

Recomendações Personalizadas de Alunos em Sistemas de Hipermedia Adaptativa Educacional usando *Data Mining*

Clarisse Celeste Cravo Matos

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Tecnologias do Conhecimento e Decisão**

Orientador: Constantino Martins

Co-orientador: Luiz Faria

Júri:

Presidente:

[Nome do Presidente, Categoria, Escola]

Vogais:

[Nome do Vogal1, Categoria, Escola]

[Nome do Vogal2, Categoria, Escola] (até 4 vogais)

Porto, Outubro 2013

Dedicatória

Dedico este trabalho à minha mãe.

Resumo

O aumento de tecnologias disponíveis na *Web* favoreceu o aparecimento de diversas formas de informação, recursos e serviços. Este aumento aliado à constante necessidade de formação e evolução das pessoas, quer a nível pessoal como profissional, incentivou o desenvolvimento área de sistemas de hipermédia adaptativa educacional - SHAE.

Estes sistemas têm a capacidade de adaptar o ensino consoante o modelo do aluno, características pessoais, necessidades, entre outros aspetos. Os SHAE permitiram introduzir mudanças relativamente à forma de ensino, passando do ensino tradicional que se restringia apenas ao uso de livros escolares até à utilização de ferramentas informáticas que através do acesso à internet disponibilizam material didático, privilegiando o ensino individualizado.

Os SHAE geram grande volume de dados, informação contida no modelo do aluno e todos os dados relativos ao processo de aprendizagem de cada aluno. Facilmente estes dados são ignorados e não se procede a uma análise cuidada que permita melhorar o conhecimento do comportamento dos alunos durante o processo de ensino, alterando a forma de aprendizagem de acordo com o aluno e favorecendo a melhoria dos resultados obtidos.

O objetivo deste trabalho foi selecionar e aplicar algumas técnicas de *Data Mining* a um SHAE, PCMAT - *Mathematics Collaborative Educational System*. A aplicação destas técnicas deram origem a modelos de dados que transformaram os dados em informações úteis e compreensíveis, essenciais para a geração de novos perfis de alunos, padrões de comportamento de alunos, regras de adaptação e pedagógicas.

Neste trabalho foram criados alguns modelos de dados recorrendo à técnica de *Data Mining* de classificação, abordando diferentes algoritmos. Os resultados obtidos permitirão definir novas regras de adaptação e padrões de comportamento dos alunos, poderá melhorar o processo de aprendizagem disponível num SHAE.

Palavras-chave: Sistemas de hipermédia adaptativa educacional, Modelação do aluno, Técnicas de *Data Mining*

Abstract

The development of technologies available on the *Web* promoted the appearance of several forms of information, resources and services. This combined with the constant need for training and development of people, on both personal and professional levels, promoted the area of Adaptive Educational Hypermedia Systems – AEHS.

These systems have the ability to adapt the kind of teaching according to the student's model, his personal characteristics, interests, needs, preferences, among other aspects. The AEHS allowed the introduction of changes in the way of teaching, from traditional education, where only school books were used, to the use of new technologies, which, through the use of the Internet, provide educational materials, emphasizing individualized learning.

The AEHS generate a large amount of data, information contained in the student model and all the data regarding the learning process of each student. These data are easily ignored and not carefully analyzed in order to improve the knowledge of the students' behavior during the teaching process, changing the form of learning according to the student and improving the results obtained.

The goal of this study was to select and apply some Data Mining techniques to an adaptive educational hypermedia system, PCMAT - Mathematics Collaborative Educational System. The use of these techniques created data models that transformed data into useful and understandable information, essential to generate new students' models, student behavior patterns, adaptation and pedagogical rules.

In this work, some models of data were created using the data mining technique of classification, using different algorithms. The results of these data models may define new adaptation rules and patterns from the behavior of students and can improve the learning process available on the AEHS.

Keywords: Adaptive Educational Hypermedia Systems, Student Model, Data Mining Techniques

Agradecimentos

Agradeço antes de tudo aos meus pais que sempre lutaram para que eu tivesse um futuro melhor, sempre me ajudaram e apoiaram a superar as fases mais difíceis.

Ao meu orientador e co-orientador de projeto, professor Constantino Martins e Luiz Faria, agradeço em especial por toda a ajuda e coordenação prestada durante a elaboração da dissertação, não esquecendo o enorme trabalho de revisão deste documento.

À professora Fátima Rodrigues pela sua disponibilidade, sugestões e ajuda prestada em relação à ferramenta *Clementine*.

A todos os meus amigos, colegas de curso e professores que me deram força e incentivaram a continuar em frente independentemente das dificuldades.

Muito obrigada, sem vocês não seria possível.

Índice

1	Introdução	1
1.1	Enquadramento	1
1.2	Motivação	2
1.3	Apresentação do tema.....	2
1.4	Objetivos.....	3
1.4.1	Objetivo geral.....	3
1.4.2	Objetivos específicos	3
1.5	Limitações	4
1.6	Estrutura do documento	4
2	Estado da arte	7
2.1	Hipermédia adaptativa	8
2.1.1	Níveis de adaptação	11
2.1.2	Técnicas de hipermédia adaptativa.....	12
2.2	Data Mining	18
2.2.1	<i>Data Mining</i> para SHA	23
2.3	Modelo de aluno	29
2.3.1	Técnicas de modelação de alunos.....	30
2.3.2	Inicialização de modelos de alunos	33
2.4	Resumo	36
3	Ferramentas de <i>Data Mining</i>.....	39
3.1	Weka.....	40
3.2	Oracle Data Mining	41
3.3	Enterprise Miner	42
3.4	Clementine	43
3.5	Estudo comparativo	43
3.6	Resumo	48
4	PCMAT - Mathematics Collaborative Educational System.....	51
4.1	Compreensão do negócio	52

4.2	Compreensão dos dados.....	54
4.3	Resumo	67
5	Modelo de dados.....	69
5.1	Preparação de dados.....	70
5.2	Exploração gráfica.....	74
5.2.1	Caracterização da amostra de dados	74
5.2.2	Classificações obtidas na avaliação de diagnóstico por sexo	75
5.2.3	Classificações obtidas na avaliação final por sexo	79
5.2.4	Comparação entre os resultados obtidos na avaliação inicial e final	83
5.2.5	Comparação das classificações nas avaliações por estilo de aprendizagem	86
5.2.6	Análise de outros fatores na distribuição das classificações	97
5.2.7	Classificações por alunos médios	100
5.3	Classificação.....	116
5.4	Avaliação dos modelos.....	120
5.5	Resumo	122
6	Conclusões	125
6.1	Resultados obtidos	125
6.2	Trabalho futuro	126
7	Bibliografia.....	129
	Anexos	133
	Anexo A	133
	Anexo B	137
	Anexo C	141
	Anexo D	153
	Anexo E	157

Lista de Figuras

Figura 1 - Ciclo clássico "Modelo de utilizador - adaptação"	10
Figura 2 - Visão geral das etapas do processo KDD	19
Figura 3 - Tipos de <i>Data Mining</i>	20
Figura 4 - Excerto de uma base de dados de controlo de ambientes à distância.....	24
Figura 5 - <i>Clustering</i> por idade e aproveitamento	27
Figura 6 - Modelo de sobreposição (<i>overlay</i>)	31
Figura 7 - Modelo de perturbação (<i>Buggy</i>).....	31
Figura 8 - Distribuição de alunos por sexo	74
Figura 9 - Distribuição de alunos por tipo de ensino	74
Figura 10 - Distribuição dos alunos por grupo	75
Figura 11 - Classificação do teste diagnóstico dos alunos por sexo.....	75
Figura 12 - Estilo de aprendizagem inicial dos alunos por sexo.....	75
Figura 13 - Classificação do conceito A1 no teste diagnóstico por sexo.....	76
Figura 14 - Classificação do conceito A2 no teste diagnóstico por sexo.....	76
Figura 15 - Classificação do conceito A3 no teste diagnóstico por sexo.....	77
Figura 16 - Classificação do conceito A4 no teste diagnóstico por sexo.....	77
Figura 17 - Classificação do conceito B1 no teste diagnóstico por sexo.....	77
Figura 18 - Classificação do conceito B2 no teste diagnóstico por sexo.....	78
Figura 19 - Classificação do conceito B3 no teste diagnóstico por sexo.....	78
Figura 20 - Classificação do conceito B4 no teste diagnóstico por sexo.....	78
Figura 21 - Classificação do conceito B5 no teste diagnóstico por sexo.....	79
Figura 22 - Classificação do teste final por sexo	79
Figura 23 - Estilo de aprendizagem final dos alunos por sexo	80
Figura 24 - Classificação do conceito A1 no teste final por sexo	80
Figura 25 - Classificação do conceito A2 no teste final por sexo	80
Figura 26 - Classificação do conceito A3 no teste final por sexo	81
Figura 27 - Classificação do conceito A4 no teste final por sexo	81
Figura 28 - Classificação do conceito B1 no teste final por sexo	81
Figura 29 - Classificação do conceito B2 no teste final por sexo	82
Figura 30 - Classificação do conceito B3 no teste final por sexo	82
Figura 31 - Classificação do conceito B4 no teste final por sexo	82

Figura 32 - Classificação do conceito B5 no teste final por sexo.....	83
Figura 33 - Classificação no teste final por sexo e estilo de aprendizagem final	84
Figura 34 - Relações entre os conceitos no teste diagnóstico	85
Figura 35 - Relações entre os conceitos no teste final.....	86
Figura 36 - Classificação inicial do A1 por estilo de aprendizagem.....	87
Figura 37 - Classificação final do A1 por estilo de aprendizagem	87
Figura 38 - Classificação inicial do A2 por estilo de aprendizagem.....	88
Figura 39 - Classificação final do A2 por estilo de aprendizagem	88
Figura 40 - Classificação inicial do A3 por estilo de aprendizagem.....	89
Figura 41 - Classificação final do A3 por estilo de aprendizagem	90
Figura 42 - Classificação inicial do A4 por estilo de aprendizagem.....	90
Figura 43 - Classificação final do A4 por estilo de aprendizagem	91
Figura 44 - Classificação inicial do B1 por estilo de aprendizagem.....	92
Figura 45 - Classificação final do B1 por estilo de aprendizagem	92
Figura 46 - Classificação inicial do B2 por estilo de aprendizagem	93
Figura 47 - Classificação final do B2 obtida pelos alunos e por estilo de aprendizagem.....	93
Figura 48 - Classificação inicial do B3 por estilo de aprendizagem	94
Figura 49 - Classificação final do B3 por estilo de aprendizagem	94
Figura 50 - Classificação inicial do B4 por estilo de aprendizagem.....	95
Figura 51 - Classificação final do B4 por estilo de aprendizagem	95
Figura 52 - Classificação inicial do B5 por estilo de aprendizagem.....	96
Figura 53 - Classificação final do B5 obtida pelos alunos e por estilo de aprendizagem.....	97
Figura 54 - Classificação no teste diagnóstico por tipo de ensino	98
Figura 55 - Classificação no teste final por tipo de ensino.....	99
Figura 56 - Classificação no teste diagnóstico por grupo.....	99
Figura 57 - Classificação no teste final por grupo	100
Figura 58- Distribuição dos alunos médios por sexo.....	101
Figura 59 - Distribuição dos alunos médios por grupo.....	101
Figura 60 - Distribuição dos alunos médios por estilo de aprendizagem inicial	101
Figura 61 - Distribuição dos alunos médios por estilo de aprendizagem final	102
Figura 62 - Relações entre os conceitos na avaliação inicial dos alunos médios.....	102
Figura 63 - Relações entre os conceitos na avaliação final dos alunos médios	103
Figura 64 - Classificação inicial dos alunos médios por sexo	103

Figura 65 - Classificação final dos alunos médios por sexo	104
Figura 66 - Classificação inicial do A1 nos alunos médios e por estilo de aprendizagem inicial	105
Figura 67 - Classificação final do A1 nos alunos médios e por estilo de aprendizagem final..	106
Figura 68 - Classificação inicial do A2 nos alunos médios e por estilo de aprendizagem inicial	107
Figura 69 - Classificação final do A2 nos alunos médios e por estilo de aprendizagem final..	107
Figura 70 - Classificação inicial do A3 nos alunos médios e por estilo de aprendizagem inicial	108
Figura 71 - Classificação final do A3 nos alunos médios e por estilo de aprendizagem final..	108
Figura 72 - Classificação inicial do A4 nos alunos médios e por estilo de aprendizagem inicial	109
Figura 73 - Classificação final do A4 nos alunos médios e por estilo de aprendizagem final..	109
Figura 74 - Classificação inicial do B1 nos alunos médios e por estilo de aprendizagem inicial	110
Figura 75 - Classificação final do B1 nos alunos médios e por estilo de aprendizagem final ..	111
Figura 76 - Classificação inicial do B2 nos alunos médios e por estilo de aprendizagem inicial	111
Figura 77 - Classificação final do B2 nos alunos médios e por estilo de aprendizagem final ..	112
Figura 78 - Classificação inicial do B3 nos alunos médios e por estilo de aprendizagem inicial	113
Figura 79 - Classificação final do B3 nos alunos médios e por estilo de aprendizagem final ..	113
Figura 80 - Classificação inicial do B4 nos alunos médios e por estilo de aprendizagem inicial	114
Figura 81 - Classificação final do B4 nos alunos médios e por estilo de aprendizagem final ..	114
Figura 82 - Classificação inicial do B5 nos alunos médios e por estilo de aprendizagem inicial	115
Figura 83 - Classificação final do B5 nos alunos médios e por estilo de aprendizagem final ..	115
Figura 84 - Taxa de confiança do modelo de classificação de dados recorrendo ao C5.0.....	116
Figura 85 - Importância dos atributos segundo o modelo de dados recorrendo ao C5.0.....	117
Figura 86 - Rede bayesiana	135
Figura 87 - Exemplo de uma rede semântica.....	137
Figura 88 - Elementos de uma rede semântica.....	139

Figura 89 - Interface do <i>Clementine</i>	141
Figura 90 - Nós Origem.....	142
Figura 91 - Nós de Registos	143
Figura 92 - Nós de Atributos.....	144
Figura 93 - Nós Gráficos	146
Figura 94 - Nós Aprendizagem	147
Figura 95 - Nós Saída	147
Figura 96 - Super-nó	148
Figura 97 - Rede neuronal	149
Figura 98 - Árvore de decisão.....	150
Figura 99 - Rede de <i>Kohonen</i>	151

Lista de Tabelas

Tabela 1 - Dados sobre um módulo de um curso à distância	24
Tabela 2 - Dados sobre o cumprimento das atividades de um curso pelos alunos.....	25
Tabela 3 - Dados para <i>clustering</i>	27
Tabela 4 - Tempo de permanência X Aproveitamento	28
Tabela 5 - Metodologia SEMMA	42
Tabela 6 - Plataformas	44
Tabela 7 – Automação	45
Tabela 8 - Técnicas de DM	46
Tabela 9 – Algoritmos	46
Tabela 10 – Visualização	47
Tabela 11 – Outros critérios.....	48
Tabela 12- Vantagens e desvantagens de ferramentas DM	49
Tabela 13 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º A.....	55
Tabela 14 - Classificações do teste diagnóstico do 7º A	56
Tabela 15 - Classificações do teste final do 7º A.....	56
Tabela 16 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º A	57
Tabela 17 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º B.....	57
Tabela 18 - Classificações do teste diagnóstico do 7º B	58
Tabela 19 - Classificações do teste final do 7º B.....	59
Tabela 20 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º B	59
Tabela 21 - Estilos de aprendizagem inicial dos alunos do grupo experimental do 7º F.....	60
Tabela 22 - Classificações do teste diagnóstico do 7º F.....	60
Tabela 23 - Classificações do teste final do 7º F	61
Tabela 24 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º F.....	62
Tabela 25 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º G.....	62
Tabela 26 - Classificações do teste diagnóstico do 7º G	63
Tabela 27 - Classificações do teste final do 7º G.....	64
Tabela 28 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º G	64
Tabela 29 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 8º F	65
Tabela 30 - Classificações do teste final do 8º F	66
Tabela 31 - Estilo de aprendizagem final dos alunos do grupo experimental do 8º F.....	66

Tabela 32 - Número total de alunos com dados em falta por ano e turma.....	73
Tabela 33 - Classificação total por conceito obtida pelos alunos	83
Tabela 34 - Mudança de estilo de aprendizagem no processo de aprendizagem.....	98
Tabela 35 - Classificação total por conceito obtida pelos alunos médios.....	104
Tabela 36 - Matriz confusão	121
Tabela 37 - Matriz confusão do modelo criado.....	121
Tabela 38 - Tabela de probabilidades condicionadas	134
Tabela 39 - Características do PCMAT.....	157
Tabela 40 - Grau de utilidade do PCMAT	158
Tabela 41 - Grau de dificuldade do PCMAT.....	158
Tabela 42 - Frequência de uso do PCMAT fora das aulas	158
Tabela 43 - Recomendação do uso do PCMAT noutras disciplinas.....	159

Acrónimos

AIX	<i>Advanced Interactive eXecutive</i>
API	<i>Application Programming Interface</i>
ARFF	<i>Attribute Relation File Format</i>
DB2	Sistema de gestão de bases de dados relacionais
DM	<i>Data Mining</i>
EaD	Ensino à distância
EM	<i>Enterprise Miner</i>
GPL	<i>General Public License</i>
HA	Hipermedia Adaptativa
IBM	<i>International Business Machines</i>
IM	<i>Intelligent Miner</i>
KDD	<i>Knowledge Discovery in Database</i>
ODBC	<i>Open Database Connectivity</i>
ODM	<i>Oracle Data Mining</i>
ROC	<i>Receiver Operating Characteristic</i>
SAS	<i>Statistical Analysis System</i>
SEMMA	<i>Sampling, Exploration, Modification, Modeling and Assessment</i>
SGBD	Sistema de Gestão de Bases de Dados
SGI	<i>Silicon Graphics Inc</i>
SHA	Sistema de Hipermedia Adaptativa
SHAE	Sistema de Hipermedia Adaptativa Educacional
SPSS	<i>Statistical Package for the Social Sciences</i>
VARK	<i>Visual, Auditory, Read/Write, Kinesthetic</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

1 Introdução

"Em todas as coisas o sucesso depende de uma preparação prévia, e sem tal preparação o falhanço é certo."

Confúcio

1.1 Enquadramento

Atualmente, o mundo é cada vez mais rigoroso e competitivo. Esta realidade exige que as pessoas sejam polivalentes a nível profissional e que estejam bem informadas a respeito da sua área de atuação e de outros setores. Nos últimos anos do século XX, o crescimento da internet permitiu o aumento da diversidade de informação, recursos e serviços disponíveis na *Web* (Alves et al., 2003). Na busca da constante formação e evolução dos indivíduos a internet tem se mostrado como uma alternativa (Bugay, 1999).

Na última década, tem se assistido a uma mudança a nível dos ambientes e ferramentas para a educação. O ensino tradicional não se restringe exclusivamente à utilização de livros escolares, as salas de aulas são providas de equipamento informático com acesso à internet e ferramentas de hipermédia com material didático.

Alguns sistemas de hipermédia são exemplos de sucesso para o ensino, permitindo a combinação adequada de várias formas de média e sua interatividade, estimulando o desenvolvimento do aluno (Rada, 1995).

A visão de um ensino individualizado, proporciona que o aluno tenha um progresso de acordo com os seus interesses, objetivos e a seu próprio ritmo. A hipermédia adaptativa dá resposta a este tipo de ensino (Brusilovsky, 2004).

Os sistemas de hipermédia adaptativa apresentam perfis para cada utilizador e aplicam as definições do perfil na adaptação de diversos aspetos do sistema, segundo as necessidades, desejos e preferências destes utilizadores. Esta adaptação ocorre através da apresentação ou omissão de partes de conteúdo e/ou hiperligações. Estes sistemas podem ser bastante úteis para qualquer tipo de aplicação que seja utilizada por pessoas com metas e conhecimentos diferentes (Brusilowsky, 1996).

Os sistemas de hipermédia adaptativa educacional produzem grande volume de dados que facilmente são ignorados, sem existir qualquer preocupação com a sua análise. Recorrendo a técnicas de *Data Mining*, os dados gerados pelos sistemas, informação dos alunos, características pessoais, conhecimentos, objetivos, preferências, entre outros, podem ser transformados em informações úteis e compreensíveis, essenciais para a geração de novos perfis de alunos, regras de adaptação e pedagógicas, padrões de comportamento de alunos e auxiliarem o interação entre o professor e aluno (Silva et al, 2001).

1.2 Motivação

A necessidade de contínua evolução e formação está, cada vez mais, presente na vida diária das pessoas, quer a nível profissional quer pessoal. Para além desta necessidade é verificado que cada indivíduo apresenta ritmos, estilos e horários de aprendizagem distintos, promovendo o ensino individualizado. O aumento de informação, recursos e serviços disponíveis na *Web* proporcionou o aparecimento de sistemas de hipermédia adaptativos educacionais, sistemas que são capazes de adaptar a aprendizagem consoante o histórico do aluno, preferências, necessidades, desejos, entre outros.

Estes sistemas geram grande volume de dados e facilmente não é atribuída a devida importância aos mesmos. É nesse sentido que surge a motivação de realizar uma análise cuidada em relação aos dados gerados, tendo como principal objetivo a otimização do processo de aprendizagem, promovendo a melhoria dos resultados obtidos pelos alunos.

1.3 Apresentação do tema

O tema deste trabalho está relacionado com um sistema de hipermédia adaptativa educacional, PCMAT - *Mathematics Collaborative Educational System* (Martins et al., 2008). Este sistema consiste numa plataforma adaptativa de aprendizagem baseada na avaliação progressiva de Matemática, disponibilizando os conteúdos e atividades consoante o perfil de

cada aluno. A implementação deste sistema está relacionada com os processos de aprendizagem da Matemática, nomeadamente no tópico "Proporcionalidade Direta". Esta unidade está incluída no programa de Matemática do 6º ano e também é abordado no 7º ano de escolaridade. Esta implementação foi utilizada com alunos cujas idades estão compreendidas entre os 12 e 14 anos, em duas escolas básicas, escola EB2/3 Dr. Ferreira de Almeida, de Santa Maria da Feira e a escola EB2/3 S. Lourenço, de Ermesinde (Martins et al., 2008).

O PCMAT armazena os dados relativos ao perfil de cada aluno e toda a informação relativa ao processo de aprendizagem. Este estudo pretende utilizar técnicas de *Data Mining* para efetuar uma análise cuidada aos dados gerados e transformá-los em informação útil, descobrindo fatos significativos e gerais, padrões e regras, favorecendo e promovendo a melhoria e adaptação do ensino.

1.4 Objetivos

1.4.1 Objetivo geral

O objetivo deste trabalho é selecionar e aplicar técnicas de *Data Mining* a serem usadas no modelo de aluno para pré-definição de perfil de alunos e regras de adaptação e pedagógicas a aplicar por cada perfil.

1.4.2 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Seleção de técnicas de *Data Mining* - Neste ponto é pretendido efetuar um estudo de algumas técnicas de *Data Mining* (DM), nomeadamente, classificação; regressão; agrupamento (*clustering*); associação de regras. Esta análise permitirá selecionar as técnicas a usar neste trabalho de acordo com os objetivos pretendidos e os dados provenientes do PCMAT - *Mathematics Collaborative Educational System*. O uso de técnicas de DM permitirá realizar análises cuidadas aos dados disponíveis pelo PCMAT, generalizar o comportamento do aluno e não esquecer as características individuais de cada um.

- Seleção da ferramenta de DM - Nesta seção é pretendido realizar um estudo comparativo entre algumas ferramentas de DM, destacando as suas vantagens e desvantagens. Após a análise dos resultados obtidos da comparação das ferramentas é selecionada a ferramenta mais adequada ao estudo em questão, permitindo efetuar uma análise cuidada aos dados armazenados pelo PCMAT e extrair regras e padrões de comportamento de alunos.
- Elaboração do modelo de dados - Neste ponto é pretendido a criação de um modelo de dados recorrendo a técnicas de DM, extraíndo regras e padrões de comportamento de alunos. O modelo de dados criado poderá melhorar o processo de aprendizagem disponibilizado pelo PCMAT.

1.5 Limitações

Inicialmente, uma das limitações sentidas foi a dificuldade de recolha de informação específica relativa a algumas ferramentas de *Data Mining*, uma vez que a informação encontrada era de carácter comercial e não técnico como se desejava. A solução utilizada para ultrapassar este obstáculo foi reduzir o conjunto de características técnicas analisadas por cada ferramenta de DM.

Por outro lado, outra dificuldade enfrentada consistiu no reduzido conjunto de dados que representava a amostra fornecida pela plataforma PCMAT - *Mathematics Collaborative Educational System*. Esta pequena amostra tornou a construção do modelo de dados *Data Mining* numa tarefa árdua, chegando a criar situações de sobre ajustamento de dados. Para solucionar este contratempo foram utilizados métodos específicos que operam com pequenas quantidades de dados.

1.6 Estrutura do documento

Este trabalho está organizado em seis capítulos e cada um pretende abordar as informações relevantes ao respetivo tema, sendo efetuada uma introdução, síntese e contendo anexos inerentes a cada tema.

O capítulo 1, "Introdução", apresenta o tema, os objetivos, motivação, limitações e organização do trabalho.

O capítulo 2, "Estado de arte", apresenta os principais conceitos bibliográficos relacionados com hipermédia adaptativa, *Data Mining* e modelação de alunos.

O capítulo 3, "Ferramentas de DM", apresenta algumas ferramentas de *Data Mining* mais utilizadas no mercado, realçando as suas vantagens e limitações. Também exhibe um estudo comparativo entre as ferramentas apresentadas e justifica a adoção da ferramenta usada neste estudo.

O capítulo 4, "PCMAT - *Mathematics Collaborative Educational System*", apresenta a plataforma PCMAT, ferramenta colaborativa e adaptativa baseada na avaliação progressiva de conteúdos matemáticos no ensino básico. Neste capítulo são ainda apresentados os resultados obtidos relativamente à avaliação da adequação dos modelos e funcionalidades do PCMAT.

O capítulo 5, "Modelo de dados", descreve o processo de preparação dos dados, exploração gráfica e elaboração do modelo de dados e a sua avaliação.

O capítulo 6, "Conclusões", exhibe as conclusões e aponta algumas direções para desenvolvimento de trabalhos futuros.

2 Estado da arte

“Aprenda com o ontem, viva o hoje, tenha a esperança no amanhã. O importante é não parar de questionar.”

Albert Einstein

Este trabalho é iniciado com o estudo da área de hipermédia, onde é abordado o aparecimento da multimédia associado ao crescimento da internet, sendo um incentivo para diversas pesquisas nesta área.

A grande diversidade de recursos disponíveis na *Web* originou a necessidade de personalizar os recursos consoante cada utilizador. Cada pessoa possui as suas necessidades e/ou exigências e, como tal, as suas expectativas relativas a um recurso são distintas de indivíduo para indivíduo (Ulbricht et al., 2007). Assim, surge a necessidade de criar grupos de utilizadores com características semelhantes (perfis de utilizadores) e consoante cada modelo exibir informações distintas. Esta necessidade pode ser colmatada com o recurso da hipermédia adaptativa (Brusilovsky, 2004).

A hipermédia adaptativa é centrada em duas áreas: hipermédia e modelação de utilizador. A segunda é responsável pela personalização ou adaptação consoante as necessidades de cada utilizador. Um sistema de hipermédia adaptativa possui um modelo que armazena objetivos, conhecimentos, experiências e preferências de cada utilizador e usa-o dinamicamente, com o objetivo de adaptar o próprio sistema às exigências do utilizador (Brusilovsky, 1996).

Apesar de diversas áreas apresentarem sistemas de hipermédia adaptativa, a área de eleição deste trabalho é a educação. Um sistema de hipermédia adaptativa educacional possui a capacidade de adaptar e apoiar o processo de aprendizagem de forma personalizada a cada

aluno, mediante a adaptação do seu comportamento. Estes sistemas têm em consideração que as pessoas são diferentes, aprendem em ritmos diferentes e apresentam necessidades de aprendizagem distintas (Oliveira et al., 2004). Os SHAE (Sistema de Hipermedia Adaptativa Educacional) exibem um modelo de utilizador que corresponde a um modelo de estudante, designado modelo de aluno, e servirá para invocar e/ou definir estratégias pedagógicas (Silva et al., 1998).

Ao longo deste trabalho também são analisados os diferentes níveis de adaptação: conteúdo, navegação e apresentação, bem como as diversas técnicas de adaptação para cada nível. Não esquecendo de abordar que o *Data Mining* consiste numa das principais fases do processo de Descoberta do Conhecimento em Base de Dados, sendo descritas cada uma das seis etapas do processo *Data Mining*.

A principal vantagem do uso desta metodologia consiste na obtenção de conhecimento, informações úteis, a partir de grandes volumes de dados provenientes de distintas fontes de dados, sendo importante para a tomada de decisões (Fayyad et al., 1996).

Os sistemas de hipermedia adaptativa educacional podem usufruir desta vantagem se recorrerem às técnicas de *Data Mining* para analisar os dados gerados por estes sistemas. Estes dados são relativos a informações dos alunos, contendo características pessoais, conhecimentos, objetivos, preferências, entre outros. Desta aliança, são obtidas informações úteis e compreensíveis, essenciais para a geração de novos perfis de alunos, regras de adaptação e pedagógicas, padrões de comportamento de alunos e auxílio no processo de aprendizagem (Silva et al., 2001).

2.1 Hipermedia adaptativa

A hipermedia é uma tecnologia que engloba dois recursos: hipertexto e multimédia (Martin, 1992). O primeiro consiste em texto eletrónico, permitindo ao utilizador interagir com os diversos conteúdos conectados entre si. O segundo representa uma tecnologia baseada no uso do computador de maneira multissensorial, integrada e intuitiva e interativa, facultando a combinação de diferentes formatos de informação, texto, imagens, animações, vídeos e sons, entre outros (Chaves, 2006).

Em meados da década de 90, surgiram os primeiros trabalhos relativos a Hipermedia Adaptativa (HA), que se centravam em duas áreas principais: hipermedia e modelação do utilizador. Estes trabalhos incentivaram o estudo e exploração desta área (Brusilovsky, 1996).

A HA é dedicada ao estudo e desenvolvimento de sistemas, arquiteturas, métodos e técnicas capazes de promover a adaptação de hiperdocumentos às necessidades, preferências e desejos dos seus utilizadores (Palazzo, 2000).

A hipermedia adaptativa possibilita a criação de sistemas que, consoante o modelo de cada utilizador, adaptam os diferentes aspetos visíveis dos próprios sistemas: conteúdo, apresentação e navegação, segundo as exigências e/ou necessidades do utilizador (Brusilovsky, 1996).

Estes sistemas promovem a adaptação de conteúdo e navegação. Na adaptação de conteúdo pode ocorrer o destaque, ocultação e/ou reorganização do material hipermedia de acordo com o modelo do utilizador. A adaptação de navegação consiste em disponibilizar ou ocultar ligações do sistema consoante os objetivos do utilizador. O recurso à HA é aconselhável quando o hiperdocumento tende a ser usado por um elevado número de utilizadores com perfis diferentes (Brusilovsky, 1996).

Para que os sistemas de hipermedia adaptativa (SHA) sejam dotados da capacidade de adaptação é necessário traçar o modelo de cada utilizador, definir características individuais do utilizador. Esta informação é representada por um modelo, designado modelo de utilizador, que pretende descrever os objetivos, conhecimentos, historial, experiência, preferências e necessidades de cada utilizador (Palazzo, 2004). A construção deste modelo pode ser realizada por várias técnicas que serão abordadas na secção 2.3.1.

Para que um sistema seja considerado um SHA deve satisfazer, no mínimo, três critérios básicos (Palazzo, 2004):

- Ser um sistema hipertexto ou hipermedia;
- Possuir um modelo de utilizador;
- Poder adaptar o sistema hipermedia a partir do modelo de utilizador.

Com o intuito de facilitar a compreensão de um SHA, Brusilovsky sugeriu um esquema que ilustra o ciclo clássico de um SHA (Brusilovsky, 1996).

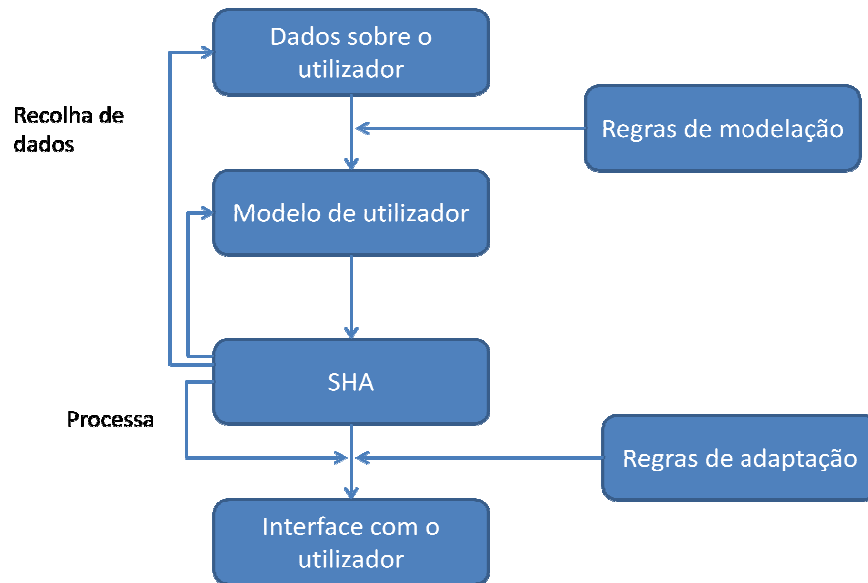


Figura 1 - Ciclo clássico "Modelo de utilizador - adaptação"

A figura 1 representa os principais processos de um SHA: a recolha de informações sobre o comportamento do utilizador, a atualização do seu modelo e, com os novos dados, a geração do efeito de adaptação de conteúdo e navegação apresentados ao utilizador (Brusilovsky, 1996).

Durante a navegação no SHA, o comportamento do utilizador é observado e são recolhidas e armazenadas informações sobre o utilizador (Wu et al., 2004). Os SHA pretendem antecipar as expectativas dos utilizadores com base nos modelos que representam o seu modelo.

Atualmente, os SHA estão dispersos por diversos setores, nomeadamente, educação, comércio eletrónico e *marketing*. Com a crescente necessidade de personalizar as aplicações, destacam-se alguns exemplos (Palazzo, 2004):

- Sistemas educacionais;
- *Workgroups*, sistemas colaborativos e corporativos;
- *Marketing* e pesquisa de opinião;
- Assistentes pessoais;
- Comércio electrónico;
- Sistemas de recuperação e filtragem de informações.

O trabalho em questão refere-se ao estudo de SHA na área da educação, sendo assim são abordados os principais conceitos de um SHA vocacionado para a educação. Os sistemas de hipermédia adaptativa educacional (SHAE) possuem a capacidade de adaptação e visam apoiar o processo de aprendizagem de forma personalizada a cada aluno, mediante a adaptação do seu comportamento (Oliveira et al., 2004).

Uma das principais tarefas de um SHAE consiste na adaptação de conteúdo e no caminho de navegação, segundo as condições didático-pedagógicas da aplicação e as características de cada aluno, definidas no modelo de aluno (Silva et al., 1998).

A adaptação final de conteúdos e navegação são o resultado de uma série de regras definidas no modelo de aluno no sistema. A forma de estruturação do SHAE é muito importante, não apenas para o desenvolvimento destes sistemas mas também para o processo de adaptação. A complexidade de um SHAE reside exatamente em definir quais são as suas regras de adaptação e pedagógicas, de forma a serem significativas em termos educacionais e em como estruturá-las, sendo ainda mais complexo o processo de modelação (Reis, 2006).

2.1.1 Níveis de adaptação

O conceito de adaptação pode ser separado em adaptação (*adaptativity*) e adaptabilidade (*adaptability*). A separação destes conceitos permite distinguir claramente sistemas adaptáveis (sistemas hipermédia personalizáveis) e os sistemas hipermédia adaptativa. Nos dois tipos de sistemas o utilizador desempenha um papel central e o objetivo principal é oferecer um sistema personalizado. Apesar disso, a diferença entre ambos consiste no responsável pela adaptação (Koch, 2000).

Um sistema de hipermédia adaptável permite que o utilizador seja o responsável direto pela mudança de diversos aspetos visíveis do próprio sistema. Esta mudança acontece tendo em conta as preferências do utilizador, que foram recolhidas através do preenchimento de um formulário aquando o registo de um utilizador. Sempre que possível o utilizador tem a possibilidade de alterar as suas preferências a qualquer momento (De Bra, 1998).

Um sistema de hipermédia adaptativa é adaptado, de acordo com os dados armazenados no modelo de utilizador, sem intervenção direta do utilizador. Neste tipo de sistema, o modelo de utilizador é alimentado através da observação do comportamento do utilizador durante a navegação pelo sistema, respostas a questionários, entre outros (De Bra, 1998). Estes sistemas são baseados em hipertexto (ou hipermédia) e usam o conhecimento e os objetivos do

utilizador, contidas num modelo, para modificar as ligações e/ou conteúdos, fornecendo uma instrução e/ou navegação individualizada. Por fim, permitem dotar as aplicações *Web* de um comportamento inteligente, tendo em conta que possuem a capacidade de antecipar as necessidades do utilizador e personalizar a aplicação (Koch, 2000).

A maioria dos SHA apresenta tantas características adaptativas como adaptáveis, seja pela iniciação do modelo de utilizador ou sistemas que permitem que os utilizadores modifiquem o modelo de utilizador quando considerarem conveniente (De Bra, 1998).

2.1.2 Técnicas de hipermédia adaptativa

A adaptação de um SHA pode ocorrer a nível de um conteúdo, apresentação e navegação, sendo composta por métodos e técnicas (Brusilovsky, 1996).

Os métodos e as técnicas são considerados conceitos distintos. O segundo conceito faz referência a técnicas de produção de adaptação em sistemas de hipermédia adaptativa, que fazem parte do nível de implementação. As técnicas podem ser caracterizadas por um tipo específico de conhecimento e um algoritmo correspondente de adaptação. Os métodos são considerados generalizações de técnicas de adaptação existentes. O mesmo método pode ser implementado por diferentes técnicas. Ao mesmo tempo, algumas técnicas são usadas para implementar diversos métodos utilizando a mesma representação do conhecimento (Brusilovsky, 1996).

A adaptação de conteúdo tem como objetivo personalizar os conteúdos para um determinado utilizador consoante os conhecimentos e objetivos definidos no seu modelo de utilizador. Por exemplo, um utilizador avançado recebe informações mais detalhadas, enquanto que um novato tem acesso a explicações adicionais sobre o determinado assunto (Palazzo, 2000).

Os métodos de adaptação de conteúdos permitem aumentar a usabilidade de uma aplicação para um determinado grupo de utilizadores com diversos graus de experiência sobre o assunto. Estes métodos permitem fornecer conteúdos adicionais, comparativos ou alternativos ou mesmo ocultar partes de conteúdos (Koch, 2000).

Conteúdo ou explicação adicional – é o método mais comum para adaptação de conteúdos e consiste em mostrar somente partes de informação (ocultando as restantes), de acordo com a experiência, conhecimento, interesses ou preferências do

utilizador. Por exemplo, omitir explicações detalhadas para utilizadores que dominam o assunto e apresentar maior número de detalhes a utilizadores inexperientes.

Variação de conteúdo – este método indica que mostrar ou esconder certas partes de informação nem sempre é suficiente para promover a adaptação uma vez que utilizadores diferentes podem necessitar de informações distintas. Através deste método o sistema armazena diversas versões de conteúdos, de forma a que o utilizador visualize os mais adequados ao seu modelo. Este método também é designado de variação da explicação (Brusilovsky, 2005).

Explicação requerida – consiste na ordenação dos conteúdos a serem apresentados ao utilizador, onde a primeira informação exibida é pré-requisito para a seguinte. Baseado nesta ideia, ao apresentar a explicação de um conceito, o sistema fornece a explicação de todos os conceitos requeridos para o seu entendimento.

Explicação comparativa – este método é baseado na semelhança entre os conteúdos. Quando o conteúdo exibido é similar ao conteúdo existente no domínio, uma explicação comparativa é utilizada para apresentar o novo conceito. O uso de explicações comparativas pode ser, especialmente, eficiente no ensino de linguagens de programação (Palazzo, 2004).

Classificação de fragmentos – considera o nível de conhecimento e experiência do utilizador para ordenar os fragmentos do conteúdo de modo a que a informação mais relevante para o utilizador seja exibida inicialmente (Brusilovsky, 1996).

As técnicas para adaptação de conteúdo permitem manipular o conteúdo de maneira a adaptá-lo às características do utilizador. A maioria destas técnicas são referidas a texto mas pode ser de igual forma aplicadas a conteúdos multimédia em geral (Brusilovsky, 1996).

Stretchtext – consiste em organizar o conteúdo utilizando uma variação de hipertexto, onde os conteúdos das hiperligações podem ser expandidos ou contraídos a partir de uma palavra-chave. O SHA determina qual o fragmento “expandido” e qual o “retraído” para a apresentação inicial ao utilizador de acordo com o seu modelo. Este pode decidir qual o item que deseja expandir ou retrain, uma vez que esta técnica permite tanto ao utilizador como ao sistema adaptar o conteúdo (Palazzo, 2000).

Texto condicional – todas as informações possíveis sobre um conceito são divididas em diversas partes de texto e, cada uma é associada a uma condição do nível de

conhecimento do utilizador. Quando a informação é apresentada, o sistema mostra apenas as partes onde a condição estabelecida é satisfeita (Brusilovsky, 1996).

Variação de páginas – esta é uma técnica simples que consiste em manter duas ou mais alternativas de páginas para cada conceito, descrevendo-o de maneiras diferentes, cada uma delas adaptada a cada nível de conhecimento: iniciante, intermédio e experiente. Esta técnica pode ser implementada no método de variação de conteúdo (Brusilovsky, 1996).

Abordagem baseada em frames – esta técnica permite a inclusão de toda a informação relacionada num *frame*. Os *frames* são estruturas de atributos e valores armazenados em *slots*, que podem conter explicações variantes sobre o conceito, hiperligações para outros *frames* e exemplos. Regras especiais de apresentação são empregadas para decidir quais os *slots* de um determinado *frame* devem ser exibidos ao utilizador e qual a ordem específica (Brusilovsky, 1996).

A navegação adaptativa tem como objetivo auxiliar o utilizador a encontrar os seus caminhos no hiperespaço, apresentando hiperligações coincidentes com os seus objetivos, conhecimentos e demais características definidas no modelo de utilizador. A adaptação consiste em mudanças da estrutura de navegação ou na forma como a estrutura é exibida ao utilizador (Brusilovsky, 1996). De seguida, são apresentados alguns exemplos de métodos que permitem adaptar a navegação:

Condução global – este método auxilia o utilizador a encontrar o caminho de navegação mais curto para a informação desejada. A condução global é o principal objetivo de suporte à navegação adaptativa em sistemas de recuperação de informações e sistemas de ajuda *online*. No caso de sistemas educacionais, onde o objetivo é aprender, é necessário capturar, acelerar ou retardar tópicos em função do desempenho do utilizador (Brusilovsky, 1996).

Condução local – tem como objetivo auxiliar o utilizador em apenas um passo da navegação, por exemplo, encontrar a “melhor” hiperligação a partir do nó atual considerando as suas preferências, conhecimento e experiência.

Orientação global – este método auxilia o utilizador na estrutura do hiperespaço e seu posicionamento nele. Os SHA recorrem a marcos visuais e mapas globais para ajudar o utilizador a localizar-se em relação ao contexto global. Os SHA oferecem um suporte maior pela aplicação de técnicas de ocultação e anotação, o que pode ocorrer de

forma sistemática, contemplando o modelo do utilizador. O método mais usado consiste em aumentar, gradualmente o número de hiperligações disponíveis à medida que o conhecimento do utilizador acerca do domínio aumenta (Koch, 2000).

Orientação local – auxilia o utilizador a distinguir entre as diferentes posições de navegação e a atual, ajudando o utilizador a seguir a hiperligação apropriada. Normalmente, implementa duas formas distintas:

- Por meio de informação adicional sobre os nós, que podem ser acedidos a partir do nó corrente;
- Pela limitação das oportunidades de navegação, visando com isso evitar o fenómeno denominado sobrecarga cognitiva, onde o excesso de informação impede ou dificulta a tomada segura de decisão na navegação.

Existem dois métodos de orientação local específicos que são utilizados em hipermédia educacional. O primeiro consiste na ocultação de nós que o aluno não está preparado para aprender, enquanto o segundo esconde as hiperligações para nós que pertencem a objetivos educacionais de outras lições, estando fora dos objetivos da lição atual (Wu et al., 2001).

Visualização personalizada – consiste na geração e atualização de visualizações personalizadas do hiperespaço, sendo utilizados agentes responsáveis por encontrar as hiperligações apropriadas ao utilizador, mantendo desta maneira a visualização personalizada. Estes agentes executam pesquisas regulares na *Web*, procurando novas hiperligações relevantes para cada utilizador e verificam a atualização ocorrida nas hiperligações recolhidas anteriormente.

As técnicas para navegação adaptativa permitem manipular as hiperligações com o propósito de adaptar a navegação dinamicamente às características do utilizador (Palazzo, 2000). Caracterizam:

Orientação direta – disponibiliza uma hiperligação para a página que o sistema considera como sendo a apropriada para o utilizador seguir. Para oferecer orientação direta o sistema pode destacar visualmente a hiperligação para o melhor nó, ou apresentar uma hiperligação adicional, normalmente denominada “*next*”, que é conectada ao melhor nó selecionado pelo sistema. Esta técnica deve responder à questão: “Em cada ponto, qual é o melhor caminho?” (Brusilovsky, 1996).

Classificação adaptativa de hiperligações – ao contrário de disponibilizar uma única hiperligação, esta técnica permite apresentar uma lista de hiperligações ordenadas de forma decrescente de importância para o utilizador. A desvantagem da ordenação adaptativa é que, cada vez que o utilizador entra na mesma página, a ordem das hiperligações pode ser diferente de acordo com o estado atual do modelo de utilizador, podendo gerar alguma confusão a nível de orientação do utilizador. A classificação adaptativa deve ser aplicada de forma limitada, pois funciona bem apenas em hiperligações não contextuais. O uso desta técnica em índices e tabelas de conteúdos é difícil e é considerado desaconselhável para hiperligações contextuais e mapas. A técnica de navegação adaptativa é muito utilizada para recuperar informação em sistemas de documentação *online* e hipermédia para o ensino. Esta técnica permite definir: “Em que ordem as hiperligações devem ser exibidas?” (Brusilovsky, 1996).

Ocultação adaptativa de hiperligações – para evitar um excesso de hiperligações disponíveis ao utilizador, o SHA oculta hiperligações de informações que não são relevantes. O uso desta técnica permite poupar o utilizador, reduzindo a sua carga cognitiva. A ocultação de hiperligações pode ser aplicada tanto a hiperligações contextuais como a não contextuais, índices ou mapas. Esta técnica tende a ser mais transparente ao utilizador, produzindo uma apresentação mais estável do que a criada com a técnica de classificação adaptativa. Esta técnica deve definir: “Quais as hiperligações que não devem ser exibidas?” (Brusilovsky, 1996).

Anotação adaptativa de hiperligações – as hiperligações são comentadas para indicar a sua importância, ou seja, as hiperligações apresentam um aspeto visível diferente para mostrar a relevância do destino. Isto pode ser feito sob a forma de texto ou de indicadores visuais, como ícones especiais, cores ou tamanhos dos caracteres. A anotação de “sinal de trânsito” é bastante usada, onde os ícones vermelho, amarelo e verde são mostrados junto à hiperligação para indicar o grau de conveniência desta. Esta técnica possibilita responder às seguintes questões: “Como agregar mais informação às hiperligações? Que hiperligações interessantes podem ser geradas?”.

Adaptação de mapas – esta técnica compreende diversas formas de adaptação de mapas de hipermédia global e local mostrados ao utilizador, aplicados à visualização gráfica da estrutura de navegação (hiperligações). Este mapa pode ser filtrado adaptativamente para apresentar uma visão geral das partes do hiperdocumento

relevantes ao utilizador. Esta técnica possibilita responder à seguinte questão: “Como apresentar mapas e índices?”.

A apresentação adaptativa tem como principal função ajustar a configuração da estrutura de navegação às preferências visuais ou necessidades do utilizador. As mudanças do aspeto da interface, geralmente acontecem em conjunto com as adaptações de conteúdo (Koch, 2000). De acordo com Rossi e Koch, são os seguinte (Koch et al., 2004):

Multi-linguagem – este método permite adaptar o sistema para a linguagem desejada do utilizador, podendo também ser dependente do contexto.

Variações de *layout* – este método inclui alterações necessárias e possíveis numa apresentação, por exemplo: cores, tipos e tamanho de fonte, tamanho máximo das imagens, orientação de texto, ordenação de fragmentos de conteúdo, etc.

A apresentação adaptativa é o termo geral para todas as técnicas que adaptam o conteúdo de hipermédia de acordo com o modelo de utilizador, distinguindo apresentação adaptativa de texto e apresentação multimédia adaptativa (Brusilovsky, 1996). A primeira é referente ao conteúdo textual das páginas que pode ser modificado em função do modelo de utilizador. Termos técnicos podem ser evitados para os iniciantes e explicações adicionais fornecidas. Do mesmo modo, detalhes adicionais podem ser fornecidos a utilizadores experientes que estão capacitados a compreendê-los. A segunda, apesar do termo sugerir que a multimédia contenha elementos que podem ser adaptados a cada utilizador, o conteúdo de uma animação, áudio ou fragmento de vídeo dificilmente pode ser adaptado (Bugay, 2006).

A apresentação adaptativa de texto exhibe três técnicas para a criação de texto adaptativo (Bugay, 2006):

Variações de páginas – podem existir versões diferentes de uma determinada página. Esta técnica ajusta o estereótipo do sistema de hipermédia adaptativa, nas quais a diversificação é limitada. Um estereótipo consiste na representação de “algo” segundo uma determinada classificação, se um utilizador for considerado iniciante/ intermédio/ experiente, este utilizador específico herda todas as características definidas para o respetivo estereótipo (iniciante/ intermédio/ experiente).

Variações de fragmentos – existem diferentes versões de fragmentos de uma página. Os fragmentos podem ser pequenos, como uma única palavra, por exemplo, um termo técnico que pode ser substituído por um termo comum com o mesmo significado, ou

grandes, como toda a página. Desta forma, a diversificação de páginas pode ser considerada como um caso especial de variação de fragmentos. O número de diversificações de uma página pode ser elevado, uma vez que cada fragmento pode ser adaptado a diferentes variáveis do modelo de utilizador.

Técnica baseada em *frames* – usando técnicas de linguagem natural as páginas são construídas a partir de pequenos itens de informação como palavras e partes de afirmações.

2.2 Data Mining

A rápida evolução dos recursos computacionais ocorrida nos últimos anos permitiu que fossem gerados grandes volumes de dados. O crescimento destes dados tem gerado uma urgente necessidade de novas técnicas e ferramentas capazes de realizar vários processos, de forma inteligente e automática, transformando os dados em informações significativas e em conhecimento. Apesar destas informações se encontrarem ocultas sob o excessivo volume de dados, podem ser descobertas e, conseqüentemente, representam uma grande valia para o planeamento, gestão e tomada de decisões (Sferra et al., 2003).

Em resposta a esta necessidade, surge o *Data Mining* (DM), processo de extração de informações de bases de dados, sendo fulcral para a tomada de decisões. Na realidade, o DM consiste numa etapa do processo *Knowledge Discovery in Database* (KDD), em português Descoberta de Conhecimento em Bases de Dados. Na década de 90, Fayyad afirma que DM é “o processo não trivial de identificação de padrões válidos, desconhecidos, potencialmente úteis e, conseqüentemente, compreensíveis em dados” (Fayyad et al., 1996). O KDD é um processo que permite a busca de conhecimento em base de dados, seleção de dados, pré-processamento, transformação e redução, *Data Mining*, análises e interpretações (Oliveira, 2005).

O processo KDD possui diversas etapas, sendo que a principal é a que forma o núcleo do processo - *Data Mining*. A figura 2 ilustra as diversas etapas mencionadas anteriormente (Fayyad et al., 1996).

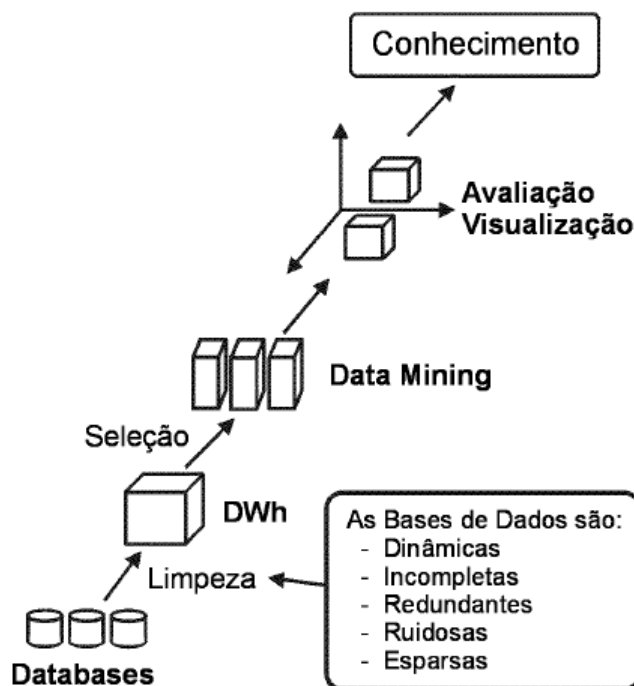


Figura 2 - Visão geral das etapas do processo KDD

A figura 2 exibe as diversas etapas do processo KDD, desde a busca de dados provenientes de distintas fontes de dados até à geração de conhecimento. A descoberta do conhecimento inicia o seu processo a partir de diversas fontes de dados (bases de dados, relatórios, logs, entre outros). Nesta fase são recolhidos os dados e estes são submetidos a uma limpeza, sendo efetuadas várias operações, tais como: preenchimento de informação em falta; avaliação da consistência de dados; e remoção de ruídos. Após a fase de limpeza, os dados podem ser organizados e estruturados em repositórios (*Data Marts* e *Data Warehouses*) e, posteriormente os dados estão preparados para o processo *Data Mining* (Oliveira, 2005). Após a fase de *Data Mining* surge a visualização gráfica da informação para refinar e conduzir à obtenção de padrões úteis.

Em suma, o processo KDD é iterativo, a busca de padrões requer que os dados sem tratamento prévio sejam sistematicamente “simplificados”, de forma a desprezar o específico e privilegiar o genérico (Oliveira, 2005).

O ciclo de vida de um projeto de DM é composto por seis fases (The CRISP-DM Consortium, 2000):

- **Regras de negócio:** visa obter o conhecimento sobre os objetivos do negócio e os seus requisitos, convertendo esse conhecimento numa definição de um

problema de *Data Mining* e num plano preliminar designado para alcançar esses objetivos;

- **Seleção dos dados:** consiste na identificação de amostras representativas dos dados e definição do domínio sobre o qual se pretende executar o processo de descoberta, tendo como objetivo a familiarização com a base de dados pelas pessoas envolvidas neste projeto;
- **Limpeza dos dados:** consiste na preparação dos dados que visa a limpeza, transformação, integração e formatação dos dados da etapa anterior, os ruídos, dados estranhos e/ou inconsistentes são tratados;
- **Modelação de dados:** tem como objetivo a aplicação de técnicas de modelação sobre o conjunto de dados preparados na fase anterior. São selecionadas e aplicadas várias técnicas e os seus parâmetros são calibrados para se obter valores otimizados. Para o mesmo tipo de problema podem ser usadas diversas técnicas e algumas requerem características específicas na forma dos dados, podendo voltar para a etapa de preparação de dados caso seja necessário;
- **Avaliação do processo:** pretende garantir que o modelo gerado atende às expectativas da organização;
- **Execução:** consiste na definição das fases de implantação do projeto de *Data Mining*.

A figura 3 ilustra os diversos tipos de *Data Mining*, bem como os vários métodos usados para extrair conhecimento útil (Rodrigues, 2008).

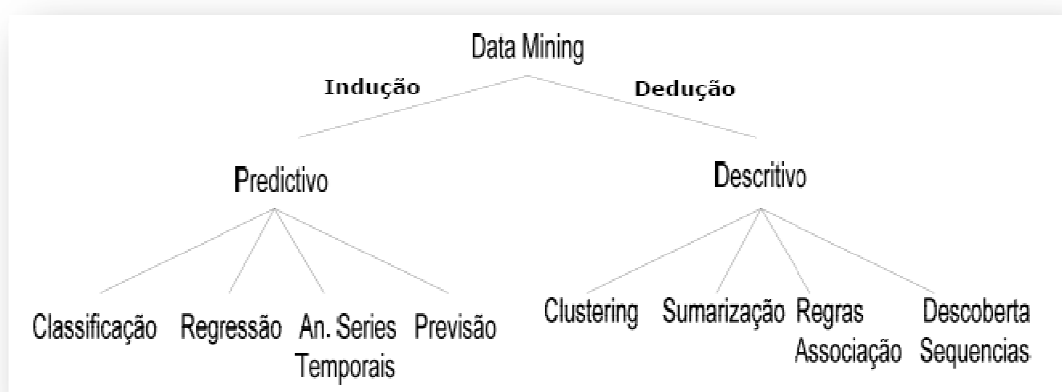


Figura 3 - Tipos de *Data Mining*

Existem dois tipos de *Data Mining*: o preditivo e o descritivo. O primeiro é orientado por um objetivo, processo direto de descoberta. No segundo, não existe um objetivo pré-definido, trata-se de um processo indireto de descoberta (Rodrigues, 2008).

O *Data Mining* preditivo recorre a diversos métodos (Rodrigues, 2008):

- **Classificação:** utiliza um conjunto de dados de entrada, pré-escolhidos, sendo gerado um novo modelo para posteriores classificações. Após o desenvolvimento do modelo de forma eficiente, este será utilizado para classificar novos registos nas classes identificadas (Martins et al., 2008). Geralmente, o método de classificação exige a comparação de um objeto ou dado com outros dados ou objetos que supostamente pertencem a classes anteriormente definidas. Nesta comparação é utilizada uma métrica que permita identificar as diferenças entre eles. Este método pode ser utilizado para compreender dados existentes bem como prever o comportamento dos dados (Euriditionhome, 2004);
- **Regressão:** metodologia estatística que utiliza a relação entre duas ou mais variáveis quantitativas, de modo a que uma variável possa ser prevista a partir de outra ou mais. É usada com duas finalidades: previsão (prever o valor de uma variável a partir de outro valor) e estimar quanto uma variável pode influenciar ou modificar outra variável (Silva et al., 2001). Esta metodologia só funciona bem com dados quantitativos como, por exemplo, peso, velocidade e idade. Quando se utilizam dados categóricos, onde a ordem não é significativa como, por exemplo, cor, nome ou sexo, é aconselhável a escolha de outra técnica (Silva et al., 2001);
- **Análise de séries temporais:** identifica regularidades e características temporais interessantes escondidas nos dados. Este método ainda analisa padrões sequenciais, periódicos, tendências, desvios e busca sequências similares ou subsequências;
- **Previsão:** dado um conjunto de dados e um modelo de previsão, prevê o valor para um atributo específico do respetivo conjunto de dados (Pinto et al., 2005).

O *Data Mining* descritivo dispõe de vários métodos (Rodrigues, 2008):

- **Agrupamento (*Clustering*):** consiste numa das primeiras etapas da metodologia de DM e visa formar grupos de objetos ou elementos mais semelhantes entre si. Pode ser estabelecido previamente um número de

grupos a ser formado ou o próprio algoritmo de *clustering* define a quantidade de grupos, sendo conhecido o número resultante somente no final do processo (Amorim, 2006). Na análise de agrupamento, os grupos ou classes são construídos com base na semelhança entre os elementos, cabendo ao analisador das classes resultantes avaliar se estas significam algo útil. Por exemplo, agrupar sintomas pode gerar classes que não representem nenhuma doença explicitamente, uma vez que doenças diferentes podem possuir os mesmos sintomas (Amorim, 2006).

Este método é utilizado quando nada ou pouco se sabe sobre os dados. Segmentar um mercado é uma típica análise de *clustering*, onde consumidores são reunidos em classes representantes dos segmentos deste mercado (Amorim, 2006).

Em geral, o *clustering* é executado por algoritmos estatísticos específicos para esse fim, porém as redes neuronais e os algoritmos genéticos são também utilizados neste sentido (Han et al., 2001):

- **Sumarização:** descrição do que caracteriza um conjunto de dados, por exemplo, conjunto de regras que descreve o comportamento e relação entre os valores dos dados (Santos, 2008);
- **Regras de associação:** permitem encontrar padrões, associações ou correlações frequentes em conjuntos de itens (objetos) de uma base de dados ou outro tipo de repositórios de informação. A descoberta destas regras pode ser utilizada como suporte à tomada de decisão. O seu objetivo é identificar uma relação entre objetos, designados por itens, passíveis de serem representados em bases de dados e que possam de alguma forma estar relacionados entre si. Um exemplo de regra de associação pode ser uma expressão do género: “De entre os utilizadores do site do INE que consultam estatísticas da saúde, cerca de 45% consultam na mesma sessão, estatísticas demográficas”. Esta afirmação pode ser representada de outro modo:
 - Estatísticas da saúde -> estatísticas demográficas (confiança =0,45) (Neves, 2002);
- **Descoberta de sequências:** é uma variação das regras associativas. As regras identificadas são usadas para reconhecer sequências relevantes que possam ser utilizadas para prever comportamentos, possuindo assim sempre o conhecimento do que pode surgir (Martins et al., 2008).

O DM também recorre a outras técnicas, tais como, a lógica *Fuzzy* e algoritmos genéticos, sendo que cada técnica se adapta melhor a um determinado problema.

Diversas áreas necessitam de recorrer à metodologia de DM para usufruir do conhecimento que podem extrair a partir de grandes volumes de dados que possuem. Este trabalho dá ênfase à área do ensino, nomeadamente à aplicação de técnicas de *Data Mining* a sistemas de hipermédia adaptativa.

2.2.1 *Data Mining* para SHA

Os SHAE geram grandes volumes de dados relativos aos alunos, desde as suas características pessoais, preferências, conhecimentos, comportamentos, entre outros. Vulgarmente, estes dados são menosprezados devido ao seu excessivo tamanho e os elevados custos que a análise a estes dados acarreta. Apesar disso, recorrendo a técnicas de DM, estes sistemas podem obter informações úteis e compreensíveis.

A aplicação de técnicas de DM sobre os dados armazenados ao longo do tempo, relativos ao acompanhamento do aluno, pode encontrar padrões de comportamento que estejam ligados ao processo de aprendizagem do aluno. Esses padrões podem ser utilizados como parâmetros para outros alunos, cabendo ao professor a decisão sobre quais aspetos aplicar as técnicas de DM, acrescentando maior flexibilidade ao acompanhamento do aluno (Silva et al., 2001).

Cada técnica de DM pode extrair um padrão distinto relacionado com a modelação de alunos, regras de adaptação e pedagógicas e comportamentos de alunos. De seguida, são apresentados alguns exemplos de extração de padrões, sendo que os dados utilizados são hipotéticos.

Classificação

Os algoritmos mais utilizados para a técnica de classificação são as árvores de decisão (Pelegri et al., 2005), regressão (Han et al., 2001) e redes neuronais (Sousa, 1998).

Para exemplificar é utilizado um excerto de uma base de dados de controlo de oferta de cursos à distância, cujo esquema é apresentado na figura 4 (Silva et al., 2001):

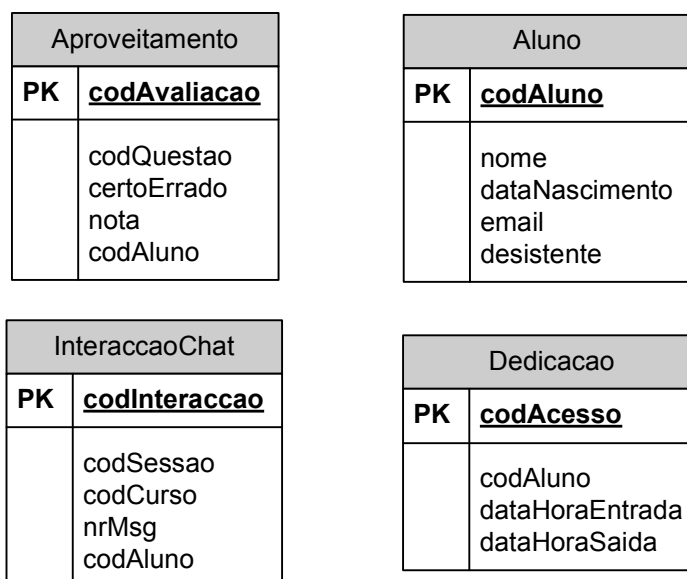


Figura 4 - Excerto de uma base de dados de controlo de ambientes à distância

Este excerto da base de dados é composto por quatro tabelas: a tabela “InteracaoChat” que armazena o número de mensagens que um aluno de um curso envia numa sessão de chat. A tabela “Aproveitamento” que contém dados sobre as questões que um aluno acerta ou erra numa avaliação formal e a nota que recebe em cada questão. A tabela “Dedicacao” exhibe os horários de cada acesso do aluno no ambiente de ensino à distância (EaD) e a tabela “Aluno” apresenta informações pessoais do aluno e se este desistiu ou não do curso (Silva et al., 2001). Para cada técnica de DM a ser executada sobre esta base de dados é necessário uma preparação de dados. Nas técnicas apresentadas posteriormente são considerados alguns cuidados, tais como, a preparação de dados.

Ao longo do curso à distância foram recolhidos e armazenados alguns dados e através destes foi gerada a tabela 1 (Silva et al., 2001).

Tabela 1 - Dados sobre um módulo de um curso à distância

Nr. De Mensagens de Chat	Aproveitamento Avaliação Formal (%)	Desistência
10	45	Sim
12	65	Não
20	65	Não
9	55	Sim
18	70	Não
43	80	Não
0	30	Sim

23	60	Não
4	50	Sim

A coluna Nr. de Mensagens de Chat indica o número total de mensagens enviadas pelo aluno nas discussões durante o curso analisado. A coluna Aproveitamento em Avaliação Formal exibe a média das notas das provas realizadas no curso e, por fim, a coluna Desistência identifica se o aluno desistiu ou não do curso. Cada linha da tabela representa um aluno. A técnica de classificação pode prever se um aluno irá desistir ou não do curso, tendo em conta algumas das suas características e dados históricos de outros alunos (Silva et al., 2001).

Após análise dos dados descritos na tabela anterior, é possível extrair algumas regras como, por exemplo:

- **Se** Nr. de Mensagens de Chat \leq 10 **Então** Desistência = sim;
- **Se** Aproveitamento em Avaliação Formal $>$ 60 **Então** Desistência = não.

Esta regra extraída permite detetar antecipadamente a possibilidade de desistência de um aluno num curso lecionado à distância.

Regras de associação

A tabela 2 representa os dados obtidos do cumprimento das atividades de um curso pelos alunos. Estes dados permitem encontrar regras de associação entre atividades de um curso cuja confiança (medida para a frequência de sucesso da regra na base de dados) seja superior ou igual a 60% e suporte (medida para a frequência de ocorrência da regra na base de dados) seja superior ou igual a 50% (Silva et al., 2001).

Tabela 2 - Dados sobre o cumprimento das atividades de um curso pelos alunos

Aluno	Atividade 1	Atividade 2	Atividade 3	Atividade 4
Aluno 1	Cumpriu	Não cumpriu	Cumpriu	Cumpriu
Aluno 2	Não cumpriu	Cumpriu	Cumpriu	Não cumpriu
Aluno 3	Cumpriu	Cumpriu	Não cumpriu	Cumpriu
Aluno 4	Cumpriu	Não cumpriu	Cumpriu	Cumpriu
Aluno 5	Cumpriu	Cumpriu	Cumpriu	Não cumpriu
Aluno 6	Não cumpriu	Não cumpriu	Cumpriu	Cumpriu

Ao analisar os dados da tabela 2, é verificada a existência de muitas associações mas só são exibidas aquelas que apresentam confiança igual ou superior a 60% e suporte igual ou superior a 50%. Por exemplo (Silva et al., 2001):

Atividade 1 → Atividade 3

- com 50% de suporte (50% dos alunos realizaram as atividades 1 e 3);
- e 75% de confiança (75% dos alunos que realizaram a atividade 1 também realizaram a 3).

Atividade 3 → Atividade 4

- com 50% de suporte (50% dos alunos realizaram as atividades 3 e 4);
- e 60% de confiança (60% dos alunos que realizaram a atividade 3 também realizaram a 4).

Atividade 4 → Atividade 1

- com 50% de suporte (50% dos alunos realizaram as atividades 4 e 1);
- e 75% de confiança (75% dos alunos que realizaram a atividade 4 e também a 1).

Estas regras de associação permitem prever o cumprimento das atividades do curso dos alunos consoante as atividades realizadas anteriormente.

Agrupamento (*Clustering*)

Recorrendo à técnica de *clustering* e aplicando-a ao acompanhamento do aluno podendo, por exemplo, agrupar os alunos por um dos seguintes atributos: faixa etária, rendimento familiar, tipo de ligação à internet, grau de escolaridade, conhecimento prévio do assunto da disciplina/curso, tempo de dedicação aos estudos, entre muitos outros atributos.

Os dados contidos na tabela 3 são obtidos a partir da base de dados descrita na tabela 1, em cada linha contém a idade e aproveitamento de um aluno (Silva et al., 2001)

Tabela 3 - Dados para *clustering*

Aproveitamento em Avaliação Formal (%)	Idade
45	16
65	20
65	19
55	32
70	29
80	25
30	17
60	20
50	18

Uma das possibilidades é agrupar os alunos tendo em conta a idade e o aproveitamento. Nesse caso, um resultado possível é agrupar os alunos em 4 categorias (Silva et al., 2001).

- Alunos com idade entre os 16 e 18 anos e aproveitamento inferior a 60%;
- Alunos com idade entre 19 e 20 anos e aproveitamento entre 60 a 65%;
- Alunos com idade entre 25 e 29 anos e aproveitamento entre 70 a 80%;
- Alunos com idade maior que 29 anos e aproveitamento inferior a 60%.

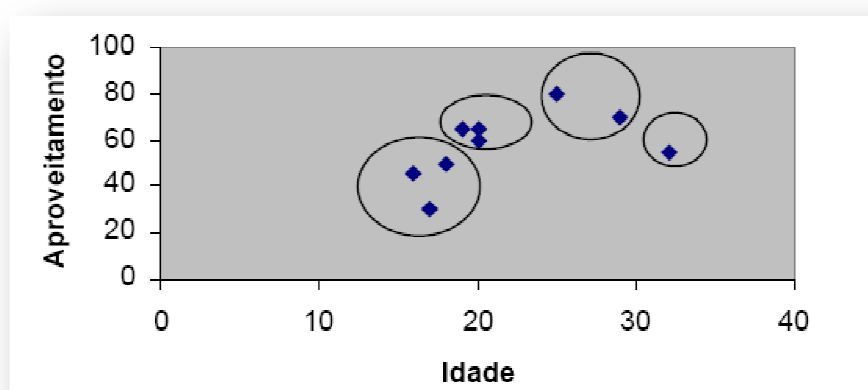


Figura 5 - *Clustering* por idade e aproveitamento

A partir destas categorias são extraídas regras de classificação:

- **Se** idade < 18 ou idade > 29 **Então** aproveitamento < 60%;

- **Se** $19 \leq \text{Idade} \leq 20$ **Então** $60\% \leq \text{aproveitamento} \leq 65\%$;
- **Se** $25 \leq \text{Idade} \leq 29$ **Então** $70\% \leq \text{aproveitamento} \leq 80\%$.

As regras anteriores permitem fazer previsões em relação ao aproveitamento do aluno, consoante a idade do aluno.

Regressão

O uso da regressão para o acompanhamento do aluno no EaD, pode prever o aproveitamento total do aluno considerando o tempo de permanência no site e os dados históricos sobre a permanência dos outros alunos.

Através de métodos matemáticos de regressão é construída a tabela 4 com uma função sobre o tempo de permanência no site do curso. Por exemplo, foi deduzida a função $f(t) = 5t$, em que t representa o tempo de permanência no site e $f(t)$ representa o aproveitamento do aluno. Recorrendo a esta fórmula hipotética, um aluno que permaneça 12,3 horas no site tem um aproveitamento estimado em 61,5% (Silva et al., 2001).

Tabela 4 - Tempo de permanência X Aproveitamento

Tempo de permanência (h)	Aproveitamento (%)
2	10
4	20
6	30
8	40
10	50
12	60
14	70
16	80
18	90
20	100

Nos SHAE o acompanhamento do aluno é um processo complexo pois carece da falta de contacto presencial entre o aluno e o professor, grande volume de dados para análise, ausência de flexibilidade na análise, entre outros. A aplicação de algoritmos de DM pode gerar padrões para a modelação e comportamento de alunos, regras de adaptação e pedagógicas,

facilitando a aprendizagem e auxiliando o professor na árdua análise dos dados de alunos fornecidos pelos sistemas (Silva et al., 2001).

A principal vantagem do uso da metodologia de DM para SHA consiste em fornecer análises mais refinadas do que as análises realizadas sem recorrer a esta metodologia, permitindo uma generalização do comportamento do aluno, por meio dos padrões obtidos, sem esquecer as suas características individuais e relacionando os padrões mais adequados a cada aluno (Silva et al., 2001).

2.3 Modelo de aluno

O termo perfil de aluno é utilizado como sinónimo de modelo de aluno, porém existe uma distinção para indicar o perfil do aluno como o modelo de um aluno específico. O perfil do aluno representa as capacidades cognitivas do utilizador, habilidades intelectuais, intenções, estilos do estudante ou preferências. Estas características são constantemente atualizadas ao mesmo tempo que o aluno navega no sistema e são utilizadas para adaptar dinamicamente o seu conteúdo, navegação e apresentação (Koch, 2000).

São identificadas cinco grupos de características associadas aos alunos, que são importantes para os sistemas adaptativos: o conhecimento, os objetivos, a história, a experiência e as preferências. Estas características são dinâmicas e necessitam que o modelo de aluno se encontre constantemente atualizado (Brusilovsky, 1996):

Conhecimento: o conhecimento do aluno sobre os assuntos abordados pelo sistema é considerado a característica mais importante do modelo de aluno. Assim, o sistema deve ser capaz de controlar dinamicamente as modificações e realimentar o modelo de aluno à medida que o aluno aprende com a utilização do sistema. Normalmente, dois modelos são usados para a modelação do aluno: modelo de sobreposição e modelo de estereótipos que serão estudados posteriormente (Brusilovsky, 1996);

Objetivos: os objetivos podem variar ao longo do tempo. Uma maneira utilizada para modelar os objetivos é por meio de pares objetivo-valor, onde o valor representa a probabilidade de que o objetivo correspondente seja o objetivo atual do aluno (Brusilovsky, 1996). Outra forma de modelar os objetivos dos alunos é apresentá-los através de uma hierarquia (árvore) de tarefas relacionadas (Vassileva, 1996);

Historial: o histórico do aluno em tópicos relevantes, não relacionados propriamente com os assuntos do sistema, é fator importante na construção do modelo de aluno. Por exemplo: conhecimentos sobre a linguagem Pascal podem ser considerados como experiência relevante para um aluno num SHAE de um curso de Java (Brusilovsky, 1996);

Experiência: é referente à experiência do aluno com sistemas baseados em hipertexto ou hipermédia, adaptativos ou não (Brusilovsky, 1996);

Preferências: os alunos de sistemas de hipermédia podem ter diferentes preferências como, por exemplo: fontes, cores, entre outros. Estas características de forma geral não podem ser deduzidas pelo sistema, sendo necessário que o aluno as declare formalmente ou por meio de resposta a questões simples (Brusilovsky, 1996).

2.3.1 Técnicas de modelação de alunos

Existem algumas técnicas que permitem a representação de modelos de alunos, também designadas como modelos de descrição:

- **Modelo diferencial:** as respostas do aluno são comparadas com a base de conhecimento. Este é dividido em duas classes: conhecimento esperado e não esperado do aluno. Nesta abordagem, é comparado o desempenho do especialista com o do aluno e não o conhecimento deles (Damico, 1995). O resultado dessa comparação deve fornecer hipóteses sobre o que o aluno não conhece.
- **Modelo de sobreposição (*overlay*):** é o paradigma padrão de representação. O modelo de aluno é considerado um subconjunto do modelo do especialista (Figura 6). Desta forma, ambos os modelos devem possuir a mesma representação. A validade deste modelo está no facto de que, considerando que o conhecimento do aluno é uma fração do especialista, fica simples identificar o material a ser apresentado. Os tipos de conhecimento que podem ser representados dentro deste modelo são: tópicos (correspondentes a elementos do conhecimento do domínio) e regras de produção, correspondendo a relações entre tópicos (Anderson, 1993).



Figura 6 - Modelo de sobreposição (*overlay*)

- **Modelo de perturbação (*buggy*):** por considerar o modelo de aluno um subconjunto do modelo do especialista, o modelo de *overlay* considera que qualquer aluno possui apenas conceitos corretos quanto ao domínio. Isto não é coincidente com a realidade, tendo em conta que eles podem ter conceitos errados sobre algum tópico. O modelo *buggy* constitui uma melhoria ao modelo *overlay*, permitindo que os alunos tenham crenças que não pertencem à base de conhecimento do especialista. A adaptação ao modelo de *overlay* pode ser percebida através da figura 7 (Giraffa, 1995).



Figura 7 - Modelo de perturbação (*Buggy*)

- **Modelo de simulação:** este modelo consegue prever os próximos passos do utilizador. Para isto, o sistema deve monitorizar o aluno, observando quais as suas reações mediante determinadas situações (Woods et al., 1995).
- **Modelo de crenças:** o sistema possui um conjunto de crenças que reflete o que considera que o aluno sabe sobre o domínio. Na verdade, essas crenças são

organizadas: crenças do professor a respeito das crenças do aluno. Este modelo deve ser revisto, porque tanto as hipóteses do tutor como as do próprio estado cognitivo do aluno podem estar incorretas e mudam com o tempo, em consequência da interação entre aluno e professor (Giraffa, 1995).

- **Modelo de agentes:** o aluno e professor são considerados duas personagens dotadas de capacidade cognitiva e interagem entre si (Giraffa, 1995).
- **Modelo estereótipo:** as propriedades e conhecimentos dos utilizadores são representados com pares de valores – itens. A diferença é que distintas combinações de pares são atribuídas a estereótipos, como iniciante, intermédio e experiente. O utilizador ao ser classificado dentro de um estereótipo, herda todas as propriedades definidas para este estereótipo (Bugay, 2006).

O conceito de estereótipo é simples, apesar disso é menos flexível e poderoso do que algumas técnicas de modelação de utilizadores. Bons resultados têm sido obtidos através da combinação da técnica de estereótipos com a de sobreposição. A inicialização do modelo é executada com a atribuição de um estereótipo para o utilizador, a qual é refinada a cada etapa de implementação da modelação por sobreposição (Rich, 1979).

Este modelo é composto por três tarefas (Dara-Abrams, 2002):

- Identificar os subgrupos de utilizadores: para definir os subgrupos dentro de toda população cujos membros têm atributos similares relevantes para a aplicação.
- Identificar características chave: para o modelo de utilizador é necessário identificar um pequeno número de características importantes que diferenciam um subgrupo de outro (Rich, 1979).
- Desenvolver a representação da hierarquia de sequência de estereótipos: as características dos subgrupos de utilizadores que são relevantes para a aplicação devem ser descritas numa representação abstrata do sistema.

Modelos de estereótipos são suficientes para modelação da interface ou para seleccionar o tipo de instrução, porém são insuficientes quando a adaptação individual requer uma descrição mais refinada do utilizador, ajuda específica ou dicas (Koch, 2000).

Além destes, é citado o uso de Redes Bayesianas (Beck, 1998). Esta técnica deriva da utilização da fórmula matemática para o cálculo de probabilidade condicional e consiste em diagramas que organizam o conhecimento numa dada área através de um mapeamento entre causas e

efeitos, estabelecendo relações prováveis ou “dependências” (Fred, 2004). O anexo A apresenta uma descrição mais detalhada acerca desta técnica.

A lógica *Fuzzy* também é considerada como uma ferramenta que pode ser usada para gerar modelos de alunos. Esta técnica permite que um conceito possa coexistir com o seu oposto, gerando soluções precisas através de informações certas ou mesmo aproximadas. Um objeto ou dado não pode ser simplesmente classificado como verdadeiro ou falso, sim ou não, etc. Por exemplo, a taxa de risco de um determinado investimento não pode ser classificada somente como alta ou baixa, uma vez que podem existir infinitas possibilidades (Bugay, 2006).

A lógica *Fuzzy* apresenta as seguintes características (Bugay, 2006):

- É baseada em palavras (linguagem natural) e não em números, os valores de verdade são expressos linguisticamente como, por exemplo, muito quente, quente, frio e muito frio;
- Utiliza modificadores de predicado (muito, pouco, mais ou menos, bastante, médio, etc.);
- Usa valores como, provável e improvável, os quais são interpretados e manipulados;
- Manuseia toda a gama de valores entre 0 e 1, tomando estes apenas como limite, ao contrário do que se passa na lógica clássica em que apenas são considerados os valores para definir o grau de pertença a um conjunto.

Por fim, as redes semânticas também podem ser usadas na representação de modelos de alunos. Uma rede semântica é definida como uma estrutura gráfica na qual os nós representam os conceitos, enquanto os arcos entre estes nós representam a relação entre eles. A partir desta perspectiva, os conceitos não têm sentido de forma isolada e somente adquirem significado quando relacionados com outros conceitos, os quais estão ligados através de arcos relacionais. Portanto, nas redes semânticas a estrutura é essencial (Rich, 1988). O anexo B apresenta uma explicação mais pormenorizada em relação à lógica *Fuzzy*.

2.3.2 Inicialização de modelos de alunos

Na inicialização de modelos de alunos é necessário proceder à inicialização dos atributos, que pode ser realizada através de uma sondagem inicial, atribuição de valores por defeito ou combinando estas duas técnicas. Também existe uma técnica que consiste na partilha de modelos de utilizadores, algumas inicializações de modelos ou, partes delas, podem ser

efetuadas através de importações de dados comuns de outros modelos de utilizadores de outro SHA (Koch, 2000).

A sondagem inicial é efetuada pelo preenchimento de questionários e/ou formulários, sendo uma técnica usada para obter dados sobre o aluno e seu conhecimento em relação ao conteúdo hipermedia. A sua dificuldade consiste na determinação do número de questões que o aluno está disposto a responder e quantas são necessárias para obter o conhecimento do mesmo, uma vez que o número e o tipo de questões podem variar consoante o conteúdo da aplicação (Dara-Abrams, 2002).

A técnica do uso de valores por defeito recorre ao modelo de estereótipos, enquadra o aluno num determinado grupo e utiliza as características deste estereótipo para inicializar o modelo (Wu, 2002). Um estereótipo de um modelo pode distinguir diversos tipos de utilizador: inexperiente, intermédio e experiente (Brusilosky, 1996).

Após a inicialização é seguida a fase de atualização do modelo de aluno. Esta etapa varia consoante o comportamento do aluno durante o período de utilização da aplicação e, tem como objetivo principal, tornar o modelo mais compatível e atualizado com o aluno, refletindo as suas necessidades e crenças. A aquisição destes dados consiste num processo de recolha de diversas formas de interação do aluno com o sistema, quer seja por um clique do rato, digitalização de texto, controlo do tempo despendido, páginas visitadas, passos usados para solucionar um determinado problema, entre outros. A partir do controlo destes itens, que podem oscilar consoante a aplicação, o sistema pode deduzir o que o aluno conhece ou não, as suas preferências e objetivos (Rossatelli et al., 2004).

As técnicas de atualização do modelo de aluno podem ser caracterizadas de diferentes maneiras (Wu et al., 2004):

- **Ativa ou passiva** – Esta classificação é baseada na participação ou não do aluno na aquisição dos dados. As técnicas ativas interagem indiretamente com o aluno, através de formulários *online*, levantando informações sobre o mesmo. Por sua vez, as técnicas passivas possibilitam a construção de um modelo baseado na conclusão de observações do comportamento do aluno, como por exemplo: páginas visitadas; análise de *logs*. Geralmente, os sistemas usam ambas as técnicas, sendo que a técnica ativa destinada à inicialização do modelo, enquanto que a passiva atualiza e aperfeiçoa o modelo durante o seu uso (Koch, 2000).

- **Automática ou determinada pelo aluno** – A definição está relacionada com quem inicia o processo de aquisição. As técnicas determinadas pelos alunos são aquelas onde é o próprio que decide quando altera o modelo. Nas técnicas automáticas, o aluno não tem influência sobre o momento da atualização do modelo. Sistemas que utilizam técnicas determinadas pelo aluno têm sido mais frequentes (Bugay, 2006).
- **Direta ou indireta** – Uma técnica direta deriva de informação direta do aluno, sendo utilizada para atualizar o modelo. Enquanto técnicas indiretas constroem o modelo de aluno com informação extraída de regras de inferências. A definição dos estereótipos é o melhor exemplo do uso de regras de inferência (Rich, 1979).
- **Explícita ou implícita** – Esta classificação depende do tipo de retorno do aluno. As técnicas explícitas são aquelas onde o aluno, conscientemente fornece as informações, enquanto as implícitas consistem na observação do comportamento do aluno independente do seu consentimento (Koch, 2000).
- **Lógica ou plausível** – A diferença entre lógica e plausível reside no grau de aceitação dos resultados, as técnicas plausíveis requerem a representação explícita da incerteza do modelo de aluno, necessitando para tal de mecanismos para manter a consistência no modelo do aluno. A manipulação destas incertezas pode ser gerida com o recurso a redes bayesianas, enquanto que as técnicas de aquisição lógica são frequentemente usadas em modelos de sobreposição (Jameson, 2003).
- **Online ou offline** – Esta classificação depende do momento em que ocorre a aquisição de dados. A maioria das técnicas é aplicada *online*, porém para a criação de estereótipos podem-se obter informações de base de dados quando o aluno está *offline* (Bugay, 2006).

Geralmente, o processo de aquisição de informações relativas ao aluno é constituído por três fases distintas: colheita de dados, diagnóstico e consistência dos dados (Rossatelli et al., 2004).

Na primeira fase, colheita de dados, o principal problema reside na fiabilidade dos dados, na quantidade de dados disponíveis e no nível de detalhe dos mesmos. A quantidade de dados necessária depende da complexidade do modelo de aluno (Koch, 2000).

Após a colheita de dados, ocorre o processo de diagnóstico que tem como função a deteção de falhas e/ou erros que possam ocorrer na fase anterior. Este processo consiste em duas etapas:

a transformação dos dados recolhidos para que possam ser utilizados pelo sistema e a avaliação de cada um destes dados resultantes do comportamento do aluno com os comportamentos considerados “corretos”. O processo de transformação consiste na separação de dados relevantes do total da informação recolhida para estabelecer as capacidades do aluno. A avaliação pretende comparar o conhecimento ou o comportamento do aluno com a conceção de conhecimento ou comportamento “correto” (Palazzo, 2004).

A última fase do processo de aquisição de informações, consistência dos dados, é referida à deteção de conflitos de consistência de dados. Como se incorporam novos dados ao modelo do aluno é frequente a geração de inconsistências com os dados existentes e, conseqüentemente, conflitos. Esta fase determina qual a hipótese a utilizar para solucionar o conflito (Bugay, 2006).

A modelação de utilizadores, neste caso de alunos, não é propriamente uma tarefa simples e trivial. Neste contexto, Brusilovsky afirma que (Brusilovsky, 1996):

“... existem alguns problemas com a modelação automática do utilizador, podendo ocorrer erros na dedução do modelo de utilizador e na execução da função de adaptação. A modelação do utilizador é uma tarefa bastante complexa e os múltiplos aspetos envolvidos no processo (modelação estática, dinâmica, comportamental, etc.) contribuem para dificultar ainda mais a sua execução, mantendo os resultados obtidos em níveis apenas parciais e restritos.”

2.4 Resumo

A hipermédia adaptativa é uma área que tem como origem a fusão da hipermédia e modelação do utilizador. Um sistema de hipermédia adaptativa possui um modelo de utilizador que armazena objetivos, preferências, conhecimentos e experiências de cada utilizador e usa-o, dinamicamente com o objetivo de adaptar os conteúdos, a navegação e apresentação do sistema às necessidades do utilizador (Brusilovsky, 1996). Durante a navegação do utilizador pela aplicação, o modelo deve ser atualizado e utilizado constante e dinamicamente, tendo em conta que as características pessoais do próprio utilizador sofrem sucessivas alterações ao longo do tempo (Koch, 2000).

No caso de sistemas de hipermédia adaptativa com finalidades educativas, o modelo de aluno serve para invocar e/ou definir estratégias pedagógicas. Desta forma, é essencial que os dados

guardados no modelo do aluno sejam usados como argumentos nas funções de decisão dos SHA (Silva et al., 1998).

Os sistemas de hipermédia adaptativos educacionais produzem grandes volumes de dados sobre os alunos que devem ser analisados, recorrendo a diversas técnicas de *Data Mining* estes dados podem ser processados e transformados em informações úteis e compreensíveis, sendo fulcrais para a obtenção de novos perfis de alunos, geração de regras de adaptação e pedagógicas, padrões de comportamentos de alunos e grande auxílio no acompanhamento entre o professor e aluno (Silva et al., 2001).

3 Ferramentas de *Data Mining*

“Só é útil o conhecimento que nos torna melhores.”

Sócrates

No mercado existem diversas ferramentas de *Data Mining* (Han et al., 2001) que apresentam diferentes técnicas de *Data Mining* (DM). No âmbito do presente trabalho, é elaborado um estudo comparativo entre diferentes ferramentas de DM existentes, com o intuito de fundamentar a escolha de uma ferramenta com características apropriadas para o trabalho a desenvolver.

Neste trabalho é pretendido utilizar uma ferramenta de *Data Mining* para extrair algumas regras e padrões de comportamento dos alunos relativamente à aprendizagem de Matemática. O principal objetivo consiste em prever o comportamento de um novo aluno consoante os dados históricos dos alunos, identificando as influências de cada característica do modelo do aluno em relação às classificações mais positivas e negativas e agrupando alunos com boas, médias e más classificações. Através de algumas técnicas de DM, nomeadamente, classificação, regras de associação e *clustering*, é possível alcançar os objetivos definidos. A técnica de classificação pode ser utilizada para discriminar os alunos em classes ou categorias, usando um modelo de dados que será construído com os dados existentes dos alunos e tendo como objetivo a previsão do comportamento de um novo aluno, em qual categoria será classificado. As regras de associação possibilitam identificar relações entre os dados dos alunos, determinar qual a influência entre as características do modelo do aluno nos resultados finais da avaliação. A técnica de *clustering* permite criar grupos de alunos com semelhanças entre si, alunos com bons, médios e bons resultados.

Após uma breve análise das soluções de DM existentes no mercado, foram selecionadas quatro ferramentas que representam a realidade da maioria das soluções e detêm as técnicas de DM necessárias ao caso de estudo, sendo destacadas formas de importação de dados, construção de modelos distintos, exibição de técnicas de DM, visualização de dados e soluções comerciais e/ou gratuitas. As ferramentas de DM analisadas são:

- *Weka (Waikato Environment for Knowledge Analysis)* - software de *Data Mining* desenvolvido pela Universidade de Waikato, Nova Zelândia;
- *Oracle Data Mining* - produto de *Data Mining* da Oracle;
- *Enterprise Miner* - ferramenta de *Data Mining* desenvolvida pelo *Statistical Analysis System Institute Inc. (SAS)*;
- *Clementine* - produto de *Data Mining* da *Statistical Package for the Social Sciences (SPSS)*.

3.1 Weka

O *Weka* é uma ferramenta *open source*, implementada na linguagem Java, dentro das especificações de *General Public License (GPL)*. Em 1993 o projeto foi iniciado e deu origem a esta ferramenta, sendo desenvolvido por um grupo de investigadores da Universidade de Waikato, Nova Zelândia. Ao longo dos anos foi consolidado e transformado numa das ferramentas de *Data Mining* mais usada em ambiente escolar (Gonçalves, 2010).

Este sistema apresenta algumas técnicas de *Data Mining*, classificação, regras de associação e *clustering* (Gonçalves, 2010). As principais características consistem na interface gráfica amigável e portabilidade, sendo possível utilizá-lo em distintos sistemas operativos (Windows, Linux e Macintosh), além de beneficiar das vantagens da orientação a objetos (Damasceno, 2010).

A ferramenta, *Weka*, pode ser utilizada em modo gráfico, linha de comandos e *Application Programming Interface (API)*. Na interface gráfica é disponibilizado um ambiente *Weka Explorer* que permite a execução de algoritmos de *Data Mining* de forma interativa (Gonçalves, 2011). A linha de comando é um meio utilizado para dar mais agilidade a processos repetitivos e acesso direto a funcionalidades que teriam inúmeros passos a serem executados na interface gráfica. A opção de acesso via API é usada por desenvolvedores de *software* porque oferece um meio mais prático para as funcionalidades implementadas no *Weka* (Damasceno, 2010).

Uma das principais desvantagens do *Weka* reside no formato específico de ficheiros de entrada de dados, formato *Attribute Relation File Format (ARFF)*, arquivo de texto que é dividido em duas partes. Na primeira são definidos o tipo de atributos e/ou valores que estes podem representar, também designado como cabeçalho. A segunda parte contém os dados referentes aos atributos declarados no cabeçalho. A estrutura deste formato não é flexível e pode inviabilizar o uso desta ferramenta em algumas aplicações reais porque outras ferramentas de DM não se restringem a um formato específico de ficheiros de dados de entrada. Outra das limitações do *Weka* consiste na baixa capacidade de processamento de dados (Gonçalves, 2010).

3.2 Oracle Data Mining

O *Oracle Data Mining (ODM)* é uma ferramenta comercial desenvolvida pela Oracle e consiste num módulo do *Oracle Database 11g Enterprise Edition* que permite criar modelos de dados que proporcionam novas abordagens aos dados e previsões antecipadas de comportamentos (Oracle Corporation, 2010). Este sistema está disponível em plataformas Windows e Unix (Cruz, 2007).

A ferramenta ODM inclui as técnicas de DM: classificação, regressão, regras de associação, *clustering* e seleção de atributos (Anacleto, 2009). O ODM pode ser disponibilizado em diversas formas, API e modo gráfico. A interface gráfica é conhecida por Oracle Data Miner e permite a interligação direta com o Sistema de Gestão de Bases de Dados (SGBD) do Oracle (Oracle Corporation, 2008).

Nesta ferramenta, a construção dos modelos de dados é dividida em duas etapas. A primeira consiste na seleção das condições gerais que são exigidas pela técnica escolhida e que são comuns aos diversos algoritmos dessa técnica. Na segunda fase são escolhidos os parâmetros específicos do algoritmo a aplicar no estudo em causa. Os modelos de dados construídos têm uma visualização própria de acordo com as técnicas e algoritmos selecionados. Os diversos resultados são apresentados de forma amigável através de gráficos ou tabelas, sempre com a possibilidade de exportação para ficheiro (Oracle Corporation, 2008).

As principais características da ferramenta residem na facilidade de utilização do sistema através de um conjunto de *wizards* que auxiliam o utilizador na criação de modelos automaticamente e na seleção dos melhores e mais apropriados, bem como, a integração com

outros produtos da Oracle (Oracle Corporation, 2008). E a sua maior limitação consiste no elevado custo do produto (Elder et al., 1998).

3.3 Enterprise Miner

O *Enterprise Miner* (EM) é uma ferramenta comercial desenvolvida pelo SAS, que segue a metodologia *Sampling, Exploration, Modification, Modeling and Assessment* (SEMMA), de acordo com as cinco fases: amostragem, exploração, modificação, modelação e avaliação (Held, 1997).

Tabela 5 - Metodologia SEMMA

Etapa	Características
Amostragem	Criação de uma amostra significativa que representa a realidade dos dados.
Exploração	Busca de padrões de comportamento, relações, tendências e anomalias, de forma a obter maior compreensão dos dados.
Modificação	Manipulação dos dados, incluindo novos dados gerados na etapa de exploração.
Modelação	Descoberta de novas informações recorrendo aos modelos de dados criados.
Avaliação	Avaliação do desempenho dos diferentes modelos de dados.

As etapas exibidas na tabela 5 correspondem a um ciclo, em que as diversas etapas podem ser executadas de forma iterativa sempre que se considere necessário (Groth, 2000).

O EM fornece diversas técnicas de *Data Mining*: classificação, regras de associação e regressão. Este *software* também permite a análise estatística, sobressaindo a sua variedade de ferramentas de análise estatística, que são baseadas na longa história do SAS no mercado de análise estatística. Estas ferramentas de análise podem ser combinadas e possibilitam a criação e comparação de vários modelos (Neves, 2007).

Os pontos principais do EM são referidos à elevada quantidade de algoritmos para manusear os dados e à interface gráfica que possibilita aos analistas organizar projetos de *Data Mining* num diagrama lógico, apresentando todos os passos do projeto num fluxo de dados e na fácil interpretação de resultados extraídos. A sua principal desvantagem reside na dificuldade de uso, sendo uma ferramenta bastante complexa (Cunha, 2009).

O EM é disponibilizado em versões cliente/ servidor para sistemas operativos Windows, Linux e Solaris (Arbex, 2010).

3.4 Clementine

O *Clementine* é uma ferramenta comercial da SPSS que incorpora funcionalidades capazes de garantir um projeto de *Data Mining*, inclusivamente de suporte metodológico, metodologia *CRoss-Industry Standard Process for Data Mining* (CRISP-DM), através de criação de um ficheiro especial, documentando todas as fases consideradas na implementação de um projeto desta natureza (Pereira, 2002). Possibilita a utilização de técnicas de DM: classificação, associação, previsão, *clustering* e regressão (Kleinschmidt, 2007).

Exibe uma interface simples, baseada em nós, sendo ligados e formando uma estrutura de fluxograma. Cada nó realiza uma tarefa como, por exemplo: importação ou extração, filtragem de dados e estimação do modelo ou erro. Esta ferramenta também permite expandir ou agrupar nós da estrutura, possibilitando uma visualização mais ou menos carregada de detalhes. A importação de dados pode ser realizada por grande variedade de tipos de ficheiros, incluindo ficheiros de texto - *flat files* (Coelho, 2005).

As características mais atrativas do *software Clementine* são a interface para programação visual que permite a construção de aplicações de DM de uma forma intuitiva e a facilidade em operar grande quantidade de volume de dados através de um modelo cliente-servidor (Capitão, 2010). Como desvantagem é identificado o elevado custo desta ferramenta (Nisbet, 2004).

O *Clementine* suporta as plataformas HP/UX, IBM AIX (*Advanced Interactive eXecutive*), Sun Solaris e Windows (Capitão, 2010).

3.5 Estudo comparativo

A escolha da ferramenta de *Data Mining* a utilizar neste trabalho não é uma tarefa fácil. Ao contrário do que se possa pensar inicialmente, a ferramenta de DM escolhida pode não representar a solução mais avançada ou a que apresenta maiores taxas de precisão na previsão. Mais importante do que estes critérios é identificar que a ferramenta selecionada é simples de usar e que apresenta as técnicas e algoritmos de *Data Mining* necessários ao estudo em questão.

Desta forma, é realizado o estudo comparativo de ferramentas de DM, abordando algumas características que são interessantes analisar para contribuir para a seleção da ferramenta de DM mais adequada ao trabalho:

- **Plataformas:** ambiente de trabalho em que a ferramenta é inserida;
- **Automação:** método que auxilia a programação visual. Por exemplo: caixa de diálogos, linguagem de programação, assistentes;
- **Técnicas DM:** técnicas que efetuam o processo de *Data Mining*;
- **Algoritmos:** algoritmos disponíveis nas técnicas de DM;
- **Visualização de dados:** representações visuais dos resultados extraídos dos dados;
- **Outros critérios:** Interface simples; Comercial/ Gratuito; Capacidade de processamento;
- **Vantagens e desvantagens:** síntese dos pontos fortes e fracos de cada ferramenta.

Plataformas

As ferramentas de DM apresentadas anteriormente são suportadas por distintas plataformas. A tabela 6 ilustra as plataformas (Cruz, 2007; Elder et al., 1998).

Tabela 6 - Plataformas

Ferramenta	PC <i>Standalone</i> (Windows)	<i>Unix</i> <i>Standalone</i>	Servidor <i>Unix</i> / Cliente PC	Servidor NT / Cliente PC	Conexão à Base de Dados
<i>Weka</i>	√	√			√
<i>ODM</i>	√	√	√	√	√
<i>EM</i>	√		√	√	√
<i>Clementine</i>	√	√		√	√

Através da tabela 6 é verificada a existência de quatro tipos de plataforma: PC *Standalone* (Windows), *Unix Standalone*, Servidor *Unix*/ Cliente PC e Servidor NT/ Cliente PC. A plataforma

PC *Standalon* Windows significa que o computador não comunica com os outros computadores em Windows, sendo também o caso do *Unix Standalone* mas para Unix.

Servidor Unix/ Cliente PC indica que o servidor funciona num ambiente Unix e o cliente é um computador normal. Servidor NT/ Cliente PC representa o servidor NT, ambiente Windows. A conexão à base de dados indica uma ligação a uma base de dados existente.

A análise dos dados exibidos na tabela anterior, tabela 6, permite concluir que todas as ferramentas de *Data Mining* estudadas apresentam ligação à base de dados e que o ODM é suportado pelo maior número de plataformas analisadas.

Automação

Ao analisar as ferramentas foi verificado que todas possuem programação visual, podendo apresentar métodos que auxiliam a programação visual, como por exemplo, caixas de diálogo, linguagem de programação e assistentes de ajuda. Após a recolha de informação relativamente ao método de automação é obtida a tabela 7 (Cruz, 2007; Elder et al., 1998).

Tabela 7 – Automação

Ferramenta	Método de Automação	Texto livre de anotações de passos
<i>Weka</i>	Programação Visual, Linguagem de Programação	?
<i>ODM</i>	Programação Visual, Linguagem de Programação, <i>Wizards</i>	?
<i>EM</i>	Programação Visual, Linguagem de Programação, Caixa de diálogo	√
<i>Clementine</i>	Programação Visual, Linguagem de Programação, <i>Wizards</i>	√

Através dos dados extraídos da tabela 7 é verificado que as ferramentas analisadas apresentam métodos auxiliares em relação à programação visual, destacando que todas exibem linguagem de programação. O ODM e *Clementine* também disponibilizam *wizards* e o EM apresenta caixas de diálogo. É importante referir que não foi possível recolher informação em relação ao texto livre de anotações de passos nos *softwares Weka* e ODM, sendo assinalado com o símbolo '?'.

Técnicas de DM

As ferramentas selecionadas exibem as seguintes técnicas de DM (Reis, 2009; Coelho, 2005).

Tabela 8 - Técnicas de DM

Ferramenta	Associação	Classificação	Clustering	Regressão	Sequências
<i>Weka</i>	√	√	√	√	
<i>ODM</i>	√	√	√	√	
<i>EM</i>	√	√	√	√	
<i>Clementine</i>	√	√	√	√	√

Todas as ferramentas estudadas possuem as técnicas de DM, classificação, regras de associação, regressão e *clustering*. O *Clementine* é a única ferramenta que apresenta a técnica de sequências.

Esta análise não é muito significativa para o estudo porque todas as ferramentas analisadas apresentam as técnicas que se pretendem utilizar neste trabalho, classificação, regras de associação e *clustering*.

Algoritmos

A tabela 9 mostra os algoritmos apresentados pelas ferramentas estudadas (Elder et al., 1998; Pozo, 2010).

Tabela 9 – Algoritmos

Ferramenta	Árvores de Decisão	Linear/ Estatística	<i>K- Means</i>	Regras de Associação
<i>Weka</i>	√		√	√
<i>ODM</i>	√	√	√	√
<i>EM</i>	√	√	√	√
<i>Clementine</i>	√	√	√	√

As árvores de decisão representam um conjunto de dados divididos segundo uma classificação de registos de acordo com o valor exibido no atributo objetivo. O algoritmo Linear/ Estatística

pretende analisar estatisticamente os dados. O *K-Means* prevê valores de conhecimentos de registos num conjunto de dados, sendo baseado na combinação dos valores dos k-registos mais próximos, mais semelhantes do conjunto de dados históricos. E as regras de associação geram todas as associações entre os itens de transações que impliquem a presença de outros itens (Pereira, 2002).

Ao analisar os resultados da tabela 9 é averiguado que as ferramentas estudadas são muito similares entre si em relação aos algoritmos, destacando o *Weka* que não exibe o algoritmo de análise estatística dos dados.

Visualização de dados

A tabela 10 representa uma comparação entre as visualizações possíveis para as ferramentas estudadas (Elder et al., 1998).

Tabela 10 – Visualização

Ferramenta	Histogramas	Gráficos circulares	Gráficos de linhas	Áreas de decisão de classificação	Gráficos dispersão
<i>Weka</i>	?	?	?	√	?
<i>ODM</i>	√	?	?	?	?
<i>EM</i>	√		√		√
<i>Clementine</i>	√	√	√	√	√

Os modos de visualização analisados são (Pereira, 2002):

- Histogramas: gráfico de barras com indicação da quantidade de cada campo, sendo fácil de compreensão.
- Gráficos circulares: gráfico em forma circular.
- Gráficos de linhas: gráfico que representa os dados segundo uma linha.
- Áreas de decisão de classificação: gráficos que reagrupam os pontos segundo características semelhantes (*clustering*).
- Gráficos de dispersão: representam pontos segundo a relação existente entre eles.

A análise dos resultados obtidos permite destacar a ferramenta *Clementine* como sendo a detentora de maior número de formas de visualização dos dados. É conveniente referir a

ausência de informação em relação a algumas ferramentas é devido à dificuldade na recolha da informação, sendo representado pelo símbolo "?".

Outros critérios

Após a análise dos critérios acima mencionados, foi considerado interessante estudar outros que apresentam um papel significativo na seleção da ferramenta de *Data Mining*. Uma das características importantes a analisar é a facilidade de utilização da ferramenta, possuir uma interface simples e amigável, sendo um dos critérios fulcrais nesta seleção. Outro aspeto relevante consiste em investigar quais as ferramentas comerciais e gratuitas, o investimento da implementação da ferramenta pode inviabilizar a sua escolha. O último critério consiste na capacidade de processamento, este pode representar um fator condicionante caso seja necessário analisar grande quantidade de dados e a ferramenta não apresente a capacidade de processamento suficiente.

Desta forma, foi criada uma tabela que contém os dados relativos aos critérios citados anteriormente (Rezende et al., 1998; Goebel et al., 1999).

Tabela 11 – Outros critérios

Ferramenta	Interface simples	Comercial / Gratuito	Capacidade de processamento
<i>Weka</i>	√	Gratuito	
<i>ODM</i>	√	Comercial	√
<i>EM</i>	√	Comercial	√
<i>Clementine</i>	√	Comercial	√

Através dos resultados obtidos da tabela 11 é verificado que as ferramentas são semelhantes relativamente aos critérios analisados, com exceção do *software Weka* que é o único gratuito e o que não exhibe grande capacidade de processamento de dados.

3.6 Resumo

Após o estudo das quatro ferramentas de *Data Mining* foi considerado conveniente realizar uma síntese das vantagens e desvantagens de cada uma. A tabela 12 representa as forças e fraquezas de cada ferramenta analisada (Elder et al., 1998).

Tabela 12- Vantagens e desvantagens de ferramentas DM

Ferramenta	Vantagens	Desvantagens
<i>Weka</i>	<ul style="list-style-type: none"> - Gratuito - Interface gráfica amigável - Portabilidade 	<ul style="list-style-type: none"> - Baixa capacidade de processamento de dados - Formato específico de ficheiros de dados (ARFF)
<i>ODM</i>	<ul style="list-style-type: none"> - Integração com os outros produtos Oracle - Fácil de usar 	<ul style="list-style-type: none"> - Preço
<i>EM</i>	<ul style="list-style-type: none"> - Interface gráfica (fluxo de dados) - Quantidade de algoritmos - Variedade de ferramentas de análise estatística 	<ul style="list-style-type: none"> - Difícil de usar
<i>Clementine</i>	<ul style="list-style-type: none"> - Interface gráfica - Quantidade de algoritmos - Opera com grande quantidade de dados 	<ul style="list-style-type: none"> - Preço

O *Weka* exibe interface gráfica amigável e portabilidade, possibilitando o suporte de vários sistemas operativos. Esta ferramenta é gratuita, sendo a sua maior vantagem e é distinta das restantes ferramentas. As suas limitações consistem na baixa capacidade de processamento de dados e no formato específico de ficheiros de dados (arquivo ARFF).

O *Oracle Data Mining* é um *software* bastante poderoso que exibe como vantagens a sua integração com outros produtos Oracle e a facilidade de uso devido à presença de *wizards*. Contudo esta ferramenta necessita de um grande investimento, elevado custo de aquisição.

O *Enterprise Miner* apresenta diversas vantagens, este sistema possui uma interface gráfica atrativa que permite representar todos os passos do projeto num fluxo de dados, detém grande quantidade de algoritmos para manusear os dados e variadas ferramentas de análise estatística. Mas, a sua utilização é demasiada complexa, sendo necessário possuir muitos

conhecimentos para utilizar esta ferramenta corretamente devido ao elevado número de funcionalidades que oferece.

O *Clementine* consiste numa ferramenta bastante completa que representa todas as fases do processo de KDD. Os seus pontos fortes são a interface gráfica que permite criar modelos de DM de forma intuitiva e a simplicidade da sua utilização. Como limitação é indicado o significativo investimento inerente a este *software*.

Após o estudo comparativo entre as ferramentas não se pode afirmar qual é a melhor ferramenta a ser utilizada, cada uma possui características distintas exibindo pontos positivos e negativos, a escolha de uma ferramenta está diretamente relacionada com os objetivos que se pretende alcançar.

De forma geral, todas as ferramentas de *Data Mining* apresentam características similares em relação aos critérios analisados anteriormente, por exemplo, interface simples, acesso à base de dados, técnicas DM, entre outros. Contudo duas das ferramentas, o *Weka* e *Clementine*, são destacadas das restantes. O *Weka* é única ferramenta gratuita analisada e que apresenta menor capacidade de processamento de dados e necessita de formato específico de ficheiros de entrada de dados. O *Clementine* é um produto comercial que apresenta diversas técnicas de DM, classificação, regras de associação, *clustering*, regressão e sequências e disponibiliza grande variedade de formas de visualização dos resultados dos modelos de DM.

A seleção recai no *software Clementine* porque consiste numa ferramenta que detém todas as técnicas de *Data Mining* essenciais a este trabalho, permitindo recorrer a outras técnicas de DM caso seja necessário, como regressão e sequências. O *Clementine* permite operar com grandes quantidades de dados e é dotado de uma interface gráfica simples e amigável. Apesar de ser um *software* comercial, o autor deste trabalho tem a licença de utilização desta ferramenta fornecida pelo departamento de Informática do Instituto Superior de Engenharia do Porto e também possui conhecimentos relativamente ao uso desta ferramenta. O anexo C apresenta uma descrição mais detalhada acerca desta ferramenta.

4 PCMAT - Mathematics Collaborative Educational System

"Assumir uma atitude responsável perante o futuro sem uma compreensão do passado é ter um objetivo sem conhecimento. Compreender o passado sem um comprometimento com o futuro é conhecimento sem objetivo."

Ronald T. Laconte

Neste capítulo são apresentados os resultados obtidos relativamente à avaliação da adequação dos modelos e funcionalidades da plataforma PCMAT, aplicada ao ensino de matemática no ensino básico.

O PCMAT é uma ferramenta colaborativa e adaptativa de aprendizagem baseada na avaliação progressiva de conteúdos matemáticos. Este sistema utiliza a metodologia construtivista, avaliando o conhecimento do utilizador e apresentando conteúdos e atividades adaptados às características e estilo de aprendizagem de cada aluno (Martins et al., 2008).

Este capítulo tem como objetivo analisar a plataforma PCMAT, abordando os seus requisitos, funcionalidades, objetivos e os dados armazenados. Esta análise permite compreender o funcionamento do PCMAT e identificar as exigências que este tipo de negócio obriga, permitindo que seja possível extrair tendências do comportamento dos alunos, regras e padrões, através dos modelos criados no processo de *Data Mining* para melhorar e personalizar o ensino fornecido por esta ferramenta.

4.1 Compreensão do negócio

Esta fase consiste na compreensão do negócio, conhecimento de como funciona o negócio, podendo abranger políticas, objetivos, requisitos, compromissos, obrigações, entre outros. O conhecimento das regras de negócio será utilizado na definição do modelo de análise, sendo traçado um plano inicial para alcançar os objetivos pretendidos (Silva, 2007).

O PCMAT é uma plataforma adaptativa de aprendizagem baseada na avaliação progressiva de Matemática, disponibilizando os conteúdos e atividades consoante o modelo de cada aluno. A primeira versão do protótipo foi implementada e avaliada nos processos de aprendizagem da Matemática, nomeadamente no tópico "Proporcionalidade Direta". Este tema está incluído no programa de Matemática do 6º ano e também é abordado no 7º ano de escolaridade. Esta implementação foi realizada em duas escolas básicas, escola EB2/3 Dr. Ferreira de Almeida, de Santa Maria da Feira e a escola EB2/3 S. Lourenço, de Ermesinde (Martins et al., 2008).

O estudo decorreu ao longo de três semanas, em cinco turmas, uma na escola de Santa Maria da Feira e as restantes quatro na escola de Ermesinde. A primeira turma apresentava 23 alunos e as restantes turmas eram constituídas por 18, 19, 20 e 21 alunos, respetivamente. A idade dos alunos estava compreendida entre os 12 e 14 anos (Martins et al., 2008).

Nenhum dos alunos detinha experiência no uso de SHAE. No entanto, mais de 85% dos alunos estavam familiarizados com o uso de computadores pessoais, nomeadamente para navegar na internet, frequência de redes sociais e para jogar (Martins et al., 2008).

Cada turma foi dividida em dois grupos: experimental e controlo. Os alunos que pertenceram ao primeiro grupo utilizaram o PCMAT no processo de aprendizagem e os alunos do segundo grupo utilizaram o método tradicional de ensino (Martins et al., 2008).

O processo de aprendizagem consistiu em diversas etapas. Inicialmente, os alunos do grupo experimental foram submetidos a um questionário VARK (Visual, Auditivo, Ler/Escriver e Prático) que avaliou o estilo de aprendizagem inicial de cada aluno, identificando um dos três estilos: teórico, visual e prático. Este estilo de aprendizagem foi utilizado no PCMAT com o intuito de personalizar e adaptar a aprendizagem de forma a que o aluno obtenha uma aprendizagem mais adequada ao seu modelo, melhorando os resultados obtidos.

Os alunos de cada turma realizaram um teste diagnóstico, avaliando o nível de conhecimentos relativos ao tema "Proporcionalidade Direta". O anexo D apresenta o teste diagnóstico.

Este teste abrangeu diversos conceitos relativos à unidade citada anteriormente, sendo classificados em dois grandes grupos:

- A - Razão;
- B - Proporções.

O grupo A foi dividido em vários conceitos relacionados com o assunto razão:

- A1 - Razões equivalentes;
- A2 - Noções de antecedente e conseqüente;
- A3 - Leitura de uma razão;
- A4 - Aplicações do conceito de razão.

O grupo B foi distribuído por diversos conceitos referentes ao assunto proporções:

- B1 - Meios e extremos;
- B2 - Leitura de proporções;
- B3 - Propriedade das proporções;
- B4 - Regras de três simples;
- B5 - Método de redução à unidade.

Após a realização do teste diagnóstico, os alunos iniciaram o estudo da unidade mencionada anteriormente, recorrendo ao uso do PCMAT ou ao método tradicional de ensino. O sucesso desta aprendizagem foi verificado com a realização do teste final, onde a evolução dos conhecimentos adquiridos por cada aluno é analisada. Finalmente, os alunos foram submetidos a outro questionário VARK para aferir o estilo de aprendizagem na fase final, analisando e comparando a evolução do estilo de aprendizagem de cada aluno ao longo do processo de aprendizagem.

Todos os dados foram armazenados num ficheiro Microsoft Excel, tendo sido importado para uma aplicação de tratamento estatístico de dados, *Clementine*.

Através do PCMAT, a evolução da aprendizagem dos alunos varia consoante diversos fatores, desde a idade até ao estilo de aprendizagem de cada aluno. Prever esta evolução passa,

essencialmente por efetuar uma análise cuidada da informação disponível por esta ferramenta. Para que na construção de modelos criados, durante o processo de *Data Mining*, sejam encontrados factos relevantes e generalizáveis é necessário extrair regras e padrões, que permitam a identificação de características e respetivos contributos em relação aos resultados obtidos no processo de aprendizagem. A previsão do comportamento de alunos neste processo permite melhorar e adaptar o ensino disponível pelo PCMAT.

4.2 Compreensão dos dados

A compreensão de dados visa a familiarização com os dados fornecidos, identificando problemas de qualidade de dados, definindo atributos de interesse, granularidades, entre outros aspetos (Silva, 2007).

Os dados fornecidos representam a informação relativa à evolução dos alunos a nível da disciplina de Matemática, especificamente a unidade de "Proporcionalidade Direta". Esta informação está distribuída por turmas distintas de 7º e 8º ano de escolaridade.

Nenhum dos alunos teve experiências anteriores relativamente ao uso de SHAE, mas cerca de 85% dos alunos estão familiarizados com o uso de computadores, nomeadamente jogos e uso de internet para pesquisa de informação (Martins et al., 2008).

Cada turma possuía dois grupos: experimental e controlo. O primeiro grupo recorreu ao PCMAT para auxiliar o processo de aprendizagem e no segundo grupo foi utilizado o método tradicional de ensino. A existência destes grupos permite comparar duas realidades distintas de ensino. A distribuição dos alunos por grupo tem em consideração as seguintes condições (Martins et al., 2008):

- se possível cada grupo apresentar o mesmo número de alunos aproximadamente;
- se possível cada grupo apresentar o mesmo número de alunos excelentes. Para isso os alunos são submetidos a um teste de diagnóstico relativo à unidade de "Proporcionalidade Direta". Para mais informações relativas ao teste diagnóstico consultar o anexo D;
- cada grupo apresentar uma quantidade similar de alunos com a mesma preferência de aprendizagem. Para este efeito, os alunos são submetidos ao questionário VARK que identifica as preferências de aprendizagem de cada aluno na receção e tratamento da

informação. Este modelo descreve três estilos de aprendizagem: teórico (aprende a ler e ouvir), visual (aprende a ver) e prático (aprende a fazer);

- cada grupo exibir uma distribuição semelhante relativamente ao sexo dos alunos.

Como foi referido anteriormente a informação do progresso matemático dos alunos é distribuída por diversas turmas. De seguida, é efetuada uma descrição de cada turma, focando as várias fases do processo de aprendizagem.

A turma 7º A apresentava 20 alunos, 8 raparigas e 12 rapazes. O grupo experimental possuía 10 alunos e os restantes alunos estavam distribuídos pelo grupo de controlo.

O grupo experimental realizou o questionário VARK, identificando o estilo de aprendizagem inicial de cada aluno. Os resultados obtidos neste questionário são apresentados seguidamente.

Tabela 13 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º A

Estilo de aprendizagem	Percentagem de alunos (%)
(1) – Teórico	20
(2) – Visual	50
(3) – Prático	30

Através dos dados da tabela 13 é verificado que 50% dos alunos do grupo desta turma exibe o estilo de aprendizagem visual.

Todos os alunos desta turma realizaram o teste diagnóstico e a maioria dos resultados obtidos neste teste foram negativos, sendo registando uma média de 15 valores percentuais no grupo de controlo, 29 valores percentuais no grupo experimental e no total dos alunos foi obtido uma média de 23 valores percentuais.

Tabela 14 - Classificações do teste diagnóstico do 7º A

Classificação	Porcentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	0	20	10
Negativa	80	80	80
Ausência	20	0	10

O conceito que exibiu maior taxa de respostas certas foi o B3 (propriedade das proporções) e os conceitos que obtiveram menor taxa de respostas certas foram o A3 (leitura de uma razão) e B2 (leitura de proporções). Foi conveniente mencionar a existência de um aluno com ensino especial nesta turma. O tipo de ensino, normal ou especial, pode influenciar o processo de aprendizagem dos alunos, por isso esta característica foi considerada um fator importante de análise e comparação da evolução de aprendizagem.

No fim do estudo, os alunos realizaram um teste com o objetivo de avaliar todos os conceitos abordados pelos grupos A e B, Razão e Proporções respetivamente. Desta forma, foi realizada uma análise e comparação entre os resultados obtidos na fase inicial (diagnóstico) e final, examinando o percurso de cada aluno. As classificações obtidas no teste final demonstraram um aumento significativo, no grupo experimental a média rondou os 62 valores percentuais, no grupo de controlo foram obtidos 27 valores percentuais e, no total, foi obtido uma média de cerca de 44 valores percentuais. De seguida, são exibidas as classificações obtidas nesta fase de aprendizagem.

Tabela 15 - Classificações do teste final do 7º A

Classificação	Porcentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	20	70	50
Negativa	70	30	45
Ausência	10	0	5

A informação contida na tabela 15 destaca que metade do número total de alunos do 7º A alcançaram classificações positivas.

Em relação aos conceitos, foi registado que o B5 (método de redução à unidade) obteve menor número de respostas corretas e o A2 (noção de antecedente e consequente) atingiu o maior número de respostas certas.

Os alunos do grupo experimental efetuaram um questionário para aferir o seu estilo de aprendizagem no fim deste processo - questionário VARK, que pretende analisar e comparar a evolução do estilo de aprendizagem de cada aluno. Os resultados deste questionário são apresentados seguidamente.

Tabela 16 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º A

Estilo de aprendizagem	Nº total de alunos
(1) - Teórico	20
(2) - Visual	60
(3) - Prático	20

Os dados da tabela 16 evidenciam um aumento do número de alunos com estilo de aprendizagem visual, uma diminuição do número de alunos com estilo prático e a quantidade de alunos com estilo de aprendizagem teórico foi mantida igual. Em suma, o processo de aprendizagem dos alunos nesta turma proporcionou a alteração de estilo de aprendizagem de alguns alunos, nomeadamente alguns alunos passaram a deter o estilo de aprendizagem visual.

A turma 7º B apresentava 21 alunos, 12 do sexo feminino e 8 do sexo masculino. Os alunos foram distribuídos por dois grupos, o grupo experimental foi constituído por 11 alunos e o grupo de controlo apresentava 10 alunos.

Inicialmente, o grupo experimental realizou o questionário VARK, para avaliar o estilo de aprendizagem inicial de cada aluno. Deste questionário foram obtidos os seguintes resultados:

Tabela 17 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º B

Estilo de aprendizagem	Percentagem de alunos (%)
-------------------------------	----------------------------------

(1) - Teórico	18
(2) - Visual	46
(3) - Prático	36

Através do conteúdo da tabela 17 é averiguado que 46% dos alunos apresentaram estilo de aprendizagem visual.

Nesta fase inicial do processo foi realizado um teste diagnóstico aos alunos, todos os conceitos relativos ao grupo A - Razão e B - Proporções foram avaliados, identificando os seus conhecimentos atuais. Os resultados obtidos neste teste foram negativos, no grupo experimental foi atingido uma média de 29 valores percentuais, 24 valores percentuais no grupo de controlo e, no total, foi alcançada uma média de 26 valores percentuais.

Tabela 18 - Classificações do teste diagnóstico do 7º B

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	20	9	14
Negativa	60	91	76
Ausência	20	0	10

Neste teste diagnóstico foram avaliados os conceitos abordados, foi verificado que o conceito que apresentou maior número de respostas corretas é o B3 (propriedade das proporções) e os conceitos que apresentaram menor taxa de respostas certas foram o B2 (leitura de proporções) e B5 (método de redução à unidade). Esta turma apresentava um aluno com necessidades especiais, ensino especial. O tipo de ensino foi considerado uma característica interessante de análise para detetar se esta condiciona a evolução de aprendizagem do aluno.

Todos os alunos foram submetidos a um teste final, sendo abordados todos os conceitos dos dois grupos citados acima, analisando e comparando as classificações alcançadas inicialmente e nesta última fase, examinando a aprendizagem de cada aluno. As classificações deste teste

evidenciaram uma melhoria considerável, no grupo experimental foi atingido uma média de 62 valores percentuais, no grupo de controlo foi alcançado a média de 45 valores percentuais e, no total dos alunos, foi obtida uma média de 54 valores percentuais. Seguidamente, são apresentados os resultados referentes a este teste.

Tabela 19 - Classificações do teste final do 7º B

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	20	82	52
Negativa	80	18	48
Ausência	0	0	0

Em relação ao conceitos, foi averiguado que o B5 (método de redução à unidade) foi o conceito que exhibe maior taxa de respostas erradas e o A2 (noções de antecedente e consequente) possuía maior número de respostas corretas.

Na última fase, o grupo experimental realizou novamente o questionário VARK, identificando o estilo de aprendizagem de cada aluno, avaliando a evolução do estilo de aprendizagem ao longo deste percurso de aprendizagem. Os resultados extraídos deste questionário evidenciaram que a maioria dos alunos possuem o estilo de aprendizagem visual, cerca de 55% dos alunos.

Tabela 20 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º B

Estilo de aprendizagem	Percentagem de alunos (%)
(1) - Teórico	18
(2) - Visual	55
(3) - Prático	27

Através da tabela 20 é verificado um acréscimo do número de alunos com estilo de aprendizagem visual, uma diminuição do número de alunos com estilo de aprendizagem prático e o número de alunos com estilo de aprendizagem teórico foi mantido constante ao

identificado na fase inicial. Este processo de aprendizagem proporciona a alteração do estilo de aprendizagem de alguns alunos, favorecendo o estilo de aprendizagem visual.

A turma 7º F foi constituída por 18 alunos, sendo 7 do sexo feminino e 11 do masculino. Cerca de 9 alunos aprenderam Matemática recorrendo ao PCMAT, grupo experimental, e os restantes alunos fizeram parte do grupo de controlo, utilizaram o método tradicional de ensino.

Na fase inicial, os alunos do grupo experimental foram submetidos ao questionário VARK, identificando a sua preferência de aprendizagem inicial. De seguida é apresentada uma tabela que indica os resultados obtidos neste questionário.

Tabela 21 - Estilos de aprendizagem inicial dos alunos do grupo experimental do 7º F

Estilo de aprendizagem	Percentagem de alunos (%)
(1) - Teórico	22
(2) - Visual	56
(3) - Prático	22

Após a observação dos dados da tabela 21, foi averiguado que cerca de 56% dos alunos apresentava estilo de aprendizagem visual.

Todos os alunos da turma realizaram o teste de diagnóstico, sendo avaliados os conceitos relativos aos dois grupos de conceitos, A e B, Razão e Proporções respetivamente. As classificações obtidas neste teste foram negativas, sendo registado uma média de 25 valores percentuais no grupo experimental, uma média de 10 valores percentuais no grupo de controlo e, no total dos alunos, foi obtida uma média de 17 valores percentuais.

Tabela 22 - Classificações do teste diagnóstico do 7º F

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total

Positiva	0	0	0
Negativa	78	100	90
Ausência	22	0	10

Em relação à avaliação dos conceitos, foi verificado que o conceito B3 (propriedade das proporções) atingiu o maior número de respostas certas e o conceito B2 (leitura de proporções) apresentou menor número de respostas certas. Não esquecendo de mencionar a existência de 2 alunos que apresentaram ensino especial. O tipo de ensino, normal ou especial, foi considerado uma característica vantajosa para a análise e comparação do progresso da aprendizagem dos alunos com necessidades distintas.

Após o estudo do tema lecionado, foi registado um aumento considerável nos resultados do teste final, no grupo experimental foi atingida uma média de 58 valores percentuais, no grupo de controlo foram obtidos 29 valores percentuais e, no total dos alunos, foi atingido uma média de 44 valores percentuais.

Tabela 23 - Classificações do teste final do 7º F

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	22	67	44,5
Negativa	56	33	44,4
Ausência	22	0	11

A tabela 23 demonstra que 45% dos alunos desta turma alcançaram resultados positivos.

Relativamente à avaliação de conceitos, foi registado que o conceito A2 (noção de antecedente e conseqüente) foi o que apresentou maior número de respostas certas e o conceito B5 (método de redução à unidade) foi o que exibiu menor número de respostas corretas.

O grupo experimental foi submetido a outro questionário VARK para analisar a evolução do estilo de aprendizagem de cada aluno ao longo da aprendizagem, verificando como a preferência de aprendizagem influencia os resultados obtidos ao longo deste processo.

Tabela 24 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º F

Estilo de aprendizagem	Percentagem de alunos
(1) - Teórico	11
(2) - Visual	67
(3) - Prático	22

Os resultados extraídos do questionário evidenciaram que a maioria dos alunos da turma, cerca de 67%, exibiu o estilo de aprendizagem visual. Foi averiguado também um acréscimo no número de alunos com estilo de aprendizagem visual, uma redução de alunos com estilo de aprendizagem teórico e a quantidade de alunos com estilo de aprendizagem prático foi mantido equivalente à fase inicial. Ao longo deste processo, alguns alunos alteraram o seu estilo de aprendizagem, foi registado maior número de alunos com estilo de aprendizagem visual e com melhores resultados nas classificações finais.

A turma 7º G apresentava 19 alunos, 12 raparigas e 7 rapazes. Esta turma foi dividida em 10 alunos no grupo experimental e os restantes 9 alunos pertenciam ao grupo de controlo.

Os alunos que constituíram o grupo experimental realizaram o questionário VARK, tendo como objetivo identificar o estilo de aprendizagem inicial dos alunos.

Tabela 25 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 7º G

Estilo de aprendizagem	Percentagem de alunos (%)
(1) - Teórico	10

(2) - Visual	40
(3) - Prático	40

Nesta turma é verificada a existência do mesmo número de alunos com estilo de aprendizagem visual e prático, cerca de 40% de alunos.

Todos os alunos desta turma foram submetidos a um teste diagnóstico para avaliar os seus conhecimentos iniciais relativos à unidade de "Proporcionalidade Direta", abordando todos os conceitos do grupo A - razão e B - Proporções. Este teste tem como objetivo analisar e comparar os conhecimentos inicial com os conhecidos adquiridos após o estudo da matéria lecionada. A maioria dos resultados do teste foram negativos, o grupo experimental apresentou uma média de 32 valores percentuais, os alunos do grupo de controlo exibiram 9 valores percentuais de média e, no total, foi registada uma média de 21 valores percentuais.

Tabela 26 - Classificações do teste diagnóstico do 7º G

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	0	20	16
Negativa	78	70	68
Ausência	22	10	16

O conceito que exibiu maior número de respostas certas foi o A2 (noção de antecedente e conseqüente) e o conceito B5 (método de redução à unidade) apresentou o menor número de respostas certas. Foi conveniente mencionar a existência de um aluno com necessidades especiais nesta turma. Esta característica foi considerada importante para analisar e comparar a evolução da aprendizagem com os restantes alunos que não exibem necessidades especiais, sendo verificado a influência deste fator no processo de aprendizagem.

Após o estudo de "Proporcionalidade Direta", os alunos realizaram um teste final, avaliando todos os conceitos abordados pelos dois grupos, A - Razão e B - Proporções. Nesta fase é pretendido efetuar uma análise e comparação entre os resultados obtidos no teste inicial e

final, avaliando o progresso de aprendizagem de cada aluno. Os resultados deste teste demonstraram uma melhoria apreciável, os alunos do grupo experimental apresentaram uma média de 53 valores percentuais, o grupo de controlo obteve uma média de 35 valores e, no total, foi atingido uma média de 44 valores percentuais.

Tabela 27 - Classificações do teste final do 7º G

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	33	70	53
Negativa	67	30	47
Ausência	0	0	0

O conceito que apresentou maior número de respostas erradas é B5 (método de redução à unidade) e o conceito que exibiu maior número de respostas certas é A2 (noção de antecedente e consequente).

Os alunos pertencentes ao grupo experimental realizaram o questionário para avaliar a evolução do estilo de aprendizagem de cada aluno - questionário VARK. Os resultados deste questionário são apresentados de seguida.

Tabela 28 - Estilo de aprendizagem final dos alunos do grupo experimental do 7º G

Estilo de aprendizagem	Percentagem de alunos (%)
(1) - Teórico	30
(2) - Visual	40
(3) - Prático	30

A tabela 28 regista um acréscimo no número de alunos com estilo de aprendizagem teórico, um decréscimo do número de alunos com estilo de aprendizagem prático e foi mantido o mesmo número de alunos com estilo de aprendizagem visual. Ao contrário do que se registou nas turmas anteriores, o processo de aprendizagem nesta turma proporcionou o aumento do número de alunos com o estilo de aprendizagem teórico, mas mesmo assim a maioria dos alunos exibe o estilo de aprendizagem visual, cerca de 40% dos alunos.

A turma 8º F era constituída por 23 alunos, 11 raparigas e 12 rapazes. Cerca de 12 alunos pertenciam ao grupo experimental e os restantes alunos, 11 alunos, representavam o grupo de controlo.

Na fase inicial do processo de aprendizagem, os alunos do grupo experimental realizaram o questionário VARK para aferir a preferência de aprendizagem inicial de cada aluno, sendo observado os seguintes resultados.

Tabela 29 - Estilo de aprendizagem inicial dos alunos do grupo experimental do 8º F

Estilo de aprendizagem	Nº total de alunos
(1) - Teórico	16
(2) - Visual	42
(3) - Prático	42

A maioria dos alunos desta turma foi distribuída pelos estilos de aprendizagem visual e prático, cerca de 42% dos alunos. Foi considerado conveniente mencionar que esta turma não realizou o teste diagnóstico, não sendo avaliados os conhecimentos iniciais relativos à unidade de "Proporcionalidade Direta".

Após o estudo da unidade citada anteriormente, todos os alunos realizaram um teste final para avaliar os conhecimentos adquiridos, foi obtida uma média de 57 valores percentuais no grupo experimental, média de 43 valores percentuais no grupo de controlo e, no total, foi atingida uma média de 50 valores percentuais.

Tabela 30 - Classificações do teste final do 8º F

Classificação	Percentagem de alunos (%)		
	Grupo Controlo	Grupo Experimental	Total
Positiva	27	50	40
Negativa	73	50	60
Ausência	0	0	0

Os dados relativos à avaliação dos conceitos indicaram que o conceito que apresentava maior taxa de respostas certas foi o B1 (meios e extremos) e o conceito que possuía menor número de respostas corretas foi B5 (método de redução à unidade).

Nesta fase final, o grupo experimental foi submetido a outro questionário VARK para identificar o estilo de aprendizagem final de cada aluno, analisando e comparando o estilo de aprendizagem e averiguando se este fator influencia o processo de aprendizagem do aluno. Seguidamente, são apresentados os estilos de aprendizagem obtidos no questionário.

Tabela 31 - Estilo de aprendizagem final dos alunos do grupo experimental do 8º F

Estilo de aprendizagem	Percentagem de alunos (%)
(1) - Teórico	16
(2) - Visual	59
(3) - Prático	25

A informação da tabela 31 evidencia um aumento do número de alunos com estilo de aprendizagem visual, uma diminuição do número de alunos com estilo de aprendizagem prático e o número de alunos com estilo de aprendizagem teórico foi mantido igual à fase inicial. Em suma, o processo de aprendizagem desta turma proporcionou a alteração de estilo de aprendizagem de alguns alunos, sendo registado maior número de alunos com estilo de aprendizagem visual e com melhores classificações finais.

Após o processo de aprendizagem da unidade de "Proporcionalidade Direta", alguns alunos do 7º e 8º ano de escolaridade foram questionados sobre a usabilidade da ferramenta PCMAT. De forma geral, a maioria dos alunos focaram alguns aspetos como pontos fortes do sistema: organização, bom grafismo e facilidade de acesso a conteúdos e atividades, considerando o PCMAT bastante útil na aprendizagem da disciplina de Matemática, indicando que facilita o estudo fora e dentro das aulas e que promove a melhoria de resultados. Por fim, os alunos recomendaram o uso de SHAE nas outras disciplinas, o que demonstra o sucesso da implementação do PCMAT na disciplina de Matemática. O anexo E apresenta informações relativas à análise de usabilidade do PCMAT.

4.3 Resumo

Este capítulo tem como principal objetivo estudar a plataforma PCMAT, analisando os seus requisitos, funcionalidades e finalidades e avaliando a adequação dos modelos e funcionalidades disponíveis do PCMAT.

O PCMAT é uma ferramenta de aprendizagem baseada na avaliação progressiva de conteúdos matemáticos em escolas do ensino básico, adaptando e personalizando a aprendizagem de cada aluno consoante o modelo do aluno. Esta aplicação armazena informação relativa ao processo de aprendizagem, nomeadamente do tema "Proporcionalidade Direta", de alunos de turmas de 7º e 8º ano de escolaridade, com idades entre os 12 e 14 anos.

Cada turma foi dividida em dois grupos, experimental e controlo. O primeiro grupo utilizou a ferramenta PCMAT para aprender a disciplina de Matemática e o segundo grupo recorreu ao método tradicional de ensino. A existência destes grupos permite estudar a evolução da aprendizagem dos alunos, tendo em conta as distintas realidades, o uso do PCMAT ou método tradicional de ensino.

O processo de aprendizagem consistiu em diversas etapas. Na fase inicial, os alunos do grupo experimental realizaram um questionário VARK, identificando o estilo de aprendizagem inicial de cada aluno: teórico, visual e prático. Todos os alunos de cada turma foram submetidos a um teste de diagnóstico, aferindo o nível de conhecimentos iniciais referentes à unidade lecionada.

Em cada turma, os dois grupos de alunos, experimental e controlo, iniciaram o estudo dos conteúdos matemáticos recorrendo ao PCMAT e ao método tradicional de ensino básico. O sucesso deste processo é avaliado com a realização do teste final, analisando a evolução dos conhecimentos adquiridos por cada aluno. Na fase final, os alunos do grupo experimental

foram submetidos a outro questionário VARK, aferindo a preferência do estilo de aprendizagem final, examinando e comparando o percurso de aprendizagem dos alunos consoante o progresso de cada aluno.

Em suma, a evolução de aprendizagem matemática dos alunos é condicionada por diversos fatores, tipo de ensino, estilo de aprendizagem, idade, entre outros. Para avaliar esta evolução é necessário proceder a uma análise cuidada da informação contida no PCMAT e avaliar os factos significativos e gerais, padrões e regras, extraídos dos modelos criados durante o processo de *Data Mining*, de forma a promover a melhoria e adaptação do ensino disponível pelo PCMAT.

5 Modelo de dados

"Aquilo que os homens de facto querem não é o conhecimento mas a certeza."

Bertrand Russell

Na fase de elaboração do modelo de dados são selecionadas e aplicadas diversas técnicas para criar um modelo de dados e os seus parâmetros são configurados para obter melhores previsões. Normalmente, existem várias técnicas para o mesmo tipo de problema de *Data Mining*, algumas delas com requisitos específicos na forma dos dados, sendo frequentemente necessário iterar algumas etapas (Rodrigues, 2008).

O objetivo nesta fase do projeto é elaborar um modelo que seja capaz de efetuar previsões sobre o progresso de aprendizagem dos alunos em relação à disciplina de matemática, descobrindo factos expressivos e gerais, de forma a contribuir para a melhoria dos resultados obtidos no processo de aprendizagem dos alunos.

Na construção do modelo de dados é necessário efetuar uma preparação dos dados, os dados iniciais são transformados em dados que possam ser utilizados em ferramentas de modelação. Após este tratamento de dados, é efetuada uma exploração gráfica que tem como intuito a análise dos dados, detetando comportamentos não considerados normais, tais como: anomalias, dados corrompidos e incompatíveis. Após a compreensão dos dados é elaborado um modelo de dados recorrendo a técnicas de *Data Mining* que permitam extrair conhecimento. Por fim, é avaliado o desempenho do modelo de dados criado tendo em conta vários fatores.

5.1 Preparação de dados

A fase de preparação de dados foca todas as atividades de transformação dos dados iniciais em informação que possa ser utilizada nas ferramentas de modelação, verificando e removendo anomalias, melhorando a qualidade dos dados. Estas atividades podem ser realizadas diversas vezes, sem qualquer ordem pré-definida (Oliveira et. al, 2004).

Existem diversas operações de transformação que podem ser executadas, tais como: extração e integração, transformação, limpeza, seleção e redução de dados (Rezende, 2005). Na extração e integração, os dados disponíveis podem ser encontrados em diferentes fontes: texto, ficheiros Excel, base de dados, entre outros. A dispersão dos dados contribui para a necessidade de obter a unificação dos dados, formando uma única fonte (Rezende, 2005).

No caso em estudo, a informação relativa ao processo de aprendizagem dos alunos foi armazenada no PCMAT e está num único ficheiro Microsoft Excel. Esta informação contém as diversas fases do processo: no início da aprendizagem, os alunos foram submetidos a uma avaliação de diagnóstico para aferir os seus conhecimentos iniciais, também responderam a questionário que permitiu identificar os estilos de aprendizagem iniciais de cada um. Após a aprendizagem, os alunos foram avaliados novamente para analisar os conhecimentos adquiridos e responderam a um questionário que identificou o estilo de aprendizagem final de cada um.

Como a questão da qualidade dos dados é um fator extremamente significativo na fase de modelação, devem ser aplicadas técnicas de limpeza com o intuito de garantir a qualidade dos mesmos. O uso da técnica de limpeza de dados permite tratar registos com valores inválidos, erros nos valores, atributos incompletos e granularidades incorretas, entre outros (Rezende, 2005).

Ao observar os atributos existentes foi constatado que a turma 8º F não apresenta dados relativos às avaliações de diagnóstico. Estes dados em falta necessitam de ser tratados para não perturbarem o modelo de dados final, sendo assim foi realizada uma análise de técnicas dedicadas a este tipo de tratamento de dados. Como apenas existe uma única turma de 8º ano de escolaridade, 8º F, foram omitidos os dados em falta recorrendo ao uso do nó *Select* da ferramenta *Clementine*, excluindo todos os registos que contêm dados incompletos.

A técnica de seleção e redução dos dados pode ser utilizada para reduzir o espaço de pesquisa da solução, tendo como objetivo selecionar os atributos relevantes e remover os que não

apresentam qualquer importância para o modelo final, tendo em conta os objetivos pretendidos (Rezende, 2005).

Neste trabalho foi identificada a presença de alguns atributos, número e nome do aluno, irrelevantes. Estes dados não apresentam nenhum significado para o estudo, apenas identificam o aluno. Os restantes atributos foram classificados como importantes ao estudo.

Outra forma de reduzir a quantidade de dados consiste na técnica de categorização, também designada discretização, substituição de um atributo contínuo (número inteiro ou real) por um discreto, por meio do agrupamento dos seus valores (Rezende, 2005).

O conhecimento desta técnica aliada à exploração gráfica dos dados relativos à evolução dos conteúdos matemáticos permitiu identificar alguns atributos que podem usufruir desta prática:

- Os atributos "LSI" e "LSF" representam o estilo de aprendizagem inicial e final respetivamente. Foi considerado conveniente classificar os valores numéricos, 1, 2 e 3, em valores discretos, sendo mais expressivo para o leitor.
 - 1 - Teórico: o aluno aprende melhor ao ouvir, por exemplo uma aula teórica;
 - 2 - Visual: o aluno aprende melhor ao ver, através de um vídeo ou uma imagem;
 - 3 - Prático: o aluno aprende melhor ao executar as tarefas.
- Os atributos:
 - PDI_total: avaliação de diagnóstico obtida por aluno;
 - A1I: avaliação inicial do conceito A1 (razões equivalentes) por aluno;
 - A2I: avaliação inicial do conceito A2 (noções de antecedente e consequente) por aluno;
 - A3I: avaliação inicial do conceito A3 (leitura de uma razão) por aluno;
 - A4I: avaliação inicial do conceito A4 (aplicações do conceito de razão) por aluno;
 - B1I: avaliação inicial do conceito B1 (meios e extremos) por aluno;

- B2: avaliação inicial do conceito B2 (leitura de proporções) por aluno;
- B3I: avaliação inicial do conceito B3 (propriedade de proporções) por aluno;
- B4I: avaliação inicial do conceito B4 (regra de três simples) por aluno;
- B5I: avaliação inicial do conceito B5 (método de redução à unidade) por aluno;
- PDF_total: avaliação final obtida por aluno;
- A1F: avaliação final do conceito A1 (razões equivalentes) por aluno;
- A2F: avaliação final do conceito A2 (noções de antecedente e consequente) por aluno;
- A3F: avaliação final do conceito A3 (leitura de uma razão) por aluno;
- A4F: avaliação final do conceito A4 (aplicações do conceito de razão) por aluno;
- B1F: avaliação final do conceito B1 (meios e extremos) por aluno;
- B2F: avaliação final do conceito B2 (leitura de proporções) por aluno;
- B3F: avaliação final do conceito B3 (propriedade de proporções) por aluno;
- B4F: avaliação final do conceito B4 (regra de três simples) por aluno;
- B5F: avaliação final do conceito B5 (método de redução à unidade) por aluno.

assumem muitos valores, sendo assim foi conveniente discretizar estes atributos em algumas categorias de classificações para facilitar a compreensão do leitor. Após a análise dos valores dos atributos foram criadas cinco categorias de classificação, uma representa os valores médios (50 a 64), duas identificam valores negativos (0 a 29 e 30 a 49) e as restantes apresentam valores positivos (65 a 89 e 90 a 100):

- Muito mau: < 30 valores percentuais;
- Mau: 30 a 49 valores percentuais;

- Médio: 50 a 64 valores percentuais;
- Bom: 65 a 89 valores percentuais;
- Muito bom: 90 a 100 valores percentuais.

Como mencionado anteriormente, a informação relativa ao 8º ano de escolaridade, 8º F, foi excluída e, conseqüentemente o atributo "ano" que representa o ano de escolaridade do aluno passa a ser encarado como um atributo irrelevante, pois todo o domínio é representado por um único valor, 7º ano de escolaridade.

A exploração gráfica dos dados permite verificar a existência de outros registos com valores nulos nas turmas do 7º ano de escolaridade, de forma a evitar constrangimentos a nível da modelação dos dados, foi considerado essencial a remoção destes registos. De seguida, é apresentada uma tabela que apresentam o número total de alunos por ano e turma que exibiam dados em falta.

Tabela 32 - Número total de alunos com dados em falta por ano e turma

Ano	Turma	Nº total de alunos
7	A	2
	B	2
	F	3
	G	3
8	F	23

A remoção de registos com dados nulos permitir detetar alguns atributos irrelevantes, nomeadamente o "Assiduidade ao PDI" e "Assiduidade ao PDF", que representam o registo de presença ou ausência dos alunos na avaliação de diagnóstico e final respetivamente. Como todos os dados em falta foram eliminados estes dois atributos apresentavam um único valor, "Presente".

Após as diversas operações de tratamento dos dados, foi realizada a exploração gráfica que irá permitir a compreensão do comportamento dos dados.

5.2 Exploração gráfica

A exploração gráfica tem como objetivo analisar os dados relativos à aprendizagem da unidade de "Proporcionalidade Direta" da disciplina de Matemática dos alunos de cinco turmas, quatro de 7º ano e uma de 8º ano de escolaridade. Esta fase permite detetar comportamentos de dados considerados não normais:

- dados corrompidos ou incompatíveis;
- anomalias;
- falta de atributos de interesse;
- diferentes níveis de granularidade.

Alguns gráficos obtidos na exploração gráfica podem contribuir para a definição de critérios, ajudando a melhorar os modelos e outros são insignificantes para o estudo em questão e os seus resultados são esquecidos. É importante mencionar que a exploração gráfica diz respeito apenas aos registos que não exibam valores nulos.

5.2.1 Caracterização da amostra de dados

Ao observar a distribuição dos alunos por sexo, figura 8, foi verificado que cerca de 54% dos alunos eram do sexo feminino e os restantes 46% dos alunos eram do sexo masculino.



Value ▲	Proportion	%	Count
Feminino		53,62	37
Masculino		46,38	32

Figura 8 - Distribuição de alunos por sexo

Os alunos foram distribuídos por tipo de ensino, figura 9, foi registado que cerca de 9% dos alunos com necessidades especiais de ensino e 91% dos alunos apresentaram ensino normal.



Value ▲	Proportion	%	Count
Especial		8,7	6
Normal		91,3	63

Figura 9 - Distribuição de alunos por tipo de ensino

Todos os alunos foram divididos em dois grupos, experimental e controle, figura 10, cerca de 43% dos alunos pertenciam ao grupo de controle e os restantes, 57% dos alunos, correspondiam ao grupo experimental.

Value ▲	Proportion	%	Count
Controle		43,48	30
Experimental		56,52	39

Figura 10 - Distribuição dos alunos por grupo

5.2.2 Classificações obtidas na avaliação de diagnóstico por sexo

Na fase inicial do processo de aprendizagem, todos os alunos realizaram um teste diagnóstico, aferindo os seus conhecimentos iniciais relativos ao tema "Proporcionalidade Direta". A figura 11 representa os resultados obtidos pelos alunos no teste diagnóstico.

Value ▲	Proportion	%	Count
Mau		26,09	18
Médio		11,59	8
Muito Mau		62,32	43

Sexo

Feminino Masculino

Figura 11 - Classificação do teste diagnóstico dos alunos por sexo

As classificações obtidas no teste diagnóstico foram negativas, apenas cerca de 12% dos alunos é que conseguiram atingir resultados positivos. Estes resultados permitiram constatar que a maioria dos alunos não possuía conhecimentos relativos a esta unidade.

Ainda na fase inicial deste processo, os alunos realizaram um questionário VARK com o intuito de aferir o estilo de aprendizagem inicial de cada aluno. Os resultados deste questionário são representados na figura 12 e identificaram a preferência de aprendizagem dos alunos, teórico, visual e prático.

Value ▲	Proportion	%	Count
Prático		60,87	42
Teórico		11,59	8
Visual		27,54	19

Sexo

Feminino Masculino

Figura 12 - Estilo de aprendizagem inicial dos alunos por sexo

A maioria dos alunos exibiu um estilo de aprendizagem inicial prático, cerca de 61% dos alunos, seguido do estilo de aprendizagem visual representado por 28% dos alunos e, por fim, o estilo de aprendizagem teórico registrando cerca de 12% dos alunos existentes.

A temática "Proporcionalidade Direta" abordava diversos conceitos relativos a dois grupos, razão e proporções. Na avaliação de diagnóstico, estes conceitos foram avaliados e o conceito A1 (razões equivalentes) obteve resultados muito negativos, sendo representado pela figura 13. Cerca de 79% dos alunos obteve classificações de "Muito Mau" e apenas 22% dos alunos atingiram classificações de "Muito Bom".

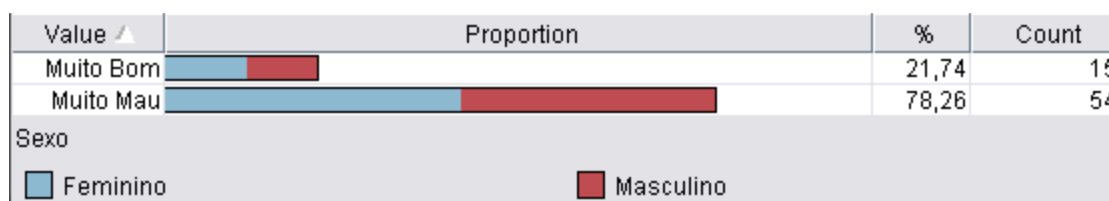


Figura 13 - Classificação do conceito A1 no teste diagnóstico por sexo

Os resultados obtidos pelo conceito A2 (noção de antecedente e consequente) são ilustrados na figura 14. Estes resultados foram distribuídos por diversos escalões de classificações, cerca de 48% dos alunos alcançaram resultados positivos e os restantes obtiveram classificações negativas. Este conceito exibiu melhores resultados em comparação com o conceito anterior, A1.

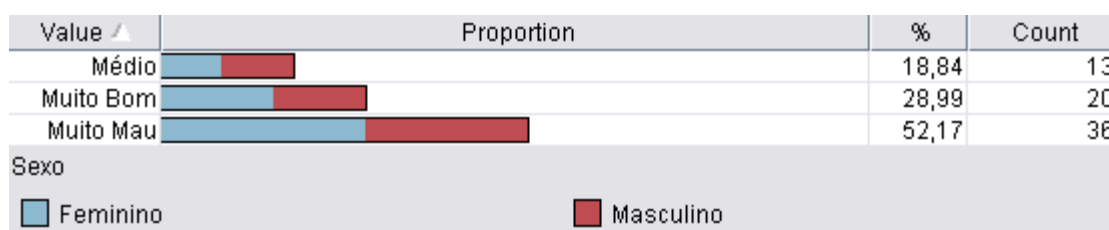


Figura 14 - Classificação do conceito A2 no teste diagnóstico por sexo

A distribuição da classificação do conceito A3 (leitura de uma razão) é representada na figura 15. Os resultados obtidos foram muito negativos, 94% dos alunos atingiram classificações de "Muito Mau" e os restantes alcançaram classificações de "Muito Bom". O conceito A3 obteve as piores classificações comparativamente com os conceitos analisados anteriormente.

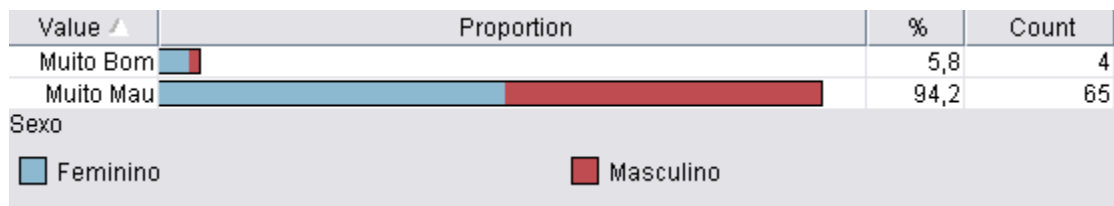


Figura 15 - Classificação do conceito A3 no teste diagnóstico por sexo

A figura 16 exibe a avaliação inicial do conceito A4 (aplicações do conceito de razão) e demonstrou que cerca de 68% dos alunos obtiveram classificações negativas de "Muito Mau", resultados inferiores a 30 valores percentuais, e os restantes alunos foram distribuídos por diversos níveis de classificação "Médio", "Bom" e "Muito Bom".

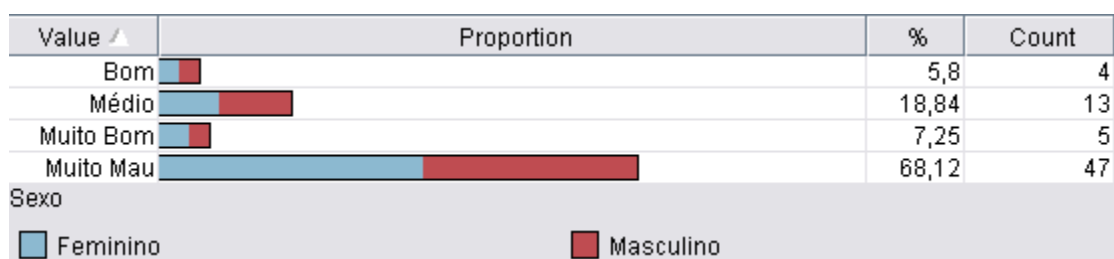


Figura 16 - Classificação do conceito A4 no teste diagnóstico por sexo

Após a análise da avaliação inicial dos conceitos do grupo A foi verificado que o conceito A3 (leitura de uma razão) exibiu maior taxa de respostas negativas e o conceito A2 (noções de antecedente e consequente) apresentou maior número de respostas certas.

Os conceitos do grupo B também foram avaliados na fase inicial. A classificação final do conceito B1 (meios e extremos) é exibida na figura 17 e foi registado que cerca de 51% de negativas, classificações de "Muito Mau" e 49% foram positivas. Os resultados positivos foram distribuídos por diferentes escalões de classificações, 12% dos alunos exibiram classificações de "Médio", 3% dos alunos apresentaram classificações de "Bom" e 35% dos alunos obtiveram classificações de "Muito Bom".

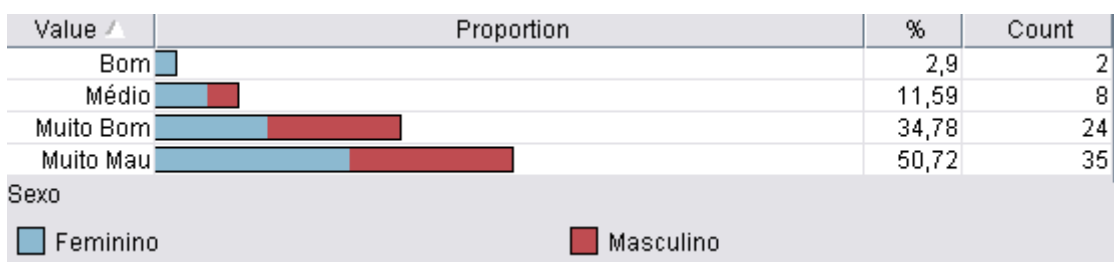


Figura 17 - Classificação do conceito B1 no teste diagnóstico por sexo

A avaliação inicial do conceito B2 (leitura de proporções) é ilustrada na figura 18 e foram resultados bastante negativos, nomeadamente 97% dos alunos exibiram classificações inferiores a 30 valores percentuais. Apenas cerca de 3% dos alunos é que alcançaram classificações de "Muito Bom", resultados compreendidos entre 90 e 100 valores percentuais.

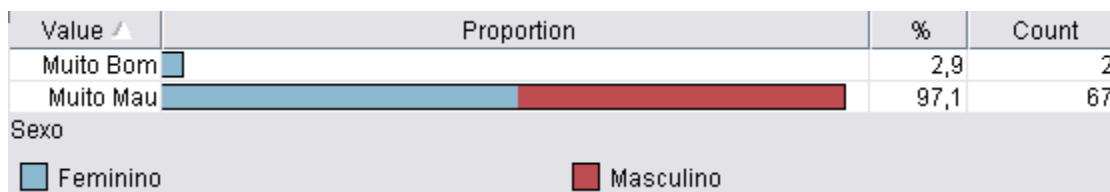


Figura 18 - Classificação do conceito B2 no teste diagnóstico por sexo

A figura 19 representa os dados do conceito B3 (propriedade de proporções) e foi registado que cerca de 20% dos alunos exibiram negativas enquanto que os restantes alunos, 80% dos alunos, apresentaram resultados positivos. Este conceito foi destacado dos conceitos avaliados anteriormente porque exibiu menor quantidade de alunos a apresentar resultados negativos.

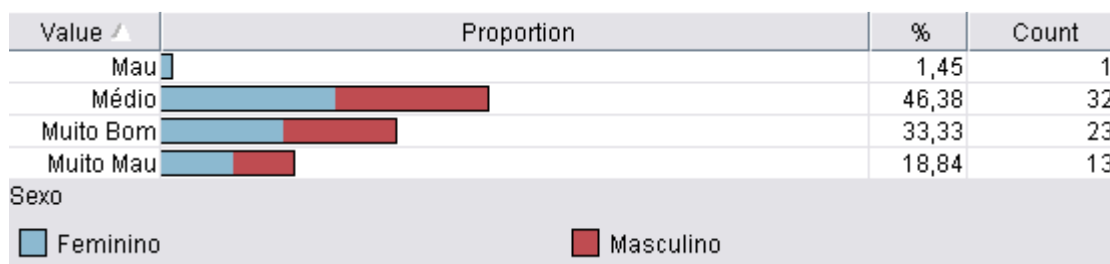


Figura 19 - Classificação do conceito B3 no teste diagnóstico por sexo

O conceito B4 (regra de três simples) é ilustrado através da figura 20 e obteve resultados negativos, concretamente 75% dos alunos alcançaram classificações inferiores a 30 valores - "Muito Mau" e os restantes alunos, cerca de 25%, atingiram valores compreendidos entre 90 a 100, classificação de "Muito Bom".

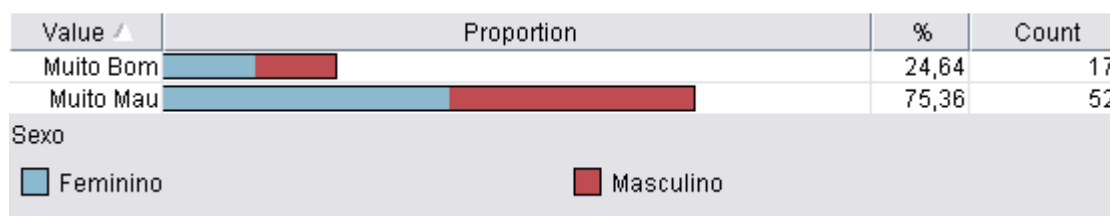


Figura 20 - Classificação do conceito B4 no teste diagnóstico por sexo

Os resultados do conceito B5 (método de redução à unidade) são exibidos na figura 21 e evidenciaram classificações bastante negativas, 93% dos alunos obtiveram resultados

negativos, valores inferiores a 30 e apenas 7% dos alunos conseguiu atingir classificações de "Muito Bom", resultados entre os 90 a 100 valores percentuais.

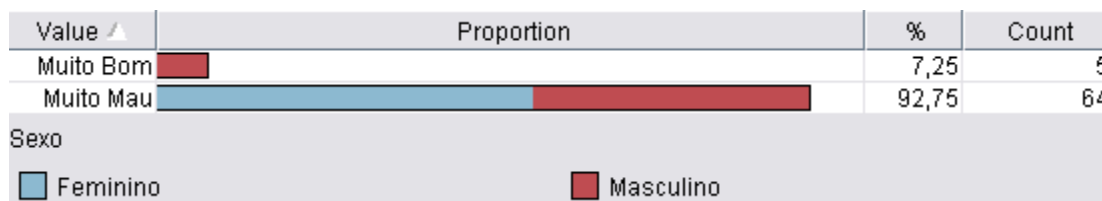


Figura 21 - Classificação do conceito B5 no teste diagnóstico por sexo

Analisando a avaliação inicial dos conceitos do grupo B, proporções, foi identificado que o conceito B3 é o que exibiu melhores resultados e o conceito B2 apresentou piores classificações.

5.2.3 Classificações obtidas na avaliação final por sexo

Após o ensino da unidade de "Proporcionalidade Direta", todos os alunos realizaram um teste final, onde foram avaliados os conhecimentos adquiridos pelos alunos. Os resultados deste teste são representados na figura 22 e demonstraram um aumento bastante significativo perante os resultados do teste diagnóstico. A maioria dos alunos conseguiram atingir classificações positivas, contudo cerca de 46% dos alunos obtiveram resultados negativos.

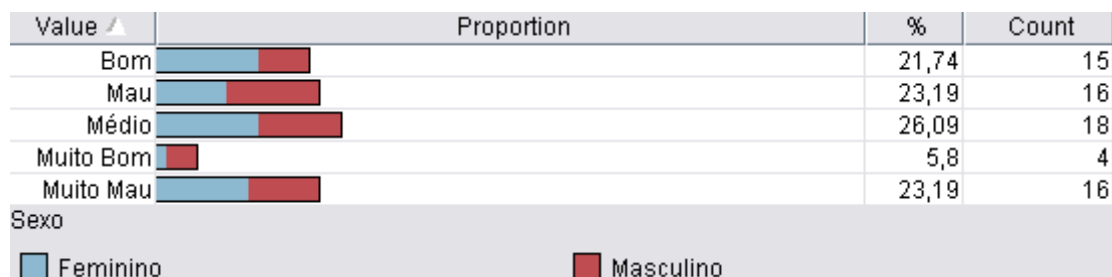


Figura 22 - Classificação do teste final por sexo

Na fase final do processo de aprendizagem, os alunos foram avaliados em relação ao estilo de aprendizagem final, figura 23, a maioria dos alunos exibiam o estilo de aprendizagem prático, nomeadamente 58% dos alunos. Comparando estes resultados com os obtidos na fase inicial, foi constatado um acréscimo de número de alunos com estilo de aprendizagem visual, uma diminuição da quantidade de alunos com estilo de aprendizagem prático e o número de alunos com estilo de aprendizagem teórico foi mantido estável.

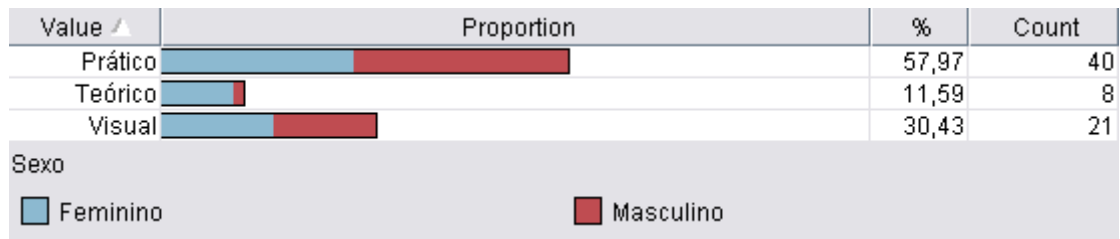


Figura 23 - Estilo de aprendizagem final dos alunos por sexo

Na avaliação final, o conceito A1 (razões equivalentes) evidenciou resultados maioritariamente negativos, sendo ilustrado na figura 24. Cerca de 52% dos alunos obtiveram classificações negativas e os restantes, 48% dos alunos alcançaram classificações positivas, sendo distribuídas por dois escalões, "Médio" e "Bom".

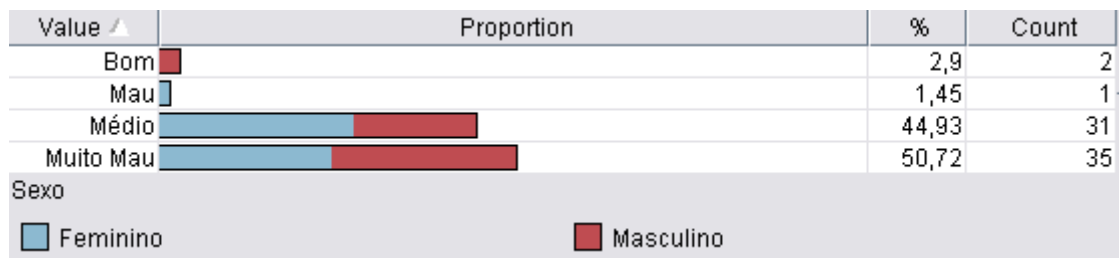


Figura 24 - Classificação do conceito A1 no teste final por sexo

A avaliação final do conceito A2 (noções de antecedente e consequente) é representada na figura 25 e demonstrou que a maioria dos alunos atingiu resultados positivos, apenas cerca de 4% dos alunos obtiveram classificações negativas, "Muito Mau".

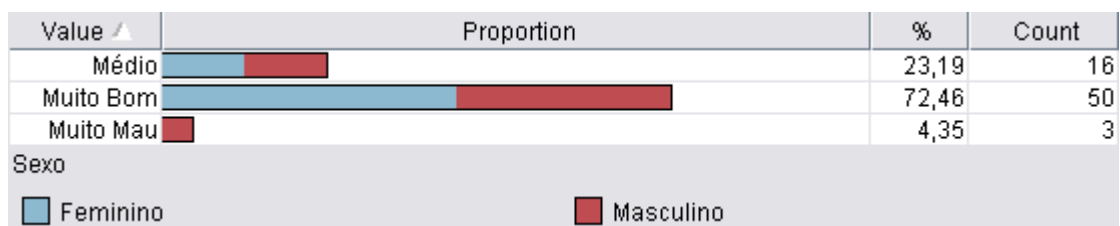


Figura 25 - Classificação do conceito A2 no teste final por sexo

A figura 26 exibe os resultados do conceito A3 (leitura de uma razão) e evidenciou que a maioria dos alunos obteve negativa, 58% dos alunos. Os restantes alunos foram distribuídos por distintos níveis de classificações positivas, "Médio", "Bom" e "Muito Bom".

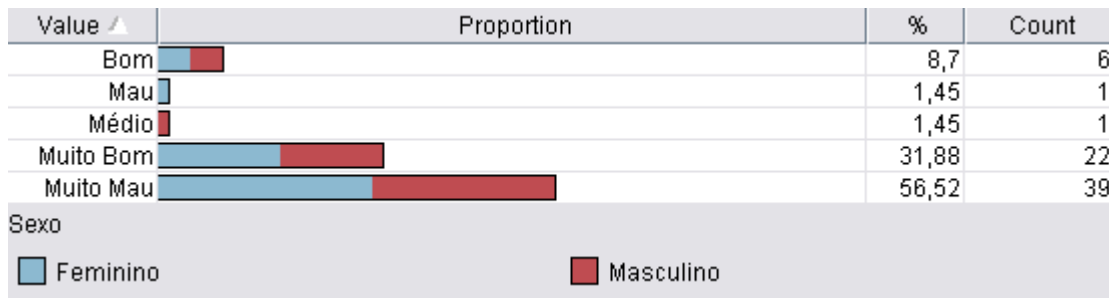


Figura 26 - Classificação do conceito A3 no teste final por sexo

Na avaliação final, a distribuição dos alunos por classificação obtida referente ao conceito A4 (aplicações do conceito de razão), figura 27, demonstrou que cerca de 61% dos alunos atingiram classificações negativas e os restantes alunos foram distribuídos por classificações de "Bom" e "Muito Bom".

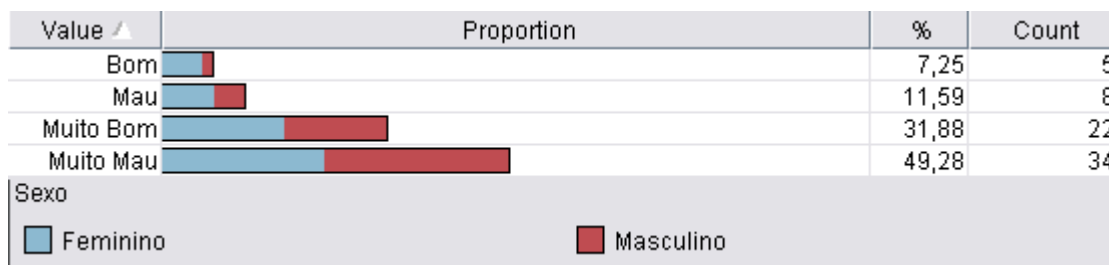


Figura 27 - Classificação do conceito A4 no teste final por sexo

Após uma breve análise aos resultados obtidos na avaliação final dos conceitos do grupo A - Razão, foi identificado que o conceito A2 obteve os melhores resultados, maior taxa de classificações positivas, e o conceito A4 exibiu maior taxa de classificações negativas.

Semelhante aos conceitos do grupo A, os conceitos do grupo B - Proporções também foram avaliados no teste final. Em relação ao conceito B1 (meios e extremos), figura 28, a maioria dos alunos alcançou resultados positivos, nomeadamente 77% dos alunos obtiveram classificações de "Médio" e "Muito Bom". Os restantes alunos, especificamente 23%, atingiram classificações negativas, inferiores a 30 valores percentuais.

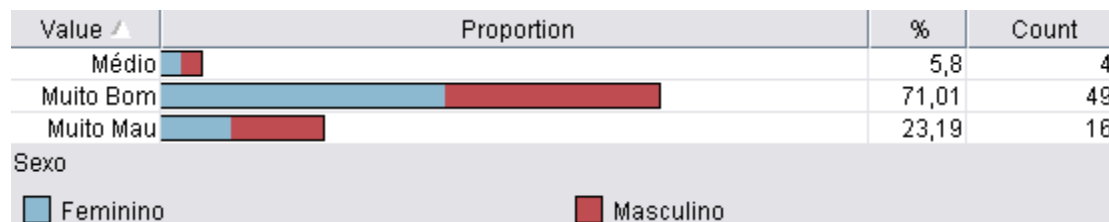


Figura 28 - Classificação do conceito B1 no teste final por sexo

A figura 29 representa a avaliação final do conceito B2 (leitura de proporções), foi verificado que a maioria dos alunos obtiveram resultados negativos. Cerca de 54% dos alunos obtiveram classificações de "Muito Mau" e "Mau", e os restantes alunos foram distribuídos por diferentes graus de classificação, "Médio", "Bom" e "Muito Bom".

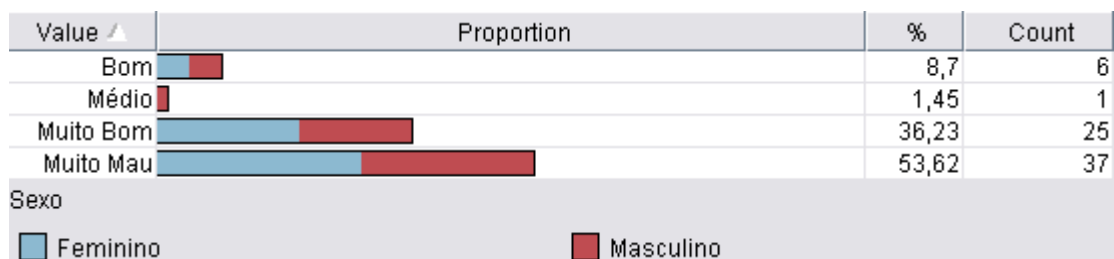


Figura 29 - Classificação do conceito B2 no teste final por sexo

Os resultados do conceito B3 (propriedade das proporções), figura 30, evidenciaram que a maioritariamente os alunos obtiveram classificações positivas, sendo distribuídos por classificações de "Médio" e "Muito Bom" e apenas cerca de 7% dos alunos atingiram classificações negativas.

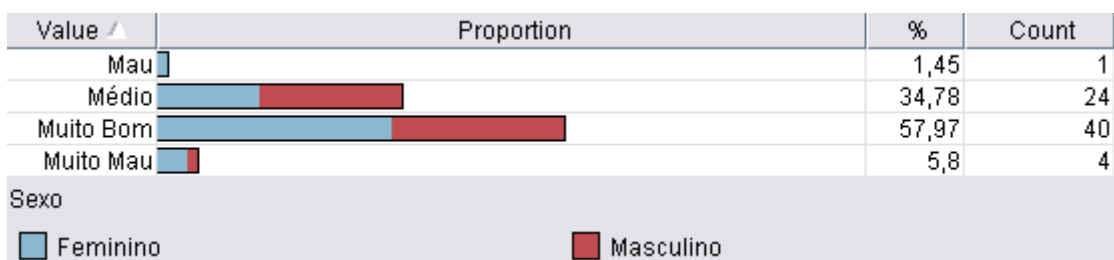


Figura 30 - Classificação do conceito B3 no teste final por sexo

A figura 31 exibe a avaliação final do conceito B4 (regra de três simples) e demonstrou que a maioria dos alunos obtiveram resultados bastante satisfatórios, cerca de 58% dos alunos alcançaram classificações de "Muito Bom" e os restantes alunos atingiram classificações negativas de "Muito Mau". Este conceito exibiu extremos de classificações, os alunos apresentavam valores muito elevados ou muito baixos a nível dos resultados obtidos.

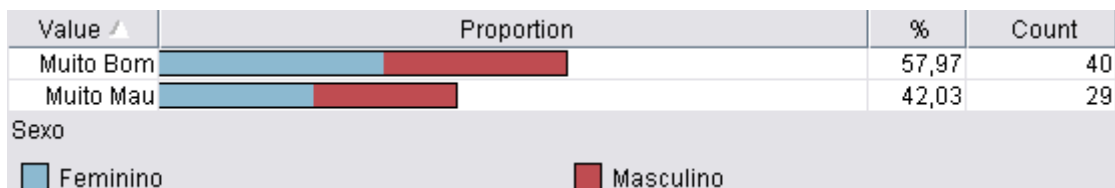


Figura 31 - Classificação do conceito B4 no teste final por sexo

O conceito B5 (método de redução à unidade), figura 32, exibiu resultados muito negativos, nomeadamente 93% dos alunos obtiveram classificações negativas e apenas 7% dos alunos atingiram classificações positivas. Este conceito apresentou um comportamento semelhante ao conceito analisado anteriormente a nível da distribuição das classificações dos alunos, sendo que as únicas classificações obtidas foram "Muito Mau" e "Muito Bom".

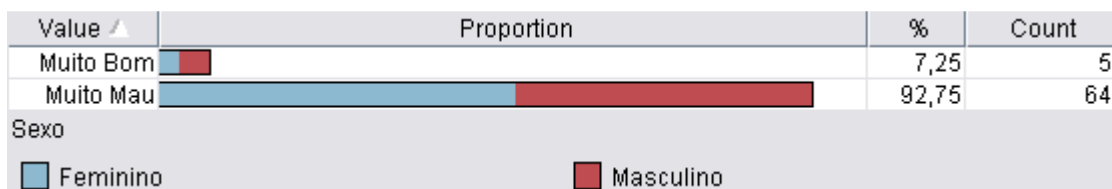


Figura 32 - Classificação do conceito B5 no teste final por sexo

Os conceitos do grupo B foram analisados e foi verificado que o conceito B3 apresentou maior número de classificações positivas e o conceito B5 foi identificado como o que exibiu maior taxa de resultados negativos.

5.2.4 Comparação entre os resultados obtidos na avaliação inicial e final

Após a análise da evolução dos resultados obtidos da avaliação inicial e final dos conceitos, tabela 33, foi possível identificar os conceitos com melhores e piores resultados a nível de aprendizagem.

Tabela 33 - Classificação total por conceito obtida pelos alunos

Conceito	Avaliação		Progresso
	Inicial	Final	
A1	1500	3430	1930
A2	2670	1784	-886
A3	400	5710	5310
A4	1575	2760	1185
B1	2975	2796	-179
B2	200	3030	2830
B3	3940	5190	1250
B4	1725	4025	2300
B5	500	525	25

Na fase inicial, foi identificado o B2 (leitura de proporções) como o conceito com piores resultados e o B3 (propriedade de proporções) com os melhores resultados. Na avaliação final

foi averiguado que o conceito B5 (método de redução à unidade) apresentava as piores classificações e o A3 (leitura de uma razão) exibiu as melhores classificações.

Comparando os dados obtidos nas duas fases distintas do processo de aprendizagem foi constatado que o conceito mais simples de ser assimilado pelos alunos foi o A3 (leitura de uma razão). Os conceitos A2 (noções de antecedente e consequente), B1 (meios e extremos) e B5 (método de redução à unidade) apresentaram as piores classificações ao longo deste processo de aprendizagem, sendo classificados como os conceitos mais difíceis de adquirir pelos alunos. Apesar das dificuldades evidenciadas na obtenção dos conceitos mencionados anteriormente, o conceito B5 apresentou um ligeiro aumento nos resultados obtidos na fase final de avaliação. Contrariamente, os conceitos A2 e B1 registaram uma diminuição nos resultados, regredindo a nível dos conhecimentos e demonstrando que a aprendizagem deste conceito não era adequada aos alunos.

Para continuar com a análise dos dados fornecidos foi considerado interessante analisar os resultados obtidos no teste final em relação ao estilo de aprendizagem e sexo dos alunos, como se observa na figura 33.

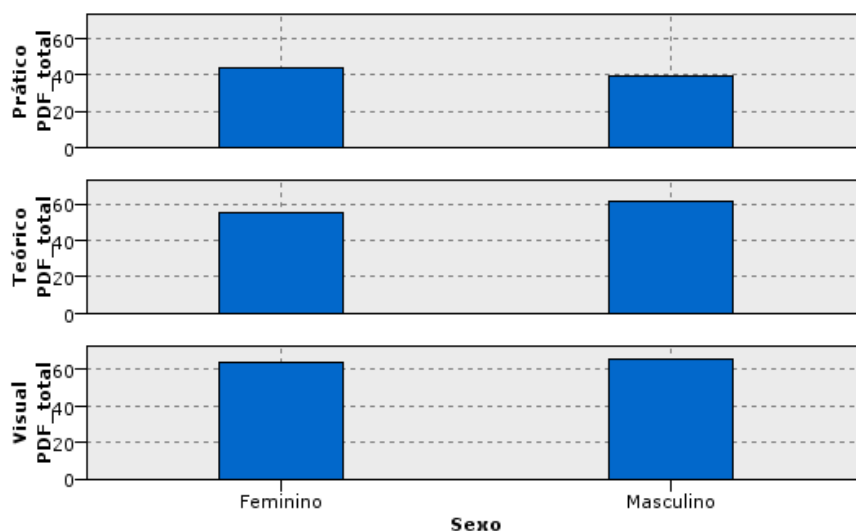


Figura 33 - Classificação no teste final por sexo e estilo de aprendizagem final

A classificação no teste final por sexo e estilo de aprendizagem final permitiu verificar que o estilo de aprendizagem e o sexo dos alunos influenciaram os resultados obtidos no teste final. Os alunos com estilo de aprendizagem prático obtiveram classificações a rondar os 40 valores percentuais, os alunos com estilo de aprendizagem teórico apresentaram classificações acima dos 50 valores percentuais e os alunos com estilo de aprendizagem visual exibiram

classificações acima dos 60 valores percentuais. Desta forma, o estilo de aprendizagem visual foi destacado dos restantes estilos de aprendizagem porque apresentou os melhores resultados. Em relação ao sexo foi verificado que os alunos masculinos obtiveram melhores classificações nos estilos de aprendizagem teórico e visual em comparação com os alunos do sexo feminino. Contudo, no estilo de aprendizagem prático, os resultados obtidos pelo sexo masculino e feminino foram muito semelhantes, apesar disso foram verificados melhores resultados nos alunos femininos.

Para compreender melhor as influências e/ou relações entre os conceitos, foram analisadas as relações entre os conceitos numa fase inicial e final do processo de aprendizagem. Desta forma, recorreu-se ao nó *web* na ferramenta *Clementine* para identificar potenciais relações entre os diferentes conceitos.

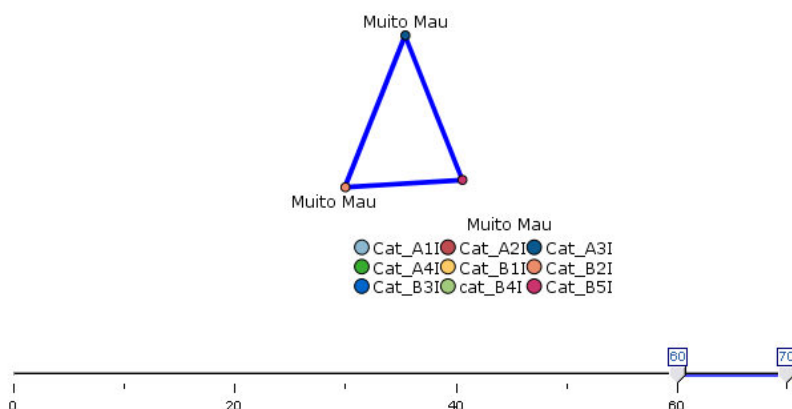


Figura 34 - Relações entre os conceitos no teste diagnóstico

A figura 34 representa as relações entre os conceitos na avaliação de diagnóstico e foram identificadas relações fortes entre as classificações de "Muito Mau" nos conceitos A3, B2 e B5.

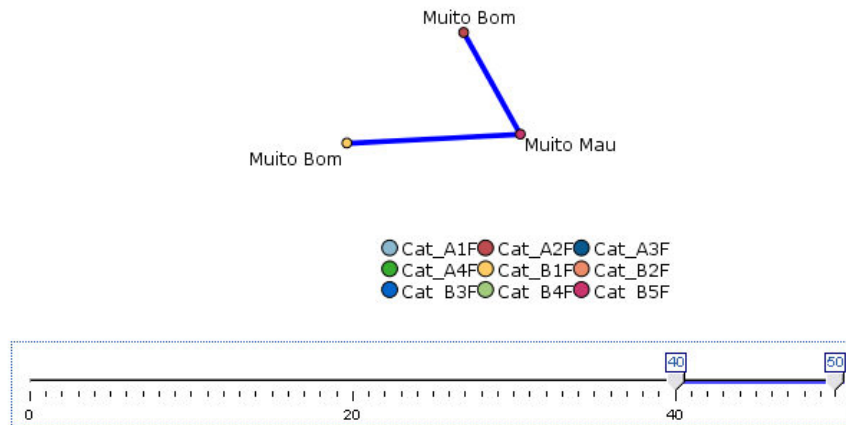


Figura 35 - Relações entre os conceitos no teste final

A figura 35 representa as relações existentes entre os conceitos no teste final e foram identificadas algumas ligações fortes entre algumas classificações de conceitos, nomeadamente a classificação de "Muito Bom" do conceito A2 e as classificações de "Muito Bom" do conceito B1 e "Muito Mau" do conceito B5.

5.2.5 Comparação das classificações nas avaliações por estilo de aprendizagem

Neste trabalho foi considerado importante estudar o progresso de aprendizagem dos conceitos, comparando os resultados obtidos na fase inicial e final da aprendizagem de cada conceito de acordo com o estilo de aprendizagem dos alunos.

Na fase inicial de avaliação do conceito A1 (razões equivalentes), figura 36, os alunos com estilo de aprendizagem prático atingiram classificações aproximadas de 22 valores percentuais, os alunos com estilo de aprendizagem teórico exibiram classificações próximas de 38 valores percentuais e os alunos com estilo de aprendizagem visual obtiveram classificações inferiores a 20 valores percentuais. Desta forma, foi identificado que os alunos com estilo de aprendizagem teórico apresentaram melhores resultados e os alunos com estilo de aprendizagem visual registaram os piores resultados.

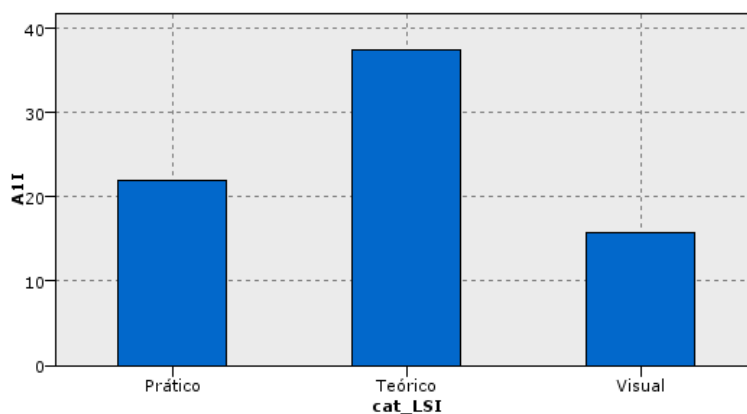


Figura 36 - Classificação inicial do A1 por estilo de aprendizagem

A figura 37 representa a classificação final do conceito A1 obtida pelos alunos e por estilo de aprendizagem. Os resultados obtidos nesta distribuição indicaram que os alunos com estilo de aprendizagem visual obtiveram melhores resultados na avaliação do conceito A1, aproximadamente 43 valores percentuais. Os alunos com estilo de aprendizagem teórico apresentaram classificações inferiores a 30 valores percentuais e, por fim, os alunos que possuíam estilo de aprendizagem prático atingiram classificações inferiores a 20 valores percentuais.

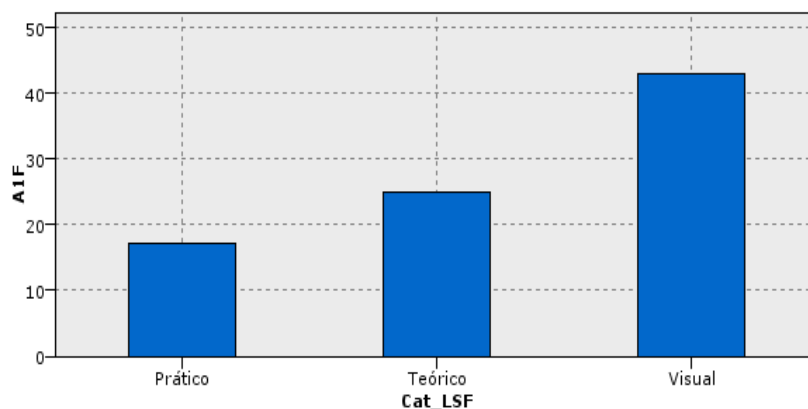


Figura 37 - Classificação final do A1 por estilo de aprendizagem

Os alunos com estilo de aprendizagem prático e teórico regrediram os conhecimentos relativos ao conceito A1. Os alunos com estilo de aprendizagem visual aumentaram os seus conhecimentos, representando os melhores resultados neste conceito. Contudo não conseguiram alcançar uma classificação positiva, situando-se acima dos 40 valores percentuais.

A avaliação inicial do conceito A2 (noções de antecedente e consequente), figura 38, registou que os alunos com estilo de aprendizagem prático obtiveram resultados próximos de 40

valores percentuais, os alunos com estilo de aprendizagem teórico apresentaram resultados superiores a 60 valores percentuais e os alunos com estilo de aprendizagem visual exibiram classificações de 40 valores percentuais.

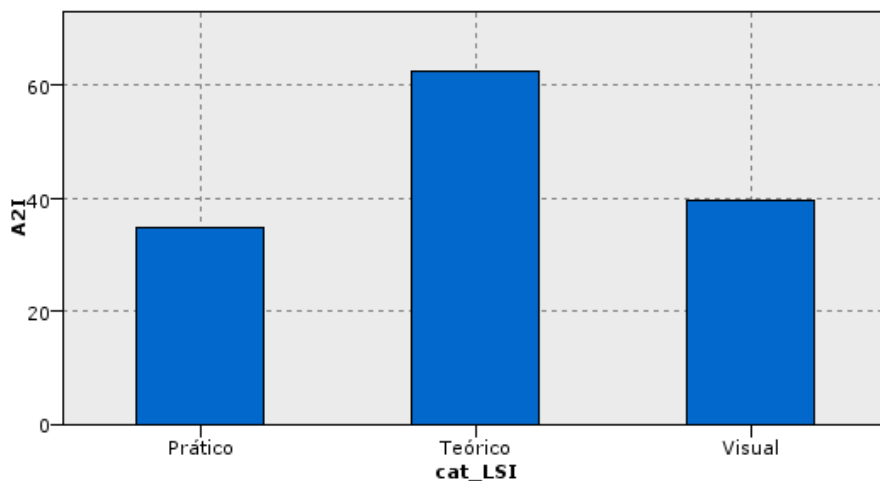


Figura 38 - Classificação inicial do A2 por estilo de aprendizagem

A figura 39 ilustra a distribuição das classificações obtidas no teste final por estilo de aprendizagem. Os alunos com estilo de aprendizagem teórico atingiram os 100 valores percentuais, os alunos com estilo de aprendizagem visual registaram resultados próximos de 90 valores percentuais e os alunos com estilo de aprendizagem prático obtiveram classificações próximas de 80 valores percentuais.

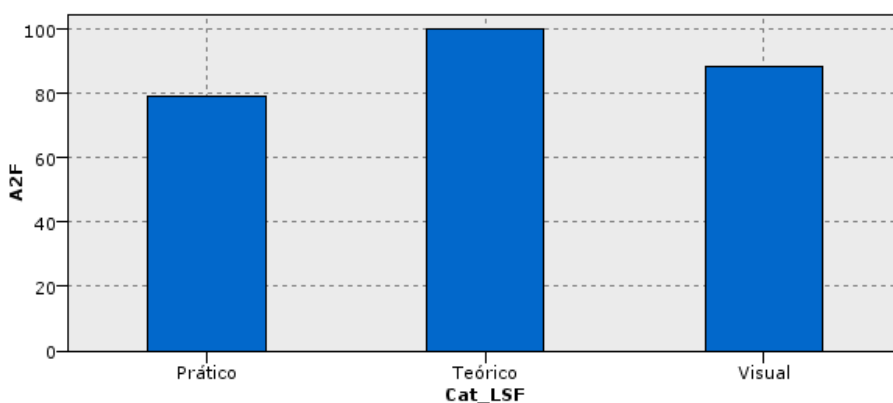


Figura 39 - Classificação final do A2 por estilo de aprendizagem

Comparando os resultados obtidos na avaliação inicial e final do conceito A2, foi concluído que a aprendizagem deste conceito foi realizada com sucesso, todos os alunos atingiram bons resultados.

Os resultados do teste diagnóstico relativamente ao conceito A3 (leitura de uma razão), figura 40, indicaram que a maioria dos alunos não tinha conhecimentos deste conceito, nomeadamente os alunos com estilo de aprendizagem teórico obtiveram 0 valores percentuais. Os alunos com estilo de aprendizagem prático apresentaram resultados inferiores a 5 valores percentuais e os alunos com estilo de aprendizagem visual atingiram classificações próximas dos 15 valores percentuais.

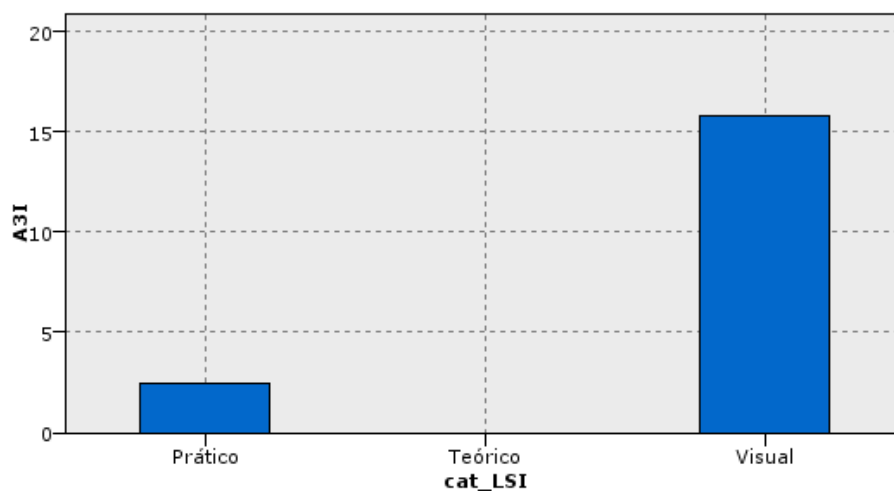


Figura 40 - Classificação inicial do A3 por estilo de aprendizagem

Após a aprendizagem do conceito A3, figura 41, foi verificado que a maioria dos alunos, nomeadamente os alunos com estilo de aprendizagem teórico e visual atingiram resultados satisfatórios, 50 e 55 valores percentuais respetivamente. Contudo os alunos com estilo de aprendizagem prático registaram mais dificuldades em atingir resultados razoáveis e obtiveram classificações próximas de 30 valores percentuais.

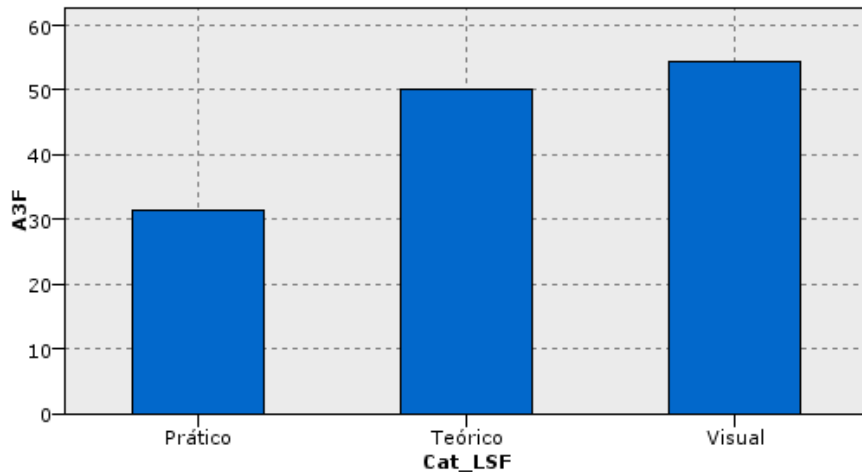


Figura 41 - Classificação final do A3 por estilo de aprendizagem

Comparando os resultados obtidos entre a avaliação inicial e final do conceito A3 foi constatado que o processo de aprendizagem deste conceito ocorreu com algum sucesso, sendo que a maioria dos alunos atingiu resultados razoáveis. Não esquecendo de mencionar que os alunos com estilo de aprendizagem prático registaram algumas dificuldades e não conseguiram alcançar resultados positivos neste conceito.

A avaliação inicial do conceito A4 (aplicações do conceito de razão), figura 42, os alunos de estilo de prático obtiveram classificações inferiores a 20 valores percentuais, os alunos com estilo teórico registaram resultados próximos de 25 valores percentuais e os alunos com estilo de aprendizagem visual atingiram resultados aproximadamente de 33 valores percentuais.

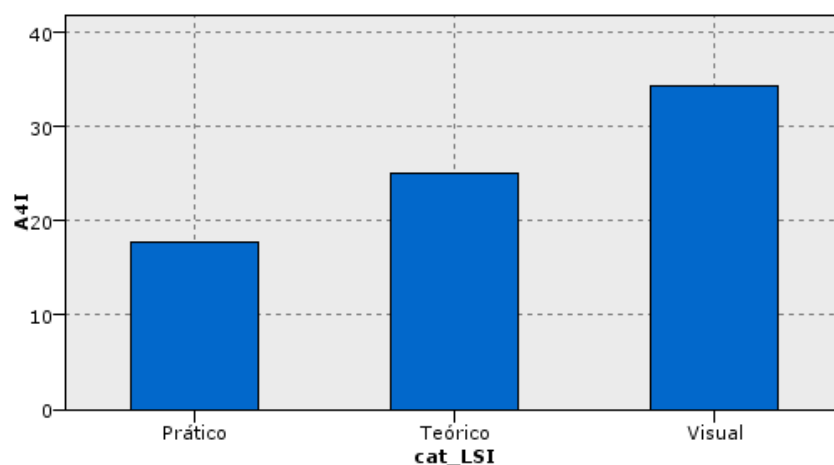


Figura 42 - Classificação inicial do A4 por estilo de aprendizagem

Na fase final de avaliação relativa ao conceito A4, figura 43, os alunos de estilo de aprendizagem prático obtiveram resultados próximos de 30 valores percentuais, os alunos com estilo de aprendizagem teórico exibiram classificações próximas de 45 valores percentuais e os alunos com estilo de aprendizagem visual apresentaram resultados próximos de 60 valores percentuais.

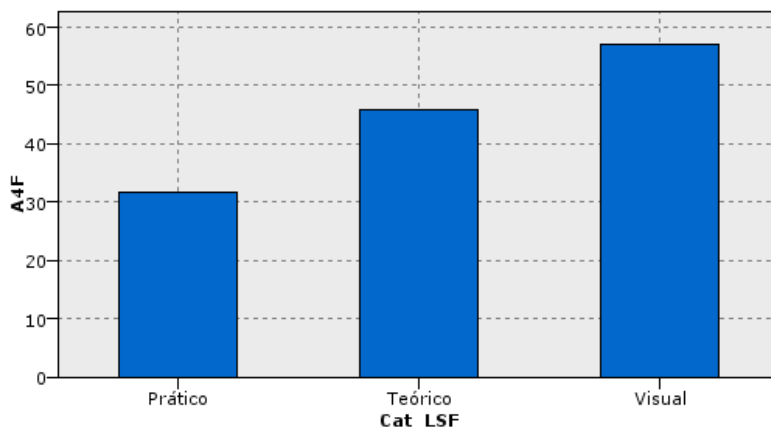


Figura 43 - Classificação final do A4 por estilo de aprendizagem

As classificações obtidas entre a fase inicial e final do processo de aprendizagem relativas ao conceito A4 demonstraram que a aprendizagem deste conceito não foi realizada com sucesso, tendo em conta que os alunos com estilo de aprendizagem prático e teórico registaram algumas dificuldades e não conseguiram alcançar classificações positivas.

Em relação ao conceito B1 (meios e extremos), figura 44, os resultados obtidos demonstraram que os alunos com estilo de aprendizagem teórico possuíam alguns conhecimentos deste conceito, registando classificações superiores a 60 valores percentuais. Os alunos com estilo de aprendizagem prático obtiveram resultados próximos de 45 valores percentuais e os alunos com estilo de aprendizagem visual alcançaram classificações de aproximadamente 35 valores percentuais.

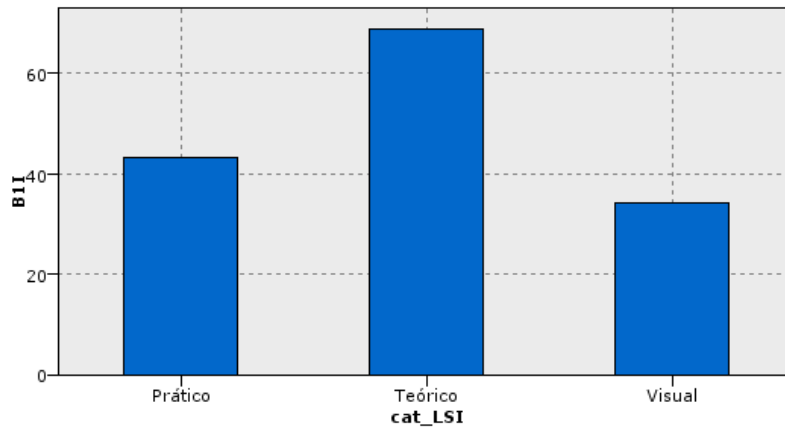


Figura 44 - Classificação inicial do B1 por estilo de aprendizagem

Após a aprendizagem do conceito B1, figura 45, os alunos com estilo prático e visual registraram classificações próximas de 80 valores percentuais e os restantes alunos, com estilo de aprendizagem teórico, exibiram resultados próximos de 70 valores percentuais.

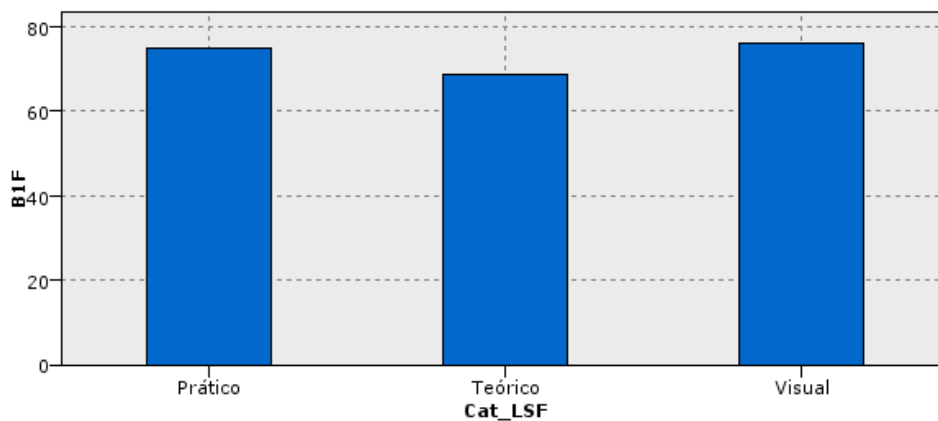


Figura 45 - Classificação final do B1 por estilo de aprendizagem

De forma geral, foi possível concluir que a aprendizagem do conceito B1 foi realizada com sucesso, tendo em conta que, independentemente da preferência de aprendizagem, os alunos obtiveram classificações satisfatórias.

A maioria dos alunos não possuía qualquer conhecimento relacionado com o conceito B2 (leitura de uma proporção), figura 46, especificamente os alunos com estilo de aprendizagem prático e teórico, registaram 0 valores percentuais na fase inicial do processo de aprendizagem. Os alunos com estilo de aprendizagem visual atingiram classificações próximas de 10 valores percentuais.

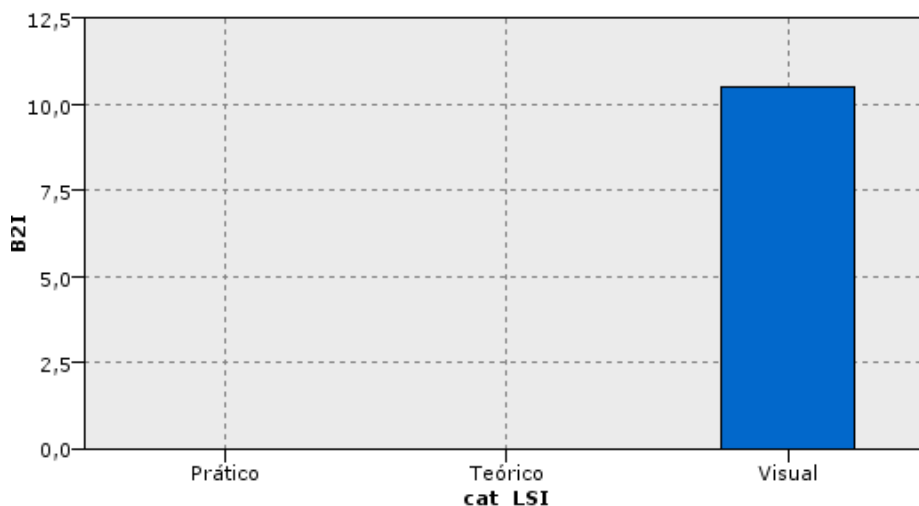


Figura 46 - Classificação inicial do B2 por estilo de aprendizagem

As classificações finais do conceito B2, figura 47, permitiram verificar que os alunos com estilo de aprendizagem prático alcançaram valores percentuais inferiores a 40, os alunos com estilo de aprendizagem teórico obtiveram resultados iguais a 50 valores percentuais e os alunos com estilo de aprendizagem visual apresentaram classificações próximas de 60 valores percentuais.

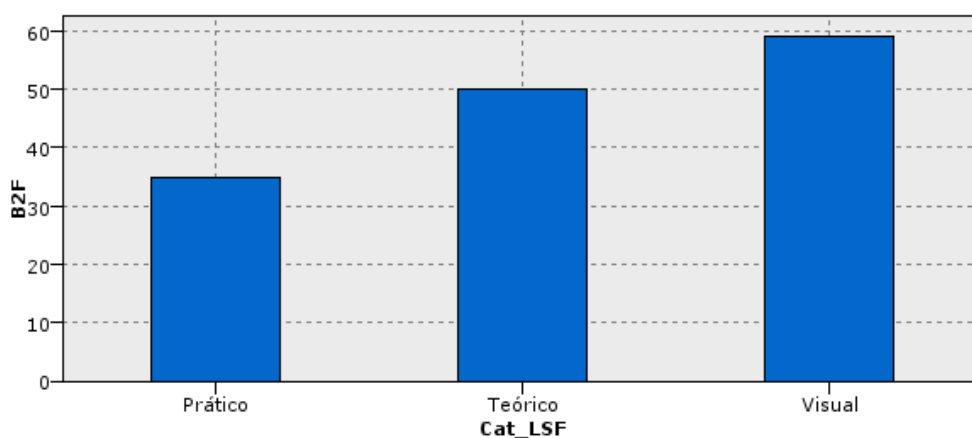


Figura 47 - Classificação final do B2 obtida pelos alunos e por estilo de aprendizagem

Comparando os resultados obtidos nas duas fases de avaliação do conceito B2, foi verificado que a maioria dos alunos, especificamente os alunos com estilo de aprendizagem teórico e visual conseguiram alcançar resultados satisfatórios, indicando o sucesso da aprendizagem deste conceito. Contudo os alunos com estilo de aprendizagem prático demonstraram algumas dificuldades e não conseguiram atingir classificações positivas, registrando resultados próximos de 35 valores percentuais.

As classificações do teste diagnóstico relativas ao conceito B3 (propriedade de proporções), figura 48, demonstraram que a maioria dos alunos possuía conhecimentos deste conceito, registrando classificações superiores a 50 valores percentuais. Os alunos com estilo de aprendizagem teórico obtiveram os melhores resultados nesta avaliação.

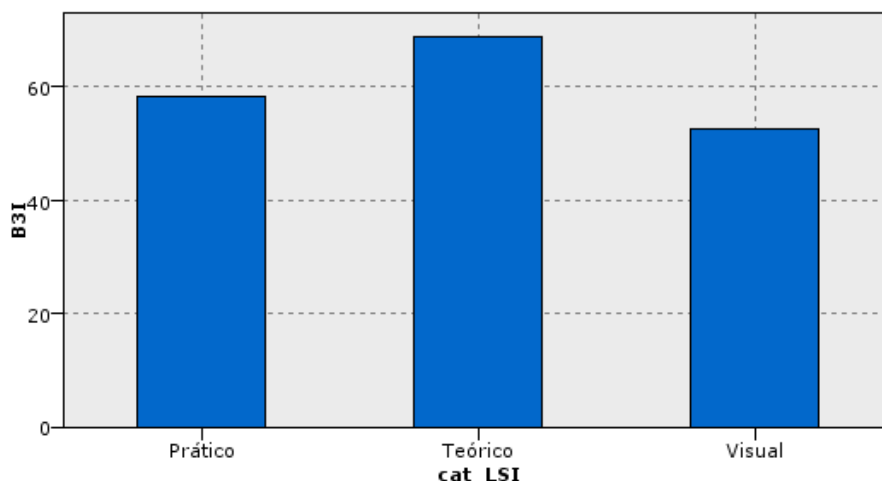


Figura 48 - Classificação inicial do B3 por estilo de aprendizagem

Em relação à avaliação final do conceito B3, figura 49, foi verificado que os alunos com o estilo de aprendizagem prático atingiram resultados próximos de 70 valores percentuais, os alunos com estilo de aprendizagem teórico exibiram resultados superiores de 70 valores percentuais e os alunos com estilo de aprendizagem visual alcançaram classificações a rondar os 90 valores percentuais.

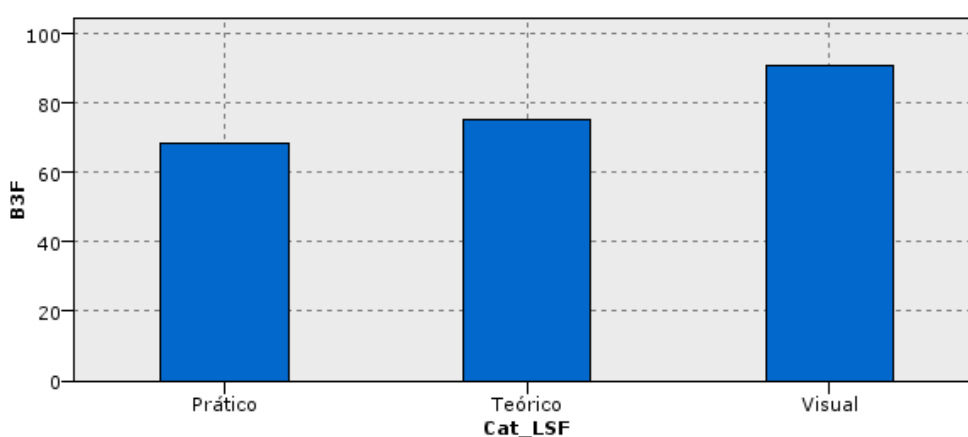


Figura 49 - Classificação final do B3 por estilo de aprendizagem

Os resultados obtidos em relação ao conceito B3 demonstraram que inicialmente os alunos apresentavam alguns conhecimentos referentes a este conceito e após a sua aprendizagem o

domínio deste conceito foi aumentando, realçando que os alunos com estilo de aprendizagem visual atingiram os melhores resultados.

A fase inicial de avaliação do conceito B4 (regra de três simples), figura 50, indicou que os conhecimentos dos alunos com estilo de aprendizagem teórico e visual eram muito similares, atingindo classificações próximas de 40 valores percentuais. E os alunos com estilo de aprendizagem práticos apenas alcançaram resultados inferiores a 20 valores percentuais.

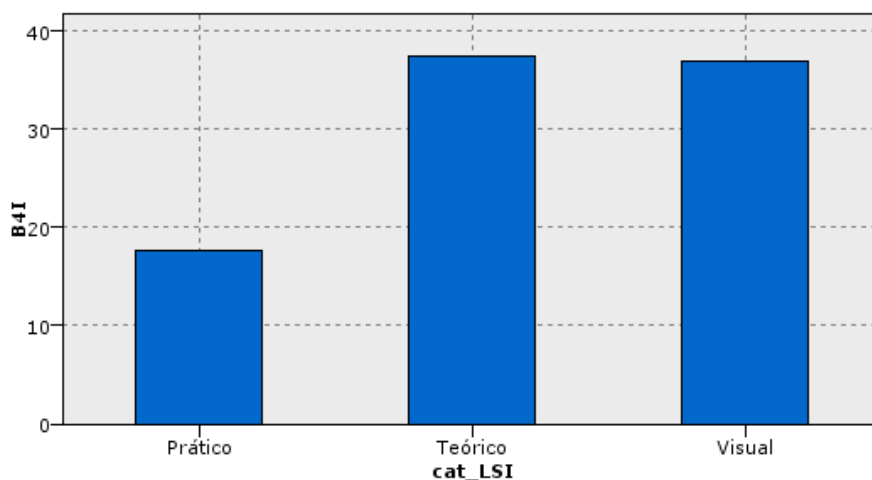


Figura 50 - Classificação inicial do B4 por estilo de aprendizagem

Os resultados do conceito B4, figura 51, demonstraram que os alunos com estilo de aprendizagem prático obtiveram classificações inferiores a 40 valores percentuais, os alunos com estilo de aprendizagem teórico atingiram valores superiores a 80 e os alunos com estilo de aprendizagem visual apresentaram classificações de 90 valores percentuais.

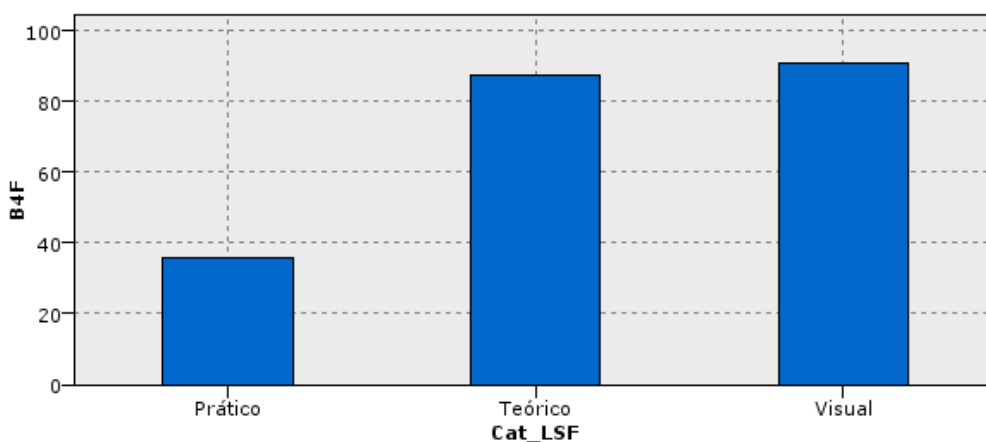


Figura 51 - Classificação final do B4 por estilo de aprendizagem

As classificações obtidas nas duas fases de avaliação do conceito B4 indicaram que a aprendizagem deste conceito foi realizada com sucesso, nomeadamente nos alunos com estilo de aprendizagem teórico e visual. Os alunos com estilo de aprendizagem prático registaram mais dificuldades na aquisição deste conceito.

A avaliação inicial do conceito B5 (método de redução à unidade), figura 52, demonstrou a ausência deste conhecimento na maioria dos alunos, os alunos com estilo de aprendizagem teórico obtiveram 0 valores percentuais, os alunos com estilo de aprendizagem prático atingiram resultados próximos de 5 valores percentuais e os alunos com estilo de aprendizagem visual alcançaram classificações superiores a 15 valores percentuais.

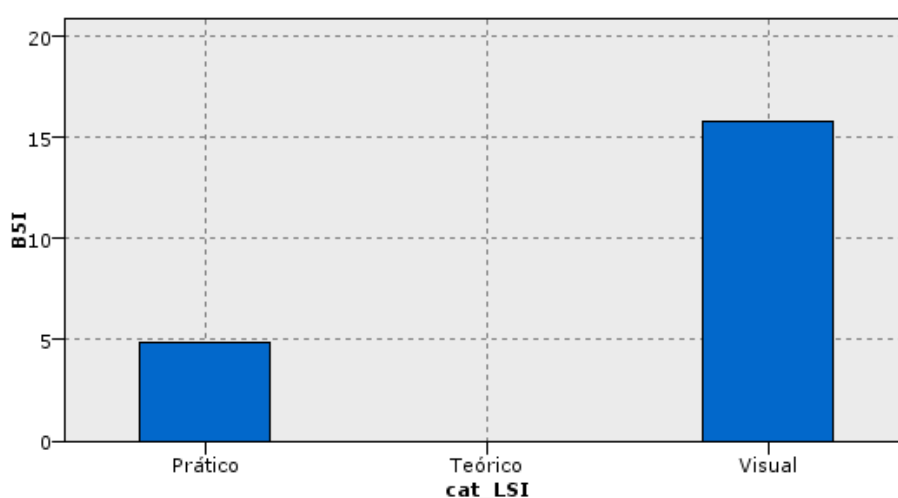


Figura 52 - Classificação inicial do B5 por estilo de aprendizagem

Em relação à avaliação final do conceito B5, figura 53, foi verificado que os alunos com estilo de aprendizagem prático apresentaram classificações inferiores a 5 valores percentuais e os alunos com estilo de aprendizagem visual atingiram resultados inferiores a 25 valores percentuais. Os alunos com estilo de aprendizagem teórico não conseguiram alcançar respostas certas em relação a este conceito.

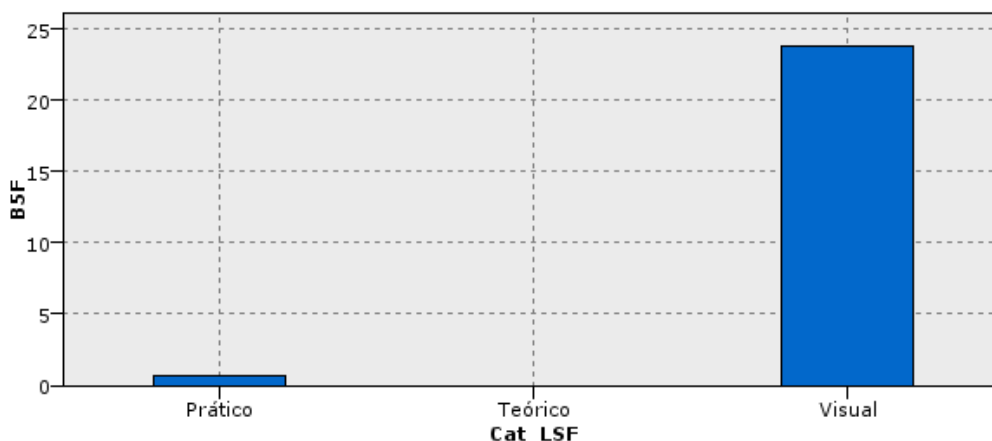


Figura 53 - Classificação final do B5 obtida pelos alunos e por estilo de aprendizagem

Os resultados obtidos nas duas fases distintas de avaliação do conceito B5 evidenciaram o insucesso da aprendizagem deste conceito. Apenas os alunos com estilo de aprendizagem visual conseguiram aumentar os seus resultados, contudo não atingiram resultados positivos. Os alunos com estilo de aprendizagem prático regrediram os seus conhecimentos e os alunos com estilo de aprendizagem teórico mantiveram os resultados obtidos na fase inicial de avaliação. Estas classificações demonstraram que a aprendizagem deste conceito não é a adequada aos alunos e não favorece a aquisição do conhecimento.

Após a análise das classificações dos conceitos por estilo de aprendizagem, foi concluído que o estilo de aprendizagem era um fator que condicionou o percurso da aprendizagem. A maioria dos alunos com estilo de aprendizagem visual obtiveram melhores resultados e os alunos com estilo de aprendizagem prático apresentaram as piores classificações. Em suma, o estilo de aprendizagem visual foi considerado o estilo mais adequado para proporcionar melhores resultados na aprendizagem.

5.2.6 Análise de outros fatores na distribuição das classificações

Também se procedeu ao estudo das mudanças de estilo de aprendizagem dos alunos ao longo deste processo. A tabela 34 representa a informação extraída deste estudo, cerca de 17 alunos alteraram a sua preferência de aprendizagem, 15 dos quais alcançaram melhores resultados com esta mudança e 2 alunos regrediram.

Tabela 34 - Mudança de estilo de aprendizagem no processo de aprendizagem

Progresso	Nº de alunos
Aumento	15
Diminuição	2
Total	17

Foi considerado útil analisar o tipo de ensino, ensino especial e normal, verificando se esta característica consistia num fator condicionante para a aprendizagem em questão. Desta forma, foi analisada a distribuição das classificações dos alunos por tipo de ensino na avaliação de diagnóstico e final.

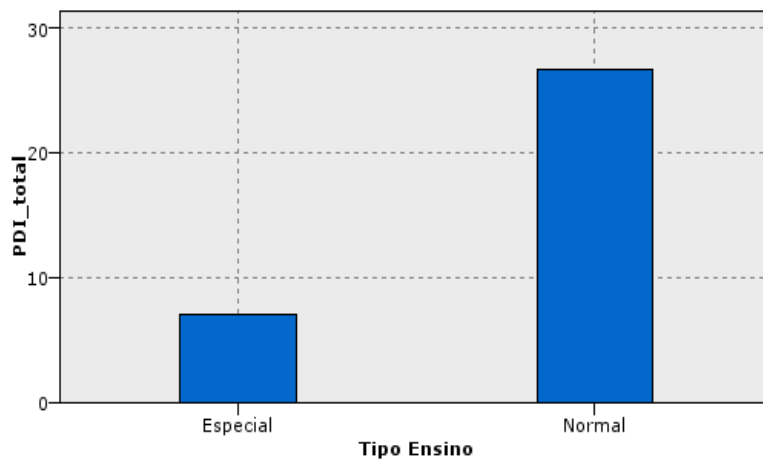


Figura 54 - Classificação no teste diagnóstico por tipo de ensino

A figura 54 exibe a avaliação dos resultados do teste diagnóstico e permitiu verificar que os alunos com necessidades especiais, ensino especial, exibiram classificações inferiores a 10 valores percentuais. Os alunos com ensino normal registaram classificações superiores a 20 valores percentuais.

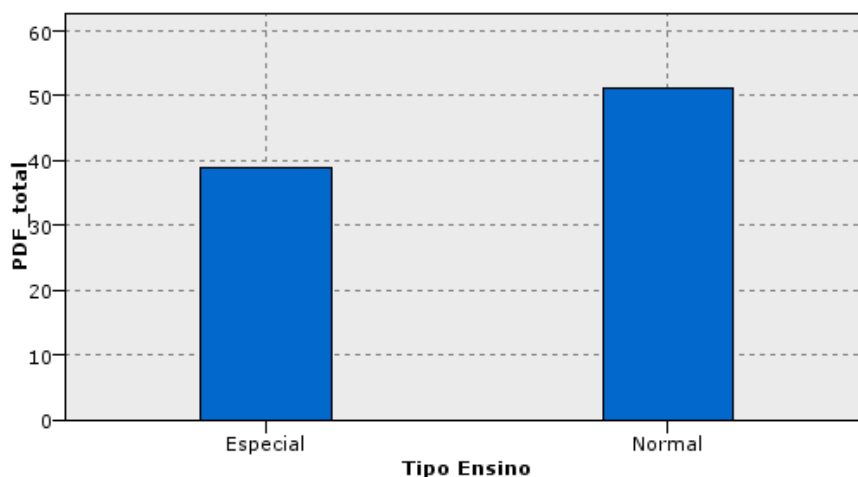


Figura 55 - Classificação no teste final por tipo de ensino

Os resultados do teste final, figura 55, demonstraram que os alunos com ensino especial obtiveram classificações próximas de 40 valores percentuais e os alunos com ensino normal atingiram classificações superiores a 50 valores percentuais.

A análise da distribuição das classificações dos alunos por tipo de ensino no teste diagnóstico e final demonstrou que, independentemente do tipo de ensino, os alunos aumentaram os seus resultados obtidos após a aprendizagem da unidade lecionada, concluindo que o tipo de ensino especial não inviabiliza a aprendizagem dos alunos, contudo o aumento de resultados obtidos não é tão expressivo como no ensino normal.

O processo de aprendizagem também pode ser condicionado pela característica grupo, grupo experimental e controlo, sendo assim foi necessário analisar a distribuição das classificações dos alunos por grupo no teste diagnóstico e final.

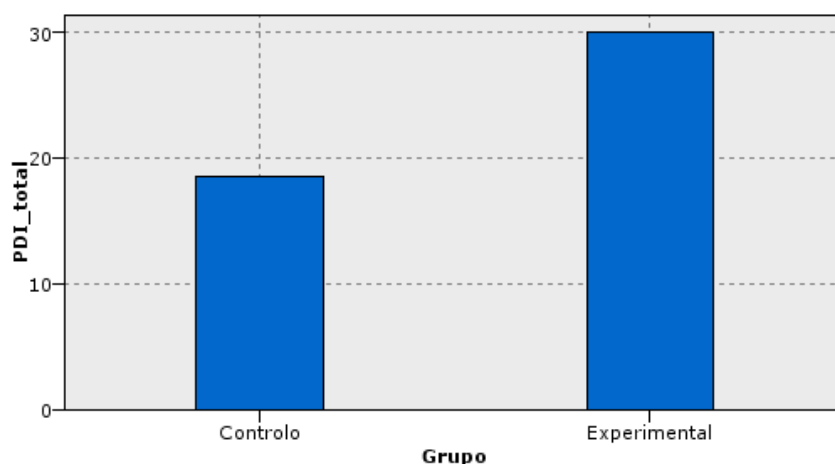


Figura 56 - Classificação no teste diagnóstico por grupo

No teste diagnóstico, figura 56, os alunos que pertenciam ao grupo de controlo obtiveram classificações inferiores a 20 valores percentuais e os alunos do grupo experimental apresentaram resultados próximos de 30 valores percentuais.

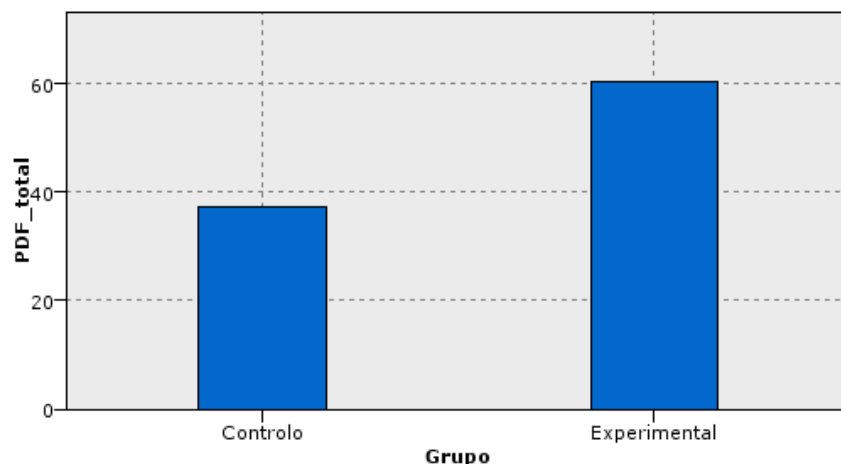


Figura 57 - Classificação no teste final por grupo

No teste final, figura 57, os alunos do grupo de controlo obtiveram classificações inferiores a 40 valores percentuais e os alunos do grupo experimental exibiram classificações próximas de 60 valores percentuais.

A distribuição das classificações dos alunos por grupo evidenciou que esta característica condicionou o processo de aprendizagem, foi verificado que o grupo experimental, os alunos que usaram o PCMAT para aprender, obtiveram melhores resultados do que os alunos do grupo de controlo, recorrendo ao método tradicional de ensino.

5.2.7 Classificações por alunos médios

Neste estudo foi considerado interessante analisar regras e padrões existentes em determinados grupos de alunos. Sendo assim, os alunos foram divididos em diferentes grupos de acordo com os resultados obtidos na avaliação final:

- Alunos maus: alunos que obtiveram classificações finais inferiores a 50 valores percentuais;
- Alunos médios: alunos que obtiveram classificações finais entre os 50 e 65 valores percentuais.

- Alunos bons: alunos que obtiveram classificações finais superiores a 66 valores percentuais.

O grupo de alunos médios foi analisado tendo como objetivo encontrar tendências e/ou relações existentes, por exemplo, identificar quais os conceitos que alcançaram melhores e piores resultados na avaliação final em relação aos alunos médios.

Ao analisar a distribuição do grupo de alunos médios, figura 58, foi verificado a existência de aproximadamente 56% de alunos do sexo feminino e os restantes 44% alunos do sexo masculino.

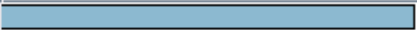

Value ▲	Proportion	%	Count
Feminino		55,56	20
Masculino		44,44	16

Figura 58- Distribuição dos alunos médios por sexo

A maioria dos alunos médios, figura 59, cerca de 80% dos alunos, pertencia ao grupo experimental, a aprendizagem era suportada pelo PCMAT, e os restantes 17% dos alunos apresentavam-se no grupo de controlo, usavam o método tradicional de ensino.

Value ▲	Proportion	%	Count
Controlo		16,67	6
Experimental		83,33	30

Figura 59 - Distribuição dos alunos médios por grupo

Na fase inicial do processo de aprendizagem os alunos foram submetidos a um questionário que identificava o estilo de aprendizagem de cada aluno, figura 60, cerca de 39% dos alunos exibiam estilo de aprendizagem prático e visual e os restantes alunos, cerca de 22%, possuem o estilo de aprendizagem teórico.




Value ▲	Proportion	%	Count
Prático		38,89	14
Teórico		22,22	8
Visual		38,89	14

Figura 60 - Distribuição dos alunos médios por estilo de aprendizagem inicial

Na fase final do processo de aprendizagem os alunos foram submetidos novamente a outro questionário para avaliar o seu estilo de aprendizagem, figura 61, foi registado que cerca de 22% dos alunos com estilo de aprendizagem teórico e visual e aproximadamente 56% de alunos com estilo de aprendizagem prático.

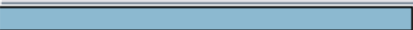


Value ▲	Proportion	%	Count
Prático		55,56	20
Teórico		22,22	8
Visual		22,22	8

Figura 61 - Distribuição dos alunos médios por estilo de aprendizagem final

Analisando e comparando a fase inicial e final do processo de aprendizagem, foi verificado que o número de alunos teóricos se manteve estável e que ocorreu uma diminuição do número de alunos com estilo de aprendizagem visual e um aumento do número de alunos com estilo de aprendizagem prático. Em suma, este processo de aprendizagem favoreceu o aumento da quantidade de alunos com estilo de aprendizagem prático.

Na perspetiva de compreender melhor as influências e/ou relações entre os conceitos, recorreu-se ao nó *web* da ferramenta *Clementine* para identificar potenciais relações entre os distintos conceitos na fase inicial e final do processo de aprendizagem.

Na avaliação inicial dos conceitos foram identificadas relações fortes entre as classificações de "Muito Mau" dos conceitos A3 e B2, sendo ilustradas na figura 62.

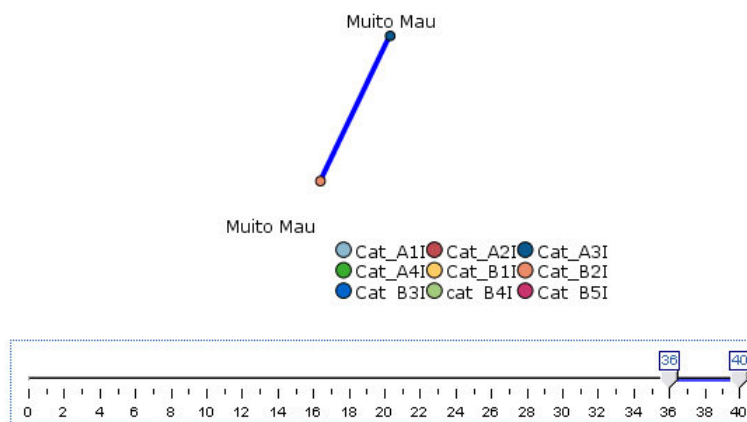


Figura 62 - Relações entre os conceitos na avaliação inicial dos alunos médios

Na avaliação final dos conceitos foram encontradas relações fortes entre as classificações de "Muito Bom" do conceito B4 e "Muito Mau" do conceito B5, sendo representadas na figura 63.

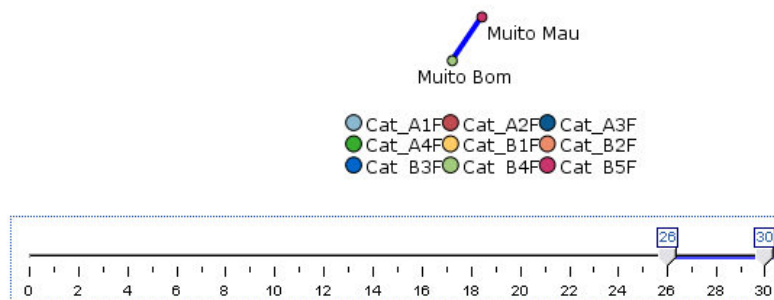


Figura 63 - Relações entre os conceitos na avaliação final dos alunos médios

Para analisar melhor a evolução dos alunos médios no processo de aprendizagem foi efetuada uma comparação entre os resultados obtidos pelos alunos no teste diagnóstico e final.

No teste diagnóstico os alunos do sexo feminino atingiram resultados inferiores a 30 valores percentuais e os do sexo masculino alcançaram resultados aproximadamente de 33 valores percentuais, sendo ilustrados na figura 64.

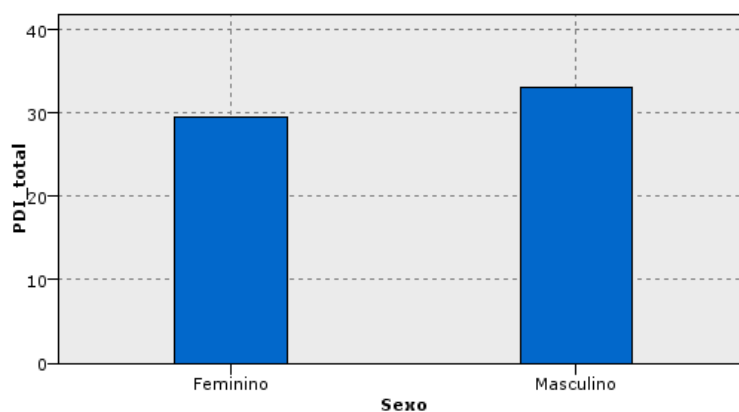


Figura 64 - Classificação inicial dos alunos médios por sexo

No teste final, figura 65, os alunos médios alcançaram resultados situados entre os 50 a 60 valores percentuais, foram verificados melhores resultados nos alunos do sexo feminino.

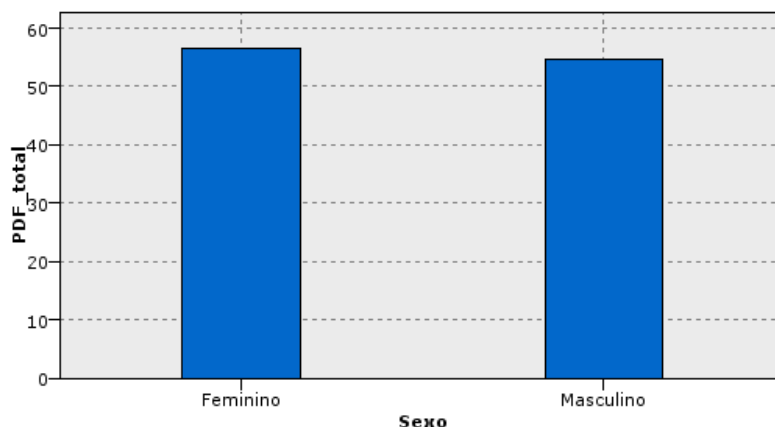


Figura 65 - Classificação final dos alunos médios por sexo

Para ajudar a compreensão do processo de aprendizagem de conceitos foi considerado importante analisar os resultados obtidos na fase inicial e final deste processo, tabela 35, identificando os conceitos com melhores e piores resultados.

Tabela 35 - Classificação total por conceito obtida pelos alunos médios

Conceito	Avaliação		Progresso
	Inicial	Final	
A1	1000	1200	200
A2	1900	3000	1100
A3	0	1680	1680
A4	1050	2736	1686
B1	2000	2100	100
B2	0	1680	1680
B3	2300	3000	700
B4	1400	2600	1200
B5	400	0	-400

Na fase inicial, foram identificados o A3 (leitura de uma razão) e B2 (leitura de proporções) como os conceitos com piores resultados e o B3 (propriedade de proporções) com os melhores resultados. Na avaliação final foi averiguado que o conceito B5 (método de redução à unidade) apresentava as piores classificações e o A2 (noções de antecedente e conseqüente) e B3 (propriedade de proporções) exibiram as melhores classificações.

Comparando os dados obtidos nas duas fases distintas do processo de aprendizagem foi constatado que o conceito mais simples de ser adquirido pelos alunos foi o A3 (leitura de uma razão), A4 (aplicações do conceito razão) e B2 (leitura de proporções). Os conceitos A1 (razões

equivalentes), B1 (meios e extremos) e B5 (método de redução à unidade) apresentaram as piores classificações ao longo deste processo, sendo classificados como os conceitos mais difíceis de assimilar pelos alunos. Apesar das dificuldades sentidas na aquisição dos conceitos mencionados anteriormente, os conceitos A1 e B1 exibiram um ligeiro aumento nos resultados obtidos na fase final de avaliação. Ao contrário, o conceito B5 registou uma diminuição de resultados, demonstrando que a aprendizagem deste conceito não era adequada aos alunos.

No sentido de entender o progresso de aprendizagem dos conceitos foi realizada a comparação entre os resultados obtidos na fase inicial e final de avaliação de cada conceito.

Na avaliação inicial do conceito A1 (razões equivalentes), figura 66, os alunos médios com estilo de aprendizagem teórico obtiveram resultados de 50 valores percentuais. Os alunos com estilo de aprendizagem prático alcançaram resultados próximos de 30 valores percentuais e os alunos com estilo de aprendizagem visual apresentaram resultados inferiores a 20 valores percentuais. Em suma, no teste diagnóstico os alunos com estilo teórico obtiveram melhores resultados e os alunos com estilo visual apresentaram os piores resultados.

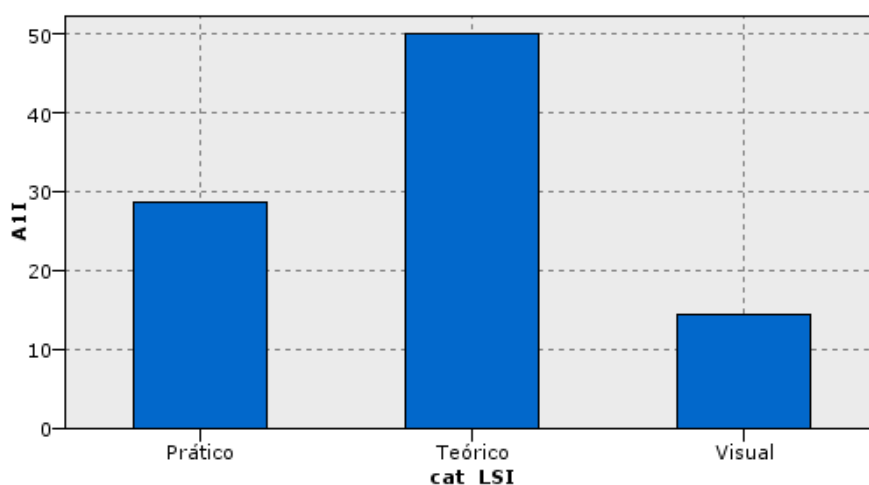


Figura 66 - Classificação inicial do A1 nos alunos médios e por estilo de aprendizagem inicial

Na avaliação final do conceito A1, figura 67, foi verificado que os alunos com o estilo de aprendizagem visual obtiveram os melhores resultados, aproximadamente 50 valores percentuais. Os alunos com estilo de aprendizagem prático obtiveram 30 valores percentuais e os alunos com estilo teórico atingiram 25 valores percentuais.

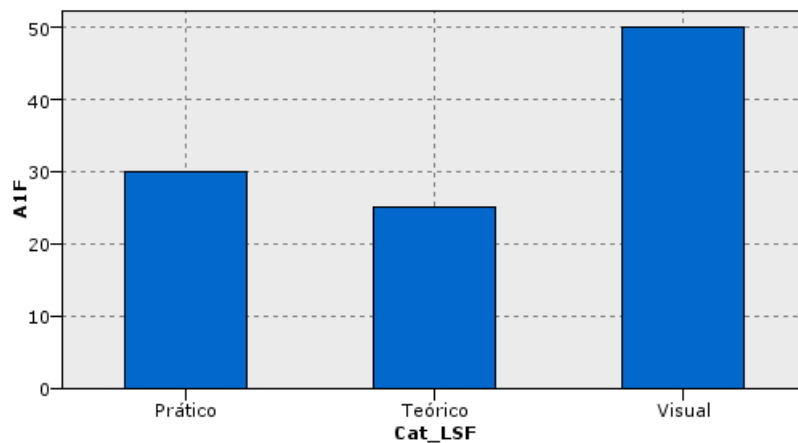


Figura 67 - Classificação final do A1 nos alunos médios e por estilo de aprendizagem final

Comparando os resultados atingidos no teste diagnóstico e final foi concluído que os alunos com estilo de aprendizagem visual obtiveram um progresso significativo, os alunos com estilo de aprendizagem prático registaram um ligeiro aumento nos resultados obtidos e, curiosamente, os alunos com estilo de aprendizagem teórico apresentaram uma significativa diminuição dos resultados obtidos. Em suma, a aprendizagem deste conceito não foi realizada com sucesso, tendo em conta que os alunos com estilo de aprendizagem prático e teórico não apresentaram classificações positivas.

O teste diagnóstico relativo ao conceito A2 (noções de antecedente e consequente), figura 68, indicou que os alunos com estilo de aprendizagem teórico obtiveram as melhores classificações, 100 valores percentuais, demonstrando total conhecimento deste conceito. Seguidamente, os alunos com estilo de aprendizagem visual exibiram resultados próximos de 50 valores percentuais e os alunos com estilo de aprendizagem prático atingiram os piores resultados, 30 valores percentuais.

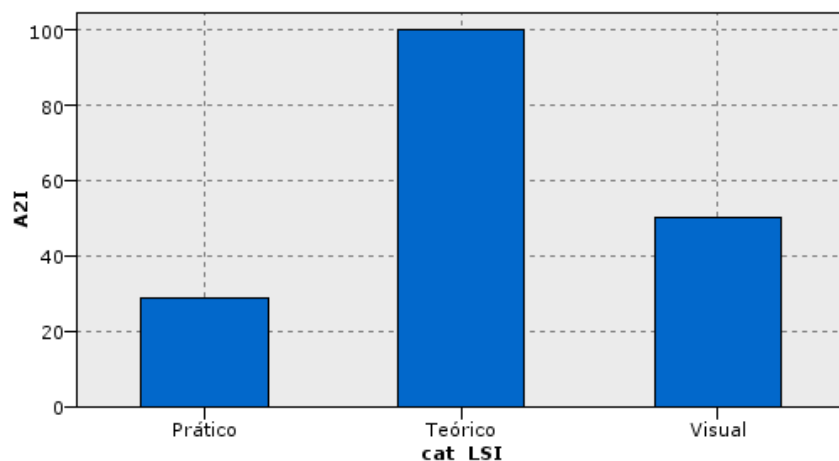


Figura 68 - Classificação inicial do A2 nos alunos médios e por estilo de aprendizagem inicial

Na avaliação final do conceito A2, figura 69, os alunos com estilo de aprendizagem teórico mantiveram as classificações obtidas na fase inicial, 100 valores percentuais, os alunos com estilo de aprendizagem prático obtiveram resultados de 90 valores percentuais e os alunos com estilo de aprendizagem visual atingiram 50 valores percentuais.

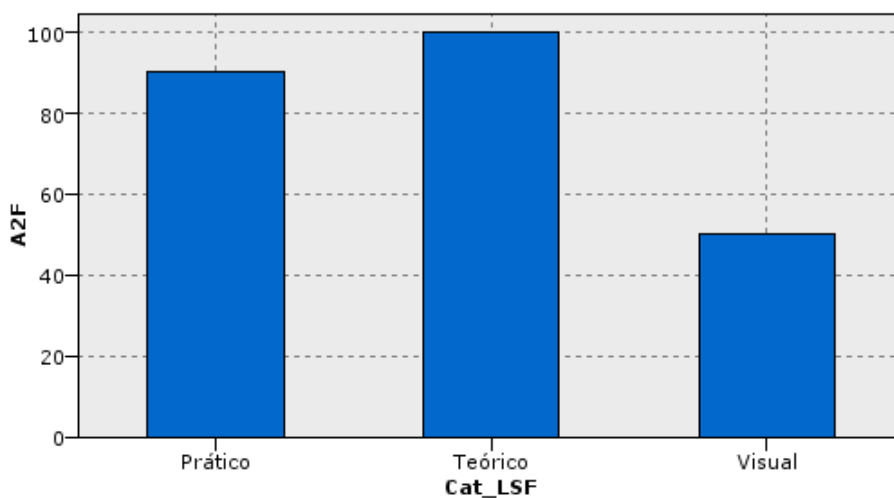


Figura 69 - Classificação final do A2 nos alunos médios e por estilo de aprendizagem final

Após a análise dos resultados obtidos nas duas fases do processo de aprendizagem, foi verificado um aumento significativo nos resultados obtidos pelos alunos com estilo de aprendizagem prático, os restantes alunos com estilo de aprendizagem teórico e visual mantiveram os resultados iguais à fase inicial. A aprendizagem deste conceito foi realizada com sucesso, nomeadamente nos alunos de estilo de aprendizagem prático e teórico, que obtiveram resultados satisfatórios. Contudo os alunos com estilo de aprendizagem visual demonstraram algumas dificuldades em alcançar classificação positiva.

A avaliação inicial do conceito A3 (leitura de uma razão), figura 70, demonstrou que os alunos médios não tinham conhecimento deste conceito. Independentemente da preferência de aprendizagem dos alunos os resultados foram semelhantes, registrando 0 valores percentuais.

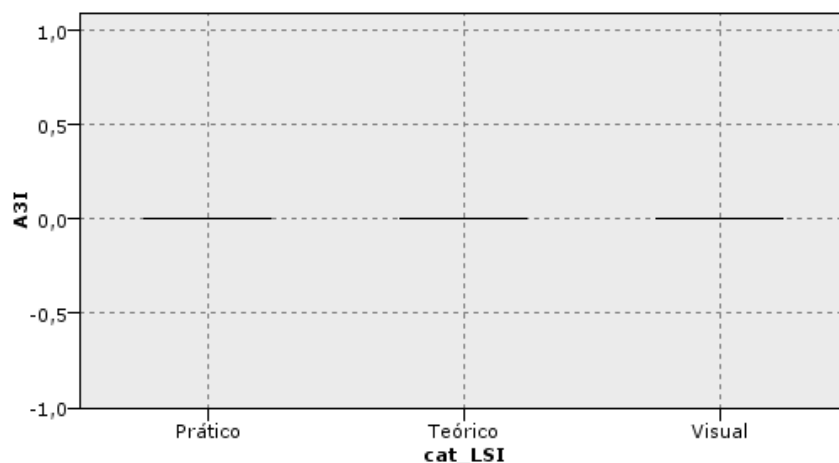


Figura 70 - Classificação inicial do A3 nos alunos médios e por estilo de aprendizagem inicial

Após a aprendizagem deste conceito, figura 71, os alunos com estilo de aprendizagem visual obtiveram os melhores resultados, registrando 60 valores percentuais, os alunos com estilo de aprendizagem teórico alcançaram resultados de 50 valores percentuais e os alunos com estilo de aprendizagem prático apresentaram resultados de 40 valores percentuais.

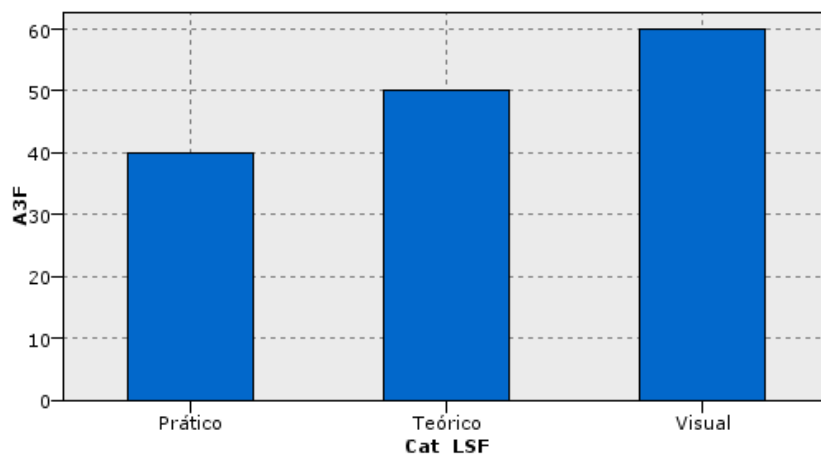


Figura 71 - Classificação final do A3 nos alunos médios e por estilo de aprendizagem final

Os resultados obtidos na fase final demonstraram o sucesso de aprendizagem do conceito A3, sendo que a maioria dos alunos não detinha conhecimento deste conceito e, após a sua aprendizagem, os alunos alcançaram resultados satisfatórios, nomeadamente os alunos com estilo de aprendizagem teórico e visual.

Em relação ao conceito A4 (aplicações do conceito razão), figura 72, os resultados obtidos indicaram que os alunos com estilo de aprendizagem teórico obtiveram as melhores classificações, cerca de 38 valores percentuais. Os alunos com estilo de aprendizagem prático e visual exibiram resultados semelhantes, aproximadamente 25 e 29 valores percentuais respectivamente.

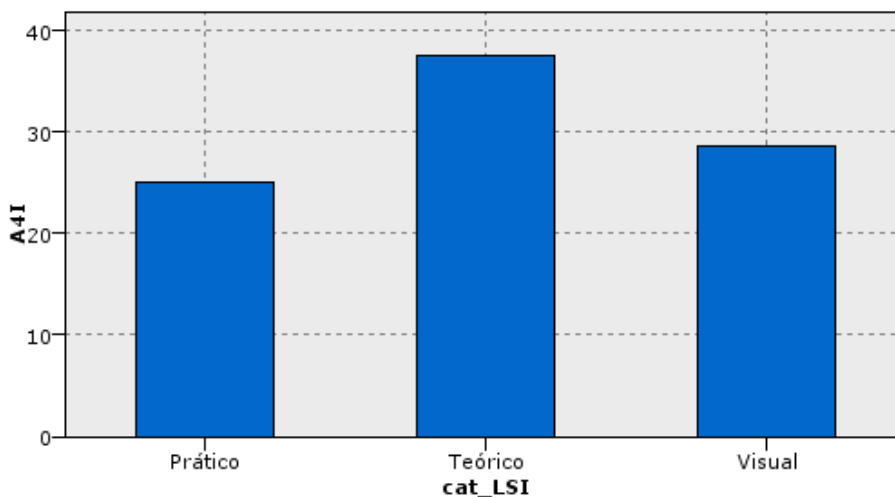


Figura 72 - Classificação inicial do A4 nos alunos médios e por estilo de aprendizagem inicial

Na fase final do processo de aprendizagem, figura 73, os alunos com estilo de aprendizagem prático e visual obtiveram resultados positivos, aproximadamente 83 e 100 valores percentuais respectivamente. E os alunos com estilo de aprendizagem teórico registaram as piores classificações, valores percentuais inferiores a 40.

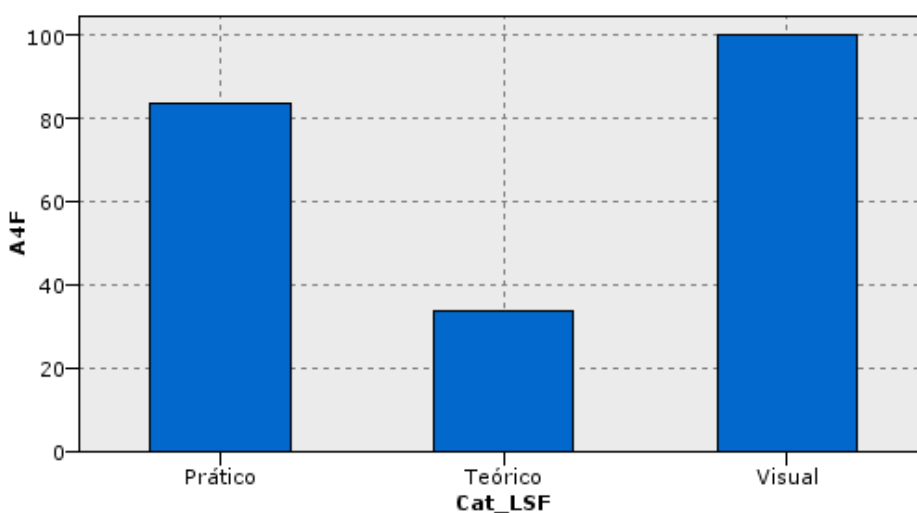


Figura 73 - Classificação final do A4 nos alunos médios e por estilo de aprendizagem final

A maioria dos alunos obtiveram sucesso na aquisição deste conceito, nomeadamente os alunos com estilo de aprendizagem prático e visual, registando um aumento bastante significativo após a aprendizagem. Os alunos com estilo de aprendizagem teórico mantiveram as classificações obtidas inicialmente, demonstrando o insucesso da aprendizagem para este estilo de aprendizagem.

Relativamente ao conceito B1 (meios e extremos), figura 74, a avaliação inicial demonstrou que a maioria dos alunos tinha algum conhecimento deste conceito, nomeadamente os alunos com estilo de aprendizagem prático e teórico, registando aproximadamente 70 e 85 valores percentuais respetivamente. Os alunos com estilo de aprendizagem visual exibiram resultados baixos, cerca de 20 valores percentuais.

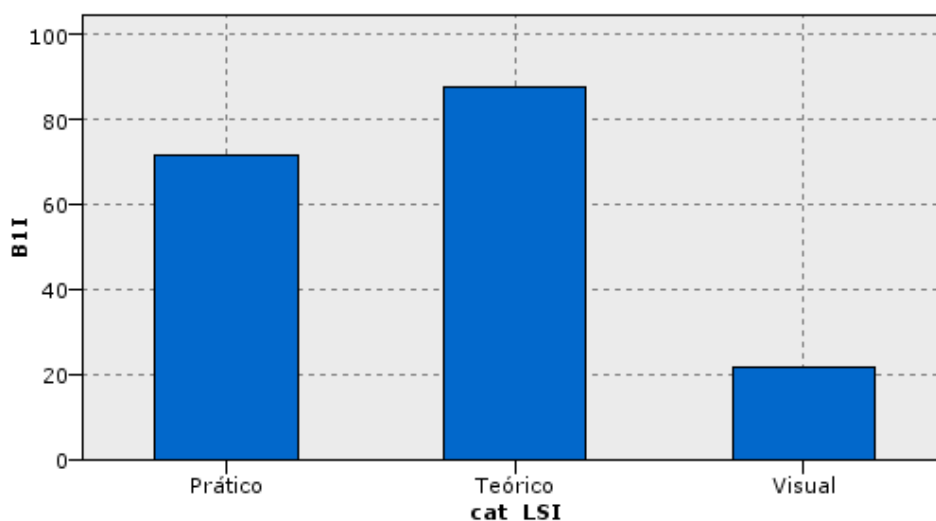


Figura 74 - Classificação inicial do B1 nos alunos médios e por estilo de aprendizagem inicial

No teste final, figura 75, os alunos com estilo de aprendizagem prático obtiveram resultados de aproximadamente 75 valores percentuais, seguidamente os alunos com estilo de aprendizagem teórico apresentaram resultados de 50 valores percentuais e os alunos com estilo de aprendizagem visual registaram resultados próximos de 25 valores percentuais.

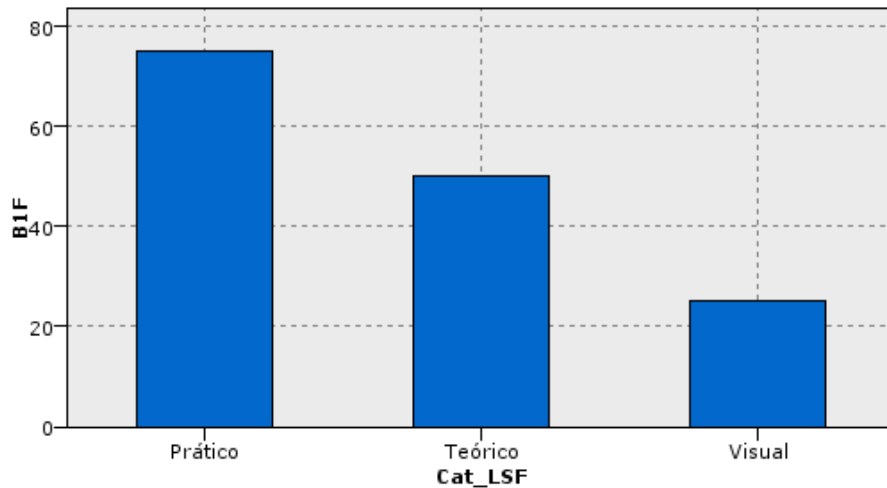


Figura 75 - Classificação final do B1 nos alunos médios e por estilo de aprendizagem final

Comparando os resultados obtidos na avaliação inicial e final do conceito B1 foi verificado que os alunos com o estilo de aprendizagem prático e visual mantiveram os mesmos resultados e, curiosamente, os alunos com estilo de aprendizagem teórico exibiram uma acentuada queda nos resultados, demonstrando que a aprendizagem deste conceito não foi realizada com sucesso.

A análise dos resultados obtidos no teste diagnóstico do conceito B2 (leitura de proporções), figura 76, demonstrou que independentemente do estilo de aprendizagem os alunos não tinham conhecimentos deste conceito.

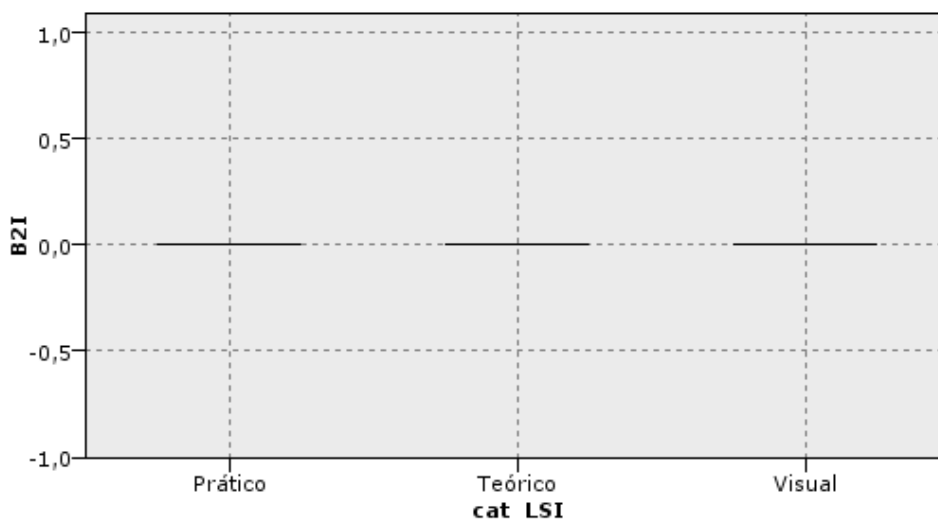


Figura 76 - Classificação inicial do B2 nos alunos médios e por estilo de aprendizagem inicial

Após a aprendizagem deste conceito, figura 77, foi constatado que a maioria dos alunos, alunos com estilo de aprendizagem teórico e prático, exibiram resultados razoáveis, 50 e 60 valores percentuais respetivamente. Os alunos com estilo de aprendizagem prático foram os que apresentaram piores resultados, 40 valores percentuais.

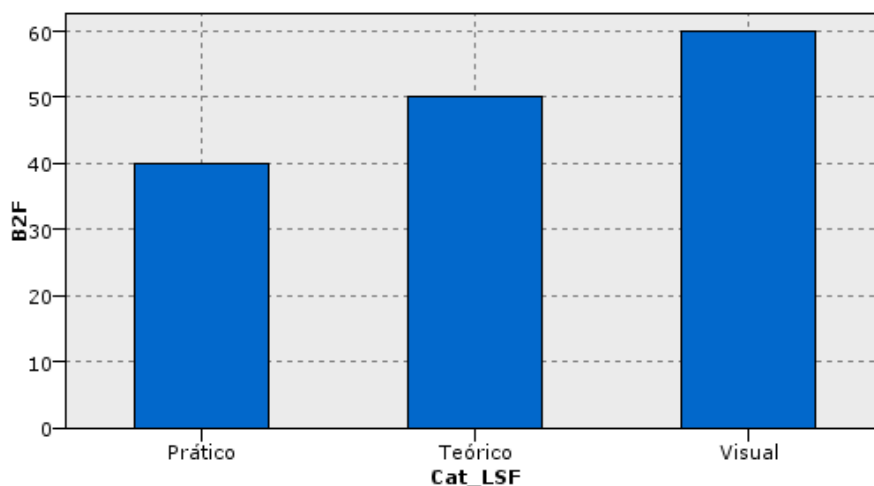


Figura 77 - Classificação final do B2 nos alunos médios e por estilo de aprendizagem final

Ao analisar a evolução da aprendizagem do conceito B2 foi averiguado que os alunos que exibiram o estilo de aprendizagem visual e teórico apresentaram melhores resultados, indicando que seria conveniente alterar a aprendizagem orientada aos alunos com estilo de aprendizagem prático, favorecendo assim a aprendizagem deste conceito.

Na fase inicial de avaliação do conceito B3 (propriedade das proporções), figura 78, os alunos com estilo de aprendizagem teórico obtiveram classificações de 100 valores percentuais, indicando o total domínio neste conceito. Seguidamente, os alunos com estilo de aprendizagem prático atingiram resultados próximos de 60 valores percentuais e os alunos com estilo de aprendizagem visual alcançaram resultados de aproximadamente 50 valores percentuais. De forma geral, foi possível afirmar que todos os alunos tinham alguns conhecimentos relativos a este conceito.

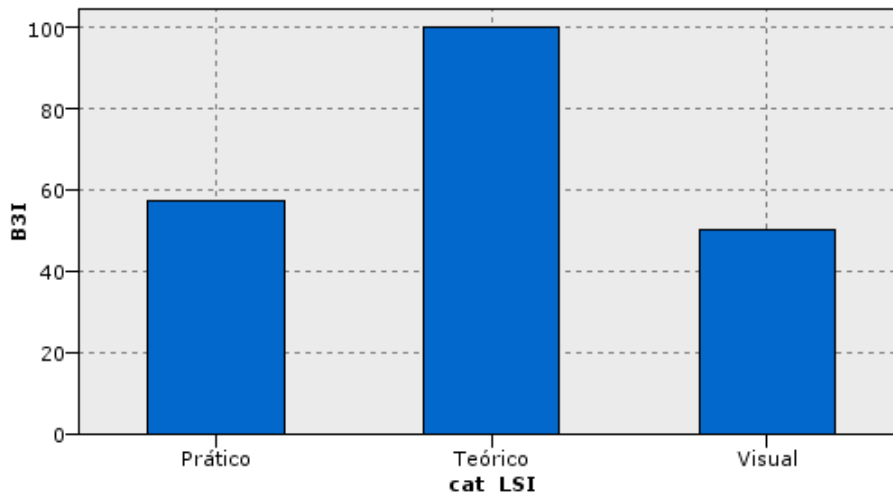


Figura 78 - Classificação inicial do B3 nos alunos médios e por estilo de aprendizagem inicial

Após a aprendizagem deste conceito, figura 79, foi observado um aumento nos resultados obtidos pelos alunos, os alunos com estilo de aprendizagem teórico mantiveram os melhores resultados, atingindo 100 valores percentuais. Os alunos com estilo de aprendizagem visual apresentaram valores próximos de 90 valores percentuais e os alunos com estilo de aprendizagem prático exibiram resultados aproximadamente de 78 valores percentuais.

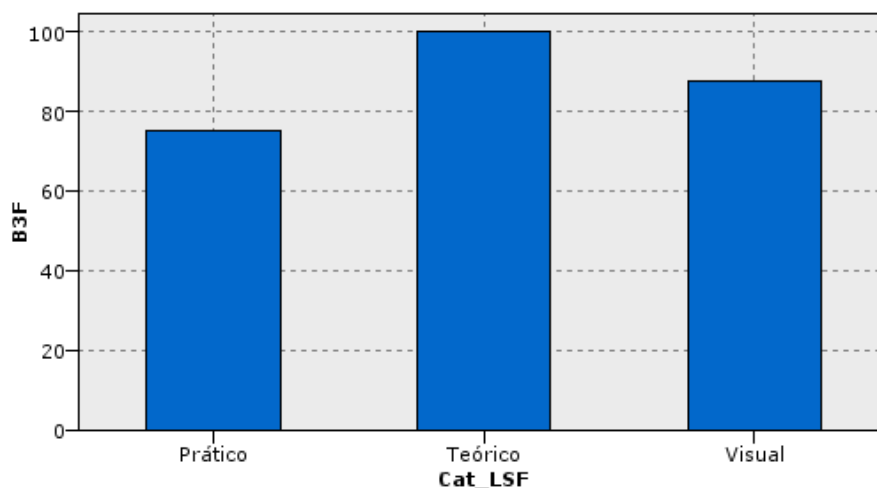


Figura 79 - Classificação final do B3 nos alunos médios e por estilo de aprendizagem final

Em suma, todos os alunos obtiveram aproveitamento no processo de aprendizagem deste conceito, realçando que os alunos já possuíam alguns conhecimentos inicialmente.

No teste diagnóstico relativo ao conceito B4 (regra de três simples), figura 80, os alunos com estilo de aprendizagem teóricos exibiram os melhores resultados, atingindo cerca de 50

valores percentuais, os alunos com estilo de aprendizagem visuais obtiveram resultados superiores a 40 valores percentuais e os alunos com estilo de aprendizagem prático alcançaram resultados próximos de 30 valores percentuais.

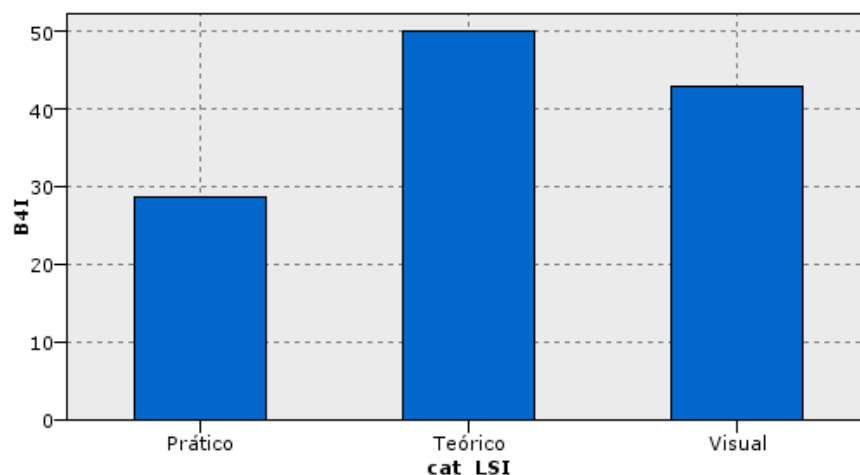


Figura 80 - Classificação inicial do B4 nos alunos médios e por estilo de aprendizagem inicial

Após a aprendizagem do conceito B4, figura 81, a maioria dos alunos, nomeadamente alunos com estilo de aprendizagem teórico e visual, atingiram classificação de 100 valores percentuais, demonstrando total domínio neste conceito. Os alunos com estilo de aprendizagem prático obtiveram classificações de 50 valores percentuais.

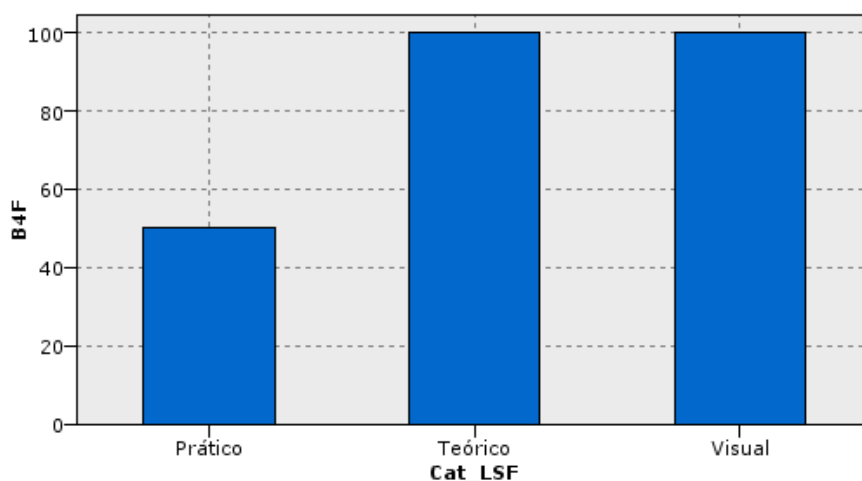


Figura 81 - Classificação final do B4 nos alunos médios e por estilo de aprendizagem final

A aprendizagem deste conceito foi realizada com sucesso, os alunos dos diversos estilos de aprendizagem conseguiram atingir resultados satisfatórios, averiguando que os alunos com estilo de aprendizagem prático exibiram maiores dificuldades na aprendizagem do conceito B4.

Em relação ao conceito B5 (método de redução à unidade), figura 82, os resultados obtidos na fase inicial indicaram que a maioria dos alunos não apresentava conhecimentos deste conceito, nomeadamente os alunos com estilo de aprendizagem teórico e visual. Os alunos com estilo de aprendizagem prático atingiram classificações próximas de 30 valores percentuais.

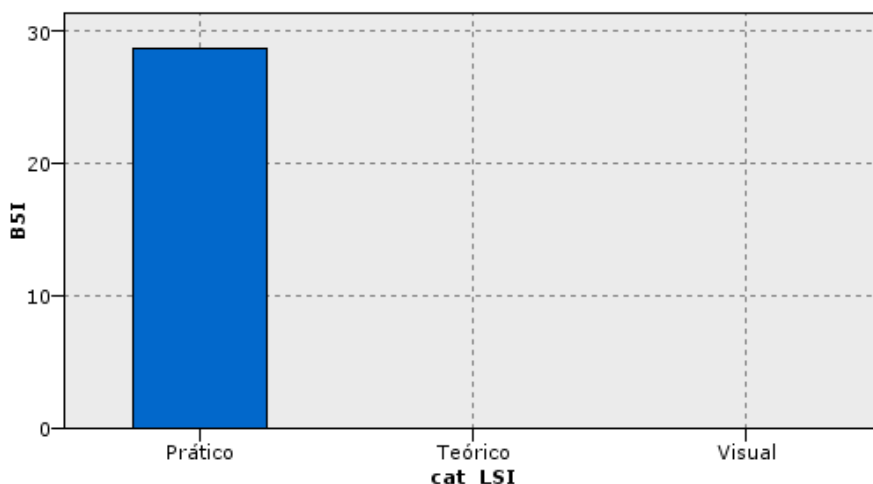


Figura 82 - Classificação inicial do B5 nos alunos médios e por estilo de aprendizagem inicial

Na fase final da avaliação, figura 83, foi averiguado que todos os alunos, independente do estilo de aprendizagem, atingiram 0 valores percentuais, demonstrando que este conceito não foi adquirido pelos alunos. Não esquecendo de mencionar que os alunos com estilo de aprendizagem prático regrediram a nível de conhecimento deste conceito.

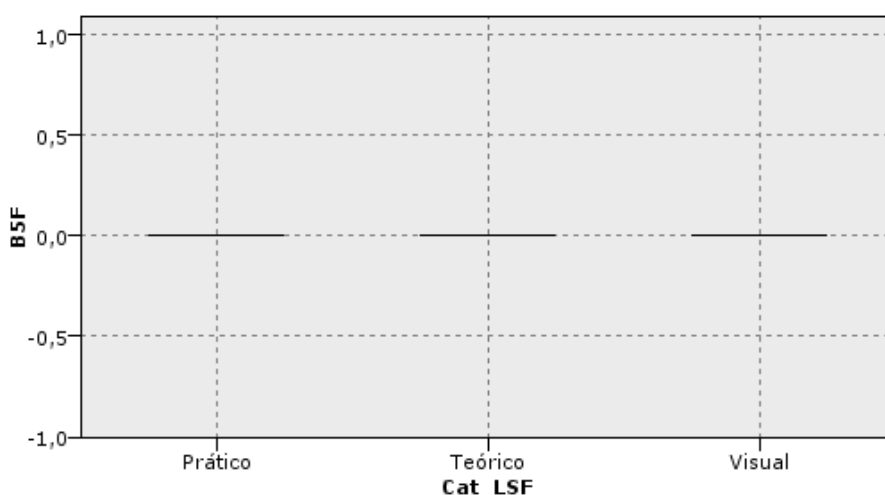


Figura 83 - Classificação final do B5 nos alunos médios e por estilo de aprendizagem final

Após a análise do progresso de aprendizagem do conceito B5 foi possível concluir que a aquisição deste conceito não foi realizada com sucesso e que, curiosamente, ocorreu uma

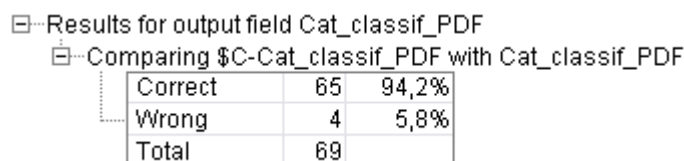
regressão de conhecimentos deste conceito em relação aos alunos com estilo de aprendizagem prático. É conveniente alterar a aprendizagem deste conceito, tornando este processo mais adequado aos alunos e favorecendo a melhoria dos resultados obtidos.

5.3 Classificação

Tendo em conta o reduzido tamanho da amostra de dados apenas foi possível utilizar a técnica de DM classificação na construção do modelo de dados.

Neste trabalho, o objetivo do modelo de classificação de dados permitiu identificar quais as características dos alunos que exerciam maior influência nos melhores resultados das classificações finais, por exemplo, estilo de aprendizagem, idade, tipo de ensino, entre outros. Para alcançar este objetivo foram criados diversos modelos de classificação de dados, recorrendo a distintos algoritmos de classificação, nomeadamente C&Rtree, C5.0 e redes neuronais.

Os modelos de classificação de dados criados obtiveram taxas de confiança de 100%, demonstrando sobre ajustamento de dados (*overfitting*) (Rodrigues, 2008). Esta situação ocorre quando o modelo criado adapta-se muito bem aos casos de aprendizagem mas responde mal a novos casos. Para ultrapassar este tipo de situação foi necessário recorrer ao uso de alguns métodos, mais indicados para pequenos conjuntos de dados, amostras de dados com elevada variância e menor número de iterações (Zupan et al., 1993). No modelo de dados gerado pelo algoritmo C5.0 foi utilizado o método de validação cruzada (*cross-validation*), divisão do conjunto inicial de dados em k-subconjuntos de igual tamanho (Rodrigues, 2008), para contornar o sobre ajustamento de dados. Desta forma, o modelo de dados obteve, aproximadamente 94% de taxa de confiança.



	Count	Percentage
Correct	65	94,2%
Wrong	4	5,8%
Total	69	

Figura 84 - Taxa de confiança do modelo de classificação de dados recorrendo ao C5.0

Após a criação do modelo de classificação de dados recorrendo ao algoritmo C5.0, foi analisada a importância das características dos alunos que mais contribuíram nos resultados obtidos nas avaliações finais.

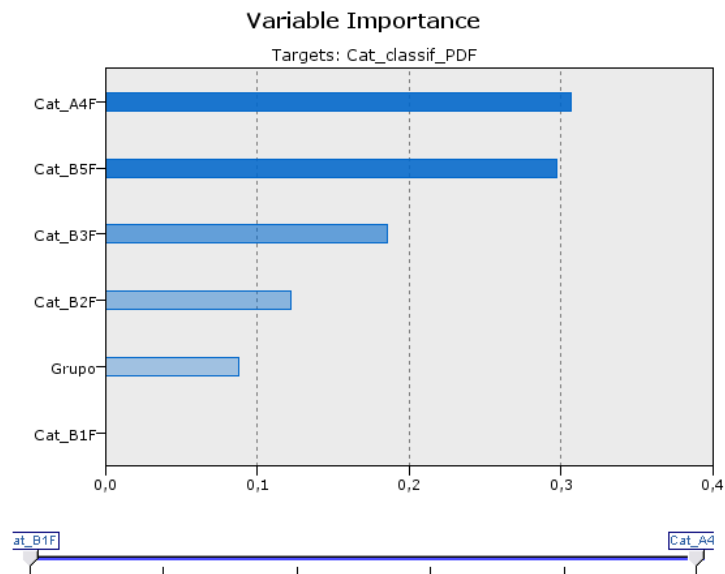


Figura 85 - Importância dos atributos segundo o modelo de dados recorrendo ao C5.0

A figura 85 identifica quais os atributos que exerceram maior influência nos resultados obtidos nas avaliações finais dos alunos. O atributo que mais contribuiu para os resultados das avaliações finais foi a classificação final do conceito A4, seguido da classificação final do conceito B5, classificação final do conceito B3, classificação final do conceito B2, o grupo (experimental ou controlo) e, por fim, a classificação final do conceito B1.

Através do modelo de classificação de dados foi possível extrair algumas regras:

- **Se** classificação final do conceito B5 = Muito Bom

Então avaliação final = Muito Bom;

Esta regra indicou que os alunos que obtiveram a classificação de "Muito Bom" na avaliação final apresentaram uma classificação final de "Muito Bom" em relação ao conceito B5. Os alunos que conseguiram adquirir este conceito obtiveram resultados muito bons na avaliação final, podendo concluir que este é o conceito mais complexo de ser aprendido pelos alunos.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Bom) **e** (grupo = Experimental)

Então avaliação final = Médio;

Esta regra indicou que os alunos do grupo experimental que obtiveram classificação de "Médio" na avaliação final e apresentaram resultados de "Muito Mau" e "Bom" em relação aos conceitos B5 e B2 respetivamente.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Bom) **e** (grupo = Controlo)

Então avaliação final = Mau;

A regra citada acima permitiu verificar que os alunos do grupo de controlo apresentaram classificação de "Mau" na avaliação final e exibiram resultados de "Muito Mau" e "Bom" relativamente aos conceitos B5 e B2 respetivamente.

As regras citadas anteriormente permitiram concluir que os alunos do grupo experimental conseguiram obter melhores resultados na avaliação final do que os alunos do grupo de controlo. Esta conclusão demonstrou que a aprendizagem dos alunos é favorecida pelo uso do PCMAT em paralelo com a aprendizagem tradicional.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Bom) **e** (classificação final do conceito B1 = Muito Bom)

Então avaliação final = Bom;

A regra anterior demonstra que os alunos que obtiveram classificação de "Bom" na avaliação final apresentaram classificações de "Muito Mau", "Muito Bom" nos conceitos B5, B2 e B1 respetivamente. Através desta regra foi possível concluir que um mau resultado obtido no conceito B5 não impede de alcançar uma classificação bastante razoável na avaliação final global.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Bom) **e** (classificação final do conceito B1 = (Muito Mau **ou** Médio))

Então avaliação final = Médio;

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Mau) **e** (classificação final do conceito A4 = (Muito Bom **ou** Bom))

Então avaliação final = Médio;

As duas regras anteriores indicaram que os alunos obtiveram classificações de "Médio" na avaliação final e apresentaram as classificações de "Muito Mau", "Muito

Mau", "Muito Mau" ou "Médio" / "Muito Bom" ou "Bom" nos conceitos B5, B2 e B1 / A4 respetivamente.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Mau) **e** (classificação final do conceito A4 = Mau)

Então avaliação final = Muito Mau;

A regra anterior indicou que os alunos apresentaram classificações de "Muito Mau" na avaliação final e obtiveram classificações de "Muito Mau", "Muito Mau", "Mau" nos conceitos B5, B2 e A4 respetivamente.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Mau) **e** (classificação final do conceito A4 = Muito Mau) **e** (classificação final do conceito B3 = Muito Mau **ou** Mau **ou** Médio)

Então avaliação final = Muito Mau;

A regra anterior demonstrou que os alunos obtiveram classificações de "Muito Mau" na avaliação final e obtiveram classificações de "Muito Mau", "Muito Mau", "Muito Mau" e "Muito Mau" ou "Mau" ou "Médio" nos conceitos B5, B2, A4 e B3 respetivamente.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Muito Mau) **e** (classificação final do conceito A4 = Muito Mau) **e** (classificação final do conceito B3 = Muito Bom)

Então avaliação final = Mau;

Esta regra indicou que os alunos obtiveram classificações de "Mau" na avaliação final e apresentaram classificações de "Muito Mau", "Muito Mau", "Muito Mau" e "Muito Bom" nos conceitos B5, B2, A4 e B3 respetivamente.

- **Se** (classificação final do conceito B5 = Muito Mau) **e** (classificação final do conceito B2 = Médio)

Então avaliação final = Mau;

Esta regra demonstrou que os alunos obtiveram classificações de "Mau" na avaliação final e apresentaram classificações de "Muito Mau" e "Médio" nos conceitos B5 e B2 respetivamente.

A análise das diversas regras permitiu identificar alguns conceitos que exerceram maior influência na avaliação final dos alunos: conceito B5 e B2.

5.4 Avaliação dos modelos

A avaliação dos modelos pretende avaliar o desempenho do modelo de dados criado. Nesta etapa é realizada a revisão de todos os passos efetuados na construção dos modelos de dados, tendo em conta os objetivos de negócio definidos inicialmente. Apesar de ser necessário verificar se alguma questão inerente ao negócio deva ser mais detalhada ou analisada, no fim deste passo os objetivos definidos para o processo de *Data Mining* devem ser atingidos (Rodrigues, 2008).

O desempenho de um modelo de dados depende de diversos fatores, tais como, o algoritmo usado na sua construção, a distribuição das classes, o tamanho dos conjuntos de dados e custos de más classificações. Devido à diversidade de fatores que condicionam o desempenho de um modelo surgem diferentes mecanismos de avaliação de modelos, métodos de amostragem, métricas e métodos para comparar modelos diferentes (Rodrigues, 2008).

Os métodos de amostragem consistem em mecanismos que permitem obter estimativas fiáveis e estão relacionados com dois conceitos distintos: exatidão e precisão. O primeiro conceito indica a proximidade de um valor verdadeiro e o segundo conceito consiste na capacidade de repetir uma medida. A exatidão dos modelos de dados está fortemente associada ao volume de dados usado na construção de um modelo. Os métodos de amostragem são distribuídos por dois grupos, alguns métodos são vocacionados para grandes conjuntos de dados e outros para pequenos conjuntos de dados (Rodrigues, 2008).

O modelo de dados gerado anteriormente, no ponto 5.3, era constituído por um pequeno conjunto de dados e, por este motivo, surgiu o sobre ajustamento de dados. Para combater este problema foi utilizado o método de validação cruzada, método de amostragem específico para pequenos conjuntos de dados.

Existem algumas métricas que permitem avaliar o desempenho de um modelo de dados, tais como (Rodrigues, 2008):

- *Accuracy*: representa a taxa de sucesso de uma classificação, razão entre a soma dos sucessos das classes a prever e o número total de instâncias ;

- Gráfico *lift*: representa a taxa de sucesso de uma classificação, razão entre a proporção de sucessos positivos no conjunto de classificações positivas e proporção de sucessos positivos no conjunto de dados global. Esta técnica é geralmente utilizada no *marketing*;
- Matriz confusão: técnica que avalia o resultado de uma classificação mapeando os valores previstos por um modelo com os valores desejados. Também designada matriz de erros. Esta técnica é representada na tabela 36.

Tabela 36 - Matriz confusão

Classe Objetivo	Classe Prevista		
		Classe +	Classe -
	Classe +	TP	FN
Classe -	FP	TN	

Legenda:

- TP - Positivos Verdadeiros (número de casos positivos corretamente classificados);
- FP - Falsos Positivos (número de casos negativos incorretamente classificados como positivos);
- TN - Negativos Verdadeiros (número de casos negativos corretamente classificados);
- FN - Falsos Negativos (número de casos positivos incorretamente classificados como negativos).

Neste estudo foi utilizada a técnica matriz confusão que representa o estudo pormenorizado da classificação dos registos efetuada pelo modelo de dados criado no ponto 5.3.

Tabela 37 - Matriz confusão do modelo criado

Cat_classif_PDF	Bom	Mau	Muito Bom	Muito Mau	Médio
Bom	14	0	1	0	0
Mau	1	15	0	0	0
Muito Bom	0	0	4	0	0
Muito Mau	0	2	0	14	0
Médio	0	0	0	0	18

A matriz confusão representada na tabela 37 permite verificar que a categoria "Muito Bom" exibiu os piores resultados e a categoria "Médio" apresentava os melhores resultados.

Existem outros métodos que permitem avaliar a comparação de desempenho de diferentes modelos de dados, o método das curvas ROC (*Receiver Operating Characteristic*) que representa graficamente a taxa dos verdadeiros positivos versus a taxa dos falsos positivos (Rodrigues, 2008).

No caso em estudo apenas se utilizou um único modelo de dados, sendo assim não foi utilizado este método, curvas ROC, para avaliar a comparação de desempenho de modelos de dados.

5.5 Resumo

A fase de modelação pretende elaborar um modelo de dados que efetue previsões sobre o progresso matemático dos alunos, descobrindo regras e padrões, que favoreçam a melhoria do processo de aprendizagem dos alunos.

A modelação teve início com a preparação dos dados. Os dados iniciais foram submetidos a diversas operações de transformação, tendo como objetivo principal aumentar a qualidade dos dados. Após a análise dos atributos existentes foram identificados alguns atributos irrelevantes ao estudo em questão, nomeadamente os atributos relativos à identificação do aluno. Foi constatada a existência de diversos atributos numéricos e foi considerado interessante discretizar alguns atributos, atribuir categorias a determinadas gamas de valores. Desta análise também se verificou a presença de alguns registos de dados com valores nulos e na tentativa de minimizar problemas na construção de um modelo de dados, foi efetuada remoção de todos os registos que apresentavam dados nulos.

Após a preparação dos dados procedeu-se à exploração gráfica dos mesmos que permitiu a extração de algumas conclusões relativas ao processo de aprendizagem dos alunos:

- Os alunos do grupo experimental exibiram melhores resultados de aprendizagem do que os alunos do grupo de controlo;

- Na fase inicial de avaliação, 88% dos alunos obtiveram resultados negativos e no final do processo de aprendizagem apenas 46% dos alunos mantiveram os resultados negativos;
- Este processo de aprendizagem proporcionou a diminuição do estilo de aprendizagem prático e o aumento do estilo de aprendizagem visual dos alunos;
- O conceito mais simples de ser adquirido pelos alunos foi o A3 (leitura de uma razão);
- Os conceitos mais difíceis de serem assimilados pelos alunos foram B1 (meios e extremos) e B5 (método de redução à unidade);
- Os alunos com estilo de aprendizagem visual apresentavam melhores resultados (o estilo de aprendizagem condicionou a aprendizagem);
- A maioria dos alunos que alteraram o estilo de aprendizagem durante o processo de aprendizagem aumentaram os resultados obtidos, com a exceção de 2 alunos;
- O tipo de ensino foi considerado um fator condicionante na aprendizagem, o ensino especial não inviabilizou o processo de aprendizagem mas o aumento dos resultados obtidos não foi tão expressivo como no ensino normal;

Na construção do modelo de dados foram utilizados três algoritmos de classificação: C&RTree, C5.0 e redes neuronais, contudo não apresentaram os resultados desejados. Como o conjunto de dados iniciais era relativamente pequeno, os modelos de dados criados obtiveram taxas de confiança muito elevadas, 100%, obtendo sobre ajustamento de dados. Para contornar este problema foi necessário utilizar o método de validação cruzada, sendo um método específico para pequenos conjuntos de dados. Através da validação cruzada, o modelo de dados criado com o algoritmo C5.0 alcançou uma taxa de confiança de 94%, sendo um valor aceitável.

Este modelo de classificação de dados permitiu a extração de algumas regras e padrões, que futuramente serão implementadas na plataforma PCMAT, tendo como intuito a melhoria do processo de aprendizagem. Por exemplo, através das regras extraídas foi verificado que o conceito B5 é o mais difícil de ser aprendido pelos alunos. Para contornar esta situação poder-se-á apresentar mais conteúdos e/ou alterar os conteúdos relativos a este conceito no PCMAT.

Antes de implementar o modelo de dados, foi realizada a avaliação do mesmo segundo métrica matriz confusão, indicando os melhores e piores resultados da classificação.

6 Conclusões

"O pesquisador que não sabe o que está a procurar não compreenderá o que encontrar."

Claude Bernard

6.1 Resultados obtidos

O estudo em questão teve como objetivo principal selecionar e aplicar técnicas de *Data Mining* a um modelo de alunos para pré-definição de perfil de alunos e regras de adaptação e pedagógicas a aplicar por cada perfil, melhorando a aprendizagem dos alunos.

Inicialmente foram analisadas várias técnicas de DM que poderiam ser aplicadas neste trabalho: classificação, *clustering*, regressão, associação de regras, entre outras. A aplicação destas técnicas permite descobrir padrões de comportamento que estejam ligados ao processo de aprendizagem do aluno. Esses padrões podem ser usados como parâmetros para outros alunos, cabendo ao professor a decisão sobre quais aspetos aplicar as técnicas de DM, acrescentando maior flexibilidade ao acompanhamento do aluno. Cada técnica de DM pode extrair um padrão distinto relacionado com a modelação de alunos, regras de adaptação e pedagógicas e comportamentos de alunos.

Neste trabalho também foi necessário selecionar a ferramenta de *Data Mining* mais adequada a este estudo, esta escolha foi baseada numa análise comparativa entre as ferramentas mais usadas no mercado, focando as vantagens e desvantagens de cada uma, mencionado no capítulo 4. A ferramenta selecionada foi o *Clementine*, ferramenta que detém todas as técnicas de DM essenciais a este trabalho, opera com grande volume de dados e apresenta uma interface gráfica simples e fácil de usar.

Após a seleção das técnicas de DM a usar foram criados modelos de dados recorrendo a distintos algoritmos de classificação: C&Rtree, C5.0 e redes neuronais. A maioria dos modelos de dados apresentaram taxas de confiança muito elevadas, valores de 100%, chegando a ocorrer situações de sobre ajustamento dos dados. Para ultrapassar este obstáculo foi utilizado um método de amostragem designado validação cruzada, específico para pequenas amostras de dados, impedindo que o modelo memorize os dados da reduzida amostra e

efetue previsões erradas. Com a utilização deste método foi possível alcançar um valor aceitável de taxa de confiança.

O modelo de classificação criado permitiu a extração de algumas regras, por exemplo:

- A classificação de "Muito Bom" foi obtida na avaliação final por alunos que exibiram classificações finais de "Muito Bom" no conceito B5. Os alunos que alcançaram o conhecimento deste conceito obtiveram muito bons resultados na avaliação final, podendo concluir que o B5 (método de redução à unidade) foi o conceito mais complexo de ser adquirido pelos alunos;
- Os alunos do grupo experimental apresentaram classificações de "Médio" na avaliação final e obtiveram classificações de "Muito Mau" e "Bom" nos conceitos B5 e B2 respetivamente. Os alunos do grupo de controlo obtiveram classificações de "Mau" na avaliação final e apresentaram classificações de "Muito Mau" e "Bom" nos conceitos B5 e B2 respetivamente. Esta situação permitiu concluir que os alunos do grupo experimental conseguiram obter melhores resultados na avaliação final do que os alunos do grupo de controlo, demonstrando que a aprendizagem dos alunos é favorecida pelo uso do PCMAT comparando com a aprendizagem tradicional.

As regras extraídas do modelo podem ser aplicadas em prol da criação de novos perfis de alunos e regras de adaptação e pedagógicas a aplicar por cada perfil.

Inicialmente estava previsto a utilização de outras técnicas de DM, nomeadamente, *clustering* e regras de associação. Contudo não se justificou o uso destas técnicas, tendo em conta o tamanho da amostra de dados e os objetivos propostos, facilmente se extraiu algumas regras e padrões através de uma simples análise estatística.

6.2 Trabalho futuro

Esta secção pretende apresentar um conjunto de direções de trabalho futuro que permitam melhorar os resultados obtidos neste estudo.

Uma possibilidade de trabalho futuro seria aumentar a dimensão da amostra de dados, resolvendo os problemas de sobre ajustamento na criação de modelos como o ocorrido durante o estudo. Também seria interessante que a amostra fosse mais representativa da realidade dos alunos, apresentando registos de alunos que frequentam o 7º e 8º ano de escolaridade, não estando exclusiva a alunos do 7º ano de escolaridade como acabou por

acontecer com a amostra utilizada. Outro aspeto que poderia ser melhorado na amostra é a qualidade dos dados, idealmente a amostra não devia apresentar dados em falta, pois esta situação reduz a quantidade de dados e perturba a construção dos modelos. Um maior volume de dados também permitiria a aplicação de outras técnicas de *Data Mining* que acabaram por não ser utilizadas na amostra devido ao pequeno tamanho da mesma, algumas regras ou padrões foram extraídos facilmente através de uma análise estatística. Os resultados do modelo de classificação também podem ser melhorados com uma amostra de maior dimensão, podendo ser testados distintos algoritmos de classificação e, conseqüentemente, melhorar o tratamento das regras obtidas.

Uma amostra de dados ideal apresentaria os seguintes pontos: maior representatividade, maior quantidade de dados e melhor qualidade dos mesmos. Uma forma possível de obter esta amostra seria reunir cerca de 3 ou mais escolas diferentes, apresentando realidades distintas. Em cada escola selecionar 3 ou mais turmas de 7º e 8º ano de escolaridade, em que cada turma apresente uma distribuição uniforme dos alunos, semelhante à existente na amostra utilizada neste estudo. Outro aspeto importante é garantir que os alunos das diferentes turmas, 7º e 8º ano de escolaridade, efetuem todos os passos inerentes ao processo de aprendizagem:

- questionário VARK para aferir o estilo de aprendizagem inicial dos alunos;
- avaliação de diagnóstico dos alunos;
- aprendizagem dos conteúdos matemáticos durante determinado período de tempo;
- questionário VARK para identificar o estilo de aprendizagem final dos alunos;
- avaliação final.

Outro ponto útil seria assegurar a reduzida existência de dados em falta pois este tipo de situação perturba a criação dos modelos de dados e, conseqüentemente, dificulta a extração de regras e padrões.

Em suma, a melhoria da qualidade dos dados e o aumento da amostra são uma alternativa para melhorar os resultados deste trabalho.

7 Bibliografia

- Alves et al., 2003 Alves, A., Moura, H. (2003) "Sistemas de hipermídia adaptativa." In: Revista do CCEI - Centro de Ciências da Economia e Informática, volume 7, número 12. Bagé: Ediurcamp, 2003.
- Amorim, 2006 Amorim, T. (2006). "Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de base de dados", Graduação em Ciência da Computação, Centro de Informática, Brasil. 2006.
- Anacleto, 2009 Anacleto, A. (2009). "Aplicação de Técnicas de Data Mining em Extração de Elementos e Documentos Comerciais". Tese de Mestrado em Análise e Sistemas de Apoio à Decisão. Faculdade de Economia do Porto. Porto. Portugal. 2009.
- Anderson, 1993 Anderson, J. (1993). "Rules of the mind". New Jersey: Lawrence Erlbaum Associates. 1993.
- Arbex, 2010 Arbex, M. (2010). "Clusterização de grupos contemporâneos com tamanho reduzido para as avaliações genéticas de rebanhos leiteiros". Universidade Federal do Rio de Janeiro. 2010
- Bastos, 2003 Bastos, P. (2003). "Inferência de propriedades químicas do algodão através de técnicas Data Mining", Universidade do Minho, Portugal. 2003.
- Beck, 1998 Beck, J., Stern, M., Haugsjaa, E. (1998). "Applications of AI in Education". The ACM's First Electronic Publication. 1998.
- Brusilovsky, 1996 Brusilovsky, P. (1996). "Methods and techniques of adaptative hypermedia". User Modeling and User-Adapted Interaction. Special issue on adaptive hypertext and hypermedia, Dordrecht, v.6, n.2-3. 1996.
- Brusilovsky, 2004 Brusilovsky, P. (2004). "Adaptative Navigation Support: From Adaptative Hypermedia to the Adaptive Web and Beyond PsychNology Journal, v. 2, n. 1, 7 -23." 2004.
- Brusilovsky, 2005 [Brusilovsky, P. \(2005\). On-line. Disponível na internet em: http://www2.sis.pitt.edu/~peterb/](http://www2.sis.pitt.edu/~peterb/) Último acesso em 23 de abril de 2013.
- Bugay, 1999 Bugay, E. (1999). "Modelagem em hipermídia de um tutorial para criação de maquetes eletrônicas". Dissertação - Universidade Federal de Santa Catarina, Florianópolis, 1999.
- Bugay, 2006 Bugay, E. (2006). "O MODELOAHAM – MI: Modelo de Hipermídia Adaptativa utilizando Inteligências Múltiplas". Tese de Doutorado. Santa Catarina – Brasil. 2006.
- Capitão, 2010 Capitão, M. (2010). "Previsão de vento baseado em Técnicas de Data Mining". Tese de Mestrado do Instituto Superior de Engenharia do Porto. Porto. 2010.
- Chaves, 2006 Chaves, E. (2006). "Ensino a Distância: Conceitos Básicos". 2006.
- Clementine, 1998 Clementine (1994-1998), Clementine Data Mining System: Reference Manual, 1994-1998
- Coelho, 2005 Coelho, P. (2005). "Um sistema para indução de modelos de predição baseados em árvores." Tese de Doutorado de Engenharia Civil na Universidade Federal do Rio de Janeiro. Brasil. 2005.
- Cruz, 2007 Cruz, A. (2007). "Data Mining via Redes Neurais Artificiais e Máquinas de Vectores de Suporte". Dissertação de Mestrado de Sistemas de Informação da Universidade do Minho. Portugal. 2007.
- Cunha, 2009 Cunha, N. (2009) "Metodologia de Seleção de Segmentações Diversificadas: Um caso de aplicação de técnicas de Data Mining em dados de consumo para Avaliação de Portfólios de Cartões Bancários". Tese de Mestrado de Gestão Comercial na Faculdade de Economia da Universidade do Porto. Porto. 2009
- Damasceno, 2010 Damasceno, M. (2010). "Introdução à Mineração de Dados usando o Weka".

- Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte/Campus Macau. 2010
- Damico, 1995 Damico, C. (1995). "Modelo de usuário para sistemas tutores inteligentes". Porto Alegre – Brasil. 1995.
- Dara-Abrams, 2002 Dara-Abrams, B. (2002). "Applying multi-intelligent adaptive hypermedia to online learning". Tese de Doutorado. 2002.
- De Bra, 1998 De Bra, P. (1998) "Adaptative hypermedia on the Web: methods, technology and applications". 1998.
- Elder et al., 1998 Elder, J, Abbott, D. (1998). "A Comparison of Leading DM Tools". Fourth International Conference on Knowledge Discovery & Data Mining. Estados Unidos da América. 1998.
- Euriditionhome, 2004 [Euriditionhome, \(2004\). "Data Mining Tutorials, Resources". Disponível em: http://datamining.eruditionhome.com](http://datamining.eruditionhome.com) Último acesso em 27 de abril de 2013.
- Fayyad et al., 1996 Fayyad, U., Piatetski-Shapiro, G., Smyth, P. (1996). "The KDD Process for Extracting Useful Knowledge from Volumes of Data." Communications of the ACM. 1996.
- Fred, 2004 Fred, A. (2004). "Redes Bayesianas". 2004.
- Giraffa, 1995 Giraffa, L. (1995). "Fundamentos de teorias de ensino-aprendizagem e sua aplicação em sistemas tutores inteligentes". Porto Alegre – Brasil.1993.
- Goebel et al., 1999 Goebel, M., Gruenwald, L. (1999). "A survey of data mining and knowledge discovery software tools". SIGKDD Explorations. 1999.
- Gonçalves, 2010 Gonçalves, E. (2010). "Mineração de Regras de Associação com a Ferramenta de Data Mining Weka". Brasil. 2010. Disponível em: <http://www.devmedia.com.br/post-20478-Mineracao-de-Regras-de-Associacao-com-a-Ferramenta-de-Data-Mining-Weka.html>
- Gonçalves, 2011 Gonçalves, E. (2011). "Data Mining com a ferramenta Weka". Rio de Janeiro. Brasil. 2011
- Groth, 2000 Groth, R. (2000) "Data Mining: Building Competitive Advantage" Prentice Hall PTR, USA. 2000
- Han et al., 2001 Han, J., Kamber, M. (2001). "Data Mining – Concepts and Techniques". 2001.
- Held et al., 1997 Held, G.; Neville, P. (1997). "Data Mining with the SAS System". SAS Institute. 1997
- Jameson, 2003 Jameson, A. (2003). "User-adaptive systems: An integrative overview". Department of Computer Science Saarland University. Alemanha. 2003.
- Kleinschmidt, 2007 Kleinschmidt, M. (2007). "Mineração de dados para Avaliação do Perfil de Usuários do Sistema de Informação da Academia da Univali." Trabalho de Conclusão de Curso de Ciência da Computação. Brasil. 2007.
- Koch et al., 2004 [Koch N., Rossi, G., \(2004\). "Patterns for adaptive Web applications." On-line. Disponível na internet em: http://www.pst.informatik.uni-muenchen.de](http://www.pst.informatik.uni-muenchen.de) Último acesso em 20 de abril de 2013.
- Koch, 2000 Koch, N. P. (2000) "Software engineering for adaptive hypermedia systems reference model, modeling techniques and development process". Tese de doutorado em engenharia de software – Ludwig-Maximilians – Universität München. Alemanha. 2000.
- Martin, 1992 , J. (1992). "Hiper documentos e como criá-los." Rio de Janeiro: Editora Campus, 1992.
- Martins et al., 2008 Martins, C.; Faria, L.; Fernandes, M.; Couto, P.; Bastos, C.; Carrapatoso, E. (2008). "PCMAT - Mathematics Collaborative Educational System" - In Smart Innovation, Systems and Technologies, Special book on: Intelligent and Adaptive Educational Learning Systems: Achievements and Trends, ed. Alejandro Peña-Ayala, 183 - 212. ISBN: 978-3-642-30170-4. Springer Berlin Heidelberg: Springer. 2008.
- Neves, 2002 Neves, J. (2002). "Ambiente de pós-processamento para regras de associação". Dissertação para obtenção do grau de Mestre em Análise de Dados e Sistemas de Apoio à Decisão. Porto. 2002.
- Neves, 2007 Neves, D. (2007). "Detecção de fraude no sistema de custas judiciais – Usando

- métodos de aprendizagem não supervisionada". Dissertação de Mestrado. 2007.
- Nisbet, 2004 Nisbet, R. (2004). "How to choose a Data Mining Suite". Information Management Special Reports. 2004
- Oliveira et al., 2003 Oliveira, F et al. (2003). "Modelo de interfaces adaptativas utilizando redes bayesianas". Pós-graduação em Ciências da Computação na Universidade Federal de Santa Catarina. Brasil. 2003.
- Oliveira et al., 2004 Oliveira, J., Fernandes, C. (2004) "Sistemas Hipermedia Adaptativos Educacionais: breve panorama e modelo de referência". Congresso Nacional de Ambientes Hipermedia para Aprendizagem. Brasil. 2004.
- Oliveira, 2005 Oliveira, A. (2005). "Aplicação de Algumas Técnicas de Data Mining em Bancos de Dados utilizando o Weka". Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte/ Campus Macau. Brasil. 2005.
- Oliveira et al., 2006 Oliveira, J., Junior, A. (2006). "Data Mining aplicado ao Ensino à Distância". Sistemas de Informação da Faculdade de Ciências Administrativas do Centro Universit Brasil. 2006.
- Oracle Corporation, 2008 Oracle Corporation. 2008. "Oracle Data Mining Concepts Release 11g Release". 2008.
- Oracle Corporation, 2010 Oracle Corporation. 2010. "Oracle Data Mining Concepts Release 1". 2010.
- Palazzo, 2000 Palazzo, L. A. M. (2000). "Modelos proativos para hipermedia adaptativa". Tese de doutoramento – Universidade Federal do Rio Grande do Sul. Porto Alegre - Brasil. 2000.
- Palazzo, 2004 Palazzo, L. A. M. (2004). "Sistemas de hipermedia adaptativa". Brasil. 2004.
- Pearl, 2004 Pearl, J. (2004). "Bayesian network". 2004.
- Pelegrin et al., 2005 [Pelegrin, D., Casagrande, D., Mattos, M., Simões, P., Charnovski, R., Bettiol, J. \(2005\) "As tarefas de Associação e de classificação na Shell de Data Mining Orion." 2005. Disponível em: http://www.dcc.unesc.net/sulcomp/artigos/sessaoOral/22103.pdf](http://www.dcc.unesc.net/sulcomp/artigos/sessaoOral/22103.pdf) Último acesso em 17 de abril de 2013.
- Pereira, 2002 Pereira, C. (2002). "Comparação de ferramentas de Data Mining". Projecto de Licenciatura do Instituto Superior Engenharia do Porto. Porto. 2002.
- Pinto et al., 2005 Pinto, C., Vieira, H., Bulhões, L. (2005) "Data Mining em R". Portugal. 2005.
- Pozo, 2010 Pozo, A. (2010). "Mineração de dados: conceitos, aplicações e experimentos com WEKA". Apontamentos da disciplina Mineração de Dados. Brasil. 2010.
- Rada, 1995 Rada, R. (1995). "Interactive media". New York USA. 1995.
- Reis, 2006 Reis, M. (2006) "Um módulo de identificação On-line do modelo de aluno baseado em estilos cognitivos de aprendizagem". Dissertação de Pós-graduação. Brasil. 2006.
- Reis, 2009 Reis, M. (2009). "Business Intelligence/ Datamining". Brasil. 2009.
- Rezende et al., 1998 Rezende, S., Paula, M., Figueiredo, L. (1998). "MineSet, Ferramenta Data Mining". Relatórios Técnicos do ICMC versão 1.0. São Carlos, Brasil. 1998.
- Rich, 1979 Rich, E. (1979). "User modeling via stereotypes". San Francisco: Morgan Kaufmann Publishers Inc. 1979.
- Rich, 1988 Rich, E. (1988). "Inteligência Artificial". São Paulo: McGraw-Hill. Brasil. 1988.
- Rodrigues, 2008 Rodrigues, F. (2008) "Descoberta do conhecimento – Data Mining". Apontamentos da disciplina Descoberta do Conhecimento – ISEP. Porto. 2008.
- Rossatelli et al., 2004 Rossatelli, M., Azevedo, P. (2004). "Diagnosticando o utilizador para criação de sistemas personalizáveis". Brasil. 2004.
- Santos, 2000 Santos, A. (2000). "Ensino à distância & Tecnologias de informação – E-learning". Editora Lidel. 2000.
- Santos, 2008 Santos, R. (2008). "Aplicações e Algoritmos de Data Mining". Brasil. 2008.
- Semantic Research, 2004 Semantic Research. (2004) "Anatomy of a semantic network." 2004.
- Sferra et al., 2003 Sferra, H., Corrêa, A. (2003). "Conceitos e aplicações de Data Mining". Revista de ciência e tecnologia. Brasil. 2003.
- Silva et al., 1998 Silva, D., Durm, R., Duval, E., Olivie, H. (1998) "Concepts and documents for

- adaptive educational hypermedia: a model and a prototype". 1998.
- Silva et al., 2001 Silva, D., Vieira, M., Seno, W. (2001) "Acompanhamento do aprendizado em Educação a distância com o uso de Data Mining". Brasil. 2001.
- Silva et al., 2003 Silva, C., Rodrigues, C., Monteiro, M. (2003) "Uso de Regras de Associação para Descoberta de Conhecimento na Produtividade de Açai no Estado do Amapá". Faculdade de Balsas e Seama. Brasil. 2003
- Silva, 2007 Silva, T. (2007). "Conceitos, técnicas, ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados" - Universidade Federal de Pernambuco - Graduação em Ciência da Computação. 2007.
- Sousa, 1998 [Sousa, M. \(1998\). "Mineração de dados: Uma implementação fortemente acoplada a um sistema gerenciador de banco de dados paralelo". 1998. Disponível em: http://www.cos.ufrj.br/~marta/papers/TeseMauroS.pdf](http://www.cos.ufrj.br/~marta/papers/TeseMauroS.pdf) Último acesso em 20 de abril de 2013.
- Sowa, 2004 Sowa, J., (2004). "Semantic networks". 2004. Disponível em: <http://www.jfsowa.com/pubs/semnet.htm> Último acesso em 25 de abril de 2013.
- The CRISP-DM Consortium, 2000 Cross Industry Standard Process for Data Mining. 2000. Disponível em <http://www.crisp-dm.org>. Último acesso em 2 de abril de 2013.
- Ulbricht et al., 2007 Ulbricht, V., Batista, C. (2007). "Informação Personalizada no Website "Diferente todo o mundo é!". InfoDesign - Revista Brasileira de Design da Informação. Brasil. 2007
- Vassileva, 1996 Vassileva, J. (1996) "A task-centered approach for user modeling in a hypermedia office documentation system: user modeling an user-adapted interaction". Special issue on adaptive hypertext and hypermedia. 1996.
- Vidal, 2002 Vidal, E. (2002). "Ensino à distância VS ensino tradicional" – Universidade Fernando Pessoa. Porto. 2002.
- Vivek et all, 2003 Vivek, S., De Roure, D. (2003). "Implementing hiperligação services via semantic Web services composition". 2003.
- Woods et al., 1995 Woods, P., Warren, J. (1995). "Adapting teaching strategies in intelligent tutoring systems". Australia. 1995.
- Wu et al., 2001 Wu, H. Kort, E., De Bra, P. (2001) "Design issues for general-purpose adaptive hypermedia systems." Aarhus, Dinamarca: Proceedings of the 12th ACM Conference on Hypertext and Hypermedia, Agosto 2001.
- Wu et al., 2004 wu, H., Houben, G., and De Bra, P. (2004). "Supporting user adaptation in adaptive hypermedia applications". Department of Computing Science Eindhoven University of Technology. Netherlands. 2004.
- Wu, 2002 Wu, H. (2002) "A reference architecture for adaptive hypermedia applications." Tese de Doutorado – Technische Universiteit Eindhoven. Netherlands. 2002.

Anexos

Anexo A

As Redes Bayesianas oferecem uma estrutura intuitiva para representar o raciocínio incerto que mistura a teoria dos grafos e probabilidades. A vantagem da sua utilização é permitir a representação e manipulação da incerteza com base em princípios matemáticos fundamentados que refletem nos valores de probabilidade a crença do especialista sobre o que espera que ocorra em determinadas situações (Oliveira et al., 2003).

O modelo Bayesiano interpreta a probabilidade condicional, onde o grau de crença de um agente causa um efeito noutra, portanto $P(A|B)$ reflete a probabilidade de A se B acontecer. A equação da regra de Bayes é dada por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A|B)P(A)}{P(B)}$$

Uma Rede Bayesiana é composta por uma parte qualitativa e outra quantitativa, ambas responsáveis pela representação do conhecimento do especialista. A parte qualitativa é um modelo gráfico que representa um “Grafo Acíclico Direcionado”, de forma a que os nós representam as variáveis de um domínio e as ligações entre nós definem o relacionamento entre eles, sendo que (Pearl, 2004):

- Estas variáveis podem pertencer ao domínio do conhecimento, conhecimento do utilizador e/ou ao modelo cognitivo;
- Cada nó representa as crenças do sistema sobre os valores possíveis (nível, estado) da variável.

Desta maneira, a distribuição probabilística condicional deve ser específica para cada nó, representando probabilidades condicionais e quantificando os efeitos que os nós “pais” exercem sobre um nó “filho” (probabilidade de o nó estar num estado específico dado os estados dos seus “pais”). A topologia da rede pode ser vista como uma base de conhecimento abstrata representando a estrutura dos processos causais do domínio (Fred, 2004).

A estrutura de uma Rede Bayesiana é ilustrada no seguinte exemplo: A casa de Carlos possui um sistema de alarme instalado para alertar possíveis assaltos e pequenos tremores de terra, tendo em conta que pequenos abalos sísmicos são responsáveis por falsos alarmes. Os seus

dois vizinhos, Maria e João, comprometeram-se a telefonar para o emprego do Carlos quando ouvissem o alarme a disparar. Apesar disso, Maria costuma ouvir música muito alta e, por vezes, não ouve o alarme. Enquanto o João telefona sempre que ouve o alarme mas, muitas vezes, ele confunde o som do alarme com o de outros sons. Dada a evidência de que alguém telefonou ou não para Carlos, é estimar a probabilidade de um assalto (Pearl, 2004).

Para construir uma rede bayesiana para o exemplo citado anteriormente, é iniciada a fase de ordenação dos nós, são adicionadas as causas (nós raiz da rede) e, em seguida, as variáveis que estas influenciam, sucessivamente até se atingir as folhas da rede (variáveis que não possuem uma relação causal com nenhuma outra). A escolha da representação dos sintomas em função das causas (causas influenciam sintomas – relações causais) conduz, em geral a redes mais compactas e de fácil definição em termos de probabilidades (Koch, 2000).

Estabelecida a topologia da rede, é necessário criar a tabela de probabilidades condicionais para cada nó, podendo ser visualizado na tabela 38 (Pearl, 2004).

Tabela 38 - Tabela de probabilidades condicionadas

Assalto	Terramoto
Verdadeiro	Verdadeiro
Verdadeiro	Falso
Falso	Verdadeiro
Falso	Falso

P (Alarme Ass., Terr.)	
Verdadeiro	Falso
0.950	0.050
0.950	0.050
0.290	0.710
0.001	0.999

Perante a observação de uma evidência ($P= 1$), é usado um mecanismo de propagação de probabilidades de forma a permitir efetuar a revisão dos valores de probabilidades dos nós influenciados (Pearl, 2004).

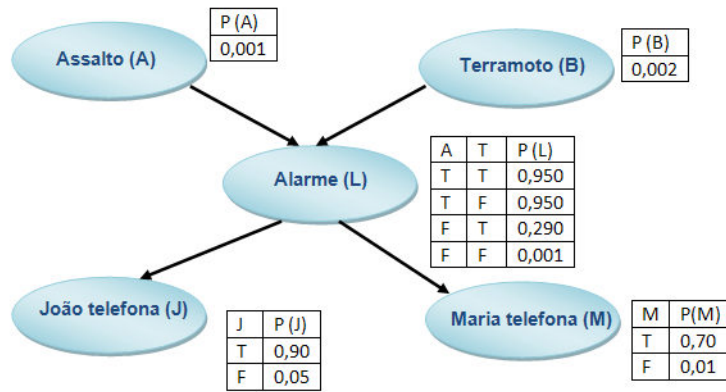


Figura 86 - Rede bayesiana

A pesquisa sobre redes bayesianas vem incentivar novos estudos sobre padrões de alunos, relações causa-efeito e previsões com base no conhecimento incompleto do mundo real.

Anexo B

A implementação computacional de redes semânticas foi, inicialmente desenvolvida na inteligência artificial e máquinas de tradução. Apesar disso, as versões mais recentes têm sido usadas nas áreas da filosofia, psicologia e linguística.

O elemento comum de todas as redes semânticas é uma representação gráfica que pode ser usada tanto para representar o conhecimento como para auxiliar sistemas automatizados de raciocínio sobre conhecimento. Algumas versões são bastante informais, enquanto outras são sistemas de lógica formalmente definidos (Pearl, 2004).

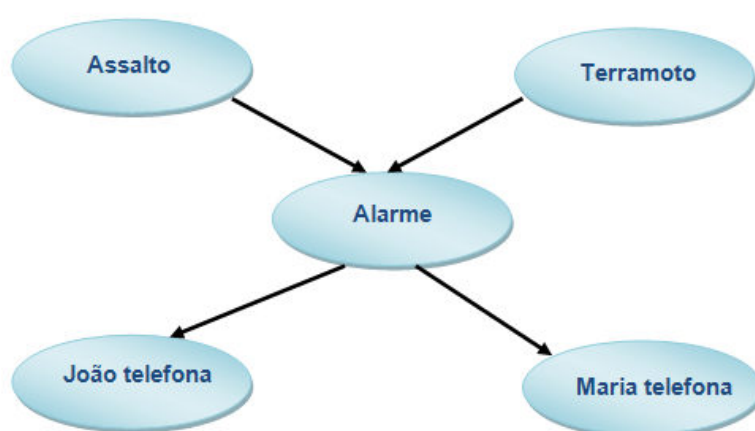


Figura 87 - Exemplo de uma rede semântica

São destacados seis tipos de redes semânticas (Sowa, 2004):

- **Redes de definição:** enfatizam a relação entre o tipo de conceito e um subtipo recém definido. A rede resultante, também denominada generalização ou subordinação hierárquica, auxilia a regra de herança para propriedades de cópia definida para um subtipo ou para todos os subtipos. Uma vez que as definições são verdadeiras, as informações destas redes são geralmente consideradas verdadeiras.
- **Redes de afirmação:** são projetadas para afirmar proposições. Diferente das redes de definição, a informação neste tipo de rede é, geralmente considerada verdadeira, a menos que isto esteja definido explicitamente com um operador modal. Algumas destas redes têm sido propostas como modelos para estruturas conceituais para linguagem natural semântica (Vivek et al., 2003).

- **Redes de implicação:** usam a implicação (IF – THEN) como relacionamento principal para conectar os nós. Podem ser usadas para representar padrões de crenças, casualidade ou inferência.
- **Redes executáveis:** incluem algum mecanismo, como um marcador de passagem ou procedimentos agregados, com os quais podem executar inferências, passar mensagens ou procurar por padrões ou associações.
- **Redes de aprendizagem:** constroem ou ampliam as suas representações pela aquisição de conhecimento de exemplos. O novo conhecimento pode mudar a rede antiga pela adição ou exclusão de nós e ligações ou pela modificação dos valores numéricos, denominados pesos, associados com os nós e ligações (Semantic Research, 2004).
- **Redes híbridas:** combinam duas ou mais técnicas anteriores, o que pode ocorrer tanto numa única rede como em redes separadas que interagem. Algumas das redes têm sido projetadas para implementar hipóteses sobre os mecanismos cognitivos humanos, enquanto outras são projetadas primariamente para a eficiência computacional (Sowa, 2004).

Uma rede semântica é constituída por três elementos básicos (Koch, 2000):

- **Conceitos:** são ideias ou pensamentos que tem significado.
- **Relações:** descrevem tipos específicos de ligações ou relacionamento entre dois conceitos.
- **Instâncias das relações:** consistem em dois conceitos ligados por um relacionamento específico.

A figura seguinte exhibe uma estrutura de uma rede semântica, constituída por muitos conceitos, relacionamentos e instâncias. A navegação numa rede semântica é realizada de modo hipertextual direto, muito parecida com um navegador Web (Semantic Research, 2004).

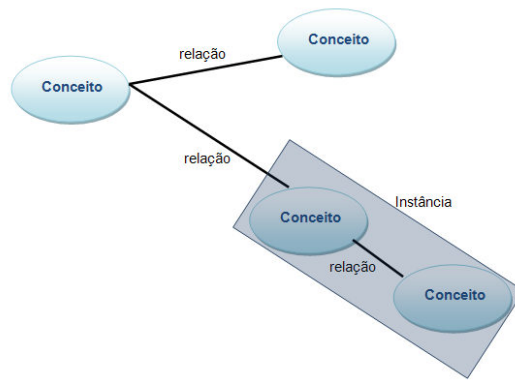


Figura 88 - Elementos de uma rede semântica

Anexo C

O *Clementine* é uma ferramenta de *Data Mining* que foi desenhada e desenvolvida com o objetivo de tirar proveito de um recurso subaproveitado, os dados. Esta ferramenta permite uma fácil navegação nos dados, utilizando gráficos para localizar relações importantes (Clementine, 1998). Esta ferramenta apresenta as suas funcionalidades integradas numa interface de programação gráfica (Rodrigues, 2008).

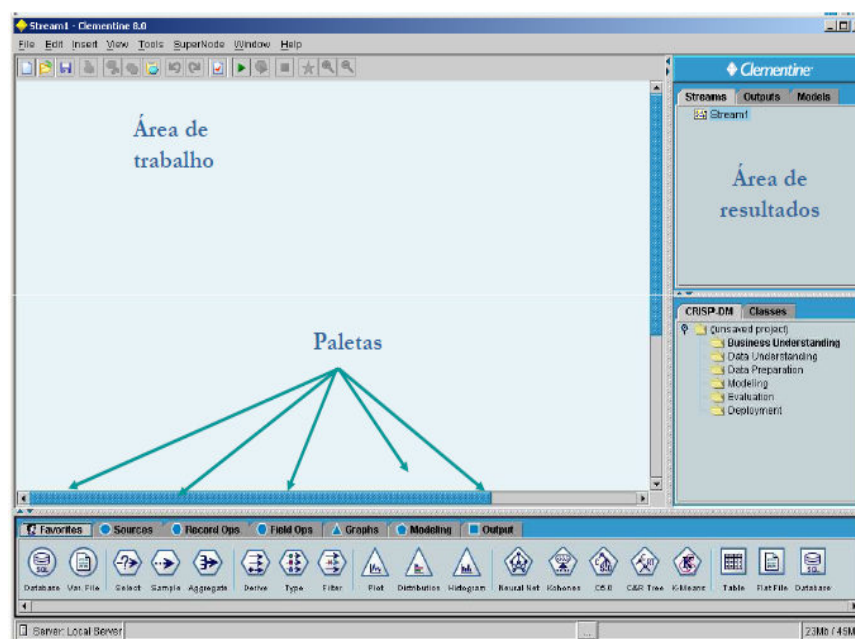


Figura 89 - Interface do *Clementine*

Na janela principal do *Clementine*, figura 11, são encontrados os seguintes componentes:

- Área de trabalho: consiste na área principal de trabalho do *Clementine*. A programação gráfica envolve a colocação e manipulação de ícones que representam nós de processamento. O conteúdo é designado de diagrama, estes podem ser carregados e gravados de e para uma unidade de armazenamento;
- Paletas: contêm famílias de ícones que representam diferentes operações possíveis no *Clementine*. Na programação gráfica, os ícones são selecionados das paletas, colocados na *stream*, conectados uns aos outros e editados;
- Área de resultados: guarda o resultado da modelação.

O *Clementine* oferece grandes potencialidades gráficas, possibilitando ao utilizador desenhar um diagrama que represente as diversas operações pelas quais os dados são submetidos: acesso aos dados, manipulação e análise dos mesmos, geração de modelos e resultados obtidos.

Cada tipo de operação é representado por ícones que, por sua vez, se encontram agrupados em paletas, estas são etiquetadas por categorias, de acordo com as funções desempenhadas pelos respetivos ícones (Bastos, 2003).

Cada nó pode ser editado, para definir parâmetros ou configurar o seu modo de execução, podendo ser usados em distintos contextos. Por exemplo, o nó tabela (Rodrigues, 2008):

- Exibe os dados em forma tabular;
- Examina os dados origem (dados a explorar);
- Mostra os resultados do processo DM.

Os nós podem ser agrupados nas diversas paletas, de acordo com as seguintes classes (Rodrigues, 2008):

- Nós origem – acesso aos dados (*input*);
- Nós de manipulação – tratamento dos dados;
- Nós de gráficos – visualização dos dados em diversos tipos de gráficos;
- Nós de modelação – criam modelos;
- Nós de saída – resultado dos dados (*output*).

Os nós de origem são circulares e efetuam a importação de dados a partir de ficheiros ou tabelas e só podem ter ligações de saída. Podem incluir nós para aceder a dados em ficheiros “ascii” em vários formatos, ficheiros fixos, bases de dados via *Open Database Connectivity* (ODBC) e permitem juntar dados de diferentes fontes de dados, bem como seleccionar ou fazer amostras de registos (Rodrigues, 2008).



Figura 90 - Nós Origem

Os nós de manipulação são hexagonais que realizam operações sobre registos e atributos, sendo divididos em:

- Nós de registo;
- Nós de atributos.

Os nós de registos realizam operações sobre registos e são divididos em oito tipos de nós (Rodrigues, 2008).



Figura 91 - Nós de Registos

Seguidamente, são descritos alguns dos mais relevantes:

- **Nó *Select*:** seleciona registos que obedecem a determinada condição booleana. Por exemplo, *Discard Age < 20*.
- **Nó *Sample*:** seleciona subconjuntos de registos, aleatoriamente ou não. Permite selecionar:
 - os *n* primeiros, ou
 - 1 em cada *n* registos, ou
 - ainda definir uma aleatoriedade, ou seja, cada registo é selecionado com uma propriedade *Random*.

Por exemplo, selecionar os 100 primeiros registos.
- **Nó *Merge*:** faz o *join* de registos de diferentes ficheiros. Por exemplo, *Medic1 (Idade, Sex, BP) Medic2 (Idade, Sex, Fe)*. Pode ser feito apenas por ordem, assim que um ficheiro termina não une mais registos, ou então por *join* de chaves.
- **Nó *Balance*:** corrige dados desequilibrados. Por exemplo, passar *BP = High* para 2,5 x.
- **Nó *Sort*:** ordenada registos por um ou mais atributos. Por exemplo, idade *sex ascending*.
- **Nó *Aggregate*:** substitui uma sequência de registos por cálculo de funções de soma, média, contagem, etc. Por exemplo, agregação pelos atributos sexo e droga:
 - Idade – Média
 - Max – Fe
 - Contar os registos agregados pela chave especificada.
- **Nó *Distinct*:** inclui ou retira registos que apresentem certos valores em atributos selecionados. Por exemplo, *Distinct Sex, Age (include)*.

Os nós de atributos realizam operações sobre atributos e são divididos em nove tipos de nós (Rodrigues, 2008).

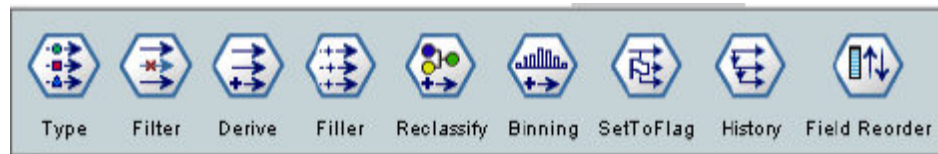


Figura 92 - Nós de Atributos

De seguida, são descritos alguns dos mais importantes:

- **Nó *Filter*:** elimina atributos, renomeá-los:
 - **Tipo:** especificam três propriedades importantes: tipo, direção e tratamento de valores anormais.
 - **Tipo:**
 - *integer* – intervalo de inteiros;
 - *real* – intervalo de reais;
 - *set* – conjunto de valores simbólicos;
 - *flag* – dois valores simbólicos (T/F, 0/1);
 - *typeless* – sem informação;
 - *auto* – tipo não instanciado, por defeito todos os atributos são auto.
 - **Direção:** relevante apenas para os nós de modelação definem se o atributo irá ser considerado na aprendizagem e como:
 - *In*;
 - *Out*;
 - *Both*;
 - *None*.
 - ***Check*:** tratamento de valores anormais.
 - *None* – opção por defeito, não é feito nenhum controle.
 - *Coerce* – os valores errados são corrigidos:
 - inteiros convertidos a reais e vice-versa;
 - um valor acima da gama de valores do atributo é corrigido com o valor máximo;
 - um valor abaixo da gama de valores de atributos é corrigido com o valor mínimo;

- um valor numérico não definido é preenchido pelo valor médio;
- um valor categórico é preenchido logo com o primeiro valor do conjunto de dados;
- útil para a limpeza de dados.
- *Discard* – os valores errados não são considerados;
- *Warm* – os valores errados são contados;
- *Abort* – se existirem valores errados é interrompida a execução da *stream*.

- **Nó *Derive*:** adiciona um novo atributo a todos os registos por ele processado.

Existem seis variantes deste nó:

- *Any* – o novo atributo resulta do cálculo de uma expressão. Por exemplo, $\text{abs}(\text{NA}-K)$;
- *Flag* – o novo atributo é um *flag* resultado de uma condição. Por exemplo, $\text{BP} == \text{Cholesterol}$;
- *Set* – o novo atributo torna valores de acordo com uma condição que pode ser válida para um conjunto de valores. Por exemplo, grupo etário $\text{Age} \geq 18 \rightarrow \text{Menor}$; $\text{Age} > 65 \rightarrow \text{Idoso}$; default $\rightarrow \text{Meia Idade}$.

Os dois nós seguintes só fazem sentido em dados sequenciais.

- *State* – semelhante à *Flag*, mas o estado que o novo atributo toma depende de duas condições independentes. Por exemplo, “On” value – Alto $\text{Na} > 0.70$; “Off” value – Baixo $\text{Na} < 0.40$.
- *Count* – conta os registos que obedecem a determinada condição, de cada vez que é encontrado um registo que obedece à condição de *reset* o contador volta ao estado inicial. Por exemplo, incrementa $\text{Age} > 25$, Reset when $\text{Age} > 50$. (Bastante útil para séries temporais).
- *Conditional* – são atribuídos valores ao novo atributo de acordo com o valor de uma expressão booleana. Por exemplo, atributo **Risco**: se $\text{BP} == \text{High and Cholesterol} == \text{High} \rightarrow \text{Elevado}$

Senão \rightarrow Médio

- **Nó *Filler*:** substitui valores de atributos.
 - **Nó *History*:** manipulação de dados sequenciais, permite adicionar, determinar valor a um novo atributo se se verificar uma sequência de valores.

Os nós de gráficos são triangulares e são usados para mostrar a informação dos atributos e suas relações. Alguns gráficos produzidos são interativos, ou seja, não mostram apenas os dados, podem ser também usados para definir novas propriedades de registo ou seleccionar subpopulações (Rodrigues, 2008);

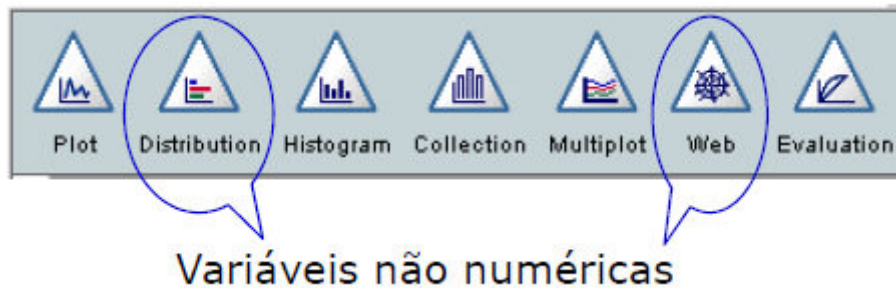


Figura 93 - Nós Gráficos

De seguida, é efetuada uma pequena descrição dos gráficos que se consideram mais importantes:

- Gráfico *Plot* - usado para relacionar valores de atributos numéricos. O campo *overlay* permite seleccionar um atributo categórico, existindo a possibilidade de seleccionar um gráfico de pontos ou linhas e comparar o gráfico com uma função conhecida – *overlay function*.
- Gráfico *Multiplot* - especialização do nó *Plot* que permite visualizar múltiplos atributos, variáveis, ao longo do eixo dos Y relativamente a uma só medida ao longo do eixo dos X. É conveniente que as diferentes medidas estejam normalizadas à mesma escala.
- Gráfico Histograma - utilizado para visualizar valores de atributos categóricos.
 - *Range*: permite especificar o intervalo de valores a considerar. Seleccionar um *overlay function* (simbólico), mostra um gráfico de barras. Permite especificar o número de barras e largura das barras.
- Gráfico Coleção – é semelhante ao nó Histograma, exceto que ao invés de mostrarem apenas as ocorrências de valores de um só atributo, mostram a distribuição dos valores de um atributo numérico relativamente aos valores de outro atributo.
- Gráfico *Web* – permite visualizar relações fortes e fracas entre vários atributos categóricos, sendo possível modificar a forma do gráfico, eliminar pontos, linhas e

altera os valores das linhas. Os limites (*thresholds*) indicam como interpretar os valores que definem as ligações entre os atributos:

- Absoluto: é o número de registos onde ocorre o par de valores;
- *Overall Percentages*: é a percentagem de ocorrência do par de valores em todo o conjunto de dados.
- Gráfico *Web Direto* – é uma variante do nó *Web* que só mostra ligações entre um atributo e outros atributos. Enquanto que o nó *Web* mostra ligações entre os outros atributos, selecionado o nó *Web* direto só mostra ligações entre os atributos *from* e os atributos *to*.
- Gráfico *Distribuição* – permite ver a distribuição de valores numéricos, sendo considerado bastante útil para detetar desequilíbrio nos dados.

Os nós de aprendizagem são algoritmos desenvolvidos com base em técnicas de inteligência artificial e estatística (Rodrigues, 2008).

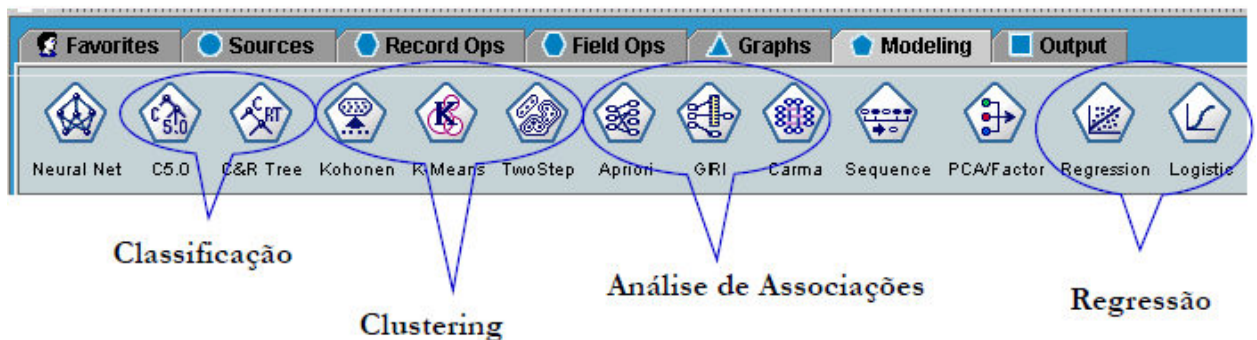


Figura 94 - Nós Aprendizagem

Os nós de saída permitem a visualização de dados e resultados das análises, e só podem ter ligações de entrada, uma só entrada (Rodrigues, 2008).



Figura 95 - Nós Saída

Uma das razões pela qual o *Clementine* é tão simples de aprender, consiste na clara definição das funções de cada nó. Este facto, no entanto, pode levar a que uma *stream* se torne muito complexa caso uma grande sequência de nós for necessária (Clementine, 1998).

Para solucionar este problema, a *stream* é dividida em várias. A primeira cria um ficheiro que será o de entrada para a segunda e, assim, sucessivamente. Esta solução é eficaz, mas implica uma constante limpeza e carregamento de *streams* (Bastos, 2003).

Em alternativa, podem ser utilizados super-nós, sendo representados por uma estrela. A sua função permite o agrupamento de fragmentos de *stream*, constituídos por vários nós, num único nó. O encapsulamento é a designação deste agrupamento (Bastos, 2003; Rodrigues, 2008).

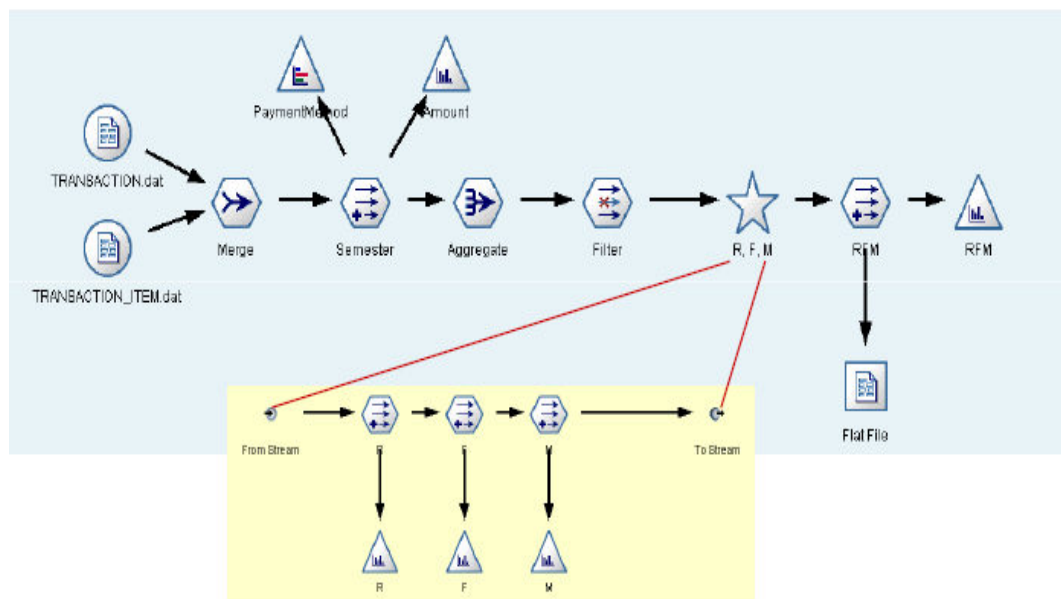


Figura 96 - Super-nó

A utilização de super-nós é vantajosa visto que mantém a *stream* clara e manuseável, permite uma fácil compreensão, mesmo para outros utilizadores, tendo em conta que os super-nós têm um baixo nível de detalhe e podem ser exportados para bibliotecas e reutilizados noutras *streams* (Bastos, 2003).

O *Clementine* também inclui uma modelação avançada ou técnicas de aprendizagem máquina que extraem inter-relacionamentos e regras de decisão a partir dos dados, permitindo automatizar aplicações, tais como: previsões, estimativas, classificações e fornecer apoio à decisão (Clementine, 1998). As técnicas de modelação mais utilizadas nesta ferramenta DM são descritas de seguida.

Redes neuronais

As redes neuronais consistem em modelos muito simples, que imitam o funcionamento do sistema nervoso humano (Clementine, 1998). A unidade básica é denominada de “neurónio” e

são, normalmente organizados em camadas (*layers*), sendo ilustrado na próxima figura (Rodrigues,2008):

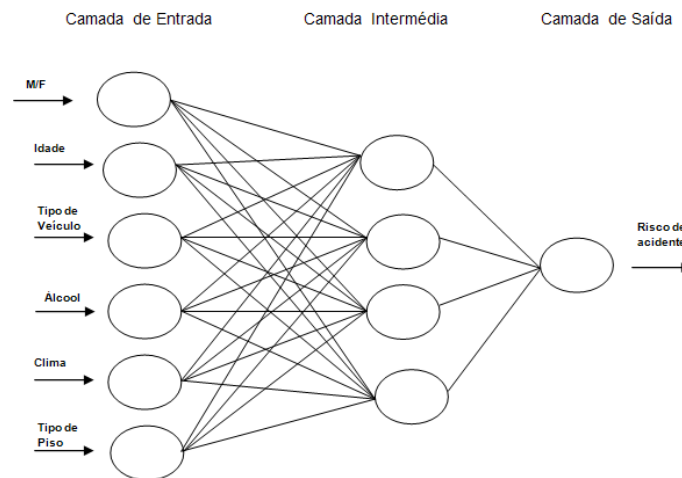


Figura 97 - Rede neuronal

Os dados de entrada são fornecidos à primeira camada e o valor de cada neurónio é propagado para todos os outros neurónios. Os valores dos pesos são alterados durante a transmissão e o resultado é devolvido pelo último neurónio. Inicialmente, todos os pesos são aleatórios e as respostas obtidas são provavelmente incorretas. A rede aprende treinando: resultados conhecidos são, constantemente apresentados à rede e comparados com o resultado desta. A informação da comparação é propagada na rede para trás, isto se a rede for deste tipo, ajustando gradualmente os pesos. Uma vez treinada, a rede pode ser aplicada a casos em que o resultado é desconhecido (Bastos, 2003).

A ferramenta tem a operação *Train Net* que permite criar e treinar uma rede neuronal, definindo os campos de entrada e saída dos dados de treino. A rede neuronal “aprende” a classificar ou prever os valores dos campos de saída através dos valores dos campos de entrada. É notório que as regras de decisão para a classificação são internas à rede neuronal, sendo difícil de acompanhar o raciocínio usado no processo de tomada de decisão (Bastos, 2003).

Indução de regras

Um dos problemas das redes neuronais é a falta de transparência das decisões tomadas pela rede. Visto que o funcionamento interno deste algoritmo é resumido a números, é difícil encontrar um raciocínio lógico no processo de tomada de decisão (Clementine, 1998).

As árvores de decisão consistem numa técnica complementar, trabalhando com a totalidade dos dados ou apenas com um subconjunto, a indução cria uma árvore de decisão que representa regras de como classificar os dados nas várias saídas. A estrutura de árvore e geração de regras são técnicas abertas e explícitas que podem ser navegadas (Bastos, 2003). Por exemplo, a regra seguinte descreve as características de um dia de ida à praia, figura 98 (Rodrigues, 2008):

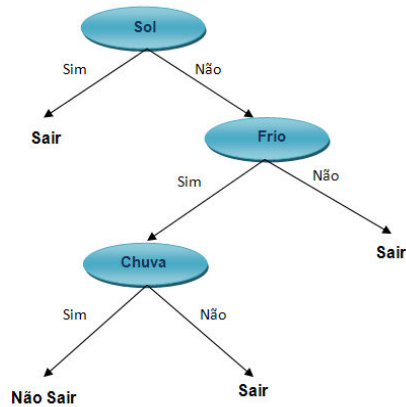


Figura 98 - Árvore de decisão

Outra vantagem da árvore reside no facto do processo incluir, automaticamente apenas os atributos que têm realmente importância na tomada de decisão, os outros são descartados. Isto faz com que se obtenha informação útil sobre os dados, sendo possível eliminar os dados que não são relevantes antes de treinar uma rede neuronal. As árvores de decisão podem ser convertidas num conjunto de regras tipo “Se-Então”, dando origem ao conjunto de regras que, na maioria dos casos, mostra a informação de uma forma inteligível (Clementine, 1998).

A estrutura em árvore é muito útil, quando se pretende saber quais os fatores que dividem a população nos vários subconjuntos. A estrutura de regras, por seu lado, mostra como um conjunto em particular se relaciona com uma conclusão (Bastos, 2003).

A ferramenta *Clementine* dispõe de duas operações, *C&RTree* e *C5.0*, que produzem regras sob a forma de árvore de decisão, representando como classificar os dados em diferentes resultados. A estrutura das árvores representa explicitamente o raciocínio da tomada de decisão (Bastos, 2003).

Redes de Kohonen

As redes de *Kohonen* são um tipo de rede neuronal que permite efetuar *clustering*, isto é, dividir a população em segmentos. A unidade básica é o neurónio e estes estão organizados em duas camadas: camada de entrada e saída. Todos os neurónios de entrada estão ligados a todos os neurónios de saída e essas ligações possuem um peso associado (Clementine, 1998). A camada de saída consiste numa tabela de neurónios bidimensional sem qualquer ligação entre as células, como se pode observar através da seguinte figura (Bastos, 2003):

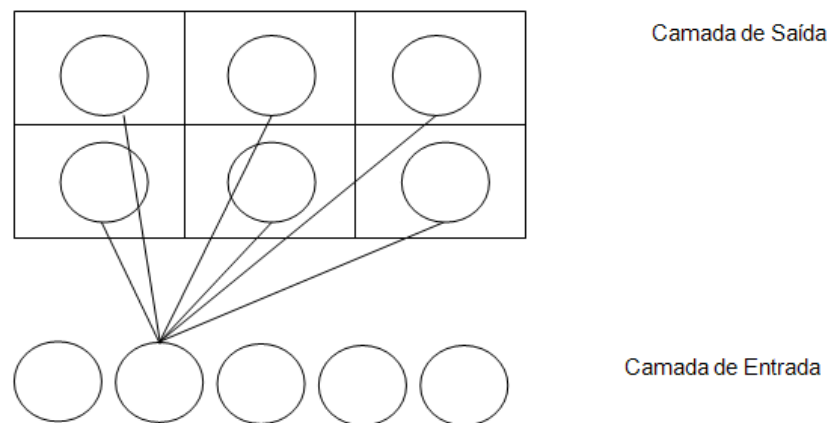


Figura 99 - Rede de *Kohonen*

Os dados de entrada são fornecidos à camada de entrada e os valores são propagados para a camada de saída. De seguida, cada neurónio de saída fornece uma resposta. O neurónio de saída que fornecer a melhor resposta é o “vencedor” e constitui a resposta à entrada fornecida (Bastos, 2003).

Inicialmente, todos os pesos são aleatórios. À medida que a rede é treinada, o peso vencedor é ajustado de forma a responder cada vez melhor, assim como os pesos dos seus vizinhos. Este processo é repetido até as alterações feitas nos pesos serem muito pequenas (Bastos, 2003).

Quando a rede estiver treinada, os registos com características semelhantes aparecerão juntos na camada de saída, ao contrário dos registos com características distintas, que se encontrarão distantes (Bastos, 2003).

A ferramenta *Clementine* dispõe de alguns algoritmos *K-Means*, *Kohonen*, *Two Step*, que criam um modelo de agrupamento, isto é, descobre semelhanças nos dados originais e agrupa-os, de

forma a maximizar a similaridade dentro do grupo e a maximizar a diferença entre os grupos (Rodrigues, 2008).

Regras de associação

As regras de associação, como o próprio nome indica, associam uma determinada conclusão a um conjunto de condições (Clementine, 1998). Por exemplo a regra:

Basquetebol ← Voleibol & Andebol (173,17.0%,0.84)


Num contexto desportivo, indica que quem pratica voleibol e andebol, também pratica basquetebol, com um grau de confiança de 84% e representando 17% da população, ou seja, 173 registos (Bastos, 2003). Para descobrir regras, o algoritmo utiliza um método de geração e testa. Inicialmente, são criadas regras simples que são validadas mediante o conjunto de dados fornecidos. As regras “boas” são armazenadas e todas as regras, após serem submetidas a várias restrições, são “especializadas”. A especialização é o processo que adiciona condições às regras, que são posteriormente validadas. O processo iterativo armazena as regras “melhores” ou “mais interessantes” que encontrou. Normalmente, o utilizador fornece um limite para o número de conjunções possíveis numa regra e utiliza um conjunto de técnicas baseadas na indexação, que permitem reduzir o tempo de procura. No fim do processo, é criada uma tabela que contém as melhores regras encontradas (Bastos, 2003).

Infelizmente, estas regras, ao contrário das árvores de decisão, não podem ser utilizadas diretamente para fazer previsões, visto existirem diferentes conclusões possíveis. Para obter a árvore de decisão associada a uma determinada regra produzida, é necessário que ela seja tratada separadamente (Clementine, 1998).

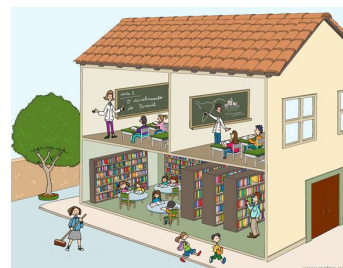
A ferramenta *Clementine* dispõe de algoritmos GRI (*Generalised Rule Induction*), APRIORI e Carma, que automaticamente, encontram associações que existam entre os atributos, explicitando essas regras associativas (Rodrigues, 2008).

Anexo D

De seguida, é apresentado o teste diagnóstico de proporcionalidade direta a que os alunos foram submetidos:

 Agrupamento Vertical de S. Lourenço	Escola EB 2,3 de S. Lourenço – Ermesinde
	Teste diagnóstico – Matemática 7º ano Proporcionalidade direta
	Nome: _____ Nº: _____ Turma: _____
	Data: ____/____/____ Classificação: _____
	Professora: _____

1. Na escola do Pedro existem 1200 alunos, dos quais 640 são rapazes. Encontra uma razão entre:
 - 1.1. O número de rapazes e o número de alunos da escola.
 - 1.2. O número de raparigas e o número de alunos da escola.
 - 1.3. O número de rapazes e o número de raparigas.



- 1.4. Escreve uma razão equivalente a $\frac{3}{4}$.

2. Considera a proporção $\frac{8}{6} \frac{12}{9}$ indica os:

2.1. extremos _____

2.2. meios _____

2.3. antecedentes _____

2.4. consequentes _____

2.5. Faz a leitura da proporção.

3. Completa de modo a obteres uma proporção: (apresenta os cálculos)

3.1. $\frac{24}{?} = \frac{10}{15}$

4. Com os número 10, 8, 5 e 4 escreve uma proporção.

5. Explica como é que fazias para constituir uma turma em que a razão das raparigas para os rapazes fosse de 3 para 4?

6. Num domingo, um parque da cidade foi visitado por 5400 pessoas, entre adultos e crianças. A razão entre o número de adultos e o número de crianças é de 1 : 3.

6.1. Quantos visitantes eram adultos?
(apresenta os cálculos)



6.2. E quantas eram as crianças?

7. Resolva o problema:

Num galinheiro obtêm-se 8 ovos por cada 10 galinhas.

Quantas galinhas são necessárias para se obterem 280 ovos?



Anexo E

Após o estudo da unidade de "Proporcionalidade Direta" na disciplina de Matemática recorrendo ao uso do PCMAT, alguns alunos foram questionados em relação à usabilidade do mesmo. Inicialmente, é apresentada a opinião dos alunos do 7º e 8º ano de escolaridade relativamente a algumas características do PCMAT.

Tabela 39 - Características do PCMAT

PCMAT	Discordo totalment.		Discordo maioritaria mente		Concordo maioritaria mente		Concordo totalment.		Sem opinião	
	7º	8º	7º	8º	7º	8º	7º	8º	7º	8º
Facilita o estudo fora das aulas				1	5	4	10	6		1
Facilita o estudo dentro das aulas			1		7	2	7	5		5
Bem organizado				2	3	3	11	7	1	
Bom grafismo/ design			1	3	8	4	6	4		1
Fácil acesso					6	3	8	8	1	1
Fácil acesso aos conteúdos da disciplina					6	5	9	6		1
Útil para melhorar os resultados da disciplina				1	7	7	7	4	1	
Recomenda o uso do PCMAT					3	4	11	6	1	2
Todos os professores usassem o PCMAT nas aulas				1	5	4	10	7		
Aceder facilmente às atividades da disciplina					6	3	9	7		2

Através dos dados exibidos na tabela 20, é averiguado que a maioria dos alunos do 7º e 8º ano de escolaridade concordaram maioritariamente ou totalmente com a caracterização do PCMAT: fácil acesso a conteúdos e atividades, bem organizado, bom grafismo, facilita o estudo e promove melhoria de resultados. A minoria dos alunos é que manifestaram a sua opinião de discordância ou mencionaram que não possuem qualquer opinião relativamente aos pontos citados anteriormente.

Seguidamente, é exibido o grau de utilidade do PCMAT segundo a opinião dos alunos das diversas turmas.

Tabela 40 - Grau de utilidade do PCMAT

PCMAT	Inútil		Pouco útil		Útil		Muito útil		Não conheço	
	7º	8º	7º	8º	7º	8º	7º	8º	7º	8º
Índice de conteúdos					11	9	4	2		1
Conteúdos					5	6	10	6		
Atividades					6	5	9	7		
Hiperligação			2	1	8	8	5	3		

A maioria dos alunos consideraram o PCMAT útil ou muito útil no processo de aprendizagem da disciplina de Matemática, nomeadamente na exposição do índice de conteúdos, conteúdos, atividades e hiperligações do sistema. Apenas cerca de 3 alunos é que manifestaram opinião contrária.

Relativamente ao grau de dificuldade do PCMAT, a maioria dos alunos demonstraram que este sistema é fácil ou muito fácil de usar. É considerado conveniente mencionar que alguns alunos do 7º ano de escolaridade sentiram alguma resistência ao uso do PCMAT.

Tabela 41 - Grau de dificuldade do PCMAT

PCMAT	Difícil		Pouco fácil		Fácil		Muito fácil		Não conheço	
	7º	8º	7º	8º	7º	8º	7º	8º	7º	8º
Índice de conteúdos			2		8	8	5	3		1
Conteúdos	1		1		8	8	5	4		
Atividades			1		11	5	3	7		
Hiperligação	2		2		7	5	4	4		3

A frequência de uso do PCMAT fora das aulas pelos alunos foi um pouco reduzida, no 7º ano de escolaridade, cerca de 8 alunos que utilizaram menos de uma vez por semana o sistema, 6 alunos pelo menos uma vez por semana e 1 aluno que utilizava todos ou quase todos os dias. No 8º ano de escolaridade, a distribuição da frequência também foi similar, apenas 4 alunos que utilizavam menos de uma vez por semana, 7 alunos pelo menos uma vez por semana e 1 aluno todos ou quase todos os dias.

Tabela 42 - Frequência de uso do PCMAT fora das aulas

Ano de escolaridade	Nunca	Menos de uma vez por semana	Pelo menos uma vez por semana	Todos ou quase todos os dias	Total
7º ano		8	6	1	15
8º ano		4	7	1	12

Os alunos gostaram de utilizar o PCMAT e recomendaram o uso desta ferramenta nas outras disciplinas, com a exceção de um aluno do 8º ano de escolaridade.

Tabela 43 - Recomendação do uso do PCMAT noutras disciplinas

Ano de escolaridade	Sim	Não	Total
7º ano	15	0	15
8º ano	11	1	12

É importante referir que a análise da usabilidade apenas se refere a três turmas, duas do 7º ano, 7º F e G, e uma do 8º ano de escolaridade, 8º F.