



# ALGORITMOS DE MACHINE LEARNING APLICADOS EM EDIFÍCIOS INTELIGENTES COM ELEVADA PENETRAÇÃO DE VEÍCULOS ELÉTRICOS

**THIAGO CESAR ROSA**

novembro de 2022

# ALGORITMOS DE *MACHINE LEARNING* APLICADOS EM EDIFÍCIOS INTELIGENTES COM ELEVADA PENETRAÇÃO DE VEÍCULOS ELÉTRICOS

Thiago Cesar Rosa



Departamento de Engenharia Eletrotécnica  
Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

**2022**



Relatório elaborado para satisfação parcial dos requisitos da Unidade Curricular de TEDSEE - Dissertação do Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia. Esta dissertação foi desenvolvida no âmbito do Projeto de Investigação Científica: PRECISE “*Power and Energy Cyber-Physical Solutions with Explainable Semantic Learning*” financiado pelo FEDER, pelo COMPETE e FCT com a referência: PRECISE (PTDC/EEI-EEE/6277/2020)

Candidato: Thiago Cesar Rosa, N° 1190042, 1190042@isep.ipp.pt

Orientação científica: Dr. João André Pinto Soares, jan@isep.ipp.pt

Coorientação científica: Prof Dr. Sérgio Ramos, scr@isep.ipp.pt



Departamento de Engenharia Eletrotécnica  
Mestrado em Engenharia Eletrotécnica – Sistemas Elétricos de Energia

**2022**



Dedico este trabalho aos meus pais Maria Nicefora e Benedito Cesar obviamente sem eles  
eu não estaria aqui.



## *Agradecimentos*

Agradecimentos que são devidos prestar ao professor Sérgio Ramos e pela oportunidade e ao Dr. João Soares pela paciência e ajuda no trabalho.

Agradeço em especial uma das minhas professoras do ensino médio (secundário que me incentivou explorar a área técnica e posteriormente de engenharia)

Agradeço também a todos os meus colegas de faculdade.



## *Resumo*

A presente dissertação discute o desenvolvimento de um método de previsão de ocupação para dois parques de estacionamentos residenciais no contexto de um edifício inteligente, a fim de se conhecer, antecipadamente, qual a taxa de ocupação desses parques de estacionamentos. Para concretizar tal objetivo, utilizaram-se dados históricos realísticos coletados por observação empírica e extrapolado para um ano.

O modelo de previsão desenvolvido utiliza técnicas de *machine learning* com diversos algoritmos testados, entre eles, *Decision Tree*, *Extra Tree*, *Logistic Regression*, *Random Forest*, *Naive Bayes*, *K-Nearest Neighbors* e *Support Vector Machine*. No modelo proposto foi identificado qual destes algoritmos obteve melhor desempenho. Vários tipos de modelos foram testados com o objetivo de melhorar os resultados obtidos, bem como compreender o impacto de cada um dos tratamentos dos dados utilizados.

A solução final teve seu desempenho validado, com métricas de avaliação com bons resultados, exatidão e precisão superiores a 80%, e se mostrou eficaz considerando os dados analisados e ainda o horizonte temporal da previsão.

### *Palavras-Chave*

Algoritmos, Inteligência Artificial, *K-Nearest Neighbors*, *machine learning*, ocupação de estacionamento, técnicas de previsão.



## *Abstract*

This dissertation is focused on the development of a prediction method for two residential car parks in the context of an intelligent building. The aim was to know in advance the occupancy rate of these car parks, using only historical data collected by empirical observation, and extrapolate for one year.

The prediction model developed uses machine learning techniques with several tested algorithms (Decision Tree, Extra Tree, Logistic Regression, Random Forest, Naive Bayes, K- Nearest Neighbors and Support Vector Machine) to identify which of these algorithms performs better. Several types of models were tested with the objective of improve the results obtained, and understand the impact of each of the data treatments used.

The final solution had its performance validated, with good evaluation metrics results. Accuracy and precision were higher than 80% and, therefore, the solution proved to be effective considering the data analyzed and the temporal horizon of the forecast.

### ***Keywords***

Algorithm, Artificial Intelligence, forecasting, K-Nearest Neighbors, machine learning, parking occupancy, prediction techniques.







# Índice

<b>AGRADECIMENTOS</b> .....	<b>I</b>
<b>RESUMO</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>V</b>
<b>ÍNDICE</b> .....	<b>IX</b>
<b>ÍNDICE DE FIGURAS</b> .....	<b>XI</b>
<b>ÍNDICE DE TABELAS</b> .....	<b>XIII</b>
<b>ACRÓNIMOS E SIGLAS</b> .....	<b>XV</b>
<b>1. INTRODUÇÃO</b> .....	<b>1</b>
1.1.OBJETIVOS .....	2
1.2.HIPÓTESES .....	3
1.3.ESTRUTURA DO TRABALHO .....	3
<b>2. ESTADO DE ARTE</b> .....	<b>5</b>
2.1.EDIFÍCIOS INTELIGENTES .....	5
2.2.SISTEMAS DE AVALIAÇÃO DE OCUPAÇÃO DE PARQUES DE ESTACIONAMENTO .....	8
2.3.TÉCNICAS DE PREVISÃO .....	12
2.4.MACHINE LEARNING .....	14
2.5.CONCLUSÕES .....	20
<b>3. METODOLOGIA DO TRABALHO</b> .....	<b>21</b>
3.1.BASE DE DADOS .....	22
3.2.ANÁLISE DOS DADOS .....	24
3.3.TRATAMENTO DOS DADOS .....	27
3.4.MODELO FINAL .....	30
3.5.CONCLUSÕES .....	31
<b>4. IMPLEMENTAÇÃO</b> .....	<b>33</b>
4.1.AVALIAÇÃO E ESCOLHA DO MODELO .....	33
4.2.DIVISÃO, TREINO, TESTE E VALIDAÇÃO.....	37
4.3.COMPARAÇÃO DOS RESULTADOS .....	38
4.4.CONCLUSÕES .....	39
<b>5. ANÁLISE DOS RESULTADOS</b> .....	<b>41</b>

5.1.PREVISÕES .....	42
5.2.ANÁLISE DAS PREVISÕES.....	50
<b>6. CONCLUSÕES.....</b>	<b>53</b>
6.1.CONTRIBUTOS .....	54
6.2.TRABALHO FUTURO.....	55
<b>7. REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>57</b>
<b>ANEXO .....</b>	<b>63</b>

## Índice de Figuras

Figura 1 - Exemplo de uma <i>DT</i> [48]	17
Figura 2 - Representação visual de uma <i>Extra Tree</i>	17
Figura 3 - Exemplo de uma <i>SVM</i> [21]	18
Figura 4 - Exemplo de <i>KNN</i> com células de aproximação [40]	19
Figura 5 - Exemplo de uma <i>Random Forest</i> [21]	20
Figura 6 - Esquema ilustrador do fluxo de trabalho do projeto	22
Figura 7 - Distribuição média de veículos estacionados x hora do dia	26
Figura 8 - Distribuição média de veículos estacionados x hora do dia feriado 10 de junho	26
Figura 9 - Distribuição média de veículos estacionados x hora do dia período de férias.	27
Figura 10 - Distribuição normal	30
Figura 11 - Exemplo de uma matriz de confusão	34
Figura 12 - Curva ROC Cave	36
Figura 13 - Curva ROC Subcave	36
Figura 14 - Comparação de 20 amostras desempenho treino x teste <i>Random Forest</i>	37
Figura 15 - Precisão média ao longo das semanas	39
Figura 16 - Lugares Estacionamento Cave	42
Figura 17 - Lugares Estacionamento Subcave	42
Figura 18 - Desempenho <i>KNN</i>	44
Figura 19 - Distribuição média de veículos estacionados x hora do dia previsão	45
Figura 20 - Distribuição média de veículos estacionados x hora do dia acumulada	46
Figura 21 - Distribuição média de veículos estacionados x hora do dia previsão	47
Figura 22 - Distribuição média de veículos estacionados x hora do dia acumulada	48
Figura 23 - Distribuição média de veículos estacionados x hora do dia previsão	49
Figura 24 - Distribuição média de veículos estacionados x hora do dia acumulada	50
Figura 25 - Desempenho de Exatidão e Precisão <i>KNN</i> (sem feriados e férias)	64
Figura 26 - Desempenho de Exatidão e Precisão <i>KNN</i> (ordem aleatória)	65



## *Índice de Tabelas*

Tabela 1 - Comparação de estudos com sistemas de aquisição de dados	11
Tabela 2 - Comparação de estudos com diferentes técnicas de previsão	14
Tabela 3 - Feriados Nacionais em 2018	23
Tabela 4 - Estrutura dos dados (cave)	24
Tabela 5 - Taxa de ocupação por dia da semana	25
Tabela 6 - Tipos de dados por coluna	28
Tabela 7 - Base de dados para criação dos modelos	31
Tabela 8 - Resultado da Exatidão na primeira avaliação	35
Tabela 9 - Comparação da Precisão dos Algoritmos	35
Tabela 10 - Comparação da precisão dos algoritmos selecionados	38
Tabela 11 - Desempenho comparativo da divisão aleatória dos dados	44
Tabela 12 - Desempenho após uma semana de previsão	44
Tabela 13 - Desempenho após oito semanas de previsão	44
Tabela 14 - Desempenho após uma semana de previsão	47
Tabela 15 - Desempenho após oito semanas de previsão	47
Tabela 16 - Desempenho após uma semana de previsão	49
Tabela 17 - Desempenho após oito semana de previsão	49
Tabela 18 - Estrutura dos dados (subcave)	63
Tabela 19 - Estrutura dos dados sem férias e feriados(subcave)	63
Tabela 20 - Estrutura dos dados sem férias e feriados(cave)	64



## *Acrónimos e Siglas*

DT	-	<i>Decision Tree</i>
FN	-	Falso Negativo
FP	-	Falso Positivo
IA	-	Inteligência Artificial
KNN	-	<i>K- Nearest Neighbors</i>
PHVE	-	<i>Plugin Hybrid Electric Vehicles.</i>
PUK	-	<i>Pearson Function Kernel</i>
RMSE	-	<i>Roots Mean Squared Error</i>
SVM	-	<i>Support Vector Machine</i>
SOC	-	<i>State of Charge</i>
SVR	-	<i>Support Vector Regression</i>
VN	-	Verdadeiro Negativo
VP	-	Verdadeiro Positivo



# 1. INTRODUÇÃO

O parque de estacionamento em edifícios residenciais ou outros é responsável por uma boa parte dos custos de energia e indiretamente das emissões de gases de efeito estufa. Nesse sentido, os edifícios inteligentes representam um passo importante para se lograr a redução de emissões e dos custos associados. No entanto, os desafios tecnológicos, sociais e regulamentares para atingir essas reduções são consideráveis e precisam ser considerados. Nesta dissertação parte-se do princípio de que a inteligência artificial (IA), nomeadamente *machine learning*, terá um papel fundamental para o futuro da rede inteligente de energia, concretamente para edifícios e veículos elétricos (VE). Sendo assim, neste trabalho propõe-se a previsão da ocupação de VEs num edifício residencial, utilizando algoritmos de *machine learning*. A antevisão com maior precisão da ocupação do parque de estacionamento de um edifício inteligente é de crucial importância, e trará benefícios para os seus gestores e para a otimização de recursos do mesmo.

Os edifícios inteligentes são construções que utilizam tecnologia para otimizar, gerenciar e operacionalizar o uso eficiente dos seus recursos e, adicionalmente, permitem que estes sejam controlados e monitorados remotamente através da internet ou rede privada de comunicação Harisson et al. (1998) [1]. Neste contexto, o principal propósito de um edifício inteligente é reduzir o consumo de energia e as suas respetivas emissões, sem colocar em causa o conforto de seus ocupantes. Contudo, este propósito é também o maior desafio atual na investigação sobre a construção e utilização inteligente de recursos dos edifícios inteligentes.

No âmbito e contexto dos edifícios inteligentes, o objetivo é realizar previsões em tempo (quase) real, próximas horas e próximo dia, com a maior exatidão possível, sendo este o principal objetivo deste projeto, com o estudo, análise e implementação de algoritmos de *machine learning* que poderão ser aplicados posteriormente para prever a taxa de ocupação do parque de estacionamento de um edifício inteligente.

Diversas técnicas de previsão surgem no desenvolvimento da área de *machine learning*, sendo que estas técnicas, quando devidamente exploradas, permitem a gestão, análise e tratamento de um elevado conjunto de dados de elevada complexidade Mitchell (1997) [33]. Sendo assim, estas técnicas permitem que máquinas aprendam e compreendam padrões do mundo real por meio de sistemas computacionais, sempre a partir de entradas (experiências anteriores), para que de forma ordenada esta técnica apresente em suas saídas soluções para os mais variados e complexos problemas.

Com as diversas vantagens que são contempladas pelos algoritmos de *machine learning* para a previsão de ocupação de um parque de estacionamento, é extremamente relevante investigar alternativas baseadas nestas técnicas para auxiliar a resolução de problemas da ocupação dos VE no interior destes edifícios, a fim de contribuir com o seu propósito de gerir, otimizar e operacionalizar seus recursos.

## **1.1. OBJETIVOS**

O objetivo é identificar e conhecer os perfis de utilização (ocupação) dos VE no estacionamento do edifício e a sua taxa de ocupação, pois isto será muito importante para dotar o edifício inteligente do futuro de uma maior quantidade de informação, para a partir daí ser possível uma gestão mais assertiva e eficaz da energia do edifício. Isto é, conhecer o perfil de carga e a quantidade de energia a armazenar em um dado dia, para suportar a carga de carregamento dos veículos.

Sendo assim, o foco é realizar uma revisão da bibliografia sobre o tema, no campo dos algoritmos de *machine learning*, para ser possível escolher o algoritmo mais adequado para realizar a precisão da taxa de ocupação de VE no parque de estacionamento do edifício e contemplando os aspetos de *hardware*, relativos aos sensores de presença em cada lugar do parque de estacionamento. Com estes dados tem-se a possibilidade de aplicação de métodos e técnicas que irão ler estes dados, técnicas estas de IA, supracitadas e utilizá-las

para realizar, estimativas, previsões e classificações de comportamento destes veículos, no contexto dos edifícios inteligentes, e assim auxiliar a gestão do edifício e a tomada de decisão.

## **1.2. HIPÓTESES**

A hipótese deste trabalho consiste em comparar técnicas e algoritmos de *machine learning* aplicados à previsão de ocupação VE no estacionamento de um edifício. Com o intuito de reproduzir, na previsão a ser realizada, uma exatidão superior a 80%, desempenho obtido em trabalhos como o de Provoost et al. (2019) [47] além de outros citados no Capítulo 2. Desta forma, é apresentada uma técnica de manipulação de dados de treinamento e validação dos modelos, além de sete algoritmos de *machine learning*.

Procurou-se, também, entender o desempenho dos modelos em diferentes cenários propostos com diferentes divisões de dados entre eles, divisão cronológica, divisão aleatória e divisão aleatória sem feriados e período de férias. Portanto, de forma geral, são analisadas e investigadas as vantagens e desvantagens do modelo proposto.

## **1.3. ESTRUTURA DO TRABALHO**

Esta dissertação está organizada da seguinte forma: no Capítulo 1, é apresentada uma visão dos objetivos gerais do presente trabalho e o tema alvo da pesquisa que envolve a pesquisa de algoritmos de *machine learning* e a comparação de diferentes técnicas de aprendizado e a avaliação de parâmetros estatísticos dos resultados.

Em seguida, o Capítulo 2, traz uma revisão técnica da bibliografia histórica relacionado aos temas abordados, edifícios inteligentes, VE e *machine learning*, além de descrever em detalhe e enfatizar os estudos que empregam os mesmos algoritmos que foram explorados neste trabalho. Também apresenta os algoritmos utilizados: *Random Forest*, *Support Vector Machine*, *Decision Tree* (DT), *K- Nearest Neighbors* (KNN), *Naive Bayes*, *Logistic Regression* e *Extra Tree*. Compreende, ainda, as formulações matemáticas, o funcionamento, vantagens e os aspectos particulares de cada algoritmo considerado.

O Capítulo 3 enumera os *softwares* e ferramentas utilizadas no desenvolvimento dos modelos de IA. São, também, discutidos os resultados obtidos pelos principais referenciais que guiaram o desenvolvimento deste estudo.

No Capítulo 4 discute-se a abordagem metodológica elaborada para os algoritmos de previsão, com a descrição de suas etapas que serão utilizadas de base para discutir e avaliar o Capítulo seguinte.

Em seguida, no Capítulo 5 apresenta-se uma análise dos principais pontos do modelo e discutiu-se os resultados obtidos, avaliados pelas métricas e sua eficiência. Os dados são apresentados de forma estruturada considerando o horizonte analisado bem como a comparação entre os modelos de previsão criados.

Finalmente, esta dissertação termina no Capítulo 6, que apresenta a conclusão, as contribuições realizadas pela pesquisa e possíveis trabalhos futuros relacionados. Os apêndices contêm os códigos dos algoritmos de previsão desenvolvidos para cada modelo de previsão proposto.

## 2. ESTADO DE ARTE

Neste Capítulo é apresentado um olhar sistemático sobre a literatura nos três temas que englobam este trabalho, porém com ênfase maior para os algoritmos de *machine learning*. A primeira seção fornece uma visão geral da pesquisa em edifícios inteligentes. A segunda traz aspectos da literatura sobre VE. A terceira apresenta a revisão sobre *machine learning*.

### 2.1. EDIFÍCIOS INTELIGENTES

A palavra "inteligente" foi usada pela primeira vez para descrever edifícios nos Estados Unidos no início dos anos 80 do século passado. Já o conceito de "edifício inteligente" foi estimulado pelo desenvolvimento da tecnologia da informação Harisson et al. (1988) [1].

A literatura traz diversas abordagens sobre edifícios inteligentes, muitas delas focadas no desenvolvimento dos aspectos construtivos e avaliação do desempenho. No entanto, pouca literatura tem sido dedicada à abordagem de otimização dos recursos energéticos em edifícios inteligentes.

Sendo os edifícios inteligentes dotados de sistemas tecnológicos que englobam *hardwares* e *softwares*, como sensores, câmaras de vigilância e as próprias técnicas de IA e sistemas de comunicação que quando analisadas e estudadas de forma científica ajudam a tomada de decisão do gestor do edifício. Decisões estas que têm por objetivo ajudar o edifício a ser

mais eficiente do ponto de vista energético, mais seguro e mais resiliente. Para viabilizar estes pontos os parques de estacionamento dos edifícios passam a ser digitalizados para permitir o controlo dos VE e perceber como o estacionamento esta sendo utilizado a cada momento e em cada dia.

O objetivo desta seção é fornecer uma revisão sucinta da pesquisa existente sobre os edifícios inteligentes, começamos com as definições de edifícios inteligentes, seguimos para o resumo das áreas de pesquisa atuais em edifícios inteligentes.

### **2.1.1. DEFINIÇÃO DOS EDIFÍCIOS INTELIGENTES**

Uma das características fundamentais presente nos edifícios inteligentes é a integração do edifício com a internet, com a possibilidade de realizar o monitoramento de diversos parâmetros, como por exemplo: emissões em tempo real e acumuladas, qualidade do ar interno, nível de iluminação, consumo de energia, número de ocupantes e vagas de estacionamentos disponíveis, todos estes parâmetros com o intuito de minimizar as emissões e economizar energia do edifício.

De acordo com a pesquisa realizada em [1], existem mais de 30 definições separadas de inteligência em relação à construção. As primeiras definições de edifício inteligente tinham como foco o aspeto tecnológico e não contemplavam indicação de interação do usuário. O Intelligent Building Institution em Washington (1988) define o edifício inteligente como *“aquele que integra vários sistemas para gerenciar recursos de forma eficaz e coordenada para maximizar: desempenho técnico, economia de investimentos e custos operacionais, flexibilidade”*.

Diversos autores descrevem ou apresentam perspectiva do ponto de vista puramente dos aspetos tecnológicos que contemplam a envolvente do edifício inteligente. Pode-se destacar de acordo com Arkin e Paiuck (1997) [2] que o novo contexto destes edifícios contempla um *"esforço multidisciplinar para integrar e otimizar as estruturas, sistemas, serviços e gerenciamento do edifício, a fim de criar um ambiente produtivo, econômico e ambientalmente aprovado para os ocupantes do edifício"*. Um ponto muito relevante é apresentado por Yang e Peng (2001) [3] consiste na capacidade de aprendizagem e adaptação do ambiente, neste contexto propôs além da capacidade de reação do edifício, este também ser capaz de aprender e otimizar o desempenho em relação à sua ocupação.

Ainda do ponto de vista energético os edifícios inteligentes, são responsáveis por aproximadamente um terço do consumo de energia mundial Joench et al. (2019) [53] e por consequência os VEs terão um papel fundamental neste consumo.

### **2.1.2. VEÍCULO ELÉTRICO NO CONTEXTO DOS EDIFÍCIOS INTELIGENTES**

Os VE tem um papel fundamental na exploração de oportunidades agregadas aos edifícios inteligentes [4]. Uma grande vantagem em utilizar os VE neste contexto é aproveitando a sua flexibilidade energética com proposto por Schuller et al. (2015) [51] e as estratégias de controlo para ocupação devem ser geridas de maneira coordenada.

Sobre o tempo de carregamento dos veículos é extremamente importante controlar e coordenar este carregamento e consequentemente otimizar a quantidade de energia para por fim minimizar o custo com energia elétrica, Rajakaruna et al. (2015) [5] apresentam um algoritmo de carregamento de VE que compreende tanto a potência de carregamento quanto o tempo. É essencial aplicar estratégias para o carregamento dos VEs e para aplicar estas estratégias é necessário conhecer o perfil de utilização dos veículos neste contexto. Começando pelo horário de entrada, saída, tempo que permanece no edifício, se está conectado e para realizar uma recarga ou eventualmente sendo utilizado para fornecer eletricidade para o edifício, na condição de veículo para o edifício (V2B). Na configuração V2B o veículo é utilizado como um vetor de energia utilizado para a troca de eletricidade com o edifício. E para tal alguns aspetos têm de ser levados em consideração, assim como tempo de carregamento, tempo de chegada, tempo de partida e potência de carga. Através da tecnologia V2B o veículo pode servir como um sistema de armazenamento de energia móvel e como uma fonte de energia renovável Rahmani (2019) [6].

Em [52] existe um exemplo de gerenciamento de energia de um edifício inteligente, considerando todos os recursos energéticos diretamente relacionados e bem como o uso intensivo dos VE, neste contexto.

Um exemplo da flexibilidade energética oferecida a um edifício quando este conta com VE é apresentada por Y. Zhou, et al, (2019) [7], que desenvolveram um modelo baseado em componentes não lineares, integrando a sistemas fotovoltaicos em edifícios e sistemas fotovoltaicos integrados em VEs, para a avaliação da flexibilidade energética. Este estudo aponta que a utilização de um sistema de gestão flexível é fundamental para o controle

energético do edifício (medida por meio de vários indicadores, como o rácio de descargas renováveis fora do pico e o rácio de descargas da rede fora do pico).

É importante observar que o objetivo é saber com precisão o perfil de utilização dos VE no estacionamento dos edifícios, que por fim será fundamental para fomentar a gestão energética e adicionalmente a preocupação com o fator econômico, como redução da fatura de eletricidade [5].

## **2.2. SISTEMAS DE AVALIAÇÃO DE OCUPAÇÃO DE PARQUES DE ESTACIONAMENTO**

A previsão da ocupação de vagas de estacionamento tem sido um tema de estudo há algum tempo, mas com um especial aumento nos últimos anos. Os avanços no poder de processamento computacional, a mobilidade, as tecnologias de veículos e de edifícios inteligentes, sensores e comunicações integradas levam a novas abordagens do problema, buscando obter resultados mais eficientes e precisos.

Os diferentes métodos utilizados variam na forma em que os dados são coletados e calculados, tendo este impacto direto sobre o tipo de infraestrutura necessária, tempo de implementação e custos.

### **2.2.1. SISTEMA BASEADO EM TEMPO REAL**

Adquirir os dados de ocupação do estacionamento e calcular a demanda em tempo real é possível com as tecnologias atuais utilizando sensores ou câmaras e sistemas em rede Klappennecker et al, (2012) [8].

Entretanto o monitoramento de cada vaga de um estacionamento é economicamente caro, em contrapartida é bastante viável monitorar o fluxo de veículos que entram e saem de um estacionamento [8].

Alguns dos sistemas baseados em dados coletados em tempo real usam câmaras inteligentes que são capazes de realizar uma deteção visual de ocupação, Amato, et al. (2017) [10] e esta deteção foi baseada em uma rede neural convolucional profunda que foi projetada especificamente para as câmaras inteligentes. Este sistema tem a capacidade de monitorar diversas vagas simultaneamente o que elimina a necessidade de inserção de um

sensor por cada lugar do estacionamento o que por consequência reduz de forma significativa os custos de instalação e manutenção dos sensores.

Com uma técnica de blocos de predição, interpolação e extrapolação um método que provou ser muito eficaz, mas também muito sensível as condições climáticas em especial as condições de iluminação. Neste contexto este tipo de sistema está sujeito a condições ambientais que podem afetar muito o desempenho desses sistemas, como sombras, variações de luz, e condições meteorológicas se instalado em um ambiente a céu aberto Delibaltov, et al. (2013) [11].

As informações dos sensores são produzidas em grande quantidade e com uma alta taxa de heterogeneidade, o que também conhecida como “*Big Data*”, o que significa que há uma necessidade de aplicar ferramentas eficientes para analisar esses dados Rong, et al. (2018) [9].

As tecnologias baseadas em sensores utilizam um sensor por vaga de estacionamento para detectar se o veículo está ou não estacionado. Consoante à finalidade e aplicação nos estacionamentos, são utilizados diferentes tipos de sensores, que podem ser divididos em duas categorias:

- Intrusiva;
- Não-intrusiva.

Os sensores intrusivos são instalados sob a área construída normalmente são sensores piezoelétricos, de peso em movimento ou magnetômetros. Os sensores não intrusivos são normalmente fixados no solo ou no teto, sendo normalmente ultrassônicos.

Uma forma muito interessante e de uso crescente nos últimos anos é um sistema baseado nos dados que os usuários fornecem. Esta forma é chamada de “*crowdsourcing*”. Como as informações dos usuários de forma estruturada contribuem para eficiência do sistema, sendo assim os usuários são encorajados a disponibilizar informações sobre vagas disponíveis naquele momento e se a informação estiver correta eles podem ser recompensados com descontos em vagas de estacionamento ou outros benefícios, inserindo assim o conceito de “*gamefication*”.

Em contrapartida, este tipo de sistema tem a necessidade de um elevado número de usuários, de modo a garantir que os dados que são produzidos são suficientes para gerar dados com informações de vagas disponíveis adequadas. Além disso, a veracidade e a qualidade das informações não são asseguradas, o que significa que há necessidade de elaborar métodos para garantir que a precisão seja atingida o que, conseqüentemente, adicionando mais complexidade ao sistema Pflugler, et al. (2016) [12].

### **2.2.2. SISTEMA BASEADO EM DADOS HISTÓRICOS**

Os sistemas baseados em dados históricos armazenados são mais econômicos e não dependem de uma base de dados de usuários. Este tipo de sistema, quando aplicado com dados suficientes, pode cobrir períodos sazonais ao longo de um ano, como estações do ano, feriados, período de férias, por exemplo Tilahun e Di Marzo (2017) [14].

Ao considerar os métodos de previsão baseados em dados históricos, estes tem-se mostrado muito úteis mesmo quando sistemas em tempo real (com sensores) estão disponíveis Caicedo (2009) [13].

Usando apenas dados históricos Liu, Guan, et al. (2010) [15] desenvolveram um modelo de previsão para ocupação de vagas de estacionamento utilizando métodos de previsão com “*chaotic time series*”. Este modelo foi testado com sucesso na cidade de Pequim, China.

Ter acesso a dados históricos é realmente importante quando se lida com este tipo de problemas, sendo mais fácil de monitorar e armazenar dados de estacionamentos fechados do que o estacionamento de rua por exemplo. Mesmo nos parques fechados, sem entrada de controle, as informações precisam ser registradas e isso pode ser feito com sensores.

É possível construir modelos de previsão a longo prazo usando somente dados históricos. O grande desafio neste caso é obter o máximo de precisão possível utilizando o menor conjunto de dados possível, para que os modelos sejam replicáveis e retrabalhados com novos dados Richter et al. (2014) [16].

De acordo com Klappenecker et al. (2014) [17], outra solução mais simples é o de monitorar o fluxo de entrada e saída no estacionamento, ao invés de monitorar cada vaga de estacionamento. Desta forma, o parque de estacionamento está sempre monitorado com

o número exato de carros no estacionamento. Entretanto, este tipo de monitoramento de focado nos carros não será capaz de avaliar se uma determinada vaga está, ou não ocupada, o que é um dos objetivos deste trabalho.

Em alguns estudos como em Karlaftis e Vlahogianni (2011) [23] foram utilizados seis meses de dados históricos, e em Pflügler et al. (2016) [12] apenas dois meses, revelando ser um curto período de tempo para considerar todos os resultados possíveis e não mostrando o impacto total das características sazonais, como as estações do ano.

Um problema levantado por Bock et al. (2017), [20] a previsão de lugares baseada em dados históricos é dependente da qualidade dos dados utilizados, pois os dados apresentam uma grande quantidade de ruído e variabilidades, como valores ausentes ou pouco confiáveis do ponto de vista estatístico, o que por consequência piora o desempenho do algoritmo, afetando sua precisão, pois é mais complicado treinar um modelo generalizado [16] Richter et al. (2014)

A tabela 1 traz um comparativo de alguns estudos com a metodologia em tempo real, baseada em dados históricos, suas limitações e alguns resultados.

**Tabela 1 - Comparação de estudos com sistemas de aquisição de dados**

<b>Trabalho</b>	<b>Tipo de Dados</b>	<b>Benefícios</b>	<b>Limitações</b>	<b>Resultados</b>
<i>Temporal and Spatial Clustering for a Parking Prediction Service</i> (Richter et al., 2014)[16]	Tempo Real	Exatidão elevada	Custo elevado e manutenção	85% de exatidão
<i>Parking lot occupancy determination from lamp-post camera images</i> (Delibaltov et al., 2013) [11]		Boa exatidão	Iluminação. Captação da imagem pode afetar resultados	76% de exatidão
<i>Unoccupied Parking Space Prediction of Chaotic Time Series</i> (LIU et al., 2010) [15]	Dados históricos	Elevada precisão	Grande quantidade de dados	2,33% de erro nas previsões
<i>A Real-Time Parking Prediction System for Smart Cities</i> (Vlahogianni et al., 2016) [23]		Baixo custo Abrangência temporal	Armazenamento de dados	3,66% de erro nas previsões

### 2.3. TÉCNICAS DE PREVISÃO

Os modelos preditivos são uma forma avançada de prever eventos no futuro com base em dados históricos de eventos do passado, várias técnicas foram aplicadas para prever vagas disponíveis e ou ocupadas.

Na literatura diversos estudos abordam modelos de previsão baseados em técnicas, como *Support Vector Machine* (SVM), *Regression Tree*, ou Cadeia de Markov, que têm sido o padrão de previsão ao longo dos anos por conta de sua simplicidade e fácil compreensão [22] Sargent (2001). Estes algoritmos podem ser rápidos, eficientes e de fácil desenvolvimento, por outro lado podem perder desempenho quando usados com dados muito complexos ou com baixa correlação [23] Karlaftis e Vlahogianni (2011). Em Silva et al. (2022) [54] foram obtidos bons resultados aproximadamente 70% de Exatidão utilizando os algoritmos *Random Forest* e *DT* para classificação de *demand response*.

A literatura sobre lugares de estacionamento disponíveis abrange muitos estudos focados em redes neuronais artificiais, enquanto outros tentam estabelecer uma comparação entre as redes neuronais artificiais e outras técnicas de *machine learning*.

Um trabalho focado em explorar redes neuronais artificiais, *Support Vector Regression* (SVR) e *Regression Tree* para previsão de taxa de ocupação de um estacionamento, considerando um horizonte de 15 minutos e  $K \cdot 15$  minutos (à frente no tempo), onde  $k$  seria o número de “*steps*” à frente de 15 minutos. Este estudo desenvolvido por Zheng et al. (2015) [24] concluiu que a técnica *Regression Tree* tem melhor desempenho nas previsões utilizando menos recursos computacionais em comparação com as redes neuronais artificiais e o SVR.

Uma solução muito útil estudada por [25] Cheira et al. (2016), utiliza regressão para construir um modelo de previsão para verificar ocupação de parques de estacionamento em áreas urbanas, foram realizados diversos métodos de regressão, nomeadamente, Regressão Linear Simples, Regressão Linear Múltipla e SVR. Conseguiram verificar que o desempenho com SVR utilizando *kernel* universal Pearson VII (PUK) melhorou, aumentando a precisão quando mais variáveis foram adicionadas aos dados.

No estudo de [26] Ionita et al. (2018) o objetivo foi criar um perfil de demanda de estacionamento usando o algoritmo de *clustering K-Means* para agrupar áreas, como uma forma de escalar sistemas de previsão onde não há dados de estacionamento. Esta solução revelou-se uma boa opção para reduzir os custos de implementação dos sensores em grandes áreas. Com a ajuda de técnicas de *machine learning*, mais especificamente *Decision trees over Random Forests*, *SVM*, *Multilayer Perceptrons* e *Extreme Gradient Boosting*, os autores foram capazes de prever a ocupação do estacionamento para cada agrupamento com a ajuda do preço médio e da capacidade média por área, e avaliar cada um dos resultados com a ajuda do sistema métrico Raiz quadrada do erro-médio RMSE.

Tamrazian et al. (2015) [27] utilizaram *K-Nearest* para desenvolver um modelo de previsão de estacionamento que foi constantemente atualizado com dados em tempo real. Este algoritmo tem a vantagem de identificar rapidamente as diferentes tendências que a ocupação está seguindo com base nos dados em tempo real que está observando, permitindo previsões mais precisas. Também tem a vantagem de ser eficiente, pois não é necessário nenhum treinamento. O modelo foi testado em dois estacionamentos da Universidade de Stanford, Califórnia, e forneceu previsões precisas com uma diminuição média e máxima das taxas de erro à medida que o tempo avançava.

Para Pflügler et al. (2016) [12] o desenvolvimento de um sistema para a previsão da situação do estacionamento utilizando apenas os dados publicamente disponíveis do sistema de estacionamento reduzindo assim os custos, tais como a necessidade de implementar sensores nos parques, ao mesmo tempo em que identifica fontes de dados importantes para ajudar no processo de previsão. Foi implementado um modelo baseado em redes neurais artificiais atingindo um quadrado médio de 0,16321. O estudo também concluiu que o dia da semana, local, temperatura e a hora do dia melhoram e aumentam a precisão da previsão, enquanto o tráfego, feriados, eventos e chuvas tem uma relevância secundária.

O foco de [29] Ramos Silva (2017) foi utilizar dados que cobriram um período total de 3 meses de dados registrados por parquímetros nas ruas de duas zonas da área central de Lisboa, em universidades, residências e comércio. Foi então desenvolvido um classificador que indica a situação do estacionamento. Durante este estudo foram testados os algoritmos J48, *Random Forest* e *Perceptron* de várias camadas, adicionando algumas variáveis qualitativas, como, horas de aula, condições climáticas, feriados, férias, eventos especiais e aleatórios. Os melhores resultados foram atingidos utilizando *Random Forest*.

**Tabela 2 - Comparação de estudos com diferentes técnicas de previsão**

<b>Trabalho</b>	<b>Técnica</b>	<b>Limitações</b>	<b>Resultados</b>
<i>Parking availability prediction for sensor-enabled car parks in smart cities</i> (Zheng et al., 2015) [24]	<i>Regression Tree</i>	Desempenho cai com mais etapas de previsão	MAEP = 5.7%
<i>Parking availability prediction for sensorenabled car parks in smart cities</i> (Zheng et al., 2015) [24]	Redes Neurais Artificiais	Desempenho inferior com correlação interna dos atributos de entrada do modelo	Mean MSE - 0.054 Mean MAE - 0.194 Mean R <sup>2</sup> - 0.059
<i>ParkGauge: Gauging the occupancy of parking garages with crowdsensed parking characteristics</i> (Cherian et al., 2016) [25]	SVR com PUK	Resultados são sensíveis ao tipo de Kernel utilizado	RSME= 0.0993
<i>Predicting Space Occupancy for Street Paid Parking</i> (Ramos Silva, 2017) [29]	<i>Random Forest</i>	Registro dos veículos e capacidade de processamento	Precisão 70% 75%

## 2.4. MACHINE LEARNING

*Machine learning*, tem como princípio no uso da lógica para modelar o planejamento, o raciocínio e a resolução de problemas humanos e por consequência da complexidade dos problemas e da sociedade em que vivemos.

Estamos hoje no domínio dos grandes dados, do volume massivo de informação e na era do *Big Data*. Em cada momento é maior a demanda e a produção de dados, dados estes que no enquadramento de soluções dinâmicas requerem explorar técnicas, o mais precisas possível, como métodos de análise de dados, exploração de dados e consequentemente a

previsão de dados e como resultante destes métodos surgem as técnicas de *machine learning*.

De uma forma resumida pode-se classificar *machine learning* como um conjunto de métodos que conseguem detetar padrões automaticamente através de dados e, conseqüentemente, utilizarem tais padrões em previsões, auxiliando o processo de tomada de decisão quando existe incerteza Murphy (2012) [18].

*Machine learning*, pode ainda ser definido como um campo de estudo que permite a aprendizagem sem uma programação explícita Aurelien Geron (2017) [32]. Ao adentrar em uma visão mais focada na engenharia Mitchell (1997) [33] afirma que um programa de computador com uma determinada experiência “*E*”, em face de uma classe de tarefas “*T*” e uma medida de desempenho “*D*”, onde se verifica uma melhora com a tarefa “*T*”, no desempenho “*D*”, resultando, assim, uma melhora na experiência “*E*”, ou seja, a experiência pode ser melhorada através através de uma tarefa. Ainda se pode acrescentar que a tarefa em si não traduz o processo de aprendizagem, mas sim a capacidade de executar tal tarefa. Por exemplo, se quiser que um computador execute um sorteio de um número aleatório o sorteio é a tarefa e a aprendizagem será o método escolhido para realização deste sorteio.

Ainda de acordo com a literatura, *machine learning* pode ser caracterizado de três formas distintas Sutton e Barto (2018) [34]:

1. Supervised Learning;
2. Unsupervised Learning;
3. Reinforcement Learning.

• *Supervised Learning*: A aprendizagem neste método consiste em dados explicitamente marcados, onde para cada entrada é conhecida uma saída desejada, ou seja, o algoritmo tem um grupo de dados com as respectivas entradas e saídas corretas correspondentes, onde o processo de aprendizagem será realizado através de comparação e procura de erros. Ao fim de ‘*n*’ iterações no processo é esperado que o modelo consiga ser capaz de classificar novos dados sem marcação. Cada exemplo é, essencialmente, a descrição de uma situação e conseqüente ação a seguir pelo sistema, sendo apropriada para aplicações de classificação [34].

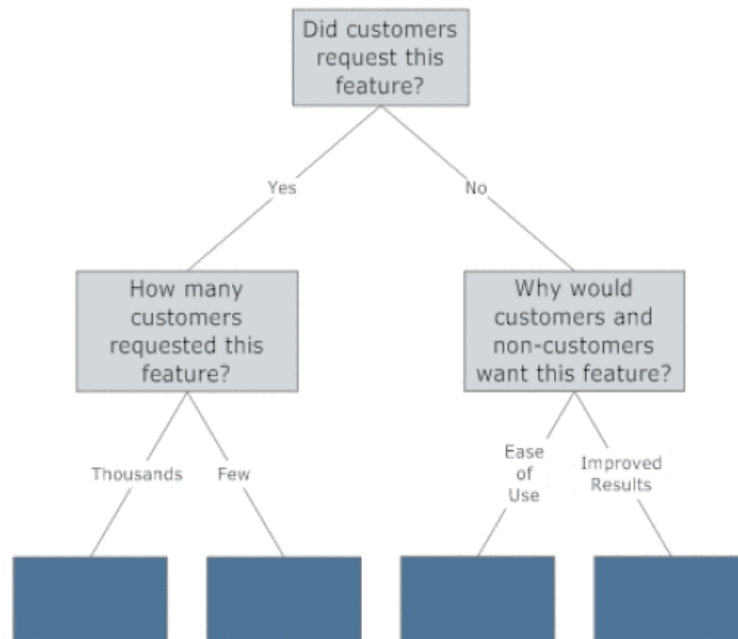
- *Unsupervised Learning*: É o tipo de aprendizagem que não rotula (marcação) os dados, ou seja, o algoritmo não deixa explícito qual é saída correta, sendo assim esperado que o algoritmo a deduza. Este é um tipo de aprendizagem amplamente utilizada para efeito de *clustering* e na identificação de *outliers* em dados. Esta aprendizagem difere da aprendizagem do tipo *supervised learning*, pois nesta não existe diferença entre dados de treino e dados de teste, espera-se que o modelo seja capaz de impor o padrão dos dados de treino e aplicá-lo corretamente para classificar os dados de teste Ben e Shai (2014) [35] Lecun et al. (2015) [36].

- *Reinforcement Learning*: Este tipo de aprendizagem, se baseia na aprendizagem do tipo tentativa e erro com base em recompensas, de forma análoga à de algumas espécies animais. Sutton e Barto (2012) [37] definem que é possível identificar quatro subelementos principais em um *Reinforcement Learning*: uma política, um sinal de recompensa, uma função de valor, e opcionalmente, um modelo de ambiente. Sendo muito utilizada em técnicas de jogos, como *gamefication* e robótica, por exemplo. O maior conjunto de recompensas resultantes de ação não explícita resulta do *feedback* de um vasto conjunto de diferentes ações [34].

#### **2.4.1. ALGORITMOS**

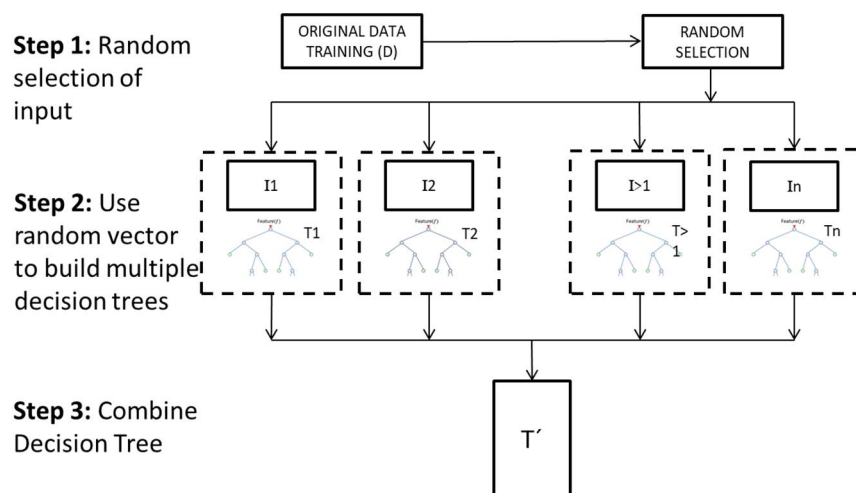
Os algoritmos em *machine learning* são a lógica por trás dos tipos de aprendizagem exemplificados anteriormente e apresentam-se alguns dos algoritmos mais utilizados.

*DT*: é um algoritmo que utiliza grafo em seus modelos de solução, e estes grafos são utilizados para problemas de classificação. *DT* é um dos algoritmos mais úteis e poderosos na mineração de dados. Este algoritmo é capaz de lidar com um número considerável de dados de entrada, como nominal, numérico e alfabético e este é o grande benefício deste algoritmo Rokach et al. (2018) [38]. Este algoritmo permite abordar um problema de uma forma sistemática e estruturada para resolvê-lo de forma lógica.



**Figura 1 - Exemplo de uma DT [48]**

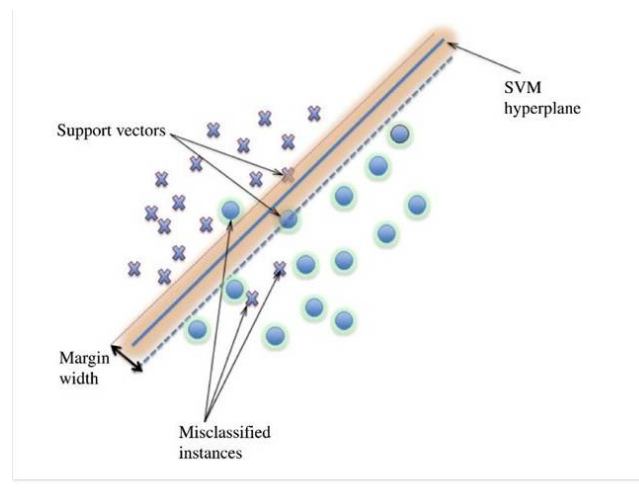
*Extra Tree*: é um algoritmo que tenta encaixar árvores de decisão aleatórias em diferentes amostras do conjunto de dados e usa o conceito de média para aumentar a precisão e também para controlar o ajuste dos dados. Este algoritmo difere das árvores de decisão clássicas na forma como são construídas. Em vez de procurar a melhor divisão para a separação das amostras de um nó em dois grupos, as divisões são aleatórias e desenhadas para cada uma das características seleccionadas aleatoriamente e a melhor divisão entre elas é escolhida Geurts et al. (2006) [49].



**Figura 2 - Representação visual de uma *Extra Tree***

*Logistic Regression*: este algoritmo de classificação, que tem como base observar um conjunto de variáveis independentes e estimar valores discretos para este conjunto. Utiliza-se distribuição logística cumulativa através de uma função logística, para estimar probabilidades e a relação entre variável dependente e categórica. Este algoritmo pode ser utilizado para previsões com certa confiabilidade.

SVM: é um algoritmo binário de classificação, Em Duin e Tax (1999) [39] foi utilizado um algoritmo de vetor de apoio para caracterizar o suporte de uma distribuição de alta dimensão. Neste algoritmo um conjunto de dados atribuído é em um vetor no lugar  $N$  dimensional, para a partir daí o SVM gerar um hiper-plano ( $N-1$ ), onde  $n$  é número de atributos que vai segregar os dados em grupos, por exemplo, em uma distribuição de dados de dois tipos, o SVM encontrará uma linha que separa estes dados em dois grupos distintos. Este algoritmo é muito utilizado para detecção de imagens, classificação de imagem e reconhecimento facial.



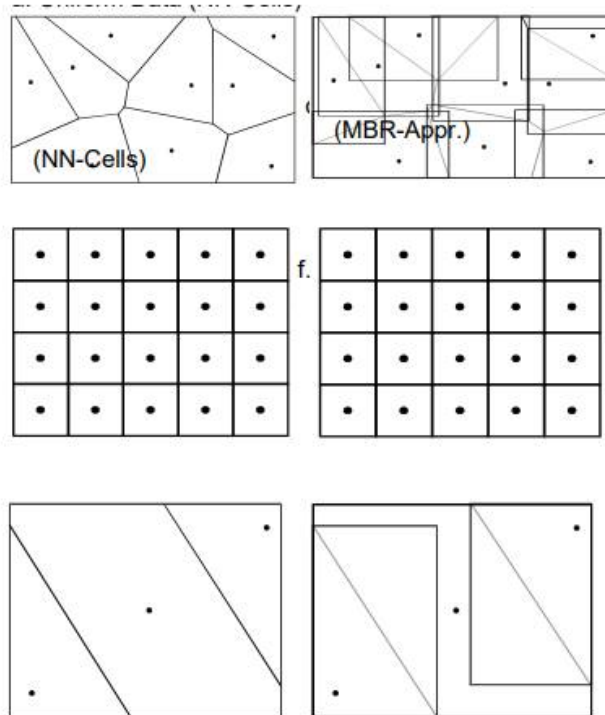
**Figura 3 – Exemplo de uma SVM [21]**

*Naive Bayes*: baseado no teorema de *Bayes*, o algoritmo utiliza uma família de classificadores probabilísticos, como é possível observar na equação 1 e equação 2, sendo um teorema com forte independência entre as características. *Naive Bayes* tem-se mostrado eficaz em muitos aspectos práticos aplicações, incluindo classificação de texto, diagnóstico médico e gestão de desempenho de sistemas Domingos e Pazzani (1997)[43].

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \quad (1)$$

$$P(B|X) = P(x_1|B) \times P(x_2|B) \times \dots \times P(x_n|B) \times P(B) \quad (2)$$

KNN: um algoritmo classificador onde o aprendizado baseado na similaridade que um vetor tem com o outro, neste algoritmo os vetores são classificados e armazenados e os novos vetores (não classificados), são comparados com todos os vetores anteriormente armazenados e classificados. A técnica baseada na estrutura KNN utiliza estruturas em árvore para representar os conjuntos de dados de treinamento. Berchtold propôs um método sobre células de Voronoi que são construídas especificamente para as consultas dos vizinhos mais próximos Berchtold et al. (1998) [40].



**Figura 4- Exemplo de KNN com células de aproximação [40]**

*Random Forest:* Nada mais é do que um conjunto de árvores de decisão que forma a “floresta aleatória” em tradução livre. Para a classificação este algoritmo utiliza atributos para cada classe, com uma estrutura similar a um fluxograma, com “nós” onde uma condição é verificada, e se atendida o fluxo segue por um ramo, caso contrário por outro, sendo assim a comparação que tem nó/classe com mais pontos é escolhida. Um vetor aleatório é gerado independentemente dos vetores aleatórios passados com a mesma

distribuição e uma árvore é gerada usando o teste de treinamento Breinan (2001) [41]. Algoritmo muito utilizado para detecção de fraudes, classificação genética.

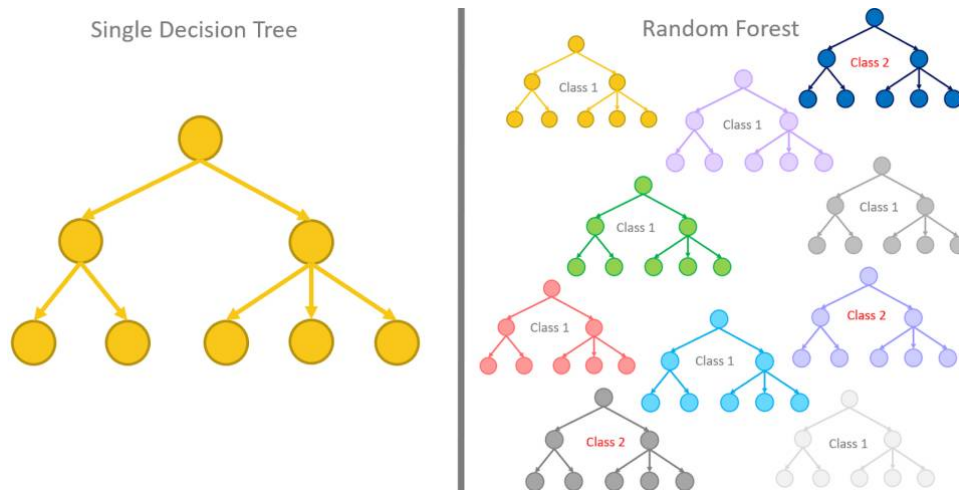


Figura 5 - Exemplo de uma Random Forest [21]

## 2.5. CONCLUSÕES

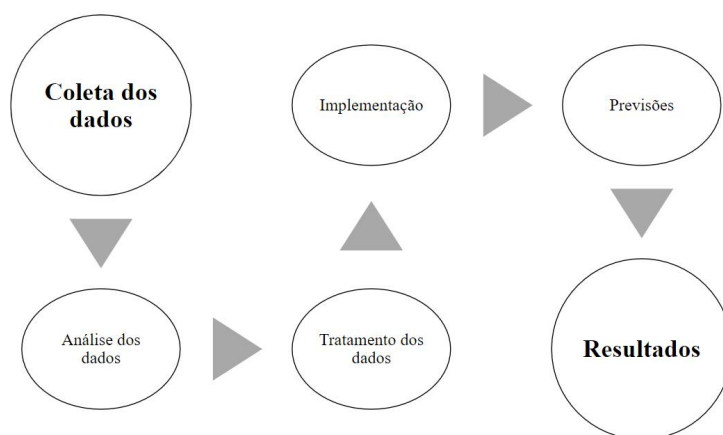
Neste Capítulo foi abordado, um panorama dos edifícios inteligentes, sistemas de avaliação de ocupação de parque de estacionamento e as técnicas de *machine learning* para previsão, no que tange a ocupação de parques de estacionamento no âmbito urbano, bem como os seus pontos mais relevantes.

É possível destacar que para os dados históricos que são os dados tratados neste trabalho, as técnicas mais utilizadas são técnicas de regressão considerando uma quantidade significativa de variáveis que contribuem para resolução do problema.

O próximo capítulo apresenta a metodologia do trabalho acerca da aquisição, exploração, tratamento e limitação dos dados com base nos trabalhos anteriores.

### 3. METODOLOGIA DO TRABALHO

Este capítulo descreve o problema e analisa os dados que foram obtidos através de observação empírica, de forma a simular um sensor típico de estacionamento que posteriormente foi incrementada com variáveis qualitativas como estado de carga da bateria e capacidade de armazenamento da bateria e por fim realizada a extrapolação, considerando o universo de um ano para análise. Após a extrapolação é realizado um trabalho de análise estatística e exploração dos dados, com o intuito de descobrir eventuais correlações entre os dados e também fomentar as hipóteses de trabalho. Essa análise/exploração é então utilizada para treinar os algoritmos propostos. A seguir na figura 6 é detalhada cada etapa.



**Figura 6 - Esquema ilustrador do fluxo de trabalho do projeto**

### **3.1. BASE DE DADOS**

Os dados utilizados neste trabalho são provenientes dos estacionamentos residenciais, e neste caso há dois conjuntos de dados no formato de Excel que foram nomeados como cave e subcave. Os dados foram reunidos a cada 30 minutos entre 1 de janeiro de 2018 e 31 de dezembro de 2018. A cave é de menor dimensão, com um total de 18 vagas de estacionamento, e a subcave tem 23 vagas de estacionamento. Todas as vagas dos estacionamentos são subterrâneas, considerando uma vaga para cada apartamento.

Os dados foram apresentados no formato de Excel composto inicialmente por abas representando cada lugar do estacionamento e cada aba com oito colunas, com data, hora, respectivo dia da semana e informação sobre o veículo a situação do veículo no estacionamento, ou seja, 1 para veículo estacionado e 0 para veículo fora.

Diversas ferramentas foram utilizadas em conjunto, tais como, Microsoft Excel e a linguagem de programação Python, utilizando bibliotecas tais como Pandas, Numpy, Seaborn.

O processo começou com a transformação dos dados com fusão inicial de todos os conjuntos de dados previamente referenciados pela coluna de data/hora que, trouxe a análise de todos os dados. O processo de fusão resultou em um novo conjunto de dados.

Procurou-se, então, fazer o tratamento dos dados para adequá-los para a preparação do modelo, o primeiro tratamento realizado foi a formatação das colunas para que fosse mais claro e mais simples de analisar os respectivos dados.

Na etapa seguinte foram criadas novas colunas, ou seja, “semana” que consistia nas semanas do ano, “soc” que indica o estado de carga da bateria e “capacidade” que mostra a capacidade da bateria dos veículos elétricos.

Depois disso, as informações derivadas da data foram adicionadas, a saber, duas colunas, “dia” que indica os dias do ano e dia semana\_n dando informações do dia atual da semana, onde o valor 1 representa o domingo, 2 representa a segunda feira, e assim por diante, até sábado que tem valor 7 e, posteriormente, foi realizada a extrapolação dos dados replicando para preencher todo o ano de 2018, considerando os feriados nacionais do ano de 2018 conforme tabela 3.

**Tabela 3 - Feriados Nacionais em 2018**

<b>Data</b>	<b>Dia da Semana</b>	<b>Feriado Nacional</b>
1 de janeiro	segunda-feira	Dia de Ano Novo
30 de março	sexta-feira	Sexta-feira Santa
01 de abril	domingo	Páscoa
25 de abril	quarta-feira	Dia da Liberdade
01 de maio	terça-feira	Dia do Trabalhador
10 de junho	domingo	Dia de Portugal
31 de Maio	quinta-feira	Corpo de Deus
15 de agosto	quarta-feira	Assunção de N.a Senhora
5 de outubro	sexta-feira	Implantação da República
01 de novembro	quinta-feira	Dia de Todos os Santos
01 de dezembro	sábado	Restauração de Independência
8 de dezembro	sábado	Dia da Imaculada Conceição
25 de dezembro	terça-feira	Natal

Foi definido um período de férias entre 01 de agosto de 2018 e 2 de setembro de 2018 onde normalmente o perfil de utilização é alterado consoante as ao período. Foram, também, considerados datas das festas de dezembro e dos feriados nacionais conforme tabela 3 onde o comportamento do uso do estacionamento é afetado. Assim, após a inserção das variáveis qualitativas os dados têm a estrutura demonstrada a seguir na tabela 4.

**Tabela 4 – Estrutura dos dados (cave)**

	Lugar	Semana	dia semana n	dia da semana	dia	hora	estacionado	soc	capacidade kWh
0	A1	1	2	segunda-feira	01/01/2018	00:10:00	0	30.0	50
1	A1	1	2	segunda-feira	01/01/2018	00:30:00	0	30.0	50
2	A1	1	2	segunda-feira	01/01/2018	01:00:00	0	30.0	50
3	A1	1	2	segunda-feira	01/01/2018	01:30:00	0	40.0	50
...	...	...	...	...	...	...	...	...	...
315356	A18	53	2	segunda-feira	31/12/2018	22:00:00	0	100.0	60
315357	A18	53	2	segunda-feira	31/12/2018	22:30:00	0	100.0	60
315358	A18	53	2	segunda-feira	31/12/2018	23:00:00	0	100.0	60

315359 linhas × 9 colunas

### 3.1.1. LIMITAÇÃO DOS DADOS

É importante observar que os dados e atributos dos dados têm algumas limitações nas informações, por tratar-se de observações que por sua vez simulam um sensor. Estas observações podem conter erros. Observando o conjunto de dados é visível que um sensor avariado ou uma leitura equivocada pode indicar ruído ou anomalias na leitura dos dados.

Esta leitura dos dados ocorreu durante o período de uma semana, considerando os dias da semana e o fim de semana de forma igual, os períodos de férias ou feriados são assumidos de forma aleatória, porém considerando que alguns veículos permanecem mais tempo estacionados e outros mais tempo fora. Portanto não existe registro para todos os veículos estacionados nestes períodos, sendo impossível calcular com exatidão todas as correlações e dados estatísticos para estes períodos.

Uma limitação importante é a pouca variabilidade dos dados, por se tratar de parques de estacionamento residenciais a tendência é que eles se repitam ao longo dos dias, semanas e meses.

### 3.2. ANÁLISE DOS DADOS

Para esta seção, o foco essencial é realizar uma análise extensiva e estatística dos dados anteriores, a fim de compreender a variação do uso dos estacionamentos e quais tipos de características mais influenciam esse uso.

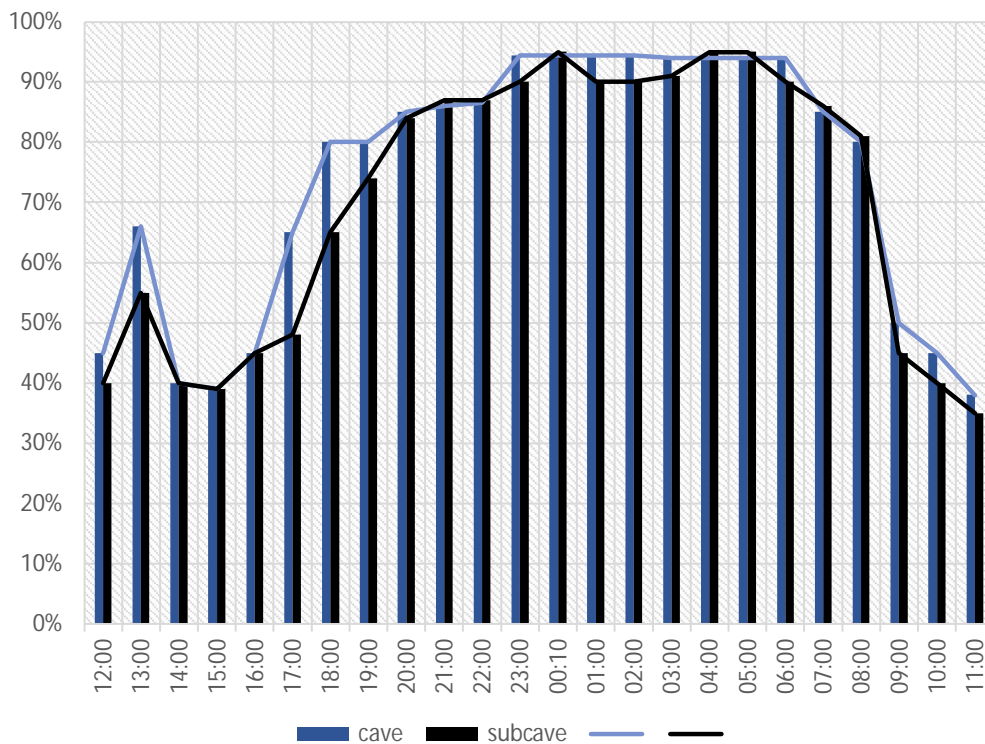
Com o objetivo de descobrir quais são as características mais relevantes e identificar quais delas tem maior influência no uso do estacionamento foram, então, comparadas

graficamente as variáveis binárias, como uma forma de observar e intuitivamente concluir o impacto que estes valores têm. Depois podemos observar a correlação entre estas variáveis e o fato do veículo estar ou não estacionado.

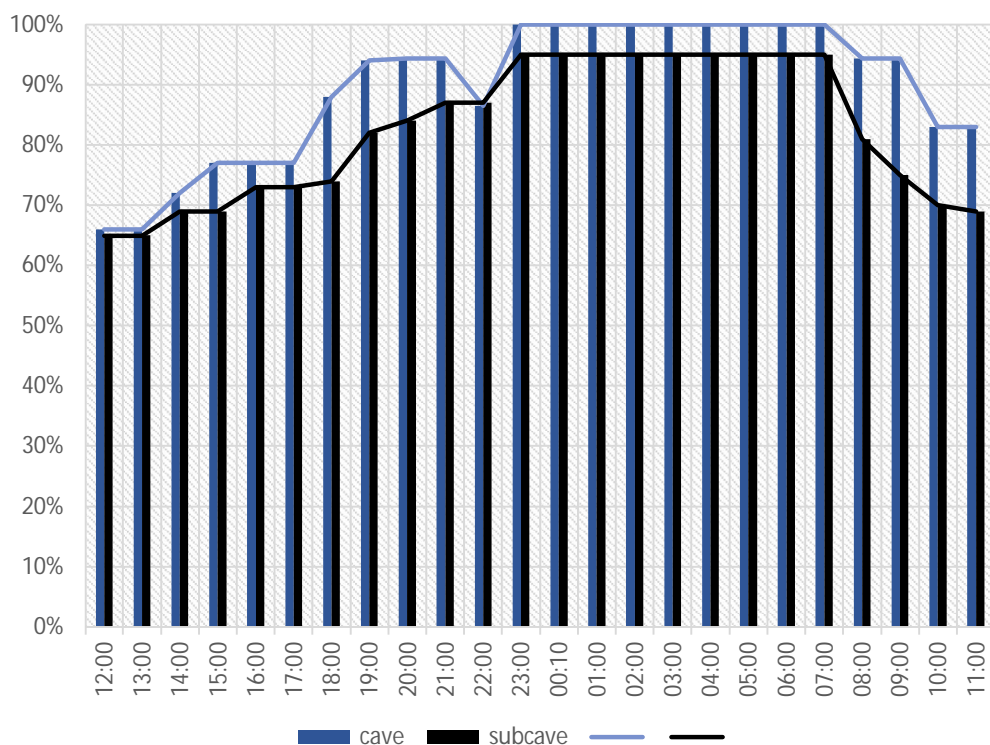
Foi então realizada uma análise estatística dos dados para identificar correlações e como os dados estão distribuídos. Analisada a taxa de ocupação relacionada com o dia da semana, de acordo com a tabela 5 e a hora de acordo com a figura 7, é possível observar que os dias em que os veículos passam mais tempo estacionados são justamente sábado e domingo e a ocupação do estacionamento começa a aumentar a partir das 18 horas e diminuir a partir das 07 horas, com um ligeiro aumento por volta as 12 e 13 horas. Esta é uma análise importante para constatar que a base de dados é constituída de forma realista e também verificar que os dados têm uma distribuição estatística normal.

**Tabela 5 – Taxa de ocupação por dia da semana**

<b>dia da semana</b>	<b>taxa de ocupação (%)</b>	
	cave	subcave
domingo	79,89	77,26
segunda-feira	63,45	60,26
terça-feira	65,49	61,92
quarta-feira	66,23	63,39
quinta-feira	66,67	63,38
sexta-feira	66,02	62,76
sábado	90,26	86,11



**Figura 7- Distribuição média de veículos estacionados x hora do dia**



**Figura 8 - Distribuição média de veículos estacionados x hora do dia feriado 10 de junho**

A ocupação dos parques de estacionamento durante um feriado tem uma mudança significativa, a título de exemplo é possível observar que figura 8 apresenta uma tendência dos veículos estarem mais tempo estacionados em relação a média dos dias de semana. Ainda é possível observar um pico de 100% de ocupação entre as 23:00 e as 07:00 horas. A taxa de ocupação está em torno da marca de 90% para cave e de 83% para subcave. Os níveis de ocupação começam a aumentar para os dois estacionamentos às 23:00 se mantendo até as 07:00.

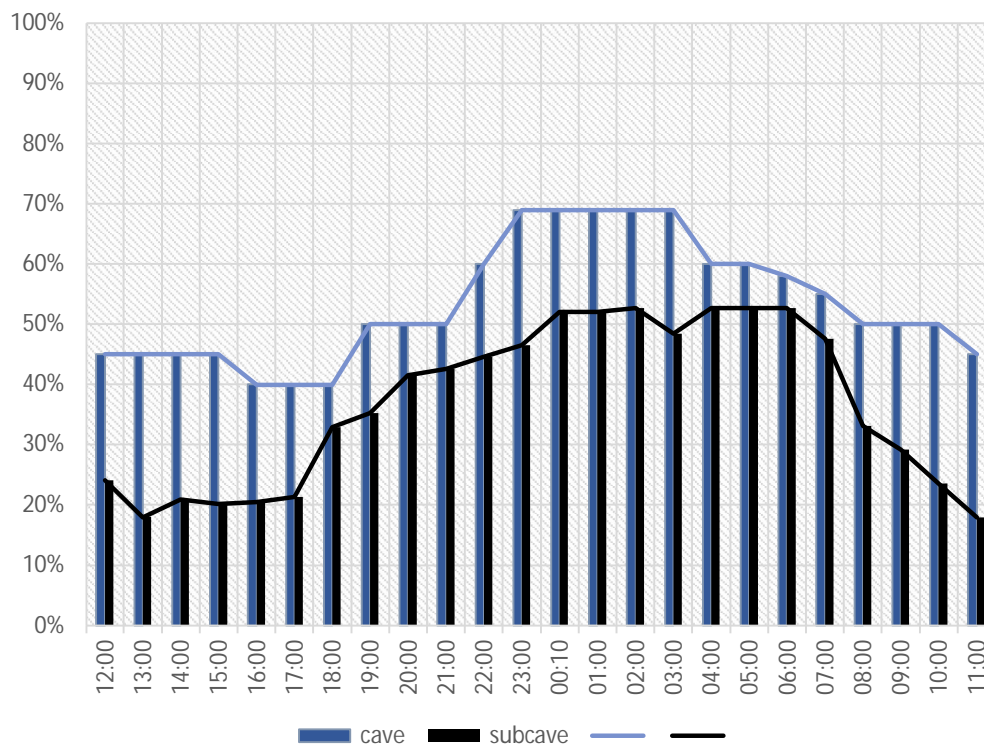


Figura 9 - Distribuição média de veículos estacionados x hora do dia período de férias.

### 3.3. TRATAMENTO DOS DADOS

Para desenvolver este estudo, os dados devem ser transformados em um conjunto de dados em série cronológica com o número de veículos dividido em cave e subcave como anteriormente citado na seção 3.1. Nesta seção, a transformação e manipulação foram realizadas no conjunto de dados previamente identificados.

Inicialmente os dados apresentavam os seguintes tipos de dado para cada coluna, conforme tabela 6, sendo então necessário realizar um tratamento para que todos os dados sejam do tipo “int64” que representam variáveis numéricas.

**Tabela 6 – Tipos de dados por coluna**

<b>variavel</b>	<b>tipo de dado</b>
Lugar	object
Semana	int64
dia semana n	int64
dia da semana	object
dia	datetime64[ns]
hora	object
estacionado	int64
soc	float64
capacidade kWh	int64

Após algumas experiências para avaliar dados e medições erradas e ou em falta, procurou-se métodos para lidar com estes erros em vez de simplesmente remover essas medidas. O primeiro método utilizado por Stolfi et al. (2019) [42] tem como finalidade lidar com medidas erradas consideradas em um determinado dia da semana ( $wm_w$ ). Este método consiste em preencher com a média dos quatro dias das semanas anteriores, quando um dia inteiro estiver errado, ou seja, se o dia 01 de novembro de 2018 (quinta-feira) estiver errado a média dos valores das quatro quintas-feiras anteriores (4, 11, 18 e 25 de outubro) será aplicada, conforme equação 3. Isto se aplica a qualquer coluna dos dados como, por exemplo, soc ou se o veículo esta ou não estacionado.

$$wm_w = \frac{wm_{w-1} + wm_{w-2} + wm_{w-3} + wm_{w-4}}{4}, w \in Weekdays \quad (3)$$

No caso de haver apenas a medida de alguma hora específica errada ou em falta ( $wm_h$ ), a média entre a hora anterior e a posterior é aplicada conforme a equação 4. Este tratamento foi aplicado para todos os casos em que a hora estava em falta. Por exemplo, se em 01 de novembro às 15:00, a medição não foi identificada, por ter um valor ausente, este valor será substituído pelo valor médio entre as 14:00 e às 16:00 horas do respetivo lugar.

$$wm_h = \frac{wm_{h-1} + wm_{h+1}}{2}, h \in Hours \quad (4)$$

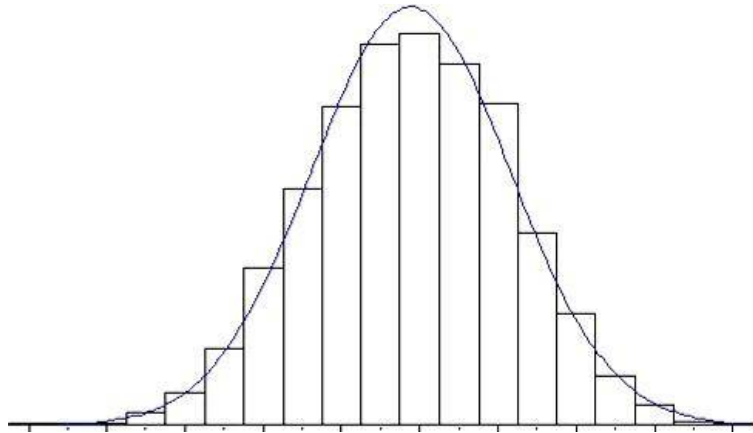
Adotando estas técnicas de tratamento, não houve necessidade de remover nenhum dado inválido ou até mesmo uma linha inteira com dados faltantes, sendo assim, terminando com um total de 315359 e 402957 linhas e 09 colunas para cave e subcave respectivamente.

O objetivo da previsão baseada em valores anteriores é obter a taxa de ocupação por um momento específico em um dia da semana utilizando os valores do mesmo dia da semana anterior ou ainda projetar um horizonte temporal mais dilatado como, por exemplo, considerar a previsão para uma ou mais semanas completas. O resultado da previsão é comparado com o resultado real para verificar sua exatidão. Os resultados são então apresentados em uma matriz de confusão para avaliar a precisão e o intervalo dessa precisão, é medida também a sua exatidão e sensibilidade. Para explorar diferentes possibilidades e verificar como a precisão muda, vários algoritmos são utilizados para obter os valores anteriores.

Assim, avaliou-se que o edifício possui carregadores de 4,4 kW Das et al. [19] e uma gama de veículos considerando de 42 a 100 kWh para a capacidade das baterias. Considerou-se apenas os carregamentos domésticos lentos neste estudo. O carregamento foi residencial, realizado após a chegada em casa, como maioria das entradas, porem considerou-se também carregamentos ao longo do dia. Com o objetivo de observar as diferentes demandas dos lugares de estacionamento/ VE, estabeleceu-se que todos os veículos são elétricos, sendo o VE, com as respectivas capacidades de bateria {42, 50, 60, 75, 100}.

Para ter uma suposição da hora em que os veículos estão estacionados além da observação empírica utilizou-se a distribuição normal cumulativa [30], desta forma foi possível obter o horário de maior procura pelas vagas de estacionamento.

A distribuição normal foi escolhida, pois este tipo de distribuição representa de uma forma mais fiel os acontecimentos da vida real. Com a observação da figura 10 pode-se verificar que o intervalo de tempo representará uma ocupação gradual dos lugares do estacionamento em função do tempo. Montgomery et al. (2002) [31].



**Figura 10 – Distribuição normal**

As horas de maior procura para os parques estacionamento, observadas no secção 3 apresetam uma distribuição normal dada pela equação 5 a probabilidade dada por  $f(X)$ , a média representada por,  $\mu$  é a valor médio dos dados e  $\sigma$  é o desvio padrão dos dados.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

Como é possível observar na figura 9, os dados considerando a frequência de estacionamento pelas horas do dia os dados da nossa base de dado possuem uma distribuição normal.

### **3.4. MODELO FINAL**

Com a análise dos dados foi possível concluir que períodos de férias e feriados tem influencia na ocupação média ao longo do dia dos parques de estacionamento. Características estas que têm grande impacto sobre a ocupação total, o que significa que essas conclusões podem ser consideradas e aplicadas para outros parques de estacionamento com o perfil de utilização e características similares.

Os resultados da análise permitem concluir que o comportamento de ocupação tem uma tendência ao longo do tempo e segue uma rotina com padrão específico de uso. Com isto pode ser possível obter precisão mais alta para os modelos considerando os dados históricos.

A tabela 7 mostra a estrutura dos dados considerada para criação dos modelos com as mesmas 315359 e 402957 linhas para cave e subcave respetivamente, porém agora com 7 colunas.

**Tabela 7 – Base de dados para criação dos modelos**

	Lugar	Semana	dia_semana	dia_code	hora_code	estacionado
0	0	1	2	0	0	0
1	0	1	2	0	0	0
2	0	1	2	0	0	0
3	0	1	2	0	0	0
...	...	...	...	...	...	...
315356	17	53	2	364	48	0
315357	17	53	2	364	48	0
315358	17	53	2	364	48	0

315359 linhas  $\times$  7 colunas

Assim os modelos serão construídos considerando a base de dados de um ano para desenvolver uma previsão precisa, o conjunto de dados deve ter classes equilibradas e fornecer instâncias suficientes para treinar, testar e validar com as conclusões obtidas da Os dados finais são:

- Semana: semana do ano de 1 a 53
- dia\_semana dia da semana de domingo a segunda feira
- estacionado: 0 para VE fora e 1 para estacionado é nosso alvo
- capacidade: capacidade da bateria em kW
- dia\_code: dia de 0 a 364
- hora-code: de 0 a 48 (representando cada 30 minutos)
- Lugar: vaga no parque de estacionamento

### 3.5. CONCLUSÕES

A abordagem deste Capítulo foi apresentar a base de dados, entender e explorar suas limitações, realizar a análise e o tratamento dos dados, com o intuito de obter o modelo final para aplicação no Capítulo 4.

O Capítulo seguinte aborda a implementação do modelo final, por este Capítulo definido e também a forma de divisão dos dados e sua validação e por fim a comparação dos resultados.



# 4. IMPLEMENTAÇÃO

Este capítulo consiste no desenvolvimento e teste de vários modelos preditivos, com os algoritmos descritos no parágrafo de secção 2. São realizados testes de cada modelo de previsão, por sua vez, os resultados obtidos são tratados. Finalmente, o modelo que apresenta os melhores resultados nas métricas de avaliação é escolhido e consequentemente otimizado.

## 4.1. AVALIAÇÃO E ESCOLHA DO MODELO

Construir os modelos de previsão, com o objetivo de prever a ocupação total do estacionamento, dois algoritmos foram escolhidos, nomeadamente o KNN e DT, que também são normalmente usadas para lidar com este tipo de problema.

Os algoritmos previamente identificados, assim como os algoritmos descartados foram aplicados a cada conjunto de dados do estacionamento, estes modelos foram desenvolvidos utilizando os conjuntos de dados contendo todas as variáveis. O objetivo é verificar o desempenho desse modelo e compará-los entre si para que possa ser feita a escolha do modelo mais adequado com as características concluídas no capítulo anterior. Esses conjuntos de dados foram divididos em 70% para treino dos dados, 20% para teste dos dados e 10% dos dados um 5-fold validação cruzada, pois ajuda a evitar o excesso de ajuste Zheng et al. (2015) [24].

A métrica de precisão pode ser calculada através da avaliação da matriz de confusão que correlaciona os valores reais com os valores previstos. A Figura 11 mostra um exemplo de uma matriz de confusão, onde existe o Verdadeiro Positivo (VP), representando os valores positivos corretamente previstos e o Verdadeiro Negativo (VN), representando os valores negativos corretamente previstos, tendo também os Falsos Negativos (FN) e os Falsos Positivos (FP), que representam os erros de previsão. O FN representa o número de valores negativos previstos como positivos e o FP representa o número de valores positivos previstos como negativos.

		Valores Reais	
		Positivo	Negativo
Valores Previstos	Positivo	VP	FP
	Negativo	FN	VN

**Figura 11 – Exemplo de uma matriz de confusão**

Cada um dos componentes mencionados colabora para calcular os valores da exatidão. A métrica de exatidão, é a medida de desempenho mais intuitiva, geralmente representando o desempenho geral do modelo, calculado usando a Equação 6.

$$Exatidão = \frac{VP+VN}{VP+FP+FN+VN} \quad (6)$$

O modelo desenvolvido para avaliação dos algoritmos neste caso é construído com base nos elementos de dados completos. A seguir, os resultados obtidos são mostrados na tabela 8, onde observar-se que a exatidão do algoritmo SVM e *Logistic Regression* apresentam desempenho visivelmente abaixo dos demais algoritmos.

**Tabela 8 - Resultado da Exatidão na primeira avaliação**

Algoritmo	Exatidão	
	Cave	Subcave
SVM	0,66495	0,76731
<i>Random Forest</i>	0,99981	0,99981
<i>Extra Tree</i>	0,93912	0,96309
<i>Logistic Regression</i>	0,64802	0,72222
<i>DT</i>	0,88128	0,94920
<i>Naive Bayes</i>	0,77492	0,87081
<i>KNN</i>	0,82154	0,84031

Para a avaliação da precisão dos algoritmos que é a medida de desempenho que avalia o percentual de preditos positivos corretos dentro de todos os casos de previsão positivos, geralmente representando o desempenho geral do modelo, calculado usando a Equação 7.

$$Precisão = \frac{VP}{VP+FP} \quad (7)$$

Os resultados obtidos e mostrados na tabela 9 ainda com os algoritmos SVM e *Logistic Regression* apresentando um desempenho inferior aos demais algoritmos, porém é possível observar que a precisão do algoritmo *Naive Bayes* também está inferior aos demais algoritmos. Por outro lado, o algoritmo *Random Forest* tem uma precisão de praticamente 100% o que leva atenção para melhor avaliar este algoritmo para compreender se o seu desempenho apresentado é real ou está super adaptado aos dados.

**Tabela 9 - Comparação da Precisão dos Algoritmos**

Algoritmo	Precisão média		Intervalo de Precisão	
	Cave	Subcave	Cave	Subcave
<i>Logistic Regression</i>	67,53%	76,66%	[66,66% ~ 68,40%]	[76,33% ~ 76,99%]
SVM	66,50%	76,73%	[66,44% ~ 66,55%]	[76,64% ~ 76,82%]
<i>Decision Tree</i>	97,73%	98,03%	[96,74% ~ 98,72%]	[97,00% ~ 98,74%]
<i>Random Forest</i>	99,88%	99,98%	[99,67% ~ 100,00%]	[99,88% ~ 100,00%]
<i>Extra Tree</i>	91,75%	95,96%	[88,93% ~ 94,58%]	[94,09% ~ 97,83%]
<i>Naive Bayes</i>	76,76%	84,47%	[75,99% ~ 77,54%]	[83,84% ~ 85,10%]
KNN	90,92%	93,70%	[89,27% ~ 92,57%]	[92,05% ~ 95,34%]

Por último foi avaliada a curva ROC (*Receiver Operating Characteristic*) e a AUC (*Area Under the Curve*), ambas são ferramentas para medir e comparar o desempenho de modelos de classificação, em *machine learning*. A curva ROC indica a relação de perdas e ganhos entre a proporção de resultados verdadeiros positivos (sensibilidade) e a taxa de falsos positivos em diferentes pontos da classificação Nettleman (1988) [28]. O algoritmo *Logistic Regression* apresentou um resultado de AUC inferior aos demais algoritmos com 0,68 e 0,76 para cave e subcave respectivamente. Também foi possível observar que o algoritmo *Random Forest* apresentou um desempenho elevado com valores de 1,0 e 0,99 para cave e subcave respectivamente, o que novamente poderia indicar que o algoritmo estava com *overfitting* dos dados.

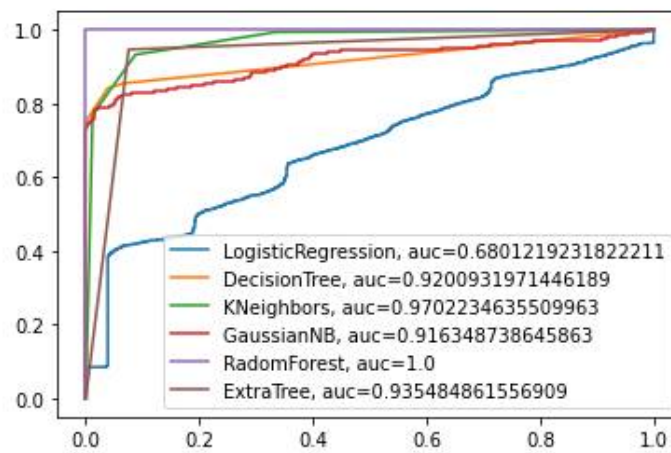


Figura 12 – Curva ROC Cave

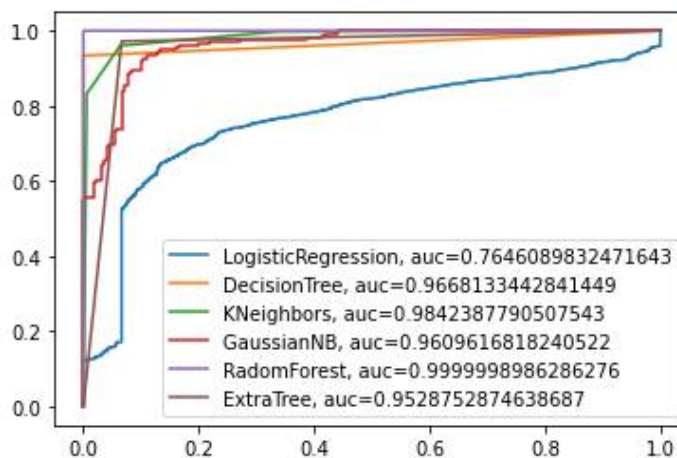


Figura 13 – Curva ROC Subcave

Foi verificado que o algoritmo *Random Forest* apresentou um desempenho extremamente elevado, porem um ponto importante de se observar é verificar se o modelo implementado não apresenta *overfitting* aos dados de entrada, ou seja, se o modelo está apenas decorando os dados e não aprendendo como eles estão distribuídos. Para isso foi realizado uma avaliação do algoritmo *Random Forest* comparando a evolução da exatidão com os dados de treino e os dados de teste, como verifica-se na figura 14 o algoritmo *Random Forest* apresentou *overfitting* aos dados, então foi descartado apesar do elevado desempenho em todas as métricas de avaliação.

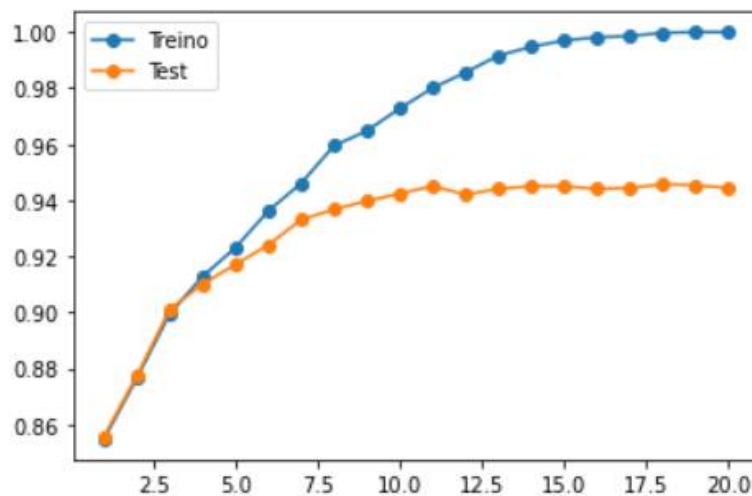


Figura 14 - Comparação de 20 amostras desempenho treino x teste *Random Forest*

Após as avaliações de desempenho decidiu-se seguir com dois algoritmos para realizar a previsão da taxa de ocupação do parque de estacionamento, como explicado no começo do parágrafo são eles KNN e DT, por terem um excelente desempenho nas métricas de avaliação tanto para cave quanto subcave, não apresentarem *overfitting* nos dados de entrada.

## 4.2. DIVISÃO, TREINO, TESTE E VALIDAÇÃO

Esta seção discute os algoritmos selecionados e o tipo de divisão do conjunto de dados usado para treinar e testar e validar o resultado de cada um dos algoritmos.

O objetivo é estabelecer uma linha de base para o desempenho dos algoritmos e os diferentes conjuntos de dados descritos na seção 4.1, todos eles são usados para treinar e testar os modelos usando método cronológico. Para este método, o conjunto de dados é dividido em três partes sendo ela 70% dos dados em treino 20% para teste e 10% para

validação. Desta forma, é possível verificar como o algoritmo funciona realizando previsões ao longo do tempo e prevendo a taxa de ocupação que muda ao longo das horas e também ao longo dos dias.

O treinamento do modelo com diferentes percentagens do conjunto de dados permite verificar como o número de dados inseridos influencia a precisão do modelo. Com um pequeno número de dados, o modelo pode não captar corretamente os padrões de ocupação, como os dados inseridos tem uma variabilidade baixa optou-se por utilizar os dados do ano inteiro para montar os modelos e realizar as previsões.

Após a análise dos resultados das experiências anteriores, os conjuntos de dados que tiveram o melhor desempenho são divididos em conjuntos de treino e teste com os resultados descritos na Tabela 10.

**Tabela 10 – Comparação da precisão dos algoritmos selecionados**

Algoritmo	Precisão média		Intervalo de Precisão	
	Cave	Subcave	Cave	Subcave
DT	92,73%	95,63%	[91,73% ~ 97,72%]	[94,40% ~ 97,96%]
KNN	87,32%	88,00%	[85,87% ~ 89,57%]	[87,09% ~ 90,21%]

### 4.3. COMPARAÇÃO DOS RESULTADOS

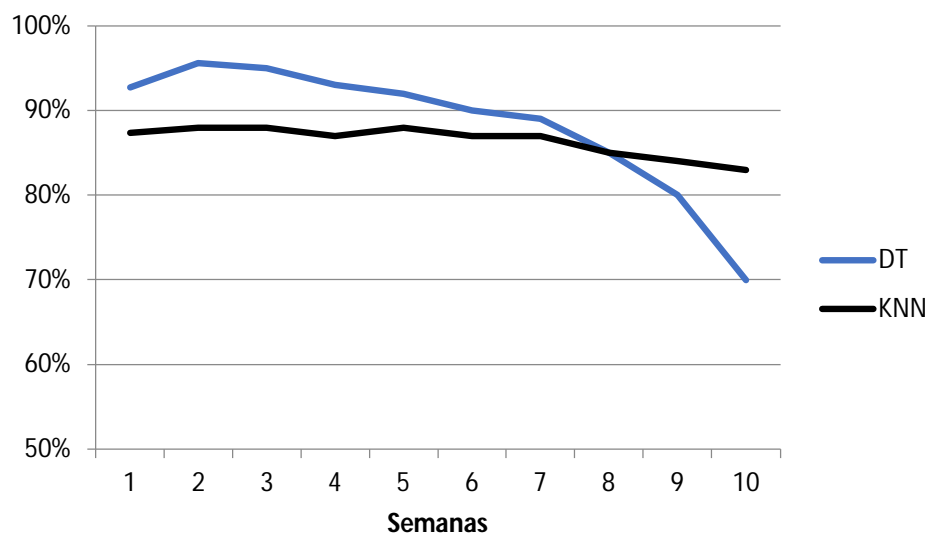
Após as avaliações de desempenho decidiu-se seguir com dois modelos para realizar a previsão da taxa de ocupação do parque de estacionamento, como explicado no começo do parágrafo são eles KNN e DT, por terem um excelente desempenho nas métricas de avaliação tanto para cave quanto subcave, e não apresentarem *overfitting* para os dados de entrada.

O modelo *Logistic Regression*, apresentou o pior desempenho com precisão e exatidão abaixo dos demais modelos, sendo assim desconsiderado para confecção do modelo final.

Ao comparar os resultados do modelo com o algoritmo SVM com o modelo *Naive Bayes*, há melhores resultados no segundo, tanto para a precisão como exatidão e observação da curva ROC. Entretanto, este modelo não é tão robusto e preparado para casos anômalos, como quando é testado com dados dos meses de férias, por exemplo, sendo também descartado nesta etapa.

A tabela 10 mostra que o modelo KNN não alcançou os melhores resultados absolutos de precisão e de exatidão, quando comparado com o DT. O modelo KNN atingiu um valor máximo de precisão de 88% com os dados da subcave e um total de 87,32% de precisão para a cave, já o modelo DT atingiu 95,63% para subcave e 92,73% para cave. Cabe ressaltar que um dos motivos dos possíveis para os valores de precisão serem tão altos nos modelos, pode ser devido à pouca variabilidade dos dados utilizados.

O modelo DT apesar de uma boa e precisão inicial apresentou uma queda no desempenho ao longo das previsões como é possível observar na figura 15, enquanto o modelo KNN, tem um desempenho também decrescente porém mais estável ao longo das previsões.



**Figura 15 – Precisão média ao longo das semanas**

#### **4.4. CONCLUSÕES**

Neste capítulo foi explorado o modelo definido no Capítulo 3, com testes realizados com todos os algoritmos previamente definidos com o objetivo de analisar seus respectivos desempenhos. Após os testes, designadamente KNN, apresentou o desempenho mais estável para previsões em um horizonte temporal mais extenso, de 8 semanas como o explorado neste trabalho.

O próximo Capítulo apresenta o resultado da análise de cenários de divisão de dados com o intuito de encontrar o tipo de divisão em que o modelo tem o melhor desempenho global.

## 5. ANÁLISE DOS RESULTADOS

Este capítulo pretende discutir sobre os resultados e as técnicas utilizadas para avaliar os modelos construídos através dos horizontes de previsão propostos. Assim, os resultados de duas técnicas de separação das amostras de treino e teste são comparados. Para cada técnica proposta, foram efetuadas diversas previsões usando os algoritmos de *machine learning* citados no capítulo 4. A avaliação do desempenho dos modelos é feita através de métodos estatísticos, os quais foram discutidos na seção anterior, as previsões foram realizadas nos dois parques de estacionamento cave e subcave que são mostrados em detalhes nas figuras 16 e 17.

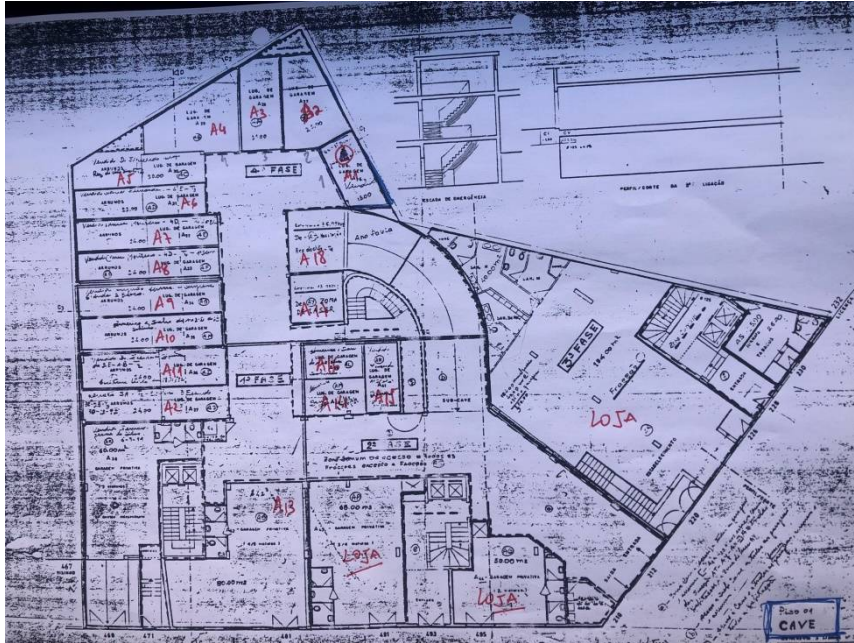


Figura 16 – Lugares Estacionamento Cave

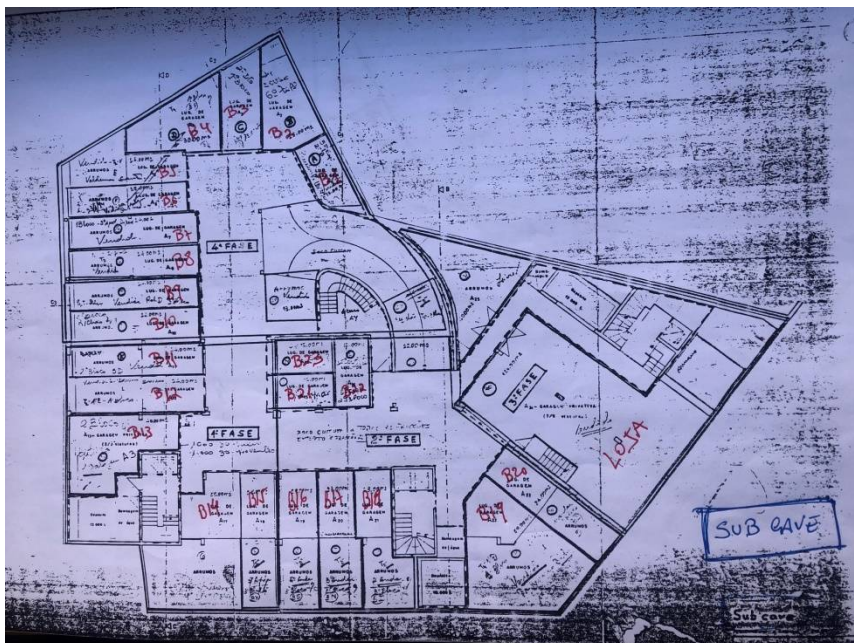


Figura 17 - Lugares Estacionamento Subcave

## 5.1. PREVISÕES

Nesta seção, o foco principal é comparar e discutir como os diferentes modelos funcionam considerando as diferentes divisões do conjunto de dados e diferentes horizontes temporais na construção dos modelos.

Ao comparar os resultados o modelo apresenta melhor desempenho global com o algoritmo KNN com uma precisão elevada sem ocorrência de *overfitting*. Os dois modelos apresentam perda de desempenho ao longo do tempo, porém o modelo DT, mostrou uma maior perda a partir da oitava semana de previsão conforme visto no capítulo 4, seção 4.3.

Dito isto, o estudo continuou com o uso do modelo KNN. Embora este não seja o modelo com os melhores valores de exatidão, demonstrou um desempenho mais estável ao longo das previsões, considerando que as previsões irão ser realizadas de uma em uma semana mantendo assim sua eficiência global ao longo do tempo.

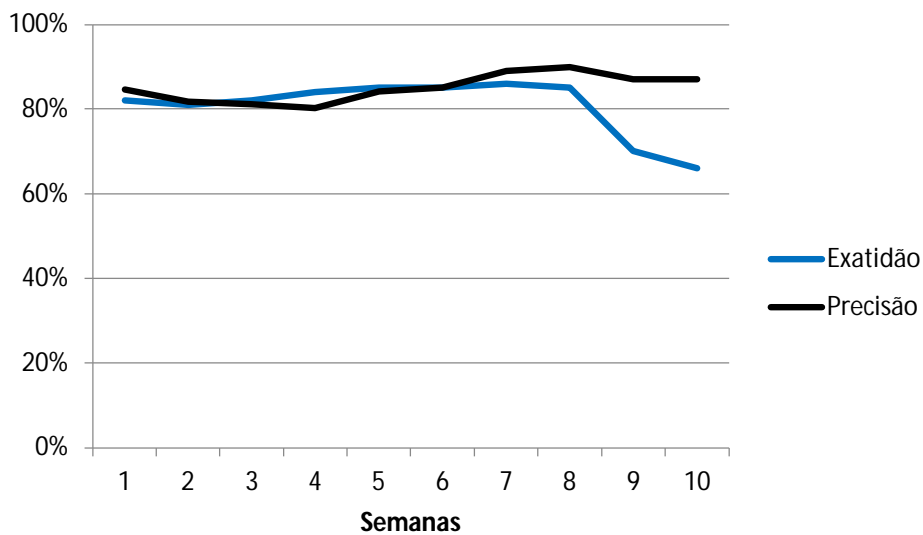
### **5.1.1. PREVISÃO COM DIVISÃO CRONOLÓGICA**

Os modelos desta seção são desenvolvidos utilizando os conjuntos de dados previamente selecionado contendo todos os meses, com divisão cronológica, ou seja, do mês de janeiro até o mês de agosto para treinar o modelo e o teste realizado com os meses de setembro, outubro, metade de novembro e posteriormente validados com a metade de novembro até o final de dezembro. O objetivo é verificar o desempenho desse modelo e compará-lo com os modelos que utilizam outras formas de divisões dos conjuntos de dados, como será estudado nas próximas seções.

Os resultados para cada conjunto de dados cave e subcave são alcançados quando os modelos são treinados com uma ordem cronológica considerando 70% dos dados para treino, 20% para teste e 10% para validação, indicando que esta percentagem oferece uma maior capacidade de generalização no modelo, sendo capaz de se adaptar melhor a diferentes condições. Os resultados quando o modelo é treinado com 80% dos dados mostram que o modelo se ajusta demais aos dados causando assim *overfitting* e respondendo pior as poucas variações presente nos conjunto de dados. Quando o modelo é treinado com 60% dos dados, tem uma queda de rendimento principalmente quando analisados os dados de verdadeiros positivos, conforme tabela 11. Os resultados desta divisão de dados podem ser observados nas tabelas 12 e 13.

**Tabela 11 - Desempenho comparativo da divisão aleatória dos dados**

	<b>Exatidão</b>	<b>Precisão %</b>	<b>Verdadeiros Positivos %</b>	<b>Divisão dos Dados</b>
cave	0,92	98,60	99,49	80%
subcave	0,94	97,50	98,60	
cave	0,82	84,60	74,44	70%
subcave	0,84	82,60	65,65	
cave	0,80	83,20	63,24	60%
subcave	0,84	81,60	56,22	



**Figura 18 – Desempenho KNN**

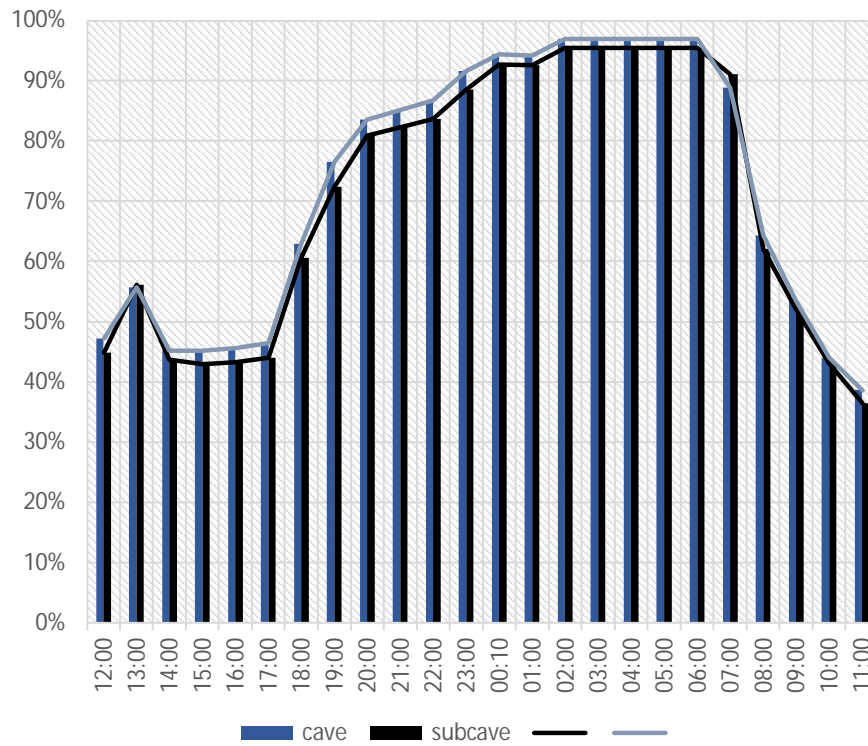
**Tabela 12 - Desempenho após uma semana de previsão**

	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0.83	87.30	71,80
Subcave	0.86	88.00	63,77

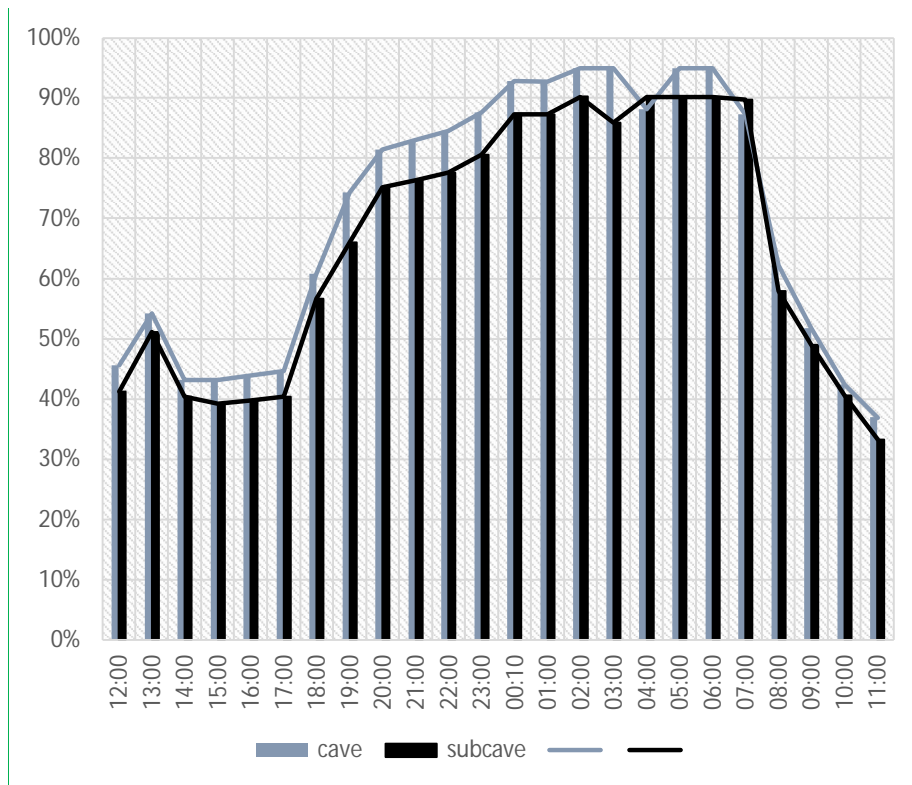
**Tabela 13 - Desempenho após oito semanas de previsão**

	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0,82	82,70	70,50
Subcave	0,87	82,10	62,85

Como comentado na seção 5.1.1 o desempenho do modelo KNN se mantém estável por um período de até oito semanas figura 18 considerando a divisão de treino que estes melhores resultados são obtidos ao treinar os modelos com 70% dos dados, indicando que esta é melhor divisão de dados para obter os melhores resultados do modelo.



**Figura 19 - Distribuição média de veículos estacionados x hora do dia previsão**



**Figura 20 – Distribuição média de veículos estacionados x hora do dia acumulada**

A Figura 19 mostra que os valores mais comuns para a taxa de ocupação dos estacionamentos cave e subcave estão em torno da marca de ocupação de 90%, entre as 22 horas e a 7 horas da manhã. Um pequeno pico pode ser observado por volta das 13 horas em torno dos 50% de níveis de ocupação, isto pode ser explicado pelo período do almoço, onde muitos veículos retornam a casa justamente neste período. Estes valores resultam em um valor médio de ocupação de 73% e 69% para cave e subcave respectivamente.

É possível observar na figura 20 a distribuição da ocupação muito semelhante da figura 19, com uma ocupação em torno da marca de 90% para um período menor de tempo, a partir da 00:10 hora para a cave e a partir de 01:00 para a subcave, também é possível observar um pico por volta as 13 horas. Esta análise resulta em um valor de ocupação médio de 70% e 67% para cave e subcave respectivamente.

### 5.1.2. PREVISÃO COM DIVISÃO ALEATÓRIA DOS DADOS

A seção 5.1.2 com as tabelas 14 e 15 que apresentam o resultado da previsão considerando a divisão aleatória dos dados para a primeira semana e após oito semanas de previsão respectivamente e demonstra que este tipo de divisão utilizando o mesmo modelo da divisão cronológica apresenta um desempenho melhor, mais consistente e mais estável ao longo do

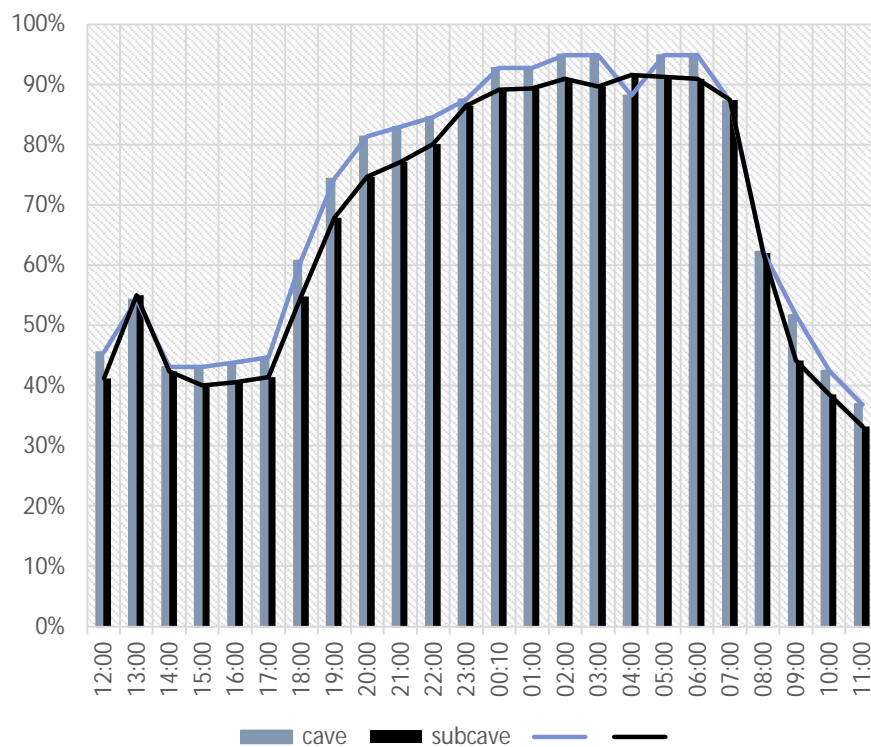
tempo, com precisão mais elevada, acompanhada da precisão e ainda da taxa de verdadeiros positivos.

**Tabela 14 – Desempenho após uma semana de previsão**

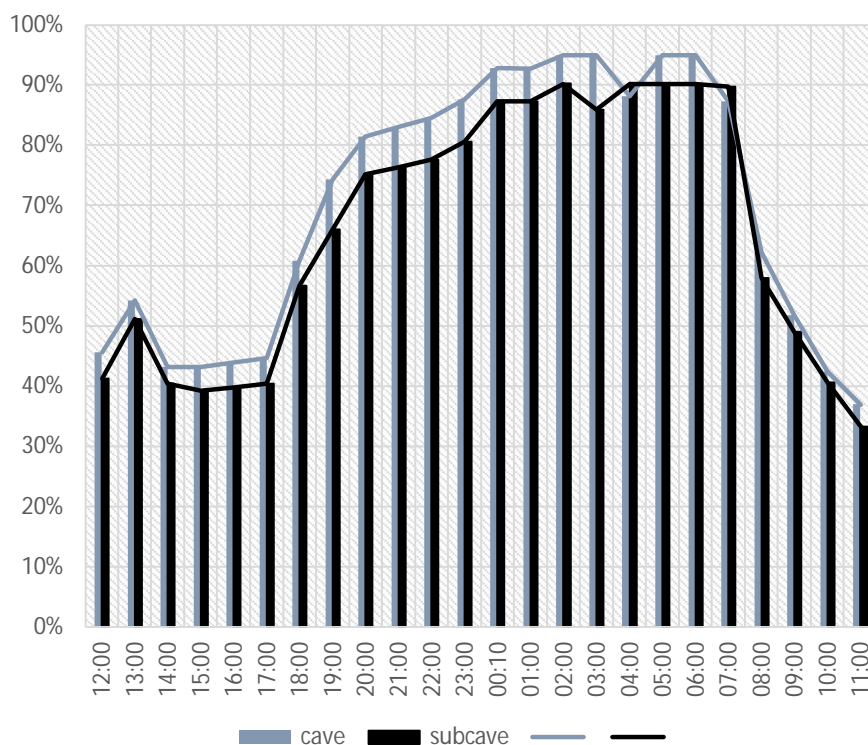
	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0,82	87,60	74,44
Sub cave	0,84	88,60	65,65

**Tabela 15 - Desempenho após oito semanas de previsão**

	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0,81	85,80	71,40
Sub cave	0,86	84,20	63,25



**Figura 21 - Distribuição média de veículos estacionados x hora do dia previsão**



**Figura 22 - Distribuição média de veículos estacionados x hora do dia acumulada**

Pode-se observar na divisão aleatória dos dados que a taxa de ocupação estão mais próximos nas duas formas de análise com uma taxa de ocupação no pico entre as 23 horas e as 7 horas em torno de 90% para cave e 85% para subcave, também é possível observar o mesmo pico de ocupação por volta das 13 horas tanto para a análise de previsão, quanto para análise acumulada. E resulta em uma ocupação média de 69% e 66% para cave e subcave respectivamente. Ainda é possível observar que a taxa de verdadeiros positivos é maior, ou seja, este modelo com a divisão aleatória tem uma quantidade maior de acertos, e uma menor variação da precisão entre a primeira e a oitava semana de previsão.

### 5.1.3. PREVISÃO SEM O PERÍODO DE FÉRIAS E FERIADOS

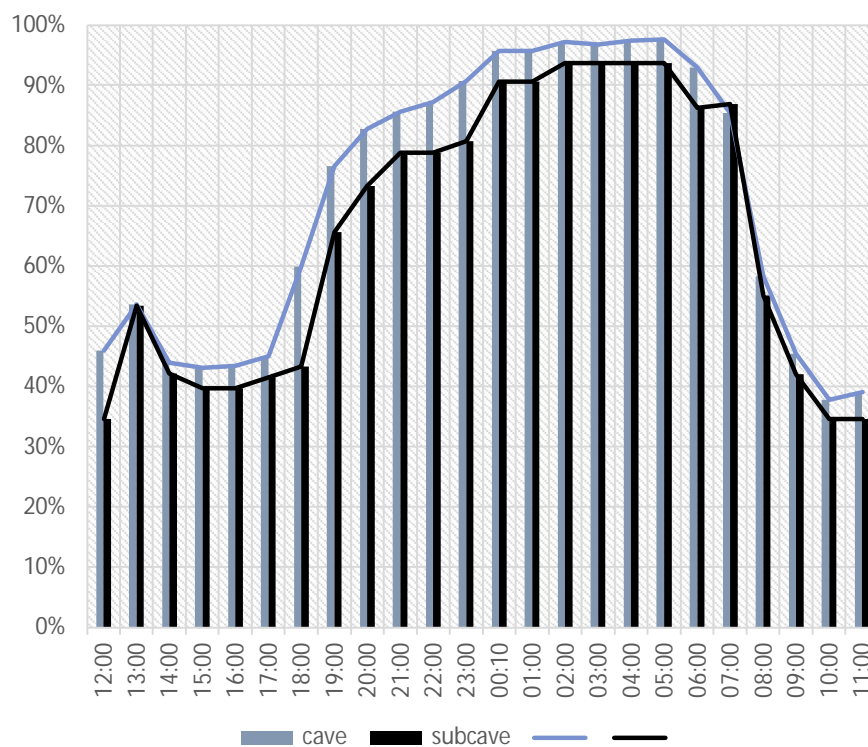
Esta seção apresenta o resultado das experiências em que os modelos são desenvolvidos para previsão sem os meses de férias e feriados, neste caso foram retirados o mes de agosto, as duas últimas semanas de dezembro e todos os feriados mostrados no Capítulo 3. Para este modelo foi utilizado a divisão aleatória, pois a mesma já tinha mostrado mais eficiente conforme análise feita na seção anterior.

**Tabela 16 - Desempenho após uma semana de previsão**

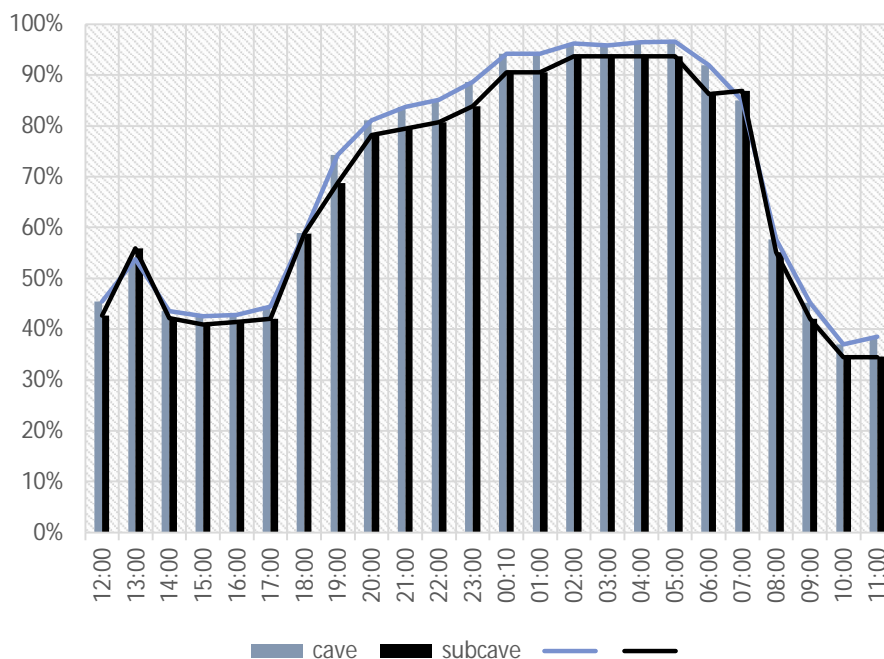
	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0.90	89.50	89,31
Sub cave	0.91	88.10	82,17

**Tabela 17 - Desempenho após oito semana de previsão**

	<b>Exatidão</b>	<b>Precisão (%)</b>	<b>Verdadeiros Positivos (%)</b>
Cave	0,89	89,10	89,21
Sub cave	0,92	88,90	80,45



**Figura 23 – Distribuição média de veículos estacionados x hora do dia previsão**



**Figura 24 - Distribuição média de veículos estacionados x hora do dia acumulada**

As tabelas 16 e 17 mostram os resultados da previsão com a exclusão dos períodos de férias e feriados, o modelo mostrou um ganho de desempenho quando comparado com o modelo utilizando todos os meses. Este resultado pode ser observado por conta da pouca variabilidade dos dados ao longo do ano, o que aponta para a direção da importância deste período para o modelo, pois ele traz a variabilidade para este período e consequentemente deixa o modelo construído mais realista.

As diferenças de comportamento de ocupação entre os meses dificultam este tipo de previsão, sendo isto mais desafiador do que as experiências feitas anteriormente e demonstrando que é difícil capturar os diferentes padrões de ocupação usando um conjunto de dados.

## 5.2. ANÁLISE DAS PREVISÕES

Os resultados obtidos através deste capítulo demonstram que é possível prever a taxa de ocupação do estacionamento dentro de um nível razoável de precisão superior a 70% e 80%, dependendo dos conjuntos de dados utilizado para tal.

Contudo o estudo desenvolvido na seção 4.2, utilizou uma divisão cronológica dos dados para prever a situação de ocupação para a semana atual e depois até sete semanas a frente, quando o modelo ainda mantém um bom desempenho e uma quantidade de verdadeiros positivos considerável. Este método provou ser menos eficaz uma vez que apesar apresentar uma boa precisão e exatidão, comparando com divisão aleatória dos dados que se mostrou mais eficientes.

Estes resultados são explicados pelos valores de ocupação inconsistentes ao longo dos dias e semanas, uma vez que demonstrado na seção 4.5, sendo difícil ter duas unidades de tempo específicas com o mesmo comportamento em dias diferentes.

O modelo com o pior desempenho foi o modelo considerando sem os meses de férias e os feriados, isto é, mesmo considerado um pequeno período em relação a toda a base de dados a sua variabilidade tem um efeito direto no desempenho do modelo, que por sua vez apresenta resultados inferiores quando são feitas previsões, mas em uma maior janela de previsão, neste caso de oito semanas.

A precisão, exatidão e verdadeiros positivos encontrados através dos diferentes experimentos mostrou que o método utilizado pode oferecer resultados razoáveis em diferentes cenários, sendo possível refinar os resultados ajustando o conjunto de dados e os parâmetros do modelo, como janela de previsão e forma de divisão dos dados.



## 6. CONCLUSÕES

O objetivo do presente trabalho foi identificar os perfis de utilização (ocupação) dos VE nos parques de estacionamento do edifício, para este fim foi realizado o estudo e desenvolvimento de um modelo de *machine learning*. Para isso, o primeiro passo foi recolher os dados, na sequência tratá-los e analisá-los, construir o modelo e avaliá-lo. Passo seguinte explorar a literatura e identificar quais técnicas, algoritmos e modelos são utilizados para resolver o problema proposto.

O modelo proposto permite saber a taxa de ocupação dos parques com uma precisão e exatidão acima dos 80% como proposto inicialmente e conseqüentemente viabiliza a melhor gestão da energia do edifício, já que o gestor poderá conhecer o perfil de utilização do parque e ainda a sua ocupação no futuro, considerando um horizonte temporal de oito semanas e assim poderá planear melhor a compra de energia e a gestão energética do edifício no médio prazo. Outra conclusão foi que o sistema proposto pode ser adaptado a outros parques de edifícios ou ainda parque de empresas privadas, revelando ser uma boa solução para facilitar a gestão de energia dos edifícios.

A falta de variáveis qualitativas como dados meteorológicos ou dados econômicos levou a uma simplificação da análise, direcionando o estudo para uma linha do tempo histórica dos parques de estacionamento, o que significa que outras variáveis poderiam ter mais impacto

na formação do modelo e conseqüentemente nas previsões. Os parques de estacionamento foram classificados como estacionamentos residenciais, o que significa que é composto exclusivamente para uso dos habitantes dos edifícios, sendo assim existe uma tendência às pessoas não se mudarem, e independentemente de agentes externos o perfil de utilização permanece constante ao longo do tempo; Em contraste, os meses, dias, horas, dias de semana, feriados, períodos de férias, tem uma importância e altera a taxa de ocupação nos dois parques estacionamentos estudados. Estas conclusões podem ser aplicadas a outros parques que são categorizados da mesma forma, como estacionamentos residenciais.

A escolha do melhor modelo de previsão atribuído para os conjuntos de dados foi feita através da construção e teste de vários tipos de modelos. Para isso, diferentes tratamentos dos dados foram realizados para alcançar a melhor taxa de precisão e conseqüentemente realizar a melhor previsão da taxa de ocupação.

Os modelos testados foram, nomeadamente, um modelo contendo todos os dados dividido em ordem cronológica, um modelo também com todos os dados dividido de forma aleatória e por fim um modelo sem dados de férias e feriados (uma vez que estes dados têm um comportamento diferente e mais complexo dos demais períodos analisados, que foi retirado para analisar seu impacto). Os três modelos foram aplicados com o algoritmo KNN que após as análises prévias mostrou o melhor desempenho para este conjunto de dados e criar o modelo de previsão mais eficiente e preciso.

## **6.1. CONTRIBUTOS**

Os principais contributos desta dissertação incluem a revisão da literatura já produzida sobre edifícios inteligentes e a previsão de ocupação de estacionamentos, sejam estes públicos ou privados, utilizando dados históricos.

Para, além disso, foi possível selecionar o algoritmo de classificação mais apropriado para a base de dados analisada (dados históricos em detrimento de dados em tempo real). Mesmo considerando as limitações de capacidade computacional e de processamento limitada, os resultados foram robustos. Destaca-se ainda a capacidade de tratamento e análise dos dados, mesmo sendo a base de dados inicial limitada. Para isso recorreu-se a vários métodos estatísticos e ferramentas, já reportados na literatura, para o tratamento de dados em falta. Foi assim possível complementar a base de dados inicial e fazer a previsão e validação final do modelo. Os testes estatísticos utilizados, como a exatidão, precisão e

análise da AUC da curva ROC também se revelaram como uma mais valia para escolher, testar e validar o algoritmo selecionado.

Considerando que no futuro, os parques de estacionamento irão corresponder a uma grande parcela dos gastos energéticos dos edifícios inteligentes, o modelo desenvolvido e validado, permitirá a previsão da taxa de ocupação dos parques de estacionamento em médio prazo. Em última análise, o trabalho aqui desenvolvido, irá auxiliar os gestores de energia dos edifícios inteligentes a fazer uma gestão mais eficiente dos recursos energéticos dos mesmos.

## **6.2. TRABALHO FUTURO**

Como dito anteriormente, poderia haver algumas melhorias na base dados, com mais dados qualitativos, ou seja, a inserção de dados socioeconómicos e climáticos, por exemplo, para verificar como essas variáveis se relacionam e como influenciam o modelo, apresentam melhores resultados ou geram mais ruídos. Outra sugestão para um trabalho futuro seria considerar a implementação do modelo em outro conjunto de dados de diferentes categorias como de um parque de estacionamento privado ou empresarial, contribuindo assim para a gestão energética destes tipos de parque de estacionamento.

- Verificar se a construção de um modelo específico para cada parque de estacionamento melhora a sua precisão global e desempenho.
- Desenvolver os modelos utilizando um conjunto de dados com uma unidade de série temporal menor, por exemplo, 10 minutos, e verificar se os resultados são melhores.
- Avaliar se os modelos tem melhor desempenho para o conjunto de dados específico categorizado de férias e feriados.



# 7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] A. HARISSON, E. LOE, J. READ Intelligent Buildings in South East Asia E & FN SPON, London, 1998
- [2] H. ARKIN, M. PACIUK Evaluating intelligent building according to level of service system integration Automation in Construction, 6 pp. 471-479 1997
- [3] J. YANG, H. PENG Decision support to the application of intelligent building technologies Renewable Energy, 22 ,pp. 67-77 2001
- [4] A. L. PEREIRA, J. F. FARDIN, and L. F. ENCARNAÇÃO, “Electric vehicles as motivators for smart grids,” SBSE 2018 - 7th Brazilian Electr. Syst. Symp., pp. 1–6, 2018 doi: 10.1109/SBSE.2018.8395791.
- [5] S. RAJAKARUNA, F. SHAHNIA, and A. GHOSH, "Plug In Electric Vehicles in Smart Grids Charging Strategies". Springer, 2015 <http://dx.doi.org/10.1007/978-981-287-317-0>
- [6] M. RAHMANI-Andebili Vehicle-for-grid (VfG): a mobile energy storage in smart grid IET Gener Transm Distrib, 13 (8) (2019), pp. 1358-1368

- [7] ZHOU, Y., CAO, S., HENSEN, J. L. M., & LUND, P. D. (2019). Energy integration and interaction between buildings and vehicles: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 114, 109337. <https://doi.org/10.1016/j.rser.2019.109337>
- [8] KLAPPENNECKER, A., LEE, H., e WELCH, J. L. (2014). Finding available parking spaces made easy. *Ad Hoc Networks*, 12 (1), 243–249. doi: 10.1016/j.adhoc.2012.03 .002
- [9] RONG, Y., XU, Z., YAN, R., e MA, X. (2018). Du-parking: Spatio-temporal big data tells you realtime parking availability. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , 646– 654. doi: 10.1145/3219819.3219876.
- [10] AMATO, G., CARRARA, F., FALCHI, F., GENNARO, C., MEGHINI, C., & VAIRO, C. (2017). Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72 , 327–334. doi: 10.1016/j.eswa.2016.10.055
- [11] DELIBALTOV, D., WU, W., LOCE, R. P., & Bernal, E. A. (2013). Parking lot occupancy determination from lamp-post camera images. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, (Itsc)*, 2387–2392. <https://doi.org/10.1109/ITSC.2013.6728584>
- [12] PFLUGLER, C., KÖHN, T., SCHEREIECK, M., WIESCHE, M., & KREMAR, H. (2016). Predicting the Availability of Parking Spaces with Publicly Available Data. *Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn*, 361 – 374
- [13] CAICEDO, F. (2009). The use of space availability information in “PARC” systems to reduce search times in parking facilities. *Transportation Research Part C: Emerging Technologies*, 17(1), 56–68. <https://doi.org/10.1016/j.trc.2008.07.001>
- [14] TILAHUN, S. L., & Di MARZO (2017). Cooperative multiagent system for parking availability prediction based on time varying dynamic markov chains. *Journal of Advanced Transportation*, 2017 . doi: 10.1155/2017/ 1760842
- [15] LIU, S., GUAN, H., YAN, H., & YIN, H. (2010). Unoccupied Parking Space Prediction of Chaotic Time Series. *Research on Influence of Aggregate Gradation on the Performance of Porous Asphalt Pavement*, c, 3738–3746. [https://doi.org/10.1061/41127\(382\)228](https://doi.org/10.1061/41127(382)228)

- [16] RICHTER, F., Di MARTINO, S., & MATTFELD, D. C. (2014). Temporal and Spatial Clustering for a Parking Prediction Service. In 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (Vol. November, pp. 278–282). IEEE. <https://doi.org/10.1109/ICTAI.2014.49>
- [17] KLAPPENECKER, A., LEE, H., & WELCH, J. L. (2014). Finding available parking spaces made easy. *Ad Hoc Networks*, 12 (1), 243–249. doi: 10.1016/j.adhoc.2012.03 .002
- [18] MURPHY, Kevin P (s.d.). Machine Learning A Probabilistic Perspective. Em: (). url: [http://noiselab.ucsd.edu/ECE228/Murphy\\_Machine\\_Learning.pdf](http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf)
- [19] H.S. DAS, M.M. RAHMAN, S. Li, C.W. TAN, Electric vehicles standards, charging infrastructure, and impact on grid integration: A technological review, *Renewable and Sustainable Energy Reviews*, , 2020 DOI: 10.1016/j.rser.2019.109618
- [20] BOCK, F., MARTINO, S. D., & ORIGLIA, A. (2017). A 2-Step Approach to Improve Data-driven Parking Availability Predictions. IWCTS'17 Proceedings of the 10th ACM SIGSPATIAL Workshop on Computational Transportation Science, 13–18. doi: 10.1145/3151547.3151550
- [21] GÉRIN, E., AYDIN, O., & Mahdavi-Amiri, A. (2019). Artificial Intelligence. *Manual of Digital Earth*, 357–385. [https://doi.org/10.1007/978-981-32-9915-3\\_10](https://doi.org/10.1007/978-981-32-9915-3_10)
- [22] SARGENT, D. J. (2001). Comparison of artificial neural networks with other statistical approaches. *Cancer*, 91(S8), 1636–1642. [https://doi.org/10.1002/1097-0142\(20010415\)91:8+3.0.CO;2-D](https://doi.org/10.1002/1097-0142(20010415)91:8+3.0.CO;2-D)
- [23] KARLAFTIS, M. G., & VLAHOGIANNI, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387–399. <https://doi.org/10.1016/j.trc.2010.10.004>
- [24] ZHENG, Y., SUTHARSHAN, R., & CHRISTOPHER, L. (2015). Parking availability prediction for sensor-enabled car parks in smart cities. In 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) (pp. 1– 6). IEEE. <https://doi.org/10.1109/ISSNIP.2015.7106902>

- [25] CHEIRAN, J., LUO, J., GUO, H., HO, S. S., & Wisbrun, R. (2016). ParkGauge: Gauging the occupancy of parking garages with crowdsensed parking characteristics. Proceedings - IEEE International Conference on Mobile Data Management, 2016–July, 92–101. <https://doi.org/10.1109/MDM.2016.26>
- [26] IONITA, A., POMP, A., COCHEZ, M., MEISEN, T., & DECKER, S. (2018). Where to Park?: Predicting Free Parking Spots in Unmonitored City Areas. Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, 22:1—22:12. doi: 10.1145/3227609.3227648
- [27] TAMZARIAN, A., QIAN, Z. (Sean), & RAJAGOPAL, R. (2015). Where Is My Parking Spot? Transportation Research Record: Journal of the Transportation Research Board, 2489, 77–85. <https://doi.org/10.3141/2489-09>
- [28] NETTLEMAN, M. D. Receiver operator characteristic (roc) curves. Infection Control and Hospital Epidemiology, v.9, n.8, p.374-377, 1988.
- [29] RAMOS Silva, M. H. (2017). Predicting Space Occupancy for Street Paid Parking (Unpublished doctoral dissertation). ISCTE-IUL.
- [30] FUMIKO Koyanagi, Yoshihisa Uriu., "Modeling Power Consumption by Electric Vehicles and Its Impact on Power Demand." Electrical Engineering in Japan Journal. Issue 4, 1997, Vol. 120
- [31] MONTEGOMERY, Douglas C. and RUNGER, George C., "Applied statistics and probability for engineers", John Wiley & Sons, Inc., USA, 2002 4,
- [32] A. GÉRON, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, 2017.
- [33] MITCHELL. Tom. Machine Learning. Mc Graw Hill Education, 1997 <https://doi.org/10.1609/aimag.v18i3.1303>
- [34] R. S. SUTTON and A. G. BARTO, Reinforcement learning: An introduction. MIT press, 2018
- [35] BEN-DAVID, Shai e Shai Shalev-Shwartz (2014). Understanding Machine Learning: From Theory to Algorithms, p. 409. isbn: 9781107057135. doi:

10.1017/CBO9781107298019. url:  
<http://www.cs.huji.ac.il/~7B~%7Dshais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>.

[36] LECUN, Yann, Yoshua BENGIO e Georey HINTON (2015). Deep learning. Em: Nature 521.7553, pp. 436444. issn: 0028-0836. doi: 10.1038/nature14539. arXiv: arXiv:1312.6184v5. url: <http://dx.doi.org/10.1038/nature14539>.

[37] SUTTON, Richard S. e Andrew G. BARTO 2012. Reinforcement learning. Em: 1.3, p. 22. issn: 18726240. doi: 10 . 1109 / MED . 2014 . 6608833. arXiv: 1603 . 02199. url: <https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf>

[38] ROKACH, Lior and MAIMON, Oded "IEEE Transaction On System Man and Cybernetics Part C", Top Down Induction Of Decision Tree Classifier-A Survey, vol. 1, no. 11, 2005.

[39] D.M.J. Tax and R.P.W. Duin. Support vector domain description. Pattern Recognition Letters, 20:1991–1999, 1999.

[40] BERCHTOLD, S., ERTL, B., KEIM, D.A., KRIEGEL, H.P., SEIDL, T.—Fast nearest neighbor search in high-dimensional space in Proceedings of the International Conference on Data Engineering, pp. 209–218 , 1998

[41] BREINAN L, Random Forests, 2001: Machine Learning, 45, 5-32.

[42] STOLFI, D. H., ALBA, E., & YAO, X. (2019). Can I Park in the City Center? Predicting Car Park Occupancy Rates in Smart Cities. Journal of Urban Technology, 0 (0), 1–15. doi: 10.1080/10630732.2019.1586223

[43] P. DOMINGOS and M. PAZZANI. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29:103–130, 1997.

[46] BEN-HUR, A.; HORN, D.; SIEGLMANN, H.T.; VAPNIK, V. Support vector clustering. J. Mach. Learn. Res, 2, 125–137. 2001

[47] PROVOOST, J. C., KAMILARIS, A., WISMANS, L. J. J., VAN DER DRIFT, S. J., & van Keulen, M. (2020). Predicting parking occupancy via machine learning in the web of things. *Internet of Things*, 12, 100301. <https://doi.org/10.1016/j.iot.2020.100301>

- [48] NAVADA, A., ASARI, A. N., PATIL, S., & SONKANBLE, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *2011 IEEE Control and System Graduate Research Colloquium*. <https://doi.org/10.1109/icsgrc.2011.5991826>
- [49] GEURTS, P., ERNST, D., & WEHENKEL, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- [50] New registrations of electric vehicles in Europe. (2020). Retrieved September 27, 2022, from Europa.eu website: <https://www.eea.europa.eu/ims/new-registrations-of-electric-vehicles>
- [51] S SCHULLER A, FLATH CM, GOTTWALT S. Quantifying load flexibility of electric vehicles for renewable energy integration. *Applied Energy*. 2015;151:335-44.
- [52] FOROOZANDEH, Z., RAMOS, S., SOARES, J., VALE, Z., & DIAS, M. (2022). Single contract power optimization: A novel business model for smart buildings using intelligent energy management. *International Journal of Electrical Power & Energy Systems*, 135, 107534. <https://doi.org/10.1016/j.ijepes.2021.107534>
- [53] R. L. JOENCH, J. SOARES, F. LEZAMA, S. RAMOS, A. GOMES and Z. VALE, "A Short Review on Smart Building Energy Resource Optimization," 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), 2019, pp. 440-445, doi: 10.1109/GTDAsia.2019.8715982.
- [54] C. SILVA, P. FARIA and Z. VALE, "Using Supervised Learning to Assign New Consumers to Demand Response Programs According to the Context," *2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe)*, 2022, pp. 1-6, doi: 10.1109/EEEIC/ICPSEurope54979.2022.9854646.

# ANEXO

**Tabela 18 - Estrutura dos dados (subcave)**

	Lugar	Semana	dia_semana	dia da semana	dia_code	hora_code	estacionado	soc	capacidade kWh
<b>0</b>	A1	1	2	segunda-feira	01/01/2018	00:10:00	1	70.0	75
<b>1</b>	A1	1	2	segunda-feira	01/01/2018	00:30:00	1	70.0	75
<b>2</b>	A1	1	2	segunda-feira	01/01/2018	01:00:00	1	80.0	75
<b>3</b>	A1	1	2	segunda-feira	01/01/2018	01:30:00	1	80.0	75
<b>4</b>	A1	1	2	segunda-feira	01/01/2018	02:00:00	1	90.0	75
...	...	...	...	...	...	...	...	...	...
<b>402953</b>	A23	53	2	segunda-feira	31/12/2018	21:30:00	1	50.0	100
<b>402954</b>	A23	53	2	segunda-feira	31/12/2018	22:00:00	1	50.0	100
<b>402955</b>	A23	53	2	segunda-feira	31/12/2018	22:30:00	1	50.0	100
<b>402956</b>	A23	53	2	segunda-feira	31/12/2018	23:00:00	1	50.0	100

402957 linhas x 9 colunas

**Tabela 19 – Estrutura dos dados sem férias e feriados (subcave)**

	Lugar	Semana	dia_semana	dia da semana	dia_code	hora_code	estacionado	soc	capacidade kWh
<b>0</b>	A1	1	3	terça-feira	02/01/2018	23:30:00	1	30.0	50
<b>1</b>	A1	1	3	terça-feira	02/01/2018	00:10:00	1	30.0	50
<b>2</b>	A1	1	3	terça-feira	02/01/2018	00:30:00	1	40.0	50
<b>3</b>	A1	1	3	terça-feira	02/01/2018	01:00:00	1	40.0	50
<b>4</b>	A1	1	3	terça-feira	02/01/2018	01:30:00	1	50.0	50
...	...	...	...	...	...	...	...	...	...
<b>356607</b>	A23	53	2	segunda-feira	31/12/2018	21:00:00	1	50.0	100

	Lugar	Semana	dia_semana	dia da semana	dia_code	hora_code	estacionado	soc	capacidade kWh
356608	A23	53	2	segunda-feira	31/12/2018	21:30:00	1	50.0	100
356609	A23	53	2	segunda-feira	31/12/2018	22:00:00	1	50.0	100
356610	A23	53	2	segunda-feira	31/12/2018	22:30:00	1	50.0	100

356612 linhas x 9 colunas

**Tabela 20 - Estrutura dos dados sem férias e feriados (cave)**

	Lugar	Semana	dia_semana	dia da semana	dia_code	hora_code	estacionado	soc	capacidade kWh
0	A1	1	3	terça-feira	02/01/2018	23:30:00	0	30.0	50
1	A1	1	3	terça-feira	02/01/2018	00:10:00	0	30.0	50
2	A1	1	3	terça-feira	02/01/2018	00:30:00	0	30.0	50
3	A1	1	3	terça-feira	02/01/2018	01:00:00	0	40.0	50
4	A1	1	3	terça-feira	02/01/2018	01:30:00	0	40.0	50
...	...	...	...	...	...	...	...	...	...
279084	A18	53	2	segunda-feira	31/12/2018	21:00:00	0	90.0	60
279085	A18	53	2	segunda-feira	31/12/2018	21:30:00	0	100.0	60
279086	A18	53	2	segunda-feira	31/12/2018	22:00:00	0	100.0	60
279087	A18	53	2	segunda-feira	31/12/2018	22:30:00	0	100.0	60

279089 linhas x 9 colunas

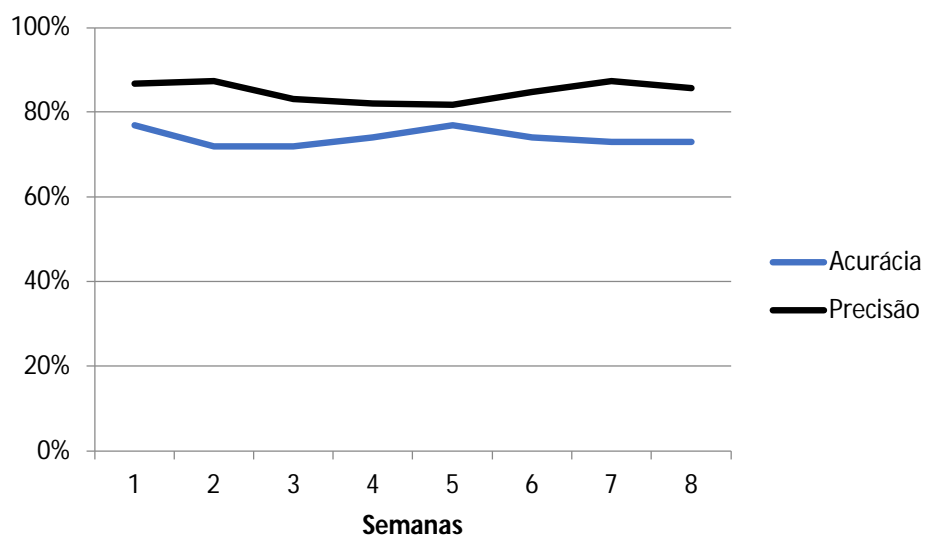


Figura 25 – Desempenho de Exatidão e Precisão KNN (sem feriados e férias)

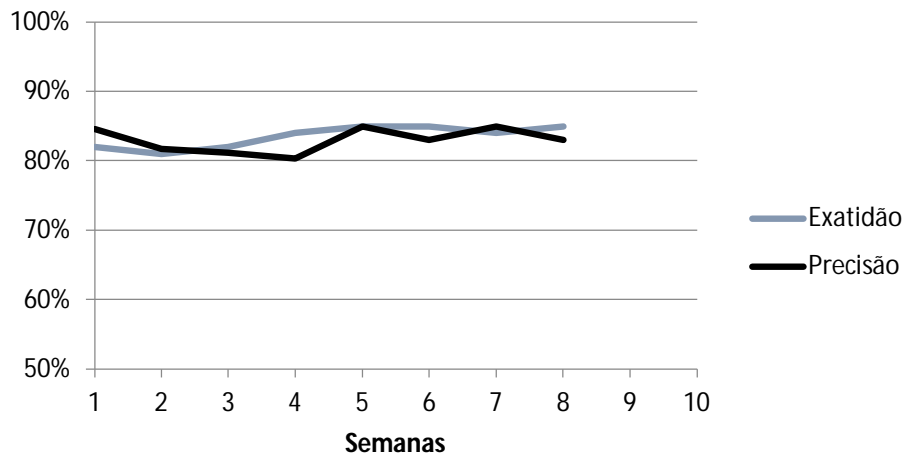


Figura 26 - Desempenho de Exatidão e Precisão KNN (ordem aleatória)