

Knowledge Discovery for Risk Assessment in Economic and Food Safety

Maria Clara Silva¹, Brigida Monica Faria^{1,2} ^a and Luis Paulo Reis^{2,3} ^b

¹ESS, Polytechnic of Porto (ESS-P.PORTO), Rua Dr. Antonio Bernardino de Almeida, 400 4200 - 072 Porto, Portugal

²Artificial Intelligence and Computer Science Laboratory (LIACC- Member of LASI LA),
Rua Dr. Roberto Frias, sn 4200-465 Porto, Portugal

³Faculty of Engineering, University of Porto (FEUP), Rua Dr. Roberto Frias, sn, 4200-465 Porto, Portugal

Keywords: Food Safety, Risk Assessment, Public Health, Knowledge Discovery.

Abstract: Foodborne diseases continue to spread widely in the 21st century. In Portugal, the Economic and Food Safety Authority (ASAE), have the goal of monitoring and preventing non-compliance with regulatory legislation on food safety, regulating the conduct of economic activities in the food and non-food sectors, as well as accessing and communicating risks in the food chain. This work purpose and evaluated a global risk indicator considering three risk factors provided by ASAE (non-compliance rate, product or service risk and consumption volume). It also compares the performance on the prediction of risk of four classification models Decision Tree, Naïve Bayes, k-Nearest Neighbor and Artificial Neural Network before and after feature selection and hyperparameter tuning. The principal findings revealed that the service provider, food and beverage and retail were the activity sectors present in the dataset with the highest global risk associated with them. It was also observed that the Decision Tree classifier presented the best results. It was also verified that data balancing using the SMOTE method led to a performance increase of about 90% with the Decision Tree and k-Nearest Neighbor models. The use of machine learning can be helpful in risk assessment related to food safety and public health. It was possible to conclude that areas regarding major global risks are the ones that are more frequented by the population and require more attention. Thus, relying on risk assessment using machine learning can have a positive influence on economic crime prevention related to food safety as well as public health.

1 INTRODUCTION

Worldwide, hazardous foods contribute to a significant number of food-related illnesses, resulting in a substantial number of deaths annually (WHO, 2023). Besides the health and social issues, food-associated diseases also provoke an enormous economic impact on society. It was registered in the United States that those food-related illnesses cause an economic impact with costs estimated to be in the billions of dollars every year (Hoffmann et al., 2015). Food safety measures also imply an expense of around seven billion dollars each year from the notification of consumers to the subsidy of the amends from lawsuits.

Economic and Food Safety Authority (ASAE)

(ASAE, 2023) is a national authority established in Portugal with administrative autonomy that pretends to supervise and prevent non-compliance with the regulatory legislation regarding food safety, govern the conduct of economic activities in the food and non-food sectors, as well as to assess and communicate risks in the food chain (Magalhães et al. 2019) (Magalhães et al. 2020). ASAE is also responsible for communicating with its counterparts at an international level. ASAE was created in 2006 and it is an entity that aims to project itself as a reference entity in consumer safety, public health, safeguarding market rules and free competition by providing public service (ASAE, 2023) (Pinto et al., 2019). ASAE is ruled by a code of conduct and ethics established in four points: common rules, inspection

^a  <https://orcid.org/0000-0003-2102-3407>

^b  <https://orcid.org/0000-0002-4709-1718>

area, scientific and laboratory area and procedure decisions area. ASAE acts on complaints, denouncements and reports that are submitted via telephone contact, e-mail, fax, website or even in person (ASAE, 2023) (Filgueiras et al., 2019). The main goal of this work is to assess the risk related to food safety, public health and economics, using the information provided by ASAE. Data was gathered over the years by ASAE's specialists and inspectors through inspections from all over the country. The risk assessment task was performed considering three risk factors given by ASAE, namely the non-compliance rate, product or service risk and consumption volume to be considered for the global risk. Another objective was to create a predictive system using classification models and employing global risk as the output variable.

This paper is organized into five sections: Introduction, Risk Assessment and Related Work, Methodology, Results and Discussion and a last section with Conclusions and Future Work.

2 RISK ASSESSMENT AND RELATED WORK

The goal of risk assessment is to help individuals or organizations make informed decisions and take appropriate actions to minimize the negative impact of potential risks. This section presents the concept of risk in public health and several approaches to assess risk based on machine learning.

2.1 Risk in Public Health

Risk has been defined in a variety of ways and it is commonly associated with words like “hazard” and “uncertainty”. Nowadays, some of the definitions are “the possibility of loss, injury, disadvantage, or destruction” or “the likelihood that harm will occur” (Jannadi and Almishari, 1999) (Dilley et al., 2001). In this work, the risk is assumed as the probability and severity of hazard outcomes caused by internal or external vulnerabilities of a certain activity and it can be avoided with preventive actions (Jannadi and Almishari, 1999) (Dilley et al., 2001). When applied the concept to the food and public health industry, can be described as a hazard present in products that cause harm of a certain magnitude. Food safety has been defined as the “concept that food will not cause harm to the consumer when it is prepared and/or eaten according to intended use” (Borchers et al., 2010). We can never be assured that a certain food is not

contaminated because it is impossible to perform every test to guarantee that every single item is free of toxins, foodborne pathogens, adulterants or contaminants. This would also have a huge economic impact as it would be extremely expensive to analyze in such detail (Jannadi and Almishari, 1999). Despite this, almost every country has a food safety agency such as ASAE in Portugal and EFSA in Europe (Fung et al., 2008) (EFSA, 2023). These agencies oversee food safety and define a “reasonable certainty of no harm” and regulate the additives that are allowed in food and the levels of unavoidable contaminants that are acceptable (Jannadi and Almishari, 1999).

2.2 Risk Assessment Approaches

According to (Hedge and Rokseth, 2020) Artificial Neural Networks, Support Vector Machines (SVM), followed by Decision Trees are the classification models that have been used the most for risk assessment approaches. Artificial Neural Networks (ANN) are advantageous since they use non-mathematical equations to develop important relationships between input and output variables (Hedge and Rokseth, 2020). ANN also require less formal statistical training to develop, and they possess the ability to encounter all possible interactions between the input and the output variable (Hedge and Rokseth, 2020) (Paltrinieri et al., 2019). These are the main reasons why so many authors use ANN for classification problems. In the study, the ANN provided good results with performances higher than 80%. However, despite the good results, the ANN is known to cause overfitting, as in, the relationship between a given input and output variable is generalized to a specific dataset (Paltrinieri et al., 2019). Also, ANN usually takes longer to train than other classification models. According to (van den Bulk et al., 2022) out of the models, the k-NN, SVM and NB are the algorithms that have been used the most frequently in the context of food safety prediction and monitoring. According to the authors, the NB structure is easier to understand, when compared to other ML models. It appears to be a promising model for analyzing data in the context of food safety since it is capable of dealing with a variety of different drivers such as economic factors, climate change and human behaviour to predict future events of food safety risks. A study for risk assessment using machine learning (Galindo and Tamayo, 2000) showed that the ANN provided the second-best results of 85% and the k-NN provided the third-best results of 83%. Decision Trees perform better than Naïve Bayes (NB) when applied to risk prediction of

healthcare accidents according to (Awad and Khanna, 2015). Decision Trees can be less effective in predicting the outcome of a continuous variable.

A study on food safety (Wu and Weng, 2021) where four different ML were implemented (Random Forest, Logistic Regression, Naïve Bayes and Support Vector Machine) showed that NB and SVM performed best out of the four, with a performance around 90%. It is defended that SVM tends to perform well in risk assessment tasks, because of their ability to generalize well in a great number of features. Despite their good performances for large datasets, according to (Paltrinieri et al., 2019), ANN still performs better. Nonetheless, more traditional methods such as the NB and SVM are still better options for classification problems, since ANNs tend to perform well only in huge datasets. SVM are more commonly used in classification problems and usually performs better than DT. Even though they are not as popular in regression tasks as they are in classification tasks.

3 METHODOLOGY

The methodology employed in this work is based on the Knowledge Discovery in Databases phases, involving risk assessment and risk classification models. The data used in this work is based on real data regarding reports, complaints and denouncements that were submitted to the Economic and Food Safety Authority, being a real use case combining risk assessment, food safety and infraction prevention.

3.1 Problem Definition

ASAE executes its inspections outlined following action guidelines based on central planning articulated with regional planning with criteria previously established in the Inspection and Oversight Plan, resulting from internal investigations carried out by the Information Analysis and Research Division or, if determined, by a superior level (PNFA, 2023). ASAE, being an authority that brings together, in terms of competencies, all aspects of risk analysis, namely risk assessment, management and communication, must assume an integrated strategy (ASAE, 2023). ASAE has to collect and analyze data that allow the characterization and assessment of risks that have a direct or indirect impact on food safety, ensuring public and transparent communication of risks and promoting

the dissemination of information on food safety with consumers (PNFA, 2023).

To help ASAE with the risk assessment matter, it is important to develop a predictive system model based on information regarding reports on infractions, complaints and denouncements collected from inspections from previous years, using data mining techniques. This classification model for risk assessment should use global risk labels previously defined.

3.2 Data Selection

The dataset is divided into three categories: the ASAE reports dataset which refers to the reports made to ASAE, the complaints submitted to ASAE and the denouncements submitted to ASAE in the indicated years.

Data were collected in Portugal from North to South, in the years ranging between 2001 and 2019. The dataset referring to the ASAE reports contained 185505 subjects and 65 variables, the dataset referring to the complaints contained 493482 subjects and five variables and lastly the dataset referring to the denouncements contained 165057 subjects and five variables. The datasets also contain information related to the number of denouncements, complaints, arrests, closed establishments and other important information collected from the ASAE inspectors that have a high prevalence for the risk evaluation. Despite the inspections-related information, there is also present information regarding the activity sectors and areas of action of the entities as well as operational areas. The number of variables in the ASAE datasets was reduced to contain the most valuable information. Table 1 presents information regarding the variables present in the dataset.

3.3 Data Pre-Processing

To achieve clean and duplicate-free data, it was performed a data pre-processing of the dataset included the complaints, denouncements and ASAE reports. A search for missing values, noisy data and duplicate values was performed. All data cleaning operations were done following the steps: search and remove lines with missing values and noisy data considering the columns “target entity ID” and “target entity name”; search for missing values and noisy data considering the columns containing numeric variables and replacing it with the median. All the duplicates were aggregated and it was observed that some entities presented a similar name only differing in one point or comma, so it was assumed as an error

and that they belong to the same entity. To correct this typo, the entities with very similar names were merged using a similarity value of 95%. Table 2 summarises the number of missing values and duplicates identified in the datasets.

Table 1: Dataset variables.

Variable	Domain	Variable	Domain
No. of arrests	0,...,311	Year of inquiry	2001,...,2019
No. of closed establishments	0,1	Month of inquiry	1,...,12
No. of infractions with offences	0,...,13593	Operational area	Public health, commercial practices
No. of infractions with crimes	0,...,5	Target type	Retail, industry, service provider
No. of complaints	1,...,1277	Principal CAE	Activities of general practice
No. of denunciations	1,..., 2035	Secondary CAE	Nursing activities
No. of notices	0,...,10	Activity designation	Coffee shop, pharmacy, supermarket
No. of proceedings with offences	0,...,102	Has infractions with offences	True, false
No. of proceedings with crimes	0,...,358	Has infractions with crimes	True, false
No. of inspections in a partial state	0,1	Activity group	Food, economic
No. of autos with offences	0,...,4	Activity sector	Food, security, production
No. of autos with crimes	0,...,2	Target entity name	Name of the entity*
No. of missing proceedings with offences	0,...,3	Target entity ID	Number with seven digits*
No. of missing proceedings with crimes	0,1	Date of inquiry	18/10/2013; 22/01/2014
No. of instant proceedings	0,...,102		

*The real ID and names of the entities cannot be disclosed because of data privacy.

Table 2: Missing values and duplicates in the datasets.

Dataset	Initial cases	Missing Values	Duplicates	Final cases
Complaints	493481	41	415819	77621
Denunciations	165056	43	61205	103808
ASAE reports	185504	4196	92549	88759

3.4 Risk Assessment

Three criteria were defined in the economics and food area to determine risk: non-compliance rate, product and service risk and consumption volume.

Non-compliance rate (NCR) is related to the legal aspects of the process. The NCR is estimated with an analysis of the total number of processes from the previous year. The risk is directly proportional to the NCR, so the higher the NCR, the bigger the risk. The non-compliance rate is described using the degree of nonconformity in the oversights, the level of risk to public health and the economic safety of the consumer.

Product or service risk (PSR) can be established using the number of denunciations and complaints received at ASAE from the previous year, with the estimation of the risk elaborated by the Division of Food Risk to the Food Area of ASAE and communications by other national and international entities like INFARMED (2023) (National Authority for Medicines and Health Products, I.P.) and the European Commission (RASFF, 2023), among others. These external indicators allow for detecting patterns in some economic activity sectors in need of an oversight intervention. The risk based on denunciations and complaints increases as the number of denunciations and complaints increases.

Consumption volume (CV) includes the analysis of the number of workers and business volume from the previous year. The data is collected from external and trustworthy entities like the National Institute of Statistics (INE, 2023), PORDATA (2023) and/ or academic institutions.

The micro risk (*MIR*) is determined by the product of the three risk factors and calculated by the entity considering the following Equation 1:

$$MIR = NCR \times PSR \times CV \quad (1)$$

where *NCR* is the non-compliance rate, *PSR* is the product or service risk and *CV* is the consumption volume. A macro risk (Equation 2) was determined by the three risk factors, an ASAE utility and the activity sector.

$$MAR_{i=1,2} = \frac{NCR}{\sum_{as} NCR} + \frac{PSR}{\sum_{as} PSR} + \frac{CV}{\sum_{as} CV} + \frac{U_i}{\sum_{as} U_i} \quad (2)$$

where, MAR_i , represents the macro risk sum, NCR is the non-compliance rate, PSR is the product or service risk, CV is the consumption volume, U_i , represents the two cases of ASAE utility and as is the activity sector.

The ASAE utility is a feature needed for the macro risk determination. It is a value that ranges between 0 and 4, where 0 represents the lowest utility or importance attributed to a particular area and 4 represents the highest utility or importance designated to a particular area. The utility was attributed differently by area for two cases to produce more variance and create distinct approaches for the macro risk. Table 3 describes how the ASAE utility was assigned.

Table 3: Missing values and duplicates in the datasets.

Area	Case 1	Case 2
Other	0	0
Cultural and well-being activities	1	1
Food and beverage	2	4
Food and health industry	3	2
Healthcare	4	3

The activity sector was assessed considering the variable “target type” since it was the most complete and accurate for the data.

The global risk was determined by the product between the micro risk and macro risk, and it is represented in Equation 3 as:

$$GR_{i=1,2} = MIR \times MAR_i \quad (3)$$

where, GR_i , $i=1,2$ represents the two cases of the global risk, MIR represents the micro risk and MAR_i , represents the two cases of the macro risk. The global risk was classified into five levels considering the different levels of risk.

3.5 Risk Assessment Using a Classification Approach

Four learning-based algorithms were applied to the data for the risk classification prediction task: a Decision Tree algorithm with a splitter number with the default “best”, to choose the best split at each node, an Artificial Neural Network algorithm with 100 iterations for the model to converge in the training step, with a learning rate of 0.001, a k-Nearest Neighbor algorithm with $k=5$ and a Naïve Bayes algorithm with an additive smoothing

parameter of 1.0. The categorical variable “target type”, which refers to the activity sector, was coded between 0 and 10. The input variables chosen by ASAE were the number of arrests, closed establishments, infractions with offences, infractions with crimes, denouncements and complaints and the target type. The results were assessed using three evaluation metrics: precision, recall and F1-score. A final experiment was performed aimed to balance the data using the Synthetic Minority Oversampling Technique (SMOTE) method (Chawla, 2002).

4 RESULTS AND DISCUSSION

ASAE provided a dataset regarding information from inspections with data collected from entities in Portugal, from North to South, within the timeline of 18 years, starting in 2001 until 2019. This section presents the achieved results, starting with an exploratory analysis.

4.1 Exploratory Data Analysis

The next analysis refers to an exploratory analysis of the numeric variables. Table 4 shows the mean, median and standard deviation of the quantitative variables.

Table 4: Statistical measures of numeric variables.

Variable	Mean	Median	SD
No. of arrests	0.121	0	0.861
No. of closed establishments	0.142	0	0.667
No. of infractions with offences	3.305	1	6.444
No. of infractions with crimes	0.322	0	2.482
No. of complaints	1.326	0	31.138
No. of denouncements	0.278	0	2.282
No. of inspections in a partial state	0.044	0	0.331
No. of notices	0.501	0	1.255
No. of proceedings with offences	1.497	1	2.043
No. of proceedings with crimes	0.203	0	1.118
No. of missing proceedings with offences	0.217	0	0.754
No. of missing proceedings with crimes	0.066	0	0.911
No. of instant proceedings	1.734	1	2.491
No. of autos with offences	1.584	1	1.972
No. of autos with crimes	0.258	0	1.844

Legend: SD – Standard Deviation.

It is possible to observe that the median is zero for variables related to the number of arrests, closed establishments, infractions with crimes, complaints, denouncements, inspections in a partial state, notices,

proceedings with crimes, missing proceedings with offences, missing proceedings with crimes and the number of autos with crimes. However, for the number of infractions with offences, proceedings with offences, instant proceedings and autos with offences, the median was 1. The mean of the variables ranged between 0.044 and 3.305, which means a dataset with a high content of zeros in each variable.

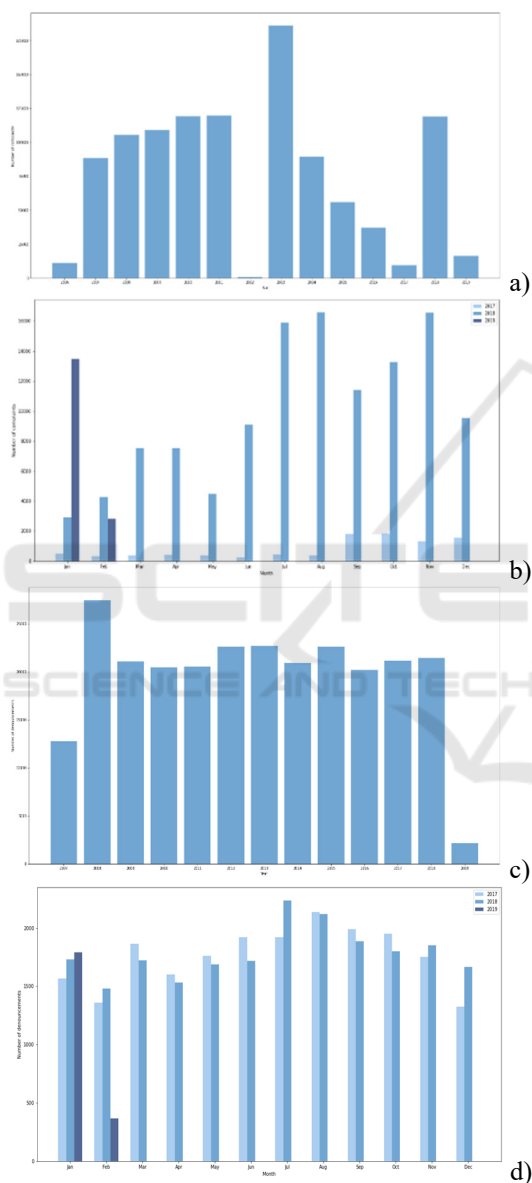


Figure 1: Complaints and denouncement distribution: a) the number of complaints through the years 2006 until 2019; b) the number of complaints by month between the years 2017 and 2019; c) the number of denouncements through the years 2008 until 2019, d) number of denouncements by month between the years 2017 and 2019.

The number of overall non-compliances, denouncements and complaints was very low, which will also lead to low risk associated with it.

The years from 2001 to 2004 were not assessed since they mostly presented zeros. The same happened in some variables for the years between 2004 and 2007. It is also important to mention that the year 2019 is incomplete and only presents the months of January and February. For the seven variables chosen (number of complaints, denouncements, arrests, closed establishments, infractions with offences and infractions with crimes) it was also determined the variation through the months between the years 2017 and 2019.

The number of initiated complaints varied through the years (Fig. 1 a.) The year 2013 was the year with more complaints (around 180000) and 2012 was the year with fewer complaints (under 1000). Fig. 1 b. shows that July, August and November presented the highest number of complaints in 2018. The number of denouncements presented small variations, being quite homogenous (Fig. 1 c.). The number of denouncements was above 10000 for every year, except for 2019 which only included the two first months of the year. Fig. 1 d. shows that every month presented more than 1000 denouncements. In 2017 the only two months with a considerable number of denouncements were January and February.

4.2 Risk Assessment Results

This section presents the results of the risk analysis performed using the three risk factors, the micro risk and the macro risk and the global risk for the selected activity sector.

The three activity sectors with a higher consumption volume and non-compliance rate were the service provider, food and beverage and retail. The product or service risk was higher for the service provider, food and beverage and storer. The distribution of the risk factors through the 10 activity sectors followed a similar pattern, however, for the non-compliance rate and product or service risk, the numbers were very low compared to the consumption volume. The micro risk represents the product between the three risk factors. Fig. 2 shows that the activity sectors with higher micro risk were the service provider, food and beverage and retail.

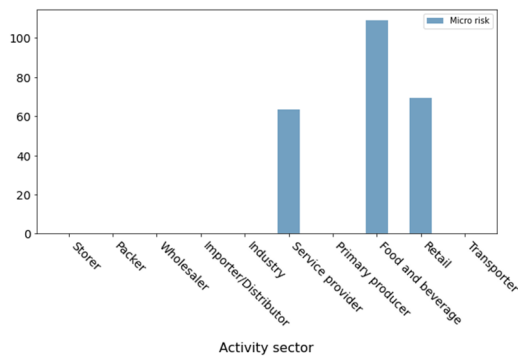


Figure 2: Micro risk by activity sector.

This is explained since the micro risk is the result of the NCR, PSR and CV, so the activity sectors with a higher micro risk would be the same as the ones with a higher NCR, PSR and CV. Analogous to the macro and the global risks. It was also possible to observe that the sectors with ASAE utility 1 overall demonstrated a higher macro risk than the sectors with ASAE utility 2. The results from the risk analysis showed that the service provider (around 60%), food and beverage (around 30%) and retail (around 20%) were the activity sectors with a higher prevalence of global risk.

4.3 Risk Analysis Classification Experiment

Table 5: Global risk (GR1,2) classification results.

Model	FS	HT	TT (s)	Prec	Recall	F1-Sc
DT	without	without	0.003	0.55	0.59	0.55
		with	0.002	0.60	0.54	0.60
	with	without	0.004	0.57	0.59	0.57
		with	0.002	0.54	0.59	0.55
NB	without	without	0.001	0.46	0.46	0.46
		with	0.001	0.46	0.46	0.46
	with	without	0.001	0.46	0.46	0.46
		with	0.001	0.46	0.46	0.46
k-NN	without	without	0.010	0.55	0.57	0.56
		with	0.005	0.55	0.57	0.56
	with	without	0.010	0.54	0.56	0.55
		with	0.005	0.54	0.56	0.55
ANN	out	without	1.563	0.57	0.61	0.58
		with	0.565	0.57	0.61	0.58
	with	without	1.582	0.54	0.58	0.56
		with	0.758	0.54	0.58	0.56

Legend: FS – feature selection; HP – hyperparameter tuning; TT – training time; Prec – precision; F1-Sc – F1-Score

A classification experiment was performed as previously described in the methodology. Four different classification models were chosen for the

global risk classification prediction: Decision Tree, k-Nearest Neighbor, Naïve Bayes and Artificial Neural Network. Feature selection, using seven features, was applied to the classification experiment, using the Chi-square feature selection method. The results (Table 5) showed that the performances of the classifiers in general did not improve with the application of feature selection. This can mainly be explained because some of the variables selected by feature selection presented a very high number of zeros. The performance of the NB was the same for all the experiments above.

By balancing the data using SMOTE, it is possible to observe in Table 6 that the results improved greatly.

Table 6: Global risk (GR1,2) classification results using SMOTE to balance data.

Model	Training Time (s)	Precision	Recall	F1-score
DT	0.007	0.972	0.972	0.972
NB	0.001	0.770	0.739	0.726
k-NN	0.003	0.930	0.928	0.927
ANN	0.614	0.873	0.867	0.840

The results from this experiment indicate, that the low performances were due to the asymmetry in data. With balanced data, it was presented results of 93% for the k-NN and 98% for the DT.

5 CONCLUSIONS AND FUTURE WORK

In this study, the risk assessment was performed using the global risk, determined using three risk factors (NCR, PSR and CV), the micro risk and the macro risk. The risk assessment was performed by activity sector, and it was possible to observe that the service provider, the food and beverage and the retail were the areas with the biggest risk associated with, when taking into consideration the number of infractions committed by the entity, the number of complaints and denouncements done by the costumers and the size of the entity in terms of the number of workers. The service provider embraces hospitals and other entities that provide services to the population. The food and beverage include coffee shops, restaurants and other similar entities. Retail includes entities such as supermarkets, among other entities that sell products to the population. These areas are usually crowded by people, and for that matter the number of dissatisfactions is higher, leading to an increase in the number of complaints and denouncements. The need

to implement a risk assessment system based on machine learning techniques has increased significantly in the past few years with the improvement of artificial intelligence and applications. Although the performances were not very high, except in the balanced dataset experiment, however, it was still possible to observe the use of machine learning in risk assessment could be very advantageous for the prediction of hazards and dangers related to food safety and public health.

For future work, it would be relevant to gather information regarding the consumption value. This would change the way the consumption volume was determined since the number of workers was the only used factor. Also as future work, different weights could be assigned to the features in the risk equations, thus creating diverse emphasis on each risk factor.

ACKNOWLEDGEMENTS

This work was supported by Base Funding UIDB- 00027-2020 of LIACC - funded by national funds through the FCT/MCTES (PIDDAC), by project IA.SAE, funded by FCT through program INCoDe.2030 and CIGESCOP – Centro Inteligente de Gestão e Controlo Operacional (POCI-05-5762-FSE-000215) by COMPETE 2020 (Fundo Social Europeu – FSE).

REFERENCES

- ASAE (2023). *Como Atua a ASAE*. Available from: <https://www.asae.gov.pt/inspecao-fiscalizacao/como-atua-a-asae.aspx>, last accessed 2023/08/15.
- Awad, M., Khanna, R. (2015). *Efficient Learning Machines - Theories, Concepts, and Applications for Engineers and System Designers*, Apress Berkeley, CA.
- Borchers, A., Teuber, S.S., Keen, C.L., Gershwin, M.E. (2010). *Food Safety*. Clin Rev Allergy I, 39(2), 95–141.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique, J Art Intelligence Research, 16.
- Dilley, M., Boudreau, T. E. (2001). *Coming to terms with vulnerability: a critique of definition of food security definition*. Food Policy., 26(3), 229–47.
- EFSA (2023). *About us | EFSA*, <https://www.efsa.europa.eu/en/aboutefsa>, last accessed 2023/03/15.
- Filgueiras, J., Barbosa, L., Rocha, G., Cardoso, H.L., Reis, L.P., Machado, J.P., Oliveira, A.M. (2019). Complaint Analysis and Classification for Economic and Food Safety. In Proceedings of the Second Workshop on Economics and Natural Language Processing, pages 51–60, Hong Kong. Association for Computational Linguistics.
- Fung, F., Wang, H-S., Menon, S. (2008). *Food safety in the 21st century*. Biomed J., 41(2), 88–95.
- Galindo J., Tamayo, P. (2000). *Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications*. Comput Econ. 15(1–2):107–43.
- Hegde, J., Rokseth, B. (2020). *Applications of machine learning methods for engineering risk assessment – A review*. Saf Sci., 122, 104492.
- Hoffmann S., Macculloch, B., Batz M. (2015). *Economic burden of major foodborne illnesses acquired in the United States*. E. Cost Foodb. Ill. in the US, pp. 1–74.
- INE (2023). *Statistics Portugal - Web Portal*, https://www.ine.pt/xportal/xmain?xpgid=ine_main&xpid=INE&xlang=en, last accessed 2023/04/01.
- INFARMED (2023). *INFARMED, I.P.*, <https://www.infarmed.pt/web/infarmed-en/>, last accessed 2023/03/15.
- Jannadi, O.A., Almishari S. (1999). Risk Assessment in Construction. J Constr Eng Manag., pp. 492–500.
- Magalhães, G., Faria, B.M., Reis, L.P., Cardoso, H.L., (2019). Text Mining applications to facilitate economic and food safety law enforcement, 4th Int. Conf. on Big DA, DM and Comp. Int., IADIS Press, pp. 199-203
- Magalhães, G., Faria, B.M., Reis, L.P., Cardoso, H.L., Caldeira, C., Oliveira, A. (2020). Automating Complaints Processing in the Food and Economic Sector: A Classification Approach. Adv in Int. Systems and Computing, vol 1160. Springer, Cham.
- Paltrinieri, N., Comfort, L., Reniers, G. (2019). *Learning about risk: Machine learning for risk assessment*. Saf Sci., 118, pp.475–86.
- Pinto, T., Faria, B.M., Reis, L.P., Cardoso H.L., Santos, T. (2019). Compliance study of hazard analysis and critical control point system. International Conference Big Data Analytics, Data Mining and Computational Intelligence, pp. 111–118. ISBN: 9789898533920.
- PNFA (2023). *Plano Nacional de Fiscalização Alimentar*, <https://www.asae.gov.pt/inspecao-fiscalizacao/plano-de-inspecao-da-asae-pif/area-alimentar/plano-nacional-de-fiscalizacao-alimentar.aspx>, last accessed 2023/03/15.
- PORDATA (2023). *Estatísticas, gráficos e indicadores*, <https://www.pordata.pt/>, last accessed 2023/04/01.
- RASFF (2007). *The Rapid Alert System for Food and Feed (RASFF) food and feed safety*. European Commission. Available from: https://food.ec.europa.eu/safety/rasff_en, last accessed 2023/07/20.
- van den Bulk, L., Bouzembrak, Y., Gavai, A., Liu, N., van den Heuvel, L., Marvin, H. (2022). *Automatic classification of literature in systematic reviews on food safety using machine learning*. Curr Res Food Sci., 5, pp. 84–95.
- WHO (2023). *Estimates of the Global Burden of Foodborne Diseases*, http://www.who.int/foodsafety/areas_work/foodborne-diseases/ferg/en/, last accessed 2023/08/15.
- Wu, L.Y., Weng, S-S. (2021). *Ensemble learning models for food safety risk prediction*. Sust.. 13(21),1–26.