Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

# Proposal of a workplace classification model for heart attack accidents from the field of occupational safety and health engineering

Alberto Sánchez-Lite [a,*], Jose Luis Fuentes-Bargues [b], Iván Iglesias [c], Cristina González-Gaya [d]

[a] Department of Materials Science and Metallurgical Engineering, Graphic Expression in Engineering, Cartographic Engineering, Geodesy and Photogrammetry, Mechanical Engineering and Manufacturing Engineering, School of Industrial Engineering, Universidad de Valladolid, P° del Cauce 59, 47011, Valladolid, Spain
[b] Project Management, Innovation and Sustainability Research Center (PRINS), Universitat Politècnica de València, 46022, Valencia, Spain
[c] EEI, School of Industrial Engineering, Univeridad de Vigo, Rúa Maxwell, n°9. 36310, Vigo-Pontevedra, Spain
[d] Department of Construction and Manufacturing Engineering, Escuela Superior de Ingeniería Industrial Universidad Nacional de Educación a Distancia (UNED), C/ Juan del Rosal 12, 28040, Madrid, Spain

## ABSTRACT

Research on occupational accidents is a key factor in improving working conditions and sustainability. Fatal accidents incur significant human and economic costs. Therefore, it is essential to examine fatal accidents to identify the factors that contribute to their occurrence. This study presents an overview of fatal heart attack accidents at work in Spain over the period 2009–2021. Descriptive analysis was conducted considering 13 variables classified into five groups. These variables were selected as predictors to determine the occurrence of this type of accident using a machine learning technique. Thirteen Naïve Bayes prediction models were developed using an unbalanced dataset of 15,616 valid samples from the Spanish Delta@database, employing a two-stage algorithm. The final model was retained using a General Performance Score index. The model selected for this study used a 70:30 distribution for the training and test datasets. A sample was classified as a fatal heart attack if its posterior probability exceeded 0.25. This model is assumed to be a compromise between the confusion matrix values of each model. Sectors with the highest number of heart attacks are 'Health and social work', 'Transport and storage', 'Manufacturing', and 'Construction'. The incidence of heart attacks and fatal heart attack accidents is higher in men than in women and higher in private sector employees. The findings and model development may assist in the formulation of surveillance strategies and preventive measures to reduce the incidence of heart attacks in the workplace.

## 1. Introduction

Occupational accidents result in significant loss of working hours and productivity for companies, and in the worst cases, the death of individuals during working hours [1,2]. These losses entail high economic and social costs for families, employers, enterprises, and society worldwide [3–6]. According to the International Labour Office, occupational safety accounts for 4 % of the global Gross Domestic Product (GDP) each year [7].

Research into occupational accidents helps identify their causes [8], thereby enabling the prevention of similar incidents in the future [9]. Such research has also facilitated the design and development of effective preventive measures [10–13].

Studies of occupational accidents can be conducted using two approaches. The first approach involves analysing accidents in

---

[*] Corresponding author.

*E-mail addresses:* asanchez@uva.es (A. Sánchez-Lite), jofuebar@dpi.upv.es (J.L. Fuentes-Bargues), ivan.iglesias.s@professor.universidadviu.com (I. Iglesias), cggaya@ind.uned.es (C. González-Gaya).

https://doi.org/10.1016/j.heliyon.2024.e37647
Received 31 May 2024; Received in revised form 9 August 2024; Accepted 6 September 2024

individual companies and extrapolating the results to the sectors in which they operate [14,15]. The second approach entails performing statistical analyses on historical accident data collected by government agencies to identify the causes of accidents [16–18]. This approach allows for the design of preventive measures using a more general perspective [19].

The second approach encompasses several lines of research. The first line focuses on the historical analysis of accidents in specific production sectors, such as the manufacturing sector in South Africa [20], the metal sector in Spain [21], and the nuclear sector in India [22]. However, the construction sector is undoubtedly the most studied. Research in this sector includes studies by Camino López et al. in Spain [23], López-Arquillos et al. in Spain [24], Cheng et al. in China [25], Antoniou and Agrafioti in Greece [26], and Kang and Ryu in South Korea [27].

Other research lines focus on specific aspects of the accident rate, such as fatalities in particular areas. For instance, Khodabandeh et al. examined construction workers in Iran [28], Barlas and Izci investigated shipyard workers in Turkey [29], and Strand et al. studied male Norwegian military peacekeepers deployed to Lebanon [30].

The analysis of fatalities in a specific country or geographical area represents another research line, similar to the works of Hansen et al. [31], Asady et al. [32], Kang et al. [33], and Gómez-García et al. [34]. Additionally, some studies have not only analysed historical accidents but have also developed fatal accident prediction models based on significant variables, as demonstrated by Santos et al. in Portugal [35] and Fuentes-Bargues et al. in Spain [36].

In the cited research, occupational accidents are associated with various factors such as personal variables (age, experience, and skills) [17,37,38], work condition variables (deviation, physical agents) [25,39], and the consequences of accidents (severity, type of injury, and part of the body affected) [23,40].

A line of research that has not been widely explored in the scientific literature on occupational accidents is the analysis from the perspective of the deviations that lead to accidents or injuries. For instance, Lafflamme et al. [41] studied overexertion injuries among Swedish nursing auxiliaries, and González-Fuentes et al. [42] focused on cleaning workers. Weisshaar et al. [43] examined dermatitis in professions such as healthcare workers, cleaners, and kitchen employees in Germany, while Violanti et al. [44] addressed law enforcement deaths, including those caused by heart attacks.

A heart attack, as a work-related accident, can result from an accumulation of adverse habits that ultimately lead to heart malfunction and subsequent death or injury. An occupational accident is defined as an event occurring during working hours that results in either a non-fatal injury with loss of working time or a fatal injury [45]. Heart attacks are included in this definition as bodily injuries encompass anatomical, functional, moral, sensory, and psychological injuries. Legally and statistically, a heart attack is considered an occupational accident, distinguishing it from a common disease because it is attributed to work circumstances and/or occurs during working hours and at the workplace [46]. However, such cases often end up in court for legal characterisation [47,48]. Myocardial infarction can result from unusual physical exertion or mental stress [49,50]. When such strain or stress occurs within a work context, it is frequently compensated by the workers' compensation system and regarded as an occupational accident [51].

The research conducted by Fuentes-Bargues et al. [36] revealed that 35 % of fatal work-related accidents in Spain between 2009 and 2021 were caused by heart attacks. This underscores the need for further analysis to identify the specific sectors, age groups, and companies where these fatal accidents occur. To date, no study in Spain has addressed this question, and only limited research has been conducted in other countries. This study aims to address the limitations identified in the existing scientific literature on this topic.

The study has two main objectives. The first is to describe the characteristics of occupational accidents due to heart attacks in the workplace in Spain over the past 15 years. The second is to analyse the main variables influencing such accidents. To achieve these objectives, the following research questions were addressed: When do heart attacks occur at work? Who is at the highest risk of experiencing a heart attack at work? How do the characteristics of the company and labour contracts affect the occurrence of heart attacks at work? Does the climate influence the number of heart attacks at work? Are there differences in the number of heart attacks at work across different regions of Spain? Which sectors in Spain have the highest number of heart attacks at work? What combination of variables could have the greatest influence on the occurrence of heart attacks at work?

The results could be utilised by stakeholders involved in health and safety management to design specific preventive measures aimed at reducing the risk of heart attacks in the workplace, thereby mitigating the social and economic impact of such accidents.

The remainder of this paper is organised as follows: Section 2 reviews the literature on accidents caused by heart attacks, while Section 3 describes the methodology. The results are presented in Section 4, and a discussion of previous research is provided in Section 5. Finally, Section 6 presents the conclusions of the study.

## 2. Literature review

In the limited scientific literature on heart attacks that is not medical in nature, there is no explicit distinction between occupational accidents and diseases. Instead, the focus is on identifying risk factors that may lead to injuries among work population-based samples or through specific investigations of at-risk occupations [44,52–54].

Kivimäki et al. [52] reviewed work stress as a risk factor for coronary heart disease and myocardial infarction. Stress was defined by Gray et al. [55] as "something that occurs when an individual confronts a situation in which his/her usual modes of coping or behaving are insufficient." In their review, Kivimäki et al. [52] compiled results from studies showing that stressful work situations led to an increased incidence of cardiovascular diseases. These situations include a high pace of work [56], long working hours (typically referring to over 48 or 55 h per week) [57], particularly among low socioeconomic status workers [58], unmet expectations between effort and reward [59], and job insecurity, especially during economic recessions [60,61]. The authors also concluded that differences in the impact of work stress on heart disease between men and women, as well as between younger and older workers, were negligible.

Other authors have observed differences in heart problems based on social class, similar to the study by Marmot et al. among British civil servants [62]. Additionally, personal variables such as sex, race, and age have been found to influence heart attack rates. For example, Leigh [53] found a positive linear relationship between heart attack accidents and white male workers, with age being a significant factor, in the USA between 1979 and 1982. Kagamimori et al. [63] conducted a comparative study between Great Britain and Japan, noting a higher frequency of cases around the age of 45, with smoking and other bad habits in the 1980s as possible explanations.

A specific study by Moser and Goldblatt [64] on occupational mortality in women aged 45–59 years in England and Wales found that mortality from coronary heart disease, including heart attacks, was not significant in the main occupations of women in this age group, such as charwomen, office cleaners, maids, canteen assistants, sales managers, and proprietors. However, higher mortality rates were observed among cooks and other less common professions for women in this age group, such as machine tool operators and workers in metal making, jewellery, and electrical production.

Contrary to what might be expected, lower-paid jobs (operators and labourers) had higher rates of heart attacks compared to higher-paid jobs (executives and managers) [54]. Additionally, workers exposed to high concentrations of environmental pollutants, such as professional drivers [65,66], welders [67], miners [68], and metal workers [69], have a higher prevalence and incidence of cardiovascular accidents. Leigh [53] found that the sectors with the highest number of heart attacks and strokes in the USA between 1978 and 1982 were wholesale and retail trade, followed by construction and manufacturing.

Among studies on occupational accidents and diseases within certain professions, heart attacks are identified as one of the leading causes of death for high-risk professionals, such as police officers and firefighters, as well as those under high stress, such as medical staff [70]. In addition to the inherent risks associated with these professions, factors such as poor diet, fatigue, and sleep disorders due to shift work also contribute to the risk [71,72].

Violanti [44], in his 2019 study on a sample of 646 police officer deaths in the USA, found that circulatory diseases were the leading cause of death, with approximately 82 % being heart attacks and an average age of 46.5 years. These findings align with the results obtained by Garbarino and Magnavita [73], who reported an average age range between 40 and 50 years. Several studies have also concluded that male police officers are at higher risk [72,74].

Another high-risk profession is that of seafarers. After deaths due to maritime disasters, occupational deaths due to disease, with heart-related deaths being predominant, are significant. Similar findings have been reported across various countries, including Singapore [75], Denmark, Germany [76], and Hong Kong [77]. Studies that analyse results by gender, such as Hansen et al.'s study on female seafarers [78], have also shown comparable results.

The temporal variables associated with heart attack accidents have not been extensively analysed in the research literature. For instance, Leigh and Miller [54] conducted a study using two national databases in the USA and concluded that heart attacks occurred more frequently on Mondays than on any other day of the week.

The consequences of heart attack accidents are significant, even in sectors with low overall accident rates. For example, Baraza and Cuguero-Escofer [79] conducted a study in the agricultural sector in Spain between 2013 and 2018. Although heart attacks accounted for only 0.17 % of the total number of accidents, they represented 47.42 % of the deaths.

The social and familial impact of a heart attack, especially when it occurs in the workplace, underscores the need for a

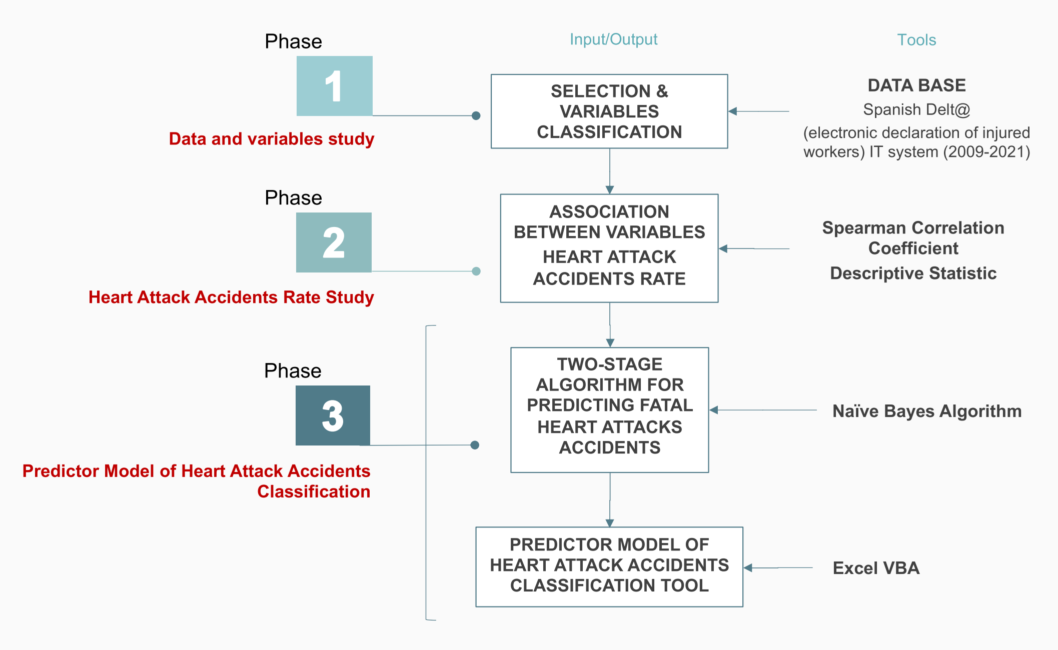


**Fig. 1.** Methodological framework: phases, tools and analysis approach. Source: Own elaboration.

comprehensive analysis of such accidents. Developing tools to identify profiles of occupational variables associated with heart attacks can lead to better health policies and ultimately save lives.

## 3. Methodology

The phases, steps, and analyses performed in this study are illustrated in Fig. 1. All statistical analyses were conducted using IBM SPSS® Statistics (version 29.0) [80]. The study covers the period from 2009 to 2021, beginning with the date of other accident studies conducted in Spain and ending with the latest published consolidated data available at the time of the research.

Phase 1 involved the selection of data and study variables. Data were sourced from the Delta@database, which is structured in accordance with Act TAS/2926, November 21, 2002 [81]. An electronic procedure was used to collect data from work accident notifications, which were provided upon request by the Ministry of Labour and Social Affairs of Spain and were fully anonymised. Between 2009 and 2021, there were 7,739,169 accidents in Spain, of which 15,984 resulted in heart attacks.

The variables related to heart attack accidents selected for this study and their groupings are listed in Table 1. Previous studies across various industrial sectors have used these variables effectively [36,82,83]. Additionally, during this phase, all reported heart attack accidents in Spain were classified as either fatal or nonfatal.

In Phase 2, an initial statistical analysis of the data was conducted to provide an overview of the occurrence of work-related heart attack accidents and to select study variables for developing a predictive model for such accidents. Spearman's rank correlation coefficient was employed to detect dependencies between variables at a 95 % confidence level, with a significance threshold of $<0.05$.

This statistical analysis of heart attack accidents was performed across various variables, including the study period, occupational status, Spanish region where the accident occurred, and the National Classification of Economic Activities (CNAE) [84] (CNAE 2009), which aligns with the European Classification of Economic Activities (NACE). Since data on the number of workers for each variable studied were not available, the analysis focused on the number of accidents rather than incidence rates.

A predictive model for heart attack accidents at work was developed in Phase 3 using the Naïve Bayes Classifier (NBC) [85,86] with the variables selected in Phase 1 (Table 1). The model was implemented using a Visual Basic for Applications (VBA) tool developed in Excel. Naïve Bayes is an effective probabilistic machine learning algorithm used in predictive modelling and classification tasks [87–91]. It is a type of supervised learning that constructs a model for the distribution of class labels with a clear definition of the target attribute. Naïve Bayes relies on Bayes' theorem, which assumes independence between each pair of predictor variables. To develop the model, the samples were sorted by class according to the selected variables (predictors) and then divided into training and validation samples for cross-validation. The objective was to obtain the most efficient and robust model. Once the training and validation samples were defined, the conditional and prior probabilities were used to calculate the posterior probability using Equation (1):

**Table 1**
Classification and description of variables and categories.

| Id | Variable group | Id | Variable | Variable Description | Number categories | Category Values |
|---|---|---|---|---|---|---|
| T | Temporal | T1 | Year | Year of the accident | 13 | 2009 to 2021 |
| | | T5 | Time of the working day | Time of the working day when the accident occurs. | 2 | 1–7:00 to 15:00 2- Rest of hours |
| P | Personal | P1 | Gender | Worker's sex | 2 | 1- Male; 2- Female |
| | | P2 | Age | Worker's age (years old) | 3 | 1- ≤35; 2–36 to 50; 3- >50 |
| | | P3 | Personal Status | Workers in the public or private sector distribution | 3 | 1- Civil servant; 2- Salaried; 3- Self-employed |
| B | Business | B1 | CNAE | Spanish National Classification of all Economic Activities (CNAE), grouped under headings | 8 | 1- Agriculture; 2- Manufacturing; 3- Construction; 4- Commercial; 5- Transport; 6- Health & Social; 7- Administrative, Scientific …; 8- Other Activities |
| | | B2 | Company staff | Company size, in terms of the number of workers. | 3 | 1- ≤ 5; 2- >5 to 25; 3- >25 |
| | | B3 | Length of Service | Length of service of the worker, in terms of months and/or years of experience | 2 | 1- ≤12 month; 2- > 12 months |
| | | B4 | Health and Safety preventive organization | Type of preventive organization regarding health and safety at work | 6 | 1- Entrepreneurial assumption; 2- Own prevention service; 3- External prevention service; 4- Designated workers; 5- Joint prevention service; 6- No preventive organization |
| | | B5 | Employment status | Worker's type of employment status. | 2 | 1- Permanent contracts; 2- Temporary contracts |
| | | B6 | Risk assessment | Risk assessment available at the company | 2 | 1- Yes; 2- No |
| C | Circumstances | C1 | Accident location | Location of accident | 4 | 1- Usual workplace; 2- Moving between work areas; 3- Going to or coming from worksite; 4- Different workplace |
| | | C2 | Usual work | The accident occurs when the worker is carrying out his/her usual work. | 2 | 1- Yes; 2- No |

$$P(C_k|x_1,\ldots x_n) = \frac{p(C_k)\times\prod_{i=1}^{n}p(x_i|C_k)}{\sum_{1}^{k}p(C_k)\times\prod_{i}^{n}p(x_n|C_k)} \tag{1}$$

where $P(C_k|x_{1,\ \ldots n})$ is the posterior probability of class $C_k$ given predictor $x_1, \ldots x_n$; $P(x_i|C_k)$ (conditional probability) is the likelihood of the predictor $x_i$ given the class $C_k$; $P(C_k)$ (class probability) is the prior probability of the class $C_k$, and P(x) is the prior probability of predictor.

Assuming independence between each pair of predictor variables, the denominator in Equation (1) is a scaling factor. The NBC algorithm assigns a class label $y = C_k$ to each sample using Equation (2) (sample Naïve Bayes classifier).

$$y = argmax_k\ p(C_k)\times\prod_{i=1}^{n}p(x_i|C_k) \tag{2}$$

After training, the accuracy of the model was evaluated in terms of the performance of the test dataset, considering the class distribution of the samples from the training and validation phases.

The NBC algorithm was selected due to its advantages [85]: High Speed: NBC operates quickly and efficiently. Accurate Classification: It can predict the classification of a test dataset with high accuracy. Multiclass Prediction: It addresses multiclass prediction problems effectively. Categorical Variables: When the input variables are categorical, NBC shows superior performance compared to numerical variables. Binary and Multiclass Classifications: It is suitable for both binary and multiclass classifications. NBC has demonstrated favourable results compared to other algorithms such as Support Vector Machine (SVM), Random Forest, Logistic

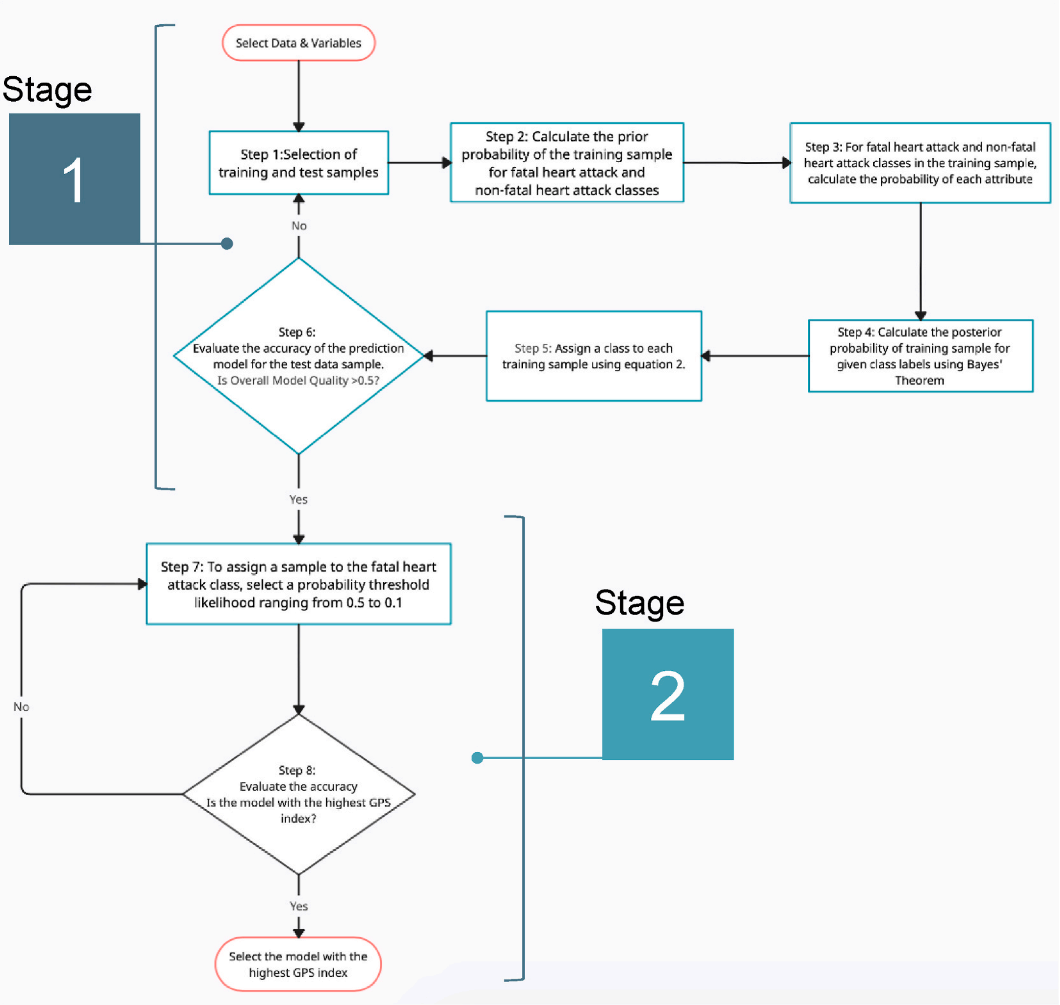


**Fig. 2.** Flowchart of the suggested NBC algorithm. Source: Own elaboration.

Regression, K-Nearest Neighbour, and decision trees [92–95].

The proposed NBC algorithm (Step 3.1) comprises two stages with eight steps (Fig. 2). The initial stages (Steps 1–6) correspond to the steps of the simple NBC algorithm. Each sample was classified as a fatal heart attack accident only if the predicted probability of belonging to the fatal heart attack accident class was higher than that of belonging to the non-fatal heart attack accident class.

However, the samples used in the study were unbalanced. There were a large number of non-fatal heart attack accidents (negative class), which made it difficult to assess how well the initial model performs for fatal heart attacks (positive class). This study focuses on identifying the highest number of true positives (high sensitivity) based on the values of the confusion matrix. Stage 2 aims to detect fatal heart attack accident samples. One strategy for achieving this is to classify as fatal heart attack accidents those samples that have a post hoc probability of belonging to the fatal heart attack accident class of less than 50 % (Step 7). Consequently, more fatal heart attack accident samples were detected at the expense of fewer correctly classified non-fatal heart attack accident samples. Since identifying fatal heart attack accidents is more crucial than identifying non-fatal heart attack accidents, a reduction in the overall classification rate can be accepted, provided that the increase in fatal heart attack accident detection does not lead to a significant reduction in the correct identification of non-fatal heart attack accidents. Therefore, all four values of the confusion matrix should be considered to assess this compromise and the quality of the model (Step 8).

In the scientific literature, several metrics are used to measure the performance of classifier models [96–101]. These metrics include the area under the Receiver Operating Characteristic (ROC AUC) analysis, Precision-Recall Curve (PRC) analysis, Error Rate (ERR), Kappa Score, Accuracy score (ACC), $F_\beta$ Score, Unified Performance Measure (UPM), Sensitivity, Youden's J Index, Precision, Likelihood Ratios (LR), Specificity, Geometric Mean (G-Mean), Class Weighted Accuracy (CWA), Matthews Correlation Coefficient (MCC), Adjusted Rand Index (ARI), Brier Score Loss, and Log Loss. These metrics can be classified into three different groups [102,103]: threshold, ranking, and probabilistic. Some of these metrics, such as F-measure curves [104], MCC [105–107], PRC curve analysis [108,109], UPM [110], and ARI [111], provide better results than the widely used accuracy, error rate, or area under the receiver operating characteristic (AUC ROC) when dealing with unbalanced classification samples. Several authors have pointed out the need to account for data imbalance in each class to identify the most appropriate metric [112,113]. In this context, De Diego et al. [113] recommended using a global index that incorporates multiple metrics and their harmonic mean to define the General Performance Score (GPS). The GPS index ranges from [0, 1], with a value of 1 indicating that all the indices considered in its definition are 1, and a value of 0 indicating that any of them is zero. It is important to note that, while the number of false positives may increase, the model should not significantly compromise its specificity and accuracy. Therefore, all four values of the confusion matrix should be considered when assessing the model [114]. In this line of research, a global index was used, considering the harmonic mean of the metrics MCC (equation (3)), UPM (equation (4)), and ARI (equation (5)). These metrics were calculated using all four values of the confusion matrix. Since the UPM provides values between 0 and 1, the ARI and MCC are normalized to the interval [0, 1]. Equation (6) shows the proposed global index ():

$$MMC = \frac{1}{2} \times \left[ \frac{TP \times TN - TP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} + 1 \right] \tag{3}$$

$$UPM = \frac{4 \times TP \times TN}{4 \times TP \times TN + (TP + TN) \times (FP + FN)} \tag{4}$$

$$ARI = \frac{1}{2} \times \left[ \frac{2 \times (TP \times TN - FP \times FN)}{(TP + FN) \times (FN + TN) + (TP + FP) \times (FP + TN)} + 1 \right] \tag{5}$$

$$GPS = \frac{3}{\frac{1}{MMC} + \frac{1}{UPM} + \frac{1}{ARI}} \tag{6}$$

True positives (TP) are positives that are correctly predicted as positives; false negatives (FN) are positives that are wrongly predicted as negatives; true negatives (TN) are negatives that are correctly predicted as negatives; and false positives (FP) are negatives that are incorrectly predicted as positives.

## 4. Results

### 4.1. Phase 1: Data and variables study: classification and selection

From the set of variables shown in Table 1, the following variables were selected for this study: Time of the Working Day (T5), CNAE (B1), Company Staff (B2), Length of Service (B3), Health and Safety Preventive Organization (HSPO) (B4), Employment Status (B5), Risk Assessment (B6), Accident Location (C1), Usual Work (C2), Gender (P1), Age (P2), and Professional Status (P3). Climate variables were also included. The climate variables have four levels: Continental, Mediterranean, Oceanic, and Tropical. Each level represents one of the climatic zones in Spain according to the classification performed by López-Arquillos et al. [24].

Analysis of the correlations between the variables showed a low or very low correlation between the selected variables. A moderate correlation was observed only for variables B3-B5 and C2-C1.

### 4.2. Phase 2: Heart attack accidents rate study. Descriptive analysis

Between 2009 and 2021, a total of 7,739,169 accidents occurred in Spain, of which 15,984 caused a heart attack, representing a very low percentage (0.207 %). However, if the mortality of this injury (3,165 deaths) is compared to the total number of fatal accidents during the study period (8,974 deaths), it represents a significantly high value (35.3 %). Regarding heart attack accidents, 19.8 % resulted in the death of workers.

Table 2 shows the evolution of accidents associated with heart attacks during the study period according to severity. The rates for total, fatal, very serious, serious, and light accidents were determined and expressed as percentages. The total accident rate (TAR) was obtained by dividing the total number of accidents in category j of variable i by the total number of accidents. The light accident rate (LAR) was obtained by dividing the number of light accidents in category j of variable i by the total number of light accidents. The serious accident rate (SAR) was obtained by dividing the number of serious accidents in category j of variable i by the total number of serious accidents. The very serious accident rate (VSAR) was obtained by dividing the number of very serious accidents in category j of variable i by the total number of very serious accidents. Finally, the fatal accident rate (FAR) was obtained by dividing the number of fatal accidents in category j of variable i by the total number of fatal accidents. The total number of accidents and the proportion of accidents among the different severity groups remained constant over the study period, with an average of approximately 240 fatalities per year.

With regard to sex, 13,454 accidents occurred in men (84.2 %) and 2,531 in women (15.8 %); this proportion was higher in the case of fatal accidents, 92.3 % in men (2,921 deaths), and 7.7 % in women (244 deaths).

When accidents leading to heart attacks were analysed according to the nationality of the workers, most of the workers were Spanish (94 %). The remaining nationalities showed residual values, with the main ones being Romania (1.4 %), and Portugal, Morocco, and Bulgaria (0.6 % each). Another point of view is the distribution of heart attack accidents according to professional status (Table 3). Notably, the number of fatal heart attack accidents was proportionally reduced (11.7 %) compared to the total number of heart attack accidents (22.2 %) in the case of public workers.

A description of heart attack accidents concerning time variables shows that there is an even distribution of accidents from Monday to Friday, with slightly higher values on Mondays in terms of both total accidents (3,161 accidents, 19.8 %) and fatalities (649 deaths, 20.5 %). The hours of the day when the highest number of heart attacks occurred were between 8:00 and 13:00 (48.3 % of total accidents and 39.7 % of fatal accidents). The highest number of accidents tended to occur in the first 4 h of the working day (64.1 % of total accidents and 57.7 % of fatal accidents).

In the last two columns of Table 4, the percentage of Fatal Heart Attack Accidents (FHAA) relative to the Total Heart Attack Accidents (THAA) and Fatal Accidents (FA) during the study period for each autonomous community is calculated. These values were compared with the mean for Spain during the study period (19.8 % FHAA/THAA and 35.3 % FHAA/FA). The high mortality rates when a heart attack occurred in the communities of the Balearic Islands (27.94 %), Valencia (24.31 %), and Castilla-La Mancha (23.15 %) should be noted. It is also worth noting the importance of mortality due to heart attacks in occupational accidents in the Canary Islands (45.73 %), Madrid (45.70 %), Melilla (42.86 %), and the Basque Country (41.45 %).

The evolution of heart attack accidents over time according to the CNAE is presented in Table 5. The economic sectors with the highest number of heart attacks are, in the first place, the "Health and Social Work Activities" sector (2,311 accidents), followed by "Transport and Warehousing" (2,245 accidents), "Manufacturing Industry" (2,098 accidents), and "Construction" (1,624 accidents). It is also important to note the subcategories "Non-perennial Crops" (86 deaths) and "Perennial Crops" (71 deaths), both belonging to the agricultural sector, where 43.4 % and 36.0 % of heart attacks, respectively, were fatal.

Table 6 provides a more detailed analysis of the CNAE (with the three-digit classification), identifying the subcategories with more than 300 accidents in the left columns and with over 60 deaths in the right columns during the study period to identify more specific subcategories that might lose their importance if a more general classification is analysed.

**Table 2**
Evolution of heart attack accidents by severity.

| Year | Total accidents | | Light | | Serious | | Very serious | | Fatal | |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | TAR(%) | N | LAR(%) | N | SAR(%) | N | VSAR(%) | N | FAR(%) |
| | 15,984 | 100 | 6,951 | 100 | 5,302 | 100 | 566 | 100 | 3,165 | 100 |
| 2009 | 1,209 | 7.6 | 437 | 6.3 | 435 | 8.2 | 73 | 12.9 | 264 | 8.3 |
| 2010 | 1,301 | 8.1 | 495 | 7.1 | 505 | 9.5 | 56 | 9.9 | 245 | 7.7 |
| 2011 | 1,198 | 7.5 | 479 | 6.9 | 427 | 8.1 | 51 | 9.0 | 241 | 7.6 |
| 2012 | 1,179 | 7.4 | 497 | 7.2 | 439 | 8.3 | 54 | 9.5 | 189 | 6.0 |
| 2013 | 1,312 | 8.2 | 564 | 8.1 | 472 | 8.9 | 53 | 9.4 | 223 | 7.0 |
| 2014 | 1,211 | 7.6 | 542 | 7.8 | 402 | 7.6 | 43 | 7.6 | 224 | 7.1 |
| 2015 | 1,163 | 7.3 | 511 | 7.4 | 374 | 7.1 | 44 | 7.8 | 234 | 7.4 |
| 2016 | 1,246 | 7.8 | 549 | 7.9 | 382 | 7.2 | 44 | 7.8 | 271 | 8.6 |
| 2017 | 1,214 | 7.6 | 552 | 7.9 | 381 | 7.2 | 25 | 4.4 | 256 | 8.1 |
| 2018 | 1,326 | 8.3 | 627 | 9.0 | 394 | 7.4 | 38 | 6.7 | 267 | 8.4 |
| 2019 | 1,358 | 8,5 | 652 | 9.4 | 440 | 8.3 | 26 | 4.6 | 240 | 7.6 |
| 2020 | 1,120 | 7.0 | 519 | 7.5 | 320 | 6.0 | 27 | 4.8 | 254 | 8.0 |
| 2021 | 1,147 | 7.2 | 527 | 7.6 | 331 | 6.2 | 32 | 5.7 | 257 | 8.1 |

**Table 3**
Heart attacks accidents (total and fatal) by professional status (2009–2021).

| Professional status | Total accidents | | Fatal accidents | |
|---|---|---|---|---|
| | N | TAR(%) | N | FAR(%) |
| | 15,984 | 100 | 6,951 | 100 |
| Private sector employees | 12,040 | 75.3 | 3,165 | 85.6 |
| Public sector employees | 3,548 | 22.2 | 370 | 11.7 |
| Self-employed with and without workers | 396 | 2.5 | 85 | 2.7 |

**Table 4**
Heart attacks accidents by autonomous communities (2009–2021).

| Autonomous communities | Total Heart Attack Accidents (THAA) | | Autonomous communities | Fatal Accidents (FA) | | Autonomous communities | Fatal Heart Attack Accidents (FHAA) | | FHAA/THAA | FHAA/FA |
|---|---|---|---|---|---|---|---|---|---|---|
| | N | % | | N | % | | N | % | % | % |
| | 15,984 | 100 | | 8,974 | 100 | | 3,165 | 100 | | |
| Andalusia | 3,034 | 19.0 | Andalusia | 1,415 | 15.8 | Andalusia | 524 | 16.6 | 17.27 | 37.03 |
| Aragon | 730 | 4.6 | Aragon | 369 | 4.1 | Aragon | 128 | 4.0 | 17.53 | 34.69 |
| Asturias | 712 | 4.5 | Asturias | 241 | 2.7 | Asturias | 81 | 2.6 | 11.38 | 33.61 |
| Balearic Islands | 136 | 0.9 | Balearic Islands | 123 | 1.4 | Balearic Islands | 38 | 1.2 | 27.94 | 30.89 |
| Canary Islands | 718 | 4.5 | Canary Islands | 293 | 3.3 | Canary Islands | 134 | 4.2 | 18.66 | 45.73 |
| Cantabria | 202 | 1.3 | Cantabria | 130 | 1.4 | Cantabria | 40 | 1.3 | 19.80 | 30.77 |
| Catalonia | 2,023 | 12.7 | Catalonia | 1,358 | 15.1 | Catalonia | 434 | 13.7 | 21.45 | 31.96 |
| Castille Leon | 1,003 | 6.3 | Castille Leon | 642 | 7.2 | Castille Leon | 208 | 6.6 | 20.74 | 32.40 |
| Castille La Mancha | 527 | 3.3 | Castille La Mancha | 462 | 5.1 | Castille La Mancha | 122 | 3.9 | 23.15 | 26.41 |
| Valencia | 1,296 | 8.1 | Valencia | 886 | 9.9 | Valencia | 315 | 10.0 | 24.31 | 35.55 |
| Extremadura | 436 | 2.7 | Extremadura | 220 | 2.5 | Extremadura | 86 | 2.7 | 19.72 | 39.09 |
| Galicia | 1,299 | 8.1 | Galicia | 818 | 9.1 | Galicia | 248 | 7.8 | 19.09 | 30.32 |
| La Rioja | 173 | 1.1 | La Rioja | 79 | 0.9 | La Rioja | 28 | 0.9 | 16.18 | 35.44 |
| Madrid | 2,010 | 12.6 | Madrid | 965 | 10.8 | Madrid | 441 | 13.9 | 21.94 | 45.70 |
| Murcia | 478 | 3.0 | Murcia | 328 | 3.7 | Murcia | 94 | 3.0 | 19.67 | 28.66 |
| Navarre | 245 | 1.5 | Navarre | 160 | 1.8 | Navarre | 44 | 1.4 | 17.96 | 27.50 |
| Basque Country | 872 | 5.5 | Basque Country | 468 | 5.2 | Basque Country | 194 | 6.1 | 22.25 | 41.45 |
| Ceuta | 47 | 0.3 | Ceuta | 10 | 0.1 | Ceuta | 3 | 0.1 | 6.38 | 30.00 |
| Melilla | 43 | 0.3 | Melilla | 7 | 0.1 | Melilla | 3 | 0.1 | 6.98 | 42.86 |

The subcategories with the most heart attack accidents were “Hospital Activities,” “Road Haulage and Removal Services,” and “Public Administration and Economic and Social Policy,” each with more than 1,000 accidents. Regarding fatal accidents, “Road Haulage and Removal Services” recorded 323 deaths, and “Public Administration and Economic and Social Policy” recorded 200 deaths during the study period.

### *4.3. Phase 3: Predictor model of heart attack accidents classification. Navïe Bayes algorithm and class prediction tool development*

Following the two-stage algorithm described in the previous section, the results obtained using the developed NBC model are presented. The study analysed 15,616 valid samples, of which 80.3 % were non-fatal heart attack accidents (HAA) and 19.7 % were fatal heart attack accidents (FHAA).

Five different NBC models were run to determine the optimal sizes of the training and test datasets, with varying training-test sample ratios of 80–20, 70–30, 50-50, 30–70, and 20–80 in the first stage. These five models classified a sample as positive for the FHAA only if the posterior probability for that sample was higher in the FHAA class than in the negative class (HAA). Fig. 4 presents the results. The dataset was ultimately split into a training sample of 70 % and a test sample of 30 % (Split 70_30 model) based on the values of the confusion matrix (see Fig. 4a), the GPS index value, and the lower bound of the confidence interval for the estimate of the area under the ROC curve (AUC ROC) for each model, with a 95 % confidence interval. The AUC ROC and GPS index values of the Split 70-30 model were higher than 0.5 (Fig. 4b).

The final selection of the NBC model occurs in the second stage. Eight NBC models were developed to detect a high number of positive cases (FHAA). A sample was assigned to the FHAA class if the posterior probability of belonging to that class, derived from the selected model in stage 1 (split 70-30 %), met or exceeded thresholds of 0.45, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, and 0.10. The selection process involved analysing the confusion matrix values and the metrics outlined in the methodology section. The model with a threshold of 0.25 (Severity 25) was deemed the most appropriate. This decision was based on the confusion matrix values, the GPS index, and the lower bound of the confidence interval for the area under the ROC curve (AUC ROC), all evaluated with a 95 % confidence interval. The AUC ROC and GPS indices of the Severity 25 model were higher than 0.5. The results are shown in Figs. 5–7.

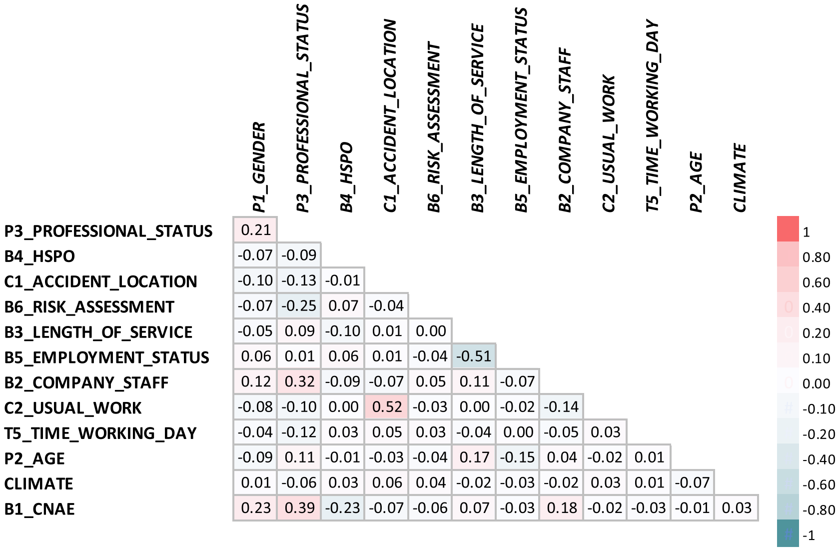
**Table 5**

Trend of the number of fatal accidents by CNAE with higher number of heart attack accidents (2009–2019).

| CNAE | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | TOTAL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Agriculture, livestock, forestry, and fisheries | 46 | 48 | 51 | 62 | 56 | 66 | 67 | 70 | 65 | 67 | 72 | 68 | 59 | 797 |
| Manufacturing Industry | 180 | 168 | 174 | 149 | 184 | 165 | 132 | 133 | 156 | 172 | 174 | 169 | 142 | 2,098 |
| Construction | 150 | 154 | 133 | 100 | 99 | 86 | 112 | 117 | 113 | 133 | 149 | 134 | 144 | 1,624 |
| Wholesale and retail trade; repair of motor vehicles and motorbikes | 118 | 119 | 126 | 127 | 135 | 112 | 124 | 135 | 109 | 129 | 127 | 104 | 128 | 1,593 |
| Transport and warehousing | 151 | 147 | 146 | 148 | 185 | 149 | 152 | 205 | 172 | 205 | 218 | 171 | 196 | 2,245 |
| Health and social work activities | 181 | 207 | 157 | 212 | 217 | 192 | 161 | 181 | 161 | 186 | 178 | 132 | 146 | 2,311 |
| Real estate, professional, scientific, technical, administrative, and support service activities | 106 | 142 | 118 | 104 | 115 | 99 | 102 | 108 | 114 | 112 | 101 | 101 | 77 | 1,399 |
| Rest of activities | 277 | 316 | 293 | 277 | 321 | 342 | 313 | 297 | 324 | 322 | 339 | 241 | 255 | 3,917 |

**Table 6**
Total and fatal heart attack accidents by CNAE (three-digit classification).

| CNAE | Description | Total accidents | | CNAE | Description | Fatal accidents | |
|---|---|---|---|---|---|---|---|
| | | N | TAR (%) | | | N | FAR (%) |
| 861 | Hospital activities | 1,639 | 10.3 | 494 | Road haulage and removal services | 323 | 10.2 |
| 494 | Road haulage and removal services | 1,287 | 8.1 | 841 | Public administration and economic and social policy | 200 | 6.3 |
| 841 | Public administration and economic and social policy | 1,148 | 7.2 | 432 | Electrical, plumbing, and other installations on construction sites | 98 | 3.1 |
| 493 | Other land passenger transport | 393 | 2.5 | 11 | Non-perennial crops | 86 | 2.7 |
| 412 | Building Construction | 622 | 3.9 | 12 | Perennial crops | 71 | 2.2 |
| 432 | Electrical, plumbing, and other installations on construction sites | 396 | 2.5 | 801 | Private security activities | 71 | 2.2 |
| 812 | Cleaning activities | 350 | 2.2 | 463 | Wholesale trade in food, beverages, and tobacco products | 67 | 2.1 |
| 801 | Private security activities | 298 | 1.9 | 493 | Other land passenger transport | 64 | 2.0 |
| 463 | Wholesale trade of food, beverages, and tobacco | 296 | 1.9 | 812 | Cleaning activities | 61 | 1.9 |



| | P1_GENDER | P3_PROFESSIONAL_STATUS | B4_HSPO | C1_ACCIDENT_LOCATION | B6_RISK_ASSESSMENT | B3_LENGTH_OF_SERVICE | B5_EMPLOYMENT_STATUS | B2_COMPANY_STAFF | C2_USUAL_WORK | T5_TIME_WORKING_DAY | P2_AGE | CLIMATE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **P3_PROFESSIONAL_STATUS** | 0.21 | | | | | | | | | | | |
| **B4_HSPO** | -0.07 | -0.09 | | | | | | | | | | |
| **C1_ACCIDENT_LOCATION** | -0.10 | -0.13 | -0.01 | | | | | | | | | |
| **B6_RISK_ASSESSMENT** | -0.07 | -0.25 | 0.07 | -0.04 | | | | | | | | |
| **B3_LENGTH_OF_SERVICE** | -0.05 | 0.09 | -0.10 | 0.01 | 0.00 | | | | | | | |
| **B5_EMPLOYMENT_STATUS** | 0.06 | 0.01 | 0.06 | 0.01 | -0.04 | -0.51 | | | | | | |
| **B2_COMPANY_STAFF** | 0.12 | 0.32 | -0.09 | -0.07 | 0.05 | 0.11 | -0.07 | | | | | |
| **C2_USUAL_WORK** | -0.08 | -0.10 | 0.00 | 0.52 | -0.03 | 0.00 | -0.02 | -0.14 | | | | |
| **T5_TIME_WORKING_DAY** | -0.04 | -0.12 | 0.03 | 0.05 | 0.03 | -0.04 | 0.00 | -0.05 | 0.03 | | | |
| **P2_AGE** | -0.09 | 0.11 | -0.01 | -0.03 | -0.04 | 0.17 | -0.15 | 0.04 | -0.02 | 0.01 | | |
| **CLIMATE** | 0.01 | -0.06 | 0.03 | 0.06 | 0.04 | -0.02 | -0.03 | -0.02 | 0.03 | 0.01 | -0.07 | |
| **B1_CNAE** | 0.23 | 0.39 | -0.23 | -0.07 | -0.06 | 0.07 | -0.03 | 0.18 | -0.02 | -0.03 | -0.01 | 0.03 |

**Fig. 3.** Spearman's correlation coefficients chart of the predictor variables used in the study. Source: Own elaboration.

The NBC procedure does not elucidate how the predictors affect the probability of a predicted response. However, examining the cross-reference tables between the predictors and the predicted values can offer a meaningful understanding. Tables 7 and 8 present the results of the contingency table analysis for each predictor variable and the predicted class values for the Severity 25 NBC model, respectively. The highest values within each category of the predictor variables are highlighted in bold.

A VBA Excel application was developed based on the selected NBC model and its results. This application allows for the selection of predictor variable levels and classifies samples into the FHAA or HAA categories. Fig. 8 displays the results obtained from applying two different samples.

Using the VBA Excel application, the combination of Gender (female), Professional Status (Civil servant), HSPO (No prevention organisation), Accident Location (Usual workplace), Length of Service (>12 months), Risk Assessment (Yes), Employment Status (Permanent contracts), Company Staff (>25), Usual Work (Usual workplace), Time of Work Day (07:00 to 15:00 h), Age (>50), Climate (Tropical), and CNAE (Real estate, professional, scientific, technical, administrative, and support service activities) predicts a probability of 0.99 for the non-fatal heart attack accident (HAA) class. Conversely, the probability of the Fatal Heart Attack Accident (FHAA) class is 0.91 for Gender (male), Professional Status (Self-employed), HSPO (Own prevention service), Accident Location (Moving between areas), Length of Service (<12 months), Risk Assessment (No), Employment Status (Temporary contracts), Company Staff (5–25), Usual Work (Usual work), Time of Work Day (Rest of hours), Age (>50), Climate (Mediterranean), and CNAE (Transport).

## 5. Discussion

Regarding the impact of company characteristics and employment contracts on the incidence of workplace heart attacks, the results

(a)

(b)

**Fig. 4.** Initial NBC model selection. (a) Values of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each model. (b) Values of the quality metrics used Source: Own elaboration.

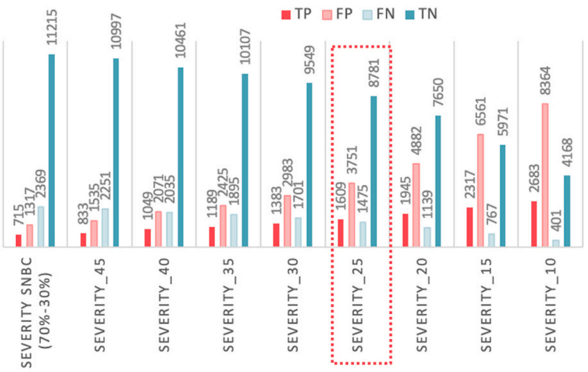


**Fig. 5.** Final NBC model selection. Values of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) for each model. Source: Own elaboration.

align with the distribution of workers across the public and private sectors. However, it is noteworthy that mortality rates do not follow the same proportionality, being higher among private sector employees (85.6 % fatal heart attacks and 75.3 % total heart attacks). Several studies have demonstrated a link between physical workload [115–117], working conditions [118–121], awareness of cardiac symptoms and appropriate responses [122,123], and the risk of heart attacks. This observation may be connected to how these factors interact with the work performed in different sectors.

In addressing which sectors in Spain have a higher incidence of myocardial infarctions at work, the study reveals that the sectors with the highest number of heart attacks are 'Health and social work', followed by 'Transport and storage', 'Manufacturing', and 'Construction'. This finding partially corroborates other studies linking risk factors for heart attacks to specific economic sectors [124,125]. Conversely, when comparing the data on total and fatal heart attacks, it is notable that of the 1,639 accidents in the "861 Hospital

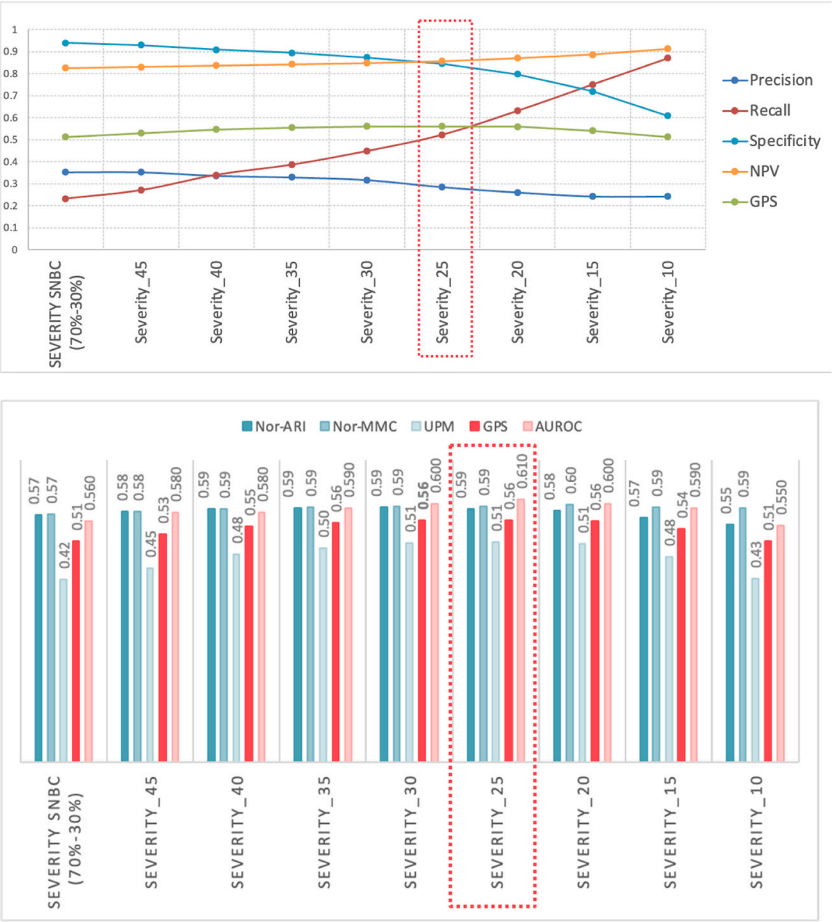


**Fig. 6.** Final NBC model selection. Values of the quality metrics used. Source: Own elaboration.

activities" sub-category, only 55 were fatal (3.4 %). This may be expected, as hospital services are more likely to detect and respond to heart attacks promptly. In contrast, the agricultural sector shows a lower overall incidence of heart attack accidents but a higher percentage of fatal incidents, as previously reported by Cuguero-Escofer et al. in Spain [79].

Regarding the question of who is most at risk of a heart attack, this study found a lower incidence rate of heart attacks and fatal heart attack accidents among females compared to males, which is consistent with findings from other research [50,60,126,127]. The study also revealed that the age group with the highest number of workplace heart attacks was individuals over the age of 50. The World Health Organization (WHO) classifies Spain as having a low risk of cardiovascular mortality; however, it also notes an increased risk for individuals over 50 years old [128].

In addressing when myocardial infarctions occur at work, the results related to the time variable suggest that the number of early-morning and Monday myocardial infarctions reported could be associated with the "weekday effect" or "Monday effect," as described by other researchers in Spain [129] and internationally [51,130].

In response to the research question regarding whether there are differences in the number of heart attacks at work across different regions of Spain, the geographical distribution of fatal accidents is a significant factor [131]. Spain is divided into 17 autonomous communities and two autonomous cities (Ceuta and Melilla). Although no direct correlation has been established between the number of employees and the number of accidents per year in each autonomous community, it can be generally stated that there is a relationship between the active population [132] and the incidence of heart attacks: the more active the population, the more heart attacks tend to occur.

Table 8 indicates that the number of FHAA is highest in the Mediterranean climate, but the distribution of heart attack accidents (HAA-FHAA) is relatively balanced across all climates. The available data suggest that further investigation may be necessary to fully address the research question concerning the climate variable. Additionally, scientific literature points to high daytime temperatures and the impact of climate change as potential risk factors for fatal heart attacks [133–135].

The VBA Excel application assists in determining which combinations of predictor variable levels forecast a higher probability for each class. It can be interpreted that a man over 50 years old, working in a small company with its own risk-prevention system but without a specific risk-prevention programme, being self-employed, with less than a year of experience in the same company, and

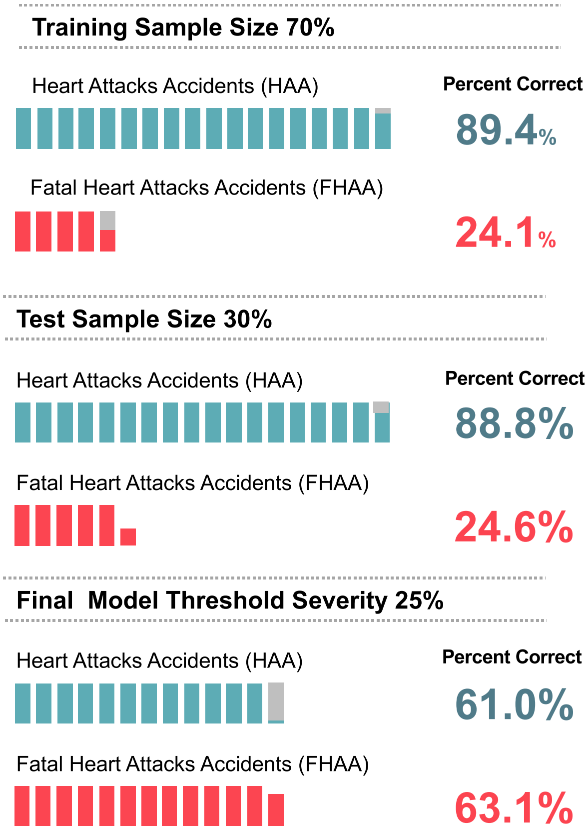


**Fig. 7.** Final NBC model selection. Comparison of initial (training and test samples) and final NBC models in terms of the percentage of true correct cases for each class. Source: Own elaboration.

working in the transport sector within a Mediterranean climate could represent the profile with the highest risk of suffering a fatal heart attack accident (FHAA).

Additionally, it is noteworthy that the analysis of heart attack accidents at work during the study period, excluding the numerous fatal accidents, shows that the average duration of sick leave is 160.3 d, with a standard deviation of 114.5 (skewness coefficient of 0.789 and kurtosis of 0.503). Consistent with other studies [136–140], this result highlights the dual social and economic incentives for companies to implement measures for detecting coronary heart disease in their workers.

Regarding the proposed Naive Bayes model, the class distribution of the dataset of valid samples (80 % HAA and 20 % FHAA) was consistent with that of the dataset available in the database, which recorded the total number of heart attack accidents for the period under consideration. The distribution of samples between the training and test datasets can influence the quality of the Naive Bayes model, and several studies recommend examining the extent of data imbalance [112,113,141,142] This is particularly crucial when dealing with imbalanced data [143]. In this study, the ratio between positive and negative cases (IR) was 4.05, which is considered slightly unbalanced ($IR < 9$) [126]. To assess the impact of the training-test ratio, five models were developed. The GPS index values and the area under the ROC curve did not reveal significant differences between the models (see Fig. 4b). It is important to note that lowering the threshold for positive results will increase the number of false positives and decrease the number of true negatives detected. Several studies have suggested that the 70–30 ratio is typically associated with favourable outcomes [144–146]. Taking all these factors into account, the Split 70-30 model was chosen as the output of the first stage of the proposed Naive Bayes classifier algorithm.

To detect the maximum number of positive cases (FHAA), the proposed two-stage NBC algorithm incorporates a second stage. Considering the imbalance of the dataset and adhering to the recommendations of several authors [103,113,147,148], a GPS index comprising three indicators that account for the four values of the Confusion Matrix (MMC, UPM, and ARI) was employed. The GPS index reached its highest value (0.56) in four of the eight proposed models: Severity_35, Severity_30, Severity_25, and Severity_20 (see Fig. 6). Severity_25 was chosen as the final model due to its superior values for the ARI and UPM indices. Additionally, this model exhibited the highest area under the ROC curve (0.61). However, the validity of using this metric with unbalanced data is contested by some researchers [107–109,149,150]. The model selected in stage 2 (Severity_25) correctly identified 39 % more true-positive cases compared to the model selected in stage 1 (Split 70-30 %). Conversely, it accurately identified 28 % fewer true-negative cases (see Fig. 7). This model represents a balance between various values of the confusion matrix.

However, the independence condition assumed by the Naive Bayes classifier (NBC) model, along with the low correlation between

**Table 7**
Cross-reference tables between B3, P3, B4, C1, C2, B5, and B2 predictor variables and predicted class value for Severity 25 NBC model.

| **B3_LENGTH_OF_SERVICE** | Severity_25 | | | | Total | |
|---|---|---|---|---|---|---|
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| <12 Months | 1,375 | 13.40 % | 2,461 | 45.90 % | 3,836 | 24.60 % |
| **>12 Months** | 8,881 | 86.60 % | 2,899 | **54.10 %** | 11,780 | **75.40 %** |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **P3_PROFESSIONAL STATUS** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Salaried** | 6,858 | 66.90 % | 5,182 | **96.70 %** | 12,040 | **77.10 %** |
| Civil servant | 3,385 | 33.00 % | 163 | 3.00 % | 3,548 | 22.70 % |
| Self-employed | 13 | 0.10 % | 15 | 0.30 % | 28 | 0.20 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **B4_HSPO** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **External Prevention Service** | 10,183 | 99.30 % | 5,026 | **93.80 %** | 15,209 | **97.40 %** |
| Joint Prevention Service | 13 | 0.10 % | 14 | 0.30 % | 27 | 0.20 % |
| Own Prevention Service | 1 | 0.00 % | 101 | 1.90 % | 102 | 0.70 % |
| Designated Workers | 28 | 0.30 % | 219 | 4.10 % | 247 | 1.60 % |
| No Preventive Organization | 31 | 0.30 % | 0 | 0.00 % | 31 | 0.20 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **C1_ACCIDENT_LOCATION** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Usual workplace** | 9,556 | 93.20 % | 2,209 | **41.20 %** | 11,765 | **75.30 %** |
| Moving between areas | 290 | 2.80 % | 2,140 | 39.90 % | 2,430 | 15.60 % |
| Going to or coming from worksite | 112 | 1.10 % | 407 | 7.60 % | 519 | 3.30 % |
| Different workplace | 298 | 2.90 % | 604 | 11.30 % | 902 | 5.80 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **C2_USUAL_WORK** | **Severity_25** | | | | **Total** | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Usual Workplace** | 9,854 | 96.10 % | 2,813 | **52.50 %** | 12,667 | **81.10 %** |
| Not Usual Workplace | 402 | 3.90 % | 2,547 | 47.50 % | 2,949 | 18.90 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **B5_EMPLOYMENT_STATUS** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Permanent contracts** | 8,538 | 83.20 % | 3,067 | **57.20 %** | 11,605 | **74.30 %** |
| Temporary contracts | 1,718 | 16.80 % | 2,293 | 42.80 % | 4,011 | 25.70 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **B2_COMPANY_STAFF** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| ≤5 | 849 | 8.30 % | 1,307 | 24.40 % | 2,156 | 13.80 % |
| >5 to 25 | 1,514 | 14.80 % | 1,932 | 36.00 % | 3,446 | 22.10 % |
| **> 25** | 7,893 | 77.00 % | 2,121 | **39.60 %** | 10,014 | **64.10 %** |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |

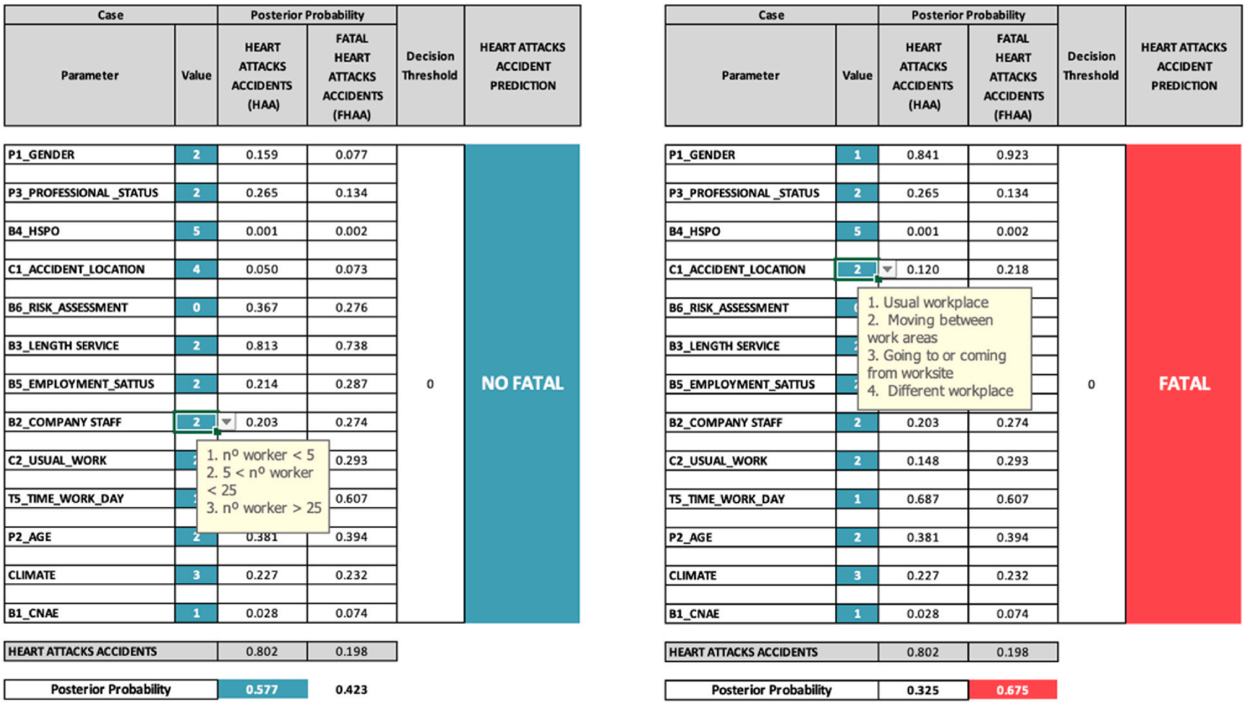
variables, mitigates issues of overfitting [151]. Fig. 3 illustrates the very low correlation among the variables in the model. The developed model showed a negligible difference between the total percentage of correct predictions for the training sample (73.30 %) and the test sample (73.2 %). The Brier Score Loss (BSL) index [152,153] is a probabilistic metric that penalises overfitting [154]. The BSL index value for the FHAA class (0.159) indicates that the model is acceptably calibrated, suggesting no significant overfitting problems.

**Table 8**

Cross-reference tables between P1, B6, T5, P2, B1, and Climate predictor variables and predicted class value for Severity 25 NBC model.

| P1_GENDER | Severity_25 | | | | Total | |
|---|---|---|---|---|---|---|
| | 0 | | 1 | | | |
| | N | % | N | % | N | % |
| **Male** | 7,833 | 76.40 % | 5,277 | **98.50 %** | 13,110 | **84.00 %** |
| Female | 2,423 | 23.60 % | 83 | 1.50 % | 2,506 | 16.00 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **B6_RISK_ASSESSEMENT** | **Severity_25** | | | | **Total** | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| No | 3,880 | 37.80 % | 1,445 | 27.00 % | 5,325 | 34.10 % |
| **Yes** | 6,376 | 62.20 % | 3,915 | **73.00 %** | 10,291 | **65.90 %** |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **T5_TIME_WORKING DAY** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **7:00 to 15:00 Hours** | 7,562 | 73.70 % | 3,136 | **58.50 %** | 10,698 | **68.50 %** |
| Rest of Hours | 2,694 | 26.30 % | 2,224 | 41.50 % | 4,918 | 31.50 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **P2_AGE** | **Severity_25** | | | | **Total** | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| ≤35 | 486 | 4.70 % | 236 | 4.40 % | 722 | 4.60 % |
| 36–50 | 3,410 | 33.20 % | 2,225 | 41.50 % | 5,635 | 36.10 % |
| **> 50** | 6,360 | 62.00 % | 2,899 | **54.10 %** | 9,259 | **59.30 %** |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **B1_CNAE** | **Severity_25** | | | | Total | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Agriculture, livestock, forestry, and fisheries** | 27 | 0.30 % | 696 | **13.00 %** | 723 | 4.60 % |
| Manufacturing Industry | 1,569 | 15.30 % | 504 | 9.40 % | 2,073 | 13.30 % |
| Construction | 436 | 4.30 % | 1,140 | **21.30 %** | 1,576 | 10.10 % |
| Commercial activities | 988 | 9.60 % | 543 | 10.10 % | 1,531 | 9.80 % |
| **Transport** | 861 | 8.40 % | 1,320 | **24.60 %** | 2,181 | 14.00 % |
| **Health and social work activities** | 2,300 | 22.40 % | 7 | 0.10 % | 2,307 | 14.80 % |
| Real estate, professional, scientific, technical, administrative, and support service activities | 845 | 8.20 % | 516 | 9.60 % | 1,361 | 8.70 % |
| **Rest of activities** | 3,230 | 31.50 % | 634 | 11.80 % | 3,864 | **24.70 %** |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |
| **CLIMATE** | **Severity_25** | | | | **Total** | |
| | **0** | | **1** | | | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| Continental | 1,244 | 12.10 % | 556 | 10.40 % | 1,800 | 11.50 % |
| **Mediterranean** | 6,332 | 61.70 % | 3,427 | **63.90 %** | 9,759 | **62.50 %** |
| Oceanic | 2,368 | 23.10 % | 1,219 | 22.70 % | 3,587 | 23.00 % |
| Tropical | 312 | 3.00 % | 158 | 2.90 % | 470 | 3.00 % |
| Total | 10,256 | 100 % | 5,360 | 100 % | 15,616 | 100 % |

Regarding the study's limitations, it is important to acknowledge the absence of data on the personal habits or health status of the casualties. Additionally, there are limitations associated with the NBC algorithm itself [151,155,156]. Although most variables showed a low correlation (Fig. 3), the model retains variables such as Length of Service-Personal Status and Usual Work-Accident Location despite their moderate correlation to ensure comprehensive information retention. Data imbalance may also be a limitation of this study. In this case, the imbalance was considered minimal. Previous studies have addressed similar issues by developing NBC models with attribute-weighted methods [157,158] or dependency-oriented aggregation methods [159], and by employing under-sampling and over-sampling balancing class algorithms to enhance models with imbalanced datasets [143,160–164].



| Case | | Posterior Probability | | | |
|---|---|---|---|---|---|
| Parameter | Value | HEART ATTACKS ACCIDENTS (HAA) | FATAL HEART ATTACKS ACCIDENTS (FHAA) | Decision Threshold | HEART ATTACKS ACCIDENT PREDICTION |
| P1_GENDER | 2 | 0.159 | 0.077 | 0 | NO FATAL |
| P3_PROFESSIONAL _STATUS | 2 | 0.265 | 0.134 | | |
| B4_HSPO | 5 | 0.001 | 0.002 | | |
| C1_ACCIDENT_LOCATION | 4 | 0.050 | 0.073 | | |
| B6_RISK_ASSESSMENT | 0 | 0.367 | 0.276 | | |
| B3_LENGTH SERVICE | 2 | 0.813 | 0.738 | | |
| B5_EMPLOYMENT_SATTUS | 2 | 0.214 | 0.287 | | |
| B2_COMPANY STAFF | 2 | 0.203 | 0.274 | | |
| C2_USUAL_WORK | 2 | [illegible] | 0.293 | | |
| T5_TIME_WORK_DAY | 1 | [illegible] | 0.607 | | |
| P2_AGE | 2 | 0.381 | 0.394 | | |
| CLIMATE | 3 | 0.227 | 0.232 | | |
| B1_CNAE | 1 | 0.028 | 0.074 | | |
| HEART ATTACKS ACCIDENTS | | 0.802 | 0.198 | | |
| Posterior Probability | | 0.577 | 0.423 | | |

1. nº worker < 5
2. 5 < nº worker < 25
3. nº worker > 25

| Case | | Posterior Probability | | | |
|---|---|---|---|---|---|
| Parameter | Value | HEART ATTACKS ACCIDENTS (HAA) | FATAL HEART ATTACKS ACCIDENTS (FHAA) | Decision Threshold | HEART ATTACKS ACCIDENT PREDICTION |
| P1_GENDER | 1 | 0.841 | 0.923 | 0 | FATAL |
| P3_PROFESSIONAL _STATUS | 2 | 0.265 | 0.134 | | |
| B4_HSPO | 5 | 0.001 | 0.002 | | |
| C1_ACCIDENT_LOCATION | 2 | 0.120 | 0.218 | | |
| B6_RISK_ASSESSMENT | 0 | [illegible] | [illegible] | | |
| B3_LENGTH SERVICE | 2 | [illegible] | [illegible] | | |
| B5_EMPLOYMENT_SATTUS | 2 | [illegible] | [illegible] | | |
| B2_COMPANY STAFF | 2 | 0.203 | 0.274 | | |
| C2_USUAL_WORK | 2 | 0.148 | 0.293 | | |
| T5_TIME_WORK_DAY | 1 | 0.687 | 0.607 | | |
| P2_AGE | 2 | 0.381 | 0.394 | | |
| CLIMATE | 3 | 0.227 | 0.232 | | |
| B1_CNAE | 1 | 0.028 | 0.074 | | |
| HEART ATTACKS ACCIDENTS | | 0.802 | 0.198 | | |
| Posterior Probability | | 0.325 | 0.675 | | |

1. Usual workplace
2. Moving between work areas
3. Going to or coming from worksite
4. Different workplace

**Fig. 8.** Example of the developed VBA Excel application that classifies a sample heart attack accident as either non-fatal (HAA) (a) or fatal (FHAA) (b). Source: Own elaboration.

## 6. Conclusions

Heart attack accidents at work, particularly fatal ones, are significant issues with considerable social and economic implications. According to the Delta@ database, men in Spain are more affected. A higher number of cases were observed in the private sector, especially in small companies lacking a specific risk management plan, and among individuals with less seniority. The sectors with the highest number of cases, both non-fatal and fatal, included health and social work, transport, manufacturing, and construction.

These findings highlight the need for enhanced supervision of working conditions and the development of effective risk management plans specifically tailored to small private companies, which constitute a significant portion of businesses in Spain. Additionally, it is crucial to provide workers with thorough training on this risk from the outset of their employment. This is particularly pertinent to the construction, healthcare, social work, transport, and manufacturing sectors. In other sectors, such as agriculture, it is essential to focus not only on the number of reported heart attacks but also on the number of fatal incidents to mitigate their impact.

A model has been developed to assist companies in estimating the likelihood of a heart attack accident at work being fatal and to classify it accordingly as fatal or non-fatal. This model considers factors such as the company's activity, sector, size, preventive organisation, as well as the age, sex, length of service, and type of contract of the employee. It allows a company to be assessed based on these variables to determine the probability of an employee experiencing a fatal or non-fatal heart attack at work. The model calculates this probability using historical data from the Delta@ database, which can be utilised in periodic health surveillance reviews by companies to monitor, prevent, and devise strategies for reducing heart attack incidents in the workplace.

In conclusion, the findings of this study can assist in the development of strategies for monitoring and preventing heart attacks at the workplace.

For future research, it would be valuable to investigate the personal habits or health status of the individuals involved in addition to the predictive factors considered in this study. Furthermore, employing Naive Bayes models with attribute-weighted methods could help relax the independence assumption, and applying under-sampling and over-sampling balancing class algorithms could improve models dealing with unbalanced datasets.

## CRediT authorship contribution statement

**Alberto Sánchez-Lite:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jose Luis Fuentes-Bargues:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Iván Iglesias:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. **Cristina González-Gaya:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This paper is based on the ongoing research of the Working Group on Risk Engineering in Manufacturing (REM) of the Manufacturing Engineering Society; the authors therefore wish to express their gratitude for the support from that institution and Universidad de Valladolid.

## References

[1] E. Cagno, G.J.L. Micheli, S. Perotti, Identification of OHS-related factors and interactions among those and OHS performances in SMEs, Saf. Sci. 49 (2011) 216–225, https://doi.org/10.1016/j.ssci.2010.08.002.

[2] B. Wang, C. Wu, L. Huang, L. Zhang, L. Kang, K. Gao, Prevention and control of major accidents (MAs) and particularly serious accidents (PSAs) in the industrial domain in China: current status, recent efforts and future prospects, Process Saf. Environ. Protect. 117 (2018) 254–266, https://doi.org/10.1016/j.psep.2018.04.025.

[3] S. Perotti, M.C. Russo, Work-related fatal injuries in Brescia County (Northern Italy), 1982 to 2015: a forensic analysis, J. Forensic & Leg. Med. 58 (2018) 122–125, https://doi.org/10.1016/j.jflm.2018.06.002.

[4] Y. Peng, S. Zhang, P. Wu, Factors influencing workplace accident costs of building projects, Saf. Sci. 72 (2015) 97–104, https://doi.org/10.1016/j.ssci.2014.08.008.

[5] F.J. Forteza, J.M. Carretero-Gomez, A. Sese, Occupational risks, accidents on sites and economic performance of construction firms, Saf. Sci. 94 (2017) 61–76, https://doi.org/10.1016/j.ssci.2017.01.003.

[6] C. Melchior, R. Ruviaro Zanini, Mortality per work accident: a literature mapping, Saf. Sci. 114 (2019) 72–78, https://doi.org/10.1016/j.ssci.2019.01.001.

[7] International Labour Office (ILO). Safety and health at work [Online]. Available from: https://ilostat.ilo.org/topics/safety-and-health-at-work/ Accessed June 5, 2023.

[8] V. Saloniemi, J. Maiti, P. Ray, Occupational injury and accident research: a comprehensive review, Saf. Sci. 50 (2012) 1355–1367, https://doi.org/10.1016/j.ssci.2011.12.015.

[9] A. Darda 'u Rafindadi, N. Shafiq, I. Othman, A. Ibrahim, M.M. Aliyu, M. Mikić, H. Alarifi, Data mining of the essential causes of different types of fatal construction accidents, Heliyon 9 (2023) e13389, https://doi.org/10.1016/j.heliyon.2023.e13389.

[10] I.D. Anyfantis, S. Leka, G. Reniers, G. Boustras, Employers' perceived importance and the use (or non-use) of workplace risk assessment in micro-sized and small enterprises in Europe with focus on Cyprus, Saf. Sci. 139 (2021) 105256, https://doi.org/10.1016/j.ssci.2021.105256.

[11] C. Johnson, C.M. Holloway, A survey of logic formalisms to support mishap analysis, Reliab. Eng. Syst. Saf. 80 (3) (2003) 271–291, https://doi.org/10.1016/S0951-8320(03)00053-X.

[12] F. Salguero-Caparros, M. Suarez-Cebador, J.C. Rubio-Romero, Analysis of investigation reports on occupational accidents, Saf. Sci. 72 (2015) 329–336, https://doi.org/10.1016/j.ssci.2014.10.005.

[13] B. Shao, Z. Hu, Q. Liu, S. Chen, W. He, Fatal accident patterns of building construction activities in China, Saf. Sci. 111 (2019) 253–263, https://doi.org/10.1016/j.ssci.2018.07.019.

[14] B. Gulhan, M.N. Ilhan, E. Fusun Civil, Occupational accidents and affecting factors of metal industry in a factory in Ankara, Turkish J. Public Health 10 (2) (2012) 76–85, https://doi.org/10.20518/tjph.173067.

[15] S.B. Batti Gonçalves, T. Mamoru Sakae, F. Liberali Magajewski, Prevalence and factors associated with work accidents in a metal-mechanic company, Rev. Bras. Med. Trab. 16 (1) (2018) 26–35, https://doi.org/10.5327/Z1679443520180086.

[16] C. Jacinto, M. Canoa, C. Guedes Soares, Workplace and organisational factors in accident analysis within the Food Industry, Saf. Sci. 47 (2009) 626–635, https://doi.org/10.1016/j.ssci.2008.08.002.

[17] J.A. Carrillo-Castrillo, J.C. Rubio-Romero, L. Onieva, A. López-Arquillos, The causes of severe accidents in the Andalusian manufacturing sector: the role of human factors in official accident investigations, Hum. Factors Ergon. Manuf. & Serv. Ind. 26 (1) (2016) 68–83, https://doi.org/10.1002/hfm.20614.

[18] D.P. Shin, Y.J. Park, J. Seo, D.E. Lee, Association rules mined from construction accident data, KSCE J. Civ. Eng. 22 (2018) 1027–1039, https://doi.org/10.1007/s12205-017-0537-6.

[19] W. Cieslewicz, K. Araszkiewicz, P. Sikora, Accident rate as a measure of safety assessment in Polish civil engineering, Saf. Now. 5 (2019) 77, https://doi.org/10.3390/safety5040077.

[20] F.H. Hedlund, Recorded fatal and permanently disabling injuries in South African manufacturing industry - overview, analysis and reflection, Saf. Sci. 55 (2013) 149–159, https://doi.org/10.1016/j.ssci.2013.01.005.

[21] J.L. Fuentes-Bargues, A. Sánchez-Lite, C. González-Gaya, V.F. Rosales-Prieto, G. Reniers, A study of situational circumstances related to Spain's occupational accident rates in the metal sector from 2009 to 2019, Saf. Sci. 150 (2022) 105700, https://doi.org/10.1016/j.ssci.2022.105700.

[22] A. Goswani, M.P.R. Mohan, An exploratory analysis of occupational accidents and risks from nuclear reactors in India, Saf. Sci. 78 (2015) 155–162, https://doi.org/10.1016/j.ssci.2015.03.023.

[23] M.A. Camino-López, D.O. Ritzel, I. Fontaneda, O.J. González-Alcantara, Construction industry accidents in Spain, J. Saf. Res. 39 (2008) 497–507, https://doi.org/10.1016/j.jsr.2008.07.006.

[24] A. López-Arquillos, J.C. Rubio-Romero, A. Gibb, Analysis of construction accidents in Spain, 2003-2008, J. Saf. Res. 43 (2012) 381–388, https://doi.org/10.1016/j.jsr.2012.07.005.

[25] C.W. Cheng, S.S. Leu, C.C. Lin, C. Fan, Characteristic analysis of occupational accidents at small construction enterprises, Saf. Sci. 48 (2010) 698–707, https://doi.org/10.1016/j.ssci.2010.02.001.

[26] F. Antoniou, N.F. Agrafioti, Meta-analysis of studies on accident contributing factors in the Greek construction industry, Sustainability 15 (3) (2023), https://doi.org/10.3390/su15032357.

[27] K. Kang, H. Ryu, Predicting types of occupational accidents at construction sites in Korea using random forest model, Saf. Sci. 120 (2019) 226–236, https://doi.org/10.1016/j.ssci.2019.06.034.

[28] F. Khodabandeh, E. Kabir-Mokamelkhah, M. Kahani, Factors associated with the severity of fatal accidents in construction workers, Med. J. Islam. Repub. Iran 30 (469) (2016) 1–7.

[29] B. Barlas, F.B. Izci, Individual and workplace factors related to fatal occupational accidents among shipyard workers in Turkey, Saf. Sci. 101 (2018) 173–179, https://doi.org/10.1016/j.ssci.2017.09.012.

[30] L. Strand, J. Martinsen, E. Fadum, E. Borud, External-cause mortality among 21 609 Norwegian male military peacekeepers deployed to Lebanon between 1978 and 1998, Occup. Environ. Med. 74 (8) (2017) 573–577, https://doi.org/10.1136/oemed-2016-104170.

[31] C.D. Hansen, Comparing fatal occupational accidents in Denmark and Sweden 1993-2012, Occup. Med. 69 (2019) 283–286, https://doi.org/10.1093/occmed/kqz064.

[32] H. Asady, M. Yaseri, M. Hosseini, M. Zarif-Yeganeh, M. Yousefifard, M. Haghshenas, P. Hajizadeh-Moghadam, Risk factors of fatal occupational accidents in Iran, Annals Occup. Environ. Med. 30 (29) (2018), https://doi.org/10.1186/s40557-018-0241-0.

[33] S.Y. Kang, S. Min, W.S. Kim, J.H. Won, Y.J. Kang, S. Kim, Types and characteristics of fatal accidents caused by multiple processes in a workplace: based on actual cases in South Korea, Int. J. Environ. Res. Publ. Health 19 (2022) 2047, https://doi.org/10.3390/ijerph19042047.

[34] A.R. Gómez-García, K.P. Córdova Falconi, P. Merino-Salazar, J. García-Arroyo, Fatal work accidents in Ecuador from 2014 to 2020: how the age of the deceased worker relates to the accidents' temporal and geographical characteristics, Arch. Environ. Occup. Health (2023) 1–7, https://doi.org/10.1080/19338244.2023.2196051.

[35] A.J.R. Santos, E.L. Rebelo, J.C. Mendes, Towards better prevention of fatal occupational accidents in Portugal, Int. Lab. Rev. 157 (3) (2018) 409–433, https://doi.org/10.1111/ilr.12114.

[36] J.L. Fuentes-Bargues, A. Sánchez-Lite, C. González-Gaya, M.A. Artacho-Ramírez, Descriptive analysis and a proposal for a predictive model of fatal occupational accidents in Spain, Heliyon (2023) e22219, https://doi.org/10.1016/j.heliyon.2023.e22219.

[37] G. Reniers, Y. Gidron, Do cultural dimensions predict prevalence of fatal work injuries in Europe? Saf. Sci. 58 (2013) 76–80, https://doi.org/10.1016/j.ssci.2013.03.015.

[38] S. Salminen, Have young workers more injuries than older ones? An international literature review, J. Saf. Res. 35 (5) (2004) 513–521, https://doi.org/10.1016/j.jsr.2004.08.005.

[39] D.-S. Chang, Y.-C. Tsai, Investigating the long-term change of injury pattern on severity, accident types and sources of injury in Taiwan's manufacturing sector between 1996 and 2012, Saf. Sci. 68 (2014) 231–242, https://doi.org/10.1016/j.ssci.2014.04.005.

[40] J.J. Zhang, K.L. Xu, G. Reniers, G. You, Statistical analysis the characteristics of extraordinarily severe coal mine accidents (ESCMAs) in China from 1950 to 2018, Process Saf. Environ. Protect. 133 (2020) 332–340, https://doi.org/10.1016/j.psep.2019.10.014.

[41] L. Lafflamme, E. Menckel, A. Stromberg, Age-related overexertion injuries among Swedish nursing auxiliaries over a 10-year period, Work 8 (2) (1997) 139–148, https://doi.org/10.3233/WOR-1997-8204.

[42] A. González-Fuentes, N.M. Busto Serrano, F. Sánchez Lasheras, G. Fidalgo-Valverde, A. Suárez-Sánchez, Work-related overexertion injuries in cleaning occupations: an exploration of the factors to predict the days of absence by means of machine learning methodologies, Appl. Ergon. 105 (2022), https://doi.org/10.1016/j.apergo.2022.103847.

[43] E. Weisshaar, M. Radulescu, S. Soder, C.J. Apfelbacher, M. Bock, J.U. Grundmann, U. Albrecht, T.L. Diepgen, Secondary individual prevention of occupational skin diseases in health care workers, cleaners and kitchen employees: aims, experiences and descriptive results, Int. Arch. Occup. Environ. Health 80 (60) (2007) 477–484, https://doi.org/10.1007/s00420-006-0154-z.

[44] J.M. Violanti, D. Fekedulegn, M. Shi, M.E. Andrew, A 22-years analysis of law enforcement deaths associated with duty-related illnesses (1997-2018), Policing: Int. J. 43 (2) (2020) 330–344, https://doi.org/10.1108/PIJPSM-07-2019-0109.

[45] Labour Office (ILO), Recording and notification accidents and diseases, An ILO code of practice. Geneva (1996).

[46] R. Poquet Catalá, La configuración jurídica del infarto como accidente de trabajo, Revista de Derecho de la Seguridad Social, Laborum 30 (2022) 105–127.

[47] R.A. Bruce, L.D. Fisher, Strategies for risk-evaluation of sudden cardiac incapacitation in men in occupations affecting public safety, J. Occup. Environ. Med. 31 (2) (1989) 124–133.

[48] M. Wojcik-Stasiak, B. Jaremin, S.E. Roberts, T. Chodnik, Sudden cardiac event on a sea-going ship and recognition of a work-related accident, Int. Marit. Health 62 (2) (2011) 110–115.

[49] V. Vaccarino, J.D. Bremner, Stress and cardiovascular disease: an update, Nat. Rev. Cardiol. (2024), https://doi.org/10.1038/s41569-024-01024-y.

[50] A. Bortkiewicz, E. Gadzicka, J. Siedlecka, A. Szyjkowska, P. Viebig, J.K. Wranicz, T. Makowiec-Dabrowska, Work-related risk factors of myocardial infarction, Int. J. Occup. Med. Environ. Health 23 (3) (2010) 255.

[51] M.S. Gotsman, J.D. Stephen Adler, Under what circumstances can an acute myocardial infarction be regarded as a work-related accident? Multi-causal diseases as work accidents, Int. J. Soc. Secur. Work. Compensat. 1 (2009) 21–32.

[52] M. Kivimäki, I. Kawachi, Work stress as a risk factor for cardiovascular disease, Curr. Cardiol. Rep. 17 (74) (2015), https://doi.org/10.1007/s11886-015-0630-8.

[53] J.P. Leigh, Occupational hazards and heart attacks, Soc. Sci. 23 (11) (1986) 1181–1185.

[54] J.P. Leigh, T.R. Miller, Occupational illnesses within two national data sets, Int. J. Occup. Environ. Health 4 (2) (1998) 99–113.

[55] M.R. Gray, et al., Occupational Stress: the Hidden Hazard (Project Module), Arizona Center for Occupational Safety and Health, Tucson. Ariz, 1980.

[56] M. Virtanen, K. Heikkila, M. Jokela, et al., Long working hours and coronary heart disease: a systematic review and meta-analysis, Am. J. Epidemiol. 176 (2012) 586–596.

[57] M. Kivimäki, G.D. Batty, J.E. Ferrie, I. Kawachi, Cumulative meta-analysis of job strain and CHD, Epidemiology 25 (3) (2014) 464–465.

[58] M. Kivimäki, M. Virtanen, I. Kawachi, et al., Long working hours, socioeconomic status, and the risk of incident type 2 diabetes: a meta-analysis of published and unpublished data from 222,120 individuals, Lancet Diabetes Endocrinol. 3 (2015) 27–34.

[59] J. Siegrist, Adverse health effects of high-effort/low-reward conditions, J. Occup. Health Psychol. 1 (1996) 27–41.

[60] N. Slopen, R.J. Glynn, J.E. Buring, T.T. Lewis, D.R. Williams, M.A. Albert, Job strain, job insecurity, and incident cardiovascular disease in the Women's Health study: results from a 10-year prospective study, PLoS One 7 (7) (2012) e40512, https://doi.org/10.1371/journal.pone.0040512.

[61] S. Lee, G.A. Colditz, L.F. Berkman, I. Kawachi, Prospective study of job insecurity and coronary heart disease in US women, Ann. Epidemiol. 14 (2004) 24–30.

[62] M.G. Marmot, S. Stanfeld, C. Patl, F. North, J. Head, I. White, et al., Health inequalities among British civil servants: the White Hall II study, Lancet 337 (1991), https://doi.org/10.1016/0140-6736(91)93068-K.

[63] S. Kagamimori, I. Matsubara, S. Sokejoma, M. Sekine, T. Matsukura, H. Nakagawa, Y. Naruse, The comparative study on occupational mortality, 1980 between Japan and Great Britain, Ind. Health 36 (1998) 252–257.

[64] K.A. Moser, P.O. O Goldblatt, Occupational mortality of women aged 45-59 years at death in England and Wales, Journal of Epidemology and Community Health 45 (1991) 117–124.

[65] A. Mannoci, B. Lojodice, A. Boccia, G. LaTorre, Systematic review of the literature regarding the risk of mortality, cancer and cardiovascular diseases in professio-nal drivers, Sanita igPubbl 69 (2) (2013) 155–170.

[66] B. Krzowski, A.E. Platek, F.M. Szymanski, A. Rys, K. Semczuk-Kaczmarek, K. Adamkiewicz, et al., Epidemiology of dysli-pidaemia in professional drivers: results of RACER-ABPM (risk of adverse cardiovascular events among professional rivers in Poland - ambulatory blood pressure monitoring) study, Kardiol. Pol. 76 (2) (2018) 396–400, https://doi.org/10.5603/KP.a2017.0229. Epub 2017 Dec 1.

[67] E. Mocevic, P. Kristiansen, J.P. Bonde, Risk of ischemic heart disease following occupational exposure to wel-ding fumes: a systematic review with meta-analysis, Int. Arch. Occup. Environ. Health 88 (3) (2015) 259–272, https://doi.org/10.1007/s00420-014-0965-2. Epub 2014 Jul 22.

[68] P. Morfeld, K.A. Mundt, L.D. DSell, T. Sorhan, R.J. McCunney, Meta-analysis of cardiac mortality in three cohorts of carbon black production workers, Int J Environ Res Public Health pii (2016) E302, https://doi.org/10.3390/ijerph13030302.

[69] R. Chowdhury, A. Ramnd, L.M. ÓKeeffe, S. Shahzad, S.K. Kunutsor, T. Muka, et al., Environmental toxic metal contaminants and risk of cardiovascular disease: systema-tic review and meta-analysis, BMJ 362 (2018) k3310, https://doi.org/10.1136/bmj.k3310.

[70] K. Szereda, J. Szymanska, Accidents at work in the health care – legal aspects in Poland, Pol. Merkur. Lek.: organ Polskiego Towarzystwa Lekarskiego 40 (235) (2016) 70–74.

[71] J.M. Violanti, Dying for the Job: Police Work Exposure and Health, Charles C. Thomas, Springfield, IL, 2014, pp. 41–51.

[72] J.E. Vena, L.E. Charles, J.K. Gu, C.M. Burchfiel, M.E. Andrew, D. Fekedulegn, J.M. Violanti, Mortality of a police cohort: 1950-2005, Journal of Law Enforcement Leadership and Ethics 1 (1) (2014) 7–30.

[73] S. Garbarino, N. Magnavita, Work stress and metabolic syndrome in police officers. A prospective study, PLoS One 10 (12) (2015) e0144318, https://doi.org/10.1371/journal.pone.0144318 article.

[74] E. Feuer, K. Rosenman, Mortality in police and firefighters in New Jersey, Am. J. Ind. Med. 9 (3) (1986) 517–529.

[75] D. Nielsen, Occupational accidents at sea, Marine Technology III 1 (2000) 553, 462.
[76] M. Oldenburg, J. Herzog, V. Harth, Seafarer deaths at sea: a German mortality study, Occupational medicine-Oxford 66 (2) (2016) 135–137, https://doi.org/10.1093/occmed/kqv153.
[77] D. Nielsen, Deaths at sea – a study of fatalities on board Hong Kong-registered merchant ships (1986-1995), Saf. Sci. 32 (2–3) (1999) 121–141.
[78] H.L. Hansen, J. Jensen, Female seafarers adopt the high risk lifestyle of male seafarers, Occup. Environ. Med. 55 (1) (1998) 49–51, https://doi.org/10.1136/oem.55.1.49.
[79] X. Baraza, N. Cugueró-Escofet, Severity of occupational agricultural accidents in Spain, 2013-2018, Saf. Sci. 143 (2021) 105422.
[80] IBM Corp. Released, IBM SPSS Statistics for Windows, Version 29.0, IBM Corp, Armonk, NY, 2022.
[81] Ministry of Labour and Social Affairs of Spain (MLSAS), Orden TAS/2926/2002, de 19 de noviembre, por la que se establecen nuevos modelos para la notificación de los accidentes de trabajo y se posibilita su transmisión por procedimiento electrónico 279 (2002) 40988–41013.
[82] X. Baraza, N. Cugueró-Escofet, R. Rodríguez-Elizalde, Statistical analysis of the severity of occupational accidents in the mining sector, J. Saf. Res. 86 (2023) 364–375, https://doi.org/10.1016/j.jsr.2023.07.015.
[83] M. del C. Rey-Merchán, J.M. Gómez-de-Gabriel, A. López-Arquillos, S.D. Choi, Analysis of falls from height variables in occupational accidents, Int. J. Environ. Res. Publ. Health 18 (24) (2021) 13417, https://doi.org/10.3390/ijerph182413417.
[84] CNAE 09. National Classification of Economic Activities in Spain. Available on: https://ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177032&menu=ultiDatos&idp=1254735976614. Accessed May 24, 2023.
[85] D. Berrar, Bayes' theorem and naive Bayes classifier, in: Encyclopedia of Bioinformatics and Computational Biology, Elsevier, 2019, pp. 403–412, https://doi.org/10.1016/B978-0-12-809633-8.20473-1.
[86] G.I. Webb, Naïve Bayes, in: Encyclopedia of Machine Learning and Data Mining, Springer US, 2016, pp. 1–2, https://doi.org/10.1007/978-1-4899-7502-7_581-1.
[87] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, Sung Hyon Myaeng, Some effective techniques for naive Bayes text classification, IEEE Trans. Knowl. Data Eng. 18 (11) (2006) 1457–1466, https://doi.org/10.1109/TKDE.2006.180.
[88] N.K. Singh, D.S. Tomar, A.K. Sangaiah, Sentiment analysis: a review and comparative analysis over social media, J. Ambient Intell. Hum. Comput. 11 (1) (2020), https://doi.org/10.1007/s12652-018-0862-8.
[89] J. Kazmierska, J. Malicki, Application of the Naïve Bayesian classifier to optimize treatment decisions, Radiother. Oncol. 86 (2) (2008) 211–216, https://doi.org/10.1016/j.radonc.2007.10.019.
[90] J. Zheng, Y. Tang, One Generalization of the Naive Bayes to Fuzzy Sets and the Design of the Fuzzy Naive Bayes Classifier (2005) 281–290, https://doi.org/10.1007/11499305_29.
[91] D. Shah, S. Patel, S.K. Bharti, Heart disease prediction using machine learning techniques, SN Computer Science 1 (6) (2020) 345, https://doi.org/10.1007/s42979-020-00365-y.
[92] I. Kumar, K. Dogra, C. Utreja, P. Yadav, A comparative study of supervised machine learning algorithms for stock market trend prediction, in: 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), IEEE, 2018, April, pp. 1003–1007.
[93] N.B. Amor, S. Benferhat, Z. Elouedi, Naive bayes vs decision trees in intrusion detection systems, in: Proceedings of the 2004 ACM Symposium on Applied Computing, 2004, March, pp. 420–424.
[94] A. Ashari, I. Paryudi, A.M. Tjoa, Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool, Int. J. Adv. Comput. Sci. Appl. 4 (11) (2013).
[95] M. Veziroğlu, E. Eziroğlu, İ.Ö. Bucak, Performance comparison between Naive Bayes and machine learning algorithms for news classification, in: Bayesian Inference-Recent Trends, IntechOpen, 2024.
[96] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1) (1950) 32–35, https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
[97] D.G. Altman, J.M. Bland, Diagnostic tests. 1: sensitivity and specificity, Br. Med. J. 308 (6943) (1994) 1552, https://doi.org/10.1136/bmj.308.6943.1552.
[98] P. Baldi, S. Brunak, Y. Chauvin, C.A.F. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview, Bioinformatics 16 (5) (2000) 412–424, https://doi.org/10.1093/bioinformatics/16.5.412.
[99] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC Med. Imag. 15 (1) (2015) 29, https://doi.org/10.1186/s12880-015-0068-x.
[100] M.J. Warrens, H. van der Hoef, Understanding the adjusted Rand index and other partition comparison indices based on counting object pairs, J. Classif. 39 (3) (2022) 487–509, https://doi.org/10.1007/s00357-022-09413-z.
[101] K. Takahashi, K. Yamamoto, A. Kuchiba, T. Koyama, Confidence interval for micro-averaged F1 and macro-averaged F1 scores, Appl. Intell. 52 (5) (2022) 4961–4972, https://doi.org/10.1007/s10489-021-02635-5.
[102] C. Ferri, J. Hernández-Orallo, R. Modroiu, An experimental comparison of performance measures for classification, Pattern Recogn. Lett. 30 (1) (2009) 27–38, https://doi.org/10.1016/j.patrec.2008.08.010.
[103] J.-G. Gaudreault, P. Branco, Empirical analysis of performance assessment for imbalanced classification, Mach. Learn. (2024), https://doi.org/10.1007/s10994-023-06497-5.
[104] R. Soleymani, E. Granger, G. Fumera, F-measure curves: a tool to visualize classifier performance under imbalance, Pattern Recogn. 100 (2020) 107146, https://doi.org/10.1016/j.patcog.2019.107146.
[105] Q. Zhu, On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset, Pattern Recogn. Lett. 136 (2020) 71–80, https://doi.org/10.1016/j.patrec.2020.03.030.
[106] D. Chicco, G. Jurman, The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation, BMC Genom. 21 (1) (2020) 6, https://doi.org/10.1186/s12864-019-6413-7.
[107] D. Chicco, G. Jurman, The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification, BioData Min. 16 (1) (2023) 4, https://doi.org/10.1186/s13040-023-00322-4.
[108] N.J. Perkins, E.F. Schisterman, The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve, Am. J. Epidemiol. 163 (7) (2006) 670–675, https://doi.org/10.1093/aje/kwj063.
[109] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PLoS One 10 (3) (2015) e0118432, https://doi.org/10.1371/journal.pone.0118432.
[110] A.R. Redondo, J. Navarro, R.R. Fernández, I.M. de Diego, J.M. Moguerza, J.J. Fernández-Muñoz, Unified Performance Measure for Binary Classification Problems, 2020, pp. 104–112, https://doi.org/10.1007/978-3-030-62365-4_10.
[111] J.M. Santos, M.J. Embrechts, On the use of the adjusted Rand index as a metric for evaluating supervised classification, in: Cesare Alippi, Marios M. Polycarpou, Christos G. Panayiotou, Georgios Ellina (Eds.), Artificial Neural Networks - ICANN 2009, 19th International Conference, Limassol, Cyprus, September 14-17, 2009, Proceedings, Part II. Lecture Notes in Computer Science, vol. 5769, Springer, 2009, pp. 175–185.
[112] L.A. Jeni, J.F. Cohn, F. de La Torre, Facing imbalanced data–recommendations for the use of performance metrics, in: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, 2013, pp. 245–251, https://doi.org/10.1109/ACII.2013.47.
[113] I.M. de Diego, A.R. Redondo, R.R. Fernández, J. Navarro, J.M. Moguerza, General performance score for classification problems, Appl. Intell. 52 (10) (2022) 12049–12063, https://doi.org/10.1007/s10489-021-03041-7.
[114] T.F. Monaghan, S.N. Rahman, C.W. Agudelo, A.J. Wein, J.M. Lazar, K. Everaert, R.R. Dmochowski, Foundational statistical principles in medical research: sensitivity, specificity, positive predictive value, and negative predictive value, Medicina 57 (5) (2021) 503, https://doi.org/10.3390/medicina57050503.
[115] S. Amiri, S. Behnezhad, Job strain and mortality ratio: a systematic review and meta-analysis of cohort studies, Publ. Health 181 (2020) 24–33, https://doi.org/10.1016/j.puhe.2019.10.030.

[116] S. Kumar, F. Hansiya Abdul Rauf, H. Rathnasekara, Working to help or helping to work? Work-overload and allocentrism as predictors of organizational citizenship behaviours, Int. J. Hum. Resour. Manag. 32 (13) (2021) 2807–2828, https://doi.org/10.1080/09585192.2019.1602549.

[117] S.A. Widodari, S.A. Sudari, D. Rachmawati, Influence of workload and organizational culture on employee performance of operational division at Toyota (Auto 2000) Grand Depok city, Depok, Lead Journal of Economy and Administration 2 (3) (2024) 139–150, https://doi.org/10.56403/lejea.v2i3.177.

[118] X. Guo, H.W. Davies, M.E. Karim, The intensity and duration of occupational noise exposure and cardiovascular disease in the United States: a nationally representative study, 2015 to 2020, Annals of Work Exposures and Health 68 (1) (2024) 36–47, https://doi.org/10.1093/annweh/wxad065.

[119] D. Hrehova, P. Ziaran, Impact of the working conditions on health and selected job-related diseases. An international comparison, SHS Web of Conferences 92 (2021) 07025, https://doi.org/10.1051/shsconf/20219207025.

[120] I. Lever, D. Dyball, N. Greenberg, S.A.M. Stevelink, Health consequences of bullying in the healthcare workplace: a systematic review, J. Adv. Nurs. 75 (12) (2019) 3195–3209, https://doi.org/10.1111/jan.13986.

[121] A. Arthur, A. Mensah-Asamoah, E.K.M. Brown, A.K. Ocansey, P.D. Deku, M.O. Moses, Work ability and anthropometric indices correlate with cardiovascular risk factors in public sector employees: cross-sectional study, Health Science Reports 6 (11) (2023), https://doi.org/10.1002/hsr2.1728.

[122] S. Noureddine, E.S. Froelicher, A.M. Sibai, H. Dakik, Response to a cardiac event in relation to cardiac knowledge and risk perception in a Lebanese sample: a cross sectional survey, Int. J. Nurs. Stud. 47 (3) (2010) 332–341, https://doi.org/10.1016/j.ijnurstu.2009.07.002.

[123] S. Zakaria, O. Nurhaya, A. Asmawi, Work responsibilities stress among academicians in private universities in Malaysia, Journal of Education and Vocational Research 6 (2) (2015) 42–47, https://doi.org/10.22610/jevr.v6i2.187.

[124] T. Lang, P. Ducimetiere, D. Arveiler, P. Amouyel, J.P. Cambou, J.B. Ruidavets, M. Montaye, V. Meyer, A. Bingham, Incidence, case fatality, risk factors of acute coronary heart disease and occupational categories in men aged 30-59 in France, Int. J. Epidemiol. 26 (1) (1997) 47–57, https://doi.org/10.1093/ije/26.1.47.

[125] C. Álvarez-Fernández, M. Vaquero-Abellán, M. Romero-Saldaña, C. Álvarez-López, Trabajadores Especialmente Sensibles al Riesgo Cardiovascular, vol. 93, 2019, pp. e1–e11.

[126] A. Groenewegen, F.H. Rutten, A. Mosterd, A.W. Hoes, Epidemiology of heart Price, A. E. (2004). Heart disease and work, Heart 90 (9) (2020) 1077–1084, https://doi.org/10.1136/hrt.2003.029298.

[127] M. Walli-Attaei, P. Joseph, A. Rosengren, C.K. Chow, S. Rangarajan, S.A. Lear, K.F. AlHabib, K. Davletov, A. Dans, F. Lanas, K. Yeates, P. Poirier, K.K. Teo, A. Bahonar, F. Camilo, J. Chifamba, R. Diaz, J.A. Didkowska, V. Irazola, S. Yusuf, Variations between women and men in risk factors, treatments, cardiovascular disease incidence, and death in 27 high-income, middle-income, and low-income countries (PURE): a prospective cohort study, Lancet 396 (10244) (2020) 97–109, https://doi.org/10.1016/S0140-6736(20)30543-2.

[128] F.L.J. Visseren, F. Mach, Y.M. Smulders, D. Carballo, K.C. Koskinas, M. Bäck, A. Benetos, A. Biffi, J.-M. Boavida, D. Capodanno, B. Cosyns, C. Crawford, C. H. Davos, I. Desormais, E. di Angelantonio, O.H. Franco, S. Halvorsen, F.D.R. Hobbs, M. Hollander, B. Williams, 2021 ESC Guidelines on cardiovascular disease. prevention in clinical practice, Eur. Heart J. 42 (34) (2021) 3227–3337, https://doi.org/10.1093/eurheartj/ehab484.

[129] I. Fontaneda, M.A. Camino López, O.J. González Alcántara, B.A. Greiner, The "weekday effect": a decrease in occupational accidents from Monday to Friday—an extension of the "Monday effect.", BioMed Res. Int. (2024) 1–12, https://doi.org/10.1155/2024/4792081, 2024.

[130] E. Ebrahimihoor, M. Karpman, J. Grover, N. Muganlinskaya, Day of the week variation in emergency department arrivals, chest pain, and acute myocardial infarction throughout 2016-2019, J. Community Hosp. Intern. Med. Perspect. 13 (5) (2023), https://doi.org/10.55729/2000-9666.1237.

[131] K. Nowacki, Accident risk in the production sector of EU countries— cohort studies, Int. J. Environ. Res. Publ. Health 18 (2021), https://doi.org/10.3390/ijerph18073618.

[132] National Statistical Institute (NSE), Actividad, ocupación y paro/Mercado laboral/Encuesta de población active/Series desde el primer trimestre de 2002/Resultados por comunidades autónomas, Available on: https://www.ine.es/jaxiT3/Datos.htm?t=4204, 2023.

[133] Y. Desai, H. Khraishah, B. Alahmad, Heat and the heart, Yale J. Biol. Med. 96 (2) (2023) 197–203, https://doi.org/10.59249/HGAL4894.

[134] M.G. Abrignani, A. Lombardo, A. Braschi, N. Renda, V. Abrignani, Climatic influences on cardiovascular diseases, World J. Cardiol. 14 (3) (2022) 152–169, https://doi.org/10.4330/wjc.v14.i3.152.

[135] P. Dilaveris, A. Synetos, G. Giannopoulos, E. Gialafos, A. Pantazis, C. Stefanadis, CLimate Impacts on Myocardial infarction deaths in the Athens TErritory: the CLIMATE study, Heart 92 (12) (2006) 1747, https://doi.org/10.1136/hrt.2006.091884.

[136] K.S. Reddy, D. Prabhakaran, V. Chaturvedi, P. Jeemon, K.R. Thankappan, L. Ramakrishnan, B.V.M. Mohan, C.S. Pandav, F.U. Ahmed, P.P. Joshi, R. Meera, R. B. Amin, R.C. Ahuja, M.S. Das, Methods for establishing a surveillance system for cardiovascular diseases in Indian industrial populations, Bull. World Health Organ. 84 (6) (2006) 461+. https://link.gale.com/apps/doc/A149514569/AONE?u=anon~2f3e5928&sid=googleScholar&xid=629cad85.

[137] P.L. Schnall, M. Dobson, P. Landsbergis, Globalization, work, and cardiovascular disease, Int. J. Health Serv. 46 (4) (2016) 656–692, https://doi.org/10.1177/0020731416664687.

[138] P. Bhatnagar, K. Wickramasinghe, E. Wilkins, N. Townsend, Trends in the epidemiology of cardiovascular disease in the UK, Heart 102 (24) (2016) 1945–1952, https://doi.org/10.1136/heartjnl-2016-309573.

[139] A.E. Price, Heart disease and work, Heart 90 (9) (2004) 1077–1084, https://doi.org/10.1136/hrt.2003.029298.

[140] R. Luengo-Fernandez, M. Walli-Attaei, A. Gray, A. Torbica, A.P. Maggioni, R. Huculeci, F. Bairami, V. Aboyans, A.D. Timmis, P. Vardas, J. Leal, Economic burden of cardiovascular diseases in the European Union: a population-based cost study, Eur. Heart J. 44 (45) (2023) 4752–4767, https://doi.org/10.1093/eurheartj/ehad583.

[141] T. Hasanin, T.M. Khoshgoftaar, J.L. Leevy, N. Seliya, Examining characteristics of predictive models with imbalanced big data, Journal of Big Data 6 (1) (2019) 69, https://doi.org/10.1186/s40537-019-0231-2.

[142] A. Rácz, D. Bajusz, K. Héberger, Effect of dataset size and train/test split ratios in QSAR/QSPR multiclass classification, Molecules 26 (4) (2021) 1111, https://doi.org/10.3390/molecules26041111.

[143] M. Buda, A. Maki, M.A. Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Network. 106 (2018) 249–259, https://doi.org/10.1016/j.neunet.2018.07.011.

[144] B. Vrigazova, The proportion for splitting data into training and test set for the Bootstrap in classification problems, Business Systems Research Journal 12 (1) (2021) 228–242, https://doi.org/10.2478/bsrj-2021-0015.

[145] S. el Hajjami, J. Malki, M. Berrada, B. Fourka, Machine Learning for anomaly detection. Performance study considering anomaly distribution in an imbalanced dataset, in: 2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech), 2020, pp. 1–8, https://doi.org/10.1109/CloudTech49835.2020.9365887.

[146] M. Revathi, D. Ramyachitra, Empirical Analysis of Preprocessing Techniques for Imbalanced Dataset Using Logistic Regression (2023) 331–339, https://doi.org/10.1007/978-981-19-2225-1_30.

[147] Q. Gu, L. Zhu, Z. Cai, Erratum to: Evaluation Measures of the Classification Performance of Imbalanced Data Sets (2009) E1, https://doi.org/10.1007/978-3-642-04962-0_55. E1.

[148] Y. Feng, M. Zhou, X. Tong, Imbalanced classification: a paradigm-based review, Stat. Anal. Data Min.: The ASA Data Science Journal 14 (5) (2021) 383–406, https://doi.org/10.1002/sam.11538.

[149] Q. Zou, S. Xie, Z. Lin, M. Wu, Y. Ju, Finding the best classification threshold in imbalanced classification, Big Data Research 5 (2016) 2–8, https://doi.org/10.1016/j.bdr.2015.12.001.

[150] C. O'Reilly, T. Nielsen, Revisiting the ROC curve for diagnostic applications with an unbalanced class distribution, in: 2013 8th International Workshop on Systems, Signal Processing and Their Applications (WoSSPA), 2013, pp. 413–420, https://doi.org/10.1109/WoSSPA.2013.6602401.

[151] I. Wickramasinghe, H. Kalutarage, Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation, Soft Comput. 25 (3) (2021) 2277–2293, https://doi.org/10.1007/s00500-020-05297-6.

[152] G.W. Brier, Verification of forecasts expressed in terms of probability, Mon. Weather Rev. 78 (1) (1950) 1–3.

[153] A.H. Murphy, E.S. Epstein, A note on probability forecasts and" Hedging", J. Appl. Meteorol. (1967) 1002–1004, 1962-1982.

[154] J. Hilden, T.A. Gerds, A note on the evaluation of novel biomarkers: do not rely on integrated discrimination improvement and net reclassification index, Stat. Med. 33 (19) (2014) 3405–3414.

[155] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Mach. Learn. 29 (2/3) (1997) 103–130, https://doi.org/10.1023/A:1007413511361.

[156] P.J.B. Pajila, B.G. Sheena, A. Gayathri, J. Aswini, M. Nalini, S.S. R, A comprehensive survey on naive Bayes algorithm: advantages, limitations and applications, in: 2023 4th International Conference on Smart Electronics and Communication (ICOSEC), 2023, pp. 1228–1234, https://doi.org/10.1109/ICOSEC58147.2023.10276274.

[157] M. Li, K. Liu, Causality-based attribute weighting via information flow and genetic algorithm for naive Bayes classifier, IEEE Access 7 (2019) 150630–150641, https://doi.org/10.1109/ACCESS.2019.2947568.

[158] C. Wu, Q. Zhang, Y. Cheng, M. Gao, G. Wang, Novel three-way generative classifier with weighted scoring distribution, Inf. Sci. 579 (2021) 732–750, https://doi.org/10.1016/j.ins.2021.08.025.

[159] Y.-L. He, G.-L. Ou, P. Fournier-Viger, J.Z. Huang, P.N. Suganthan, A novel dependency-oriented mixed-attribute data classification method, Expert Syst. Appl. 199 (2022) 116782, https://doi.org/10.1016/j.eswa.2022.116782.

[160] [a] L. Jiang, D. Wang, Z. Cai, X. Yan, Survey of Improving Naive Bayes for Classification (2007) 134–145, https://doi.org/10.1007/978-3-540-73871-8_14; [b] L.A. Sevastyanov, E.Y. Shchetinin, On methods for improving the accuracy of multi-class classification on imbalanced data, ITTMM 20 (2020) 70–82.

[161] G.I. Webb, J.R. Boughton, Z. Wang, Facing imbalanced data–recommendations for the use of performance metrics, Mach. Learn. 58 (1) (2005) 5–24, https://doi.org/10.1007/s10994-005-4258-6.

[162] U. Norinder, S. Boyer, Binary classification of imbalanced datasets using conformal prediction, J. Mol. Graph. Model. 72 (2017) 256–265, https://doi.org/10.1016/j.jmgm.2017.01.008.

[163] S. Taheri, M. Mammadov, Learning the naive Bayes classifier with optimization models, Int. J. Appl. Math. Comput. Sci. 23 (4) (2013) 787–795, https://doi.org/10.2478/amcs-2013-0059.

[164] L. Jiang, C. Li, S. Wang, L. Zhang, Deep feature weighting for naive Bayes and its application to text classification, Eng. Appl. Artif. Intell. 52 (2016) 26–39, https://doi.org/10.1016/j.engappai.2016.02.002.