Understanding evacuation behavior for efective disaster preparedness: a hybrid machine learning approach

Article *in* Natural Hazards · June 2024

DOI: 10.1007/s11069-024-06759-y

5 authors, including:

Evangelos Karampotsis
University of the Aegean

Kitty Kioskli
University of Essex

Georgios Dounias
University of the Aegean

Amalia Polydoropoulou
University of the Aegean

Natural Hazards
https://doi.org/10.1007/s11069-024-06759-y

LETTER TO THE EDITOR

Understanding evacuation behavior for effective disaster preparedness: a hybrid machine learning approach

Evangelos Karampotsis[1] · Kitty Kioskli[2,3] · Athina Tsirimpa[2,4] · Georgios Dounias[1] · Amalia Polydoropoulou[2]

Received: 17 November 2023 / Accepted: 21 June 2024

Abstract
This paper delves into the pivotal role of machine learning in responding to natural disasters and understanding human behavior during crises. Natural disasters, from earthquakes to floods, have profound consequences for both the environment and society, impacting health, the economy, and mental well-being. Prevention and preparedness are key components of disaster management, yet the psychological challenges faced by affected individuals are equally significant. Psychosocial support and educational programs play a vital role in aiding individuals in their recovery. Machine learning, in this context, offers the ability to predict the evolution of natural disasters, providing early warnings that can save lives and reduce losses. It further extends to analyzing data related to human behavior during disasters, enhancing readiness for future calamities. This study specifically addresses the challenge of understanding human behavior during a snowstorm that struck Greece in 2023, employing artificial intelligence techniques to develop classification models categorizing individuals into three distinct groups based on socio-economic characteristics and is one of the few machine learning approaches that have been performed to date on data derived from corresponding questionnaire surveys. Artificial intelligence methodologies were harnessed to construct these classification models, with a focus on categorizing individuals into three specific classes: "Did not travel at all", "Traveled only as necessary", or "Did not limit travel". The dataset employed in this study was collected through a survey conducted within the framework of the AEGIS+research project, concentrating on assessing the mental health of individuals impacted by natural disasters. The goal was to generalize the optimal classification model and extract knowledge applicable in natural disaster scenarios. Three methodological frameworks for data analysis were proposed, incorporating combinations of Simple Logistic Regression and Inductive Decision Trees with the SMOTE data balancing method and a new data balancing method called LCC (Leveling of Cases per Class), within the context of validation procedures like "Use Train Set," "10-fold Cross Validation," and "Hold Out." This paper's contribution lies in the development of hybrid classification models, highlighting the significance of data balancing with LCC method throughout the modeling process. The results were deemed satisfactory, with the inductive decision tree method demonstrating superior performance (Classification accuracy near to 90%). This approach, offering strong classification rules, holds potential for knowledge application in natural disaster risk management. Knowledge Mining and Metadata

Extended author information available on the last page of the article

Published online: 27 June 2024

Analysis further revealed the socio-economic characteristics influencing the decision to move during a natural disaster, including age, education, work-status, and workstyle. Crucially, this work, in addition to providing knowledge through the data mining process that can be used to estimate evacuation probability, develop targeted emergency information messages, and improve evacuation planning, is also used as a catalyst for future research efforts. It encourages the collection of relevant data, the exploration of new challenges in data analysis related to natural disasters and mental health, and the development of new data balancing methods and hybrid data analysis methodological frameworks.

Keywords Evacuation · Trips-related decisions · Behavioral response · Compliance · Natural disasters · Artificial intelligence · Inductive machine learning · Data mining

1 Introduction

The frequent natural disasters (or natural hazard impacts) that happen worldwide have recently significantly impacted society, the economy, and security. Natural disasters are ecosystem-wide phenomena that can cause imbalances in the supply and demand of social resources and socioeconomic system instability. According to the literature, there are six different types of natural disasters: biological, geological, fire, meteorological, environmental pollution disasters, and maritime disasters (Liu et al. 2020).

Regularly monitoring natural and manufactured catastrophes and their effects on the EU, the European Commission (EC) assists Member States in their attempts to implement the necessary measures via various policies and recommendations, such as the EU Civil Protection Mechanism. The European Commission (EC) actively oversees the impact of natural and human-induced disasters on the European Union (EU) and collaborates with Member States to facilitate the implementation of necessary measures through various policies and recommendations, including the EU Civil Protection Mechanism. According to official European statistics from 2001 to 2020, floods constituted the largest share of weather-related disasters at 41%, storms at 27% and extreme heat at 23%. The remaining 9% of such incidents were associated with various phenomena, including wildfires, droughts, and landslides (Fig. 1) (Weilnhammer et al. 2021; Peterson 2023).

These calamities inflict severe damage upon local economies, landscapes, cultural heritage, human lives, and overall well-being. Over the period from 1980 to 2020, natural disasters incurred an annual cost of over €12 billion for EU countries, affecting the lives of more than 50 million people within the EU (Yanatma 2023). Research into the impact of natural

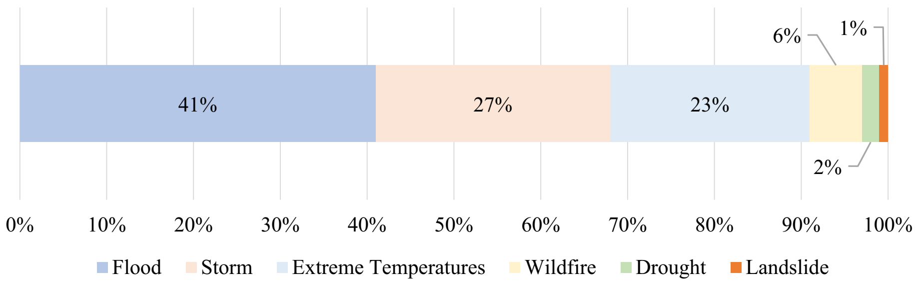

Fig. 1 Distribution of Weather-related disaster incidents in Europe between 2001 and 2020

disasters on individuals is a prominent area of study, focusing on the potential short- and long-term psychological consequences. Ongoing research efforts aim to gather and analyze existing data to enhance our understanding of human factors and their effects during and after natural catastrophes. These endeavors are crucial for optimizing strategies and aligning risk management techniques. The EU introduced the Flood Directive to mitigate the impact of disasters, while the United Nations established the Sendai Framework for Disaster Risk Reduction from 2015 to 2030. These initiatives are essential steps towards minimizing the consequences of such events (European Disaster Risk Management 2023).

Analyzing the statistics above in the context of climate change underscores the critical need to prioritize and enhance preparedness against natural disasters. The central objective remains the protection of life and the assurance of human safety. Acute awareness and understanding of evacuation behavior during such calamities are crucial for ensuring comprehensive and effective human life preparation and protection.

Moreover, optimal resource management emerges as a strategy that stems from delving into evacuation behavior, ensuring the efficient use of available resources when preparing for natural disasters.

Building upon these identified needs, this paper aims to advance the comprehension of disaster preparedness, seeking to contribute to developing improved action plans. Specifically, the innovative machine learning approaches adopted in this study center around using hybrid data analysis systems for more accurate predictions of individuals' behavior during evacuations in the wake of natural disasters.

Delving into a 3-class classification problem, this work focuses on developing computational intelligence models for categorizing individuals based on their adherence to emergency service recommendations during snowstorms, as communicated through the emergency line (112). The identified classes include "I did not make any trip," "I made only the necessary trips," and "I did not limit my trip."

The integration of these methods lays the foundation for advanced warning systems, enabling swift and effective responses to emergencies. The hybrid nature of the machine learning methods allows for a detailed analysis and understanding of factors influencing evacuation behavior in various environments and conditions. The results obtained facilitate the adaptation and personalization of preparedness plans, accounting for the unique characteristics of each region or situation.

Moreover, the study delves into the impact of information systems on the evacuation decision-making process. This analysis can contribute to developing improved warning systems, considering how individuals perceive and respond to the information they receive.

Within the Artificial Intelligence (AI) domain, this study makes a substantial contribution to advancing data analysis derived from questionnaire surveys. Using intelligent methodologies, this paper deliberately addresses the challenges posed by heightened statistical noise, uncertainty, and ambiguity inherent in such survey data. Specifically, this study introduces an innovative and comprehensive methodological framework encompassing essential procedures for data analysis, including data pre-processing, pattern classification, and result evaluation. The ultimate objective is to facilitate documented knowledge mining, thereby enhancing the overall efficacy of the data analysis process.

This work represents a concerted effort to devise innovative methods to enhance natural disaster preparedness. The approach involves seamlessly integrating cutting-edge technology with a detailed understanding of people's behavioral patterns in emergencies arising from natural disasters. The primary purpose of this work is to strengthen the effectiveness of protection measures and to improve risk management practices in cases of natural disasters.

2 Literature review

In this section, we present a comprehensive literature review that encompasses two distinct domains of inquiry. The first domain delves into the intersection of artificial intelligence and its significant contributions to addressing real-world challenges stemming from psychology and risk management within the context of natural disasters. The second domain focuses on analyzing questionnaire data through the prism of machine learning methodologies.

The primary objectives of this dual literature review are twofold. Firstly, it aims to shed light on the invaluable role played by machine learning in tackling real-world issues that closely parallel the problem under consideration in this paper. Secondly, it seeks to chart the landscape of computational intelligence approaches employed in prior research endeavors when dealing with analogous data, specifically, questionnaire data.

2.1 From data to decisions

Emergency evacuation is individuals' swift and efficient relocation to secure locations during crises, such as natural disasters (Joo et al. 2013; Zhao and Wong 2021). While the initial impact of such events may appear minor, their potential escalation and the resulting increase in casualties underscore the importance of promptly evacuating residents from affected areas following a disaster (Dulebenets et al. 2019). Emergency evacuation studies are systematic and intricate, considering factors such as organization, behavioral aspects, First Responders (FRs), logistics, and travel preferences (Liu et al. 2020). Emphasizing the effectiveness of emergency evacuation is crucial for reducing the impact of disasters and safeguarding lives and property.

Understanding evacuation behavior is paramount for enhancing disaster preparedness and response plans. Recent studies highlight the vital role of this knowledge in saving lives and mitigating the impacts of emergencies. For instance, a survey conducted by Johnson et al. (2022) emphasizes the necessity for disaster management officials to comprehensively understand human behavior during evacuations, enabling the development of more effective and targeted evacuation strategies.

Additionally, research by Smith and Lee (2023) demonstrates the creation of predictive models based on analyzing past evacuation scenarios and identifying recurring patterns. These models facilitate the anticipation of potential issues and the optimization of evacuation routes.

Furthermore, research conducted by Brown and Williams (2024) underscores the significance of incorporating elements of human behavior, including responses to panic and decision-making processes, in crafting more astute emergency responses and improving communication strategies. This emerging body of research highlights the critical role of understanding evacuation behavior in devising proactive and adaptable approaches to safeguarding communities during emergencies and disasters (Wong 2020).

The consequences of natural disasters are highly stressful experiences that affect people and society daily (Warsini et al. 2014). Whether a disaster is artificial or natural, it manifests psychosocial symptoms such as anxiety, depression, grief, and stress (Reyes and Elhai 2004). Disrupted daily routines, social support systems, property losses, and relocation intensify the psychosocial impact documented (Mitchell et al. 2008). The acute emotions initially experienced by affected individuals are gradually replaced by chronic psychological and psychiatric problems requiring ongoing care and behavioral interventions (Madrid and Grant 2008).

Clinical depression, post-traumatic stress disorder (PTSD), and substance misuse are among the long-term psychosocial effects of natural disasters (Yzermans et al. 2005; Lin et al. 2002). According to earlier research, the quality of life in afflicted communities has decreased (Mental Health Assistance to the Populations Affected by the Tsunami in Asia - Indonesia | ReliefWeb 2023). After a natural disaster, an individual's quality of life largely depends on their level of psychological and psychiatric impairment. Consequently, the decline in quality of life becomes increasingly pronounced as psychosocial and mental issues become more evident (Papanikolaou et al. 2012; Norris et al. 1994).

Recent research has shown promising results when applying Machine Learning (ML) techniques to analyze and predict evacuation behavior during emergencies. For instance, a study conducted by Smith et al. (2022) demonstrated the effectiveness of deep learning algorithms in assessing real-time data sourced from sensors and social media during evacuations, enabling more accurate and timely predictions of crowd movements and congestion patterns.

Similarly, a study conducted by Johnson and Lee (2023) showcased the benefits of utilizing ML models to identify potential bottlenecks and obstacles in evacuation routes. This empowers emergency responders to optimize evacuation plans, ensuring safer outcomes. These studies underscore the advantages of harnessing ML to extract valuable insights from vast datasets, enhancing the planning and decision-making process during emergency evacuations.

Utilizing the information from the specific section of the literature review described in this work, it becomes clear that the incorporating of Machine Learning (ML) in data analysis processes derived from questionnaires about natural disasters and the behavior of people who have experienced similar events holds immense potential for enhancing evacuation protocols. Machine learning's specific contribution lies in providing emergency management agencies with indispensable means to anticipate and respond to swiftly changing conditions, leading to more efficient disaster response with greater accuracy.

2.2 Knowledge discovery from questionnaire data

Survey data present several challenges, including biases, missing data, and ambiguity, which can impact data quality. In addressing these challenges, machines equipped with advanced data analysis and machine learning capabilities prove indispensable for survey and social science researchers. These machines efficiently process extensive survey data, swiftly identifying hidden trends. Their contribution lies in ensuring accuracy and consistency while managing large datasets, thus minimizing human errors. Beyond essential data management, machines excel in complex data analysis, pattern recognition, and other tasks, allowing researchers to focus on interpretation and analysis. These highlights the crucial role machines play in survey and social science research (Buskirk and Kirchner 2020).

For instance, Terano and Ishino (1995) employ a combination of inductive learning (Flener and Schmid 2008) and genetic algorithms (Deb 1999) with interactive and automated phases to analyze questionnaire data related to consumer goods for marketing decision-making. The core concept of this method involves Integrating inductive learning to generate decision trees or sets of decision rules and using genetic algorithms to extract compelling features, resulting in simple, easily understandable—accurate knowledge extraction from noisy data. This unique approach utilizes both a human-in-the-loop phase (simulated breeding) and an automated genetic algorithm-based phase to evaluate the offspring (decision trees). The study's effectiveness was validated qualitatively and quantitatively using a case study involving consumer product questionnaire data comprising 2400 patterns with 16 attributes.

Similarly, Robertson et al. (1998) assess subjective aspects of dental trauma through questionnaires using various computerized inductive techniques within artificial intelligence. The questionnaires encompassed descriptive variables and questions that reflected the functional, personal, and social impacts of patients' oral conditions following dental trauma. While this methodology might be novel to many in dentistry, the study explains the processes and terminology involved. Initially, a neural network was employed to identify potential relationships within the data. However, the network couldn't make these relationships explicit, so other inductive methods were necessary. Inductive methods (Flener and Schmid 2008) can derive rules from a set of examples, and when combined with domain knowledge, they can reveal connections between variables. The study concludes that artificial intelligence-based methods can significantly enhance explanatory value and clarify database understanding.

In another instance, Deng et al. (2012) present a machine learning approach for deriving trip purposes from GPS track data in passive GPS travel surveys, offering an innovative alternative to traditional paper-and-pencil methods. The method leverages various attributes such as time stamps, land-use types at trip endpoints, spatiotemporal indices, and demographic information to construct a decision tree for classification. Each feature contributes partial evidence to determine a trip's purpose, and they work collaboratively in a reasoning process aided by adaptive boosting. Multiple decision trees (Podgorelec et al. 2002) are generated through voting mechanisms, and their construction is guided by gain ratios computed for relevant attributes. The approach was evaluated with 226 GPS trip records from thirty-six respondents, achieving a promising overall classification accuracy of 87.6% after ten iterations of adaptive boosting. This technique demonstrates the potential of machine learning in accurately categorizing trip purposes from GPS data, providing a more efficient and less burdensome survey method.

Babić (2017) addresses the critical connection between academic motivation and academic performance, emphasizing the need to detect both low and high levels of academic motivation in students. The research aims to create a classification model that predicts student academic motivation based on their behavior in a learning management system (LMS) course. Participants from the Faculty of Education in Osijek were involved, and three machine learning classifiers (neural networks, decision trees, and support vector machines) (Mustapha et al. 2020) were employed. A t-test of the difference in proportions was used to assess the performance of these models. While all classifiers yielded successful results, the neural network model demonstrated the highest success in identifying student academic motivation based on their LMS course behavior.

In their work, Sánchez-Maroon et al. (2017) explore the application of decision trees, a ML algorithm commonly used in data mining, as behavioral models for agents in agent-based models, especially in empirical contexts. Decision trees (Podgorelec et al. 2002) offer transparency and accessibility for domain experts without a computing or artificial intelligence background. However, they are sensitive to construction methods, particularly preprocessing. The paper outlines the processes used to derive decision trees within a model of everyday pro-environmental behavior at work. It compares different preprocessing ways and examines their effects on the models.

Jain et al. (2019) introduce a system for depression analysis and suicidal ideation detection, focusing on predicting suicidal tendencies based on depression levels. Real-time data from students and parents was collected through questionnaires, processed into meaningful data, and used to identify depression severity levels (minimal, mild, moderate, moderately severe, and severe) through machine learning algorithms. The XGBoost (Chen and Guestrin 2016) classifier achieved the highest accuracy at 83.87% in this dataset. Additionally, tweets were collected and classified to determine whether the author is experiencing depression, with the Logistic Regression classifier achieving the highest accuracy at 86.45%.

Chien et al. (2020) with their study, aimed to develop a predictive model for online learners' learning outcomes using ML techniques. The model considered factors such as student motivation, learning tendencies, online learning-motivated attention, supportive learning behaviors, and final test scores. The study involved 225 college students enrolled in online courses over three semesters. Data from the third semester were used as training data. Stepwise logistic regression and random forest (RF) analysis methods (Liu et al. 2012) were employed. The RF approach proved more accurate in predicting final grades with fewer items. Additionally, it identified four things that could potentially remember at-risk learners even before they enroll in an online course.

Babić (2017) present the profound impact of mental health issues, focusing on depression in children and adolescents and its consequences for individuals, families, and society. Early and accurate detection of depression in this age group is paramount. This research pioneers machine learning for detecting depression in 4–17-year-olds, leveraging the robust Young Minds Matter (YMM) dataset. The study's objectives include creating a predictive model for depression, evaluating machine learning algorithm performance, and investigating the relationships between family activities and socioeconomic factors contributing to depression. To achieve these goals, the study utilizes the Boruta algorithm with a Random Forest (RF) classifier to extract vital features for depression detection from highly correlated variables. The Tree-based Pipeline Optimization Tool (TPOT classifier) selects appropriate supervised learning models. In the depression detection phase, RF, XGBoost (XGB), Decision Tree (DT), and Gaussian Naive Bayes (GaussianNB) (Ontivero-Ortega et al. 2017) are employed. Ultimately, this research aims to advance the early identification and prevention of depression in children and adolescents, offering valuable insights into ML methodologies and critical features for effectively predicting this mental health condition.

Dabhade et al. (2021) focus on predicting students' academic performance in a technical institution in India. A dataset was collected through questionnaires and academic records. Data preprocessing and factor analysis were applied to remove anomalies, reduce data dimensionality, and identify the most correlated features. ML algorithms were compared using Python 3, and the support vector regression linear algorithm (Gu et al. 2015) demonstrated superior predictive performance.

The work of Kim et al. (2021) assessed the effectiveness of the Patient Health Questionnaire-9 (PHQ-9) in identifying suicidal ideation. Data from 8,760 completed questionnaires from college students were analyzed, scoring the PHQ-9 in conjunction with four categories (PHQ-2, PHQ-8, PHQ-9, and PHQ-10). Suicidal ideation was evaluated using the Mini-International Neuropsychiatric Interview suicidality module, and ML (ML) algorithms, including k-nearest neighbors (Kramer 2013), linear discriminant analysis (LDA) (Xanthopoulos et al. 2013), and random forest, were used. The results demonstrated that random forest, employing the nine items of the PHQ-9, achieved an excellent area under the curve of 0.841, with 94.3% accuracy. The positive and negative predictive values were 84.95% and 95.54%, respectively. This study affirms that machine learning algorithms using the PHQ-9 in primary care settings are highly dependable in screening individuals with suicidal ideation.

García et al. (2021) identify risk factors for future suicide attempts in the general population using a data-driven ML approach. The analysis involved over 2500 questions from an extensive, nationally representative survey of US adults. Data was collected from two waves of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) conducted with a nationally representative sample of adults in the US. The study employed a balanced random forest model trained using cross-validation to develop a suicide attempt risk model. The results revealed numerous factors associated with suicide attempts, including previous suicidal thoughts and behaviors, functional impairment due to mental disorders, socioeconomic disadvantage, younger age, and recent financial crisis. This information may inform future clinical assessments and the development of suicide risk assessment tools.

Wang et al. (2022) explore using machine learning techniques on self-reported questionnaire data to predict the 10-year risk of cataract surgery in middle-aged and older Australians. The research collected baseline data, including demographic, socioeconomic, medical, lifestyle, and dietary factors, and self-rated health status as risk factors. Cataract surgery events were confirmed using Medicare Benefits Schedule Claims data. Three machine learning algorithms (random forests, gradient boosting machine (Ayyadevara 2018), and deep learning) were compared to a traditional logistic regression model predicting cataract surgery risk. Cross-validation was used for evaluation, with primary outcome measures being the areas under receiver operating characteristic curves. The study included 207,573 participants aged 45 and above without prior cataract surgery. The machine learning algorithms outperformed the traditional model, and critical predictors included age, self-rated vision, and health insurance. In summary, this research demonstrates that machine learning models can accurately predict cataract surgery risk using questionnaire data, with a slight advantage over conventional logistic models.

Sun et al. (2022) study investigates risk factors associated with positive mammographic findings using questionnaire data and machine learning techniques. The goal is to improve breast cancer detection rate and enable early advanced diagnostic studies and treatments. Two machine learning approaches, XGB-SRVCs and Lasso-SRVCs (Chen and Guestrin 2016), are compared to identify the most effective method for predicting positive mammographic findings. The study incorporates demographic and clinical information into the analysis using machine learning.

This comprehensive literature review highlights the diversity of machine learning methods applied to questionnaire data analysis, including Genetic Algorithms, Neural Networks, Support Vector Machines, Random Forest, XGBoost, Logistic Regression, Gaussian Naive Bayes, k-nearest neighbors, linear discriminant analysis, and others. These methods serve various purposes, from data analysis to prediction and classification across domains.

In conclusion, this piece of the total literature review emphasizes the significant potential of machine learning in enhancing the analysis and comprehension of questionnaire data. While comparable approaches are limited in the current literature, these studies provide compelling evidence of machine learning's capacity to unlock valuable insights and deliver practical benefits across diverse domains. As technology and methodologies evolve, the prospect of encountering more innovative approaches promises to enhance further our ability to extract profound knowledge from questionnaire data.

3 Data presentation

This section provides an exhaustive account of the data collection procedures via questionnaires and the subsequent compilation of the dataset used in the following data analysis phases. Furthermore, we present the outcomes and inferences derived from the statistical analysis conducted on the initial dataset; that dataset emerged after undergoing thorough data preprocessing steps given in this section.

The statistical analysis aimed to extract information that was utilized while validating the conclusions emerged by the further data analysis with computational intelligence methods presented in this work.

3.1 Data collection

The data collection process outlined in this section was conducted with the primary objective of acquiring a more expansive and inclusive sample of individuals with direct experience in natural disasters. This data is to be subjected to analysis using statistical methodologies and computational intelligence techniques, with the overarching goal of extracting valuable insights into human behavior. These insights, in turn, can inform critical emergency decision-making during evacuations in the context of natural disasters, such as snowstorms.

Regarding data acquisition, we rigorously obtained ethical clearance from the Institutional Review Board of the American College of Greece and ensured that eligible participants provided informed consent. Notably, the dataset used in this study was sourced from the ANDREAS program of the AEGIS+ research project.

Furthermore, our research received ethical approval and obtained informed consent from individuals who met specific inclusion criteria, including being 18 years of age or older, residing in Greece, having experienced the impact of a natural hazard, and being able to participate and provide informed consent.

In the context of the data collection process, we employed a range of data collection strategies to ensure a diverse and representative sample. These strategies included utilizing various online platforms such as Facebook, Twitter, and LinkedIn, as well as popular online forums like Reddit and Quora. Additionally, we leveraged mailing lists associated with prominent emergency management, disaster response, and resilience organizations, including esteemed entities like Federal Emergency Management Agency (FEMA) and the International Committee of the Red Cross (ICRC).

Finally, after completing the phases of the data collection process (Fig. 2), the primary data set (1197 cases | 62 variables) emerged. The initial dataset was drawn and constructed from the preliminary data set ((1197 cases | 62 variables)) that was used in further data analysis (both statistical and machine learning) and knowledge mining processes.

3.2 Construction of final dataset

The data set used in this work consists of 16 observations (variables) of 525 cases and resulted from a complex process of preprocessing the original data (1197 cases | 65 variables). The specific data preprocessing process includes several corresponding steps described later in this section.

The aim of the overall data pre-processing (Fig. 3) described in the continuation of this section was the construction of a homogeneous and complete data set concerning the cases

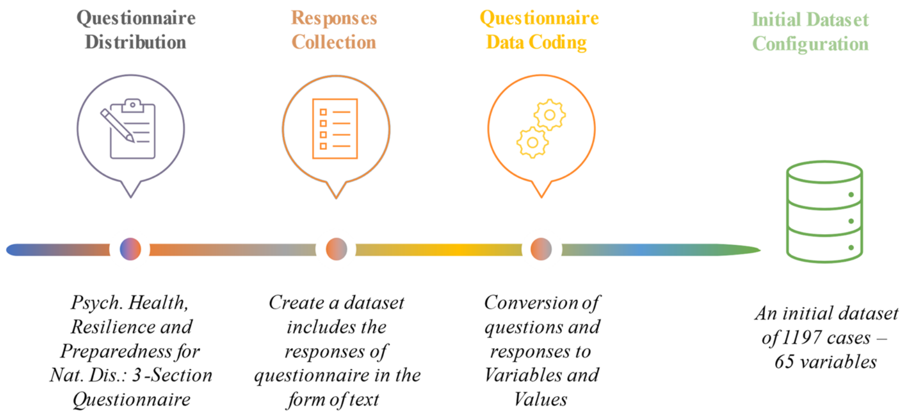

Fig. 2 Data collection process

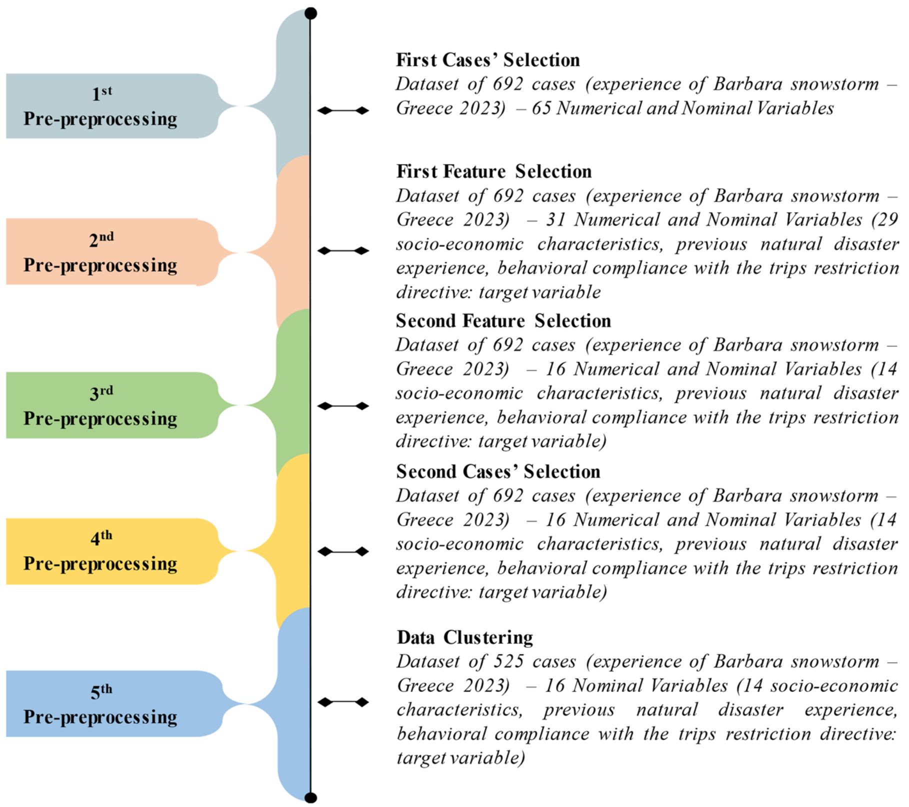

Fig. 3 Construction of final dataset: presentation of data pre-processing

that had experienced the snowstorm known as "Barbara", which affected Greece during the winter of 2023.

According to the relevant report of the National Meteorological Service of Greece (2024), severe weather "Barbara" swept through Greece from January 29 to February 3, 2023, leaving a landscape covered in snow and frost. The intense harsh weather was accompanied by various dangerous phenomena that affected citizens' daily lives. Specifically, during the harsh weather "Barbara", there were heavy snowfalls in mountainous and semi-mountainous areas, while the snow even reached lowland areas such as Attica and Thessaloniki. Temperatures registered a noticeable drop, going as low as − 10 °C in some areas. The frost was intense, creating problems in the road network, especially in the morning. At the same time, fierce winds swept the Aegean Sea, with gusts reaching up to 10 Beaufort. However, the effects of this lousy weather were felt in many areas. Power outages were reported in several areas due to falling snow-covered cables. The road network in mountainous and semi-mountainous regions was closed as a precaution, with snow chains being a necessary accessory for many roads. Ferry and plane cancellations affected travel, while schools in several areas were closed for a day or two due to the snowfall. In conclusion, "Barbara" is a prime example of the need for proper preparation and response to extreme weather conditions, as early warning and coordinated action can avoid serious accidents and disruption to society.

Following the presentation of the data concerning the specific bad weather ("Barbara"), the purpose of homogenization (converting numerical variables, such as age, to nominal variables based on observations of the frequencies of the specific variable) and the completeness of the final data set (i.e., the data set used in the further data analysis of this paper) was to reduce the statistical noise of the data (Gupta and Gupta 2019) that will yield the optimal fit of the machine learning algorithms.

In the initial phase of data pre-processing, we began with the original dataset containing 1197 cases and 65 variables in the initial data pre-processing phase. Our objective was to construct a refined dataset focusing on instances linked to the recent snowstorm known as "Barbara," which affected Greece during the winter of 2023. To accomplish this, we meticulously filtered the data to retain only those cases in which respondents had confirmed their experience with the specific natural disaster (Barbara—Snowstorm) as indicated by their affirmative responses in the collected questionnaires. This rigorous selection process yielded a new dataset of 692 cases, each retaining the original set of 65 variables.

Considering the objective of linking socio-economic characteristics as well as previous natural disaster experience with a person's behavior during the trips restriction order in an area affected by a natural phenomenon (such as a snowstorm), during the second data pre-processing phase, in data set resulting from the first stage of data pre-processing (692 cases | 65 variables/features) feature/variable selection process was performed. After the specific feature selection process, a new data set of 31 variables emerged (29 socio-economic characteristics, previous natural disaster experience, behavioral compliance with the trips restriction directive: target variable).

Then, from the second data set (692 cases | 31 variables) resulting from the second data pre-processing phase, the variables derived from questionnaire questions whose answers were unique (questions to which all 692 participants gave a specific answer. After completing the third data pre-processing stage, a new data set was obtained, including 692 cases | 16 variables (14 socio-economic characteristics, previous natural disaster experience, behavioral compliance with the trips restriction directive: target variable).

During the next stage of data pre-processing, from the data set resulting from the corresponding third stage (692 cases | 16 variables), the cases (i.e., the questionnaires) where even one question was not answered were removed. That is, cases with at least one missing value were deleted. Thus, completing the specific data pre-processing phase, a further data set was obtained, including 525 cases | 16 variables (14 socio-economic characteristics, previous natural disaster experience, behavioral compliance with the trips restriction directive: target variable).

Finally, in the final stage of data pre-processing, the numerical variables "Age" and "Residence Time" values were grouped according to their frequency analysis. This data pre-processing stage resulted in the creation of the data set used in the further data analyses, which includes 525 cases | 16 nominal variables (14 socio-economic characteristics, previous natural disaster experience, behavioral compliance with the trips restriction directive: target variable).

As previously mentioned, the overarching objective of the data pre-processing procedure outlined in this paragraph was to culminate in the creation of a definitive dataset primed for effective utilization in subsequent data analyses of this paper. This ultimate dataset comprises data extracted from 16 nominal variables, including (1) Gender, (2) Age Group, (3) Education, (4) Residence Time, (5) Residence Type, (6) Owned Residence, (7) Household with Child, (8) Household with Seniors, (9) Household with Pets, (10) Net Annual Household Income, (11) Work Status, (12) Work from Home, (13) Transporting Mean, (14) GPS Usage, (15) Natural Disaster Experience, and the target variable, (16) Trip Restriction. This target variable, or dependent variable, encapsulates an individual's response to instructions were issued by the emergency authorities, concerning travel restrictions during the snowstorm season, codenamed "Barbara" in Greece during the winter of 2023.

The detailed presentation of this conclusive dataset, encompassing 525 cases and 16 variables, is expounded upon in the subsequent paragraph.

3.3 Statistical analysis of final dataset

This subsection includes the statistical analysis performed on the final dataset (525 cases | 16 variables), obtained after completing the overall pre-processing of the original dataset (1197 cases | 65 variables).

The statistical analysis of the final data set (525 Cases | 16 Variables) includes two respective phases: (i) the statistical analysis of the frequencies of the values of each variable and (ii) the Chi-Square test.

The purpose of the frequency analysis of specific categorical data includes describing the dataset, identifying anomalies, extracting exciting information and supporting decision-making based on this analysis (Bartholomew 1980).

The Chi-Square test (χ^2) (Franke et al. 2012) is a statistical tool employed to evaluate the relationship between categorical variables. It enables the examination of independence among categories, the validation of data fitness, and the derivation of insights regarding the associations between variables. This statistical test is a powerful instrument for analyzing categorical data and provides valuable insights into the interplay of distinct categories. It has been effectively employed to examine the correlation between the target (or dependent) variable and the other variables present in the final dataset, which served as input (or independent) variables in subsequent data analysis.

The overall statistical analysis presented in this subsection was to obtain helpful information for further data analysis using computational intelligence methods and for the stratified statistical validation of the results of this study.

3.3.1 Frequencies analysis

This paragraph describes in detail all the variables that form the final data set (525 Cases | 16 Variables) and some critical observations from the statistical frequency analysis of the values of each variable.

Specifically, the variable "Gender" represents the gender of each participant in the survey through a questionnaire. It is a categorical variable with the values Male and Female and with respective percentages of 53.70% and 46.30% in each value. As observed in the frequency analysis of this variable, there is high participation from both genders, with the male gender leading with 7.4% over the female gender.

The categorical variable "Age Group" includes the age groups to which the participants of the questionnaire survey may belong. Specifically, it is formed by four age groups: under 36, 37 to 49. 50 to 58 and Over 59, to which the percentages 24.80, 25.30, 23.60 and 26.30% correspond. According to the above frequencies of the values of this variable, it is observed that there is almost equal participation from all age groups.

The variable "Education" expresses the level of education of each participant. It is a categorical variable of five categories. According to the frequency analysis of the values of this variable, although there is participation from almost all levels of education, most of the participants (41.9%) have university education.

The variables "Residence Time", "Residence Type", and "Owned Residence" refer to the residence of the participants and express respectively the time of residence in their residence (Under 10, 11 to 23, 24 to 39 or Over 40), the type of residence (Detached house or Apartment) and whether their residence is owned (Yes or No). Looking at the frequencies of each of these three variables, we conclude that there is almost equal participation according to the time of residence; most participants (62.10%) live in an apartment as well as the dwelling (apartment or detached house) for most participants (69.90%) is privately owned.

Next, the variables "Household with Child". "Household with Senior, "Household with Pets" and "Net Annual Household Income" refer to the household from which each participant comes. In particular, the categorical variables "Household with Child". "Household with Senior, "Household with Pets" express the fact (Yes or No) whether the household includes at least one child, at least one senior and at least one pet, while the categorical variable "Net Annual Household Income" categories the participants into five corresponding monetary categories (Euros): Under 10,000, 1000–2000, 20,000–50,000, Over 50,000. The leading information provided by the frequency analysis of the values of these variables is that the most significant percentage of participants (40.80%) have an annual household income between 20,000 and 50,000 Euros.

Then, the categorical variables "Work Status" and "Work Style" form the work profile of each participant. Specifically, the variable "Work Status" categorizes each participant into three categories (full-time, Retired or Other) according to their working hours and the variable "Work Style" categorizes the participants into four types (Telecommuting, Via living, Via living and telecommuting or neither) according to their working style. According to the analysis of the frequencies of the values of these variables, it is observed that most of the participants (58.90%) are employed full time, and the working style of most of the participants (55.60%) is telecommuting.

The variables "Transporting Mean" and "GPS Usage" refer to the usual mode and means of transport and the use of GPS by each participant. Specifically, the categorical variable "Transporting Mean" represents whether a participant commutes by car or uses other modes or means of transport (walking, public transportation, bicycle, etc.). In contrast, the variable "GPS Usage" represents the frequency (Yes, Rarely or No) of GPS use exhibited by each participant, regardless of their mode and means of transport. Looking at the frequencies of these variables, most participants (65.50%) use their car for their commutes and when they travel (regardless of the means and mode of travel), most (61.70%) use and follow the GPS instructions.

Finally, the variables "Natural Disaster Experience" and "Trip Restriction" are related to the natural disasters experienced by the participants. Specifically, the categorical variable "Natural Disaster Experience" represents whether a participant has experienced a natural disaster in the past (before 2023), and the variable "Trip Restriction" represents each participant's behavior (I did not make any trip, I made only the necessary trips or I did not limit my trips) in complying with the travel restriction guideline. According to the analysis of the frequencies of the values of these variables, it is observed that most of the participants (74.70%) have experienced a past (before 2023) natural disaster. Of all the participants who experienced the snowstorm "Barbara" (Greece, Winter 2023), they did not follow the directive of limiting their trips as the highest percentage (73. 5%) either made the (subjectively) necessary trips (57.10%) or did not follow at all (16.40%) the specific instruction given by the emergency line.

Of the variables described in detail above and summarized in Table 1, the target variable (or dependent variable) of further data analysis with machine learning methods is the variable "Trip Restriction", which is formed by three categories of classification of the cases (i.e., the participants of the survey through a questionnaire: I did not make any trip, I made only the necessary trips, or I did not limit my trips.

3.3.2 Results of chi-square test

This subsection presents the results of the independence test between the target variable (Trip Restriction) and each input variable. According to the Chi-Square independence test, the null hypothesis was defined as complete independence (Asymptotic Significance over 5%) of the target variable from the input variable under test. Also, it is noteworthy to mention that before performing this statistical test, the prerequisite test (minimum number of cases in each category of each variable over five) was performed and verified to produce safe conclusions. Specifically, the Table 2 and the corresponding graph (Fig. 4) present the asymptotic significance (%) that each input variable shows with the target variable.

According to the results of the Chi-square statistical test (Table 2 and Fig. 4), the variables that show complete dependence (asymptotic significance under 5%) with the target variable are the variables "Work Style", "Age Group", "Work Status", "Owned Residence", "Gender" and "Household with Seniors" and the variables showing complete independence (asymptotic significance over 5%) are the variables "Net Annual Household Income", "Household with Pets", "Natural Disaster Experience", "Education", "Household with Child, "Residence Time" and "Residence Type".

Finally, the variables, with which the target variable tends to show some dependence, are "Transporting Mean" and "GPS Usage". The information and conclusions obtained from the specific statistical control are noteworthy and are used in the process of the statistical evaluation of the proposed data analysis models presented in this work.

Table 1 Summary table of variables of the final dataset (525 Cases|16 Variables)

S.n.	Variable	Values	Frequency	Percent (%)
1	Gender	Male	282	53.70
		Female	243	46.30
2	Age Group	Under 36	130	24.80
		37–49	133	25.30
		50–58	124	23.60
		Over 59	138	26.30
3	Education	High School Part 2	97	18.50
		Outstanding Educational Institution	50	9.50
		University or College	220	41.90
		Master	135	25.70
		PhD	23	4.40
4	Residence Time	Under 10	138	26.30
		11–23	137	26.10
		24–39	116	22.10
		Over 40	134	25.50
5	Residence Type	Detached house	199	37.90
		Apartment	326	62.10
6	Owned Residence	Yes	367	69.90
		No	158	30.10
7	Household with Child	Yes	237	45.10
		No	288	54.90
8	Household with Seniors	Yes	105	20.00
		No	420	80.00
9	Household with Pets	Yes	176	33.50
		No	349	66.50
10	Net Annual Household Income	Under 10,000	94	17.90
		1000–2000	185	35.20
		20,000–50,000	214	40.80
		Over 50,000	32	6.10
11	Work Status	Full time	309	58.90
		Other	124	23.60
		Retired	92	17.50
12	Work Style	Telecommuting	38	7.20
		Via living	292	55.60
		Via living and telecommuting	91	17.30
		Neither	104	19.80
13	Transporting Mean	Car	344	65.50
		Other	181	34.50
14	GPS Usage	Yes	324	61.70
		Rarely	101	19.20
		No	100	19.00
15	Natural Disaster Experience	Yes	392	74.70
		No	133	25.30

Table 1 (continued)

S.n.	Variable	Values	Frequency	Percent (%)
16	Trip Restriction	I did not make any trip	139	26.50
		I made only the necessary tips	300	57.10
		I did not limit my trips	86	16.40

Table 2 Summary Table of Results of the Ch-Square Statistical Test (Targe Variable: Trip Restriction)

Variable	Asymptotic significance (%)
Work Style	0.00
Age Group	0.10
Work Status	0.10
Owned Residence	1.50
Gender	3.50
Household with Seniors	4.30
Transporting Mean	5.40
GPS Usage	8.70
Net Annual Household Income	32.80
Household with Pets	34.50
Natural Disaster Experience	39.70
Education	62.00
Household with Child	71.70
Residence Time	80.80
Residence Type	84.30

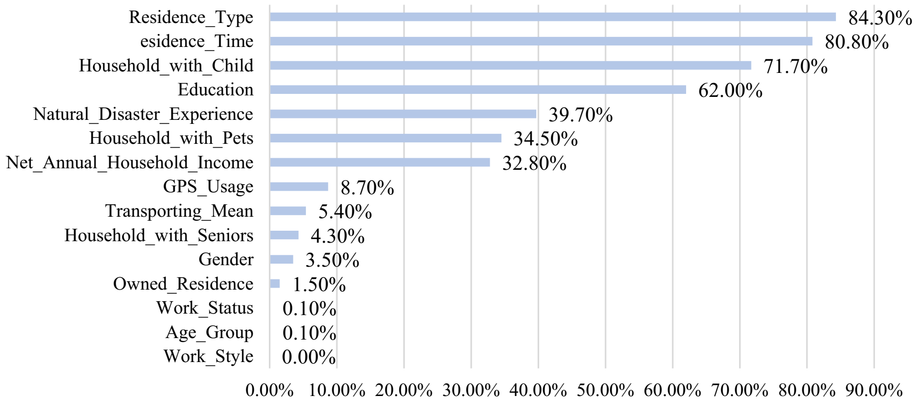

Fig. 4 Summary graph of results of the Ch-Square statistical test (Targe Variable: Trip Restriction)

4 Experimental methodology

In this section, we introduce the methodological frameworks devised within this study to address the challenge of categorizing compliance behavior concerning the movement restriction directive, as observed during a snowstorm event like "Snowstorm Barbara," which impacted Greece during the winter of 2023.

In particular, we have designed and applied three distinct methodological frameworks that encompass combinations of (i) data balancing, (ii) classification techniques, and (iii) validation methods rooted in the broader realm of machine learning. The distinguishing factor among these three frameworks lies in how data balancing methods are integrated into the overall modelling process.

Within the initial methodological framework (Fig. 5), the data balancing process is omitted. Both model development and evaluation are conducted using the original dataset, and the assessment of classification models takes place through established validation methods, namely: (i) utilizing the Training Set, (ii) Use Test Set, and (iii) k-fold Cross Validation.

In the second methodological framework (Fig. 6), we incorporate a data balancing step, executed before developing and evaluating classification models. This means that the classification models are developed and assessed using a dataset in which cases have been equalized across the categories of the target variable. The validation procedure for these refined classification models remains consistent, employing the same classification methods previously mentioned.

The third methodological framework (Fig. 7) introduces an innovative approach by simultaneously developing and validating classification models. This approach incorporates balancing, classification, and k-fold Cross Validation methods. In this third framework, data balancing is exclusively applied to the training set generated during each random partition of the original dataset into k folds as part of the k-fold Cross Validation process. This entire process is iterated k times.

For instance, if we set k = 4 in the context of the k-fold Cross Validation, four distinct experimental phases are executed. In the first experimental phase, a 4-fold Cross Validation process is conducted. During each of the four stages within this Cross

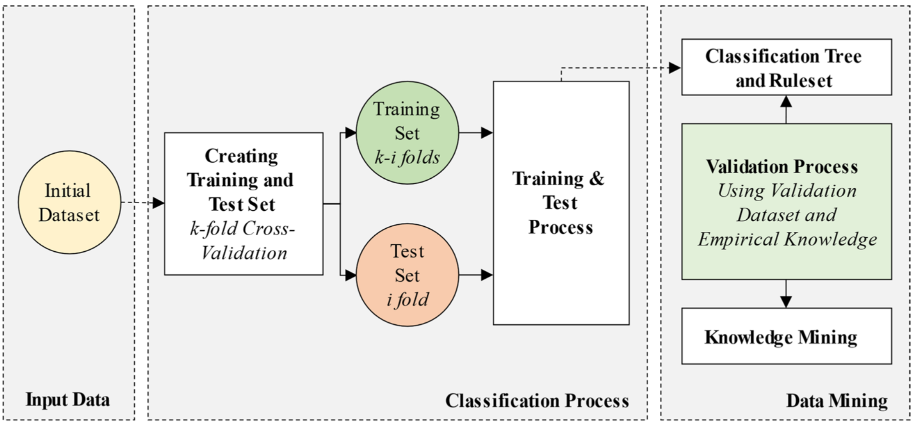

Fig. 5 Fiest Methodological Framework

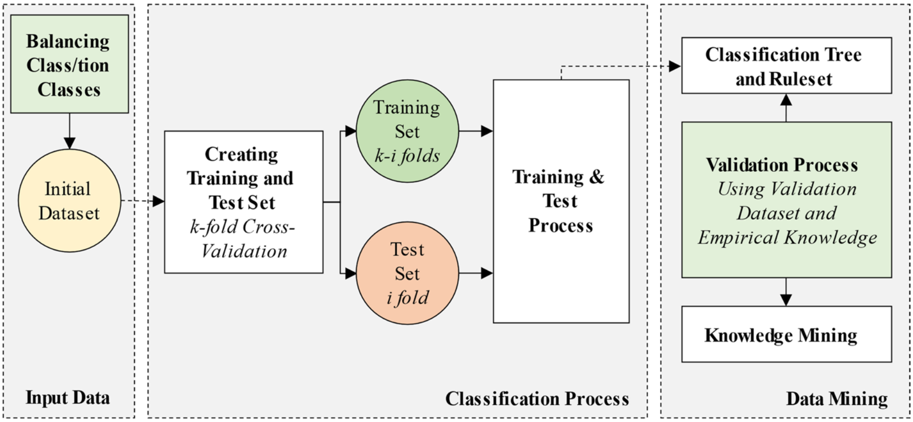

Fig. 6 Second Methodological Framework

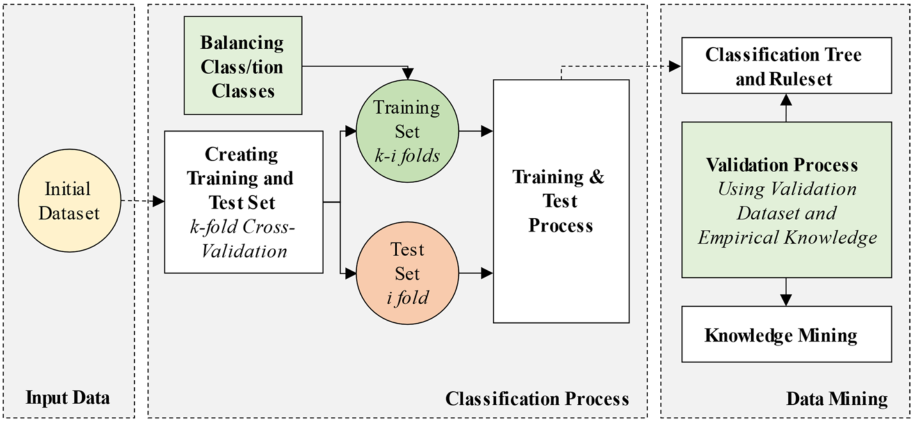

Fig. 7 Third Methodological Framework

Validation, classification models are trained using k-i folds of data, with each train set (k-i folds of data) being balanced using the corresponding method and then evaluated with the i-to fold. This sequence is repeated in the subsequent three experimental phases, where the original data set is randomly divided into four new folds each time. The model's final validation is accomplished by aggregating the statistical evaluation measures across all experimental phases to yield comprehensive insights into its performance.

Finally, the methods of (i) data balancing and (ii) classification used in all the above methodological frameworks were respectively (i) the SMOTE method, (ii) Inductive Decision Trees, (iii) Simple Logistic Regression and the statistical measure evaluation of the generated classification models was the classification accuracy they exhibited in each validation phase.

4.1 Balancing with SMOTE

SMOTE, an acronym for Synthetic Minority Over-sampling Technique, stands out as a robust method in machine learning and data analysis. Its primary purpose is to tackle class imbalance concerns within datasets. When one class significantly surpasses another—typically referred to as the minority class and the majority class—conventional machine learning algorithms can exhibit a bias toward the majority class, thereby leading to suboptimal performance for the minority class (Elreedy and Atiya 2019).

SMOTE leverages a clever approach to rectify this imbalance: it generates synthetic instances of the minority class. This is achieved by selecting a data point from the minority class and strategically creating new artificial data points. These synthetic data points are strategically positioned along the line segments connecting the chosen topic with its nearest neighbors (Kramer 2013). In effect, this process effectively balances the class distribution, empowering machine learning models to learn from the minority class without being overwhelmed by the presence of the majority class (Elreedy and Atiya 2019) (Fig. 8).

SMOTE is widely used to counter class imbalance in machine learning. However, it comes with certain drawbacks. One significant concern is that introducing synthetic samples may contribute noise to the dataset, hampering the model's ability to generalize effectively. The interpretability of the model is also compromised, as the inclusion of synthetic instances increases complexity (Fernandez et al. 2018; Batista et al. 2004; Chawla et al. 2002; Agrawal et al. 2015).

Additionally, SMOTE's dependency on the density of the feature space and its sensitivity to the choice of the k-value can limit its applicability, particularly in sparse or high-dimensional datasets. The computational burden of generating synthetic samples poses another challenge, impacting training times, especially for larger datasets. Despite these disadvantages, SMOTE remains a valuable tool, but its application should be carefully considered based on the specific characteristics of the dataset at hand (Fernandez et al. 2018; Agrawal et al. 2015).

4.2 Balancing with LCC

In this research paper, utilizing SMOTE, we introduce a novel data levelling approach known as LCC (Leveling of Cases per Class). We assess its impact on the performance of classification models and compare it with the effects of the SMOTE method. This balancing technique addresses imbalanced data, contributing to the development of intelligent classification models that demonstrate optimal performance and generalization. The proposed data balancing methodology thoroughly analyses the frequency distribution of

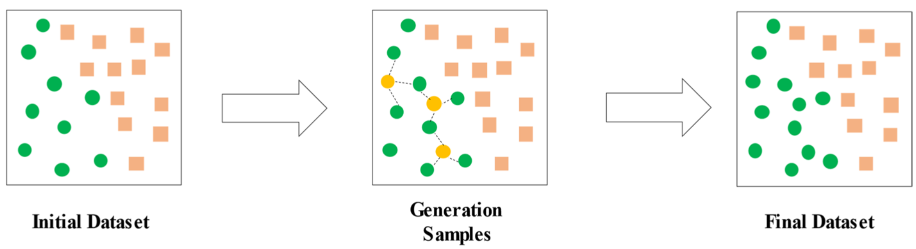

Fig. 8 Synthetic Minority Oversampling Technique (SMOTE)

classification class values within a given classification problem. Its primary goal is to uphold the information gain present in the initial dataset, ensuring parity between the information gain of the initial dataset and that of the final dataset.

The proposed data balancing method (Fig. 9) is about nominal data balancing and includes the following steps: (i) Frequency analysis of classification class values, (ii) Finding the classification class that gathers the biggest number of hits, (iii) Dividing the number of cases of the largest classification class by the number of cases of each smaller classification class, (iv) Multiplying each smaller classification class by the corresponding integer quotient from the above step and (v) In each classification class, adding several cases, selected from the data set by random draw, equal to the remainder of the division of the largest class by the corresponding (quotient) smallest class.

$$\mathrm{d_b} = \mathrm{d_m} \cdot \Pi + \upsilon_{\mathrm{D}/\mathrm{d_m}} \mathrm{d_b}, \mathrm{d_m}, \Pi, \upsilon_{\mathrm{D}/\mathrm{d_m}} \in N \tag{1}$$

Where D: the integer number of cases of the largest class (C_{max}), $\mathrm{d_m}$: the integer number of cases of class C_v with $C_v \leq C_{max}$, Π: the integer quotient of the division $\frac{\mathrm{D}}{\mathrm{d_m}}$, $\upsilon_{\mathrm{D}/\mathrm{d_m}}$: the integer remainder of the division $\frac{\mathrm{D}}{\mathrm{d_m}}$ and $\mathrm{d_b}$: the number of cases of class C_v resulting after the balancing process.

Finally, to examine and observe potential biases, both when using the specific balancing method and when applying the SMOTE method, cross-validation was performed, where at each stage of the cross-validation, the data of the training set was balanced as described in more detail in the presentation of the Third Methodological Framework.

4.3 Inductive machine learning

Inductive machine learning, a cornerstone in artificial intelligence and data science, is centered around the art of deriving overarching insights from particular data points, anticipating future outcomes or categorizing previously unobserved instances. This methodology is advantageous when grappling with intricate, unstructured data within real-world applications (Michalski 1983; Gahegan 2003).

Inductive machine learning involves extracting patterns, rules, or models from a dataset using concrete examples. It operates by discerning shared attributes and associations within the training data and employing these insights to forecast fresh, uncharted data characteristics. The true strength of inductive machine learning lies in its adaptability and capacity to extrapolate from past experiences, equipping systems with the capability to make well-informed, real-time decisions and predictions. This methodology has transformed many sectors, from healthcare and finance to autonomous vehicles and recommendation systems. As technological progress marches forward, inductive machine learning continues to play

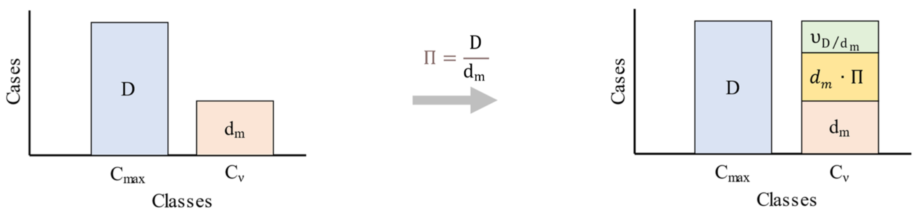

Fig. 9 Leveling of Cases per Class Balancing Method

a pivotal role in shaping the future of AI, enhancing our ability to harness the immense potential latent within vast and intricate datasets (Dutton and Conroy 1997).

Basic inductive machine learning methods are the foundational building blocks for more advanced and intricate machine learning models. These techniques find application in various tasks, including classification, regression, and clustering. Notable fundamental methods encompass Linear Regression, Logistic Regression, K-Nearest Neighbors, Decision Trees, Random Forest, Support Vector Machines, K-Means Clustering, Principal Component Analysis, and Association Rule Mining (Michalski 1983; Lopez de Mantaras and Armengol 1998).

In this study, two basic inductive learning methods, Simple Logistic Regression, and Inductive Decision Trees, were employed to analyze categorical data emanating from questionnaire responses.

Simple logistic regression and inductive decision trees, also known as inference trees, each bring distinct advantages when dissecting categorical data. Logistic regression offers interpretability by furnishing transparent coefficient-based results, excels in handling categorical predictors, issues probabilistic outputs, and boasts regularization capabilities to combat overfitting. Conversely, decision trees harness non-linearity to capture intricate interactions, demonstrate grace in handling missing data, facilitate ease of interpretation, autonomously select essential features, and serve as foundational components in powerful ensemble methods. The choice between these methods should pivot on the specific characteristics of the dataset and the nature of the problem at hand (Michalski 1983).

4.3.1 Simple logistic regression

Simple Logistic Regression is a fundamental statistical technique for analyzing the relationship between a binary dependent variable and one or more independent variables. In its basic form, it's employed when the outcome variable is dichotomous, meaning it has only two possible outcomes: yes/no, success/failure, or 0/1 (Peng et al. 2002).

The primary goal of Simple Logistic Regression is to model the probability of one of the binary outcomes as a function of the independent variable(s). It accomplishes this by fitting a logistic curve (S-shaped) to the data, which can capture the non-linear relationship between the independent variable(s) and the probability of the outcome occurring (Peng et al. 2002; Sperandei 2014).

This technique is widely used in various fields, including medicine, economics, and social sciences, for tasks like predicting disease occurrence, analyzing customer behavior, or studying the impact of a particular factor on a binary event. Simple Logistic Regression provides valuable insights into the probability of an event happening, making it a crucial tool in statistical analysis and predictive modelling (Sperandei 2014).

In this work, the SimpleLogistic algorithm was used. This algorithm is designed to build linear logistic regression models using LogitBoost (Li 2012) with simple regression functions as its foundational base learners. To ensure model efficiency, the optimal number of LogitBoost iterations is determined through a cross-validation process, which also enables automatic attribute selection (Landwehr et al. 2005).

4.3.2 Inductive decision trees

As mentioned above, inductive decision trees are a versatile machine learning method renowned for their adeptness in data-driven decision-making and prediction. They offer the unique capability to handle categorical and numerical data, making them indispensable tools in diverse domains like finance, healthcare, and marketing (Dahan et al. 2014; Perner 2015).

The inductive decision trees algorithm used in this work was the C4.5 algorithm. The C4. algorithm is an updated commercial version of the C4.5 algorithm (Quinlan, 1993). Generally, it addresses the primary challenge of induction decision trees, which corresponds to finding the root and the branches of a tree, choosing the most suitable feature (variable) in every development stage of a decision tree, using entropy information criteria, precisely the information gain criterion (Quinlan 1993).

$$\mathrm{Gain}(\mathrm{S},\mathrm{A}) = E(\mathrm{S}) - \sum_{\mathrm{i}=1}^{\mathrm{m}} \mathrm{Pr}\left(\mathrm{A_i}\right) \cdot \mathrm{E}\left(\mathrm{S_{Ai}}\right) \tag{2}$$

where: S: cases set, A: variable, m: number of values of variable A in S, $\mathrm{Pr}(\mathrm{A_i})$|: frequency of cases that have Ai value in S, $\mathrm{E}(\mathrm{S_{Ai}})$|: subset of S with items that have Ai value, E(S): information entropy of S and Gain (S, A): gain of S after a split on attribute A.

Entropy is calculated before information gain is calculated and used to deter-mine how "informative" an attribute is. The basic formula of entropy:

$$\mathrm{E}(\mathrm{S}) = -\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Pr}\left(\mathrm{G_i}\right) \cdot \log_2 \mathrm{Pr}\left(\mathrm{G_i}\right) \tag{3}$$

where: S: cases set, $\mathrm{G_i}$: frequency of class $\mathrm{C_i}$ in S, n: number of classes in S and E(S): information entropy of S.

The C4.5 algorithm, over other inductive decision trees algorithms, can trans-form a classification tree into rules called rulesets. The tree's transformation into a ruleset consists of converting the tree paths into simple "IF / THEN" rules and pruning each rule (having as an evaluation criterion the classification accuracy of each rule) to yield the final ruleset. The ruleset's main advantage is that it is more understandable than trees, as it describes with simple logic sentences a specific context associated with a class. Other benefits of the ruleset are that they contain fewer classification rules than the rules derived from tree paths and, in some cases, are more accurate predictors than trees.

In the context of improving the performance of the classification model presented in this work, the inductive learning algorithm (algorithm C4.5) was optimized with the adaptive boosting method (Quinlan 1993, 1987).

Adaptive Boosting (or AdaBoost) is a supervised ensemble learning algorithm developed by Freund and Schapire in 1995. Adaptive Boosting reduces the error of any ML algorithm (such as Inductive Decision Trees) by sequentially turning many weak classifiers into one robust classifier. This can be accomplished with sequential weight adjustments, individual voting powers and a weighted sum of the final algorithm classifiers (Freund et al. 1999).

4.4 Validation process

Validation processes in machine learning are crucial for assessing and improving the performance of a model. They help ensure that the model generalizes well to new, unseen data. Some standard validation processes in machine learning are Train-Test Split or Use Train Set, k-fold Cross Validation, Stratified Cross Validation, Leave-One-Out Cross Validation (LOOCV), Nested Cross Validation, Time Series Cross Validation, Bootstrapping, Hold-Out Validation and Monte Carlo Cross Validation (Polyzotis et al. 2019; James et al. 2023).

Each validation process has advantages and is suited to different scenarios. The choice of validation method depends on the nature of the data, the problem at hand, and the goals of the machine learning project.

In this research study, considering the unconditional nature of the dataset, the relatively small sample size comprising 525 cases, and the overarching goal of knowledge discovery, three fundamental validation methodologies were employed:

- Training Set Evaluation: This involved assessing the model's performance using a designated training set, providing initial insights into its capabilities.
- k-fold Cross-Validation: To mitigate the limitations of a small dataset, k-fold cross-validation was implemented. This technique systematically partitioned the data into subsets, iteratively training and evaluating the model, ensuring a comprehensive assessment of its robustness.
- Hold-Out Validation: Hold-out validation was utilized in addition to the other methods. This involved reserving a distinct validation set to fine-tune the model before the final evaluation, optimizing its performance.

These validation procedures were strategically selected to cater to the unique characteristics of the data and the objectives of knowledge extraction, collectively ensuring the reliability and generalization of the machine learning approach.

Finally, the validation of the ruleset was conducted both with the use of specific statistical criteria, as well as with the use of the knowledge and experience of the collaborating experts.

To use the following statistical evaluation criteria, the rules of the ruleset were transformed into association rules of the $X \rightarrow Y$ format, where: X: the antecedent (the combination of input variables with the corresponding values, for each rule) and Y: the consequent (the class of each rule) from the ruleset.

Specifically, the statistical criteria used to evaluate the ruleset presented in this paper was the confidence for a $X \rightarrow Y$ rule, which defines how many of the cases containing X, also contain Y as a percentage of the total number of cases containing X. The C4.5 algorithm uses the Eq. (4) (Laplace ratio) to estimate the confidence of each rule (Robertson et al. 1998).

$$\mathrm{conf}(X \rightarrow Y) = \frac{(X \cup Y) + 1}{X + 2} \tag{4}$$

4.5 Knowledge mining

Knowledge mining, or knowledge discovery or extraction, is a pivotal process rooted in data science and machine learning. It involves extracting valuable insights from extensive datasets unveiling hidden patterns and trends. Knowledge mining transforms raw data into structured information by employing techniques such as data preprocessing, pattern recognition, clustering, and classification, facilitating informed decision-making and a competitive edge. Its applications span diverse fields, from business intelligence and scientific research to healthcare and finance, and it remains increasingly vital as data volumes surge, fueling innovation and enhancing decision-making capabilities (Maimon and Rokach 2010).

The final knowledge mining process of this work is based on the validation process and aims to produce new knowledge in the form of IF/THEN rules.

Specifically, utilizing the ruleset of the final classification model and using the statistical criterion of the confidence presented by each classification rule and the number of cases gathered by a classification rule, the strongest classification rules of the final classification model were collected and presented.

Any rule that exceeds the 90% confidence limit and contains more cases than the average of cases collected by all rules was characterized as strong.

5 Result and findings

This paper addresses a 3-class classification problem by employing simple logistic regression and the inductive decision tree method optimized through the AdaBoost algorithm. The specific classification task pertains to categorizing research participants into three distinct classes based on their responses to a questionnaire titled "Psychological Health, Resilience, and Preparedness for Natural Disasters." This survey was conducted as part of the ANDREAS service within the broader AEGIS+ project.

The three classification categories are derived from the primary research question. It examines whether participants adhered to the emergency line's (112) recommendation to limit their travel during the snowstorm "Barbara" that affected Greece in 2023. The questionnaire inquired: "During the recent snowstorm, Barbara, there was a strong recommendation through 112 to avoid trips. How did you respond?" Participants could choose from the following responses: (i) I refrained from travelling while the recommendation was in effect. (ii) I only travelled when necessary and (iii) I did not restrict my trips (Table 4).

This section unveils the results of implementing the methodological frameworks delineated in Sect. 4. The data utilized at each stage of the proposed methodologies originates from the final dataset, comprising 525 cases and 16 variables, as elaborated in the data preprocessing phase detailed in Sect. 4.

To elucidate, the initial dataset for each methodological framework underwent the following structure transformation:

Beginning with the final dataset of Sect. 3, which comprised 525 cases and 16 variables, a subset of 35 cases | 16 Variables was randomly selected to constitute the test set employed in the "Hold Out" evaluation procedure. Consequently, the training dataset for each methodological framework, comprising 490 cases and 16 variables, was formulated.

The size of the test set, comprising 35 cases, may seem notably small, constituting approximately 7% of the original dataset, which contained 525 patients. This limited test set size is primarily attributed to the inherent scarcity of cases within the original dataset.

Furthermore, the specific count of 35 cases was meticulously selected to facilitate the validation process, particularly within the first methodological framework discussed in Sect. 4. For this purpose, a "10-fold Cross Validation" approach was employed, which leveraged the remaining 490 cases. This deliberate choice created 10 equally sized folds, each containing 49 patients, ensuring robust and reliable validation.

In the classification phase within each methodological framework, the variable "Trip Restriction" was designated as the output variable. In contrast, the remaining 15 variables were categorized as input variables, as illustrated in Table 3.

Table 3 Input and output variables of 3-class classification problem

Variable	Type
Gender	Input
Age Group	
Education	
Residence Time	
Residence Type	
Owned Residence	
Household with Child	
Household with Seniors	
Household with Pets	
Net Annual Household Income	
Work Status	
Work Style	
Transporting Mean	
GPS Usage	
Natural Disaster Experience	
Trip Restriction	Output

Table 4 Classes of output variable (Trip Restriction)

Class	Training set (N)	Training set (%)	Test set (N)	Test set (%)
I refrained from travelling	128	26.12	11	31.43
I only travelled when necessary	287	58.57	15	42.86
I did not restrict my trips	75	15.31	9	25.71
Total:	490	100.00	35	100.00

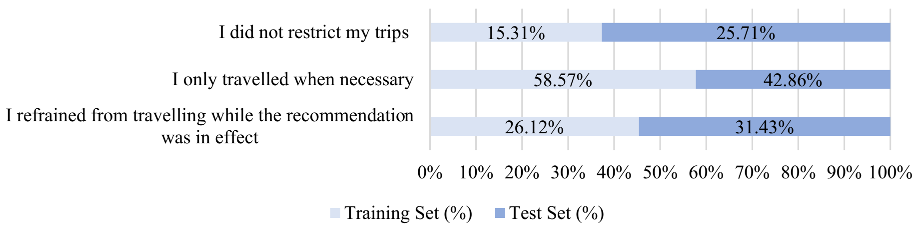

Fig. 10 Classes of output variable (Trip Restriction) (490 Cases)

As seen in Table 4 and the diagram in Fig. 10, there is a high degree of imbalance of classification classes, as half and more cases are concentrated in one (of the three) classification classes.

Table 5 Framework 01—classification accuracy of the best classification models

Validation method	Use training set (490 Cases) (%)	10-fold cross validation (%)	Hold out (35 Cases) (%)
Decision Tree	99.60	46.90	54.30
RuleSet of Decision Tree	99.60	57.10	53.70
Simple Logistic Regression	58.30	56.70	37.14

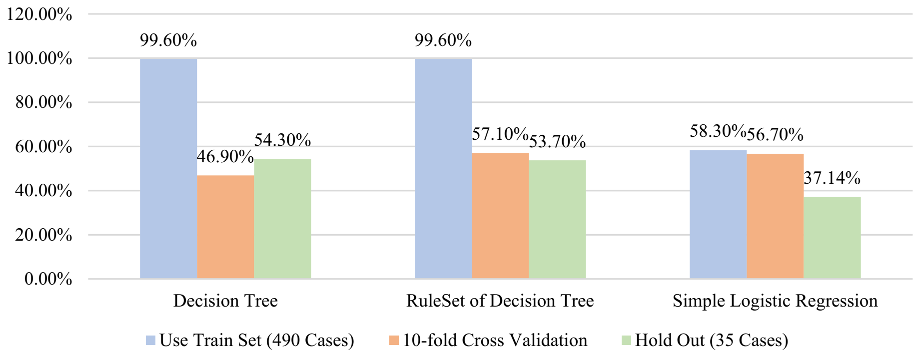

Fig. 11 Framework 01—classification accuracy of the best classification models

5.1 Results of the first framework

In line with the initial methodological framework (Framework 01) presented in Sect. 4, the construction of classification models was undertaken using a dataset consisting of 490 cases and 16 variables. This involved the application of both the logistic regression method (as detailed in Sect. 4) and the inductive decision tree method (also discussed in Sect. 4).

To evaluate the effectiveness of these classification models, results were validated using three distinct techniques: "Use Training Set," "10-fold Cross Validation," and "Hold Out." The outcomes of these validations are summarized in Table 5, while Fig. 11 visually represents the performance metrics of the top-performing models for each specific methodological case.

Based on the performance evaluation of the classification models detailed in Table 5, it's evident that the inductive decision tree method consistently yields the highest accuracy rates across all three validation methods. Nonetheless, it's noteworthy that this method demonstrates relatively lower accuracy rates during the "10-fold Cross Validation" and "Hold Out" validation procedures.

The diminished performance observed in the "10-fold Cross Validation" and "Hold Out" methods can be attributed to several factors. First, the imbalance within the classification classes of the target variable exerts a considerable influence, making it more challenging for the models to classify instances in the minority class correctly. Furthermore, statistical noise stemming from the abundance of variables, the number of cases, and potential biases and random responses within the questionnaire data used to

Table 6 Framework 02—classification accuracy of the best classification models (Balancing with SMOTE)

Validation method	Use training set (490 Cases) (%)	10-fold cross validation (%)	Hold out (35 Cases) (%)
Decision Tree	99.80	73.10	57.30
RuleSet of Decision Tree	99.60	72.90	54.20
Simple Logistic Regression	63.20	59.40	34.20

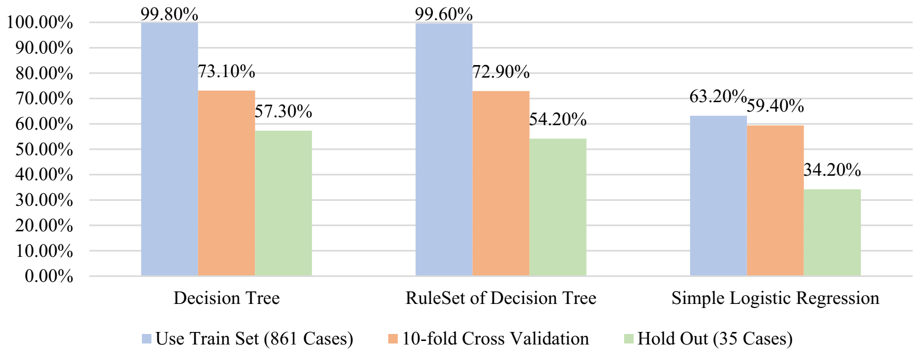

Fig. 12 Framework 02—classification accuracy of the best classification models (Balancing with SMOTE)

create the dataset all contribute to the variance in performance. These factors collectively impact the models' ability to generalize effectively and consistently across different validation scenarios.

5.2 Results of the second framework

Given the suboptimal performance of the classification models when applying the initial methodological framework outlined in Sect. 4, a subsequent methodological framework, denoted as "Framework 02," was implemented. This revised approach encompasses a data balancing step utilizing the SMOTE method, as detailed in Sect. 4. This balancing process occurred before the development and validation of the classification models.

Specifically, when applying the SMOTE or LCC method to the data set (490 Cases | 16 Variables) used in the previous methodological framework (Framework 01), a new data set of 861 Cases | 16 Variables with 287 Cases in each classification class of the output variable (Table 4).

Table 6 and the corresponding graphical representation in Fig. 12 offer an insightful analysis of each model's performance within the context of the three validation procedures employed in the initial methodological framework. These results shed light on the effectiveness of "Framework 02" in addressing the performance limitations observed in the first methodological framework.

As per the insights gleaned from Table 6 and the accompanying visual representation in Fig. 12, a notable enhancement in the classification models' performance becomes apparent when the data balancing method is implemented before developing

Table 7 Framework 02—classification accuracy of the best classification models (Balancing with LCC)

Validation method	Use training set (490 Cases) (%)	10-fold cross validation (%)	Hold out (35 Cases) (%)
Decision Tree	99.80	89.70	62.30
RuleSet of Decision Tree	99.70	88.40	60.20
Simple Logistic Regression	71.20	62.40	39.40

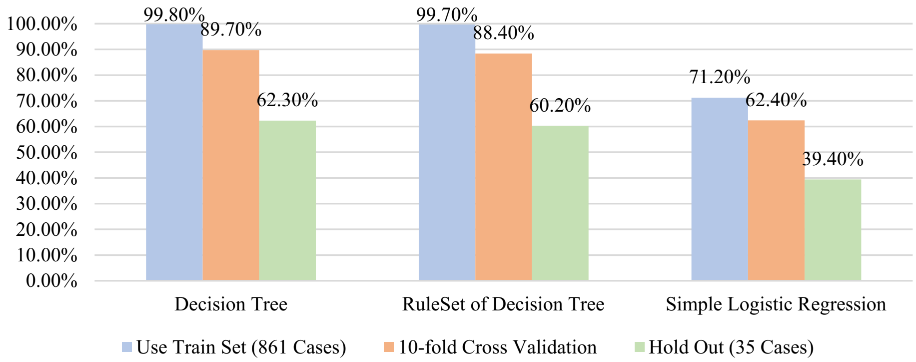

Fig. 13 Framework 02—Classification accuracy of the best classification models (Balancing with LCC)

and validating these models. Within the specific testing phase encapsulated in "Framework 02," it's evident that the inductive decision tree model, along with its RuleSet counterpart, excels in achieving significantly higher classification accuracy when contrasted with the simple logistic regression model.

This discrepancy underscores that there may be more suitable approaches for addressing classification classes from similar questionnaires than the simple logistic regression method.

However, it's worth noting that the performance of the inductive decision tree model and its RuleSet counterpart registers lower accuracy levels during the "10-fold Cross Validation" and "Use Test Set" validation procedures. This observation hints at the continued presence of statistical noise within the dataset, which may emanate from both the data balancing process—potentially introducing bias into the classification models—and the dataset, which encompasses many variables.

Subsequently, when using the new balancing method called LCC, the generated classification models present the same as well as higher performances (Table 7 and Fig. 13).

As evidenced by the corresponding table and diagram (Table 7 and Fig. 13), the superior method, even in the case of using the LCC balancing method, is the inductive decision tree method. This is confirmed by the fact that the classification accuracy during each validation process shows a significantly higher statistical difference than the classification accuracy obtained by applying the simple logistic regression method.

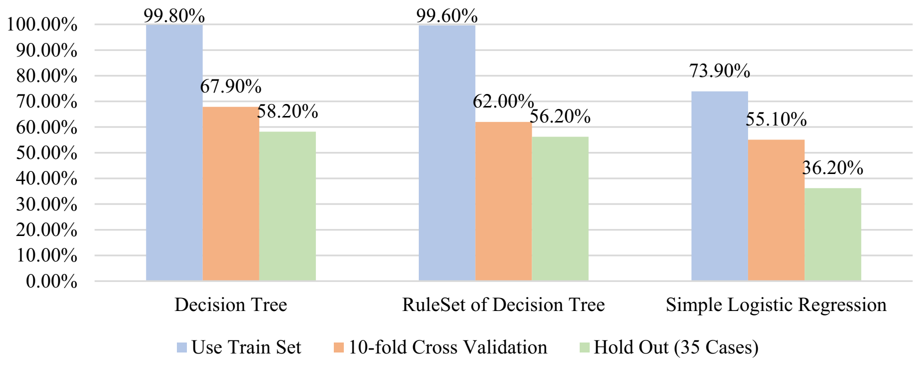

Fig. 14 Framework 03—classification accuracy of the best classification models (Balancing with SMOTE)

Table 8 Framework 03—classification accuracy of the best classification models (Balancing with SMOTE)

Validation method	Use training set (490 Cases) (%)	10-fold cross validation (%)	Hold out (35 Cases) (%)
Decision Tree	99.80	67.90	58.20
RuleSet of Decision Tree	99.60	62.00	56.20
Simple Logistic Regression	73.90	55.10	36.20

5.3 Results of the third framework

The third methodological framework, introduced to investigate the potential emergence of biased classification models resulting from applying the second framework (Framework 02), presents a novel balancing-data approach. As outlined in Sect. 4, this distinctive approach seamlessly integrates data balancing, classification, and the 10-fold Cross Validation method.

Within this third framework, the process of data balancing is notably constrained to the training set formed during each random partition of the original dataset, which takes place as an integral component of the 10-fold Cross Validation procedure. This process is rigorously repeated ten times, aligning with the tenfold partitioning, to comprehensively assess the impact of data balancing on model performance and potential biases across various iterations.

The outcomes of these validations are summarized in Table 7, while Fig. 13 visually represents the performance metrics of the top-performing models for each specific methodological case (Fig. 14) (Table 8).

As observed in Table 9 and the diagram in Fig. 15, using the LCC balancing method contributes equally significantly to improving the performance of the final classification models.

Finally, as in the case of using the SMOTE balancing method, so in the case of the LLC balancing method, the classification method that yields the highest classification accuracy is the inductive decision tree method.

Table 9 Framework 03—classification accuracy of the best classification models (Balancing with LCC)

Validation method	Use training set (490 Cases) (%)	10-fold cross validation (%)	Hold out (35 Cases) (%)
Decision Tree	99.80	89.10	60.70
RuleSet of Decision Tree	99.50	88.10	59.70
Simple Logistic Regression	70.00	58.40	38.20

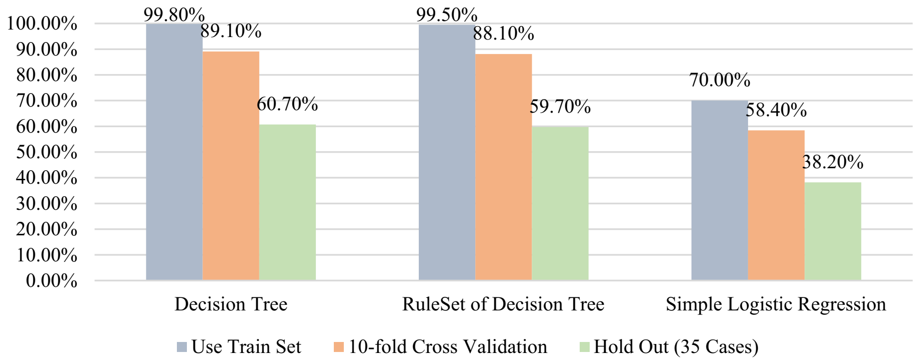

Fig. 15 Framework 03—classification accuracy of the best classification models (Balancing with LCC)

5.4 Comparisons and conclusions

In the context of the comparison of the Third Methodological Framework—which uses parallel data balancing and training within the 10-fold Cross Validation process—with the Second Methodological Framework—which applies data balancing before the training process within the same process—it is observed that both case of using the SMOTE balancing method, as well as in the case of using the LCC balancing method, the performance of the classification models in the Third Methodological Framework shows only limited reductions compared to those created in the Second Methodological Framework.

Then, comparing the performance of the best classification model produced by the combination of the SMOTE balancing method with the Inductive Decision Tree method with the best model resulting from the combination of the LCC balancing method with the Inductive Decision Tree method, it is observed that the model, which uses the LCC balancing method, performs almost 20% better than the model, which uses the SMOTE balancing method. This suggests that the LCC method mitigates the statistical noise of the data. In contrast, the SMOTE method, which increases the statistical noise of the data, adds new technically produced patterns.

Finally, observing the performance of the best classification model of the global approach to the 3-class classification problem (i.e., the model resulting from the combination of the LCC data balancing method with the Inductive Decision Tree method) in the context of its evaluation in new (unknown for the model) data (35 patterns), it is observed that the performances remain relatively stable, both when applying the balancing process before the training process (Second Methodological Framework), and when applying the data balancing process alongside the training process (Third Methodological Framework).

In conclusion, these findings suggest that the second source of statistical noise, arising from the inherent characteristics of the classification problem and the data itself, exerts a major influence on the model's performance. Nevertheless, we underscore the potential for alleviating this impact by refining the suggested data balancing method (LCC method) or exploring novel alternative data balancing techniques, which merit further exploration and investigation.

5.5 Knowledge mining and metadata analysis

Upon a comprehensive evaluation of the models resulting from applying the methodological frameworks expounded in Sect. 4, it becomes evident that the simple logistic regression method needs to be better suited for addressing the classification challenge outlined in this study. This method consistently yields relatively low classification accuracy, falling below the 60% threshold, across all validation procedures, including "Use Train Set," "10-fold Cross Validation," and "Hold Out." Furthermore, it's worth noting that simple logistic regression exhibits diminished accuracy and tends to generate intricate classification models that are challenging to comprehend and employ.

In contrast, the inductive decision tree method excels with significantly higher classification accuracy, nearly touching the 90% mark when juxtaposed with the simple logistic regression method. It offers the distinct advantage of yielding easily interpretable classification rules in the form of "IF/THEN" statements. These rules encapsulate valuable knowledge that can be effectively harnessed in disaster management, particularly in evacuation during natural calamities like snowstorms.

Below, we present the robust classification rules that have emerged from this study's analysis of the inductive decision tree models. These rules are characterized by their high confidence level and represent valuable insights for decision-making and planning.

5.5.1 Classification Rule 1

If a person is over 58 years of age, lives in a private residence, has no children in the household, is fully employed through teleworking or through a combination of teleworking and living, their primary means of transportation is a car while driving follows GPS directions and has experienced a natural disaster (fire, flood or earthquake) in the past, then 96.20% will follow the recommendation to limit trip given by the emergency line in the event of a snow storm.

The above classification rule covers 18 cases from the data set used to develop and evaluate the presented classification model and shows confidence equal to 96.20%.

This classification rule is strong as it covers a more considerable number of cases than the average number of cases covered by the ruleset as a whole and shows greater confidence than the moderate confidence of the ruleset.

According to the specific classification rule, a person with an independent friend over the age of 58 who has experienced a natural disaster in the past and who is allowed to work alternately (telecommuting) in a snowstorm will follow the travel restriction recommendations. That is, both the experience and the way of working play a decisive factor in limiting a person's trips during a snowstorm.

5.5.2 Classification Rule 2

If a person of any gender and over the age of 58 with a master education who has experienced a natural disaster in the past, who has lived for 24–39 years in the area where his residence is located, works both with the telework method and with the combination of telecommuting and living, when driving uses GPS and faithfully follows its instructions, then 92.30% will follow the recommendation to limit travel in a possible snow storm.

A robust classification rule (Covering 11 cases with 92.3% confidence) verifies Classification Rule 1, providing the information that work mode is a determining factor in limiting trip in a snowstorm as corresponding determining factors are the experience of a natural disaster and the ability to follow GPS directions.

5.5.3 Classification Rule 3

A college-educated woman working full-time in the form of a living, who has experienced a natural disaster in the past and rarely or hardly ever follows GPS directions, then 92.90% will partially ignore the recommendation trip restriction and make the necessary trips.

A strong classification rule (covering 18 cases with 92.90% confidence), in combination with Rules 1 and 2, confirms the fact that the way of working as well as the ability to follow GPS instructions are decisive factors that directly affect the travel decision of a citizen to a snowstorm.

5.5.4 Classification Rule 4

A head-educated man who is full-time employed and whose primary mode of transportation is a car, then 89.50% of the time, will completely ignore the travel restriction recommendations given by the emergency line in a potential snowstorm.

This classification rule, although less robust than the above classification rules (covers 17 cases with 89.50% confidence), that a male who works exclusively for life will not follow any travel restriction recommendation. A fact that verifies the knowledge provided by previous classification rules about the influence that work style has on a person's compliance with travel restriction recommendations in a potential snowstorm.

5.5.5 Classification Rule 5

If a person under the age of 36, whose annual family income is between €10,000 and €20,000, works exclusively for life, then, despite not having experienced a natural disaster in the past, 85.00% will faithfully follow the trip restriction instructions to be given by the emergency line in a snowstorm.

This rule, although it is the least strong of the set of strong classification rules presented in this paragraph (it collects 17 cases and shows 85.00% confidence), in combination with the above classification rules, provides the information that younger people (under the age of 36) more easily adapt to the travel restriction recommendations that can be given by the emergency line in a possible snow-storm.

Incorporating insights derived from the Chi-Square statistical test, as detailed in Sect. 3, we can discern a distinct pattern where variables exhibiting substantial

Table 10 Variables with the highest degree of dependence (Chi-Square Test)

Variable	Asymptotic significance (%)
Work Style	0.00
Age Group	0.10
Work Status	0.10
Owned Residence	1.50
Gender	3.50
Household with Seniors	4.30
Transporting Mean	5.40
GPS Usage	8.70

dependence on the target variable of the classification problem, as outlined in Table 8, manifest a corresponding frequency of occurrence within the robust classification rules (Table 10).

More specifically, when we scrutinize the target variable, "Trip Restriction," following Table 8, it becomes evident that it shares a notable degree of dependence with the variables "Work Style," "Age Group," and "Work Status." These variables surface with higher frequencies in the robust sorting rules.

This interrelationship between these variables underscores their significance in shaping the outcome of the classification model. The variables "Work Style," "Age Group," and "Work Status" appear to play a crucial role in the classification process, as they are not only closely associated with the target variable but are also recurrently featured in the strong classification rules, which, in turn, contribute to the model's overall effectiveness.

6 Conclusion

Natural disasters, including earthquakes, floods, fires, and other events, have significant impacts on both the environment and society. They affect health, the economy, and the environment, underscoring the importance of prevention and preparedness for community protection. However, natural disasters also profoundly affect people's mental health, leading to stress and psychological challenges. Psychosocial support and educational programs are crucial in addressing these issues and helping individuals return to normalcy.

Machine learning is emerging as a critical tool in responding to natural disasters and understanding human behavior during evacuations. Machine learning algorithms enable the prediction of disaster evolution, offering early warnings that can save lives and reduce losses. Moreover, applying machine learning to analyze corresponding data aids in understanding human behavior during crises, enhancing preparedness for future disasters. Machine learning is essential in responding to natural disasters and gaining insights into human behavior during emergencies.

This research specifically focused on the challenge of estimating human behavior during a natural disaster, such as a snowstorm. Artificial intelligence techniques were used to develop classification models that categorize people's responses into three groups (no travel, essential travel only, unrestricted travel) based on socio-economic characteristics during a snowstorm that struck Greece in the winter of 2023.

As mentioned, artificial intelligence methods were applied to develop classification models in the present effort. These models seek to categorize the behavior of individuals, considering their socioeconomic characteristics, into three discrete categories: "Did not travel at all," "Traveled only as necessary," or "Did not limit travel." This categorization was based on instructions to restrict movement during the recent "Barbara" snowstorm, which affected Greece in the winter of 2023. The data used in this study came from a parallel survey conducted within the framework of the AEGIS+ research project. This research focused on assessing the mental health of individuals who had experienced natural disasters and collected data through a questionnaire.

Utilizing machine learning to analyze survey data opens up new horizons for comprehending individuals' perspectives and preferences across various domains, encompassing social, political, and healthcare realms. While this methodology offers valuable insights, it grapples with two fundamental challenges. Firstly, safeguarding privacy is paramount since questionnaires often contain sensitive personal information. Secondly, ensuring that the sampled responses are a true reflection of the broader population is crucial, as variations in answers and non-responses can significantly impact the analysis outcomes. This approach provides indispensable insights but necessitates meticulous handling of privacy and data representation issues.

Protecting personal data in this work is upheld by the corresponding privacy protocols, which safeguard the dataset utilized in this research. This dataset comprises 525 Cases and encompasses 16 Variables, offering a robust foundation for the exploration and analysis of survey data.

The aim of this specific work was the development of capable machine learning models that will be able to classify with high accuracy the cases of the initial data set into the classification as mentioned above classes (I did not make any trip; I made only the necessary tips, or I did not limit my trips).

The goal of this work was to generalize the optimal classification model and extract knowledge that can be appropriately used in a natural disaster situation. Specifically, three methodological frameworks for data analysis with machine learning methods are proposed that include combinations of Simple Logistic Regression and Inductive Decision Trees, both using the SMOTE method and a new proposed data balancing method called LCC in the context of the "Use Train Set", "10-fold Cross Validation" and "Hold Out" validation procedures. Out".

The contribution of this article, through the proposed methodological frameworks of data analysis with computational methods, is located both in the area of natural disaster risk management through the mining of new relevant knowledge and in the broader field of data analysis with machine learning methods through the development of hybrid models classification models that include the data balancing process at various stages of modelling and confirm its importance and positive influence on the performance of respective classification models.

The results of the overall approach were deemed satisfactory and encouraging for the continuation of corresponding approaches. Specifically, through statistical evaluation procedures and comparing the performance (classification accuracy) of the models resulting from the simple logistic regression method with the models resulting from the inductive decision trees method, it was observed that the most appropriate method was the inductive decision trees method optimized with AdaBoost algorithm and combined with LCC balancing method. With this method, classification models were developed with satisfactory classification accuracy (almost 100% during the Uset Test validation process and nearly 90% during the 10-fold cross Validation and Hold Out validation processes) that produce strong classification rules, which can be exploited as knowledge in risk management in a natural disaster situation.

During the process of Knowledge Mining and Metadata Analysis, it emerged and crossed with the conclusions of the Chi-Square performed on the original data set (525 Cases | 16 Variables) that the socio-economic characteristics that influence the decision to move during a natural disaster are the age, education as well as a person's work profile.

In summary, this work focuses on predicting people's evacuation behavior during natural disasters. Understanding the factors influencing evacuation decisions can contribute to more effective risk management and enhanced disaster preparedness. More in detail, in this article, we adopt a hybrid machine learning approach, combining respective algorithms to analyze data from various sources, such as (i) Demographic data (Age, gender, socio-economic status, etc.), (ii) Psychological characteristics (Stress level, sense of vulnerability, trust in the authorities, etc.), (iii) Environmental factors (Characteristics of the area, information about the disaster, etc.) as well as (iv) Behavior during previous disasters (Evacuation experiences, participation in preparedness plans, etc.). Data analysis with machine learning methods presented in this paper can reveal patterns and correlations influencing evacuation decisions. This can lead to (i) Predicting the likelihood of an evacuation, (ii) Developing targeted information messages during natural disasters, and (iii) Improving evacuation planning.

Our work offers a promising approach for predicting evacuation behavior and enhancing natural disaster risk management. Adopting research findings can help reduce loss of life and protect communities from the devastating effects of natural disasters.

Acknowledgements The authors would like to extend their sincere gratitude to the participants for their valuable time and cooperation. They would also like to express their appreciation to The American College of Greece for their support and approval of the ethical procedures. The research described in this paper was funded by the Greek national project "Development of the 'Coastal Environmental Observatory and Crisis Management in Island Areas' Infrastructure (AEGIS+)" conducted at the University of the Aegean. Additionally, the second author (KK) would like to acknowledge the financial support provided for the following projects: 'Collaborative, Multi- modal and Agile Professional Cybersecurity Training Program for a Skilled Workforce In the European Digital Single Market and Industries' (CyberSecPro) project, which has received funding from the European Union's Digital Europe Programme (DEP) programme under grant agreement No 101083594. The 'Human-centered Trustworthiness Optimisation in Hybrid Decision Support' (THEMIS 5.0) project, which has received funding from the European Union's Horizon Programme under grant agreement No 101121042. The 'advaNced cybErsecurity awaReness ecOsystem for SMEs' (NERO) project, which has received funding from the European Union's DEP programme under grant agreement No 101127411. And 'Fostering Artificial Intelligence Trust for Humans towards the optimization of trustworthiness through large-scale pilots in critical domains' (FAITH) project, which has received funding from the European Union's Horizon Programme under grant agreement No 101135932. It is important to note that the views expressed in this paper solely represent the opinions of the authors and not those of the University of the Aegean, the American College of Greece, trustilio B.V., the European Commission, or the partners involved in the mentioned projects. Finally, the authors declare that there are no conflicts of interest, including any financial or personal relationships, that could be perceived as potential conflicts.

Author contributions The authors confirm contribution to the paper as follows: study conception and design: AP, AT, KK, EK; data collection, recruitment process and ethics approval: AP, AT, KK; data analysis and production of the first draft of the manuscript: EK, GD. Discussion and conclusions: AP, EK, KK, AT, GD. All authors reviewed the results and approved the final version of the manuscript.

Funding The authors have not disclosed any funding.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Ethical approval The study received ethical clearance from the Institutional Review Board of the American College of Greece (reference number: 202212333). As the survey was carried out anonymously, direct assistance was not feasible for participants who obtained high scores on the psychometric tests and displayed potential indications of clinical problems. To tackle this concern, the information sheet included contact details for accessible support services.

References

Agrawal A, Viktor HL, Paquet E (2015) SCUT: multi-class imbalanced data classification using SMOTE and cluster-based undersampling. In: 2015 7th international joint conference on knowledge discovery, knowledge engineering and knowledge management (IC3K), pp 226–234. Accessed 19 Feb 2024. Available: https://ieeexplore.ieee.org/abstract/document/7526924?casa_token=RuN4fIOHSEEAAAAA:FfW0KJeqkV4uB7VX69pyV2RPSkv72vOQtyebN2fsQOkaDLhHdGb4W1it1Lkj7EYFDhW211Ls

Ayyadevara VK (2018) Gradient boosting machine. In: Ayyadevara VK (ed) Pro machine learning algorithms: a hands-on approach to implementing algorithms in Python and R. Apress, Berkeley, pp 117–134

Babić IĐ (2017) Machine Learning methods in predicting the student academic motivation. Croat Oper Res Rev 8:443–461

Bartholomew DJ (1980) Factor analysis for categorical data. J Roy Stat Soc Ser B Methodol 42(3):293–312. https://doi.org/10.1111/j.2517-6161.1980.tb01128.x

Batista GEAPA, Prati RC, Monard MC (2004) A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl 6(1):20–29. https://doi.org/10.1145/1007730.1007735

Brown M, Williams R (2024) human behavior factors in emergencies: implications for emergency response and communication strategies. J Disaster Stud 21(1):12–28

Buskirk TD, Kirchner A (2020) Why machines matter for survey and social science researchers. Big data meets survey science. John Wiley & Sons Ltd, Hoboken, pp 9–62

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/jair.953

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, in KDD '16. Association for Computing Machinery, New York, pp 785–794

Chien HY, Kwok OM, Yeh YC, Sweany NW, Baek E, McIntosh W (2020) Identifying at-risk online learners by psychological variables using machine learning techniques. Online Learn 24:131–146

Dabhade P, Agarwal R, Alameen KP, Fathima AT, Sridharan R, Gopakumar G (2021) Educational data mining for predicting students' academic performance using machine learning algorithms. Mater Today Proc 47:5260–5267. https://doi.org/10.1016/j.matpr.2021.05.646

Dahan H, Cohen S, Rokach L, Maimon O (2014) Proactive data mining with decision trees. SpringerBriefs in Electrical and Computer Engineering. Springer-Verlag, New York

De La Garza ÁG, Blanco C, Olfson M, Wall MM (2021) Identification of suicide attempt risk factors in a national US survey using machine learning. JAMA Psychiat 78:398–406. https://doi.org/10.1001/jamapsychiatry.2020.4165

Deb K (1999) An introduction to genetic algorithms. Sadhana 24(4):293–315. https://doi.org/10.1007/BF02823145

Deng Z, Ji M (2012) Deriving rules for trip purpose identification from GPS travel survey data and land use data: a machine learning approach, pp 768–777

Dulebenets MA, Abioye OF, Ozguven EE, Moses R, Boot WR, Sando T (2019) Development of statistical models for improving efficiency of emergency evacuation in areas with vulnerable population. Reliab Eng Syst Saf 182:233–249. https://doi.org/10.1016/j.ress.2018.09.021

Dutton DM, Conroy GV (1997) A review of machine learning. Knowl Eng Rev 12(4):341–367. https://doi.org/10.1017/S026988899700101X

Elreedy D, Atiya AF (2019) A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Inf Sci 505:32–64. https://doi.org/10.1016/j.ins.2019.07.070

European Disaster Risk Management (2023) Accessed 29 Jul 2023. Available: https://civil-protection-humanitarian-aid.ec.europa.eu/what/civil-protection/european-disaster-risk-management_en

Fernandez A, Garcia S, Herrera F, Chawla NV (2018) SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Intell Res 61:863–905. https://doi.org/10.1613/jair.1.11192
Flener P, Schmid U (2008) An introduction to inductive programming. Artif Intell Rev 29(1):45–62. https://doi.org/10.1007/s10462-009-9108-7
Franke TM, Ho T, Christie CA (2012) The chi-square test: often used and more often misinterpreted. Am J Eval 33(3):448–458. https://doi.org/10.1177/1098214011426594
Freund Y, Schapire R, Abe N (1999) A short introduction to boosting. J Jpn Soc Artif Intell 14(771–780):1612
Gahegan M (2003) Is inductive machine learning just another wild goose (or might it lay the golden egg)? Int J Geogr Inf Sci 17(1):69–92. https://doi.org/10.1080/713811742
Gu B, Sheng VS, Tay KY, Romano W, Li S (2015) Incremental support vector learning for ordinal regression. IEEE Trans Neural Netw Learn Syst 26(7):1403–1416. https://doi.org/10.1109/TNNLS.2014.2342533
Gupta S, Gupta A (2019) Dealing with noise problem in machine learning datasets: a systematic review. Procedia Comput Sci 161:466–474. https://doi.org/10.1016/j.procs.2019.11.146
Jain S, Narayan SP, Dewang RK, Bhartiya U, Meena N, Kumar V (2019) A machine learning based depression analysis and suicidal ideation detection system using questionnaires and twitter. In: 2019 IEEE students conference on engineering and systems (SCES), Feb. 2019, pp 1–6
James G, Witten D, Hastie T, Tibshirani R, Taylor J (2023) An introduction to statistical learning: With applications in python. Springer Nature [Online]. Available: https://books.google.com/books?hl=el&lr=&id=ygzJEAAAQBAJ&oi=fnd&pg=PR7&dq=An+Introduction+to+Statistical+Learning,+second+edition,+springer&ots=WzlbjSH_Tj&sig=s63qpVbiofehpQ4nSzsblcGSmds. Accessed 26 June 2024
Johnson R, Lee S (2023) Predictive models for identifying evacuation bottlenecks using machine learning. Disaster Prev Manag 18(1):75–90
Johnson A, Smith J, Lee J (2022) The importance of understanding evacuation behavior: enhancing disaster preparedness and response strategies. J Emerg Manag 5:78–94
Joo J et al (2013) Agent-based simulation of affordance-based human behaviors in emergency evacuation. Simul Model Pract Theory 32:99–115. https://doi.org/10.1016/j.simpat.2012.12.007
Kim S, Lee H-K, Lee K (2021) Which PHQ-9 items can effectively screen for Suicide? Machine learning approaches. Int J Environ Res Public Health 18(7):3339. https://doi.org/10.3390/ijerph18073339
Kramer O (2013) K-Nearest Neighbors. In: Kramer O (ed) Dimensionality reduction with unsupervised nearest neighbors. Springer, Berlin, pp 13–23
Landwehr N, Hall M, Frank E (2005) Logistic model trees. Mach Learn 59(1):161–205. https://doi.org/10.1007/s10994-005-0466-3
Li P (2012) Robust LogitBoost and adaptive base class (ABC) LogitBoost. arXiv https://doi.org/10.48550/arXiv.1203.3491
Lin M-R, Huang W, Huang C, Hwang H-F, Tsai L-W, Chiu Y-N (2002) The impact of the chi-chi earthquake on quality of life among elderly survivors in Taiwan—a before and after study. Qual Life Res 11(4):379–388. https://doi.org/10.1023/A:1015543113448
Liu Y, Wang Y, Zhang J (2012) New Machine learning algorithm: random forest. Information computing and applications. Lecture Notes in Computer Science. Springer, Berlin, pp 246–252
Liu H, Chen H, Hong R, Liu H, You W (2020) Mapping knowledge structure and research trends of emergency evacuation studies. Saf Sci 121:348–361. https://doi.org/10.1016/j.ssci.2019.09.020
Lopez de Mantaras R, Armengol E (1998) Machine learning from examples: inductive and lazy method. Data Knowl Eng 25(1):99–123. https://doi.org/10.1016/S0169-023X(97)00053-0
Madrid PA, Grant R (2008) Meeting mental health needs following a natural disaster: lessons from Hurricane Katrina. Prof Psychol Res Pract 39(1):86–92. https://doi.org/10.1037/0735-7028.39.1.86
Maimon O, Rokach L (2010) Introduction to knowledge discovery and data mining. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook. Springer, Boston, pp 1–15
Mental Health Assistance to the Populations Affected by the Tsunami in Asia - Indonesia | ReliefWeb (2023) Accessed 29 Jul 2023. Available: https://reliefweb.int/report/indonesia/mental-health-assistance-populations-affected-tsunami-asia
Michalski RS (1983) A theory and methodology of inductive learning. In: Michalski RS, Carbonell JG, Mitchell TM (eds) Machine learning. Morgan Kaufmann, San Francisco, pp 83–134
Mitchell MJ, Witman M, Taffaro C (2008) Reestablishing Mental Health Services in St. Bernard Parish, Louisiana, following Hurricane Katrina. Prof Psychol Res Pract 39(1):66–76. https://doi.org/10.1037/0735-7028.39.1.66

Mustapha A, Mohamed L, Ali K (2020) An overview of gradient descent algorithm optimization in machine learning: application in the ophthalmology field. In: Hamlich M, Bellatreche L, Mondal A, Ordonez C (eds) Smart applications and data analysis. Communications in Computer and Information Science. Springer International Publishing, Cham, pp 349–359

National Meteorological Service of Greece (2024). Accessed 18 Feb 2024. Available: http://www.emy.gr/emy/el/meteorology/meteorological_news?name=2302031536

Norris FH, Phifer JF, Kaniasty K (1994) Individual and community reactions to the Kentucky floods: findings from a longitudinal study of older adults. Individual and community responses to trauma and disaster: the structure of human chaos. Cambridge University Press, New York, pp 378–400

Ontivero-Ortega M, Lage-Castellanos A, Valente G, Goebel R, Valdes-Sosa M (2017) Fast Gaussian Naïve Bayes for searchlight classification analysis. Neuroimage 163:471–479. https://doi.org/10.1016/j.neuroimage.2017.09.001

Papanikolaou V, Adamis D, Kyriopoulos J (2012) Long term quality of life after a wildfire disaster in a rural part of Greece. Open J Psychiatry 02(02):164–170. https://doi.org/10.4236/ojpsych.2012.22022

Peng C-YJ, Lee KL, Ingersoll GM (2002) An introduction to logistic regression analysis and reporting. J Educ Res 96(1):3–14. https://doi.org/10.1080/00220670209598786

Perner P (2015) Decision tree induction methods and their application to big data. In: Xhafa F, Barolli L, Barolli A, Papajorgji P (eds) Modeling and processing for next-generation big-data technologies: with applications and case studies, Modeling and optimization in science and technologies. Springer International Publishing, Cham, pp 57–88

Peterson MJ (2023) The UN general assembly. In: International organization and global governance, 3rd edn. Routledge. ISBN: 9781003266365

Podgorelec V, Kokol P, Stiglic B, Rozman I (2002) Decision trees: an overview and their use in medicine. J Med Syst 26(5):445–463. https://doi.org/10.1023/A:1016409317640

Polyzotis N, Zinkevich M, Roy S, Breck E, Whang S (2019) Data validation for machine learning. Proc Mach Learn Syst 1:334–347

Quinlan JR (1987) Generating production rules from decision trees. In: IJCAI, vol 87, pp 304–307

Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kaufmann, Burlington

Reyes G, Elhai JD (2004) Psychosocial interventions in the early phases of disasters. Psychother Theory Res Pract Train 41(4):399–411. https://doi.org/10.1037/0033-3204.41.4.399

Robertson A, Sillén R, Norén JG (1998) Inductive analysis methods applied on questionnaires. Acta Odontol Scand 56:268–275. https://doi.org/10.1080/000163598428437

Sánchez-Maroño N, Alonso-Betanzos A, Fontenla-Romero O, Gary Polhill J, Craig T (2017) Empirically-derived behavioral rules in agent-based models using decision trees learned from questionnaire data. In: Alonso-Betanzos A, Sánchez-Maroño N, Oscar Fontenla-Romero J, Polhill G, Craig T, Bajo J, Corchado JM (eds) Agent-based modeling of sustainable behaviors. Springer International Publishing, Cham, pp 53–76

Smith J, Lee S (2023) Predictive models for optimizing evacuation routes: insights from past evacuation incidents. Disaster Prev Manag 18(3):145–160

Smith J, Johnson A, Brown M (2022) Application of deep learning algorithms for real-time analysis of evacuation behavior. J Emerg Manag 15(3):125–140

Sperandei S (2014) Understanding logistic regression analysis. Biochem Med 24(1):12–18. https://doi.org/10.11613/BM.2014.003

Sun CK, Tang YX, Liu TC, Lu CJ (2022) An integrated machine learning scheme for predicting mammographic anomalies in high-risk individuals using questionnaire-based predictors. Int J Environ Res Public Health 19:9756. https://doi.org/10.3390/ijerph19159756

Terano T, Ishino Y (1995) Marketing data analysis using inductive learning and genetic algorithms with interactive and automated phases. In Proceedings of 1995 IEEE international conference on evolutionary computation, vol 2, pp 771–776

Wang W et al (2022) Predicting the 10-year risk of cataract surgery using machine learning techniques on questionnaire data: findings from the 45 and up study. Br J Ophthalmol 106(11):1503–1507. https://doi.org/10.1136/bjophthalmol-2020-318609

Warsini S, West C, Ed GD, Res Meth GC, Mills J, Usher K (2014) The psychosocial impact of natural disasters among adult survivors: an integrative review. Issues Mental Health Nurs 35:420–436. https://doi.org/10.3109/01612840.2013.875085

Weilnhammer V et al (2021) Extreme weather events in europe and their health consequences—a systematic review. Int J Hyg Environ Health 233:113688. https://doi.org/10.1016/j.ijheh.2021.113688

Wong SD (2023) Compliance, congestion, and social equity: tackling critical evacuation challenges through the sharing economy, joint choice modeling, and regret minimization. Accessed 29 Jul 2023. Available: https://escholarship.org/uc/item/9b51w7h6

Xanthopoulos P, Pardalos PM, Trafalis TB (2013) Linear discriminant analysis. In: Xanthopoulos P, Pardalos PM, Trafalis TB (eds) Robust data mining. Springer, New York, pp 27–33

Yanatma S (2023) How much has climate change cost the EU in the last 40 years? Euronews. Accessed 23 Oct 2023. Available: https://www.euronews.com/green/2022/11/13/climate-change-has-cost-the-eu-almost-almost-500-billion-in-the-last-40-years

Yzermans CJ, Donker GA, Kerssens JJ, Dirkzwager AJ, Soeteman RJ, ten Veen PM (2005) Health problems of victims before and after disaster: a longitudinal study in general practice. Int J Epidemiol 34(4):820–826. https://doi.org/10.1093/ije/dyi096

Zhao B, Wong SD (2021) Developing transportation response strategies for wildfire evacuations via an empirically supported traffic simulation of Berkeley, California. Transp Res Rec 2675(12):557–582. https://doi.org/10.1177/03611981211030271

Authors and Affiliations

Evangelos Karampotsis[1] · Kitty Kioskli[2,3] · Athina Tsirimpa[2,4] · Georgios Dounias[1] · Amalia Polydoropoulou[2]

✉ Evangelos Karampotsis
ekarampotsis@aegean.gr

Kitty Kioskli
kitty.kioskli@trustilio.com

Athina Tsirimpa
atsirimpa@acg.edu

Georgios Dounias
g.dounias@aegean.gr

Amalia Polydoropoulou
polydor@aegean.gr

1 Department of Financial and Management Engineering, University of the Aegean, Chios, Greece

2 Department of Shipping, Trade and Transport, University of the Aegean, Chios, Greece

3 Trustilio B.V., Amsterdam, The Netherlands

4 Department of Maritime Transport and Logistics, The American College of Greece, Athens, Greece