ISBN: 978-628-96613-1-6. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2025.1.1.630

Data Analysis of Missing People in Ecuador

Pamela Quinga[1]; Esteban Jara[1]; Francisco Cela[1]; Christopher Coba[1]; Andrea Pilco[1]; Viviana Moya[1]
[1]Facultad de Ciencias Técnicas, Universidad Internacional Del Ecuador UIDE, Quito 170411, Ecuador,
alquingagu@uide.edu.ec, esjarapa@uide.edu.ec, frcelabu@uide.edu.ec, chcobaqu@uide.edu.ec, anpilcoat@uide.edu.ec, vimoyago@uide.edu.ec

Abstract*–. The phenomenon of missing people has negative consequences for the individuals affected, their families, and society at large. This issue has become increasingly prominent in recent years, partly due to the influence of social media, which has emerged as a vital tool in the search for missing persons. Given the growing importance of addressing this challenge, this research contributes to the expanding field of machine learning applications for social issues. Supervised machine learning models were applied to open data on missing persons in Ecuador between 2021 and 2024 to predict the status of individuals as either "Found" or "Deceased." Using personal, social, and event-specific variables, two models Random Forest (RF) and Support Vector Machine (SVM) were implemented and evaluated. The models were assessed using key performance metrics, including accuracy, precision, recall, F1 score, and confusion matrices, to determine their effectiveness. The analysis revealed that the RF model achieved superior performance on the test data compared to SVM, with an accuracy of 89%, demonstrating its suitability for the dataset. These findings provide valuable insights into the factors influencing the outcomes of disappearance cases, allowing decision-makers to optimize resource allocation, improve search strategies, and support evidence-based decision-making. Predicting the status of a missing person offers an innovative approach to addressing this critical social issue.

***Keywords*—Missing persons, Machine learning, Classification, Predictive model, Supervised learning.**

I. Introduction

Missing people is relevant as a critical global problem that concerns individuals, families and societies, as it has significant emotional, social and organizational consequences. Even with the efforts that are being made by the policing authorities and different organizations, the nature and variety of incidents are still complex, taking into consideration the time consumed in searching for lost persons [1]. In missing person cases, Artificial Intelligence (AI) and Machine Learning (ML) have developed into innovative resources that comprise sophisticated models for analyzing big data, recognizing trends and enhancing the efficiency of decision-making tasks.

They allow for finding the optimum set of data sources that can include face recognition, sentiment analysis, and models for improving search quality and managing resources efficiently. On this account, the paper under development aims at deploying the application of the AI and ML frameworks particular to Ecuador's data needs and issues that are not sufficiently met by existing models, and which can be enhanced through the use of these technologies. The status of similar research in other countries has been reviewed to assess global developments and design a sound local solution.

Vulnerable persons' disappearance has positively influenced research development, especially where multiple technological advancements are adopted. They added basic knowledge by majoring in users' behaviour and decisions when searching for information. In [2] critically examined decision-making processes in SAR using CDM or critically examined the CDM about SAR. Their experiences also identified the mental processes and critical decision points rescue teams face, from which subsequent decision-support technologies for enhancing search effectiveness can be designed.

From these considerations, the study of [3] proposed an agent-based model for the behaviour of missing persons in wilderness areas. Including factors within the perception framework and overall goals of the developed model established an improved accurate model compared to the generalized diffusion models, thus reducing the search time within environments by up to 40 per cent. This work emphasized including environmental and behavioural processes in the frameworks for prediction. Another step forward in predictive modelling was made with the use of Bayesian Networks, according to [4]. Their work involved a Misper-Bayes system, which estimated the likely areas where a missing person or persons could be found, especially those groups of people that are usually vulnerable, such as children or those with dementia. This probabilistic model outperformed traditional manual methods used by UK police forces, offering a robust decision-support framework for search operations.

Furthermore,[5] then developed predictive systems and introduced the MP-Net framework that used natural language processing and the deep learning method to analyse the NGO data for missing person predictions in China. Thus, using oral information, the system had an 87.18% recall rate and proved the further potential of qualitative data in predictive models. Concerning the older adult population, [6] used gradient-boosting algorithms to examine missing person information in Colombia. Its main milestones are also clear: Choosing features is given paramount importance by the authors proposed, and the features offered include time, age, and size of the municipality, and, according to the authors, the AUC of their model reached 0.79. This research has also provided insight into how the formulation of such models may need to be adapted for unique or at-risk demographics and geographic locations.

Another type of source has also been acknowledged, namely social media. The research of [7] proposed using an r-instance learning model to recognise missing persons' tweets. As a result, although the shared posts came in sparse and almost unstructured, the model utilising the word embeddings and homophily from social networks was able to detect proper Tweets successfully. Facial recognition technologies have gone even further to bring radical changes in the field as seen in [8] which developed the Suhail system as an Android application incorporating TensorFlow to search for face similarities in public images with the missing persons database. The system improved identification speed and accuracy, emphasising the role of mobile applications in decentralised search processes.

Going further into more detail, [9] put forward a framework for combining facial recognition with video surveillance. By using convolutional neural networks (CNNs), the system rapidly identifies individuals in high-traffic areas like airports, improving real-time search capabilities while addressing privacy concerns. Some of them are the ReUnite AI system by [10] for facial recognition. This system incorporates PCA and GANs to improve face and age progression detection functions. Its ability to predict age-related facial changes over time significantly improved the identification of long-missing individuals. Meanwhile, the authors of [11] discussed the existing facial recognition methods and created the combined Haar Cascade-CNN model. Their approach has a 90% accuracy, which was further improved from traditional machine learning algorithms, including the KNN and the SVM. This work focused on enhancing the scalability of a model fit for practical application.

Computational methods applied in Colombia concern pattern recognition of missing individuals. For instance, [12] used data mining to determine the socio-demographic factors of disappearances and develop decision rules on how age, gender, and location led to disappearances. These findings provided recommendations for local police in Colombia about what to do. Tracking systems have also come out as Key Performance Indicators with vast coverage. The Missing Persons Comprehensive Tracking System: MPCTS was designed by [13] using a multi-area CCTV police station mapping interface with secure login. This platform improves data communication and ramps up response time should there be a need to search for missing persons. The work of [14] synthesised the literature about the procedural and ethical considerations of police in missing person cases. Their study underscored the importance of balancing procedural rigour with compassion, setting the stage for integrating advanced technologies into traditional investigative workflows. Last, [15] investigated demographic and behavioural characteristics associated with spatial types of missing individuals. Specifically, based on the observational studies of archival cases, they defined the range of choice influencers, including planning behaviours and vulnerability. These include improvements to the search techniques based on the findings and dynamics of human actions regarding the geography of the missing persons' cases.

Various studies have highlighted the importance of technological integration as a key tool for improving citizens' quality of life [16]. One particularly relevant field is the multidisciplinary approach to the search for missing persons in today's world, where AI-based solutions play a fundamental role in addressing these challenges more efficiently and effectively.

This research highlights the application of supervised machine learning models to predict the future status of missing persons in Ecuador. The specific contributions of this study include:

1) Utilization of Ecuadorian Data: Leveraging data provided by the "Ministerio del Interior" of the Government of Ecuador, this study focuses on a real-world dataset relevant to the country's context.
2) Implementation of Machine Learning Models: This research applies two machine learning models, Random Forest (RF) and Support Vector Machine (SVM), to analyse and accurately predict the status of missing persons.

This work is structured as follows: Section II describes the methodology, Section III presents the results and evaluation metrics, Section IV discusses the findings, and Section V summarizes the conclusions and implications.

II. Methodology

Two machine learning models were developed to determine the status of missing persons, categorizing them as either 'deceased' or 'found.' The methodology employed for this purpose is illustrated in Fig. 1 and described in detail below.

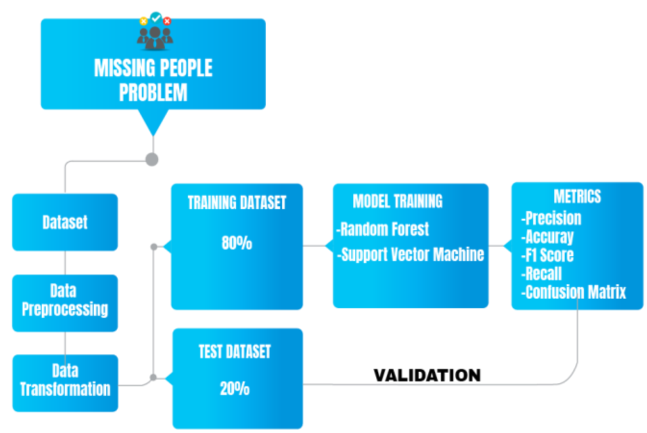

Fig. 1 Flowchart of data analysis of missing people in Ecuador.

A. *Dataset*

The study utilized open data provided by the "Ministerio del Interior" of the Government of Ecuador [17]. The data were downloaded on October 21, 2024. The original database consisted of 4,570 records spanning the year of 2024, with information presented in Spanish. Each record corresponds to a reported case of a missing person within Ecuador's territory, categorized across its 24 provinces. The database includes the following variables:

1) Provincia (in English means "Province"): Represents the geographic location where the disappearance occurred.
2) Latitud (in English means "Latitude"): Latitude coordinates of the province where the case occurred.

3) Longitud (in English means "Longitude"): Longitude coordinates of the province where the case occurred.
4) Edad_aproximada (in English means "Approximate Age"): The estimated age of the missing person.
5) Sexo (in English means "Gender"): Indicates the gender of the missing person.
6) Motivo_desaparición (in English means "Reason for Disappearance"): The initial classification of the reported reason for the disappearance.
7) Motivo_desaparición_observada (in English means "Observed Reason for Disappearance"): A refined classification of the reported reason.
8) Fecha_desaparición (in English means "Disappearance Date"): The specific date when the incident was reported.
9) Situación_actual (in English means "Current Situation"): Refers to the status of the case, indicating whether the person has been located and, if so, whether they were declared deceased.
10) Fecha_localización (in English means "Date of Location"): Information collected with specific dates when the person was located.

The dependent output variable in this study is the Current Status, which includes two possible outcomes: Deceased or Found. The remaining nine variables have been categorized into specific groups, as detailed in Table I. This study does not focus on the geographical location of the reported cases. Instead, the analysis prioritizes personal, social, and temporal characteristics that are relevant for predicting the status.

TABLE I
CLASSIFICATION OF VARIABLES BY CATEGORY

Category	Variable
Personal	Approximate Age Gender
Social	Reason for Disappearance Observed Reason for Disappearance
Date of the event	Disappearance Date Date of Location
Location of the event	Province Latitude Longitude

B. *Data preprocessing*

From the original dataset, the variables related to the location of the event were removed, resulting in a refined dataset with six variables or attributes for analysis. During data pre-processing, a thorough data cleaning process was performed to address missing values and inconsistencies [18], as detailed below:

1) Missing Values: Missing values were identified in the dataset, which occur when no value is stored in certain cells. This issue was particularly prominent in the columns 'Reason for Disappearance', 'Observed Reason for Disappearance', and 'Date of Location'. For the categorical variables 'Reason for Disappearance' and 'Observed Reason for Disappearance', the missing values were imputed using the most frequent class in each column: 'Personal causes' for 'Reason for Disappearance' and ' Circumstantial ' for 'Observed Reason for Disappearance'. Missing values were replaced with the most frequently occurring category, determined by analyzing the most common month in the dataset, which was February.
2) Inconsistencies: An inconsistency was found in the 'Approximate Age' variable, where an unrealistic value of 160 years was recorded. Since this is a numerical variable, the erroneous value was corrected by replacing it with the average of all valid entries in the column.

TABLE II
VARIABLES DESCRIPTION

Variable	Type	Values
Approximate Age	Numerical	Childhood = 0 Adolescence = 1 Adulthood = 2 Senior = 3
Gender	Numerical	Female = 1 Male = 2
Reason for Disappearance	Categorical	Personal causes, violence, deceased, social causes, family causes, missing-temporary absence, missing-disability/illness/disorders, academic causes, economic causes, discrimination, possible link to crime, closed by the prosecutor's office/crime reformulated, family problems, missing, social problems.
Observed Reason for Disappearance	Categorical	Circumstantial, physical violence, murder, psychological violence, alcoholism, robbery, drug addiction, not related to violence, not determined, accident, intellectual disability, sensory disability, homicide murder, psychological disability, migration, academic/school, hospitalized, accidental, organized criminal groups, affective relationship, accidental due to carelessness, homicide, natural death, physical disability, extortion, sexual orientation and gender identity, detention, suicide, extortive kidnapping, food trial, femicide, human trafficking, rape, sexual harassment, sexual abuse, closed by the prosecutors, suicidal behavior disorder, illicit trafficking of migrant, digital sexual violence, extortive kidnapping, hit sucker, harassment, dysfunctional family, relational problems, missing person, influence of friends.
Disappearance Date	Categorical	January to September
Date of Location	Categorical	January to September

C. Data transformation

The dataset analyzed contains categorical variables that were transformed into numerical variables to facilitate analysis. For instance, the gender variable was encoded as 1 for female and 2 for male. Similarly, the age variable, which originally had multiple distinct values, was grouped into the following categories: Childhood (0–12 years), Adolescence (13–17 years), Adulthood (18–59 years), and Senior (60+ years). These categories were numerically encoded as follows: Childhood = 0, Adolescence = 1, Adulthood = 2, and Senior = 3. Additionally, the variables Reason for Disappearance and Observed Reason for Disappearance retained their categorical values, while the columns Disappearance Date and Date of Location preserved their date format. A description of each variable or attribute is provided in Table II.

D. Machine learning modeling

This research focuses on the application of machine learning to address a social problem. Two supervised machine learning algorithms were employed. The dataset was divided into two subsets: a training set, used to fit the model to the data, and a test set, used to evaluate the model's ability to generalize to unseen data. Specifically, 80% of the dataset, comprising 3,656 data points, was allocated for training, while the remaining 20%, consisting of 914 data points, was reserved for testing. In this study, two predictive models were implemented: Random Forest (RF) and Support Vector Machine (SVM).

1) Random Forest is a widely used supervised machine learning algorithm that combines the outputs of multiple decision trees to produce a single result. Decision trees aim to identify the best splits for data subsets and are typically trained using the classification tree algorithm [19].
2) Support Vector Machine is a supervised machine learning algorithm that classifies data by identifying an optimal line or hyperplane that maximizes the separation between classes in an N-dimensional space. SVMs are widely employed for classification tasks and work by distinguishing between two classes through the determination of the optimal hyperplane, which maximizes the margin between the nearest data points of opposing classes [20].

E. Training

A testing environment was established after defining the datasets, and the training process was conducted using Google Colab. A detailed description of the training environment is provided in Table III.

TABLE III
TRAINING ENVIRONMENT

Environment	Component	Capacity
Google Colab	RAM	12.7 GB
	GPU	T4 GPU 15 GB VRAM

The two models, Random Forest (RF) and Support Vector Machine (SVM), were employed to predict the status of missing persons due to their well-established in classification tasks, particularly within complex and high-dimensional datasets. SVMs are adept at handling non-linear decision boundaries by leveraging kernel functions, making them highly adaptable to various data structures. Meanwhile, RFs, as ensemble learning methods, mitigate overfitting by aggregating the predictions of multiple decision trees. The specific parameters applied during the training phase are detailed in Table IV.

TABLE IV
TRAINING PARAMETERS

Model	Parameter	Value
Random Forest (RF)	Number of estimators	20
	Maximum Depth	20
	Minimum samples to split	2
	Minimum samples per leaf	1
	Class weights	balanced
Support Vector Machine (SVM)	Regularization parameter (C)	10
	Kernel	Radial Basis Function (rbf)
	Gamma	scale

III. RESULTS

In this section, descriptive statistics of the dataset used are presented to provide a comprehensive understanding of the data distribution and characteristics. Additionally, the performance of the models utilized, specifically the Random Forest (RF) and Support Vector Machine (SVM) models, was thoroughly evaluated using metrics to ensure a reliable and detailed assessment.

A. Descriptive statistics

The distribution of “Found” and “Deceased” cases in the dataset used for this research is presented in Fig. 2, highlighting the trends across months and the number of cases for both categories**.** There were 582 cases in February.

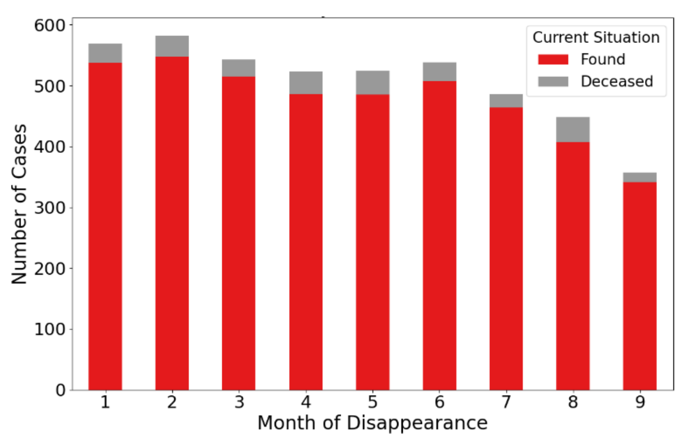

Fig. 2 Case distribution by month and current situation

Fig. 3 shows the age distribution of individuals categorized by their current situation ("Found" and "Deceased"). The "Found" group has a higher average age (35.03 ± 17.90) compared to the "Deceased" group (23.16 ± 15.86), indicating that younger individuals are more likely to be in the "Deceased" category. The "Found" group also shows variability in age, with several outliers above 80 years, while the "Deceased" group has a narrower range. This suggests a potential relationship between age and the likelihood of being found or deceased, with younger individuals at higher risk of being in the latter category.

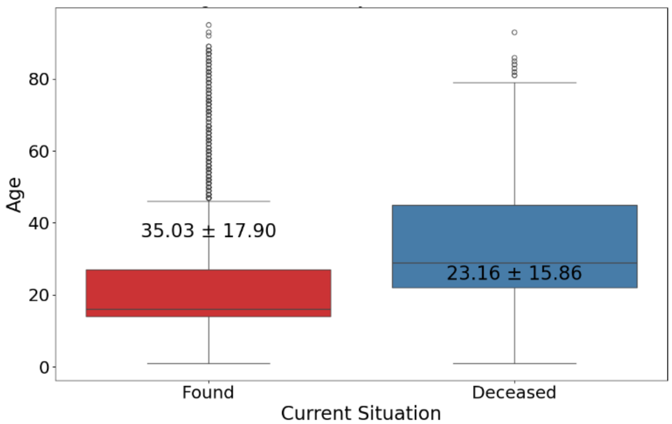

Fig. 3 Age distribution by current situation

Moreover, as illustrated in Fig. 4, females were predominantly represented in the "Deceased" category, with only a minimal proportion categorized as "Found." Their total frequency was significantly higher compared to males. While males were also primarily in the "Deceased" category, they exhibited a slightly higher proportion of individuals categorized as "Found" relative to females. This observation highlighted a potential gender disparity, where females were more frequently recorded as "Deceased," whereas males appeared to have a marginally better likelihood of being "Found."

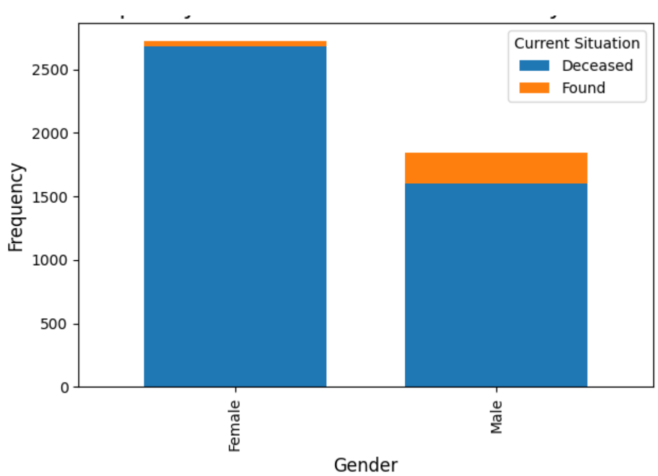

Fig. 4 Frequency distribution of "Found" and "Deceased" by gender

In the dataset, the variable Reason for Disappearance represents the initial assessment of the disappearance, made without prior study or investigation. Fig. 5 shows the five main reasons for disappearance. The leading causes were personal reasons (2,374 cases).

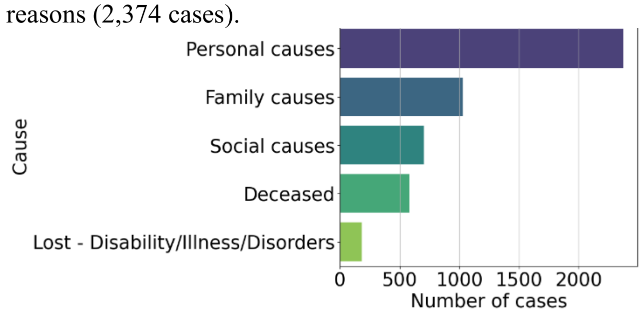

Fig. 5 Most frequent reasons for disappearance – based on 'reason for disappearance'.

Additionally, the dataset includes the variable Observed Reason for Disappearance, which refers to the cause officially assigned by competent judicial authorities following observation and investigation. These causes do not necessarily match the initial reasons reported at the time of disappearance. Fig. 6 presents the five most common observed causes.

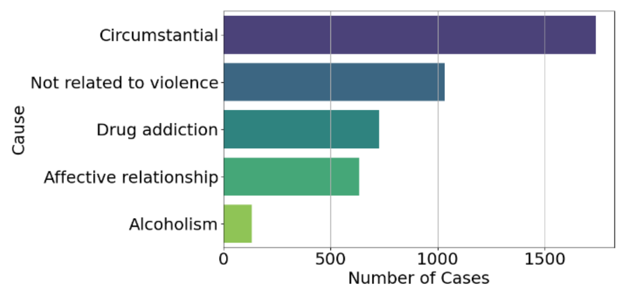

Fig. 6 The most frequent reasons for disappearance, based on the 'Reason for Disappearance' variable.

B. *Random Forest Performance*

The results of the metrics obtained from the test data using the Random Forest (RF) model are presented in Table V. Fig. 7 illustrates the confusion matrix generated by the model, providing valuable insights into its classification performance for two categories: "Found" and "Deceased." According to the matrix, the model accurately classified 741 cases as "Found" (True Positives) and 790 cases as "Deceased" (True Negatives). However, it also misclassified 119 cases as "Deceased" when they were actually "Found" (False Positives) and 66 cases as "Found" when they were actually "Deceased" (False Negatives).

TABLE V
METRICS OBTAINED FROM RANDOM FOREST MODEL

Metrics	Appears Dead
Accuracy	0.89
Precision	0.86
F1-score (Found)	0.89
F1-score (Deceased)	0.90
Recall	0.91

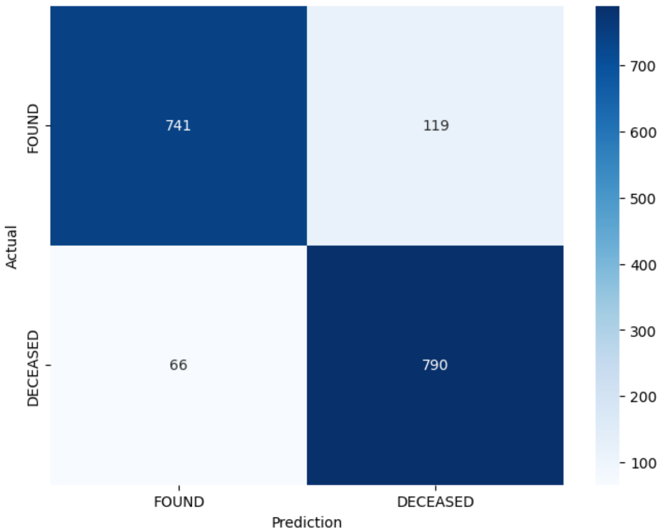

Fig. 7 Confusion Matrix for the RF model.

C. *Support vector machine performance*

The metrics obtained from the SVM model are presented in Table VI. Fig. 8 illustrates the confusion matrix, which shows that the model correctly identified 572 cases as "Found" (True Positives) and 712 cases as "Deceased" (True Negatives). However, it also misclassified 288 cases as "Deceased" when they were actually "Found" (False Positives) and 144 cases as "Found" when they were actually "Deceased" (False Negatives).

TABLE VI
METRICS OBTAINED FROM SUPPORT VECTOR MACHINE MODEL

Metrics	Appears Dead
Accuracy	0.75
Precision	0.66
F1-score (Found)	0.73
F1-score (Deceased)	0.77
Recall	0.79

IV. DISCUSSION

The phenomenon of disappearances represents a complex and urgent challenge for judicial, police, and humanitarian institutions. The study [21], which examined how Latin American countries specifically Colombia, Chile, Mexico, and Peru have incorporated international standards for the search for missing persons, as well as for the exhumation, recovery, protection, identification, and restitution of human remains, offers insights into institutional responses across the region. In [22] focused on the reasons, places and circumstances associated with forced disappearances in Latin America, revealing the complexity of quantifying these events due to the clandestine and systematic nature of the violence involved. However, these studies do not directly address the specific context of Ecuador. According to the Office of the Attorney General of Ecuador, approximately 10,000 missing persons reports are filed each year [23].

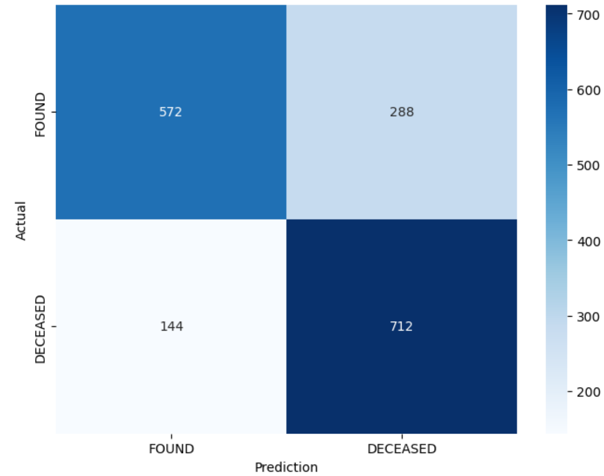

Fig. 8 Confusion Matrix for the SVM model.

To improve the dissemination and coordination of search efforts in Ecuador when a person had gone missing, the study in [23] proposed the SiGPro system, which automated early alerts through a progressive notification mechanism and geographic organization using the MQTT protocol.

Previous research has examined the extent to which international standards for the search and identification of missing persons are being implemented or has focused on analysing the underlying causes and contexts of disappearances. In contrast, this study addresses a critical gap by exploring the social and demographic profiles of missing persons in Ecuador. It distinguishes between those found alive ("Found") and those found deceased ("Deceased"). The findings reveal concerning patterns: the average age of individuals in the "Found" group is significantly higher (35.03 ± 17.90) than in the "Deceased" group (23.16 ± 15.86). Moreover, females are disproportionately represented in the "Deceased" category, while they are a minority among those found alive. A peak in disappearances occurs between January and February, with a high incidence of cases linked to circumstantial causes referring to situations where a person goes missing due to circumstances not necessarily associated with crime, violence, or coercion by third parties. Personal causes predominate during this period, typically involving voluntary disappearances related to personal conflicts, emotional distress, or intentional decisions to sever contact with family and social networks.

Currently, the agency responsible for this issue in Ecuador is DINASED (National Directorate of Crimes Against Life, Violent Deaths, Disappearances, Extortion, and Kidnappings). Recognizing socio-demographic and contextual patterns among missing persons not only improves search systems but also provides a solid foundation for prevention. This study suggests that it is possible to coordinate efforts between researchers, academic institutions, and DINASED to develop predictive models or early warning systems based on risk profiles.

V. CONCLUSIONS

This study analyzes the problem of disappearances in Ecuador by incorporating key information such as personal, social, and event-specific data. These categories were derived from the variables within the dataset. The research evaluates the performance of two machine learning models—Random Forest (RF) and Support Vector Machine (SVM)—to predict the status of missing persons, categorized as "Found" and "Deceased." The models were assessed using metrics such as precision, accuracy, F1 score, recall, and confusion matrices.

The RF model outperformed the SVM model, achieving a recall of 91% and a precision of 86%, identifying both status categories. This research provides a quantitative perspective on the status of missing persons in Ecuador, demonstrating the potential of machine learning models to assist decision-makers in tackling this highly complex and uncertain issue. In particular, the analysis reveals key characteristics and patterns among individuals reported missing in Ecuador during 2024, as well as the number of individuals found deceased or located alive based on these characteristics.

The findings underline the value of utilizing such datasets to identify key factors influencing the outcomes of disappearance cases. These insights can assist social organizations in allocating and optimizing resources, enhancing search strategies, and reimagining the role of law enforcement, particularly in better understanding and addressing the status of missing persons in Ecuador.

REFERENCES

[1] E. Nuzzolese, "Missing People, Migrants, Identification and Human Rights," *JOURNAL of FORENSIC ODONTO-STOMATOLOGY*, vol. 30, no. 1, pp. 47–59, 2012.

[2] K. Harrington, M. Brown, J. Pinchin, and S. Sharples, "Decision-making within missing person search," *Cognition, Technology and Work*, vol. 20, no. 4, pp. 665–680, Nov. 2018, doi: 10.1007/s10111-018-0515-x.

[3] W. Mohibullah and S. J. Julier, "Developing an agent model of a missing person in the wilderness," in *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, 2013, pp. 4462–4469. doi: 10.1109/SMC.2013.759.

[4] D. Reilly, M. Taylor, P. Fergus, C. Chalmers, and S. Thompson, "Misper-Bayes: A Bayesian network model for missing person investigations," *IEEE Access*, vol. 9, pp. 49990–50000, 2021, doi: 10.1109/ACCESS.2021.3069081.

[5] A. Dong *et al.*, "Predicting the locations of missing persons in China by using NGO data and deep learning techniques," *Int J Digit Earth*, vol. 17, no. 1, 2024, doi: 10.1080/17538947.2024.2304076.

[6] A. L. Ruiz-Rizzo, M. E. Archila-Meléndez, and J. J. F. González Veloza, "Predicting the probability of finding missing older adults based on machine learning," *J Comput Soc Sci*, vol. 5, no. 2, pp. 1303–1321, Nov. 2022, doi: 10.1007/s42001-022-00171-x.

[7] Y. Yang *et al.*, "r-Instance Learning for Missing People Tweets Identification," May 2018, [Online]. Available: http://arxiv.org/abs/1805.10856

[8] W. K. Aljohani *et al.*, "Suhail: A Deep Learning-Based System for Identifying Missing People," *Computer and Information Science*, vol. 16, no. 2, p. 36, Apr. 2023, doi: 10.5539/cis.v16n2p36.

[9] A. Syed Musthafa, S. Dinesh, K. Dinesh Kumar, C. Jeeva, and S. Madesh, "Digital Vigilance: AI Solutions in the Quest for Missing Persons using face recognition with deep learning algorithms," in *2nd International Conference on Artificial Intelligence and Machine Learning Applications: Healthcare and Internet of Things, AIMLA 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/AIMLA59606.2024.10531327.

[10] Y. Galphat, G. Amarnani, C. Bajaj, K. Mulchandani, and J. Repale, "ReUnite AI: Harnessing Face Detection and Age Progression for missing person identification," in *2024 2nd World Conference on Communication and Computing, WCONF 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/WCONF61366.2024.10692237.

[11] B. Vinavatani, M. R. Panna, P. H. Singha, and G. J. W. Kathrine, "AI for Detection of Missing Person," in *Proceedings - International Conference on Applied Artificial Intelligence and Computing, ICAAIC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 66–73. doi: 10.1109/ICAAIC53929.2022.9792672.

[12] G. E. R. Agudelo, C. E. M. Marín, and P. A. G. García, "Computational model to support the detection of profiles of missing person in Colombia," *Inteligencia Artificial*, vol. 24, no. 67, pp. 121–128, 2021, doi: 10.4114/intartif.vol24iss67pp121-128.

[13] V. Sudeksha, G. Lahari, A. N. S. Srinivas, K. Mannem, and K. Jamal, "Missing Persons Comprehensive Tracking System," in *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems, ICITEICS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICITEICS61368.2024.10625192.

[14] N. R. Fyfe, O. Stevenson, and P. Woolnough, "Missing persons: the processes and challenges of police investigation," *Policing Soc*, vol. 25, no. 4, pp. 409–425, Jul. 2015, doi: 10.1080/10439463.2014.881812.

[15] F. O'Brien, S. Giles, and S. Waring, "Relationships between demographic and behavioural factors and spatial behaviour in missing persons' cases," *Criminology and Criminal Justice*, vol. 23, no. 3, pp. 409–430, Jul. 2023, doi: 10.1177/17488958211060475.

[16] J. Vásconez, E. Schotborgh, I. Vásconez, V. Moya, A. Pilco, O. Menéndez, R. Guamán-Rivera, and L. Guevara, "Smart Delivery Assignment through Machine Learning and the Hungarian Algorithm," *Smart Cities*, vol. 7, no. 3, pp. 1109-1125, 2024, doi: org/10.3390/smartcities7030047

[17] "Personas Desaparecidas," Datos Abiertos Ecuador, https://datosabiertos.gob.ec/dataset/personas-desaparecidas/resource/7329fc75-94c3-4611-94ee-d84cd18fef1b (accessed Jan. 20, 2025).

[18] P. Li, S. Rao, J. Blase, Y. Zhang, X. Chu, and C. Zhang, "CleanML: A Study for Evaluating the Impact of Data Cleaning on ML Classification Tasks," in *Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE*), 2021, pp. 13–24, doi: 10.1109/ICDE51399.2021.00009.

[19] B. Charbuty, and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20-28, 2021.

[20] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[21] P. S. Taibi and M. C. Núñez, "The search for missing persons: Comparative analysis of the incorporation of international standards into national search plans implemented in Latin American countries," *International Review of the Red Cross*, vol. 1, pp. 1–23, 2025, doi: 10.1017/S1816383125000104

[22] O. Bernasconi, J. Jaramillo, and M. López, "The number of disappearance: trajectories in the tally of victims of forced disappearance in Latin America," *Tapuya: Latin American Science, Technology and Society*, vol. 5, no. 1, pp. 1–23, 2022, doi: 10.1080/25729861.2022.2090486.

[23] M. Zambrano, E. L. O. Mejía, and X. Calderón, "SIGPRO: A real-time progressive notification system using MQTT bridges and topic hierarchy for rapid location of missing persons," *IEEE Access*, vol. 8, pp. 149190–149198, 2020, doi: 10.1109/ACCESS.2020.3016476.