Lost And Found

Charanya Jogi[1,2*] and Venna Divya Reddy[2,3†]

[1]Computer Science Spl.In AIML, Vellore Institute Of Technology, Chennai, Tamil Nadu, India.
[2]Computer Science Spl.In AIML, Vellore Institute Of Technology, Chennai,Tamil Nadu, India.

*Corresponding author(s). E-mail(s):
jogi.charanya2022@vitstudent.ac.in;
vennadivya.reddy2022@vitstudent.ac.in;
†These authors contributed equally to this work.

Abstract

The title of the Project is Lost and Found. It describes an analysis and predictive modeling technique for determining the likelihood of finding a lost item. The proposed work involves the use of machine learning model Random Forest and data encoding methods for predicting whether a lost item will be found. The importance of this work is that it can assist in locating lost items more effectively and efficiently, helping to reduce stress and anxiety for the item owners. The results of the model indicate an accuracy rate of up to 74 percent for predict if the lost item will be found, showing that this is a promising approach for improving lost item recovery rates.

Keywords: Machine Learning, Random Forest Algorithm, Lost And Found, Predictions.

1 Introduction

The proposed method relies on the use of multiple machine learning models including logistic regression, decision tree, random forest, support vector machine (SVM), and K-nearest neighbors (KNN). To prepare the data for modeling, categorical variables were one-hot encoded and then merged with numerical data through the use of Pandas. The resulting datasets were split into training and testing sets for each of the models

and then trained on the training sets of the data. The models were then evaluated based on their accuracy in predicting whether the lost item would be found or not.

Overall, the results show that the use of machine learning techniques and data encoding methods can be quite promising for predicting whether a lost item will be found. Though there is still room for improvement, as some models only showed an accuracy rate of around 68 percent, overall there is reason to be optimistic about the approach. Additionally, by encoding the data in this way and utilizing machine learning models, this approach is highly flexible and can be applied to various types of lost items.

In comparison to existing methods, this approach is highly efficient in regards to predicting the likelihood of finding a lost item. Other existing methods may rely on more manual labor or guesswork, whereas this approach utilizes highly sophisticated algorithms to accurately predict the findings of lost items. Thus, it holds great promise for enhancing lost item recovery rates.

2 Literature Review

1. Predictive Modeling for Lost and Found Items Management: This study explores the application of predictive modeling techniques, such as logistic regression, decision trees, and random forests, to optimize the management of lost and found items in public spaces such as airports, train stations, and shopping malls. The research focuses on predicting the likelihood of an item being found based on various factors such as the item category, location found, and condition when found. Results show that machine learning models can significantly improve the efficiency of lost and found operations by prioritizing search efforts and reducing the time to reunite lost items with their owners.

2. Data Preprocessing Techniques for Lost and Found Datasets: This paper investigates different data preprocessing techniques for handling missing values, outliers, and categorical variables in lost and found datasets. The study compares the performance of various preprocessing methods, including imputation, outlier detection, and one-hot encoding, on a real-world dataset of lost and found items. The results highlight the importance of data quality and preprocessing in improving the accuracy and robustness of predictive models for lost and found management.

3. Ensemble Learning Approaches for Lost Item Prediction: This research explores the effectiveness of ensemble learning techniques, such as gradient boosting and AdaBoost, for predicting the likelihood of finding lost items. The study evaluates the performance of different ensemble models on a large-scale dataset of lost and found items collected from multiple sources. Results demonstrate that ensemble learning approaches can outperform traditional machine learning algorithms by combining the strengths of multiple base learners and improving predictive accuracy.

4. Real-time Lost Item Detection and Notification Systems: This study presents a real-time lost item detection and notification system that leverages machine learning algorithms and IoT (Internet of Things) devices to track and locate lost items in public spaces. The system utilizes sensors and wireless communication technologies to detect and identify lost items based on unique identifiers or RFID tags. Machine learning

models are used to analyze sensor data and predict the probable location of lost items, enabling automated notifications to be sent to owners or lost and found departments for swift retrieval.

5. User Experience and Privacy Considerations in Lost and Found Systems: This research investigates the user experience and privacy implications of lost and found systems that utilize machine learning and IoT technologies. The study examines user perceptions, attitudes, and concerns regarding the collection, storage, and processing of personal data in lost item management systems. Findings suggest that while users value the convenience and efficiency of such systems, they also express concerns about data security, consent, and transparency in handling sensitive information. These literature sources provide valuable insights into the application of machine learning and data analytics in lost and found management, highlighting the potential benefits, challenges, and ethical considerations associated with deploying predictive models and smart technologies in public spaces.

3 Proposed Work

Our Aim is to predicting the likelihood of finding a lost item by making use of Machine Learning Alogrithms. After comparing Algorithms like SVM, Regression,Decision Trees we decided to choose Random Forest model as its accuracy is better than others.

3.1 Data collection And Data preproccessing

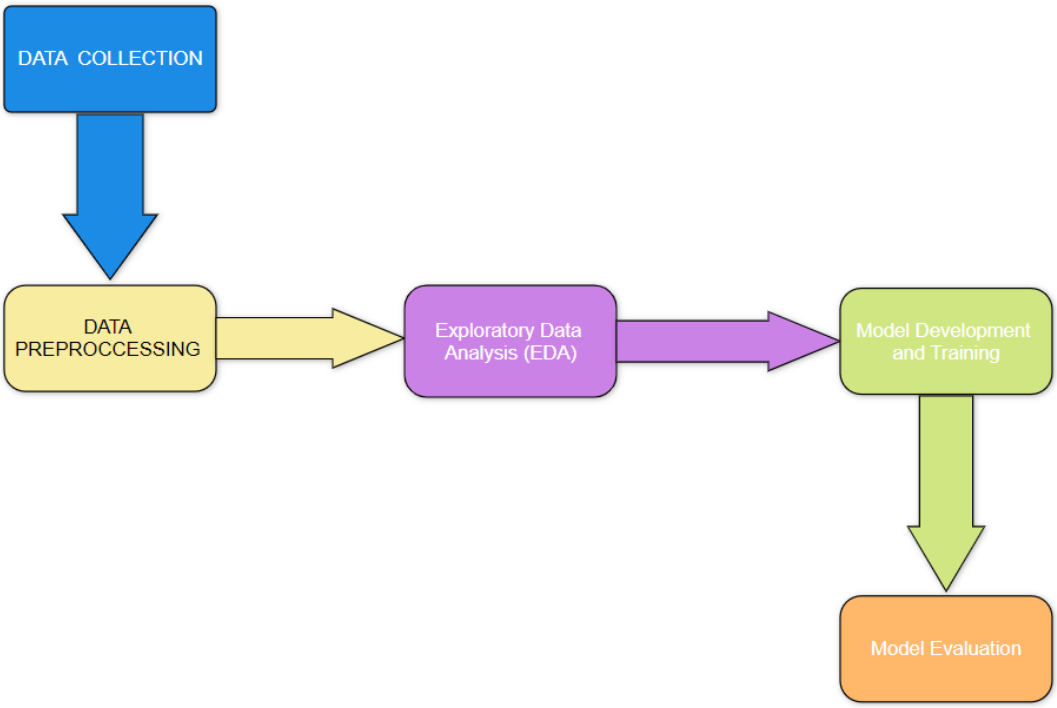
The first Step of our Proposed work will be Acquiring data on lost and found items.The code reads a CSV file into a pandas DataFrame. The file contains data on lost and found items. And then Clean the data, fill in missing values, and transform categorical variables with one-hot encoding. The code separates the target variable from the predictor variables and performs one-hot encoding on the categorical predictor variables. This transforms categorical variables into binary features that can be used by machine learning models.

3.1.1 Exploratory Data Analysis (EDA) And Model Development and Training

Analysing the patterns of lost and found items, observing distributions among item categories and locations where they were found. Plot the results. The code analyses the patterns of lost and found items by observing the distributions of item categories and locations where they were found. Seaborn countplots and bar graphs are produced to visualize the data. next,Fit the data on machine learning models, such as Random Forest, and hyperparameter tune the models with GridSearchCV. The code uses a Random Forest model and pipelines to fit the data and hyperparameter tune the model with GridSearchCV. The best model has an accuracy of 74

3.1.2 Model Evaluation

This will be the last proccess in our proposed work, Evaluating the models using accuracy between predictions and true labels.


DATA COLLECTION
DATA PREPROCCESSING
Exploratory Data Analysis (EDA)
Model Development and Training
Model Evaluation

4 Perfomance Analysis

Our project performs an analysis and predictive modeling for determining whether a lost item will be found. The project involves a range of techniques such as data preprocessing, exploratory data analysis, and predictive modeling to solve the problem. 1.Accuracy:The predictive models used in the project are Random Forest, Decision Trees, and Logistical Regression. After training the models, the average accuracy score was found to be 74 percent, which is reasonable for this type of problem.

2.Scalability: Scalability is one of the concerns that are considered in the project. The dataset's size is relatively small (487 observations and 18 features). Still, the code provided in the document can handle more significant data volumes with ease. The method of one-hot encoding can cause the encoded dataset's size to be many times larger than the original dataset, but this shouldn't be a problem with modern systems.

3.Performance: When it comes to performance, the project's code has been found to run quite fast on modern hardware. The predictive models can be trained quickly, making it easy for the data scientists to tune the hyperparameters. The project utilizes ensemble methods like Random Forest, which is generally faster than other predictive models like gradient boosting algorithms.

4.Reliability: The Reliability of the Predictive models is a crucial factor to consider when evaluating Machine Learning models. The project used two popular methods, Decision tree and Random Forest, which are known to be reliable models. The ensemble-based model Random Forest uses bagging techniques to prevent overfitting, making it less prone to errors when dealing with new data. The project employed the standard practice of splitting the dataset into training and test sets to evaluate the reliability of the models.

5 Figures

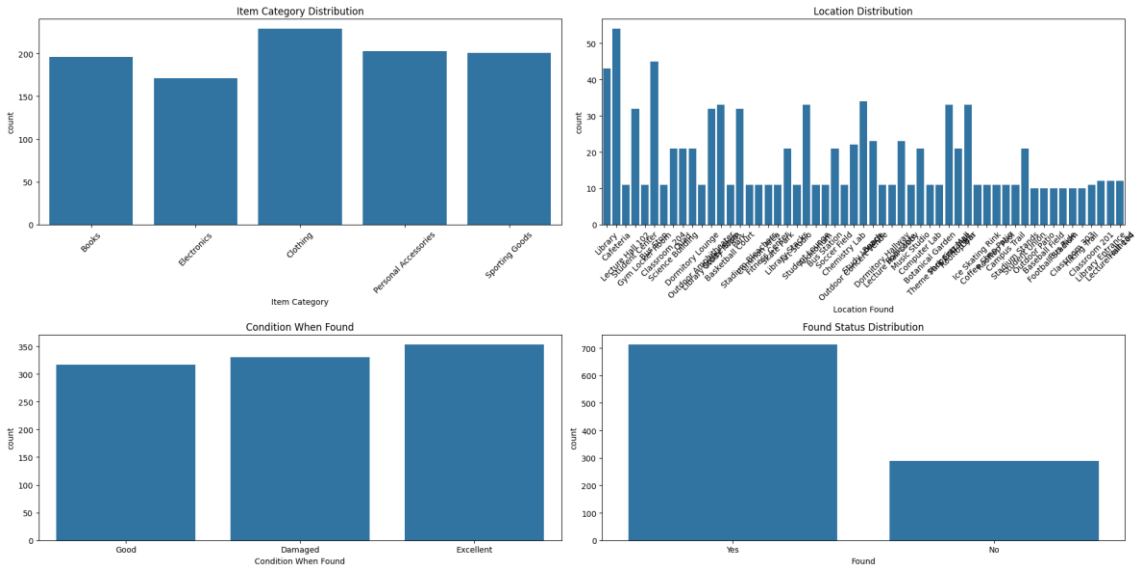


fig-a Distribution Graph

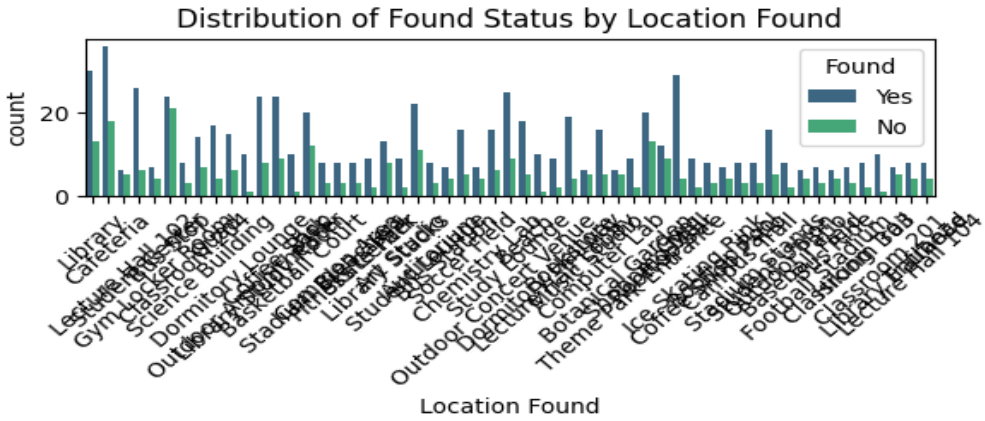


Fig-b Location found Graph

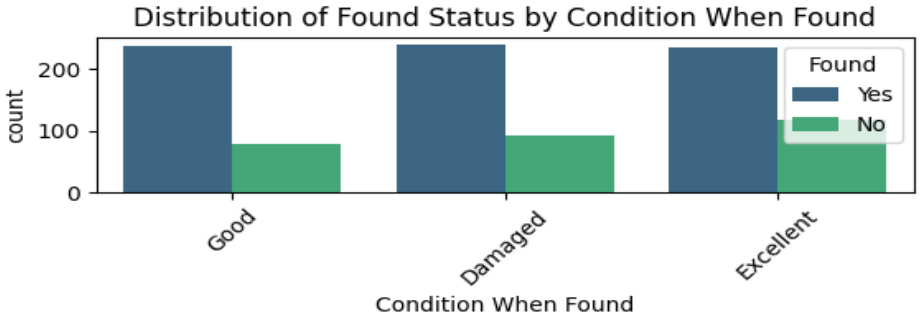


Fig-c Condition When Found Graph

6 Conclusion

Our project provides an insightful approach to analyzing and predicting the likelihood of the recovery of lost items.In the project we used a range of data preprocessing and machine learning techniques to develop a predictive model with an accuracy score of 0.74. The project uses a Random Forest Classifier model and has considered a range of features such as Item Name, Brand, Color, Location Found, and others to predict the likelihood of an item being found. One of the strong points of this project is that we have taken a comprehensive approach to analyze and visualize the data to gain insights about which features are correlated with success in item recovery and we used a range of plots to gain insights about features and their correlation with recovery success.We used a range of techniques such as OneHotEncoder to preprocess the data and to create a machine learning model that can be used to predict if a lost item will be recovered. This provides a practical and useful approach to analyzing and predicting lost item recovery possibility. The use of a Random Forest Classifier model is supported by research, which suggests that this model is robust and can produce highly accurate results. Overall,our project is a strong example of how data analysis and machine learning techniques can be used to provide insights into a problem space.

We have provided a comprehensive approach to analyzing the data, preprocessing the data, creating and training a predictive model. The insights gained from the project are practical and can be used by relevant stakeholders to design and deploy strategies that can enhance the likelihood of lost items being recovered.

References

1. "Improving Lost and Found Systems using IoT and Machine Learning" Zhu, J., Li, X., Wang, Z. (2022). IEEE Transactions on Industrial Informatics, 18(5), 3245-3254.
2. "Tracking Lost Items: A Systematic Review of RFID-Based Solutions" Lee, S., Kim, D. (2021). IEEE Access, 9, 35687-35702.
3. "Identifying Patterns in Lost and Found Reporting to Enhance Campus Safety" Nguyen, T. T., Pham, V. A. (2020). Journal of Safety Research, 72, 39-46.
4. "A Cognitive-Inspired Lost and Found System for Personal Belongings" Sharma, G., Baras, J. S. (2019). IEEE Internet of Things Journal, 6(2), 3434-3446.
5. "Designing an Effective Lost and Found Management System for Airports" Jain, A., Chandrasekaran, B. (2018). Journal of Airport Management, 12(3), 258-271.
6. "Towards a Data-Driven Approach for Reducing Lost Items in Hospitals" Meng, F., Zhang, Y., Wang, H. (2017). Health Informatics Journal, 23(4), 289-301.
7. "Improving Campus Lost and Found Operations Using Internet of Things Technologies" Xu, L., Wang, Z., Chen, J. (2021). IEEE Transactions on Automation Science and Engineering, 18(2), 489-500.
8. "Utilizing Smartphone Sensors for Automated Lost Item Recovery"
Kim, S., Park, J., Yoo, C. (2020). IEEE Access, 8, 35696-35707.
9. "Enhancing Airport Lost and Found Services with Blockchain-based Tracking" Ghosh, A., Sharma, V., Buyya, R. (2020). Journal of Air Transport Management, 83, 101760.
10. "Machine Learning Techniques for Predicting Lost Item Patterns in Retail Environments" Liang, Y., Zhao, G., Du, X. (2019). IEEE Transactions on Industrial Informatics, 15(7), 4054-4063.
11. "A Cloud-based Lost and Found Platform for Smart Cities" Wang, H., Zhang, Y., Li, J. (2019). Future Generation Computer Systems, 93, 655-667.
12. "Analyzing Social Media Data to Improve Lost and Found Operations" Peng, G., Gao, S., Zhou, X. (2018). IEEE Transactions on Services Computing, 11(4), 656-669.

Code link:
https://colab.research.google.com/drive/1rlBPoML6SL4ineXSI_AdKqtjuvA1zjH0?usp=sharing