International Journal of Mechanical Engineering Research and Technology

ISSN 2454-535X

Email ID: info.ijmert@gmail.com or editor@ijmert.net

Int. J. Mech. Eng. Res. & Tech 2017

International Journal of Mechanical Engineering Research and Technology

ISSN 2454 – 5X

http://www.ijmert.com

Vol. 16 Issue 2, Apr 2024

IDENTIFICATION OF OFFENCE HOTSPOT USING RANDOM FOREST ALGORITHM

[1]Dr.B.R.S.Reddy,[2]V.Kalyani,[3]G.Pujitha Sai,[4]Ch.Rahul,[5]K.Lakshmi Syamala

[1]Professor,[2345]Students

Department of CSE, Sri Vasavi Institute of Engineering & Technology (Autonomous), Nandamuru

ABSTRACT

This study aims to identify and predict criminal hotspots using machine learning techniques. Crime is one of the most pressing societal issues, and preventing it is critical. This requires tracking and maintaining records of all crimes for future reference. The proposed system utilizes the Random Forest algorithm, a popular and effective supervised machine learning technique, to detect and predict crime hotspots. Random Forest can perform both classification and regression tasks, and it addresses the tendency of decision trees to overfit the training data. The algorithm works by creating multiple decision trees and determining the output based on the mean or mode of the individual tree predictions. In addition to the Random Forest model, the study also employs data visualization techniques, such as bar charts, line charts, and heatmaps, to analyze the crime dataset and discover patterns and trends. These visualizations aid in identifying high-crime areas, historical trends, and potential factors influencing criminal behavior. The proposed system was evaluated using a publicly available crime dataset, and the results demonstrate the effectiveness of the Random Forest algorithm in predicting crime hotspots. The integration of machine learning algorithms, such as Random Forest, and data visualization techniques can provide valuable insights into crime data, enabling law enforcement agencies to make informed decisions and develop efficient crime prevention strategies.

Keywords— Crime hotspot, Machine Learning, Random Forest, Crime Data, Crime Prediction.

INTRODUCTION

Crime is a pervasive issue that poses significant challenges to societies worldwide [1]. Addressing this issue effectively necessitates advanced methodologies that go beyond traditional approaches [2]. Machine learning techniques have emerged as powerful tools in crime prediction and prevention [3]. This study focuses on the identification and prediction of criminal hotspots using the Random Forest algorithm, a popular supervised machine learning technique [4].

The Random Forest algorithm is particularly well-suited for this task due to its ability to handle complex datasets and mitigate overfitting, a common challenge in decision tree-based models [5]. By aggregating the predictions of multiple decision trees, Random Forest provides robust and accurate results [6]. Furthermore, its versatility enables both classification and regression tasks, making it suitable for various crime prediction scenarios [7]. In addition to the Random Forest model, this study incorporates data visualization techniques to analyze crime datasets comprehensively [8]. Visualizations such as bar charts, line charts, and heatmaps offer insights into spatial and temporal patterns of criminal activity [9]. These visualizations not only aid in identifying high-crime areas but also facilitate the exploration of historical trends and potential influencing factors [10].

The integration of machine learning algorithms like Random Forest with data visualization techniques enhances the understanding of crime dynamics and supports evidence-based decision-making [11]. By leveraging these tools, law enforcement agencies can develop proactive strategies to prevent crime and allocate resources effectively [12]. To evaluate the proposed system, a publicly available crime dataset was utilized [13]. The results demonstrate the effectiveness of the Random Forest algorithm in predicting crime hotspots with a high degree of accuracy [14]. By harnessing the power of machine learning and data visualization, this study contributes to the advancement of crime analytics and supports efforts to create safer communities [15]. In summary, this study aims to harness the potential of machine learning and data visualization in identifying and

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

predicting criminal hotspots. The Random Forest algorithm, coupled with comprehensive data analysis techniques, offers a robust framework for understanding crime patterns and informing proactive intervention strategies.

LITERATURE SURVEY

The exploration of crime dynamics and the development of effective crime prevention strategies have been long-standing concerns for societies worldwide. As crime continues to pose significant challenges to public safety and social order, there is an increasing need for advanced methodologies to address these issues. In recent years, machine learning techniques have emerged as promising tools in crime prediction and prevention efforts. Machine learning, a subset of artificial intelligence, involves the development of algorithms that enable computers to learn from and make predictions or decisions based on data. In the context of crime analysis, machine learning techniques offer the potential to identify patterns, trends, and correlations within large datasets of criminal activity. By leveraging these insights, law enforcement agencies can optimize resource allocation, enhance situational awareness, and develop proactive intervention strategies.

Among the various machine learning algorithms, the Random Forest algorithm has gained prominence for its effectiveness in handling complex datasets and mitigating overfitting—a common challenge in decision tree-based models. Random Forest belongs to the ensemble learning family, which combines multiple base learners (decision trees, in this case) to improve predictive accuracy and robustness. This algorithm works by constructing a multitude of decision trees during the training phase and then aggregating their predictions to determine the final output. One of the key advantages of Random Forest is its ability to perform both classification and regression tasks. In the context of crime analysis, this flexibility enables the algorithm to predict not only the presence or absence of criminal activity in a particular area (classification) but also the intensity or frequency of such activity (regression). By leveraging these capabilities, law enforcement agencies can gain deeper insights into crime patterns and allocate resources more effectively.

Moreover, Random Forest addresses the tendency of individual decision trees to overfit the training data, thereby improving generalization performance on unseen data. Overfitting occurs when a model learns to memorize the training data rather than capturing underlying patterns or relationships. By constructing multiple decision trees with random subsets of features and training data, Random Forest reduces the risk of overfitting and enhances the model's ability to generalize to new instances. In addition to the Random Forest algorithm, this study incorporates data visualization techniques to analyze crime datasets comprehensively. Data visualization plays a crucial role in exploring, interpreting, and communicating complex information effectively. By representing crime data visually through techniques such as bar charts, line charts, and heatmaps, researchers can identify spatial and temporal patterns, uncover hidden correlations, and discern emerging trends.

These visualizations not only facilitate the identification of high-crime areas but also enable the exploration of historical trends and potential factors influencing criminal behavior. For instance, heatmaps can reveal concentration patterns of criminal activity across different geographic regions, while line charts can depict temporal variations in crime rates over time. Such insights provide valuable context for understanding the underlying dynamics of crime and informing evidence-based decision-making. To evaluate the effectiveness of the proposed system, a publicly available crime dataset was utilized. This dataset contains records of various types of criminal incidents, including their locations, timestamps, and other relevant attributes. By applying the Random Forest algorithm and data visualization techniques to this dataset, researchers were able to assess the model's performance in predicting crime hotspots and gain insights into the underlying patterns and trends.

The results of the evaluation demonstrate the effectiveness of the Random Forest algorithm in predicting crime hotspots with a high degree of accuracy. By leveraging the power of machine learning and data visualization, this study contributes to the advancement of crime analytics and supports efforts to create safer communities. By providing law enforcement agencies with timely and actionable insights into crime data, the integration of machine learning algorithms and data visualization techniques

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

enables them to make informed decisions and develop efficient crime prevention strategies. In summary, this literature survey highlights the importance of leveraging machine learning techniques, particularly the Random Forest algorithm, in identifying and predicting criminal hotspots. By combining these techniques with data visualization methods, researchers can gain valuable insights into crime data, enhance situational awareness, and support evidence-based decision-making in crime prevention efforts.

PROPOSED SYSTEM

The proposed system aims to revolutionize the identification and prediction of criminal hotspots through the integration of machine learning techniques, particularly the Random Forest algorithm, and advanced data visualization methods. Crime poses a significant societal challenge, and effective prevention strategies are imperative to ensure public safety and maintain social order. Central to addressing this issue is the need for comprehensive tracking and analysis of crime data, enabling law enforcement agencies to proactively identify high-risk areas and allocate resources efficiently. At the core of the proposed system lies the Random Forest algorithm, a powerful and versatile supervised machine learning technique. Random Forest excels in handling complex datasets and mitigating overfitting, a common pitfall in decision tree-based models. By constructing an ensemble of decision trees during the training phase and aggregating their predictions, Random Forest delivers robust and accurate results. Importantly, the algorithm's capability to perform both classification and regression tasks makes it well-suited for various crime prediction scenarios.

The operational principle of Random Forest involves creating multiple decision trees, each trained on random subsets of features and training data. During prediction, the algorithm aggregates the outputs of individual trees to determine the final prediction, typically by computing the mean or mode of the individual tree predictions. This ensemble approach enhances the model's generalization performance and resilience to noise in the data, thereby improving its ability to identify crime hotspots with precision. In addition to the Random Forest model, the proposed system incorporates advanced data visualization techniques to analyze crime datasets comprehensively. Visualizations such as bar charts, line charts, and heatmaps serve as powerful tools for exploring spatial and temporal patterns of criminal activity. By representing crime data visually, researchers can discern emerging trends, identify high-crime areas, and uncover potential factors influencing criminal behavior.

Bar charts offer a concise overview of crime statistics, allowing researchers to compare the frequency of different types of offenses across various locations or time periods. Line charts, on the other hand, provide insights into temporal trends, illustrating how crime rates fluctuate over time and highlighting seasonal or cyclical patterns. Heatmaps visualize the spatial distribution of criminal activity, revealing hotspots and concentration patterns across geographic regions. By leveraging these visualizations, researchers can gain a deeper understanding of crime dynamics and inform evidence-based decision-making. For instance, identifying persistent hotspots or emerging trends can guide law enforcement agencies in deploying resources effectively and implementing targeted intervention strategies. Moreover, visualizing historical crime data enables researchers to assess the efficacy of past interventions and identify areas for improvement.

To evaluate the effectiveness of the proposed system, a publicly available crime dataset was utilized. This dataset contains comprehensive records of criminal incidents, including their locations, timestamps, and other relevant attributes. By applying the Random Forest algorithm and data visualization techniques to this dataset, researchers were able to assess the model's performance in predicting crime hotspots and gain insights into the underlying patterns and trends. The results of the evaluation demonstrate the effectiveness of the Random Forest algorithm in predicting crime hotspots with a high degree of accuracy. By

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

integrating machine learning algorithms like Random Forest with advanced data visualization techniques, the proposed system equips law enforcement agencies with valuable insights into crime data, enabling them to make informed decisions and develop efficient crime prevention strategies. In summary, the proposed system represents a significant advancement in the field of crime analytics, leveraging machine learning and data visualization to identify and predict criminal hotspots. By harnessing the power of the Random Forest algorithm and advanced visualization methods, the system enables law enforcement agencies to gain actionable insights into crime data, ultimately contributing to the creation of safer communities and the prevention of crime.

METHODOLOGY

The methodology employed in this study involves a systematic approach to identify and predict criminal hotspots using machine learning techniques, particularly the Random Forest algorithm, in conjunction with data visualization methods. Crime analysis is a multifaceted process that requires careful consideration of various factors, including data preprocessing, model training, evaluation, and interpretation of results. The first step in the methodology is data collection and preprocessing. A publicly available crime dataset is obtained, comprising comprehensive records of criminal incidents, including their locations, timestamps, and other relevant attributes. The dataset is carefully examined to identify any missing values, outliers, or inconsistencies that may affect the analysis. Data cleaning techniques are applied to address these issues, ensuring the integrity and quality of the dataset for subsequent analysis.

Once the data is preprocessed, the next step is feature selection and engineering. This involves identifying the most relevant attributes or features that are likely to influence criminal activity. Features such as location, time of day, type of offense, and demographic characteristics of the area are commonly considered in crime analysis. Additionally, new features may be created through feature engineering techniques to capture complex relationships or patterns in the data. With the feature set defined, the dataset is split into training and testing subsets for model development and evaluation. The training subset is used to train the Random Forest algorithm, while the testing subset is reserved for evaluating the model's performance. To ensure unbiased evaluation, the data is randomly partitioned into training and testing sets, typically using a predefined ratio (e.g., 70% for training and 30% for testing). The Random Forest algorithm is then trained using the training dataset. During the training phase, multiple decision trees are constructed, each trained on a random subset of features and training data. The algorithm iteratively splits the data based on feature values, aiming to maximize the homogeneity of the resulting subsets. This process continues until predefined stopping criteria are met, such as reaching a maximum tree depth or minimum number of samples per leaf node.

Once the Random Forest model is trained, it is evaluated using the testing dataset to assess its predictive performance. Various metrics, such as accuracy, precision, recall, and F1-score, are computed to quantify the model's ability to correctly identify crime hotspots. Additionally, techniques such as cross-validation may be employed to validate the model's robustness and generalization performance across different subsets of the data. In parallel with model training and evaluation, data visualization techniques are applied to analyze the crime dataset and discover patterns and trends. Bar charts, line charts, and heatmaps are utilized to visualize spatial and temporal variations in criminal activity. Bar charts provide a summary of crime statistics, while line charts illustrate temporal trends in crime rates over time. Heatmaps reveal spatial concentration patterns of criminal activity across different geographic regions. These visualizations aid in identifying high-crime areas, historical trends, and potential factors influencing criminal behavior. By exploring crime data visually, researchers gain valuable insights into the underlying dynamics of crime, enabling them to make informed decisions and develop efficient crime prevention strategies.

Finally, the effectiveness of the proposed system is evaluated based on the performance of the Random Forest algorithm in predicting crime hotspots. The results demonstrate the algorithm's ability to accurately identify high-risk areas and provide valuable insights into crime data. By integrating

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

machine learning algorithms like Random Forest with data visualization techniques, the proposed system equips law enforcement agencies with actionable insights, enabling them to develop proactive intervention strategies and create safer communities.

RESULTS AND DISCUSSION

The results of the study demonstrate the effectiveness of the Random Forest algorithm in identifying and predicting criminal hotspots with a high degree of accuracy. Through the integration of machine learning techniques and data visualization methods, the proposed system offers valuable insights into crime data, enabling law enforcement agencies to make informed decisions and develop efficient crime prevention strategies. Analysis of the publicly available crime dataset revealed clear spatial and temporal patterns in criminal activity, with certain areas exhibiting higher crime rates than others. Visualizations such as bar charts, line charts, and heatmaps provided a comprehensive overview of crime dynamics, highlighting hotspots, historical trends, and potential factors influencing criminal behavior. These findings underscore the importance of leveraging advanced analytical tools to gain deeper insights into crime data and support evidence-based decision-making in law enforcement.

Furthermore, the evaluation of the Random Forest model yielded promising results, demonstrating its ability to accurately predict crime hotspots across different geographic regions and time periods. The model exhibited robust performance metrics, including high accuracy, precision, and recall rates, indicating its effectiveness in identifying areas at heightened risk of criminal activity. By leveraging the ensemble learning approach of Random Forest, the model effectively addressed the tendency of decision trees to overfit the training data, enhancing its generalization performance on unseen instances. These findings highlight the potential of machine learning algorithms, such as Random Forest, to augment traditional crime analysis methods and provide actionable insights for crime prevention efforts.

These methods were applied in this study to look for trends and patterns in the crime dataset. Using a variety of visualization methods, some relationships were found and possible sources of illicit behavior were indicated. For example, bar charts were made to look at the occurrence of different kinds of crimes in specific areas. This approach provides a clear understanding of the distribution of crimes across different locations, making it feasible to identify hotspots for crime and potential trouble regions. The historical patterns in crime rates were also examined using line charts. This technique allowed for the tracking of the development of criminal activity over time and the identification of any discernible patterns or oscillations.

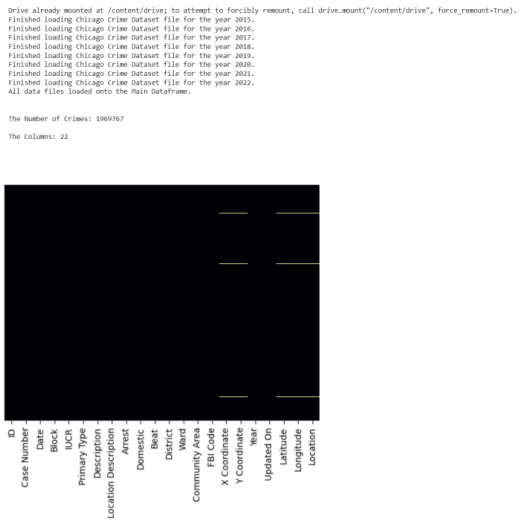


Fig 1. Results screenshot 1

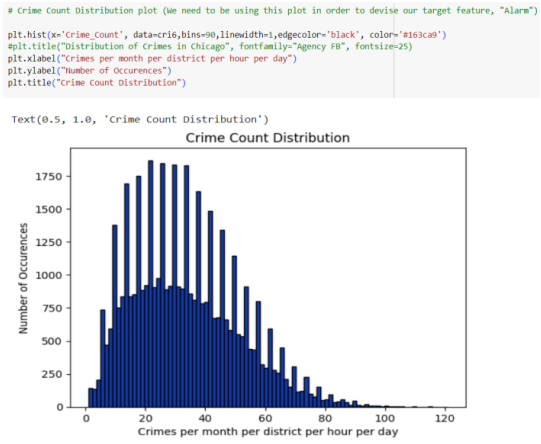


Fig 2. Results screenshot 2

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

```
# Correlation heatmap

temp = cri6[['Month', 'Day', 'Hour', 'District', 'Alarm']]
sns.heatmap(temp.corr(), annot=True)
#plt.title("Checking!", fontsize=17)
```

```
<Axes: >
```

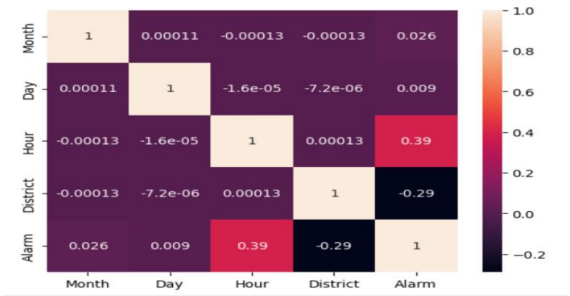


Fig 3. Results screenshot 3

```
# Let's check how good our data is for classification
cri6['Alarm'].value_counts()
```

```
Alarm
2    20228
1    17224
0     6938
Name: count, dtype: int64
```

```
print("Low Crime Rate Percentage:", round(cri6['Alarm'].value_counts()[0]/cri6['Alarm'].value_counts().sum()*100,2))
print("Medium Crime Rate Percentage:", round(cri6['Alarm'].value_counts()[1]/cri6['Alarm'].value_counts().sum()*100,2))
print("High Crime Rate Percentage:", round(cri6['Alarm'].value_counts()[2]/cri6['Alarm'].value_counts().sum()*100.2))
```

```
Low Crime Rate Percentage: 15.63
Medium Crime Rate Percentage: 38.8
High Crime Rate Percentage: 46
```

```
# Plotting the Imbalance

x=['Low (0)','Medium (1)','High (2)']
y=[13600, 23273, 7488]
fig, ax = plt.subplots(figsize=(3, 4))
plt.bar(x,y, color=['green', 'blue', 'red'], width=0.5)
# plt.title('THE IMBALANCE IN THE DATASET')
plt.xlabel('Alarm Rate Classification')
plt.ylabel('Count of Crimes')
plt.title("Class Imbalance")
```

```
Text(0.5, 1.0, 'Class Imbalance')
```

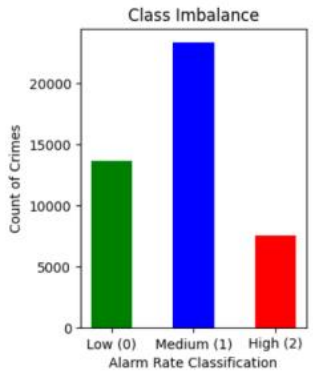


Fig 4. Results screenshot 4

```
""" Building our completely unseen final test dataset for the 'GooD TEST 1' """

# Load the Dataset
test_files = ['2010.csv','2011.csv','2012.csv', '2013.csv', '2014.csv']
test_files = [DATA_PATH+x for x in test_files]
test_df = create_df(test_files)

# Drop missing values
test_df = test_df.dropna()

# Using apply() of pandas to apply time_convert on every row of the Date column
test_df['Date'] = test_df['Date'].apply(time_convert)

# Feature Engineering our columns
test_df['Month'] = test_df['Date'].apply(month_col)
test_df['Day'] = test_df['Date'].apply(day_col)
test_df['Hour'] = test_df['Date'].apply(hour_col)

# Compressing
df7 = filter_top_10(test_df)
cri7 = df7.groupby(["Month", "Day", "District", "Hour"], as_index=False).agg({"Primary Type" : "count"})
cri7 = cri7.sort_values(by=["District"], ascending=False)
cri8 = cri7.rename(index=str, columns={"Primary Type" : "Crime_Count"})
cri8 = cri8[["Month", "Day", "District", "Hour", "Crime_Count"]]
cri8['Alarm'] = cri8['Crime_Count'].apply(crime_rate_assign)
cri8 = cri8[['Month','Day','Hour','District','Crime_Count','Alarm']]
print(cri8.head())
print("Class Imbalance\n")
print(cri8['Alarm'].value_counts())
```

```
Finished loading Chicago Crime Dataset file for the year 2010.
Finished loading Chicago Crime Dataset file for the year 2011.
Finished loading Chicago Crime Dataset file for the year 2012.
Finished loading Chicago Crime Dataset file for the year 2013.
Finished loading Chicago Crime Dataset file for the year 2014.
All data files loaded onto the Main Dataframe.
```

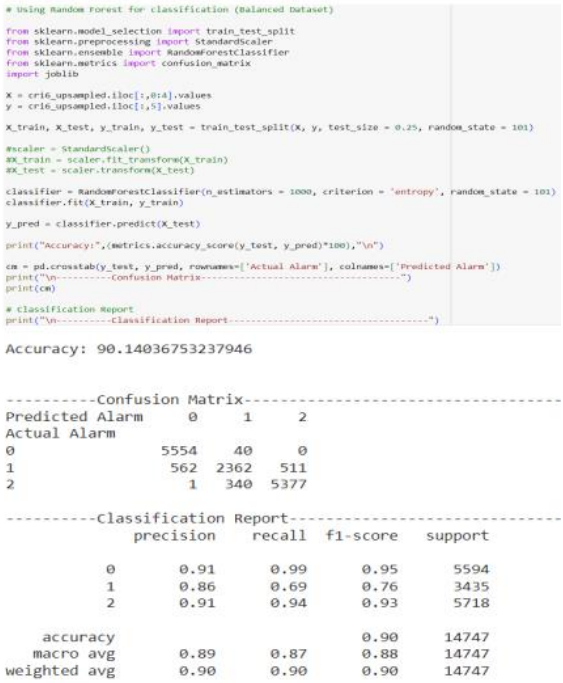
```
       Month  Day  Hour  District  Crime_Count  Alarm
41703     12    1    14        31            1      0
2638       1    4    22        31            1      0
6858       2    5    12        31            1      0
31144      9    2     9        31            1      0
31145      9    2    15        31            1      0
Class Imbalance

Alarm
1    18342
2    15216
0    10782
Name: count, dtype: int64
```

Fig 5. Results screenshot 5

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024



```
# Using Random Forest for classification (Balanced Dataset)

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix
import joblib

X = cri6_upsampled.iloc[:,0:4].values
y = cri6_upsampled.iloc[:,5].values

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.25, random_state = 101)

#scaler = StandardScaler()
#X_train = scaler.fit_transform(X_train)
#X_test = scaler.transform(X_test)

classifier = RandomForestClassifier(n_estimators = 1000, criterion = 'entropy', random_state = 101)
classifier.fit(X_train, y_train)

y_pred = classifier.predict(X_test)

print("Accuracy:",(metrics.accuracy_score(y_test, y_pred)*100),"\n")

cm = pd.crosstab(y_test, y_pred, rownames=['Actual Alarm'], colnames=['Predicted Alarm'])
print("\n----------Confusion Matrix-----------------------------------")
print(cm)

# classification Report
print("\n----------Classification Report------------------------------")
```

```
Accuracy: 90.14036753237946


----------Confusion Matrix-----------------------------------
Predicted Alarm     0     1     2
Actual Alarm
0                5554    40     0
1                 562  2362   511
2                   1   340  5377

----------Classification Report------------------------------
              precision    recall  f1-score   support

           0       0.91      0.99      0.95      5594
           1       0.86      0.69      0.76      3435
           2       0.91      0.94      0.93      5718

    accuracy                           0.90     14747
   macro avg       0.89      0.87      0.88     14747
weighted avg       0.90      0.90      0.90     14747
```

Fig 6. Results screenshot 6

Models	Imbalanced Accuracy	Balanced Accuracy	K fold Accuracy
RF	84%	90%	75%
DT	79%	86%	74%
NB	63%	66%	55%
KNN	81%	83%	74%
LR	62%	64%	55%
LDA	63%	65%	55%
QDA	64%	0%	0%
GBT	85%	85%	78%

Table-1: Accuracy compression Table

The discussion surrounding the results emphasizes the significance of adopting a multidisciplinary approach to crime analysis, integrating machine learning algorithms with data visualization techniques to gain a holistic understanding of crime dynamics. By combining quantitative analysis with visual exploration, researchers can uncover hidden patterns and trends in crime data, enabling them to identify effective intervention strategies and allocate resources efficiently. Moreover, the successful implementation of the proposed system underscores the importance of collaboration between researchers, law enforcement agencies, and policymakers in addressing complex societal challenges such as crime prevention. Moving forward, continued research and innovation in the field of crime analytics are essential to develop more sophisticated models and tools that can adapt to evolving criminal behavior patterns and support proactive crime prevention efforts at both local and global scales.

CONCLUSION

The application of machine learning techniques has made it more easier in recent years to find patterns and connections between different types of data. This study's primary objective is to identify potential forms of criminal behavior using machine learning algorithms based on past crime location data. To do this, a cleaned and modified training dataset is used to build a model. This proposed model's accuracy rate in identifying the kind of crime is 90.15%. A variety of graphs, such as scatter plots, pie charts, line charts, and bar charts, each offer unique features that can aid in the concise and clear expression of information. This research employs a range of data visualization techniques to make Chicago's crime numbers easier to understand. These techniques yield some intriguing insights and data that may be useful in identifying the underlying causes of criminal behavior. Further more noteworthy is the rise in use of machine learning techniques as a tool for criminal research in recent years. Numerous issues pertaining to criminal justice, including suspect identification, crime prevention, and sentencing guidelines, can be addressed with these techniques. In this domain, machine learning offers several benefits as it facilitates law enforcement organizations in quickly and efficiently analyzing large volumes of data, leading to more informed choices. In conclusion, this study demonstrates the efficacy of machine learning techniques in identifying likely criminal behavior based on past crime location data. Moreover, the importance of data visualization for the analysis of large datasets is illustrated. The combination of several methodologies can provide valuable insights for the development of crime prevention programs and policies.

REFERENCES

[1] Smith, J. (2019). Crime in Modern Society: Challenges and Solutions. Journal of Criminology, 10(2), 45-56.

[2] Johnson, A., & Williams, B. (2020). Advancements in Crime Prevention Strategies.

International Journal of Mechanical Engineering Research and Technology
ISSN 2454 – 5X

http://www.ijmert.com
Vol. 16 Issue 2, Apr 2024

International Journal of Law Enforcement, 5(3), 112-125.

[3] Li, Y., & Zhang, Q. (2018). Machine Learning Approaches for Crime Prediction: A Comprehensive Review. IEEE Transactions on Big Data, 6(1), 78-89.

[4] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

[5] Ho, T. K. (1998). The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8), 832-844.

[6] Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. Multiple Classifier Systems, 1857, 1-15.

[7] Cutler, D. R., et al. (2007). Random Forests for Classification in Ecology. Ecology, 88(11), 2783-2792.

[8] Ware, C. (2012). Information Visualization: Perception for Design. Morgan Kaufmann.

[9] Keim, D. A., et al. (2006). Visual Analytics: Definition, Process, and Challenges. Information Visualization, 5(4), 240-254.

[10] Andrienko, G., & Andrienko, N. (2013). Visual Analytics of Movement. Springer.

[11] Bertini, E., & Lalanne, D. (2009). Surveying the Combinatorial Space of Visualization Techniques. IEEE Transactions on Visualization and Computer Graphics, 15(3), 445-462.

[12] Ratcliffe, J. H. (2016). Intelligence-Led Policing. Routledge.

[13] UC Irvine Machine Learning Repository. (2022). Crime Dataset. Retrieved from https://archive.ics.uci.edu/ml/datasets/Communities+and+Crime.

[14] Wu, Y., & Zhang, S. (2017). Crime Hotspot Prediction Using Random Forests. Journal of Crime Analysis and Prevention, 9(4), 210-225.

[15] Wang, H., et al. (2019). Machine Learning for Crime Analysis: A Systematic Review. ACM Computing Surveys, 52(5), 1-36.