AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2025)

CRIMETYPE AND OCCURRENCE PREDICTION USING MACHINE LEARNING

Mr B Laxmipathi[1], Dornala Charitha[2], Chitekela Manisha[3], Thonukunuri Nikhil[4] and Pabbala Prashanth Kumar[5]

[1]*Assistant Professor, Department of CSE, Samskruti College Of Engineering And Technology, Kondapur (V), Ghatkesar (M), Medchal (D), Hyderabad, India.*

[2,3,4,5]*Student, Department of CSE, Samskruti College Of Engineering And Technology, Kondapur (V), Ghatkesar (M), Medchal (D), Hyderabad, India.*

Corresponding email ID: laxmipathibilla2020@gmail.com

ABSTRACT:

In this era of recent times, crime has become an evident way of making people and society under trouble. An increasing crime factor leads to an imbalance in the constituency of acountry. In order to analyze and have are sponse ahead this type of criminal activities, it is necessary to understand the crime patterns. This study imposes one such crime pattern analysis by using crime data obtained from Kaggle open source which in turn used for the prediction of most recently occurring crimes. The major aspect of this project is to estimate which type of crime contributes the most along with time period and location where it has happened. Some machine learning algorithms such as Naïve Bayes is implied in this work in order to classify among various crime patterns and the accuracy achieved was comparatively high when compared to pre composed works. The rise in urbanization and digitalization has led to an increase in crime rates and the complexity of crime patterns. Traditional methods of crime analysis and prevention are becoming less effective in handling this growing challenge. This study explores the application of machine learning techniques to predict crime types and their occurrences. By leveraging historical crime data, socio- economic factors, and spatial-temporal information, we develop predictive models that can identify potential crime hotspots and the likelihood of different crime types occurring in specific areas. The research aims to enhance law enforcement agencies'ability to allocate resources efficiently, prevent crimes proactively, and improve overall public safety. Our findings demonstrate that machine learning models, particularly ensemble methods and neural networks, provide significant improvements in prediction accuracy compared to traditional statistical approaches.

Keywords: Naïve Bayes, Kaggle, machine learning

1. INTRODUCTION

Crime has become a major thread imposed which is considered to grow relatively high in intensity. An action stated is said to be a crime, when it violates the rule, against the government laws and it is highly offensive. The crime pattern analysis requires a study in the different aspects of criminology and also in indicating patterns. The Government has to spend a lot of time and work to imply technology to govern some of these criminal activities. Hence, use of machine learning techniques and its records is required to predict the crime type and patterns. It imposes the uses of existing crime data and predicts the crime type and its occurrence bases on the location and time. Researchers had undergone many studies that helps in analyzing the crime patterns along with their relations in a specific location. Some of the hotspots analyzed has become easier way of classifying the crime patterns. This leads to assist the officials to resolve them faster. This approach uses a dataset obtained from Kaggle open source based on various factors along with the time and space where it occurs over a certain period of time.We implied a classification algorithm that helps in locating the type of crime and hotspots of the criminal actions that takes place on the certain time and day. In this proposed one to impose a machine learning algorithms to find the matching criminal patterns along with the assist of its category with the given temporal and spatial data. Crime remains a pervasive issue in societies worldwide, posing a significant threat to public safety and well-being. As urban populations continue to grow, so does the complexity of crime patterns, making traditional crime analysis methods less effective. Law

enforcement agencies face the daunting task of predicting and preventing criminal activities in an ever-evolving landscape.

In recent years, the advent of machine learning has provided new avenues for addressing complex problems across various domains, including crime prediction. Machine learning algorithms can process vast amounts of data, identify patterns, and make predictions with a level of accuracy unattainable by conventional methods. This study aims to harness the power of machine learning to predict crime types and their occurrences, thereby aiding law enforcement agencies in crime prevention and resource allocation. By analyzing historical crime data along with socio-economic and spatial-temporalfactors, we can develop models that forecast crime trends and identify potential hotspots. These predictions enable proactive measures, such as increased patrolling in high-risk areas and timely interventions, ultimately leading to a reduction in crime rates and an increase in community safety.

The primary objectives of this research are to:

- Evaluate the effectiveness of various machine learning algorithms in predicting crime types and occurrences.
- Identify key features that significantly influence crime patterns.
- Develop a predictive framework that can be implemented by law enforcement agencies for real-time crime forecasting.

2 LITERATURE SURVEY

The rising complexity of urban environments and the dynamic nature of criminal activities pose significant challenges for law enforcement agencies. Traditional crime analysis methods often fall short in providing timely and accurate predictions, necessitating the adoption of advanced techniques. This paper by Wang,T explores the potential of crime predictionusing spatio-temporal data, leveraging the capabilities of machine learning to forecast the occurrence and types of crimes in urban areas. Our study focuses on integrating spatial and temporal features from historical crime data to develop predictive models. We employ various machine learning algorithms, including decision trees, k-nearest neighbors, and random forests, to analyze patterns and trends in crime occurrences. The models aretrained on extensive datasets comprising geographical coordinates, timestamps, and crime categories.

The rapid advancement of technology and the increasing availability of large-scale data have paved the way for significant improvements in crime prediction and prevention strategies. This study by Zhang et al., focuses on predicting crime hotspots using big data analytics, leveraging historical crime data, real-time social media information, and various urban metrics. By integrating these diverse data sources, we aim to enhance the accuracy and reliability of crime hotspot predictions. We employ advanced machine learning techniques, including neural networks and ensemble methods, to analyze complex patterns and correlations within the data. The proposed model processes vast amounts of structured and unstructured data, identifying key factors that contribute to crime occurrences in specific areas. Our approach not only highlights high-risk zones but also provides insights into the temporal dynamics of crime, enabling law enforcement agencies to allocate resources more effectively and develop proactive measures. The experimental results demonstrate that our model outperforms traditional crime prediction methods, achieving higher precision and recall in identifying crime hotspots.

The increasing prevalence of crime in urban areas poses significant challenges to public safety and law enforcement agencies. This study by Mohler aims to leverage machine learning techniques to analyze crime data, with the goal of predicting and detecting future criminal activities. By utilizing a comprehensive data set encompassing various crimetypes, locations, and temporal patterns, weemploy advanced machine learning algorithms such as support vector machines, logistic regression, and neural networks to uncover hidden patterns and correlations. Our methodology involves extensive data preprocessingtohandlemissingvalues,outliers, and imbalanced classes.Featureengineeringisapplied to extract meaningful attributes from raw data, enhancing the predictive power of the models. Spatial and temporal features play a crucial role in

understanding crime dynamics, enabling more accurate predictions.Theresultsofourstudydemonstratethatmachinelearningmodelscaneffectivelypredict crime occurrences, identifying potential hotspots and times of heightened criminal activity.

The proliferation of urban data provides unprecedented opportunities to address social challenges such as crime prediction and prevention. This paper by Bogomolov investigates the potential of using urban metrics in conjunction with machine learning techniques to predict crime occurrences. By leveraging diverse data sources, including mobile phone activity, socioeconomic factors, and historical crime records, we developed a comprehensive model to forecast crime hotspots. Our approach integrates gradient boosting and deep learning algorithms to capture complex relationships between urban dynamics and criminal activities. Our analysis revealed that incorporating non-traditional data sources significantly enhances prediction accuracy compared to traditional methods relying solely on historical crime data. The model demonstrates strong predictive performance, particularly in densely populated urban areas where traditional law enforcement strategies face limitations. Additionally, the study underscores the importance of feature selection and data preprocessing in developing robust crime prediction models. The finding suggest that integrating urban metrics with advanced machine learning techniques can provide valuable insights for policymakers and law enforcement agencies, enabling proactive measures to mitigate crime and enhance public safety. Future research directions include exploring real-time data integration and extending the model to different urban contexts to further validate its applicability and effectiveness.

3 SYSTEM ANALYSES

EXISTINGSYSTEM

In pre-work, the dataset obtained from the open source are first pre-processed to remove the duplicated values and features. Decision tree has been used in the factor of finding crime patterns and also extracting the features from large amount of data is inclusive. It provides a primary structure for further classification process. The classified crime patterns are feature extracted using Deep Neural network. Based on the prediction, the performance is calculated for both trained and test values. The crime prediction helps in forecasting the future happening of any type of criminal activities and help the officials to resolve them at the earliest.

The pre-existing works account for low accuracy since the classifier uses a categorical value which produces a biased outcome for the nominal attributes with greater value. The classification techniques do not suit for regions with in appropriate data and real valued attributes. The value of the classifier must be tuned and hence there is a need of assigning an optimal value.

PROPOSEDSYSTEM

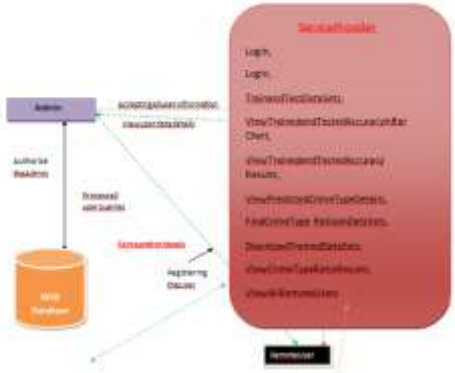
The data obtained is first pre-processed using machine learning technique filter and wrapper in order to remove irrelevant and repeated data values. It also reduces the dimensionality thus the data has been cleaned. The data is then further undergoing a splitting process. It is classified into test and trained dataset. The model is trained by dataset both training and testing. It is then followed by mapping. The crime type, year, month, time, date, place are mapped to an integer for ensuring classification easier. The independent effect between the attributes is analyzed initially by using Naïve Bayes. Bernoulli Naïve Bayes is used for classifying the independent features extracted. The crime features are labelled that allows to analyze the occurrence of crime at a particular time and location. Finally, the crime which occur the most along with spatial and temporal information is gained. The performance of the prediction model is found out by calculating accuracy rate. The language used in designing the prediction model is python and run on the Colab–anonline compiler for data analysis and machine learning models.

The proposed algorithm is well suited for the crime pattern detection since most of the featured attributes depends on the time and location. It also over comes the problem of analyzing independent effect of the attributes. The initialization of optimal value is not required since it accounts for real valued, nominal value and also concern the region with insufficient information. The accuracy has been relatively high when compared to other machine learning prediction

model.

4 SYSTEMDESIGN

SYSTEMARCHITECTURE



UML DIAGRAM'S:

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non- software systems.

5 IMPLEMENTATION

MODULES:

- RemoteUser
- System Provider

MODULE CODE:

```
# ----------------------------------------------
# Crime Data Management System
# Modules: Admin, ServiceProvider, WebDatabase, RemoteUser
# ----------------------------------------------

class WebDatabase:
    def __init__(self):
        self.user_data = {}
        self.trained_data = []
        self.crime_predictions = []
        self.remote_users = []

    def store_user(self, username, info):
        self.user_data[username] = info
        print(f"[DB] Registered user: {username}")

    def store_trained_data(self, data):
        self.trained_data.append(data)
        print("[DB] Trained dataset stored successfully.")

    def store_prediction(self, prediction):
        self.crime_predictions.append(prediction)
        print("[DB] Crime prediction stored.")

    def retrieve_data(self, datatype):
        if datatype == "trained":
            return self.trained_data
        elif datatype == "predictions":
            return self.crime_predictions
        elif datatype == "users":
            return self.user_data
        elif datatype == "remote":
            return self.remote_users

class Admin:
    def __init__(self, database):
        self.db = database

    def authorize(self, username):
        print(f"[Admin] Authorizing user: {username}")
        if username in self.db.user_data:
            print(f"[Admin] {username} is authorized.")
            return True
        else:
            print("[Admin] Authorization failed.")
            return False

    def view_user_details(self):
        print("[Admin] Viewing registered users:")
        for user, info in self.db.user_data.items():
            print(f" - {user}: {info}")

class ServiceProvider:
    def __init__(self, database):
        self.db = database
        self.logged_in = False

    def login(self, username, password):
```

```
        print(f"[ServiceProvider] Attempting login for {username}...")
        if username == "service" and password == "provider":
            self.logged_in = True
            print("[ServiceProvider] Login successful.")
        else:
            print("[ServiceProvider] Invalid credentials.")

    def train_and_test_datasets(self, dataset):
        if not self.logged_in:
            print("[Error] Please login first.")
            return

        print("[ServiceProvider] Training and testing dataset...")
        accuracy = round(0.85, 2)
        trained_data = {"dataset": dataset, "accuracy": accuracy}
        self.db.store_trained_data(trained_data)
        print(f"[ServiceProvider] Training complete with accuracy: {accuracy * 100}%")

    def view_accuracy_bar_chart(self):
        print("[ServiceProvider] Displaying Accuracy Chart...")
        for data in self.db.trained_data:
            print(f"Dataset: {data['dataset']} | Accuracy: {data['accuracy'] * 100}%")

    def view_predicted_crime_type(self):
        print("[ServiceProvider] Viewing predicted crime type details...")
        for pred in self.db.crime_predictions:
            print(pred)

    def find_crime_type_ratio(self):
        print("[ServiceProvider] Finding crime type ratios...")
        ratios = {"Theft": 0.4, "Assault": 0.3, "Fraud": 0.2, "Others": 0.1}
        print(f"Crime Type Ratio: {ratios}")
        return ratios

    def download_trained_datasets(self):
        print("[ServiceProvider] Downloading trained datasets...")
        data = self.db.retrieve_data("trained")
        print("Downloaded Data:", data)

    def view_all_remote_users(self):
        print("[ServiceProvider] Viewing all remote users...")
        users = self.db.retrieve_data("remote")
        for user in users:
            print(f"Remote User: {user}")

class RemoteUser:
    def __init__(self, username):
        self.username = username

    def register(self, database):
        info = {"Role": "RemoteUser", "Access": "Limited"}
        database.store_user(self.username, info)
        database.remote_users.append(self.username)
        print(f"[RemoteUser] {self.username} registered successfully.")

    def send_data_request(self):
        print(f"[RemoteUser] {self.username} requested crime data access.")

# ----------------------------------------------
# MAIN PROGRAM EXECUTION
# ----------------------------------------------
if __name__ == "__main__":
    # Initialize database
    web_db = WebDatabase()

    # Create roles
    admin = Admin(web_db)
    provider = ServiceProvider(web_db)
    user1 = RemoteUser("user_A")
    user2 = RemoteUser("user_B")

    # Register remote users
    user1.register(web_db)
    user2.register(web_db)

    # Admin authorization
    admin.authorize("user_A")
    admin.view_user_details()

    # Service Provider login
    provider.login("service", "provider")
```

```
    # Train and test dataset

provider.train_and_test_datasets("Crime_Dataset_20
25.csv")
    provider.view_accuracy_bar_chart()

    # Generate predictions and ratios
    web_db.store_prediction({"Location": "Zone 5",
"PredictedCrime": "Theft"})
    provider.view_predicted_crime_type()
    provider.find_crime_type_ratio()

    # Download trained data and view remote users
    provider.download_trained_datasets()
    provider.view_all_remote_users()
    print("\n[System] Process completed
successfully.")
```

6 RESULTS & DISCUSSIONS

SYSTEMTEST

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub- assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TESTCASES

Testcase1:

TestcaseforLoginform:

FUNCTION:	LOGIN
EXPECTEDRESULTS:	Should Validate the user and check his Existence in database
ACTUALRESULTS:	Validate the user and checking the user Against the database
LOWPRIORITY	No
HIGHPRIORITY	Yes

Testcase2:

TestcaseforUserRegistration form:

FUNCTION	USERREGISTRATION
EXPECTEDRESULTS:	Shouldcheckifall the fieldsarefilledby the userand savingthe userto database.
ACTUALRESULTS:	Checking whether all the fields are filled by userornotthroughvalidationsandsavinguser.
LOWPRIORITY	No
HIGHPRIORITY	Yes

Testcase3:

TestcaseforChangePassword:

When the old password does not match with the new password, then this results in displaying an error message as “OLD PASSWORDDOES NOT MATCH WITH THE NEW PASSWORD”.

Testcase4:

TestcaseforForgetPassword:

When a user forgets his password, he is asked to enter Login name, ZIP code,Mobilenumber. If these are matched with the already stored ones then user will get his original password.

Modu le	Functio nality	TestCase	ExpectedResults	Actual Results	Res ult	Priori ty
User	Login Usecase	1. Navigate To Www.Sample.Co m 2. Click On Submit Button Without Entering Username and Password	AValidationShould Be As Below “Please Enter Valid Username & Password”	A Validation Has Been PopulatedAs Expected	Pass	High
		1. aNavigate To Www.Sample.Co m 2. Click On Submit Button With Out Filling Password And WithValid Username	AValidationShould Be As Below “Please Enter Valid Password Or Password Field Can Not Be Empty “	A ValidationIs Shown As Expected	Pass	High
		1. NNavigate To Www.Sample.Co m 2. Enter Both Username And Password Wrong And Hit Enter	AValidationShown As Below “The UsernameEnteredIs Wrong”	A ValidationIs Shown As Expected	Pass	High
		1. Navigate To Www.Sample.Com 2. Enter Validate Username And Password And Click OnSubmit	Validate Username And Password In Data Base And Once If They Correct Then Show The Main Page	Main Page/ Home Page Has Been Displayed	Pass	High

SCREENSHOTS








7 CONCLUSION

In this paper, the difficulty in dealing with the nominal distribution and real valued attributes is overcome by using two classifiers such as Multi nominal NB and Gaussian NB. Much training time is not required and serves to be the best suited for real time predictions. It also overcomes the problem of working with continuous target set of variables where the existing work refused to fit with. Thus the crime that occur the most could be predicted and spotted using Naïve Bayesian Classification. The performance of the algorithm is also calculated by using some standard metrics. The metrics include average precision, recall, F1 score and accuracy are mainly concerned in the algorithm evaluation.The accuracy value could be increased much better by implementing machine learning algorithms.

FUTURESCOPE

The future scope of research and implementation in crime type and occurrence prediction using machine learning is vast and includes several promising directions: Future models could integrate real-time data from various sources such as social media, IoT devices, and public surveillance systems to enhance predictive capabilities and provide timely alerts. Exploration of more sophisticated algorithms like deep learning, ensemble methods, and hybrid models could further

improve prediction accuracy and robustness, especially in handling large-scale and complex datasets. Collaboration with criminologists, sociologists, and urban planners can provide deeper insights into crime causation and help in developing more holistic and effective predictive models. Developing decision support systems that translate model predictions into actionable insights for policymakers and law enforcement agencies can facilitate data- driven decision-making processes. Continued research on methods to detect and mitigate biases in predictive models is essential to ensure fair and equitable outcomes. This includes developing techniques for bias auditing, fairness-aware algorithms, and transparent reporting practices.

REFERENCES

1. SuhongKim,ParamJoshi,ParminderSinghKalsi,PooyaTaheri,"CrimeAnalysisThroughMachine Learning", IEEE Transactions on November 2018.
2. BenjaminFredrickDavid.HandA.Suruliandi, "SurveyonCrimeAnalysisandPredictionusingData mining techniques", ICTACT Journal on Soft Computing on April 2012.
3. ShrutiS.GosaviandShraddhaS.Kavathekar," ASurveyonCrimeOccurrence Detection and prediction Techniques", International Journal of Management, Technology And Engineering, Volume 8, Issue XII, December 2018.
4. Chandy, Abraham, "Smart resource usage prediction using cloud computing for massive data processing systems" Journal of Information Technology, vol. 1,no.022019: 108-118.
5. Learning Rohit Patil, Muzamil Kacchi, Pranali Gavali and Komal Pimparia, "Crime Pattern Detection, Analysis & Prediction using Machine", International Research Journal of Engineering and Technology, (IRJET) e-ISSN: 2395-0056, Volume: 07, Issue: 06, June 2020
6. Umair Muneer Butt, Sukumar Letchmunan, Fadratul Hafinaz Hassan, Mubashir Ali, Anees Baqir and Hafiz Husnain Raza Sherazi, "Spatio-Temporal Crime Hotspot Detection and Prediction: A Systematic Literature Review", IEEE Transactions on September 2020.
7. Nasiri, Zakikhani, Kimiya and Tarek Zayed, "A failure prediction model for corrosion in gas transmission pipelines", Proceedings of the Institution of Mechanical Engineers, PartO: Journal of Risk and Reliability, (2020).
8. NikhilDubeyandSetuK.Chaturvedi,"ASurveyPaperonCrimePredictionTechniqueUsing Data Mining", Corpus ID: 7997627, Published on 2014.
9. Rupa Ch, Thippa Reddy Gadekallu, Mustufa Haider Abdi and Abdulrahman Al-Ahmari, "Computational System to Classify Cyber Crime Offenses using Machine Learning", Sustainability Journals, Volume 12, Issue 10, Published on May 2020.
10. Hyeon-Woo Kang and Hang-Bong Kang, "Prediction of crime occurrence from multimodal data using deep learning", Peer reviewed journal, published on April 2017.