A Project Report on

# CRIME ANALYSIS AND PREDICTION

A Dissertation submitted to JNTU Hyderabad in partial fulfillment of the academic requirements for the award of the degree.

## Bachelor of Technology

## in

## Computer Science and Engineering

Submitted by

M Guru Sai Chawan
(20H51A0517)
T Manohar
(20H51A05D3)
M Meghana
(20H51A05P5)

Under the esteemed guidance of
Mr. M. Shiva kumar
(Assistant Professor)

**Department of Computer Science and Engineering**

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY**
(UGC Autonomous)
*Approved by AICTE *Affiliated to JNTUH *NAAC Accredited with A$^+$ Grade
KANDLAKOYA, MEDCHAL ROAD, HYDERABAD - 501401.

**2020- 2024**

# CMR COLLEGE OF ENGINEERING & TECHNOLOGY

KANDLAKOYA, MEDCHAL ROAD, HYDERABAD – 501401

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## CERTIFICATE

This is to certify that the Major Project report entitled **" Crime Analysis and Prediction "** being submitted by M**.** Guru Sai Chawan (20H51A0517), T. Manohar (20H51A05D3), M. Meghana (20H51A05P5) in partial fulfillment for the award of **Bachelor of Technology in Computer Science and Engineering** is a record of bonafide work carried out under my guidance and supervision.

The results embodies in this project report have not been submitted to any other University or Institute for the award of any Degree.

**Mr. M. Shiva Kumar**
**Assistant Professor**
**Dept. of CSE**

**Dr. Siva Skandha Sanagala**
**Associate Professor and HOD**
**Dept. of CSE**

**EXTERNAL EXAMINER**

# ACKNOWLEDGEMENT

With great pleasure we want to take this opportunity to express our heartfelt gratitude to all the people who helped in making this project a grand success.

We are grateful to **Mr. M. ShivaKumar,** Assistant Professor, Department of Computer Science and Engineering for his valuable technical suggestions and guidance during the execution of this project work.

We would like to thank**, Dr. Siva Skandha Sanagala,** Head of the Department of Computer Science and Engineering, CMR College of Engineering and Technology, who is the major driving forces to complete our project work successfully.

We are very grateful to **Dr. Ghanta Devadasu**, Dean-Academics, CMR College of Engineering and Technology, for his constant support and motivation in carrying out the project work successfully.

We are highly indebted to **Major Dr. V A Narayana,** Principal, CMR College of Engineering and Technology, for giving permission to carry out this project in a successful and fruitful way.

We would like to thank the **Teaching & Non- teaching** staff of Department of Dept Name for their co-operation

We express our sincere thanks to **Shri. Ch. Gopal Reddy**, Secretary& Correspondent, CMR Group of Institutions, and Shri Ch Abhinav Reddy, CEO, CMR Group of Institutions for their continuous care and support.

Finally, we extend thanks to our parents who stood behind us at different stages of this Project. We sincerely acknowledge and thank all those who gave support directly or indirectly in completion of this project work.

| | |
|---|---|
| M Guru Sai Chawan | 20H51A0517 |
| T Manohar | 20H51A05D3 |
| M Meghana | 20H51A05P5 |

# TABLE OF CONTENTS

## List of Figures

# ABSTRACT

Crime analysis and prediction represent a crucial approach in identifying and addressing criminal activities systematically. By leveraging data mining techniques, this system can discern patterns and extract valuable insights from unstructured data, thereby predicting regions with a higher likelihood of crime occurrences and visualizing areas prone to criminal activities. The extraction of previously unknown information from existing datasets facilitates the anticipation of emerging crime trends, enhancing the effectiveness of law enforcement efforts. Given the pervasive impact of crimes on societal well-being, economic prosperity, and national reputation, the development of advanced systems and innovative methodologies for crime analytics is imperative to safeguard communities effectively.

Various types of criminal analysis and crime prediction can be achieved through the application of diverse data mining techniques. These include but are not limited to association rule mining, clustering analysis, and classification algorithms. Association rule mining uncovers relationships between different variables, aiding in identifying factors contributing to criminal behaviors. Clustering analysis helps in segmenting geographical regions based on similarities in crime patterns, facilitating targeted interventions. Classification algorithms predict the likelihood of specific crimes occurring in a given area based on historical data, empowering law enforcement agencies to proactively allocate resources and implement preventive measures. By harnessing the power of data mining, law enforcement agencies can enhance their capabilities in analyzing, detecting, and predicting various crime probabilities, thereby fostering safer and more secure communities.

# CHAPTER 1
# INTRODUCTION

# CHAPTER 1
# INTRODUCTION

## 1.1. Problem Statement

As technology continues to advance, criminals are increasingly equipped with sophisticated tools and methods, resulting in a concerning rise in crime rates across various categories. Notably, crimes such as burglary and arson are on the uptick, paralleled by a disturbing surge in severe offenses like murder, sexual assault, abuse, and gang-related activities. Recognizing the urgency of this issue, our project on "crime analysis and prediction" endeavours to harness the power of advanced computer techniques to confront and mitigate the impact of criminal activities.

To address the multifaceted challenge of rising crime rates, we have embarked on a comprehensive initiative to collect and analyze crime data sourced from diverse platforms including blogs, news outlets, and websites. By aggregating this data into a centralized repository, we aim to construct a robust database of crime reports. Subsequently, leveraging sophisticated computer algorithms and techniques, such as machine learning and time series forecasting, our project seeks to uncover underlying patterns and trends within the data. Through this analytical process, we endeavor to empower law enforcement agencies with actionable insights to enhance their capabilities in crime prevention and detection.

Ultimately, our goal is to not only aid in the swift apprehension of criminals but also to proactively identify areas at heightened risk of criminal activity. By leveraging data-driven approaches, we aspire to contribute to the collective efforts aimed at reducing crime rates and fostering safer communities.

## 1.2. Research Objective

This research Project aims to develop a crime analysis and prediction Model, we utilize Jupyter Notebook as our primary tool for analyzing crime data extracted from Kaggle in CSV format. Our research focuses on preprocessing the data, building machine learning models using algorithms available in libraries like scikit-learn, and evaluating their performance on testing datasets. Through visualization and analysis within the Jupyter environment, we aim to identify the most effective models for crime prediction, considering factors such as accuracy and interpretability. Although we do not deploy our models using Flask, we discuss potential applications for our findings in informing law enforcement strategies and proactive crime prevention efforts.

In conclusion, our research demonstrates the effectiveness of leveraging machine learning techniques within the Jupyter Notebook environment for crime prediction tasks. By analyzing crime data and developing predictive models, we contribute to the advancement of strategies aimed at enhancing community safety and combating criminal activities. Our findings provide valuable insights for law enforcement agencies, facilitating informed decision-making and proactive measures to address the challenges posed by crime.

### 1.3. Project Scope and Limitations

**Scope:**

Our project primarily focuses on leveraging machine learning techniques within the Jupyter Notebook environment to analyze crime data sourced from Kaggle. The research study entails preprocessing the collected data, constructing predictive models using various algorithms from libraries like scikit-learn, and assessing their efficacy. The primary aim is to identify optimal models for crime prediction through comprehensive visualization and analysis within the Jupyter environment. While the current focus does not involve deploying models using Flask, future extensions aim to explore more advanced classification algorithms to enhance prediction accuracy and implement enhanced privacy measures for safeguarding the dataset. Additionally, potential directions for further research include incorporating facial recognition technology to predict criminal behavior based on behavioral patterns, contributing to proactive crime prevention strategies.

**Limitations:**

1. Data Quality: The effectiveness of our predictive models heavily relies on the quality and completeness of the crime data obtained from Kaggle. Inaccurate or incomplete data may compromise the reliability of our predictions.
2. Algorithm Selection Bias: The choice of machine learning algorithms utilized in our project may introduce bias and affect the accuracy of crime predictions. Different algorithms may perform better or worse depending on the nature of the crime data and its underlying patterns.
3. Privacy Concerns: While the project aims to enhance public safety, the utilization of sensitive crime data raises privacy concerns. Implementing appropriate measures to protect the privacy and confidentiality of individuals involved in criminal incidents is paramount.
4. Real-world Implementation: Deploying predictive models developed in a controlled

environment to real-world settings may encounter challenges such as integration with existing law enforcement systems, user adoption, and acceptance by stakeholders.

5. Ethical Considerations: The use of predictive analytics in law enforcement raises ethical questions regarding potential biases, fairness, and accountability in decision-making processes. Addressing these ethical considerations is essential for responsible and equitable deployment of predictive crime prediction systems.
6. Computational Resources: The computational resources required for preprocessing large volumes of crime data and training complex machine learning models could pose limitations, particularly for researchers with limited access to high-performance computing resources.

Acknowledging the limitations, we emphasize the ongoing efforts to enhance data quality, mitigate biases, and address privacy concerns, while ensuring transparent communication and ethical considerations in our project.

# CHAPTER 2
# BACKGROUND WORK

# CHAPTER 2

# BACKGROUND WORK

## 2.1 Social Media Analysis for Crime Prediction

### 2.1.1. Introduction

In recent years, the exponential growth of social media platforms has revolutionized how individuals communicate, share information, and express opinions. This digital transformation has not only altered the landscape of social interaction but has also opened new avenues for understanding and addressing societal challenges, including crime prevention. The concept of Social Media Analysis for Crime Prediction has emerged at the intersection of technology, social science, and law enforcement, offering novel approaches to anticipate, mitigate, and respond to criminal activities.

Social media platforms serve as virtual arenas where users openly discuss a wide range of topics, including their experiences with crime, perceptions of safety, and observations of suspicious behavior. These digital conversations, coupled with the abundance of geotagged content, provide valuable insights into the dynamics of criminal activities within communities. By leveraging advanced computational techniques such as natural language processing (NLP), machine learning, and network analysis, researchers and law enforcement agencies can extract actionable intelligence from social media data to enhance crime prediction efforts.

The potential of social media analysis for crime prediction lies not only in its ability to detect ongoing criminal activities but also in its capacity to identify emerging trends and anticipate future threats. By analyzing patterns of user behavior, sentiment, and communication networks, predictive models can forecast the likelihood of various types of crimes occurring in specific locations and timeframes. Moreover, social media analysis enables law enforcement agencies to engage with communities more effectively, fostering collaboration, trust, and proactive measures to prevent crime. However, amidst the promises of this approach, challenges such as data privacy concerns, algorithmic biases, and ethical considerations underscore the need for responsible and transparent practices in its implementation.

### 2.1.2. Merits, Demerits and Challenges

**Merits:**

**Timely Insights:** Social media analysis provides real-time monitoring of public sentiment and emerging trends, enabling proactive crime prevention measures based on up-to-date information.

**Wide Data Availability:** Social media platforms offer a vast amount of publicly available data, facilitating comprehensive analysis for predictive modeling and crime trend identification.

**Cost-effectiveness:** Compared to traditional methods, social media analysis for crime prediction can be more cost-effective, utilizing readily available digital data and automated analytical techniques to inform law enforcement strategies.

**Community Engagement:** By analyzing social media discussions, law enforcement agencies can engage with the community more effectively, fostering trust, cooperation, and collaboration in crime prevention efforts.

**Demerits:**

**Data Privacy Concerns:** Utilizing social media data raises ethical concerns regarding user privacy and data protection, as extracting information may inadvertently expose personal details or sensitive information about individuals.

**Data Quality and Bias**: Social media data may contain noise, misinformation, and biases, affecting the accuracy of predictive models. Biases present in user-generated content can lead to skewed results and erroneous conclusions.

**Legal and Ethical Considerations**: Legal and ethical challenges arise concerning surveillance, data ownership, and freedom of expression when using social media data for law enforcement purposes, requiring careful navigation to ensure compliance and respect for individual rights.

**Algorithmic Fairness**: Predictive models trained on social media data may perpetuate biases, leading to unfair outcomes, particularly for marginalized communities. Addressing algorithmic biases is crucial to uphold fairness and equity in crime prediction efforts.

**Challenges:**

**Data Volume and Complexity:** Managing and analyzing the vast volume and diverse formats of social media data pose challenges, requiring scalable infrastructure, advanced analytical tools, and expertise in data science.

**Verification and Validation**: Ensuring the accuracy and reliability of information extracted from social media data is challenging, requiring robust verification and validation procedures to distinguish between factual information and misinformation.

**Interdisciplinary Collaboration**: Successful implementation requires collaboration between computer scientists, social scientists, and law enforcement professionals to bridge technical expertise with domain knowledge for effective predictive modeling and actionable insights.

**Dynamic Nature of social media**: Social media platforms are dynamic environments with rapidly changing user behavior and content trends, necessitating continuous monitoring and adaptation of predictive models to address emerging threats effectively.

### 2.1.3. Implementation

Implementing social media analysis for crime prediction involves a multifaceted approach encompassing several key steps. Firstly, data collection is paramount, which typically involves retrieving relevant data from social media platforms using APIs or web scraping techniques. This data may encompass various forms of user-generated content, including text posts, images, videos, and geolocation information, providing a rich source of information for analysis. Preprocessing is then undertaken to clean and standardize the collected data, removing noise, handling missing values, and normalizing text through tasks such as sentiment analysis and geocoding.

Following preprocessing, the next step involves feature extraction, where meaningful features are derived from the processed data. This may involve extracting keywords, identifying user behavior patterns, and analyzing temporal trends to uncover valuable insights relevant to crime prediction. Subsequently, predictive models are developed using machine learning algorithms such as classification, clustering, or regression. These models are trained on historical data to predict various crime-related outcomes, such as crime hotspots, types of crime, or characteristics of perpetrators.

Evaluation of the predictive models is crucial to assess their performance and generalization capabilities. Metrics such as accuracy, precision, recall, and F1-score are employed to evaluate the models, validating their efficacy on historical data and assessing their suitability for real-world applications. Finally, the trained models are deployed in operational settings to assist law enforcement agencies in decision-making processes. This may involve integrating the models into existing crime prediction systems, developing standalone applications for real-time monitoring and alerting, or providing actionable insights to inform proactive crime prevention strategies. Overall, successful implementation requires interdisciplinary collaboration between computer scientists, social scientists, and law enforcement professionals, alongside careful consideration of ethical and legal implications to ensure responsible and transparent practices.

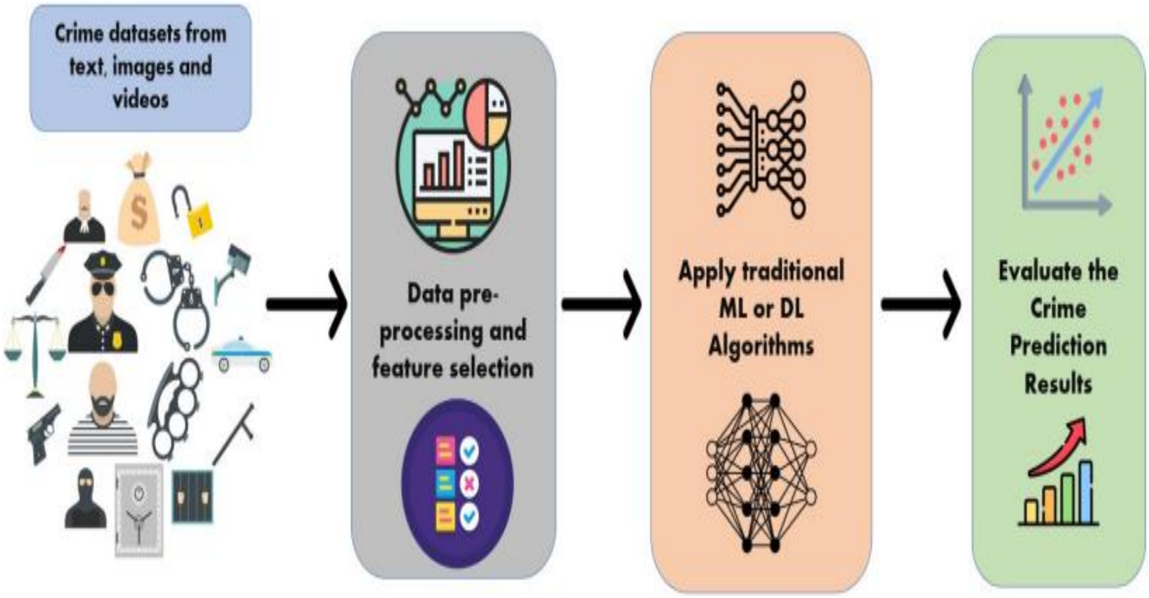

Crime datasets from text, images and videos
Data pre-processing and feature selection
Apply traditional ML or DL Algorithms
Evaluate the Crime Prediction Results

Figure.2.1: The architecture of existing system.

## 2.2. Crime Trend Analysis using ML

### 2.2.1. Introduction

Crime Trend Analysis using Machine Learning (ML) represents a significant advancement in the field of law enforcement and public safety. Leveraging the power of data-driven methodologies, this approach aims to identify, analyze, and predict patterns in criminal activities. By harnessing various machine learning algorithms, such as classification, clustering, and regression, alongside advanced statistical techniques, law enforcement agencies can gain valuable insights into crime trends, contributing to more effective crime prevention and intervention strategies.

The proliferation of digital technologies has led to the accumulation of vast amounts of data related to criminal incidents, including crime reports, witness statements, and demographic information. Crime Trend Analysis using ML capitalizes on this wealth of data to uncover hidden patterns, correlations, and anomalies within the crime data. By analyzing historical crime records and contextual factors such as socio-economic indicators, geographical features, and temporal trends, ML-based approaches enable law enforcement agencies to proactively identify high-risk areas, allocate resources efficiently, and anticipate emerging threats.

Furthermore, Crime Trend Analysis using ML fosters a data-driven approach to decision-making in law enforcement, empowering agencies to prioritize resources, optimize patrol routes, and deploy personnel based on evidence-based insights. By moving beyond traditional reactive methods of crime analysis, ML-based approaches offer the potential to shift towards proactive and preventive measures, ultimately enhancing public safety and community well-being.

### 2.2.2. Merits, Demerits and Challenges

**Merits:**

**Predictive Insights**: ML-based crime trend analysis provides predictive insights into future crime patterns, enabling law enforcement agencies to anticipate and pre-empt criminal activities.

**Data-driven Decision Making**: By leveraging large volumes of historical crime data, ML facilitates evidence-based decision-making in law enforcement, leading to more targeted and effective resource allocation.

**Efficiency and Resource Optimization**: ML algorithms can automate the analysis of crime data, reducing the time and effort required for manual analysis and enabling agencies to optimize resource allocation and patrol strategies.

**Scalability and Adaptability:** ML models can be scalable and adaptable to different types of crime data and geographical regions, making them versatile tools for crime trend analysis across diverse contexts.

**Demerits:**

**Data Quality and Bias**: ML-based crime trend analysis is susceptible to biases and inaccuracies present in the underlying crime data, which can lead to skewed or erroneous predictions. Biases in data collection, reporting practices, and sampling methods may inadvertently perpetuate systemic inequalities and distort the analytical outcomes.

**Ethical and Privacy Concerns**: The utilization of sensitive personal data in crime trend analysis raises ethical concerns regarding privacy, consent, and the potential for misuse of data. Striking a balance between public safety imperatives and individual privacy rights poses complex ethical dilemmas, necessitating robust safeguards and transparent accountability mechanisms.

**Interpretability and Transparency**: ML models employed in crime trend analysis, particularly complex ones like deep learning algorithms, may lack interpretability, making it challenging to comprehend the underlying factors driving predictions. The opacity of black-box algorithms hampers accountability and may undermine public trust in the decision-making process.

**Resource Constraints:** Implementing ML-based crime trend analysis demands substantial investments in data infrastructure, computational resources, and specialized expertise. Small law enforcement agencies with limited resources may face challenges in acquiring, managing, and leveraging the requisite technologies and human capital for effective implementation.

**Challenges:**

**Data Volume and Complexity**: Managing and analyzing large volumes of heterogeneous crime data pose challenges in data processing, storage, and scalability. Addressing data quality issues, integrating disparate data sources, and accommodating evolving data formats necessitate robust data management strategies and scalable analytical frameworks.

**Algorithmic Bias Mitigation**: Mitigating biases inherent in ML algorithms and crime data requires concerted efforts to enhance algorithmic fairness, transparency, and accountability. Employing bias detection techniques, fairness-aware learning algorithms, and diverse training data can help mitigate biases and promote equitable outcomes.

**Ethical and Legal Compliance**: Ensuring ethical and legal compliance in ML-based crime trend analysis entails navigating a complex landscape of regulatory frameworks, privacy laws, and ethical guidelines. Upholding principles of fairness, transparency, and accountability while balancing competing interests poses significant challenges for law enforcement agencies and data practitioners.

**Interdisciplinary Collaboration**: Successful implementation of ML-based crime trend analysis hinges on interdisciplinary collaboration between data scientists, criminologists, policymakers, and community stakeholders. Bridging disciplinary silos, fostering mutual understanding, and co-creating solutions that address societal needs and concerns are essential for the effective deployment of ML-based crime analysis tools.

### 2.2.3. Implementation

Implementing Crime Trend Analysis using Machine Learning (ML) involves a systematic approach encompassing several key phases. Firstly, data acquisition and preprocessing are imperative, involving the collection of diverse crime data sources such as police reports, incident logs, and emergency calls. This data undergoes rigorous preprocessing steps, including data cleaning, normalization, and feature engineering, to ensure consistency and relevance for subsequent analysis.

Following data preparation, the selection and training of ML models are pivotal. Law enforcement agencies must choose suitable ML algorithms, considering the nature of the crime data and the analytical objectives. Training these models on historical crime data enables them to discern patterns, correlations, and anomalies within the data. Various ML techniques, ranging from decision trees to deep learning algorithms, may be employed based on the complexity and nuances of the crime data.

Validation and evaluation of ML models are critical to assess their predictive performance and generalizability. Techniques such as cross-validation and holdout validation are utilized to validate the models, while metrics such as precision, recall, and F1-score gauge their accuracy and robustness. Deploying the trained ML models within operational settings involves integrating them into existing crime analysis workflows and decision support systems. User-friendly interfaces and tools are provided to law enforcement personnel, facilitating interaction with the models and interpretation of results.

Continuous monitoring and iteration are essential components of ML-based crime trend analysis implementation. Law enforcement agencies must monitor the performance of deployed ML models in real-time, collecting feedback from users and stakeholders. Iterative refinement of the models based on new data and evolving crime trends enhances their predictive accuracy and usability over time. Moreover, ongoing training and capacity-building initiatives are crucial to equip law enforcement personnel with the necessary skills and knowledge to leverage ML-based crime analysis tools effectively. Overall, successful implementation requires a collaborative effort between data scientists, law enforcement professionals, policymakers, and community stakeholders, guided by ethical considerations and a commitment to transparency and accountability.

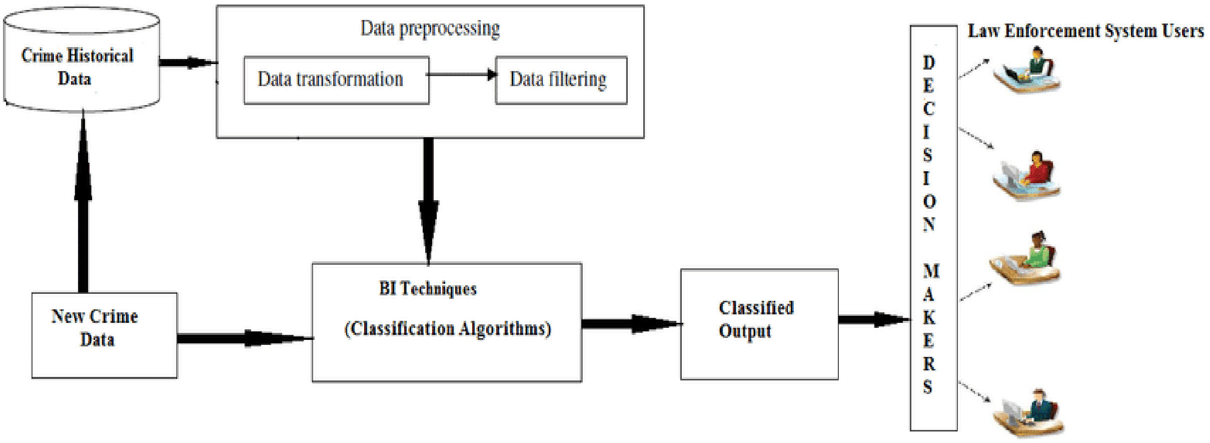


Figure.2.2:crime trend analysis using ML existing system.

## 2.3. Crime prediction using Sentiment Analysis in Police Reports using NLP

### 2.3.1. Introduction

Crime prediction using Sentiment Analysis in Police Reports using Natural Language Processing (NLP) represents a cutting-edge application of computational linguistics in law enforcement. In an era where vast volumes of textual data are generated daily, the digitization of police reports offers a rich source of information ripe for analysis. Leveraging NLP techniques, particularly Sentiment Analysis, enables law enforcement agencies to delve deeper into the linguistic nuances of these reports, uncovering implicit sentiments and attitudes embedded within the text. By deciphering the underlying emotions expressed in police narratives, this approach seeks to illuminate potential patterns and correlations indicative of criminal activities, thereby empowering authorities with proactive tools for crime prediction and prevention.

The advent of Sentiment Analysis presents a paradigm shift in crime analysis methodologies, augmenting traditional approaches with data-driven insights derived from textual data. By discerning the sentiment conveyed within police reports—whether positive, negative, or neutral—law enforcement agencies gain access to a wealth of contextual information that transcends mere statistical data. This granular understanding of the emotional undercurrents surrounding criminal incidents allows for a more nuanced assessment of risk factors and threat levels, enabling agencies to tailor their response strategies accordingly. Moreover, by adopting a proactive stance towards crime prevention, informed by sentiment analysis findings, law enforcement agencies can anticipate emerging threats and allocate resources preemptively, fostering safer communities and more effective public safety initiatives.

### 2.3.2. Merits, Demerits and Challenges

**Merits:**

**Granular Insights:** Sentiment Analysis furnishes nuanced insights into the emotions and attitudes conveyed within police reports, enabling a deeper understanding of the contextual factors surrounding criminal incidents. This granular understanding empowers law enforcement agencies to tailor intervention strategies more effectively.

**Proactive Intervention:** By detecting fluctuations in sentiment trends within police reports, law enforcement agencies can establish proactive intervention mechanisms to identify potential crime hotspots or emerging threats. This early warning system facilitates timely responses and preemptive measures to mitigate criminal activities.

**Data-driven Decision Making**: Sentiment Analysis facilitates data-driven decision-making by transforming unstructured textual data from police reports into quantifiable metrics. These metrics provide actionable intelligence, enabling law enforcement agencies to optimize resource allocation and operational strategies based on empirical evidence.

**Resource Optimization:** Prioritizing resources based on sentiment analysis findings allows law enforcement agencies to allocate personnel and resources more efficiently. By focusing resources on areas with heightened risk or negative sentiment indicators, agencies can maximize the impact of their interventions and enhance public safety outcomes.

**Demerits:**

**Data Quality Concerns**: The reliability and accuracy of sentiment analysis in police reports may be compromised by inconsistencies or inaccuracies in the data. Factors suchas varying reporting practices, linguistic nuances, and subjective interpretations can introduce noise and bias into the analysis, affecting the validity of predictive outcomes.

**Contextual Understanding**: Sentiment analysis algorithms may struggle to capture the contextual nuances and complexities inherent in police reports. Ambiguous language, sarcasm, or cultural references may confound the algorithms, leading to misinterpretations or inaccurate sentiment classifications.

**Bias and Fairness**: Sentiment analysis models may inadvertently perpetuate biases present in the data, leading to disparities in predictive outcomes. Factors such as demographic biases, linguistic biases, and cultural biases can influence the sentiment analysis results, potentially resulting in inequitable or unfair predictions.

**Ethical Considerations**: Analyzing sensitive textual data from police reports raises ethical concerns regarding privacy, consent, and data protection. Law enforcement agencies must navigate these ethical considerations carefully, ensuring compliance with legal regulations and ethical guidelines to safeguard the privacy rights of individuals mentioned in the reports.

**Challenges:**

**Data Volume and Complexity:** Managing and analyzing large volumes of textual data from police reports pose challenges in data processing and analysis. The sheer volume and complexity of the data require robust NLP techniques, scalable infrastructure, and computational resources to derive meaningful insights effectively.

**Algorithmic Interpretability**: Interpreting the results of sentiment analysis algorithms and understanding the rationale behind their predictions can be challenging. Complex machine learning models, particularly deep learning algorithms, may lack interpretability, hindering the ability to scrutinize and validate the analysis effectively.

**Contextual Understanding:** Achieving a nuanced understanding of the context surrounding criminal incidents in police reports is essential for accurate sentiment analysis. Incorporating domain-specific knowledge, understanding linguistic nuances, and adapting to evolving language trends pose challenges for sentiment analysis algorithms.

**Resource Constraints**: Implementing sentiment analysis in police reports requires investments in data infrastructure, computational resources, and expertise. Small law enforcement agencies with limited resources may face challenges in acquiring, managing, and leveraging the requisite technologies and human capital for effective implementation.

### 2.3.3. Implementation

Implementing crime prediction using Sentiment Analysis in police reports involves a systematic approach encompassing several key steps. Initially, data acquisition and preprocessing are pivotal, encompassing the collection of textual data from police reports and preprocessing it to enhance readability and consistency. This may involve techniques such as text normalization, tokenization, and removal of stop words to prepare the data for sentiment analysis.

Following data preprocessing, Sentiment Analysis techniques are applied to classify the sentiment expressed in police reports as positive, negative, or neutral.

These techniques may include lexicon-based approaches, machine learning algorithms, or deep learning models trained on labeled datasets. The sentiment analysis results provide insights into the emotional context surrounding criminal incidents, facilitating a deeper understanding of the underlying dynamics.

Subsequently, predictive modeling is employed to forecast future crime trends or identify high-risk areas based on sentiment analysis findings. This involves feature extraction from the sentiment analysis results, such as sentiment scores, sentiment trends over time, or sentiment distributions across different categories of crimes. Machine learning algorithms, such as regression, classification, or clustering, are then utilized to develop predictive models trained on historical data. These models enable law enforcement agencies to anticipate emerging threats, allocate resources efficiently, and implement targeted intervention strategies to prevent crime. Continuous monitoring and evaluation of the predictive models ensure their effectiveness and reliability in real-world applications, supporting evidence-based decision-making in law enforcement operations.

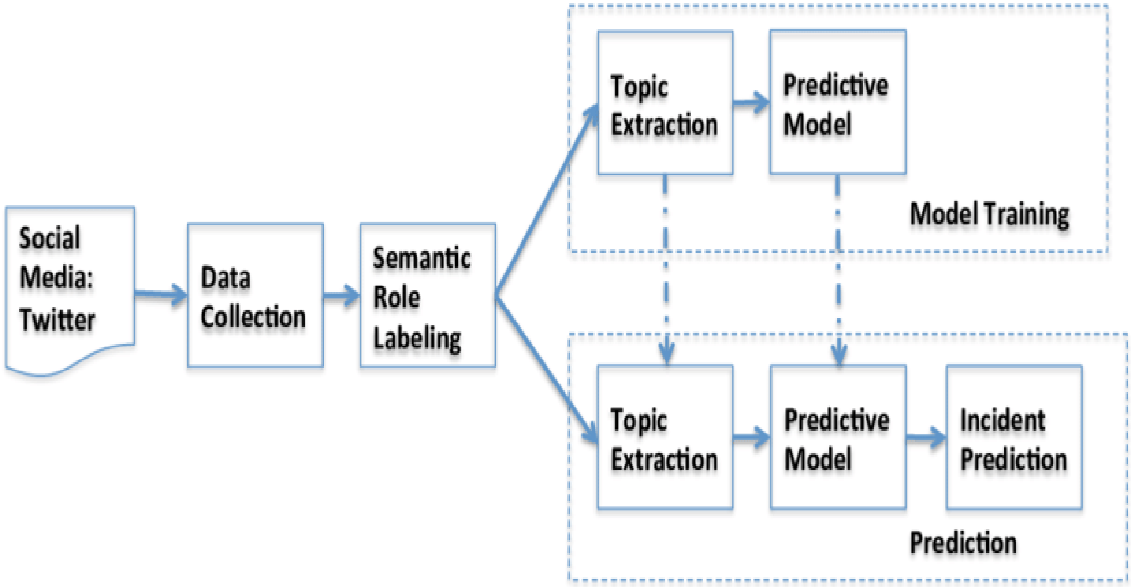


Figure.2.3.1: NLP crime Analysis and prediction block diagram

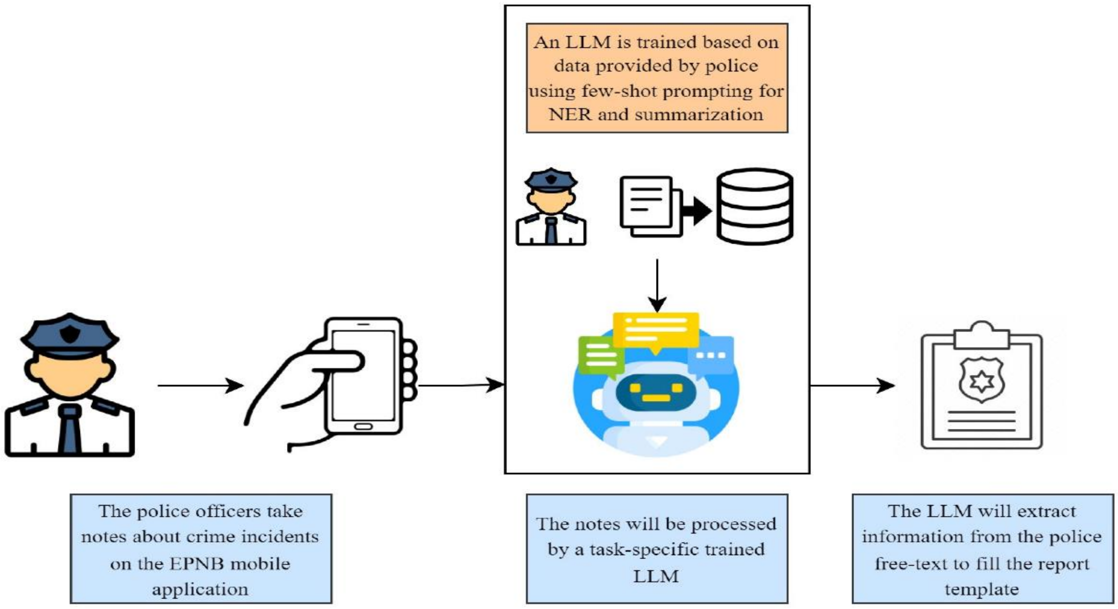

An LLM is trained based on data provided by police using few-shot prompting for NER and summarization
The police officers take notes about crime incidents on the EPNB mobile application
The notes will be processed by a task-specific trained LLM
The LLM will extract information from the police free-text to fill the report template

Figure.2.3.2: NLP Crime analysis and prediction model overview diagram

# CHAPTER 3
## PROPOSED SYSTEM

# CHAPTER 3

## 3.1. Objective of Proposed Model:

The main objective of the proposed model for crime analysis and prediction using machine learning is to develop a predictive system that can effectively analyze historical crime data and generate accurate forecasts regarding future criminal activities. This model aims to leverage various machine learning algorithms and techniques to identify patterns, trends, and correlations within the data to predict when and where crimes are likely to occur. By doing so, law enforcement agencies and policymakers can allocate resources more efficiently, deploy preventive measures proactively, and ultimately work towards reducing crime rates and enhancing public safety. Additionally, the model may also assist in identifying high-risk areas or demographics, understanding underlying factors contributing to criminal behavior, and evaluating the effectiveness of intervention strategies.

In this project, we will be using the technique of machine learning for crime prediction of crime data sets. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training the model, data preprocessing will be done following this feature selection and scaling will be done so that the accuracy obtained will be high. Here We are Using the different classical time Series forecasting Methods like AR (Auto Regression), SARIMA (Seasonal Autoregressive Integrated Moving Average). The K-Nearest Neighbor (KNN) classification and various other algorithms (Decision Tree and Random Forest) will be tested for crime and propose one with better query-based use for training.

Overall, the main objective of the proposed crime analysis and prediction model using machine learning is to leverage data-driven insights to enhance law enforcement efforts, improve resource allocation, and ultimately contribute to the reduction of crime rates and the promotion of public safety.

## 3.2. Algorithms Used for Proposed Model:

The proposed model utilizes K-Nearest Neighbor, Time Series forecasting-SARIMA and Prophet, Random Forest, K-means Clustering, Principle Component Analysis.

**3.2.1. K-Nearest Neighbor (KNN):** K-Nearest Neighbors (KNN) is a simple and intuitive machine learning algorithm used for both classification and regression tasks. KNN is easy to understand and implement, making it a popular choice for beginners in machine learning. However, its performance can be sensitive to the choice of the value of 'k' and the distance metric used. Additionally, KNN can be computationally expensive, especially with large datasets, as it requires calculating distances to all points in the dataset for each prediction.

Here is how it works:

- **Select a value for K**: This is the number of nearest neighbors that the algorithm will consider when making a prediction.
- **Calculate distances**: Measure the distance between the new point and all other points in the dataset. The most common distance metric used is Euclidean distance, but other metrics like Manhattan distance can also be used.
- **Find the K nearest neighbors**: Identify the 'k' points in the dataset that are closest to the new point based on the calculated distances.
- **Majority voting (for classification) or averaging (for regression):** For classification tasks, the algorithm assigns the class label that is most common among the 'k' nearest neighbors. For regression tasks, it calculates the average value of the target variable among the 'k' nearest neighbors.
- **Make a prediction**: Assign the predicted class or value to the new point based on the results of the majority voting or averaging.

### ❖ SARIMA and Prophet

**SARIMA** (Seasonal Autoregressive Integrated Moving Average): It is specifically designed to handle time series data that exhibit seasonal patterns. SARIMA incorporates four key components:

- **Seasonal Component**: SARIMA includes a seasonal component that captures periodic fluctuations in the data. This component allows the model to account for recurring patterns that occur at regular intervals, such as daily, weekly, or yearly seasonality.
- **Autoregressive (AR) Component**: The AR component models the relationship between the current observation and its past values.
- **Integrated (I) Component:** The integrated component represents the differencing of the time series data to make it stationary

- **Moving Average (MA) Component**: It helps capture short-term fluctuations and irregularities in the data.

**SARIMA** is defined by several parameters, including p, d, q, P, D, Q, and m, which control the non-seasonal and seasonal components of the model. Tuning these parameters is crucial for achieving accurate forecasts with SARIMA.

**Prophet**: Prophet is a forecasting model developed by Facebook's Core Data Science team. It is designed to handle time series data with strong seasonal patterns, multiple seasonality, and holiday effects. Prophet offers several key features:

- **Seasonality Detection**: Prophet automatically detects and handles seasonal patterns in the data, including daily, weekly, and yearly seasonality. This allows the model to capture complex seasonal effects without manual intervention.

- **Trend Modeling**: Prophet captures both short-term fluctuations and long-term trends in the data. It uses a piecewise linear model to fit the trend, allowing for changes in slope over time.

- **Holiday Effects:** Prophet allows users to specify holidays and events that may impact the time series data. It includes built-in support for holidays in various countries and regions, as well as the option to define custom holidays. This feature enables the model to account for irregularities in the data caused by holidays and special events.

- **Uncertainty Estimates:** Prophet provides uncertainty estimates for the forecasts, allowing users to assess the reliability of the predictions. It generates uncertainty intervals around the forecasted values, indicating the range within which the true values are likely to fall.

  SARIMA and Prophet are two powerful forecasting models that are widely used for time series analysis and prediction. SARIMA is well-suited for data with clear seasonal patterns, while Prophet is ideal for handling multiple seasonality’s and holiday effects. Both models offer advanced features for capturing complex patterns in time series data and generating accurate forecasts. By understanding the principles and capabilities of

SARIMA and Prophet, analysts and practitioners can make informed decisions and effectively forecast future trends in their data. Time series forecasting plays a crucial role in various fields, including finance, economics, and weather prediction.

In recent years, advanced forecasting models such as SARIMA and Prophet have gained popularity due to their ability to capture complex patterns in time series data.

### ❖ RANDOM FOREST:

Random Forest is a versatile and widely used ensemble learning algorithm in machine learning. It is known for its robustness, flexibility, and effectiveness across various types of datasets and predictive tasks.

Components of Random Forest: Random Forest consists of several components that contribute to its effectiveness:

- **Decision Trees:** Each decision tree in the Random Forest ensemble is constructed using a subset of the training data and a subset of the input features. This randomness ensures diversity among the trees and helps prevent overfitting.
- **Bootstrap Aggregating (Bagging):** Random Forest employs a technique called bagging, which involves training each decision tree on a random sample of the training data, drawn with replacement. This sampling process introduces diversity among the trees and reduces the variance of the final model.
- **Random Feature Selection**: In addition to sampling the training data, Random Forest also randomly selects a subset of features to consider when splitting each node of the decision tree. This further diversifies the trees and prevents them from relying too heavily on any single feature.
- **Voting or Averaging Mechanism:** The final prediction of the Random Forest is determined by aggregating the predictions of all individual trees through a voting mechanism for classification tasks or averaging for regression tasks.

Applications of Random Forest: Random Forest has wide-ranging applications across various domains, including:

- **Classification**: Random Forest is commonly used for classification tasks, such as spam detection, medical diagnosis, and customer churn prediction. Its ability to handle high-dimensional data and complex decision boundaries makes it well-suited for a wide range of classification problems.

- **Regression**: Random Forest can also be applied to regression tasks, such as stock price prediction, housing price estimation, and demand forecasting. Its robustness to outliers and nonlinear relationships in the data makes it an effective tool for predicting continuous variables.
- **Feature Importance**: By analyzing the relative importance of features based on their contribution to the performance of the model, practitioners can gain insights into the underlying factors driving the prediction.

Random Forest is a powerful ensemble learning algorithm that combines the strengths of decision trees with the benefits of ensemble learning. Its ability to handle complex datasets, prevent overfitting, and provide insights into feature importance makes it a valuable tool in predictive modeling and data analysis.

### ❖ K-Means Clustering:

At its core, K-means clustering operates on the principle of partitioning a dataset into 'K' clusters, where 'K' is a predefined number chosen by the user. The algorithm aims to minimize the within-cluster variance, which is the sum of squared distances between data points and their respective cluster centroids. The key principles of K-means clustering include:

- **Centroid-Based Clustering:** K-means clustering adopts a centroid-based approach, where each cluster is represented by a centroid, which is the mean of all data points assigned to that cluster. The algorithm iteratively adjusts the positions of the centroids to minimize the within-cluster variance.
- **Initialization:** The algorithm begins by randomly initializing 'K' centroids in the feature space. These initial centroids serve as the starting points for the clustering process.
- **Assignment Step:** In this step, each data point is assigned to the nearest centroid based on its Euclidean distance. The assignment of data points to clusters is based on the principle of minimizing the within-cluster variance.
- **Update Step**: After assigning all data points to clusters, the centroids are recalculated as the mean of all data points belonging to each cluster. This step aims to update the positions of the centroids to better represent the clusters.
- **Convergence:** The assignment and update steps are repeated iteratively until convergence, i.e., until the centroids no longer change significantly or a predefined

convergence criterion is met. At convergence, the algorithm has partitioned the dataset into 'K' clusters, and each data point is assigned to the cluster represented by the nearest centroid.

Methodology of K-Means Clustering: The methodology of K-means clustering can be summarized as follows:

- **Initialization**: Randomly initialize 'K' centroids in the feature space.
- **Assignment Step**: Assign each data point to the nearest centroid based on Euclidean distance.
- **Update Step**: Recalculate the centroids as the mean of all data points assigned to each cluster.
- **Repeat**: Iterate the assignment and update steps until convergence is reached.
- **Convergence**: Stop the algorithm when the centroids no longer change significantly or a convergence criterion is met.
- **Output**: Obtain the final clusters, with each data point assigned to the cluster represented by the nearest centroid.

## 3.3. Designing:

### 3.3.1. Architecture:

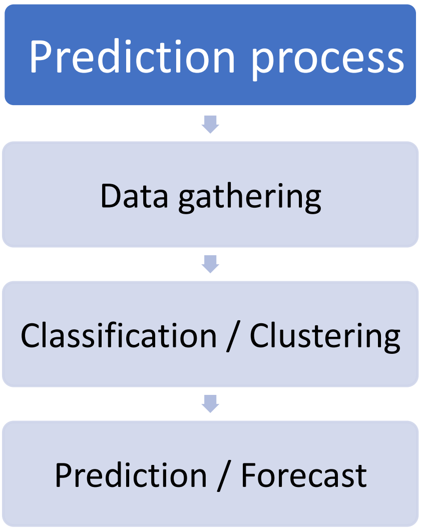


Figure.3.3.1: Architecture of the Proposed System

### 3.3.2. Sequence Diagram:

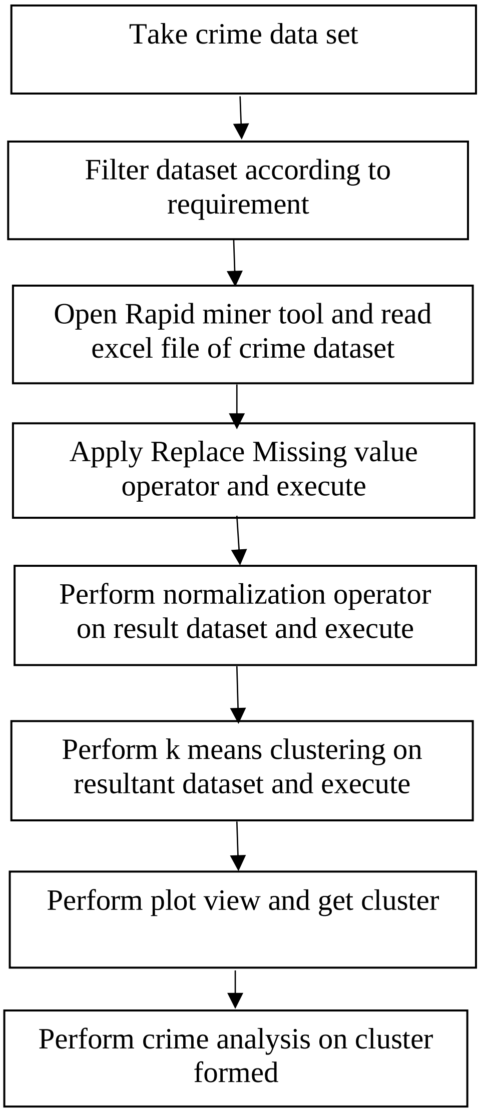

Take crime data set
Filter dataset according to requirement
Open Rapid miner tool and read excel file of crime dataset
Apply Replace Missing value operator and execute
Perform normalization operator on result dataset and execute
Perform k means clustering on resultant dataset and execute
Perform plot view and get cluster
Perform crime analysis on cluster formed

Figure.3.3.2: Sequence flow of Proposed System

### 3.4. Stepwise Implementation:

In this work, we will build a machine learning module. The model works on the concept of Time Series Forecasting and Clustering. After successful running of the module the analysis and the forecasting results are shown through graphs and plots. We intend to provide an analysis study by combining our findings of a particular crimes' dataset with its demographic's information.

### 3.4.1 Data Collection and Preprocessing:

The data which we have used in this model has been collected from Toronto Police

Service- Public safety portal. This collected data can be used to extract meaningful information which can help both the police department and the public to maintain safe surroundings. Identifying crime and predicting dangerous hotspots at a certain time and place could provide a better visualization for both public and authorities.

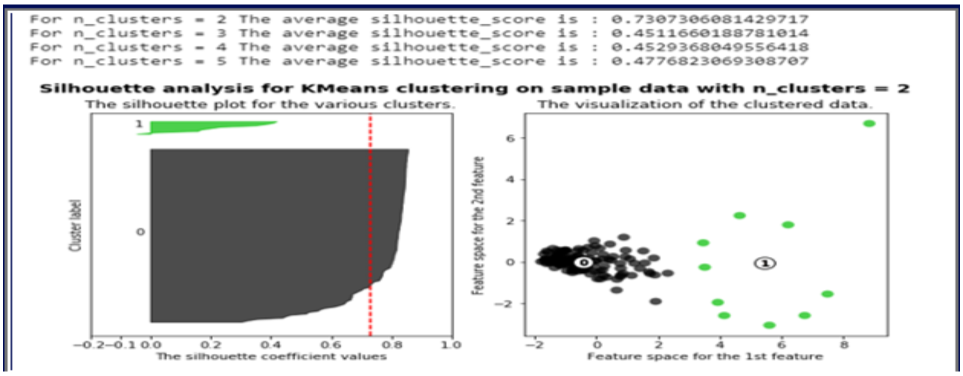
In this dataset we have columns such as- Occurrence date, month, reporting date, Neighborhood, type of offence, MCI (or Major Crime Indicators). This would help in clearly showcasing which neighborhood are dangerous and require more focus of police agencies. It would also supplement to the general public's knowledge for their own well-being and safety. We have used several time series forecasting along with classification and clustering algorithms to validate the results.

### 3.4.2 Classification:

Classification refers to a predictive modelling problem where a class label is predicted for a given example of input data. In this project, we have used Random Forest Classification algorithm which operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean/average prediction of the individual trees and have got an accuracy of 58.38%.

### 3.4.3 Clustering:

Clustering, is an unsupervised machine learning task. It involves automatically discovering natural grouping in data. In this model we have used K-Means Clustering technique that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. We have chosen Elbow method to determine number of clusters in the dataset.



```
For n_clusters = 2 The average silhouette_score is : 0.7307306081429717
For n_clusters = 3 The average silhouette_score is : 0.4511660188781014
For n_clusters = 4 The average silhouette_score is : 0.4529368049556418
For n_clusters = 5 The average silhouette_score is : 0.4776823069308707
```

Silhouette analysis for KMeans clustering on sample data with n_clusters = 2
The silhouette plot for the various clusters.
The visualization of the clustered data.
Cluster label
1
0
−0.2
−0.1
0.0
0.2
0.4
0.6
0.8
1.0
The silhouette coefficient values
Feature space for the 2nd feature
6
4
2
0
−2
−2
0
2
4
6
8
Feature space for the 1st feature

Figure.3.4.1: silhouette Analysis

### 3.4.4 Time Series Analysis:

In our Crime Visualization component, we utilize Time Series Analysis techniques, specifically ARIMA and Prophet, to uncover and visualize patterns in crime data over time. ARIMA (Auto-Regressive Integrated Moving Average) helps identify trends and seasonality in historical crime data, enabling us to make predictions about future crime rates. Prophet, a forecasting tool, enhances our analysis by capturing daily, weekly, and yearly patterns in the data and providing more accurate predictions. Through visualizations generated by these analyses, we gain insights into the temporal dynamics of crime, aiding law enforcement in developing targeted strategies and making informed decisions to enhance public safety.

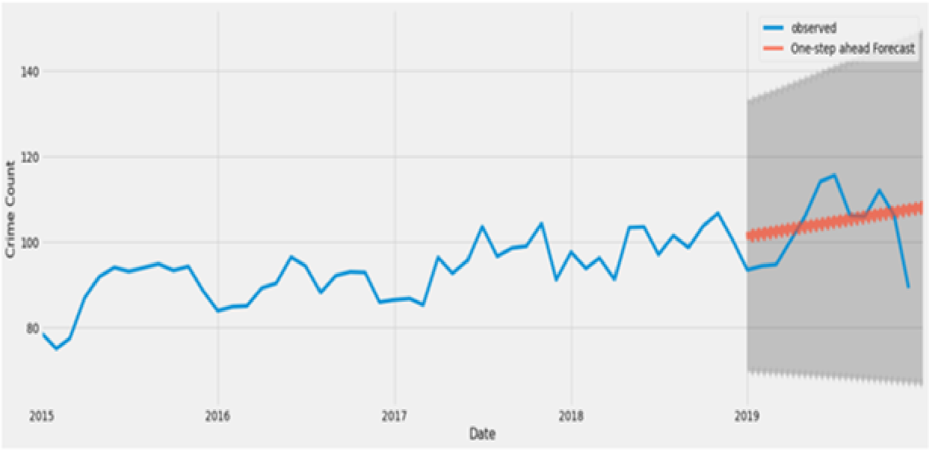


Figure.3.4.2: SARIMAX Forecasting

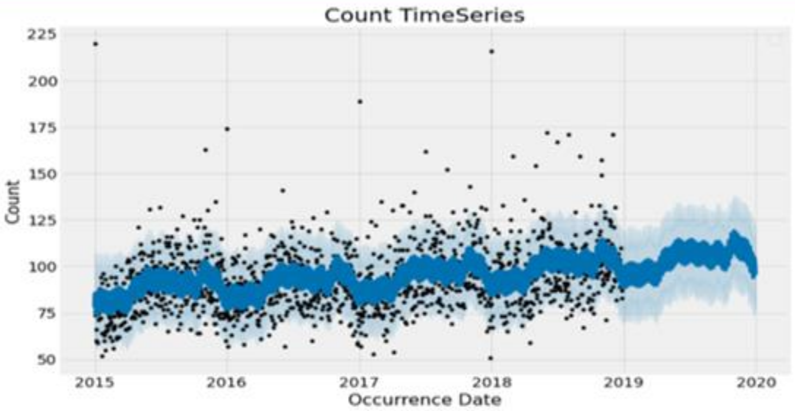


Figure.3.4.3: Prophet Time Series Forecasting

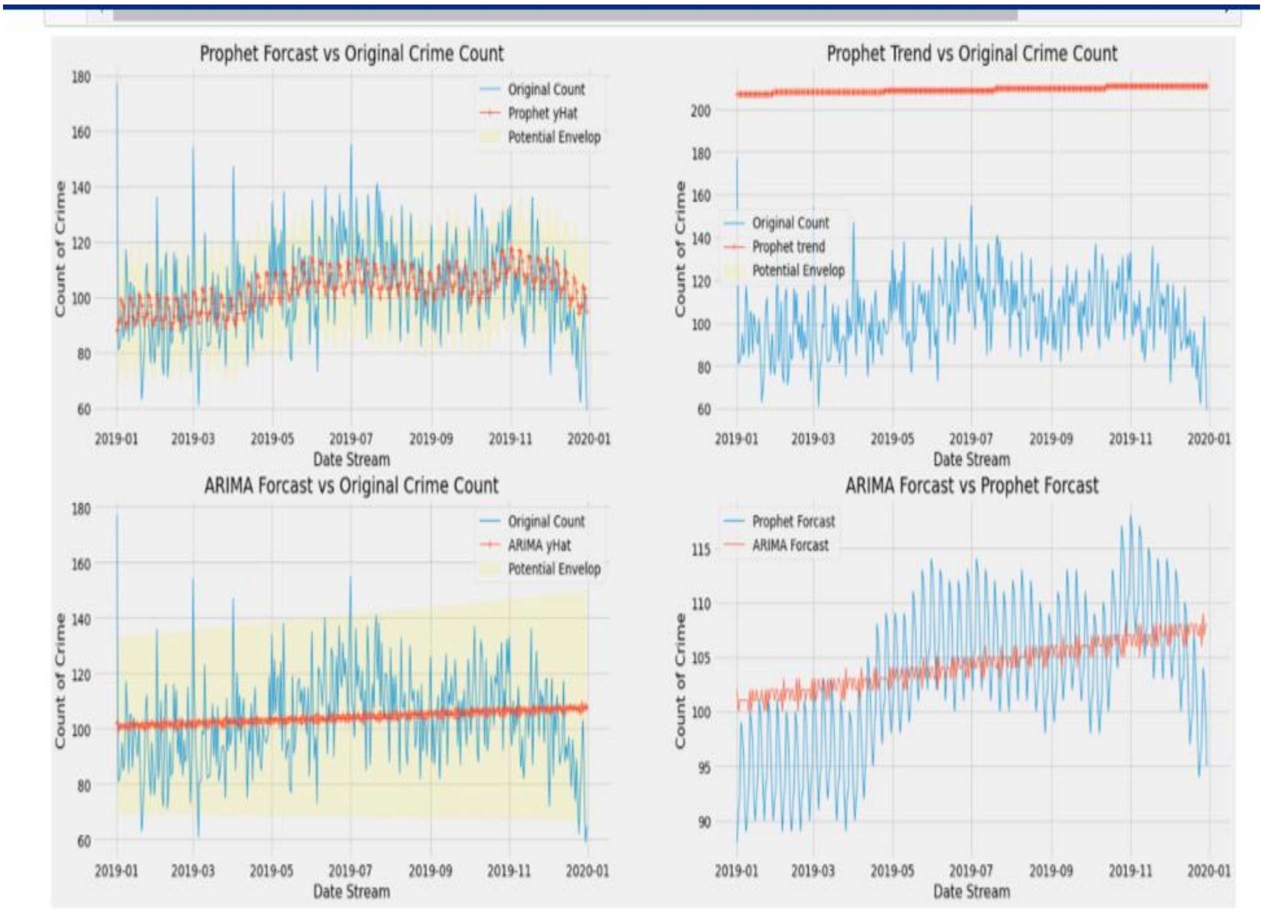


Figure.3.4.4: Differences between the original &the predictive model plotted in graphical manner.

### 3.4.5 Visualization:

In our Crime Visualization component, we use advanced time series analysis techniques like ARIMA and Prophet to uncover meaningful patterns and trends in crime data over time. By leveraging these tools, we create visual representations that highlight the temporal dynamics of criminal activities. These visualizations offer a clear and intuitive understanding of how crime rates fluctuate, allowing law enforcement and decision-makers to identify peak periods, seasonal variations, and potential future trends. The goal is to provide a comprehensive and visually accessible overview of crime patterns, facilitating effective decision-making and resource allocation for crime prevention and intervention strategies.

## 3.5 Implementation Code and output:

```
plt.figure(figsize=(30,10))
MCI= df.groupby('MCI',as_index=False).size()
assaultTypes= df[df.MCI=='Assault'].groupby('offence', as_index=False).size()
autoTheftTypes= df[df.MCI=='Auto Theft'].groupby('offence', as_index=False).size()
plt.subplot(221)
sns.barplot(x='MCI', y='size', data=MCI.sort_values(by='size', ascending=False))
plt.title('Major Crime Indicator', fontsize='xx-large')
plt.xlabel('Types of Crime', fontsize='x-large')
plt.ylabel('Crime Count', fontsize='x-large')
plt.subplot(222)
sns.barplot(x='offence', y='size', data=assaultTypes.sort_values(by='size', ascending=False))
plt.title('Offence Distribution by Assult', fontsize='xx-large')
plt.xlabel('Types of Offence', fontsize='x-large')
plt.ylabel('Offence Count', fontsize='x-large')
plt.xticks(rotation=90)
plt.show()
```

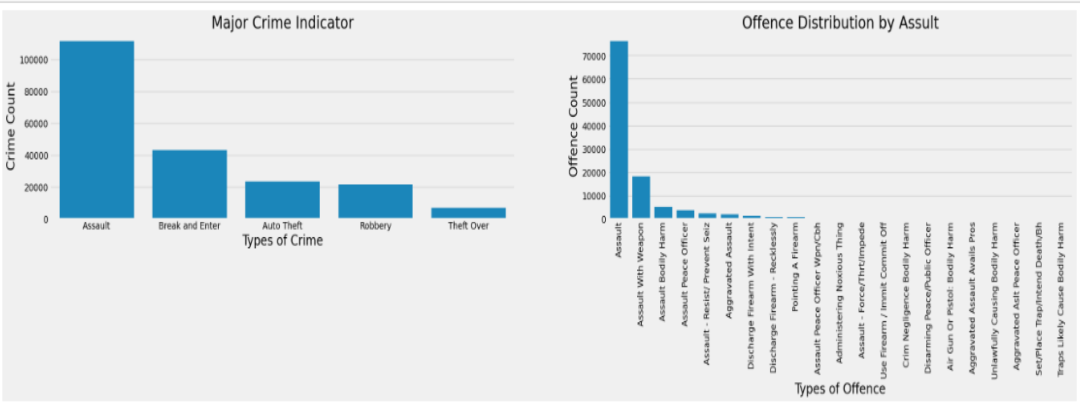


Fig.3.5.1: Type of crime vs major crime indicators.

```
plt.figure(figsize=(30,8))
breakEnterTypes= df[df.MCI=='Break and Enter'].groupby('offence', as_index=False).size()
robberyTypes= df[df.MCI=='Robbery'].groupby('offence', as_index=False).size()
theftOverTypes= df[df.MCI=='Theft Over'].groupby('offence', as_index=False).size()
plt.subplot(131)
sns.barplot(x='offence', y='size', data=breakEnterTypes.sort_values(by='size',
ascending=False))
plt.title('Offence Distribution by Break and Enter', fontsize='xx-large')
plt.xlabel('Types of Offence', fontsize='x-large')
plt.ylabel('Offence Count', fontsize='x-large')
plt.xticks(rotation=90)
plt.subplot(132)
sns.barplot(x='offence', y='size', data=robberyTypes.sort_values(by='size', ascending=False))
plt.title('Offence Distribution by Robbery', fontsize='xx-large')
plt.xlabel('Types of Offence', fontsize='x-large')
plt.ylabel('Offence Count', fontsize='x-large')
plt.xticks(rotation=90)
plt.subplot(133)
```

```
sns.barplot(x='offence', y='size', data=theftOverTypes.sort_values(by='size',
ascending=False))
plt.title('Offence Distribution by Theft Over', fontsize='xx-large')
plt.xlabel('Types of Offence', fontsize='x-large')
plt.ylabel('Offence Count', fontsize='x-large')
plt.xticks(rotation=90)
plt.show()
```

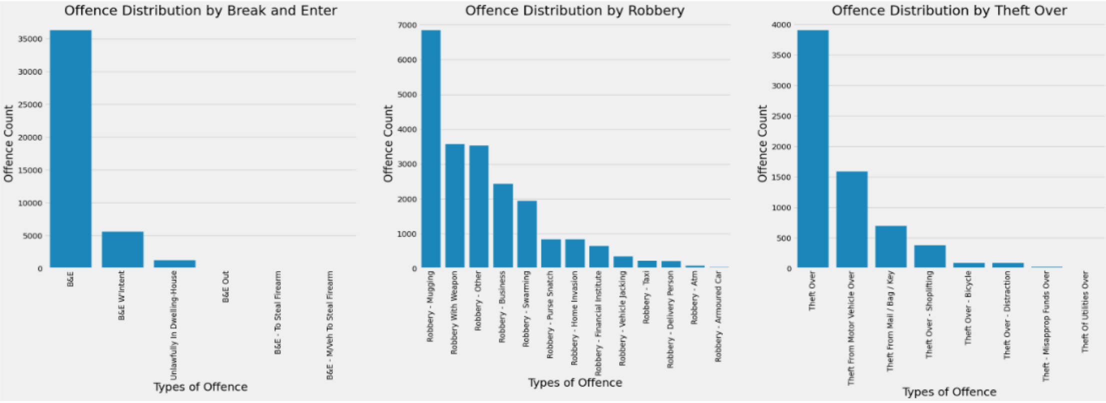


Fig.3.5.2: Type of crimes vs various indicators Graphs.

```
mci_distribution =
df.groupby(['occurrenceyear','occurrencemonth'],as_index=False).agg({'count':'sum'})
mci_distribution['occurrenceyear']=
mci_distribution['occurrenceyear'].astype('int').astype('str')
mci_distribution['monthYear']= mci_distribution['occurrencemonth'] +', '+
mci_distribution['occurrenceyear']
plt.grid('on')
plt.plot(mci_distribution['monthYear'], mci_distribution['count'])
plt.xlabel('Month Stream')
plt.ylabel('Count of Crime')
plt.title('Time Series Distribution of Crime [Monthwise]')
plt.xticks(np.arange(0, mci_distribution['monthYear'].shape[0], 3), rotation=90)
plt.show()
```

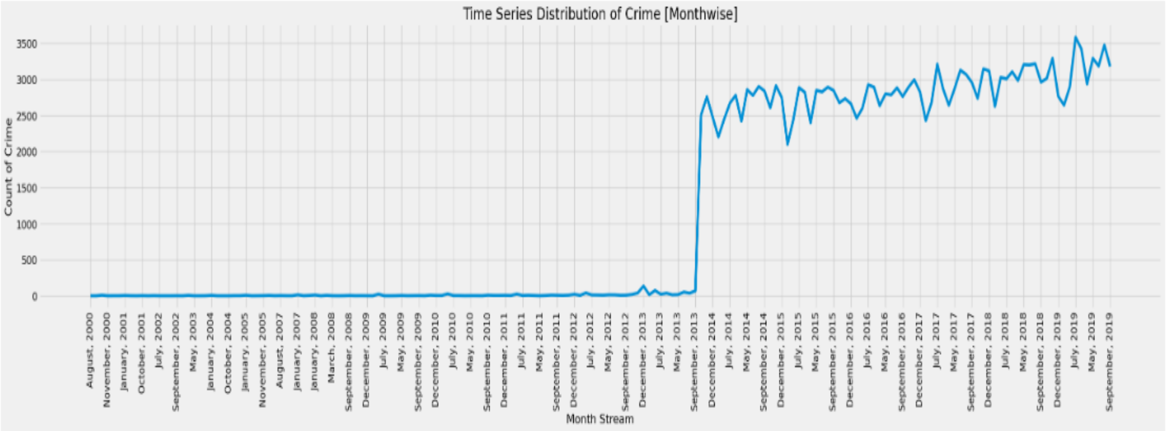


Fig.3.5.3: Time series Distribution of crime [month-wise]

```
decomposition=sm.tsa.seasonal_decompose(trimmedOccuranceDateCount[trimmedOccuranceDateCount.index > pd.to_datetime('2018-12-31')], model='additive')
plt.rcParams['figure.figsize'] = 24, 8
decomposition.plot()
plt.show()
```

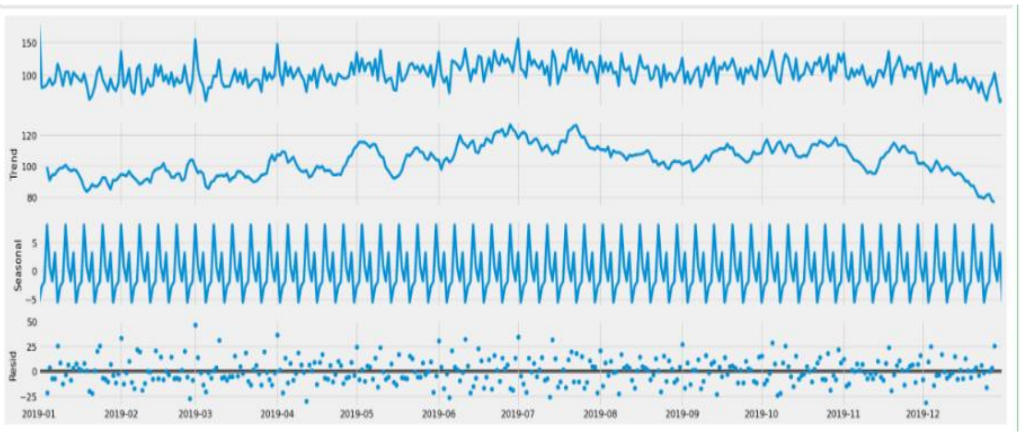


Fig.3.5.4: crime Time series indicator graph year wise.

```
plt.rcParams['figure.figsize']= 24, 12
plt.subplot(221)
plt.grid('on')
plt.title('Prophet Forcast vs Original Crime Count')
plt.plot(prophetForcast_yhat.index, testData['count'].iloc[:-1], lw=1)
plt.plot(prophetForcast_yhat.index, prophetForcast_yhat['yhat'], lw=1, marker='+')
plt.fill_between(prophetForcast_yhat.index, prophetForcast_yhat['yhat_upper'],
prophetForcast_yhat['yhat_lower'], color='#f7ed25', alpha=.15)
plt.xlabel('Date Stream')
plt.ylabel('Count of Crime')
plt.legend(['Original Count', 'Prophet yHat', 'Potential Envelop'])
plt.subplot(222)
plt.grid('on')
plt.title('ARIMA Forcast vs Original Crime Count')
plt.plot(testData.index, testData['count'], lw=1)
plt.plot(testData.index, testData['predValues'], lw=1, marker='+')
plt.fill_between(testData.index, testData['upperCount'], testData['lowerCount'],
color='#f7ed25', alpha=.15)
plt.xlabel('Date Stream')
plt.ylabel('Count of Crime')
plt.legend(['Original Count', 'ARIMA yHat', 'Potential Envelop'])
plt.subplot(223)
plt.grid('on')
plt.title('ARIMA Forcast vs Prophet Forcast')
plt.plot(prophetForcast_yhat.index, prophetForcast_yhat['yhat'], lw=1)
plt.plot(testData.index[:-1], testData['predValues'][:-1], lw=1)
plt.xlabel('Date Stream')
plt.ylabel('Count of Crime')
plt.legend(['Prophet Forcast', 'ARIMA Forcast'])
plt.show()
```

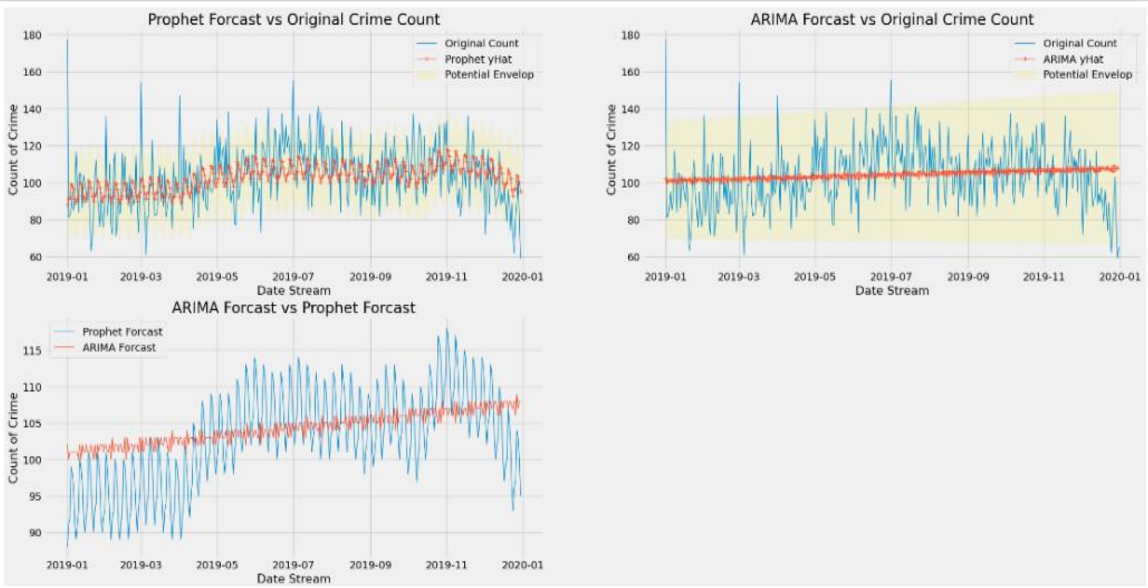
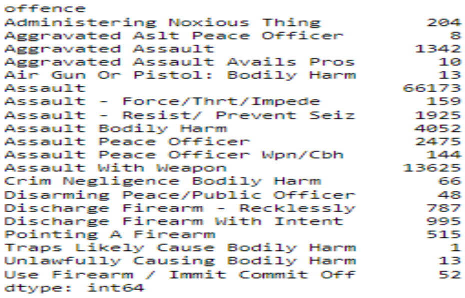


Fig.3.5.5: Prophet Forecast vs original crime count graph.

```
assault = df[df['MCI'] == 'Assault']
assault_types = assault.groupby('offence',as_index=False).size()
print(assault_types)
ct = assault_types.sort_values(ascending = False)
ax = ct.plot.bar()
ax.set_xlabel('Types of Assault')
ax.set_ylabel('Number of occurences')
ax.set_title('Assault crimes in Toronto',color = 'green',fontsize=20)
plt.show()
```



```
offence
Administering Noxious Thing           204
Aggravated Aslt Peace Officer           8
Aggravated Assault                   1342
Aggravated Assault Avails Pros         10
Air Gun Or Pistol: Bodily Harm         13
Assault                             66173
Assault - Force/Thrt/Impede           159
Assault - Resist/ Prevent Seiz       1925
Assault Bodily Harm                  4052
Assault Peace Officer                2475
Assault Peace Officer Wpn/Cbh         144
Assault With Weapon                 13625
Crim Negligence Bodily Harm            66
Disarming Peace/Public Officer         48
Discharge Firearm - Recklessly        787
Discharge Firearm With Intent         995
Pointing A Firearm                    515
Traps Likely Cause Bodily Harm          1
Unlawfully Causing Bodily Harm         13
Use Firearm / Immit Commit Off         52
dtype: int64
```

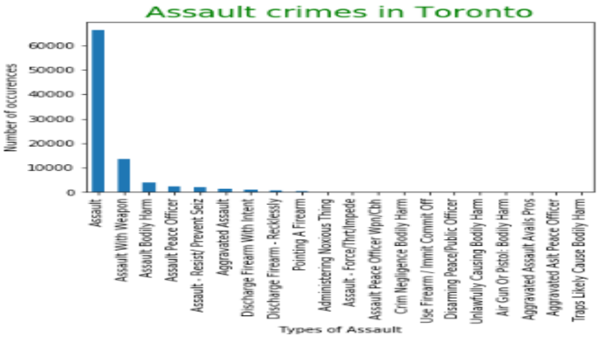


Fig.3.5.6: Type of assault vs no of occurrences graph.

```
import seaborn as sns
mci_monthwise =
df.groupby(['occurrencemonth','MCI'],as_index=False).agg({'Total':'sum'})
plt.figure(figsize=(15, 7))
crime_count = mci_monthwise.pivot("MCI","occurrencemonth","Total" )
plt.yticks(rotation=1)
ax = sns.heatmap(crime_count,cmap="YlGnBu", linewidths=.5)
plt.title("Major Crime Indicators by Month",color = 'red',fontsize=14)
plt.show()
```

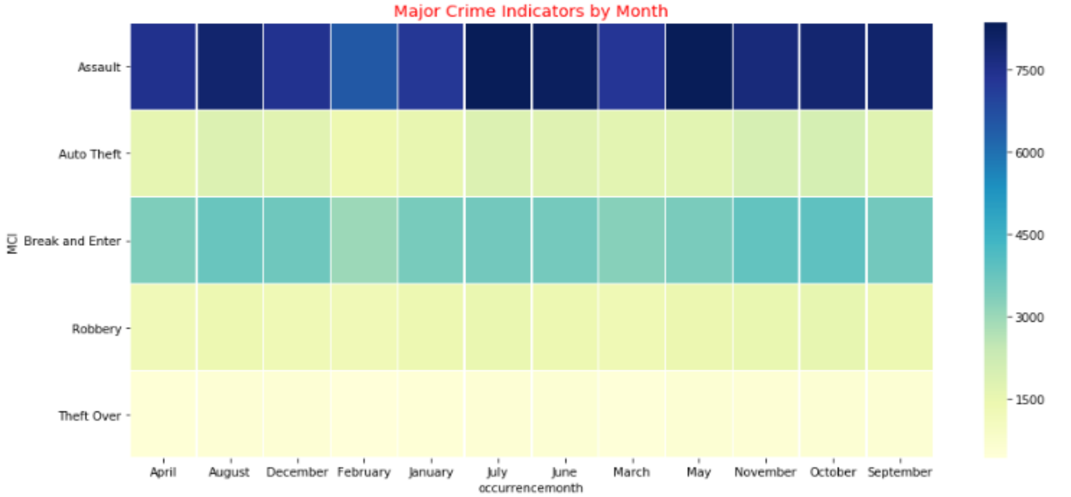


Fig.3.5.7: Major crime indicator vs month visualization graph.

```
hour_crime_group =
df.groupby(['occurrencehour','MCI'],as_index=False).agg({'Total':'sum'})
fig, ax = plt.subplots(figsize=(15,10))
hour_crime_group.groupby('MCI').plot(x="occurrencehour", y="Total",
ax=ax,linewidth=5)
ax.set_xlabel('Hour')
ax.set_ylabel('Number of occurences')
ax.set_title('Crime Types by Hour of Day in Toronto',color =
'red',fontsize=25)
```

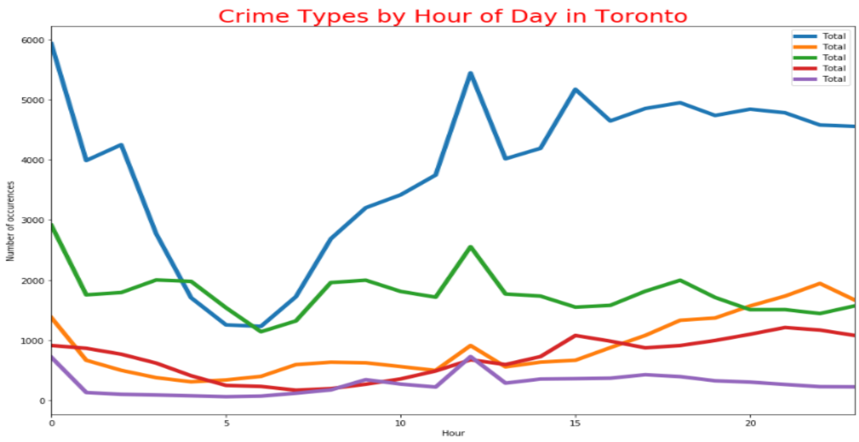


Fig.3.5.8: Graph for no of occurrences in hour.

```
plt.plot(K, Sum_of_squared_distances0, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```

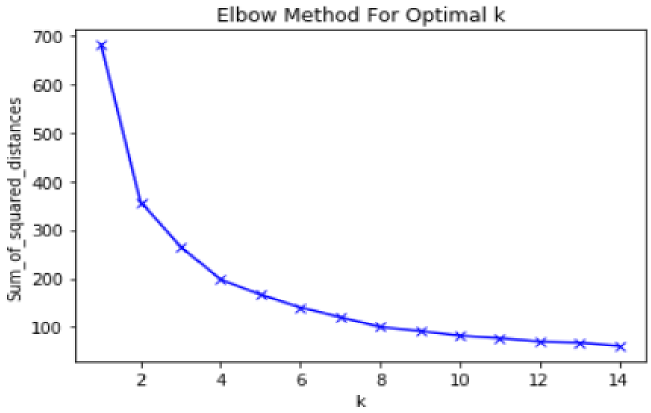
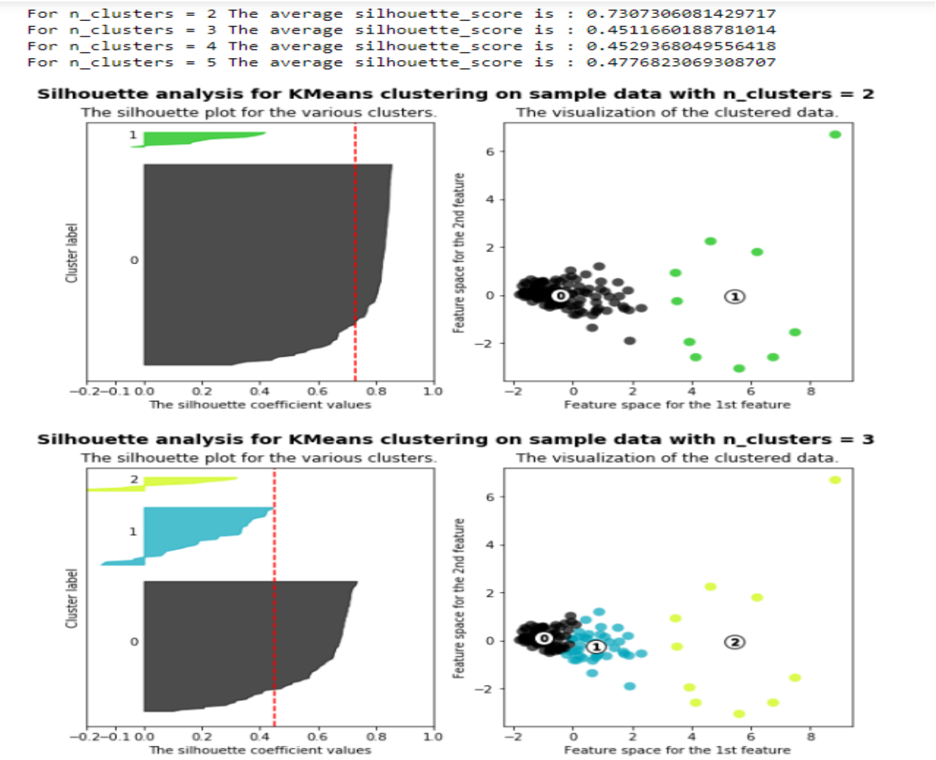


Fig.3.5.9: Elbow method for Optimal K-means algo

```
for n_clusters in range(2,6):
    kmeans = KMeans(n_clusters=n_clusters , random_state=3425)
    cluster_labels = kmeans.fit_predict(X0)
    silhouette_avg = silhouette_score(X0, cluster_labels)
    print("For n_clusters =", n_clusters, "The average silhouette_score is :", silhouette_avg)
    sample_silhouette_values = silhouette_samples(X0, cluster_labels)
    fig, (ax1, ax2) = plt.subplots(1, 2)
    fig.set_size_inches(10, 5)
    ax1.set_xlim([-0.1, 1])
    ax1.set_ylim([0, len(X) + (n_clusters + 1) * 10])
    y_lower = 10
    for i in range(n_clusters):
       # samples belonging to
        # cluster i, and sort them
        ith_cluster_silhouette_values = \
            sample_silhouette_values[cluster_labels == i]
        ith_cluster_silhouette_values.sort()
        size_cluster_i = ith_cluster_silhouette_values.shape[0]
        y_upper = y_lower + size_cluster_i
        color = cm.nipy_spectral(float(i) / n_clusters)
        ax1.fill_betweenx(np.arange(y_lower, y_upper),
                          0, ith_cluster_silhouette_values,
                          facecolor=color, edgecolor=color, alpha=0.7)
        # Label
        ax1.text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
        # Computing the new y_lower
        y_lower = y_upper + 10  # 10 for the 0 samples
    ax1.set_title("The silhouette plot for the various clusters.")
    ax1.set_xlabel("The silhouette coefficient values")
    ax1.set_ylabel("Cluster label")
```

```
  # vertical line
  ax1.axvline(x=silhouette_avg, color="red", linestyle="--")
  ax1.set_yticks([])
  ax1.set_xticks([-0.2,-0.1, 0, 0.2, 0.4, 0.6, 0.8, 1])
  # showing the actual cluster
  colors = cm.nipy_spectral(cluster_labels.astype(float) / n_clusters)
  ax2.scatter(X0[:, 0], X0[:, 1], marker='.', s=300, lw=0, alpha=0.7,
            c=colors, edgecolor='k')
  # Labeling the clusters
  centers = kmeans.cluster_centers_
  #  cluster centers
  ax2.scatter(centers[:, 0], centers[:, 1], marker='o',
            c="white", alpha=1, s=200, edgecolor='k')
  for i, c in enumerate(centers):
    ax2.scatter(c[0], c[1], marker='$%d$' % i, alpha=1,
              s=50, edgecolor='k')
  ax2.set_title("The visualization of the clustered data.")
  ax2.set_xlabel("Feature space for the 1st feature")
  ax2.set_ylabel("Feature space for the 2nd feature")
  plt.suptitle(("Silhouette analysis for KMeans clustering on sample data "
               "with n_clusters = %d" % n_clusters),fontsize=14,
fontweight='bold')
plt.show()
ax1.legend()
ax2.legend()
```



```
For n_clusters = 2 The average silhouette_score is : 0.7307306081429717
For n_clusters = 3 The average silhouette_score is : 0.4511660188781014
For n_clusters = 4 The average silhouette_score is : 0.4529368049556418
For n_clusters = 5 The average silhouette_score is : 0.4776823069308707
```

Silhouette analysis for KMeans clustering on sample data with n_clusters = 2
The silhouette plot for the various clusters.
The visualization of the clustered data.
Cluster label
1
0
−0.2 −0.1 0.0 0.2 0.4 0.6 0.8 1.0
The silhouette coefficient values
Feature space for the 2nd feature
6
4
2
0
−2
−2 0 2 4 6 8
Feature space for the 1st feature
Silhouette analysis for KMeans clustering on sample data with n_clusters = 3
The silhouette plot for the various clusters.
The visualization of the clustered data.
Cluster label
2
1
0
−0.2 −0.1 0.0 0.2 0.4 0.6 0.8 1.0
The silhouette coefficient values
Feature space for the 2nd feature
6
4
2
0
−2
−2 0 2 4 6 8
Feature space for the 1st feature

Fig.3.5.10: Silhouette analysis for K-means algorithm for clustering.

# CHAPTER 4
## RESULTS AND DISCUSSION

# CHAPTER 4

# RESULTS AND DISCUSSION

The exploration of existing solutions sheds light on the diverse approaches and methodologies available to enhance the capabilities of integration of multiple models.

**Social Media Analysis for Crime Prediction**

Approach:

The project employs advanced natural language processing and machine learning techniques for social media analysis, extracting valuable insights to predict potential crime hotspots. Through sentiment analysis, entity recognition, and geospatial mapping, the approach aims to discern patterns and indicators of criminal activities within the vast landscape of social media data. By combining these techniques, the project enhances law enforcement capabilities to proactively address emerging threats and allocate resources efficiently.

Benefits:

Integrating social media insights with existing systems contributes to more accurate predictive analytics, aiding in the development of robust crime prediction models. Furthermore, ethical data collection practices and privacy protection ensure responsible implementation, crucial for maintaining public trust and adherence to ethical standards.

**Crime Trend Analysis using ML**

Approach:

In Crime Trend Analysis using Machine Learning (ML), the project involves collecting comprehensive historical crime data, including crime types, locations, and contextual factors. Utilizing machine learning algorithms, the approach includes temporal analysis, crime type classification, geospatial mapping, and consideration of demographic and socio-economic factors, the project aims to predict future crime trends, providing law enforcement with actionable insights to strategically allocate resources and implement targeted interventions.

Benefits:

By considering demographic and socio-economic factors, the project offers insights into the root causes of crime. Overall, Crime Trend Analysis using ML contributes to evidence-based decision-making, proactive law enforcement, and the development of effective crime prevention strategies.

**Crime prediction using Sentiment Analysis in Police Reports using NLP**

Approach:

Crime prediction using Sentiment Analysis in Police Reports employs Natural Language Processing (NLP) to analyze the emotional tone in law enforcement documentation.

The approach involves data collection from police reports, text preprocessing, sentiment analysis, and entity recognition. By understanding the emotional context surrounding criminal incidents, the project aims to uncover patterns indicative of potential escalations or de-escalations in criminal activities.

Benefits:

This approach offers significant benefits in crime prediction. Firstly, by leveraging sentiment analysis, it provides law enforcement with early insights into the emotional dynamics of reported incidents, aiding in the identification of potential future criminal trends.

**Comparison:**

Each solution brings its own merits to the table. While the first solution focuses on precision and task-specific understanding, Social media analysis for crime prediction enhances valuable insights from the data from social media platforms and predicts the crime, and crime trend analysis used to get the trending crime from the any selected place and helps us to find root cause of the crimes, while the crime prediction using the sentiment analysis in police reports uses the NLP which does the sentiment analysis using trained data and gets us the accurate criminal trends and accurate crime prediction ,but this model only works when we trained with large data combining aspects of all three solutions presents an ideal approach, Crime Prediction and analysis, gives us the pictorial visualization of the probability of happening of the crimes based on the certain provided data by analyzing these visualizations one can easily predict the crimes.

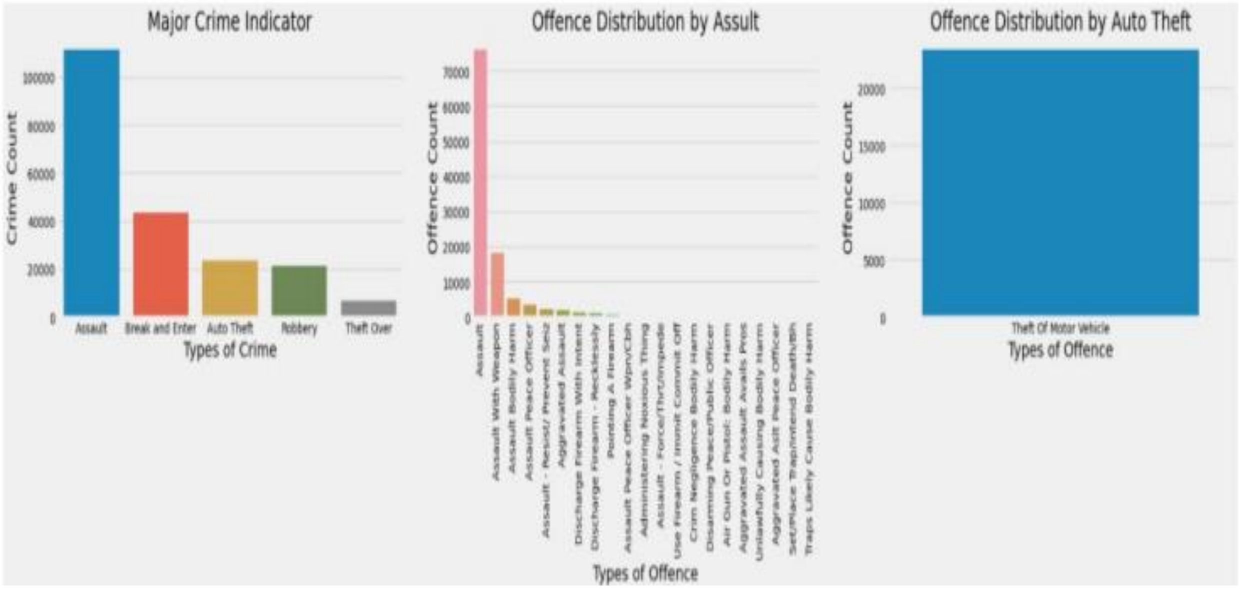


Figure 4.1.1: Plotting and Analysis of different Crime

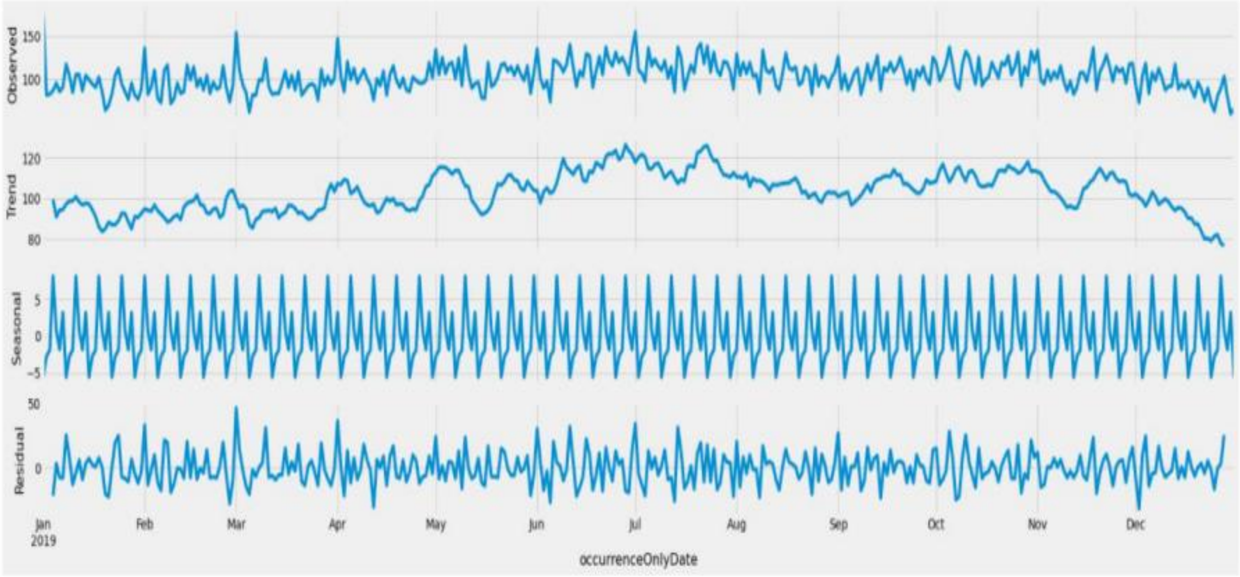
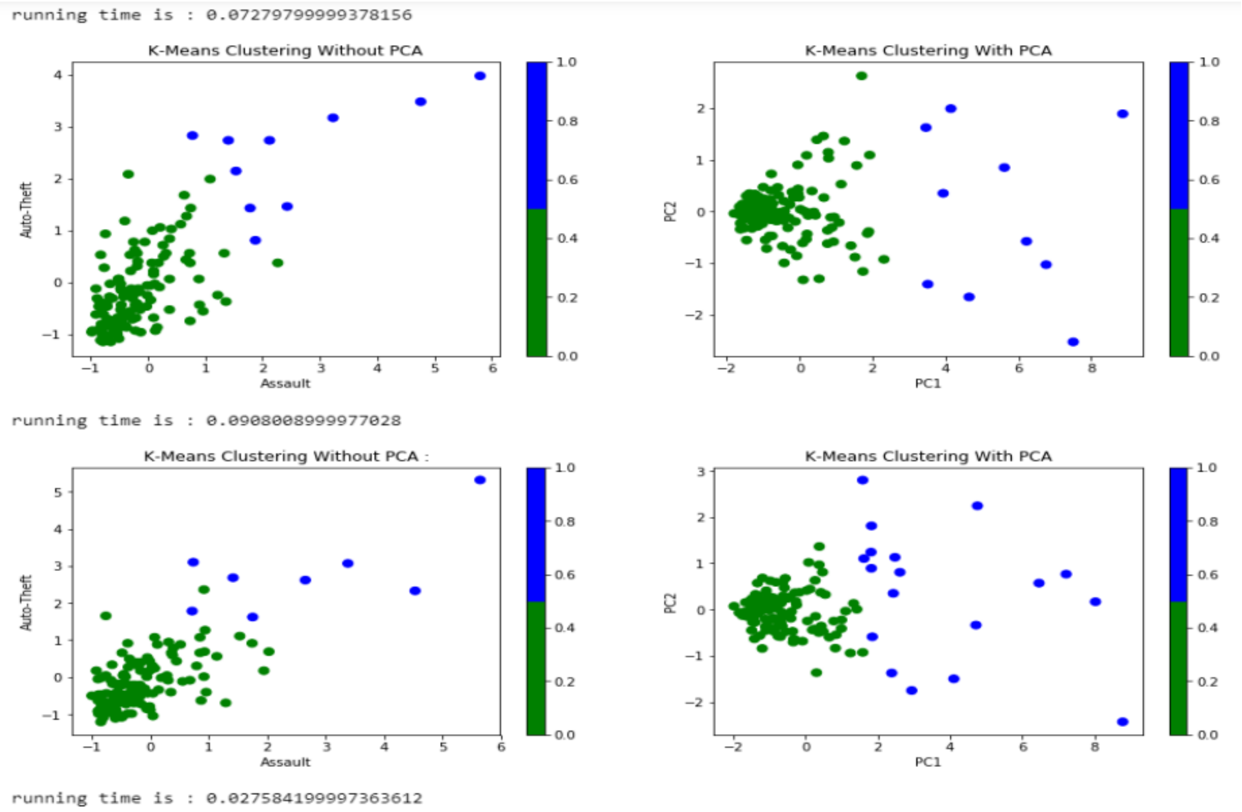


Figure.4.1.2: Time Series Analysis for Total Crime Count



```
running time is : 0.07279799999378156
```

```
running time is : 0.0908008999977028
```

```
running time is : 0.027584199997363612
```

Figure.4.1.3: K-Means Clustering

# CHAPTER 5

# CONCLUSION

# CHAPTER 5
# CONCLUSION

In conclusion, the purpose of this project is to analyze and predict the crimes so that the police can be given an act of hint to take proactive measures before the situation could worsen up. This model helps to analyze and predict crime. Using machine learning approaches, the areas and hotspots can be predicted based on the type of crime and give the overall prediction of any crime. The paper also focuses on building this approach by importing machine learning modules, and time series analyzing methods and also giving the visualizations of the crime.

This project not only enhances our understanding of crime patterns but also establishes a robust system for proactive decision-making and community safety. The combination of forecasting, clustering, and machine learning presents a powerful approach to address the dynamic nature of criminal activities, contributing to effective crime prevention strategies and informed law enforcement initiatives.

The proposed model is very useful for both the investigating agencies and the police officials in taking necessary steps to reduce crime. The project helps the crime analysis to analyze these crime networks by means of various interactive visualizations. Future enhancement of this research work on training bots to predict the crime prone areas by using machine learning techniques. Since, machine learning is similar to data mining, advanced concepts of machine learning can be used for better prediction. The data privacy, reliability, accuracy can be improved for enhanced prediction.

# REFERENCES

# REFERENCES

[1] Kiani, Rasoul, Siamak Mahdavi, and Amin Keshavarzi. "Analysis and prediction of crimes by clustering and classification." International Journal of Advanced Research in Artificial Intelligence 4, no. 8 (2015): 11-17.

[2] Akansha A Chikhale, Ankita K Dhavale, Aparna P Thakre, Diksha B Herode, Nikita D Nasre, Pracheta D Patrikar, Prof. Milind Tote. “A Review on Crime Rate Analysis Using Data Mining” International Journal of Scientific Research in Science, Engineering and Technolgoy 5, no. 5 (2019): 119 – 125.

[3] J. Zhou, Z. Li, J. J. Ma and F. Jiang, "Exploration of the Hidden Influential Factors on Crime Activities: A Big Data Approach," in IEEE Access, vol. 8, pp. 141033141045,2020,doi:10.1109/ACCESS.2020.3009969.

[4] Das, Priyanka, Asit Kumar Das, Janmenjoy Nayak, Danilo Pelusi, and Weiping Ding. "A graph-based clustering approach for relation extraction from crime data." IEEE Access 7 (2019): 101269-101282.

[5] Li Ding, Dana Steil, Matthew Hudnall, Brandon Dixon, Randy Smith, David Brown, Allen Parrish, "PerpSearch: An integrated crime detection system," 2009 IEEE International Conference on Intelligence and Security Informatics, 2009, pp. 161-163, doi: 10.1109/ISI.2009.5137289.

[6] https://bard.google.com/chat

[7] https://ai.meta.com/llama/

[8] GoogleLaMDA | Discover AI use cases (gpt3demo.com)

## GitHub Link:

https://github.com/manohar-3112/Crime-Analysis-And-Prediction

ISBN No. 978-81-967875-5-4 Proceedings of ICRTEM - 2024

# 1st INTERNATIONAL CONFERENCE

## On Recent Trends in Engineering & Management Science

## (ICRTEM - 2024)

Organised by :

**SAI SPURTHI INSTITUTE OF TECHNOLOGY**

on March 11th & 12th 2024.

**Editors :**
**Dr. V.S.R. Kumari**
**Dr. Kishor Kumar .G**
**Dr. K. Bhaskar Mutyalu**

**PROCEEDINGS BOOK**

www.icrtem.com

1st International Conference

on

# Recent Trends in Engineering & Management Science

(ICRTEM-2024)

**11th &12th March 2024, Hybrid Mode**

## PROCEEDINGS BOOK

**Organized by**

**SAI SPURTHI INSTITUTE OF TECHNOLOGY**

**SATHUPALLY**

**In association with**

**NEWZEN INFOTECH, HYDERABAD**

www.icrtem.com | www. saispurthi.ac.in

**ISBN No: 978-81-967875-5-4**

Paper ID: ICRTEM24_197

# CRIME ANALYSIS AND PREDICTION

#1 M. Guru Sai Chawan, UG Student,

#2 T. Manohar, UG Student,

#3 M. Meghana, UG Student,

#4 M. Shiva Kumar, Professor,

Department of CSE,

**CMR COLLEGE OF ENGINEERING & TECHNOLOGY, HYDERABAD,**

***Abstract*—**Crime is a pervasive issue in our society, demanding proactive measures for prevention. With a multitude of crimes occurring frequently, maintaining an accurate database is imperative. This database serves as a crucial resource for future reference and analysis. The core objective of this project is to utilize machine learning and data science techniques to analyze crime datasets and predict potential criminal activities. Extracted from official police portals, the dataset contains essential information such as crime descriptions, types, dates, locations, and times. Prior to model training, rigorous data preprocessing, feature selection, and scaling will be undertaken to enhance predictive accuracy. Various algorithms, including K-Nearest Neighbor (KNN) classification, will be evaluated for crime prediction, with the most accurate one selected for training. Additionally, graphical visualization of the dataset will be employed to discern patterns, such as peak crime times and months of heightened criminal activity. Ultimately, this project aims to showcase how machine learning can empower law enforcement agencies in detecting, predicting, and addressing crimes swiftly, thereby reducing overall crime rates.

***Keywords*—** *Python, K-Nearest Neighbor, Time Series forecasting-SARIMA and Prophet, Random Forest, K-means Clustering, Principle Component Analysis.*

# CRIME ANALYSIS AND PREDICTION.

[#1] M. Guru Sai Chawan, [#2] T. Manohar, [#3] M. Meghana, [#4] M. Shiva Kumar

[1,2,3] UG Student, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana

[4] Professor, Department of CSE, CMR College of Engineering & Technology, Hyderabad, Telangana

***Corresponding Author: T. Manohar, manohartangallapalli225@gmail.com***

***Abstract*—Crime is a pervasive issue in our society, demanding proactive measures for prevention. With a multitude of crimes occurring frequently, maintaining an accurate database is imperative. This database serves as a crucial resource for future reference and analysis. The core objective of this project is to utilize machine learning and data science techniques to analyze crime datasets and predict potential criminal activities. Extracted from official police portals, the dataset contains essential information such as crime descriptions, types, dates, locations, and times. Prior to model training, rigorous data preprocessing, feature selection, and scaling will be undertaken to enhance predictive accuracy. Various algorithms, including K-Nearest Neighbor (KNN) classification, will be evaluated for crime prediction, with the most accurate one selected for training. Additionally, graphical visualization of the dataset will be employed to discern patterns, such as peak crime times and months of heightened criminal activity. Ultimately, this project aims to showcase how machine learning can empower law enforcement agencies in detecting, predicting, and addressing crimes swiftly, thereby reducing overall crime rates.**

***Keywords— Python, K-Nearest Neighbor, Time Series forecasting-SARIMA and Prophet, Random Forest, K-means Clustering, Principle Component Analysis.***

## I INTRODUCTION

Crime is a pervasive issue affecting communities of all sizes, with incidents ranging from minor offenses to serious felonies like robbery, murder, and assault occurring at an alarming rate. As the frequency and spread of crimes increase, there's a pressing need for expedited case resolution. Law enforcement agencies, particularly the police, bear the responsibility of curbing and mitigating these rising crime rates. However, the sheer volume of crime data poses significant challenges for effective crime prediction and criminal identification. Recognizing the urgency for swift case resolution, this project aims to leverage technology, specifically machine learning algorithms implemented in Python, to predict the likelihood of different types of crimes occurring in specific areas. By analyzing datasets sourced from official platforms, this initiative seeks to enhance law enforcement's ability to proactively address and prevent criminal activities.

## II RELATED WORK

In the quest for innovation and efficiency, modern projects frequently rely on existing solutions as fundamental building blocks for development. This approach not only recognizes the expertise and advancements of those who came before us but also nurtures a collaborative ecosystem where ideas can evolve and confront new challenges. In our project, we wholeheartedly embrace this ethos, conscientiously integrating elements from existing solutions to enrich our endeavor. These existing solutions serve as guiding lights, offering insights and frameworks that shape the direction of our project.

### A. Social Media Analysis for Crime Prediction.

Social media platforms have become integral sources of information in today's digital age. This project aims to harness the power of social media data for crime prediction by employing advanced natural language processing (NLP) techniques and machine learning algorithms. The objective is to sift through vast amounts of unstructured data on platforms like Twitter, Facebook, and Instagram to identify potential indicators of criminal activities.

B. **Crime Trend Analysis using ML.** The project on Crime Trend Analysis focuses on utilizing historical crime data to uncover patterns, correlations, and trends that can inform law enforcement agencies and policymakers. Through advanced data analytics and visualization techniques, the project aims to provide actionable insights into the dynamics of criminal activities over time. It aims to understand how crimes happen and when, so that police can create effective strategies to prevent and reduce crime in communities. The project follows ethical guidelines, values transparency, and involves working closely with others, like community members and other organizations, to ensure everyone is on the same page throughout the project.

C. **Crime prediction using Sentiment Analysis in Police Reports using NLP.** This project delves into the exploration of the emotional tone and contextual nuances within law enforcement documentation. By applying natural language processing (NLP) techniques, the project aims to extract sentiment information from police reports, shedding light on the emotional context surrounding criminal incidents. This project aims to enrich the understanding of the human element in law enforcement by uncovering the emotional undertones within police reports. By integrating sentiment analysis into crime analysis workflows, the project strives to enhance communication, transparency, and community trust in law

## III METHODS AND EXPERIMENTAL DETAILS

### A. *Dataset:*

The crime dataset obtained from Kaggle contains valuable information such as crime location, city, time, longitude, latitude, and more. This dataset serves as a crucial resource for our crime prediction and analysis project. By analyzing patterns in this data, we aim to predict potential criminal activities and gain insights into crime dynamics over time. fields like crime location and time, along with geographical coordinates, enable us to identify hotspots and trends. ethical considerations guide our use of this data, ensuring privacy and responsible analysis. through collaboration with relevant stakeholders, including law enforcement and communities, our project seeks to contribute to proactive and evidence-based decision-making for enhancing public safety.

### B. *Data Clustering*:

In our Crime Prediction and Analysis Project, we utilize K-means clustering to group similar patterns within the crime data. K-means is a statistical method that helps identify natural clusters in the dataset based on certain features such as crime location, time, and geographical coordinates. By applying K-means clustering to the crime data, we aim to uncover distinct crime patterns and hotspots. This allows law enforcement to focus resources more effectively, identifying areas with similar crime characteristics. The insights gained from K-means clustering contribute to a more targeted and efficient approach in predicting and preventing criminal activities within communities.

### C. *Time series Analysis:*

In our Crime Visualization component, we utilize Time Series Analysis techniques, specifically ARIMA and Prophet, to uncover and visualize patterns in crime data over time. ARIMA (Auto-Regressive Integrated Moving Average) helps identify trends and seasonality in historical crime data, enabling us to make predictions about future crime rates. Prophet, a forecasting tool, enhances our analysis by capturing daily, weekly, and yearly patterns in the data and providing more accurate predictions. Through visualizations generated by these analyses, we gain insights into the temporal dynamics of crime, aiding law enforcement in developing targeted strategies and making informed decisions to enhance public safety.

enforcement practices. Ethical considerations, privacy protection, and the responsible use of data are paramount throughout the project's development and implementation.

### D. *Crime Visualization:*

In our Crime Visualization component, we use advanced time series analysis techniques like ARIMA and Prophet to uncover meaningful patterns and trends in crime data over time. By leveraging these tools, we create visual representations that highlight the temporal dynamics of criminal activities. These visualizations offer a clear and intuitive understanding of how crime rates fluctuate, allowing law enforcement and decision-makers to identify peak periods, seasonal variations, and potential future trends. The goal is to provide a comprehensive and visually accessible overview of crime patterns, facilitating effective decision-making and resource allocation for crime prevention and intervention strategies.

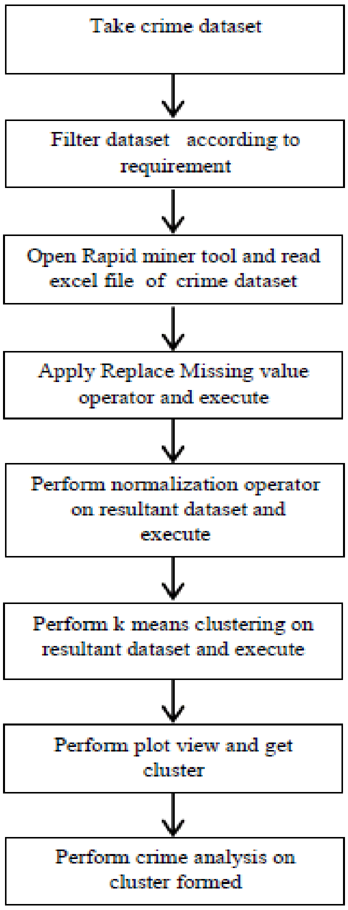

Take crime dataset
Filter dataset according to requirement
Open Rapid miner tool and read excel file of crime dataset
Apply Replace Missing value operator and execute
Perform normalization operator on resultant dataset and execute
Perform k means clustering on resultant dataset and execute
Perform plot view and get cluster
Perform crime analysis on cluster formed

Fig. Architecture of the Model

## IV RESULTS AND DISCUSSIONS

The exploration of existing solutions sheds light on the diverse approaches and methodologies available to enhance the capabilities of integration of multiple models.

### Social Media Analysis for Crime Prediction.

**Approach:**

The project employs advanced natural language processing and machine learning techniques for social media analysis, extracting valuable insights to predict potential crime hotspots. Through sentiment analysis, entity recognition, and geospatial mapping, the approach aims to discern patterns and indicators of criminal activities within the vast landscape of social media data. By combining these techniques, the project enhances law enforcement capabilities to proactively address emerging threats and allocate resources efficiently.

**Benefits**:

Integrating social media insights with existing systems contributes to more accurate predictive analytics, aiding in the development of robust crime prediction models. Furthermore, ethical data collection practices and privacy protection ensure responsible implementation, crucial for maintaining public trust and adherence to ethical standards.

### Crime Trend Analysis using ML

**Approach:**

In Crime Trend Analysis using Machine Learning (ML), the project involves collecting comprehensive historical crime data, including crime types, locations, and contextual factors. Utilizing machine learning algorithms, the approach includes temporal analysis, crime type classification, geospatial mapping, and consideration of demographic and socio-economic factors, the project aims to predict future crime trends, providing law enforcement with actionable insights to strategically allocate resources and implement targeted interventions.

**Benefits:**

By considering demographic and socio-economic factors, the project offers insights into the root causes of crime. Overall, Crime Trend Analysis using ML contributes to evidence-based decision-making, proactive law enforcement, and the development of effective crime prevention strategies.

### Crime prediction using Sentiment Analysis in Police Reports using NLP.

**Approach:**

Crime prediction using Sentiment Analysis in Police Reports employs Natural Language Processing (NLP) to analyze the emotional tone in law enforcement documentation. The approach involves data collection from police reports, text preprocessing, sentiment analysis, and entity recognition. By understanding the emotional context surrounding criminal incidents, the project aims to uncover patterns indicative of potential escalations or de-escalations in criminal activities. Integrating sentiment-based indicators into predictive models enhances the accuracy of crime prediction, allowing for a more nuanced and proactive approach by law enforcement.

**Benefits**:

This approach offers significant benefits in crime prediction. Firstly, by leveraging sentiment analysis, it provides law enforcement with early insights into the emotional dynamics of reported incidents, aiding in the identification of potential future criminal trends. Integrating NLP into crime prediction models enhances the contextual understanding of crime data, contributing to more accurate predictions.

### Comparison:

Each solution brings its own merits to the table. While the first solution focuses on precision and task-specific understanding, Social media analysis for crime prediction enhances valuable insights from the data from social media platforms and predicts the crime, and crime trend analysis used to get the trending crime from the any selected place and helps us to find root cause of the crimes, while the crime prediction using the sentiment analysis in police reports uses the NLP which does the sentiment analysis using trained data and gets us the accurate criminal trends and accurate crime prediction ,but this model only works when we trained with large data combining aspects of all three solutions presents an ideal approach, Crime Prediction and analysis ,gives us the pictorial visualization of the probability of happening of the crimes based on the certain provided data by analyzing this visualizations one can easily predict the crimes.

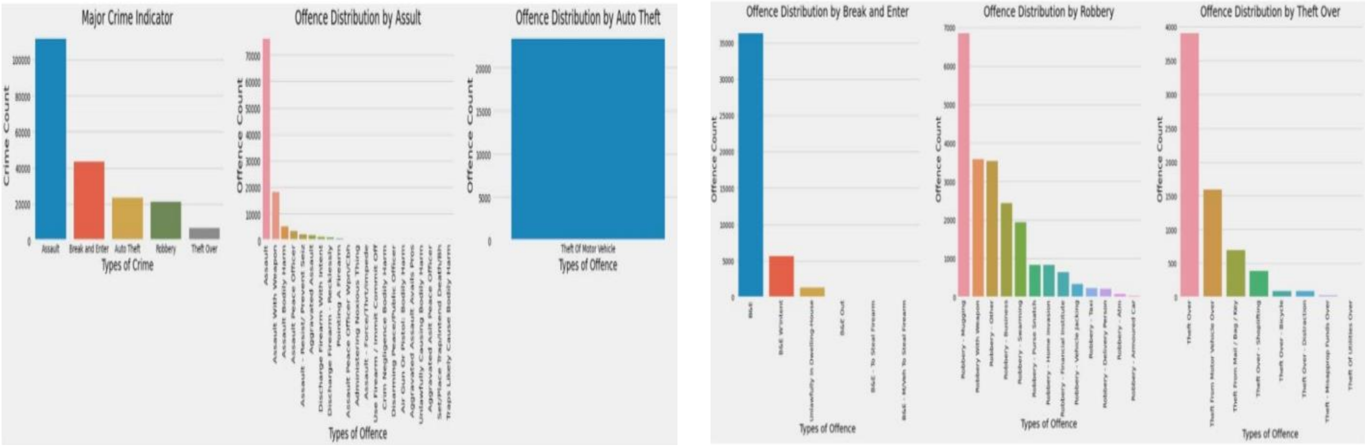
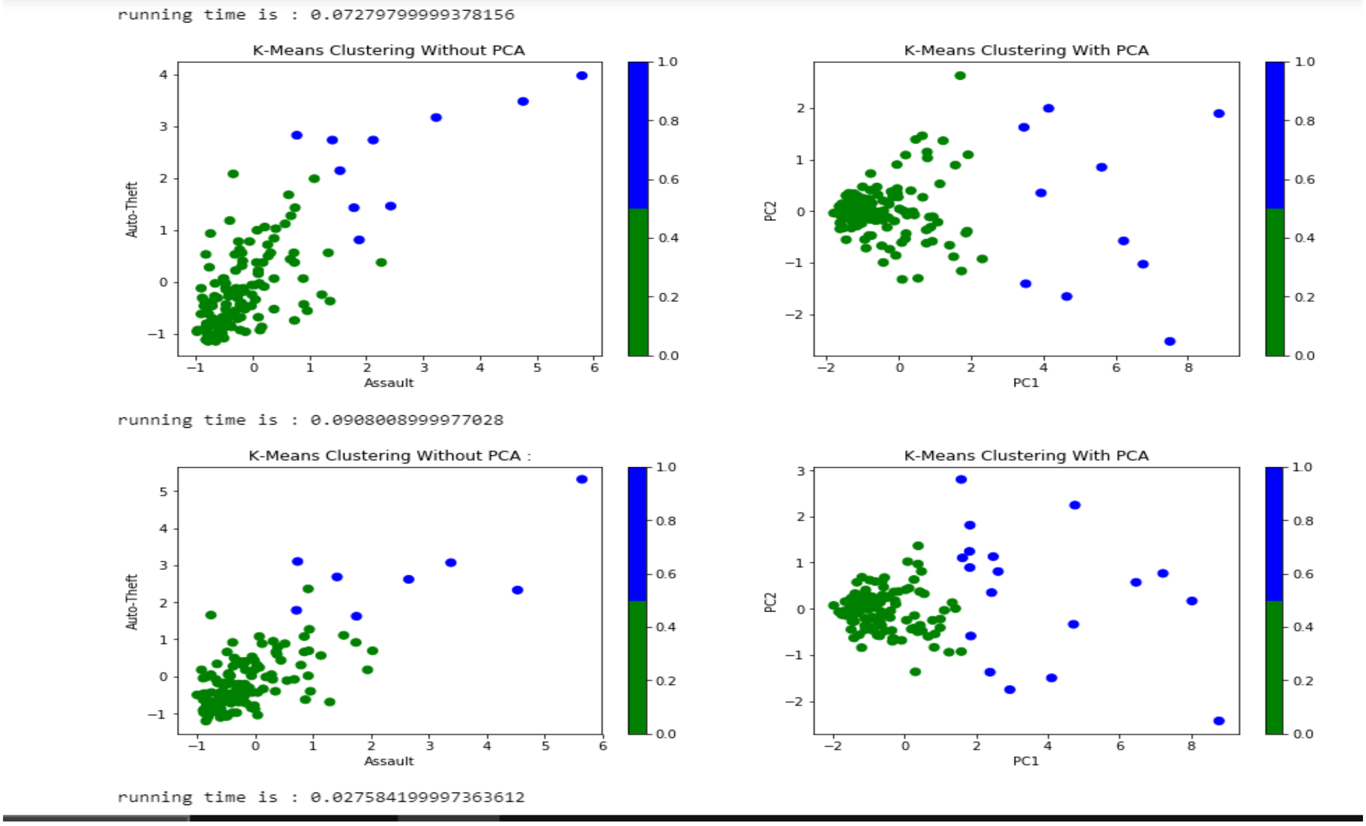


Fig. Visualization Images.



```
running time is : 0.07279799999378156
```

K-Means Clustering Without PCA
Auto-Theft
Assault
K-Means Clustering With PCA
PC2
PC1
1.0
0.8
0.6
0.4
0.2
0.0

```
running time is : 0.0908008999977028
```

K-Means Clustering Without PCA :
Auto-Theft
Assault
K-Means Clustering With PCA
PC2
PC1

```
running time is : 0.027584199997363612
```

Fig. Clustering Images.

## V CONCLUSION

The purpose of this project is to analyze and predict the crimes so that the police can be given an act of hint to take proactive measures before the situation could worsen up. This model helps to analyze and predict crime. Using machine learning approaches, the areas and hotspots can be predicted based on the type of crime and give the overall prediction of any crime. The paper also focuses on building this approach by importing machine learning modules, and time series analyzing methods and also giving the visualizations of the crime.

## REFERENCES

[1] Kiani, Rasoul, Siamak Mahdavi, and Amin Keshavarzi. "Analysis and prediction of crimes by clustering and classification." International Journal of Advanced Research in Artificial Intelligence 4, no. 8 (2015): 11-17.

[2] Akansha A Chikhale, Ankita K Dhavale, Aparna P Thakre, Diksha B Herode, Nikita D Nasre, Pracheta D Patrikar, Prof. Milind Tote. “A Review on Crime Rate Analysis Using Data Mining” International Journal of Scientific Research in Science, Engineering and Technolgoy 5, no. 5 (2019): 119 – 125.

[3] J. Zhou, Z. Li, J. J. Ma and F. Jiang, "Exploration of the Hidden Influential Factors on Crime Activities: A Big Data Approach," in IEEE Access, vol. 8, pp. 141033141045,2020,doi:10.1109/ACCESS.2020.3009969.

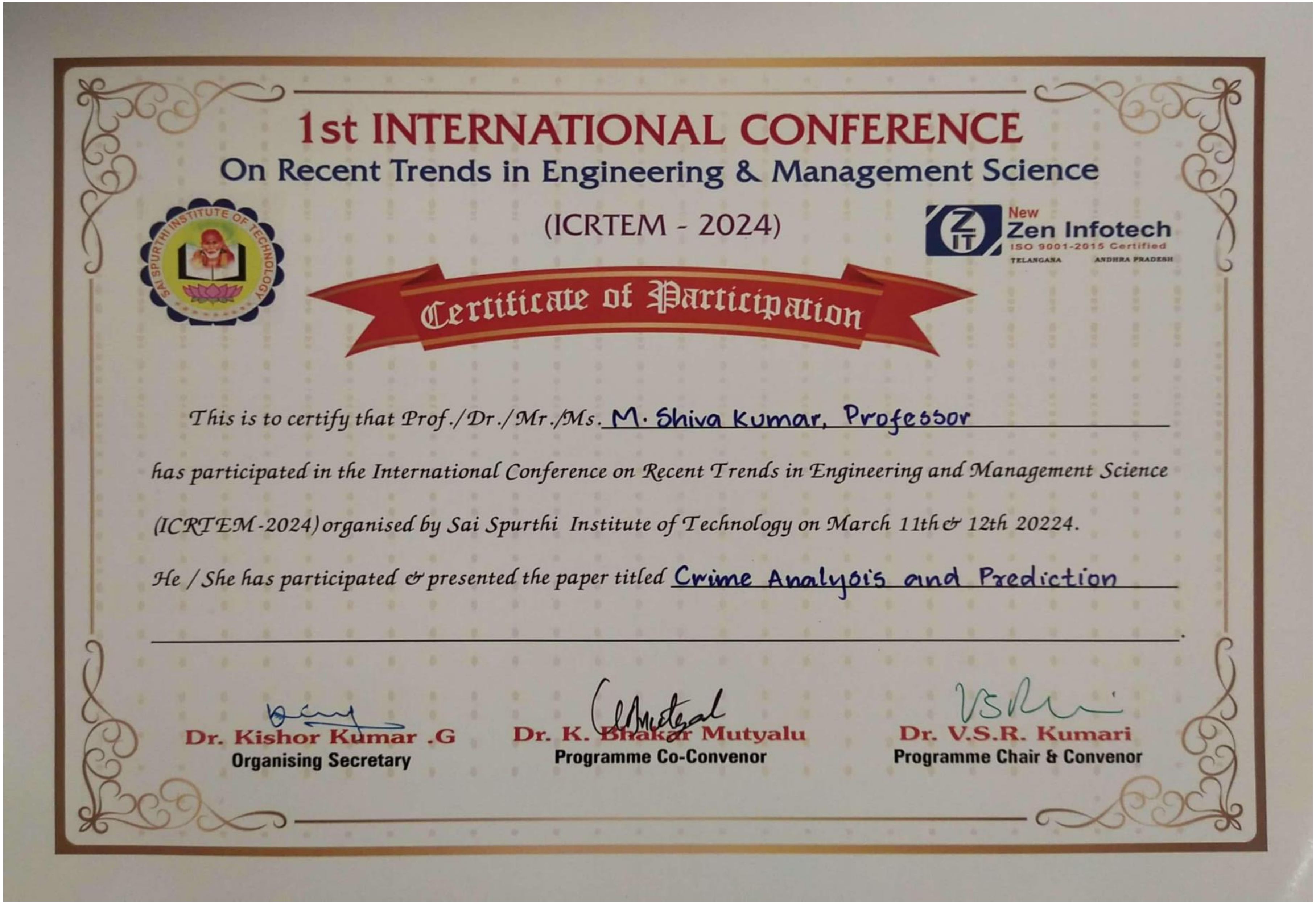
[4] Das, Priyanka, Asit Kumar Das, Janmenjoy Nayak, Danilo Pelusi, and Weiping Ding. "A graph-based clustering approach for relation extraction from crime data." IEEE Access 7 (2019): 101269-101282.

[5] Li Ding, Dana Steil, Matthew Hudnall, Brandon Dixon, Randy Smith, David Brown, Allen Parrish, "PerpSearch: An integrated crime detection system," 2009 IEEE International Conference on Intelligence and Security Informatics, 2009, pp. 161-163, doi: 10.1109/ISI.2009.5137289.

[6] https://bard.google.com/chat

[7] https://ai.meta.com/llama/

[8] Google LaMDA | Discover AI use cases (gpt3demo.com)



# 1st INTERNATIONAL CONFERENCE

## On Recent Trends in Engineering & Management Science

(ICRTEM - 2024)

SAI SPURTHI INSTITUTE OF TECHNOLOGY

New Zen Infotech
ISO 9001-2015 Certified
TELANGANA ANDHRA PRADESH

## Certificate of Participation

This is to certify that Prof./Dr./Mr./Ms. M. Shiva Kumar, Professor

has participated in the International Conference on Recent Trends in Engineering and Management Science

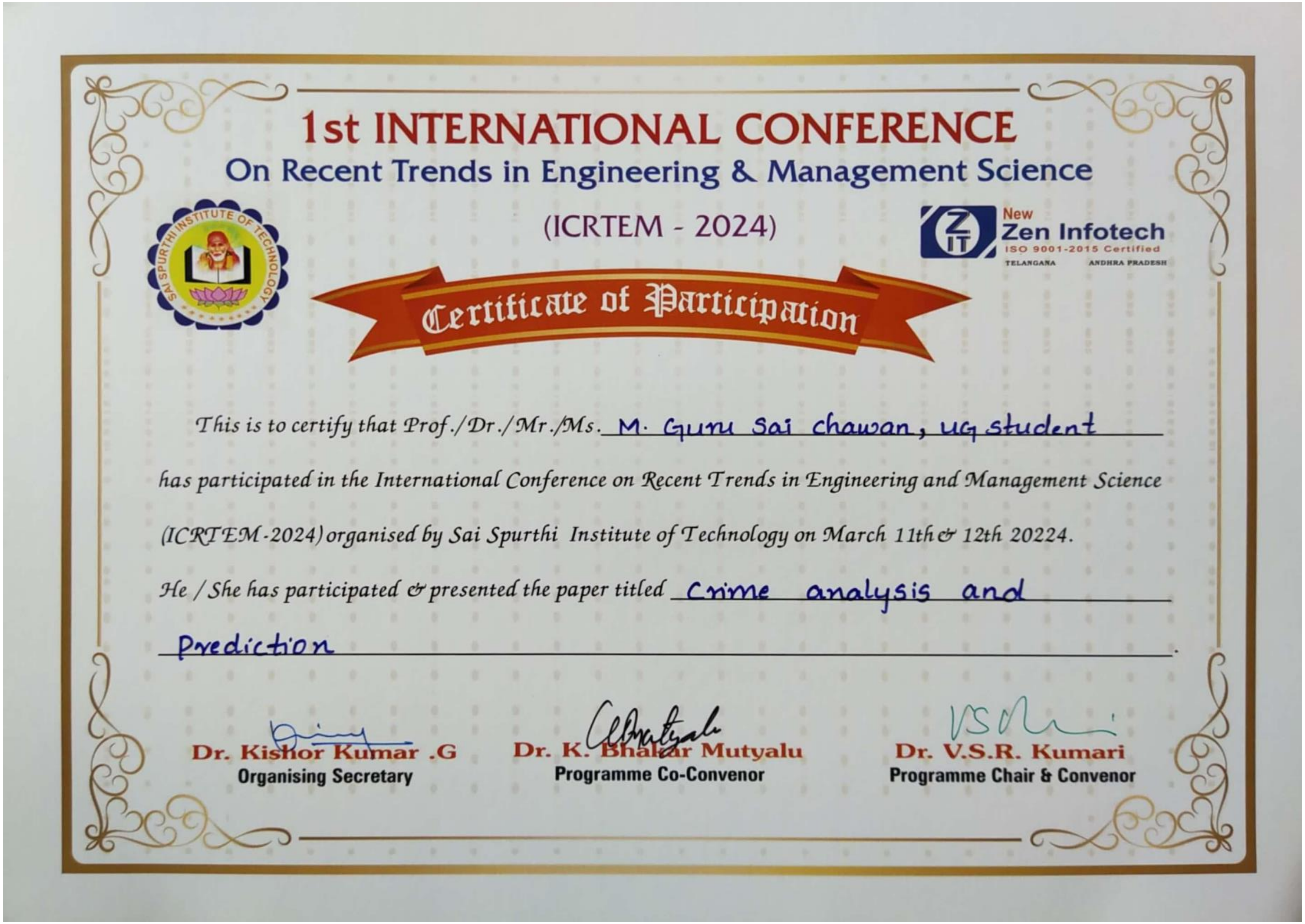
(ICRTEM-2024) organised by Sai Spurthi Institute of Technology on March 11th & 12th 20224.

He / She has participated & presented the paper titled Crime Analysis and Prediction.

Dr. Kishor Kumar .G
Organising Secretary

Dr. K. Bhakar Mutyalu
Programme Co-Convenor

Dr. V.S.R. Kumari
Programme Chair & Convenor



# 1st INTERNATIONAL CONFERENCE

## On Recent Trends in Engineering & Management Science

(ICRTEM - 2024)

SAI SPURTHI INSTITUTE OF TECHNOLOGY
New Zen Infotech
ISO 9001-2015 Certified
TELANGANA ANDHRA PRADESH

## Certificate of Participation

*This is to certify that Prof./Dr./Mr./Ms.* M. Guru Sai chawan, UG student

*has participated in the International Conference on Recent Trends in Engineering and Management Science (ICRTEM-2024) organised by Sai Spurthi Institute of Technology on March 11th & 12th 20224.*

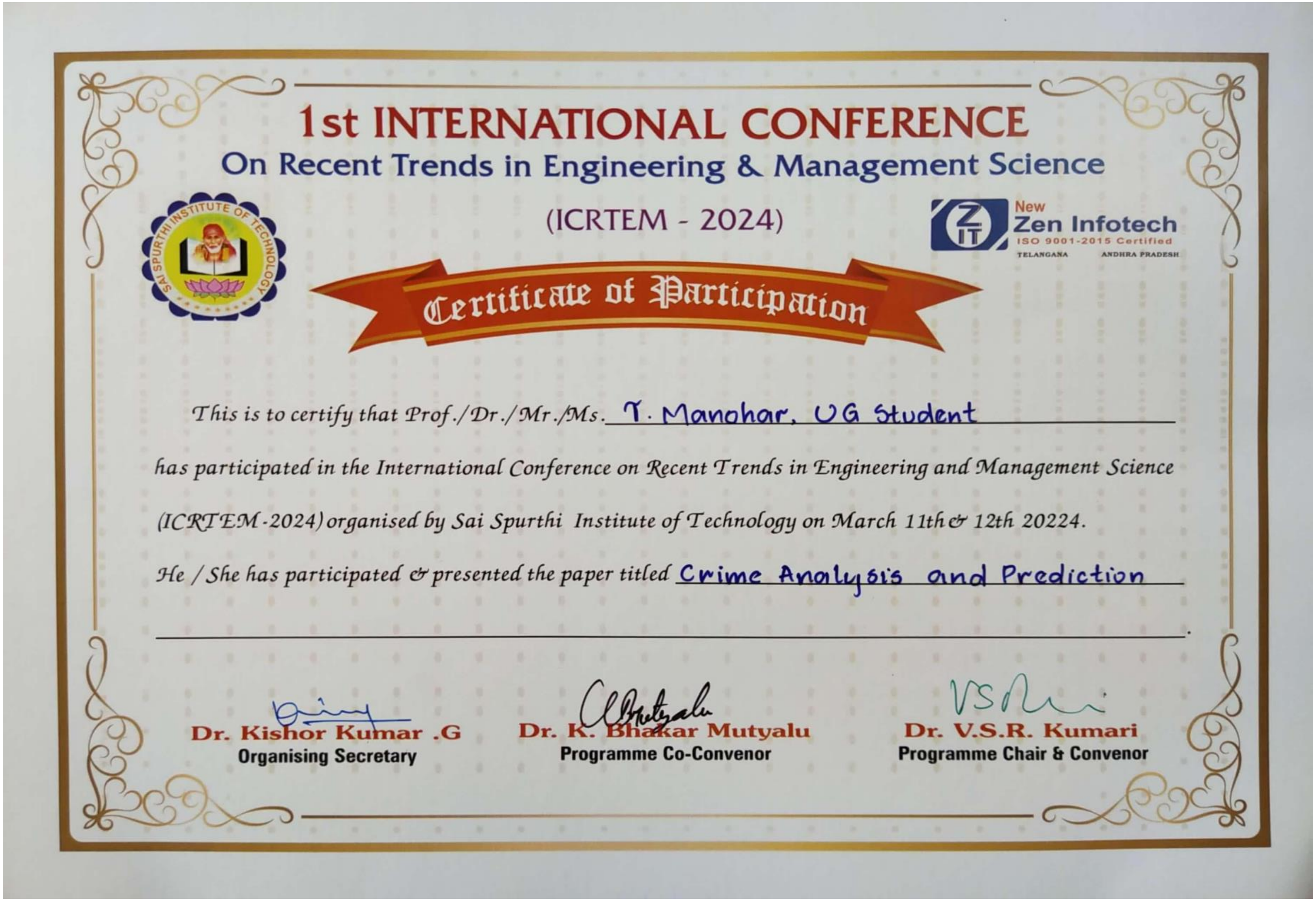
*He / She has participated & presented the paper titled* Crime analysis and Prediction.

VS

**Dr. Kishor Kumar .G**
**Organising Secretary**

**Dr. K. Bhaskar Mutyalu**
**Programme Co-Convenor**

**Dr. V.S.R. Kumari**
**Programme Chair & Convenor**



# 1st INTERNATIONAL CONFERENCE

On Recent Trends in Engineering & Management Science

(ICRTEM - 2024)

SAI SPURTHI INSTITUTE OF TECHNOLOGY

New Zen Infotech
ISO 9001-2015 Certified
TELANGANA ANDHRA PRADESH

## Certificate of Participation

This is to certify that Prof./Dr./Mr./Ms. T. Manohar, UG Student

has participated in the International Conference on Recent Trends in Engineering and Management Science

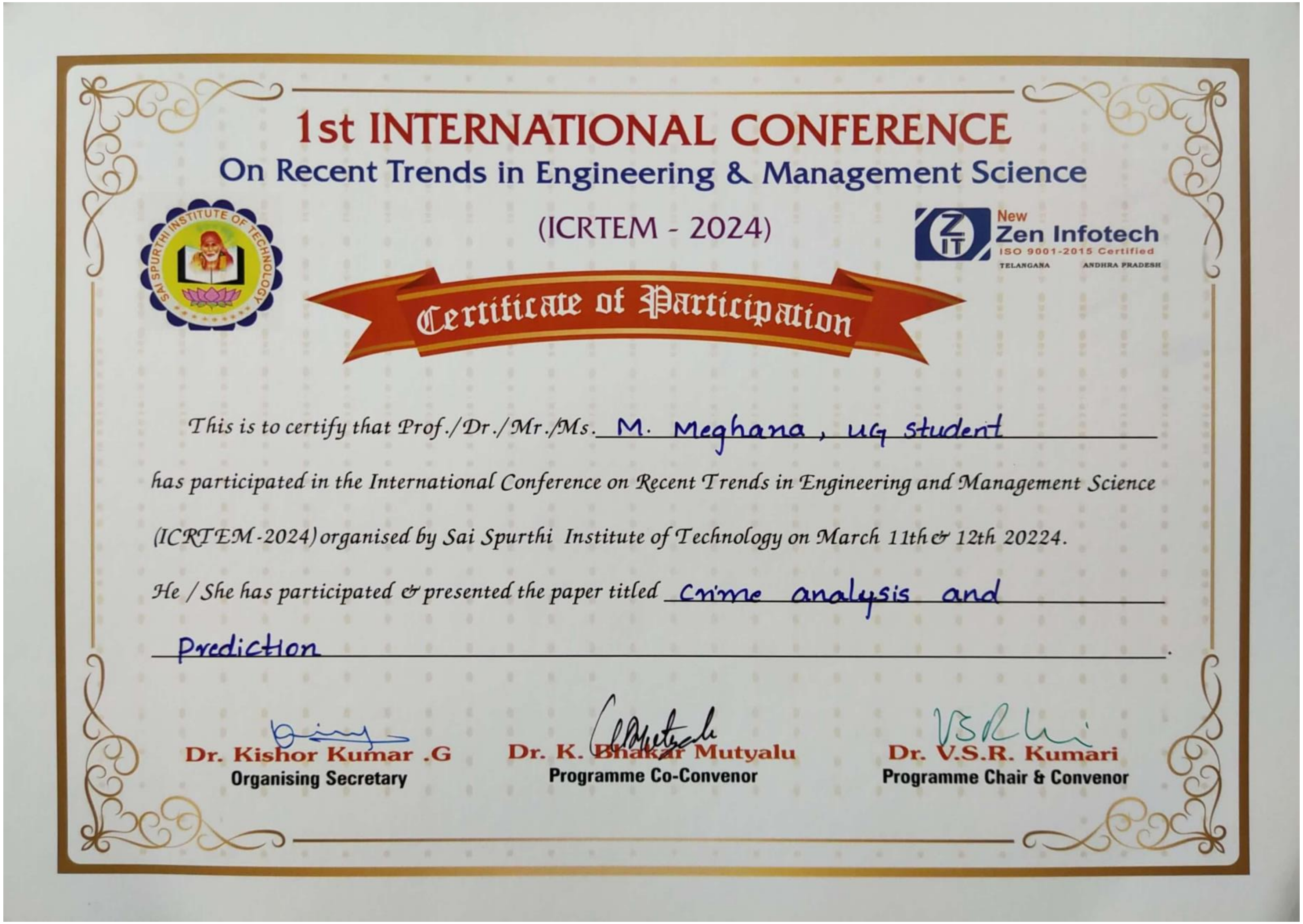
(ICRTEM-2024) organised by Sai Spurthi Institute of Technology on March 11th & 12th 20224.

He / She has participated & presented the paper titled Crime Analysis and Prediction

**Dr. Kishor Kumar .G**
Organising Secretary

**Dr. K. Bhakar Mutyalu**
Programme Co-Convenor

**Dr. V.S.R. Kumari**
Programme Chair & Convenor



# 1st INTERNATIONAL CONFERENCE

## On Recent Trends in Engineering & Management Science

(ICRTEM - 2024)

New Zen Infotech
ISO 9001-2015 Certified
TELANGANA ANDHRA PRADESH

## Certificate of Participation

*This is to certify that Prof./Dr./Mr./Ms.* M. Meghana , UG student

*has participated in the International Conference on Recent Trends in Engineering and Management Science (ICRTEM-2024) organised by Sai Spurthi Institute of Technology on March 11th & 12th 20224.*

*He / She has participated & presented the paper titled* Crime analysis and Prediction.

Dr. Kishor Kumar .G
Organising Secretary

Dr. K. Bhakar Mutyalu
Programme Co-Convenor

Dr. V.S.R. Kumari
Programme Chair & Convenor