FBI Time Series Forecasting

ct Description

ess Context The FBI Crime Investigation Project is a strategic initiative designed to harness the power of analytics to predict crime patterns and improve public safety. In recent years, urban centers across the d States have faced challenges related to rising crime rates and the complex dynamics of crime rrences. To address these issues effectively, law enforcement agencies require advanced tools and odologies that enable them to anticipate crime trends, allocate resources strategically, and implement tive measures to prevent criminal activities.

project focuses on developing a predictive model that estimates the number of crime incidents on a monthly using detailed data at the day and hour level. The data includes information such as crime types, raphical coordinates (latitude and longitude), neighborhood details, and time stamps, allowing for a rehensive analysis of both spatial and temporal patterns in crime data. By understanding when and where s are most likely to occur, law enforcement agencies can optimize patrol schedules, allocate personnel efficiently, and deploy resources where they are needed most.

mportance of predicting crime incidents extends beyond law enforcement. Urban planners, policy makers, community leaders can leverage these insights to create safer and more resilient communities. For instance, rstanding high-risk areas and times for specific types of crime can guide the placement of public safety ures such as street lighting, surveillance cameras, and community policing initiatives. Additionally, these ts can inform public awareness campaigns, emergency response planning, and collaboration with nunity organizations to address underlying social issues that contribute to crime.
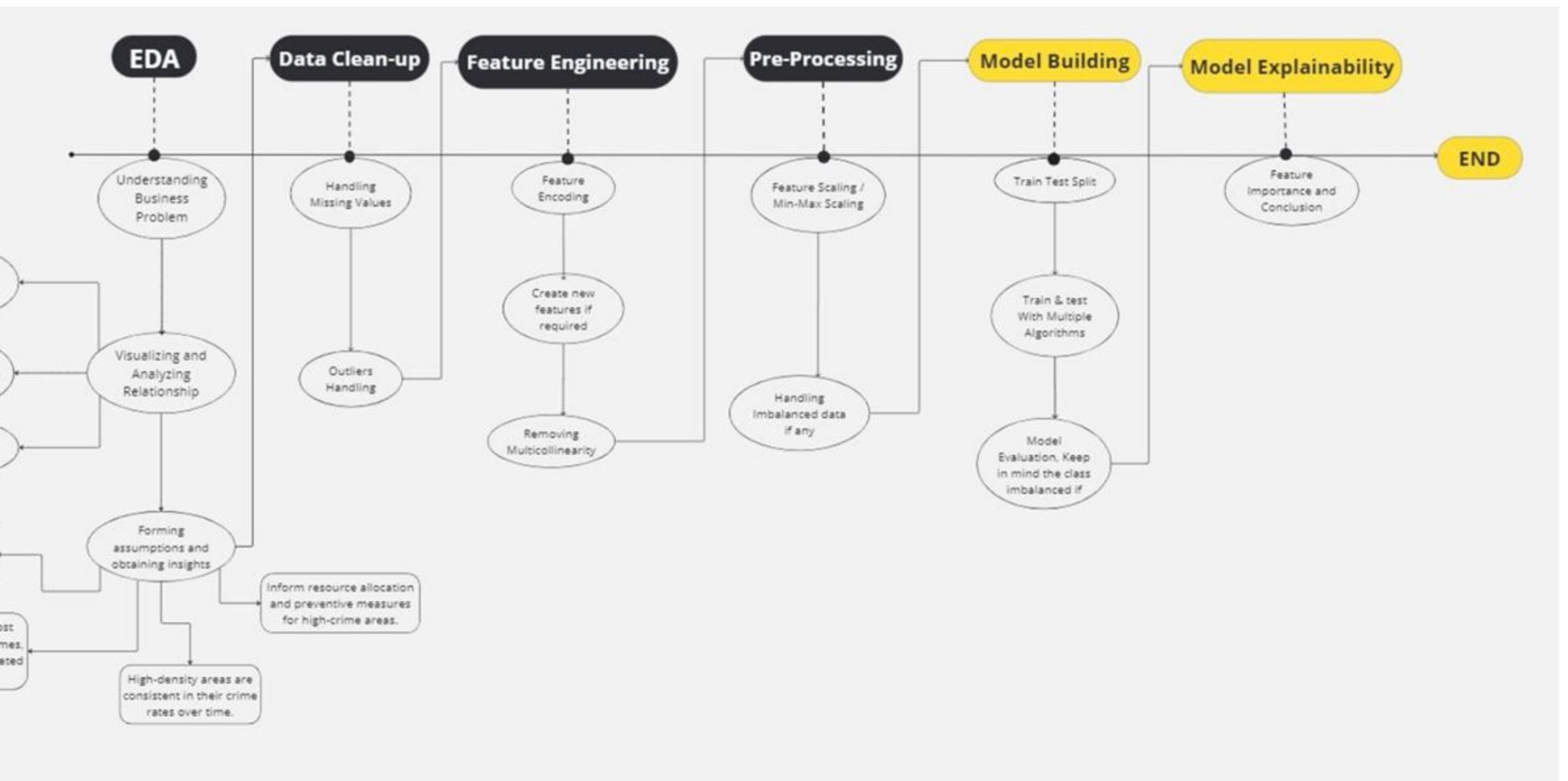
Dataset Description

Train Data:

TYPE	Category of the crime (e.g., "Other Theft").
HUNDRED_BLOCK	Street block where the crime occurred.
NEIGHBOURHOOD	Neighborhood where the crime took place.
X	X-coordinate of the crime location.
Y	Y-coordinate of the crime location.
Latitude	Latitude of the crime location
Longitude	Longitude of the crime location
HOUR	Hour of the day when the crime occurred
MINUTE	Minute of the hour when the crime occurred
YEAR	Year of the crime
MONTH	Month of the crime
DAY	Day of the month when the crime occurred
Date	Full date (YYYY-MM-DD) when the crime occurred

Test Data:

YEAR	The year in which the incidents occurred
MONTH	The month in which the incidents occurred
TYPE	The category of crime
Incident_Counts	The number of crime incidents to be predicted

Main Libraries used:

- **Pandas:** For efficient data manipulation and cleaning, handling large datasets with features like crime types, locations, and time stamps.

- **NumPy:** Used for numerical computations and managing large arrays and matrices, essential for preparing data for machine learning.

- **Scikit-Learn:** A versatile library for implementing machine learning models, preprocessing data, and evaluating model performance.

- **Statsmodels:** Useful for statistical modeling and time series analysis, including ARIMA and SARIMA models for understanding temporal crime trends.

- **XGBoost: **An advanced library for efficient gradient boosting, ideal for predictive modeling with structured data.

- **Matplotlib and Seaborn:** For visualizing spatial and temporal distributions of crime data, and displaying model performance and feature importance.

- **Geopandas:** An extension of Pandas for handling and visualizing geospatial data, enabling analysis of spatial patterns and relationships in crime data.

EDA
- Understanding Business Problem
- Visualizing and Analyzing Relationship
- Forming assumptions and obtaining insights
  - Inform resource allocation and preventive measures for high-crime areas.
  - High-density areas are consistent in their crime rates over time.

Data Clean-up
- Handling Missing Values
- Outliers Handling

Feature Engineering
- Feature Encoding
- Create new features if required
- Removing Multicollinearity

Pre-Processing
- Feature Scaling / Min-Max Scaling
- Handling Imbalanced data if any

Model Building
- Train Test Split
- Train & test With Multiple Algorithms
- Model Evaluation, Keep in mind the class imbalanced if

Model Explainability
- Feature Importance and Conclusion

END

Project Evaluation Criteria

- Understanding the Dataset and Problem Statement

- Efficient Exploratory Data Analysis (EDA)

- Understanding of how to prep the data and make it ready for training.

- Understanding the target feature and its distribution

- Assessing target features for class imbalance.

- Selecting the Approach and Algorithm

- Modeling - which algorithm to use?

- Evaluation while keeping class imbalance in mind.

- Feature Importance and Conclusion

- Understanding how your project is useful to stakeholders

Project Evaluation Criteria

Rubrics

Rubrics	Weightage
Summary and Technical Documentation in Collab Notebook	10
EDA and Visualization	5
Looking for and Handling NaN/ Null/ Missing Values and Outliers	2.5
Finding Correlation in Variables (Both Dependent and Independent, Visuali	10
Pick Appropriate Independent Variables, Test Train Split, Train Model	10
Prediction and Calculate Some Evaluation Metrics for Model	10
Number of Models Experimented (At least 2)	5
Hyperparameter Tuning	5
Final Summary of Conclusion	2.5
Commented Code	5
Proper Output Formatting	5
Modularity of Code	5
Video Presentation	20
Fluency and Grammatical Accuracy in Video	5