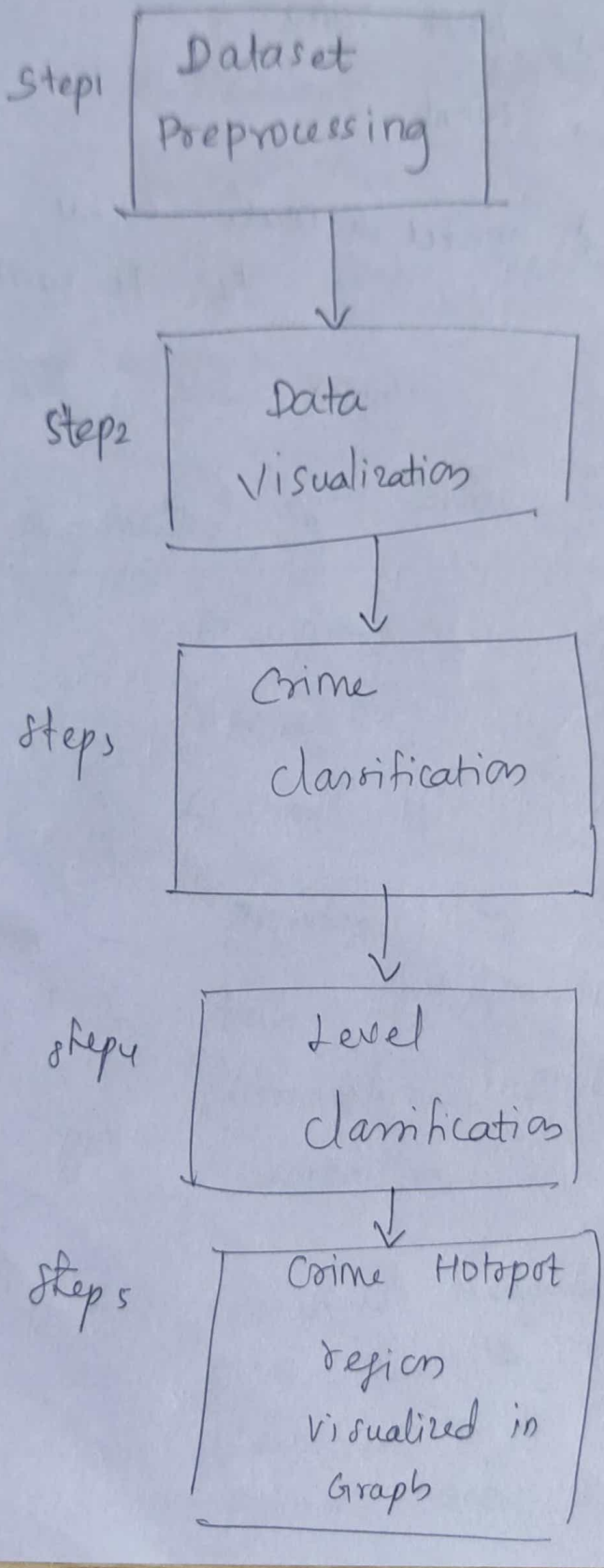
Project Summary

(K. Keerthana
1817123)



Data Prepocessing

* Data preprocessing done to reduce the noise, incomplete and inconsistent data.
* If any missing value exist in the dataset is found and removing the missing values.

Data Visualization:

* Data visualization helps to better understand the crime status happening over the period of time.
* We have done visualization on crime type to identify top 5 crime happening in the city.
* crime happening during hours, month. done to get the insights about unsafe hours.
* crime happened in across the district will be analysed to find the crime rate high in which places and plotted in the map view to see district with crime rate.

Crime Classification:

* Crime classification is done prior to the level classification.

* The purpose of the crime classification is to predict the type of crime happen in that location.

* In the dataset itself the attribute value "primary type" holds the value of the type of crime happen in that area at that time.

* In training phase we have Choose the best algorithm suitable for multiclass classification and trained the algorithmic model with this dataset, with no alteration in the dataset 80 % of the dataset taken for training.

* In testing phase, the trained model is taken to predict the remaining 20 % of data in the dataset, in that the primary type is going to predict which is already present that is removed

* Our model going to predict the value in primary type with the help of other attribute value which are location, time, description and.

Level classification:

* After classifying the crime in the location. Level classification which is representing the level column which is not present in the dataset.

* We are going to add that column and do classification under the 3 categories low, medium, high

* This level classification done based on the crime count in the location. Based on the range of the crime count the level which it comes under will be taken automatically.

* The Algorithms used here to obtain best accuracy against various algorithm approach.

* Algorithms used are will be Random forest classifier, Decision Tree classifier, svm classifier. Naive Bayes classifier, KNN classifier.

Crime Hotspot visualization:

* After predicting the level of the crimes in the cities, the Hotspot regions which are high in crime rates that is mentioned by the "level" attribute with values "high" are visualized as a Hotspot region in the map.