Safer Map Navigation using Crime Prediction in Machine Learning

A

report submitted

in the partial fulfilment of the requirements for the award of the degree of

**Bachelor of Technology**

in

**Information Technology**

By

**Anshika Awasthi 1805213011**

**Ashutosh Yadav 1805213016**

**Shadan Amir 1805213052**

Under the guidance of

## Supervisor Co-supervisor

[Dr. Pavan Kumar Tiwari] [Miss. Mudita Sharan]

**Contents**

DECLARATION (i)

CERTIFICATE (ii)

ACKNOWLEDGEMENT (iii)

ABSTRACT

MOTIVATION

1. INTRODUCTION
2. RELATED WORKS
3. METHODOLOGY
   1. TYPES OF RECOMMENDATION SYSTEM
      1. CONTENT- BASED RECOMMENDOR
      2. COLLABORATIVE - BASED RECOMMENDOR
      3. HYBRID- BASED RECOMMENDOR SYSTEM
   2. K-NN ALGORITHM
   3. COSINE SIMILARITIES
   4. SENTIMENTAL ANALYSIS
   5. NEURAL NETWORK
   6. MATRIX FACTORIZATION
4. EXPERIMENTAL RESULTS
5. CONCLUSIONS
   1. CONCLUSIONS
   2. FUTURE WORKS
6. REFERENCES
7. ANNEXURE

# 

# DECLARATION

We hereby declare that this submission is our own work and that, to the best of our belief and knowledge, it contains no material previously published or written by another person or material which to a substantial error has been accepted for the award of any degree or diploma of university or other institute of higher learning, except where the acknowledgement has been made in the text. The project has not been submitted by us at any other institute for the requirement of any other degree.

Submitted by: - Date-:

(1) Name: Anshika Awasthi

Roll No.: 1805213011

Branch: IT

Signature:

(2) Name: Ashutosh Yadav

Roll No.: 1805213016

Branch: IT

Signature:

(3) Name: Shadan Amir

Roll No.: 1805213052

Branch: IT

Signature:

# CERTIFICATE

This is to certify that the project report entitled “Safer Map Navigation using Crime Prediction in Machine Learning” presented by, Anshika Awasthi, Ashutosh Yadav and Shadan Amir in the partial fulfillment for the award of Bachelor of Technology in Information Technology, is a record of work carried out by them under my supervision and guidance at the Department of Computer Science and Engineering at Institute of Engineering and Technology, Lucknow.

It is also certified that this project has not been submitted at any other Institute for the

award of any other degrees to the best of my knowledge.

Dr. Pawan Kumar Tiwari Miss. Mudita Sharan

# 

# ACKNOWLEDGMENT

We would like to express our deepest sense of gratitude towards our mentors, Dr. Pawan Kumar Tiwari, and Miss. Mudita Sharan for their guidance throughout the project development phase. Their views have always been a guiding factor for us in creating a earning stack and acting accordingly. This project has given us great opportunities to push ourselves ahead of our respective comfort zones and work with completely unfamiliar technologies. Lastly, we would love to thank our families and friends for their motivation and help whenever required.

Regards,

Anshika Awasthi

Ashutosh Yadav

Shadan Am

**Abstract**

Crimes are treacherous and common social problem faced throughout the world. Crimes affect the quality of living, economic growth, and reputation of a nation. There has been an great increase in crime rate in the last few years. In order to cut down the crime rate, the law enforcements need to take strong preventive measures. With the aim of saving the society from crimes, there is a need for advancement in systems and new approaches for improving the analysis of crime for protecting their communities. Accurate real-time c predictions of crime help to reduce the crime rate but remains challenging problem for the scientific community as crime occurrences depend on enormous number of complex factors. In this work, various visualizing techniques and ML algorithms are adopted for predicting the crime distribution over an area depending on multiple factors. In the first step, the raw and the basic datasets were processed and visualized based on the needs. Afterwards, machine learning algorithms were used to extract the knowledge out of these large datasets and discover the hidden relationships among the data which is further used to report and discover the crime patterns that is valuable for crime analysts to analyze these crime networks by the means of various interactive visualizations for crime prediction and hence is supportive in prevention of crimes.

## 

## Motivation

Ever-increasing crime levels in our country make communities decline in manyfold ways. To reduce and stop crimes it is very important to identify the major reasons behind crimes, predict crimes, and prescribe solutions. Through effective preparation, it is only possible that necessary precautions can be taken on time. The recent advent of data has given us opportunities to harness its powers in ways that were only imaginable a couple of decades earlier. Due to large volumes of data and the number of algorithms needed to be applied to crime data, it has always been unrealistic to do manual analysis. Therefore, itis necessary to have a platform that can apply any algorithm required to do a descriptive, predictive, and prescriptive analysis on large volumes of crime data. Through those three methodologies, it would be possible for law enforcement authorities to take suitable actions to prevent the crimes. Moreover, by predicting the highly likely targets to be attacked, during a specific period and specific geographical location, police will be able to identify better ways to deploy the limited resources and to find and fix the problems leading to crimes. Several applications are already being developed for crime analysis and reporting. Most of these tools are developed to help the police to identify different crime patterns and even to predict criminal activities and work accordingly. These are mostly complex software that needs a lot of training before use. Incorporating this helping hand into a user-friendly Google Map would make this technology extremely accessible and reliable for the public.

# Report

## Introduction

Diversity defined as a core characteristic of Indian culture, unfortunately, is not limited only to languages, religions, politics, or ideology. The term very well extends itself to the variety and volume of crimes committed in the national parameter day today. As the world locked itself up in the light of a global pandemic, India reported an increase of 28.0% (14,45,127) in the registration of cases over 2019. The crime rate registered per lakh population showed a dramatic increase from 385.5 in2019 to

487.8 in 2020.

As much as these enormous digits are intimidating, they can be harnessed to prevent their onset in the first place. This crime data can be analyzed through rigorous machine learning models to identify intricate patterns and predict the likelihood of such upcoming events with high precision. This can aid individuals to avoid being caught up in such scenarios and at the same time help government bodies to be better prepared to respond to such threats.

While the scope of such projects is multifaceted, their accessibility remains only a promise. Reliable software is still limited to very few multinational tech companies. The prospect of such solutions being available to the public is farfetched to the date.

Through this paper, we aim to develop through extensive research, a usable and reliable solution to the aforementioned problem using Maps. The current map applications employ the best of world- class technologies to suggest routes that reduce the navigation duration to the minimum. While they incorporate factors like real-time traffic, they fail to consider safety as the most crucial aspect of this experience. We attempt to bridge this gap and make navigation more comfortable and safer for billions of people that rely on maps to guide their ways.

This paper proposes a web mapping & visualization-based crime prediction tool which is built in RgoogleMaps, googleVis, etc. The proposed model uses different visualization techniques to show the defined analysis of crimes and various ways that can predict the crimes using machine learning algorithms. The work has followed the steps that are mostly used in Data Analysis, where the important phases are the Data collection, Data pre-processing, Data Visualization and Model building which are discussed more in detail in the following sections. In brief, in data collection phase the data is obtained by scrapping data from different news websites using python libraries. The data pre-processing phase consists of cleaning and transformation of data as needed. The visualization phase generates various reports and maps for diagnosis and for proper analysis and finally, in model building phase various machine learning algorithms are used for classification of crime into different categories that can happen in a particular location.

# OBJECTIVES OF THE PROJECT

→The main objective of the project is to predict the crime rate and analyze the crime rate to be happened in future. Based on this Information the officials can take charge and try to reduce the crime rate.

→The concept of Multi Linear Regression is used for predicting the graph between the Types of Crimes (Independent Variable) and the Year (Dependent Variable)

→The system will look at how to convert crime information into a regression problem, so that it will help detectives in solving crimes faster.

→Crime analysis based on information available to find crime patterns. Using different multi linear regression techniques, frequency of occurrence of crime can be predicted based on territorial distribution of existing data and Crime recognition.

## 

## Related works

Though the advent of the Big Data era and the availability of fast, efficient algorithms for data analysis has created a spur in the research work being done in the field of understanding patterns in this crime data, the greatest hurdle is the collection of data. Major database problems can be compiled as:

* For extracting relevant information, one must sift through piles of data manually.
* More time is needed in searching for the required information.
* A large number of hard copies are required to be referred.
* Access to it is not available anywhere.
* It is not possible to keep all the records in memory or a file because of the large data which they comprise.
* A large amount of data maintenance is the big problem; searching and predicting a particular data manually is not possible from the files.

Several projects have employed used Apriori and KNN algorithms to effectively predict upcoming events, but they’re limited to complex web apps. There is no existing system for this project

## Methodology

## MACHINE LEARNING:

Machine Learning is a sub-area of artificial intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by searching and matching patterns in databases. In other ways or words: Machines. Learning enables IT systems to recognize patterns based on existing algorithms and data sets and to develop adequate solution concepts. Therefore, in Machine Learning, artificial knowledge is made based on experiences.

Independently generate solutions, the prior action of people is necessary. For example, the algorithms which are required and the data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be pre-defined. Once these two steps have been completed, the system can perform the following tasks by Machine learning.

Diagram

Description automatically generated

Fig. Machine learning processes.

**Modelling**

The following are the objectives of the proposed framework:

* + To investigate the current crime rate, examine data framework, and recognize limitations and issues.
  + To structure another or improved crime examination framework for data.
  + To test the new or improved framework.

To distinguish conditions for effective execution of the framework. The following figure shows the structure of the proposed framework.

## SYSTEM ARCHITECTURE & METHODOLOGY



Data set

Data Cleaning

Data Pre- processing

Data transformation

Converting to Numerical Dataset

Train Data

Data selection

Using KNN

Algorithm

Predicting Result

The methods used in crime rate prediction are:

* Data Collection
* Data Pre-processing
* Using KNN Algorithm
* Result

## DATA COLLECTION:

We have chosen datasets for predicting crime rates, and we'll use the necessary attributes for training our model over it. We obtained the data from various online sources. The dataset is in the form of 878049\*9. The different columns in our dataset depict the specific date on which the crime was reported, the types of crimes, the summary or significance of the crime, the days on which the crime occurred, the district, and the address where the crime has occurred.

## DATA PREPROCESSING:

The process of transforming raw data into the appropriate format is known as data preprocessing. Data cleaning, data reduction, and data integration are all examples of data preprocessing methods. The data is gathered from different sources, each with its collection of attributes and data. We convert the raw data into the appropriate format because some data is either replicated or duplicated, contains some attributes that are not for our interest, and contains NULL values.

**VISUALIZATION OF DATA USING GOOGLE MAP:**

This module extracts the recent crime data from the dataset and based on longitude-latitude it tags the specific location of the city. This tagging also displays the crime location name, the type of crime that happened. This information is useful for an individual in knowing dangerous and risky areas and it thus can help them to avoid such areas. The picture can help the law enforcement to improve the security in the areas. Fig.1 shows that locations, where crimes occurred, are very near to each other. From this, we can analyze that if a location feasible to a criminal attack, then the nearby locations are also feasible for the crime to occur. This module also provides the facility to enquire about a specific location to show what type of crime is feasible to happen in that location.

**CRIME PREDICTION***:*

Data mining involves exploring the datasets and extracting the fruitful information to transform into an understandable form for the further use. The data mining techniques were applied to the crime data for the crime prediction based on theories in Criminology. The criminology mainly focuses on the Rational Choice Theory and Routine Activity Theory. The Rational choice theory focuses on the understanding of crimes from offender**’**s perspective which is directly concerned with thinking process of the offender and how they evaluate their opportunities.

The Routine Activity Theory states that for a crime to occur, a likely offender finds a suitable target with capable guardians absent and states that crimes are unaffected by social causes such as poverty, inequality, and unemployment. The criminals repeat their activities by choosing the targets which are under similar conditions. Based on this information the work provides the use of the following algorithms.

## CLUSTERING ALGORITHM:

Because of the variety of applications, clustering algorithm has become and gained a rising research problem in the field of data processing. Many data and their meaning have been implemented by cluster algorithmic programs in recent years, and they are now used in a broad range of applications such as image manipulating, processing micro-objects, visual converse, narcotics, and political economy, which contributes to the quality of the algorithms. Cluster algorithms have the main advantage as that they cannot be ordered. Also, the algorithmic program progress will produce spectacular results when used in conjunction with an information set, but it will not produce accurate results. Clustering algorithms are usually unsupervised learning algorithms. It is more like data pre-processing. There is an update daily. Only a few people can work with data. It is helpful at the beginning of every module and is also known as data pre-processing. Whenever there is any chance for clustering as many data items as possible, clustering in data mining is a very crucial algorithm as the result which is going to be obtained is in the form of a scalar.

**Logistic Regression:**

In statistics, the logistic model is used for modelling the probability of a certain class or event existing in a manner such as a pass or fail, win or lose, alive or dead, or healthy or sick. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, considering many more complex extensions exist.

In regression analysis, logistic regression or logit regression can be said as estimating the parameters of a logistic model which is a form of binary regression. Mostly Mathematically, a binary logistic model has a dependent variable with two possible values, such as pass or fail which is represented by an indicator variable, where the two values are labelled "0" and "1".

**Multiple linear regression (MLR):**

Also known simply as multiple regression, is a statistical technique that uses several explanatory variables to presume the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the linear relationship between the number of predictor variables.

In essence, multiple regression is the extension of ordinary least-squares (OLS) regression that involves more than one explanatory variable.

The goal of multiple linear regression is to model the linear relationship between the explanatory/independent variables and response/dependent variables in the system. So far, we have seen that simple linear regression is used at places where a single predictor variable X was used to model the response variable Y. In many applications, mostly multiple factors influence the response. Multiple regression models thus describe how a single response variable Y depends linearly

Table

Description automatically generated with medium confidence

**K-NEAREST NEIGHBOUR**

Steps to take when using the K-means algorithm are as follows:

Assume that X represents the data clusters as a vector and the V represents the points of centers assets.

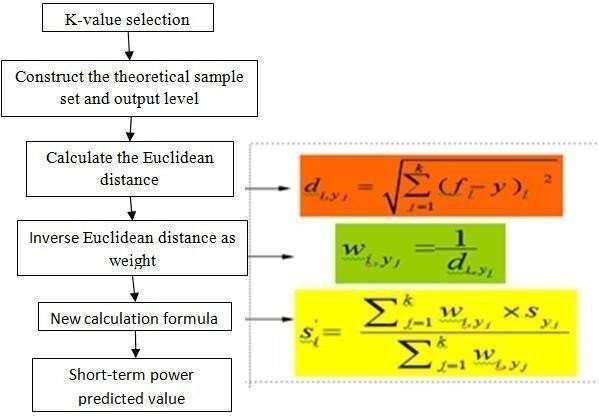
1. Clustering centers are denoted by the letter ‘c.'
2. Measure the distance between cluster centers and each data point.
3. Assign the information objects to the cluster center with the shortest distance from all other cluster centers.
4. Build a new cluster core using the formula.
5. Measure the distance between each cluster center and each data point.
6. Repeat from step 3 onwards again, if no data points are assigned.

**K-NN vicinity calculation map**

Diagram

Description automatically generated

K-means is one of the most effective methods for clustering problems. The k clustering centroids can be used to mark new data entries. This approach uses a specific clustering technique, let us assume k clusters here to classify a given data set. The first step is to designate k centers, each for a single cluster. Because of the absolute difference in position and the completely different outcome, the centers should be positioned correctly. The best choice is to position them as far apart as possible from each other clusters. The next step is to assign each aim happiness to a specific data set and link it to the nearest center. The next move is taken until no intention remains unfulfilled. We tend to re-calculate k new centroids now. There is a loop formed. As a result, the k centers change the positions one by one until all of them have been changed. Subsequently, this formula focuses to minimize the associate objective function which grasps as a square error function.



**1.Weighted KNN Algorithm (WKNN)-**

**W**eighted KNN is a modified version of [k nearest neighbors](https://www.geeksforgeeks.org/k-nearest-neighbours/). One of the major issues that affects the performance of the kNN algorithm is the choice of the hyperparameter k. If k is too small, the algorithm would be more sensitive to outliers. If k is too large, then the neighborhood may include too many points from other classes.  
Another issue is the approach to combining the class labels. The simplest method is to take the majority vote, but this can be a problem if the nearest neighbors vary widely in their distance and the closest neighbors more reliably indicate the class of the object.

**Algorithm:**

* Let L = { ( xi , yi ) , i = 1, . . . ,n } be a training set of observations xi with given class yi and let x be a new observation(query point), whose class label y has to be predicted.
* Compute d(xi, x) for i = 1, . . . ,n , the distance between the query point and every other point in the training set.
* Select D’ ⊆ D, the set of k nearest training data points to the query points
* Predict the class of the query point, using distance-weighted voting. The v represents the class labels. Use the following formula

Text

Description automatically generated with low confidence

## GAUSSIAN KERNEL:

## Gaussian Kernel uses the formula below-

## Text Description automatically generated

## w is the weight, d(a,b) is distance between a and b. σ is a parameter we set. The query point is the point we are trying to estimate, so we take the distance of one of the K-nearest points and give its weight to be as Figure.

Gaussian kernel [10] is a prevalent portion work that is regularly utilized in different factual example acknowledgment looks into and their applications to quantify the similitude between two examples in a dataset. Nonetheless, the scalar parameter beta (β) in Gaussian portion work as appeared in the accompanying condition fundamentally influences the last outcomes. This online information driven device is utilized to compute the ideal beta (β) parameter dependent on the given dataset. When utilizing this online administration, the client is required to present their dataset with a record name in.txt.

Diagram

Description automatically generated

#### 

#### 

#### Flow Chart of Train and Test Information of Sample Kernel

## SYSTEM CLASSIFICATION:

The outline of the aforementioned Machine Learning techniques is as follows–

* Regulated Learning
* Unaided Learning
* Semi-managed Learning
* Support Learning

## Regulated Learning:

In different cases, we can prepare the calculation for these as the marked models as information where the ideal outcome is already known in advance. Along with contrasting exact results, a learning estimate can give a lot of knowledge guidelines. If any error occurs, the genuine outcome is being contrasted by the learning calculation which conflicts with the exact result and shows the warning. Driven learning uses a variety of ways, such as relapse, arrangement, angle boosting, and predicting, to proactive validation of the estimation of a name on additional unlabeled

Type of data. This is the main reason why this approach is widely used in environments where prerequisite knowledge is used to forecast events that are most likely to occur in the future. Consider prediction whether a credit card transaction will most likely be corrupt, or which insurance clients will most likely file claims.

## Unaided Learning:

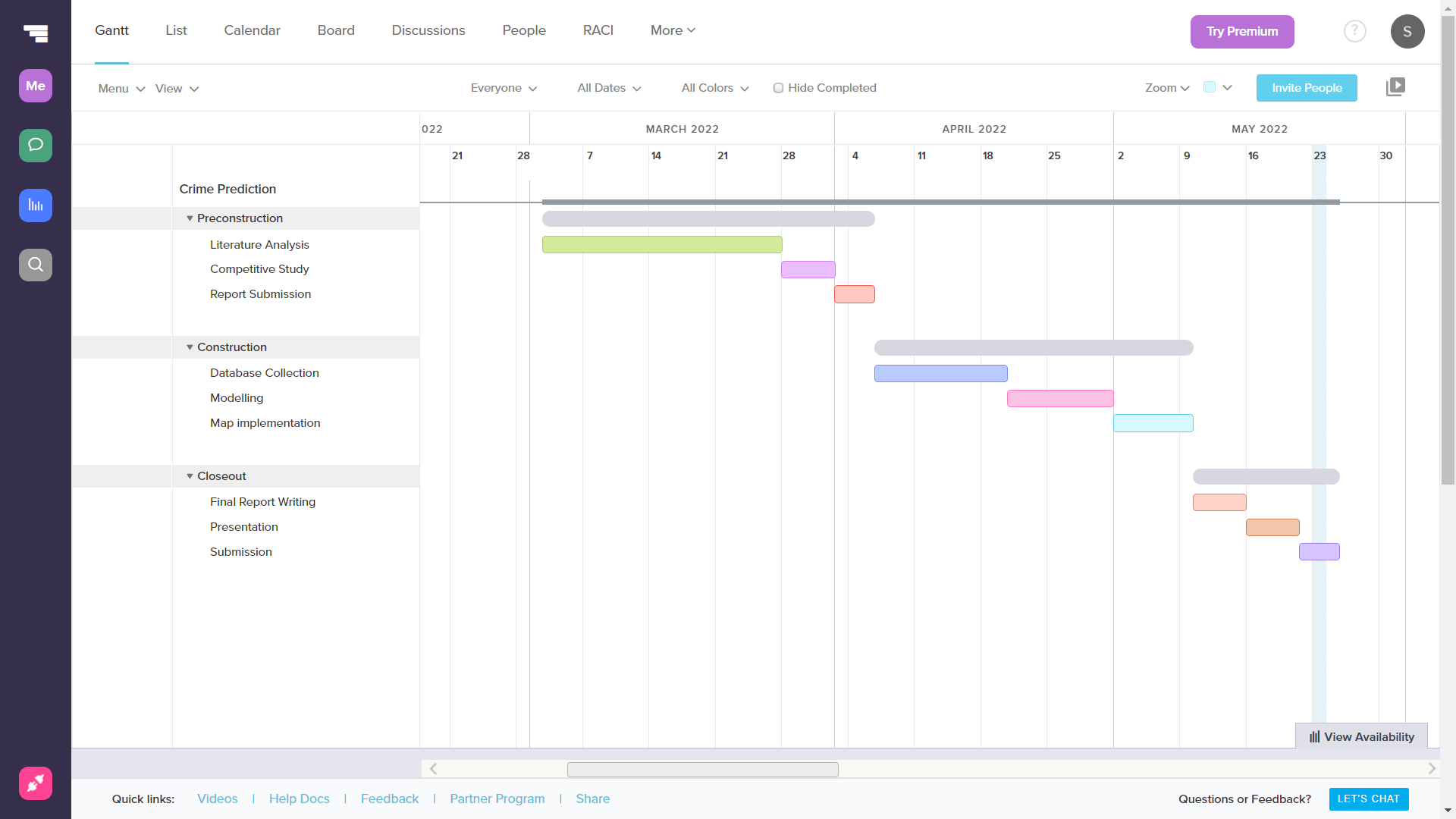
This machine learning method finds its use in places where knowledge is undocumented. In this case, the framework wouldn’t be provided with the "valid answer" and those of calculation must discern what is shown. Thus, the main goal here is to look over the data and thus find an example as well as structure within the available information indexes. For instance, such learning figures out client segments with the most similar characteristics and identity and also helping the organization to treat them equally in regard to the marketing efforts. Also, it can judge properties that differentiate client fragments from each other. In any case, it has to be done with identifying a similar structure in the available informational set. Additionally, these calculations may be figured out the exceptions which are available in informational indexes.

## Semi-Managed Learning:

Semi-managed learning is used and implemented in almost similar circumstances where supervised learning is required. Nonetheless, what is worth noting is that its strategy prepares using both unlabeled and labeled data. In an ideal world, a set which is small and is of named data will be used alongside a large amount of unlabeled data, so getting data unlabeled requires less time, money, and effort. This form of AI is commonly used in combination with techniques such as relapse, characterization, and expectation. Organizations that believe it would be difficult to meet the substantial costs associated with a structured learning process opted for semi-directed learning.

## Plan of work

The work schedule that we followed for the proper execution of the project is as follows:



**• Literature Analysis:** This includes a detailed analysis of the prominent work being done in the field. This accounts us an in-depth idea about the procedure and outcomes that people are achieving throughout the globe working on similar problem statements. This also included the primary and the secondary research. The secondary research included intricate analysis at the external level- works and interviews already done in the field. And the initial research included local research and taking interviews of people we did personally to develop a personal stance of the problem.

* **Problem statement:** We employed the ‘Double diamond’ approach to define our problem statement more specifically. Initially, accumulated a large range of knowledge about the subject and the problem and then finally converged on the aspects of the problem specifically of our attention.

**• Database collection:** This phase included a collection of crime data from different sources such as government websites and police/crime bureau portals. This is a large task consisting of unorganized and scattered data points that are available online.

* **Modeling:** This includes structuring and planning of the database collection and application of all the algorithms to predict potential threats as explained in the previous sections of the report.
* **Map Implementation:** The final stage of the project includes the application of predicted data to the map API to develop an easily accessible tool for safer navigation.

Chart, waterfall chart

Description automatically generated

# Experimental Setup

Data has 800k observation of 9 different variables. There are atotal of 39 categories or sections of the crime in India. The Data seems largely distributed in terms of frequency of the occurrences. As the top 4 categories of the data are composed of almost 53% of the data and the rest of the 47% is comprised of the remaining categories. The data has the maximum occurrence of 174900 for larceny and just 6 for TREA (trespassing and loitering in industrial area). The results came when running on R tool with i3 processor and 4 GB Ram system.

1. Jupyter Notebook Version

Version 5.7.8 is used in our Project.

The Jupyter notebook is a tool which we can use for our machine learning project and analysis statistically. We can also download anaconda from the web source and within it Jupyter notebook most useful tool for machine learning purpose.

2. Python Version

Python 3.7.7 is used for this project.

Python one of the most useful programming language. It is object oriented and interpreted. It is a high-level language. There are lots of in-built libraries in Python for machine learning purpose which we can use easily.

1. Text Editor

Atom Version 1.47.0 is used here for better visualization of the code structure and understanding of the code.

1. Windows Version

Jupyter notebook and python 3 can only be used is all the operating systems including Windows, iOS and Linux.

It is most useful in Linux but can be used in windows also.

It can be run on windows XP, vista,7,8 and the latest version windows 11 as well.

# 

# Results

Indore is probably one of the busiest cities around the globe and it is unfathomably hard to tell what is happening around the city. Given the quality of police officers, we cannot just send the police in same manner around the city for set rounds. A portion of the areas shown have a high wrongdoing rate while others are not as high.

To explain the occurrences of crime, we excessively enhance the crime episodes longitude and latitude on the Indore map.

Results can be obtained in five models i.e., KNN with weighted, KNN with Optimal, KNN with Gaussian, KNN with Cosine and KNN with Inverse. The given below table gives the comparison of many different models. The given below table shows the comparison results of various methods

#### Table 1. Results comparison of various methods

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Training Accuracy** | | **Cross Validation Accuracy** | |
|  | **TURE** | **FALSE** | **TURE** | **FALSE** |
| KNN with Weighted | 0.6771347 | 0.3228653 | 0.2458039 | 0.7541961 |
| KNN with Optimal | 0.6151592 | 0.3848408 | 0.2493673 | 0.7506327 |
| KNN with | 0.5663381 | 0.4336619 | 0.2622219 | 0.7377781 |
| Gaussian KNN with Cosine | 0.626049 | 0.373951 | 0.2526642 | 0.7473358 |
| KNN with inverse | 0.7670175 | 0.2329825 | 0.2634874 | 0.7365126 |

**Timeline

Description automatically generated**

**Graphical user interface, text

Description automatically generated**

**TRAINING AND TESTING ANALYSIS OF EXPERIMENTAL RESULTS**

During Classification, KNN with various kernels connected on Indore Crime dataset. Information is approved through Training and Cross-Validation criteria.

Machine learning is an extremely ground-breaking field with regards to Artificial intelligence and, if a model is done well, the degree of precision that a few calculations accomplish can be perplexing and confusing. Unquestionably, the present and eventual fate of wise frameworks experiences ML and enormous analysis of information. During classification, KNN with multiple portions connected on Indore Crime dataset. Information is proven through Training and Cross-Validation of data criteria.

## 

## 

## CONCLUSION

Machine learning is an extremely ground-breaking field with regards to Artificial intelligence and, if a model is done well, the degree of precision that a few calculations accomplish can be perplexing and confusing. Undoubtably, the present and eventual fate of wise frameworks experience ML and huge analysis of information. During classification, KNN with multiple portions connected on SF Crime dataset.

Information is proven through Training and Cross-Validation of the data.

In Training criteria, Training accuracy of the KNN model with Inverse piece (76% accuracy) would give more accuracy than other portion models, for example, Weighted, Gaussian and Cosine, though during (Testing) Cross-Validation process, Validation accuracy is the practically comparative for Gaussian and Inverse kernels – 26% accuracy when contrasted and other kernel models.

**FUTURE SCOPE**

## 

## This paper presents the visualization techniques and classification algorithms that can be used for predicting the crimes and helps the law firms. In coming future, there will be a plan for applying other classification algorithms on the crime data and improving the accuracy in prediction. In another direction, we will be trying to build an Android App for the live capture of the realistic data and updating the results by using this new data frequently, that would help us in better prediction and providing the general information to the public for the awareness of trends in the crime.

#### While, there is little reason to believe that the crime rate will increase dramatically in the first decade of the 21st Century, given that the anticipated increases in the globalization, sophistication, and organization of crime, one may come to a conclusion that the impact of crime on Western societies may be more severe than the one witnessed under a similar rate of crime in the past. The goal of any society should not be just to catch criminals but to prevent crimes from happening in the first place.

1. Predicting future crime spots.

2. Predicting who will commit the crime.

## References

* ToppiReddy, Hitesh Kumar Reddy, Bhavna Saini, and Ginika Mahajan. "Crime prediction & monitoring framework based on spatial analysis." Procedia computer science 132 (2018): 696- 705.
* https://ncrb.gov.in/en/Crime-in-India-2020 - Volume I Statistics.
* Way Better! - Find the Safe Way App Service- https://[www.behance.net/gallery/67893883/Way-Better-Find-the-Safe-Way-AppService](http://www.behance.net/gallery/67893883/Way-Better-Find-the-Safe-Way-AppService)
* Aghababaei, Somayyeh and Masoud Makrehchi. “Mining social media content for crime prediction.” 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE[4] , 20
* *Oren Anava, Kr Y. Levy* ***“****k\_-Nearest Neighbors: From Global to Local”, arXiv:1701.07266v1 [stat.ML] (****2017****).*
* [*Ming Zhao*](http://ascidatabase.com/author.php?author=Ming&last=Zhao)*, and* [*Jingchao Chen*](http://ascidatabase.com/author.php?author=Jingchao&last=Chen)*, “Improvement and Comparison of Weighted k Nearest Neighbors Classifiers for Model Selection”, Volume 10 (1): 109-118, (****2016****).*