A
Project Report
On
"Crime Detection"

Prepared by

Krupali Dobariya (18DIT014)
Kevin Khunt (18DIT028)
Yash Mistry (D19DIT081)

Under the guidance of

Prof. Dipak Ramoliya

A Report Submitted to

Charotar University of Science and Technology

For Partial Fulfillment of the Requirements for the

5th Semester Summer Internship-I (IT346)

Submitted at

Department of Information Technology

Devang Patel Institute of Advance Technology and Research

At: Changa, Dist: Anand – 388421

June 2020

CHARUSAT
CHAROTAR UNIVERSITY OF SCIENCE AND TECHNOLOGY

CERTIFICATE

This is to certify that the report entitled "**Crime Detection**" is a bonafide work carried out by **Ms. Krupali Dobariya(18DIT014), Mr. Kevin Khunt(18DIT028) ,Mr. Yash Mistry (D19DIT081)** under the guidance and supervision of **Prof. Dipak Ramoliya** for the subject **IT346 Summer Internship-I** of 5th Semester of Bachelor of Technology in **Department of Information Technology, DEPSTAR** at Faculty of Technology & Engineering – CHARUSAT, Gujarat.

To the best of my knowledge and belief, this work embodies the work of candidate himself, has duly been completed, and fulfills the requirement of the ordinance relating to the B.Tech. Degree of the University and is up to the standard in respect of content, presentation and language for being referred to the examiner.

Prof. Dipak Ramoliya
Assistant Professor
Department of Computer Science & Engineering
DEPSTAR, CHARUSAT, Changa, Gujarat.

Dr. Amit Nayak
I/C Head of Department
Department of Information Technology
DEPSTAR, CHARUSAT, Changa, Gujarat

Dr. Amit Ganatra
Principal, DEPSTAR
Dean, FTE
CHARUSAT, Changa, Gujarat.

TABLE OF CONTENTS

ACKNOWLEDGEMENT

The presentation of this report gives me the feeling of fulfillment. With immense pleasure we would like to present this report on this Dissertation report of **"Crime Detection".**

We would like to take this opportunity to bestow my acknowledgement to the entire person who have directly or indirectly availed me in making my project feasible and to turn it up in to successful piece of work.

We take this opportunity to express our gratitude to the following personalities who have helped us throughout our project. Without their support this project would never have been completed. we are extremely thankful to them. we are thankful to our Department HOD **Dr. Amit Nayak**, who has provided me resources from the college and we are also very thankful Internal Guide **Prof. Dipak Ramoliya** who has provided their time and never-ending support. For their support and help, we heartily thank them. This project would be incomplete without them.

ABSTRACT

The objective of this project is to tackle a vital issue in the society - Crimes. Analyzing and examining of crimes happening in the world will give us a Broadview in understanding the crime regions and can be used to take necessary precautions to mitigate the crime rates this will also extend the quality life of all citizen. Identifying Crime patterns will allow us to tackle problems with unique approaches in specific crime category regions and improve more security measures in society. Current studies show the reason of increase in crime rates is more in areas that are economically backward. In few decades' property crime will be a target. Physical hardware like tv sets, mobiles continue to be a target for thefts. The following approach involves predicting crimes classifying, pattern detection and visualization with effective tools and technologies. Use of beneficial data entity like location, timestamps, and number of crimes make it possible to define a well programmed approach. Use of past crime data trends helps us to correlate factors which might help understanding the future scope of crimes.

Chapter 1: Introduction

1.1. Project definition:

As machine learning is currently the most famous technology, it is used in various applications. One such application that it is being used is in Crime Detection and Prevention. The idea behind this are used is that crimes are relatively predictable; it just requires being able to sort through a massive volume of data to find patterns that are useful to law enforcement. This kind of data analysis was technologically impossible a few decades ago, but the hope is that recent developments in machine learning are up to the task.

Our try to use machine learning is to predict when and where crime will take place. A program based on the observation that certain crime types tend to cluster in time and space. By using past data and observing where recent crimes took place, they can predict where future crimes will likely happen.

1.2. Description:

Crime is one of the most predominant and alarming aspects in our society and its prevention is a vital task. Crime detection is a systematic way of detecting and investigating patterns and trends in crime. The aim of this model is to increase the efficiency of crime prediction. This model will predict as per location, number of crimes and timestamp of the crimes committed. These predictions are done based on the data stored in the dataset from the crime scene. The model predicts using algorithms like, K- Neighbors Classifier. It will be trained and tested using dataset with past crime records in it and implemented using python.

Chapter 2 : Requirements

2.1. Software and Hardware Requirements:

For developing the model:

- Operating System: Windows 7 or higher, Linux.
- 1GB HDD/SSD memory space requirement for software installation.
- 2GB RAM for execution of testing model.
- Python IDLE(v3.7)/ Jupyter Notebook - for interpreting, debugging code and testing model.
- Google colab with good Internet connectivity (Optional/Alternative).
- Intel UHD Graphics 620 or GPU(best suitable)
- Intel core i3-3200 or higher.

For executing the model:

- 2MB RAM for execution of classification.
- Intel UHD Graphics 620 or GPU(best suitable)
- Intel core i3-3200 or higher.

Chapter 3 : System Design

3.1. Major Functionality:

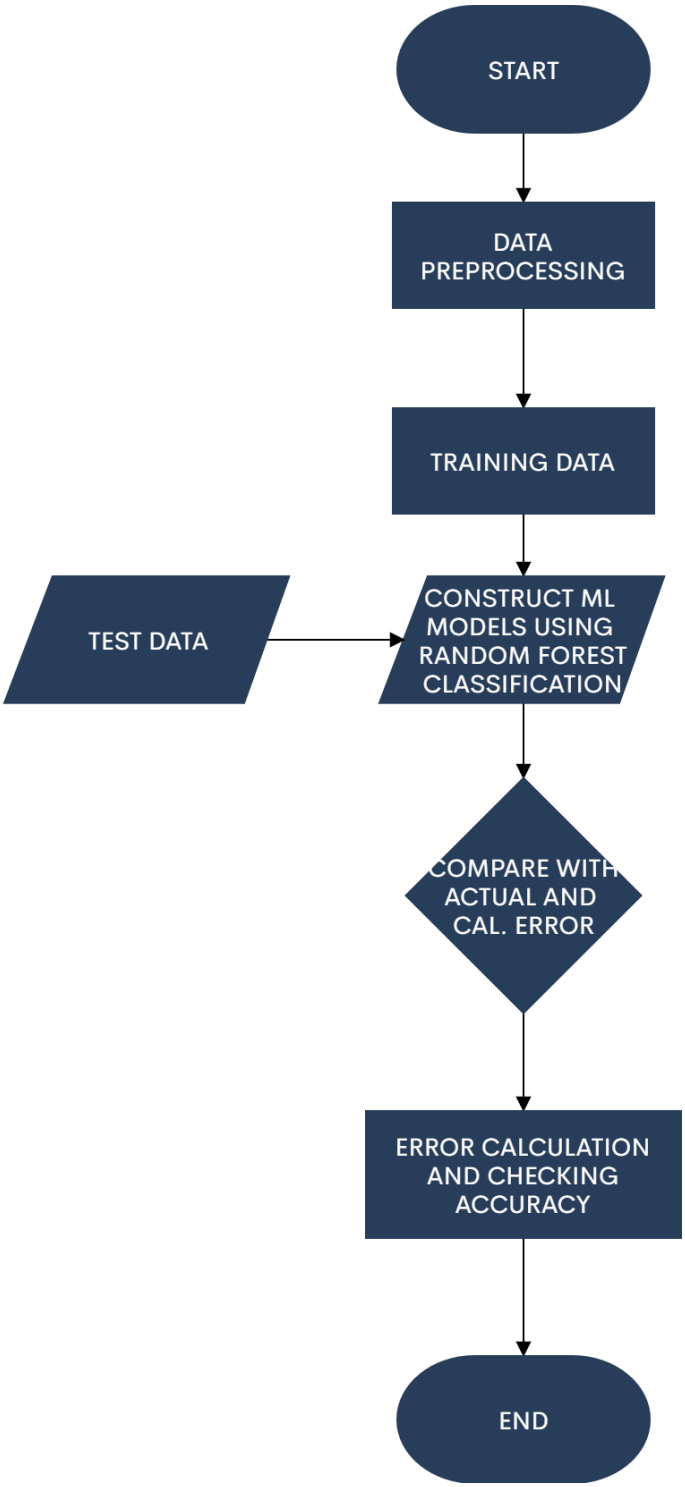
With help of Pandas library we pre-process the data to filter Not Applicable data such as null values which keeps the dataset clean and maintain efficiency while working with huge data records.

We used Matplotlib for data visualisation that is plotting various model output in a vertical bar graph.

Numpy library for easy and efficient data calculation within the dataset which also helps easy plotting of charts.

Sklearn for training model using the Random Forest and predict based on classification obtained by using the Random Forest of sklearn python library.

3.2. Flow Chart:



3.3. Screenshots:

- Using read_csv function of pandas for dataset import

530648	Mischief	2017	5	29	22.0	30.0	14XX E 7TH AVE	Grandview-Woodland	494533.97	5456824.97	49.264163	-123.075129
530649	Offence Against a Person	2017	4	13	NaN	NaN	OFFSET TO PROTECT PRIVACY	NaN	0.00	0.00	0.000000	0.000000
530650	Theft from Vehicle	2017	6	5	17.0	0.0	8XX HAMILTON ST	Central Business District	491487.85	5458385.78	49.278168	-123.117031
530651	Vehicle Collision or Pedestrian Struck (with I...	2017	6	6	17.0	38.0	13XX BLOCK PARK DR	Marpole	490204.00	5451444.00	49.215706	-123.134512

530652 rows × 12 columns

- Using dropna function of pandas to remove rows with any NA (Not Applicable values) in it.

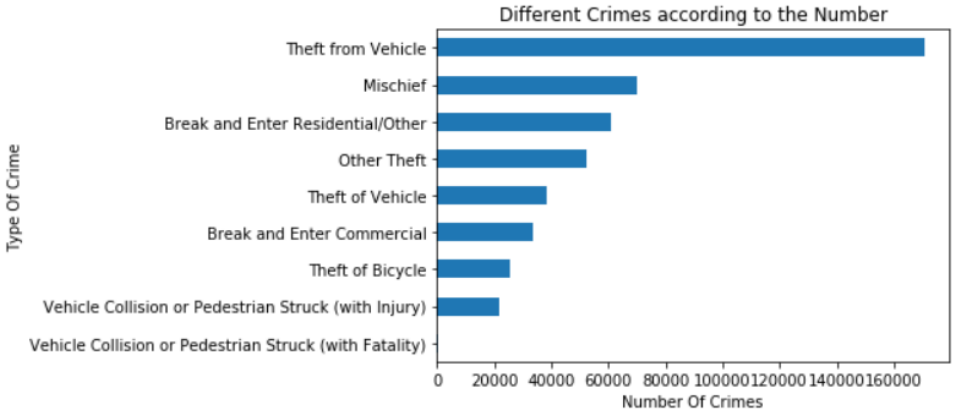
530646	Mischief	2017	1	18	14.0	44.0	14XX E HASTINGS ST	Grandview-Woodland	494563.75	5458727.40	49.281276	-123.074746
530647	Break and Enter Residential/Other	2017	3	3	9.0	16.0	31XX ADANAC ST	Hastings-Sunrise	497265.49	5458296.71	49.277420	-123.037595
530648	Mischief	2017	5	29	22.0	30.0	14XX E 7TH AVE	Grandview-Woodland	494533.97	5456824.97	49.264163	-123.075129
530650	Theft from Vehicle	2017	6	5	17.0	0.0	8XX HAMILTON ST	Central Business District	491487.85	5458385.78	49.278168	-123.117031
530651	Vehicle Collision or Pedestrian Struck (with I...	2017	6	6	17.0	38.0	13XX BLOCK PARK DR	Marpole	490204.00	5451444.00	49.215706	-123.134512

474015 rows × 12 columns

- Counting values of each particular crime for getting a clear cut image view.

```
Out[4]: Theft from Vehicle                                    170889
        Mischief                                               70157
        Break and Enter Residential/Other                      60856
        Other Theft                                            52160
        Theft of Vehicle                                       38351
        Break and Enter Commercial                             33841
        Theft of Bicycle                                       25620
        Vehicle Collision or Pedestrian Struck (with Injury)   21887
        Vehicle Collision or Pedestrian Struck (with Fatality)   254
        Name: TYPE, dtype: int64
```

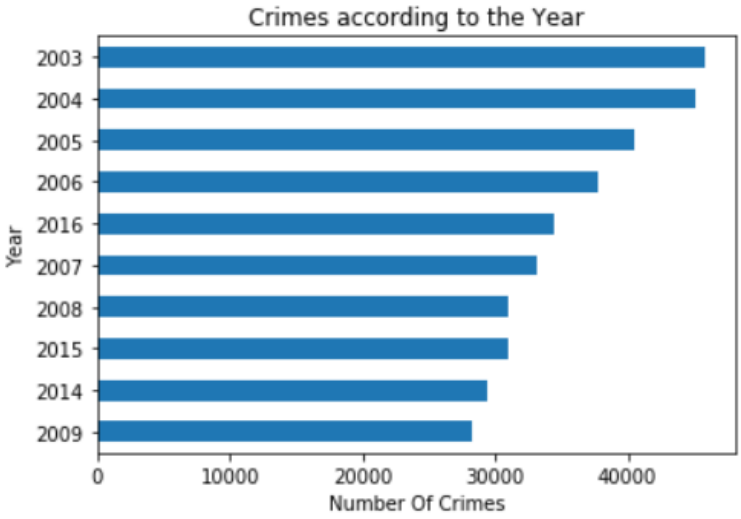
- Plotting Horizontal bar graph for each crime.



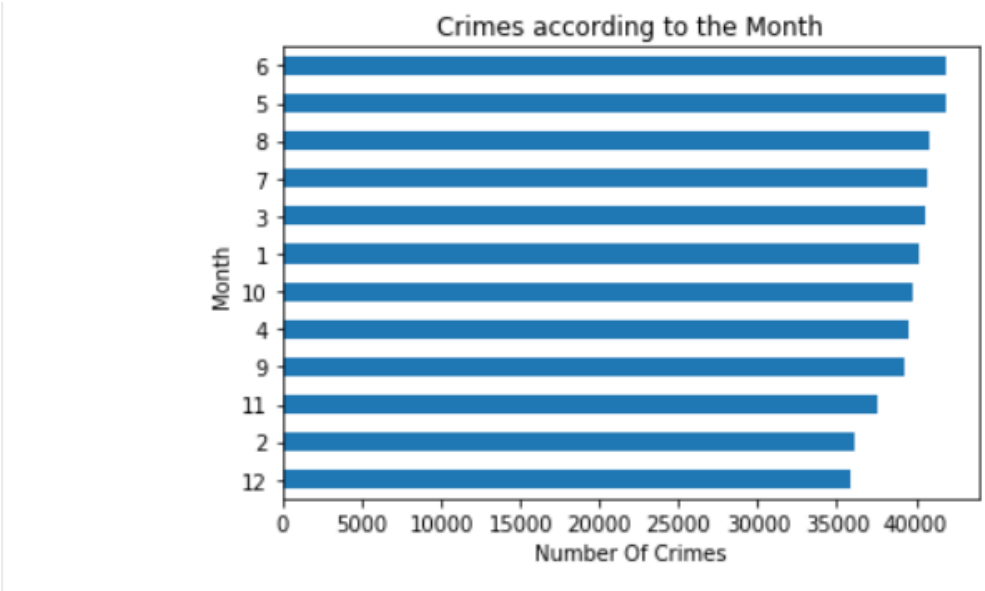
- Concluding the data types for further model testing.

```
Out[6]: TYPE              object
        YEAR               int64
        MONTH              int64
        DAY                int64
        HOUR             float64
        MINUTE           float64
        NEIGHBOURHOOD     object
        X                float64
        Y                float64
        Latitude         float64
        Longitude        float64
        dtype: object
```

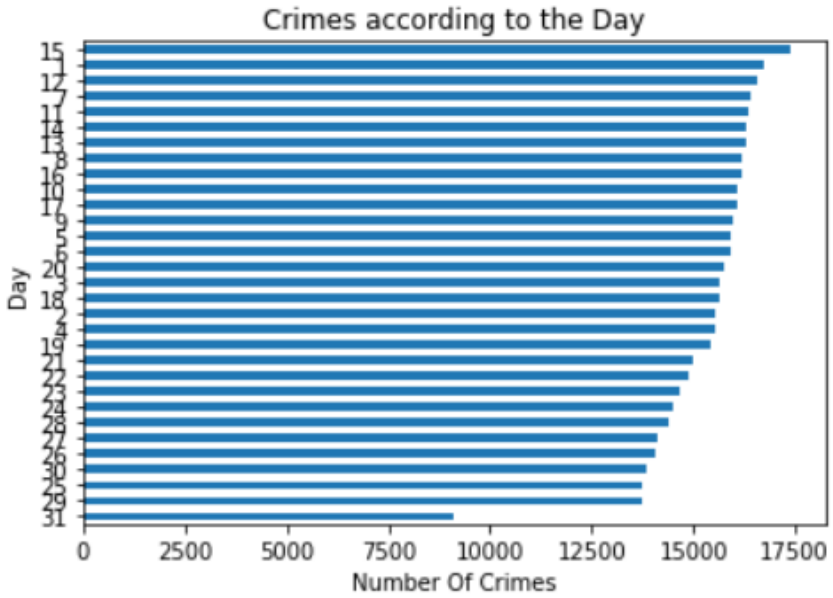
- Plotting crimes according to the years and number



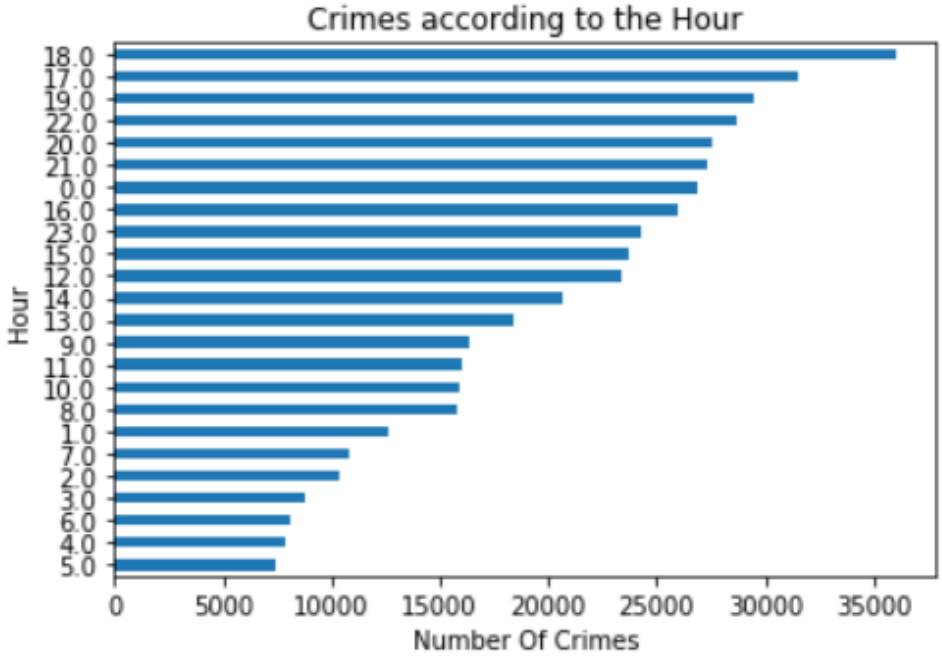
- Plotting crimes according to month.



- Plotting crimes according to days.



- Plotting crimes according to hours.



- Removing points outside bounding box.

Out[12]:

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	NEIGHBOURHOOD	X	Y	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	Strathcona	493906.5	5457452.47	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	Strathcona	493906.5	5457452.47	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	40.0	Strathcona	493906.5	5457452.47	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	15.0	Strathcona	493906.5	5457452.47	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	45.0	Strathcona	493906.5	5457452.47	49.269802	-123.083763

- Using head function of pandas for quickly testing if your object has the right type of data in it.

Out[11]:

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	NEIGHBOURHOOD	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	Strathcona	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	Strathcona	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	40.0	Strathcona	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	15.0	Strathcona	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	45.0	Strathcona	49.269802	-123.083763

- Avoiding dummy trap variable so that predictions will do the good job if categorical values are converted into numerical (binary vectors) values.

Out[16]:

	0	1	2	3	4	5	6	7	8	9	...	20	21	22	23	24	25	26	27	28	29
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	5.0	12.0	16.0	15.0	49.269802	-123.083763
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	5.0	7.0	15.0	20.0	49.269802	-123.083763
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	4.0	23.0	16.0	40.0	49.269802	-123.083763
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	4.0	20.0	11.0	15.0	49.269802	-123.083763
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	4.0	12.0	17.0	45.0	49.269802	-123.083763

5 rows × 30 columns

- Splitting data into train set and test set using train_test_split of sklearn.model_selection library.

Out[17]:

	0	1	2	3	4	5	6	7	8	9	...	20	21	22	23	24	25	26	27	28	29
125965	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2005.0	5.0	7.0	15.0	4.0	49.262326	-123.074456
151408	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2006.0	12.0	12.0	11.0	0.0	49.269447	-123.137994
362936	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2014.0	2.0	21.0	17.0	3.0	49.269348	-123.046810
29417	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2003.0	9.0	9.0	9.0	30.0	49.283419	-123.112772
287543	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	2011.0	1.0	10.0	19.0	0.0	49.257340	-123.160538

5 rows × 30 columns

- Using Random Forest Classifier of sklearn library of python for classification of our model.

```
Out[19]: RandomForestClassifier(bootstrap=True, class_weight=None, criterion='entropy',
                                max_depth=None, max_features='auto', max_leaf_nodes=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, n_estimators=95,
                                n_jobs=None, oob_score=False, random_state=0, verbose=0,
                                warm_start=False)
```

- Using metrics function of sklearn library for getting the accuracy of our model.

```
Accuracy: 0.4897524339947048
```

Chapter 4 : Constraints and Future Enhancement

4.1. Limitations:

- By far our model is able to classify on specific dataset selected by us.
- By far our model is able to produce only accuracy of tested model.
- By selection of classification method it is still quite less accurate.
- No future prediction.
- Limited feature.

4.2. Outcome:

- Learning information gathering phase while selecting specific dataset from different dataset assets.
- Learning different libraries of Python for developing the model.
- Learning various methods of data visualization.
- Learning techniques /methods of data visualization.
- Distribution of work load among teammates.

4.3. Future Enhancement:

- Testing and implementing with different classification techniques to increase accuracy of our model.
- Further pre-processing of data for clean and efficient data usage.

CONCLUSION

With the help of machine learning technology, it has become easy to find out relation and patterns a mong various data's. The work in this project mainly revolves around predicting the type of crime which may happen if we know the location of where it has occurred. Using the concept of machine learning we have built a model using training data set that have undergone data cleaning and data tr ansformation. The model predicts the type of crime with accuracy of 0.4897524339947048. Data visualization helps in analysis of data set. The graphs include bar, pie, line and scatter graphs each having its own characteristics. We generated many graphs and found interesting statistics that helpe d in understanding Chicago crimes datasets that can help in capturing the factors that can help in ke eping society safe.

REFERENCE

- https://www.geeksforgeeks.org/python-introduction-matplotlib/
- https://www.udemy.com/course/machinelearning/
- https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761
- https://towardsdatascience.com/data-preprocessing-concepts-fa946d11c825
- https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/