**Ramaiah Institute of Technology**

(An Autonomous Institute, Affiliated to VTU)

MSR Nagar, MSRIT POST, Bengaluru-560054

A Dissertation Report on

Project Title: Classification and Prediction by using Lazy Learner’s Algorithm

Submitted by

Shrikant R karki 1MS15CS120

Pavankumar G L 1MS15CS085

Vinoda N 1MS16CS423

Praissy Joy 1MS16CS411

Pavankumar K H 1MS11CS073

# *Bachelor of Engineering in Computer Science & Engineering*

Under the guidance of

Sowmya B J & Sanjeetha R

Assistant Professor’s

Computer Science Engineering



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**M.S.RAMAIAH INSTITUTE OF TECHNOLOGY**

**(Autonomous Institute, Affiliated to VTU)**

**Bengaluru-560054**

[www.msrit.edu](http://www.msrit.edu), **2017**

Acknowledgement

We would like to sincerely show our Gratitude to Sowmya B J ,Assistant Professor computer science and engineering, for the stimulating guidance to do the Data mining project on the topic Classification and Prediction by Lazy learner’s algorithm, Also we would like to thank all our friends and facaulty who helped us in finalizing this Project.

INTRODUCTION

Lazy learning is a learning method in which generalization beyond the  Training data is delayed until a query is made to the system, as opposed to the Eager learning where the system tries to generalize the training data before receiving queries.

Lazy : less time in training but more time in predicting.

In case of Eager learners, when given a set of training tuples, will construct a generalization model before receiving new tuples to classify.In eager learners,we can think of the learned model as being ready and eager to classify previously unseen tuples.In lazy approach,the learner instead waits until the last minute before doing any model construction in order to classify a given test tuple.That is, when given a training tuple,the lazy learners algorithm simply stores it and waits until it is given a test tuple.Only when it is given a test tuple,lazy learners algorithm performs the generalization in order to classify the tuple based on its similarity to the stored training tuples.Unlike eager learning algorithms,lazy learners algorithm performs less work when a training tuple is presented and more work when making a classification or prediction. Because lazy learners stores the training tuples or ‘Instances’ they are also referred to as “Instance Based Learners”.

Classification: Classification is a data mining technique that assigns categories to a collection of data in order to aide in more accurate predictions and analysis. Also called sometimes called a Decision Tree, classification is one of several methods intended to make the analysis of very large data sets effective.

When making a classification or prediction, lazy learners can be computationally expensive.They require efficient storage techniques and are well suited to Implementation on parallel hardware .Lazy learners support incremental learning.

Data Set Description

Source of Dataset: We have selected our dataset from ‘[www.kaggle.com](http://www.kaggle.com)’.

Our data set name is “10\_Property\_stolen\_and\_recovered”.The data set has a big information about the crimes in all the states in India in every year since 2001.It also provides the information about the number of crimes in which property is stolen,the value of the property stolen in all the years and the number of cases in which property is recovered,the value of the recovered property.Our Data set is as follows.

Attributes Description: In our dataset we have totally eight attributes. They are :

1)Area\_Name : :In our data set,We can call ‘Area\_Name’ a key attribute. It Specifies the place where the crime has occurred. In our dataset it specifies the name of the state where the crime occurred.

2)Year : It Specifies year of crime. In our case we have taken the crimes happened between 2001 & 2010

3)Group\_Name : It Specifies name of the groups which did the crime.

4)Sub\_Group\_Name : It Specifies the Alias names of the crime groups.

5)Cases\_Property\_Recovered : It Specifies the number of Properties recovered in a Particular case in a state.

6)Cases\_Property\_Stolen : It Specifies the number of Properties stolen in a Particular case in a state.

7)Value\_of\_Property\_Recovered : It Specifies the Cost of the Property recovered in a crime,in a particular state.

8)Value\_of\_Property\_Stolen : It Specifies the cost of the Property stolen in a crime in a particular state.

Data Set size in terms of Bytes : 200kb

Number of Tuples : 2500

Inference

By considering the taken dataset,following inference has been drawn:

We have considered four attributes in the dataset. Those are

1.area\_name,

2.year,

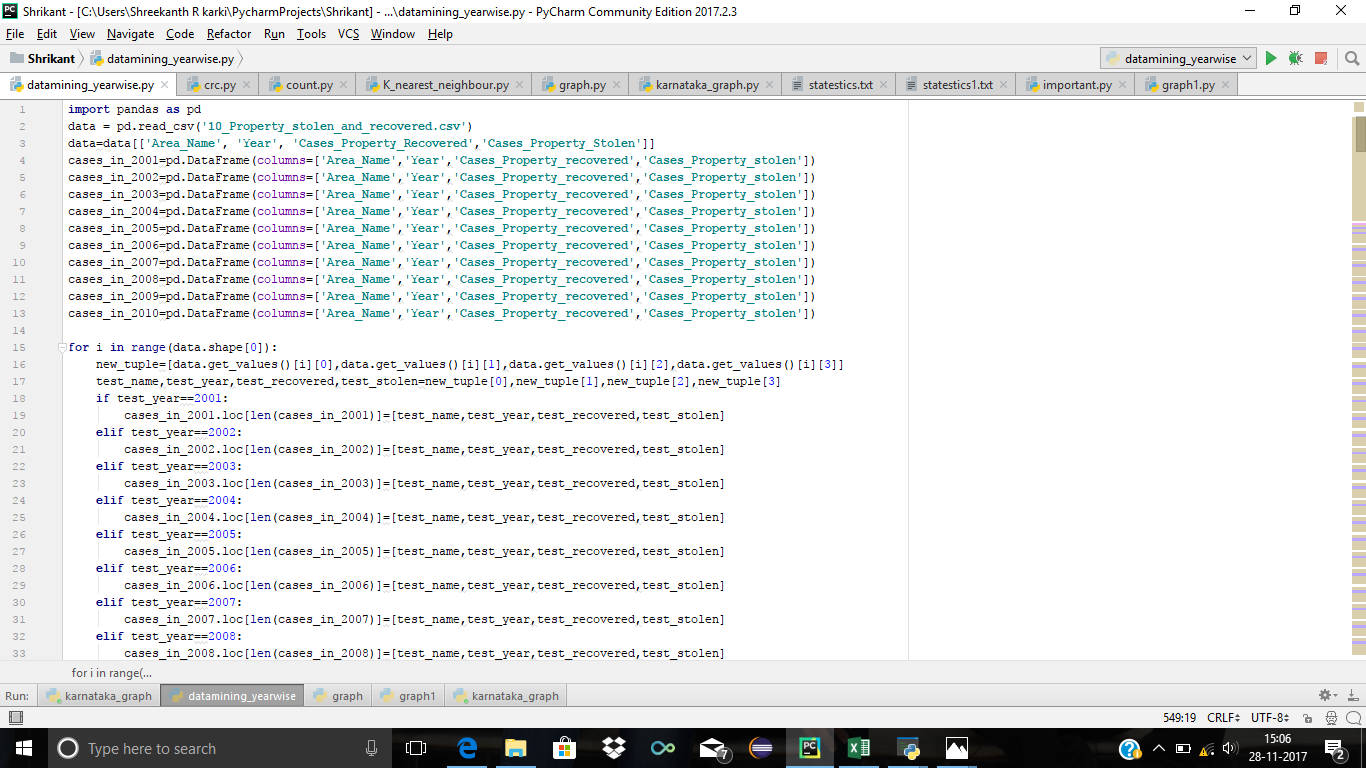
2. cases stolen and

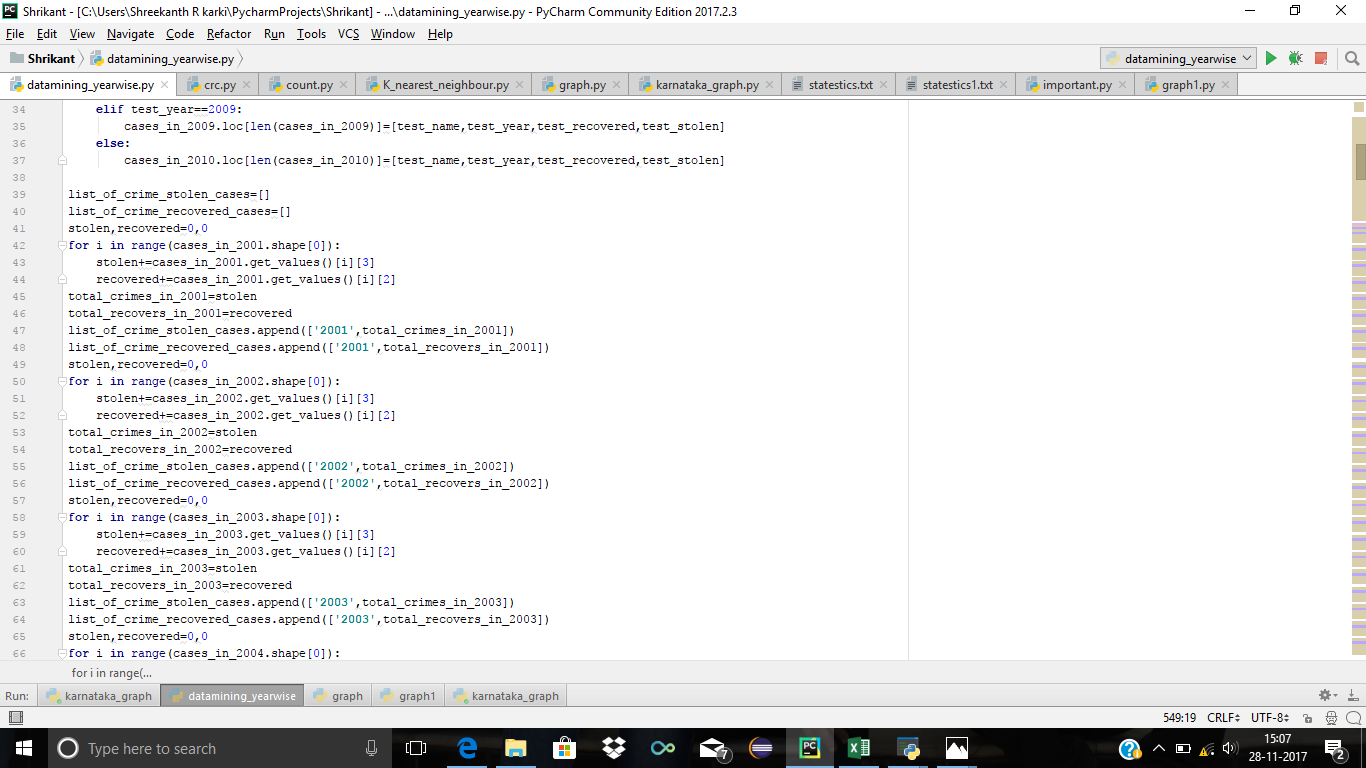
3.cases property recovered.

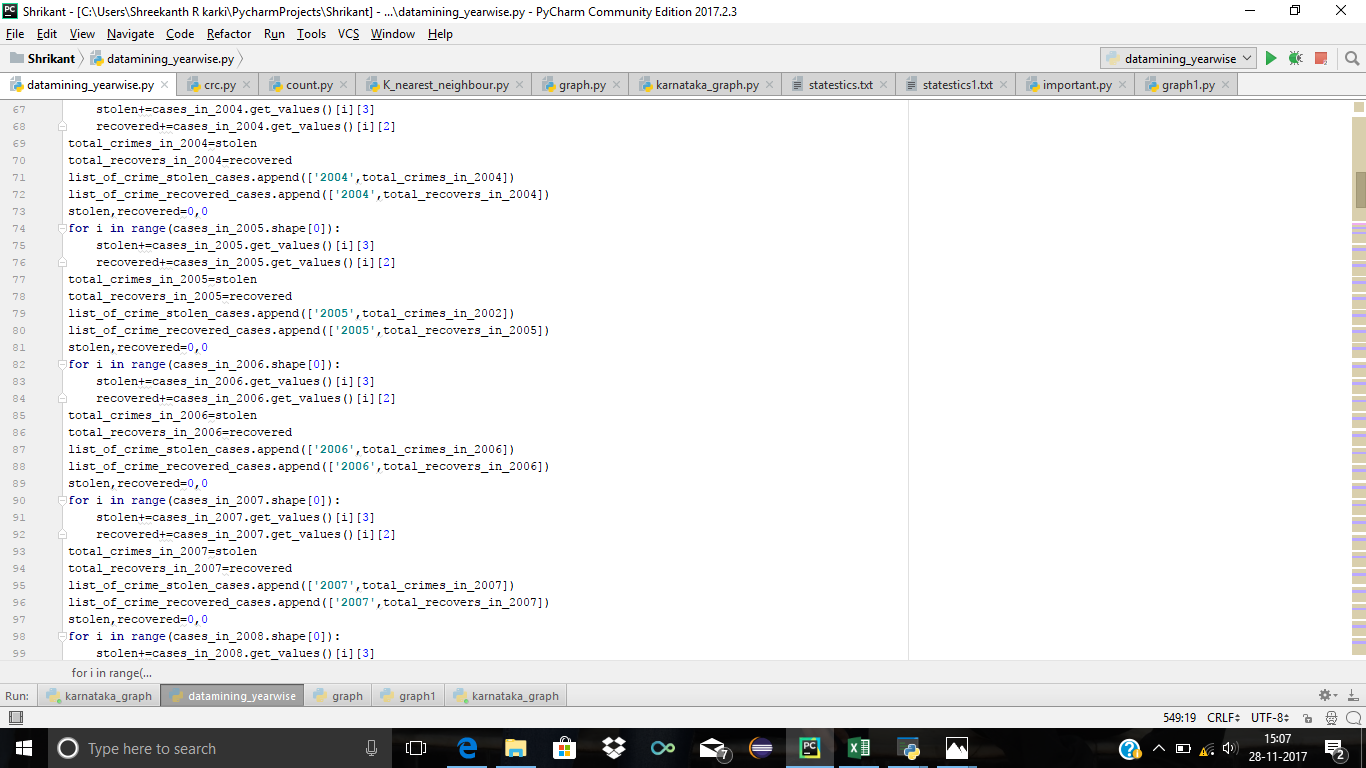
The attribute ‘area name’helps us to find where the crime is happened. The attribute ‘Year’ helps us to find the year in which the crime is happened. And we can find the total number of crimes by summing up the values of ‘cases property stolen’. And we can also retrieve the information about the number of ‘cases solved or property recovered’.

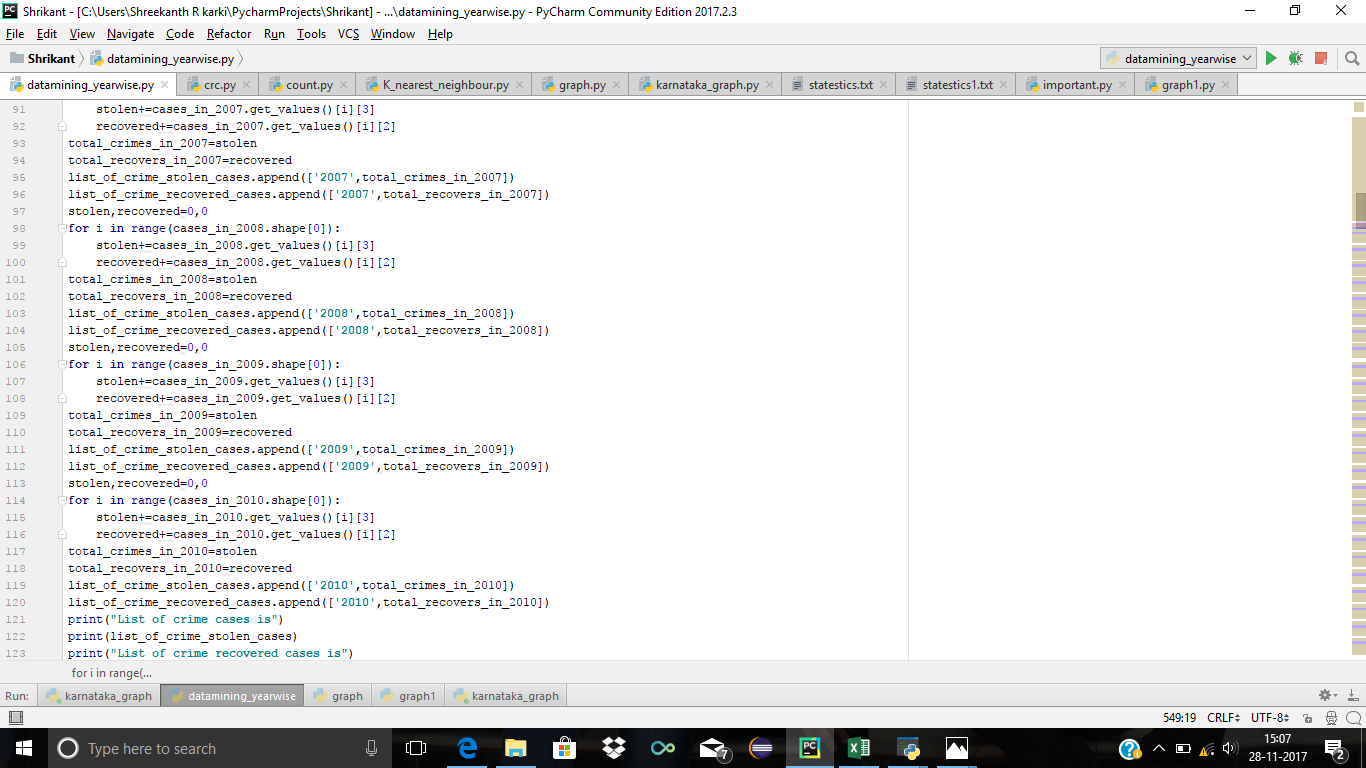
Algorithm Description

Snapshot of the code:



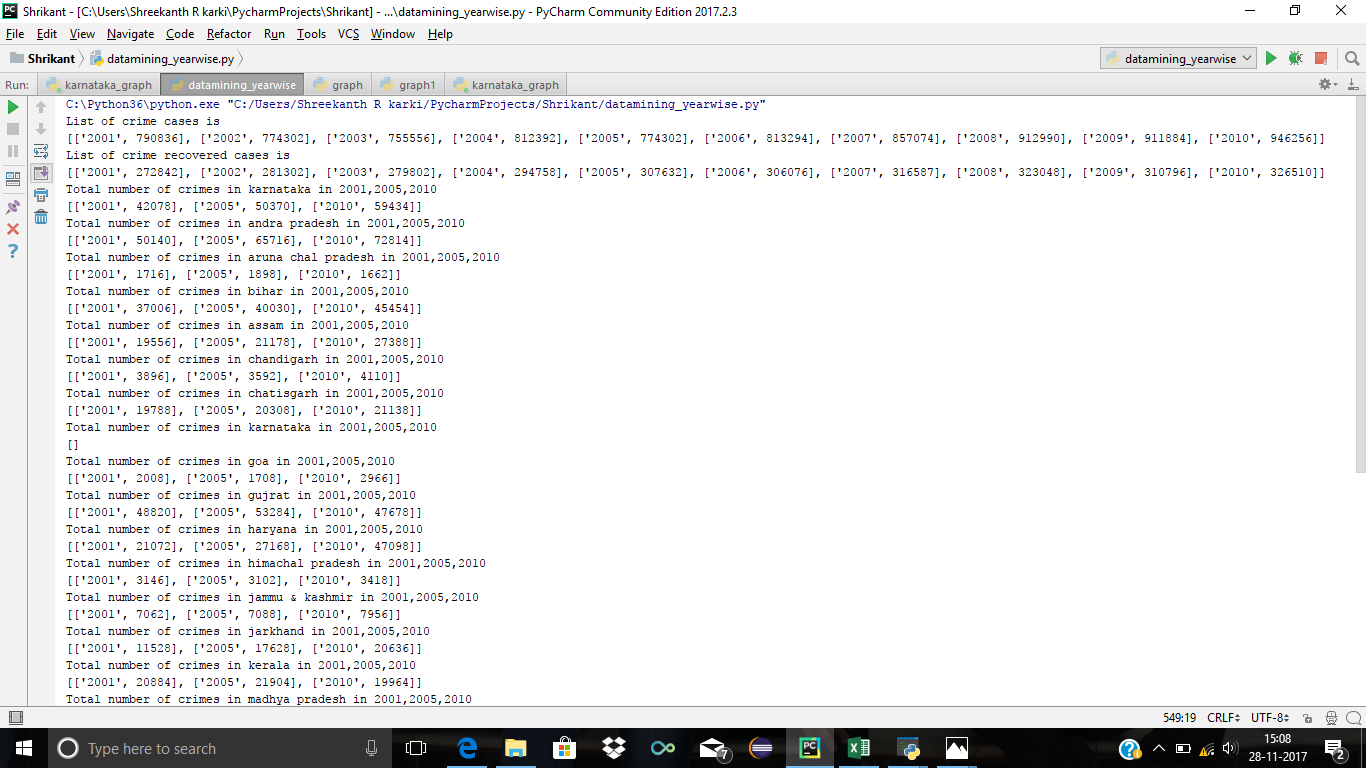






Result

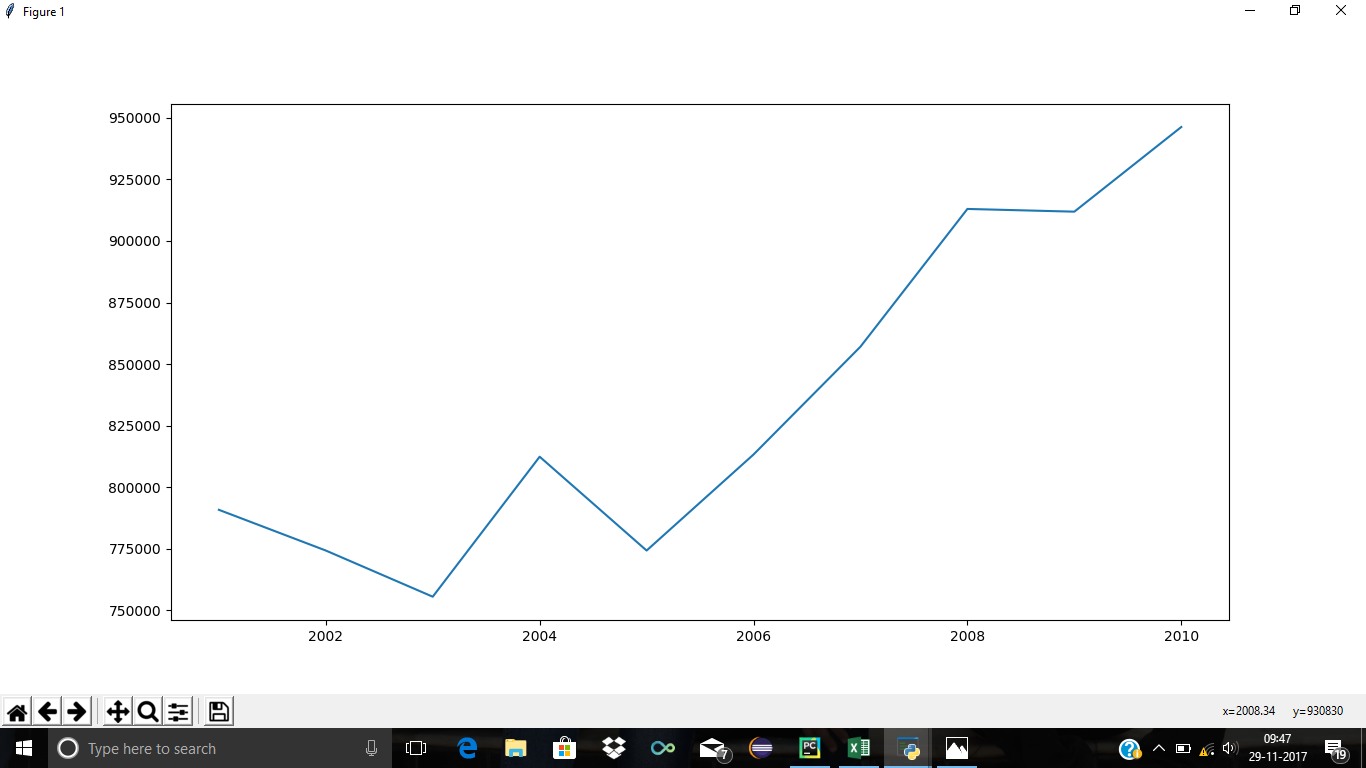
Snapshot and its description:



Description of the Result:

Our algorithm classifies the dataset based on the year in which the crimes have been done, We have classified the dataset with respect to the crimes in 2001 to 2010.Analso we have made classes for each state and the number of crimes happened in those particular states.

The graph between Year and number of crimes in all the states

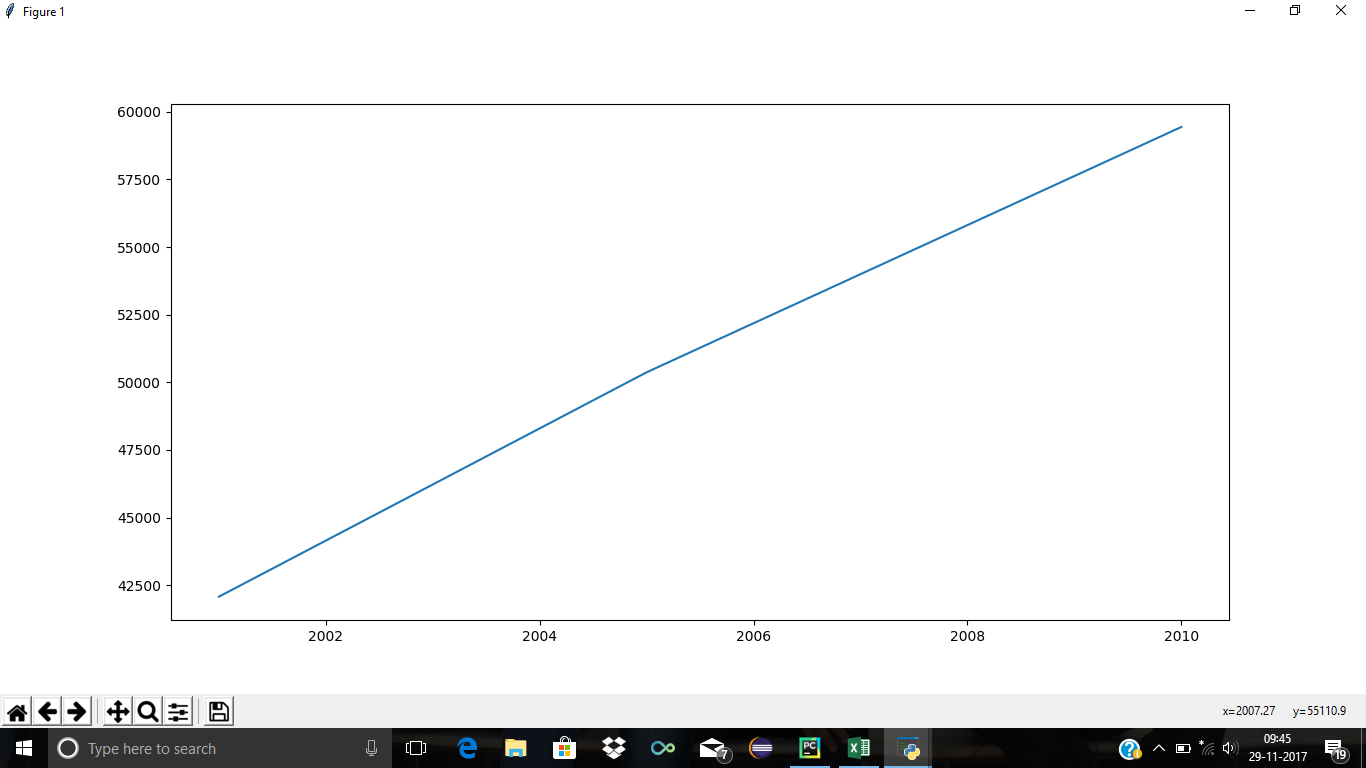


x-axis : Year of crime

y-axis: Number of crimes

Inference:

The graph shows the drastic increase in the number of crimes as the year goes up.



x-axis: year of crime

y-axis: Number of crimes in Karnataka

Inference:

The above graph indicates that the number of crimes has been Linearly increased over the period of years 2001-2010 in Karnataka

Advantages:

* The target function will be approximated locally, such as in the [k-nearest neighbour algorithm](https://en.wikipedia.org/wiki/K-nearest_neighbor_algorithm).
* Lazy learning systems can simultaneously solve multiple problems and deal successfully with changes in the problem domain.
* It is used for large datasets with few attributes.

Disadvantages :

* Large space requirement to store the entire training dataset. Particularly noisy training data increases the case base unnecessarily, because no abstraction is made during the training phase.
* Lazy learning methods are usually slower to evaluate, though this is coupled with a faster training phase
* .Lazy classifiers are most useful only for large datasets with few attributes.

SOCIAL IMPACT

Number of crimes has been increased drastically over the period of years,so the society need to be more secured.