**Project Report**

**Predictive Analytics**

**Crime Rate Prediction**

**Made by -**

|  |  |  |
| --- | --- | --- |
| SAPID | Roll No | Name |
| 500109573 | R2142221174 | Suraj Tuteja |
| 500105913 | R2142220378 | Aaditya Vikram Joshi |
| 500106133 | R2142220918 | Harshvardhan Singh |

**Submitted to – Dr Achala Shakya**



SCHOOL OF COMPUTER SCIENCE

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES, BIDHOLI, DEHRADUN, UTTRAKHAND, INDIA

TABLE OF CONTENT

|  |  |  |
| --- | --- | --- |
| S. No. | Content | Page |
| 1 | Introduction |  |
| 1.1 | Problem Statement |  |
| 1.2 | Motivation |  |
| 1.3 | Objectives |  |
| 2 | LR |  |
| 3 | Methodology |  |
| 4 | Design Diagrams |  |
| 5 | Working |  |
| 6 | Result and Output |  |
| 7 | References |  |

1. **Introduction**

The criminal cases in India are increasing rapidly due to which number of cases pending are also piling up. This continuous increase in the criminal cases is proving to be difficult to be classified and to be solved. Recognizing the criminal activity patterns of a place is important in order to prevent it from happening. The crime solving agencies can do a better work if they have a good idea of the pattern of criminal activities that are happening in a particular area. This can be done by using machine learning by employing different algorithms to find the patterns of the criminal activities in a particular area. Crime rate is increasing now-a-days in many countries.In today's world with such higher crime rate and brutal crime happening, there must be some protection against this crime. Here we introduced a system by which crime rate can be reduced. Crime data must feed into the system. We introduced data mining algorithm to predict crime. K-means algorithm plays an important role in analyzing and predicting crimes. K-means algorithm will cluster co-offenders, collaboration and dissolution of organized crime groups, identifying various relevant crime patterns, hidden links, link prediction and statistical analysis of crime data. This system will prevent crime occurring in society. Crime data is analyzed which is stored in the database. Data mining algorithm will extract information and patterns from database. System will group crime. Clustering will be done based on places where crime occurred, gang who involved in crime and the timing crime took place. This will help to predict crime which will occur in future. Admin will enter crime details into the system which is required for prediction. Admin can view criminal historical data. Crime incident prediction depends mainly on the historical crime record and various spatial and demographic information.

* 1. **Problem Statement**

Criminals are nuisance for the society in all corners of world for a long time now and measures are required to eradicate crimes from our world. Our mission is to offer crime prevention application to keep public safe.

Current policing strategies work towards finding the criminals, basically after the crime has occurred. But, with the help of technological advancement, we can use historic crime data to recognize crime patterns and use these patterns to predict crimes beforehand. We are using clustering algorithms to predict crime prone areas.

* 1. **Motivation**

The foundation of this project is built on the notion that crimes can often be predicted when patterns in criminal behavior are identified. Although processing large amounts of data to uncover these patterns was unthinkable a few decades ago, advancements in machine learning have made it possible today.

Take Gujarat, for example. Despite an overall decline in the crime rate to 11.9 per 100,000 people, there are serious challenges that cannot be ignored. Gujarat has consistently topped the nation in custodial deaths for three years straight, with 24 deaths recorded in 2022. This disturbing trend raises concerns about the practices of law enforcement. In major cities like Ahmedabad and Surat, crime rates remain among the highest in the country, indicating that urbanization continues to fuel criminal activity.

Equally troubling is the rise in crimes against women and children, particularly sexual violence. In 2022 alone, Gujarat reported over 2,209 cases of rape, a significant increase from the previous year. Drug-related crimes are also escalating, with large drug seizures pointing to increased trafficking. These factors have contributed to growing insecurity among the population.

On the bright side, there have been positive developments. Law enforcement reforms, like the implementation of the VISHWAS CCTV surveillance project, have helped prevent some crimes and improved public safety in key areas. However, more work is needed to address the rising number of thefts, sexual violence, and other serious offenses.

This project seeks to apply machine learning techniques to analyze crime data and predict crime-prone areas, giving law enforcement the tools they need to act before crimes occur and, ultimately, make cities safer for everyone.

* 1. **Objectives**
* The objective of our work is to:

 Predicting crime before it takes place.

 Understanding crime pattern.

 Analysis of crime in Indore

* Some of the goals :-

Much of the current work is focused in:-

Understanding patterns of criminal behavior that could help in solving criminal investigations.

1. **Literature Review**

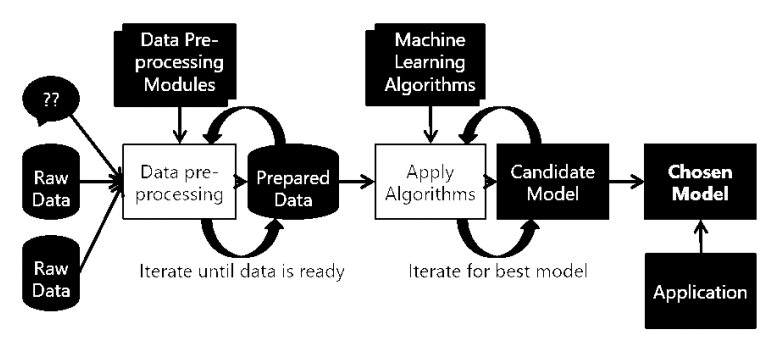
There is a great deal of literature on crime prediction, both scientific and lay. Much of the scientific literature on crime prediction has focused on the development of statistical models that can be used to predict crime rates in specific geographic areas. These models typically take into account a variety of factors, including population density, demographics, economic conditions, and the like. While such models can be quite accurate, they are often criticized for being too general and for failing to take into account the specific circumstances of individual crimes.

The lay literature on crime prediction is often much more speculative, and often focuses on the idea of using psychological profiling to predict who is likely to commit a crime. This approach has been popularized in books and movies, but has little scientific basis.

Shiju Sathyadevan, Devan M.S, proposed that Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder, sex abuse, gang rape etc have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it .It is only within the last few decades that the technology made spatial data mining a practical solution for wide audiences of Law enforcement officials which is affordable and available. Since the availability of criminal data or records is limited we are collecting crime data from various sources like web sites, news sites, blogs, social media, RSS feeds etc. This huge data is used as a record for creating a crime record database. So the main challenge in front of us is developing a better, efficient crime pattern detection tool to identify crime patterns effectively

1. **Methodology**

The term machine learning refers to the automated detection of meaningful patterns in data. In the past couple of decades it has become a common tool in almost any task that requires information extraction from large data sets.We are surrounded by a machine learning based technology: search engines learn how to bring us the best results (while placing pro\_table ads), anti-spam software learns to filter our email messages, and credit card transactions are secured by a software that learns how to detect frauds. Digital cameras learn to detect faces and intelligent personal assistance applications on smart-phones learn to recognize voice commands. Cars are equipped with accident prevention systems that are built using machine learning algorithms.



The inputs to our algorithms are time (hour, day, month, year), place (latitude and longitude), class of crime

* Act 379-Robbery,
* Act 13-Gambling,
* Act 279-Accident,
* Act 323-Violence,
* Act 302-Murder,
* Act 363-Kidnapping

Dataset which we are using is scraped daily from website of Gujrat police which is publically

available. But the dataset is Hindi and in order to perform machine learning this data cannot be used as it is.

Hence the data needs to be processed

Features of this dataset

 थाना : Police Station

 थाना अपराध/मगगक्रमाांक : Police Station identification number

 धारा : I.P.C. act number

 फररयादी का नाम एवां पता : Complainant name & address

 आरोपी का नाम एवां पता : Accused name & address

 घटना स्थल : Incident place

 घटना ददनाांक व समय : Incident date & time

 कायमी ददनाांक व समय : Reporting date & time

 दवलांब सेकायमी का कारण : Reason of Time delay in reporting to police

 घटना के कारण सदित दववरण : Incident information in brief

Before implementing machine learning algorithms on our data, we went through a series of preprocessing steps with our classification task in mind. These included:

Dropping features such police station, station number, Complainant name & address,Accused name & address

Dropping features such as Resolution, Description and Address: The address was dropped because we had information about the latitude and longitude, and, in that context, the address did not add much marginal value.

The output is the class of crime that is likely to have occurred.We try out multiple classification algorithms, such as KNN (K-Nearest Neighbors), Decision Trees, and Random Forests.

The timestamp contained the year, date and time of occurrence of each crime. This was decomposed into five features: Year (2018), Month (1-12), Date (1-31), Hour (0- 23) and Minute (0-59).

.

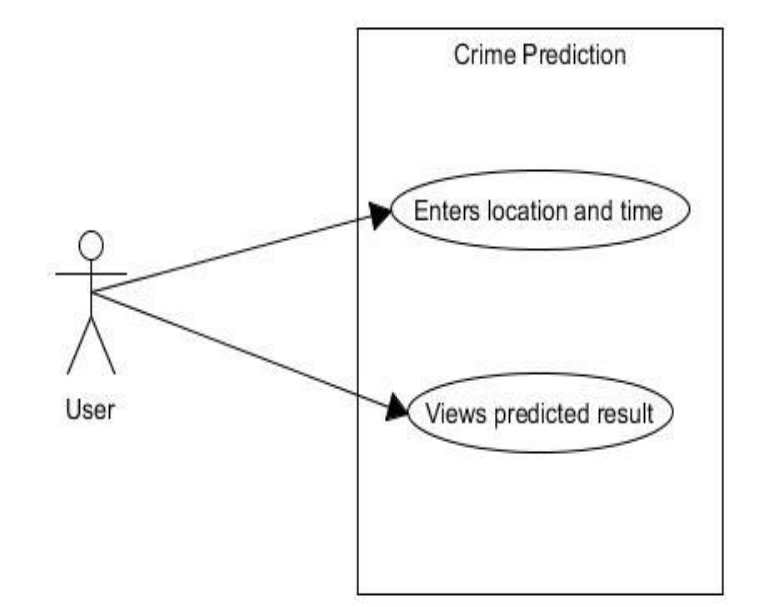
1. **Design Diagrams**

Use case diagram represent the overall scenario of the system. A scenario is nothing but a

sequence of steps describing an interaction between a user and a system.

Thus use case is a set of scenario tied together by some goal. The use case diagram are drawn

for exposing the functionalities of the system.



1. **Working.**

* Th working done under the steps firstly we have imported the packages:

Flask(0.12.2); Pandas(0.22.1); Numpy(1.14.2); Sklearn(0.19.1); Geopy (1.13.0)

* Then after that we have inserted our data set which was in csv file format and read the data .
* Then we have use the timestamp function where we can use our data more effectively and have more accurate results for that
* Then we have done the data visualization and analysis than testing by three algorithms:

**KNN (K-Nearest Neighbors),**

A powerful classification algorithm used in pattern recognition K nearest neighbors stores all

available cases and classifies new cases based on a similarity measure (e.g. distance

function).One of the top data mining algorithms used today. A non-parametric lazy learning

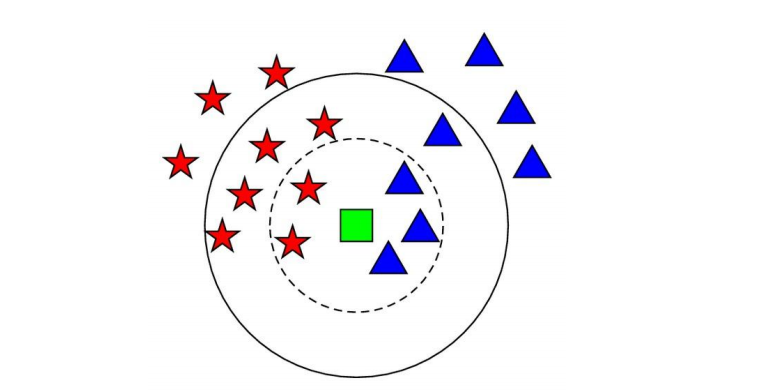
algorithm (An Instance based Learning method).

KNN: Classification Approach

 An object (a new instance) is classified by a majority votes for its neighbor classes.

 The object is assigned to the most common class among-st its K nearest

neighbors.(measured by distance function)



**Decision Trees,**

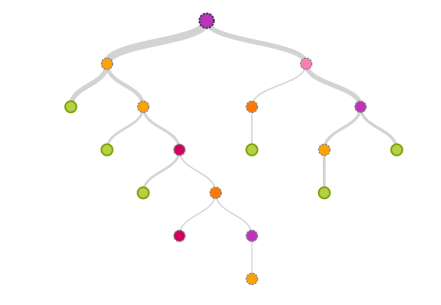
As the name says all about it, it is a tree which helps us by assisting us in decision-making.

Used for both classification and regression, it is a very basic and important predictive learning algorithm.

 It is different from others because it works intuitively i.e., taking decisions one-by-one.

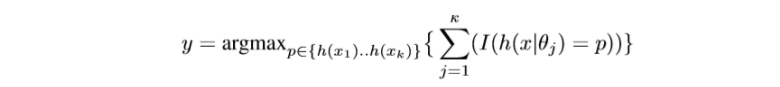
 Non-parametric: Fast and efficient.

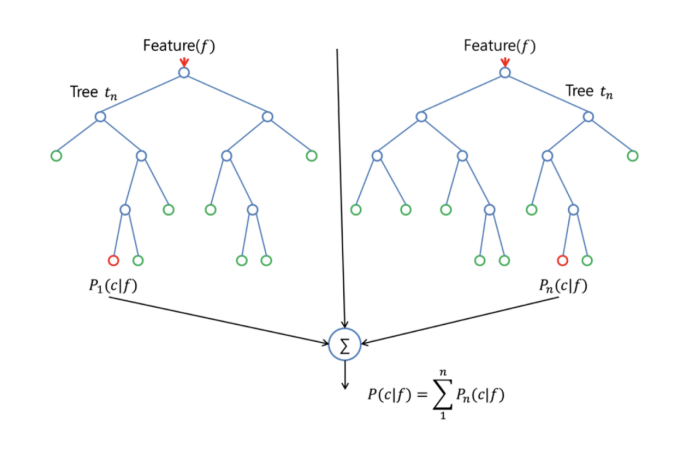
It consists of nodes which have parent-child relationships:



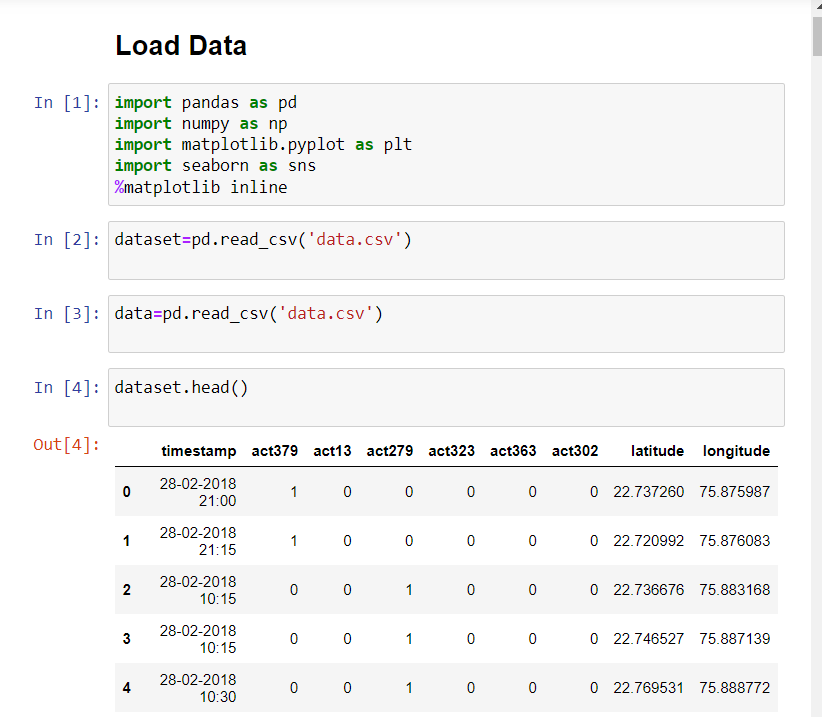
**Random Forests.**

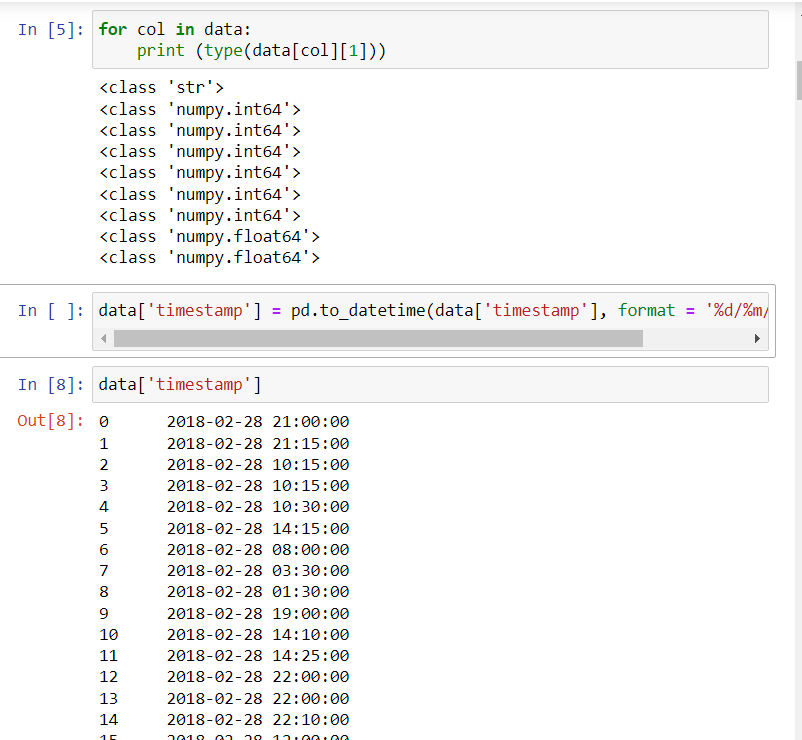
Random Forests is a very popular ensemble learning method which builds a number of classifiers on the training data and combines all their outputs to make the best predictions on the test data. Thus, the Random Forests algorithm is a variance minimizing algorithm that uses randomness when making split decision to help avoid over fitting on the training data. A random forests classifier is an ensemble classifier, which aggregates a family of classifiers h(x|θ1),h(x|θ2),..h(x|θk). Each member of the family, h(x|θ), is a classification tree and k is the number of trees chosen from a model random vector. Also, each θk is a randomly chosen parameter vector. If D(x,y) denotes the training datasets, each classification tree in the ensemble is built using a different subset Dθk(x,y) ⊂ D(x,y) of the training datasets. Thus, h(x|θk) is the kth classification tree which uses a subset of features xθk ⊂ x to build a classification model. Each tree then works like regular decision trees: it partitions the data based on the value of a particular feature (which is selected randomly from the subset), until the data is fully partitioned, or the maximum allowed depth is reached. The fifinal output y is obtained by aggregating the results thus:

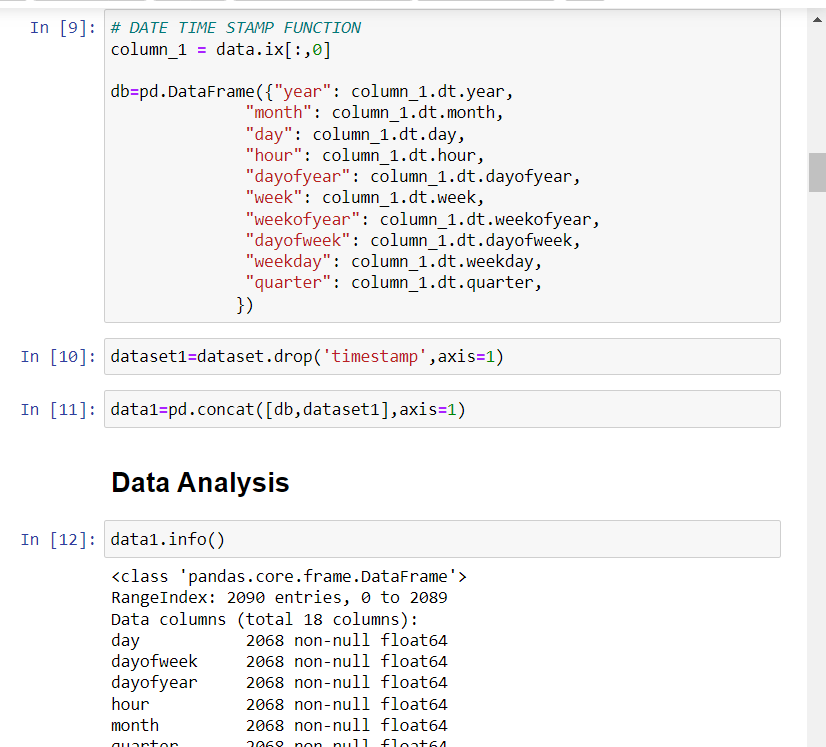


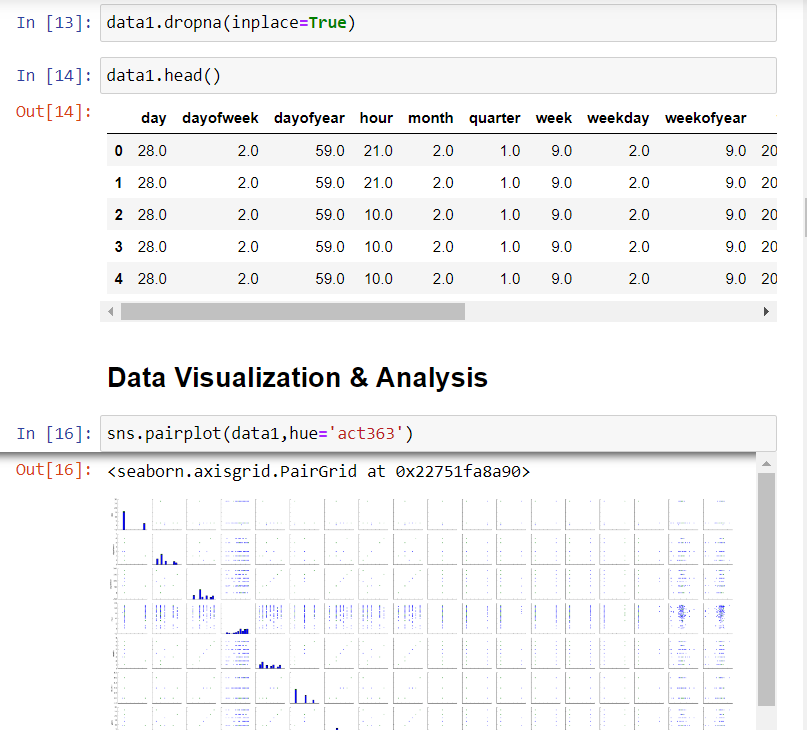


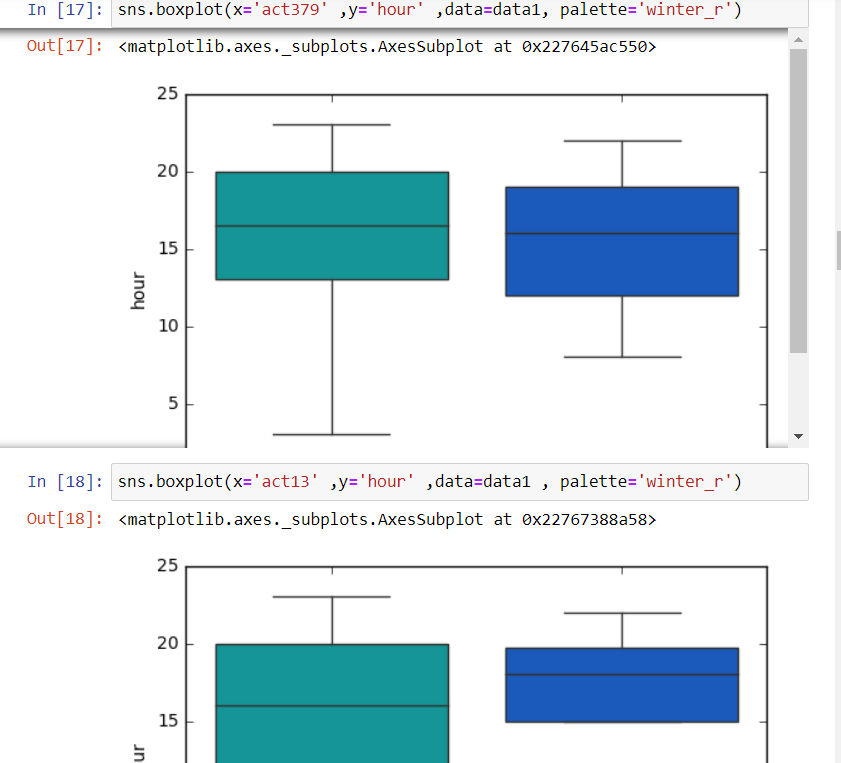
**6. Result and Output**

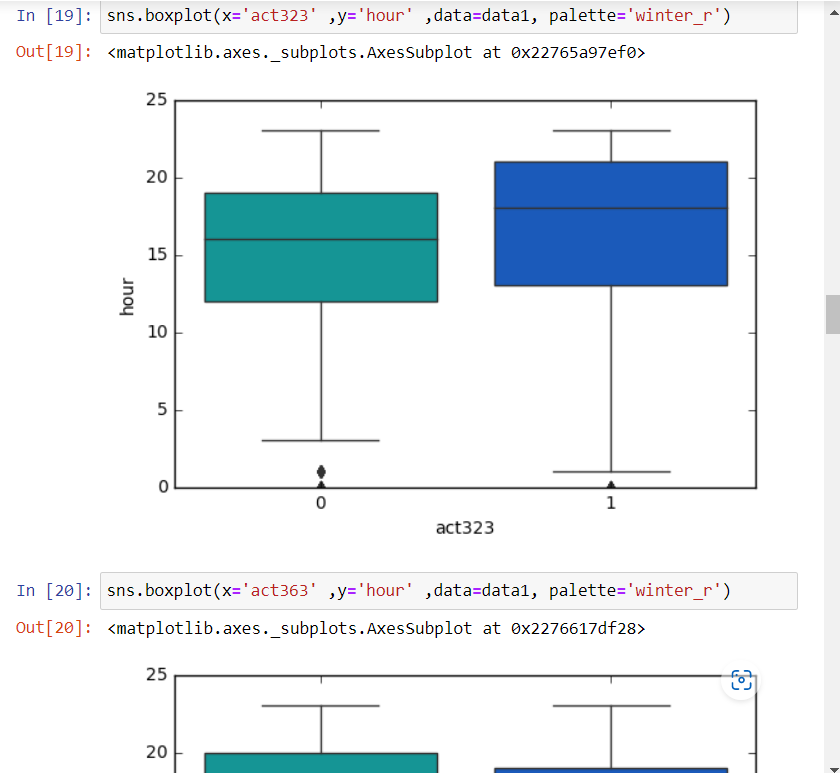


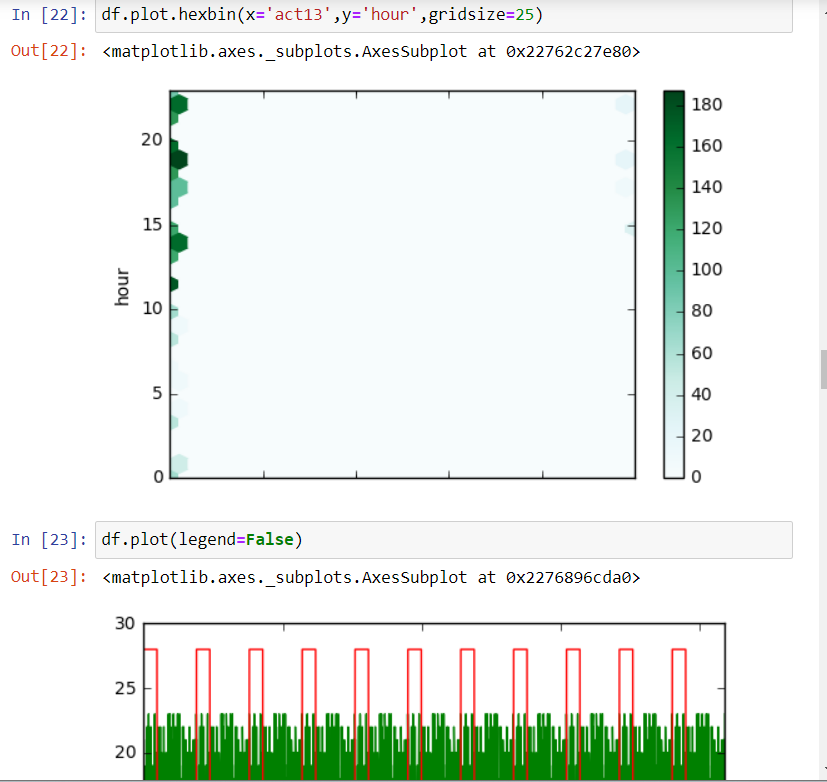


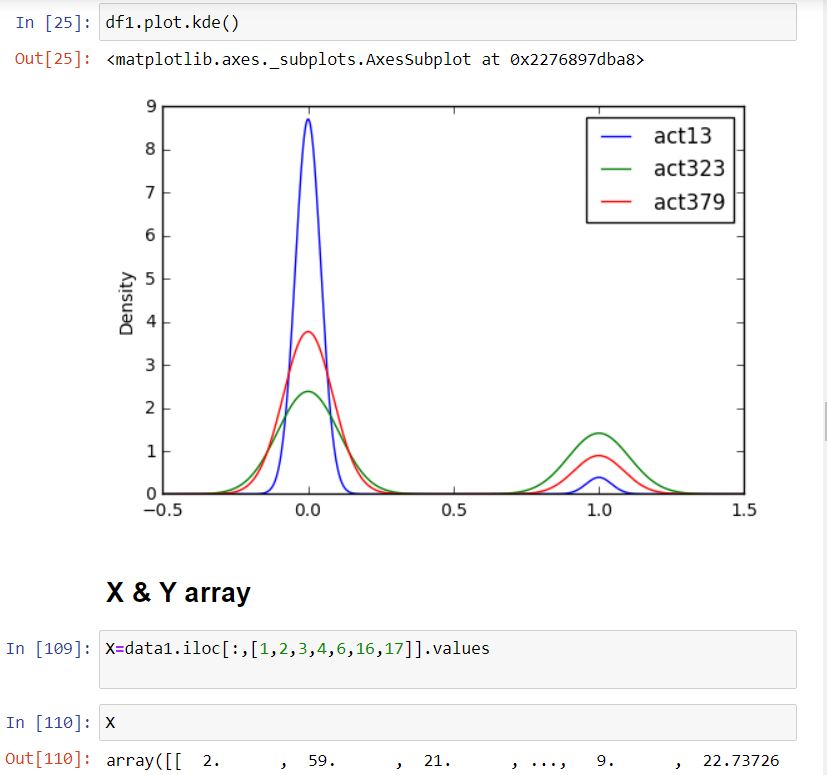


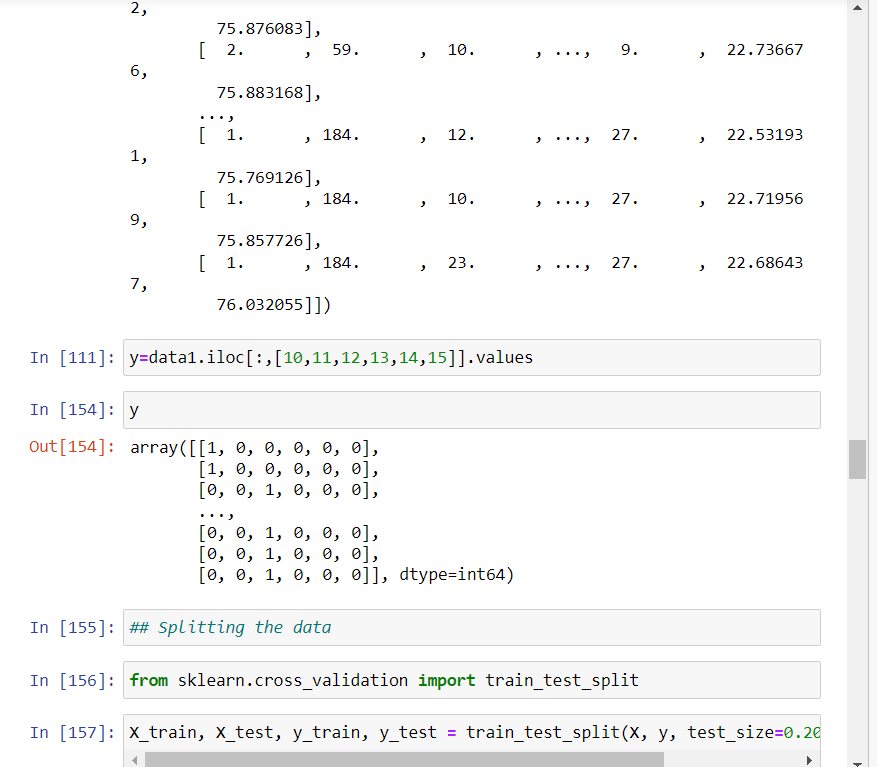


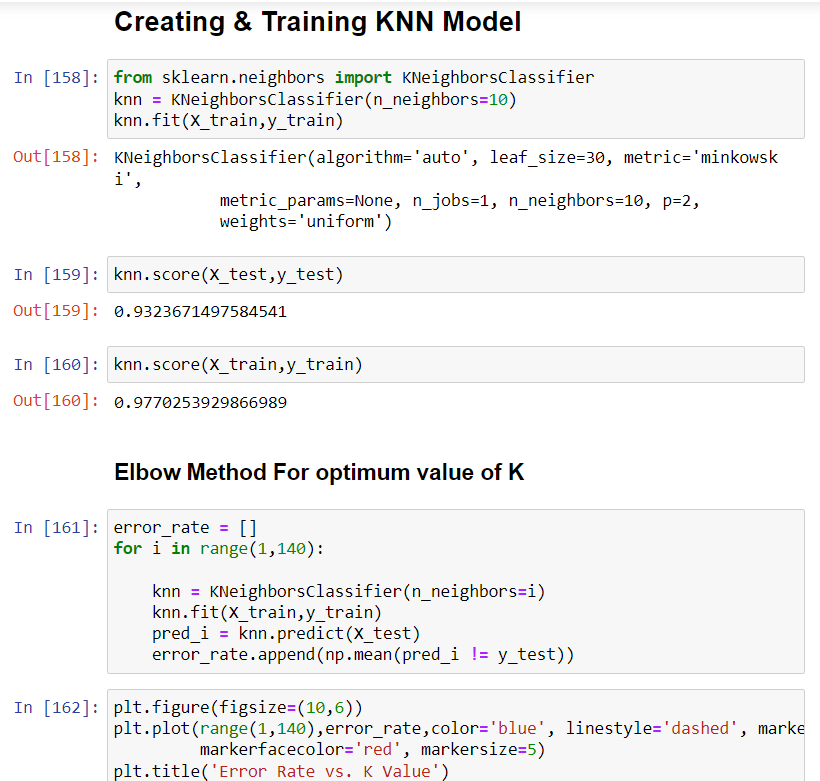


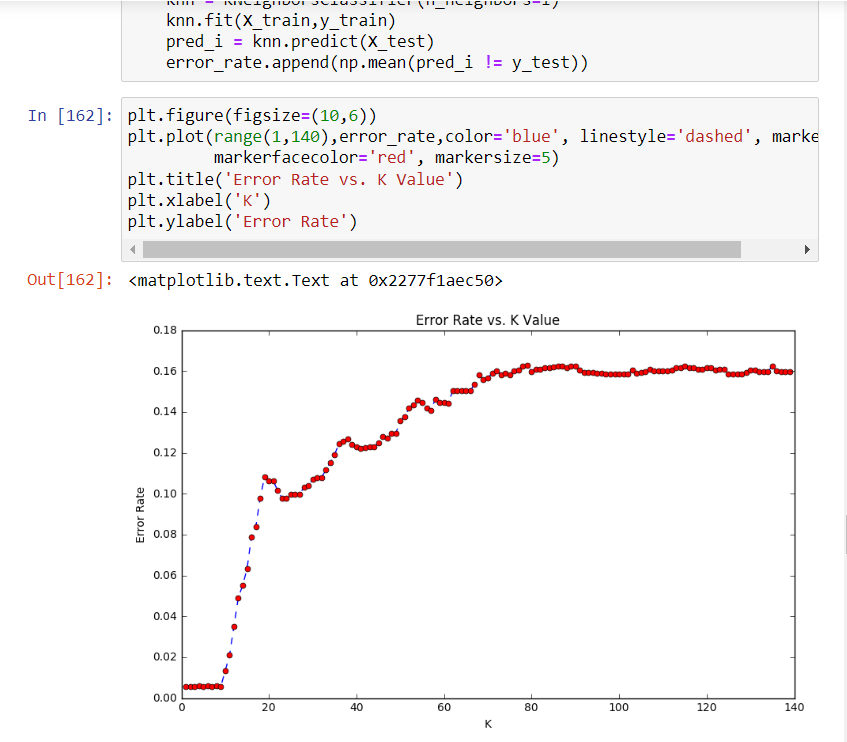


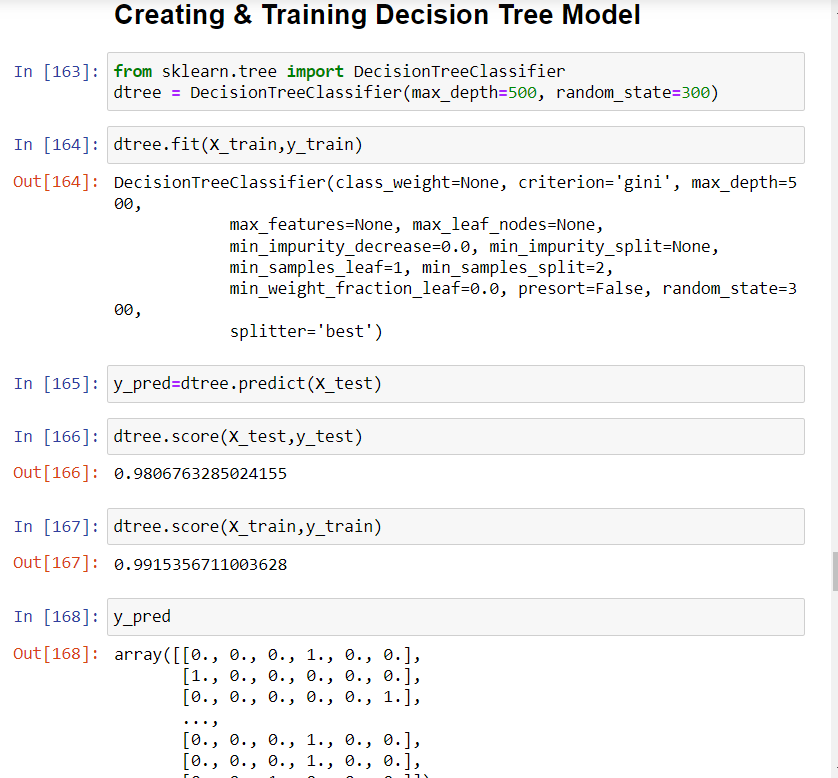


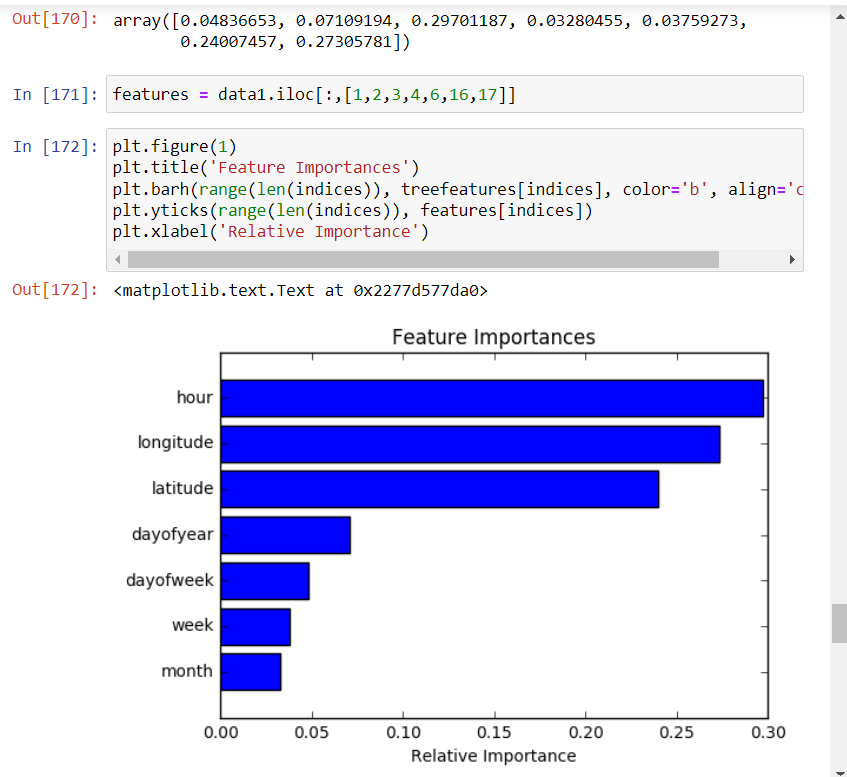


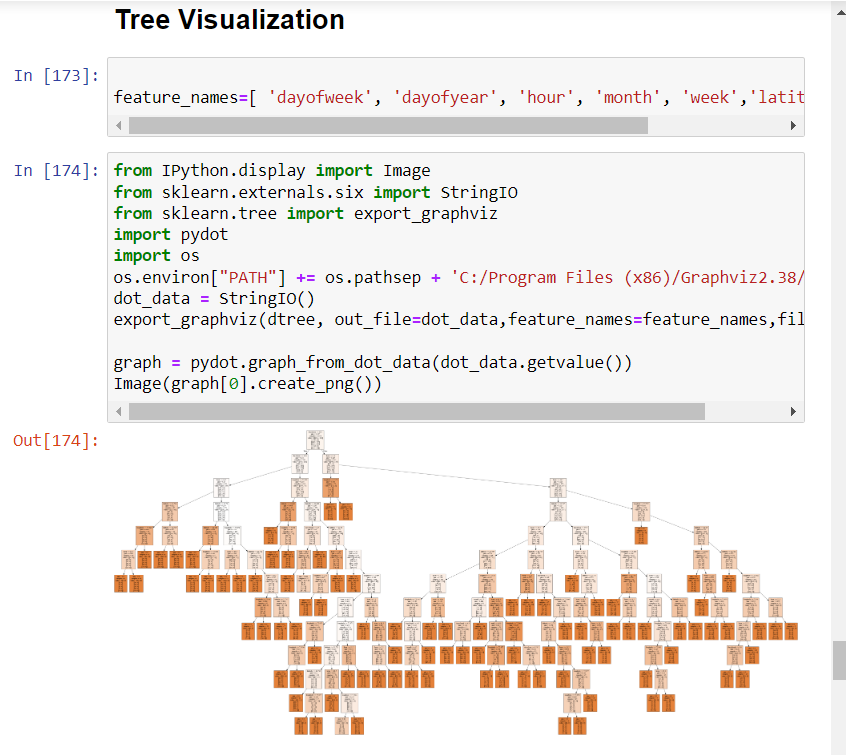


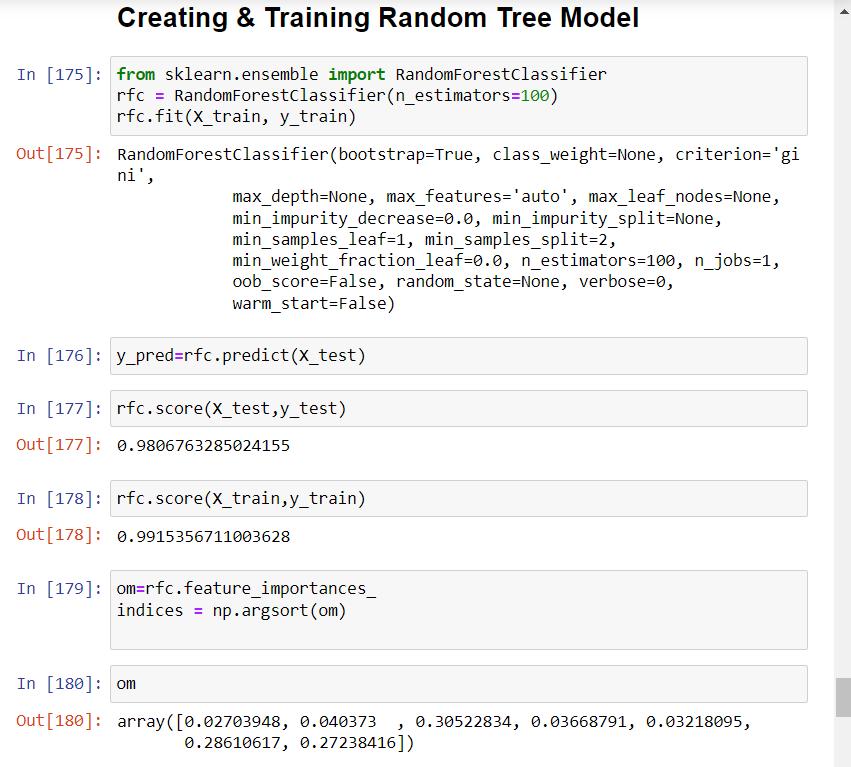


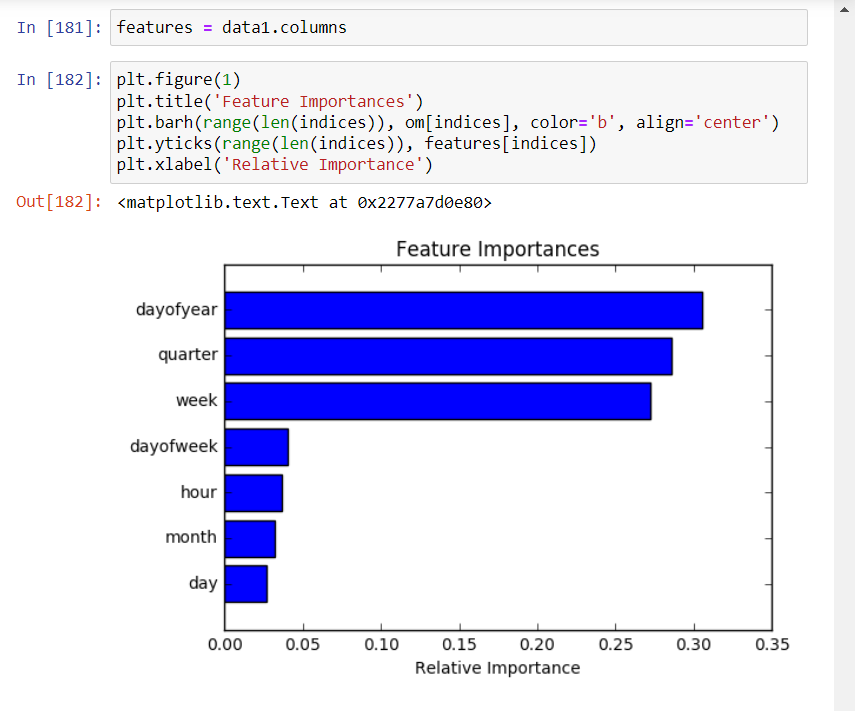












**7. References**

|  |  |
| --- | --- |
| **Textbooks:** | **Websites:** |
| Programming Python, Mark Lutz | <https://www.w3schools.com/python/> |
| Head First Python, Paul Barry | <https://www.tutorialspoint.com/python/index.htm> |
| Core Python Programming, R. Nageswara Rao | <https://www.javatpoint.com/python-tutorial> |
| Learning with Python, Allen B. Downy | <https://www.learnpython.org/> |
|  | <https://www.pythontutorial.net> |