# PROFESSIONAL TRAINING REPORT

**at**

**Sathyabama Institute of Science and Technology**

**(Deemed to be University)**

Submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

By

## ALAPATI TEJASWI

**REG. NO: 39110040**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI – 600119, TAMIL NADU**

**APRIL 2022**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the Bonafide work of **ALAPATI TEJASWI** **(Reg. No: 39110040)** who carried out the project entitled **“PREDICTING THE GOLD PRICE”** under our supervision from November 2020 to March 2021.

## Internal Guide

**Dr. M.D Anto Praveena , M.E., Ph.D.,**

**Head of the Department**

**Submitted for Viva voce Examination held on** 

**Internal Examiner External Examiner**

**DECLARATION**

I **ALAPATI TEJASWI (Reg. No: 39110040)** hereby declare that the Project Report entitled **“PREDICTING THE GOLD PRICES”** done by me under the guidance of **Dr. M.D Anto Praveena** is submitted in partial fulfillment of the requirements for the award of Bachelor of Science degree in Computer Science.

| **DATE: 10/4/2022**  **PLACE: CHENNAI** | ALAPATI TEJASWI |
| --- | --- |
|  | **SIGNATURE OF THE CANDIDATE** |

## ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to the **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of **Computer Science and Engineering** for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr.M.D Anto Praveena, M.E., Ph.D., for** her valuable guidance, suggestions and constant encouragement paved the way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**TABLE OF CONTENTS**

| *S.NO* | TITLE | PAGE NO |
| --- | --- | --- |
|  | ABSTRACT | **vi** |
|  | LIST OF FIGURES | **ix** |
|  | LIST OF ABBREVIATIONS | **x** |
| 1 | INTRODUCTION | **01** |
|  | 1.1 OUTLINE OF THE PROJECT | **01** |
|  | 1.2 BACKGROUND OF THE PROJECT | **02** |
|  | 1.3 PROBLEM STATEMENT | **02** |
|  | 1.4 OBJECTIVES | **03** |
|  | 1.5 SCOPE | **03** |
| 2 | ALGORITHMS AND METHODS | **04** |
|  | 2.1 GENERAL | **04** |
|  | 2.2 PYTHON LIBRARIES USED | **05** |
|  | 2.3 OVERVIEW | **07** |
|  | 2.4 HARDWARE AND SOFTWARE REQUIREMENTS | **08** |
|  | 2.5 DESIGN CRITERIA | **08** |
|  | 2.6 EVALUATION OF INDICES | **09** |
|  | 2.7SUMMARY | **10** |
| **3** | SYSTEM IMPLEMENTATION | **11** |
|  | 3.1 GENERAL | **11** |
|  | 3.2 IMPLEMENTATION PHASE | **13** |
| 4 | RESULTS AND DISCUSSION | **15** |
| 5 | CONCLUSION | **16** |
| 6 | SOURCE CODE | **17** |
| 7 | REFERENCES | **19** |

**ABSTRACT**

Machine Learning plays a virtual role from past years in normal speech command, product recommendation as well as in the medical field also. Instead of this it provides better customer services and a safer automobile system. All of the things show that ML is trending technology in almost all fields, so we are trying to coined ML in our project. Nowadays the real estate market is a standout amongst the most focused regarding pricing and keeps fluctuating. People are looking to buy metals like gold and silver with their budgets and by analyzing market strategies. But the main disadvantage of the current system is to calculate a price of gold without necessary prediction about future market trends and result is price increase. So, the main aim of our project is to predict the accurate price of gold without any loss. There are many factors that must be taken into consideration for predicting the gold price and try to predict efficient gold pricing for customers with respect to their budget as well as also according to their priorities. So, we are creating a gold price prediction model by using the linear regression algorithm of machine learning. This model will help people to get the information about the gold prices variation and will be informative for investors. The result of this research provides that our linear regression model is a success and is useful.

Indexed Terms- Linear regression, machine learning

**LIST OF FIGURES**

| Fig No | Fig Name | Page No |
| --- | --- | --- |
| 1 | Formula images | 14 |
| 2 | Screenshot of the result | 15 |

**LIST OF ABBREVIATIONS**

| ML | MACHINE LEARNING |
| --- | --- |
| RMSE | ROOT MEAN SQUARED  ERROR |
| MSE | MEAN SQUARED ERROR |

**CHAPTER 1**

**INTRODUCTION**

The demand for gold never ends. The gold rate trend shows that gold is one of the best investment plans. So, it is wise to predict the trend of the gold rate. Different statistical models can be used for suitable modeling and prediction of data. Machine learning plays a major role from past years in image detection, spam reorganization, normal speech command, product recommendation and medical diagnosis. Present machine learning algorithms help us in enhancing security alerts, ensuring public safety and improving medical enhancements.

**1.1 OUTLINE OF THE PROJECT**

Machine learning has often been applied to the prediction of financial variables, but usually with a focus on stock prediction rather than commodities.

In my project, I chose to apply supervised learning to the prediction of gold prices in order to see what kind of success I could achieve. For the purposes of profitability, it is more important to predict the relative change in price tomorrow (i.e., whether the price will go up or down) than to predict the absolute price tomorrow.

**1.2 BACKGROUND OF THE PROJECT**

Gold has been the original store of value and medium of exchange to mankind for centuries till paper currency took over a couple of centuries ago. However, most of the sustainable paper currencies were backed by Gold till as late as 1971, when the Bretton Woods agreement was scrapped, and world currencies became a true ′𝐹𝑖𝑎𝑡′ currency.

Gold, however, continues to garner interest not only as a metal of choice for jewelry, but also as a store of value and often advisable part of a diversified investment portfolio as it tends to be an effective inflation hedge and haven when economies are going through a rough patch.

**1.3 PROBLEM STATEMENT**

Forecasting is a function in management to assist decision making. It is also described as the process of estimation in unknown future situations. In a more general term, it is commonly known as prediction which refers to estimation of time series or longitudinal type data. Gold is a precious yellow commodity once used as money. It was made illegal in the USA 41 years ago but is now once again accepted as a potential currency. The demand for this commodity is on the rise. The gold prices are time series data of gold prices fixed twice a day in London. Factors influencing gold prices are many and we must be selective in this study to ensure that the model developed is significant.

**1.4 OBJECTIVE**

Objective of this study was to develop a forecasting model for predicting gold prices based on economic factors such as inflation, currency price movements and others. Following the melt-down of US dollars, investors are putting their money into gold because gold plays an important role as a stabilizing influence for investment portfolios. Due to the increase in demand for gold in Malaysia and other parts of the world, it is necessary to develop a model that reflects the structure and pattern of the gold market and forecast movement of gold price. The most appropriate approach to the understanding of gold prices is the Linear Regression model

**1.5 SCOPE OF THE PROJECT**

We are required to model the gold price prediction model with the available independent variables and with gold price as the single dependent variable. Many factors determine the price of gold and several economic factors have been identified to have influence on the gold prices. Variables such as Commodity Research Bureau future index (CRB); USD/Euro Foreign Exchange Rate (EUR USD); Inflation rate (INF); Money Supply (M1); New York Stock Exchange (NYSE); Standard and Poor 500 (SPX); Treasury Bill (T-BILL) and US Dollar index (USDX) were considered to have influence on the prices.

**CHAPTER 2**

**ALGORITHMS AND METHODS**

**2.1 GENERAL**

Python is an interpreted, object oriented, high -level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for rapid application development, as well as for use as scripting or glue language to connect existing components together. Python’s simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The python interpreter and the extensive standard library are available in source or binary form without change for all major platforms and can be freely distributed.

Often programmers are more compatible with python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging python programs is easy; a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program does not catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in python itself, testifying to python’s introspective power. On the other hand, often the quietest way to debug a program is to add a few print statements to the source. The fast edit-test-debug cycle makes this simple approach very effective.

**2.2 PYTHON LIBRARIES USED**

**2.2.1 Sci-kit Learn or Sklearn for Machine Learning**

Scikit-learn (formerly scikits.learn and also known as sklearn) is a [free softwar](https://en.wikipedia.org/wiki/Free_software)e [machine](https://en.wikipedia.org/wiki/Machine_learning) [learnin](https://en.wikipedia.org/wiki/Machine_learning)g [library](https://en.wikipedia.org/wiki/Library_(computing)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) [programming language](https://en.wikipedia.org/wiki/Programming_language). It features

various [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support vector](https://en.wikipedia.org/wiki/Support_vector_machine) [machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [*k*-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN](https://en.wikipedia.org/wiki/DBSCAN), and is designed to interoperate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy) and [SciPy](https://en.wikipedia.org/wiki/SciPy).

Scikit-learn is largely written in Python and uses [NumPy](https://en.wikipedia.org/wiki/Numpy) extensively for high-performance linear algebra and array operations. Furthermore, some core algorithms are written

in [Python](https://en.wikipedia.org/wiki/Cython) to improve performance. Support vector machines are implemented by a Python wrapper around [LIBSVM](https://en.wikipedia.org/wiki/LIBSVM); logistic regression and linear support vector machines by a similar wrapper around [LIBLINEAR](https://en.wikipedia.org/wiki/LIBLINEAR). In such cases, extending these methods with Python may not be possible.

Scikit-learn integrates well with many other Python libraries, such as [matplotlib](https://en.wikipedia.org/wiki/Matplotlib) and [plotly](https://en.wikipedia.org/wiki/Plotly) for plotting, [NumPy](https://en.wikipedia.org/wiki/NumPy) for array vectorization, [pandas](https://en.wikipedia.org/wiki/Pandas_(software)) data frames, [SciPy](https://en.wikipedia.org/wiki/SciPy), and many more.

### 2.2.2 Importing libraries and modules necessary for this project

Python code in one [module](https://docs.python.org/3/glossary.html#term-module) gains access to the code in another module by the process of [importing](https://docs.python.org/3/glossary.html#term-importing) it. The [import](https://docs.python.org/3/reference/simple_stmts.html#import) statement is the most common way of invoking the import machinery, but it is not the only way. Functions such as [importlib.import\_module()](https://docs.python.org/3/library/importlib.html#importlib.import_module) and built- in [import ()](https://docs.python.org/3/library/functions.html#__import__) can also be used to invoke the import machinery.

The [import](https://docs.python.org/3/reference/simple_stmts.html#import) statement combines two operations; it searches for the named module, then it binds the results of that search to a name in the local scope. The search operation of the import statement is defined as a call to the [import ()](https://docs.python.org/3/library/functions.html#__import__) function, with the appropriate arguments. The return value of [import()](https://docs.python.org/3/library/functions.html#__import__) is used to perform the name binding operation of the import statement. See the import statement for the exact details of that name binding operation.

**IMPORTING LIBRARIES**

import pandas as pd

import numpy as np

from math import sqrt

from numpy import log

from pandas import Series

import matplotlib.pyplot as plt

from matplotlib.dates import date2num

import seaborn as sns

from datetime import datetime

import subprocess

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import r2\_score

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import mean\_squared\_error

**2.3 OVERVIEW**

* Python is interpreted - python is processed at runtime by the interpreter.
* You do not need to compile your program before executing it. This is like PERL and PHP
* Python is interactive - you can sit at a python prompt and interact with the interpreter directly to write your programs.
* Python is object-oriented - python supports object-oriented style or technique of programming that encapsulates code within the projects.’
* Python is a beginner’s language - python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

As shown in the above PYTHON plays a major role in all applications development and we are doing projects based on the basic syntax. As we can send anything with our convenience.

**2.4 HARDWARE AND SOFTWARE REQUIREMENTS**

Programming language: Python 3.7.0 and above versions

Hardware requirements: Computer /Laptop with good internet connection

Software requirements: Microsoft Windows 10, Anaconda, Jupyter notebook or Google Collaboratory

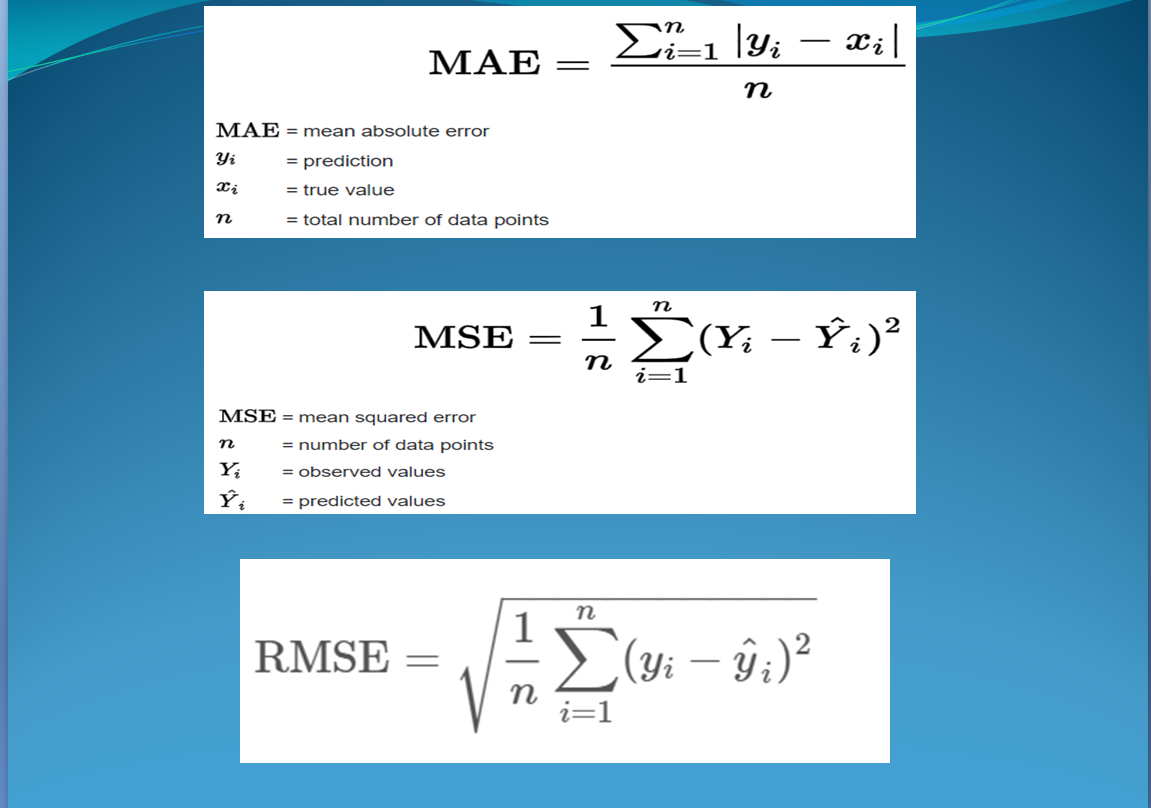
**2.5 DESIGN CRITERIA**

* simplicity: easily understood
* efficiency: uses minimal resources
* completeness: solves the problem
* not independent
* simplicity by default

**2.6 EVALUATION OF INDICES**

In our project we are going to evaluate some indices.

Multiple evaluating criteria are used for comparing the performance of regression models. The performance evaluation indices used here are Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Mean absolute Error (MAE), Coefficient of determination (R2 score).



**2.7 SUMMARY**

As we need to connect to internet for opening websites and gather information about the skills required for the jobs, we can also use other programming languages like C, C++, Java but its effective to use and have predefined functions to use in the python language since the syntax for this programming language is simple to use and is more effective comparatively. So, we have used python for this project for accessing sites and recommended skills as per the present trending technologies.

# CHAPTER 3

**SYSTEM IMPLEMENTATION**

## 3.1 GENERAL

Implementation is the stage in the project where the theoretical design is turned into a working system. The implementation phase constructs, installs and operates the new system. The most crucial stage in achieving a new successful system is that it will work efficiently and effectively. In order to implement this Skills Recommender project firstly we need to install anaconda software in our system. There are some steps to be followed to while installing anaconda in our system and the steps are as follows:

## STEP 1:

<https://repo.anaconda.com/archive/Anaconda3-2021.05-Windows-x86_64.exe>

Download and Install anaconda from the above link

Select the default options when prompted during the installation of Anaconda.

**STEP 2:**

Open “Anaconda Prompt” by finding it in the Windows (Start) Menu.

**STEP 3:**

 Type the command to verify that Anaconda was installed.

**python --version**  
Python 3.7.3

**STEP 4:**

Type the below command to update Anaconda.

conda update --all --yes

**STEP 5:**

Type the below command to start the Jupyter Notebook.

**jupyter notebook**

**STEP 6:**

Later, we don’t have to do all the above steps.

Directly open anaconda then from the given options select Jupyter notebook and do your work.

**3.2 IMPLEMENTATION PHASE**

### 3.2.1 Load data from CSV files

CSV (comma-separated value) files are a common file format for transferring and storing data. The ability to read, manipulate, and write data to and from CSV files using Python is a key skill to master for any data scientist or business analyst. In this post, we’ll go over what CSV files are, how to read CSV files into [Pandas Data Frames](https://www.shanelynn.ie/using-pandas-dataframe-creating-editing-viewing-data-in-python/), and how to write Data Frames back to CSV files post analysis.

[Pandas](https://pandas.pydata.org/) is the most popular data manipulation package in Python, and [Data Frames](https://www.datacamp.com/community/tutorials/pandas-tutorial-dataframe-python) are the Pandas data type for storing tabular 2D data.

The basic process of loading data from a CSV file into a [Pandas Dataframe](https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.html) (with all going well) is achieved using the “read\_csv” function in Pandas:

gold\_data= pd.read\_csv("gld\_price\_data.csv")

Once we selected our best algorithm (i.e., Linear regression), we proceeded to experiment with fine-tuning our model to improve prediction accuracy. By fine-tuning the model, we are referring to modifications other than tweaking the parameters of the earning algorithm.

### 3.2.2 Train and test splitting

For this section we will take the gold dataset and split the data into training and testing subsets. Typically, the data is also shuffled into a random order when creating the training and testing subsets to remove any bias in the ordering of the dataset.

For this will use StratifiedShuffleSplit.

StratifiedShuffleSplit cross-validator provides train/test indices to split data in train/test sets.

This cross-validation object is a merge of Stratified Fold and Shuffle Split, which returns stratified randomized folds. The folds are made by preserving the percentage of samples for each class.

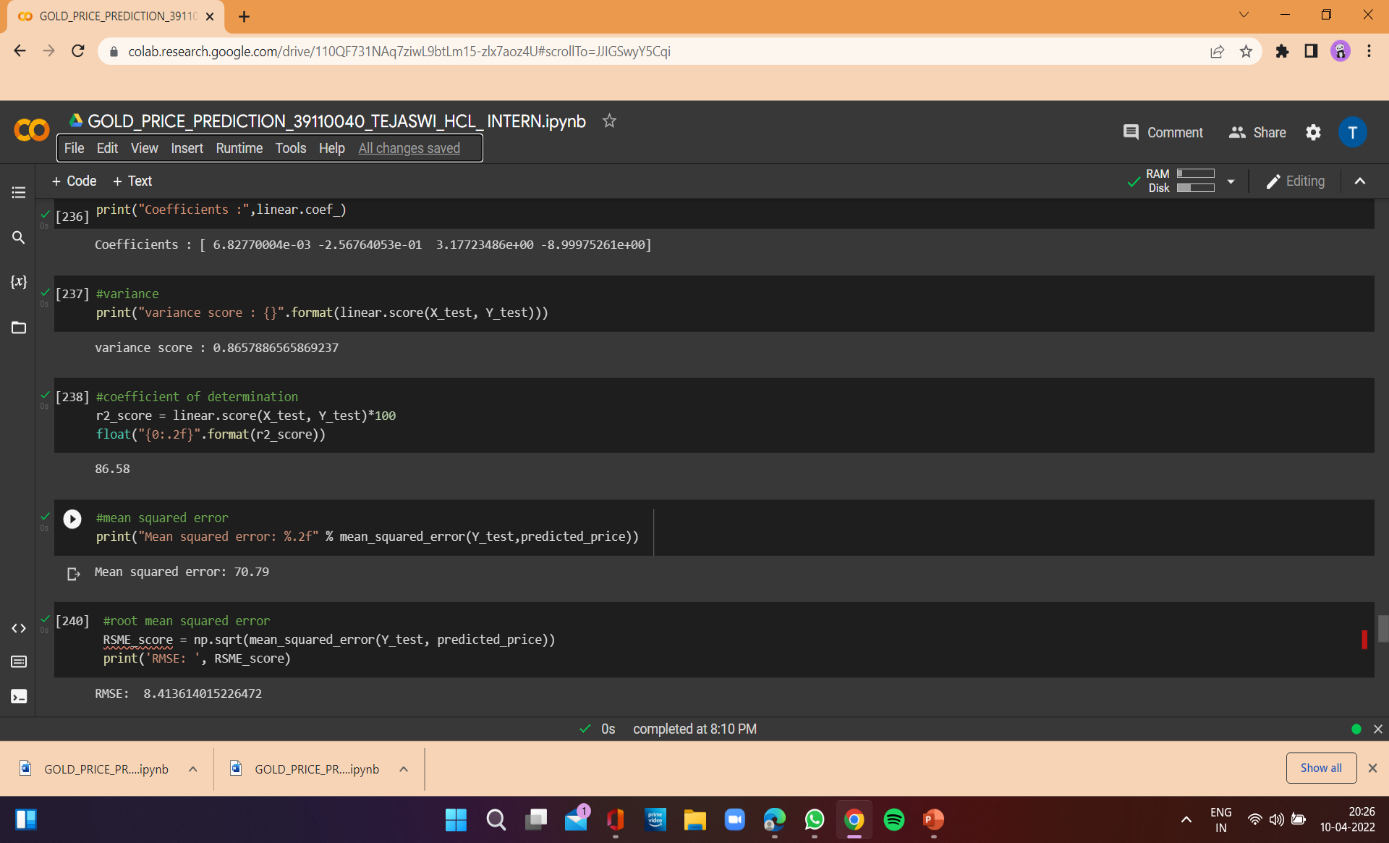
Generally splitting the data into testing and training set:

X\_train, X\_test, Y\_train, Y\_test = train\_test\_split (X, Y, test\_size = 0.2, random\_state=2)

**CHAPTER 4**

**RESULTS AND DISCUSSION**

On executing the python code of predicting rented bike count , we will get MSE, RMSE, Coefficient of determination as follows :



This is output of the executed python code

MSE = 70.79

RMSE = 8.413614015226472

Coefficient of determination = 86.58

**CHAPTER 5**

**CONCLUSION**

As you saw in this project, we first train a machine learning model, then use the trained model for prediction.​

Similarly, any model can be made much more precise, by feeding a very large dataset, to get a very accurate score (but it will be time-consuming).​

Here we used linear regression to predict the gold prices and compared the actual and predicted prices.​

This is very informative for investors and people who tends to buy or sell gold and can predict the values beforehand.

**CHAPTER 6**

**SOURCE CODE**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, box and whisker chart

Description automatically generated

**Graphical user interface, text, application, email

Description automatically generated**

**CHAPTER 7**

**REFERENCES**

Altman, N.S. (1992). An introduction to kernel and nearestneighbor nonparametric regression.European Conference on Complex Systems, University of Warwick in UK (vol. 2009). Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. doi:10.1023/A:1010933404324 Breiman, L. (2017). Classification and regression trees. Routledge, Taylor and Francis, FL. Chen, T. (2003). A fuzzy back propagation network for output time prediction in a wafer fab. Applied Soft Computing, 2(3), 211–222. doi:10.1016/S1568-4946(02) 00066-2 Chen, T. (2007). Advanced Engineering Informatics, 21(1), 55–65. doi:10.1016/j.aei.2006.10.002 Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. Genomics, 99(6), 323–329. doi:10.1016/ j.ygeno.2012.04.003 Corcoran, J., Li, T., Rohde, D., Charles-Edwards, E., & Mateo-Babiano, D. (2014). doi:10.1016/j.jtrangeo.2014.09.003 DeMaio, P. (2009).. Deng, H., & Runger, G. (2013). Gene selection with guided regularized random forest. Pattern Recognition, 46(12)