D.Y.PATIL COLLEGE OF ENGINEERING

AND TECHNOLOGY, KOLHAPUR.

Domain Specific Mini Project

Title : Recession Prediction

Submitted By:

Name	Roll no.
1.Nuren Imtiyaz Pathan	**20**
2.Aryaa Prashant Hanamar	**21**
3.Heena Mashak Mulla	**22**
4.Akshaya Appaso Shinde	**23**
5.Siddharth Yogesh Nashikkar	**24**

UNDER THE GUIDANCE OF:

<u>Mrs. D. R. Patil</u>

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Year: 2019-2020

D. Y. PATIL COLLEGE OF ENGINEERING AND TECHNOLOGY, KOLHAPUR

CERTIFICATE

This is to certify that the project group consisting of following members have satisfactorily completed the Domain Specific Mini Project work entitled " RECESSION PREDICTION " at TE (CSE) semester – VI as prescribed in the syllabus of Shivaji University for the academic year 2019-2020.

Roll No.	Name	Exam No.
20	Nuren Imtiyaz Pathan	2017078166
21	Aryaa Prashant Hanamar	2017078181
22	Heena Mashak Mulla	2017078197
23	Akshaya Appaso Shinde	2017078168
24	Siddharth Yogesh Nashikkar	2017078071

Date:24-05-2020

Place:Kolhapur

Project Guide   Project Coordinator   HOD

Examiner   Principal

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ACKNOWLEDGEMENT

We the students of Third year from Department of Computer Science and Engineering of D. Y. Patil College of Engineering and technology have successfully completed the Domain Specific Mini Project on the topic "Recession Prediction" for the academic year 2019-2020.We whole heartedly express our sincere gratitude to our project guide Mrs. D. R. Patil for the valuable guidance towards the completion of this project.

Name of student   Sign

1.Nuren Imtiyaz Pathan

2.Aryaa Prashant Hanamar

3.Heena Mashak Mulla

4.Akshaya Appaso Shinde

5.Siddharth Yogesh Nashikkar

INDEX:

INTRODUCTION:

A recession is when the economy experiences a marked slippage in economic activity. There's a drop in the following five economic indicators: real gross domestic product, income, employment, manufacturing, and retail sales. People often say a recession is when the GDP growth rate is negative for two consecutive quarters or more. These are the indicators to watch if you want to know when the economy is in a recession.

(a)Real income measures personal income adjusted for inflation. Transfer payments, such as Social Security and welfare payments, are also removed. When real income declines, that reduces consumer purchases and demand.

(b)Employment as measured by the monthly jobs report. Here's an analysis of the current jobs statistics.

(c)The health of the manufacturing sector, as measured by the Industrial Production.

(d)Manufacturing and wholesale-retail sales adjusted for inflation.

(e)The NBER also looks at monthly estimates of GDP provided by Macroeconomic Advisers.

Recessions are characterized by a rash of business failures and often bank failures, slow or negative growth in production, and elevated unemployment. The economic pain caused by recessions, though temporary, can have major effects that alter an economy. Hence analysis is required to give advance warning of a recession and avoid ill effects on the economy.

Objectives:

1. To predict the future recession that is it will tell the period of recession in future.

2. Period of current recession.

REQUIREMENT SPECIFICATION:

A recession is commonly understood to be a widespread and prolonged decline in economic activity which leaves its impact on that economy for a very long time. Predicting this recession before its occurrence can help in taking up the preventive measures before time and avoid the loss to a great extent.

The prime requirement of this project is to predict the probability of recession of the entered year on the basis of the data of previous years that will be provided to train the model.

The project will be an window based application where the user will be asked to enter the year for which prediction should be done and click on predict. Once the year is taken the model will predict the recession possibility in that year.

Technologies used-

-   Python 3:version 3.6
-   Machine learning algorithms

Software Requirements-

-   Microsoft Windows 8 or later
-   Pycharm
-   Microsoft Excel

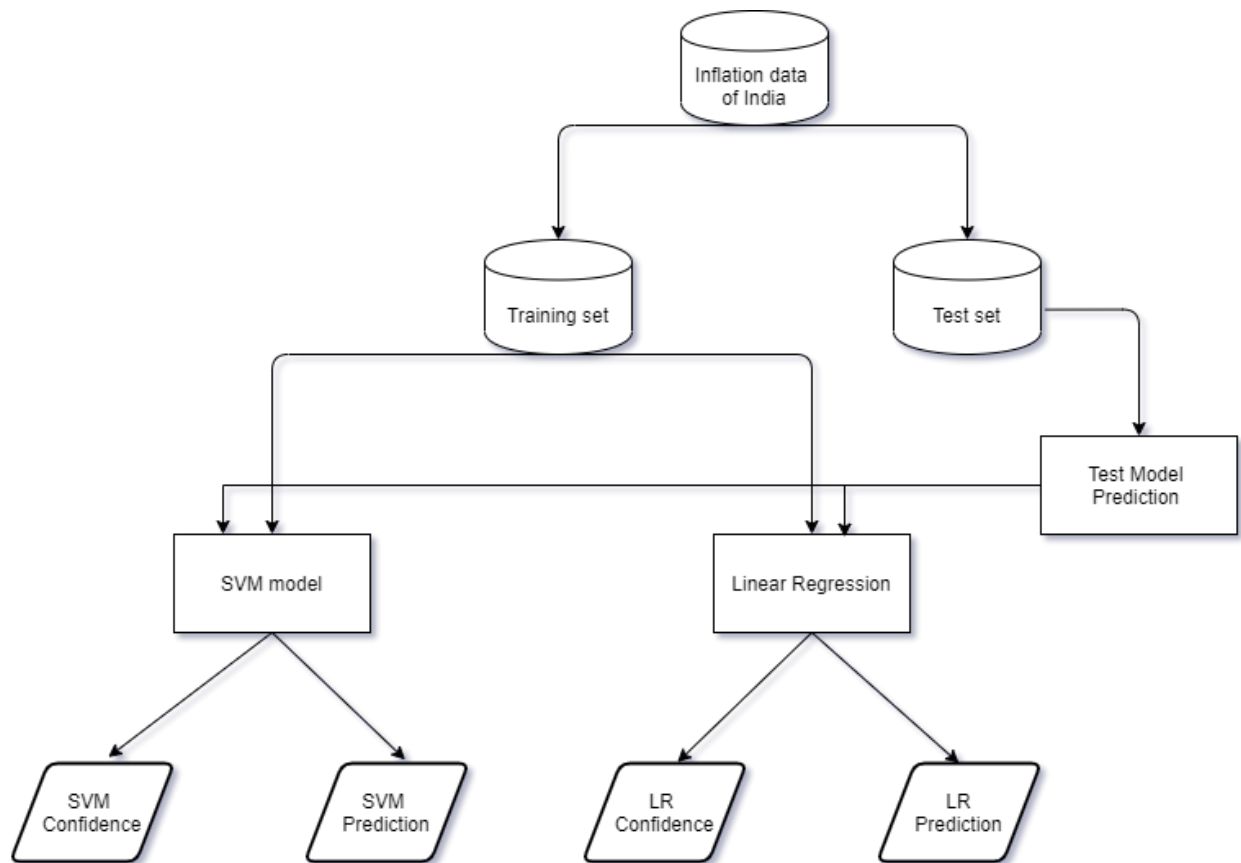Hardware requirements-
-   Intel Core i3 processor
-   RAM:4GB
-   Disk space: 2.5GB and another 1 GB for caches
-   Monitor Resolution: 1024*768

Libraries used-
-   Pandas & Matplotlib
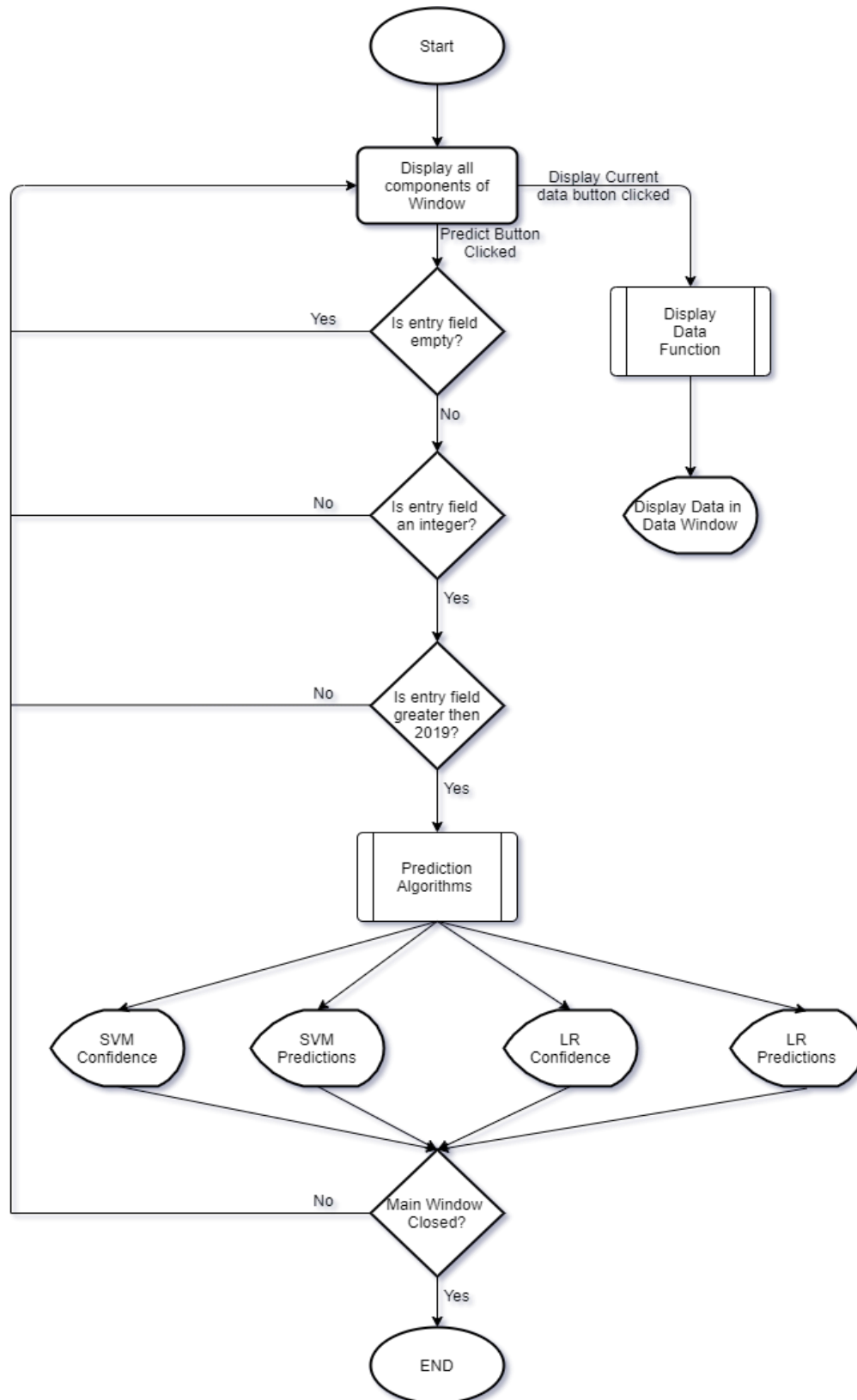-   Numpy
-   Sklearn
-   tkinter

DESIGN:

System Architecture:

Explanation:

The dataset used here consists of the data about inflation rate in India from the year 1958 to 2018. After data cleaning it was split into train data and test data with the help of train_test_split function from sklearn. We have used two different algorithms for model training i.e. SVM and linear regression. On the basis of test model prediction and the specific model(SVM or LR) a prediction of the rate of inflation for the entered year is made and displayed along with the confidence score of that model so that we can consider the results of the model with more accuracy.

Flow chart:

```
                          ┌─────────┐
                          │  Start  │
                          └────┬────┘
                               │
                               ▼
        ┌──────────────┐  Display Current
        │  Display all │  data button clicked
        │ components of │──────────────────────┐
        │    Window     │                       │
        └──────┬────────┘                       ▼
          Predict Button              ┌──────────────┐
            Clicked                   │   Display    │
               │                      │    Data      │
               ▼                      │   Function   │
   Yes    ◇ Is entry field ◇          └──────┬───────┘
  ◄───────◇     empty?    ◇                  │
          ◇              ◇                   ▼
               │ No               ┌──────────────┐
               ▼                  │ Display Data in│
   No     ◇ Is entry field ◇      │  Data Window   │
  ◄───────◇ an integer?   ◇       └────────────────┘
          ◇              ◇
               │ Yes
               ▼
   No     ◇ Is entry field ◇
  ◄───────◇ greater then   ◇
          ◇    2019?       ◇
               │ Yes
               ▼
        ┌──────────────┐
        │  Prediction  │
        │  Algorithms  │
        └──────┬───────┘

  SVM        SVM         LR          LR
Confidence Predictions Confidence Predictions

          ◇ Main Window ◇
   No     ◇   Closed?    ◇
  ◄───────◇             ◇
               │ Yes
               ▼
          ┌─────────┐
          │   END   │
          └─────────┘
```

Explanation:

1. Start.
2. A window window with all the components.
3. On clicking the Display current dataset button the entire dataset is displayed in a separate window.
4. Next the user is asked to enter the required year and click on the predict button.
5. After entering the valued is checked if the entry field is empty.
   a. If yes then an exception occurs which has been handled in the code and goto step 4.
   b. If no then proceed to step 6.
6. Now we have to check if the entered value is an integer.
   a. If yes the proceed to step 7.
   b. If no a message "Enter valid number is displayed" and goto step 4.
7. If the entered year is not greater than 2019 then the user is asked to re-enter the value which should be greater then 2019.
8. Later in the flow the once the correct input from the user is obtained prediction for the entered year is done on the basis of our training algorithms.
9. The predicted result is then displayed on the window along with the confidence score.
10. Main window closed
    a. If yes then END
    b. If no then goto step 4.

Algorithms used:

1. Linear Regression: Linear regression is a linear model, e.g. a model that assumes a linear relationship between the input variables (x) which is the date of the year in our case and the single output variable (y) which is the inflation rate. More specifically, that y can be calculated from a linear combination of the input variables (x)

2. Support vector regression(SVR): Support Vector Regression(SVR) is quite different than other Regression models. It uses the Support Vector Machine(SVM, a

classification algorithm) algorithm to predict a continuous variable which is GDP in our case.

IMPLEMENTATION DETAILS:

Function used:

1. def displayData():
   This function is used for creating buttons and labels. when we click on the button it will run the function with whom it is initialized and display the output.
2. def displayPredict():
   This function is used for predicating the inflation value after 2019.It takes valid input from users and gives the required results.

Pseudo code of algorithms used in project:

Import the required libraries like quandl, numpy,

LinearRegression, SVR, train_test_split, train_test_split, pyplot and tkinter.

Read the dataset using read_csv() and store it in variable 'df'

// For tinkteri.ewindow based application:

copy the values from 'df' into 'InitialData'

initialize :Tkinter

//Create the frames and pack it using pack function

  TopFrame, MidFrame, LowFrame = Frame(MainWindow).pack()

//Generate a SVR and LR Frame, give name and pack it

  svrFrame, lrFrame = Lable(). pack()

// function to display the data

DisplayData()

create a button name "databutton" and initialize its parameters

databutton(cmd)= DisplayData // when we click on the button it will run the function with whom it is initialized

create a label for the entry of "NoOfMonths" and initialize with its required parameters

[END OF DisplayData()]

// Function to Display the Prediction

DisplayPredict()

if NoOfMonths is null:

then : display error

else if NoOfMonths< 2019 or NoOfMonths = string

then: return enter valid number

else :

create a new col "prediction" and shift the values

//create X and Y var and store the values in array format

split data into x_train, x_test, y_train, y_test using train_test_split() function

// For SVR (Support vector machine using regression):

initialize svr_rbf = SVR()

fit the svr_rbf using fit()

svr_confidence = svr_rbf.score() to calculate the accuracy

// Similarly for LR (Linear regression):

initialize lr = LinearRegression()

fit the lr using fit()

lr_confidence = lr.score() to calculate the accuracy

use lr.predict() and svr.predict() to calculate the predictions using SVR or LR

[END OF DisplayPredict()]

// To present the predicted values on window based app

for i in the range of NoOfMonths:

display LR and SVR Prediction values

[END FOR]

//Declaring the graph

create a button "PredictButton" and set its parameters.

GraphX = yrs

GraphY = predicted values of LR

GraphY1 = predicted values of SVR

plot(GraphX, GraphY)

show()

plot(GraphX, GraphY1)

show()

[END]

Development tools used:

1. PyCharm: It is an integrated development environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains. It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django as well as Data Science with Anaconda. PyCharm is cross-platform, with Windows, macOS and Linux versions.

2. Libraries-
   - Pandas: pandas is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language. Data analysis was done using pandas.
   - Numpy: NumPy was used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined.
   - Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Data visualization was done using pyplot from matplotlib.
   - Sklearn: This module was used to implement the different machine learning algorithms used in the project.
   - Tkinter: Tkinter is the Python interface to the Tk GUI toolkit shipped with Python. The GUI of the projected id created using the library.

3. Microsoft Excel: Used Microsoft's product Excel to cover range of machine learning tasks such as data mining, data analytics, smart visualization

Supporting tools used:

1. train_test_split is a function in Sklearn model selection for splitting data arrays into two subsets: for training data and for testing data. With this function, you don't need to divide the dataset manually.

2. matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB . Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

TESTING AND RESULTS:

Testing is a process of executing a program or application with the intent of finding the software bugs. Software Testing is necessary because we all make mistakes. Some of those mistakes are unimportant, but some of them are expensive or dangerous. We need to check everything and anything we produce because things can always go wrong – humans make mistakes all the time. So there are some test cases provided below:
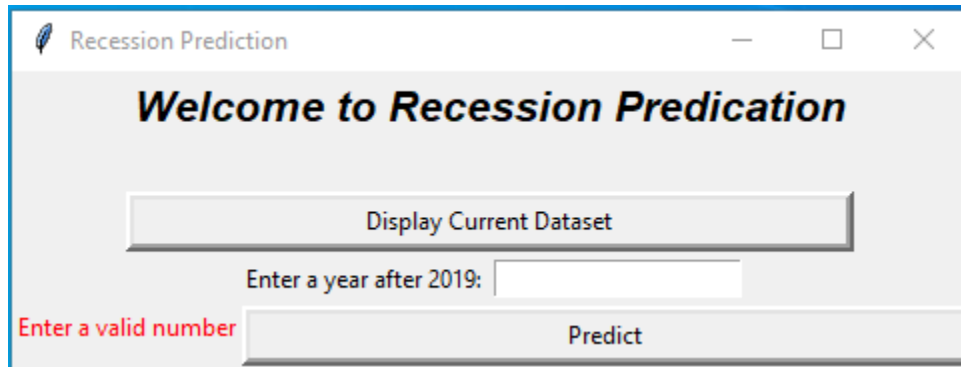
Test case 1:

If we enter invalid integer or string then it will gives the error and asking for entering the valid value.
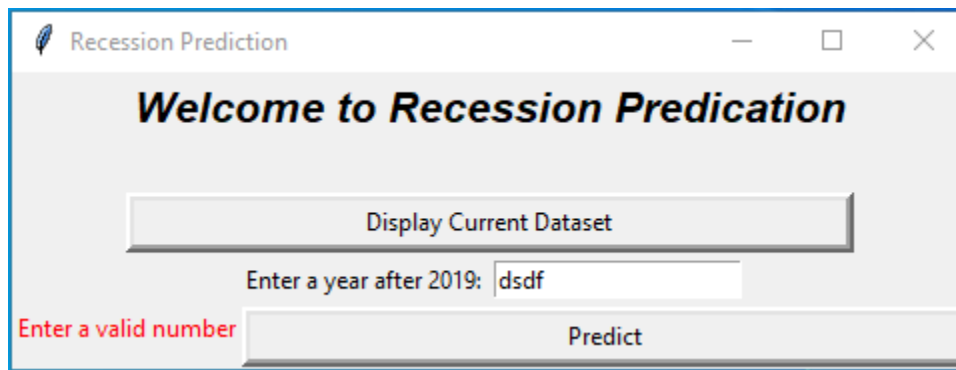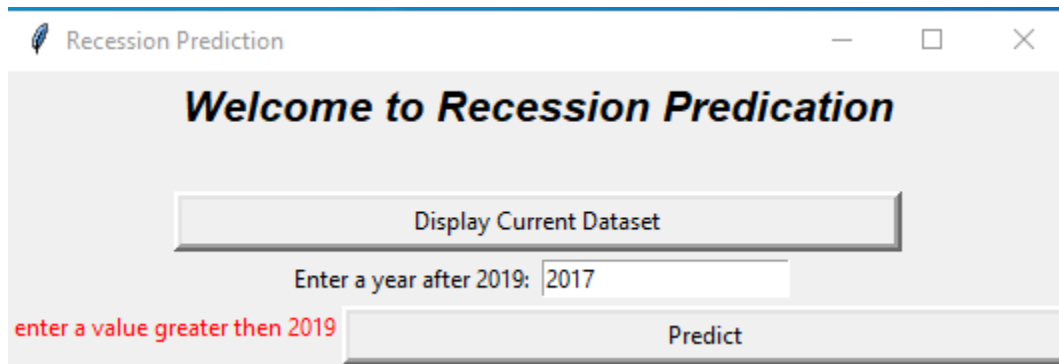

Test case 2:

If we don't give any value and push the predict button then it will give the error and ask to enter a value.
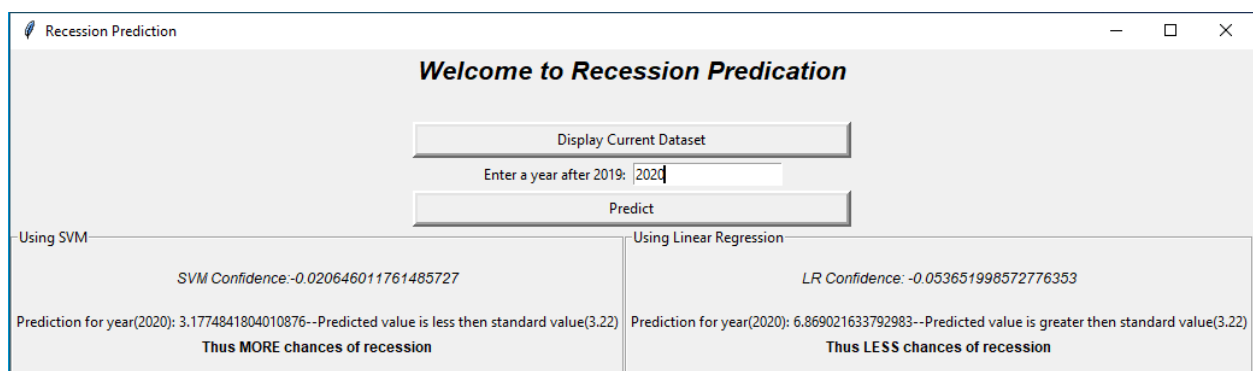
Test case 3:

If we give the previous year to predict then it also gives the error and asking for the years after 2019

Test case 4:

If we enter the valid value then we will get the results.
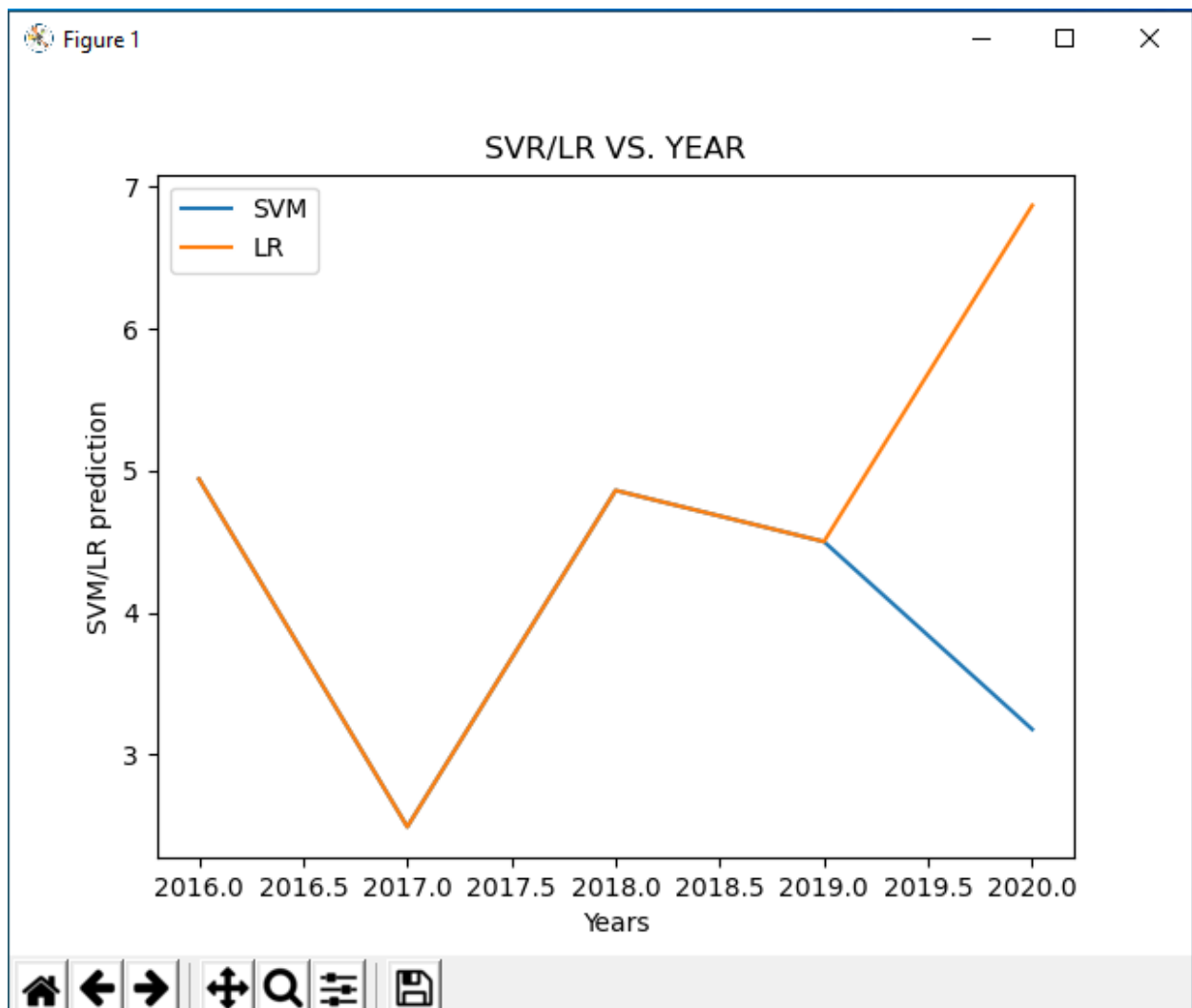

Input: year after 2019

Output: SVM confidence: -0.020   LR confidence :-0.0536

          SVM predication: 3.177484   LR predication: 6.86902

SVR/LR VS. YEAR

As SVM confidence is greater than LR confidence so we will consider SVM predication values. So there is more chances of recession in 2020.

CONCLUSION:

Recession has been a major problem for any economy today. The department hit by this problem needs a lot of time and efforts to cope up with the loss. The increased rates of unemployment , inflation and many other difficulties that may also lead to other major problems which include sucide rates among youth due too unemployment and many more. As a preventive measure we need to predict this situation before hand to take the necessary steps. Hence we have created this model where we have made use of Supervised learning algorithms like Linear regression and

support vector machine using regression(SVR) . We have successfully predicted recession on the basis of past data. The users of this application will be able to see the rate of inflation and accordingly understand the condition of future recession which will be a great boon for any economy.

FUTURE SCOPE:

To predict the duration of recession predicted in the specified year.

REFERENCES:

1. Dataset: https://www.google.com/url?q=http://machinelearningmastery.com/confidence-interval-for-machine-learning/&sa=D&source=hangouts&ust=1590228693540000&usg=AFQjCNEf7cK60-l3moe1IScQ19yEm5nPhg

2. https://towardsdatascience.com/linear-regression-understanding-the-theory-7e53ac2831b5

3. https://medium.com/coinmonks/support-vector-regression-or-svr-8eb3acf6d0ff