A PROJECT REPORT ON

"ONLINE SHOPPING PREDICTION USING SVM"

SUBMITTED BY

Miss. Kirti Chiplunkar	Exam No. 15
Miss. Sonal Dimbar	Exam No. 19
Miss. Stuti Thorat	Exam No. 74

UNDER THE GUIDANCE OF

PROF. PRADNYA GULHANE

DEPARTMENT OF COMPUTER ENGINEERING

ALL INDIA SHRI SHIVAJI MEOMRIAL SOCIETY'S

INSTITUTE OF INFORMATION TECHNOLOGY

KENNEDY ROAD, NEAR R.T.O. PUNE-411001

SAVITRIBAI PULE PUNE UNIVERSITY

2019-2020

ACKNOWLEDGMENT

It gives us immense pleasure in presenting the project report on **"ONLINE SHOPPING PREDICTION USING SVM"** The success and the outcome of this report required a lot of guidance. We are very grateful to our guide **Prof. Pradnya Gulhane** who has provided expertise and encouragement. We thank ma' am who provided vision and knowledge that was very helpful throughout the research. All that we have done is only due to the great guidance. We also express our gratitude to **Dr. S.N. Zaware** Head of Computer Engineering Department, AISSMS Institute of Information Technology, for the valuable support.

We would also like to extend our sincere thanks to Principal Dr.P.B. Mane, for his dynamic and valuable guidance throughout the project and providing the necessary facilities that helped us to complete our dissertation work.

Miss. Kirti Chiplunkar
Miss. Sonal Dimbar
Miss. Stuti Thorat

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Overview

1.2 Problem Definition

A Python program which makes use of Machine Learning Algorithm for predicting the needs and behavior of the customers in online shopping

1.3 Objective

To use Support Vector Machine (SVM) to predict needs and behavior of the customers in online shopping.
To effectively predict whether purchase will be made or not during next visit .

1.4 Project Scope

- This system will be useful for stock marketers.
- Currently, System predicts whether purchase will be made or not during next visit of the customer.

1.5 Methodology

We consider an online store to be available as a website where a web user may browse the site’ s pages, search for products, add them to a virtual shopping cart, and confirm a purchase. To build and evaluate SVM classifiers we used real data from commercial Web server access logs and records

.

Chapter 2

SOFTWARE REQUIREMENTS SPECIFICATION

2.1 Dependencies

Dependencies

Numpy, Spicy, Pandas, Scikit-Learn, Matplotlib.

2.2 System Requirements

2.2.1 Software Requirements(Platform Choice)

1. **Operating System:** Windows7 and higher / Linux
2. **IDE :** PyCharm
3. **Language Support :** Python 3.6 and higher

2.2.2 Hardware Requirements

1. **Disk Space:** Minimum disk space of 500 GB is expected for computations and storage means.
2. **Processor:** i5 CPU @1.60 GHz 1.80 GHz, 32-bit x32 OR 64-bit x64 pro- cessor is preferable.
3. **Memory:** 4 GB RAM and above , .
4. **Display:** 1600 * 900 minimum display resolution for better display.

Chapter 3

PROJECT IMPLEMENTATION

3.1 Tools and Technologies Used

1. **Pandas:** It's a package in python useful for scientific computation. It's a strong N-dimensional collection entity. It consists of tools for integrating C/C++. Useful for operating on linear algebra, Fourier transform, and random number capabilities.
2. **Scikit-learn:** It's a machine learning open source library supporting support vector machine. Regression and clustering via clubbing similar objects together.
3. **SciPy:** It's a package build on numpy for scientific and technical computations. It has functions for optimization.
4. **Pillow:** It's a open source library supporting the image manipulation. It provides the ease to access different types of formats of the image files. It is supported in Windows, Linux OS.
5. **NumPy:** NumPy is a package in Python used for Scientific Computing. NumPy package is used to perform different operations. The ndarray (NumPy Array) is a multidimensional array used to store values of same datatype. These arrays are indexed just like Sequences, starts with zero. NumPy uses much less memory to store data.

6. **Matplotlib:** Matplotlib is a part, rather say a library of python. Using Matplotlib you can plot graphs, histogram and bar plot and all those things. It is the Python's equivalent of MATLAB. And after that you can go with cufflinks and plotly as they'll give you an interactive feature to your plots.

3.2 Algorithm Details

3.1.1 Algorithm 1 : Support Vector Machine(SVM)

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. An SVM model is basically a representation of different classes in a hyperplane in multidimensional space.In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

In the SVM classifier, it is easy to have a linear hyper-plane between these two classes. But, another burning question which arises is, should we need to add this feature manually to have a hyper-plane. No, the SVM algorithm has a technique called the kernal trick. The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e. it converts not separable problem to separable problem.

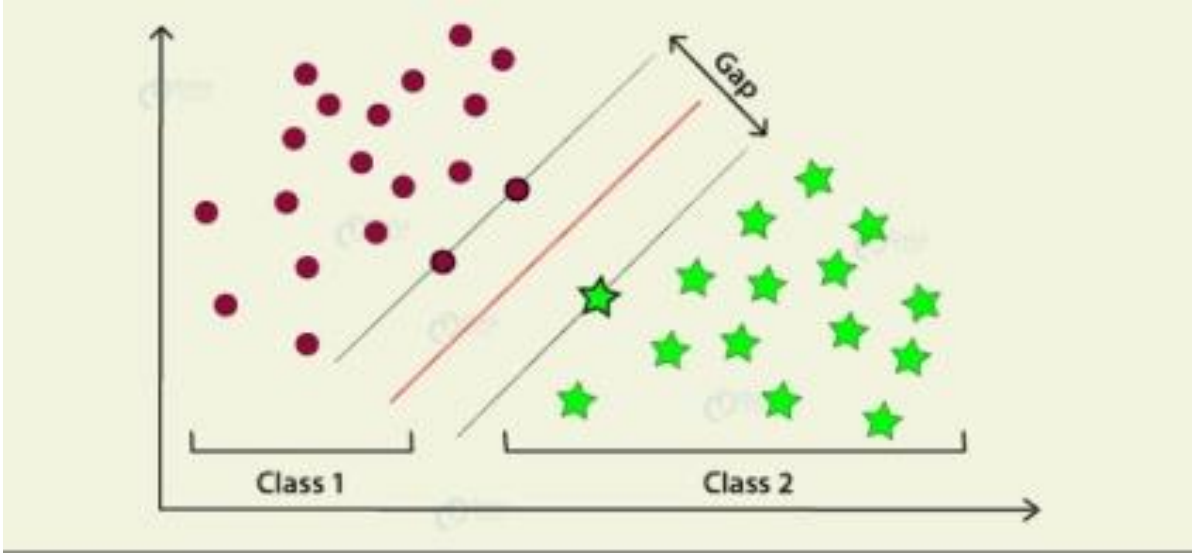

Figure 3.1.1:. Support Vector Machine(SVM)

Steps:

1. Importing the libraries.
2. Reading the dataset.
3. Splitting the dataset into Training set and testing set.
4. Feature Scaling.
5. Fitting the classifier into Training set.
6. Predicting result.
7. Visualizing the SVM results.

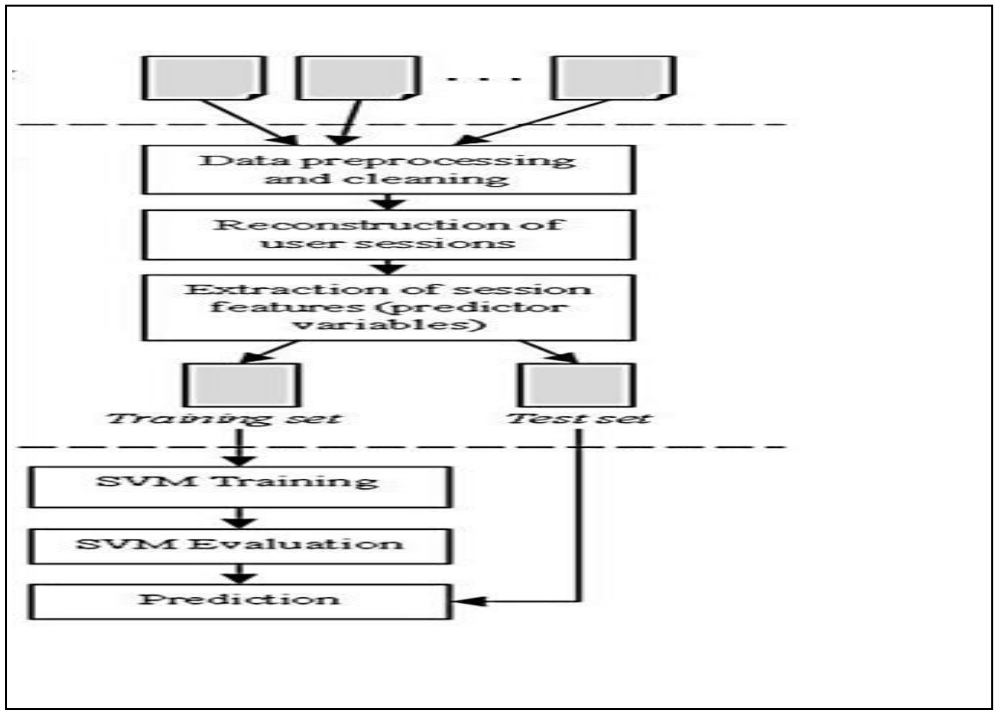

Figure 3.1.2:.Steps for Computing

Chapter 4

RESULTS

4.1 Screenshots

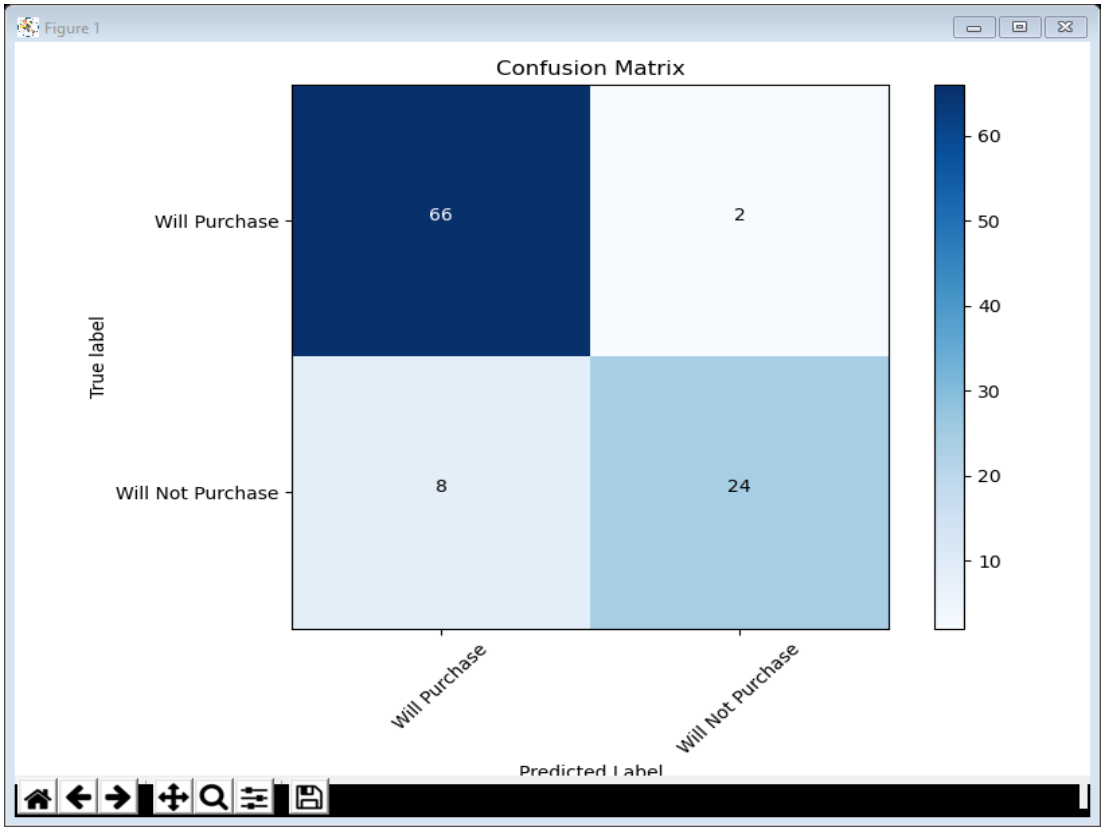

Figure 4.1:. Confusion matrix

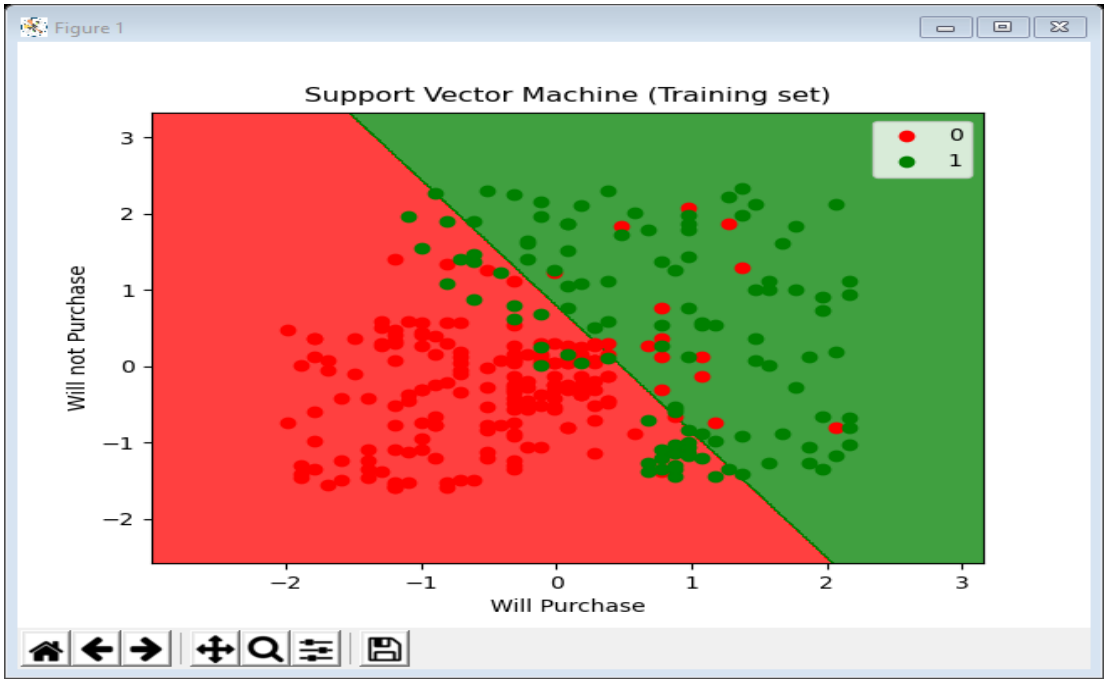

Figure 4.2:. Support Vector Machine (Training Set)

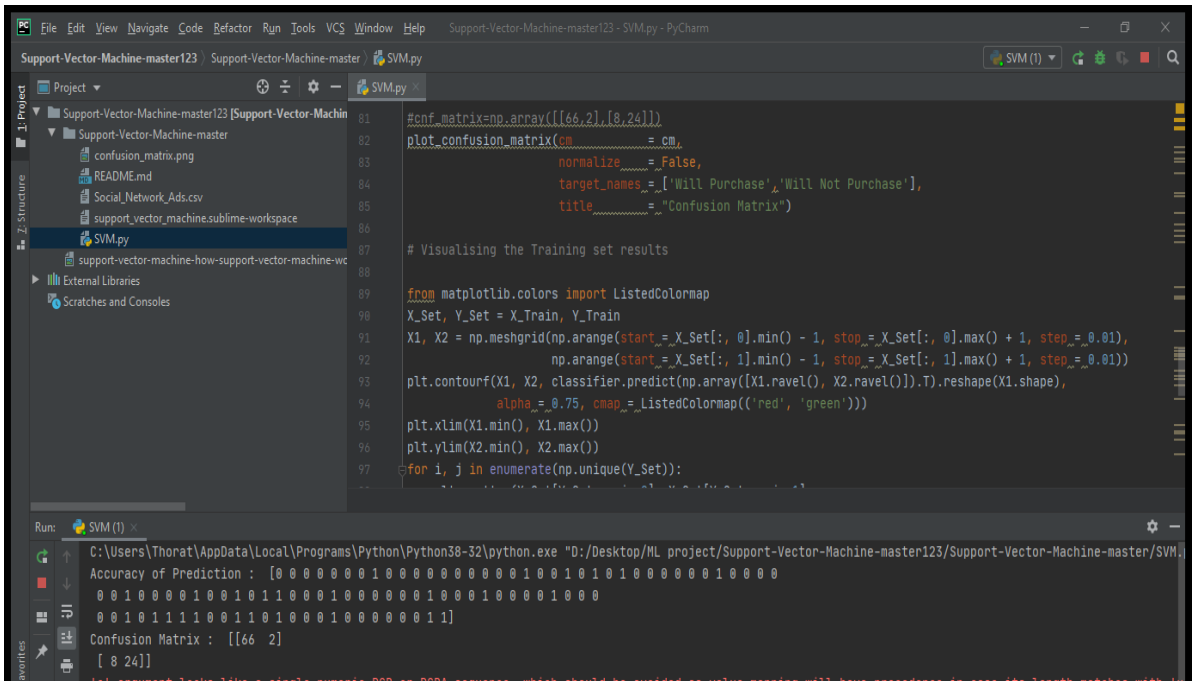

Figure 4.3:. Accuracy of Predication

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusion

Thus, we have implemented Support Vector Machine for Prediction of needs and behavior of the customers in online shopping

5.2 Future work

In future work, more different types of datasets could be used to see the results. Now, only limited dataset is used here, further larger datasets could be used. It would be interesting to integrate the proposed classifier into the process of managing active user sessions in the web store and to verify its efficiency in real time.