**AMAZON DATA RATING REVIEW ANALYSIS**

**AND PREDICTING RATING FROM REVIEW**

**USING LINEAR SVM ALGORITHM**

**MINI PROJECT**

**B. Tech in Computer Science and Engineering**

**5th semester**

2021-2022



**Submitted by:** **Guided By:**

Mr. Shaik Farhathulla Dr Saurabh Mishra

Roll. No.: 2015303 (Resource Person)

CSE-ML-5thSem

Session: 2021-2022

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**GRAPHIC ERA UNIVERSITY, DEHRADUN**

# ACKNOWLEDGEMENT

I would like to thank Saurabh Mishra sir, without whom, I might not have known about this technology.

Also, I would like to thank my parents for their support and encouragement and allowing me to reach this far.

Shaik Farhathulla

Roll No.:2015303

SEC: ML

## Session: 2021 – 2022

## GEU, Dehradun

### 1.1 About Project

I have created a SVM model, the model will predict the rating from reviews. In the sense, it simply says weather the review is positive or negative and predict the rating for that review. It is useful in our daily life for recommendation of a product or any other thing. I will be analyzing the reviews, and classifying the reviews weather they are negative reviews or positive reviews. After that I will be splitting the data into training and testing data. Then model will be evaluated and predicts the value.

### 1.2 Requirement of Project

**1.2.1 Hardware Requirement**

1.Laptop

#### 1.2.2 Software Requirement

Operating System : Windows 10 an application

Language : Python 3.8

Application : Anaconda, Jupyter Notebook

### 1.3 Introduction

I have used Machine learning to create my model.

**Machine learning** is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning** focuses on the development of computer programs that can access data and use it learn for themselves. Machine learning is of three types:

1. **Supervised learning:**

It is the **machine learning** task of **learning** a function that maps an input to an output based on example input-output pairs. A **supervised learning** algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

1. **Unsupervised learning:**

It is a type of **machine learning** that looks for previously undetected patterns in a data set with no pre-existing labels and with a minimum of human supervision.

1. **Reinforcement learning** (RL):

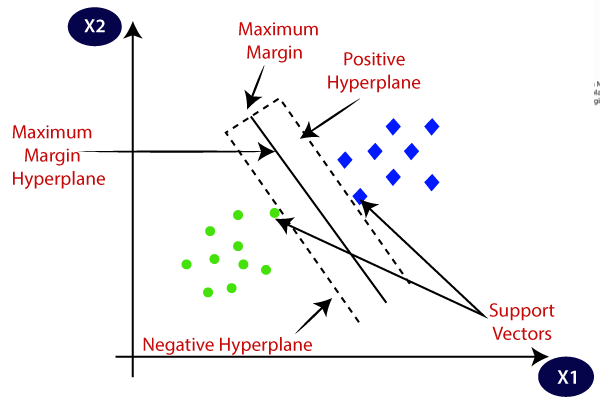
It is an area of **machine learning** concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward. **Reinforcement learning** is one of three **basic machine learning** paradigms, alongside supervised **learning** and unsupervised **learning**.

**Support Vector Machine (SVM) :**

SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

SVM algorithm can be used for Face detection, image classification, text categorization, etc.



#### 1.4 Working of the Project.

* **We will be importing Scikit Learn, Pandas, Seaborn, Matplotlib and NumPy.**

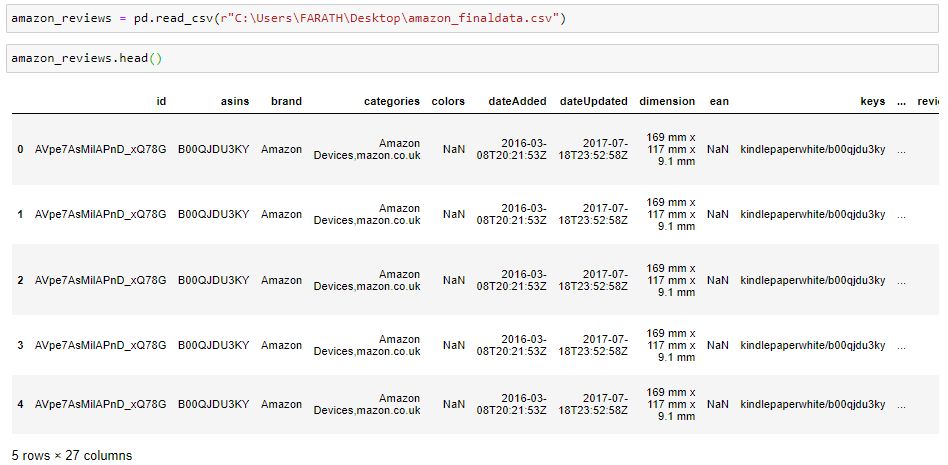
import numpy as np

import seaborn as sns

import pandas as pd

import matplotlib.pyplot as plt

* **Importing and reading the data**



amazon\_reviews.keys()

amazon\_reviews.info()

amazon\_reviews.describe()

### Sentiment Analysis on the data



I will be using VADER analyzer and analyzing my data and classifying the reviews which are present in my data.

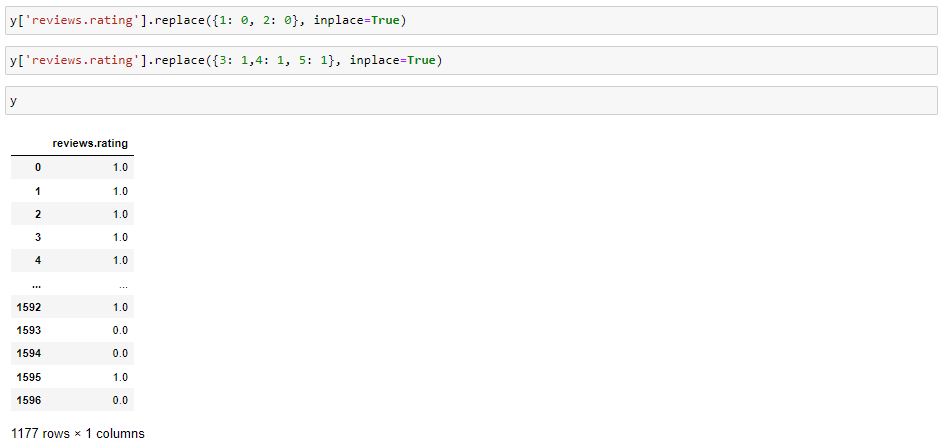
By this I would get the compound, negative, positive and neutral values which are used while training my model.

### Creating data frames



Here, we will be creating two data frames, which are used in training and evaluating the model.

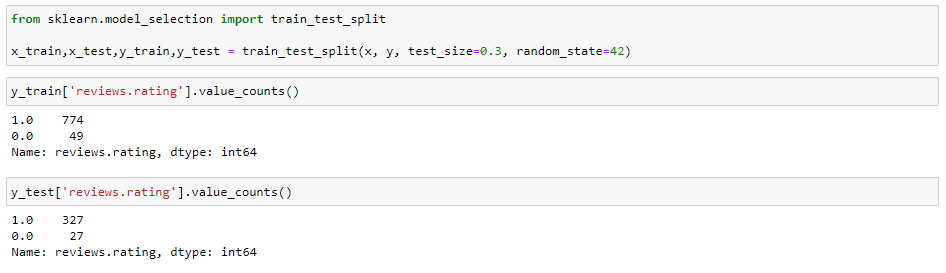
* **Replacing the values in data**



I would be replacing the 1 and 2 rating with 0.

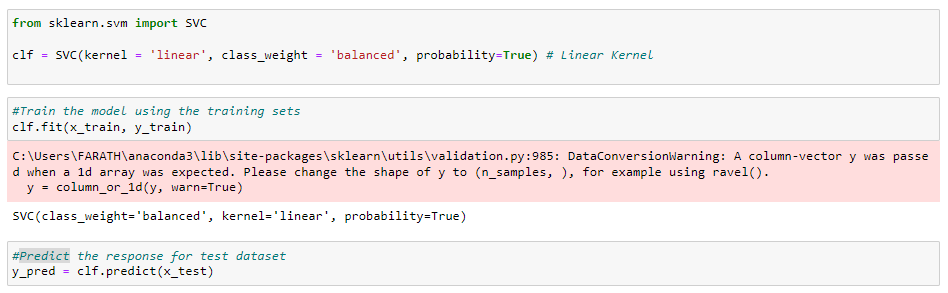
And 3, 4 and 5 with 1.

* + **Splitting the data**



I will be splitting the data into two parts one for training and other for testing

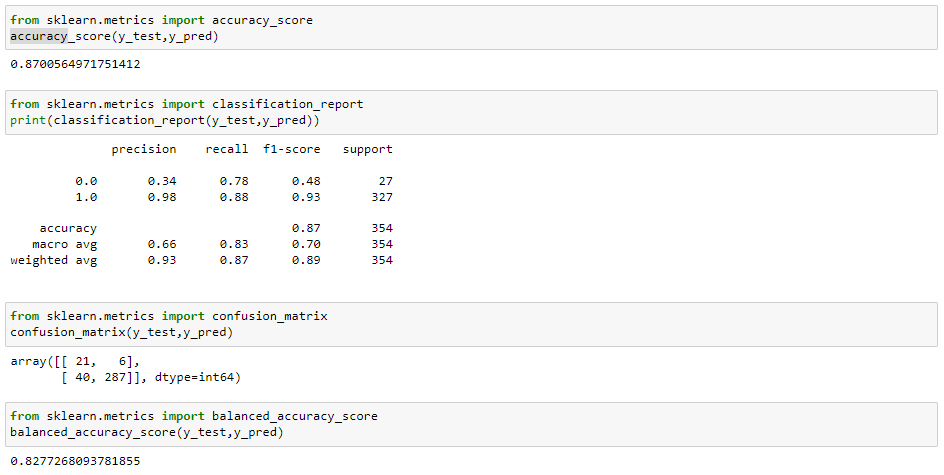
* **Training and Testing the model**



Here I will be calling the SVM model for training and evaluation of the model.

Then the value will be predicted.

* **Accuracy testing**



At last I will be calculating the accuracy. Hence, A good accuracy results a good mode

**1.5 Conclusion :**

We have created a model which can predict the rating from the reviews by using SVM algorithm. We have also learn how to used the Linear kernel while using the SVM algo. This model is used for recommending the product.

## APPENDIX

**Find the code in the given link:**

[**https://drive.google.com/file/d/13X0eYx28bspooBss82faUb7m6HiqqcEZ/view?usp=sharing**](https://drive.google.com/file/d/13X0eYx28bspooBss82faUb7m6HiqqcEZ/view?usp=sharing)

## REFERENCE

* Machine Learning, Tom Mitchell, McGraw Hill, 1997
* Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow