**Sentiment Analysis of movie reviews.**

**Team SC :T 95.**

* **Name: Abdalla Osama Abdelrahman Abdel Hady.**

**ID: 20191700379.**

* **Name: Ahmed Osama Maher Hussein.**

**ID: 20191700012.**

* **Name: Adham Ahmed Saied Farag.**

**ID: 20191700083.**

* **Name: Sara Mohammed Saad Ahmed.**

**ID: 20191700278.**

* **Name: Manal Gerges Zaki Bakheet.**

**ID: 20191700660.**

**1-Data Preparation**

* We have Data consist from text documents:

1000 text in positive.

1000 text in Negative.

We read them by OS library to read all text in the file of pos & neg.

* We read each line in text then we combine all lines in one sentence and give it target

1 for positive ones.

0 for negative ones.

* Combine All positive in one Data frame and All negative in one Data frame and then combine both data frames in one data frame.
* The Data frame has 2 columns 1 for reviews and 1 for Targets.

**2-Data Preprocessing.**

1. We make tokenization for words.
2. Then we remove all punctuation marks in sentence.
3. We make all words in lower case.
4. We make Lemmatization to the words.
5. We concatenate the words to make sentence.

**3- Feature Extraction**

Use TF-IDF vectorizer to convert text data into a numerical feature vector.

**4-Data Split**

We split the data to train, validate and test.

Train = 70%.

Validate = 15%.

Test = 15%.

**5- Model Selection**

We choose 3 Models:

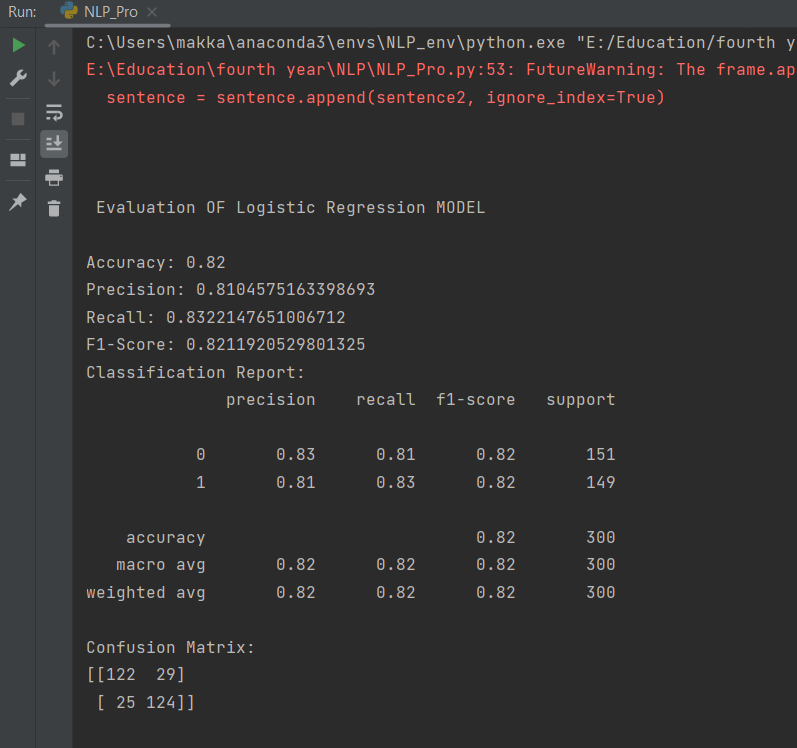
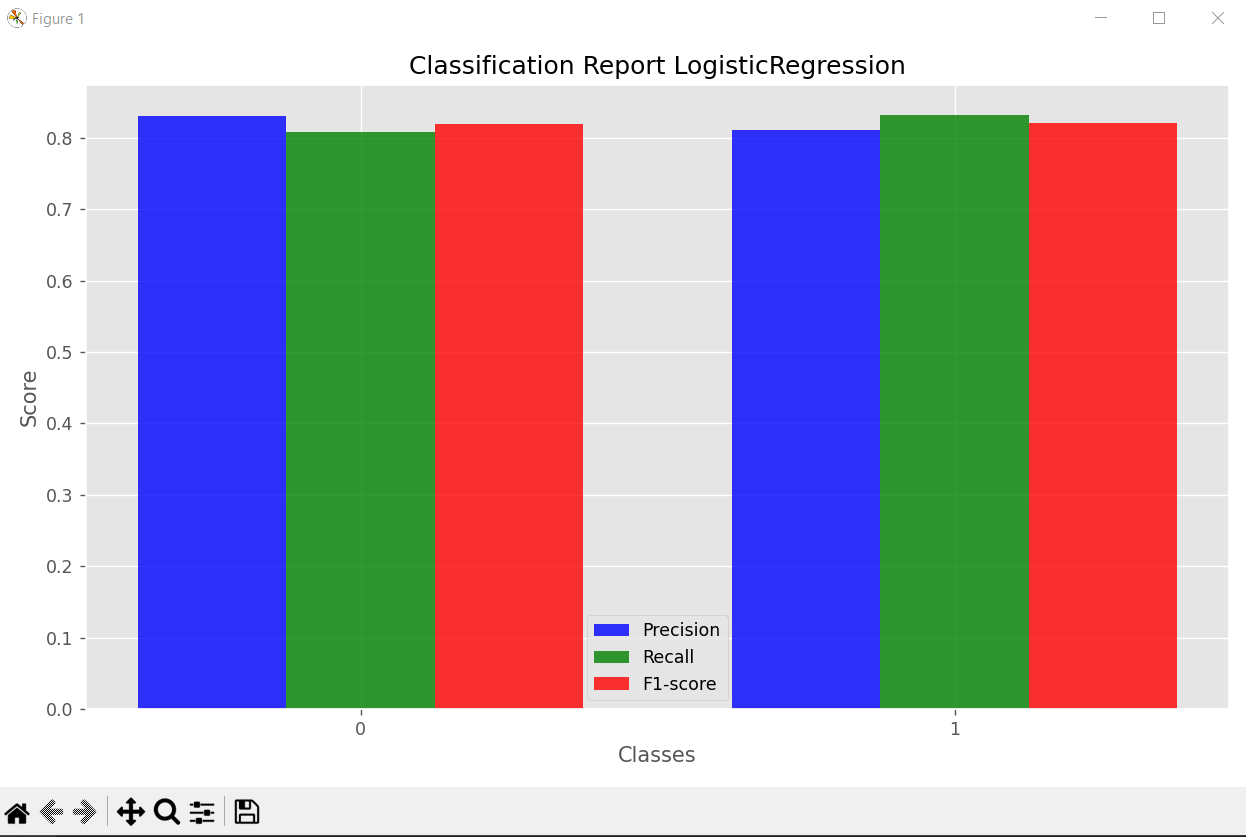
1-Logistic Regression.

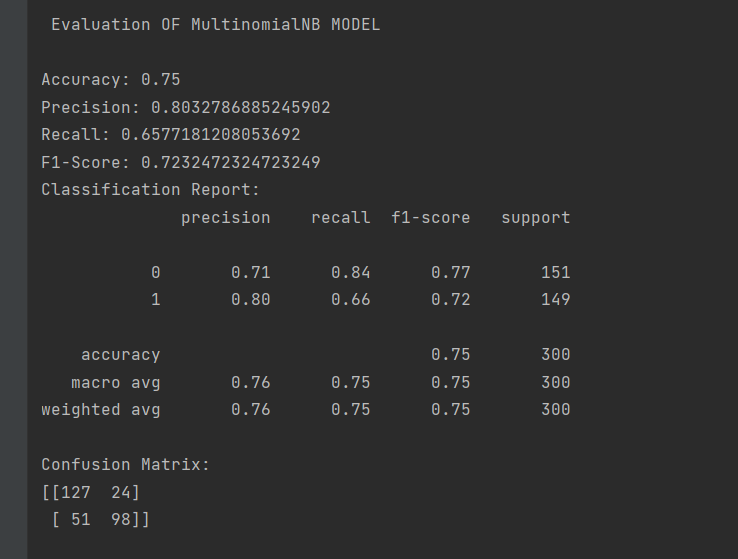
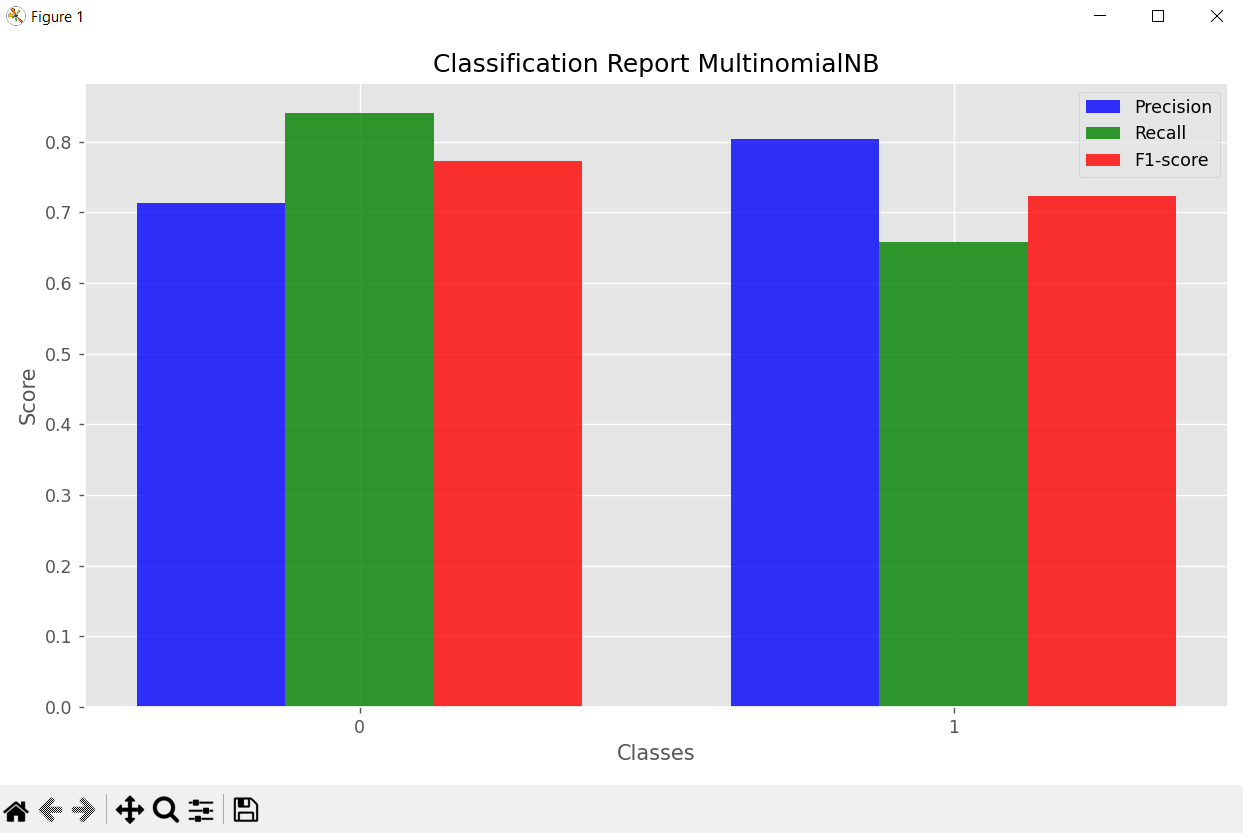
2-Multinomial NB.

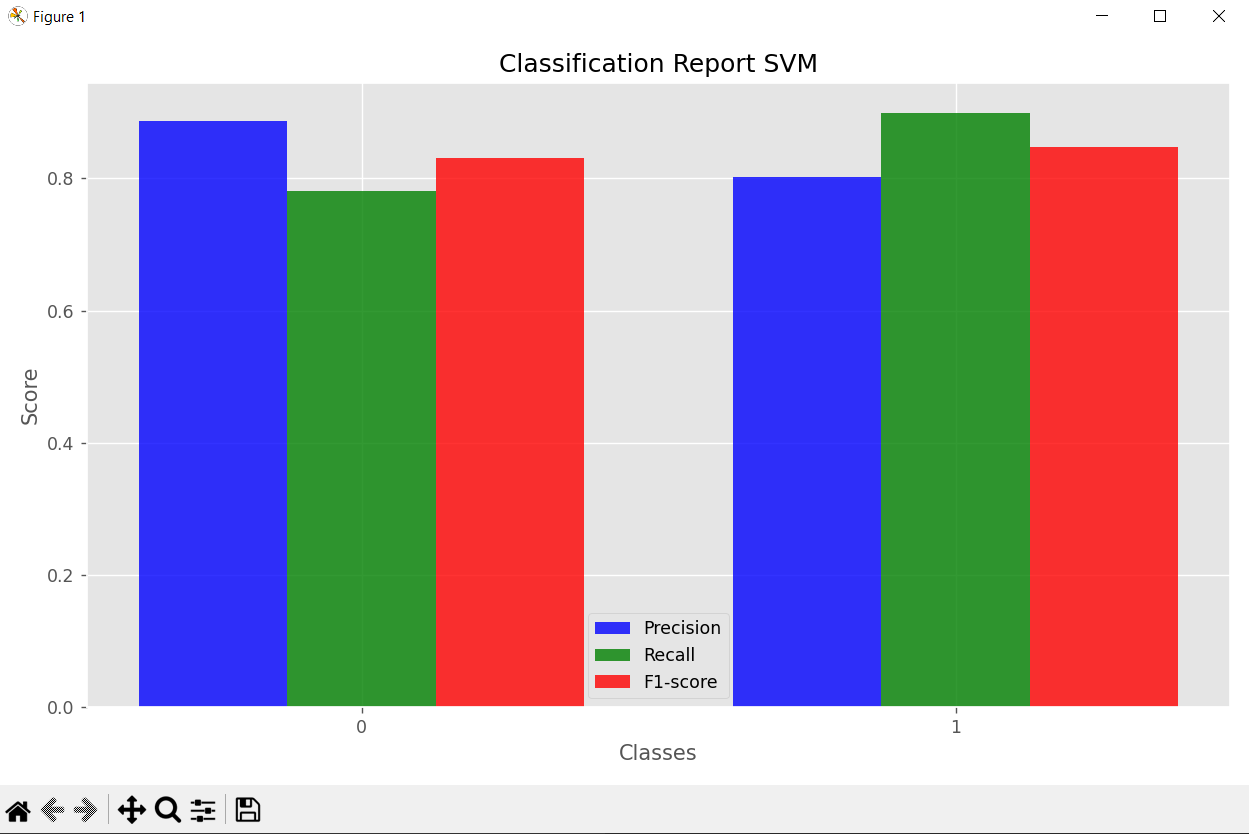
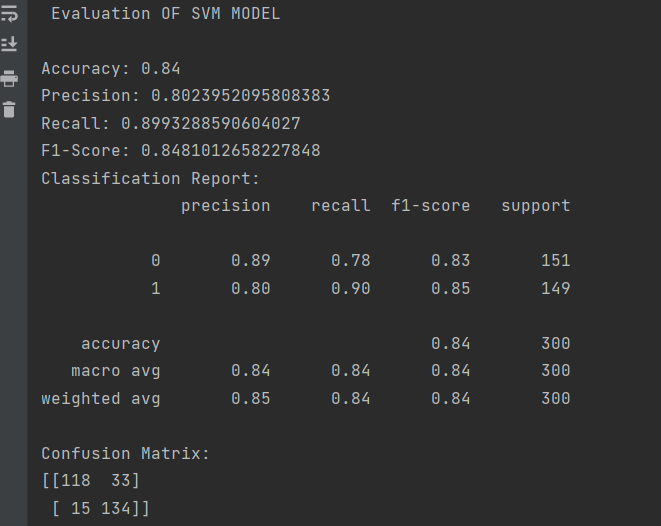
3-SVM Model.

Then we training the model by this three models and evaluate by the validate data and we get an good result from Logistic Regression And SVM.

But the best one is SVM.

The Logistic Regression Results. 

The results of Multinomial NB. 

The results of SVM. 

The Results of test by SVM model.