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.

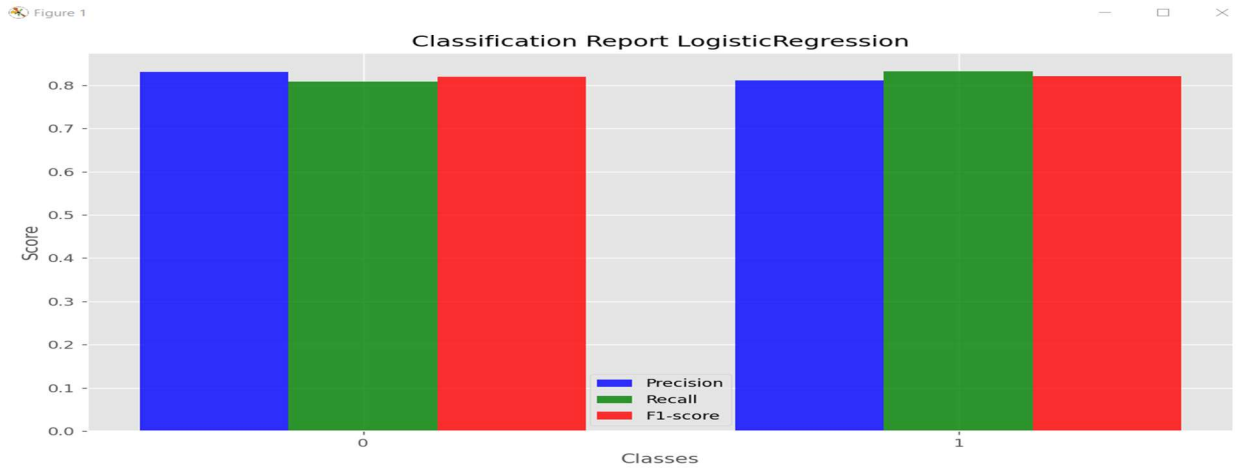
2-Multinomial NB.

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.

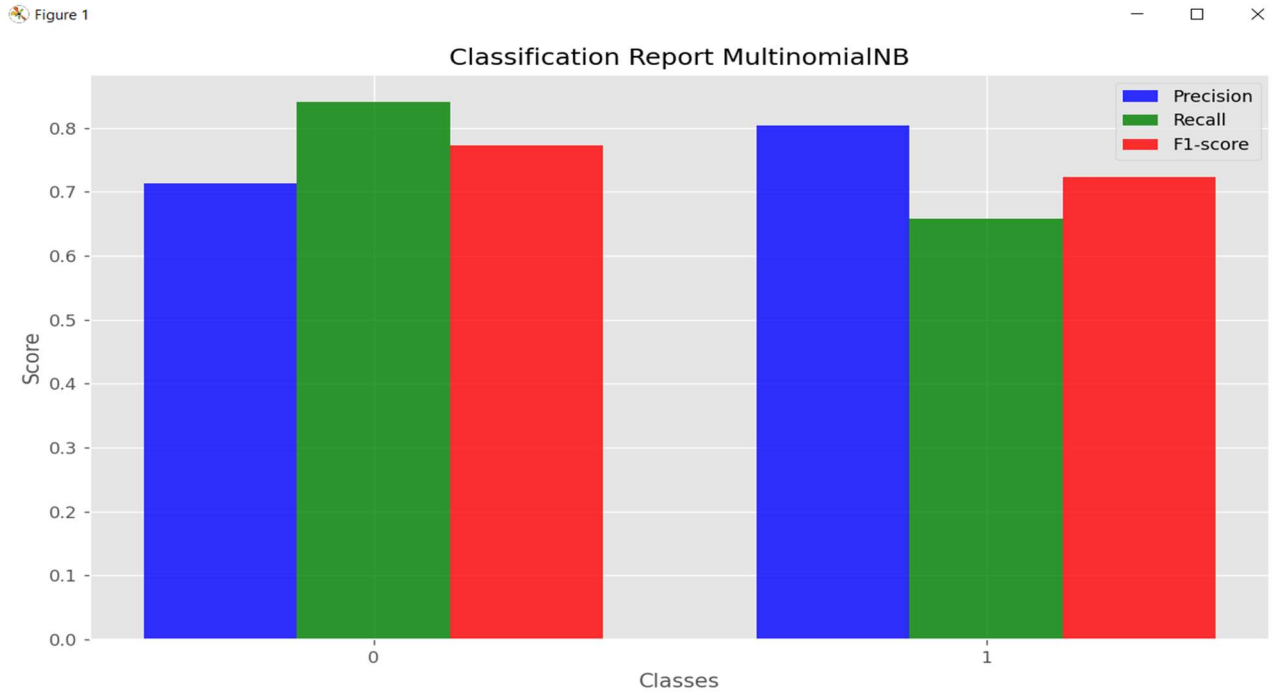
But the best one is SVM.

The Logistic Regression Results.




```
Run:  NLP_Pro
C:\Users\makka\anaconda3\envs\NLP_env\python.exe "E:/Education/fourth y
E:\Education\fourth year\NLP\NLP_Pro.py:53: FutureWarning: The frame.ap
  sentence = sentence.append(sentence2, ignore_index=True)



 Evaluation OF Logistic Regression MODEL

Accuracy: 0.82
Precision: 0.8104575163398693
Recall: 0.8322147651006712
F1-Score: 0.8211920529801325
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.81      0.82       151
           1       0.81      0.83      0.82       149

    accuracy                           0.82       300
   macro avg       0.82      0.82      0.82       300
weighted avg       0.82      0.82      0.82       300

Confusion Matrix:
[[122  29]
 [ 25 124]]
```

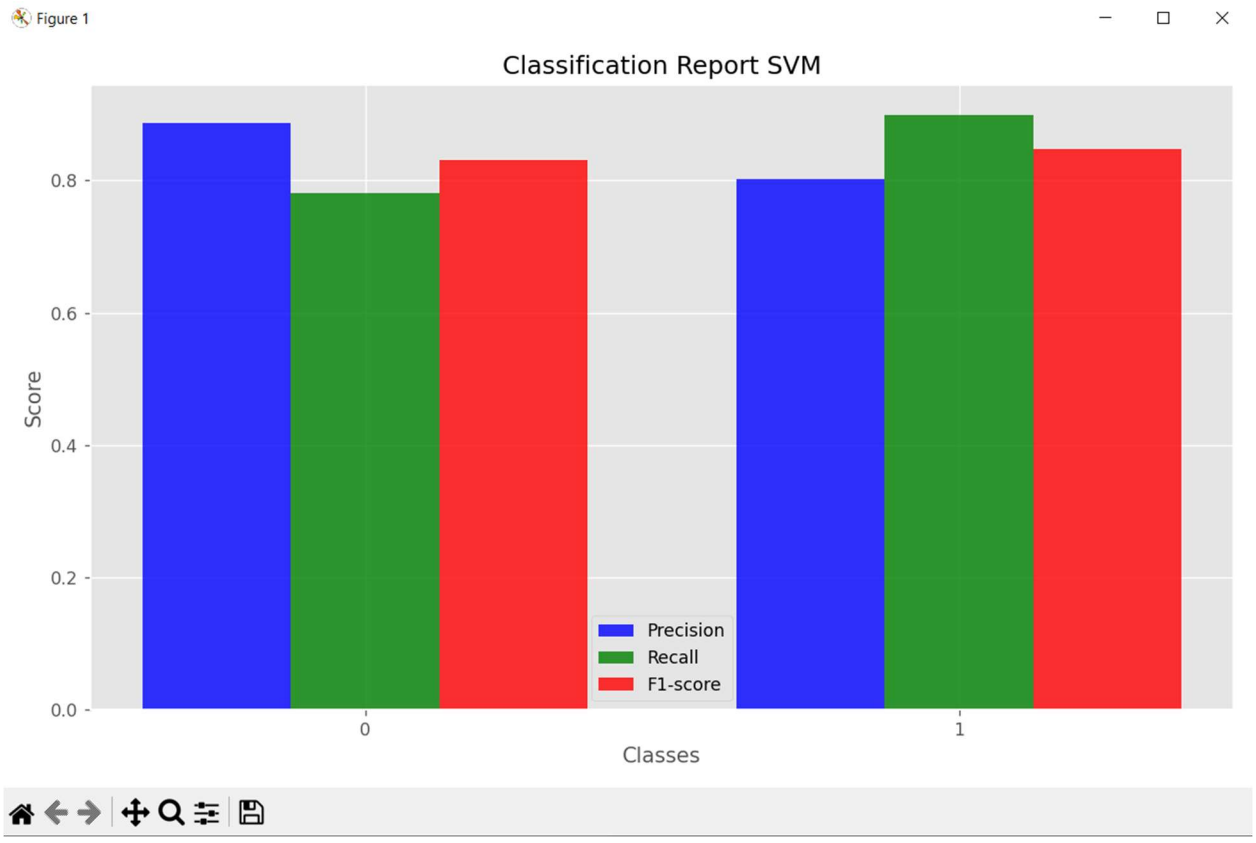
The results of Multinomial NB.



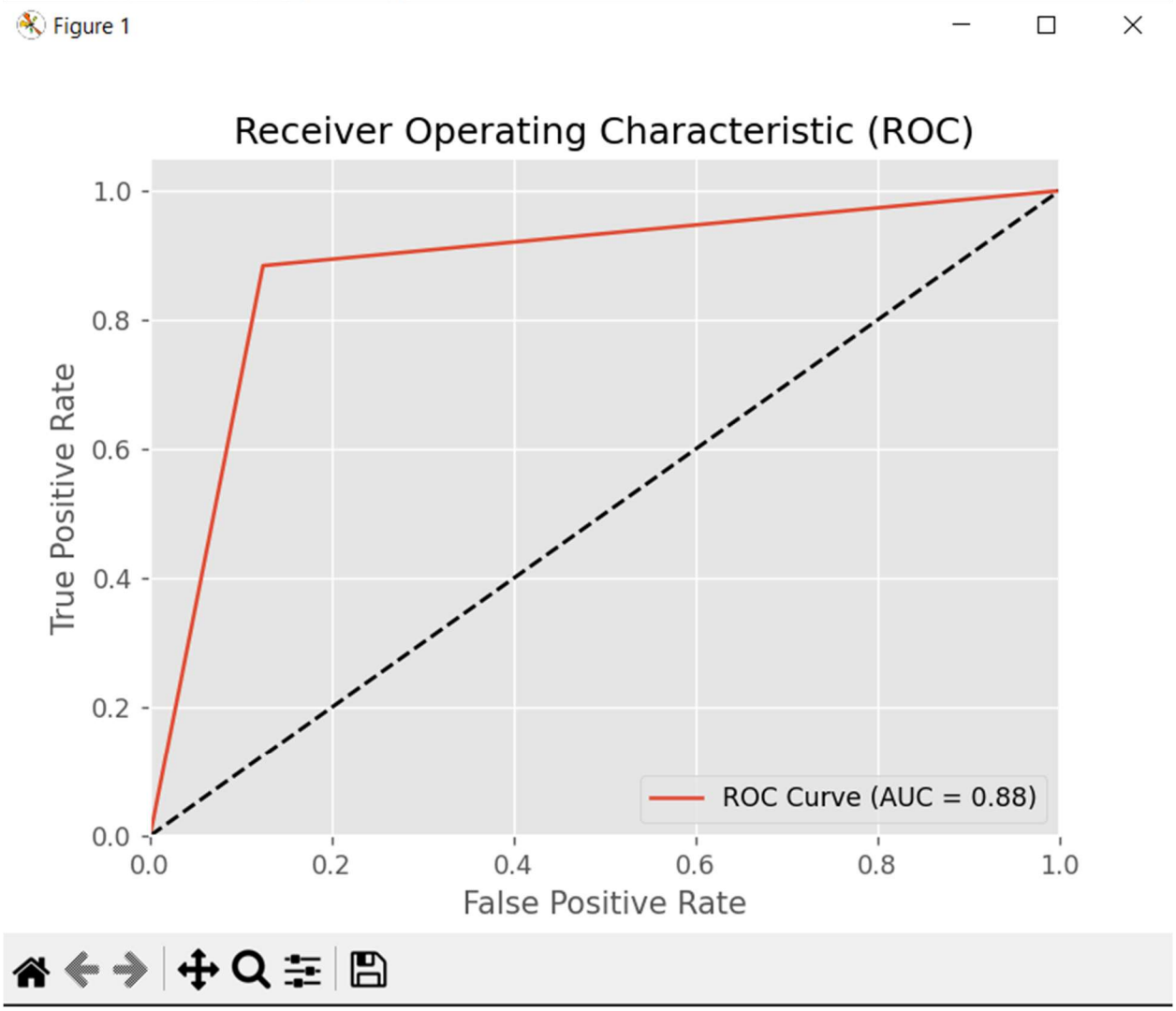

```
 Evaluation OF MultinomialNB MODEL

Accuracy: 0.75
Precision: 0.8032786885245902
Recall: 0.6577181208053692
F1-Score: 0.7232472324723249
Classification Report:
              precision    recall  f1-score   support

           0       0.71      0.84      0.77       151
           1       0.80      0.66      0.72       149

    accuracy                           0.75       300
   macro avg       0.76      0.75      0.75       300
weighted avg       0.76      0.75      0.75       300

Confusion Matrix:
[[127  24]
 [ 51  98]]
```

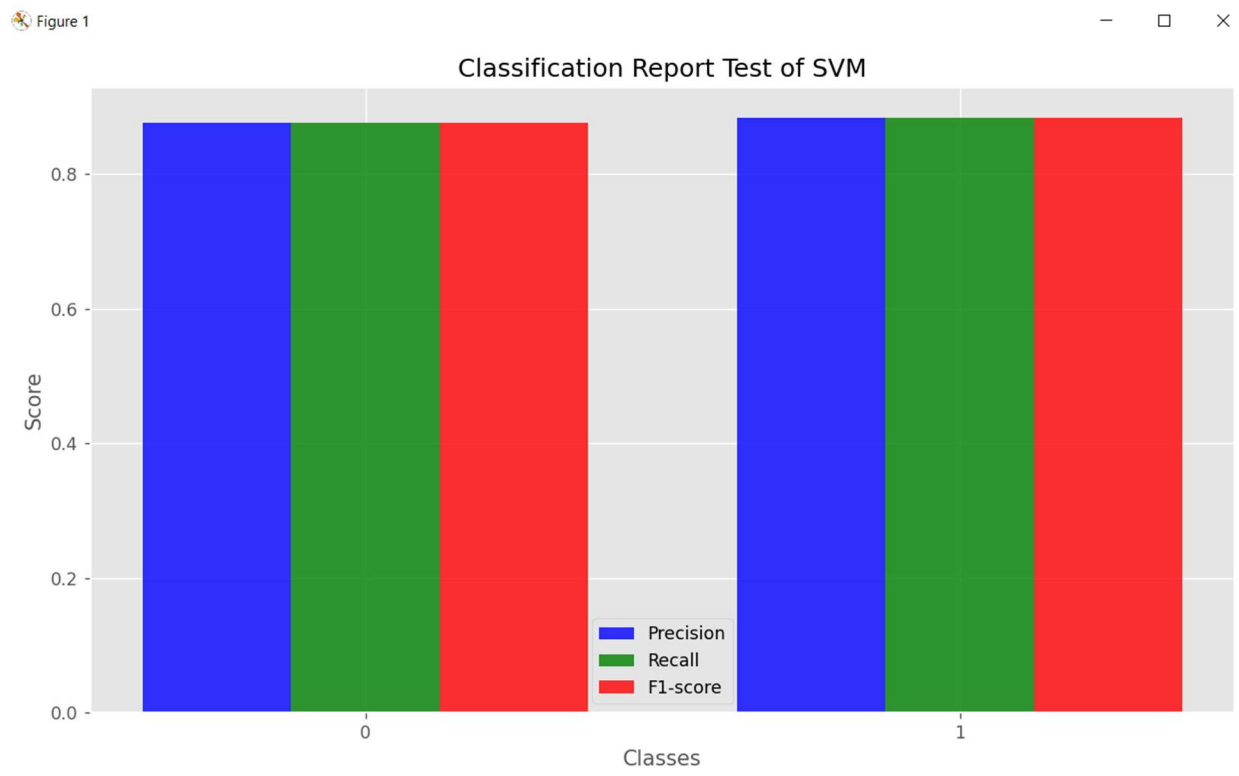
The results of SVM.

```
 Evaluation OF SVM MODEL

Accuracy: 0.84
Precision: 0.8023952095808383
Recall: 0.8993288590604027
F1-Score: 0.8481012658227848
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.78      0.83       151
           1       0.80      0.90      0.85       149

    accuracy                           0.84       300
   macro avg       0.84      0.84      0.84       300
weighted avg       0.85      0.84      0.84       300

Confusion Matrix:
[[118  33]
 [ 15 134]]
```



The Results of test by SVM model.






```
Accuracy: 0.88
Precision: 0.8838709677419355
Recall: 0.8838709677419355
F1-Score: 0.8838709677419355
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.88      0.88       145
           1       0.88      0.88      0.88       155

    accuracy                           0.88       300
   macro avg       0.88      0.88      0.88       300
weighted avg       0.88      0.88      0.88       300

AUC: 0.8798665183537262
```