

# Sentiment Analysis of Movie Reviews

ECE - 569A

Jonaki Medda (V00950300)  
Karthiga Thangavelu (V00925048)  
Mandeep Kaur (V00971821)

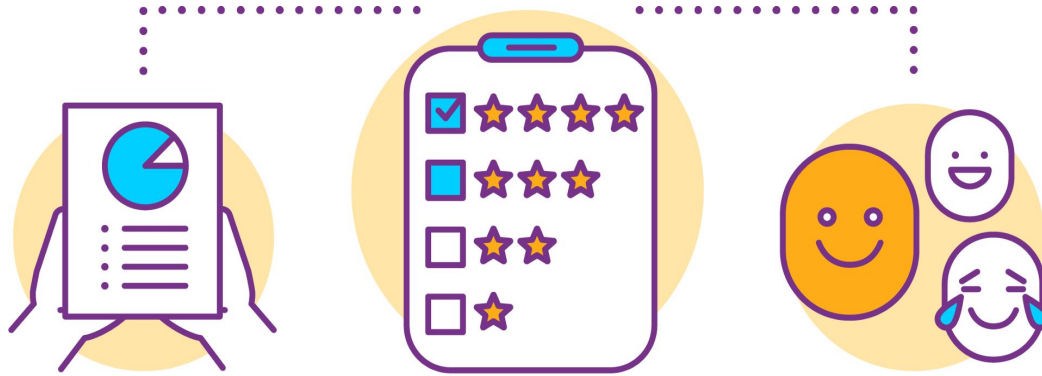
# Presentation Outline

1. Project Description
2. Motivation
3. Related Work
4. Data and Language Used
5. Data Preprocessing
6. Implementation
7. Evaluation Approach
8. Results
9. Limitation
10. Conclusions



# Project Description

- Building up a machine learning model by training it with a huge amount of data.
- Model will analyze the sentiments of the audience from their feedback.
- Apply the knowledge to determine which ones are popular.



# Motivation

Since the opinions and feedback of the public is a major concern, to provide some recommendations to the public to choose which movies to watch through an app.



# Related Work

S. No.	Year	Author	Dataset Used	Algorithm used
1	2021	Vihaan Nama	IMDB	Naive-Bayes
2	2020	A. Sheik Abdullah	Twitter API	SVM
3	2019	Mamtesh	IMDB	Naive Bayes
4	2018	V. Uma Ramya	Twitter API	Multinomial Naïve Bayes
5	2018	Palak Baid	IMDB	Naive-Bayes

# Dataset and Language Used

- IMDB dataset
  - 50000 movies
  - 70% for training and 30% for testing
  - For streamlit, we are using another IMDB dataset of 110 movies which we collected manually
- Language Used
  - Python
  - NLP Libraries



# Data Preprocessing

- Lower case
- Removing URL
- Removing Punctuations
- Removing Stop words
- Removing HTML tags
- Lemmatization
- Tokenization



# Implementation

## Word Embedding

- **Bag of Words:** It generates a collection of vectors representing the count of word occurrences in the text (reviews).
- **TF-IDF:** This technique uses numerical statistics to indicate the value of a certain term in a text and includes information on both the more significant and less important words.
- **Standard Scaling:** In order to compute the tf-idf representation, we have to compute the inverse document frequencies based on the training data and use these statistics to scale both the training and test data.





# Implementation

## Classifiers Used

- **Naïve Bayes:** It is easy & simple to build and is highly useful for classifying text in very large data sets.
- **Multinomial Naïve Bayes:** This model is available for Naïve Bayes classifier in scikit-learn library and is used in case of disjoint counts.
- **SVM:** The algorithm deploys the mechanism of non-linear mapping to transform the original observed training samples to a level of higher order dimensions.

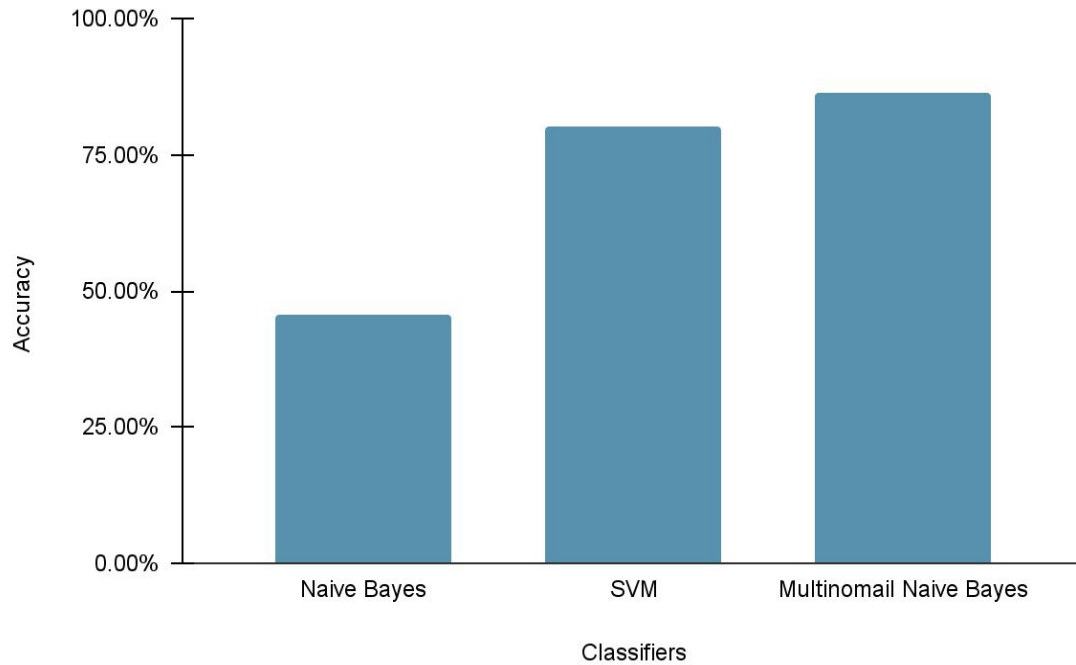


# Evaluation Approach

- **Accuracy:** The number of real findings among the entire number of instances investigated is the accuracy.
- **Precision:** The number of positive class predictions that really belong to the positive class is measured by precision.
- **F1 score:** It is a measure of a test's accuracy in binary classification statistical analysis
- **Recall:** The number of positive class predictions made out of all positive examples in the dataset is measured by recall
- **Confusion Matrix:** It is a table that shows how well a classification model performs on a set of test data for which the real values are known.



# Comparison of Different Approaches



## Results: SVM

```
0.8014
```

```
Accuracy score 0.801
```

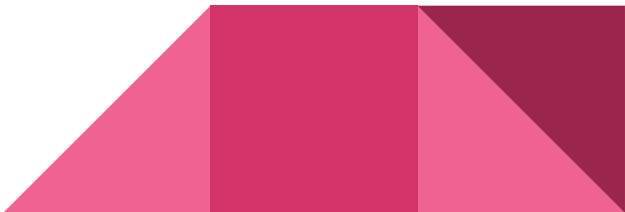
```
Precision: 0.7942715145173947
```

```
Recall: 0.8120069528011766
```

```
F1 score: 0.8013862073755122
```

```
[[5948 1573]
```

```
[1406 6073]]
```



## Results: Multinomial Naive Bayes

0.8642666666666666

Accuracy score 0.864

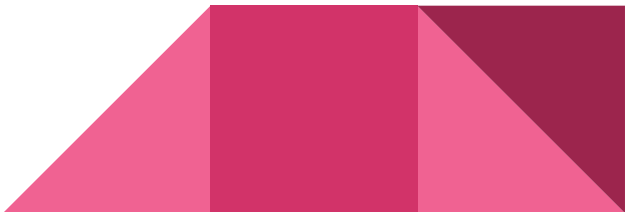
Precision: 0.8729614910237083

Recall: 0.8517181441369167

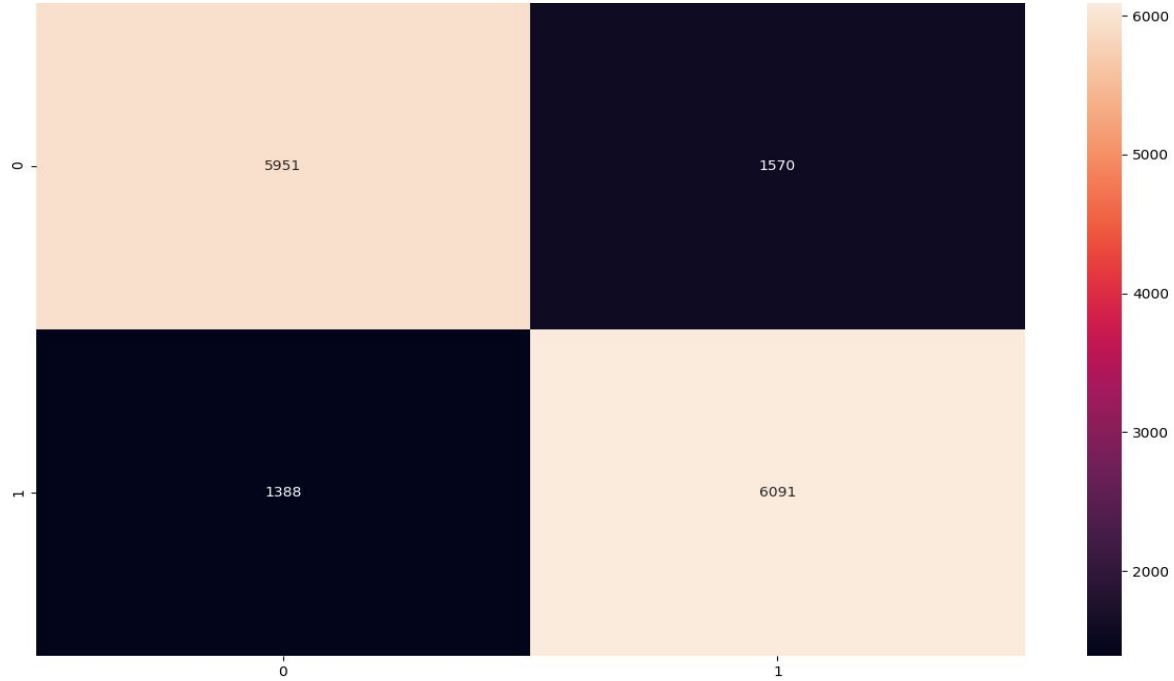
F1 score: 0.8642363907784164

[[6594 927]

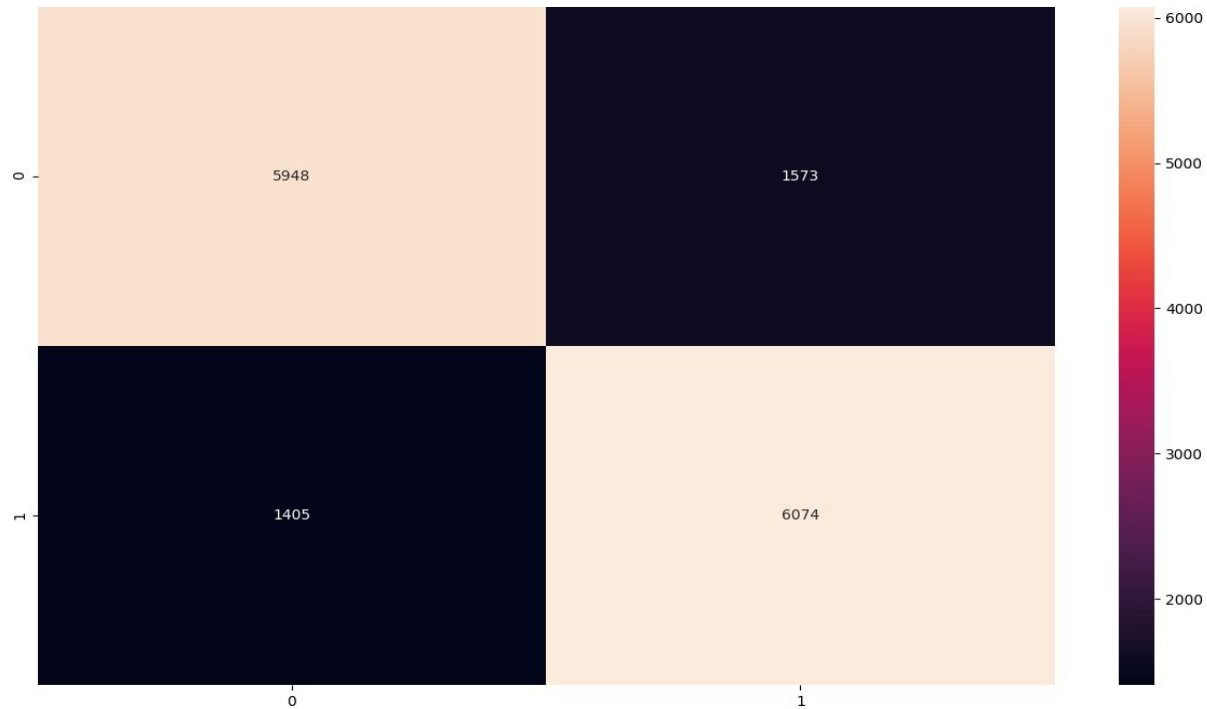
[1109 6370]]



# Heat Map of SVM



# Heat Map of Multinomial Naive Bayes







# Limitations

1. Dataset was not good enough because we could not able to get the movie names along with reviews. That's why we used only 100 movies data for the stream lit.
2. Accuracy for the Naive Bayes was only 45% because preprocessing was not adequate as compared to other algorithms.



# Conclusion

We conclude that Multinomial Naive Bayes Machine Learning algorithm produces better accuracy as compared to the SVM and Naive Bayes classifier algorithms

We successfully implemented a Streamlit where one can easily search for must watch movies as good movies or average movies as bad movies before watching.



# References

- [1] Vihaan Nama, Vinay Hegde, B. Satish Babu, “Sentiment Analysis of Movie Reviews: A Comparative Study between the Naive-Bayes Classifier and a Rule-based Approach”, ICITIIT, IEEE, 2021.
- [2] B Pang, L Lee- “Opinion mining and sentiment analysis” Foundations and Trends in Information Retrieval Vol. 2, No 1-2 (2008) 1–135
- [3] A. Sheik Abdullah, K. Akash, J. ShaminThres, S. Selvakumar, “Sentiment Analysis of Movie Reviews Using Support Vector Machine Classifier with Linear Kernel Function”, Evolution in Computational Intelligence, Springer, pp 345-354, 2020.
- [4] Rahul, Vasundhara Raj, “Online Reviews Over Sentiment Analysis using Machine Learning: A Systematic Review” JOURNAL OF CRITICAL REVIEWS, ISSN- 2394-5125 VOL 7, ISSUE 05, 2020
- [5] Mamtesh, Seema Mehla, “Sentiment Analysis of Movie Reviews using Machine Learning Classifiers”, International Journal of Computer Applications (0975 – 8887), Volume 182 – No. 50, April 2019.
- [6] Lakshmi Surekha P, Jayanthi A, “A Movie Review Sentiment Analysis Using Machine Learning Techniques”, International Journal of Recent Scientific Research, Vol. 10, Issue 8, pp. 34492-34497, 2019
- [7] Mais Yasen, Sara Tedmori, “Movies Reviews Sentiment Analysis and Classification” IEEE, 2019.
- [8] H. Swathi, S. S. Aravinth, V. Nivethitha, T. Saranya, R. Nivethanandhini, “Sentiment Analysis of Movie Review using data Analytics Techniques”, IRE Journals | Volume 2 Issue 9 | ISSN: 2456-8880, 2019.
- [9] Beatrice Lopez, Minh Anh Nguyen, Xavier Sumba, “IMDb Sentiment Analysis”, McGill Univ, CA, February 2019.
- [10] V.Uma Ramya, K. Thirupathi Rao, “Sentiment Analysis of Movie Review using Machine Learning Techniques”, International Journal of Engineering & Technology, Volume 7, pp 676-681, 2018.
- [11] Gurshabad Singh Brar, Ankit Sharma, “Sentiment Analysis of Movie Review Using Supervised Machine Learning Techniques”, International Journal of Applied Engineering Research, ISSN 0973-4562, Volume 13, Number 16, pp. 12788-12791, 2018.
- [12] Palak Baid, Apoorva Gupta, Neelam Chaplot, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques” International Journal of Computer Applications (0975 – 8887), Volume 179 – No.7, December 2017.

THANK YOU