Sentiment Analysis on Movie Review

Using ML Approach

Team: BinaryBeasts
Mohee, Rohith, and Sridhar
[2018112005, 2020201042, 2018111021]

TABLE OF CONTENTS

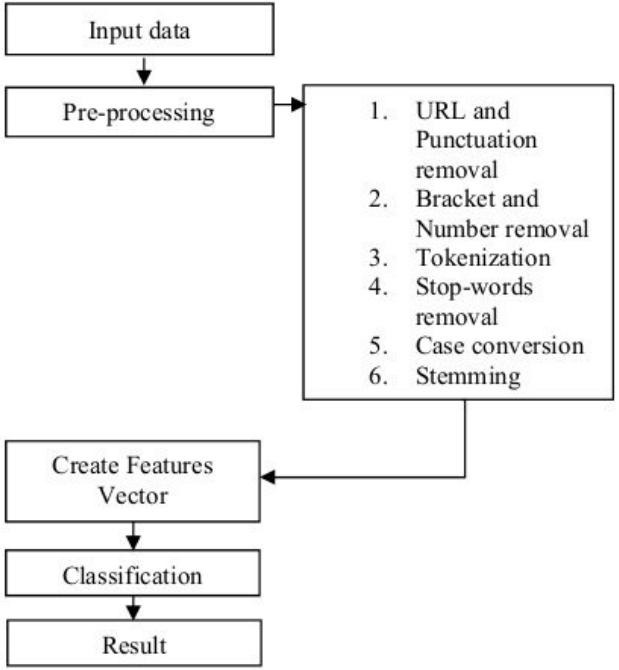
Introduction

Movies are one of the most popular form of entertainment for mankind and it is common for to people watch and express their opinions about it on the internet. Everyday huge amounts of data are generated of such opinions.

People often tend to check rating and reviews before watching the movie.

By analyzing movie review data we can learn about the strong and weak point of a movie and tell us if the movie meets the expectation of the user.

Sentiment analysis (SA) helps in analysing the review of that movie.



Scope of the research paper

The research paper aims to compare different machine learning classification techniques over a single feature extraction method

Feature extractor: TF-IDF

Classifiers: MNB, BNB, Maximum Entropy(LR), SVM, and Decision Tree

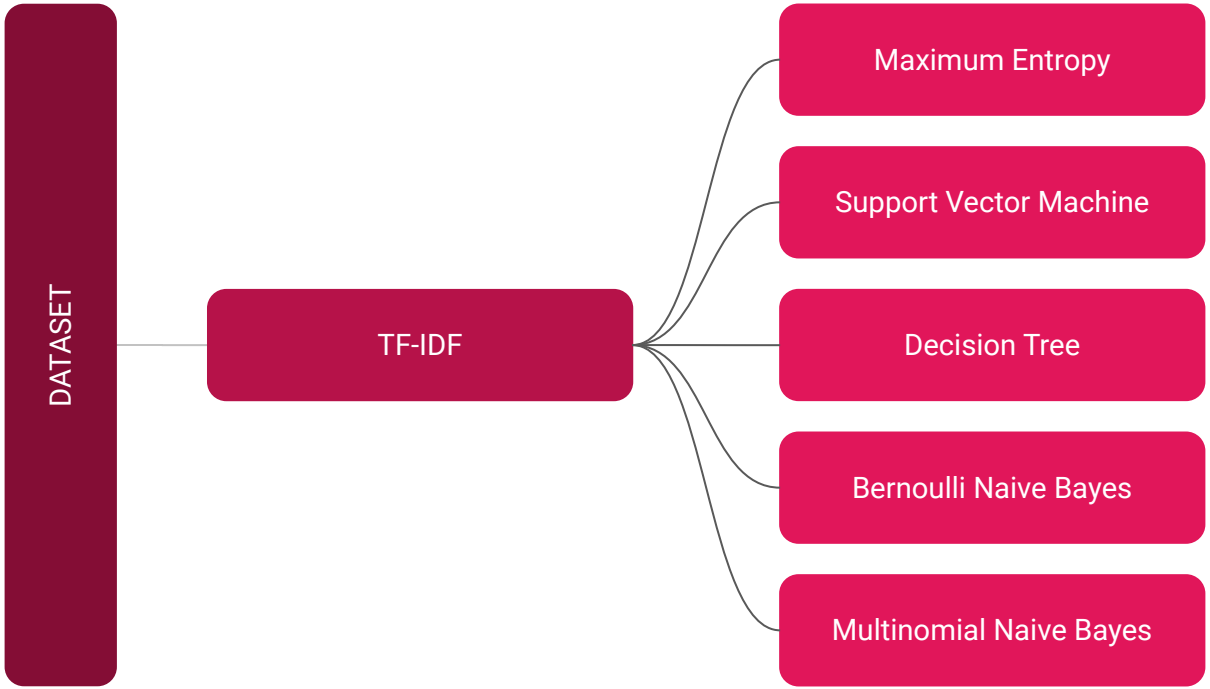
Increasing the scope of the project

We have compared some DL and ML classification techniques to classify the movie reviews for a set of feature extraction methods

Feature extractor: TF-IDF, GloVE, Word2Vec, Doc2Vec, and Bag of words

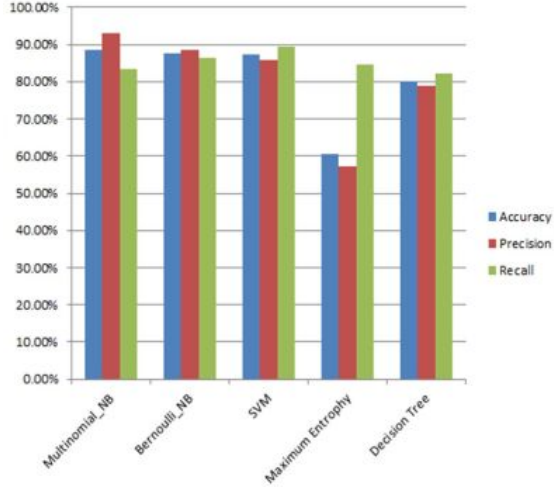
Classifiers: MNB, BNB, ME(LR), SVM, Decision Tree, CNN, and LSTM

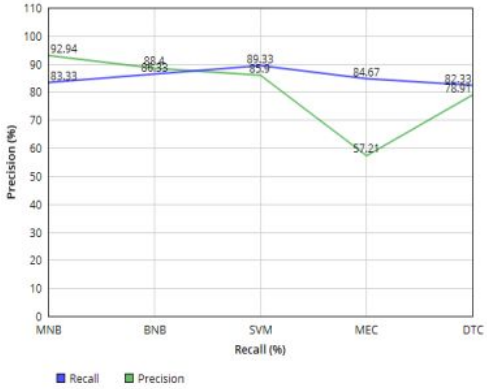
Scope of the Research Paper :-



Paper's Results

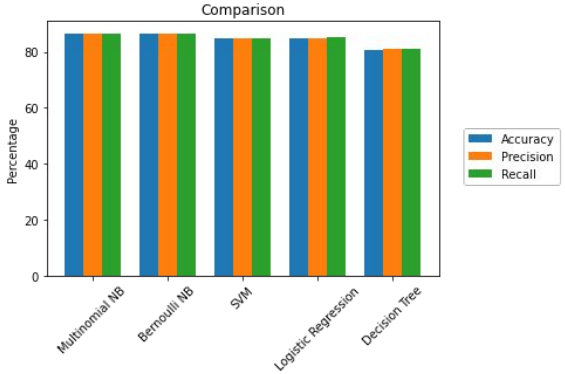
Method	Accuracy	Precision	Recall	F-score
Multinomial NB	88.50%	92.94%	83.33%	87.87%
Bernoulli NB	87.50%	88.40%	86.33%	87.35%
SVM	87.33%	85.90%	89.33%	87.58%
Maximum Entropy	60.67%	57.21%	84.67%	68.28%
Decision Tree	80.17%	78.91%	82.33%	80.58%

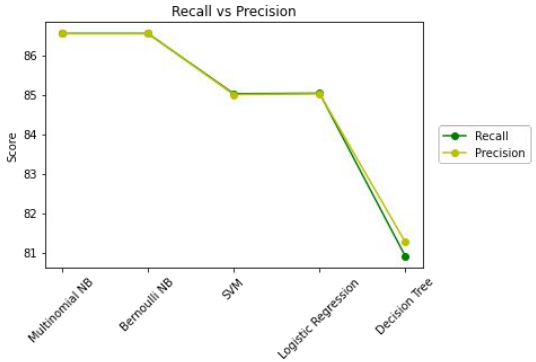




Our Results [link to dataset]

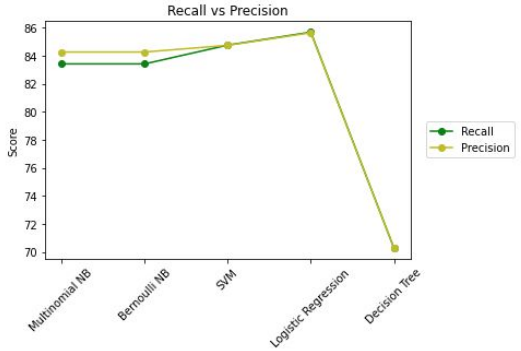
Model Name	Accuracy	Precision	Recall
Multinomial Naïve Bayes	85.5	85.5535	85.5535
Bernoulli Naïve Bayes	86.5	86.5541	86.5541
Support Vector Machine	85	85.0071	85.0281
Maximum Entropy	85	85.0266	85.0407
Decision Tree	81.75	82.2781	81.9074

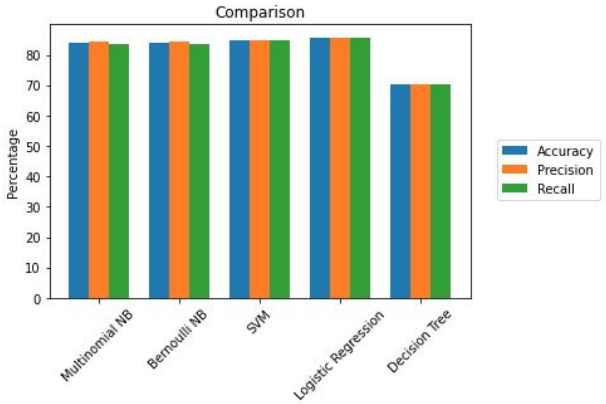




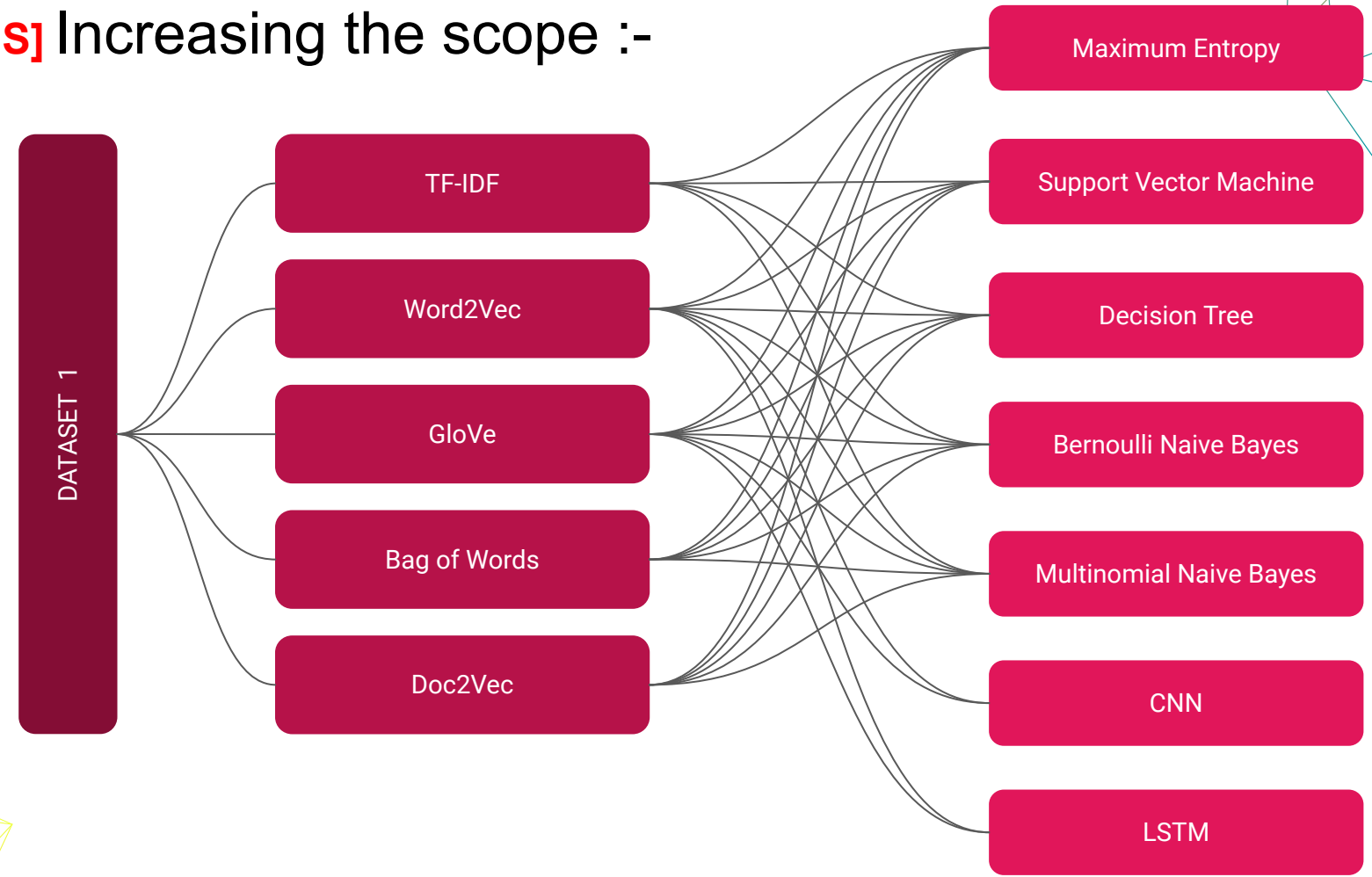
Results for another dataset [link to dataset]

Model Name	Accuracy	Precision	Recall
Multinomial Naïve Bayes	83.8	84.263	83.4163
Bernoulli Naïve Bayes	83.8	84.263	83.4163
Support Vector Machine	84.8	84.7409	84.7571
Maximum Entropy	85.7	85.6391	85.6784
Decision Tree	70.4	70.2911	70.281



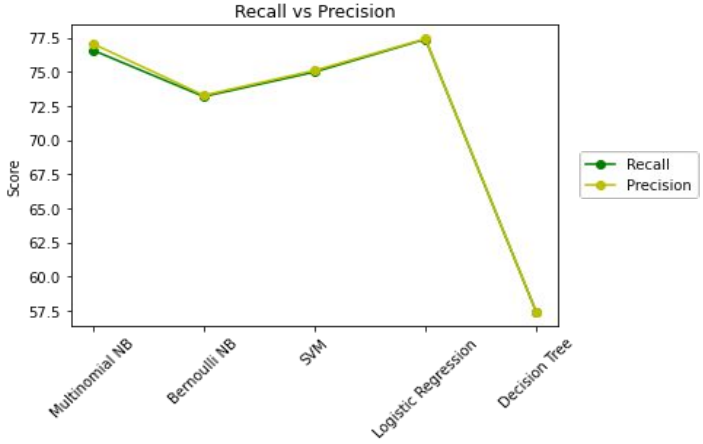


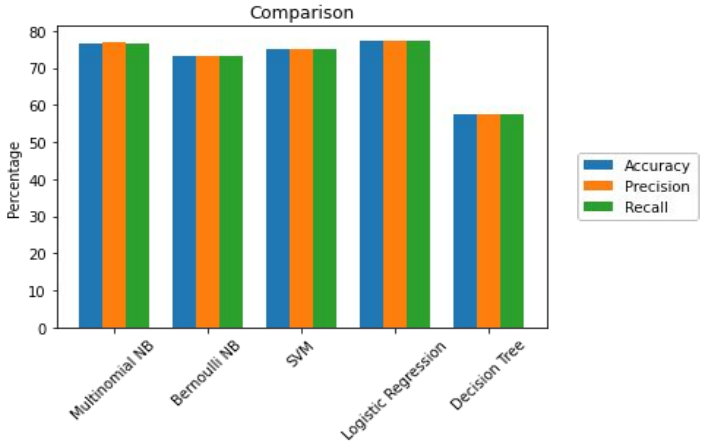
[BONUS] Increasing the scope :-



Doc2Vec

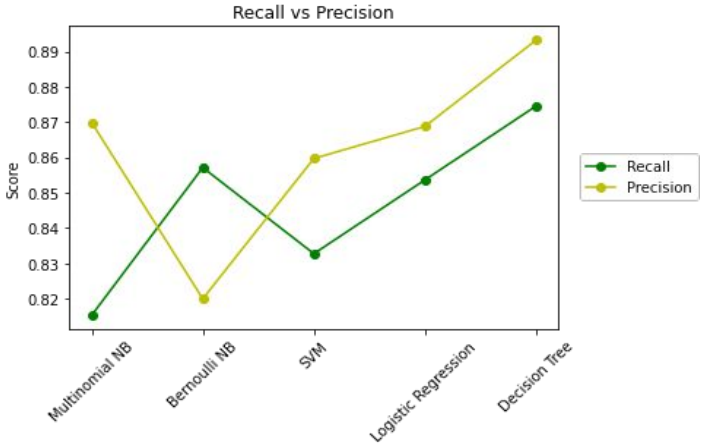
Model Name	Accuracy	Precision	Recall
Bernoulli Naïve Bayes	73.2	73.2954	73.2
Multinomial Naïve Bayes	76.6	77.0717	76.6
Support Vector Machine	75	75.1161	75
Maximum Entropy	77.4	77.4356	77.4
Decision Tree	57.4	57.4001	57.4

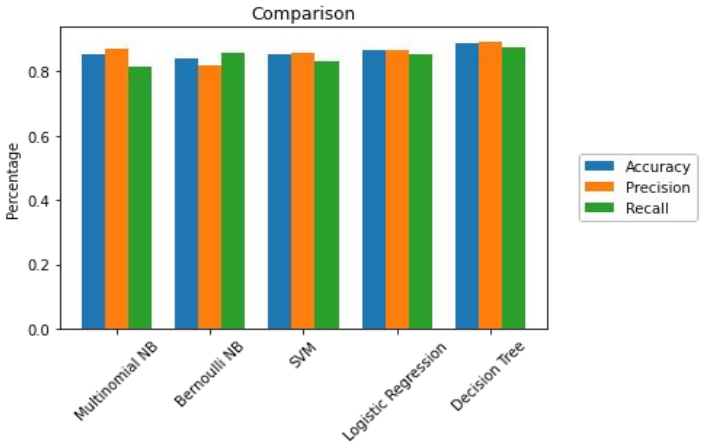




Bag of words

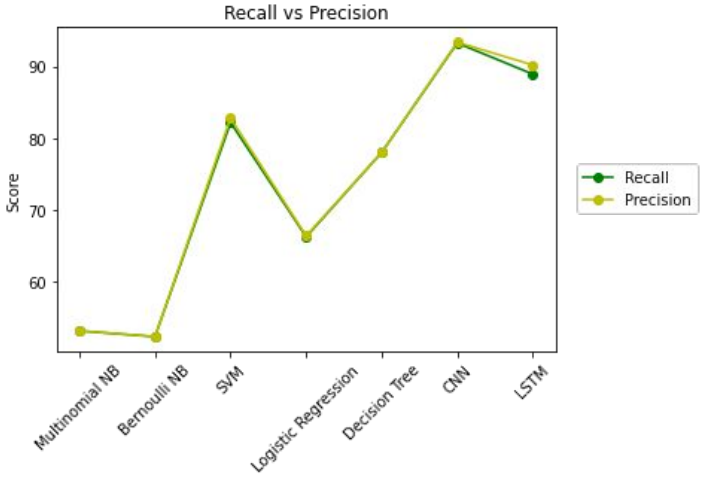
Model Name	Accuracy	Precision	Recall
Bernoulli Naïve Bayes	0.82	0.857143	0.841667
Multinomial Naïve Bayes	0.869888	0.815331	0.853333
Support Vector Machine	0.859712	0.832753	0.855
Maximum Entropy	0.868794	0.853659	0.868333
Decision Tree	0.893238	0.874564	0.89

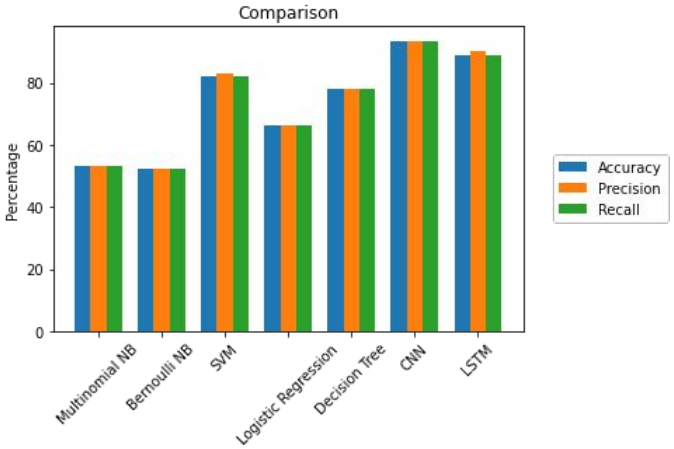




Word2Vec

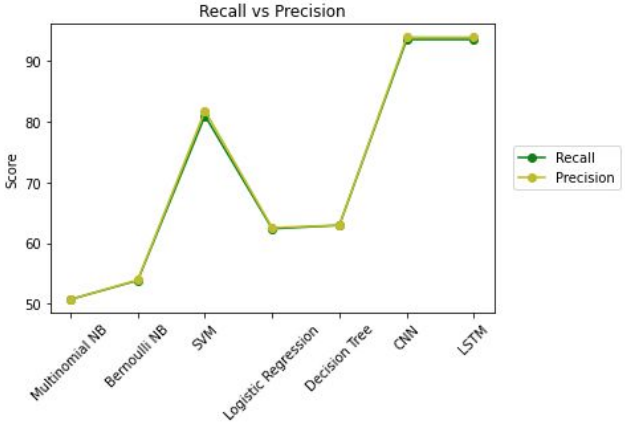
Model Name	Accuracy	Precision	Recall
Multinomial Naïve Bayes	53.3333	53.2323	53.2023
Bernoulli Naïve Bayes	52.3333	52.4113	52.4062
Support Vector Machine	82	82.8301	82.2096
Maximum Entropy	66.1667	66.4197	66.2983
Decision Tree	78	78.0228	78.0377
CNN	93.1667	93.3815	93.273
LSTM	88.6667	90.2095	88.9344

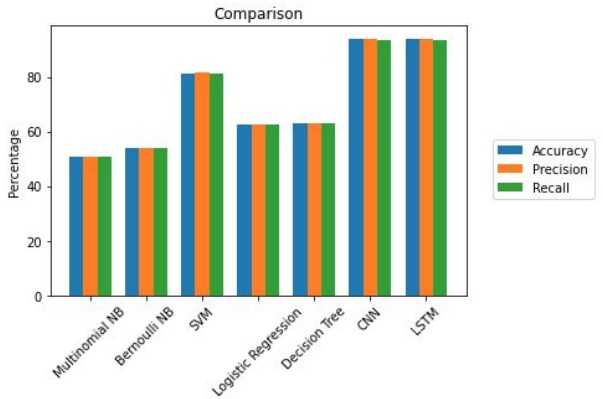




GloVe

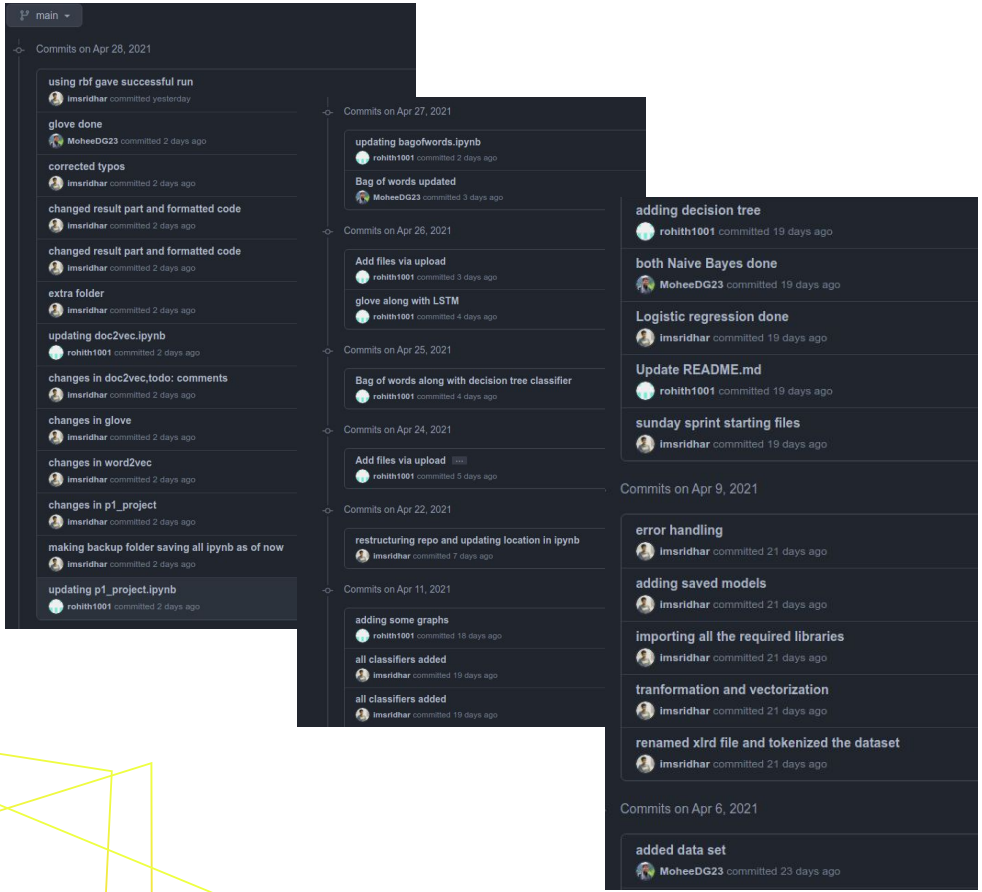
Model Name	Accuracy	Precision	Recall
Multinomial Naïve Bayes	50.6667	50.7366	50.7336
Bernoulli Naïve Bayes	54	53.9372	53.8615
Support Vector Machine	81.1667	81.8531	81.0157
Maximum Entropy	62.5	62.5542	62.3883
Decision Tree	63	62.9836	62.9518
CNN	93.6667	93.9562	93.5908
LSTM	93.6667	93.9562	93.5908





TEAM and WORK

[link to our GitHub repo]



Name	Contribution
Mohee	TF-IDF, Word2Vec,BoW, Data Preprocessing ,Slides and Documentation
Rohith	Glove, TF-IDF, Word2Vec,CNN,LSTM
Sridhar	CNN, LSTM , Classifiers, Slides and Documentation

References

Reference Papers

- A. Rahman and M. S. Hossen, "Sentiment Analysis on Movie Review Data Using Machine Learning Approach," 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), 2019, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201470.
- Yessenov, Kuat, and Saša Misailovic. "Sentiment analysis of movie review comments." *Methodology* 17 (2009): 1-7.
- T. P. Sahu and S. Ahuja, "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms," *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*, 2016, pp. 1-6, doi: 10.1109/MicroCom.2016.7522583.

Blogs Referred

- https://towardsdatascience.com/multiclass-classification-with-word-bags-and-word-sequences-4fffd4d62e0c
- https://www.kaggle.com/dormann/cnn-vs-lstm-movie-review-classification
- https://monkeylearn.com/sentiment-analysis/
- https://github.com/linanqiu/word2vec-sentiments/blob/master/word2vec-sentiment.ipynb
- https://medium.com/swlh/sentiment-classification-using-word-embeddings-word2vec-aedf28fbb8ca

THANK YOU!

Contact us at-
mohee.datta@research.iiit.ac.in
rohith.ramesh@students.iiit.ac.in
sridhar.m@research.iiit.ac.in