IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

Data Analysis using WORD2VEC to Suggest to Newbies

Swetha Ramana D V, Sai Karthik Reddy N S, Thanikanti Bhuvaneshwar, Sundeep Hiremat, Sharana Basava M
Department of Computer Science and Engineering
Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, Karnataka, India

Abstract: *Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviours. People post comments in social media mentioning their experience about an event and are also interested to know if the majority of other people had a positive or negative experience on the same event. Sentimental Analysis takes unstructured text comments, reviews about an event, etc., from all reviews posted by different users and classifies comments unto different categories as either positive or negative opinion. This is also known as polarity classification. This work aims at comparing the performance of different machine learning algorithms in performing sentimental analysis of Reviews.*

I. INTRODUCTION

There is a lot of research that proposed numerous techniques to analyze and classify the sentiment of the text, these techniques can be categorized into three approaches: supervised, unsupervised, and hybrid. The supervised approach is a machine learning approach involving the use of classification algorithms such as Naïve Bayes and support vector machines,maximum entropy, etc. The unsupervised approach includes various methods that are based on sentiment lexicons. The hybrid approach combines both machine learning and lexicon-based approaches.
The current research paper covers the analysis of the contents on the web covering lots of areas which are growing exponentially in numbers as well as in volumes as sites are dedicated to specific types of institutions

II. LITERATURE SURVEY

[1]SameerchandPudaruth, Sharmila Moheeputh ,NarmeenPermessur and AdeelahChamroo on " Sentiment Analysis from Facebook Comments using Automatic Coding in NVivo 11", This paper shows how QSR NVivo 11 can be used to extract and analyse posts and comments from a Facebook page. Our dataset is currently limited to only text data, that is, emoticons, images, audio files and videos were not taken into consideration. The auto code facility in NVivo 11 was used to tag the comments with the appropriate emotion. A comment can be tagged with both positive and negative sentiments. Stemmed words and synonyms are also used in the comparison process.
For the page 'Opposing Views', we have seen that the percentage of negative comments is more than twice the number of positive comments. Our aim in this paper is not to show that people tend to express more negative views than positive views but rather it should be considered more as a description of the use of sentiment analysis to extract human expressions on social media. Businesses can use sentiment analysis to understand the voice of the market, improve their brand management strategies, gain competitive advantage and develop new improved products. In the future, we intend to repeat the same experiment on several public pages on Facebook and then compare the results. We also intend to compare the NVivo auto coder with other sentiment classification tools.

[2] Aijith S Poriyath1, Amitha Joseph on "Sentiment Analysis on Facebook Comments ", the NLTK and the VADER analyser were applied to conduct a sentiment analysis. This approach is easy to use and time-consuming method. The results indicated that the VADER Sentiment Analyser was an effective choice for sentiment analysis classification

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

using Facebook Comments. VADER is quickly and easily classified huge amounts of data. However, the present approach has the following limitations.

First a general lexicon was used to categorise specific data. Third, the data were not trained. The world now accumulated with huge amount of data. Each day there is tremendous amount of data that is accumulated in the social media sites. The methods to find useful information from this data that help to speedy decision-making higher productivity in various fields. Facebook is Multilingual social media platform. This is one of biggest challenge to perform the sentiment analysis on Facebook comment. Sentiment analysis on Facebook data provide variety opportunities in business, politics and for the individuals. Sentiment Analysis can apply in any form of data collected from the social media not only the comments. You can apply in your status set, chat, etc.

Apply sentimental analysis on the chat will also help to understand the mentality of each chat group. This will be helpful when social media that will uses as a e-learning platform VADER classifies the sentiments very well. It is always available and easy to use, ready-made model which can be used across multiple sectors such as, social-media texts, analysing reviews etc. Advantage of using VADER is that it does not require any training data. Well, we can see that the results obtained are very much accurate.

III. METHODOLOGY

Data Clean-Up: Data extracted from Reviews comes out with a lot of data the name of the person or entity making the comment, For the purpose of this study, only the actual texts were used. Even this field contained some extra characters that had to be removed.

Vectorization: The general process of turning a collection of tokens into numerical feature vectors.

Feature Extraction: Text data demands a special measure before you train the model, Words after vectorization are encoded as integers or floating-point values for feeding input to machine learning algorithm. This practice is described as feature extraction. Scikit-learn library offers TF-IDF vectorizer to convert text to word frequency vectors.

Fitting Data to Classifier and Predicting Test Data: Train data is fitted to a suitable classifier upon feature extraction, then once the classifier is trained enough then we predict the results of the test data using the classifier, then compare the original value to the value returned by the classifier.

Result Analysis: Here the accuracy of different classifiers are shown among which the best classifier with highest accuracy percent is the chosen. Some factors such as f-score, mean, variance etc., also accounts for consideration of the classifiers

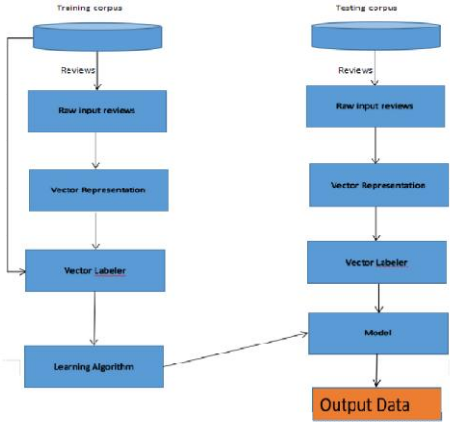
IV. PROPOSED SYSTEM

Sentiments are the words or sentences that represent view or opinion that is held or expressed that can be Positive, Negative, or neutral. We are going to propose a novel hybrid approach involving both corpus-based and dictionary-based techniques, which will find the semantic orientation of the sentiments words in reviews. We also consider features like emotions, neutralization, negation handling and capitalization as they have recently become a huge part of the internet language. The proposed Sentiment analysis on reviews is based on two important parts via data extraction, pre-processing of extracted data and classification.

- Dataset is taken and loaded.
- The data is pre-processed to clean the data and understand the dataset.
- For each review we apply NPL techniques like word vectorization.
- Data is spitted as train and test data.
- Model is built using ML algorithms like Logistic regression, Random Forest, SVM and NLP techniques like Word2vec embeddings.
- The model is trained with the pre-processed data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

- The model is tested and accuracy is calculated for different ML algorithms.
- The algorithm with best accuracy is finalized and that model will predict whether the review is positive or negative.



System Architecture

V. RESULTS

EVALUATION OF MODELS

Random Forest (RF): The classification report for the RF model. In this case, the total F1-score obtained is 94.7 %.

Naïve bayes: The classification report for the Naïve Bayes model. In this case, the total F1-score obtained is 92 %.

Neural Network: The classification report for the Neural Network model. In this case, the total F1-score obtained is 89%.

KNN: The classification report for the KNN model. In this case, the total F1-score obtained is 84 %.

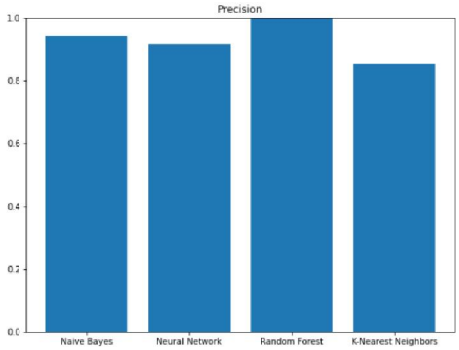


Fig 1: Accuracy of Algorithms

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

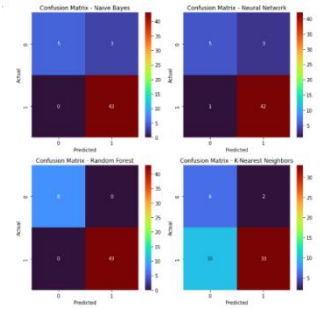


Fig 2: Confusion Matrix of Algorithms

```
Random Forest Classifier Accuracy:  0.9473684210526315
Confusion Matrix:
[[ 1  0]
 [ 1 17]]
Classification Report:
              precision    recall  f1-score   support

           0       0.50      1.00      0.67         1
           1       1.00      0.94      0.97        18

    accuracy                           0.95        19
   macro avg       0.75      0.97      0.82        19
weighted avg       0.97      0.95      0.96        19
```

Fig 3: Accuracy & Classification Report of Random Forest

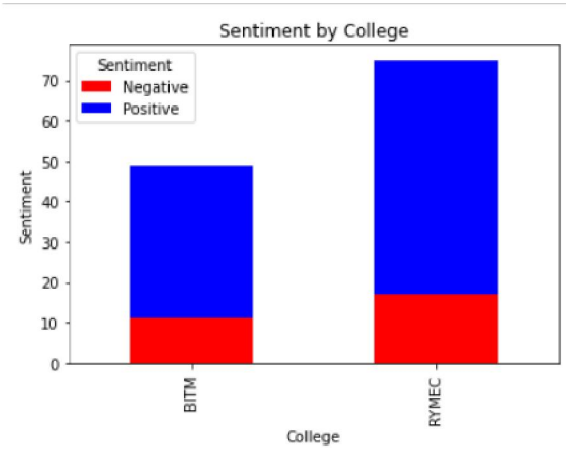


Fig 4: Final Result

VI. CONCLUSION

Summary

Institutional recommendation system can be especially useful for new joiners who are willing to pursue the course they want. Overall, institutional recommendation systems for new joiners can offer many potential benefits, including increased efficiency, reduced risk, and improved investment outcomes. These systems can help new joiners build a well-diversified portfolio that aligns with their investment goals and risk tolerance, while also providing guidance on how to rebalance their portfolios over time.

We furnished results for Sentiment and Emotional Analysis on Institution Reviews. On applying Logistic regression, Random Forest, Naïve Bayes, Random Forest stands out with 94% accuracy.In conclusion, institutional recommendation systems can be a valuable tool for new joiners in the industry, as they can help them make more informed decisions about their knowledge and build a well-diversified portfolio.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 6, May 2023

Future Scope

In future work, we aim to handle Institutional Recommendation for major of our Institutions present in our State, here this kind of a system is in the developing phase and lots of future enhancements are planned and are undergoing. This application can be expanded with many new other schemes and areas.

REFERENCES

[1] Yallanki Vikas, K. Nagendra chary on Sentiment Analysis on 'Facebook Comments International Conference on Intellectual Property Rights' (Feb 2021).
[2] The Use of Word2vec Model in Sentiment Analysis, International Conference on Artificial Intelligence, Robotics and Control (December 2019).
[3] SameerchandPudaruth, Sharmila Moheeputh, NarmeenPermessur and AdeelahChamroo on 'Sentiment Analysis from Facebook Comments using Automatic Coding in NVivo 11'(2019) Conference on Distributed Computing and Artificial Intelligence.
[4] Sentiment Analysis using Word2vec-CNN-BiLSTM Classification International Conference on Social Networks Analysis (December 2020).
[5] Ravichandran, M., Kulanthaivel, G. and Chellatamilan, T., (2015). On 'Intelligent Topical Sentiment Analysis for the Classification of E-Learners and their Topics of Interest'The Scientific World Journal.
[6] Omkar Borade, Kaushik Gosavi, Ajay Shinde, Avinash Gowda 'Sentiment Analysis of College Reviews'(2018)
[7] Shiv Naresh Shivhare and Prof. Saritha Khetawat on 'Emotion Detection from Text' Azad National Institute of Technology(2017).