Sophisticated Machine Learning Methods to Accomplish Analogous Customer Behavioural Analysis

Md Jaber Ali Taha [1] Subramanian K.M [2] Muntaj Begum Shaik [3]

[1]Research Scholar, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

[2]Professor, Department of Computer Science and Engineering, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

[3]Assistant Professor, Department of Information Technology, Shadan College of Engineering and Technology, Hyderabad, Telangana, India – 500086.

-- ***--

Abstract:- Data is growing in leaps and bounds day by day with the rapid growth in cloud and internet technologies. Social media has become an integrated part of life as it is used to share content, opinions, reviews, suggestions etc. Anybody can voice out their opinions through social media. This data can be reused or analyzed to understand the customer and cater to his needs. In particular, e-Commerce platform enable the users to write product reviews to help other buyers purchase better products and be benefited from their opinions. The e-Commerce platform can also mine the data for sentiment analysis to understand the customer opinion and build recommendation systems to understand the customer needs and suggest products. Recommendation systems give a personalized shopping experience to users and there by increases sales and profits. In this project we use n-gram analysis to understand the data in the dataset and then apply multiple machine learning algorithms to the selected dataset to identify the appropriate classification algorithm that gives the best accuracy. We also try to build a recommender system that uses collaborative filtering to recommend appropriate products as per the customer needs.

Keywords:- *Data Visualization, Sentiment Analysis, Machine Learning, Natural Language Processing*

---***---

I INTRODUCTION

Data is growing at a rapid rate due to social networking. Many users tend to voice their opinions on multiple topics and events in the form of reviews, blogs, tweets, posts etc. E-Commerce platforms enable their customers to post reviews about the performance and quality are various products that are being sold on the website. When customers give feedback about the products, it will help other customers make purchasing decisions. E-Commerce platforms and also take appropriate actions to maintain the quality of products that are being sold on their website. By analyzing the multiple reviews that are given to a product, the e-Commerce platform can find out the overall sentiment of the customers associated with that product. The e-Commerce platform can thus make a decision to keep a product are not based on the user sentiment. This will help to maintain the goodwill, quality of products which in turn improves the business of the e-Commerce platform. However, the reviews associated with the products might be thousand in numbers and it is not possible to manually find the overall user sentiment as the system need to understand human language. Machine learning techniques and natural language processing techniques help in achieving this tedious task. In this project, we try to build a web application and a machine learning model

that can accurately predict the sentiment from the review and a recommender system that can recommend suitable products to the customers.

Aim of the Project

We aim to extract the sentiment of the user for a particular product from the user review and get recommendations for users based on their purchases on the e-commerce platforms. The dataset used for this purpose is Amazon Reviews dataset and it is given to the developed machine learning model for training.

Scope of the Project

In this project, the accuracy of the developed machine learning model is calculated and compared with other machine learning algorithms and list of recommended products based on user processes is prepared. The admin of the application uploads and analyzes the data set and compare set with multiple machine learning algorithms. The process of maintaining users and their purchases on the e-Commerce platform does not fall under the purview of this project.

II LITERATURE SURVEY

"Consumer Behavior Modeling Based on Social Psychology and Complex Networks"

Bringing novelty in the services that an organization offers requires the clients to accept the changes. It is difficult to assess if a client would accept changes are not. This paper aims to bring about a new customer behaviour framework which contains agents in a consumer network. This framework uses snowball sampling techniques. The remarkable of differentiating technologies that are used in this paper are social psychology and versatile networks.

"Modeling prediction in recommender systems using restricted Boltzmann machine"

Combined or collaborative filtering is a powerful technique that forms the core of recommender systems which in turn used in marketing of products to the internet users. This paper elaborates on controlled Boltzmann machine for combined filtering and suggests neighborhood and condition based model based on similarity and relevance scores. The evaluation of this model has been done based on hidden units rate of information and initializing functions. When a data set that has 22 million rows has been fed to the proposed model, it has achieved an accuracy of 78.5% in predicting user priorities in the recommended commercials.

III SYSTEM ANALYSIS

.
Existing system:

Now a days, if an organization has to increase its profits, it has to understand the user intentions while trying to search or buy a product on the e-commerce websites. By understanding the user intentions and suggesting products, the organizations will be able to increase their sales and profits. However, in the existing scenario, the e-commerce websites are unable to accurately predict the user requirements and unable to collate the user opinions about products based on reviews. We need to make use of sophisticated machine learning procedures like as Natural Language Processing (NLP)

Disadvantages:
- Not accurate
- Not capable of gathering consolidated user opinion or suggest products.

Proposed System:

In this project, we propose to develop an application that has the capabilities to suggest products based on consumer opinions and interests. We utilize the sophisticated machine learning techniques such as NLP to identify the sentiment of the user for a particular product based on the review. We also use collaborative filtering technique in our recommendation system to suggest products to users based on user purchases. These recommendations or suggestions can help them buy more products and thereby increase the sales.

Advantages:
- Automated recommendation system
- Improves sales and profits

IV IMPLEMENTATION

In order to implement the proposed application, we have divided it into 2 modules:

1. Admin

2. Consumer

Admin Module:

The admin of the application is accountable for performing the below actions:

1. Uploading the dataset

2. Statistical analysis of user opinions to gather the sentiment.

Research Paper   © 2012 IJFANS. All Rights Reserved

3. Development of machine learning model for sentiment analysis and recommender systems.

4. Evaluate the performance of various machine learning algorithms on the amazon products dataset.
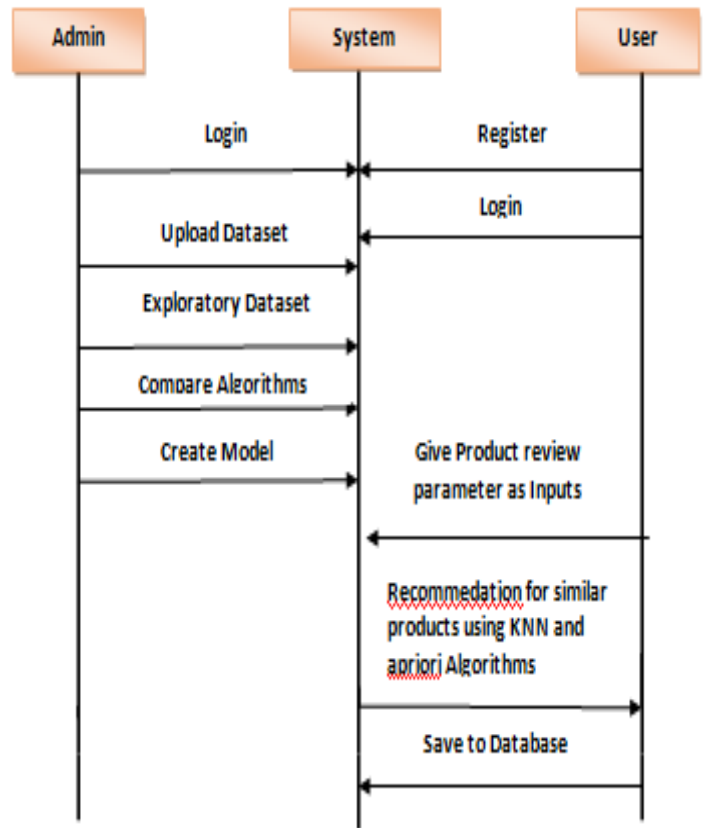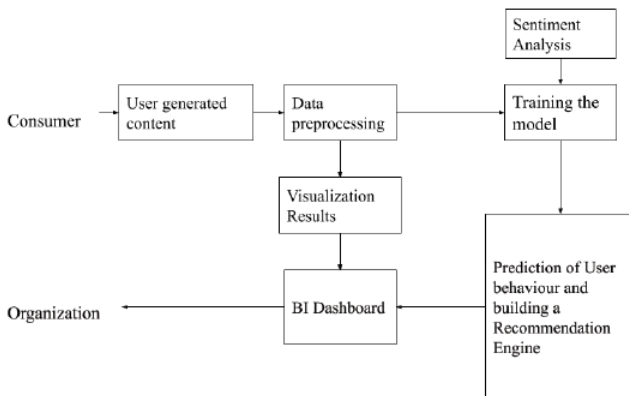
Consumer Module:

The consumer of the system can utilize the machine learning services that are offered like:

1. Logging into the system

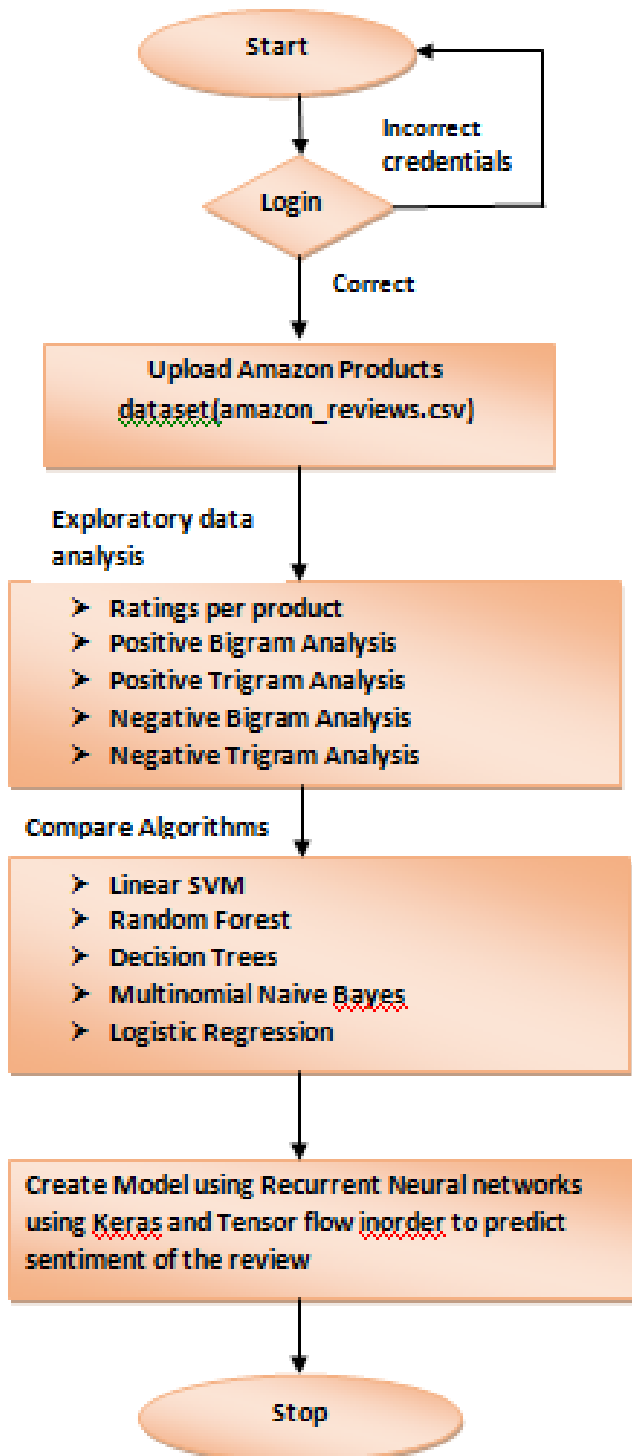2. Write reviews for products and get the sentiment of the review.

3. Get product suggestions for purchase based on userId

IV SYSTEM DESIGN

Architecture diagram:

Data Flow Diagram: Admin

SEQUENCE DIAGRAM:

Start

Incorrect credentials

Login

Correct

Upload Amazon Products dataset(amazon_reviews.csv)

Exploratory data analysis

- Ratings per product
- Positive Bigram Analysis
- Positive Trigram Analysis
- Negative Bigram Analysis
- Negative Trigram Analysis

Compare Algorithms

- Linear SVM
- Random Forest
- Decision Trees
- Multinomial Naive Bayes
- Logistic Regression

Create Model using Recurrent Neural networks using Keras and Tensor flow inorder to predict sentiment of the review

Stop

Research Paper

Data flow Diagram: User

V PROJECT EXECUTION

Home page

This is the application's home page when PyCharm is used to run the programme. The programme is hosted on a web server, and a URL is produced to access it. When a user clicks on the URL, the page shown below opens in a browser.
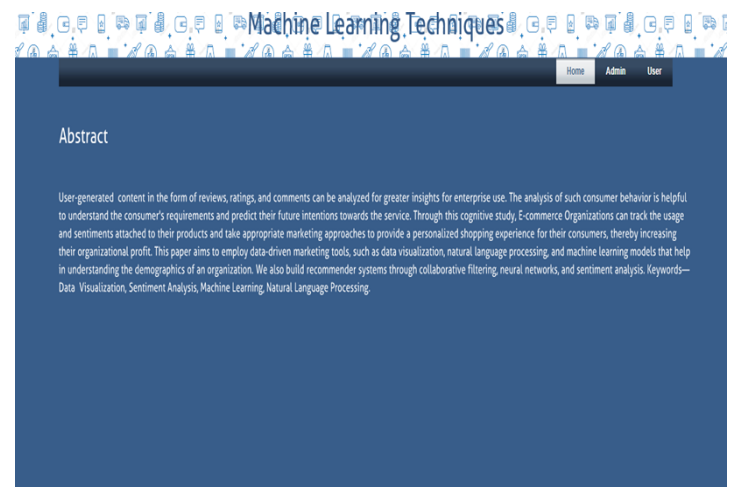
Home page screenshot

Admin Login

The admin module's login page is located here. In order to perform tasks like uploading the dataset, training the dataset, exploratory data analysis of the dataset, feeding the dataset to different machine learning algorithms to find the algorithm that can meet the best accuracy, and creating a model that can be hosted on the Flask Application to be used by the users, the administrator must log into the system using his credentials.

472

Research Paper
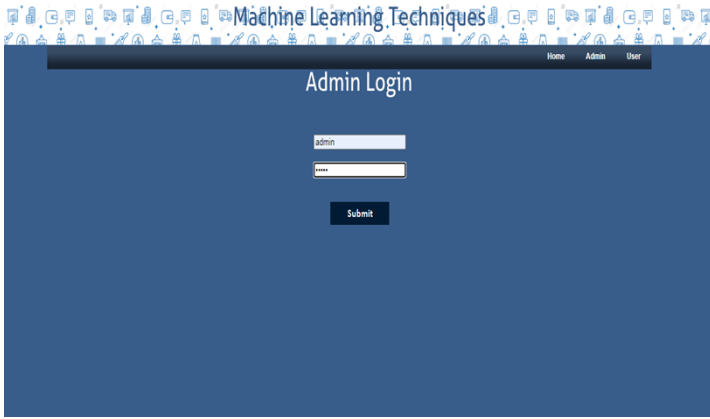
Admin login

Upload Dataset

The system administrator can upload datasets that are used to train machine learning models on this page. To upload a file to a server, an administrator must first choose the file by clicking the Choose file button, then click the Upload button. A success message indicating that the file was successfully uploaded would be shown once the upload was finished. We are using the dataset amazon reviews for this project.
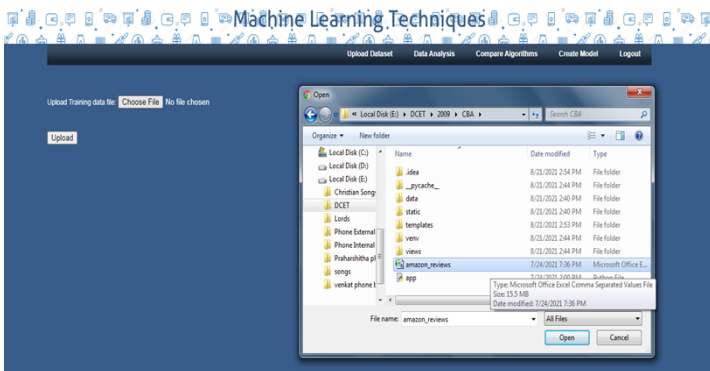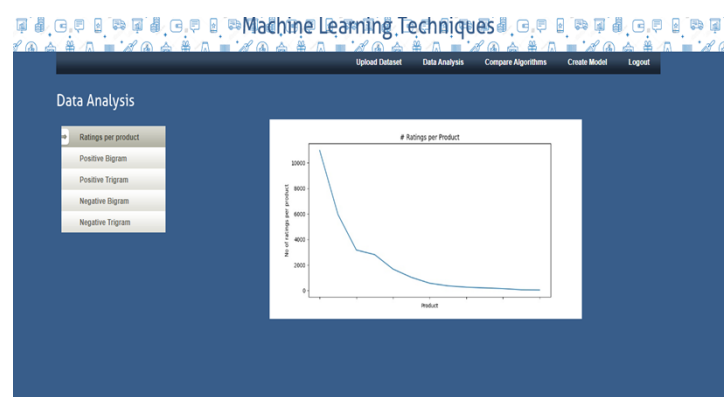

Upload Dataset

Data Analysis

Exploratory data analysis is done on the dataset to uncover patterns, find missing data, and establish links between different output characteristics using graphs, statistics, etc.

Ratings per product
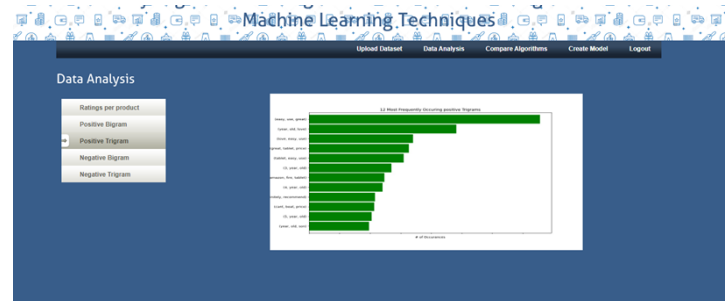

Ratings per product

Positive Bigram

The below graph shows the Most frequently occurring Positive Bigram.
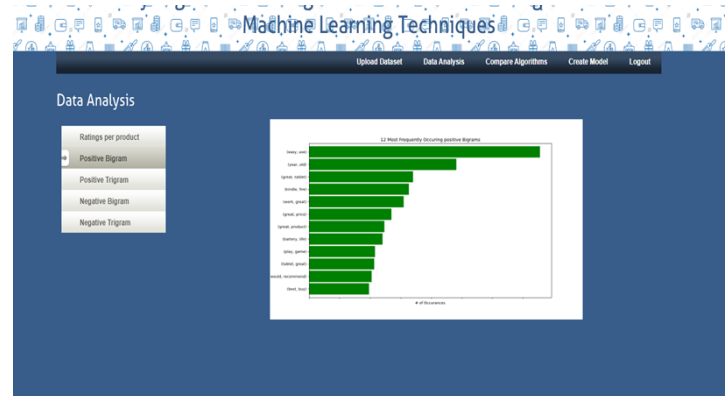
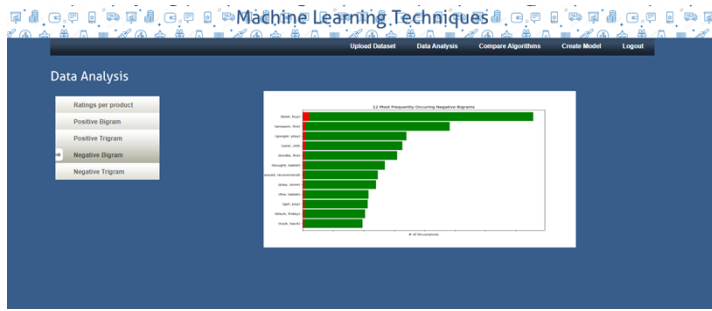Positive Bigram

Positive Trigram

The below graph shows the Most frequently occurring Positive Trigram.

Positive Trigram
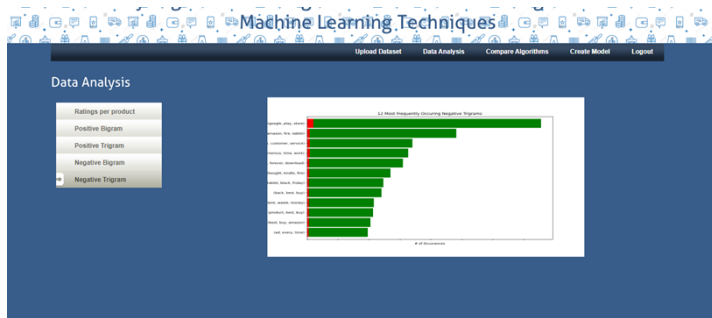
Negative Bigram

The below graph shows the Most frequently occurring Negative Bigram.

Negative Bigram

Negative Trigram

The below graph shows the Most frequently occurring Negative Trigram.
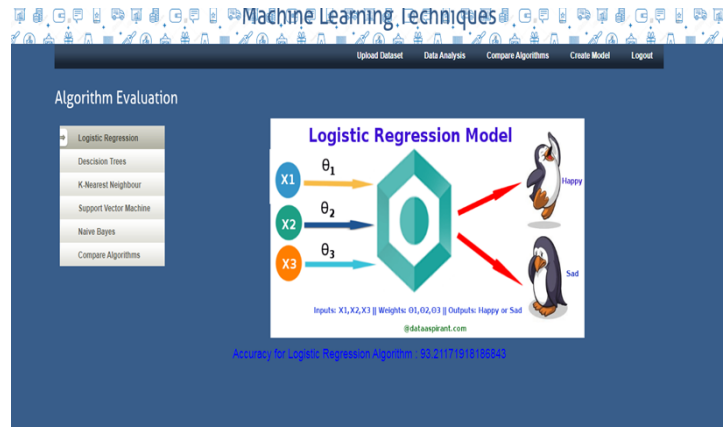

Negative Trigram

Compare Algorithms

On this page, the admin can feed the dataset to various Algorithms to train them and get the test accuracy for each algorithm.
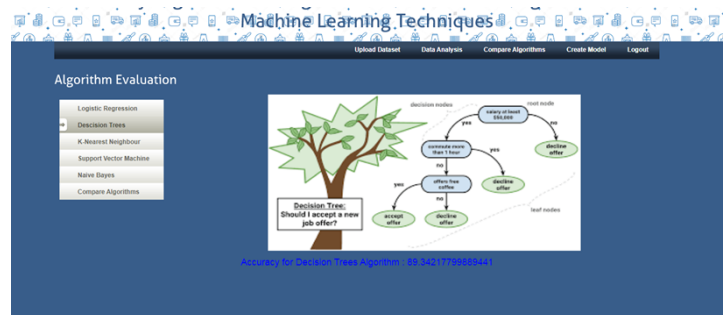
Logistic Regression

When the dataset is feed to Logistic regression algorithm, we observe that the test accuracy is 93.21171918186843%.

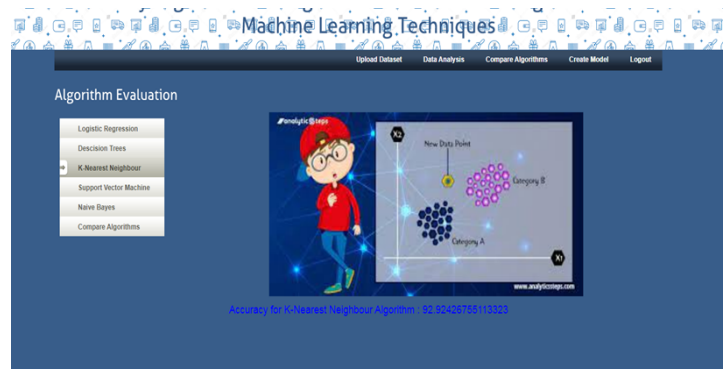Logistic Regression

Decision Trees

When the dataset is feed to Decision Trees algorithm, we observe that the test accuracy is 89.4416804864566%.

Decision Trees

K-Nearest Neighbour

When the dataset is feed to K-Nearest Neighbour algorithm we observe that the test accuracy is 92.92426755113323%.
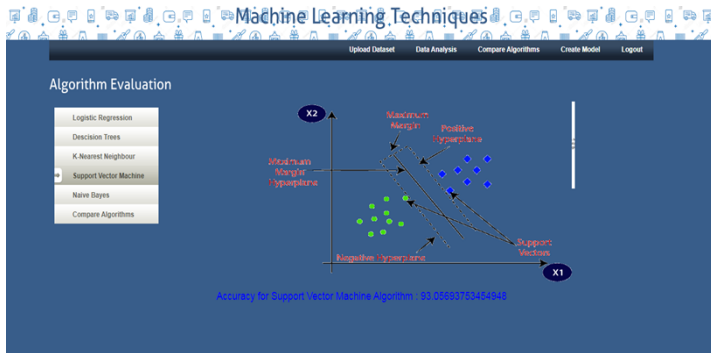

K-Nearest Neighbour

Support Vector Machine

When the dataset is feed to Support Vector Machine algorithm, we observe that the test accuracy is 93.05693753454948%.
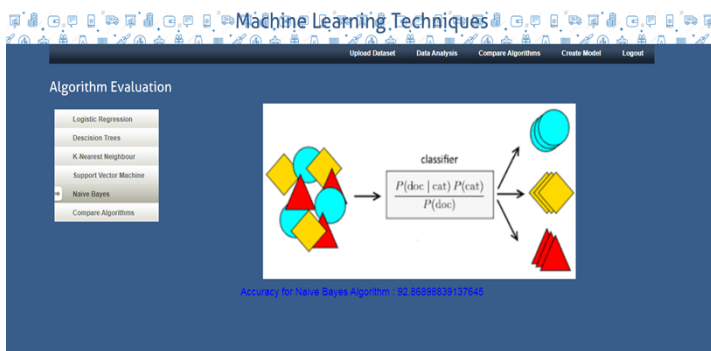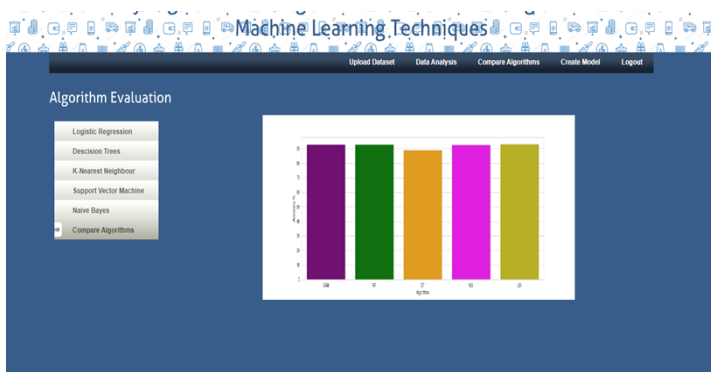

Support vector machine

Naive Bayes

When the dataset is feed to Naive Bayes algorithm, we observe that the test accuracy is 92.86898839137645%.

Naive Bayes

Compare Algorithms

This screen shows the comparison of various test accuracies of the Algorithms.

Compare Algorithms

VI CONCLUSION

The major goal of this study is to assist e-commerce companies in the analysis of their sales as well as in the comprehension of user preferences and feedback in order to help them create marketing strategies that will increase their revenue and bottom line. Users' evaluations and feedback provide firms with information about what the public expects from products and services, and these insights can be used to develop marketing strategies that specifically target end users. To do this, recommender systems must be used to find comparable products that a client would like to buy once the feelings of the reviews have been assessed. In this project, we used machine learning techniques like sentiment analysis and natural language processing to extract the sentiment of the products from the Amazon product review data set. We then fed the data set to various machine learning algorithms to determine which algorithm can provide the highest accuracy. The Naive Bayes algorithm, which we built using KNN and apriori algorithm, provides the best accuracy, as we have found. According to experimental findings, the system provides the highest level of accuracy in comparison to other systems. We have also integrated this model into the web application so that users can access the list of suggested products for a certain user and assess the tone of a product review.

Future Scope:

It is necessary to further improve the model's structure because the item-based collaborative filtering algorithm's performance analysis is not straightforward. The suggested algorithm is required to execute a new "revolution," and the suggested algorithm's future development will take a new turn. We cannot be left behind in the wave of social evolution unless the proposed algorithm is continuously improved and its course adjusted.

REFERENCES

[1] T. Yoshida, M. Hasegawa, T. Gotoh, H. Iguchi, K. Sugioka and K. Ikeda, "Consumer behavior modeling based on social psychology and complex networks," The 9th IEEE International Conference on E-Commerce Technology and The 4th IEEE International Conference on Enterprise Computing, E-Commerce and E-Services (CEC-EEE 2007), Tokyo, 2007, pp. 493-494. DOI: 10.1109/CEC-EEE.2007.36

[2] R. He, J. McAuley. Modeling the visual evolution of fashion trends with one-class collaborative filtering. WWW, 2016

[3] J. McAuley, C. Targett, J. Shi, A. van den Hengel. Image-based recommendations on styles and substitutes. SIGIR, 2015

[4] Ben Yedder, Hanene & Zakia, Umme & Ahmed, Aly & Trajkovic, Ljiljana. (2017). Modeling prediction in

recommender systems using restricted boltzmann machines. 2063-2068. 10.1109/SMC.2017.8122923.

[5] Bengio, Y. & Delalleau, Olivier. (2008). Justifying and Generalizing Contrastive Divergence. Neural computation. 21. 1601-21. 10.1162/neco.2008.11-07-647.

[6] Abdollahi, Behnoush & Nasraoui, Olfa. (2016). Explainable Restricted Boltzmann Machines for Collaborative Filtering.

[7] Chamlertwat, & Bhatarakosol, & Rungkasiri, (2012). "Discovering consumer insight from twitter via sentiment analysis. journal of universal computer science." 18. 973-992.

[8] S. K. Khatri and A. Srivastava, "Using sentimental analysis in prediction of stock market investment," 2016 5th International Conference on Reliability, Infocom Technologies, and Optimization (Trends and Future Directions) (ICRITO), Noida, 2016, pp. 566-569. DOI: 10.1109/ICRITO.2016.7785019

[9] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019. DOI: 10.1109/ACCESS.2019.2909919

[10] Statnikov, Alexander I., Constantin F. Aliferis, Douglas P. Hardin, and Isabelle Guyon. "A gentle introduction to support vector machines in biomedicine: case studies." (2011).