International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Ad Analysis using Machine Learning

Yashas C N[1], Kishan V[2], Pranay Reddy Juturu[3] and Prof. Pavithra H[4]

[1,2,3]Department of Computer Science and Engineering, R.V College of Engineering, Bengaluru, Karnataka , India
[4]Professor, Department of Computer Science and Engineering, R.V College of Engineering, Bengaluru, Karnataka, India

---

Abstract - *In this paper we understand the concept of recommendation system and how beneficial it is to understand and analyze the advertisements as per user. As the audience keep evolving and there are different generation of users who are interested in different type of content. So, it is really important to recommend the right type of ad's in this evolving system so that the situation can be a win-win situation for the user and for the ad agencies as well. By using right type of recommendation system, this provides a solution of dynamically providing the information to the user and helps in promoting the product to the user and increasing their sales and business. In this paper, we'll discuss and analyze the type of approach used in solving this problem.*

1. INTRODUCTION

Machine Learning is a field of computer science which uses statistical models to classify or predict the data .We are separating the data as test data and train data, where we use the train data to train the model and test data to test and see how that works and check the result meets the expectations and to improve the accuracy of the model.

In this paper, we have briefly explained a model which could recommend ads to the user in the most optimal way and recommends the content in the social media according to the user's interest and has a higher percentage of accuracy.

According to the recent study it is shown that, an average millennial spends around 8 hours a day om social media such (Facebook, Instagram etc..) and in YouTube around 4 billion videos are being watched every day. So social media, is a great source of obtaining the data. So, there are techniques incorporated by large corporations such as Facebook, Google who recommendation system to display the content according to the user's interest. But also, there are some websites by using the right type of recommendation system will be able to provide the right type of content to the right users.

2. RELATED WORK

There are lot of related work with respect to ad analysis using machine learning. One of the most used work is by video and sound analysis. Where the sound and video is segregated into frames and hashing is used so that there are no duplicates and naïve Bayes and sentiment analyzer is used for further recommendation. This method is not optimal as this method introduces to a lot of noise and accuracy will be low. The approach used in this journal is feasible as the accuracy will be high and solves the above problem.

3. ANALYSIS METHODOLOGY

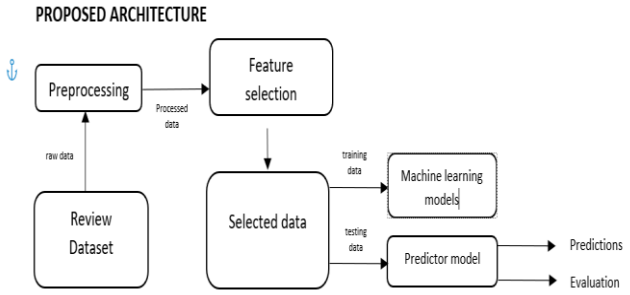

Fig 1.1

In the fig 1.1 the architecture of the proposed model has been explained. The following steps have been explained to achieve this process. They are: -

3.1 Data Visualization

In this method, the data is obtained from different e-commerce website. The data 100k rows and 25 columns and has different fields. The dataset which is chosen 80% test data and 20% train data. The data visualization is shown in the Fig. 1.2 and Fig 1.3.

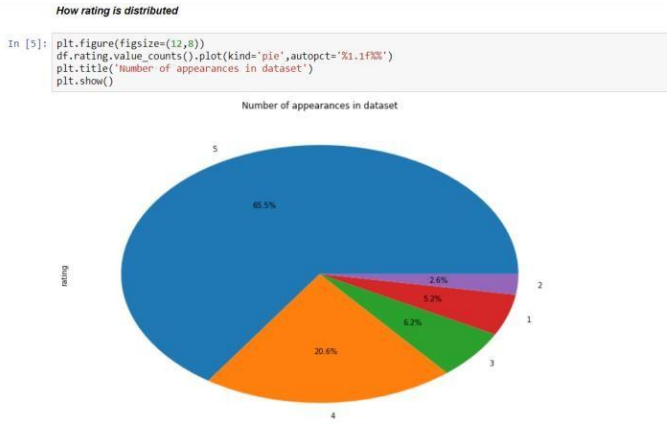

Fig 1.2. Distribution of Rating

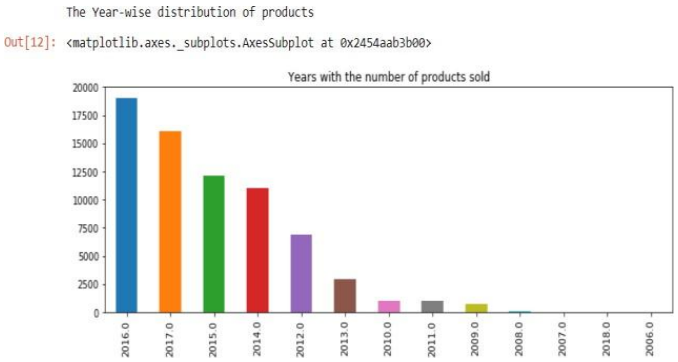

Fig 1.3. Year wise distribution of the product.

3.2 Data Preprocessing

Data pre-processing is the next step in this process as the data is obtained. There will be columns which have any kind of rows of content if there are rows with any type of non-relevant information, that particular row is dropped. And if there are NaN values, relevant information will be in those columns. The code snippet for this process is given below: -

Fig 1.4 Dropping the values which are not relevant

Fig 1.5 Fill Nan with suitable values.

3.3 Feature Selection

This is the next step in the process where various machine learning algorithms are incorporated to obtain the desired result: -

3.31 Gaussian Naïve Bayes

Naïve Bayes algorithm is based on Bayes' theorem with the assumption of independence with every pair of features. This requires extremely small amount of data to estimate the parameters but is fast compared to the more methods. However, it is a bad classifier but is used in real world application such as spam filtering and document classification.

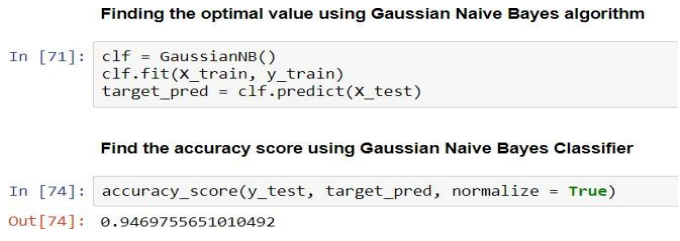

Fig 1.6 Accuracy using Gaussian Naïve Bayes

3.32 Natural language Processing (NLP)

Natural Language processing is a term which is used for automatic computation processing of human languages. This includes both algorithms that take human-produces text as input, and algorithms that produce natural looking text as input.

We have analyzed the review column for our dataset to understand customer review for each product. We have done preprocessing for removal special characters, digits, punctuations etc..

Fig 1.7 Removal punctuation and stop words.

This step (natural language processing) is achieved by 3 algorithms: -

A. Bag of Words.

B. TF-IDF.(Term Frequency, Inverse Document Frequency)

C. Hashing

The code snippet for each of these algorithm is given below : -

Fig 1.8 Bag of Words using CountVectorizer

Fig 1.9 TF-IDF using TfidfVectorizer

Fig 1.10 Hashing using HashingVectorizer

3.33 Sentiment Analysis

Sentiment is like a combination of tone of voice, word choice, and writing style all rolled into one. Natural language with labels about positivity or negativity, we can develop agents that can learn to understand any sentiment. The code snippet for sentiment analysis is given below.

Fig 1.11 Sentiment Analysis for the text

3.34 Stack Ensemble Model

Ensemble modeling is the process of running two or more related but different analytical models and then synthesizing the results into a single score or spread in order to improve the accuracy of predictive analysis and data mining applications. Stacking is a way of combining multiple models, that introduces the concept of a meta learner. Applying stacked models to real-world big data problems can produce greater prediction accuracy and robustness than do individual models.

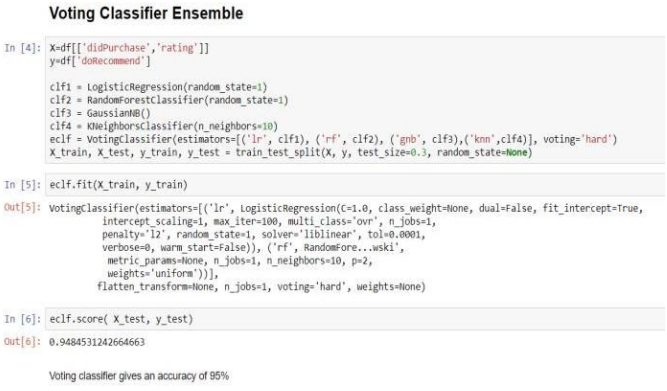

Fig 1.12 Accuracy using Stacked Ensemble for Voting Classifier.

4. Results

Keeping the independent and dependent variables same across various algorithms, we found acceptable results of various algorithms used for our E-commerce dataset.

Dependent variable: doRecommend Independent variable: ProductId, didPurchase, Username, Rating.

Since we take into consideration various independent variables and find the correlation between them, our aim to predict the likelihood for the Advertisement being legitimate to the user and percentage if correctness.

Gaussian Naïve Bayes	94.69%
Natural Language Processing for Sentiment Analysis	Bag of Words TF-IDF Hashing, and sentiment analysis
Stacked Ensemble model	Voting Classifier ~94%

5. Conclusion

By building this analysis system ,we are able to find products which are able to find the products which are similar and can be recommended to a set of users who have similar buying pattern through a set of ads generating based on the current analysis, After using various algorithms, we concluded that, Gaussian Naïve Bayes, NLP technique are most suitable for our dataset and for performing the required sentiment analysis serves as a way to let us know products demand and the kind of Ad based on them.

REFERENCES

[1] R Vinit Kaushik, Raghu R, Maheshwar Reddy, Ankita Prasad, Sai Prasanna M S, "Ad analysis using Machine learning", PESIT – Bangalore South Campus.

[2] Anitha anandhan, Liyana Shubu, Maizatul Akmar Ismail and Ghulam Mujtaba "Social Media Recommender System: Review and Open Research Issues", IEEE Transaction on knowledge and data Engineering, date of publication February 27,2018.

[3] Ashutosh Bansal, Saleena B, Prakash B, "Using Data mining techniques to analyze the customer reaction toward social media."

[4] Bengio, Samy; Goodfellow, Ian J.; Kurakin, Alexey (2017)," Adversial Learning at Scala" Google AI.