Opinion Mining from a Food App

Ajay Sakthivel. C[1], Ajith kumar. R[2], Ajith Kumar. R[3], Karthick. M[4], Janani. I[5]

[1-4]Student, Sona College of Technology, Salem
[5]Assistant Professor, Sona College of Technology, Salem

---***---

Abstract: In today's digital world, a food delivery apps are widely used because it provides a platform for people to share their opinion about the restaurants and cafes they have visited. This project includes analysis of client ratings and reviews in food delivery app's utilizing content mining. Utilizing content mining, we break down the content audits / reviews from the client with a specific end goal to create productive results and legit surveys. We gather survey dataset and use it to check the reliability of the rating given and audit composed by the client. This procedure looks at the client audit premise on their content setting and it demonstrates that how they feel about their visit to that place. To predict the reviews of the restaurant given by customers as positive or negative or neutral on the basis of collected dataset of food delivery app's .

INTRODUCTION

Nowadays people all over the world prioritize using delivery apps to buy the food instead of going to restaurant and they share their opinions or reviews about the food they eaten. Different customers give their different reviews based on their taste and quality of the food and their experience with the delivery app. The basic idea of analyzing the food delivery app's dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant. Restaurant is a domain which are traversed by small as well as big players. Data mining provides a way for both to improve their business with minimum effort. Restaurant business rely on the taste of food, the variety of cuisine that is provided in the restaurant, environment, availability of home delivery, online booking, price etc. When any of the factor is improved or included it is possible to increase customer attention and thus increase productivity in business. A Web-application is developed to predict the rating or category for a given input review using NLP, ANN and FLASK. The dataset is acquired and the aforementioned steps in NLP are performed. The vectorized text is the input is the input to the ANN model. Then the model is trained and tested. The remaining sections of this paper consists of related work on Natural Language Processing (NLP), Artificial Neural Networks (ANN), Proposed work with Architecture diagram and result, future enhancements, conclusion and future work.

NLP

NLP is a subset of AI where customers are provided with the ability to understand a language And perform a specific task.

NLP is widely used in sentiment analysis to determine the customer's ability to determine the customer's say on a company to reflect person's mood.

There are lot of models to implement a above but here we use a bag of words model where the word occurrence is used as a input feature for the model.

NLP pipeline:

STEP 1: Lower case conversation and retain only the characters a-z.

STEP 2: Tokenization.

TOKENIZATION:

Tokenization is the act of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. ... In the process of Tokenization, some characters like punctuation marks are discarded. The tokens become the input for another process like parsing and text mining.

STEP 3:Stop words removal.

STOP WORDS:

Stop words are word**s** which are filtered out before or after processing of natural language data (text). ... Other search engines remove some of the most common words—including lexical words, such as "want"—from a query in order to improve performance.

STEP 4: Stemming /lematization.

STEMMING:

Stemming is the process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma. Stemming is important in natural language understanding (NLU) and natural language processing (NLP). ... When a new word is found, it can present new research opportunities.

LEMATIZATION:

Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.

STEP 5:Vectorization.

VECTORIZATION:

vectorization is a methodology in NLP to map words or phrases from vocabulary to a corresponding vector of real numbers which used to find word predictions, word similarities/semantics. The process of converting words into numbers are called Vectorization.

ANN:

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyzes and processes information. It is the foundation of artificial intelligence (AI) and solves problems that would prove impossible or difficult by human or statistical standards. Artificial neural networks (ANN) are used for modelling non-linear problems and to predict the output values for given input parameters from their training values. Artificial Neural Network(ANN) uses the processing of the brain as a basis to develop algorithms that can be used to model complex patterns and prediction problems. ... In our brain, there are billions of cells called neurons, which processes information in the form of electric signals.

FLASK:

Flask is a popular Python web framework, meaning it is a third-party Python library used for developing web applications.

LITERATURE SURVEY

1. Improving restaurants by extracting subtopics from reviews

"Improving restaurants by extracting subtopics from yelp reviews". (Social Media Expo) (2014) in conjunction with has been enclosed any insights towards generating a wise and additional economical recommendation system that recommends restaurants to users supported their review comments announce on YELP for various business.

2. An overview of different approaches to recommendations

Many popular Ecommerce sites widely use RSs to recommend news, music, research articles, books, and product items. Recommendation systems use personal, implicit and local information from the Internet. Recommender systems or recommendation systems (RSs) are a subset of information filtering system and are software tools and techniques providing suggestions to the user according to their need.

3. Opinion reviews for Thai restaurant review using NLP

To help the businesses, the model of opinion mining is proposed for classifying the reviews and to analyze the attitude of customers for improving their products and services. In this research, the artificial neural network is applied to classify the positive and negative reviews. Experiments showed that the accuracy of the opinion feature identification is 83.33% and the accuracy of the opinion polarity classification is 81.47% with a testing data which be manually collected from a hotel review on the agoda website.

4. Opinion Mining and Summarization of Hotel Reviews

It is difficult for web users to read and understand contents from large number of reviews. Important and useful information can be extracted from reviews through opinion mining and summarization process. We presented machine learning and Senti Word Net based method for opinion mining from hotel reviews and sentence relevance score based method for opinion summarization of hotel reviews. We obtained about 87% of accuracy of hotel review classification as positive or negative review by machine learning method.

5. Extensive hotel reviews classification using long short term memory

Text Mining has a wide variety of applications such as sentiment analysis, spam detection, sarcasm detection, and news classification. Reviews classification using user sentiments is an important and collaborative task for many organizations. This research is a classification study on the hotel review sentiments given by guests of the hotel. The results reveal that the proposed model performs better as compared to the existing state-of-the-art models when combined word embedding with LSTM and shows an accuracy of 78%, precision 80%, recall 71%, and F1-score 76.53%.

6. Statistical Analysis for Customer Product Reviews in Russian Internet Segment Using Text Mining

In the paper, we analysis customer product reviews on the base of statistical approaches of Text Mining and machine learning models. The reviews were collected from Russian sites of popular online store. Data analysis to solve classification problems was carried out in the R environment for statistical analysis. Despite not taking into account word order and relations between words within the bag-of-words model, the results show that a statistical approach to analyzing reviews in the case of classification allows achieving high results.

7. Understanding and Detecting Fake Reviews in App Stores

This paper studies fake reviews, their providers, characteristics, and how well they can be automatically detected. This inspired the development of a simple classifier to automatically detect fake reviews in app stores. On a labelled and imbalanced dataset including one-tenth of fake reviews, as reported in other domains, our classifier achieved a recall of 91% and an AUC/ROC value of 98%.

We discuss our findings and their impact on software engineering, app users, and app store operators.

8. Online Customer Reviews of Products Using Opinion Mining with Clustering

Internet shopping is a method for powerful exchange among cash and merchandise which is finished by end clients without investing a huge energy spam. The goal of this paper is to dissect the high-recommendation web-based business sites with the help of a collection strategy and a swarm-based improvement system. At first, the client surveys of the items from web-based business locales with a few features were gathered and, afterward, a fuzzy c-means (FCM) grouping strategy to group the features for a less demanding procedure was utilized

9. Fine Grained Sentiment Analysis of App Reviews

The star ratings are given to the whole app and developers do not have a mean to analyze the feedback for the single features. In this paper we propose an automated approach that helps developers filter, aggregate, and analyze user reviews. We use natural language processing techniques to identify fine-grained app features in the reviews. We then extract the user sentiments about the identified features and give them a general score across all reviews. The extracted features were coherent and relevant to requirements evolution tasks.

10. Opinion mining from online hotel reviews – A text summarization approach

This study proposes a novel multi-text summarization technique for identifying the top-k most informative sentences of hotel reviews. The results indicate that the proposed approach outperforms the other two, and most of the subjects believed that the proposed approach can provide more comprehensive hotel information. Previous studies on review summarization have primarily examined content analysis, which disregards critical factors like author credibility and conflicting opinions. We considered such factors and developed a new sentence importance metric.

11. Predicting Helpfulness of Online Reviews

The traditional way of determining the helpfulness of a review is through the feedback from human users. However, such a method does not necessarily cover all reviews. Moreover, it has many issues like bias, high cost, etc. Thus, there is a need to automate this process. This paper presents a set of machine learning (ML) models to predict the helpfulness online reviews. This paper presents a set of machine learning (ML) models to predict the helpfulness online reviews.

12. Using Deep Learning for Online Reviews Helpfulness Prediction

This thesis proposes a system for online reviews helpfulness prediction using Recurrent Convolutional Neural Networks (RCNN) within two approaches. The first one is a supervised learning approach, which is similar in spirit to existing works, while the second one, which is the unique aspect of our work, tackles this problem as a semi-supervised learning problem taking advantage of the large number of reviews with no helpfulness feedback from the customers. Both approaches are evaluated on a subset of the famous Amazon product reviews dataset and the results show that both approaches are better than existing approaches

13. Employing Deep Learning Methods for Predicting Helpful Reviews

In This paper we present a set of machine/deep learning models, especially using Recurrent Convolutional Neural Network (RCNN) to predict the helpfulness of reviews. The results show that both approaches are better than existing approaches. Moreover, the results show that the second approach has a remarkably better performance compared with the first one, which is in accordance with recent trends in machine/deep learning that focus on benefiting from the huge amount of unlabeled data to enhance the performance of supervised models.

PROPOSED SYSTEM

FLOWCHART
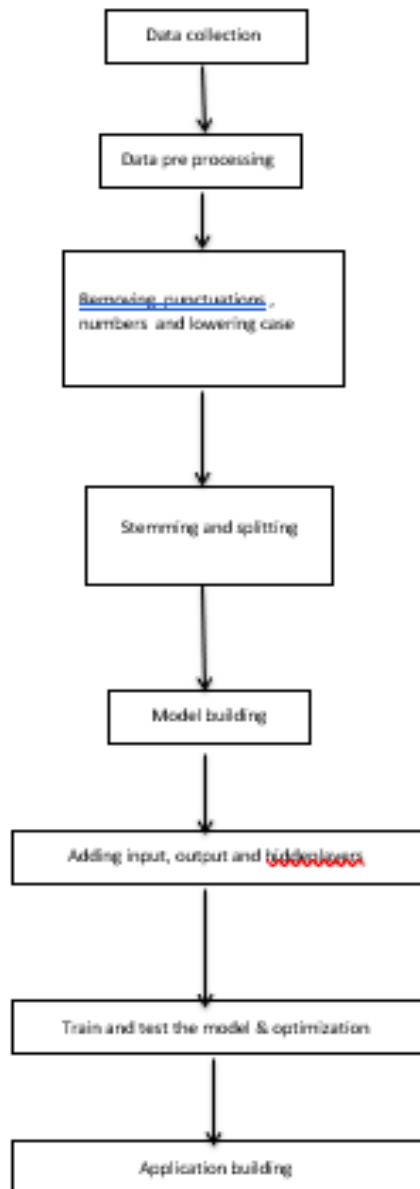
RESULT:

Our project tested 78.7% accurate, which is good measure. Given a review our model displays the performance appraisal on scale

of (1 to 5 stars).

1 - Poor,

2 – fair,

3 – Average,

4 – good,

5 – excellent.

CONCLUSION

Our present model displays the performance appraisal based on the review or costumer opinion given. Our future enhancement would be encouraging the costumers what went well or asking suggestions and how likely they would recommend this to others.

FUTURE ENHANCEMENT

Food delivery app's dataset is used to a create classification model for restaurant rating. It was found that Multilayer perception work well with this dataset. In this paper an attempt is made to predict the process which would further enhance the rating of the review. This project can be further extended to create a tool to evaluate the trigger to improve the ratings.

REFERENCES:

1. https://www.semanticscholar.org/paper/Improving-Restaurants-by-Extracting-Subtopics-from-Huang-Rogers/923d01e0ff983c0ecd3fde5e831e5faff8485bc5

2.https://www.semanticscholar.org/paper/The-groupon-effect-on-yelp-ratings%3A-a-root-cause-Byers-Mitzenmacher/6802843f1741a552f240a1afd8fe16ab4cc9650b

3.https://cs224d.stanford.edu/reports/WuJi.pdf

4.http://www.acadpubl.eu/hub/2018-119-15/2/373.pdf

5. https://smallbiztrends.com/2016/10/mobile-ordering-app-benefits.html

6.https://www.computerworld.com/article/2487774/mobile-apps/how-apps-are-changing-fast-food.html

7.https://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/inf2b-learn-note07-2up.pdf

8.https://www.researchgate.net/publication/313803362_The_Analysis_and_Summarizing_System_of_Thai_Hotel_Reviews_Using_Opinion_Mining_Technique

9.https://www.researchgate.net/publication/343535027_Statistical_Analysis_for_Customer_Product_Reviews_in_Russian_Internet_Segment_Using_Text_Mining

10.https://www.researchgate.net/publication/332750476_Towards_Understanding_and_Detecting_Fake_Reviews_in_App_Stores

11. https://www.kaggle.com/laowingkin/amazon-fine-food-review-sentiment-analysis

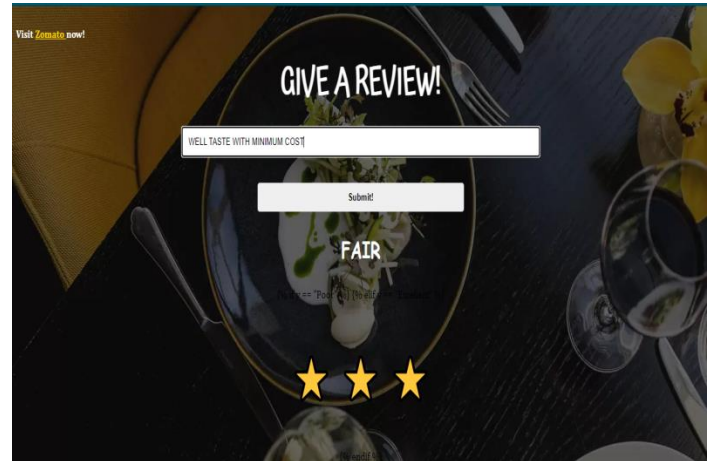12.https://thesai.org/Downloads/Volume11No7/Paper_55-Effective_Opinion_Words_Extraction_for_Food_Reviews.pdf

13.https://www.irjet.net/archives/V6/i4/IRJET-V6I4134.pdf

14.https://www.ijana.in/Special%20Issue/S71.pdf

15.https://archive.ics.uci.edu/ml/datasets/Bag%2Bof%2BWords

16.https://github.com/yanndupis/RNN-Amazon-Fine-Food-Reviews

17.https://github.com/monkeylearn/hotel-review-analysis

18.https://www.youtube.com/watch?v=X2vAabgKiuM

19.https://www.youtube.com/watch?v=MK6T6bjNO90

20.https://www.youtube.com/watch?v=S3q-ow0Ktly

OUTPUT SCREENSHOTS