Survey on Online Product Review Analysis and Recommendation using Machine Learning Technique

Akshatha S[1], Madhurya J[2], S Lakshmi Priya[3], Manasa Manjunath[4]

[123]UG Students, The National Institute Of Engineering, Mysuru

[4]Assistant Professor, Department of ISE, The National Institute of Engineering, Mysuru

Abstract- In these days understanding client sentiments is of predominate importance in promoting ways. This not solely offer the business insights however customers perceive their product and/or services, it'll jointly offer them a thought on a way to boost their offers. This paper makes an attempt to grasp the correlation of various variables in client reviews on a women clothing e-commerce and to classify every review and check whether it recommends the reviewed product or not and whether or not it consists of positive, negative, or neutral sentiment. To attain these goals, we employed univariate and multivariate analyses on dataset features other than review titles and review texts. Results have shown that a recommendation is a robust indicator of a positive or negative sentiment score. On the opposite hand, ratings in product reviews are fuzzy indicators of sentiment scores, hence we are attempting to find out whether we have a tendency to reach a maximum F1-score of recommendation classification and sentiment classification.

Keywords- E-commerce; Bayesian Algorithm; Text Processing; sentiment classification; supervised learning.

I.   INTRODUCTION

Companies are beginning to use social media as a tool for understanding their clients, so as to order to any other further improvements in their products and/or services. Because of this movement, text analysis has become a popular field of research in linguistics and linguistic communication process. One of the most popular problems in the field is text classification, a task that makes an attempt to categorize documents to one or more classes which will be done manually or computationally. Customer reviews and ratings are taken from 'Women's Clothing E-Commerce' dataset which is taken from Kaggle.

Sentimental analysis

The process of characterizing and categorizing opinions specified in a piece of text, particularly to check whether the perspective towards a selected topic, product, etc. is either positive, negative or neutral.These days people are expressing and exchanging their thoughts through on-line internet forums, blogs, and in different social media platforms. Someday they're giving reviews and opinions on completely different products, brand and their services. Their reviews towards a product isn't solely improves the merchandise of the product quality but also influence purchase selection of the customers. Thus, product review analysis is wide accepted platform where customer can easily be aware about their requirements. Sentiment mining is one of the necessary aspects of knowledge mining. Sentiment analysis are often performed using machine learning approach or lexicon based approach or a combination of both.

II.   IMPLEMENTATION METHODOLOGY

Machine Intelligence Library Used

As for the data preprocessing and handling, the numpy and pandas Python libraries were used. For data visualization, the matplotlib and seaborn Python libraries were used. Word Cloud is one of the data visualization technique used for representing text data in which the size of each word indicates its frequency or importance

Dataset

The Women's Clothing E-Commerce Reviews was used as the dataset for this study. This dataset consists of reviews written by real customers, thus it's been anonymized, i.e. client names weren't enclosed, and references to the corporate were replaced with "Retailer".

TABLE 1 Dataset columns

Features
Clothing ID
Age
Title
Review Text
Rating
Recommender IND
Positive Feedback Count
Division Name
Department Name
Class Name

Data Analysis

Here we analyze the dataset features and their implications on user recommendation and review sentiments.

Sentiment mining is an important aspects of data mining. This can be performed using machine learning approach or lexicon based approach or a combination of both. In this experiment we have made use of Machine Learning approach.

Machine learning approach

It involves both supervised and unsupervised techniques. Supervised techniques are implemented by building a classifier. The commonly used algorithms are Naïve Bayes, Support Vector Machines(SVM). It's stated that SVM is more appropriate for sentiment classification as they perform better when the reviews contain both positive and negative words. But SVM works with large data set. Naïve Bayes works quite efficiently with a small set of training data. The appropriate set of features must be selected for sentiment classification.

The commonly used features are: -

● **Term presence and frequency-** This includes unigrams or n-grams and their frequency. For sentiment analysis unigrams, have proved to be more effective.

● **POS information-** Part of speech is used to disambiguate sense to guide feature selection. POS is used for Text indexing and retrieval. This POS tags could be for identifying and treating differently the different meanings of polysemous word.

● **Negations-** These are important as they have the capability to reverse the sentiment of a review.

    Example-"like", "don't like".

Naive Bayes

Naive Bayes could be a straightforward technique for constructing classifiers. There's not one algorithmic rule for coaching such classifiers, however a bunch of algorithms supported a typical principle: all naive Bayes classifiers assume that the worth of a selected feature is freelance of the worth of the other feature, given the category variable.

For example, a fruit that could be thought to be an orange, if it's orange, round, and regarding ten cm in diameter. A naive Bayes classifier considers all of those options to contribute severally to the likelihood that this fruit is an orange, in spite of any attainable correlations between the colour, roundness, and diameter options.

Naive Bayes classifiers are often trained terribly with efficiency in a very supervised learning method. An added advantage of Naive Bayes is that it solely needs a little variety of information to estimate the parameters necessary for classification. Here we've used Multinomial Naive Bayes algorithm.

III. DESIGN

In this experiment, we have used Naive Bayes algorithm. Our dataset i.e. Women's E-commerce Clothing has 11 columns in which we have used recommendation, rating, review text columns. Word cloud is used in extraction of most used words and classifying their polarity. Design includes four phases, gathering all user reviews for products, pre-processing the gathered reviews using NLP, analyzing those reviews, generation of final review of product.

● All the reviews present in text format were gathered.

● Data preprocessing was performed to remove all the stop words, white spaces and other miscellaneous data was cleaned.

● Analyzing was done using Naive Bayes algorithm and supporting libraries like pandas, numpy, matplotlib, seaborn, word cloud,etc.,

● On implementing the above idea and following all the previously mentioned steps we will be able to get the final review of product which can be used.

IV. CONCLUSION

This paper tackles the drawback of sentiment analysis, sentiment polarity categorization. During this experiment, we have found that there have been a lot of reviews given by females compared to males. From the analysis made, the general division products were more sold out than others. Age group between30-40 gave more rating when compared to other age groups. Word cloud is used for classifying words based on sentiment and polarity. After predicting the values and evaluating our model against actual ratings using confusion_matrix and classification report we obtain F1 score.

V. REFERENCES

[1]. X. Fang and J. Z. Fang, "Sentimental analysis using product review," in Journal of Big Data, 2015.

[2]. J. N and Liu, "Opinion spam and analysis," in International Conference on Web Search and Data Mining, New York;NY;USA, 2008.

[3]. J. L, Y. M, Z. M, l. X and Z. T, "Target-dependent twitter sentiment classification," in Association of Computational Linguistics, PA;USA.

[4]. T. LK-W, T. Y.-L. Na J-C and C. K, "Sentiment-level sentiment polarity classification using a linguistic approach," in Digital Libraries for , Germany, 2011.

[5]. L. Y, Z. J, W. X and Z. A, "An information theoritic approach to sentiment polarity classification," in Proceedings of the 2Nd Joint WICOW/AIRWeb Workshop on Web Quality, 2012.

Akshatha S
Studying Information Science and Engineering at The National Institute of Engineering, Mysuru**,** graduating in May 2019
Research area Machine Learning

Madhurya J
Information Science and Engineering Undergrad Student at The National Institute of Engineering, Mysuru Graduating in May 2019
Research area Machine Learning

S Lakshmi Priya
Studying Information Science and Engineering at The National Institute of Engineering, Mysuru,graduating in May 2019
Research area Machine Learning