

# Sentimental Analysis on E-commerce Women's Clothing

<sup>1</sup>Anuj Kumar Prajapati, <sup>2</sup>Abhishek Pandey, <sup>3</sup>Mradul Kumar Jain

<sup>3</sup>Associate Professor, <sup>1,2</sup> Department of Computer Science and Engineering, ABES Engineering College, Ghaziabad 201009, Uttar Pradesh, India.

<sup>1</sup>anuj.16bcs1088@abes.ac.in, <sup>2</sup>abhishek.16bcs1069@abes.ac.in, <sup>3</sup>mradul.jain@abes.ac.in

**Abstract**— From the last few years people seem to be more influenced to work in the area of sentiment analysis. Sentiment analysis provides a way to various businesses (here, e-commerce) to views of customers to their product or services. Not only this it also helps businesses to improve the quality of their product. The purpose of the article is to understand the correlation of various variables in customer reviews of women's clothing and their classification into positive, negative or neutral. To achieve the desired result we've used various Machine Learning libraries such as Pandas, python's re module for data, a natural language processing library called NLTK and to the result we've used a library known as.

**Keywords**— Data Science, Data Analysis, Classification, Data Visualization, e-commerce, Sentiment Analysis, Natural Language Processing(NLP).

## I. INTRODUCTION

In this era of internet business are opting online methods to increase their reach to the customers and in this process reviews on a product become a key for the businesses to understand their customers and to make their product reach the mark of customer's requirement. Review analysis comes in handy to serve the purpose. One of the most important problems in this area is text classification. By classifying the reviews we can perform what is the most required task for today i.e., Sentiment Analysis (computational method using statistics and natural language processing to categorize the opinions given in the text) which can further help the businesses to improve their product or services and hence, they can provide better user experience to their customers. In this paper, we'll be making an attempt to understand the sentiments of the customers on the basis of their reviews on e-commerce websites so that companies can develop a strong, robust recommendation system for their customers to improve their business.

## II. RELATED WORK

In paper the author has discussed the importance of sentiment analysis and how it can be used to understand the customers choice and author also tries to find out the age group of the customers that are satisfied by purchasing things online. The author first tries to analyze the non-text review function, such as age, class of costumes purchased, etc. and then tries to find the relation between them and the product recommended., he tries to implement a two-way

recursive neural network with long-term memory to determine whether the review text recommends the purchased product or not. Review sets and customer generated content are of very unpredictable nature so various text analysis techniques such as topic modeling, opinion mining, sentiment analysis, classification etc. play a role. So to achieve the purpose with Google have been used to implement a neural network for the study. Python libraries like and Panda have handled the task of data and lastly the use of and has made the data visualization easy in form of data clouds. Only reviews that are thoughtful and insight can be used for the interpretation of results and develop a robust system for recommendation of products. indicators of the usefulness of the review vary in different areas of research due to availability. In the area of online travel sites, opinions with more votes are considered more helpful to consumers. Identifying helpful reviews through feature engineering can reduce consumer search costs. This paper is very helpful in our project and gives an idea of the challenges that we could face. It also gives us the brief idea of how it is difficult to make a reputation in the online market. The paper the data on two basis firstly, analysis on the basis of distribution where it has made analysis based on Age and number of positive feedback, Department and name of division, Rating Recommendation, Class Name and Labels etc. and secondly, analysis based on multidimensional distribution, which performs analysis based on the department name by department name, class name by department name, class name by department name, age by the number of positive reviews, etc. The Univariate Distribution reveals that the age group that were

most satisfied by online purchase were of the age 35 to 44 years because a huge chunk of positive reviews were from this age group only. It also suggests the apparel size and clothing being mostly reviewed. Here the author classified the reviews with rating more than three as positive and less than three as negative. Multivariate Distribution reveals the generalized view of the size of clothes that were most positively reviewed, and also gives a brief view of which class of the fashion was most liked. The paper has made use of recurrent neural network (RNN) and LSTM gate equation which they have implemented using Google TensorFlow.

Figure 1 shows the used Recurrent Neural Network (RNN) and Figure 2 shows the used LSTM gate equation which is implemented using Google Tensorflow.

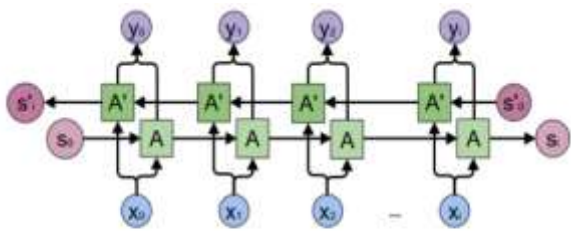


Figure 1: Image from [10]. Computation of a conventional Bidirectional RNN maps input sequences  $x$  to target sequences  $y$ , with loss  $L(t)$  at each time step  $t$ . The RNN cells  $s$  propagate information forward in time (towards the right) while the RNN cells  $s'$  propagate information backward in time (towards the left). Thus at each time step  $t$ , the output units  $o(t)$  (before applying an activation function to get  $y$ ) can benefit from a relevant summary of the past in its  $s(t)$  input, and from a relevant summary of the future in its  $s'(t)$  input.

Figure 1. The Recurrent Neural Network (RNN)

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned}$$

Figure 2. The LSTM gate equation

It is observed that we can underestimate the popularity of a product when we are about to begin the analysis in an online available product as the matter of fact most popular products must have a massive number of ratings and we cannot just ignore. Our model can give the business an idea of estimating the quantity of the popular product to up with the demand of the customer. Given a data set we can give the insight of the fashion trends and can help in.

### III. PROPOSED METHODOLOGY

The process of sentiment analysis begins by the extraction of useful data from the given data set and ends

when the final goal is achieved. The entire process will go in following steps:

Step 1: In this step we'll collect the right data set for classification and sentiment analysis. The data set has been taken from in format.

Step 2: This step involves the cleaning of data from the given raw data set. In this process of cleaning data to extract the desired review text and other parameters like clothing ID, age, title etc. and removing all the entries that have null entries.

Step 3: In this step we'll identify the popular data based on the number of reviews on a unique parameter i.e., .

Step 4: In this step we'll normalize the review text and perform sentiment analysis on filtered review.

Step 5: In this step we will classify the sentiments into positive, negative and neutral and will visualize the data of all three categories in the form of words cloud each for individual categories.

## IV. IMPLEMENTATION & RESULTS

Figures showing the results after implementing the project.

1. Figure 3 shows the variation of number of reviews in accordance with age group to conclude which age group is most satisfied with the product purchase.

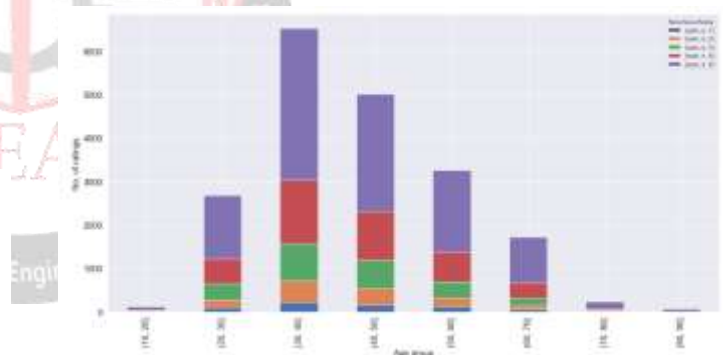


Figure 3. Graph showing the number of reviews in accordance of age group.

2. Figure 4 depicts a box plot of compound sentiment score across age groups.

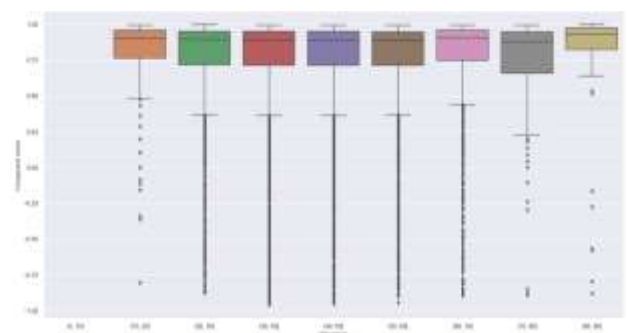
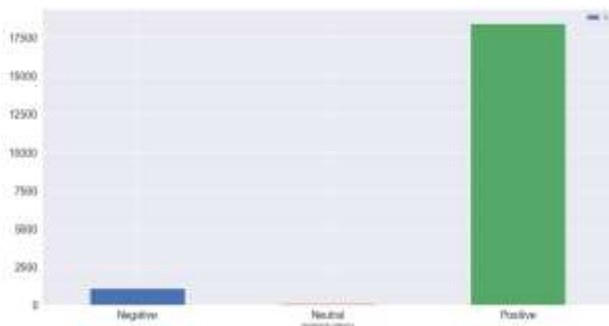


Figure 4. Boxplot of compound sentiment score across age groups

3. Figure 5 is a plot showing the Number of Sentiment across Sentiment Category.



**Figure 5. Plot showing the Number of Sentiment across Sentiment Category**

## V. CONCLUSIONS

This project shows the huge progress achieved thanks to machine learning technology. The main part of the project is the classification of reviews into the three categories given. The model is accurate to 83.8%, but 14.2% of the time, the model does not anticipate that the customer would recommend her product based on her sentiments, and not really have. Given that this is a general set of data, ongoing analysis can only identify trends that will help companies make decisions. To build a model that adapts customer recommendations based on their history, you would need to query a database that contains customer purchase history. In this way, you can apply collaborative filtering and personalize your shopping experience online. This recommendation system is gaining popularity on many e-commerce sites. Therefore, an e-commerce company may consider adopting such a model to increase the convenience of customers using the Internet and increase revenues.

## REFERENCES

- [1] Abein Fred M. , Paul M. , .D. (2018) Statistical Analysis on E-Commerce Reviews, with Sentiment Classification using Bidirectional Recurrent Neural Network
- [2] Nick Brooks. 2018. Women's E-Commerce Clothing Reviews. (2018). <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>
- [3] Fred M. , Paul M. , .Department of Computer Science Adamson University Manila, Philippines <https://arxiv.org/pdf/1805.03687.pdf>
- [4] François Chollet et al. 2015. Keras. <https://github.com/keras-team/keras>. (2015).
- [5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. Deep Learning. MIT Press. <http://www.deeplearningbook.org>.
- [6] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. Neural computation 9, 8 (1997), 1735–1780
- [7] J. D. Hunter. 2007. Matplotlib: A 2D graphics environment. Computing In Science & Engineering 9, 3 (2007), 90–95. <https://doi.org/10.1109/MCSE.2007.55>