Consumer Purchase Intention Predictor

Submitted in partial fulfillment of the requirements of

**Mini Project (CSP605)**

for

Third Year of Computer Engineering

By

Naresh Alwala 18102B0021

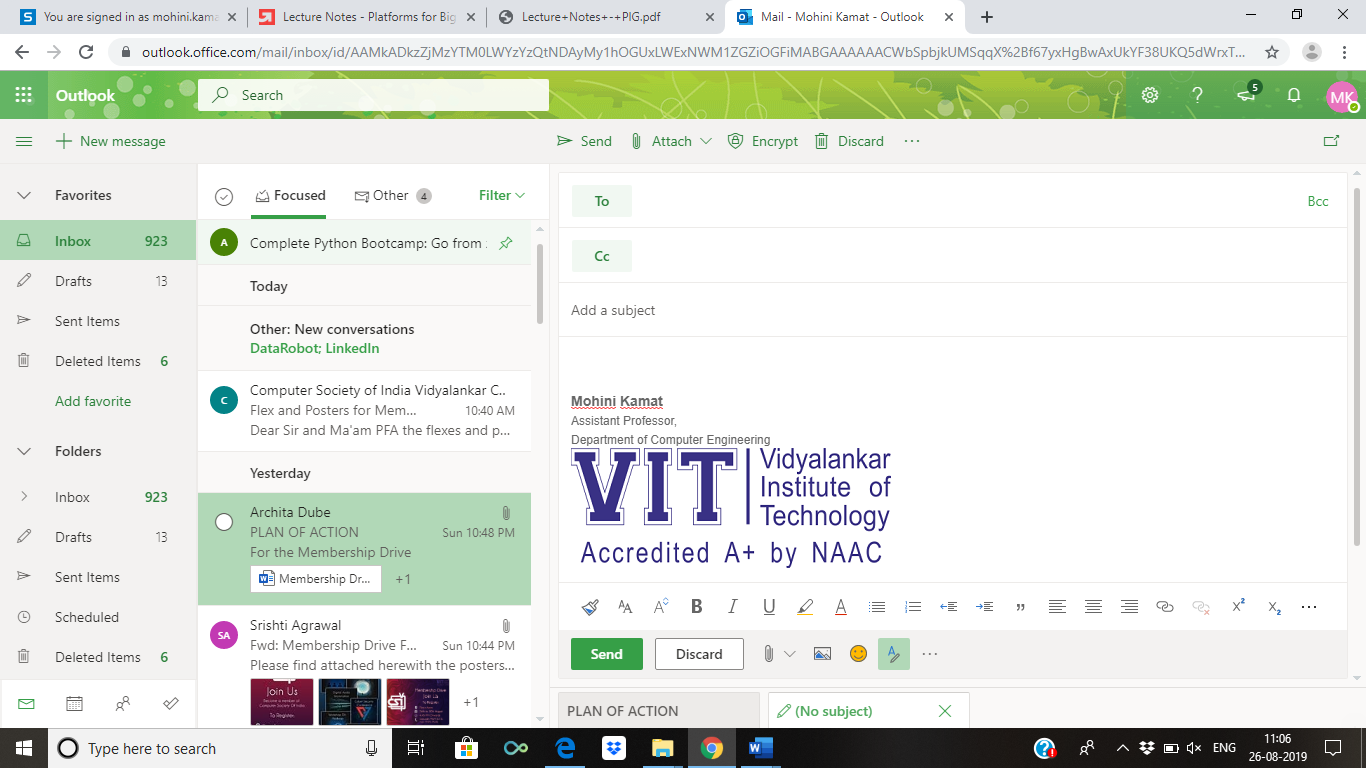
Sumith Pevekar 18102B0031

Aayush Pandey 18102B0032

Under the Guidance of

Prof. Prakash Parmar

Department of Computer Engineering



Vidyalankar Institute of Technology

Wadala(E), Mumbai-400437

University of Mumbai

2020-21

**CERTIFICATE OF APPROVAL**

This is to certify that the project entitled

**“Consumer Purchase Intention Prediction”**

is a bonafide work of

**Naresh Alwala 18102B0021**

**Sumith Pevekar 18102B0031**

**Aayush Pandey 18102B0032**

submitted to the University of Mumbai in partial fulfillment of

**Mini Project (CSP605)**

for

Third Year of Computer Engineering

Guide Head of Department Principal

(Prof. Prakash Parmar) (Dr. Sachin Bojewar) (Dr. Sunil Patekar)

Mini Project Report Approval

This project report entitled ***Consumer Purchase Intention Prediction*** by

1. ***Naresh Alwala 18102B0021***
2. ***Sumith Pevekar 18102B0031***
3. ***Aayush Pandey 18102B0032***

is approved for Mini Project (CSP605) for Third Year of Computer Engineering.

|  |  |
| --- | --- |
| Internal Examiner | External Examiner |

Date: 17/05/2021

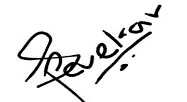
Place: Mumbai

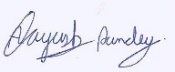
Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Name of student Roll No. Signature



1) Naresh Alwala 18102B0021

2) Sumith Pevekar 18102B0031

3) Aayush Pandey 18102B0032

Date: 17/05/2021

Place: Mumbai

Acknowledgements

Firstly, we would like to express our sincere gratitude to several individuals and organization for supporting us to work on such an interesting and challenging project. We wish to express our sincere gratitude to our supervisor, Professor Prakash Parmar, for his enthusiasm, patience, insightful comments, practical advice and unceasing ideas that have helped us tremendously in our research on the project. His immense knowledge, profound experience and professional expertise in data analysis and algorithms has enabled us to understand the different machine learning algorithms for this project.

We also wish to express our sincere thanks to Vidyalankar Institute Of Technology for giving us the opportunity to research and develop the project of our interest, and also aided us with all the necessary resources required to work on this project.

Abstract

Recently, there has been a significant rise in the ecommerce industry and more specifically in people buying products online. There has been a lot of research being done on figuring out the buying patterns of a user and more importantly the factors which determine whether the user will buy the product or not. In this study, we will be researching on whether it is possible to identify and predict the purchase intention of a user for a product and target that user towards the product with a personalized advertisement or a deal. Further, we wish to develop a software that will help the businesses identify potential customers for their products by estimating their purchase intention in measurable terms from their tweets and user profile data on twitter. After, performing the initial analysis on the twitter data we have found that customers that show purchase intentions have some common keywords in their tweets. We need to apply some text analytical models to tweets data in order to predict purchase intention of users based on their tweets

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr No** | **Description** | **Page No** |
| 1 | Introduction | 8 |
| 2 | Problem Definition | 8 |
| 3  4 | Literature Survey  Proposed Solution | 9  10 |
| 5 | Conclusion | 11 |
| 6 | References | 11 |

Introduction

There have been several research studies for analyzing the insights of online consumers buying behavior. However, only a few have addressed the customers buying intention for products. We want to develop a machine learning approach that will identify potential customers for a product by estimating the purchase intention in measurable terms from tweets on twitter. We have used a text analytical machine learning approach because although text analytics can be performed manually, it is inefficient. By using text mining and natural language processing algorithms it will be much faster and efficient to find patterns and trends. In a way we can say that Purchase Intention detection task is close to the task of identifying wishes in product reviews.

Problem Definition

Purchase intentions are frequently measured and used by marketing managers as an input for decisions about new and existing products and services. Up till now many companies still use customer survey forms in which they ask questions like how likely you are to buy a product in a given time frame and using that information they calculate the purchase intention. We want to see if we can use Twitter tweets to train a model to identify tweets which show purchase intention for a product.

Literature Survey

There have been several research studies for analyzing the insights of online consumers buying behavior. However, only a few have addressed the customers buying intention for products. These include suggestions for a product or a desire to buy a product. They used linguistic rules to detect these two kinds of wishes. Although rule-based approaches for identifying the wishes are effective, but their coverage is not satisfactory, and they can’t be extended easily. Purchase Intention detection task is close to the task of identifying wishes in product reviews. Here we don’t use the rule-based approach, but we present a machine learning approach with generic features extracted from the tweets.

Past studies have shown that it is possible to apply Natural Language Processing (NLP) and Named Entity Recognition (NER) to tweets. However, applying NER to tweets is very difficult because people often use abbreviations or (deliberate) misspelled words and grammatical errors in tweets. Other studies used these techniques to apply sentiment analysis to tweets. The first studies used product or movie reviews because these reviews are either positive or negative. These studies merely analyze the sentiment of a tweet about a product after the author has bought it. We will however be extracting features from tweets to find whether the user has shown purchase intention towards the product or not.

Moreover, research articles that we found online like *Identifying Purchase Intentions by Extracting Information from Tweets* ( February 8, 2017, RADBOUD U NIVERSITY NIJMEGEN) and *Tweetalyst: Using Twitter Data to Analyze Consumer Decision Process* (The Berkeley Institute of Design) investigate if an artificial intelligence approach can predict (from existing user created content on twitter) if someone is a potential customer for a specific company or product and identify users at different stages of the decision process of buying a given product. Further looking at research reports online of different universities and researcher, give us an insight of the impact of social network marketing on consumer purchase intention and how it is affected by the mediating role of consumer engagement.

Some of NLP libraries for preprocessing techniques commonly used for twitter data are the TweetNLP library (a tokenizer, a part-of-speech tagger, hierarchical word clusters, and a dependency parser for tweets), unigrams, bigrams and stemming. There are also some dictionary-based approaches such as using the textBlob library (TextBlob is a Python (2 and 3) library for processing textual data. It provides a consistent API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, and more).

The common machine learning algorithms that we are planning to use for text analysis are Random Forest, Naive Bayes, Support Vector Machine and K-nearest neighbor algorithm.

Proposed Solution

There has been several researches made by different institution on the same field by using NLP sentiment analysis that merely finds the sentiment of a tweet. However, there are several limitation related to it like grammatical errors and complex sentence formations. We are planning to use the dataset of about 3200 tweets that has been scraped from the twitter using web crawler software and official twitter API. After scraping out the tweets we will manually annotate the tweets to classify them as purchase intent tweet or not and then store it csv formatted file. Once the dataset is ready we will preprocess it using different NLP libraries to tokenize each tweet and also perform stemming to reduce each word to its word stem. After the preprocessing task, we will train the text analytical models like SVM, Random Forest, etc algorithms to predict the class of the tweets that it will read from the twitter and finally we will also prepare a dashboard to show the list of all potential users and provide graphical overview.

Conclusion

Up till now for the project, we have successfully created our own database using the web crawler software to extract the tweets from twitter and annoted the data by classifying the tweets as a purchase intent tweet or not. We had to create our own dataset because there does not exist a publicly available dataset for purchase intention based on twitter tweets.

Also Looking at the other researches that are done in the similar field, our project also stands apart since we are planning to implement 5 different models and after evaluating them, we will choose the best one customized to the product data for best accuracy.

References

1. Websites:
   1. <https://scikit-learn.org/stable/>
   2. <https://github.com/jonbakerfish/TweetScraper>
   3. <https://monkeylearn.com/blog/sentiment-analysis-of-twitter/>
2. Relevant Papers
   1. Identifying Purchase Intentions by Extracting Information from Tweets, February 8, 2017, RADBOUD U NIVERSITY NIJMEGEN, BACHELOR ’S THESIS IN ARTIFICIAL INTELLIGENCE.
   2. Tweetalyst: Using Twitter Data to Analyze Consumer Decision Process, The Berkeley Institute of Design.
   3. Mining Purchase Intent by Dublin City University, School of Computing, ADAPT Centre, Dublin, Ireland