

Review analysis of product based on consumer needs using SVM(Support Vector Machine)

Submitted in partial fulfillment of the requirements for
the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

SHREEKESH ANAND (Reg. No. 37110710)

DEEPAK ANAND (Reg. No. 37110176)



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

**Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600 119**

MARCH - 2021



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with "A" grade by NAAC

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **SHREEKESH ANAND (Reg. No. 37110710)** and **DEEPAK ANAND (Reg. No.37110176)** who carried out the project entitled "**Review analysis of product based on consumer needs using SVM(Support Vector Machine)**" under my supervision from **August 2019 to March 2020**.

Internal Guide

Mrs. R. Sathya Bama Krishna, M.E.,Ph.D

Head of the Department

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **SHREEKESH ANAND** hereby declare that the Project Report entitled “**Review analysis of product based on consumer needs using SVM(Support Vector Machine)**” is done by me under the guidance of **Mrs. R. Sathya Bama Krishna, M.E.,Ph.D** Department of Computer Science and Engineering at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in Computer Science and Engineering.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala, M.E., Ph.D., Dean**, School of Computing, **Dr. S. Vigneswari, M.E., Ph.D.,** and **Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Mrs. R. Sathya Bama Krishna, M.E., Ph.D Assistant Professor**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

Recently, there has been an enormous rise among the eCommerce business and loads of specifically in people buying merchandise online. loads of and loads of people have started posting on-line regarding whether or not or not they have to buy for the merchandise or asking whether or not or not they must purchase the merchandise or not. There has been an excellent deal of study being done on determinant the buying patterns of a user and loads of considerably the factors that verify whether or not or not the user will get the merchandise or not. One such platform is Twitter that has become quite widespread in recent years. throughout this study, we tend to our planning to be exploring the matter of characteristic and predicting the acquisition intention of a user for a product. once applying varied text analytical models to tweets information, we've found that it is so accomplishable to predict if a user has shown purchase intention towards a product or not, and once doing a bit analysis we've found that people global organization agency had initially shown purchase intention towards the merchandise have in most cases to boot bought the merchandise

TABLE OF CONTENTS

ABSTRACT	v
LIST OF FIGURES	viii

CHAPTER No.	TITLE	PAGE No.
1.	INTRODUCTION	1
	1.1 MANAGEMENT OF PRODUCT	2
	1.2 RESEARCH AND SIGNIFICANCE	3
2.	LITERATURE SURVEY	4
3.	AIM AND SCOPE	6
	2.1 AIM OF PROJECT	6
	2.1.1 OBJECTIVES	6
	2.2 SCOPES	6
	2.2.1 ADVANTAGES	6
	2.3 INCREASING IN SELLS OF PRODUCT	7
4.	SYSTEM DESIGN & METHODOLOGY	9
	4.1 PROPOSED APPROACH	9
	4.1.1 WEB APPLICATION	9
	4.1.2 DASHBOARD SCREEN	9
	4.1.3 UPLOAD ANNOTATED DATA	9
	4.1.4 ANALYSIS FUNCTION	9
	4.2 DATA COLLECTION AND ANNOTATION	10
	4.3 DATA PREPERATION	10
	4.3.1 DATA PREPROCESSING TECHNIQUES	10
	4.3.2 FORMATION OF DOCUMENT VECTOR	11
	4.4 MODELLING	11

5.	RESULTS AND DISCUSSION	13
6.	CONCLUSION AND FUTURE WORK	22
	REFERENCES	24
	APPENDIX	
	A. PAPER ACCEPTANCE MAIL	46
	B. PLAGIARISM REPORT	47
	C. JOURNAL PAPER	48
	D. SOURCE CODE	59

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE No.
5.1	HOME PAGE	13
5.2	UPLOAD ANNOTATED DATA	14
5.3	NUMBER OF ALGORITHM	15
5.4	COUNT THE PERCENTAGE OF PEOPLE	16
5.5	RESULT OBTAIN BY SVM	17
5.6	TOTAL NUMBER OF TWEETS	18
5.7	TESTING PAGE	19
5.8	SHOWING THE POTENTIAL CUSTOMER	20
5.9	DASHBOARD AND KEYWORD	21

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

This project will be a internet application that predicts the intension that a client can obtain a product that he's fascinated by supported his social media posts like Twitter tweets and user profile information. This will facilitate the business target a selected client a lot of expeditiously and boost their sales. First, we tend to hunt for Twitter tweets of potential customers needing to obtain a product. And supported those tweets we tend to predict the probability that the client can obtain the merchandise. we tend to then build a model by gathering tweets from users WHO have already expressed intention to shop for the merchandise mistreatment their tweet history and if attainable, their internet search history additionally so coaching the text analytical model supported those tweets. mistreatment the model, we tend to input potential customers WHO have tweeted regarding the merchandise however haven't bought it. And supported the coaching information the model estimates a prediction/likelihood of whether or not the client can exit or not. we've restricted the scope of our information to solely mobile phones. Our model predicts the buyer intention for the most recent coming mobile phones. we've tested it on the most recent iPhone X variants and our model has provided promising results with accuracies of up to eightieth considering that we've restricted annotated information

1.2 MANAGEMENT OF PRODUCT

Our project is a web application that can predict the likelihood/certainty that a customer will buy a product that he is interested in based on his social media posts such as Twitter tweets and user profile data. This will help the company/business target that particular customer more efficiently and boost its sales.

We will be searching for Twitter tweets of potential customers wanting to buy a product. And based on those tweets we will estimate/predict the likelihood that the customer will buy the product. We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it. And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not. We will limit the scope of our data to only mobile phones. And our model will predict the consumer intention for the latest upcoming mobile phones by gathering data and updating the model at fixed intervals. We plan to generate a trending list of phones on twitter which will be launching soon or have been recently launched and the users can use that list of products on our application to find a customer for that product/mobile phone.

1.3 RESEARCH AND SIGNIFICANCE

The primary aim of this proposed system is to assist the company or a particular sells person in such a way that they can get all the necessary and detailed information about their product on the given website which will help the company to target on that particular person in much faster way. The proposed system also provides information

or status of their product and the people intension whether they like new or existing product or not. There are several intents of this proposed system. Some of the important intents are as follows:

- This model provides detailed information about the customer review.
- This system has separate website which is used by the company so that they easily find the people who like their product.
- It makes the whole process efficient.
- This process is based on people review on social media platform.
- It provides authentication and keeps the system secured.

CHAPTER 2

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. Studies on identification of wishes from texts, specifically Ramanand et al. (Ramanand, Bhavsar, and Pedanekar 2010) consider the task of identifying 'buy' wishes from product reviews. These wishes 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 (Li et al., 2012) (Liu et al., 2011). However, applying NER to tweets is very difficult because people often use abbreviations or (deliberate) misspelled words and grammatical errors in tweets. Nonetheless, Finin et al. (2010) tried to annotate named entities in tweets using crowdsourcing. 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. Wang et al. (2011) and Anta et al. (2013) analyzed the sentiment of tweets filtered on a certain hashtag (keywords or phrases starting with the symbol that denote the main topic of a tweet). 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.

More recently, research articles 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 like *The Impact of Social Network Marketing on Consumer Purchase Intention in India: Consumer Engagement as a Mediator* (Asian

Journal of Business and Accounting 10(1), 2017) 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. Based on UGT theory (Uses and Gratification Theory).

Some preprocessing techniques commonly used for twitter data are the sentiment140 API (Sentiment140 allows you to discover the sentiment of a brand, product, or topic on Twitter), 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 are used for text analysis are Linear Regression, Random Forest, Naive Bayes and Support Vector Machine. We will be looking at these models later in detail.

CHAPTER 3

AIM AND SCOPE

2.1 AIM OF PROJECT

The primary aim of this proposed system is to find about the specific product, their review, their merits and demerits. It also gives the specific company to easily reach out to the person who likes their product and target them so they can buy their product. This system also shows whether the product of the particular company is still relevant in the market or not.

2.1.1 OBJECTIVES

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. We aim to analyze the tweets related to a product and identify the purchase intention in it. In this way we can rank the tweets which have high purchase intention and report the name of the person who tweeted as potential customer of product. We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it. And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not.

2.2 SCOPE

We want to develop a machine learning model which can predict the numerical value for the consumer intention for a tweet. By doing this we can prove that social media such as Twitter is also an important tool which marketers can use when deciding to target a customer. We believe that our work can be valuable to applications focusing on exploiting purchase intentions from social media.

2.2.1 ADVANTAGES

- Time saving as it is operated online.
- People can see the review , advantage or disadvantage of the product before buying int.
- Company can target the people who wants to buy their product.
- It is very easy and simple to use and saves lot of mone

2.2.2 Increasing in Sells of the Product

Currently we have many recommendation systems available which recommend different products to the user, most of which are not efficient. No such effective model for businesses to identify potential customers. Further, 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 an application 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. In a way we can say that Purchase Intention detection task is close to the task of identifying wishes in product reviews.

We aim to analyze the tweets related to a product and identify the purchase intention in it. In this way we can rank the tweets which have high purchase intention and report the name of the person who tweeted as potential customer of product.

We will make a model by gathering tweets from users who have already expressed intention to buy the product and see their tweet history and if possible, their web search history as well. Using this model, we will input potential customers who have tweeted about the product but have not bought it yet! And based on the training data the model will estimate a prediction/likelihood of whether the customer will buy it or not.

CHAPTER 4

SYSTEM DESIGN & METHODOLOGY

4.1 Proposed Approach

In this section, we describe the details of our approach to tackle the problem of purchase intention detection. We will begin by describing our data collection and annotation process. Then we will describe our approach for data preprocessing and transforming the data to train text analytical models.

Following are the main components of this model:-

4.1.1 Web Application :-

We have developed a web application so that users can easily access our application. A web application also provides portability, speed and is easy to use. It also does not require much hardware to be setup on the user's end.

4.1.2 Dashboard screen :-

We have developed a dashboard which shows the relevant statistics according to the dataset of the product which the user can use to evaluate the results and check the trend for the product.

4.1.3 Upload annotated dataset function :-

The user will press the upload annotated dataset button through which he will be asked to upload the dataset he wants to test for his product and find the relevant customers from that dataset. The output will be in a form of a pie chart and a table showing the different level of customers and the scores assigned to each customer.

4.1.4 Analysis function :-

When the user presses the analysis button, he will be taken to the screen which will show the detailed analysis of our dataset that we have built containing word clouds, positive vs negative tweets and the most used

words for that product.

4.2 Data collection and annotation

As there are no annotated Twitter tweets corpora available publicly for detection of purchase intent, we had to create our own. This was done using a web crawler developed by JohnBakerFish which crawled the website to collect the data. We had collected over 100,000 tweets but since they were not annotated, we had to cut down to just 3200 tweets which were randomly selected out of the dataset and we manually annotated them using a basic criterion we had defined:

Criteria for Labelling of tweets

	Tweet	Class
1	Comparing iphone x with other phone and telling other phone are better?	No PI
2	Talking about good features of iphone x?	PI
3	Talking about negative features of iphone x?	No PI
4	liked video on Youtube about iphone x?	PI

We used just 3200 tweets out of such a large dataset as we were limited by time. We defined definition of Purchase Intention as object that is having action word like (buy, want, desire) associated with it. Each tweet was read by 3 people and final class was decided by maximum voting.

4.3 Data preparation

4.3.1 Data preprocessing techniques:

We processed the tweets using these techniques in chronological order. First, we started our groundwork by converting our text into lower case, to get case uniformity. Then we passed that lower case text to punctuations and special characters removal function. Text may contain unwanted special characters, spaces, tabs and etcetera which has no significant use in text classification. Next step was stop words removal since the tweets also contains useless words which are routine part of the sentence and grammar but do not contribute to the meaning of the sentence. Likes of “the”, “a”, “an”, “in” and etcetera are the words mentioned above. So, we do not need these words, and it is better to remove these. Further we also removed the top 2 most common words because their recurrence does not contribute to the meaning in the sentence. This can also be the result of mistake as the

data we are analyzing is an informal data where formal sentence norms are not taken into consideration. We also removed some rare words like names, brand words (not iphone x), left out html tags etc. These are unique words which do not contribute much to interpretation in the model. Finally we stemmed the words to their root. Stemming works like by cutting the end or beginning of the word, considering the common prefixes or suffixes that can be found in that word. For our purpose, we used Porters Stemmer, which is available with NLTK. We also experimented with lemmatization. The analysis is performed in morphological order. A word is traced back to its lemma, and lemma is returned as the output. But it did not yield a considerable change in the corpus.

After preprocessing the tweets, we were left with about 1300 tweets for training data and remaining for testing.

4.3.2 Formation of Document Vector

We made 3 types of document vectors for the purpose of experimentation. First, is the term frequency document vector. We have stored text and its labeled class in data frame. And we have constructed a new data frame with columns as the words and document count as the rows. So, individual frequency of words in a document count is recorded. Second, is the inverse document frequency vector which is a weighting method to retrieve information from the document. Term frequency and inverse document frequency scores calculated and then product of $TF \times IDF$ is called TF-IDF. IDF is important in finding how relevant a word is. Normally words like 'is', 'the', 'and' etc. have greater TF. So IDF calculated a weight to tell how important least occurring words are. Lastly, we also used the textblob library to help create the document vector. With the help of textblob library we calculated sentiments of individual word and then multiplied the sentiment score with TF and TF-IDF of that word.

4.4 Modelling

For the modelling part, we used different text analytical models to experiment with. Since there are already prebuilt libraries with predefined machine learning models, we chose the Scikit-learn library in python and experimented with these prebuilt models using these configurations.

1. Support Vector Machine (SVM): Simply put, SVM is a supervised machine learning algorithm which does complex transformation on the data. And then it tries to separate data on classes we have defined on our data.

2. Naive Bayes: Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of conditional independence between every pair of features given the value of the class variable.
3. Logistic Regression: Logistic regression is the appropriate regression analysis to conduct when the dependent variable is dichotomous (binary). Like all regression analyses, the logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.
4. Decision Tree: Decision Trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.
5. Neural Network: It is deep learning machine algorithm, which is arranged in a layer of neuron. There is an input layer, output layer and hidden layers of neurons. Neuron network is adaptive as neurons in these layers learn from their initial input and subsequent runs.

CHAPTER 5

RESULTS AND DISCUSSION

As a result of this system getting introduced in our defence system, it will act like a bridge to cover the gap of communication between the customer and the company. The company or person will be able to understand about their product better which will further result in taking quick actions.

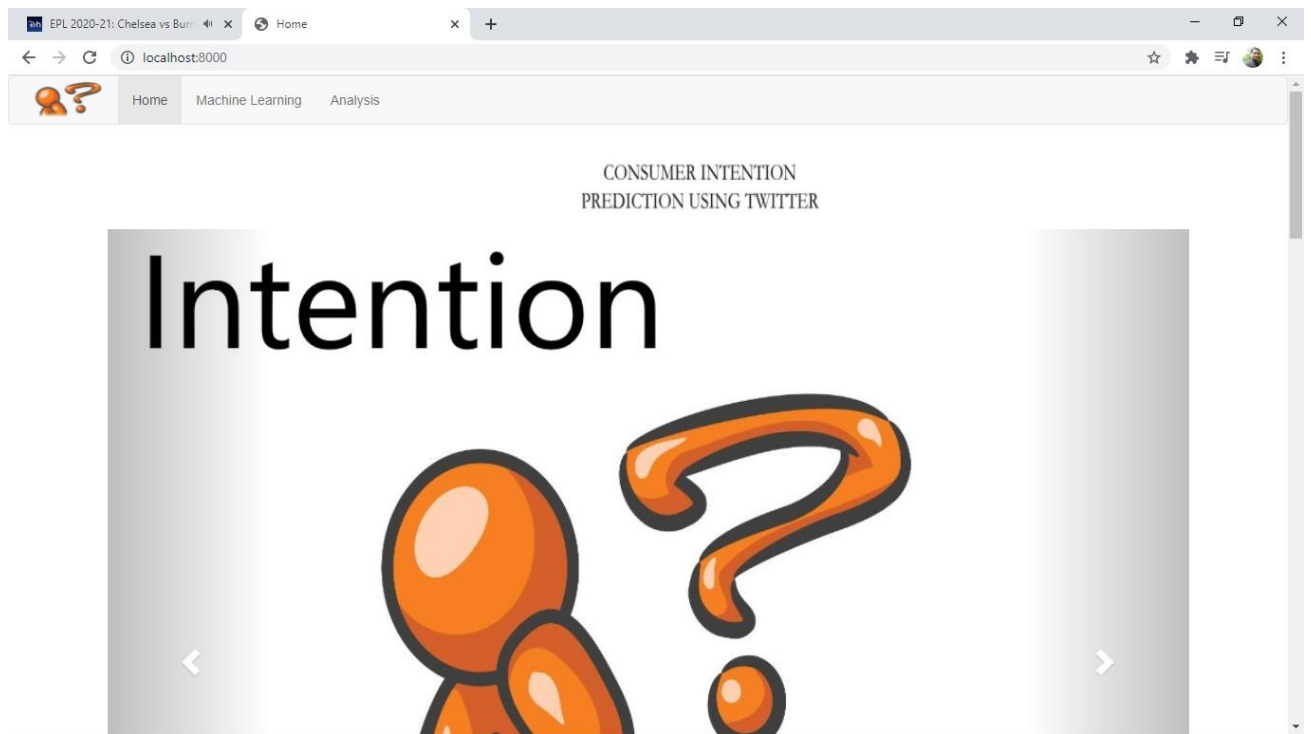


Fig 5.1: Home Page

The Fig 5.1 is the home page of the project. This page will appear once the user or the admin make a log-in into the portal. It is the frontend of the project or it can also be termed as user interface. Here the user the gets the multiple options to execute or access their task as per need.

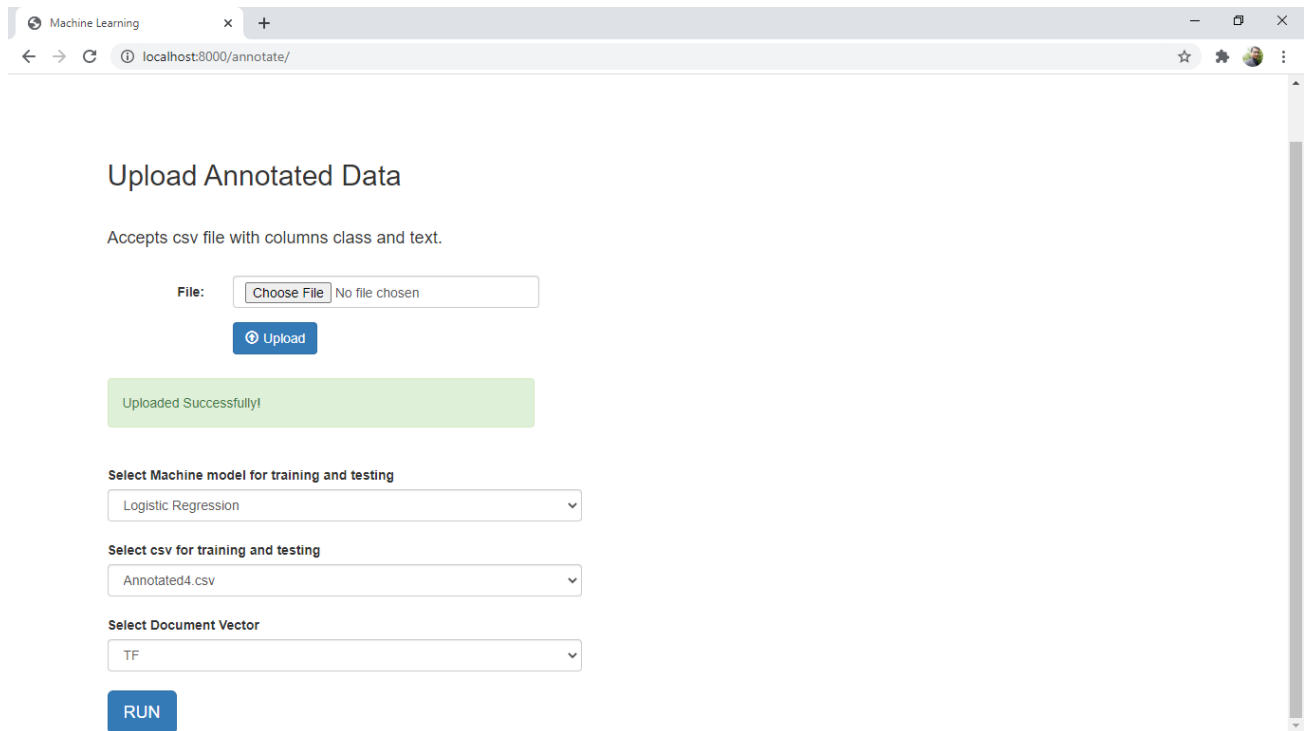


Fig 5.2: upload Annotated data

The Fig 5.2 is the part of the project where user can upload their data . This page will appear once the user or the admin make a log-in into the portal. This page will appear when the given user go through the pages and want to see the result of their product . In order to know more about there product the user must have set of data in the xml format. After uploading the data user have option to check or test the data with four different algorithm . It has five specific sections. The first one is afcose to upload the data, the second one is to select the algorithm which you want to check, the third one is testing data which is the reference of your data and the fourth is document vector

Machine Learning

localhost:8000/annotate/

Upload Annotated Data

Accepts csv file with columns class and text.

File: No file chosen

Uploaded Successfully!

Select Machine model for training and testing

- Logistic Regression
- SVM
- Naive Bayes
- Neural Network
- Decision Tree

Select Document Vector

TF

Fig 5.3: Number of Algorithm

The Fig 5.3 is the annotated data page which basically shows the number of algorithm present in this given project. User can check any of them in order to get the desire answer. Whoever give the best result is the best suited for the product and their respective company.

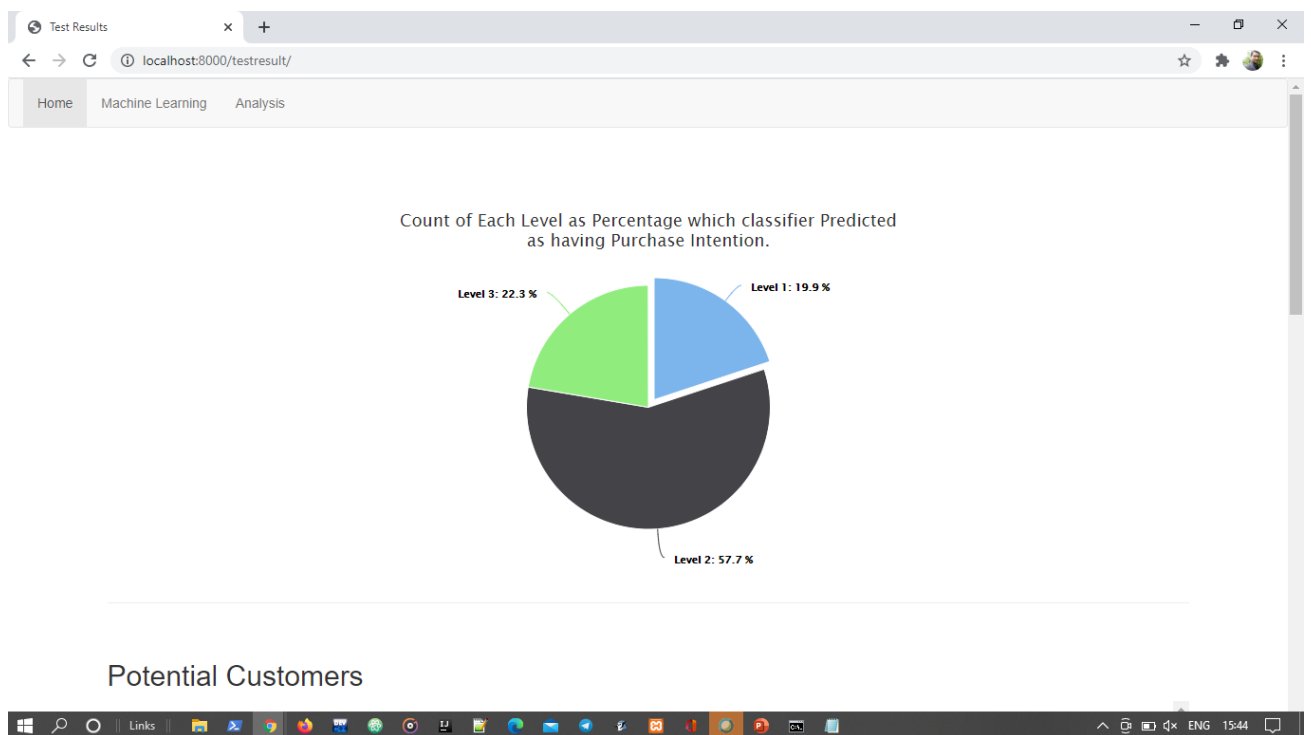


Fig 5.4: Count the percentage of people

The Fig 5.4 shows that the number or percentage of the people who wants or shows the intension to buy the product. It is a bar graph which consist of three level of people and shows the intension of the people in each level. According to their intension user can easily understand whether his product is really wants by people or not. If it likes by a good number of people then the user can easily target them .

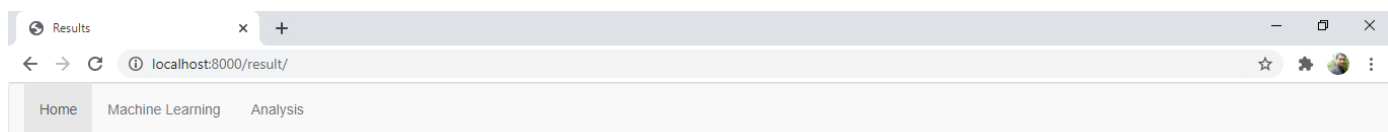


Fig 5.5: Result obtain by svm

The Fig 5.5 is the stats page which is obtain when you want to test your data . After testing the data it shows the accuracy, precision, true negative and true positive values and recall values. Similarly it will shows the value of each and every algorithm which user wants to check.

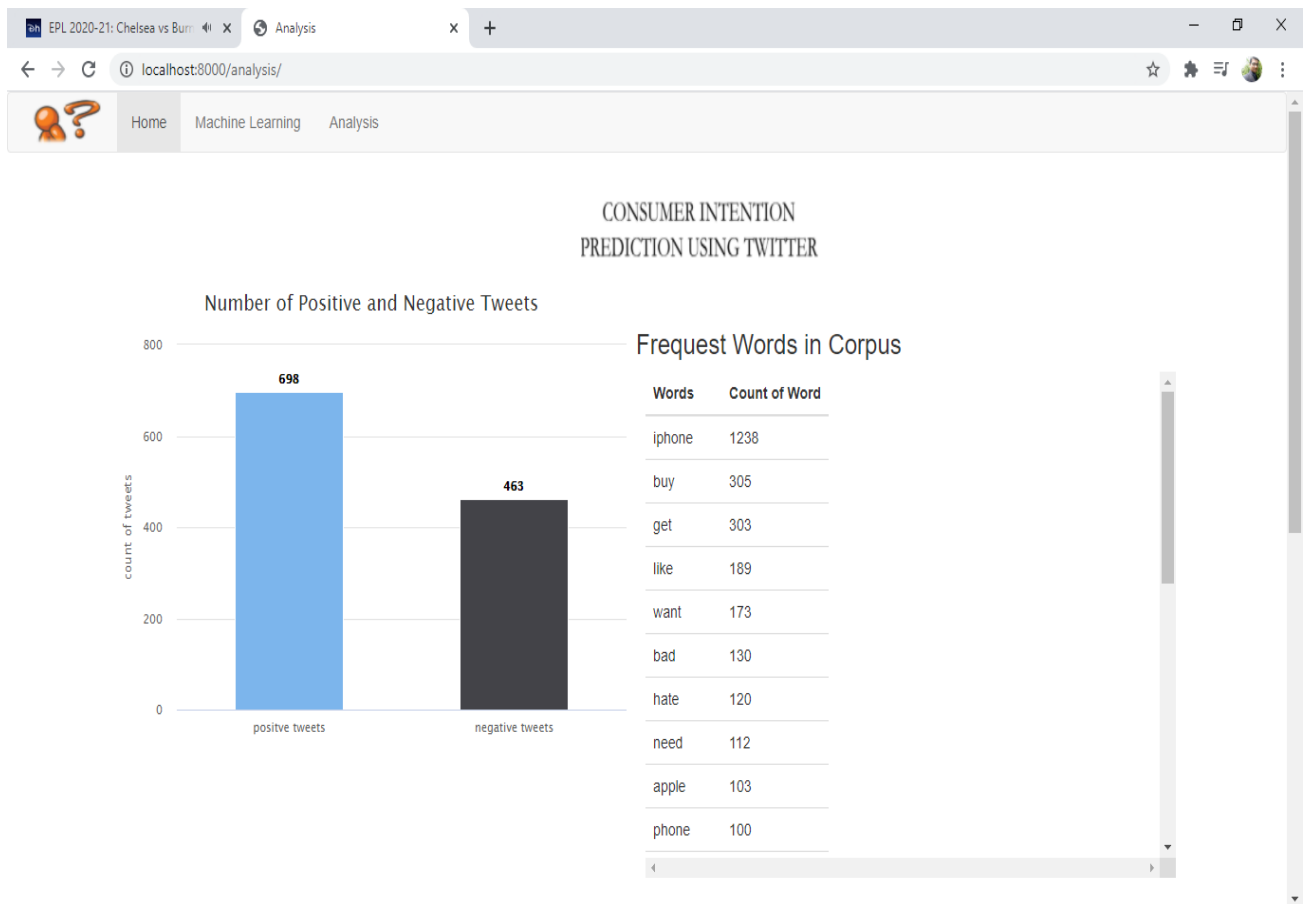


Fig 5.6: Total number of tweets

The Fig 5.6 shows that the total number of tweets which user has whether it's negative or positive . In this linear diagram it shows the different number of positive and negative tweets and it also shows the keywords by which it decide whether the tweets is positive or negative. Like if the tweets consist of buy or good or get or like then it regard as a positive tweets but if it has bad or disgusting or hate or waste then it thinks as a negative .

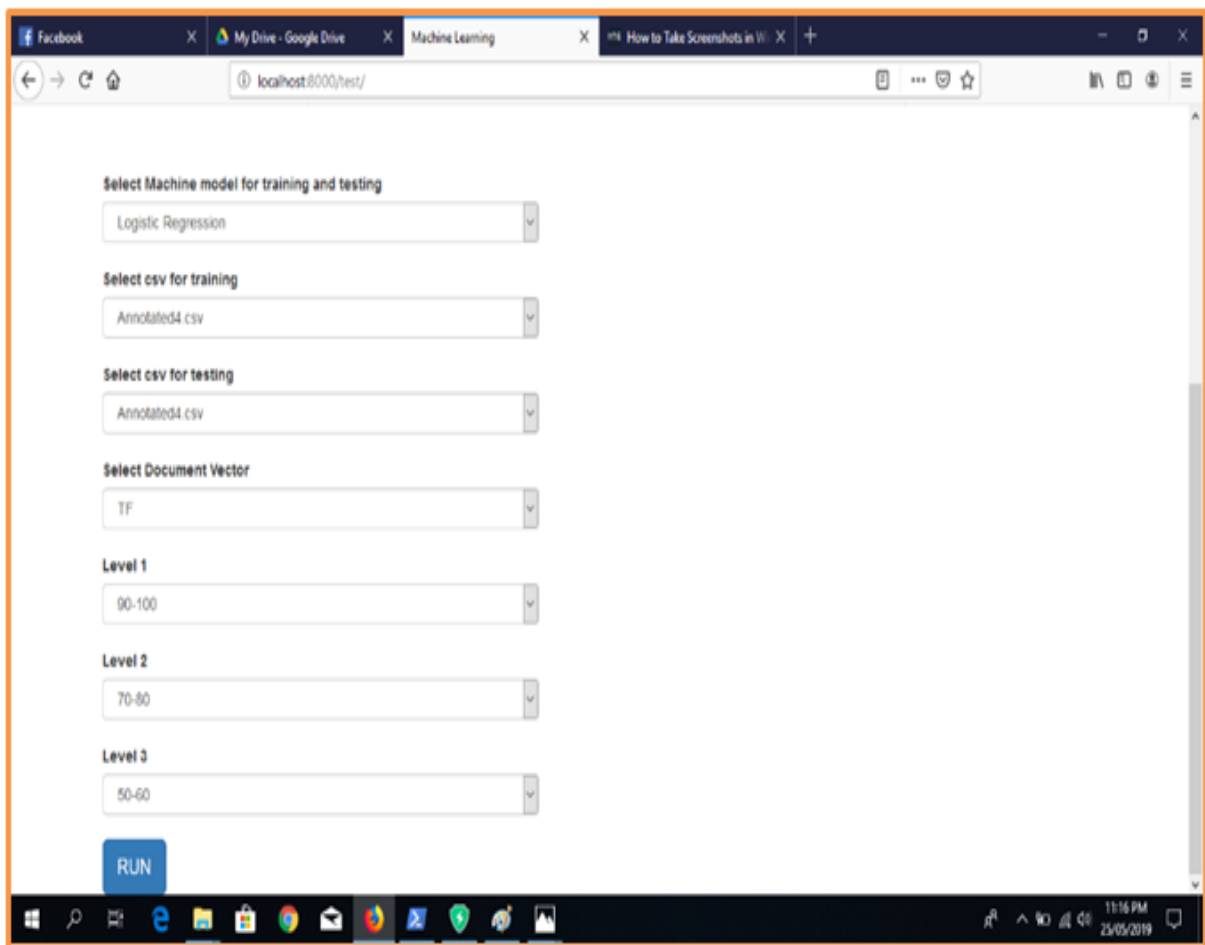


Fig 5.7: Testing page

The Fig 5.7 is the testing page where user can test their data or train the data. It consists of seven layer or you can say seven places where user has to choose what's the best suited for him. The first layer consists of different kinds of algorithm, the second one is used for selecting the csv file in order to train their data, the third one is consist of training the data, we have to upload the data which is present which must be in csv form and upload on the website, the fourth one is of documentation vector, there are three option tf, idf and tf-idf, on the basis of user demand they can select their own preference, the next three column is if level of the tweets and with its accuracy, The user can select the people on the basis of this level. Like if the user wants, he wants the percentage of people whose intension to buy the product is more than ninety percent then he can choose the value between 90 to 100 and move to the next level, if he wants in level two he needs people having intension of 80-90 percent then he can choose those options, if level three he needs the people having intension more than 60 percent he can easily choose those option. it means according to the preference of the user we can select the level of the people according to their intension. It helps the company enormously so that they can easily categorize the people and target them according to that, it simply saves their a lot of time and resources.

#	text	class	Predicted Class	score	Levels
10	@Mnihosetheboss i really wish i get to meet you one day you do such good tech videos i hope i win your iphone x giveaway in that way i maybe able to talk to you big fan of you""BOSS"" i watch every video first @Mnihosetheboss #1000000	yes	yes	0.9925247526020748	Level 1
368	Want to win SENA Cases iPhone Xs Giveaway Enter to win 1st Prize: iPhone X ? i just entered to win and you can too. http://gvwy.io/emu3ld0	yes	yes	0.9920151425428116	Level 1
404	Yeah, so far Apple has done a good job with OLED, i haven't read stuff about iPhone X or earlier Apple Watch having UI elements stuck. i just hope to see Promotion elsewhere (it might get tough in the Mac with most of them using iGPUs though).	yes	yes	0.9883664829887914	Level 1
764	When you purchase the iPhone X but it doesn't come to that store till December 1st so you have to use and 8 till then but idk i kinda like the 8... might have to just get the plus and refund that X	yes	yes	0.9837933047330661	Level 1
265	We finally got to Cali and i had to go go to Best Buy so we are in San Francisco and a homeless man with a stack of at least 5k runs into the Best Buy and day look i have money and throws it on the table and says i need an iPhone X	yes	yes	0.982989068231601	Level 1
798	i just saw someone's iphone x and idk man i think i gotta keep rolling w my 7+ like the screen is the same but i can't play no tiny phone game i'm a grown woman. but we'll see. we'll see.	yes	yes	0.9820500785226293	Level 1
807	i really need to go to sprint and look at the iphone 8 and iphone x and play w / them to see which one i really want cause ive been hearing mixed reviews	yes	yes	0.9796289504916967	Level 1
676	i need a sugar daddy to buy me the new iPhone X for Christmas. Fuck independent woman for now!	yes	yes	0.9791651403343848	Level 1
954	Can anybody with the iPhone X or iPhone 8 Plus DM me and let me know if you like your phone or if you wish	yes	yes	0.9765741209630547	Level 1

Fig 5.8: Showing the potential customer

The Fig 5.8 shows the number of tweets which gives us the idea of the people who wants to buy or who shows their intension to buy the product who also wish or give some positive feedback about the product . After running our data these tweets are divided into three parts , the first part shows the potential customer who shows their intension to buy the product , the second part shows the people who does not like the product and are not happy with the product at all, the third part is consist of the people who are neither too happy nor in deep trouble with the product . It gives the company the clear idea how and when to target those people and what is their strategy to counter those different kind of people having different intension.

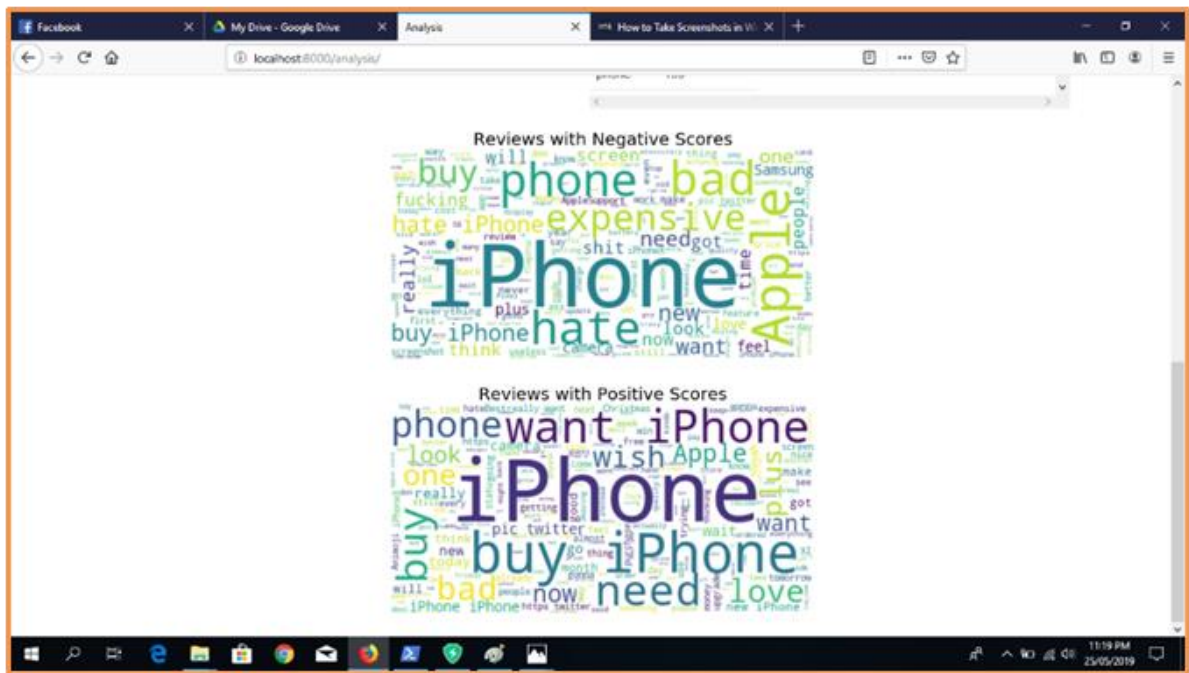


Fig 5.9: Dashboard and Keyword

The Fig 5.9 is all about the dashboard which can consist of keywords and the words which is used in the tweets. It shows the keyword which basically decide the review either belong to positive score or it belongs to negative score. It is the final presentation of the website.

CHAPTER 6

CONCLUSION AND FUTURE WORK

CONCLUSION

Our results were quite promising since we had created our own dataset and were building the model from scratch. We had to create our own dataset because there does not exist a publicly available dataset for purchase intention based on twitter tweets.

The 2 major problems that we faced were:

1. The imbalance class problem: Since our dataset was manually annotated by us, we had about 2000 positive tweets and 1200 negative tweets. Due to this we were getting a very low True Negative Rate and our model was not accurately predicting the negative class.
2. Limited annotated data: Since we had to manual annotate each tweet in the dataset and this process takes a lot of time, we were only able to annotate about 3200 tweets.

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

We were not able to get more than 80% accuracy because of the two problems highlighted above. To achieve even 80% accuracy with an imbalance class data and such a small dataset is a victory.

After showing the website to a few potential clients we have received positive remarks about our product and people seem interest in this new approach to customer identification and targeted marketing.

FUTURE WORK

In future we work on the different kind of product not specially belongs to only one product. We try to access the data directly through the twitter in order to get more and more data and gather bulk of data .To continue our work forward, it is worth trying out the dataset on deep learning models such as RNNs (recurrent neural

networks), convolutional NN, and deep belief networks. Further, we can also use the dataset to find the intention shown towards specific features of the product rather than the product as a whole and target the user towards the specific feature of the product to increase the likeliness to purchase the product.

REFERENCES

Books:

1. Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin.

Inspirations for code and designs:

1. Building a prediction model, <https://www.kaggle.com/gpayen/building-a-prediction-model>
2. Sentiment analysis, <https://www.kaggle.com/laowingkin/amazon-fine-food-review-sentiment-analysis>.
3. TEXT PREPROCESSING USING PYTHON, <https://www.kaggle.com/shashanksai/text-preprocessing-using-python>.

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. The Impact of Social Network Marketing on Consumer Purchase Intention in Pakistan: Consumer Engagement as a Mediator, Asian Journal of Business and Accounting 10(1), 2017.
4. Using Twitter Data to Infer Personal Values of Japanese Consumers, 29th Pacific Asia Conference on Language, Information and Computation pages 480 – 487 Shanghai, China, October 30 - November 1, 2015, Copyright 2015 by Yinjun Hu and Yasuo Tanida.

Datasheets:

1. Amazon Fine Food Reviews, 500,000 food reviews from Amazon, Stanford Network Analysis Project, <https://www.kaggle.com/snap/amazon-fine-food-reviews>