Review Analysis of Products Based on Consumer Needs Using SVM (Support Vector Machine)

Prof.Sachin Wakurdekar
Department of Computer Engineering
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

Arpit Mishra
Department of Computer Engineering
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

Pranav Kumar
Department of Computer Engineering
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

Kshitij Chuahan
Department of Computer Engineering
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

Abstract

As of late, there has been a critical ascent in the web based business industry and all the more explicitly in individuals purchasing items on the web. An ever increasing number of individuals have begun posting on the web about whether they need to purchase the item or finding out if they should purchase the item or not. There has been a ton of exploration being done on sorting out the purchasing behaviors of a client and all the more significantly the components which decide if the client will purchase the item or not. One such stage is Twitter which has gotten very mainstream as of late. In this examination, we will investigate the issue of recognizing and foreseeing the buy aim of a client for an item. Subsequent to applying different content insightful models to tweets information, we have discovered that it is to be sure conceivable to foresee if a client have shown buy aim towards an item or not, and in the wake of doing some examination we have discovered that individuals who had at first shown buy aim towards the item have much of the time likewise purchased the item.

Keywords: AI Approach, Support Vector Machine, Customer Needs, Buyer Goal, Tweet Activities

I. INTRODUCTION

We need to foster an AI approach that will distinguish possible clients for an item by assessing the buy aim in quantifiable terms from tweets on twitter. We have utilized a content logical AI approach on the grounds that despite the fact that text examination can be performed physically, it is wasteful. By utilizing text mining and normal language preparing calculations it will be a lot quicker and effective to discover examples and patterns. In a manner we can say that Purchase Intention discovery task is near the errand of recognizing wishes in item audits.

Buy goals are every now and again estimated and utilized by promoting supervisors as a contribution for choices about new and existing items and administrations. Up till now numerous organizations actually use client overview structures in which they pose inquiries like that you are so liable to purchase an item in a given time period and utilizing that data they ascertain the buy expectation. We need to check whether we can utilize Twitter tweets to prepare a model to distinguish tweets which show buy aim for an item.

A. *Complexity*

The intricacy of our methodology is that we need to figure how to gauge the buy goal from a tweet. Investigating the distinctive sort of text insightful techniques and picking the best one for our errand will be very difficult. Estimating the consequences of our AI model and afterward choosing the best one will include a ton of elements which we should compute.

B. *Motivation*

We need to foster an AI model which can anticipate the mathematical incentive for the buyer goal for a tweet. By doing this we can demonstrate that we online media, for example, Twitter is likewise a significant apparatus which advertisers can utilize when choosing to focus on a client. We accept that our work can be important to applications zeroing in on misusing buy expectations from web-based media.

C. *Challenges*

The main test we confronted was that we couldn't track down any open dataset with respect to buy expectation. We needed to scrap the information from Twitter utilizing a web scrubber. Also, since we ourselves accumulated the information we needed to physically comment on the tweets. Once more, this cycle was amazingly tedious as we needed to go through each tweet and choose the buy goal. Thirdly, we had restricted commented on information due to the protracted cycle of manual comment and time requirement.

D. *Organization of the paper*

The remainder of this paper is coordinated as follows:
We survey related work on buy aim and web based purchasing conduct in Section 2. In Section 3, we clarify our information assortment and comment measure, trailed by model creation. In Section 4, we present the tests and their outcomes. At last, Section 5 finishes up the paper and gives the extent of future work.

II. LITERARY REVIEW

There have been a few exploration reads for examining the bits of knowledge of online shoppers purchasing conduct. In any case, a couple have tended to the clients purchasing expectation for items. Studies on recognizable proof of wishes from messages, explicitly Ramanand et al. (Ramanand, Bhavsar, and Pedanekar 2010) consider the errand of recognizing 'purchase' wishes from item surveys. These desires incorporate ideas for an item or a longing to purchase an item. They utilized etymological standards to distinguish these two sorts of wishes. In spite of the fact that standard based methodologies for recognizing the desires are viable, yet their inclusion isn't good, and they can't be expanded without any problem. Buy Intention recognition task is near the errand of recognizing wishes in item audits. Here we don't utilize the standard based methodology, yet we present an AI approach with conventional highlights extricated from the tweets.

Past investigations have shown that it is feasible to apply Natural Language Processing (NLP) and Named Entity Recognition (NER) to tweets (Li et al., 2012) (Liu et al., 2011). Nonetheless, applying NER to tweets is extremely challenging in light of the fact that individuals regularly use shortened forms or (intentional) incorrectly spelled words and syntactic blunders in tweets. In any case, Finin et al. (2010) attempted to clarify named substances in tweets utilizing publicly supporting. Different investigations utilized these methods to apply assessment examination to tweets. The primary investigations utilized item or film surveys in light of the fact that these audits are either certain or negative. Wang et al. (2011) and Anta et al. (2013) dissected the feeling of tweets sifted on a certain hashtag (watchwords or expressions beginning with the image that mean the principle subject of a tweet). These examinations simply dissect the notion of a tweet about an item after the creator has gotten it. We will anyway be separating highlights from tweets to discover whether the client has shown buy aim towards the item or not.

All the more as of late, 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) explore if a man-made reasoning methodology can anticipate (from existing client made substance on twitter) in case somebody is an expected client for a particular organization or item and recognize clients at various phases of the choice interaction of purchasing a given item. Further seeing examination reports like 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) give us an understanding of the effect of interpersonal organization promoting on buyer buy goal and how it is influenced by the interceding job of customer commitment. In light of UGT hypothesis (Uses and Gratification Theory).

Some preprocessing methods commanly utilized for twitter information are the sentiment140 API (Sentiment140 permits you to find the feeling of a brand, item, or theme on Twitter), the TweetNLP library (a tokenizer, a grammatical form tagger, various leveled word bunches, and a reliance parser for tweets), unigrams, bigrams and stemming. There are additionally some word reference based methodologies like utilizing the textBlob library (TextBlob is a Python (2 and 3) library for preparing literary information. It gives a reliable API to jumping into normal regular language preparing (NLP) assignments, for example, grammatical form labeling, thing phrase extraction, feeling investigation, and the sky is the limit from there).

The normal AI calculations that are utilized for text examination are Linear Regression, Random Forest, Naive Bayes and Support Vector Machine. We will be taking a gander at these models later exhaustively.

A. *Model Portrayal*

III. PROPOSED APPROACH

In this segment, we depict the subtleties of our way to deal with tackle the issue of procurement aim recognition. We will start by portraying our information assortment and explanation measure. Then, at that point we will portray our methodology for information preprocessing and changing the information to prepare text logical models.

A. *Data Assortment and Explanation*

As there are no explained Twitter tweets corpora accessible freely for recognition of procurement purpose, we needed to make our own. This was finished utilizing a web crawler created by JohnBakerFish which slithered the site to gather the information. We had gathered more than 100,000 tweets however since they were not commented on, we needed to chop down to only 3200 tweets which were arbitrarily chosen out of the dataset and we physically explained them utilizing an essential rule we had characterized:

Measures for Labeling 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 feature so of iphone x?*	*No PI*
4	*liked video on Youtube about iphone x?*	*PI*

We utilized only 3200 tweets out of an enormous dataset as we were restricted by time. We characterized meaning of Purchase Intention as item that is having activity word like (purchase, need, want) related with it. Each tweet was perused by 3 individuals and last class was chosen by most extreme democratic.

B. Data readiness

1) Data preprocessing strategies:

We prepared the tweets utilizing these methods in sequential request. To begin with, we began our foundation by changing over our content into lower case, to get case consistency. Then, at that point we passed that lower case text to accentuations and extraordinary characters evacuation work. Text might contain undesirable unique characters, spaces, tabs and etcetera which has no huge use in text arrangement. Following stage was stop words evacuation since the tweets additionally contains futile words which are standard piece of the sentence and language however don't add to the importance of the sentence. Preferences of "the", "a", "an", "in" and etcetera are the words referenced previously. In this way, we needn't bother with these words, and it is smarter to eliminate these. Further we likewise eliminated the best 2 most normal words in light of the fact that their repeat doesn't add to the significance in the sentence. This can likewise be the aftereffect of error as the information we are examining is a casual information where formal sentence standards are not thought about. We likewise eliminated some uncommon words like names, brand words (not iphone x), left out html labels and so forth These are exceptional words which don't contribute a lot to translation in the model. At last we stemmed the words to their root. Stemming works like by cutting the end or start of the word, considering the normal prefixes or additions that can be found in that word. For our motivation, we utilized Porters Stemmer, which is accessible with NLTK. We additionally tried different things with lemmatization. The examination is acted in morphological request. A word is followed back to its lemma, and lemma is returned as the yield. Yet, it didn't yield a significant change in the corpus.

In the wake of preprocessing the tweets, we were left with around 1300 tweets for preparing information and staying for testing.

C. Formation of Document Vector

We made 3 sorts of archive vectors with the end goal of experimentation. To begin with, is the term recurrence record vector? We have put away content and its named class in information outline. Furthermore, we have developed another information outline with segments as the words and record consider the columns. In this way, singular recurrence of words in a report tally is recorded. Second, is the opposite archive recurrence vector which is a weighting technique to recover data from the record. Term recurrence and reverse record recurrence scores determined and afterward result of TF*IDF is called TF-IDF. IDF is significant in discovering how applicable a word is. Regularly words like 'is', 'the', 'and' and so on have more noteworthy TF.

Accuracy Table

	Naive Bayes	Logistic Regression	Support Vector Machine	Decision Tree	Artificial Neural Network
TF + neg handling + kfold	75.2	76.9	74	69	74.2
TF-IDF + neg handling + kfold	70.2	74.4	77.7	70.4	67.8
TF + neg handling + lemmatization + kfold	75.4	77.4	74.4	70.9	72.7
TF-IDF + neg handling + lemmitization + kfold	69.6	72.8	75.9	70.4	73.7
TF + lemmitization	75.6	76.9	73.6	73.6	71.3
TF-IDF + lemmitization	73.9	74.2	79.2	69.3	73.6

For our second attempt after reorganizing the data preprocessing steps and adding code for negation handling, we got these results:

True Negative Rate

	Naive Bayes	Logistic Regression	Support Vector Machine	Decision Tree	Artificial Neural Network
TF + neg handling + kfold	45.6	47	48.6	48.6	51
TF-IDF + neg handling + kfold	11.4	26.9	49.1	46.2	0
TF + neg handling + lemmatization + kfold	43.3	47.6	48.3	51.3	51
TF-IDF + neg handling + lemmitization + kfold	11.4	24.9	46	52.7	49.3
TF + lemmitization	49.4	46	47.1	57.5	51.7
TF-IDF + lemmitization	13.8	24.1	46	47.1	52.9

Accuracy Table

	Naive Bayes	Logistic Regressio	Support Vector Machin	Decision Tree	Artificial Neural Networ	Naive Bayes CUSTOM
TF	78.2	80.2	80.5	69.3	76	79.4
TF-IDF	65.6	78.2	78.2	72.3	77.6	76.7
binary doc	77.5	80.8	80.2	72.6	78.9	79.4
text-blob + TF		79.5	78.5	66	75.2	72.7
text-blob + TF-IDF		78.9	76.9	69.6	75.6	70.75
text-blob + binary doc		79.5	78.5	72.3	79.2	73.12

True Negative Rate

	Naive Bayes	Logistic Regressio	Support Vector Machin	Decision Tree	Artificial Neural Networ	Naive Bayes CUSTOM
TF	32.8	29.7	42.2	45.3	43.8	46
TF-IDF	48.4	37.5	48.4	46.9	46.9	52
binary doc	28.1	32.8	45.3	48.4	46.9	46
text-blob + TF		31.2	39.5	54.7	40.6	48
text-blob + TF-IDF		40.6	43.7	51.6	50	54
text-blob + binary doc		31.2	39	48.4	32.8	50

Precision

	Naive Bayes	Logistic Regressio	Support Vector Machin	Decision Tree	Artificial Neural Networ	Naive Bayes CUSTOM
TF	83.4	83.2	85.4	83.8	84.9	86.8
TF-IDF	83.5	84.2	86.2	84.7	85.8	87.5
binary doc	82.5	83.8	85.9	85.1	86	86.8
text-blob + TF		83.4	83.9	85	84.2	86
text-blob + TF-IDF		84.8	85	85.2	86	86.85
text-blob + binary doc		83.4	84.5	85	83.6	86.48

Recall

	Naive Bayes	Logistic Regressio	Support Vector Machin	Decision Tree	Artificial Neural Networ	Naive Bayes CUSTOM
TF	90.3	93.7	90.8	75.7	84.5	87.7
TF-IDF	70.3	89.1	86.2	79.1	85.8	82.8
binary doc	90.7	93.7	89.5	79.1	87.5	87.7
text-blob + TF		92.5	89.9	69	84.5	78.8
text-blob + TF-IDF		89.1	85.8	74.5	82.4	74.87
text-blob + binary doc		92.4	89.1	78.6	91.6	78.81

IV. Conclusion

Our outcomes were very encouraging since we had made our own dataset and were building the model without any preparation. We needed to make our own dataset in light of the fact that there doesn't exist an openly accessible dataset for buy goal dependent on twitter tweets.

The 2 significant issues that we confronted were:

1) The irregularity class issue: Since our dataset was physically clarified by us, we had around 2000 positive tweets and 1200 negative tweets. Because of this we were getting an exceptionally low True Negative Rate and our model was not precisely foreseeing the negative class.
2) Limited commented on information: Since we needed to manual comment on each tweet in the dataset and this interaction requires some investment, we were simply ready to explain around 3200 tweets.

Taking a gander at the other investigates that are done in the comparative field, our venture additionally stands separated since we have executed 5 distinct models and subsequent to assessing them, we pick the best one altered to the item information.

We couldn't get over 80% precision on account of the two issues featured previously. To accomplish even 80% exactness with an irregularity class information and a little dataset is a triumph.

V. Future work

To proceed with our work forward, it merits giving a shot the dataset on profound learning models like RNNs (intermittent neural organizations), convolutional NN, and profound conviction organizations. Further, we can likewise utilize the dataset to discover the goal displayed towards explicit highlights of the item as opposed to the item all in all and focus on the client towards the particular component of the item to expand the likeliness to buy the item.

REFERENCES

A. *Books:*

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

B. *Inspirations for code and designs:*

[1] Building a pridction model for the given data we found https://www.kaggle.com/gpayen/building-a- prediction-model
[2] Sentiment analysis of the given data is found from, 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.

C. *Relevant Papers:*

[1] Identifying Purchase Intentions by Extracting Information from Tweets, February 8, 2017, RAD BOUDU 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.

D. *Websites:*

[1] https://www.kaggle.com/snap/amazon-fine-food-reviews
[2] https://scikit-learn.org/stable/