

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 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 applying various text analytical models to tweets data, we have found that it is indeed possible to predict if a user have shown purchase intention towards a product or not, and after doing some analysis we have found that people who had initially shown purchase intention towards the product have in most cases also bought the product.

Introduction:

Our project is a web application that predicts 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 a particular customer more efficiently and boost their sales.

First, we search for Twitter tweets of potential customers wanting to buy a product. And based on those tweets we estimate/predict the likelihood that the customer will buy the product. We then make a model by gathering tweets from users who have already expressed intention to buy the product using their tweet history and if possible, their web search history as well and then training the text analytical model based on those tweets. Using the model, we input potential customers who have tweeted about the product but have not bought it. And based on the training data the model estimates a prediction/likelihood of whether the customer will buy it or not. We have limited the scope of our work to only mobile phones. Our model predicts the consumer intention for the latest upcoming mobile phones. We will be testing it on the latest iPhone X variants and check with its accuracies.

Rationale:

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.

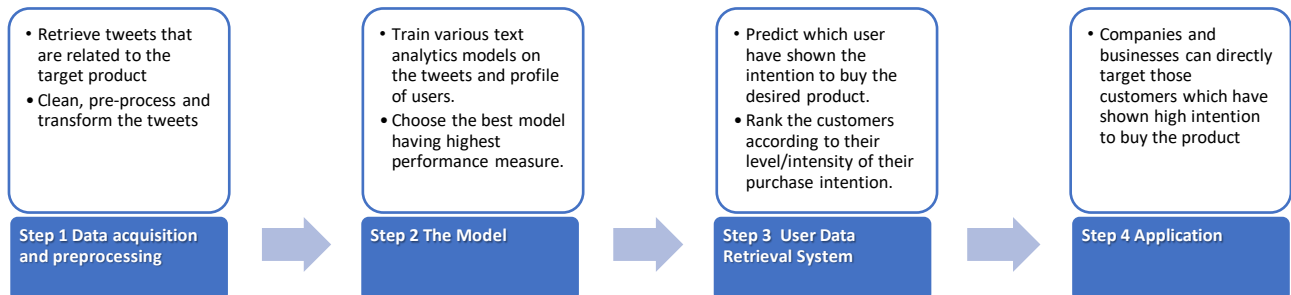
Objectives:

- To create Web application we need to create various Dashboards.
- To update or to test the data we need to have the annotated Data.
- The annotated data is the collect the twitter data in the form of csv file.
- based on the training data, the model will estimate a prediction/likelihood of whether the customer will buy it or not.

Literature Review:

[1] **Our inference out of first paper:** Explains about Social platforms like Twitter and Facebook allow users to share and publish information with their users. In addition to this, users use them to answer very precise and highly contextualized queries, or queries for which the relevant content has not been authored yet, e.g., asking about a conference event using its hashtag on Twitter. Even if the profile is private, we can still derive some attributes. Also, when we tried as an external user to extract profiles from different online social network e.g. Twitter and Facebook , either no data or minimal data was made available

[2] **Our inference out of second paper:** The common machine learning algorithms that are used for text analysis are Linear Regression, Random Forest, Naive Bayes and Support Vector Machine. Applying Named Entity Recognition (NER) and Natural Language Processing (NLP) to tweets is very difficult because people often use abbreviations or deliberately misspelled words and grammatical errors in tweets.

Methodology: Steps involved in User Interest Based Social Media Data Retrieval System**Evaluation:**

To evaluate models, we will be using the following techniques:

1. Confusion Matrix
2. Accuracy
3. Precision
4. Recall
5. F-Measure
6. True Negative Rate

Software Requirements:

1. Python 3.6
2. Python Django Framework
3. Mongo DB
4. scikit-learn library for Python
5. Internet browser eg: Google Chrome
6. A Python code editor

References:

1. Data Retrieval from Online Social Network Profiles for Social Engineering Applications, Sophia Alim, Ruqaya Abdul-Rahman, Daniel Neagu and Mick Ridley *Department of Computing, University of Bradford, BD7 1DP {S.Alim, R.S.H Abdul-Rahman, D.Neagu, M.J.Ridley}@bradford.ac.uk*.
2. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms, *Information Systems* 56(2016)1–18.