Twitter Sentiment Analysis Applied to Customer Loyalty Prediction

In recent days, the popularity of social media is increasing significantly. More and more people share their ideas, information, and opinions with others using social media. Understanding of people opinion present in social media has tremendous applicability. The business world wants to take advantage of information shared by the people on social media. The sentiment analysis can be used to estimate the emotions of the users objectively. The research focuses on Twitter sentiment analysis and application of the sentiment information for customer loyalty prediction. The service provider wants to promote loyal the customer at the same time to retain the loyal customers. The airline service is considered for the study in this research work and tweets related to airlines gathered from Twitter. Four region airlines are selected which are European, Indian, American and Australian. Tweepy a package from Python is used to collect the airline tweets. The airline services have help desk twitter handle, which is used to collect the tweets. In some case, the official Twitter handle is also used to gather the tweets. The twitter sentiment analysis is performed using TextBlob, which provides the sentiment score for the tweets.

Airline Consumer Loyalty

There are two types of tweets collected using the search term such as "left airline" and "loyal to airline" or "loyal flyer". People who left the airline service have tweeted with the term "left airline" and they are not loyal to airlines. Tweets with terms and "loyal to airline" or "loyal flyer" are consumers who are loyal to airline services and continue taking services from the airlines.

Using these terms around 10,000 tweets are collected specifically to the airline service. It beneficial to know whether a consumer is loyal or not to an airline service. The machine learning techniques can be used to predict whether a consumer is loyal or not to a service based on the twitter data as shown in figure 1.
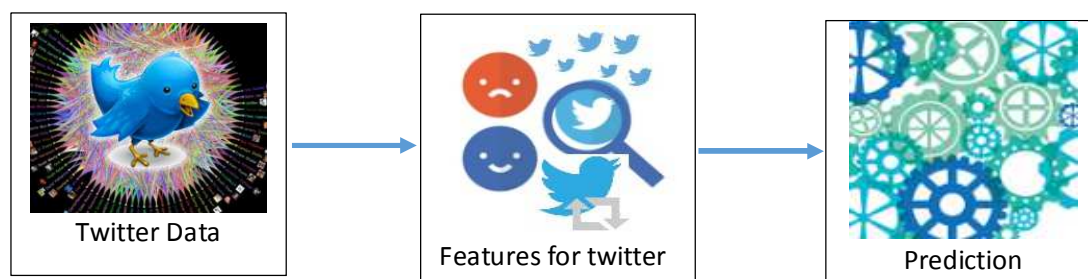

Figure 1. Twitter data prediction

Twitter Features for Loyalty Prediction

The Twitter data has been used to predict whether a consumer is loyal or not. Feature analysis and selection play a vital in machine learning. Feature analysis is carried out using 524 not-loyal tweets and 524 loyal tweets out of 10,000 tweets. Sentiment analysis on the tweets is performed and the positive and negative

sentiment scores are collected. The positive and negative sentiment score for each category of not-loyal and loyal are visualized using scatter plot. Figure 2 shows the distribution of not-loyal and loyal tweets against the positive and negative sentiment score. Along x-axis positive sentiment score and along y-axis negative sentiment score are shown. Tweets are color-coded with yellow for not-loyal and purple for loyal.
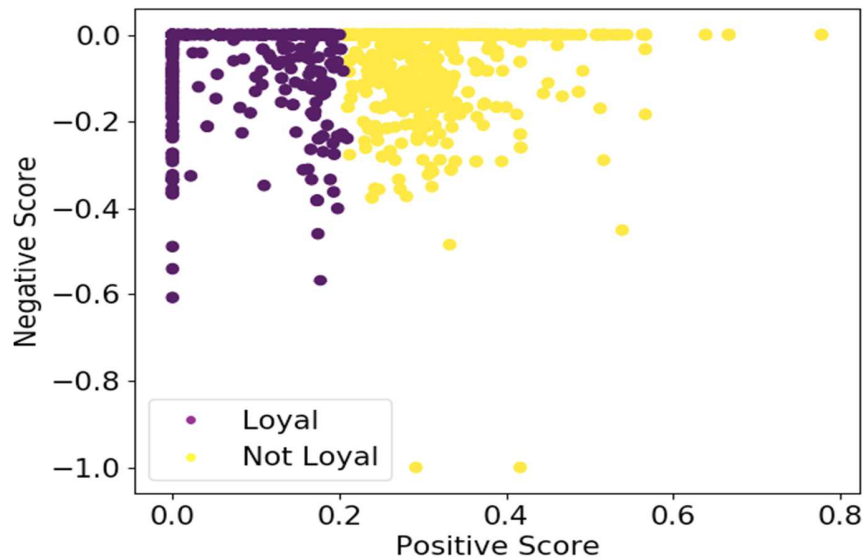

Figure 2. Distribution of tweet with respect positive and negative sentiment score.

The number of followers for the user are collected from the Twitter account. The following figure 3 shows the distribution of not-loyal and loyal tweets with respect to positive sentiment score (y-axis) and the number of followers (x-axis).
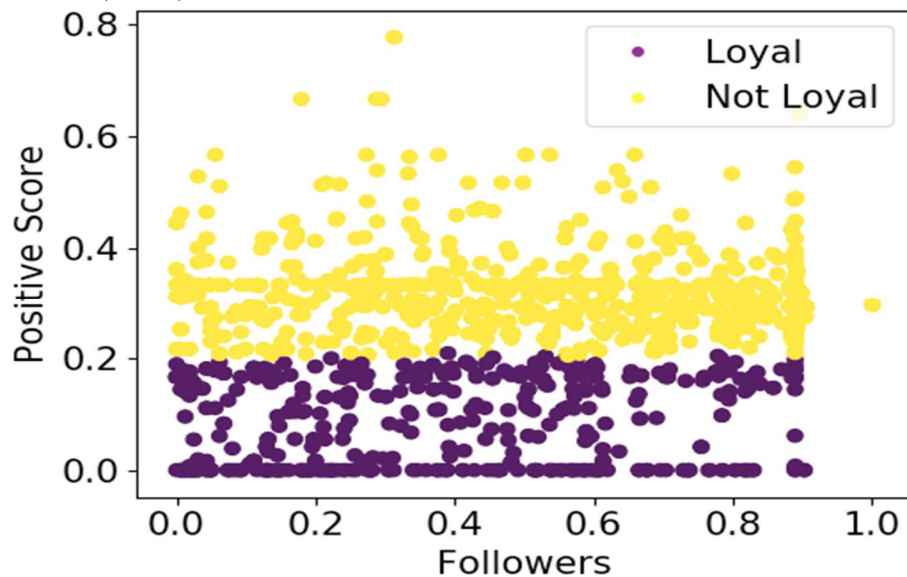

Figure 3. Distribution of tweet with respect positive and the number of followers.

A 3D visualization of the positive, negative sentiment score and the number of followers is depicted in figure 4. Along x-axis, the positive sentiment score, along y-axis the negative sentiment score and along z-

axis the number of followers are shown. This 3D scatter plot shows each point as a tweet either loyal or not-loyal category. The 3D view reveals vital information about the influence of the passengers. This graph indicates based on their followers and the sentiment scores where they stand and how much they can influence to the other passengers. Identifying the influential passengers in an airline service is important for market perspectives.
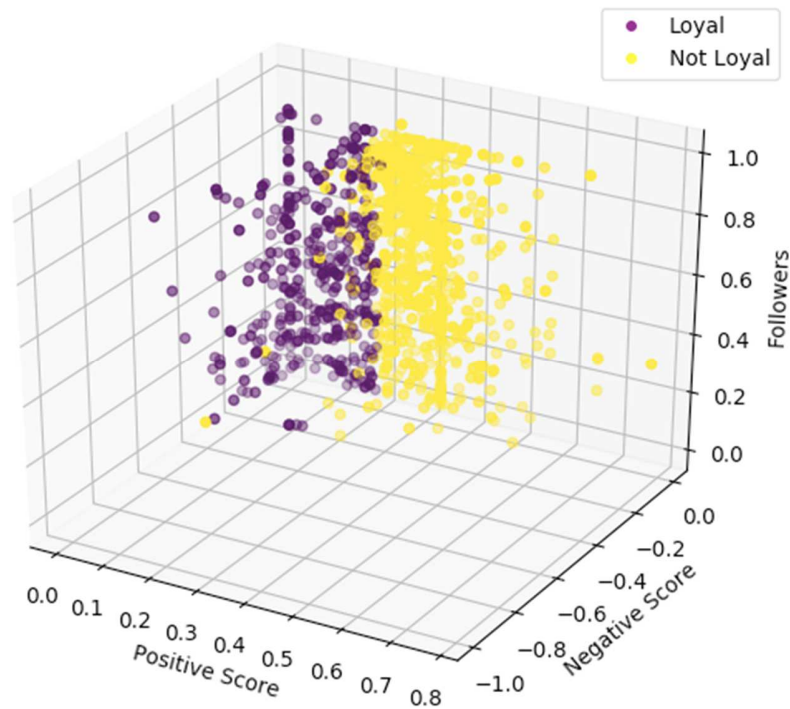

Figure 4. 3D scatter plot for the not-loyal and loyal customers.

Loyalty Prediction

The machine learning techniques can be utilized to predict whether a consumer is loyal or not. The Twitter-related information is used as the features such as the positive sentiment score, negative sentiment score, mean of the retweet, mean of likes and the number of followers. Three different classifiers are used such as Random Forest classifier, Decision Tree classifier, and Logistic Regression classifier. The cross-validation is performed on 10,000 tweets and the results are given in the below table.

Model	Accuracy	Misclassification
Random Forest	99.05	0.95
Decision Tree	98.97	1.03
Logistic Regression	91.10	8.90

Research Paper

The research on consumer loyalty evaluation using twitter data is appeared in:

http://thesai.org/Publications/ViewPaper?Volume=9&Issue=6&Code=IJACSA&SerialNo=52

Cite the Work

Please cite the following research paper:

Rida Khan, Siddhaling Urologin, "Airline Sentiment Visualization, Consumer Loyalty Measurement and Prediction Using Twitter Data", in International Journal of Advanced Computer Science and Applications (IJACSA), volume 9, issue 6, pp. 380-388, 2018.

Further Projects and Contact

For further reading and other projects please visit

Dr. Siddhaling Urolagin,
Department of Computer Science
BITS Pilani, Dubai Campus, Academic City
Dubai, United Arab Emirates
Email: siddhaling@dubai.bits-pilani.ac.in
www.researchreader.com
siddesh_u@yahoo.com