

# Brand Listening using Twitter Sentiment Analysis

Bundidth Dangsawang

Dept. Information Technology

Faculty of Information Technology

King Mongkut's University of Technology North Bangkok

Bangkok, Thailand

bundidth\_d@hotmail.com

Tiphathai Thongthammachart

Dept. Information Technology

Faculty of Information Technology

King Mongkut's University of Technology North Bangkok

Bangkok, Thailand

tipthong@gmail.com

Nitigan Nakjuatong

Dept. Information Technology

Faculty of Information Technology

King Mongkut's University of Technology North Bangkok

Bangkok, Thailand

nitigan.n@fitm.kmutnb.ac.th

Maleerat Sodanil

Dept. Information Technology

Faculty of Information Technology

King Mongkut's University of Technology North Bangkok

Bangkok, Thailand

maleerat.s@it.kmutnb.ac.th

**Abstract**— A text mining technique toward sentiment analysis on Twitter users to sneaker by considering from #hashtag can explain the brand popularity, sneaker model that has impact on buyer's buying decision process or distribution decision of the distributor. A text mining has 3 steps, namely: firstly, it is the brand popularity consideration which examining from number of tweets that relevant to sneaker #hashtag in which resulting in 1,553 tweets of 20 sneaker brands and 799 tweets when excluding the missing value; and pursuant to the stated result, it can be concluded that the most popular brand is Nike with 34% of tweets which beating Adidas that has 26% of tweets, secondly, using text mining to explore the brand satisfaction—both Nike and Adidas, and lastly utilizing sentiment analysis with Alien Text Analysis as a tool in order to extract the opinions from 1,000 tweets in which the outcome is divided into 2 polarities, namely: positive and negative. The abovesaid outcome can suggest the satisfaction trends of Twitter users toward the top 2 most popular brands using SVM, which provides an accuracy of 73.91%.

**Keywords**—Twitter ; Popularity ; Text Sentiment

## I. Introduction

Currently, internet enables people to easily communicate and connect including share information to others around the world, particularly social media sites that users can exchange and choose to consume news that they wanted to hear as well as sharing such news to other users whom are in the same network. As a result, the social network sites are turn into an important and very effective tool for marketing and public relations purposes since the strength point of social network sites are spreading the messages rapidly the to the target group, especially the popular microblog called Twitter that users can sharing their messages to many people in such a short time. Most Twitter users use #hashtag in order to follow the news and information that they are interested in since such hashtag symbol (#) is a tool for searching latest information and news such as tourism, artist, actor and actress, currently popular drama or tv series including the breaking news that later become talk of the town. According to the foregoing benefit of Twitter, many sellers and distributors use Twitter platform to connect them with customers, widely, so that customers can conveniently search promotional campaign or

make pre-order. It can be said that such method effectively influences customer's decision-making process and can create word of mouth that impact on other customer buying decision as well. Furthermore, user can retweet some message for sharing that message to other user, and when user needs product information, user can post question on seller account about the product. Currently, since sneaker become fashion trends worldwide and most Twitter users want to know about the most popular sneakers information, therefore the online stores use relevant #hashtag as a tool for attract the attention of the consumers, such as name of popular brands, sneaker model, in order to increase their sales, and likewise the consumers. According to analyzing the opinions from Twitter posts by using data mining process [1-4], it indicates the popularity trends and the customer opinion to brands which such result can become the deciding factor of purchasing and selling for buyer and seller respectively. This research creates a model for analyzing the sneaker trends and the reason for sneaker shopping of brands such as Nike and Adidas by extracting Twitter posts with classification and text mining techniques. In addition, this research aims at studying and collecting data for both popularity analysis and customer feedback analysis from Twitter users on sneakers.

## II. Literature review

### A. Naïve Bayes

Naïve Bayes [5,6] technique is based on Bayes' Theorem which is the statistic theory that using probability to evaluate the uncertain data and turn such data into numbers. It explains the probability of an event (A) occurrence when another event (B) occurred in which can describe in easy equation called equation 1 [7] as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

where  $P(A|B)$  is the probability of an event A occurrence if an event B occurred

$P(B|A)$  is probability of an event B occurrence if an event A occurred

$P(A)$  is the probability of an event A

$P(B)$  is the probability of an event B

Categorizing by using probability principle is called classification method. It can predict and explain the result including analyzing the relationship between variables in order to create the conditional probability of each relationship. The classification algorithms deemed one of an effective method for classifying the data and it performs well working with text classification. The process is not complicated which suitable for multiple classes and without any functional dependency by giving the probability of the given data as describes in equation 2 [8] as follows:

$$P(A_1, A_2, \dots, A_n | C_j) = \prod_{i=1}^n P(A_i | C_j) \quad (2)$$

Group  $C_j$  for data with attribute  $n$   $X = \{A_1, A_2, \dots, A_n\}$  or using symbol as  $P = (A_1, A_2, \dots, A_n | C_j)$  with  $\prod$  means the product of  $P = \text{all } (A_i | C_j)$ ,  $i = 1, 2, 3, \dots, n$  and  $j = 1, 2, 3, \dots, n$  and thus the Bayesian Classification will be as equation 3 as follows:

$$V_{NB} = \text{argmax} P(C_j) \prod_{i=1}^n P(A_i | C_j) \quad (3)$$

#### B. SVM (Support Vector Machine)

SVM technique is one of machine learning algorithm which classifying mixed data and making decision boundary for categorizing data into two classes by drawing a line or a hyperplane with maximum distance separation and the support vector machine will use the mapping function to transform data from the input space to feature space and then create Kernel function on feature space in order to measure the similarity of the data. Generally, the examples of Kernel function are Linear, Polynomail Kernel, Radial Basic Function and Sigmoid [9]

#### C. Related Research

Ankit and Nabizath Saleena [10] Twitter Sentiment Analysis is the way of identifying sentiments and opinions in tweets. The main computational steps in this process are determining the polarity or sentiment of the tweet and then categorizing them into the positive tweet or negative tweet. The primary issue with Twitter sentiment analysis is the identification of the most suitable sentiment classifier that can correctly classify the tweets. Generally, base classification technique like Naive Bayes classifier, Random Forest classifier, SVMS and Logistic Regression are being used. In this paper, an ensemble classifier has been proposed that combines the base learning classifier to form a single classifier, with an aim of improving the performance and accuracy of sentiment classification technique. The results show that the proposed ensemble classifier performs better than stand-alone classifiers and majority voting ensemble classifier. In addition, the role of data pre-processing and feature representation in sentiment classification technique is also explored as part of this work.

### III. Research Methodology

This research aims at studying and collecting data for popularity analysis from Twitter users on sneakers which such data is unstructured data by analyzing the frequency of

mention toward sneaker brands and extracting the Twitter post with sneaker #hashtag. Then analyzing data and interpreting result into tweet frequency which enable the researcher to analyze the brand popularity. The researcher collects information of the 2 most popular sneaker brands for using text mining based on the analysis of Twitter users opinion toward both brands with text sentiment analysis in order to provide the result into 3 polarities, namely: positive, neutral, and negative by utilizing Aylien Text Analysis as a tool so that it results in doing text sentiment from Twitter posts.



Fig. 1. The example of Tweet with sneaker #hashtag

Nowadays, sneaker is very popular in Thailand because there is great demand from buyers and thus sellers gaining the benefit from Twitter platform by using it as a marketing channel and communication channel with buyers. The data from Twitter is unstructured data which this paper classified such data into two classes, namely: 1) determining the popularity from frequency by extracting the data from Twitter with sneaker #hashtag in which resulting in 1,553 tweets or 799 tweets when considering tweet frequency or merely from tweets with brand name therewith and 2) doing text sentiment in order to recognize the opinion of Twitter users toward sneaker brands. The researcher analyzes the opinion by choosing 2 most popular sneaker brands in order to use #hashtag with sneaker brands so that it will allow researcher to extract the data from Twitter for analysis purpose.

#### A. Data Preparation

According to the data from Twitter, it indicates that such data is unstructured [6] data since the message with #hashtag has no sneaker brands even though the message has the word of sneaker. Consequently, the researcher hence determines such unstructured data to Missing Value and select merely the message with sneaker brands in Tweet text so that it can determine the frequency of each brand.

##### 1) Data collection

In accordance with the data from Twitter, it reveals that the data from Twitter with sneaker #hashtag has 1,553 tweets as a result, but there are merely 799 tweets that can calculate the frequency in which the researcher has collect the

data from 1 August 2019-31 August 2019.

2) *Data transformation*

The data transformation has many tools for analysis purpose as follows: 1) Tokenizing data 2) transforming cases: converting all text cases into merely lower case 3) Filter stop words for removing the article word and preposition due to unnecessary detail for analysis. The data from Twitter with sneaker #hashtag has 1,553 tweets as a result whereas merely 799 tweets that can calculate the frequency since it is explicitly mentioned 20 sneaker brands. The most tweets mention of sneaker brands is Nike with 261 tweets or 34% of the total tweets while Adidas is runner up with 201 tweets or 26% of the total tweets, and other brands such as Fila, Vans, Skechers, Converse, Domba have lower tweet percentage at 12%, 7%, 4%, 4.% and 4.% respectively.

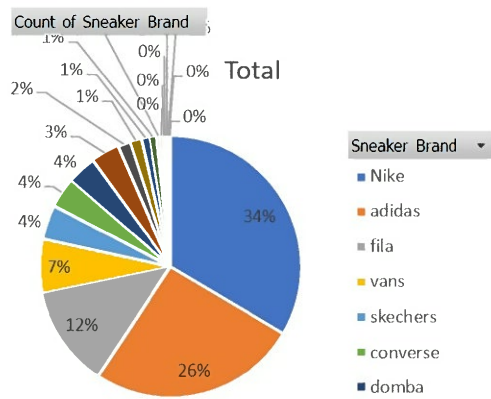


Fig.2 The frequency of sneaker brand mention from examples collection

IV. Model creation

A. *Model creation*

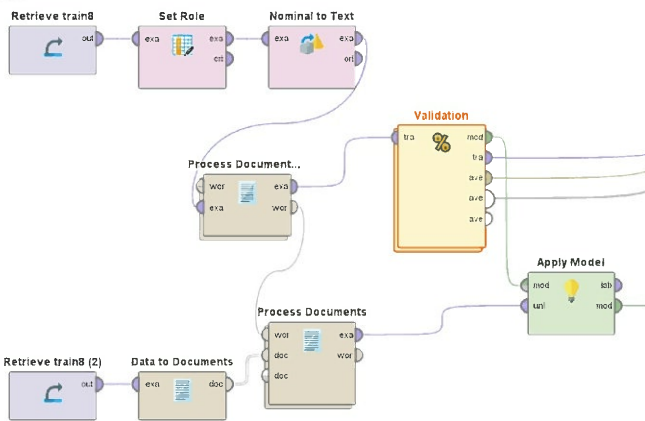


Fig 3. Model Creation for Analysis by Rapid Miner Program

The sentiment analysis divided tweets into 2 classes: positive sentiment and negative sentiment, from Twitter data extraction with #hashtag, toward 2 most popular sneaker brands. First, analyzing the brand popularity by extracting the data with sneaker #hashtag and then discovering the popularity ranking determining by the number of tweets

toward the brands. Afterwards, extracting the data with each brand name #hashtag in which the result shows that the 2 most popular sneaker brands are Nike and Adidas. Thus, doing data preprocessing, for text sentiment analysis, in order to explore the sentiment toward sneaker brands for the next step.

V. Experiment and Results

Not only the model testing is necessary for accuracy and effectiveness purposes, but also for suitability for data analysis. This research uses SVM technique and Naïve Bayes technique in order to compare them and thereupon select the most effective model for further data analysis.

A. *Experiment*

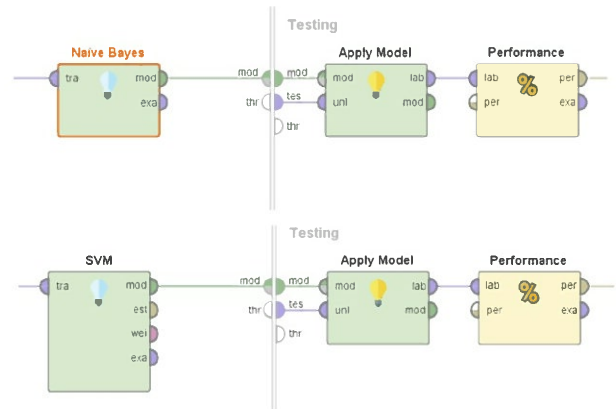


Fig.4. Data Analysis Model on comparison result of Twitter Users' opinion and sentiment by using Naïve Bayes technique and SVM technique

B. *Results of evaluation*

In accordance with simulation modeling between Naïve Bayes technique and SVMs technique with 70:30 ratio, the finding reveals that Naïve Bayes technique reaches accuracy percentage of 63.34% while SVMs technique has accuracy percentage of 69.45%. Nevertheless, when changing the ratio to 80:20, the result shows that Naïve Bayes technique reaches accuracy percentage of 61.35% while SVMs technique has accuracy percentage of 73.91%. In conclusion, the aforementioned result indicates explicitly that SVM technique has greater accuracy than Naïve Bayes technique.

TABLE I. Result of evaluation

Table classification technique		
Percentage	70:30	80:20
Naive Bayes	63.34%	61.35%
SVM	69.45%	73.91%

C. *Text Sentiment Analysis*

It is a process of identifying the opinion or feeling of Twitter users toward both brands through text sentiment analysis method to determine the result of 2 polarities, namely: positive or favorable feeling, neutral or nonaligned state, and negative or unfavorable feeling by

using text sentiment analysis method from tweets.

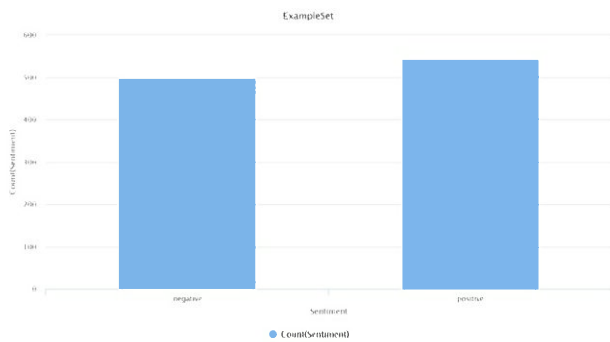


Fig.5. Sentiment of Sneaker Brands when Using Text Sentiment Analysis

The result of Text Sentiment Analysis suggests that the Twitter users have more favorable feeling (positive) to product than unfavorable feeling based on samples of 100 tweets extraction by using sneaker brand #hashtag.

TABLE II. Result of Sentiment

Index	Nominal value	Absolute count	Fraction
1	Positive	541	0.521
2	Negative	497	0.479

## VI. Conclusion

Pursuant to the Twitter data analysis, the result indicates that not only the sneaker #hashtag has impact on Twitter users but it also recognizes Twitter users sentiments toward top 2 most popular sneaker brands. The researcher would like to propose the suggestions- on the basis of development purpose- as follows: 1) Twitter data analysis on Thai language needs proper and suitable text sentiment analysis process since the data warehouse of Thai language is insufficient which resulting in difficulty for further analysis 2) using sneaker #hashtag allows researcher to discover the brand popularity from the frequency of tweets with the mention of brand names, but in the case of specifically refer to the shoe models shoes of each brand, it is necessary to choose the method of data extraction with detailed information. According to the basic information survey before conducting research, there is no means to specify the popular shoe models clearly since there are many shoe models of each brand, and this research thereof construct the hypothesis of the brand popularity based

on information from single source-Twitter- merely 3) the data extraction from Twitter reveals that there are many parties involved: for instances, the official distributor, small supplier, retailer or buyer whom want to buy sneaker. Hence, it is necessary to increase the process and use #hashtag properly in order to analyze the data with more effectiveness and resulting in accuracy and meeting the requirements.

## Acknowledgment

The research was completed with kind cooperation of the research advisor from the Department of Information Technology, Faculty of Information Technology, King Mongkut's University of Technology North Bangkok.

## References

- [1] Chumwatana, Todsana (2015) "Using sentiment analysis technique for analyzing Thai customer satisfaction from social media." In: 5th International Conference on Computing and Informatics (ICOI) 2015, 11-13 August 2015, Istanbul, Turkey.
- [2] R. Othman, R. Belkaroui, and R. Faiz, "Extracting Product Features for Opinion Mining Using Public Conversations in Twitter," *Procedia Computer Science*, vol. 112, pp. 927-935, 2017/01/01, 2017.
- [3] M. Hammad and S. R. El-Beltagy, "Towards Efficient Online Topic Detection through Automated Bursty Feature Detection from Arabic Twitter Streams," *Procedia Computer Science*, vol. 117, pp. 248-255, 2017/01/01, 2017.
- [4] D. Yu, D. Xu, D. Wang, and Z. Ni, "Hierarchical Topic Modeling of Twitter Data for Online Analytical Processing," *IEEE Access*, vol. 7, pp. 12373-12385, 2019.
- [5] R. A. Plunz et al., "Twitter sentiment in New York City parks as measure of well-being," *Landscape and Urban Planning*, vol. 189, pp. 235-246, 2019/09/01, 2019.
- [6] Ankit and N. Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis," *Procedia Computer Science*, vol. 132, pp. 937-946, 2018/01/01, 2018.
- [7] A. Kongthon, C. Haruechaiyasak, J. Pailai, and S. Kongyoung, "The role of Twitter during a natural disaster: Case study of 2011 Thai Flood," *Proceedings of PICMET: Technology Management for Emerging Technologies*, vol. 12, pp. 2227-2232, 2012.
- [8] G. Backfried, J. Gollner, G. Qirchmayr, K. Rainer, G. Kienast, G. Thallinger, C. Schmidt, and A. Peer, "Integration of Media Sources for Situation Analysis in the Different Phases of Disaster Management: The QuOIMA Project," *Intelligence and Security Informatics Conference (EISIC)*, pp. 143-146, 2013.
- [9] Ali, W., Shamsuddin and Ismail A.S, "Web Proxy Cache Content Classification based on Support Vector Machine," *Journal of Artificial Intelligence*, vol. 4[1], pp. 100-109, 2011.
- [10] Ankit and Nabizath Saleena, "An Ensemble Classification System for Twitter Sentiment Analysis," *International Conference on Computational Intelligence and Data Science (ICCIDS 2018)*, pp. 937-946, 2018.