Proceedings Paper:

Altammami, SH orcid.org/0000-0002-3801-8236 and Rana, OF (2017) Topic Identification System to Filter Twitter Feeds. In: Proceedings of the 2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI). 2016 3rd International Conference on Soft Computing & Machine Intelligence (ISCMI), 23-25 Nov 2016, Dubai, United Arab Emirates. IEEE , pp. 206-213.

https://doi.org/10.1109/ISCMI.2016.14

Topic Identification System to Filter Twitter Feeds

Shatha Hamad Altammami
Department of Information Systems
King Saud University
Riyadh, Saudi Arabia
Shaltammami@ksu.edu.sa

Omer F. Rana
Computer Science & Informatics
Cardiff University
Cardiff, UK
O.F.Rana@cs.cardiff.ac.uk

Abstract— Twitter is a micro-blogging service where users publish messages of 140 characters. This simple feature makes Twitter the source for concise, instant and interesting information ranging from friends' updates to breaking news. However, a problem emerge when a user follows many accounts while interested in a subset of its content, which leads to overwhelming tweets he is not interested in receiving. We propose a solution to this problem by filtering incoming tweets based on the user's interests, which is accomplished through a classifier. The proposed classifier system categorizes tweets into generic classes like Entertainment, Health, Sport, News, Food, Technology and Business. This paper describes the creation and evaluation of the classifier until 89% accuracy obtained.

Keywords— Short Text Classification; Classifier; Twitter.

I. INTRODUCTION

Social media has become an important part in our daily life, specifically Twitter[1]. This is due to its nature as a micro-blogging service that sets a 140 character limit in a tweet, which encouraged users to share information in the least words possible. This simple feature has attracted millions of users to make Twitter the source for rich and various information ranging from critical news to personal updates by friends, celebrities or organizations.

When users log in to Twitter, they typically see a chronological stream of tweets in their feed as sent by the people they chose to follow. Thus when a user follows many people, known as friends or resources, he is faced with information overload where it is impossible to read thousands of tweets arriving in his feed every day. To solve this problem, two fundamental questions should be answered. What incoming messages do users value? And how do users manage this flood of messages?

For the first question, the real problem exists when a user is interested in a subset of tweets his friends present. For example, consider a hypothetical user 'A' who follows user 'B' because of the latter's tweets about business. However, 'B' does not limit his tweets on that topic, but also tweet about sport. Currently, 'A' is interested in a subset of 'B' tweets and has few tools to filter non-business tweets from 'B'. This is because Twitter assumes that all tweets from the people 'A' follows contain information he is interested in receiving. In other words, users tend to receive unwanted tweets due to their non-overlapping interests from the people they follow; therefore, filtering the user's feed to present only the relevant and interesting tweets to the user is essential.

To answer the second question, an investigation on the existing Twitter feature and third party tools was conducted. Twitter provides 'List creation' that aids in organizing incoming tweets. Although it organizes the tweet feed, the user still receives every tweet sent by his friends including tweets he is not interested in receiving. Another application that aims to solve this problem is TweetDeck[2], which provides a filtering algorithm that enables the user to filter his feed based on a set of keywords. This application works well when the user knows exactly what he wants to see in his feed by creating filters for specific topics. However, this does not automatically cope with the evolving nature of Twitter, requiring the user to manually update the created filters.

The existence of this problem is further demonstrate through a past study which estimates that only 36% of Twitter's feed is worth reading [1] since many tweets are irrelevant, superfluous, or too difficult to understand without context. Therefore, users can benefit from tools that help them sort the "wheat from the chaff" by analyzing and filtering their tweet feed.

The reminder of this paper organized as follows. Section II presents related work. Then we introduce our system in Section III. After that we thoroughly explain building and training the classifier in Section IV. An experiment of the proposed classifier is tested in Section V. Then we evaluate the results in Section VI, and discuss it in Section VII.

II. RELATED WORK

Large number of studies have been conducted on Twitter for a variety of purposes [2]. A subsit of these studies focused on provideing better experince to the user by filtering tweets based on his interest. One of the traditional methods to discover a user's interests is by analyzing the content of his timeline. This approach was reviewed by past research [3], and determined that profiling users' personal interests in this way is infeasible, because users do not necessarily tweet about all of their interests. Another approach to infer user's interests is by applying the theory used in recommender systems. These systems are classified into collaborative filtering (CF) and content-based filtering (CB) [4]. Studies show that the former technique has two issues, sparsity and scalability [5]. In the other hand, CB technique detects similarities between items that share the same characteristic, which causes overspecialized recommendations that only include items very similar to those of which the user came across[6]. Another idea to obtain users interests was mentioned by Ramage et al. In their work, Twitter users were asked to rate the quality of

[1] https://Twitter.com/
[2] https://about.Twitter.com/products/tweetdeck

posts from users they follow[7]. However, this requires a great deal of time and effort, and becomes infeasible when the data set is large. Other studies focused on reordering a user's feed to place the most important tweets on the top based on specific features. Some of the proposed techniques include sorting tweets according to author influence score, number of followers or retweets [8]. However, current influence metrics are susceptible to be fooled by things like bots [9]. In a different study, a Twitter client called Eddi organizes tweets in a user's feed into groups based on tweet topic[10]. Their topic detection algorithm uses search engine as an external knowledge base. Although they claim Eddi system outperforms comparable topic detection algorithms, Zhang et al proved otherwise by exposing noisy documents to the system, and concluded that Eddi fails to provide accurate results with such documents[11]. Another study that uses the web, involves the automatic generation of multi-domain personalized user profiles[12]. However, this approach require collecting information from the user various social networks, which rises privacy concerns.

III. Our System

In the proposed system, users explicitly identify their interests to prevent cases of cold starts. Then tweets in their feed can be filtered accordingly by determining the tweet's topic. This can be accomplished by designing a system that consists of two parts, user Interface and a Classifier. This is illustrated in Figure1 where the Interface interact with the user and obtain his preference, then collect and filter his tweet feeds. The filtering process is done by sending the tweets to the backend of the system, the classifier, which will use machine learning techniques to classify tweets into a set of predefined classes. These classes were identified after a survey conducted on 380 twitter users, where 78% of them stated they do not read all tweets received due to their overlapping interests with the users they follow. Therefore, these users were asked to identify the most common tweet topics they look for on twitter which were Technology, Sport, Health, Entertainment, Food, News and Business. Once the tweet is classified, it is sent along with its class (topic) to the Interface, which places the tweets into 'Like' or 'Dislike' category according to the user's specified topics of interest. In this paper we focus on the system core, the classifier.

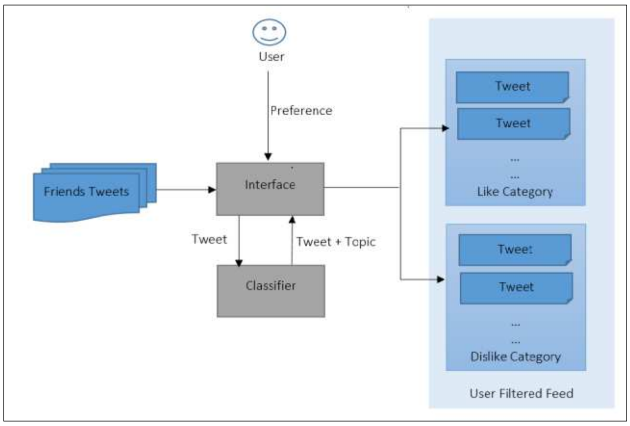

Figure 1. Proposed system

A. Data Collection

The performance of the classifier rely greatly on the amount and quality of the training data. Some researchers used manually-labelled data for training. However, this approach is time consuming and does not produce reliable results since categorization is susceptible to human past experience, and therefore same document can be categorized differently by different people[13]. Another approach is to use the available lexical databases like WordNet. This was successful in a study conducted to identify sentiment in blogs about specific products[14]. However, Twitter has a dynamic nature where new terms are coined, and the data used for classification needs to be up to date. Therefore, relying on lexical knowledge for categorization may not produce as high results. A different approach that considers the evolving nature of Twitter is to use tweets as training data. However, to overcome manually coding tweets into their topics, tweets are obtained from Twitter users who dedicate their timeline to one distinct topic[13]. This approach forms our corpus, which is a collection of labelled tweet that is used as training and testing data for the classifier. We identified at least 10 Twitter accounts for each of the seven predefined topics. For example, TechCrunch is a Twitter account that tweets about technology. After that, we created a crawler to collect 154,905 tweets from 80 Twitter accounts.

B. Preprocess Data

To train a classifier with the collected tweets, we have to present these tweets in a specific way. We chose the bag-of-words (BOW) model, where the frequency of each word in the collection of tweets for a specific topic is used as a feature for training the classifier. However, tweets have to be pre-processed first to improve accuracy in the classification stage. The collected tweets were pre-processed through the following steps.

1- Remove URLs and @username
2- Remove punctuation and special characters
3- Removing repeated letters. E.g. coooool to cool
4- Remove words starting with a number
5- Remove 'RT'
6- Remove stop-words
7- Convert text to lowercase
8- Tokenize words using whitespace

C. Examine Corpus

The proposed system rely on frequency of terms in the collected tweets as features to train the classifier; therefore, we must avoid the notorious "garbage in, garbage out". To do that, training data must be representative for each of the seven predefined classes. There are more than 2,000 tweets for each class, and examining them manually is infeasible. Therefore, a visual representation of tweets for every class was built using TF-IDF scheme to identify the most frequent terms for each class or topic. Then these terms were plotted in a word cloud where the most frequent terms are shown in larger fonts while less frequent terms are shown in smaller fonts. An example of this is shown in Figure 2, which illustrate the word-cloud for the

Technology class. After examining each class word-cloud we were ready to build and train the classifier.

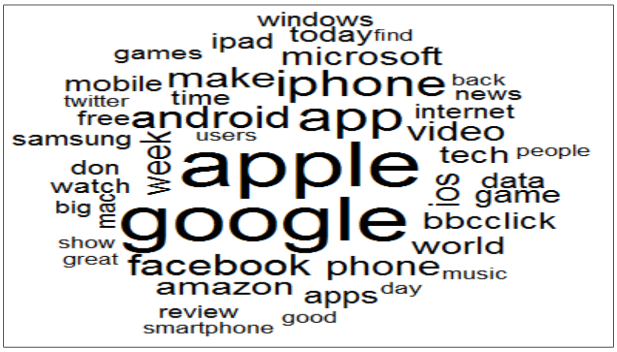

Figure 2. Technology word cloud

IV. BUILDING AND TRAINING CLASSIFIER

To build the classifier, we experimented with different machine learning algorithms, and identified factors affecting the results. The first factor is the steps taken in pre-processing the training data; the second factor is training the classifier with different machine learning algorithms. In fact, enhancing performance required three trails until we reached an accuracy of 94% in the training phase.

A. Evalution Measures

The metrics used to measure the classifier performance are accuracy, precision, recall or F-score. Specifically, we use the formulas listed below, where l is the number of topics which equals seven. Additionally, tp, tn, fp, fn are true positive, true negative, false positive and false negative respectfully. F-score uses a value of β =1 to give an equal weight to recall and precision.

- Precision

$$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l}$$

- Recall

$$\frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l}$$

- F-Score

$$\frac{(\beta^2 + 1)\,\text{precision} * \text{recall}}{\beta^2\,\text{precision} + \text{recall}}$$

- Accuracy

$$\frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + fn_i + fp_i + tn_i}}{l}$$

B. Trail 1

The collected tweets were pre-processed and removed tweets which became empty, thus obtaining a final set of 154304 tweets distributed over the seven classes as shown in Figure 3. We have shuffled the data to ensure randomness for better performance as proved in previous work [13]. Then we divided the collected tweets into two sets, training with 70% of the data and testing with the remaining 30%. Finally, these tweets were fed into the classifier, which applied two algorithms, Support Vector Machine (SVM) and Maximum Entropy

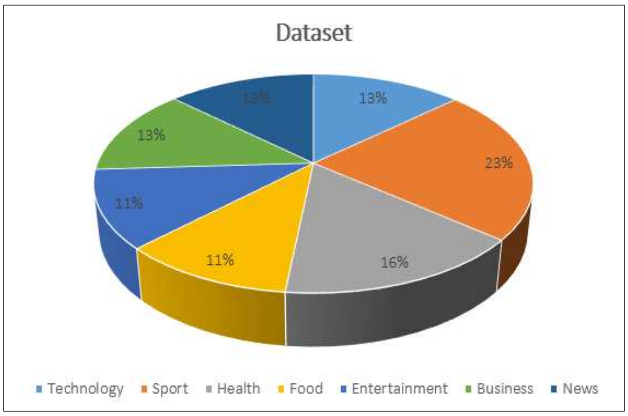

Figure 3. Data distribution

To test the classifier performance for the first trial, we calculate the precision, recall, F-score and accuracy per class as shown in Table 1.

	Technology	Sport	Health	Food	Entertainment	Business	News
FP	1660	7339	2352	1234	2158	2476	1620
TP	2854	8635	4056	2820	2566	2905	3545
FN	3110	1376	2764	1960	2523	2886	2141
TN	36517	26791	34969	38127	36894	35874	36835
precision	0.63	0.54	0.63	0.7	0.54	0.54	0.69
Recall	0.48	0.86	0.59	0.59	0.5	0.5	0.62
F-score	0.54	0.66	0.61	0.64	0.52	0.52	0.65
Accuracy	0.89	0.8	0.88	0.93	0.89	0.88	0.91

TABLE I. TRAIL 1 CONFUSION MATRIX

After that, we obtain the overall performance of the classifier as illustrated in Table 2. Although the accuracy was acceptable, we aimed for a better performance. We notices some tweets had only one or two words that were misclassified, so we try to improve the result by improving the data set as we explain in Trail 2 next.

	Precision	Recall	F-score	Accuracy
Trail 1	61%	59%	59%	88%
Trail 2	62%	59%	60%	89%
Trail 3	83%	83%	83%	94%

TABLE II. TRAILS RESULTS

Glmnet							
	Tech.	Sport	Health	Food	Ent.	Business	News
FP	950	10493	1345	840	547	1322	1598
TP	2349	8848	3560	2405	1691	2239	3356
FN	3442	956	3365	2400	3345	3624	2561
TN	37400	23844	35871	38496	38558	36956	36626
Prec.	0.71	0.46	0.73	0.74	0.76	0.63	0.68
Recall	0.41	0.9	0.51	0.5	0.34	0.38	0.57
Fscore	0.52	0.61	0.6	0.6	0.47	0.47	0.62
MaxEnt							
	Tech.	Sport	Health	Food	Ent.	Business.	News
FP	1932	5098	2095	1676	1937	2145	2212
TP	2949	7714	4097	2952	2620	2801	3764
FN	2842	2090	2828	1853	2416	3062	2153
TN	36418	29239	35121	37660	37168	36133	36012
Prec.	0.6	0.6	0.66	0.64	0.57	0.57	0.63
Recall	0.51	0.79	0.59	0.61	0.52	0.48	0.64
Fscore	0.55	0.68	0.62	0.62	0.54	0.52	0.63
SVM							
	Tech.	Sport	Health	Food	Ent.	Business	News
FP	2027	5948	2375	1172	1718	2178	1677
TP	2986	7929	4222	2797	2482	2832	3640
FN	2805	1875	2703	2008	2554	3031	2277
TN	36323	28389	34841	38164	37387	36100	36547
Prec.	0.6	0.57	0.64	0.7	0.59	0.57	0.68
Recall	0.52	0.81	0.61	0.58	0.49	0.48	0.62
Fscore	0.56	0.67	0.62	0.63	0.54	0.52	0.65

TABLE III. ALGORITHMS CONFUSION MATRIX

C. Trail 2

In this trail, tweets with less than three words were removed. This is because such tweet show very little information and can hardly be classified. After removing short tweets, the dataset content decreased by 5%, however, the overall performance increased by one percent to produce 89% accuracy as shown in Table 2.

D. Trail 3

For this trail, we used an additional algorithm, Glmnet, in which each tweet was classified using three algorithms. This improved performance to reach 94% accuracy as shown in Table 2. The three algorithms performance for each topic is demonstrated in Table3; while the overall performance of algorithms is shown in Table 4. Although Glmnet performance was lower, it boosted the overall performance of the classifier.

	SVM	MaxEnt	Glmnet
Precision	0.62	0.61	0.65
Recall	0.59	0.59	0.52
F-score	0.60	0.60	0.54

TABLE IV. ALGORITHMS PERFORMANCE

V. EXPERIMENT

The system classifier produced high accuracy using the testing data, however, it is essential to test the system on real users and get a better understanding of its functionality in the real world. Therefore, an experiment was conducted on four Twitter users to classify their Twitter feeds then ask them to validate it. The volunteer would read the tweet and decide if the assigned topic is valid or not. If not, they choose the appropriate topic from the seven predefined topics. Moreover, the volunteer can assign more than one topic to a tweet, or indicate the tweet is not clear, or tweet topic is not among the seven predefined topics.

	Tweets	Correct	Incorrect	Not Clear
User 1	200	97	78	25
User 2	220	147	47	26
User 3	202	78	99	25
User 4	220	128	70	14

TABLE V. EXPERIMENT RESULT

The volunteers evaluated a total of 842 tweet. Table 5 demonstrated the distribution of these tweets among the volunteers and how many of these were classified correctly by the classifier. Additionally, the table shows the tweets that were incorrectly classified while their topic was clear to the user. In Table 6, we calculate the average accuracy of all classes per user. The average accuracy obtained for this experiments is 88%, which is lower than the accuracy we obtained using the testing data in Trail 3.

	User1	User2	User3	User4	Avg.
Accuracy	85%	91%	86%	88%	**88%**

TABLE VI. PERFORMANCE PER USER

A. Trail 4

After conducting the experiment, the correctly classified tweets and the tweets reclassified by the users are fed into the classifier as training data. Although the number of new tweets, which is 842, is very small compared to our dataset, we wanted to determine if this would improve accuracy. However, the performance achieved is 89% accuracy.

VI. RESULT AND EVALUTION

A. Statistical Analysis

We evaluate and compare manual and automated categorization techniques by asking two questions. First, how close are the results of the automated method when compared to the manual method? Second, can the result of the automated method be considered accurate enough to be used as an approximation to the manual one?

To answer these questions, statistical analysis is carried to compare the results of both techniques. For the manual annotation technique, we obtain tweets that represents the gold standard or ground truth by using five annotators. This enabled us to draw good human judgement of each tweet. Each annotator can agree with the automatic categorization result, disagree and reassign the tweet to another category, or state that the tweet is not clear and cannot be categorized. Tables 7, 8 and 9 show a breakdown of the agreements among annotators for tweets.

Manual classification Agreement with Automatic classification	Number of Tweets	Percentage
5 Annotators	49	11.7%
4 Annotators	107	25.5%
3 Annotators	52	12.4%
2 Annotators	55	13.1%
1 Annotator	58	13.8%
no Annotator	99	23.6%
Total	**420**	**100%**

TABLE VII. MANUAL CALSSIFICATION AGREEMENT WITH AUTOMATIC CLASSIFIER

Manual classification Disagreement with Automatic Classification	Number of Tweets	Percentage
5 Annotators	18	4.3%
4 Annotators	63	15.0%
3 Annotators	59	14.0%
2 Annotators	67	16.0%
1 Annotator	78	18.6%
no Annotator	135	32.1%
Total	**420**	**100%**

TABLE VIII. MANUAL CLASSIFICATION DISAGREEMENT WITH AUTOMATIC CLASSIFIER

Manual Classifiers Find the Tweet Not Clear	Number of Tweets	Percentage
5 Annotators	3	0.7%
4 Annotators	4	1.0%
3 Annotators	21	5.0%
2 Annotators	40	9.5%
1 Annotator	161	38.3%
no Annotator	191	45.5%
Total	**420**	**100%**

TABLE IX. MANUAL CLASSIFICATION STATE THE TWEET IS NOT CLEAR

To form the gold standard, the decision of three or more annotators for each tweet is taken into account. For example, if three or more annotator agreed with the assigned topic by the classifier, we assume the tweet is correctly categorized. However, if two annotators agree and the other two annotators disagree with the assigned topic, while one annotator find the tweet not clear, we assume the tweet is not clear.

The annotators classified 420 tweets, however 76 tweets were removed since annotators found them not clear. Then we conducted the Chi-square test for the following null and alternative hypotheses:

H0: There is no significant difference between manual and automated categorization analysis tools.

Ha: There is significant difference between manual and automated categorization analysis tools.

The level of statistical significance determines whether to reject the null hypothesis and accept the alternative one or fail to reject it if there is no evidence to prove it. The table below shows the automatic and manual annotation result for 344 tweets.

	Automatic	**Manual**
Health	42	33
Food	43	53
Entertainment	25	35
Sport	106	85
News	46	58
Technology	50	50
Business	32	28
Other	0	2
Total	**344**	**344**
Chi-Square P-value	**0.014764018**	

TABLE X. AUTOMATIC AND MANUAL ANNOTATION OF TWEETS

The p-value in Table 10 indicates that there is a minimum significant difference between manual and automatic classification methods, which mean the null hypothesis can be accepted.

B. Evaluation

Once an acceptable performance was obtaining, an investigation on the reasons behind misclassified tweets is done. However, before explaining these reasons, understanding the classifiers functionality is essential. The classifier must assign the tweet to one of the classes, so it will never indicate that a tweet cannot be classified. Therefore, misclassified tweets fall into the following cases.

a. Tweet has more than one topic.

b. Tweet is extremely noisy or topic is not clear
c. Tweet does not fall under any of the seven topics

An example of these tweets are shown in the table below as indicated by the 'Type' column. To solve this, we can use the additional information provided by the classifier. When the classifier assigns a class to the tweet, it provides the probability of it belonging to that class.

Type	Tweet	Class	Classifier Result
A	New treatment may offer hope for injured Olympian http://t.co/TIneX6otl8	News	Health
B	YMCMB-Young Mo'Ne Cash Mo'Ne Baseball!!!	Entertain.	Business
C	@RichOnOWN @realrobbell loved that Oldsmobile analogy.	Sport	Entertain.

TABLE XI. TYPES OF MISCLASSIFIED TWEETS

We manually vetted the misclassified tweets and observed that tweets can have three classification cases.

1- All three algorithm agree on the assigned class
2- Two algorithm agree on one class while the third algorithm assigns a different class
3- No agreement, in which all algorithms disagree and assign three different classes to the tweet.

The number of instances for each of the above cases are shown in Table 12. The majority of the tweets were classified into one class by all algorithms, and few were classified into three different classes. In the following sections we further elaborate on these cases.

Consensus	Tweets	Correct	Incorrect
3 Algorithms	30463	21601	8862
2 Algorithms	12060	4967	7093
1 Algorithm	1618	1140	478

TABLE XII. ALGORITHMS CONSENSUS

Consensus - Three Algorithms: About 70% of tweets were classified into one class by all algorithms, which implies the tweet can only have one topic. However, even when three algorithm agree on a class, there were 29% misclassification of tweets. Therefore, before assigning the topic, a method to ensure that it really belong to that class is required. One observation, if one of the algorithms gave a probability of 0.5 or more, then it is more likely to be classified correctly. In other words, if all algorithms agree on a class for a particular tweet with probability less than 0.5 for all algorithms, then there is a high chance it is incorrect.

Consensus - Two Algorithms: The second case occurs when the algorithms do not agree on one class for the tweet, which implies the tweet can be classified into two classes. This occurs when two algorithm agree on one class for the tweet, and the third algorithm assign the tweet to a different class. However, the current classifier assigns one and only one class to the tweet. The decision of the class is determined by taking average probability of the two algorithms that agreed on the class and compare it to the probability of the class chosen by the third algorithm. Then it will assign the class with the higher probability.

To enhance performance of the current classifier when dealing with tweets that have several topics, we observe the probability given by the three algorithms. If a tweet was classified into two classes by the algorithms with probability above 0.5 for both classes, then the tweet must be assigned under these two classes.

No Consensus: Although the table above show some correct classification without the algorithms agreement, the highest probability observed was 0.4. Moreover, most tweet were observed to be classified into two topics at most. Therefore, if there was no algorithm agreements on the tweet topic, then topic more likely cannot be identified.

VII. DISCUSION

Although replicating the ability of a human coder to interpret the nuances of a text in context cannot be done by machine learning algorithms, this work proves the ability of the proposed system to classify tweets with 89% accuracy. In fact, due to the unbiased performance of machines, it can outperforms the human coder in interpreting some tweets. This is because the user interpretation of a tweet is based on his experience and knowledge, which was observed through our experiment when one user stated that a tweet about 'yoga' could not be classified. Again, this is due to the person lack of knowledge on this particular sport. Although this is true, many tweets in this experiment were misclassified by the system.

The figure below shows how our system performance improved throughout the first three trails. Then it decreased in the last, when we fed into the classifier different sets of tweets that were manually classified by three different users. This implies the users might have different perceptions and different cultural backgrounds that affected their classification decisions. In fact, the classifier performance was degraded due to the inconsistent classification of similar tweets in the training data. Therefore, performance may increase if all tweets were classified by one person to obtain consistent classified tweets as training data.

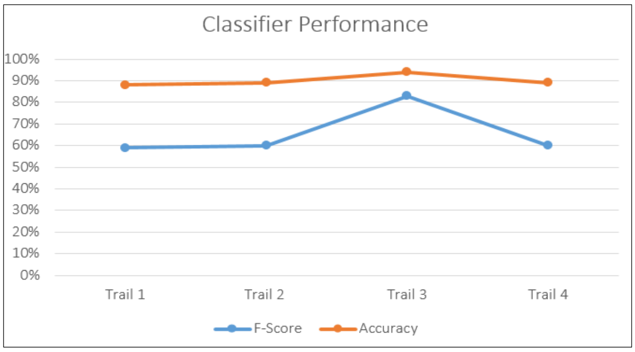

Figure 4 - Classifier trails performance

To enhance the performance, we first look at the factors affecting it. We observed that classifiers nature makes it classify every tweet even if it cannot be sure about its class. Therefore, we can build a method around that to make the classifier assign a class only when it is sure. To do this, we look back at the observation made in previous section. The classification of tweets is decided based on the three algorithms agreement on the class assigned. Another factor affecting the classification, is the probability of a tweet belonging to a class, which is given by each algorithm. Therefore, a tweet can be assigned to one or two classes based on the probability of the algorithms as explained in the previous section. Finally, we acknowledge the limitation of our system and recommend improvement for future work. The improvement include using other features to determine tweet's importance if topic cannot be determined. For example, the presence of URLs or hashtags. Additionally, classification time required by our system is not practical in the real world and we plan for improvement in the future.

VIII. CONCLUSION

This paper tackles a problem faced by Twitter users who receive irrelevant tweets. The problem exists because Twitter assumes if a user follows an account, then he is interested in all of its tweets. This leads to overwhelming users with hundreds of messages they are not interested in receiving. To solve this problem, we propose a system to filter the user's Twitter feeds by knowing the user interests and delivering tweets matching them. Our system uses a classifier, which is the focus of this study.

The classifier categorizes tweets into one of seven predefined topics. It applies three algorithms, SVM, Maximum Entropy and Glmnet; while the data was presented using the Bag-Of-Word approach and TF-IDF feature selection technique. With respect to the short, sparse and noisy tweets, the classifier produced 89% accuracy.

IX. REFERENCE

[1] P. André, M. Bernstein, and K. Luther, "Who gives a tweet?: evaluating microblog content value," in Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 2012, pp. 471-474.

[2] S. A. Williams, M. M. Terras, and C. Warwick, "What do people study when they study Twitter? Classifying Twitter related academic papers," Journal of Documentation, vol. 69, pp. 384-410, 2013.

[3] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and tweet: experiments on recommending content from information streams," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010, pp. 1185-1194.

[4] D. Riaño, E. Onaindia, and M. Cazorla, Artificial Intelligence Research and Development: Proceedings of the 15th International Conference of the Catalan Association for Artificial Intelligence, 2012.

[5] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of dimensionality reduction in recommender system-a case study," DTIC Document2000.

[6] Y. Blanco-Fernández, J. J. Pazos-Arias, A. Gil-Solla, M. Ramos-Cabrer, M. López-Nores, J. García-Duque, et al., "A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems," Knowledge-Based Systems, vol. 21, pp. 305-320, 2008.

[7] D. Ramage, S. T. Dumais, and D. J. Liebling, "Characterizing Microblogs with Topic Models," ICWSM, vol. 10, pp. 1-1, 2010.

[8] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in Proceedings of the 19th international conference on World wide web, 2010, pp. 591-600.

[9] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: when bots socialize for fame and money," in Proceedings of the 27th Annual Computer Security Applications Conference, 2011, pp. 93-102.

[10] M. S. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. H. Chi, "Eddi: interactive topic-based browsing of social status streams," in Proceedings of the 23nd annual ACM symposium on User interface software and technology, 2010, pp. 303-312.

[11] J. Zhang, D. Liu, K.-L. Ong, Z. Li, and M. Li, "Detecting topic labels for tweets by matching features from pseudo-relevance feedback," in Proceedings of the Tenth Australasian Data Mining Conference-Volume 134, 2012, pp. 9-19.

[12] P. Kapanipathi, F. Orlandi, A. P. Sheth, and A. Passant, "Personalized Filtering of the Twitter Stream," in SPIM, 2011, pp. 6-13.

[13] C. F. Simoes, "Catweetegories: machine learning to organize your Twitter stream," 2013.

[14] P. Melville, W. Gryc, and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 1275-128