Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPARATIVE ANALYSIS OF ML POS ON ARABIC TWEETS

[1]MUSTAFA ABDULKAREEM, [2]SABRINA TIUN

Faculty of Information Science and Technology

E-mail: [1]mustafa_aldayni@yahoo.com, [2]sabrinatiun@ukm.edu.my

ABSTRACT

One of the challenges of natural language processing is social media text like tweets. Conversational text in contrast to genres that are highly edited (standard language) which traditional NLP tools have been developed for contains many syntactic patterns and non-standard lexical items. These are the outcomes of dialectal variation, diversity in topic, orthography, unintended errors, conversational errors and creative language use. The fact that twitter text is characterized by idiosyncratic style, noise and linguistic errors makes it difficult to part-of-speech tag. The aim of this paper is to design and implement models of speech tagging for Arabic tweets by investigating numerous models of machine learning like K-Nearest Neighbour, Naïve Bayes and Decision tree models. In this paper, a novel Arabic Twitter corpus is introduced while assessing various state-of-the-art POS taggers which retrained on the given corpus. A state-of-the-art accuracy of 87.97% is achieved when tagging twitter.

Keywords: *Arabic part of speech tagging, Arabic tweets Classification, Feature Extraction*

1. INTRODUCTION

Twitter which has more than 20 million users monthly, offers a wealth of text which is uncoated [1] thereby producing messages reaching a peak rate of more than 230000 per minute [2]. Studies have shown that useful information which can be used for different purpose can be obtained from twitter (for instance monitoring earthquakes [3] and flu prediction [4]. However, the accuracy of major downstream NLP techniques such as term extraction and identification of named entity as well as general application results are impaired by the absence of quality part-of-speech taggers specifically designed for this emerging genre.

Large amount of text in electronic form is being produced by the increasing popularity of social media and the creation of web content by users. Twitter which is one of the commonly used microblogging service is also one of the productive sources of user-created content that has attracted the attention of researchers that seek to exploit and understand these user-created data [5]; [6]. Noisy user-generated text is often described as the difference in language style which is being used by users on social media platforms like Twitter. Processing text obtained from twitter using tools that have been specifically tailored to process well-edited text like that contained in newswire, official documents and literature may be very challenging due to the presence of typographical errors and ungrammatical structures found in tweets. The effect of language style used in user-created web content on the performance of standard NLP tools have been investigated by researchers in previous studies [7] and [8]. Other studies have indicated that it is important for NLP tools and resources to be designed in a way that carters for such linguistic variations that can be found in such text [9-16] (POS-tagging).

The aim of this paper is to design and implement models of speech tagging for Arabic Tweets based on the investigation of different models of machine learning such as KNN, naïve Bayes, and decision tree models. The major contributions of this work are as follows:

(i) An extensive comparative assessment of machine learning POS taggers that have been trained on MSA corpus on tagging Arabic tweet datasets.

(ii) An extensive comparative assessment of machine learning POS taggers which have been trained on a novel Arabic Twitter corpus on tagging Arabic tweet datasets.

There are several motivations which are motivated us to do research for Arabic Twitter POS Tagging:

(1) If we directly apply Arabic NLP tools that are developed for standard language social media text which, the performance will be very low due to the differences between Modern Standard Arabic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(MSA) and Arabic spoken dialects. (2) The need for Arabic Twitter POS tagging in advanced NLP system such as Sentiment analysis on Arabic twitter data. As it is known most of the words that describe the opinions are adjectives and adverbs. Thus, to extract the features for sentiment analysis. (3) The language style variation used on social media platforms, such as Twitter for example, is often referred to as noisy user-generated text. Tweets can contain typographical errors and ungrammatical structures that pose challenges for processing tools that have been designed for and tailored to high quality, well-edited text such as that found in newswire, literature and official documents.

2. RELATED WORK

Only very few studies have been conducted on POS tagging for tweets with only one of such proposed tagging for Arabic tweets [17]. An implementation of Brill's Transformation-Based

Part-of-Speech (POS) was presented by [17] for POS tagging of Egyptian Arabic tweets their accuracy was 87%, This work does not cover other Arabic dialects, whereas in our work. Other Arabic dialects were not included in this work. A first attempt was made by [9] to design and implement POS tagging of English tweets. These researchers were able to design and develop POS inventory specifically for twitter but the level of accuracy was far lower than that of conventional genres. The system2 can be found online for the purpose of research. This tagger which was designed by [9] was CRF based with the major features including checking of words that contain suffix features up to length 3, words with digits or hyphen and capitalization word pattern. Other features such as regular expressions for the identification of hashtags, at-mentions as well as URLs based on frequently-capitalised tokens, phonetic normalization, traditional tag dictionary based on Penn Treebank (PTB), distributional similarity features for the limited data condition were added by other authors for the purpose of improvisation. [11] made an attempt to improve the POS tagging for Twitter as well as Internet Relay Chat (IRC) possessing with unsupervised word clusters; compared to the system which was developed by [9], twitter tagging was improved by 3%. [11], were able to improve the system through the evaluation of large-scale lexical characteristics and unsupervised word clustering. A novel dataset of English tweets was released by the authors who annotated the new dataset using their own guidelines for POS annotation. Through the use of discriminative sequence labelling model: CRF, a French POS tagging system was proposed by [15] who achieved 91.9% accuracy on a target corpus obtained from different kinds of French SMT user such as Twitter, Facebook, medical web forums and Video games. A total of 28 POS tags obtained from Treebank have been proposed by these researchers. Accuracies of 92.7% and 90.1% were reported for similar system setup evaluated on a dataset which contained 800 English social media data like NPS chat with PTB tags and English tweets respectively.

3. PROPOSED METHOD

An automatic Feature-Rich Part-of-Speech Tagging and analysis of Tweets using Machine Learning Classifiers is being designed in this work. The overall architecture of the POS tagging and disambiguation of Arabic Tweets system is presented in Figure.1. The following phases are the phases contained in the machine learning architecture:

- Language resources
- Data pre-processing phase
- Features extraction Phase
- Part of speech tagging phase.
- Evaluation phase.

3.1 Language Resources Description

Using a supervised machine learning technique relies on the existence of annotated training data. Such data is usually created manually by humans or experts in the relevant field. The Arabic POS-labeled Tweets datasets are currently not available. The main objective of this phase is to compile a small annotated Arabic corpus which is primarily intended for training machine learning models for POS tagging of Arabic tweets. Each word in our training corpus is labeled with one of the POS tagging. The data set are collected from Arabic twitter accounts (https://twitter.com/?lang=ar) focusing on those who write in Iraqi, Egyptian and Lebanese dialects. The research aim to design data sets from Arabic tweets. The ATC (Arabic tweets corpus) will contains tweets from many Arabic spoken dialects complied over the period of time. The percentage of Iraqi dialect is 60% of the corpus, the percentage of Egyptian dialect is 25% of the corpus and the percentage of Lebanese dialect is 15% of the corpus.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

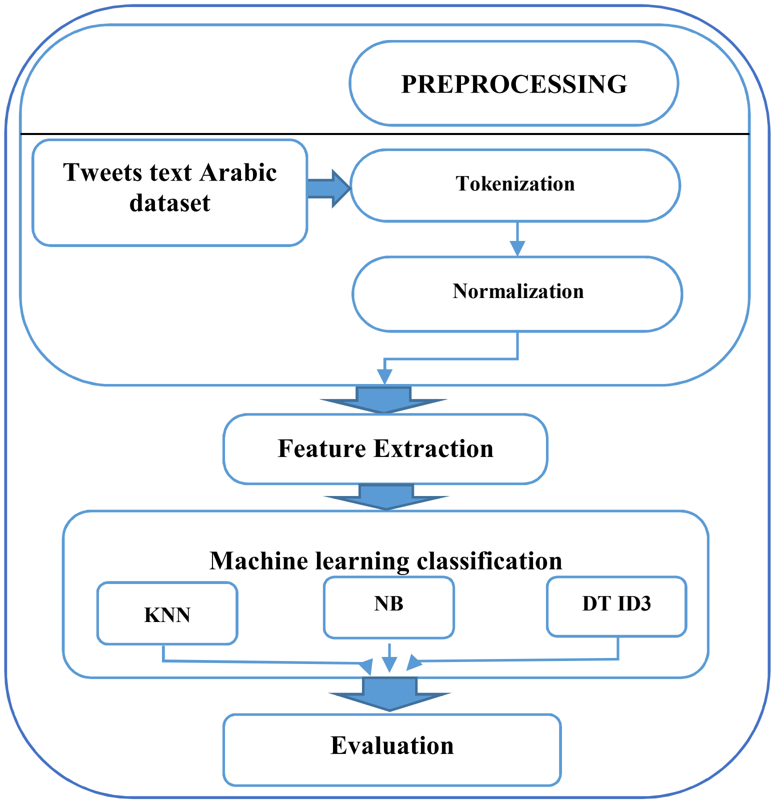

Figure 1. The proposed method framework

3.2 Pre-Processing

Pre-processing is an important activity in the designing of dataset which is done using approaches of machine learning. Prior to part-of-speech tagging, all tweets must be subjected to pre-processing phase. All new tweets in this system were subjected to the following pre-processing steps:

1) Tokenization: this is the initial step which is taken in the processing of any natural language. It is a very important preliminary step in NLP system because it helps in splitting sentences into tokens which can be fed into a POS tagger or morphological analyzer.

2) Normalization: normalization is a step which is also required because of the noisy user-generated text involving variation in language style used on social media platforms like twitter. It is also required because such user-generated text contains ungrammatical structures and typographical errors which cannot be processed by processing tools which have been specially designed to process well-edited text like documents, literature and the kind of text found on newswire. Therefore, the Arabic tweets in this study are being subjected to the process of normalization.

3.3 Feature extraction

This step which helps in improving the effectiveness of tagging tasks in relation to accuracy and speed of the learning is crucial in any natural language processing task including part of speech tagging. In this phase of feature extraction, the processing of test and training data is done using one or more pieces of software so as to extract descriptive information. The task determines the selection of features to be extracted. The identification of the major features of POS tagging has been done using the various possible combination of tag context and available word. The prefixes and suffixes of all words are included in the features. In this work, the concept prefix/suffix is a sequence of first/last few characters of a word, which may not necessarily be a prefix/suffix that is

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

linguistically meaningful. The aim of this phase is to transform each word to a vector of feature values. A set of features for POS tagging of Arabic tweets is defined in this work. A summary of these feature sets is presented in Table.1. The following feature vector represents word in the corpus.

Table 1 summary of feature extraction

Feature category	Feature name	Feature
Word affixes	F3	Prefix 1 (first character)
	F4	Prefix 2 (first two characters)
	F5	Prefix 3 (first three characters)
	F6	Suffix 1 (last character)
	F7	Suffix 2 (last two characters)
	F8	Suffix 3 (last three characters)
Context-based features	F1	Previous word
	F2	Next word
	F9	Word length
	F10	Is the word containing digit

3.4. Machine Learning And Part Of Speech Tagging

This phase involves the designing and implementation of POS analysis and disambiguation algorithm for Arabic tweets using different models of machine learning namely KNN, Naïve Bayes, and Decision Tree models

- **Naïve Bayes**

The naive Bayes technique is exhaustively used for part of speech tagging. Given a table of feature vectors, the technique decides the rear possibility, where the term is related to multiple part of speech tags, and assigns it to the appropriate tag with the maximum rear possibility. There are two used approaches: multinomial models and multivariate Bernoulli models. Naïve Bayes is a stochastic model of generating documents makes use of Bayes' rule. To classify as the best named entity class n* for a new term w, it computes: as in Equation (1)

$$p(c_j|w_i) = \frac{p(c_j)p(c_j|w_i)}{p(w_i)} \qquad (1)$$

- **K-Nearest Neighbour (KNN) Classifier**

The K-nearest neighbor (KNN) is a typical example-based classifier that does not build an explicit, declarative representation of the tag, but rely on the tags attached to the training words Similar to the test words. Given a test word w, the system finds the K-nearest neighbors among training words. The similarity score of each nearest neighbor's words to the test words is used as the weight of the tags in the neighbor's words. The weighted sum in KNN categorization can be written as in Equation (2):

$$\text{score}(w, t_i) = \sum_{d_j = KNN(d)} sim(w, w_j)\, \delta(w_j, t_i) \qquad (2)$$

Where KNN (d) indicates the set of K- Nearest Neighbors of word w If w_j Belongs to 〖 t〗_i, δ (w_j,t_i) equals 1, or other-wise 0. For test word w, it should belong to the tag that has the highest resulting weighted sum.

- **Decision Tree Learning**

Decision tree learning is an approach which learns an abstract but human-readable data structure from the training data – a simple decision tree. Figure 2 shows a simple decision tree that could be used for trip planning. Every node represents a decision required for the final tagging result. The tags are represented by the tree's leaves. The purpose of a decision tree is to order the examples and the algorithm should start with the best ordering features to reach a fast convergence. So a decision tree algorithm has its own feature subset selection. However, to find the best ordering features we have to rank all available features. Therefore, we can use an entropy-based measure like the information gain

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

based on the input training set S and a single feature F: as in Equations (3) and (4)

$$InformationGain(S, F) = Entropy(S) - Average\ entropy\ (S, F). \quad (3)$$

The average entropy is defined by the following formula

$$Average\ entropy\ (S, F) = \sum_i \frac{|s_i|}{s} Entropy(S_i) \quad (4)$$

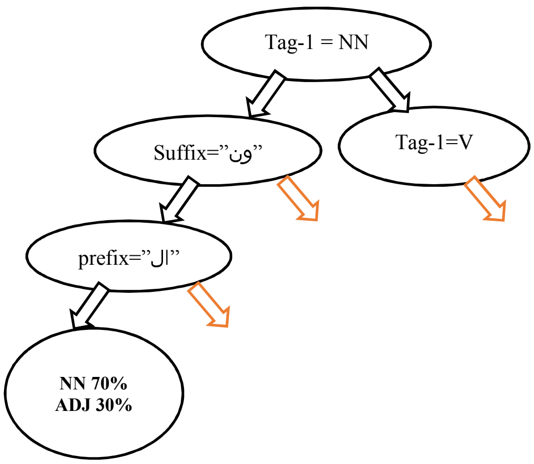

Figure 2: A simple decision tree for trip planning.

3.5. Evaluation

In order to conduct an empirical comparison of 10 different features and four classification techniques (Key Nearest Neighbour, naïve Bayes, decision tree and learning algorithm ID3) numerous experiments have been carried out for models of part of speech tagging for Arabic Tweets. Two datasets were used for this purpose; a properly anotated data set obtained from Arabic tweets and MSA corpus in this study for evaluation as standard corpus.

The overall effectiveness of the POS classification has been measured by recall (R), precision (P), and Macro-average (F1). The opportunity to evaluate the effieciency of the proposed prototype is given by this while the efficiency of the suggested tag is being evaluated. Response time is reduced and smaller space is created by the perfect system. Let the number of words which are tagged manually as appropriate be TP and true positives by the classifier. Let the number of words which are classified manually as relevant be FN and classified as insignificant by the classifier (false negatives). Let FP be the number of words that are classified manually as irrelevant, and classified as appropriate by the classifier (false positives). Finally, the number of words which are classified manually as insignificant be TN and the classifier (true nrgative). as in Equations (5),(6) and (7)

$$\mathrm{Pr_i} = \frac{\mathrm{TP_i}}{\mathrm{PT_i} + \mathrm{FP_i}} \quad (5)$$

$$\mathrm{R_i} = \frac{\mathrm{TP_i}}{\mathrm{TP_i} + \mathrm{FN_i}} \quad (6)$$

$$\mathrm{F - \ measure} = \frac{2\mathrm{Pr} \times \mathrm{Re}}{\mathrm{Pr} + \mathrm{Rp}} \quad (7)$$

4. EXPERIMENTAL RESULTS

In each experiment, each classifier is applied on testing set using 10-fold cross-validation. In this work, we used 10 features which means 2^10 different experiments can be performed. However, the results here are obtained for selected experiments from these 2^10 experiments. Each classifier has been tested in two setting. The first setting is when the classifier is trained on dataset from MSA and tested with data from Arabic tweets. The second setting is when the classifier is trained on dataset from Arabic tweets and tested with data from Arabic tweets.

In the first experiments, NB classifier is applied on testing set using 10-fold cross-validation. The idea is to show the best results obtained when the NB classifier is applied. Table 2 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic Tweets by applying the NB classifier with different set of features. In Table 2 the training data is from MSA and the testing data from Arabic

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

tweets. The highest result yield by NB classifier trained with MSA is with 83.9% f-measure. On the other hand, Table 3 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic Tweets by NB trained using tweets data with different set of features. The highest result yield by NB classifier trained using tweets data with 87.97% f-measure. This means that using labeled data from Arabic tweets has an obvious positive effect on the quality of part of speech tagging of Arabic Tweets.

Table 2 The Performance Of NB Trained Using MSA For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

NB (MSA with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
1	***1***	***0***	***0***	***1***	***0***	***0***	***0***	***0***	***0***	***78.81181***	***93.949***	***82.58894***
0	1	0	0	0	0	1	1	1	1	79.56437	86.87945	80.08154
0	0	0	1	0	0	0	1	0	0	80.7245	71.96321	69.66069
0	0	0	0	1	1	0	0	0	1	77.44526	64.45319	66.38483
0	0	1	0	0	0	0	1	0	1	76.00246	76.61385	68.82677
1	0	0	0	1	0	0	1	0	0	79.05748	85.64269	76.24142
1	0	0	0	0	1	0	1	0	0	77.78286	80.30516	74.61942
1	0	1	0	0	0	0	1	0	1	78.04876	84.44702	75.00429
0	1	1	0	0	0	1	1	0	1	79.62794	86.97598	80.20325
0	0	0	0	0	1	0	1	0	1	75.89075	70.48515	64.63474

Table 3 The Performance Of NB Trained Using Tweets Data For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

NB (Tweets with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
0	1	0	1	0	0	0	1	0	0	86.61097	87.359	84.63746
0	1	0	0	1	0	0	0	0	1	86.12043	87.30279	84.62728
0	1	0	0	0	1	0	0	0	0	69.71856	75.4514	69.7452
0	1	0	1	0	1	0	0	1	0	83.76097	85.15414	82.03383
1	1	0	1	0	0	1	0	0	0	84.97469	91.46045	86.30918
0	1	0	1	1	0	1	0	0	0	86.15121	86.39678	83.63401
1	1	0	0	0	0	0	0	1	0	70.78458	77.3053	72.52478
0	1	0	0	1	1	0	1	1	0	86.21272	86.87654	83.8621
1	***1***	***0***	***0***	***1***	***0***	***0***	***0***	***0***	***0***	***86.7617***	***92.52319***	***87.97295***
0	1	0	0	0	0	1	1	1	1	85.78521	86.18027	83.44461

In the second experiments, KNN classifier is applied on testing set using 10-fold cross-validation. . The idea is to show the best results obtained when the KNN classifier is applied.

Table 4 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic Tweets by applying the KNN classifier with different set of features. In Table 4 the training data is from MSA and the testing data from Arabic tweets. The highest result yield by KNN classifier trained with MSA is with 82.68% f-measure. On the other hand, Table 5 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic

Tweets by KNN trained using tweets data with different set of features. The highest result yield by KNN classifier trained using tweets data with

Journal of Theoretical and Applied Information Technology
31st January 2017. Vol.95. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

87.22% f-measure. This means that using labeled data from Arabic tweets has an obvious positive effect on the quality of part of speech tagging of Arabic Tweets. However, the results obtained by KNN is lower than NB.

Table 4 The Performance Of KNN Trained Using MSA For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

KNN (MSA with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
0	0	0	1	0	0	0	1	0	0	73.92004	78.43478	71.01011
0	0	0	0	1	1	0	0	0	1	70.76638	79.5538	71.46202
0	0	1	0	0	0	0	1	0	1	66.39909	82.11295	69.20732
1	0	0	0	1	0	0	1	0	0	82.80418	85.5127	82.40275
1	***0***	***0***	***0***	***0***	***1***	***0***	***1***	***0***	***0***	***82.12173***	***86.83178***	***82.90667***
1	0	1	0	0	0	0	1	0	1	81.46168	86.52744	82.51141
0	1	1	0	0	0	1	1	0	1	78.98619	73.62976	73.67393
0	0	0	0	0	1	0	1	0	1	71.77119	76.98677	69.10089
1	1	1	1	1	0	0	0	1	0	80.90607	76.00859	76.31637
0	1	1	1	1	1	0	1	0	0	81.89884	79.79442	79.60156

Table 5 The Performance Of KNN Trained Using Tweets Data For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

KNN (Tweets with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
0	0	0	1	0	0	0	1	0	0	77.75862	77.48199	74.56372
0	0	0	0	1	1	0	0	0	1	77.94619	81.73978	76.80717
0	0	1	0	0	0	0	1	0	1	77.8483	81.14662	75.25871
1	0	0	0	1	0	0	1	0	0	86.60175	86.64235	85.45914
1	0	0	0	0	1	0	1	0	0	86.97545	85.76012	84.91924
1	***0***	***1***	***0***	***0***	***0***	***0***	***1***	***0***	***1***	***87.45767***	***88.77528***	***87.22467***
0	1	1	0	0	0	1	1	0	1	84.39596	86.5424	83.53671
0	0	0	0	0	1	0	1	0	1	77.45483	76.20688	71.46441
1	1	1	1	1	0	0	0	1	0	86.84627	85.59959	83.81477
0	1	1	1	1	1	0	1	0	0	85.48172	87.72464	85.04308

In this experiments, decision tree ID3 classifier is applied on testing set using 10-fold cross-validation. The idea is to show the best results obtained when the decision tree ID3 classifier is applied.

Table 6 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic Tweets by applying the decision tree ID3 classifier with different set of features. In Table 6 the training data is from MSA and the testing data from Arabic tweets. The highest result yield by decision tree ID3 classifier trained with MSA is with 84.2% f-measure. On the other hand, Table 7 shows the performance in terms of the precision, recall, F-measure of the part of speech tagging of Arabic Tweets by decision tree ID3 trained using tweets data with different set of features. The highest result yield by decision tree ID3 classifier trained using tweets data with 86.4% f-measure. This means that using labeled data from Arabic tweets has an obvious positive effect on the quality of part of speech tagging of Arabic Tweets. However, the results obtained by decision tree ID3 is lower than that of NB and KNN when they are trained using tweets data. However, it achieves the best result among all classifiers when they are trained using MSA.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 6 The Performance Of Decision Tree ID3 Trained Using MSA For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

DT ID3 (MSA with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
0	0	0	1	0	0	0	1	0	0	72.98419	84.64775	72.60638
0	0	0	0	1	1	0	0	0	1	72.79086	80.1938	72.34908
0	0	1	0	0	0	0	1	0	1	72.77758	83.00053	70.96095
1	0	0	0	1	0	0	1	0	0	82.76144	89.30386	85.03685
1	0	0	0	0	1	0	1	0	0	76.89493	89.35266	79.8919
1	***0***	***1***	***0***	***0***	***0***	***0***	***1***	***0***	***1***	***82.74937***	***92.1463***	***85.54746***
0	1	1	0	0	0	1	1	0	1	81.50712	77.09031	75.00431
0	0	0	0	0	1	0	1	0	1	72.98419	84.64775	72.60638
1	1	1	1	1	0	0	0	1	0	80.08212	78.23234	75.30516
0	1	1	1	1	1	0	1	0	0	80.35231	77.05936	74.45854

Table 7 The Performance Of Decision Tree ID3 Trained Using Tweets Data For Part Of Speech Tagging Of Arabic Tweets With Different Feature Sets

DT DT3 (Tweets with Tweets)												
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	PRECISION	RECALL	FMEASURE
0	0	0	1	0	0	0	1	0	0	79.69624	80.65906	77.21724
0	0	0	0	1	1	0	0	0	1	79.82846	81.24758	77.5745
0	0	1	0	0	0	0	1	0	1	79.71629	80.71306	77.24203
1	0	0	0	1	0	0	1	0	0	89.4908	84.81385	85.41921
1	0	0	0	0	1	0	1	0	0	86.47725	82.99104	83.118
1	0	1	0	0	0	0	1	0	1	89.47923	84.79828	85.4051
0	1	1	0	0	0	1	1	0	1	87.02468	85.59804	84.76463
0	0	0	0	0	1	0	1	0	1	79.69624	80.65906	77.21724
1	***1***	***1***	***1***	***1***	***0***	***0***	***0***	***1***	***0***	***89.53983***	***87.5208***	***86.78573***
0	1	1	1	1	1	0	1	0	0	87.0565	85.64003	84.80334

According to the experiments of part of speech tagging of Arabic Tweets when classifiers trained using tagged data from MSA, the highest result yield by decision tree id3 classifier with 84.3% f-measure. According to the experiments of part of speech tagging of Arabic Tweets when classifiers trained using tagged data from
tweets, the highest result yield by NB classifier with 87.97% f-measure.

5. CONCLUSION

In this work the experiments which were conducted to empirically compare three approaches (naïve Bayes, key nearest neighbor, decision tree learning algorithm
ID3) for models of part of speech tagging for Arabic Tweets was described. The main objective of this work is to produce a new methodology for POS tagging for Arabic tweets, In order to examine the performance of machine learning techniques so as to determine the most suitable method for POS tagging of Arabic dialects, different experiments were carried out. Results of the experiments showed that in POS tagging of Arabic Tweets the highest result yield is obtained by decision tree ID3 when classifiers are trained using tagged data from MSA while the highest result yield is obtained by NB classifier if training is done using tagged data from tweets with 87.97% f-measure. The methodology successfully meets these objectives. In conclusion, based on the results of the experiments it is recommended that data from the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

same domain be used in order to obtain the best part-of-speech tagging for Arabic tweets.

ACKNOWLEDGEMENT

This research project is partially funded by Malaysia Government under research grant FRGS/1/2016/ICT02/UKM/02/14.

REFERENCES:

[1] L. O'Carroll, "Twitter active users pass 200 million," *The Guardian,* vol. 18, 2012.

[2] O. Ashtari, "The super tweets of# sb47," *Twitter. com Blog,* 2013.

[3] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 851-860.

[4] A. Culotta, "Towards detecting influenza epidemics by analyzing Twitter messages," in *Proceedings of the first workshop on social media analytics*, 2010, pp. 115-122.

[5] A. Ritter, S. Clark, and O. Etzioni, "Named entity recognition in tweets: an experimental study," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 1524-1534.

[6] M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment in Twitter events," *Journal of the American Society for Information Science and Technology,* vol. 62, pp. 406-418, 2011.

[7] J. Foster, Ö. Çetinoglu, J. Wagner, J. Le Roux, S. Hogan, J. Nivre*, et al.*, "# hardtoparse: POS Tagging and Parsing the Twitterverse," in *AAAI 2011 Workshop on Analyzing Microtext*, 2011, pp. 20-25.

[8] M. Osborne, S. Petrovic, R. McCreadie, C. Macdonald, and I. Ounis, "Bieber no more: First story detection using Twitter and Wikipedia," in *SIGIR 2012 Workshop on Time-aware Information Access*, 2012.

[9] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein*, et al.*, "Part-of-speech tagging for twitter: Annotation, features, and experiments," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, 2011, pp. 42-47.

[10] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, and N. Schneider, "Part-of-speech tagging for Twitter: Word clusters and other advances," *School of Computer Science, Carnegie Mellon University, Tech. Rep,* 2012.

[11] O. Owoputi, B. O'Connor, C. Dyer, K. Gimpel, N. Schneider, and N. A. Smith, "Improved part-of-speech tagging for online conversational text with word clusters," 2013.

[12] L. Derczynski, A. Ritter, S. Clark, and K. Bontcheva, "Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data," in *RANLP*, 2013, pp. 198-206.

[13] F. Godin, B. Vandersmissen, A. Jalalvand, W. De Neve, and R. Van de Walle, "Alleviating Manual Feature Engineering for Part-of-Speech Tagging of Twitter Microposts using Distributed Word Representations," in *Workshop on Modern Machine Learning and Natural Language Processing, NIPS 2014Workshop on Modern Machine Learning and Natural Language Processing (NIPS 2014)*, 2014.

[14] T. Lynn, K. Scannell, and E. Maguire, "Minority Language Twitter: Part-of-Speech Tagging and Analysis of Irish Tweets," *ACL-IJCNLP 2015,* p. 1, 2015.

[15] F. Nooralahzadeh, C. Brun, and C. Roux, "Part of Speech Tagging for French Social Media Data," 1764.

[16] A. Jamatia and A. Das, "Part-of-Speech Tagging System for Indian Social Media Text on Twitter," स*ocial-*ई*ndia 2014,* vol. 2014, p. 21, 2014.

[17] R. Al-Sabbagh and R. Girju, "A supervised POS tagger for written Arabic social networking corpora," 2012.