

# ARABIC NEWS CREDIBILITY ON TWITTER: AN ENHANCED MODEL USING HYBRID FEATURES

SAHAR F. SABBEH<sup>1</sup>, SUMAIA Y. BAATWAH<sup>2</sup>

<sup>1</sup>Information systems Dept. Faculty of Computing and information technology, King AbdulAziz University- KSA.

<sup>1</sup>Information systems Dept. Faculty of Computing and information technology, Banha university - Egypt.

<sup>2</sup>Computer science Dept. Faculty of Computing and information technology, King AbdulAziz University- KSA.

E-mail: <sup>1</sup>ssabbeh@kau.edu.sa, <sup>2</sup>sbaatwah@stu.kau.edu.sa

## ABSTRACT

Recently, social media and specially Twitter has become a main source for news consumption and sharing among millions of users. Those platforms enable users to author, publish and share content. Such environments can be used to publish and spread rumors and fake news whether unintentionally or even maliciously. That is why credibility of information in such platforms has been increasingly investigated in many domains (i.e. information sciences, psychology, sociology...etc). This paper proposes a machine learning - based model for Arabic news credibility assessment on Twitter. It uses hybrid set of features that are topic and user related to evaluate news credibility. In addition to the traditional content-related features, Content verifiability and users' replies polarity analysis used for a more accurate assessment. The proposed model consists of four main modules: a) content parsing and features extraction module, b) content verification module, c) users' comments polarity evaluation and d) credibility classification module. A data set of 800 Arabic news that are manually labeled is collected from Twitter. Three different classification techniques were applied (Decision tree, support vector machine (SVM) and Naive Bayesian(NB)). For model training and testing, 5-fold cross validations were performed and performance diagnostics were calculated. Results indicate that decision tree achieves TRP higher than SVM by around 2% and 7% than NB, also FPR almost 9% lower than SVM and 10% lower than NB. For precision, recall, f-measure and accuracy, decision tree achieves almost 2% higher than SVM and 7% higher than NB for the tested data-set. Experiments also revealed that the proposed system achieves accuracy that outperforms the system proposed by Hend.et.al [29] and TweetCred [2].

**Keywords:** *News Credibility, Arabic News, Machine Learning, Twitter, Verifiability, Text Polarity*

## 1. INTRODUCTION

Twitter belongs to a category of web applications that support user generated content (UGC). Those types of platforms do not require any user – side design or publishing skills. They provide a channel where users are allowed to create, publish and share information easily[21]. In recent years, Twitter has grown vastly in terms of users and content. According to [www.Alexa.com](http://www.Alexa.com) ranking for the top 500 we sites, Twitter is globally ranked the 13<sup>th</sup> and approximately about 73 million global Internet users visit Twitter daily. Tweets represent users' personal opinions and discussions or news headlines. This makes Twitter one of the main sources for news publishing, sharing and spreading between users.

However, those types of platforms suffer from the lack of supervision over content which can lead to misleading, and inaccurate (fake) information either unintentionally or intentionally for malicious purposes of misleading consumers[1,3].

Thus the main characteristics of fake news are: a) intent (non-credible news are written with a dishonest intention to mislead other users) and b) authenticity (verified as fake)[4,6]. That is why, there is a pressing need for tools that can differentiate between credible and non-credible information.

We can divide the research in the literature in to four categories; a) the first includes works that try to identify the most informative attributes/features for higher precision credibility. These features are at different levels (i.e. user level features, content and/or message level features...etc [22], [23], [24],

[25], [31], [32]. b) The second on the other hand, argues that the most influential feature is user reputation, credibility and trustworthiness [9], [19], [20]. c) Another point of view relies only on textual, visual and contextual features at topic/post level [1], [2], [5],[8],[11],[12],[13]. In those studies, text analysis and natural language processing (NLP) techniques used to identify content features to assess credibility. Finally the fourth category, c) the hybrid model that combines features at more than one level arguing that neither source- related nor content related features can solely guarantee high accuracy rates.

In this paper, and after surveying the research in the literature, we can outline the key limitations as follows:

1. Most of the proposed models for news credibility target English language, while a huge amount of Arabic news is available.
2. The majority of the literature focuses on either source – based or content-based credibility and the available resources (especially ones that target the Arabic language) that adopt hybrid features use only two or three combined features for the evaluation.
3. Sentiment analysis techniques were used in the context of credibility assessment [18],[20]. The work in [20] used sentiment of content to determine its subjectivity, while [18] latter used sentiment analysis for identifying the source/user sentimental state that can influence his judgment of tweets. None of them targeted the analysis of replies to get their polarity as an indicator of credibility.

#### **Our Contributions:**

- (1) The proposed model targets news in Arabic Language.
- (2) The model utilizes a hybrid set of features that relates to both user and topic as well as content verifiability against trusted external sources. Based on a survey of the most informative features we have chosen a set of previously investigated topic related and source related features. One new added feature includes the analysis of users' replies polarity as an indication for credibility.
- (3) We apply different machine learning techniques for classifying 800 news extracted from Twitter and record their performance. The experimental results showed that the proposed model is more accurate in comparison of [29] and [1].

**Paper Organization:** This paper is organized as follows: Related work is presented in Section 2. The proposed model architecture and details are presented in section 3. Next, we describe the used data set and discuss our experimental results in Section 4. Section 5 presents the evaluation of the proposed model in comparison with two other systems. Finally, we conclude our work and present our insights in Section 6.

## **2. RELATED WORK**

Information credibility has been investigated in the context of social media due to the lack of supervision in such environments. The work on credibility assessment can be classified based on the features used for credibility assessment to:

- a) User/source- related features,
- b) Topic/post - related features.
- c) Hybrid features.

### **2.1 User/Source -Related Features**

This direction of research adopts the claims that inaccurate news can probably be created and spread by automated software agents or fake accounts created only for this sake. Thus, source/user-related features (i.e. users' account/profile, demographics, age, account age, followers, photo,...etc) can be extracted and used to evaluate source credibility. Such features alone are not enough as assessing user reputation is also important in order to filter malicious users[14], [15].

User trustworthiness and reputation were investigated in the context of online knowledge repositories like Wikipedia [16], [17] where they were evaluated in order to predict the quality of users' new contributions. On social media, face book posts were classified as hoaxes or not based on users who liked them[9] rather than its content. The work assumes that posts can be classified as hoax or non-hoax based on the analysis of users' polarization. The work divided users into three categories i) users liked hoax posts only, ii) users liked non-hoax posts only, and iii) users with mixed likes. Thus, the category to which users who like a post belong could indicate the nature of the post.

User behavior of tweeting and retweeting was also analyzed for assessing the credibility of tweets and classifying them as credible or non-credible[20]. Graph-based analytical techniques were applied on user networks in which users can follow each other and receive each other's' posts, which results in a directed graph which can be analyzed to identify trusted sources/users[19]. The

work in [18], tried to identify credible users/sources among Twitter users based on analyzing users' reputation on a given topic as well as measuring users sentiment to identify topically relevant and credible sources of information.

## 2.2 TOPIC-RELATED FEATURES

Researchers in this literature base their work on the assumption that topic/post - related features can help identify non-credible content. Those features can be extracted either from: (i) content or (ii) context. Content - related features include visual and textual features which can be collected and analyzed using standard NLP and text analysis techniques (i.e. images included, Hash-tags, URLs, sentiments/subjective content...etc).

In TweetCred [1], which is a real time credibility assessment tool based on semi-supervised ranking model. It extracts content-related features in real time and feed them together with the annotated posts into SVM-Ranking algorithm. Upon new feeds, the model predicts credibility and displays rating scaled from 1 (low) to 7 (high). Web and text mining techniques were utilized in [8] to detect rumors. This work assumed the existence of different copies for the same piece of information with original and fake copies. It tries to locate candidate text source and compare it with a given text using parse thicket graph analysis. In [2], they tried to identify trending rumors by text analysis techniques to find signature phrases used to express disbelief and/or uncertainty of a certain piece of information. Clusters of similar posts were ranked based on the probability of containing rumors.

On the other hand, contextual information were also analyzed including topic headlines, users comments, rates, likes, emotional reactions, number of shares,...etc. For example, topic/post headline may be misleading (known as "clickbaits") which implies non-credible content or at least irrelevant content. NLP techniques can be utilized to identify clickbaits. For example the work in [13], identifies clickbaits based on matching n-gram of the topic/headline. Another work differentiated between ambiguous and misleading headlines[5], as ambiguity does not necessarily mean inaccuracy. This work utilized sequential rules to detect headline ambiguity whereas misleading headlines were identified based on the similarity between headline and topic content/body.

Another contextual feature is user rating and/or tagging which was used in [11],[12] to classify topics' credibility. The availability of tools that enable users' feedback and tagging to identify fake

news (i.e. Facebook tool) enabled researchers to identify credibility as well as to analyze the accuracy of users' flagging over time.

## 2.3 HYBRID FEATURES

This category employed both source and topic related features for more accurate identification of credible information in social media.

The work in [7] relied on hybrid set of features to identify credibility of rumors in Sina Weibo platform which is a Chinese microblog. This work introduced two new sets of features including client-based features (web application client and mobile client program type) and location based feature to identify the place where the event in the topic took place (domestic (in China) or foreign). Other set of features were applied in [20]. They chose a set of features at different levels: a) message-based, b) user-based, and c) propagation-based features. Both content-level features and network structure were used in [26]. This work tried to rank users/source based on an estimated expertise to a certain topic.

Due to the increasing amount of Arabic content and huge number of Arab world users, Arabic content on social media was target to credibility analysis. Hend. et.al [28],[29] based their model for credibility assessment on verification of news against external sources and more three features (the existence of inappropriate words, is user account verified? and user grade based on grade given for Twitter account by TwitterGrader.com). The source of the article and the time of occurrence are two features used to analyze news articles in [33]. This work considered the lake of those features as a violation that indicates article inaccuracy. CAT[30], which is an assessment tool for Arabic tweets credibility relied on content and source - related features.

In summary, the bulk of the research in the literature and especially ones which target the Arabic content disregarded users' comments as an important credibility indication. In the next section we propose a machine learning - based model for credibility assessment for Arabic news on Twitter.

## 3. ARABIC NEWS CREDIBILITY ASSESSMENT FOR TWITTER

According to a report[27] released in March 2017, the number of active users on Twitter (monthly) is approximately 11.1 million and 29% of all active Twitter users are in Saudi Arabia. Totally, the Arabic countries generate 27.4 million tweets per day where Saudi Arabia generates 33%

of all tweets in the Arab region. This makes Twitter a tool for broadcasting news in Saudi Arabia and other Arab countries. As previously stated, the work in Arabic language suffer from the shortcoming of relying only on user - based and content- based features while neglecting other contextual data especially users' comments.

In this work we adopt the hybrid- level feature model, as we think that the short length of tweets and replies makes it necessary to combine features at content, context and source level. After surveying the literature and experiments to identify the most informative features we based our work on a set of features investigated in [20],[24],[28],[31],[32]. The chosen features include topic - related (textual, visual and contextual), source - related and verifiability. Additionally, users' comments/replies

polarity is added to the features as we think that it will help improve credibility evaluation.

**1) Source -based features include:** Is user verified? , Account demographics (location) near event? , number of followers, account age, is account active?, bio available? Bio contains URL?, account has real name?

**2) Content - based features include:** contains a picture or URL?, Contains Hashtags?, number of retweets?, Content is verified against verified sources?, users' replies polarity. The selected features are summarized in figure 1(a,b).






Type	Feature	Description	Hypothesis	Implementation	ref
User	1. Author Has A "Verified" Account		Authors with verified accounts have a high credibility	Twitter API	Meredith Morris et.al [24]
	2. Author Location Near News Event Topic		Authors with a similar location to the location mentioned in the content indicate higher credibility	- We should find if there is a city name in the content then compare it with the user location - We can get user location from Twitter API	Meredith Morris et.al [24]
	3. Author Has High Number Of Followers		Authors who have >10k followers is considered an influencer and have a high credibility	Twitter API	Castillo et.al [20]
	4. Author Use real Name		Authors with real names indicates higher credibility	Extracted using Twitter API and checked manually	Meredith Morris et.al [24]
	5. account is old		If the account registered before 2013 and active, it considered credible.	-Twitter API	Amal Almansour [32]
	3. Account is active			There exist tweets/retweets in past 30 days	-Twitter API

Figure 1(a): Selected Feature set for Credibility assessment

Type	Feature	Description	Hypothesis	Implementation	ref
	7. Author bio include url		If the author have a web link in his/her bio, it is considered more credible.	Twitter API	John O'Donovan et.al [31]
	3. Author has a bio		If the author have a bio, it is considered more credible	Twitter API	Amal Almansour [32]
Content	9. Content with URL		Tweets that contain URL may redirects to a trusted website which indicate a high credibility	- We should make sure that the URL is a valid link. - If the URL is shortened, we unshorten it first. - The url should be valid. It should redirect to official .gov or credible newspaper.	Carlos Castillo et.al [20]
	10. Content with event "image" attached or video		Tweets that contain event image or video indicates higher credibility	Extracted using Twitter API and checked manually	Amal Almansour [32]
	11. Contain hashtag #		Tweets that contain hashtags indicates higher credibility	Twitter API, REGEX tools	John O'Donovan et.al [31]
	12. Tweet highly retweeted		If the tweet retweeted more > 1000 it considered more credible	Twitter API	John O'Donovan et.al [31]
	13. Verifiability		Similarity with trusted news websites	TF/IDF	Hend S. Al-Khalifah [28]

Figure 1(b): Selected Feature set for Credibility assessment

3.1 System Architecture

The architecture of the proposed system shown in Figure.2 consists of four modules

- a) **Content parsing and feature extraction** which is responsible for extracting news content and features via Twitter API.
- b) **Content verification module**: which verifies news content against verified sources.
- c) **Users' comments polarity evaluation**: which analyzes users' comments to determine their polarity.
- d) **Credibility classification module**: which evaluates any piece of news as credible or not credible.

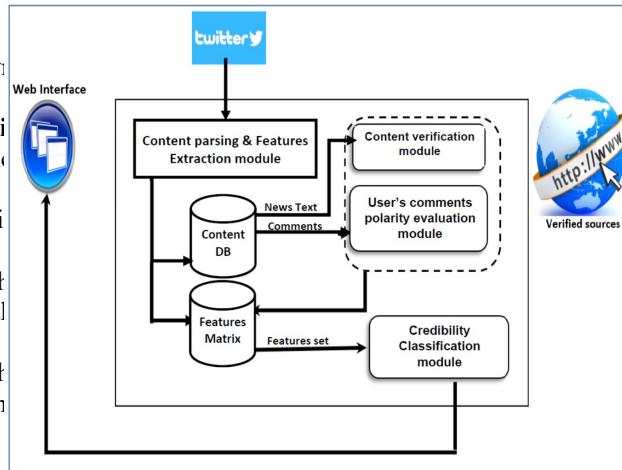


Figure 2. The Proposed System Architecture

The system first receives a URL of the content to be checked. The content parsing and feature extraction module then parses news content, account - related features, users' comments using Twitter API.

Content is then processed to extract its related feature using text processing and regular expression techniques. Topic and account - related features are then stored in the feature matrix.

Content is then verified against external trusted sources via the content verification module. Extracted users' comments are then analyzed for polarity evaluation assessment. Feature matrix is updated and fed into the credibility classification module which then sends back its results to the web interface. The process of assessment is shown in figure 3.

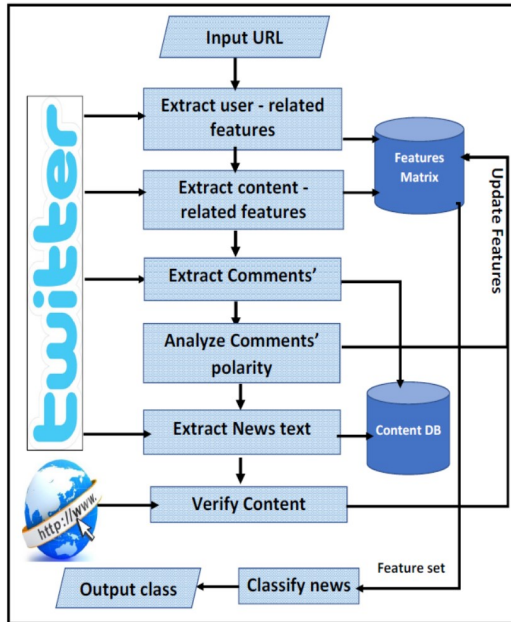


Figure 3: Steps For Credibility Assessment.

### 3.1.1 Content parsing and extraction

This module is mainly responsible for parsing news content (textual/visual) and storing them in content database. After parsing content, its features are extracted (content has images, URLs contained, hashtag, no of retweets) and stored in a features matrix. Source-related features are also extracted using Twitter API (is account verified, account demographics, is account active, number of followers, account has bio, image, real name?, account age) and then stored in features matrix.

### 3.1.2 Content verification module

One important feature to ensure credibility of any piece of information is **verifiability**, which enables users to verify content. In the proposed system, this module is responsible for verifying content against credible and reliable external sources (i.e. Saudi press agency: <https://www.skynewsarabia.com>), <http://www.spa.gov.sa>, <https://arabic.cnn.com>, <https://www.alarabiya.net/>). Based on the work in [29], verification is calculated using cosine similarity between news and external content.

### 3.1.3 Users' comments polarity evaluation module

Users' comments are an integral part of all social media platforms. Comments are powerful feedback mechanisms that enable users to share their experiences and opinions. Analyzing those comments have been target in many research

domains (i.e product reviews, political events...etc). In the proposed system, we believe that the polarity of users' comments can have a significant effect on credibility assessment. Top k comments are evaluated for their polarity. Each comment is tokenized into a list of words. All words in the comment are then processed using an Arabic NLP library [36]. Comments' polarization can be calculated with the help of Arabic sentiment library for Twitter "AraSenTi" [35]. A set of standard Arabic and Saudi variant of Arabic words that indicate falsifying or supporting information are used as an input to the polarity calculation algorithm (i.e. كذب, مو كذب, صحيح غير صحيح, كذب...etc)

The total weight is given for each comment based on the occurrence of negative/positive words based on the following algorithm.

**Inputs:** five vectors (*dec\_wordlist*, *inc\_wordlist*, *inv\_wordlist*, *negative\_wordlist* and *positive\_wordlist*) where:

- inc\_wordlist* : list of Arabic words that indicate exaggerate assurance [i.e. كثيرا, فعلا, جدا, كثر]
- dec\_wordlist*: a list of words that indicate exaggerate contradiction {i.e. بالكاد, مرة..etc}.
- inv\_wordlist* : a list of words that indicate denial {i.e. لا, ليس, لن, غير, ما..etc}.
- negative\_wordlist* : a list of words that indicate negative/false content.
- positive\_wordlist* : a list of words that contains positive/true content.

**Output:** polarity: integer value of text polarity.

- Initialize *polarity* = 0
- Repeat the following steps for *index*, *word* in *enumerate(text)*
- if *word* in *inc\_wordlist*:
- if *text[index + 1]* in *positive* then *polarity* = *polarity* + 1
- if *text[index + 1]* in *negative* then *polarity* = *polarity* - 1
- if *word* in *dec\_wordlist*:
- if *text[index+1]* in *pos\_wordlist* then *polarity* = *polarity* - 1
- if *text[index+1]* in *neg\_wordlist* then *polarity* = *polarity* + 1
- if *word* in *inv\_wordlist*:
- ant* = *replace(text[index + 1])*
- if *ant* then
- if *ant* in *pos\_wordlist* then *polarity* = *polarity* + 2
- if *ant* in *neg\_wordlist* then *polarity* = *polarity* - 2
- if *word* in *pos\_wordlist* then *polarity* = *polarity* + 1

15. if word in neg\_wordlist then polarity =polarity+ 1
16. return polarity
17. End

Algorithm 1.Steps For Credibility Assessment.

After the polarity of each word is retrieved, polarity of each comment is calculated by the summation of polarities of all words in a comment divided by the number of words in the comment. Then the total polarity of the topic is calculated by the weighted sum of polarity of all comments, mathematically calculated as follows:

$$\text{Total polarity} = \sum_{k=1}^K \sum_{w=1}^N \frac{\text{polarity}_i}{N(\text{words})} \quad (1)$$

Where: *k* is comment in all comments *K*.  
*w* is word in the comment *k*.

### 3.1.4 Credibility classification module

As indicated previously, we make use a set of hybrid features for credibility evaluation. This module accepts the vector of features of the post/new and these features are then fed into a trained classifier. The J48 decision tree classifier was chosen based on the observed values of our experiments (section 4). J48 is the implementation of ID3 algorithm that builds tree in a top-down approach. It uses entropy and information gain to construct the tree. Information Gain is the difference between the base entropy and the conditional entropy of the attribute. The attribute/predictor with the highest information gain is said to be the "most informative attribute" and used for tree branching/partitioning[37]. The algorithm proceeds as follows:

- 1- Select the most informative attribute based on entropy:

$$H = - \sum_c P(c) * \text{Log}_2 p(c) \quad (2)$$

Where:

*H* is the entropy for classes

*p(c)* is the probability of a given class *c*.

Entropy for a certain attribute is calculated using the weighted sum of the class entropies for each attribute value.

$$H_{attr} = \sum_v p(v) \sum_c p(c|v) * \text{Log}_2 p(c|v) \quad (3)$$

where:

*H<sub>attr</sub>* : is the entropy for attribute.

*p(v)* :is the probability of a given attribute value.  
*P(c|v)*: is the probability of class (*c*) given attribute value (*v*).

- 2-Calculate information gain of an attribute using the following equation:

$$\text{Infogain}_{attr} = H - H_{attr} \quad (4)$$

## 4. EXPERIMENTAL RESULTS

Our experiments included training three different classifiers using data set of 800 Arabic news collected from Twitter. The ground truth data set were manually labeled as credible and non-credible. Five - fold cross validations were performed so that all data set were used for both training and testing to avoid over-fitting. Different diagnostics were recorded for each of the three classifiers [Decision tree(J48), Support vector machine(SVM) and Naive Bayesian classifier(NB)]. Diagnostics include precision, Recall and F-Measure, TPR, FPR and accuracy.

- a) **Precision**: it indicates the number of correctly classified news (True positives) in proportion to all classified instance (true positives + false positives). Mathematically, precision is calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

where:

*TP*: is the number of news correctly identified as credible.

*FP*: is the number of news incorrectly identified as credible.

- b) **Recall/Sensitivity**: which is used to indicate capability of the system to classify inputs correctly. Recall is expressed by the following equation:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

where:

*TP*: is the number of news correctly identified as credible.

*FN*: is the number of news incorrectly identified as non-credible.

- c) **F-measure**: a harmonic mean of both precision and recall and calculated as follows:

$$F - \text{measure} = 2 * \frac{\text{precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

- d) **False Positive Rate(FPR)**: used to indicate the ratio between false positives and the total

number of news that do not belong to the class *c* and this needs to be minimized. FPR can be calculated as followed:

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

- e) **Accuracy:** indicates the ability to differentiate the credible and non-credible cases correctly. It's the proportion of true positive(TP) and true negative(TN) in all evaluated news:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where:

*TP:* is the number of news correctly identified as credible.

*FP:* is the number of news incorrectly identified as credible.

*TN:* is the number of news correctly identified as non-credible.

*FN:* is the number of news incorrectly identified as non-credible.

Results of the performance measures for all the three classifiers during five iteration are shown in figure 4. The results indicate that performance of the decision tree achieves a higher ratios compared to SVM and NB as shown in Figure5.

The observed results show that the decision tree achieves higher rates in all diagnostics as it achieves: 90% TPR, 12.96% FPR, 90.28% Precision, 90.18% recall, 90.10% f-measure and 90.18% accuracy. While SVM achieves: 88.58% TPR, 21.16% FPR, 88.62% precision, 88.58% recall, 88.26% f-measure and 88.58% accuracy. Finally, Naive Bayesian classifier achieved: 83.80% TPR, 22.30% FPR, 84.32% precision, 83.86% recall, 83.84% f-measure and 83.80% accuracy.

TPR						FPR					
	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5		Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
Decision tree	0.899	0.92	0.97	0.92	0.8	Decision tree	0.108	0.08	0.056	0.203	0.201
SVM	0.899	0.92	0.9	0.91	0.8	SVM	0.108	0.069	0.474	0.206	0.201
Naïve Bayesian	0.78	0.88	0.87	0.92	0.74	Naïve Bayesian	0.208	0.115	0.328	0.206	0.258
precision						Recall					
	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5		Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
Decision tree	0.899	0.922	0.971	0.922	0.8	Decision tree	0.899	0.92	0.97	0.92	0.8
SVM	0.899	0.924	0.898	0.91	0.8	SVM	0.899	0.92	0.9	0.91	0.8
Naïve Bayesian	0.791	0.885	0.873	0.923	0.744	Naïve Bayesian	0.78	0.88	0.87	0.923	0.74
F-measure						Accuracy					
	Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5		Iteration 1	Iteration 2	Iteration 3	Iteration 4	Iteration 5
Decision tree	0.899	0.92	0.97	0.916	0.8	Decision tree	89.90%	92%	97%	92%	80%
SVM	0.899	0.921	0.886	0.907	0.8	SVM	89.90%	92%	90%	91%	80%
Naïve Bayesian	0.78	0.881	0.872	0.92	0.739	Naïve Bayesian	78%	88%	87%	92%	74%

Figure 4: Performance Results of the chosen classifiers.





Figure 5: Performance measures for the chosen classifiers

Decision tree achieves around 2% higher TRP than TRP and 7% than NB, approximately 9% lower FPR than SVM and 10% lower than NB. For precision, recall, f-measure and accuracy, decision tree is almost 2% higher than SVM and 7% higher than NB as shown in Figure 6.

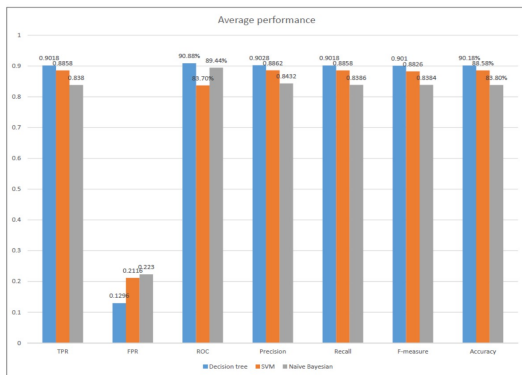


Figure 6: Average Performance For Three Classifiers

## 5. EVALUATION

A higher-level experiment was conducted to evaluate the performance of the proposed system using the chosen set of features compared to the performance using only features used by Hend et al. [29]. For this experiment, classifiers were trained using the data set with features in[29]. A data set of 320 piece of news were input for both systems for assessment. Results in figure 7 indicates that the accuracy achieved by the system using the chosen features outperforms the accuracy using only the features in[29] as shown in figure 8.

	Decision tree	SVM	Naive bays
HIND et. al	87.30%	61.30%	65.40%
The proposed system	89.90%	80.80%	78%

Figure 7: Accuracy Results Compared With Hend. Et. Al.[29]

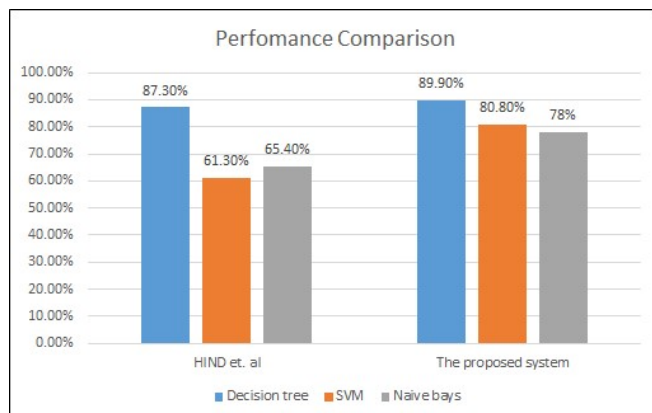


Figure 8: Accuracy Comparison With Hend. Et. Al.[29]

Another comparison was held between the performance of the proposed system and TweetCred[1] over the same data set. a data set of 320 piece of news were evaluated using TweetCred browser plug-in and were fed into our classification module. Credibility assessment obtained from the two systems were recorded and accuracy was calculated. Accuracy of the systems is shown in figure 9.

Results show that the accuracy of the proposed system is 89.9% where TweetCred only achieves 51.4%. This significant difference is due to the fact that TweetCred does not target news, it's a general purpose credibility assessment tool that collects features in real time. the issue is that the used features are not applicable to news (i.e. presence of stock symbol, presence of happy smiley, presence of sad smiley, presence of swear words, presence of negative emotion words, presence of positive emotion words, presence of pronouns, mention of self words in tweet (i; my; mine)). These set of features cannot be found in news as even in fake news, formal language is always used.

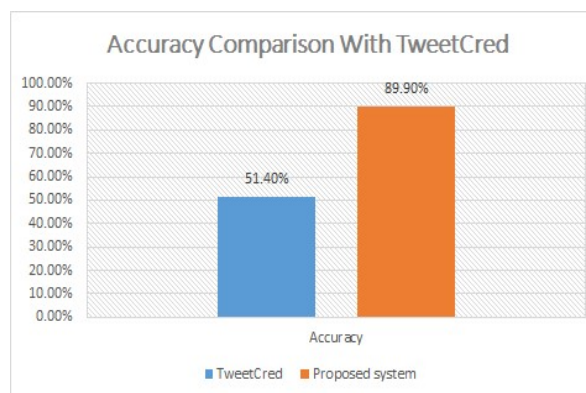


Figure 9. Accuracy Comparison With Tweetcred[1]

## 6. CONCLUSION AND FUTURE WORK

User generated content platforms such as Twitter enable users to author and publish their own generated content without prior expertise. These platforms are rich environments for incorrect, fake information and rumors. Credibility analysis is a necessity in today's social media platforms as they turned to a vital broadcasting source of information and news. In this paper, we presented a credibility assessment model for Arabic news credibility assessment on Twitter. Different set of features were used to assess credibility at content, post and source- level. The polarity of users' comments was added to the chosen features as a significant indicator for credibility assessment. The evaluation in comparison with other works in the literature shows higher accuracy of the proposed system. results showed that the proposed system outperforms two of the systems in the literature. However more experiments need to be conducted with larger data-set and the effect of adding/removing features need to be evaluated for more accurate performance.

## REFERENCES:

- [1] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, Patrick Meier, Tweetcred: "Real-time credibility assessment of content on twitter". In the proceedings of the international Conference on Social Informatics. Springer,2014.
- [2] Zhao, Z.; Resnick, P.; and Mei, Q., "Enquiring minds: Early detection of rumors in social media from enquiry posts", In the Proceedings of the 24th International Conference on World Wide Web, 1395-1405, 2015.
- [3] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, Anupam Joshi, "Faking Sandy:Characterizing and Identifying Fake Images on Twitter During Hurricane Sandy". In Proceedings of the 22nd International

- Conference on World Wide Web (WWW 13 Companion). ACM, New York, NY, USA, 2013.
- [4] Niall J Conroy, Victoria L Rubin, Yimin Chen. “Automatic deception detection: Methods for finding fake news”, Proceedings of the Association for Information Science and Technology, 52(1):14, 2015.
- [5] Wei Wei, Xiaojun Wan, “Learning to identify ambiguous and misleading news headlines”, in the Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), 41724178, 2017.
- [6] Klein, David O, Wueller, Joshua R., “Fake News: A Legal Perspective”, Journal of Internet Law (Apr. 2017). Available at SSRN: <https://ssrn.com/abstract=2958790>
- [7] Fan Yang, Yang Liu, Xiaohui Yu, Min Yang, “Automatic detection of rumor on Sina Weibo”. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics (MDS’12). ACM, New York, NY, USA, , Article 13 , 7 pages.
- [8] GALITSKY B., “Detecting Rumor and Disinformation by Web Mining”, AAAI Spring Symposium Series, North America, mar. 2015.
- [9] Eugenio Tacchini, Gabriele Ballarin, Marco L. Della Vedova, Stefano Moret, and Luca de Alfaro, “Some Like it Hoax: Automated Fake News Detection in Social Networks”, CoRR,abs/1704.07506, 2017.
- [10] Eons Seo, Prasant Mohapatra, Tarek Abdelzaher. “Identifying Rumors and Their Sources in Social Networks”, SPIE 2012.
- [11] Sebastian Tschatschek, Adish Singla, Manuel Gomez-Rodriguez, Arpit Merchant, Andreas Krause, “Detecting Fake News in Social Networks via Crowdsourcing”, CoRR volume abs/1711.09025, 2-17.
- [12] Jooyeon Kim, Behzad Tabibian, Alice Oh, Manuel Gomez- Rodriguez, “Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation”, In the Proceedings of the 11th ACM International Conference on Web Search and Data Mining (WSDM 2018).
- [13] Peter Bourgonje, Julian Moreno Schneider, Georg Rehm, “From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles”, Proceedings of the 2017 EMNLP Workshop on Natural Language Processing meets Journalism, pages 8489 Copenhagen, Denmark, 2017.
- [14] L. Mui, M. Mohtashemi, and A. Halberstadt, A Computational Model of Trust and Reputation for E-businesses, In Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS02)-Volume 7, Washington, DC, USA, IEEE Computer Society. 2002.
- [15] Jennifer Golbeck and James Hendler, Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks, In Engineering Knowledge in the Age of the Semantic Web , pages 116131. Springer, Berlin, Heidelberg, October 2004.
- [16] B. Thomas Adler and Luca de Alfaro. A Content-driven Reputation System for the Wikipedia, In Proceedings of the 16th International Conference on World Wide Web , WWW 07, pages 261270, Banff, Alberta, Canada, 2007. ACM.
- [17] Tabibian, B., Valera, I., Farajtabar, M., Song, L., Scholkopf, B., Gomez-Rodriguez, M, “Distilling information reliability and source trustworthiness from digital traces”. In proceedings of the International World Wide Web Conference Committee (IW3C2), perth, Australia. ACM, 2017.
- [18] Majed Alrubaian, Muhammad Al-Qurishi, Mabrook Al-Rakhami, Mohammad Mehedi Hassan, and Atif Alamri, “Reputation-based credibility analysis of Twitter social network users”, Concurrency and Computation Practice and Experience, 2016.
- [19] Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, Christos Faloutsos, “Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow”. In the Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Pages 886-893. 2013.
- [20] C. Castillo, M. Mendoza, and B. Poblete, “Information credibility on twitter”, In the Proceedings of the 20th international conference on World wide web, Hyderabad, India, 2011.
- [21] Akshay Java, Xiaodan Song, Tim Finin, Belle Tseng, “Why We Twitter: Understanding Microblogging Usage and Communities”, Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007, Springer, 2007.
- [22] Alexandra Olteanu, Stanislav Peshterliev, Xin Liu, Karl Aberer, “Web Credibility: Features Exploration and Credibility Prediction”, in the proceedings of European Conference on Information Retrieval. ECIR 2013: Advances in Information Retrieval pp 557-568, 2013
- [23] Ruohan Li, Ayoung Suh , “Factors Influencing Information credibility on Social Media Platforms: Evidence from Facebook Pages”, In the proceedings of the 3rd Information Systems International Conference (ISICO2015), 2015.

- [24] Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, Julia Schwarz, “Tweeting is Believing? Understanding Microblog Credibility Perceptions”, CSCW 2012, USA.
- [25] Kanda Runapongsa Saikaew, Chaluemwut Noyunsan, “Features for Measuring Credibility on Facebook. Information”. In the proceedings of the XIII International Conference on Computer Science and Information Technology (ICCSIT 2015), Thailand, 2015.
- [26] Canini, K. R., Suh, B., Pirolli, P. L., “Finding credible information sources in social networks based on content and social structure”, in Proceedings of the IEEE Second International Conference on Social Computing, SocialCom11, 18 (2011).
- [27] Fadi Salem, “The Arab Social Media Report series provides in-depth each year analysis on social media trends, growth and demographic breakdowns across 22 Arab countries”. 7th Edition of the Arab Social Media Report Feb 2017.
- [28] Rasha Mohammad Bin Sultan, Hend AlKhalifa, Abdul-Malik Al-Salman. “Measuring the credibility of Arabic text content in twitter”. In the proceedings of digital Information Management (ICDIM).2010
- [29] Hend S. Al-Khalifa, Rasha M. Al-Eidan, “An experimental system for measuring the credibility of news content in Twitter”, International Journal of Web Information Systems Vol. 7 No. 2, pp. 130-151. 2011.
- [30] Rim El Ballouli, Wassim El-Hajj, Ahmad Ghandour, Shady Elbassuoni, Hazem Hajj and Khaled Shaban, “CAT: Credibility Analysis of Arabic Content on Twitter”, Proceedings of The Third Arabic Natural Language Processing Workshop (WANLP), pages 6271, Valencia, Spain, April 3, 2017.
- [31] John ODonovan, Byungkyu Kang, Greg Meyer, Tobias Hollerer, Sibel Adal, “Credibility in Context: An Analysis of Feature Distributions in Twitter”, In the proceedings of the International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust, 2012.
- [32] Amal AlMansour, Costas S. Iliopoulos, “Using Arabic Microblogs Features in Determining Credibility”, Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2015.
- [33] Mohamed Hammad and Elsayed Hemayed, “Automating Credibility Assessment of Arabic News”, In the proceedings of the International Conference on Social Informatics (SocInfo) Social Informatics pp 139-152, 2013.
- [34] Mohammed N. Al-Kabi, Izzat M. Alsmadi, Amal H. Gigieh, Heider A. Wahsheh, Opinion Mining and Analysis for Arabic Language, International Journal of Advanced Computer Science and Applications (IJACSA), Vol.5, No.5, 2014.
- [35] Nora Al Twaresh, Hend Al Khalifa, AbdulMalik Al Salman, AraSenTi: Large Scale Twitter Specific Arabic Sentiment Lexicons, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 697705, Berlin, Germany, 2016.
- [36] Maha Althobaiti, Udo Kruschwitz, Massimo Poesio, AraNLP: A Java-based Library for the Processing of Arabic Text, In the Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.
- [37] Ian H. Witten; Eibe Frank; Mark A. Hall. “Data Mining: Practical machine learning tools and techniques”, 3rd Edition. Morgan Kaufmann, San Francisco. p. 191. 2011.