

Text Mining in Educational Objectives for Arabic Twitter Educator Accounts

Mohammed A. Alasmrai, Adel R. Alharbi
University of Tabuk, Tabuk, Saudi Arabia

Abstract

Twitter is a social networking site that offers micro blogging services to interact via Twitter posts, also called tweets, on any devices with Internet access. In Saudi Arabia, Twitter users occupy the second place worldwide. The vast majority of educators and learners students, providers, and educational organizations utilize Twitter, for facilitator-learner interactions and lifelong learning experiences. In the past few years, researchers have significantly studied online social networking sites to examine the relationship between educational outcomes and social networking. However, there is little research exploring integration of micro-blogging tool, Twitter, into adult and continuing education as well as facilitating for adult learners how to find the right twitter accounts based on learning outcomes. In addition, there is too much shared public and educational information, which is foolproof to spread unrelated content. Subsequently, this counterproductive and prevent adult learners from achieving the learning outcomes. This work purposes to develop a model that could classify Arabic tweets in adult educators' accounts into one of three types of learning outcomes: knowledge, skill, and affective in order to predict tweets belonging to each of these outcomes. Our results show the highest accuracy classification rates through applying support vector machine and random forest algorithms with above 0.75 for all learning outcomes.

1. Introduction

Twitter is regarded as one of the most famous social network platforms with users throughout the entire international community. Part of its appeal is that users can follow other users with a public profile [1]. Of the half of a billion tweets that are posted every day, over 4.2 million are associated with education [1]. These tweets range in content, significantly, and incorporate many aspects and issues pertaining to education. Commonly, adult educators use the platform to share ideas and resources, ask and answer questions, collaborate on problems, critique curricula and textbooks, and engage in many other practices using 140 characters or less, per the Twitter guidelines. Educational users of Twitter enjoy many educational-based tweets that

help them to become better, more well-rounded, and more informed teachers.

Muijs and Reynolds [2] note that, considering the utter importance of classifying educational objectives for the purpose of dictating the selection of instructional methods, media, and evaluation in the classroom setting, educators are able to manually classify tweets in terms of educational objectives as either cognitive, skill, or affective domains. However, using the method of educational data mining provides more classification results for a set of verbs than using manual-based system. According to the International Educational Data Mining Society [3], educational data mining is described as "an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data obtained from educational settings, and uses those methods to better understand students and the settings in which they learn." Basically, educational data mining serves to translate words and expressions into numerical values and qualities. Text analysis is complex and involves information retrieval, distributions, lexical examination to study word repetition realization, labeling annotation, data extraction, information mining methods including connection, and affiliation investigation, conception, and prescient examination. Although rather complex, the overall objective is clear and straightforward to convert content into information for investigation.

To be effective at teaching, educators must have a clear, concise, and precise idea of the information they want their students to learn. According to Muijs and Reynolds in [2], adult education requires highly specific skills and educators must be able to convey to their students precisely what each student is to take away from the course. One way this task is accomplished is through having goals and objectives for each topic, lesson, or course of study. Objectives ought to be written in such a way that learners are able to demonstrate their understanding. Furthermore, objectives need to be measurable to gauge the degree to which students have acquired the skills and information of the course or lesson. Krutka and Jeffrey [4] add that, in increased numbers, educators have relied on social media, such as Twitter, to engage students and shape students' professional and practical learning. Twitter is also commonly used as a platform for uniting educators with their peers for the purpose of exchanging ideas

and creating a solid community of educational professionals. For instance, some educators participate in weekly Twitter chats with their students and/or with their fellow educators. These social media networking activities serve to establish a strong online learning network which can augment educational discourse and be exciting. It is a means for allowing educators to become an active part of a passionate and collaborative collegial group of professionals.

In the past several years, there has been a substantial body of scholarly literature published on social networking sites such as Facebook, YouTube, Instagram, and Twitter [5]. The purpose of most of these investigations is to explore and analyze the relationship between educational outcomes and social networking. The literature which focuses on this relationship is rather detailed and a great deal is known concerning it. However, there is very little scholarly research published on exploring the integration of Twitter, a micro-blogging tool, into adult and continuing education. There is not nearly enough scholarly material concerning how adult learners can find the right Twitter accounts based on learning goals, objectives, and outcomes. Additionally, there is way too much shared public and educational information, which is likely to cause the spreading of unrelated content. It is counterproductive to adult learners and only serves to impede their search to achieve their learning outcomes [5]. Therefore, the purpose of this research is to help overcome these challenges and create a means for providing adult learners and educators with an effective, reliable, and efficient means of finding the most appropriate tweets for their specific needs.

There are some primary objectives associated with our research project.

- The first target is to collect a set of Arabic tweets data from several educator accounts.
- The second is to develop a model that can classify Arabic tweets by accounts held by adult educators into one of three different types of learning outcomes: cognitive (knowledge), psychomotor (skill), and affective (attitude) through using different classification algorithms such as Support Vector Machine and Random Forests.
- The final target is predicting tweets which belong to each of these three outcomes with above 0.75 accuracy classification rates.

The following sections present related work, Twitter and sentiment analysis in Arabic education and learning theories, the proposed methodology, experimental results, and the conclusion and future work.

2. Related Work

There is a wide and diverse body of related work that helps to shed light on the research purpose and objectives of this study. Murugiah [6] discusses some of the major learning theories applicable to adults and their application to synchronous learning tools. The purpose of this research is to understand how these tools better support stakeholders' needs and preferences while interacting online. The research reveals that, to improve adult learners' satisfaction, adult educators should be well attuned to their needs. They should be privy to the way in which learners make meaning and create learning from their unique experiences. To be successful at accomplishing these goals for adult learners, online synchronous learning tools have to be interactive, promote a conducive learning environment, and provide stakeholders with additional opportunities to construct knowledge.

Alqarni et al. [7] dedicated their research towards identifying and measuring the security and educational challenges Saudi society deals with. To do so, they analyze users' sentiments via Twitter. Arabic tweets that contained certain keywords were scanned and scrutinized. A model was proposed based on the tweets retrieved that was able to predict which tweets belonged to "the cross-cultural and ethics of dialogue and rules of difference challenges than those for the dominant negative social values challenge" [7]. The study's findings showed that tweet's sentiments analysis was reliable and, therefore, provided a more economical and faster alternative to traditional methods of gauging societal psychological wellness.

King [8] explored educators' experiences using Twitter as a network for professional learning. The study found that traditional learning for professionals is inadequate at addressing the needs of educators in the 21st global context. One of the major issues is that educators are becoming more isolated and do not have the social and collaborative networks in place they need to sustain best practices within the classroom context. Twitter and other social networking platforms were revealed to be excellent at overcoming these challenges and providing educators with the academic and social support they need to excel in their fields. Twitter was found to be the most appropriate because it allows for both passive and active interactions and is easy to search via keywords and hash-tags.

Leudy [9] explores the perceptions of adult students who are enrolled in some form of distance higher education. The theory used for this study was Adult Learning Theory, which provided the framework for the data analysis. The findings revealed that these students were comfortable using new devices for communication purposes provided that they had sufficiently clear and ample

instructions on how to use them, in addition to perceiving a need to use them. On the other hand, when students viewed the technology as too tedious or arduous to learn, they were less likely to accept it. The authors conclude that if educators want to introduce technologies to adult students in the classroom setting, then they should meet two criteria. First, the technologies have to be explained, in detail, to students so that they feel comfortable using them. Second, the students have to perceive a needed reason to learn how to use them, preferably one that will augment their learning.

Taylor and Hamdy [10] state that there are numerous theories and frameworks associated with adult learners and learning, each with their own merits. The authors propose a model which combines aspects of several theories into a flow diagram that is well-suited for planning purposes, such as curriculum planning or individual learning. Importantly, at each phase of the model, the proposed guide offers suggestions on the responsibilities of both the educator and the student. Basically, the guide maintains that the role of the institution of higher education is to ensure that both time and resources are amply available to promote effective learning. The model is useful because it combines the best aspects of each of the major adult learning theories and lays out responsibilities for all stakeholders.

Kidd [11] synthesized new perspective on online learning for adult students. The author explores the historical development of online learning, looking specifically at two contexts-business and education. The research showed that for nearly 30 years there have been varying degrees of online learning available to students, but the online opportunities have increased steadily over the past 15 years. For the most part, adult students greatly appreciate the ability to learn remotely because, typically, the classes are less expensive and are more flexible concerning time. Also, travel time is not a factor since learners do not have to leave their homes. One of the challenges for educators and students alike is overcoming the feeling of isolation. To help curb this challenge, it was revealed that online social networking platforms, such as Twitter, can be highly effective and beneficial for all stakeholders.

Dunlap and Lowenthal [12] point out that online learning must serve to facilitate a social process of acquiring new knowledge in order to be appropriate and effective. To accomplish this goal, educators must provide students with space, challenges, and opportunities to engage in social actions with each other and the educator. Based on these premises, the authors recommend "Twitter to encourage free flowing just-in-time interactions" [12]. Twitter provides many benefits to learners because it is organic and facilitates interactions, collaborations, and critiques of other posts. As a platform, it can

help to overcome some of the challenges and shortcomings associated with online learning because it promotes interactions and communications between all vested stakeholders.

Finally, Alzafiri [13] investigated the impact of web-based instruction on two types of learning: cognitive and psychomotor. The research hypothesized that there would be no appreciable difference between the two learning types. Moreover, it was predicted that there would not be significant interaction effects between the types of instruction and participant's gender. Both hypotheses were proven true suggesting that cognitive and psychomotor types of learning are helpful to students and the variable of gender studied does not impact outcomes.

3. Twitter and Sentiment Analysis in Arabic Education and Learning Theories

This section is dedicated to concern Twitter in education, adult learning theories, types of educational objectives, related verbs, the uniqueness of the Arabic language, and sentiment analysis. The adult learning theories to be explored include andragogy, transformative learning, and self-directed learning. The types of educational objectives to be discussed are cognitive domain, skill domain, and affective domain.

3.1. Twitter in Education

The research shows that educators are increasingly relying on social media platforms for teaching and learning purposes [14]. Deaton in [15] explains that one of the major benefits of social media in general and Twitter in particular is that it is flexible. Therefore, users can be passive, active, or both. Educators can use it as either a passive or active tool for learning. For example, learners can watch or surf (passive) or tweet and produce (active). Educators, because of the flexibility, are able to plan for a variety of learner roles or use it as a means of conducting individualized and personalized education.

Oh [5], another benefit of using Twitter for educational purposes is that it is a great source of primary and secondary data. It is authentic and consists of original, diverse, real-world content that adds a level of realism that is not possible through a textbook. Twitter offers a wide and diverse range of digital content that is excellent for primary research. Students cannot just read and repeat the information on Twitter (or any other social media site for that matter); rather, they must analyze, evaluate, synthesize, and make sense out of the raw materials. It is an exceptional source for providing young

researchers with unfiltered material for research purposes.

Also, Deaton [15] add that Twitter is familiar, and most students already know how to use it. Because it is so common, the students do not have to focus their attention on learning procedural knowledge. Instead, they can dedicate their efforts towards learning new content and tasks instead of worrying about format and layout of new programs and platforms. Moreover, Twitter accounts do not cost users any money and allow for customization which can make assignment differentiation easier.

According to Tess [16], another benefit of using Twitter for educational purposes is that it leaves a trail or a record. Old tweets can be archived for a variety of purposes, such as grading, referencing, or revisiting. It is also one of the most flexible platforms with many features that allow educators to easily modify it to meet their needs. Users can search by topics, hash-tags, and lists. It can also be utilized as a read-only account without requiring students to actually publish their own material. Due to these numerous benefits, the overarching flexibility of the platform, and the familiarity of it among many students, educators continue to use Twitter as a supplement for education and as a means of connecting with their peers and other education professionals.

3.2. Adult Learning Theories

Adult learning theories are excellent for helping individuals to understand the unique needs of people as they learn and make meaning out of the information being conveyed to them. Brockett and Hiemstra [17] explain that learners at all levels have their own tacit beliefs and worldviews and, then, apply them to the learning process. Within a non-profit organization, learners are required to create spaces for sharing of ideas with people from different generations, genders, sexes, socioeconomic statuses, races, and ethnicities. In this collective environment, learners are able to enrich their experiences with one another through conversations. This enables them to locate common qualities and aspirations which bring them together and serve to unite them. The two primary purposes of these forms of interactions are: to take advantage of the fact that adults are cognitive beings who have feelings, awareness, and creativity; and, to learn not simply for the sake of learning but, moreover, for the purpose of making meaning, understanding, and transforming.

The work by Illeris in [18] states that several learning theories are applicable for understanding distance learning. Some of the more common ones include Independent Study, transactional Distance Theory, Industrialization of Teaching, Equivalency Theory, and Interaction and Communication Theory.

However, because this research focus on the ways in which learners perceive communication when learning at a distance, adult Learning Theory [19] was selected, to learn more about the choices of adult learners' preferred adult educators' twitter accounts when they are engaged in collaborative efforts at a distance, as well as to incorporate and understand adult learning characteristics in interaction. The three adults learning theories to be explored are andragogy, transformative learning, and self-directed learning.

3.3. Andragogy

Andragogy dates back to the work of Malcolm Knowles in the 1980 who proposed the theory. According to Knowles andragogy is "the art and science of helping adults learn" [19]. It is, therefore, both a theoretical and practical framework for ensuring that adults learn in a welcoming, conducive, and age-appropriate environment. This theory includes six learner-centered guidelines utilized by adult educators. First, adults have to recognize that it necessary for them to learn something in order for them to put the time and effort into undertaking to learn it. In general, adults are not going to be complacent and just learn something if they do not have a direct reason for learning it.

Second, Smith in [19] adds that adults have to recognize that they are responsible for their own learning and each decision that they make in their lives. No one is responsible for their situation or position and, furthermore, they must take ownership of their education and not rely on their teachers to do the work for them. Third, adults who enter an educational environment mostly do so with more and different experience than younger people do. Most of the time, adults have worked in the real world, bought houses, raised families, and been in marriages. These life experiences place them in a different category than younger learners.

Fourth, Knowles et al. [20] show that adult have high level of eagerness to learn new things. They want to gain the knowledge and skillsets they need to cope effectively with their issues and problems experienced in the real world. They are not wasting their time in a classroom; instead, they are using their time to gain additional education-based resources than they had previously. Fifth, adults tend to be life-centered, task-centered, career-centered, and problem-centered. This applies to their orientation to learning and impacts how they tackle the challenges involved in the educational process. Fifth, most adults are highly responsive to some type of external-based motivation such as the prospect of a better job, promotion, or pay raise; however, the most important and potent motivations are internal and include a desire to improve their job satisfaction,

a desire for a higher quality of life, or a pursuit for more self-confidence and self-esteem.

Significantly, the work by Ozuah in [21] argue that Knowles' theory is directly applicable to web-based platforms and synchronous learning because it is focused on understanding the characteristics and needs of learners who are inter-acting with others. Adults who participate in a synchronous learning activity, such as white board, online group chats, or application-sharing, create a safe, inclusive, comforting, and respectful environment. The ambience is learner-centered and is appropriate for allowing learners to flourish and grow.

3.4. Transformative Learning

EW Taylor [22], the Transformative Learning Theory is attributed to Jack Mezirow who defined it as being "constructivist, an orientation which holds that the way learners interpret and reinterpret their sense experience is, central to making meaning and hence learning" [23]. The theory is dedicated to adult learning and looks at it predominantly within the context of higher education such as college or university. It is grounded on four basic principles. The first is that adults tend to exhibit two distinct types of learning.

The first is instrumental or cause and effect. The second type is communicative and based on feelings and emotions. The second premise is that learning involves change particularly to the individual's meaning structure. These include both perspectives and worldviews of the learner.

The work by EW Taylor in [22] explains that the third premise of Transformative Learning Theory is that change to structures of meaning occur through a process which involves reflection. Learners take the time to reflect on content, course material, processes, and even premises introduced to them by the educator, class readings, and other initiatives. The final premise is that learning is complex and dynamic. Therefore, it involves rethinking, refining, and elaborating on meaning schemes. It also entails learning and transforming new schemes as well as perspectives.

3.5. Self-Directed Learning

Self-Directed Learning Theory studies the relationship between adult learning and self-directed learning. Mezirow [23] notes that "no concept is more central to what adult education is all about than self-directed learning." According to Knowles, self-directed learning is "a process in which individuals take the initiative without the help of others in diagnosing their learning needs, formulating goals, identifying human and material resources, and evaluating learning outcomes." [19]

Brockett and Hiemstra [17] add that adult learners, within this theoretical context, are any person who is performing social rules that are assigned, normally, by a culture to those who are considered to be adults. An adult learner recognizes that they are responsible for their lives and learning outcomes. According to them, self-directed Learning Theory is built around five key assumptions. Firstly, adults are motivated to learn when they experience needs and interests that learning will satisfy. Secondly is that adults' orientation to learning is centered on their lives. Thirdly is that the experience is, for adult learning, the richest and most valuable source of education. Fourthly is the adults have a deep and abiding internal need to be self-guiding and self-directing, autonomous entities. Finally, individual differences among adults, per this theory, increase with age. They also explain that the first three of the five assumptions highlight that adults typically possess a highly pragmatic orientation and approach towards the educational process and learning in general. The fourth assumption is what is most important for this research project because of its obvious relation to self-directed learning. The last assumption is also key and demonstrates the internal value of self-directed learning where there is variation in both the form and the speed that learning takes on.

3.6. Types of Educational Objectives

There are various types of learning and, based on these forms of learning, there are specific goals and objectives associated with them. These outcomes are referred to, often-times, as the three different types of learning and having a deep and holistic understanding of them is vital to facilitating synchronous learning. Adams [24], the three types of learning are cognitive domain, skill domain, and affective domain. The following looks at each of these three domains and their related verbs.

1) Cognitive Domain: Adams [24] notes that learning involves the ability to discern certain concepts, ideas, and rules. For instance, one of the overarching objectives might be to oversee synchronous chats and content during a web-based meeting or conference. This learning must be delivered in a suitable way, such as delivering it through a web-based synchronous learning tool. This is because this method is appropriate for communicating to the learner via the use of language, texts, numbers, and symbols.

2) Skill Domain: Adams [24] explains that learning includes the translation of mental ideas and concepts into tangible, physical actions. For instance, an educator might use shared whiteboards along with web-based chat tools simultaneously to communicate with students. This type of learning can be supported by synchronous learning applications in a web-based

environment by providing verbal directions, reinforcing practice for mastering the use of the keyboard, understanding how to use the cursor and audio, and the list goes on.

Table 1. Knowledge domain sample verbs

Arabic Verb	English Verb	Verb in English letters
يعرف	define	yerrf
قائمة	list	yueadu qayima
يبين	state	yubin
يسمى	name	yasmi
يستذكر معلومة	Recall info.	yastadei maelumatan
يكبر	repeat	yakarur
يشير إلى	point out	yushir 'iilaa
يخبر	tell	yukhbir
يلخص	summarize	yalkhus
يستشهد	cite	yastashhid
يحدد	identify	yuhadid
يشرح	explain	yashrah
يتعرف على	recognize	yataaraf ealaa
يناقش	discuss	yunaqish
يصف	describe	yasf
يعيد توضيح	restate	yeed twdih
يترجم	translate	yutarjam

Table 2. Skill domain sample verbs

Arabic Verb	English Verb	Verb in English letters
يصنف	differentiates	yasnf
يميز	distinguishes	yamiz
ينسخ	copies	yansukh
يعرض	shows	yuearid
يكتب	write	yaktub
يتحرك	moves	yataharak
يعرض	displays	yuearid
يجيب	responds	yujib
يعيد إنتاج	reproduce	yueid 'iintaj
يتفاعل	react	yatafaal
يتابع	follows	yutabie
يتبع أثر	traces	yatabie 'athar
يصنع	originates	yasnae
يصمم	designs	yasmim

3) Affective Domain: According to Adams [24], this final category implies behavior which is based on an underlying set of attitudes or beliefs. For instance, one goal of interaction through synchronous chat is to use tools such as hand raising, instant surveys, yes / no polls, and the list goes on. These tools serve to reinforce these behaviors and provide students with verbal guidance in synchronous learning.

Table 3. Affective Domain Sample Verbs

Arabic Verb	English Verb	Verb in English letters
يقر بالعرفان	acknowledge	yuqaru bialearfan
يشارك معلومة	Shares info.	yusharik maeluma
يقدر	appreciates	yuqdr
يعتز به	cherish	yaetazu bihi
يثمن	treasure	yuthmn
يبادر بـ	initiates	yubadr bi
يقدم دعوة	invites	yuqadim daewatan
ينضم	joins	yandam
يبرر	justifies	yubarir
يقدم مقترح	proposes	yuqadim muqtarah
يحترم	respect	yahtarim
يجامل	courteous	yajamil
يحبس بواجب	dutiful	yahs bwajib
يتابع باهتمام	follows	yutabie biaihitam

3.7. Related Verbs

According to various researchers in [25, 26, 27, 28], the important element of an educational objective is the action verb that identifies the performance required. When creating and recognizing objectives, the more precise the action verb the better. When choosing verbs for educational objectives, the emphasis is on verbs they must be unambiguous and specific to be measure accurately. The three domains of educational objectives are: cognitive (knowledge), psychomotor (skill), and affective (attitude). The tables 1, 2, and 3 show the sample of related verbs for each domain that were used in our project. The verbs were chosen based on Bloom's work for taxonomy of educational objectives in [24, 25].

3.8. The Uniqueness of the Arabic Language

The structure and arrangement of the Arabic language is vi-tally different comparing to other language families, including English, Spanish, and Chinese. First, the writing rule is based on the standard of 'right to left', while the formula depends on their place in a sentence. Second, gender plays a focal role in formative the affixes of words. Third, capital letters do not subsist in the Arabic language and there are syntactic standards that expect to distinguish units,, and acronyms [29]. Every one of these complexities bring about challenges in in Arabic text analysis, which are additionally combined with the absence of research in that field.

3.8.1. Sentiment Analysis. Sentiment analysis (SA) is part of Natural Language Processing (NLP), a field of computer science and artificial intelligence that deals with interactions between people and computers [30, 31].

SA has three primary levels: the level of the document, the level of the phrase, and the level of aspect. The primary job at the document level is to identify whether a favorable or a negative polarity is expressed throughout the opinion paper. The sentence level deals with the feeling of each phrase [32]. The aspect level is about extracting feelings from various parts of objects [33]. Most sentiment analysis studies focus on English-written text using the method of sentiment lexicons. Using the lexicon method to other languages, however, will cause an issue of domain adaptation [34]. Also, for Arabic sentiment phrases, there is no lexicon and therefore machine learning or corpus-based method is generally used. Both techniques are controlled methods in which a set of information is first classified as adverse, positive or neutral and appears through vectors of features. The classifier then utilizes these vectors as training data to recognize comparable characteristics and group the information in a certain way [35].

4. Methodology

Our methodology was implemented using the R machine learning and statistical software tool set. R is an open-source programming language that has several ML algorithmic packages and libraries [36]. Figure 1 presents the process of this work approach from obtaining the Arabic sentiments from Twitter for the educator accounts based on the educational objectives, then collect the tweets data into the data

set. The first step after gathering the data, the data per-processing step should take place to clean the data from the white spaces, emotions, #, URL, and the stop words. The send step should the data transformation in order to convert the data to matrix and partition it into containers to fed them to the classification algorithms. Finally, part of the data was used for train the classification model and the rest were used for testing the model to predict the types of educational objectives.

4.1. Data Collection

1) Construct Data set: To gather the Arabic tweets, an application was created to snatch tweets from around 48 Arabian accounts were used for education purposes. Some of these accounts were for educational organization accounts, and others were for well-known educators in the media for example: 3bdullahAl3amri, academicres3, ArabAcademia, profAMAl-hussein, and psych4yourchild. The data was collected by using a developed Python program that implemented the Twitter API method for data streaming. Each of the collected accounts were have more than a hundred tweets. For the purpose of our project, we collected only the text feature excluding any other features such as: the time the tweet was posted, when the account was created, the number of followers and friend, and the number of re-tweets. We gathered total of 142,183 observations.

2) Labeling: To make our research problem as a supervised learning, we had to label the collected

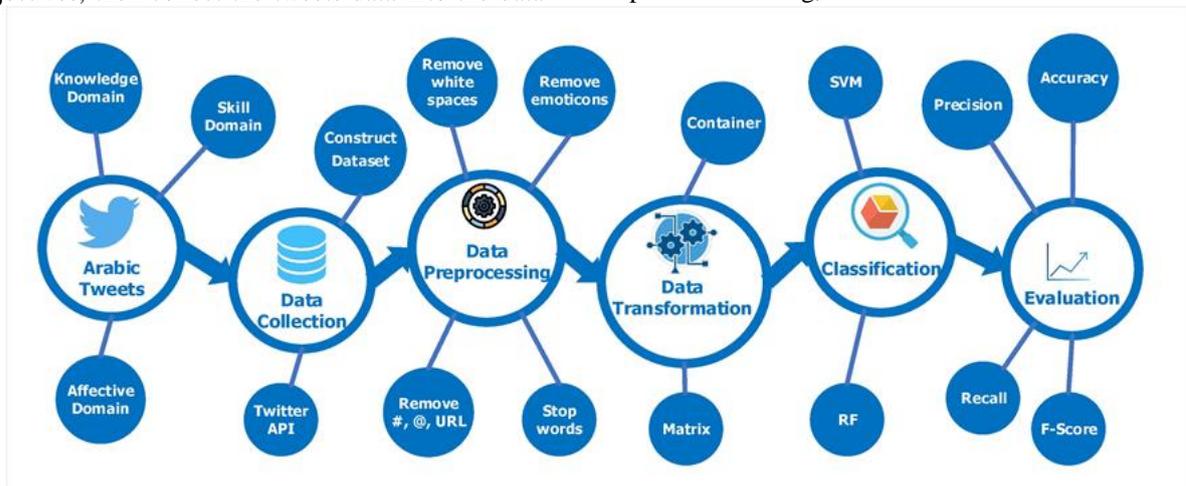


Figure 1. Methodology Overview Processes

tweets in order to classify them based on their educational objective aspects. Our approach developed a simple programmed function that able to find a particular word in a tweet. If a word is existed in single tweet the function return a true value. Otherwise, it will return a false value. We used this function to scan over the tweets of each individual

educator tweets based on the educational objectives terminologies, then label the tweets, accordingly.

4.2. Data Pre-processing

It is likely that pre-processing methods could be used to standardize certain tweet tokens due to the variable and un-predictable nature of the language

used in tweets. Most tweets are extremely probable to contain some type of grammar or spelling errors, acronyms, colloquialisms, and slangs. Our pre-processing phase of information includes the following significant steps before passing the text to the classifier as:

1) Removing characters and spaces: special characters, numbers, new line chars, and white spaces are unnecessary for the classification.

2) Removing re-tweets: sometime twitter users would like to resend other user messages which called re-tweets which creates the (RT) symbol within the tweet text. All the (RT) symbol are removed from the tweets.

3) Removing emoticons: emoticons are used in micro-blogging to express users' emotions in a brief way. All the emoticons are removed to have clean text for the classification.

4) URL extracting: URLs are used to share more explanations than can be given by a short message. Some trials replace this with an equivalence class < URL > to reduce the task size while others simply remove URLs from the document.

5) Detection of pointers: using the () token in front of the username, messages on Twitter can point to another user. In addition, to categorize the post, a hashtag with a (#) symbol is used.

6) Identification of punctuation: in micro-blogging, punctuation is used to prevent proper grammar and to facilitate the communication of feelings. Removing punctuation from the text will lower the training set's redundant function.

7) Removal of stop words: deleting undervalued significant phrases like "a," "an" and "the" in the classification process is a popular technique.

4.3. Classification Methods

1) Support Vector Machine (SVM): SVMs are the class of algorithms that are based on kernel substitution. They can be characterized as systems which use hypothesis space of linear functions in a high dimensional element space. This will be trained with a learning algorithm that executes a learning bias derived from statistical learning hypothesis. It is conceivable to build exceptionally nonlinear order classification technique utilizing SVMs without stalling out in local minima [37].

2) Random Forest (RF): RF comprises of numerous classification trees known as tree classifiers, which are utilized to predict the class dependent on the categorical dependent variable [38]. Each tree gives a class for the input vector and the class with most noteworthy turns will be picked. This current classifier's error rate relies on the correlation between any two trees in the forest and the quality of every individual tree in the forest. Hence, to limit the error rate the trees ought to be solid and independent of one another [39].

4.4. Data Transformation

Before implementing the data by machine learning classifiers, we are required to represent the text in a format that suitable for the classifier to deal with it, therefore the text is transformed into a document-term matrix. It is probably the most common format to represent texts for computation. It applied to generate a document-term matrix [40]. The document-term matrix is then partitioned into a container, which is essentially a list of objects that will be fed to the machine learning algorithms. The output is of the container includes separate train and test sparse matrices. In our case, the text of tweets was converted to create document-term matrix, then the container had a feature vector of document labels taken from the data set. Moreover, train and test sizes were partitioned indicating which documents to put into the training set and test set, respectively [41].

5. Experimental Results

This work investigates the classification of Arabic text as an educational objective patterns with various algorithms such as SVM and RF based on the collected data set. All classifiers' parameters such as the seed and tune length number were set up under the same situation to avoid biasing with respect to those parameters. The results were based on several classification metrics as follow:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Where: (TP) is the True positive, (FP) is the False positive, (TN) is the True negative, and (FN) is the False negative.

5.1. Model Validation and Comparison Results

This experiment carryout the evaluation of each classifier according to the educational objectives based on ten-fold cross validation strategy. Figure 2

shows the accuracy results compression between SVM and RF classifiers. SVM classifier performed 0.81, 0.70, and 0.72, while RF classifier performed 0.84, 0.80, and 0.75 classifiers for knowledge, skill, and affective objectives, respectively. RF classifier achieve higher rates than SVM classifier for all educational objectives. These results guided our decision to make the RF classifier be the chosen classifier.

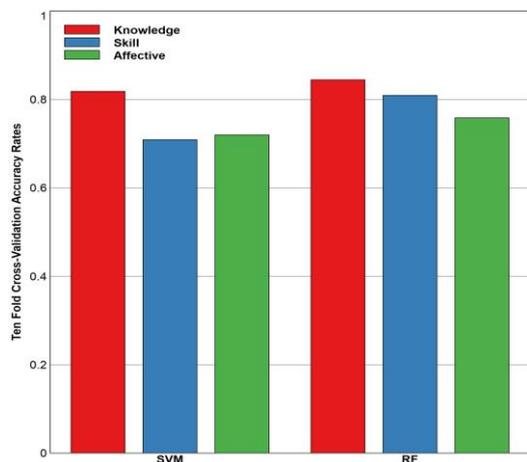


Figure 2. Classification Models Compression Accuracy Results

5.2. Different Percentage of Training Set Results

One of the important validation strategies is to evaluate the classification model in real situation, where the classifier is trained in part of the date set and being tested in the remained part. In this experiment, we trained the RF classifier into various percentage of the data from 0.1 until 0.9 and test it individually for each time.

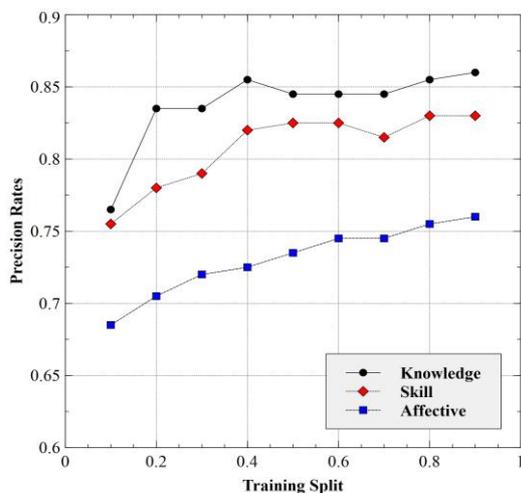


Figure 3. Precision rate results

Figure 3 presents the precision, recall, and f-score rate results for each educational objectives in different percentage of training sets. In Figure 3, the knowledge objective performed the highest precision result rates compering to the other two objectives with 0.86 in the trained split of 0.9. The affective objective performed lower starting with 0.68 in the trained split of 0.1 until 0.75 in in the trained split of 0.9.

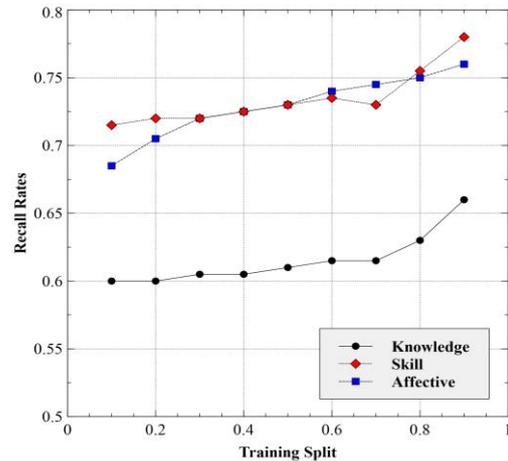


Figure 4. Recall rate results

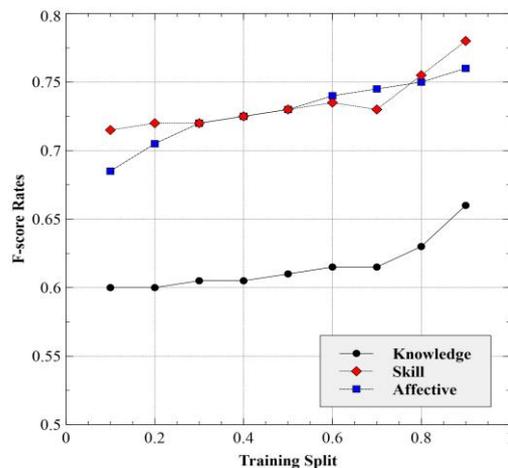


Figure 5. F-score rate results

Figures 4 and 5 show the recall and F-score performance results. Both figures generated a highly comparable plot results. The skill and affective objective performed higher than knowledge objective with about 0.76 and 0.75 in trained split of 0.9, meanwhile the knowledge started up with 0.6 in the trained split of 0.1 with little of improvement to ended up with about 0.67 in trained split of 0.9.

5.3. Different Number of Verbs

This experiment studies the impact of the number of verbs on the classification model predictions. In our work, the number of words of each educational objectives were as: knowledge (77), skill (256), and affective (295). The experiment was build based on

single user tweets, which about 3135 observations and a five cross-validation in order to calculate the accuracy rates for various number of affective verbs, where it starts with 10 verbs until reaches the full collected verbs (251). Figure 4 shows the accuracy rates achieved the highest in less verbs (10) with 0.99 accuracy rates, then start to decrease as long as the number of the verbs are increased until 0.77 rate. The reason behind of these decreasing rates is the classification model can get confused with many of verbs because it generates large predictions for these verbs. A good exercise is to have considerable verbs that suited for the size of a particular data set.

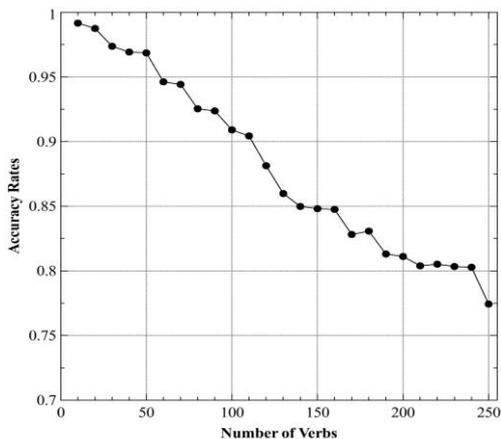


Figure 6. Accuracy Rates with Different Number of Verbs

6. Conclusion and Future Work

This paper addresses the concept and implementation of a characterization of Arabic text for the purpose of predicting and classifying the Twitter accounts based on their educational objectives such as: knowledge, skill, and effective. The data is collected from about 48 twitter Arabian educator accounts. Our purposed method started with cleaning the collected text, then used various algorithms such as: SVM and RF. We conducted several experimental results. The first experiment compares between two classification models were RT perform better than SVM classifier with above 0.75 for the all learning outcomes based on educator Arabic tweets. The second experiment used different percentage of training sets started from 0.1 until 0.9 for the RT classifier based on three classification metrics: precision, recall, and F-score rate results for each educational objectives. The precision rate results perform the best when educational objective is knowledge with 0.86 accuracy rates. Where the recall and F-score rate results perform the best when educational objective is skill with about 0.76 accuracy rates. The third experiment shows results the RF classifier performed with different number of affective verbs where it was build based on large

single user tweets. The accuracy rates is decreasing as the number of verbs is increased from 0.1 until 0.77. This happens because the classifier generates large predictions for the verbs.

In future the work, we intend to collect more text data size in order to refine our results. One of the improvements could be is to use the ensemble learning which is a combination of multiple classifiers to boost the overall predictability. In terms of educational objectives and related verbs, a future study could be a comparison between the classifier results and educational experts' perspectives on tweet text classification. Knowing that data are capable of yielding direct value and are useful across all levels of the education system is critical to ensuring data quality. In fact, the extent of data use at least partly determines data quality. After all, the more useful data is to an adult educator, the more incentive there is for that adult educator to ensure that the data is produced in an accurate and timely manner. In addition, the machine learning model can provide an automated way of solving problems like classifying learning objectives. It helps beginning adult educators to classify and verify the learning objectives of their contents in twitter. Finally, a similar study could be conducted within the kingdom of Saudi Arabia standards of learning outcomes verbs.

7. Acknowledgment

The authors would like to thank the University of Tabuk for their support and helpful comments.

8. References

- [1] T. Statista, "Twitter statistics & facts," 2017.
- [2] D. Reynolds and D. Muijs, *Effective teaching: Evidence and practice*. SAGE, 2005.
- [3] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12–27, 2013.
- [4] D. G. Krutka and J. P. Carpenter, "Participatory learning through social media: How and why social studies educators use twitter," *Contemporary Issues in Technology and Teacher Education*, vol. 16, no. 1, pp. 38–59, 2016.
- [5] J. Oh, "Educators and social media: learning motivators in creative media courses," *Asia Pacific Journal of contemporary education and communication technology*, 2018.
- [6] S. S. Murugiah, "Adult learning theories and their application in select-ing the functionality of synchronous learning tools." Online Submission, 2005.
- [7] H. A. Alqarni, Y. AlMurtadha, and A. O. Elfaki, "A twitter sentiment analysis model for measuring security and educational challenges: A case study in Saudi Arabia." *JCS*, vol. 14, no. 3, pp. 360–367, 2018.

- [8] V. King, "A little birdy told me: Educators-experiences with twitter as a professional learning network," 2017.
- [9] S. J. Lendy, "Distance learning: adult learners and computer-mediated communication," 2014.
- [10] D. C. Taylor and H. Hamdy, "Adult learning theories: Implications for learning and teaching in medical education: A mee guide no. 83," *Medical teacher*, vol. 35, no. 11, pp. e1561–e1572, 2013.
- [11] T. T. Kidd, *Online education and adult learning: new frontiers for teaching practices: new frontiers for teaching practices*. IGI Global, 2009.
- [12] J. Dunlap and P. Lowenthal, "Tweeting the night away: Using twitter to enhance social," 2009.
- [13] F. M. Alzafiri, *An experimental investigation on the effects of Web-based instruction/training on cognitive and psychomotor learning*. University of North Texas, 2000.
- [14] A. Al-Bahrani, D. Patel, and B. J. Sheridan, "Have economic educators embraced social media as a teaching tool?" *The Journal of Economic Education*, vol. 48, no. 1, pp. 45–50, 2017.
- [15] S. Deaton, "Social learning theory in the age of social media: Implications for educational practitioners." *Journal of Educational Technology*, vol. 12, no. 1, pp. 1–6, 2015.
- [16] P. A. Tess, "The role of social media in higher education classes (real and virtual)—a literature review," *Computers in human behavior*, vol. 29, no. 5, pp. A60–A68, 2013.
- [17] R. G. Brockett and R. Hiemstra, *Self-direction in adult learning: Perspectives on theory, research and practice*. Routledge, 2018.
- [18] K. Illeris, *Contemporary theories of learning: learning theorists... in their own words*. Routledge, 2018.
- [19] M. K. Smith, *Malcolm Knowles, informal adult education, self-direction and andragogy*. Infed, the Encyclopedia of informal education, 2002.
- [20] M. Knowles, E. Holton, and R. Swanson, "The adult learner woburn," MA: Heinemann-Butterworth, 1998.
- [21] P. O. Ozuah, "First, there was pedagogy and then came andragogy," *Einstein journal of Biology and Medicine*, vol. 21, no. 2, pp. 83–87, 2016.
- [22] E. W. Taylor, "Transformative learning theory," in *Transformative learning meets building*. Brill Sense, 2017, pp. 17–29.
- [23] J. Mezirow, *Transformative dimensions of adult learning*. ERIC, 1991.
- [24] N. E. Adams, "Bloom- taxonomy of cognitive learning objectives," *Journal of the Medical Library Association: JMLA*, vol. 103, no. 3, p. 152, 2015.
- [25] B. S. Bloom, M. D. Engelhart, E. J. Furst, W. H. Hill, D. R. Krathwohl et al., "Taxonomy of educational objectives: the classification of educational goals. handbook i: cognitive domain. new york: David mckay company," Inc. (7th Edition 1972), 1956.
- [26] D. R. Krathwohl, "Taxonomy of educational objectives: The classification of educational goals," *Affective Domain*, 1964.
- [27] R. Dave, *Developing and writing behavioral objectives*. Educational Innovators Press, 1970.
- [28] A. J. Harrow, *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*. Addison-Wesley Longman Ltd, 1972.
- [29] G. Fu and X. Wang, "Chinese sentence-level sentiment classification based on fuzzy sets," in *Proceedings of the 23rd international conference on computational linguistics: posters*. Association for Computational Linguistics, 2010, pp. 312–319.
- [30] B. Liu et al., "Sentiment analysis and subjectivity." *Handbook of natural language processing*, vol. 2, no. 2010, pp. 627–666, 2010.
- [31] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information retrieval*, vol. 12, no. 5, pp. 526–558, 2009.
- [32] N. Farra, E. Challita, R. A. Assi, and H. Hajj, "Sentence-level and document-level sentiment mining for arabic texts," in *2010 IEEE international conference on data mining workshops*. IEEE, 2010, pp. 1114–1119.
- [33] H. Zhou and F. Song, "Aspect-level sentiment analysis based on a generalized probabilistic topic and syntax model," in *The Twenty-Eighth International Flairs Conference*, 2015.
- [34] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent systems*, vol. 28, no. 2, pp. 15–21, 2013.
- [35] A. Shoukry and A. Rafea, "Sentence-level Arabic sentiment analysis," in *2012 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2012, pp. 546–550.
- [36] N. Matloff, *The art of R programming: A tour of statistical software design*. No Starch Press, 2011.
- [37] K. P. Bennett and C. Campbell, "Support vector machines: hype or hallelujah?" *Acm Sigkdd Explorations Newsletter*, vol. 2, no. 2, pp. 1–13, 2000.
- [38] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [39] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*. IEEE, 2012, pp. 182–188.
- [40] D. Meyer, K. Hornik, and I. Feinerer, "Text mining infrastructure in r," *Journal of statistical software*, vol. 25, no. 5, pp. 1–54, 2008.
- [41] L. Collingwood, T. Jurka, A. E. Boydston, E. Grossman, W. van Atteveldt et al., "Rtexttools: A supervised learning package for text classification," 2013.