

2016 Olympic Games on Twitter: Sentiment Analysis of Sports Fans Tweets using Big Data Framework

Azam Seilsepour¹, Reza Ravanmehr², Hamid Reza Sima³

1,3- Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, IRAN.

2- Department of Computer Engineering, Central Tehran Branch, Islamic Azad University, Tehran, IRAN.

(r.ravanmehr@iauctb.ac.ir)

Received (2019-01-05)

Accepted (2019-05-04)

Abstract: Big data analytics is one of the most important subjects in computer science. Today, due to the increasing expansion of Web technology, a large amount of data is available to researchers. Extracting information from these data is one of the requirements for many organizations and business centers. In recent years, the massive amount of Twitter's social networking data has become a platform for data mining research to discover facts, trends, events, and even predictions of some incidents. In this paper, a new framework for clustering and extraction of information is presented to analyze the sentiments from the big data. The proposed method is based on the keywords and the polarity determination which employs seven emotional signal groups. The dataset used is 2077610 tweets in both English and Persian. We utilize the Hive tool in the Hadoop environment to cluster the data, and the Wordnet and SentiWordnet 3.0 tools to analyze the sentiments of fans of Iranian athletes. The results of the 2016 Olympic and Paralympic events in a one-month period show a high degree of precision and recall of this approach compared to other keyword-based methods for sentiment analysis. Moreover, utilizing the big data processing tools such as Hive and Pig shows that these tools have a shorter response time than the traditional data processing methods for pre-processing, classifications and sentiment analysis of collected tweets.

Keywords: Big Data, Sentiment Analysis, Hadoop, Social Network, Twitter.

How to cite this article:

Azam Seilsepour, Reza Ravanmehr, Hamid Reza Sima. 2016 Olympic Games on Twitter: Sentiment Analysis of Sports Fans Tweets using Big Data Framework. J. ADV COMP ENG TECHNOL, 5(3) Summer : 143-160

I. INTRODUCTION

One of the new trends in the field of the computer is big data. Today, big data analytics is one of the fundamental challenges in computer science. From the early years of computer science, researchers have attempted to provide optimal and appropriate methods for analyzing the data. So far, many methods and approaches have been proposed for big data analytics such as data mining techniques, statistical methods, machine learning approaches and etc.

One of the new frameworks introduced to process and analyze the big data is Apache Hadoop. This framework makes it possible to divide a job into smaller pieces, then process them on different nodes, and eventually display the final results of the nodes in the output. In addition to introducing the Hadoop framework, Apache has introduced another software called Hive. The Hive software installs on a Hadoop framework and provides different analytics tools. In the era of social communications, people are keen to engage, share and collaborate through social networks, blogs, wikis, and other online Media. In recent



This work is licensed under the Creative Commons Attribution 4.0 International Licence.

To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>

years, the swarm intelligence has spread to many different regions with a particular focus on the areas of daily life, such as commerce, tourism, education, and health, and expands the size of the social web. Extracting the knowledge of such a large amount of unstructured information is a very difficult task. Taking into account the feelings that social network users are experiencing, they can be used to analyze various events. One of these events was the 2016 Olympics, which has attracted many people to social networks. The Twitter social network, which is one of the most widely used social networks and generates a huge amount of data every day, has been used in this research. Tweets of fans about the Iranian athletes have been collected and analyzed during the one-month period of the 2016 Olympic Games. In this paper, a new big data framework for clustering and extraction of information is presented to analyze the sentiment of the tweets collected from Twitter. The proposed method is based on keywords and polarity determination. The Hive tool on the Hadoop environment has been utilized to cluster the data, and the Wordnet and SentiWordnet 3.0 tools are used to analyze the sentiments of fans of Iranian athletes. For this purpose, 2077610 tweets were collected both in English and Persian about three Iranian athletes in the Olympic and Paralympics 2016. The remainder of this paper is organized as follows: in Section 2, the fundamentals of this research will be provided, and the proposed big data approach for 2016 Olympic sentiment analysis is explained in Section 3. The simulation results and evaluations will be presented in Section 4. Conclusions and future works will be provided in Section 5.

II. RESEARCH BACKGROUNDS

The term big data has long been used to refer to large volumes of data stored and analyzed by large organizations such as Google or NASA. Recently, this term has been used to denote large data sets that are so immense and bulky that cannot be managed by traditional management tools and databases. Big data is a new trend in computer science, which has attracted much attention over the past few years. It is possible to discover and use patterns by searching within these big data [1].

1. Twitter

In the proposed method, Twitter knowledge base was used to analyze sentiments of similar big data. Twitter is a social network created in 2006. A feature of Twitter is that it allows users to send up to 140 characters, text messages or videos, photos, and audios in the shortest possible time, which are called Tweets. Twitter had 41 million users in 2009, and growing popularity of social networking has led to a continuing increase in the number of users of this network. Many users generate large amounts of data every day, which has led many researchers to concentrate on processing such data.

2. Big Data

The term big data was first introduced in the late 1990s by scientists who were not able to store and analyze the growing amounts of data generated by digital technology. In 2005, big data became a research field in large companies such as Google, Yahoo, Amazon, and Netflix, as they had huge amounts of web-based data. Along with these matters, RFID devices and related equipment were developed for faster processing of input data. These trends led to the introduction of MapReduce programming model in 2004. In 2008, Apache Corporation developed the Hadoop project, a parallel processing system for big data in cluster form using MapReduce programming model in a high level. In 2012, Gartner provided a more precise definition of the concept of big data as follows: "big data includes high volume, speed, and variety of information that requires a new form of processing for decision-making, discovering perspective, and optimal processing" [2].

3. Sentiment Analysis in Big Data

Sentiments analysis is usually associated with social media and is widely related to big data. For example, Twitter is a popular microblog for people around the world. This social network is an event-driven network in which the users report their status on a regular basis. Analysis of sentiments in big data on a specific topic can be used as a process for reviewing text or speech to find comments, viewpoints, or feelings of the author or speaker. All the words in this section are strongly related to sentiments and describe highly subjective and ambiguous concepts. It can

easily be concluded that the sentiments depend on the context and scope of review. An automated effort to extract information from big data using a computer system increases complexity because software requires specific boundaries to eliminate ambiguity. In contrast, a Tweet lacks sentence structure, uses colloquial words, including emoji (☺) and sometimes words with repeating letters to enhance the sentiment ("I looooooooooooooooooove chocolate"). Also, there is a higher probability of typos due to the nature of devices used to create Tweets [3]. In the following, the application of sentiment analysis in big data is explained. Detection of tendency is of importance for large enterprises since knowing the tendency of customers can change the perspective of an enterprise. People's orientations will also be useful for cultural and political institutions, and parties exploit the process of changing users' tendency for their own ends. Generally speaking, understanding the tendency of people is important to everyone: to advance their economic goals, to achieve their political aims, and to plan for the future. In recent years, this enormous amount of data has turned into a hotbed for data mining research to discover facts, trends, proceedings, and even predictions of events. In some studies, it has been attempted to analyze the sentiments and emotions of Twitter users. Discovery of trends means to detect the interests of users in different periods. A majority of tendency detection methods attempt to find the tendency using keywords in user posts. A sudden increase in the use of a group of words will be indicative of a new tendency or occurrence of a new event [3].

III. RELATED WORKS

Birjali et al. combined machine learning algorithms with semantic sentiment analysis to predict suicidal ideation using Twitter4j to extract the data from Twitter as well as an algorithm of computing semantic analysis based on WordNet [4]. Then, Weka tool was employed to perform machine learning algorithms. The experimental result showed that the proposed method can identify suicidal ideation. Öztürk et al. investigated the public opinions and sentiments about the Syrian crisis [5]. They collected 2381,297 relevant

Tweets in Turkish and English using twitterR package written in R programming language. Also, they developed a Turkish analysis lexicon. In order to visually summarize this text data, they used wordcloud package written in R programming language in which the words in the cloud were located in accordance with their frequency. The experimental results demonstrated that Turkish Tweets were mostly positive sentiments towards Syrians and refugees relative to neutral and negative sentiments. Pandey et al. proposed a novel metaheuristic method (CSK) based on K-means and cuckoo search to find the optimum cluster-heads from the sentimental contents of Twitter dataset [6]. The experimental results showed that the proposed method outperforms the existing methods. Existing studies in analysis of Tweets assume that all words of a tweet sentiment have polarity, so the word's sentiment polarity will be ignored. To solve this problem, Xiong et al. proposed a multi-level sentiment-enriched word embedding learning method that combined a parallel asymmetric neural network to model n-gram, word-level sentiment, and tweet-level sentiments in the learning process [7]. Experimental results demonstrated that the proposed method performed better than pioneer methods. Molinera et al. proposed a new way to determine how the debate is progressing; for instance, whether there was a consensus among the participants and which alternatives were preferred [8]. They used the sentiment analysis to measure the preference level of social media users with respect to a certain set of alternatives. The Tweets from Twitter have been used as the dataset, and group decision-making methods were applied in this paper. A deep learning based sentiment classifier has been proposed in [9]. This classifier works based on a word embedding model and a linear machine learning algorithm, which serves as a baseline classifier. Then, two ensemble techniques have been proposed to aggregate the baseline classifier with other surface classifiers. The surface classifiers are widely used classifiers in sentiment analysis. In addition, two models were proposed to combine both surface and deep features to merge information from several sources. A fuzzy logic approach for sentiment analysis has been proposed in [10]. The purpose of this paper was to build a model analyzing the content of a microblog like a

tweet, which analyses it to understand customer opinion. Then, this model can be used as a basis for a computer application. In the proposed method, the first step was to extract and separate the emojis, hashtags, and text from a tweet. The emojis can be easily classified by a lookup table, and the hashtags can be used to create a classifier. The second step is to parse the natural text of tweet lexically. In the last step, the combination of emoji classification, hashtag classification, and textual meaning will be imported to the fuzzy logic as input. The fuzzy logic will classify the tweet and store the result in its database. When the application finishes the analysis of a number of Tweets, the mean average of the Tweet can be calculated. The fuzzy logic module classifies each Tweet into these categories: strongly positive Tweets, positive Tweets, negative Tweets, and strongly negative Tweets. It assigns a number to each Tweet and stores the number in a database. Wang et al. proposed a model based on data mining and sentiment analysis to detect a person's depression [11]. Their method consisted of two steps. The first step was to propose a sentiment analysis method based on vocabulary and man-made rules to calculate the depression inclination of each microblog. In the second step, a depression detection model was developed according to the proposed method, and 10 features were derived from psychological research on depressed users.

IV. PROPOSED APPROACH

This section presents the proposed method that includes modules for reception, clearance, preprocessing, classification, and reprocessing of Tweets for sentiment analysis. The Tweets reception module first receives the big data related to the 2016 Olympic and Paralympic games. Then, Tweets Clearance module runs from unnecessary items and includes removing extra items like links, segmenting sentences into words, eliminating writing problems, and changing emoji to word on collected Tweets. After trimming the Tweets, the Tweets Processing Module runs from the entire Tweets to extract the Tweets of Iranian athletes in English and Persian. After extracting the Tweets related to the mentioned Iranian athletes in 2016 Olympic and Paralympic games, the Tweets Classification

Module tags and classifies the Tweets by determining polarity based on observation of keywords. Finally, the Tweets from the previous module are categorized by the Tweets Re-Processing Module to analyze sentiments in three intervals: one day before the competition, during competition, and one day after it, calculating the measure of precision and recall criteria. In Figure 1, workflow of the proposed method is visible, and further descriptions of each module are given below.

1. Tweets receiving module

A large amount of data can be accessed using the data derived from Tweets. The big data related to 2016 Olympic Games has been collected over a specific period. This collection has been completed in the beginning and end of Olympic Games from August 6, 2016 to August 21, 2016 and that of Paralympics from September 7, 2016 to September 17, 2016. Tweets are retrieved from twitter.com using Twitter Search API during 2016 Olympic Games. Twitter Search API plans for access to read and write Twitter data using 1-2% of a random sample from all Tweets. This collection is done using Web Scraper designed by Python Ruby. This web browser is used to collect and analyze English and Farsi Tweets in real time through a predefined list of hashtags (such as Table 1) by ignoring considerations. 1042795 Olympic-related Tweets and 1034815 Paralympic related ones have been collected. For each Tweet, the collected information includes the user id of the person who Tweets, tweet text, date, time, number of replies to the tweet, retweet times of the tweet, and the number of its likes. The Tweet type column has been added to be used in analysis. Tweets received through API Streaming are stored in HDFS format by selecting the saved default items to use information contained in Hadoop. HDFS is a distributed, scalable, and portable file system. Data are stored in data nodes and the information related to them is stored in the name node. This file system uses the Internet Protocol Setup layer to communicate, and the servers use a remote procedure call to talk to each other. HDFS file system is not limited to MapReduce tasks but can also be used for many other apps running and developing in Apache, including HBase database, Apache Mahout machine learning system, and Apache Hive data storage system.

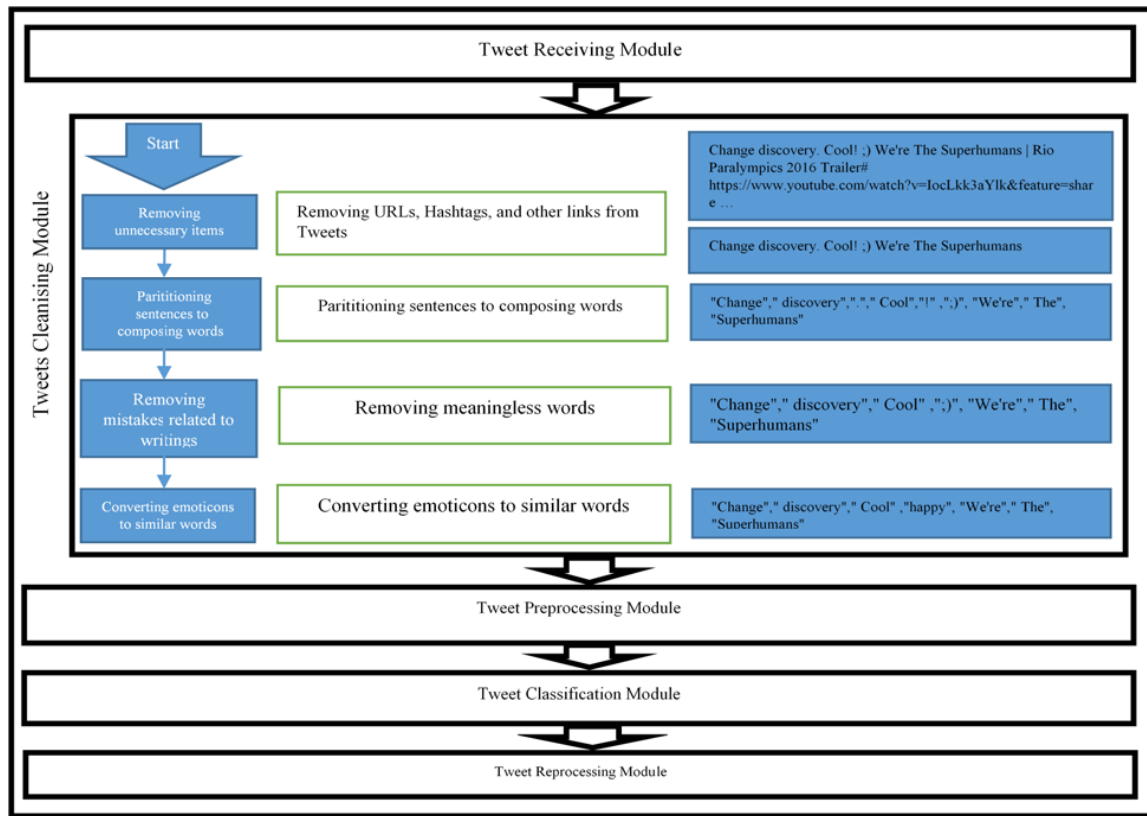


Fig.1. Workflow of proposed system

TABLE 1 sample hashtags

Farsi Hashtags Samples				
#محمود_گودرزی	#وزیر_ورزش	#ناداری	#المپیک	#ریو
#المپیک_ریو	#المپیک_2016	#ریو_2016	#احسان_حدادی	#کیمیا_علیزاده
#والیبال	#لیلا_رجبی	#المپیک_ریو_2016	#کیانوش_رستمی	#ورزش
#ورزش_ایران	#تکواندو	#کشتی_آزاد	#وزنه_برداری	#کشتی_فرنگی
#ورزش	#ورزش_ایران	#محمد_بنا	#کشتی_ایران	#بسکتبال
#بهداد_سلیمی	#بهمن_گلبار_نژاد	#سیامند_رحمان	#کامیل_قاسم	#کاروان_ورزش_ایران
#فدراسیون_جهانی_کشتی	#رضا_یزدانی	#پارالمپیک_ریو_2016	#کاروان_پارالمپیک_ایران	#تیراندازی
#پارالمپیک_ریو	#پارالمپیک	#والیبال_نشسته	#زهره_نعمنی	
English Hashtags Samples				
#behdad_salimi	#Rio_Olympics_2016 #Nadavary	#Kyanosh_Rostamy	#Leyla_Rajaby	#bahman_golbar_nejad
#Kymya_alyzadeh	#Sport	#2016Olympics	#Rio2016	#ehsan_Hadady
#siamandrahman	#Iran	#WorldRecord	#Paralympics	#Mohammad_Godarzy
#Vazyr_Varzesh	#Mojtaba_Abedy	#fencing	#Rio	#Rio_Olympics
#Olympics	#Volleyball	#Rio2016	#wrestling	#powerlifting

2. Tweets Cleansing Module

In this module, Tweets are cleaned of unnecessary items. As shown in Figure 1 (Tweets Cleansing Module section), the module has four steps to work on Tweets. The first step is to clear additional items to remove URL, hashtag, and other links from the text of Tweets. The second step is to parse the sentence into its constituent words, and the third step is to delete items related to writing to remove meaningless words such as full stop and comma. Finally, the fourth step is to convert the emoji to word in order to convert emojis to similar words like ";)" to "happy" or "joy" [11]. This step should be done on Tweets in both Farsi and English.

3. Tweets Preprocessing Module

After removing extra items from Tweets, Tweets Preprocessing Module is applied on Tweets in both English and Farsi from among all Tweets to extract Tweets of Iranian athletes ("Kianooshi", "Kimia Alizadeh", and "Siamand Rahman"). Using hashtags (#) including the names of these athletes in English and Farsi, the Tweets associated with these athletes during 2016 Olympic and Paralympic Games are separated. An example of the pseudo-code of this module is shown in Figure 2.

```
Select all field from tbl_tweet_total where text like '%athlete1_name%'
Select all field from tbl_tweet_total where text like '%athlete2_name%'
Select all field from tbl_tweet_total where text like '%athlete3_name%'
Select all field from tbl_tweet_total where text like '%نام ورزشکار1%'
Select all field from tbl_tweet_total where text like '%نام ورزشکار2%'
Select all field from tbl_tweet_total where text like '%نام ورزشکار3%'
Select all field from tbl_tweet_total where text like '#WorldRecord'
Select all field from tbl_tweet_total where text like 'نیزه یرداری#'
```

Fig. 2. Pseudo-code of Preprocessing Module

4. Tweets Classification Module

In this module, polarity determination and tagging are used to classify the Tweets that were obtained from the output of the previous model. Survey of Tweets is based on the words in the library. The methodology of this module is based on observing the keywords in Tweets. This classification is achieved according to a library that contains words (adjective or verb) expressing

feelings in Farsi (ترس, خشم, تعجب, غمگینی, شادی, خنثی, پیش بین)

and English (Fear, Anger, Surprise, Sadness, Joy, Neutral, Anticipation). The Tweets are surveyed, and the code for each Tweet in which the word associated with the mentioned sentiments is listed is recorded in database (for example, the words in "ترس" and "Fear" take the code 1), and

if the word associated with the mentioned sentiment group is not visible in it, the "خنثی"

or "Neutral" group code is recorded for it. This process is performed as follows. For example, the word groups "ترس" and "fear" are searched for

in the entire Tweets. Any tweet containing these words is updated in Tweet Type column with value 1. An example from the pseudo-code of this module is shown in Figure 3. Following the mentioned preparation steps, analysis can be done on the created data. The output of this module consists of Tweets categorized based on the sentiment signal contained in it. This module can be run with WordNet or SentiWordNet 3.0 tools, which are a library for processing text data. It should be noted that the entity detection system can only detect the listed entities by default; for this purpose, a manually developed dictionary is used to extract other entities.

```
UPDATE siamand_tweet SET tweet_type = '1' WHERE text like '%Fear%';
UPDATE siamand_tweet SET tweet_type = '2' WHERE text like '%Anger%';
UPDATE siamand_tweet SET tweet_type = '3' WHERE text like '%Surprise%';
UPDATE siamand_tweet SET tweet_type = '4' WHERE text like '%Sadness%';
UPDATE siamand_tweet SET tweet_type = '5' WHERE text like '%Joy%';
UPDATE siamand_tweet SET tweet_type = '6' WHERE text like '%Neutral%';
UPDATE siamand_tweet SET tweet_type = '7' WHERE text like '%Anticipation%';
UPDATE siamand_tweet SET tweet_type = '1' WHERE text like '%ترس%';
UPDATE siamand_tweet SET tweet_type = '2' WHERE text like '%خشم%';
UPDATE siamand_tweet SET tweet_type = '3' WHERE text like '%تعجب%';
UPDATE siamand_tweet SET tweet_type = '4' WHERE text like '%غمگینی%';
UPDATE siamand_tweet SET tweet_type = '5' WHERE text like '%شادی%';
UPDATE siamand_tweet SET tweet_type = '6' WHERE text like '%خنثی%';
UPDATE siamand_tweet SET tweet_type = '7' WHERE text like '%پیش بینی%';
```

Fig. 3. Pseudo-code of Classification Module

As shown in Table 2, the data.txt file has three parts: the first part is the name of entity (sentiment word), the second part is the type of entity (sentiment signal type), and the third part presents the entity code (sentiment signal code).

In this file, the attribute assignment process is such that whenever any of the entities listed in the above file exists in the input sentence, the type appearing against it is considered as the entity type. For instance, if the word "worry" is included in the entry, its type is considered as "Fear".

TABLE 2 KEY WORD SAMPLES

Sentiment word	Sentiment Signal Word	Sentiment signal Code
English Keyword		
worry	Fear	1
shit	Anger	2
goggle	Surprise	3
sadly	Sadness	4
cheer	Joy	5
hope	Anticipation	6
Persian Keyword		
مور مور	ترس	1
جهنم	خشم	2
حیرت آور	تعجب	3
رنج	غمگینی	4
ماشالا	شادی	5
ایمان داشتن	پیش بینی	6

5. Tweets Reprocessing Module for sentiment analysis

This module is implemented based on the output of the previous module, namely the classified Tweets. To analyze the information resulting from Tweets, the Tweets have to be reprocessed. The next processing criterion is the competition time of these samples. Tweets should be divided into three periods in order to recognize the sentiment impact of competition by Iranian athletes on Iranian fans. For example, the competition of "Siamand Rahman" was held on "16.9.2016" at "23:30" local time, and the Tweets related with this athlete have been surveyed in three periods as follows: the first period one day before competition, the second during the competition, and the third one day after it. The next survey criterion is the presence of a word associated with each group of sentiments based on calculation of precision, recall, and accuracy. For instance, if the Tweets of a majority of fans are in anticipation group before competition of Siamand Rahman, most Tweets are in joy group during the competition as well as after

it. Hence, the sentiment created among fans by this tournament can be distinguished based on the type of Tweet group. After reprocessing the Tweets and their separation based on the mentioned periods, the total number of Tweets in these periods is calculated. Then, after calculating the total Tweets number, their number in each sentiment group is calculated in smaller one-hour intervals. A sample of the pseudocode of this module can be seen in Figure 4.

```
Select all field from kianosh_tweet where date between '08/12/2016' and '08/14/2016'
Select all field from kimiya_tweet where date between '08/17/2016' and '08/19/2016'
Select all field from siamand_tweet where date between '10/14/2016' and '10/16/2016'
Select total COUNT (kianosh_tweet) from kianosh_tweet where date between '08/13/2016' and '08/13/2016' and time between '03:25'and'02:25';
Select total COUNT (kimiya_tweet) from kimiya_tweet where date between '08/17/2016' and '08/18/2016' and time between '17:25'and'16:25';
Select total COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/15/2016' and time between '00:00'and'23:00';
Select Fear COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '1' and time between '02:30'and'02:35';
Select Anger COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '2' and time between '02:30'and'02:35';
Select Surprise COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '3' and time between '02:30'and'02:35';
Select Sadness COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '4' and time between '02:30'and'02:35';
Select Joy COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '5' and time between '02:30'and'02:35';
Select Neutral COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '6' and time between '02:30'and'02:35';
Select Anticipation COUNT (siamand_tweet) from siamand_tweet where date between '10/14/2016' and '10/14/2016' and_type = '7' and time between '02:30'and'02:35';
```

Fig. 4. Pseudo-code of Reprocessing Module

For example, 1379 Tweets are related to "Siamand Rahman". After surveying these Tweets based on the three mentioned periods, there are 286 Tweets related to the period before "Siamand Rahman" competition. Within the first hour

before the competition of "Siamand Rahman", there are a total of 5 Tweets. From these Tweets, 1 Tweet is in sentiment group "Surprise" and 4 Tweets in sentiment group "Anticipation." Calculation of precision, recall, and accuracy (measurement error) of the proposed method is based on the corresponding formulas [13]. In the following, the comparison is performed on the basis of precision, recall, and accuracy criteria with available methods, as described in the next section.

V. EVALUATION OF RESULTS

This section describes the results and experiments conducted to categorize Twitter data and analyze sentiments. It has two main parts that are separately presented below. The first part explains the tested cases, and the results obtained by the proposed algorithm have been discussed in the previous chapter. The second part of the proposed methodology for data classification and sentiment analysis is discussed using the second method. Sections 3 and 4 will compare the results obtained in the proposed method with other methods.

In general, the clustering process was as follows. First, Twitter data were extracted using a crawler. Then, the sentimental entities in each of the extracted Tweets (if present) were tagged. Finally, these data were classified into different clusters using big data processing tools such as Mahout or Hive. The purpose of this process is to collect similar data in the context of Twitter. For instance, decision-makers in enterprises can produce or advertise their products based on the data in these categories.

1. Configuration environment

Apache Hadoop, Apache Mahout, Wordnet, SentiWordnet 3.0, Twitter 4j, and Eclipse Mars 1 configuration environment were used in this study. The implementation languages were Python Ruby, Java JDK, and Hive. In this implementation, the Twitter Knowledge Base was used; 1042795 Olympic-related Tweets and 1034815 Paralympics-related ones were collected over a specific period of time. There is also a

high capacity to process a higher amount of data due to the implementation on Hadoop and Hive frameworks. The assessment was conducted in a Hadoop cluster with two nodes. Ubuntu 17.04 Linux operating system and 2.7.2 version of Hadoop were used. We used the Apache Hive tool for analysis process in Hadoop cluster. To compare the response time of queries in this tool relative to peer tools (Pig, SQL), the sample queries mentioned in Table 3 were linked to the three end modules of Figure 1 on the tables associated with "Kianoosh Rostami", "Kimia Alizadeh", and "Siamand Rahman". Based on mathematical functions, uniform distribution has been used to run 1000 random queries on the mentioned tables. Finally, the response time is calculated based on Poisson distribution function and plotted by exponential function. Based on the results related to "Siamand Rahman", the average response time for HIVE tool in Tweets preprocessing module, Tweets Classification Module, and Tweets Reprocessing Module for sentiments analysis is 155 seconds, 215 seconds, and 666 seconds, respectively. The results are shown in Figure 5. According to the findings, the response time of HIVE tool is shorter than the mentioned cases.

TABLE 3. QUERY SAMPLES

Select query 1	SELECT * FROM SIAMAND_RAHHMAN_TWEETS T WHERE T.tweet like '%ocheer%';
Select query 2	SELECT T.date,T.time FROM SIAMAND_RAHHMAN_TWEETS T WHERE T.date > '25/06/1395';
Aggregation query	SELECT T.id,COUNT(T.Tweet_type) FROM SIAMAND_RAHHMAN_TWEETS T GROUP BY T.id ;
Join query	SELECT INTO TEMP TWEETS_COUNT, COUNT(T.Tweet_type) as S_tweet_type, COUNT(K.Tweet_type) as K_tweet_type FROM SIAMAND_RAHHMAN_TWEETS AS T,KIANOSH_ROSTAMI_TWEETS AS K WHERE T.Tweet_type=K.Tweet_type AND T.date BETWEEN '20/06/1395' AND '26/06/1395' GROUP BY S_tweet_type.TWEETS_COUNT;

2. Results obtained from classification of Twitter data and sentiment analysis based on the proposed method

"Siamand Rahman" has participated in weightlifting at Rio Paralympic caravan. The total number of Tweets collected for Rio Paralympics from Tweets Reception Module is 1034815. The Tweets Clearance Module runs on this number of

Tweets. Subsequently, Tweets Preprocessing Module separates the Tweets related to samples based on hashtags such as "#siamandrahman #" and "سیامند رحمان #". From a total of 1034815

Paralympic Tweets, 1379 Tweets are related to "Siamand Rahman", including 1156 Tweets in Farsi and 223 Tweets in English. The next step is to survey each Tweet using Tweets classification module based on sentiment-related word and Tweet tagging. This classification is based on a library that includes words (adjective or verb) that express feelings in Farsi (ترس، خشم، تعجب، غمگینی، شادی، خنثی، پیش بین) and

English (Fear, Anger, Surprise, Sadness, Joy, Neutral, Anticipation). The Tweets are surveyed, and the code for any tweet containing the word associated with the listed sentiment group is recorded in the database (for example, the words

in "ترس" and "Fear" group are assigned code 1), and if the word associated with sentiment group is not seen in it, the "خنثی" or "Neutral" group code is recorded for it. This process is performed in such a way that the word group "ترس" and "Fear" (for example) is searched for in the entire Tweets from Tweets Preprocessing Module. Any Tweet containing these words is updated in Tweet Type column with the value of 1. Afterward, all Tweets related to "Siamand Rahman" competition are counted according to the number of sentiment signals. The output of this Tweets Classification Module is divided into seven sentiment signal groups: group 1 (ترس | Fear), group 2 (خشم |

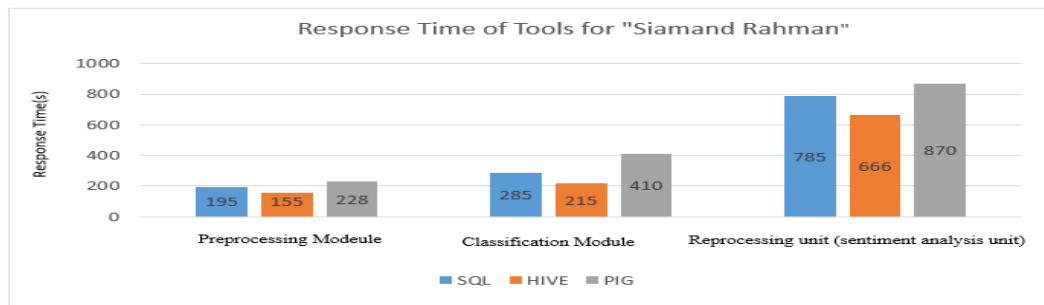


Fig. 5. Response Time of Different Tools for "Siamand Rahman"

TABLE 4 Tweet analysis based on sentiment signals before competition of "Siamand Rahman"

Time	Tweets	Fear	Anger	Surprise	Sadness	Joy	Neutral	Anticipation	Precision	Recall
00:00	5	0	0	1	0	0	0	4	80/00	1/40
01:00	11	0	0	1	0	2	0	8	72/73	2/80
02:00	6	0	0	1	0	5	0	0	0/00	0/00
03:00	13	3	0	1	2	0	1	6	46/15	2/10
04:00	8	0	0	0	0	3	0	5	62/50	1/75
05:00	15	3	0	0	0	3	0	9	60/00	3/15
06:00	10	0	0	0	0	0	0	10	100/00	3/50
07:00	13	0	0	2	0	3	0	8	61/54	2/80
08:00	16	3	0	1	0	3	0	9	56/25	3/15
09:00	6	0	0	2	2	0	0	2	33/33	0/70
10:00	1	1	0	0	0	0	0	0	0/00	0/00
11:00	12	2	2	2	2	2	2	0	0/00	0/00
12:00	9	0	0	1	0	0	0	9	100/00	3/15
13:00	6	0	0	0	0	2	0	4	66/67	1/40
14:00	16	2	0	3	3	0	0	8	50/00	2/80
15:00	11	2	0	0	0	0	0	9	81/82	3/15
16:00	9	0	0	0	0	0	0	8	88/89	2/80
17:00	8	0	0	0	0	0	0	7	87/50	2/45
18:00	7	2	0	0	0	0	0	5	71/43	1/75
19:00	8	3	0	0	0	0	0	4	50/00	1/40
20:00	17	0	0	0	0	0	0	16	94/12	5/59
21:00	18	0	0	1	0	0	0	17	94/44	5/94
22:00	20	1	0	0	0	0	0	19	95/00	6/64
23:00	41	0	0	2	0	2	2	35	85/37	12/24

Anger), group 3 (تعجب | surprise), group 4 (غمگینی | sadness), group 5 (شادی | Joy), group 6 (خنثی | Neutral), group 7 (پیش‌بینی | Anticipation).

After surveying these Tweets based on Tweets classification module, the Tweets Re-Processing Module runs for analysis of sentiments within the three mentioned periods.

286 Tweets are related to the period before "Siamand Rahman" competition, 702 Tweets to the time "Siamand Rahman" competes, and 391 Tweets to the period after his competition. The results of this count, calculation of precision and recall criteria are shown in Table 4 (Tweet analysis based on sentiment signal before the competition of "Siamand Rahman").

TABLE 5
MEASUREMENT ERROR FOR ANTICIPATION SIGNAL BEFORE COMPETITION OF "SIAMAND RAHMAN"

Time	Tweets	Anticipation	Measurement error
00:00	5	4	20.00
01:00	11	8	27.27
02:00	6	0	100.00
03:00	13	6	53.85
04:00	8	5	37.50
05:00	15	9	40.00
06:00	10	10	0.00
07:00	13	8	38.46
08:00	16	9	43.75
09:00	6	2	66.67
10:00	1	0	100.00
11:00	12	0	100.00
12:00	9	9	0.00
13:00	6	4	33.33
14:00	16	8	50.00
15:00	11	9	18.18
16:00	9	8	11.11
17:00	8	7	12.50
18:00	7	5	28.57
19:00	8	4	50.00
20:00	17	16	5.88
21:00	18	17	5.56
22:00	20	19	5.00
23:00	41	35	14.63

Fig. 6. Measurement error before "Siamand Rahman"

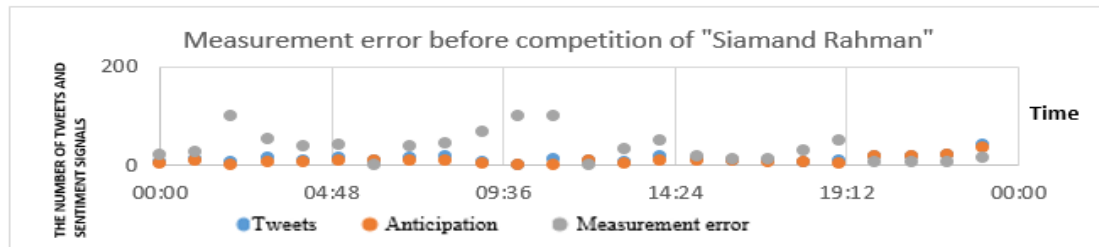


TABLE 6
Tweet analysis based on sentiment signal during competition of "Siamand Rahman"

Time	Tweets	Fear	Anger	Surprise	Sadness	Joy	Neutral	Anticipation	Precision	Recall
00:00	5	0	0	1	0	0	0	4	80/00	1/40
01:00	11	0	0	1	0	2	0	8	72/73	2/80
02:00	6	0	0	1	0	5	0	0	0/00	0/00
03:00	13	3	0	1	2	0	1	6	46/15	2/10
04:00	8	0	0	0	0	3	0	5	62/50	1/75
05:00	15	3	0	0	0	3	0	9	60/00	3/15
06:00	10	0	0	0	0	0	0	10	100/00	3/50
07:00	13	0	0	2	0	3	0	8	61/54	2/80
08:00	16	3	0	1	0	3	0	9	56/25	3/15
09:00	6	0	0	2	2	0	0	2	33/33	0/70
10:00	1	1	0	0	0	0	0	0	0/00	0/00
11:00	12	2	2	2	2	2	2	0	0/00	0/00
12:00	9	0	0	1	0	0	0	9	100/00	3/15
13:00	6	0	0	0	0	2	0	4	66/67	1/40
14:00	16	2	0	3	3	0	0	8	50/00	2/80
15:00	11	2	0	0	0	0	0	9	81/82	3/15
16:00	9	0	0	0	0	0	0	8	88/89	2/80
17:00	8	0	0	0	0	0	0	7	87/50	2/45
18:00	7	2	0	0	0	0	0	5	71/43	1/75
19:00	8	3	0	0	0	0	0	4	50/00	1/40
20:00	17	0	0	0	0	0	0	16	94/12	5/59
21:00	18	0	0	1	0	0	0	17	94/44	5/94
22:00	20	1	0	0	0	0	0	19	95/00	6/64
23:00	41	0	0	2	0	2	2	35	85/37	12/24

Fig. 7. Tweet analysis based on sentiment signals before competition of "Siamand Rahman"

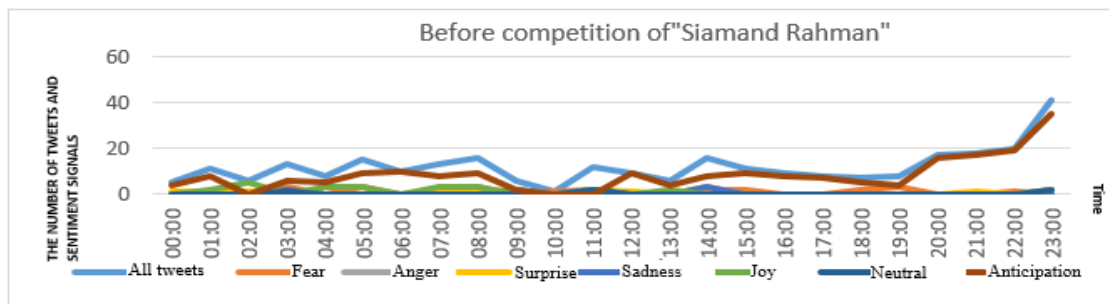


TABLE 7
Tweet analysis based on sentiment signal after competition of "Siamand Rahman"

Time	Tweets	Fear	Anger	Surprise	Sadness	Joy	Neutral	Anticipation	Precision	Recall
02:00	15	0	0	1	0	14	0	0	93/33	3/58
03:00	13	3	0	1	2	6	1	0	46/15	1/53
04:00	18	0	0	0	0	15	0	0	83/33	3/84
05:00	15	3	0	0	0	12	0	0	80/00	3/07
06:00	10	0	0	0	0	9	0	1	90/00	2/30
07:00	13	0	0	2	0	11	0	0	84/62	2/81
08:00	45	3	0	1	0	36	0	5	80/00	9/21
09:00	43	0	0	0	0	40	0	0	93/02	10/23
10:00	38	1	0	0	0	37	0	0	97/37	9/46
11:00	29	0	2	2	0	24	0	0	82/76	6/14
12:00	27	0	0	1	0	26	0	0	96/30	6/65
13:00	16	0	0	5	0	11	0	0	68/75	2/81
14:00	16	2	0	3	3	8	0	0	50/00	2/05
15:00	11	2	0	0	0	9	0	0	81/82	2/30
16:00	9	0	0	2	0	7	0	0	77/78	1/79
17:00	8	0	0	0	0	6	0	0	75/00	1/53
18:00	7	2	0	0	0	5	0	0	71/43	1/28
19:00	8	3	0	0	0	5	0	0	62/50	1/28
20:00	14	0	0	0	0	11	0	3	78/57	2/81
21:00	12	0	0	1	0	11	0	0	91/67	2/81
22:00	13	1	0	0	0	11	0	1	84/62	2/81
23:00	11	0	0	2	0	6	2	0	54/55	1/53

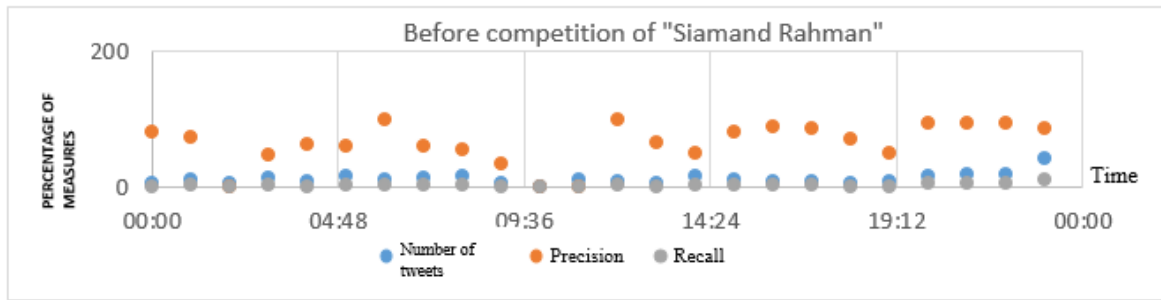


Fig. 8. Percentage of precision and recall based on sentiment signals before competition of “Siamand Rahman”

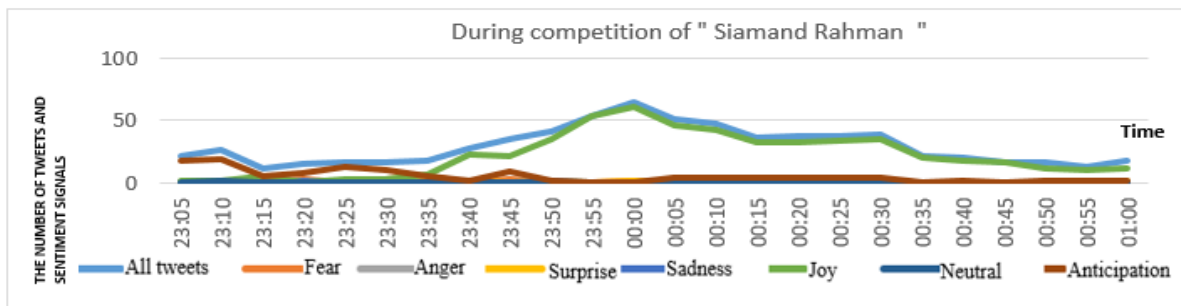


Fig. 9. Tweet analysis based on sentiment signals during competition of “Siamand Rahman”

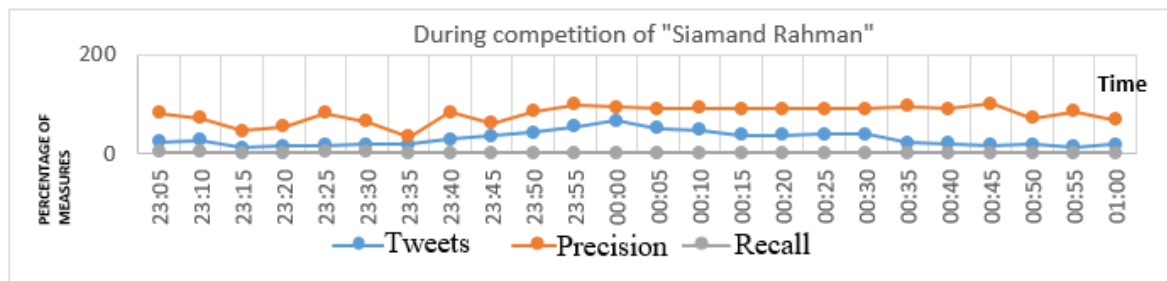


Fig. 10. Percentage of precision and recall based on sentiment analysis during competition of “Siamand Rahman”

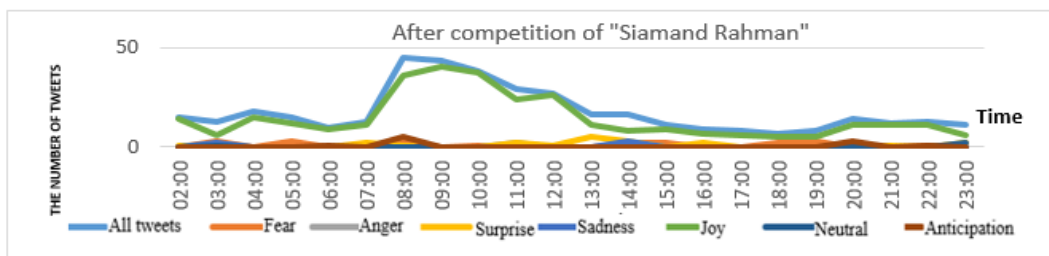


Fig. 11. Tweet analysis based on sentiment signals after competition of “Siamand Rahman”

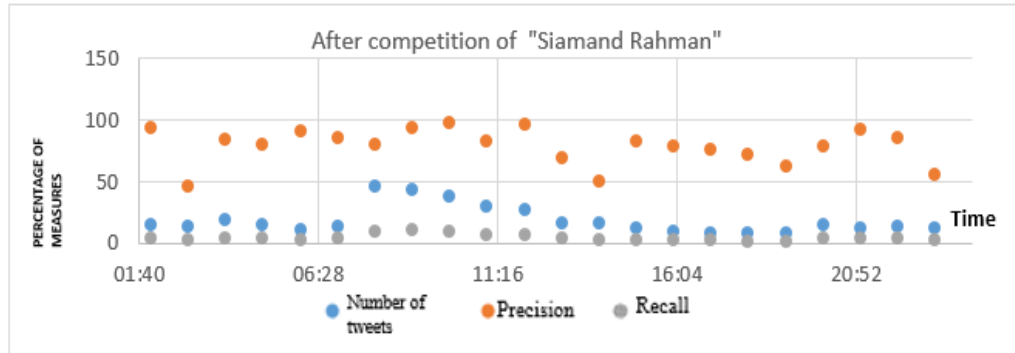


Fig. 12. Percentage of precision and recall based on sentiment analysis after competition of “Siamand Rahman”

According to the results presented in Table 4 and the majority of sentiment signals in relevant interval (before athlete's competition), the type of sentiment signal could be predicted in the next period (during athlete's competition). Table 5 results from Table 4, and the measurement error criterion for sentiment signal is "anticipated". There are 5 Tweets associated with this athlete; for example, at 12 AM, one day before "Siamand Rahman" competition. From these 5 Tweets, there are 4 Tweets with 20% measurement error criterion and a sentiment signal "anticipating" athlete's success in the competition. According to 24 records in Tables 4 and 5 with an average measurement error criterion of 35.93%, it can be predicted that if this athlete succeeds in the next interval, most sentiment signals of athlete's fans are written in "Joy" group. The diagram of calculation error measurement for "Anticipation" sentiment signal in the interval before competition of "Siamand Rahman" can be seen in Figure 6. According to the prediction made according to Table 5 with an average measurement error of 35.93, which anticipated that in case this athlete succeeds in the next interval, the majority of sentiment signals of athlete's fans are placed in the "Joy" group, Table 6 (Tweet analysis based on sentiment signal during competition of "Siamand Rahman's") and Table 7 (Tweet analysis based on sentiment signal after competition of "Siamand Rahman's") indicate that Tweets are placed in sentiment signal group of "Joy" with average precision of 78.99 and 74.94, respectively. Based on the mentioned tables, Figure 7 (Tweets analysis based on sentiment signal before the competition of "Siamand Rahman"), Figure 8 (percentage of

precision and recall criterion related to Tweets number based on sentiment signal before the competition by "Siamand Rahman"), Figure 8 (survey of Tweets based on sentiment signal during "Siamand Rahman's" competition), Figure 9 (percentage of precision and recall criteria from sentiment signal during "Siamand Rahman's" competition), Figure 10 (survey of Tweets according to sentiment signal after the competition by "Siamand Rahman"), and Figure 11 (percentage of precision and recall criterion from sentiment signal after "Siamand Rahman's" competition) are obtained.

3. Classification of Twitter data and sentiment analysis based on Pak et al. approach

Pak et al. method is based on observing the keywords for sentiment analysis [14] according to extracted Tweets are only available in English. This approach does not use sentiment analysis tools (like Hive) in Hadoop environment for big data. In this method, polarity determination and classification are based on three sentiment signals: "positive", "negative", and "neutral". The presence of n-gram in this approach is used as a binary feature, while for general information retrieval purposes, the keyword occurrence frequency is a more appropriate attribute because overall sentiments are not necessarily reflected through repeated use of keywords. The results of bigrams and trigrams have had a better performance for polarity classification. In this study, it has been attempted to provide an optimal setting for Twitter microblogging data. On the one hand, high-level n-grams, including trigrams, should have a better ability to record the patterns of

sentiment signals.

On the other hand, unigrams should provide good coverage of data. N-grams are created by a set of n-grams from consecutive negative words (such as the word "no" and "negative" letters) connected to a preceding or following word.

This approach improves the classification precision relative to the one in which negation plays a role in the expression of opinion and sentiment.

In this way, a sentimental classification is created using Newbies polynomial, and classification is also based on SVM and CRF for comparison. However, the Newbies classification has achieved the best results. Finally, the probability of each sentiment is calculated. Sentiment analysis results on 2016 Olympic and Paralympic Tweets using this method are presented for ease of comparison in the intervals of the proposed approach.

TABLE 8
Average precision and recall in three time periods

Time Period	Average precision by proposed method	Average precision by Pak method	Average recall by proposed method	Average recall by Pak method
Before competition "Siamand Rahman"	64.07	47.35	97.06	95.83
During competition "Siamand Rahman"	78.99	45.81	99.25	95.83
After competition "Siamand Rahman"	74.94	45	96.44	95.65

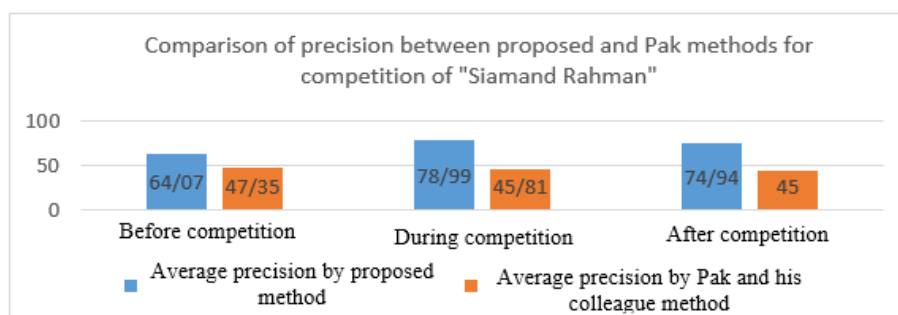


Fig. 13. Comparison of precision between proposed and Pak methods for competition of “Siamand Rahman”

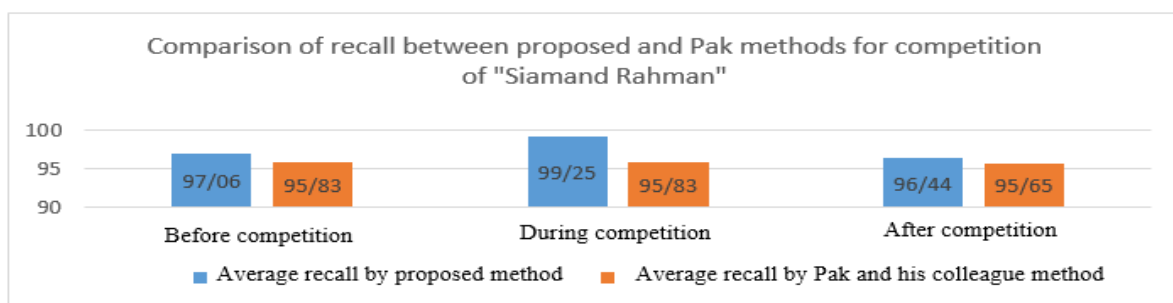


Fig. 14. Comparison of recall between proposed and Pak methods for competition of “Siamand Rahman”

4. Comparison of the proposed method results with Pak et al. method

Based on Figure 7 and 8, in the interval before "Siamand Rahman's" competition, the majority of fans' Tweets are in sentiment signal "anticipation" with a precision criterion of "64.07" and a recall criterion of "2.94" due to anticipation of the athlete's success in the competition ahead. Obviously, in a number of Tweets, cases of sentiment signal groups "Joy" and "Sadness" are observed but cannot be raised. Based on Figures 9 and 10, during "Siamand Rahman's" competition, most Tweets are in sentiment signal group "Anticipation" and more "Joy" with a precision criterion of "78.99" and recall criterion of "0.75" because of the finalization of "Siamand Rahman's" success in winning the gold medal in this tournament. Based on Figures 11 and 12, in the period after "Siamand Rahman's" competition, the majority of Tweets are in sentiment signal "Joy" with precision criterion of "74.94" and recall criterion of "3.56", a trend caused by success of "Siamand Rahman" in winning the gold medal of tournament. Based on the above-mentioned Figures, the initial sentiment signal has been based on "anticipation" and the feelings of the athlete's fans and sports caravan of Iran have been changed to "Joy" sentiment signal after the completion of competitions by this athlete. The above statements along with the results of Pak et al. method for "Siamand Rahman" at similar intervals can be seen in Table 8. Figures 13 and 14 compare the precision and recall criteria of the two methods in competition of "Siamand Rahman", respectively. Based on these two Figures and the inverse relationship between precision and recall criteria, improvement of results is visible according to the proposed method. The average precision criterion based on proposed method within the three periods of "Siamand Rahman's" competition is equal to "72.67", which has been improved by 26.62% compared to the average precision criterion of Pak et al., which is equal to 46.05.

The average recall criterion according to the proposed method in the three calculated periods of "Siamand Rahman's" competition is equal to 2.42, which has been improved by 1.81% in comparison to the average recall criterion of Pak et al. method, which is "4.23".

5. Classification of Twitter data according to dataset of Yang et al. and sentiment analysis based on the proposed method

Yang et al. [12] received the Tweets related to US soccer team fans during the five FIFA 2014 soccer games (three matches between US national team and another team and two matches among other teams) using the Twitter search API. Analysis of emotions was used to examine the emotional responses of US fans to Twitter, especially emotional changes after the goal (US national team to opponent or opponent to the US national team). In the matches in which the US team was engaged, there were more fear and anger in negative sentiments signal, and in general, these feelings diminished when the US team scored goals to opponent's team. Anticipation and joy were also generally matched to match scores and conditions throughout the tournament.

In addition, US national team Tweets in matches between other teams showed more joy and anticipation than negative feelings (such as anger and fear), and the patterns were not clear in response to goals or loss of goals. This technique showed that sports fans use Twitter for emotional purposes and a big data approach to analyze the sentiments of sports fans indicated that the results had good condition in terms of the anticipated results. In total, in this method, there were 1007, 1295, and 2135 Tweets for three US team matches (from a total of 26,881, 26,014, and 49,576 Tweets) as well as 461 and 468 Tweets for the two games of France-Nigeria and Brazil-Colombia (out of a total of 21, 901 and 25,494 Tweets). The total number of Tweets listed in brackets includes Tweets without location information and Tweets with type seal from other countries. In the remainder of this section, the results are presented on the data set used in Yang et al. as well as the analysis of these Tweets using the proposed method. Finally, the comparison of the results of this analysis with the proposed method is presented. For ease of comparison, only Tweets for three US games are used in this section.

6. Comparison of results obtained in the proposed method based on data set of Yang et al.

In previous section, the sentiment survey and analysis results on Yang et al. data in three matches of the US national team have been

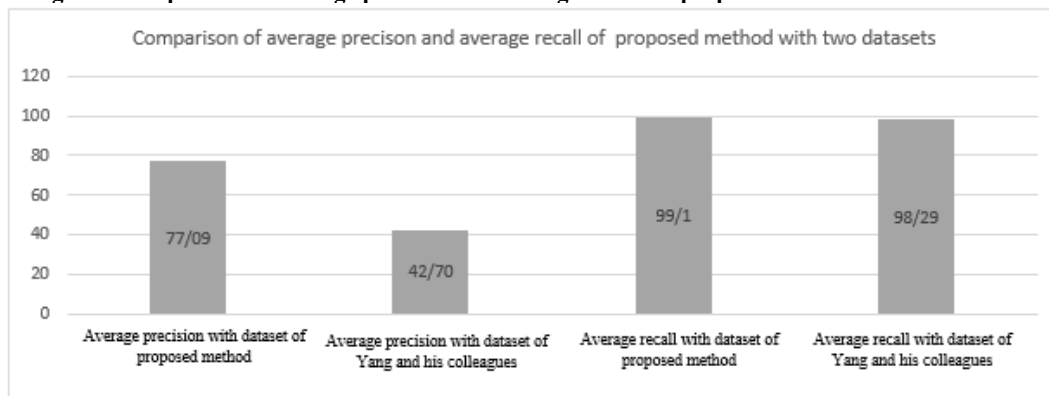
presented. Among the three mentioned periods, only the match time has been reviewed because of the appropriate number of Tweets. According to Figures 13, 14 and 10, during the completion of “Siamand Rahman” competition, the majority of

Tweets are in sentiment signal group “anticipation” and more “Joy” with precision criterion of “78.99” and recall criterion of “0.75”, a trend caused by conclusion of “Siamand Rahman’s” success to win gold medal in this competition. According

TABLE 9
Comparison of Average precision and recall with two datasets

Time period	During competition of "Siamand Rahman"	During competition of "Kianoosh Rostami"	During competition of "Kimia Alizade"	During competition US against Portugal	During competition US against Germany	During competition US against Belgium
Average precision with the dataset of proposed method	78.99	71.41	80.87
Average precision with the dataset of Yang method	46.76	40.52	40.43
Average recall with the dataset of proposed method	99.25	98.88	99.17
Average recall with the dataset of Yang method	98.19	98.34	98.35

Fig. 15. Comparison of average precision and average recall of proposed method with two datasets



to the mentioned Figures, the initial sentiment signal is "Anticipation" and after completion of the competitions, the sentiments of athlete's fans as well as Iranian sports caravan have changed to "Joy". The above statements and other results related to competition times of "Kianoosh Rostami" and "Kimia Alizadeh" can be seen in Table 9.

During the "US vs. Portugal" match, the Tweets of a majority of fans are in sentiment signal "Anticipation" with precision and recall criteria of "46.76" and "1.81", respectively because of predicting the success of this team in the next match. However, the sentiment signals "joy" and "anger" are observed in some Tweets but cannot be validated. During the match of "US vs. Germany", most Tweets are in sentiment signal "Anticipation" with precision and recall criteria of "40.42" and "1.66", respectively. During the "US-Belgium" match, the majority of Tweets were in sentiment signal group "Anticipation" with precision and recall criteria of "45.93" and "1.65", respectively. On the basis of the above Figures, the initial sentiment signal is "Anticipation" and after completion of these competitions, the sentiments of athlete's fans as well as US sports caravan have changed to sentiment signal of "Joy". The above-mentioned results can be seen in Table 9. In Figure 15, the precision and recall criteria have been compared between the proposed method with the dataset of Yang et al. Based on this Figure and the inverse relationship between precision and recall criteria, the improvement of results according to the proposed method is visible.

The average precision criterion based on the proposed method and its data set in the calculated competition period is equal to "77.09", which shows "34.39%" improvement relative to precision criterion of Yang et al. method that was equal to "42.70".

The average recall criterion based on the proposed method and its data set in the calculated competition period is equal to "0.9", which shows "0.81%" improvement relative to

precision criterion of Yang et al. method that was equal to "1.71".

VI. CONCLUSION

This paper deals with the sentiments of Iranian sports fans in 2016 Rio Olympic and Paralympic based on "Natural Test" and a big data approach for analyzing real-time sentiments on Twitter of Iranian sports fans. Our study showed that the use of big data analysis of sentiments in Tweets is in line with expectations.

Big data analysis allows for analysis of a large number of Tweets and examination of natural sentimental reactions of real-life sports fans. However, our analysis of sentiments has its limitations. First, we analyzed sentiments by exploring each word. In this study, we did not consider the meaning of the whole Tweets. Second, we examined Tweets as a way of exposing sentiments but many sports fans may not trust the media. Third, we do not know the percentage of fans of athletes engaged in a competition. Big data analysis may only show some sports fans. Fourth, commenting on sentiment signals in Tweets of Twitter users is a difficult task. In general, numerical indicators based on the number of words with sentiment signals are comparable with sentiment signals and whole Tweets for a competition. When there are fewer Tweets, the conclusion will be difficult.

Future works can be conducted in the following ways:

- Sentiment analysis on data from LinkedIn Job Opportunities site.
- Big data analysis to forecast the future of the market, customer orientation, and general tastes of customers.
- Analysis of sentiments in stores as well as financial, communications, marketing, and other companies to help discover the relationship between internal factors (including price, product placement, and employee skills) with external ones (including economic status, market competition, and customer geographical location).
- Discovering the pattern of sentiment expressed in various areas of media sources and among different social strata.

REFERENCES

1. López, V., Del Río, S., Benítez, J.M. and Herrera, F., 2015. Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data. *Fuzzy Sets and Systems*, 258, pp.5-38.
2. Chen, C.P. and Zhang, C.Y., 2014. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, pp.314-347.
3. Pang, B. and Lee, L., 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1-2), pp.1-135.
4. Birjali, M., Beni-Hssane, A. and Erritali, M., 2017. Machine learning and semantic sentiment analysis based algorithms for suicide sentiment prediction in social networks. *Procedia Computer Science*, 113, pp.65-72.
5. Öztürk, N. and Ayvaz, S., 2018. Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35(1), pp.136-147.
6. Pandey, A.C., Rajpoot, D.S. and Saraswat, M., 2017. Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, 53(4), pp.764-779.
7. Xiong, S., Lv, H., Zhao, W. and Ji, D., 2018. Towards Twitter sentiment classification by multi-level sentiment-enriched word embeddings. *Neurocomputing*, 275, pp.2459-2466.
8. Morente-Molinera, J.A., Kou, G., Peng, Y., Torres-Albero, C. and Herrera-Viedma, E., 2018. Analysing discussions in social networks using group decision making methods and sentiment analysis. *Information Sciences*, 447, pp.157-168.
9. Araque, O., Corcuera-Platas, I., Sanchez-Rada, J.F. and Iglesias, C.A., 2017. Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications*, 77, pp.236-246.
10. Howells, K. and Ertugan, A., 2017. Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia computer science*, 120, pp.664-670.
11. Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L. and Bao, Z., 2013, April. A depression detection model based on sentiment analysis in micro-blog social network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 201-213). Springer, Berlin, Heidelberg.
12. Yu, Y. and Wang, X., 2015. World Cup 2014 in the Twitter World: A big data analysis of sentiments in US sports fans' tweets. *Computers in Human Behavior*, 48, pp.392-400.
13. https://link.springer.com/referenceworkentry/10.1007%2F978-0-387-30164-8_652, Last accessed on April 2019.
14. Pak, A. and Paroubek, P., 2010, May. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, Vol. 10, No. 2010, pp. 1320-1326.