

A real-time big data sentiment analysis for iraqi tweets using spark streaming

Nashwan Dheyaa Zaki¹, Nada Yousif Hashim², Yasmin Makki Mohialden³,
Mostafa Abdulghafoor Mohammed⁴, Tole Sutikno⁵, Ahmed Hussein Ali⁶

¹University of Information Technology and communications, College of Engineering, Iraq

²Computer Science Dep, College of Education, Al-Iraqia University Baghdad, Iraq

³Department of Computer Science, College of Science, Mustansiriyah University, Iraq

⁴Faculty of Automatic Control and Computers, University Polytechnic of Bucharest, Romania

⁵Department of Electrical Engineering, Universitas Ahmad Dahlan, Indonesia

⁶Embedded System and Power Electronics Research Group, Yogyakarta, Indonesia

⁶Al-Salam University College Computer Science Department Baghdad, Iraq

Article Info

Article history:

Received Nov 12, 2019

Revised Feb 6, 2020

Accepted Mar 15, 2020

Keywords:

Big data

Online processing

Sentiment analysis

Spark streaming

Twitter platform

ABSTRACT

The scale of data streaming in social networks, such as Twitter, is increasing exponentially. Twitter is one of the most important and suitable big data sources for machine learning research in terms of analysis, prediction, extract knowledge, and opinions. People use Twitter platform daily to express their opinion which is a fundamental fact that influence their behaviors. In recent years, the flow of Iraqi dialect has been increased, especially on the Twitter platform. Sentiment analysis for different dialects and opinion mining has become a hot topic in data science researches. In this paper, we will attempt to develop a real-time analytic model for sentiment analysis and opinion mining to Iraqi tweets using spark streaming, also create a dataset for researcher in this field. The Twitter handle Bassam AlRawi is the case study here. The new method is more suitable in the current day machine learning applications and fast online prediction.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Mostafa Abdulghafoor Mohammed,
Faculty of Automatic Control and Computers,
University Polytechnic of Bucharest, Romania.
Email: alqaisy86@gmail.com

1. INTRODUCTION

The much attention has been given to real-time data processing and big data analytics in the recent era [1, 2]. The increasing volume of information generated daily has made it necessary that organizations must devise ways of handling this information since the existing techniques are not efficient in handling such volume of data created at such a high rate [3-6]. The concept of big data does not only relate to the data volume; it also encompassed data velocity and variety[7]. Data can be either structured or unstructured; it can also be in the forms of a dumped file or in real-time streaming format with high velocity [8-10]. The recent advancements in social network websites have made sentiment analysis and opinion mining a trending research area [11-13]. Twitter is one of the social network applications which is used daily by millions of people to express their opinion. Iraqi dialect stream data has witnessed a tremendous level of growth [14-17] and social media applications remain the commonest source of this volume of stream data. Iraqi users express their opinion and attitude towards issues using different applications [18, 19]. The current opinion of Arabic users can be understood through real-time sentiment analysis of social media streams. Sentiment analysis

(SA) refers to the identification of the hidden subjectivity in a text and highlighting the human emotions. During SA, the opinions' contextual polarity (such as positive, neutral, or negative) is portrayed [14, 20-22]. The current researchers on Iraqi sentiment depend on supervised or lexicon-based methods; however, until now, real-time SA has not been embarked by the existing Iraqi studies.

There are some key challenges before developing any online analytics framework; first is developing reliable and efficient frameworks to distributed data without losing accuracy [23, 24]. Example of this framework is Apache Storm and Apache Kafka [25-27]. The problem of streaming data is that it contains high-velocity information in continues form. Therefore, the process of text analysis in current machine tools will be a bottleneck. All the previous work on sentiment analysis are limited to batch analysis of data. Hence, this study used Spark and decision tree (DT) to propose and develop SA solution for Iraqi dialect. In this study, we want to design an online analytics framework that ingests data from Twitter API with fast processing and ready for predictions at any time [28-30]. The above system is missing in the current Arabic research of SA. This study aims to develop a real-time analytic model for sentiment analysis and opinion mining of Iraqi tweets using spark streaming (Bassam Al-Rawi is the case study). This organization of this paper is as follows: a general description of Spark Streaming was presented in section 2 while section 3 presented the concept of resilient distributed dataset (RDD). Section 4 described sentiment analysis (text analytics). Section 5 explained the proposed method in this study while the experiment and results were detailed and discussed in section 6. The last section (section 7) presented the conclusion drawn from the study.

2. SPARK STREAMING

Apache Spark is an open source framework which consists of an engine for programs distribution across machine clusters and a sophisticated model for writing programs [31-33] ment, it has contributed to the Apache Software Foundation, making it possible for a data scientist to access distributed programming. The initialization of the Spark engine is shown in Figure 1.

```

holden@hosp2: ~/Downloads/spark-1.1.0-bin-hadoop1.2 /bin/pyspark
Python 2.7.6 (default, Mar 22 2014, 22:59:56)
[GCC 4.8.2] on linux2
type 'help' for 'copyright', 'credits' or 'license' for more information.
Spark assembly has been built with Hive, including Dataclass jars on classpath
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
14/11/19 14:33:49 WARN Utils: Your hostname, hosp2 resolves to a loopback address: 127.0.1.1; using 172.17.42.1 instead (on interface docker0)
14/11/19 14:33:49 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
14/11/19 14:33:49 INFO SecurityManager: Changing view acls to: holden.
14/11/19 14:33:49 INFO SecurityManager: Changing modify acls to: holden.
14/11/19 14:33:49 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(holden, )
c users with modify permissions: Set(holden, )
14/11/19 14:33:49 INFO Slf4jLogger: Slf4jLogger started
14/11/19 14:33:49 INFO Remoting: Starting remoting
14/11/19 14:33:49 INFO Remoting: Remoting started; listening on addresses : [akka.tcp://sparkDriver@172.17.42.1:35821]
14/11/19 14:33:49 INFO Utils: Successfully started service 'sparkDriver' on port 35821.
14/11/19 14:33:49 INFO SparkEnv: Registering MapOutputTracker
14/11/19 14:33:49 INFO DiskBlockManager: Created local directory at /tmp/spark-local-20141119143349-5776
14/11/19 14:33:49 INFO Utils: Successfully started service 'Connection manager for block manager' on port 57218.
14/11/19 14:33:49 INFO ConnectionManager: Bound socket to port 57218 with id = ConnectionManagerId(172.17.42.1:57218)
14/11/19 14:33:49 INFO MemoryStore: MemoryStore started with capacity 265.4 MB
14/11/19 14:33:49 INFO BlockManagerMaster: Trying to register BlockManager
14/11/19 14:33:49 INFO BlockManagerMasterActor: Registering block manager 172.17.42.1:57218 with 265.4 MB RAM
14/11/19 14:33:49 INFO BlockManagerMaster: Registered BlockManager
14/11/19 14:33:49 INFO HttpFileServer: HTTP file server directory is /tmp/spark-399c53ec-0be8-4843-9a7d-9345e9705766
14/11/19 14:33:49 INFO HttpServer: Starting HTTP Server
14/11/19 14:33:49 INFO Utils: Successfully started service 'HTTP file server' on port 49088.
14/11/19 14:33:49 INFO Utils: Successfully started service 'SparkUI' on port 4040.
14/11/19 14:33:49 INFO SparkUI: Started SparkUI at http://172.17.42.1:4040
14/11/19 14:33:49 INFO AkkaUtils: Connecting to HeartbeatReceiver: akka.tcp://sparkDriver@172.17.42.1:35821/user/HeartbeatReceiver
Welcome to
  _____
 /_ _ _ _ _ \
| | | | | |
| |_|_|_|_|_|
 \_ _ _ _ _ /
  version 1.1.0

Using Python version 2.7.6 (default, Mar 22 2014 22:59:56)
SparkContext available as sc.
>>>

```

Figure 1. Initializing spark

The first thing a Spark program must do is to create a SparkContext object, which tells Spark how to access a cluster. To create a SparkContext you first need to build a SparkConf object that contains information about your application Spark as shown in Figure 1. Streaming is a component of Spark which facilitates live stream data processing [34, 35]. Instances of data streams are the generated log files by production web servers or message queues which contains status updates posted by a web service user [36, 37]. The API for the manipulation of data streams which closely matches the RDD API of

the Spark Core is provided via Spark Streaming, giving programmers the chance of learning the project and switching between apps which manipulate the stored data in memory, on disk, or data arriving in real-time. Spark Streaming was also developed to provide an equivalent level of throughput, fault tolerance, and scalability as Spark Core [38-43]. Stream processing at a high level is all about the incessant processing of unbounded data streams. However, it is a difficult task to do this in a consistent and fault-tolerant manner [44]. However, there have been improvements in the stream processing engines such as Spark, Heron, Kafka, Flink, and Samza over the past few years which enables the development and operation of complex stream processing apps by businesses [45, 46]. Spark revolves around the concept of a resilient distributed dataset (RDD), which is a fault-tolerant collection of elements that can be operated on in parallel. The number of data pieces in a batch depends on the rate of incoming data and on the batch interval. The concept of DStream at a high level is shown in Figure 2.



Figure 2. DStream RDD

3. RESILIENT DISTRIBUTED DATASET (RDD)

The concept of RDD is pretty unique in the domain of distributed data processing as they are introduced to address the problems of complexity and efficiency of both interactive and iterative data processing instances [47-50]. Spark 2.0 gives Spark users the leverage of not having to be having a direct interaction with RDD, but it is important to provide them with the robust mental model of the concept of RDD. Figure 3 showed the Spark Executor window. In brief, Spark depends on the RDD concept where both the idea of a large dataset representation in Spark and the idea for working with it are presented. As immutable, fault-tolerant, parallel data structures, RDD allows users to clearly persist intermediate results in memory, optimize data placement via partitioning, and use a set of rich operators to manipulate them [51, 52].

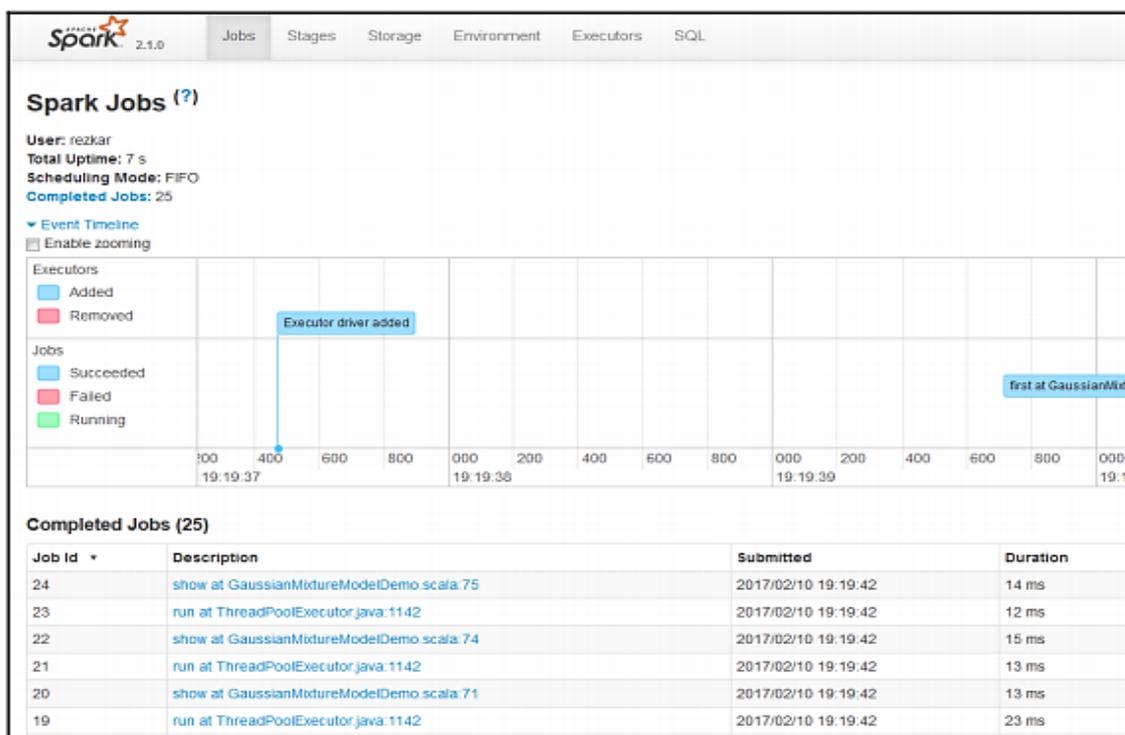


Figure 3. Spark executor

As showing in Figure 3 Spark Executer, so to execute jobs, Spark breaks up the processing of RDD operations into tasks, each of which is executed by an executor. Prior to execution, Spark computes the task's closure. The closure is those variables and methods which must be visible for the executor to perform its computations on the RDD (in this case for each()). This closure is serialized and sent to each executor.

4. SENTIMENT ANALYSIS (TEXT ANALYTICS)

Text analytics refers to the ways of extracting information from a text collection [45, 53]. The patterns and themes in a given dataset can be uncovered using several data processing and analysis algorithms and techniques. The major aim of this process is to make the unstructured text meaningful in order to extract the relationships and contextual meaning [54]. The analysis of peoples' political opinions on social networks is a perfect instance of sentiment analysis. The recent trend of tweets is shown in Figure 4. Similarly, the analysis of restaurants reviews on Yelp is another instance of SA [55, 56]. Sentiment analysis is typically implemented using Natural Language Processing (NLP) libraries and frameworks, such as OpenNLP and Stanford NLP.

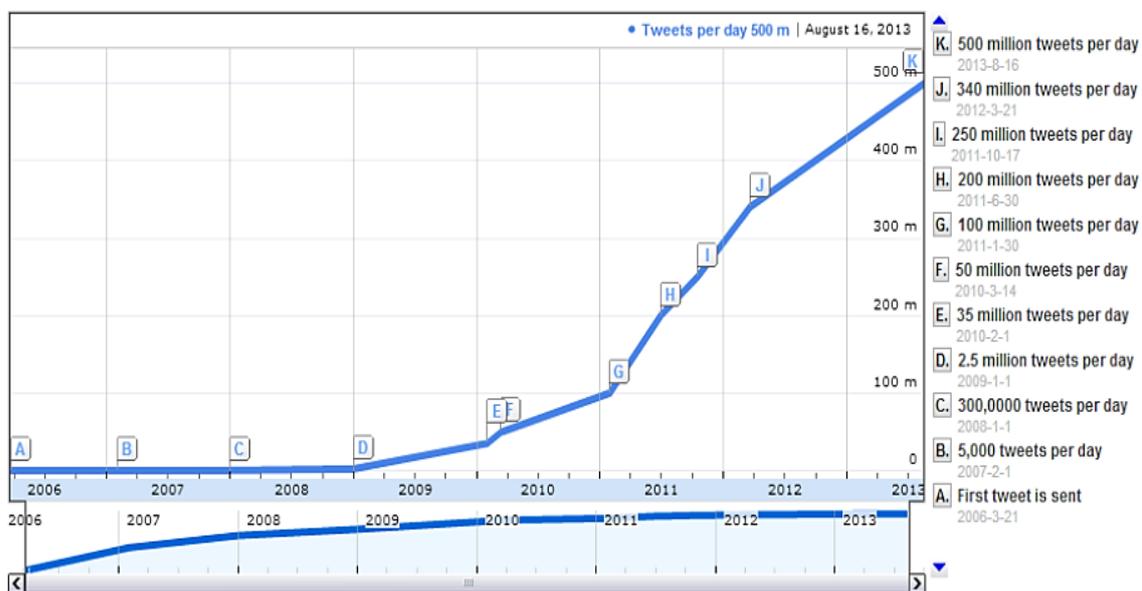


Figure 4. The recent trends of tweets

5. PROPOSED METHOD

The proposed method is to design a fast analysis framework for gathering, processing, prediction, and visualization of Twitter data. There are many advantages of the proposed method that distinguished it from previous sentiment analysis frameworks. The proposed system breaks the challenge of analyzing hundreds of tweets that arrived the system memory per second. The framework provides a solution to the volume of data by HDFS storage of Spark. We provide parallel data gathering nodes and parallel processing nodes for scalable stream data. The challenge of managing the arrived tweets was solved by Kafka. The proposed framework is shown in Figure 5. The tweets arrive from twitter API in batches and are arranged by Kafka into data streams. The data stream is forwarded to Spark engine (Spark Streaming). The new method is based on lexicon-based algorithm using Apache Spark. The case study of the new method is Bassam Al-Rawi Tweets through the hashtag Bassam Al-Rawi on Twitter.

Before data processing through the Spark engine, Spark Streaming will convert the data into RDD form and transfer it to the system memory. The sentiment based on tweets will be categorized into positive, negative, and neutral as Figure 6. If the summations of sentiment of all the words of a tweet are positive, the event will be categorized as positive and placed in the positive bucket in HDFS. The negative sum of sentiment is placed in the negative bucket of HDFS, and the same for neutral. With this process, sentiment of the whole messages can be easily captured and processed and used eventually to derive the live dashboard for monitoring the trends as Figure 7. Apache Kafka was preferred in this study than the original Twitter

streamer in Apache Spark because of the lack of support for many options by the original streamer; hence, we strive to include support for buffering incoming stream for later processing or for when a certain condition is met. It also does not support streaming tweets in certain languages.

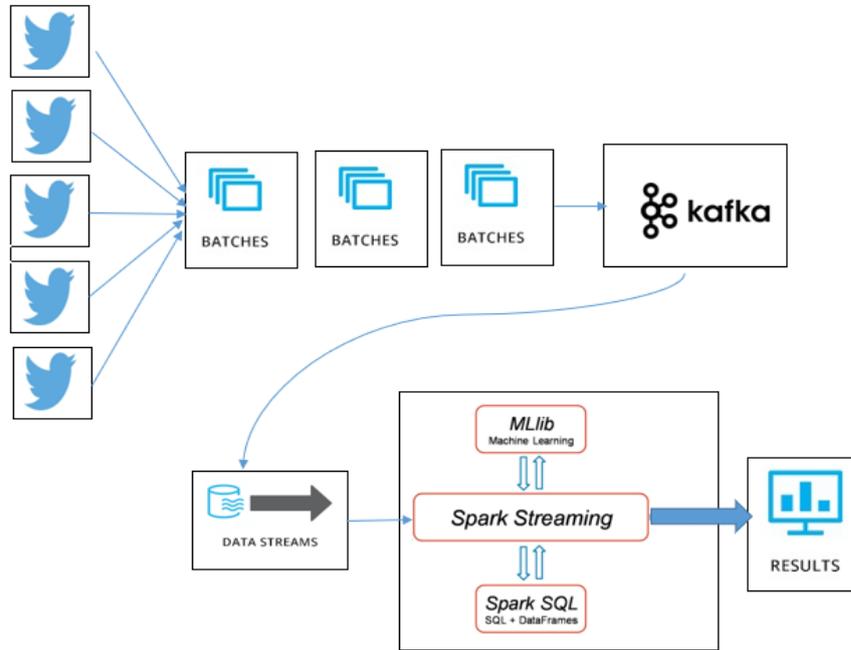


Figure 5. The proposed framework

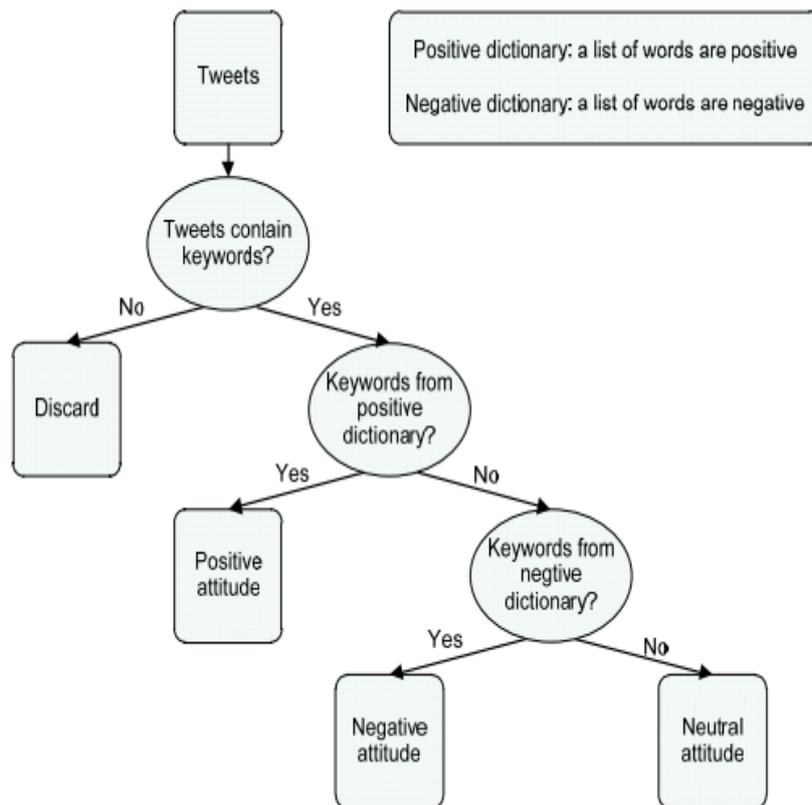


Figure 6. DT of tweets

k-Means - Summary

Number of Clusters: 2
 Distance Measure: Squared Euclidean Distance
 Average Cluster Distance: 9.823
 Davies-Bouldin Index: 0.805

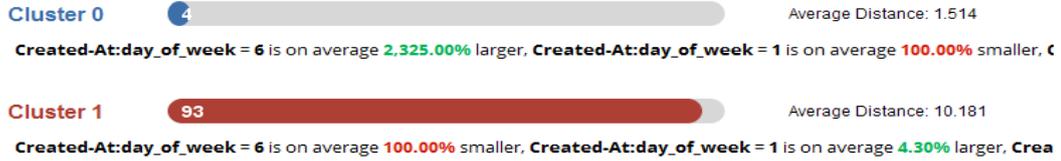


Figure 7. K-Means clusters of tweets

Figure 7 showing the ability of K-Means algorithm can be used to identifies unknown groups in complex and unlabeled data sets. The general parameters that used in this experiment was as follow: the number of clusters 2, the distance measure is squared Euclidean distance, the average cluster distance is 9.823, and the Davies-Bouldin index is 0.805.

6. EXPERIMENT AND RESULTS

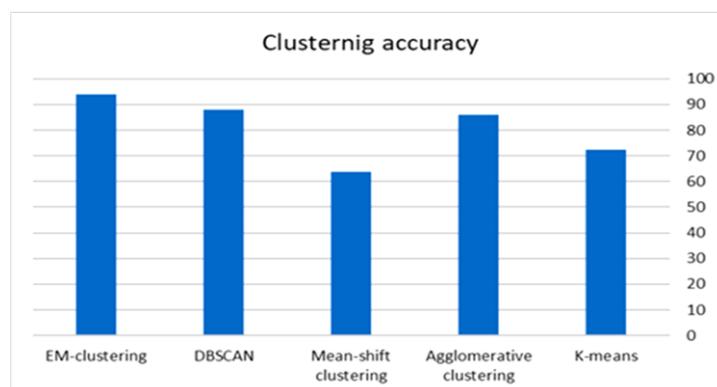
The experiment of our proposed system was based on two implementations. First, we implemented the framework in Weka platform on 30% of Bassam Al-Rawi Twitter dataset. Second, we implemented the same percentage of data in the Spark platform. We also made a comparison in term of time consumption through the processing stage and ingestion stage. The real data sets collected from twitter API according to Bassam Al-Rawi Hashtag. The Qatari player of Iraqi origin, Bassam Al-Rawi, raised controversy in the social networking sites after scoring the only goal of Qatar's quarter-final team that led to the exclusion of Iraq from the Asian Cup 2019 in Emirates as Figure 8.

#####	محمد طوران	756570160769753088	-1	ar	<a href=	والله المفروض بسام الراوي ميحتفل بيهل بطريقة من كول عليه	0.0	1112015215934275592
#####	رحيم	11018611chealth	Bl109604837	ar	<a href=@health_Blogger1	اي????????اي والله كلش يقهون همة حرامات بسام الراوي راحت عليه حسنة	0.0	1111952429376028672
#####	نون الورد	7953186114555968	-1	ar	<a href-RT @al_karbalaeia:	وسخة ماكتبهلوشفته بمنشن بسام الراوي بيوم خسرا متبرد بما معناه انتقوا كلامكم مع الناس	126.0	1111950868063182848
#####	سُ	709519341012828160	-1	ar	<a href=ZyDTH4GSs	صديقي لآعب الدجيل الطغري بسام هشام الراوي يقولك هارلك يا اتحاديه	0.0	1111788432622272512
#####	ترقيوي	820086890	-1	ar	<a href=2	أهداف بسام الراوي الصاروخية على التيبال في تصفيات كأس اسيا 2020 تحت 23 سنة	0.0	1111689588931788880
#####	ReNgo	445333665	ReNgo_Sp528065625	ar	<a href=@ReNgo_Sport	بسام الراوي	0.0	1111609366569013249
#####	Jemma	110849675752740044	-1	ar	<a href-RT @Fjr8557463:	لوما من حطنا نفسي العراق ونفت اسم الوطنيةالوطن حيث نلن لا حيث فولد بسام الراوي	1.0	1111542504216805379
#####	بنين الصاكة	110287574	YounisMa461112276	ar	<a href=@YounisMahmoud10	بسام الراوي كلب ابن كلب	0.0	1111495212256378885
#####	محمود الحراة	781296056	basam_9724520825c	ar	<a href=@basam_97	منور بسام الراوي ربي يحفظك ويرعك ويسهل لك أمرك	0.0	1111313263164379136
#####	DHURGH	287299912	waassnnn285248752	ar	<a href=@waassnnn	بسام الراوي يحفل	0.0	1111307486185820160
#####	Abdulrh	93300262942853297	-1	ar	<a href-RT @alkasschannel:	بي بعد التاهل لكاس اسيا تحت 23 سنة: سنذهب الي تايلاند من اجل اللقب وبطاقة الالمنياد	6.0	1111267671440572418
#####	Bo Saeed	967465176168521728	-1	ar	<a href-RT @alkasschannel:	بي بعد التاهل لكاس اسيا تحت 23 سنة: سنذهب الي تايلاند من اجل اللقب وبطاقة الالمنياد	6.0	1111242486582861824
#####	علي الجويدي	108925318412675072	-1	ar	<a href-RT @galeb_vb:	بي ماعنده شرف.. باع وطنه.. واحتفل كلما احنا ..اعداء @_Momel313	2.0	1111203769562345473
#####	rayahi	761398907721089024	-1	ar	<a href-RT @galeb_vb:	بي ماعنده شرف.. باع وطنه.. واحتفل كلما احنا ..اعداء @_Momel313	2.0	1111202761956302850
#####	خالد الرتيدي	100384216900479795	-1	ar	<a href-RT @Alwsl_123:	??? جوميز ??? جوفينكو ??? الهلال.. اقوى فريقين في غرب قارة آسيا من وجهة نظري	74.0	1111199144926945280
#####	غلب المنصور	737373408	kareems320671391	ar	<a href=@kareemsayed @_	الراوي ..الي ماعنده شرف.. باع وطنه.. واحتفل كلما احنا ..اعداء ابن ا @_Momel313	2.0	1111185115655430144
#####	Ayman Sa	11083150c	kooora285735127	ar	<a href=@kooora	بن تعلية - بسام الراوي	0.0	1111168483771994118
#####	almataree	135662239	-1	ar	<a href=6dsF42loDA	بسام الراوي..رونالد كومان منتخب قطر	0.0	1111061484858011649
#####	bojasseem	965979548	-1	ar	<a href-RT @alkasschannel:	بي بعد التاهل لكاس اسيا تحت 23 سنة: سنذهب الي تايلاند من اجل اللقب وبطاقة الالمنياد	6.0	1111019871129743365
#####	العربي الجديد	2399311781	-1	ar	<a href=t.co/5W5Qj1Xmi	ما وجه التبيه بين بسام الراوي ورونالد كومان يا ترى! الرابط البديل:	1.0	1111016291450519552
#####	قنوات الكاس	304982427	-1	ar	<a href=	بسام الراوي قائد منتخبنا الالمني بعد التاهل لكاس اسيا تحت 23 سنة: سنذهب الي تايلاند	6.0	11110991726133555201
#####	hussein gi	712635065990266880	-1	ar	<a href=	من فترة حكى الكك عن بسام الراوي وانو بييشه كبار منندي الصريرات الحرة بالعالم تحدياً	0.0	11110985851004174337
#####	Melina	891732312	kareems320671391	ar	<a href=@kareemsayed @YShaf	ميشن لل #عل بسام الراوي	0.0	1111090680219484162

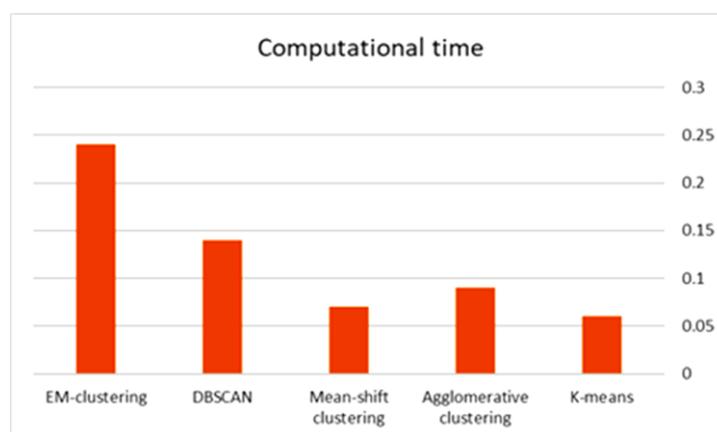
Figure 8. Sample of Iraqi tweets dataset (bassam case study)

This method ensures that all nearby points are in the same cluster. According to the results of k-means, agglomerative clustering and the other clustering method, we can claim that the accuracy of EM method is higher than others. However, the computational time is more, especially for the dataset with higher dimension. The future work to be done is to reduce the computational time of this algorithm to make it more suitable for high dimensional datasets. As well as testing it on more clustering problems and comparing its

performance with other clustering methods. In this study we implement different data clustering algorithms for online dataset (Bassam Al-Rawi dataset) acquired from twitter API. We made a comparison of this algorithms interm of accuracy and computational time. Figure 9 shows the clustering results. Figure 9 (a) and Figure 9 (b) show the clustering result accuracy and the clustering result in term of computational time of EM-clustreing, DBSCAN, mean-shift clustering, agglomerative clustering and K-means function from statistical and Machine Learning toolbox of MATLAB, respectively.



(a)



(b)

Figure 9. (a) The clustering result accuracy, (b) The clustering result in term of computational time

7. CONCLUSION

Sentiment analysis is one of the most significant areas of text analysis. This study suggested a distributed approach to real-time sentiment analysis for Iraqi dialect sentiment on Twitter using a lexicon-based algorithm. This framework was proposed to gather, filter, and mine streams of data in three main phases of ingestion, processing, and visualization. All the components of the proposed framework have been tested and discussed based on Bassam Al-Rawi dataset. The significant improvement is the speed of processing tweets and implementation of some machine learning algorithms, such as DT and K-Means clustering based on lexicon algorithm in Weka compared to Spark. Another output of this framework is the presentation of a method to collect Twitter datasets for future studies in data analysis and other machine learning approaches.

REFERENCES

- [1] H. Luu, "Beginning Apache Spark 2: with resilient distributed datasets, Spark SQL, structured streaming and Spark machine learning library," *Apress*, 2018.

- [2] I. E. Alaoui and Y. Gahi, "The Impact of Big Data Quality on Sentiment Analysis Approaches," *Procedia Computer Science*, vol. 160, pp. 803-810, 2019
- [3] Q. Kang, "Sentiment analysis of big social data with Apache Hadoop," [Honolulu]:[University of Hawaii at Manoa],[December 2014], 2014.
- [4] M. A. Ahmed, R. A. Hasan, A. H. Ali, and M. A. Mohammed, "The classification of the modern arabic poetry using machine learning," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, pp. 2667-2674, 2019.
- [5] T. R. Rao, P. Mitra, R. Bhatt, and A. Goswami, "The big data system, components, tools, and technologies: a survey," *Knowledge and Information Systems*, pp. 1-81, 2019.
- [6] F. Bravo-Marquez, M. Mendoza, and B. Poblete, "Meta-level sentiment models for big social data analysis," *Knowledge-Based Systems*, vol. 69, pp. 86-99, 2014.
- [7] A. Rajput, "Chapter 3 - Natural Language Processing, Sentiment Analysis, and Clinical Analytics," in *Innovation in Health Informatics*, M. D. Lytras and A. Sarirete, Eds., ed: Academic Press, pp. 79-97, 2020.
- [8] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review," *Procedia Computer Science*, vol. 161, pp. 707-714, 2019.
- [9] E. Georgiadou, S. Angelopoulos, and H. Drake, "Big data analytics and international negotiations: Sentiment analysis of Brexit negotiating outcomes," *International Journal of Information Management*, vol. 51, p. 102048, 2020.
- [10] B. Haryanto, Y. Ruldeviyani, F. Rohman, J. D. T.N, R. Magdalena, and Y. F. Muhamad, "Facebook Analysis of Community Sentiment on 2019 Indonesian Presidential Candidates from Facebook Opinion Data," *Procedia Computer Science*, vol. 161, pp. 715-722, 2019.
- [11] S.-U. Hassan, N. R. Aljohani, N. Idrees, R. Sarwar, R. Nawaz, E. Martínez-Cámara, et al., "Predicting literature's early impact with sentiment analysis in Twitter," *Knowledge-Based Systems*, vol. 192, p. 105383, 2020.
- [12] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, and H. Mora, "A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making," *Industrial Marketing Management*, 2019.
- [13] A. Kumar, K. Srinivasan, W.-H. Cheng, and A. Y. Zomaya, "Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data," *Information Processing & Management*, vol. 57, p. 102141, 2020.
- [14] A. Assiri, A. Emam, and H. Al-Dossari, "Real-time sentiment analysis of Saudi dialect tweets using SPARK," in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 3947-3950.
- [15] A. H. Ali and M. Z. Abdullah, "Recent trends in distributed online stream processing platform for big data: Survey," in *2018 1st Annual International Conference on Information and Sciences (AiCIS)*, pp. 140-145, 2018.
- [16] A. H. Ali and M. Z. Abdullah, "A novel approach for big data classification based on hybrid parallel dimensionality reduction using spark cluster," *Computer Science*, vol. 20, 2019.
- [17] M. A. H. Ali, "An Efficient Model for Data Classification Based on SVM Grid Parameter Optimization and PSO Feature Weight Selection," *International Journal of Integrated Engineering*, pp. 1-12, 2020.
- [18] A. H. Ali, "Fuzzy generalized Hebbian algorithm for large-scale intrusion detection system," *International Journal of Integrated Engineering*, vol. 12, pp. 81-90, 2020.
- [19] Z. F. Hussain, H. R. Ibraheem, M. Alsajri, A. Hussein Ali, M. A. Ismail, S. Kasim, et al., "A new model for iris data set classification based on linear support vector machine parameter's optimization," *International Journal of Electrical & Computer Engineering*, pp. 2088-8708, vol. 10, 2020.
- [20] R. A. Hasan, I. Alhayali, A. Royida, N. D. Zaki, and A. H. Ali, "An adaptive clustering and classification algorithm for Twitter data streaming in Apache Spark," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 17, p. 3086, 2019.
- [21] M. A. Mohammed, R. A. Hasan, M. A. Ahmed, N. Tapus, M. A. Shanan, M. K. Khaleel, et al., "A Focal load balancer based algorithm for task assignment in cloud environment," in *2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1-4, 2018.
- [22] R. A. Hasan, M. A. Mohammed, N. Țăpuș, and O. A. Hammood, "A comprehensive study: Ant Colony Optimization (ACO) for facility layout problem," in *2017 16th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, pp. 1-8, 2017.
- [23] R. O. Sinnott, H. Duan, and Y. Sun, "Chapter 15 - A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the Weather," in *Big Data*, R. Buyya, R. N. Calheiros, and A. V. Dastjerdi, Eds., ed: Morgan Kaufmann, pp. 357-388, 2016.
- [24] C. Song, X.-K. Wang, P.-f. Cheng, J.-q. Wang, and L. Li, "SACPC: A framework based on probabilistic linguistic terms for short text sentiment analysis," *Knowledge-Based Systems*, p. 105572, 2020.
- [25] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Information Processing & Management*, p. 102212, 2020.
- [26] F. Mehraliyev, A. P. Kirilenko, and Y. Choi, "From measurement scale to sentiment scale: Examining the effect of sensory experiences on online review rating behavior," *Tourism Management*, vol. 79, p. 104096, 2020.
- [27] J. Ramsingh and V. Bhuvaneswari, "An efficient Map Reduce-Based Hybrid NBC-TFIDF algorithm to mine the public sentiment on diabetes mellitus-A big data approach," *Journal of King Saud University-Computer and Information Sciences*, 2018.
- [28] G. A. Ruz, P. A. Henriquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Future Generation Computer Systems*, vol. 106, pp. 92-104, 2020.

- [29] K. Sailunaz and R. Alhaji, "Emotion and sentiment analysis from Twitter text," *Journal of Computational Science*, vol. 36, p. 101003, 2019.
- [30] H. Shirdastian, M. Laroche, and M.-O. Richard, "Using big data analytics to study brand authenticity sentiments: The case of Starbucks on Twitter," *International Journal of Information Management*, vol. 48, pp. 291-307, 2019.
- [31] O. A. Hammood, M. N. M. Kahar, and M. N. Mohammed, "RElay Suitability-based Routing Protocol (RESP) for Video Streaming in Vehicular Ad Hoc Networks," *International Journal of Computers Communications & Control*, vol. 14, pp. 21-38, 2019.
- [32] M. Petrov, N. Butakov, D. Nasonov, and M. Melnik, "Adaptive performance model for dynamic scaling Apache Spark Streaming," *Procedia Computer Science*, vol. 136, pp. 109-117, 2018.
- [33] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction," *Computers & Electrical Engineering*, vol. 65, pp. 393-399, 2018.
- [34] S. K. Bharti, B. Vachha, R. K. Pradhan, K. S. Babu, and S. K. Jena, "Sarcastic sentiment detection in tweets streamed in real time: a big data approach," *Digital Communications and Networks*, vol. 2, pp. 108-121, 2016.
- [35] A. Bifet, G. Holmes, B. Pfahringer, and R. Gavaldà, "Detecting sentiment change in Twitter streaming data," *Journal of Machine Learning Research - Proceedings Track*, vol. 17, pp. 5-11, 2011.
- [36] S. Das, R. K. Behera, M. kumar, and S. K. Rath, "Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction," *Procedia Computer Science*, vol. 132, pp. 956-964, 2018.
- [37] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, *et al.*, "A performance evaluation of machine learning-based streaming spam tweets detection," *IEEE Transactions on Computational social systems*, vol. 2, pp. 65-76, 2015.
- [38] A.-H. A. Salih, A. H. Ali, and N. Y. Hashim, "Jaya: An Evolutionary Optimization Technique for Obtaining the Optimal Dthr Value of Evolving Clustering Method (ECM)," 2018.
- [39] M. A. Mohammed, A. A. Kamil, R. A. Hasan, and N. Tapus, "An Effective Context Sensitive Offloading System for Mobile Cloud Environments using Support Value-based Classification," *Scalable Computing: Practice and Experience*, vol. 20, pp. 687-698, 2019.
- [40] M. A. Mohammed, I. A. Mohammed, R. A. Hasan, N. Tãpuş, A. H. Ali, and O. A. Hammood, "Green Energy Sources: Issues and Challenges," in *2019 18th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, pp. 1-8, 2019.
- [41] M. A. Mohammed, Z. H. Salih, N. Tãpuş, and R. A. K. Hasan, "Security and accountability for sharing the data stored in the cloud," in *2016 15th RoEduNet Conference: Networking in Education and Research*, pp. 1-5, 2016.
- [42] O. A. Hammood, M. N. M. Kahar, M. N. Mohammed, W. A. Hammood, and J. Sulaiman, "The VANET-Solution Approach for Data Packet Forwarding Improvement," *Advanced Science Letters*, vol. 24, pp. 7423-7427, 2018.
- [43] O. A. Hammood, M. N. M. Kahar, W. A. Hammood, R. A. Hasan, M. A. Mohammed, A. A. Yoob, *et al.*, "An effective transmit packet coding with trust-based relay nodes in VANETs," *Bulletin of Electrical Engineering and Informatics*, vol. 9, pp. 685-697, 2020.
- [44] J. Zou, F. Fekri, and S. W. McLaughlin, "Mining streaming tweets for real-time event credibility prediction in twitter," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pp. 1586-1589, 2015.
- [45] K. Shuang, Q. Yang, J. Loo, R. Li, and M. Gu, "Feature distillation network for aspect-based sentiment analysis," *Information Fusion*, vol. 61, pp. 13-23, 2020.
- [46] N. J. Shoumy, L.-M. Ang, K. P. Seng, D. M. M. Rahaman, and T. Zia, "Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals," *Journal of Network and Computer Applications*, vol. 149, p. 102447, 2020.
- [47] Z. Nabi, "Pro Spark Streaming: The Zen of Real-Time Analytics Using Apache Spark," *Apress*, 2016.
- [48] M. A. Mohammed and N. TãPUŞ, "A Novel Approach of Reducing Energy Consumption by Utilizing Enthalpy in Mobile Cloud Computing," *Studies in Informatics and Control*, vol. 26, pp. 425-434, 2017.
- [49] N. Q. Mohammed, M. S. Ahmed, M. A. Mohammed, O. A. Hammood, H. A. N. Alshara, and A. A. Kamil, "Comparative Analysis between Solar and Wind Turbine Energy Sources in IoT Based on Economical and Efficiency Considerations," in *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, pp. 448-452, 2019.
- [50] Z. H. Salih, G. T. Hasan, and M. A. Mohammed, "Investigate and analyze the levels of electromagnetic radiations emitted from underground power cables extended in modern cities," in *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 1-4, 2017.
- [51] S. Vashishtha and S. Susan, "Fuzzy rule based unsupervised sentiment analysis from social media posts," *Expert Systems with Applications*, vol. 138, p. 112834, 2019.
- [52] J. Zhang, M. Chen, H. Sun, D. Li, and Z. Wang, "Object semantics sentiment correlation analysis enhanced image sentiment classification," *Knowledge-Based Systems*, vol. 191, p. 105245, 2020.
- [53] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, pp. 30-38, 2011.
- [54] M. Solaimani, "Design and Development of Real-Time Big Data Analytics Frameworks," 2017.
- [55] Z. H. Salih, G. T. Hasan, M. A. Mohammed, M. A. S. Klib, A. H. Ali, and R. A. Ibrahim, "Study the Effect of Integrating the Solar Energy Source on Stability of Electrical Distribution System," in *2019 22nd International Conference on Control Systems and Computer Science (CSCS)*, pp. 443-447, 2019.
- [56] S. A.-b. Salman, A.-H. A. Salih, A. H. Ali, M. K. Khaleel, and M. A. Mohammed, "A New Model for Iris Classification Based on Naïve Bayes Grid Parameters Optimization," *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, vol. 40, pp. 150-155, 2018.