Smart City Services Monitoring Framework using Fuzzy Logic Based Sentiment Analysis and Apache Spark

Bahra Mohamed
LIST Department of Computer Science
faculty of sciences and techniques
Tangier, Morocco
bahra002@gmail.com

Fennan Abdelhadi
LIST Department of Computer Science
faculty of sciences and techniques
Tangier, Morocco
afennan@gmail.com

Bouktaib Adil
LIST Department of Computer Science
faculty of sciences and techniques
Tangier, Morocco
bouktaib.adil1@gmail.com

Hmami Haytam
LIST Department of Computer Science
faculty of sciences and techniques
Tangier, Morocco
hmami.h@gmail.com

***Abstract*—Smart city technologies are reshaping our everyday life. New application and systems architectures appears to enhance information and communications technologies, allowing cities to harness their resources more intelligently, also to include citizens in the development of their environments by analyzing the tremendous amount of data generated by people, which cover wide diversity of domains, thus, providing valuable contribution to tackle environmental and socio-economic problems efficiently. The problem encountered under these circumstances is that continuously analyzing users content, especially text content to extract expressed sentiment in real-time is a challenging task that needs scalable and high performance distributed systems, also the opinion of a person concerning a particular city service, politic decision or even a street reputation may not be binary, e.g., positive or negative, due to the fuzzy character of our language. To this end we propose an architecture that will take advantage of distributed and streaming computing framework combined with applying fuzzy logic to classify expressed sentiments in real time, thus to offer insight to city governance and to expand citizen participation in the evolution of their cities.**

Keywords— Smart City, Big Data, Sentiment analysis, Fuzzy logic, Apache Spark, Social media Analysis, Streaming analysis, Lexicon based technique, Classification.

I. INTRODUCTION

Smart cities are the future for our men kind, every day new intelligent technologies appear, taking advantages of all the resources available in order to enhance information and communication systems, and broadly link all facets of society along with tackling the majority of city problems. these technologies can benefit more than ever before from massive and openly available data, which is being generated everyday by different sources like sensors and social media, this latter bring together researcher, analysts, developers and business entities in order to create systems architectures and applications that aims to extract valuable information leading to better and innovative approach for decision making.

Sentiment analysis or opinion mining is one of the challenging processes due to, first the complexity of language and the expressed opinion which may not be always binary (like or don't like) and second, the enormous and continuously generated social data, that touch different topics, languages, and cultures. Given this, we propose a smart city framework that combines new Big Data technologies such as Apache Spark Streaming, in order to process tremendous amounts of data generated by Twitter in real time, and also to take into account the fuzzy aspect of opinions expressed by users, using Fuzzy logic functions that introduces the effect of linguistic hedges.

The remainder of our work is organized as follows: in section 2, we highlight our motivation. Section 3 presents the main concepts. In Section 4, we discuss related work. Moreover, Section 5 details the proposed framework and in section 6, we conclude our work and present future steps.

II. MOTIVATION

Seeking to live a better life is one of the reasons that keep pushing humans toward thinking of new ideas and innovative approaches that take advantage of all kind of data being generated every day and which will become the heartbeat of future cities, i.e., "Smart cities" [1].

Intensive researches and approaches were introduced to tackle sentiment analysis problem and deal with real-time processing, however, the majority of this approaches combines real-time frameworks with traditional machine learning techniques, and consider the expressed sentiment of a user as being binary. Fuzzy logic conversely, can deal with the uncertainty and vagueness character of human language, which makes it a good candidate for sentiment analysis task.

For this aim, we introduce a novel approach that implements fuzzy logic in a distributed and high-performance computing framework (i.e., Apache Spark) to run a fuzzy sentiment classification and extend expressed opinion classes.

III. MAIN CONCEPTS

A. Sentiment Analysis

Sentiment analysis, also stated as opinion mining, is the process of identifying people's expressed opinions, sentiments and preferences toward entities from a written text by applying natural language techniques and machine learning approaches, this analysis is considered a decisive factor to enhance cities services, government, and economic strategies.

This analysis is categorized as a classification problem in which person shared content is placed under different classes that describe his or her sentiment in favor of a subject. Three classification levels exist to tackle the problem [2]:

978-1-7281-4368-2/19/$31.00 ©2019 IEEE

- Document Level Sentiment Analysis: consider the whole opinionated text as being about a single topic and classify it as positive or negative [3].
- Sentence Level Sentiment Analysis: can be seen a short document, this analysis consists of two main steps, where the first aim to determine if the document contains subjective or objective content and the second focus on extracting sentiment related to each sentence within the document [4].
- Feature Level Sentiment Analysis: It is considered a fine-grained analysis of text content that take into account the characteristics/features of an entity, e.g., "the lightning is good in this area" [5].

Also, to address sentiment analysis of user's content, different algorithms and approaches were developed and categorized as: machine learning approaches and lexicon-based approaches which they could be also combined.

B. *Fuzzy Logic*

The reality of our world does not correspond to mathematical precision as Albert Einstein said in "Geometrie und Erfahrung" lecture [6], "As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."

Our thinking as humans does not work as numerical calculations but as approximate reasoning because of manipulating the imprecise nature of information, this was one of the reasons of igniting a trend of researches about vagueness and integration of imprecision in measurements, thus the development of fuzzy logic where Lotfi A. Zadeh was its father. He presents the fuzzy set concept in 1965 seminal paper [7], that aim to formalize vague notions and apply membership functions to associate classes to numbers in order to describe complexes systems behavior.

Fuzzy logic is a moving step from a bivalence system or absolute truth (i.e., true, false) to partial truth, in our example of sentiment classification, it means moving from positive or negative sentiment to more human reasoning formalism that can measure the degree of positivity/negativity. Hereafter we will present a brief definition of fuzzy logic vocabulary and methods.

1) Fuzzy sets: It is a form of a set where elements have a degree of membership by using a membership function that is valued in real unit interval [0,1] which extend the classical crisp sets. If we consider U a universe of discourse[1] and its elements are represnted by x, then for a fuzzy set F of U defined by a membership function $M_F(x)$:

$$M_F(x): U \rightarrow [0, 1], \text{ where } \begin{cases} M_F(x) = 1 \text{ if } x \text{ is totally in } F \\ M_F(x) = 0 \text{ if } x \text{ is not in } F \\ 0 < M_F(x) < 1 \text{ if } x \text{ is partly in } F \end{cases}$$

2) Membership functions: Is a function that represents the fuzziness of a fuzzy set and assigns to each x of U a value from the interval [0, 1], so that for x of the universe U, the value of $M_F(x)$ is considered as a degree of membership of x in F. Some of most used membership functions in fuzzy logic are, Triangular, Trapzoidal, Gaussian functions.

3) Linguistic variables: Was introduced by Zadeh [8, 9, 10], it a variable that has words or sentences as values not numbers, this values are called linguistic values of that variable, (e.g., strongly negative, negative, neutral, positive and strongly positive for a sentiment variable).

4) Linguistic hedges: Named also qualifiers, represent terms that extend the values of a linguistic variable. According to Zadeh [11] , values of a linguistic variable involves primary terms which may be associated with terms considered as operators that can modify the intensity and meaning of their operands such as negation "Not", concentrators (e.g., very, definitely, precisely) and dilators (e.g., rather, quite, nearly).

5) Fuzzy If-Then rules: Represents statements used to formulate statements that comprise fuzzy logic [12]. A form of a fuzzy if-then rule is formulated as:

IF<*Fuzzy proposition*> THEN <*Fuzzy proposition*>

where the IF part is defined as the antecedent and the second part associated with THEN, represent the consequent.

6) Fuzzification and Defuzzification: Are two way process in which the fuzzification process aims to convert a crisp or real value into a fuzzy input by the mean of a membership functions [13], whereas the defuzzification is the reverse process that returns a crisp value from a fuzzy quantity, some of the used methods for defuzzification are: center of gravity, Weight average and max-membership [14].

7) Fuzzy inference: Is the process that integrates all the membership functions, If-Then rules, and operators in order to formulate the mapping from a given input to produce an output, some of the known fuzzy inferences are: Mamdani [15], Sugeno [16], Tsukamoto [17], a comparison is in [18].

C. *MapReduce Model*

To manage massive amount of data and provide high performance computation, MapReduce principle [19] is considered as one of the major programming paradigm that allows the process of large datasets across distributed nodes, the principle behind this paradigm is that model is designed based on two major steps called *Map* and *Reduce* and its execution pass through three phases, map, shuffle and reduce:

- **Map phase**: Each worker node within the cluster applies Map() function to its data partition and generate key-value pairs
- **Shuffle phase**: Is considered as an intermediate phase that uses the key-value pairs from the map phase and organizes them so that each worker node can have all the data associated with the same key.
- **Reduce phase:** In this step, the worker executes the Reduce() function on the data grouped by key and finally generate the output results.

MapReduce Model is adopted by the Apache Hadoop[2] Framework, which is an open source software widely used in distributed and batches processing.

[1]: https://en.wikipedia.org/wiki/Domain_of_discourse
[2]: https://hadoop.apache.org/

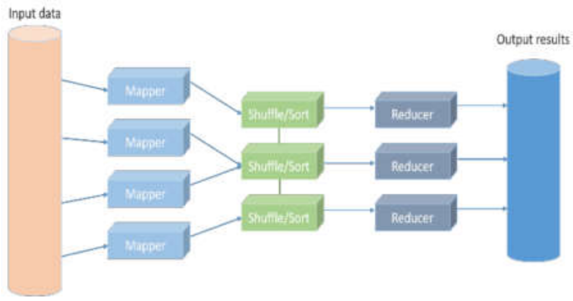

Fig. 1 MapReduce Model

D. *Apache Spark Framework*

Apache Spark[3], is the open source project that is considered as an evolution of Hadoop, also follows the MapReduce principle of Hadoop but with the introduction of a Distributed Acyclic Graph (DAG) that help to optimize the execution plan of more complex jobs, further, it is known by providing a stack of high-level APIs which could be used by different programming languages and it outperforms Hadoop by using a caching system enabling data to be cached and reused in order to faster performance of algorithms that needs many operations. The framework provides in-memory primitives that allow users to load data in memory to be processed with up to 100 times faster than other application.

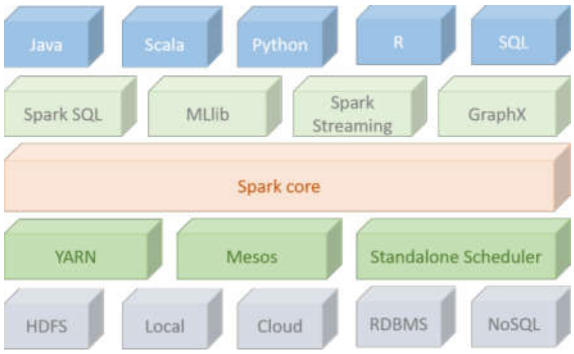

Fig. 2 Apache Spark Framework

IV. Related Work

A study on sentiment analysis has been made by Abhishek Kaushik et al [20] in which they present used methods and techniques and highlight the workflow of opinion mining. A survey by Walaa Medhat el al [21] is yet another rich document that gives an overview of the main classification levels: document level [22], sentence level [23, 24] and aspect level [25, 26] also a wide range of used algorithms in machine learning and lexicon based approaches that were applied to sentiment analysis tasks.

A. Al Nuaimi et al [27] investigate the benefits of social media events analysis in favor of smart city development. In [28] the authors classify microblogging content based on six primary emotions (e.g., “anger”, “disgust”, “fear”, “joy”, “sadness”, and “surprise” introduced by Eckman et al. 1982), to capture crowd pulse in smart city mobile applications, while Amit Gupte et al [29] guided a comparative study of Naive Bayes, Maximum entropy, Boosted Trees and random forest in term of usage scenarios and performances.

Alexandros Baltas et al introduced in [30], a sentiment analysis tool using Apache Spark cloud framework with three supervised machine learning algorithms of MLlib library (Naive Bayes, Logistic regression, and decision trees), moreover, the authors of [31] propose a novel method that use hashtags and emoticons in a tweet to perform sentiment classification and using an adjustment of AkNN to permit a parallel and distributed computation.

An application of Fuzzy logic in sentiment analysis is presented by Karen Howells and Ahmet Ertugan [32] where they depict a fuzzy logic model that use emojis, hashtags and natural text to classify a tweet in a range of five classes, from strongly negative to strongly positive, [33] is another Fuzzy logic based approach for sentiment analysis that make use of linguistic hedges which we will base our model upon to extract citizens opinions in a distributed and real-time architecture.

V. Proposed Architecture

Information is much more important when it comes in the right time and place, e.g., a pre-decision phase in a political environment, people opinions about an announcement of a new product feature, developing a strategic plan or even a reputation of an area while traveling.

Big Data and real-time analysis are nowadays critical factors for developing smart city solutions or support business strategies. In our work, we focus on sentiment analysis field and we present a novel approach in which we combine fuzzy logic with big data technologies such as Kafka, Apache Spark Streaming and Cassandra in a sentiment analysis framework.

The proposed framework is consist of five layers: Data streams, Data integration, Streaming classification, External storage, and Real-time Monitoring.

A. *Data Stream Layer*

In every big data environment, data acquisition is a crucial phase that defines the sources of data that will be processed by the system. In our framework, we consider Twitter Streaming API as a source of streaming data, which allows collecting user’s tweets about deferent topics, this stream of tweets will then be published to Apache Kafka[4] Topic through the use of a Kafka producer.

B. *Data Integration Layer*

This layer involves combining data collected from deferent sources in a high throughput, unified manner.

In this stage, we used Apache Kafka, an open source high-performance software capable of handling real-time data streaming through a distributed publish-subscribe message queue system that makes it massively scalable.

C. *Streaming Classification*

This layer includes two main component: Apache Spark Streaming and Fuzzy logic engine defined as follow.

a) Apache Spark Streaming: is an extension provided by Apache Spark Core API, this component enables real-time processing of data stream in a high-throughput and fault-tolerance manner, the streaming process of the aforementioned extension relies on DStream units that represent a series of RDDs (i.e., Resilient Distributed Datasets).

Spark uses RDDs as a fundamental representation of Data Structure which consists of collections of immutable objects,

[3]: https://spark.apache.org/, [4]: https://kafka.apache.org/

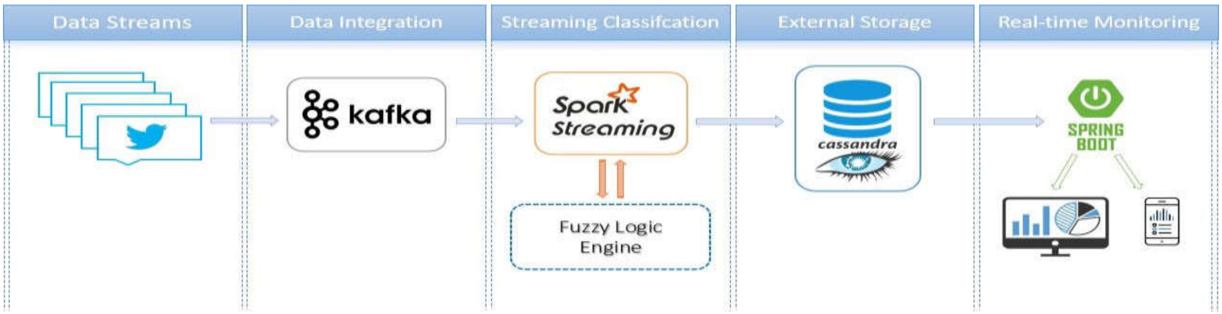

Fig. 3 Proposed sentiment analysis framework

(e.g., a set of tweets), that can be computed in parallel in deferent nodes of the cluster with fault-tolerance.

b) Fuzzy Logic Engine: this component includes the core of fuzzy sentiment analysis process, we will describe the real-time classification algorithm as implemented in Apache Spark.

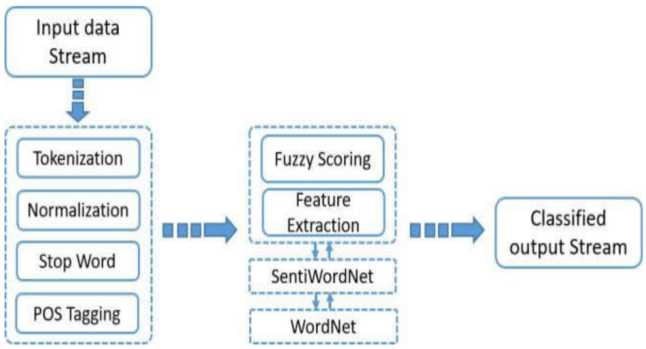

Fig. 4 Fuzzy Sentiment Analysis Process

By using Apache Spark Streaming, Data received from Kafka stream is transformed to DStreams, which are mini-batches of RDDs that contains tweets data from a certain interval.

As illustrated in Fig.4, tweets from stream data source pass through a preprocessing phase that involves:

a) Tokenization: where a tweet text *t* is broken up into words called tokens, $t = \{w_1, w_2, \ldots, w_n\}$. for further processing.

b) Cleaning text and normalization: in this phase, we convert found imoticones into their text string equivalent, and we process with removing URLs and punctuation then we lowercase the text, we also used a dictionary Tab.1 to handle negation expressed through apostrophes to avoid disambiguation.

TABLE I. DICTIONARY FOR WORDS SENS DISAMBIGUATION

Expressed Words	Uniform Structure
hasn't	has not
don't	do not
wasn't	was not

c) Stop word filtering: Consistes of eliminating the most common words in a language that has no contribution to the final sentiment polarity of a given tweet. A stopword dictionary is used for this task.

d) Part Of Speech Tagging: we used for this task Stanford's POS tagger[5]. Tweet words tokens are assigned a part-of-speech tag to define nouns, adjectives, and adverbs which will be then used as features for sentiment scoring.

e) Featue Extraction and Fuzzy Scoring: Part-of-speech tags, bigrams, and trigrams are used as features for fuzzy sentiment scoring, We identify opinionated words as well as their linguistic hedges. We used SentiWordNet[6] as a dictionary for associating polarity to opinionated words features as initial score value, this later is modified by the presence of linguistic hedges. SentiWordNet is a lexicon database consisting of words with their positive, negative and objective score value.

For an opinioned word, we search in SentiWordNet for corresponding initial score $\mu(s)$, if not found we used WordNet[7] dictionary to extract associated synonyms which are then searched in SentiWordNet for sentiment score. The resulting initial score of features can be inverted, increased or decreased if a complement (e.g., Not), concentrator (e.g., Very) or dilator (e.g., somewhat, more or less) respectively is present. We use fuzzy logic functions proposed by Zadeh to calculate the final fuzzy score as follow:

- If the feature is preceded with a linguistic hedge "Not(negation)", the fuzzy sentiment score is obtained using (1):

$$f(\mu(s)) = 1 - \mu(s) \qquad (1)$$

- If the hedge is a concentrator we use the following rule (2):

$$f(\mu(s)) = [\mu(s)]^2 \qquad (2)$$

- And if the preceding hedge is a dilator, the modified score is calculated using (3) function:

$$f(\mu(s)) = [\mu(s)]^{½} \qquad (3)$$

The final output score for the tweet is normalized and the class label (*t-label*) is assigned to the tweet based on the final fuzzy sentiment score (*fscore*) that is:

$$
\begin{aligned}
fscore \in [0, 1],\ & t\text{-}label = \text{"strongly negative" if } 0 \le fscore \le 0.25 \\
& t\text{-}label = \text{"negative" if } 0.25 < fscore < 0.5 \\
& t\text{-}label = \text{"neutral" if } fscore = 0.5 \\
& t\text{-}label = \text{"positive" if } 0.5 < fscore \le 0.75 \\
& t\text{-}label = \text{"strongly positive" if } 0.75 < fscore \le 1
\end{aligned}
$$

[5]: https://nlp.stanford.edu/software/tagger.shtml,
[6]: http://ontotext.fbk.eu/sentiwn.html, [7]: https://wordnet.princeton.edu/

The output results of the processed data is a continuous DStreams that will be parsed and written to Cassandra database.

D. External Storage

In this layer, we store the labeled tweets obtained from the previous layer to Cassandra database, this later is categorized as one of NoSQL databases that manage massive data storage, easily scalable and offer a near real-time performance. The stored results are then consumed by Spring Boot to create micro-services that will deliver near real-time analytics.

E. Real-time Monitoring

This layer represents the final stage in our proposed framework, we used Spring Boot as backend server that sends latest computed results through web sockets to update analytics dashboards in near real-time, which will help in monitoring smart city services as well as notify citizens based on their personal interest.

VI. Conclution and Futur Work

Sentiment analysis will keep being one of the most interesting fields of research and applications, and which can largely contribute to the development of smart city services, and enhance information and communication systems, however, taking into account most of recent work related to sentiment analysis using high-scale and distributed framework, we can notice that the majority of the implemented models, belongs to traditional machine learning models, and which classify users expressed sentiment based on first-order logic (i.e., an opinionated text document can belong to one class or the other) without considering the nature of human language that is characterized as being most of the time imprecise and non-categorical.

Fuzzy logic has the advantage of dealing with uncertainty and lexical imprecision, and it can measure the degree of an expressed sentiment toward a particular topic, resulting in offering multiple sentiment class levels and more rich information.

In our work we presented a novel sentiment analysis computing approach, in which we attempt to use the power of Apache spark streaming extension to address microblogging data stream and real-time processing and analytics, also we apply fuzzy logic to take into account the vagueness and uncertainty character of human language, and take advantage of using fuzzy linguistic functions to dial with negation challenge as well as to reveal the intensity of sentiment expressed by a citizen towards city services.

As future work, we will continue with implementing the proposed architecture. As this is a work in progress, we need to experiment with the results of combining Apache Spark Streaming and fuzzy logic for sentiment analysis and compare our results with other used approaches in term of performance and quality of results. Moreover, we plan to combine neural network and fuzzy logic and run a comparative study of the two approaches.

References

[1] Kitchin, "The real-time city? big data and smart urbanism". GeoJournal vol. 79, Issue 1, pp. 1–14, February 2014.

[2] S. Kolkur, G. Dantal, R. Mahe, "Study of Different Levels for Sentiment Analysis", International Journal of Current Engineering and Technology, vol. 5, no. 2, April 2015.

[3] R. Moraes, J. F. Valiati, W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN", vol. 40, pp. 621-633, Issue 2, 1 February 2013.

[4] V. S. Jagtap, K. Pawar, "Analysis of different approaches to Sentence-Level Sentiment Classification", International Journal of Scientific Engineering and Technology, vol. 2 Issue 3, pp. 164-170, April2013.

[5] P. P. Tribhuvan, S.G. Bhirud, A. P. Tribhuvan, "A Peer Review of Feature Based Opinion Mining and Summarization", (IJCSIT) International Journal of Computer Science and Information Technologies, vol. 5 (1), pp. 247-250 , February 2014.

[6] A Einstein, "Geometrie und Erfahrung" . Sitzungsberichte der PreussischenAkademie der Wissenschaften, pp. 123-130, Januar 1921.

[7] L.A.Zadeh, "Fuzzy sets", Information and control 8 (3), vol. 8, Issue 3, pp. 338-353, June 1965.

[8] L.A.Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I", Information Sciences, vol. 8, Issue 3, pp. 199-249. 1975

[9] L.A.Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II", Information Sciences, vol. 8, Issue 4, pp. 301–357. 1975

[10] L.A.Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—III", Information Sciences, vol. 9, Issue 1, pp. 43–80. 1975.

[11] L. A. Zadeh, "A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges", Journal of Cybernetics, vol 2, Issue, pp. 4-34, 1972.

[12] V. Novák, S Lehmke, "Logical structure of fuzzy IF-THEN rules", Fuzzy Sets and Systems, vol. 157, Issue 15, pp 2003-2029, 1 August 2006.

[13] F. Díaz-Hermida, D.E. Losada, A. Bugarín, S. Barro, "A Probabilistic Quantifier Fuzzification Mechanism: The Model and Its Evaluation for Information Retrieval", IEEE Transactions on Fuzzy Systems, vol. 13 , Issue. 5 , October 2005.

[14] T.A Runker, "Selection of appropriate deffuzification methods using application specific properties", ieee transactions on fuzzy systems, vol. 5, no. 1, pp. 688 – 700, february 1997.

[15] E.H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant", Proceedings of the Institution of Electrical Engineers vol. 121, Issue. 12 , pp. 1585-1588. December 1974.

[16] T. Takagi, M. Sugeno, "Fuzzy Identification of Systems and Its Applications to Modeling and Control", IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-15, Issue. 1, pp. 116-132, Jan.-Feb. 1985.

[17] S.Murakami, "Application of Fuzzy Controller to Automobile Speed Control System", IFAC Proceedings Volumes, vol 16, Issue 13, pp. 43-48, July 1983.

[18] S. Chaudhari, M. Patil,"Study and Review of Fuzzy Inference Systems for Decision Making and Control", Am Int J Res Sci, Technol, Eng Math, 5 (1), pp. 88-92, 2014.

[19] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters", Proceedings of the 6th conference on Symposium on Opearting Systems Design & Implementation, vol. 6, pp. 137-150, January 2004.

[20] A. Kaushik, A. Kaushik, S. Naithani,"A Study on Sentiment Analysis: Methods and Tools", International Journal of Science and Research (IJSR), vol. 4, Issue 12, December 2015.

[21] W. Medhat, A. Hassan, H. Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal, vol. 5, Issue 4, pp. 1093-1113, December 2014.

[22] A. Yessenalina, Y. Yue, C. Cardie, "Multi-level Structured Models for Document-level Sentiment Classification", Conference on Empirical Methods in Natural Language Processing, pp. 1046-1056, MIT, Massachusetts, USA, 9-11 October 2010.

[23] R Jayashree, M. K Srikanta, "An analysis of sentence level text classification for the Kannada language", International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2011.

[24] N. Farra, E. Challita, R. Abou Assi, H. Hajj, "Sentence-level and Document-level Sentiment Mining for Arabic Texts", IEEE International Conference on Data Mining Workshops, 2010.

[25] K. Schouten, F. Frasincar, "Survey on Aspect-Level Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, vol. 28, Issue: 3, pp. 813 – 830, March 1 2016.

[26] V.K. Singh, R. Piryani, A. Uddin , P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification", Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International MultiConference on. IEEE, pp. 712-717, 2013

[27] A. Al Nuaimi, A. Al Shamsi, A. Al Shamsi, Elarbi Badidi, "social media analytics for sentiment analysis and event detection in smart cities", 4th International Conference on Natural Language Computing, 2018.

[28] A. Vakali, D. Chatzakou, V. Koutsonikola, G. Andreadis, "Social data sentiment analysis in smart environments extending dual polarities for crowd pulse capturing", in DATA, pp. 175–182, 2013.

[29] A. Gupte, S. Joshi, P. Gadgul, A. Kadam, "Comparative Study of Classification Algorithms used in Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 (5), pp. 6261-6264, 2014.

[30] A. Baltas, A. Kanavos, A. K. Tsakalidis, "An Apache Spark Implementation for Sentiment Analysis on Twitter Data", International Workshop on Algorithmic Aspects of Cloud Computing (ALGOCLOUD) vol 10230, pp. 15–25, 2016.

[31] N. Nodarakis, S. Sioutas, A. Tsakalidis, and G. Tzimas, "Large Scale Sentiment Analysis on Twitter with Spark", EDBT/ICDT Workshops, 2016.

[32] K. Howells, A. Ertugan, "Applying fuzzy logic for sentiment analysis of social media network data in marketing", Procedia Computer Science, vol. 120, pp. 664-670, 2017.

[33] K. Indhuja, R. P C Reghu, "Fuzzy Logic Based Sentiment Analysis of Product Review Documents", First International Conference on Computational Systems and Communications (ICCSC), 2014.