

Detecting and Fact-checking Misinformation using “Veracity Scanning Model”

Yashoda Barve¹, Jatinderkumar R. Saini^{2*}, Ketan Kotecha³, Hema Gaikwad⁴

Suryadatta College of Management, Information Research & Technology, Pune, India¹

Symbiosis Centre for Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India³

Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University), Pune, India^{2, 4}

Abstract—The expeditious flow of information over the web and its ease of convenience has increased the fear of the rampant spread of misinformation. This poses a health threat and an unprecedented issue to the world impacting people’s life. To cater to this problem, there is a need to detect misinformation. Recent techniques in this area focus on static models based on feature extraction and classification. However, data may change at different time intervals and the veracity of data needs to be checked as it gets updated. There is a lack of models in the literature that can handle incremental data, check the veracity of data and detect misinformation. To fill this gap, authors have proposed a novel Veracity Scanning Model (VSM) to detect misinformation in the healthcare domain by iteratively fact-checking the contents evolving over the period of time. In this approach, the healthcare web URLs are classified as legitimate or non-legitimate using sentiment analysis as a feature, document similarity measures to perform fact-checking of URLs, and incremental learning to handle the arrival of incremental data. The experimental results show that the Jaccard Distance measure has outperformed other techniques with an accuracy of 79.2% with Random Forest classifier while the Cosine similarity measure showed less accuracy of 60.4% with the Support Vector Machine classifier. Also, when implemented as an algorithm Euclidean distance showed an accuracy of 97.14% and 98.33% respectively for train and test data.

Keywords—Document similarity; fact-checking; healthcare; incremental learning; misinformation; sentiment analysis

I. INTRODUCTION

The exponential growth of the internet and World Wide Web (WWW) and its ease of convenience, has led to an information flow expeditiously. Social media, especially Facebook and Twitter have become major sources for information sharing. The expediency, diversified knowledge, and reasonable cost attract the users of the internet to access and share information online, leading to a rapid generation of information [1]. In the healthcare domain, an enormous volume of health and medical-related material is accessible online. It was observed that physicians choose the web as a valuable information resource for medical practice, education, or learning as well as decision support while patients surf the internet for information on diseases, infections, and their indications. For example, 65% of users prefer the internet to search health-related topics [2, 3, 4]. According to the survey in 2017, by Pew Research Center, 88% of American people have quick access to the internet at home and 81% of them get updates of news from the internet [5]. Therefore, it can be

determined that the users make maximum usage of the internet for information access.

However, the material made available online doesn’t guarantee quality as well as correctness. The credibility and veracity of information is a major concern as it may lead to the rampant spread of misinformation [3, 4]. Misinformation is inaccurate or incorrect information that can be verified with available facts. The misinformation or false information may appear in various forms like fake news, rumor, satire news, hoaxes, misinformation, disinformation, etc. This massive spread of misinformation over the web has detrimental effects on people’s life [6].

Apart from the existing health crisis, the spread of the ubiquitous problem of misinformation poses additional health threats and presents another unprecedented issue to the world [7, 8]. This creates a severe effect on people’s life and medical experts as well [9]. For example, during the recent Covid-19 pandemic, misinformation about ingesting fish tank cleaning products can cure the virus or 5G networks generate radiations that triggers the virus or statement like “coronavirus is just like the flu” or “coronavirus is an engineered bioweapon” had an impact on people that they started believing the misinformation. Such misinformation causes panic amongst citizens and may lead to death [5, 10]. During 2014, Ebola outbreak, misinformation on the web and social media about some products which can cure Ebola had led to deaths [2]. Another example, misconception about the measles, mumps, and rubella (MMR) vaccine producing autism had a negative societal impact. Therefore, detecting misinformation has become a necessity to provide timely, verified, and credible information to the users in a way that can benefit society as a whole. Failure to meet this requirement can promote the misuse of misinformation which has adverse effects [1, 5, 10].

Researchers have been passionate about finding solutions to the misinformation detection problem. For example, recently, big social media companies like Facebook, Twitter, and Google have developed machine learning and deep learning-based models to detect misinformation on Covid-19 related posts and ads. In this, Facebook reported that they have had identified and deleted around 50 million posts on Covid-19 while Google and Twitter have taken corrective actions to remove scammer ads on face masks, hand sanitizers, etc. [10]. However, simply detecting misinformation cannot guarantee the veracity or credibility of information. Hence, fact-checking has an increasing demand for veracity scanning of information that can classify information as true or false [11].

*Corresponding Author.

Fact-checking is assessing the truthiness of information that is under investigation in an attempt to identify whether the information is factual [11, 12]. Automatic fact-checking refers to checking the truthiness of the information repeatedly based on all available data and classifying it into True, False, Mostly True, Mostly False, and Half True. According to [12] the process of fact-checking involves identifying the context of the claim, identifying new and previously fact-checked claims, and performing fact-checking with existing verified. According to the literature, there are three main techniques to perform fact-checking based on the evidence used to determine the veracity of the information. The first is the reference approach, these are based on valid or recognized sources and claims which are fact-checked beforehand. Second, knowledge graph approaches, which are based on subject-predicate-object triples for fact-checking [2]. The third category is contextual approaches, which involve perceptive about societal and other context-related claims. Many researchers have developed models and fully automated tools like VERA and Claimbuster to fact-check claims in all three categories. The knowledge graphs and contextual approaches showed higher accuracy values on author-generated datasets but accuracy decreased on different datasets like FEVER and HeroX reaching between 50% and 65% whereas reference approaches succeeded between 77% and 82%. These results reveal that automated fact-checking is a challenging task to resolve fully [12].

The main challenges in the automated fact-checking process involve 1) unavailability of standard annotated datasets in a specific domain. For example, fact-checking websites like politifact.com or fact-check.org mainly focus on specific domains viz. politics, news, etc. Thus, obtaining comprehensive datasets from these websites in a certain domain is not possible. 2) Expert and human annotation is extremely time-consuming and costly. It was studied that the most reliable approach in misinformation detection is to perform human expert-based fact-checking of data. However, with the large volume of data and the haste with which the misinformation is generated and disseminated uncontrollably, manual fact-checking is become time-consuming and might not be able to stop the impact of misinformation in its early stages [7]. 3) Verifying the truthiness of contents with the knowledge base. Therefore, there is a pressing need to design a dynamic and automatic fact-checking model to detect and verify healthcare misinformation [7, 12, 13].

Hence, to deal with data drift occurring in the model and to detect misinformation by iteratively performing fact-checking the authors have proposed a Veracity Scanning Model (VSM) using a combination of techniques viz. incremental learning, sentiment analysis, and standard document similarity measures.

A hybrid approach of incremental learning, sentiment analysis, and document similarity can help to detect misinformation as well as perform fact-checking with already verified data, and also handle the newly arriving chunk of data on the web automatically. Following are the research objectives.

1) To develop a methodology to perform automatic fact-checking using standard document similarity measures viz.

Euclidean Distance, Jaccard Distance, and Cosine Similarity and classify healthcare URLs as Legitimate or Non-Legitimate using Veracity Scanning Model (VSM).

2) To evaluate and validate the performance of the proposed model.

The remaining section of the paper is structured as follows: Section II discusses the literature, Section III explains the methodology and Section IV highlights the results and discussion followed by Section V conclusion and future enhancements.

II. LITERATURE SURVEY

This section discusses various techniques used in the literature to tackle above mentioned challenges. Section A describes the reason behind using incremental learning, section B elaborates on misinformation detection techniques and section C focuses on fact-checking methods.

A. Incremental Learning Approach

The classic problem of false information or misinformation detection and fact-checking deals with the static data and does not consider the streaming nature of the data. The profile of information classified as true and false may change over time. This results in a phenomenon called concept drift of data drift. In the literature, the researchers have fingered such problems using techniques like ensemble learning, or incremental learning [14, 15]. The technique of ensemble learning involves dividing the data stream into small chunks and then training each of the data chunks with different classifiers and ultimately choosing the best classifier. These types of algorithms are recommended to handle sudden or rigorous concept drift and are not much suitable for incremental drift of data [14]. In [15] the authors have used ensemble learning technique with online Bagging with classifiers viz. multi-layer perceptron, Gaussian Naïve Bayes Hoeffding Tree. Incremental Learning (IL) techniques iteratively learn knowledge from newly arriving data without forgetting previously learned knowledge without retraining the model on a complete dataset. Thus, the necessity of the availability of whole labeled data vanishes. Hence, the incremental learning approach is considered to be more suitable to handle smooth concept drifts, and have better performance on efficiency [16, 17, 18 19]. In the literature, researchers have used incremental learning techniques to detect fake news using Artificial Neural Networks (ANN). However, ANNs suffer from catastrophic forgetting which lowers the performance of the model as data streams arrive [20]. The deep learning and neural network-based techniques can classify short text appearing sequentially but require large memory space and training time, thus reducing the performance of the model [19]. Therefore, a novel incremental approach of VSM can be efficiently used to classify the textual data of false information or misinformation.

B. Detecting Misinformation using Sentiment Analysis

Detecting misinformation has gained researchers' attention and is widely focused on politics and mass communication areas. However, less attention is paid to the healthcare domain. The healthcare-related misinformation is studied in five different categories mainly communicable diseases,

infections like Zika Virus, Ebola, influenza, etc., chronic non-communicable diseases, diet and nutrition, smoking, and water safety. The selection of the right features plays a key role in detecting misinformation. In the literature, researchers have focused on several types of features like syntactical, user-specific, image-specific, sentimental, etc. However, sentimental features are found to be the most effective in determining the percentage of misinformation in a document [6, 21]. In [22, 23] authors have focused on sentimental features for healthcare misinformation detection. Thus, in this research authors have considered the sentimental features as a central feature.

C. Fact-Checking using Incremental Learning

The baseline approach for automatic fact-checking using referencing is finding the resemblance among new statements with already fact-checked statements such as Jaccard Distance, Cosine Similarity, Euclidian Distance, Manhattan Distance, etc. [24]. However, the static model can't cope up with incremental data popping up over a period of time. Thus techniques like incremental learning should be adopted. Incremental learning is the process of adapting to the newly arriving data, without the need to reprocess the old instance but remembering previously learned knowledge. In a research incremental learning was adopted to identify new features and new classes as the documents evolve over the time period with the help of incremental neural network based on neural perceptron [25]. To classify documents based on security, a methodology consisting of the combination of incremental learning and similarity features was proposed. Incremental learning is achieved through documental representation and similarity is measured by fetching sentence features. The classification process is based on security labels of already classified documents [26]. In another research, incremental learning for Hierarchical Dirichlet Process (HDP) was used with partial supervision i.e. the training data contains a mixture of labeled and unlabeled documents. Incremental learning is considered for newly arriving data without referring to previously learned data and also maintaining the robustness and consistency of the model. It was observed that the partially labeled dataset makes an important contribution to achieving good accuracy. The model also introduces granular computing to handle unlabeled data [27]. Thus, to update the model automatically after the arrival of new data and consequently classify/cluster the newly arriving documents either into a fixed number of classes or identify new classes or generate new features for document classification or clustering, it is a good approach to combine incremental learning and document similarity techniques [24, 28, 29]. In the author's previous work [1], a fact-checking model was proposed to find misinformation in the healthcare domain. In this research, authors have proposed a new technique of threshold computation function to classify URLs as legitimate or non-legitimate along with incremental learning to deal with data drifts occurring over the period of time and perform fact-checking.

D. Potential Research Gaps Identified

Following is the summarized list of potential research gaps identified through extensive literature from section A, B and C:

1) The research work conducted previously does not tackle the problem of incremental data appearing at different interval of time while dealing with the misinformation detection problem.

2) To the best of the author's knowledge, detecting misinformation via fact-checking is not studied extensively in the literature.

3) Recent techniques in this area focus on extracting features from the text and classifying the text as true or false. However, the authors found that the veracity of information plays a significant role in the classification of misinformation.

III. METHODOLOGY

The Veracity Scanning Model (VSM) detects and performs fact-checking of healthcare data using an incremental learning approach. VSM consists of three main phases viz. Monitoring, Spotting, and Checking. These phases are generated based on the fact-checking process model defined in the literature [12]. The Monitoring phase consists of fetching healthcare URLs and generating sentimental Bag-of-Words (s-BoW). The spotting phase includes extracting features and detecting misinformation based on features extracted. The checking Phase consists of performing fact-checking and ultimately classifying URLs into True or False. This section elaborates on the working of every phase in detail. Fig. 1 displays the detailed methodology of Veracity Scanning Model (VSM) architecture, comprising of both iterations, diagrammatically.

1) *Fetch healthcare-related web URLs*: In this research, the authors have considered a document as a web page or URL text. Thus, to fetch the web URLs the authors have collected URLs from the Google search engine by using the set of keywords related to the healthcare domain. The list of 25 predefined keywords related to the healthcare domain along with their synonyms is maintained to get appropriate search results. The authors have collected 1000 URLs. Apart from these 1000 URLs, authors have fact-checked 200 URLs from healthcare based on expert opinions, existing valid datasets, and manual checking. This dataset of 200 URLs is used for Fact-checking.

2) *Sentimental Bag-of-Words (s-BoW)*: In this phase, the textual contents of each URL are scrapped using a web scraper developed as a part of this research. The newly designed web scrapper can fetch only healthcare-related contents from the URL and remove non-healthcare-related contents. In the pre-processing stage, punctuations, single characters, stop words and duplicate data are removed thus reducing the size of the corpus and removing unwanted information appearing in the text. Further, a sentimental Bag-of-Words (s-BoW) related to the healthcare domain is developed. Initially, s-BoW contains manually identified and labeled sentimental words from the healthcare domain. This s-BoW evolves and grows as the model fetches and extracts new URLs incrementally.

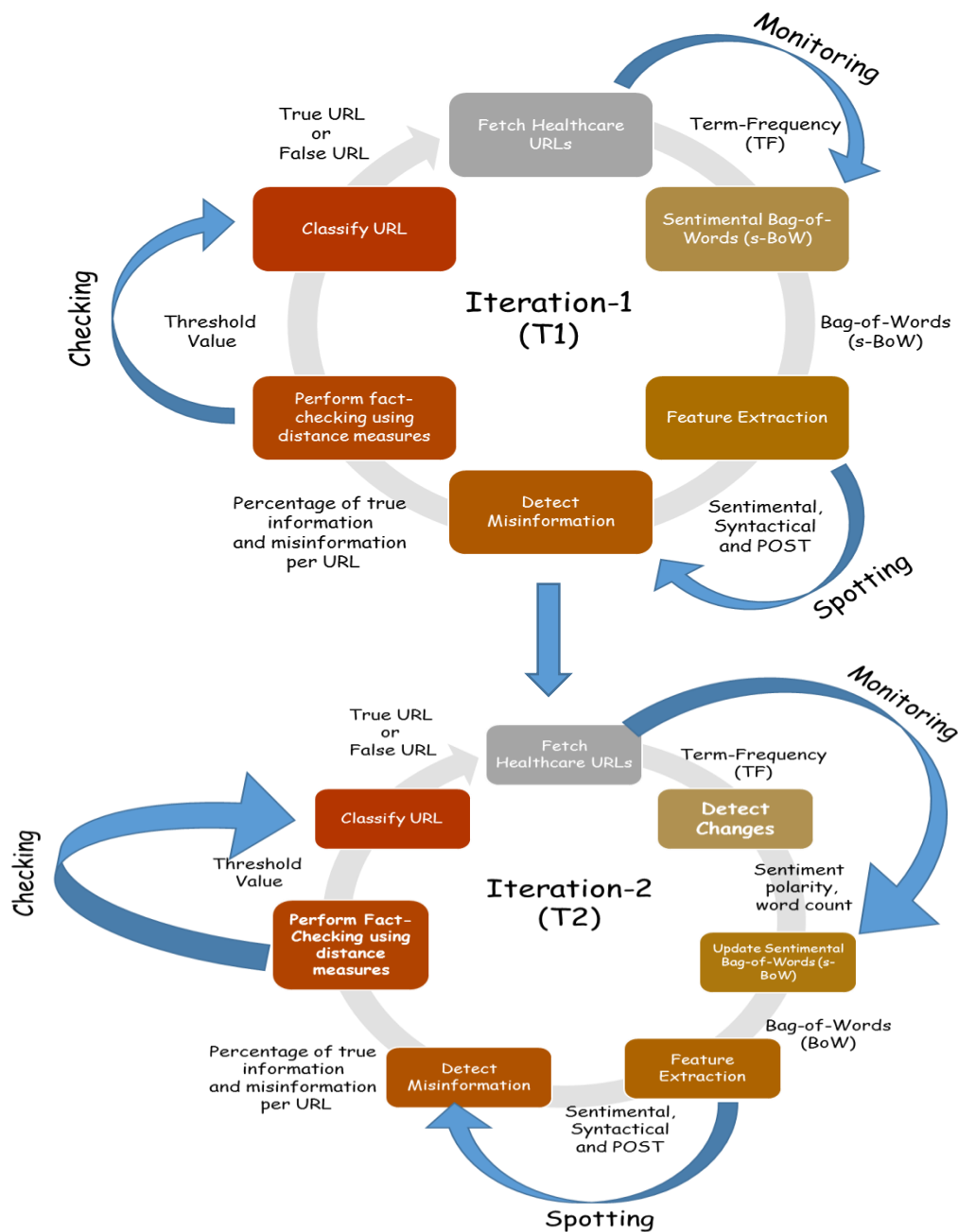


Fig. 1. Diagrammatic Representation of Methodology with the use of 'Veracity Scanning Model (VSM).

3) *Feature extraction*: In this phase, required features are extracted from the text using Term-Frequency (TF). The features extracted include a number of positive and negative words, count of words, nouns, pronouns, adjectives, and distances. The final list of features is the same as that used in the author's previous work.

4) *Change detection*: In the second and subsequent iterations, URLs are fetched to detect changes in the contents. The changes are detected based on the change in the word count, sentimental words, and sentence polarity. These changes are recorded and features are updated accordingly. Also, the sentimental Bag-of-Words is updated based on the

newly arriving sentimental words. This helps to identify misinformation on new content.

5) *Detecting misinformation, perform fact-checking, and classify URLs*: This phase involves detecting the percentage of misinformation in URLs and categorizing them into True or False using a state-of-the-art classifier. The methodology to perform fact-checking is based on the author's previous work [1]. In this research, authors have devised a threshold-based fact-checking algorithm to perform fact-checking. The classification of URLs is based on the threshold value generated. An algorithm to compute the threshold value is shown in Fig. 2. Once the URL is fetched, the distance

between the incoming URL is computed with one of the URLs from the legitimate URL set using standard distance measure formulas of Jaccard Distance, Euclidean Distance, and Cosine Similarity [1]. The process is repeated for all the 2000 URLs at time T2. Further, a threshold value is computed by finding an average of all the distances of all the URLs. Thus, URLs are classified based on this threshold value. Apart from these classifications, Euclidean distance, Jaccard distance, and cosine similarity are used as a feature. The five state-of-the-art classifiers are used for classification viz. Logistic regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT) and Random Forest (RF).

Input: URL
 Output: 1. Classification of URLs into Legitimate or Non-Legitimate
 Algorithmic Steps:

1. Compute distance of URL w.r.t Legitimate URLs using Jaccard, Euclidean, and Cosine similarity measures
2. Compute threshold value T1 using distance with respect to Legitimate URLs separately for each distance measure.

$$T1 = \frac{\sum \text{distance of all incoming URLs}}{\text{Total number of URLs}}$$

3. if (distance > T1)
4. Class= Non-Legitimate URLs
5. else
6. Class= Legitimate URLs

Fig. 2. A Threshold-based Fact-checking Algorithm for Classifying URLs.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section elaborates on the analysis of the results and performance of the model. The classification of URLs as Legitimate (True URLs) and Non-Legitimate (URLs with Misinformation) is performed using state-of-the-art classifiers viz. Logistic Regression (LR), Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and also, threshold-based distance measure algorithms.

A. Performance Evaluation

The performance evaluation is measured through accuracy, precision, recall and F1-score and presented graphically respectively through Fig. 3 to Fig. 6 for document similarity measures on various classifiers. It can be seen that the RF Classifier outperformed the other 79.2% accuracy for the JD measure followed by LR Classifier 78.1% accuracy for JD Measure. The SVM model showed the least performance with an accuracy of 60.4% on the Cosine Similarity measure.

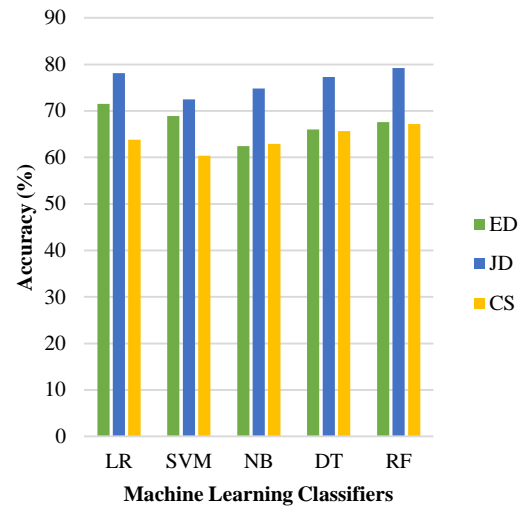


Fig. 3. Accuracy of the Proposed Model in Comparison with Standard Distance Measures.

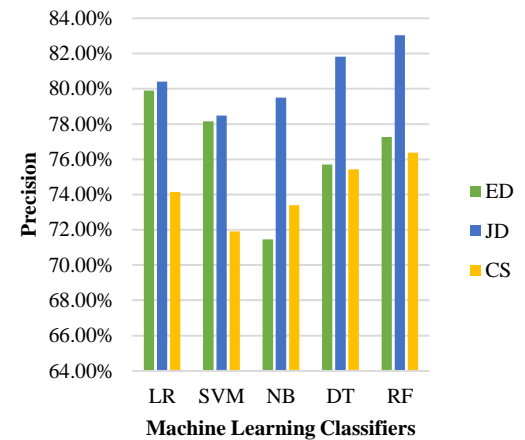


Fig. 4. The Precision of the Proposed Model in Comparison with Standard Distance Measures using Machine Learning Classifiers.

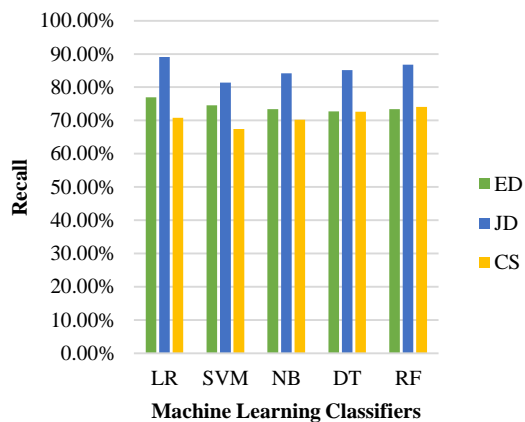


Fig. 5. Recall Matrix of the Proposed Model in Comparison with Standard Distance Measures using Machine Learning Classifiers.

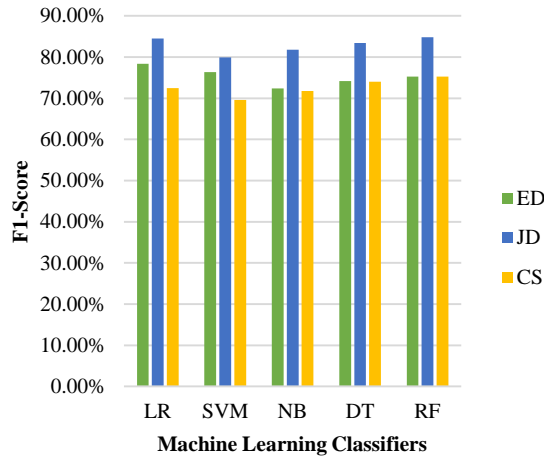


Fig. 6. F1-Score for the Proposed Model in Comparison with Standard Distance Measures.

Tables I to IV show the performance measures of accuracy, precision, recall, and F1-score, all in percentage and concerning distance measures for the five different classifiers, viz. LR, SVM, NB, DT and RF.

TABLE I. ACCURACY IN PERCENTAGE FOR THE PROPOSED MODEL FOR THE DISTANCE MEASURES

Accuracy			
	ED	JD	CS
LR	71.5	78.1	63.8
SVM	68.9	72.5	60.4
NB	62.4	74.8	62.9
DT	66	77.3	65.7
RF	67.6	79.2	67.2

TABLE II. PRECISION IN PERCENTAGE FOR THE PROPOSED MODEL FOR THE DISTANCE MEASURES

Precision			
	ED	JD	CS
LR	79.91%	80.40%	74.14%
SVM	78.16%	78.48%	71.90%
NB	71.45%	79.49%	73.41%
DT	75.70%	81.83%	75.43%
RF	77.27%	83.05%	76.38%

TABLE III. RECALL IN PERCENTAGE FOR THE PROPOSED MODEL FOR THE DISTANCE MEASURES

Recall			
	ED	JD	CS
LR	76.93%	89.14%	70.83%
SVM	74.55%	81.40%	67.41%
NB	73.36%	84.23%	70.24%
DT	72.77%	85.12%	72.62%
RF	73.36%	86.76%	74.11%

TABLE IV. F1-SCORE IN PERCENTAGE FOR THE PROPOSED MODEL FOR THE DISTANCE MEASURES

F1-Score			
	ED	JD	CS
LR	78.39%	84.54%	72.45%
SVM	76.31%	79.91%	69.59%
NB	72.39%	81.79%	71.79%
DT	74.20%	83.44%	74.00%
RF	75.27%	84.86%	75.23%

B. Analysis of the Proposed Model (VSM)

To evaluate the performance of VSM, the authors have analyzed the results of VSM at three different time intervals T1, T2, and T3. For time T1, the data of 2000 URLs were collected and analyzed to detect the percentage of misinformation in URLs and classify them into Legitimate or Non-Legitimate URLs after performing fact-checking. At time T2, once again the 2000 URLs are scrapped to detect any changes in the data. The changes are detected based on the count total number of words, sentimental words, and sentence polarity. Fig. 7 shows the number of URLs changed at time T2 and T3. It can be seen that at time T2 52 URLs have changed while at time T3 22 URLs have shown changes in data. Fig. 8 and Fig. 11 shows the change in percentage of misinformation due to change in incoming data at different time interval T1, T2, and T3. Thus, it can be seen from the figures that 50% of the URLs show a major increase in the percentage of misinformation at times T2 and T3. Another observation is that around 20% of the URLs showed a decrease in the percentage of misinformation. Also, 6 such URLs showed a change in data throughout the three iterations. The fluctuation in the percentage of these 6 URLs is shown in Fig. 10. It can be seen that 50% of URLs have increased in percentage of misinformation. Fig. 12, Fig. 13, Fig. 14 shows the confusion matrix of the VSM model for three-time intervals T1, T2, and T3. Fig. 9 displays the statistical analysis of the VSM model in terms of mean, mode, and standard deviation of legitimate and non-legitimate URLs. Hence, it has become a need of time to track the changes occurring in the data and update the model accordingly to detect and fact-check the newly changed data for misinformation increase or decrease. Incremental learning plays a key role to handle such a situation.

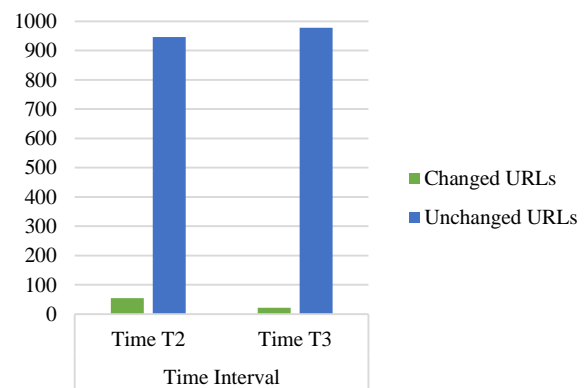


Fig. 7. Number of URLs changed at Time Interval T2 and T3.

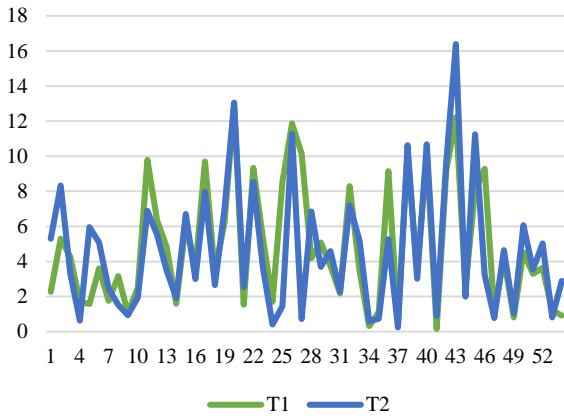


Fig. 8. URLs showing the change in Percentage of Misinformation at T1 and T2.

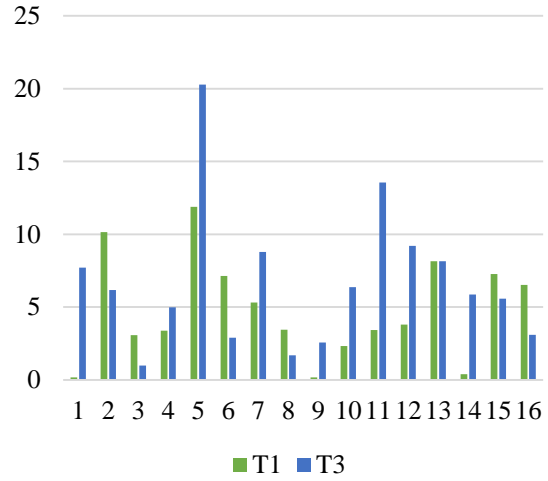


Fig. 11. URLs showing the change in Percentage of Misinformation at T1 and T3.

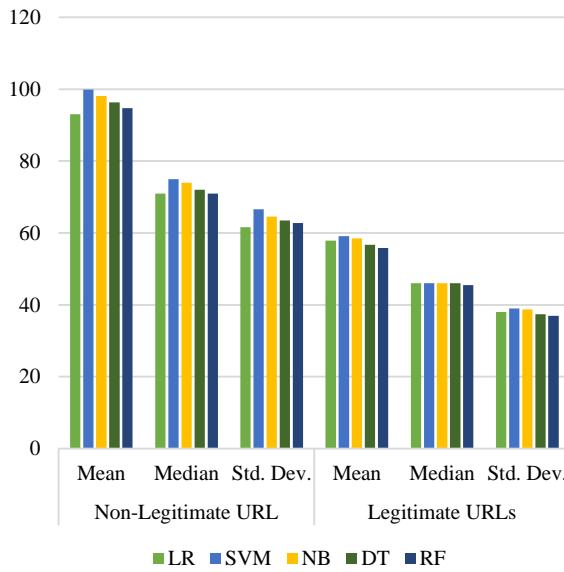


Fig. 9. Jaccard Distance Measure based on Mean, Mode, and Standard Deviation of the URLs.

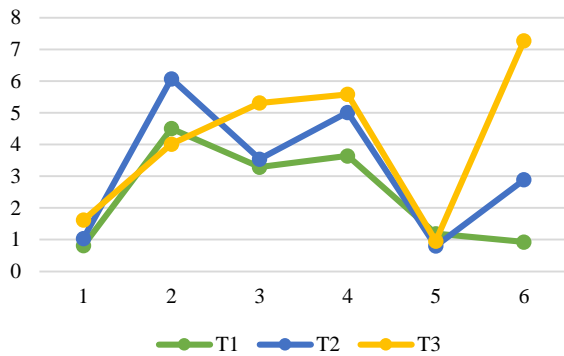


Fig. 10. URLs showing the change in Percentage of Misinformation in Three Iterations.

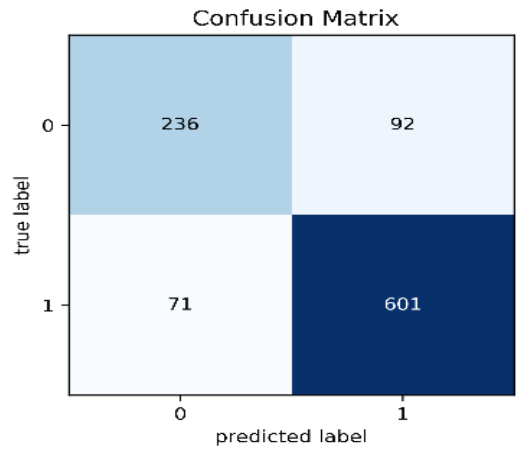


Fig. 12. Confusion Matrix at Time T1.

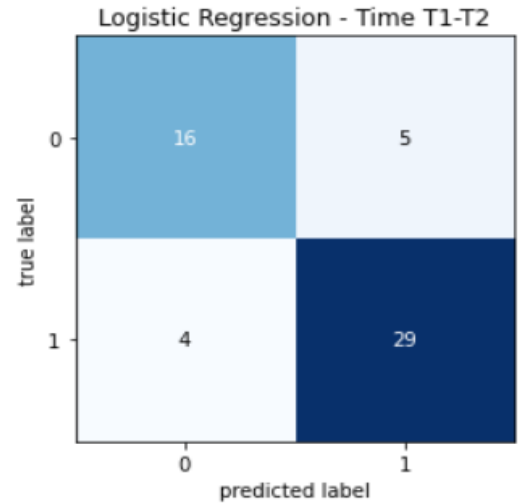


Fig. 13. Confusion Matrix at Time T2.

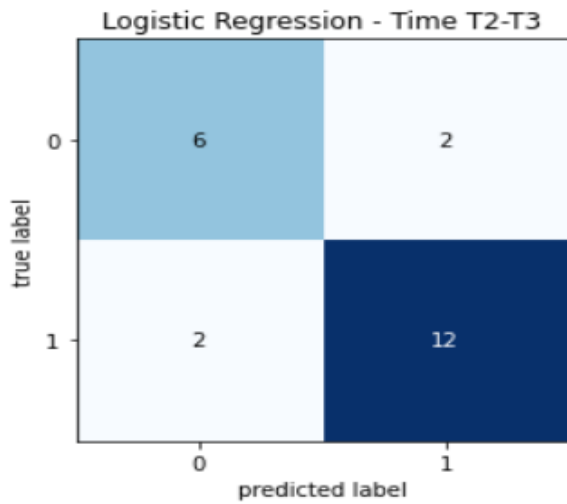


Fig. 14. Confusion Matrix at Time T3.

To evaluate the performance of the model on incremental data, 1000 URLs data was converted into five iterations. Each

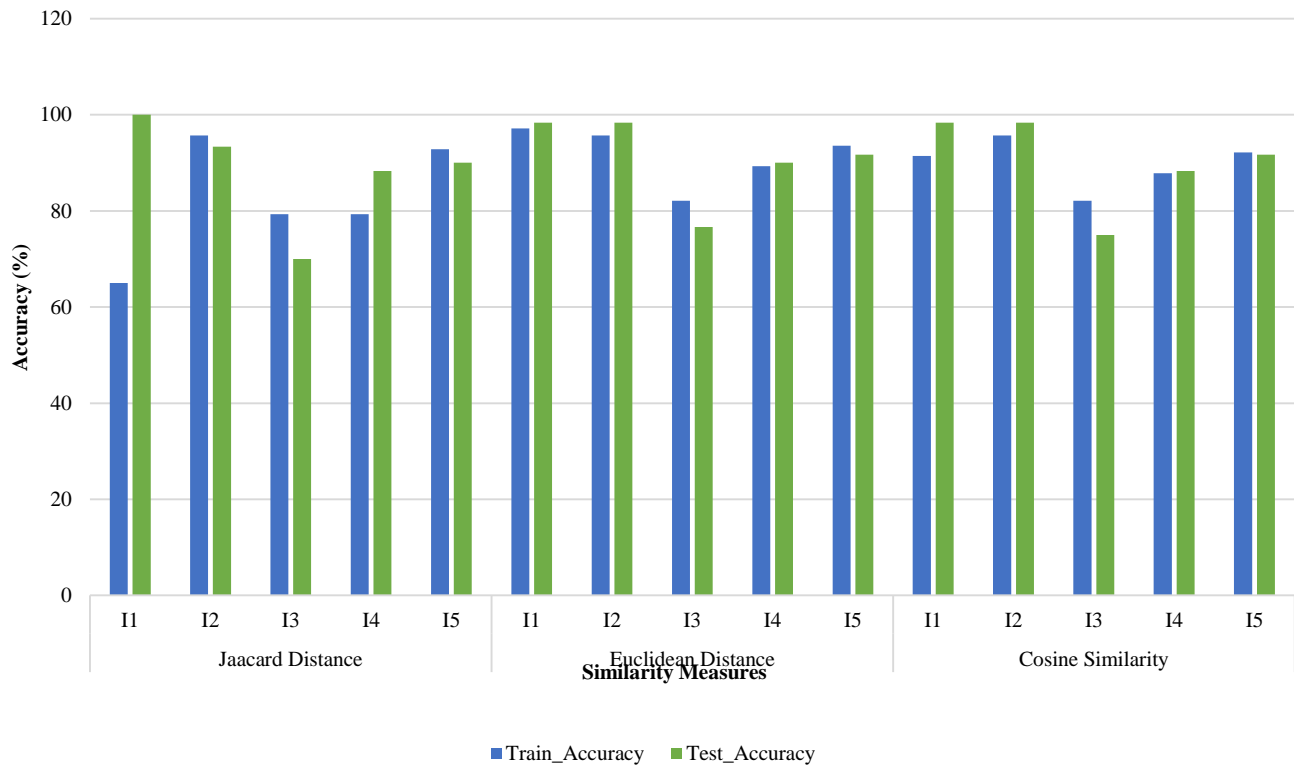


Fig. 15. Accuracy of Similarity Measures based Algorithms for Five Iterations of Incremental Data.

V. CONCLUSION AND FUTURE ENHANCEMENTS

In this research authors have proposed a Veracity Scanning Model (VSM) using incremental learning, sentiment analysis, and document similarity approach. VSM overcomes the limitations of static models which fail to record changes at different time intervals. It was observed that URLs keep changing the contents with time and that fluctuates the percentage of misinformation in URLs. Therefore, to identify

iteration has 200 URLs with a train and test split of 60% and 40% respectively. The threshold values for three distance measures viz. Jaccard, Euclidean, and Cosine are 0.177, 92.06, and 0.27 respectively. It can be seen from Fig. 15 that Jaccard distance showed max accuracy of 100% on test data, thus leading to model overfitting. However, in successive iterations, it showed an accuracy of about 90% to 93%. Euclidean distance measure showed maximum accuracy of 97.14% on training data followed by Jaccard distance and Cosine similarity with approx. 95% of accuracy. Also, the minimum accuracy for training and test data is 65% and 70% respectively using the Jaccard distance measure. Thus overall, the Euclidean distance measure performed well compared to others for both train and test data with an accuracy of 97.14% and 98.33%, respectively.

C. Comparison of VSM Model with Existing Technique

The proposed VSM model outperformed [22] in terms of accuracy. The work [22] showed the highest accuracy of 87.6% with random forest classifier using topic, linguistic, sentiment, and behavioral features while VSM showed an accuracy of 91.67%.

trustworthy URLs, especially in the healthcare domain there is a need for techniques like incremental learning to be adopted. The experimental results show that the Jaccard distance measure outperformed other distance measures with an accuracy of 79.2% with the Random Forest classifier, whereas the cosine similarity measure showed less performance of 60.4% accuracy with Support Vector Machine Classifier. Also, when implemented as an algorithm Euclidean distance

showed an accuracy of 97.14% and 98.33% respectively for train and test data.

In the future, the author wants to propose a new distance measure algorithm to classify URLs into legitimate and non-legitimate URLs and compare the performance with standard distance measures.

REFERENCES

- [1] Y. Barve and J. R. Saini, "Healthcare Misinformation Detection and Fact-Checking: A Novel Approach", *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 10, pp. 295–303, 2021.
- [2] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "DETERRENT: Knowledge Guided Graph Attention Network for Detecting Healthcare Misinformation", in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2020, pp. 492–502, doi: 10.1145/3394486.3403092.
- [3] J. J. McGowan and E. S. Berner, "Proposed Curricular Objectives to Teach Physicians Competence in Using the World Wide Web", *Acad. Med.*, vol. 79, no. 3, pp. 236–240, 2004, doi: 10.1097/00001888-200403000-00007.
- [4] S. Weinstein, "Internet for pathologists", *Pathology*, vol. 30, no. 4, pp. 364–368, 1998, doi: 10.1080/00313029800169646.
- [5] X. Chen, F. Zhou, F. Zhang, and M. Bonsangue, "Catch me if you can: A participant-level rumor detection framework via fine-grained user representation learning", *Inf. Process. Manag.*, vol. 58, no. 5, p. 102678, 2021, doi: 10.1016/j.ipm.2021.102678.
- [6] P. Meel and D. K. Vishwakarma, "Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities", *Expert Syst. Appl.*, vol. 153, 2020, doi: 10.1016/j.eswa.2019.112986.
- [7] S. Shabani, Z. Charlesworth, M. Sokhn, and H. Schuldt, "SAMS: Human-in-the-loop approach to combat the sharing of digital misinformation", in *CEUR Workshop Proceedings*, 2021, vol. 2846, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85104639416&partnerID=40&md5=325c6e737487df99352cf9c7de1ac333>.
- [8] B. Pecher, I. Srba, R. Moro, M. Tomlein, and M. Bielikova, "FireAnt: Claim-Based Medical Misinformation Detection and Monitoring", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12461 LNAI, pp. 555–559, 2021, doi: 10.1007/978-3-030-67670-4_38.
- [9] P. Lara-Navarra, H. Falciani, E. A. Sánchez-Pérez, and A. Ferrer-Sapena, "Information management in healthcare and environment: Towards an automatic system for fake news detection", *Int. J. Environ. Res. Public Health*, vol. 17, no. 3, 2020, doi: 10.3390/ijerph17031066.
- [10] J. Ayoub, X. J. Yang, and F. Zhou, "Combat COVID-19 infodemic using explainable natural language processing models", *Inf. Process. Manag.*, vol. 58, no. 4, p. 102569, 2021, doi: 10.1016/j.ipm.2021.102569.
- [11] X. Zeng, A. S. Abumansour, and A. Zubiaga, "Automated fact-checking: A survey", no. August, pp. 1–21, 2021, doi: 10.1111/lnc3.12438.
- [12] E. Saquete, D. Tomás, P. Moreda, P. Martínez-Barco, and M. Palomar, "Fighting post-truth using natural language processing: A review and open challenges", *Expert Syst. Appl.*, vol. 141, 2020, doi: 10.1016/j.eswa.2019.112943.
- [13] S. Bhatt, N. Goenka, S. Kalra, and Y. Sharma, "Fake News Detection: Experiments and Approaches Beyond Linguistic Features", *Lect. Notes Data Eng. Commun. Technol.*, vol. 71, pp. 113–128, 2022, doi: 10.1007/978-981-16-2937-2_9.
- [14] W. Zang, P. Zhang, C. Zhou, and L. Guo, "Comparative study between incremental and ensemble learning on data streams: Case study", *J. Big Data*, vol. 1, no. 1, pp. 1–16, 2014, doi: 10.1186/2196-1115-1-5.
- [15] P. Ksieniewicz, P. Zybiewski, M. Choraś, R. Kozik, A. Gielczyk, and M. Woźniak, "Fake News Detection from Data Streams", *Proc. Int. Jt. Conf. Neural Networks*, 2020, doi: 10.1109/IJCNN48605.2020.9207498.
- [16] A. Habib, M. Z. Asghar, A. Khan, A. Habib, and A. Khan, "False information detection in online content and its role in decision making: a systematic literature review", *Soc. Netw. Anal. Min.*, vol. 9, no. 1, 2019, doi: 10.1007/s13278-019-0595-5.
- [17] A. Chefrour, "Incremental supervised learning: algorithms and applications in pattern recognition", *Evol. Intell.*, vol. 12, no. 2, pp. 97–112, 2019, doi: 10.1007/s12065-019-00203-y.
- [18] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification", *Expert Syst. Appl.*, vol. 169, 2021, doi: 10.1016/j.eswa.2020.114171.
- [19] Y. Barve and P. Mulay, "Bibliometric Survey on Incremental Learning in Text Classification Algorithms for False Information Detection", *Libr. Philos. Pract.*, vol. 2020, no. November, pp. 2388–2392, 2020.
- [20] M. Umer, G. Dawson, and R. Polikar, "Targeted Forgetting and False Memory Formation in Continual Learners through Adversarial Backdoor Attacks", 2020, doi: 10.1109/IJCNN48605.2020.9206809.
- [21] Z. Xu and H. Guo, "Using Text Mining to Compare Online Pro- and Anti-Vaccine Headlines: Word Usage, Sentiments, and Online Popularity", *Commun. Stud.*, vol. 69, no. 1, pp. 103–122, 2018, doi: 10.1080/10510974.2017.1414068.
- [22] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches", *Inf. Process. Manag.*, vol. 58, no. 1, 2021, doi: 10.1016/j.ipm.2020.102390.
- [23] L. Kinkead, A. Allam, and M. Krauthammer, "Autodiscern: Rating the quality of online health information with hierarchical encoder attention-based neural networks", *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, 2020, doi: 10.1186/s12911-020-01131-z.
- [24] H. Zhang, X. Xiao, and O. Hasegawa, "A load-balancing self-organizing incremental neural network", *IEEE Trans. Neural Networks Learn. Syst.*, vol. 25, no. 6, pp. 1096–1105, 2014, doi: 10.1109/TNNLS.2013.2287884.
- [25] Z. Chen, L. Huang, and Y. L. Murphey, "Incremental learning for Text document classification", in *IEEE International Conference on Neural Networks - Conference Proceedings*, 2007, pp. 2592–2597, doi: 10.1109/IJCNN.2007.4371367.
- [26] Y. Liang, Z. Wen, Y. Tao, G. Li, and B. Guo, "Automatic security classification based on incremental learning and similarity comparison", in *Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019*, 2019, pp. 812–817, doi: 10.1109/ITAIC.2019.8785798.
- [27] D. Wang and A. Al-Rubaie, "Incremental learning with partial-supervision based on hierarchical Dirichlet process and the application for document classification", *Appl. Soft Comput. J.*, vol. 33, pp. 250–262, 2015, doi: 10.1016/j.asoc.2015.04.044.
- [28] T. F. Rodrigues and P. M. Engel, "Probabilistic clustering and classification for textual data: An online and incremental approach", in *Proceedings - 2014 Brazilian Conference on Intelligent Systems, BRACIS 2014*, 2014, pp. 288–293, doi: 10.1109/BRACIS.2014.59.
- [29] T. Doan and J. Kalita, "Overcoming the challenge for text classification in the open world", in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference, CCWC 2017*, 2017, doi: 10.1109/CCWC.2017.7868366.