Higher Classification of Fake Political News Using Decision Tree Algorithm Over Naive Bayes Algorithm

T. Dinesh[1]; Dr.T. Rajendran[2*]

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.

[1]dinesharmi143@gmail.com

[2*]Associate Professor, Project Guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India.

[2*]rajendrant.sse@saveetha.com

Abstract

*Aim: The main aim of the study proposed is to perform higher classification of fake political news by implementing fake news detectors using machine learning classifiers by comparing their performance. **Materials and Methods:** By considering two groups such as Decision Tree algorithm and Naive Bayes algorithm. The algorithms have been implemented and tested over a dataset which consists of 44,000 records. Through the programming experiment which is performed using N=10 iterations on each algorithm to identify various scales of fake news and true news classification. **Result:** After performing the experiment the mean accuracy of 99.6990 by using Decision Tree algorithm and the accuracy of 95.3870 by using Naive Bayes algorithm for fake political news in. There is a statistical significant difference in accuracy for two algorithms is p<0.05 by performing independent samples t-tests. **Conclusion:** This paper is intended to implement the innovative fake news detection approach on recent Machine Learning Classifiers for prediction of fake political news. By testing the algorithms performance and accuracy on fake political news detection and other issues. The comparison results shows that the Decision Tree algorithm has better performance when compared to Naive Bayes algorithm.*

Key-words: Innovative Fake News Detection, Decision Tree Algorithm, Naive Bayes Algorithm, Machine Learning, Statistical Analysis.

1. Introduction

The primary initiation of the study is to implement a fake news detector to detect the fake political news that is published or shared over the social media (Giełczyk, Wawrzyniak, and Choraś

2019). Fake news is not new to this world. It has been growing through centuries. Previous there used to be only limited ways to transfer the fake news, either by people rumoured about it or through letters. But as now decades passed, social media has become a primary platform for exchange of news. Fake news is being shared over social media. Social media is one of platforms where most of the fake news is being shared (Ibrishimova and Li 2020). The application of the research is to improve fake political news prediction and to implement innovative fake news detectors (Paschalides et al. 2019) (Reis et al. 2019).

There are nearly 208 articles published in google scholar and 134 articles published in IEEE Xplore related to fake news deletection. In paper (Shu et al. 2017) a data mining procedure is performed on social media to collect the fake news and convert it into a dataset which can be used for analysis and is cited about 1057 times as reference for research. There are many other machine learning classifiers implemented earlier to detect fake news published or shared over social media has minimal accuracy (Ahmad et al. 2020). In this paper, the implementing a machine learning classifier can provide a better accuracy for fake political news published over social media than the previously implemented classifiers. A fake news detector is implemented which is used to detect the fake news published on social media using multiple sources and various classes, proposed in paper (Karimi et al. 2018). The classifiers which are used in previous proposed papers have less accuracy rate so the implemented Decision Tree classifier to give improved accuracy and comparing it with Naive Bayes algorithm. It is a supervised machine learning algorithm which is similar to the classifiers implemented in paper (Reis et al. 2019) like SVM, Naive Bayes. This paper (Shu et al. 2017) is best for future researchers who are interested in fake news detection as a reference, a data mining procedure is performed on social media to collect the fake news and convert it into a dataset which can be used for analysis.

Previously our team has a rich experience in working on various research projects across multiple disciplines (Sathish and Karthick 2020; Varghese, Ramesh, and Veeraiyan 2019; S. R. Samuel, Acharya, and Rao 2020; Venu, Raju, and Subramani 2019; M. S. Samuel et al. 2019; Venu, Subramani, and Raju 2019; Mehta et al. 2019; Sharma et al. 2019; Malli Sureshbabu et al. 2019; Krishnaswamy et al. 2020; Muthukrishnan et al. 2020; Gheena and Ezhilarasan 2019; Vignesh et al. 2019; Ke et al. 2019; Vijayakumar Jain et al. 2019; Jose, Ajitha, and Subbaiyan 2020). Now the growing trend in this area motivated us to pursue this project.

The methods which are used before have less accuracy rate, less reliable and not much effective in prediction of fake political news. I have experience in research of fake political news detection. The main aim of the study is to perform higher classification of fake political news by

implementing fake news detectors using machine learning classifiers like Decision Tree algorithm and Naive Bayes algorithm and comparing their performance. The significance is less than 0.05 shows that our hypothesis holds true. While performing an independent sample t test, if the significance is 0.022 then our assumptions are true.

2. Materials and Methods

The paper study was done at CISCO Lab, Saveetha School of Engineering, SIMATS. The two supervised learning algorithms as two groups, Decision Tree and Naive Bayes. By performing two iterations on each group, one for fake news detection and other for true news detection using these two algorithms. Through the programming experiment that have performed N=10 iterations on each algorithm with sample size=10 to identify various scales of fake news and true news classification (Ahmad et al. 2020). The G-power test is about 80%. Alpha error rate is a type-I error considered as 0.05 which gives the difference between two algorithms considered. Enrollment ratio is about 1. The significance rate is 0.022 shows that our hypothesis holds true.

Dataset Description

The dataset which that used in this paper is "fake news and real news dataset". The Dataset was collected from the open source Kaggle platform (https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset). This dataset consists of data related to the US elections held in 2016. The dataset contains two files "true.csv" and "fake.csv". Both the files contain four major and relevant attributes named as "Title", "Text", "Subject" and "Date". Considering only the text attribute as dependent for analysis and classification.

Naive Bayes Algorithm

Naive Bayes algorithm is a classifier that works based on Bayes theorem with an overview of independence among predictors. This model is easy to build and also used for large datasets. The only drawback of Naive Bayes algorithm is to assume all factors as dependent on each variable. It is mainly based on the theorem formulated by Bayes's:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

(1)

Where,

A, B = events

P(A|B) = probability of A given B is true

P(B|A) = probability of B given A is true

P(A), P(B) = the independent probabilities of A and B

Pseudocode: Naive Bayes Algorithm

Input: Training Set

Output: Classifiers trained accuracy

1. Read the trained dataset into the classifier.
2. Calculate the mean and standard deviation for predictions.
3. Repeat

    Calculate gauss density for each iteration

    Until probability of fake political news texts are calculated

4. Define class

    def MultinomialNB()

    if(condition satisfy)

    return accuracy

    else

    return to previous step

    end

5. Predicted Accuracy

Decision Tree Algorithm

Decision tree is a supervised learning algorithm. It is used for both regression and classification. The goal of using a Decision Tree is to create a training model that can be used to predict the class or value of the target variable by learning simple decision rules developed from training data. Equations which are required to perform classification using decision tree:

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2 \tag{2}$$

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i \tag{3}$$

$$Information\ Gain\ =\ Entropy(before) - \sum_{j=1}^{K} Entropy(j,\ after) \tag{4}$$

Where,

Gini: Gini Index

E(S): Entropy

Pseudocode: Decision Tree Algorithm

Input: Trained dataset

Output: Classifier trained accuracy

1. Read the training dataset into the classifier

2. Define a class dtree

Class dtree

3. Get all the required stuff from previous inputs

4. Define another class to test the attribute

def evaluate(test attribute)

if(end iteration is leaf)

return accuracy

else

return children[test attribute].evaluate(test attribute)

end

5. Classifiers predicted accuracy.

Experiment Setup

The platform used to evaluate the machine learning algorithms was jupyter lab. The hardware configurations were intel core i5 processor with a RAM size of 4GB was used. The system type used

was 64-bit, OS, X64 based processor with HDD of 917 GB. The operating system used was Windows and the tool used was jupyter lab with python programming language.

The dataset is fake and real news is collected. Data preprocessing has to be done. Data cleaning like removing the unnecessary attributes from the dataset and concatenating and shuffling also need to be done. Data exploration shows the contents present in the dataset. Convert the dataset that it contains only the data needed for the classifier. Split the dataset into a training set and testing set. Now implement the machine learning classifier and use the training dataset to train the classifier. After training the classifier uses a testing dataset to test the trained classifier to get the predicted accuracy from the classifier.

The SPSS tool is used to perform the statistical calculations for the results that are obtained from classifiers for various test sizes. The text part in the training dataset is independent variable whereas the text part in testing dataset are dependent on training dataset. The comparison of the performances of Decision Tree algorithm and Naive Bayes algorithm is done.

3. Results

Accuracy Table (DST, NBA), the accuracy of the Decision Tree algorithm is approximately 99% and Naive Bayes algorithm is approximately 95%. The accuracy varies for different test sizes in decimals. The accuracy varies due to random change in the test size of the algorithm (from Table.1).

Table 1 - Accuracy Table (DST, NBA), the Accuracy of the Decision Tree Algorithm is Approximately (99%) and Naive Bayes Algorithm is Approximately (95%)

Test Size	0.1	0.2	0.3	0.4
Naive Bayes Algorithm	94.86	95.33	95.07	95.03
Decision Tree Algorithm	99.73	99.60	99.58	99.58

Group Statistics, the mean accuracy and standard deviation for Decision Tree algorithms is 99.6990 and 0.10577. For Naive Bayes algorithm is 95.3870 and 0.00061. In performing statistical analysis of 10 samples, Decision Tree obtained 0.10577 standard deviation with 0.33 standard error while Naive bayes obtained 0.00061 standard deviation with 0.12 standard error (from Table.2).

Table 2 - Group Statistics, The Mean Accuracy and Standard Deviation for Decision Tree Algorithms is 99.6990 and 0.10577. For Naive Bayes Algorithm is 95.3870 and 0.00061

	DST, NBA	N	Mean	Std. Deviation	Std. Mean Error
Accuracy	DST	10	99.6990	0.10577	0.3345
	NBA	10	95.3870	0.00061	0.12668

Independent Samples Test, the comparison of accuracy for fake political news classification using the Decision Tree algorithm and Naive Bayes algorithm with significance less than p<0.05 and standard error difference 0.13130. When compared with the other algorithms performance of the proposed Decision tree classifier achieved better performance than Naive Bayes classifier (from Table.3).

Table 3 - Independent Samples Test, the Comparison of Accuracy for Fake Political News Classification Using Decision Tree Algorithm and Naive Bayes Algorithm with Significance 0.022 and Standard Error Difference 0.13130.

		Levene's Test for Equality of Variances (1)	Levene's Test for Equality of Variances (2)	T-test for Equality of Means (3)	T-test for Equality of Means (4)	T-test for Equality of Means (5)
		F	Sig.	Std.Error Difference	95% Confidence lower	95% Confidence upper
Accuracy	Equal Variances assumed	6.243	0.022	0.13103	4.03672	4.58728
	Equal Variances not assumed			0.13103	4.02101	4.60299

It is known as the fake political news detector architecture. The architecture defines the steps which are performed to develop a fake political news detector. It consists of the steps as Data Pre-processing, Database, Data Extraction, Modelling Classifer, Implementation and Predicted Accuracy (from Fig.1).

Fig. 1 - Machine Learning Classifier Architecture
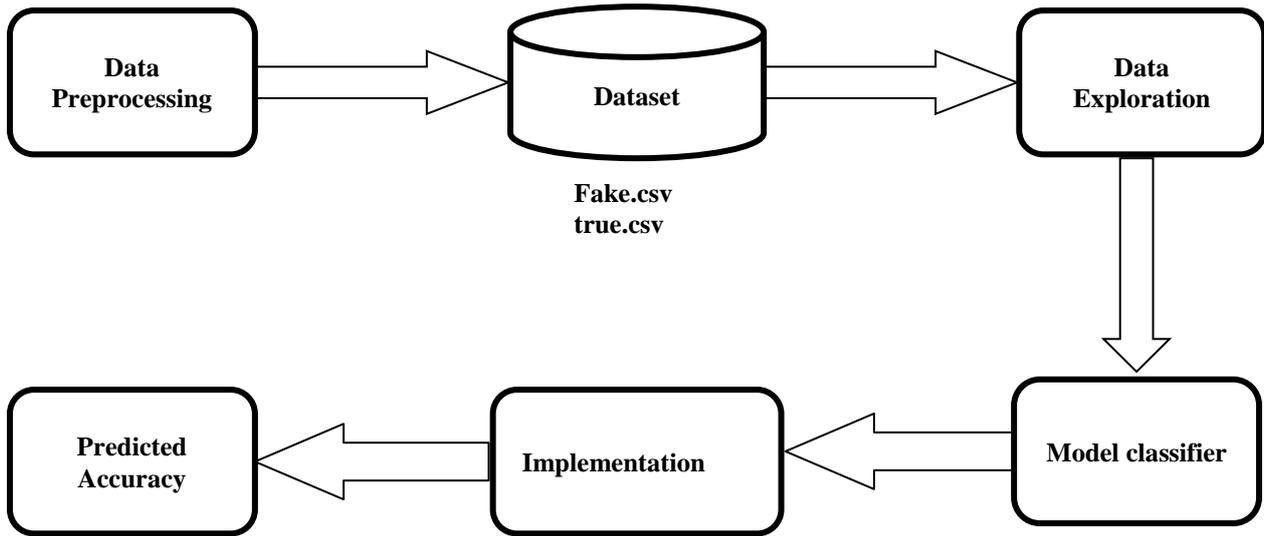
Simple Bar Mean of Accuracy by DST, NBA, the bar chart representing the comparison of mean accuracy of Decision Tree algorithm is 99.6990 and Naive Bayes algorithm is 95.3870. Decision tree algorithm with error rate of 0.3345 and Naive Bayes algorithm have error rate about 0.12668. Independent t-test was used to compare the accuracy of two algorithms and a statistically significant difference was noticed $P < 0.05$. The Decision tree model obtained 99.69% accuracy (from Fig.2). When compared with the other algorithms performance of the proposed Decision tree classifier achieved better performance than Naive Bayes classifier.

Fig.2 - Simple Bar Mean of Accuracy by DST, NBA, the bar chart representing the comparison of mean accuracy of Decision Tree algorithm is 99.6990 and Naive Bayes algorithm is 95.3870. X-Axis: Decision tree algorithm vs Naive Bayes algorithm. Y-Axis: Mean accuracy of detection ± SD.
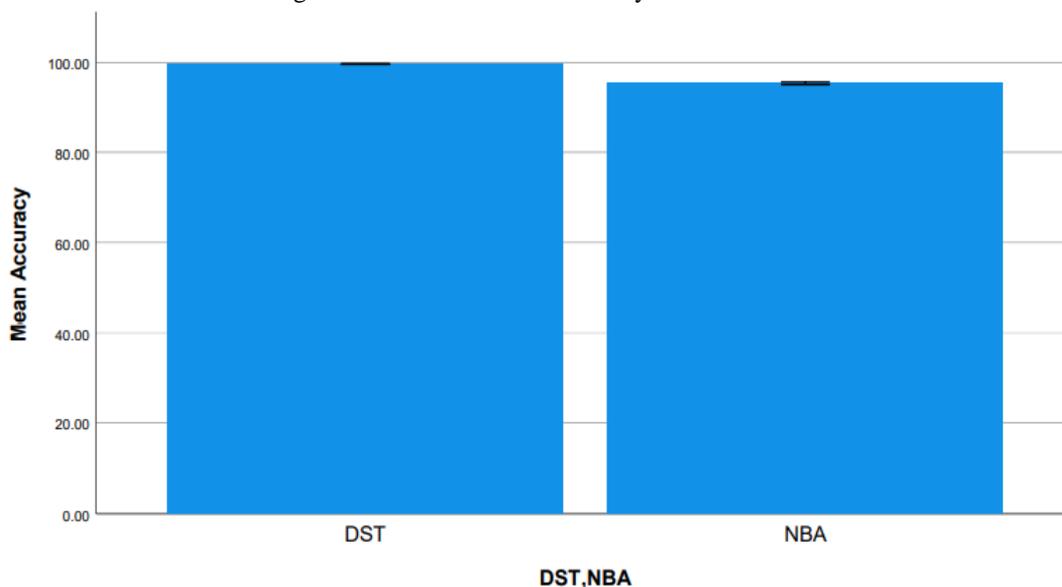
4. Discussion

Decision Tree algorithm have better accuracy rates than Naive Bayes algorithm. The results which are collected by performing multiple iterations of the experiment for identifying different scales of accuracy rate. Further, performed the statistical calculations using the SPSS tool with the results that are obtained from the experiment. Independent samples t-test is performed. In this study of fake political news classification, the Decision Tree has higher accuracy approximately (99%) in comparison to Naive Bayes algorithm approximately (95%). The significance rate is 0.022 which indicates that our hypothesis holds true.

The mean accuracy and standard deviation for the Decision Tree algorithm is 99.6990 and 0.10577. For Naive Bayes, the algorithm is 95.3870 and 0.00061. Decision tree algorithm appears to produce the most consistent results with minimal standard deviation. In paper (Agarwal et al. 2020), feed forward neural networks which are developed using deep learning used to identify the fake news on social media which gives an accuracy of 97%. In paper (Ahmad et al. 2020), Random forest and SVM(Support Vector Machine) machine learning algorithms are implemented to identify the fake news with accuracy of 91% and 96%. From these two papers we can observe that the Decision Tree algorithm proposed has better accuracy. But according to (Poddar, D., and Umadevi 2019), SVM has 91% higher accuracy than the decision tree. On the basis of literature survey it is proved that the Decision Tree algorithm has better accuracy compared with Naive Bayes algorithm.

There is a statistical significant difference in accuracy for two algorithms is p<0.05, by performing independent samples tests in the SPSS statistical tool. Mean and standard deviation are also calculated using the SPSS statistical tool. Standard error difference defines the error level, the Decision tree algorithm with error rate of 0.3345 and Naive Bayes algorithm have error rate about 0.12668. In the previous work the error rate of classifier decision trees stands best by achieving a recall score of 0.942. Followed by XGBoost classifier which achieved a recall of 0.94 (Ahmad et al. 2020).

Our institution is passionate about high quality evidence based research and has excelled in various fields ((Vijayashree Priyadharsini 2019; Ezhilarasan, Apoorva, and Ashok Vardhan 2019; Ramesh et al. 2018; Mathew et al. 2020; Sridharan et al. 2019; Pc, Marimuthu, and Devadoss 2018; Ramadurai et al. 2019). We hope this study adds to this rich legacy.

The main limitation in our experiment is that the attributes in the dataset contain very few to predict accuracy(%) for fake political news classification. The more the independent and dependent variables the more accuracy will be improved.

The future work, the dataset contains many attributes the classifier can work efficiently and can improve the prediction accuracy. Attributes like profile, source, proofs can result in improved accuracy and exact precision values.

5. Conclusion

The approach of classifying the fake political news manually requires more knowledge of the domain. In this research, the problem of classifying fake political news articles using machine learning models is discussed. The accuracy of an innovative fake news detection for political news detection using Decision Tree algorithms have better accuracy in comparison with Naive Bayes algorithms. The significance rate is 0.022 which indicates that our hypothesis holds true.

Declarations

Conflict of interests

No conflict of interest in this manuscript.

Author Contribution

Author T. Dinesh was involved in data collection, data analysis, manuscript writing. Author Dr. T. Rajendran was involved in conceptualization, data validation and critical review of manuscript.

1. InfySEC.

2. Saveetha University

3. Saveetha Institute of Medical And Technical Sciences

4. Saveetha School of Engineering

References

Agarwal, A., Mittal, M., Pathak, A., & Goyal, L.M. (2020). Fake news detection using a blend of neural networks: an application of deep learning. *SN Computer Science*, *1*(3), 1-9.

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M.O. (2020). Fake News Detection Using Machine Learning Ensemble Methods. *Complexity*, *2020*. https://doi.org/10.1155/2020/8885861

Ezhilarasan, D., Apoorva, V.S., & Ashok Vardhan, N. (2019). Syzygium cumini extract induced reactive oxygen species-mediated apoptosis in human oral squamous carcinoma cells. *Journal of Oral Pathology & Medicine*, *48*(2), 115-121.

Gheena, S., & Ezhilarasan, D. (2019). Syringic acid triggers reactive oxygen species–mediated cytotoxicity in HepG2 cells. *Human & experimental toxicology*, *38*(6), 694-702.

Giełczyk, A., Wawrzyniak, R., & Choraś, M. (2019). Evaluation of the existing tools for fake news detection. *In IFIP International Conference on Computer Information Systems and Industrial Management*, 144-151.

Ibrishimova, M.D., & Li, K.F. (2019). A machine learning approach to fake news detection using knowledge verification and natural language processing. *In International Conference on Intelligent Networking and Collaborative Systems*, 223-234.

Jose, J., & Subbaiyan, H. (2020). Different treatment modalities followed by dental practitioners for ellis class 2 fracture–A questionnaire-based survey. *The Open Dentistry Journal*, *14*(1), 59–65.

Karimi, H., Roy, P., Saba-Sadiya, S., & Tang, J. (2018). Multi-source multi-class fake news detection. *In Proceedings of the 27th international conference on computational linguistics*, 1546-1557.

Ke, Y., Al Aboody, M.S., Alturaiki, W., Alsagaby, S.A., Alfaiz, F.A., Veeraraghavan, V.P., & Mickymaray, S. (2019). Photosynthesized gold nanoparticles from Catharanthus roseus induces caspase-mediated apoptosis in cervical cancer cells (HeLa). *Artificial cells, nanomedicine, and biotechnology*, *47*(1), 1938-1946.

Krishnaswamy, H., Muthukrishnan, S., Thanikodi, S., Antony, G.A., & Venkatraman, V. (2020). Investigation of air conditioning temperature variation by modifying the structure of passenger car using computational fluid dynamics. *T hermal Science*, *24*(1 Part B), 495-498.

Malli Sureshbabu, N., Selvarasu, K., Nandakumar, M., & Selvam, D. (2019). Concentrated growth factors as an ingenious biomaterial in regeneration of bony defects after periapical surgery: A report of two cases. *Case reports in dentistry*, *2019*.

Mathew, M.G., Samuel, S.R., Soni, A.J., & Roopa, K.B. (2020). Evaluation of adhesion of Streptococcus mutans, plaque accumulation on zirconia and stainless steel crowns, and surrounding

gingival inflammation in primary molars: Randomized controlled trial. *Clinical oral investigations*, *24*(9), 3275-3280.

Mehta, M., Tewari, D., Gupta, G., Awasthi, R., Singh, H., Pandey, P., & Satija, S. (2019). Oligonucleotide therapy: an emerging focus area for drug delivery in chronic inflammatory respiratory diseases. *Chemico-biological interactions*, *308*, 206-215.

Muthukrishnan, S., Krishnaswamy, H., Thanikodi, S., Sundaresan, D., & Venkatraman, V. (2020). Support vector machine for modelling and simulation of Heat exchangers. *Thermal Science*, *24* (1 Part B), 499-503.

Paschalides, D., Christodoulou, C., Andreou, R., Pallis, G., Dikaiakos, M. D., Kornilakis, A., & Markatos, E. (2019, October). Check-It: A plugin for detecting and reducing the spread of fake news and misinformation on the web. *In IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 298-302. https://doi.org/10.1145/3350546.3352534.

PC, J., Marimuthu, T., Devadoss, P., & Kumar, S.M. (2018). Prevalence and measurement of anterior loop of the mandibular canal using CBCT: A cross sectional study. *Clinical implant dentistry and related research*, *20*(4), 531-534. https://europepmc.org/article/med/29624863.

Poddar, K., & Umadevi, K.S. (2019). Comparison of various machine learning models for accurate detection of fake news. *In Innovations in Power and Advanced Computing Technologies (i-PACT)*, 1, 1-5. https://doi.org/10.1109/i-pact44901.2019.8960044.

Ramadurai, N., Gurunathan, D., Samuel, A. V., Subramanian, E., & Rodrigues, S. J. (2019). Effectiveness of 2% Articaine as an anesthetic agent in children: randomized controlled trial. *Clinical oral investigations*, *23*(9), 3543-3550.

Ramesh, A., Varghese, S., Jayakumar, N.D., & Malaiappan, S. (2018). Comparative estimation of sulfiredoxin levels between chronic periodontitis and healthy patients–A case-control study. *Journal of periodontology*, *89*(10), 1241-1248.

Reis, J.C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, *34*(2), 76-81. https://doi.org/10.1109/mis.2019.2899143.

Samuel, M.S., Bhattacharya, J., Raj, S., Santhanam, N., Singh, H., & Singh, N.P. (2019). Efficient removal of Chromium (VI) from aqueous solution using chitosan grafted graphene oxide (CS-GO) nanocomposite. *International journal of biological macromolecules*, *121*, 285-292.

Samuel, S.R., Acharya, S., & Rao, J.C. (2020). School Interventions–based Prevention of Early-Childhood Caries among 3–5-year-old children from very low socioeconomic status: Two-year randomized trial. *Journal of public health dentistry*, *80*(1), 51-60.

Sathish, T., & Karthick, S. (2020). Wear behaviour analysis on aluminium alloy 7050 with reinforced SiC through taguchi approach. *Journal of Materials Research and Technology*, *9*(3), 3481-3487.

Sharma, P., Mehta, M., Dhanjal, D.S., Kaur, S., Gupta, G., Singh, H., & Satija, S. (2019). Emerging trends in the novel drug delivery approaches for the treatment of lung cancer. *Chemico-biological interactions*, *309*, 108720.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter, 19*(1), 22-36.

https://doi.org/10.1145/3137597.3137600.

Sridharan, G., Ramani, P., Patankar, S., & Vijayaraghavan, R. (2019). Evaluation of salivary metabolomics in oral leukoplakia and oral squamous cell carcinoma. *Journal of Oral Pathology & Medicine*, *48*(4), 299-306.

Varghese, S.S., Ramesh, A., & Veeraiyan, D.N. (2019). Blended Module-Based Teaching in Biostatistics and Research Methodology: A Retrospective Study with Postgraduate Dental Students. *Journal of dental education*, *83*(4), 445-450.

Venu, H., Raju, V.D., & Subramani, L. (2019). Combined effect of influence of nano additives, combustion chamber geometry and injection timing in a DI diesel engine fuelled with ternary (diesel-biodiesel-ethanol) blends. *Energy*, *174*, 386-406.

Venu, H., Subramani, L., & Raju, V.D. (2019). Emission reduction in a DI diesel engine using exhaust gas recirculation (EGR) of palm biodiesel blended with TiO2 nano additives. *Renewable Energy*, *140*, 245-263.

Vignesh, R., Ditto Sharmin, C., Annamalai, S., & Baghkomeh, P.N. (2019). Management of complicated crown-root fracture by extra-oral fragment reattachment and intentional reimplantation with 2 years review. *Contemporary clinical dentistry*, *10*(2), 397-401.

Jain, S.V., Muthusekhar, M.R., Baig, M.F., Senthilnathan, P., Loganathan, S., Wahab, P.A., & Vohra, Y. (2019). Evaluation of three-dimensional changes in pharyngeal airway following isolated lefort one osteotomy for the correction of vertical maxillary excess: a prospective study. *Journal of maxillofacial and oral surgery*, *18*(1), 139-146.

Vijayashree Priyadharsini, J. (2019). In silico validation of the non-antibiotic drugs acetaminophen and ibuprofen as antibacterial agents against red complex pathogens. *Journal of periodontology*, *90*(12), 1441-1448.