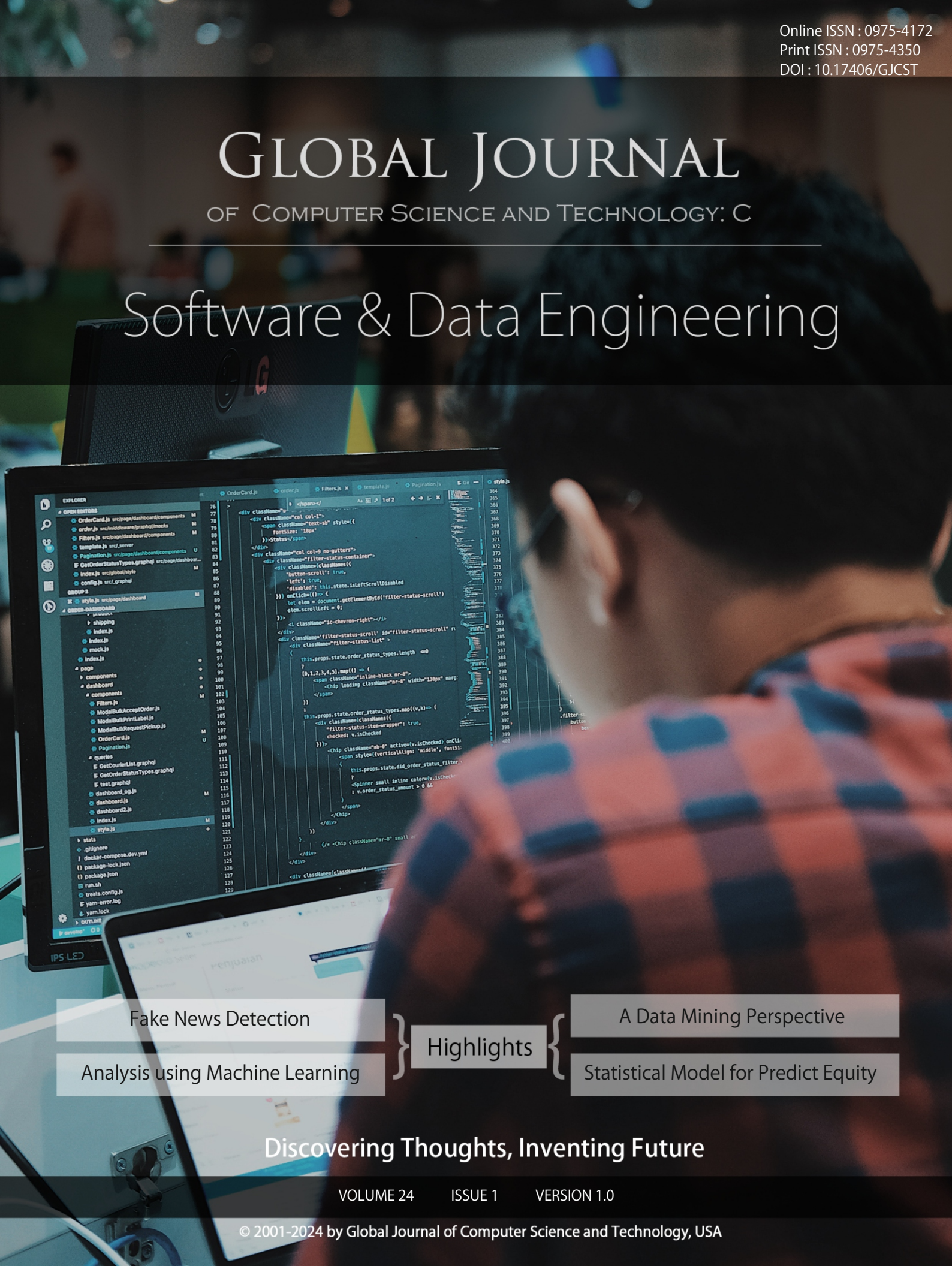
Online ISSN : 0975-4172
Print ISSN : 0975-4350
DOI : 10.17406/GJCST

Global Journal

of Computer Science and Technology: C

Software & Data Engineering



Fake News Detection
Analysis using Machine Learning

Highlights

A Data Mining Perspective
Statistical Model for Predict Equity

Discovering Thoughts, Inventing Future

VOLUME 24 ISSUE 1 VERSION 1.0

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Volume 24 Issue 1 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals® Headquarters
945th Concord Streets,
Framingham Massachusetts Pin: 01701,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals Pvt Ltd
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Excluding Air Parcel Charges):

Yearly Subscription (Personal & Institutional)
250 USD (B/W) & 350 USD (Color)

Editorial Board

Global Journal of Computer Science and Technology

Dr. Stefano Berretti

Ph.D. in Computer Engineering and Telecommunications, University of Firenze Professor Department of Information Engineering, University of Firenze, Italy

Dr. Abdurrahman Arslanyilmaz

Computer Science & Information Systems Department Youngstown State University Ph.D., Texas A&M University University of Missouri, Columbia Gazi University, Turkey

Dr. Lamri Sayad

Ph.d in Computer science, University of BEJAIA, Algeria

Dr. Tauqeer Ahmad Usmani

Ph.D in Computer Science, Oman

Dr. Hazra Imran

Ph.D in Computer Science (Information Retrieval), Athabasca University, Canada

Dr. Magdy Shayboub Ali

Ph.D in Computer Sciences, MSc in Computer Sciences and Engineering, BSc in Electronic Engineering, Suez Canal University, Egypt

Dr. Nurul Akmar Binti Emran

Ph.D in Computer Science, MSc in Computer Science, Universiti Teknikal Malaysia Melaka, Malaysia

Dr. Asim Sinan Yuksel

Ph.D in Computer Engineering, M.Sc., B.Eng., Suleyman Demirel University, Turkey

Dr. Anis Bey

Dept. of Computer Science, Badji Mokhtar-Annaba University, Annaba, Algeria

Alessandra Lumini

Associate Researcher Department of Computer Science and Engineering University of Bologna Italy

Dr. Rajesh Kumar Rolen

Ph.D in Computer Science, MCA & BCA - IGNOU, MCTS & MCP - MIcrosoft, SCJP - Sun Microsystems, Singhania University, India

Dr. Rajneesh Kumar Gujral

Ph.D in Computer Science and Engineering, M.TECH in Information Technology, B. E. in Computer Science and Engineering, CCNA Certified Network Instructor, Diploma Course in Computer Servicing and Maintenance (DCS), Maharishi Markandeshwar University Mullana, India

Dr. Aziz M. Barbar

Ph.D. IEEE Senior Member Chairperson, Department of Computer Science AUST - American University of Science & Technology Alfred Naccash Avenue Ashrafieh, Lebanon

Dr. Federico Tramarin

Ph.D., Computer Engineering and Networks Group, Institute of Electronics, Italy Department of Information Engineering of the University of Padova, Italy

Dr. Chutisant Kerdvibulvech

Dept. of Inf. & Commun. Technol., Rangsit University Pathum Thani, Thailand Chulalongkorn University Ph.D. Thailand Keio University, Tokyo, Japan

Dr. Roheet Bhatnagar

Ph.D in Computer Science, B.Tech in Computer Science, M.Tech in Remote Sensing, Sikkim Manipal University, India

Contents of the Issue

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 24 Issue 1 Version 1.0 Year 2024
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Innovative Approaches to Fake News Detection: A Data Mining Perspective

By Mustapha Ismail, Yayale Isihaka Muhammad
& Abdullahi Modibbo Abdullahi

Gombe State University

Abstract- Fake news becomes a major concern in the era of social media, as it can spread rapidly and has significant impacts on individuals and society. Society and individuals are negatively influenced both politically and socially by the widespread increase of fake news either generated by humans or machines. In the era of social networks such as Facebook, X (twitter) and WhatsApp, the quick rotation of fake news makes it challenging to evaluate its reliability promptly. Therefore, automated fake news detection tools have become a crucial requirement. To address the aforementioned issues, two data mining classification techniques were used as Extreme Gradient Boosting and Decision Tree with some python features. This study is designed to use Decision Tree and Extreme Gradient Boosting methods to develop an effective approach for detecting and classifying news as real or fake to obtain a reliable model performance. These models are trained on a labeled dataset consisting of both real and fake news.

Index Terms: *fake news, detection, data mining, social media, classification.*

GJCST-C Classification: *DDC Code: 006.312, 302.23, 384.3*

Strictly as per the compliance and regulations of:

Innovative Approaches to Fake News Detection: A Data Mining Perspective

Mustapha Ismail [α], Yayale Isihaka Muhammad [σ] & Abdullahi Modibbo Abdullahi [ρ]

Abstract- Fake news becomes a major concern in the era of social media, as it can spread rapidly and has significant impacts on individuals and society. Society and individuals are negatively influenced both politically and so cially by the widespread increase of fake news either generated by humans or machines. In the era of social networks such as Facebook, X (twitter) and WhatsApp, the quick rotation of fake news makes it challenging to evaluate its reliability promptly. Therefore, automated fake news detection tools have become a crucial requirement. To address the aforementioned issues, two data mining classification techniques were used as Extreme Gradient Boosting and Decision Tree with some python features. This study is designed to use Decision Tree and Extreme Gradient Boosting methods to develop an effective approach for detecting and classifying news as real or fake to obtain a reliable model performance. These models are trained on a labeled dataset consisting of both real and fake news. The performance of the models was evaluated using standard evaluation metrics such as accuracy, precision, recall, and F1-score. The proposed approach achieved 100% accuracy in distinguishing between real and fake news. It revealed and highlighted the potential of utilizing data mining techniques to combat the spread of fake news and provide valuable insights for researchers and practitioners in the field of information confirmation/verification and media literacy. We hope to use a different dataset to test the proposed model.

Index Terms: fake news, detection, data mining, social media, classification.

I. Introduction

Have you heard or feel dilemma after confirming a particular news is false? False information is not new, however it has become a hot topic. Traditionally, we got our news from trusted sources, journalists and media outlets that are required to follow strict codes of practice. However, the internet has enabled a whole new way to publish, share and consume information and news with very little regulation or editorial standards. Many people now get news from social media sites and networks and often it can be difficult to tell whether stories are credible or not. Information overload and a general lack of understanding about how the internet works by people has also contributed to an increase in fake news or hoax stories (Yates, 2017). Also, (Tandoc, 2019) found that only about 1% of examined Twitter account inspired 80% volume of sources of fake news.

Fake news refers to falsified news or propaganda disseminated through traditional media platforms such as print and television, as well as non-traditional media platforms such as social media [31]. The primary objective of disseminating such information is to deceive readers, harm a company's reputation, or profit from sensationalism (clickbait). It is widely regarded as one of the most serious threats to democracy, free speech, and social order [32]. Fake news is rapidly being disseminated through social media platforms such as twitter, Instagram, and Facebook, according to [33]. These platforms provide an avenue for the public to express their thoughts and opinions in an unfiltered and uncensored manner. Compared to conventional method views from media publishers' platforms, some news pieces hosted or shared on social media sites receive more views. According to researchers [32] who researched the speed with which fake news spreads on Twitter, tweets containing misleading information reach individuals six times faster than factual tweets. A few facts about fake news in the United States are as follows: 62% of Americans get their news from social media [34] making Fake news had a higher Facebook share than legitimate news [35].

Fake news detection is a subtask of text classification [38] and is often defined as the task of classifying news as real or fake. The term 'fake news' refers to the false or misleading information that appears as real news. It aims to deceive or mislead people. Fake news comes in many forms, such as clickbait (misleading headlines), disinformation (with malicious intention to mislead the public), misinformation (false information regardless of the motive behind), hoax, parody, satire, rumor, deceptive news, and other forms as discussed by [39].

Nowadays, information is easily accessible online, from articles by reliable news agencies to reports from independent reporters to extreme views published by unknown individuals. Moreover, social media platforms are becoming increasingly important in everyday life, where users can obtain the latest news and updates, share links to any information they want to spread, and post their own opinions. Such information may create difficulties for information consumers as they try to distinguish fake news from genuine news. The

Author α σ: Department of Computer Science, Gombe State University, Gombe, Nigeria.
Author ρ: Department of Computer and Information Science, Towson University, Towson Maryland, USA.
e-mail: aabdull2@students.towson.edu

wide spread of fake news on online social media has influenced public trust Knight Foundation, (2018), Naeem and Bhatti, (2020), etc. Under such severe circumstances, automatically detecting fake news has been an important countermeasure in practice. Based on a systematic review of recent literature published over the last five years, we synthesized different views dealing with fake news. We investigated machine learning (ML) applications to detect fake news, focusing on the characteristics of the different approaches and techniques, conceptual models for detecting fake news and the role of cognitive agents in this context as they have gained great popularity in the last few years. Data mining refers to extracting useful insights from large datasets, feature extraction is a technique that reduces raw data through extraction of most pertinent information [38], while ML algorithms are algorithms employed to learn from data and generalize to unobserved data [7].

This study was proposed to use a Decision Tree (DT) and Extreme Gradient Boost (XGBoost) machine learning algorithm to develop an accurate and reliable detection model that gives correct detection that is better than the one found in the published literature. Among other things the study seeks to address and show that a boosting algorithm outperforms other detection or prediction classifiers; design a framework for classification of news as fake or real using Boosting algorithms; describe what is fake news and how an individual can take preventive measures to avoid being contracted or being a victim; evaluate detection performances using evaluation metrics and compare with other results found in the published articles. The rest of the paper is organized as follows: Section II presents the existing relevant literature review; in section III the methodology employed was stated. Section IV provides the results obtained and discussion and finally, conclusion was drawn in section V.

II. Literature Review

The term fake news has recently become widespread. Even though there is no generally accepted definition of fake news, it continues to evolve day-by-day daily. Traditional fake news is generally defined as intentional behavior that harm a person or group, which could make it difficult for the victim to defend himself or herself. From the traditional definition of fake news, fake news could be explained as the use of information technology platforms especially social media to communicate wrong information about an individual or group, either intentional or otherwise. The use of news environment perception (NEP) to observe news environments for fake news detection on social media, designed popularity- and novelty-oriented perception modules to assist fake news detectors was proposed by (Sheng et al, 2022). Experiments on offline and online data show the effectiveness of NEP in boosting the performance of existing models and drew insights on how NEP helps to interpret the contribution of macro and microenvironment in fake news detection [1]. T. SU (2022), proposed the use of a User Network Embedding Structure (UNES) model, which performs fake news classification on Twitter through the use of graph embedding to represent Twitter users' social network structure, Compared to the existing approach of using user networks with handcrafted features, UNES does not require any pre-annotated data (e.g., user type (individual users or publishers), users' stance, and if they have engaged with fake news before) and observed that using the user network embedding trained on a combined user network of two datasets is on par with or outperforms the user network embedding trained for the single experimental dataset on the MM-COVID and the SD datasets, respectively, which indicates the robustness of our proposed framework, FNDF. Thus, we showed that the three task models are all important components of our end-to-end fake news detection framework, and that the FNDF is robust when applied to news involving unseen users, if the user friendship network embedding is updated with the unseen users and their friends. Combining embedded entities with the language model results in as much as 177.6% increase in MAP on ranking check-worthy tweets, and a 92.9% increase in ranking check-worthy sentences [2]. In their study, Ali et al (2022), proposed and investigated several cutting-edge fake news detecting systems and associated problems. Methods for detecting and identifying false news, such as credibility-based, temporal-based, social context based, and content-based, were also thoroughly examined. Finally, the research investigates several datasets used to identify false news and proposed an algorithm [3]. Ahmad et al (2020), used ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. Ensemble techniques along with Linguistic Inquiry and Word Count (LIWC) feature set used in this research are the novelty of the proposed approach. There are numerous reputed websites that post legitimate news content, and a few other websites such as Politi-Fact and Snopes which are used for fact checking. In addition, there are open repositories which are maintained by researchers, accuracies of the techniques are: the accuracy achieved by each algorithm on the four considered datasets. It is evident that the maximum accuracy achieved on DS1 (ISOT Fake News Dataset) is 99%, achieved by random forest algorithm and Perez-LSVM. Linear SVM, multilayer perception, bagging classifiers, and boosting classifiers achieved an accuracy of 98%. The average accuracy attained by ensemble learners is 97.67% on DS1, whereas the corresponding average for individual learners is 95.25%. The absolute difference between

individual learners and ensemble learners is 2.42% which is not significant. Benchmark algorithms Wang-CNN and Wang-Bi-LSTM performed poorer than all other algorithms. On DS2, bagging classifier (decision trees) and XGBoost are the best performing algorithms, achieving an accuracy of 94%. Interestingly, linear SVM, random forest, and Perez-LSVM performed poorly on DS2. Individual learners reported an accuracy of 47.75%, whereas ensemble learners' accuracy is 81.5%. A similar trend is observed for DS3, where individual learners' accuracy is 80% whereas ensemble learners' accuracy is 93.5%. However, unlike DS2, the best performing algorithm on DS3 is Perez-LSVM which achieved an accuracy of 96%. On DS4 (DS1, DS2, and DS3 combined), the best performing algorithm is random forest (91% accuracy). On average, individual learners achieved an accuracy of 85%, whereas ensemble learners achieved an accuracy of 88.16 %.) Worst performing algorithm is Wang-Bi-LSTM which achieved an accuracy of 62% [4].

Althabiti et al (2022) examined an English dataset labelled as whether a particular article is 'true', 'false', 'partially false' and 'other', investigated four ML algorithms and pre-trained transformers to solve this multi-classification problem and attempted to use an external dataset from Kaggle to help improve the model. However, the additional dataset did not increase the performance, even though we used a different number of samples in each attempt. Finally, their findings from over 30 experiments show that the BERT model outperforms other models. The obtained testing results on the leader board indicate that we got an F1 of around 0.305, which slightly differs from the highest participant's score with only about 0.03. Future work recommended finding an additional dataset with a similar format may help improve the model. Also, using an ensemble method, which considers both rule-based and deep learning methods, could significantly enhance the proposed system [5]. The study by (Johnson1 et al, 2021), used random forest and decision tree algorithms on a dataset containing both fake and real news to do classification. The software used for the experiment was WEKA and the result generated showed that random forest correctly classified instance is 100% and incorrectly classified instance is 0% while the decision tree correctly classified instance is 93.6364% and incorrectly classified instance is 6.3636%. The results are a proof that random forest algorithm is a better classification tool as compared to decision tree. The results obtained show that Random Forest is a better classification tool with correctly classified instance of 100% and incorrectly classified instance of 0% as compared to the decision tree with correctly classified instance of 93.6364% and incorrectly classified instance of 6.3636%. It is recommended that future studies be carried out in the area of fake news prevention so that fake news after being detected can be blocked from gaining access into the society. They used a classification report and confusion matrix to assess their model during the validation phase [6]. The work by (Sharma et al, 2020) aimed to perform binary classification of various news articles available online with the help of concepts pertaining to Artificial Intelligence, Natural Language Processing and ML. They also aimed to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news, various NLP and ML Techniques have to be used. The model is trained using an appropriate dataset and performance evaluation is also done using various performance measures. The best model, i.e. the model with highest accuracy is used to classify the news headlines or articles. As evident above for static search, our best model came out to be Logistic Regression with an accuracy of 65%. Hence they used grid search parameter optimization to increase the performance of logistic regression which then gave us the accuracy of 75%. As a result, they can say that if a user feed a particular news article or its headline in our model, there are 75% chances that it will be classified to its true nature. The user can check the news article or keywords online; he can also check the authenticity of the website. The accuracy for dynamic system is 93% and it increases with every iteration. We intent to build our own dataset which will be kept up to date according to the latest news [7]. E. K. Qalaja et al (2022), the authors employed supervised ML techniques on our newly developed dataset. Specifically, the proposed system categorizes fake news related to COVID-19 extracted from the Twitter platform using four ML-based models, including decision tree, Naïve Bayes (NB), artificial neural network (ANN), and k-nearest neighbors (KNN) classifiers), our experimental evaluation reported that DT based detection model had achieved the highest detection performance scoring 99.0%, 96.0%, 98.0%, and 90.0% in ACC, FSC, AUC, and MCC, respectively. The second set of experiments employs the small dataset (i.e., 700 tweets); their experimental evaluation reported that DT based detection model had achieved the highest detection performance scoring 89.5%, 89.5%, 93.0%, and 80.0% in accuracy, f1-score, area under the curve, and MCC, respectively. The results obtained for all experiments have been generated for the best-selected features [8].

Shu et al (2017), proposed the use of TriFN to detect fake news on social media where focused on using news contents and social contexts. For news content-based approaches, features are extracted as linguistic-based and visual-based. Linguistic-based features aim to capture specific writing styles and sensational headlines that commonly occur in fake news content Potthast et al. (2017), Afroz, Brennan, and Greenstadt (2018). For social context-based

approaches, the features include user-based, post-based and network-based. User-based features from user profiles to measure their characteristics and credibility (Castillo et al, 2011) and (Kwon et al, 2013). Finally came out with the "accuracy of 80%". [9]. In their study (Unirio et al, 2019), applied neural network using WEIBO dataset to detect fake news on social media achieving the degree of accuracy seventy five percent [10]. Orellana et al (2018), the authors proposed the use of ML, text analytics and network models – to understand the factors underlying audience attention and news dissemination on social media (e.g., effects of popularity, type of day) and also provide new tools/guidelines for journalists to better disseminate their news via these social media [11]. According to (Bondielli et al, 2019), the use of tree like network using Breath First Search (BFS) strategy to analyze and summarize the approaches for source detection of rumor and misinformation in social network and provides an intense research contribution for further exploration of source detection of rumor in a social network [12].

According to (Shelke et al, 2019), the use of Data mining, ML, Classification application with automated fact-checking applications developed to tackle the need for automation and scalability and came out with the accuracy performance of classification models 88.2%[13]. The study published by (Nyow et al, 2019), proposed the use of Artificial Intelligence, Natural Language Processing and ML to provide the user with the ability to classify the news as fake or real and also check the authenticity of the website publishing the news with multiple models trained and also some pre-trained model extracted from Felipe Adachi. The accuracy of the model is around 95% for the entire self-made model and 97% for this pre-trained model [14]. Aphiwongsophon et al (2018), explored the used of ML techniques to detect fake news by using four popular methods in the experiments: - Naïve Bayes, neural network, SVM and the normalization method for cleaning data before using the ML method to classify data. The result shows that the Naïve Bayes used to detect fake news has accuracy. Two other more advanced methods which are neural network and SVM achieved the accuracy of 99.90% [15]. Rubin et al (2016), proposed the use of satire method, satire is a type of deception that purposely incorporates cues revealing its own deceptiveness, the deception detection was quite challenging. However, the method was able to integrate word level features using an established ML approach in text classification and SVM. The style-based deception detection method reaches relatively high accuracy rates of 90%, precision of 84% and recall of 87% [16]. Ray et al, (2017), considered the use of naïve Bayes classifier to detect fake news by Naive Bayes. This method has performed as a software framework and experimented it with various records from the Face book, etc., resulting in an accuracy of 74% [17]. The paper neglected the punctuation errors, resulting in poor accuracy [18].

Gil, P (2019), the estimated various ML algorithms and made the researches on the percentage of the prediction. The accuracy of various predictive patterns included bounded decision trees, gradient enhancement, and SVM were assorted. The patterns are estimated based on an unreliable probability threshold with 85-91% accuracy [19]. Tandoc et al, (2017), utilized the Naive Bayes classifier, discussed how to implement fake news discovery to different social media sites. They used Face book, Twitter and other social media applications as a data source for news. Accuracy is very low because the information on this site is not 100% credible [20]. Sharma et al (2019), presented feedback-based approaches for fake news detection. In content-based approaches, the text of an article is regarded as the primary source of information. However, rich secondary information in the form of user responses and comments on articles and patterns of news propagation through social media can likely be more informative than article contents that are crafted to avoid detection. These secondary information sources form the basis of the works discussed [21]. Devi et al (2019), proposed the use of text processing and Naïve Bayes for training model and analyzed detection of fake news which is now prevalent in social media platforms and websites, used Therefore by using ML techniques and concluded that any news from large or small dataset can be classified as fake or not fake with previous data set values in less time which helped the user to believe in particular news that appears on social media or other sources [22].

Kesarwani et al (2020), proposed the use of a simple approach for detecting fake news on social media with the help of K-Nearest Neighbor classifier and achieved a classification accuracy of this model approximate 79% tested against Face book news posts dataset [24]. Khanam et al (2021), the authors proposed the use of six algorithms used for the detection are as: XGboost, Random Forests, Naive Bayes, K-Nearest Neighbors (KNN), Decision Tree, and SVM. The confusion matrix is automatically obtained by Python code using the cognitive learning library when running the algorithm code in Anaconda platform. Three common methods are utilized through their researches Naïve Bayes, Neural Network and SVM. Naïve Bayes has an accuracy of 96.08% for detecting fake messages. The neural network and the support vector machine (SVM) reached an accuracy of 99.90%. The scope of this paper is to cover the political news data, of a dataset known as Liar-dataset, it is a New Benchmark Dataset for Fake News Detection and labeled by fake or trust news. We have performed analysis on "Liar" dataset. The results of the analysis of the datasets using

the six algorithms have been depicted using the confusion matrix [25]. The current methods are reviewed for detection and classification of fake news using different supervised learning algorithms and a few unsupervised learning.

III. Materials and Method

Several classification algorithms can be used to classify whether given news is real or fake. But for this study, since we want to make a thorough detection, two different ML classification algorithms were chosen based on the published articles reviewed. These include eXtreme Gradient Boosting and Decision tree (DT). The eXtreme Gradient Boosting was employed as it has been optimized to increase GMB's speed and prediction performance; it is scalable and integrated into a different platform. It is faster than other algorithms due to less resource usage. It has a new tree learning algorithm to handle fewer data while Decision Tree (DT) was used because it minimizes the chance of missing crucial information or taking the wrong steps. Which could lead to unnecessary escalation. Moreover, it equips frontline agents with the necessary knowledge to handle a range of inquiries confidently, mitigating the need to involve supervisors or specialized teams.

Model built was subjected on the training data to learn from it and evaluated on the testing data. The results obtained is evaluated on performance evaluation metrics for further determination and investigation of best performing model for fake and real detection. Figure 1 Demonstrates the Framework of the Study.

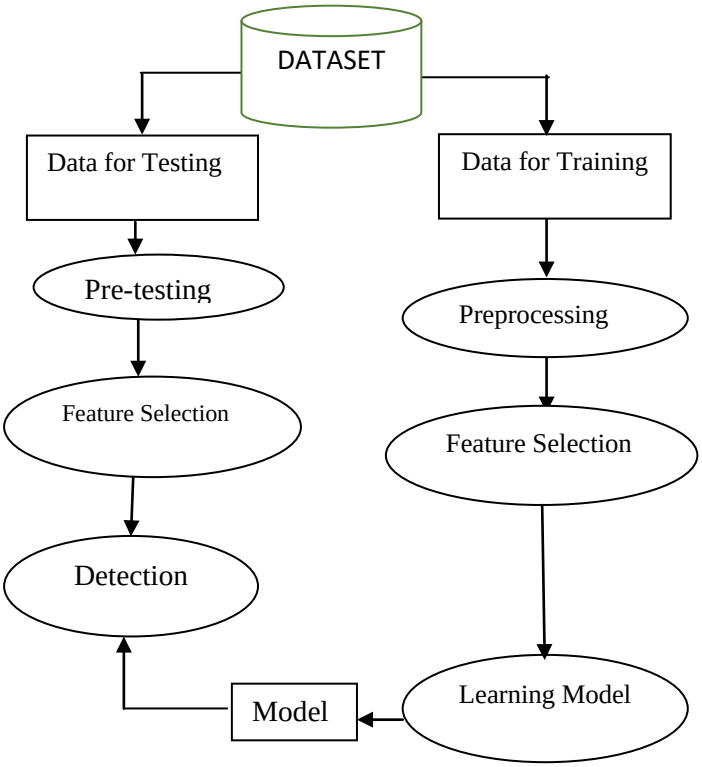


Figure 1: Framework of the Study

Experimental Setup

First of all, it was ensured that all necessary programs, tools, and techniques that would be needed to perform this experiment were downloaded and installed to obtain a good result. The python libraries used for various operations and functions are also installed. These include NumPy which is used for numerical python; pandas are used for data loading and analysis is also acquired and installed to set the environment ready. Jupyter notebook was used because it presents codes and data very well. Scikit-learn python machine learning library is used for designing, building; the system and software used for this experiment is Windows 10, Python 3.7 and colab notebook were used to run the experiment.

Data Collection

Data collection is a critical step in the research process, as the quality and accuracy of the data collected impacts the validity and reliability of the findings. It is important to ensure that the data collection process is well-designed, carefully planned, and effectively executed to minimize errors and biases. The data is Downloaded in Comma Separated Value (Csv) Format.

Data Description

Data description is important because it helps to understand the dataset and its properties, which can guide the choice of appropriate techniques and methods for analyzing the data. It also helps to identify potential issues or problems in the data, such as outliers, missing data, or measurement errors.

Data Preparation

After importing the necessary supporting programs, files, and dataset for the research work. Data preparation is paramount which is used to set the data ready to go for a machine learning project. The data collected was loaded into the google colab notebook using a panda's command read data as follows.

```
#Import the data
from google.colab import drive
drive. mount('/mntDrive')
```

It is common knowledge that most data collected or downloaded must be prepared or preprocessed to make it fit the proposed model to obtain an accurate and even dependable result. Therefore, the data obtained has undergone data preprocessing, feature extraction and feature engineering as briefly explained.

The goal of data preprocessing is to improve the quality of the data, remove any inconsistencies or errors, and make it easier to work with. Data preprocessing is a crucial step in the data analysis process, as it has a significant impact on the accuracy and validity of the results. Properly preprocessed data helps to improve the performance of machine learning models and other analytical tools, leading to better insights and decisions preprocessing: In any Machine Learning process, data preprocessing is the step in which the data gets transformed, or encoded, to bring it to such a state that now the machine can easily parse it. In other words, the features of the data can now be easily interpreted by the algorithm. In this fake news detection, preprocessing is the major thing that should be done. Firstly, as the dataset is collected from kaggle.com. Therefore, unnecessary pieces of information were removed, converted to lower case, removed punctuation, symbols and stop words and so on.

a) Dataset and Data Preprocessing

Dataset was collected from a popular ML repository called kaggle with 44919 rows and 6 columns. It is the one of the largest community of data scientist in the world. Pre-processing refers to the transformations that were applied to data before feeding it to the ML algorithm. Data Preprocessing is a technique that is used to convert the raw data into a clean data set. Some of the important data preprocessing techniques used in the study was data cleaning, dimensionality reduction, and feature engineering.

This is an essential phase that is used to enhance the quality of data to promote the extraction of meaningful insights. To also ensure that too much noise is minimized in the dataset to avoid over-fitting or under-fitting the proposed designed model. Improve the computational efficiency and accuracy of the model performance. To prepare the dataset for appropriate prediction, overfitting was avoided by training the model with sufficient number of rows and columns while under fitting was avoided by amping up model complexity, down regularization and data collection.

b) Feature Extraction

The feature extraction techniques was used in this research to reduce dataset over fitting the prediction model, improve prediction accuracy and reduce model training time. The dataset features that would be used to train the ML models have a great influence on the performance of the algorithm. Irrelevant, inappropriate or partially relevant features can undesirably influence model performance. Having unrelated features in the data can decrease the accuracy of the models, especially linear algorithms like linear and logistic regression. This feature extraction step is a process of dimensionality reduction process by which an initial set of data is reduced by identifying only the most relevant key features from the dataset that affects the detection machine learning model. In this study, data mining classification techniques were employed (DT and XBootst Algorithms) due to; DT is being faster than other algorithms due to less resource usage. It has a new tree learning algorithm to handle fewer data. Parallel and distributed computing accelerate learning, allowing for faster model discovery. Its prediction success is quite high while Gradient boosting decision trees are relatively easy to implement. Many include support for handling categorical features, don't require data preprocessing and streamline the process of handling missing data. Feature extraction was used to extract necessary relevant features for fake news detection and classification model.

Feature Selection

This feature is also used primarily to improve or enhance model detection performance accuracy. It is a technique of machine learning that leverages data to create new variables that are not in the training dataset. It generally produces new features for both types of machine learning projects, supervised and unsupervised learning. It is used for simplifying and increasing the speed of dataset transformation and manipulation aside from improving precision, recall, F1-score, and accuracy of model performance. Developing machine learning classifiers like extreme gradient boost and decision tree are also set in place.

Filter Method: Filter feature selection method and intrinsic techniques were employed to evaluate the relationship between each input variable and the target variable, and these scores are used as the basis to choose (filter) those input variables that were used in the model meanwhile intrinsic Algorithms that perform automatic feature selection during training (Decision Trees).

Model Application

There are a lot of already developed machine learning algorithms that has been rightly used and applied them directly for detection and classification purposes but for this study work, the extreme gradient boosting and decision tree algorithm were developed to fit the proposed model for this work. Hence, the modified extreme gradient boosting and decision tree algorithm models have been applied to the already preprocessed or prepared dataset for the detection and classification of news as fake or real.

Train-Test Split

Machine learning classification algorithms were used in training the model because the accuracy of the machine learning model mostly depends on the model training on the dataset. After the model has been developed and trained then testing is necessary to measure how accurate the detection performance of the model is.

At this stage, the dataset was divided into two sets, seventy-five and twenty-five percent; the first one for training and the former for testing, this was done to evaluate the model performance on the dataset that is not known to the model.

A Learning Model

It is a program that can find patterns or make decisions from a previously unseen dataset. For example, in natural language processing, machine learning models can parse and correctly recognize the intent behind previously unheard sentences or combinations of words.

Model

A model an informative representation of an object, example, pattern, exemplar, ideal mean someone or something set before one for guidance or imitation. Model applies to something taken or proposed as worthy of imitation.

It was observed that the actual parameters and their impact may vary depending on the specific implementation and version of the algorithms used. Additionally, the choice of Decision Trees and XGBoost in this study depends on the nature of the problem, the dataset size, and the desired balance between interpretability and predictive performance. Table 1 summarizes the comparisons of model parameters.

c) Fine Tuning

Fine-tuning typically refers to the process of taking a pre-trained ML model and training it further on a specific task or dataset to improve its performance. This is to take advantage of the knowledge and information the model has already learned from the large amount of data in the pre-training process. It allows you to transfer pre-existing knowledge to a new task where one can continue to train and adapt the model to improve its accuracy and efficiency.

The fine-tuning techniques consist of the following steps:

1. Loading the pre-trained model.
2. Freeze most if not all layers in the model to prevent them from further training.
3. Swap final layer or layers of the model with a new one that are specific to the tack.
4. Train the model on dataset using a lower learning rate than in the pre-trained phase.
5. Evaluate performance of the fine-tuned model and adjust the hyper-parameters as necessary.

Table 1: Summary and Comparison of the Model Parameters

Parameter	Decision Trees	XGBoost
Learning Algorithm	Greedy recursive partitioning	Gradient boosting
Ensemble Method	Not an ensemble method	Boosting ensemble method
Regularization	Prone to overfitting	Includes regularization (L1, L2 penalties)
Handling Missing Values	Not naturally handled	Can handle missing values natively
Feature Importance	Provides feature importance	Provides feature importance
Parallel Processing	Generally, not parallelizable	Can be parallelized
Speed	Can be slower for large datasets	Faster due to parallelization and optimization
Robustness	Sensitive to noise and outliers	More robust due to ensemble and regularization
Hyperparameter Tuning	Fewer hyperparameters to tune	More hyperparameters to tune
Memory Usage	Lower memory usage	Higher memory usage
Suitable for Large Datasets	Limited scalability	Well-suited for large datasets

d) Prediction Tools

The python programming language would be throughout this study. Sci-kit learn libraries would be employed to help experiment. Python has a huge set of libraries and extensions, which are specifically designed for prediction models. Sci-kit learn is one of the best sources for ML algorithms https://scikit-learn.org where nearly all types of ML algorithms are readily available, easy, and quick evaluation of ML algorithms is possible. Numpy and Pandas will be used to deal with the data. For debugging and its ability to present code nicely Jupyter Notebooks was used.

e) Performance Evaluation Technique

The generality of the training datasets is the major goal of building a prediction model using ML techniques. ML models should be able to perform pretty well on real data. The dataset will be divided into two categories; training data and testing data. Training data will be used to train ML classifiers whereas testing data to test ML classifiers.

f) Evaluation Metrics

These are tools that are used to measure the effectiveness of the proposed model, to determine whether the built model can accurately make the required prediction. Many of these tools are in existence but the most commonly used for future predictions are; accuracy, precision, recall, and f1-score. They are calculated as follows:

g) Accuracy, Precision, Recall, And F-Score

These evaluation metrics were used to evaluate fake news detection models in (Poddar & D, 2019), (Ahmad et al., 2020), and (Ghafari et al., 2020). They are calculated as follows:

Accuracy is simply defined as the measure of the ratio of all testing samples which is classified as correct.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

Precision means the ratio of relevant classified samples among the total retrieved samples.

$$Precision = \frac{tp}{tp + fp}$$

Recall is defined as the ratio of relevant classified samples among the total amount of relevant samples.

$$Recall = \frac{tp}{tp + fn}$$

F1-Score is the harmonic average of the precision and recall.

$$F1 - Score = 2\left(\frac{precision * recall}{precision + recall}\right)$$

Where:

TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative values.

IV. Result and Discussion

Figure 1, presents the chart showing the number of real news (1) and fake news (0) available in the dataset. The fake news data has slightly outperformed the real news data in accuracy as shown in figure 2.

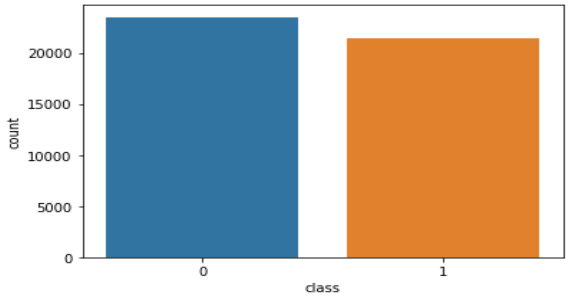


Figure 2: Count of Fake and Real News

A word cloud is a graphical representation of textual data that display the most frequently occurring words in a given text or dataset. The words are usually displayed in a randomized way as shown in figures 3 and 4, with their font size and color determined by their frequency of occurrence. Word clouds are often used to provide a quick visual summary of a large text or dataset, highlighting the most important or relevant terms. Figure 3 and 4 displayed the word cloud of real news and word cloud of fake news respectively. The word with the highest frequency on figure 3 is "said" while on figure 4 is "trump". The word "trump" appears frequently because the USA elections was trending then and made it appears much in the dataset used.

Figure 3: Word Cloud of Real News

Figure 4: Word Cloud of Fake News

Figure 5 displayed the bar chart of words frequency, it refers to the number of times a word appears in a given text or dataset. It is a measure of the importance or relevance of a word within the context of the text. The bar chart just as the word cloud indicated that the word said appeared the highest number of times in the data.

Confusion metrics are useful for evaluating the performance of classification algorithms, as they provide a more detailed understanding of the accuracy and errors of the model. From the confusion matrix, various performance metrics such as accuracy, precision, recall, and F1 score can be calculated. Figure 6 and figure 7 are confusion matrices of results obtained from decision tree classification and extreme gradient boosting classifier.

As presented in the confusion matrix figure 6, the decision tree classifier has correctly made 5861 true positive prediction and correctly made 5323 true negative prediction. It however, failed to make 14 false positive predictions and 32 false negative predictions. Hence, the decision tree algorithm performed very good prediction.

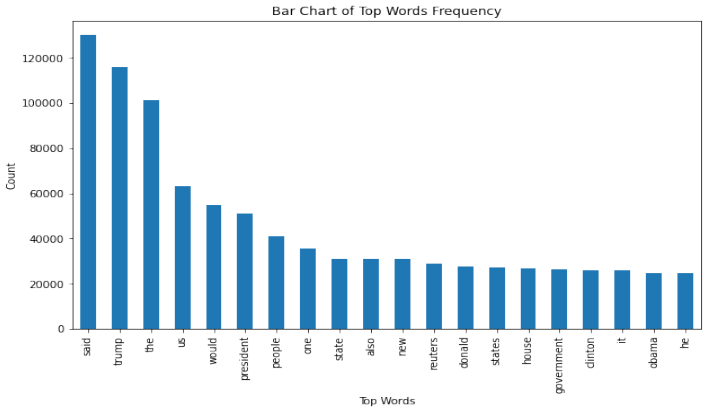


Figure 5: Bar Chart of Top Words Frequency

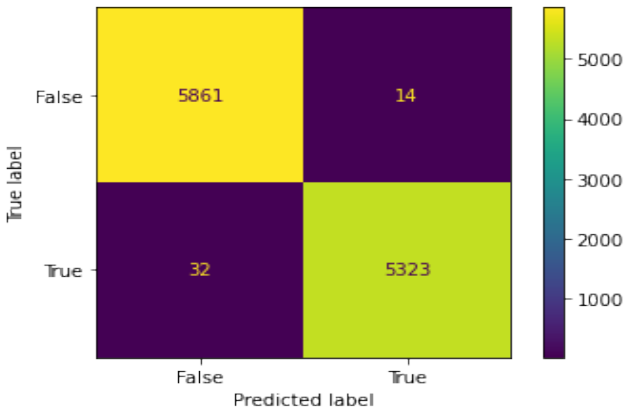


Figure 6: Confusion Matrix of Decision Tree Model

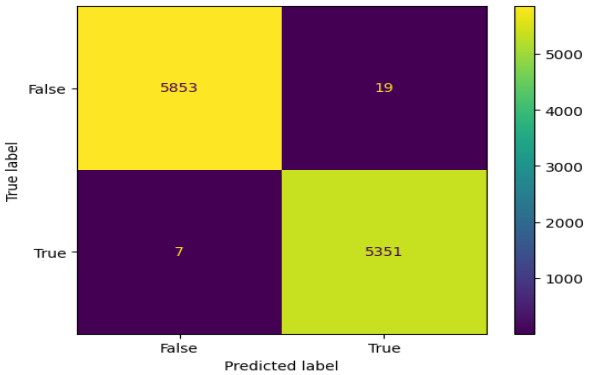


Figure 7: Confusion Matrix of Xgboost Model

As presented in the confusion matrix in figure 6 above, the extreme gradient boosting model has correctly made 5853 true positive prediction and also correctly 5351 true negative prediction. It only makes 19 false positive prediction and 7 false negative prediction. Therefore, this model has perform extremely well based on the result presented.

Table 2: Evaluation Results

	Model	Accuracy	Precision	Recall	F1-Score
Proposed	XGBoost	1.00	1.00	1.00	1.00
	DT	0.99	0.99	1.00	1.00
Existing	KNN	0.89	0.90	0.88	0.89
	DT	0.89	0.90	0.85	0.88

The simplest intuitive performance metric is accuracy, which is the ratio of properly predicted observations to all observations. Figure 7 showed the comparison of the entire various ML model used in the study as also shown in Table 2. As displayed in figure 7, accuracy, figure 8, precision, figure 9, recall and figure 10, f1-score. The extreme gradient boosting model have the highest prediction performance for all evaluation metrics which is also higher than the existing result published by [43]. Therefore, for efficient prediction performance and decision-making, a precise forecast and classification of fake and real news is highly required. XGBoost is a powerful ML algorithm that had been employed in various applications and fake news detection with several contributing factors like gradient boosting, handling high dimensional data, regularization etc [29]. which possibly made it perform better that DT and KNN.

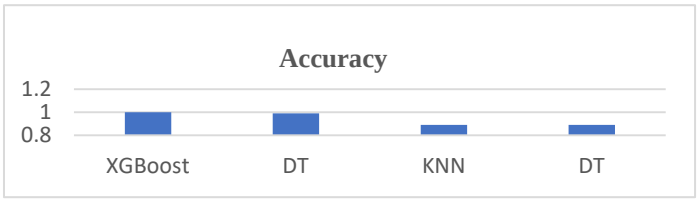


Figure 7: Accuracy of the Prediction Model

The ratio of accurately predicted positive observations to total expected positive observations is known as precision. As indicated in figure 8, the XGBoost has the highest performance followed by decision tree.

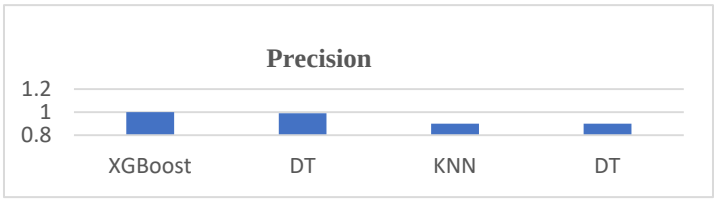


Figure 8: Precision of the Models

Figure 9 present the recall which is defined as the proportion of accurately predicted positive observations to all observations in the class. Again, the proposed model that is extreme gradient boosting model has performed better than the existing model.

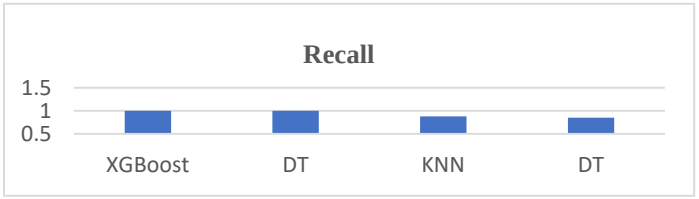


Figure 9: Recall Performance of the Models

Figure 10 shows the f1-score of the performance model. It considers both false positives and false negatives. Although it is not as intuitive as accuracy, F1-score is frequently useful than accuracy, especially if the class distribution is unequal. Our proposed model also outperformed highly incredible compared to the existing model. From the results in table 2 F1 score and a recall of 1 respectively shows an excellent performance.

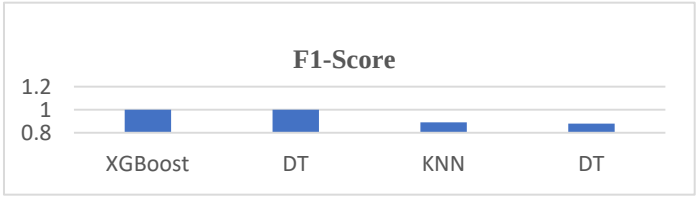


Figure 10: F1-Score Performance of the Models

The confusion matrix result presented in table 2 is a useful tool for understanding the performance of a classification model. It also allowed us to calculate several metrics that can be used to evaluate the model. Two ML models were designed; the results obtained from the proposed model outperformed the existing model found in [43]. The accuracy performance of extreme gradient boost was superb as shown in table 1 with 100% accuracy, precision, recall and f1-score. As for decision tree, it performance is also commendable, 99 percent accuracy and precision while hundred percent for recall and f1-score. From the generated result our proposed models with extreme gradient boost and decision tree algorithms have demonstrated an accurate and reliable performance that is more than what was found in the existing work.

Overall, the results suggest that the XGBoost model is a more effective and reliable model for detecting fake news, and its performance is statistically significantly better than the Decision Tree and KNN models. In general, this research work contributes to the ongoing efforts in addressing the challenges of fake news by utilizing data mining techniques to detect and classify misleading information. The findings have implications for media organizations, social media platforms, and individuals seeking reliable information in the digital age. Limitations of the work emanated from the complex aspect of the fake news detection like Data quality and availability, class imbalance, lack of context, Continuous evolution of fake news, language and cultural barriers, false positives and false negatives, limited domain knowledge and adversarial attacks.

These limitations highlight the challenges and complexities of fake news detection and the need for ongoing research and development to improve the accuracy, effectiveness, and transparency of fake news detection models.

V. Conclusion

This work developed an effective approach to identify and categorize fake news articles using data mining techniques. Through the utilization of text preprocessing, feature extraction, and ML algorithms, the proposed approach was capable of distinguishing between real and fake news articles. The performance of the models was evaluated and results demonstrated the effectiveness of data mining techniques in fake news detection and classification. The proposed approach achieves 100% accuracy and performance in distinguishing between real and fake news articles from running the proposed model and obtained results. Media organizations can benefit from incorporating data mining techniques into their fact-checking processes, improving the overall accuracy and reliability of news content. The practical implications of fake news detection involve using specific strategies and tools, such as fact-checking websites, machine learning algorithms, and social media monitoring, to improve accuracy, efficiency, and credibility, and reduce the spread of misinformation. Fake news research can contribute to restoring public trust in media by developing effective fact-checking methods, improving media literacy, and promoting transparency and accountability in journalism. Also, Social media platforms can utilize these techniques to identify and mitigate the impact of fake news on their platforms, thereby enhancing the trustworthiness of shared information. Moreover, individuals can leverage the outcomes of this study to enhance their media literacy skills and make informed judgments about the credibility of news articles they encounter with. The issue of fake news if not curtailed is causing more harm demanding an imperative action from tech industries and policy makers.

References Références Referencias

1. Q. Sheng, J. Cao, X. Zhang, R. Li, Wang, Y. Zhu (2022), "*News Environment Perception for Fake News Detection*" Key Lab of Intelligent Information Processing of Chinese Academy of Sciences, Institute of Computing Technology, Chinese Academy of Sciences University of Chinese Academy of Sciences.
2. S. Ting (2022), "*Automatic Fake News Detection on Twitter*" Submitted in Fulfillment of The Requirements for the Degree of Doctor of Philosophy School of Computing Science College of Science and Engineering University of Glasgow.
3. I. Ali, M. Nizam, P. Shivakumara, N. Binti (2022), *"Fake News Detection Techniques on Social Media"* Wireless Communications and Mobile Computing Volume 2022, Article ID 6072084, 17 page.
4. I. Ahmad, M. Yousaf, S. Yousaf1, M. Ahmad (2020), *"Fake News Detection Using Machine Learning Ensemble Methods"* Volume 2020, Article ID 8885861, 11 pages.
5. S. Althabiti a, b, A. Alsalkac, E. Atwell d (2022*), "SCUoL at CheckThat! 2022: Fake News Detection Using Transformer-Based Models"*. The fifth edition of the "Check That! Lab" is one of the 2022 Conference and Labs of the Evaluation Forum (CLEF).
6. E. Johnson1, J. Inyangetoh2, M. Esang3 (2021), *"An Experimental Comparison of Classification Tools for Fake News Detection"* International Journal of Advanced Research in Computer and Communication Engineering. Vol. 10, Issue 8, August 2021 DOI 10.17148/IJARCCE.2021.10820 ©IJARCCE This work is licensed under a Creative Commons Attribution 4.0 International License 135 ISSN (O) 2278-1021, ISSN (P) 2319-5940.

7. U. Sharma, S. Saran, M. Shankar (2020), "*Fake News Detection using Machine Learning Algorithms"* International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NTASU - 2020 Conference Proceedings.
8. K. Emad, Q. A Al-Haija*, A. Tareef3, M.M. Al-Nabhan4 (2022),*"Inclusive Study of Fake News Detection for COVID-19 with New Dataset using Supervised Learning Algorithms"*. International Journal of Advanced Computer Science and Applications, Vol. 13, No. 8, 2022.
9. K. Shu, S. Wang, H. Liu(2017) "*Exploiting Tri-Relationship for Fake News Detection*" Computer Science and Engineering, Arizona State University, Tempe, 85281, USA.
10. F. Unirio, R. Unirio, A. Unirio (2019) *"Can Machines Learn to Detect Fake News? A Survey Focused on Social Media"* Proceedings of the 52nd Hawaii International Conference on System Sciences.
11. C. Rodriguez, M. T Keane (2018)" *Attention to news and its dissemination on Twitter: A survey"* Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Ireland, Computer Science Review 29 (2018) 74–94.
12. A. Bondiellia & F, Marcelloni b (2019) *"A survey on fake news and rumour detection techniques"*. a Dipartimento di Ingegneria dell'Informazione, University of Pisa, Largo Lucio Lazzarino, 1, Pisa, Italy. b Dipartimento di Ingegneria dell'Informazione, University of Florence, Italy. m
13. S. Shelke a & V. Attar b(2019) *"Source detection of rumor in social network"* a Ph.D. Research Scholar, Department of Computer Engineering & IT, College of Engineering, Pune (COEP), 411005, Department of Computer Engineering & IT, College of Engineering, Pune (COEP), 411005, India. Journal homepage.
14. N. Nyow & H. Chua (2019) *"Detecting Fake News with Tweets' Properties*". Department of Computing and Information Systems Sunway University Selangor.
15. S. Aphiwongsophon & P. Chongstitvatana (2018) *"Detection of fake news with machine learning method*" 15th International Conference on Electrical Engineering/Electronic, Computer, Telecommunications and Information Technology.
16. V. Rubin, N. Conroy, Y. Chen & S. Cornwell (2016) *"Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News,"* Proc Second Work Comput. Approaches to Decept. Detect., pp. 7–17, 2016.
17. S. Ray (2017). *"common-machine-learning-algorithms"*/https://www.analyticsvidhya.com/blog//
18. Economic and Social Research Council. Using Social Mmedia. Available at: https://esrc.ukri.org/research/impact-toolkit/social-media/using-social-media
19. P. Gil (2019), Available at: https://www.lifewire.com/what-exactly-is-twitter-2483331. ASCI-2020 IOP Conf. Series: Materials Science and Engineering 1099 (2021) 012040 IOP Publishing doi:10.1088/1757-899X/1099/1/01204012
20. E. C. Tandoc Jr.et al (2017). *"Defining fake news a typology of scholarly definitions".* Digital Journalism.
21. K. Sharma, F. Qian, H. Jiang, N. Ruchansky (2019), *"Combating Fake News: A Survey on Identification and Mitigation Techniques "*University of Southern California MING ZHANG, Peking University YAN LIU, University of Southern California.
22. B. Devi, A. Soni, S. Kapkoti, S, Shankar (2019*) "Fake News Detection Based on Machine Learning by using TFIDF*" Department of Computer Science and Engineering SRM Institute of Science and Technology, Ramapuram, Chennai, India.
23. Z. Mahid, S. Manickam, S. Karuppayah (2018) *"Fake News on Social Media: Brief Review on Detection Techniques"* National Advanced IPV6 Centre (Universiti Sains Malaysia) Pulau Pinang, Malaysia.
24. A. Kesarwani, S. Chauhan, A. Nair (2020) "*Fake News Detection on Social Media using K-Nearest Neighbor Classifier"* School of VLSI and ESD National Institute of Technology Kurukshetra, India.
25. Z. Khanam, B N Alwasel, H. Sirafi1 and M. Rashid (2021) "*Fake News Detection Using Machine Learning Approaches*"1College of Computing and Informatics, Saudi Electronic University, Dammam, KSA2 School of Computer Science and Engineering, Lovely Professional University, Jalandhar, India.
26. https://science.sciencemag.org/content/359/6380/1094.summary Science 09 Mar 2018: Vol. 359, Issue 6380, pp. 1094-1096 DOI: 10.1126/science. Aao 2998.
27. Z. Khanam & M.N. Ahsan (2017*) "Evaluating the effectiveness of test-driven development: advantages and pitfalls* "International. J. Appl. Eng. Res. 12, 7705–7716, 2017.
28. V. Agarwala, H. Sultanaa, S. Malhotraa, A. Sarkarb (2019), *"Analysis of Classifiers for Fake News Detection"* International Conference on Recent Trends in Advanced Computing 2019, Icrtac 2019.
29. J. Chevallier, D. Guégan, S. Goutte (2021). *"Is It Possible to Forecast the Price of Bitcoin?"* 377–420.
30. K. Poddar., & D, G. B. A. (2019). *"Comparison of Various Machine Learning Models for Accurate Detection of Fake News".* 1–5.
31. A. Thota, P. Tilak, S. Ahluwalia, N. Lohia *"Fake news detection: a deep learning approach"*, SMU Data Science Review 1 (2018) 10.

32. K. Langin *((2018) "Fake news spreads faster than true news on twitter"*- thanks to people, not bots, Science magazine.
33. H. Allcott, M. Gentzkow ((2017) "*Social media and fake news"* in the 2016 election, Journal of economic perspectives 31 211–36.
34. J. Gottfried, E. Shearer (2016) *"News use across social media platforms",* http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/(2016)
35. C. Silverman, L. Alexander (2016) *"How teens in the balkans are duping trump supporters with fake news"*. buzzfeed, 14 november, 2016.
36. DataReportal – global digital insights', Data Reportal –Global Digital Insights. Online. Available: https://datareportal.com/. Reached access: 11- Dec-2022.
37. J. M. B (2018), *"Fake News: Real Lies, Affecting Real People*". North Charleston, SC: Createspace Independent Publishing Platform,
38. C. Liu, X. Wu, M. Yu, G. Li, J. Jiang., W. Huang, X. Lu (2019) *"A two-stage model based on BERT for short fake news detection"*. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 11776 LNAI, pp. 172–183 (2019). https://doi.org/10.1007/ 978-3-030-29563-9_17 2.
39. X. Zhouk, R. Zafarani (2020) *"A survey of fake news*": fundamental theories, detection methods, and opportunities. ACM Comput. Surv. https://doi.org/10.1145/3395046
40. E. K. Qalaja, Q. A. Al-Haija, A. Tareef, A. A. Al-Nabhan. (2022), *"Inclusive Study of Fake News Detection for COVID-19 with New Dataset using Supervised Learning Algorithms".* (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 13, No. 8, 2022.
41. G. Mavridis (2018) "Fake news and Social Media". How Greek users identify and curb misinformation online" Media and Communication Studies.
42. Tandoc Jr, E. C. (2019). The facts of fake news: A research review. *Sociology Compass*, *13*(9), e12724.

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 24 Issue 1 Version 1.0 Year 2024
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Market Basket Analysis using Machine Learning

By Atul Sharma, Dr. Mohammad Salim Hamidi & Yousuf Hotak

Jahan University

Abstract- Market Basket Analysis is a technique used to analyse the items which are most likely to be purchased together mostly in the retail and economic sector. This technique is especially beneficial for optimization purpose. In this research paper we have used the market basket dataset from the Kaggle repository. This database has been analysed for very well know topic Market Basket Analysis using Python language.

Keywords: *market basket analysis, machine learning, python.*

GJCST-C Classification: *DDC Code: 006.31*

Strictly as per the compliance and regulations of:

Market Basket Analysis using Machine Learning

Atul Sharma [α], Dr. Mohammad Salim Hamidi [σ] & Yousuf Hotak [ρ]

Abstract- **Market Basket Analysis is a technique used to analyse the items which are most likely to be purchased together mostly in the retail and economic sector. This technique is especially beneficial for optimization purpose. In this research paper we have used the market basket dataset from the Kaggle repository. This database has been analysed for very well know topic Market Basket Analysis using Python language.**

Keywords: market basket analysis, machine learning, python.

I. Introduction

Market Basket Analysis is a valuable tool for businesses seeking to optimize their product offerings, increase cross-selling opportunities, and improve marketing strategies. Market basket analysis can be used to enhance the profitability of any business. Machine Learning is rewarding the retail industry in a unique way. It supports the retail sector in all areas, from predicting sales success to locating customers. Market basket analysis (MBA) is one such top retail application of machine learning. It helps retailers know what products people are purchasing together so that the store/website layout can be designed in the same manner.[1]

We have followed the below mentioned process for the task of Market Basket Analysis research project

- Gather transactional data, including purchase history, shopping carts, or invoices.
- Analyse product sales and trends.
- Use algorithms like Apriori or FP-growth to discover frequent item sets and generate association rules.
- Interpret the discovered association rules to gain actionable insights.
- Develop strategies based on the insights gained from the analysis.

II. Review of Literature

(Chaudhary, S. (2022, February 11) has talked about the importance of Market Basket Analysis in his research; (Stevens, S. (2023, September 7) has talked critically about the Data Analysis implication using Machine Learning; (Simplilearn. (2022, November 22) has discussed about the key components of the Market Basket Analysis; (McColl, L. (2022, March 1) has discussed about the Market Basket Analysis using Python; (*How to use market basket analysis for retail and marketing*. (2023, December 19) talks about the analysis of Market Basket analysis for retail sector; *Overview of market basket analysis*. (n.d.) discusses about the overview related to the Market basket analysis; Predoiu, O. (2024, April 2) talks about customer behavior analysis; Elnahla, N. (2021) discusses about Retail lance and its Marketing Implications with reference to Market Basket Analysis.

III. Research Methodology

We have worked on the Quantitative research. The past (historical) research data has been downloaded from the Kaggle repository for analysis. Now this data has been analyzed very effectively using Python language. According to Dawson (2019), a research methodology is the primary principle that will guide your research. It becomes the general approach in conducting research on your topic and determines what research method you will use. A research methodology is different from a research method because research methods are the tools you use to gather your data (Dawson, 2019). You must consider several issues when it comes to selecting the most appropriate methodology for your topic. Issues might include research limitations and ethical dilemmas that might impact the quality of your research.[2]

IV. Data Analysis & Interpretation

Even with years of professional experience working with data, the term "data analysis" still sets off a panic button in my soul. And yes, when it comes to serious data analysis for your business, you'll eventually want data scientists on your side. But if you're just getting started, no panic attacks are required.[3]

Author α: Asstt. Professor. e-mail: atul.sharma@jpemgzb.ac.in
Author σ: Vice Chancellor. Jahan University, Kabul, Afghanistan.
e-mail: avc@jahan.edu.af
Author ρ: Dean Jahan University, Kabul, Afghanistan.
e-mail: dean_cs@jahan.edu.af

```
Python 3.12.0 (tags/v3.12.0:0fb18b0, Oct  2 2023, 13:03:3
Type "help", "copyright", "credits" or "license" for more
>>> import pandas as pd
>>> import plotly.express as px
>>> import plotly.io as pio
>>> import plotly.graph_objects as go
>>> pio.templates.default = "plotly_white"
>>> data = pd.read_csv("E:/market_basket_dataset.csv")
>>> print(data.head())
   BillNo  Itemname  Quantity  Price  CustomerID
0    1000    Apples         5   8.30       52299
1    1000    Butter         4   6.06       11752
2    1000      Eggs         4   2.66       16415
3    1000  Potatoes         4   8.10       22889
4    1004   Oranges         2   7.26       52255
>>> _
```

Figure 1: Importing Utilities & Reading Dataset

Figure 1 above shows us steps to import utilities in Python which would be required for our Data analysis.

```
>>> print(data.isnull().sum())
BillNo        0
Itemname      0
Quantity      0
Price         0
CustomerID    0
dtype: int64
```

Figure 2: Verification of the Consistency of Data

Figure 2 above shows that we do not have any null data in our dataset.

Further we go ahead to check for Summary Statistics of the dataset as shown below (Figure 3).

```
>>> print(data.describe())
            BillNo    Quantity       Price    CustomerID
count   500.000000  500.000000  500.000000    500.000000
mean   1247.442000    2.978000    5.617660  54229.800000
std     144.483097    1.426038    2.572919  25672.122585
min    1000.000000    1.000000    1.040000  10504.000000
25%    1120.000000    2.000000    3.570000  32823.500000
50%    1246.500000    3.000000    5.430000  53506.500000
75%    1370.000000    4.000000    7.920000  76644.250000
max    1497.000000    5.000000    9.940000  99162.000000
```

Figure 3: Statistics for the Dataset

Figure 3 above shows the Statistical results of dataset.

Now let us look at the pictorial representation Sales Distribution of the items as.

```
>>> print(rules[['antecedents', 'consequents', 'support',
  antecedents consequents   support  confidence      lift
0    (Apples)     (Bread)  0.045752    0.280000  1.862609
1     (Bread)    (Apples)  0.045752    0.304348  1.862609
2    (Apples)    (Butter)  0.026144    0.160000  0.979200
3    (Butter)    (Apples)  0.026144    0.160000  0.979200
4    (Apples)    (Cereal)  0.019608    0.120000  0.592258
5    (Cereal)    (Apples)  0.019608    0.096774  0.592258
6    (Apples)    (Cheese)  0.039216    0.240000  1.311429
7    (Cheese)    (Apples)  0.039216    0.214286  1.311429
8    (Apples)   (Chicken)  0.032680    0.200000  1.530000
9   (Chicken)    (Apples)  0.032680    0.250000  1.530000
```

Figure 4: Sales Distribution

- *Antecedents:* These are the items that are considered as the starting point or "if" part of the association rule. Here is our case we have Bread, Butter, Cheese, and Chicken as the antecedents in our analysis. The entities or "itemsets" produced from the data are called antecedents. To put it another way, it's the IF element on the left. In the situation before, bread serves as the antecedent.[4]
- *Consequent:* These are the items that tend to be purchased along with the antecedents or the "then" part of the association rule. The term "consequent" refers to an item or group of items that are encountered along with the antecedent. The THEN part of the sentence is displayed on the right-hand side. The result in the aforementioned case is butter.[5]
- *Support:* Support measures how frequently a particular combination of items (both antecedents and consequents) appears in the dataset. It refers to the proportion of transactions in which the items are expected to be bought together. For example, the first rule indicates that Bread and Apples are bought together in approximately 4.58% of all transactions. Support refers to the frequency or occurrence of a specific combination of items in the dataset. Thus indicates frequency of item set appearing in the transactions being analyzed.[6]
- *Confidence:* Confidence quantifies the likelihood of the consequent item being purchased when the antecedent item is already in the basket. Alternately it shows the probability of buying the subsequent item wherein the antecedent item is already in the basket. Figure above shows that there is a 30.43% chance of buying Apples when Bread is already kept in the basket after purchase. The probability that a transaction that contains the items on the left hand side of the rule (in our example, pencil and paper) also contains the item on the right hand side (a rubber). The higher the confidence, the greater the likelihood that the item on the right hand side will be purchased or, in other words, the greater the return rate you can expect for a given rule.[7]
- *Lift:* Lift measures the degree of association between the antecedent and consequent items, while considering the baseline purchase probability of the consequent item. If we find a lift with a value greater than 1 then this would indicate a positive association between the antecedent and the consequent item then it would indicate that the items are most likely to be bought together rather than independently. If we obtain a value which is less than 1 then it would indicate a negative association between the two. We can find a lift of 1.86 suggests a positive association between Bread and Apples. Lift is the measure of the effect of purchasing item A on purchasing item B. It is used to determine whether the combination of items has practical value. In other words, it is used to see if the combination of items is purchased more frequently than the individual items. If the value is greater than 1, it means that the combination is effective, while if it is less than 1, it means that it is ineffective.[8]

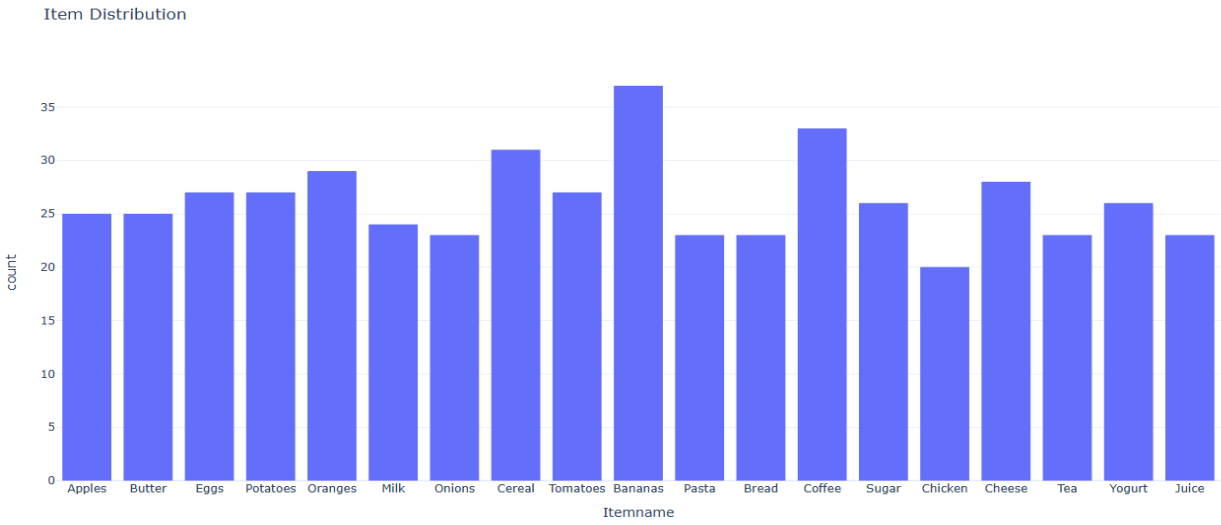


Figure 5: Item Distribution

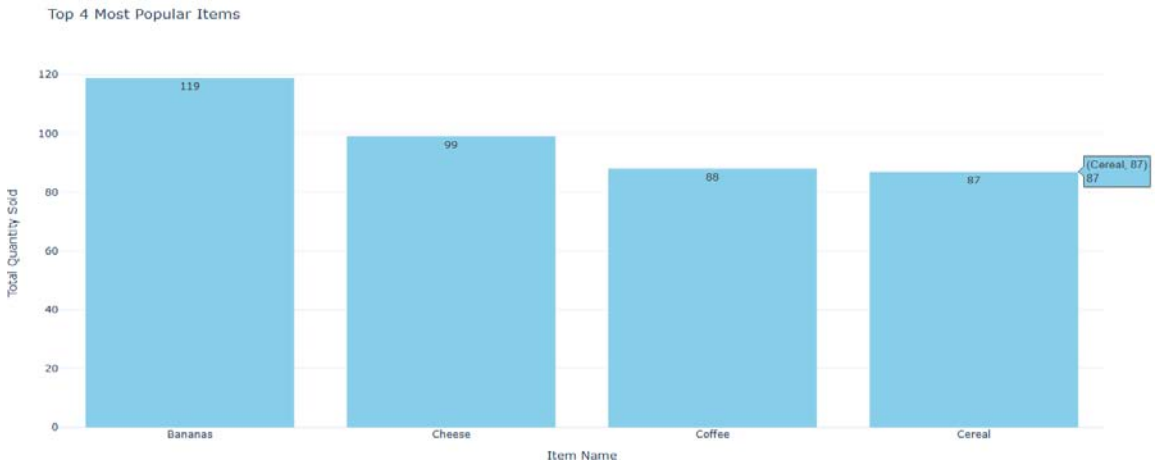


Figure 6: Top four Most Popular Items

It is observed that bananas are the most popular item sold in the store.

Understanding Customer Behavior.

Figure 7: Understanding Customer Behavior

By the term customer behavior, we understand the trends in the buying habits and factors which influence the decision to buy something else along with previous item. Here in Figure 7 above we explore customer behavior by comparing average quantity and total spending. Customer Behavior Analysis represents the study of how people make buying decisions concerning a product, service, and /or organization.[9]

V. Conclusion, Implications, and Scope For Future Research

Henceforth it may be concluded that the historic data can be analyzed very effectively using Python language which is highly flexible and simple. This data analysis would be highly beneficial to end users in terms of decision-making in the future. They can very easily plan out their investment based upon the results that have been obtained with the help of this application. It would help them to have a better decision-making which would result in generating more profits. Since Market Basket Analysis is a highly productive tool to optimize the selling opportunities hence this project becomes utmost important. In the near future we would design a model wherein the predictions can be made beforehand. Artificial intelligence has revolutionized market basket analysis by automating the process of data analysis and rule discovery.[10]

Acknowledgement

We would like to express our deepest gratitude to my adviser, Professor Mamta Bansal, for her invaluable guidance and support throughout this research. Her expertise and dedication have been a source of inspiration and motivation.

References Références Referencias

1. Chaudhary, S. (2022, February 11). Understanding market basket analysis in data mining. Hire the World's Most Deeply Vetted Developers & Teams Turing. https://www.turing.com/kb/market-basket-analysis.
2. Stevens, S. (2023, September 7). What is data analysis? Examples and how to get started. https://zapier.com/blog/data-analysisexam- ple/
3. Simplilearn. (2022, November 22). *What is market basket analysis? Overview, uses, types, and examples* Simplilearn.com. https://www.simplilearn.com/what-is-market-basket-analysis-article.
4. Simplilearn. (2022, November 22). *What is market basket analysis? Overview, uses, types, and examples*.Simplilearn.com. https://www.simplilearn.com/what-is-market-basket-analysis-article.
5. *How to use market basket analysis for retail and marketing*. (2023, December 19). Thought Spot. https://www.thoughtspot.com/data-rends/analytics/market-basket-analysis.
6. McColl, L. (2022, March 1). *Market basket analysis: Understanding customer behaviour*. Select Statistical Consultants. https://select-statistics.co.uk/blog/market-basket-analysisunderstanding-customer-behaviour/
7. *Overview of market basket analysis*. (n.d.). Modern Big Data Analytics & BI Software - Fine BI. https://intl.finebi.com/blog/market-basket-analysis

8. Predoiu, O. (2024, April 2). Customer Behavior analysis. Omniconvert Ecommerce Growth Blog. https://www.omniconvert.com/blog/customer-behavior-analysis/
9. Elnahla, N. (2021). Your Retailer Needs You: Retail lance and its Marketing Implications. https://doi.org/10.22215/etd/2021-14633

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 24 Issue 1 Version 1.0 Year 2024
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A Statistical Model for Predict Equity Price of the Apple Inc. Company in the Information Technology Sector

By Most Tajmary Mahfuz

Science. St. John's University

Abstract- The goal of this paper was to produce predictions regarding the future stock price of the massive corporation Apple Inc. The Information Technology (IT) sector comprises six distinct industries, among which Apple Inc., which is a large-cap company, has been the mainstay of my research. Apple Inc. is one of the most valuable companies in the world, with a market capitalization of over $2 trillion. In the study, a predictive model has been devised to estimate the equity price of the Apple Inc. company, employing diverse statistical techniques. A variety of statistical methods to examine 23 observations using time series data from FactSet-2022 has been used in this study. Graphical techniques including histograms, dissipate plots, and line plots were used to outwardly investigate the dataset, while elucidating insights were utilized to sum up the attributes of versatile factors like Value, EPS, BVPS, CR, DTA, EBIT, and SPS.

Keywords: statistical model, equity, information technology, descriptive statistics.

GJCST-C Classification: LCC Code: HG4515.95.A6

Strictly as per the compliance and regulations of:

A Statistical Model for Predict Equity Price of the Apple Inc. Company in the Information Technology Sector

Most Tajmary Mahfuz

Abstract- The goal of this paper was to produce predictions regarding the future stock price of the massive corporation Apple Inc. The Information Technology (IT) sector comprises six distinct industries, among which Apple Inc., which is a large-cap company, has been the mainstay of my research. Apple Inc. is one of the most valuable companies in the world, with a market capitalization of over $2 trillion. In the study, a predictive model has been devised to estimate the equity price of the Apple Inc. company, employing diverse statistical techniques. A variety of statistical methods to examine 23 observations using time series data from FactSet-2022 has been used in this study. Graphical techniques including histograms, dissipate plots, and line plots were used to outwardly investigate the dataset, while elucidating insights were utilized to sum up the attributes of versatile factors like Value, EPS, BVPS, CR, DTA, EBIT, and SPS. Relationship investigation was directed to survey the straight relationship between factors. The essential scientific strategy utilized was relapse examination, with the practical detail of the model framed in Conditions 1, 2, and 3. These conditions portray Cost as a component of book esteem per share (BVPS), income before interest and duties (EBIT), and sell per share (SPS). Two types of the example relapse condition were introduced, one including a catch term (α) and coefficients (β) for every autonomous variable, and another barring the block term. Based on the values of the independent variables, the regression models attempt to predict Price, the dependent variable. The examination was directed utilizing the R programming language. This exploration adds to the field by offering bits of knowledge into the prescient demonstration of value costs inside the IT area, with explicit pertinence to Apple Inc.

Keywords: statistical model, equity, information technology, descriptive statistics.

I. Introduction

The goal of this paper was to produce predictions regarding the future stock price of the massive corporation Apple Inc. by using statistical models. The Information Technology (IT) sector comprises six distinct industries, among which Apple Inc., which is a large-cap company has been the mainstay of my research. Apple Inc. is one of the most valuable companies in the world, with a market capitalization of over $2 trillion. In the study, a predictive model has been devised to estimate the equity price of the Apple Inc. company, employing diverse statistical techniques.

The model can be used by investors and analysts to make informed decisions about investing in Apple Inc. and to predict its future performance.

II. Previous Research

Predicting equity prices is a significant issue in financial research, especially in the Information Technology (IT) sector, where businesses such as Apple Inc. hold significant influence. This literature study evaluates current research on constructing statistical models for predicting Apple Inc.'s share price using regression analysis in the IT industry.

The impact of market opinion on the stock returns of Apple Inc., providing an understanding of issues affecting its equity price has been considered in this study [1]. For developing statistical models, especially in financial forecasting, [2] this original textbook offers foundational knowledge on econometrics, including regression analysis. The association between Apple Inc.'s financial performance and its stock price movements, providing perceptions to prospective forecasters for regression analysis has been demonstrated in this study [3]. This study [4] combined regression modeling to forecast changes in the company's stock price based on company performance metrics and its financial data. Regression analysis has been applied in this empirical study [5] to explore the relationship between several financial variables and Apple Inc. stock returns. By concentrating on challenges associated with regression analysis, such as multicollinearity and model overfitting a comprehensive guide has been provided this this paper [6]. This review [7] explores the use of sentiment analysis of social platforms like Twitter data to predict stock prices, offering potential further predictors for regression analysis. This study [8] provides on the parameters that

Author: Data Science, College of Collins Coll Prof. Studies Department of Computer Science, Math, and Science. St. John's University, Queens, New York.

affect volatility and stock returns, which may assist in determining which variables should be considered in a regression analysis. Tsay's book [9] provides sophisticated approaches for evaluating financial time series data, including methods relevant to regression analysis for predicting equities prices. Not only emphasizing Apple Inc., but this research paper also predicted the trend pattern by using machine learning techniques such as Support Vector Machine (SVM) that could be perfect for regression analysis in forecasting equity prices [10].

III. Methodology

The data has been taken from FactSet-2022 and it is time series data. The data is secondary data and there were 23 observations in the dataset. Generally, several statistical techniques have been used in this research project. For graphical techniques, histograms, scatter plots, line plots, and as analytical methods, three types of statistical tools have been used. I have used descriptive statistics for scalable variables (Price, EPS, BVPS, CR, DTA, EBIT, SPS) and to measure the linear association, I have used correlation, and finally to predict the model regression analysis. I have stated R as the scripting language.

The functional specification, population regression equation, and sample regression equation have been shown in the following Eqns. 1, 2 and 3.

$$+ \qquad + \qquad +$$

Eqn. 1 $$\text{Price} = f(\text{BVPS}, \text{EBIT}, \text{SPS})$$

Eqn. 2 $$\text{Price} = \alpha + \beta_{BVPS}*\text{BVPS} + \beta_{EBIT}*\text{EBIT} + \beta_{CR}*\text{SPS} + \varepsilon$$

Eqn. 3 $$\text{Price} = a\ b_{BVPS}*\text{BVPS} + \beta_{EBIT}*\text{EBIT} + \beta_{CR}*\text{SPS} + e$$

In the above multiple regression equations, price is the dependent variable which was hypothesized to be a positive function on independent variables respectively book value per share (BVPS), earnings before interest and taxes (EBIT) and sell per share(SPS).

IV. Results

For graphical analysis, a histogram (fig.1), scatter plot (fig. 2), and line-plot (fig. 3) have been developed for all dependent and independent variables.

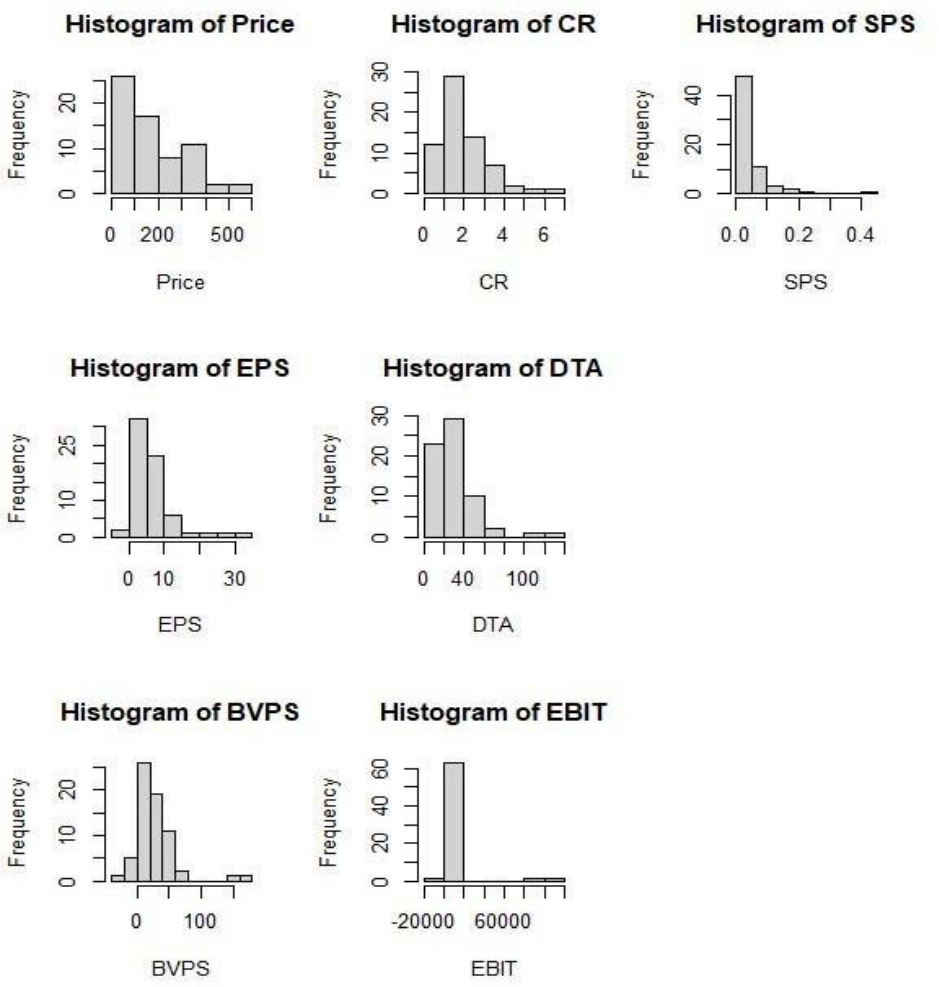


Figure1

From *fig.1* we can see that the histogram of Price, EPS, CR, SPS, DTA are positively skewed which is skewed to the right BVPS, and EBIT's histogram is relatively symmetric.

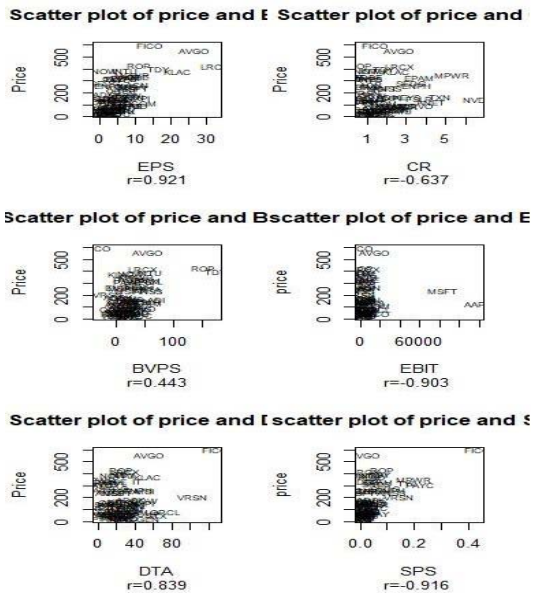


Figure 2

The scatter plot of dependent and independent variables is shown in *fig. 2.* Price and EPS are positively correlated (r=0.692) with very strong linear association. Whereas price and CR (r=-0.637) show a negative but moderate linear relation. The Price-BVPS scatterplot suggests a positive correlation (r=0.443).While price and EBIT (r=-0.903) shows a negative but strong linear relation. Price-DTA scatterplot (r=0.839) and Price-SPS scatterplot (r=0.916) show strong positive linear associations.

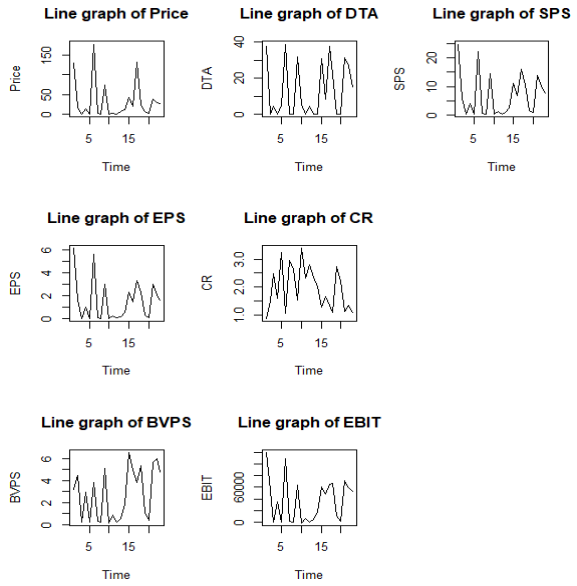


Figure 3

The line graph in *fig. 3* shows the trend of Price, EPS, BVPS, DTA, CR, EBIT and SPS with respect to month. Cyclical positive trend has been seen from the line graph of Price, EPS, BVPS, DTA, CR, EBIT and SPS. Whereas relatively secular trend with seasonality has been from DTA line graph.

Table 1: Descriptive Statistics

Name	n	Mean	Median	Std.Dev	Skewness	Kurtosis
Price	23	33.43	14.46	48.89	1.72	5.22
EPS	23	1.51	0.99	1.77	1.18	3.90
BVPS	23	2.73	2.94	2.29	0.16	1.44
CR	23	1.95	1.68	0.77	0.32	1.81
DTA	23	13.05	5.02	15.07	0.59	1.70
EBIT	23	36886.22	34494.00	36577.37	0.59	2.44
SPS	23	6.85	4.17	7.40	0.94	2.99

Table 1

Shows the descriptive statistics of the variables. From the table.1, the value of skewness indicates that the distribution of Price, EPS, BVPS, CR, DTA, EBIT, and SPS are positively skewed to the right and Price, EPS, and SPS have extreme values with high kurtosis, whereas others are significantly more peaked than symmetric leptokurtic distributions.

Table 2: Correlation Matrix

	Price	EPS	BVPS	CR	DTA	EBIT	SPS
Price	1.000	0.921	0.443	0.637	0.839	0.837	0.916
EPS	0.921	1.000	0.643	-0.801	0.832	0.971	0.993
BVPS	0.443	0.643	1.000	-0.866	0.703	0.786	0.689
CR	-0.637	-0.801	-0.866	1.000	-0.725	-0.893	-0.820
DTA	0.839	0.892	0.703	-0.725	1.000	0.864	0.929
EBIT	0.837	0.971	0.786	-0.893	0.864	1.000	0.970
SPS	0.916	0.993	0.689	-0.820	0.929	0.970	1.000

Table 2:

Shows the correlation matrix among the variables. Almost all variables showed a positive and strong correlation. Price and SPS show the highest with (r=0.921) while Price and CR indicate the lowest correlation(r=-0.637). The correlation coefficients of all variables agreed with the original hypothesis.

Table 3: Regression Results

Eqn. 4 Price = 1.433-6.15*BVPS+9.34*SPS-0.001*EBIT

	Price	BVPS	SPS	EBIT
t-stat	0.272	-2.234*	4.281***	-0.798
p-value	0.788	0.038	0.000	0.435
r(corr.)		0.443	0.916	0.837

n=23 , r^2 =0.91 F=62.93*** on cutoff 3 SE= 15.91

Our hypothesis is, Ho: βBVPS=βSPS=βEBIT=0

Ha: $\beta_i \neq 0$.

The overall equation is significant at 5% level of significance as F statistics is (62.93>3.00), so we can conclude that, null hypothesis (H0) is rejected. Since the coefficient of determination is r^2 =0.91, so 91% variation in price is explained by BVPS, SPS and EBIT. For significant test of the regression coefficient individually t-test has been done. SPS (4.281>2.99) and BVPS (|-2.234|>1.28) statistically significant at 1% and 10% levels respectively. Whereas EBIT (|-0.798|<1.28) is not statistically significant.

V. Conclusion

The study presents a statistical model for predicting the equity price of Apple Inc. in the Information Technology sector. The Research was successful, and the explanatory power was high. Price has been considered as the response variable concerning explanatory variables BVPS SPS and EBIT. All explanatory variables were significant except DTA. Price was a successful function of the explanatory variables, and all agreed with the hypothesis. The model can be used by investors and analysts to make informed decisions about investing in Apple Inc. and to predict its future performance. The research might be improved by considering adding more variables.

References Références Referencias

1. Chen, W., Lai, Y., & Su, C. (2018). The impact of market sentiment on stock returns: A case study of Apple Inc. International Journal of Business and Economics, 17(2), 167-182.
2. Gujarati, D. N., & Porter, D. C. (2009). Basic econometrics (5th ed.). McGraw-Hill Education.
3. Chandra, P., & Kumar, S. (2020). Financial performance and stock price movements: A study of Apple Inc. Journal of Commerce and Management Thought, 11(4), 885-892.
4. Lee, H., & Lee, J. (2017). Predicting stock price movements: A study of Apple Inc. Journal of Financial Research, 40(3), 361-376.
5. Li, C., & Tseng, Y. (2016). An empirical study of Apple Inc.'s stock returns: A regression analysis approach. Journal of Business and Economic Policy, 3(4), 108-116.
6. Kennedy, P. (2008). A guide to econometrics (6th ed.). Wiley-Blackwell.
7. Kim, J., & Kim, T. (2019). Predicting stock prices using sentiment analysis of Twitter data: An empirical analysis of the IT sector. Journal of Computational Science, 30, 36-46.
8. West, K. D. (2003). Stock returns and volatility: A firm-level analysis. Journal of Financial Economics, 67(2), 281-310.
9. Tsay, R. S. (2010). Analysis of financial time series. John Wiley & Sons.
10. Chen, J., Jiang, Z., & Wu, H. (2015). Stock trend prediction with technical indicators using SVM. Neurocomputing, 187, 220-235.

Global Journals Guidelines Handbook 2024

www.GlobalJournals.org

Memberships

Fellows/Associates of Computer Science Research Council

FCSRC/ACSRC Memberships

Introduction

FCSRC/ACSRC is the most prestigious membership of Global Journals accredited by Open Association of Research Society, U.S.A (OARS). The credentials of Fellow and Associate designations signify that the researcher has gained the knowledge of the fundamental and high-level concepts, and is a subject matter expert, proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice. The credentials are designated only to the researchers, scientists, and professionals that have been selected by a rigorous process by our Editorial Board and Management Board.

Associates of FCSRC/ACSRC are scientists and researchers from around the world are working on projects/researches that have huge potentials. Members support Global Journals' mission to advance technology for humanity and the profession.

FCSRC

Fellow of Computer Science Research Council

FELLOW OF COMPUTER SCIENCE RESEARCH COUNCIL is the most prestigious membership of Global Journals. It is an award and membership granted to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Fellows are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Fellow Members.

Benefits

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A FCSRC member gets access to a closed network of Tier 1 researchers and scientists with direct communication channel through our website. Fellows can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Fellows receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Fellows can use the honored title of membership. The “FCSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FCSRC or William Walldroff, M.S., FCSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Fellow members of FCSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation. All fellows get a dedicated page on the website with their biography.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Fellows receive discounts on future publications with Global Journals up to 60%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Fellows get secure and fast GJ work emails with unlimited forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, fellows receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Fellows are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All fellows receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 60% of sales proceeds

Fellows can publish articles (limited) without any fees. Also, they can earn up to 70% of sales proceeds from the sale of reference/review books/literature/publishing of research paper. The FCSRC member can decide its price and we can help in making the right decision.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Fellow members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

Access to Editorial Board

Become a member of the Editorial Board

Fellows may join as a member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. Additionally, Fellows get a chance to nominate other members for Editorial Board.

Career | Credibility | Exclusive | Reputation

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 5 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 10 GB free secure cloud access for storing research files.

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL

ASSOCIATE OF COMPUTER SCIENCE RESEARCH COUNCIL is the membership of Global Journals awarded to individuals that the Open Association of Research Society judges to have made a 'substantial contribution to the improvement of computer science, technology, and electronics engineering.

The primary objective is to recognize the leaders in research and scientific fields of the current era with a global perspective and to create a channel between them and other researchers for better exposure and knowledge sharing. Members are most eminent scientists, engineers, and technologists from all across the world. Associate membership can later be promoted to Fellow Membership. Associates are elected for life through a peer review process on the basis of excellence in the respective domain. There is no limit on the number of new nominations made in any year. Each year, the Open Association of Research Society elect up to 12 new Associate Members.

Benefits

To the institution

Get letter of appreciation

Global Journals sends a letter of appreciation of author to the Dean or CEO of the University or Company of which author is a part, signed by editor in chief or chief author.

Exclusive Network

Get access to a closed network

A ACSRC member gets access to a closed network of Tier 2 researchers and scientists with direct communication channel through our website. Associates can reach out to other members or researchers directly.They should also be open to reaching out by other.

Career | Credibility | Exclusive | Reputation

Certificate

Certificate, LoR and Laser-Momento

Associates receive a printed copy of a certificate signed by our Chief Author that may be used for academic purposes and a personal recommendation letter to the dean of member's university.

Career | Credibility | Exclusive | Reputation

Designation

Get honored title of membership

Associates can use the honored title of membership. The “ACSRC” is an honored title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., ACSRC or William Walldroff, M.S., ACSRC.

Career | Credibility | Exclusive | Reputation

Recognition on the Platform

Better visibility and citation

All the Associate members of ACSRC get a badge of "Leading Member of Global Journals" on the Research Community that distinguishes them from others. Additionally, the profile is also partially maintained by our team for better visibility and citation.

Career | Credibility | Reputation

Future Work

Get discounts on the future publications

Associates receive discounts on future publications with Global Journals up to 30%. Through our recommendation programs, members also receive discounts on publications made with OARS affiliated organizations.

Career | Financial

GJ Account

Unlimited forward of Emails

Associates get secure and fast GJ work emails with 5GB forward of emails that they may use them as their primary email. For example, john [AT] globaljournals [DOT] org.

Career | Credibility | Reputation

Premium Tools

Access to all the premium tools

To take future researches to the zenith, associates receive access to all the premium tools that Global Journals have to offer along with the partnership with some of the best marketing leading tools out there.

Financial

Conferences & Events

Organize seminar/conference

Associates are authorized to organize symposium/seminar/conference on behalf of Global Journal Incorporation (USA). They can also participate in the same organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent. Additionally, they get free research conferences (and others) alerts.

Career | Credibility | Financial

Early Invitations

Early invitations to all the symposiums, seminars, conferences

All associates receive the early invitations to all the symposiums, seminars, conferences and webinars hosted by Global Journals in their subject.

Exclusive

Publishing Articles & Books

Earn 30-40% of sales proceeds

Associates can publish articles (limited) without any fees. Also, they can earn up to 30-40% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.

Exclusive | Financial

Reviewers

Get a remuneration of 15% of author fees

Associate members are eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get a remuneration of 15% of author fees, taken from the author of a respective paper.

Financial

And Much More

Get access to scientific museums and observatories across the globe

All members get access to 2 selected scientific museums and observatories across the globe. All researches published with Global Journals will be kept under deep archival facilities across regions for future protections and disaster recovery. They get 5 GB free secure cloud access for storing research files.

Associate	Fellow	Research Group	Basic
$4800 **lifetime designation**	$6800 **lifetime designation**	$12500.00 **organizational**	APC **per article**
Certificate, LoR and Momento **2** discounted publishing/year **Gradation** of Research **10** research contacts/day **1 GB** Cloud Storage **GJ** Community Access	**Certificate,** LoR and Momento **Unlimited** discounted publishing/year **Gradation** of Research **Unlimited** research contacts/day **5 GB** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**Certificates,** LoRs and Momentos **Unlimited** free publishing/year **Gradation** of Research **Unlimited** research contacts/day **Unlimited** Cloud Storage **Online Presense** Assistance **GJ** Community Access	**GJ** Community Access

Preferred Author Guidelines

We accept the manuscript submissions in any standard (generic) format.

We typeset manuscripts using advanced typesetting tools like Adobe In Design, CorelDraw, TeXnicCenter, and TeXStudio. We usually recommend authors submit their research using any standard format they are comfortable with, and let Global Journals do the rest.

Alternatively, you can download our basic template from https://globaljournals.org/Template.zip

Authors should submit their complete paper/article, including text illustrations, graphics, conclusions, artwork, and tables. Authors who are not able to submit manuscript using the form above can email the manuscript department at submit@globaljournals.org or get in touch with chiefeditor@globaljournals.org if they wish to send the abstract before submission.

Before and during Submission

Authors must ensure the information provided during the submission of a paper is authentic. Please go through the following checklist before submitting:

1. Authors must go through the complete author guideline and understand and *agree to Global Journals' ethics and code of conduct,* along with author responsibilities.
2. Authors must accept the privacy policy, terms, and conditions of Global Journals.
3. Ensure corresponding author's email address and postal address are accurate and reachable.
4. Manuscript to be submitted must include keywords, an abstract, a paper title, co-author(s') names and details (email address, name, phone number, and institution), figures and illustrations in vector format including appropriate captions, tables, including titles and footnotes, a conclusion, results, acknowledgments and references.
5. Authors should submit paper in a ZIP archive if any supplementary files are required along with the paper.
6. Proper permissions must be acquired for the use of any copyrighted material.
7. Manuscript submitted *must not have been submitted or published elsewhere* and all authors must be aware of the submission.

Declaration of Conflicts of Interest

It is required for authors to declare all financial, institutional, and personal relationships with other individuals and organizations that could influence (bias) their research.

Policy on Plagiarism

Plagiarism is not acceptable in Global Journals submissions at all.

Plagiarized content will not be considered for publication. We reserve the right to inform authors' institutions about plagiarism detected either before or after publication. If plagiarism is identified, we will follow COPE guidelines:

Authors are solely responsible for all the plagiarism that is found. The author must not fabricate, falsify or plagiarize existing research data. The following, if copied, will be considered plagiarism:

- Words (language)
- Ideas
- Findings
- Writings
- Diagrams
- Graphs
- Illustrations
- Lectures

- Printed material
- Graphic representations
- Computer programs
- Electronic material
- Any other original work

Authorship Policies

Global Journals follows the definition of authorship set up by the Open Association of Research Society, USA. According to its guidelines, authorship criteria must be based on:

1. Substantial contributions to the conception and acquisition of data, analysis, and interpretation of findings.
2. Drafting the paper and revising it critically regarding important academic content.
3. Final approval of the version of the paper to be published.

Changes in Authorship

The corresponding author should mention the name and complete details of all co-authors during submission and in manuscript. We support addition, rearrangement, manipulation, and deletions in authors list till the early view publication of the journal. We expect that corresponding author will notify all co-authors of submission. We follow COPE guidelines for changes in authorship.

Copyright

During submission of the manuscript, the author is confirming an exclusive license agreement with Global Journals which gives Global Journals the authority to reproduce, reuse, and republish authors' research. We also believe in flexible copyright terms where copyright may remain with authors/employers/institutions as well. Contact your editor after acceptance to choose your copyright policy. You may follow this form for copyright transfers.

Appealing Decisions

Unless specified in the notification, the Editorial Board's decision on publication of the paper is final and cannot be appealed before making the major change in the manuscript.

Acknowledgments

Contributors to the research other than authors credited should be mentioned in Acknowledgments. The source of funding for the research can be included. Suppliers of resources may be mentioned along with their addresses.

Declaration of funding sources

Global Journals is in partnership with various universities, laboratories, and other institutions worldwide in the research domain. Authors are requested to disclose their source of funding during every stage of their research, such as making analysis, performing laboratory operations, computing data, and using institutional resources, from writing an article to its submission. This will also help authors to get reimbursements by requesting an open access publication letter from Global Journals and submitting to the respective funding source.

Preparing your Manuscript

Authors can submit papers and articles in an acceptable file format: MS Word (doc, docx), LaTeX (.tex, .zip or .rar including all of your files), Adobe PDF (.pdf), rich text format (.rtf), simple text document (.txt), Open Document Text (.odt), and Apple Pages (.pages). Our professional layout editors will format the entire paper according to our official guidelines. This is one of the highlights of publishing with Global Journals—authors should not be concerned about the formatting of their paper. Global Journals accepts articles and manuscripts in every major language, be it Spanish, Chinese, Japanese, Portuguese, Russian, French, German, Dutch, Italian, Greek, or any other national language, but the title, subtitle, and abstract should be in English. This will facilitate indexing and the pre-peer review process.

The following is the official style and template developed for publication of a research paper. Authors are not required to follow this style during the submission of the paper. It is just for reference purposes.

Manuscript Style Instruction (Optional)

- Microsoft Word Document Setting Instructions.
- Font type of all text should be Swis721 Lt BT.
- Page size: 8.27" x 11'", left margin: 0.65, right margin: 0.65, bottom margin: 0.75.
- Paper title should be in one column of font size 24.
- Author name in font size of 11 in one column.
- Abstract: font size 9 with the word "Abstract" in bold italics.
- Main text: font size 10 with two justified columns.
- Two columns with equal column width of 3.38 and spacing of 0.2.
- First character must be three lines drop-capped.
- The paragraph before spacing of 1 pt and after of 0 pt.
- Line spacing of 1 pt.
- Large images must be in one column.
- The names of first main headings (Heading 1) must be in Roman font, capital letters, and font size of 10.
- The names of second main headings (Heading 2) must not include numbers and must be in italics with a font size of 10.

Structure and Format of Manuscript

The recommended size of an original research paper is under 15,000 words and review papers under 7,000 words. Research articles should be less than 10,000 words. Research papers are usually longer than review papers. Review papers are reports of significant research (typically less than 7,000 words, including tables, figures, and references)

A research paper must include:

a) A title which should be relevant to the theme of the paper.
b) A summary, known as an abstract (less than 150 words), containing the major results and conclusions.
c) Up to 10 keywords that precisely identify the paper's subject, purpose, and focus.
d) An introduction, giving fundamental background objectives.
e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition, sources of information must be given, and numerical methods must be specified by reference.
f) Results which should be presented concisely by well-designed tables and figures.
g) Suitable statistical data should also be given.
h) All data must have been gathered with attention to numerical detail in the planning stage.

Design has been recognized to be essential to experiments for a considerable time, and the editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned unrefereed.

i) Discussion should cover implications and consequences and not just recapitulate the results; conclusions should also be summarized.
j) There should be brief acknowledgments.
k) There ought to be references in the conventional format. Global Journals recommends APA format.

Authors should carefully consider the preparation of papers to ensure that they communicate effectively. Papers are much more likely to be accepted if they are carefully designed and laid out, contain few or no errors, are summarizing, and follow instructions. They will also be published with much fewer delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and suggestions to improve brevity.

Format Structure

It is necessary that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

All manuscripts submitted to Global Journals should include:

Title

The title page must carry an informative title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) where the work was carried out.

Author details

The full postal address of any related author(s) must be specified.

Abstract

The abstract is the foundation of the research paper. It should be clear and concise and must contain the objective of the paper and inferences drawn. It is advised to not include big mathematical equations or complicated jargon.

Many researchers searching for information online will use search engines such as Google, Yahoo or others. By optimizing your paper for search engines, you will amplify the chance of someone finding it. In turn, this will make it more likely to be viewed and cited in further works. Global Journals has compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Keywords

A major lynchpin of research work for the writing of research papers is the keyword search, which one will employ to find both library and internet resources. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining, and indexing.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy: planning of a list of possible keywords and phrases to try.

Choice of the main keywords is the first tool of writing a research paper. Research paper writing is an art. Keyword search should be as strategic as possible.

One should start brainstorming lists of potential keywords before even beginning searching. Think about the most important concepts related to research work. Ask, “What words would a source have to include to be truly valuable in a research paper?” Then consider synonyms for the important words.

It may take the discovery of only one important paper to steer in the right keyword direction because, in most databases, the keywords under which a research paper is abstracted are listed with the paper.

Numerical Methods

Numerical methods used should be transparent and, where appropriate, supported by references.

Abbreviations

Authors must list all the abbreviations used in the paper at the end of the paper or in a separate table before using them.

Formulas and equations

Authors are advised to submit any mathematical equation using either MathJax, KaTeX, or LaTeX, or in a very high-quality image.

Tables, Figures, and Figure Legends

Tables: Tables should be cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g., Table 4, a self-explanatory caption, and be on a separate sheet. Authors must submit tables in an editable format and not as images. References to these tables (if any) must be mentioned accurately.

Figures

Figures are supposed to be submitted as separate files. Always include a citation in the text for each figure using Arabic numbers, e.g., Fig. 4. Artwork must be submitted online in vector electronic form or by emailing it.

Preparation of Eletronic Figures for Publication

Although low-quality images are sufficient for review purposes, print publication requires high-quality images to prevent the final product being blurred or fuzzy. Submit (possibly by e-mail) EPS (line art) or TIFF (halftone/ photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Avoid using pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings). Please give the data for figures in black and white or submit a Color Work Agreement form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution at final image size ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs): >350 dpi; figures containing both halftone and line images: >650 dpi.

Color charges: Authors are advised to pay the full cost for the reproduction of their color artwork. Hence, please note that if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a Color Work Agreement form before your paper can be published. Also, you can email your editor to remove the color fee after acceptance of the paper.

Tips for writing a good quality Computer Science Research Paper

Techniques for writing a good quality computer science research paper:

1. Choosing the topic: In most cases, the topic is selected by the interests of the author, but it can also be suggested by the guides. You can have several topics, and then judge which you are most comfortable with. This may be done by asking several questions of yourself, like "Will I be able to carry out a search in this area? Will I find all necessary resources to accomplish the search? Will I be able to find all information in this field area?" If the answer to this type of question is "yes," then you ought to choose that topic. In most cases, you may have to conduct surveys and visit several places. Also, you might have to do a lot of work to find all the rises and falls of the various data on that subject. Sometimes, detailed information plays a vital role, instead of short information. Evaluators are human: The first thing to remember is that evaluators are also human beings. They are not only meant for rejecting a paper. They are here to evaluate your paper. So present your best aspect.

2. Think like evaluators: If you are in confusion or getting demotivated because your paper may not be accepted by the evaluators, then think, and try to evaluate your paper like an evaluator. Try to understand what an evaluator wants in your research paper, and you will automatically have your answer. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

3. Ask your guides: If you are having any difficulty with your research, then do not hesitate to share your difficulty with your guide (if you have one). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work, then ask your supervisor to help you with an alternative. He or she might also provide you with a list of essential readings.

4. Use of computer is recommended: As you are doing research in the field of computer science then this point is quite obvious. Use right software: Always use good quality software packages. If you are not capable of judging good software, then you can lose the quality of your paper unknowingly. There are various programs available to help you which you can get through the internet.

5. Use the internet for help: An excellent start for your paper is using Google. It is a wondrous search engine, where you can have your doubts resolved. You may also read some answers for the frequent question of how to write your research paper or find a model research paper. You can download books from the internet. If you have all the required books, place importance on reading, selecting, and analyzing the specified information. Then sketch out your research paper. Use big pictures: You may use encyclopedias like Wikipedia to get pictures with the best resolution. At Global Journals, you should strictly follow here.

6. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right? It is a good habit which helps to not lose your continuity. You should always use bookmarks while searching on the internet also, which will make your search easier.

7. Revise what you wrote: When you write anything, always read it, summarize it, and then finalize it.

8. Make every effort: Make every effort to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in the introduction—what is the need for a particular research paper. Polish your work with good writing skills and always give an evaluator what he wants. Make backups: When you are going to do any important thing like making a research paper, you should always have backup copies of it either on your computer or on paper. This protects you from losing any portion of your important data.

9. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several unnecessary diagrams will degrade the quality of your paper by creating a hodgepodge. So always try to include diagrams which were made by you to improve the readability of your paper. Use of direct quotes: When you do research relevant to literature, history, or current affairs, then use of quotes becomes essential, but if the study is relevant to science, use of quotes is not preferable.

10.Use proper verb tense: Use proper verb tenses in your paper. Use past tense to present those events that have happened. Use present tense to indicate events that are going on. Use future tense to indicate events that will happen in the future. Use of wrong tenses will confuse the evaluator. Avoid sentences that are incomplete.

11. Pick a good study spot: Always try to pick a spot for your research which is quiet. Not every spot is good for studying.

12. Know what you know: Always try to know what you know by making objectives, otherwise you will be confused and unable to achieve your target.

13. Use good grammar: Always use good grammar and words that will have a positive impact on the evaluator; use of good vocabulary does not mean using tough words which the evaluator has to find in a dictionary. Do not fragment sentences. Eliminate one-word sentences. Do not ever use a big word when a smaller one would suffice.

Verbs have to be in agreement with their subjects. In a research paper, do not start sentences with conjunctions or finish them with prepositions. When writing formally, it is advisable to never split an infinitive because someone will (wrongly) complain. Avoid clichés like a disease. Always shun irritating alliteration. Use language which is simple and straightforward. Put together a neat summary.

14. Arrangement of information: Each section of the main body should start with an opening sentence, and there should be a changeover at the end of the section. Give only valid and powerful arguments for your topic. You may also maintain your arguments with records.

15. Never start at the last minute: Always allow enough time for research work. Leaving everything to the last minute will degrade your paper and spoil your work.

16. Multitasking in research is not good: Doing several things at the same time is a bad habit in the case of research activity. Research is an area where everything has a particular time slot. Divide your research work into parts, and do a particular part in a particular time slot.

17. Never copy others' work: Never copy others' work and give it your name because if the evaluator has seen it anywhere, you will be in trouble. Take proper rest and food: No matter how many hours you spend on your research activity, if you are not taking care of your health, then all your efforts will have been in vain. For quality research, take proper rest and food.

18. Go to seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

19. Refresh your mind after intervals: Try to give your mind a rest by listening to soft music or sleeping in intervals. This will also improve your memory. Acquire colleagues: Always try to acquire colleagues. No matter how sharp you are, if you acquire colleagues, they can give you ideas which will be helpful to your research.

20. Think technically: Always think technically. If anything happens, search for its reasons, benefits, and demerits. Think and then print: When you go to print your paper, check that tables are not split, headings are not detached from their descriptions, and page sequence is maintained.

21. Adding unnecessary information: Do not add unnecessary information like "I have used MS Excel to draw graphs." Irrelevant and inappropriate material is superfluous. Foreign terminology and phrases are not apropos. One should never take a broad view. Analogy is like feathers on a snake. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Never oversimplify: When adding material to your research paper, never go for oversimplification; this will definitely irritate the evaluator. Be specific. Never use rhythmic redundancies. Contractions shouldn't be used in a research paper. Comparisons are as terrible as clichés. Give up ampersands, abbreviations, and so on. Remove commas that are not necessary. Parenthetical words should be between brackets or commas. Understatement is always the best way to put forward earth-shaking thoughts. Give a detailed literary review.

22. Report concluded results: Use concluded results. From raw data, filter the results, and then conclude your studies based on measurements and observations taken. An appropriate number of decimal places should be used. Parenthetical remarks are prohibited here. Proofread carefully at the final stage. At the end, give an outline to your arguments. Spot perspectives of further study of the subject. Justify your conclusion at the bottom sufficiently, which will probably include examples.

23. Upon conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print for the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects of your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form which is presented in the guidelines using the template.
- Please note the criteria peer reviewers will use for grading the final paper.

Final points:

One purpose of organizing a research paper is to let people interpret your efforts selectively. The journal requires the following sections, submitted in the order listed, with each section starting on a new page:

The introduction: This will be compiled from reference matter and reflect the design processes or outline of basis that directed you to make a study. As you carry out the process of study, the method and process section will be constructed like that. The results segment will show related statistics in nearly sequential order and direct reviewers to similar intellectual paths throughout the data that you gathered to carry out your study.

The discussion section:

This will provide understanding of the data and projections as to the implications of the results. The use of good quality references throughout the paper will give the effort trustworthiness by representing an alertness to prior workings.

Writing a research paper is not an easy job, no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record-keeping are the only means to make straightforward progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear: Adhere to recommended page limits.

Mistakes to avoid:

- Insertion of a title at the foot of a page with subsequent text on the next page.
- Separating a table, chart, or figure—confine each to a single page.
- Submitting a manuscript with pages out of sequence.
- In every section of your document, use standard writing style, including articles ("a" and "the").
- Keep paying attention to the topic of the paper.
- Use paragraphs to split each significant point (excluding the abstract).
- Align the primary line of each section.
- Present your points in sound order.
- Use present tense to report well-accepted matters.
- Use past tense to describe specific results.
- Do not use familiar wording; don't address the reviewer directly. Don't use slang or superlatives.
- Avoid use of extra pictures—include only those figures essential to presenting results.

Title page:

Choose a revealing title. It should be short and include the name(s) and address(es) of all authors. It should not have acronyms or abbreviations or exceed two printed lines.

Abstract: This summary should be two hundred words or less. It should clearly and briefly explain the key findings reported in the manuscript and must have precise statistics. It should not have acronyms or abbreviations. It should be logical in itself. Do not cite references at this point.

An abstract is a brief, distinct paragraph summary of finished work or work in development. In a minute or less, a reviewer can be taught the foundation behind the study, common approaches to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Use comprehensive sentences, and do not sacrifice readability for brevity; you can maintain it succinctly by phrasing sentences so that they provide more than a lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study with the subsequent elements in any summary. Try to limit the initial two items to no more than one line each.

Reason for writing the article—theory, overall issue, purpose.

- Fundamental goal.
- To-the-point depiction of the research.
- Consequences, including definite statistics—if the consequences are quantitative in nature, account for this; results of any numerical analysis should be reported. Significant conclusions or questions that emerge from the research.

Approach:

- Single section and succinct.
- An outline of the job done is always written in past tense.
- Concentrate on shortening results—limit background information to a verdict or two.
- Exact spelling, clarity of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else.

Introduction:

The introduction should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable of comprehending and calculating the purpose of your study without having to refer to other works. The basis for the study should be offered. Give the most important references, but avoid making a comprehensive appraisal of the topic. Describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will give no attention to your results. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here.

The following approach can create a valuable beginning:

- Explain the value (significance) of the study.
- Defend the model—why did you employ this particular system or method? What is its compensation? Remark upon its appropriateness from an abstract point of view as well as pointing out sensible reasons for using it.
- Present a justification. State your particular theory(-ies) or aim(s), and describe the logic that led you to choose them.
- Briefly explain the study's tentative purpose and how it meets the declared objectives.

Approach:

Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done. Sort out your thoughts; manufacture one key point for every section. If you make the four points listed above, you will need at least four paragraphs. Present surrounding information only when it is necessary to support a situation. The reviewer does not desire to read everything you know about a topic. Shape the theory specifically—do not take a broad view.

As always, give awareness to spelling, simplicity, and correctness of sentences and phrases.

Procedures (methods and materials):

This part is supposed to be the easiest to carve if you have good skills. A soundly written procedures segment allows a capable scientist to replicate your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order, but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt to give the least amount of information that would permit another capable scientist to replicate your outcome, but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section.

When a technique is used that has been well-described in another section, mention the specific item describing the way, but draw the basic principle while stating the situation. The purpose is to show all particular resources and broad procedures so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step-by-step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

Materials may be reported in part of a section or else they may be recognized along with your measures.

Methods:

- Report the method and not the particulars of each process that engaged the same methodology.
- Describe the method entirely.
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures.
- Simplify—detail how procedures were completed, not how they were performed on a particular day.
- If well-known procedures were used, account for the procedure by name, possibly with a reference, and that's all.

Approach:

It is embarrassing to use vigorous voice when documenting methods without using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result, when writing up the methods, most authors use third person passive voice.

Use standard style in this and every other part of the paper—avoid familiar lists, and use full sentences.

What to keep away from:

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings—save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part as entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Use statistics and tables, if suitable, to present consequences most efficiently.

You must clearly differentiate material which would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matters should not be submitted at all except if requested by the instructor.

Content:

- Sum up your conclusions in text and demonstrate them, if suitable, with figures and tables.
- In the manuscript, explain each of your consequences, and point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation of an exacting study.
- Explain results of control experiments and give remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or manuscript.

What to stay away from:

- Do not discuss or infer your outcome, report surrounding information, or try to explain anything.
- Do not include raw data or intermediate calculations in a research manuscript.
- Do not present similar data more than once.
- A manuscript should complement any figures or tables, not duplicate information.
- Never confuse figures with tables—there is a difference.

Approach:

As always, use past tense when you submit your results, and put the whole thing in a reasonable order.

Put figures and tables, appropriately numbered, in order at the end of the report.

If you desire, you may place your figures and tables properly within the text of your results section.

Figures and tables:

If you put figures and tables at the end of some details, make certain that they are visibly distinguished from any attached appendix materials, such as raw facts. Whatever the position, each table must be titled, numbered one after the other, and include a heading. All figures and tables must be divided from the text.

Discussion:

The discussion is expected to be the trickiest segment to write. A lot of papers submitted to the journal are discarded based on problems with the discussion. There is no rule for how long an argument should be.

Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implications of the study. The purpose here is to offer an understanding of your results and support all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of results should be fully described.

Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact, you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved the prospect, and let it drop at that. Make a decision as to whether each premise is supported or discarded or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."

Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work.

- You may propose future guidelines, such as how an experiment might be personalized to accomplish a new idea.
- Give details of all of your remarks as much as possible, focusing on mechanisms.
- Make a decision as to whether the tentative design sufficiently addressed the theory and whether or not it was correctly restricted. Try to present substitute explanations if they are sensible alternatives.
- One piece of research will not counter an overall question, so maintain the large picture in mind. Where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

When you refer to information, differentiate data generated by your own studies from other available information. Present work done by specific persons (including you) in past tense.

Describe generally acknowledged facts and main beliefs in present tense.

The Administration Rules

Administration Rules to Be Strictly Followed before Submitting Your Research Paper to Global Journals Inc.

Please read the following rules and regulations carefully before submitting your research paper to Global Journals Inc. to avoid rejection.

Segment draft and final research paper: You have to strictly follow the template of a research paper, failing which your paper may get rejected. You are expected to write each part of the paper wholly on your own. The peer reviewers need to identify your own perspective of the concepts in your own terms. Please do not extract straight from any other source, and do not rephrase someone else's analysis. Do not allow anyone else to proofread your manuscript.

Written material: You may discuss this with your guides and key sources. Do not copy anyone else's paper, even if this is only imitation, otherwise it will be rejected on the grounds of plagiarism, which is illegal. Various methods to avoid plagiarism are strictly applied by us to every paper, and, if found guilty, you may be blacklisted, which could affect your career adversely. To guard yourself and others from possible illegal use, please do not permit anyone to use or even read your paper and file.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350