Global Journal of Advanced Computer Science and Technology

Volume No. 13

Issue No. 2

May - August 2025

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

Global Journal of Advanced Computer Science and Technology

Aims and Scope

Global Journal of Advanced Computer Science & Technology deals with information technology, its evolution and future prospects and its relationship with the Business Management. It addresses technological, managerial, political, economic and organizational aspects of the application of IT in relationship with Business Management. The journal will serve as a comprehensiveresource for policy makers, government officials, academicians, and practitioners. GJACST promotes and coordinates the developments in the IT based applications of business management and presents thestrategic roles of IT and management towards sustainable development.

Global Journal of Advanced Computer Science and Technology

Global Journal of Advanced Computer Science and Technology

(Volume No. 13, Issue No. 2, May - August 2025)

Contents

Sr. No	Article/ Authors	Pg No
01	Innovative Approaches to Fake News Detection: A Data Mining Perspective - *Mustapha Ismail α, Yayale Isihaka Muhammad σ& Abdullahi Modibbo Abdullahi ρ*	1 - 22
02	Cyber Security Test Platform Establishments and Cyberattacks Practice *- Dr. Chetanpal Singh α, Ass. Professor Rahul Thakkar σ, Jatinder Warraich ρ, Numan Ahmed Ꞥ & Vimal B. Patel ¥*	23 - 44
03	Market Basket Analysis using Machine Learning *- Atul Sharma α, Dr. Mohammad Salim Hamidi σ & Yousuf Hotak ρ*	45 - 51
04	Survey: Artificial Intelligence Ethics - *Shuxi Wang α & Zengyan Xia σ*	52 - 64

Innovative Approaches to Fake News Detection: A Data Mining Perspective

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

A B S T R A C T

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 nontraditional 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

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 MMCOVID 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 WangCNN 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 pretrained model extracted from Felipe Adachi. The accuracy of the model is around 95% for the entire selfmade 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 itwith 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 feedbackbased approaches for fake news detection. In contentbased 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 toincrease GMB's speed and prediction performance; it is scalable and integrated into a different 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.

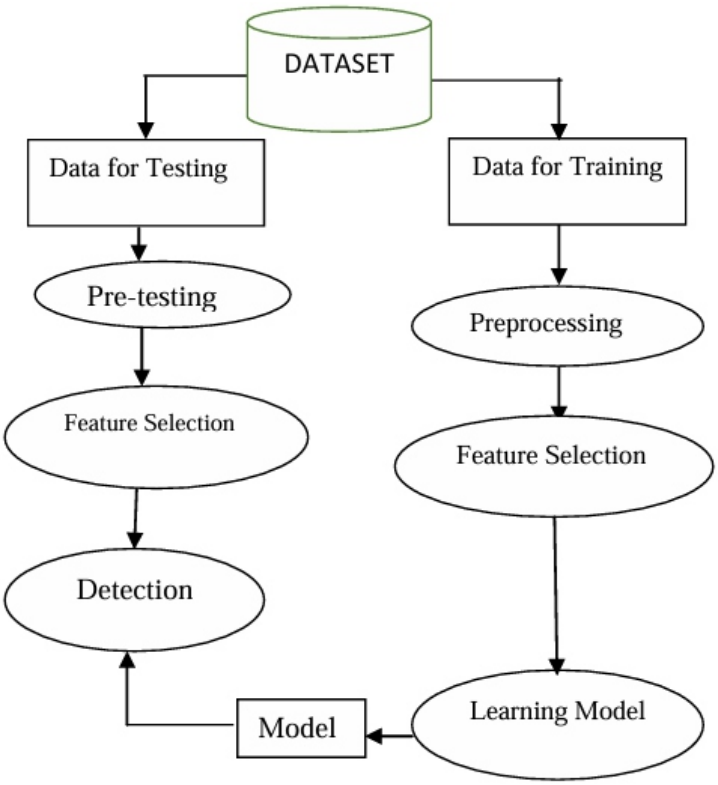


Figure 1: Framework of the Study

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. Scikitlearn python machine learning library is used for designing, building; thesystem 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 whichis 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
fromgoogle.colabimportdrive
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

ake 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 andstop 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 underfitting 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 andcolumns 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 beingfaster 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 includesupport 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, F1score, 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 isa 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 meansomeone 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 specificto 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.orgwhere 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 thechart 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.

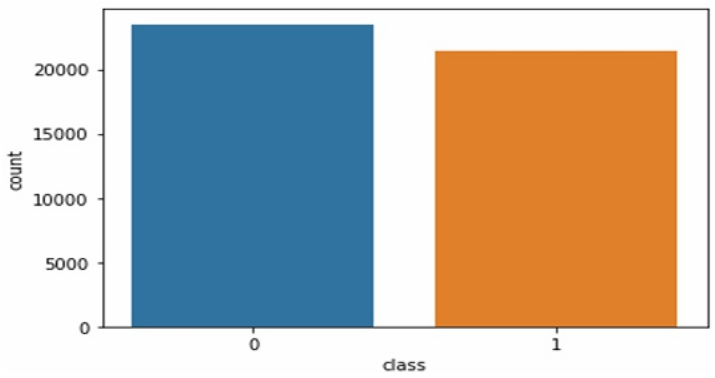


Figure 2: Count of Fake and Real News

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 14false positive predictions and 32 false negative predictions. Hence, the decision tree algorithm performed very good prediction.

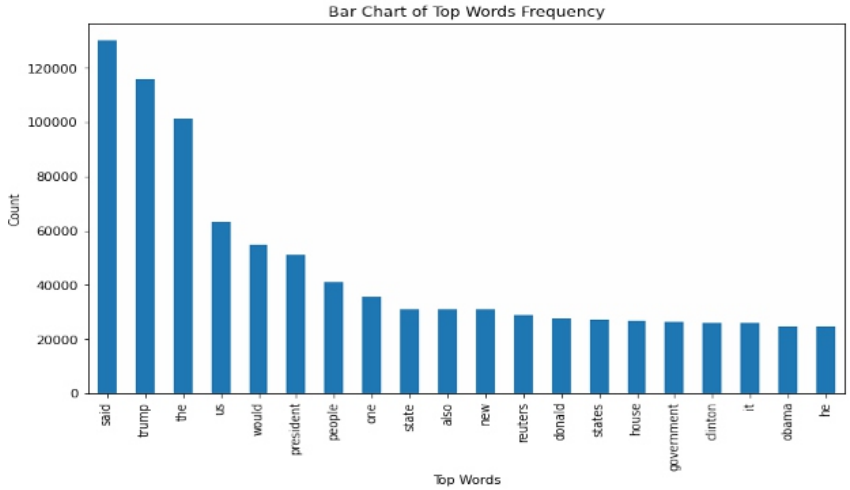


Figure 5: Bar Chart of Top Words Frequency

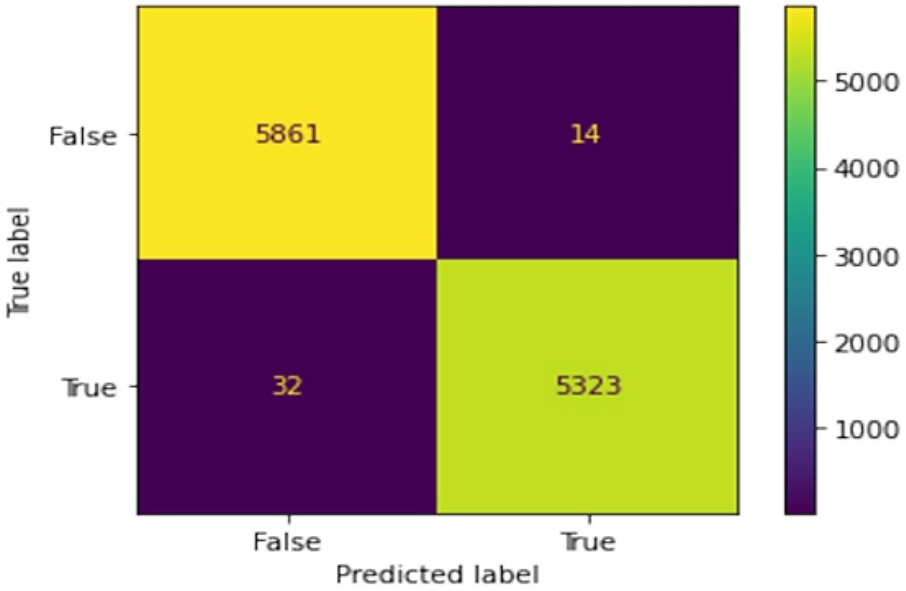


Figure 6: Confusion Matrix of Decision Tree Model

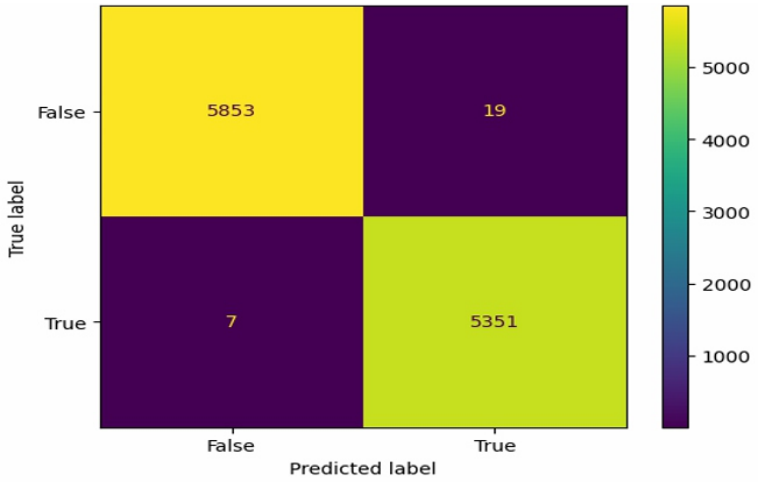


Figure 7: Confusion Matrix of Xgboost Model

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.

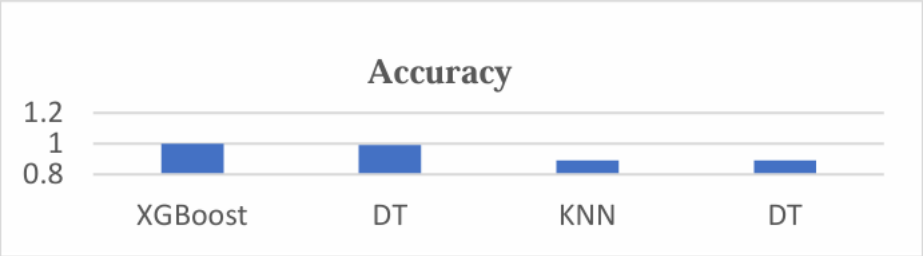


Figure 7: Accuracy of the Prediction Model

The ratio of accurately predicted positive observations to total expected positive observations is known as precision. As indicated in figure 8, the

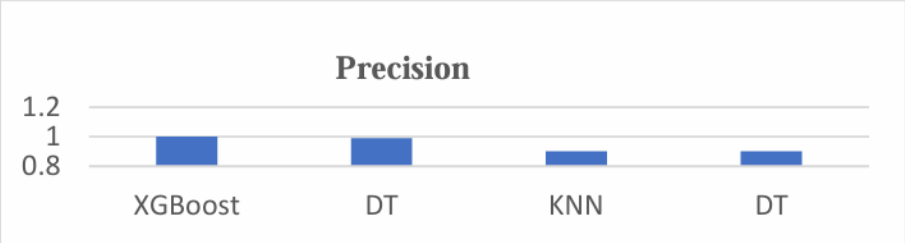


Figure 8: Precision of the Models

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

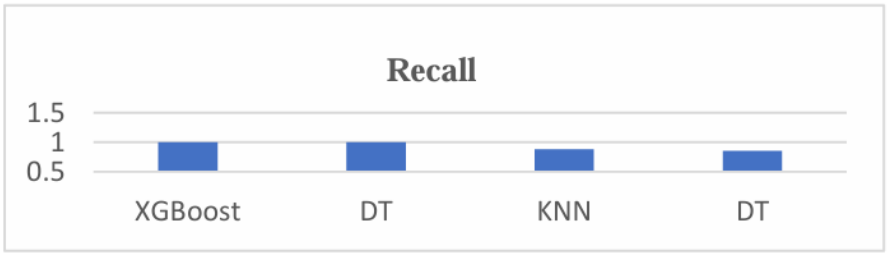


Figure 9: Recall Performance of the Models

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.

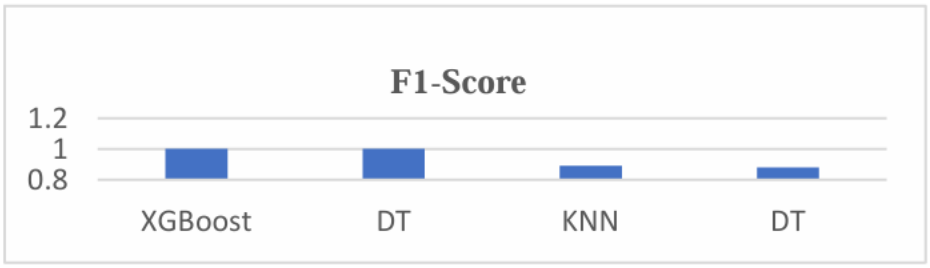


Figure 10: F1-Score Performance of the Models

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

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 TriRelationship 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-learningalgorithms"/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-socialmedia

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 CaliforniaMING ZHANG, Peking UniversityYAN 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 Approaches1College 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/359/6380/1 094.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-platform s-2016/(2016) Year 2024 14 Global Journal of Computer Science and Technology (C) XXIV Issue I Version I

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. AlNabhan. (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.

Cyber Security Test Platform Establishments and Cyberattacks Practice

Dr. Chetanpal Singh α, Ass. Professor Rahul Thakkar σ, Jatinder Warraich ρ, Numan Ahmed Ꞥ & Vimal B. Patel ¥

A B S T R A C T

In this study, cyber security test platform aims to evaluate the vulnerabilities, of cyber-attack exercises to review cyber security challenges. In the introduction section, brief overview of the research context has been provided by developing research questions, and determining the problem statements. In the Literature review section, different articles will be analyzed for gathering better information about the research questions. Secondary research methodology will be utilized in this research paper, and brief explanation of chosen research methodology has been mentioned in the third section. The main purpose of this research paper is to conduct a proper platform, which can detect cyber-attack, and decrease the attack numbers. This paper will provide several improvement of the proposed platform for developing the scalability.

Index Terms:cyber exercise, test platform, cyber-physical system, security applications.

I. INTRODUCTION:

a) Research background

Acyber security platform is basically a key solution that has been used to secure and control an organisation's data and network systems. The cyber security testing platform is a privileged access and security audit system that is performed to identify vulnerabilities, weakness, as well as misconfigurations of the targeted hosts [1]. In this modern digitised era, for every organisation, cybersecurity is an important area that helps to provide safeguards from all types of possible cybersecurity risks. Effective cyber security measurements help the organisation to reduce the possibilities of successful attacks as well as minimise the damages that a cyberattack can cause. In every organisation, the importance of several cybersecurity practices is more relevant, and the possibility of a data breach is reduced through the implementation of the measurements of security practices that utilise effective authentication mechanisms [1]. The chances of cybersecurity risks increase because of too much involvement in the latest technologies.

This research study will help the researchers to know the importance of the establishment of cyber security test platforms against any kinds of important information leakages, software & hardware damages, data thefts, as well as interruption of various services. The capability to understand and evaluate the threat data assists in reducing any damage as well as realising the flaws [2]. The internet system is completed in the company of priceless information as well as technical facilities which have eased the individuals with so much malicious information. The quality of the data can degenerate unintentionally with the assistance of data integrity tasks. Moreover, cyber security proposes a process to protect the entire information system of a company that is connected through modern internet systems. There are so many solutions for cyber security tests, and those are network security, mobile security, data security, application security, operational security, identity management, database and infrastructure security [2]. The two important tools that are used to do the security testing tools are "static application security testing" (SAST) along with "dynamic application security testing" (DAST). This research work will help the researcher to continue the research to highlight all the essential areas of the research work.

As the number of cybercrimes is increasing day by day, cyber security test platforms ensure the community continuously depends on their activities and services. The main goal behind cyber security testing is to point out the threats within the system and calculate the effective vulnerabilities in which way all the dangers can be easily encountered and where the system pauses its functions [3]. In this research, both the readers and the future researchers, after reviewing this research paper, can easily come to know the reason for which cyber security test platforms are arranged in most organisations or firms. In this process, various machine learning algorithms are used to maintain the validity, reliability, and generalizability of the organisation's information security system. Security testing proposes a software testing form that has been performed to analyse the entire system against any security-based expectations [3]. The purpose behind continuing this research is to make applications impenetrable to the possible security threats in the vicinity of identifying both vulnerabilities and weaknesses of the security systems. The security system is basically used for the identification of potential vulnerable threats and the measurements of the overall security systems. After continuing the research work, the researcher needs to be controlled to continue the research in the insight of the general risks of facing the software. The actionable insight from these proposed topics is used to complete the security risks and gaps.

b) Problem Statement

This research work is conducted based on highlighting the importance of the threats as well as measurements of the effective vulnerabilities that encountered the threats and issues faced in the information system. This research work will help to identify the vulnerabilities which are actively used

within the organisation that used to lead an entirely insured security-based incident. As nowadays cyber crimes are rapidly increasing, the in-depth reason behind the importance of the establishment of cyber security test platforms and cyber security practices is needed, which will help the future researcher to carry on future research on numerous important areas.

c) Research Aims & Objectives Aim:

This research study is continued with the aim of continuing the research work to highlight the importance of cyber security test platforms establishments along with cyberattack performances.
Objectives:

This research work will be conducted focusing on the following objectives and those objectives are,

• To point out the roles and responsibilities of cyber security test platforms as well as various cyberattack practices.
• To identify the required tools that are used as effective cybersecurity software tools for maintaining information security.
• To discover the current trends of the latest cybersecurity best activities, measurements, and techniques to strengthen the information security system of a company.
• To identify the issues that can be solved through cyber security test platforms and cyberattack practices.

d) Research Questions

RQ1: What are the roles and importance of the establishment of cyber security platforms and cyberattack practices to ensure an effective information system?

RQ2: What are the current trends of the latest cybersecurity best activities, measurements, and techniques to strengthen the information security system?

RQ3: What are the issues that can be solved through cyber security test platforms and cyberattack practices? RQ4: What are the required tools which are used as effective cybersecurity software tools for maintaining information security?

e) Research Significance

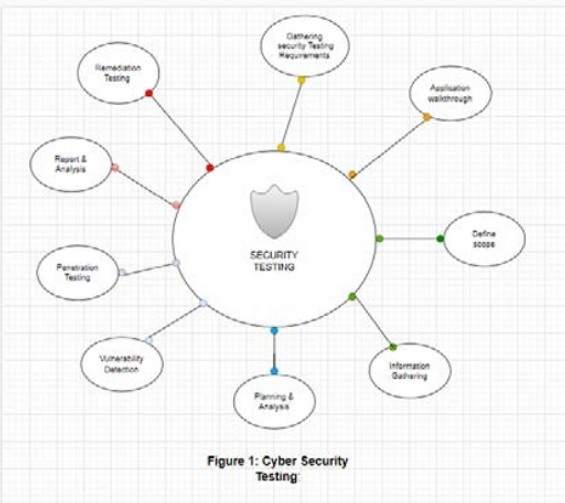
The security advisories issued every year by the ICS-CERT (‘Industrial Control System- Computer

Emergency Response Team') are rapidly increasing. So the significance of this research work is to highlight the important areas and tools that are used for the establishment of cyber security test platforms and various cyber attack practices to ensure organisational activities [4]. As nowadays almost all organisations become aware of their data securities, it should be necessary to continue the entire research work by detecting and understanding both security vulnerabilities as well as weaknesses in various source codes.

II. Literature Review

a) Significance of Cyber Security and Testing Platforms

Businesses across all sectors have been experiencing an increase in threats in the platforms of cybersecurity for the last few years. In 2022, most numbers of cybersecurity companies have seen the highest development in cyber attacks. According to [8], more than half of business companies in the country have reported breaches in cybersecurity in one year. Today, business companies can only conduct business with the involvement of hackers. The nature of attacks on cyber platforms has changed drastically in a few years, the percentage of malware practices has decreased, and phishing numbers increased to more than 85%. Business organisations across the globe have tried to implement cyber security to protect computers, mobile phones, servers and networks from malicious intent attacks. It is essential to implement protective measures to protect the systems and important information of the business. After the application of GDPR, it is important to cover the personal data of the companies and their employees. Some components of cyber security have been designed to strike the cyber attackers early, although cybersecurity professionals today are keen on defending the assets of the companies at first. It has been utilised as the process to protect everyone from cybercrime, and it can provide help from finding theft to identifying threats at the international level.



Source: [9]

Figure 1: Cyber Security Testing

As per the opinion of [9], a breach in the security of the servers can expose the personal and essential information of companies across the wo; It is considered a serious issue and has a strong impact on the financial conditions of the companies. Cybersecurity is very much essential to protect the business operations of companies in the time of globalisation and digital technology. It encompasses several technologies and approaches to protect servers, official and personal data, and computer systems from various cyber-attackers. There are some subdomains of cybersecurity, such as application security, cloud security, Data security, mobile security and Network security. As per [10], application security helps the server to implement defences that are different and put them into the software of the organisation to protect the server from a range of threats. To implement this application successfully, it needs a cybersecurity expert to assess secure code, design the application securely and apply full information to reduce the rate of unauthorised access to the server. Cloud security helps companies to secure their servers by creating an architecture of the cloud; several service providers of cloud systems utilise this application, such as AWS, Google and Azure. The subdomain of data security helps companies to maintain authentication protocols that can be two or multi-factor. Mobile security is considered essential to the new generation, and this security application protects personal and official information gathered on the mobile device and guides them from unauthorised access, loss of device and virus attacks.

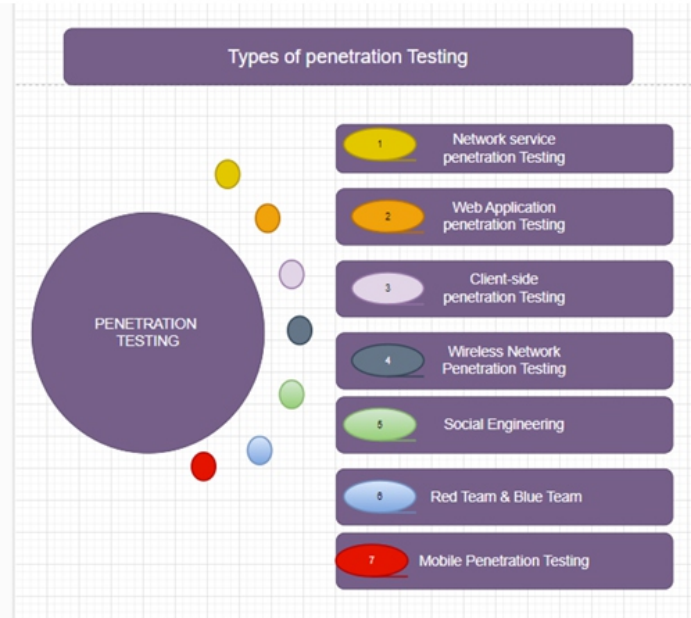
Three recognised examples of cybercrime are crimes that are assisted by the puters, when hackers get into the system of a computer and where computers are used incidentally. There are several kinds of cyber threats; Malware attacks, trojan attacks, cyber-terrorism, SQL injection, Phishing and Service denial. Companies need to use some essential software testing to prevent these cyber attacks; penetration testing, security testing, usability testing, configuration testing, SAAS and fuzzing are considered significant testing software that can prevent cyberattacks.

According to [11], the system of penetration testing is usually considered a pre-planned attack against the infrastructure of Information and technology, website and applications of several companies. It is essential to provide real-time experience to the business management employees and the hackers' working process tools. Security testing is also necessary for every stage of the software development process, and it helps to contain security vulnerabilities and high turnover rates. Usability testing is also essential at the time of developing products of the company, such as new websites and applications of mobile devices of IoT. It helps gain more customer base as they can understand the effects and efficiency. A business server must not be hacked at the time of conducting business, and the application of cloud computing, such as SAAS and IAAS, is important to the company as it is an advanced technology.

Companies use advanced and late applications of this cloud computing software to ignore vulnerabilities. By utilising software testing in cyber security, business organisations can develop more secure systems, and it is essential to prevent online threats.

b) Essentiality of Cyber Security Test Platforms aand Cyber Attack Practice to Prevent Cyber Attacks

Penetration testing forms are considered to be an essential part of assessing the risk of security for all businesses, rectifying the clear defects and eliminating the subtle susceptibility from the perspective of hackers. Besides this, the cyber attacks practice is considered to be practice to defend the servers, computers, electronic systems, data and networks from malicious attacks. Here from the opinion of the researchers [12], a lot of research effort has been conducted to develop the cyber-security of the smart grids by utilising various kinds of techniques. The current power systems consist of the generations and sensors that give permission to two-way communication with the infrastructure of the system with reliable energy production via the combination of "Distributed Energy Resources (DERs)" and "Advanced Metering Infrastructure (AMI)". This complicated communication system bears major benefits; by developing reliability, manageability and energy efficiency, it creates the vulnerabilities of the system to cyber attacks for the huge numbers of access points and devices that do its operation outside the administrative domain considered to be traditional. Since the power grid can lead to disastrous events, it is optional to research the effects or consequences of cyber attacks on the power system.



Source: [13]

Figure 2: Penetration Testing

From the opinion of some authors [13], in North American blackouts, the lack of system awareness is considered to be the main reason behind the blackouts, which highlights the essentiality of the analysis of cyberattacks in terms of maintaining a reliable and stable power supply operation. The cyber attack could damage or destroy the equipment or request false demands that might result in a huge rate of energy generated. Additionally, the spiteful attack also bears the dangerous capability of causing false negatives or a condition that is a wrong overload in the power system. Another disruption is also running the potential conduction in the various parts of the smart grid and electric vehicle infrastructure. Spiteful attacks can stop the services in the substation computers by obstructing communications with the device. The real-time detection of cyber attacks is supreme for the authentic performance of the vital infrastructure involving smart grids. Constant and online system observation is needed to detect the cyber attacks that have been targeted to see and gain attack pliability. The individual sensors in a wide-scale network are considered to be the primary target of security understanding. It is possible for the compromised insider to access the data stored easily in a compromised confluence. In theory, the key cancellation of the compromised node is possible by the application of a proper or authentic mechanism to the sensor network. However, the approach of authentication on the basis of security gateway structure or cryptography could be more practical for the storage constraints and computation of the system.

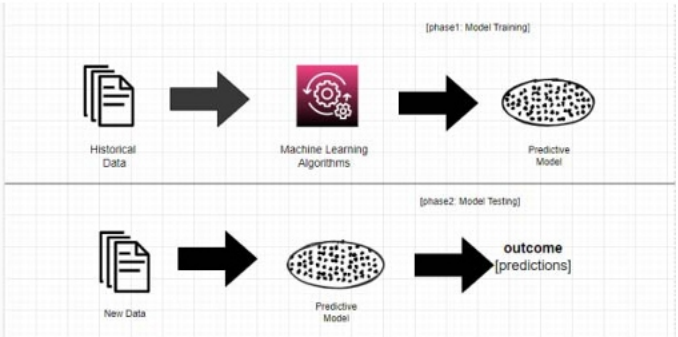
According to the opinion of another researcher [14], it is optimised that the techniques of advanced anomaly detection and security control theories on the basis of various methods of state estimation are very capable of immunising the power system where the major part of this is physically impractical, mathematically expensive and unscalable for the network which is complicated in a large-scale. In the present day, a large amount of information is produced on all of the grids that develop the entrance ability for the monitoring of the realtime system. The historical information describes the operation of the system that bears the capability to rectify the possible and anomalies attacks. Although the traditional techniques of "Bad Data Detection (BDD)" are not ready for the purpose of real-time computation, and the difficulties related to storing the great volume of the information generated in the smart grid. Such kind of difficulties enlarges the potentiality of the utilisation of techniques of data analysis, like ML, in terms of handling the data set that is structured in a complicated way with Artificial Intelligence in terms of preventing and detecting cyber attacks. Here from the opinion of other researchers [15], ML algorithms are possible to use in evaluating different types of measurement combinations via states, AMI and control actions by understanding their structures of them, where they can detect the "False Data Injection (FDI)" attack by understanding the non-linear and complicated connection among the measurements. Several ML algorithms are compared and tested for the matter of detecting the FDI attacks, where machine learning has got success in classifying the attacks related to FDI.

A method of hybrid intrusion detection has been suggested on the basis of a process of common path miming in terms of detecting the unusual power system events from the PMU relays, information and energy management system logs. Additionally, the techniques of cyber attack detection on the basis of a correlation between the two parameters of PMU utilising the Pearson correlation coefficient have also been suggested. Such methods evaluated the transformation of correlation between the two parameters using the Pearson correlation coefficient.

c) Present Trends of the Modern Cyber Security Activities, Measurements and Techniques

Automation has developed its essentiality in the matter of cyber security. The Automated procedures of security bear the capability to decrease the time that it has taken to give a response and detect threats and develop the exactness to detect threats. Automation has also reduced the dependency on manual procedures that can be prone to human error and time-consuming. Here according to [16], in the present day, the Fourth Industrial Revolution is famous as 4.0, which visualises the rapid change in industries, procedures, social patterns and technology as an outcome of developed smart automation and interconnectivity. This type of revolution has influenced most industries all over the world and caused an enormous transformation in a manner that is non-linear at an unrivalled rate, with the inference for all the economies, industries and disciplines. Industry 4.0 has been described as a term that is utilised to define the current trend of the industrial exchange of data and technology automation, which involves the Internet of Things, cognitive computing and cyber-physical systems with the improvement of the smart factory. The start of the digital revolution to Industry 4.0 has taken place with data gathering, obeyed by the AI in terms of interpreting the information. So, the "intelligence revolution" is able to be considered in the matter of servicing and computing, as AI has reshaped the world that includes intelligence and human behaviour into systems or machines.

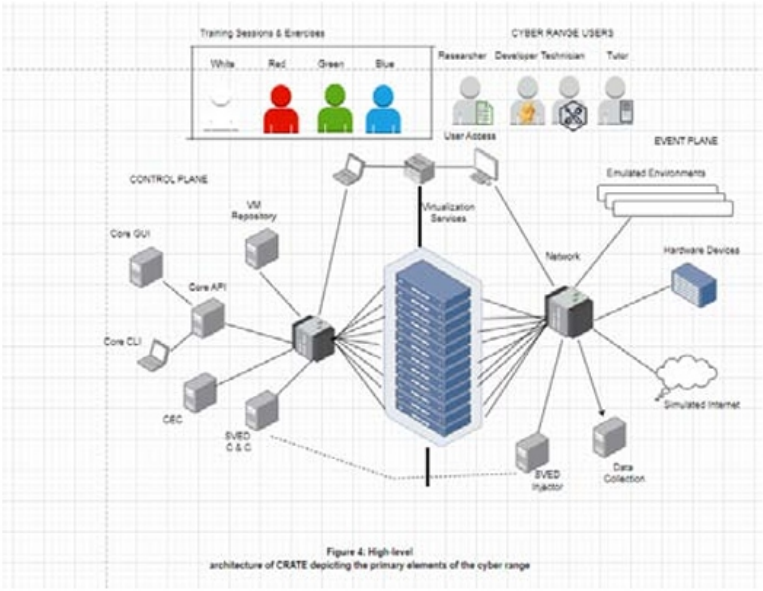
From the opinion of [17], in the present days, machine learning modelling has been applied in a practical way, especially in the matter of cyber security.



Source: [17]

Figure 3: General Structure of the ML-Based Predictive Model Considering both the Training and Testing Phase

For instance, the application of the ML strategy in order to get the covid 19 assistance to the people who actually need it. Several cyber-attacks and anomalies have the chance in terms of being detected by utilising the approaches of machine learning in the part of cyber security. Additionally, the strategy based on ML has the ability to improve an effective smart parking system for the environments of smart cities. Besides this, AI is considered the buzzword as it has prepared to influence businesses of all sizes and shapes in all industries. The AI of the sector can develop the available services or products to make all these more safe, reliable and effective.



Source: [18]

Figure 4: High-level Architecture of CRATE Depicting the Primary Elements of the Cyber Range

The above figure shows a high level of CARTE's architecture, with the servers considered to be visualisation that abode the imitated environments in the centre part. On the left side, the control plane is used to manage the cyber range, and on the right side, the event plane is utilised for the system where the execution of the experiments or research and training is conducted. The plane planes are depicted as two zones of security, which are separated from one another, which is important in terms of ensuring that the execution of the events in the event plan has not affected the control plane.

The server for virtualisation gives abode to the virtual machine utilised in the imitated environment. Currently, there are approximately 500 virtualisation servers existing in CARTE, where the virtualisation servers generally operate a customised, tiny operating system on the basis of Linux that is renowned by the name CarteOS. In order to facilitate the maintenance of cyber range and ensure the honour of the servers, CarteOS operate in a read-only domain and cover the file systems that are utilised in storing the configuration and virtual machine.This has enabled the server's operating system to be

replaced without impacting or affecting the organised virtual machine or its composition, permitting CarteOS to be upgraded as the latest software versions and updates regarding security become available.

d) Cyber Security Testing Tools and their Usage

Since the beginning of 2020, organisations across the world have been facing several cybersecurity problems. Ransomware attack rates have increased by more than 140% after the pandemic. Companies have hired many cybersecurity analysts in business management to assess security-related issues, and they are responsible for reporting any security breaches and evaluating the servers' weaknesses of the respective companies. Several types of cyber security tools have been used to find any vulnerabilities in the web applications and servers of the companies. According to [19], cyber security tools can enhance the possibility of identifying threats to servers. Some important cyber security testing tools are;

Burp Suite is a well-known software, and it is considered one of the best toolboxes that can provide testing of web security. The application of this tool is designed to use by click with a point process. It is a graphical tool, and it works to conduct security testing on any online application. The application of this tool helps the entire process of testing from the mapping of the initials, and it can analyse the attack surface of an application by discovering any flaws of security in the application. It is a security solution for web applications, and it helps companies to test any vulnerability manually, and it also helps to Intercept messages of HTTP. Burp Suite is used to conduct several activities, such as trying a web application, web crawling and web application analysis. This tool can be built into the browser of Chrome.

Vega is a web security scanner and a web security testing platform, and it helps to test web application securities. The application of Vega helps to identify any SQL injection and also other vulnerabilities in the server of any company. It helps to find cross-site scripting which can be reflected or stored. As per [20], The application of Vega provides TLS security settings and sees all the possible opportunities to enhance the security of the TLS servers. It has an automated scanner that helps to test the servers quickly and can intercept proxy servers at the time of tactical inspection. This application can be updated by utilising the application of artificial intelligence in the javascript language.

OpenVas software is a vulnerability scanner, and it can be helpful in conducting unauthenticated testing and authenticated testing of several industrial protocols. According to [21], It also can help to improve the tuning of performance at the time of scanning large scales and provide the language of internal programming to conduct vulnerability of any type. This application has been used for many years and is a process that can find any vulnerabilities in the servers. It classifies the system resources and allocates

all the enumerable values. Then it detects all the probable threats and reduces the vulnerabilities by giving proper priorities.

The intruder is an essential vulnerability scanner to prevent the issues of cyber security and identify any weaknesses in the system servers. This tool helps to save time as it proactively scans any new threats in the server and offers an interpretation system to identify all the unique threats. This tool's main positive aspect is the support staff's quality. It has a chat app that can ease various quotations, and the device has the comprehensiveness of all the outputs. It helps to identify all the vulnerabilities, and it takes several actions to fix them.

According to [22], Zed Attack proxy is a great tool for analysing static code, and cybersecurity companies have used it to detect all the security problems in principle; it helps to fix the issues of vulnerability. This tool can highlight several suspicious codes that have developed in the server system, and it provides feedback on security during the review of the code. It also can identify several technical debts and fix the vulnerabilities in the application in the code. It can detect bugs faster and give feedback to the developers to enhance the quality of the code.

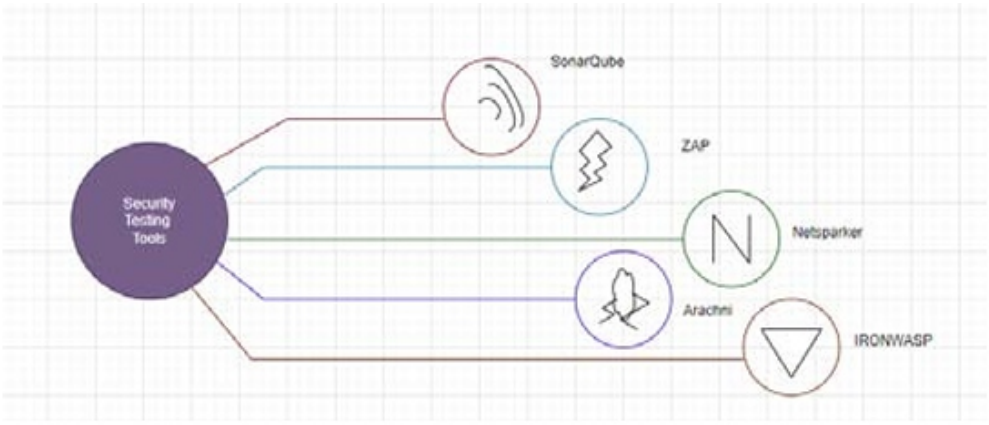


Figure 5: Security Testing Tools

III. RESEARCH Methodology

a) Research Overview

Research methodology basically follows the measurements of the research processes, and it perfectly channels both the identification and completion of possible important areas which have been considered in the proposed research topic. The methodology of this research work has been discussed here by the researchers to implement specific data and continue the flow of narration of this research.

The researcher has continued this research by following proper research philosophy, research approach, research design, data collection processes, and data analysis process [23]. Different techniques and tools that are used in this research are also proposed in this methodology section. The researcher in this research follows “the positivism research philosophy”, “the deductive research approach”, “the descriptive research design”, “the secondary data collection processes”, and “the qualitative data analysis processes" to continue the research work. So, the research project is completely organised through sequential processes that are defined on behalf of the scope of the project, research method, and analysis of all the collected data.

b) Research Methods

The method in this research work has been followed in the vicinity of the mixed research method because both primary as well as secondary data have been used to carry on the research work. All the information that has been collected is the secondary qualitative data. This secondary methodology has helped the researchers to accumulate, classify, and evaluate all the published articles which will be available from various internet resources and libraries. The secondary research proposes certain questions as well as focuses on some hypotheses [23]. This research topic is based on the establishment of cyber security test platforms and cyberattack practices, which perform with the assistance of internet connectivity. The secondary research work relates to internet connectivity, and the data analysis process points out the critical viewpoints used in security implementation measurements. The entire research method highlights the specific orientation of the current research issues.

c) Research Philosophy

The researcher in this research work follows the “positivism research philosophy" are not, and it will propose a clear, brief, and concise discussion that does not use any kinds of descriptive stories. Any interpretation is not allowed because of its value-free nature. Some common theories and basic concepts are applied based on the research objects. Nowadays, cyber security attacks are increasing day by day, and it covers a range of situations within very short periods. The main concept for which the researcher carries on their knowledge consists of genuine decisions [23]. The key feature of this positivist research philosophy is to use clear, brief, and concise discussion, which does not utilise any descriptive stories. It dismisses an individual's importance which proposes subjective values and experiences. Finally, positivist research philosophy ensures that researchers make perfect predictions based on both social and society-based changes. Positivism basically holds the idea that empiricists observe natural processes. The basic characteristics of positivism are to propose valid knowledge and identify the facts of the collected information. So the strength of this positivist research philosophy is to be a pioneer in the first scientific study of the proposed topic.

d) Research Approach

The “deductive research approach" has been used to highlight the procedures that the researcher selects to analyse, collect, as well as interpret the data. It helps the researchers to determine the success behind the research work and maintain the overall standards of the research. This research approach has been used to support the researchers to remain confirmed about the existing theories. The deductive research approach proposes the possibility of delineating casual relationships in the middle of variables and concepts. In the case of qualitative research work, the researcher applies the theory with a "top-down approach" for analysing the collected data. It basically helps to continue the research works from general to more specific [24]. The benefit of this kind of deductive approach is to explain the variables and concepts which are interrelated with both causes and effects of the research. It also helps to measure both concepts and ideas of the research work that are possibly reached to a broader extent.

e) Research Design

Research design is basically the blueprint of the entire research process. The researcher in this research study follows the “descriptive research design" to point out and address all the possible issues which may arise during the research and data analysis processes. The proposed research design is basically a type of research design that focuses on obtaining any systematic information to describe a situation, phenomenon, or population. This descriptive research design provides permission to the researchers to explain and learn the value of more variables in the absence of any casual and valuable hypotheses. The researcher proposes this research design with the aim of systematically and accurately explaining the situation of the current research work [24]. The main purpose behind this research work is to define, describe, and validate the findings of the research works, which helps both the researchers and future readers of this research work to obtain a focused description of the current phenomena along with proper analysis and interpretation of the research findings.

f) Research Data Collection

For this research study, the mixed research approach has been utilised, and for that reason, both primary and secondary data have been employed. The main research has yet to be evaluated as the essential part of obtaining innovative data; rather, the narration has been followed via a secondary literature review, and primary data will be analysing those extracted parts. For gathering secondary data, articles have been chosen from various secondary resources, concluding IEEE Xplore, google scholar, and with this other internet resources. Research topic-based suitable keywords have been used so that it can be easy to obtain relevant information and provide proper justification based on the cyber security-based platform development [25].

The utilised keywords have been chosen, such as cyberattack, platform, cyber threats and so on. The articles published before 2019 have yet to be considered relevant for this research process. As technological innovations are constantly developing, due to that reason, to provide current information, it is important to obtain current data also. The citation index of every research article was measured properly to ensure research credibility. The strategies for the primary research approach have been analysed by utilising different models that support the prevention of cyber attack activities. The main purpose of operating the secondary data collection process

g) Used Tools and Techniques

The model has been utilised in simulating the power grid utility in terms of tools and techniques. As the power system simulator, it will help in creating the simulation environment in constructing the models through flow cases of the power system. For the graphical user interface, the RSCAD can be used in developing the power system models with the help of a simulator. Within the submission level, these IEDs can communicate with the RTDS using digital inputs. Referring to the "IEC 61850 GOOSE protocols", the RTAC process can be found with SCADA measurements. This RTDS can also communicate with the control servers in compiling the DNP3 and IEC 61850 protocols [30]. This RTDS also can be interfaced along with the substation control while using the hardwired connections. Therefore the ethernet connection also has been used in managing the hardware or similar type of communication also.

h) Research Data Analysis

For a research process, the data analysis technique is an essential part that should be followed properly stepwise. The researcher follows a "qualitative data analysis" process to evaluate all the collected data. It has come to know that this type of research data analysis process In choosing the secondary data analysis process, the related testing results in terms of cyber security also have been compared. The main purpose of data visualisation is to depict the observation result properly through various graphs, charts, and with these other types of visualisation tools. This ISSAC setup also delivers the SCADA network within the enterprise level along with the computing nodes [29]. Nevertheless, this ISAAC has been used in simulating organisational models consisting of CPS. Similarly, connectivity can also be made between branch campuses and research laboratories.

Comparison of 5 to 6 Research Papers

Citation	Title	Results
(Khandkeret al., 2021)	Cybersecurity Attacks on Software Logic and Error Handling Within ADS-BImplementations: System-atic Testing of Resilience and Countermeasures	In this research paper, the concern has been laid on detailing the test platform and attack along with the utilization and experimental set reflected in the result. In the process of experiments, 36 varied ADS-B. In combination with host, hardware and software. Even around 2107 test samples were accumulated, among them, 966 of which were actual aeroplanes while 11141 were spoofed aeroplanes of attackers. There was a clear evaluation of the high-power attacks that were much easier to detect. On the other hand, low-power ones were critical to being detected and even erroneously prone.
(Oyewumiet al., 2019)	ISAAC: The Idaho CPS Smart Grid Cybersecurity Testbed	In the paper, the concern has been laid on utilizing the ISAAC's SCADA visualization, along with the cybersecurity abilities, to form the experimental results. The experiments manage the evaluation of the network information and incorporate accumulation packet stream via ISAACs interaction channel at a DoS attack. This experiment leads to the ML framework for data-related health monitoring. The result reflected the process of developing resilience and threat assessment of CPS, detecting stealth cyber attacks against state removal as well as application. In the process, a dignitary visible wall-mounted display has been implemented within the "Power lab-tested firm" while utilizing the "IRIG-B" synchronization" digital clock having SEL-3401. This tends to give time with an accuracy rate of around ±100ns. The result outlined the current use of ISAAC; when significantly integrated, ISAAC will form CPS research as well as the educational capability of the regions around Idaho. Idaho CPS Smart Grid Cybersecurity Testbed of surrounding that emulated the strength utility.
(Ramirez et al., 2023)	PLC Cybersecurity Test Platform Establishment and Cyberattack Practice †	The "PLC Cybersecurity Test Platform" has been analyzed in this research paper. In the test platform, different cybersecurity tools are utilized. "Personal computer running Kali 2022.3" as kernel operating system, which plays like an attacker, and with this ", a personal computer running Ubuntu 22.04" is utilized as the target device. The target device races *ModbusPal v1.6* for stimulating Modbus communication. Modbus utilizes port 502 for communication, which can be a target for attack exploitation. To identify the Modbus register, Metaploit has the capability to provide requests on individual addresses. Metasploit utilization supports register modification in the chosen target.
(Kim et al., 2019)	Cyber-Physical Battlefield Platform for Large-Scale Cybersecurity Exercises	In this research paper, *a cyber-physical battlefield* (CPB) platform has been developed that can provide scalability in cybersecurity exercises. For developing the platform, it is essential to conduct an on-site visit to gain better information about the security threats, as well as the working phenomenon of the individual sectors. In operation, CPB can stimulate ICS/SCADA system. This platform's successful application within "Locked Shields 2018" (LS18) provides a shred of strong evidence.
(Munaiahet al., 2019)	Characterizing Attacker Behavior in a Cybersecurity Penetration Testing Competition	For measuring attacker mindset, proper security software should be developed. In this research paper, a multimodal dataset has been chosen during "the 2018 National Collegiate Penetration Testing Competition" to understand the attacker's mind. *MITRE ATT & CK* framework is utilized to codify tactics, as well as techniques. Attackers applied various unregistered accesses to handle the user's account. Through the proposed framework, at first, it can be easy to identify ATT&CK's tactics for decreasing attack numbers.

I) Data Validity and Reliability

The procedure of the research has been related to the mixed approach to the performance of the study. The researched information has been separated into two kinds of efforts such as primary and secondary. The secondary literature review was established on the basis of the journals of Google scholars. The

researchers have healthily ignored the store of available data and intentional incorporation the data in terms of getting advantages about the research efforts. The dependency of the developed elements is possible to be justified by the efficiency of the considered datasets. The effectiveness of the model could be improved by nourishing a higher amount of information to the logical system within the model of the device. The utilisation of the classification logic has allowed a huge amount of data to actuate and streamline the model into the perfect procedures of detection.

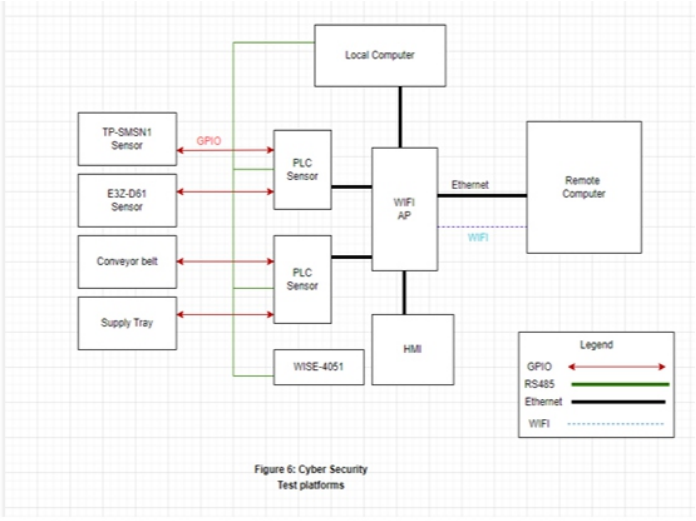
j) Research Limitation

The procedure of the research has followed an approach of mixed methodology, where it collects its own limitations by gathering the previous hindrances and limitations underlined or highlighted in the articles that are already published. The restrictions or regulations of the secondary literature review are highlighted in the absence of the statistical establishment. The procedure of the ideological similarities to the articles that are published has been erased by the application and establishment of the primary research methodology. The strategy of the primary research has been lacking the statistical justifications steps to describe the efficacy of the model that has been developed, and the process of development and rectification of the types of threats from the dataset have been executed to give the development effort incremental success a justification.

IV. Result & Comparison

For testing the different kinds of cyber attacks, the ISAAC facilities have been found to develop the capacity by developing both realistic and practical CPS. Through the potential utilisation, the real-time simulation of cyber attacks has been determined through RADICAL. The different researching areas of modern networks and computing platforms have facilitated it. By delivering a contained and secure environment, it has been used for securing critical infrastructure and experimental analysis also. It has been used in conducting sophisticated cyber attacks by strengthening the necessary infrastructure. in terms of visualisation, the SCANVILLE has been used in delivering real-time data analyses through a total ISAAC testbed [28]. For emulating real-world enterprise, this SCANVILLE has been found as important for infrastructure utility. Therefore, this data visualisation can also be used for trend identification, monitoring the overall system and detecting real-time attacks. This can be assessed in regulating the violations by simulating threats and negative incidents along with their happenings. ISSAC has been found in the delivery of realistic emulation environments in case of comparative validation and testing. Combined with the multiple CPS research approaches, it has been used in investigating vulnerabilities and assessing and exploiting their impact.

Emulating the SCADA network has also been used in facilitating the experimental environment, which has been used within the cyber-defence training curriculum. Regarding remote utilisation, the ISAAC testbed can be used to expand the completed designs and plans across the State of Idaho. In connection with this, the OSI layer two can be referred to as the Tunneling protocol of the Idaho Regional Optical Network (IRON). Integrating the IRON and VLAN has helped enable the testing through the growth of the additional laboratories. In terms of CIA confidentiality, integrity and availability have been installed within the ISAAC network, which has played an important role against cyber attacks.

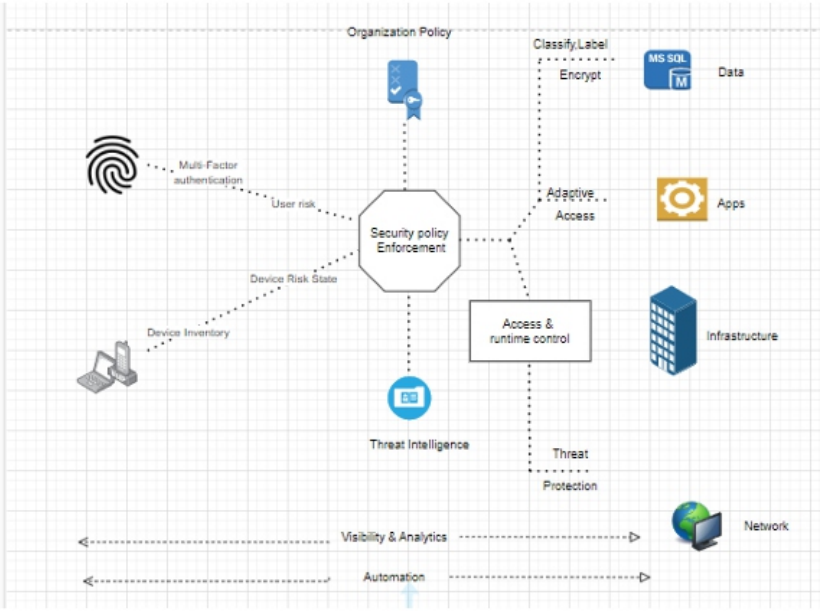


Source: [28]

Figure 6: Cyber Security Test Platforms

This ISAAC network might also be classified into subtypes to create a digital-level defence. Both the DeMilitarized and VLANs have been found to have stringent access control policies for developing performance and security. Hence each of the VLANs has been found to have explicitly- defined access control. Nevertheless, for delivering perimeter-level defence, the ISAAC firewalls also can be used to prevent direct ingress and egress connections between the external networks. In defending both infiltration and exfiltration, the configured control list can be shared [26]. Within the reverse proxy mode, the statistical engine features and web proxy can be refereed with restricted and controlled access. This web proxy can eliminate the risk of direct exposure of ISSAC to the internet. The IDS has been used for intrusion detection to integrate the Switched Port Analyzer and port monitoring. By using the NIDS, an authorised network might be reviewed in revalidating the experimentation. In the case of node re-imagination, virtual machine endpoints and instances can be combined along with the operation system. Through maintaining the modified security updates, a web proxy server can be used within the infrastructure of PoT, SCANVILLE and RADICAL.

In contrast to this, there are multiple studies also have been found on the securing of the DAS- B security. Considering the different studies, this can be categorised between two kinds of studies: broadcast authentication and localization verification. The RF communication defences have been explored in measuring the effectiveness of the PLS techniques. Referring to the RSS- Distance model, the signal attenuates in travelling through space. To distinguish the real aircraft, the spoofing unit has been set up through random transmission of the fake "ADS-B 1090ES" signals by encoding the random positions. Letting the spoofing set up, the receiver has been found to receive both the real and spoofed calls. For a defined ADS- B massage, a setup also has been calculated in calculating the 3D distance between the receiver and the aircraft. In case the real-time RSS and retrieved RSS have been as close enough, then only the aircraft can be considered legitimate. Regarding this, the RFF has been found to suffer from both fluctuations and noise. According to the tolerance level, the attacks are found in using multiple power levels, which are medium power attacks, low power attacks and high power attacks [27]. For defending, the Doppler shift can be used in measuring the frequency wave motion between the receiver and transmitter.



Source: [18]

Figure 6: Cyber Security Statistics

Additionally, the Doppler shift can be added with an ADS-B signal for verification of the velocity along with the aircraft position. For the coordinated attack types, the ADS- B messages can be found with the bodies such as FAA, RTCA and ICAO. For making defences against other types of attacks, effective software can be detected through data fluctuation. In implementing the developed logic for alerts, the above notification can occur through aerospace- arrived ways of handling and notifying of alerts. Configuring the signals can be displayed by signals can be displayed by displaying the threshold and delivering the sensible defaults. [7].

V. Conclusion

In the present day, most organisations are facing a lot of threats due to the dangerous effects of cyber attacks, where such threats have strengthened the potentiality of losing or misplacing vital and confidential organisational information. So, with a major focus on such difficulties, the paper has done its best to describe the importance of cyber security testing and practising cyber attacks. The purpose of the cyber security testing utilised several tactics and methodologies in order to measure the effectiveness of the cyber security strategy against the possible risk that can be faced by the security system of the systems. Here the paper has identified the vital vulnerabilities that have been utilised in the industry and organisation in an active way in terms of launching cyber attacks. The report has discussed the significance of the automation system, giving reference to ML, AI and other methods and techniques. ML is estimated to provide support in analysing and predicting dangerous activities like malware, phishing, authentication attacks, application attacks and so on; while considering this, several companies have improved their systems with the implications of machine learning. The integration of AI has the ability to attack the way to test cyber security attacks as a mandatory process to utilise the cyber security tools. As technology is becoming regularly updated, the establishment of AI-based cyber-attack platforms also needs to be periodically updated, and for this reason, more future work on this topic has been required to point out the roles and significance of future research works.

Besides this, the importance of AI in the current cyber security testing and cyber attack practice is small. AI in cyber security has erased the tasks that are meant to be time-consuming; here, by scanning information, rectifying the possible threats and decreasing the false positives to extract the non-threatening activities, AI technologies are supporting organisations. The paper has reflected the usage of AI and ML technologies in such a way that it has got evidence of the expertise of the technologies in terms of focusing on the more vital or critical tasks. However, the paper has found it quite difficult to fix the number of environments that will be run simultaneously due to its dependency on size and organisational complexity to be emulated. So, in future research of the paper, this matter will get a concentration while researching and discussing the topic or study.

ACKNOWLEDGMENT

I would prefer to express my absolute gratitude towards my professor, for the valuable advice, supervision and guidance from early phase of this research. I am extremely grateful for attainment of the constant support throughout the research. I am additionally thankful to all the participants those contributed in several ways towards this research, also humbly acknowledge all contribution.

REFERENCES

1. Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2023). Towards Insighting Cybersecurity for Healthcare domains: A comprehensive review of current practices and trends. Cyber Security and Applications, 100016.

2. Ahmadi, A., Nabipour, M., Taheri, S., Mohammadi Ivatloo, B., &Vahidi Nasab, V. (2022). A new false data injection attack detection model for cyberattack resilient energy forecasting. IEEE Transactions on Industrial Informatics, 19(1), 371381.

3. Sun, C. C., Cardenas, D. J. S., Hahn, A., & Liu, C. C. (2020). Intrusion detection for cybersecurity of smart metres. IEEE Transactions on Smart Grid, 12(1), 612622.

4. Gasiba, T., Lechner, U., Pinto-Albuquerque, M., & Porwal, A. (2020). Cybersecurity awareness platform with a virtual coach and automated challenge assessment. In Computer Security: ESORICS 2020 International Workshops, CyberICPS, SECPRE, and ADIoT, Guildford, UK, September 14–18, 2020, Revised Selected Papers 6 (pp. 67-83). Springer International Publishing.

5. Karagiannis, S., Maragkos-Bel Maps, E., & Magkos, E. (2020). An analysis and evaluation of open source capture the flag platforms as cybersecurity e-learning tools. In Information Security Education. Information Security in Action: 13th IFIP WG 11.8 World Conference, WISE 13, Maribor, Slovenia, September 21–23, 2020, Proceedings 13 (pp. 6177). Springer International Publishing.

6. Ukwandu, E., Farah, M. A. B., Hindy, H., Brosset, D., Kavallieros, D., Atkinson, R., ... & Bellekens, X. (2020). A review of cyber-ranges and test-beds: Current and future trends. Sensors, 20(24), 7148.

7. Aldawood, H., & Skinner, G. (2019, January). An academic review of current industrial and commercial cyber security social engineering solutions. In Proceedings of the 3rd International Conference on Cryptography, Security and Privacy (pp. 110-115).

8. Hanif, Y., & Lallie, H. S. (2021). Security factors on the intention to use mobile banking applications in the UK older generation (55+). A mixed-method study using modified UTAUT and MTAM-with perceived cyber security, risk, and trust. Technology in Society, 67, 101693.

9. Renaud, K., & Ophoff, J. (2021). A cyber situational awareness model to predict the implementation of cyber security controls and precautions by SMEs. Organisational Cybersecurity Journal: Practice, Process and People, 1(1), 24-46.

10.Muthuppalaniappan, M., & Stevenson, K. (2021). Healthcare cyber-attacks and the COVID-19 pandemic: an urgent threat to global health. International Journal for Quality in Health Care, 33(1), mzaa117.

11. Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola,A., Epiphaniou, G., Maple, C., & Bellekens, X. (2021). Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Computers & security, 105, 102248.

12. Karimipour, H., Dehghantanha, A., Parizi, R. M., Choo, K. K. R., & Leung, H. (2019). A deep and scalable unsupervised machine learning system for cyber-attack detection in large-scale smart grids. IEEE Access, 7, 80778-80788.

13.Nifakos, S., Chandramouli, K., Nikolaou, C. K., Papachristou, P., Koch, S., Panaousis, E., & Bonacina, S. (2021). Influence of human factors on cyber security within healthcare organisations: A systematic review. Sensors, 21(15), 5119.

14.Ukwandu, E., Farah, M. A. B., Hindy, H., Brosset, D., Kavallieros, D., Atkinson, R., ... & Bellekens, X. (2020). A review of cyber-ranges and test-beds: Current and future trends. Sensors, 20(24), 7148.

15. Awamleh, A. M. A., Alorfi, A. S. M., Al-Gasawneh, J. A., & Al-Rawashdeh, G. (2020). Cyber security and ethical hacking: The importance of protecting user data. Solid State Technology, 63(5), 7894-7899.

16. Stevens, T. (2020). Knowledge in the grey zone: AI and cybersecurity. Digital War, 1, 164-170.

17. Sarker, I. H. (2022). Ai-based modelling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science, 3(2), 158.

18.Gustafsson, T., & Almroth, J. (2021, March). Cyber range automation overview with a case study of CRATE. In Secure IT Systems: 25th Nordic Conference, NordSec 2020, Virtual Event, November 23–24, 2020, Proceedings (pp. 192-209). Cham: Springer International Publishing.

19. Thaqi, R., Vishi, K., & Rexha, B. (2022). Enhancing Burp Suite with Machine Learning Extension for Vulnerability Assessment of Web Applications. Journal of Applied Security Research,1-19.

20. Altulaihan, E. A., Alismail, A., & Frikha, M. (2023). A Survey on Web Application Penetration Testing. Electronics, 12(5), 1229.

21. Ardo, A. A., Bass, J. M., & Gaber, T. (2022, February). An empirical investigation of agile information systems development for cybersecurity. In Information Systems: 18th European, Mediterranean, and Middle Eastern Conference, EMCIS 2021, Virtual Event, December 8–9, 2021, Proceedings (pp. 567-581). Cham: Springer International Publishing.

22. Lallie, H. S., Shepherd, L. A., Nurse, J. R., Erola, A., Epiphaniou, G., Maple, C., & Bellekens, X. (2021). Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic. Computers & security, 105, 102248.

23. Newman, M., & Gough, D. (2020). Systematic reviews in educational research: Methodology, perspectives and application. Systematic reviews in educational research: Methodology, perspectives and application, 3-22.

24. Ryder, C., Mackean, T., Coombs, J., Williams, H., Hunter, K., Holland, A. J., & Ivers, R. Q. (2020). Indigenous research methodology–weaving a research interface. International Journal of Social Research Methodology, 23(3), 255-267.

25. Hafidz, M. A., & Elihami, E. (2021). Learning The Nonformal Education Through Research Methodology: A Literature Review. Jurnal Edukasi Nonformal, 2(1), 47-55.

26. Fowler, D.S., Bryans, J., Cheah, M., Wooderson, P. and Shaikh, S.A., 2019, July. A method for constructing automotive cybersecurity tests, a CAN fuzz testing example. In 2019 IEEE 19th International Conference on Software Quality, Reliability and Security Companion (QRS-C) (pp. 1-8). IEEE.

27. Oyewumi, I.A., Jillepalli, A.A., Richardson, P., Ashrafuzzaman, M., Johnson, B.K., Chakhchoukh, Y., Haney, M.A., Sheldon, F.T. and de Leon, D.C., 2019, February. Isaac: The idaho cps smart grid cybersecurity testbed. In 2019 IEEE Texas Power and Energy Conference (TPEC) (pp. 1-6). IEEE.

28. Khandker, S., Turtiainen, H., Costin, A. and Hämäläinen, T., 2021. Cybersecurity attacks on software logic and error handling within ADS-B implementations: Systematic testing of resilience and countermeasures. IEEE Transactions on Aerospace and Electronic Systems, 58(4), pp.27022719.

29. Munaiah, N., Rahman, A., Pelletier, J., Williams, L. and Meneely, A., 2019, September. Characterizing attacker behavior in a cybersecurity penetration testing competition. In 2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 1-6). IEEE.

30. Matheu-García, S.N., Hernández-Ramos, J.L., Skarmeta, A.F. and Baldini, G., 2019. Risk-based automated assessment and testing for the cybersecurity certification and labelling of IoT devices. Computer Standards & Interfaces, 62, pp.64-83.

31. Ramirez, R., Chang, C.K. and Liang, S.H., 2023. PLC Cybersecurity Test Platform Establishment and Cyberattack Practice. Electronics, 12(5), p.1195.

32. Kim, J., Kim, K. and Jang, M., 2019, May. Cyber physical battlefield platform for large-scale cybersecurity exercises. In 2019 11th International Conference on Cyber Conflict (CyCon) (Vol. 900, pp. 1-19). IEEE.

Market Basket Analysis using Machine Learning

Atul Sharma α, Dr. Mohammad Salim Hamidi σ & Yousuf Hotakρ

A B S T R A C T

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. Itsupports 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]

```
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]

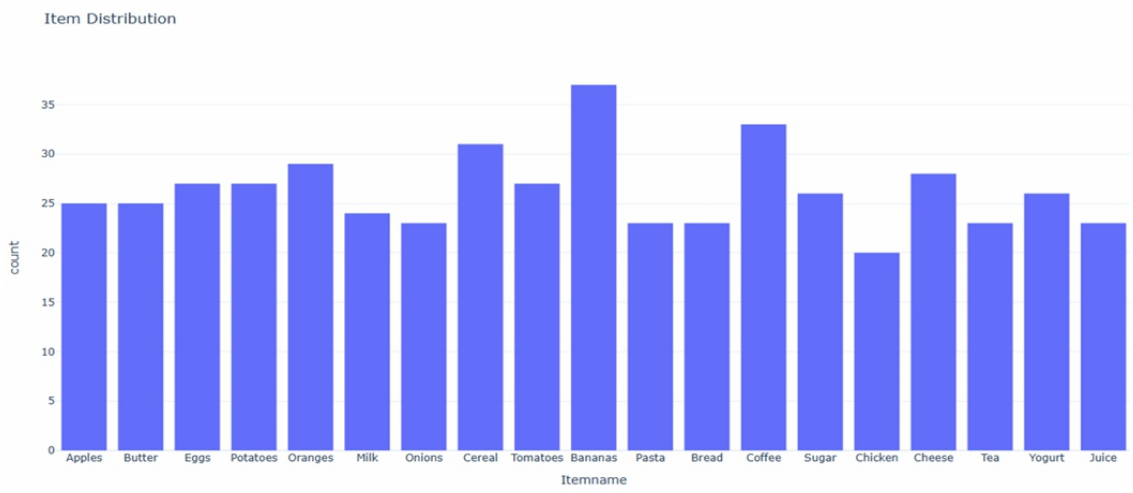


Figure 5: Item Distribution

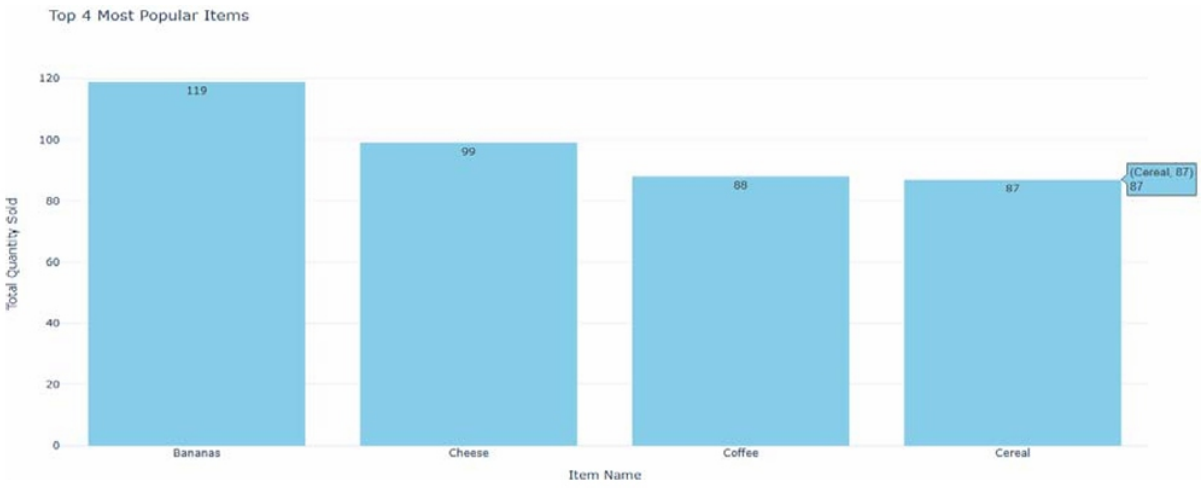
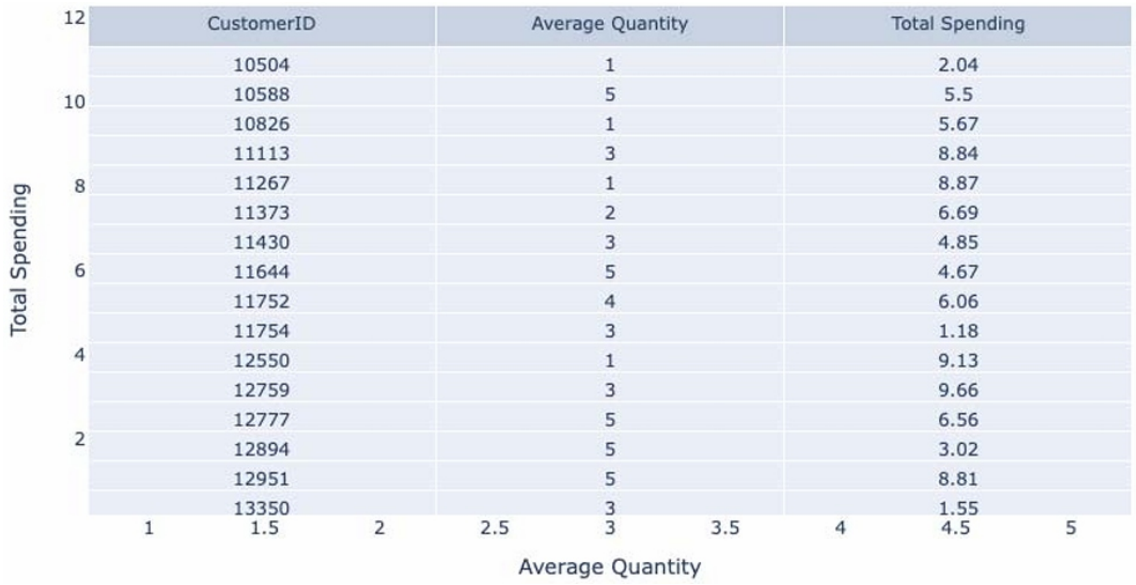


Figure 6: Top four Most Popular Items

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

Understanding Customer Behavior.



CustomerID	Average Quantity	Total Spending
10504	1	2.04
10588	5	5.5
10826	1	5.67
11113	3	8.84
11267	1	8.87
11373	2	6.69
11430	3	4.85
11644	5	4.67
11752	4	6.06
11754	3	1.18
12550	1	9.13
12759	3	9.66
12777	5	6.56
12894	5	3.02
12951	5	8.81
13350	3	1.55

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

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/marketbasket-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.simpli learn.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://selectstatistics.co.uk/blog/market-basket-analysisunder standing-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-beha vior-analysis/

9. Elnahla, N. (2021). Your Retailer Needs You: Retail lance and its Marketing Implications. https://doi.org/10.22215/etd/2021-14633

Survey: Artificial Intelligence Ethics

Shuxi Wang α & Zengyan Xia σ

A B S T R A C T

Artificial Intelligence Ethics is playing an important role with the development of Artificial Intelligence (AI). It is popular recognized that obeying to Artificial Intelligence Ethics guidelines and principles may resolve so many problems caused by Artificial Intelligence. This paper reviewed the development history of Artificial Intelligence Ethics, listed the main guidelines and principles of Artificial Intelligence Ethics, proposed the methods of Artificial Intelligence Ethics governance, discussed related algorithms to solve Artificial Intelligence Ethics problems.

Keywords: artificial intelligence; artificial intelligence ethics; artificial intelligence ethics guidelines and principles; artificial intelligence ethics algorithms; artificial intelligence ethics governance.

I. INTRODUCTION:

With the development of Artificial Intelligence, more and more ethical problems are caused by Artificial Intelligence, and Artificial Intelligence Ethics are drawing more and more attention. To solve Artificial Intelligence Ethics problems, this paper reviewed the development history of Artificial Intelligence Ethics, listed the main guidelines and principles of Artificial Intelligence Ethics, proposed the methods of Artificial Intelligence Ethics governance, discussed related algorithms to solve Artificial Intelligence Ethics problems.

Artificial intelligence ethics is not only a social problem, but also a philosophical problem. This paper has the viewpoint that Artificial intelligence ethics should be computed. This paper attempt to use mathematics and algorithms to solve Artificial intelligence ethics problems. In this paper, one Artificial Intelligence Ethics model will be proposed to solve Artificial intelligence ethics problems.

II. WHAT IS ARTIFICIAL INTELLIGENCE ETHICS

Artificial intelligence ethics is an academic hotspot. Artificial intelligence ethics mainly include the following aspects: (1). Whether or not Artificial Intelligence should own moral awareness. (2). Whether or not Artificial Intelligence should own the sense of responsibility. (3). Should Artificial Intelligence

should own the sense of responsibility. [3]. Should Artificial Intelligence make moral and ethical judgments regarding decisions related to human life and safety. [4]. If Artificial Intelligence can learn and createindependently, should Artificial Intelligence own intellectual property rights. [5]. Whether or not the application of Artificial Intelligence meets ethical and moral standards.Forexample, whethertheweaponization of artificial intelligence is acceptable. The debate in the academic community regarding the moral judgment of artificial intelligence requires distinguishing between two issues: first, the moral evaluation of artificial intelligence itself; Secondly, the evaluation of the good and evil consequences of the development and application of artificial intelligence. The key issue of moral judgment on artificial intelligence has not been resolved, that is, the issue of evaluating the good and evil of artificial intelligence itself has not been distinguished from the evaluation of the good and evil of the consequences of the development and application of artificial intelligence. The key to solving the latter problem still lies in humanity itself. However, in order to solve the previous problem, we cannot judge it based on the existing ethical and moral framework, but should critically reflect on traditional technological ethics. Overall, there are three main positions and viewpoints in the academic community.

a) The First Optimistic Stance

Experts and scholars who hold this position believe that artificial intelligence is just a means and tool, and it does not matter whether it is good or bad. The key lies in the human beings who use it. They hold an optimistic attitude towards the future development prospects of artificial intelligence. Overall, the research and widespread application of artificial intelligence have more advantages than disadvantages for human development, and can generate huge economic and social benefits.

Generally speaking, the optimistic stance is mostly upheld by some artificial intelligence professionals who are related to the research and application of artificial intelligence, or consider their own interests, or by scientists who blindly worship science and technology. Its flaw lies in the one-sided and isolated view of the positive aspects of artificial intelligence, such as its ability to generate huge economic and social benefits, reconstruct almost all industries including finance, healthcare, education, transportation, etc., and promote overall changes in human lifestyles. They intentionally or unintentionally ignore or conceal the negative effects of artificial intelligence, such as the emergence of killer robots that will pose a security threat to humanity and the potential degradation of human civilization caused by excessive reliance on artificial intelligence.

b) The Second Type of Neutral Stance

Experts and scholars who hold this position acknowledge that artificial intelligence itself has the

potential to "do evil", and its research and application pose a potential threat to humanity and may bring serious consequences. However, for some reasons, they still strongly support the development of artificial intelligence technology. Artificial intelligence is currently in its early stages of development, and its harm is far from strong enough, so there is no need to worry too much; For example, Tom Austin, a global leading analyst in the field of artificial intelligence, stated that Hawking's warning that "the complete development of artificial intelligence will lead to the complete destruction of humanity" is "very foolish", citing the reason that "artificial intelligence is still very low-level". "Artificial things cannot surpass humans"; This viewpoint stems from a certain religious sentiment, which states that "the Creator is always superior to what he has created," so there is no need to worry excessively. It is ironic that the scientists who should have had the most atheistic spirit seek intellectual resources from religious beliefs. Human beings can set moral standards for artificial intelligence, but there has never been an effective argument that artificial intelligence will inevitably comply with the moral standards set by humans; Some people believe that as long as moral education is provided to artificial intelligence, it can ensure that they are dedicated to goodness and serve humanity wholeheartedly. However, the question is how moral education can prevent artificial intelligence from developing in an unethical direction.

c) The Third Pessimistic Stance

Experts and scholars who hold this position believe that artificial intelligence is no longer a tool. It has a sense of life and learning ability, and has two moral possibilities of "doing evil": one is that the powerful power of artificial intelligence may trigger "human evil", and the other is that artificial intelligence itself has the ability to "do evil", and humans cannot cope with the "evil" of artificial intelligence, ultimately leading to nothingness and destruction. Therefore, they expressed concerns that artificial intelligence may lose control or harm humanity in the future. In early 2015, Stephen Hawking, Bill Gates, Elon Marks, and others signed an open letter calling for control over artificial intelligence development. Max believes that artificial intelligence can "summon demons", posing a greater threat to humanity than nuclear weapons. Hawking explicitly asserts that the complete development of artificial intelligence may lead to the extinction of humanity. People who hold a pessimistic stance believe that technology is not the path to human liberation. It is "not liberated from nature by controlling it, but rather a destruction of nature and humans themselves. The process of constantly killing living beings will ultimately lead to overall destruction."

In the application of Artificial Intelligence, balancing the development of technology with ethical and moral standards is an important issue. We need to consider technological innovation and development, while also paying attention to the interests and rights of society and humanity. For example, in the field of autonomous driving, it is necessary to consider vehicle safety and pedestrian safety, establish sound safety standards and regulations, and strengthen the supervision and management of artificial

intelligence to avoid safety and ethical issues. Dealing with ethical issues related to Artificial Intelligence requires multifaceted work and measures. Firstly, we need to strengthen research and discussion on ethical issues related to artificial intelligence, and establish ethical and moral standards for artificial intelligence. Secondly, it is necessary to strengthen the supervision and management of artificial intelligence, standardize its application and development, and prevent safety and ethical issues from arising. In addition, it is necessary to strengthen public participation and supervision, establish a sound feedback mechanism, and ensure the safety and ethics of Artificial Intelligence. Artificial intelligence ethics must conform to human morality: not infringing on privacy, not harming humanity, not influencing the political situation, especially in the field of AI weapons, more caution should be exercised.

In short, the development and application of Artificial Intelligence will inevitably face ethical and moral challenges. We need to strengthen research and discussion on artificial intelligence, establish ethical and moral standards for artificial intelligence, and strengthen supervision and management of artificial intelligence to avoid safety and ethical issues. In the long run, only by achieving a balance between technological development and ethical standards can the sustainable development and application of artificial intelligence technology be achieved.

III. THE PRINCIPLES OF ARTIFICIAL INTELLIGENCE ETHICS

Artificial intelligence ethics must ensure that AI does not make decisions that contain bias or discrimination, and establish mechanisms to interrogate AI to ensure that they comply with human ethical standards. Serving the interests of humanity and never harming humanity. The principles of Artificial Intelligence Ethics includes:

1. Developing artificial intelligence is for the common benefit of humanity.
2. Artificial intelligence should ensure fairness and be easy to understand.
3. Artificial intelligence should not be used to infringe on people's privacy.
4. All citizens have the right to receive education, enabling them to prosper and develop spiritually, emotionally, and economically alongside artificial intelligence.
5. Artificial intelligence should not be endowed with the autonomy to harm, destroy, or deceive humans.

Artificial intelligence is not without risks, and the formulation of these principles helps to mitigate risks. If there are methods to prevent the misuse of artificial intelligence technology, the public will trust AI more and better apply this technology.

AI will eliminate old jobs and create new ones. During the transition from old to new, the government must do a good job in vocational re-education to ensure that those who have been robbed of their jobs find new jobs.

IV. SPECIFIC ARTIFICIAL INTELLIGENCE ETHICAL IMPLICATIONS

Artificial Intelligence ethics include so many aspects. The following are some specific Artificial Intelligence ethical implications.

Medical Field

With the rapid development of technology, artificial intelligence technology has made breakthrough applications in the medical field. The development of artificial intelligence technologies such as artificial intelligence assisted diagnosis, intelligent drug design, and artificial intelligence assisted surgery has brought unprecedented development opportunities to the medical industry. However, the development of artificial intelligence in the medical field has also brought ethics and challenges. How to ensure that the use of artificial intelligence in the medical field does not cause potential harm to human health has become an urgent issue that needs to be addressed.

The basic principles and applications of artificial intelligence in the medical field.

a) The basic principles of artificial intelligence in the medical field.

Artificial intelligence (AI) technology is the ability to simulate human intelligent activities and achieve certain tasks. Its core is based on algorithms and big data, using techniques such as deep learning and machine learning to enable computers to recognize, analyze, and process large amounts of data, thereby achieving the acquisition, understanding, and application of information.

Harmful effects: exploring the ethics and challenges of artificial intelligence in the medical field. The application of artificial intelligence in the medical field mainly includes the following aspects:

1. Auxiliary Diagnosis

By analyzing a large amount of medical data, artificial intelligence can assist doctors in disease diagnosis and improve the accuracy of diagnosis. For example, artificial intelligence can recognize and analyze medical images through technologies such as depth learning, assist doctors in detection and

localization, and reduce surgical risks.

2. Intelligent Drug Design

Artificial intelligence can assist scientists in drug development, improving drug efficacy and reducing drug side effects. By deeply mining a large amount of bioinformatics data, artificial intelligence can predict the molecular structure and properties of drugs, providing scientists with targeted directions for drug development.

3. Artificial Intelligence Assisted Surgery

Artificial intelligence can assist doctors in surgical simulation and planning, improving the safety and efficiency of surgery. For example, artificial intelligence can provide real-time feedback to doctors by simulating surgical scenarios, assisting them in performing precise operations during the surgical process and reducing surgical risks.

The Ethics and Challenges of Artificial Intelligence in the Medical Field.

b) Privacy Protection

The data in the medical field is highly sensitive and involves patient privacy information. The application of artificial intelligence in the medical field requires strict protection of patient privacy to avoid the leakage of patient personal information.

1. *Data Security*

In the medical field, artificial intelligence needs to process a large amount of data, including sensitive information such as patient medical records, images, genes, etc. The confidentiality and security of these data are crucial, and artificial intelligence enterprises need to take strict security measures to ensure that these data will not be leaked during transmission, storage, and use.

2. Anonymity

In the medical field, artificial intelligence needs to process a large amount of anonymous data, such as patient medical records, medication usage records, etc. Although these data do not contain personal identification information, they are still sensitive. Therefore, artificial intelligence enterprises need to adopt strict policies and measures to ensure the security and confidentiality of these anonymous data.

c) Moral Responsibility

The application of artificial intelligence in the medical field has a strong moral responsibility. Thedevelopment of artificial intelligence technology requires adherence to ethical standards and respect for human dignity and rights.

1. Respect Individual Rights

The application of artificial intelligence in the medical field may have an impact on the personal rights of patients, such as infringement of patient privacy and leakage of genetic information. Therefore, artificial intelligence enterprises need to respect the individual rights of patients, protect their dignity and privacy.

2. Adhere to ethical standards

The application of artificial intelligence in the medical field requires adherence to medical ethical standards, respect for medical ethics and professional ethics. For example, artificial intelligence technology needs to ensure that there is no misdiagnosis or missed diagnosis in the detection and positioning process, to avoid unnecessary harm to patients.

d) Publicity

The application of artificial intelligence in the medical field needs to ensure fairness and avoid unfair distribution of medical resources due to factors such as race and gender.

1. Public Allocation of Medical Resources

The application of artificial intelligence in the medical field needs to follow the principle of public allocation of medical resources, ensuring that everyone can access public and reasonable medical resources.

2. Oppose Discrimination

The application of artificial intelligence in the medical field needs to oppose discrimination and ensure that everyone can receive equal medical treatment. Harmful effects: exploring the ethics and challenges of artificial intelligence in the medical field.

The use of artificial intelligence in the medical field has great potential and development space.

However, in order to ensure that the use of artificial intelligence in the medical field does not cause potential harm to human health, it is necessary to comply with relevant ethical standards and legal regulations. Artificial intelligence enterprises need to shoulder moral responsibilities, protect the privacy and dignity of patients, ensure the allocation of public medical resources, and make positive contributions to human health and the development of the medical industry.

Field of Human Rights

The risks of basic rights include personal data and privacy protection, as well as non-discrimination. The use of artificial intelligence may affect the fundamental values of the European Union and lead to the infringement of fundamental rights, including freedom of speech, freedom of assembly, human dignity, non-discrimination based on gender, race or ethnicity,

freedom of religious belief, or non-discrimination based on disability age, sexual orientation (applicable in certain fields), protection of personal data and private life, and the right to effective judicial remedies and fair trials, and consumer protection rights, etc. These risks may be due to flaws in the overall design of artificial intelligence systems (including human supervision), or may be due to possible biases not being corrected when using data (for example, the system only uses or primarily uses data from men for training, resulting in poor results related to women).

Prejudice and discrimination are inherent risks in any social or economic activity. Human decision-making cannot avoid errors and biases. However, when the same bias appears in artificial intelligence, it may have a greater impact, and without social governance mechanisms that control human behavior, it can affect and discriminate against many people. This situation also occurs when artificial intelligence systems are "learning" during operation.

Field of War

For a long time, discussions on the application of artificial intelligence in military have been limited to autonomous weapons and the ethical issues they bring. With the development of technology, attention should now be paid to the impact of artificial intelligence on security and other aspects of the military field. Artificial intelligence is greatly changing the civilian sector, such as improving efficiency, reducing costs, and automating processes, and the military will also usher in an AI revolution. At present, all countries must obtain human permission before using weapons, which is also in line with human values. However, what problems will occur when opponents deploy autonomous weapons without human permission needs to be urgently discussed.

There is reason to believe that even countries that have imposed some restrictions on artificial intelligence capabilities will encounter such opponents, which puts the countries that have imposed restrictions at a disadvantage. Therefore, countries must have a comprehensive understanding of what artificial intelligence can do. Although autonomous weapons have attracted a lot of attention, most conversations about this technology are negative, leading people to overlook the positive applications of artificial intelligence in areas such as military protection and reducing civilian casualties.

The Advantages of Artificial Intelligence

Artificial intelligence has broad application potential in optimizing human-machine collaboration in fields such as command chain communication and logistics, as well as predicting opponent maneuvers. Numerous countries, including the Israeli Defense Force, are conducting corresponding research.

Military commanders will use artificial intelligence to solve the dilemmas of war. Artificial intelligence will enhance the decision-making ability of commanders by providing more accurate battlefield situational awareness and higher responsiveness, thanks to constantly updated sensor data.

Artificial intelligence technology will also help decision-makers and analysts cope with the impact of information overload, better organize and process evergrowing opponent data, and enable troops to make predictions about future events and outcomes, enabling them to better prepare for combat. Better understanding of opponents is becoming one of the most promising application directions for artificial intelligence. Artificial intelligence will achieve faster and more real-time information collection, detection patterns, communication network drawing, and even better sensing of opponent morale by analyzing their language on social media and other platforms. These new AI features are equivalent to Intelligence Gathering 2.0. This type of analysis can be extended to the military communication and social media activities of civilians in hostile countries, in order to better understand a country's willingness to war at any time, which is the most critical factor in human warfare and will have a huge impact on decision-makers in both civilian and military fields.

In the field of military logistics and maintenance, artificial intelligence can create revolutionary cost saving efficiency, which is why most military forces prioritize conducting research in this area. Logistics support may lead to the most fundamental changes in the military. Artificial intelligence systems can also optimize the procurement process and achieve supply chain automation, predict the demand for maintenance equipment and order supplies, while minimizing costs. Artificial intelligence can also be used for personnel allocation, helping the military identify which soldiers are most suitable for which unit. Unlike other aspects of artificial intelligence, these applications are unlikely to raise any significant legal or ethical issues.

Artificial intelligence based technology can also enhance the capabilities of individual soldiers, which should not be seen as an unethical or dangerous way.

What is the Application of Artificial Intelligence in National Defense?

At the strategic level, artificial intelligence can enhance the capabilities of air defense systems. Emerging weapons, such as hypersonic missiles, are difficult to detect by existing defense systems due to their speed, while air defense systems that use artificial intelligence processing capabilities can detect and intercept such missiles. In addition, in the field of information warfare, artificial intelligence has great potential in quickly verifying information or identifying opponents.

v. ARTIFICIAL INTELLIGENCE ETHICS MODEL AND ALGORITHM

As an architecture system of autonomous intelligent agents, artificial intelligence's subjectivity or subject structure is somewhat similar to how we move certain functions of the human brain (or functions similar to the human brain) into machines. If the biological basis of human subjectivity is "neural", then the subjectivity of artificial intelligence can only be an imitation of human subjectivity. The scientific basis and manifestation of this imitation is the "algorithm". Setting aside the extent to which artificial intelligence agents are similar to human agents, researchers point out that the key element in making artificial intelligence products intelligent agents is the "moral algorithm" - an algorithm that teaches autonomous artificial intelligence devices to act responsibly, embedded in the algorithm system of artificial intelligence.

The subject mode of artificial intelligence faces significant ethical challenges on this issue. Taking autonomous robots in healthcare and the battlefield as an example, how should robots make decisions when facing the dilemma of life and death for human life? When inappropriate decisions lead to avoidable harm, whose responsibility is that? On this issue, although the subject mode of artificial intelligence highlights the importance of moral algorithms, its deeper and more important dependence is undoubtedly the "good law" established by humans for themselves.

Currently, there are roughly three algorithms that endow artificial intelligence with moral abilities.

One is to expand the moral logic through semantic networks, forming the concepts of obligation and permission;

The second is to establish association rules through knowledge graphs to detect moral judgment situations;

The third is to explore relevant relationships through cloud computing, evaluate or predict the consequences of actions.

Moral algorithms are algorithmic programs embedded in the algorithmic system that need to be improved. It itself is constantly changing and developing, rather than a specific existing thing, nor is it an ultimate assumption that can be achieved overnight or once and for all. It, as an artificial construction, is a "manual goodness" that leads to the "purpose goodness" and therefore depends on the human subject pattern. At this scale, algorithms can only promote the evolution of moral algorithms and their embedding in machines in a responsible manner by reflecting or following the "good law" of human subjectivity. This is the basic principle that moral construction in the era of artificial intelligence should follow, that is, algorithms follow the "good law". In this principle, although the term "good law" is abstract and ambiguous, the scale of human subjectivity it represents may also cause controversy in specific content, but it clearly points to two moral forms on the human scale in form.

The first form of morality is dominated by common human subject patterns and involves all ethical issues that humans may bring when expanding artificial intelligence. Specifically, when people view artificial intelligence as a tool, its moral specificity and importance always call for the return of the moral responsibility of human subjects. This is a simple normative orientation, which means that humans should plan and embrace the advent of the artificial intelligence era in a responsible way. A prudent ethics suggests that the greatest threat that artificial intelligence may face is not from machines, but from humans or their intentions and actions. Considering that the algorithm that endows robots with moral abilities is essentially an algorithm that mimics human morality, how is it possible to present human morality in the form of algorithms in machines if humans cannot obtain clarity on moral issues? The problem paradoxically illustrates the moral construction of artificial intelligence implosion. It responds in some way to James Moore's demand for ethical intelligence subjects to have moral clarity, that is, as autonomy increases, artificial intelligence with autonomous moral abilities must be able to make clear rational decisions when facing moral dilemmas or conflicts of different moral principles. This demand for moral clarity, in turn, constructs or depicts the characteristics of "good law" at the human scale, forcing the human subject model to do everything possible to break out of various moral ambiguity zones that may lead to dark consequences (or even disasters).

The second form of morality is dominated by the "inter subject" mode of interaction between human subjects and artificial intelligence subjects, involving the moral construction of the dependent relationship between human subjects and intelligent subjects. This is a new field. Moral algorithms can

only continuously correct biases or errors, further upgrade and improve in the repeated game between human machine interaction subjects. Autonomous robots may make decisions that we believe are morally wrong - such as being authorized not to provide pain relievers to patients, or biased artificial intelligence may self reinforce and harm society. However, this should not be a reason for humans to reject robots, but rather an opportunity for robots or artificial intelligence to improve and enhance their moral form. With the establishment of interdependence between human subjects and artificial intelligence subjects, autonomous robots with self decision-making ability, once they learn to develop decision-making algorithms from a moral perspective in their interaction with human subjects, can become a "good law" of interdependence between humans and machines to avoid harm.

vi. CONCLUSION AND FUTURE WORK

This paper has the viewpoint that Artificial intelligence ethics should be computed. This paper attempt to use mathematics and algorithms to solve Artificial intelligence ethics problems. In this paper, one Artificial Intelligence Ethics model will be proposed to solve Artificial intelligence ethics problems. The future work will focus on related algorithms about Artificial intelligence ethics.

References

1. D. L. Medin, "Structural principles in categorization", in T. J. Tighe & B. E. Shepp (eds.) Perception, Cognition and Development: Interaction Analyses, Hillsdale, N. J.: Lawrence Erlbaum, 1983, p.1469.

2. S. Shanker, Wittgenstein's Remarks on the Foundations of AI, London & New York: Routlege,1998, p.187.

3. Beauchamp, T. L. and Childress, J. F. 2012. Principles of Biomedical Ethics 8th. Oxford University Press, Oxford.

4. Biller-Andorno, N.; Aebi-Mueller, R.; (...); Sedlakova, J.2021. Ineffectiveness and unlikelihood of benefit: Dealing with the concept of futility in medicine. Swiss Academy of Medical SciencesBern.

5. Biller-Andorno, N; Ferrario, A; (...); Krauthammer, M. Mar 2022. Mar 2021 (Early Access). JOURNAL OF MEDICAL ETHICS 48 (3), pp.175-183.

6. Biller-Andorno, N and Biller, A. Oct 10 2019. NEW ENGLAND JOURNAL OF MEDICINE 381 (15), pp.1480-1485.

7. Giubilini, Alberto and Savulescu, Julian. 2018. Philosophy & technology 31 (2) , pp.169-188.

8. Hermann, H; Feuz, M; (...); Biller-Andorno, N. Jun 2020. Apr 2020 (Early Access). MEDICINE HEALTH CARE AND PHILOSOPHY 23 (2) , pp.253-259.

9. Loi, M; Ferrario, A and Viganò, E. Sep 2021. Oct 2020 (Early Access). ETHICS AND INFORMATION TECHNOLOGY 23 (3), pp.253-263.

10. Meier, LJ; Hein, A; (...); Buyx, A. Jul 3 2022. Mar 2022 (Early Access). AMERICAN JOURNAL OF BIOETHICS 22 (7) , pp.4-20.

11. van de Poel, I. Sep 2020 | Sep 2020 (Early Access). MINDS AND MACHINES 30 (3), pp.385-409.

12. Shaw, D; Trachsel, M and Elger, B. Jul 2018. BRITISH JOURNAL OF PSYCHIATRY. 213 (1), pp.393-395.

13. den Hartogh, G. Mar 2016. MEDICINE HEALTH CARE AND PHILOSOPHY. 19 (1) , pp.71-83.

14. Hermann, H; Trachsel, M and Biller-Andorno, N. Sep 2015. JOURNAL OF MEDICAL ETHICS. 41 (9) , pp.739-744.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

-

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Notes: