**“REAL TIME FORGERY ANALYSIS USING NOVEL MULTILAYER PERCEPTRON”**

**A Project Report submitted in partial fulfillment of the requirements for the award of the degree of,**

**BACHELOR OF TECHNOLOGY IN**

**COMPUTER SCIENCE AND ENGINEERING**

**Submitted by**:

#### B.MUNI KUMAR 321910302004

#### D.SUPRIYA 321910302005

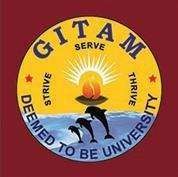
#### K.SATYA SAI VENKAT 321910302026

#### A.KOTI MYTHREYI 321910302036

**Under the esteemed guidance of**

## Dr.TANVIR H. SARDAR

## ASSISTANT PROFESSOR



**Department of Computer Science & Engineering, GITAM SCHOOL OF TECHNOLOGY**

## GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT

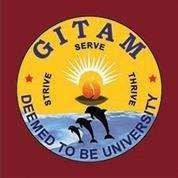
**(Deemed to be University) Bengaluru Campus.**

## March 2023

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GITAM SCHOOL OF TECHNOLOGY**

## GITAM

**(Deemed to be University)**



**CERTIFICATE**

This is to certify that the project report entitled **“REAL TIME FORGERY ANALYSIS USING NOVEL MULTILAYER PERCEPTRON”** is a bonafide record of work carried out by **D.SUPRIYA(321910302005), B.MUNI KUMAR(321910302004), K.SATYA SAI VENKAT(321910302026), A.MYTHREYI(321910302036)** submitted in partial fulfillment of the requirement for the award the of degree of **Bachelors of Technology in Computer Science and Engineering**.

**Project Guide. Head of the Department.**

**SIGNATURE OF THE GUIDE SIGNATURE OF THE HOD**

**Dr.Tanvir H. Sardar**

Assistant Professor.

.

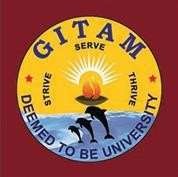
**Dr. VAMSIDHAR YENDAPALLI,**

Professor.

## DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING GITAM SCHOOL OF TECHNOLOGY

**GITAM**

**(Deemed to be University)**



## DECLARATION

We, hereby declare that the project report entitled **“REAL TIME FORGERY ANALYSIS USING NOVEL MULTILAYER PERCEPTRON”** is an original work done in the **Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University)** submitted in partial fulfillment of the requirements for the award of the degree of **B.Tech.** in Computer Science and Engineering. The work has not been submitted to any other college or University for the award of any degree.

#### Date:

**Registration No(s). Name(s) Signature(s)**

#### 321910302004

**321910302005**

#### 321910302026

**321910302036**

#### B. MUNI KUMAR

#### D.SUPRIYA

#### K. SATYA SAI VENKAT

#### A. KOTI MYTHREYI

I

## ACKNOWLEDGEMENT

The Satisfaction and Euphoria accompanying any Minor Project's successful completion would be incomplete without mentioning the people who made it possible, whose constant guidance and encouragement and our efforts with successful. We take this opportunity to express the deepest gratitude and appreciation to all those who held us directly or indirectly towards the successful completion of the Project.

I would like to thank **Dr. Nageswara Rao T**, Ph.D., Director, GITAM School of Technology, for providing all the facilities.

I would like to thank HOD **Dr. VAMSIDHAR YENDAPALLI**, Ph.D., Professor, Department of Computer Science and Engineering for his support and encouragement that went a long way in the successful completion of this Project.

I consider that as a privilege to express our gratitude and respect to **Dr.Tanvir H. Sardar,** Assistant professor, department of Computer Science and Engineering for being our internal guide for his integral and incessant support offered to us throughout the course of this project, and for constant source of inspiration throughout the Seminar.

I would like to thank our project coordinator **Mr. RATI BHAN** for his integral and incessant support offered to us throughout the course of this project and for constant source of inspiration throughout the Project.

I would like to thank my parents and my friends for their support and encouragement in the completion of the project on time.

Finally, I would like to thank all the teaching and non-teaching staff members of the Computer Science and Engineering Department, for their support in the completion of the project on time.

|  |  |
| --- | --- |
| **Student Name’s** | **Registration No**. |
| B. MUNI KUMAR | 321910302004 |
| D.SUPRIYA | 321910302005 |
| K. SATYA SAI VENKAT | 321910302026 |
| A. KOTI MYTHREYI | 321910302036 |

**ABSTRACT**

To improve the accuracy in finding out the forgery analysis using Novel Multi-Layer Perception and Random Forest Algorithm. Novel Multi-Layer Perception and Random Forest Algorithm are classifier algorithms under machine learning. Materials and Methods: The two groups in this study are the Random Forest Algorithm and Novel Multi-Layer Perception. A sample size of 10 is used for each group and the parameters of the present study include an alpha value of 0.05 and a beta value of 0.2.

The significance level of the dataset was predicted using the SPSS tool with the assumption that G-Power is 80%. Each algorithm's output is distinct, and each one does it using a different sample size. Results: In defining the forgery analysis with significance value 0.001, the Novel Multi-Layer Perception produces an output with accuracy 84.2%, which is more accurate than the Random Forest Algorithm with accuracy value 77.4% (p<0.05). Conclusion: The Novel Multi-Layer Perception is more accurate than the Random Forest Algorithm in classifying forgeries. This algorithm gives the best result. Novel Multilayer Perception obtained an improved accuracy of 4.7% compared to the Random Forest Algorithm.

Keywords: Forgery Analysis, Novel Multi-Layer Perception, Random Forest Algorithm, Machine learning, Classifier, Fake data.

## TABLE OF CONTENTS

**Title Page No.**

Declaration I

[Acknowledgement II](#_TOC_250023)

[Abstract III](#_TOC_250022)

[Table of Contents IV](#_TOC_250021)

[List of Figures V](#_TOC_250020)

[List of Tables V](#_TOC_250019)

1. [INTRODUCTION 1](#_TOC_250018)-3
2. [LITERATURE SURVEY 4](#_TOC_250017)
3. [SOFTWARE AND HARDWARE SPECIFICATIONS 5](#_TOC_250016)
   1. [Hardware Requirements 5](#_TOC_250015)
   2. [Software Requirements 5](#_TOC_250014)
4. [OBJECTIVES 6](#_TOC_250013)
5. [LIBRARIES 7-9](#_TOC_250011)
6. [PROBLEM STATEMENT 10](#_IMPLEMENTATION)
7. [DESIGNING 11-14](#_TOC_250004)
8. [DATASET 15](#_EXPERIMENTAL_RESULTS)
9. [IMPLEMENTATION 15-17](#_CONCLUSION)
10. [CALUCLATIONS 18-23](#_FUTURE_WORK)
11. OUTPUTS 24-26
12. CONCLUSION 27
13. FUTUREWORK 28
14. REFERENCES 29-30

## LIST OF FIGURES

|  |  |  |
| --- | --- | --- |
| **Fig no** | **Title** | **Page no** |
| 1 | Real time forgery | 2 |
| 2 | Advantages of real time processing | 3 |
| 3 | Hardware requirements | 5 |
| 4 | Software requirements | 5 |
| 5 | Objectives | 6 |
| 6 | What is classifier | 11 |
| 7-9 | Designing of classifiers | 12-14 |
| 10 | Dataset | 15 |
| 11-13 | Implementation | 15-17 |
| 14-24 | Calculation of classifiers | 18-23 |
| 25-30 | Outputs of classifiers | 24-26 |

## LIST OF TABLES

|  |  |  |
| --- | --- | --- |
| **Table** | **Table title** | **Page no** |
| 1 | Literature Survey | 4 |
| 2 | Output of all data sets | 24-26 |

# INTRODUCTION

# What is Real time forgery analysis

In earlier days, people used to purchase their products through offline methods and there were no online methods recently. They used to buy their product on the buying place itself with the other customer reviews who were standing beside them. As technology is increasing or growing day by day, people's power is becoming flawless nowadays (Nguyen et al. 2019). By the help of people's weakness, the marketing is flattering and current online products are trending into the real world. Not only the pros for delivering online products to the end users, here cons are also marked. The crucial major feature is that it will be helpful for measuring products from different localities in one online shopping, which reduces the work, provides opinion from a number of individuals and many things to be added, as well as cons. But the major area of the research is about online fake data reviews (Provided, n.d.; Al-Sultany and Hussain 2019). As people will shop according to their needs and will buy through online and save their time but by buying through online many people will follow the reviews of the product and go to the next step of billing. The key role of every human psychology is reviews or feedback from different individuals for a particular product (Bhavitha and Thangaraj 2022). This is also another platform for the hackers to change the clarity of the product, with the fake data reviews they may increase the level of the product or they can decrease the status of the consequence. So, by finding fake data reviews using different machine learning algorithms and finding whether the review which is given by the human is fake data or genuine (Malte and Martin 2020). Forgery analysis is used in many different areas, not only in particular online methods, there are many frauds like fingerprint scanning, fake data news, fake data documents and many others. These many areas come under forgery analysis.

There are about 36 articles in IEEE Xplore and 19 in Scopus related to this study. Novel Multilayer Perception is derived from structured data algorithms (Malte and Martin 2020; Chiluwa and Samoilenko 2019). The observation of forgery analysis using machine learning was performed. Analysis of reviewing of product applications techniques for classification models were evolved (Lenain 2003) This first Report was developed based on forgery analysis using machine learning techniques. Automatic finding of forgery analysis was derived with different algorithms with finding of higher accuracy, as study says my finding shows that fraudulent reviews typically have lengthier sentences, more pauses, and duplicate phrases. The most cited article is (“Figure 1: Evolution of Searches about Fake News and Fake Reviews,” n.d.; Banerjee and Chua 2019). Using these language variables, it is able to distinguish false reviews from actual reviews with great accuracy, according to the application of different machine learning classification algorithms. The overall best cited article is (Kaur 2016) Forgery Analysis using Machine Learning Techniques.

As many datasets contain 1000 of datasets but to detect which is genuine dataset regarding the algorithm and finding of higher accuracy comes to final discussion of the article (Kingsnorth 2016). The disadvantage or the trouble is that finding out which fake data is genuine or fake is extremely difficult to find because the majority of currently used standard feature extraction procedures are created for quick analysis (Stokes 2009; Wang, Aisihaer, and Aihemaiti 2022). The aim of this article is to increase the accuracy to find forgery analysis using a multi machine learning algorithm.

“Analysis” literally means a detailed examination of the elements or structure of something. In a nutshell, Data Analytics is the process of analyzing data from the past in order to make appropriate decisions in the future by utilizing valuable insights. Data Analysis, on the other hand, aids in understanding the data and provides necessary insights from the past to comprehend what has occurred so far. So, Data Analysis is an umbrella term that encompasses Data Analytics in its subset.

The term "multilayer perceptron" does not refer to a single perceptron that has multiple layers. Rather, it contains many perceptron’s that are organized into layers. An alternative is "multilayer perceptron network". Moreover, MLP "perceptron’s" are not perceptrons in the strictest possible sense. True perceptron’s are formally a special case of artificial neurons that use a threshold activation function such as the Heaviside step function. MLP perceptron’s can employ arbitrary activation functions. A true perceptron performs binary classification, an MLP neuron is free to either perform classification or regression, depending upon its activation function. The term "multilayer perceptron" later was applied without respect to nature of the nodes/layers, which can be composed of arbitrarily defined artificial neurons, and not perceptron’s specifically. This interpretation avoids the loosening of the definition of "perceptron" to mean an artificial neuron in general.



Fig:1

# Features of Real time forgery analysis

• Unreal discounts.

• Flimsy packaging.

• Grammatical & spelling mistakes.

• Fake websites.

• Poor quality of products.

# Advantages and Disadvantages of Real time forgery analysis

**Advantages:**

* Real reviews build trustworthiness.
* Provide social proof.
* Improve your brand's online visibility.
* Online reviews help confirm authority.
* Online reviews help consumers make purchasing decisions

**Disadvantages:**

* some process needs to be optimized.
* Lack of data.

Diagram

Description automatically generated

Fig:2

# What is Novel Multi-Layer Perceptron

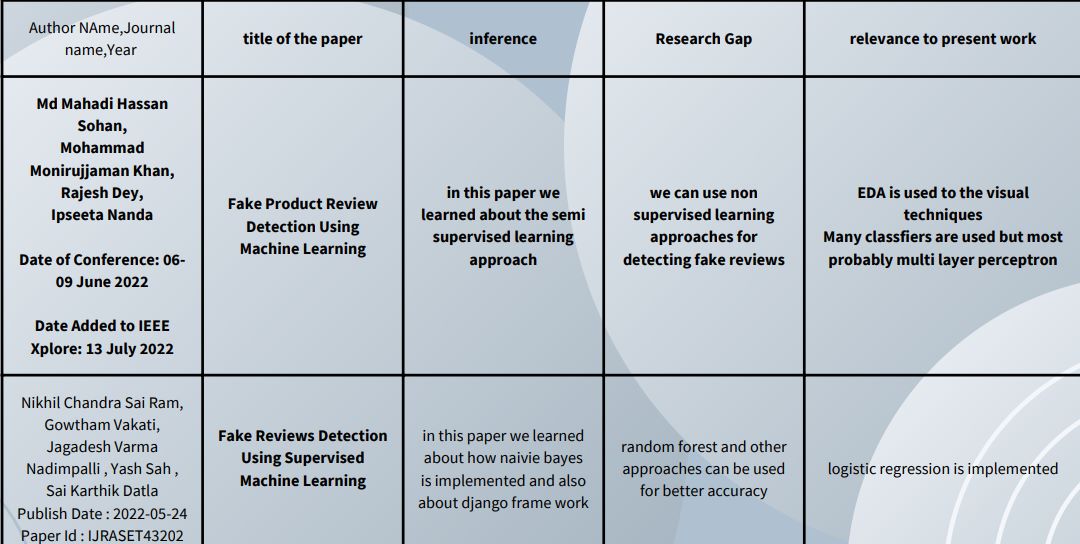
Novel MultiLayer Perception and Random Forest Algorithm are two groups of algorithms under machine learning which are used. These two algorithms belong to the classifier. The given number of samples in research are two in which Group 1 is Novel Multilayer Perception which is compared with the Group 2 of Random Forest Algorithm. With the help of accuracy output, it is iterated 10 times to get improved accuracy with G-power 80%, threshold 0.05% and Confidence Interval(CI) 95%.

# This algorithm is one of the machine learning algorithms in which it is used to detect the forgery analysis to predict the future profits. Nowadays, forgery analysis is becoming an effect to the economy, so this article predicts the output.

# LITERATURE SURVEY

Literature survey is the most important step in any kind of research. Before start developing, we need to study the previous papers of our domain which we are working and based on study we can predict or generate the drawback and start working with the reference of previous papers. “In this section, we briefly review the related work Product Fake Reviews Detection System using Novel Multi-layer perceptron based on Deep Learning.

“yelp.csv” dataset is used to predict the real time forgery analysis. The execution is done in the Python environment using the main packages pandas, sklearn, NumPy, keras, matplotlib, seaborn, countVectorizer. Analysis is done using google Collaboratory.



# Table 1

# SOFTWARE AND HARDWARE SPECIFICATIONS

* 1. **Hardware Requirements:**

Memory: 128 Gb

Processor: Intel or Ryzen

Operating System: Windows 10 or Windows11

Hard Disk: 2 Gb or Higher



Fig:3

**3.2. Software Requirements:**

IDE: Google colab

Language: Python

Internet Browser: Internet Explorer 6.0 or above, Google chrome, MozillaFirefox



Fig:4

# OBJECTIVES

## Step by Step process:

# Step 1: data set gathering

# Step 2: analyzing the data collected

# Step 3: start an approach for data minimization

# Step 4: importing all the packages needed

# Step 5: loading the dataset

# Step 6: creating a new column in the data set for the number of words in the review

# Step 7: Compare text length to stars

# Step 8: Calculating mean values of voted columns

# Step 9: Classification

# Step 10: Splitting the dataset into two parts and removing the punctuation marks.

# Step 11: implementing countVectorizer ()

# Step 12: implementing classifiers.

# 

Fig:5

# LIBRARIES

* 1. **Pandas**

Pandas is defined as an open-source library that provides high-performance data manipulation in Python. The name of Pandas is derived from the word Panel Data, which means an Econometrics from Multidimensional data. It is used for data analysis in Python and developed by Wes McKinney in 2008.

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with “relational” or “labelled” data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.

* 1. **Numpy**

NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transform, and matrices. NumPy was created in 2005 by Travis Oliphant. It is an open-source project and you can use it freely.

NumPy is a python library mainly used for working with arrays and to perform a wide variety of mathematical operations on arrays. NumPy guarantees efficient calculations with arrays and matrices on high-level mathematical functions that operate on these arrays and matrices.

* 1. **Matplotlib**

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Create publication quality plots. Make interactive figures that can zoom, pan, update.

Matplotlib is a cross-platform, data visualization and graphical plotting library for Python and its numerical extension NumPy. As such, it offers a viable open-source alternative to MATLAB. Developers can also use matplotlib's APIs (Application Programming Interfaces) to embed plots in GUI applications

* 1. **Pyplot**

Pyplot is an API (Application Programming Interface) for Python's matplotlib that effectively makes matplotlib a viable open-source alternative to MATLAB. Matplotlib is a library for data visualization, typically in the form of plots, graphs and charts.

Pyplot is a sub-module of the matplotlib library for Python. It is a library consisting of a collection of functions/methods used for plotting simple 2D graphs using Python. Pyplot can be imported using import matplotlib

* 1. **Seaborn**

Seaborn is a library that uses Matplotlib underneath to plot graphs. It will be used to visualize random distributions.

Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized easily..

* 1. **SciPy**

SciPy is a scientific computation library that uses NumPy underneath. SciPy stands for Scientific Python. It provides more utility functions for optimization, stats and signal processing. Like NumPy, SciPy is open source so we can use it freely

* 1. **Nltk**

The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for applying in statistical natural language processing (NLP). It contains text processing libraries for tokenization, parsing, classification, stemming, tagging and semantic reasoning

* 1. **String**

Python string module contains a single utility function – capwords (s, sep=None). This function split the specified string into words using str. Split (). Then it capitalizes each word using str. Capitalize () function.

* 1. **Math**

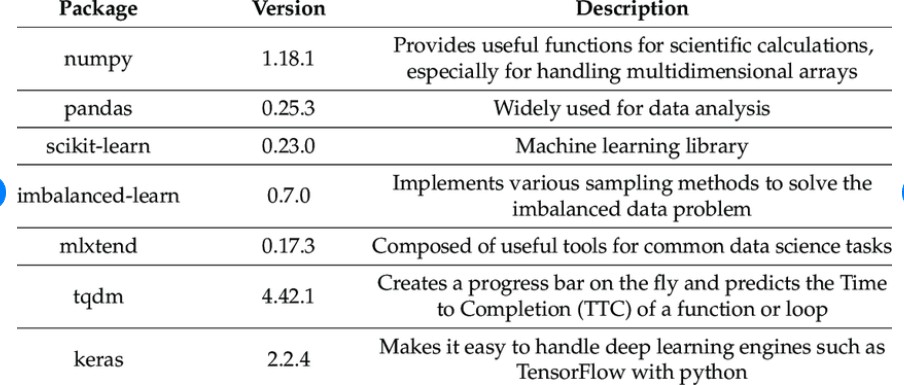
math is a built-in module in the Python 3 standard library that provides standard mathematical constants and functions.

* 1. **Count Vectorizer**

CountVectorizer is used to transform a corpora of text to a vector of term / token counts. It also provides the capability to preprocess your text data prior to generating the vector representation making it a highly flexible feature representation module for text.

from sklearn.feature\_extraction.text import CountVectorizervectorizer = CountVectorizer()

* 1. **Uses of Libraries**



# PROBLEM STATEMENT

The problem statement for fake review detection is to identify and classify reviews as genuine or fake. The task involves analyzing a massive collection of reviews with different categories and labels and detecting reviews that are computer-generated (CG) and, therefore, fake.

The objective is to develop an algorithm that can accurately distinguish between genuine and fake reviews, thus providing users with reliable information and preventing businesses from being unfairly reviewed. The algorithm must be scalable, robust, and able to handle different categories and types of reviews. The ultimate goal is to improve the credibility of online reviews and promote trust and transparency in online platforms.

Detection of fake reviews out of a massive collection of reviews having various distinct categories like Home and Office, Sports etc. with each review having a corresponding rating, label i.e.CG(Computer Generated Review) and OR(Original Review generated by humans) and the review text. Main task is to detect whether a given review is fraudulent or not. If it is computer generated, it is considered fake otherwise not.

# DESINING

* 1. **What is a Classifier?**

## A classifier in machine learning is an algorithm that automatically orders or categorizes data into one or more of a set of “classes.” One of the most common examples is an email classifier that scans emails to filter them by class label: Spam or Not Spam.

## In data science, a classifier is a type of machine learning algorithm used to assign a class label to a data input. An example is an image recognition classifier to label an image (e.g., “car,” “truck,” or “person”).

Fig:6

## Classifiers used in project

## K nearest neighbor classifier.

## Support vector machine.

## Decision tree classifier.

## Gradient boosting classifier.

## Random forest classifier.

## Novel multi-layer perceptron.

* 1. **K-nearest neighbor Classifier**

K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique. KNN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category This KNN algorithm is used for regression as well as for the classification but mostly it is used for the classification problems

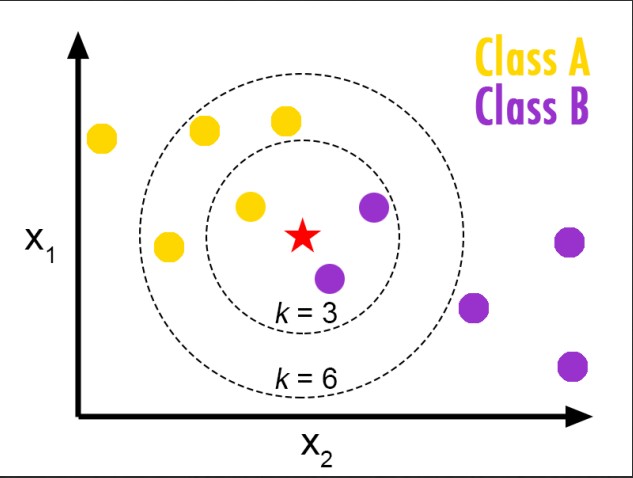
sklearn.neighbors.KNeighborsClassifier(neighbours int, default=5)Testing the IP Address for to communicate between nodes.

Fig:7

* 1. **Support vector machine**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.

sklearn.svm import SVC

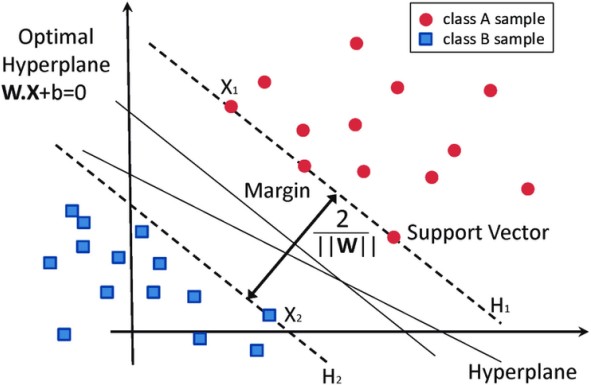


Fig:8

## Decision tree Classifier

## Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree structured classifier, where internal nodes represent the features of a dataset, branches

## represent the decision rules and each leaf node represents the outcome

## fromsklearn. treeimport DecisionTreeClassifier dtree=DecisionTreeClassifier (max\_depth =6, random\_state=123,criterion='entropy')

## Random Forest Classifier

## Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

## from sklearn.ensemble import RandomForestClassifier

## rfc=RandomForestClassifier()

## Gradient boosting Classifier

## Gradient boosting classifiers are a group of machine learning algorithms that combine many weak learning models together to create a strong predictive model. Decision trees are usually used when doing gradient boosting.

## from sklearn.ensemble import GradientBoostingClassifier

## gbc=GradientBoostingClassifier()

* 1. **Novel multi-layer perceptron**

A Multilayer Perceptron has input and output layers, and one or more hidden layers with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.

from sklearn.neural\_network import MLPClassifier

mlp = MLPClassifier()

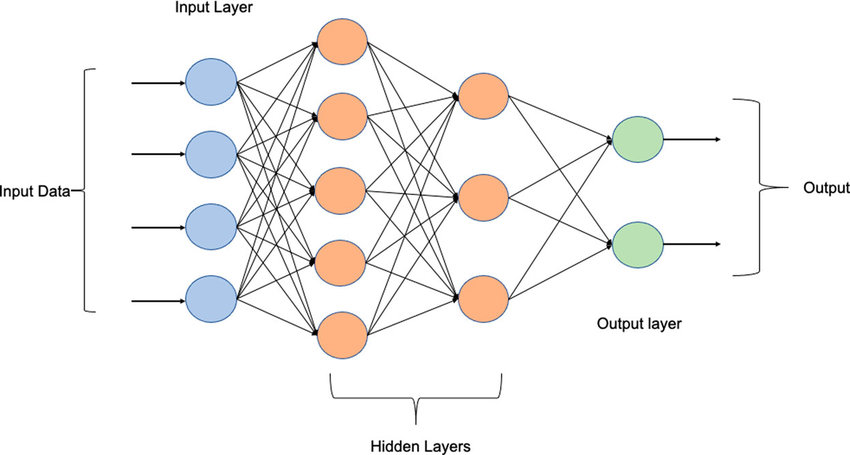


Fig:9

# DATASET

The challenging situation we faced during the experiments was to find a suitable sales dataset. We used yelp.csv dataset. This is available publicly available in Kaggle.

it comprises ten parameters.

The shape of the data set is (1000, 10)

Information of the data set is follows

Graphical user interface, text, application, email

Description automatically generated

Fig:10

# Implementation

# Python provides various libraries for data processing. Some of them are pandas numpy matplotlib. pyplot seaborn scipy. These are vast libraries which helped us to create applications and models in a variety of fields

# Text, letter Description automatically generated

Fig:11

* 1. **Comparison of data**

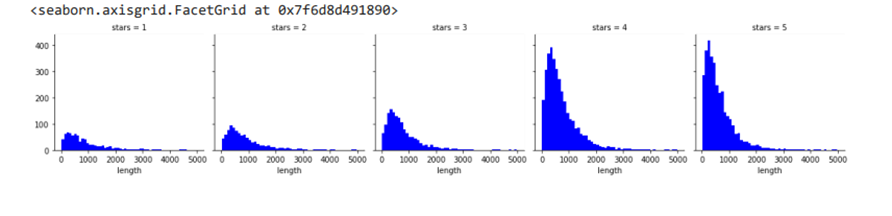
****

Fig:12

* 1. **Training and testing the data**

Train/Test is a method to measure the accuracy of your model. It is called Train/Test because you split the data set into two sets: a training set and a testing set. 80% for training, and 20% for testing. You train the model using the training set. This is the phase where we the results that is nothing but the accuracy, precision, recall and f1 score

* + 1. **Precision**

Precision is how good the model is at predicting a specific category.

* + 1. **Recall**

## Recall measures the proportion of actual positive labels correctly identified by the model.

## 

## 

* + 1. **F1 score**

## F1 score is a machine learning evaluation metric that measures a model's accuracy.

## 

* + 1. **Weighted Avg**

## Weighted average or weighted sum ensemble is an ensemble machine learning approach that combines the predictions from multiple models, where the contribution of each model is weighted proportionally to its capability or skill.

## We have weighted average for precision, recall and for f1 score.

## 

* + 1. **Macro Avg**

## The macro average is the arithmetic mean of the individual class related to precision, memory, and f1 score.

## We have macro average for precision recall and f1 score.

* + 1. **Accuracy**

Accuracy is how close a given set of measurements are to their true value, while precision is how close the measurements are to each other.

* + 1. **Overall Accuracy values:**

**Table

Description automatically generated**

Fig:13

# Calculations

# We got confusion matrices for classifier, with the help of this confusion matrix we calculated precision recall ,f1 score, weighted avg, macro avg, accuracy with above mentioned formulas.

# A confusion matrix presents a table layout of the different outcomes of the prediction and results of aclassification problem and helps visualize its outcomes. It plots a table of all the predicted and actual values of a classifier.

# K-nearest neighbor Classifier

# 

Fig:14

# 

Fig:15

# Support vector machine

# 

Fig:16

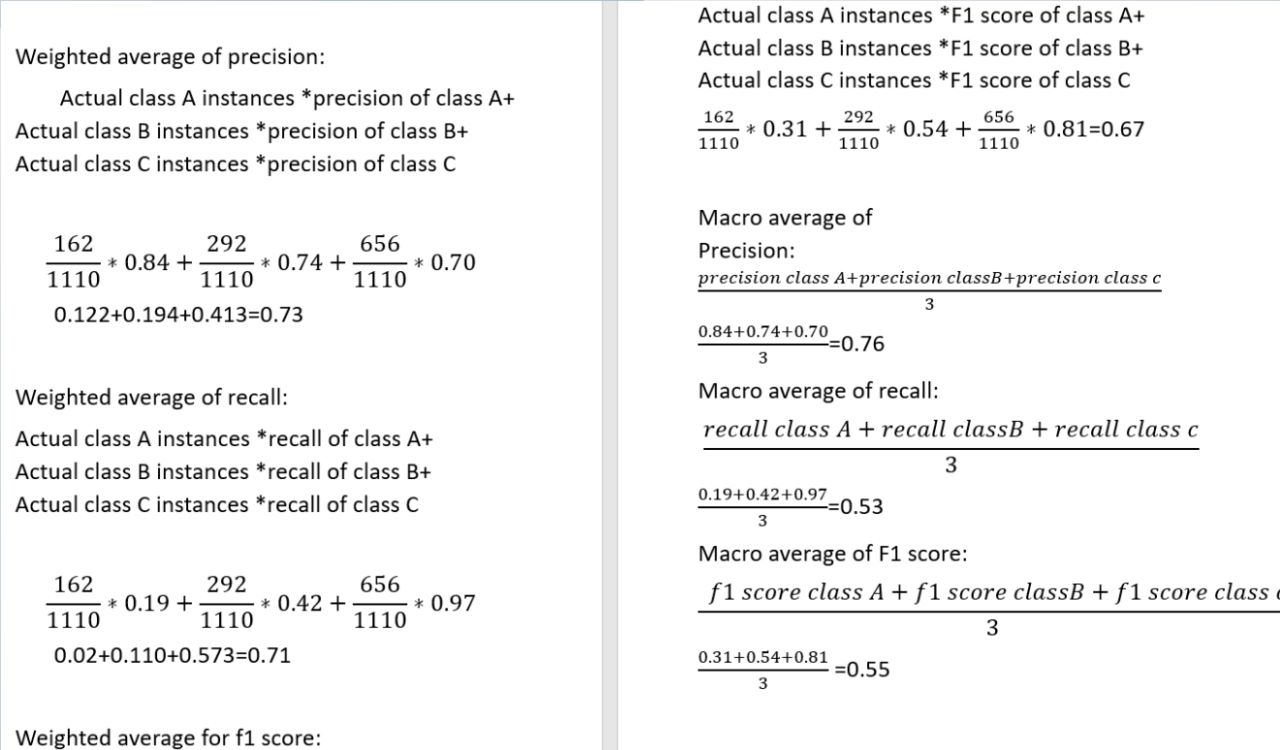


Fig:17

# Random Forest

# 

Fig:18

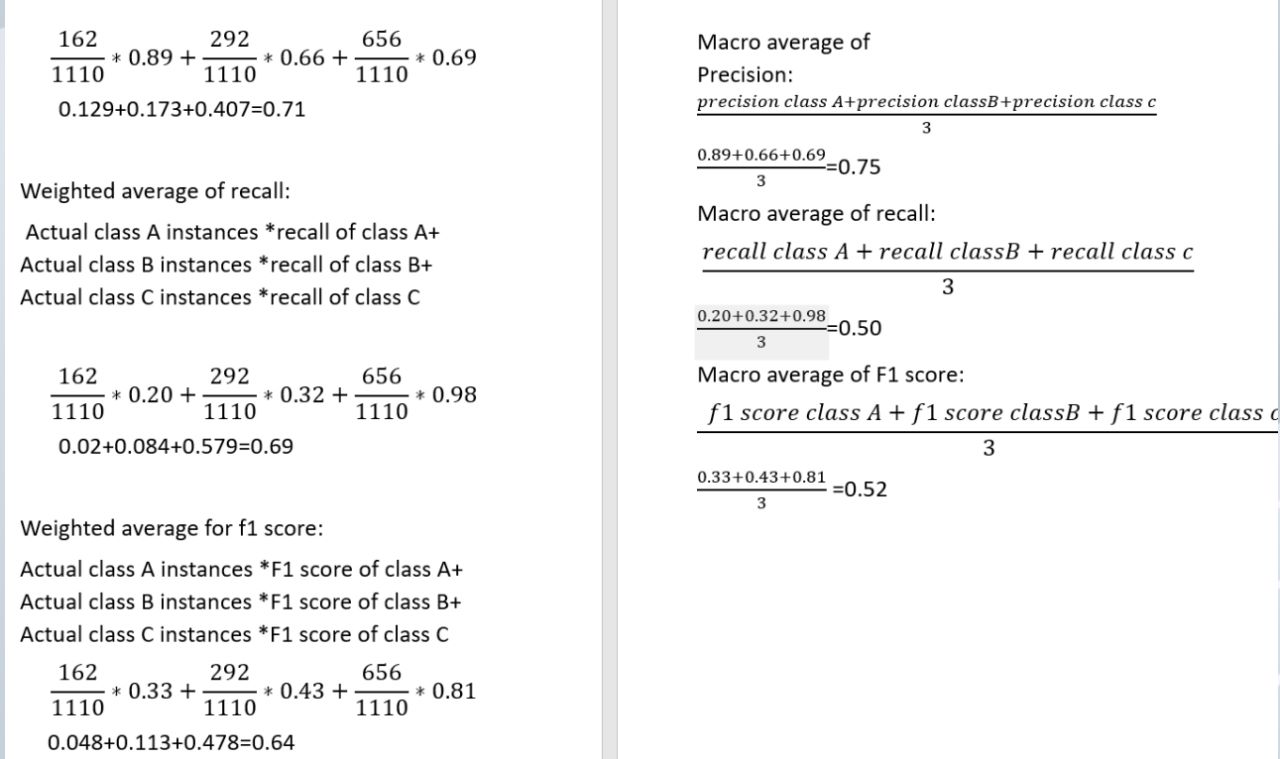


Fig:19

# Decision tree Classifier

# 

# Fig:20

# 

# Fig:21

* 1. **Gradient Boosting Classifier**

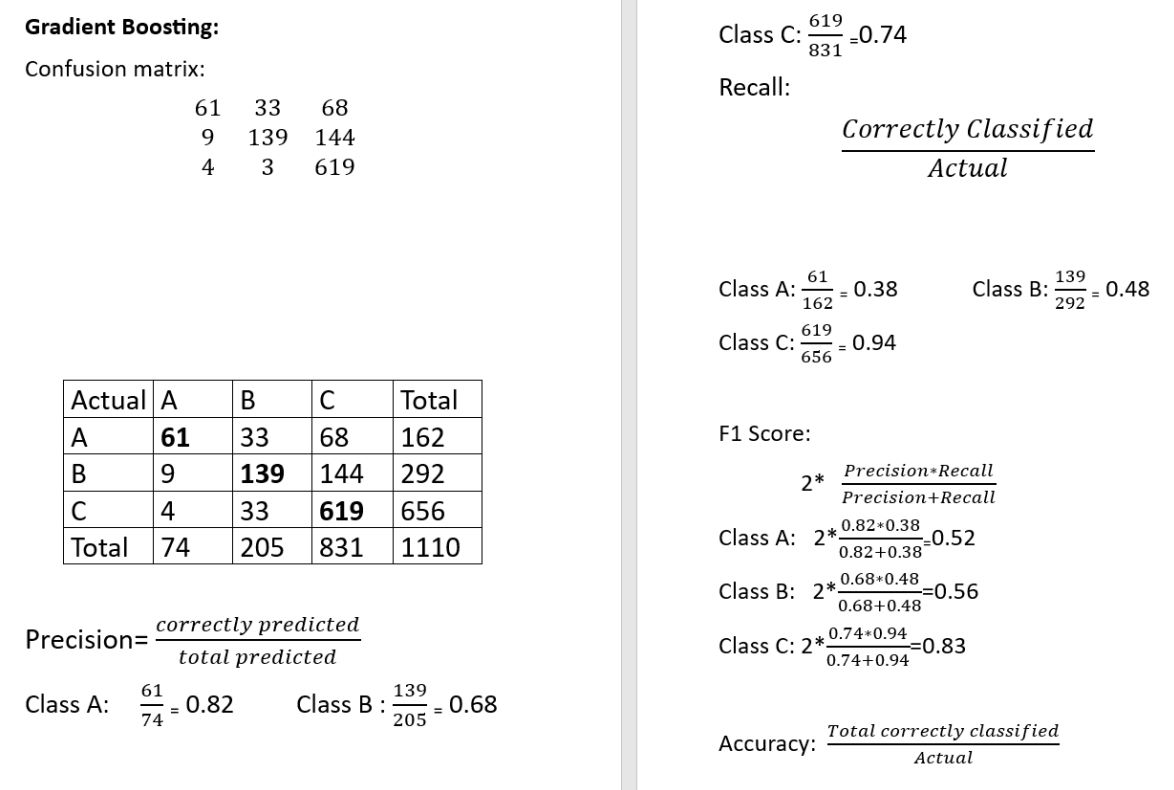


Fig:22

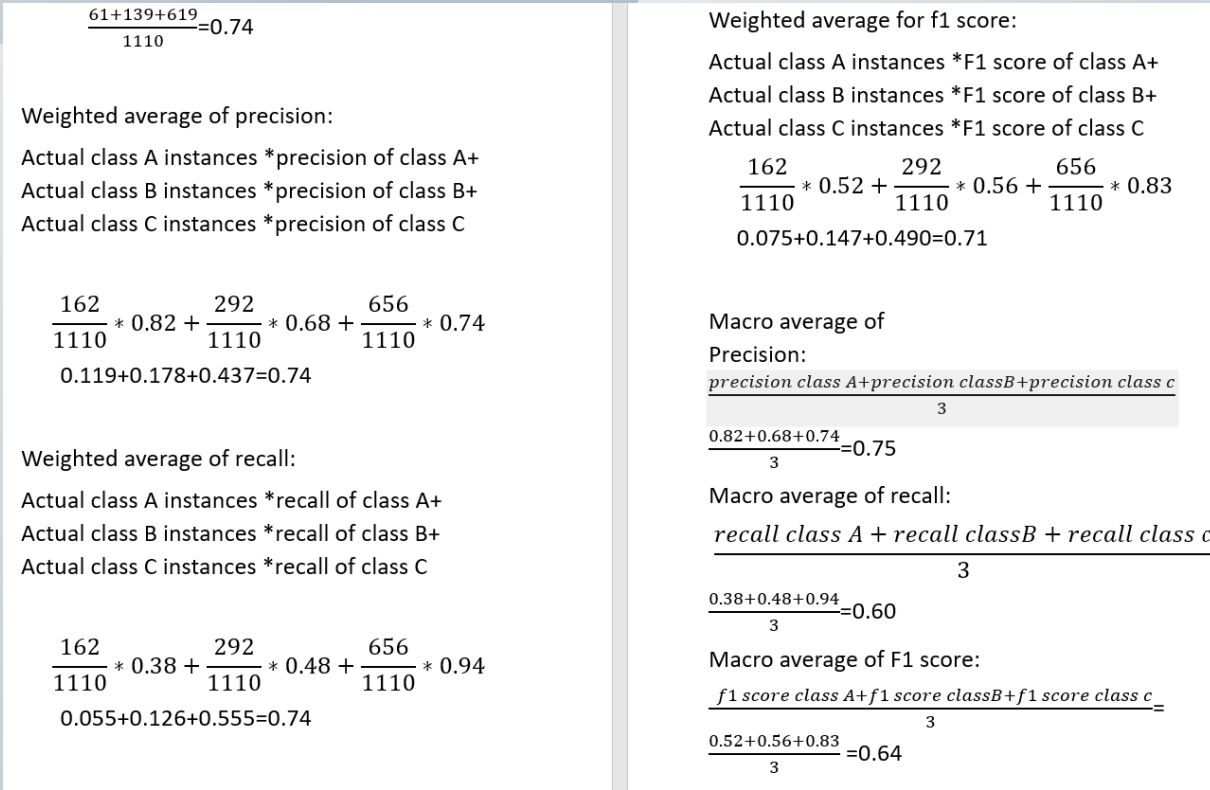


Fig:23

* 1. **Novel Multi-layer perceptron**

# 

Fig:24

# We got same and accurate values in both execution and manual calculations. We got highest

# accuracy for novel multi-layer perceptron.

# 

# Output

# KNN Output

# 

# Fig:25

# Support vector machine

# 

# Fig:26

# Random Forest classifier

# 

# Fig:27

# Decision Tree classifier

# 

# Fig:28

# 

# Gradient Boosting classifier

# 

# Fig:29

# Novel Multi- layer perceptron

# 

# Fig:30

# CONCLUSION

# In this study we used deep learning concept which we did not used in our mini project. This makes our project to stood up in different level and it also gave us the knowledge of deep learning techniques. The technique that we have used in our major project is Novel Multi-Layer Perceptron. This gave us best accuracy compared to machine learning classifiers although we have used them but we just compared between the ML and DL classifiers.

# We like to conclude that the Novel Multi-Layer Perception used in this study to detect counterfeit

# of reviews on applications. The Novel Multi-Layer Perception has an accuracy score of 77.00%

# and next we got highest accuracy is gradient boosting classifier Compared to gradient boosting,

# Novel Multi-Layer Perception appears to be more accurate at spotting forgery detection.

# The forgery Analysis detection is made for the purpose of removing fake reviews. Novel Multilayer Perceptron classifier outperformed gradient boosting classifier in this study's testing dataset in terms of classification accuracy.

# Future Work

# We look forward to use more parameters in our dataset to improve our model so that.It can predict more accurately To use a real time dataset which will allow us to compare the user’s timestamps of the reviews to find if a certain user is posting too many reviews in a short period of time. To use and compare other machine learning algorithms like logistic regression to extend the research

# to deep learning techniques.

# To build real time models with the help of Django frame work along with deep learning

# We look forward by implementing different preprocessing techniques which helps to make model more accurate and user friendly.

# REFERENCES

Al-Sultany, Ghaidaa A., and Saba Mohammed Hussain. 2019. “Fake Reviews Detection through Users Behavior Analysis.” Journal of Advanced Research in Dynamical and Control Systems. <https://doi.org/10.5373/jardcs/v11sp10/20192864>.

Banerjee, Snehasish, and Alton Y. K. Chua. 2019. “Toward a Theoretical Model of Authentic and Fake User-Generated Online Reviews.” Advances in Media, Entertainment, and the Arts.

<https://doi.org/10.4018/978-1-5225-8535-0.ch007>.

Bhavitha, K. V., and S. John Justin Thangaraj. 2022. “Novel Detection of Accurate Spam Content Using Logistic Regression Algorithm Compared with Gaussian Algorithm.” In 2022 International Conference on Business Analytics for Technology and Security (ICBATS). IEEE.

<https://doi.org/10.1109/icbats54253.2022.9759003>.

Chiluwa, Innocent E., and Sergei A. Samoilenko. 2019. Handbook of Research on Deception, Fake News, and Misinformation Online. IGI Global.   
Deng, Yuchao. 2022. “Image Forgery Detection Using Deep Learning Framework.” 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE). <https://doi.org/10.1109/iciscae55891.2022.9927668>.

“Figure 1: Evolution of Searches about Fake News and Fake Reviews.” n.d.

<https://doi.org/10.7717/peerjcs.219/fig-1>.

Kaur, Ramandeep. 2016. “Image Forgery and Detection of Copy Move Forgery in Digital Images: A Survey of Recent Forgery Detection Techniques.” International Journal of Computer Applications. <https://doi.org/10.5120/ijca2016909164>.

Kingsnorth, Simon. 2016. Digital Marketing Strategy: An Integrated Approach to Online Marketing. Kogan PagePublishers.   
K, Kiranashree B., and B. K. Kiranashree. 2020. “Survey on Image Forgery Detection Using Machine Learning Classifier.” International Journal for Research in Applied Science and Engineering Technology. <https://doi.org/10.22214/ijraset.2020.6330.>

**All the links related to project:**

**Data set link:**

<https://www.kaggle.com/datasets/omkarsabnis/yelp-reviews-dataset>

**Codelink:** <https://colab.research.google.com/drive/1SQcJz4MKqsEqWOYDYo7HkSyhWFLRMGN7?authuser=1>

**Drivelink:**

[https://drive.google.com/drive/folders/1Mi3Bghn\_admGhAfbMmQ6pMgW-tHSqg9u?usp=sharing](%20https:/drive.google.com/drive/folders/1Mi3Bghn_admGhAfbMmQ6pMgW-tHSqg9u?usp=sharing)