

# FAKE NEWS DETECTION USING MACHINE LEARNING EMPLOYING DECISION TREE ALGORITHM

Dr.M. RAJESWARI<sup>1</sup>

Assistant Professor,

Department of B. Com (Business Analytics)

PSGR Krishnammal College for Women Coimbatore, India.

[rajeswarim@psgrkcw.ac.in](mailto:rajeswarim@psgrkcw.ac.in)

G. SAKTHI<sup>2</sup>

UG Scholar,

Department of B. Com (Business Analytics)

PSGR Krishnammal College for Women Coimbatore, India.

[sakthiopalkrishnan01@gmail.com](mailto:sakthiopalkrishnan01@gmail.com)

## ABSTRACT

These days a lot of information is being shared over social media and we are not able to differentiate between which information is fake and which is real. People immediately start expressing their concern or sharing their opinion as soon as they come across a post, without verifying its authenticity. This further results in spreading of it. Fake news and rumors are the most popular forms of false and unauthenticated information and should be detected as soon as possible for avoiding their dramatic consequences. This paper is a review and comprehensive analysis of the articles in recent literatures which were about detecting fake news over social media. The aim of this work is to create a system or model that can use the data of past news reports and predict the chances of a news report being fake or not. Various researchers have attempted solving this challenge in a multitude of ways to test which method works and get desirable results.

**Key Words:** Multi-layer Perceptron, Accuracy Level, Prediction and Analyzing

## I. INTRODUCTION

The main purpose of the project is to detect the fake news in social media under the problem of Russia vs Ukraine. World is changing rapidly. No doubt we have a number of advantages of this digital world but it also has its disadvantages as well. There are different issues in this digital world. One of them is fake news. Someone can easily spread a fake news. Fake news is spread to harm the reputation of a person or an organization. It can be a propaganda against someone that can be a political party or an organization. There are different online platforms where the person can spread the fake news. This includes the Facebook, Twitter etc. Machine learning is the part of artificial intelligence that helps in making the systems that can learn and perform different actions. A variety of machine learning algorithms are available that include the supervised, unsupervised, reinforcement machine learning algorithms. The algorithms first have to be trained with a data set called train data set. After the training, these algorithms can be used to perform different tasks. Machine learning is using in different sectors to perform different tasks. Most of the time machine learning algorithms are used for prediction purpose or to detect something that is hidden.

Fake News contains misleading information that could be checked. This maintains lie about a certain statistic in a country or exaggerated cost of certain services for a country, which may arise unrest for some countries like in Arabic spring. There are organizations, like the House of Commons and the Crosscheck project, trying to deal with issues as confirming authors are accountable. However, their scope is so limited because they depend on human manual detection, in a globe with millions of articles either removed or being published every minute, this cannot be accountable or feasible manually. A solution could be, by the development of a system to provide a credible automated index scoring, or rating for credibility of different publishers, and news context.

## II. OBJECTIVE

To find the accurate level of the Fake News spread using the Decision Tree Algorithm.

## III. RELATED WORKS

Fake news detection research appeared for a couple of years and is a relatively new and difficult research field. The difficulties come from the semantics of natural languages and manual identification via human beings, let along machines. In this project, we propose to analyze the performance of several machine learning algorithms integrating tools such as Fake News Tracker [1], doc2vec, Support Vector Machine (SVM) and decision trees. Our preliminary results indicate that the SVM and the Decision Tree are suitable to identify fake news with an acceptable accuracy of 95 percent. Typically, the Decision Tree method shows a better result than SVM. [1]

This research proposed utilizing two different machine learning algorithms (random forest and decision tree (J48)) to detect the fake news. In this paper, the full dataset size equals 20,761 samples, while the testing sample size equals 4,345 samples. The pre-processing steps start with cleaning data by removing unnecessary special characters, numbers, English letters, and white spaces, and finally, removing stop words is implemented. After that, the most popular feature extraction method (TF-IDF) is used before applying the two suggested classification algorithms. The results show that the best accuracy achieved equals 89.11% using the decision tree model while using the random forest; the accuracy achieved equals 84.97 %. [2]

This paper proposes a machine learning approach to identify fake reviews. In addition to the features extraction process of the reviews, this paper applies several features engineering to extract various behaviours of the reviewers. The paper compares the performance of several experiments done on a real Yelp dataset of restaurants reviews with and without features extracted from user's behaviours. In both cases, we compare the performance of several classifiers; KNN, Naive Bayes (NB), SVM, Logistic Regression and Random Forest. [3]

In this survey, we focus on one particular data mining algorithm -- decision trees -- and how differential privacy interacts with each of the components that constitute decision tree algorithms. We analyse both greedy and random decision trees, and the conflicts that arise when trying to balance privacy requirements with the accuracy of the model. [4]

This paper aims to use machine learning algorithms to detect fake news. We compare various classification algorithms, namely Logistic Regression, Multinomial Naïve Bayes, Linear SVC, Decision Tree, and K-Nearest Neighbours and find out which algorithm works best in detecting fake news correctly. We will be using different percentages of data from the same dataset and determine which algorithm gives best results when 1%, 25%, 50%, 75%, and 100% of the data is available for training and testing. [5]

Nowadays, most researchers focus on creating a model that is capable of accurate prediction to distinguish whether a particular article is classified as real or false news. It is necessary to determine what makes the new site "legitimate" and define it objectively. The harmful effects of inaccurate information will make people believe that Hillary Clinton has a foreign child, trying to convince readers that President Trump is trying to cancel the first amendment to kill India's crowds because false rumours spread in WhatsApp application. [6]

The experiments conducted in this study evaluated the results using F1-Measure and compared them to results obtained via other existing methods, such as utilizing stemmers and part-of-speech taggers, where it indicated an increment of more than 12.6% for the novel method using semantic relation over other methods. [7]

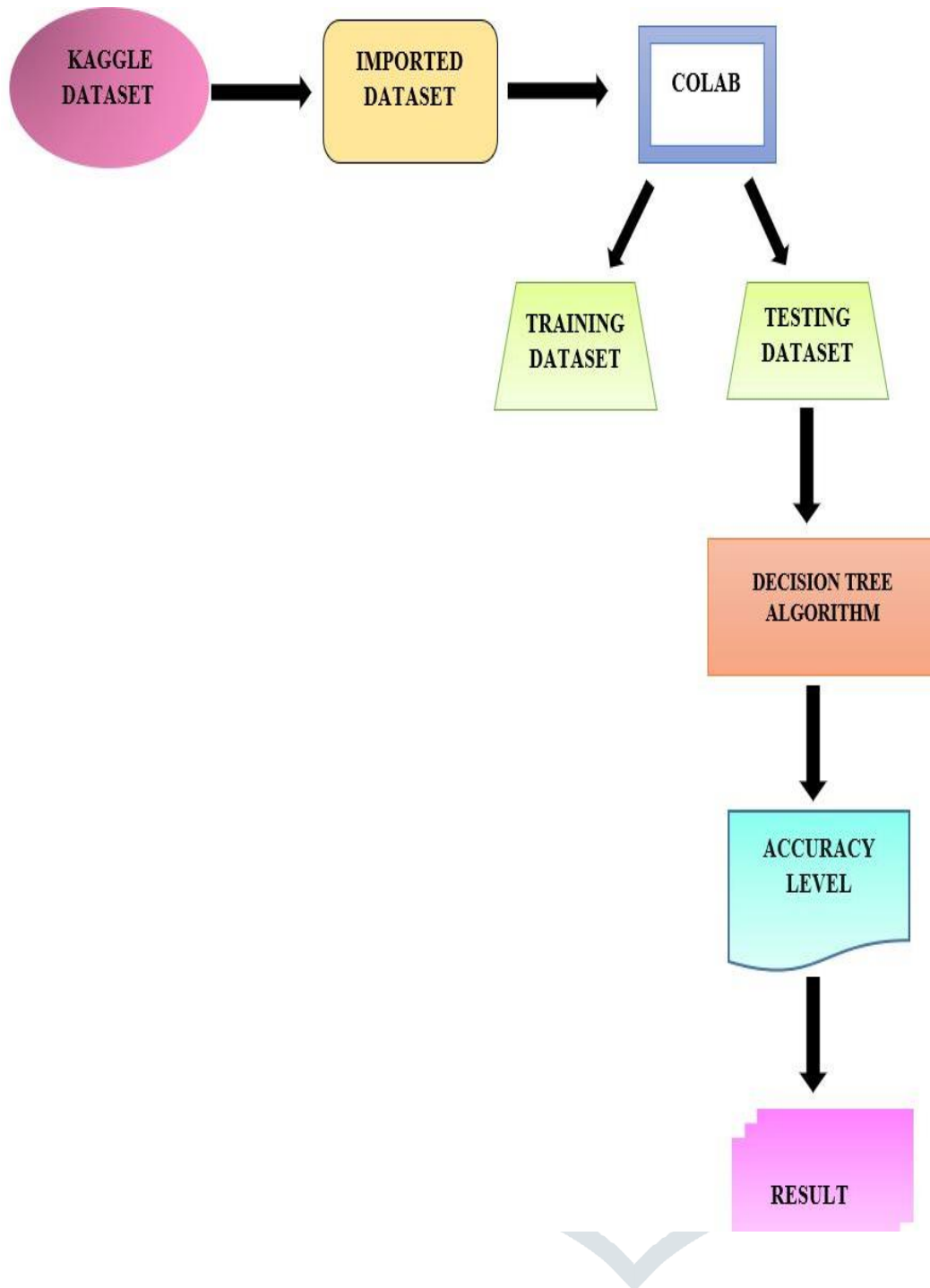
## IV. METHODOLOGY

### A. DECISION TREE:

One of the most common algorithms used in classification is the J48 algorithm. It is based on the C4.5 algorithm in which all the data to be studied must be of the type numeric and categorical kind. Therefore, a continuous type of data will not be examined [15,16,17]. J48 utilizes two different pruning ways. The first method, named subtree replacement, which denotes the possibility of replacement nodes in a decision tree with its leaves to minimize the number of tests in the convinced path. Usually, the subtree raising is of a modest impact on the models of the decision tree. Typically, there is no exact way to predict an option's utility, although it can be advisable to turn it off when the induction procedure takes longer because of the subtree's raising being relatively computationally complicated.

- 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.
- In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.
- The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions.

**B. FLOWCHART**



**V. RESULT**

```

from sklearn.tree import DecisionTreeClassifier

# vectorizing and applying TF-IDF
pipe = Pipeline([('vect', CountVecToBae()),
                 ('tfidf', TfidfTransformer()),
                 ('model', DecisionTreeClassifier(criterion='entropy',
                                                max_depth=20,
                                                splitter='best',
                                                random_state=42))])

# fitting the model
model = pipe.fit(X_train, y_train)

# Accuracy
prediction = model.predict(X_test)
print("accuracy: {}".format(round(accuracy_score(y_test, prediction)*100,2)))
del(['decision tree'] = round(accuracy_score(y_test, prediction)*100,2)

accuracy: 99.26%
  
```

**FIG 1**

From the above diagram we had run the program of Fake News Detection by using Decision Tree Algorithm.



FIG 2

The output is been visualized by plotting with the use of Decision Tree Algorithm and hence verified successfully.

## VI. CONCLUSION

In these days more people are continuously consuming news from social media rather than the traditional media. This fake news develops a strong negative impact on individual users and the society. Therefore, for detecting the fake news we analyze different research papers and identifies Word Embedding, Tokenization and Parts of speech tagging are best for Pre-Processing of data and also identifies TF-IDF and Count Vectorizer are best for feature extraction. So, for better approach Further we want to use those methods for Pre-Processing, feature extraction and also, we want to implement the Random Forest classifier, Convolutional Neural Networks, Long Short-Term Memory for high accuracy and an Ensemble Learning Approach for high accuracy.

## REFERENCE

- [1] Ahmed M. Elmogy<sup>1</sup>, Usman Tariq<sup>2</sup>, Atef Ibrahim<sup>4</sup> Fake Reviews Detection using Supervised Machine Learning College of Computer Engineering and Sciences Prince Sat tam Bin Abdul-Aziz University, KSA<sup>1,2,4</sup> Faculty of Eng., Tanta University, Egypt<sup>1</sup>.
- [2] Jehad Ali, R. K., Nasir Ahmad, Imran Maqsood Fake News Classification Using Random Forest and Decision Tree (J48).
- [3] Sam F., "Decision Tree Classification with Differential Privacy: A Survey", 2019.
- [4] Sarthak Vats Comparing classification algorithms for fake news detection on different amounts of data from same dataset ISSN No: 1006-7930.
- [5] Shikun\_Lyu, D.Lo Fake News Detection by Decision Tree Computer Science Southeast on 28 March 2020
- [6] Suhad A. Yousif, Hussam Y. Abdul-Wahed, Nadia M. G. Al-Saidi, "Extracting a new fractal and semi-variance attributes for texture images", 2019.
- [7] Suhad A. Yousif, Islam Elkabani, "The Effect of Combining Different Semantic Relations on Arabic Text Classification", Vol. 5, No. 6, 112-118, 2015.