FAKE NEWS DETECTION USING MACHINE LEARNING ALGORITHMS

Utkarsh Meena

Computer Science and Engineering

Amity University

Noida, Uttar Pradesh, India utkarshmeena03@gmail.com

*Abstract***— *Fake news came in lime light few years back with the evolvement of era of internet and social media. Social media gave it a powerful base to establish and it created problem for every researcher. After Artificial intelligence came into action, it has created a new level of approach and with the help of machine learning detection of fake news became easy. In this report we talked about these algorithm and their role. Algorithms like NLP, Decision trees, KNN, SVM, N-Gram analysis, Naïve bays classifier. Out of these KNN showed the most amount of accuracy but if we take the average of all accuracy it came out to be 79.83% which is remarkable and can be improved by adding more algorithm and testing to train the model.***

***Keywords—* *machine learning, mental health, covid-19, cognitive learning.***

# INTRODUCTION

Everything in today’s time works with the help of internet and social media also plays a very important part in this. Now a day’s mass media also holds huge stakes to present themselves on social media and brings a challenge to world with fake news occurrence on large rate on social media. These fake news are very dangerous but many organizations do it intentionally to make profit out of these. Platforms like Facebook, Twitter, YouTube came across many fake news articles daily but they all track them very easily with the help of artificial intelligence.

So therefore experts and researchers have used artificial intelligence to tackle this issue. This decade is going to be the most important one for the technology to reach huge heights in every field. Tech in real life applications has made everyone dependent on it, and why not as it has made our work easy and fast. Technology like artificial intelligence will play an important role in everyday life. Artificial intelligence is the most talked topic now a days and experts believe that it will definitely impact positively on world economy.

Artificial intelligence is a branch of computer science which deals with enabling a machine/computer for intelligent decision making which typically requires human-level intelligence. Its sub set is machine learning and deep learning. The process were the computer is learning ON ITS OWN from EXPERIENCE is known as machine learning. The machine learning algorithms plays a very important role in determining the fake news by its algorithms. Applications

# XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

of machine learning like NLP [natural language processing], Computer vision etc. The machine learning techniques which are used in determining fake news are NLP natural language processing, decision trees, neural networks etc. In 2020, the whole world is going through a biggest recession with the impact of corona virus epidemic. The pandemic costed many lives due to many reasons and one of the reason is fake news on this epidemic which resulted worst. The fake news on pandemic like “we can go out without masks “,”this virus do not last in higher temperature “,”or that it will end in march ”.These news were fake and created stampede all around the world and changed people thinking about the virus.

# METHODOLOGY

Fake news are creating lots of fuss all around the world and to cope with it machine learning has come with its intelligent algorithms to detect their sources. Machine learning is a part of Artificial intelligence which predicts or learn from its input data and train its model to give high accuracy. Machine Learning algorithms are made in such a manner that they train data and give their own result which is somewhat similar and accurate.

There are five main steps for building a model with artificial intelligence. 1. Problem scoping

1. Data extraction
2. Data exploration
3. Model implementation
4. Evaluation

In problem scoping we answer the five question “why, who, where, what”. In data extraction we collect relevant data and use it. In data exploration we explore data and divide it into specific domain. Then we find a perfect model for project which will provide accurate results. After that model evaluate it self and learn from its own actions and present result with more accuracy.

# DATA COLLECTION

Machine Learning manipulate data provided to it and then learns from that data, so data collection becomes one of the most important step in building an model.

In case of fake news, they are formed in each and every sector for sake of earning profit and defaming others. They come from sources or organisation which earn their face values through it and make money out of it. If we take example of politics each and every party create such news of their opposition to defame them with the help of media which promote these kind of articles worldwide. So extracting of data is done by two methods by survey or by searching articles from well-known organization. In this model, data of articles from world-wide news websites are collected and then explored by the means of graphs ,tables, bars so that machine works more accurately. These articles are inserted into the machine and algorithms works on them.

**ALGORITHM USED**

**Some major algorithms which were used in are explained below:-**

1. Support Vector Machines- A assist vector machine (SVM), which can be used interchangeably with a guide vector network (SVN), is also taken into consideration to be a supervised learning algorithm. SVMs work via being trained with specific facts already prepared into one of a kind categories. Hence, the model is constructed after it has already been educated. Furthermore, the goal of the SVM method is to distinguish which class any new facts falls under, in addition, it must also maximize the margin between the two classes. The optimal goal is that the SVM will discover a hyperplane that divides the

dataset into groups. (1)

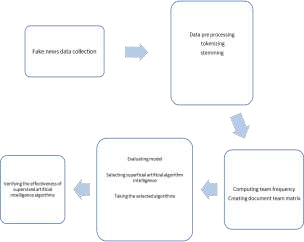
1. Naïve Bayes- It is a conditional possibility model which may be used for labelling. The purpose is to discover a way to expect the class variable (B) using a vector of independent variables (A), i.e., locating the characteristic f: A-->B. In chance terms, the goal is to find P (B) opportunity of B belonging to a certain magnificence A. B is typically assumed to be a categorical variable with or greater discrete values. It is a mathematically simple manner to consist of contributions of many factors in predicting the elegance of the next facts instance inside the testing set. The problem of Naïve Bayes is that they count on that all features are not established on every other. (1)

1. Decision Trees – They have a predefined target variable. In terms of their shape, they provide a top-down strategy. A decision tree is a shape used to divide a statistics set containing a massive quantity of information into smaller clusters via applying a fixed of decision rules. In different words, it's far a shape utilized by dividing big amounts of facts into very small agencies of facts by applying simple decision-making steps.

1. KNN- [K-Nearest Neighbours algorithm] this method is used for classification and regression types. These two regression and classification are most unique techniques which helps a lot for making results.

1. NLP [Natural Language Processing] - The most interesting algorithm for someone to work upon and provides a very good base of learning. This algorithm works by analysing letters, words, and sentences and then give review on it. In case of fake news detection it goes through the articles and tells it authentication.

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Advantages** | **Disadvantages** |
| SVM | Delivers a unique output,  More efficient, Scale-up with highdimensional data. | Lack of transparency of results, Difficult to select suitable kernel data. |
| KNN | Calculation time is very quick,  Uses both calculation and regression, New data can be added seamlessly. | Need feature scaling, Computation is very high,  Large storage requirements. |
| NAÏVE BAYES | Easy for massive  datasets,  Handles discrete and continuous data, Used for both binary and multiclassification, Not sensitive to irrelevant features. | Computationally intensive especially, for models including many variables,  Sometimes, models that are appropriately trained and tuned outperform Naïve Bayes models as they are too simple. |
| RANDOM FOREST | Used for regression and classification problem,  Solve overfitting problem in the decision tree. | The Prediction rate is relatively slow in complex Random forest problems. |
| GBM | No data preprocessing required,  Predictive accuracy. | Its implementation may be more difficult due to the higher computational demand. |
| LOGISTIC  REGRESS  ION | Computational efficiency,  Ease of regularization, For input features, no scaling is required. | Difficult to solve a nonlinear problem,  Overfitting issues. |



## LITERATURE REVIEW

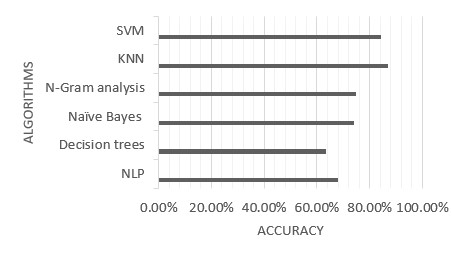
In this section, several research articles have been reviewed and assessed, out of which, few of them are related to mental health and Covid-19. The articles have been grouped as per the type of mental illness, the ML technique used to the kind of illness, and the algorithm's accuracy. Table II shows the review results conducted.

|  |
| --- |
| **Results** |
| Shlok Gilda (1),Pune Institute of Computer Technology in her journal in 2017, has applied NLP [Natural language processing] as her technique to detect fake news all over the internet. She applied this machine learning algorithm and came up to an accuracy of 67.89% through naïve model.  Dataset used in this research was from different open sources. |
| Mykhailo Granik (2), Computer Science Department Vinnytsia National Technical University Vinnytsia, Ukraine in their journal in 2017, has applied naïve Bayes classifier model to detect fake news. The accuracy of 74% using such a basic model was a huge success for them as by using basic mathematical theorem they were accomplished. This gave a big break through that basic Artificial Intelligence can also help in finding accurate results. |
| Feyza Altunbey Ozbay (3), Department of Software Engineering, Faculty of Engineering, Firat University in their journal has proposed the use of supervised artificial intelligence algorithms to achieve a result. They use algorithms like decision trees, ZeroR and more in their model. They quoted that their precision may be improved if more accurate algorithms are used in this model. |
| Devyani Keskar (4), in their journal, used algorithms like decision trees, N-gram analysis, flume with an precision of 75%. They used flume to extract a sufficient amount of data of autherised resources. |
| Hadeer Ahmed1 (5),in their journal obtained the exceptional accuracy of 92 % by using N-gram analysis and some algorithms of artificial intelligence. They also used two terms TF-IDF (term frequency – inverted document frequency) and LSVM (linear support vector machine). |
| Shubham Bauskar (6), in their research paper used NLP [natural language processing] as their main machine learning algorithm and proposed a model which gave a accuracy of 90.63%. they quoted that by adding more authentic material to the model. |
| Lilapati Waikhoma (7), in their journal in 2019 used technique N- Gram analysis and obtained a an average accuracy of 75%. |

RESULT

The algorithms used in this model of fake news detection shows amazing result and remarkable precision in their performances. It is seen that every algorithm with same data shown more than 60% of accuracy rate.

All the algorithm works with highest precision but KNN with this data set worked very well and got more than 90 percent of accuracy.



## REFERENCES

1. *A Study on Fake News Detection Using Naïve Bayes, SVM, Neural Networks and LSTM.* Tijare, Poonam. 2019, Jour of Adv Research in Dynamical & Control Systems.
2. *Evaluating Machine Learning Algorithms for Fake News Detection.* Gilda, S. 2017, IEEE 15th Student Conference on Research and Development (SCOReD), pp. 110-115.
3. *Fake news detection using naive Bayes classifier.* Granik, M., & Mesyura, V. 2017, IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903.
4. *Fake news detection within online social media using supervised artificial intelligence algorithms.* Ozbay, F. A., & Alatas, B. 2020, Physica A: Statistical Mechanics and its Applications.
5. *Fake News Classification on Twitter Using Flume, N-Gram Analysis, and Decision Tree Machine Learning Technique.* Keskar, D., Palwe, S., & Gupta, A. 2020, International Conference on Computational Science and Applications, pp. 139-147. 6. *Detection of online fake news using N-gram analysis and machine learning techniques.* Ahmed, H., Traore, I., & Saad, S. 2017, October, In International Conference on Intelligent,

Secure, and Dependable Systems in Distributed and Cloud Environments , pp. 127-138.

1. *Natural Language Processing based Hybrid Model for*

*Detecting Fake News Using Content-Based Features and Social Features.* Bauskar, S., Badole, V., Jain, P., & Chawla, M. 2019, International Journal of Information Engineering and Electronic .

1. *Fake News Detection Using Machine Learning.* Waikhom,

Lilapati and Goswami, Rajat Subhra. 2019, International Conference on Advancements in Computing & Management (ICACM) 2019.