International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(16s) 2026 pp. 842-864
https://doi.org/10.70917/ijcisim-2026-4518
Article

Social Media Fake news detection Utilizing a variety of datasets with probabilistic latent semantic analysis & k-means algorithms

Shubha Mishra[1]

[1]Department of Center for AI, MITS DU Gwalior
Email: mis.shubha@gmail.com

Abstract: Social media has become usefull and most important platform for individuals to access news due to its speed and cost-effectiveness of disseminating information on a particular channel. However, these platforms also make it a breeding ground for the spread of fake news, which can impact society and individuals. With the advent of modern technology that comes with the fourth industrial revolution, promoting openness and increased engagement, social media has evolved to serve multiple purposes. It has become an integral part of our lives, beyond what was originally intended for just a small group of people. Consequently, identifying such fake news has become a crucial task for researchers and remains a major concern. This paper aims to examine the methods of publishing and distributing fake news, and outlines the approach of classifying, organizing, and developing algorithms. Our proposed solution is called the "Fake News Detection using Hybrid-kMeans (FNDHKM)" algorithm that utilizes Natural Language Processing (NLP) and machine learning (ML) techniques to detect fake news on social media platforms, specifically Twitter. The experiments were performed on three different datasets obtained from Twitter and involved dimensionality reduction on the extracted data. The highest accuracy achieved was 85% and 90% for precision and accuracy, respectively. The proposed FNDHKM approach showed significantly high accuracy with low overhead.
Keywords: Machine Learning, k-means Clustering, Deepfake Analysis, Deep Learning, fake news detection, Cluster Computing, Natural Language Processing (NLP),hybrid k-Means, fourth industrial transformation

1. INTRODUCTION

Social media provides numerous benefits to people globally, connecting those who are geographically separated and serving as a tool for activists to organize. However, it also has its downsides, such as instances of bullying, division caused by political entities, and the spread of false news. These "dark sides" cannot be ignored and pose challenges to the positive aspects of social media.Companies like Facebook, Google, and Twitter allow users to post content, including false information that can spread rapidly without proper verification.Fake news has become a widespread problem, particularly in Europe and among the business community, and can have a significant impact on people's opinions and decision-making [10].People have been used to a system that provides them with phoney news, and as a result, they suffer pressure.

1.1 Social Media Fake News Detection

This is followed later by the Look Motor Control Impact and the Look Proposal Impact. This has resulted in the advancement for the term “post-truth,” which in 2016 was followed by referrals to the Oxford Dictionary’s entry for the term. [1]. From false information, it is recognised as the component that is presented as a small amount of real news; nonetheless, it is incorrect most of the time or fully. [2] [3] [4]. The primary reason why firms called Facebook

were featured in false news was utilised to speed up the social status in media. This was followed by the propagation of the news that was phoney, which further grew dangerously. [5]. Nearly every social media hub provided an option for data sharing, encouraging the dissemination of that content for a few specified purposes. The locations of social media platforms provide a fast and easy option for sharing information; this means that data may

be transmitted without restrictions of either time or space. There is a growing awareness among the citizens of Cambridge, and they are working toward the goal of putting an end to the spread of false news. [6]. Various biometric system assessment parameters may be used to evaluate the data and information about false news utilising parameters as measurements. A select few of the assessment parameters for biometric systems are outlined as follows:

FAR:The False Acceptance Rate (FAR) [40] refers to the probability that the system wrongly recognizes an unauthorized individual by comparing the biometric input with a template. This occurs when a team reaches the biometric input with a layout.

$$FAR(\eta) = \frac{false\_accept(\eta)}{length(impsc)} \text{.....(1)}$$

FRR:The False Rejection Rate (FRR) [40], also known as the False Negative Rate, is the probability that the system wrongly denies access to an authorized individual by failing to match the biometric input with the template. This happens when an authorised individual tries to coordinate the biometric input with a design.

$$FRR(\eta) = \frac{false\_reject(\eta)}{length(gensc)} \text{..... (2)}$$

EER:The Equal Error Rate (EER) [40] is the value at which the FAR and FRR are equivalent. It signifies the point where the system's sensitivity is balanced in such a way that both FAR and FRR are minimized.

$$EER(\eta) = \frac{FAR(\eta_1)+FRR(\eta_1)}{2} \text{.....(3)}$$

GAR:The Genuine Acceptance Rate (GAR) [41] is a measure used to evaluate the performance of a biometric authentication system, as an alternative to FRR. The GAR represents the likelihood that a biometric authentication system accurately accepts an authorized individual based on a comparison between the biometric input and a stored template.

$$GAR = 1 - FAR(\eta_1) \text{.....(4)}$$

DI:Decidability Index [42] is an excellent measure to calculate separate genuine and imposter classes.

$$DI = \frac{|\mu_g|-|\mu_i|}{\sqrt{\sigma_g^2+\sigma_i^2}} \text{.....(5)}$$

The Social Media Fake News Detection sections will be presented in a structured format as follows: The subsequent section conducts an analysis of the methods discussed in previous sections and presents the obtained results. Further clarification on the implementation of the calculations can be found in Section III. The experimental study is presented in Section IV. The results of tests conducted on benchmark datasets are comprehensively discussed using Accuracy and Recall metrics in Section V. the conclusions are drawn in Section VI.

2. RELATED WORK

Coordinated techniques are widely used for identifying fake news at present. These algorithms were used to classify distinct thoughts for different collections of data known as news content [7], user profiles [8], information propagation [9], and various settings [10]. Despite their promising results, there are certain limitations that should be taken into account, such as the need for pre-annotated datasets for classification and the increase in time and overhead due to meticulous examination of genuine and fake news reports. However, there are ways to minimize the stack

related to manual checking and improve the quality of classification through perseverance. (El-Latif, Quantum-Inspired Blockchain-Based Cybersecurity: Securing Smart Edge Utilities in IoT-Based Smart Cities,

2021)[11].

The main reason behind fake news being spread in society is to mislead the news readers, and some individuals are unable to distinguish between genuine and fake news due to a lack of expertise [12][13].Unsupervised learning has emerged as an alternative to supervised learning in the effort to identify fake news. The reason behind this is the ability to gather information on reader reactions to fake news published on social media. These reactions allow researchers to understand the behavior and mindset of the readers. As seen in [14], a study was conducted and showed a remarkable 90% accuracy in detecting false news. The study also managed to categorize false news into distinct groups. To address the problem of fake news, a modular approach was taken where the news was analyzed and processed to verify its authenticity.The proposed solution involves using a module for data analysis and processing to determine the truthfulness of the news. On the other hand, Akil.io. Aldwairi [15] provides a method for checking the effectiveness of clickbait. Helmstetter [16] proposed a weakly supervised technique that collects a large scale of data. The data collected consists of both trustworthy and non-trustworthy titles, and this data is used to generate a list of classifiers. By using these classifiers, it is possible to differentiate between fake and authentic news. However, it is important to note that not all news from misleading sources is fake, and classification should be done based on factors such as the impact of the news and the category.In the study by Yang [17], an unsupervised learning method was proposed to identify fake news by considering the trustworthiness of the news and the legitimacy of the readers as static variables. The methodology also takes into account the user's behavior on such news and analyzes their fake news. Bayesian orchestration is used to establish the conditional dependence between the news trustworthiness, the client's perception, and the user's legitimacy. The Gibbs sampling approach, known as the collapsed approach of assessment, is used to infer the news genuineness and the client's legitimacy with unlabeled data. During later times, Nicollas [18] used a computational approach based on NLP and Positive Naive Bayes (PNB) to identify fake news obtained from social media. The method analyzed the text content and sentiments expressed in the news and compared them with the information available in reliable sources to determine the authenticity of the news. The results showed that the proposed method was able to effectively identify fake news with high accuracy.Denis utilized Positive Naive Bayes (PNB) [29] and applied it to the Unlabelled Positive (PU) concept related to knowledge, linking it to the classification and functioning of the content.The Posterior Probability of the Positive Naive Bayes (PNB) Model:

Given certain x

$$P(X = x) \propto P(C = c) \prod_{i=1}^{n} \quad P(C = c) \text{.....(6)}$$

Conditional probability for uncertain instance x^{u_e}

$$P(X^{u_e} = x^{u_e}) \propto P(C = c) \prod_{i=1}^{n} \quad \sum_{k=1}^{|V(x_i{}^{u}c)|} \quad P(C = c) P_{ik} \text{..... ...(7)}$$

V is the probability vector here,

For certain data $P(0)$

$$P(0) = \frac{N(x_{ij}, U) - P(x_{ij}(1)|D|P - |P|)}{|D|(1-P)} \text{.....(8)}$$

$|D| \rightarrow$ Cardinality of the dataset

$P \rightarrow$ Positive examples

$N \rightarrow$ Negative examples

$U \rightarrow$ Hidden positive samples/examples

The equations [30] are explained below, which form the basis of our proposed algorithm:

The notations which are shown above are givne in details below (1)-(3) :

$n(d_i)$ shows the length of th document , i.e. token in a certain document . di,

$n(d_i, w_j)$ show how many time the word is repeated d_i,

$P(d_i)$ show the chace that di is related to the cluster k which is given by variable z_k,

$P(z_k)$ shows the word distribution related to the cluster z_k,

$P(z_k|d_i, w_j)$ shows the details of the cluster distributin with reference to a certain set of document di and word w_j.

To conclude clusterk is specific for a particular document $d_i$with respect to its magnitude of $P(d_i)$:

This study aimed to perform dimensionality reduction and data compression on the raw data using passive semantic analysis. After reducing 85% of the features, three different classification approaches were implemented, including two utilizing cascading or unique learning algorithms, and one utilizing statistical analysis to compare the types of news. We have utilized [18] to learn the machine learning process related to news classification, which is prone to errors according to our sources. The proposed algorithm is based on False News Detection using Hybrid-k Means (FNDHKM). Our proposed methodology has been shown to be effective in successfully extracting data and reducing complexity for news classification using Twitter and NLP pre-processing. The experimental results highlight the effectiveness of our categorization approaches in accurately classifying the validity of news. A comparison will then be made between our approach and previously used techniques [18].

3. METHODOLOGY

The current research has established that a novel FNDHKM algorithm was developed to detect false news using a hybrid K-Means approach. This algorithm consists of three distinct steps: the PLSA calculation for dimensionality reduction [19], the classifiers based on Positive Naive Bayes [20, 21], and the hybrid k-means [22] by network adaptation. The focus of this approach is on incorporating innovative techniques, with the key technique being the Expand method. This method is based on [23],[26] NLP pre-processing and leverages data specific to the user, while taking into account recent news. Additionally, the Expand method is dependent on the user receiving updated information.This method is a crucial aspect of the algorithm as it relies on utilizing data specific to the end-user. The end-user does not have access to the parameters of content distribution or the reputation models, and the proposed method addresses this issue by simplifying the process through the use of natural language processing (NLP). The overall structure of the proposed FNDHKM algorithm is depicted in Figure 1.

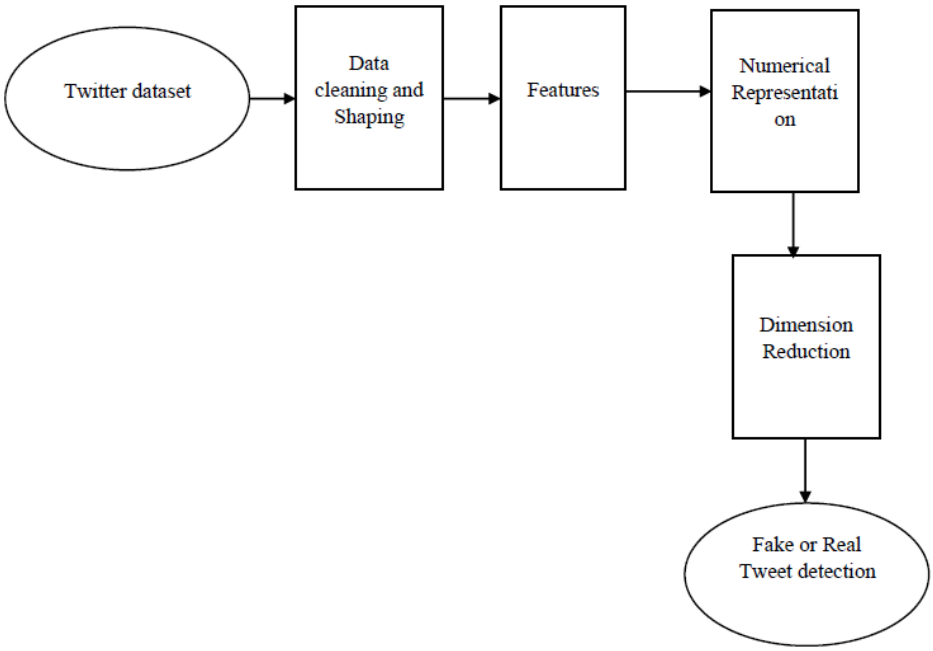


Figure 1: The architecture of the proposed FNDHKM algorithm

A. Collection and Management of Datasets:

This study used the corpus of false information as described in Rubin et al. [13]. The news was collected from Twitter and included both false and authentic reports of recent events. A python script was used to retrieve the most recent news from the Twitter API. However, this method resulted in the elimination of certain details such as the report and social network, and the format could be changed. Despite these limitations, Barreto et al.'s study [13] showed the correlation between the request and the imposed time limit. The use of a specific API has several drawbacks, including a decrease in the number of tweets that can be recorded. The window of opportunity for data collection was between the two time periods, starting from the second month. To overcome the restrictions on data recovery, an alternative is to update the source of the news and collect it from a specific journal. According to experts, this is the preferred approach. Scholarly research has found that the tweets collected from the website "Boatos.org" are unreliable sources of information.

B. Data Set Visualization and Preprocessing: Cleaning and further shaping data [43] are crucial steps in any data extraction plan. They employ language-specific techniques such as checking the spelling of end words and ensuring that the substances are what they appear to be sourced from.

1.Tokenization: After preprocessing, each line of the corpus undergoes tokenization, which converts a bounded sequence into a list of tokens, providing finer control [44]. The string can then represent an event token [24]. This process eliminates distinct spelling variations, such as capitalization and rare characters, from each token.

2.Stop Word Removal: After removing punctuation and special characters, we proceed to remove the stop words, which are the most frequent and less informative words in the corpus, such as articles, pronouns, and conjunctions [45]. This is necessary as the more frequently a word occurs in the corpus, the less information it carries. In the case of tweets, it is also important to exclude specialized Twitter terms, such as possible hashtags, links, and mentions [46][50].

3.Correction of spelling: checking for accurate spelling may be accomplished by comparing the token and its closet writer inside to the information provided in the dictionary[25, 47, 48].

4.. Recognizing named entities: Recognizing named entities involves finding real names in the text and removing them. The Levenshtein distance is used to measure the difference between two strings, in this case, the difference

between words in the dataset and the reference words. The goal is to find the smallest number of operations needed to transform one string into another.

5.Stemming: Stemming is a process of reducing words to their base or root form. The idea behind stemming is to remove affixes (suffixes, prefixes, infixes, etc.) from a word to obtain its root form, which is called the stem. This helps to reduce the dimensionality of the data and to group words based on their meanings, even if they have different forms. The goal of stemming is to normalize the words in a text so that different forms of the same word can be treated as the same for analysis purposes.

6.Numerical representation: The fundamental academic structure of each phrase is stripped away, and the remaining words are presented as the numerical representation. Despite adhering to strict standards, no single phrase would be scientifically useful, as sentences are still composed of word roots rather than measurable values. To create an algebraic representation, we use the vector space model. This illustration demonstrates that phrases can be viewed as arrays of vectors, with each expression capable of being related to a different set of goals. We have obtained evidence using the repeated tf-IDF method, which provides a true measure of the centrality of a particular piece of work within a given phrase in relation to the database as a whole [27][51].

Because many of the restricted words are removed throughout the process, the estimate of the vector is connected to the very high number of restricted terms that are included inside the whole database. The work is employed in a certain area of the phrase with such significance, exposing itself to be fundamental for thoughtfully considering the text's most important notion. Additionally, by utilizing a vector space representation, we are able to perform mathematical operations and manipulations on the data. This is because, in this representation, each word or term in a document is treated as a numerical value, which makes it possible to perform mathematical calculations on them. Furthermore, by using the tf-IDF methodology, we are able to determine the importance or centrality of a particular word or term in a document, relative to the entire corpus. In conclusion, the numerical demonstration of the data enables us to perform meaningful analysis and gain valuable insights into the data set.

7.Dimensionality reduction is a technique used in natural language processing to simplify the complexity of large datasets. There are several methods for dimensionality reduction, including PCA(Principal Component Analysis), LSA (Latent Semantic Analysis), PLSA(Probabilistic Latent Semantic Analysis), Multidimensional Scale, positive LSA, and fast outline.[28][30] PLSA was chosen in this case because it is a type of integrated vector from a natural perspective, and it doesn't over-centralize the data like recent singular value decomposition methods. The flow of activities was checked over to ensure that the data was transformed into the appropriate format for processing by the algorithm. There are three approaches to classifying news stories as fake or real, including classification and unsupervised clustering. The next technique is based on the premise that the news is truthful, given the procedure and conditions. This approach is referred to as the procedure-and-circumstances-based technique. It's seen in the illustration below.

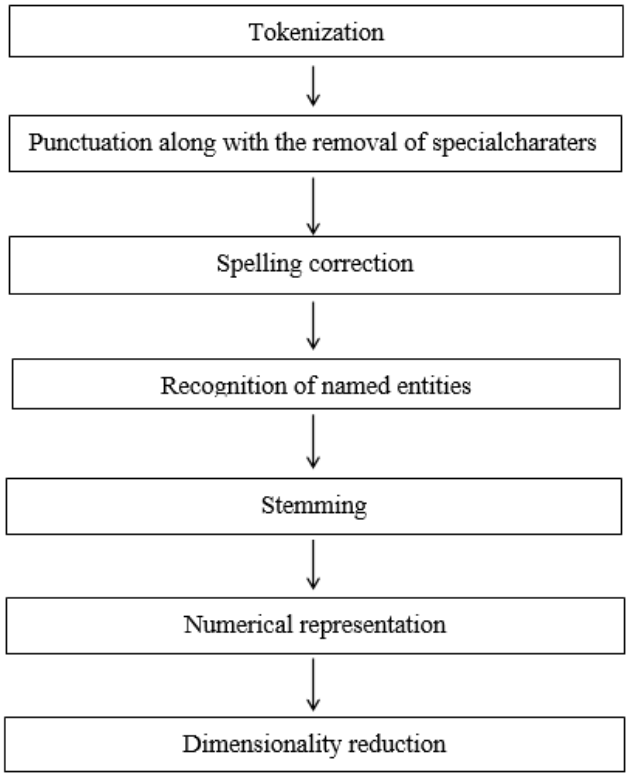


Figure 2: Steps of data representation

B. Proposed Method for Detecting Fake News Using Hybrid K-Means

1.Dimensionality Reduction with a Systematic Approach - Detecting false news involves coordinating computational processes. This method is based on a dataset that contains both fake and genuine news. The understanding of the presence of both types of news sets a standard for what is considered "genuine" and "fake" information. The Positive Naive Bayes (PNB) approach and the Positive Unlabeled (PU) method are used for acquiring data and storing the news in memory. The PU learning process involves organizing and learning from unlabeled data, such as images, using techniques like One-Class SVM if they are considered positive occurrences [31]. In this process, the details of the unlabeled material are not taken into account. The aim is to develop a classifier that can distinguish between positive and unlabeled data [32]. The computations involved in PU learning can be divided into three different classes, with 90% of the tests used for planning and 10% used for data.

2.The Network Change Strategy - When all of the tests' dimensions have been decreased, the D grid stores them all in the database and explains the huge number of columns j that break when the number of tests exceeds 2,000. On other words, the more highlights in a test there are, the more relevant this explanation becomes. D is projected across T, resulting in a system S that is actually more condensed, despite the fact that T's estimates are j*i and I is significantly smaller than j. The following formula must be used to organize the changes in sharpen:

$D_{(k\times i)}\times T_{(i\times j)}=S_{(k\times j)}$(5) ……Equ.(5)

The centroids are important in defining the structure of the transformation of data. The decline of arrangement M (incorporate arrangement) has been observed in the past. The hybrid k-means computation is used as a heuristic method to divide the data into k different types of clusters. The initial centroids are optimized for grouping the data points and then multiple phases of large clusters are formed. This is followed by the removal of noise, leading to the creation of purer clusters.In the following step two, the agglomerative settling (AGNES) calculation is utilized again to merge with relatively small clusters and refer to a comparative subject. In layer 3, we preserve a strategic distance

from clamour to decrease their help influence on the K-means, which corrects the incorrect combining that is occurring in AGNES [33]. To conceal this uncertainty, we used the Elbow to analyze the congruence of the findings and the numerous wholes of clusters followed by data as the outcomes. To do this, we have planned to deceive it.The data compactness can be assessed by applying the Elbow [34] method to the clusters and creating a valid relationship for the number of bunches that effect the overall diversity of the data contained inside the bunch. Graphically, the driving regard of k can be discovered by selecting the place where the desired twist diminishes unmistakably, then staying around indefinitely after that.

3.Radial Perimeter Procedure:The radial perimeter procedure is a method for widespread control in the D network, which has 2,000 features. Condition 2 was created as a result of reducing the PLSA dimensions in the M framework. Formula 2 tests the position of the x_ij^2 variables in the D network, and the output of this process is an irregular variable called R_i^2, which is the square of the length of a hypersphere that contains the actual news.

This approach involves expanding the analysis to include the entire network while considering the information related to both fake and genuine news. By doing so, a set of unique features specific to each type of news is obtained, providing a comprehensive view of the information being analyzed.

$$R_i^2 = x_{i0}^2 + x_{i1}^2 + \cdots x_{ik}^2 \text{.....Equ.(6)}$$

4. APPROACH FOR THE FNDHKM METHOD

genuineThis is a description of the three steps involved in the FNDHKM process. In the first step, the material is cleaned and organized. In the second step, the material is transformed into a numerical representation using vector space illustration. In the third step, dimensionality reduction is performed using PLSA, and the identification of false news is carried out using the PNB classifier and the Hybrid k-Means algorithm..

$$m_j^{(r)} = \frac{\sum_{x_i \in C_j^{(r)}} x_i + m_j^{(r-1)}}{\left|C_j^{(r)}\right| + 1} \text{.....Equ(7)}$$

The process of the FNDHKM method is depicted in Figure 3 and comprises of several stages. The first step involves tokenizing the content corpus T to extract individual words and phrases. In the next step, tweets from the content corpus are collected and organized. Step 3 transforms the collected data into a numerical representation by using a vector space model. Steps 4 and 5 further refine the numerical representation of the data using the PLSA technique. The purpose of these steps is to reduce the dimensionality of the data, resulting in a more manageable and simplified numerical representation. In step 6, the reduced dimension numerical representation is returned, which is then used in step 7. The final step involves using the hybrid-k Means clustering approach to accurately detect and cluster false news profiles.

Data Cleaning and Shaping

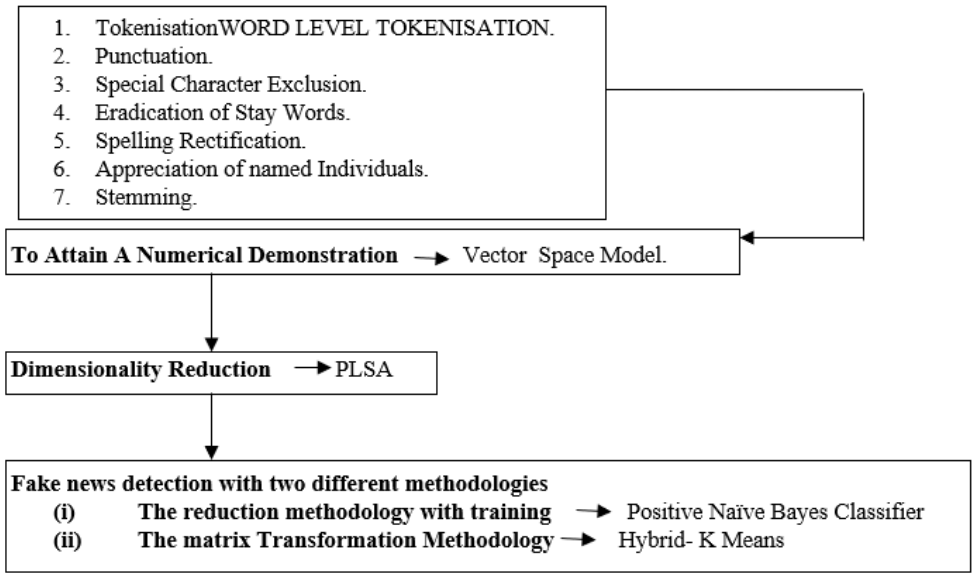


Figure 3: Workflow of the proposed FNDHKM approach

The method for identifying false news is as follows:

Algorithm 1: FNDHKM

1: Begin( ) {

2. Load the text corpus with tweets

3. Do the following for each word in the text corpus:

4. Apply word-level tokenisation, compute tokens, and store them in the word vector:

5: **Calculate**V⇒ $\frac{\partial_j(\theta)}{\partial v_c} = -\sigma(-v_o^t - v_c)u_o + \sum_{k=1}^{k} \sigma(v_k^t v_c)v_k$

Improve loss f^n by:

$$\frac{\partial_j(\theta)}{\partial v_o} = -\sigma(-v_o^t - v_c)v_c$$

$\frac{\partial_j(\theta)}{\partial v_k} = \sum_{k=1}^{k} \sigma(v_k^t v_c)v_c$; Where,v_cThe vector is the numerical vector of words.

6: **Apply** dimension reduction v_c calculate log-likelihood of v_c .

Then,

7: **Call** EM Algorithm()

Returns reduced $v_c$using PLSA

8: **Apply**KMeans clustering {

8.1 Calculate initial centroid by:

$$u_e = \{1, \quad if\ \left(C - f_p^{t-1}\right) < 0\ -\frac{\left|f_p^t - f_p^{t-1}\right|}{T}, \quad otherwise$$

Where, $f_p = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} \left|x_i^j - c_p^k\right|^2}{w_f}$

$$f_p^k = \frac{\sum_{i=1}^{n} \left|x_i^j - c_p^k\right|^2}{w_f}$$

}

Then,

9: v_cThe vectors are divided into two groups: the first group contains actual twitter text, while the second group contains phoney tweet text. } }

10: **End ()**

5. UPDATED HYBRID- K-MEANS ALGORITHM

The k-means algorithm is an efficient and widely used unsupervised learning method for partitioning data into clusters [37], [38]. It is often used for numerical data but can also be applied to non-numerical data by converting them into d-dimensional vector representations [39]. This allows the algorithm to work with numerical values and effectively partition the data into different clusters based on similarity or proximity. The algorithm works by iteratively finding the centroids of each cluster and reallocating data points to the nearest centroid until convergence. This process results in the formation of distinct clusters, allowing the data to be effectively organized and analyzed:

$\{x_i = [x_{i1} x_{i2} \quad \ldots. \; x_{id}]^T \epsilon \, R^d \mid i = 1 \ldots n\}$, ……..Equ(8)

The superscript "T" in the equation represents the transpose of a matrix. The transpose of a matrix is obtained by flipping the matrix over its diagonal, so that the row and column indices are interchanged. The "n" in the equation represents the number of records in the database.

The k-means algorithm works by minimizing the within-cluster sum of squared distances (WCSS) between the data points and their cluster centroid. The objective is to minimize the WCSS to find the optimal number of clusters, k, that represent the data. This is achieved by iteratively updating the centroids based on the average of the data points in the cluster until convergence is reached, or a stopping criterion is met. In other words, the k-means method aims to find the cluster centroids that produce the best grouping of data points in terms of minimizing the WCSS.

$c^* = V(c) \; = \sum_{j=1}^{k} \quad \sum^{n} \quad \|x_i - c_j\|$……Equ(9)

where: k – the number of clusters C_j , $j = 1 \ldots . k$,

$$c_j = \left[c_{j1} c_{j2} \quad \ldots . . c_{jd}\right]^T \; \epsilon \, R^d \quad - \; the \; centroid \; of \; the \; cluster C_j ,$$

$$c^* = \left[(c_1^*)^T (c_2^*)^T \quad \ldots. \; \left(c_p^*\right)^T\right]^T \in \; R^{dk}$$

The k-means method computes its results by using a specific mathematical formula. Using the k mean procedure, which has a mathematical formula for each record x_i and a centroid c_j, one may get the distanced $(x_i c_j)$ value. Now, if we take into consideration the Euclidian distance, we have:$d_{x_i c_j}$.

$d_{x_i c_j} = \; \|x_i - c_j\| = \sqrt{\sum_{l=1}^{d} \quad \left(\left(x_i l - c_{jl}\right)^2\right)}$,……..……Equ(10)

The formula for the Euclidian distance can be added to equation (9) to optimize the optimization problem. The result will be a more manageable and solvable issue, as the formula for the distance calculation will be included in the equation:

$$c^* = V(c), V(c) = \sum_{j=1}^{k} \quad \sum^{n} \quad \sqrt{\sum_{l=1}^{d} \quad (x_{il} - c_{jl})^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(11)}$$

The shape of the final cluster can be determined by using the formula given in (11), which is used for calculating between the centroid cj and the record xi

There are 5 stages in the k-means algorithm, out of which the first four are given below. The first step is step zero and is done once, and the remaining are performed cyclically till the cluster is not obtained in the final form.

Step 0.In this step, we select four things: the quantity of clusters p, the preliminary centroids, the determinedinteger of iterations and the acknowledged error. The μand denotes it is used for further calculations.

$\mu = 1 \ldots . . \mu_{max}$, where μ_{max}It is the thoroughgoingquantity of iterations which needs to be adjusted. I am starting from μ=1.

Various approaches are often used when it comes to the beginning of the centroid vector c. Earlier in this piece, we spoke about the initial centroid vector, denoted by the number c (1), and how it contains the potential location of the final centroid. The vector representing the beginning centroid may be written as:

$$c^{(1)} = \left[c_1^{T(1)} c_2^{T(1)} \quad \ldots. \quad c_p^{T(1)}\right]^T \epsilon R^{dp} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(12)}$$

In addition to this, the first centroids will be derived from this:

$$\left\{c_1^{(1)}, \ c_2^{(1)} \ , \ldots, c_p^{(1)}\right\} \subset R^d \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(13)}$$

The next step is to set the starting value for the acceptable error $\varepsilon > 0$, which should be greater than zero and will serve as the stop condition while the algorithm is executed.

Step 1: We may compute the distance ($d_{x_i c_j}$) between each vector xi and a centroid cj of the cluster, Cj where the for I = 1....n and. j=1...k respectively. We do this by using the formula (10) as our guide.

Step 2: The results collected in Step 1 will be used to create a new set of clusters in the next step. The vectors xm that are located close to the centre, also known as the centroid $C_j^{(\mu)}$, will be included in each new cluster $C_j^{(\mu)}$. We can begin the cluster production process by using the distance determined in step 1, which is the first step. Each point xm has several k distances $d_{x_m c_i^{(\mu)}}, i = 1 \ldots k$ that was computed, the vector xm will be included in a cluster $C_j^{(\mu)}$ if the distance $d_{x_m c_i^{(\mu)}}$ is minimum, which means:

$$c_j^{(\mu)} = \left\{x_m | d_{x_m c_j^{(\mu)}} \leq d_{x_m c_j^{(\mu)}}, \forall i = 1 \ldots k\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(14)}$$

where each vector xm is exclusively associated with a single cluster $C_j^{(\mu)}$,

Step 3. The new centroid $C_j^{(\mu+1)}$For each cluster, the fundamentals must be established $C_j^{(\mu)}$ based on the vectors xi that has been assigned to the cluster $C_j^{(\mu)}$. The new centroid is determined based on the bsee of the approaching formuxta, which contributes to the determination of the average of the coordinates for all vectors $x_j \epsilon C_j^{(\mu)}$.

$$C_j^{(\mu)} = \frac{1}{\left|C_j^{(\mu)}\right|}\sum_{x_i \epsilon C_j^{(\mu)}} \quad x_i, j = 1 \dots p, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(15)}$$

where $\left|C_j^{(\mu)}\right|$is the number of vectors that have been distributed across the cluster $C_j^{(\mu)}$. We may get the approaching centroid vector by utilising equation (15), which you can find below:

$$c^{(\mu+1)} = \left[c_1^{T(\mu+1)} c_2^{T(\mu+1)} \quad \ldots. \quad c_p^{T(\mu+1)}\right]^T \epsilon R^{dp} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\text{Equ (16)}$$

The scalar representation of the number (8) is:

$$C_{jl}^{(\mu+1)} = \frac{1}{\left|C_j^{(\mu)}\right|}\sum_{x_i \epsilon C_j^{(\mu)}} \quad x_{il},\ l = 1 \dots d, j = 1 \dots p. \qquad \ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(17)}$$

Step 4.It is determined whether or not the stop condition of the algorithm is met:

$$\left|c^{(\mu+1)} - c^{(\mu)}\right| < \varepsilon \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(18)}$$

This means that the k-means algorithm stops when the centroid vectors no longer change from one iteration to the next and when the optimization condition described in equation (16) is met. The final centroid vector c obtained is considered the optimal solution for the optimization problem:

$$c^* = c^{(\mu+1)}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(19)}$$

This is equivalent to the final centroids:

$$\left\{c_1^*,\ c_2^*, \dots, c_p^*\right\} = \left\{c_1^{(\mu+1)},\ c_2^{(\mu+1)}, \dots, c_p^{(\mu+1)}\right\} \subset R^d \ldots\ldots\ldots\ldots\ldots\text{Equ(20)}$$

If the conditions in equation (18) are not met, then the iteration will be increased by replacing with 1, and the process will continue from step 1. This iteration will be repeated until the conditions are met and the optimal solution is found.

$$eq^8 \rightarrow mean$$

To update centroid compute$c_i(x_i y_i)$.

$$C_{ix_i y_i} = \frac{Point\ which\ is\ centroid + mean}{|[[mean]^{itreation\ number}]| + 1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{Equ(21)}$$

The centred vector has now become,

$$[C_i(x_i y_i), \quad C_{i+1}(x_{i+1} y_{i+1}), \dots\dots\dots C_i + n(x_{i+n} y_{i+n})].$$

Until stopping condition is met.

The aim of the k-means method is to make each cluster more compact and clearly distinguishable from other clusters. In the conventional k-means approach, the centroid of a cluster is updated using only the mean value. However, in our case, the centroid of the cluster was updated using equation 21, which improved the compactness and separability of the cluster from other clusters.

6. EXPERIMENTAL ANALYSIS

A. The experiment's environment:

This system was selected to perform the exploratory evaluation due to its robust configuration, which includes the high-performing Intel Core i7 4770 processor, a generous amount of 16 GB of RAM, and fast storage capabilities

provided by the 512 GB SSD. These components ensure efficient and speedy processing of the data, which is crucial for accurate and reliable results.

B. Details of the Dataset

1. Financial tweets: The financial tweets dataset consists of over 28,000 tweets related to publicly traded companies. Each tweet is annotated with the company name being discussed and the profile picture of the user. The dataset features eight attributes and contains tweets from verified users of firms listed on the NYSE, NASDAQ, and NSP exchanges. The user interface for this dataset was inspired by the website https://www.kaggle.com/davidwallach/financial-tweets.

2. Political news - This dataset includes items relating to political news marked with phrases indicating that they check it as politically untrustworthy.This dataset has a total of 25117 tweets that make up the preparation set and5881 tweets that make up the testing set. The general audience on Kaggle may see Dataset. The collection included a variety of news tweets that were associated with Trump. https://www.kaggle.com/c/fake-news/data was the URL of the interface that was utilised for the incorrect news.

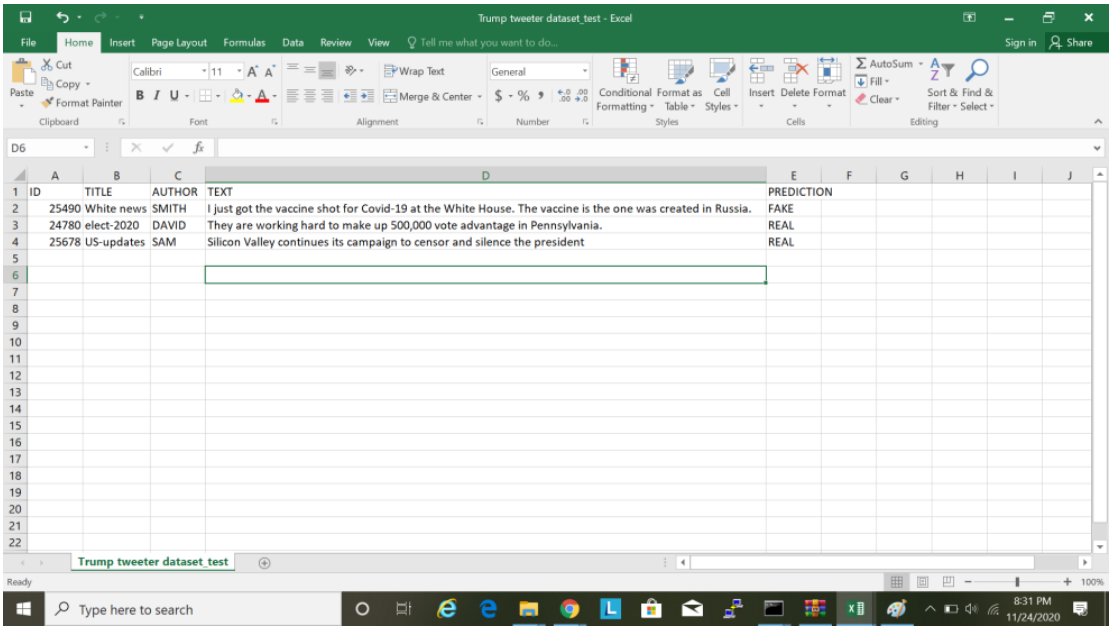
3. The COVID-19 tweets dataset [35] contains information on tweets related to the COVID-19 pandemic, including their IDs and the relevance scores. The data was collected by monitoring real-time tweets sent to Twitter that were related to the coronavirus and used over 90 specific keywords and hashtags. These keywords and hashtags were widely used when discussing the pandemic. The dataset was updated on March 20, 2020, to ensure that the data was accurate. The dataset contains a total of 182,857,703 tweets, with keywords such as "hospital", "covid19", "coronavirus", "sector", "vaccination", "people", "home", "pandemic", and "quarantine". The data was collected using the SMMT toolkit, a social media mining toolkit, which is used to extract data from various social media sites.

C. Performance evaluation

In other words, accuracy measures the proportion of correctly classified instances over all instances, while recall measures the proportion of correctly classified instances among all the relevant instances. A perfect classifier has an accuracy and recall equal to 1[36].

Illustrative example -

A graph is given to illustrate how the FNDHKM algorithm works to identify false news and demonstrate its efficacy.The performance of the FNDHKM algorithm was evaluated using the Political News and COVID-19 Tweets datasets. The algorithm was trained using these datasets and the results were tested using the same datasets. The results of the testing were presented in the form of graphs, with Figure 3 showing the performance results and Figure 4 showcasing the predictions made by the algorithm. The results were saved in a .csv file for future reference and analysis.



ID	TITLE	AUTHOR	TEXT	PREDICTION
25490	White news	SMITH	I just got the vaccine shot for Covid-19 at the White House. The vaccine is the one was created in Russia.	FAKE
24780	elect-2020	DAVID	They are working hard to make up 500,000 vote advantage in Pennsylvania.	REAL
25678	US-updates	SAM	Silicon Valley continues its campaign to censor and silence the president	REAL

Figure 4: Illustrative examples ofthe Political News dataset

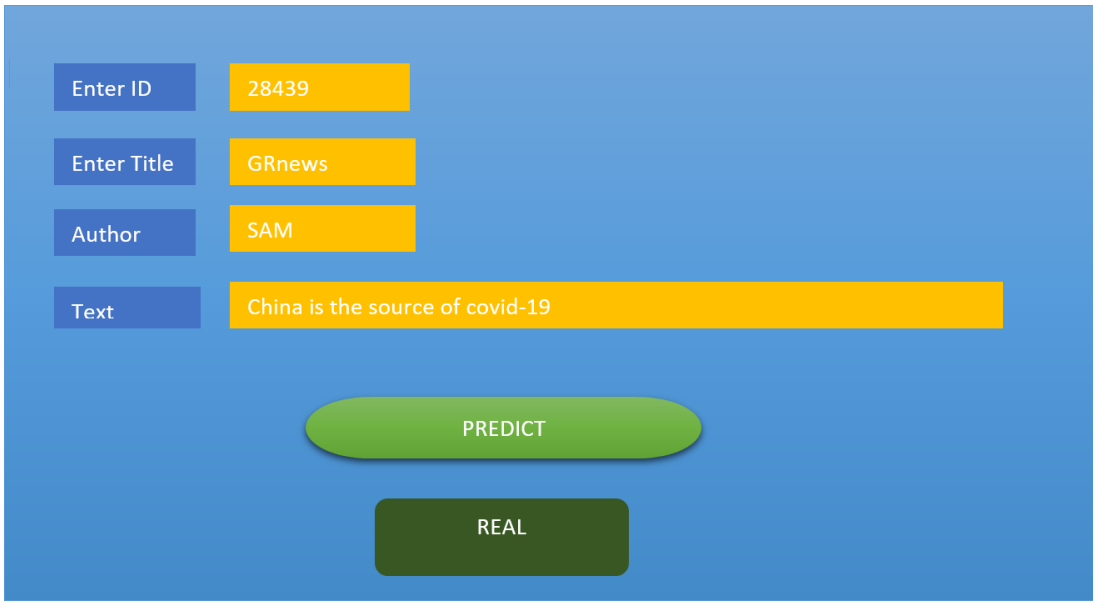
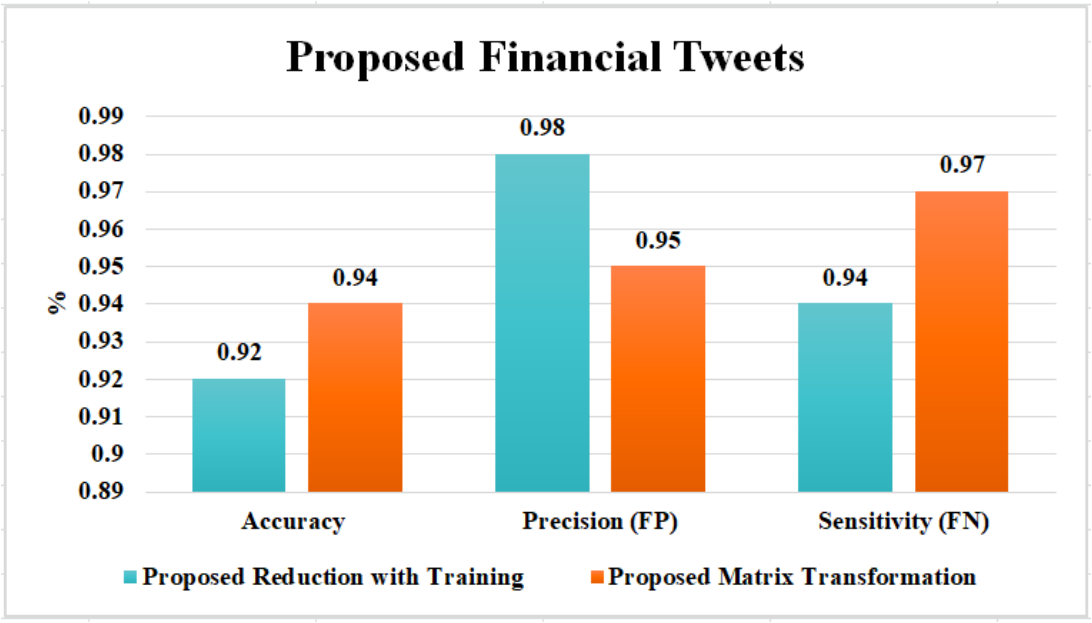


Figure 5: Illustrative examples on COVID-19 Tweets dataset

A. A dataset of tweets on the consequences of the budget Figure 6 is a representation of the data obtained from the audit and the exactness of the Budgetary Tweet dataset. Figure 6(a) provides information on the total number of clusters formed from the political dataset depicted forms. There are three clusters, with one cluster containing genuine tweets, another cluster containing false tweets, and the third grouping comprehending irregularities or exceptional instances. It represents how well the model performed on the dataset used in a scientific study. Based on



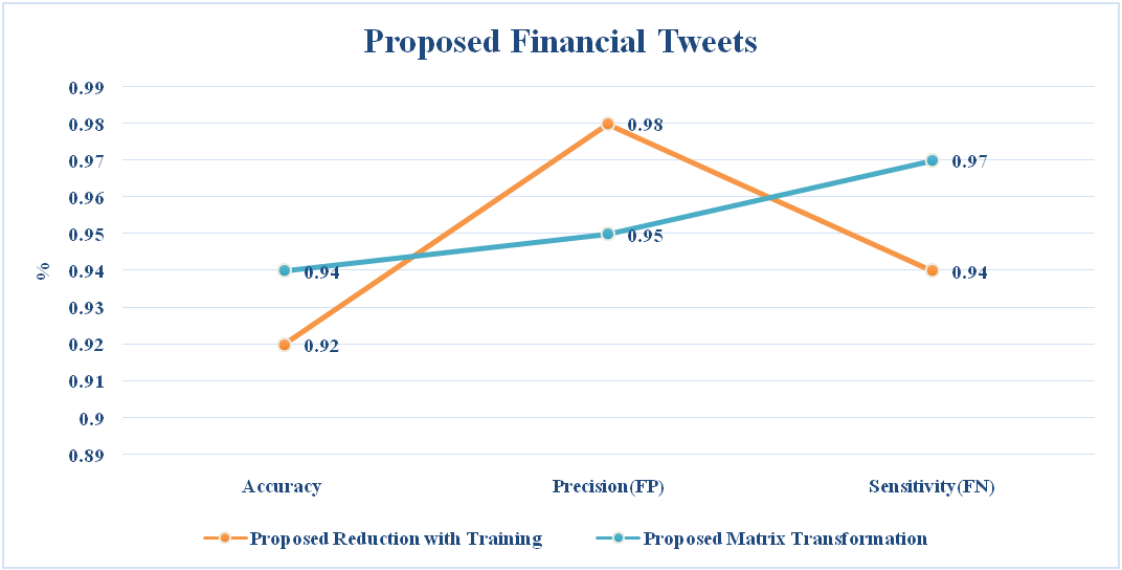


Figure 6: (a): The outcome of the ProposedFNDHKM algorithm applied to financial tweets regarding accuracy and recall.

this method, reduction with training has an accuracy of more than 92%, while matrix transformation has an accuracy of 94%, with a high recall rate. Financial Tweets were also analysed using k Mean [18] reduction with training, which obtained an accuracy of 86%, and k Mean [18] matrix transformation, which achieved an accuracy of 91%, combined with the ability to accurately remember (sensitivity) tweets, as shown in Figure 6(b).

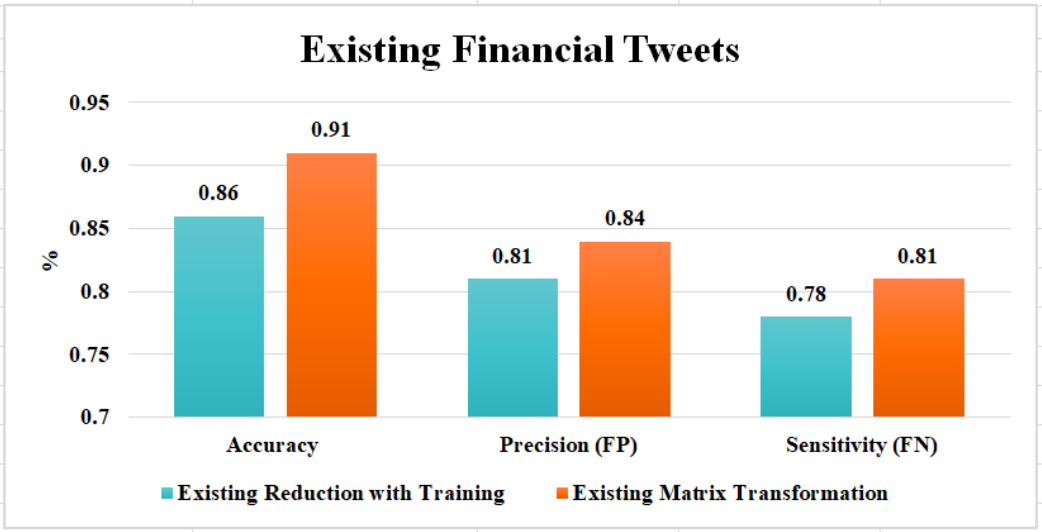


Figure 6(b): The result of applying k-means to previously published tweets on the economy in terms of accuracy and recall.

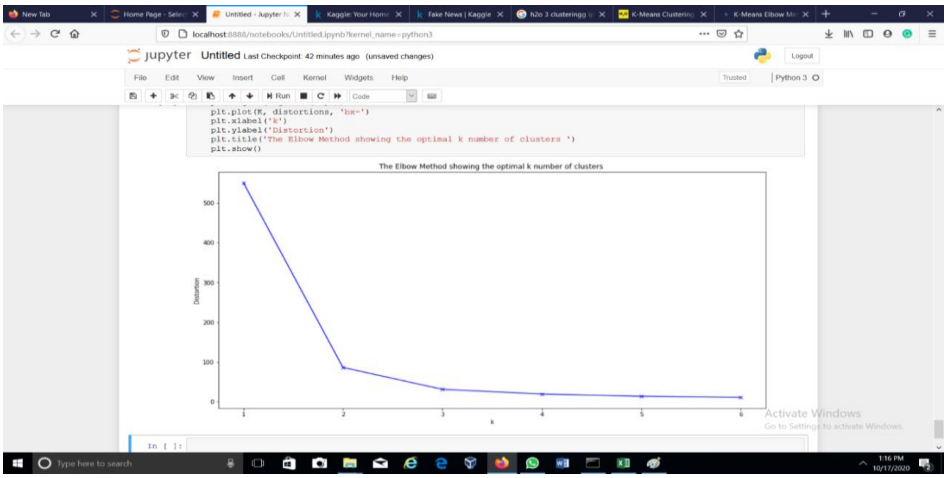
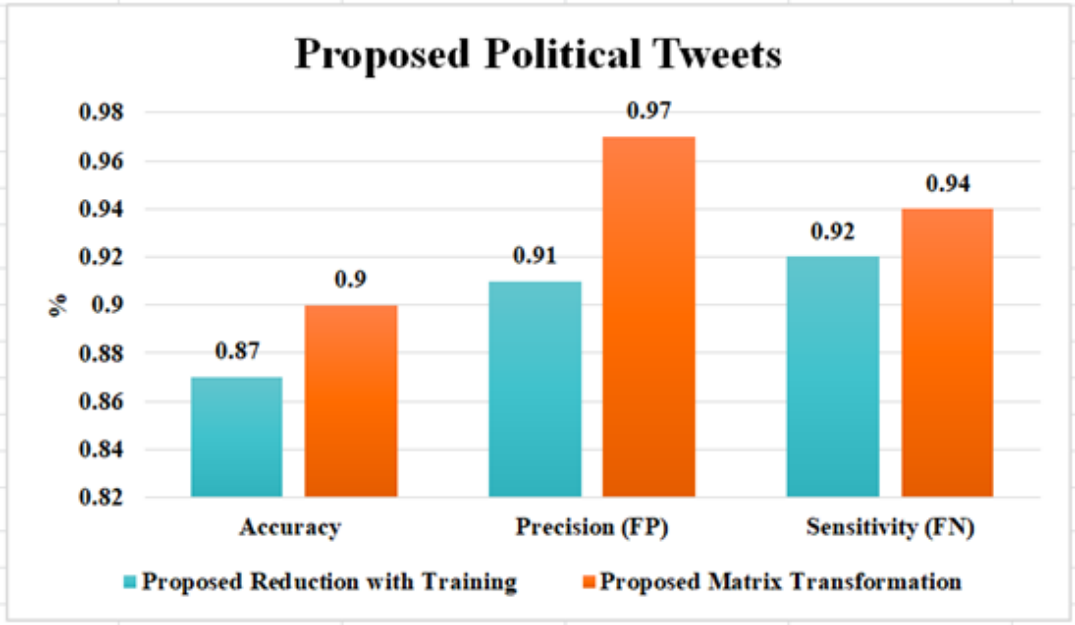


Figure 7: Elbow curve for Financial Tweets dataset in terms of the optimal number of cluster k.

Figure 7: Regarding the ideal number of clusters k, the elbow curve for the Financial Tweets dataset.

Dataset of political News

The recall data can be obtained from the accuracy and recall data presented in the graphic. Figuare 7 depict the Elbow curve for Financial Tweets dataset in terms of the optimal number of cluster, Figure 8(a) provides an illustration of the number of clusters created from the political news dataset. The dataset was separated into three clusters: one that contains authentic tweets, another that consists of false tweets, and a third cluster that encompasses either idiosyncratic or exceptional cases. Figure 9 displays the efficiency of the model when applied to the political news dataset. It shows that both the training and matrix transformation reductions had an accuracy rate of over 86 percent, along with a robust recall (sensitivity) rate. Furthermore, Figure 8(b) shows the results of the Political k Means [18] reduction with training, which achieved an accuracy rate of more than 78 percent, and the k Means [18] matrix transformation, which achieved an accuracy rate of over 80 percent, with a good recall (sensitivity) rate.



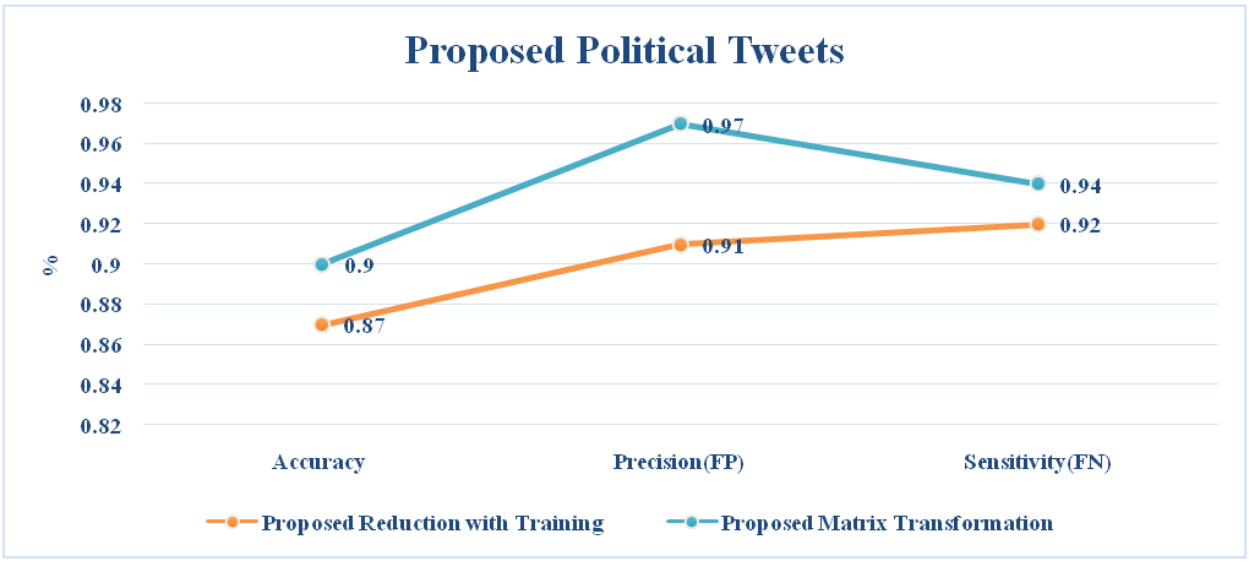


Figure 8: (a) : The outcome of the Proposed FNDHKM algorithm applied to the Political News dataset in terms of accuracy and recall.

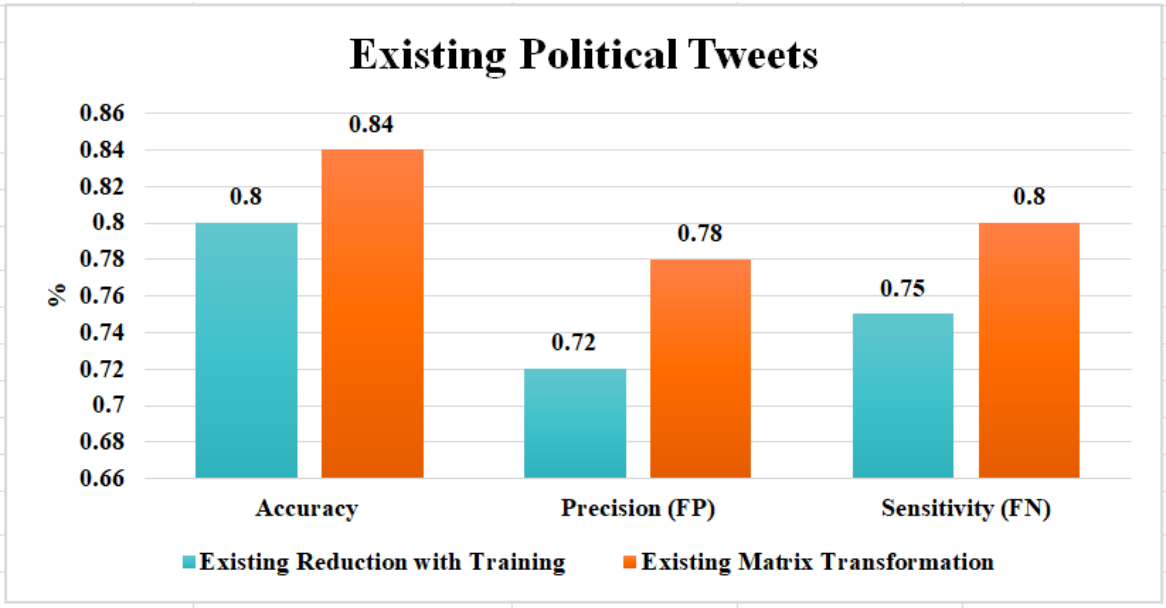


Figure 8(b): The result of applying an existing k-means algorithm to the Political News dataset concerning accuracy and recall.

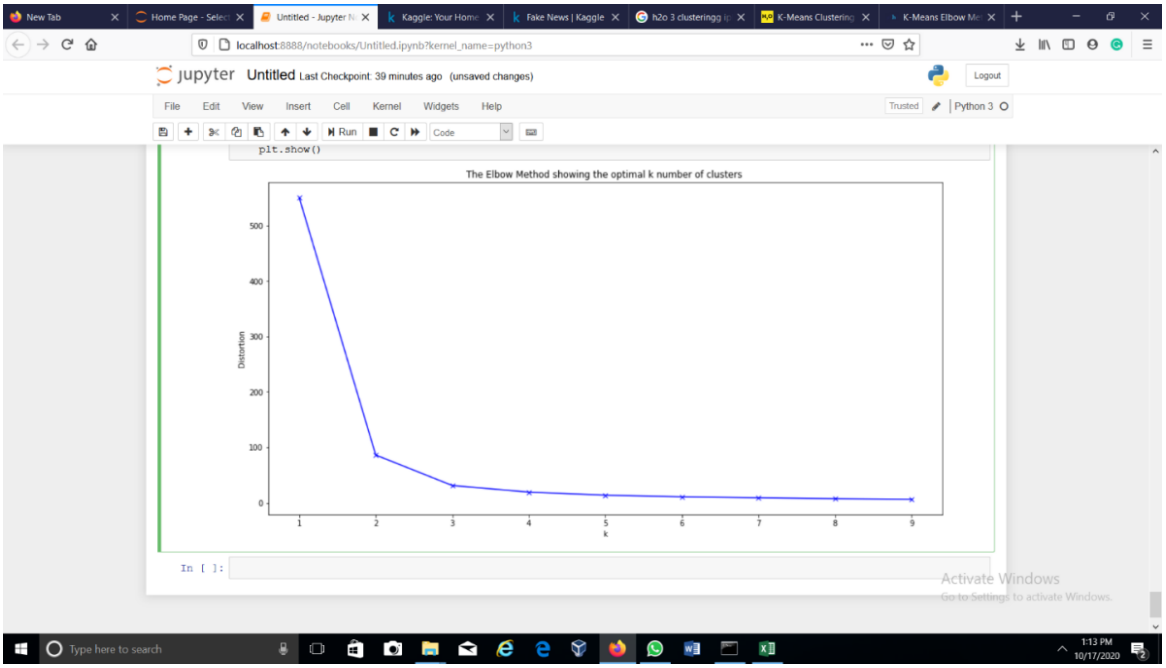


Figure 9: Elbow curve for a political dataset regarding the optimal number of cluster k.

Dataset of COVID-19 Tweets

Figure 9 demonstrates the effectiveness of the model on the political news dataset, Figure 10a displays the results of a precision survey conducted on the financial Twitter dataset. Figure 11 provides a closer examination of the clusters formed from the political dataset. These clusters are represented as three distinct shapes, with the first cluster containing genuine tweets, the second cluster containing fake tweets, and the third cluster containing inconsistent tweets or exceptions. It reveals that the reduction achieved through training had an accuracy rate of over 94 percent and the lattice change achieved an accuracy rate of over 95 percent, along with a high recall (sensitivity)

rate. Furthermore, Figure 10(b) displays the results of the COVID-19 Tweets dataset using the k Means [18] reduction with training, which achieved an accuracy rate of over 90 percent, and the k Means [18] matrix transformation, which achieved an accuracy rate of over 93 percent, while still maintaining a good recall (sensitivity) rate.

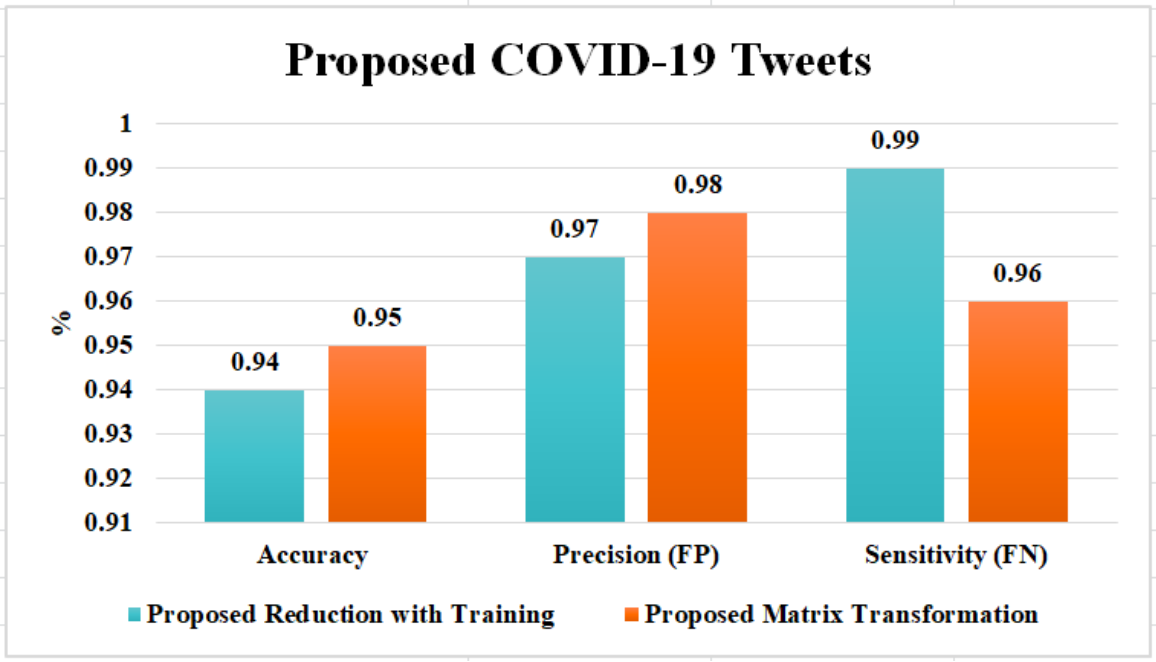


Figure 10: (a): In terms of accuracy and recall, the results of applying the ProposedFNDHKM algorithm on the Covid-19 Tweet dataset are as follows.

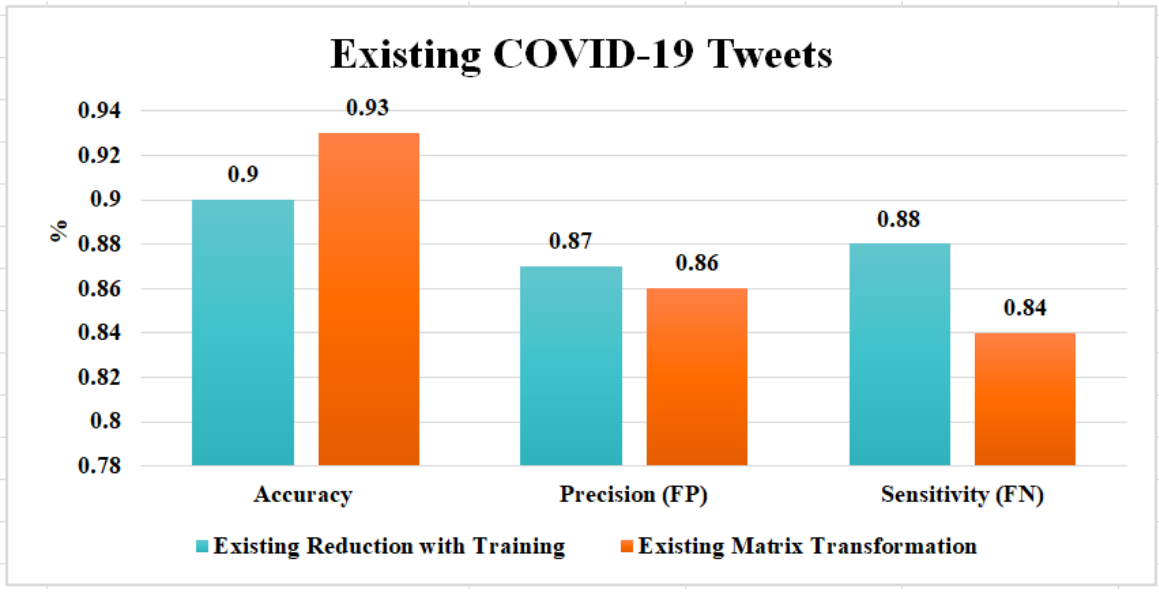


Figure 10: (b): Existing k-means dataset's results in terms of accuracy and recall for the 19 Tweets in the dataset.

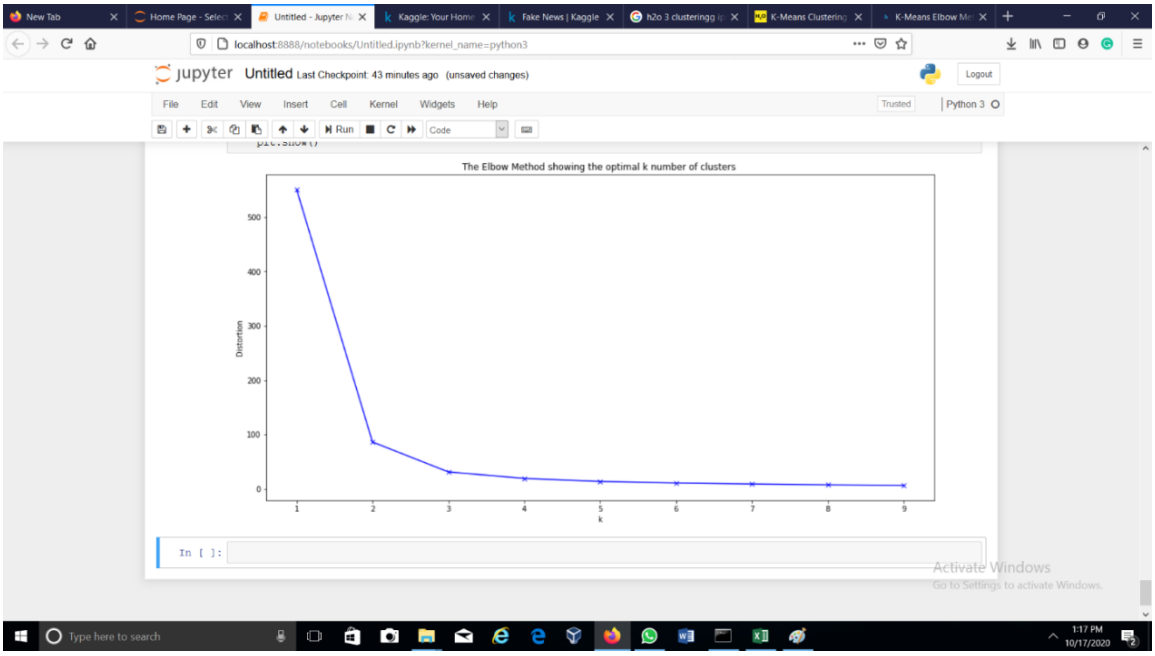


Figure 11: Regarding the ideal number of clusters k, the elbow curve for the COVID-19 tweets dataset.

7. DISCUSSION

The current model for detecting false news uses a combination of established knowledge, information from the profiles of individuals who created the news, and natural language processing techniques to extract meaningful insights. However, the proposed model takes a different approach by utilizing probabilistic latent semantic analysis (PLSA) for dimensionality reduction. This approach introduces a probabilistic weight to the model, which helps to optimize the randomness in the clustering component. Additionally, the proposed model uses HK-means and a vector space model for numerical representation, which enhances its overall performance compared to the current model that primarily focuses on textual analysis and probabilistic integration of texts. In short, the proposed model combines various techniques to provide a more comprehensive and effective method for detecting false news.

8. CONCLUSION

In this research, we introduced the "False News Disclosure using Hybrid-k Means (FNDHKM)" approach, which utilizes NLP and ML techniques to effectively identify fake news obtained from media sources. The FNDHKM approach employs three methods to detect false news, including classification and unsupervised clustering techniques. The approach evaluates the veracity of news and its potential for spreading by analyzing the region procedure. Our results showed that the FNDHKM approach performed well in terms of precision and recall on three different datasets. We also found that the Elbow approach can be applied to additional datasets. In the future, we plan to extend our work to larger datasets of fake news as the number of false news instances continues to increase at an exponential rate. Therefore, it is important to develop a flexible and scalable method to address this issue.

9. FUTURE SCOPE

For future research, expanding the experimentation to real-time datasets from various social media sites, such as Facebook and LinkedIn, can enhance the results presented in this research. It may also be valuable to consider

messages frequently shared within WhatsApp groups, which could provide insight into real-time traffic patterns. The development of a flexible and scalable computation that can handle large datasets is crucial in effectively addressing the issue of false news.

Conflicts of Interest:

The authors declare that they have no conflicts of interest.

Data Availability:

This dataset includes around 28 thousand or more tweets about virtually publicly listed organisations. These tweets are tagged with the name of the company about which the user is tweeting as well as their image in one or more zones. The dataset features eight highlights and includes tweets from confirmed customers on companies traded on the NYSE, NASDAQ, and NSP. The user interface of the dataset is modelled after the one found at https://www.kaggle.com/davidwallach/financial-tweets.

REFERENCES

1. Farkas, J., &Schou, J. (2018). Fake news as a floating signifier: Hegemony, antagonism and the politics of falsehood. Javnost-The Public, 25(3), 298-314
2. Sukhodolov, A. P. (2017). The phenomenon of "fake news" in the modern media space. EURASIAN COOPERATION: HUMANITIES ASPECTS, 36.
3. Fox, J. (2020). 'Fake news'–the perfect storm: historical perspectives. Historical Research, 93(259), 172-187.
4. Zhuk, D., Tretiakov, A., &Gordeichuk, A. (2018, May). Methods to Identify Fake News in Social Media Using Machine Learning. In Proceedings of the 22st Conference of Open Innovations Association FRUCT (pp. 401-404)
5. S. P. Binguac, "On the compatibility of adaptive controllers (Published Conference Proceedings style)," in Proc. 4th Annu. Allerton Conf. Circuits and Systems Theory, New York, 1994, pp. 8–16. doi:10.1011/sp.1994.312Smith, J., Leavitt, A., & Jackson, G. (2018). Designing new ways to give context to news stories. Facebook Newsroom. https://newsroom. fb. com/news/2018/04/inside-feed-article-context.
6. Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social media as information source: Recency of updates and credibility of information. Journal of computer-mediated communication, 19(2), 171-183.
7. Wang, W. Y. (2017). " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.
8. Castillo, C., Mendoza, M., &Poblete, B. (2011, March). Information credibility on twitter. In Proceedings of the 20th international conference on World wide web (pp. 675-684).
9. Wu, L., & Liu, H. (2018, February). Tracing fake-news footprints: Characterizing social media messages by how they propagate. In Proceedings of the eleventh ACM international conference on Web Search and Data Mining (pp. 637-645).
10. Ma, J., Gao, W., Wei, Z., Lu, Y., & Wong, K. F. (2015, October). Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (pp. 1751-1754).
11. Kim, J., Tabibian, B., Oh, A., Schölkopf, B., & Gomez-Rodriguez, M. (2018, February). Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 324-332).
12. Bond Jr, C. F., &DePaulo, B. M. (2006). Accuracy of deception judgments. Personality and social psychology Review, 10(3), 214-234.
13. Rubin, V. L., Conroy, N., Chen, Y., & Cornwell, S. (2016, June). Fake news or truth? using satirical cues to detect potentially misleading news. In Proceedings of the second workshop on computational approaches to deception detection (pp. 7-17).
14. Zhuk, D., Tretiakov, A., &Gordeichuk, A. (2018, May). Methods to Identify Fake News in Social Media Using Machine Learning. In Proceedings of the 22st Conference of Open Innovations Association FRUCT (pp. 401-404).
15. Aldwairi, M., &Alwahedi, A. (2018). Detecting fake news in social media networks. Procedia Computer Science, 141, 215-222.
16. Helmstetter, S., &Paulheim, H. (2018, August). Weakly supervised learning for fake news detection on Twitter. In 2018IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 274-277). IEEE.
17. Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., & Liu, H. (2019, July). Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 5644-5651).
18. de Oliveira, N. R., Medeiros, D. S., &Mattos, D. M. (2020). A Sensitive Stylistic Approach to Identify Fake News on Social Networking. IEEE Signal Processing Letters, 27, 1250-1254.

19. Kennedy, C. (2019, August). Expectation maximisation on unsupervised web mined data using probability latent semantic analysis (PLSA) algorithm. In 2019 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD) (pp. 1-9). IEEE.
20. Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence (Vol. 3, No. 22, pp. 41-46).
21. He, J., Zhang, Y., Li, X., & Wang, Y. (2010, April). Naive bayes classifier for positive unlabeled learning with uncertainty. In Proceedings of the 2010SIAM International Conference on Data Mining (pp. 361-372). Society for Industrial and Applied Mathematics.
22. Geng, X., Zhang, Y., Jiao, Y., & Mei, Y. (2019). A novel hybrid clustering algorithm for topic detection on chinesemicroblogging. IEEE Transactions on Computational Social Systems, 6(2), 289-300.
23. Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019, January). Fake news: Fundamental theories, detection strategies and challenges. In Proceedings of the twelfth ACM international conference on web search and data mining (pp. 836-837).
24. Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., &McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations (pp. 55-60).
25. de Oliveira, N. R., Reis, H. L., Fernandes, N. C., Bastos, A. C., Medeiros, S. D., &Mattos, M. D. (2020, February). Natural Language Processing Characterization of Recurring Calls in Public Security Services. In 2020 International Conference on Computing, Networking and Communications (ICNC) (pp. 1009-1013). IEEE.
26. Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.
27. Raposo, F., Ribeiro, R., & de Matos, D. M. (2014). On the application of generic summarization algorithms to music. IEEE Signal Processing Letters, 22(1), 26-30.
28. Halko, N., Martinsson, P. G., &Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. SIAM review, 53(2), 217-288.
29. Ratnesh Kumar Dubey, Dilip Kumar Choubey, “Deconstructive human Face Recognition using deep Neural Network”, Multimedia Tools and Applications Springer (ISSN NO: 1380-7501/1573-7721) published 28 March 2023. https://doi.org/10.1007/s11042-023-15107-4. Volume 82, 31800-31850.
30. Kuta, M., & Kitowski, J. (2015). Comparison of latent semantic analysis and probabilistic latent semantic analysis for documents clustering. Computing and Informatics, 33(3), 652-666.
31. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural computation, 13(7), 1443-1471.
32. Ratnesh Kumar Dubey, Dilip Kumar Choubey, “Efficient Prediction of Blast Disease in Paddy Plant using optimized Support Vector Machine”, IETE Journal of Research, Taylor & Francis Ltd (ISSN NO: 0377-2063 / 0974-780X)published 10 April 2023. https://doi.org/10.1080/03772063.2023.2195842. Volume 69.
33. Geng, X., Zhang, Y., Jiao, Y., & Mei, Y. (2019). A novel hybrid clustering algorithm for topic detection on chinesemicroblogging. IEEE Transactions on Computational Social Systems, 6(2), 289-300.
34. Liu, F., & Deng, Y. (2020). Determine the number of unknown targets in Open World based on Elbow method. IEEE Transactions on Fuzzy Systems.
35. Ratnesh Kumar Dubey, Dilip Kumar Choubey, “An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification”, Multimedia Tools and Applications Springer (ISSN NO: 1380-7501/1573-7721) published 07 August 2023. https://doi.org/ 10.1007/s11042-023-16247-3. Volume 82, issue 20.
36. R. K. Dubey, Dilip Kumar Choubey, “Reliable Detection of Blast Disease in Rice Plant Using Optimized Artificial Neural Network”, Agronomy Journal Wiley (ISSN NO: 1435-0645) published 23 August 2023. https://doi.org/10.1002/agj2.21449.
37. J. Mac Queen, “Some methods for classification and analysis of multivariate observations,” in Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 1967, vol. 1, pp. 281–297.
38. S. Lloyd, “Least squares quantization in PCM,” IEEE Trans. Inf. Theory, vol. 28, pp. 129–137, Mar. 1982. doi:10.1109/TIT.1982.1056489.
39. I.-D. Borlea, R.-E. Precup, F. Dragan, A.-B. Borlea, "Centroid Update Approach to K-Means Clustering," Advances in Electrical and Computer Engineering, vol.17, no.4, pp.3-10, 2017, doi:10.4316/AECE.2017.04001.
40. Gautam, Prakash, and Pankaj Raj Dawadi. "Keystroke biometric system for touch screen text input on android devices optimization of equal error rate based on medians vector proximity." 2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA). IEEE, 2017.

41. Pandit, Yogesh Sharad, and Chandan Singh D. Rawat. "Biometric Personal Identification based on Iris Patterns." Journal of Telematics and Informatics 2.1 (2014): 7-14.
42. Williams, Gerald O. "The use of d'as a “decidability” index." 1996 30th Annual International Carnahan Conference on Security Technology. IEEE, 1996.
43. Piyush Kumar Shukla Manish Agrawal, Asif Ullah Khan, “Stock Price Prediction using Technical Indicators: A Predictive Model using Optimal Deep Learning”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-2, July 2019, 2297-2305, DOI: 10.35940/ijrteB3048.078219.
44. Ruby Bhatt, Priti Maheshwary, Piyush Shukla, Prashant Shukla, Manish Shrivastava, Soni Changlani,Implementation of Fruit Fly Optimization Algorithm (FFOA) to escalate the attacking efficiency of node capture attack in Wireless Sensor Networks (WSN),Computer Communications,Volume 149,2020,Pages 134-145,ISSN 0140-3664,https://doi.org/10.1016/j.comcom.2019.09.007.
45. Prashant Kumar Shukla, Piyush Kumar Shukla, Poonam Sharma, Paresh Rawat, Jashwant Samar, Rahul Moriwal, Manjit Kaur, “Efficient prediction of drug–drug interaction using deep learning models”, Volume 14, Issue 4, August 2020, p. 211 – 216, DOI: 10.1049/iet-syb.2019.0116 , Print ISSN 1751-8849, Online ISSN 1751-8857, The Institution of Engineering and Technology , 23/04/2020.
46. Pannu, H.S., Singh, D. & Malhi, A.K. Multi-objective particle swarm optimization-based adaptive neuro-fuzzy inference system for benzene monitoring. Neural Comput & Applic 31, 2195–2205 (2019). https://doi.org/10.1007/s00521-017-3181-7.
47. Ratnesh Kumar Dubey, Dilip Kumar Choubey, “Adaptive feature selection with deep learning MBi-LSTM model-based paddy plant leaf Disease classification”, Multimedia Tools and Applications Springer (ISSN NO: 1380-7501/1573-7721. published 21 August 2023. https://doi.org/10.1007/s11042-023-16475-7.
48. Singh, D., Kumar, V. A Comprehensive Review of Computational Dehazing Techniques. Arch Computat Methods Eng 26, 1395–1413 (2019). https://doi.org/10.1007/s11831-018-9294-z.
49. Nguyen, G.N., Viet, N.H.L., Elhoseny, M., ...Gupta, B.B., El-Latif, A.A.A. Secure blockchain enabled Cyber–physical systems in healthcare using deep belief network with ResNet model Journal of Parallel and Distributed Computingthis link is disabled, 2021, 153, pp. 150–160.
50. Zhang, W.-Z., Elgendy, I.A., Hammad, M., ...Guizani, M., El-Latif, A.A.A. Secure and Optimized Load Balancing for Multitier IoT and Edge-Cloud Computing Systems IEEE Internet of Thing, 2021, 8(10), pp. 8119–8132, 9279239.
51. Abd El-Latif, A.A., Abd-El-Atty, B., Mehmood, I., ...Venegas-Andraca, S.E., Peng, J. Quantum-Inspired Blockchain-Based Cybersecurity: Securing Smart Edge Utilities in IoT-Based Smart Cities Information Processing and Management 2021, 58(4), 102549.