Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777814

IMPROVE ANALYSING ACCURACY OF IMPLICIT TOURIST FEEDBACK USING EFFECTIVE HYBRID CLASSIFIER MODEL

N. Saraswathi
Research Scholar, Dept. of CSE., Annamalai University, Annamalainagar – 608002.

Dr.T. Sasirooba
Assistant Professor, Dept. of CSE., Annamalai University, Annamalainagar – 608002.

Dr.S. Chakaravarthi
Associate Professor, Dept. of AI and DS., Panimalar Engineering College, Chennai - 600123.

Abstract:
One of the most significant communications for knowledge exchange is social media. The detrimental effect of false information on everyone using social networking sites is growing at the same time. For countries all around the world, the growth of the travel and tourism sector is essential to their economies. The sharing of tourist reviews on social media is exponentially growing at the same time. This sharing impacts psychological decisions regarding travel. Over the decades, a variety of social networking methods for analysis have been developed. However, they came into a number of issues as a result of short-text social media comments, which left them with a serious shortage of data. This research proposed a recommendation system for social networks to identify if the content is false or true using a hybrid SVM classifier using the LSTM in order to get around these challenges. The recommended method begins with gathering data using a tool like Kaggle and producing a dataset. Pre-processing of the acquired data includes segmentation, stop word elimination, lemmatization, and spelling correction. Following pre-processing, a count vectorizer is used to extract features from the data and transform them into binary. A hybrid SVM classifier using the LSTM model is used to further classify the collected features. Data that has already been processed and contains real information is compared to predicted data. It will be clear if the message is true or false if the predicted data matches the pre-processed data. To attain greater performance, the recommended approach is put into action. The overall projected result of the hybrid SVM classifier using the LSTM approach used for recommendations is superior to the current methods. The hybrid SVM classifier using the LSTM model was more accurate at identifying untrue from real communications.

Introduction:

The usage of social networking sites, where users may express their thoughts, interact with other users, and share their opinions with others, has become an ongoing fixture in people's everyday life. Opinions may spread quickly on social networking sites, where people share their emotions and opinions based on their own life experiences. Social media platforms offer crucial information in the proper formats and at a reasonable price. Inadequate by-products are produced as a result, and misleading information, including views, is conveyed to tourists. The economy of individuals, the economy of the whole country, and the growth of the travel and tourism sector will all be significantly impacted by this. Many machine learning and data mining approaches are now being utilized to quickly identify fraudulent information. The truth about the news is useful. Recent methods depend on n-grams and a bag of words to extract structural attributes. Following that, to distinguish between real and fraudulent communications, several machines or deep learning techniques like support vector machines, random forests, etc., were used. True or definite verbs, subjectivity, writing style, and consistency are examples of advanced qualities utilized for natural language processing. Researchers are currently looking at new tools to identify false information on the web.

It's still challenging to spot false comments on social networking sites. Long sentences must be processed in several existing content-oriented techniques. In contrast to longer written content, phrases and words may be learnt with greater efficiency. The short content is frequently observed on social networking sites, particularly Twitter, Facebook, and Google's viewer forums, which poses serious issues with data limitation. For each news circumstance, a false news prediction algorithm needs a variety of users' perspectives. This makes it simple to spot fraudulent information. However, a lot of people choose to share their viewpoints by reposting comments that have been posted previously. It may also be used to detect false news in advance.

On social networking sites, sources of misleading traveller comments might be detrimental. Recently, the situation known as the economic crisis has affected everyone around the globe. One of the most frequently discussed topics of conversation on the web's networks is surveys regarding tourist destinations. As a result, the number of international tourists has substantially decreased due to the dissemination of false information about the advantages and disadvantages of popular tourist locations. This knowledge has a negative impact on society and presents major difficulties. In addition to harming people's leisure-related thoughts and mindsets, fake news has a detrimental effect on people's mental health by inducing excessive worry or anxiety about travelling. In order to solve this issue, a hybrid SVM using an LSTM classifier is used in the recommendation system for social networks to determine if the information is true or false. This study focuses on identifying false information on social networking sites that are associated with travel websites and user reviews. The proposal is divided into three stages: pre-processing, feature selection, and classification to determine if providing information about travel destinations and deals is accurate or not. This model's primary goal is to increase its ability to accurately identify true and false news.

The primary contribution of the research is the development of a dynamic, lightweight recommender system for social network analysis utilizing the SVM classifier method and a hybrid LSTM block and preprocessing technique. Additionally, for efficient feature selection, Count Vectorizer turns text features into binary features. Additionally, a hybrid LSTM block-based SVM classification model is presented to increase prediction accuracy and generate more accurate predictions than other current models. As a result, the hybrid LSTM module used in the SVM classification model is intended to prevent the propagation of negative information about tourism.

The subsequent part of the study is organized as follows, and Section 2 discusses some research publications that use various machine learning and deep learning approaches to develop false news detection models. The recommended technique and procedure in the proposed model are thoroughly described in Section 3. The outcomes of applying the recommended false communication identification model are explained in Section 4. The full study project is eventually concluded in Section 5.

Literature Review:

Identifying fake news through online is still very difficult. Several existing content-based methods need a long text for processing. However, when compared to long text the sentences and words can be learned better [4]. Fake news is wrong information that spread on social media that is obviously false. Political, economic parties, or psychological profit can mislead the fake news and it will widespread among a huge population. Recently various machine learning and data mining techniques have been used to identify fake news easily. It can help to know the real news [1]. Recent techniques depend on extracting texture attributes using a bag of words and n-gram. Subsequently, several machines or deep learning approaches, such as support vector machine, random forest, etc were also used for classifying real news from fake news. Factive or assertive verbs, subjectivity, writing style, and consistency are the advanced features used for Natural Language Processing. Presently, Scholars learn improved features to detect fake news on social media [2, 3]. Many users share their comments in simple re-share the source scenario. It can also be used for effective prediction of fake news. Classifying the misinformation is expensive due to diffusion structure of re-tweets because of privacy concerns like social media. Many of them delete or hide the information [5]. Various researchers developed different prediction approaches [6, 7] to deal with these concerns. The source of misleading information has a harmful influence on social interactions [8]. In recent years, the worldwide has faced a situation is corona virus pandemic. COVID-19 has become an important discussed topic on social platforms. Hence, the pandemic of the COVID-19 issue produces various spread of misinformation like fake news [9]. The COVID-19 pandemic information leaks extremely on social media. This information causes significant difficulties and influences society [11]. Fake news not only endangers people's physical health but also creates negative impression on people's mental health, causing excessive anxiety or terror due to disinformation. In [12], the researcher had designed an Ontology and Context-Based Recommendation System (OCBRS) to consider the review's context to determine the opinion. In this model, the Neuro-Fuzzy Classification method was used to extract the context of the review. The reviews were automatically classified in this method by using the fuzzy rule. Ontology facilitates the systematic and hierarchical methodology to cluster the context and act as a repository of context. This method approaches an improvement in the accuracy of recommendation systems. The researcher of [13] presented an automatic prediction of COVID-19 regarding deliberations from social media. A natural language process approach depends on topic modeling to reveal multiple difficulties about COVID- 19 from public sentiments. In addition, the model revealed the sentimental classification of COVID-19 comments using LSTM recurrent neural network. In [14], the researcher had designed a recommendation approach utilizing the SVM machine learning technique for attribute extraction as well as combined ontology strategy. In this method, the verbal demand from the inactivated user was captured by a humanoid robot and translated into a full-text query. This technology analyzed the comprehensive queries to extract the demands of impaired users and convert them into search engine-friendly formats. The SVM was used to locate important data and eliminate useless information. Combined ontology-based sentiment analysis was used to determine the item's positive polarity for recommendation to the user. Java was used to create the intelligent model and combined ontology. In [15], geometric, deep learning was presented for an automatic fake news identification model. The fundamental algorithm simplified the classical CNN to graphical representation, which enables the fusion of varied data such as new propagation, social graph, user activity, profile, and content. The developed model used a Twitter dataset to train and test the information. In addition, fake news spread within a few hours of propagation the model predicted the fake news at an early stage. The model tested and trained the data separately at different times. The model utilized the content-based strategy to predict fake news. In [16], the researcher had presented an automatic fake news detection strategy for Facebook fake news prediction in a chrome environment. The developed model extrzseveral features from a Facebook account with a lot of news content attributes to investigate the characteristics of the account over deep learning. The implementation study used real-time data from Facebook. The researcher in [17] had developed a fake news identification model based on Graph-aware Co- Attention Networks (GCAN). The developed model detected fake news issues in real-time events on social media networks. The dataset execution used the short-text and relevant sequence of retweet users without text comments from Twitter social media. GCAN classifier predict whether the source Twitter comment was fake or not. According to the above-discussed literature, the existing strategies still contain several difficulties such as rapid propagation, access method and high cost. Various recommendation systems are designed based on Neuro-Fuzzy [16], CNN [17] and SVM [18] techniques. The performance of the existing model is less compared to the proposed model. Automatic fake news detection models on social media still have challenges because usually, social media comments are too short text, which causes major data sparsity issues [15]. Fake news detection model needs a diverse amount of user comments for each news scenario but, many users share their comments in simple and re-share the source scenario. This also causes difficulty for achieving effective prediction of fake news [16]. The fake news detection issues are a more realistic scenario on social media [17]. The proposed model designed a fake news detection model for Social Networking Analysis using a Hybrid LSTM-SVM Classifier to overcome these difficulties. Therefore, a hybrid LSTM-SVM model is proposed to recommend real or fake news effectively. Long Short Term Memory (LSTM) networks are a specialization of Recurrent Neural Network (RNN), and it can be used to detect long-term dependency in a single sentence. This kind of neural network was employed in the various fake news detecting methods. This shows a better classification performance than several existing [18]. LSTM is a kind of recurrent neural network (RNN) and it is developed to solve stability and speed difficulties [19]. The initial step is responsible for recognize the data that is no need for this process. The sigmoid functions are responsible for this process and consider the current Xt input at time and output as Vt-1 at time t-1. Based on the previous output the sigmoid function makes a decision for eliminate the unwanted portion in the output [20]. The SVM technique's goal is to discriminate categories each new data that enters within it, as well as to enhance the margin between the two labels. The SVM algorithm can find a hyperplane that splits the dataset into two groups [21]. Furthermore, combining multiple binary SVMs to form a multi-label classification model. Let be the training instances each case consists of an input as well as a label. Every hyper plane has a bias and a weight vector that may be calculated using the equation below (7) [22].

Proposed Methodology:

Using SVM classification along with LSTM as a hybrid classifier, a lightweight recommendation system is provided for analyzing travel perceptions posted by people on various social networking platforms in order to identify false information. Social networking websites have become one of the primary platforms for information sharing and collecting during the past ten years. Due to the convenience of its access ways for users, social media's continuous development continues to advance significantly depending on the number of users and their demands. Tourist users do this to take advantage of social media's easy access, but they also demonstrate that some noise and misinformation, especially fake news, is spread there. Misinformation about tourism services is still common on many tourism websites. Many natural language programming researchers have developed several methods to detect fake messages from users online. The architecture of the proposed model is shown in Figure 1, and it uses SVM classification along with LSTM as a hybrid classifier to develop an efficient false message detection model for accurate prediction.

The three different phases of the proposed false news detection architectures are Preprocessing phase, the Classification phase, and the Data Comparison phase for fake news prediction.

Segmentation, stop word removal, lemmatization, spelling correction, and the count vectorizer approach are all part of the pre-processing phase. The binary of text is feature-extracted using a count vectorizer. A hybrid SVM classification along with LSTM model is used in the classification phase to classify the extracted characteristics. The projected and preprocessed data are compared with the actual data during the data comparison phase.

It is either a true message or a fraudulent message if the anticipated data and the preprocessed data are equivalent.

Data collection:

A dataset is first created by collecting consumer opinions and information on tourism from many different websites. The dataset is entirely based on data on the tourist location, services, users' own experiences, and the preferences of individuals, and it consists of more than 10,000 pieces of information collected through the website, including both false and true news. Pre-processing is used for transforming this data from its original form into one that is computer-readable.

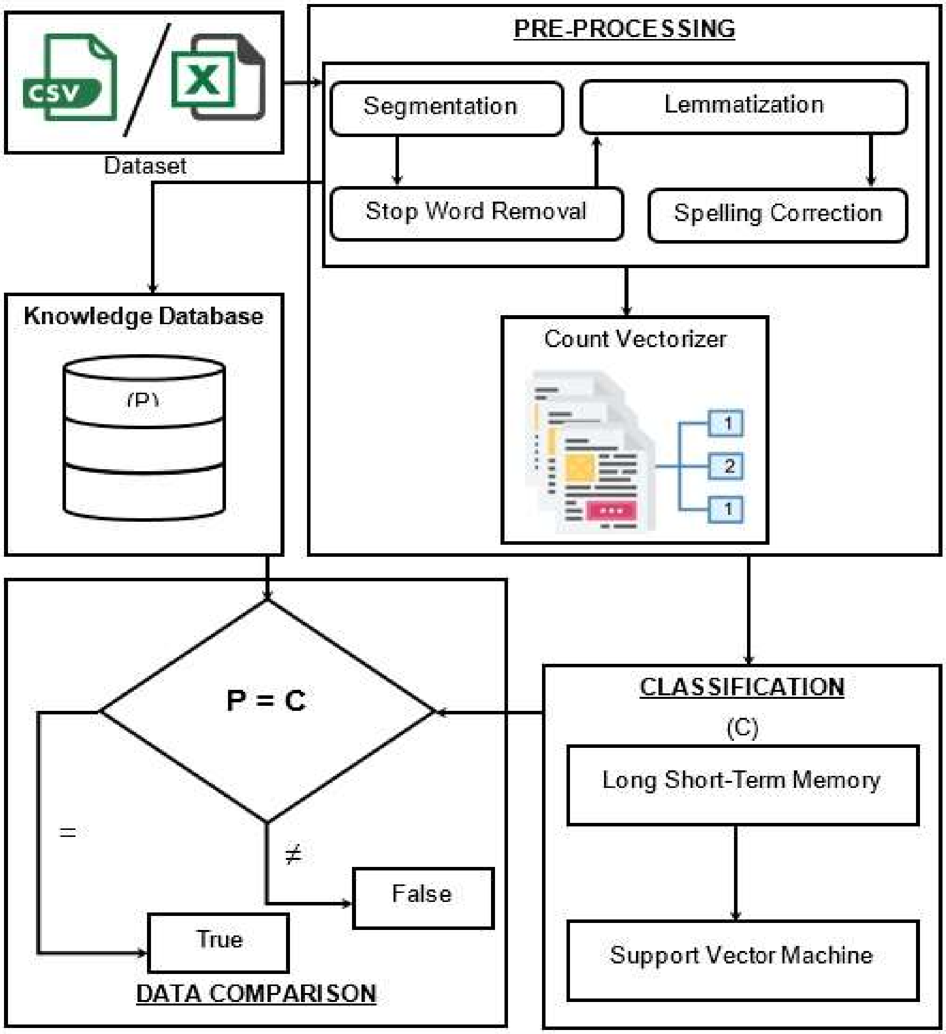

Figure 1. System Architecture

Pre-Processing:

The five preprocessing techniques in this suggested model are segmentation, stop word elimination, lemmatization, spelling correction, and count vectorizer. These methods pre-process the gathered data and get it ready for the subsequent stage of feature extraction. Through normalisation or the elimination of redundant information, this procedure eliminates noise from the collected dataset.

Segmentation:

Sentences represent the basic form of user data about tourism that has been collected. In order to digest them, sentences are divided up into their component words. The process of carefully splitting a text string into discrete, processable components is known as string parsing.

Stop Word Elimination:

Natural language processing (NLP), which is frequently utilised, uses the output from the splitting process to provide input back into the stop word removal procedure. Applications may now focus on what matters by removing stop words from texts like "is," "of," "the," "in," "a," and "on" that are frequently used in place of words.

Lemmatization:

Lemmatization is the systematic process of identifying the 'lemma' of a word based on its meaning. Lemmatization is the process of correcting errors utilising lexical and morphological analyses of words with the intention of identifying a word's lemma, which is its most fundamental or lexical form.

Spelling Correction:

Each word is examined by the spell checker using thousands of correctly spelt words. For automatic spelling correction, the majority of systems employ data from various sources of noisy and accurate word mappings as training data.

Count Vectorizer:

Natural language processing's foundational idea is the count vectorizer, a machine-learning model for feature extraction. Data frames are then processed once the data has been loaded into them to make sure that only the necessary fields are taken into consideration. The count vectorizer uses the training dataset to build a dictionary following the completion of the preprocessing. Features of the technique are obtained by extracting specific features. In order to create binary vectors based on word frequency, a count vectorizer preprocesses the word vectors from a bag of words. Based on the Sklearn package, the count vectorizer method operates. Representation vectors that include features are produced. It is necessary for the features to look at the estimation of words with higher frequencies prior to performing the count vectorizer. Finally, binary vectors may be created from text vectors to be usefully categorised.

Algorithm 1: Pseudo Code for Feature Extraction

Input data = X,
Preprocessed data = P,
 # For each text input in P, apply count vectorizer.
max(t) = Threshold #Initialize the threshold value max(t) for frequencies of word.
Sm = sparse matrix #the text input is transformed into a sparse matrix.
freq(t) = frequency Count #count the frequency of text freq(t) represented in sparse matrix.
mf(t) = data_comparison #compare the freq(t) with threshold value max(t) for entire dataset.
Select texts from matrix which met maximum threshold value.
Represent the selected text as most relevant features.

Output = Features are extracted and represented

Classification:

In this technique, false news linked to the travel and tourism sector can be prevented from spreading by using a hybrid SVM classification along with LSTM classification model to anticipate false information on websites. A hybrid SVM classification along with LSTM classifier is given the binary vectors from the count vectorizer. As a result, a thorough explanation of the proposed hybrid classification tactics is provided.

LSTM:

Recurrent neural network (RNN) long-short-term memory (LSTM) networks are specialised to identify long-term relying on a single sentence. Many false news detecting techniques have employed this kind of neural network. Compared to many other existing ones, it has a more powerful classification performance. A significant feature can be communicated by LSTM units up to a long way from the beginning of the input sequence, capturing potential long-range relationships. The expression of a document may be determined from its semantic description at various levels of refinement according to the state-of-the-art learning approach known as

LSTM. The phrase "embedding of words" refers to the representation of all words as a continuous, low-dimensional real-valued vector, while "embedding matrix" refers to the storage of each word vector.

LSTM, a kind of RNN, was created to solve stability and speed restrictions. A novel idea is provided by engaging with a block or a cell in essentially the same manner as how RNNs operate. The LSTM network is made up of memory units called cells, and this cell state is recognised as an essential state that permits and maintains forward data flow. Data removal and addition are made possible using sigmoid function gates. The connections between these gates, each with a unique set of weights, resemble a sequence of matrix operations. Implementing gates can lead to issues with long-term dependencies, which can be resolved by memorising these LSTs. The following list of stages describes the LSTM process.

Stage1: The initial stage is in the position of identifying information that is not necessary for this procedure. This process takes into account both the current input I_t in time t and the output O_{t-1} in time t-1, which is controlled by a sigmoid function. The sigmoid function chooses to exclude the redundant portion of the output based on the preceding output. Oblivion Gate (G_t) is the name of this procedure. Additionally, the cell's L_{t-1} almost all-inclusive value range falls inside the range of 0 and 1.

$H = wg_f [O_{t-1}, I_t]$ (1)

$G_t = \sigma (H + s_f)$ (2)

where wg_f stands for weight, s_f represents forgetting gate bias and is a sigmoid function.

Stage2: An ignore stage and a store stage from I_t in the cell state are both included in this stage. The needed information obtained from the concealed state O_{t-1} may be applied to the present input I_t because of the connection between I_t and O_{t-1}. The sigmoid layer and the tan_h layer are two further layers in this process. Using 0s and 1s, the sigmoid layer is in the position of deciding whether new data is updated or discarded. The weights are updated by providing values between -1 and 1 to the tan_h layer. Values are chosen according to their level of significance.

$p_t = \sigma (H + s_p)$ (3)

$IN_t = tan_h (H + s_n)$ (4)

Stage3: The cell status information is remembered in this instance. A memory gate that connects the hidden state O_{t-1} and I_t can update the cell state information L_{t-1}. The memory gate is created by fusing the current input I_t with the previous hidden level output O_{t-1}.

$L_t = L_{t-1}G_t + IN_tp_t$ (5)

Stage4: The output of the current input is merged with the important information from the previous hidden state O_{t-1} and the preceding cell state L_{t-1} at this point to create the undetectable state O_{t-1} next cell.

$V_t = \sigma (H_q + s_q)$ (6)

$V_t = V_t tan_h(L_t)$ (7)

where the output gate's weight and bias, respectively, are designated by the letters wg_q and s_q. The SVM classifier receives the output from the LSTM as input.

SVM:

One of the supervised machine learning techniques that can interact with an SV network is the support vector machine (SVM). SVM outperforms other classification techniques and is a very effective classification approach. Due to this, it is more challenging to identify associations between phrases, which are frequently the most crucial factor in evaluating the overall emotion polarity of texts. Multivariate characteristics are used by the SVM classifier at various levels of language description. The SVM approach aims to maximise the margin between two labels while also differentiating the kinds of each new piece of data that enters it. The hyper plane that splits the dataset into two groups may be located using the SVM method. To further explain, it can be said that while SVM discards "data points near the hyper plane" and the altered location of the separating hyper plane, support vectors are regarded as significant elements in the dataset. The highest likelihood at the data points that properly classify the data may be identified as a hyper plane as "the further from the hyper plane our data points lie "and" a line that linearly divides and classes a set of data". SVM provides more benefits since it performs better, makes predictions with more accuracy, and uses a smaller, more condensed dataset. Additionally, it may be used in a variety of ways to categorise or determine numbers. Support vector machines can also handle high-dimensional spaces and are more memory-efficient.

Utilising the Support Vector Machine (SVM) machine learning approach, binary classification is carried out. The key concept is to identify a superior hyper plane that appropriately divides data into two labels.

Also, a multi-label classification model is created by combining many binary SVMs.

Allowing for a_i and b_i as the training events, each example has input a_i and label $b_i \in (-1,1)$. Equation (8) below may be used to get the bias s and weight vector w_g of each hyperplane.

$w_g a + s = 0$ (8)

A function called the hypothesis function can be defined from the hyperplane equation describing the training and test data (9).

$g(a) = \text{sign}\,(w_g a + s)$ (9)

Equation (10), which deals with the kernel function, may be used to represent the previous function.

$g(x) = \text{sign}\,[(\gamma_1 b_1 k(a,a_1) + \gamma_2 b_2\, k(a,a_2) + \gamma_3 b_3\, k(a,a_3) + \ldots + \gamma_{IN} b_{IN}\, k(a,a_{IN})) + s]$ (10)

where function s is used The kernel function $K(a,a_i)$ is used to transfer the input vectors to the extended feature space where a_i is the training input, IN is the number of training occurrences, and b_i represents the related label. The constraint which is specified in Eqs (10) determines the coefficients of γ. Also, γ_i works with the constraint that the value can be zero or greater, where i =1 to IN.

$\gamma_1 b_1 + \gamma_2 b_2 + \gamma_3 b_3 + \ldots + \gamma_{IN} b_{IN} = 0$ (11)

Hybrid SVM classification along with LSTM:

LSTM and SVM are two approaches that are combined in the suggested model. The framework for the hybrid SVM classification along with LSTM is shown in Figure 2. The layer1, layer2, layer3, and final layer are the four layers that make up the proposed hybrid SVM classification along with LSTM classifier. The LSTM input layer receives the 80% feature extracted data from the count vectorizer and feeds it to the layer2. The buried layer of an LSTM is represented by a gated cell or gated unit. It has four layers, each of which interacts with the others to determine the cell's output and state. These two items are then sent to the subsequent concealed layer. While LSTM contains one $\tan_h$ layer and three logistic sigmoid gates, RNN only has one tanh layer in its neural network. These gates choose which data is retained for the subsequent cell and which data is eliminated. The discovery scale runs from 0 to 1, where "0" means "reject information" and "1" means "include information." The output from the layer2 is given to the layer3, which uses it to determine the proper absolute structures of the features picked up by the network's lower levels. The input is then reduced to a detailed description of characteristics because they are normally at the top of the network structure. Every node in the layer3 determines its own weight among the other nodes. The final layer is then fed the output from the layer3.

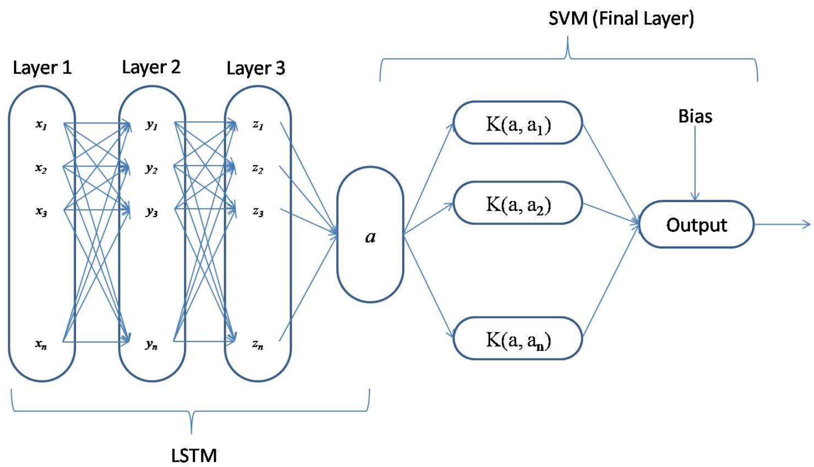

Fig. 2. Architecture of the Hybrid LSTM-SVM.

Typically, the LSTM network's final layer is its top layer. A final function frequently employed for multi-class classification, the final layer's formula is essentially identical to the sigmoid function used for logistic regression. Only when the classes are mutually exclusive is it possible to utilise the final function in a classifier. In order to achieve accurate prediction, the suggested technique substitutes the LSTM network's final layer with an SVM classifier. SVM carries out classification by looking for hyper planes that distinguish between classes. In this case, the categorization and training phases use both strategies.

The under consideration document is fed into the LSTM network as input, a statistical model of the LSTM neural network is built, and for each vector in the training set a feature vector is produced using the weights gained from the final network layer. The SVM classifier was used to finish the training phase. The embedding vector is created using a previously trained LSTM network, and the classification steps are applied in the same manner to each vector that has to be categorised. The trained model is tested by retesting 20% of the feature extracted data once the training phase is complete. In order to offer predicted classes, the SVM classifier combines the embedding vector with additional classification characteristics.

Data Comparison:

The projected data is compared to the pre-processed data, which has correct information on the tourist destinations and their services, after the classification procedure is complete. Accurate data regarding tourism and its services, local people' lifestyles, World Health Organisation (WHO) recommendations, and other issues are provided by the databases obtained from these statistics. If the recommended SVM classification along with LSTM classifier's output resembles the preprocessed data, it is regarded as a true message from the collected users; otherwise, it is treated as a false message. This enhanced the capacity of the proposed hybrid prediction model to detect fake news about tourist sites.

Algorithm 1: Pseudo Code for Hybrid classifier model based on LSTM-SVM

Input data = *Features are extracted and represented & P*

Cv = Features are extracted

Bv = Features as Binary vector #Represented the features as binary vector for classification.

C = classified_data #Apply hybrid LSTM-SVM classifier model to get C.

#Compared C with P.

if both numeric values equal (C = P)

{

real news

}

Else

{

False news

}

Output = Actual News represented

RESULT AND DISCUSSION:

To assess its effectiveness, MATLAB software is used to develop the proposed dynamic lightweight recommender system for social network data analysis utilising a hybrid SVM classification along with LSTM classification model. The retrieved information was compiled from user comments on tourism-related websites, and it contained both false and true news. Capital letters, special characters, and keywords are used in commands. Sentence extraction, segmentation, stop word elimination, lemmatization, and spelling correction are all common preprocessing techniques. The count vectorizer receives the results of the preprocessing, extracts the features, and turns the text into binary vectors. Two sets of the count vectorizer's output are separated for the classifier's training and testing. The proposed hybrid SVM classification along with LSTM classifier is initially fed 2000 users' review datasets using feature extraction to train the model. The remaining 8000 opinion data are then input into the trained model to test the classifier when the training is complete. The pre-processed data is then contrasted with the expected classifications. The reviews are recognised as real if the anticipated class matches the preprocessed data, else the collected user feedback data is handled as false reviews.

Table 1 explains how the input is retrieved during the preprocessing phase and how the segmentation step may segment each word after receiving the data. Following segmentation, stop word removal is applied to the data to get rid of terms like "is," "of," "the," "in," "a," and "on".

Table 1. Input and Output attained in pre-processing

INPUT	SPLITTING	STOP_WORD	LEMENTIZE	SPELL_CORRECTION
chennai marina beach is longest natural	“Chennai” “marina” ”is” “beach” “longest” “natural”	“Chennai” “marina” “beach” “longest” “natural”	“Chennai” “marina” “beach” “longest” “natural”	“Chennai marina beach longest natural beaches india beautiful morning evening seeing joggers early

beaches in india, beautiful in morning and evening seing. joggers early morning walk pavement	"beaches" "in" "india" "," "beautiful" "morning" "and" "evening" "seing" "." "joggers" "early" "morning" "walk" "pavement"	"beaches" "india" "," "beautiful" "morning" "evening" "seing" "." "joggers" "early" "morning" "walk" "pavement"	"beaches" "india" "beautiful" "morning" "evening" "seing" "joggers" "early" "morning" "walk" "pavement"	morning walk pavement"
Very good stay. Nice location. Suggest to offer breakfest included in that offer. Also check to offer for complimentery pick up	"Very" "good" "stay" "." "Nice" "location" "." "Suggest" "to" "offer" "breakfest" "included" "in" "that" "offer" "." "Also" "check" "to" "offer" "for" "complimentery" "pick" "up"	"good" "stay" "." "Nice" "location" "." "Suggest" "offer" "breakfest" "included" "offer" "." "check" "offer" "complimentery" "pick"	"good" "stay" "Nice" "location" "Suggest" "offer" "breakfest" "included" "offer" "check" "offer" "complimentery" "pick"	"good stay Nice location Suggest offer breakfast included offer check offer complimentary pick"
Its value for money. Its a pleasent experience. They have to improve the reception. Otherwise nothing bad about the hotql.	"Its" "value" "for" "money" "." "It's" "a" "pleasent" "experience" "." "They" "have" "to" "improve" "the" "reception" "." "Otherwise" "nothing" "bad" "about" "the" "hotql" "."	"value" "money" "." "pleasent" "experience" "." "improve" "reception" "." "Otherwise" "nothing" "bad" "about" "hotql" "."	"value" "money" "pleasent" "experience" "improve" "reception" "Otherwise" "nothing" "bad" "about" "hotql"	"value money pleasant experience improve reception Otherwise nothing bad about hotel"

Lemmatization is the process of removing inflectional suffixes and prefixes to reveal the word's lexical form. The data is then generated with findings that have already been pre-processed and errors fixed. These results are then used to select features in a count vectorizer. These features are then provided in the classification procedure after that. In comparison to other models, the proposed framework performed better.

Experimental Analysis:

The proposed web-based networking analysis model was put to the test utilising a number of performance parameters. Several parameters, including accuracy, error, sensitivity, specificity, false positive rate, f1_score, and kappa, were taken into account for this research. On the basis of user reviews on travel, the proposed algorithm (LSTM-SVM) and false communication detection model have been evaluated. For comparison analysis, a number of currently available strategies based on false information detection models are taken into consideration.

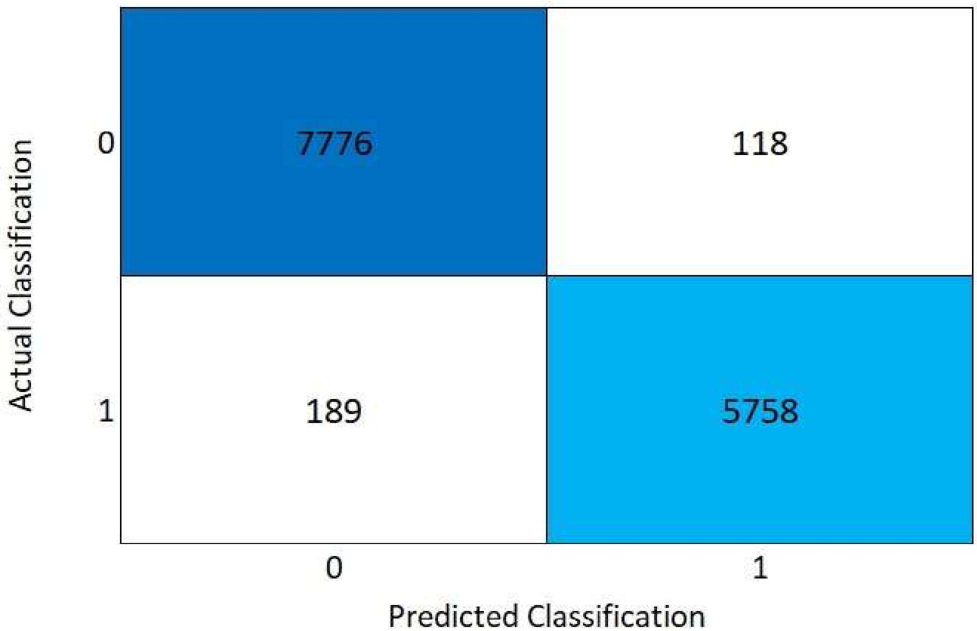

Fig. 3. Confusion matrix obtained for the proposed model.

The confusion matrix found in this suggested model is shown in Figure 3. Here, the projected class and the actual class are denoted by the labels X and Y. A 2-by-2-dimensional confusion matrix is employed since the proposed framework classifies two classes as false and real.

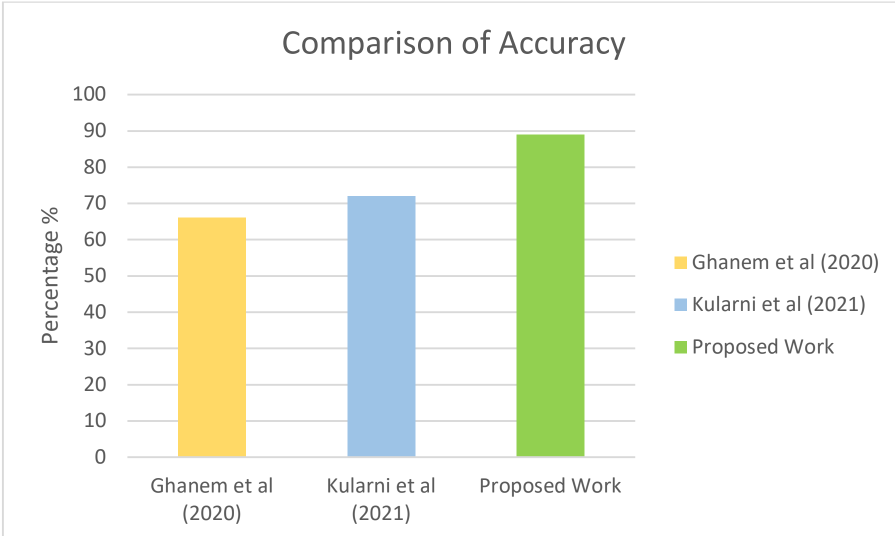

Fig. 4. Comparison of Accuracy among Proposed and Existing Approaches.

Figure 4 displays a comparison of the proposed and current methodologies for identifying false information on social media platforms based on accuracy (%). The value of accuracy and several machine learning and deep learning approaches are the foundations of graphical representation. percentages with regard to X and Y labels. The suggested method's accuracy is 90% compared to other approaches like LSTM's 71% accuracy, and SVM's 65% accuracy. It demonstrates that the suggested model performs better than others.

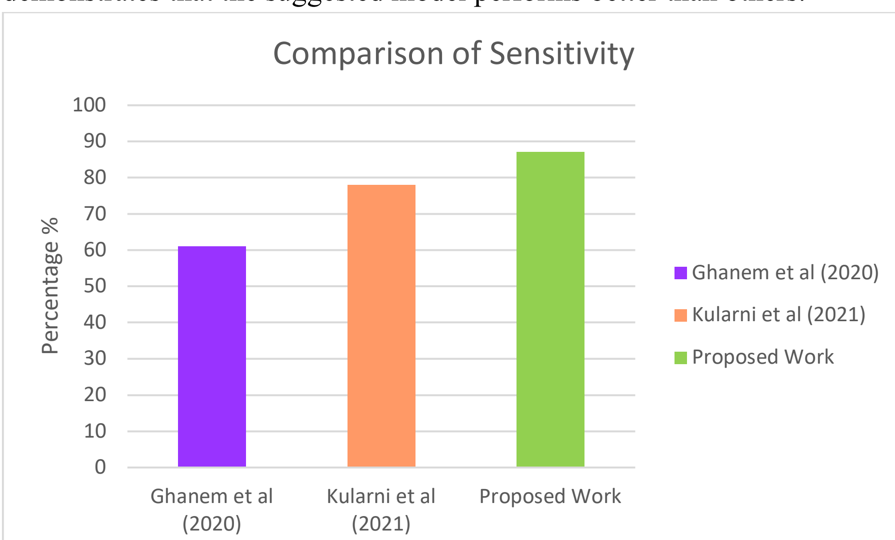

Fig. 5. Comparison of Sensitivity among proposed and existing approaches.

Figure 5 displays a comparison of the proposed and current methodologies for detecting false information in order on social media in terms of sensitivity (%). Multiple machine learning and deep learning approaches, as well as sensitivity values, are used to interpret images. Percentages with regard to X and Y labels. The proposed method's 88% sensitivity is high when compared to other approaches, such as LSTM's 78%, and SVM's 60%. This can demonstrate that the proposed approach performs better than others.

Fig. 6. Comparison of Error among Proposed and Existing Approaches.

In Figure 6, a comparison of the accuracy of proposed and current approaches for spotting false information on social media is presented. A number of machine learning and deep learning approaches, as well as the value of error in percentage for both, are used to analyse graphics both the X and Y labels. Comparing the suggested technique to current ones like LSTM 0.28, and SVM 0.35, the detection error is 0.1. This can indicate that the proposed approach performs better than others.

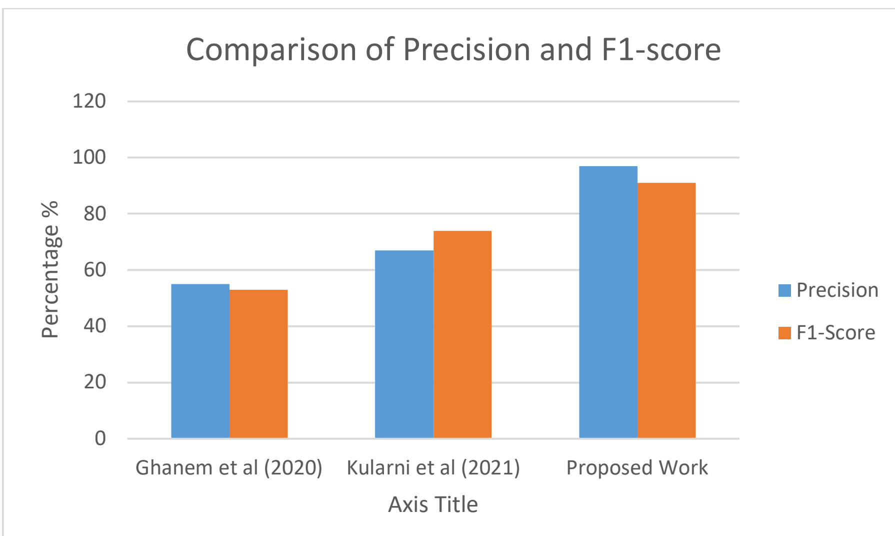

Fig. 7. Comparison of Precision among Proposed and Existing Approaches.

In Figure 7, the proposed and current solutions for spotting false information in social media are compared in terms of accuracy (%). The value of accuracy in % on the X and Y labels, respectively, and various machine learning and deep learning approaches are the foundations for graphical interpretation. In comparison to other approaches like LSTM (71%), and SVM (52%), the suggested method has an accuracy of 98%. It demonstrates that the suggested model performs better than others.

Figure 7 shows the comparison of the proposed and current methodologies for detecting false information in social networks using F1_score (%). The value of F1_score in % on the X and

Y labels, respectively, and various machine learning and deep learning algorithms are the basis of the graphical representation. The proposed technique's F1_score is 90%, which is high when compared to other approaches like LSTM's 71%, and SVM's 52%. This can indicate that the proposed approach performs better than others.

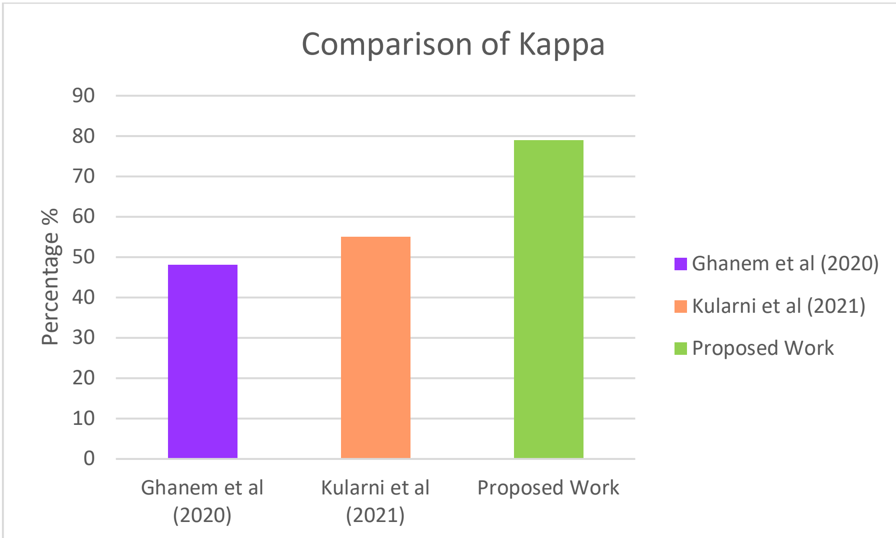

Fig. 8. Comparison of Kappa among Proposed and Existing Approaches.

Figure 8 shows the comparison of the proposed and current methodologies for spotting false information in social media based on kappa (%). The value of accuracy in % on the X and Y labels, respectively, and various machine learning and deep learning approaches are the foundations for graphical interpretation. The proposed method's accuracy is 80%, which is greater than that of other approaches like LSTM (55%), and SVM (48%). This demonstrates how the proposed approach performs better than others.

CONCLUSIONS

To enhance the accuracy of false information predictions, a hybrid SVM classification along with LSTM classifier-based false information detection model has been developed. The information is obtained from tourism-related websites. 20% of the data was first used to train the hybrid SVM classification along with LSTM classification model, which conducted segmentation, stop word elimination, lemmatization, and spelling features on the input. Following testing with the remaining 80% of the collected features, the trained model successfully produces classes. The actual information is compared between the pre-processed data and the estimated data. The information provided is true or false depending on whether the anticipated information matches the pre-processed data. The suggested model is used to calculate performance indicators like accuracy (90%) and sensitivity (88%) and Error (98%) and precision (98%) and F1_score (90%) and kappa (80%). In comparison to current methods, the overall predicted effect of the hybrid SVM classification along with LSTM classification approach used in recommendation methods like LSTM, and SVM is better. The proposed social network analysis methodology offers efficient detection of false information, which can be used to determine if reviews submitted by visitors about the tourism sector are true or false. It can also be used to determine the authenticity of messages for user awareness.

REFERENCES:

[1] Dong, X., Victor, U., and Qian, L., Two-path deep semisupervised learning for timely fake news detection, IEEE Trans. Comput. Soc. Syst., 2020, vol. 7, no. 6, pp. 1386–1398.

[2] Choudhary, M., Chouhan, S.S., Pilli, E.S., and Vipparthi, S.K., BerConvoNet: A deep learning framework for fake news classification, Appl. Soft Comput., 2021, vol. 110, p. 107614.

[3] Sabeeh, V., Zohdy, M., Mollah, A., and Al Bashaireh, R., Fake news detection on social media using deep learning and semantic knowledge sources, Int. J. Comput. Sci. Inf. Secur. (IJCSIS), 2020, vol. 18, no. 2.

[4] Agarwal, A., Mittal, M., Pathak, A., and Goyal, L.M., Fake news detection using a blend of neural networks: an application of deep learning, SN Comput. Sci., 2020, vol. 1, no. 3, pp. 1–9.

[5] Kaliyar, R.K., Goswami, A., and Narang, P., DeepFakE: improving fake news detection using tensor decomposition-based deep neural network, J. Supercomputing, 2021, vol. 77, no. 2, pp. 1015–1037.

[6] Jadhav, S.S. and Thepade, S.D., Fake news identification and classification using DSSM and improved recurrent neural network classifier, Appl. Artif. Intell., 2019, vol. 33, no. 12, pp. 1058–1068.

[7] Ahmad, I., Yousaf, M., Yousaf, S., and Ahmad, M.O., Fake news detection using machine learning ensemble methods, Complexity, 2020.

[8] Akinyemi, B., Adewusi, O., and Oyebade, A., An Improved Classification Model for Fake News Detection in Social Media, Int. J. Inform. Technol. Comput. Sci. (IJITCS), 2020, vol. 12, no. 1, pp. 34–43.

[9] Abdullah, A., Awan, M., Shehzad, M., and Ashraf, M., Fake news classification bimodal using convolutional neural network and long short-term memory, Int. J. Emerg. Technol. Learn., 2020, vol. 11, pp. 209–212.

[10] Hakak, S., Alazab, M., Khan, S., Gadekallu, T.R., Maddikunta, P.K.R., and Khan, W.Z., An ensemble machine learning approach through effective feature extraction to classify fake news, Future Gener. Comput. Syst., 2021, vol. 117, pp. 47–58.

[11] Khan, J.Y., Khondaker, M.T.I., Afroz, S., Uddin, G., and Iqbal, A., A benchmark study of machine learning models for online fake news detection, Mach. Learn. Appl., 2021, vol. 4, p. 100032.

[12] Sulthana, A.R. and Ramasamy, S., Ontology and context based recommendation system using neuro-fuzzy classification, Comput. Electr. Eng., 2019, vol. 74, pp. 498–510.

[13] Jelodar, H., Wang, Y., Orji, R., and Huang, S., Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach, IEEE J. Biomed. Health Inf., 2020, vol. 24, no. 10, pp. 2733–2742.

[14] Ali, F., Kwak, D., Khan, P., Ei-Sappagh, S.H.A., Islam, S.R., Park, D., and Kwak, K.S., Merged ontology and SVM-based information extraction and recommendation system for social robots, IEEE Access, 2017, vol. 5, pp. 12364–12379.

[15] Monti, F., Frasca, F., Eynard, D., Mannion, D., and Bronstein, M.M., Fake news detection on social media using geometric deep learning, arXiv preprint arXiv:1902.06673, 2019.

[16] Sahoo, S.R. and Gupta, B.B., Multiple features based approach for automatic fake news detection on social networks using deep learning, Appl. Soft Comput., 2021, vol. 100, p. 106983.

[17] Lu, Y.J. and Li, C.T., GCAN: Graph-aware co-attention networks for explainable fake news detection on social media, arXiv preprint arXiv:2004.11648, 2020.

[18] Kulkarni, A. and Shivananda, A., Converting text to features, in Natural Language Processing Recipes, Berkeley, CA: Apress, 2021, pp. 63–106.

[19] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., and On, B.W., Fake news stance detection using deep learning architecture (CNN-LSTM), IEEE Access, 2020, vol. 8, pp. 156695–156706.

[20] Lv, S., Zhang, H., He, H., and Chen, B., Microblog rumor detection based on comment sentiment and CNN-LSTM, in Artificial Intelligence in China, Singapore: Springer, 2020, pp. 148–156.

[21] Ghanem, B., Rosso, P., and Rangel, F., An emotional analysis of false information in social media and news articles, ACM Trans. Internet Technol. (TOIT), 2020, vol. 20, no. 2, pp. 1–18.

[22] Tharwat, A., Parameter investigation of support vector machine classifier with kernel functions, Knowl. Inf. Syst., 2019, vol. 61, no. 3, pp. 1269–1302.