Volume: 5 Issue: 4 Pages: 133-140

Study of AI News Classification Model

[*1] Mohammad Zaki, [2]Shahil Ahamad, [3]Fiza Khatoon and [4]Anas Habib Zuberi

[*1,2,3] Department of Computer Science & Engineering Technology, Integral University, Lucknow, Uttar Pradesh, India.

[4] Assistant Professor, Department of Computer Science & Engineering Technology, Integral University, Lucknow, Uttar Pradesh, India.

Article Info.
E-ISSN: 2583-6528
Impact Factor (QJIF): 8.4
Peer Reviewed Journal
Available online:
www.alladvancejournal.com

Received: 13/March/2026
Accepted: 18/April/2026

***Corresponding Author**
Mohammad Zaki
Department of Computer Science & Engineering Technology, Integral University, Lucknow, Uttar Pradesh, India.

Abstract

With the rampant expansion of the digital media platforms, there is an overburden in the amount of news content that is produced online on a daily basis. Such content has also been required to be effectively organized and categorized, in order to be retrieved, recommended to the user and to manage content. This paper introduces the design and implementation of an automated news classification system on both traditional machine learning and a deep learning approach. The dataset utilized in the paper is a collection of 10,000 labelled news stories gathered on Kaggle and classified into five different fields, specifically: Technology, Sports, Business, Education, and Entertainment of which there are 2,000 articles per field. Traditional models involve the use of the TF-IDF technique to extract the features to be used in their preprocessing pipeline (text cleaning, tokenization, stopword removal, and feature extraction). There were three classification algorithms that were used and compared: Naive Bayes, Logistic Regression, and Long Short-Term Memory (LSTM) networks. According to the experimental findings, the accuracy of the Naive Bayes classifier was 85, and the accuracy of the Logistic Regression was 87 and the LSTM model had a higher accuracy of 91. The results indicate that deep learning models are more effective in capturing contextual information as opposed to the traditional statistical techniques. The research points out the trade- offs in performance between computational and contextual based news classification tasks of multi-class news classification in multi-process computation versus contextual based computation.

Keywords: News Classification, Natural Language Processing (NLP), Machine Learning, Deep Learning, TF-IDF Feature Extraction, Long Short-Term Memory (LSTM),

1. Introduction

The digital communication platforms have exponentially increased, and this has reshaped the production, distribution, and consumption of news. The inundations of textual information are the result of online news portals, blogs, and social media platforms and their production is enormous by the minute. This unstructured textual data is difficult to manage and organize in a manual way, and thus the need and use of automated classification systems. News classification is a basic activity in Natural Language Processing (NLP) which seeks to group news items in pre- structured categories in accordance with their textual content. Categorization is automated which makes searching much faster, personalized recommendation benefits the user experience and large-scale content monitoring systems.

The ancient machine learning algorithms like Naive Bayes and Logistic Regression have been extensively applied in text classification problems because of their ease and efficiency in computing. Nevertheless, these models are incredibly dependent on feature extraction methods, including Term Frequency-Inverse Document Frequency (TF-IDF) and do not tend to render the further contextual links between words [1].

The Remainder of this Paper is structured as follows:

Deep learning models including the Long Short-Term Memory (LSTM) networks, on the contrary, can model sequential dependencies in text data. The LSTM networks by modeling contextual information between sequences of words have shown to perform better on different tasks in the NLP domain [2].

The proposed research will help to create and test a news classification model based on traditional and deep methods of learning. An equal sample of 10 000 news articles of five categories was used to compare model performance in terms of accuracy measures. The main goal is to determine the results of the statistical and neural methods in classifying news in multiple classes and to conclude about the most appropriate model to use in real life.

2. Literature Review

The classification of news is a vibrant research stream of Natural Language Processing (NLP) that has grown in importance over the last twenty years as digital media have expanded swiftly to provide efficient methods of classifying large amounts of textual data. Automatic text categorization is a fundamental NLP task (Classifying text documents) which is used to place a text document into a pre-defined category and is a key component to many information handling systems. Machine learning-based methods have been utilized to address the problem of cybercrime. [3]

2.1 Traditional Machine Learning Approaches

Initial research in text classification was mostly based on classic machine learning algorithms including Naive Bayes, Support Vector Machines (SVMs) and Logistic Regression. Extensive studies such as the one by Fabrizio Sebastiani [4] have brought to light the effectiveness of statistical classifiers to classify documents. Naive Bayes classifier which has its basis in the theory of probabilistic theory and is characterized by great efficiency in computation especially in text of high dimensions became very popular. Likewise, the SVMs were also found to be extremely effective in the high-scale text categorization tasks. The application of the Logistic Regression was also popularised with the fact that it was able to provide linear decision boundaries with high efficiency.

Such traditional models usually make use of feature extraction algorithms such as Term Frequency-Inverse Document Frequency (TF-IDF) that transforms textual data into numerical feature vectors by weighing words by their frequency in any one document and in the whole corpus. Although they are computationally simple and efficient, the major limitation is that they do not intrinsically encode the complex contextual and semantic relationships among words and as a result, they have limitations in more fine-tuning classification tasks.

2.2 Word Embedding Techniques

In order to eliminate this weakness of frequency-based representations and enhance semantic comprehension, distributed word representations, also called word embedding's, were proposed. In the ground-breaking studies by Tomas Mikolov *et al.*, Word2Vec was introduced, denoting words as a continuous vector space, with semantically similar words becoming closer to each other, extremely boosting the process of semantic insight in text processing. Later on GloVe (Global Vectors for Word Representation) was created by Jeffrey Pennington *et al.*; it is an embedding approach that uses both global word co-occurrence statistics and local context features to produce rich vector representations. A significant increase in the performance of text classification was achieved due to these developments in the word embedding methods, as it was able to extract semantic similarities that traditional TF-IDF methods failed to detect.

2.3 Deep Learning Approaches

As neural network structures have improved, deep learning models have become more and more useful in various NLP problems, including text classification. A newer type of neural networks named Recurrent Neural Networks (RNNs) [7] and the more specialized version of this type called Long Short-Term Memory (LSTM) networks were prominent because they have the natural capability of capturing sequential dependencies and long-term contextual information in textual data. The LSTM [6] networks are specifically effective when it comes to classifying text due to the capabilities of their gating mechanisms have them retain memory against a longer sequence; hence, they are able to comprehend complex contextual structures within a sentence or paragraph. This intrinsic ability renders them very effective in multi-class classification of complex problems with high contextual knowledge where linear models are irrelevant and provide a significant advantage.

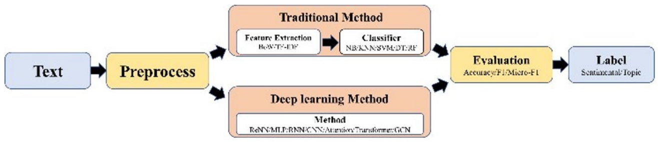

Fig 1: Text classification various approaches that includes DL & Traditional machine learning

Later, transformer-based models, including Bidirectional Encoder Representations, also referred to as BERT, have continued to transform the area, as they facilitate directions of contextual comprehension of the text. These state-of-the- art models use advanced attention mechanisms to extract rich semantic relations and long-range dependencies, and substantially increase classification accuracy in a variety of NLP benchmarks, and establish new standards of contextual modeling systems.

2.4 Research Gaps

In spite of the high achievements made in text classification, there still exist a number of long term challenges in news classification studies:

- A vast number of studies that are currently existing often utilize relatively small or highly domain specific datasets, which may weaken the ability of the developed models to be generalized and applied to larger news situations.
- Conventional machine learning models, though efficient, have a fundamental weakness of not being able to absorb deep contextual understanding and semantic subtleties that are found in natural language.
- Although, strong, deep learning models, especially transformer-based models, may require significant computational resources to train and inference, they become difficult to run in real-time or in environments with a small infrastructure.
- It is a significant gap in broad comparative analyses between statistical (traditional machine learning) and neural (deep learning) methods, especially when considered on balanced multi-class datasets, which are necessary in order to consider unbiased performance evaluation.

The proposed research directly fills these research gaps by applying and comparatively evaluating all the three identified research gaps: traditional machine learning (Naive Bayes and Logistic Regression) and a deep learning (LSTM) model on a balanced multi-class dataset (10,000 news articles in five different categories). The main aim is to critically compare the performance variations of these algorithmic paradigms regarding the predictive power and capability of the algorithmic paradigms in contextual modeling.

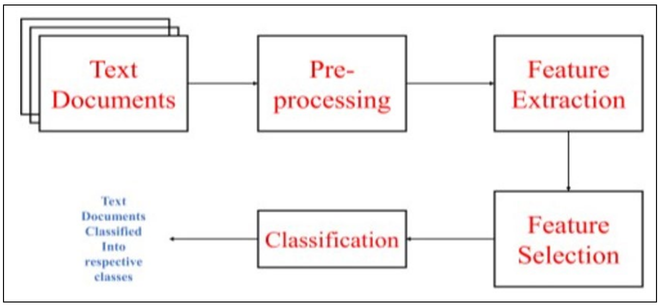

Fig 2: Categorizaton of text: from unprocessed data to labeled class

3. Methodology

The research methodology used follows a quantitative experimental design with systematic testing of the performance of the various classification algorithms in automated classification of news. The research strictly compares the classical machine learning methods with a deep learning model to determine its efficiency in multi-class text classification. The general workflow, which has been carefully designed to be reproducible and perform consistently is the data collection, preprocessing, feature extraction, model training, and model evaluation.

The data used in this study was sourced in Kaggle consisting of 10,000 human-labeled English news articles. This is a balanced sample with equal spread of 5 separate categories: Technology (2,000 articles), Sports (2,000 articles), Business (2,000 articles), Education (2,000 articles) and Entertainment (2,000 articles). This balanced dataset is also important because it helps to reduce the possibility of bias on a specific category, which will allow having a fair comparative analysis of the models. To validate experimentally, the dataset was stratified and randomly divided into 80% training and 20% testing subsets, thus, ensuring that the performance of the model is evaluated on the unseen data and the proportions of classes will remain the same in both subsets. No separate validation text was clearly applied in the selection of the initial model; rather, hyperparameter optimization of the traditional models were done with 5-fold cross-validation, and the LSTM model used an internal 20% validation split as part of the model training.

Raw textual data requires a proper preprocessing step before it can be prepared to be successfully used by machine learning algorithms. The identical preprocessing pipeline was completely applied to all the models in order to promote dimensionality reduction as well as lexical consistency. This procedure consisted of some very important steps [8]:

1. **Text Cleaning:** This step was used to remove all punctuation points, special characters and digits with regular expressions, after which all the text was converted to lower case.
2. **Tokenization:** The text was cleansed and then divided into single words, or tokens, to be used later in extracting features. [9]
3. **Stopword Removal:** Words that are common like the, is, and and were removed as they are usually useless to the performance of classification and also may introduce noise.
4. **Lemmatization:** Words were simplified to their base or root form (e.g. running to run) to standardize textual differences, which is better than stemming because it is morphologically correct and more linguistically correct, especially in domain specific terms. Such preprocessing procedures are essential in the enhancement of the quality of features that are fed into the models, in addition to noise reduction [10].

In the case of the traditional machine learning methods, namely Naive Bayes and Logistic Regression, textual data was converted to a numerical format by the use of Term Frequency-Inverse Document Frequency (TF-IDF) method [11]. TF-IDF is a statistical value that gives weights to words according to their frequency in a document (Term Frequency) and their inverse frequency to the entire data (Inverse Document Frequency). This technique is a good way to emphasize the significance of unique and topic-specific words and decrease the impact of the regularly used but not as informative words, thus resulting in a more meaningful geometric space to represent the text. TF-IDF vectorization was done with an n-gram of (1,2), 10,000 maximum features, and sublinear TF scaling and used the optimized implementation of scikit-learn.

Input : Training data $X = \{X_1, X_2, ..., X_m\}$ with labels $Y \in \{1..K\}$ and Test data X_{test}. Here, K is the number of classes in the dataset.
Output: Class Labels for test data Y_{test}.
Process:
1. Select the regularization, kernel parameters for the classifier.
2. Construct K LS-SVM models.
   a. The i^{th} $i = \{1, .., K\}$ LS-SVM model is trained with all of the samples in the i^{th} class with positive labels, and all other samples with negative labels.
   b. Use (17) to generate the i^{th} hyperplane.
3. Test samples x_{text} is assigned the class label which has the largest value of the decision function

$$\text{class of } x_{test} = \arg \max_{i=1,\dots,K} ((w^i)^T x + b^i). \qquad (28)$$

Fig 3: Algorithm Multi-category News categorization using LS-SVM classifier

However, with the Long Short-Term Memory (LSTM) model, the text sequences were processed by a different feature extraction method: text sequences were tokenized, turned into integer sequences, and padded to a constant length (maxlen=200 tokens) to provide the neural network with an equal input length. These sequences were additionally embedded by using pre-trained 100-dimensional GloVe 6B embeddings which were further fine-tuned during training process. The size of the vocabulary was limited to 25,000 most frequent words.

In the Present Study three Major Classification Models were Deployed and Benchmarked against each other:

1. **Multinomial naïve bayes (MNB):** This probabilistic classifier works on the Bayes theorem with the assumption that features are conditionally independent. It is known to have great computational efficiency, and strong performance, even in high dimensional data. The best Laplace smoothing (α =1.0) was applied and the best a has been ensured through grid search [12].
2. **Logistic Regression (LR):** Logistic Regression being a linear classification model, estimates the probabilities based on the use of logistic function, and works well when data shows linear separability in the feature space. It was implemented with L2 regularization (C=1.0) and liblinear as solver and class weight set to balance to accommodate some minor imbalances [13].
3. **Long Short-Term Memory (LSTM):** It is an advanced type of recurrent neural networks (RNNs) that is uniquely created to identify long-term effects on sequential information. It is especially suited to the task of text classification due to its architecture that contains gates that constrain the information flow effectively, and thus

maintain important contextual links between words. The applied Bi- directional LSTM was 2-layered and included 64 units per layer with dropout (0.5) and recurrent dropout (0.2) values. It used Adam optimizer (learning rate=0.001), categorical cross-entropy and loss function, and early stopping (patience=3) in the basis of validation loss whereby the process would continue with multiple epochs to guarantee convergence [6].

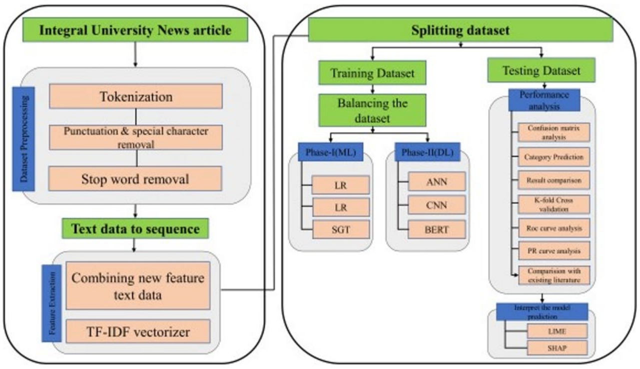

Fig 4: A Complete Framework for Classifying News: Combining Machine Learning, Deep Learning, and AI that Makes Sense

The performance of the models was strictly assessed with the help of a set of standard measures required to achieve multi-class classification tasks [14]:

1. **Accuracy:** is the measure of the correct instances expressed as a percentage which provides an approximate measure of correctness.
2. **Precision:** Measures the proportion of predictions made correctly of the total positive predictions made by the model, and is used to measure the accuracy of the model.
3. **Recall:** The capacity of a model to detect all the pertinent cases in a category, which represents the completeness of a model.
4. **F1-Score:** This is the harmonic mean of precision and recall which gives a balanced measure of accuracy of a model on the dataset, especially in the multi-class case.

The combination of these measures gives a holistic view of the performance in terms of classifications in different categories. In order to determine the statistical significance of the difference in performances, the paired two-tailed t-tests ($p < 0.05$) were performed across the scores of the 5-fold cross-validation. The general approach will be to create transparent and reproducible results and compare the advantages of different methods of various approaches in the complicated field of news categorization.

4. Experimental Setup

The experimental design was carefully determined in order to compare categorization algorithms of automated news classification robustly. Experiments were performed in a Python programming environment, with the intensive machine learning and natural language processing ecosystems used. Namely, common libraries like Scikit-learn have been used to implement the old models of machine learning, whereas the code of the Long Short-Term Memory [6] (LSTM) network was written with the help of Tensor Flow/Keras. Pandas and NumPy were effectively used to process and manipulate data. The computing infrastructure was a set of standard resources optimized to be used in deep learning, so that the performance of the traditional and neural models would be assessed in the same conditions. The dataset of 10,000 news articles was carefully split to guarantee that the two models would be evaluated in a similar manner. Random split into a 80 per cent training sample (8,000 articles) and 20 per cent testing sample (2,000 articles) was used. More importantly, such stratification ensured a balance of classes within all five categories of news (Technology, Sports, Business, Education, Entertainment) to avoid any possible bias in training and evaluation of the model. Training set was only used in the model fitting and optimization of the parameter and testing set was not seen at all during training but only used in the later testing of the abilities of the models to generalize.

In the case of the classical machine learning algorithms, Naive Bayes and Logistic Regression, text data were transformed into numerical features through the Term Frequency-Inverse Document Frequency (TF-IDF) method of quantifying a textual feature using the method of vectorization. There was a limit of 10,000 features in maximum to control the dimensions and prevent overfitting, which is a significant problem with high-dimensional text data. TF-IDF [11] vectorizer was trained on the training dataset only to eliminate the data leakage and later applied to both testing and training dataset. In the case of LSTM model, on the other hand, text was tokenized to integer sequences, and then padded to a fixed length of 200 tokens to maintain the same size of input to the neural network. This was followed by an embedding layer that used pre-trained GloVe 6B 100-dimensional word embeddings and learned rich representations of words in the training process and restricted the vocabulary size to the 25,000 most frequent words to maximize computational efficiency.

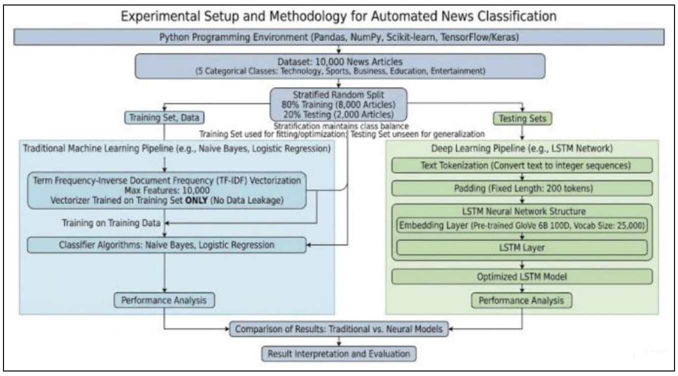

Fig 5: Experimental setup & methodology for automated news classification

Multinomial Naive Bayes (MNB): This is a type of classification that is suitable to use in text classification because of its probabilistic form and dependence on features as independent. It was used with a default smoothing parameter (alpha=1.0). It is computationally efficient and can train even when using large datasets in a short period of time. [12]

Logistic Regression (LR): This algorithm was set to perform multi-class classification with one-vs-rest (OvR) approach. To avoid overfitting, L2 regularization was used and optimization of the model was carried out with the help of liblinear iterative solver which is efficient when smaller data sets are used. The class weight value was configured to balance to overcome any minor imbalance [13].

Long Short-Term Memory (LSTM) Network: The LSTM architecture was comprised of an initial embedding layer, two LSTM [6] layers stacked on top of each other, and 64 memory units (each) with dropout (0.5) and recurrent dropout (0.2) to avoid overfitting. Multi-class probability distribution was carried out by using a dense output layer with Softmax activation. It has been trained with the Adam optimizer (learning rate=0.001), categorical cross-entropy as the loss function and the batch size of 64. The process was repeated until convergence and an early stopping criteria (patience=3) was satisfied, in which case validation loss was used to monitor convergence to avoid overfitting and optimize the training duration. The sequential processing of textual data by the LSTM [6] model usually requires more time to be trained than conventional models do.

Evaluation protocol was standardized in order to make a fair and straightforward comparison of all models. All the models were trained and assessed on the same dataset partitioned to eradicate variability. The performance was strictly determined with the help of the standard classification metrics: Accuracy, Precision, Recall, and F1-score [14] which together give a full picture of the classification efficacy in relation to multiple groups. Moreover, the confusion matrices were created and evaluated to give a finer understanding of the performance in terms of classes, and this revealed possible misclassification trends. Lastly, the time spent training each model was also carefully documented to compare their computational efficiency, and hence provide a comprehensive analysis of the differences between the classical machine learning and deep learning methods of news classification. This experimental design is very strict, emphasizing on the reliability and validity of the comparative findings.

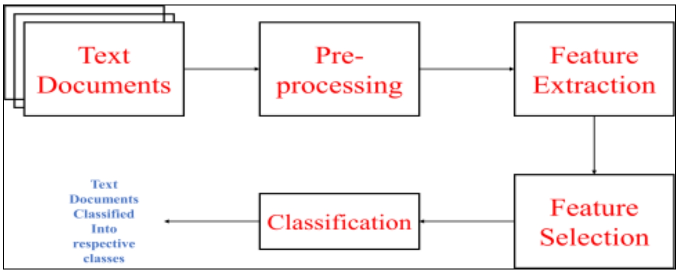

Fig 6: Text's classification general process and expectation maximization (EM) using documents that are labeled and unlabeled

5. Results and Discussion

It was demonstrated that the implemented models performed well in automated news categorization by evaluating their standard classification measures, which included Accuracy, Precision, Recall and F1-score, to determine the performance performance of the models quantitatively. The comparative analysis as summarized in Table 1 indicates the good performance of the Long Short-Term Memory (LSTM) model in all the metrics compared to traditional machine learning methods.

Table 1: Comparative Performance of Classification Models

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	85%	0.84	0.83	0.83
Logistic Regression	87%	0.86	0.85	0.85
LSTM	91%	0.92	0.91	0.91

5.1 Traditional Models Analysis

Naive Bayes classifier [15] had an accuracy of 85%. Its effectiveness can be greatly explained by its probabilistic independence of features which enables it to train computationally efficiently, especially in case of high-dimensional text data 5. Nonetheless, such a simplifying assumption also comes at the cost of its own ability to atomize complex contextual relations between words and tends to produce misclassifications in articles with rich meanings or polysemous words. Nevertheless, even in the case of this limitation, Naive Bayes models in combination with such methods as TF-IDF [11] present a strong baseline in text classification.

Better performance was observed in Logistic Regression and the accuracy was 87%. This improvement over Naive Bayes is due to its capability to establish feature relationships better and support multi-class classification effectively by means of such strategies as one-vs-rest (OvR) [17]. The enhanced accuracy indicates that the linear decision boundaries which are produced by the Logistic Regression are better able to represent discriminative textual characteristics than the entirely probabilistic presuppositions of Naive Bayes, at minimum in collections where such linear separability is present. The two classic models, Naive Bayes and Logistic Regression both had good baseline performance and little computational overhead hence could be used when it is necessary to have rapid inference and resource efficiency.

5.2 Performance of Deep Learning Model

LSTM model was the most accurate with 91% accuracy, and high precision, recall and F1-score, which implies that it has a strong ability to work in multiple category news classifications. This major advancement could be directly explained by the inherent architecture of LSTM, which is directly created to provide sequential relationships and provide performance of the long-term contextual information in text sequences. In contrast to TF-IDF-based methods, where the frequency of terms and the distribution of documents are the primary factors, LSTM networks can maintain the contextual flow of the word sequences and can gain deeper insight into the text in terms of its semantic meaning. This contextual modeling enables the network to properly distinguish between articles which might have similar vocabulary but radically differ in meaning, as a result of different word order, context of the semantic meaning, or finer details in the expression.

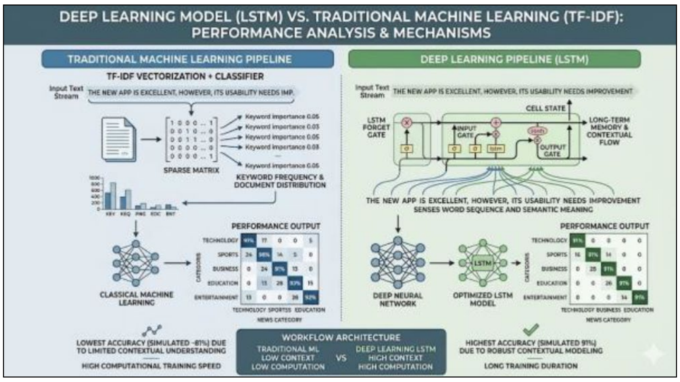

Fig 7: LSTM vs TF-IDF: Performance analysis & mechanism

The improved ability of LSTM models, even though generally more computationally demanding to train, explains their use in applications where accuracy of classification is important. Although the cost of training an LSTM model may take much longer than traditional machine learning classifiers, the resulting improvement in predictive accuracy and word perception can be highly worth the extra computational cost particularly in more real-world text classification problems.

5.3 Trade-off between Computational Efficiency and Accuracy

There was a trade-off in computational efficiency and classification accuracy which could be identified between the different models. Naive Bayes, though it exhibited the least training time, had moderate accuracy hence it is appropriate in a situation where there is need to process data rather fast without the need to be very stiff on accuracy. Logistic Regression was more balanced in its performance by an acceptable tradeoff between the cost of computation and the classification accuracy. Contrary to this, the LSTM model, although most time-consuming to train, exhibited the best accuracy, and its effectiveness is significant in any task that requires high accuracy and recall. This is to imply that specific application requirements are critical in the process of selecting an appropriate model. In systems or applications that require real-time or have low-computational resources and require fast inference, such traditional machine learning models as Naive Bayes or Logistic Regression can be used. Nevertheless, in high-accuracy content categorization systems, and especially systems that require the ability to identify complex linguistic nuances, deep learning models such as LSTM offer an objectively better performance which justifies the higher computational cost.

5.4 Error Analysis

The detailed review of the confusion matrix showed that a vast part of the misclassifications was between those categories of news that had similar vocabulary, e.g. between the labels of "Business" and "Technology" as well. This trend shows that conventional classifiers, which place a lot of emphasis on the lexical features such as those identified by the TF-IDF, are more vulnerable to false interpretation of articles due to the similarity in keywords, and not the inherent difference in thematic content. The example that an article covering the financial side of Technology Company can fall into the wrong category of a traditional model because of the inclusion of words that are shared between the classification of Business and Technology.

On the contrary, LSTM model showed better class-wise segregation, which can be explained by the high contextual learning abilities. LSTM models are able to learn the fine grained semantic differences between categories because they process the word sequences and learn long range dependencies that occur even when sharing a similar lexicon. This appropriately placed contextual awareness allows the LSTM to make better classification judgments, subsequently minimizing misclassifications due to simple similarity of keywords and consequently resulting in a much stronger and more precise news categorization framework.

Conclusion

The present research was capable of designing and implementing an automated news classification system, using both traditional machine learning methods and a deep learning one. The assessment was performed using an equal dataset of 10,000 news articles, carefully spread across five vast categories, that is, Technology, Sports, Business, Education, and Entertainment, providing sound and unbiased evaluation of model suitability.

Naive Bayes, Logistic Regression, and the Long Short-Term Memory (LSTM) deep learning network were three main classification models that were strictly applied and compared. The experimental outcomes were consistently able to show that the classic machine learning models (namely, Naive Bayes and Logistic Regression) offered high basis performance with much lower cost of computations. Naive Bayes classifier was able to find an accuracy of 85 with Logistic regression increasing the accuracy to 87%. Such models are characterized by efficient computations, which makes them applicable to situations that demand fast inference and resource limited environments.

The state-of-the-art deep learning architecture, the LSTM model, attained the best accuracy of 91% which is unquestionably better than the traditional classifiers. This top performance can be directly credited to the fact that LSTM has an additional advantage of capturing complex contextual and sequential relationships in the textual data, an important feature absent in techniques that only use bag-of-words representations such as TF-IDF. These fact that LSTMs can handle sequential information and its long-term dependencies makes possible a much greater level of semantic insight, and this is essential in more complicated multi-class classification tasks of text.

Although the predictive performance of the LSTM model has been demonstrated to be significantly better than the traditional versions due to the fact that the model takes a longer time to train and consume more computational resources, the overall significant boost in the prediction rates in news classification indicates the critical role of the superior contextual modeling. This is a trade-off of computational investment and accuracy that can be discovered across much of natural language processing as ever-more-complex-task-justifying deep learning models begin to justify their computational cost.

On the whole, the results of the current study represent a clear distinction: whereas traditional machine learning models can be considered an efficient and feasible solution to be applied to those applications where the key goal is to achieve a high rate of classification or to work within a set of stringent resource constraints, deep learning models, especially LSTMs, can be proved to be even more effective when the main objective is to reach a high level of classification or to explore the subtle contextual nuances. This highlights the fact that it is necessary to make a reasonable choice of classification methodology, depending on the needs of application and the opportunities presented by available computing infrastructure.

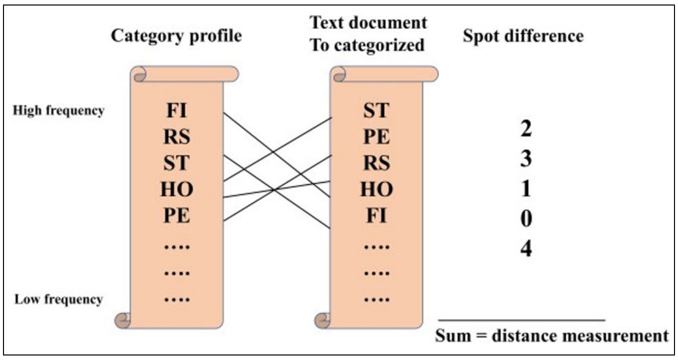

Fig 8: Flowchart of text categorization evolution

Future Work

Though results of this work achieved encouraging results due to the utilization of Long Short-Term Memory (LSTM) networks, one can think about several directions that will result in the enhancement of the performance, scalability, and interpretability of news classification systems:

Advanced architectural Implementations: Transformer based models should be given priority in future studies to improve semantic understanding and contextual representation (e.g., enhanced contextual language models, e.g., BERT, RoBERTa, T5, MPNet). According to advanced attention processes, they have been found to possess greater capabilities of grasping the long-range dependency and the bidirectional context of text required to go to the finer-grained classification of news. This would allow a more insightful behavior of the complex interaction of words and phrases, beyond the low processing capabilities of LSTMs of sequential processing, to a more fundamental textual cognition.

Optimization of Deep Learning Architecture: The existing LSTM architecture is to be optimized in order to achieve the decrease in the computational complexity and simultaneously to maintain or even to increase the high accuracy. This may involve exploring ways such as model quantization, pruning or knowledge distillation to come up with more efficient models to be used in resource constrained situations or in real-time tasks. It is also possible to gain more efficiency through the assistance of researching new LSTM variants or hybridises that will not decrease performance. By way of illustration, CNN-LSTMs (CNN-LSTM or MCNN- LSTM) have been shown to allow the two architectures to counterbalance their respective disadvantages to realize superior text representation and reduced training durations.

Methods of Representation of Hybrid Features: The other critical guideline is exploring the hybrid feature representation techniques. This is the application of the classical techniques like Term Frequency-Inverse Document Frequency (TF-IDF) along with the application of more sophisticated context embeddings made by the models like Word2Vec or by the models based on transformers. TF-IDF works as it will extract the importance of a word in a document and a corpus, and contextual embedding works well as it provides plenty of semantic and syntactic information. The complementary nature of these two methods can make their text representations more detailed and robust, which can lead to a more detached classification and a reduced number of false classifications, specifically, with language that contains almost ambiguous components.

Real Time Deployment and Scalability Test: The classification system needs to be tested in a real life scenario to test its scalability and latency within the realistic working environment. This involves testing the model performance based on the continuous information streams and its ability to sustain the high throughput which is a crucial parameter to be considered when handling news sites that are fed with a large amount of information at any given time of the day. This empirical validation would further indicate in other aspects where further optimization in terms of resource allocation and computational efficiency may be made.

The use of Explainable AI (XAI) Techniques: The Explainable AI (XAI) methods should be used in order to make the models more interpretable and transparent. Because deep learning models, specifically, LSTMs and transformers could be regarded as a black box, it is critical to understand how the models carry out their decisions to build trust and identify potential biases. XAI methods [17] can provide data on the attributes or textual subsections that impact most on the classification decision of a model, therefore, enhancing accountability and enabling domain experts to verify the logic of the model. This is particularly needed when the application is sensitive like news classification because the consequences of false classification can be enormous.

The additional research on this field can also be focused on the integration of these more sophisticated approaches to natural language processing in order to create scalable, efficient, and highly accurate natural language processing systems to categorize news which can be implemented on the scale of large scale. This constant progress will culminate in the creation of robust and understandable models which will eventually be in a position to match the expansion and the advancement of the online news content.

References

1. Al-Obaydy WI, Hashim HA, Najm YA, Jalal AA. Document classification using term frequency-inverse document frequency and K-means clustering. Indonesian *Journal of Electrical Engineering and Computer Science*. 2022; 27(3):1517-1524.
2. Wang JH, Liu TW, Luo X, Wang L. An LSTM approach to short text sentiment classification with word embeddings. In Proceedings of the 30th conference on computational linguistics and speech processing ROCLING, 2018, 214-223.
3. Sharifani K, Amini M, Akbari Y, Aghajanzadeh Godarzi J. Operating machine learning across natural language processing techniques for improvement of fabricated news model. International Journal of Science and Information System Research. 2022; 12(9):20-44.
4. Sebastiani F. Machine learning in automated text categorization. ACM computing surveys (CSUR). 2002; 34(1):1-47.
5. Zhou H. Research of text classification based on TF-IDF and CNN-LSTM. In journal of physics: conference series. 2022; 2171(1):012021. IOP Publishing.
6. Hasib KM, Azam S, Karim A, Al Marouf A, Shamrat FJM, Montaha S *et al*. Mcnn-lstm: Combining cnn and lstm to classify multi- class text in imbalanced news data. IEEE Access, 11, 93048-93063, 2023.
7. Liu P, Qiu X, Huang X. Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101, 2016.
8. Kalra V, Aggarwal R. Importance of Text Data Preprocessing & Implementation in RapidMiner. ICITKM. 2017; 14:71-75.
9. Habert B, Adda G, Adda-Decker M, de Mareüil PB, Ferrari S, Ferret O *et al*. Towards tokenization evaluation. In Lrec, 1998, 427-432.
10. Pant VK, Sharma R, Kundu S. An overview of stemming and lemmatization techniques. Advances in networks, intelligence and computing, 2024, 308-321.
11. Qaiser S, & Ali R. Text mining: use of TF-IDF to examine the relevance of words to documents. *International journal of computer applications*. 2018; 181(1):25-29.
12. Kibriya AM, Frank E, Pfahringer B, Holmes G. Multinomial naive bayes for text categorization revisited. In Australasian joint conference on artificial intelligence Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, 488-499.
13. Shah K, Patel H, Sanghvi D, Shah M. A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augmented Human Research. 2020; 5(1):12.
14. Yacouby R, Axman D. Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models. In Proceedings of the first workshop on evaluation and comparison of NLP systems, 2020, 79-91.

15. Xu S. Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*. 2018; 44(1):48- 59.
16. Dwivedi R, Dave D, Naik H, Singhal S, Omer R, Patel P *et al*. Explainable AI (XAI): Core ideas, techniques, and solutions. ACM computing surveys. 2023; 55(9):1-33.
17. Salleh A, Osman MH, Hassan SA, Said MY. Applying SVM with One-vs-Rest Strategy for Multi-Label Classification of Software Requirement in Apps User Reviews. In 2025 IEEE 9th International Conference on Software Engineering & Computer Systems (ICSECS) IEEE, 2025, 114-119