International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 22s, 2025
https://theaspd.com/index.php

A Natural Language Process based LSTM Framework for Keyword Extraction

Vishnu Teja Karumanchi[1], Venubabu Rachapudi[2]
[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India, kvishnutej1@gmail.com
[2]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India, venubabu.r@gmail.com

Abstract: *The keyword extraction means document contains collection of terms by keyword extraction, which is an automated process. It is the process of identifying keyword from essential content of a given document. The functionality of many NLP systems are enhanced by keyword extraction, which is crucial problem in many Natural Language Processing (NLP) applications. In keyword extraction by using traditional methods use small amounts of data and produce inaccurate results. The main strategy used to analyze a large number of documents and extract relevant data is Keyword extraction. Especially young people, spend a lot of time searching the internet for relevant data for gaining and absorbing knowledge is time-consuming process of learning from a different sources. Therefore, Natural Language Process Based Long Short-Term Memory (LSTM) Framework for Keyword Extraction is introduced. In this analysis, the Neural Information Processing Systems (NIPS) dataset was used from Kaggle.The suggested system is primarily focused on scraping data from websites that provide summaries and keywords from the information extracted from multiple websites. It also provides ability for users to choose website or text of their choosing. The system is trained and evaluated using the NIPS Dataset, which extracts high-quality keywords from user-generated reviews. The LSTM model compares with existing methods. Experimental results show that the proposed model achieves higher precision, recall, and F1-score. The results demonstrate its effectiveness in capturing contextual and semantic relevance over baseline methods. Therefore, this proposed system helps to identify the main topic and summarize the content of a text. Hence, this system shows better results interms of accuracy, precision, recall and F1-score.*

Keywords: *Keyword Extraction, Summarize, Long Short-Term Memory (LSTM), Natural Language Processing (NLP), Documents, Websites*

INTRODUCTION

In the modern world most valuable resource is data. The valuable information contained in vast data for different purposes by researchers. For textual data, there are two primary methods for presenting or extracting important insights. The processing of data and beginning to record the information manually will appears to be significant at first option which is manual analysis [1].

The relevant documents with interests are identifying huge volumes of textual data is quickly filtered and amount of documents is expanding at an increasing rate every day. These days, text files with hundreds of thousands and several million web pages are frequently stored [2]. The primary characteristics, concept, theme, etc. of the document will make it easier to analyze such huge amounts of data with subset of words (keywords). The document summarization can make easily by using appropriate keywords, which also make it easier to organize and retrieve documents according to their content. The reader understands the content with the help of keywords in academic articles. The key concepts in a given section of a textbook, it is helpful in helping readers recognize and remember. To represent the main idea of text, Keywords are used to measure similarity for text clustering [3].

By using computer tools, text is analyzed by automated text analysis, which is different method. The next method is more important because when a significant amount of text cannot be processed by manual process [4]. The annotating article is popular technique that reflects their primary content is Automatic Keyword/Keyphrase Extraction (KE) [5]. The data source is appropriate and pertinent to other important domains and/or subjects, it is crucial in the field of information retrieval is determined with help keywords from any piece of data [6]. The text's content is contained in the keyword is primary information that helps readers. The effectiveness of Information Retrieval (IR) will be increased by examining keywords to determine user's ability [7]. A several number of NLP applications, including text classification, text clustering, IR, and question answering systems have benefits because of keywords that are concise and refined, they can be used to calculate text correlation with little complexity. A several number of NLP methods are required for keyword extraction [8].

The process of automatically identifying keywords that can be used to model topics and represent the text is known as keyword extraction [9]. The amount of textual information must rapidly scan through these days to identify documents relevant to interests that are increasing every day. Several million web pages and tens of thousands of text files are frequently stored nowadays. Having subset of words (Keywords) that can provide major characteristics, concept, subject, etc. of document will make it easier to analyze such huge amounts of data. Appropriate keywords can act as very good summary of document and make it easier to organize and find materials based on their content [10]. In academic writing, keywords are utilized to inform the reader about the article's substance. In textbook, they are helpful for readers to recognize and remember key ideas in given section. Keywords can be utilized as gauge of similarity for text clustering because they capture essential idea of document [11].

The relevant data, which is extracted from a huge amount of data, and the process is known as keyword extraction [12]. The text classification, text clustering, tracking, topic detection, and summarization, are field that extracts keywords. The social media sites like Facebook and Twitter analyzes data for making decisions that depends extensively on social media analysis by keyword extraction. The users can broadcast and share information by Twitter, which is significant online social networking site for microblogging. The data with 140 characters or tweets length are generated by Twitter. In some cases an excessive amount of data is received, like on Twitter that makes it difficult for users to read and it becomes difficult to extract relevant information in this situation. Therefore, keyword extraction concept is implemented [13].

The language used in computer-human communication is NLP (Natural Language Processing), and it allows computers to understand human input languages are two of NLP's many challenges efficiently. NLP is particular branch of artificial intelligence, which is interested, difficult to represent and develop is natural language processing in one of the field [14]. The system which communicates with users in the manner rather than requiring them to learn a new language is developed by NLP. A huge amount of information is recorded by using numerous natural languages. The data that are available online at any time and from any location may be published in books, journals, research papers, reports, and other materials.

The different natural languages using data that is stored on systems is need lot of information to manipulate. Following is the arrangement of the remaining paper: in section II, the literature survey is provided, A Natural Language Process Based LSTM Framework for Keyword Extraction is described in section III; in section IV explains result analysis; section V concludes the paper and references are in VI.

LITERATURE SURVEY

W. Guo, Z. Wang and F. Han, et.al [15] BERT semantics and K-Truss graph(BSKT) is combined by multi-feature fusion keyword extraction method is proposed. The BSKT algorithm acts as basis for TextRank method, which integrates K-Truss features, BERT (Bidirectional Encoder Representations from Transformers) semantic features, and additional features. The BSKT method first collects word vectors from BERT pretraining model in order to compute semantic difference to optimize iterative process of the TextRank word graph. The BSKT model determines its BSKT and word's truss level feature by reducing TextRank word graph. By integrating truss level features and word (Inverse Document Frequency) IDF, BSKT model scores words to extract keywords. The most advanced keyword extraction algorithm (Semantic Clustering TextRank) extracts 1–10 keywords exceeded by BSKT algorithm in this experimental results. Therefore, it also observers 11.2% increase in F1-score.

Licciardo.G.D, Benedetto.L.D, Liguori.R, Rubino.A and Vitolo.P, et.al [16] introduces audio feature extraction in Keyword Spotting (KWS) using convolutional autoencoder- method that hasn't been investigated in the literature. The automate audio feature extraction process, maintain a low computational complexity for suggested method's strengths include its ability, and enable accuracy values of the entire KWS systems that are paring with the new methods. The Google speech command dataset which is available publicly evaluates efficacy of proposal by contrasting it with popular MFC (Mel Frequency Cepstrum) framewok in terms of number of necessary operators and classification metrics in noisy conditions. The MFC achieves 5.2%, average classification accuracy of twelve classes ranging from (81.84-90.36)% when signal-to-noise ratio spans range (0 -40 dB) in suggested audio feature extractor.

Shen.L, Li.R, Mao.X and Huang.S, et.al [17] individual documents keywords are extracted by using unsupervised framework that enhances keyword extraction in two ways are proposed. It uses length filtering, regular matching, and stopword removal in the candidate keyword selection step to decrease the quantity of candidate keywords while increasing their quality. The three different methods are combined for word co-occurrence and semantic relationships to determine weight of edges in graph model. It also uses word co-occurrence, semantic relationships (Normalized Google Distance, WordNet and Word Embedding,). In evaluation criteria, keyword extraction techniques suggested with other strong baseline techniques in two datasets using F1-Score, Recall and Precision, values. Therefore, suggested framework produce positive outcomes.

G. -W. Kim, W. -H. Kim, K. Chung and J. -C. Kim, et.al [18] extraction of high-quality video metadata with existing recommendation systems are combined to develop a new recommendation system. The Extraction of Meta-Data for Recommendation by using keyword mapping is suggested to construct contextualized data using object detection models and STT (Speech-To-Text) models. The public dataset MovieLens, is applied to hybrid recommendation system to extract and map keywords. The contextualized data is created by using process of Google's Speech-to-Text API (Application Programming Interface) and YOLO (You Only Look Once). The keywords are extracted by using TextRank algorithm, which are then mapped to the MovieLens dataset, and uses Hybrid Recommendation System.

Kadoch.M, Xiong.A, Yu.P, Tian.H, Liu.Z, and Liu.D, et.al [19] using TextRank, semantic clustering news keyword extraction framework is proposed by SCTR (Semantic Clustering TextRank). The BERT (Bidirectional Encoder Representation from Transformers) model produces k-means clustering is performed to represent semantic clustering using the word vectors, initially. By using the clustering results TextRank weight transfer probability matrix is created. The word graph computation and keyword extraction are done continuously, experiment's test target uses Chinese news library. The conventional TextRank and Term Frequency-Inverse Document Frequency (TF-IDF) methods in terms of recall, F1 score and precision, performs better by using SCTR algorithm.

He.D, Chen.B, Pu.L and Lin.C, et.al [20] Public-key encryption with keyword search (PEKS) is authenticated encryption of cryptographic, which is primitive that was suggested. The most existing schemes involve time-consuming bilinear pairing operations that are not suitable for IIoT (Industrial Internet of Things). The bilinear pairing operations are completely avoided by PAEKS scheme while creating the trapdoor and keyword ciphertext. The random oracle model uses decisional q-ABDHE and computational Diffie-Hellman for assumptions. Therefore, while performing theoretical and experimental comparisons it uses MCI (Multi-Ciphertext Indistinguishability) and trapdoor privacy. The proposal's computational overhead is greatly decreased without resulting in additional communication expenses or security compromises when compared to the majority of current classical PAEKS schemes.

Liu.H, Tang.J Yu.H, and Yang.Z, et.al [21] spatial distribution of specific text is proposed by using novel keyword extraction metric to enhance retrieval performance. A cloud-based vehicle social information retrieval method is created as solution in suggested framework. The retrieval of data with strict security and privacy preserving policies that conceptualizes constrained information retrieval in vehicle social networks. The comparing data sets and various evaluation metrics have been carried out in accordance with the proposed framework. The findings shows better and have a wider range of applications in this paper.

Z. -Z. Hu, J. –R Lin, L. -M. Chen, and J. -L. Li et.al [22] topic modeling and keyword extraction uses novel text mining technique to improve the decision-making process and also identifies main ideas and their dynamics of on-site problems. In a real-world project, suggested method was then tested with seventy two thousand fifty problems records. The suggested method could effectively extract important issues hidden in data and detect how they changed over time by using on-site inspection and data-centric decision-making process with more effective. This study provides (1) identifying important issues by developing a new framework and explains way they develop over time in texts, and (2) to minimize on-site problems with effective decision-making by advancing field and providing insights on current topics and way they develop over time.

S. Chang, G. -J. Ahn and S. Park, et.al [23] keyword extraction is integrated for data augmentation driven presents weakly supervised learning approach. The pseudo-queries by extracting keywords from corpus passages are developed. The pseudo-labels are generated by using well-known weak supervised learning techniques for developing relevance between these pseudo-questions and passages. The keyword extraction techniques can effectively formulate queries and train neural Information Retrieval (IR) systems demonstrates current synthetic query generation approach. The potential of pseudo-labeling techniques as useful alternatives when large amounts of ground truth data are unavailable for particular, models trained with pseudo-labels perform very similarly to models trained with ground truth data.

J. Duan, X. Liao, Y. An and J. Wang, et.al [24] Event Keywords Extraction's (EKE) low-resource event extraction is enhanced by multi-prompt learning technique KeyEE (Key Event Extraction). The EE (Event Extraction) and EKE are trained simultaneously by using shared pre-trained language model, and also uses an auxiliary EKE sub-prompt. The auxiliary sub-prompt, KeyEE reduces its dependence on annotated data by learning event keywords. The different EKE sub-prompt strategies are examined and evaluated to stimulate more research. The KeyEE uses ACE(Automatic Content Extraction)2005 and ERE (Event and Relation Extraction) benchmark datasets that significantly improves performance under low-resource conditions.

A. Amin, et.al [25] documents has been introduced for topic prediction in the Urdu language by using new unsupervised method that can extract more important information. The extracting keywords from the text and ranking by suggested TOP-Rank system that is based on where they appear in a sentence. These keywords and their ranking scores are used to create keyphrases by using syntactic rules to extract more significant topics. Based on the keyword scores keyphrases are re-ranked based on their positions within the document and ranked. The keyphrase with the

highest score is chosen as the document's topic by top-ranked keyphrases that are identified by suggested model with topical significance. The suggested system's performance is compared with the new methods currently in use by two distinct datasets is used in this experiment.

FRAMEWORK OF A NATURAL LANGUAGE PROCESS BASED LSTM FRAMEWORK FOR KEYWORD EXTRACTION

In this section, framework of an a natural language process based LSTM framework for keyword extraction is observed in figure.1. In this proposed system Neural Information Processing Systems (NIPS) dataset was used from Kaggle. For starting the process, input is given by input sentence. Then from that data unwanted data is removed in pro-processing step. Then the data is converted into digital form and in word segmentation the words are segmented. Therefore, by using Gensim (Generate Similar), the keywords are extracted. From the database, data is given to neural network as well as the data from extracted data is also given to neural network of LSTM algorithm. The twelve long input sequences are managed by the LSTM in this system. The size of vocabulary and embedding dimension used in the LSTM input layer is nine.Finally summarization is done by that extracted word.

The NIPS (now NeurIPS) dataset refers to the collection of papers published at the Neural Information Processing Systems conference, a leading machine learning and computational neuroscience conference. This dataset typically includes the title, authors, abstracts, and full text of papers from the conference proceedings, spanning from its inception in 1987 to the most recent years. In preprocessing original data table is converted into a new one, and the process is known as preprocessing. It involves various methods to modify the data before extracting latent vectors from it. In data processing, data that initially get ready for primary processing or additional analysis. The multiple steps which are needed to get the data ready for the user can be referred to by this term in any initial or pre-processing step. The NIPS dataset is preprocessed as text is cleaned by removing stopwords. It converts all words to lowercase and applies stemming or lemmatization. The TF-IDF or word embeddings methods will change text into number form. The LSTM sequences are padded to equal length and it is encoded for training. The class distribution of dataset is includes optimization, neural networks, reinforcement learning, NLP, and computer vision.

The data which is gathered and recorded from actual physical phenomena and transforming it into digital format that computer can analyze, the process is known as DAQ (Data Acquisition). Data can be acquired in four ways: sharing/exchanging information, converting/transforming legacy information, gathering new information, and purchasing information. The automated collection (such as sensor-derived data), and gathering existing data from other sources that includes manual recording of experimental observations. Word segmentation is the process of identifying word forms from a string of characters. It's a key step in Natural Language Processing (NLP) tasks like parsing, tags part-of-speech, and machine translation. Word segmentation can also involve adding spaces between words.

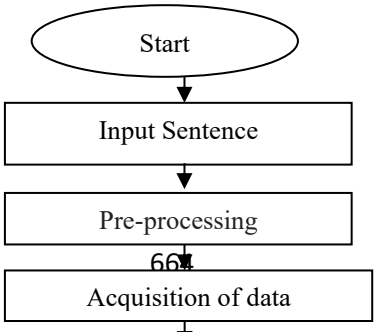

Figure 1: Framework of a Natural Language Process Based LSTM Framework for Keyword Extraction

In Python, open source Gensim library is created by Radim Rehurek that is utilized in unsupervised and NLP topic modeling. It is developed to extract semantic topics from documents, and it can handle large text collections. It mainly focuses on memory processing, so it is different from other machine learning software packages. The effective multicore implementations of a number of algorithms to increase processing speed is offered by Gensim. The practical text processing features are more while compared with packages like Scikit-learn, R(R is a statistical programming language with numerous packages for data analysis and machine learning), and others.

The field which integrates Artificial Intelligence (AI), linguistics, and computer science, is NLP (Natural language processing), which is interesting and quickly developing. The interaction between computers and human language is monitored by NLP. It also allows machines to understand, interpret, and produce meaningful practical human language. The automating various tasks extracts valuable information from the increasing volume of text data generated everyday from research articles to social media posts by NLP and it became an important tool.

A neural network is machine learning algorithm that uses interconnected nodes, called neurons, to solve complex problems. The human brain inspires neural networks are capable of tasks like image recognition, NLP and machine translation. It is a computational learning system that utilizes network of functions to process data input and produces the desired output, often in a different format. It consists of interconnected nodes, or perceptrons, which apply non-linear activation functions to the input data.

The architecture of a Long Short-Term Memory (LSTM) network revolves around its core component, the LSTM cell, which is designed to manage long-term dependencies in sequential data. Each LSTM layer is composed of multiple such cells. The operation within an LSTM cell can be understood through its three primary gates and the cell state. It consists of four layers that interact with one another in a way to produce the output of that cell along with the cell state. These two things are then passed onto the next hidden layer.The LSTM (Long Short-Term Memory) is Kind of RNN (Recurrent Neural Network) that uses gates to capture both short-term as well as long-term memory. LSTMs are designed to process and retain data over several steps. They are widely used in deep learning and are ideal for sequence prediction tasks. Three gates like forget, input and output gate controls memory cell in LSTM architectures. These gates determine adding, removal and output data from memory cell. The input gate manages data which is added to memory cell, and forgets gate regulates data which is removed from memory cell.

The memory cell outputs regulate the data from output gate. The selected long-term dependencies maintaining or removing data as it moves through network enables LSTM network. The LSTM in a hidden state maintains short-term memory network. The hidden state is updated by using input, earlier hidden state, and current state of memory cell.

Summarization is process of condensing a longer passage into a shorter, more concise version that captures the most important points. It can be done for a different purposes, such as in book reports, legal arguments, or to help someone decide on an investment. The process of reducing and clearly expressing the significant facts or ideas about subject or person, or text that contains these facts or ideas is automatic text summarization.

RESULT ANALYSIS

In this section, analysis for natural language process based LSTM framework for keyword extraction is observed.

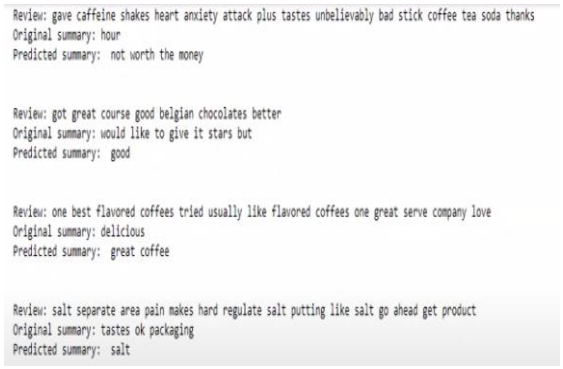

Figure 2: Summary Extraction Based on Reviews using Text Summarization Model

The above figure shows 4 different reviews with their summary in figure 2. The summary is extracted from reviews of review, original summary and predicted summary. The original user review text is unstructured and lengthy, but original summary is a truth summary by a human and the predicted summary is generated by the NLP model. Therefore, the proposed will summarizes the user reviews.

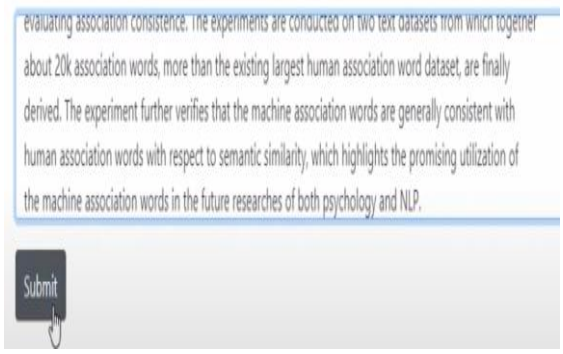

Figure 3: Word Extraction using NLP

By giving a paragraph as input and after submitting it, the word is extracted using NLP in figure 3. By using NLP, specific words are identified and extracted. The figure below shows a simple user interface for an NLP (Natural Language Processing) system, and it uses a large text box where a user can enter or paste some text. This type of tool is used to test or analyze keyword words.

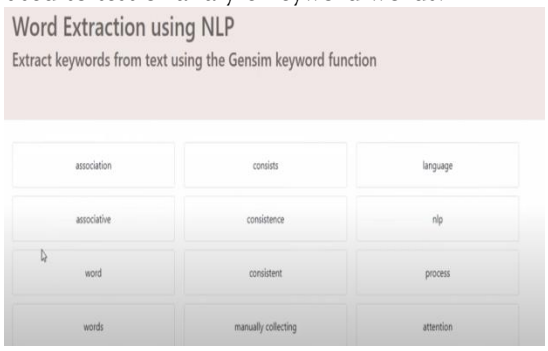

Figure 4: Keyword Extraction using NLP with Gensim

In figure 4, keywords are extracted from the text using Gensim keyword function. The Gensim keyword function automatically extracts significant terms from a given text. This figure shows keyword extraction and reduces manual effort. Figure 5 shows the training and testing graph. The figure shows training and testing graph. The blue line represents the training loss, while the orange line represents the testing (or validation) loss. Both training and testing decrease over time. The optimal number of training epochs is determined.

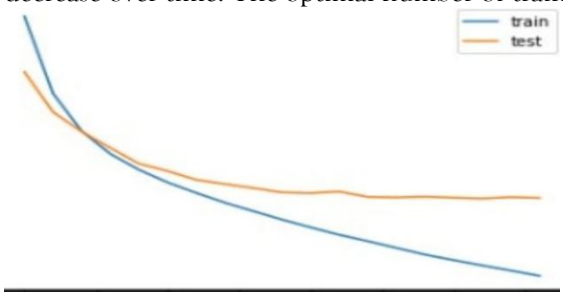

Figure 5: Training and Testing Graphs of Proposed Model

The performance analysis is observed in table.1 for natural language-process based LSTM framework for keyword extraction. In this comparision, LSTM (proposed system) is compared with BERT semantics and K-Truss graph(BSKT) [15] and Semantic Clustering TextRank (SCTR) [19]. The proposed system shows better performance in all parameters interms of accuracy, precision, recall and F1-Score.

Table.1: Performacne Comparison

Parameters	BSKT model [15]	SCTR [19]	LSTM (Proposed)
Accuracy	80	92	95
Precision	61.3	90	92
Recall	71	91	94
F1-Score	76	75	84

Comparative accuracy and precision values are represented in Figure 4 and Figure 5 for the natural language process for keyword extraction using LSTM and compared with the existing systems of BERT semantics and K-Truss graph(BSKT) [15] and Semantic Clustering TextRank (SCTR) [19]. From the graph, it is observed that the accuracy and precision of the proposed method is high when compared with other existing methods. In this graph, the X-axis describes methods and the Y-axis describes percentage.

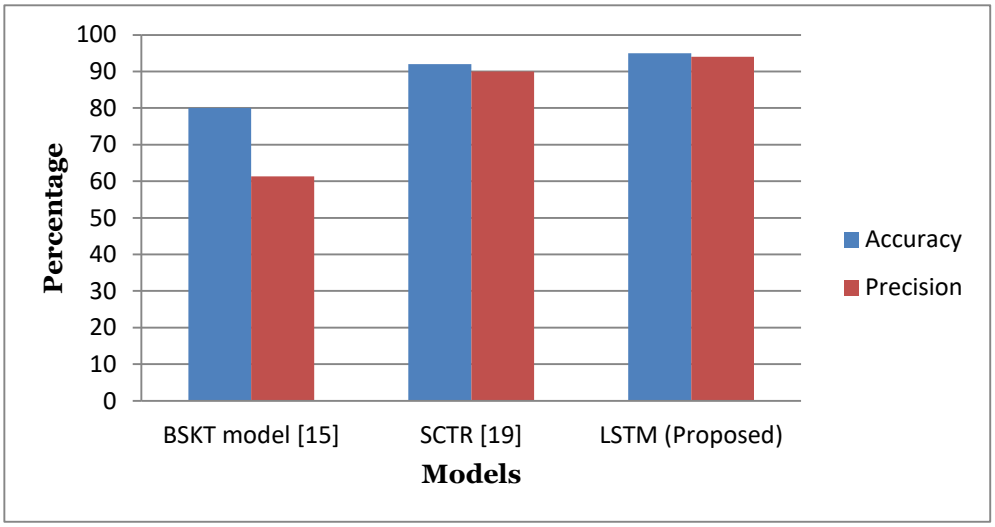

Figure 6: Accuracy and Precision Comparison Graph

Recall and F1-Score parameters comparative analysis is represented in Figure 6 and Figure 7 below for different existing models. In this comparative analysis representation, the X-axis presents methods and the Y-axis presents the percentage. It is observed that the recall and F1-score of the described model are high compared with BERT semantics and K-Truss graph(BSKT) [15] and Semantic Clustering TextRank (SCTR) [19].

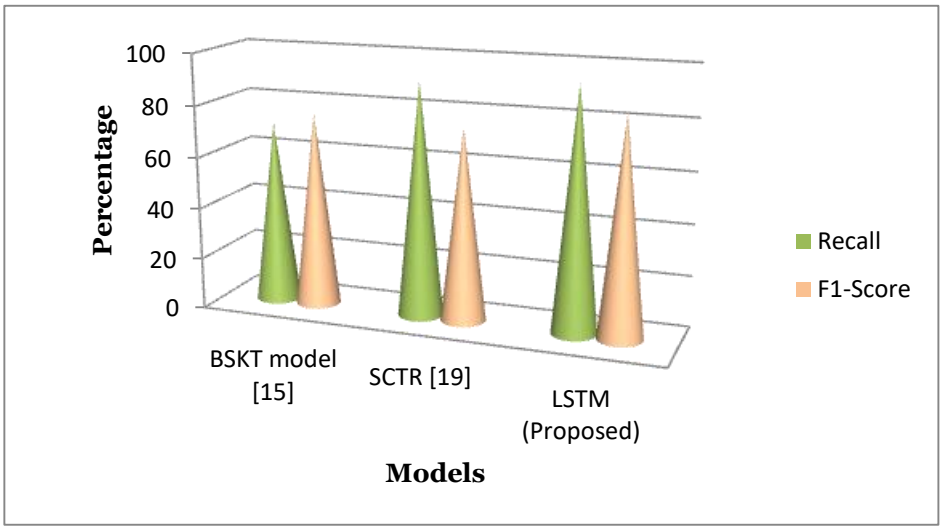

Figure 7: Recall and F1-Score Comparison Graph

CONCLUSION

Hence, NLP (Natural Language Process) based LSTM framework for keyword extraction is concluded in this section. The users can choose the website of their choice that scrapes data from websites. . This study demonstrates the proposed method by using the NIPS dataset for document classification and shows better accuracy in identifying keywords. It provides a summary and keywords from data that extracts from multiple websites by suggested system. The suggested system starts with data extraction from a website link for summarized text and extraction of keywords by number of stages that include removing outliers and data which is not relevant, demonstrating significance of specific data taken from website, and produces summary of extracted data. The relevant data is selected from extracted data by NLP. However, the model's performance may vary with imbalanced class distributions. In this system, keyword extraction is integrated with LSTM for summarization. This helps in capturing long-term dependencies in text and provide more accurate summary. In future, this system will integrate with advanced deep learning models, to improve classification accuracy and scalability across different research domains.

REFRENCES

[1] B. Armouty and S. Tedmori, "Automated Keyword Extraction using Support Vector Machine from Arabic News Documents," 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 2019, pp. 342-346, doi: 10.1109/JEEIT.2019.8717420.

[2] A. Payak, S. Rai, K. Shrivastava and R. Gulwani, "Automatic Text Summarization and Keyword Extraction using Natural Language Processing," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 98-103, doi: 10.1109/ICESC48915.2020.9155852.

[3] M. Zhang, X. Li, S. Yue and L. Yang, "An Empirical Study of TextRank for Keyword Extraction," in IEEE Access, vol. 8, pp. 178849-178858, 2020, doi: 10.1109/ACCESS.2020.3027567.

[4] M. Nadim, D. Akopian and A. Matamoros, "A Comparative Assessment of Unsupervised Keyword Extraction Tools," in IEEE Access, vol. 11, pp. 144778-144798, 2023, doi: 10.1109/ACCESS.2023.3344032.

[5] W. Fu and S. Akbar, "Expert Profile Identification From Community Detection on Author-Publication-Keyword Graph With Keyword Extraction," in IEEE Access, vol. 12, pp. 27918-27930, 2024, doi: 10.1109/ACCESS.2024.3368003.

[6] Y. Wei and Y. Ding, "Application of Text Rank Algorithm Fused With LDA in Information Extraction Model," in IEEE Access, vol. 11, pp. 84301-84312, 2023, doi: 10.1109/ACCESS.2023.3296141.

[7] Z. Jiang, C. Miao and X. Li, "Application of keyword extraction on MOOC resources," in International Journal of Crowd Science, vol. 1, no. 1, pp. 48-70, March 2017, doi: 10.1108/IJCS-12-2016-0003.

[8] R. Harakawa and M. Iwahashi, "Ranking of Importance Measures of Tweet Communities: Application to Keyword Extraction From COVID-19 Tweets in Japan," in IEEE Transactions on Computational Social Systems, vol. 8, no. 4, pp. 1030-1041, Aug. 2021, doi: 10.1109/TCSS.2021.3063820.

[9] R. Devika, S. Vairavasundaram, C. S. J. Mahenthar, V. Varadarajan and K. Kotecha, "A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data," in IEEE Access, vol. 9, pp. 165252-165261, 2021, doi: 10.1109/ACCESS.2021.3133651.

[10] D. P. Joseph and P. Viswanathan, "SDOT: Secure Hash, Semantic Keyword Extraction, and Dynamic Operator Pattern-Based Three-Tier Forensic Classification Framework," in IEEE Access, vol. 11, pp. 3291-3306, 2023, doi: 10.1109/ACCESS.2023.3234434.

[11] A. Gupta, A. Chadha and V. Tewari, "A Natural Language Processing Model on BERT and YAKE Technique for Keyword Extraction on Sustainability Reports," in IEEE Access, vol. 12, pp. 7942-7951, 2024, doi: 10.1109/ACCESS.2024.3352742.

[12] A. E. Blanchard et al., "A Keyword-Enhanced Approach to Handle Class Imbalance in Clinical Text Classification," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 6, pp. 2796-2803, June 2022, doi: 10.1109/JBHI.2022.3141976.

[13] Y. Tao, Z. Cui and Z. Jiazhe, "Research on Keyword Extraction Algorithm Using PMI and TextRank," 2019 IEEE 2nd International Conference on Information and Computer Technologies (ICICT), Kahului, HI, USA, 2019, pp. 5-9, doi: 10.1109/INFOCT.2019.8711099.
[14] G. S. K. Kurniawan and K. M. Lhaksmana, "Keyword Extraction from Scientific Publications Using Local Features and Embedding Model," 2023 9th International Conference on Signal Processing and Intelligent Systems (ICSPIS), Bali, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICSPIS59665.2023.10402723.
[15] W. Guo, Z. Wang and F. Han, "Multifeature Fusion Keyword Extraction Algorithm Based on TextRank," in IEEE Access, vol. 10, pp. 71805-71813, 2022, doi: 10.1109/ACCESS.2022.3188861.
[16] P. Vitolo, R. Liguori, L. D. Benedetto, A. Rubino and G. D. Licciardo, "Automatic Audio Feature Extraction for Keyword Spotting," in IEEE Signal Processing Letters, vol. 31, pp. 161-165, 2024, doi: 10.1109/LSP.2023.3346280.
[17] Shen.L, Li.R, Mao.X and Huang.S, "Automatic Keywords Extraction Based on Co-Occurrence and Semantic Relationships Between Words," in IEEE Access, vol. 8, pp. 117528-117538, 2020, doi: 10.1109/ACCESS.2020.3004628
[18] G. -W. Kim, W. -H. Kim, K. Chung and J. -C. Kim, "Extraction of Meta-Data for Recommendation Using Keyword Mapping," in IEEE Access, vol. 12, pp. 103647-103659, 2024, doi: 10.1109/ACCESS.2024.3430375.
[19] Kadoch.M, Xiong.A, Yu.P, Tian.H, Liu.Z, and Liu.D, "News keyword extraction algorithm based on semantic clustering and word graph model," in Tsinghua Science and Technology, vol. 26, no. 6, pp. 886-893, Dec. 2021, doi: 10.26599/TST.2020.9010051.
[20] He.D, Chen.B, Pu.L and Lin.C, "User-Friendly Public-Key Authenticated Encryption With Keyword Search for Industrial Internet of Things," in IEEE Internet of Things Journal, vol. 10, no. 15, pp. 13544-13555, 1 Aug.1, 2023, doi: 10.1109/JIOT.2023.3262660.
[21] Liu.H, Tang.J Yu.H, and Yang.Z, "Toward Keyword Extraction in Constrained Information Retrieval in Vehicle Social Network," in IEEE Transactions on Vehicular Technology, vol. 68, no. 5, pp. 4285-4294, May 2019, doi: 10.1109/TVT.2019.2906799.
[22] Z. -Z. Hu, J. –R Lin, L. -M. Chen, and J. -L. Li, "Understanding On-Site Inspection of Construction Projects Based on Keyword Extraction and Topic Modeling," in IEEE Access, vol. 8, pp. 198503-198517, 2020, doi: 10.1109/ACCESS.2020.3035214.
[23] S. Chang, G. -J. Ahn and S. Park, "Improving Performance of Neural IR Models by Using a Keyword-Extraction-Based Weak-Supervision Method," in IEEE Access, vol. 12, pp. 46851-46863, 2024, doi: 10.1109/ACCESS.2024.3382190
[24] J. Duan, X. Liao, Y. An and J. Wang, "KeyEE: Enhancing Low-Resource Generative Event Extraction with Auxiliary Keyword Sub-Prompt," in Big Data Mining and Analytics, vol. 7, no. 2, pp. 547-560, June 2024, doi: 10.26599/BDMA.2023.9020036.
[25] A. Amin et al., "TOP-Rank: A Novel Unsupervised Approach for Topic Prediction Using Keyphrase Extraction for Urdu Documents," in IEEE Access, vol. 8, pp. 212675-212686, 2020, doi: 10.1109/ACCESS.2020.3039548.