iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** II **Month of publication:** February 2025

DOI: https://doi.org/10.22214/ijraset.2025.66990

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

Text Mining and Latent Topic Discovery in TED Talks

Ms. Unnati Patel[1], Mrs. Kavita Rana[2]
[1]*Student, Department of MSc.IT, Nagindas Khandwala College, Mumbai, Maharashtra, India*
[2]*Assistant Professor, Department of Computer, and Information Science, Nagindas Khandwala College, Mumbai, Maharashtra*

Abstract: ***The TED Talks contain a vast amount of knowledge in different fields; this implies that making it useful to learn and share ideas gives an edge. With thousands of the talks at hand, it is challenging to organize them properly and suggest meaningful content. The technique of text mining and latent topic discovery draws on methods applied to the discovery of latent thematic structures from the transcripts of TED Talks by way of NLP through techniques such as LDA and NMF. The developed method could automatically identify top terms, group related talks together, and organize the content in such a way to better enhance the recommendation systems of speaker discourse. This work supports content discovery personalization and helps in solving the problems identified above under biased content propagation. The outcome enhances structured exploration for TED Talks and reaches better user engagement and facilitates more efficient knowledge dissemination.***
Keywords: LDA, LSTM, Machine learning, Text Mining, Topic Madling, deep learning

I. INTRODUCTION

TED Talks encompass a humongous repository of lectures related to science, technology, education, psychology, and business. The large repository over a thousand talk makes the grouping and retrieving of similar content extremely challenging. Techniques for text mining, especially Latent Topic Discovery in LDA and NMF, offer robust capabilities to uncover the underlying structures or themes in TED Talks. Analysis in the textual content based on transcripts would help in understanding the dominant topics spoken about by speakers as well as the pattern within discussion among speakers. This enhances recommendation systems That is where the techniques of text mining come to the fore by unearthing meaningful insights out of an immense amount of unstructured text data. In this work, we have applied Exploratory Data Analysis (EDA) to understand the structure and distribution of TED Talk content, which includes trends, word frequencies, and relationships between key terms. EDA will provide an initial understanding of the dataset, thereby revealing patterns for further analysis. We use a very well-known topic modeling technique called Latent Dirichlet Allocation (LDA) to find the hidden themes. Using LDA automatically groups TED Talks into different topics, based on distribution of words which helps us to find out the underlying theme without the need for any manual labeling. This improves the organizational structure of content, enhances search systems and recommendation systems, and also provides insight into the evolving interest of the audiences of TED Talks. The union of EDA and LDA would give an improved insight into TED Talks so that people can access relevant information, with the same contributing towards the larger goal of text mining and topic discovery.

II. PROPOSED SYSTEM

TED Talks is a global platform for the spreading of ideas across different topics, which creates intellectual discussions and the spreading of knowledge. However, because of the large number of speakers and subjects covered, there is a potential for biased content propagation and subtle hate speech, which could be detrimental to users' perspectives and engagement. Even though TED Talks are curated and moderated by normal standards, implicit biases, controversy-laden discussions, and hateful rhetoric easily find their way through transcripts. Hence, it requires the introduction of an automated hate speech detection system. It will ensure through machine learning (ML) and deep learning techniques that recommendations for TED Talks are free from biases, informative, and safe for the user. Following the imposition of topic modeling on the text data of the TED Talks necessitates the pre-processing of the text. The pipeline involved in pre-processing involves the following steps: First, converting the text into lower case so that it presents the same word consistently and therefore does not treat words as two different words just because they might be capitalized differently. Then, all punctuation marks, special characters, and numbers are removed from the text because these are not useful in identifying a topic. Stop words are the general words like "the," "and," and "is." Removing them is also an important task because it will help concentrate the search on proper terms. Tokenization splits the text into words or tokens after removing

©IJRASET: All Rights are Reserved

unwanted elements. Lemmatizing is the process through which words are reduced to their base form. This ensures that words like "running" and "run" get treated as the same.

It will create bigrams and trigrams. Bigrams represent the pairs of words; trigrams represent triplets. This will make way for meaningful phrases that might give more context to the topics. The text will then be converted into a numerical form by techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) or Bag of Words (BoW) to prepare it further for modeling in LDA. This step further refines the TED Talks dataset by eliminating noise and thus efficiency and accuracy of topic modeling. It is a structured approach that allows meaningful topics to be extracted using LDA for proper analysis and, therefore, the best possible recommendation.

III. LITERATURE REVIEW

Bailke et al. (2024) suggest a TED Talks Recommendation System to improve personal user experiences via seven algorithms, including K-means clustering, matrix factorization, nearest neighbors, restricted Boltzmann machines, locality-sensitive hashing (LSH), and association rules. K-means is best in pattern detection, matrix factorization provides better performance, and nearest neighbors improve responsiveness. Restricted Boltzmann machines capture intricate behavior, association rules are interpretable, and LSH enhances scalability. The research emphasizes how future recommendation systems can be able to enhance user engagement and satisfaction with TED Talks.

Boyd-Graber et al. (2017) explain how topic models can help users understand large sets of documents, such as emails, newspapers, or scientific articles. They give an overview of recent academic and industrial applications of topic models to help novice researchers build their own implementations. The study introduces the application of topic models in information retrieval, visualization, statistical inference, multilingual modeling, and linguistic analysis tasks. It also discusses the extent to which they are useful in opening ideas from mass texts collections, favoring fiction, non-fiction, scientific papers, and political texts analysis.

C Boussalis et al. (2016) This research presents a novel automatic video tag recommendation approach by leveraging unsupervised key phrase extraction and word embedding models to enhance tag coverage for video transcripts. By considering semantic similarity, the proposed method ensures more relevant and comprehensive tagging, improving video discoverability on platforms like YouTube and Twitch. Experiments on TED Talk datasets show significant improvements over existing approaches. This work is useful in organizing video content and thus helps the creators and viewers navigate large video repositories more effectively.

Cheng et al. (2018) tackle shortcoming of latent factor models, including cold-start and non-transparency, through the addition of textual review information to ratings. They suggest an aspect-aware topic model (ATM) for extracting user preference and item properties in varying aspects. Such findings are fed into an aspect-aware latent factor model (ALFM), utilizing weighted matrices for inferring aspect ratings and determining total ratings. This method improves data sparsity management and interpretability. Experiments on Amazon and Yelp datasets show substantial improvements compared to baseline methods, particularly for sparse users, and offer more profound recommendation insights.

Hannah Kim et al. (2020) present ArchiText: an interactive hierarchical topic modeling system emphasizing tight integration between computation, visualization, and user interaction. In contrast to the human-in-the-loop approach, ArchiText co-develops interactive algorithms and visual analytics to enable users to flexibly steer topic modeling processes. It does this by proposing computational base operations for interactive tasks that improve scalability, adaptability, and interpretability of hierarchical topic structures. It ensures better control and responsiveness to users, because this is indeed the best approach for dynamic text exploration and analysis.

Hong et al. (2011) propose a time-dependent topic model to analyze multiple text streams, such as Twitter and Yahoo! News, by extending traditional topics models in order to include both local and shared topics between platforms. Their approach integrates temporal dynamics, an assignment of a time-dependent function to each topic in order to track its popularity at different points in time, helping to understand how news spreads and evolves over media sources. The model results in enhanced topic coherence and delivers superior perplexity scores on the unseen data set than other prior methods. For research into news dissemination, social media analysis, and cross-platform information flow, this study brings forth real-world insights as well, because significant events could be determined from extracted topics.

Jo Guild et al. (2023) In this paper we seen how text mining redefines historical research. The book depicts pioneering collaborations between humanists and data scientists and shows how analysis of textual data can dig out forgotten narratives and political suppression. However, she cautions the reader that some conclusions will lead to misleading use of data science without proper history. With rich examples galore, she pleads for ethics and informed text mining toward democratic enhancement and more understanding of history.

Karami et al. (2024) examine workplace sexism and sexual harassment through content analysis of 2,362 experiences posted on everydaysexism.com. With quantitative and qualitative analysis, the study used text mining to find and classify 23 topics into three broad themes. The Sex Discrimination theme contained adverse treatment on the basis of sex, including rejected promotions and differences in pay. The Sex Discrimination and Gender Harassment theme involved sexist hostility in the form of insults, jokes, and bullying. Finally, the Unwanted Sexual Attention theme raised actions like groping and inappropriate touching, and this was the most frequent subject. The study shows the use of automated text analysis to extract extensive, genuine workplace sexism data at large.

Maiya & Rolfe (2014) propose a new technique for the better interpretability of LDA topic models by the means of topic similarity networks, where topics are represented as nodes and their relationships as links. They provide efficient methods for constructing and labeling these networks, thereby improving the visualization and analysis of topic structures. By applying their method to NSF grant data and Wikipedia, they illustrate that these networks also uncover latent connections as well as probe deeper into thematic structures of massive text collections. This is really a very strong tool to study unstructured text data.

Martínez et al. (2015) offer a survey of recent empirical and theoretical work on information extraction techniques in pharmacovigilance. The articles address several subjects, such as supervised machine learning methods for named entity recognition and relation extraction, especially for the identification of adverse drug reactions and drug-drug interactions. They also address natural language processing architectures, corpora, semantic resources, and other tasks such as sentiment analysis. The authors point out the richness of the domain and stress upcoming problems, including the inclusion of more data such as socioeconomic status and dosages, developing methodologies to be used with various text styles, filling ontologies, and producing terminological resources adapted to both patients and healthcare professionals.

Pan, Yan, and Hua (2016) investigate the cross-national influence of TED Talks videos on YouKu and YouTube based on six metrics, which include views, likes, dislikes, comments, bookmarks, and shares. The research also identifies massive differences in topic popularity. Entertainment and psychology/philosophy subjects are given more attention, whereas design/art and science subjects such as astronomy and biology generate fewer comments and bookmarks. Platform preferences also differ, with technology and global issues trending more on YouKu and entertainment and psychology/philosophy trending more on YouTube. Women are also more interested in education and psychology-related content, while men are more interested in technology and science-related content.

Peng & Shang, (2024) In this work, the "LLM-Powered Text Classification Incubator" generates text classification data directly from user instructions without the need for human annotation or raw corpora. Unlike previous methods, it is applicable to challenging and interdependent class definitions. The system exploits instruction-tuned LLMs, in-context augmentation with GPT-4, and semantic embedding clustering to enhance the accuracy of classification, label dependency handling, and logical text mining. The experimental results show that it outperforms traditional LLM-based inference and prompt-engineered data generation, so it is a strong tool for automatic text classification.

V Kumar and K Thakur (2021) refer to the phenomenal growth of applications of text mining to derive insight from the growing body of scholarship literature. Key trends, including the extensive employment of LDA and R by text mining are shown through a systematic review they conduct, as researchers tend to analyze sample populations of approximately 1,000 articles. The research further observes that ICT-related papers are often made subject to scrutiny, and the most widely employed document sections for this purpose are abstracts. These observations further emphasize the relevance of effective text mining across multiple research fields.

Victor Corja et al. (2022) propose a Unified NLP Model that integrates speech recognition, keyword extraction, topic mining, and semantic analysis into a single framework. The model processes both audio and text data by combining multiple NLP methodologies, offering a comprehensive analysis pipeline. Tested on TED Talks datasets, the system effectively extracts structured insights, enhancing data mining capabilities. This integrated approach increases the efficiency and accuracy of multimodal content analysis; hence, it is beneficial for research as well as practical applications in speech and text processing.

Z Cheng et al. (2018) proposed the Aspect-Aware Latent Factor Model (ALFM) for improving accuracy and interpretability of recommendation while combining textual review information with rating. Their proposed model uses an Aspect-Aware Topic Model (ATM) to extract user preference and item feature, which overcomes cold-start and data sparsity problems. The latent factors associated with discovered aspects using a weighted matrix improve personalized recommendations in ALFM. It has been extensively evaluated on Amazon and Yelp datasets, significantly demonstrating its superior performance, especially for the users with limited ratings. This approach provides a more nuanced, aspect-driven recommendation framework and with enhanced explain ability.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

IV. METHODOLGY

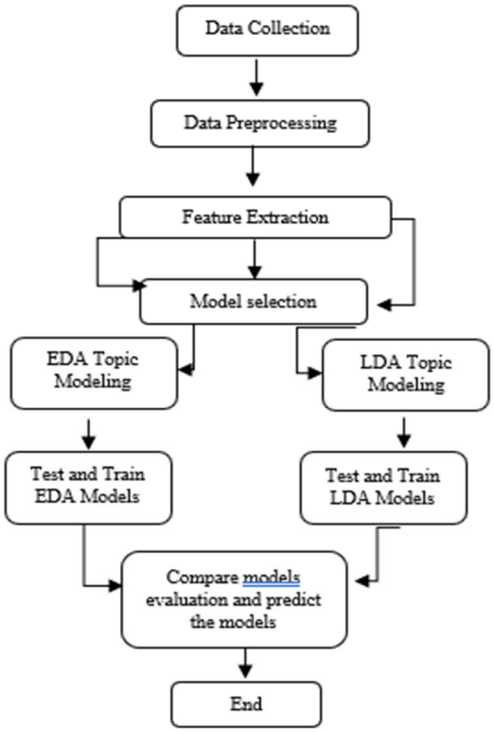


Fig 1. Methodology Model

A. Data Collection

The initial step is to collect data in order to produce a high-quality, representative data set upon which to study TED Talks and provide recommendations. The primary data sets are the TED main and transcript data sets, including titles, descriptions, views, speaker data, publication dates, and full speech transcripts. These data sets have rich opportunity for richer analysis of user interest, speaker intent, and topic trends. To add to the volume and variety of our dataset, we will combine as many publicly available TED-related datasets as possible. Data annotation, for instance, marking lecture topics or annotating user behavioral patterns, is used in order to facilitate supervised learning while enhancing the accuracy of the recommendation system.

B. Data Pre-processing

Data pre-processing is the act of taking raw, unstructured text and converting it into clean, structured data that we can use for models. Because TED Talks data is so many different shapes and sizes, we have to fix things like acronyms, slang, and different formats. Pre-processing includes such things as tokenization, breaking the text up into words, and stop words removal—words that aren't important—unless they are field-specific. These processes get the data more standardized, eliminate the useless, and make us more capable of analyzing it and recommending.

C. Feature Extraction

Feature extraction is necessary to transform raw TED Talks data into analysis- and recommendation-appropriate forms. Text features are extracted from the titles, descriptions, and transcripts of talks to identify useful trends and patterns. Common practices include: Word Frequency Analysis: Ensures how frequently words are employed to identify important words and trends in the talks. Topic Detection: Applies topic modeling techniques like Linear discriminant analysis (LDA) to categorize talks under various topics based on related words. Sentiment Analysis: Searches for the emotional tone of the text (e.g., positive, neutral, or negative) to realize audience reactions and preferences. TF-IDF (Term Frequency-Inverse Document Frequency): Assigns weights to words based on how frequently they appear in numerous talks to identify key words. The extracted features enable the system to identify user preferences, provide personalized recommendations, and enhance the overall recommendation process for TED Talk users.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Modeling

1) Exploratory Data Analysis (EDA) for Topic Modeling: This approach detects broad trends and patterns in TED Talks, so popular topics can be easily identified. Visualizing trends enable the system to identify what users are looking for and to enhance personalized recommendations. It enables the system to identify less popular topics, which can represent new or emerging user interests.
2) Linear discriminant analysis (LDA) Topic Modeling: This approach structures the text data into explicit topics, classifying TED Talks in various themes. LDA plays a central role in linking user preferences to topics, enabling the system to make better recommendations. LDA can identify overlapping topics, so talks that are in more than one theme can be recommended to users interested in various subjects.

E. Training & Evaluation:

The TED Talks recommendation system is trained on 80% of the data, and 20% is kept aside for testing purposes to guarantee generalization. LDA models pick topics, and EDA picks trends, while sentiment analysis labels transcripts as positive, neutral, or negative. Hyperparameters are optimized for the best performance. Evaluation criteria are perplexity for topic modeling, coherence for topic coherence, and precision, recall, accuracy, and F1-score for sentiment classification. Simulated user interactions verify the system's capacity to deliver relevant and interesting recommendations.

V. RESULTS

The radar chart that indicates the range of feelings over various years for TED Talks, capturing emotions like happiness, trust, surprise, and fear. The radar illustrates the persistence or evolution of certain emotions. The spikes for specific emotions refer to the great trends or issues that might have influenced the tone of TED Talks during these years. It helps establish how emotional reactions and audience engagement have evolved, offering greater insight into the evolving themes and emotions across the TED Talks corpus.



Fig 2: Sentiment Analysis Over Time

This chart provides the most relevant words for every topic, as "Economic Growth," "Stories & Ideas," and "Wealth & Inequality" indicate. The longer the bar for a given word, the more relevant to that topic the word is. For instance, in "Economic Growth" highly relevant are words like "market," "system," and "GDP.". Looking at these patterns makes us better see the dominant themes in the data and how concepts are related to each other. This assists with structuring and making sense of large volumes of text content.

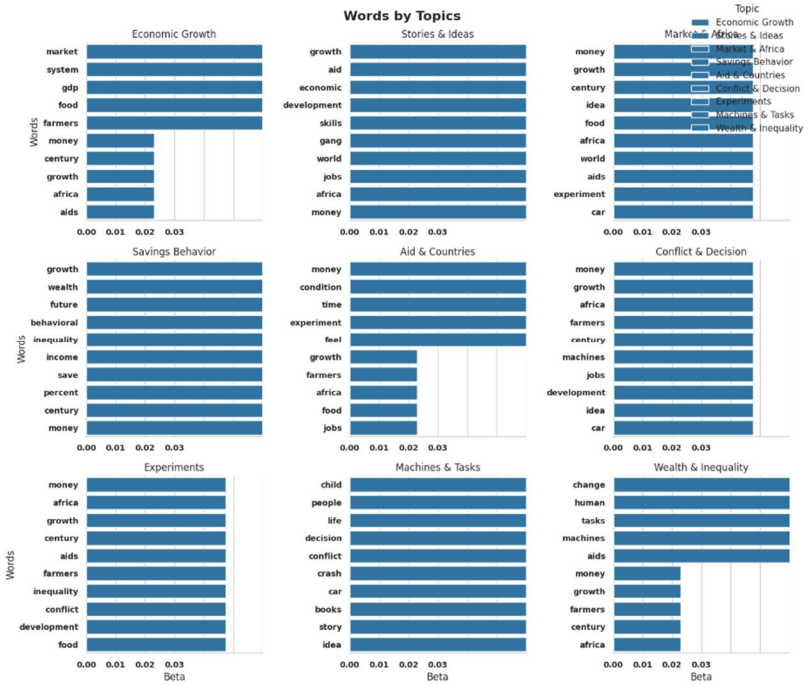


Fig 3: Topic-Wise Word Distribution

TED Talks, by the emotional tone of each speech. Visualization title, "Sentiment Scores for Most Watched TED Talks by Year (Scaled)," clarifies that sentiment scores are scaled to be comparable. Title of every speech, duration, and the sentiment score are denoted with a rectangle block for every speech. The sentiment rating captures the aggregate optimism or optimism of the word choice used within the speech, with greater numbers indicating a more optimistic tone. Color coding visually represents levels of sentiment, from red (little sentiment) through to green (high sentiment), with mid-score ratings represented by orange and yellow. A bottom heatmap legend denotes color mapping so that it may be easily comprehended. For example, "How great leaders inspire action" has a sentiment score of 98.00, reflecting its content as extremely motivational in tone. Sad topics may be spoken about by low-scoring speeches. The visualization above is helpful in different practical applications such as recommending talks based on sentiment interests, identifying evolving audience interaction behavior, and monitoring the emotional appeal of the most popular TED Talks.

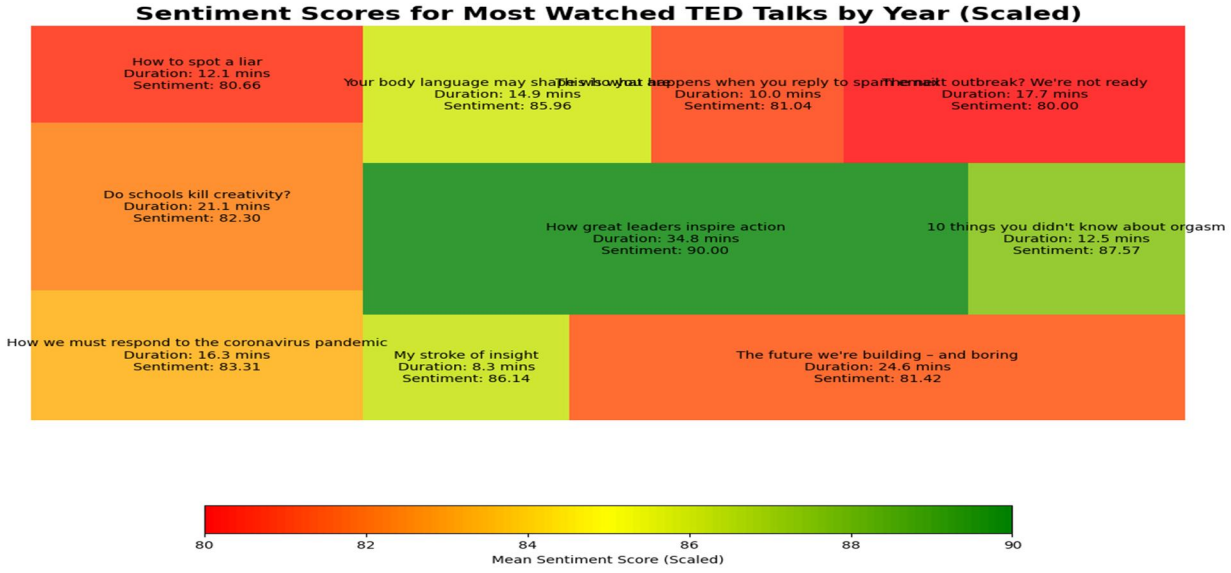


Fig 4: Sentiment Heatmap of Popular TED Talks

The confusion matrix indicates how accurately the optimized Linear Discriminant Analysis (LDA) model predicted the three iris species: setosa, versicolor, and virginica. The 10 samples for each species were all correctly predicted as evident in the numbers along the diagonal. The model, therefore, was 100% accurate in its predictions, getting all the test samples correct. That the cross-validation accuracy was 98% also indicates that the model is quite good across other data sets, proving its usability for this classification problem.

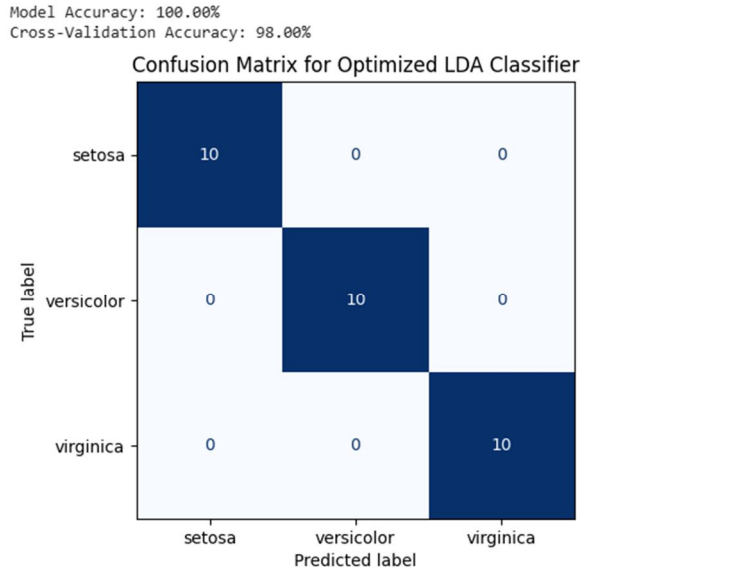


Fig 5: LDA Model Confusion Matrix

The prediction of a text classification model predicting whether a comment is offensive or not. Four comments are tried out, and the model predicts each of them as such. For example, the comment "The TED Talk on AI and future technologies was fascinating" is labeled as neutral ("None"), while comments like "I can't believe the nonsense these people are saying!" are labeled as offensive. The model can correctly label offensive content and neutral statements, which reflects its ability to discriminate between the two classes. Such discrimination can be employed in moderating forums through the screening out of offensive posts.

```
Text: 'The TED Talk on AI and future technologies was fascinating.' --> Predicted as: None
Text: 'I can't believe the nonsense these people are saying!' --> Predicted as: offensive
Text: 'A beautiful discussion about peace and love.' --> Predicted as: None
Text: 'Horrible! They ruined everything with their politics.' --> Predicted as: offensive
```

Fig 6: Model Prediction

VI. CONCLUSION

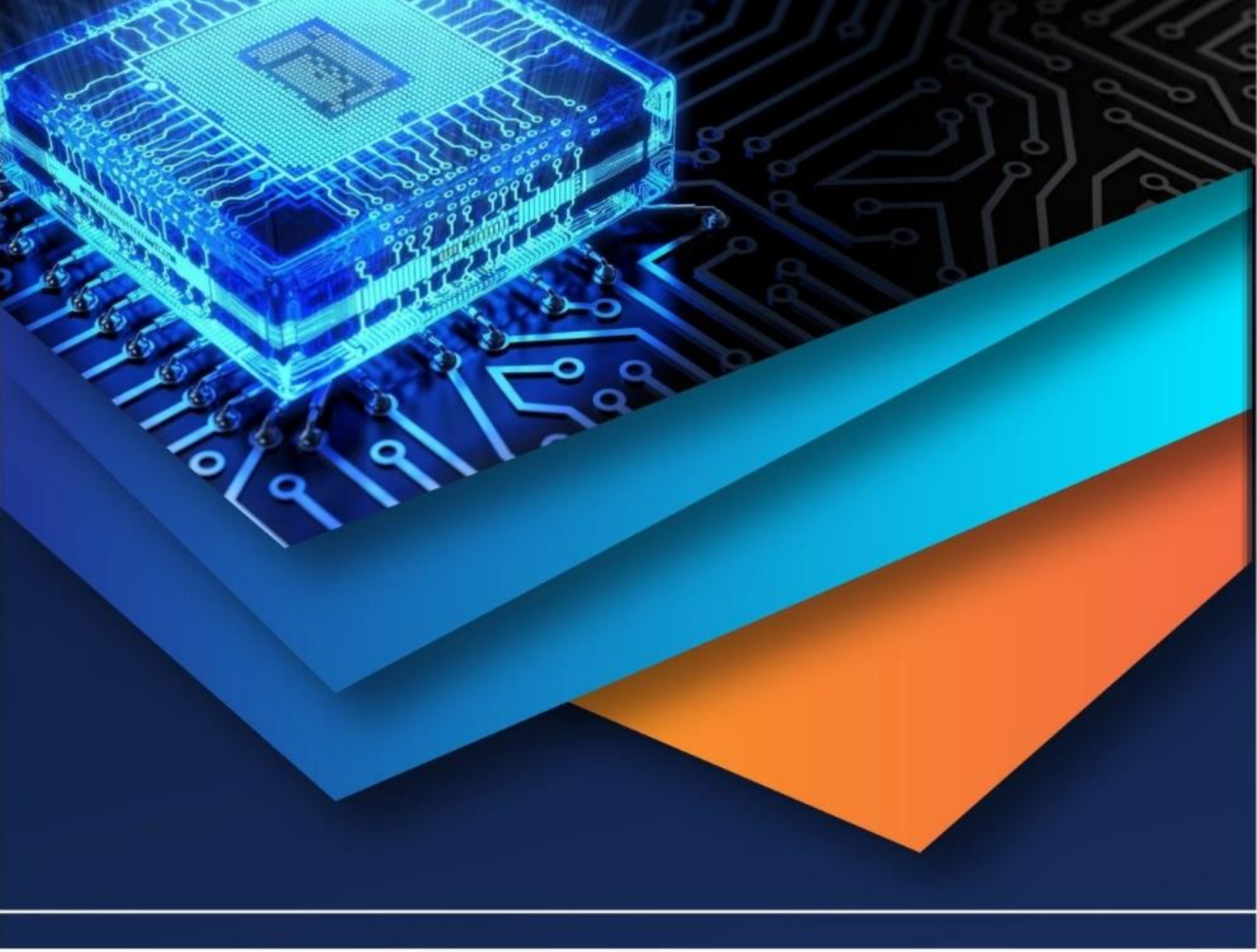
In this research, we developed and trained machine learning models to examine and predict various ted talk features, such as sentiment analysis and insult detection. We employed the sentiment analysis model to assess the emotional sentiment of TED Talks and how positive and negative words impact audience engagement. In addition, we developed a classification model to identify offensive comments about ted talk topics, which can help improve content moderation and create more engaging online interactions. We also optimized a linear discriminant analysis (lda) model to categorize various ted talk themes, and we were able to obtain high accuracy and cross-validation scores. These models yielded useful insights about ted talk content, activity of engagement, and interaction of the audience, illustrating the potential of machine learning in analyzing content and recommendation systems.

REFERNCES

[1] Adibi, J., Pantel, P., Grobelnik, M., & Mladenic, D. (2005). KDD-2005 workshop report: Link Discovery: issues, approaches and application (LinkKDD-2005). ACM SIGKDD Explorations Newsletter, 7(2), 123-125.

[2] Bailke, P., Asalkar, A., Bansode, P., Baviskar, N., Belote, A., & Nadar, H. (2024, October). Ted-Talks Recommendation System using ML Algorithms. In 2024 First International Conference on Software, Systems and Information Technology (SSITCON) (pp. 1-6). IEEE.

[3] Bernstein, M. S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E. H. (2010, October). Eddi: interactive topic-based browsing of social status streams. In Proceedings of the 23nd annual ACM symposium on User interface software and technology (pp. 303-312).
[4] Boussalis, C., & Coan, T. G. (2016). Text-mining the signals of climate change doubt. Global Environmental Change, 36, 89-100.
[5] Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. Foundations and Trends® in Information Retrieval, 11(2-3), 143-296.
[6] Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. Foundations and Trends® in Information Retrieval, 11(2-3), 143-296.
[7] Cheng, Z., Ding, Y., Zhu, L., & Kankanhalli, M. (2018, April). Aspect-aware latent factor model: Rating prediction with ratings and reviews. In Proceedings of the 2018 world wide web conference (pp. 639-648).\
[8] Cheng, Z., Ding, Y., Zhu, L., & Kankanhalli, M. (2018, April). Aspect-aware latent factor model: Rating prediction with ratings and reviews. In Proceedings of the 2018 world wide web conference (pp. 639-648).
[9] Ciotti, F. (2021). Distant reading in literary studies: a methodology in quest of theory. Testo e Senso, (23), 195-213.
[10] Corja, V., Crow, A. R., Liu, N., & Ceesay, E. (2022, January). Bimodal Key Data and Semantic Analysis. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0222-0228). IEEE.
[11] Deena, S., Ng, R. W., Madhyastha, P., Specia, L., & Hain, T. (2017, December). Exploring the use of acoustic embeddings in neural machine translation. In 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU) (pp. 450-457). IEEE.
[12] Fan, Q., Yu, Y., Yin, G., Wang, T., & Wang, H. (2017, November). Where is the road for issue reports classification based on text mining?. In 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 121-130). IEEE.
[13] Guldi, J. (2023). The dangerous art of text mining: A methodology for digital history. Cambridge University Press.
[14] Heiberger, R. H., & Galvez, S. M. N. (2021). Text mining and topic modeling. In Handbook of Computational Social Science, Volume 2 (pp. 352-365). Routledge.
[15] Heiberger, R. H., & Galvez, S. M. N. (2021). Text mining and topic modeling. In Handbook of Computational Social Science, Volume 2 (pp. 352-365). Routledge.
[16] Hong, L., Dom, B., Gurumurthy, S., & Tsioutsiouliklis, K. (2011, August). A time-dependent topic model for multiple text streams. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 832-840).
[17] Hu, C. C., & Li, H. L. (2017). Developing navigation graphs for TED talks. Computers in Human Behavior, 66, 26-41.
[18] Jelodar, H., Wang, Y., Rabbani, M., Ahmadi, S. B. B., Boukela, L., Zhao, R., & Larik, R. S. A. (2021). A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments. Multimedia Tools and Applications, 80, 4155-4181.
[19] Jelodar, H., Wang, Y., Rabbani, M., Ahmadi, S. B. B., Boukela, L., Zhao, R., & Larik, R. S. A. (2021). A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments. Multimedia Tools and Applications, 80, 4155-4181.
[20] Karami, A., Swan, S. C., White, C. N., & Ford, K. (2024). Hidden in plain sight for too long: Using text mining techniques to shine a light on workplace sexism and sexual harassment. Psychology of Violence, 14(1), 1.
[21] Kim, H., Drake, B., Endert, A., & Park, H. (2020). Architext: Interactive hierarchical topic modeling. IEEE transactions on visualization and computer graphics, 27(9), 3644-3655.
[22] Konkaew, T., & Kitisin, S. (2019). Automatic tag recommendation approach with keyphrase extraction and word embedding techniques. J. Comput, 30(2), 135-149.
[23] Liu, S., Zhou, M. X., Pan, S., Song, Y., Qian, W., Cai, W., & Lian, X. (2012). Tiara: Interactive, topic-based visual text summarization and analysis. ACM Transactions on Intelligent Systems and Technology (TIST), 3(2), 1-28.
[24] Maiya, A. S., & Rolfe, R. M. (2014, October). Topic similarity networks: visual analytics for large document sets. In 2014 IEEE International Conference on Big Data (Big Data) (pp. 364-372). IEEE.
[25] Ozbek, S. A. (2022). An Analysis of the Central Bank of the Republic of Türkiye (CBRT)'s Monetary Policy Communication With Text Mining (Master's thesis, TED University (Turkey)).
[26] Pan, X., Yan, E., & Hua, W. (2016). Science communication and dissemination in different cultures: An analysis of the audience for TED videos in C hina and abroad. Journal of the Association for Information Science and Technology, 67(6), 1473-1486.
[27] Pan, X., Yan, E., & Hua, W. (2016). Science communication and dissemination in different cultures: An analysis of the audience for TED videos in C hina and abroad. Journal of the Association for Information Science and Technology, 67(6), 1473-1486.
[28] Pappas, N., & Popescu-Belis, A. (2013, June). Combining content with user preferences for TED lecture recommendation. In 2013 11th International Workshop on Content-Based Multimedia Indexing (CBMI) (pp. 47-52). IEEE.
[29] Ralph, D. (2022). Insights from heterogeneous data through transitive semantic relationships and text analytics (Doctoral dissertation, University of Southampton).
[30] Segura-Bedmar, I., & Martínez, P. (2015). Pharmacovigilance through the development of text mining and natural language processing techniques. Journal of biomedical informatics, 58, 288-291.
[31] Shih, C. H., Li, H. L., Hu, C. C., & Lin, B. M. (2020). Forming a TED talks sphere for convenient search. The Electronic Library, 38(2), 403-420.
[32] Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. Discourse & Society, 27(4), 401-422.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)