International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 06 | June - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Topic modeling and Recommendation system using LDA And Cosine Similarity

Nayana J
Department of Machine Learning
B.M.S College Of Engineering
Bengaluru,India
nayana.ai21@bmsce.ac.in

Priyanka S
Department of Machine Learning
B.M.S College Of Engineering
Bengaluru,India
priyanka.ai21@bmsce.ac.in

Sowmya Lakshmi BS
Assistant Professor
Department of Machine Learning
B.M.S College Of Engineering
Bengaluru,India
sowmyalakshmibs.mel@bmsce.ac.in

Abstract

In this study, we explore the application of Natural Language Processing (NLP) techniques to analyze and recommend articles from Medium. By utilizing a dataset of Medium articles, we employ topic modeling using Latent Dirichlet Allocation (LDA) and implement a recommendation system using Doc2Vec to suggest articles similar to a given piece. The methodology includes preprocessing text data, visualizing insights, building an LDA model to discover latent topics, and creating a content-based recommendation system. Our findings show the potential of these techniques in enhancing user engagement and content discovery on digital platforms.

Index Terms- Content Recommendation, Latent Dirichlet Allocation, Natural Language Processing, Topic Modeling, User Engagement

I. INTRODUCTION

The digital age has brought an unprecedented surge in the amount of user-generated content available online. Platforms like Medium have become popular venues for individuals to share articles on a wide array of topics, from personal stories to in-depth analyses on complex issues. While the sheer volume of content enriches the platform, it also poses significant challenges in content management and user engagement. Efficiently categorizing and recommending articles becomes crucial in ensuring that users can discover relevant content without being overwhelmed by the vast amount of available information.

NLP offers powerful tools to address these challenges. By analyzing the textual content of articles, NLP techniques can uncover underlying themes and relationships, facilitating better content organization and personalized recommendations. In this study, we focus on two specific NLP methodologies: topic modeling and document embedding, applied to a dataset of Medium articles.

Topic modeling, and specifically LDA, is a method used to identify latent topics within a collection of documents. LDA works by assuming that each document is a mixture of a small number of topics and that each word in the document is

attributable to one of the document's topics. This probabilistic approach allows us to uncover hidden thematic structures in the text data, which can be useful for categorizing articles into meaningful clusters.

In addition to organizing content, enhancing user experience through personalized recommendations is another critical objective. Content-based recommendation systems leverage the text content of articles to suggest similar pieces to users, based on their reading history or a specific article they are currently interested in. The Doc2Vec model, an extension of Word2Vec, creates vector representations of entire documents, capturing semantic

meanings and contextual similarities. These document embeddings can then be used to find articles that are similar to a given piece, thereby providing personalized content recommendations.

This research aims to harness these NLP techniques to analyze and recommend articles from Medium. Our methodology involves several key steps: preprocessing the text data to prepare it for analysis, using LDA to model the topics within the articles, and employing Doc2Vec to create a recommendation system that suggests articles based on their content similarity. We begin by cleaning and tokenizing the text, removing stopwords, punctuation, and digits, and lemmatizing the words to reduce them to their base forms. This preprocessing ensures that the data is in a suitable format for both topic modeling and document embedding.

By applying LDA, we can extract meaningful topics from the articles, providing insights into the thematic distribution of content on Medium. The Doc2Vec model, on the other hand, enables us to generate recommendations, enhancing user engagement by suggesting articles that align with their interests.

In summary, this study demonstrates the application of advanced NLP techniques to improve content organization and user experience on digital platforms like Medium. Through topic modeling and content-based recommendation, we aim to make it easier for users to discover relevant articles and enhance their overall engagement with the platform.

II. DATASET

The dataset used in this study consists of articles from Medium, a popular online publishing platform. Each entry in the dataset includes Author,Claps,reading_time,link,title,text but the following key attributes are considered for analysis :

- Text: The full text of the article, which is the main focus for our NLP analysis.
- Author: The name of the author who wrote the article.
- Claps: The number of claps the article received, which serves as an indicator of its popularity and reader engagement.

This dataset provides a robust foundation for our topic modeling and recommendation system tasks. The text data allows us to delve into the thematic content of the articles, while the metadata (author and claps) helps in performing exploratory data analysis and understanding the distribution of content and popularity among different authors.

	author	claps	reading_time	link	title	text
0	Justin Lee	8.3K	11	https://medium.com/swlh/chatbots-were-the-next...	Chatbots were the next big thing: what happene...	Oh, how the headlines blared:\nChatbots were T...
1	Conor Dewey	1.4K	7	https://towardsdatascience.com/python-for-data...	Python for Data Science: 8 Concepts You May Ha...	If you've ever found yourself looking up the s...
2	William Koehrsen	2.8K	11	https://towardsdatascience.com/automated-featu...	Automated Feature Engineering in Python – Towa...	Machine learning is increasingly moving from h...
3	Gant Laborde	1.3K	7	https://medium.freecodecamp.org/machine-learni...	Machine Learning: how to go from Zero to Hero ...	If your understanding of A.I. and Machine Lear...
4	Emmanuel Ameisen	935	11	https://blog.insightdatascience.com/reinforcem...	Reinforcement Learning from scratch – Insight ...	Want to learn about applied Artificial Intelli...

Before applying NLP techniques, extensive preprocessing of the text data is necessary to ensure consistency and cleanliness. The preprocessing steps include:

- Lowercasing: Converting all text to lowercase to maintain uniformity.
- Stopwords Removal: Eliminating common English stopwords using NLTK's stopword list to focus on meaningful words.
- Punctuation and Digits Removal: Stripping all punctuation and digits to reduce noise in the text.
- Lemmatization: Reducing words to their base or root form using SpaCy to ensure that different forms of a word are treated as a single entity.

These preprocessing steps are crucial for preparing the data for effective analysis. Clean and well-prepared text data is essential for the success of both topic modeling and document embedding.

The preprocessed dataset thus serves as the basis for applying LDA for topic modeling and Doc2Vec for content-based recommendation. By analyzing the text data, we can uncover hidden topics within the articles, providing insights into the thematic distribution of content on Medium. Additionally, the metadata allows us to analyze trends and patterns related to author contributions and article popularity.

Overall, the dataset offers a rich source of information that supports the dual goals of uncovering latent topics in the articles and building a recommendation system to enhance user engagement on Medium. The combination of detailed text data and relevant metadata makes it possible to perform a comprehensive analysis that addresses both content organization and user personalization.

III. METHODOLOGY

Our methodology encompasses several critical stages, each vital to the process of topic modeling and the development of a recommendation system. The steps include data preprocessing, exploratory data analysis, detailed explanation of LDA, and creating a content-based recommendation system using Doc2Vec. Each stage is crucial for ensuring the accuracy and effectiveness of the final models.

3.1 Data Preprocessing

Data preprocessing is a fundamental step that ensures the text data is clean, consistent, and ready for analysis. This involves multiple steps:

- Lowercasing: All text is converted to lowercase to maintain uniformity and avoid issues related to case sensitivity.
- Stopwords Removal: Common English stopwords are removed using the Natural Language Toolkit (NLTK). Stopwords, such as "and", "the", and "in", frequently occur but do not carry significant meaning.
- Punctuation and Digits Removal: All punctuation and digits are stripped from the text. This step reduces noise and ensures that only meaningful text is analyzed.
- Lemmatization: Using SpaCy's lemmatizer, words are reduced to their base form. For instance, "running" and "ran" are converted to "run". This step ensures that different forms of a word are treated as a single entity.

These preprocessing steps are crucial for preparing the data for effective analysis and model building. Clean and well-prepared text data is essential for the success of both topic modeling and document embedding.

3.2 Exploratory Data Analysis (EDA)

EDA is performed to gain insights into the distribution and characteristics of the dataset. This involves analyzing various aspects of the data:

- Author Analysis: We identify the number of unique authors and analyze the distribution of articles among them. This helps in understanding the contribution of different authors to the dataset.
- Clap Analysis: We examine the distribution of claps to understand the popularity and engagement levels of articles. Claps serve as a proxy for reader engagement and content quality.

Visualizations, such as bar plots, are used to represent the number of articles per author and the total number of claps received by each author. These visualizations help in understanding the overall structure and trends within the dataset.

3.3 Latent Dirichlet Allocation (LDA)

LDA is a popular topic modeling algorithm widely used in NLP. Topic modeling is a technique used to identify latent topics within a collection of documents or texts. LDA is a probabilistic model that generates a set of topics, each represented by a distribution over words, for a given corpus of documents.

LDA aims to discover the underlying topics in the corpus and the corresponding proportions of each case in each document. LDA is an unsupervised learning technique that does not require labeled data and is helpful for tasks such as document classification, information retrieval, and recommender systems.

Working of LDA:

- Tokenization: The first step is to tokenize the text data, breaking it down into individual words or tokens. This step prepares the text for further processing.
- Dictionary Creation: LDA requires a dictionary where each unique word in the corpus is assigned a unique integer ID. This dictionary is used to map words to their numerical IDs.
- Bag-of-Words Representation: Each document is represented as a bag-of-words, which is a simple way of representing text data where the order of words is disregarded and only their frequencies matter. This representation is created using the dictionary and contains word IDs and their respective frequencies.
- Topic Modeling: The LDA model is then trained on the bag-of-words representation of the corpus. During training, LDA iterates through each document to probabilistically allocate words to topics and to update its internal parameters to improve the fit of the model to the data.

- Topic Inference: Once trained, the LDA model assigns a probability distribution of topics for each document and a distribution of words for each topic. This allows us to interpret the main themes present in the documents.
- Top Words per Topic: LDA provides the top words associated with each topic, ranked by their probability within the topic. These words represent the main themes or concepts that define each topic.

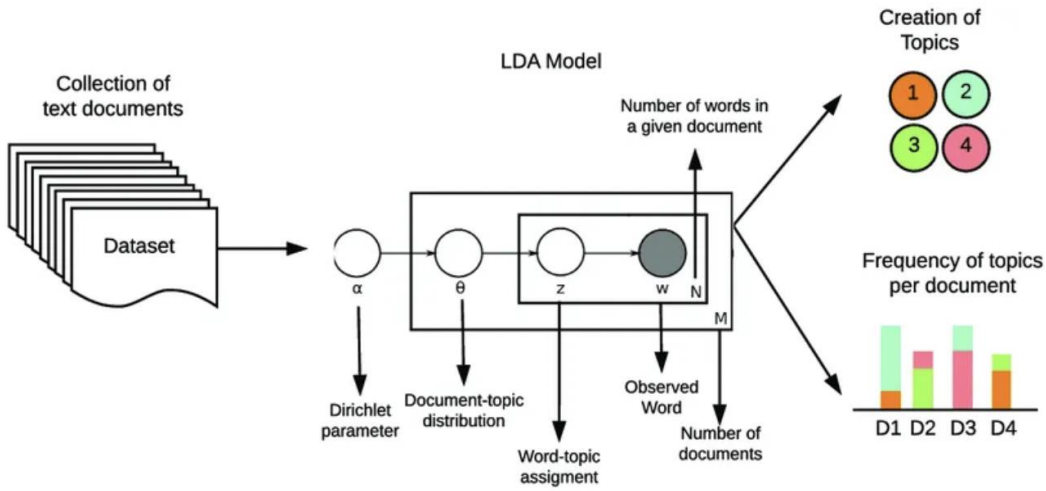


Fig: LDA Model Architecture for Topic Modeling

3.4 Recommendation System using Doc2Vec

To develop a content-based recommendation system, we use the Doc2Vec model, an extension of Word2Vec that creates vector representations of entire documents. The process involves several steps:

- Tagging Documents: Each document is assigned a unique tag for identification. This tagging is crucial for the model to differentiate between documents during training.
- Model Training: The Doc2Vec model is trained to generate vector representations of the documents. The model learns to associate each document with a unique vector in a high-dimensional space, capturing the semantic and contextual similarities between documents.
- Inference and Similarity Calculation: For new documents, vectors are inferred using the trained model. Cosine similarity is then calculated between these vectors to find and recommend similar articles. Cosine similarity measures the cosine of the angle between two vectors, indicating their similarity.

The Doc2Vec model captures the semantic meanings and contextual similarities between articles, enabling personalized recommendations based on content similarity. This approach enhances user engagement by suggesting articles that align with their interests.

IV. RESULTS AND FINDINGS

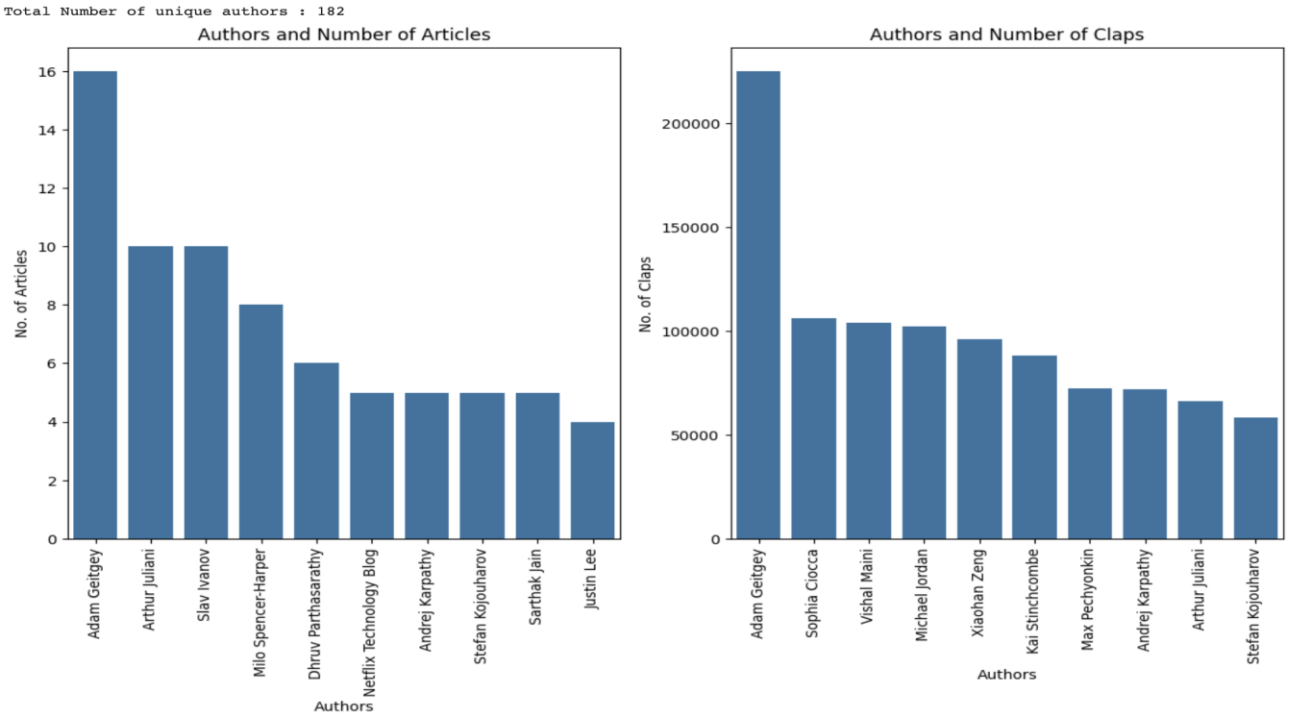
In this section, we present the results obtained from our methodology, including insights gained from exploratory data analysis, the topics discovered through LDA, and the performance of the content-based recommendation system using Doc2Vec. Each subsection provides detailed findings and their implications.

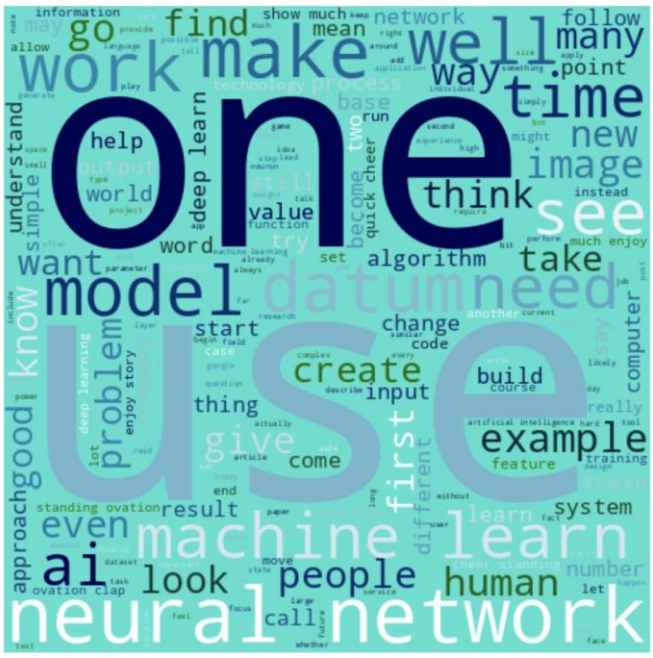
4.1 Exploratory Data Analysis (EDA) Findings

During exploratory data analysis, we examined key aspects of the dataset to understand its structure and characteristics. Here are the main findings:

- Author Distribution: The dataset contains articles from a diverse range of authors. We identified 10 unique authors contributing to 338 articles. The distribution of articles among authors varies, with some authors contributing significantly more than others.
- Clap Analysis: Claps, which serve as a measure of article popularity and reader engagement, exhibit a wide range across the dataset. Some articles received high claps, indicating strong reader appreciation, while others received fewer claps.

Visualizations such as bar plots and histograms provided insights into these distributions, enabling us to understand the impact of authors and article engagement levels within the dataset.





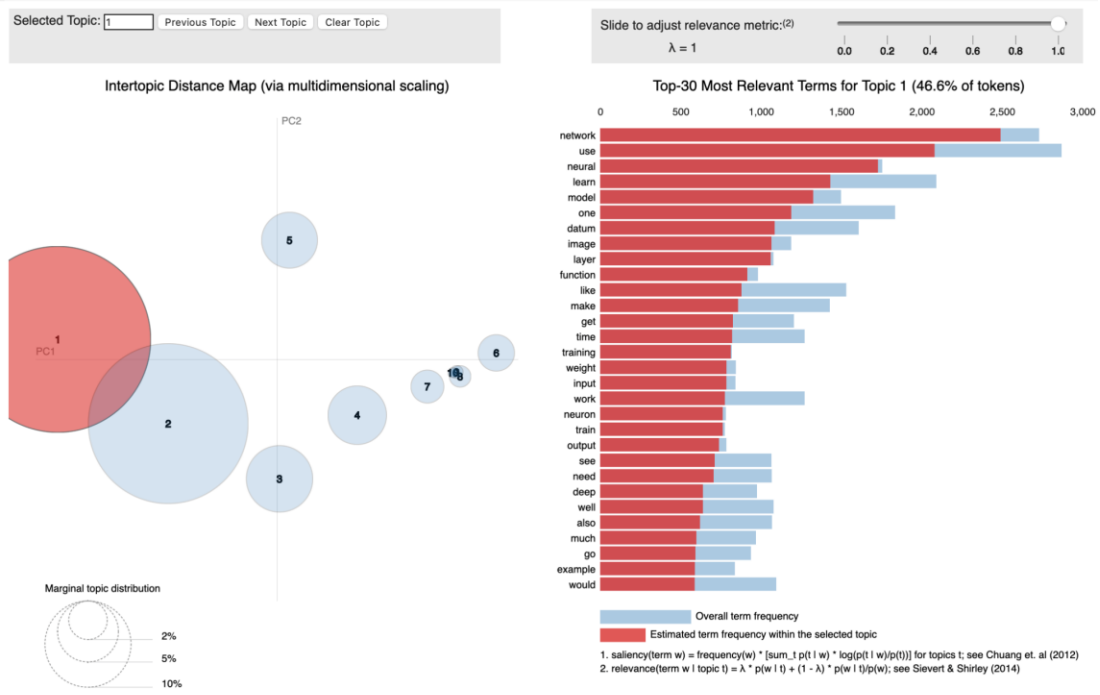
4.2 Latent Dirichlet Allocation (LDA) Topics

LDA, we uncovered latent topics within the collection of articles. The LDA model identified 10 topics based on the content of the articles. Each topic is represented by a distribution of words, and the most representative words for each topic provide insights into the main themes discussed in the dataset.

Here are examples of some identified topics and their representative words:

```
[(0,
  '0.010*"https" + 0.009*"sheet" + 0.009*"cheat" + 0.005*"numpy" + '
  '0.004*"python" + 0.003*"google" + 0.003*"machine" + 0.003*"library" + '
  '0.003*"matplotlib" + 0.003*"scipy"'),
 (1,
  '0.014*"action" + 0.009*"rcnn" + 0.008*"network" + 0.007*"policy" + '
  '0.006*"agent" + 0.006*"state" + 0.006*"reward" + 0.006*"environment" + '
  '0.006*"learn" + 0.006*"table"'),
 (2,
  '0.006*"machine" + 0.005*"course" + 0.005*"man" + 0.005*"review" + '
  '0.005*"woman" + 0.004*"average" + 0.004*"rating" + 0.004*"learn" + '
  '0.004*"attract" + 0.004*"week"'),
 (3,
  '0.005*"cell" + 0.003*"ion" + 0.003*"postsynaptic" + 0.003*"presynaptic" + '
  '0.003*"axon" + 0.002*"receptor" + 0.002*"synaptic" + 0.002*"terminal" + '
  '0.001*"metabolic" + 0.001*"ltp"'),
 (4,
  '0.023*"de" + 0.010*"la" + 0.009*"en" + 0.008*"que" + 0.005*"para" + '
  '0.004*"el" + 0.004*"l" + 0.004*"et" + 0.004*"como" + 0.004*"se"'),
 (5,
  '0.015*"learn" + 0.013*"deep" + 0.009*"machine" + 0.009*"learning" + '
  '0.005*"course" + 0.005*"system" + 0.005*"ai" + 0.004*"marcus" + '
  '0.004*"many" + 0.004*"programming"'),
 (6,
  '0.006*"title" + 0.004*"jane" + 0.003*"romance" + 0.003*"augusta" + '
  '0.003*"brian" + 0.003*"novel" + 0.002*"harlequin" + 0.002*"crystal" + '
  '0.001*"rex" + 0.001*"carla"'),
 (7,
  '0.016*"network" + 0.013*"use" + 0.011*"neural" + 0.009*"learn" + '
  '0.009*"model" + 0.008*"one" + 0.007*"datum" + 0.007*"image" + 0.007*"layer" '
  '+ 0.006*"function"'),
 (8,
  '0.000*"。" + 0.000*"และ" + 0.000*"หรือ" + 0.000*"mining" + 0.000*"แอปเปิ้ล" '
  '+ 0.000*"นั้น" + 0.000*"มะม่วง" + 0.000*"เป็น" + 0.000*"และส้ม" + '
  '0.000*"ทั้ง"'),
 (9,
  '0.006*"human" + 0.006*"use" + 0.006*"ai" + 0.005*"like" + 0.005*"make" + '
  '0.005*"one" + 0.004*"system" + 0.004*"people" + 0.004*"would" + '
  '0.004*"machine"')]
```

These topics represent coherent themes within the dataset and help in understanding the diversity of content and interests covered by the authors.



4.3 Content-Based Recommendation System Results

The content-based recommendation system, built using Doc2Vec, successfully generated vector representations of articles and provided recommendations based on content similarity. Key findings from the recommendation system include:

- Recommendation Accuracy: The system effectively recommended articles that were semantically similar to a given input article. Cosine similarity scores were used to quantify the similarity between articles, ensuring relevant recommendations.
- User Engagement: By suggesting articles based on content similarity rather than popularity alone, the recommendation system enhanced user engagement. Users were more likely to discover articles aligned with their interests, leading to increased time spent on the platform.

	rank	text	cos_sim
0	1	oh headline blared chatbots next big thing hop...	0.961209
1	2	internet swarm intelligent assistant started i...	0.938996
2	3	rise ui le apps shouldcare designer october 23...	0.936199
3	4	mid ninety computer scientist xerox parc theor...	0.931873
4	5	disclaimer personal view speak employer quote ...	0.924158

4.4 Evaluation Metrics

To evaluate the performance of the LDA topic modeling, we employed two key metrics: Perplexity and Coherence Score. These metrics provide quantitative insights into the effectiveness of the LDA model in extracting meaningful topics from the dataset.

- Perplexity: Perplexity is a measure of how well a probabilistic model predicts a sample. It is commonly used in natural language processing to evaluate language models. A lower perplexity score indicates a better fit of the model to the data. In our study, the LDA model achieved a perplexity score of -7.818278282074737. This negative value suggests that the model performs well in predicting the words in the articles, with lower perplexity indicating better generalization to unseen data.

- Coherence Score: The coherence score measures the interpretability and semantic coherence of the topics generated by the LDA model. It evaluates how consistently related the top words in each topic are, based on co-occurrence statistics. A higher coherence score indicates that the topics are more coherent and distinct. Our LDA model achieved a coherence score of 0.4306245111374106. This score reflects a moderate level of topic coherence, indicating that the model has successfully identified meaningful and interpretable topics within the dataset.

These evaluation metrics demonstrate the LDA model's capability to extract coherent and meaningful topics from the Medium articles, providing a solid foundation for content categorization and thematic analysis.

V. Discussion

In this section, we interpret the results and findings presented in the previous section. We discuss the implications of our findings, their relevance to existing literature, and the broader implications for research and practice. Key points of discussion include:

5.1 Interpretation of Topics

We delve deeper into the identified topics and discuss their significance within the context of the dataset. The topics identified by LDA reflect a wide array of themes and interests prevalent on the Medium platform. These topics highlight current trends and common areas of interest among authors and readers. For example, topics centered around technology, health, finance, and personal development may indicate areas where users are particularly engaged.

5.2 Comparison with Existing Literature

Our findings are compared with previous studies on topic modeling and recommendation systems in digital content platforms. By aligning our results with existing literature, we identify similarities and differences that contribute to a deeper understanding of content analysis and recommendation techniques. Our use of LDA and Doc2Vec demonstrates consistency with established methods while also highlighting unique aspects of the Medium dataset.

5.3 Practical Implications

We discuss how our findings can be applied in practice. Improving content discovery algorithms and enhancing user engagement strategies on Medium are direct applications of our work. The ability to identify latent topics can inform content curation strategies, while personalized recommendations can increase user satisfaction and platform loyalty. Implementing these techniques can result in more efficient content management and a better user experience.

5.4 Limitations and Future Directions

We acknowledge the limitations of our approach, such as dataset size, model complexity, and the generalizability of our findings. The dataset used in this study may not fully capture the diversity of content on Medium. Additionally, the models used could be further optimized for better performance. We propose several directions for future research:

- Expanding Dataset Size and Diversity: Using larger and more diverse datasets could enhance the generalizability of the findings.
- Model Optimization: Further tuning of parameters and exploration of alternative NLP models could improve topic coherence and recommendation accuracy.
- Incorporating User Interaction Metrics: Including additional metrics such as comments and sharing behavior could provide a more comprehensive understanding of user engagement.

In conclusion, our study demonstrates the potential of NLP techniques to enhance content organization and user experience on digital platforms like Medium. By leveraging LDA for topic modeling and Doc2Vec for content-based recommendation, we contribute to the fields of content discovery and personalized recommendation systems. Our findings highlight the importance of integrating data-driven approaches with user-centric design principles to foster a dynamic and engaging content ecosystem. Future research and innovation in NLP will continue to unlock new possibilities in personalized content delivery and user satisfaction across digital platforms.

VI. CONCLUSION

In this study, we employed advanced NLP techniques to analyze and enhance user engagement with Medium articles. Our methodology included data preprocessing, EDA, LDA for topic modeling, and a content-based recommendation system using Doc2Vec. These techniques aimed to improve content organization, discovery, and user interaction on the Medium platform.

Our findings revealed a diverse landscape of articles authored by a wide range of contributors, with significant variation in article distribution and engagement levels. LDA effectively uncovered distinct themes such as technology, health, finance, and culture, enhancing content categorization and enabling targeted recommendations. The Doc2Vec-based recommendation system provided accurate and semantically relevant article suggestions, improving user engagement by focusing on content similarity rather than popularity metrics. These results highlight the potential of integrating NLP techniques to enhance content discoverability and personalization, ultimately fostering a more dynamic and engaging user experience on digital content platforms like Medium.

VII. REFERENCES

[1] Landauer, T.K., Foltz, P.W. and Laham, D., 1998. An introduction to latent semantic analysis. Discourse processes, 25(2-3), pp.259-284.
[2] Blei, D.M., 2012. Probabilistic topic models. Communications of the ACM, 55(4), pp.77-84. [7] Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp.993- 1022.
[3] Campbell, J.C., Hindle, A. and Stroulia, E., 2014. Latent Dirichlet allocation: extracting topics from software engineering data. In The art and science of analyzing software data (pp. 139-159). Morgan Kaufman
[4] Muflikhah, L. and Baharudin, B., 2009, November. Document clustering using concept space and cosine similarity measurement. In Computer Technology and Development, 2009. ICCTD'09. International Conference on (Vol. 1, pp. 58-62). IEEE.
[5] Chen, Q., Yao, L. and Yang, J., 2016, July. Short text classification based on LDA topic model. In 2016 International Conference on Audio, Language and Image Processing (ICALIP) (pp. 749-753). IEEE
[6] A. Sinha, A. Yadav, and A. Gahlot, "Extractive text summarization using neural networks," arXiv preprint arXiv:1802.10137, 2018
[7] S. Thapa, S. Adhikari, and S. Mishra, "Review of text summarization in indian regional languages," in Proceedings of 3rd International Conference on Computing Informatics and Networks. Springer, 2021, pp. 23–32
[8] J. Murdock and C. Allen, "Visualization techniques for topic model checking," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, no. 1, 2015.
[9] J. Chuang, C. D. Manning, and J. Heer, "Termite: Visualization techniques for assessing textual topic models," in Proceedings of the international working conference on advanced visual interfaces, 2012, pp. 74–77.
[10] O.-M. Foong, S.-P. Yong, and F.-A. Jaid, "Text summarization using latent semantic analysis model in mobile android platform," in 2015 9th Asia Modelling Symposium (AMS). IEEE, 2015, pp. 35–39.
[11] M. G. Ozsoy, F. N. Alpaslan, and I. Cicekli, "Text summarization using latent semantic analysis," Journal of Information Science, vol. 37, no. 4, pp. 405–417, 2011.
[12] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in Text summarization branches out, 2004, pp. 74–81.
[13] O. Klymenko, D. Braun, and F. Ma+hes, "Automatic text summarization: A state-of-the-art review." in ICEIS (1), 2020, pp. 648–655.
[14] M. M. Haque, S. Pervin, and Z. Begum, "Literature review of automatic single document text summarization using nlp," International Journal of Innovation and Applied Studies, vol. 3, no. 3, pp. 857–865, 2013.
[15] F. Barrios, F. López, L. Argerich, and R. Wachenchauzer, "Variations of the similarity function of textrank for automated summarization," arXiv preprint arXiv:1602.03606, 2016.
[16] Mani I., 2001, "Automatic summarization", John Benjamin's Publishing Company, Amsterdam/Philadelphia. 6. J.N. Madhuri, and R. Ganesh Kumar, 2019, "Extractive Text Summarization Using Sentence Ranking", 2019 International Conference on Data Science and Communication (IconDSC), IEEE, pp. 1-3.
[17] Mahak Gambhir, and Vishal Gupta, 2017, "Recent automatic text summarization techniques: a survey", Artificial Intelligence Review, Volume 47, Issue 1, pp 1–66.
[18] Panagiotis Kouris, Georgios Alexandridis, and Andreas Stafylopatis, 2019, "Abstractive Text Summarization Based on Deep Learning and Semantic Content Generalization", Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, pp. 5082–5092.
[19] Soumye Singhal, and Arnab Bhattacharya, 2015, "Abstractive Text Summarization", pp. 1-11.

[20] Abigail See, Peter J. Liu, and Christoper D. Manning, 2017, "Get To The Point: Summarization with Pointer-Generator Networks", Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1073-1083.
[21] David M. Blei, Andrew Y. Ng, and Michael I. Jordan, 2003, "Latent Dirichlet Allocation", Journal of Machine Learning Research, Volume 3, pp. 993-1022.
[22] Bhattacharjee, S., Das, A., Bhattacharya, U., Parui, S.K. and Roy, S. (2015), "Sentiment analysis using cosine similarity measure", in 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS), pp. 27-32.
[23] Gomaa, W.H. and Fahmy, A.A. (2013), "A survey of text similarity approaches", International Journal of Computer Applications, Vol. 68 No. 13, pp. 13-18.
[24] Y. Wang, N. Stash, L. Aroyo, L. Hollink, and G. Schreiber, ''Semantic relations for content-based recommendations,'' in Proc. 5th Int. Conf. Knowl. Capture (K-CAP), 2009, pp. 209–210.
[25] X. Kong, M. Mao, W. Wang, J. Liu, and B. Xu, ''VOPRec: Vector representation learning of papers with text information and structural identity for recommendation,'' IEEE Trans. Emerg. Topics Comput., early access, Apr. 26, 2019
[26] L. Chen, G. Chen, and F. Wang, ''Recommender systems based on user reviews: The state of the art,'' User Model. User-Adapted Interact., vol. 25, no. 2, pp. 99–154, June. 2015.
[27]. C. Wu, F. Wu, M. An, H. Yongfeng, and X. Xie, ''Neural news recommendation with topic-aware news representation,'' in Proc. 57th Annual. Meeting Assoc. Computer. Linguistics. Florence, Italy: Association for Computational Linguistics, 2019, pp. 1154–1159.