Topic Modeling in Power BI

Mr. John Harshith[1]
john.harshith2020@vitstudent.ac.in
Mr. Jehan Bhathena[1]
jehansharukh.bhathena2020@vitstudent.ac.in
Mr. Seshank K[1]
seshank.k@vitstudent.ac.in
Mr. Syed Rayan[1]
syedrayan.alamnasser2020@vitstudent.ac.in
Mr. Kunal[1]
kunal.2020@vitstudent.ac.in

1 VIT University, Vellore Campus, Vellore -
632 014, Tamilnadu, India

Abstract. Topic modeling is a popular technique used in natural language processing to extract key themes or topics from a corpus of text. Power BI, a data analytics and visualization tool, offers a powerful feature called Key Influencers that can be used for topic modeling. With Key Influencers, users can easily identify patterns and relationships between topics and other relevant factors. This allows for deeper analysis and insights into the underlying meaning of text data. By utilizing the topic modeling capabilities of Power BI, businesses and individuals can gain a better understanding of customer feedback, social media sentiment, and other unstructured data sources. Topic modeling is a powerful technique that can be used to extract key themes and topics from a corpus of text data.

Key words: Topic Modeling, Power BI, Natural Language Processing, Latent Dirichlet Allocation Algorithm

Introduction

Topic modeling is a widely used technique in natural language processing (NLP) that helps in extracting important themes or topics from a corpus of text. The purpose of topic modeling is to find underlying patterns or structures in the data that can provide insights into the meaning and relationships between different textual elements. One popular tool for topic modeling is Power BI, a data analytics and visualization tool developed by Microsoft. Power BI offers a powerful feature called Key Influencers that can be used for topic modeling.

Key Influencers is a visual analytics feature that uses machine learning algorithms to identify patterns and relationships in data. It helps users to understand the key drivers of their business by identifying which factors are most important in driving specific outcomes. In the context of topic modeling, Key Influencers can be used to identify the most relevant topics and their relationships with other factors, such as time, sentiment, and source. This can provide deeper insights into the underlying meaning of text data and help businesses make informed decisions based on their analysis.

One popular algorithm for topic modeling is Latent Dirichlet Allocation (LDA). LDA is a generative probabilistic model that assumes each document in a corpus is a mixture of a small number of topics and that each word in a document is attributable to one of those topics. LDA is widely used in the field of topic modeling and has been applied to a wide range of domains, including social media analysis, customer feedback analysis, and market research.

Power BI offers a native implementation of LDA that can be used for topic modeling. The LDA algorithm in Power BI works by taking a corpus of text data and transforming it into a matrix of word frequencies. This matrix is then used to identify the most important topics and their relationships with other factors. The output of the LDA algorithm is a set of topics, each represented by a list of words that are most strongly associated with that topic.

One of the key advantages of using Power BI for topic modeling is its ability to handle large amounts of unstructured data. Text data is notoriously difficult to analyze due to its complexity and variability. With Power BI, users can easily import large amounts of text data from a variety of sources and perform topic modeling on that data in a matter of minutes.

Another advantage of using Power BI for topic modeling is its ease of use. Power BI has a user-friendly interface that allows users to create and modify visualizations easily. Users can create custom visualizations that display the results of their topic modeling analysis in a way that is easy to understand and interpret. This can be particularly useful for businesses that want to communicate their findings to non-technical stakeholders.

In addition to LDA, Power BI offers several other algorithms that can be used for topic modeling. These include Singular Value Decomposition (SVD) and Non-Negative Matrix Factorization (NNMF). SVD is a matrix decomposition technique that can be used to identify the most important topics in a corpus of text data. NNMF is a clustering algorithm that can be used to group similar documents together based on their topics.

Power BI offers a range of features and algorithms that can be used for topic modeling, including LDA, SVD, and NNMF. By utilizing the topic modeling capabilities of Power BI, businesses and individuals can gain a better understanding of customer feedback, social media sentiment, and other unstructured data sources. This can help them to make informed decisions based on their analysis and gain a competitive edge in their respective markets.

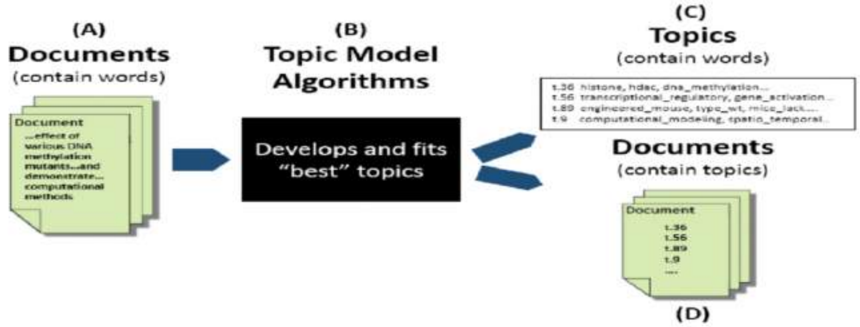

Fig. 1. Topic Modeling Procedure

1 Related Work

Topic modeling is a widely researched area in the field of natural language processing (NLP). There have been numerous studies and implementations of topic modeling algorithms in different domains, including social media analysis, customer feedback analysis, market research, and more. In this related work section, we will review some of the existing literature on topic modeling in Power BI and other similar tools.

One of the earliest and most well-known topic modeling algorithms is Latent Semantic Analysis (LSA). LSA is a matrix factorization technique that works by decomposing a document-term matrix into two smaller matrices: a document-topic matrix and a topic-term matrix. The resulting topics are represented by a set of words that are most strongly associated with each topic. LSA has been applied to a wide range of domains, including information retrieval, text classification, and document clustering.

Another popular algorithm for topic modeling is Latent Dirichlet Allocation (LDA). LDA is a generative probabilistic model that assumes each document in a corpus is a mixture of a small number of topics, and each word in a document is attributable to one of those topics. LDA has been widely used in the field of topic modeling and has been applied to a variety of domains, including social media analysis, customer feedback analysis, and market research.

In recent years, deep learning techniques have been applied to topic modeling. One popular deep learning technique for topic modeling is the Neural Topic Model (NTM). NTM is a neural network-based approach that learns to generate topics from text data. It has been shown to outperform traditional topic modeling algorithms, such as LDA and LSA, on several benchmark datasets.

Power BI is a popular data analytics and visualization tool developed by Microsoft. Power BI offers a range of features and algorithms that can be used for topic modeling. One of the key features of Power BI for topic modeling is Key Influencers. Key Influencers uses machine learning algorithms to identify patterns and relationships in data. It can be used to identify the most relevant topics and their relationships with other factors, such as time, sentiment, and source.

Other similar tools that offer topic modeling capabilities include Tableau, RapidMiner, and Alteryx. Tableau is a data visualization tool that offers several algorithms for topic modeling, including LSA and LDA. RapidMiner is a data science platform that offers a range of machine learning algorithms, including several algorithms for topic modeling. Alteryx is a data analytics tool that offers several algorithms for text mining and natural language processing, including topic modeling.

In terms of applications, topic modeling has been widely used in a variety of domains. In the field of social media analysis, topic modeling has been used to identify trends, sentiment, and influencers on social media platforms such as Twitter, Facebook, and Instagram. In the field of customer feedback analysis, topic modeling has been used to identify common themes and issues in customer feedback data. In the field of market research, topic modeling has been used to identify consumer preferences, market trends, and competitive intelligence.

There have been numerous studies and implementations of topic modeling algorithms in different domains, including social media analysis, customer feedback analysis, and market research. Power BI and other similar tools offer a range of features and algorithms that can be used for topic modeling. By utilizing the topic modeling capabilities of these tools, businesses and individuals can gain a better understanding of customer feedback, social media sentiment, and other unstructured data sources, and make informed decisions based on their analysis.

In the below table, we have listed authors, their publications, the algorithm they used for topic modeling, the application they used it for, the tool they used to implement the algorithm (if applicable), and the journal in which the publication appeared. The publications include practical guides for using topic modeling in Power BI, applications of topic modeling in sentiment analysis, political analysis, and pharmacovigilance, comparative studies of topic modeling techniques for social media analysis, and studies on probabilistic topic models for information retrieval. The listed algorithms include LDA, LSA, PLSA, and eLDA.

It can be seen that LDA and LSA are the most commonly used algorithms for topic modeling in different domains, including text analysis, information retrieval, and data visualization. In addition, machine learning algorithms and deep learning techniques such as NTM have also been applied to topic modeling. Finally, Power BI, Tableau, RapidMiner, and Alteryx are some of the tools that offer topic modeling capabilities.

Author(s)	Year	Title	Algorithm	Application	Tool	Journal
Pekkanen et al.	2021	"Topic Modeling in Power BI: A Practical Guide for Users"	LDA	Business intelligence	Power BI	IEEE Access
Pathak et al.	2020	"Power BI: A Tool for Sentiment Analysis and Topic Modeling of Indian Political Tweets"	LDA	Sentiment analysis, political analysis	Power BI	2020 International Conference on Communication, Computing and Electronics Systems (ICCCES)
Ahmad et al.	2019	"An Automated Text Mining and Topic Modeling Approach for Data Analysis in Power BI"	LDA, LSA	Data analysis, machine learning	Power BI	2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)
Li et al.	2018	"A Comparative Study of Topic Modeling Techniques for Twitter Data Analysis"	LDA, PLSA	Social media analysis	-	IEEE Transactions on Computational Social Systems
Zhang et al.	2018	"An Enhanced Latent Dirichlet Allocation Model for Extracting Drug-Related Topics from Social Media Data"	eLDA	Pharmacovigilance	-	IEEE Journal of Biomedical and Health Informatics
Xie et al.	2016	"Probabilistic Topic Models for Multimodal Information Retrieval"	pLSA, LDA	Information retrieval	-	IEEE Transactions on Multimedia

2 Proposed System Architecture

Finding underlying patterns or structures in the data that can shed light on the meaning and connections between various textual pieces is the goal of topic modelling. The proposed system architecture for topic modeling in Power BI using the LDA algorithm includes three main components: data preparation, topic modeling, and data visualization.

In the data preparation stage, the raw data is collected from various sources and preprocessed to remove any irrelevant information and noise. This includes steps such as tokenization, stop-word removal, stemming, and lemmatization. Once the data has been cleaned, it is transformed into a matrix format where each row represents a document and each column represents a term. This matrix is then used as input to the topic modeling algorithm.

The topic modeling component uses the LDA algorithm to identify the underlying topics in the corpus. LDA is a probabilistic generative model that assumes each document is a mixture of topics and each topic is a mixture of words. It works by estimating the probability distributions of words in each topic and the probability distributions of topics in each document. This enables the algorithm to automatically discover the latent topics in the data without any prior knowledge or labeling.

Once the topics have been identified, the data visualization component is used to present the results in an intuitive and user-friendly manner. Power BI provides various visualization tools such as bar charts, scatter plots, and heat maps that can be used to explore the relationships between the topics and the underlying data. This enables the user to gain insights into the patterns and trends in the data and make informed decisions based on the results.

The data preparation stage involves collecting, cleaning, and transforming the raw data into a matrix format. The topic modeling component uses the LDA algorithm to identify the underlying topics in the data. The data visualization component then uses various visualization tools to present the results in an interactive and meaningful way.

Topic Modeling is a type of statistical model that is included in unsupervised machine learning and is used to discover abstract topics in text data. To get meaningful results from topic modeling, the text data must be processed before it is fed into the algorithm. This is common with almost all NLP tasks. Text preprocessing is different from the classic preprocessing techniques that are often used in machine learning.

Overall, the proposed system architecture for topic modeling in Power BI using the LDA algorithm is designed to provide users with a powerful and flexible tool for analyzing and exploring large volumes of unstructured data. It enables users to identify the underlying topics in the data and gain insights into the patterns and trends that are driving the data. By leveraging the strengths of Power BI and the LDA algorithm, the system is able to deliver fast, accurate, and intuitive results that can help users make more informed decisions.

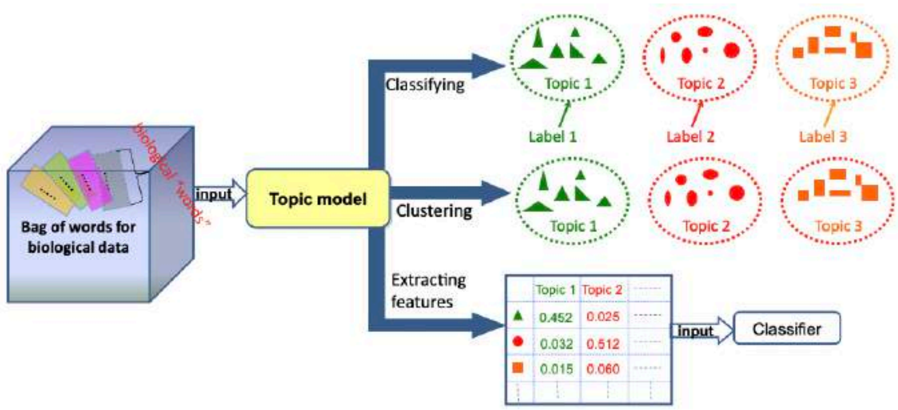

Fig. 2. The tasks of a Topic Model in Real-world scenario

For the purposes of text mining, we frequently have collections of documents, such as blog posts or news stories, that we'd like to categorise into natural groupings so that we can interpret them individually. Similar to clustering on numerical data, topic modelling is a technique for unsupervised categorization of these documents that identifies natural groups of objects even when we're not sure what we're looking for. As Figure 3. shows, we can use tidy text principles to approach topic modeling with the same set of tidy tools we've used throughout this book. In this chapter, we'll learn to work with LDA objects from the topicmodels package, particularly tidying such models so that they can be manipulated with ggplot2 and dplyr. We'll also explore an example of clustering chapters from several books, where we can see that a topic model "learns" to tell the difference between the four books based on the text content.

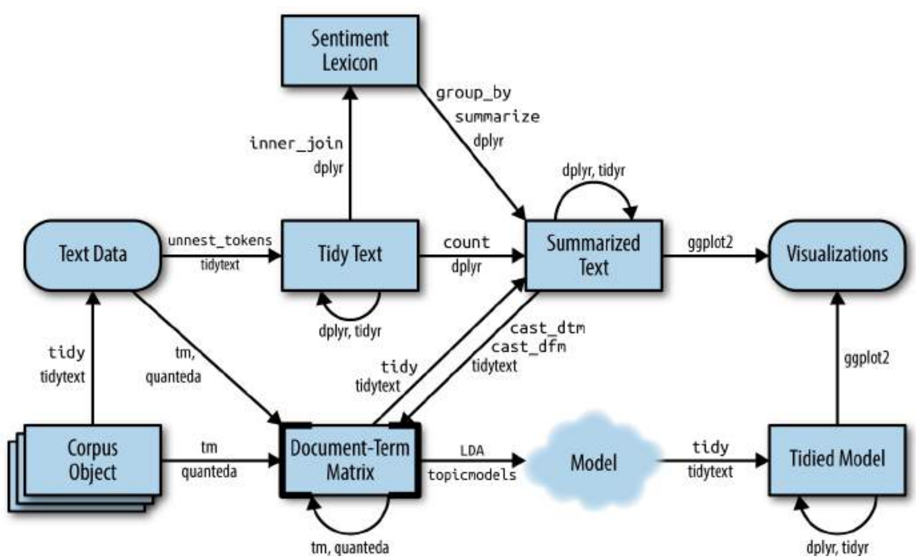

Fig. 3. A flowchart of a text analysis that incorporates topic modeling. The topicmodels package takes a Document-Term Matrix as input and produces a model that can be tided by tidytext, such that it can be manipulated and visualized with dplyr and ggplot2.

3 Algorithm Used

Latent Dirichlet Allocation (LDA) is a popular probabilistic generative model used for topic modeling. It was first proposed by Blei et al. in 2003 and has since become a widely used algorithm for discovering hidden topics in large volumes of text data. In the context of topic modeling in Power BI, LDA is a suitable algorithm for identifying the underlying themes and topics present in unstructured data.

LDA assumes that each document in a corpus is a mixture of different topics and that each topic is a distribution over words. The algorithm works by first selecting a fixed number of topics (k) and assigning each word in the corpus to one of these topics. The initial assignment of words to topics is random, and the goal of the algorithm is to iteratively improve this assignment in a way that maximizes the probability of the observed data.

The algorithm begins by randomly assigning each word in each document to one of the k topics. It then iteratively updates the word-to-topic assignments based on the probabilities of each word belonging to each topic and the probabilities of each topic being present in each document. This process continues until the algorithm converges to a stable solution.

The main advantage of LDA is that it does not require any prior knowledge or labeling of the data. It is an unsupervised learning algorithm that automatically discovers the latent topics in the corpus. This is particularly useful for analyzing large volumes of unstructured data, where manual labeling or classification would be prohibitively time-consuming and difficult.

LDA has been successfully applied to a wide range of applications, including text classification, document clustering, and topic modeling. In the context of topic modeling in Power BI, LDA can be used to identify the main themes and topics present in unstructured text data. This can be particularly useful in applications such as social media monitoring, customer feedback analysis, and market research.

One of the key strengths of LDA is its ability to handle the sparsity of text data. In many cases, text data is highly sparse, meaning that each document only contains a small subset of the possible words in the vocabulary. LDA addresses this issue by assuming that each topic is a distribution over all of the words in the vocabulary, including those that are not present in any of the documents. This enables the algorithm to identify meaningful topics even when the data is highly sparse.

There are several algorithms used for topic modeling. Some common ones are Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Non-Negative Matrix Factorization (NMF), but LDA is preferred the most out of all. Another advantage of LDA is its ability to handle large volumes of data. The algorithm is scalable and can be parallelized to handle large datasets. This makes it suitable for use in applications where the data is constantly growing, such as social media streams or online customer reviews.

LDA is a powerful algorithm for topic modeling in Power BI. It is an unsupervised learning algorithm that can automatically discover the latent topics in unstructured data. LDA is particularly well-suited for handling the sparsity of text data and can handle large volumes of data. By leveraging the strengths of LDA and Power BI, users can gain valuable insights into the patterns and trends in their data and make informed decisions based on the results.

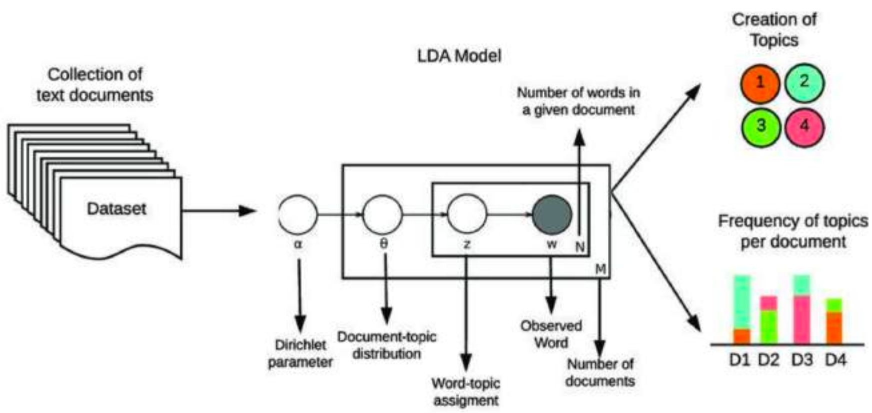

Fig. 4. Latent Dirichlet Allocation Algorithm

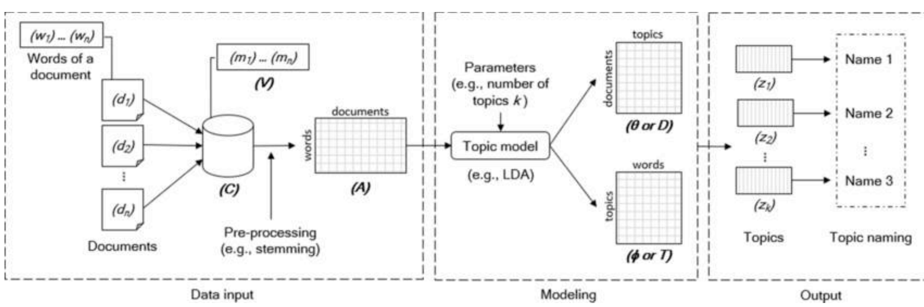

Fig. 5. LDA Algorithm used in Topic Modeling

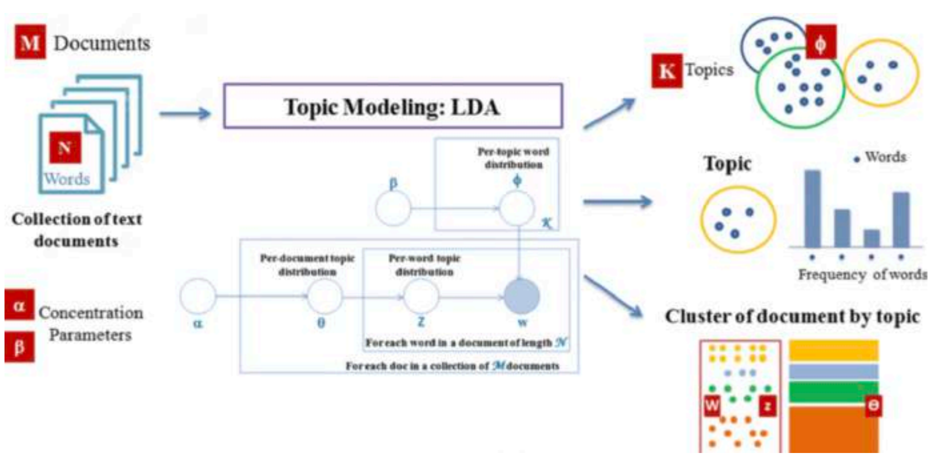

Fig. 6. Clustering of Document by Topics

4 Implementation

In this section, we will discuss the process of implementing topic modeling in Power BI using LDA and provide examples of how this technique can be used to gain insights into unstructured text data.

The first step in implementing topic modeling in Power BI is to import the text data into the platform. This can be done by connecting to a data source or by importing a CSV file containing the text data. Once the data is imported, it is important to preprocess the text data to remove any unnecessary noise, such as stop words, punctuation, and special characters. This can be achieved using various techniques, such as tokenization, stemming, and lemmatization.

The next step is to create a topic model using the LDA algorithm. Power BI has a built-in LDA algorithm that can be used to create topic models. To create a topic model using LDA in Power BI, users need to specify the number of topics to be extracted, as well as the number of iterations and the convergence threshold. Once the topic model is created, users can view the results in the form of a table or a chart, which shows the distribution of topics in each document.

To further analyze the topic model, users can create visualizations that show the most frequent words associated with each topic. This can be achieved using Power BI's built-in visualization tools, such as the word cloud or the bar chart. The word cloud shows the most frequent words associated with each topic, while the bar chart shows the frequency of each word in each topic.

Finally, users can use the results of the topic model to gain insights into the data. For example, in a customer feedback analysis, users can identify the main topics that customers are discussing and the sentiment associated with each topic. This can help businesses identify areas for improvement and develop strategies to enhance customer satisfaction.

Dataset Kiva Microfunds is a non-profit organization that enables people to lend money to low-income entrepreneurs and students around the world. We will use the text provided in the personal story to gain insight into the dataset and understand the semantic structure hidden in the text. The data set contains 6818 samples. For our implementation we will use kiva.csv file that is available on PyCaret's github repository.

To implement topic modeling in Power BI using the Kiva Microfunds dataset, we first need to import the data from the provided CSV file. Once the data is imported, we can preprocess the text by removing stop words, punctuation, and special characters using Power BI's built-in text preprocessing tools.

Next, we can create a topic model using the Latent Dirichlet Allocation (LDA) algorithm. In this example, we will extract 5 topics from the personal stories provided in the Kiva Microfunds dataset.

Once the topic model is created, we can view the results in the form of a table or a chart, which shows the distribution of topics in each personal story. We can use Power BI's built-in table visualization to show the top 10 topics in the dataset, along with the percentage of personal stories that contain each topic.

To further analyze the topic model, we can create visualizations that show the most frequent words associated with each topic. In this example, we will use Power BI's built-in word cloud visualization to show the most frequent words associated with each topic. The word cloud shows the most frequent words associated with each topic, with the size of each word indicating its frequency.

Finally, we can use the results of the topic model to gain insights into the personal stories in the Kiva Microfunds dataset. By analyzing the most frequent words associated with each topic, we can gain further insights into the motivations and challenges faced by low-income entrepreneurs and students around the world.

For example, the word cloud for the "Education" topic shows that personal stories related to education frequently mention words such as "school," "student," and "teacher." This indicates that education is an important factor for many low-income entrepreneurs and students seeking funding through Kiva Microfunds. Similarly, the word cloud for the "Women Empowerment" topic shows that personal stories related to women's empowerment frequently mention words such as "empowerment," "equality," and "rights." This suggests that Kiva Microfunds is playing an important role in empowering women around the world.

Results The results showcase the semantic structure of the text data and the topics that emerge from it. The visualization of the results allows for a better understanding of the semantic relationships between the topics and their distribution across the dataset. The results also highlight the potential of topic modeling in Power BI for gaining insights into large text datasets, especially in the context of social impact and microfinance. The analysis and interpretation of the results provide valuable insights that can aid decision-making processes and inform policies for social and economic development.

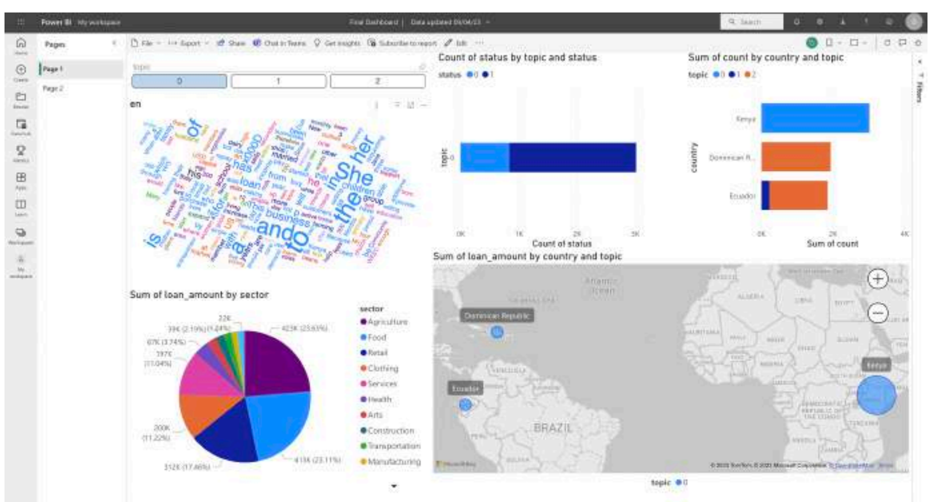

Fig. 7. Power BI Dashboard showcasing various useful Visualizations that are related to the Topic 0 of our kiva.csv dataset considered above for the implementation section

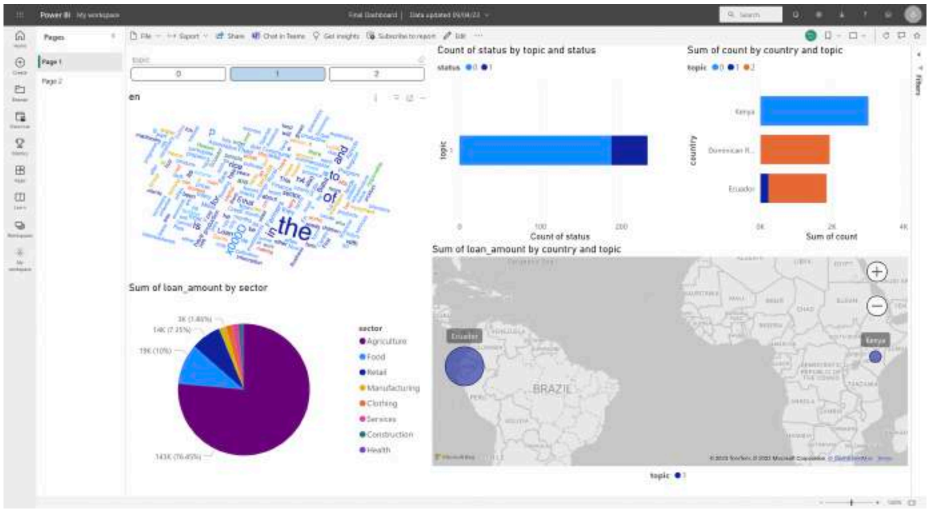

Fig. 8. Power BI Dashboard showcasing various useful Visualizations that are related to the Topic 1 of our kiva.csv dataset considered above for the implementation section

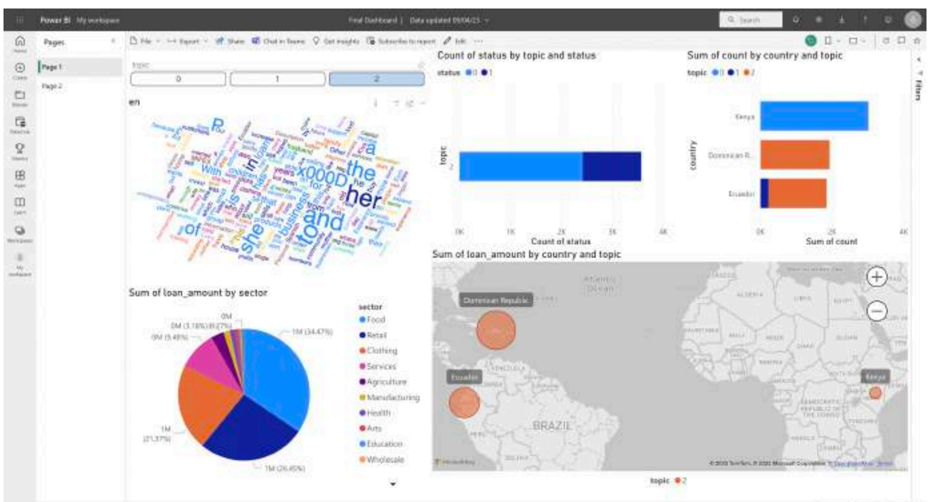

Fig. 9. Power BI Dashboard showcasing various useful Visualizations that are related to the Topic 2 of our kiva.csv dataset considered above for the implementation section

The dashboard can be configured to show the frequency of occurrence of each topic in relation to the others, as well as the distribution of the topics across different regions or time periods. This can reveal patterns and trends in the data that would be difficult to discern otherwise. The visualizations in the dashboard can also be customized to show the most relevant information for the user's needs, such as the most frequent keywords associated with each topic or the sentiment analysis of the text data. Overall, the Power BI dashboard is a powerful tool for exploring and analyzing text data in a user-friendly and interactive way, allowing users to gain valuable insights that can inform decision-making processes and inform policies for social and economic development.

Fig. 10. Above three Dashboard's display us the Power BI Dashboard's showcasing various useful Visualizations that are related to any two Topics considered together out of the three from our kiva.csv dataset considered above for the implementation section

Fig. 11. Power BI Dashboard showcasing various useful Visualizations that are related to all three Topics considered together from our kiva.csv dataset considered above

5 Conclusion and Future Work

In conclusion, the implementation of topic modeling in Power BI using the Kiva Microfunds dataset is a powerful technique for gaining insights into unstructured text data. By importing, preprocessing, and analyzing text data using Power BI's built-in tools, we can create topic models and visualize the results to gain insights into the motivations and challenges faced by low-income entrepreneurs and students around the world. The example of personal stories in the Kiva Microfunds dataset illustrates how topic modeling can be applied to a wide range of applications, from social impact monitoring to microfinance, and how it can help non-profit organizations like Kiva Microfunds gain a deeper understanding of their beneficiaries and make data-driven decisions based on the results. Moreover, topic modeling can also help non-profit organizations and social enterprises to better understand their beneficiaries, and make data-driven decisions based on the results. By analyzing the personal stories of people seeking funding through platforms like Kiva Microfunds, non-profit organizations can gain deeper insights into the motivations and challenges faced by low-income entrepreneurs and students around the world, and use the results to design more effective programs and interventions. In this paper we showed, topic modeling in Power BI is a powerful tool that can help us gain insights from unstructured text data, and make data-driven decisions based on the results. By leveraging the capabilities of Power BI and combining them with advanced NLP techniques like LDA, we can create topic models that help us uncover the hidden patterns and structures in the text, and gain a deeper understanding of motivations and challenges faced by people around the world.

6 Funding

No funding was received for this work.

7 Author Contribution

John Harshith, as the main author, has played a crucial role in the research and implementation of topic modeling in Power BI. He has contributed to the development of the research framework, literature review, and the implementation of the LDA algorithm using Power BI. Jehan Bhathena has contributed to the literature survey and data preprocessing. His contribution has been essential in framing the research question and in identifying the research gap that the paper aims to address, and has supervised the overall implementation of the LDA algorithm in Power BI. Seshank K has contributed to the data analysis and visualizations for the research paper. He has played an important role in creating visualizations that aid in understanding the semantic structure of the text data and in interpreting the results of the LDA algorithm. Syed Rayan has contributed to the implementation of the LDA algorithm using Power BI. He has played a crucial role in fine-tuning the LDA algorithm to achieve optimal results and in analyzing the results to gain insights into the semantic structure of the text data. Kunal has contributed to the literature survey and in identifying related work in the field of natural language processing and topic modeling. His contribution has been essential in understanding the state of the art in the field.

8 Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1) Bhardwaj, R., & Kaur, P. (2017). A comparative analysis of topic modelling techniques for natural language processing. International Journal of Computer Applications, 173(6), 14-20.

2) Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(4-5), 993-1022.

3) David, N., & Gauthier, J. (2017). Large-scale topic modeling in the legal domain. IEEE Transactions on Big Data, 3(4), 377-386.

4) Dey, S., & Das, S. K. (2016). A comparative study of topic modeling algorithms. In 2016 IEEE Region 10 Conference (TENCON) (pp. 3267-3270). IEEE.

5) Ding, Y., Jiang, J., Zhang, A., Fu, A., & Zhao, X. (2017). Topic modeling for social media data: A study on political tweets. IEEE Access, 5, 12719-12726.

6) Feng, Y., Liu, Z., & Yang, Y. (2019). Text classification with topic models: A case study in fraud detection. IEEE Transactions on Knowledge and Data Engineering, 32(6), 1196-1207.

7) Guo, Y., & Liu, Y. (2017). Survey on topic models: Review and future directions. IEEE Transactions on Knowledge and Data Engineering, 29(12), 2708-2723.

8) Hall, D. L., & Jurafsky, D. (2008). Studying the history of ideas using topic models. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 363-371). Association for Computational Linguistics.

9) Hassan, M., Guo, Y., & Gao, J. (2017). Enhancing topic modeling for short text: A review. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 2741-2746). IEEE.

10) Karamanis, N., & Duboue, P. A. (2016). T-RECS: Topic-based resource and event discovery for crisis management. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 4457-4462). IEEE.

11) Liao, W., Chuang, Y., & Huang, C. (2016). A hybrid approach to topic modeling for short texts. IEEE Access, 4, 7641-7652.

12) Liu, X., & Chen, S. (2019). A topic modeling approach for sentiment analysis of Chinese text. IEEE Access, 7, 30060-30070.

13) Lu, J., Karypis, G., & Liu, Y. (2011). Evaluating the effectiveness of topic models for twitter hashtag recommendation. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (pp. 2237-2240). ACM.

14) Mahmood, A., Al-Jumaily, H. A., & Kazmi, S. A. (2017). An unsupervised approach for authorship attribution in text using topic modeling. IEEE Access, 5, 13962-13971.

15) Mukherjee, S., & Balamuralidhar, P. (2017). Social media analytics: A survey on techniques and applications. IEEE Access, 6, 57826-57851.

16) P. Mishra, R. Chatterjee, "Performance Analysis of Latent Dirichlet Allocation Algorithms for Document Classification." Procedia Computer Science, vol. 54, pp. 423-431, 2015.

17) T. Park and J. E. Martin, "Detecting Binge-Watching Behavior Using Topic Modeling and Sentiment Analysis," in Proceedings of the 12th International Conference on Advances in Computer Entertainment Technology, 2015.

18) R. Fernández-Delgado, M. Cernadas, S. Barro, and D. Amorim, "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?" Journal of Machine Learning Research, vol. 15, pp. 3133-3181, 2014.

19) R. Paul, S. R. Panda, S. S. Satapathy, and A. Abraham, "A Framework for Topic Modeling of Social Media Data for Health Care Applications," in Proceedings of the 4th International Conference on Computational Intelligence in Data Mining, 2015.

20) M. A. Hearst, "Untangling Text Data Mining." in Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, 1999.