**Literature review**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Articles | Objective | Methods | Variables | Results | Research gaps |
| Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey  Hamed Jelodar, Yongli Wang , Chi Yuan , Xia Feng , Xiahui Jiang , Yanchao Li and Liang Zhao | * This paper, investigates highly scholarly articles (between 2003 to 2019) related to topic modeling based on LDA to discover the research development, current trends and intellectual structure of topic modeling. | * This paper investigates scholarly articles (from 2003 to 2019) which are related to Topic Modeling based on LDA to discover the research development, current trends and intellectual structure of topic modeling based on LDA. * This paper also investigates topic modeling applications in various sciences. * The challenges in topic modeling, such as image processing, Visualizing are also discussed * This paper also introduces some of the most famous data and tools in topic modeling. | * LDA * Different sciences used for topic modelling * Different researchers | * Topic models have an important role in computer science for text mining. In Topic modeling, a topic is a list of words that occur in statistically significant methods. A text can be an email, a book chapter, a blog posts, a journal article and any kind of unstructured text. * Topic modeling can provide a useful view of a large collection in terms of the collection as a whole, the individual documents, and the relationships between the documents. | * No researches included from before 2003 |
| Topic Modelling Twitter Data with Latent Dirichlet Allocation Method  Edi Surya Negara,  Dendi Triadi and Ria Andryani | * In this study using the LDA method as an algorithm to produce topic modeling, each topic similarity, and visualization of topic clusters from the tweet data generated as many as 4 topics (Economic, Military, Sports, Technology) in Indonesian, where each topic has a number different tweets | * Data: sport, economic, military and technology related Indonesian tweets * Topic modelling | * Language * Topic modelling method * The amount of clusters | * The LDA method used in the processing of tweet data is successfully carried out and works optimally, in each topic extraction, topic modeling, generating index words that are in each topic cluster and computer visualization in the topic. * LDA output shows optimal performance in the process of word indexing in Sport topics with 1260 tweets with an accuracy of 98% better than the LSI method in Topic Modeling. | * It takes a new breakthrough to be able to create a large corpus / library in Indonesian to process Indonesian text data in the natural science field Language Processing (NLP). |
| Twitter Sentiment Analysis on Coronavirus using Textblob  Chinder Kaur and Anand Sharma | * This paper, analyzes the sentiments regarding coronavirus disease(COVID-19). This is a very critical issue in these days. fetching the twitter streaming tweets related to coronavirus using twitter API and analyze these tweets using machine learning techniques and tools as positive, negative and neutral | * Data: Tweets * Machine learning techniques * Sentiment analysis with the Textblob technique. | * Time * Sentiment(Positive, Negative and Negative) | * Positive tweets percentage 24.0% Negative tweets percentage 32.1% Neutral tweets percentage 43.9% * There is also decreasing time sentiment trend visible | * It has been realized that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis. |
| Bike-sharing: the good, the bad, and the future -an analysis of the public discussion on Twitter  David Duran-Rodas,  Dominic Villeneuve,  Gebhard Wulfhorst | * Due to the dilemma of bike-sharing concerning its benefits and drawbacks, and its unclear future, we focused on a mixed-methods approach to analyze this public discussion through posts or “tweets” from the social media channel Twitter. | * Data : 12,000 tweets were collected in English around the world related to bike-sharing for a period of about six months. * Topic clustering * Sentiment analysis | * Tweets * Complements * Roots | * There were 3.7 times more original tweets including the term “bike–sharing” classified as positive than negative. * Strongly negative tweets focused on poor performance, vandalism, theft, and oversupply of BSS, especially with regard to dockless Asian BSS start-ups with low-quality bikes | * the comparison of tweets regarding BSS to other shared modes such as car-sharing, scooter or ride-sharing lacked. * Interviews with people who weren’t positive about BSS. |
| Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naive Bayes  Rachmawan Adi Laksono and  Kelly Rossa Sungkono | * This study tries to classify Surabaya restaurant customer satisfaction using Naive Bayes and to compare to accuracy with textblob method. | * Data is retrieved from reviews from Tripadvisor. * Data sampling is crawling by using WebHarvy Tools * Naïve Bayes and text blob is used | * Sentiment * The amount of clusters * The 2 different methods | * The result from this research shows that these two methods get the customer response accurately and Naive Bayes method is less accurate than TextBlob sentiment analysis with a different accuracy of 2.9%. | * Further research can be done by increasing the number and variety ofreview data, or by other methods, to increase the value of accuracy. |
| Technological developments and socio-economic issues of wireless mobile communications  [Ronald Beaubrun](https://www.sciencedirect.com/science/article/abs/pii/S0736585300000265" \l "!) and [Samuel Pierre](https://www.sciencedirect.com/science/article/abs/pii/S0736585300000265#!) | * This paper examines the technological developments, as well as the worldwide social-economic impacts of wireless mobile communications. More specifically, it gives an overview on the technological developments of wireless mobile communications, describes the evolution towards the next-generation systems, analyzes the reasons for the growth rate of subscribers, and the related social development. | * General review of the impact of wireless mobile services on society. |  | * There is a increase in time of people using wireless mobile services * This has a positive influence on the economy. Also on the job market, it created new jobs. * Another positive societal effect was the increase in personal safety by this technology. | * There is no analytic reasoning behind this paper. |
| Sentiment Analysis in Social Networks through Topic Modeling  Debashis Naskar , Sidahmed Mokaddem , Miguel Rebollo , Eva Onaindia | * This paper, analyzes the sentiments derived from the conversations that occur in social networks. The goal is to identify the sentiments of the users in the social network through their conversations. | * Data: Social media(Mainly twitter) * ANEW, a lexical dictionary is used to identify affective emotional feelings associated to a message according to the Russell’s model of affection * A topic modeling mechanism called Sent LDA, based on the Latent Dirichlet Allocation (LDA) is used to find the topic distribution in a general conversation * Sentiment analysis |  | * Experimentation showed that topic modeling is very helpful for sentiment classification of twitter messages since different contextual views of sentiments are obtained. We used various levels of sentiments and we observed that the primary sentiment identification is an appropriate level to analyze users’ emotional tendencies. | * No comparison between methods. |
| Topic modeling and sentiment analysis of global climate change tweets  Biraj Dahal ,Sathish A. P. Kumar and Zhenlong Li | * In this study, a large dataset of geotagged tweets containing certain keywords relating to climate change is analyzed using volume analysis and text mining techniques such as topic modeling and sentiment analysis | * Data: Tweets * Latent Dirichlet allocation was applied for topic modeling to infer the diferent topics of discussion * Valence Aware Dictionary and sEntiment Reasoner was applied for sentiment analysis to determine the overall feelings and attitudes found in the dataset | * Location * Users per country * The amount of clusters | * The discussion of climate change on Twitter is negative overall; it seems that users discuss climate change primarily as part of a usually negative reaction to current events. * The only regular peaks in positive sentiment are during warm autumn and winter days. This suggests that as the climate changes over time, Twitter users will speak happily when there are warmer days in the winter but speak more negatively during most other events. | * Not a lot of data of USA twitter users. For the development of mitigation policies, Twitter users in the USA should discuss more about their opinions on policy or potential plans. |
| Topic based Sentiment Analysis for COVID-19 Tweets  Manal Abdulaziz and Mashail Alsolamy | * The paper specifically focuses on twitter and extracts the most discussed topics during and after the first wave of the Coronavirus pandemic. | * The extraction was based on a dataset of English tweets pertinent to COVID-19. The research study focuses on two main periods with the first period starting from March 01,2020 to April 30, 2020 and the second period starting from September 01,2020 to October 31, 2020.’ * The Latent Dirichlet Allocation (LDA) was adopted for topics extraction whereas a lexicon based approach was adopted for sentiment analysis. | * Time * Amount of clusters * Sentiment | * The research findings revealed the appearance of conflicting topics throughout the two Coronavirus pandemic periods. Besides, the expectations and interests of all individuals regarding the various topics were well represented. | * This study has some limitation, it appeared when working with a big data volume, the research spent much time to do that comparing to the time given to finish this research. Furthermore, separating tweets of each day in a file caused long time to collecting them and merge in one dataset |
| Impact of Smartphone’s on Society  Muhammad Sarwar and  Tariq Rahim Soomro | * The intention of this study is to investigate how Smartphone‘s are impacting the society and also how Smartphone‘s are going to transform the culture, social life, technology landscape and other diverse aspects of modern society. The intention of this study is to understand all the positive and negative aspects of Smartphone on the society. | * Impact is measured by using surveys for the topics business, health and learning | * Characteristis of the survey participants | * The key impacts are : to be always-connected, addiction to phone, single device with all required features, business edge, convenient educational features, apps as new technology, entertainment, best utilization of time, disrespectful behaviour, privacy issues, impact on culture, distraction at work & at education. | * A research gap is that this research is only conducted for one country. The results can differ for other countries. |
|  |  |  |  |  |  |

**FLOWCHART:** DATA

Data extraction:

Both companies

Features

Feature extraction

Data Preparation

**FLOWCHART:** TOPIC MODELLING

Evaluation Topic modelling model

Topic of the tweets

**FLOWCHART:** Sentiment analysis

Sentiment Time analysis

Sentiment of the tweet

Evaluation

Sentiment model

**Data extraction:**

Getting the data for the two companies(Xfinity and Mint mobile), by extracting tweets from twitter for time period 01-01-2019 till 01-06-2021.

**Data Preparation:**

Merging the different datasets for each company into one big dataset and saving it as a csv file. Then cleaning the data(only English tweets , no retweets, no duplicates, checking the data on missing data).

**Feature extraction:**

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). So in this case getting rid of the words that are not important for the text analysis. The first step is to tokenize the tweets. The next step is to get rid of the stop words for each tweet.

**Topic modelling &Evaluation:**

After the features are extracted, the data is ready for topic modelling. This topic modelling step requires an evaluation step. In this step the accuracy of the model needs to be tested, by adjusting the amount of clusters and the amount of iterations. This results in a loop in the flowchart. When the optimum performance of the model is reached then the topic per tweet will be available.

**Sentiment analysis &Evaluation:**

After the topic of the tweets are known, the data is ready for sentiment analysis. This step will be done only for the tweets with the topics that are relevant for the thesis (User satisfaction, affordability and willingness). This sentiment analysis step requires an evaluation step. In this step the accuracy of the model needs to be tested. This results in a loop in the flowchart. When the optimum performance of the model is reached then the sentiment per tweet will be available.

**Sentiment Time analysis:**

Now the sentiment per tweet is known, the data is ready for the time analysis. The date column in combination with polarity(sentiment) column will result in the sentiment over time. Now the Covid effect will be clear for each of the different topics (User satisfaction, affordability).

**FLOWCHART:** THESIS(WRITING)

Literature

Review

Introduction

Appendix

References

Further Research

Discussion

Conclusion

Methodology

Theoretical Framework(TM)

**Timeline:**

|  |  |
| --- | --- |
| 31-05-2021-04-06-2021 | *Data collection part 1* |
| 04-06-2021-11-06-2021 | *Data collection part 2 + cleaning + literature review and flowchart* |
| 11-06-2021-18-06-2021 | *Topic modelling + Structure thesis file* |
| 18-06-2021-25-06-2021 | *Sentiment analysis + introduction* |
| 25-06-2021-02-07-2021 | *Time analysis + theoretical framework*  ***Deadline Data analysis*** |
| 02-07-2021-09-07-2021 | *Methodology + Results + checking data analysis on mistakes* |
| 09-07-2021-16-07-2021 | *Methodology + Results + + Rewriting parts after feedback* |
| 16-07-2021-23-07-2021 | *Conclusion, discussion further research , Appendix + Rewriting parts after feedback* |
| 23-07-2021-30-07-2021 | *Lay out , checking all the parts of the thesis , spelling check + Rewriting parts after feedback* |
| 02-08-2021-06-08-2021 | **Deadline thesis** |
| 09-08-2021-13-08-2021 | **Thesis defence** |

**Literature review**

**Topic modelling:**

Several conclusions have been drawn from various papers on topic modeling. These papers range from papers that examine scholarly articles (from 2003 to 2019) which are related to topic modeling to studies that use topic modelling. A summary of the results is presented below:

1. Topic models have an important role in computer science for text mining. Topic modeling can provide a useful view of a large collection in terms of the collection as a whole, the individual documents, and the relationships between the documents.
2. The LDA method used in the processing of tweet data is successfully carried out and works optimally, in each topic extraction, topic modeling, generating index words that are in each topic cluster and computer visualization in the topic.
3. LDA output shows optimal performance in the process of word indexing in topics with 1260 tweets with an accuracy of 98% better than the LSI method in Topic Modeling.
4. Experimentation showed that topic modeling is very helpful for sentiment classification of twitter messages since different contextual views of sentiments are obtained.
5. Another study showed that Topic modelling has some limitation, it appeared when working with a big data volume, the research spent much time to do that comparing to the time given to finish this research.
6. Topic modelling has also some limitations during clustering of data from not well known languages (for example Indonesian tweets).

**Sentiment Analysis**

Several conclusions have been drawn from various papers on topic modeling. These papers range from papers that reviewed on papers that used sentiment analysis to studies that used sentiment analysis for their own project. A summary of the results is shown below:

1. Research shows that both Naïve Bayes and Textblob get the sentiment response accurately but the Naive Bayes method is less accurate than TextBlob sentiment analysis with a different accuracy of 2.9%.
2. Research also shows that the primary sentiment identification is an appropriate level to analyze users’ emotional tendencies.
3. Multiple researches show that sentiment analysis over time can be directly connected to world events or seasonal changes (Winter/ Summer). This validates the results of this analysis.
4. It has been realized that if the neutral sentiments are significantly high, there is a need to improve Twitter sentiment analysis.
5. Another study showed that Sentiment analysis has some limitation, it appeared when working with a big data volume, the research spent much time to do that comparing to the time given to finish this research.
6. Sentiment analysis has also some limitations during clustering of data from not well known languages (for example Indonesian tweets).

**Effect Wireless mobile technology on society:**

The following effects of Wireless mobile services on society and economy are stated in the papers:

1. There is an increase in time of people using wireless mobile services
2. This has a positive influence on the economy. Also on the job market, This market grows and therefore creates new jobs.
3. Another positive societal effect was the increase in personal safety by this technology