Journal of Cyberspace Studies
eISSN: 2588-5502
http://jcss.ut.ac.ir
University of Tehran

# Modelling Procedure while Assessing the Impact of News Articles on Cryptocurrency (Bitcoin) Market Movement

**T.O. Maku[1], Monday Osagie Adenomon[2*], Mary U Adehi[2]**
1. Department of Statistics, Federal University, Otuoke, Nigeria.
2. Department of Statistics, Nasarawa State University, Kefi, Nigeria.
(*Corresponding author: adenomonmo@nsuk.edu.ng, https://orcid.org/0000-0002-9523-8032)

| Article Info | Abstract |
|---|---|

Original Article

Main Object: Business & Economics, Data science, Statistics, computer science

Received: 09 April 2025
Revised: 02 June 2025
Accepted: 02 June 2025
Published online: 13 July 2025

**Keywords:**
Bitcoin,
CNBC's market section website,
LDA,
prediction,
sLDA,
Topic modelling.

**Background:** Cryptocurrencies have a variety of unique qualities, from cutting-edge technology to highly secure architecture. Additionally, the ability to invest in cryptocurrency, as an asset or a function of its prosperity has made crypto-currencies attractive to venture capitalists, computer scientists, and statisticians.
**Aims:** In this study, we concentrated on a collection of documents web-scrapped from the market section of CNBC, where each document is associated with a response variable.
**Methodology:** These documents contain preprocessed words/terms of day-to-day reportage on cryptocurrency (Bitcoin). The corresponding response variables are the daily opening and closing price of Bitcoin prices. The Supervised Latent Dirichlet Allocation(sLDA), a statistical model of labeled documents, was used to analyze the textual data alongside their corresponding response variables, since our study aims to predict the response variable for unlabeled new documents.
**Results:** Hidden Topics with their unique terms from the preprocessed articles were exposed through a Natural language processor. Mean absolute error (MAE), Mean absolute percentage error (MAPE), and Root mean square error (RMSE) graphs were constructed for the sLDA models with 'k = 3,10,20,30,50,75,100 and 200 Topics' values where the model with the best evaluation metric, was selected for prediction purpose.
**Conclusion:** It was discovered that the sLDA model with k = 20. A posterior covariance matrix which shows the proportion of terms from the documents, making up a Topic. Coefficient values were generated in other to graphically visualize how important the discovered topics are and how they affect the market trend. Finally, the prediction of new labels (numeric-decoded closing prices) for the unlabeled documents was done and comparisons were made; the predicted labels follow a similar pattern to that of the time series closing price trend.

**Cite this article:** Maku TO, Adenomon MO, Adehi MU. (????). "Modelling Procedure while Assessing the Impact of News Articles on Cryptocurrency (Bitcoin) Market Movement". *Cyberspace Studies*. ?(?): 1-20. doi: https://doi.org/10.22059/jcss.2025.393177.1139.


Website: https://jcss.ut.ac.ir/| Email: jcss@ut.ac.ir |
EISSN: 2588-5502
Publisher: University of Tehran

## 1. Introduction

Cryptocurrencies have a variety of unique qualities, from cutting-edge technology to highly secure architecture. Additionally, the ability to invest in cryptocurrency, as an asset or a function of its prosperity has made crypto-currencies attractive to venture capitalists, computer scientists, and statisticians. Aside from all these aspects, there are other troubling issues such as the absence of financial institution regulation (Buchholz et al., 2012). Therefore, sentiments and ideologies have an impact on the movement of Bitcoin's price. The way that large-scale Bitcoin owners behave, as well as society norms, political beliefs, and emotional states, all influence the price of Bitcoin. Financial analysts and scholars have become increasingly interested in digital currency because of its widespread acceptance. Financial analysts and researchers enjoy the challenge of market price extrapolation, but the developing cryptocurrency marketplaces lack in-depth study (Yao et al., 2019).

The lack of this in-depth study prompted us to investigate other hidden factors contributing to the market price trend. In clear terms, we identified a factor which is textual crowd-trading-knowledge. Textual crowd-trading-knowledge is believed to have a strong root in articles, especially News articles. Through several mobile applications, those interested in Bitcoin transactions can obtain the latest information about price movements, causes, and news at any time and location. They could use this knowledge to decide whether to buy or sell. It has been shown by numerous studies that a significant portion of the public uses' websites, applications, or social networking sites to receive information in some capacity. According to Kaya and Karsligil (2010), web channels are now the second most important source of information for people who prefer reading news articles over watching or listening to the news. Television is the most important source of information.

The above resource is what we have taken advantage of to analyze and give it a meaning numerically. The way Bitcoin market price trend gets affected by the daily News articles on traders' activities was investigated. Eklund & Bejerholm (2004) postulated that sources of textual data include both independently developed and corporately produced materials. Sources generated by the organization, such as quarterly and yearly reports, can provide a rich language structure; when carefully examined, can forecast future performance.

Our chosen model in analyzing this phenomenon is a supervised Topic model known as the Supervised Latent Dirichlet Allocation (sLDA), simply because, our target is prediction. According to Blei et al. (2017), the sLDA is a variant of the Latent Dirichlet Allocation (LDA).

A supervised topic model is required for document collections that contain response variables where the objective is to predict the response variable given the document in the collection. The Supervised Latent Dirichlet Allocation (sLDA), created by Blei et al. (2017), aims to infer latent themes that predict the response variable. First, they used a publicly available dataset of newspaper movie reviews introduced by Pang & Lee (2008), which contains movie reviews paired with a certain number of stars (ratings). The analysis was treated as a regression problem rather than a classification task. Next, they applied the sLDA to the analysis of two real-world tasks. The second application was to study amendment texts from the 109th and 110th U.S. senates. The discriminating parameters from an ideal-point analysis based on voting history which is utilized in quantitative political science in this case to map senators to a real-valued point on a political spectrum, make up the response variables (Clinton et al., 2004). The authors observed that the sLDA offered a better prediction on all datasets when compared to the results of linear regression and the Least Absolute Shrinkage and Selection Operator (LASSO) regression of the unsupervised LDA model.

Zang and Kjellström (2014) researched on the ways to supervise a topic model. The research suggested two factorized supervised topic models that factorize the topics into signal and noise. They also offered a detailed analysis of the behaviour of supervised topic models using Supervised Latent Dirichlet Allocation (SLDA). The findings of experiments conducted on synthetic and real-world data for computer vision tasks indicate that factorized topic models can improve performance and that increased supervision is necessary for optimal outcomes.

Survival-supervised latent Dirichlet allocation models were developed by Dawson and Kendziorski (2012) for the genomic study of time-to-event outcomes. With the use of their approach, groups of clinical and genomic characteristics that are common to patient subgroups can be efficiently identified, and each patient is then uniquely defined by these qualities. A useful patient subgroup was discovered by applying survLDA to The Cancer Genome At-las (TCGA) ovarian research. These patient subgroups are distinguished by varying propensities to display aberrant mRNA expression and methylations, which correlate to varying rates of survival from first therapy. The study demonstrates how technological developments such as supervised topic modeling continue to improve the ease and precision of measuring the genome and phenome; as a result, genomic-based studies of the disease frequently involve the collection of a wide variety of data types from large patient populations.

Wilcox et al. (2021) researched on Supervised Latent Dirichlet Allocation with Covariates which is a Bayesian Structural and Measurement Model of Text and Covariates. They proposed a novel statistical model, supervised latent Dirichlet allocation with covariates (SLDAX) that jointly incorporates a latent variable measurement model of text and a structural regression model to allow the latent topics and other manifest variables to serve as predictors of an outcome. Using a simulation study with data characteristics consistent with psychological text data, they found that SLDAX estimates were generally more accurate and more efficient. To illustrate the application of SLDAX and a two-stage approach, they provided an empirical clinical application to compare the application of both the two-stage and SLDAX approaches.

Perotte et al. (2011) presented a model for hierarchically and multiple labeled bag-of-word data called hierarchically supervised latent Dirichlet allocation (HSLDA). Their work focused mostly on out-of-sample label prediction, but it was also interesting to see better lower-dimensional representations of the bag-of-word data. Using large-scale data from retail product classification and healthcare document labeling tasks, they showcased HSLDA. They demonstrated that, as compared to models that do not use the structure from hierarchical labels, out-of-sample label prediction is much improved.

Mohan et al. (2019) gathered a significant quantity of time series data and used deep learning models to analyze it in connection to linked news stories, improving the accuracy of stock price predictions. The study shows that there is difficulty in forecasting stock values due to their extremely volatile character, which is influenced by a wide range of political and economic issues, shifts in leadership, investor attitude, and numerous other factors. The study was able to reveal a high association between the movement of stock prices and the release of news stories with the use of topic modeling. The quantity of training data supplied determined how accurate deep learning algorithms were. However, by compiling a sizable amount of time series data and applying deep learning models to analyze it for relevant news stories, they were able to increase the accuracy of stock price predictions.

Yap et al. (2012) used computer methods to forecast stock values using financial data. They relied on intricate mathematical models and historical market data, which was limited to evaluating data that was available to them. Because of this flaw, they were unable to respond to unforeseen circumstances that deviated from previous patterns. Weighted terms were assigned to a new piece by the study's prediction to ascertain its expected direction of movement. To their credit, these more straightforward techniques have demonstrated a limited but significant ability to forecast price direction but not actual price. The popularity of Quantitative funds, or Quants, has grown in the last few years. Quants automatically sort through financial data using numbers and select stocks. These systems vary in the degree of trading control they possess, from basic stock recommenders to transaction executors, despite being built on proprietary technology.

Sharma (2020) compared stock price prediction models using news articles, using global indices, exchange rates, historical stock prices, global news, and technical indicators. They utilized an ensemble of Long Short-Term Memory (LSTM) models. To compare the results, their investigation was also expanded to include some benchmark categorization models. According to their results, an ensemble of three LSTM models predicted growing and falling trends equally well and steadily, yielding an accuracy of 60% with the best recall and true negative score.

Sahut et al. (2024) evaluated five models based on cutting-edge machine-learning techniques. These models were picked from the literature on crude oil forecasting to evaluate the impact of news-based sentiment on crude oil price prediction. The COVID-19 pandemic era and the years from 1990 to the start of the pandemic were included in the results for each approach. This made it possible for them to investigate how news-based mood functions in various stages of economic growth and disaster. Compared to other eras, such as the 2008–2009 financial crisis, a notable impact of news-based sentiment was seen on the forecasting performance of machine learning techniques during the Covid-19 period.

Loughran et al. (2019) investigated if investors respond to news stories about oil that disclose supply and demand information in a timely and logical manner. To facilitate investors' and researchers' assessment of the information value of oil tales, their study generated a unique keyword list of 130 terms and modifiers connected to the oil industry. They discovered a notable overreaction in the short run to the news stories on oil connected to the Dow Jones. Lower oil prices the next trading day were linked to phrases in delayed news items like output cut, production reduced, scarcity, and demand rising. The data supports the theory that oil traders exaggerate their reactions to news articles that are extensively read.

The major challenge of using structured textual data to predict traders’ interest is because of its extremely volatile character of such data, which is influenced by a wide range of political and economic issues, shifts in leadership, investor attitude, and numerous other factors (Mohan et al., 2019; Fataliyev et al., 2021).

The baseline models for textual data are social media content, News content, official company announcements, traditional textual representation techniques, deep learning based advanced NLP techniques, statistical models, machine learning techniques and deep learning models (Frank et al., 2017; Li et al., 2018; Shah et al., 2019; Kumar et al., 2020; Thakkar & Chaudhari, 2021; Fataliyev et al., 2021).

Hence, our suggested text-feature extraction of the sLDA a supervised Latent Dirichlet Allocation aims to predict traders’ interest in cryptocurrency and raise traders' awareness when they rely on news stories for trading information.

This paper contributes to existing literatures on the potentials and capabilities of sLDA (supervised Latent Dirichlet Allocation) to predict traders’ interest in cryptocurrency.

## 2. Materials and Method

News articles on cryptocurrency-related activities published in foreign media between 2016 and 2022 were the research population's focus. The Consumer News and Business Channel is where these news stories are found (CNBC). Because Bitcoin is so popular compared to other cryptocurrencies, it will be the only cryptocurrency sampled in this study.

As secondary data, the corpus of news items serves as the data set for this study. From 2016 until 2022, every one of the more than 6,000 news pieces or articles was written in English. Leveraging a custom Python script called "Beautiful Soap" created with the Jupyter Notebook, quick scrapping of the text data was accomplished by leveraging relevant meta-data from the previously mentioned source. The following meta-data was included in the pages, which were kept in a comma-separated (CSV) format: (i) article summary, (ii) article section, (iii) article link, (iv) article date, (v) article summary, (vi) article body, (vii) opening price, (viii) closing price.

We used the query "Daily Bitcoin reports". Furthermore, we made use of the articles' corresponding closing prices as our response variable.

The bag-of-words document representation is assumed by topic models (Blei et al., 2017). Each document is represented here as a bag of its terms/words, with no respect for word order or grammar. Many Natural Language Processing and Information Retrieval algorithms use this simplified model. The NLP stage consisted primarily of four broad steps: (1) loading the input data (News articles), (2) pre-processing the data, (3) transforming texts into bag-of-words vectors, and (4) training the sLDA models (Mckinney, 2010).

The news items (which will now be referred to as documents) cannot be simply fed into the model as raw data or free-text but must instead be transformed into a suitable form for the modeling framework. Normalization, tokenization, stemming/lemmatization, and stop-word removal are common text data pre-processing techniques. Following the collection and collation of articles, the text will be pre-processed in Python using the SpaCy, Gensim, and Pandas modules (Mckinney, 2010). Pre-processing is required before performing NLP on the text. The texts of the articles were subsequently normalized by making them lowercase. Then, word elongations and foreign characters that weren't words, such as punctuation and other non-ASCII characters, were eliminated. Next, non-informative stop-words that regularly occur, such as "the", "is", "I", and "did" were eliminated (using stop-words provided in the genism module of Python). Following that, token words were lemmatized in Python using the genism module. Lemmatization is a type of text normalization that involves grouping inflected forms of words into their base or dictionary root terms, known as lemma. For example, lemmatizing the terms 'trouble', 'troubling', and 'troubled' yields the lemma 'trouble'. The traditional stemming of tokens will be avoided, based on Schofield et al. (2017)'s recommendation that topic coherence is rarely enhanced between the pre-stemming and post-steaming Topic models. Finally, whitespaces were removed to make the document more compact. Documents containing fewer than 50 words were eliminated. Words that appeared in less than 70% of the corpus were pruned as well.

**2.1. Specifications and Estimation of the sLDA**

For response-document pairs, the supervised latent Dirichlet allocation model (sLDA) shows better strength in executing such a plan.

Distributions over document collections are known as topic models, in which each document is represented by a set of discrete random variables, $W_{1:n}$, which are its words. The words in a document are treated as emerging from a set of latent themes in topic models, which are a set of unknown distributions over the vocabulary. Each document in a corpus uses a mixture of subjects with topic proportions that are specific to it, but all documents in the corpus share the same K topics. Different from traditional document mixing models, which link every document to a single, unidentified topic, topic models isolate each document. Erosheva et al. (2004) describe them as mixed-membership models. Each document has a corresponding response as covariates which are jointly modelled for prediction when determining labels for unlabeled new documents. By allowing a response to be associated with each document and jointly modeling the response variable and the corpus of documents, the LDA model is extended to a supervised learning environment using Supervised Latent Dirichlet Allocation (sLDA). According to Blei et al. (2017), this enables it to identify the latent topics that are most predictive of the response variables in the training set and even to make predictions about future unlabeled documents. Following the LDA model's notation, let y represent a response variable from a generalized linear model with parameters η and $\delta$. Should we consider the subsequent fixed;$\beta_{1:K}$: the k topics with each $\beta_k$ a vector of term probability, η and $\delta$ and the Dirichlet hyperparameter for the per-document topic proportion θ. For every document and response variable, the generative process assumed by the sLDA is as follows:

1. Draw topic proportion, $\theta |\ \alpha \sim Di(\alpha)$;
2. for each word,
   a) Draw a topic assignment $Z_n \mid \theta \sim Mult(\theta)$
   b) Draw word $w_n \mid Z_n, \beta_{1:K} \sim Mult(\beta_{z_n})$;
3. Draw a response variable $y|\ z_{1:N}, \eta, \delta \sim \text{GLM}\ (\bar{z}, \eta, \delta\ )$, where $\bar{z} = \frac{1}{N}\sum_{n=1}^{N} z_n$.

For sLDA, the Dirichlet distribution (Di) is used as a prior for the topic distributions. Each document is assumed to have a distribution over topics drawn from a Dirichlet distribution. This helps in capturing the variability in topic proportions across different documents. Each topic is represented by a multinomial distribution over the vocabulary, indicating the probability of each word appearing in that topic. This allows the model to not only discover Topics but also predict the response variable based on the document's topic distribution.

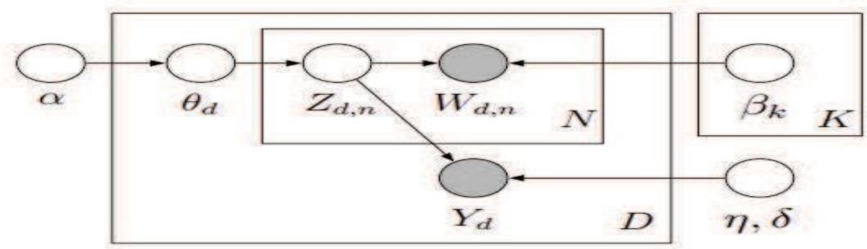


**Figure 1.** Graphical Model representation of Supervised Latent Dirichlet Allocation (sLDA)

### 2.2. Fitting the model

The method used for fitting the model is the variational expectation-maximization (VEM) algorithm. The sLDA model was equipped with parameters such as the "Document-Term matrix", "K" (the number of topics), "vocab" (vocabulary words associated with word indices used in documents), "e.iteration", "m.iteration", "alpha", "eta", "var" (variance of the response variable), "annotation" (response variable) and various parameters was used to adjust the speed of convergence of the algorithm. Though there are no standards for the choice of parameters the parameters were used to set the convergence tolerance for the variance and E-M algorithms, respectively. Also, the maximum number of iterations for the conjugate gradient algorithm was set by one of the parameters which iterates between the E-step and M-step in a bid to maximize the likelihood of the corpus. The procedure finds the maximum bound about the latent variables (the topic proportions and the topic assignments Z) in the E-step, and the M-step, finds the maximum bound for the model parameters (the topics and the multivariate normal parameters).

Application of variational EM was done until the relative change in the likelihood bound was achieved. The iterations convergence ranged between 1hr to 6hrs 45minutes on a core i7 HP laptop of 2.7GHz with 8 GB RAM for each of the K values.

A data split was done on the 5000 plus preprocessed documents/articles in the proportion of 70% training data and 30% test data. The training data set was used to train our sLDA model while the test data set was used as a test for prediction of our response variable. The response variable is the Timeseries closing price of Bitcoin corresponding to the documents. Additionally, we categorized our response variable into a numeric classification of "low= 1", "fairly low= 2", "fairly high= 3" and "high= 4" with a threshold of "less than or equal to 10000", "less than or equal to 20000", "less than or equal to 40000" and "less than or equal to 60000", respectively.

Using the subsequent initializations: $\eta$ to a grid on [-1, 1] with increments of $1/K$, and $\beta_k$ to randomly perturbed uniform topics. Different sLDA models with varying numbers of topics as 3, 10, 20, 30, 50, 75,100 and 200 were trained so that the variational EM method was executed for the per-document ELBO at the E-steps as well as until a relative change in the corpus-like likelihood bound was less than 0.021% (Blei et al., 2017).

## 3. Result and Discussion

The experiment below was conducted to assess the quality of prediction using the sLDA model. We carried out a sensitivity analysis juxtaposing a latent Dirichlet allocation (LDA) with a regression model with our proposed Supervised latent Dirichlet allocation (sLDA) model. This experiment was carried out using two datasets of an equal number of documents. One dataset was from the CNBC market section (as stated in section 2), and the other was from the Bitcoinist website. Both datasets went through the same preprocessing stages (as seen in section 2). We will refer to the CNBC dataset as 'Dataset(1)' and the Bitcoinist dataset as 'Dataset(2)'.

We compared the behaviors of sLDA and LDA + Regression models while trying to investigate the extrapolative strengths of the two models on the two different datasets. The LDA on its own cannot predict the Bitcoin price category. The usual practice when LDA is to be used for prediction has always been joint modeling of the LDA with a regression model (Clinton et al., 2004; Blei et al., 2017). To better understand the experiment, Figures 2 and 3 show some metric evaluations used to adjudge the better model in the topic range of 3 to 200. This explains how our proposed model (sLDA) for prediction purposes behaved against the conventional model (LDA+ regression) for prediction. In a quest for the optimal model between the LDA and the sLDA models, model evaluation metrics were employed to pick an optimal model that best fits the prediction task. The choice of our model evaluation was born out of the type of classification we used, i.e. the multinomial classification techniques. The evaluation metrics are Mean absolute Error (MAE), Mean absolute percentage Error (MAPE) Root mean squared Error (RSME), and the coefficient of Determination ($R^2$ ).

As seen in Figure 2, the sLDA model across the 'K' values, performed better with K=20 and K= 30 from Figure 3. Likewise, from Figure 2, the LDA + regression model across the 'K' values, performed better with K= 75 and K=100. Figure 3 shows that the sLDA model with K=20 has the lowest MAE, MAPE, and RSME values but has the highest $R^2$ value when compared with the LDA model. Also, the sLDA model with K=30 has the lowest MAE, MAPE, and RSME values and the highest $R^2$ value when compared with the LDA model. The result in Figure 2 shows that LDA at K=75 has an $R^2$ value of 0.3246 while sLDA at K= 20 has an $R^2$ value of 0.6262. The sLDA model performance has an improvement of 90% when we consider the $R^2$ values for both models. From Figure 3, LDA at k =100 has an $R^2$ value of 0.4462 while sLDA at K= 30 has an $R^2$ value of 0.736. The sLDA model performance also had an improvement of 64%. According to Zang and Kjellström (2014), the improvement in classification results

*Cyberspace Studies*, Vol ?, No ?, ????

is always significant through different noise levels when the number of topics is small.

The subsequent subsections show the practical application of the above experiment using ‘Dataset(1)’.

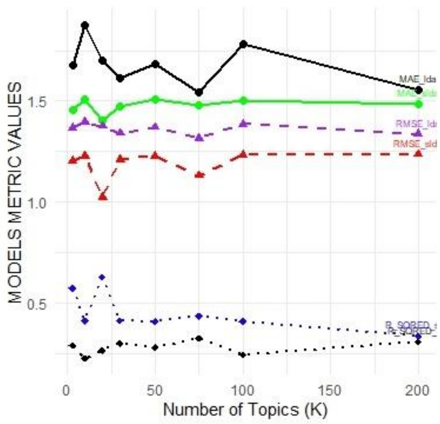


**Figure 2.** sLDA vs LDA behaviour with Dataset(1)

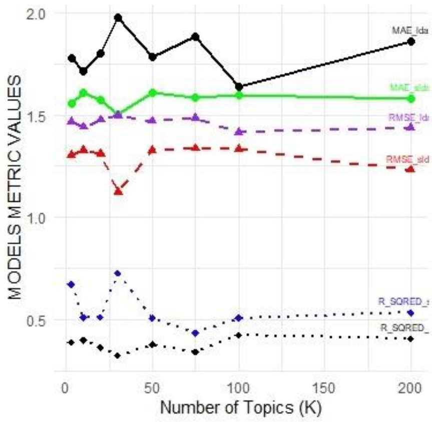


**Figure 3.** sLDA vs LDA behaviour with Dataset(2)

**Table 1.** Tabular summary of the split corpus

| Number of documents | Total word count | Total number of unique words |
|---|---|---|
| 4073(training data)<br>1746(testing data) | 1,151,424<br>494,535 | 17,330<br>15,591 |

### 3.1. sLDA model evaluation

In a quest for the optimal sLDA model, model evaluation metrics were employed to pick an optimal model that best fits our prediction task on the first dataset. The evaluation metrics are Mean absolute error, mean absolute percentage error, and root mean squared error.

Table 2 shows the sLDA model with its numeric evaluation values was picked. A good look across the table shows that the sLDA model with K= 3 has the lowest numeric values across the three metrics but choosing that of K= 3 may lead to underfitting our model. We settled for the sLDA model with the next lowest numeric value, which is the K= 20 sLDA model, as the optimal model.

Exploring Figure 4, gives an insight into the topics that are predictive of the reported News from each of the documents. Our Topics ranges from Minner’s investment, Crypto ransom, Economic policies, Scandals, exchange rates etc. We will briefly discuss the first four Topics. Topic 1 depicts simply, talked about the community of bitcoin miners in Texas who thinks that investing in the energy sector can aid in resolving the state's (Texas) power grid issues. This news was heavily reported by the media and appeared in some of the web scraped documents at different dates (Figure 4).

**Table 2.** Table showing numeric values of three sLDA model evaluation metric

| K | MAE | RMSE | MAPE |
|---|---|---|---|
| 3 | 1.05841 | 1.202434 | 0.6576 |
| 10 | 1.11 | 1.228 | 0.72566 |
| 20 | 1.070072 | 1.223507 | 0.639 |
| 30 | 1.076113 | 1.220401 | 0.6587 |
| 50 | 1.11068 | 1.227168 | 0.7296 |
| 75 | 1.073 | 1.231 | 0.645 |
| 100 | 1.11445 | 1.2332 | 0.7268 |
| 200 | 1.082776 | 1.235816 | 0.6508 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Topic 1 | mining | bitcoin | miner | power | energy | mine | state | electricity | cost | texas |
| Topic 2 | attack | account | hacker | security | data | wallet | computer | criminal | informatic | ransomwa |
| Topic 3 | crypto | cryptocurr | token | musk | digital | dogecoin | ether | asset | market | tesla |
| Topic 4 | china | country | chinese | governme | bank | south | korea | yuan | world | north |
| Topic 5 | stock | percent | market | year | growth | price | investor | average | analyst | equity |
| Topic 6 | blockchair | technolog | network | transactio | token | ripple | ledger | project | applicatio | coin |
| Topic 7 | card | credit | account | income | gain | cash | inflation | youre | return | rate |
| Topic 8 | president | trump | house | russia | bill | senate | russian | governme | state | federal |
| Topic 9 | exchange | security | company | million | commissic | offering | agency | comment | customer | statement |
| Topic 10 | dollar | rate | inflation | bank | market | index | week | high | yield | reserve |
| Topic 11 | bank | currency | digital | financial | central | cryptocurr | payment | regulator | regulatior | libra |
| Topic 12 | investor | fund | asset | investmer | bitcoin | investing | market | invest | financial | advisor |
| Topic 13 | people | money | student | home | life | woman | family | inflation | school | university |
| Topic 14 | long | trader | kelly | brian | short | fast | call | buyer | spread | adami |
| Topic 15 | bitcoin | currency | bitcoins | gold | value | digital | transactio | price | read | virtual |
| Topic 16 | bitcoin | price | cryptocurr | percent | trading | exchange | digital | currency | market | according |
| Topic 17 | company | coinbase | million | square | venture | startup | billion | customer | product | business |
| Topic 18 | inflation | going | money | cramer | really | thats | thing | want | people | know |
| Topic 19 | share | stock | premarke | revenue | cnbc | earnings | company | quarter | tesla | estimate |
| Topic 20 | buffett | apple | read | berkshire | warren | google | amazon | food | book | billionaire |

**Figure 4.** The twenty probable (20) Topics with Ten (10) words/terms from the sLDA model with K= 20

Topic 2 is all about (crypto security) such as ransomware attacks posing a severe risk, thus companies in the financial services, healthcare & pharmaceuticals, and employee personally identifiable information (PII) and HR sectors need to implement strong security measures to safeguard their data and systems. The news was also reported by the media and appeared in some of the web scraped documents at different dates, either by reference to it or just a few mentions.

Topic 3 shows deep thought about Bitcoin’s market competition amidst other cryptocurrencies. That is, despite its humorous beginnings, Dogecoin has grown into an intriguing aspect of the cryptocurrency world, mostly because of Elon Musk's lighthearted engagement. Observing Musk's tweets as the market moves forward may be just as important as examining technical charts.

Topic 4 is about how some countries’ economic circumstances affect the market trend of Bitcoin. All the topics had words that were either heavily mention, partially mentioned or not mentioned at all in the news encapsulated in our various published documents.

*Cyberspace Studies*, Vol ?, No ?, ????

Figure 5 explains how important the topics are to the documents. Topic 3, shows a high level of importance to the documents while Topic 20, shows the lowest level of importance to the documents.

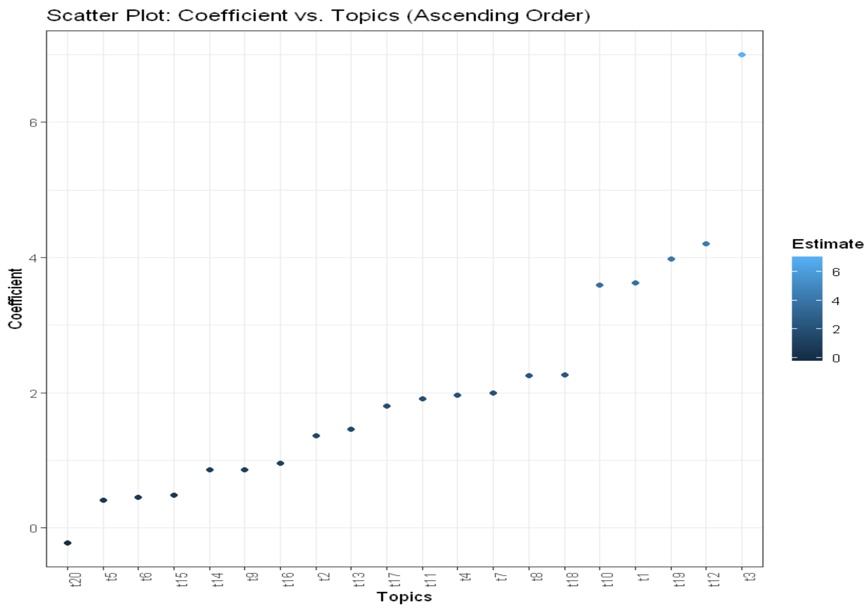


**Figure 5.** Graph of the coefficients of the sLDA model's probable Topics

Table 3 is a 4073 (row) by 20 (column) matrix which shows the proportions of terms/words from the documents/articles that make up the sLDA Topics. It is a posterior covariance matrix. However, because of the volume of this result and for clarity's sake, just the first eight Topics and the first Fourteen documents were selected and displayed in Table 4, while the full information can be seen in the appendix.

A critical look into the first four documents, Table 4 shows how the topic-proportion in the News known as a document and the response of the market price trend, 24 hours after publication. With reference to Table 4, Topic 1, 6, 7 and 8 from Document 0 had its impact on the market trend while Topic 2, 3, 4, 5 had their proportional impact on the market trend from Document 0. Topic 6 had no impact on the market trend from Document 1 while Topic 1, 2, 3, 4, 5, 7 and 8 had their proportional impact on the market trend from Document 1. All the Topics had their proportional effects on the market trend from Document 2. Also, all the topics had their proportional effect on the market trend from document 3 except for Topic 5. The effects were either positive or negative change in the market trend. While there are many factors that can cause price fluctuations in financial markets, the function of the media, is a consistent source of information and sentiment.

**Table 3.** The sLDA model per topic-document proportion output

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 |
|---|---|---|---|---|---|---|---|---|
| Doc 0 | 0 | 0.011905 | 0.02381 | 0.011905 | 0.154762 | 0 | 0 | 0 |
| Doc 1 | 0.013699 | 0.034247 | 0.013699 | 0.136986 | 0.068493 | 0 | 0.006849 | 0.020548 |
| Doc 2 | 0.008 | 0.054 | 0.002 | 0.02 | 0.022 | 0.286 | 0.018 | 0.008 |
| Doc 3 | 0.003236 | 0.038835 | 0.074434 | 0.116505 | 0 | 0.223301 | 0.029126 | 0.016181 |
| Doc 4 | 0 | 0.00463 | 0.125 | 0.023148 | 0.101852 | 0.041667 | 0 | 0.013889 |
| Doc 5 | 0.048421 | 0.065263 | 0.008421 | 0.012632 | 0.176842 | 0.002105 | 0.027368 | 0 |
| Doc 6 | 0.021277 | 0.035461 | 0.014184 | 0.070922 | 0.060284 | 0.049645 | 0.035461 | 0.053191 |
| Doc 7 | 0 | 0 | 0.22 | 0 | 0.28 | 0.02 | 0 | 0 |
| Doc 8 | 0.601027 | 0.005137 | 0.02226 | 0.119863 | 0.025685 | 0.003425 | 0.001712 | 0.035959 |
| Doc 9 | 0.008086 | 0 | 0.040431 | 0.016173 | 0.040431 | 0.002695 | 0.016173 | 0.105121 |
| Doc 10 | 0.012346 | 0.018519 | 0.074074 | 0.018519 | 0.098765 | 0.006173 | 0 | 0.030864 |
| Doc 11 | 0 | 0.018127 | 0.003021 | 0.039275 | 0.009063 | 0.069486 | 0.012085 | 0.003021 |
| Doc 12 | 0.002899 | 0.005797 | 0.023188 | 0.156522 | 0.034783 | 0.008696 | 0.017391 | 0 |
| Doc 13 | 0 | 0.003876 | 0.003876 | 0 | 0.027132 | 0.003876 | 0.023256 | 0.027132 |
| Doc 14 | 0.009479 | 0.009479 | 0 | 0.004739 | 0.530806 | 0.080569 | 0.018957 | 0.004739 |

**Table 4.** The price market trend for the few selected published documents (see Appendix)

| Doc | Date published | open | High | Low | Close | Volume | % Change |
|---|---|---|---|---|---|---|---|
| Doc 0 | 28/8/2022 | 20032.73 | 20133.36 | 19606 | 19606 | 4.16E+10 | -2.13% |
| Doc1 | 24/8/2022 | 21513.00 | 21826.33 | 21172.0 | 21378.1 | 5.06E+10 | -0.63% |
| Doc 2 | 25/8/2022 | 21367.72 | 21773.15 | 21330.7 | 21591.28 | 5.04E+10 | 1.05% |
| Doc 3 | 18/8/2022 | 23338.393 | 23575.61 | 23140.4 | 23206.1 | 5.1E+10 | -0.57% |

Figure 6 is a graph of an in-sample forecast which we did to predict response variables for a Document; i.e., a document could be read and inferences made as to how it is going to affect the market price. This could be made useful by a would-be miner or cryptocurrency analysts. The sLDA prediction model was tested using the testing data which is assumed to be unlabeled documents/articles. The predicted classification in line with the earlier mentioned classification in section (2.2), was juxtaposed with the testing data's initial response variable (classification). For the case of our voluminous data, 200 documents of the testing data and its response variable, alongside the corresponding predicted response, were captured. The result shows that the two trends appear in a near-similar pattern.

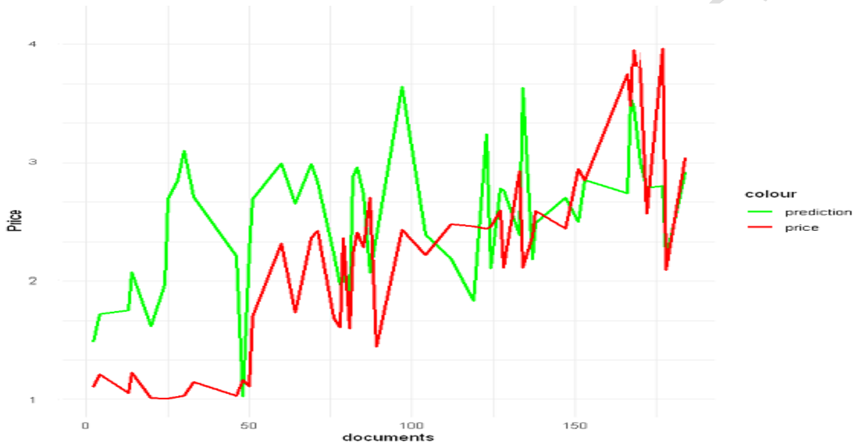


**Figure 6.** The initial price classification and the sLDA predicted classification

## 4. Conclusion

A helpful technique for examining huge text corpora is topic modeling. Topic modeling has applications outside of text mining, linguistics, and language modeling; it has been utilized in computer vision, population genetics analysis, social sciences, and humanities research.

Topic modeling, by revealing the topic structure buried in the text collection, can offer a higher-level exploration strategy to have a solid grasp of big text corpora when combined with other text-mining and machine-learning technologies. Additionally, the outputs of topic models can strengthen tasks like information retrieval, collaborative filtering, classification/categorization, and recommendation systems.

To understand how latent crowd knowledge affects the bitcoin market and to further leverage its predictive power, the research has looked at the significance and use of the sLDA-VEM and sLDA-predict for textual data mining and analysis. Hidden predictive topics used by traders in our textual data were exposed. Similarly, the links between Documents/ articles and topics were disclosed. Finally, we investigated the sLDA model's predictive power by extrapolating traders' interest or response variables for unlabeled documents.

## Conflict of interest

The authors declared no conflicts of interest.

## Ethical considerations

The authors have completely considered ethical issues, including informed consent, plagiarism, data fabrication, misconduct, and/or falsification, double publication and/or redundancy, submission, etc. This article was not authored by artificial intelligence.

## Authors Contributions

TM, MOA and MUA: Developed the conceptual framework; TM-Source for the data and analyze the data; MOA and MUA-Supervised the work; TM-Came up with the first draft; TM, MOA and MUA- came up with the final draft; MOA and MUA- Do the editing of the workData availability.

## Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

## Acknowledgements

We wish to thank the effort of the reviewers and the Editor.

# Appendix 1. The sLDA model Per Topic-Document proportion output (doc0-doc101

| | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 | Topic 11 | Topic 12 | Topic 13 | Topic 14 | Topic 15 | Topic 16 | Topic 17 | Topic 18 | Topic 19 | Topic 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| doc 0 | 0 | 0.011905 | 0.02381 | 0.011905 | 0.154762 | 0 | 0 | 0 | 0 | 0.095238 | 0 | 0.083333 | 0.035714 | 0 | 0 | 0.214286 | 0 | 0.130952 | 0.238095 | 0 |
| doc 1 | 0.013699 | 0.034247 | 0.013699 | 0.136986 | 0.068493 | 0 | 0.006849 | 0.020548 | 0.061644 | 0.006849 | 0.041096 | 0.027397 | 0 | 0.041096 | 0.089041 | 0.363014 | 0 | 0.041096 | 0.034247 | 0 |
| doc 2 | 0.008 | 0.054 | 0.002 | 0.02 | 0.022 | 0.286 | 0.018 | 0.008 | 0.006 | 0.06 | 0.042 | 0.002 | 0.388 | 0.006 | 0.046 | 0.008 | 0.006 | 0.004 | 0 | 0.014 |
| doc 3 | 0.003236 | 0.038835 | 0.074434 | 0.116505 | 0 | 0.223301 | 0.029126 | 0.016181 | 0.12945 | 0.019417 | 0.110032 | 0.035599 | 0.006472 | 0.003236 | 0 | 0.110032 | 0.058252 | 0.022654 | 0 | 0.003236 |
| doc 4 | 0 | 0.00463 | 0.125 | 0.023148 | 0.101852 | 0.041667 | 0 | 0.013889 | 0.037037 | 0.00463 | 0.00463 | 0.064815 | 0.023148 | 0.00463 | 0.074074 | 0.337963 | 0.009259 | 0.12037 | 0.00463 | 0.00463 |
| doc 5 | 0.048421 | 0.065263 | 0.008421 | 0.012632 | 0.176842 | 0.002105 | 0.027368 | 0 | 0.002105 | 0 | 0 | 0.008421 | 0.018947 | 0 | 0.008421 | 0.006316 | 0.035789 | 0.416842 | 0.109474 | 0.052632 |
| doc 6 | 0.021277 | 0.035461 | 0.014184 | 0.070922 | 0.060284 | 0.049645 | 0.035461 | 0.053191 | 0.028369 | 0.010638 | 0.439716 | 0 | 0 | 0 | 0.099291 | 0.053191 | 0 | 0.010638 | 0.01773 | 0 |
| doc 7 | 0 | 0 | 0.22 | 0 | 0.28 | 0.02 | 0 | 0 | 0 | 0.02 | 0.04 | 0 | 0 | 0.02 | 0.02 | 0.32 | 0 | 0 | 0.06 | 0 |
| doc 8 | 0.601027 | 0.005137 | 0.02226 | 0.119863 | 0.025685 | 0.003425 | 0.001712 | 0.035959 | 0 | 0.063356 | 0 | 0 | 0.035959 | 0.001712 | 0.025685 | 0.027397 | 0.001712 | 0.02226 | 0.001712 | 0.005137 |
| doc 9 | 0.008086 | 0 | 0.040431 | 0.016173 | 0.040431 | 0.002695 | 0.016173 | 0.105121 | 0.070081 | 0.221024 | 0.013477 | 0.002695 | 0.008086 | 0.002695 | 0.002695 | 0.110512 | 0.010782 | 0.005391 | 0.312668 | 0.010782 |
| doc 10 | 0.012346 | 0.018519 | 0.074074 | 0.018519 | 0.098765 | 0.006173 | 0 | 0.030864 | 0.006173 | 0 | 0.006173 | 0.006173 | 0.018519 | 0.006173 | 0.320988 | 0.074074 | 0 | 0.117284 | 0 | 0.185185 |
| doc 11 | 0 | 0.018127 | 0.003021 | 0.039275 | 0.009063 | 0.069486 | 0.012085 | 0.003021 | 0.012085 | 0.018127 | 0.181269 | 0.015106 | 0.006042 | 0 | 0.084592 | 0.453172 | 0.003021 | 0.021148 | 0 | 0.05136 |
| doc 12 | 0.002899 | 0.005797 | 0.023188 | 0.156522 | 0.034783 | 0.008696 | 0.017391 | 0 | 0.005797 | 0.002899 | 0.002899 | 0.055072 | 0 | 0.002899 | 0 | 0.605797 | 0.005797 | 0.06087 | 0.002899 | 0.005797 |
| doc 13 | 0 | 0.003876 | 0.003876 | 0 | 0.027132 | 0.003876 | 0.023256 | 0.027132 | 0.069767 | 0.015504 | 0.042636 | 0.341085 | 0.015504 | 0.015504 | 0.104651 | 0.267442 | 0.01938 | 0.003876 | 0 | 0.015504 |
| doc 14 | 0.009479 | 0.009479 | 0 | 0.004739 | 0.530806 | 0.080569 | 0.018957 | 0.004739 | 0 | 0.037915 | 0 | 0 | 0.004739 | 0 | 0 | 0.033175 | 0.004739 | 0.180095 | 0.052133 | 0.028436 |
| doc 15 | 0 | 0.014493 | 0.007246 | 0.021739 | 0 | 0.014493 | 0.007246 | 0.188406 | 0 | 0.5 | 0.014493 | 0 | 0.007246 | 0 | 0.014493 | 0.036232 | 0 | 0.15942 | 0 | 0.014493 |
| doc 16 | 0.010101 | 0.020202 | 0.020202 | 0.010101 | 0.010101 | 0.010101 | 0.050505 | 0.050505 | 0 | 0.010101 | 0.090909 | 0.020202 | 0.010101 | 0.020202 | 0.10101 | 0.191919 | 0.30303 | 0.070707 | 0 | 0 |
| doc 17 | 0.00692 | 0.00346 | 0 | 0.013841 | 0.287197 | 0 | 0 | 0 | 0 | 0.256055 | 0.034602 | 0.027682 | 0.038062 | 0 | 0.058824 | 0.027682 | 0.017301 | 0.183391 | 0.038062 | 0.00692 |
| doc 18 | 0 | 0 | 0.048077 | 0 | 0.019231 | 0.028846 | 0.009615 | 0 | 0.346154 | 0 | 0 | 0.076923 | 0.019231 | 0.009615 | 0.173077 | 0 | 0.105769 | 0.048077 | 0.019231 | 0.096154 |
| doc 19 | 0.006849 | 0 | 0.205479 | 0 | 0 | 0 | 0.013699 | 0.020548 | 0.431507 | 0 | 0.020548 | 0.006849 | 0 | 0 | 0.013699 | 0.006849 | 0.006849 | 0.027397 | 0.232877 | 0.006849 |
| doc 20 | 0.022346 | 0 | 0 | 0 | 0.072626 | 0.005587 | 0.03352 | 0.145251 | 0 | 0.03352 | 0.340782 | 0.027933 | 0.005587 | 0.005587 | 0.039106 | 0.122905 | 0.01676 | 0.122905 | 0 | 0.005587 |
| doc 21 | 0.013889 | 0.023148 | 0.037037 | 0 | 0.00463 | 0.018519 | 0 | 0.00463 | 0.226852 | 0.013889 | 0.097222 | 0 | 0.041667 | 0.013889 | 0.027778 | 0.384259 | 0.009259 | 0.032407 | 0.037037 | 0.013889 |
| doc 22 | 0.007117 | 0 | 0.071174 | 0.181495 | 0.003559 | 0.032028 | 0.007117 | 0.003559 | 0.010676 | 0 | 0.330961 | 0.071174 | 0 | 0.003559 | 0.014235 | 0.199288 | 0.021352 | 0.039146 | 0 | 0.003559 |
| doc 23 | 0 | 0.077586 | 0.008621 | 0 | 0.012931 | 0.017241 | 0 | 0.008621 | 0.517241 | 0.017241 | 0.017241 | 0.099138 | 0.025862 | 0.00431 | 0.168103 | 0 | 0.00431 | 0.012931 | 0.00431 | 0.00431 |
| doc 24 | 0.000866 | 0.019913 | 0.017316 | 0.049351 | 0.057143 | 0.35671 | 0.050216 | 0.001732 | 0.045022 | 0 | 0.117749 | 0.01039 | 0.021645 | 0 | 0.018182 | 0.038095 | 0.08658 | 0.104762 | 0.002597 | 0.001732 |
| doc 25 | 0.022222 | 0.066667 | 0.007407 | 0.02716 | 0.019753 | 0.004938 | 0.012346 | 0.116049 | 0.083951 | 0.037037 | 0 | 0.002469 | 0 | 0.002469 | 0.017284 | 0.009877 | 0.002469 | 0.024691 | 0.516049 | 0.02716 |
| doc 26 | 0.004082 | 0.008163 | 0 | 0.028571 | 0.004082 | 0.008163 | 0.012245 | 0.073469 | 0.004082 | 0.730612 | 0.008163 | 0.004082 | 0.008163 | 0.004082 | 0 | 0.004082 | 0 | 0.085714 | 0 | 0.012245 |
| doc 27 | 0 | 0 | 0.002632 | 0.013158 | 0.005263 | 0 | 0.007895 | 0.013158 | 0.002632 | 0.036842 | 0.013158 | 0 | 0.010526 | 0.010526 | 0.015789 | 0.018421 | 0.039474 | 0.042105 | 0.747368 | 0.021053 |
| doc 28 | 0.003125 | 0.003125 | 0.025 | 0 | 0.028125 | 0.025 | 0.021875 | 0.003125 | 0.225 | 0.034375 | 0.1375 | 0.059375 | 0.015625 | 0 | 0.025 | 0.3125 | 0.034375 | 0.040625 | 0.003125 | 0.003125 |
| doc 29 | 0 | 0.009569 | 0.009569 | 0 | 0.004785 | 0.181818 | 0.033493 | 0 | 0.004785 | 0 | 0.009569 | 0.062201 | 0 | 0.004785 | 0 | 0.148325 | 0.009569 | 0.07177 | 0.028708 | 0.421053 |
| doc 30 | 0.002188 | 0.015317 | 0.002188 | 0 | 0.013129 | 0 | 0.875274 | 0.013129 | 0.006565 | 0.004376 | 0 | 0.006565 | 0 | 0.010941 | 0.021882 | 0.002188 | 0.008753 | 0.015317 | 0 | 0.002188 |
| doc 31 | 0.005848 | 0.005848 | 0.017544 | 0.023392 | 0.064327 | 0.017544 | 0.017544 | 0.011696 | 0.023392 | 0.02924 | 0.017544 | 0.005848 | 0 | 0 | 0.023392 | 0.707602 | 0 | 0.005848 | 0.023392 | 0 |
| doc 32 | 0 | 0 | 0.011905 | 0 | 0 | 0.041667 | 0.136905 | 0 | 0.005952 | 0.071429 | 0.291667 | 0.071429 | 0.005952 | 0 | 0.208333 | 0.107143 | 0 | 0.017857 | 0.011905 | 0.017857 |
| doc 33 | 0.005025 | 0.005025 | 0.01005 | 0.025126 | 0.045226 | 0.01005 | 0.085427 | 0 | 0.005025 | 0.015075 | 0 | 0.01005 | 0.030151 | 0.015075 | 0.055276 | 0.190955 | 0.39196 | 0.015075 | 0.085427 | 0 |
| doc 34 | 0.506127 | 0.019608 | 0.001225 | 0.023284 | 0.002451 | 0.019608 | 0.007353 | 0.055147 | 0.001225 | 0.008578 | 0.003676 | 0.012255 | 0.084559 | 0 | 0.017157 | 0.007353 | 0.046569 | 0.126225 | 0.001225 | 0.056373 |
| doc 35 | 0.041509 | 0.026415 | 0.041509 | 0.003774 | 0.241509 | 0 | 0.007547 | 0.003774 | 0 | 0 | 0.003774 | 0.003774 | 0 | 0.011321 | 0.003774 | 0.003774 | 0.071698 | 0.358491 | 0.177358 | 0 |
| doc 36 | 0.002012 | 0.006036 | 0.199195 | 0.002012 | 0.040241 | 0.076459 | 0.014085 | 0.026157 | 0.012072 | 0.020121 | 0.01006 | 0.034205 | 0.026157 | 0.004024 | 0 | 0.01006 | 0.406439 | 0.100604 | 0.002012 | 0.008048 |
| doc 37 | 0.815934 | 0 | 0.002747 | 0.021978 | 0.002747 | 0.008242 | 0.010989 | 0.019231 | 0 | 0.016484 | 0 | 0.016484 | 0.03022 | 0.008242 | 0.002747 | 0.013736 | 0.002747 | 0.019231 | 0.008242 | 0 |
| doc 38 | 0.010582 | 0.716931 | 0 | 0.079365 | 0 | 0.002646 | 0.002646 | 0.087302 | 0.015873 | 0.02381 | 0.013228 | 0 | 0.002646 | 0 | 0 | 0.002646 | 0.015873 | 0.010582 | 0 | 0.015873 |
| doc 39 | 0.017699 | 0.0059 | 0.00295 | 0 | 0.011799 | 0.294985 | 0.00295 | 0.00295 | 0.085546 | 0.023599 | 0.076696 | 0.053097 | 0 | 0 | 0.014749 | 0.053097 | 0.351032 | 0 | 0.00295 | 0 |
| doc 40 | 0.134021 | 0 | 0.061856 | 0.006873 | 0.024055 | 0.151203 | 0.003436 | 0 | 0.013746 | 0.006873 | 0.268041 | 0.054983 | 0.006873 | 0 | 0.034364 | 0.072165 | 0.120275 | 0.027491 | 0 | 0.013746 |
| doc 41 | 0.006849 | 0.020548 | 0.006849 | 0.013699 | 0.068493 | 0.013699 | 0 | 0 | 0.013699 | 0.006849 | 0 | 0.013699 | 0.006849 | 0 | 0.452055 | 0.020548 | 0 | 0.30137 | 0.020548 | 0.034247 |
| doc 42 | 0 | 0 | 0.00625 | 0 | 0.075 | 0.025 | 0 | 0 | 0.025 | 0.05625 | 0.0125 | 0.0125 | 0.01875 | 0 | 0.18125 | 0.3 | 0.0375 | 0.20625 | 0.01875 | 0.025 |
| doc 43 | 0.027435 | 0.001372 | 0.001372 | 0.056241 | 0.010974 | 0.001372 | 0.030178 | 0.109739 | 0.043896 | 0.010974 | 0.024691 | 0.012346 | 0.030178 | 0.004115 | 0.00823 | 0.045267 | 0.050754 | 0 | 0.474623 | 0.056241 |
| doc 44 | 0 | 0.005386 | 0.005386 | 0.001795 | 0.001795 | 0 | 0.003591 | 0.08079 | 0.010772 | 0.057451 | 0.025135 | 0 | 0.016158 | 0.001795 | 0.001795 | 0.021544 | 0.043088 | 0.012567 | 0.70377 | 0.007181 |
| doc 45 | 0.020661 | 0 | 0.082645 | 0 | 0.099174 | 0.061983 | 0.004132 | 0.008264 | 0.012397 | 0.016529 | 0.004132 | 0.041322 | 0 | 0.004132 | 0.008264 | 0.061983 | 0.347107 | 0.020661 | 0.206612 | 0 |
| doc 46 | 0.005405 | 0.037838 | 0.07027 | 0.032432 | 0 | 0.005405 | 0.027027 | 0.010811 | 0.010811 | 0.081081 | 0.243243 | 0.027027 | 0.07027 | 0.005405 | 0.010811 | 0.135135 | 0.07027 | 0.118919 | 0.021622 | 0.016216 |
| doc 47 | 0.003205 | 0.009615 | 0.019231 | 0 | 0.044872 | 0.259615 | 0.022436 | 0 | 0.064103 | 0.012821 | 0.102564 | 0.112179 | 0.003205 | 0 | 0 | 0.022436 | 0.13141 | 0.169872 | 0.016026 | 0.00641 |
| doc 48 | 0.02968 | 0.004566 | 0.002283 | 0.027397 | 0.203196 | 0.018265 | 0.009132 | 0.086758 | 0 | 0.369863 | 0.006849 | 0 | 0.011416 | 0 | 0 | 0.027397 | 0.004566 | 0.070776 | 0.10274 | 0.025114 |
| doc 49 | 0 | 0.032787 | 0 | 0.004098 | 0.008197 | 0.012295 | 0.020492 | 0.008197 | 0.028689 | 0.036885 | 0.020492 | 0.028689 | 0 | 0.016393 | 0 | 0.008197 | 0.053279 | 0 | 0.713115 | 0.008197 |
| doc 50 | 0.020408 | 0 | 0.010204 | 0.010204 | 0.193878 | 0 | 0.020408 | 0 | 0.010204 | 0 | 0 | 0 | 0 | 0 | 0.030612 | 0.153061 | 0.173469 | 0.193878 | 0.122449 | 0.061224 |
| doc 51 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.014493 | 0 | 0 | 0 | 0 | 0 | 0.971014 | 0 | 0 | 0 | 0 | 0 | 0.014493 |
| doc 52 | 0.004202 | 0 | 0 | 0.063025 | 0 | 0.012605 | 0.033613 | 0 | 0.415966 | 0.016807 | 0.205882 | 0.004202 | 0 | 0 | 0.222689 | 0 | 0.021008 | 0 | 0 | 0 |
| doc 53 | 0.046784 | 0.007797 | 0.007797 | 0.068226 | 0.007797 | 0.025341 | 0.005848 | 0.204678 | 0.015595 | 0.136452 | 0.003899 | 0 | 0 | 0 | 0.007797 | 0.02924 | 0.005848 | 0 | 0.413255 | 0.013645 |
| doc 54 | 0.005848 | 0.002924 | 0.026316 | 0.040936 | 0.362573 | 0.005848 | 0.020468 | 0.017544 | 0.011696 | 0.377193 | 0.046784 | 0.011696 | 0.002924 | 0.002924 | 0 | 0.011696 | 0.011696 | 0.011696 | 0.017544 | 0.011696 |
| doc 55 | 0.003344 | 0.010033 | 0.006689 | 0.036789 | 0.073579 | 0 | 0.013378 | 0.023411 | 0.006689 | 0.685619 | 0.020067 | 0.003344 | 0.003344 | 0.003344 | 0 | 0.0301 | 0.006689 | 0.056856 | 0.013378 | 0.003344 |
| doc 56 | 0.003597 | 0.014388 | 0.02518 | 0.122302 | 0.003597 | 0.438849 | 0.010791 | 0 | 0.032374 | 0.010791 | 0.215827 | 0 | 0 | 0.003597 | 0.003597 | 0.039568 | 0.064748 | 0.007194 | 0 | 0.003597 |
| doc 57 | 0.006494 | 0 | 0 | 0.064935 | 0.019481 | 0.019481 | 0.012987 | 0.006494 | 0.181818 | 0 | 0 | 0 | 0 | 0 | 0.032468 | 0.603896 | 0.012987 | 0 | 0.038961 | 0 |
| doc 58 | 0.004566 | 0 | 0 | 0.009132 | 0.045662 | 0 | 0.073059 | 0.045662 | 0.004566 | 0.525114 | 0.068493 | 0.027397 | 0 | 0 | 0.009132 | 0.150685 | 0.009132 | 0.004566 | 0.004566 | 0.018265 |
| doc 59 | 0.040404 | 0.03367 | 0.060606 | 0 | 0 | 0.037037 | 0.13468 | 0.043771 | 0.003367 | 0.094276 | 0.208754 | 0.003367 | 0.013468 | 0.013468 | 0.191919 | 0.006734 | 0.003367 | 0.087542 | 0.006734 | 0.016835 |
| doc 60 | 0.011019 | 0.030303 | 0.082645 | 0 | 0.300275 | 0 | 0.002755 | 0.00551 | 0.013774 | 0 | 0.022039 | 0.016529 | 0.030303 | 0.002755 | 0.002755 | 0.404959 | 0 | 0.057851 | 0.016529 | 0 |
| doc 61 | 0 | 0.424332 | 0.002967 | 0.002967 | 0 | 0.002967 | 0.062315 | 0.020772 | 0.035608 | 0.002967 | 0.008902 | 0.005935 | 0.136499 | 0.002967 | 0.047478 | 0.008902 | 0 | 0.21365 | 0.008902 | 0.011869 |
| doc 62 | 0.086207 | 0.034483 | 0.045977 | 0.017241 | 0.413793 | 0 | 0.051724 | 0 | 0.005747 | 0.028736 | 0 | 0 | 0.005747 | 0 | 0 | 0.137931 | 0 | 0 | 0.155172 | 0.017241 |
| doc 63 | 0 | 0.013072 | 0.084967 | 0.078431 | 0.019608 | 0.013072 | 0.019608 | 0.03268 | 0.333333 | 0.019608 | 0.006536 | 0.006536 | 0 | 0 | 0 | 0.352941 | 0 | 0.013072 | 0.006536 | 0 |
| doc 64 | 0.003861 | 0.011583 | 0.494208 | 0.015444 | 0.003861 | 0.084942 | 0.027027 | 0.042471 | 0.096525 | 0 | 0.027027 | 0.027027 | 0.011583 | 0 | 0.003861 | 0.003861 | 0.084942 | 0.015444 | 0.019305 | 0.027027 |
| doc 65 | 0 | 0 | 0 | 0.019553 | 0.011173 | 0.022346 | 0.055866 | 0 | 0.019553 | 0.002793 | 0.365922 | 0.03352 | 0.22067 | 0.005587 | 0.086592 | 0.002793 | 0.00838 | 0.136872 | 0 | 0.00838 |
| doc 66 | 0.004167 | 0 | 0 | 0 | 0.241667 | 0 | 0 | 0.004167 | 0 | 0.058333 | 0 | 0.083333 | 0.016667 | 0.008333 | 0.075 | 0.245833 | 0.004167 | 0.245833 | 0 | 0.0125 |
| doc 67 | 0.015385 | 0.553846 | 0 | 0.005128 | 0.005128 | 0.020513 | 0.035897 | 0.015385 | 0 | 0.015385 | 0.025641 | 0 | 0.020513 | 0 | 0.051282 | 0 | 0.128205 | 0.051282 | 0 | 0.05641 |
| doc 68 | 0.010256 | 0.005128 | 0.051282 | 0.010256 | 0.169231 | 0.066667 | 0.010256 | 0 | 0.117949 | 0.005128 | 0.010256 | 0.030769 | 0 | 0 | 0.082051 | 0.328205 | 0.071795 | 0.025641 | 0.005128 | 0 |
| doc 69 | 0.046512 | 0.162791 | 0.005814 | 0.023256 | 0.087209 | 0 | 0.023256 | 0 | 0.040698 | 0 | 0.052326 | 0.05814 | 0.174419 | 0 | 0.22093 | 0.052326 | 0.011628 | 0.017442 | 0.005814 | 0.017442 |
| doc 70 | 0.022556 | 0.022556 | 0 | 0.015038 | 0.06015 | 0.015038 | 0 | 0.067669 | 0 | 0.142857 | 0 | 0.090226 | 0 | 0.097744 | 0.090226 | 0.18797 | 0.007519 | 0.165414 | 0.015038 | 0 |
| doc 71 | 0.007463 | 0.037313 | 0.014925 | 0.007463 | 0 | 0.671642 | 0.007463 | 0.022388 | 0.022388 | 0 | 0.029851 | 0.007463 | 0.022388 | 0.007463 | 0.007463 | 0.007463 | 0.029851 | 0.037313 | 0.052239 | 0.007463 |
| doc 72 | 0 | 0.007692 | 0 | 0.007692 | 0 | 0 | 0.023077 | 0 | 0.023077 | 0.015385 | 0.023077 | 0 | 0.007692 | 0.684615 | 0 | 0 | 0.007692 | 0.030769 | 0.169231 | 0 |
| doc 73 | 0.012397 | 0 | 0 | 0.024793 | 0 | 0.582645 | 0.012397 | 0.008264 | 0.028926 | 0 | 0.066116 | 0.012397 | 0.008264 | 0.012397 | 0.008264 | 0 | 0.202479 | 0 | 0.008264 | 0.012397 |
| doc 74 | 0.021739 | 0.504348 | 0 | 0.017391 | 0 | 0 | 0.147826 | 0.008696 | 0.013043 | 0 | 0 | 0.008696 | 0.06087 | 0 | 0.004348 | 0 | 0.030435 | 0.13913 | 0.026087 | 0.017391 |
| doc 75 | 0.027477 | 0.000901 | 0 | 0.011712 | 0.060811 | 0.000901 | 0.05045 | 0.118919 | 0.002703 | 0.196396 | 0.276577 | 0.015315 | 0.004955 | 0.00045 | 0.005405 | 0.001351 | 0.002252 | 0.215766 | 0.001802 | 0.005856 |
| doc 76 | 0.004032 | 0.004032 | 0.028226 | 0.028226 | 0.004032 | 0.072581 | 0 | 0.008065 | 0.008065 | 0.028226 | 0.104839 | 0.120968 | 0.008065 | 0.004032 | 0.004032 | 0.483871 | 0.012097 | 0.028226 | 0.020161 | 0.028226 |
| doc 77 | 0 | 0 | 0.013274 | 0.004425 | 0.134956 | 0 | 0 | 0.112832 | 0.017699 | 0.019912 | 0.108407 | 0.103982 | 0.024336 | 0.006637 | 0.037611 | 0.00885 | 0.002212 | 0.05531 | 0.068584 | 0.280973 |
| doc 78 | 0 | 0 | 0.012048 | 0.012048 | 0.036145 | 0.048193 | 0 | 0.012048 | 0 | 0.168675 | 0 | 0 | 0.024096 | 0 | 0 | 0.60241 | 0.060241 | 0 | 0 | 0.024096 |
| doc 79 | 0.092628 | 0.003781 | 0.035917 | 0.009452 | 0.037807 | 0.034026 | 0.007561 | 0.005671 | 0.00189 | 0.192817 | 0.035917 | 0.10775 | 0 | 0.00189 | 0.126654 | 0.277883 | 0 | 0.024575 | 0 | 0.003781 |
| doc 80 | 0 | 0 | 0 | 0 | 0.006667 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.993333 | 0 | 0 | 0 | 0 | 0 | 0 |
| doc 81 | 0.006452 | 0.002151 | 0.215054 | 0.036559 | 0.187097 | 0.017204 | 0.017204 | 0.002151 | 0.025806 | 0.019355 | 0.062366 | 0.053763 | 0 | 0 | 0.021505 | 0.24086 | 0.015054 | 0.006452 | 0.066667 | 0.004301 |
| doc 82 | 0.008889 | 0.022222 | 0.017778 | 0.044444 | 0.008889 | 0.004444 | 0.017778 | 0.022222 | 0.004444 | 0.751111 | 0.004444 | 0.004444 | 0.017778 | 0.004444 | 0 | 0.048889 | 0 | 0.013333 | 0.004444 | 0 |
| doc 83 | 0.015326 | 0.030651 | 0.005747 | 0.074713 | 0.019157 | 0.011494 | 0.070881 | 0.266284 | 0.007663 | 0.009579 | 0.01341 | 0.02682 | 0.02682 | 0.001916 | 0.009579 | 0.003831 | 0.068966 | 0.005747 | 0.241379 | 0.090038 |
| doc 84 | 0.012397 | 0.008264 | 0.014463 | 0.016529 | 0 | 0 | 0 | 0.086777 | 0.030992 | 0.144628 | 0.012397 | 0.002066 | 0.018595 | 0 | 0.008264 | 0.070248 | 0.028926 | 0.004132 | 0.485537 | 0.055785 |
| doc 85 | 0.008547 | 0.031339 | 0.005698 | 0.02849 | 0.011396 | 0.407407 | 0.005698 | 0.002849 | 0.008547 | 0.031339 | 0.088319 | 0.019943 | 0.011396 | 0 | 0.005698 | 0.011396 | 0.202279 | 0.091168 | 0.025641 | 0.002849 |
| doc 86 | 0.038194 | 0 | 0.013889 | 0.003472 | 0.402778 | 0.003472 | 0.006944 | 0.010417 | 0.010417 | 0.076389 | 0.034722 | 0.052083 | 0.003472 | 0.006944 | 0.065972 | 0.006944 | 0.0625 | 0.152778 | 0.006944 | 0.041667 |
| doc 87 | 0.009346 | 0 | 0.261682 | 0 | 0 | 0 | 0.014019 | 0.014019 | 0 | 0 | 0.004673 | 0.065421 | 0.116822 | 0.004673 | 0.228972 | 0 | 0.051402 | 0.182243 | 0.023364 | 0.023364 |
| doc 88 | 0.020101 | 0.040201 | 0.180905 | 0 | 0.110553 | 0.005025 | 0.01005 | 0.040201 | 0.005025 | 0.060302 | 0.030151 | 0.221106 | 0 | 0.020101 | 0.040201 | 0.005025 | 0 | 0.030151 | 0.055276 | 0.125628 |
| doc 89 | 0 | 0 | 0.040541 | 0 | 0.351351 | 0 | 0.013514 | 0 | 0 | 0.22973 | 0.040541 | 0.121622 | 0 | 0 | 0.013514 | 0.148649 | 0 | 0.013514 | 0.013514 | 0.013514 |
| doc 90 | 0.006329 | 0.006329 | 0.018987 | 0 | 0.06962 | 0.012658 | 0 | 0.050633 | 0.006329 | 0.018987 | 0.012658 | 0.037975 | 0.012658 | 0.664557 | 0 | 0.044304 | 0.006329 | 0.006329 | 0.012658 | 0.012658 |
| doc 91 | 0.007463 | 0 | 0 | 0 | 0.007463 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.977612 | 0.007463 | 0 | 0 | 0 | 0 | 0 |
| doc 92 | 0.053977 | 0.011364 | 0.002841 | 0.025568 | 0.073864 | 0.005682 | 0.059659 | 0.099432 | 0.153409 | 0.005682 | 0.022727 | 0.215909 | 0.017045 | 0.008523 | 0 | 0.014205 | 0.139205 | 0.068182 | 0.005682 | 0.017045 |
| doc 93 | 0.005814 | 0 | 0.087209 | 0.011628 | 0.241279 | 0.011628 | 0 | 0.052326 | 0.005814 | 0.177326 | 0 | 0.011628 | 0.002907 | 0.002907 | 0.020349 | 0.252907 | 0 | 0.046512 | 0.063953 | 0.005814 |
| doc 94 | 0 | 0 | 0.035398 | 0 | 0.353982 | 0 | 0 | 0 | 0 | 0.079646 | 0 | 0.061947 | 0 | 0.053097 | 0.141593 | 0.070796 | 0.053097 | 0.150442 | 0 | 0 |
| doc 95 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.032787 | 0 | 0.942623 | 0.008197 | 0 | 0 | 0.008197 | 0 | 0.008197 |
| doc 96 | 0.004049 | 0.016194 | 0.218623 | 0.004049 | 0.076923 | 0.008097 | 0.012146 | 0 | 0.012146 | 0.036437 | 0.012146 | 0.109312 | 0.036437 | 0.004049 | 0.080972 | 0 | 0.060729 | 0.105263 | 0.008097 | 0.194332 |
| doc 97 | 0 | 0.003891 | 0.163424 | 0.019455 | 0.003891 | 0.003891 | 0 | 0.011673 | 0 | 0.058366 | 0.587549 | 0.042802 | 0 | 0 | 0.023346 | 0.023346 | 0.011673 | 0.038911 | 0 | 0.007782 |
| doc 98 | 0.009524 | 0 | 0.019048 | 0.009524 | 0.228571 | 0.07619 | 0 | 0.019048 | 0.114286 | 0.009524 | 0 | 0.047619 | 0 | 0.038095 | 0.038095 | 0.266667 | 0.057143 | 0 | 0.066667 | 0 |
| doc 99 | 0.026667 | 0.02 | 0 | 0.026667 | 0.18 | 0.013333 | 0.04 | 0.033333 | 0.006667 | 0.046667 | 0.006667 | 0.22 | 0.073333 | 0.04 | 0.006667 | 0.04 | 0.026667 | 0.18 | 0.013333 | 0 |
| doc 100 | 0.011696 | 0.555556 | 0 | 0.005848 | 0 | 0.02924 | 0.035088 | 0.233918 | 0.005848 | 0 | 0.05848 | 0 | 0 | 0.005848 | 0 | 0.011696 | 0 | 0.017544 | 0.017544 | 0.011696 |
| doc 101 | 0 | 0.163743 | 0.239766 | 0 | 0.011696 | 0.005848 | 0.035088 | 0.023392 | 0.134503 | 0.052632 | 0 | 0 | 0.005848 | 0.005848 | 0.087719 | 0 | 0 | 0.152047 | 0.035088 | 0.046784 |

# Appendix 2. The sLDA model Per Topic-Document proportion output (doc3970-doc4072)

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| doc 3970 | 0.015584 | 0.711688 | 0.015584 | 0.083117 | 0.012987 | 0.005195 | 0.018182 | 0.012987 | 0.012987 | 0.012987 | 0.018182 | 0 | 0.025974 | 0 | 0.012987 | 0 | 0.020779 | 0.012987 | 0.005195 | 0.002597 |
| doc 3971 | 0.008547 | 0.006838 | 0.003419 | 0.02735 | 0.051282 | 0.02735 | 0.068376 | 0.025641 | 0 | 0.054701 | 0.557265 | 0.044444 | 0.008547 | 0.006838 | 0.054701 | 0.003419 | 0.010256 | 0.02906 | 0.003419 | 0.008547 |
| doc 3972 | 0.008247 | 0.002062 | 0.02268 | 0.063918 | 0.02268 | 0 | 0.03299 | 0.090722 | 0.018557 | 0.230928 | 0.002062 | 0.004124 | 0.018557 | 0 | 0 | 0.026804 | 0.024742 | 0.012371 | 0.412371 | 0.006186 |
| doc 3973 | 0.004587 | 0.077982 | 0.009174 | 0.013761 | 0.018349 | 0.009174 | 0.105505 | 0.087156 | 0.027523 | 0.087156 | 0.013761 | 0.03211 | 0.041284 | 0.004587 | 0.013761 | 0.041284 | 0.09633 | 0 | 0.307339 | 0.009174 |
| doc 3974 | 0.006211 | 0.012422 | 0.31677 | 0.006211 | 0.012422 | 0 | 0.012422 | 0.012422 | 0.018634 | 0.055901 | 0 | 0.006211 | 0.018634 | 0.006211 | 0.049689 | 0.26087 | 0 | 0.130435 | 0.055901 | 0.018634 |
| doc 3975 | 0.020408 | 0.020408 | 0.006803 | 0.013605 | 0 | 0.006803 | 0.006803 | 0.122449 | 0.013605 | 0.006803 | 0.346939 | 0 | 0 | 0 | 0.244898 | 0.006803 | 0.040816 | 0.136054 | 0.006803 | 0 |
| doc 3976 | 0.05283 | 0.003774 | 0.158491 | 0 | 0 | 0.154717 | 0.015094 | 0.015094 | 0 | 0.003774 | 0 | 0.015094 | 0.018868 | 0.003774 | 0.007547 | 0.007547 | 0.275472 | 0.215094 | 0.003774 | 0.049057 |
| doc 3977 | 0.017668 | 0.007067 | 0.007067 | 0.024735 | 0.04947 | 0.021201 | 0.109541 | 0.010601 | 0.007067 | 0.014134 | 0.017668 | 0.053004 | 0.007067 | 0.003534 | 0.007067 | 0.084806 | 0.508834 | 0.035336 | 0.010601 | 0.003534 |
| doc 3978 | 0 | 0.395833 | 0.2125 | 0.004167 | 0.008333 | 0.029167 | 0.045833 | 0.05 | 0.025 | 0.004167 | 0.029167 | 0 | 0.004167 | 0 | 0.0875 | 0.066667 | 0.008333 | 0.025 | 0.004167 | 0 |
| doc 3979 | 0.006154 | 0.009231 | 0.003077 | 0.209231 | 0.055385 | 0.018462 | 0.04 | 0.175385 | 0.021538 | 0.024615 | 0.021538 | 0.024615 | 0.08 | 0 | 0.018462 | 0.061538 | 0.006154 | 0.153846 | 0.049231 | 0.021538 |
| doc 3980 | 0 | 0.072607 | 0.0033 | 0.026403 | 0.016502 | 0.135314 | 0.092409 | 0.0033 | 0 | 0.009901 | 0.029703 | 0.039604 | 0 | 0 | 0.026403 | 0.0033 | 0.148515 | 0.148515 | 0.042904 | 0.20132 |
| doc 3981 | 0.003049 | 0.006098 | 0.04878 | 0.02439 | 0.094512 | 0 | 0.033537 | 0.085366 | 0.006098 | 0.079268 | 0.060976 | 0.039634 | 0 | 0.015244 | 0.003049 | 0.036585 | 0.039634 | 0.277439 | 0.128049 | 0.018293 |
| doc 3982 | 0 | 0.015038 | 0.007519 | 0.015038 | 0.007519 | 0 | 0.421053 | 0.052632 | 0.030075 | 0.007519 | 0 | 0.030075 | 0.007519 | 0 | 0.285714 | 0.045113 | 0.037594 | 0.022556 | 0.007519 | 0.007519 |
| doc 3983 | 0 | 0 | 0 | 0 | 0.008696 | 0.008696 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.93913 | 0.008696 | 0.008696 | 0.008696 | 0.008696 | 0 | 0.008696 |
| doc 3984 | 0.02935 | 0.012579 | 0.392034 | 0 | 0.035639 | 0.010482 | 0.020964 | 0 | 0.188679 | 0.018868 | 0.02935 | 0.092243 | 0.025157 | 0 | 0.010482 | 0.016771 | 0.031447 | 0.067086 | 0.004193 | 0.014675 |
| doc 3985 | 0.044118 | 0 | 0.073529 | 0 | 0.176471 | 0 | 0.088235 | 0 | 0.014706 | 0.323529 | 0 | 0.058824 | 0 | 0 | 0.058824 | 0.014706 | 0 | 0.029412 | 0.088235 | 0.029412 |
| doc 3986 | 0 | 0.005882 | 0 | 0.035294 | 0.035294 | 0.005882 | 0 | 0.029412 | 0.035294 | 0.005882 | 0 | 0.064706 | 0 | 0 | 0.252941 | 0.470588 | 0.035294 | 0 | 0.005882 | 0.017647 |
| doc 3987 | 0.03908 | 0.006897 | 0.055172 | 0 | 0.03908 | 0.027586 | 0.025287 | 0.062069 | 0.052874 | 0.009195 | 0.062069 | 0.055172 | 0.025287 | 0.002299 | 0.002299 | 0.022989 | 0.314943 | 0.186207 | 0.009195 | 0.002299 |
| doc 3988 | 0.014572 | 0.045537 | 0.003643 | 0.081967 | 0.016393 | 0.081967 | 0.105647 | 0.010929 | 0.001821 | 0.005464 | 0.428051 | 0 | 0.061931 | 0.003643 | 0.041894 | 0.020036 | 0 | 0.069217 | 0.001821 | 0.005464 |
| doc 3989 | 0.006173 | 0.012346 | 0.009259 | 0 | 0.046296 | 0.003086 | 0.018519 | 0 | 0.003086 | 0.824074 | 0.009259 | 0 | 0.012346 | 0 | 0.009259 | 0.009259 | 0.009259 | 0.027778 | 0 | 0 |
| doc 3990 | 0 | 0.018868 | 0.09434 | 0 | 0.018868 | 0.037736 | 0 | 0.075472 | 0 | 0 | 0.018868 | 0.169811 | 0.018868 | 0 | 0.264151 | 0.113208 | 0 | 0.150943 | 0.018868 | 0 |
| doc 3991 | 0.042105 | 0.005263 | 0.257895 | 0.005263 | 0.005263 | 0.010526 | 0 | 0 | 0.042105 | 0.057895 | 0.026316 | 0 | 0 | 0 | 0.010526 | 0.494737 | 0 | 0.026316 | 0 | 0.015789 |
| doc 3992 | 0 | 0 | 0.021739 | 0 | 0.01087 | 0.01087 | 0.032609 | 0.01087 | 0.076087 | 0.01087 | 0.195652 | 0.141304 | 0 | 0 | 0 | 0.336957 | 0 | 0.065217 | 0.032609 | 0.054348 |
| doc 3993 | 0.008065 | 0 | 0.004032 | 0 | 0.03629 | 0.004032 | 0.294355 | 0.076613 | 0.016129 | 0.012097 | 0.217742 | 0.016129 | 0.032258 | 0.004032 | 0.012097 | 0.092742 | 0.092742 | 0 | 0.024194 | 0.056452 |
| doc 3994 | 0 | 0.022472 | 0.022472 | 0.011236 | 0.280899 | 0.033708 | 0.05618 | 0 | 0.033708 | 0.269663 | 0.022472 | 0.05618 | 0 | 0.022472 | 0.011236 | 0.134831 | 0 | 0 | 0 | 0.022472 |
| doc 3995 | 0.012658 | 0.063291 | 0.004219 | 0 | 0.004219 | 0.21097 | 0 | 0.004219 | 0.012658 | 0.042194 | 0.025316 | 0.092827 | 0.059072 | 0.037975 | 0.244726 | 0 | 0.008439 | 0.008439 | 0.025316 | 0.14346 |
| doc 3996 | 0.014085 | 0.016432 | 0.00939 | 0.002347 | 0 | 0.007042 | 0.00939 | 0.070423 | 0.023474 | 0.018779 | 0.481221 | 0.023474 | 0.002347 | 0 | 0.039906 | 0.025822 | 0.124413 | 0.126761 | 0.002347 | 0.002347 |
| doc 3997 | 0 | 0 | 0.117155 | 0 | 0.004184 | 0.004184 | 0.012552 | 0 | 0.020921 | 0 | 0.008368 | 0.225941 | 0.004184 | 0.004184 | 0.025105 | 0.292887 | 0.096234 | 0.167364 | 0.016736 | 0 |
| doc 3998 | 0 | 0 | 0.008929 | 0 | 0 | 0 | 0 | 0 | 0 | 0.008929 | 0 | 0 | 0 | 0.982143 | 0 | 0 | 0 | 0 | 0 | 0 |
| doc 3999 | 0 | 0.038278 | 0.019139 | 0.002392 | 0.004785 | 0.011962 | 0.619617 | 0.004785 | 0.019139 | 0.014354 | 0.004785 | 0.114833 | 0.016746 | 0.002392 | 0.026316 | 0.033493 | 0 | 0.033493 | 0.028708 | 0.004785 |
| doc 4000 | 0.229437 | 0 | 0.008658 | 0.002165 | 0.015152 | 0.028139 | 0.073593 | 0.010823 | 0 | 0.008658 | 0.008658 | 0.004329 | 0.004329 | 0.004329 | 0.002165 | 0 | 0.060606 | 0.038961 | 0.493506 | 0.006494 |
| doc 4001 | 0.008368 | 0.004184 | 0.196653 | 0.008368 | 0.066946 | 0.004184 | 0 | 0.020921 | 0.016736 | 0.355649 | 0.008368 | 0.066946 | 0.004184 | 0.012552 | 0.008368 | 0.200837 | 0.004184 | 0.008368 | 0.004184 | 0 |
| doc 4002 | 0.003333 | 0.006667 | 0.49 | 0 | 0 | 0.003333 | 0.093333 | 0.006667 | 0.02 | 0 | 0.016667 | 0 | 0.056667 | 0 | 0.096667 | 0.016667 | 0.026667 | 0.146667 | 0.006667 | 0.01 |
| doc 4003 | 0.0125 | 0 | 0 | 0.01875 | 0.04375 | 0.00625 | 0.00625 | 0.00625 | 0.00625 | 0.0125 | 0.025 | 0.13125 | 0.15625 | 0 | 0.075 | 0.00625 | 0.4875 | 0 | 0 | 0.00625 |
| doc 4004 | 0.082051 | 0.005128 | 0.487179 | 0 | 0 | 0 | 0.015385 | 0.020513 | 0 | 0.015385 | 0.025641 | 0.030769 | 0.005128 | 0.010256 | 0.025641 | 0.082051 | 0.041026 | 0.112821 | 0.041026 | 0 |
| doc 4005 | 0.02965 | 0.005391 | 0.018868 | 0 | 0 | 0 | 0.598383 | 0 | 0.048518 | 0.005391 | 0.018868 | 0.037736 | 0.010782 | 0 | 0.005391 | 0.153639 | 0.016173 | 0.03504 | 0.002695 | 0.013477 |
| doc 4006 | 0.014493 | 0.014493 | 0.003623 | 0.003623 | 0.01087 | 0.014493 | 0.003623 | 0.021739 | 0.003623 | 0 | 0.01087 | 0.003623 | 0.025362 | 0.003623 | 0.036232 | 0.126812 | 0.264493 | 0.293478 | 0.01087 | 0.134058 |
| doc 4007 | 0 | 0.007634 | 0.076336 | 0.015267 | 0.274809 | 0.007634 | 0 | 0.030534 | 0 | 0.091603 | 0.10687 | 0.038168 | 0.038168 | 0.007634 | 0.007634 | 0.10687 | 0 | 0.167939 | 0.007634 | 0.015267 |
| doc 4008 | 0.005952 | 0.041667 | 0.005952 | 0 | 0.029762 | 0.005952 | 0 | 0.011905 | 0 | 0.172619 | 0.041667 | 0 | 0 | 0.005952 | 0 | 0 | 0 | 0.446429 | 0.214286 | 0.017857 |
| doc 4009 | 0.019704 | 0.172414 | 0.152709 | 0.018062 | 0.036125 | 0.024631 | 0.004926 | 0 | 0.006568 | 0.006568 | 0.009852 | 0.082102 | 0.175698 | 0.004926 | 0.123153 | 0.045977 | 0.003284 | 0.090312 | 0.022989 | 0 |
| doc 4010 | 0.004587 | 0 | 0.082569 | 0.009174 | 0.091743 | 0.004587 | 0.073394 | 0 | 0 | 0.004587 | 0 | 0.119266 | 0.013761 | 0.004587 | 0.004587 | 0.004587 | 0.050459 | 0.059633 | 0.050459 | 0.422018 |
| doc 4011 | 0.018828 | 0.002092 | 0.043933 | 0.002092 | 0.062762 | 0.228033 | 0.008368 | 0.016736 | 0.012552 | 0.008368 | 0.100418 | 0.041841 | 0.016736 | 0.002092 | 0.228033 | 0.140167 | 0.002092 | 0.060669 | 0.004184 | 0 |
| doc 4012 | 0.001887 | 0.037736 | 0.003774 | 0.056604 | 0.035849 | 0.00566 | 0.007547 | 0.3 | 0.009434 | 0.039623 | 0.003774 | 0.026415 | 0.013208 | 0.018868 | 0.001887 | 0.011321 | 0.018868 | 0.015094 | 0.260377 | 0.132075 |
| doc 4013 | 0.015625 | 0.003906 | 0.042969 | 0.001953 | 0.052734 | 0.304688 | 0.017578 | 0.003906 | 0.228516 | 0.025391 | 0.003906 | 0.046875 | 0.019531 | 0 | 0.021484 | 0.095703 | 0.044922 | 0.052734 | 0.001953 | 0.015625 |
| doc 4014 | 0.017778 | 0.004444 | 0.062222 | 0 | 0.066667 | 0 | 0.004444 | 0.004444 | 0.004444 | 0.008889 | 0.004444 | 0.004444 | 0.004444 | 0.546667 | 0 | 0.004444 | 0.066667 | 0.102222 | 0.093333 | 0 |
| doc 4015 | 0 | 0 | 0.006944 | 0 | 0.013889 | 0 | 0.013889 | 0.006944 | 0 | 0 | 0 | 0.055556 | 0.006944 | 0.625 | 0.006944 | 0.006944 | 0.020833 | 0.006944 | 0.159722 | 0.069444 |
| doc 4016 | 0.020619 | 0.010309 | 0 | 0.010309 | 0.046392 | 0.118557 | 0 | 0 | 0.041237 | 0.030928 | 0.010309 | 0.010309 | 0.015464 | 0 | 0.154639 | 0.489691 | 0 | 0.030928 | 0.005155 | 0.005155 |
| doc 4017 | 0.020833 | 0.015625 | 0.505208 | 0.020833 | 0.010417 | 0.005208 | 0.005208 | 0.096354 | 0.067708 | 0.005208 | 0.002604 | 0.005208 | 0.013021 | 0 | 0.015625 | 0.080729 | 0.007813 | 0.033854 | 0.015625 | 0.072917 |
| doc 4018 | 0.017928 | 0.011952 | 0 | 0.095618 | 0.085657 | 0.023904 | 0.02988 | 0.181275 | 0.023904 | 0.047809 | 0 | 0 | 0.003984 | 0.045817 | 0 | 0.011952 | 0.043825 | 0 | 0.318725 | 0.057769 |
| doc 4019 | 0.061611 | 0.018957 | 0.080569 | 0.004739 | 0 | 0 | 0.004739 | 0.004739 | 0.042654 | 0.018957 | 0.004739 | 0.061611 | 0.047393 | 0 | 0.07109 | 0.028436 | 0.042654 | 0.075829 | 0.009479 | 0.421801 |
| doc 4020 | 0.009709 | 0.009709 | 0.02589 | 0 | 0.022654 | 0.145631 | 0.003236 | 0 | 0.022654 | 0.003236 | 0 | 0.05178 | 0 | 0.003236 | 0.330097 | 0.132686 | 0.016181 | 0.200647 | 0.003236 | 0.019417 |
| doc 4021 | 0.027273 | 0.063636 | 0.009091 | 0.013636 | 0.022727 | 0 | 0 | 0.004545 | 0.031818 | 0 | 0 | 0 | 0 | 0.009091 | 0.045455 | 0.009091 | 0.018182 | 0.004545 | 0.722727 | 0.018182 |
| doc 4022 | 0 | 0.021739 | 0.006211 | 0.009317 | 0.009317 | 0.009317 | 0.009317 | 0.021739 | 0 | 0.009317 | 0 | 0 | 0.096273 | 0.012422 | 0.015528 | 0.009317 | 0.068323 | 0.121118 | 0.015528 | 0.565217 |
| doc 4023 | 0 | 0.140351 | 0 | 0.013158 | 0 | 0.035088 | 0.188596 | 0.013158 | 0.214912 | 0 | 0.087719 | 0.065789 | 0.149123 | 0.004386 | 0.013158 | 0 | 0.039474 | 0.017544 | 0.017544 | 0 |
| doc 4024 | 0 | 0.018116 | 0.007246 | 0 | 0.007246 | 0 | 0.003623 | 0.086957 | 0.108696 | 0.014493 | 0.608696 | 0.021739 | 0.021739 | 0 | 0.036232 | 0 | 0.003623 | 0.050725 | 0.003623 | 0.007246 |
| doc 4025 | 0.032143 | 0.010714 | 0 | 0.028571 | 0.007143 | 0.610714 | 0.010714 | 0 | 0.007143 | 0.003571 | 0.017857 | 0.014286 | 0.014286 | 0 | 0.021429 | 0.014286 | 0 | 0.185714 | 0.021429 | 0 |
| doc 4026 | 0.012915 | 0.035055 | 0.005535 | 0.00369 | 0.053506 | 0.027675 | 0.188192 | 0.184502 | 0.300738 | 0.001845 | 0.119926 | 0.009225 | 0.001845 | 0 | 0.001845 | 0.00369 | 0 | 0.00369 | 0.023985 | 0.02214 |
| doc 4027 | 0.021898 | 0 | 0.072993 | 0 | 0.138686 | 0 | 0.029197 | 0 | 0 | 0.058394 | 0.043796 | 0.029197 | 0 | 0.007299 | 0.036496 | 0.437956 | 0.007299 | 0.065693 | 0.043796 | 0.007299 |
| doc 4028 | 0.01581 | 0.12253 | 0.035573 | 0.003953 | 0 | 0.019763 | 0.003953 | 0.011858 | 0.501976 | 0 | 0.003953 | 0 | 0 | 0 | 0.118577 | 0.059289 | 0.047431 | 0.039526 | 0.007905 | 0.007905 |
| doc 4029 | 0 | 0 | 0.014815 | 0.155556 | 0.022222 | 0.003704 | 0.014815 | 0.014815 | 0.003704 | 0.618519 | 0.096296 | 0.033333 | 0 | 0 | 0.007407 | 0.007407 | 0 | 0.003704 | 0.003704 | 0 |
| doc 4030 | 0.006289 | 0 | 0.283019 | 0.012579 | 0.125786 | 0 | 0.006289 | 0 | 0 | 0.220126 | 0 | 0.056604 | 0.031447 | 0 | 0.006289 | 0.194969 | 0.018868 | 0.012579 | 0.006289 | 0.018868 |
| doc 4031 | 0 | 0.036667 | 0.056667 | 0.053333 | 0.04 | 0.05 | 0.006667 | 0.03 | 0 | 0.013333 | 0.636667 | 0.003333 | 0 | 0 | 0.003333 | 0.003333 | 0.03 | 0.02 | 0.003333 | 0.013333 |
| doc 4032 | 0.05098 | 0.003922 | 0.019608 | 0.117647 | 0 | 0.498039 | 0 | 0.015686 | 0.015686 | 0.015686 | 0.086275 | 0.015686 | 0.027451 | 0 | 0 | 0 | 0.12549 | 0 | 0.003922 | 0.003922 |
| doc 4033 | 0.004386 | 0.030702 | 0.013158 | 0.039474 | 0.114035 | 0.035088 | 0.048246 | 0.013158 | 0 | 0.048246 | 0.04386 | 0.008772 | 0.04386 | 0.074561 | 0 | 0.267544 | 0.02193 | 0.131579 | 0.061404 | 0 |
| doc 4034 | 0.014035 | 0.319298 | 0.003509 | 0.017544 | 0.007018 | 0.007018 | 0.031579 | 0.031579 | 0.017544 | 0 | 0.094737 | 0.003509 | 0.045614 | 0.003509 | 0.017544 | 0 | 0.010526 | 0.364912 | 0.010526 | 0 |
| doc 4035 | 0 | 0.052356 | 0.026178 | 0.041885 | 0.015707 | 0.429319 | 0.010471 | 0 | 0.036649 | 0 | 0.17801 | 0.015707 | 0.005236 | 0.005236 | 0.010471 | 0.052356 | 0.04712 | 0.062827 | 0.010471 | 0 |
| doc 4036 | 0.023952 | 0.011976 | 0.209581 | 0.011976 | 0.005988 | 0 | 0.48503 | 0.005988 | 0.047904 | 0 | 0.071856 | 0 | 0.041916 | 0 | 0.023952 | 0.005988 | 0.011976 | 0.005988 | 0.017964 | 0.017964 |
| doc 4037 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.010417 | 0 | 0.979167 | 0.010417 | 0 | 0 | 0 | 0 | 0 |
| doc 4038 | 0.014124 | 0.022599 | 0.5 | 0 | 0.002825 | 0 | 0.016949 | 0.016949 | 0.211864 | 0.050847 | 0.008475 | 0.056497 | 0.00565 | 0 | 0.00565 | 0.002825 | 0.002825 | 0.081921 | 0 | 0 |
| doc 4039 | 0.030812 | 0.016807 | 0.014006 | 0.005602 | 0.011204 | 0.008403 | 0.070028 | 0.008403 | 0.002801 | 0.028011 | 0.002801 | 0 | 0.62465 | 0 | 0 | 0.002801 | 0.005602 | 0.131653 | 0.002801 | 0.033613 |
| doc 4040 | 0.003906 | 0.007813 | 0.007813 | 0.007813 | 0.003906 | 0.191406 | 0.007813 | 0 | 0.039063 | 0.003906 | 0.003906 | 0.109375 | 0.003906 | 0 | 0.066406 | 0.488281 | 0.03125 | 0.003906 | 0.011719 | 0.007813 |
| doc 4041 | 0 | 0.005952 | 0.136905 | 0.035714 | 0 | 0.047619 | 0.005952 | 0 | 0.047619 | 0.005952 | 0 | 0.017857 | 0.02381 | 0.011905 | 0.041667 | 0.107143 | 0.011905 | 0 | 0 | 0.5 |
| doc 4042 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.012195 | 0 | 0 | 0 | 0.97561 | 0.012195 | 0 | 0 | 0 | 0 | 0 |
| doc 4043 | 0.009804 | 0.029412 | 0.009804 | 0 | 0.019608 | 0.009804 | 0.009804 | 0 | 0.029412 | 0.058824 | 0.009804 | 0.117647 | 0.009804 | 0 | 0.009804 | 0.490196 | 0 | 0.117647 | 0.029412 | 0.039216 |
| doc 4044 | 0.193182 | 0 | 0.007576 | 0 | 0.299242 | 0.056818 | 0.018939 | 0 | 0.079545 | 0.007576 | 0.05303 | 0.003788 | 0.015152 | 0 | 0 | 0.064394 | 0.034091 | 0.003788 | 0.151515 | 0.011364 |
| doc 4045 | 0.003401 | 0.001701 | 0 | 0.001701 | 0.246599 | 0.001701 | 0.028912 | 0.003401 | 0.006803 | 0.02551 | 0.003401 | 0.018707 | 0.02381 | 0.028912 | 0 | 0.008503 | 0.05102 | 0.353741 | 0.119048 | 0.073129 |
| doc 4046 | 0.001669 | 0.015025 | 0.001669 | 0.081803 | 0.158598 | 0.001669 | 0 | 0 | 0.010017 | 0.005008 | 0 | 0.225376 | 0.011686 | 0 | 0.003339 | 0 | 0.370618 | 0.098497 | 0.010017 | 0.005008 |
| doc 4047 | 0.016949 | 0.002825 | 0.081921 | 0.002825 | 0.10452 | 0.002825 | 0.050847 | 0.079096 | 0.022599 | 0.016949 | 0.008475 | 0.011299 | 0.011299 | 0.002825 | 0.00565 | 0.09322 | 0.435028 | 0 | 0.036723 | 0.014124 |
| doc 4048 | 0 | 0.004032 | 0.016129 | 0.032258 | 0.008065 | 0.008065 | 0.008065 | 0.153226 | 0.008065 | 0.193548 | 0.040323 | 0.008065 | 0.032258 | 0 | 0.004032 | 0.096774 | 0.008065 | 0.012097 | 0.366935 | 0 |
| doc 4049 | 0 | 0 | 0.005747 | 0.094828 | 0.002874 | 0.091954 | 0 | 0.014368 | 0.008621 | 0.054598 | 0.491379 | 0 | 0.005747 | 0.005747 | 0.04023 | 0.020115 | 0.020115 | 0.143678 | 0 | 0 |
| doc 4050 | 0.009063 | 0.003021 | 0.021148 | 0.003021 | 0.126888 | 0.084592 | 0.003021 | 0.018127 | 0.356495 | 0 | 0.003021 | 0.069486 | 0 | 0.02719 | 0.066465 | 0.048338 | 0.003021 | 0.054381 | 0.090634 | 0.012085 |
| doc 4051 | 0 | 0 | 0.113636 | 0 | 0 | 0.022727 | 0 | 0.022727 | 0 | 0.045455 | 0.045455 | 0.022727 | 0 | 0 | 0.045455 | 0.454545 | 0 | 0.204545 | 0 | 0.022727 |
| doc 4052 | 0 | 0.014423 | 0 | 0 | 0.014423 | 0.129808 | 0 | 0.004808 | 0.014423 | 0.004808 | 0.028846 | 0.355769 | 0.038462 | 0 | 0.009615 | 0.163462 | 0.168269 | 0.048077 | 0 | 0.004808 |
| doc 4053 | 0 | 0.040876 | 0.018978 | 0.00146 | 0.021898 | 0.10365 | 0.007299 | 0.013139 | 0.017518 | 0 | 0.309489 | 0.084672 | 0.013139 | 0 | 0.059854 | 0.09781 | 0.109489 | 0.071533 | 0.007299 | 0.021898 |
| doc 4054 | 0 | 0 | 0 | 0.021672 | 0.241486 | 0.009288 | 0 | 0.03096 | 0.012384 | 0.455108 | 0.01548 | 0.006192 | 0.024768 | 0 | 0.003096 | 0.040248 | 0.003096 | 0.024768 | 0.068111 | 0.043344 |
| doc 4055 | 0.02381 | 0 | 0.228571 | 0.014286 | 0.071429 | 0.014286 | 0.047619 | 0.009524 | 0 | 0.004762 | 0 | 0.019048 | 0 | 0 | 0 | 0.290476 | 0.138095 | 0.014286 | 0.12381 | 0 |
| doc 4056 | 0.012698 | 0.015873 | 0.073016 | 0.006349 | 0 | 0.003175 | 0.019048 | 0.006349 | 0.114286 | 0.003175 | 0.044444 | 0.003175 | 0.066667 | 0 | 0.457143 | 0.006349 | 0.092063 | 0.060317 | 0.006349 | 0.009524 |
| doc 4057 | 0.4 | 0.002985 | 0.140299 | 0.077612 | 0.023881 | 0.01791 | 0.01791 | 0.00597 | 0 | 0.044776 | 0.01194 | 0 | 0.008955 | 0 | 0.098507 | 0.050746 | 0.020896 | 0.047761 | 0.026866 | 0.002985 |
| doc 4058 | 0.048711 | 0.017192 | 0.011461 | 0.008596 | 0.197708 | 0.005731 | 0.005731 | 0.100287 | 0.008596 | 0.088825 | 0.020057 | 0.011461 | 0.022923 | 0.002865 | 0 | 0.005731 | 0.025788 | 0.386819 | 0.025788 | 0.005731 |
| doc 4059 | 0.023438 | 0.523438 | 0 | 0.054688 | 0 | 0.007813 | 0.007813 | 0.023438 | 0.070313 | 0 | 0.039063 | 0 | 0 | 0 | 0.015625 | 0.046875 | 0 | 0.132813 | 0.007813 | 0.046875 |
| doc 4060 | 0.003257 | 0.016287 | 0.006515 | 0.035831 | 0.026059 | 0.081433 | 0.045603 | 0.009772 | 0.04886 | 0 | 0.214984 | 0.003257 | 0.084691 | 0.009772 | 0.172638 | 0 | 0.179153 | 0.029316 | 0.006515 | 0.026059 |
| doc 4061 | 0 | 0.018987 | 0.037975 | 0.012658 | 0 | 0.006329 | 0.006329 | 0.012658 | 0 | 0.018987 | 0.031646 | 0 | 0.563291 | 0 | 0 | 0.075949 | 0.012658 | 0.170886 | 0.006329 | 0.025316 |
| doc 4062 | 0.010204 | 0.010204 | 0 | 0 | 0.255102 | 0 | 0.056122 | 0.010204 | 0.045918 | 0.056122 | 0.02551 | 0.030612 | 0.010204 | 0 | 0.122449 | 0.05102 | 0.005102 | 0.295918 | 0 | 0.015306 |
| doc 4063 | 0 | 0 | 0 | 0 | 0 | 0.011364 | 0 | 0.022727 | 0 | 0 | 0 | 0.022727 | 0 | 0.806818 | 0.011364 | 0 | 0.011364 | 0.068182 | 0.022727 | 0.022727 |
| doc 4064 | 0.342857 | 0.009524 | 0.052381 | 0.038095 | 0.02381 | 0.014286 | 0 | 0.114286 | 0.047619 | 0.028571 | 0.019048 | 0.02381 | 0.07619 | 0.004762 | 0.033333 | 0.028571 | 0.028571 | 0.080952 | 0 | 0.033333 |
| doc 4065 | 0.003937 | 0.023622 | 0.047244 | 0.007874 | 0.019685 | 0.015748 | 0.015748 | 0.043307 | 0 | 0.625984 | 0.003937 | 0.007874 | 0 | 0 | 0.043307 | 0.007874 | 0.011811 | 0.094488 | 0.027559 | 0 |
| doc 4066 | 0.010336 | 0.002584 | 0.436693 | 0.01292 | 0.002584 | 0.015504 | 0.377261 | 0 | 0.036176 | 0.005168 | 0 | 0.01292 | 0.005168 | 0.01292 | 0.046512 | 0.005168 | 0 | 0.007752 | 0 | 0.010336 |
| doc 4067 | 0 | 0.052632 | 0.035088 | 0.017544 | 0.017544 | 0 | 0 | 0 | 0.017544 | 0.087719 | 0.017544 | 0.035088 | 0.017544 | 0.035088 | 0 | 0.614035 | 0 | 0 | 0.052632 | 0 |
| doc 4068 | 0.026596 | 0.031915 | 0.005319 | 0.015957 | 0 | 0.143617 | 0.005319 | 0.026596 | 0.010638 | 0.026596 | 0.489362 | 0.010638 | 0.015957 | 0 | 0.026596 | 0.021277 | 0 | 0.12766 | 0.005319 | 0.010638 |
| doc 4069 | 0.017544 | 0 | 0.131579 | 0.008772 | 0 | 0.061404 | 0.026316 | 0.004386 | 0 | 0 | 0.078947 | 0.013158 | 0.057018 | 0.004386 | 0.179825 | 0.087719 | 0.307018 | 0.017544 | 0.004386 | 0 |
| doc 4070 | 0.02819 | 0.023739 | 0.008902 | 0.014837 | 0.037092 | 0.048961 | 0.005935 | 0.011869 | 0.01632 | 0.002967 | 0 | 0.007418 | 0.235905 | 0 | 0.002967 | 0.017804 | 0.360534 | 0.068249 | 0.008902 | 0.099407 |
| doc 4071 | 0.003165 | 0.003165 | 0.006329 | 0.142405 | 0.193038 | 0.003165 | 0 | 0 | 0.050633 | 0.452532 | 0 | 0.003165 | 0.009494 | 0.003165 | 0.009494 | 0.012658 | 0.003165 | 0.018987 | 0.079114 | 0.006329 |
| doc 4072 | 0 | 0 | 0.035242 | 0.008811 | 0.022026 | 0 | 0.026432 | 0.118943 | 0 | 0 | 0 | 0.008811 | 0.004405 | 0.590308 | 0 | 0.013216 | 0.07489 | 0.004405 | 0.092511 | 0 |

# Appendix 3. Preprocessed data randomly attached from Document 0 to Document 5818


| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Unnamed | article_he | article_se | article_lin | article_fir | article_las | article_su | article_bo | Date | Open2 | Close2 |
| 2 | doc 0 | BitcoinÂ b | Markets | https://w | 2022-08-2 | 2022-08-2 | Investors | bitcoin bri | 8/28/2022 | 20032.73 | 19606.74 |
| 3 | doc 1 | A closely- | Cryptocur | https://w | 2022-08-2 | 2022-08-2 | A measure | bitcoin pc | 8/24/2022 | 21513 | 21378.1 |
| 4 | doc 2 | Bitcoin ha | Cryptocur | https://w | 2022-08-2 | 2022-08-2 | Bitcoin ma | crypto wir | 8/25/2022 | 21367.72 | 21591.28 |
| 5 | doc 3 | Sudden cr | Crypto Wo | https://w | 2022-08-1 | 2022-08-2 | Bitcoin hit | bitcoin fri | 8/18/2022 | 23338.39 | 23206.19 |
| 6 | doc 4 | Ether is up | Technolog | https://w | 2022-08-1 | 2022-08-1 | Since find | since find | 8/18/2022 | 23338.39 | 23206.19 |
| 7 | doc 5 | Skybridge | Crypto | https://w | 2022-08-1 | 2022-08-1 | Bitcoin fu | bitcoin fu | 8/14/2022 | 24443.21 | 24323 |
| 8 | doc 6 | Bitcoin to | Cryptocur | https://w | 2022-08-1 | 2022-08-1 | Bitcoin bri | bitcoin bri | 8/14/2022 | 24443.21 | 24323 |
| 9 | doc 7 | BlackRock | Markets | https://w | 2022-08-1 | 2022-08-1 | The larges | blackrock | 8/10/2022 | 23145.39 | 23940.2 |
| 10 | doc 8 | MicroStrat | Crypto Wo | https://w | 2022-08-0 | 2022-08-0 | MicroStrat | microstrat | 8/1/2022 | 23311.56 | 23311.59 |
| 11 | doc 9 | Bitcoin bri | Cryptocur | https://w | 2022-07-2 | 2022-07-2 | Bitcoin's r | bitcoin to | 7/28/2022 | 22955 | 23847.01 |
| 12 | doc 10 | Tesla has | Technolog | https://w | 2022-07-2 | 2022-07-2 | Tesla CEO | early year | 7/19/2022 | 22517.1 | 23407.35 |
| 13 | doc 11 | Bitcoin jur | Markets | https://w | 2022-07-2 | 2022-07-2 | Bitcoin ros | bitcoin ros | 7/27/2022 | 21279.62 | 22943.12 |
| 14 | doc 12 | Crypto mi | Crypto Wo | https://w | 2022-07-1 | 2022-07-1 | New data | data block | 7/17/2022 | 21207.85 | 20769.4 |
| 15 | doc 13 | Bitcoin to | Cryptocur | https://w | 2022-07-1 | 2022-07-1 | Bitcoin bo | bitcoin bo | 7/17/2022 | 21207.85 | 20769.4 |
| 16 | doc 14 | Bitcoin to | Technolog | https://w | 2022-07-2 | 2022-07-2 | Bitcoin to | bitcoin br | 7/19/2022 | 22517.1 | 23407.35 |
| 17 | doc 15 | Bitcoin Fa | Crypto Wo | https://w | 2022-07-0 | 2022-07-0 | The "Bitco | bitcoin far | 7/1/2022 | 19922.82 | 19336.1 |
| 18 | doc 16 | Worldâ‚¬â | ETF Edge | https://w | 2022-07-0 | 2022-07-0 | Grayscale | digital cur | 7/6/2022 | 20209.98 | 20557.54 |
| 19 | doc 17 | For bitcoi | Technolog | https://w | 2022-07-1 | 2022-07-1 | Industry p | improvem | 7/13/2022 | 19330.03 | 20187.15 |
| 20 | doc 18 | Coinbase | Technolog | https://w | 2022-07-1 | 2022-07-1 | Coinbase | share coir | 7/17/2022 | 21207.85 | 20769.4 |
| 21 | doc 19 | El Salvado | Crypto Wo | https://w | 2022-06-2 | 2022-06-2 | The gover | salvador e | 6/24/2022 | 21097.66 | 21279.53 |
| 22 | doc 20 | Billions in | Crypto Wo | https://w | 2022-07-0 | 2022-07-1 | The bad d | crypto len | 7/5/2022 | 20219.67 | 20185.96 |
| 23 | doc 21 | Bitcoin he | Markets | https://w | 2022-07-0 | 2022-07-0 | Bitcoin is | price bitco | 7/7/2022 | 20562.57 | 21647.17 |
| 24 | doc 22 | Bitcoin jus | Crypto Wo | https://w | 2022-06-3 | 2022-06-3 | Bitcoin jus | bitcoin fir | 6/29/2022 | 20278.9 | 20101.12 |
| 25 | doc 23 | Five reaso | Technolog | https://w | 2022-07-0 | 2022-07-0 | Bitcoin los | bitcoin wo | 6/30/2022 | 20117.92 | 19816.75 |
| 26 | doc 24 | Bitcoin po | Cryptocur | https://w | 2022-06-3 | 2022-06-3 | Bitcoin ha | bitcoin th | 6/29/2022 | 20278.9 | 20101.12 |
| 27 | doc 25 | Bitcoin fal | Cryptocur | https://w | 2022-06-3 | 2022-06-3 | Cypto hed | bitcoin th | 6/29/2022 | 20278.9 | 20101.12 |
| 28 | doc 26 | Charts sug | Mad Mone | https://w | 2022-06-2 | 2022-06-2 | "Bitcoin c | cnbcs cran | 6/22/2022 | 20722.23 | 19976.05 |
| 29 | doc 27 | ProShares | Markets | https://w | 2022-06-2 | 2022-06-2 | Eight mon | eight mor | 6/19/2022 | 19010.48 | 20554.9 |
| 30 | doc 28 | Bitcoin bri | Crypto Wo | https://w | 2022-06-2 | 2022-06-2 | Bitcoin fel | bitcoin bri | 6/28/2022 | 20715.2 | 20293.45 |
| 31 | doc 29 | Bitcoin bo | Cryptocur | https://w | 2022-06-2 | 2022-06-2 | The crypto | bitcoin jur | 6/19/2022 | 19010.48 | 20554.9 |
| 32 | doc 30 | Bitcoin bri | Crypto Wo | https://w | 2022-06-1 | 2022-06-1 | Bitcoin fel | bitcoin plu | 6/17/2022 | 20406.02 | 20468.21 |
| 33 | doc 31 | Bitcoin's p | Technolog | https://w | 2022-06-1 | 2022-06-1 | A $4 billio | lost billior | 6/13/2022 | 26599.66 | 22509.1 |
| 34 | doc 32 | Bitcoin co | Technolog | https://w | 2022-06-2 | 2022-06-2 | Ian Harnet | cryptos pa | 6/21/2022 | 20577.6 | 20739.94 |
| 35 | doc 33 | Bitcoin ha | Invest in Y | https://w | 2022-06-1 | 2022-06-1 | Bitcoin's r | cryptocurr | 6/14/2022 | 22456.78 | 22209.52 |
| 36 | doc 34 | New York | Crypto Wo | https://w | 2022-06-0 | 2022-06-0 | Lawmaker | following | 6/2/2022 | 29795.17 | 30441.47 |
| 37 | doc 35 | Bitcoin ca | Technolog | https://w | 2022-06-1 | 2022-06-1 | Bitcoin's p | selloff cry | 6/14/2022 | 22456.78 | 22209.52 |
| 38 | doc 36 | MicroStrat | Cryptocur | https://w | 2022-06-1 | 2022-06-1 | MicroStrat | aggressive | 6/14/2022 | 22456.78 | 22209.52 |
| 39 | doc 37 | Bitcoin rel | Markets | https://w | 2022-06-2 | 2022-06-2 | Bitcoin bo | bitcoin cli | 6/20/2022 | 20580.04 | 20672.8 |
| 40 | doc 38 | Cramer du | Crypto | https://w | 2022-06-1 | 2022-06-1 | CNBC's Jin | cnbcs cran | 6/12/2022 | 28415.54 | 26704.73 |
| 41 | doc 39 | Bitcoin dr | Crypto Wo | https://w | 2022-06-1 | 2022-06-1 | Factors in | bitcoin tu | 6/12/2022 | 28415.54 | 26704.73 |
| 42 | doc 40 | Human rig | Crypto Wo | https://w | 2022-06-0 | 2022-06-0 | Human rig | washingto | 6/7/2022 | 31380.14 | 31193.07 |
| 43 | doc 41 | Bitcoin dr | Cryptocur | https://w | 2022-06-1 | 2022-06-1 | BitcoinÂ b | bitcoin bri | 6/13/2022 | 26599.66 | 22509.1 |
| 44 | doc 42 | Binance p | Markets | https://w | 2022-06-1 | 2022-06-1 | Binance te | binance m | 6/12/2022 | 28415.54 | 26704.73 |
| 45 | doc 43 | Bitcoin bri | Crypto Wo | https://w | 2022-06-0 | 2022-06-0 | BitcoinÂ f | bitcoin sli | 6/6/2022 | 29930.46 | 31367.42 |
| 46 | doc 44 | Grayscale | Technolog | https://w | 2022-06-3 | 2022-06-3 | The U.S. S | grayscale | 6/29/2022 | 20278.9 | 20101.12 |
| 47 | doc 45 | Despite th | Empower | https://w | 2022-06-1 | 2022-06-1 | CNBC spol | even cryp | 6/12/2022 | 28415.54 | 26704.73 |
| 48 | doc 46 | Bitcoin sir | Technolog | https://w | 2022-06-1 | 2022-06-1 | Crypto inv | bitcoin dr | 6/17/2022 | 20406.02 | 20468.21 |
| 49 | doc 47 | Bitcoin ris | Crypto Wo | https://w | 2022-06-0 | 2022-06-0 | Bitcoin ros | bitcoin mo | 6/5/2022 | 29847.33 | 29909.21 |
| 50 | doc 48 | Bitcoin fal | Cryptocur | https://w | 2022-06-1 | 2022-06-1 | Bitcoin is | bitcoin fel | 6/15/2022 | 22174.8 | 22490.59 |
| 51 | doc 49 | MicroStrat | U.S. Marke | https://w | 2022-05-1 | 2022-05-1 | MicroStrat | microstrat | 5/11/2022 | 31000.98 | 28984.54 |
| 52 | doc 50 | Bitcoin wa | Next Gen | https://w | 2022-07-0 | 2022-07-0 | Bitcoin wa | bitcoin plu | 7/7/2022 | 20562.57 | 21647.17 |
| 53 | doc 51 | Bitcoin pr | Technolog | https://w | 2022-05-1 | 2022-05-1 | New resea | bitcoin mi | 5/17/2022 | 29833.05 | 30452.62 |
| 54 | doc 52 | Bitcoin bil | Technolog | https://w | 2022-06-2 | 2022-06-2 | Bankman- | central ba | 6/21/2022 | 20577.6 | 20739.94 |
| 55 | doc 53 | Coinbase | Markets | https://w | 2022-06-2 | 2022-06-2 | Coinbase | coinbase s | 6/21/2022 | 20577.6 | 20739.94 |
| 56 | doc 54 | A $3.5 bill | Crypto Wo | https://w | 2022-05-0 | 2022-05-1 | South Kor | multibillio | 5/8/2022 | 34062.89 | 33933.54 |
| 57 | doc 55 | New York | Crypto Wo | https://w | 2022-05-0 | 2022-05-0 | Lawmaker | state york | 5/4/2022 | 34062.89 | 39699.03 |
| 58 | doc 56 | Bitcoin bo | Markets | https://w | 2022-05-1 | 2022-05-1 | The price | price bitco | 5/12/2022 | 29044.26 | 29130.04 |
| 59 | doc 57 | Bitcoin co | Davos WE | https://w | 2022-05-2 | 2022-05-2 | Guggenhe | bitcoin dr | 5/22/2022 | 29445.06 | 30366.04 |
| 60 | doc 58 | Jack Dorse | Bitcoin | https://w | 2022-05-1 | 2022-05-1 | At Block's | block exe | 5/17/2022 | 29833.05 | 30452.62 |
| 61 | doc 59 | Luna Foun | Crypto Wo | https://w | 2022-05-0 | 2022-05-0 | The Luna | luna foun | 5/4/2022 | 34062.89 | 39699.03 |
| 62 | doc 60 | This real e | Crypto De | https://w | 2022-04-2 | 2022-04-2 | Compass | taing busi | 4/22/2022 | 34062.89 | 39757.55 |
| 63 | doc 61 | Bitcoin inv | Crypto Wo | https://w | 2022-05-1 | 2022-05-1 | Investors | investor b | 5/9/2022 | 34062.89 | 30363.87 |
| 64 | doc 62 | Warren Bu | 2022 Berk | https://w | 2022-04-3 | 2022-05-0 | Despite a | bitcoin ste | 4/29/2022 | 34062.89 | 38592 |
| 65 | doc 63 | $3 billion | Crypto Wo | https://w | 2022-05-1 | 2022-05-1 | Luna Foun | investor e | 5/15/2022 | 30076.93 | 31333.77 |
| 66 | doc 64 | Grayscale | Finance | https://w | 2022-05-1 | 2022-05-1 | A spot-ba | grayscale | 5/10/2022 | 30106.43 | 31024.7 |
| 67 | doc 65 | Crypto ex | Technolog | https://w | 2022-06-2 | 2022-06-2 | CoinFlex i | digital ass | 6/28/2022 | 20715.2 | 20293.45 |
| 68 | doc 66 | El Salvado | Crypto Wo | https://w | 2022-05-0 | 2022-06-2 | El Salvado | salvador a | 5/8/2022 | 34062.89 | 33933.54 |
| 69 | doc 67 | 40% of bit | Crypto Wo | https://w | 2022-05-0 | 2022-05-1 | Bitcoin is | bitcoin ne | 5/8/2022 | 34062.89 | 33933.54 |
| 70 | doc 68 | Bitcoin dr | Cryptocur | https://w | 2022-05-0 | 2022-05-0 | Bitcoin cra | bitcoin br | 5/5/2022 | 34062.89 | 36568.76 |
| 71 | doc 69 | Fort Wort | Crypto Wo | https://w | 2022-04-2 | 2022-04-2 | Three Bitr | fort worth | 4/25/2022 | 34062.89 | 40475.28 |
| 72 | doc 70 | Bitcoin dr | Crypto Wo | https://w | 2022-05-0 | 2022-05-0 | Cryptocur | bitcoin dr | 5/8/2022 | 34062.89 | 33933.54 |
| 73 | doc 71 | Michael S | Crypto Wo | https://w | 2022-04-2 | 2022-04-2 | MicroStrat | miami mic | 4/20/2022 | 34062.89 | 41363.5 |
| 74 | doc 72 | Bitcoin dr | Crypto Wo | https://w | 2022-05-1 | 2022-05-1 | The price | bitcoin dr | 5/10/2022 | 30106.43 | 31024.7 |
| 75 | doc 73 | Itâ‚¬â„¢s a | ETF Edge | https://w | 2022-04-1 | 2022-04-1 | Crypto ex | crypto exp | 4/16/2022 | 34062.89 | 40386.51 |
| 76 | doc 74 | What to c | Invest in Y | https://w | 2022-04-2 | 2022-04-2 | Before ad | investor s | 4/25/2022 | 34062.89 | 40475.28 |
| 77 | doc 75 | Bitcoin jur | Crypto Wo | https://w | 2022-05-0 | 2022-05-0 | The crypto | price bitco | 5/3/2022 | 34062.89 | 37732.3 |
| 78 | doc 76 | Bitcoin bri | Crypto Wo | https://w | 2022-05-1 | 2022-05-1 | Bitcoin fel | bitcoin dr | 5/9/2022 | 34062.89 | 30363.87 |
| 79 | doc 77 | Bitcoin dr | Cryptocur | https://w | 2022-05-0 | 2022-05-0 | The crypto | bitcoin co | 5/7/2022 | 34062.89 | 35552.56 |
| 80 | doc 78 | Bitcoin Piz | Davos WE | https://w | 2022-05-2 | 2022-05-2 | The indus | davos swi | 5/23/2022 | 30295.93 | 29114.08 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5792 | doc 5790 | 5 signs yo | Money\|M | https://w | 2018-01-2 | 2018-01-2 | Here's ho | horrible h | 1/25/2018 | 34062.89 | 11219.44 |
| 5793 | doc 5791 | If you inve | The Begin | https://w | 2018-01-1 | 2018-01-1 | Google's s | search en | 1/10/2018 | 34062.89 | 15065 |
| 5794 | doc 5792 | Travis Kal | Money\|M | https://w | 2018-01-0 | 2018-01-0 | Wealth m | travis kala | 1/7/2018 | 34062.89 | 16515.29 |
| 5795 | doc 5793 | Meet the | Careers\|M | https://w | 2018-01-1 | 2018-01-1 | Millennial | werent da | 1/9/2018 | 34062.89 | 14710.6 |
| 5796 | doc 5794 | The 10 mc | Careers\|M | https://w | 2018-01-2 | 2018-01-2 | Education | though co | 1/22/2018 | 34062.89 | 10917.31 |
| 5797 | doc 5795 | Legendary | Entrepren | https://w | 2018-01-2 | 2018-01-2 | Bill Miller | legendary | 1/21/2018 | 34062.89 | 11481.03 |
| 5798 | doc 5796 | Men most | Money\|M | https://w | 2018-01-1 | 2018-01-1 | A new GO | financial r | 1/15/2018 | 34062.89 | 13769.22 |
| 5799 | doc 5797 | Make It's | Money\|M | https://w | 2017-12-3 | 2017-12-3 | These we | bitcoin cra | 12/30/2017 | 34062.89 | 12732.36 |
| 5800 | doc 5798 | Self-made | Money\|M | https://w | 2018-01-0 | 2018-01-0 | If you hav | want supe | 1/4/2018 | 34062.89 | 15163.9 |
| 5801 | doc 5799 | GOP tax p | Money\|M | https://w | 2017-12-2 | 2017-12-2 | "More tha | middle cla | 12/19/2017 | 34062.89 | 17700.78 |
| 5802 | doc 5800 | 3 commor | Careers\|M | https://w | 2017-12-2 | 2017-12-2 | Your essay | college ap | 12/20/2017 | 34062.89 | 16599.69 |
| 5803 | doc 5801 | Mark Cuba | Money\|M | https://w | 2017-12-2 | 2017-12-2 | "Shark Tar | year oppo | 12/25/2017 | 34062.89 | 13965.21 |
| 5804 | doc 5802 | The 8 best | Entrepren | https://w | 2017-12-1 | 2017-12-1 | Here are t | heart shar | 12/18/2017 | 34062.89 | 18931.2 |
| 5805 | doc 5803 | Vanguard | Money\|M | https://w | 2017-12-1 | 2017-12-1 | "I don't lik | jack bogle | 12/12/2017 | 34062.89 | 16798.03 |
| 5806 | doc 5804 | Here's hov | Careers\|M | https://w | 2017-12-2 | 2017-12-2 | The bill cc | morning s | 12/19/2017 | 34062.89 | 17700.78 |
| 5807 | doc 5805 | Here's hov | Money\|M | https://w | 2017-11-1 | 2017-11-1 | With Andr | billionaire | 11/14/2017 | 34062.89 | 6647.346 |
| 5808 | doc 5806 | Departme | Careers\|M | https://w | 2017-10-2 | 2017-10-2 | Criminals | safe cyber | 10/23/2017 | 34062.89 | 5946.927 |
| 5809 | doc 5807 | Mark Cuba | Money\|M | https://w | 2017-10-2 | 2017-12-2 | Even billic | come sma | 10/20/2017 | 34062.89 | 6000.943 |
| 5810 | doc 5808 | Success ha | Leadershi | https://w | 2017-11-0 | 2017-11-0 | Not every | charles bu | 11/1/2017 | 34062.89 | 6751.17 |
| 5811 | doc 5809 | Billionaire | Entrepren | https://w | 2017-10-2 | 2017-10-2 | â,¬Å“I wo | peter thie | 10/25/2017 | 34062.89 | 5725.302 |
| 5812 | doc 5810 | Mark Cuba | Money\|M | https://w | 2017-10-2 | 2017-10-2 | It's "a far k | prefer pay | 10/24/2017 | 34062.89 | 5526.929 |
| 5813 | doc 5811 | The 9 sure | Leadershi | https://w | 2017-09-0 | 2017-09-0 | "Take any | resume te | 9/4/2017 | 34062.89 | 4224.22 |
| 5814 | doc 5812 | Looks like | Money\|M | https://w | 2017-08-1 | 2017-11-1 | "We live i | rising hea | 8/10/2017 | 34062.89 | 3391.23 |
| 5815 | doc 5813 | For the bu | Careers\|M | https://w | 2017-06-2 | 2017-06-2 | Traveling | corporate | 6/20/2017 | 34062.89 | 2718.49 |
| 5816 | doc 5814 | This start- | South by S | https://w | 2017-03-1 | 2017-03-1 | A former | marijuana | 3/14/2017 | 34062.89 | 1240.06 |
| 5817 | doc 5815 | Want to tr | Money\|M | https://w | 2017-03-2 | 2017-03-2 | An easy, a | brick farm | 3/19/2017 | 34062.89 | 1038.05 |
| 5818 | doc 5816 | 'Shark Tan | Careers\|M | https://w | 2017-02-2 | 2017-02-2 | The cyber | tempting | 2/27/2017 | 34062.89 | 1179.03 |
| 5819 | doc 5817 | The Five F | Money\|M | https://w | 2016-10-2 | 2016-10-2 | If you and | bought so | 10/26/2016 | 34062.89 | 678.486 |
| 5820 | doc 5818 | Turn your | How I Mac | https://w | 2016-05-0 | 2016-07-0 | The Collis | john collis | 5/8/2016 | 34062.89 | 458.6 |

## References


Blei, D.M.; Kucukelbir, A. & McAuliffe, J.D. (2017). "Variational inference: A review for statisticians". *Journal of the American Statistical Association*. 112(518): 859-877. https://doi.org/10.1080/01621459.2017.1285773.

Buchholz, M.; Delaney, J.; Warren, J. & Parker, J. (2012). *Bits and Bets, Information, Price Volatility, and Demand for Bitcoin*. Economics. 312. https://www.reed.edu/economics/parker/s12/312/finalproj/Bitcoin.pdf.

Clinton, J.; Jackman, S. & Rivers, D. (2004). "The statistical analysis of roll call data". *American Political Science Review*. 98(2): 355-370. https://doi.org/10.1017/S0003055404001194.

Dawson, J. & Kendziorski, C. (2012). "Survival-supervised latent Dirichlet allocation models for genomic analysis of time-to-event outcomes". *arXiv*. TR 225. https://doi.org/10.48550/arXiv.1202.5999.

Eklund, M. & Bejerholm, U. (2004). "Time use and occupational performance among persons with schizophrenia". *Occupational Therapy in Mental Health*. 20: 27-47. https://doi.org/10.1300/J004v20n01_02.

Erosheva, E.; Fienberg, S. & Lafferty, J. (2004). "Mixed-membership models of scientific publications". *Proceedings of the National Academy of Science*. 101(1): 5220-5227. https://doi.org/10.1073/pnas.0307760101.

Fataliyev, K.; Chivukula, A.; Prasad, M. & Liu, W. (2021). "Text-based stock market analysis: A review". 1(1). July. https://arxiv.org/pdf/2106.12985.

Frank, X.; Cambria, E. & Welsch, R.E. (2017). "Natural language based financial forecasting: A survey". *Artificial Intelligence Review*. 50(3): 49-73. https://link.springer.com/article/10.1007/s10462-017-9588-9.

Kaya, M.Y. & Karsligil M.E. (2010). "Stock price prediction using financial news articles". *2nd IEEE International Conference on Information and Financial Engineering*. IEEE: 478-482.

Kumar, G.; Jain, S. & Singh, U.P. (2020). "Stock market forecasting using computational intelligence: A survey". *Archives of Computational Methods in Engineering*. 28(6): 1-33. http://dx.doi.org/10.1007/s11831-020-09413-5.

Li, Q.; Chen, Y.; Wang, Y.; Chen, Y. & Chen, H. (2018). "Web media and stock markets: A survey and future directions from a big data perspective". *IEEE Transactions on Knowledge and Data Engineering*. 30: 381-399. http://dx.doi.org/10.1109/TKDE.2017.2763144.

Loughran, T.; Mcdonald, B. & Pragidis, I. (2019). "Assimilation of oil news into prices". *International Review of Financial Analysis*. 63. https://doi.org/10.1016/j.irfa.2019.03.008.

Mckinney, W. (2010). “Data structures for statistical computing in Python”. *Pyton in Science Conference*. http://dx.doi.org/10.25080/Majora-92bf1922-00a.

Mohan, S.; Mullapudi, S.; Sammeta, S.; Vijayvergia, P. & Anastasiu, D. (2019). “Stock price prediction using news sentiment analysis”. *IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. 205-208. http://dx.doi.org/10.1109/BigDataService.2019.00035.

Pang, B. & Lee, L. (2008). “Opinion Mining and Sentiment Analysis”. *Found. Trends Inf. Retr.* 2(1-2): 1-135. http://dx.doi.org/10.1561/1500000011.

Perotte, A.; Bartlett, N.; Elhadad, N. & Wood, F. (2011). “Hierarchically supervised latent dirichlet allocation”. *Neural Information Processing Systems*. 24. https://www.researchgate.net/publication/228449895_Hierarchically_Supervised_Latent_Dirichlet_Allocation.

Sahut, J.; Hájek, P.; Olej, V. & Hikkerova, L. (2024). “The role of news-based sentiment in forecasting crude oil price during the Covid-19 pandemic”. *Annals of Operations Research*. 345(2): 861-884. http://dx.doi.org/10.1007/s10479-024-05821-z.

Schofield, A.; Magnusson, M.; Thompson, L. & Mimno, D. (2017). “Understanding text pre-processing for latent dirichlet allocation”. *EMNLP*. https://www.cs.cornell.edu/~xanda/winlp2017.pdf.

Shah, D.; Isah, H. & Zulkernine, F. (2019). “Stock market analysis: A review and taxonomy of prediction techniques”. *International Journal of Financial Studies*. 7(2): 26. https://doi.org/10.3390/ijfs7020026.

Sharma, R.K. (2020). “Comparison of stock price prediction models using news articles, currency exchange rates and global indicator performance”. *Journal of Advanced Research in Dynamical and Control Systems*. 12(7). http://doi.org/10.5373/JARDCS/V12SP7/20202273.

Thakkar, A. & Chaudhari, K. (2021). “Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions”. *Information Fusion*. 65: 95-107. https://doi.org/10.1016/j.inffus.2020.08.019.

Wilcox, K.T.; Jacobucci, R.; Zhang, Z. & Ammerman, B.A. (2021). “Supervised latent dirichlet allocation with covariates: A bayesian structural and measurement model of text and covariates”. *Psychological Methods*. 28: http://dx.doi.org/10.31234/osf.io/62tc3.

Yao, W.; Xu, K. & Li, Q. (2019). *Exploring the influence of news articles on Bitcoin Price with Machine Learning*. Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. http://dx.doi.org/10.1109/ISCC47284.2019.8969596.

Yap, A.Y.; Schumaker, R. & Chen, H. (2012). “Predicting Stock price movement from financial news articles”. *Information Systems for Global Financial Markets*. http://dx.doi.org/10.4018/978-1-61350-162-7.ch006.

Zang, C. & Kjellström, H. (2012). “How to supervise topic models”. *European Conference on Computer Vision*. pp. 500-515. Springer.