Applied Mathematics and Computational Intelligence
Volume 12, No. 2, Aug 2023 [59-119]

Peer-to-peer Online Lending Sentiment Analysis

Reem Saleh Ali[1], Khai Wah Khaw[2]*, Amira Gaber[3], Alhamzah Alnoor[4]

[1,2]School of Management, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[3]System and Biomedical Engineering Department, Cairo University, Egypt
[4]Management Technical College, Southern Technical University, Basrah, Iraq

* Corresponding author : khaiwah@usm.my

Received: 27 Jan 2023
Accepted: 23 June 2023

ABSTRACT

Online Peer-to-Peer Lending has emerged as a popular fintech application in recent years, captivating numerous individuals with its user-friendly approach and the absence of stringent regulations typically associated with traditional banks. Consequently, this rise in popularity has piqued the interest of researchers, prompting them to explore the phenomenon from various angles. This study specifically focuses on sentiment analysis and topic modeling techniques to gain insights into people's perceptions of online lending. The ubiquity of the internet has turned it into an everyday tool, enabling people to freely share recommendations and warnings on diverse subjects. Leveraging RapidMiner, the researchers extracted data from Twitter, obtaining up to 500 rows per keyword. Subsequently, the same tool was employed for data preprocessing and text preparation, followed by sentiment analysis and topic modeling on the extracted data. The analysis led to a dataset of 1800 rows, used for evaluating the sentiment analysis model through cross-validation, revealing an impressive 92% accuracy for the chosen model. The findings indicated that the majority of individuals hold positive opinions about most Peer-to-Peer lending online platforms.

Keywords: Lending Online, Peer-to-Peer, RapidMiner, Sentiment Analysis, Topic Modelling

1 INTRODUCTION

The Industrial Revolution 4.0 gives paramount importance to the swift progress of technology, largely supported by Artificial Intelligence (AI). This has become a focal point of discussion for every nation, with a wide array of technological advancements already underway or in prospect during this revolution. Notably, the financial or banking industry is experiencing significant changes [1]. The financial sector is rapidly embracing digitization and interconnectedness, with real-time payments infrastructure reshaping its operations. Advancements in ICT technologies like BigData, the Internet of Things (IoT), AI, blockchains, mobile apps, cloud services, and web infrastructures integrated with financial technology innovations have resulted in a notable increase in financial transactions. While these advancements benefit both customers and financial institutions, they also bring forth increased risks and vulnerabilities for financial services. Consequently, the attack surface broadens, making attack detection more challenging. As a result, the financial sector is now recognized as one of the most vulnerable infrastructures to cyber threats [2].

Fintech represents a breakthrough in financial services, employing technology to enhance accessibility and reduce barriers for people seeking financial products and services. The progression of fintech is reshaping the financial business model, enabling it to emulate the operations of regulated institutions, despite operating with a lower barrier to entry and potentially leading to unregulated behaviors.

1.1 Background

Starting in 2008, the collapse of the financial system shattered the public's trust in the traditional intermediaries, the regulated banks. This not only led to the implosion of the mainstream financial system, burdening countless borrowers with substantial debts, but also resulted in limited access to new credit sources for individuals and small businesses. As a response to this crisis, the concept of "disintermediation," obtaining credit without involving banks, gained political momentum among those advocating for financial system reforms that prioritize consumer interests. Peer-to-peer lending companies, aiming to "revolutionize credit by eliminating banks from the lending process," emerged as a promising solution according to the Financial Times. While the amount of credit offered through peer-to-peer lending remains modest compared to traditional channels, public interest in this alternative approach remains high [3].

Peer-to-peer lending operates much like the banking sector, offering financial transactions to the general public. Additionally, due to their shared objective of promoting financial inclusion, fintech peer-to-peer lending has the potential to challenge the traditional banking model. This emerging form of lending is generating significant interest among potential users, including both lenders and individuals seeking investment opportunities. Consequently, peer-to-peer lending presents an alternative financial solution for entrepreneurs aiming to expand their enterprises, thus establishing it as a novel banking alternative [1].

Financial fraud presents a significant worldwide concern, creating substantial challenges in both well-established and developing economies. In 2018, the estimated global expenses attributed to fraud amounted to a staggering $4 trillion (approximately $12,000 per person in the US). Developing markets, lacking the necessary sophisticated legal and regulatory expertise to effectively combat financial crimes, may face market meltdowns as a consequence. Adding to the complexity, financial crime regulation is further hindered by varying country-specific political regimes, regulatory cultures, and government agency structures. Additionally, technological advancements have exacerbated the issue by facilitating the rapid spread of fraudulent schemes and financial transactions over the Internet. Considering the pervasive nature of white-collar crime, conducting comparative research becomes indispensable in comprehending these offenses within a globalized context [4].

According to statistics, borrowing through P2P platforms is consistently on the rise, particularly in certain countries [5]. CNBC (2021) reported that peer-to-peer lending is gaining global momentum, with the industry experiencing a steady 17% annual growth rate and showing promising prospects for future expansion [6].

1.2 Goal of Study

This paper examines users' viewpoints regarding the website. The website's decision-making can influence users' perspectives and, consequently, their choices in using the site. Conducting sentiment analysis will yield insights into how users perceive the website, and whether user-generated social media data will have a positive or negative impact on the website's reputation.

Peer-to-peer (P2P) lending has made its way into the market, allowing borrowers to access loans directly from individuals, often with more favorable terms, especially for those with a poor credit history. Consequently, numerous users now prefer microblogging platforms like Twitter to express their satisfaction with the website and its services. Among the plethora of online lending websites available, this paper focuses on the top six platforms selected based on the "money under 30" websites, which cover topics like loans, banking, and investing. Notably, Prosper emerges as the leading platform due to its high credit scores, followed closely by BlockFi, Upstart, SOLO Funds, FundingCircle, and Kiva, each serving a distinct purpose. The paper will delve into an analysis of these six platforms, leveraging big data from Twitter to explore their impact on the websites' performance.

This study aims to analyze public opinion about P2P lending online platforms by using Twitter as a data source and RapidMiner to obtain a mining process. Specifically, the objectives of this study are twofold:

1- To determine the sentiment of users' opinions towards P2P lending online platforms and evaluate the performance of the sentiment analysis model using cross-validation. The study will use the confusion matrix to analyze the accuracy, precision, recall, and F1-score of the sentiment analysis model.

2- To identify the key topics discussed by users related to P2P lending online platforms using topic modeling and word cloud visualization.

The main objective of this study is to examine the impact of a claim stating that the social lending market will reach 705.81 billion by 2031 on people's perceptions of peer-to-peer lending. To achieve this, the researchers utilized RapidMiner to conduct sentiment analysis. Sentiment analysis has been a valuable tool for businesses for many years, enabling them to grasp trends, address customer feedback, and assess data from social media platforms like Facebook, Instagram, and Twitter. These platforms serve as hubs where people openly share their thoughts and opinions about the quality of services and the benefits they receive. Leveraging social media can help businesses identify issues, enhance their services, and target the appropriate audience effectively. Thus, this paper aims to offer lending platforms a comprehensive understanding of their customers' sentiments, allowing them to cater to their needs more effectively. Further investigation is necessary to gain insights into users' perspectives on flourishing platforms such as Prosper, BlockFi, Upstart, SOLO Funds, FundingCircle, and Kiva.

2 LITERATURE REVIEW

Recently, there has been rapid growth in Online Peer-to-Peer lending, with many individuals seeking alternatives to traditional banks due to the ease and user-friendliness of these platforms.

Consequently, extensive research has been conducted on social lending, focusing on analyzing the Fintech Business Model. This analysis encompasses various aspects such as default risk, interest rates, and repayment predictions. Furthermore, researchers have delved into risk control, creditworthiness, credit scores, and overall risk evaluation. Additionally, they have explored opportunity costs, the time it takes to receive returns, and repayment predictions. The research has resulted in the development of several models to aid borrowers and lenders, enabling lenders to make informed investment choices and borrowers to select suitable loans. Recommendations systems have also been designed to assist borrowers in finding the most favorable loan offers. The studies have investigated factors influencing the acquisition of loans and identified key factors that significantly impact Peer-to-Peer lending. Lastly, sentiment analysis has been employed to understand the impact of words on lenders and borrowers, influencing their decision-making when applying for investments or loans.

2.1 Analyzing Fintech Business Model

The research investigates the impact of opportunity cost on a FinTech business model involving various entities in the supply chain. These entities include a manufacturer, other supply chain participants, and parameters associated with them. The study introduces an online peer-to-peer lending platform, two manufacturers, a retailer, and the supply chain FinTech business model. It analyzes both Stackelberg and Nash equilibrium strategies for the supply chain participants. Specifically, the study explores a pre-selling strategy aimed at raising funds for a producer offering subpar goods. Through pre-selling, manufacturers of low-quality products can seize opportunities with retailers. One of the primary objectives of the research is to assess how pre-selling contributes to increased profits for supply chain participants. In this context, the study determines crucial factors such as the service level of the online P2P platform, the percentage of the retailer's pre-ordering quantity, and the optimal prices for all participants [7]. To further enhance the model, the research proposes incorporating three key actors: a manufacturer acting as the Stackelberg leader, a P2P lending platform as the sub-leader, and a capital-constrained retailer as the follower. The objective of this model is to identify the optimal lending platform service rate [8].

2.2 Analyzing the Reasons Behind the Rise in Default.

Examine the impact of asymmetric information on China's peer-to-peer lending market, revealing noteworthy trends. Over time, both interest rates and default rates experience substantial growth. The rise of interest rates beyond the legal limit of 15.4% for private lending leads to a higher likelihood of borrower defaults. The absence of interest rate caps, which could prevent adverse selection, emerges as a key factor contributing to the collapse of China's peer-to-peer lending market. However, the evidence does not strongly support the moral hazard effect concerning interest rates and loan size. Moreover, utilizing applicant characteristics-based credit scoring shows promise in mitigating the asymmetric information problem, though it does not entirely resolve it [9].

A novel approach for peer-to-peer lending default prediction was introduced, leveraging soft information extracted from textual descriptions. The method involved developing a topic model to extract valuable features from loan descriptions, followed by creating four default prediction models to showcase the effectiveness of these features. Additionally, a two-stage method was employed to curate an optimal feature set encompassing both soft and hard data. Analyzing real-world data from a prominent P2P lending platform in China, the proposed method demonstrated superior performance compared to existing approaches that relied solely on hard information for loan default prediction [10].

Data from Lending Club (LC) was collected, covering the period from the third quarter of 2018 to the second quarter of 2019. Subsequently, the collected data underwent analysis using several statistical and AI models. The statistical models included Logistic Regression, Bayesian Classifier, and Linear Discriminant Analysis (LDA), while the AI models comprised Decision Tree, Random Forest, LightGBM, Artificial Neural Network (ANN), and Convolutional Neural Network (CNN). Evaluation of these models involved employing various metrics such as the confusion matrix series, the AUC-ROC curve, the Kolmogorov-Smirnov chart (KS), and the student's t-test.

Empirical studies revealed that LightGBM outperformed the other models, exhibiting a notable 2.91% increase in accuracy. This improvement resulted in a substantial revenue surge of nearly USD 24 million for Lending Club [11]. Additionally, a Boosted Tree classification approach was employed to predict delinquency likelihood in the LC dataset. The results indicated that XGboost demonstrated exceptional performance, achieving 99% accuracy on both the training and test sets [12].

In late 2016, LightGBM and XGboost, two contemporary machine learning algorithms, were developed in Asia using real P2P transaction data from LC. These algorithms effectively predicted the default risk on the platform in an innovative manner. A comparison of their outcomes revealed that the LightGBM algorithm, leveraging multiple observational datasets, achieved the most accurate classification predictions. As a result, the average performance rate of LC's historical transaction data increased by 1.28 percentage points, leading to a significant $117 million reduction in loan defaults. The study concludes by offering recommendations for LC and other P2P platforms, as well as suggesting future directions for development in this field across different countries, considering the factors that influence default rates [13].

Using TensorFlow, an empirical analysis was conducted on loan data issued by LC between 2007 and 2015, categorized into three classes: safe loans, risky loans, and bad loans. Due to an imbalanced distribution with the majority of data in the safe loan class, the Synthetic Minority Over-Sampling Technique (SMOTE) was employed to enhance the accuracy of DNN predictions. The results show that the proposed approach achieved an impressive test data classification accuracy of 93%, significantly surpassing the 75% accuracy obtained using MLP with just one hidden layer [14].

An investment decision-making framework was developed using Artificial Neural Networks (ANN) and Logistic Regression to estimate the internal rate of return and default probability for loans in the LC dataset. The study concludes that adopting the profit scoring model can lead to higher profits, but it also increases the investment's risk. Additionally, a sensitivity analysis was conducted to assess the model's responsiveness to changes in the total investment amount [15].

Furthermore, the research utilized machine learning techniques to calculate the platform's default risk and compared different models' predictive capabilities. Three types of variables were employed to evaluate a platform's default risk: operating characteristics, customer feedback, and compliance capability. The findings revealed that abnormal returns significantly elevate the risk of default. Nevertheless, platforms with positive customer recommendations and transparent information disclosure can mitigate default risk. Among the models tested, the CART model outperformed both Random Forests and Logit regression models. These empirical results provide valuable insights into default risk prediction, aiding the government in more effective regulation [16].

The predictive capability of XGBoost in determining P2P loan defaults was confirmed. To achieve this, they utilized LC's loan data and feature engineering techniques to create a P2P loan default prediction model using the XGBoost algorithm. To assess its performance, they compared it with Logistic Regression and Decision Trees using five metrics: accuracy, AUC value, error rate, model robustness, and model run time. The results revealed that the XGBoost algorithm achieved a prediction accuracy rate of 97.705%, outperforming the other models and effectively minimizing loss costs due to prediction errors. Additionally, leveraging the XGBoost algorithm, they identified the top ten features with the greatest influence on loan default rates [17].

Using the LightGBM algorithm, data from Renrendai, a prominent platform in the P2P industry, was analyzed. The study explored both the fundamental LightGBM model and its integration with linear blending to create an optimal model for identifying default risk. The proposed technique is applicable to diverse multidimensional data samples. The combined approach of LightGBM and linear blending achieved impressive precision (91.36%), recall (75.90%), and accuracy (84.36%) values. In comparison to conventional machine learning models, the established LightGBM algorithm proves highly efficient in recognizing loan default on the P2P platform. Additionally, the LightGBM algorithm demonstrates effective judgment of default risk for a vast array of users on P2P platforms, encompassing numerous multidimensional data samples [18].

2.3 Models that help borrowers and Lenders.

Researchers carried out an empirical investigation utilizing public datasets sourced from Paipaidai, the largest online P2P lending platform in China. Their aim was to explore the determinants affecting the probability of securing a loan through online P2P lending. The findings revealed that various factors, including the annual interest rate, repayment period, description, credit grade, successful loan number, failed loan number, gender, and borrowed credit score, significantly influence the success of loan funding on the Paipaidai platform [19].

An iterative computation model was developed for evaluating unknown loans. Extensive experiments were conducted using real-world data from Prosper, the largest American P2P lending marketplace, to validate this proposed model. The results of the experiments showed that the computation model was effective in aiding borrowers to select good loans and helping lenders make sound investment decisions. Additionally, the experimental findings revealed that integrating the Logistic classification model with the iterative computation model created a hybrid classification model, which demonstrated that the two models complemented each other. Based on the results, the researchers concluded that the hybrid model proved to be more efficient and stable compared to either individual model [20].

An innovative reject inference model called OD-LightGBM is introduced, integrating an outlier detection technique (isolation forest) with a state-of-the-art gradient-boosting decision tree algorithm. The model's performance is evaluated using real-world P2P lending datasets, LC and We.com, demonstrating superior predictive capabilities compared to benchmark models. The proposed framework exhibits robustness across various parameter settings, consistently yielding stable results while effectively mitigating information asymmetry [21].

Presenting a novel and transparent approach for establishing peer-to-peer lending platforms based on blockchain technology, ensuring trust and credibility. This proposed method utilizes a distributed ledger to record cryptographically secured transactions, creating an immutable and publicly verifiable record of all security-related aspects [22]. Additionally, a borrower recommendation system is developed to enhance funding opportunities while reducing interest rates. The methodology involves three main steps [23]:

1. Employing an RF with a recursive borrower feature selection model to predict bidding interest rates.

2. Predicting the success rate of bidding loans and improving the sentiment analysis of borrowing reasons.

3. Comparing the results to determine the best loan option for borrowers. Remarkably, the proposed method surpasses the current state-of-the-art technique, achieving an impressive accuracy increase from 67% to 91% in predicting the success rate of bidding loans.

An innovative approach is presented in this study, combining multifractal dimension (MFD), probit regression, artificial prior knowledge, and fireworks coevolution binary glowworm swarm optimization (FCBGSO) to address investment risk in P2P lending. Initially, the relevant influencing factors are identified, and irrelevant attributes are eliminated using probit regression. To enhance the subset of influencing factors, artificial prior knowledge is employed, progressively adding extracted factors from the original dataset of P2P lending. The resulting subsets are concise and easily manageable. Employing an extreme learning machine, the subset with the highest classification accuracy is determined. Experimental results on the Renrendai platform's P2P lending dataset demonstrate the superiority of this proposed approach over other state-of-the-art methods in terms of validity and effectiveness [22].

Presented is a goal programming approach designed to construct an optimal P2P lending portfolio that takes into account both anticipated returns and the opportunity cost for individual investors. The initial step involves utilizing a logistic regression model to predict the likelihood of loan default for each loan proposal. Additionally, a Weibull regression is employed to assess the opportunity cost resulting from the time needed to secure the loan. By employing goal programming techniques, a portfolio is created to minimize the deviation between the desired return on investment and the surplus from the predetermined opportunity cost caused by an extended bidding period. The data obtained from these methods was subsequently applied to the Prosper platform [24].

2.4 Deep Learning and Machine Learning.

Multiple machine learning algorithms were evaluated for credit scoring in peer-to-peer lending, using a dataset sourced from LC's official website. The algorithms included single classifiers such as logistic regression, decision tree, and multilayer perceptron, as well as homogeneous ensembles like XGBoost, GBM, and Random Forest. Additionally, heterogeneous ensemble classifiers like Stacked Ensembles were employed. The results demonstrated that ensemble classifiers outperformed single classifiers, with Stacked Ensemble and XGBoost emerging as the top performers [25].

A Convolutional Neural Network (CNN) architecture was employed to predict repayment in P2P social lending, enabling automatic feature extraction and performance enhancement. By classifying the borrower's loan status through robust feature capturing and pattern learning, the method demonstrated effective repayment prediction in 5-fold cross-validation results. The standard CNN outperformed other machine learning methods with an accuracy of 75.86%. Moreover, utilizing different CNN models like Inception, ResNet, and Inception-ResNet achieved a state-of-the-art performance of 77.78%. The study indicated that projecting samples into the feature space further improved the performance of the extracted features by the model [26].

A three-stage game was proposed to examine the optimal risk control ability and corresponding prices of P2P lending platforms under various tariffs. To gauge the impact of risk control ability on prices, risk price coefficients were introduced for lenders and borrowers, wherein higher coefficients denoted increased sensitivity of prices to risk control ability. Additionally, the study explored the influence of platform scales in determining optimal risk control ability, prices, and market shares. The analysis revealed the following findings: In equilibrium, risk price coefficients were inversely related to optimal risk control ability. The risk-price coefficient for lenders had a relatively smaller effect on optimal risk control ability compared to that for borrowers if some lenders were multi-home. Moreover, under specific conditions, smaller platforms exhibited higher risk control ability and prices, and they attracted more borrowers than larger platforms [27].

Between 2007 and 2016, an investigation was conducted into the creditworthiness of LC borrowers. The study focused on the impact of supervised classification models and techniques used to address class imbalance in predicting creditworthiness rates. The researchers combined ensemble, cost-sensitive, and sampling methods with various models such as Logistic Regression, Decision Tree, and Bayesian learning models. The findings revealed that sampling techniques performed better than ensemble and cost-sensitive approaches [28]. To create a ranking score system, the researchers utilized the LC dataset from 2009 to 2013. They incorporated several models, including Logistic Regression, ANN, LDA, Linear SVM, RF, Bayesian net, radio SVM, Naïve Bayes, classification and regression tree, and k-Nearest Neighbor. Based on the results, Logistic Regression achieved the highest rank in terms of Percentage Correctly Classified [29].

Presented is an innovative data-analysis approach for the P2P market, involving the utilization of LDA-based feature selection combined with restricted Boltzmann machines (RBMs) to construct credit scoring models. The proposed credit risk model harnesses advanced machine learning techniques. Subsequently, credit datasets were employed to assess the model's performance. During the feature selection process, twelve features from the Australian credit dataset, twenty-two features from the German credit dataset, and eighty features from the LC dataset were selected. The resulting accuracy rates achieved were 86.09%, 76.7%, and 81.53%, respectively [30].

An innovative data-analysis method for the P2P market is presented, which involves using LDA-based feature selection along with restricted Boltzmann machines (RBMs) to create credit scoring models. The credit risk model proposed utilizes cutting-edge machine learning techniques. To evaluate the model's performance, credit datasets were employed. During the feature selection, twelve features from the Australian credit dataset, twenty-two features from the German credit dataset, and eighty features from the LC dataset were chosen. The achieved accuracy rates were 86.09%, 76.7%, and 81.53%, respectively [30].

2.5 Sentiment Analysis.

Sentiment analysis was performed on the IGROW platform's Google Play Store data, utilizing the K-Nearest Neighbor (K-NN) Algorithm and the Naive Bayes classification algorithm. The evaluation yielded accuracy, precision, and recall values of (73.85%, 76.60%, 85.71%) and (75.38%, 80.95%, and 80.95%) for K-NN and Naive Bayes classification, respectively. Notably, the Naive Bayes classification exhibited slightly superior accuracy and precision values compared to K-NN [31].

The study delved into ways in which borrowers on a P2P platform can enhance the probability of successfully securing funding for their loans. It identified a collection of loan description features and assessed their impact on the success of funding. Utilizing sentiment analysis, an unstructured P2P loan dataset was analyzed to extract emotions. The research suggests that borrowers can increase their likelihood of obtaining funding by improving the textual quality of their loan descriptions, focusing on readability and linguistic correctness. Additionally, the study found that incorporating specific emotions into the loan descriptions can make them more attractive to potential lenders [32].

Based on a dual-processing persuasion theory, the Elaboration Likelihood Model, and a comprehensive dataset from Renrendai, a Chinese P2P lending platform, this study presents four compelling arguments related to the Completeness, Sentiment, and Intensity of Language. By analyzing the number of certificates containing central and peripheral cues in voluntary information, the research reveals two distinct approaches to persuade borrowers in P2P lending. Both central and peripheral cues significantly influence lenders' decision-making. Interestingly, while previous findings in fund-raising appeals indicated that negative emotions could evoke "empathy-helping," this study shows that negative sentiment is specifically associated with reduced funding success. Moreover, it identifies a negative interaction effect between the number of certificates and Completeness on funding success [33].

This paper presents an overview of various online peer-to-peer lending studies and subsequently prioritizes sentiment analysis due to the significant impact of people's words on decision-making processes. By employing sentiment analysis and topic modeling techniques, the paper aims to gauge the polarity of individuals' opinions and identify the prevalent vocabulary and subjects that concern them the most.

3 METHODOLOGY

Chapter 3 will cover the methodology and procedures employed in this project. The data for sentiment analysis and topic modeling, aimed at obtaining insights into the online peer-to-peer lending platform, was extracted and analyzed using RapidMiner. The process stages, as shown in Figure 1, encompass data collection, data pre-processing, topic modeling and sentiment analysis, along with opinion summarization.

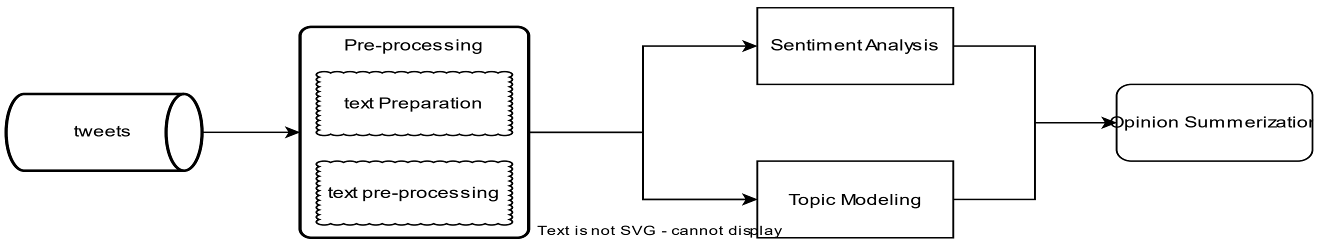


Figure 1: Opinion Mining

Developed by the same company, RapidMiner is an intuitive data analysis platform that provides a comprehensive integrated environment catering to machine learning, data preparation, text analysis, model deployment, business analytics, and predictive analytics. With a wide array of over a thousand drag-and-drop operators at its disposal, this tool enables swift and effortless execution of statistical mining operations. Its applications span across various domains, including business, commerce, education, research, and training, while also serving as a valuable tool for rapid prototyping, application development, and model deployment.

3.1 Data collection

To begin the project, data collection was the first phase. RapidMiner was utilized for this purpose, enabling the extraction of tweets related to peer-to-peer online lending and specific keywords from Twitter. The Twitter search operator facilitated the query-based retrieval of tweets containing the chosen keywords. To enhance the context of the tweets, additional operator settings were applied in expert mode. The search was also restricted to English language tweets. It is essential to note that the Twitter API imposes strict rate limits on data scraping.

To ensure comprehensive data coverage, the entire process was repeated with various keywords, including "Online lending," "P2P lending," "prosper lending," "BlockFi," "Upstart," "SOLO Funds," "Funding Circle," and "Kiva." This approach was necessary as RapidMiner only offers five hundred raw tweets per word. The objective was to collect and analyze every relevant tweet about these platforms. For a visual representation of the process, refer to Figure 2.

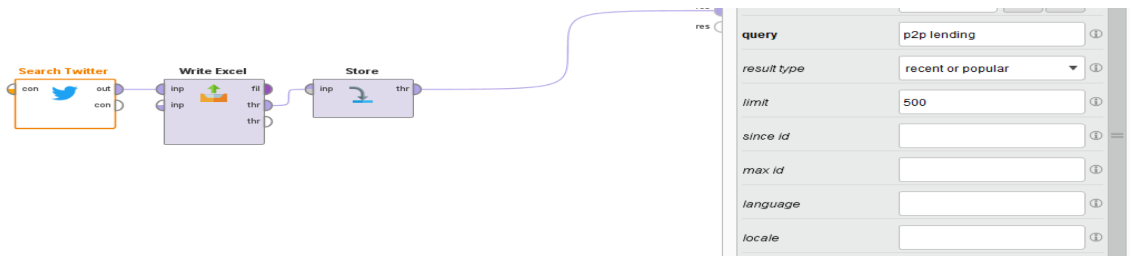


Figure 2: Search and store process

Figure 2 shows how we added 'write excel' and 'store' to save search results in an excel sheet and in the RapidMiner repository.

3.2 Data Exploration

Since RapidMiner was utilized for data extraction from Twitter, we had the freedom to select the preferred data storage format, which, in this case, was Excel. The extracted data possessed the following attributes: "Created-At," representing the timestamp; "Form-User," indicating the user's name; "From-User-ID," denoting the User account ID; "To-User," capturing the name of any person tagged in the tweet, if applicable; "To-User-Id," holding the ID of the tagged account; "Language," specifying the language of the tweet; "Source," providing the hyperlink to the tweet; "Geo-Location-Latitude" and "Geo-Location-Attitude," referring to the geographic coordinates from where the user composed the tweet; and finally, "Id," which corresponds to the tweet's unique identifier.

3.3 Data Pre-Processing

In this paper, the tool employed for data preprocessing on raw data gathered from social media platforms was RapidMiner. The raw data from Twitter will undergo a two-stage processing approach: text preparation and text pre-processing. Within the cleaning process, several steps are involved, starting with the selection of an attribute.

Figure 3: First step in the cleaning process

Figure 3 depicts the use of Select Attributes to remove unnecessary columns that do not contribute to the outcome of this paper. Only a few columns are required, such as where it was created, the user, and the text content.

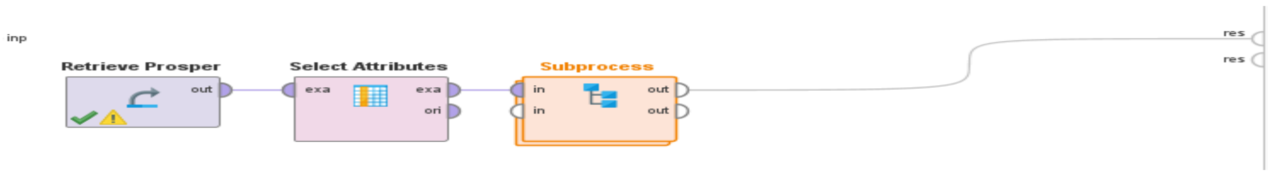


Figure 4: Second step in the cleaning process

Figure 4 depicts a subprocess that includes some steps for removing superfluous words and characters.

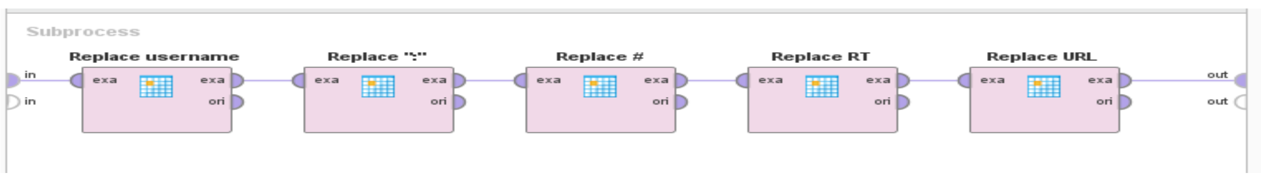


Figure 5: Text pre-processing process

Figure 5 depicts a subprocess in which a few elements are deleted during cleaning.

Figure 6: The text preparation process

The next crucial task involves eliminating white spaces and eliminating duplicates. These essential processes are necessary for effective sentiment analysis, ensuring accuracy and eliminating redundant data during data processing. Given that Twitter data may sometimes include missing values, the Replace Missing Values operator can be employed to substitute these gaps with appropriate characters.

After completing the data cleaning process, it becomes crucial to safeguard and process the data by including the 'write excel' and 'store' functionalities. This essential step allows for a thorough review, enabling the identification and resolution of any potential errors in the displayed results. By meticulously revisiting each step, one can guarantee that any issues encountered initially are prevented from reoccurring, thereby ensuring optimal and error-free outcomes.

3.4 Sentiment Analysis

After all data has been thoroughly cleaned and checked for errors that may have occurred during the cleaning process, sentiment analysis will begin. Sentiment analysis is the process of categorizing and categorizing tweets as positive or negative. The method is depicted in the figure below:

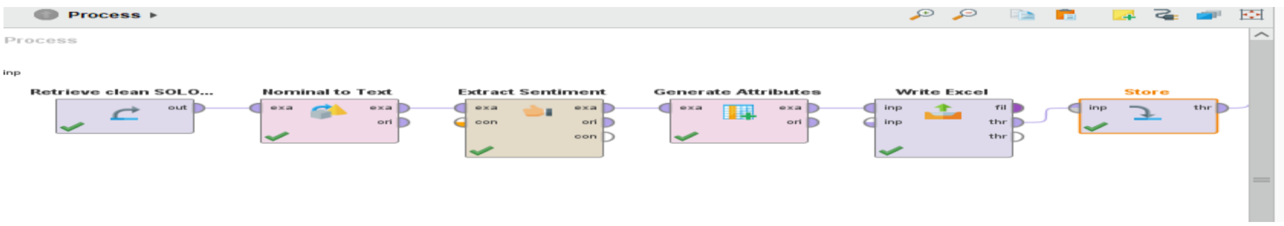


Figure 7: Sentiment analysis process

In Figure 7, the process involves retrieving the clean data once again and then conducting sentiment extraction. For this paper, the sentiment analysis utilizes the Valence Aware Dictionary for Sentiment Reasoning (VADER) model, which is a text sentiment analysis tool capable of detecting both polarity (positive/negative) and emotional intensity. VADER's sentimental analysis relies on a dictionary that maps lexical features to emotion intensities, providing sentiment scores. These scores are obtained by summing up the intensity of each word in the text. Additionally, VADER is well-suited for analyzing social media data, especially Twitter, and is known to yield dependable results.

The Generate attribute is used to generate sentiment (positive, neutral, or negative) based on the result score. Set function expressions for the generated attribute as shown below:

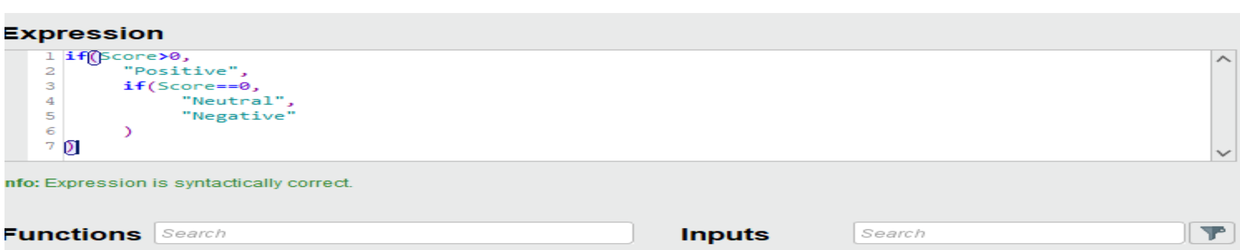


Figure 8: Expression to Form Sentiment Analysis

As shown in Figure 8 above, the expression that should be used is that if the score is greater than zero, the sentiment is positive; if the score is equal to zero, the sentiment is neutral; and anything else is considered negative. As a result, this formula will assist RapidMiner in determining and classifying sentiment based on the scoring number generated by the process and this step. After using the 'write excel' and 'store' operators in this process, the result will be saved in Excel.

After the sentiment analysis process, we need to check the accuracy of the sentiment analysis model through the cross-validation process. The process and the sub-process are shown in the following figures:

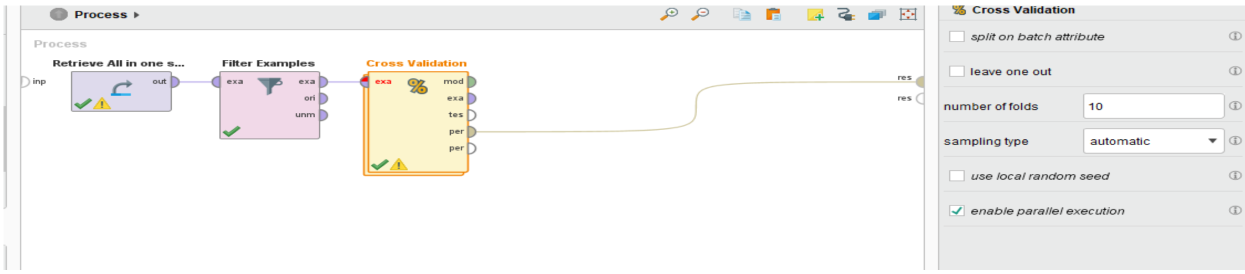


Figure 9: The cross-validation process

Cross-validation is a popular machine learning technique for assessing a model's performance and estimating its accuracy on unknown data. The available data is divided into two sets: a training set and a testing set. To gauge the model's performance on brand-new, untested data, it is trained on the training set and then assessed on the testing set. As shown in Figure 9 above, we collected all data sheets in one big sheet after done sentiment analysis and add a filter example operator to choose the column that we mainly depend on and add the cross-validation operator.

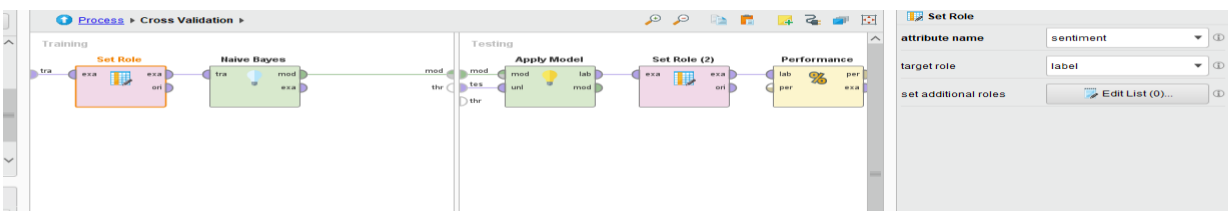


Figure 10: The sub-process of cross-validation

As depicted in Figure 10, the cross-validation process subprocess began with setting up the role to choose the column sentiment that we would use in our analysis and then add the model that would best fit our dataset. For the training section, we selected Naive Bayes because it would work well with the data we obtained from Twitter. Finally, for the testing section, we added the applied model and performance to check the applied model's confusion matrix.

3.5 Generate Word List

The final method for this process is called generate word list, and it is used to tokenize the sentences that have been processed word by word. Tokenization means that any sentences, statements, or a few words given to the tools will be split into words, and each word will have its own scoring based on how many times the word is used its popularity, and the sentiment of the word itself. This process is important to know how the exact perspective of the customer towards every word that they have tweeted and how-to sentiment it. The following figures will show the process of generating the list.

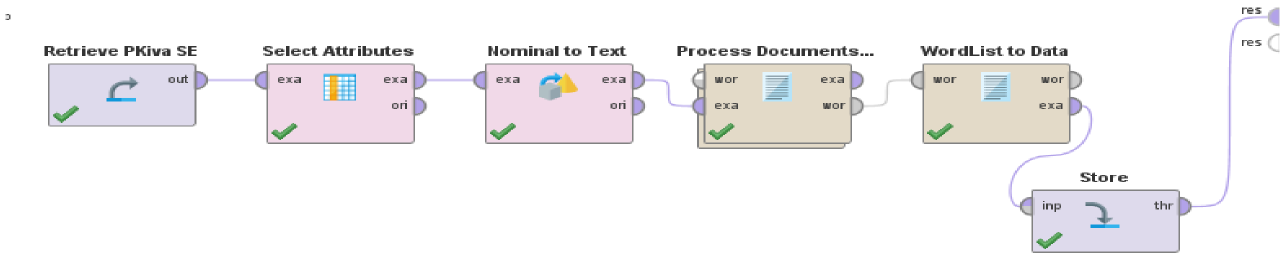


Figure 11: Part 1 of the word cloud process

In Figure 11, the 'Nominal to Text' operator is utilized to transform attributes into string format. Each nominal value is directly converted into a string value for the corresponding new attribute. When a value is absent in the nominal attribute, the resulting value in the new attribute will also be missing.

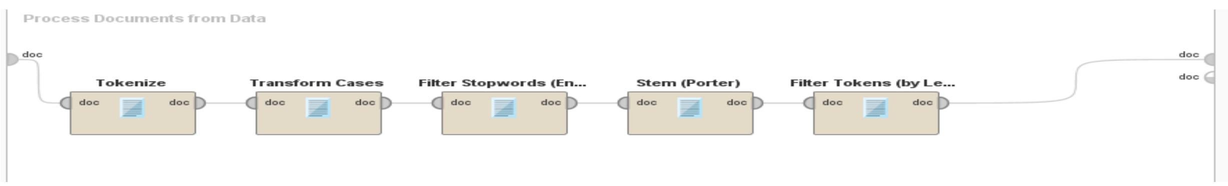


Figure 12: Process Documents from data

Figure 12 shows a subprocess that basically from an operator named 'process document from data', this operator, has a few subprocesses which are tokenized; to split the text into a single as mentioned earlier in this page, transform cases; to standardize all words to lower case, filter Stopwords (English); to remove any English stop word such as 'the', 'and', etc...., stem (portal); reduce the length of the word, filer tokens (by length); will remove any too short or too long words.

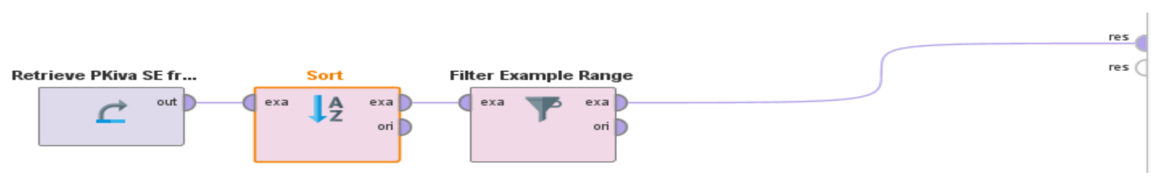


Figure 13: Part 2 of the word cloud process

Following the extraction of the word from each tweet and saving it in an Excel file, the "Sort" operator is used to sort each class by descending because this project wants to find the greatest number of times the word appeared for each class, and the "Filter Example Range" operator is used to set the top twenty words.

3.6 Topic Modelling

Topic modelling is employed to extract information from documents by generating conceptual "topics" from a collection of words that better represent the details of the document set. This task consists of two parts: identifying key themes in the text data set and determining which documents cover these topics. To achieve this, we followed several steps. First, we obtained the cleaned data and applied the "select attributes," "Filter Example," "Nominal to Text," and "data to documents" operators. The latter operator generates documents based on values in the data set, and we utilized the "loop collection" operator to loop over its subprocess for each object in the input collection. The output of this operator is also a collection, and additional results from the subprocess can be presented as collections through its output ports. Inside the operator, we added various tasks like "Tokenize," "Transform Cases," "Filter Stop words (English)," "Stem (Porter)," and "Filter Tokens (by Length)," as depicted in Figure 3.3. Finally, we applied the "Extract Topics from Documents (LDA)" operator, which utilizes the LDA technique to assume that words in each document are linked. To determine the number of topics to generate, we set it to ten, with the top five words per topic.

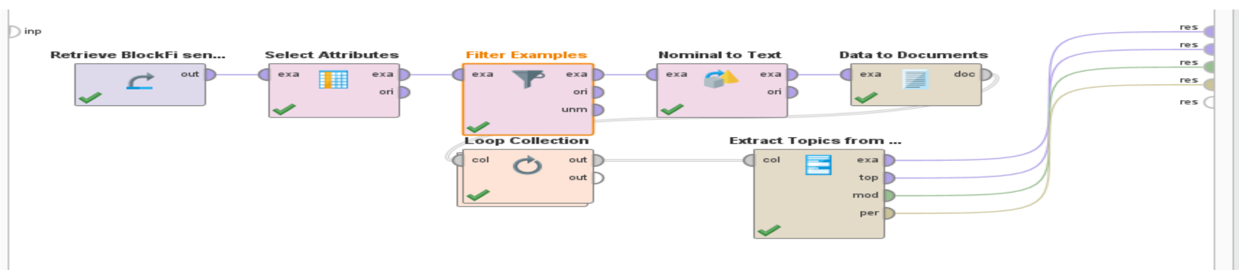


Figure 14: The process of topic modelling using the LDA method.

3.7 Opinion Summarization

Finally, the modelling results are presented in table, graph, and word cloud formats to aid comprehension and answer the two research questions. Furthermore, the RapidMiner results are shown in the Appendices.

4 RESULTS

The majority of raw data used in this paper is five hundred rows, and after data cleaning for each platform, p2p lending, and online lending, the P2P lending number is 181, the online lending number is 127, Prosper is 327, BlockFi is 249, Upstart is 371, Solo Funds is 64, Funding Circle is 108, and Kiva is 407. After completing the sentiment analysis, or text mining, three types of results were generated for each platform: sentiment overview, word overview, and topic modelling. In terms of a general overview of customer attitudes toward each platform expressed in tweets, the results are shown in the figure below:

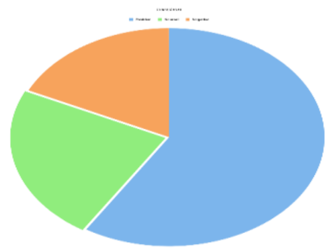


Figure 15: Kiva

The figure shows that the Scored positive with 240 (58.9%) from 407 data that completed the sentiment process after the cleaning process, neutral with 94 (23%), and negative with 73 (17.9%) from total data.

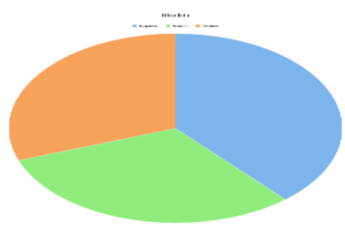


Figure 16: BlockFi

Here The highest sentiment toward BlockFi is Negative with 96 (38.5%), followed by Neutral with 77 (31%) and positive with 76 (30.5%) from 249 data points that complete the sentiment process. From 127 data that complete the sentiment analysis for Online Lending, the highest sentiment towards Online Lending is Positive with 80 (62%), while Neutral sentiment with scoring 33 (25.9%) and 14 (11%) Negative. For P2P Lending, from the total of 181 data after the cleaning process, the highest sentiment towards this kind of website is Positive with 93 (51.3%), Neutral as much as 70 (38.6%) and Negative is 18 (10%) from the total data. for prosper, from a total of 327 data after the cleaning process, the highest sentiment toward this platform is Positive with 157 (48%), Negative with 91 (27.8%) and Neutral is 79 (24.1%). from 64 data that completed sentiment analysis after the cleaning process for SOLO Funds, 32 (50%) had Positive sentiment, 18 (28.1%) is Neutral and Negative with 14 (21.8%). From 371 data that result from the cleaning process for Upstart the highest sentiment is positive 184(49.5%), neutral as much as 94 (25.3%), and negative 93 (25.1%). For Kiva, from a total of 407 data after the cleaning process, the highest sentiment toward this platform is Positive with 240 (58.9%), while Neutral is 94 (23.1%), and Negative is 73 (17.9%) from the total data. From 108 data that complete the sentiment analysis for Funding Circle 58 (53.7%) data is Positive sentiment, while 26 (24.1%) is Negative and the rest of the data is Neutral with 24 (22.2%). For BlockFi, from the total of 249 data after cleaning, the highest sentiment is Negative with 96 (38.5%), Neutral with 77 (30.9%), and lastly, Positive sentiment is 76 (30.5%).

After showing the results of the sentiment analysis process by numbers and percentage for each keyword we will show the result of cross-validation as we collect all those sheets in one sheet and here are the results for them.

Criterion: accuracy, kappa

Table View / Plot View

accuracy: 91.99% +/- 2.47% (micro average: 91.98%)

	true Positive	true Negative	true Neutral	class precision
pred. Positive	805	27	0	96.75%
pred. Negative	115	398	5	76.83%
pred. Neutral	0	0	484	100.00%
class recall	87.50%	93.65%	98.98%	

Figure 17: Cross-validation results

The figure shows the number of true positive, true negative, false positive, and false negative for each class positive, negative, and neutral. The model gets 91.99% as the total correct prediction.

Accuracy for the model = (TP + TN) / (TP + TN + FP + FN) (805+398+484)/(805+115+27+398+5+484) = 1687/1834= 91.9%

Precision = TP / (TP + FP) for positive = 805/920 = 87.5%, For negative = 398/425 = 93.65%, and for neutral = 484/489 = 98.9%

Recall = TP / (TP + FN) For positive class = 805/ (805+27) = 96.75%, For negative class = 398/(398+115+5) = 76.83%, and for neutral class = 484/484 = 100%.

F1-score = 2 * (Precision * Recall) / (Precision + Recall)

For positive class = 2 * (87.5% * 96.75%) / (87.5% + 96.75%) = 91.89%.

For negative class = 2*(93.65% * 76.83%) / (93.65% + 76.83%) = 84.41%.

For neutral class = 2*(98.9% * 100%) / (98.9% + 100%) = 99.44%.

Due to the fact that the cleaning sheets used in the sentiment analysis process do not all contain the same number of rows of data, we can compare sentiment results by percentage. According to the results that are displayed above, we used eight different words, seven of which have positive sentiment analysis, with Online Lending having the highest percentage (62%) and Kiva having the highest number of rows and positive tweets (240), respectively. The final word, BlockFi, is the only one with a negative sentiment analysis of 38.5% and a positive sentiment analysis of 30.5% for third place. The word cloud results for positive and negative sentiment analysis show the most frequent word combinations that occur together; as a result, the scoring string column used in this process for each keyword is shown in the following figures, and the remaining ones are shown in the appendix.

Figure 18: BlockFi Negative word cloud

Here are the most negative sentiment words, which can be seen clearly in a few words such as 'contagion,' 'unaware,' 'uncertainties,' and 'falling,' where the scoring between one word and another is nearly identical.

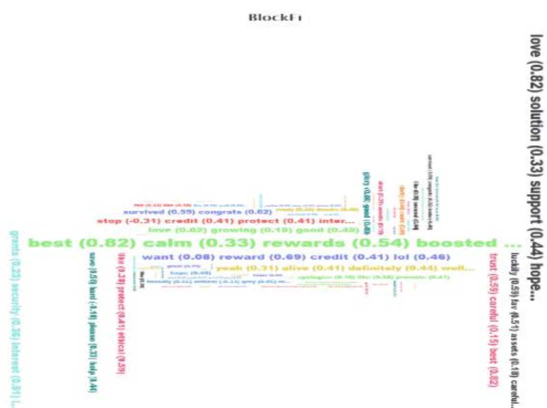


Figure 19: BlockFi positive word cloud

Here are the most positive sentiment words, which can clearly be seen a few Positive words as 'love', 'solution', 'support', and 'hope'. but the frequency is quite low when compared to the negative sentiment.

Figure 20: Kiva Positive Word Cloud

The most positive sentiment words which can be seen clearly are 'easy', 'join', 'helping,' and 'dream' the scoring between one word to another word is almost the same.

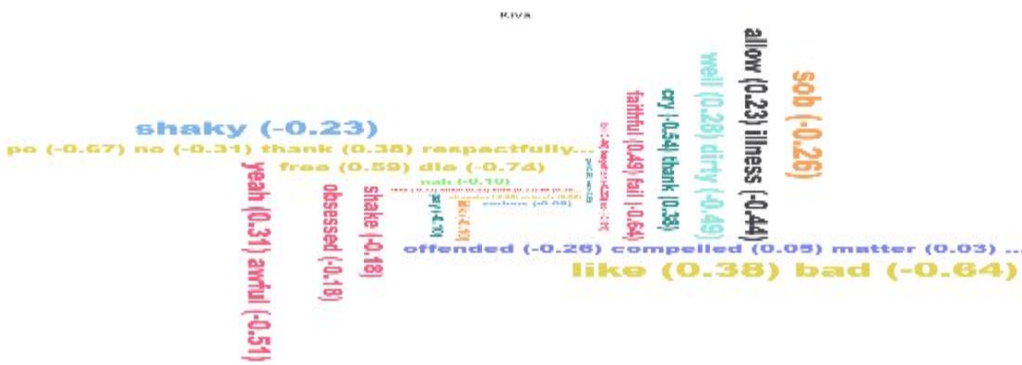


Figure 21: Kiva Negative Word Cloud

Here are the most negative sentiment words which can see clearly which are 'bad', 'awful', illness', 'fail', and 'sob'.

Table 1: BlockFi Word Generate

Positive			Negative		
Word	Document	Total	Word	Document	Total
BlockFi	34	35	BlockFi	66	70
Crypto	15	18	Withdraw	35	38
Withdraw	12	18	Crypto	26	27
Credit	8	9	Collapse	16	17
Bitcoin	8	8	Halt	16	16
Voyage	8	8	Lender	12	12
Card	7	7	Stop	8	10
Hope	5	6	Contagion	9	9
Lender	5	6	Platform	8	8
People	6	6	Fail	4	6

Table 1 displays the generated word from the list of Tweets gathered for sentiment analysis. In this table, the BlockFi word has the most occurrences among positive words and negative words, but the difference is in the total number of occurrences for each sentiment. Crypto has the second highest occurrence for positive words generated by eighteen times in fifteen documents, while withdraw has thirty-eight occurrences in a total of thirty-five documents for the BlockFi platform. There are a few services that are offered on the platform mentioned in positive sentiments, such as credit, bitcoin, and voyage, while other words may be because of customer service, which affects the users' perspective of the platform.

Table 2: Kiva Word Generate

Positive			Negative		
Word	Document	Total	Word	Document	Total
Kiva	105	109	Kiva	24	24
Help	48	50	People	11	12
Make	45	47	Work	5	7
Join	37	37	Student	3	6
Easi	36	36	Time	5	6
Differ	33	34	Debt	3	5
Life	34	34	Start	3	5
Loan	18	29	Exacting	4	4
Pursue	29	29	Live	4	4
Dream	28	28	Plot	2	4

Table 2 shows that the words generated for the Kiva sentiment itself, which is Kiva platform, occur 109 times in 105 documents followed up by help, join; where one of the websites that provide information about online lending platforms stated that the Kiva platform popular among first-time borrowers as it has hybrid lender with a mix of lending and crowdfunding and they offer zero percentage loans to small businesses and a lot of users mention it on tweets, which occurs 50 times in total 48 documents and 37 times in total 37 document respectively. Then there are some other words like easy, differ, life and dream.

In contrast, for the negative word generated for Kiva, the platform itself is mentioned twenty-four times in a total of twenty-four documents. Some users also used words like work, student, debt and live in their tweets. The remaining will be added in the appendix.

Table 3: Part of LAD Analysis

Topic / Document	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
0	0.023	0.052	0.040	0.033	**0.293**	0.037	0.058	0.022	0.034	**0.408**

1	0.035	0.064	0.049	0.061	**0.529**	0.051	0.053	0.042	0.059	0.056
2	0.040	0.061	0.061	0.144	**0.272**	**0.280**	0.037	0.020	0.047	0.040
3	0.033	0.066	0.045	0.047	**0.580**	0.044	0.052	0.023	0.038	0.070
4	0.056	**0.117**	0.080	0.066	**0.333**	0.044	0.055	0.092	0.045	**0.112**

Calculating the extracted topic from the document makes use of the LDA analysis. For example, according to Table 4.3, Documents 0, 1, 2, 3, and 4 have high confidence values for Topic 4 of 0.29, 0.52, 0.27, 0.58, and 0.33, respectively. The topic is manually defined by referring to the document and keyword with the most weights, so Documents 0, 1, 2, 3, and 4 are related to how people express their dissatisfaction with how the BlockFi platform works. Furthermore, there are multiple Documents with high confidence in two Topics, such as Documents 0 and 4, which have confidence values for Topic 9 of 0.408 and 0.112, respectively with high weighted words 'deposit', and 'people'.

Table 4: part Topic discovered using LDA.

Topic	Major keywords and representative data
BlockFi	‘BlockFi,’ ‘Withdraw,’ ‘Crypto’,’ people,’ and ‘deposit’
	1- ‘Constant and never-ending warnings ignored. **BlockFi** Celsius ftx’ 2- ‘**BlockFi** Restricts Platform Activities Also Suspended **Withdrawals**’ 3- ‘**BlockFi** blocking **withdrawals**. Meanwhile, you can **deposit** liquidity in a single-sided fashion within our web app for more…’ 4- ‘Looks like **BlockFi** is completely dead. Internal Slack deactivated; **people** tell me they stored almost all customer **deposits** on…’
Bad	‘Cryptocurr,’ ‘bad,’ ‘stop,’ and ‘troubl’
	1- ‘**Cryptocurrency** lender BlockFi said it was pausing withdrawals and limiting activity on its platform, becoming the latest casualty of the sudden collapse of Sam Bankman-Fried’s crypto empire’ 2- ‘Market down so **bad** it isn’t even reacting to BlockFi news...’ 3- ‘Maybe it just needs to revisit 16k, let us see how **bad** it is with BlockFi?? ‘ 4- ‘BlockFi next in line **stopped** withdrawals’

Bankrupt	‘Bankrupt,’ ‘loan,’ ‘shut,’ ‘collaps,’ and ‘case’

1- ‘Uncertainty related to the **collapse** of FTX, corresponding defaults, and lending platform BlockFi halting of withdrawals as they face **bankruptcy** is leading to a vicious withdrawal cycle on other exchanges so let us do Crypto_systemic_stress_test by withdrawing our funds!! ‘
2- ‘Bitcoin is falling, FTX is **bankrupt**, BlockFi halting all activity, and we just received the news Twitter can go **bankrupt**...all in one day. The side effects of Eclipse are in progress. Bring it on baby!’
3- ‘Was at dinner last night when **shut** off their card. Got rugged trying to pay because they just **shut** it off after losing everyone’s money. Then send this email reminding us to pay them back?? God damn fkn clown show!’

The topic discovered by using the LDA method is shown in Table 4.5, and the findings of this project show that the main topic discussed by people on Twitter related to peer-to-peer lending online, specifically the BlockFi platform, is linked to BlockFi, Bad, and Bankrupt.

Table 5: part of LAD Analysis

Topic / Document	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
0	0.001	0.000	0.000	0.000	0.000	0.000	0.000	**0.996**	0.001	0.000
1	0.001	0.000	0.000	0.000	0.000	0.000	0.000	**0.996**	0.001	0.000
2	0.001	0.000	0.000	0.000	0.000	0.000	0.001	**0.995**	0.001	0.001
3	0.001	0.000	0.000	0.000	0.000	0.000	0.000	**0.996**	0.001	0.000
4	0.001	0.000	0.001	0.000	0.000	0.000	0.000	0.003	**0.995**	0.000

The LDA analysis is used to compute the extracted topic from the document. For example, according to Table 5, Documents 0, 1, 2, and 3 have a high confidence value for Topic 7 of 0.996. The topic represented is manually defined by referring to the document and keyword with the highest number of weights, so Documents 0, 1, 2, and 3 are related to how Kiva helps people through loans. Furthermore, Document 4 has a high confidence value for Topic 8 of 0.995, indicating that Document 4 has been linked to kiva and good.

Table 6: Part Topic discovered using LDA.

Topic	Major keywords and representative data
Help	‘Help,’ ‘join,’ ‘differ,’ ‘life,’ ‘make’ and ‘good’ 1- ‘Kiva is an easy way to make a real **difference** in someone's **life**. Will you **join** me in **helping** Jiraporn in Sikhoraphum, Surin to pursue their dream?’ 2- ‘Kiva lenders are about to fund their largest ever loan (!) which will provide solar power for over 100,000 people in the DRC. Just 3 hours to go and only 1% of the loan amount is still needed!!’ 3- ‘I have been supporting women via microloans for a few years now. Small change **makes** a **difference**. Kiva is an easy way to make a real **difference** in someone's **life**. Will you **join** me in **helping** Clairia in Kigali to pursue their dream?
Loan	‘Loan,’ ‘lend,’ ‘help,’ and ‘kiva’ 1- ‘Kiva **loan** no 25 Frietou's **loan** is 40% funded. We were happy to **help** her buy supplies for her spice and vegetable business, and to adjust the stock to her customer’s needs.’ 2- ‘Banks won’t **lend** if you cannot pay it back, so what has happened here is the government covering much of the risks and letting banks **lend** unfettered.’ 3- ‘**Thanks** for sharing this. Never heard of it, but this is something I can do. Hard to determine if some of these orgs are legit. Want to **help**, but do not have money to throw away. Will do more reading about it.’ 4- ‘**Kiva** is really good’
Giving	‘Give,’ ‘moder,’ ‘experi,’ ‘social,’ and ‘socialist’ 1- ‘Got Velocitron Override and now I want to finish the Animated design. Though, I would make it a slight nod to MEGAS XLR, because for some reason Animated Override makes me think of Kiva ‘ 2- ‘We put some of our tithing money into Kiva instead of **giving** it to the Mormon church, and it just keeps **giving** and **giving**!!! Way better experience, Imo.’ 3- ‘All he has to do to **moderate** harder is to balance **moderation**, which to this point has been completely lopsided in favour of woke socialists. We do not care if there are standards, they just need to be applied equally, and not serve one (idiot) special interest group.’

Financial	'Loan,' 'financi,' 'peopl,' and 'borrow'	1- 'So, then the interest would balloon until the total amount owed is way more than the original amount **borrowed**? That seems like a **financial instrument** misused and improperly regulated.' 2- 'Members have made over one million **interest-free loans** in support of Kiva's Agriculture (540,143) and Food (462,476) sectors. They represent almost half of the Team's 2.12 million **loans**. Retail, Services, Clothing and Education are the next most popular categories.' 3- '**Borrowers** like Neriman are shining examples of how Kiva **loans** help **people** improve their lives and create **financial** independence....'

The topic discovered by using the LDA method is shown in Table 4.6, and the findings of this project show that the main topic discussed by people on twitter's tweets related to peer-to-peer lending online, specifically the Kiva platform, is linked to help, loan, give, and financial. The appendix shows the results of the topic modelling in RapidMiner using the LDA method.

5 DISCUSSION

The details of the paper's sentiment analysis findings will be discussed in this topic. As shown in the previous topic's tables, there is a comparison of both positive and negative sentiment, as well as a comparison of these major platforms, which are Upstart, BlockFi, Kiva, and Funding Circle.

The first objective is to know the opinion of people about peer-to-peer lending online.

Figure 22: BlockFi Sentiment

Figure 22 shows that Twitter users have a negative perspective through their Tweets; positive is the second highest, and neutral is the lowest among the sentiment analyses. The possibility for why the negative is increasing is that many Tweets are about making statements about the BlockFi platform by displaying expressions on it. As an example, 'Looks like BlockFi is completely dead. Internal Slack deactivated, people tell me they stored almost all customer deposits on FTX', and 'BlockFi Restricts Platform Activities Also Suspended Withdrawals' These tweets express their feelings about the platform, as they discuss how the BlockFi platform facing a lot of problems.

Because the occurrence of positive sentiment for BlockFi is low, only a few tweets contribute to it, such as some customers who were satisfied with the service did mention in their tweets like Yeah that was keeping alive and that was most definitely Bitcoin as well.'. They also state, 'It is currently working.' which makes no exception for the Lexicon Based Approach to classify as positive or negative.

The details on how to generate a list of words, as mentioned in the previous topic, will be discussed further in this topic, and the top ten words with meaning will be chosen as a result. The comparison in words is important in this paper because it ensures that the sentiment analysis results are focused on a much smaller area. Table1 The top ten positive and negative word occurrences in this generated word list are shown. The most common are BlockFi words, which express both positive and negative emotions. Despite the fact that both words are frequently used, the total number of times BlockFi words are mentioned in negative sentiments outnumbers the number of times they are mentioned in positive sentiments. The total occurrence occurred for negative sentiments because most users who speak about BlockFi will demonstrate to others how dissatisfied and frustrated they are with the platform's services and how the platform can stop withdrawal and other services.

According to the second objective, we discovered that most people speak about topic four, and the most commonly used words in this topic are BlockFi and withdrawal. Tables 3 and 4 in the result chapter show a portion of the process. The remainder of the process and results will be shown in an appendix.

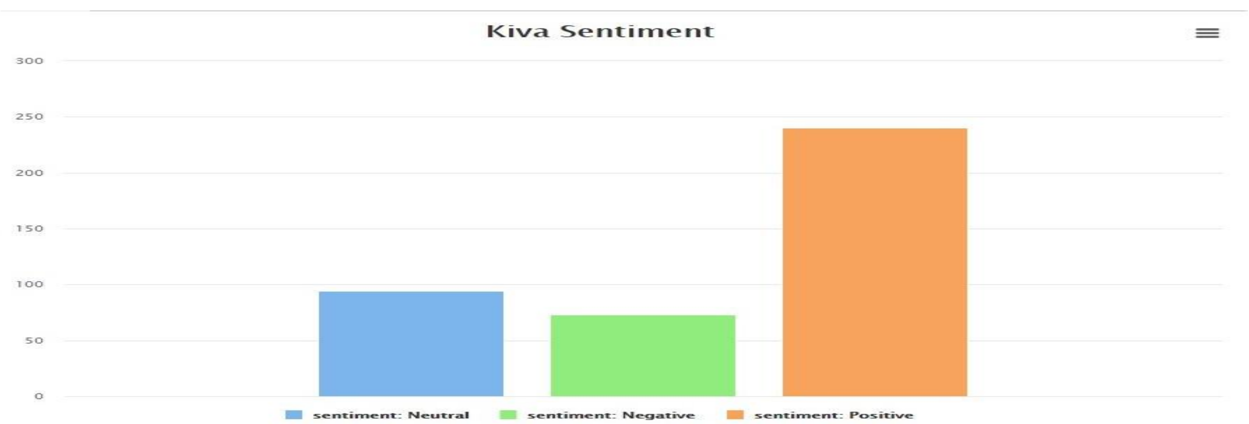


Figure 23: Kiva Sentiment

Figure 23 shows that Twitter users have a positive perspective through their Tweets, with negative coming in second and neutral coming in last among the sentiment analyses. The fact that many Tweets are about making statements about the Kiva platform by displaying expressions on it could explain why the positive is increasing. Kiva, for example, is a simple way to make a real difference in someone's life. Will you join me in assisting Senegal's 08-Yarang Group in pursuing their dream?' and 'Help Save the World Visit' These tweets express their feelings about the platform by discussing how the Kiva platform helps others make a difference in their lives.

Because the occurrence of negative sentiment for Kiva is low, only a few tweets contribute to it, such as some customers who were dissatisfied with the service did mention in their tweets Now it is at the expense of taxpayers because they are left paying for student debt they did not sign up for & an education they did not receive. It is nonsense.' They managed the flashback plot far better than Kiva did.' These have scores of -0.8 and -1.2, respectively. Those who also say, "And the teachers' The person to turn that frown upside down is yourself boo stay woke," implying that the Lexicon Based Approach makes no distinction between positive and negative classification.

As mentioned in the previous topic, the details on how to generate a list of words will be discussed further in this topic, and the top ten words with meaning will be chosen as a result. In this paper, the comparison in words is important because it ensures that the sentiment analysis results are focused on a much smaller area. Table 2 This generated word list displays the top ten positive and negative word occurrences. Kiva words, which express both positive and negative emotions, are the most common. Despite the fact that both words are frequently used, the total number of times Kiva words are mentioned in positive sentiments outnumbers negative sentiments. The total occurrence occurred for positive sentiments because most Kiva users who speak about it will recommend it to others, as they demonstrate how Kiva makes a difference and helps others improve their lives.

According to the second objective, we discovered that the majority of people talk about topic one, and the most frequently used words in this topic are Kiva, help, and simple to apply. A portion of the process is depicted in Tables 5 and 6 of the result chapter. The remaining steps and results will be included in an appendix.

5.1 Limitation

RapidMiner is based on Sentiment Analysis. The majority of the processes are related to the Lexicon-based Approach, particularly when using Vader. Because Vader is used to analyzing words rather than sentences, the limitation that can be found here is a limitation in processing sarcastic sentences, which means that the words used do not mean what they are. Some users use this method to attack, defend, or simply uniquely express themselves. As an example, consider the sentence Follow all Funding Circle social media accounts. We offer financial services with zero stress. SME Business loan. This sentence expresses how good this is and how you can invest and get a loan without worrying about anything, but as Vader cannot see the meaning behind it and only sees the word stress, it is interpreted negatively.

Aside from that, RapidMiner has a limitation regarding scoring words, where the analysis of the words that can be shown in a diagram or graph is limited to five hundred words only, which means that a large project or extraction of data cannot be contained or processed in a diagram by RapidMiner. In fact, the data that has been gathered must be analyzed in other tools or manually to ensure that all of the data provided can be processed in a timely manner. This limitation may limit the process of analysis to five hundred data points only, necessitating the use of other tools if the number exceeds five hundred. We cannot simply extract data whenever we want. As far as data extraction from Twitter is concerned, we extract data from dates prior to two weeks.

Furthermore, RapidMiner has a limitation in the languages that can be extracted, tokenized, or processed using Vader. Even though RapidMiner has stated on their website that users can request that RapidMiner add up some other language by providing them with a dictionary and so on, it still takes a long time for certain languages to be extracted by the RapidMiner. For example, RapidMiner can extract Mandarin, but most users have difficulty doing so and require the assistance of a RapidMiner developer to ensure that this process occurs. This type of constraint will slow down the extraction process.

RapidMiner is also based on a lexicon-based approach, which means that it cannot directly teach the tools to understand specific words. Some languages have words that can have complicated meanings just by one word, which makes RapidMiner difficult to process when compared to Machine Learning Approach, where the words can be trained to be positive, negative, or neutral. In Japanese, for example, the term "mononaware" refers to an awareness of the impermanence of all things and a gentle sadness at their passing. This one word has a complicated meaning, making it difficult for the Lexicon-based Approach to process and provide a sentiment toward it.

6 CONCLUSION

Sentiment analysis is a technique that is simple and convenient to apply, saves money and time, and aids in the study of large amounts of data or samples of data gathered from online sources like Twitter or social media networks. RapidMiner, a possible tool for sentiment analysis, was used to analyze user and customer sentiment. It is useful for predicting sentiment and for focusing on improving user and customer satisfaction with products and services.

Improvement work is still being done to increase the accuracy of the data analysis. Since Twitter makes it difficult to determine the country from a remark or tweet, Additionally, a very specialized keyword search returns very little information, and occasionally none. The textual tweet data we obtained from Twitter constitutes the sole subject matter of this paper. In the future, Twitter will be able to incorporate multimedia data like images, music, and videos, making the output more appealing and user-friendly. Finally, a future suggestion would be to leverage the expression meaning of the sentences or explanations in addition to the hashtags and keywords to collect the data.

This study's influence was meant to assist future analysts in developing a practical model that can quickly and accurately predict user ratings through sentiment analysis. Having a better grasp of one's own and one's competitors' positions can also be beneficial to business. Based on the unfavorable criticism, additional work on the real model can be built and improved to offer better services and serve the users better.

The findings of this study have important implications for businesses operating in the P2P lending industry. By understanding user opinions and perspectives, businesses can better position themselves in the market and improve their services. The use of sentiment analysis can also help businesses to test the response of customers to new announcements and to learn more about how the public feels about their services.

Furthermore, the use of RapidMiner as an analytical tool to build sentiment analysis models was found to be simple and effective compared to traditional machine learning techniques. The use of the confusion matrix to evaluate the performance of the sentiment analysis model can provide valuable insights for businesses looking to improve their online reputation.

Based on the findings and conclusions of the study, there are some recommendations to be considered such as: (1) The platforms should consider the customers' opinion and their viewpoints so, the people will be satisfied with their service, (2) The platforms should justify the downtime for any services to make customers on the same page, and (3) Customers should not blame platforms for any inconvenient situation. They should wait for justification from their side.

In future research, it is recommended to incorporate more advanced techniques such as natural language processing and deep learning to further improve the accuracy of the sentiment analysis model. Additionally, further research could explore the use of other social media platforms and multimedia data to gain a more comprehensive understanding of public opinion.

Overall, this study provides a valuable contribution to the field of sentiment analysis and topic modeling for P2P lending online platforms. The findings offer insights for businesses to improve their services and for researchers to further explore the potential of sentiment analysis in understanding public opinion.

ACKNOWLEDGEMENT

This work is supported by Ministry of Higher Education Malaysia, Fundamental Research Grant Scheme Grant No. FRGS/1/2022/STG06/USM/02/4, for the project entitled "Efficient Joint Process Monitoring Using a New Robust Variable Sample Size and Sampling Interval Run Sum Scheme".

REFERENCES

[1] H. S. Disemadi, M. A. Yusro, and W. G. Balqis, "The problems of Consumer Protection in Fintech peer to peer lending business activities in Indonesia," *Sociological Jurisprudence Journal*, vol. 3, no. 2, pp. 91–97, 2020.

[2] S. Boudko, H. Abie, M. Boscolo, and D. Ferrario, "Predictive analytics service for security of blockchain and peer-to-peer payment solutions," *Lecture Notes in Electrical Engineering*, pp. 71–81, 2021.

[3] S. Reddy and K. Gopalaraman, "Peer to peer lending, default prediction-evidence from lending club," *Internet Banking and Commerce,* 2016. [online] Available: https://www.icommercecentral.com/open-access/peer-to-peer-lending-default-predictionevidence-from-lending-club.php?aid=81766&view=mobile

[4] L. Huang and H. N. Pontell, "Crime and crisis in China's P2P online lending market: A comparative analysis of fraud," *Crime, Law and Social Change*, 2022.

[5] E. Yeo and J. Jun, "Peer-to-peer lending and bank risks: A closer look," *Sustainability*, vol. 12, no. 15, p. 6107, 2020.

[6] K. Taujanskaitė and E. Milčius, "Accelerated growth of peer-to-peer lending and its impact on the consumer credit market: Evidence from Lithuania," *Economies*, vol. 10, no. 9, p. 210, 2022.

[7] A. A. Taleizadeh, A. Z. Safaei, A. Bhattacharya, and A. Amjadian, "Online peer-to-peer lending platform and supply chain finance decisions and strategies," *Annals of Operations Research*,

vol. 315, no. 1, pp. 397–427, 2022.

[8] G.-X. Gao, Z.-P. Fan, X. Fang, and Y. F. Lim, "Optimal Stackelberg Strategies for financing a supply chain through online peer-to-peer lending," *European Journal of Operational Research*, vol. 267, no. 2, pp. 585–597, 2018.

[9] J. Wang and R. Li, "Asymmetric information in peer-to-peer lending: Empirical evidence from China," *Finance Research Letters*, vol. 51, p. 103452, 2023.

[10] C. Jiang, Z. Wang, R. Wang, and Y. Ding, "Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending," *Annals of Operations Research*, vol. 266, no. 1-2, pp. 511–529, 2017.

[11] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, 2015.

[12] S. (F. Pengnate and F. J. Riggins, "The role of emotion in P2P microfinance funding: A sentiment analysis approach," *International Journal of Information Management*, vol. 54, p. 102138, 2020.

[13] V.-S. Ha, D.-N. Lu, G. S. Choi, H.-N. Nguyen, and B. Yoon, "Improving credit risk prediction in online peer-to-peer (P2P) lending using feature selection with Deep Learning," *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 2019.

[14] J. Duan, "Financial system modeling using Deep Neural Networks (dnns) for effective risk assessment and prediction," *Journal of the Franklin Institute*, vol. 356, no. 8, pp. 4716–4731, 2019.

[15] G. Babaei and S. Bamdad, "A multi-objective instance-based decision support system for investment recommendation in peer-to-peer lending," *Expert Systems with Applications*, vol. 150, p. 113278, 2020.

[16] H. Guo, K. Peng, X. Xu, S. Tao, and Z. Wu, "The prediction analysis of peer-to-peer lending platforms default risk based on comparative models," *Scientific Programming*, vol. 2020, pp. 1–10, 2020.

[17] Y. Zhang, H. Li, M. Hai, J. Li, and A. Li, "Determinants of loan funded successful in online P2P lending," *Procedia Computer Science*, vol. 122, pp. 896–901, 2017.

[18] B. Gao and V. Balyan, "Construction of a financial default risk prediction model based on the LIGHTGBM algorithm," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 767–779, 2022.

[19] Y. Xia, "A novel reject inference model using outlier detection and gradient boosting technique in peer-to-peer lending," *IEEE Access*, vol. 7, pp. 92893–92907, 2019.

[20] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, "Study on a prediction of P2P network loan default based on the machine learning lightgbm and xgboost algorithms according to different high dimensional data cleaning," *Electronic Commerce Research and Applications*, vol. 31, pp. 24–39, 2018.

[21] Y. Aleksandrova, "Comparing performance of machine learning algorithms for default risk prediction in peer-to-peer lending," *TEM Journal*, pp. 133–143, 2021.

[22] P. Xia, Z. Ni, X. Zhu, and L. Ni, "A novel key influencing factors selection approach of P2P lending investment risk," *Mathematical Problems in Engineering*, vol. 2019, pp. 1–12, 2019.

[23] K. Ren and A. Malik, "Recommendation engine for lower interest borrowing on peer-to-peer lending (P2PL) platform," *IEEE/WIC/ACM International Conference on Web Intelligence*, 2019.

[24] S. Adarsh, V. S. Anoop, and S. Asharaf, "Distributed consensus mechanism with novelty classification using proof of immune algorithm," *International Conference on Innovative Computing and Communications*, pp. 173–183, 2022.

[25] X. Zeng, L. Liu, S. Leung, J. Du, X. Wang, and T. Li, "A decision support model for investment on P2P lending platform," *PLOS ONE*, vol. 12, no. 9, 2017.

[26] Kim and Cho, "Towards repayment prediction in peer-to-peer social lending using Deep Learning," *Mathematics*, vol. 7, no. 11, p. 1041, 2019.

[27] H. Liu, H. Qiao, S. Wang, and Y. Li, "Platform competition in peer-to-peer lending considering risk control ability," *European Journal of Operational Research*, vol. 274, no. 1, pp. 280–290, 2019.

[28] L. E. Boiko Ferreira, J. P. Barddal, H. M. Gomes, and F. Enembreck, "Improving credit risk prediction in online peer-to-peer (P2P) lending using imbalanced learning techniques," *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017.

[29] P.-C. Ko, P.-C. Lin, H.-T. Do, and Y.-F. Huang, "P2P lending default prediction based on AI and statistical models," *Entropy*, vol. 24, no. 6, p. 801, 2022.

[30] T. Alloway, "Big Banks Muscle in on Peer-to-Peer Lending." Financial Times, October 28, 2013, accessed March 23, 2015.

[31] E. Mussalimun, H. Khasby, G. I. Dzikrillah, and Muljono, "Comparison of k- Nearest neighbor (K -NN) and naïve Bayes algorithm for sentiment analysis on Google Play Store Textual Reviews," *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, 2021.

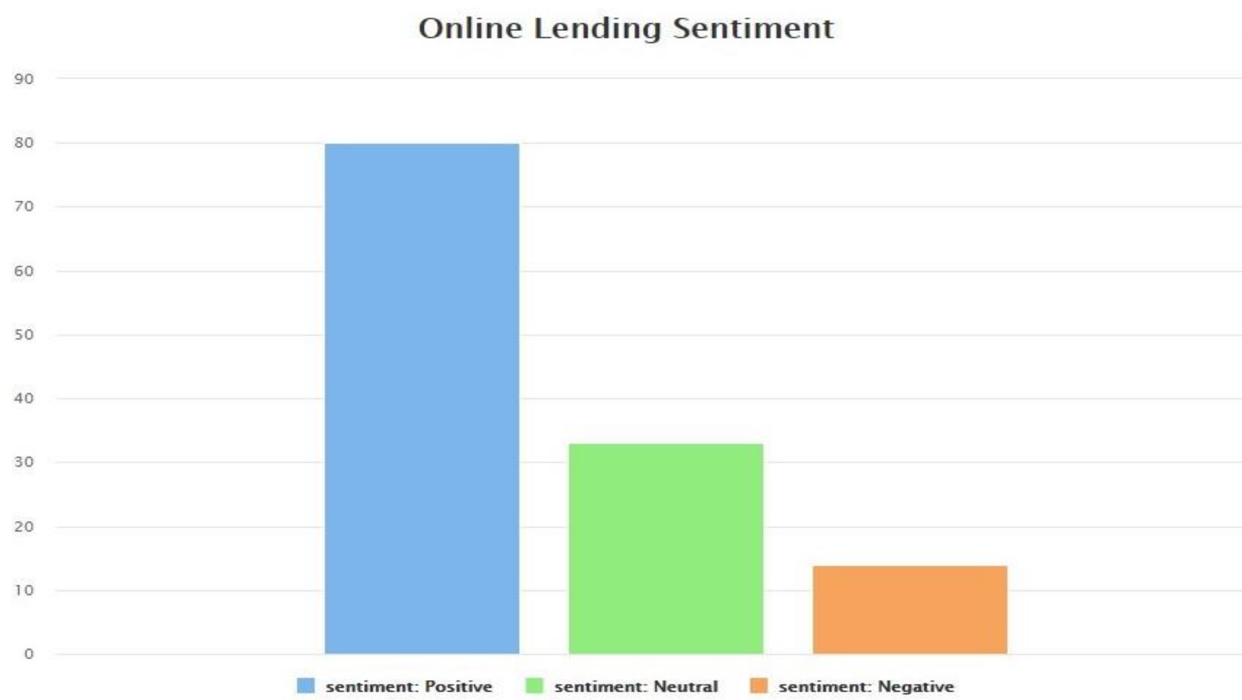
[32] D.-H. Shih, T.-W. Wu, P.-Y. Shih, N.-A. Lu, and M.-H. Shih, "A framework of global credit-scoring modeling using outlier detection and machine learning in a P2P lending platform," *Mathematics*, vol. 10, no. 13, p. 2282, 2022.

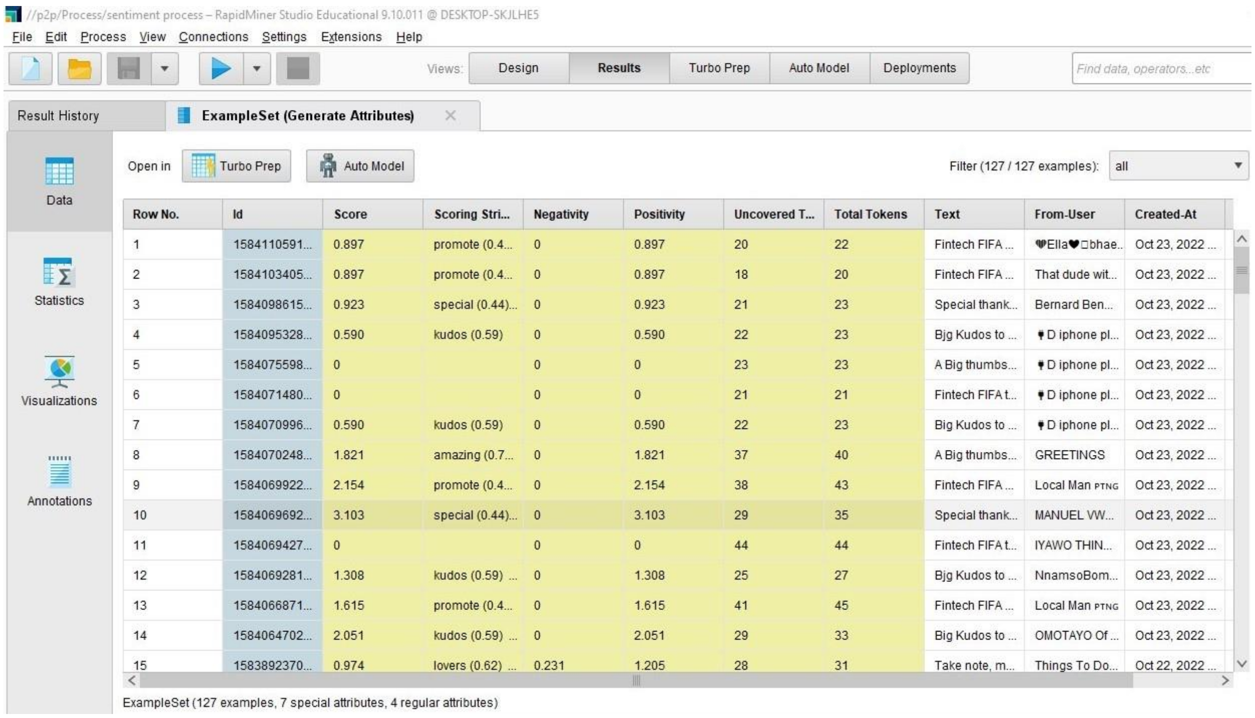
[33] J.-T. Han, Q. Chen, J.-G. Liu, X.-L. Luo, and W. Fan, "The persuasion of borrowers' voluntary information in peer-to-peer lending: An empirical study based on elaboration likelihood model," *Computers in Human Behavior*, vol. 78, pp. 200–214, 2018.

APPENDIX

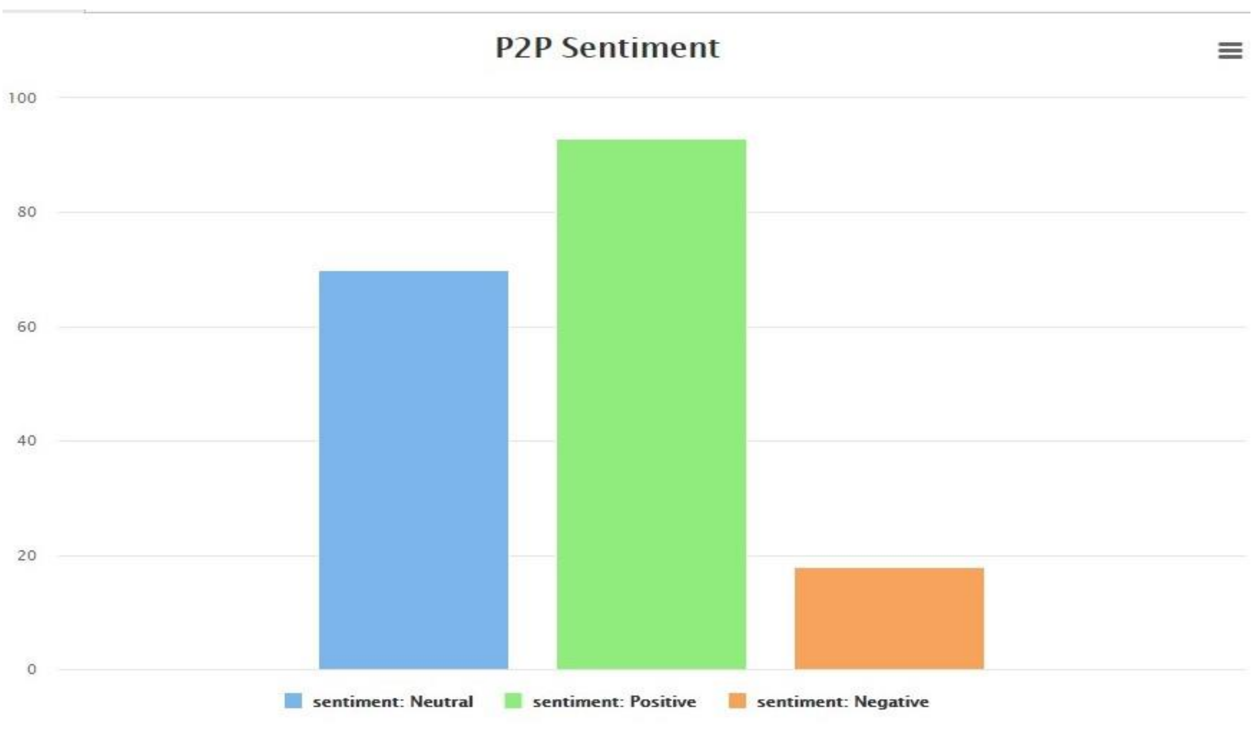
Sentiment Output

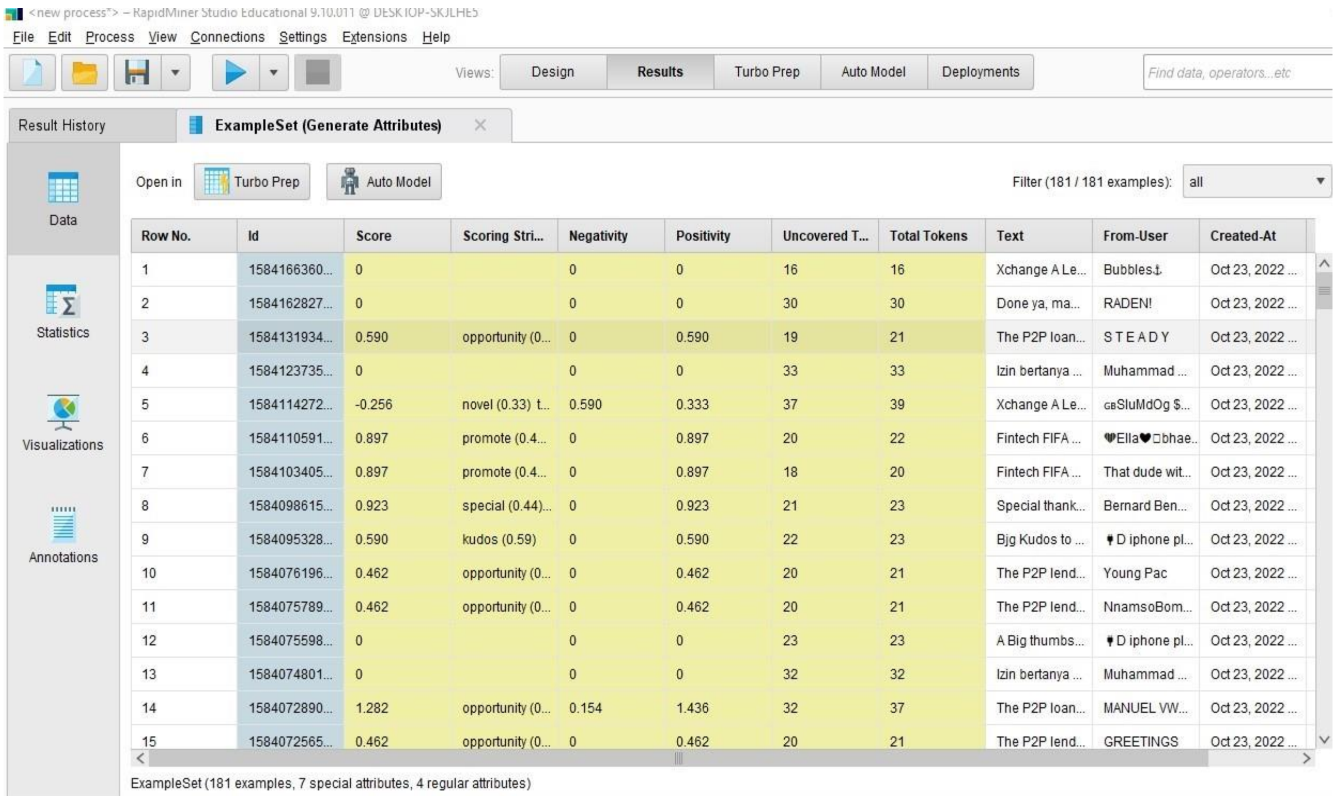
Online Lending





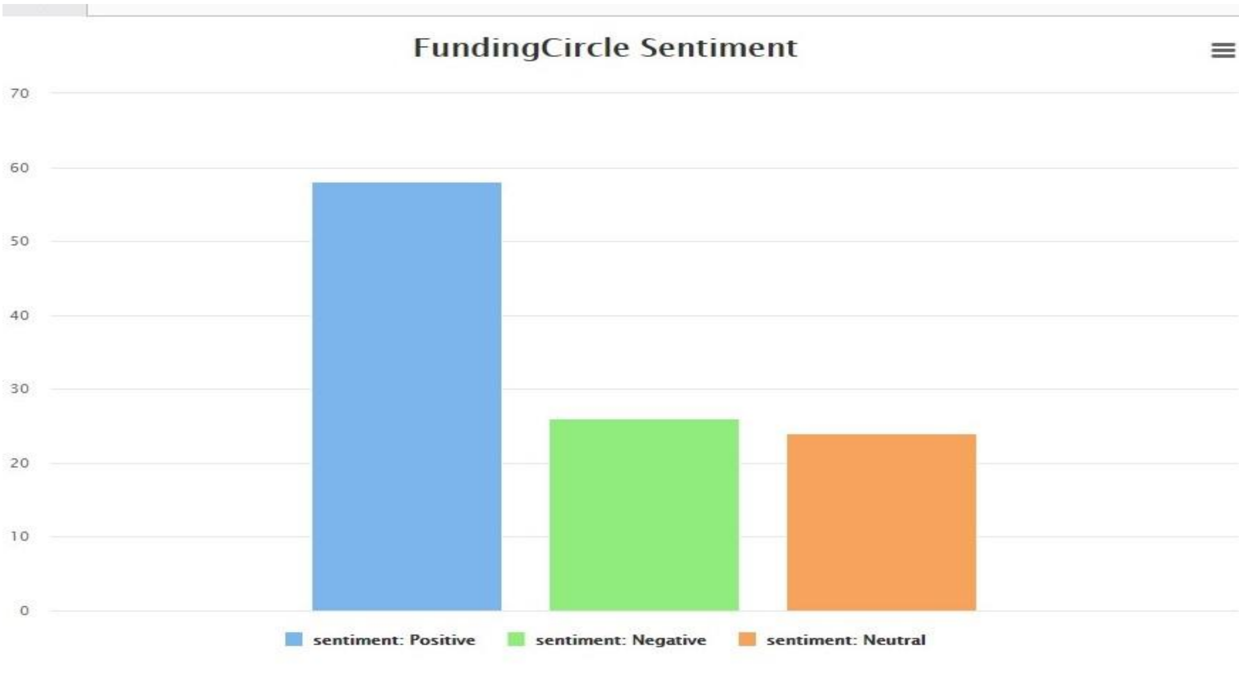
Peer-to-Peer Lending

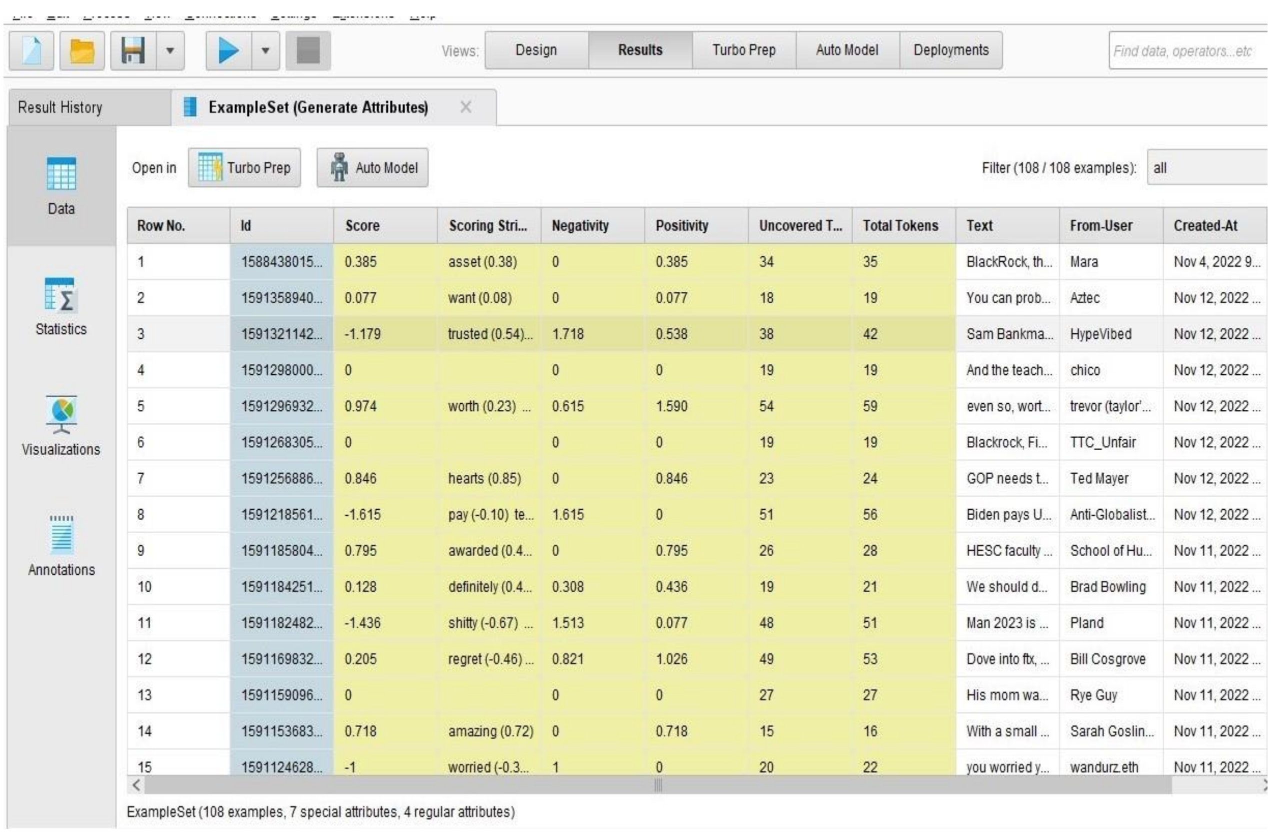




Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1584166360...	0		0	0	16	16	Xchange A Le...	Bubbles⚓	Oct 23, 2022 ...
2	1584162827...	0		0	0	30	30	Done ya, ma...	RADEN!	Oct 23, 2022 ...
3	1584131934...	0.590	opportunity (0...	0	0.590	19	21	The P2P loan...	S T E A D Y	Oct 23, 2022 ...
4	1584123735...	0		0	0	33	33	Izin bertanya ...	Muhammad ...	Oct 23, 2022 ...
5	1584114272...	-0.256	novel (0.33) t...	0.590	0.333	37	39	Xchange A Le...	GBSluMdOg $...	Oct 23, 2022 ...
6	1584110591...	0.897	promote (0.4...	0	0.897	20	22	Fintech FIFA ...	♥Ella♥□bhae...	Oct 23, 2022 ...
7	1584103405...	0.897	promote (0.4...	0	0.897	18	20	Fintech FIFA ...	That dude wit...	Oct 23, 2022 ...
8	1584098615...	0.923	special (0.44)...	0	0.923	21	23	Special thank...	Bernard Ben...	Oct 23, 2022 ...
9	1584095328...	0.590	kudos (0.59)	0	0.590	22	23	Bjg Kudos to ...	D iphone pl...	Oct 23, 2022 ...
10	1584076196...	0.462	opportunity (0...	0	0.462	20	21	The P2P lend...	Young Pac	Oct 23, 2022 ...
11	1584075789...	0.462	opportunity (0...	0	0.462	20	21	The P2P lend...	NnamsoBom...	Oct 23, 2022 ...
12	1584075598...	0		0	0	23	23	A Big thumbs...	D iphone pl...	Oct 23, 2022 ...
13	1584074801...	0		0	0	32	32	Izin bertanya ...	Muhammad ...	Oct 23, 2022 ...
14	1584072890...	1.282	opportunity (0...	0.154	1.436	32	37	The P2P loan...	MANUEL VW...	Oct 23, 2022 ...
15	1584072565...	0.462	opportunity (0...	0	0.462	20	21	The P2P lend...	GREETINGS	Oct 23, 2022 ...

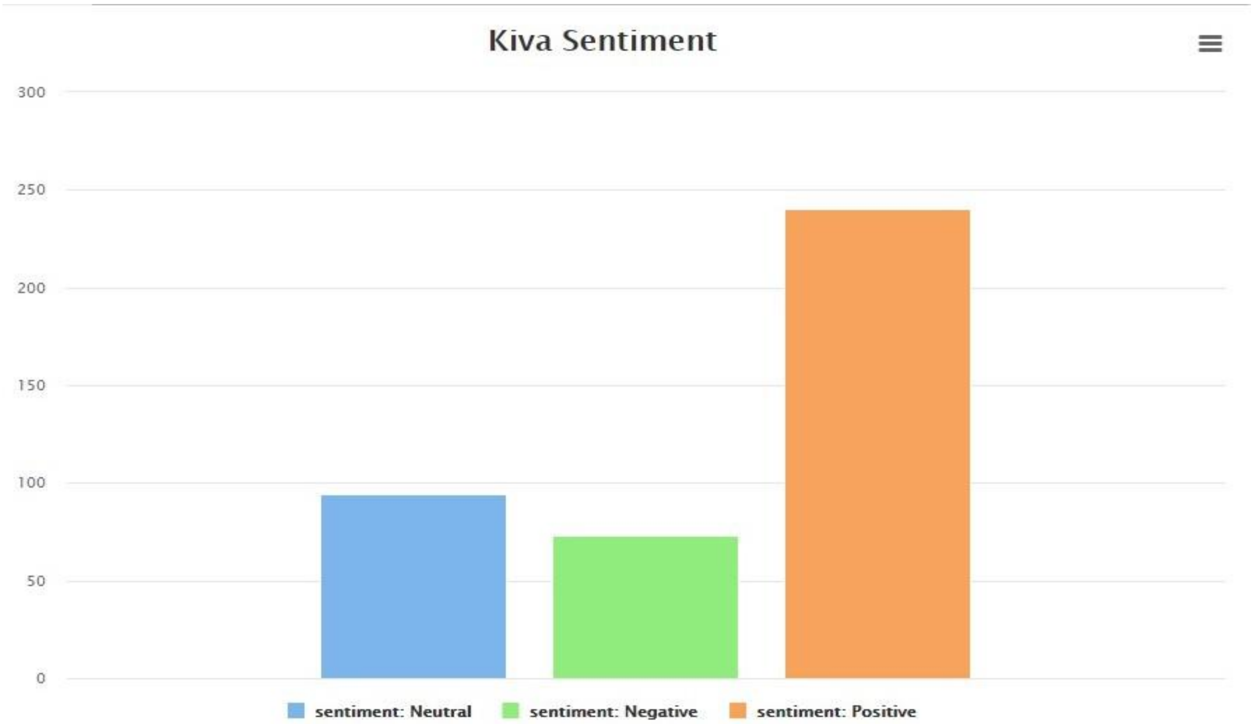
Funding Circle

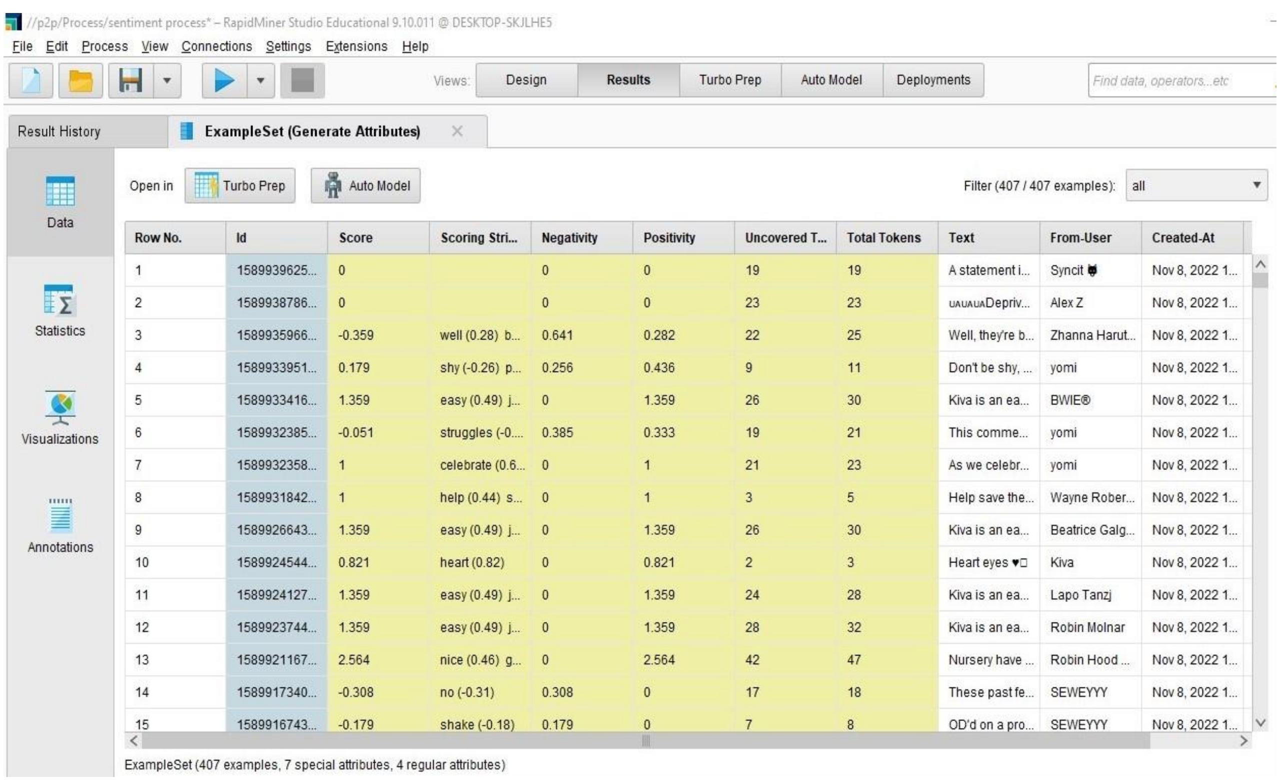




Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1588438015...	0.385	asset (0.38)	0	0.385	34	35	BlackRock, th...	Mara	Nov 4, 2022 9...
2	1591358940...	0.077	want (0.08)	0	0.077	18	19	You can prob...	Aztec	Nov 12, 2022 ...
3	1591321142...	-1.179	trusted (0.54)...	1.718	0.538	38	42	Sam Bankma...	HypeVibed	Nov 12, 2022 ...
4	1591298000...	0		0	0	19	19	And the teach...	chico	Nov 12, 2022 ...
5	1591296932...	0.974	worth (0.23) ...	0.615	1.590	54	59	even so, wort...	trevor (taylor'...	Nov 12, 2022 ...
6	1591268305...	0		0	0	19	19	Blackrock, Fi...	TTC_Unfair	Nov 12, 2022 ...
7	1591256886...	0.846	hearts (0.85)	0	0.846	23	24	GOP needs t...	Ted Mayer	Nov 12, 2022 ...
8	1591218561...	-1.615	pay (-0.10) te...	1.615	0	51	56	Biden pays U...	Anti-Globalist...	Nov 12, 2022 ...
9	1591185804...	0.795	awarded (0.4...	0	0.795	26	28	HESC faculty ...	School of Hu...	Nov 11, 2022 ...
10	1591184251...	0.128	definitely (0.4...	0.308	0.436	19	21	We should d...	Brad Bowling	Nov 11, 2022 ...
11	1591182482...	-1.436	shitty (-0.67) ...	1.513	0.077	48	51	Man 2023 is ...	Pland	Nov 11, 2022 ...
12	1591169832...	0.205	regret (-0.46) ...	0.821	1.026	49	53	Dove into ftx, ...	Bill Cosgrove	Nov 11, 2022 ...
13	1591159096...	0		0	0	27	27	His mom wa...	Rye Guy	Nov 11, 2022 ...
14	1591153683...	0.718	amazing (0.72)	0	0.718	15	16	With a small ...	Sarah Goslin...	Nov 11, 2022 ...
15	1591124628...	-1	worried (-0.3...	1	0	20	22	you worried y...	wandurz.eth	Nov 11, 2022 ...

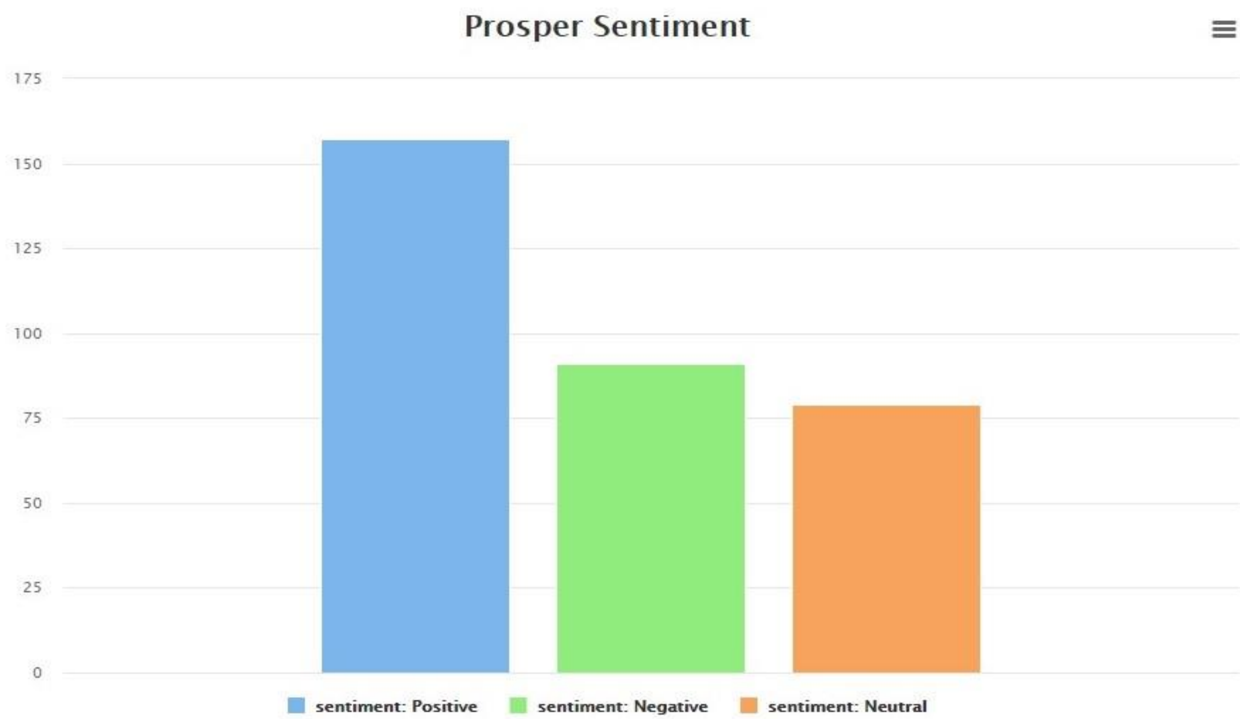
Kiva

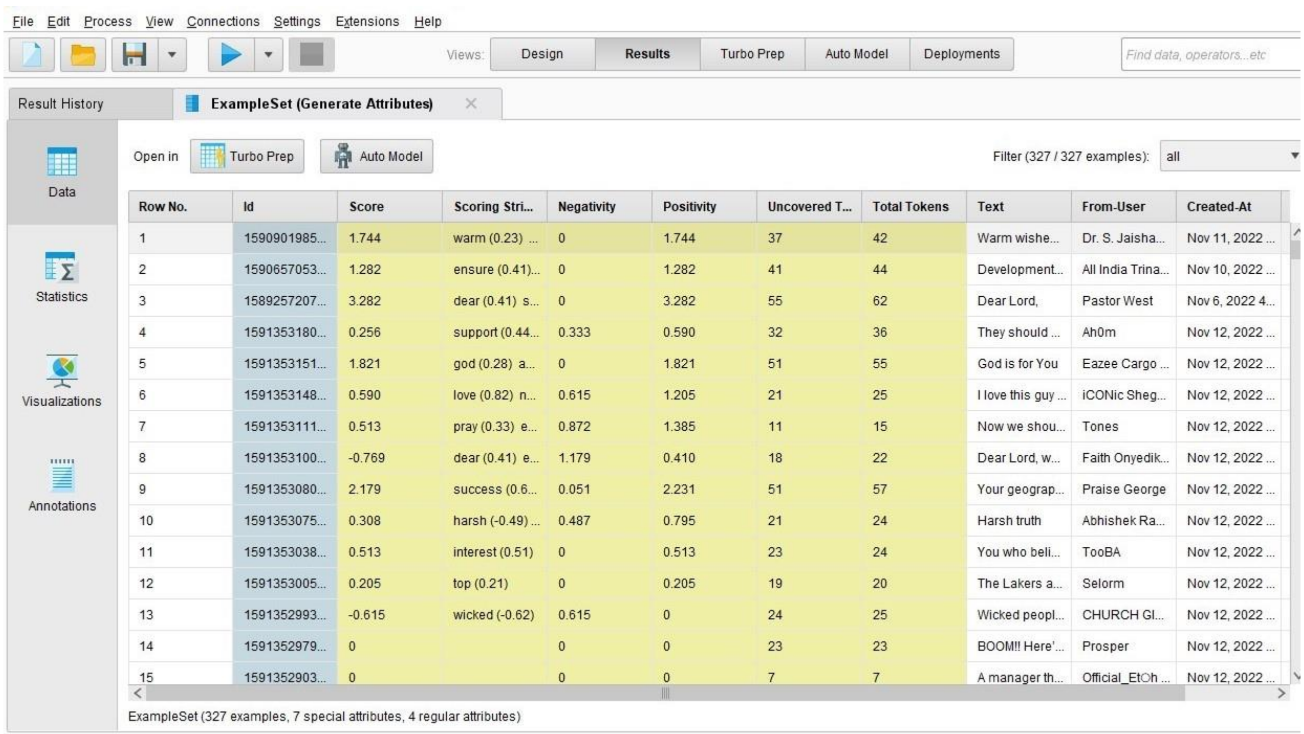




Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1589939625...	0		0	0	19	19	A statement i...	Syncit	Nov 8, 2022 1...
2	1589938786...	0		0	0	23	23	UAUAUADepriv...	Alex Z	Nov 8, 2022 1...
3	1589935966...	-0.359	well (0.28) b...	0.641	0.282	22	25	Well, they're b...	Zhanna Harut...	Nov 8, 2022 1...
4	1589933951...	0.179	shy (-0.26) p...	0.256	0.436	9	11	Don't be shy, ...	yomi	Nov 8, 2022 1...
5	1589933416...	1.359	easy (0.49) j...	0	1.359	26	30	Kiva is an ea...	BWIE®	Nov 8, 2022 1...
6	1589932385...	-0.051	struggles (-0....	0.385	0.333	19	21	This comme...	yomi	Nov 8, 2022 1...
7	1589932358...	1	celebrate (0.6...	0	1	21	23	As we celebr...	yomi	Nov 8, 2022 1...
8	1589931842...	1	help (0.44) s...	0	1	3	5	Help save the...	Wayne Rober...	Nov 8, 2022 1...
9	1589926643...	1.359	easy (0.49) j...	0	1.359	26	30	Kiva is an ea...	Beatrice Galg...	Nov 8, 2022 1...
10	1589924544...	0.821	heart (0.82)	0	0.821	2	3	Heart eyes ♥□	Kiva	Nov 8, 2022 1...
11	1589924127...	1.359	easy (0.49) j...	0	1.359	24	28	Kiva is an ea...	Lapo Tanzj	Nov 8, 2022 1...
12	1589923744...	1.359	easy (0.49) j...	0	1.359	28	32	Kiva is an ea...	Robin Molnar	Nov 8, 2022 1...
13	1589921167...	2.564	nice (0.46) g...	0	2.564	42	47	Nursery have ...	Robin Hood ...	Nov 8, 2022 1...
14	1589917340...	-0.308	no (-0.31)	0.308	0	17	18	These past fe...	SEWEYYY	Nov 8, 2022 1...
15	1589916743...	-0.179	shake (-0.18)	0.179	0	7	8	OD'd on a pro...	SEWEYYY	Nov 8, 2022 1...

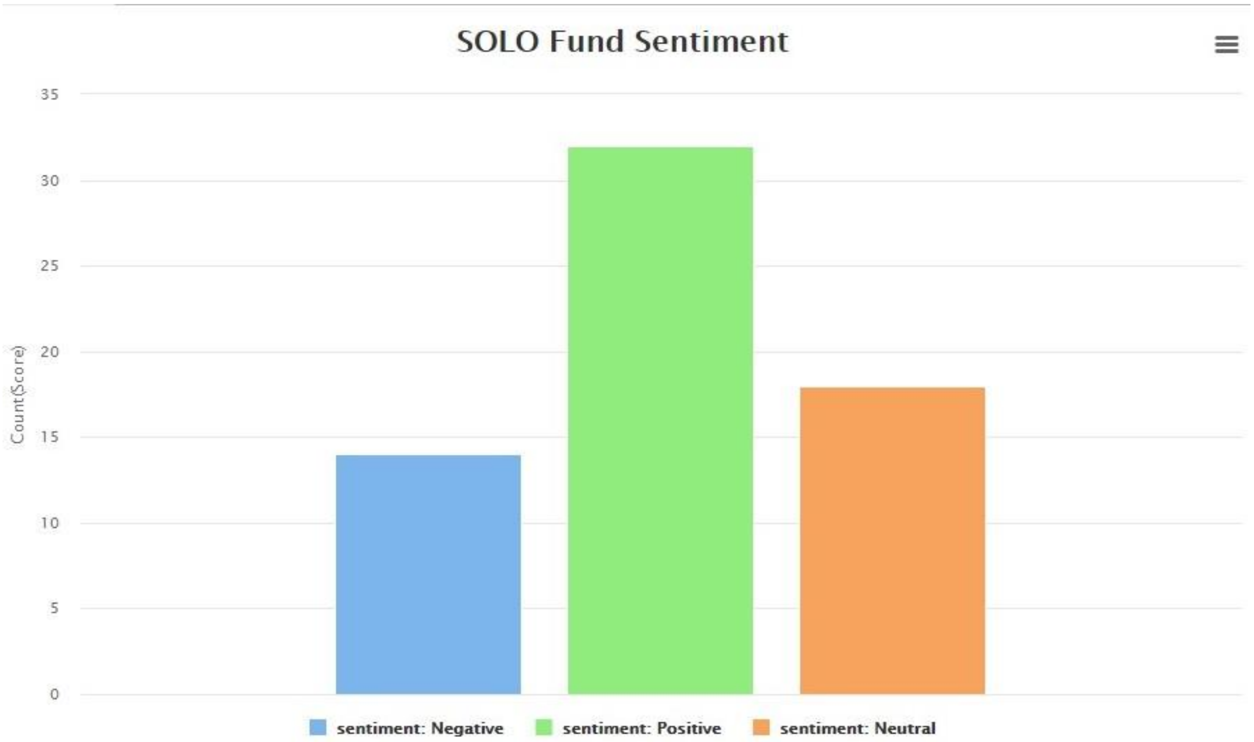
Prosper

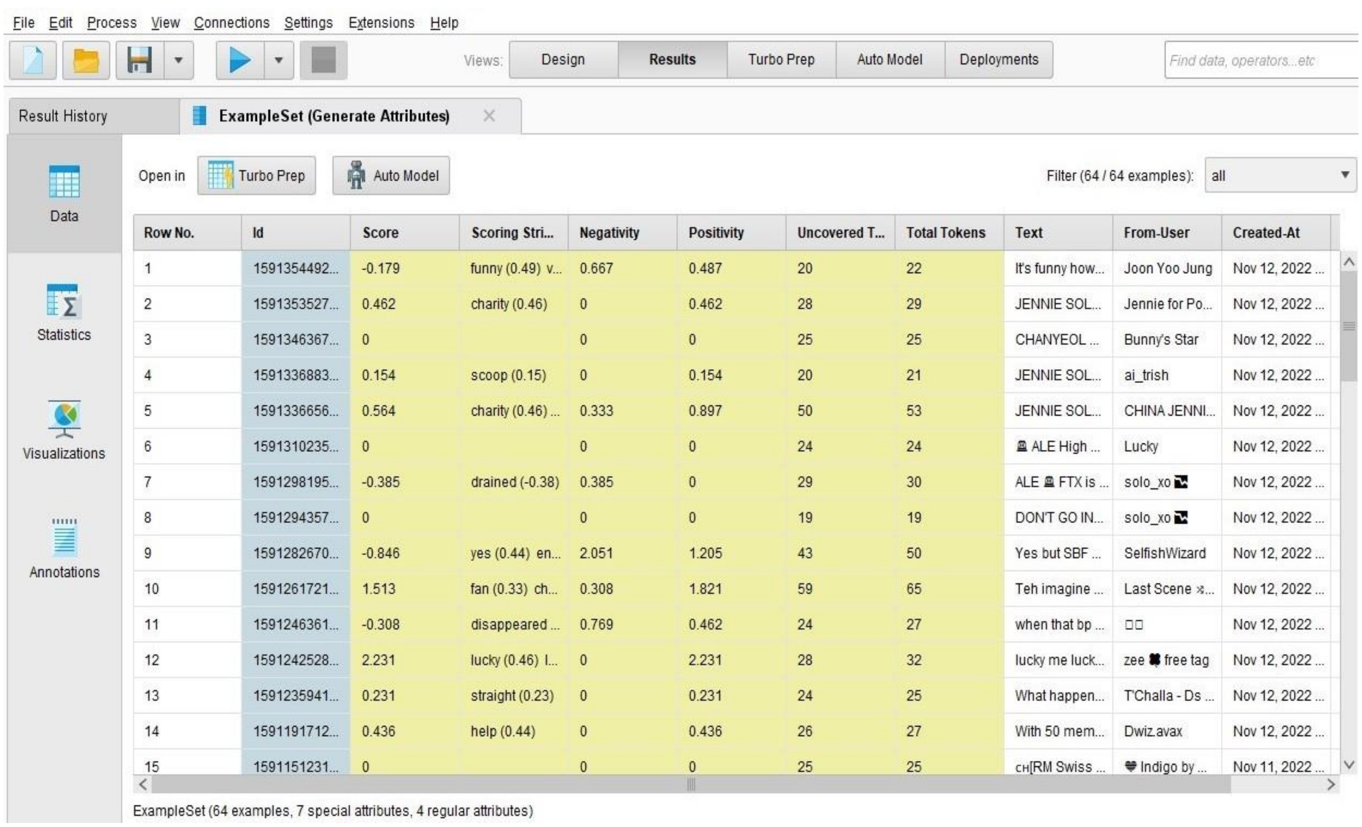

Prosper Sentiment
175
150
125
100
75
50
25
0
sentiment: Positive
sentiment: Negative
sentiment: Neutral



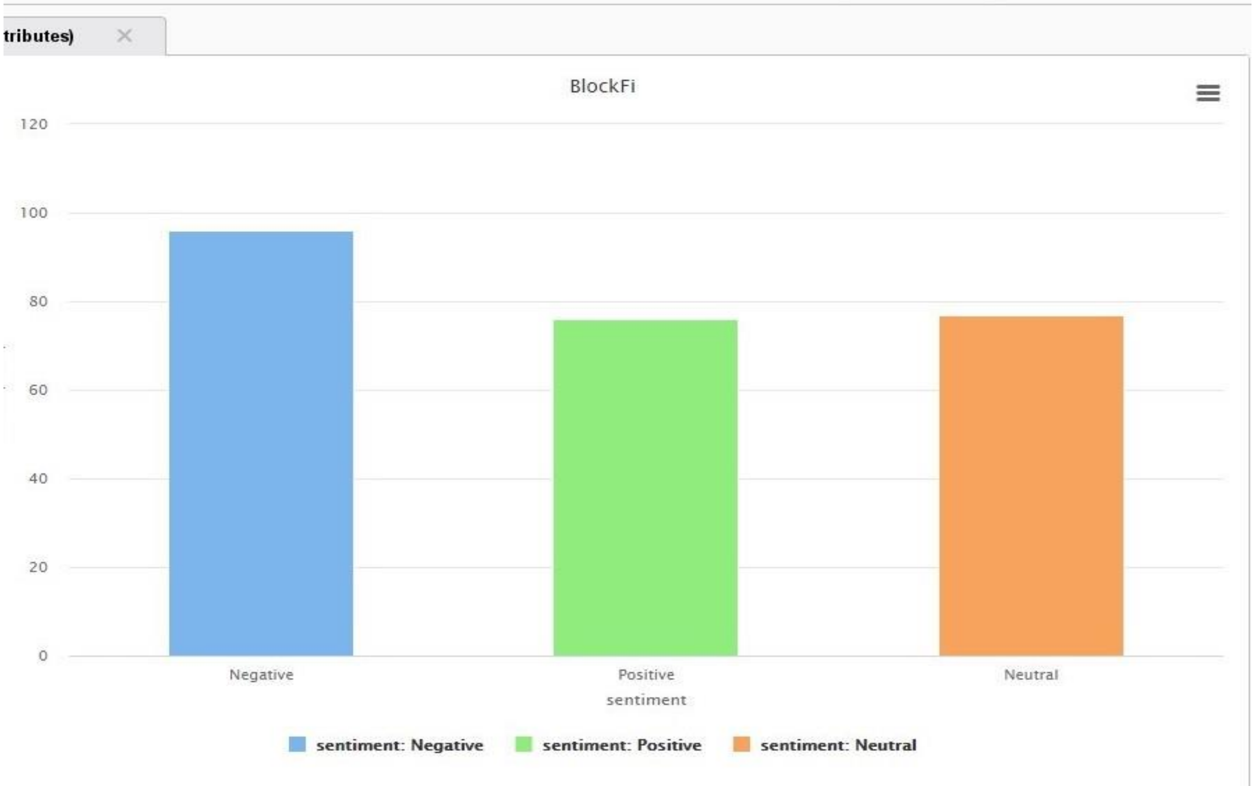
Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1590901985...	1.744	warm (0.23) ...	0	1.744	37	42	Warm wishe...	Dr. S. Jaisha...	Nov 11, 2022 ...
2	1590657053...	1.282	ensure (0.41)...	0	1.282	41	44	Development...	All India Trina...	Nov 10, 2022 ...
3	1589257207...	3.282	dear (0.41) s...	0	3.282	55	62	Dear Lord,	Pastor West	Nov 6, 2022 4...
4	1591353180...	0.256	support (0.44...	0.333	0.590	32	36	They should ...	Ah0m	Nov 12, 2022 ...
5	1591353151...	1.821	god (0.28) a...	0	1.821	51	55	God is for You	Eazee Cargo ...	Nov 12, 2022 ...
6	1591353148...	0.590	love (0.82) n...	0.615	1.205	21	25	I love this guy ...	iCONic Sheg...	Nov 12, 2022 ...
7	1591353111...	0.513	pray (0.33) e...	0.872	1.385	11	15	Now we shou...	Tones	Nov 12, 2022 ...
8	1591353100...	-0.769	dear (0.41) e...	1.179	0.410	18	22	Dear Lord, w...	Faith Onyedik...	Nov 12, 2022 ...
9	1591353080...	2.179	success (0.6...	0.051	2.231	51	57	Your geograp...	Praise George	Nov 12, 2022 ...
10	1591353075...	0.308	harsh (-0.49) ...	0.487	0.795	21	24	Harsh truth	Abhishek Ra...	Nov 12, 2022 ...
11	1591353038...	0.513	interest (0.51)	0	0.513	23	24	You who beli...	TooBA	Nov 12, 2022 ...
12	1591353005...	0.205	top (0.21)	0	0.205	19	20	The Lakers a...	Selorm	Nov 12, 2022 ...
13	1591352993...	-0.615	wicked (-0.62)	0.615	0	24	25	Wicked peopl...	CHURCH Gl...	Nov 12, 2022 ...
14	1591352979...	0		0	0	23	23	BOOM!! Here'...	Prosper	Nov 12, 2022 ...
15	1591352903...	0		0	0	7	7	A manager th...	Official_EtOh ...	Nov 12, 2022 ...

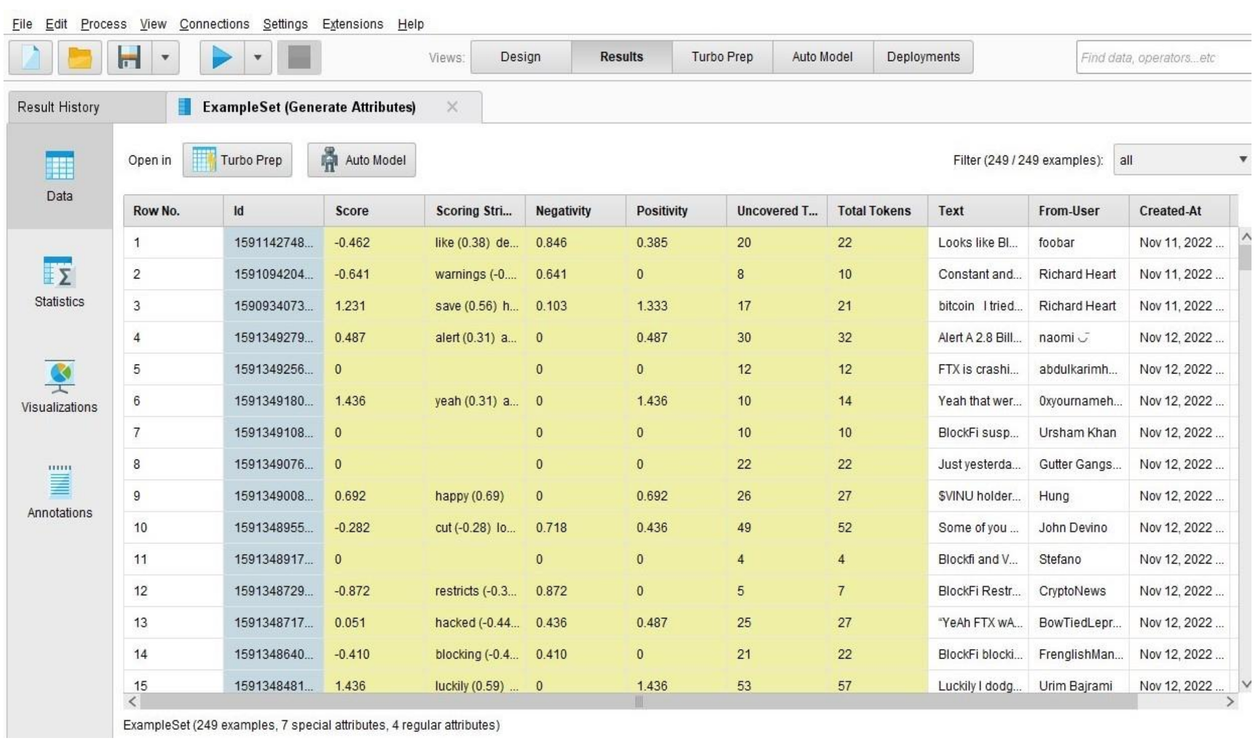
Solo Fund


SOLO Fund Sentiment
Count(Score)
35
30
25
20
15
10
5
0
sentiment: Negative
sentiment: Positive
sentiment: Neutral


File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
Result History
ExampleSet (Generate Attributes)
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (64 / 64 examples): all
Row No. | Id | Score | Scoring Stri... | Negativity | Positivity | Uncovered T... | Total Tokens | Text | From-User | Created-At
1 | 1591354492... | -0.179 | funny (0.49) v... | 0.667 | 0.487 | 20 | 22 | It's funny how... | Joon Yoo Jung | Nov 12, 2022 ...
2 | 1591353527... | 0.462 | charity (0.46) | 0 | 0.462 | 28 | 29 | JENNIE SOL... | Jennie for Po... | Nov 12, 2022 ...
3 | 1591346367... | 0 | | 0 | 0 | 25 | 25 | CHANYEOL ... | Bunny's Star | Nov 12, 2022 ...
4 | 1591336883... | 0.154 | scoop (0.15) | 0 | 0.154 | 20 | 21 | JENNIE SOL... | ai_trish | Nov 12, 2022 ...
5 | 1591336656... | 0.564 | charity (0.46) ... | 0.333 | 0.897 | 50 | 53 | JENNIE SOL... | CHINA JENNI... | Nov 12, 2022 ...
6 | 1591310235... | 0 | | 0 | 0 | 24 | 24 | ALE High ... | Lucky | Nov 12, 2022 ...
7 | 1591298195... | -0.385 | drained (-0.38) | 0.385 | 0 | 29 | 30 | ALE FTX is ... | solo_xo | Nov 12, 2022 ...
8 | 1591294357... | 0 | | 0 | 0 | 19 | 19 | DON'T GO IN... | solo_xo | Nov 12, 2022 ...
9 | 1591282670... | -0.846 | yes (0.44) en... | 2.051 | 1.205 | 43 | 50 | Yes but SBF ... | SelfishWizard | Nov 12, 2022 ...
10 | 1591261721... | 1.513 | fan (0.33) ch... | 0.308 | 1.821 | 59 | 65 | Teh imagine ... | Last Scene ... | Nov 12, 2022 ...
11 | 1591246361... | -0.308 | disappeared ... | 0.769 | 0.462 | 24 | 27 | when that bp ... | | Nov 12, 2022 ...
12 | 1591242528... | 2.231 | lucky (0.46) l... | 0 | 2.231 | 28 | 32 | lucky me luck... | zee free tag | Nov 12, 2022 ...
13 | 1591235941... | 0.231 | straight (0.23) | 0 | 0.231 | 24 | 25 | What happen... | T'Challa - Ds ... | Nov 12, 2022 ...
14 | 1591191712... | 0.436 | help (0.44) | 0 | 0.436 | 26 | 27 | With 50 mem... | Dwiz.avax | Nov 12, 2022 ...
15 | 1591151231... | 0 | | 0 | 0 | 25 | 25 | [RM Swiss ... | Indigo by ... | Nov 11, 2022 ...
ExampleSet (64 examples, 7 special attributes, 4 regular attributes)

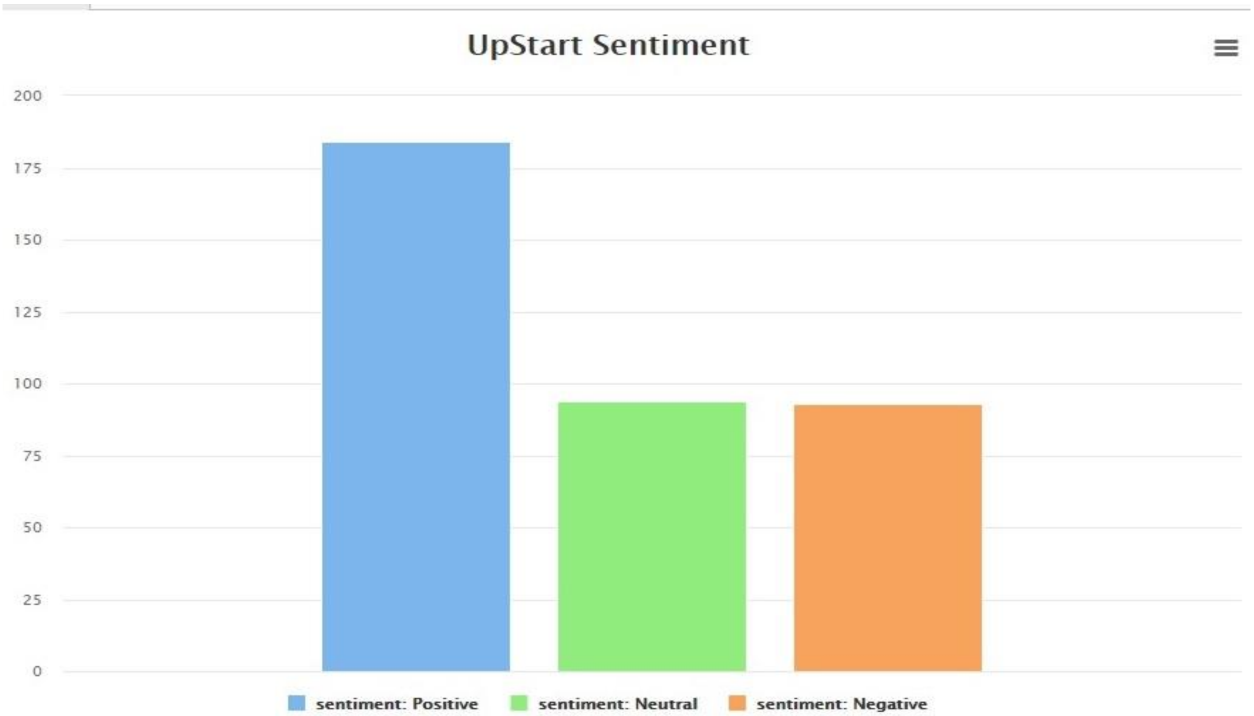
BlockFi

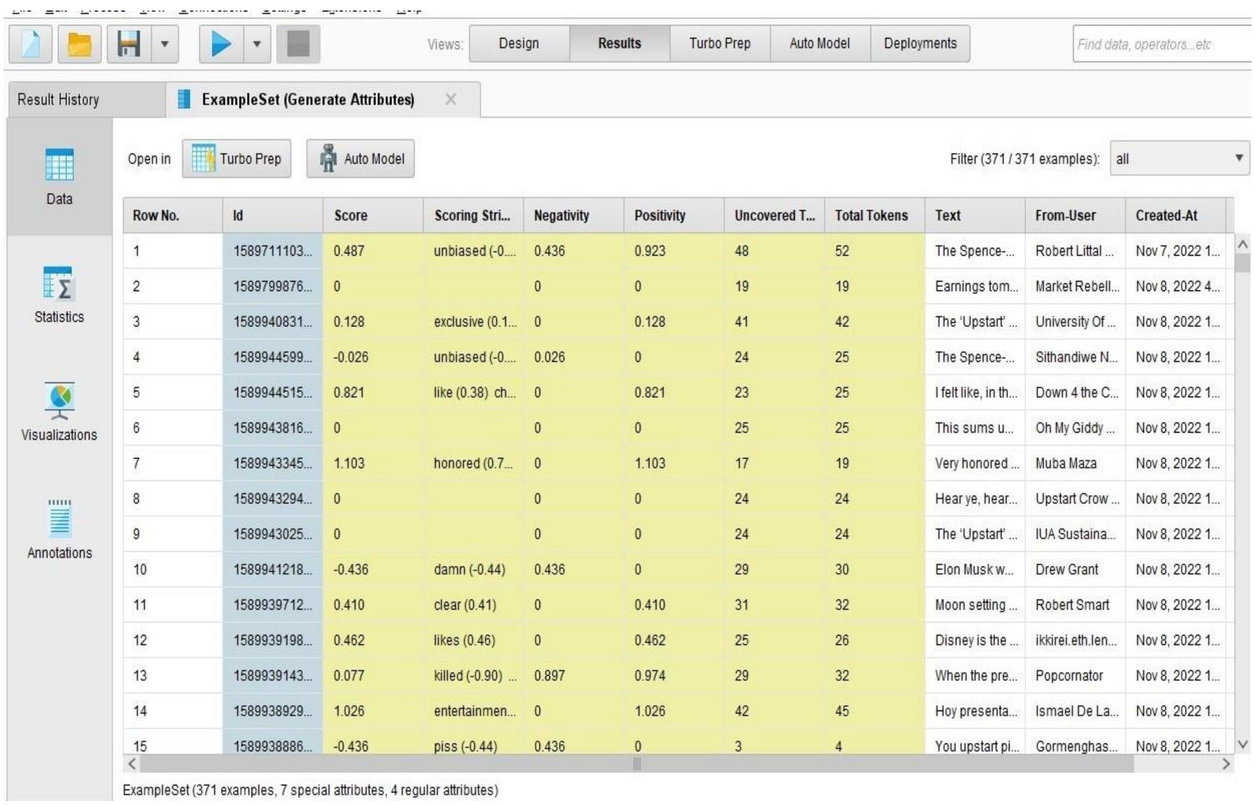




Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1591142748...	-0.462	like (0.38) de...	0.846	0.385	20	22	Looks like Bl...	foobar	Nov 11, 2022 ...
2	1591094204...	-0.641	warnings (-0....	0.641	0	8	10	Constant and...	Richard Heart	Nov 11, 2022 ...
3	1590934073...	1.231	save (0.56) h...	0.103	1.333	17	21	bitcoin I tried...	Richard Heart	Nov 11, 2022 ...
4	1591349279...	0.487	alert (0.31) a...	0	0.487	30	32	Alert A 2.8 Bill...	naomi ◡̈	Nov 12, 2022 ...
5	1591349256...	0		0	0	12	12	FTX is crashi...	abdulkarimh...	Nov 12, 2022 ...
6	1591349180...	1.436	yeah (0.31) a...	0	1.436	10	14	Yeah that wer...	0xyournameh...	Nov 12, 2022 ...
7	1591349108...	0		0	0	10	10	BlockFi susp...	Ursham Khan	Nov 12, 2022 ...
8	1591349076...	0		0	0	22	22	Just yesterda...	Gutter Gangs...	Nov 12, 2022 ...
9	1591349008...	0.692	happy (0.69)	0	0.692	26	27	$VINU holder...	Hung	Nov 12, 2022 ...
10	1591348955...	-0.282	cut (-0.28) lo...	0.718	0.436	49	52	Some of you ...	John Devino	Nov 12, 2022 ...
11	1591348917...	0		0	0	4	4	Blockfi and V...	Stefano	Nov 12, 2022 ...
12	1591348729...	-0.872	restricts (-0.3...	0.872	0	5	7	BlockFi Restr...	CryptoNews	Nov 12, 2022 ...
13	1591348717...	0.051	hacked (-0.44...	0.436	0.487	25	27	"YeAh FTX wA...	BowTiedLepr...	Nov 12, 2022 ...
14	1591348640...	-0.410	blocking (-0.4...	0.410	0	21	22	BlockFi blocki...	FrenglishMan...	Nov 12, 2022 ...
15	1591348481...	1.436	luckily (0.59) ...	0	1.436	53	57	Luckily I dodg...	Urim Bajrami	Nov 12, 2022 ...

Upstart


UpStart Sentiment
200
175
150
125
100
75
50
25
0
sentiment: Positive
sentiment: Neutral
sentiment: Negative

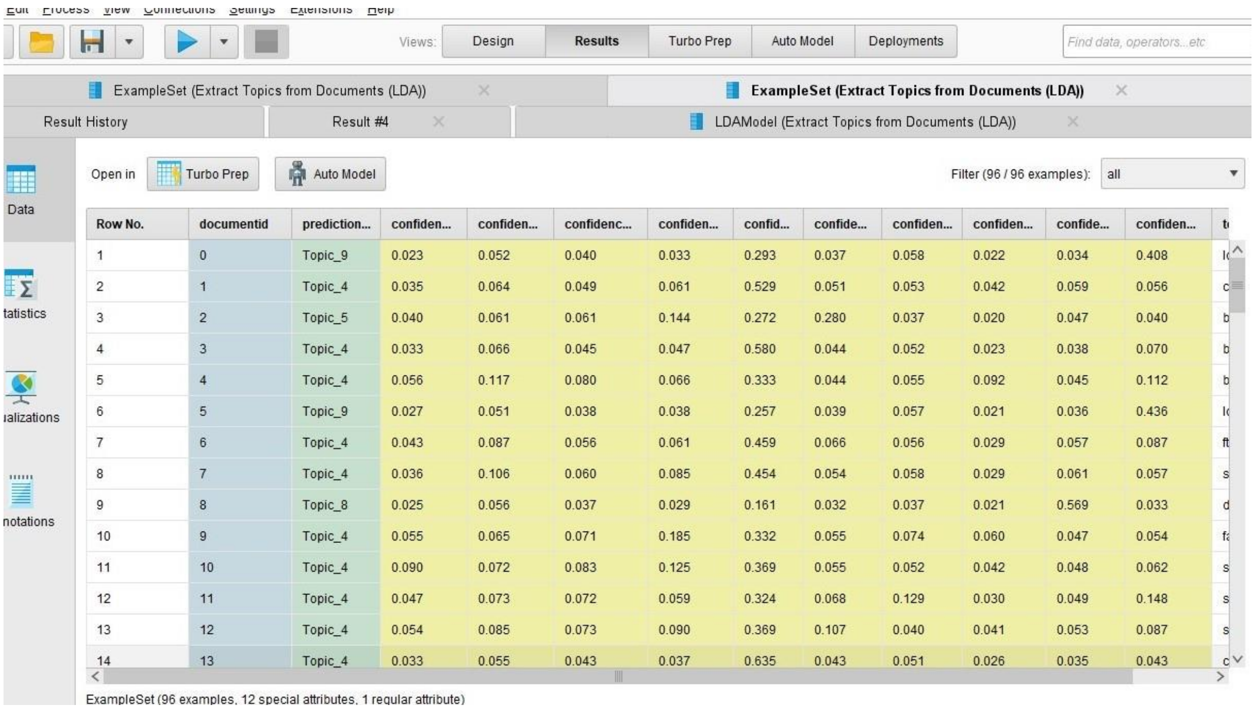


Row No.	Id	Score	Scoring Stri...	Negativity	Positivity	Uncovered T...	Total Tokens	Text	From-User	Created-At
1	1589711103...	0.487	unbiased (-0....	0.436	0.923	48	52	The Spence-...	Robert Littal ...	Nov 7, 2022 1...
2	1589799876...	0		0	0	19	19	Earnings tom...	Market Rebell...	Nov 8, 2022 4...
3	1589940831...	0.128	exclusive (0.1...	0	0.128	41	42	The 'Upstart' ...	University Of ...	Nov 8, 2022 1...
4	1589944599...	-0.026	unbiased (-0....	0.026	0	24	25	The Spence-...	Sithandiwe N...	Nov 8, 2022 1...
5	1589944515...	0.821	like (0.38) ch...	0	0.821	23	25	I felt like, in th...	Down 4 the C...	Nov 8, 2022 1...
6	1589943816...	0		0	0	25	25	This sums u...	Oh My Giddy ...	Nov 8, 2022 1...
7	1589943345...	1.103	honored (0.7...	0	1.103	17	19	Very honored ...	Muba Maza	Nov 8, 2022 1...
8	1589943294...	0		0	0	24	24	Hear ye, hear...	Upstart Crow ...	Nov 8, 2022 1...
9	1589943025...	0		0	0	24	24	The 'Upstart' ...	IUA Sustaina...	Nov 8, 2022 1...
10	1589941218...	-0.436	damn (-0.44)	0.436	0	29	30	Elon Musk w...	Drew Grant	Nov 8, 2022 1...
11	1589939712...	0.410	clear (0.41)	0	0.410	31	32	Moon setting ...	Robert Smart	Nov 8, 2022 1...
12	1589939198...	0.462	likes (0.46)	0	0.462	25	26	Disney is the ...	ikkirei.eth.len...	Nov 8, 2022 1...
13	1589939143...	0.077	killed (-0.90) ...	0.897	0.974	29	32	When the pre...	Popcornator	Nov 8, 2022 1...
14	1589938929...	1.026	entertainmen...	0	1.026	42	45	Hoy presenta...	Ismael De La...	Nov 8, 2022 1...
15	1589938886...	-0.436	piss (-0.44)	0.436	0	3	4	You upstart pi...	Gormenghas...	Nov 8, 2022 1...

ExampleSet (371 examples, 7 special attributes, 4 regular attributes)

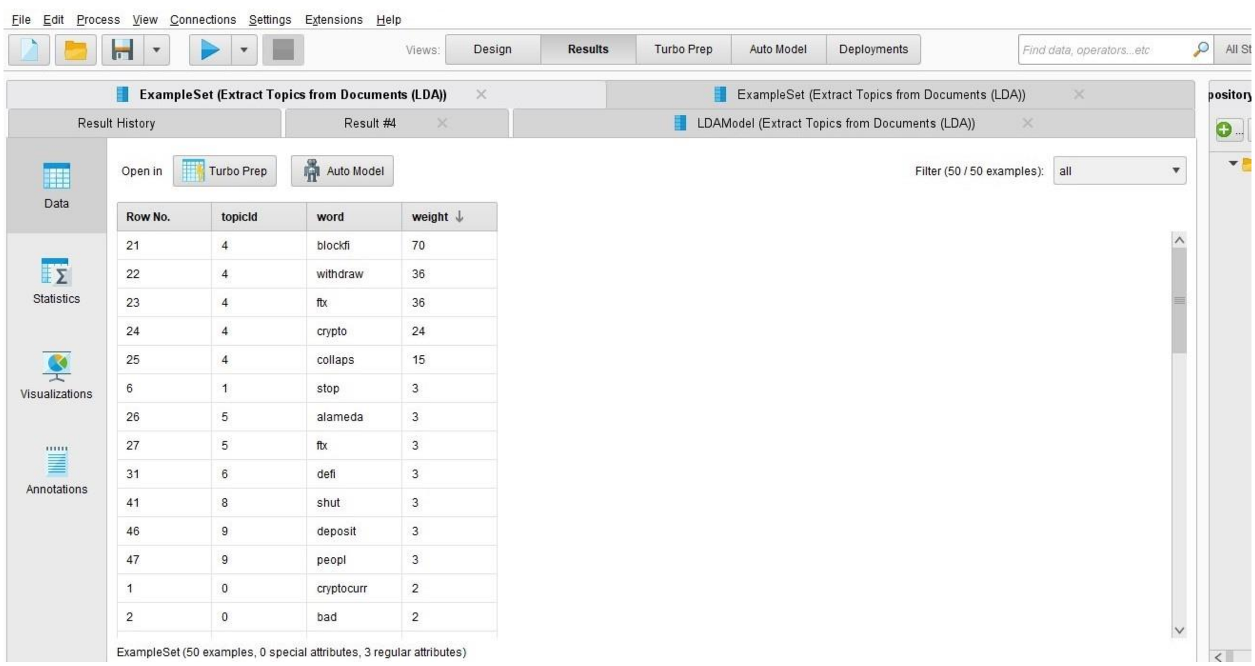
Topic Modelling

BlockFi



Row No.	documentid	prediction...	confiden...	confiden...	confidenc...	confiden...	confid...	confide...	confiden...	confiden...	confide...	confiden...	t
1	0	Topic_9	0.023	0.052	0.040	0.033	0.293	0.037	0.058	0.022	0.034	0.408	l
2	1	Topic_4	0.035	0.064	0.049	0.061	0.529	0.051	0.053	0.042	0.059	0.056	c
3	2	Topic_5	0.040	0.061	0.061	0.144	0.272	0.280	0.037	0.020	0.047	0.040	b
4	3	Topic_4	0.033	0.066	0.045	0.047	0.580	0.044	0.052	0.023	0.038	0.070	b
5	4	Topic_4	0.056	0.117	0.080	0.066	0.333	0.044	0.055	0.092	0.045	0.112	b
6	5	Topic_9	0.027	0.051	0.038	0.038	0.257	0.039	0.057	0.021	0.036	0.436	l
7	6	Topic_4	0.043	0.087	0.056	0.061	0.459	0.066	0.056	0.029	0.057	0.087	ft
8	7	Topic_4	0.036	0.106	0.060	0.085	0.454	0.054	0.058	0.029	0.061	0.057	s
9	8	Topic_8	0.025	0.056	0.037	0.029	0.161	0.032	0.037	0.021	0.569	0.033	d
10	9	Topic_4	0.055	0.065	0.071	0.185	0.332	0.055	0.074	0.060	0.047	0.054	f
11	10	Topic_4	0.090	0.072	0.083	0.125	0.369	0.055	0.052	0.042	0.048	0.062	s
12	11	Topic_4	0.047	0.073	0.072	0.059	0.324	0.068	0.129	0.030	0.049	0.148	s
13	12	Topic_4	0.054	0.085	0.073	0.090	0.369	0.107	0.040	0.041	0.053	0.087	s
14	13	Topic_4	0.033	0.055	0.043	0.037	0.635	0.043	0.051	0.026	0.035	0.043	c

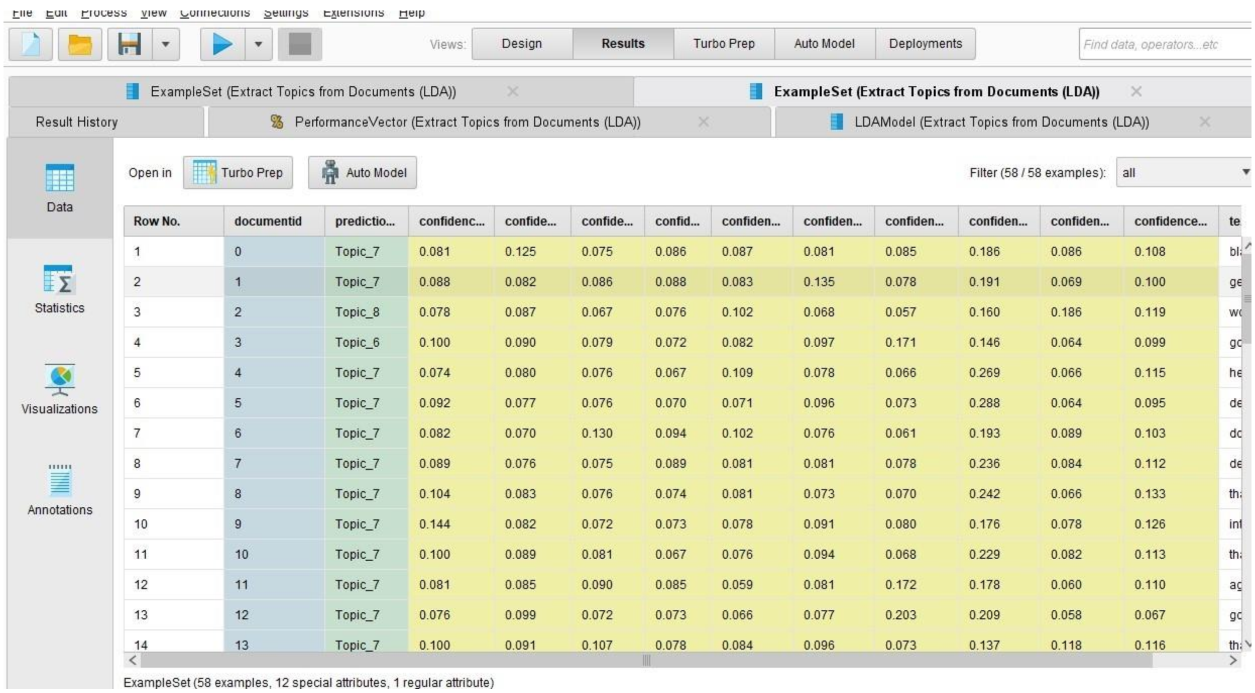
ExampleSet (96 examples, 12 special attributes, 1 regular attribute)

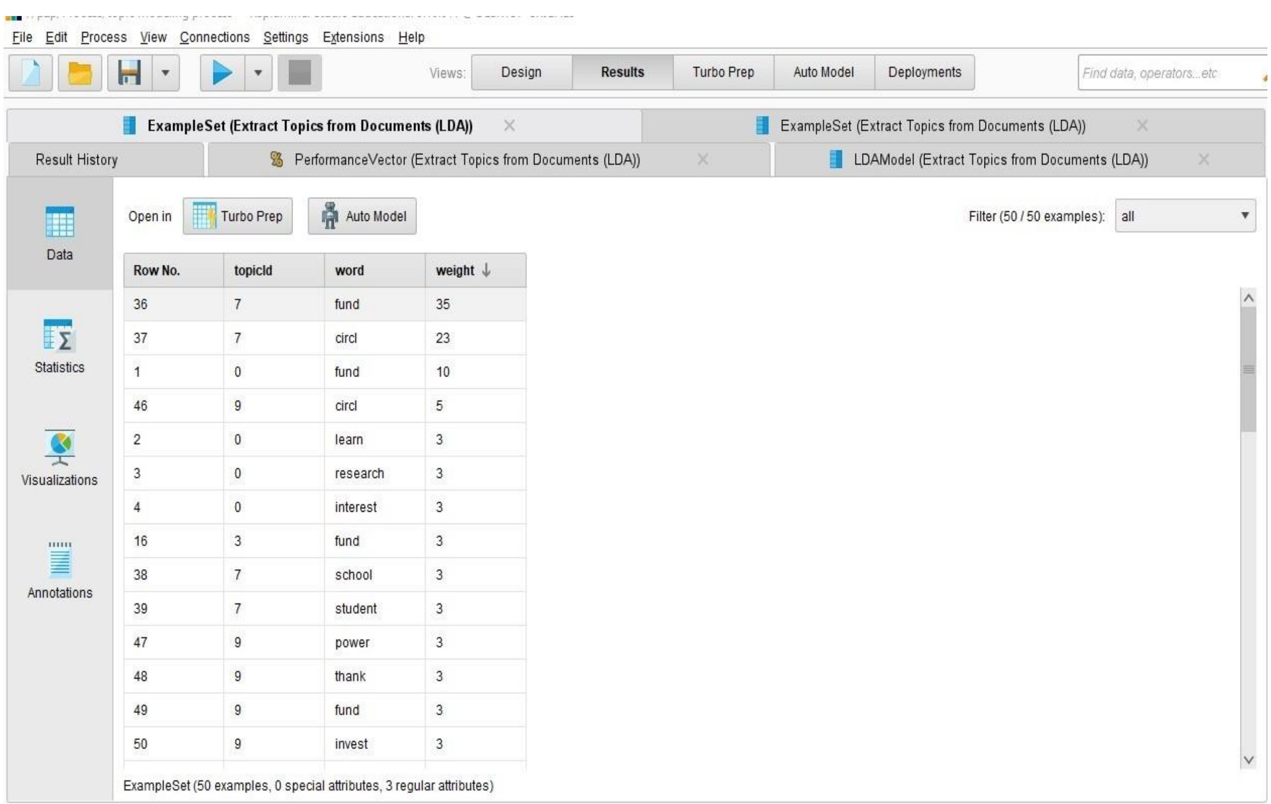


Row No.	topicId	word	weight
21	4	blockfi	70
22	4	withdraw	36
23	4	ftx	36
24	4	crypto	24
25	4	collaps	15
6	1	stop	3
26	5	alameda	3
27	5	ftx	3
31	6	defi	3
41	8	shut	3
46	9	deposit	3
47	9	peopl	3
1	0	cryptocurr	2
2	0	bad	2

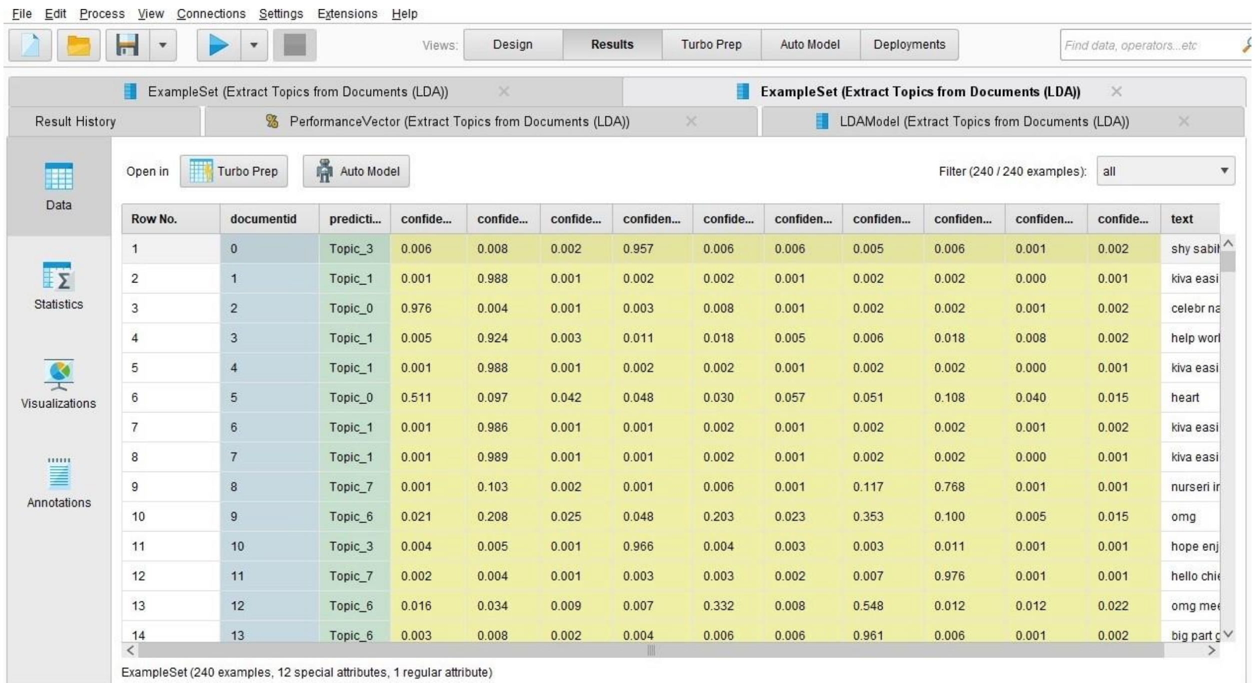
ExampleSet (50 examples, 0 special attributes, 3 regular attributes)

Funding Circle

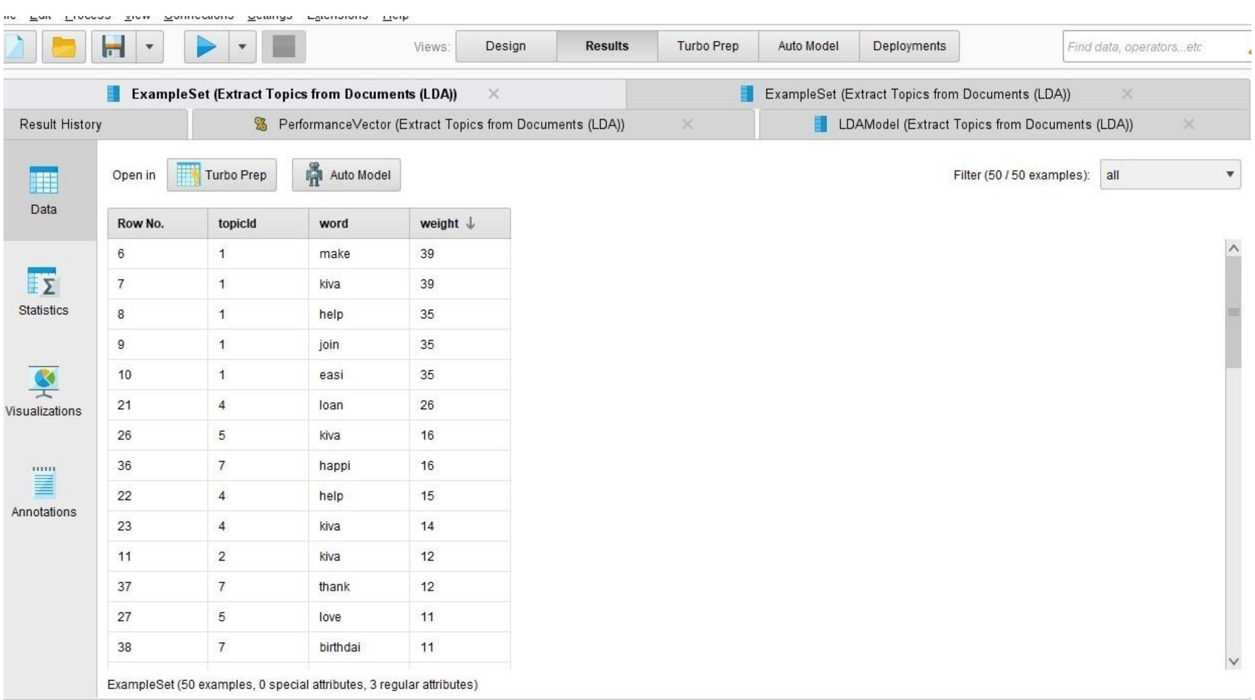

Views: Design Results Turbo Prep Auto Model Deployments
ExampleSet (Extract Topics from Documents (LDA))
ExampleSet (Extract Topics from Documents (LDA))
Result History
PerformanceVector (Extract Topics from Documents (LDA))
LDAModel (Extract Topics from Documents (LDA))
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (58 / 58 examples): all
Row No. documentid predictio... confidenc... confide... confide... confid... confiden... confiden... confiden... confiden... confiden... confidence...
1 0 Topic_7 0.081 0.125 0.075 0.086 0.087 0.081 0.085 0.186 0.086 0.108
2 1 Topic_7 0.088 0.082 0.086 0.088 0.083 0.135 0.078 0.191 0.069 0.100
3 2 Topic_8 0.078 0.087 0.067 0.076 0.102 0.068 0.057 0.160 0.186 0.119
4 3 Topic_6 0.100 0.090 0.079 0.072 0.082 0.097 0.171 0.146 0.064 0.099
5 4 Topic_7 0.074 0.080 0.076 0.067 0.109 0.078 0.066 0.269 0.066 0.115
6 5 Topic_7 0.092 0.077 0.076 0.070 0.071 0.096 0.073 0.288 0.064 0.095
7 6 Topic_7 0.082 0.070 0.130 0.094 0.102 0.076 0.061 0.193 0.089 0.103
8 7 Topic_7 0.089 0.076 0.075 0.089 0.081 0.081 0.078 0.236 0.084 0.112
9 8 Topic_7 0.104 0.083 0.076 0.074 0.081 0.073 0.070 0.242 0.066 0.133
10 9 Topic_7 0.144 0.082 0.072 0.073 0.078 0.091 0.080 0.176 0.078 0.126
11 10 Topic_7 0.100 0.089 0.081 0.067 0.076 0.094 0.068 0.229 0.082 0.113
12 11 Topic_7 0.081 0.085 0.090 0.085 0.059 0.081 0.172 0.178 0.060 0.110
13 12 Topic_7 0.076 0.099 0.072 0.073 0.066 0.077 0.203 0.209 0.058 0.067
14 13 Topic_7 0.100 0.091 0.107 0.078 0.084 0.096 0.073 0.137 0.118 0.116
ExampleSet (58 examples, 12 special attributes, 1 regular attribute)


File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
ExampleSet (Extract Topics from Documents (LDA))
ExampleSet (Extract Topics from Documents (LDA))
Result History
PerformanceVector (Extract Topics from Documents (LDA))
LDAModel (Extract Topics from Documents (LDA))
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (50 / 50 examples): all
Row No. topicId word weight
36 7 fund 35
37 7 circl 23
1 0 fund 10
46 9 circl 5
2 0 learn 3
3 0 research 3
4 0 interest 3
16 3 fund 3
38 7 school 3
39 7 student 3
47 9 power 3
48 9 thank 3
49 9 fund 3
50 9 invest 3
ExampleSet (50 examples, 0 special attributes, 3 regular attributes)

Kiva

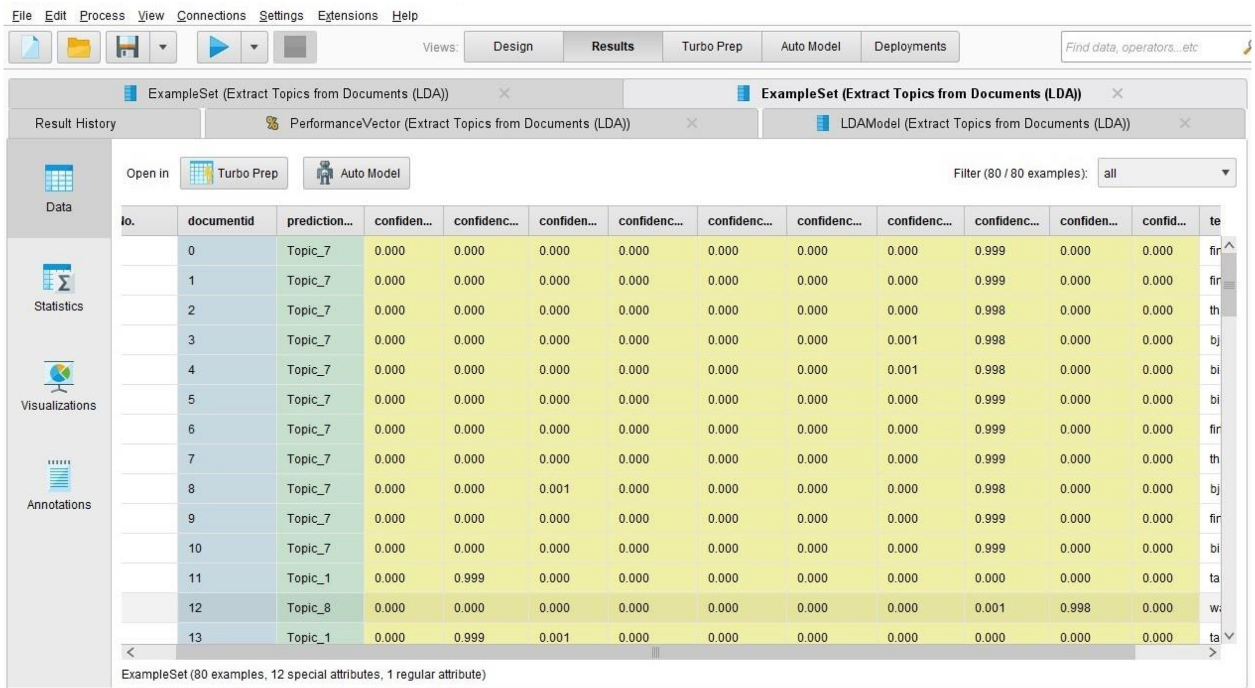


Row No.	documentid	predicti...	confide...	confide...	confide...	confiden...	confide...	confiden...	confiden...	confiden...	confiden...	confide...	text
1	0	Topic_3	0.006	0.008	0.002	0.957	0.006	0.006	0.005	0.006	0.001	0.002	shy sabi
2	1	Topic_1	0.001	0.988	0.001	0.002	0.002	0.001	0.002	0.002	0.000	0.001	kiva easi
3	2	Topic_0	0.976	0.004	0.001	0.003	0.008	0.001	0.002	0.002	0.001	0.002	celebr na
4	3	Topic_1	0.005	0.924	0.003	0.011	0.018	0.005	0.006	0.018	0.008	0.002	help worl
5	4	Topic_1	0.001	0.988	0.001	0.002	0.002	0.001	0.002	0.002	0.000	0.001	kiva easi
6	5	Topic_0	0.511	0.097	0.042	0.048	0.030	0.057	0.051	0.108	0.040	0.015	heart
7	6	Topic_1	0.001	0.986	0.001	0.001	0.002	0.001	0.002	0.002	0.001	0.002	kiva easi
8	7	Topic_1	0.001	0.989	0.001	0.001	0.002	0.001	0.002	0.002	0.000	0.001	kiva easi
9	8	Topic_7	0.001	0.103	0.002	0.001	0.006	0.001	0.117	0.768	0.001	0.001	nurseri ir
10	9	Topic_6	0.021	0.208	0.025	0.048	0.203	0.023	0.353	0.100	0.005	0.015	omg
11	10	Topic_3	0.004	0.005	0.001	0.966	0.004	0.003	0.003	0.011	0.001	0.001	hope enj
12	11	Topic_7	0.002	0.004	0.001	0.003	0.003	0.002	0.007	0.976	0.001	0.001	hello chi
13	12	Topic_6	0.016	0.034	0.009	0.007	0.332	0.008	0.548	0.012	0.012	0.022	omg mee
14	13	Topic_6	0.003	0.008	0.002	0.004	0.006	0.006	0.961	0.006	0.001	0.002	big part c

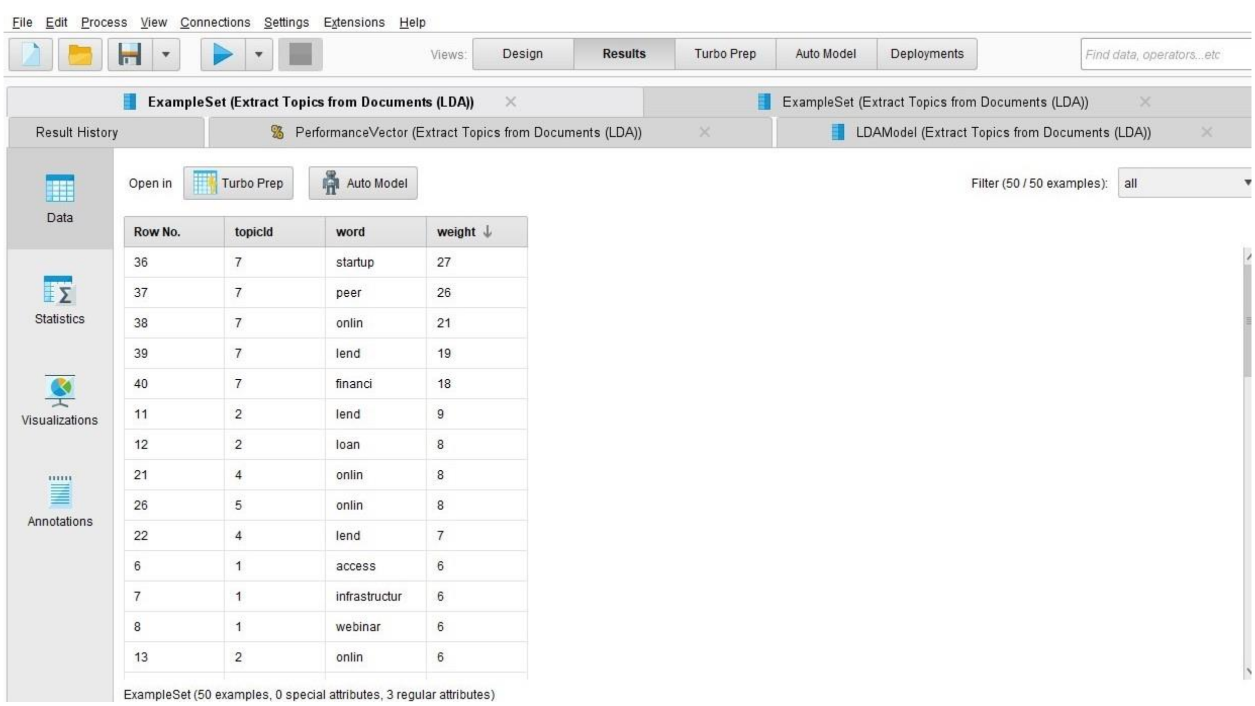


Row No.	topicId	word	weight
6	1	make	39
7	1	kiva	39
8	1	help	35
9	1	join	35
10	1	easi	35
21	4	loan	26
26	5	kiva	16
36	7	happi	16
22	4	help	15
23	4	kiva	14
11	2	kiva	12
37	7	thank	12
27	5	love	11
38	7	birthdai	11

Lending online

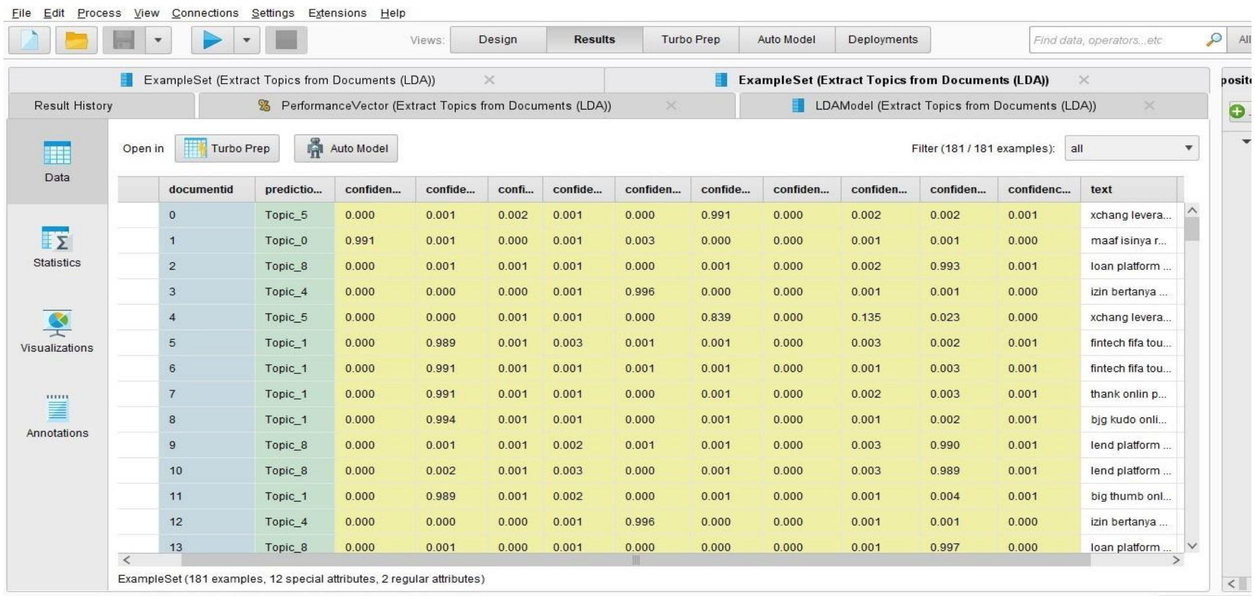


lo.	documentid	prediction...	confiden...	confidenc...	confiden...	confidenc...	confidenc...	confidenc...	confidenc...	confidenc...	confiden...	confid...	te
	0	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	fir
	1	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	fir
	2	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.998	0.000	0.000	th
	3	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.998	0.000	0.000	bj
	4	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.998	0.000	0.000	bi
	5	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	bi
	6	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	fir
	7	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	th
	8	Topic_7	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.998	0.000	0.000	bj
	9	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	fir
	10	Topic_7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	bi
	11	Topic_1	0.000	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	ta
	12	Topic_8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.998	0.000	wa
	13	Topic_1	0.000	0.999	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	ta



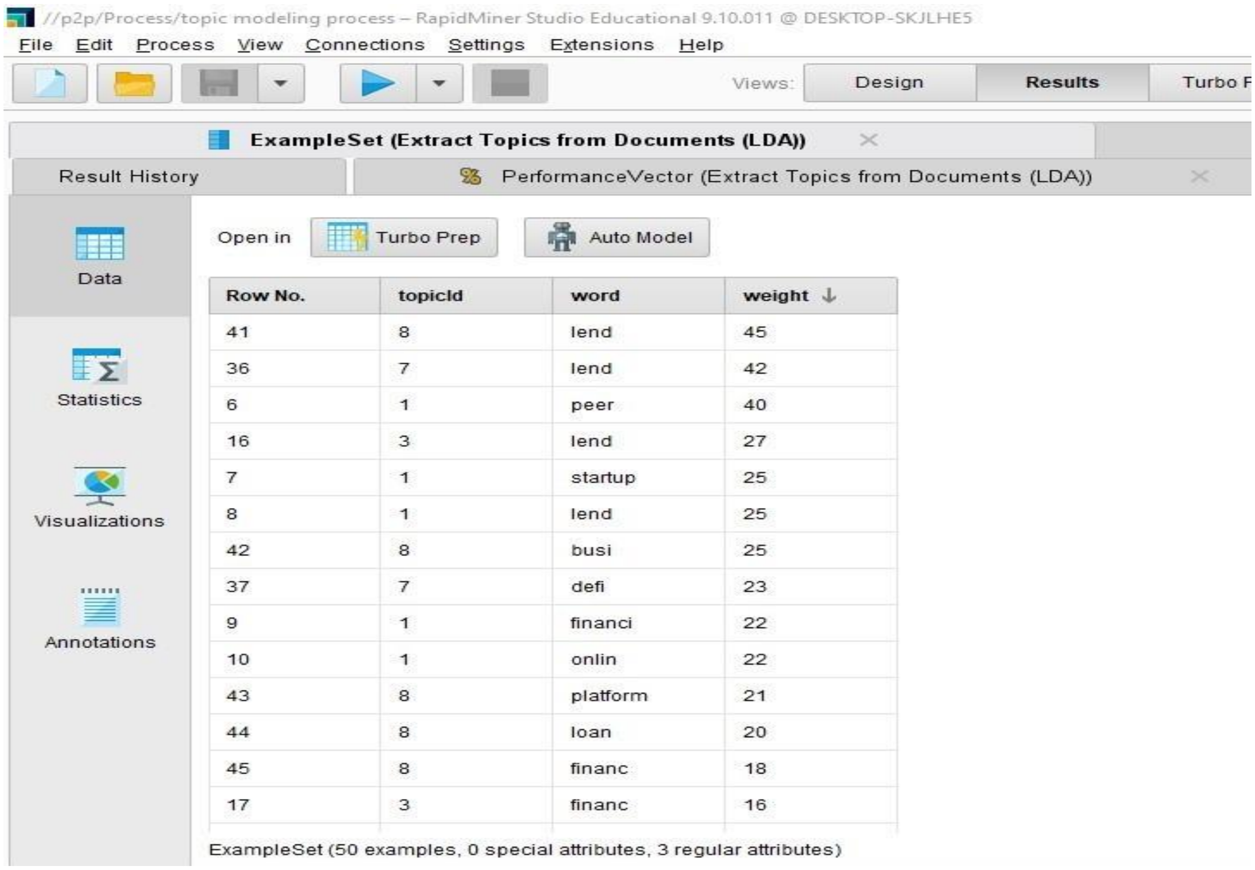
Row No.	topicId	word	weight ↓
36	7	startup	27
37	7	peer	26
38	7	onlin	21
39	7	lend	19
40	7	financi	18
11	2	lend	9
12	2	loan	8
21	4	onlin	8
26	5	onlin	8
22	4	lend	7
6	1	access	6
7	1	infrastructur	6
8	1	webinar	6
13	2	onlin	6

Peer-to-peer



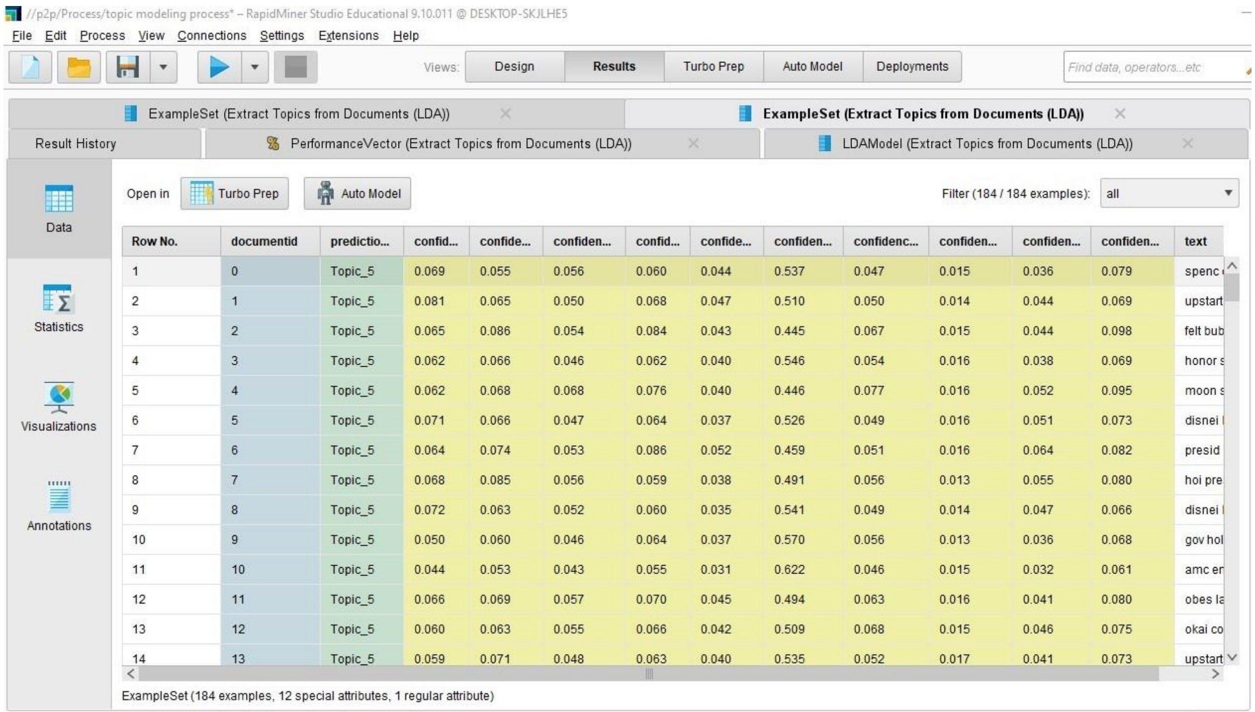
	documentid	predictio...	confiden...	confide...	confi...	confide...	confiden...	confide...	confiden...	confiden...	confiden...	confidenc...	text
	0	Topic_5	0.000	0.001	0.002	0.001	0.000	0.991	0.000	0.002	0.002	0.001	xchang levera...
	1	Topic_0	0.991	0.001	0.000	0.001	0.003	0.000	0.000	0.001	0.001	0.000	maaf isinya r...
	2	Topic_8	0.000	0.001	0.001	0.001	0.000	0.001	0.000	0.002	0.993	0.001	loan platform ...
	3	Topic_4	0.000	0.000	0.000	0.001	0.996	0.000	0.000	0.001	0.001	0.000	izin bertanya ...
	4	Topic_5	0.000	0.000	0.001	0.001	0.000	0.839	0.000	0.135	0.023	0.000	xchang levera...
	5	Topic_1	0.000	0.989	0.001	0.003	0.001	0.001	0.000	0.003	0.002	0.001	fintech fifa tou...
	6	Topic_1	0.000	0.991	0.001	0.001	0.001	0.001	0.000	0.001	0.003	0.001	fintech fifa tou...
	7	Topic_1	0.000	0.991	0.001	0.001	0.000	0.001	0.000	0.002	0.003	0.001	thank onlin p...
	8	Topic_1	0.000	0.994	0.001	0.001	0.000	0.001	0.000	0.001	0.002	0.001	bjg kudo onli...
	9	Topic_8	0.000	0.001	0.001	0.002	0.001	0.001	0.000	0.003	0.990	0.001	lend platform ...
	10	Topic_8	0.000	0.002	0.001	0.003	0.000	0.001	0.000	0.003	0.989	0.001	lend platform ...
	11	Topic_1	0.000	0.989	0.001	0.002	0.000	0.001	0.000	0.001	0.004	0.001	big thumb onl...
	12	Topic_4	0.000	0.000	0.000	0.001	0.996	0.000	0.000	0.001	0.001	0.000	izin bertanya ...
	13	Topic_8	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.001	0.997	0.000	loan platform ...

ExampleSet (181 examples, 12 special attributes, 2 regular attributes)

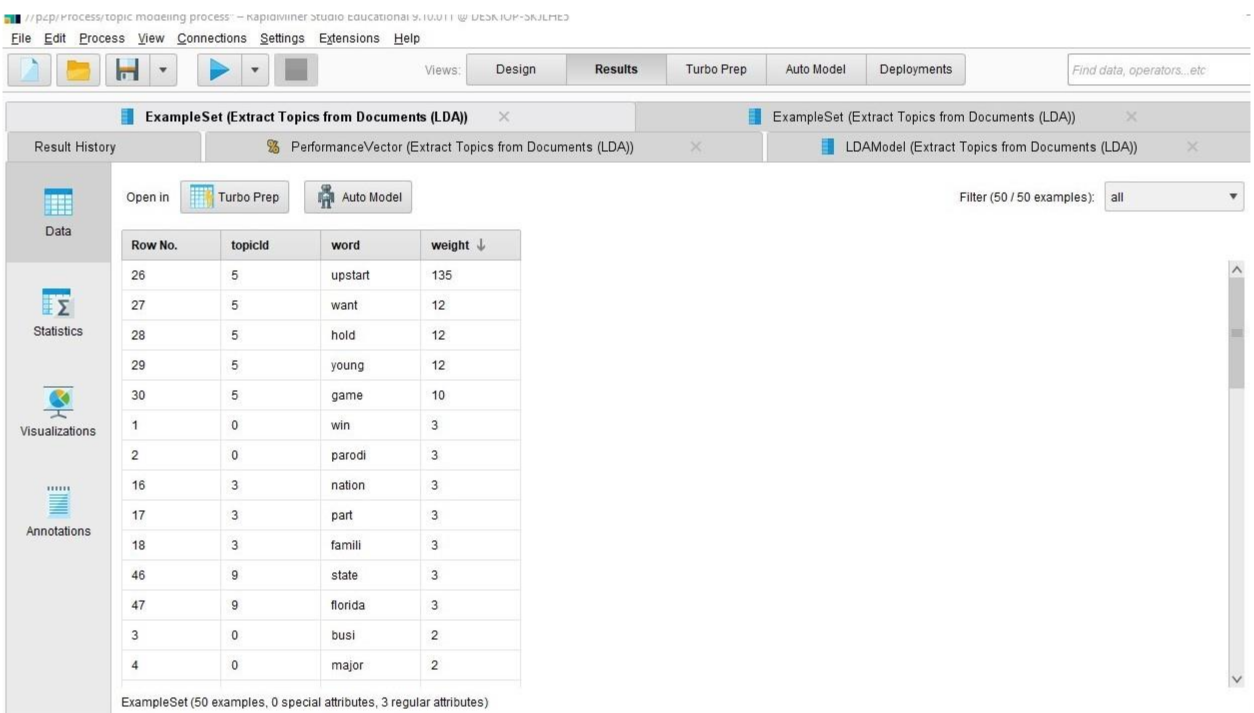


Row No.	topicId	word	weight
41	8	lend	45
36	7	lend	42
6	1	peer	40
16	3	lend	27
7	1	startup	25
8	1	lend	25
42	8	busi	25
37	7	defi	23
9	1	financi	22
10	1	onlin	22
43	8	platform	21
44	8	loan	20
45	8	financ	18
17	3	financ	16

Upstart

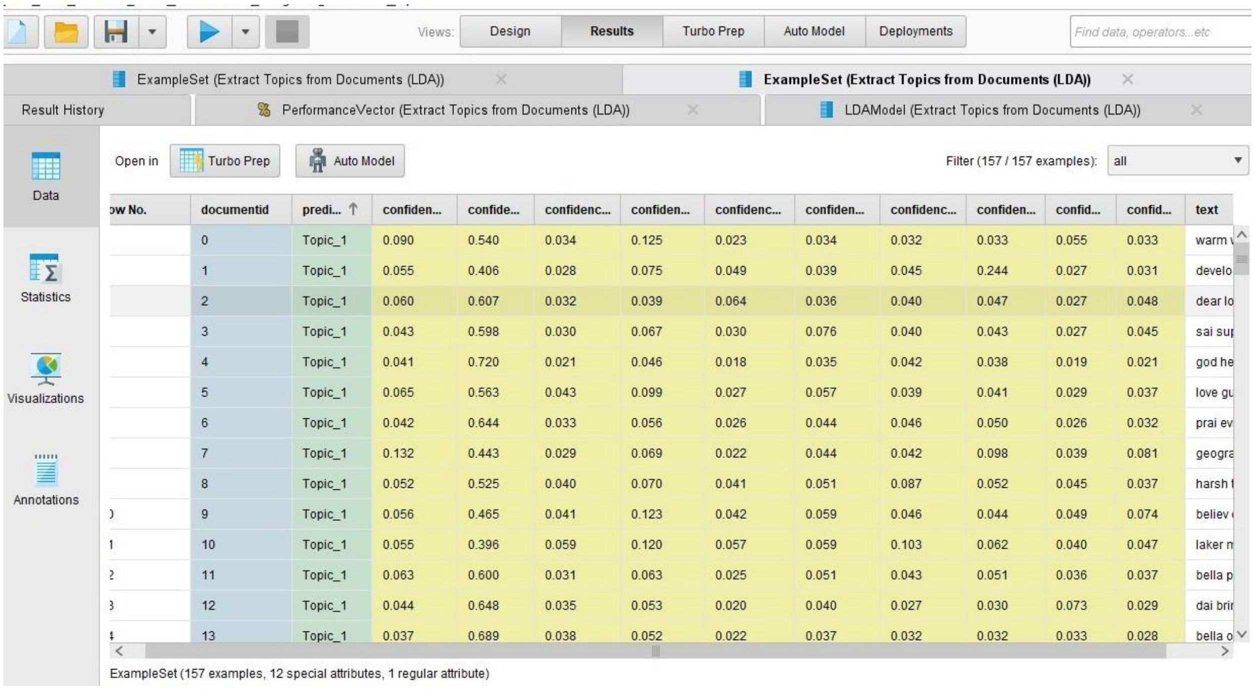


Row No.	documentid	predictio...	confid...	confide...	confiden...	confid...	confide...	confiden...	confidenc...	confiden...	confiden...	confiden...	text
1	0	Topic_5	0.069	0.055	0.056	0.060	0.044	0.537	0.047	0.015	0.036	0.079	spenc
2	1	Topic_5	0.081	0.065	0.050	0.068	0.047	0.510	0.050	0.014	0.044	0.069	upstart
3	2	Topic_5	0.065	0.086	0.054	0.084	0.043	0.445	0.067	0.015	0.044	0.098	felt bub
4	3	Topic_5	0.062	0.066	0.046	0.062	0.040	0.546	0.054	0.016	0.038	0.069	honor s
5	4	Topic_5	0.062	0.068	0.068	0.076	0.040	0.446	0.077	0.016	0.052	0.095	moon s
6	5	Topic_5	0.071	0.066	0.047	0.064	0.037	0.526	0.049	0.016	0.051	0.073	disnei
7	6	Topic_5	0.064	0.074	0.053	0.086	0.052	0.459	0.051	0.016	0.064	0.082	presid
8	7	Topic_5	0.068	0.085	0.056	0.059	0.038	0.491	0.056	0.013	0.055	0.080	hoi pre
9	8	Topic_5	0.072	0.063	0.052	0.060	0.035	0.541	0.049	0.014	0.047	0.066	disnei
10	9	Topic_5	0.050	0.060	0.046	0.064	0.037	0.570	0.056	0.013	0.036	0.068	gov hol
11	10	Topic_5	0.044	0.053	0.043	0.055	0.031	0.622	0.046	0.015	0.032	0.061	amc er
12	11	Topic_5	0.066	0.069	0.057	0.070	0.045	0.494	0.063	0.016	0.041	0.080	obes la
13	12	Topic_5	0.060	0.063	0.055	0.066	0.042	0.509	0.068	0.015	0.046	0.075	okai co
14	13	Topic_5	0.059	0.071	0.048	0.063	0.040	0.535	0.052	0.017	0.041	0.073	upstart

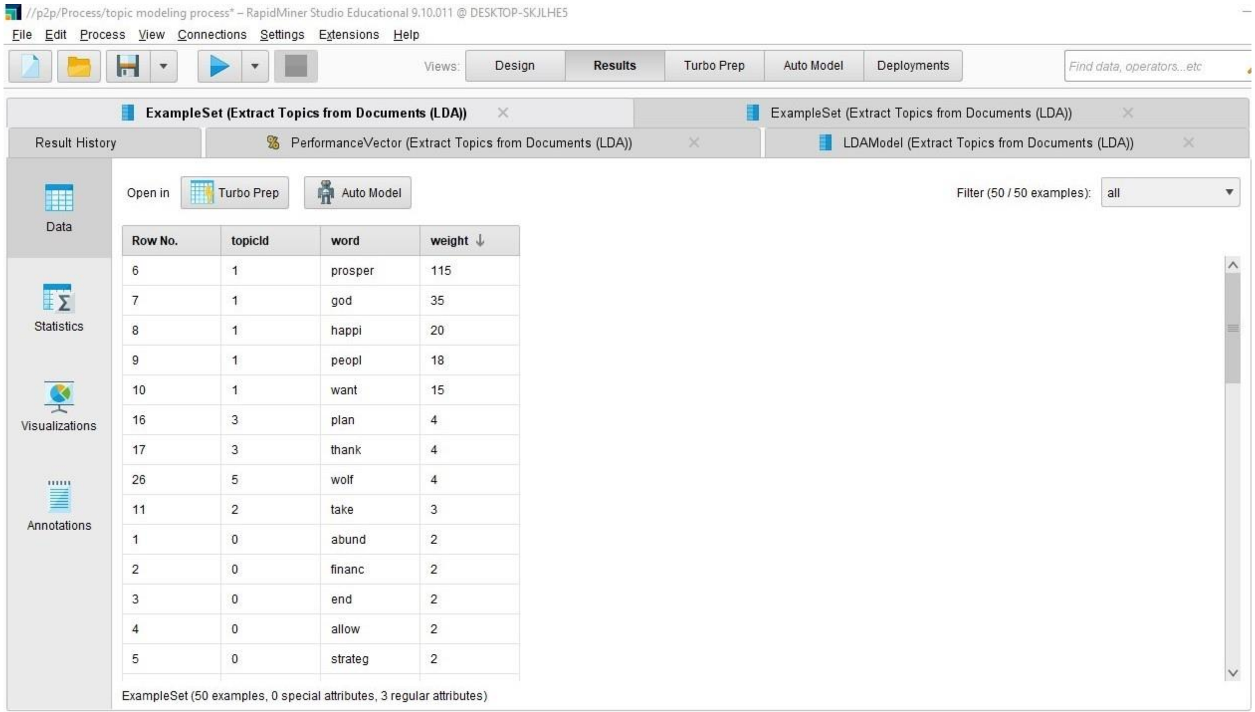


Row No.	topicId	word	weight
26	5	upstart	135
27	5	want	12
28	5	hold	12
29	5	young	12
30	5	game	10
1	0	win	3
2	0	parodi	3
16	3	nation	3
17	3	part	3
18	3	famili	3
46	9	state	3
47	9	florida	3
3	0	busi	2
4	0	major	2

Prosper

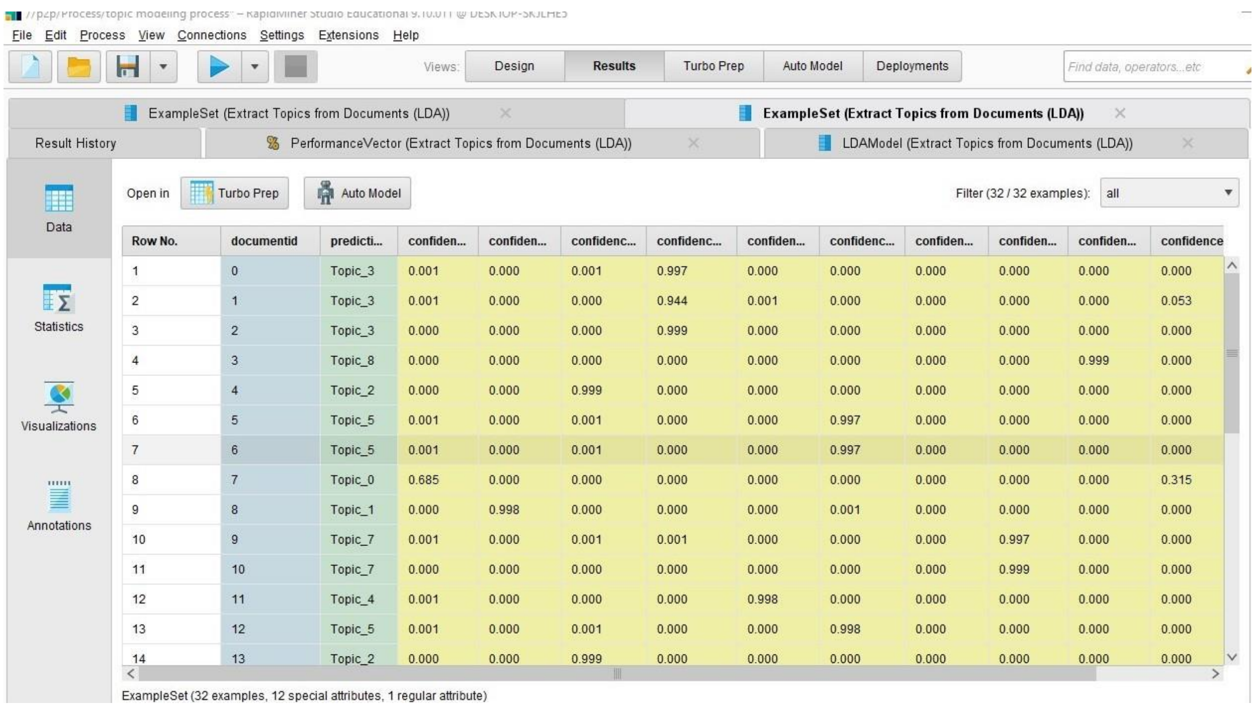


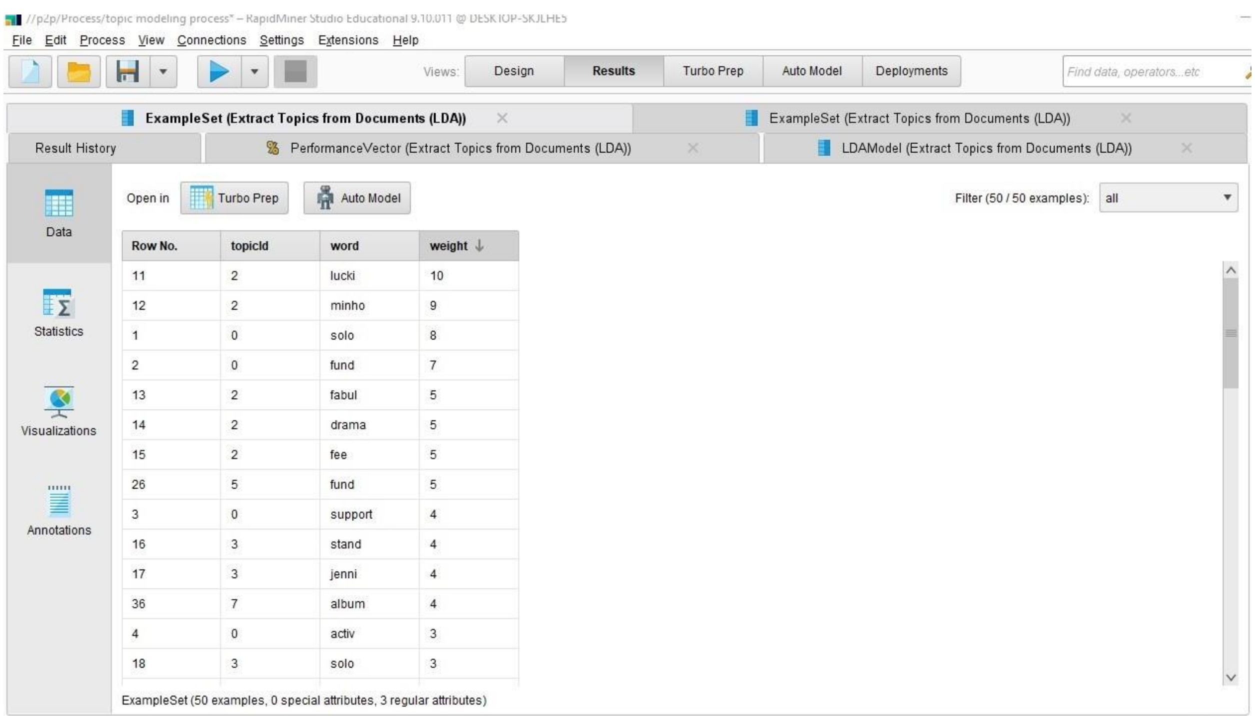
ow No.	documentid	predi...	confiden...	confide...	confidenc...	confiden...	confidenc...	confiden...	confidenc...	confiden...	confid...	confid...	text
	0	Topic_1	0.090	0.540	0.034	0.125	0.023	0.034	0.032	0.033	0.055	0.033	warm v
	1	Topic_1	0.055	0.406	0.028	0.075	0.049	0.039	0.045	0.244	0.027	0.031	develo
	2	Topic_1	0.060	0.607	0.032	0.039	0.064	0.036	0.040	0.047	0.027	0.048	dear lo
	3	Topic_1	0.043	0.598	0.030	0.067	0.030	0.076	0.040	0.043	0.027	0.045	sai sup
	4	Topic_1	0.041	0.720	0.021	0.046	0.018	0.035	0.042	0.038	0.019	0.021	god he
	5	Topic_1	0.065	0.563	0.043	0.099	0.027	0.057	0.039	0.041	0.029	0.037	love gu
	6	Topic_1	0.042	0.644	0.033	0.056	0.026	0.044	0.046	0.050	0.026	0.032	prai ev
	7	Topic_1	0.132	0.443	0.029	0.069	0.022	0.044	0.042	0.098	0.039	0.081	geogra
	8	Topic_1	0.052	0.525	0.040	0.070	0.041	0.051	0.087	0.052	0.045	0.037	harsh t
)	9	Topic_1	0.056	0.465	0.041	0.123	0.042	0.059	0.046	0.044	0.049	0.074	believ
1	10	Topic_1	0.055	0.396	0.059	0.120	0.057	0.059	0.103	0.062	0.040	0.047	laker n
2	11	Topic_1	0.063	0.600	0.031	0.063	0.025	0.051	0.043	0.051	0.036	0.037	bella p
3	12	Topic_1	0.044	0.648	0.035	0.053	0.020	0.040	0.027	0.030	0.073	0.029	dai brir
4	13	Topic_1	0.037	0.689	0.038	0.052	0.022	0.037	0.032	0.032	0.033	0.028	bella o



Row No.	topicId	word	weight
6	1	prosper	115
7	1	god	35
8	1	happi	20
9	1	peopl	18
10	1	want	15
16	3	plan	4
17	3	thank	4
26	5	wolf	4
11	2	take	3
1	0	abund	2
2	0	financ	2
3	0	end	2
4	0	allow	2
5	0	strateg	2

Solo Fund


File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
ExampleSet (Extract Topics from Documents (LDA))
ExampleSet (Extract Topics from Documents (LDA))
Result History
PerformanceVector (Extract Topics from Documents (LDA))
LDAModel (Extract Topics from Documents (LDA))
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (32 / 32 examples): all
Row No. documentid predicti... confiden... confiden... confidenc... confidenc... confiden... confidenc... confiden... confiden... confiden... confidence
1 0 Topic_3 0.001 0.000 0.001 0.997 0.000 0.000 0.000 0.000 0.000 0.000
2 1 Topic_3 0.001 0.000 0.000 0.944 0.001 0.000 0.000 0.000 0.000 0.053
3 2 Topic_3 0.000 0.000 0.000 0.999 0.000 0.000 0.000 0.000 0.000 0.000
4 3 Topic_8 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.999 0.000
5 4 Topic_2 0.000 0.000 0.999 0.000 0.000 0.000 0.000 0.000 0.000 0.000
6 5 Topic_5 0.001 0.000 0.001 0.000 0.000 0.997 0.000 0.000 0.000 0.000
7 6 Topic_5 0.001 0.000 0.001 0.000 0.000 0.997 0.000 0.000 0.000 0.000
8 7 Topic_0 0.685 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.315
9 8 Topic_1 0.000 0.998 0.000 0.000 0.000 0.001 0.000 0.000 0.000 0.000
10 9 Topic_7 0.001 0.000 0.001 0.001 0.000 0.000 0.000 0.997 0.000 0.000
11 10 Topic_7 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.999 0.000 0.000
12 11 Topic_4 0.001 0.000 0.000 0.000 0.998 0.000 0.000 0.000 0.000 0.000
13 12 Topic_5 0.001 0.000 0.001 0.000 0.000 0.998 0.000 0.000 0.000 0.000
14 13 Topic_2 0.000 0.000 0.999 0.000 0.000 0.000 0.000 0.000 0.000 0.000
ExampleSet (32 examples, 12 special attributes, 1 regular attribute)

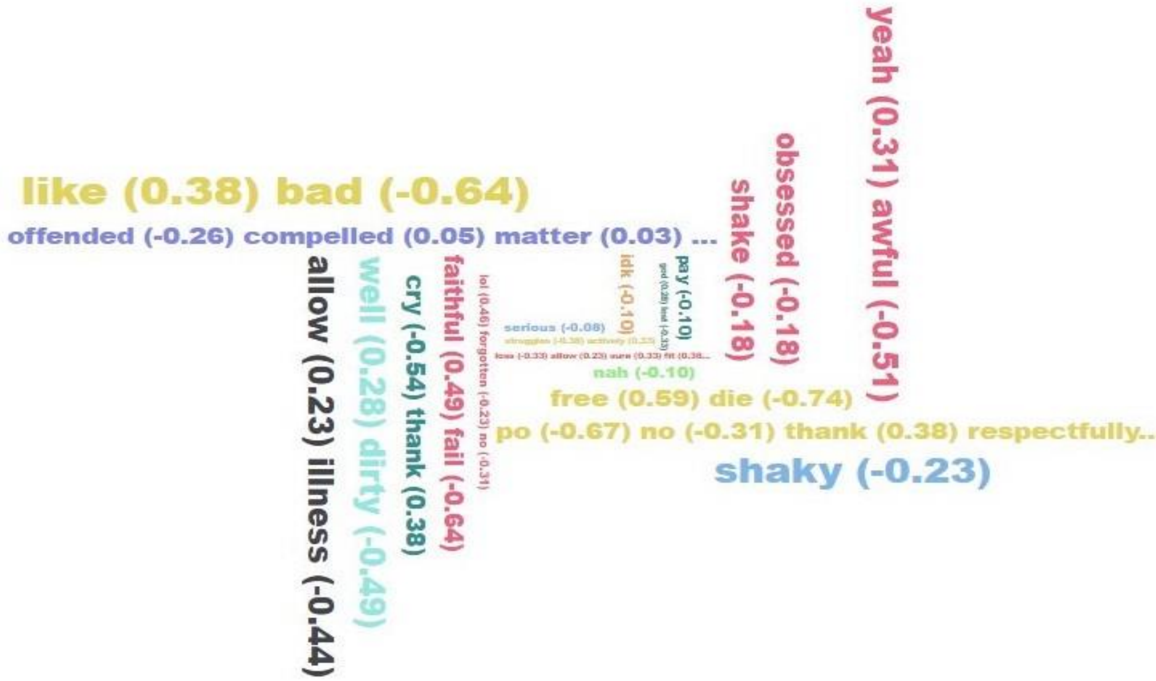

File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
ExampleSet (Extract Topics from Documents (LDA))
ExampleSet (Extract Topics from Documents (LDA))
Result History
PerformanceVector (Extract Topics from Documents (LDA))
LDAModel (Extract Topics from Documents (LDA))
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (50 / 50 examples): all
Row No. topicId word weight
11 2 lucki 10
12 2 minho 9
1 0 solo 8
2 0 fund 7
13 2 fabul 5
14 2 drama 5
15 2 fee 5
26 5 fund 5
3 0 support 4
16 3 stand 4
17 3 jenni 4
36 7 album 4
4 0 activ 3
18 3 solo 3
ExampleSet (50 examples, 0 special attributes, 3 regular attributes)

Most using word in Positive and Negative Statements:

Kiva

Positive

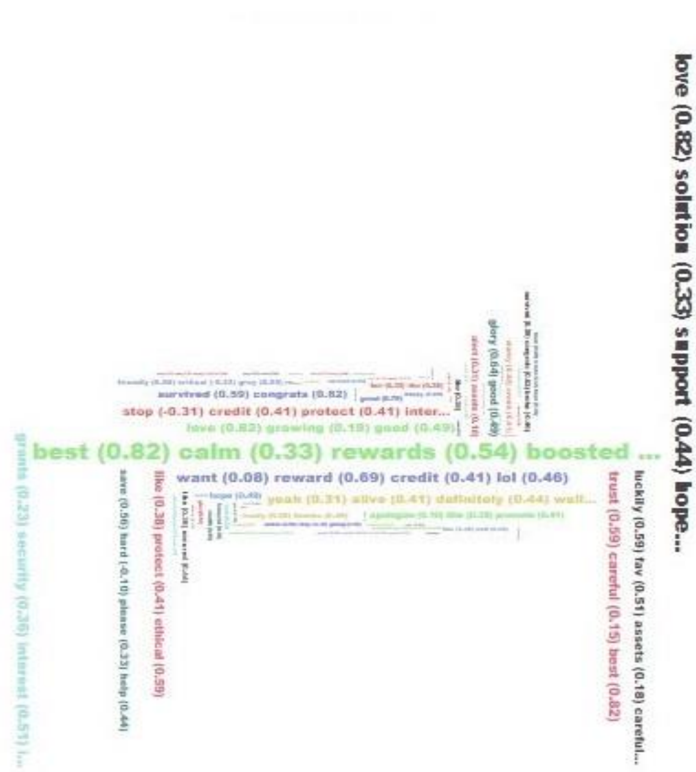
Negative



BlockFi

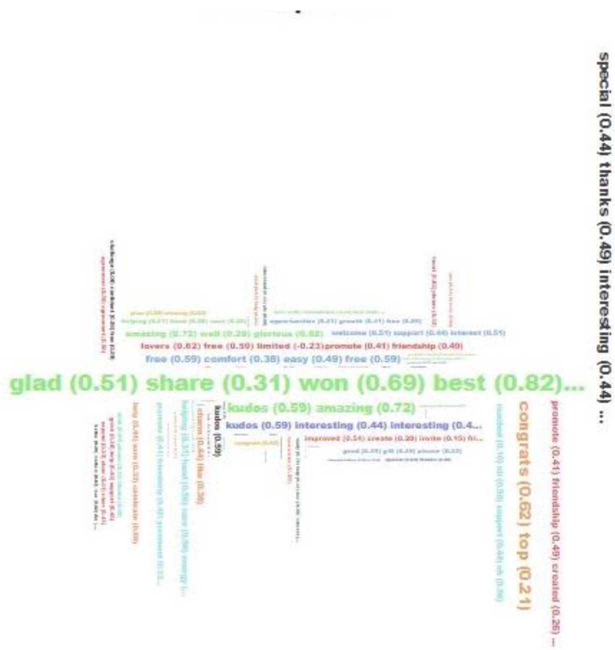
Negative

Positive

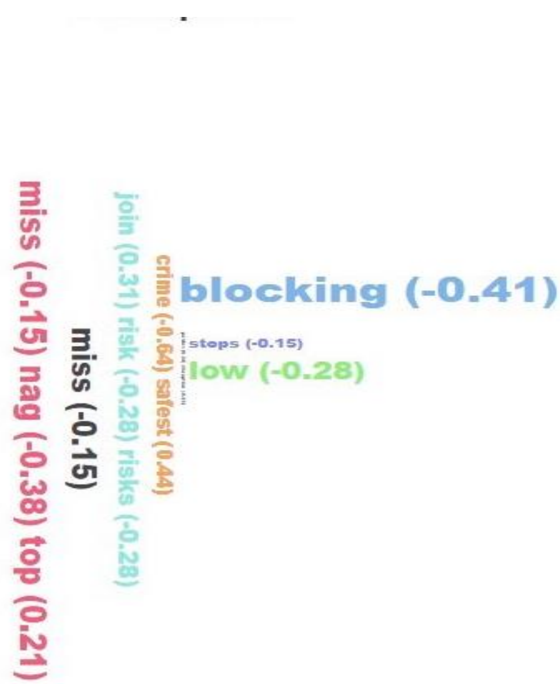


Online Lending

Positive

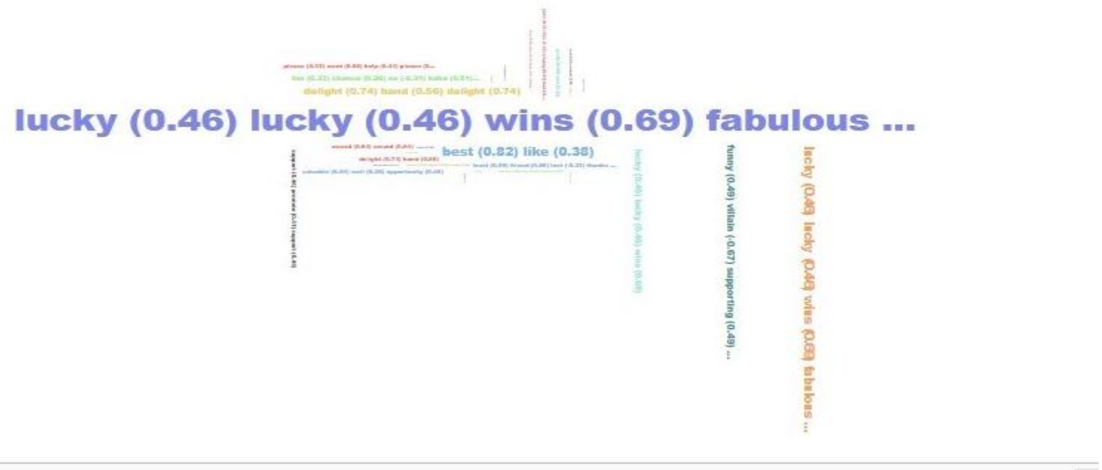


Negative



SOLO Fund

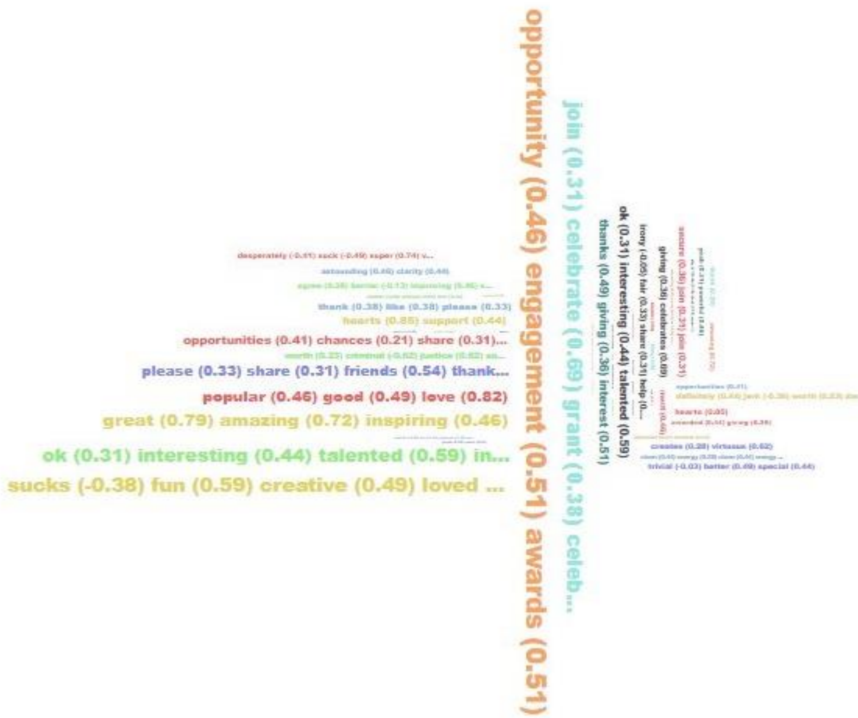
Positive



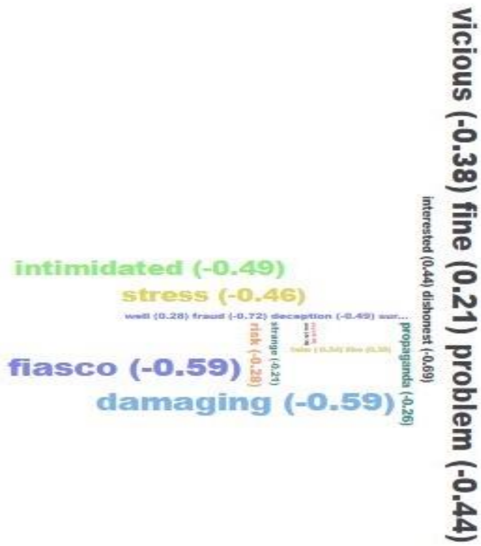
Negative

Funding Circle

Positive

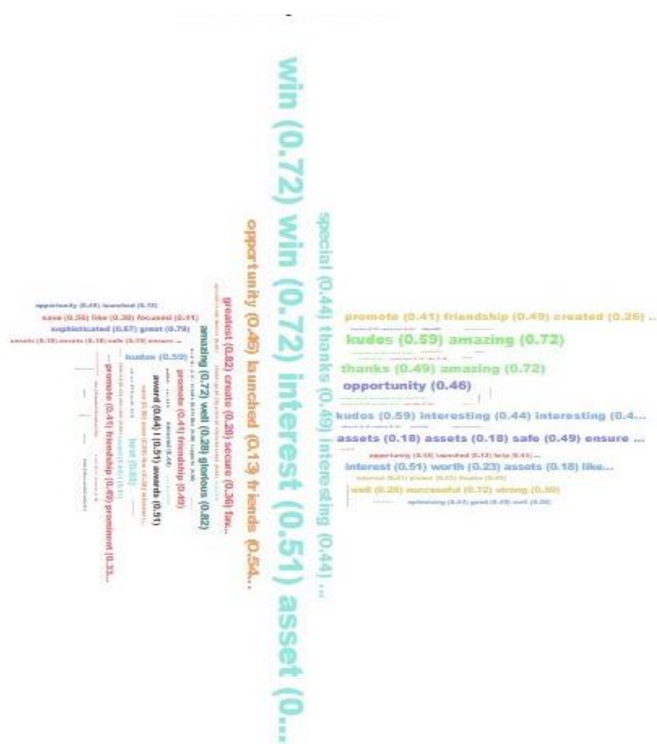


Negative



Peer-to-Peer

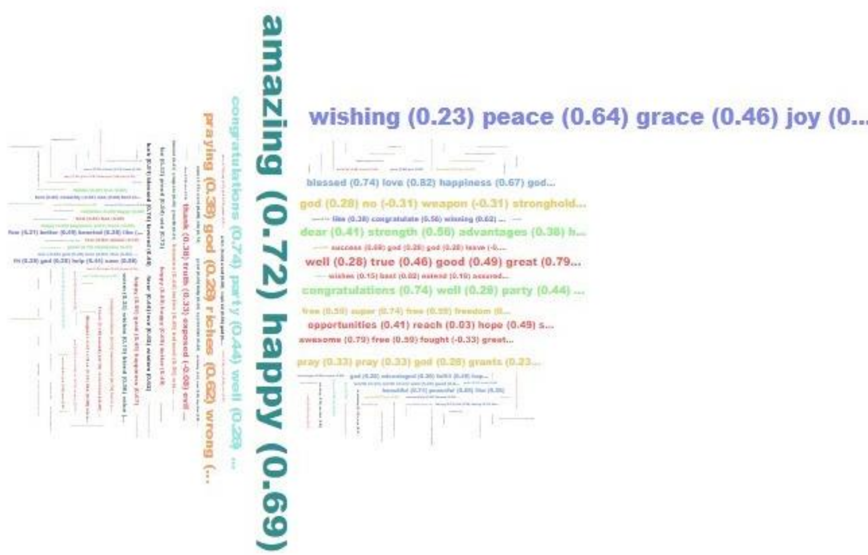
Positive



Negative

Prosper

Positive



Negative

Upstart

Positive

Negative


bother (-0.36) worth (0.23)
want (0.08) miss (-0.15)
strike (-0.13)
waste (-0.46) praying (0.38)
avoids (-0.18)
wish (0.44) luck (0.51) upset (-0.41) threateni...
straight (0.23) no (-0.31)
good (0.48) bad (-0.64) good (0.48) embarrassin...
challenged (-0.10)
hard (-0.10)
top (0.21) shit (-0.67) shitted (-0.44) agree (...