

## Teaching Case

# Using Topic Modeling to Discover Reasons for Employee Attrition

Frank Lee  
flee@gsu.edu

Alexander Algarra  
aalgarra@marketsource.com

J. Mack Robinson College of Business  
Georgia State University  
Atlanta, Georgia 30303, USA

## Abstract

This teaching case explores employee attrition, its impact on businesses, and how data analysis can be a weapon against it. It highlights the importance of pinpointing why employees leave and building strategies to keep them. By understanding attrition, businesses can make plans and stay competitive. The case also dives into Latent Dirichlet Allocation (LDA), a Natural Language Processing (NLP) technique to uncover hidden themes in text data. It explains how LDA can be used to identify the root causes of employee attrition. Additionally, the case explores RapidMiner, a data science platform that allows users to handle data prep, machine learning, and create predictive models to forecast employee churn.

**Keywords:** Attrition, Turnover, Latent Dirichlet Allocation, Human Resources, Retention, RapidMiner

## 1. INTRODUCTION

### Employee Attrition

Understanding why employees leave empowers organizations to retain their top talent. By addressing employee concerns and improving the work environment, organizations can keep their best people, giving them a competitive advantage. Employee turnover data (attrition) is a valuable tool for future planning. It helps identify potential workforce gaps and allows for targeted development programs. This way, organizations can groom successors for critical roles, ensuring a smooth transition and continued success (Demirkaya, et al., 2022).

Understanding the root causes of employee attrition equips organizations with a powerful weapon: targeted retention strategies. This reduces the financial burden of replacing lost talent and maintains a competitive edge by

keeping your best people on board. Researchers can leverage this knowledge to identify industry-specific attrition trends. This allows organizations to benchmark their own situation, comparing their attrition rates and practices to industry standards and best practices. This valuable comparison helps identify areas for improvement and implement the most successful retention strategies within their specific industry.

While attrition's impact on businesses is clear, its reach extends far beyond the bottom line. It can lead to broader social issues like higher unemployment and job insecurity. By studying attrition's contributing factors, researchers can offer valuable insights to policymakers, informing them on policies that promote job creation and retention. But it is not just jobs on the line. Attrition can also affect employee well-being, affecting both physical and mental health. Researchers can help organizations develop

healthier and more supportive work environments by understanding the reasons behind employee departures, ultimately benefiting everyone involved.

### Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) is a probabilistic generative model used in Natural Language Processing (NLP) for topic modeling (Blei, et al., 2003). LDA aims to discover the hidden thematic structure of text data and to represent this structure in an interpretable form. This method allows people to identify the underlying topics in documents. Now, to understand how LDA works, it is helpful to break down the model into its parts:

- Word: An atomic unit of meaning.
- Document: A collection of words.
- Corpus: A collection of documents
- Topic: A set of words likely to co-occur in documents.

The basic idea behind Latent Dirichlet Allocation is that each document in the constructed corpus can be a mixture of topics (Blei, et al., 2003). A topic is a distribution over sets of words, and a document is an amalgam of topics, each contributing a particular weight. Each topic is a probabilistic distribution over a set of words. Strictly speaking, the LDA algorithm assumes each document is a probabilistic mixture of hidden topics. The LDA model assumes that each document is created as follows:

1. Choosing the number of words in the document.
2. Choosing a distribution over topics for the document from the Dirichlet distribution.
3. And, for each word in the document:
  - Choosing a topic from the topic distribution.
  - Choosing a word from the topic distribution.

### RapidMiner

RapidMiner is a data science platform that allows users to perform a wide range of data preparation, machine learning, and predictive modeling tasks—providing a drag-and-drop interface that allows users to quickly build and perform complex analyses without extensive programming knowledge.

One of the benefits of RapidMiner is its ability to handle large datasets, as it allows the user to process data in memory or on disk. The platform also provides visualization and reporting tools to help users understand the data and thus communicate their findings. Businesses and organizations in various industries, including

finance, healthcare, and retail, use RapidMiner. It is also popular among data scientists and researchers who must perform complex analyses on large datasets.

### Learning Objectives

By completing this assignment, you will be able to:

- Understand the methodology used to collect and analyze text data.
- Understand the concept of Latent Dirichlet allocation (LDA).
- Analyze the attrition of the organization's employees through data analytics.
- Identify the key topics and themes that are driving this trend.
- Develop critical thinking skills and interpret the findings to recommend employee retention and engagement solutions.

## 2. CASE BACKGROUND

### Case Text

Even before working in human resources, Maggie has always been deeply passionate about data. She knew that data could provide valuable insights into human behavior and could be used to improve decision-making across a variety of fields. She has a degree in statistics and has completed several online courses in data science and machine learning. Though Maggie was passionate about data from the start, her admiration for analytics grew tenfold when she began working in HR. After several years in various roles, Maggie realized the power of data analytics and how it could be applied to ensure that the employees within her company were more satisfied with their own work. Maggie combined her love of data and desire to support her peers in the workplace to solve some of the industry's most pressing challenges, such as employee turnover and engagement.

Maggie stays current on industry blogs, attends conferences, and networks with other passionate HR professionals who share her interest in data-driven decision-making. So, when her boss, the Chief Human Resources Officer, tasked her with investigating the company's high attrition rates, Maggie was thrilled to apply her data background to the employees she worked alongside. She knew this was a complex problem requiring a deep understanding of the underlying factors driving employee turnover.

Maggie began by gathering all the data on employee turnover, including exit interviews, employee satisfaction surveys, and performance

data. She also took a comprehensive approach by contacting other departments across the company to understand better the broader context in which the attrition occurred. After all, how did attrition and turnover impact the company as a whole?

As she began to dive into the data, Maggie noticed several patterns emerging. She saw a higher turnover rate among employees who had been with the company for less than a year. Many of these employees cited a lack of career development opportunities as a reason for leaving. She also saw a strong correlation between employee engagement and turnover, with disengaged employees more likely to leave.

Armed with these insights, Maggie began to develop a plan to address the underlying issues driving the attrition problem that her peers faced. She proposed a series of initiatives to improve career development opportunities, enhance employee engagement, and strengthen the company's culture and values. For example, Maggie recommended that the company introduce a formal mentorship program to help new employees navigate their roles and develop their careers. Additionally, the program would foster engagement for new employees.

She also suggested that the company invest in recurring training programs and skill-building workshops to help employees acquire new skills and advance in their careers based on their interests. Additionally, she recommended that the company improve its communication and feedback processes to help employees feel more engaged and connected to the company's goals and values.

With the support of her boss and other senior leaders, Maggie implemented these initiatives, closely monitoring their impact on employee turnover rates. Over time, she saw a significant reduction in attrition, with employees reporting higher levels of engagement, more significant career development opportunities, and a stronger sense of connection to the company's mission and values.

Maggie felt immensely proud of her work and its impact on the company and its employees, especially because it included individuals she worked to support every day. She realized that her passion for data analytics had not only helped her to solve a critical business problem but had also enabled her to make a meaningful difference in people's lives. Maggie continued to explore new

ways to apply data analytics to HR, and her work became a model for other companies seeking to improve employee retention and engagement.

### The Data Source

The data set contains 1,155 responses from employees who have quit the organization labeled "Quit\_Reason" and their respective "Employee\_ID," "Hire\_Date," and "Quit\_Date." The responses were collected through an exit survey conducted by the organization. This data set can be used for various analyses, such as identifying the most common reasons for employee attrition, exploring patterns in the reasons for quitting, and predicting which employees are likely to leave the organization based on their tenure. The attributes within the data set are as follows:

- Employee\_ID: A unique identifier of each employee.
- Quit\_Reason: A text field where employees could explain why they quit the organization.
- Hire\_Date: The date the organization hired the employee.
- Quit\_Date: The date the employee quit the organization.

The research problem of the organization's attrition is to investigate why employees leave their jobs in record numbers. This phenomenon has been observed and has caused organizational concerns regarding employee retention and engagement, particularly in the Human Resources field. The research gap in this area is the lack of understanding of the critical topics and themes driving the trend, making it difficult for organizations to develop effective retention strategies.

The importance of this research problem is twofold. First, it provides a better understanding of the factors driving the attrition trend, which can help organizations develop more effective employee retention strategies. Second, it contributes to the body of knowledge in the Human Resources field, particularly in data analytics, and can be used as a reference for future research.

### 3. PROJECT ACTIVITY

This exercise aims to identify the key reasons behind the resignation trend by analyzing the content of 1,155 employee responses when asked why they have decided to leave the company. Respectively, the exercise aims to:

- Identify the key topics and themes related to the resignations within the company.
- Develop a better understanding of the factors driving the trend, such as job satisfaction, work-life balance, compensation, and employee benefits.
- Provide recommendations to organizations on improving employee retention and engagement based on the study's findings.

The following steps will help guide you along during the data analysis. Please make sure to follow each of the steps in order.

1. Open RapidMiner Studio on your computer. The RapidMiner Studio window will appear. A new window will ask you to select a template for your new process. Select the "Blank Process" option to open a blank process and click the "Create" button. (Appendix A, Figure 1).
2. Once the blank process is open, you will see tabs in the console, including the repository, process, operators, parameters, background monitor, and help. (Appendix A, Figure 2)
3. To import the "Employee\_Attrition" dataset into RapidMiner, click on the "File" option in the top menu bar. In the dropdown menu that appears, click on "Import Data." A new window will appear; select the "Import File" option in this window. In the "Select File" window that appears, navigate to the location where the "Employee\_Attrition" dataset is saved to your computer. Then, click the "Open" button, and in the "Select Table" window that appears, make sure the "First row contains attribute names" checkbox is selected. Finally, click on the "Finish" button. The "Employee\_Attrition" dataset should now appear in the Repository tab under the "Data" folder. We can drag it onto the process canvas to start working with it in the workflow. (Appendix A, Figure 3).
4. Before we get started building the LDA process, we must save the process. In the RapidMiner Studio window, ensure the process is open and visible on the process canvas. In the top menu bar, click on the "File" option; in the dropdown menu that appears, click "Save As.". A new window will appear, asking for the location and name of the saved process. Choose the location under the "Processes" tab and name the file "Attrition\_LDA." Once the process is saved, it is accessible moving forward by navigating to

the file's location and double-clicking. (Appendix A, Figure 4).

5. The process canvas is the large central area in the RapidMiner Studio window where the data analysis workflows are built. Click and drag the "Employee\_Attrition" dataset from the Repository panel under the data folder onto the process canvas. The dataset will now appear as a blue box on the canvas, as a "Retrieve" operator with its name displayed at the top. It is possible to connect the "Employee\_Attrition" dataset to operators on the canvas to build your data analysis workflow (Appendix A, Figure 5).
6. Connect the "Employee\_Attrition" dataset to the results port of the process canvas on the right-hand side. A blue circle on the left side of the operator box will represent the input port. To do this, click on the blue circle on the right side of the dataset box and drag the connection arrow to the results port (Appendix A, Figure 6).
7. We can examine the dataset attributes once the dataset is connected to the results port. To do this, click on the "Run Process" button in the console's upper left corner, displaying a data preview with columns representing the dataset attributes. We can see the dataset contains the text attribute "Quit\_Reason," which are the responses given by employees when asked why they have decided to leave the company (Appendix A, Figure 7).
8. The Select Attributes operator is used to specify which attributes to include or exclude from the dataset and can help simplify the dataset and remove unnecessary or redundant columns. To specify the Select Attributes operator to select the "Quit\_Reason" attribute of the "Employee\_Attrition" dataset, follow these steps: (Appendix A, Figure 8)
  - Search for the operator and drag a Select Attributes operator onto the process canvas from the "Operators" panel on the left side of the screen.
  - Connect the "Employee\_Attrition" dataset to the input port of the Select Attributes operator. To do this, click on the blue circle on the right side of the dataset box and drag the connection arrow to the input port of the Select Attributes operator.
  - Double-click on the Select Attributes operator to open its configuration window.

- In the configuration window, a list of all the attributes in the dataset appears. To select the "Quit\_Reason" attribute, click the checkbox next to its name in the list.
  - Once the attribute has been selected, click "Apply" to save your changes and close the configuration window.
  - The output of the Select Attributes operator will be a new dataset containing only the selected attributes.
9. The Nominal to Text operator converts nominal values (categorical data) in a dataset into text values and can help prepare data for certain types of analysis or visualization. To add the Nominal to Text operator into the process, follow these steps: (Appendix A, Figure 9).
- Search for the operator and drag the Nominal to Text operator onto the process canvas from the "Operators" panel.
  - Connect the output of the Select Attributes operator to the input port of the Nominal to Text operator.
  - Double-click on the Nominal to Text operator to open its configuration window.
  - The configuration window shows a list of all the attributes in the input dataset. Please select the "Quit\_Reason" attribute we want to convert by checking its checkbox.
  - Next, select the "Conversion" tab in the configuration window. Here, it is possible to specify how to convert the nominal values to text. For example, replace each value with a specific text string or a corresponding numerical code.
  - Once the conversion options have been specified, click "Apply" to save the changes and close the configuration window.
  - The output of the Nominal to Text operator will be a new dataset with the selected attribute converted from nominal to text.
10. The "Multiply" operator takes the example set from the input port and delivers copies to the output ports. Each connected port creates an independent copy. So, changing one copy does not affect other copies. For example, suppose there is an input example set with 1,000 examples. In that case, use the "Multiply" operator with a factor of 3 to create a new set of 3,000 examples. This operation can generate data for testing or training purposes. To add the Multiply operator in the process, follow these steps: (Appendix A, Figure 10)
- Search for the operator and drag and drop the "Multiply" operator from the "Operators" panel onto the canvas.
  - Connect the output of the "Nominal to Text" operator to the input of the "Multiply" operator. To do this, drag the output port of the "Nominal to Text" operator to the input port of the "Multiply" operator.
11. The "Process Documents from Data" operator in RapidMiner is used to preprocess text data and extract meaningful information from it. This operator takes a text input file or example set as input and generates word vectors from string attributes as output. To add the Process Documents from Data operator into the process, follow these steps: (Appendix A, Figure 11)
- Search for the operator and drag and drop the "Process Documents from Data" operator from the "Operators" panel onto the canvas.
  - Connect the output of the "Multiply" operator to the input of the "Process Documents from Data" operator. To do this, click and drag the output port of the "Multiply" operator to the input port of the "Process Documents from Data" operator.
  - Double-click the "Process Documents from Data" operator to open its parameters dialog. It is possible to configure the settings for text preprocessing, feature extraction, and other options.
12. The Process Documents from Data operator contains a subprocess and a subprocess is a smaller, self-contained process that can be used as a building block within a larger process. Subprocesses can be thought of as subroutines or functions in programming. We will be adding the "Tokenize", "Transform Cases", "Filter Stopwords", "Filter Tokens", and "Stem (Porter)" operators to the "Process Documents from Data" subprocess. To begin building the subprocess within the operator, follow these steps: (Appendix A, Figure 12)
- Double-click the "Process Documents from Data" operator to open its subprocess.
  - In the "Search for Operators" section, click and drag the "Tokenize", "Transform Cases", "Filter Stopwords", "Filter Tokens", and "Stem (Porter)" operators and connect them to results tab.
13. Furthermore, we must specify the parameters for each operator. To do this click on each operator and navigate to the parameters

section. The following operator's parameters are listed from left to right of the subprocess.

- The "Tokenize" operator splits a text string into individual words or tokens, takes a text attribute as input, and generates a new attribute containing the tokens as separate examples. (Appendix A, Figure 13).
  - The "Transform Cases" operator in RapidMiner converts the case of a text attribute. It takes a text attribute as input and generates a new attribute that contains the text with the case transformed according to the selected mode, either uppercase or lowercase. (Appendix A, Figure 14).
  - Stopwords are words that are commonly used in a language but do not carry much meaning or information, such as "the," "and," "a," and so on. Removing stopwords from a text attribute can reduce noise in the data and improve the accuracy of subsequent analyses. (Appendix A, Figure 15).
  - The "Filter Tokens (by Length)" operator in RapidMiner removes tokens outside a specified range of character lengths. This can be useful for removing very short or very long tokens that may not be relevant to the analysis. (Appendix A, Figure 16).
  - The "Stem (Porter)" operator in RapidMiner reduces words to their base or root form, also known as stemming, and is essential for text mining because it can help to normalize the text and reduce the dimensionality of the data. (Appendix A, Figure 17).
14. Now, we need to navigate back to the main process window. The "Wordlist to Data" operator converts a list of words into a data set and can be helpful when there is a pre-defined list of words you want to filter before reporting it. To add the Wordlist to Data operator into the process, follow these steps: (Appendix A, Figure 18).
- Locate the "Wordlist to Data" operator in the "Text Processing" category of the RapidMiner Operators panel.
  - Drag the "Wordlist to Data" operator onto the Process canvas.
  - Connect the output of the "Process Documents from Data" operator to the "Wordlist to Data" operator by clicking and dragging the output arrow of the former to the input port of the latter.
15. The "Set Role" operator defines attributes' roles in a data set and allows you to specify whether each attribute is a regular, label, id, or weight attribute. Here are the steps to add the "Set Role" operator to your process and connect it to the output of the "Multiply" operator: (Appendix A, Figure 19).
- Locate the "Set Role" operator in the "Data Transformation" category of the Operators panel.
  - Drag the "Set Role" operator onto the Process canvas.
  - Connect the output of the "Multiply" operator to the "Set Role" operator by clicking and dragging the output arrow of the former to the input port of the latter.
  - In the "Set Role" operator, it is possible to specify the role of each attribute in the input data set by selecting the appropriate role from the drop-down menu next to each attribute.
  - Specify the role of "Quit\_Reason" to "label" and click apply.
16. The "Extract Topics from Data (LDA)" operator in RapidMiner finds topics using the LDA method. To add the "Extract Documents from Data (LDA)" operator into your RapidMiner process and connect it to the "Set Role" operator, you can follow these steps: (Appendix A, Figure 20).
- Drag the "Extract Topics from Data (LDA)" operator from the "Text Processing" folder in the "Operators" panel and drop it onto the process canvas.
  - Connect the output of the "Set Role" operator to the input of the "Extract Topics from Data (LDA)" operator.
17. Once you have completed your process, you can connect the outputs of the "Extract Topics from Data (LDA)" and the "Wordlist to Data" operators to the results output and run the process (Appendix A, Figure 21).
18. After the process has completed running, RapidMiner should automatically show the Results tab. There should be six individual results, including two example sets, performance vector, LDA prediction, WordList, and result history (Appendix A, Figure 22).
- Performance Vector (Extract Topics from Data (LDA)): Provides several metrics to evaluate the performance of the topic model

generated by the operator. These metrics include (Appendix A, Figure 25)

- LDA Prediction (Extract Topics from Data (LDA)): Provides several performance metrics and information about the topics and documents in the dataset. (Appendix A, Figure 26)
- WordList (Process Documents from Data): Generates a list of all the unique words present in the input documents along with their total occurrences across all the documents and document occurrences, i.e., the number of documents in which a particular word occurs. (Appendix A, Figure 27)

#### 4. LDA TOPICS

Topic 0	Events, Social, Listen, Action, Take, Activities, Hold, Satisfaction, Improve, Follow
Topic 1	Pay, Pension, Careerpromotions, Better, Incentives, Benefit, Communication, Higher, Perk, Raises
Topic 2	Insurance, Package, Compensation, Healthcare, Better, Personal, Less, Get, Rid, Care
Topic 3	Employees, Work, Good, Management, Keep, Works, Recognize, Could, Communication, Team
Topic 4	Employees, Treat, Respect, Offer, Listen, Benefits, Reward, Listen, Childcare, Support
Topic 5	People, Reimbursement, Hours, Education, Training, Flexible, Need, Career, Working, Job
Topic 6	Office, Really, Going, Organization, Part, Would, Outside, See, Feel, Previous
Topic 7	Company, Outings, Like, Would, Get, Employees, Sponsored, Benefits, Think, Offer
Topic 8	Time, Vacation, Rollover, Increase, Bonus, Raises, Standard, Industry, Nothing, Performance
Topic 9	Jobs, Raises, Performance, Employees, Management, Give, Better, Done, Good, Year

**Table 1.** LDA Topic Table

#### Topic Descriptions

##### *Topic 0 (Employee Engagement and Development)*

This topic seems to be centered around employee engagement and satisfaction, with words like "satisfaction," "improve," and "follow" suggesting a focus on improving employee well-being. The inclusion of words like "events," "social," and "activities" suggests that the topic also encompasses opportunities for social and professional development, which could help to improve employee engagement and job satisfaction. Finally, the use of words like "listen," "action," and "hold" suggests a focus on taking action to address employee concerns and hold events or activities that are tailored to their needs and interests.

##### *Topic 1 (Motivation and Retention)*

This topic seems to be centered around employee compensation and benefits, with words like "pay," "pension," "incentives," and "perks" suggesting a focus on providing attractive compensation packages to retain and motivate employees. The inclusion of words like "career promotions," "better," and "higher" suggests that the topic also encompasses opportunities for career advancement and growth. Finally, using words like "communication" suggests a focus on transparent and effective communication between employees and management regarding compensation and benefits.

##### *Topic 2 (Healthcare for Improved Well-being)*

This topic seems to be centered around employee benefits and healthcare, with words like "insurance," "compensation," "healthcare," and "personal," suggesting a focus on providing employees with access to healthcare services and comprehensive benefits packages. Words like "better" and "less" suggest that the topic also encompasses improving employee benefits' quality and affordability. Finally, using words like "get rid" and "care" suggests a focus on addressing employee concerns related to healthcare and well-being.

##### *Topic 3 (Effective Management and Team Communication)*

This topic seems to be centered around the importance of effective management and team communication in retaining employees, with words like "employees," "work," "management," and "team" suggesting a focus on teamwork and collaboration. The use of words like "keep," "works," and "recognize" suggests that the topic also encompasses efforts to retain and recognize good employees. Finally, using words like "could" and "communication" suggests a focus on

improving communication and feedback channels between employees and management for increased retention and job satisfaction.

#### *Topic 4 (Employee Well-being and Supportive Culture)*

This topic seems to be centered around creating a supportive work culture that prioritizes employee well-being, with words like "employees," "treat," "respect," "benefits," and "childcare," suggesting a focus on offering benefits and support to employees. Using words like "reward" and "listen" suggests that the topic also encompasses efforts to recognize and respond to employee needs and concerns. Finally, using words like "support" suggests a focus on creating a workplace environment that supports and empowers employees to succeed, which can lead to improved retention rates.

#### *Topic 5 (Career Development and Work-Life Balance)*

This topic seems to be centered around offering opportunities for career development and work-life balance to employees, with words like "education," "training," "career," "flexible," and "working," suggesting a focus on employee growth and development. The use of words like "reimbursement," "hours," and "need" suggests that the topic also encompasses efforts to provide employees with the resources they need to advance their careers, such as financial support and flexible scheduling. Finally, using words like "job" suggests a focus on retaining employees by offering fulfilling and rewarding work experiences that align with their career goals and priorities.

#### *Topic 6 (Organizational Culture and Employee Satisfaction)*

This topic seems to be focused on the overall culture and environment within the organization, with words like "office," "organization," and "feel" suggesting a focus on the physical workplace and the atmosphere within it. The use of words like "going," "see," and "outside" suggests a focus on how employees experience and interact with the organization beyond the physical workspace. Using words like "really" and "previous" suggests that this topic may also encompass employee perspectives on organizational changes or improvements over time. Finally, using words like "part" and "would" suggests that this topic may also encompass employee opinions on their role within the organization and how well it aligns with their interests and goals.

#### *Topic 7 (Benefits and Perks)*

This topic seems to be focused on the various benefits and perks offered to employees by the

company, with words like "company," "employees," and "sponsored" suggesting a focus on the employer-employee relationship. The use of words like "outings," "like," and "get" suggests a focus on recreational or social activities that the company may offer as a benefit or perk. Additionally, the use of words like "think" and "offer" suggests that this topic may also encompass employee opinions on the adequacy and effectiveness of the benefits and perks offered by the company.

#### *Topic 8 (Compensation)*

This topic seems to be focused on various forms of compensation and benefits offered to employees, with words like "time," "vacation," "roll over," "increase," "bonus," and "raises" suggesting a focus on financial compensation. Using words like "standard" and "industry" suggests that this topic may also encompass comparisons to industry standards or expectations for compensation and benefits. Additionally, using words like "performance" and "nothing" suggests that this topic may include employee opinions on how their compensation and benefits are tied to their performance and contribution to the company.

#### *Topic 9 (Performance and Career Growth)*

The topic contains words related to employee performance, career growth, and rewards for good performance, such as jobs, raises, performance, better, done, good, and year. The words suggest that employees seek opportunities to improve their performance and advance in their careers and expect to be rewarded for their efforts.

## **5. PROJECT REPORT**

In this project report, you will comprehensively analyze employee attrition and its impact on businesses. Ensure the report includes an executive summary, introduction, data collection, data preparation, methodology, conclusion, reference, and appendix. Focus on interesting, surprising, or important results in your discussion, interpret the results with detailed information, and explore the consequences or implications of your findings. In the results and discussion section, concentrate on the most interesting, surprising, or important results obtained from the analysis. Interpret the 10 topics found in the LDA modeling process and explain your reasoning behind the interpretation. Discuss the consequences or implications of these results and, if the findings are unexpected, explore possible explanations, such as other factors not included in the analysis.



Conclude by summarizing the project's main points and key takeaways, highlighting the value of data analytics in addressing HR challenges and improving employee retention and engagement. Cite all sources used in the references section and include any supplementary materials, such as tables, charts, or code, in the appendix. By following this structure and focusing on the most relevant results, your project report will thoroughly analyze employee attrition and offer valuable insights for improving employee retention and engagement.

## 6. REFERENCES

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- Demirkaya, H., Aslan, M., Güngör, H., Durmaz, V., & Rodoplu Şahin, D. (2022). Covid-19 and quitting jobs. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.916222>

## APPENDIX A Guided Step-by-Step Figures

Due to the file size limit, not all Figures in the Appendix are displayed here. You can download the full Appendix at <https://tinyurl.com/iscap2024>.

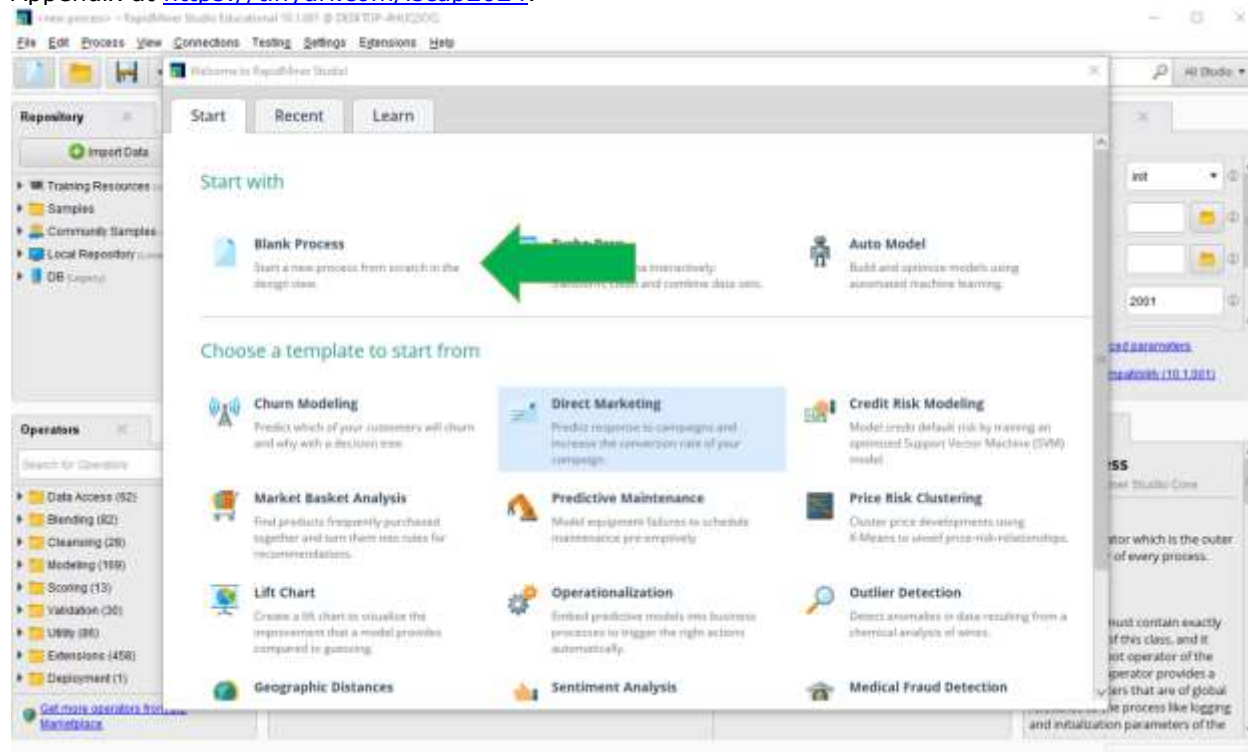


Figure 1. Open blank process

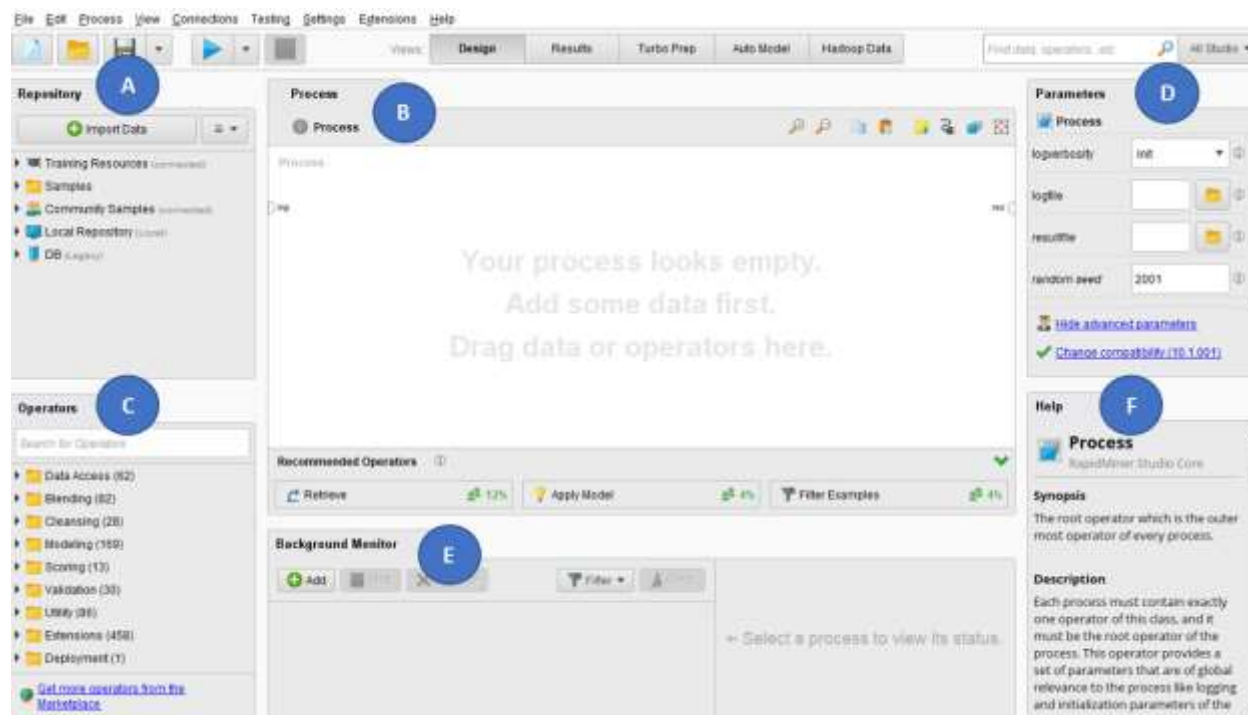


Figure 2. Introduction to the RapidMiner console.

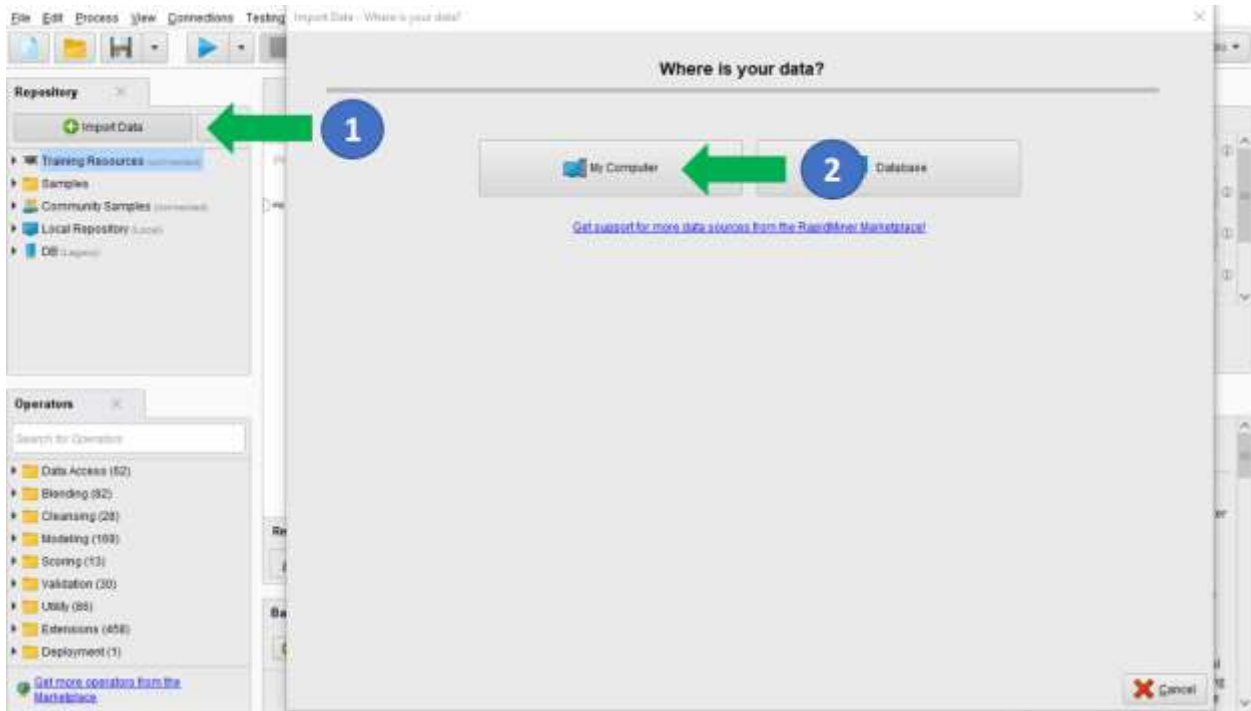


Figure 3. Import the "Employee\_Attrition" data set into RapidMiner.

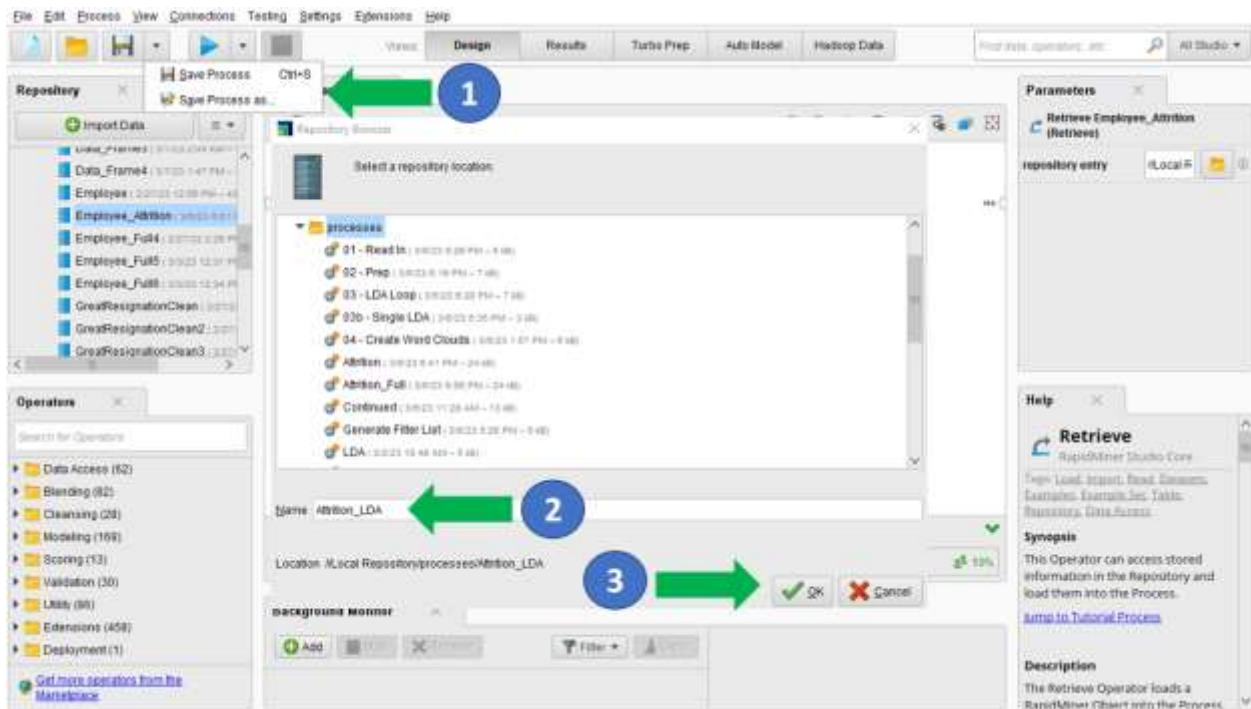
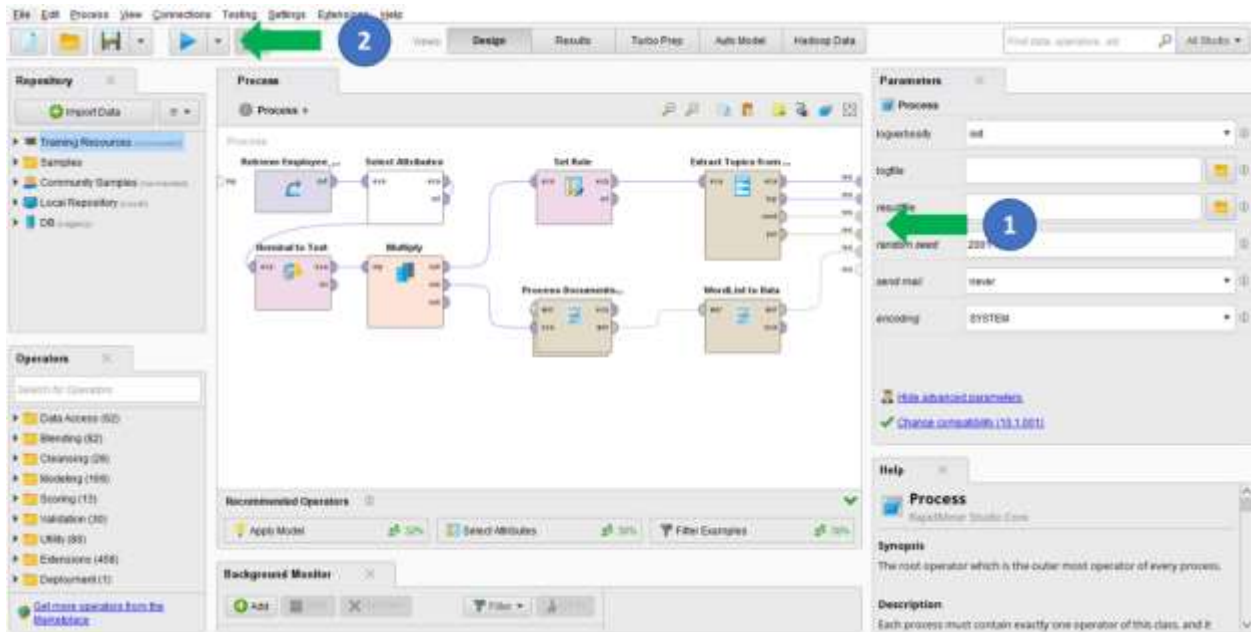
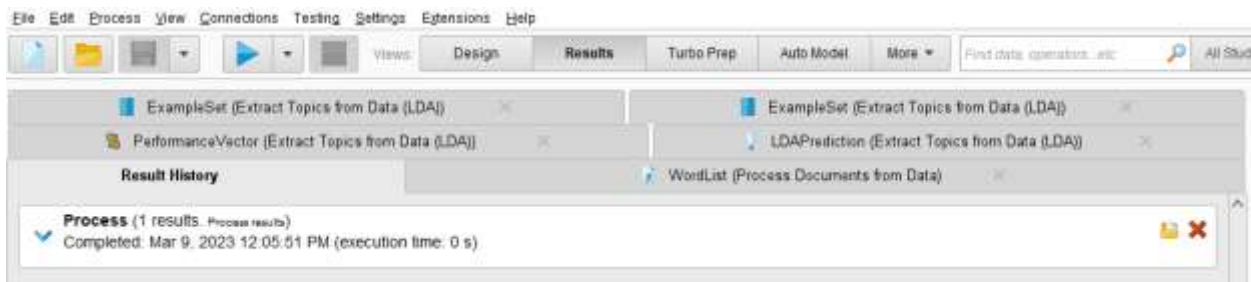


Figure 4. Save the process as "Attrition\_LDA"



**Figure 21.** Connect all outputs to the result ports and click "Run Process."



**Figure 22.** Process Results

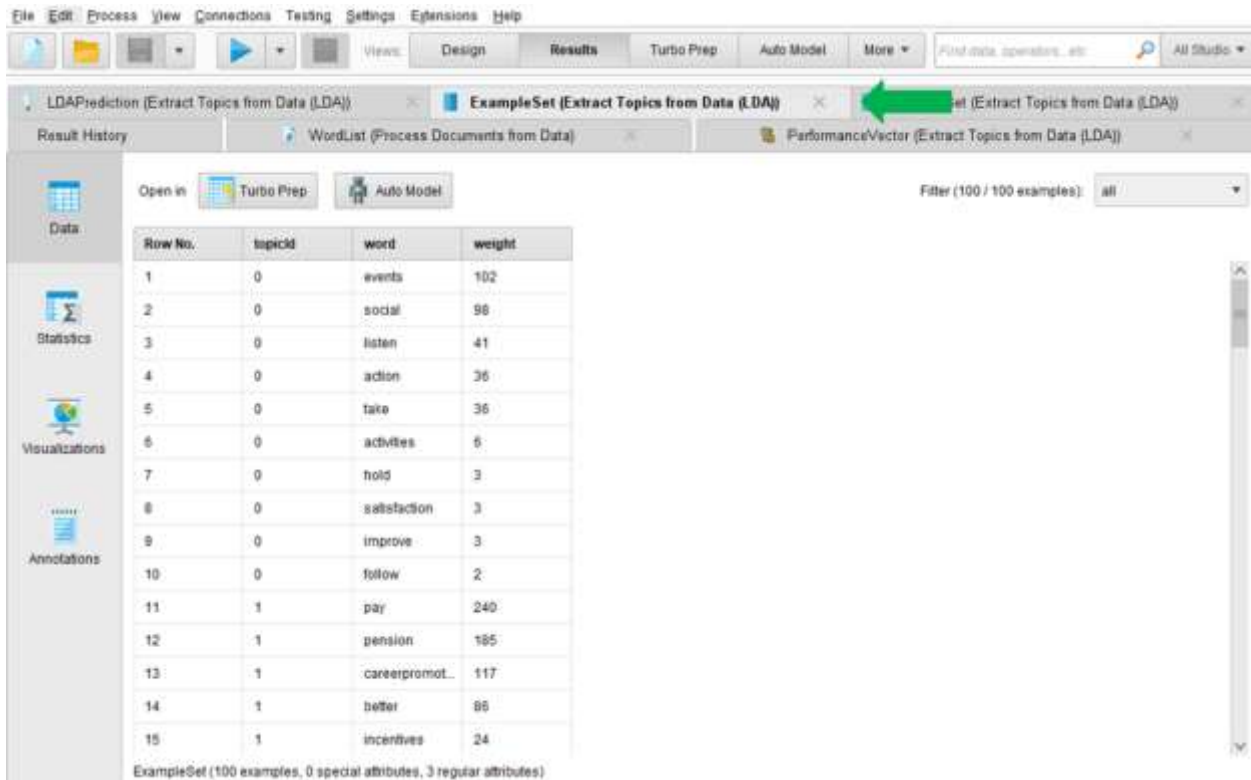


Figure 23. ExampleSet (Extract Topics from Data (LDA))

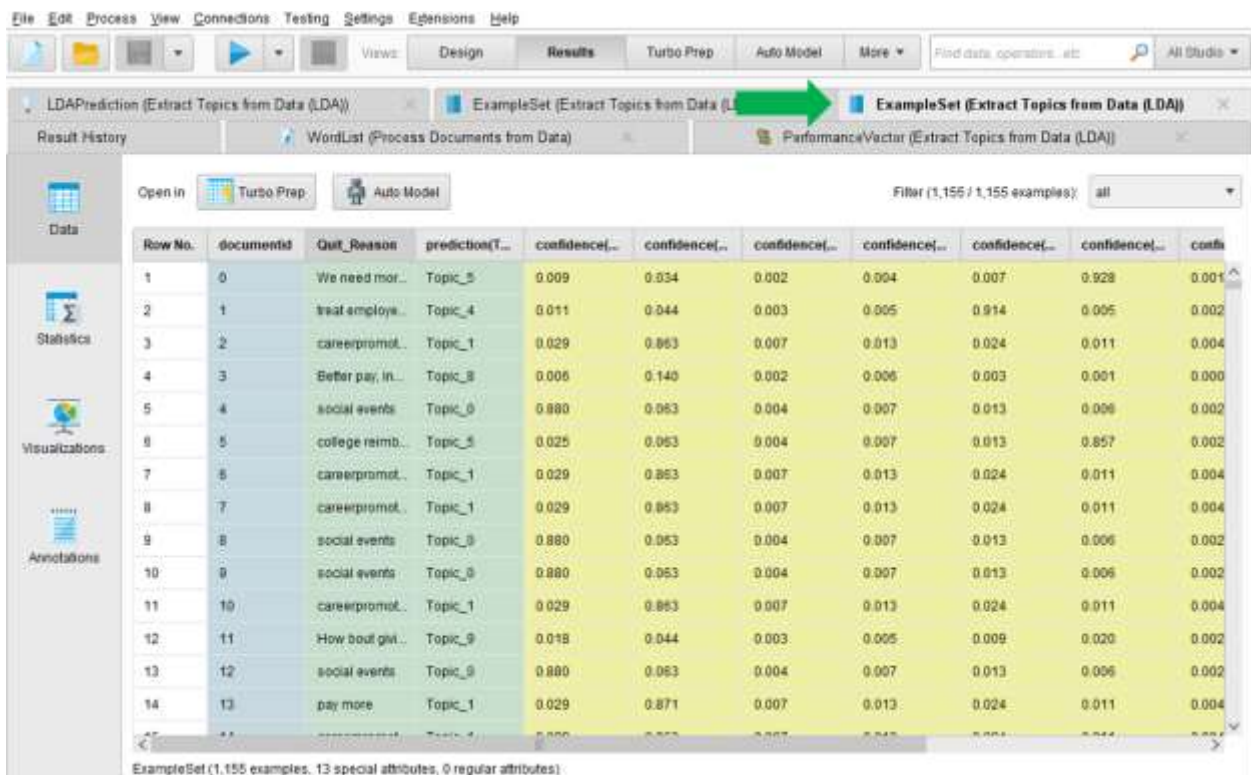


Figure 24. ExampleSet (Extract Topics from Data (LDA))



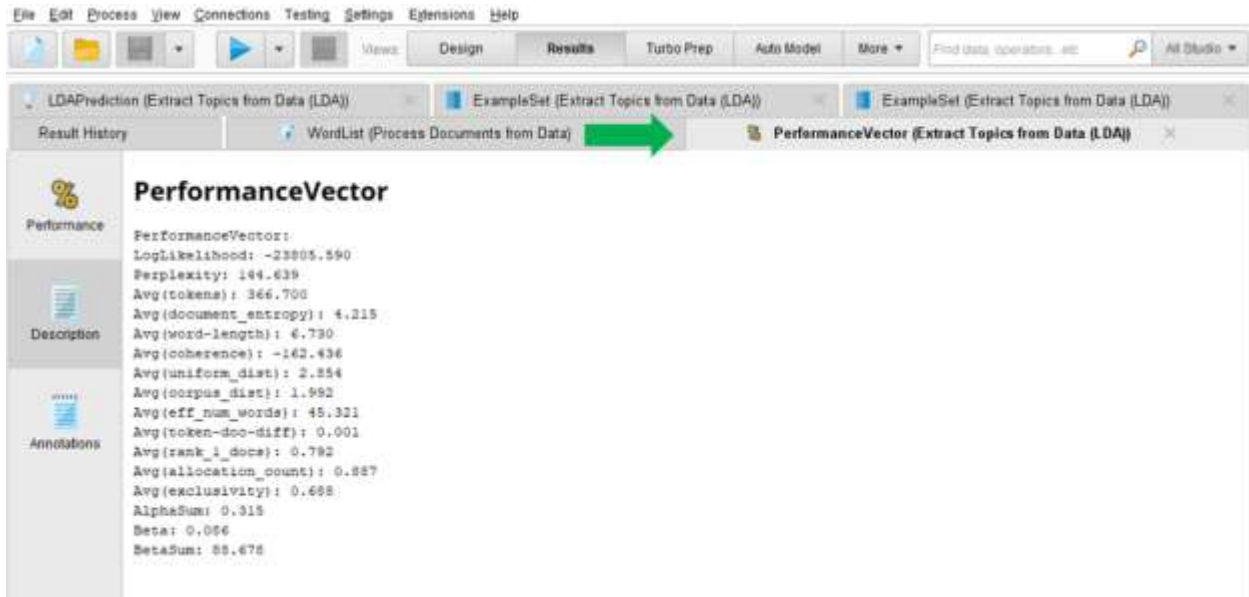


Figure 25. Performance Vector (Extract Topics from Data (LDA))

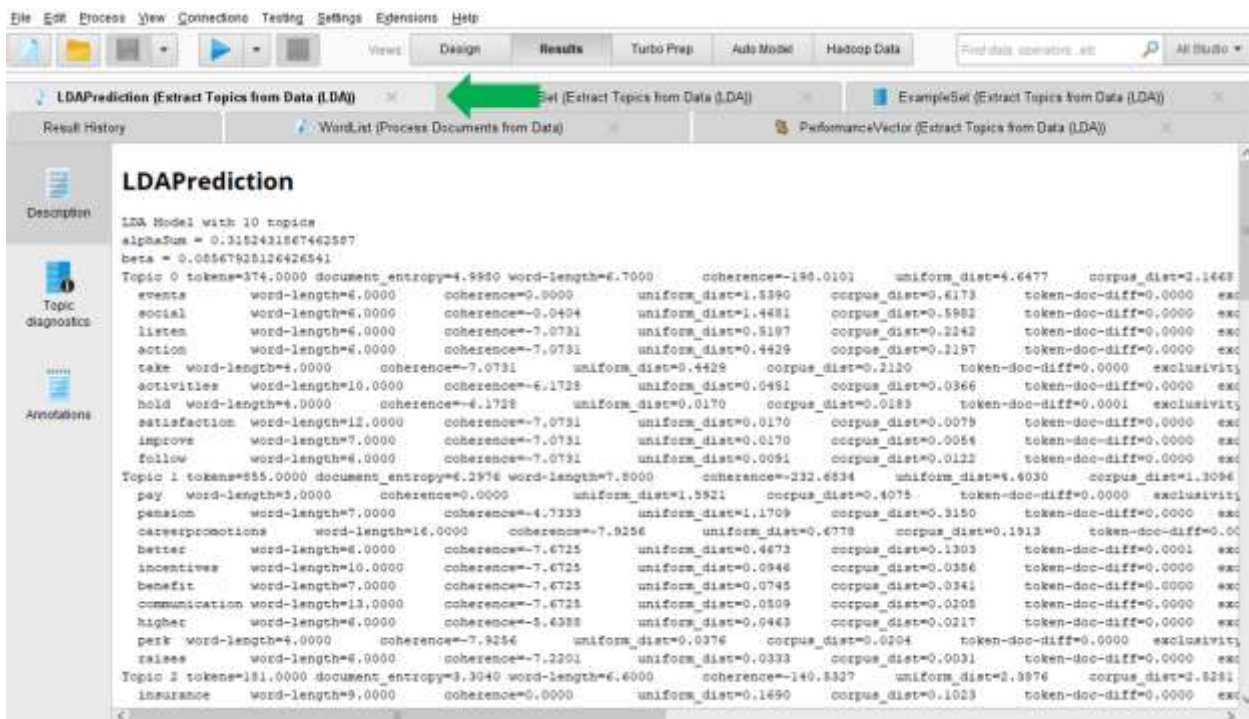
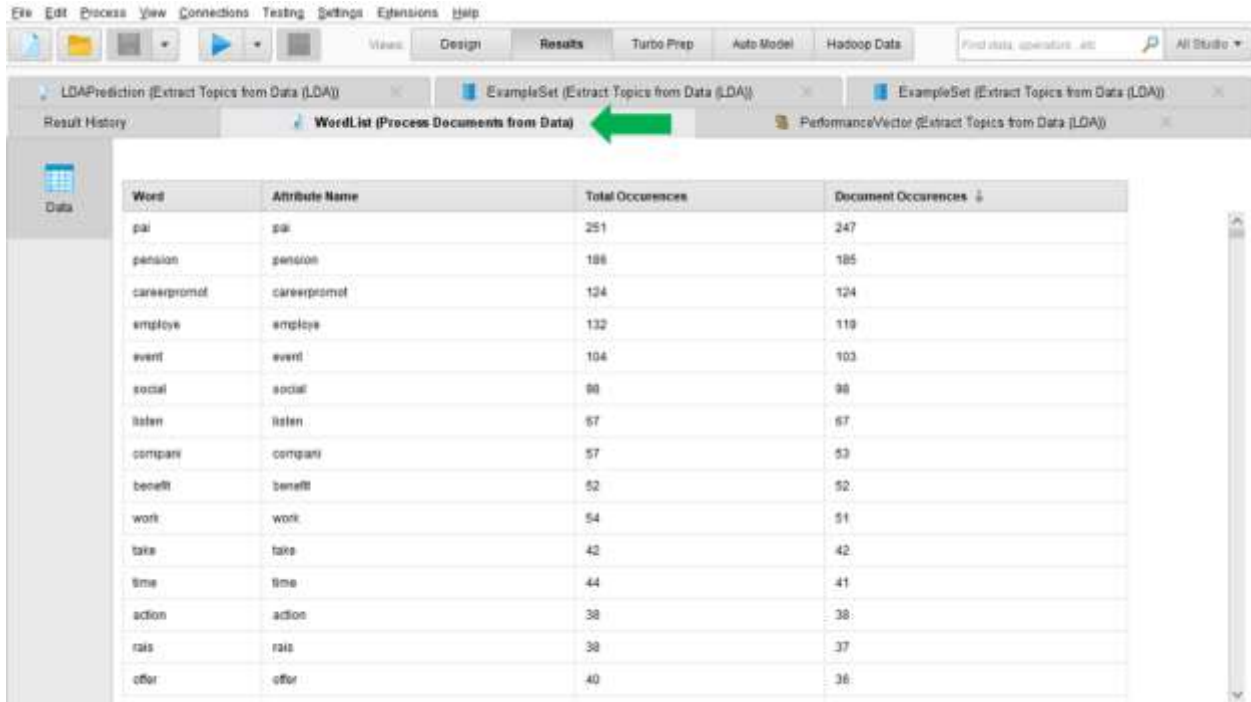


Figure 26. LDA Prediction (Extract Topics from Data (LDA))



The screenshot shows a software interface with a menu bar (File, Edit, Process, View, Connections, Testing, Settings, Extensions, Help) and a toolbar. Below the toolbar, there are several tabs: 'LDAPrediction (Extract Topics from Data (LDA))', 'ExampleSet (Extract Topics from Data (LDA))', 'WordList (Process Documents from Data)', and 'PerformanceVector (Extract Topics from Data (LDA))'. A green arrow points to the 'WordList' tab. The 'WordList' tab displays a table with the following data:

Word	Attribute Name	Total Occurrences	Document Occurrences
pal	pal	251	247
pension	pension	188	185
careerpromot	careerpromot	124	124
employe	employe	132	119
event	event	104	103
social	social	98	98
listen	listen	67	67
compani	compani	57	53
benefit	benefit	52	52
work	work	54	51
take	take	42	42
time	time	44	41
action	action	38	38
rais	rais	38	37
offer	offer	40	36

Figure 27. WordList (Process Documents from Data)