Discovering Business Processes from Organizational Email: Text Classification and Process Mining

Yaghoub Rashnavadi[1], Sina Behzadifard[2], Reza Farzadnia[3], Sina Zamani[4]

[1,2,4] Kharazmi University, Tehran, Iran
[3]Pars Investment Casting co. Tehran, Iran

Abstract. Communication is indispensable for today's lifestyle, and thanks to technology, millions of people can communicate as quickly as possible. The effect of this breakthrough has transformed organizations to the degree that they generate billions of emails daily to facilitate their operations. There is implicit information behind this vast corpus of human-generated content that can be mined and used for their benefit. This paper tries to address the opportunity that email logs can bring to organizations and propose an approach to discover process models by combining supervised text classification and process mining. This framework consists of two main steps, text classification, and process mining. First, Emails will be classified with supervised machine learning, and to mine, the processes fuzzy Miner is used. To further investigate the application of this framework, we also applied this framework over a real-life dataset from a case study organization.

Keywords: Process Mining, Business Processes, Natural Language Processing, Machine Learning

1 Introduction

Email plays a crucial role in today's lifestyle, and with over 124.5[1] Billion business emails sent or received daily, the importance of this communication medium cannot be neglected. Although email is mainly used for personal purposes, many organizations use this tool to facilitate both internal and external communications and even manage complex projects through a large team in different locations.

In this process, the recipients generate a large amount of unstructured natural language documents [1] and are stored in databases or deleted without any further use. However, this corpus of textual data has more to offer about organizations' daily operations as it contains different information like tasks, activities, and processes. However, regarding the business processes and tasks, there is no explicit demonstration of them which can be helpful to manage them.

The lack of invisibility over such processes can cause intentional or unintentional deviations from the business processes' central goal. It may lead to losing the

[1] https://www.campaignmonitor.com/blog/email-marketing/2019/05/shocking-truth-about-how-many-emails-sent/

organizations' competency as the Email contents are flexible and easy to change through time. Even steps from a defined process will be neglected or not followed in the defined sequence. This flexibility makes users act in a way that is not exactly a match from the designed business processes.

So, we assume that the implicit business process and workflows followed in an Email loop can be discovered by mining the unstructured textual data generated by employees through their daily communication. This research investigates an approach to discover the underlying business processes by leveraging a supervised machine learning technique and process mining. The supervised machine learning technique allows to discover activities based on the combination of words in an Email and map them to a defined set of process activities. We used fastText as a classifier of activities, an efficient text classification tool provided by Facebook research[2]. This tool improves the reliability and reproducibility of empirical findings. Then the classified emails are used as an activity, and the consequence event log to be mined with fuzzy Miner.

To analyze this approach's feasibility, we experiment with a real dataset from an organization and provide findings and results.

This research's main contribution is its experimentation with the combination of supervised text classification with a machine learning technique(fastText) and process mining to discover a process model. While there are researches that investigate the same problem, this research tries to propose a new methodology.

Section 2 provides an overview of related work and literature. In Section 3, we will introduce the method in detail. Section 4 presents the experimental results with real-world data, and the conclusion and future research are summarized in section 5.

2 Related Work

The business process discovery problem is one of the popular topics among researchers in the business process management domain. However, the challenge of extracting business processes is to make them explicitly visible to others. Solutions like process mining can facilitate this process by providing algorithms that synthesize a process model from data and event logs [2].

The research area of process mining deals with techniques designed to extract knowledge from event logs [3], and this information can be illustrated as process models. These techniques provide new tools for a wide range of practices, Process Discovery, Monitoring, and Enhancement, with applications that can bring value to the businesses, like facilitating the process alignment and bottleneck analysis while predicting the problems in the execution of processes [4].

Extracting workflows using Natural Language Processing (NLP) and sequence mining techniques is another area that researchers tried to explore with unstructured texts from emails like the research conducted by Shing [1]. They used latent semantic indexing, an unsupervised technique, and density-based spatial clustering of applications with noise to determine the number of clusters and label events.

[2] https://github.com/facebookresearch/fastText

The method of email classification is a well-discovered area. Corston-Oliver and his colleagues from Microsoft Research demonstrated a use case for the Email classification method as a summarizer of emails to create a "to-do list" [5]. In their research, a dataset of 15,741 Email messages was collected, and with the help of human annotators, 146 messages were tagged independently. The prepared dataset was used to train Linear Support Vector Machines (SVMs) to classify emails to tasks in the to-do list. Boldt and Borg also tried to propose an approach to classify customer support Emails to increase service speed. The method they found with the best performance in F1-Score was a Long-Short Term Memory (LSTM) Network to classify Emails into 33 different classes [6].

In 2007, Aalst et al. presented a tool for ProM (Process Mining Tool), EmailAnalyzer. This tool analyzes and transforms Email messages in MS-Outlook into a format used in process mining tools. This research's main goal was to create a social network map from email logs [7].

Banziger et al. [3] have investigated the use of unsupervised machine learning to detect and assign activity labels automatically to messages in a CRM (Customer Relationship Management) tool. Jlailaty et al. also tried to address the same problem with the unsupervised clustering machine learning technique to automatically label Emails with the related activity and mine the respected process model [4].

In another research, J Jlailaty et al. proposed a method that extracts business activity from the Email information. They used a binary supervised classification to eliminate Emails not related to the processes (not process-oriented). They found Gradient Boosting Classifier as the most efficient classifier among the seven tested classification methods. Each relevant sentence was clustered in seven clusters (each representing an activity type) by applying hierarchical clustering [5].

Most of the related literature to this research assumed that each email is associated with only one activity, while this assumption cannot be valid in most real-life situations. Mavaddat et al. proposed a three-step method to extract business processes related to emails. They created a model to demonstrate the interactions among different role instances in a process: Email categorization, conversation network fining, and conversation network tagging [11]. In the first step, they divided Emails into two different categories: business process-related and non-business process-related. To achieve this goal and find the best text mining algorithm, they tried different algorithms like Naïve Bayes and Support Vector Machine (SVM). The authors used WEKA (Waikato Environment for Knowledge Analysis) tool to automatically categorize emails through learning from the previously prepared training dataset (manually labeled email's dataset). The output of this binary classification of emails is used to find email threads, conversations about a similar topic through semantic similarity measurement of each email to other emails. Finally, to label the interactions between each role instance, authors used the Speech Act Theory [12] to classify instances to one of these labels: "assertive, directive, commissive, expressive, and declaration" (illocutionary speech acts). The authors suggested that it is possible to discover business process fragments by exploring the resulted conversation networks and their patterns.

Like Mavaddat et al., Jlailaty et al. tried to address the challenge of non-related Emails and filter the related emails to be mined for activities and related information

[10]. After data preprocessing and cleaning, the sentences of each email are classified into two categories: process-oriented and non-process oriented (a binary classification). This classification is conducted through creating a vocabulary dictionary from a process model repositories and ontologies [5]. For the next step, the business-related Emails were clustered with clustering techniques, and the label was defined in a "semi-automatic way." Finally, each activity instance's metadata was extracted through the linked information in a cluster, like the recipients' organizational role. Researchers tested the proposed approach on the Enron dataset [13].

Rather than physical processes, mental business processes are harder to discover and are executed by "Knowledge workers" [7]. Di Ciccio and his colleagues addressed such processes to discover automatically and mine the corresponding implicit process model. First, extract Email messages and structure communication threads. In the second step, essential parts, activities, and tasks were identified, and finally, the consequence data was mined with Process Describing Grammar. This approach was named "MAILOFMine" [7].

Nassim Laga et al. presented an approach that facilitates labeling Emails and automatically classifies them into process instances, activities IDs, and actors (process-related items) using machine learning techniques [14]. In this approach, the authors created a platform that helps users collaborate to create annotated data and increase data volume through time. They also created a proof of concept with a dataset containing 1026 Emails that used logistic regression classifier as a multiclass predictor to predict process-related tags to validate their approach.

This literature review's main focus was to provide a brief review of the related studies investigating machine learning in Emails' classification to discover the implicit business process inside them. The previous studies' main focus was to practice unsupervised machine learning like clustering techniques and discover the latent activity inside each cluster. On the other hand, some studies experimented with the supervised Email classification methods to classify Emails into non-related and related Emails (Binary Classification) or predict a set of process-related labels (Multiclass Classification).

In this study, we address process discovery with machine learning and process mining. Simultaneously, we tried to use fastText as a classifier of emails and fuzzy Miner to discover processes from a real-world dataset.

3 Method

This research's main objective is to propose a framework that can extract activities from the Email corpus and mine the process models hidden in them. To achieve this goal, we assumed that the information needed to discover activities is in the Email messages, and all the Emails are related to the processes.

This framework consists of three steps: Data Preprocessing, Activity Discovery, and Process Mining. An overview of the approach is depicted in Figure 1.

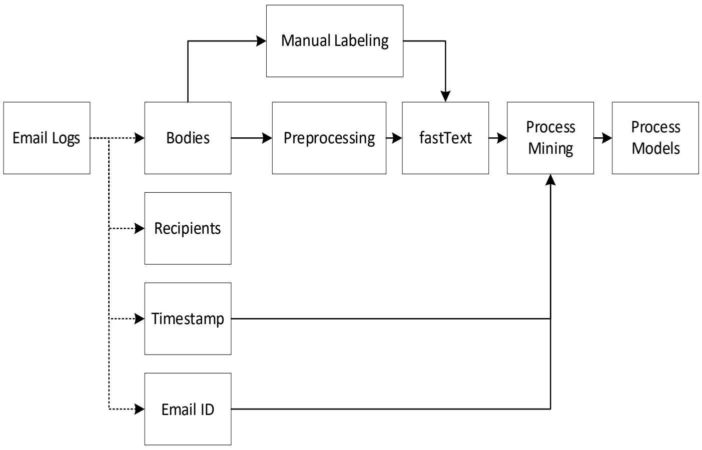

Fig. 1. the proposed framework

3.1 Data collection and preprocessing

Email logs contain four main attributes: Body, Subject, Recipients, and timestamp. While email attachments can also be considered the main attributes, we assume attachments are redundant and out of this research scope.

The email body carries the main message and intention of the sender to the receivers, Recipients. The timestamp is a digital record of the time of occurrence of sending the Emails.

Like any other human-generated data, the body text of the email contains many noises and irrelevant characters. A structured process is needed to reduce noise and its effects by removing or changing them with the appropriate character. Punctuations and symbols are one kind of noise, and all such data should be removed before training or testing the model. This process can vary depending on the Emails' language and the domain, as there may be cases that special symbols can have special meaning.

In this step, an expert human agent investigated sample data from email bodies that his/her primary responsibility is to assign a proper label to each email body based on the workflow's steps and activities defined earlier. This dataset will be used to create the classifier. This human agent must be aware of the organizational environment that Emails are communicated. For instance, emails from finance departments can have contents, words, and acronyms that a non-expert agent cannot easily perceive the activity mentioned in them.

In this step, we had to consider the limitations and challenges of email communication services, Like the challenges regarding the "reply," "reply-all," and "forward. In this project, we did not face such a challenge as the Email databases were designed to log every instance as a separate event and store them separately while connecting them through an ID, LetterID. For situations where emails are stored in a different logic,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 September 2021 doi:10.20944/preprints202005.0007.v2

there are solutions to this challenge. Most Email services add the sequence of such actions in the email thread, and a log of previous communications is stored in the latest Email message.

Additionally, the challenge of mixed messages in emails is widespread and can affect similar researches. While this issue is prevalent, research has investigated how to extract each sentence in an Email and classify each into relevant classes. In this research, we did not face such a challenge, as each email was related to only one activity due to the case study company's communication style to avoid ambiguity in their organizational communication.

3.2 Text Classification

Text Classification is one of the challenges that has many applications in real life. In this research, we use fastText to classify Emails. fastText is an open-source, free, linear-based model, lightweight library that allows users to create text representations and text classifiers [6].

fastText uses a hierarchal softmax function that reduces the computational complexity, leading to a faster search for the predicted class [7]. It works on standard, generic hardware [8].

Unlike deep learning models, fastText is a standard and industrial tool that helps transparency and reproducibility. To prepare data for supervised classification with fastText, we had to add labels to each Email message of the training dataset by adding a "__label__ "to the starting point of messages and the input: __label__ <Email Body>.

3.3 Process Mining

Process mining aims to extract information about event logs (Van Der Aalst, Weijters, and Maruster, 2004). Process Mining is a relatively new research discipline and can be assumed as a bridge between data mining and business process modeling [10]. This step uses the generated event log to discover the process models with the fuzzy Miner. Fuzzy Miner is an algorithm that is focused on unstructured behavior and large event logs. Its output is configurable to reach a desired level of abstraction, and this can be visualized in a fuzzy model. This technique is one of the most valuable tools in case study applications [11].

We preferred to mine the event logs with Disco due to its user-friendly interface. Disco is a complete process mining toolkit from Fluxicon that facilitates process mining [12]. We try fuzzy Miner because of its ability to deal with unstructured processes due to abstraction and clustering techniques and attempts to make more understandable models from unstructured processes [13]. Datasets can be imported in Disco in CSV or Excel, and the processes will be automatically discovered through fuzzy Miner.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 September 2021 doi:10.20944/preprints202005.0007.v2

4 Experiment

In this section, we validate the designed experiment over the case study organization[3]. Pars Investment Casting was officially established in January 2001. The companies' primary focus area is the production of industrial turbo-machineries for Oil and Gas market. The process of manufacturing these complicated machines needs a strong procurement team to manage purchasing materials and services. The importance of the procurement department's processes and their need to discover as-is processes from their organizational email made us focus on only the department's processes. This focus was crucial to our research as communications with more than 100 employees can be very challenging.

The Email log used in this research was extracted from the primary database of organizational correspondence. The dataset contained 100,000 rows of data across all the departments. To narrow down the Emails to the procurement department and, consequently, the related process, we filtered out Emails concerning the "Procurement Department" by searching relevant emails that Purchase experts were involved in at least one email, leading to a dataset of 1933 Email instances.

A sample of 1087 rows of data was labeled based on the existing workflow's primarily designed activities. Fig. 2 demonstrates the workflow of the organization's purchasing process that each purchase requester has to follow to procure the materials through the procurement department.

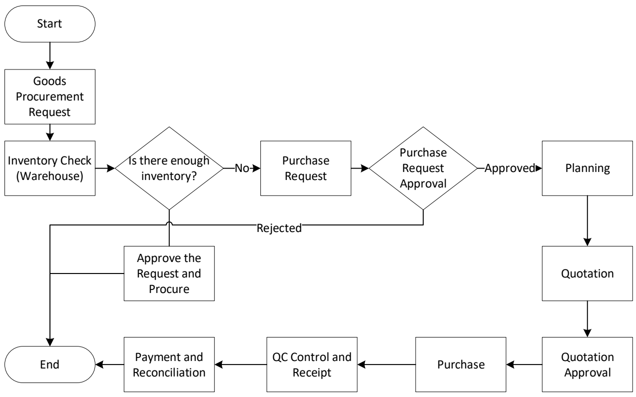

Fig. 2. The process of procurement

The workflow models the process with nine main activities. However, as the above activities could not represent all the communications carried in email, we expanded the 19 label activities. This expansion helped us follow the as-is state of processes in the log. All the labels and their definitions are provided in Table 1. While manual labeling was in progress, we also check for cases that Emails can have several activities, but we could not find any Email message carrying two or more activities. As we find out, the

[3] Pars Investment Casting co., www.parscasting.com

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 September 2021 doi:10.20944/preprints202005.0007.v2

application of emails in the case study organization only facilitates internal communication, process, and paperwork. So, each thread of email is only related to one instance.

Table 1. Labels and their Definitions

Labels	Iterations	Definitions
Comment	273	Informing others about comments on the process or activities conducted
Approval	233	Approval is given when direct managers are agreed to purchasing of a product or service
FYI	208	(For-your-information) Informing others in an Email loop about an event (non-related to the primary process)
Approval Request	84	Requesting for Approval from direct managers
Prioritization Request	55	Requesting for prioritization for the previous activity
Payment Request Approval	54	Approval of a payment request
Payment Approval	50	Approving the initiation of the payment process
Payment	34	Finalization of the payment
Follow up	21	Follow up notification to ask for the latest state of the process
Purchase Request	17	Sending the purchase request to procurement officers
Out-Sourcing	12	Initiation of the out-sourcing process
Payment Request	11	Request for payment
Quotation	10	Quoting process from the market from a similar product
Planning	8	Reference for the planning process
Documentation (Receipt)	3	Documentation of the Goods/Service Receipts
QC approval request	2	Requesting for Quality Control Approval
Documentation (Invoice)	1	Documentation of the Invoices
QC Ok	1	Receiving the Approval of the QC team
Product Approval Request	1	Requesting for Product Approval from stakeholders

In Table 2, a sample from the data has been provided. Pars Investment Casting is an Iranian Company, so the communications are in Persian.

Table 2. A sample of the data

Letter ID	Employee Id Sender	Employee Id Receiver	Receive Date	Email	Label	Letter ID	Employee Id Sender	Employee Id Receiver
81177	7557	6448	2018-05-28 15:44:58	باسلام احتراما جهت استحضار	FYI	81177	7557	6448

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 September 2021 doi:10.20944/preprints202005.0007.v2

81177	8448	8532	2018-05-29 09:37:26	باسلام لطفا پس از مذاکره اقدام شود	Approval	81177	8448	8532

To prepare the data for activity classification, we first had to preprocess the data. All the steps from data preprocessing to labels prediction were followed in the Python[4] environment and Google Collaboratory. The steps of the preprocessing process are as below:

1. Removing redundant symbols and punctuations (“!”#$%&()*+,-;⇔?@[\]^_`{|}~<.*?>|&([")
2. Removing redundant spacings
3. Removing numbers
4. Removing HTML codes

4.1 Data slicing, training, and testing dataset

To build the classifier, we experimented with 15 different feature settings of fastText. Learning rate, embedding dimension, n-grams, epoch were the parameters that we tried to experiment with, and the best configuration was found as below:

1. lr(learning rate) = 0.9
2. embedding dimension=150
3. epoch=50
4. n-grams= (2,10)

With these settings, a training dataset containing 900 rows, and a test dataset with 176 rows of data, we could achieve the best training accuracy of 98.8% and validation accuracy of 85.8%. With this model, we labeled all the remaining datasets to create the final event log. To better understand the generated event log, we provided a few data analyses in the following. The distribution of the labeled data is demonstrated in Fig 3.

[4] https://www.python.org/

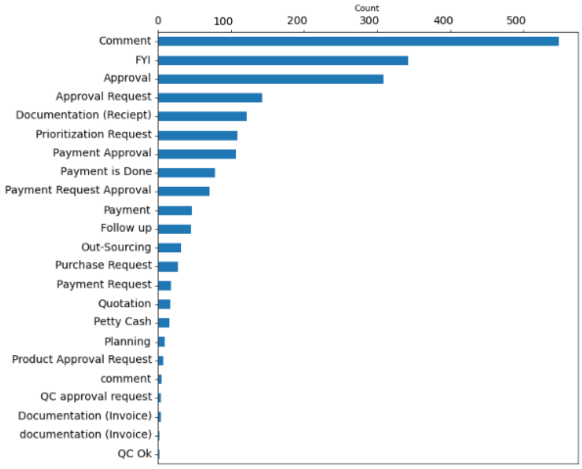

Fig. 3. the distribution of the Activities after Activity classification

As it is clear from Fig. 3, the distribution of labels is not normal, and most of the Emails are only to comment on the process or inform others about the status of the process. The second most repeated emails were about the Approvals.

In Fig 4, the Emails' string's length is provided in a histogram with bins of 50. The maximum length of emails is 62 words, and the minimum has a length of one word. In figure 5, the frequency of unique words used in emails is demonstrated. Most Emails had 10 to 40 unique words.

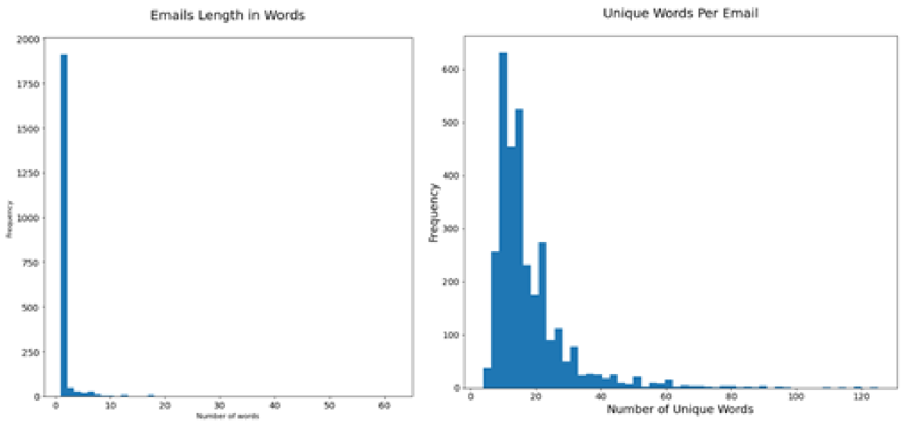

Fig. 4. Emails' lengths and frequency of unique words used in emails

We used Disco to apply the fuzzy Mining algorithm to the extracted event log and discover the process model in the next step. The output of this step is illustrated in Fig. 6. Based on the Disco's statistics, this dataset contained 1933 events in 204 cases. The median duration was 18.2 days, and it took 54.1 days for each case to complete. Seventy people were involved in generating this dataset, and only 81 cases could be captured from their start.

At first glance, in Fig. 6, the discovered model seems complicated, showing how complex communication can be in only one department of an organization. Based on the information from the mined email log, 57.27% of activities were not directly connected to the procurement process, but their role is crucial to transfer information. The frequency of such activities is illustrated through the darkness of colors.

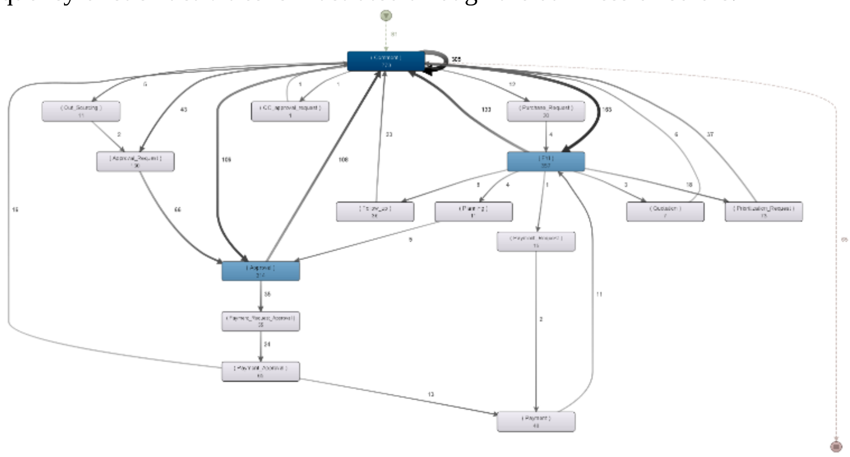

Fig. 5. The Discovered process model based on Email Log with the fuzzy Miner in Disco

With more exploration in the mining results, we figured out that there are other processes that people follow, which are not followed the standard workflow or only communicated to initiate or close a process. In contrast, the rest were followed offline, without any record in this system. This discovery was beneficial for the organization as the role of such a system has never been so clear for the management team, and its ability to enable them to audit processes was also attractive.

With an increase in the abstraction levels by the fuzzy Miner, Fig. 7, we were able to see that "Approvals" are the central part of the workflow that employees follow in their emails, and the rest of the workflow is not logged as much these activities. From this model and analysis of the designed workflow, it can be perceived that the primary application of their Email-based communication is only to communicate managerial approvals and decisions about the process. At the same time, there is more potential in performing all the communications in the email.

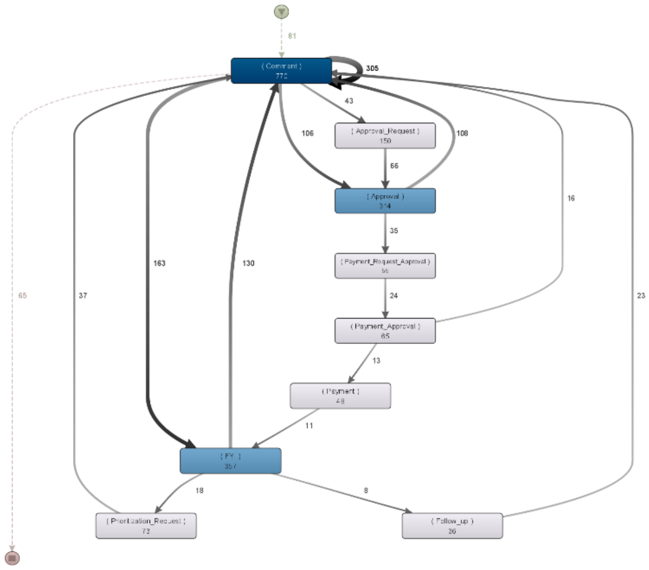

Fig. 6. Higher abstraction of the discovered process model with the cutoff of 55%

5 Conclusion and future work

This paper examined the proposed framework for business process discovery from an organization's email logs. We combined the supervised text classification technique with fastText and the fuzzy Mining to mine processes to accomplish this goal.

We experimented with the framework over a case study to show its practicality. The information mined showed it could be beneficial for organizations with email infrastructure to monitor their processes and improve them based on the discovered knowledge.

In the case study, we figured out that emails were only about receiving approvals from the managers and not following the rest of the procurement process. Such a tool can increase organizations' understanding of their processes, mostly followed through unstructured data, like text, making their organizations more transparent. Managers can be aware of the processes that are conducted and avoid possible deviations from the designed processes. At the same time, they can monitor business processes and solve real-time bottlenecks. This capability can help them take measures, solve pain points, and increase agility in business processes [22]. In the future, we will try to experiment with more data and implement semi-supervised methodologies to classify Emails. We also suggest experimenting with pre-trained word embeddings like Glove or any other word embedding and text feature extraction techniques for further research.

Our proposed methodology's primary limitation is its inflexibility to change activities, which is the most critical drawback of supervised machine learning. The model can be trained through time intervals with new datasets to solve this challenge, or like

 doi:10.20944/preprints202005.0007.v2

Laga et al., there can be a platform to facilitate the labeling process with the help of users in an organized structure [14].

During this research, we figured out that many factors affect the organization's communications, making process discovery very challenging with analyzing contents generated by them, so this topic can also be investigated with enthusiastic researchers.

The other research path that we can propose is to investigate the role of organizational culture in communication as we figured out the culture of the case study organization was unique in its case. The vocabulary domain used through the organization can have a meaning beyond the ordinary meaning of them in the ordinary contexts outside of the organization.

References

1. [1] L. Shing et al., "Extracting Workflows from Natural Language Documents: A First Step," in Lecture Notes in Business Information Processing, vol. 342, Cham: Springer International Publishing, 2019, pp. 294–300.
2. [2] B. F. Van Dongen, A. K. Alves De Medeiros, and L. Wen, "Process mining: Overview and outlook of Petri net discovery algorithms," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5460 LNCS, Springer, 2009, pp. 225–242.
3. [3] W. Van Der Aalst et al., "Process mining manifesto," Lect. Notes Bus. Inf. Process., vol. 99 LNBIP, no. PART 1, pp. 169–194, 2012, doi: 10.1007/978-3-642-28108-2_19.
4. [4] C. J. Turner, A. Tiwari, R. Olaiya, and Y. Xu, "Process mining: From theory to practice," Bus. Process Manag. J., vol. 18, no. 3, pp. 493–512, Jun. 2012, doi: 10.1108/14637151211232669.
5. [5] S. Corston-Oliver, E. Ringger, M. Gamon, and R. Campbell, "Task-Focused Summarization of E-mail BT - ACL-WS2004A," Proc. ACL Work., pp. 1–8, 2004, [Online]. Available: http://tangra.si.umich.edu/clair/intranet/bibs/summbib/Corston-Oliver&al Task-Focused Summarization of E-mail.pdf%5Cnhttp://tangra.si.umich.edu/clair/intranet/bibs/summbib/Corston-Oliver&amp%5Cnal Task-Focused Summarization of E-mail.pdf.
6. [6] A. Borg, M. Boldt, O. Rosander, and J. Ahlstrand, Email classification with machine learning and word embeddings for improved customer support, vol. 33, no. 6. 2021.
7. [7] C. Di Ciccio, M. Mecella, M. Scannapieco, D. Zardetto, and T. Catarci, "MailOfMine - Analyzing mail messages for mining artful collaborative processes," in Lecture Notes in Business Information Processing, Jun. 2012, vol. 116 LNBIP, pp. 55–81, doi: 10.1007/978-3-642-34044-4_4.
8. [8] R. Banziger, A. Basukoski, and T. Chaussalet, "Discovering Business Processes in CRM Systems by Leveraging Unstructured Text Data," Proc. - 20th Int. Conf. High Perform. Comput. Commun. 16th Int. Conf. Smart City 4th Int. Conf. Data Sci. Syst. HPCC/SmartCity/DSS 2018, pp. 1571–1577, 2019, doi: 10.1109/HPCC/SmartCity/DSS.2018.00257.
9. [9] D. Jlailaty, D. Grigori, and K. Belhajjame, "Business Process Instances Discovery from Email Logs," in Proceedings - 2017 IEEE 14th International Conference on Services Computing, SCC 2017, Jun. 2017, pp. 19–26, doi: 10.1109/SCC.2017.12.
10. [10] D. Jlailaty, D. Grigori, and K. Belhajjame, "On the elicitation and annotation of business activities based on emails," in Proceedings of the ACM Symposium on Applied Computing, 2019, vol. Part F147772, pp. 101–103, doi: 10.1145/3297280.3297534.

11. [11] M. Mavaddat, I. Beeson, S. Green, and J. Sa, "Facilitating Business Process Discovery using Email Analysis," in BUSTECH 2011 : The First International Conference on Business Intelligence and Technology, 2011, no. c, pp. 40–44, [Online]. Available: http://www.thinkmind.org/index.php?view=article&articleid=bustech_2011_2_30_90043.
12. [12] A. Koller and J. R. Searle, "Speech Acts: An Essay in the Philosophy of Language," Language (Baltim)., vol. 46, no. 1, p. 217, 1970, doi: 10.2307/412428.
13. [13] B. Klimt and Y. Yang, "The enron corpus: A new dataset for email classification research," in Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 2004, vol. 3201, pp. 217–226, doi: 10.1007/978-3-540-30115-8_22.
14. [14] N. Laga, M. Elleuch, W. Gaaloul, and O. A. Smaili, "Emails analysis for business process discovery," CEUR Workshop Proc., vol. 2371, no. 2, pp. 54–70, 2019.
15. [15] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," 15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf., vol. 2, pp. 427–431, Aug. 2017, doi: 10.18653/v1/e17-2068.
16. [16] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, and T. Mikolov, “FastText.zip: Compressing text classification models,” arXiv1612.03651 [cs], pp. 1–13, Dec. 2016, [Online]. Available: http://arxiv.org/abs/1612.03651.
17. [17] A. Alessa, M. Faezipour, and Z. Alhassan, "Text classification of flu-related tweets using FastText with sentiment and keyword features," in Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018, Jun. 2018, pp. 366–367, doi: 10.1109/ICHI.2018.00058.
18. [18] W. Van Der Aalst, T. Weijters, and L. Maruster, “Workflow mining: Discovering process models from event logs,” IEEE Trans. Knowl. Data Eng., vol. 16, no. 9, pp. 1128–1142, 2004, doi: 10.1109/TKDE.2004.47.
19. [19] C. W. Günther and W. M. P. Van Der Aalst, "Fuzzy mining - Adaptive process simplification based on multi-perspective metrics," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4714 LNCS, pp. 328–343, 2007, doi: 10.1007/978-3-540-75183-0_24.
20. [20] C. W. Günther and A. Rozinat, “Disco: Discover your processes,” CEUR Workshop Proc., vol. 936, pp. 40–44, 2012.
21. [21] W. M. P. Van Der Aalst and C. W. Günther, "Finding Structure in Unstructured Processes: The Case for Process Mining," in Proceedings - 7th International Conference on Application of Concurrency to System Design, ACSD 2007, 2007, pp. 3–12, doi: 10.1109/ACSD.2007.50.
22. [22] R. L. Raschke and J. S. David, "Business process agility," in Association for Information Systems - 11th Americas Conference on Information Systems, AMCIS 2005: A Conference on a Human Scale, 2005, vol. 3, pp. 1522–1527.