IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 August 2022, accepted 10 October 2022, date of publication 21 October 2022, date of current version 8 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3216418

RESEARCH ARTICLE

Augmenting Business Process Model Elements With End-User Feedback

SANAM AHMED[1] **AND KHURRAM SHAHZAD**[2]
[1]Department of Software Engineering, University of the Punjab, Lahore 54590, Pakistan
[2]Department of Data Science, University of the Punjab, Lahore 54590, Pakistan
Corresponding author: Sanam Ahmed (sanam.ahmad@pucit.edu.pk)

ABSTRACT COVID-19 has imposed unprecedented restrictions on society which has compelled organizations to work ambidextrously. Consequently, organizations need to continuously monitor the performance of their business process and improve them. To facilitate that, this study has put forth the idea of augmenting business process models with end-user feedback and proposed a machine learning based approach (AugProMo) to automatically identify correspondences between end-user feedback and elements of process models. Furthermore, we have generated three valuable resources, process models, feedback corpus and gold standard benchmark correspondences. Also, 2880 experiments are performed to identify the most effective combination of word embeddings, data balancing techniques, feature vectors and machine learning techniques. The study concludes that the proposed approach is effective for augmenting business process models with end-user feedback by identifying correspondences between them.

INDEX TERMS Business process management, business process innovation, explorative business process management, user feedback analysis, word-embeddings, machine learning.

I. INTRODUCTION

Business processes represent the way organizations produce products or deliver services to end-users. Formally, it is defined as a set of activities that are performed in a certain order to achieve a business goal [1]. Business Process Management (BPM) is the discipline that deals with designing, implementing, monitoring, controlling and redesigning business processes [2]. The graphical representation of a business process is referred to as a business process model or process model. Business Process Modeling Notation (BPMN) is the de jure standard for designing business process models [3].

Nowadays, businesses are following an ambidextrous approach to achieve a competitive edge. As per the ambidextrous theory, the exploration approach involves searching and experimentation to innovate new products and services. This is contrary to the exploitation approach which aims to refine and improve existing features in products and services. Recent studies have discussed that ambidexterity theory has been used in the BPM discipline both theoretically and practically [4]. Likewise, another notable study has endorsed that the convergence of digital innovation and BPM would make the BPM discipline more ambidextrous [5].

Digital innovation has revolutionized traditional products and services. For instance, during the last few years Zoom, Google Meet and Microsoft Teams have changed the way meetings are conducted. Similarly, COVID-19 has paved the way to rethink the business processes of various domains for the pandemic, as well as the post-pandemic era. Likewise, we contend that the advancements in natural language processing and deep learning can be leveraged to the benefit of the BPM discipline.

Customer feedback has a vital role for service-oriented companies where customer satisfaction has higher significance [6]. Several studies have been conducted to extract information from customer feedback using disruptive technologies. Some studies of these studies have proposed topic modeling and machine learning based approaches to identify informative customer feedback for mobile applications [7], [8]. Furthermore, another study [9] has proposed an approach for summarizing customer feedback about mobile applications. Similarly, leading researchers have argued that the end-user feedback about business processes contains valuable information which can be used for improving

The associate editor coordinating the review of this manuscript and approving it for publication was Zhangbing Zhou.

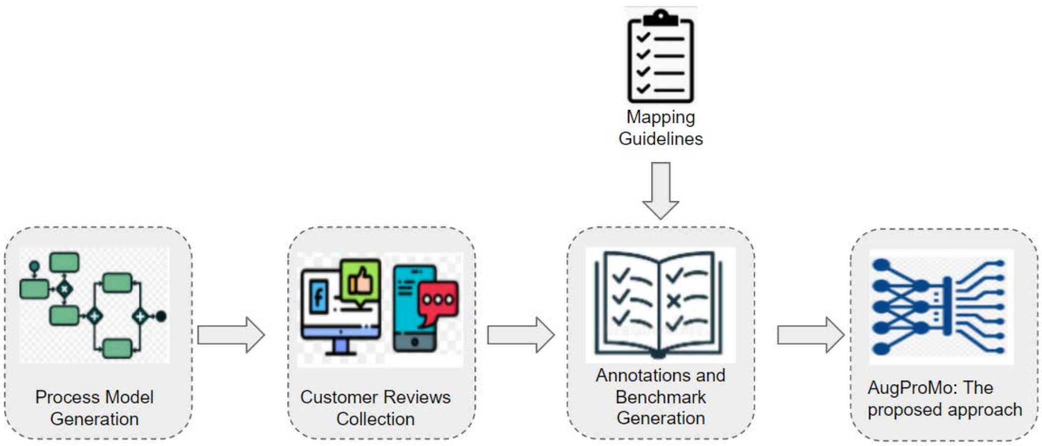

FIGURE 1. The framework of the study.

business processes [10]. Also, a recent attempt has been made to automatically extract business process redesign suggestions from end-user feedback [11].

In this study, we propose the notion of augmenting process models with end-user feedback that will pave the way for various process innovation initiatives. The augmentation of process models requires identifying correspondences between end-user feedback and process model elements. To that end, this study has proposed a novel approach (AugProMo) to identify correspondences between the two artifacts. AugProMo employs a machine learning based approach coupled with state-of-the-art word embeddings and data balancing techniques to automatically identify correspondences between end-user feedback and process model elements. Figure 1 presents an overview of the proposed approach. In particular, the key contributions are the following.

- *Real-world process models:* We have designed six real-world process models from three different domains, education, food and transport, and made them publicly available for the community. Two of these process models are designed from scratch, whereas the remaining four process models are designed by tracing the functionality of mobile apps. The process models are designed in Camunda, a free and open workflow platform which allows exporting models in multiple machine readable formats [12]. The designed models are composed of 93, 89, 87, 85, 67 and 41 process model elements, which represents the substantial effort involved in designing these process models. These models are publicly available for the research and development.
- *User feedback corpora and benchmark datasets:* For each process model a corpus of end-user feedback is generated. Two of these feedback corpora are generated by conducting a survey from end-users, whereas the remaining four corpora are generated by scrapping the social media and Google Play Store. The use of these different sources for generating the end-user feedback corpora demonstrates the suitability of diverse ways of generating the corpora. Furthermore, six gold-standard benchmark corpora are generated in which correspondences between end-user feedback and process model elements is defined. For generating the benchmark, all combinations of process model elements and feedback sentences are generated and subsequently, the mappings among these combinations are manually assessed.
- *A mapping approach:* Finally, we have proposed a novel approach (AugProMo) for the automatic identification of correspondence between the end-user feedback and elements of process models. The proposed approach relies on the most effective combination of word embeddings, data balancing techniques and novel feature vectors, which are fed to machine learning techniques for learning and prediction. Finally, 2880 experiments are performed on the six benchmark datasets using six machine learning and deep learning techniques, four balancing techniques, and four types of word embeddings.

The rest of the paper is organized as follows: the research problem of identifying correspondences between end-user feedback and process model elements is illustrated in Section II. Literature review and dispositions from the previous work are discussed in Section III. The details of the process model generation process and their specifications are presented in Section IV, whereas the details of the end-user feedback corpora are presented in Section V. The procedure used to generate the Element-Feedback benchmark corpora and their specifications are presented in Section VI. The proposed AugProMo approach is presented in Section VII, and the details of the experiments that we have conducted are presented in Section VIII. The results of the experiments and their analysis are presented in Section IX. Section X discusses how the types of organizations that can benefit from the proposed approach. Finally, the conclusions of the study are presented in Section XI.

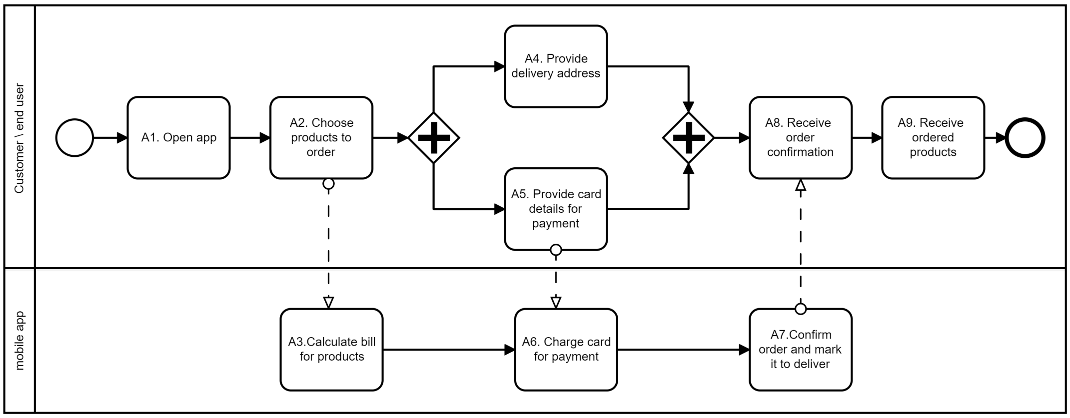

FIGURE 2. Excerpt version of an example process.

TABLE 1. Illustration of correspondences between end-user feedback and process model elements.

ID	Feedback	Elements
R1	I opened the app it took too long to load.	A1
R2	I selected items to buy in few clicks.	A2
R3	I selected the products to order and provided my delivery address its supper easy.	A2, A4
R4	I received my order within a day superb.	A9
R5	Overall good.	NA

II. PROBLEM ILLUSTRATION

This section illustrates the problem of identifying correspondences between end-user feedback and activities of business process models. Consider a business process of buying products using a mobile app. An excerpt of the process model for purchasing a product is presented in Figure 2. The model is designed using Business Process Modeling and Notation (BPMN) which is the de jure standard for business process modeling. In the figure, the start and end events are represented by a circle, actors are represented by horizontal swimlanes, activities are presented by rounded-edged rectangles, whereas the diamonds having (+) symbols represent the parallel gateways between activities.

It can be observed from the figure that the process starts when a customer opens the app to order products. The customer uses this app to choose the products to be ordered. In response, the app then calculates the bill for the chosen products. Subsequently, the customer provides the billing and delivery information after which the payment is deducted. Finally, the product is delivered to the customer. Table 1 presents a sample of end-user feedback sentences and their corresponding elements. For instance, it can be observed from Table 1 that the feedback sentence R1 contains an expression about starting the mobile app. Due to the presence of the label of activity A1 (''Open app'') in R1, it is clear that R1 corresponds to activity A1.

While conducting the study, it is observed that there are several challenges in identifying correspondences between the end-user feedback and process model elements. A key challenge is that the feedback sentences use the vocabulary that is different from the activity labels. For example, feedback sentence R2 is about choosing products to be ordered but there is no activity label in the process model that uses the vocabulary used in R2. However, if the semantics are considered it becomes clear that R2 corresponds to A2 (Choose products to order). Another notable challenge is that a feedback sentence may correspond to more than one activity or it can be ambiguous. For instance, R3 corresponds to two activities, A2 and A4. Furthermore, it is also possible that the feedback sentences are generic and may not correspond to any individual activity. R5 is an example feedback sentence that does not correspond to any activity of the process model. Therefore, we contend that the identification of correspondences between end-user feedback and process model activities is a challenging task which requires the attention of researchers.

III. RELATED WORK

A considerable number of studies have been conducted to bridge the two areas of research, Natural Language Processing (NLP) and BPM. Some of these studies analyze the labeling style of process elements to identify incorrect labeling styles [13], whereas others develop techniques to automatically generate textual process descriptions [14]. Similarly, work has been done to automatically generate process models from textual process descriptions [15], as well as on checking process compliance of process models against textual descriptions [16]. Also, notable studies have been conducted on the automatic detection of inconsistencies between the two descriptions [17]. Furthermore, some studies have employed NLP based techniques for finding relevant process models from a collection of process models using their textual descriptions [18], [19]. For brevity, this section summarizes the three directions of research that are most relevant to this study. These three directions are: a) process model matching, b) aligning textual descriptions and process models, and c) analysis of customer feedback on business processes.

A. PROCESS MODEL MATCHING (PMM)

Process Model Matching (PMM) refers to the identification of correspondences between elements of two process models [20]. The PMM problem gained attention in 2013 when the first PMM Contest was organized to evaluate the effectiveness of the matching techniques [21]. Seven matching techniques participated in the contest and the results of all the techniques were publicly released. Subsequently, the second edition of the PMM contest was organized in 2015 in which twelve teams participated [20]. Similar to the first contest, the datasets and results of the techniques were publicly released. The resources released from the two contests served as a catalyst for advancements in PMM techniques. For instance, some of the subsequent studies evaluated the effectiveness of syntactic and semantic similarity measures [22], whereas other studies, such as [23], enhanced the datasets for a deeper evaluation of the matching techniques. Another study has used state-of-the-art word embeddings to achieve a very high F1 scores for all the publicly available datasets [24].

Identifying correspondences between elements of two process models is a significantly different and less challenging research problem than identifying correspondences between process model elements and end-user feedback due to two reasons. Firstly, the labels of elements of the two process models are of comparable length, whereas the length of end-user feedback can be significantly different from process model elements making the identification more challenging. Secondly, the labels of process model elements typically comply a labeling style, verb-object or action-noun, whereas the end-user feedback is free-text with does not comply with any writing style.

B. ALIGNING TEXTUAL DESCRIPTIONS AND PROCESS MODELS (TEXT-MODEL-ALIGNMENT)

Identifying correspondences between aligning textual process descriptions and process models is another related direction of research. A notable study [25] proposed an Integer Linear Programming (ILP) based approach to align textual process description with process model elements. In contrast, a recent study [26] has introduced a framework for the automatic extraction of annotated process descriptions and analysis with the process model and the corresponding process event logs.

The task of identifying correspondences between end-user feedback and process models is substantially different from aligning textual descriptions and process models due to two reasons. Firstly, the textual process descriptions are formally written text which includes relevant business vocabulary, in contrast, the end-user feedback is a casually written informal text without the relevant business vocabulary which makes it a challenging research problem. Another difference between the two problems is that the textual process description is essentially the description of the workflow of the organization, whereas the end-user feedback primarily contains their experiences, sentiment, suggestions, etc., while ambiguously referring to the activities of a workflow.

C. ANALYSIS OF PROCESS FEEDBACK (FEEDBACK-ANALYSIS)

Recent studies have highlighted that end-user feedback about business processes is of higher significance as it includes information on the problems faced by end-users, their sentiments, as well as suggestions for process redesign [11], [27]. The initial work in this research direction proposed to classify process feedback across the widely used process performance dimensions, time, cost, quality and flexibility [27]. Another study has performed sentiment analysis on end-user feedback to assess the level of customer satisfaction [28]. The most recent work [28] has proposed to extract suggestions for explorative process redesign. However, the key limitation of these studies is that they utilize user feedback for various types of analysis, while the corresponding process model remains an auxiliary component of the analyses.

To the best of our knowledge, merely a single study has been conducted which aims to map end-user feedback to process model elements [29]. However, this study significantly differs from our previous work [29] having the following dispositions. Firstly, this study includes process models from three domains, education, food and transport, whereas the previous study considered process models from the education domain only. Secondly, the feedback corpora developed in this study are manifolds larger in size than the ones developed in the previous study. Thirdly, the existing study used basic syntactic and semantic string matching measures to identify the correspondences which achieved a low F1 score. In contrast, this study has proposed a novel approach for identifying correspondences which achieved a very high F1 score.

IV. PROCESS MODEL GENERATION

This study has used six process models (PM1 - PM6) from three diverse domains education, food and transport. The first two models, PM1 and PM2, are generated using the classical approach that is used in our previous study [29], whereas the remaining four models (PM3 - PM6) are the newly generated process models. Where, PM1 is the admission process model of a university, and PM2 is the course registration process model of another university. The remaining four models are generated by tracing back the real-world mobile applications from two domains, food and transport. For generating these four models, we selected three active users of each app for tracing the underlying business process. Each user has used the respective mobile app for more than 6 months which represents their familiarity with its functionalities. The users were asked to perform multiple end-to-end transactions using the mobile app and record each step of the process. The recorded steps were discussed by the users and the pathways were synthesized to define the control flow. Finally, the generated process models were designed using BPMN. Accordingly, the four process models (PM3 - PM6) are

TABLE 2. Specifications of the designed process models.

Item	PM1	PM2	PM3	PM4	PM5	PM6
No of Activities	36	68	54	58	64	72
No of Actors	3	4	3	3	3	3
No of Gateway Transitions	2	17	10	24	20	18
Total Count of Process Elements	41	89	67	85	87	93

generated. The reason for tracing-back process models from mobile apps is that the underlying process models are not publicly available, whereas the end-user feedback of the mobile apps is publicly available. Therefore, if the process models are generated, we can leverage the end-user feedback which is publicly available on the Google play store for these mobile apps.

An overview of the specifications of the six process models is presented in Table 2. It can be observed from the table that the process models are of different sizes, including a small process model of 36 elements, as well as a large process model of 93 elements. It can also be observed from the table that a simple process has merely 2 gateways, whereas the complex process model has 24 gateways. These numbers represent that the collection includes a process model having a large diameter, as well as a process model with ample breath. These specifications represent that significant effort is involved in designing these process models.

V. END-USER FEEDBACK CORPORA

This section presents the second contribution of this study, six corpora of end-user feedback for the corresponding six process models. For generating the feedback corpora, feedback is collected from multiple sources which include, scrapping Facebook pages, scrapping the feedback available on the Google play store and conducting end-user surveys. The details of the protocol used for generating the end-user corpora are discussed below.

A. DATA COLLECTION

As discussed earlier, two approaches were employed for generating process models, the classical approach and tracing the mobile apps. For each type of process model, a different approach is used for collecting end-user feedback. That is, for the manually generated process models, PM1 and PM2, the end-user feedback is collected by scrapping Facebook pages and by conducting a survey from the available students. Accordingly, a feedback corpus of 2000 feedback is generated for PM1 and another feedback corpus of 2742 feedback for PM2. During the feedback collection, the ethical considerations about the informed consent and anonymity proposed by [30] are ensured.

In contrast to the above, the end-user feedback corpora of the remaining four process models, PM3, PM4, PM5 and P6, are generated by scrapping user feedback from the Google play store. This is accomplished by using a python script which takes as input the identifier of a mobile app and scraps the latest 2000 user feedback from the Google play store.

B. DATA CLEANING

The raw corpora generated from the diverse sources were screened to elicit the data cleaning requirements. It was observed that the scrapped feedback include redundant information, such as identifiers, numeric scores, usernames, etc. For anonymity, the identification information is omitted and the textual content was separated. Similarly, the identifier and the numeric score were also omitted to separate the cleaned text. It was also observed that the feedback contained non-English content which was identified by employing a lookup based approach using the WordNet dictionary. As the last step, a lookup approach was used for performing the spelling corrections.

Accordingly, six end-user feedback corpora, C1 - C6, are generated. The feedback corpora on process model PM1 comprises of 2000 feedback, PM2 comprises of 2741 feedback, whereas PM3, PM4, PM5 and PM6 comprise of 802, 1334, 1471 and 1664 feedback, respectively.

VI. ELEMENT-FEEDBACK MAPPING BENCHMARK

A key contribution is the development of a benchmark in which the user feedback is mapped to the elements of process models, i.e. correspondences between the feedback corpus and the respective elements of the process model are defined. The developed benchmark can be useful in a number of ways. For instance, it can be used to evaluate the effectiveness of an automatic technique for finding correspondences between textual feedback and process model elements. Also, the developed benchmark can be used by the supervised learning techniques for learning and prediction.

For the development of the benchmark, all the N sentences from the end-user feedback are paired with the M activities of the process model to generate $N \times M$ pairs of user feedback comments and elements of process models. Due to this pairing, the correspondence between each feedback and all the element labels is assessed. That is, it is determined whether feedback and process element corresponds to each other or not. In order to ascertain the consistency of correspondences we have developed mapping guidelines. The details of the guidelines used for assessing the correspondences and the specifications of the development benchmark are presented in the following subsections.

A. CORRESPONDENCE ASSESSMENT GUIDELINES

It is widely acknowledged that the use of benchmark generation guidelines ensures the consistency of the annotations. In our case, the use of guidelines will ensure the consistent assessment of correspondence between end-user feedback and process models. Another benefit of using guidelines is that it reduces the effort involved in developing the benchmark. Furthermore, the use of guidelines improves the quality of the developed benchmark which is of higher significance for supervised learning techniques. Hence, we contend that the developed benchmark and the guidelines are valuable resources and they will serve as a catalyst for the automation of business process innovation.

For the development of the guidelines, a random collection of pairs was independently assessed by two human experts and the key observations were recorded. Subsequently, the correspondences were compared and conflicts were discussed. Also, the observation was synthesized to develop an initial set of guidelines. The process was repeated multiple times to develop a consensus about the observations. Correspondence assessment guidelines are designed based on these observations. The mutually finalized observations by the two experts are presented in Appendix A.

Essentially, there are two cases of correspondences between end-user feedback text and process model elements, *corresponds* and *doesn't-correspond*. Furthermore, we contend that the correspondences between feedback and process model elements can be at three levels: Syntactic, Semantic and Business Semantics. For each level, a high-level guideline and illustrative examples are defined. Table 3 presents the guidelines for assessing the correspondences and also includes an illustrative example for each case.

According to the guidelines, the correspondence is called *syntactic* mapping if a majority of the word tokens of a process model element or its root forms are also used in the user feedback. The correspondence is *semantic* if a majority of the word tokens in a process model element are semantically similar to the ones used in the feedback. Finally, the type of correspondence is *business semantics* if the word tokens of a process model element have a complex relationship with the tokens used in the user feedback. That is, the terms used in the two strings have the same business semantics.

For a further understanding, additional examples of the three types of mappings are presented in Table 4. It can be observed from the table that in the S1-E2 pair the element label text(''Document verification'') and the feedback phrase (''Document verification is a very lengthy process'') have common word tokens ''document'' and ''verification'', which makes it a syntactic mapping. In the example pair S2-E1, the label text (''Document verification'') and feedback sentence (''Your clerk is very lazy he took almost 10 minutes to check my documents'') have few words that are semantically similar, ''verify'' and ''check''. However, there are other word tokens including a pronoun, noun, preposition and adjective, that are used by users to express their feelings about process elements.

In the S3-E1 pair, the element label (''Visits Director office'') and the feedback sentence (''I went to the head of the department office)'' has complex relation (polysemy). Similarly, ''Head of department'' is related to the word token ''Director'' due to the polysemy nature of the word token. Therefore, it is categorized as business semantics. Similarly, in the example pair S3-E2 the token ''reached'' and ''appear'' has a cause/effect relation.

B. THE MAPPING BENCHMARK GENERATION

This section discusses the approach used to generate the Element-Feedback mapping benchmark which correspondences between the feedback corpus and the elements of a process model. Note that we have generated a benchmark for the six corpora C1 - C6 and the elements of the six process models PM1 - PM6 that we designed.

For generating the benchmark, N feedback of corpus C1 was paired with M labels of process model elements to generate $N \times M$ pairs. A human expert manually analyzed each pair to determine whether the correspondence in a pair exists or not, using the guidelines discussed earlier. If there is a syntactic, semantic or business semantic mapping, it is marked as mapped pair, else it is marked as un-mapped pair.

Note that a human expert manually annotated $N \times M$ pairs of feedback corpus C1 and process model PM1. It is important to note that the quality and consistency of the annotation are ensured by involving a second expert who randomly verified the assessments. Similarly, the Element-Feedback pairs of the remaining five corpora and the respective process models are manually assessed.

The details of the pairs generated for the six datasets, D1 to D6, are presented in Table 5. It can be observed from the table that the dataset D1 comprising of 2000 feedback comments paired with 38 labeled elements to form 76000 total pairs. Similarly, D2 is composed of 232,985 pairs, D3 has 51,328 pairs, D4 has 109,388 pairs, D5 has 123,564 pairs and D6 has 149,760 pairs. All these pairs are manually assessed by human experts. These large number of pairs represent the significant amount of effort involved in the development of the benchmark.

C. SPECIFICATIONS OF THE BENCHMARK

The specifications of the six benchmark datasets, D1 to D6 are presented in Table 5. It can be observed from the table that the dataset D1 comprises of 76, 000 pairs which include 588 corresponding and 75412 non-corresponding pairs. Similarly, D2 contains 1770 corresponding and 231215 non-corresponding pairs. Furthermore, D3, D4, D5 and D6 contain 231, 406, 259 and 337 corresponding pairs, respectively, whereas the remaining are non-corresponding pairs. These benchmark datasets are publicly available [31].

It can be observed from the table that there is a very high imbalance in the count of corresponding and non-corresponding pairs in the dataset D1. Furthermore, the imbalance exists in all six benchmark datasets. There are two key reasons for this imbalance. Firstly, the feedback corpora contain a large number of feedback sentences, whereas most of the feedback is generic and they are about the overall process. The second reason for the imbalance stems from the fact that we generated a cross-product of Element-Feedback pairs to be exhaustive in our approach to finding correspondences.

TABLE 3. Guidelines for mapping end-user feedback to process model elements.

Types	Examples	Mapping guidelines
Syntactic	*Element label:* Receives product. *Feedback:* He Receives product.	*Guideline 1:* A feedback is mapped to the process element label if all the word tokens of the feedback in exact same linguistic form are present in process element label text. Although feedback text may contain extra word tokens
Semantic	*Element label:* Receives product. *Feedback:* My maid received the product it was good.	*Guideline 2:* A feedback is mapped to the process element label and word tokens of the feedback are semantically similar or related. Although feedback text may contain extra word tokens.
Business Semantics	*Element label:* Complete application. *Feedback:* It was easy to fill the form, It took just 10 minutes.	*Guideline 3:* A feedback is mapped to process element labels if word token or combination of word tokens in feedback text have complex lexicon relation like hyponym, polysemy, cause/effect and business domain specific relation with word token or combination of word tokens in process element label text.

TABLE 4. Examples of feedback guidelines.

Type	Example Id	Example
Syntactic	S1-E1	Process element label: Document verification Feedback: Document verification is a very lengthy process. Mapped Category: Syntactic
	S1-E2	Process element label: Service Resumed Feedback: Service resumed in few minutes Mapped Category: Syntactic
Semantic	S2-E1	Process element label: Document verification Feedback: Your clerk is very lazy he took almost 10 minutes to check my documents Mapped Category: Semantic
	S2-E2	Process element label: Service Resumed Feedback: Its so annoying that I repeatedly checked your service and it took hours to resume. Mapped Category: Semantic
Business Semantic	S3-E1	Process element label: Visits Director office Feedback: I went to the head of the department office. Mapped Category: Business Semantic
	S3-E2	Process element label: Appear in exam Feedback: I reached the examination hall on time. Mapped Category: Business Semantic
	S3-E3	Process Element Label: Pays cash Feedback: I paid 200 rupees for the product. Mapped Category: Business Semantic

TABLE 5. Specifications of the benchmark datasets.

Dataset	Name	Elements Count	Feedback Count	Total Pairs	Corresponding Pairs	Non-Corresponding Pairs
D1	PM1	38	2000	76,000	588	75,412
D2	PM2	85	2741	232,985	1770	231,215
D3	PM3	64	802	51,328	231	51,091
D4	PM4	82	1334	109,388	406	108,982
D5	PM5	84	1471	123,564	259	123,305
D6	PM6	90	1664	149,760	337	149,423

VII. AugProMo: THE PROPOSED APPROACH

In this study, we have proposed a novel technique for the automatic identification of mappings between user feedback and process model elements to augment process models. The proposed approach is a unique combination of machine learning, data balancing and word representation techniques. A key feature of the approach is that it relies on generating sentence-level word embeddings and uses them as features for learning and prediction of mappings. The reason for the choice of word embeddings stems from the fact that they have achieved groundbreaking results for various natural language processing tasks. For instance, [32] has established that word embeddings are useful for named entity recognition in text. Furthermore, the effectiveness of word embeddings has been proven in the BPM domain for process model matching [24].

Figure 3 presents an overview of our proposed approach. It can be observed from the figure that the input to the proposed approach is a set of Element-Feedback pairs, whereas the output is the set of predictions about correspondences or non-correspondences in Element-Feedback pairs. According to the approach, firstly, feature vectors are generated which are to be fed to machine learning. Secondly, a data balancing approach is used to address the data imbalance problem, Finally, training and predictions are performed on the balanced dataset. The details of the feature vector generation, data balancing, machine learning model training and prediction are discussed in the following subsections.

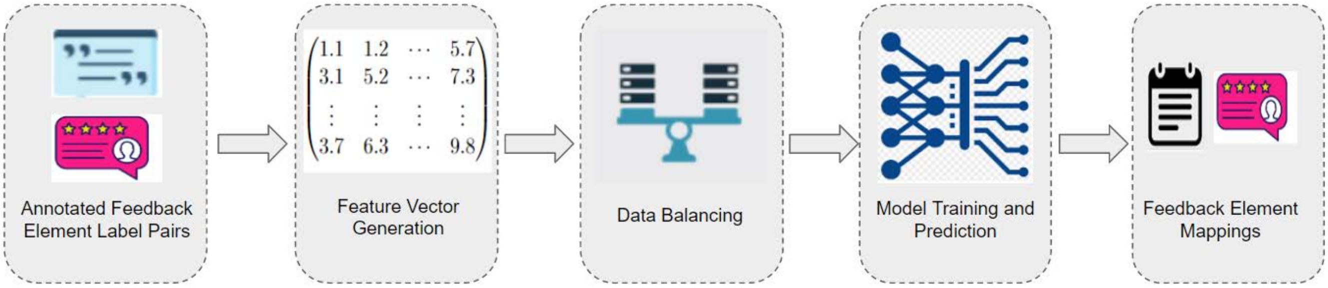

FIGURE 3. Overview of the proposed approach.

A. FEATURE VECTOR GENERATION

The generation of feature vectors in a multi-phase process. It includes tokenization, word level embeddings generation, sentence level embeddings generation and vector operations to generate the features to be fed to machine learning techniques. Figure 4 shows the steps for generating the feature vectors.

- Tokenization: Feedback R_{text} and process element label L_{text} are tokenized using the Gensim library's built-in tokenizer to generate two lists of tokens R_{tokens} and L_{tokens}.
- Word level embedding generation: For each token in R_{tokens} and L_{tokens} word-level embedding vectors are generated to produce collections of word embedding vectors W_R and W_L.
- Sentence-level embedding generation: All embedding vectors in W_R are averaged to generate a sentence-level embedding vector, $\overrightarrow{R}$. Likewise, all embedding vectors in W_L are averaged to generate a sentence-level embedding vector, $\overrightarrow{L}$.
- Vector operations: Finally, vector operations (sum, subtract, concatenation, union and intersect) are performed on $\overrightarrow{R}$ and $\overrightarrow{L}$ to generate five feature vectors. In particular, Union and Intersect are performed on 1-dimensional vectors. Whereas, for the N-dimensional vectors, $\overrightarrow{L}$ and $\overrightarrow{R}$, are converted to 1-dimensional $\overrightarrow{L}$ and $\overrightarrow{R}$ vectors. This conversion is performed keeping in view the complexity of Union and Intersect operations for N-dimensional vectors. The feature vectors and the details of their operations are presented in Table 6.

B. DATA BALANCING

The feedback in natural language offers the flexibility to express their sentiments or expression about the overall process or a certain fragment of the process. Therefore, the feedback includes a large number of comments that do not correspond to any process element. This is also endorsed by the differences between the corresponding and non-corresponding pairs presented in Table 5. The presence of an imbalance in the dataset impedes the learning abilities of machine learning techniques and ultimately impedes their prediction accuracy. To that end, the proposed approach employs a data balancing approach.

TABLE 6. Feature vectors and their operations.

No.	Feature Vector	Representation
1	$\overrightarrow{F}_{sum}$	$\overrightarrow{F}_{sum} = \overrightarrow{R} + \overrightarrow{L}$
2	$\overrightarrow{F}_{sub}$	$\overrightarrow{F}_{sub} = \overrightarrow{R} - \overrightarrow{L}$
3	$\overrightarrow{F}_{con}$	$\overrightarrow{F}_{con} = \overrightarrow{R} \parallel \overrightarrow{L}$
4	$\overrightarrow{F}_{union}$	$\overrightarrow{F}_{union} = \overrightarrow{R} \cup \overrightarrow{L}$
5	$\overrightarrow{F}_{intersect}$	$\overrightarrow{F}_{intersect} = \overrightarrow{R} \cap \overrightarrow{L}$

C. MODEL TRAINING AND PREDICTION

Finally, a machine learning technique is used for the training and prediction of the mappings. We chose machine learning models due to their ability to learn data patterns for predicting whether they a corresponding or non-corresponding pairs. Machine learning models would be trained on the feature vectors to evaluate the effectiveness of the developed human benchmark datasets.

VIII. EVALUATION

This study has performed a comprehensive evaluation of the proposed approach using a large number of experiments. That is, 2880 experiments are performed using a variety of word embeddings, data balance and machine learning techniques to identify the most suitable settings. In the following subsections, the details of these types of word embeddings, data balancing and machine learning techniques are discussed, followed by the details of the experimental setup.

A. TYPES OF WORD EMBEDDINGS

Experiments are performed using four types of word embeddings. The key reason for choosing these word embeddings is that they represent the state-of-the-art in the domain. Furthermore, these embeddings have achieved groundbreaking results for various NLP tasks. The four types of word embeddings used for experiments are Word2Vec, FastText, GloVe and BERT. Word2Vec is the collection bag-of-word and skip-gram based models used to generate the word embedding vectors [33]. FastText is a model for word embedding vector generation developed by Facebook AI [34]. GloVe

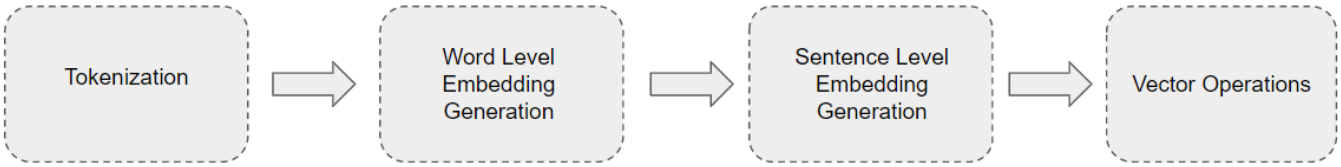

FIGURE 4. Approach for feature vector generation.

embedding vectors are generated based on the co-occurrence probability ratio between words [35]. And, BERT is transformer based deep learning model to generate contextual word embedding vectors [36].

B. DATA BALANCING TECHNIQUES

This study has used four diverse data balancing techniques, Random Over sample (RU), Random Under Sample (RD), NearMiss (NM) and SMOTE (SM). Where, Random Over Sample is a non-heuristic approach that randomly selects the minority class instances and over-sample them to resolve the imbalance problem [37]. In contrast, Random Under Sample is a non-heuristic method which removes random samples from the majority class [38]. NearMiss is a K-Nearest Neighbors(KNN) based down sampling method which discards K samples which are nearest to each other and retains fewer samples in data point cluster [39]. Finally, SMOTE deals with imbalanced datasets by over-sampling the borderline minority class instances [40].

C. MACHINE LEARNING TECHNIQUES

Six traditional machine learning techniques are used for the experimentation. It includes Logistical Regression (LR), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN) and Multi-layer Perceptron (MLP). LR is a supervised machine learning model which models the training feature vectors using the Sigmoid functions to predict the class as corresponding or non-corresponding [41]. NB is a classification algorithm which classifies feature vectors using probabilistic Bayes's theorem assuming feature vectors of corresponding or non-corresponding classes are independent of each other [42]. DT classifies feature vectors as corresponding or non-corresponding using multi-stage decision functions in a tree-like hierarchical manner [43]. RF uses multiple tree-like decision units which vote for the given random feature vector as corresponding or non-corresponding feature vector [44]. KNN classifier predicts the label of feature vector as corresponding or non-corresponding based on the label of the k most similar feature vectors in the learning data [45]. MLP is a deep neural network which consists of an input layer, output layer and multiple hidden layers [46]. Hidden layers in MLP learn using rectified Linear units ReLU as activation function then its output layer predicts the class as corresponding or non-corresponding for the input feature vectors.

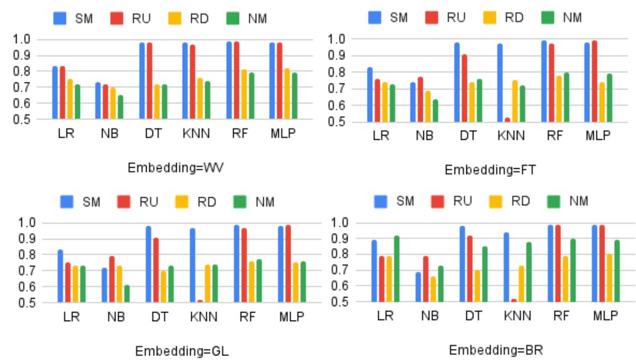

FIGURE 5. Comparison of balancing techniques of F_{Sub} for dataset D1.

D. EXPERIMENTAL SETUP

Experiments are performed for the six benchmark datasets that are discussed in an earlier section. In particular, we generated five types of feature vectors using the proposed approach and evaluated the effectiveness of five types of feature vectors using each of the four word-level embeddings and balancing techniques, as well as for all six machine learning techniques. That is, ($5 \times 4 \times 4 \times 6$) 480 experiments are performed using each of the six datasets, making the total number of experiments to (480×6) 2880 experiments.

For the experiments, each dataset was randomly divided into 10 equal parts, where 70% was used for training and the remaining 30% was used for testing. For each experiment, Precision, Recall and F1 scores were computed, however, due to space limitations only the results of the F1 scores are presented. Another reason for choosing the F1 score over Precision and Recall scores is that it is the harmonic mean of the two measures which provides a more pragmatic evaluation.

IX. RESULTS AND ANALYSIS

As discussed in the preceding section, 2880 experiments are performed to evaluate the most effective combination of features, data balancing and machine learning techniques which makes it challenging to analyze these many results. Therefore, the notable results are presented in this section and the remaining results are presented in the Appendix. Below, the key observations about the balancing techniques, machine learning techniques and most importantly the features vectors are discussed. Table 7 presents the results of experiments for the dataset D1.

TABLE 7. Results of the dataset D1.

	$\overrightarrow{F}_{Sum}$				$\overrightarrow{F}_{Sub}$				$\overrightarrow{F}_{Concat}$				$\overrightarrow{F}_{Union}$				$\overrightarrow{F}_{intersect}$			
Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
LR	0.82	0.81	0.81	0.76	0.83	0.83	0.83	0.89	0.84	0.84	0.84	0.76	0.69	0.69	0.71	0.74	0.58	0.41	0.51	0.75
NB	0.66	0.68	0.67	0.64	0.73	0.74	0.72	0.69	0.68	0.68	0.69	0.64	0.59	0.59	0.58	0.52	0.3	0.37	0.36	0.52
DT	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.77	0.96	0.47	0.45	0.77
KNN	0.98	0.97	0.97	0.74	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.74	0.97	0.96	0.97	0.75	0.96	0.46	0.45	0.75
RF	0.99	0.99	0.99	0.77	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.77	0.99	0.99	0.99	0.77	0.98	0.47	0.45	0.77
MLP	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.77	0.97	0.95	0.96	0.76	0.96	0.44	0.4	0.76

A. COMPARISON OF BALANCING TECHNIQUES

Figure 5 presents a comparison of the balancing techniques for $\overrightarrow{F}_{Sub}$ in the dataset D1. It can be observed from the figure that the F1 score achieved by SM and RU balancing techniques are higher than RD and NM for all types of embeddings and machine learning techniques. Furthermore, this observation is valid for a majority of the datasets. It represents that SM and RU are the two most effective data balancing techniques. It is due to the reason that these techniques over-sample the minority class instances, simply downsampling the majority class affects model learning and prediction for a majority of datasets. Furthermore, SM is a better choice than RU for four feature vectors $\overrightarrow{F}_{Sum}$, $\overrightarrow{F}_{Sub}$, $\overrightarrow{F}_{Concat}$, $\overrightarrow{F}_{Union}$. One possible explanation to the effectiveness of SM is that it is a heuristics-based technique which over-samples the borderline instances and performs better regardless of machine learning and word embedding techniques.

It can also be observed that the KNN machine learning technique when used with RU balancing produces the lowest F1 score. It is because, KNN classifies feature vectors based on their distance from the neighbouring feature vectors, whereas RU balancing technique randomly over-samples feature vectors which adversely affects the ability of the technique to predict the class based on the neighbouring feature vectors.

B. COMPARISON OF MACHINE LEARNING TECHNIQUES

For conciseness, the results of the SM balancing technique are discussed in this section, as SM is the most effective balancing technique. Figure 6 presents a comparison of machine learning techniques. It can be observed from the results that DT, KNN, RF, and MLP achieved a very high F1 score of 0.9. Furthermore, these techniques outperformed LR and NB for all the feature vectors. It is due to the reason that our feature vectors are multidimensional and complex in nature, whereas KNN is a non-parametric model which performs better with a large number of features. DT and RF are tree-based machine learning models that are suitable for non-linear complex features. In contrast, MLP is a deep learning model which is effective for large and complex datasets. From the results, we concluded that DT, KNN, RF and MLP are the most effective machine learning techniques for our proposed approach.

Note, that the performance of LR and NB is relatively below-par for a majority of datasets. It is due to two reasons: firstly, LR is useful for linear datasets, whereas our feature vectors are multidimensional. Secondly, NB is a probabilistic model based on independence assumptions between the features, whereas feature vectors may have complex dependence.

C. COMPARISON OF THE FEATURES VECTORS

To establish the effectiveness of the proposed approach, a detailed analysis of the feature vectors is performed. Below, firstly, a comparison of the types of word embeddings is performed followed by a comparison of the types of feature vectors.

It can be observed from Figure 6 that the use of three types of embeddings WV, FT and GL achieved a very high F1 score of 0.9 for the D1 dataset. This indicates that WV, FT and GL are more effective than BR embeddings. A possible reason for the below-par performance of BR embeddings is that these embeddings are generated on a smaller corpus which impedes the effectiveness of representation. Similarly WV, FT and GL embeddings outperformed BR for a majority of the datasets.

F1 scores of $\overrightarrow{F}_{Sub}$ are relatively high regardless of the embeddings and machine learning techniques for the D1 dataset. Hence, $\overrightarrow{F}_{Sub}$ is the most effective feature vector and $\overrightarrow{F}_{Intersect}$ is the least effective feature vector, especially with embeddings GL and FT.

The feature vector $\overrightarrow{F}_{Intersect}$ generated using GL and FT embeddings have reduced information due to two factors, dimension reduction and the nature of the intersect operation, which affect the learning ability of models. The results presented in the appendix further endorse the observation. $\overrightarrow{F}_{Sub}$ is the most effective feature while $\overrightarrow{F}_{Intersect}$ is the least effective feature across a majority of the datasets.

X. APPLICATION AND OUTLOOK

It is widely acknowledged that a large number of organizations are embracing BPM [47]. Some of these organizations are at an early stage of maturity as they have merely modelled their business processes, while others are at a higher level of maturity as they have implemented the complete BPM life-cycle. In order to better understand the effort involved in benefiting from the proposed approach, the organization needs to be classified into three types depending upon their preparedness for benefiting from the proposed approach.

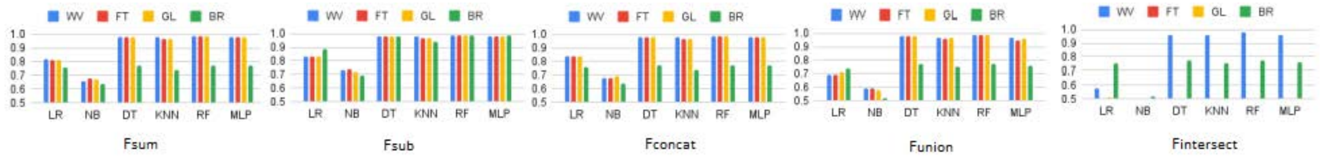

FIGURE 6. Comparison of feature vectors for dataset D1.

TABLE 8. Results of the dataset D1.

		$\vec{F}_{Sum}$				$\vec{F}_{Sub}$				$\vec{F}_{Concat}$				$\vec{F}_{Union}$				$\vec{F}_{intersect}$			
Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
LR	SM	0.82	0.81	0.81	0.76	0.83	0.83	0.83	0.89	0.84	0.84	0.84	0.76	0.69	0.69	0.71	0.74	0.58	0.41	0.51	0.75
	RU	0.81	0.74	0.74	0.72	0.83	0.76	0.75	0.79	0.84	0.78	0.78	0.72	0.71	0.63	0.63	0.7	0.41	0.63	0.63	0.7
	RD	0.76	0.75	0.73	0.69	0.75	0.74	0.73	0.79	0.76	0.76	0.74	0.69	0.63	0.66	0.68	0.66	0.54	0.35	0.48	0.66
	NM	0.75	0.77	0.76	0.73	0.72	0.73	0.73	0.92	0.71	0.79	0.78	0.81	0.66	0.54	0.63	0.66	0.74	0.35	0.48	0.66
NB	SM	0.66	0.68	0.67	0.64	0.73	0.74	0.72	0.69	0.68	0.68	0.69	0.64	0.59	0.59	0.58	0.52	0.3	0.37	0.36	0.52
	RU	0.67	0.88	0.83	0.87	0.72	0.77	0.79	0.79	0.68	0.89	0.84	0.87	0.52	0.92	0.94	0.63	0.3	0.99	0.99	0.63
	RD	0.66	0.64	0.69	0.61	0.7	0.69	0.73	0.66	0.63	0.64	0.67	0.61	0.6	0.63	0.6	0.54	0.3	0.35	0.36	0.54
	NM	0.72	0.69	0.67	0.64	0.65	0.64	0.61	0.73	0.7	0.66	0.68	0.71	0.59	0.52	0.61	0.58	0.54	0.46	0.47	0.58
DT	SM	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.77	0.96	0.47	0.45	0.77
	RU	0.98	0.93	0.92	0.99	0.98	0.91	0.91	0.92	0.98	0.99	0.99	0.99	0.73	0.99	0.99	0.99	0.44	0.99	0.99	0.99
	RD	0.74	0.74	0.71	0.71	0.72	0.74	0.7	0.7	0.74	0.74	0.75	0.71	0.67	0.73	0.68	0.71	0.88	0.38	0.4	0.71
	NM	0.72	0.78	0.77	0.73	0.72	0.76	0.73	0.85	0.72	0.74	0.75	0.81	0.98	0.72	0.7	0.75	0.59	0.46	0.47	0.75
KNN	SM	0.98	0.97	0.97	0.74	0.98	0.97	0.97	0.94	0.98	0.97	0.97	0.74	0.97	0.96	0.97	0.75	0.96	0.46	0.45	0.75
	RU	0.97	0.53	0.51	0.33	0.97	0.53	0.52	0.52	0.97	0.54	0.54	0.33	0.67	0.37	0.45	0.46	0.45	0.34	0.47	0.46
	RD	0.76	0.76	0.75	0.71	0.76	0.75	0.74	0.73	0.77	0.76	0.75	0.71	0.71	0.74	0.68	0.71	0.6	0.39	0.38	0.71
	NM	0.74	0.75	0.71	0.77	0.74	0.72	0.74	0.88	0.74	0.77	0.75	0.81	0.98	0.71	0.69	0.62	0.89	0.46	0.47	0.62
RF	SM	0.99	0.99	0.99	0.77	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.77	0.99	0.99	0.99	0.77	0.98	0.47	0.45	0.77
	RU	0.99	0.98	0.97	0.99	0.99	0.97	0.97	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.45	0.99	0.99	0.99
	RD	0.8	0.77	0.76	0.71	0.81	0.78	0.76	0.79	0.8	0.79	0.77	0.71	0.71	0.76	0.73	0.71	0.63	0.39	0.4	0.69
	NM	0.79	0.81	0.8	0.73	0.79	0.8	0.77	0.9	0.76	0.79	0.81	0.81	0.75	0.71	0.74	0.75	0.91	0.46	0.47	0.75
MLP	SM	0.98	0.98	0.98	0.77	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.77	0.97	0.95	0.96	0.76	0.96	0.44	0.4	0.76
	RU	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.97	0.98	0.99	0.96	0.39	0.99	0.99	0.96
	RD	0.8	0.79	0.75	0.74	0.82	0.74	0.75	0.8	0.82	0.77	0.79	0.74	0.67	0.67	0.67	0.62	0.54	0.35	0.36	0.62
	NM	0.79	0.78	0.78	0.77	0.79	0.79	0.76	0.89	0.72	0.77	0.79	0.86	0.71	0.64	0.69	0.5	0.88	0.45	0.47	0.51

TABLE 9. Results of the dataset D2.

		$\vec{F}_{Sum}$				$\vec{F}_{Sub}$				$\vec{F}_{Concat}$				$\vec{F}_{Union}$				$\vec{F}_{intersect}$			
Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
LR	SM	0.8	0.78	0.78	0.8	0.8	0.79	0.79	0.87	0.82	0.82	0.81	0.8	0.73	0.71	0.7	0.75	0.62	0.39	0.38	0.76
	RU	0.73	0.72	0.73	0.79	0.74	0.74	0.74	0.81	0.76	0.76	0.76	0.8	0.67	0.66	0.65	0.69	0.59	0.25	0.25	0.69
	RD	0.71	0.73	0.74	0.79	0.75	0.74	0.74	0.82	0.75	0.75	0.76	0.79	0.66	0.66	0.67	0.74	0.58	0.33	0.36	0.74
	NM	0.81	0.77	0.79	0.79	0.79	0.81	0.76	0.86	0.79	0.81	0.78	0.88	0.7	0.69	0.72	0.75	0.78	0.32	0.32	0.75
NB	SM	0.59	0.61	0.61	0.63	0.59	0.6	0.62	0.66	0.7	0.68	0.68	0.65	0.59	0.56	0.54	0.56	0.5	0.35	0.33	0.56
	RU	0.91	0.88	0.85	0.9	0.89	0.82	0.82	0.82	0.94	0.9	0.9	0.91	0.9	0.74	0.75	0.72	0.98	0.76	0.99	0.72
	RD	0.61	0.62	0.61	0.63	0.61	0.6	0.59	0.66	0.62	0.64	0.62	0.63	0.63	0.62	0.59	0.54	0.34	0.34	0.33	0.54
	NM	0.68	0.65	0.66	0.58	0.71	0.68	0.65	0.62	0.71	0.67	0.68	0.62	0.6	0.61	0.63	0.53	0.5	0.41	0.33	0.53
DT	SM	0.98	0.98	0.98	0.81	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.81	0.98	0.98	0.98	0.81	0.95	0.39	0.49	0.81
	RU	0.98	0.97	0.97	0.99	0.97	0.97	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.33	0.99	0.99
	RD	0.72	0.71	0.71	0.79	0.7	0.68	0.72	0.79	0.72	0.7	0.73	0.79	0.67	0.68	0.7	0.79	0.62	0.39	0.4	0.8
	NM	0.84	0.84	0.81	0.79	0.84	0.84	0.84	0.73	0.83	0.82	0.82	0.88	0.8	0.78	0.72	0.77	0.9	0.32	0.4	0.77
KNN	SM	0.97	0.97	0.97	0.81	0.97	0.97	0.97	0.94	0.97	0.97	0.97	0.65	0.97	0.96	0.96	0.72	0.94	0.44	0.38	0.71
	RU	0.4	0.41	0.4	0.99	0.4	0.41	0.4	0.41	0.46	0.45	0.46	0.33	0.38	0.36	0.43	0.5	0.51	0.99	0.37	0.5
	RD	0.73	0.72	0.74	0.79	0.72	0.7	0.73	0.84	0.75	0.72	0.75	0.77	0.65	0.74	0.75	0.77	0.57	0.39	0.36	0.77
	NM	0.79	0.78	0.76	0.79	0.8	0.78	0.76	0.81	0.78	0.75	0.75	0.86	0.82	0.76	0.76	0.72	0.9	0.41	0.4	0.72
RF	SM	0.98	0.98	0.98	0.81	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.81	0.99	0.99	0.99	0.81	0.96	0.44	0.43	0.81
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.8	0.78	0.8	0.8	0.79	0.79	0.78	0.84	0.79	0.8	0.8	0.79	0.75	0.77	0.77	0.79	0.64	0.39	0.41	0.79
	NM	0.87	0.85	0.84	0.79	0.87	0.86	0.86	0.82	0.86	0.84	0.83	0.88	0.84	0.81	0.8	0.77	0.9	0.41	0.4	0.77
MLP	SM	0.97	0.97	0.97	0.81	0.97	0.97	0.97	0.99	0.97	0.97	0.97	0.81	0.96	0.94	0.95	0.81	0.9	0.42	0.4	0.81
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99
	RD	0.79	0.73	0.76	0.79	0.79	0.76	0.76	0.85	0.79	0.74	0.76	0.79	0.67	0.71	0.73	0.64	0.57	0.42	0.32	0.64
	NM	0.86	0.85	0.83	0.79	0.86	0.85	0.85	0.85	0.86	0.84	0.84	0.88	0.83	0.73	0.79	0.66	0.86	0.41	0.39	0.66

A. TYPE A

These are the organizations that have modelled their business processes and also have some mechanism in place to gather end-user feedback about their processes or services. The mechanism for collecting feedback could be in the form of social media pages, feedback campaigns, or any other mechanism for recording end-user feedback. The organizations that have speech-based feedback systems in place may

TABLE 10. Results of the dataset D3.

		$\vec{F}_{Sum}$				$\vec{F}_{Sub}$				$\vec{F}_{Concat}$				$\vec{F}_{Union}$				$\vec{F}_{intersect}$			
Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
LR	SM	0.9	0.9	0.91	0.86	0.91	0.91	0.91	0.95	0.91	0.92	0.92	0.86	0.81	0.8	0.83	0.84	0.74	0.32	0.32	0.84
	RU	0.82	0.83	0.82	0.81	0.84	0.84	0.84	0.88	0.87	0.87	0.87	0.82	0.72	0.71	0.75	0.74	0.68	0.66	0.66	0.74
	RD	0.84	0.82	0.81	0.8	0.84	0.83	0.84	0.84	0.87	0.82	0.83	0.79	0.73	0.73	0.71	0.76	0.73	0.32	0.32	0.76
	NM	0.72	0.72	0.66	0.81	0.62	0.66	0.64	0.85	0.71	0.74	0.7	0.93	0.67	0.7	0.68	0.6	0.72	0.32	0.32	0.61
NB	SM	0.74	0.73	0.73	0.7	0.76	0.76	0.77	0.74	0.74	0.75	0.75	0.68	0.69	0.69	0.7	0.64	0.35	0.33	0.66	0.64
	RU	0.87	0.8	0.82	0.92	0.79	0.76	0.76	0.81	0.69	0.69	0.7	0.93	0.71	0.76	0.76	0.73	0.99	0.99	0.99	0.73
	RD	0.75	0.69	0.66	0.64	0.75	0.74	0.68	0.7	0.73	0.67	0.66	0.64	0.71	0.67	0.59	0.64	0.36	0.34	0.32	0.64
	NM	0.7	0.64	0.59	0.78	0.69	0.69	0.64	0.78	0.7	0.67	0.68	0.93	0.52	0.6	0.66	0.58	0.88	0.44	0.33	0.58
DT	SM	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.88	0.96	0.47	0.45	0.88
	RU	0.96	0.96	0.95	0.99	0.96	0.96	0.95	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.81	0.75	0.73	0.78	0.87	0.75	0.82	0.8	0.83	0.8	0.76	0.74	0.75	0.79	0.76	0.79	0.69	0.46	0.43	0.79
	NM	0.7	0.69	0.64	0.79	0.64	0.68	0.69	0.72	0.7	0.68	0.65	0.92	0.74	0.75	0.67	0.68	0.91	0.44	0.33	0.68
KNN	SM	0.98	0.98	0.98	0.88	0.98	0.98	0.99	0.98	0.99	0.98	0.99	0.88	0.99	0.99	0.99	0.87	0.95	0.46	0.45	0.87
	RU	0.5	0.57	0.57	0.41	0.57	0.58	0.58	0.56	0.64	0.66	0.64	0.41	0.49	0.44	0.5	0.49	0.38	0.33	0.49	0.49
	RD	0.79	0.74	0.75	0.73	0.84	0.69	0.78	0.76	0.83	0.7	0.78	0.73	0.65	0.73	0.68	0.79	0.63	0.45	0.43	0.79
	NM	0.65	0.59	0.64	0.77	0.6	0.62	0.62	0.75	0.64	0.58	0.59	0.93	0.71	0.72	0.72	0.64	0.91	0.32	0.33	0.64
RF	SM	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.82	0.96	0.46	0.45	0.87
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.9	0.83	0.83	0.8	0.9	0.82	0.87	0.82	0.85	0.83	0.88	0.79	0.79	0.78	0.79	0.79	0.72	0.46	0.43	0.79
	NM	0.73	0.7	0.69	0.81	0.71	0.7	0.72	0.76	0.71	0.71	0.72	0.93	0.75	0.79	0.72	0.68	0.91	0.44	0.33	0.68
MLP	SM	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.88	0.99	0.97	0.99	0.87	0.94	0.46	0.44	0.87
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.99	0.99	0.98
	RD	0.86	0.84	0.8	0.78	0.88	0.83	0.8	0.87	0.88	0.81	0.81	0.79	0.72	0.72	0.76	0.64	0.74	0.32	0.46	0.64
	NM	0.72	0.74	0.71	0.81	0.75	0.7	0.72	0.83	0.74	0.74	0.72	0.93	0.8	0.77	0.77	0.45	0.87	0.34	0.33	0.45

TABLE 11. Results of the dataset D4.

		$\vec{F}_{Sum}$				$\vec{F}_{Sub}$				$\vec{F}_{Concat}$				$\vec{F}_{Union}$				$\vec{F}_{intersect}$			
Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
LR	SM	0.87	0.88	0.89	0.8	0.87	0.88	0.88	0.92	0.88	0.89	0.89	0.8	0.74	0.76	0.78	0.75	0.7	0.33	0.32	0.77
	RU	0.77	0.79	0.77	0.77	0.75	0.77	0.75	0.83	0.81	0.82	0.81	0.76	0.68	0.69	0.69	0.6	0.65	0.66	0.67	0.61
	RD	0.78	0.8	0.8	0.78	0.73	0.74	0.79	0.8	0.7	0.8	0.8	0.77	0.7	0.73	0.69	0.7	0.6	0.32	0.31	0.7
	NM	0.71	0.72	0.76	0.74	0.68	0.73	0.75	0.8	0.74	0.74	0.83	0.77	0.69	0.76	0.74	0.64	0.71	0.32	0.32	0.64
NB	SM	0.66	0.68	0.68	0.68	0.69	0.7	0.7	0.72	0.73	0.74	0.74	0.7	0.65	0.67	0.67	0.5	0.3	0.33	0.33	0.52
	RU	0.71	0.66	0.69	0.85	0.71	0.64	0.63	0.67	0.65	0.66	0.67	0.86	0.67	0.69	0.68	0.66	0.9	0.99	0.99	0.66
	RD	0.62	0.69	0.68	0.67	0.62	0.69	0.67	0.65	0.68	0.71	0.7	0.67	0.6	0.71	0.68	0.45	0.32	0.32	0.34	0.45
	NM	0.69	0.6	0.71	0.65	0.69	0.6	0.71	0.64	0.73	0.71	0.73	0.71	0.68	0.67	0.63	0.5	0.91	0.45	0.4	0.5
DT	SM	0.99	0.99	0.99	0.83	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.83	0.99	0.99	0.99	0.83	0.96	0.45	0.45	0.83
	RU	0.95	0.96	0.95	0.99	0.95	0.95	0.94	0.95	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.71	0.7	0.79	0.78	0.75	0.79	0.75	0.76	0.74	0.71	0.74	0.78	0.7	0.72	0.8	0.78	0.62	0.42	0.42	0.78
	NM	0.78	0.81	0.83	0.74	0.81	0.81	0.81	0.73	0.8	0.78	0.83	0.77	0.84	0.79	0.81	0.7	0.94	0.45	0.4	0.7
KNN	SM	0.98	0.98	0.98	0.78	0.98	0.98	0.98	0.97	0.98	0.98	0.98	0.77	0.99	0.98	0.98	0.75	0.95	0.41	0.45	0.79
	RU	0.45	0.42	0.43	0.42	0.45	0.43	0.44	0.45	0.49	0.48	0.49	0.33	0.39	0.33	0.5	0.42	0.33	0.33	0.41	0.42
	RD	0.7	0.8	0.75	0.7	0.68	0.8	0.73	0.79	0.69	0.83	0.76	0.7	0.63	0.72	0.75	0.7	0.6	0.42	0.33	0.7
	NM	0.71	0.72	0.74	0.7	0.74	0.7	0.69	0.78	0.69	0.7	0.77	0.73	0.8	0.72	0.81	0.66	0.95	0.24	0.4	0.66
RF	SM	0.99	0.99	0.99	0.8	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.83	0.99	0.99	0.99	0.83	0.97	0.45	0.45	0.83
	RU	0.99	0.99	0.93	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.79	0.8	0.83	0.77	0.81	0.82	0.85	0.84	0.81	0.83	0.85	0.78	0.7	0.79	0.75	0.78	0.67	0.42	0.42	0.77
	NM	0.83	0.85	0.86	0.74	0.83	0.81	0.83	0.8	0.82	0.82	0.85	0.77	0.84	0.84	0.84	0.7	0.95	0.45	0.4	0.7
MLP	SM	0.98	0.99	0.98	0.82	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.82	0.98	0.98	0.98	0.83	0.93	0.43	0.44	0.82
	RU	0.99	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.94	0.99	0.99	0.99
	RD	0.76	0.82	0.83	0.78	0.78	0.79	0.8	0.81	0.78	0.81	0.78	0.78	0.72	0.74	0.76	0.78	0.68	0.41	0.44	0.53
	NM	0.81	0.81	0.84	0.74	0.84	0.8	0.86	0.8	0.84	0.8	0.83	0.76	0.82	0.81	0.8	0.7	0.91	0.45	0.32	0.66

also benefit from speech-to-text conversion to generate user feedback in the textual form. The organizations that fall in this class are at the higher stage of readiness as the two necessary resources, process models and end-user feedback, required for the proposed approach are readily available. In such cases, the effort is limited to generating the training dataset (gold standard annotations) for input to machine learning techniques. Given that generating a training dataset is a manual task which involves a substantial amount of effort, there are established mechanisms that can minimize the manual effort. One possible mechanism is to generate a seed benchmark and subsequently employ a bootstrapping approach for generating a large training dataset.

B. TYPE B

The organizations in this class have modeled their business processes but do not have any mechanism for collecting feedback about their processes or services. It is abundantly established that numerous organizations have modeled their business processes [48], [49]. These organizations are at an early stage of readiness as the end-user feedback, which is a necessary resource for the proposed approach, is not

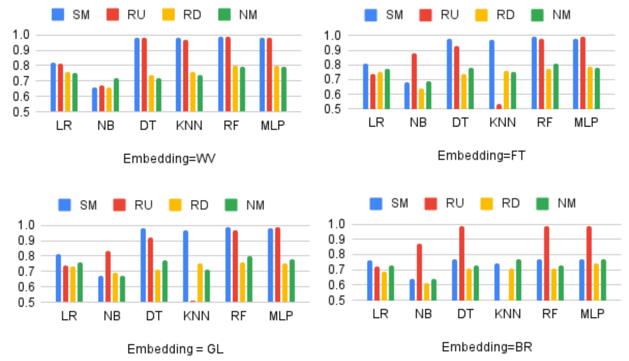

FIGURE 7. Comparison of F_{Sum} for dataset D1.

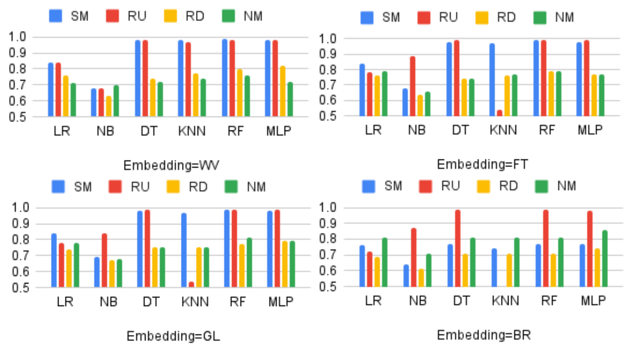

FIGURE 8. Comparison of F_{Concat} for dataset D1.

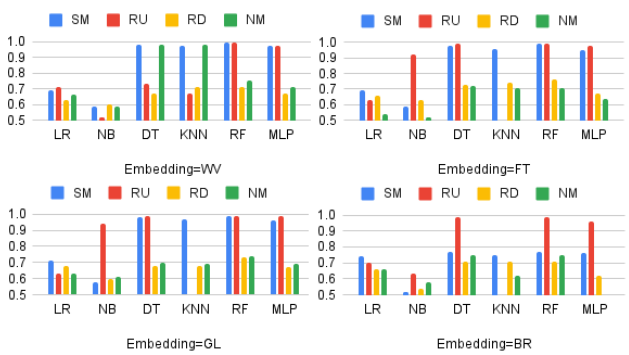

FIGURE 9. Comparison of F_{Union} for dataset D1.

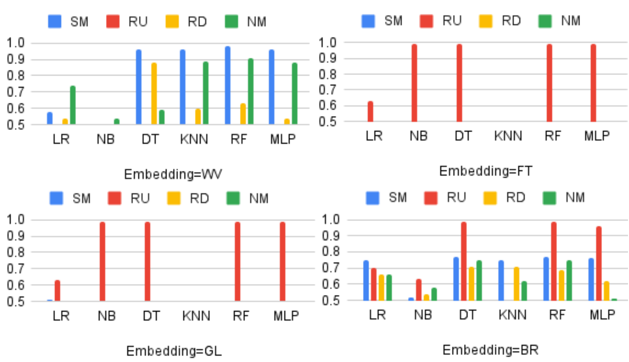

FIGURE 10. Comparison of $F_{Intersect}$ for dataset D1.

available. Hence, a substantial amount of effort along with the commitment of the management is a pre-requisite to benefit

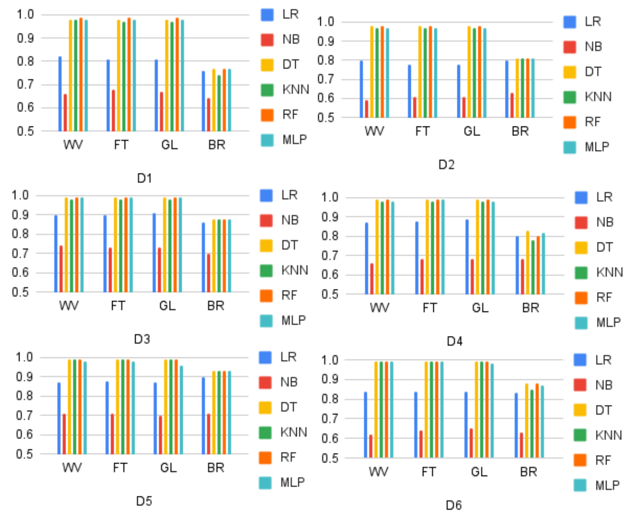

FIGURE 11. Comparison of machine learning techniques for F_{Sum}.

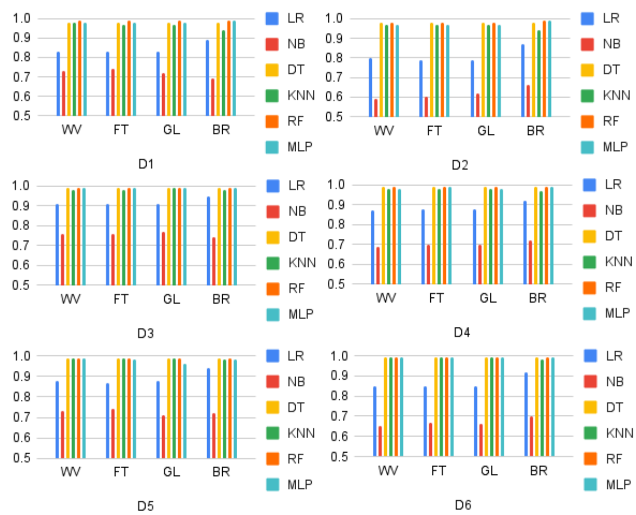

FIGURE 12. Comparison of machine learning techniques for F_{Sub}.

from the proposed approach. It includes setting a mechanism for collecting end-user feedback and the effort involved in the development of the benchmark that can be used for training machine learning techniques.

C. TYPE C

The organizations that have not modeled their business processes are classified as Type C. Typically, these are the organizations that are yet to embrace BPM. Such organizations require a significant amount of effort and management commitment for benefiting from the proposed approach, as it involves the development of all three types of resources, i.e., designing process models, employing a mechanism for collecting end-user feedback and gold standard benchmark to be used for training machine learning techniques.

XI. CONCLUSION

Business process management can become practically ambidextrous with digital innovation. End-user feedback

TABLE 12. Results of the dataset D5.

Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
			$\overrightarrow{F}_{Sum}$				$\overrightarrow{F}_{Sub}$				$\overrightarrow{F}_{Concat}$				$\overrightarrow{F}_{Union}$				$\overrightarrow{F}_{intersect}$		
LR	SM	0.87	0.88	0.87	0.9	0.88	0.87	0.88	0.94	0.88	0.88	0.88	0.9	0.78	0.82	0.79	0.82	0.56	0.35	0.35	0.81
	RU	0.81	0.82	0.82	0.86	0.83	0.82	0.82	0.87	0.84	0.84	0.84	0.87	0.61	0.67	0.67	0.7	0.51	0.69	0.71	0.7
	RD	0.8	0.83	0.78	0.91	0.81	0.82	0.8	0.86	0.81	0.83	0.78	0.91	0.6	0.65	0.71	0.77	0.49	0.32	0.32	0.77
	NM	0.8	0.83	0.77	0.9	0.81	0.82	0.76	0.93	0.8	0.82	0.77	0.95	0.83	0.81	0.8	0.8	0.32	0.32	0.32	0.8
NB	SM	0.71	0.71	0.7	0.71	0.73	0.74	0.71	0.72	0.65	0.65	0.64	0.71	0.59	0.6	0.6	0.54	0.39	0.33	0.33	0.55
	RU	0.65	0.67	0.73	0.92	0.63	0.66	0.69	0.66	0.55	0.54	0.54	0.93	0.56	0.59	0.6	0.56	0.35	0.99	0.99	0.56
	RD	0.63	0.71	0.69	0.65	0.63	0.71	0.7	0.69	0.58	0.56	0.61	0.65	0.52	0.57	0.62	0.54	0.58	0.34	0.34	0.54
	NM	0.83	0.85	0.77	0.72	0.8	0.8	0.73	0.88	0.87	0.88	0.83	0.82	0.67	0.87	0.8	0.61	0.85	0.38	0.36	0.61
DT	SM	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.93	0.86	0.39	0.39	0.93
	RU	0.96	0.97	0.97	0.99	0.97	0.97	0.97	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.999	0.99	0.99	0.99	0.99
	RD	0.76	0.78	0.79	0.9	0.77	0.74	0.72	0.83	0.67	0.76	0.71	0.9	0.7	0.78	0.72	0.89	0.57	0.37	0.38	0.89
	NM	0.88	0.9	0.87	0.9	0.9	0.92	0.78	0.88	0.9	0.91	0.8	0.95	0.85	0.91	0.81	0.9	0.89	0.38	0.36	0.9
KNN	SM	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.9	0.99	0.99	0.99	0.92	0.85	0.39	0.37	0.92
	RU	0.6	0.65	0.6	0.7	0.6	0.61	0.6	0.67	0.6	0.61	0.61	0.69	0.38	0.39	0.42	0.62	0.65	0.33	0.35	0.62
	RD	0.82	0.85	0.79	0.9	0.82	0.83	0.77	0.84	0.83	0.85	0.79	0.9	0.55	0.78	0.76	0.89	0.49	0.37	0.35	0.89
	NM	0.82	0.87	0.78	0.9	0.79	0.84	0.76	0.91	0.81	0.84	0.76	0.94	0.91	0.92	0.85	0.89	0.88	0.32	0.32	0.89
RF	SM	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.93	0.99	0.99	0.99	0.93	0.87	0.39	0.39	0.93
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.82	0.87	0.8	0.91	0.84	0.85	0.79	0.87	0.83	0.85	0.79	0.92	0.7	0.78	0.78	0.89	0.6	0.37	0.38	0.88
	NM	0.89	0.91	0.88	0.9	0.91	0.89	0.87	0.91	0.93	0.93	0.85	0.95	0.85	0.91	0.85	0.9	0.89	0.38	36	0.9
MLP	SM	0.98	0.98	0.96	0.93	0.99	0.98	0.96	0.98	0.99	0.99	0.99	0.92	0.98	0.98	0.98	0.92	0.81	0.38	0.99	0.92
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.38	0.99
	RD	0.76	0.83	0.8	0.92	0.8	0.83	0.8	0.85	0.8	0.8	0.76	0.92	0.66	0.81	0.73	0.65	0.53	0.32	0.36	0.65
	NM	0.9	0.91	0.85	0.9	0.9	0.89	0.84	0.94	0.91	0.9	0.83	0.95	0.9	0.87	0.87	0.82	0.81	0.32	0.32	0.82

TABLE 13. Results of the dataset D6.

Tech	B.Tech	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR	WV	FT	GL	BR
			$\overrightarrow{F}_{Sum}$				$\overrightarrow{F}_{Sub}$				$\overrightarrow{F}_{Concat}$				$\overrightarrow{F}_{Union}$				$\overrightarrow{F}_{intersect}$		
LR	SM	0.84	0.84	0.84	0.83	0.85	0.85	0.85	0.92	0.87	0.87	0.88	0.83	0.74	0.75	0.75	0.74	0.64	0.35	0.37	0.74
	RU	0.75	0.76	0.76	0.79	0.77	0.78	0.77	0.82	0.8	0.8	0.8	0.8	0.62	0.64	0.64	0.61	0.59	0.7	0.7	0.61
	RD	0.73	0.78	0.75	0.77	0.74	0.74	0.76	0.78	0.75	0.78	0.77	0.77	0.62	0.6	0.67	0.61	0.59	0.33	0.35	0.61
	NM	0.75	0.76	0.74	0.85	0.76	0.76	0.7	0.86	0.76	0.76	0.73	0.9	0.8	0.74	0.74	0.65	0.7	0.35	0.32	0.65
NB	SM	0.62	0.64	0.65	0.63	0.65	0.67	0.66	0.7	0.68	0.68	0.69	0.63	0.6	0.61	0.63	0.51	0.49	0.36	0.34	0.51
	RU	0.81	0.88	0.78	0.86	0.79	0.79	0.77	0.62	0.59	0.6	0.61	0.87	0.57	0.61	0.63	0.66	0.99	0.99	0.99	0.66
	RD	0.56	0.77	0.63	0.64	0.57	0.77	0.66	0.7	0.58	0.7	0.63	0.64	0.58	0.55	0.64	0.54	0.33	0.55	0.33	0.54
	NM	0.77	0.77	0.7	0.68	0.78	0.77	0.67	0.76	0.79	0.8	0.79	0.81	0.78	0.78	0.69	0.55	0.82	0.37	0.35	0.55
DT	SM	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.88	0.92	0.41	0.41	0.88
	RU	0.96	0.97	0.96	0.99	0.96	0.96	0.96	0.96	0.99	0.99	0.99	0.99	0.99	0.999	0.99	0.99	0.99	0.9	0.99	0.99
	RD	0.69	0.65	0.69	0.76	0.75	0.66	0.68	0.69	0.76	0.72	0.75	0.77	0.62	0.62	0.71	0.73	0.58	0.41	0.38	0.74
	NM	0.82	0.85	0.83	0.85	0.86	0.83	0.78	0.83	0.83	0.84	0.82	0.87	0.93	0.86	0.82	0.82	0.87	0.36	0.35	0.82
KNN	SM	0.99	0.99	0.99	0.85	0.99	0.99	0.99	0.98	0.99	0.99	0.99	0.85	0.99	0.98	0.99	0.85	0.91	0.36	0.34	0.85
	RU	0.54	0.542	0.54	0.58	0.54	0.54	0.54	0.56	0.63	0.56	0.57	0.57	0.36	0.36	0.44	0.59	0.47	0.34	0.41	0.59
	RD	0.72	0.74	0.69	0.7	0.72	0.74	0.68	0.74	0.73	0.74	0.67	0.7	0.59	0.66	0.65	0.72	0.54	0.33	0.38	0.72
	NM	0.77	0.77	0.75	0.83	0.76	0.77	0.7	0.82	0.75	0.77	0.75	0.88	0.9	0.83	0.81	0.79	0.88	0.36	0.33	0.79
RF	SM	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.88	0.99	0.99	0.99	0.88	0.92	0.41	0.41	0.88
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
	RD	0.75	0.76	0.75	0.78	0.78	0.75	0.77	0.81	0.76	0.76	0.78	0.77	0.69	0.66	0.69	0.74	0.61	0.41	0.38	0.73
	NM	0.88	0.86	0.87	0.85	0.87	0.87	0.79	0.9	0.88	0.88	0.87	0.91	0.92	0.84	0.8	0.82	0.88	0.36	0.35	0.82
MLP	SM	0.99	0.99	0.98	0.87	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.87	0.98	0.98	0.98	0.86	0.88	0.41	0.41	0.86
	RU	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.87	0.99	0.98	0.99	0.99	0.99	0.98
	RD	0.77	0.74	0.78	0.78	0.73	0.74	0.74	0.8	0.8	0.72	0.74	0.78	0.68	0.62	0.68	0.5	0.55	0.41	0.33	0.5
	NM	0.86	0.83	0.81	0.84	0.85	0.81	0.77	0.9	0.84	0.81	0.82	0.89	0.93	0.76	0.8	0.54	0.81	0.33	0.33	0.54

about a business process is recognized as a valuable asset for both exploratory and exploitative approaches for business process innovation. This study has proposed a novel concept of augmenting process models with end-user feedback by identifying correspondences between feedback and process model elements. These augmented models will facilitate the domain experts for both business process redesign and business process innovation initiatives. This study has proposed AugProMo which offers a step-by-step approach to augment process models with end-user feedback by determining correspondences between them. It employs a machine learning based approach coupled with state-of-the-art word embeddings, novel feature vector representation and data balancing techniques. The effectiveness of the AugProMo is evaluated through extensive experiments. The analysis of the results revealed that SMOTE is the most effective balancing technique. Traditional machine learning techniques (KNN, DT, RF, MLP) are very effective for identifying correspondence between end-user feedback and process model elements. Furthermore, $\overrightarrow{F}_{Sub}$ is the most effective feature vector as it achieved the highest F1 scores for all word embeddings and machine learning techniques across all datasets. This novel idea of augmenting feedback with process model elements has several future research directions.

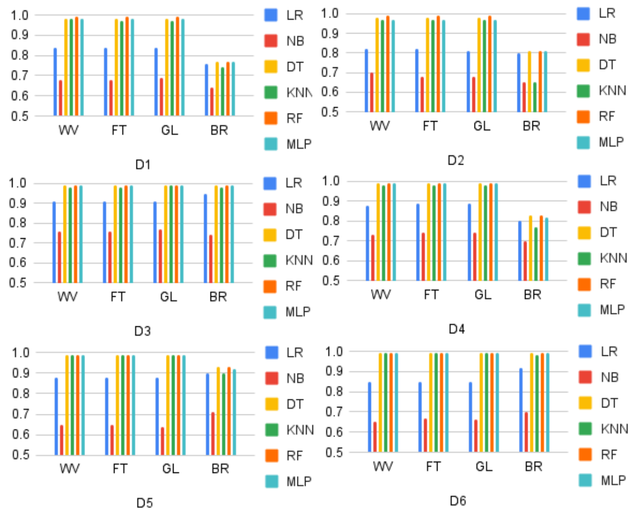

FIGURE 13. Comparison of machine learning techniques for F_{Concat}.

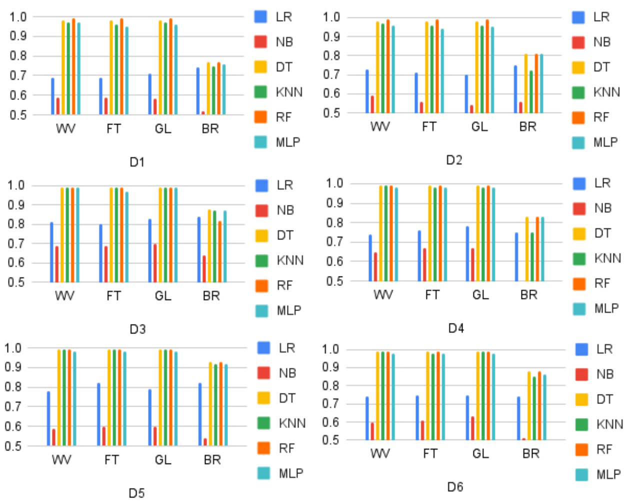

FIGURE 14. Comparison of machine learning techniques for F_{Union}.

- Develop NLP based approaches to rank the mapped feedback automatically based on their content.
- Perform sentiment analysis on these mapped feedback and process model element label pairs to develop a perception of the user feedback about a process model or its fragment.
- Lastly, to visually augment the mapped feedback to process model elements.

APPENDIX A CORRESPONDENCE ASSESSMENT GUIDELINES

Correspondence assessment guidelines are designed based on these observations.

- There are cases in which the vocabulary used in the user feedback is identical to the one used in the labels of process model elements.

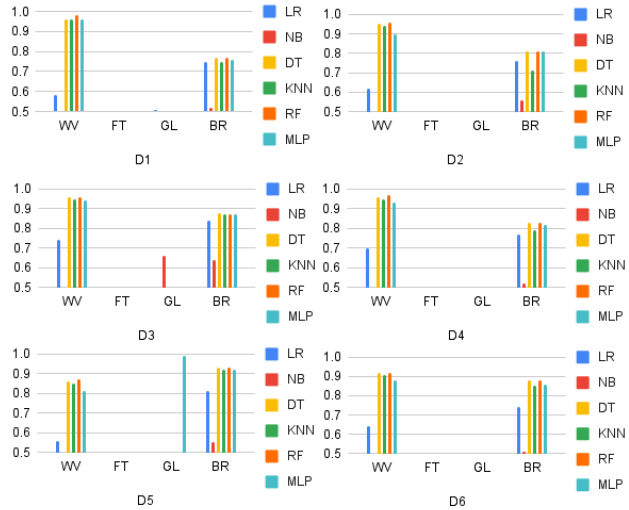

FIGURE 15. Comparison of machine learning techniques for $F_{Intersect}$.

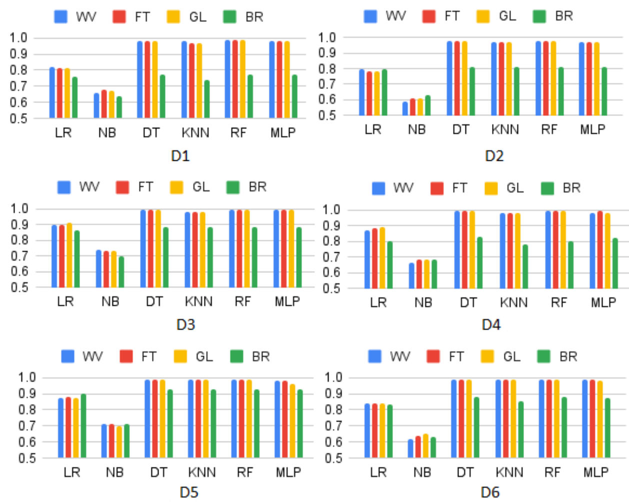

FIGURE 16. Comparison of word embeddings F_{Sum}.

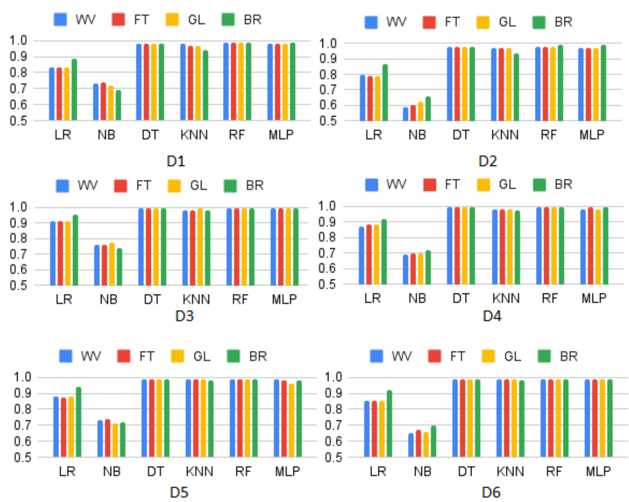

FIGURE 17. Comparison of word embeddings for F_{Sub}.

- In numerous cases, the vocabulary used for process model elements was different from the one used by the

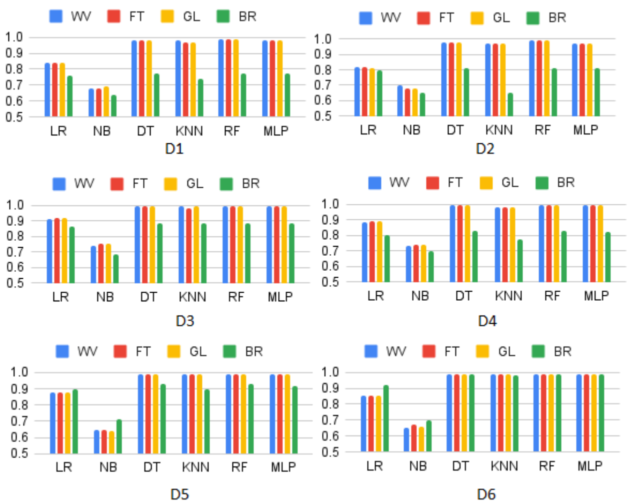

FIGURE 18. **Comparison of word embeddings for F_{Concat}.**

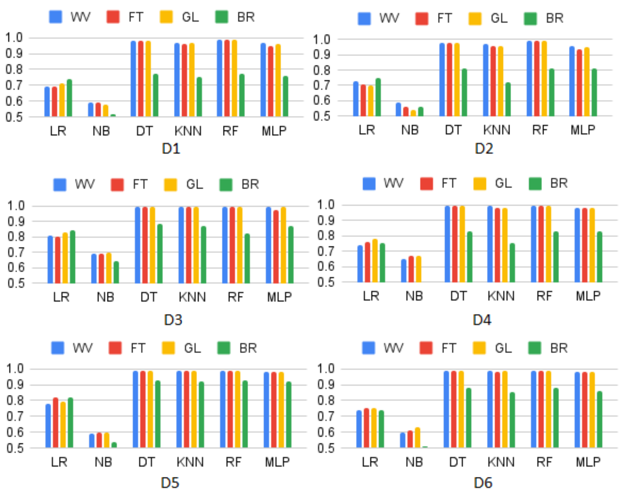

FIGURE 19. **Comparison of word embeddings for F_{Union}.**

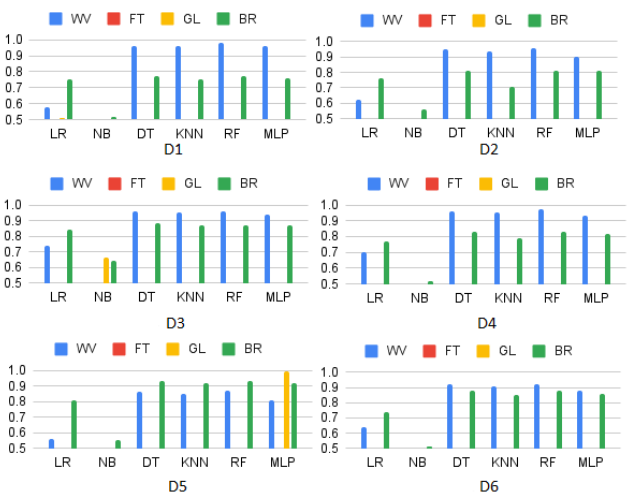

FIGURE 20. **Comparison of word embeddings for $F_{Intersect}$.**

end-users in their feedback, however they have the same semantics.

- There were several cases in which words used in the process labels have complex relations, such as cause - effect, hyponym and polysemy, with the word used in the user feedback.
- Also, the feedback text includes word tokens that describe sentiments of the end-users, whereas such expressions are not present in the labels of process model elements.
- Finally, the word count of the process elements varies between 3 and 5, whereas there is not restriction on the number of tokens in the end-user feedback.

REFERENCES

[1] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, "Introduction to business process management," in *Fundamentals of Business Process Management*. Springer, 2013, pp. 1–31.

[2] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*, vol. 1. Springer, 2013.

[3] M. Chinosi and A. Trombetta, "BPMN: An introduction to the standard," *Comput. Standards Interfaces*, vol. 34, no. 1, pp. 124–134, Jan. 2012.

[4] A. Van Looy, "How the COVID-19 pandemic can stimulate more radical business process improvements: Using the metaphor of a tree," *Knowl. Process Manag.*, vol. 28, no. 2, pp. 107–116, Apr. 2021.

[5] J. Mendling, B. T. Pentland, and J. Recker, "Building a complementary agenda for business process management and digital innovation," *Eur. J. Inf. Syst.*, vol. 29, no. 3, pp. 208–219, May 2020.

[6] P. J. Danaher and J. Mattsson, "Customer satisfaction during the service delivery process," *Eur. J. Marketing*, vol. 28, no. 5, pp. 5–16, May 1994.

[7] N. Chen, J. Lin, S. C. H. Hoi, X. Xiao, and B. Zhang, "AR-miner: Mining informative reviews for developers from mobile app marketplace," in *Proc. 36th Int. Conf. Softw. Eng.*, May 2014, pp. 767–778.

[8] S. Panichella, A. D. Sorbo, E. Guzman, C. A. Visaggio, G. Canfora, and H. C. Gall, "How can I improve my app? Classifying user reviews for software maintenance and evolution," in *Proc. IEEE Int. Conf. Softw. Maintenance Evol. (ICSME)*, Sep. 2015, pp. 281–290.

[9] A. Di Sorbo, S. Panichella, C. V. Alexandru, J. Shimagaki, C. A. Visaggio, G. Canfora, and H. C. Gall, "What would users change in my app? Summarizing app reviews for recommending software changes," in *Proc. 24th ACM SIGSOFT Int. Symp. Found. Softw. Eng.*, Nov. 2016, pp. 499–510.

[10] M. Rosemann, "Explorative process design patterns," in *Proc. Int. Conf. Bus. Process Manag.* Springer, 2020, pp. 349–367.

[11] A. Mustansir, K. Shahzad, and M. K. Malik, "Towards automatic business process redesign: An NLP based approach to extract redesign suggestions," *Automated Softw. Eng.*, vol. 29, no. 1, pp. 1–24, May 2022.

[12] Camunda Services. (2022). *BPMN Tool: Cawemo Camunda*. [Online]. Available: https://https://cawemo.com/

[13] A. Koschmider, K. Figl, and A. Schoknecht, "A comprehensive overview of visual design of process model element labels," in *Proc. Int. Conf. Bus. Process Manag.* Springer, 2016, pp. 571–582.

[14] H. Leopold, J. Mendling, and A. Polyvyanyy, "Generating natural language texts from business process models," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Springer, 2012, pp. 64–79.

[15] H. van der Aa, C. D. Ciccio, H. Leopold, and H. A. Reijers, "Extracting declarative process models from natural language," in *Proc. Int. Conf. Adv. Inf. Syst. Eng.* Springer, 2019, pp. 365–382.

[16] H. van der Aa, H. Leopold, and H. A. Reijers, "Checking process compliance against natural language specifications using behavioral spaces," *Inf. Syst.*, vol. 78, pp. 83–95, Nov. 2018.

[17] H. van der Aa, H. Leopold, and H. A. Reijers, "Comparing textual descriptions to process models—The automatic detection of inconsistencies," *Inf. Syst.*, vol. 64, pp. 447–460, Mar. 2017.

[18] H. Leopold, H. van der Aa, F. Pittke, M. Raffel, J. Mendling, and H. A. Reijers, "Searching textual and model-based process descriptions based on a unified data format," *Softw. Syst. Model.*, vol. 18, no. 2, pp. 1179–1194, Apr. 2019.

[19] M. Rana, K. Shahzad, R. M. A. Nawab, H. Leopold, and U. Babar, "A textual description based approach to process matching," in *Proc. IFIP Work. Conf. Pract. Enterprise Modeling*. Springer, 2016, pp. 194–208.

[20] G. Antunes, M. Bakhshandeh, J. Borbinha, J. Cardoso, S. Dadashnia, C. D. Francescomarino, M. Dragoni, P. Fettke, A. Gal, C. Ghidini, and P. Hake, *The Process Model Matching Contest 2015*, vol. 248. Berlin, Germany: Geellschaft für Informatik, 2015.
[21] U. Cayoglu, R. Dijkman, M. Dumas, P. Fettke, L. García-Bañuelos, P. Hake, C. Klinkmüller, H. Leopold, A. Ludwig, P. Loos, J. Mendling, A. Oberweis, A. Schoknecht, E. Sheetrit, T. Thaler, M. Ullrich, I. Weber, and M. Weidlich, "Report: The process model matching contest 2013," in *Proc. Int. Conf. Bus. Process Manag.* Springer, 2013, pp. 442–463.
[22] F. Jabeen, H. Leopold, and H. A. Reijers, "How to make process model matching work better? An analysis of current similarity measures," in *Proc. Int. Conf. Bus. Inf. Syst.* Springer, 2017, pp. 181–193.
[23] M. Ali, K. Shahzad, S. I. Muzaffar, and M. K. Malik, "Deep analysis of process model matching techniques," *IEEE Access*, vol. 8, pp. 99239–99253, 2020.
[24] K. Shahzad, S. Kanwal, K. Malik, F. Aslam, and M. Ali, "A word-embedding-based approach for accurate identification of corresponding activities," *Comput. Electr. Eng.*, vol. 78, pp. 218–229, Sep. 2019.
[25] J. Sànchez-Ferreres, H. van der Aa, J. Carmona, and L. Padró, "Aligning textual and model-based process descriptions," *Data Knowl. Eng.*, vol. 118, pp. 25–40, Nov. 2018.
[26] J. Sànchez-Ferreres, A. Burattin, J. Carmona, M. Montali, L. Padró, and L. Quishpi, "Unleashing textual descriptions of business processes," *Softw. Syst. Model.*, vol. 20, no. 6, pp. 2131–2153, Dec. 2021.
[27] A. Mustansir, K. Shahzad, S. I. Muzaffar, and K. Malik, "Utilizing customer feedback for business process performance analysis," in *Proc. Int. Conf. Bus. Informat. Res.* Springer, 2018, pp. 235–249.
[28] A. Mustansir, K. Shahzad, and M. K. Malik, "Sentiment analysis of user feedback on business processes," in *Proc. 36th IEEE/ACM Int. Conf. Automated Softw. Eng. Workshops (ASEW)*, Nov. 2021, pp. 204–210.
[29] S. Ahmed and A. Mustansir, "Mapping textual feedback to process model elements," in *Proc. 35th IEEE/ACM Int. Conf. Automated Softw. Eng. Workshops*, Sep. 2020, pp. 131–137.
[30] L. D. Roberts and P. J. Allen, "Exploring ethical issues associated with using online surveys in educational research," *Educ. Res. Eval.*, vol. 21, no. 2, pp. 95–108, Feb. 2015.
[31] K. S. S. Ahmed. (2022). *Revmap Datasets Repository*. [Online]. Available: https://www.kaggle.com/datasets/sanamahmed/revmapdatasets
[32] S. Kanwal, K. Malik, K. Shahzad, F. Aslam, and Z. Nawaz, "Urdu named entity recognition: Corpus generation and deep learning applications," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 19, no. 1, pp. 1–13, Jan. 2020.
[33] Y. Goldberg and O. Levy, "word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method," 2014, *arXiv:1402.3722*.
[34] I. Santos, N. Nedjah, and L. de Macedo Mourelle, "Sentiment analysis using convolutional neural network with fastText embeddings," in *Proc. IEEE Latin Amer. Conf. Comput. Intell. (LA-CCI)*, Nov. 2017, pp. 1–5.
[35] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
[36] G. Wiedemann, S. Remus, A. Chawla, and C. Biemann, "Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings," 2019, *arXiv:1909.10430*.
[37] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. ICML*, vol. 97. Princeton, NJ, USA: Citeseer, 1997, pp. 179–186.
[38] S. Kotsiantis and P. Pintelas, "Mixture of expert agents for handling imbalanced data sets," *Ann. Math., Comput. Teleinformatics*, vol. 1, no. 1, pp. 46–55, 2003.
[39] I. Mani and I. Zhang, "KNN approach to unbalanced data distributions: A case study involving information extraction," in *Proc. Workshop Learn. Imbalanced Datasets*, vol. 126, Washington, DC, USA, 2003, pp. 1–7.
[40] N. V. Chawla, "Data mining for imbalanced datasets: An overview," in *Data Mining and Knowledge Discovery Handbook*. 2009, pp. 875–886.
[41] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *J. Biomed. Inform.*, vol. 35, nos. 5–6, pp. 352–359, 2002.
[42] J. Ren, S. D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, "Naive Bayes classification of uncertain data," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 944–949.
[43] M. J. Aitkenhead, "A co-evolving decision tree classification method," *Expert Syst. Appl.*, vol. 34, no. 1, pp. 18–25, Jan. 2008.
[44] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.
[45] S. Zhang, X. Li, M. Zong, X. Zhu, and D. Cheng, "Learning *k* for kNN classification," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, pp. 1–19, 2017.
[46] Z.-X. Ye and Z.-H. Ling, "Multi-level matching and aggregation network for few-shot relation classification," 2019, *arXiv:1906.06678*.
[47] H. A. Reijers, S. V. Wijk, B. Mutschler, and M. Leurs, "BPM in practice: Who is doing what? in *Proc. Int. Conf. Bus. Process Manag.* Springer, 2010, pp. 45–60.
[48] J. Recker, M. Rosemann, M. Indulska, and P. Green, "Business process modeling—A comparative analysis," *J. Assoc. Inf. Syst.*, vol. 10, no. 4, p. 1, 2009.
[49] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, "Process-aware information systems," in *Fundamentals of Business Process Management*. Springer, 2018, pp. 341–369.

SANAM AHMED received the B.S. degree in IT and the M.Phil. degree in CS from the University of the Punjab, where she is currently pursuing the Ph.D. degree with the Department of Computer Science. She is a full-time Faculty Member at the Department of Software Engineering, University of the Punjab. She has industry experience of software development in silicon valley company for payment processing.

KHURRAM SHAHZAD received the master's and Ph.D. degrees from the KTH Royal Institute of Technology, Stockholm. He has been associated with the Information Systems Groups, Technical University Eindhoven, Eindhoven, and the University of Fribourg, Fribourg. He is currently a Tenured Associate Professor at the Department of Data Science, University of the Punjab. He has published more than 50 papers in international conferences and journals.