2022-11-16

Behavioral Mapping by Using NLP to Predict Individual Behaviors

Jafari, Reyhaneh

UNIVERSITY OF CALGARY

Behavioral Mapping by Using NLP to Predict Individual Behaviors

by

Reyhaneh Jafari

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN ELECTRICAL ENGINEERING

CALGARY, ALBERTA

NOVEMBER, 2022

Abstract

"What candidates or team members will do in specific circumstances" has always been an important piece of information for most employees or team leaders to consider when making a decision. It takes a significant amount of time to determine who is the best candidate for a particular job position. Companies are looking for the most efficient method for making this decision. Most of the time, personality assessments are used to identify an individual's character traits. Regardless of personality type, individuals will behave differently in a positive atmosphere than in a stressful one. Hence, characteristics alone can not predict behavior. Thus, text analysis and the identification of candidate behaviors (behaviorism) now enable companies to understand how people think, feel, and act in a given situation and then choose from a vast pool of candidates the best candidate for the job. By leveraging the existing intellectual property data associated with the behavioral mapping in AccuMatch Behavior Intelligence as well as expert data and using tools such as Amazon Comprehend Service (ACS), IBM Watson Natural Language Understanding (NLU), and Machine Learning (ML) techniques, various methods have been developed and analyzed in order to predict how individuals in a team become motivated, what their individual decision reference is, and what their execution style is. Therefore, this dissertation presents multiple proposed methods to predict Towards/Away, Internal/External, and Option/Procedure behavior and discusses the rationale behind the selection of these methods along with the results obtained.

Preface

For the degree of Master of Science at the University of Calgary, this thesis is submitted. The work presented here is an original work by Reyhaneh Jafari, supervised by Dr. Behrouz H. Far, except where references to previous work are made. This research was supported partly by Mitacs and AccuMatch Behavior Intelligence.

Part of this work has been presented in the following publication:

R. Jafari, B. H. Far, "Behavioral Mapping, Using NLP to Predict Individual Behavior,"

Acknowledgements

I would like to express my gratitude to my supervisor, Dr. Behrouz H. Far, for his encouragement, expertise, support, knowledge, patience, and kindness.

I would also like to thank Yousef Mehrdad Bibalan for his mentorship, wisdom, and expertise in using different online services; Mahmood Khalghollah for his insight and support; and Behnam Nikbakht for his patience in reading my thesis several times.

I am grateful to Nagui Bihelek (Founder and CEO of AccuMatch) and Sandra Bihelek (Master Trainer of AccuMatch), who explained the domain and provided good samples, and also for their useful suggestions.

I would like to extend my sincere appreciation to all my instructors, lab colleagues, and friends who have always tried to support me in difficult situations and who have supported this project.

Dedication

Dedicated to all team leaders who strive to assemble the best team for the task.

Table of Contents

List of Figures

List of Tables

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
ACS	Amazon Comprehend Service
NLU	Natural Language Understanding
NLP	Natural Language Processing
ML	Machine Learning
$MBTI$	Myers–Briggs Type Indicator
CV	Curriculum Vitae
LR	Logistic Regression
RFC	Random Forest Classifier
MNB	Multinomial Naive Bayes
GNB	Gaussian Naive Bayes
CNB	Complement Naive Bayes
ROC	Receiver operating characteristics
AUC	Area Under (ROC) Curve
TC	Text Classification
AWS	Amazon Web Service
$TF-IDF$	Term-Frequency times Inverse Document-Frequency
GUI	Graphical User Interface
DA	Data Augmentation
FSL	Few Shot Learning
$BERT$	Bidirectional Encoder Representations from Transformers

Chapter 1

Introduction

1.1 Background

Matching the right person with the job requirements is a challenge on most of the e-recruitment platforms. Organizing and planning may be necessary for one job, whereas another might require that a group of people collaborate to come to a decision [1]. While selecting the right candidate is a challenging task, hiring the right employees can have a lasting impact on a company's image [2]. According to Gamage's study, recruitment and selection are strongly related to the performance of small and medium-sized enterprises (SMEs) [3]. To increase the quality of employees, a number of studies have been conducted to assist organizations in making better hiring decisions using resumes, interviews, assessment centers, work samples, cognitive tests, and personality tests [4]. Several online assessment solutions (i.e., Humanic AI [5], Expert.ai [6], Crystal Knows [7], Receptiviti [8]) uses the target person's name, email address, LinkedIn profile ID, or social media account to find the publicly available social media text for that person, analyze it [9] [10], and produce scores on common personality scales (e.g. Big Five, Myers–Briggs Type Indicator (MBTI), Enneagram, DISC Assessment (Table 1.1 provides a brief description of these personality trait scales.))[11]. Previous approaches have some limitations. In the absence of using social media, it cannot be used as a framework for predicting whether the person in question is right for the job. Secondly, according to Novikov et al., the Big Five model, which has been mostly used for personality prediction, needs change. Because personality traits are not distinct from each other nor from demographic characteristics, and they can be affected by the instrument used [12]. Therefore, it is important to have a model that reflects both the personality of the candidate and is helpful and beneficial to the recruitment process [13]. Thirdly, although the MBTI is commonly employed in human resource management, it is widely used in research and counseling in higher education [14]. Based on Randall et al.'s research, the MBTI is less than 76% test-retest reliable (Test-retest reliability measures how consistent test scores are from one test administration to the next.), and they caution that the MBTI is reliable over time [15]. Lastly, for inferring personality traits reliably, the textual content of answers to standard interview questions can be used [16].

By asking behavioral interview questions, you can ensure responses that will offer insight into each candidate's natural strengths, blind spots, and motivators, all of which are essential when choosing the right person for the job [17]. However, one major criticism of the job interview is the possibility of bias introduced by the interviewer's prejudices [16]. Structured interviews where the same questions are asked of every candidate and evaluated using a well-defined rubric have been shown to reduce bias [18]. As a result, the textual content of AccuMatch Behavior Intelligence interview responses has been employed, which provides vital information for recognizing candidate particular behaviors such as motivation direction, decision reference, planning style, and focused attention, among others. This study focuses on behavior, which depicts how a person behaves in certain circumstances and responds to their environment as a whole. Others [19] [20] concentrate on personality, which refers to a person's intrinsic qualities that have been developed during their lifetime by heredity and the environment. In other words, examining behavior reveals what individuals do, but assessing personalities reveals their ideas, emotions, and motivations for behaving a specific way.

Table 1.1: Personality scales definition

Name	definition
Big Five	According to psychologists, a person's personality can be assessed on the basis of five major traits or dimensions, called the "Big Five personality traits." The acronym for these five traits is O.C.E.A.N., which stands for openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism [10].
MBTI	The MBTI is a self-report tool that is designed to identify a person's personality type. In MBTI, four dichotomies are used as indicators of personality traits. These four dichotomies are: introversion vs extroversion (I vs E), sensing vs intuition (S vs N), thinking vs feeling (T vs F), perception vs judging (P vs J) [21].
Enneagram	The Enneagram identifies nine fundamental personality types, and the Enneagram system of personality is designed to determine how closely an individual resembles each type. The nine types are: reformer, helper, achiever, individualist, investigator, loyalist, enthusiast, challenger, and peacemaker [22].
DISC	The DISC is a model that maps the personality of a person into assertive, openness, receptivity, and control types. An individual's perception of their surroundings and their ability to control them is the basis of this model [23].

1.2 Statement of Problem

Recruiting, or the selection of prospective employees from a vast pool of applicants, has always been a time-consuming issue for companies. In order to address this issue, many companies have started to use e-recruitment platforms. The use of online recruitment in SAT telecoms, for instance, led to a 44% cost reduction and an appreciable decrease in the hiring process [24]. Therefore, using online assessment tools will reduce the time it takes to find the best candidate. Analyzing answers to AccuMatch's online questionnaire will provide insight into some behaviors that will help organizations choose the most suitable candidate. In other words, AccuMatch looks at the answers to the questionnaire to see if the person is qualified for a certain job or not. One behavior that this work has been focusing on is whether the job requires someone to make good decisions and plan tasks, or whether it is acceptable to have someone who has trouble managing priorities because of how he/she reacts (Toward/Away). Another is: do they prefer a self-motivated person or a person who needs feedback (Internal/External) or do they need someone to seek options or follow a process (Option/Procedure)? These behavioral indices are beneficial not only to employers in finding the right applicant for the job but also to employees who want to be perceived as effective and exhibit these behaviors [1]. In this research, the aim is to provide answers to the following questions:

- Is it possible to predict the Towards/Away, Internal/External, and Option/Procedure behavioral indices automatically? In other words, is it possible to build an automatic classifier (the learner) that can inductively classify text without knowing much about the domain?
- Is there any algorithm to evaluate behavior on a scale of one to ten?

Having answered the first question, the second has been left for future research due to a lack of data.

1.3 Presumptions

AccuMatch Behavior Intelligence sponsored this research. Consequently, the aim of this research is to predict the behaviors defined by the company's experts. To do the research, several confidential data sources were used, including the candidates' responses to the company's questionnaire, as well as expert data. While the author tries to explain the domain and its benefits, its intention is to resolve the company's problem, which is to find an accurate way to predict behavior automatically. In order to achieve the solution, all methods of feature extraction have been confirmed by experts.

1.4 Research Objectives

The purpose of this research is to determine what candidates do (behavior analysis) rather than who they are (personality analysis). The behavior is analyzed from different angles in other research as well. In Furnham et al.'s work, they study how narcissism, psychopathy, and machiavellianism pertain to organizational behavior and success [25]. Using three dimensions of listening behavior—attention, perceptivity, and responsiveness—Ruyter and Wetzels analyzed the influence of moral behavior on the relationship between frontline employees and customers [26]. Román et al., recognized that the employee's ethical sales behavior, including fair play, honesty, and complete disclosure, impacts the customer's satisfaction [27]. Hence, various jobs require different types of behavior. In this work, different methodologies with same dataset have been proposed to determine which model best predicts the specific behaviors like the candidate motivation index (Towards/Away), decision reference index (Internal/External), and execution style index (Option/Procedure). In other words, the following methods have been compared in this study in order to determine which one is more effective in classifying behaviors. So, the end result of this research is a list of different methods that have been suggested and a comparison of them.

- 1. IBM sentiment analysis
- 2. IBM emotional analysis
- 3. Amazon comprehend service
- 4. Bag of words model (Towards/Away, Internal/External)
- 5. Vector-Base model
- 6. Breaking sentences (Option/Procedure).
- 7. Few shot learning (Towards/Away)

1.5 Thesis Outline

This thesis is organized as follows: The workflow for comparing the algorithms and the methods is outlined in chapter 2. Afterward, definitions of each behavior and some examples of behavior have been provided. Data collection and the results of each proposed method are discussed in chapter 3. Also, the work limitations have been explained in this chapter too. Chapter 4 presents the experimental results, conclusion, and future work.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

In recent years, much research has been conducted to remove irrelevant cases while matching applicants with relevant jobs. It is challenging for recruiters to predict the work behaviors of job candidates because there is limited information recruiters have at their disposal during the selection process, and those involved in the recruitment process are likely to perform well during the selection process, which can mislead them. As a result of a lack of data from applicants at first, using an online questionnaire to infer specific behaviors will be an alternative to traditional psychological testing. Therefore, the automatic assessment should be able to identify the behaviors like motivation (Towards/Away), decision reference (Internal/External), and execution style (Option/Procedure) of a person. Based on Pennebaker et al., a person's choice of words can reveal a great deal about their social status, age, gender, and motives [28]. Due to advances in computer technology, text analysis has become a powerful method of analyzing what people say as reliably and rapidly as possible [28]. Also, with the help of NLP, unstructured data can be turned into insights and make information easy to access and understand.

The following research has been done to predict the performance of applicants: Using demographic data such as age, gender, marital status, education background, and work experience, Chien and Chen were able to predict behaviors such as the work performance and retention of applicants [4]. According to Davies and Mcdonald, the combination of the Integrity tests (direct and indirect) and the General Mental Ability (GMA) test has the most predictive validity for the recruitment process in terms of predicting performance levels [29]. In their research, the direct test of integrity may include questions about attitudes towards dishonesty, beliefs about the prevalence of dishonesty, and, in some cases, direct questions about past dishonest behavior. While a person's indirect assessment attempts to uncover factors in their character that may underlie dishonest behavior.

The most common method of predicting an individual's personality is to review their curriculum vitae (CV). A system was developed by Menon et al. that allows the applicant to upload a CV and a Twitter account, after which a list of candidates is ranked by compatibility score and sent to the employer who posted the job offer [20]. A problem with this was inferred personality scores based on digital data are considerably lower than those calculated by examining distinct self-reported questionnaires. It is therefore unreasonable to expect that predicted personality traits will reproduce known relationships with life outcomes regularly [12]. To overcome this issue, Sudha et al. designed an online multiple-choice question (MCQ) test that included personality questions. The system determines professional eligibility by analyzing uploaded CV training datasets and quiz results. The final results of the personality quizzes will be given to both candidates and the administration [19]. Through personality analysis, employers can distinguish between individuals and understand the "why" behind employees' behaviors within a specific context, such as the workplace. On the other hand, employers do not need to examine their employees' personalities to understand how they approach risk in the workplace; they simply need to analyze their behavior and understand what the employees will do. Also, personality is vague and needs to be untangled and analyzed before it can be changed in any way. A psychiatrist is probably the best person to talk to about this.

2.2 Selection of tool to study

The primary purpose of this research is to determine which behavior index type the data belongs to in any dichotomous behavior. In other words, it has been assumed that this is a classification problem. For first six methods, to make a better prediction, different models have been applied which are Logistic Regression (LR), Random Forest Classifier (RFC), Multinomial Naive Bayes (MNB), Complement Naive Bayes (CNB) and Gaussian Naive Bayes (GNB). Since the goal is to classify the dichotomous (binary) variables; hence, LR is the appropriate model to analyze the data [30]. Also, in the case of text classification problems, linear classifiers are often considered to be strong baseline solutions. When the right features are used, they are often able to achieve state-of-the-art results in spite of their simplicity [31]. On the other hand, based on A. Alshamsi et al.'s research, RFC is one of the methods that performed well with unbalanced datasets [32]. In addition, naive bayes classifiers show promising results on text classification [33] and some psychological problems [9]. Furthermore, the CNB classifier is well suited to the current dataset, which is unbalanced. Because the single run of the predictive model may result in a noisy estimate of model performance, 5-fold cross-validation has been applied to estimate the model reliably. The process consists of repeating the cross-validation procedure multiple times and reporting the accuracy mean (The accuracy of predicting test data labels has been considered.), the area under the Receiver operating characteristics (ROC) curve (AUC), precision, recall, and F1-score across all runs. Since simple classification accuracy is seldom a good measure of performance [34] the AUC has been used to evaluate the classifiers, and it shows the most accurate and successful model with the best performance [35][36]. Hence, the final model has been chosen with highest AUC in each method. As mentioned in [37], text search systems are generally evaluated experimentally as opposed to analytically. To put it in another way, an experimental evaluation of a classifier is usually based on its ability to make the right classification decisions. Consequently, each method has been evaluated based on its prediction of unseen data to determine which classifier (All the parameters of classifiers have their default value, but to control the randomness of the sample and to produce the same results in each run, the random_state has been set to 0.) is most effective at predicting the right label. Figure 2.1 shows a summary of implemented procedure.

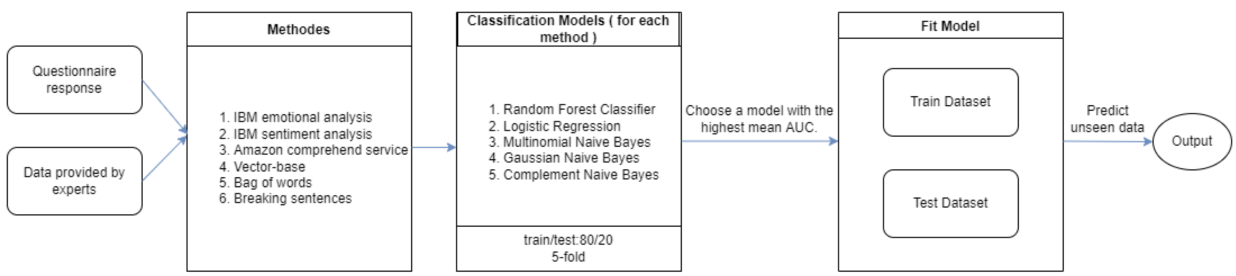

Figure 2.1: The workflow of the implemented procedure

Briefly, the inputs for all proposed methods are the candidate responses to the questionnaire and the data provided by the experts, and all five ML models were considered for each proposed method. For each proposed method, the ML model with the highest mean AUC has been fitted for a specific training and testing dataset. Ultimately, all proposed methods have been compared based on their false positives and false negatives.

2.3 IBM emotional analysis

Several IT companies, like Google, IBM, Amazon, and Microsoft, now offer AI technologies that can be used right away through their cloud platforms. Technologies such as these allow us to use powerful techniques to solve a wide range of problems. For instance, understanding how people feel, what they think, and who they are from what they write (e.g., extracting sentiment, emotions, and psychographics from text) [38]. IBM Watson NLU is a cloud-based platform that aims to infer emotions based on texts; its use cases include personal and business communications, market research, self-branding, and automated customer service agents [39]. It uses deep learning to extract emotions, sentiments, categories, classifications, entities, keywords, relations, and syntax from the written text. Emotions identified include anger, fear, joy, sadness, and disgust; identified social tendencies include the Big Five personality traits. Because IBM Watson NLU is capable of detecting emotions, as mentioned in [40], it can be considered an automatic classifier. Therefore, IBM Watson NLU has been used to predict behavioral indices based on emotions, which are features of our solution.

2.4 IBM sentiment analysis

Several studies have been conducted on emotion recognition and sentiment analysis based on social media text [41][42][43]. Essentially, the goal of these studies is to classify user or group attention, sentiment, or emotion towards a subject or without a specific target. A company can, for example, use user-generated content to predict and/or explain various aspects of its performance, such as sales, profits, brand perception, customer satisfaction, and market performance [44]. There is a triplet that describes a sentiment: ($\boldsymbol{y}$, $\boldsymbol{o}$, $\boldsymbol{i}$), where $\boldsymbol{y}$ describes the type of sentiment, $\boldsymbol{o}$ describes the sentiment's orientation, and $\boldsymbol{i}$ describes the sentiment's intensity [45]. An orientation (also known as polarity, tone, or semantic orientation) defines a level of granularity that can be at the document, sentence, or aspect level [46]. There are three types of sentiment: positive, negative, and neutral. Usually, neutrality refers to the absence of any sentiment. Additionally, the intensity of a sentiment can vary within the same sentiment polarity (e.g., the use of perfect vs. good) [44].

The purpose of this study is to analyze human behavior using text, which is a challenging task. In order to accomplish this, sentiment analysis has been combined with text classification (TC). In contrast to emotion, which captures feelings such as anger, sadness, joy, fear, disgust, and sadness, sentiment is a higher-level classification of emotions that is divided into positive, negative, and neutral groups. Thus, the highest sentiment score has been used as an input, and the model has to predict the behavior. In other words, the relationship between the dependent variable, which is a category such as Towards/Away (what we want to predict) and the independent variable, which is sentiment (the highest positive or negative sentiment) (the features), is measured. Because of having negative values in the dataset, CNB and MNB cannot be used, and the model has been chosen from the other three models, which are RFC, LR, and GNB. In this approach, a data with 60% negative and 40% positive sentiment was counted as -60% sentiment. Additionally, since the answers may consist of multiple sentences, the orientation is at the document level.

2.5 Amazon comprehend service

In order to utilize IBM Watson NLU, some data had to be deleted; because it does not accept data with a length of less than 3 and more than 2000. Furthermore, IBM is difficult to use. The ACS, which is an NLP service, has been considered to be one of the easiest APIs to deploy [47]. ACS uses machine learning to uncover patterns and relationships in text. It can even provide insight into the sentiments of the customers. It analyzes the text using tokenization and parts of speech and organizes a collection of texts by topic. The autoML capabilities of ACS have been employed to create a custom set of entities or TC models that are tailored specifically for this project [48]. According to Kaggle [49], cloud computing has been dominated by Amazon Web Services (AWS), which is the first player in terms of usage, and IBM is in last place. Due to having promising accuracy with IBM Watson NLU, I reasoned that the result of custom classification might be good.

2.6 Vector-base method

Since the task is TC and classifiers, or algorithms that construct classifiers, cannot interpret texts directly. In view of this, it is necessary to employ a uniform indexing procedure that maps a text into a compact representation of its content. There are two ways to accomplish this. The first is by associating terms with words. This is sometimes called the bag of words approach or the set of words approach to represent a document. The second approach involves computing term weights and term-frequency times inverse document-frequency (TF-IDF) can be used to represent text as a vector of weighted terms [37]. In other words, to transform a collection of text documents into numerical feature vectors, TF-IDF can be used. In order to classify the dataset, the TF-IDF has been applied, which aims to extract features from the text. By doing so, the text will be converted into feature vectors that can be processed while preserving its original information. In this approach, feature vectors are used as inputs and dichotomous behaviors as outputs for creating a classification model. Since, ML algorithms have been widely used in the prediction of psychological characteristics [50]. This research used a python-based ML library, Scikit-learn, that is simple, easy to use, and effective [51] for creating a classification model.

2.7 Using Bag of words method

One prominent method of quantitative text analysis focuses on word counts. Psychological word-counting strategies can be used to evaluate both content (what is being said) and style (how it is being said) [28]. Pennebaker, who is a social psychologist, discovers that the words people use reflect their psychological state. He realized that function words correlate with personality characteristics (e.g. pronouns (I, she, it), articles (a, an, the), prepositions (up, with), auxiliary verbs (is, don't), negations (no, never), conjunctions (but, and), quantifiers (few, most) and common adverbs (very, really).) [52]. Also, during meetings, experts usually mentioned certain words to identify certain behaviors. For instance, the pronoun "I" often appears in the sentences of people who are internal and are usually focused on themselves. Individuals who use "team" and "feedback" to measure their success are considered external. By providing a graphical user interface (GUI) [53] for the Towards/Away and Internal/External behavioral indices to the experts , the experts were able to assign each word or collocation to a behavioral index or even disgard the word. The system is able to save the assigned words to each behavioral index and then assess the data. So, the system is going to predict the dichotomous behaviors, and the features are word counts for each category in this method.

2.8 Breaking sentences method

To overcome the lack of data, the experts proposed to punctuate the questionnaire answers for the Option/Procedure behavior index and break the punctuated sentences. It has been decided to use two punctuators to ensure that sentences are punctuated more accurately. By doing so, the number of records was raised from 103 to 350. Then experts labelled the sentences. They assumed option sentences as 1, procedure sentences as 2, option and procedure sentences as 3, and meaningless sentences as 0.

After removing neutral and meaningless data, the model is going to predict the label for each sentence, whether it is an option or a procedure. Since the vector-base model has the best results for Option and Procedure, the features of this method are sentences, which have been converted to vectors.

2.9 Few shot learning method

The human brain can distinguish the sentiment of a text after viewing a few examples. However, computers require large amounts of data in order to classify what they "read" and differentiate between text sentiments. Furthermore, providing a large amount of data to computers has its own challenges. Firstly, it is important to note that the data that requires human labelling is by definition more difficult. Because when making decisions regarding labeling, humans are inclined to inject their own biases and judgments. Furthermore, the addition of a single labelled data does not have much impact on models since they are trained using batches of data. Moreover, as each data must be retrained until convergence, this can become a costly endeavor, especially when viewed in terms of performance improvement versus acquisition cost (time and money). A new model that is few-shot learning (FSL) has been developed to solve all the above problems.

By implementing it, computers will be able to learn from a few samples. In other words, FSL aims to bridge the gap between artificial intelligence and human learning. By incorporating prior knowledge, it can learn new tasks containing only a few examples with supervised information [54].

In this method, prior knowledge is provided by a popular embedding model, Bidirectional Encoder Representations from Transformers (BERT), and also other vector or glove models like **word2vecgooglenews300**, **glovewikigigaword50**, **glovetwitter25**, and also other pretrained models like **fasttextwikinewssubwords300**, and **conceptnetnumberbatch170630**. On top of that, a classifier is used to categorize the data. This proposed method has been implemented for the Towards/Away behavior index. The main difference between this method and the previous ones is the workflow. However, at the end, the results are evaluated based on their ability to predict the unseen data too. In other words, the model with a lower number of false positives and false negatives is considered a better model. Therefore, the number of misclassified data that is classified as Towards but is actually Away and vice versa has been determined. Consequently, in the conclusion part, all methods were assessed based on the number of incorrect predictions for unseen data, and those with the lowest number were considered to be the most effective.

2.10 Defining Behaviors

As a general definition, behavior can be summarized as follows:

Every experience we have is based on behavior. There are many ways in which our behavior can impact our world or others, whether it is our own behavior or observations of others' behavior that lead to judgments. All of our interactions and activities are characterized by our behavior. We may react to something we observe or take action in response. In either case, these are still behaviors. Non-action is also a form of behavior.

For the purposes of evaluating broad aspects of an employee's suitability for a given position, AccuMatch considered specific behaviors such as their ability to perform under stressful conditions, their time management skills, and so on. Six of the 52 behaviors will be the focus of this study. Below is a description of the focused behaviors.

2.10.1 Defining Towards / Away

The goal of the motivation index is to predict what the candidates' direction is towards the desired value (Toward) and what their direction is away from the undesirable value (Away from). Toward people, have trouble seeing what to avoid and are good at making choices and planning tasks. While Away people struggle to manage priorities because they are reacting to problems as they come up, and they are easily sidetracked by unpleasant situations. Table 2.1 shows some key factors and samples for Towards and Away index.

Table 2.1: Key to recognizing Towards or Away responses

Towards indicators	***Towards sample***	***Away indicators***	***Away sample***
Involve inclusion words	Being part of a franchise has its benefits	Involves Exclusion words	Without integrity, nothing works well
Stays focused on and is energized by goals	The final product is important to increased sales and reputation	Have trouble maintaining focus on goals and is energized by threats	I can't lose my client contracts
Tend to reach, achieve, get, attain	I feel comfortable, secure and encourage to always give extra	Tend to avoiding, not having, staying away, getting rid of	Minimizing stresses in both work and personal environments makes me more productive

The following questions have been asked to understand the applicant direction:

- What factors are important to you in a work environment? (name at least 3)
- Which of these factors is the most important to you? (your most important criteria)
- Why this factor is so important to you?
- And, what is it about this factor that makes it important to you?
- And how does this factor affect you when you have that?

2.10.2 Defining Internal / External

Predicting whether a candidate is Internal or External is intended to provide information about the candidate's decision-making process. Are they influenced by their own thoughts or do they turn to some external source? How do they determine what is right and appropriate? External people need feedback to know if they are doing well, whereas Internal people are confident about their level of performance because they

Table 2.2: Key to recognizing Internal or External responses

Internal indicators	***Internal sample***	***External indicators***	***External sample***
Believe they have the right to decide	I am the decision maker	No internal standard to gage from	I need to ask someone
Interpret "finger pointing" as someone else is deciding. Causes polarity	Don't tell me what to do	Comfortable learning from others experience	When my team members share their ideas with me
Closed to outside opinions	I am right	Open minded, accept others opinion	When I get positive feedback

internally carry all. External brain's circuits evaluate using external standards, whereas Internal brain's circuits maintain their own standards. Table 2.2 presents some key factors and samples relating to the Internal and External Index. The following question has been posed in order to understand the applicant's decision reference:

- How do you know you have done a good job in your role?

2.10.3 Defining Option/ Procedure

When a person is asked, "Why," it is necessary to consider the two planning-style behavioral indices, namely, the Procedure and the Option, to determine whether he/she will tell a story or give reasons. People who have Option behavior are effective at identifying new solutions, whereas people who have Procedure behavior are most effective at following a specific process. Options people enjoy starting new projects, whereas Procedure people are content to repeat the same tasks over and over again. Option people hear the word "Why" and provide a list of reasons, while Procedure people ignore the "why" and distort it into an answer to a "how" question. These questions were asked in order to determine the applicant's direction:

- Why did you choose the work that you do? OR
- Why did you choose the last job you had?

Chapter 3

METHODOLOGY

3.1 Data Collection

The Accumatch piloted its model in two different companies in 2015 and 2017. A total of 103 data sets have been evaluated on a scale of one to ten by experts. The data was unbalanced , and there was also not enough data to examine the model. Initially, there was 10 Away data and 93 Towards data, which was an unacceptably uneven balance for training. A data augmentation (DA) method, paraphrasing, has been used to fix the class imbalance that is undersampling in this domain [55]. In other words, to overcome a lack of training data and help improve the model's generalization capabilities and regularize the aim, the creation of artificial training data for machine learning by transformations using DA is used [56]. Hence, to increase the diversity of training data without directly collecting new data or balancing the distribution of classes, the DA was employed. Other scholars have also said that creating extra data may increase the quality of a solution (classifier) [57][58][59]. QuillBot, which is a state-of-the-art paraphrasing tool, has been used to generate output text data while preserving the meaning of input text with variations in word choice and grammar. It uses machine learning to rewrite phrases while retaining their original meaning [60]. A GUI has been developed so that experts can evaluate and assess data [61] and ensure that the data has been labelled appropriately. In other words, this GUI was used to perform the preprocessing step. A variety of technologies have been considered in the development of this GUI, including NodeJS for the backend, Angular (TypeScript) for the frontend, as well as Firestore Database (NoSQL) as a dataset. Through this GUI, the experts were able to search for a person and see their responses to each question (Figure 3.2), as well as the paraphrased versions of their responses. By this GUI (Figure 3.1), the experts were also able to select a question from the question list and see all the candidates' responses with their paraphrase.

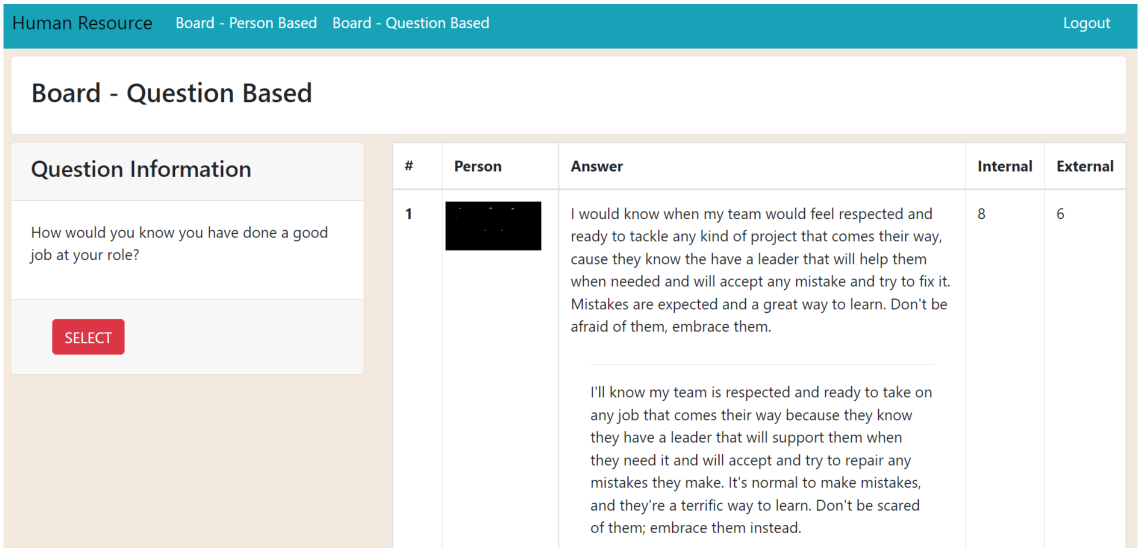

Figure 3.1: GUI for displaying all candidate responses and their paraphrased to a particular question

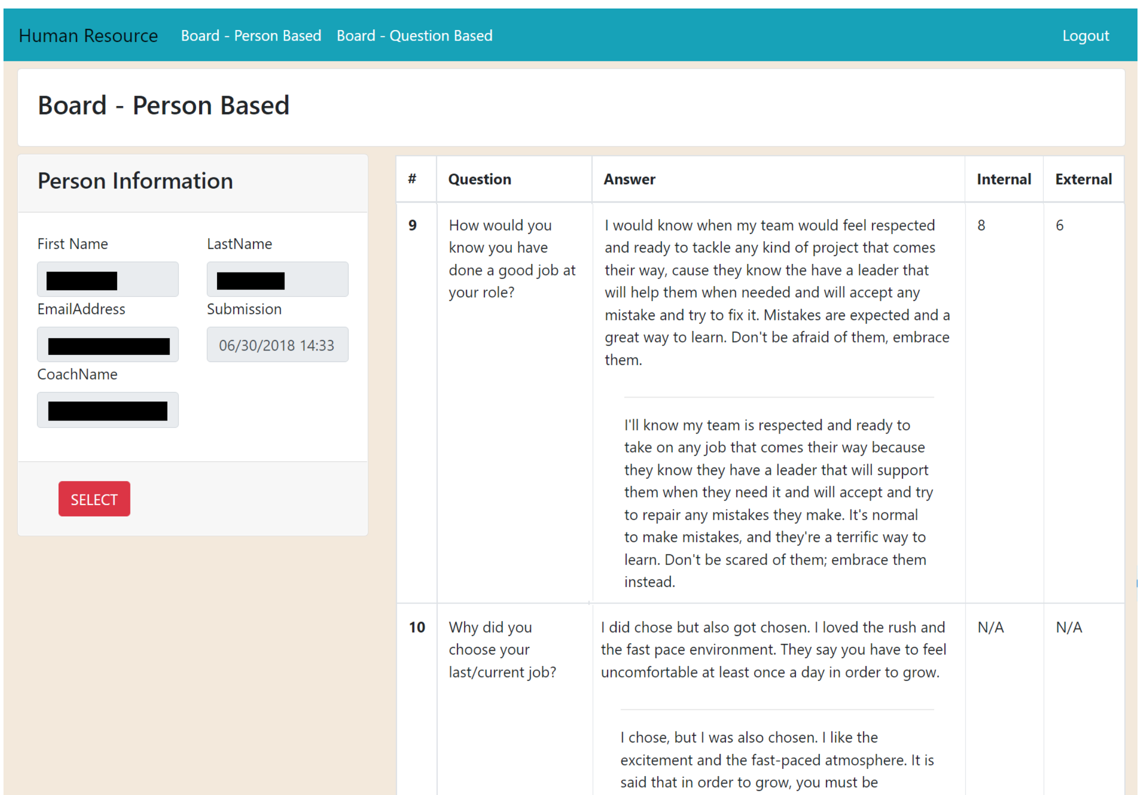

Figure 3.2: GUI for searching each person and viewing their responses with paraphrased responses

Typically, augmentation is used in computer vision. In this field, the creation of an augmented image is relatively straightforward. Even when noise is added or a portion of the image is cropped, a model can still classify the image. Due to the complexity of language, it is difficult to augment text in the field of NLP. Because it is not possible to replace every word with another and also, there are not all synonyms for every word. Thus, it can be challenging to define textual transformations that preserve labels in DA [56]. However, experts, who are master coaches and trainers at AccuMatch Behavior Intelligence, rejected the paraphrased data because, in their opinion, changing a word to its synonym changes the labelling significantly.

Finally, to balance classes in this low-resource domain, experts have been asked to provide new data. Because it takes a lot of time to ask and label the answers of new applicants, experts have created 106 new labelled data sets, which are answers to Towards/Away and Internal/External questions. They jointly answered the questions with one or more sentences. By doing so, they hoped to minimize bias since, if one expert answered a question, he/she would be revealing only their own behavior, but together they tried to provide different scenarios for each behavioral index. 106 new data have been created for Towards/Away and 84 new data have been created for Internal/External. With the addition of new data, the imbalance is still present but there are fewer differences (e.g., for Towards and Away is 25, it means there are 90 Towards 65 Away).

The dataset has been labelled based on expert evaluation. In other words, the behavior that has the highest score in dichotomous behavior is considered to be the label for that data. For example, data with a Towards score higher than an Away score has been categorised as Towards. In cases of equal scores, the data has been removed.

3.2 Analysis Model

This research is concerned with supervised classification problems in which the aim is to devise a method or construct a rule for assigning data to a dichotomous behavioral index. Thus, only two classes have been considered in this classification problem. After implementing the methods, machine learning models, which are at the core of many recent advances in science and technology, have been used. Apart from the accuracy of model predictions, there is also a need to evaluate the models. To make this decision, after five fold, the mean of the area under the ROC curve, which is a well-known measure of performance for the multiple class cases, has been used [62][63][64][65]. For visualizing classifiers, ROC graphs [34], which represent the relationship between the true positive rate TPR (or sensitivity) of the test and its false positive rate (FPR) (or 1–specificity)[63], have been shown for Towards/Away behavioral index.

For each behavioral index (except Option/Procedure), 20% of the total data was used to evaluate predictions (This data was not used for training or testing. So, it is unseen data). Firstly, because real data may be different and the evaluation metric may not accurately reflect the product's goals. Secondly, experts must be sure that the model will perform well on real-world data.

From section 3.3 to 3.8, all methods have been implemented using explained workflow. Each section begins with an explanation of the method, followed by an description of the results. A different approach has been used in section 3.9, which is the FSL approach. The main reason for employing this approach is that, since there are a few data, it has been assumed that an ML model can learn from prior knowledge to classify each response. Therefore, different pre-trained models have been used to evaluate the results. These models have been described as a subsection of this section, and then the results of this method have been provided.

3.3 IBM emotional analysis

With the IBM Watson NLU, meaning and metadata can be extracted from unstructured text data using deep learning. It can identify hidden meanings in the data by processing categories, classifications, entities, keywords, sentiments, emotions, relationships, and syntax [38]. In this approach to predicting the behavior indices, the emotions of sentences have been detected as features. According to [40], an emotion detector is an automatic classifier. Hence, to create an automatic classifier, the data has been gathered and the features, which are emotions, have been extracted by calling the IBM Watson NLU API, and then a classification model has been trained to recognize and classify specific patterns.

3.3.1 Results

IBM Watson NLU, as mentioned previously, has a limitation in that it does not accept data with a length of less than or equal to 3 and more than 2000. Therefore, the following 6 sentences have been deleted from the analysis dataset for Internal/External: "I just know", "Let me see", "I know", "I don't know", "I am a winner" and "I never lose". Also, one data with a length greater than 2000 for Towards and Away has been deleted from the dataset.

IBM Watson NLU considers the level of emotions for both the Internal and External behavioral index to be almost equal. The mean of sadness for Internal data is 0.23, but for External data it is 0.24, which corresponds to low accuracy as shown in Table 3.1. The AUC for Towards/Away and Internal/External is higher than accuracy, which indicates a lower true negative than true positive. For Option/Procedure, the total true positive is higher than the false negative, which is the reason behind having a high recall.

Unlike Towards/Away, where there were 112 differences in data lengths, the Internal/External lengths

Table 3.1: Results of applying IBM emotional analysis to different behavioral indices

Behavior index	Towards/Away	Internal/External	Option/Procedure
model	CNB	RFC	LR
accuracy	74%	50%	71%
F1	74%	50%	68%
Precision	74%	50%	72%
Recall	74%	50%	71%
AUC	81%	55%	58%
# wrong prediction on the Test Set	8 of 31	14 of 28	11 of 21
#Wrong prediction on Unseen Dataset	7 of 53	19 of 42	N.A.

were quite close. Internal was 30 and External was 32. In the case of Option/Procedure behavior, a dataset was not provided by the experts. Therefore, just the answers to the questionnaire was used for training and testing. One data had the same score for Option and Procedure, so it was deleted, and "I am not" was not processed by IBM. In total, there are 101 data in the dataset, 60 Options, and 41 Procedures. The mean length of Option is 36, while the mean length of Procedure is 40. The score of joy is higher than others among other emotions for the Option/Procedure behavior index. For Option people, it is 0.65, while for Procedure people, it is 0.52.

Unlike Towards/Away, where there were 112 differences in data lengths, the Internal/External lengths were quite close. Internal was 30 and External was 32. In the case of Option/Procedure behavior, a dataset was not provided by the experts. Therefore, just the answers to the questionnaire were used for training and testing. One data had the same score for Option and Procedure, so it was deleted, and "I am not" was not processed by IBM. In total, there are 101 data in the dataset, 60 Options, and 41 Procedures. The mean length of Option is 36, while the mean length of Procedure is 40. The score of joy is higher than others among other emotions for the Option/Procedure behavior index. For Option people, it is 0.65, while for Procedure people, it is 0.52.

The point-biserial correlation has been employed in order to determine the relationship between the category Towards or Away (binary variable) and the emotions. There is a negative correlation between category and joy of -0.63, which reflects the fact that, as the category increases from 1 to 2, the joy decreases. To put it in another way, the mean of joy emotion for Toward is 0.59 and for Away is 0.23. The highest positive correlation score is for sadness, which is 0.5, showing that Away data has a higher score for sadness emotion than Toward.

The highest positive correlation between emotions and Internal or External behavior is for sadness, which is 0.026. The sadness score for Internal is lower than the sadness score for External. Joy is correlated positively with this behavioral index, while other emotions (anger, fear, disgust) are negatively correlated.

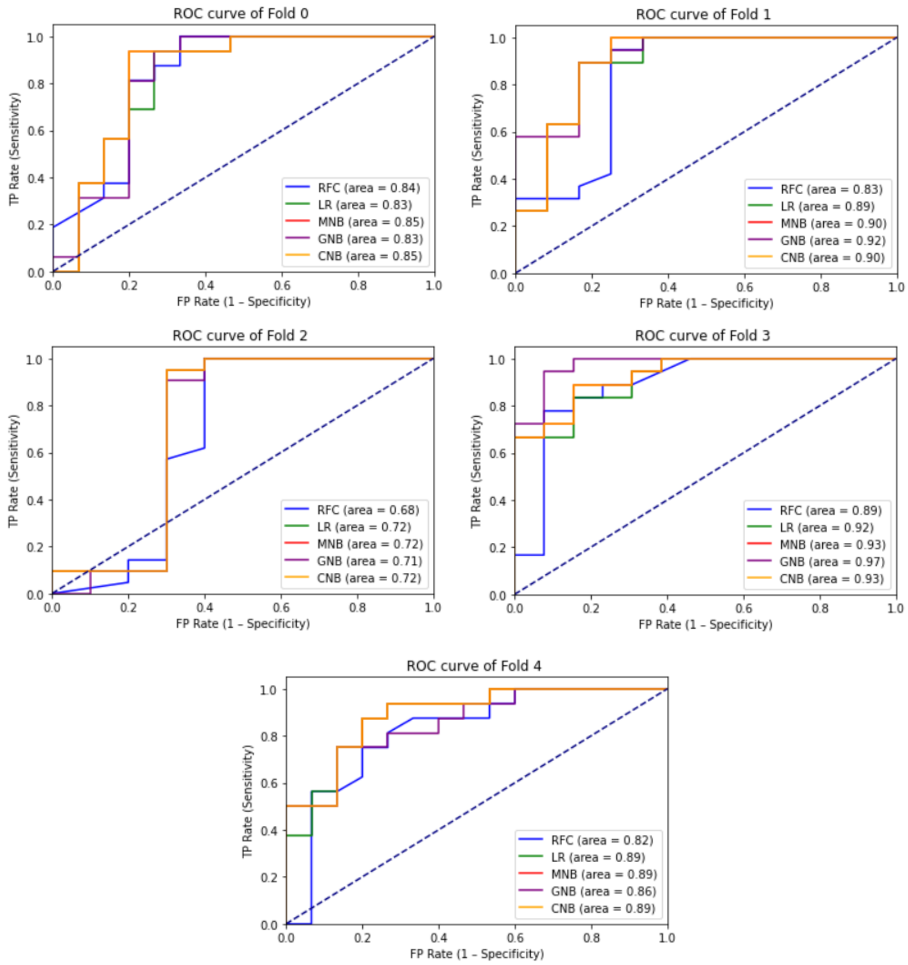

Figure 3.3: ROC graph of IBM emotional analysis for Towards/Away

Except for joy, the correlation between Option/Procedure with other emotions is positive. In other words, the mean value of emotions for Procedure is greater than the mean value of emotions for Option; for example, the mean value of disgust for Procedure is 0.022 while it is 0.019 for Option.

The ROC curve has been presented for Towards/Away to show which ML model was considered for each proposed method, as well as the results of each model in each fold. As Figure 3.3 shows, although the MNB and CNB diagrams coincide and the mean AUC for both diagrams is 85, the model's ability to predict test data was not good after choosing the MNB model. There were 18 incorrect predictions out of 31 while there were 8 incorrect predictions for CNB.

3.4 IBM sentiment analysis

The study of words is integral to social, clinical, personality, and cognitive psychology. Also, taking words a bit more seriously will help us better understand who we are and what we do [28]. By detecting the hidden subjective expression in the text, sentiment analysis recognizes the matching relationship between characteristic words and emotional words in the text. Hence, using sentiment analysis can help to perform a variety of tasks such as subjective classification, opinion summary, etc [66]. Thus, sentiment analysis has been employed to do a typical TC task. In other words, sentiment analysis has been employed as a second approach to uncover hidden behaviors.

3.4.1 Results

Table 3.2: Results of applying IBM sentiment analysis to different behavioral indices

Behavior index	Towards/Away	Internal/External	Option/Procedure
model	RFC	LR	GNB
accuracy	80%	57%	57%
F1	80%	47%	50%
Precision	81%	76%	50%
Recall	80%	57%	57%
AUC	83%	45%	69%
# wrong prediction on the Test Set	6 of 31	12 of 28	9 of 21
#Wrong prediction on Unseen Dataset	8 of 53	21 of 42	N.A.

For Toward data, the mean sentiment score is 0.78, while for Away is -0.36. One reason for the negative sentiment for Away data is that Away people tend to focus on negative events. The rate of neutral sentiment for Towards data was more than Away data; it was 10 to 6.

As shown in Table 3.2, with the sentiment analysis, Towards and Away behavior has the best answers, compared to other behavior types. Although the accuracy of this model is higher than the previous one, it seems that this approach is not yet a good approach for Internal and External data because it cannot predict unseen data well (the number of wrong predictions is higher than the previous approach). Because there is not much sentimental difference between Internal and External sentences (the difference between the mean sentiment score of Internal and External behavior is 0.09). In other words, out of 63 and 74, IBM regarded 23 Internal data and 30 External data as neutral, respectively. The biggest difference between the mean scores of sentiments was the Towards and Away behavior, with a difference of 1.14, followed by the behavior of Option and Procedure, with a difference of 0.20, and finally Internal and External. This difference shows that it's easier to put data into groups when there is a bigger difference between their average scores.

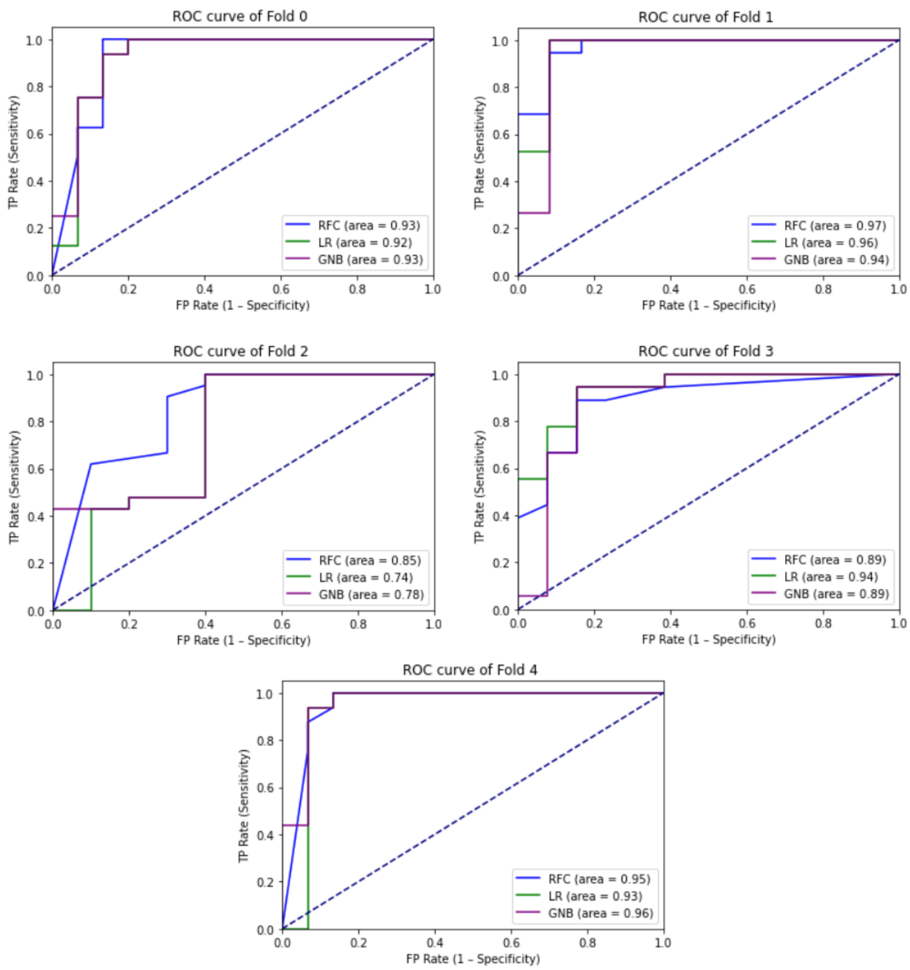

Figure 3.4: ROC graph of IBM sentiment analysis for Towards/Away

Internal/External and Option/Procedure data are people's responses to why they think they have done a good job and, mainly, why they chose their job. While for Towards/Away, the response contains the reasons that are based on people's feelings. For example, "I feel a duty to myself to keep proving to myself what my best is and how far I can take my abilities." is an answer to why is this factor so important to you. Hence, the emotion and sentiment analysis is working well for Towards and Away.

Figure 3.4 illustrates the ROC curves for each fold of the Towards/Away behavior index. This indicates that the overall AUC of RFC is higher than others. As can be seen, the RFC and GNB diagrams coincide at some points. However, the ability to distinguish between classes, in RFC is better than GNB because its true positives are higher in fewer thresholds.

3.5 Amazon comprehend service

With ACS, text analytics can be utilized to extract insights regarding document content using a fully managed natural language processing service. In order to analyze text accurately, ACS uses deep learning technology. As well as gathering insights about a document or set of documents. To improve accuracy, its models are continually trained with new data across multiple domains (the software uses a pre-trained model). Based on existing data, Comprehend Custom builds customized NLP models on behalf of the user using automatic machine learning (AutoML). In other words, ACS is categorized as a machine learning service. It is capable of discovering insights from text and attaching labels or tags to textual units such as sentences and queries [48][67][68]. In comparison to other service providers, Amazon's approach is the simplest: you provide a text string and the language to be used for analysis. It is only necessary to have two columns in Comprehend to perform multiclass classification: one for text and one for classes [67].

There are two modes of document processing available in Amazon comprehend: synchronous and asynchronous. Asynchronous jobs have been used to process unseen data in this project.

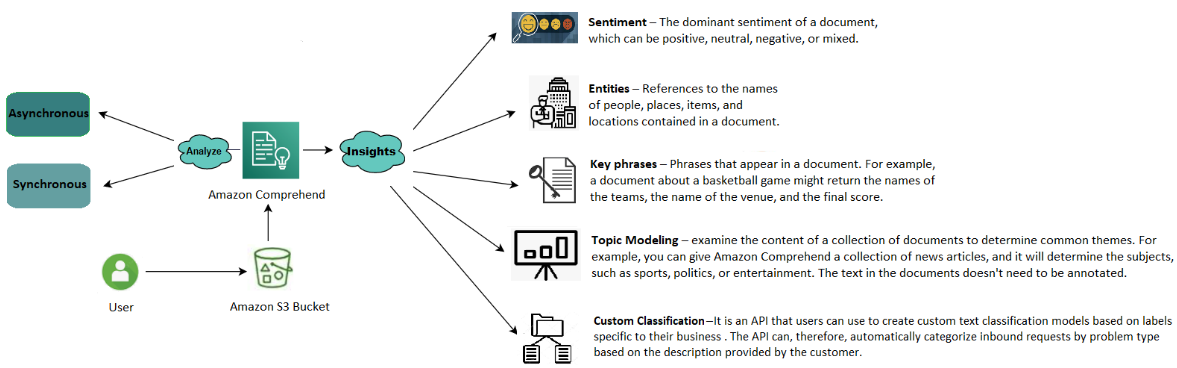

Figure 3.5: ACS workflow and insights

In order to run the ACS, the user must sign in to the AWS console and create a bucket. As a next step, upload the data and create a model within the same bucket. After that, the user can use features like keyphrase extraction, sentiment analysis, entity recognition, topic modelling, and custom classification APIs [48]. Figure 3.5 has been included as a demonstration of how ACS works and what it is able to perform.

3.5.1 Results

ACS calculates the average without taking into account the proportion of each label in the dataset (macro). But due to the imbalance in the dataset, the weighted average metrics were calculated for other proposed methods. Since ACS uses a different metric than the other methods, the F1, accuracy, precision, and recall of this method cannot be compared with those of other methods. The wrong predictions can, however, be

Table 3.3: Results of applying ACS to different behavioral indices

Behavior index	Towards/Away	Internal/External	Option/Procedure
accuracy	87%	60%	61%
F1	87%	59%	61%
Precision	87%	62%	61%
Recall	87%	60%	62%
AUC	N.A	N.A	N.A
# wrong prediction on the Test Set	4 of 31	11 of 28	8 of 21
#Wrong prediction on Unseen Dataset	39 of 53	19 of 42	N.A.

compared to those. As Table 3.3 provides an overview of the data, the prediction for test data is good, whereas the prediction for unseen data is poor. A poor prediction can be explained by the assumption that the unseen data was shorter than the train and test data. However, in comparison with IBM's model, this model does not offer as good a level of capability to predict the labels.

3.6 Vector-base method

As mentioned in section 2.6, raw data, which consists of a sequence of symbols, cannot be directly fed into text analysis algorithms because these algorithms are more suited to numerical feature vectors, which have a fixed size, rather than raw text documents with variable lengths. So, to transform an entire phrase into a vector format of real numbers that are used to predict the words and similarities, the vector base method has been used. In fact, vectorization can be used to classify texts, identify similar words, cluster the documents, or extract the features [69]. For converting data to vectors, sklearn TfidfVectorizer has been used. This vectorizer considers both unigrams and bigrams. Also, because of having phrases of different sizes, the data with less length than the maximum length has been padded by value zero.

Table 3.4: Results of applying Vector-base method to different behavioral indices

Behavior index	Towards/Away	Internal/External	Option/Procedure
model	LR	RFC	CNB
accuracy	61%	64%	71%
F1	56%	64%	59%
Precision	79%	64%	51%
Recall	61%	64%	71%
AUC	91%	69%	50%
# wrong prediction on the Test Set	12 of 31	10 of 28	6 of 21
#Wrong prediction on Unseen Dataset	17 of 53	20 of 42	N.A.

3.6.1 Results

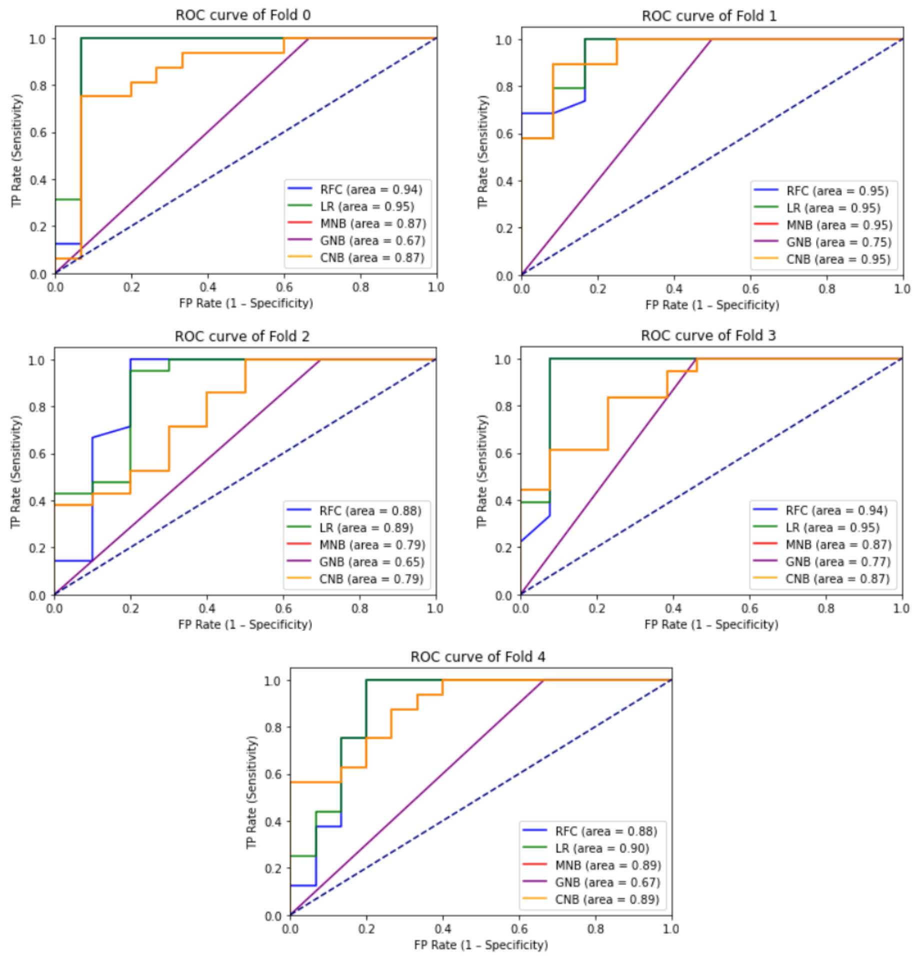

Figure 3.6: ROC graph of Vector-base method for Towards/Away

Since the experts provided the majority of the Away data (The data provided by them is shorter than the questionnaire answers.), the mean length for the Towards is 144.57 whereas the Away is 34.12. Consequently, the majority of vectors for Away data in the Vector-base method are zero, resulting in inaccurate predictions. The difference in length between the train/test data and the unseen data also caused poor predictions for Internal/External. When it comes to predicting test data, Option/Procedure result for this method is superior to other proposed methods. The Table 3.4 provides an overview. As can be seen, the precision is higher than accuracy, which shows that the true positive rate is greater than the false positive rate. In addition, not only is the accuracy of models for Towards/Away and Internal/External low, but also the number of incorrect predictions for unseen and tested data is unacceptable. In other words, the model is underfitted. Alternatively, the models are not capable of identifying simple patterns within datasets. In order to overcome underfitting, it is possible to increase the complexity of the model. As part of an effort to increase the complexity of RFC, the number of trees has been increased from 100 to 200. Unfortunately, no difference has been observed in the results. Also, the maximum number of iterations required by the solvers to reach convergence has been increased from 100 to 200 for LR. Once again, the problem was not resolved.

Figure 3.6 illustrates the perfect fit of the LR model to a specific sample of test data for Toward/Away. Because in comparison to others with lower thresholds, it has a higher true positive rate.

3.7 Bag of words method

Word counts are a popular method of quantitative text analysis. A number of prediction problems, such as language modelling and document classification, have been solved successfully using this technique. Also, Joulin et al. demonstrated that the representation of sentences as bags of words (BoW) and the training of a linear classifier, such as logistic regression, is a simple and efficient baseline for sentence classification [31].

Khalghollah et al. address the challenges of designing a learning control for system dynamics with significant nonlinearity and considerable bias. They do this using a graphical quantitative theory of human personalities, which assumes that their personalities represent how people interact with the world around them using feedback. The model is designed in the form of coordinates, which on the horizontal axis is empowerment/manipulation, which illustrates how individuals attempt to influence the world and other individuals, and on the vertical axis is emotion/logic, which illustrates how individuals make choices. In this model, choosing a personality leads straight to selecting numerical weightings [53]. For this purpose, they created a GUI that allows them to choose empowerment, manipulation, emotion, and logical words in individuals' speech and count them for each person. The same GUI has been used for the Towards, Away, Internal, and External behavioral indexes. Because, words were also considered for analyzing behavior by experts. They assumed that External people rely on others for their motivation and feedback. So, their answers to "how do you know you have done a good job at .....?" contain words to express feedback from others, such as, "When our clients are happy and praise us for our good work with them." or "When my manager shows a positive reaction to the work.". However, counting words is based on the premise that the words people use convey psychological information beyond their literal meanings and independent of their semantic contexts [28]. In other words, this approach is different from the way we communicate with each other. It treats the language as if it were merely a bag of words, and every message is just a random assortment of those words.

Hence, by providing a GUI to the experts, the experts were able to assign each word or collocation to a behavioral index or even discard the word. The system was able to save the assigned words to each behavioral index and then assess each dataset. For example, the scores that are generated by the system for the sentence "The clients are successful in accomplishing their goals, and they express their appreciation either in words or through their actions." are 3 Internal and 7 External. In other words, this sentence is External.

3.7.1 Results

Table 3.5: Results of applying Bag of words method to different behavioral indices

Behavior index	Towards/Away	Internal/External
model	LR	RFC
accuracy	87%	42%
F1	87%	42%
Precision	87%	42%
Recall	87%	42%
AUC	87%	54%
# wrong prediction on the Test Set	4 of 31	16 of 28
#Wrong prediction	13 of 53	14 of 42

The interesting point about the Bag-of-word method is that the experts have chosen mostly words with negative sentiments as Away, such as bad, bitter, avoid, blame, complacent, etc., while for Towards they have considered words with positive sentiments such as achieve, successfully, admire, amazing. These findings support the positive response provided by the IBM sentiment analysis results.

The experts have chosen mostly neutral words and words related to the outside world for External like budget, career, cash, check, and client. For Internal, some chosen words are positive, like desire; some neutral, like duty; and some negative, like fear. Table 3.5 shows the results of this method. However, the poor results for Internal and External indicate that this is not a proper method for this behavior because some parts of the dataset have not been evaluated by the expert (In the Work Limitation 3.10 section, the reason why experts did not collaborate on this part and considered it an unreliable method has been explained.).

Like other methods, the ROC curve for each fold of Towards/Away has been shown in Figure 3.7. LR, RFC, and GNB had similar AUCs in most folds. However, LR had a higher mean AUC than the others. Furthermore, the true positive rate for LR is higher than others. Hence, the LR model has been chosen for Towards and Away.

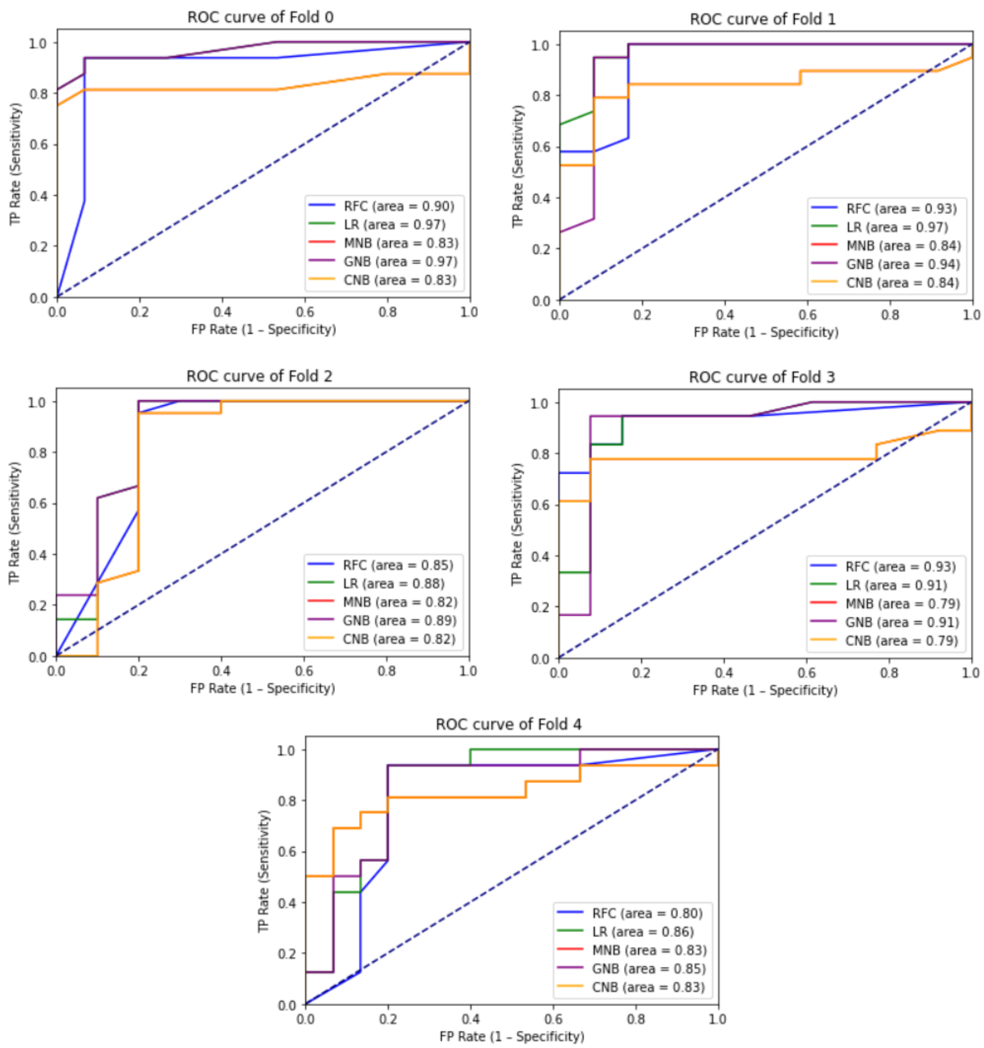

Figure 3.7: ROC graph of Bag of words method for Towards/Away

3.8 Breaking sentences method

Experts recommended breaking down the sentences, scoring each sentence, and labeling the sentences with the highest scores in order to collect more information regarding the Option or Procedure. Thus, after calling `https://pypi.org/project/punctuator/`, followed by `https://github.com/notAI-tech/fastPunct` and breaking the sentences, 350 data were available. In total, 50 records were removed that did not fall under either an Option or a Procedure. In addition, there were 71 neutral data, which were either Options or Procedures, and those were also removed. There are 103 Options and 126 Procedures following that. An example of the work that experts have done is: "I chose but also got chosen. I loved the rush and the fast-paced environment. They say you have to feel uncomfortable at least once a day in order to grow." There are 3 sentences in this sample. The first one is neutral because the first part "I did choose" is Option and "got chosen" is Procedure. The second and third sentences are options.

The reason behind using punctuator is that, for Option/Procedure, syntax plays an important role in prediction. For example, people with Option behavior tend to use bullet points or numbers in their writing. Those who are Procedures write more and their writings are longer than Options.

The vector-Base method 3.6 with the proposed workflow was implemented for this method as well, but unfortunately, the accuracy of this method by using the model with the highest AUC, MNB, was only 63% which was not acceptable since experts expected an accuracy greater than 80%.

3.9 Few shot learning method

In data-intensive applications, machine learning has been highly successful. However, when the data set is small, machine learning often fails to achieve its full potential. In other words, with a small amount of data, in most cases, there are prediction errors and the model is unable to make perfect predictions [54]. Also, many artificial intelligence (AI) applications that use large datasets are proving to be more powerful than humans in many fields, such as AlphaGo, which beats human champions, and image recognition, where the residual network (ResNet) performs better on ImageNet than human classification.

The ability of humans to learn new tasks rapidly is attributed to the fact that they have acquired knowledge over the years. In other words, humans are capable of using their prior knowledge to solve a new task, which is a subfield of FSL, called Meta Learning. Hence, if a person wants to label a text based on its sentiment, a few examples must be shown, the task must be explained to the labeler, or even the answers might be discussed. Furthermore, humans are already familiar with a variety of languages and words. Thus, unlike machine learning models, they do not begin from scratch. Due to this, a similar approach can be applied to ML problems by using pre-trained models, which are models that have already been developed by others for similar problems. Also, it has been shown that many NLP tasks can be improved effectively by using language model pre-training [70].

FSL, which is a new ML paradigm, has been proposed to learn from a limited number of examples just as humans do. In other words, the objective is to bring AI and humans closer together, which can help to construct computer programs that automatically improve with experience. It is mainly used in computer vision, where computers are able to gain a high-level understanding of digital images or videos through a process referred to as FSL. The purpose of this approach is to address problems such as the classification of images, the recognition of objects, the recognition of gestures, the location of scenes, and the prediction of motion. So, there are a lot of datasets related to these tasks that researchers can train their models on. However, there has been a recent increase in interest in the use of FSL in NLP. A few examples of applications include parsing, translation, sentence completion, sentiment classification, user intent classification for dialog systems, and criminal charge prediction. Unfortunately, there is not enough data or models for text, and also, in some cases, augmentation rules are specific to a particular data set, which makes it difficult to apply them to other data sets. Additionally, it is unlikely that a human being is capable of enumerating all possible invariants. As a result, manual data augmentation is not capable of solving the FSL problem in its entirety [54]. Also, as it has been discussed in 3.1, the experts did not accept augmented data. They also believe that for Towards or Away, the concept is close to sentiment analysis, but other data sets like IMDB Movie Reviews, or Twitter US Airline Sentiment can not be used. Because, as demonstrated by sentiment analysis, the data may be Away but not negative, such as ”I can’t ensure that I will always achieve this goal if I am not growing on a regular basis.”, or even an Away data may be neutral, such as ”one drama-creating team member has the potential to lower the level at which the entire team operates at.”. In other words, according to experts in this domain, using neither augmentation nor other data sets is not a valid approach to solve the problem.

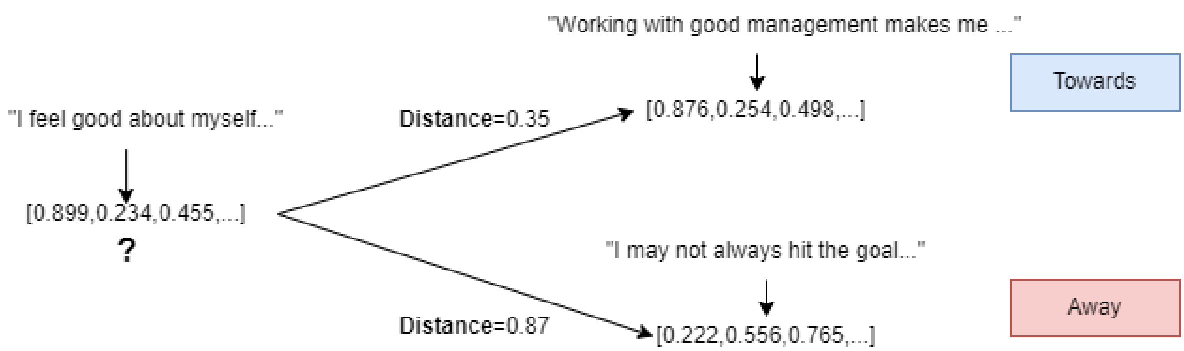

Figure 3.8: Distance similarity sample

Parnami et al. [71] have categorized the FSL into two groups. Non-Meta Learning and Meta Learning In the case of Non-Meta Learning, it is possible to transfer the knowledge from an already learned task (Transfer Learning). The idea is to find a similar task for which there is plenty of data, train a network, and then fine-tune the network based on the few-shot data. As part of their categorization, Meta Learning is classified as metric based, optimization based, and model based. The metric learning task is to learn the distance through data samples. In other words, its goal is to assign the predicted label to a shorter distance. As an example, in the figure 3.8, the data whose distance is going to be labeled is closer to the Towards data, so it will be labeled Towards. An optimization-based approach aims to maximize generalization performance with limited training data. To do so, the first step is to have a dataset that is closely related to the task, and then the meta-learner model aims to initiate the parameters by looking at similar tasks, followed by the base-learner aiming to have a minimum loss on the current task by using those parameters. Therefore, to achieve the lowest amount of loss by having the least amount of data, the learning process begins by looking at similar tasks and trying to find the best parameters for them, then proceeding through those parameters for the current task. Model-based meta-learning has no assumptions for learning or providing the parameters. It uses internal or external architecture to change the parameters rapidly. As an example, Memory-Augmented Neural Networks (MANN) is a kind of metric-based meta-learning that uses memory modules to encode and generalize the entire support set into memory slots [71].

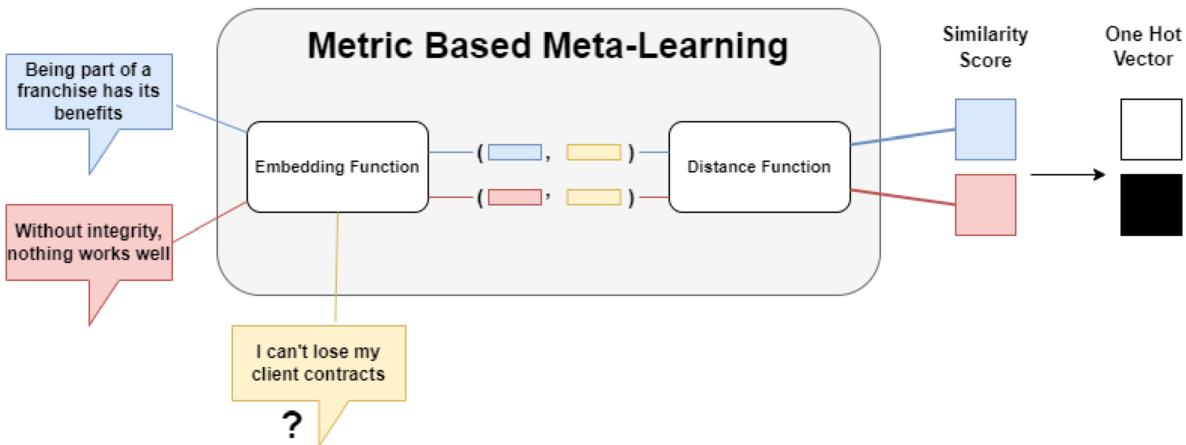

Figure 3.9: Metric-based meta-learning method workflow

Since the current problem is supervised learning, and according to Huisman et al., metric-based techniques are simple and effective for supervised learning [77]; the reason is that metric based meta learning is a pairwise comparison, which relies on labeled examples as input. I have used the metric based on which a pre-trained model has been used to convert the text to vectors. In this work, a 2-way-62-shot learning approach has been employed in this TC problem for training. Hence, 62 examples have been used to discriminate between 2 classes. Hence, for each data in the test set, the distance between the current data and the whole training data has been calculated. Finally, the train data label that has the shortest distance with the test data has been taken into account. Fig 3.9 shows the whole process of employing meta-learning. Here, the blue data indicates Towards and the red data indicates Away, and the task is to predict the label for the yellow data. After converting all of the data into a vector, the distance between the yellow data and the red data was calculated; since the yellow one is closer to the red one, it was considered as Away data.

In the next few sub-sections, different pre-trained models like BERT, or vector or glove models like word2vecgooglenews300, glovewikigigaword50, glovetwitter25, and also other pre-trained models like fast-textwikinewssubwords300, and conceptnetnumberbatch170630, which have been used in this approach, have been described. The purpose of this section is to understand where the knowledge from these pre-trained models comes from and then to understand which one is the best fit for this domain.

3.9.1 word2vecgooglenews300 model

Word vectors allow computers to interpret the meaning of words in human-written text. It is intended to represent words as lists of numbers, where small changes to the numbers correspond to small changes in the meaning of the word. Using word vectors, an algorithm can compare words by their meaning rather than just their spelling, which is helpful for building AI algorithms for NLP. As an example, vector ('Paris')-vector ('France')-vector ('Italy') results in a vector that is very close to vector ('Rome') [72]. In other words, by grouping similar words into vector representations, distributed representations of words enable learning algorithms to do algebraic operations on words and perform well in many NLP tasks.

Thus, word2vec is a tool that takes a text corpus as input and produces a vector of words as an output. First of all, it constructs a vocabulary based on the training text data, and then it learns how to represent words in a vector format. In many NLP and ML applications, the resulting word vector file is used as a feature. Hence, word2vec appears to be capturing important statistics about the training corpus based on their impressive capability at transferring to new problems. Hence, a model's representations can be more accurately transferred to other NLP problems the more data it receives to be trained on [73].

There are many applications of NLP and ML that rely on pre-trained word representations derived from large corpora of text, such as news collections, Wikipedia, and web crawlers. In general, pre-trained representations provide distributional information about words, which enhances the generalization of models learned from a limited number of examples. A large labelled corpus of text data is typically used to gather this information [73]. As an example, word2vecgooglenews300 is a pretrained model that uses a portion of the Google News dataset (about 100 billion words). Three million words and phrases (data-driven approaches are used to obtain the phrases) are represented by 300-dimensional vectors in the model.

Results

As a result of the loss or alteration of research datasets over time, the fact that they become obsolete, or the fact that they cannot be accessed and processed due to the lack of an appropriate implementation. Hence, Gensim has developed a standardized API for processing unstructured text (without images or audio) and launched its own dataset storage, committed to long-term support. For this project, word2vect or other pre-trained models (except BERT) were loaded using the Gensim downloader API.

The output from using this pre-trained model was promising. There were 11 incorrect predictions out of 53, which is good, but it is still less than IBM sentiment and emotional analysis. In the predicted test set, there are five errors out of 31. However, as the number of samples increased, the accuracy increased too. For 5 samples for each behavior index, the accuracy was less than 65 percent, then it increases.

3.9.2 glovetwitter25 and glovewikigigaword50 model

Using vector arithmetic, recent methods for learning vector space representations of words have been able to capture fine-grained semantic and syntactic regularities. GloVe, derived from Global Vectors, which is an open-source project at Stanford, is a distributed word representation model. For obtaining vector representations of words, the model uses an unsupervised learning algorithm. The way to achieve this is to map words into a meaningful space in which the distance between words is related to the semantic similarity. In other words, during training, word-to-word correlation statistics from a corpus are aggregated, and the resulting representations showcase interesting linear substructures of the word vector space. First, let's discuss what a co-occurrence matrix is before describing Glove usage. It is defined as follows:

*In the case of a corpus comprising **V** words, the co-occurrence matrix **X** will be a **V** * **V** matrix, where the i_{th} row and j_{th} column of **X** (X_{ij}), denotes how often word **i** has been seen together with word **j***

Consider the following two sentences as a corpus: "The fast cat wears no hat." and "The cat in the hat ran fast.". There are eight words in this case: cat, fast, hat, in, no, ran, the, and wear. A matrix of 8x8 will result. Assuming that column and row indices begin at zero, column 0 and row 1 will have a value of two because cat and fast happened to be together twice.

It is possible to use GloVe to find relationships between words, such as synonyms, company-product relationships, zip codes and cities, etc. In addition, it can also be used to develop online and offline systems designed to detect psychological distress in patient interviews. However, the unsupervised learning algorithm does not accurately identify homographs, i.e., words with the same spelling but different meanings. As an example of homographs, consider the word "content", which means happy or satisfied in the sentence "Pedro was content with his life." and is used as an adjective, whereas it means all that is contained within something in, "This program includes inappropriate content" and it has been used as a noun. The cosine similarity of these two sentences by using glovetwitter25 pre-trained model is 0.84, which means the model assumed that the sentences have similar meaning because of word content but they do not.

There are two pre-trained glove vectors that have been used. The first one is **glovetwitter25**, which is based on 2 billion tweets, 27 billion tokens, and 1.2 million vocab, and whose text has been lowercased before any further step (uncased). This model contains 1193514 vectors. The other one is, **glovewikigigaword50** which is based on Wikipedia 2014 plus Gigaword 5 (English Gigaword Fifth Edition is a comprehensive archive of newswire text data that has been compiled by the Linguistic Data Consortium (LDC) over several years.) contains 400000 vectors.

Results

Although the accuracy of both of these models is high, they overfit on their first attempt. In other words, the model is not capable of identifying basic patterns in a dataset that has not previously been trained. To solve this problem first, I have tried to reduce the number of samples from 62 (for each behavior index) to 40 in 12 steps. Unfortunately, the accuracy was still high and the number of wrong predictions was high too. So, in the next step, the length of the vectors has been reduced incrementally from 10 to its max dimension (glove wiki dimension is 50 while glove twitter is 25). For glove wiki, its best result was 17 incorrect predictions with a vector size of 33, and for glove twitter, it was 18 with a vector length of 22. For glove wiki, its accuracy is 87% and for glove twitter, it is 85%, respectively. As a result of reducing the number of vectors, glove wiki's number of incorrect predictions has dropped to 5 and glove twitter's number has dropped to 2. These results indicate that even after cutting the vectors, the number of incorrect predictions for unseen data remains high.

3.9.3 fasttextwikinewssubwords300

FastText, written in C++, is an extension to Word2Vec that supports multi-processing during training, which was proposed by Facebook in 2016. This library is intended to facilitate the learning of word representations and sentence classifications in an efficient manner. Both supervised and unsupervised problems can be solved using FastText. A FastText word is defined as a bag of character n-grams plus the word itself. For example, the FastText representations of the character n-grams for the word mathematics are $\prec$ ma, at, mat, math, the,...,mathematic,mathematics $\succ$. $\prec$ and $\succ$ are used to differentiate the ngram of a word from the word itself, so for instance, if the word math is part of the vocabulary, it is represented as $\prec$math$\succ$. By doing this, short words that may appear as ngrams of other words can retain their meaning. Also, this allows to capture the underlying meaning of suffixes and prefixes inherently. All the above n-grams together will make up mathematics's word embedding vector. Upon training the neural network, word embeddings will be available for all the n-grams given the training dataset. Since some n-grams of rare words are likely to appear in other words, it is now possible to properly represent rare words. There are 999999 vectors in FastText, and its sources are:

- **UMBC webBase corpus**: It was derived from the February 2007 crawl. Over 100 million web pages from over 50,000 different websites are included in this collection, making it one of the largest. It contains a collection of English paragraphs with over three billion words. Although extracting textual content from HTML tags was a huge success for the Stanford WebBase project, there are still numerous instances of duplicated texts, truncated texts, non-English texts, and strange characters in this database.

- **statmt.org**: The site provides resources for the study of statistical and neural machine translation, i.e. the computerized translation of text from one human language to another based on the translation of vast amounts of data.

- **Wikipedia 2017**: For each page in the English Wikipedia, which is a knowledge source till 2016-12-21, only the plain text is extracted and all structured data sections such as lists and figures are stripped. A total of 5,075,182 articles with 9,008,962 unique uncased token types were retained after eliminating internal disambiguation, list, index, and outline pages [75].

Results

From 10 to 62 samples, the accuracy has been monitored. The minimum accuracy for this model is with 55 samples, which is 58 percent, and the maximum is with 62 samples, which is 83 percent. However, just like previous models, this model is not able to predict the unseen data well. At first, there were 25 incorrect predictions. Then, by reducing the length of the vectors from 300 to 170, the number of incorrect predictions is reduced to 17 with an accuracy of 80. In the end, this pre-trained model isn't able to correctly predict Towards/Away using information from different sources.

3.9.4 conceptnet-numberbatch-17-06-300 model

The ConceptNet semantic network is a freely available model that helps computers comprehend the meaning of words that people use. In 1999, at the MIT Media Lab, an initiative called Open Mind Common Sense was launched, which led to the conception of ConceptNet. As a result, knowledge has been gathered from other crowdsourced resources, expert-created resources, and games with a purpose.

In contrast to Google News and GloVe 1.2 embeddings trained on 840 billion words in the Common Crawl, ConceptNet is a huge open-source knowledge graph that focuses on the meanings of words that are common-sense (not named entities). Due to its focus on words, it is particularly suitable for representing word meanings as vectors [72]. It is clear that ConceptNet has continued to play an important role in a field that has come to focus on word embeddings, since word embeddings have the benefit of ConceptNet's experience. As demonstrated by ConceptNet Numberbatch's state-of-the-art results in matching human annotators on multiple evaluations, it is possible to improve word embeddings using ConceptNet as a means of improving their robustness and correlation with human judgment [75]. ConceptNet resources include:

- A number of facts have been gathered from Open Mind Common Sense (OMCS) (an artificial intelligence project based) and sister projects in other languages (By utilizing the contributions of many thousands of individuals on the web, these projects utilize a large commonsense knowledge base.).

- By using a custom parser (”Wikiparsec”), the information is extracted from multiple languages of the Wiktionary database, which is a web-based dictionary of terms in all natural languages and several artificial languages. Wiktionary is a project of the English language that aims to describe all words from all languages using definitions and descriptions that are in English.

- The Open Multilingual WordNet is a linked-data representation of the original WordNet in multiple languages and its parallel projects in multiple languages.

- JMDict which is a Japanese-multilingual dictionary.

- The OpenCyc hierarchy of hypernyms is based on Cyc, a predicate logic system used to represent commonsense knowledge.

- The DBPedia, which is a network of facts derived from Wikipedia infoboxes.

Since it is possible to enhance machine learning about the language if it is provided with specific knowledge and information from external sources of information, ConceptNet combines these sources to create a knowledge graph in which words and phrases of natural language are connected by labelled edges. ConceptNet contains over 21 million edges and over 8 million nodes. Approximately 1,500,000 nodes are contained in its English vocabulary, and at least 10,000 nodes are contained in 83 languages. An important factor influencing how a knowledge graph is used is what each node represents. In addition, it has implications for linking and importing other resources. Due to the fact that different resources present their content in different ways, the process of linking and importing other resources has become challenging. Nodes in ConceptNet represent words or phrases in natural language, usually common words written in their clarified form. For instance, the noun form of the word ”balance” in English is represented in ConceptNet by the URI /c/en/balance/n[72].

Result

Although similar to Fasttext, the ConceptNet knowledge is from different sources, and its vectors are more than glove models, but its results are way better than the others. Initially, it had an accuracy of 90%, with 14 incorrect predictions. By reducing the length of the vector to 20, its accuracy has been reduced to 85 and its incorrect prediction to 13. In the current model, all of the wrong predictions are Away data. In other words, it predicts all the Away data as Towards. Another characteristic of this model is that, like Word2vec, as samples increase, accuracy increases as well.

3.9.5 BERT

In 2018, BERT was published as a language model that converts words into numbers. An algorithm that converts words into numbers is crucial because machine learning models accept input from numbers (not words) so that you can train them on original-textual data. To do so, the original model of Bert has been trained on BooksCorpus (a dataset containing more than 10,000 books of different genres), which has about 800 million words. It has several advantages over the others (i.e., word2vect, glove, etc.). Firstly, it is a bidirectional training model, which can provide a deep understanding of language context and flow. Hence, it can completely differentiate between two different meanings of the bank in "bank deposit" and "river bank", in which the first one is a human-made building and the second one is land alongside a body of water. In other words, it is capable of capturing the relationships in a bidirectional way. Secondly, it is a pre-trained model with a large dataset, which means its ability to predict is usually better than others.

Halder et al. have proposed TARS (Task-Aware Representation of Sentences) [76] a method in which they add a binary classifier on top of BERT. They have shown that the ability of their algorithm is higher than BERT when the number of samples is small (FSL). According to them, by providing a label for each text, the model has both the examples, which are text, and the descriptions, which are labels. As a result, they assume that they will be able to add the ability of humans to their algorithm. which is, describing the label to the person when they wish to perform a classification task. They also demonstrated that their algorithm's ability is superior for some datasets than others. For doing classification, they have created a similarity measure between the label names based on distances between their BERT encoded embeddings. A similar methodology has been employed in this study as previously mentioned.

Small BERT has been called in this method, which is the original BERT with fewer layers. There are 4 transformer layers with an embedding size of 512 and 8 attention heads. In other words, Small BERT is a black box that generates a vector shaped 512 for each input token in a sequence. The attention layers are learning layers; they look at different parts of the sentence and try to discover more semantic or syntactic information.

Result

Even though BERT is one of the most popular models for solving a variety of problems, for the current domain it does not work properly. After cutting its vector to 35, its ability to predict unseen data is 18.

The majority of scholars, such as Halder et al. [76], consider the accuracy of their results. However, since this research is concerned with the ability to predict unseen data, the results are not satisfactory.

3.9.6 Conclusion

A few-shot learning approach is useful when training examples are hard to find (e.g., rare diseases), when data annotation is costly, or when there are ethical issues. It can also be utilized when time is limited to gather new data, which in this domain was one of the reasons for the lack of data. Thus, few-shot training differs from traditional methods of training machine learning models, which typically require large amounts of training data. In other words, few-shot learning approaches are intended to solve a task using only a few samples. Hence, for Towards and Away behavior, this algorithm has been implemented because of the rarity of samples, the need to reduce data collection effort, or to mimic human learning.

Table 3.6: Results of applying different word embeddings

Model Name	***Source***	***#Vectors***	***Acc***	***Wrong pred***
word2vecgooglenews300	Google News (about 100 billion words)	3000000	%83	11
conceptnetnumberbatch170630	OMCS, Wiktionary, JMDict, DBPedia	1917247	%85	13
fasttextwikinewssubwords300	Wikipedia 2017, UMBC webbase corpus and statmt.org news dataset (16B tokens)	999999	%80	17
glovewikigigaword50	Wikipedia 2014 + Gigaword 5 (6B tokens, uncased)	400000	%87	17
glovetwitter25	Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased)	1193514	%85	18
BERT	BooksCorpus	N.A.	%83	18

In this proposed method, through the use of the trained model, all of the text was transformed into embeddings. Then, for each data point in the test set, its distance from the train set was calculated and measured. The label of data with the closest distance has been assumed as the label of test data. In the first step, an assessment of the accuracy and number of incorrect predictions was made. Since some models had high accuracy, their ability to predict unseen data was weak. Hence, in the second attempt, the accuracy and number of wrong predictions were calculated for samples ranging from 10 to 62 and vectors ranging from 10 to the maximum length of the vector.

M. Huisman et al. [77] point out that metric-based meta-learning is unable to incorporate new task information into network weights when there is a significant distance between meta train and meta test tasks. This may lead to a reduction in performance. Since final results were determined by the minimum number of incorrect predictions, the mean length for train is 99.20 and for test is 94.48 for unseen data is 11.62. In other words, the accuracy is high because the model can predict test data with few mistakes, but when it wants to predict unseen data where there is a lot of distance between its length and the training, the model cannot predict well.

The results of using a pre-trained model with this algorithm are shown in Table 3.6. In this table, the model name that has been used for calling the Gensim downloader API, the source of each model, the number of vectors, and the accuracy for each model have been mentioned. In addition, the number of incorrect predictions for unseen data, which is out of 53, has been provided. The models have been sorted based on their number of incorrect predictions, from low to high. According to the results, the word2vecgooglenews300 had the best result, which was 11, which predicted 6 Towards and 5 Away data incorrectly. After that, conceptnetnumberbatch170630 is the best model.

Due to the higher number of vectors for word2vect, its output was superior to the others. As can be seen, the results have not been affected by the variety of resources or even the architecture used to generate the vectors (BERT).

3.10 Work Limitation

Lack of data and imbalanced data are the main challenges. In addition, it should be noted that the text provided by experts is shorter than the actual data. In other words, the variety of data for only a few hundred data is overwhelming.

Although the results of the Bag of Words method are better than Vector-Base and ACS, But there are some threats to the validity of this method. The first is that experts stated that the model is time-consuming, so they did not assign all the words in the dataset to each behavior indicator. The other one is that some words might have more weight in the expert's mind while the system weights all the words equally. For analyzing each behavior, some questions needed to be asked from the applicant. Therefore, the experts emphasized that if they wanted to add another question to the questionnaire in the future for behaviors, assigning each word to each behavior index would need to be repeated. As a result of the non-automatic prediction of behavior, experts rejected this strategy. Also, as noted by Pennebaker et al., all text analysis programs that use word counts are incapable of considering context, irony, sarcasm, or even the possibility that a word can have more than one meaning [28].

Chapter 4

Conclusion

4.1 Summary and conclusion

In the workplace, team environment, or relationship, knowing one's personality traits has limited value. The behavior assessment is much more valuable in improving the effectiveness of human interaction in any environment. A pure behavior assessment differs from a personality assessment in this respect [28]. Furthermore, in addition to hard skills, an employee's behavior plays an important role in maintaining a stable organization, since behavior is determined by the circumstances in which they are employed [78]. One job may necessitate an individual's detailed attention to organizing and planning; another may require that a number of people work together to come to decisions as one. These are very different types of behaviors, but if an employee wants to be perceived as effective, he or she must exhibit these behaviors [1].

Global access to electronic technology and the internet, as well as the advancement of modern information and digital technologies, have made the universal HR management system more useful. However, implicit information such as individual behaviors, which is essential in understanding applicants' career-evolving patterns, is most often overlooked in most e-recruitment methods. Policies and practices that optimize employee-job fit can help employers hire and place motivated, committed employees. In this regard, by analyzing specific behaviors, such as a candidate's motivation or direction, this study can be used to determine the most suitable candidate for a particular position. AccuMatch Behavior Intelligence allows companies to classify candidates based on their capabilities and ability to adapt to their company's requirements. As a result, the company will be able to select the most qualified candidate for each job description, leading to an expert workforce. In other words, with the aid of this research, the HR department

will be able to simplify the daunting task of selecting the most appropriate candidate from a large pool

Table 4.1: An overview of the Towards/ Away data

Train/Test	***#Towards***	***#Away***	**Unseen**	***#Towards***	***#Away***
155	90	65	53	39	14

Table 4.2: An overview of the Internal/External data

Train/Test	***#Internal***	***#External***	**Unseen**	***#Internal***	***#External***
137	63	74	42	17	26

Table 4.3: An overview of the Option/Procedure data

Train/Test	***#Option***	***#Procedure***	**Unseen**	***#Option***	***#Procedure***
102	60	42	0	0	0

of applicants for a given position, resulting in a highly-skilled workforce. To achieve this goal, six different behaviors, which are: Towards, Away, Internal, External, Option, and Procedure, have been analyzed by the given data.

The objective of this research is to classify the data between each of these dichotomous behaviors. Since the task is TC, it refers to the assignment of predefined classes to free texts [37]. Hence, the performance of different text representation methods has been provided and analyzed to predict the mentioned behaviors. In order to make the comparison of classifiers more accurate, the following assumptions have been used:

- 1. Same dataset (An overview of the data can be seen in the tables 4.1, 4.2, 4.3.)
- 2. One particular way to compare the results (When the model is trying to predict unseen data, the sum of false positives and false negatives is taken into account)

This study requires the models to learn the trends in the training data and then apply this knowledge to the evaluation of new observations. Consequently, the models must be able to generalize well to the underlying structure rather than focus too much on the details of the training data. To find out if overfitting is happening, the models were compared based on how well they could predict data they had never seen before with the least number of mistakes.

Due to the use of off-the-shelf technology and ML classifiers, which have an automatic classifier builder (the learner), the author was able to select inductively between existing classifiers without having much knowledge of the underlying domain. In this respect, this work offers an advantage over knowledge engineering in which most of the effort is devoted to developing a classifier cite [37].

Table 4.4: Results of applying different methods for Towards/Away

Model Name	***Accuracy***	***#Wrong pred* (53)**
IBM emotional analysis	%74	7
IBM sentiment analysis	%80	8
ACS	%87	39
Vector-Base	%61	17
Bag of words	%87	13
conceptnetnumberbatch170630	%85	13
fasttextwikinewssubwords300	%80	17
glovetwitter25	%85	18
glovewikigigaword50	%87	17
word2vecgooglenews300	%83	11
BERT	%83	18

Table 4.5: Results of applying different methods for Internal/External

Model Name	IBM emotional analysis	IBM sentiment analysis	ACS	Vector-Base	Bag of words
Accuracy	%50	%57	%60	%64	%42
***#Wrong pred* (42)**	19	21	19	20	14

Table 4.6: Results of applying different methods for Option/Procedure

Model Name	IBM emotional analysis	IBM sentiment analysis	ACS	Vector-Base	Breaking sentences
Accuracy	%71	%57	%61	%71	%63
***#Wrong pred* (21)**	19	9	8	6	17

4.2 Experimental results

Table 4.4, 4.5, and 4.6 show the results for Towards/Away, Internal/External, and Option/Procedure, respectively, and the methods that have been implemented for them. The results show that for Towards/Away, IBM emotional analysis and IBM sentiment analysis have the best performance because the number of wrong predictions for unseen data is 7 and 8 out of 53, which is far less than the others. This result shows that there is a relationship between a statement's emotions and sentiment and the Toward and Away behavior. When considering the best proposed method with the minimum number of incorrect predictions, the following order appears: Bag of Words (Due to the fact that the experts did not evaluate all of the sentences, the output is not as valid as it could be.), ACS, IBM emotional analysis, Vector-Base, IBM sentiment analysis. As previously mentioned, the accuracy of this behavior is not acceptable.

It should be noted that the experts did not provide any new information regarding options or procedures. As a result, there is no unseen data concerning this dichotomous behavior. For this reason, the ability to predict test data has been considered for Option and Procedure. On the basis of the minimum number of errors in the test set, the order of methods for Option/Procedure is Vector-base, ACS, IBM sentiment analysis, Breaking sentences, and IBM emotional analysis. To sum up, the results for Internal/External and Option/Procedure were not as promising as those for Towards and Away. Therefore, the results of this research indicate that the proposed method is effective in determining the Towards/Away behavior index. To put it another way, experts can utilize the implemented proposed methods for Towards/Away when they are seeking to find the most suitable candidate based on their motivation index.

Two different AI technologies, IBM Watson NLU and ACS, were applied to the same problem in this research, and some advantages and disadvantages were identified. The reasons behind choosing these technologies were first, according to Ermakova et al. [44], IBM Watson NLU is the most accurate solution among other services (Amazon comprehend, Google Cloud Natural Language, Microsoft Azure, etc.). The other reason is that, although IBM personality assessment for any individual (or for a body of text) had been discontinued by the time this research was done, the technology was used because this research was attempting to solve a problem nearly the same. Also, ACS was the dominant player in cloud computing by the time this research was going on. Here is a comparison of the advantages and disadvantages of these services:

The first advantage of the ACS is that it is easy to use and the documents are easy to follow. In addition, the ACS does not have a limitation on the length of input text, whereas IBM Watson NLU will not accept text that is longer than 2000 characters and shorter than 3 characters. Furthermore, ACS support a greater variety of languages [79] than IBM Watson NLU [80]. IBM Watson NLU has some advantages, such as being free for 250K items per month, whereas Amazon does not. For this domain, the IBM Watson results for Towards and Away were better than those from Amazon.

a) Towards b) Away

Figure 4.1: Towards/Away WordCloud

In order to determine which words have been used more frequently for Towards and Away, figure 4.1 has been provided. Since the questions relate to the work, the criteria, and the experience, the most frequent words are "job", "work", "company", and "people". Figure 4.1 illustrates an interesting point; the word "without" is one of the most prevalent words used by people who want to get rid of something, which is Away people. Additionally, it shows that Towards and Away do not simply refer to being positive or negative. They refer to showing certain subsets of behaviors, such as Toward people are focused on goals and managing priorities, while Away people have difficulty maintaining focus on goals.

4.3 Value proposition

Table 4.7: Behavioral mapping research benefits

Value	***Target audience***	***clarification***
Assist in finding the right candidate	HR, Employer, project managers	Answers are categorized according to behaviors
Assist in adjusting for a specific position	employee	it can be used as a self-report tool
In contrast with other research [19][20][8], this study focuses on what people do rather than who they are	project manager, team leader	The focus of this research is on behavior while others is on personality
It has been observed that other researchers [26][27] focus on the impact of specific behaviors such as fair play and honesty on specific job roles, such as front-line employees and sales representatives, whereas the results of this study can be applied across all job roles.	project manager, team leader, HR, employer	Employers can determine the applicant's behavior by analyzing the answers to the AccuMatch questionnaire if they are seeking a candidate with specific behavior.
In comparison to others[19][20], this work considered more methods for the classification.	Researchers	in this work, 7 feature extractions (Proposed methods) have been considered for classifying the behaviors, while Menon et al. [20] consider two which are linguistic category and profile-centric category and Sudha et al. [19] consider an individual's personality features (Big five) and linguistic information.

Several works, including Humanic AI [5], Expert.ai [6], Crystal Knows [7], and Receptiviti (it is based on Prof. James W. Pennebaker's research, who is a prof in psychology) [8], have focused on personality traits and then the company's sales teams. A major advantage of this work is its ability to encompass all teams in a company. Secondly, they tend to focus on what people write on social media accounts or in their emails, which is incorrect. Because people often use tools for correcting their email in order to make it more effective, or even help others create great social media accounts, such as LinkedIn, so they have more opportunities in their lives. As a result, their work has some limitations that need to be considered.

In this study, standard questions were used to capture the behavior of all teams and also have the same input for everyone. Also, in this work, 7 feature extractions (proposed methods) have been considered for classifying behaviors, while Menon et al. [20] consider two, which are linguistic category and profile-centric category, and Sudha et al. [19] consider an individual's personality features (the Big five) and linguistic information. Therefore, a lot of effort has been put into finding the best method for categorizing these behaviors. Table 4.7 has illustrated briefly how this research differs from previous work and who can benefit from this research.

4.4 Future work

In this research, there have been no considerations made of demographic factors such as nationality, or difficulties associated with second language learning (where most of us make errors in particle usage rather than content words). There have been no mentions of language proficiency or intelligence. Also, linguistic categories such as average length of text, the average number of question marks, the average number of commas, etc. have not been considered. Moreover, different languages may also be considered in the future as inputs to this TC task.

Although to solve the problem, 7 methods have been proposed, but the results are close together. The best ones were for the Towards/Away behavior index by sentiment analysis, followed by emotional, and finally by the word2vectgooglenews method. I am assuming one reason behind that might be the number of sentences that exist in each dataset, and another is differences in length of data between expert data and questionnaire responses. Individuals were asked multiple questions in order to determine whether they were Toward or Away; hence, separating the answers to each question and labeling each one might improve the results.

The study of neurolinguistic programming examines how our thoughts influence our behavior. It examines how the brain interprets the signals it receives and how these interpretations influence what we do. A neurolinguistic programming technique accomplishes this by utilizing language. As a result of examining how the brain processes information, NLP techniques can help people view their thoughts, feelings, and emotions as things they can influence rather than passive events. However, there are still unresolved challenges in the technology of NLP (sentiment analysis), such as linguistic complications [44], which may have caused these results. Hence, it is possible that running the code in the next few months or years will produce better results.

AccuMatch Behavior Intelligence had designed 52 behaviors that can predict the applicant's working style or does the candidate like to work in a team or independently? What is the planning style of the candidate? Do they like rigid procedures? Do they need clearly defined parameters? In this study, I have focused on six behaviors, these behaviors can show how an applicant thinks about what is going on around. To compare the applicants with each other, the experts assess each applicant by scoring each behavior between 1 and

10. Because there were a few data, the models were not able to predict the correct score. I think that by having more data, a classification method is able to correctly predict the scores of behaviors.

Bibliography

[1] Bihelek, N., 2022. Behavioral intelligence history and benefits for coaches and their clients. [ebook] Available at: https://naguibihelek.com/get-my-ebook/ [Accessed 25 April 2022].

[2] O. Henry and Z. Temtime, "Recruitment and selection practices in SMEs: Empirical evidence from a developing country perspective," Journal of Advances in Management, 2010.

[3] A. S. Gamage, "Recruitment and Selection Practices in Manufacturing SMEs in Japan: An analysis of the link with business performance," Ruhuna Journal of Management and Finance, vol. 1, no. 1, 2014, Accessed: May 04, 2022. [Online]. Available: http://www.mgt.ruh.ac.lk/rjmf/pdfs/RJMF0101_JA_p37.pdf

[4] C.-F. Chien and L.-F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry," Expert Systems with Applications, vol. 34, no. 1, pp. 280–290, Jan. 2008, doi: 10.1016/j.eswa.2006.09.003.

[5] Humantic, "humantic AI — personality based prospect And talent intelligence," humantic.ai. https://humantic.ai/ (accessed Jun. 22, 2022).

[6] "Expert.ai Natural Language Understanding," Expert.ai. https://www.expert.ai/ (accessed Jun. 22, 2022).

[7] "Crystal - the app that tells you anyone's personality.," Crystalknows.com, 2019. https://www.crystalknows.com/ (accessed Jun. 22, 2022).

[8] "Receptiviti: The language and social psychology insights platform," Receptiviti. https://www.receptiviti.com/ (accessed Jun. 22, 2022).

[9] V. R, "Comparative Analysis for Personality Prediction by Digital Footprints in Social Media," June 2021, vol. 3, no. 2, pp. 77–91, May 2021, doi: 10.36548/jitdw.2021.2.002.

[10] M. Hassanein, W. Hussein, S. Rady, and T. F. Gharib, "Predicting Personality Traits from Social Media using Text Semantics," 2018 13th International Conference on Computer Engineering and Systems (ICCES), Dec. 2018, doi: 10.1109/icces.2018.8639408.

[11] Janz, T., 2022. Short, Shorter, Shortest – Online Tests vs. Candidate Social Media Analytics. [online] hrexaminer. Available at: https://www.hrexaminer.com/short-shorter-shortest-online-tests-vs-candidate-social-media-analytics/ (Accessed 25 April 2022).

[12] P. Novikov, L. Mararitsa, and V. Nozdrachev, "Inferred vs traditional personality assessment: are we predicting the same thing?," arXiv:2103.09632 [cs], Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.09632

[13] O. Alhendi, "Personality Traits and Their Validity in Predicting Job Performance at Recruitment: a Review," International Journal of Engineering and Management Sciences, vol. 4, no. 3, pp. 222–231, Sep. 2019, doi: 10.21791/ijems.2019.3.21.

[14] M. Hojat, J. B. Erdmann, and J. S. Gonnella, "Personality assessments and outcomes in medical education and the practice of medicine: AMEE Guide No. 79," Medical Teacher, vol. 35, no. 7, pp. e1267–e1301, Apr. 2013, doi: 10.3109/0142159x.2013.785654.

[15] K. Randall, M. Isaacson, and C. Ciro, "Validity and Reliability of the Myers-Briggs Personality Type Indicator," Journal of Best Practices in Health Professions Diversity, vol. 10, no. 1, pp. 1–27, 2017, doi: 10.2307/26554264.

[16] Jayaratne, M. and Jayatilleke, B., 2020. Predicting Personality Using Answers to Open-Ended Interview Questions. IEEE Access, 8, pp.115345-115355, https://doi.org/10.1109/ACCESS.2020.3004002

[17] Skloot, G., n.d. Skloot G., Behavioural interview questions. Asking the right behavioral questions to find the best candidate. [ebook] Available at: https://www.crystalknows.com/hubfs/eBooks/Behavioral%20Interview%20Questions/Behavioral%20Interview%20Questions.pdf [Accessed 25 April 2022].

[18] Levashina, J., Hartwell, C., Morgeson, F. and Campion, M., 2013. The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature. Personnel Psychology, 67(1), pp.241-293.

[19] G. Sudha, S. K. K, S. J. S, N. D, S. S, and K. T. G, "Personality Prediction Through CV Analysis using Machine Learning Algorithms for Automated E-Recruitment Process," IEEE Xplore, Dec. 01, 2021. https://ieeexplore.ieee.org/abstract/document/9711787

[20] V. M. Menon and H. A. Rahulnath, “A novel approach to evaluate and rank candidates in a recruitment process by estimating emotional intelligence through social media data,” 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Sep. 2016, doi: 10.1109/icngis.2016.7854061

[21] S. S. Keh and I.-T. Cheng, “Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models,” arXiv:1907.06333 [cs, stat], Jul. 2019, Accessed: Jun. 20, 2022. [Online]. Available: `http://arxiv.org/abs/1907.06333`

[22] R. A. Newgent, P. H. Parr, I. Newman, and K. K. Wiggins, “The Riso-Hudson Enneagram Type Indicator: Estimates of Reliability and Validity,” Measurement and Evaluation in Counseling and Development, vol. 36, no. 4, pp. 226–237, Jan. 2004, doi: 10.1080/07481756.2004.11909744.

[23] A. A. G. Agung and I. Yuniar, “Personality assessment website using DISC: A case study in information technology school,” 2016 International Conference on Information Management and Technology (ICIMTech), Nov. 2016, doi: 10.1109/icimtech.2016.7930305.

[24] S. Pande, “E-recruitment creates order out of chaos at SAT Telecom,” Human Resource Management International Digest, vol. 19, no. 3, pp. 21–23, Jun. 2011, doi: 10.1108/09670731111125899.

[25] A. Furnham, G. Trickey, and G. Hyde, “Bright aspects to dark side traits: Dark side traits associated with work success,” Personality and Individual Differences, vol. 52, no. 8, pp. 908–913, Jun. 2012, doi: 10.1016/j.paid.2012.01.025

[26] K. de Ruyter and M. G. M. Wetzels, “The Impact of Perceived Listening Behavior in Voice-to-Voice Service Encounters,” Journal of Service Research, vol. 2, no. 3, pp. 276–284, Feb. 2000, doi: 10.1177/109467050023005.

[27] S. Román and S. Ruiz, “Relationship outcomes of perceived ethical sales behavior: the customer’s perspective,” Journal of Business Research, vol. 58, no. 4, pp. 439–445, Apr. 2005, doi: 10.1016/j.jbusres.2003.07.002.

[28] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, “Psychological Aspects of Natural Language Use: Our Words, Our Selves,” Annual Review of Psychology, vol. 54, no. 1, pp. 547–577, Feb. 2003, doi: 10.1146/annurev.psych.54.101601.145041.

[29] W. Davies and A. McDonald, “The Science Behind Predicting Job Performance at Recruitment,” Pearson TalentLens, May 2018, Accessed: May 06, 2022. [Online]. Available: `https://www.talentlens.co.uk/wp-content/uploads/sites/5/White-Paper-The-Science-Behind-Predicting-Job-Performance-at-Recruitment.pdf`

[30] J. Feng, H. Xu, S. Mannor, and S. Yan, "Robust logistic regression and classification." , Accessed: Jun. 20, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2014/file/6cdd60ea0045eb7a6ec44c54d29ed402-Paper.pdf

[31] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of Tricks for Efficient Text Classification," Jul. 2016, Accessed: Aug. 03, 2022. [Online]. Available: https://arxiv.org/abs/1607.01759

[32] A. Alshamsi, R. Bayari, and S. Salloum, "Sentiment Analysis in English Texts," Advances in Science, Technology and Engineering Systems Journal, vol. 5, no. 6, pp. 1683–1689, Dec. 2020, doi: 10.25046/aj0506200.

[33] Y. Ko, "How to use negative class information for Naive Bayes classification," Information Processing & Management, vol. 53, no. 6, pp. 1255–1268, Nov. 2017, doi: 10.1016/j.ipm.2017.07.005.

[34] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861–874, Jun. 2006, doi: 10.1016/j.patrec.2005.10.010.

[35] M. Yuvalı, B. Yaman, and Ö. Tosun, "Classification Comparison of Machine Learning Algorithms Using Two Independent CAD Datasets," Mathematics, vol. 10, no. 3, p. 311, Jan. 2022, doi: 10.3390/math10030311.

[36] P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, and D. S. Lee, "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes," Journal of Clinical Epidemiology, vol. 66, no. 4, pp. 398–407, Apr. 2013, doi: 10.1016/j.jclinepi.2012.11.008.

[37] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, Mar. 2002, doi: 10.1145/505282.505283.

[38] "Watson Natural Language Understanding - Overview," www.ibm.com.https://www.ibm.com/cloud/watson-natural-language-understanding (accessed May 10, 2022).

[39] M. Mostafa, T. Crick, A. C. Calderon, and G. Oatley, "Incorporating Emotion and Personality-Based Analysis in User-Centered Modelling," arXiv:1608.03061 [cs], Aug. 2016, Accessed: May 12, 2022. [Online]. Available: http://arxiv.org/abs/1608.03061

[40] "Types of machine learning algorithms — en.proft.me," Proft.me, Dec. 24, 2015. https://en.proft.me/2015/12/24/types-machine-learning-algorithms/ (accessed May 10, 2022).

[41] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition From Text," IEEE Access, vol. 7, pp. 111866–111878, 2019, doi: 10.1109/ACCESS.2019.2934529.

[42] Y. M. Tun and P. Hninn Myint, "A Two-Phase Approach for Stance Classification in Twitter Using Name Entity Recognition and Term Frequency Feature," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), Jun. 2019, doi: 10.1109/icis46139.2019.8940282.

[43] S. Panichella, A. Di 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," IEEE Xplore, Sep. 01, 2015. https://ieeexplore.ieee.org/document/7332474

[44] T. Ermakova, M. Henke, and B. Fabian, "Commercial Sentiment Analysis Solutions: A Comparative Study," 2021.

[45] B. Liu, "Introduction," in Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, 2nd ed., Cambridge: Cambridge University Press, 2020, pp. 1–17.

[46] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," ACM Computing Surveys, vol. 50, no. 2, pp. 1–33, May 2017, doi: 10.1145/3057270.

[47] "Text Analytics Tools — 10 Text Analysis Software Reviews," Datamation, Apr. 09, 2021. `https://www.datamation.com/big-data/text-analysis-tools/` (accessed May 13, 2022).

[48] R. K. Mohanta, "Discover Insights and Relationships in Text Using Amazon Comprehend," International Journal of Trend in Innovative Research, no. 2582-0354, Accessed: May 12, 2022. [Online]. Available: `http://ijtiir.com/wp-content/uploads/2021/07/IJTIIR783402.pdf`

[49] "State of data science and machine learning 2021," www.kaggle.com. `https://www.kaggle.com/kaggle-survey-2021` (accessed Jun. 22, 2022).

[50] C. Basto, "Extending the Abstraction of Personality Types based on MBTI with Machine Learning and Natural Language Processing," arXiv preprint arXiv:2105.11798, May 2021, Accessed: May 09, 2022. [Online]. Available: `https://arxiv.org/abs/2105.11798`.

[51] L. Buitinck et al., "API design for machine learning software: experiences from the scikit-learn project," arXiv:1309.0238 [cs], Sep. 2013, Accessed: April. 24, 2022. [Online]. Available: `https://arxiv.org/abs/1309.0238`

[52] J. W. Pennebaker, "The secret life of pronouns," New Scientist, vol. 211, no. 2828, pp. 42–45, Sep. 2011, doi: 10.1016/s0262-4079(11)62167-2.

[53] M. Khalghollah and C. J. B. Macnab, “A high-level design process for neural-network controls through a framework of human personalities,” Learning Control, pp. 1–34, 2021, doi: 10.1016/b978-0-12-822314-7.00006-7.

[54] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a Few Examples,” ACM Computing Surveys, vol. 53, no. 3, pp. 1–34, Jul. 2020, doi: 10.1145/3386252.

[55] S. Y. Feng et al., “A Survey of Data Augmentation Approaches for NLP,” arXiv:2105.03075 [cs], Dec. 2021, Accessed: Jun. 06, 2022. [Online]. Available: `http://arxiv.org/abs/2105.03075`

[56] M. Bayer, M.-A. Kaufhold, and C. Reuter, “A Survey on Data Augmentation for Text Classification,” arXiv:2107.03158 [cs], Mar. 2022, Accessed: Jun. 06, 2022. [Online]. Available: `https://arxiv.org/abs/2107.03158`

[57] M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “Synthetic data augmentation using GAN for improved liver lesion classification,” IEEE Xplore, Apr. 01, 2018. `https://ieeexplore.ieee.org/abstract/document/8363576`

[58] X. Zhu, Y. Liu, J. Li, T. Wan, and Z. Qin, “Emotion Classification with Data Augmentation Using Generative Adversarial Networks,” Advances in Knowledge Discovery and Data Mining, pp. 349–360, 2018, doi: 10.1007/978-3-319-93040-4_28.

[59] M. Banko and E. Brill, “Scaling to very very large corpora for natural language disambiguation,” Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL ’01, 2001, doi: 10.3115/1073012.1073017.

[60] F. Georgiadis, “The Future Is Now: ‘Now AI can rephrase entire sentences and maintain their original meaning’ With…,” Authority Magazine, Dec. 16, 2018. `https://medium.com/authority-magazine/the-future-is-now-now-ai-can-rephrase-entire-sentences-and-maintain-their-original/-meaning-with-f98eb54d974b` (accessed Jun 06, 2022)

[61] L. Palen and K. M. Anderson, “SOCIAL MEDIA RESEARCH. Crisis informatics–New data for extraordinary times,” Science (New York, N.Y.), vol. 353, no. 6296, pp. 224–225, Jul. 2016, doi: 10.1126/science.aag2579.

[62] D. J. Hand and R. J. Till, “A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems,” Machine Learning, vol. 45, no. 2, pp. 171–186, 2001, doi: 10.1023/a:1010920819831.

[63] S. D. Walter, “The partial area under the summary ROC curve,” Statistics in Medicine, vol. 24, no. 13, pp. 2025–2040, 2005, doi: 10.1002/sim.2103.

[64] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?,’” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16, 2016, doi: 10.1145/2939672.2939778.

[65] P. Flach, J. Hernández-Orallo, C. Ferri, “A Coherent Interpretation of AUC as a Measure of Aggregated Classification Performance.” Accessed: Jun. 09, 2022. [Online]. Available:http://www.icml-2011.org/papers/385_icmlpaper.pdf

[66] H.-C. K. Lin, T.-H. Wang, G.-C. Lin, S.-C. Cheng, H.-R. Chen, and Y.-M. Huang, “Applying sentiment analysis to automatically classify consumer comments concerning marketing 4Cs aspects,” Applied Soft Computing, vol. 97, p. 106755, Dec. 2020, doi: 10.1016/j.asoc.2020.106755.

[67] H. Singh, “Machine Learning Use Cases in AWS,” Practical Machine Learning with AWS, pp. 207–225, Nov. 2020, doi: 10.1007/978-1-4842-6222-1_10.

[68] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning–based Text Classification,” ACM Computing Surveys, vol. 54, no. 3, pp. 1–40, May 2021, doi: 10.1145/3439726.

[69] Zamani, H.,Croft, W. B, (2016). Estimating embedding vectors for queries.ICTIR 2016 - Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval, pp.123–132.

[70] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018, Accessed: Aug. 04, 2022. [Online]. Available: http://arxiv.org/abs/1810.04805

[71] A. Parnami and M. Lee, “Learning from Few Examples: A Summary of Approaches to Few-Shot Learning,” Mar. 2022, Accessed: Aug. 18, 2022. [Online]. Available: http://arxiv.org/abs/2203.04291

[72] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” Jan. 2013, Accessed: Aug. 04, 2022. [Online]. Available: http://arxiv.org/abs/1301.3781

[73] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, “Advances in Pre-Training Distributed Word Representations,” Dec. 2017, Accessed: Aug. 03, 2022. [Online]. Available: http://arxiv.org/abs/1712.09405

[74] D. Chen, A. Fisch, J. Weston, and A. Bordes, “Reading Wikipedia to Answer Open-Domain Questions,” Mar. 2017, Accessed: Aug. 09, 2022. [Online]. Available: http://arxiv.org/abs/1704.00051

[75] R. Speer, J. Chin, and C. Havasi, “ConceptNet 5.5: An Open Multilingual Graph of General Knowledge,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.11164.

[76] K. Halder, A. Akbik, J. Krapac, and R. Vollgraf, “Task-Aware Representation of Sentences for Generic Text Classification,” Proceedings of the 28th International Conference on Computational Linguistics, 2020, doi: 10.18653/v1/2020.coling-main.285.

[77] M. Huisman, J. N. van Rijn, and A. Plaat, “A survey of deep meta-learning,” Artificial Intelligence Review, vol. 54, no. 6, pp. 4483–4541, Apr. 2021, doi: 10.1007/s10462-021-10004-4.

[78] K. Cherry, “Behavior Analysis in Psychology,” Verywell Mind, Aug. 06, 2006. `https://www.verywellmind.com/what-is-behavior-analysis-2794865` (accessed May 10, 2022).

[79] “Dominant language - Amazon Comprehend,” docs.aws.amazon.com. `https://docs.aws.amazon.com/comprehend/latest/dg/how-languages.html` (accessed Jun. 23, 2022)

[80] “IBM Watson Natural Language Understanding - FAQ,” www.ibm.com. `https://www.ibm.com/cloud/watson-natural-language-understanding/faq#:~:text=View%20What%20language%20does%20NLU` (accessed Jun. 23, 2022)