electronics

Article

Using a Profiling System to Recommend Employees to Carry Out a Project

Abderrahim Rafae * and Mohammed Erritali

Data4earth Laboratory, Sultan Moulay Slimane University, Beni Mellal 23000, Morocco; m.erritali@usms.ma
* Correspondence: rafae.abderrahim@gmail.com

Abstract: In this paper, we introduce the application of a profiling system to suggest appropriate employee profiles for project assignments based on task specifications. The primary objective of this system is to assist managers in gaining a comprehensive understanding of their employees' profiles and motivations. Our research introduces a recommendation system that relies on a profiling approach, analyzing messages and publications shared within a professional network. The proposed system is composed of two main components. The first component focuses on profiling, extracting relevant information from the company's Human resources (HR) data, identifying interests, and establishing a psychological profile from publications exchanged within the professional platform. The second component is dedicated to recommending profiles that closely align with the specific requirements of each project. Our system yields promising results in predicting favored candidates for projects, achieving an accuracy of 0.92 and an F-score of 0.94. By integrating message-based profiling and leveraging data from professional networks, our approach proves to be effective in recommending well-suited candidates for various projects.

Keywords: recommendation system; profiling; interests; psychological profile

Citation: Rafae, A.; Errital, M. Using a Profiling System to Recommend Employees to Carry Out a Project. *Electronics* **2023**, *12*, 3388. https://doi.org/10.3390/electronics12163388

Academic Editors: Dimitris Apostolou, Gurjot Singh Gaba, Yassine Sadqi, Najlae Idrissi and Abdul Wahid

Received: 2 July 2023
Revised: 28 July 2023
Accepted: 8 August 2023
Published: 9 August 2023

1. Introduction

In the quest for continuous improvement and enhanced profitability, businesses have embarked on a transformative journey to optimize their management structures and explore innovative solutions. One pivotal aspect of this evolution involves identifying ways to improve the productivity of their employees through effective collaboration and interaction within clusters. The creation of a conducive environment that fosters teamwork and trust yields numerous advantages, including heightened productivity, stimulated innovation, efficient problem-solving, and strengthened team dynamics. Consequently, successful project management necessitates the implementation of a sound employee selection process.

Traditionally, employee selection for projects relied heavily on leaders' subjective judgments, with emphasis placed on an individual's perceived confidence and skill set to handle the tasks at hand [1]. However, this approach had its limitations, as human confidence remained subjective and lacked concrete practical foundations. Subsequently, businesses began developing human resources (HR) processes that utilized competency evaluations and limited historical data to propose employee recommendations for project assignments. Nonetheless, this method often resulted in sub-optimal team compositions and less-than-optimal project outcomes.

Our conviction lies in the notion that, while competencies and experience are essential attributes for project success, these alone are not sufficient. Beyond the acquisition of skills over time, an employee's intrinsic motivation and determination to overcome challenges play a critical role. As authors Dyer Jr, W. G., and Dyer, J. H. [2] aptly stated, "For a team to succeed, its members need two things: (1) the skills to accomplish goals laid out for the team, and (2) "fire in the belly", that is, the motivation to succeed".

In response to this pressing challenge, we propose the implementation of a profiling system that capitalizes on advanced data analytics and machine learning techniques to recommend employees for project execution. The profiling system seeks to streamline the employee selection process by providing data-driven insights into the characteristics, skills, experiences, and expertise of individual employees, while also uncovering their centers of interest and gaining insight into their psychological motivations. By harnessing the power of Natural Language Processing (NLP), the system will empower organizations to make informed decisions in forming project teams, thereby enhancing project performance and overall organizational success.

The primary objective of this study is to design and develop an innovative profiling system capable of accurately evaluating and matching employees to specific project roles based on their unique centers of interest and individual competencies. To achieve this goal, we propose the integration of diverse data sources, including HR data, and the analysis of employees' messages and publications in their professional networks within the organization.

Through a fusion of data analytics algorithms and machine learning, our proposed solution will enable the profiling system to identify patterns, correlations, and insights that facilitate optimal employee-to-project assignments. The system's intelligent recommendation engine will consider various project requirements, employee interests, and individual profiles, ensuring the formation of balanced and high-performing project teams.

In this paper, we present an in-depth study of the utilization of a profiling system to recommend employees for project assignments. We commence with a comprehensive state-of-the-art analysis, exploring existing methodologies and frameworks related to employee selection and profiling in project management. This review reveals the limitations of current practices and highlights the need for a more systematic, data-driven approach to employee recommendations. In conclusion, this paper introduces a novel contribution to the field of project management by presenting a data-driven profiling system capable of enhancing employee selection for project execution. By empowering organizations to assemble well-matched and motivated project teams, the proposed solution has the potential to revolutionize project management practices, leading to higher success rates and improved organizational performance. In the subsequent parts of this paper, we will delve into different aspects of our research. Section 2 offers an overview of the current state of the field, exploring existing methodologies and frameworks. In Section 3, we introduce our proposed model in detail. Moving forward, Section 4 presents the outcomes of our study, shedding light on the significance and advantages of our model. Finally, we conclude our research in Section 5, providing valuable insights into the importance and impact of our profiling system in the realm of project management.

2. Related Work

The team is the fundamental pillar of any project, and the judicious selection of its members is of paramount importance in the optimal management of a project. Forming an effective team within the organization becomes one of the management's top priorities, as it is a critical condition for the success of the company [3]. Indeed, a well-constituted team possesses the necessary strengths to face challenges, achieve set objectives, and exceed expectations. Therefore, it is crucial to carefully select the team, with meticulous consideration of collaborators' skills and personalities, to ensure the success of a project within a company. This guarantees a high-performing, cohesive, and motivated team capable of overcoming challenges and delivering exceptional results, thereby reinforcing the positive image of the company in the market.

In the past, it was not uncommon for the process of selecting collaborators for project tasks to be made arbitrarily, primarily based on the project manager's subjective ideas and preferences. This senseless and unmethodical selection approach could lead to numerous problems and risks for the project and the entire organization. Choosing collaborators without following a structured approach or objective criteria could result in ill-matched teams, where the necessary skills and experiences to accomplish tasks were lacking. This could lead to reduced operational efficiency, as team members might not be able to fulfill their responsibilities optimally. Moreover, when collaborators are selected arbitrarily, they might feel a lack of recognition and appreciation for their skills and expertise. As a consequence, their motivation and commitment to the project would diminish, leading to a decline in the quality of work delivered. Additionally, the absence of an objective selection process could trigger conflicts within the team. Collaborators might perceive inequalities or unfair treatment, creating tensions and harming team collaboration and cohesion.

To avoid these problems, companies have gradually adopted more systematic and equitable approaches to selecting collaborators for their projects, implementing more structured Human Resources (HR) management processes based on collaborators' skills. Using collaborators' Curriculum Vitae (CV) to evaluate their skills and qualifications offers an objective and transparent approach to forming high-performing teams tailored to the project's needs [4,5].

However, recruiter assessments are not entirely reliable. Moreover, with a large number of CVs received during the selection processes to find the right collaborators for projects, it has become impossible for project managers to thoroughly examine all the details in a limited time. Therefore, despite the use of CVs, there may still be limitations in project managers' ability to fully identify the skills and aptitudes of each candidate. Hence, automating this process becomes essential. One study relies on objective criteria extracted from candidates' LinkedIn profiles and subjective criteria extracted from their social presence to estimate candidates' relevance scores and deduce their personality traits. The ranking of candidates is based on machine learning algorithms that learn the scoring function from training data provided by human recruiters. However, this approach primarily focuses on technical skills based on LinkedIn profiles, potentially overlooking other important qualities such as interpersonal skills or adaptability [6].

Nevertheless, even with the automation of selection processes based on objective criteria, a crucial aspect that may be outsourced or poorly accounted for in collaborator profiles is the psychological aspect. Technical skills and professional experience can be relatively objectively evaluated based on CVs and online platforms, but assessing personality traits, soft skills, and cultural compatibility remains more complex. A multi-agent team formation mechanism is proposed for a task, considering parameters such as team cardinality, willingness, capability, trust, and reciprocity of the agent. The system is simulated with a case study of forming a virtual medical council using JADE, and the results show that our proposed algorithm produces an efficient team of expert agents from a community of agents best suited to the task. Team efficiency is determined by the average willingness, trust, and reciprocity of the team as a whole. By using specific parameters for each agent, the approach allows customizing the formation of each team based on each member's unique characteristics and skills. This can lead to teams better suited for the specific task at hand. However, while the approach may be effective in the specific context of the case study, its ability to generalize to other scenarios or application domains remains to be verified. The approach's effectiveness may vary depending on specific tasks and contexts [7].

Several studies have been proposed on this subject, and others have integrated the profiling system into the selection process.

A profiling system based on the DISC theory divides a person's personality based on dominance, influence, steadiness, and compliance. Based on a sample used for five Facebook and Twitter accounts, this approach can complement traditional recruitment methods by providing additional information about candidates, allowing for a more comprehensive evaluation. However, the sample of five Facebook and Twitter accounts used in the study may not represent the entire population of potential candidates. The results obtained from this sample may not apply universally to a larger population of candidates. Additionally, while the DISC theory is a popular tool for personality profiling, it has some limitations. Human personality is complex and multidimensional, and reducing it to only four dimensions may not capture its full diversity [8]. It is also a system that contributes to suggesting employee engagement in the organizational environment. It aims to understand the impact of an employee's demographic profile on their engagement within the organization to improve productivity by applying better strategies, using logistic regression techniques on a data set collected from questionnaires. Logistic regression is a powerful statistical method for modeling relationships between variables, allowing for meaningful predictions and insights into employee engagement. However, relying solely on an employee's demographic profile to assess their engagement may oversimplify the complexity of personality and individual motivations. Other factors, such as values, interests, professional aspirations, and the work environment, may also influence engagement. Furthermore, while logistic regression is useful for modeling relationships between variables, it cannot explain all the complexities of employee engagement. Other analysis methods, such as machine learning, may be required to deepen the analysis [9]. On the other hand, research proposes problem-solving methods to determine the best employees to reward using the profile-matching method. This evaluation depends on work discipline and administration, reliability, maturity, personal integrity, morale, communication, work principles, quantity and quality of work, and interest in interconnected organizations. The approach can quickly provide a list of potential candidates, which can be useful for optimizing the rewards and talent management process. However, by focusing solely on the defined criteria, the approach may neglect other important factors that influence employee performance, such as the work context, organizational environment, development opportunities, etc. [10]. It is also found that poor work allocation can lead to less-than-optimal organizational performance. Placement within an organization is mainly made subjectively. A study proposes applying the profile-matching method to determine the suitability of professional placements based on employees' skills. This placement adequacy is based on eight criteria, namely, years of service, level of education, professional experience in the structural domain, suitability of the scientific domain with the work, professional experience in the academic domain, professional experience in the non-academic domain, academic rank, and teaching experience. While it can provide recommendations on job compatibility with employees' profiles, this research is still limited to employees at the structural level of deans, program heads, and directors. Thus, the eight criteria do not consider individuals' aspirations, passions, and intrinsic motivations, which can play a crucial role in their job satisfaction and organizational engagement. Neglecting these aspects may result in placing some employees in positions that may match their professional skills but are not aligned with their interests and aspirations [11]. A new approach to behavioral profiling is proposed to analyze and better understand employees. This method exploits vast amounts of data collected in real-time from employees' daily work, using the WeChat Robotic Platform, which is well designed to collect data on employees' daily professional behavior, transfer raw data into structured records, and obtain key features through data analysis. This approach is likely to be more objective as it relies on concrete data and observable facts rather than subjective evaluations. However, while the We-chat Platform may be suitable for data collection, it may not be entirely suitable for all contexts or types of organizations. Additionally, this approach primarily focuses on daily professional behavior, which may reduce the complexity and richness of employees' skills and characteristics [12].

Moreover, an approach is offered with an alternative to sorting and merging vast amounts of data in relational databases. It uses a top-k query algorithm to produce a reduced result containing only the k best carefully ranked elements. The use of the top-k query algorithm allows for quickly generating a reduced result, which can be advantageous when dealing with large amounts of data. However, the choice of the value of k can significantly influence the results. An inappropriate value of k can lead to biased or incomplete results. Furthermore, the approach may not be suitable for all types of data or queries, limiting its versatility in some scenarios [13].

Various studies have tried to incorporate other criteria and techniques for team formation. An article illustrates methods of using ranked preference lists of students/instructors/mentors for project selections to form teams during the 2018/2019 academic year, using the Gale–Shapely algorithm. The Gale–Shapely algorithm can be used as an efficient team formation algorithm based on the collected data. Since most students are placed in teams they prefer, even when project-side preference data are random, and efficiency increases with team size, instructors using this tool in the context of CATME will create teams of students with a strong preference for the project they were assigned. The approach is based on ranked preference lists of students, instructors, and mentors, allowing for the consideration of individual choices in an ordered and structured manner. Moreover, the Gale–Shapely algorithm can be used in various situations where team formation is necessary, whether for academic projects or other professional contexts. However, the approach focuses on students' preferences for projects, but it does not consider the specific skills of team members, which could be important for project success [14].

Furthermore, a TF-HMM (Team Formation with HMM: Hidden Markov Models) solution to the problem of team formation within a work process, aiming to improve execution efficiency, is presented. Individual expertise and transfer relationships are taken into consideration, providing an appropriate modeling method. Experiments conducted on a real data set showed that TF-HMM is more efficient and practical. The TF-HMM approach considers both the individual expertise of team members and transfers relationships between them, providing a better understanding of each member's skills and how they can be shared and used effectively within the team. However, using hidden Markov models may be computationally resource-intensive, which could increase the time required for team formation. This could be problematic in environments where decisions need to be made quickly. Additionally, if work conditions or team members change frequently, the TF-HMM approach may require continuous adaptation to remain relevant. Updating models and data could be a challenge [15].

The various approaches for collaborator selection and profiling have strengths and weaknesses as already mentioned. None of the mentioned approaches can perfectly capture all dimensions of a candidate or team. One of the essential aspects often overlooked is the importance of individuals' aspirations, passions, and intrinsic motivations. These factors play a crucial role in job satisfaction and organizational engagement. Therefore, it is essential to consider these aspects in the collaborator selection and profiling process, as we will elaborate on in our study.

3. Proposed Model and Algorithms

The selection of team members in the context of project execution is a crucial factor in ensuring project success. The effectiveness of a team is greatly enhanced when the right individuals are assigned tasks that align with their strengths, particularly if the tasks align with their interests and present challenges that stimulate them.

In this research, the methodological needs revolve around the construction of a profiling system, which forms the core of our approach. We began by constructing a test corpus, which were utilized throughout the various stages of developing our profiling system. This corpus was composed of two main data sets: the HR data set and the data set containing messages exchanged within the professional network of the organization.

Upon the completion of the corpus design, we proceeded to extract the emotions and centers of interest of the collaborators from their messages. By combining these two metrics with the HR data of each collaborator, we could generate comprehensive profiles for each individual. This integration allowed us to obtain a holistic understanding of each collaborator's characteristics.

To conclude our study, we analyzed the matching between the constructed profiles and the project's specifications outlined in the project brief. This analysis aimed to recommend a ranked list of favored collaborators for team formation to undertake the project in descending order of suitability. The proposed architecture for our user profiling method based on professional social networks is presented in Figure 1.

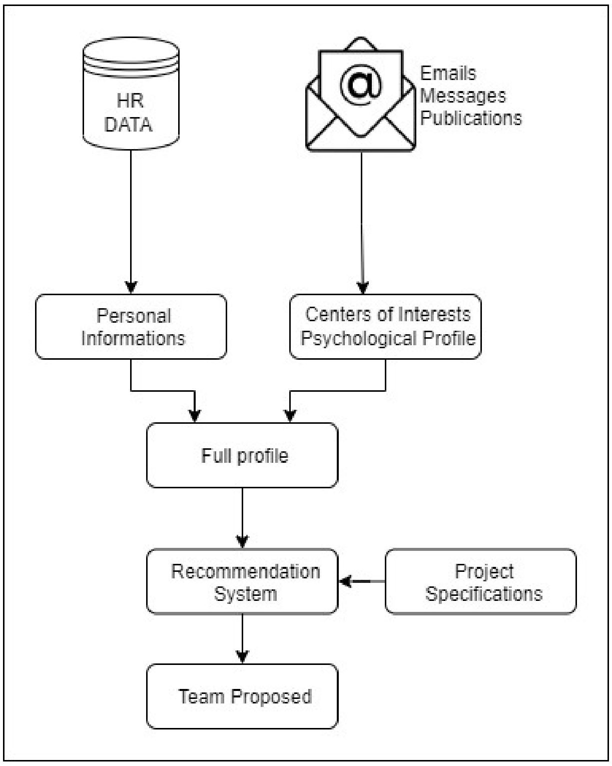

Figure 1. The architecture of the methodological needs of the proposed recommendation system.

In this study, our project endeavors to address the challenge of team member selection, specifically to align with the project's requirements. To achieve this goal, we leveraged the capabilities of a profiling system designed to extract the centers of interest of platform users from digital messages and emails exchanged on an online professional platform. However, before delving into this aspect, our initial step involved constructing a comprehensive database tailored to the specific requirements of our research. This database plays a pivotal role in supporting various stages of our profiling system's development process.

3.1. Data Collection

As a result of the unavailability of application data, we were faced with the task of seeking out data that closely resemble the application data. This left us with two viable options: the creation of synthetic data or the search for open-source data that exhibit similarities to the application data. In our specific scenario, we collected data that align with the specified characteristics and resorted to generating fabricated data for the missing information. The subsequent list outlines the various steps we undertook to construct our data set.

3.1.1. Data Set Characteristics

In this section, we explore the key features and characteristics of the data set used in this study. Our primary objective revolves around the meticulous selection of project collaborators, leveraging their comprehensive profiles that comprise two crucial components. The initial segment encompasses vital HR data that can be found on the global base of the company (name, age, gender, etc.), while the second part delves into obtaining centers of interest and psychological profiles from users' messages and emails. We required a data set showcasing the messages and pertinent information to achieve this:

- User_Id: The sender's user ID.
- date_send: The date the message was sent.
- Message: The message sent from sender to receiver.
- User_Receiver_Id: The receiver's user ID.

- response_date: The Message reply date.

3.1.2. Open Source Data

The most optimal data for a social application, particularly for predicting links or relationships between users, are derived from applications that share a similar context, such as Facebook and Twitter. Below, we provide insights into the data sets we utilized.

One of the data sets we utilized is the "Customer Support Twitter" data set. This data set consists of directed networks and contains valuable information about Twitter users. The data are presented in a CSV format, where each line represents a tweet. The data set is structured with seven columns, each serving a specific purpose. Notably, each conversation in this data set involves at least one request from a consumer and at least one response from a company. The data set contains a substantial 2,811,774 lines and 7 columns [16]. Furthermore, we generated the HR data set using Python's random generation functionality.

3.1.3. Data Augmentation

During this phase, we leveraged the tweet data set to generate synthetic data based on the real data collected. Subsequently, the data obtained from the customer support tweets data set were utilized to construct our message database. To achieve this, we proceeded through a series of sequential steps as outlined in Figure 2.

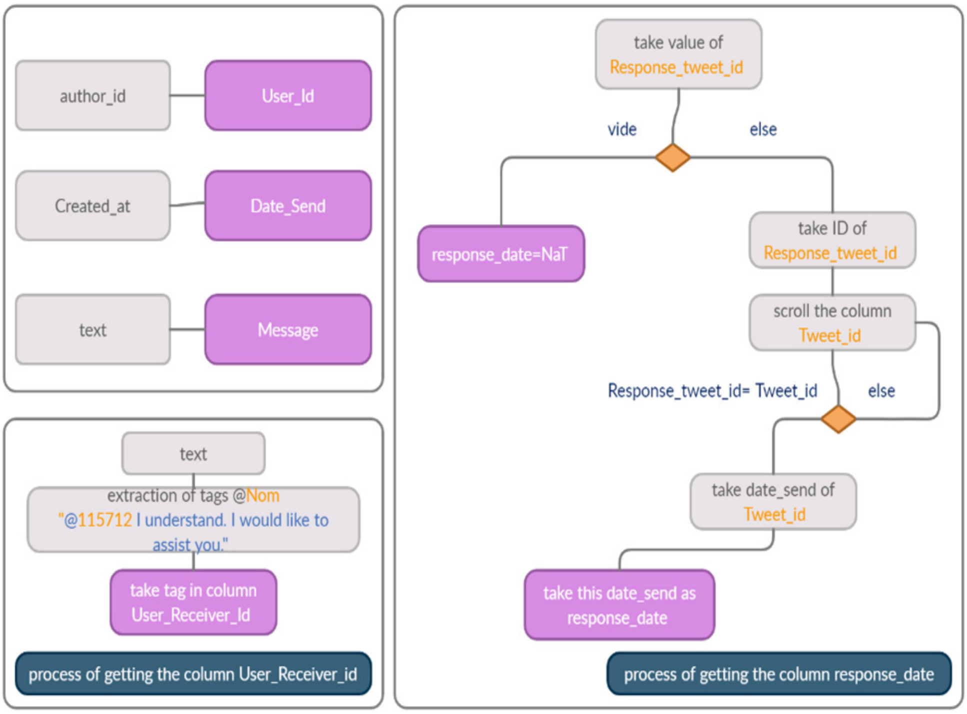

Figure 2. Data collection.

After completing the entire process, we successfully acquired our message database from the tweets data set.

However, we found that the resulting messages generated from tweets were inadequate and lacked meaningful content. Since our focus is on professional messages, it became evident that we needed to obtain data from a more reliable source. To address this, we decided to compile a collection of messages from Wikipedia, which is known for its professional content.

To achieve this, we employed web scraping techniques using the Wikipedia Python library. This enabled us to retrieve relevant articles on specific topics from the list of previously defined topics. For message extraction, we adopted a sentence-level approach, treating each sentence in the article as an individual message.

As a result of this approach, we obtained satisfactory results in terms of obtaining professional and meaningful messages from Wikipedia data. The use of web scraping and the Wikipedia Python library proved to be effective in meeting our requirements for a high-quality message database.

3.1.4. Data Pre-Processing

In the analysis of publications and messages, data cleaning plays a crucial role, serving as a potent tool for handling textual data in general. In this study, we have devised a pre-processing function tailored to the specific characteristics of the messages we collected, with a particular focus on preserving semantic significance. The pre-processing function can be divided into two main parts: one dedicated to the column's date send and response date, and the other dedicated to the content of the message column, as shown in Figure 3.

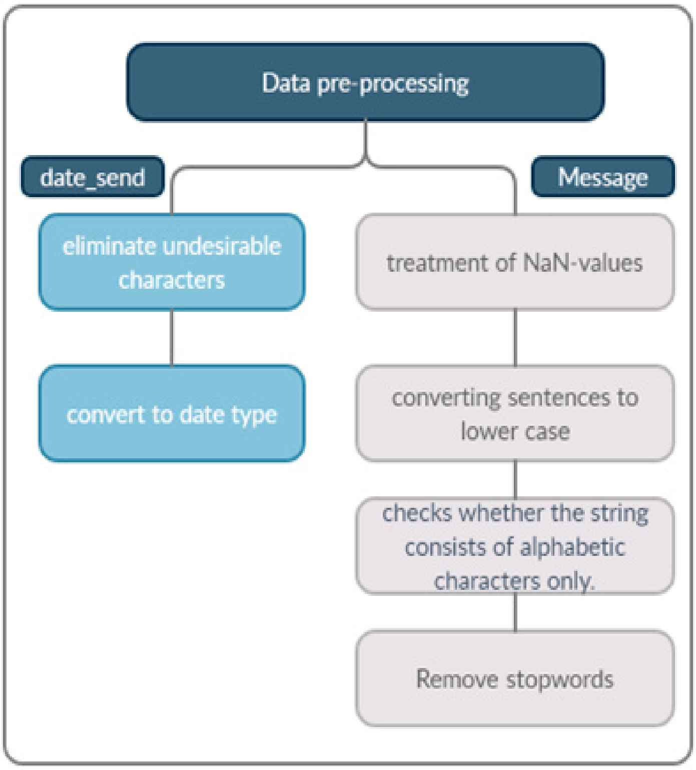

Figure 3. Data Cleaning.

3.2. Psychological Profile

In this section, we focus on the analysis of the psychological profile of users, specifically by extracting the top emotions from the text generated by each user. To accomplish this, we employed EmoNet [17], a neural network tool for emotion recognition, which predicts emotions from text and categorizes them into eight primary emotions: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation, as illustrated in Figure 4.

However, our research further developed these emotion categories to align with the 24 emotions defined by the Plutchik wheel. The Plutchik wheel expands on the basic emotions and encompasses a broader spectrum of emotional states, as presented in Figure 5, taking into account the complex and nuanced nature of human emotions.

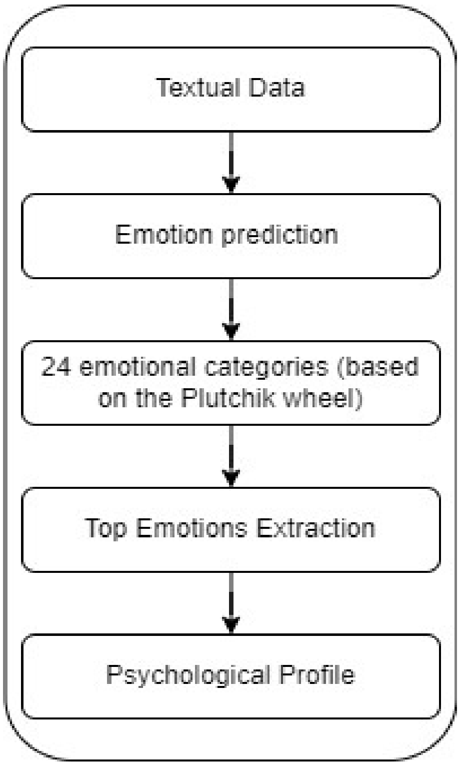

Figure 4. Psychological Profile Process.

Figure 5. Plutchik's wheel of emotion.

By utilizing EmoNet and extending its categories to the 24 emotions based on the Plutchik wheel, we aimed to capture a more comprehensive understanding of users' psychological profiles. This allowed us to delve beyond the basic emotional states and gain insights into the subtle variations and combinations of emotions that users express in their text.

The EmoNet tool analyzes the textual data provided by users and assigns a score or intensity to each of the 24 emotional categories. These scores indicate the strength or prominence of each emotion within the user's text. By aggregating the scores across all the text generated by a user, we can determine their top emotions—the emotions that appear with the highest intensity.

This analysis of the psychological profile through emotion extraction provides valuable insights into the emotional disposition and tendencies of users. It offers a deeper understanding of their emotional states, preferences, and inclinations, which can be useful in various applications such as personalized recommendations, targeted marketing strategies, and understanding user behavior.

In essence, through the utilization of EmoNet and the expansion of its emotion categories to encompass the 24 emotions outlined in the Plutchik wheel, we can effectively identify the primary emotions expressed in users' text, thereby constructing a comprehensive psychological profile. This analytical approach enriches our comprehension of users' emotional states at the moment of writing their messages, shedding light on their sentiments towards the subject matter or the individual to whom the message is directed.

3.3. Centers of Interests

In the quest to identify the motivations and elements that can passionately drive a collaborator to integrate and devote their utmost efforts towards achieving their goals, as Steve Jobs once stated, "Passion is the force that propels us forward, even when everything seems impossible". We believe that these motivations are synonymous with a collaborator's centers of interest. Specifically, we posit that each message written by a user, conveying positive emotions, can be considered an indication of the subjects that genuinely interest the sender.

To determine a person's centers of interest, we defined them as their favorite topics and those most frequently repeated in their messages and publications, particularly those conveying positive emotions. To initiate this process, we collected all messages containing positive emotions from the user whose centers of interest we sought to discover. Subsequently, we observed the occurrences of each topic in their messages to compile a list of centers of interest, ranked in descending order based on the high frequencies of the topics, until we reached the final topic, as depicted in Figure 6.

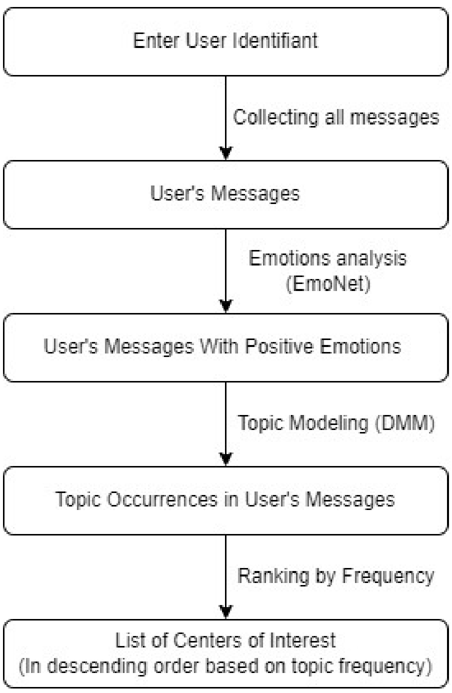

Figure 6. Center of Interest Extraction.

3.3.1. Topic Modeling

Topic modeling is a technique used to extract topics from an unlabeled collection of text in order to reveal the underlying subject or structure of the corpus. Instead of labeling the documents beforehand, topic modeling provides an interpretable representation of the content. Researchers are exploring novel approaches using various algorithms in the realm of short-text topic modeling. They trace the historical progress in this field and propose a taxonomy of methods for modeling subjects in short texts, including multinomial mixtures and self-aggregation-based methods. The taxonomy categorizes topic modeling approaches into three main classes:

- Dirichlet Multinomial Mixture (DMM) [18] Model: This method assumes that each text originates from a unique latent topic. Several enhancements to DMM have been suggested, such as Gibbs collapsed sampling or the Twitter-LDA model, to improve its performance.
- Global Word Co-occurrences: Certain models leverage the proximity of words in the corpus to gauge their relevance. These approaches use rich global word co-occurrence models, employing a sliding window to retrieve word occurrences.
- Self-aggregation Methods: To address scarcity in short texts, these methods merge multiple short texts into larger pseudo-documents (P). The recently developed Self-Aggregation-based Topic Model (SATM) [19] combines topic clustering and modeling simultaneously. Short texts are fused into pseudo-long documents before inferring topics. Probabilistic Topic Modeling (PTM) [20] is a Bayesian approach that summarizes data, such as text, using a small set of latent variables representing the underlying themes or topics. It is a statistical generative model that represents documents as a mixture of probabilistic topics and topics as a mixture of words.

We opted for the Dirichlet Mixture Model (DMM) as our basis, which is a generative probabilistic model with the hypothesis that a document is generated from a unique subject, i.e., all words in a document are generated using the same thematic distribution.

3.3.2. Comparison of Topic Modeling Algorithms

The issue of the rating of short text topic patterns is still an open question. Several measures have been suggested to measure the coherence of subjects in texts [21,22]. While some metrics are generally appropriate for long texts, they may be questionable for short texts [23]. Many of the conventional measures (e.g., perplexity) try to approximate the likelihood of retaining test data based on some parameters from training data. However, this probability does not necessarily give a strong indication of the quality of the subjects that are retrieved [24]. To provide a good assessment of our situation, we evaluated all the patterns below as the text forms using the STTM (Short Text Topic Modeling) library on Java. Before starting the comparison, we built a labeled database from the data set of messages from which we created a topics column ourselves. After that, we applied the 4 algorithms to the message column from which we obtained the following result. Images-y presents a part of the labeled data set with the columns of the algorithms we want to be compared (Figure 7).

Once we built our database and extract topics from the messages using the four algorithms, the next step was to select the most suitable algorithm for our use case. For this purpose, we evaluated the algorithms based on three common metrics: recall, accuracy, and F1-Score. These metrics are widely used to assess the performance of classification models, including machine learning and automatic learning models. They provide valuable insights into different aspects of the model's prediction quality and measure its ability to accurately classify the data. Below are the equations for calculating these metrics.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Samples}} \tag{1}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

Upon applying the metrics, the following results were obtained:

	Modèle	Accuracy	F1-score	Recall
0	DMM	0.878788	0.902951	0.878788
1	LDA	0.616162	0.629550	0.616162
2	BTM	0.616162	0.629550	0.616162
3	PTM	0.616162	0.629550	0.616162

Figure 7. Comparison of topic modeling algorithms by the accuracy, recall, F1-score.

Based on the evaluation results of the four algorithms concerning accuracy, recall, and F-score, as shown in Figure 7. We observed that DMM performed well in modeling short-text subjects. Therefore, we decided to utilize DMM as the foundation for extracting topics from the messages in our study.

3.4. Recommendation System

In this section, we outline the functionality of our recommendation system, which focuses on extracting the required skills from project specifications and analyzing the correspondence of these skills with the profiles of the collaborators we have constructed. This process enables us to generate a list of profiles ranked according to their level of correspondence with the project requirements, as illustrated in Figure 8.

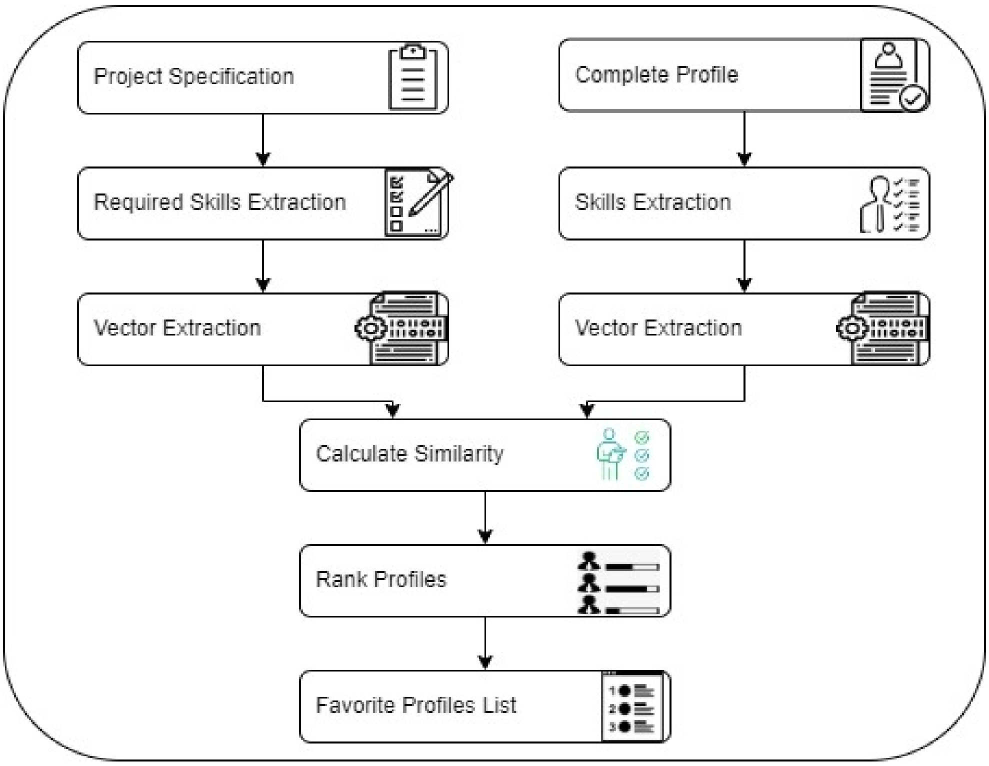

Figure 8. Recommendation System Process.

Our recommendation system serves as a valuable tool for managers, providing them with a comprehensive and ordered list of employees based on their profile alignment with project specifications. By extracting the necessary skills from the project requirements, our system identifies the key competencies and qualifications required for successful project execution.

We chose to represent Skills by converting the skills of employees and the required skills of the project into numerical vectors. Each skill can be represented as a binary value (1 if the employee possesses the skill; 0 otherwise) or as a numerical value indicating the level of proficiency.

We used the cosine similarity formula to calculate the similarity between the skills vector of each employee and the required skills vector of the project. The cosine similarity value indicates the degree of match between the employee's skills and the project requirements.

We sorted the employees' profiles based on their cosine similarity scores in descending order. Profiles with higher similarity scores indicate a closer match to the project's skill requirements. The recommendation system then generates a list of profiles, sorted in descending order of correspondence.

This sorted list offers managers a clear vision of the complete profiles of employees who are potential candidates for the project. It enables them to easily identify and prioritize employees whose skills closely align with the project's specific needs. By having access to this ordered list, managers can make informed decisions and effectively allocate employees to tasks that are most suitable for their skill sets and expertise.

The recommendation system's ability to provide managers with a ranked list of employees enhances the efficiency and effectiveness of the employee assignment process. It eliminates the need for managers to manually review individual profiles and simplifies the decision-making process. This system empowers managers to optimize project team composition and enhance project outcomes by selecting employees with the most relevant skills and qualifications.

In summary, our recommendation system extracts the required skills from project specifications and compares them with the profiles of employees. The resulting list of profiles, ordered by correspondence, enables managers to gain a clear understanding of each employee's complete profile and make well-informed decisions in the assignment process. By leveraging this system, managers can efficiently match employees to projects, ensuring a higher likelihood of success and maximizing the utilization of available skills within the organization.

4. Results and Discussion

In this section, we present the results obtained from our research and discuss the implications and limitations of our profiling system for team member selection in project execution.

4.1. Experiment

Test Environment: For the development of our system, we used Python as our programming tool and FastAPI to create an easy-to-use and efficient model, allowing it to be integrated seamlessly into the professional networks of various companies.

Our experiments were divided into two parts:

4.1.1. Extracting User Profiles

We initiated the experiment by extracting the profile of each user based on the corpus constructed in Section 3.1. This involved extracting the subject and emotions from each message, as illustrated in Figure 9.

User_Id	115712
date_send	2017-10-31 21:49:35+00:00
Message	im happy to use machine learning in our project
User_Receiver_Id	sprintcare
response_date	*31/10/2017 17:54:49*
topic	machine learning
Emotion	joy

Figure 9. Emotions and Subject Analysis Extracted from Messages.

After extracting the subject and main emotion from each message, as shown in the previous Figure 9, this message presented "machine learning" as the subject and "joy" as the emotion. We then moved on to the second step of extracting a psychological profile based on the majority of messages from each collaborator, as we can see in the following JSON results.

'Psychological profile': 'joy': 0.8571428571428571, 'serenity': 0.8571428571428571, 'ecstasy': 0.8571428571428571, 'Positive': 1.0

We can observe from the last example that this user has a positive psychological profile, which leads us to the next step. Here, we have identified the subjects of messages containing positive emotions and sorted them in ascending order to obtain the results presented in the following JSON output.

'Center of interest': 'machine learning': 0.3333333333333333, 'study mathematical': 0.16666666666666666,'proposal characteristics': 0.16666666666666666, 'definition tasks': 0.16666666666666666, 'application business': 0.16666666666666666

Furthermore, to complete the extraction of each collaborator's profile, we integrated their interests and psychological profile with the company's HR data. This resulted in a comprehensive profile generated by FastAPI, in the form of a JSON file, as shown in the following example:

"115712": "Date Between": "Date debut": "2011-08-29", "Date fin": "2017-12-02",
"Center of interest": "machine learning": 0.3333333333333333, "artificial intelligence": 0.16666666666666666, "definition tasks": 0.16666666666666666, "deep learning": 0.16666666666666666,
"Psychological profile": "joy": 0.8571428571428571, "serenity": 0.8571428571428571, "ecstasy": 0.8571428571428571,
"HR DATA": "age": 27, "gender": "Male", "Family situation": "Single", "nationality": "Serbia", "years of experience": 4

In the last example, we were able to obtain a comprehensive profile of this collaborator, which we utilize in the next section.

4.1.2. Proposed List of Project Members

After constructing the candidates' profiles, our next step involved studying the profiles that closely match the project specifications to predict the profiles that best fit this project. To achieve this, we employed cosine similarity, which allows us to rank the profiles in descending order based on their highest similarity percentage. As we can observe from the provided JSON example, it represents a project specification entry in the project database. Each line corresponds to the specifications of a specific project.

"Proj1": "Domain of interest": "machine learning", "artificial intelligence", "python", "deep learning",
"preferences": "age range ": 23–30, "gender": "Male", "Female", "Family situation": "Single","Married", "nationality": "Serbia", "years of experience range": "+3"

After calculating the cosine similarity between Project 1 specifications and User 115712, we obtained a percentage that classifies them among the favored members to undertake this project. This is because User 115712 possesses centers of interest that closely align with the project's domain requirements and also meets the conditions related to age and years of experience, making them a perfect candidate for this project.

For the evaluation of our module, we did not find labeled data that directly correspond to our study. Therefore, we created a sample data set consisting of 20 projects with their specifications and a column representing the top-rated candidate favored for each project. Additionally, we gathered a sample of 50 candidate profiles. In this data set, we achieved an accuracy of 0.92 and an F1-score of 0.94. However, it is important to note that these results are not entirely conclusive as we relied on a manually labeled data set created by ourselves. We look forward to obtaining real-world data to test our approach in a more realistic setting.

4.2. Discussion: Comparison of the Proposed Approach with the Literature

The proposed approach to employee selection for project execution holds paramount importance in contrast to the traditional methods outlined in the existing literature. While businesses have long relied on subjective judgments and limited historical data to recommend employees for projects, the proposed profiling system leverages advanced data analytics and machine learning techniques to revolutionize this process. This comparison section highlights the significance of the proposed approach by outlining its key advantages over traditional practices.

4.2.1. Data-Driven Decision Making

The traditional approach to employee selection primarily relied on subjective evaluations by leaders, leading to potential biases and sub-optimal team compositions. In contrast, the proposed profiling system adopts a data-driven approach, integrating diverse data sources, including HR data, and analyzing employees' messages and professional network publications. This data-driven decision-making process ensures a more accurate evaluation of individual employees, resulting in well-informed project team formations.

4.2.2. Intrinsic Motivation and Determination

While competencies and experience are essential attributes, the proposed approach recognizes the significance of an employee's intrinsic motivation and determination to overcome challenges. The literature has previously emphasized the importance of both skills and motivation for team success. By incorporating insights into employees' centers of interest and psychological motivations, the proposed system ensures a more comprehensive evaluation, leading to the formation of highly motivated and productive project teams.

4.2.3. Natural Language Processing (NLP) Empowerment

The proposed approach harnesses the power of NLP to analyze and gain insights from employees' messages and publications. This advanced technology empowers organizations to delve deeper into employees' interests and expertise, uncovering hidden talents and potential matches for specific project roles. Traditional methods lacked such sophisticated analyses, limiting the depth of understanding and potential for optimal employee-to-project assignments.

4.2.4. Enhanced Project Performance

Traditional employee selection methods often resulted in sub-optimal team compositions and less-than-optimal project outcomes. In contrast, the proposed profiling system's intelligent recommendation engine considers various project requirements, employee interests, and individual profiles to ensure the formation of balanced and high-performing project teams. This emphasis on team compatibility and motivation significantly enhances project performance and overall organizational success.

4.2.5. Revolutionary Impact on Project Management

The literature review highlights the limitations of current practices and the need for a more systematic, data-driven approach to employee recommendations. The proposed approach aligns with this need and introduces a novel contribution to the field of project management. By revolutionizing the way organizations select employees for projects, the proposed system has the potential to drastically improve project success rates and elevate organizational performance.

4.3. Limits and Future Work

This paper proposes an approach aimed at providing a comprehensive selection of profiles by identifying centers of interest from messages with positive emotions. However, there are differences in the levels of interest, and we plan to develop a method that defines interest categories based on extracted emotions, such as admiration and acceptance, in addition to the two positive emotions. For example, admiration may have a higher degree of interest than acceptance, and we intend to establish interest levels based on emotions. Our profiling system, based on this approach, aims to address the gaps in the literature regarding project team member selection. However, we recognize that there may be challenges in team dynamics if the selected members are not harmonious, which could affect task correlations. Therefore, we believe it is essential to pay attention to the interpersonal relationships among team members. To further enhance our recommendation system, we plan to incorporate an analysis of the relational aspect.

This addition aims to provide insights into the collaboration dynamics among employees and enable recommendations based on team member compatibility. By identifying employees who can collaborate effectively, our system strives to ensure unity and efficiency within project teams.

Integrating the relational profile aspect holds great promise for improving team collaboration and project outcomes. By considering interpersonal dynamics and employee compatibility, managers can form teams that work cohesively and efficiently. Taking relational profiles into account is expected to further enhance the accuracy and success of our recommendation system.

5. Conclusions

In conclusion, this article presents an approach aimed at providing businesses with a solution to assist them in selecting team members for project execution. The proposed solution is based on a recommendation system that offers a list of favored candidates for project realization, relying on the alignment between project specifications and candidate profiles. These profiles are constructed using three key elements: HR data, centers of interest, and emotions extracted from messages exchanged on the company's professional network.

We believe that the centers of interest extracted from the messages represent hidden skills that employees develop discreetly, as they are motivated to work on projects aligned with their passions. We strongly advocate that motivation and passion toward a goal empower individuals to overcome any obstacles they encounter on their path to success. Our model's evaluation of the test sample yielded promising results, with an accuracy of 0.92 and an F-score of 0.94. However, we acknowledge the need to validate our approach with real-world data to further strengthen its reliability.

Moving forward, future work should focus on exploring the possibility of testing our solution on real data to validate its effectiveness. Additionally, improvements can be made by refining the levels of centers of interest based on emotions and studying the relational aspect between candidates to ensure team member compatibility.

In summary, our approach demonstrates the potential in assisting businesses with effective team member selection for project success. By considering hidden competencies and emotional factors, we believe that our solution can contribute significantly to enhancing team dynamics and project outcomes. Further research and experimentation with real-world data will validate the robustness of our approach and open doors for continuous improvement in this area.

Author Contributions: Conceptualization, A.R.; methodology, A.R.; software, A.R.; validation, A.R. and M.E.; formal analysis, A.R.; investigation, A.R.; data curation, A.R.; writing—original draft preparation, A.R.; writing—review and editing, A.R.; visualization, A.R.; supervision, M.E.; project administration, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Aga, D.A.; Noorderhaven, N.; Vallejo, B. Transformational leadership and project success: The mediating role of team-building. *Int. J. Proj. Manag.* **2016**, *34*, 806–818. [CrossRef]
2. Dyer, W.G., Jr.; Dyer, J.H. *Beyond Team Building: How to Build High Performing Teams and the Culture to Support Them*; John Wiley & Sons.: Hoboken, NJ, USA, 2019.
3. Nikitenko, G.V.; Zvyagintseva, O.S.; Sergienko, E.G.; Babkina, O.N.; Chernikova, L.I. Development of human resources of the organization with the help of team-building model. *Calitatea* **2017**, *18*, 132.
4. Cole, M.S.; Rubin, R.S.; Feild, H.S.; Giles, W.F. Recruiters' perceptions and use of applicant résumé information: Screening the recent graduate. *Appl. Psychol.* **2007**, *56*, 319–343. [CrossRef]
5. Smart, K.L. Articulating skills in the job search: Proving by example. *Bus. Commun. Q.* **2004**, *67*, 198–205. [CrossRef]
6. Faliagka, E.; Ramantas, K.; Tsakalidis, A.; Tzimas, G. Application of machine learning algorithms to an online recruitment system. In Proceedings of the International Conference on Internet and Web Applications and Services, Stuttgart, Germany, 27 May–1 June 2012; pp. 215–220.
7. Singh, A.J.; Acharya, S.; Dutta, A. Agent based task specific team formation for effective distributed decision making. In Proceedings of the 2013 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, Thailand, 15–17 May 2013; IEEE: New York City, NY, USA, 2013; pp. 1–6.
8. Utami, E.; Luthfi, E.T. Profiling analysis based on social media for prospective employees recruitment using SVM and Chi-Square. *J. Phys. Conf. Ser.* **2018**, *1140*, 012043.
9. Raza, D.M.; Hasan, F. Employee Engagement and Turnover utilizing Logistic Regression. In Proceedings of the 2021 IEEE 8th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Dehradun, India, 11–13 November 2021; IEEE: New York City, NY, USA, 2021; pp. 1–6.
10. Tanti, L.; Puspasari, R.; Triandi, B. Employee performance assessment with profile matching method. In Proceedings of the 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 7–9 August 2018; IEEE: New York City, NY, USA, 2018; pp. 1–6.
11. Soetanto, H.; Budiyanto, U. Analysis of Job Placement Based on Employee Competency Using Profile Matching. In Proceedings of the 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Jakarta, Indonesia, 6–7 October 2022; IEEE: New York City, NY, USA, 2022; pp. 394–398.
12. Ni, X.; Zeng, S.; Qin, R.; Li, J.; Yuan, Y.; Wang, F.Y. Behavioral profiling for employees using social media: A case study based on wechat. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; IEEE: New York City, NY, USA, 2017; pp. 7725–7730.
13. Paoletti, A.L.; Martinez-Gil, J.; Schewe, K.D. Top-k matching queries for filter-based profile matching in knowledge bases. In Proceedings of the International Conference on Database and Expert Systems Applications, Porto, Portugal, 5–8 September 2016; Springer International Publishing: Cham, Switzerland, 2016; pp. 295–302.
14. Meulbroek, D.; Ferguson, D.; Ohland, M.; Berry, F. Forming more effective teams using catme teammaker and the gale-shapley algorithm. In Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE), Covington, KY, USA, 16–19 October 2019; IEEE: New York City, NY, USA, 2019; pp. 1–5.

15. Lin, S.; Luo, Z.; Yu, Y.; Pan, M. Effective team formation in workflow process context. In Proceedings of the 2013 International Conference on Cloud and Green Computing, Washington, DC, USA, 30 September–2 October 2013; IEEE: New York City, NY, USA, 2013; pp. 508–513.
16. Kaggle. Available online: https://www.kaggle.com/datasets/thoughtvector/customer-support-on-twitter (accessed on 7 August 2023).
17. Abdul-Mageed, M.; Ungar, L. Emonet: Fine-grained emotion detection with gated recurrent neural networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 718–728
18. Nigam, K.; McCallum, A.K.; Thrun, S.; Mitchell, T. Text classification from labeled and unlabeled documents using EM. *Mach. Learn.* **2000**, *39*, 103–134. [CrossRef]
19. Quan, X.; Kit, C.; Ge, Y.; Pan, S.J. Short and sparse text topic modeling via self-aggregation. In Proceedings of the 24th International Joint Conference on Artificial Intelligence, IJCAI 2015. AAAI Press/International Joint Conferences on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015; pp. 2270–2276.
20. Zuo, Y.; Wu, J.; Zhang, H.; Lin, H.; Wang, F.; Xu, K.; Xiong, H. Topic modeling of short texts: A pseudo-document view. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 2105–2114.
21. David, N.; Grieser JH, L.K.; Timothy, B. Automatic evaluation of topic coherence, in: Human Language Technologies. In Proceedings of the 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, USA, 2–4 June 2010; pp. 100–108.
22. Mimno, D.; Wallach, H.; Talley, E.; Leenders, M.; McCallum, A. Optimizing semantic coherence in topic models. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Edinburgh UK, 27–31 July 2011; pp. 262–272.
23. Chen, J.; Nairn, R.; Nelson, L.; Bernstein, M.; Chi, E. Short and tweet: Experiments on recommending content from information streams. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, Atlanta, GA, USA, 10–15 April 2010; pp. 1185–1194.
24. Chang, J.; Gerrish, S.; Wang, C.; Boyd-Graber, J.; Blei, D. Reading tea leaves: How humans interpret topic models. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; pp. 288–296.