A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

Refining IEFP's Recommender Systems: A Comparative Exploration of Project Based Learning and a Modernized Framework

Silje Marie Østberg Mogen

Work project carried out under the supervision of:

Patrícia Xufre
Susana Lavado

20/12/2023

Abstract

This study explores an alternative approach to skill-based recommender systems to match job candidates and opportunities, avoiding the usage of dimensionality reduction methods on skills. Using several evaluation metrics, it compares the results of this new approach with results obtained by the previous models. The proposed approach of splitting the skills data into one dataset per major group of a job proves to be more efficient when comparing to the previous methods. Despite modest evaluation results, the model shows promise. Overall, this research concludes that such a recommendation system has great potential to enhance matching diversity within Public Employment Centers.

Keywords: recommender system, job recruitment, skills, similarity, Jaccard similarity.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

The Instituto do Emprego e Formação Profissional (IEFP) is the Portuguese Public Employment Service. A vital part of IEFP´s mission is to promote the creation and quality of jobs and to fight unemployment through the implementation of active employment policies, such as professional training (IEFP n.d.). Together with IEFP, me and a group of three other students worked on a Project Based Learning (PBL) project with the goal of making a recommender system to recommend available job opportunities to job candidates registered at IEFP, based on the skills defined in the European Standard for Competencies and Occupations (ESCO). The aim of the project was to surpass the existing absolute matching system by constructing an advanced recommender system that included fuzzy matching based on skills, and therefore enhanced the matching diversity.

The existing workflow at IEFP within their information system consists of a counselor cross-referencing a candidate´s demographic information, professional experience and/or desired job codes, as defined in the Classificação Portuguesa das Profissões (CPP) with job offers CPP codes and demographic information, conducting searches for 100% matches. This system could also be accessed by a candidate himself or herself through the online interface of IEFP. The main issue with this workflow is that the recommendations are too conservative in the jobs it suggests considering that similar jobs with different job codes are not suggested. Consequently, this solution may be time-consuming for counselors, as they might have to perform the search several times with different filters and CPP codes to locate a suitable position.

With the recommender system developed in this thesis, the candidate can define and input their own skills, as well as personalize the filters applied. Alternatively, a counselor could handle this. The system will look for jobs that best suit the candidate based on their skills set, enabling it to suggest jobs from different codes, rather than just one. Up until this

project, IEFP used the national standard CPP codes to identify jobs but moving forward they intend to use ESCO codes instead. By providing a "common language" on jobs and skills that organizations like IEFP can use, ESCO aims to promote job mobility throughout Europe and, consequently, a more integrated and effective labor market (European Commission 2022).

The recommender system suggested in this thesis does not change the candidate's current journey within IEFP systems, but it might help them discover more offers that fit their characteristics. In the event that the candidate's characteristics do not match the requirements of available job offers, the candidate may be recommended by a counselor to undergo professional training to learn new skills. When the characteristics of a candidate align with the requirements of a job offer, the candidate has the option to apply for the position. The outcome of this application may result in the candidate either securing or not securing the job. In the latter case, the process starts over as demonstrated in figure 1.

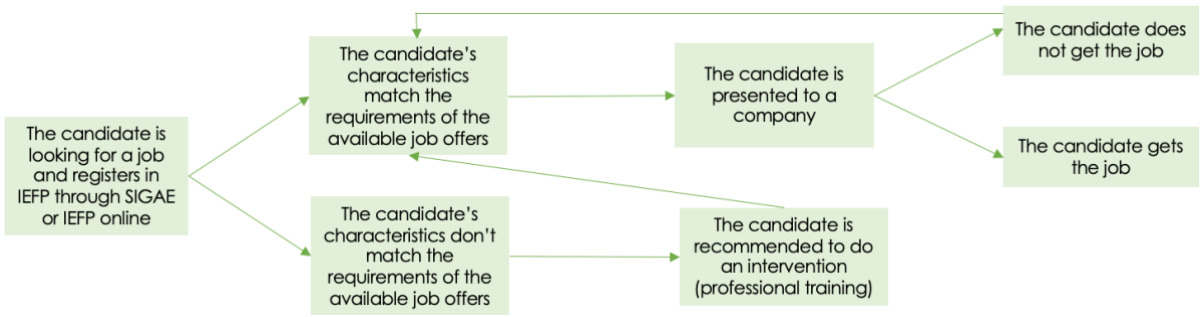

Figure 1. Journey of a candidate within the IEFP systems.

This work project is a continuation of a PBL-project at Nova SBE, which is a student-oriented teaching method where students can work on a real-world problem together with a company. This method of teaching allows students to apply the theoretical knowledge they have learned in class to real-world situations, helping the student develop critical thinking skills and teamwork abilities. The purpose of this thesis is to determine whether developing the recommender system through an alternative approach—that is, without using dimensionality reduction techniques—leads to better results than the model developed in PBL.

The following sections present the work studying if a new approach to the development of a recommender system for IEFP is beneficial. The literature review section will include theory about recommender systems and similarity measures. The data curation section describes the data used in this project and decisions made regarding that data, while the Exploratory Data Analysis (EDA) section include information to better understand it. Following the EDA is a section about evaluation metrics and how they work. The next sections are about the baseline model and the proposed model in this thesis, both including model description, evaluation, and a comparison between the models. I will end the thesis with a discussion about the models and their limitations, before a section about future work to be researched.

2 Literature Review

2.1 Recommender systems

The study of modern recommender systems has been ongoing since the early 1990s, and from those studies two main systems have emerged (Jannach, et al. 2021). Content-based recommendation systems create profiles of items and then recommend items whose profile is similar to those a user has previously liked. Meanwhile, collaborative filtering identifies users similar to a given user and recommends items the former users have liked to the latter user (Balabanovic and Shoham 1997).

Both methods have several advantages and disadvantages. For instance, collaborative filtering often struggles with a cold start problem, meaning that it cannot recommend any items for initial users due to insufficient information about their interaction history. The lack of interactions makes it difficult to identify similar users. In content-based systems the users do not rely on each other, and recommendations can be made with only one active user. However, these systems encounter a cold start problem as well, as they struggle to learn user

preferences without enough ratings from that one active user (Lops, Marco and Semeraro 2011).

One of the biggest disadvantages with the content-based system is that a user is only recommended items similar to the ones the user has already liked and will therefore not be exposed to new, interesting, fields. This is referred to as the serendipity problem, to highlight the content-based systems limitations to produce novel and unexpected recommendations. This is not an issue in collaborative filtering since similar users may like items across fields as well (Lops, Marco and Semeraro 2011).

Moreover, the feature selection in content-based system is, to some extent, a manually designed process that underscores the importance of domain knowledge. For instance, in a job recommendation system, relevant features might include skills, experience level, industry, and location. To identify the key features that influence job preferences, it is essential to understand the job market and industry requirements. In addition, enough relevant information needs to be incorporated into the content to distinguish items that the user prefers from those they dislike (Lops, Marco and Semeraro 2011).

The transparency of a content-based system is a clear advantage, since it can be explained by explicitly listing the content features or descriptions that led to an item appearing on the recommendation list. In collaborative filtering, an item's appearance on the recommendation list is attributed to the liking of an unknown user or a group of unknown users, making the system more like a black box (Lops, Marco and Semeraro 2011).

2.2 Similarity Measures

A way to identify job opportunities that are a good fit for a job candidate is to check if the set of skills required by the job and the set of skills owned by the candidate are similar. Sets of skills are more similar the more there is an overlap between the skills they contain.

These sets of skills can be represented as vectors of 0 and 1, where 0 indicates that a skill is not included in the skills set, and 1 indicates that it is included in that skills set. This way, all the skills sets are represented by vectors of the same size.

To measure similarity between binary vectors, two measures used in machine learning are the Jaccard coefficient and the Sørensen-Dice coefficient. The Jaccard coefficient is calculated by dividing the intersection by the union of the two asymmetric vectors. The vectors are asymmetric if presence of a skill (1) is more important than absence (0). If we define the variables *a*, *b* and *c*, along with the vectors *i* and *j*, consisting of *n* binary attributes, as follows:

- *a* is the number of attributes that equal 1 for both vectors *i* and *j*.
- *b* is the number of attributes that equal 0 for vector *i* but equal 1 for vector *j*.
- *c* is the number of attributes that equal 1 for vector *i* 0 but equal for vector *j*.

Then the Jaccard coefficient is calculated by the following equation (Karabiber, Jaccard Similarity 2020):

$$(1)\ J(i,j) = sim(i,j) = \frac{a}{a+b+c}$$

The Sørensen-Dice coefficient is calculated as twice the intersection between the two vectors *i* and *j*, divided by the sum of cardinalities (Cardinal 2022). The Sørensen-Dice is similar to the Jaccard coefficient but gives double weight to the true matches both in the numerator and denominator. Similar, the Sørensen-Dice is also used for asymmetric vectors, as there can in both methods be an infinite amount of true negative matches, meaning that both measures are negative matches exclusive. The Sørensen-Dice is calculated by the following equation, using the *a*, *b* and *c* defined above (Tappert, Cha and Choi 2009):

$$(2)\ D(i,j) = sim(i,j) = \frac{2a}{2a+b+c}$$

Both Sørensen–Dice and Jaccard coefficients evaluate the proportion of matches in relation to the total membership. This highlights that the measures will impose penalties for differences between vectors, even when one is entirely contained within the other. Both measures range from 0 to 1, where 1 signifies an identical match and 0 denotes no similarity. The greater the disparity between the vectors, the closer the value is to the lower end of this range (Cardinal 2022). Sørensen-Dice does not satisfy the triangle inequality and is therefore not to be considered a proper distance metric, while Jaccard is. The triangle inequality ensures that the distance metric behaves consistently and in a way that aligns with our geometric intuition (Gallagher 1996).

A third common similarity measure is cosine similarity, which calculates the cosine of the angle between two vectors. The cosine similarity between two binary vectors also ranges from 0 to 1. Utilizing cosine similarity is advantageous when handling sparse data due to its capacity to disregard 0-0 matches, a characteristic it shares with Jaccard and Sørensen-Dice. Including 0-0 matches in sparse data computations can lead to an inflation of similarity scores (Karabiber, Cosine Similarity 2021). Both the Jaccard coefficient and cosine similarity is common to use in recommender systems, while Sørensen-Dice is often used as a supplement (Tilores 2023).

3 Data Description

This project relies on two primary sources for data, the data provided by IEFP and data about ESCO skills and occupations – retrieved from the European Commission´s official website (https://esco.ec.europa.eu/en).

Data provided by IEFP consists of three tables with information pertaining to job candidates registered with IEFP, job offers available within the system, and job applications that were submitted. The datasets spanned from 2017 to 2021. Given the substantial volume of data, the approach involved selecting a representative 10% of the candidate data and then

applying a filter to isolate the data specific to the year 2019, making the data more manageable. For the job offers, IEFP was unable to provide a crucial feature regarding the state of a job offer (Active or Passive), meaning that it was not possible to define an accurate end date of the offer. In the PBL project, the end date of a job offer was set to be the end of the dataset. Because of this, the number of active jobs would accumulate over time. To avoid ending up with an unnaturally high number of matches for the recommender system in this thesis, the end date of an offer was set to three months after the start date, where an accurate end date was not already provided.

Furthermore, scraping the ESCO website was necessary to gather information about ESCO skills and occupations. At the time the web scraping was conducted, the ESCO website did not fully reflect all the ESCO job codes and skills due to ongoing updates. Scraping the website an additional time would prove to be too time consuming, meaning that, for a few occupations, there is missing data. Both the ESCO skills and the ESCO job codes are built in a hierarchical way, with jobs codes comprising up to six distinct levels. Each level in the hierarchy represents a progressively specialized subset of the broader category. For each of these levels in the job codes, the European Commission denotes 'essential' and 'optional' skills tied to a particular job.

The ESCO dataset on skills consists of 8574 skills in the fourth level of the hierarchy, which is the predominant level for skill definitions. The hierarchy of skills reaches ten different levels, however beyond the fourth one, each level has fewer and fewer unique skills. Level ten has only two distinct skills, compared to level five's 2917. The skills in the fourth level were converted to a binary table, with ESCO job code as index and skills as columns. If a skill were present in a job the value would be 1 and 0 otherwise.

To deal with highly sparse datasets, the binary table of skills related to each job code was split into ten different datasets, one for each major group of the job (see appendix A for

description). The major group is the highest level in the job hierarchy, defined by the first digit in the ESCO code. For instance, the dataset for major group two, which is the group with the highest number of unique skills, would include all the job codes related to this major group and skill sets associated with these jobs. It is now easier to handle because, for major group two, the dataset only contains 3748 skills instead of the entire 8574 original skills. This reduction in the skills sets not only simplifies the dataset but could also significantly improve processing time, allowing for faster and more efficient computations.

As of now, IEFP does not have any information about the skills of a job candidate, or skills required by a specific job offer. As a result, the skills related to the candidate or job offer are assumed by default to be the skills associated with a job occupation in ESCO. For the candidate, the skills would be retrieved from both the desired job of the candidate and their previous job. After consulting with IEFP, more weight is given to the skills related to the desired job, since it is highly likely that the candidate already has these skills, and such a job would be a better match for the candidate. The weighting is 80% to the desired job and 20% to the previous job, as agreed upon with IEFP. Once the skills are implemented in IEFP´s systems, this weighting would not be necessary, considering the candidates and the employers would define the skills themselves.

4 Exploratory Data Analysis

An Exploratory Data Analysis (EDA) was carried out to obtain insights into the features of the datasets received from IEFP and from the data retrieved from ESCO. Visualizations from the EDA can be seen in appendix B. This helped to uncover the characteristics of the typical candidate profile and job offers, giving a fundamental understanding that was crucial for the recommender system's later development and optimization. For instance, the analysis showed a sharp disparity in the types of jobs that candidates preferred. Of the candidates, 84% indicated that they wanted permanent jobs, but

only 38% of the job listings that were available fit this description. This indicates that there is a significant gap between the types of jobs that candidates want and the jobs that are currently available in the dataset.

A detailed analysis of the user demographics was also included in the EDA. The analysis showed that the most common employment category for people who have registered with IEFP is "unemployed and seeking employment," which accounts for 76% of candidates. Additionally, 36% of candidates report that they registered as a result of their non-permanent job ending. Gender and age also have a significant role in the candidate's educational background and the type of work that the user wants. Women and younger people tend to have a higher education than men and the older generation. The most popular major group for men is Craft and Related Trades, with 19% of the men, but only 3% of women wanting a job in this field. 70% of the women want a job in the major groups Professionals, Service and Sales Workers, and Elementary Occupations, while men are more equally distributed amongst the groups. In addition, two thirds of the candidates desire a job in the same major group as their previous job. Each candidate sends an average of two applications to job offers a year, whilst the median is one application sent per candidate per year.

The majority of the job offers available in the dataset are located in the major groups Service and Sales Workers, and Elementary Occupations, which aligns with the candidate demands. There is a great similarity between the distribution of job offers across major groups and the distribution of the desired major groups by candidates, revealing a significant alignment in the percentages within each category. The only two major categories with a discrepancy over 5% are Professionals and Craft and Related Trades Workers.

In terms of skill distribution per occupation, the median number stands at 20, whereas the average is slightly elevated at 23.36. The major group also affects the median and average, with some major groups having higher median and average values than others. Furthermore,

the EDA showed some intriguing trends regarding the ESCO skills. Most of the skills, namely 4698 out of 8574, are only found in one major group, while 2068 can be found in two. Only 7 skills can be found in all major groups, with one such skill being "Manage staff". This highlights how the skills tend to be closely aligned to the occupational context of major groups, and these insights served as a key motivator for splitting the skills into their respective major groups.

5 Evaluation Metrics

The models are evaluated based on three metrics throughout the modeling process: recall@k, precision@k and hit rate. These metrics are commonly used in the evaluation of recommendation system, to measure the relevance of the recommended items (Kapre 2021). A relevant item is, in this context, a job offer to which the candidate has applied. In addition, the processing time of the system was also measured. These metrics will be used to assess whether the baseline model—the model created during the PBL—was outperformed by the model proposed in this thesis.

5.1 Recall@K

In recommendation systems, recall@K is a metric that shows out of all the relevant items, how many of them are recommended to the candidate in the top K (Kapre 2021). In the context of job recommendations, recall shows the number of recommended jobs the candidate has applied to divided by the number of jobs applied to by the candidate. For example, if the candidate has applied to six offers, and the recommendation list with a length of ten shows two of them, the recall@K would be 2/6 = 0.33. In this project, where we use historical data, a high recall would mean that the model is able to replicate the historical matches correctly.

$$(3)\ Recall@K = \frac{Number\ of\ Relevant\ Items\ in\ Top\ K}{Total\ Number\ of\ Relevant\ Items}$$

5.2 Precision@K

Precision@K shows “out of K” items recommended to a user, how many of them are relevant (Aher 2023). For the job recommendations in this project, precision shows out of all the jobs on the recommendation list, how many did the candidate apply to. For instance, if on the top ten recommendation list, the candidate applied to two of them, the precision@K would be 0.2.

$$(4)\ Precision@K\ = \frac{Number\ of\ Relevant\ Item\ in\ Top\ K}{K}$$

5.3 Hit Rate

Hit rate measures how often we are able to recommend a relevant item (Li 2019). If a relevant item is on the recommendation list, the hit rate is 1 for that candidate. In this context, this would mean that if a candidate has applied to at least one of the offers on the recommendation list it would result in a match.

$$(5)\ Hit\ Rate = \frac{Count\ of\ users\ with\ at\ least\ one\ match}{Count\ of\ all\ users}$$

5.4 Processing time

To assess and compare the efficiency of the models, a random sample of 100 candidates will be selected for an examination of the average processing time per candidate, which represents the duration the model takes to formulate and present a recommendation for each candidate. The random sample follows the same major group distribution as the original dataset. Evaluating processing time can help to ensure that the responsiveness of the recommendation system is aligned with users' expectations and operational needs.

6 Baseline Model (PBL models)

6.1 Model Description

As baseline model for this project, I will use the models created during the PBL phase. The models will act as a benchmark to conduct performance evaluation on the proposed

model in this thesis. In the approach of enhancing IEFP´s matchmaking process, the PBL team developed two different recommender systems. Both models calculated the cosine similarity between the candidate vector and the job vector to determine compatibility, but with different variations to the vectors. We utilized the candidate registration date to simulate providing recommendations as of that particular date. While recommendations could theoretically be generated at different stages of the candidate journey, this project specifically takes the candidate's registration date into account as the matching date.

The skills in the baseline model were transformed using dimensionality reduction methods to reduce the number of features from 8574 unique skills to 188 components that still contained most of the information and 60% of the variability that the original data had. This was done using a method called Principal Component Analysis (PCA). PCA reduces feature dimensions through the generation of linear combinations of the original features, referred to as components (Sharma 2023). PCA has certain disadvantages in addition to its many benefits. Specifically, PCA can help reduce noise and feature count, but at the cost of losing some potentially important information. Furthermore, it could be challenging to interpret the components, particularly when combining different skills sets (Chawla 2023). Two entirely different skills ended up in the same component during the PBL phase, for example the skills “clean jewelry pieces” and “adjust engineering designs”, illustrating the potential disadvantages of using PCA to capture subtle relationships between skills. Despite the possible drawbacks, we chose PCA as a tactical approach for tackling big data challenges.

The first model we developed considered the candidate vector as the skill components of the candidate and the job vector as the skill components of a job code. The candidate profile was built using a weighted average approach, giving 80% weight to the components derived from the desired job of the candidate and 20% weight to the components derived from the past job. After the cosine similarity was calculated, the job code was merged back with

each unique job offer, and the ten job offers with highest cosine similarity was recommended to the candidate, given that the offer was active during the candidate´s registration date in IEFP.

The second model developed was a hybrid model considering both skills, demographic information, and the similarity between the ESCO code of the desired job of the candidate and the job recommended. The hybrid model is composed of three independent models that are ultimately given distinct weights to form a single collected model.

The first independent model is the same as the skills model above, while the second one is a model which considered only the demographic information of a candidate and the demographics of a job. This includes the district where the job or candidate is located, the education level, the amount of work experience, the type of contract and the regime of the job. To define the vectors used in the similarity calculation and make sure the data was standardized, categorical features such as district, regime, and contract type underwent One-Hot encoding. Concurrently, numerical features like minimum education level and minimum experience level were subjected to transformation using MinMaxScaler. To make sure the same scaler was used on both the candidate and the job, maximum education and maximum work experience was excluded from the model. The decision between minimum and maximum requirements was made based on the assumption that the minimum requirement was more crucial to meet, and that the candidate's level was closer to the minimum than the maximum. Cosine similarity was calculated based on the standardized vectors.

The last independent model calculated the similarity between the ESCO code of the desired job of the user and the job recommended. It works by using the first five digits in the code. If the first digit of the two ESCO codes does not match, they receive a similarity of 0, under the premise that disparate major groups imply complete dissimilarity. Conversely, if the first digit’s match, a foundational similarity score of 0.2 is allocated. With each subsequent

matching digit, the similarity score increments by 0.2. This is because more digit matches point to increased specificity and, consequently, job role similarity. As a result, the higher the similarity score, the closer the recommended job aligns with the candidate's desired job.

All three independent models were combined into one hybrid model based on weighted similarity scores. Several weighting options were considered, but the best one turned out to be 50% weight to the skills, 25% weight to the demographic model and 25% to the ESCO codes similarity.

Furthermore, we split the hybrid model into sorted and unsorted, the difference being that the sorted output would recommend jobs closer to the candidate´s matching date, leaving us with two different hybrid models. This was necessary due to the missing data on end dates, considering we would have an abnormally high number of offers available, some of which would be made well in advance of the candidates' matching date. Different lengths of the recommendation list were tested, but for comparison reasons, I will focus on the results from the top ten recommendations.

6.2 Evaluation

Table 1 presents the evaluation results for the PBL models. The low precision, recall and hit rate in the skills model may suggest that the model is not being capable of matching recommendations effectively to candidate´s needs, due to its restrictions. For example, most candidates may have preferences regarding the location of the job, which the skills model will not be able to detect. Preferences like that are part of the hybrid model, which shows a substantial increase in all metrics compared to the skills model, indicating that using a hybrid model enhances the model’s ability to identify relevant recommendations among all the possibilities. Even though the results are better than the skills model, they are still quite low. The hybrid model was only able to recommend at least one of the jobs the candidate has

applied to in 4% of the cases. In addition, the huge difference in the metrics between the sorted and unsorted dates in the hybrid model tells us that the model is highly sensitive to temporal order of the job offers. Both models operate efficiently in terms of processing time. However, it is important to note that the processing time does not entirely encapsulate the entire model workflow, given that the PCA and the definition of the candidate profile is performed in advance.

	Precision@10	**Recall@10**	**Hit Rate**	**Processing Time**
Skills model	0.000615%	0.019%	0.033%	3.56s
Hybrid model with skills focus	0.092% (0.025%)	2.815% (0.775%)	4.264% (1.208%)	2.99s

Table 1. Evaluation - Baseline Models (sorted and unsorted dates) – Numbers in parentheses represent unsorted dates.

7 Model

7.1 Model Description

To address the shortcomings of the model developed during the PBL phase, an alternative model is being proposed in this thesis. The first problem of the PBL model is the reliance on PCA, which can be challenging to interpret and full of subjective decisions such as the number of components. PCA was conducted because the team were not able to handle the large amount of data we had on skills in its binary form. For each candidate we would have one column for each skill in addition to their demographic information, making a total of over 8500 columns.

To get around the big data issues without using dimension reduction techniques, ESCO job codes and their respective skills were split into their major group, allowing for a more interpretable and manageable format. Using smaller data frames would also significantly improve the processing time for each candidate due to reduced computational

complexity.

The model developed is a content-based recommender system, which focuses on the skills to recommend jobs to a candidate. A job is recommended to a candidate if the skills of the job are somewhat similar to the skills of the candidate. As a starting point the model first included similarity calculation between skills in terms of both Jaccard coefficient and Sørensen-Dice coefficient, but since Sørensen-Dice does not satisfy the triangle inequality, only Jaccard were considered in the end.

The model calculates the Jaccard similarity of the skills associated to the job code of the candidate to all the jobs in the same major group, for the desired job and past job separately. If the Jaccard similarity is over 0.30, the job code is kept in a dictionary along with the similarity. The threshold of 0.30 was determined based on the observation that it is uncommon for the available jobs to have a similarity exceeding 0.50. For example, in major group eight, the average number of similar jobs when the threshold is 0.50 is 2.12, whereas the average is 5.36 with a threshold of 0.30. By setting the threshold at 0.30, a substantial range of jobs can be recommended. Jobs are weighted with an 80-20 split between the desired and past job, giving jobs similar to the desired job more weight than jobs associated with the past job of the candidate. All similar jobs are added to each candidate profile, giving one row in the data frame per job code, as seen in table 2. The table also demonstrates the similarity between the job codes 9112.2 and 9112.6, namely “building cleaner” and “train cleaner”. Both occupations share a lot of the same skills, for example “clean class surfaces”, meaning that having the skills needed in one of the jobs, could also make you fit to work in the other.

CANDIDATE ID	START DATE	END DATE	DISTRICT	…	DESIRED JOB	PAST JOB	JOB CODE	JACCARD SCORE
123456	01-01-2019	31-03-2019	4	…	9112.2	4223.1	9112.2	0.80
123456	01-01-2019	31-03-2019	4	…	9112.2	4223.1	9112.6	0.52
123456	01-01-2019	31-03-2019	4	…	9112.2	4223.1	4223.1	0.20
123456	04-08-2019	13-09-2019	4	…	9112.2	9112.2	9112.2	1
123456	04-08-2019	13-09-2019	4	…	9112.2	9112.2	9112.6	0.65

Table 2. Example candidate profile for candidate 123456.

Candidates often have different preferences when it comes to the filters applied to the recommender system. For example, some may place more value on the job's location, while others may place greater value on it being a permanent position. Therefore, the candidate themselves (or the counselors) can, based on their individual preference, define the filter criteria applied to the system. This ensures that the recommended jobs not only match their skills set, but also adhere to their distinct priorities and preferences. Furthermore, the model makes sure that only offers that are available during the candidates' matching date are recommended. When the available job offers are defined, the job offers are merged with the candidate profile to create a matching table, consisting of one row per job offer that matches the candidate profile. The matching table is then trimmed to only display the top ten recommendations, determined by the highest Jaccard scores, as illustrated in table 3.

CANDIDATE ID	MATCHING DATE	DISTRICT CANDIDATE	…	OFFER ID	START DATE JOB	END DATE JOB	DISTRICT JOB	JOB CODE
123456	01-01-2019	4	…	567891	15-12-2018	31-01-2019	4	9112.2
123456	01-01-2019	4	…	567257	26-12-2018	15-02-2019	4	9112.2
123456	01-01-2019	4	…	579436	01-01-2019	13-03-2019	4	9112.2

Table 3. Example matching table for candidate 123456 (and top three recommendations).

To explore the impact the different filters had on the result of the model, the model was divided into five new models, each incorporating different combinations of filter criteria but following the same methodology. In addition, a model incorporating all skills in one data

frame rather than splitting them up per major group was included to examine the effect this had on the model's processing time. See table 4 for the models and their filter combinations. This method not only makes it possible to analyze the subtle differences between the filter options, but it also makes it easier to determine which filter option has the biggest influence on the recommendation list.

Model number	**Filters applied**
MODEL 1	No filters
MODEL 2	All filters
MODEL 3	Job requirements; work experience and level of education
MODEL 4	Demographic information; district, work regime and type of contract
MODEL 5	Work experience, level of education and district
MODEL 6	Work experience, level of education, district, and considering all skills instead of per major group

Table 4. Models developed and their filter option.

7.2 Evaluation

From table 5 we can see that the fastest model is the model without filters, but it comes at a cost of precision, recall, and hit rate. Since model 1 does not account for user preferences, there is little chance that the candidate applied to any of the jobs that are recommended. Model 2, which considers all filters, performs only a bit better than the first in all metrics, except processing time and the average number of recommendations, indicating that while the filters are refining the recommendations, they may still not be aligned with the candidate's behavior. A shorter recommendation list would also explain a higher precision, given that there is a lower chance of an irrelevant recommendation on the list.

Model 3's narrow focus on job requirements does not significantly shift the metrics, implying that there may be other factors beyond job requirements that influence what candidates apply to. A noticeable improvement in model 4's recall and hit rate may suggest that introducing demographic information in the filters plays a role in the candidate's

application behavior. The precision does not improve more than the model with all filters, even though the average number of recommendations is higher, meaning that leaving out job requirement filters from the model does not significantly improve the relation of relevant items in the list.

The clearly best model is model 5, including district as a filter along with work experience and level of education. Comparing the results to models 2, 3 and 4, this demonstrates that district is an important factor to consider when making recommendations, while type of contract and the regime of the work is not. Including type of contract and regime in the filters, which is the difference between model 2 and model 5, significantly decreases both recall and hit rate in over 50%, while precision also has a drastic decrease. The average number of recommendations is also higher than model 2, speaking to the fact that model 5 is able to recommend more of the relevant offers.

The performance of model 6, which includes all skills, regardless of the job group, and filters on district, work experience and level of education, is comparable to model 5, but with a severe increase of processing time, suggesting that using all the skills adds computational complexity without proportional gains in the evaluation results. This implies that splitting the skills sets into major groups and therefore reducing the noise of irrelevant skills ensures that the model processes quickly and aligned with user's expectation while not losing any important information. In addition, the marginal differences between sorted and unsorted dates reveals that the chronological arrangement of job offers according to their start date does not significantly impact the results.

	Precision@10	Recall@10	Hit Rate	Processing Time	Average number of recommendations
MODEL 1 - No filters	0.024% (0.024%)	0.748% (0.747%)	1.161% (1.160%)	2.22s	9.839
MODEL 2 - All filters	0.092% (0.092%)	1.467% (1.467%)	2.314% (2.315%)	3.18s	5.409
MODEL 3 - Job requirement filters	0.025% (0.025%)	0.767% (0.767%)	1.151% (1.150%)	2.25s	9.666
MODEL 4 - Demographic filters	0.092% (0.092%)	1.763% (1.762%)	2.742% (2.742%)	2.35s	6.378
MODEL 5 – District, work experience and education level	0.154% (0.154%)	3.716% (3.717%)	5.602% (5.601%)	2.25s	7.961
MODEL 6 - All skills	0.152% (0.152%)	3.730% (3.730%)	5.609% (5.609%)	36.72s	8.075

Table 5. Evaluation – Models (sorted and unsorted dates) – Numbers in parentheses represent unsorted dates.

7.3 Comparison with baseline models

All the proposed models in this thesis outperformed the skills model from the PBL phase, even model 1 without filters where the focus is on the skills similarity. This enhances the idea that using only PCA components to recommend jobs is not effective, and that you lose important information doing so. Models 1 and 3 are the only models with results lower than the baseline hybrid model with unsorted dates, suggesting that setting the end date of offers three months after the start date still ensures a comprehensive coverage of relevant offers.

When it comes to the baseline hybrid model with sorted dates, models 1 and 3 are again the only ones with lower precision, which means that the other models perform equally well or better as the baseline hybrid model in terms of including relevant job offers in the recommendation list. In terms of recall, the baseline hybrid model has a result 1-2% better than models 1 to 4. This suggests that the baseline hybrid model is more effective than the individual models 1 to 4 in terms of recommending offers to which the candidate has applied.

However, models 5 and 6 outperform the hybrid model in all metrics. This means the models with district, work experience, and education level filters are better at recommending relevant offers to the candidate than the baseline hybrid model. This could be because district is such an important filter, and in models 5 and 6 the list only shows the specified district, while in the baseline hybrid model the district is measured in terms of similarity.

The baseline hybrid model had a processing time of 3 seconds per candidate on average when the PCA components were already incorporated. In a real-world situation the skills components would also need to be calculated for each user, which would increase the processing time because of the added calculations needed. Models 1, 3, 4 and 5 are all quicker than the baseline hybrid model, meaning that avoiding dimensionality reduction methods also leads to faster processing of the models.

8 Discussion

This work project aimed to research whether a different approach to the development of a recommender system for IEFP would be beneficial compared to the models developed during PBL. More precisely, this thesis investigated whether employing a strategy without dimensionality reduction techniques would produce improved results. This thesis's methodology involved dividing the skills data into smaller datasets, one for each major job group. The conclusion of this analysis is that it does work better to split the skills data compared to a PCA model, especially when only focusing on the skills. When also adding demographic information into the equation, the proposed models in this project work better than the hybrid model from PBL, provided that the district, work experience and education level is part of the filters, and work regime and type of contract are not.

Although the proposed models yielded modest evaluation results, they demonstrate significant potential. The proposed models shows faster processing times compared to the baseline models, which aligns more with user expectations. Furthermore, unlike in PCA,

where skills are hidden within components, the transparency for users is improved. The users can more easily determine which skills contributed to the recommendation of a particular offer. Another advantage offered by the models is the ability for users to define filter criteria according to their preferences. This not only boosts the flexibility of the recommender system but also has the potential to increase satisfaction and engagement, as users have direct control over the filters.

To ensure the effectiveness of the model it is crucial to address the limitations of the current dataset, especially for the status of a job offer. The matchmaking process would improve with this knowledge, concentrating on offers that were truly available to the candidate. An assumption of the job offers availability was needed in this project to avoid recommending those that were not realistically accessible. The updating of the data collection should not only include the status of the offer, but also the skills set of an offer or of the candidates registered at IEFP. Not all skills related to a job code may be relevant for the candidate, or they may possess more skills than what the job code does. Thus, it is crucial to collect data on the accurate skills of a candidate. Further, considering the web scraping of the ESCO website did not adhere with a full dataset, this needs to be repeated to get the missing data.

Additionally, in the proposed models a threshold of 0.30 was chosen to determine the minimum Jaccard similarity between two jobs for it to be recommended to a candidate. However, this threshold was not validated with IEFP. Though the threshold of 0.30 ensures that a wider range of jobs can be recommended, it does not guarantee that each candidate will receive ten recommendations. More study and validation on the threshold are necessary to make sure the models meet IEFP requirements for the recommender system.

One of the limitations with evaluating the model based on historical applications is that matchmaking is only employed once per user registration. The matching is carried out

using the offers that were available in the system on the date of the candidate's registration; however, in reality, the recommendation process might take longer, and new offers might become available while the candidate is registered at IEFP. The low results of the models highlight this, as relevant offers may not have been recommended simply because they were not available during the start date of the candidate. When applying the recommendation system in real-life situations, the matching date can be set using a trigger. This means that the model can rerun on any day during the candidate's IEFP period and provide fresh recommendations with all the active job offers.

The skewed results between precision and recall could also be explained by the low number of applications per candidate. If a candidate on average only applied to two offers, but the recommendation list is always a length of ten, the average precision would be at best around 0.2. But if both of those applications were on the list, the average recall would be 1. In addition, with the current solution in IEFP, candidates have the option to apply for all kinds of offers available, not only those represented in a top ten list. However, the system searches for 100% matches, which might lead to candidates not being exposed to offers who might also fit their skill set. Lower results for the suggested model in this thesis may result from the possibility that the historical applications do not accurately reflect all the offers in which a candidate might be interested.

Furthermore, proving the model's superiority requires more than just determining whether it performs better than the baseline model. It might be advantageous to include subjective assessments of the effectiveness and applicability of the recommendations. Counselors may have years of experience recommending suitable job offers to candidates, and they might notice details that the model misses. This could work as a supplement to the objective classification methods used in the project to evaluate the model.

To conclude, the proposed models in this thesis show improvement compared to the baseline models, especially regarding processing time and transparency for the users of the recommender systems. These improvements indicate a great potential for improving the matching diversity within IEFP. However, it is important to address the limitations of the current dataset and the models before applying the system in a real-world situation, to ensure the models reliability and effectiveness.

10 Future work

A possibility for future work is to expand the model to not only provide recommendation to job seekers, but also for employers seeking suitable candidates. By expanding the system to suggest well-matched candidates to a job offer, companies could receive targeted recommendations, enhancing their recruitment process. Additionally, it could be beneficial to include work training as a recommendation for candidates who might not find any suitable offers. This could help them with their skills development, and therefore broaden the system’s scope to support people at various stages in their career. Adding these extensions could increase the system's versatility and offer both job seekers and employers a more comprehensive solution.

References

Aher, Pratik. 2023. *Evaluation Metrics for Recommendation Systems — An Overview.* August 09. Accessed December 05, 2023. https://towardsdatascience.com/evaluation-metrics-for-recommendation-systems-an-overview-71290690ecba.

Balabanovic, Marko, and Yoav Shoham. 1997. "Fab: Content-Based, Collaborative Recommendation." *Communications of the ACM* 40 (3): 66-72.

Cardinal, James Scott. 2022. *Similarity Measures and Graph Adjacency with Sets.* October 28. Accessed October 13, 2023. https://towardsdatascience.com/similarity-measures-and-graph-adjacency-with-sets-a33d16e527e1.

Chawla, Awi. 2023. *The Advantages and Disadvantages of PCA To Consider Before Using It.* May 10. Accessed December 01, 2023. https://www.blog.dailydoseofds.com/p/the-advantages-and-disadvantages.

European Commission. 2022. *Start using ESCO.* September 26. Accessed December 04, 2023. https://esco.ec.europa.eu/en.

Gallagher, Eugene D. 1996. *COMPAH 96 Documentation.* Boston, MA: University of Massachusetts at Boston.

IEFP. n.d. *A Instituição.* Accessed September 15, 2023. https://www.iefp.pt/instituicao.

Jannach, Dietmar, Pearl Pu, Francesco Ricci, and Markus Zanker. 2021. "Recommender Systems: Past, Present, Future." *AI Magazine* 42 (3): 3-6.

Kapre, Sangram. 2021. *Common metrics to evaluate recommendation systems.* February 21. Accessed December 05, 2023. https://flowthytensor.medium.com/some-metrics-to-evaluate-recommendation-systems-9e0cf0c8b6cf.

Karabiber, Fatih. 2021. *Cosine Similarity.* July 21. Accessed October 23, 2023. https://www.learndatasci.com/glossary/cosine-similarity/.

—. 2020. *Jaccard Similarity.* December 08. Accessed October 23, 2023. https://www.learndatasci.com/glossary/jaccard-similarity/.

Li, Susan. 2019. *Evaluating A Real-Life Recommender System, Error-Based and Ranking-Based.* January 16. Accessed December 05, 2023. https://towardsdatascience.com/evaluating-a-real-life-recommender-system-error-based-and-ranking-based-84708e3285b.

Lops, Pasqale, Gemmis de Marco, and Giovanni Semeraro. 2011. "Content-Based Recommender Systems: State of the Arts and Trends." In *Recommender Systems Handbook*, by Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B. Kantor, 73-105. Boston: Springer.

Sharma, Pulkit. 2023. *The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes).* November 05. Accessed December 01, 2023. https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/.

Tappert, Charles C., Sung-Hyuk Cha, and Seung-Seok Choi. 2009. "A Survey of Binary Similarity and Distance Measures." *Journal of Systemics Cybernetics and Informatics*, 11.

Tilores. 2023. *Sørensen–Dice Coefficient Algorithm.* Accessed November 29, 2023. https://tilores.io/sorensen-dice-online-tool.

Waggener, Bill. 1995. "Pulse Code Modulation Techniques." 206. Solomon Press.

Appendix A

Major Group	Name	Description
0	Armed Forces Occupations	Armed forces occupations include all jobs held by members of the armed forces.
1	Managers	Managers plan, direct, coordinate and evaluate the overall activities of enterprises, governments and other organizations, or of organizational units within them, and formulate and review their policies, laws, rules and regulations.
2	Professionals	Professionals increase the existing stock of knowledge; apply scientific or artistic concepts and theories; teach about the foregoing in a systematic manner; or engage in any combination of these activities.
3	Technicians and associate professionals	Technicians and associate professionals perform technical and related tasks connected with research and the application of scientific or artistic concepts and operational methods, and government or business regulations.
4	Clerical support workers	Clerical support workers record, organize, store, compute and retrieve information, and perform a number of clerical duties in connection with money-handling operations, travel arrangements, requests for information, and appointments.
5	Services and sales workers	Services and sales workers provide personal and protective services related to travel, housekeeping, catering, personal care, protection against fire and unlawful acts; or demonstrate and sell goods in wholesale or retail shops and similar establishments, as well as at stalls and on markets.
6	Skilled agricultural, forestry and fishery workers	Skilled agricultural, forestry and fishery workers grow and harvest field or tree and shrub crops; gather wild fruits and plants; breed, tend or hunt animals; produce a variety of animal husbandry products; cultivate, conserve and exploit forests; breed or catch fish; and cultivate or gather

		other forms of aquatic life in order to provide food, shelter and income for themselves and their households.
7	Craft and related trades workers	Craft and related trades workers apply specific technical and practical knowledge and skills to construct and maintain buildings; form metal; erect metal structures; set machine tools or make, fit, maintain and repair machinery, equipment or tools; carry out printing work; and produce or process foodstuffs, textiles, wooden, metal and other articles, including handicraft goods.
8	Plant and machine operators and assemblers	Plant and machine operators and assemblers operate and monitor industrial and agricultural machinery and equipment on the spot or by remote control; drive and operate trains, motor vehicles and mobile machinery and equipment; or assemble products from component parts according to strict specifications and procedures.
9	Elementary occupations	Elementary occupations involve the performance of simple and routine tasks which may require the use of hand-held tools and considerable physical effort.

Table 6. Explanation of ESCO Major Groups.

Source: European Commission. 2022. *Occupations*. September 26. Accessed December 12, 2023. https://esco.ec.europa.eu/en/classification/occupation_main.

Appendix B

No Education	Cannot read or write
Basic Education	1st to 9th grade
Secondary Education	10th to 12th grade
Higher Education	Licentiate, post-secondary education, bachelor's degree, master's degree, and doctorate

Table 7. Explanation of what the levels of education consists of.

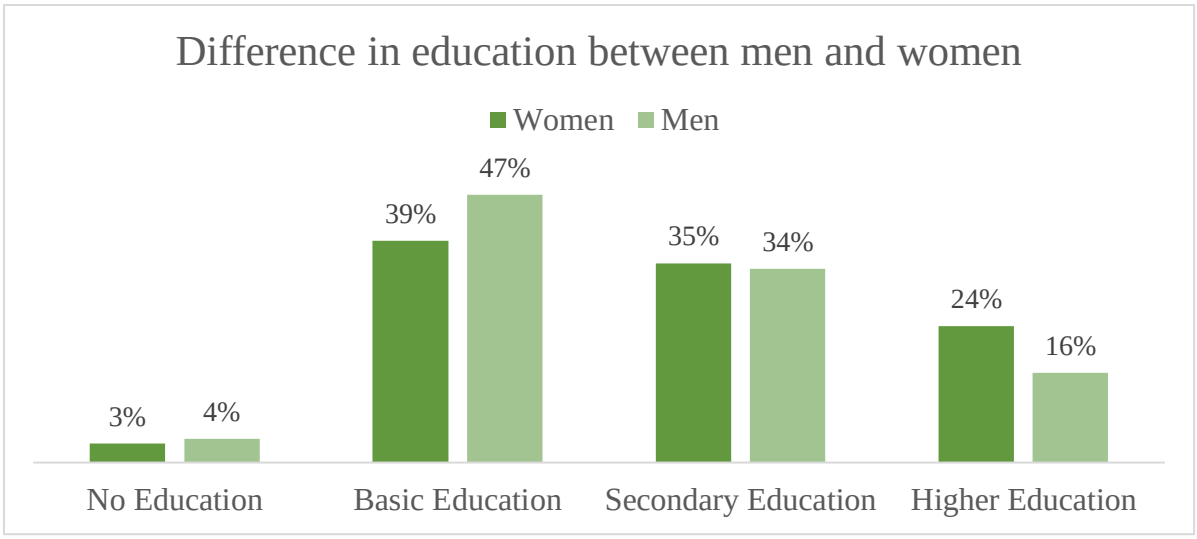

Figure 2. Diagram illustrating the difference in education between men and women.

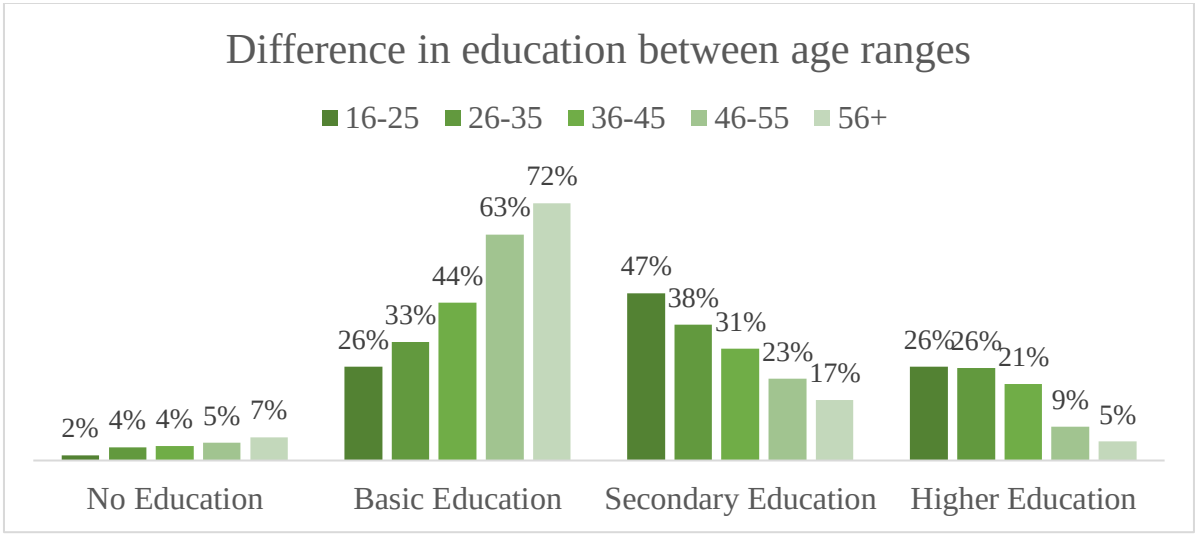

Figure 3. Diagram illustrating the difference in education between age ranges.

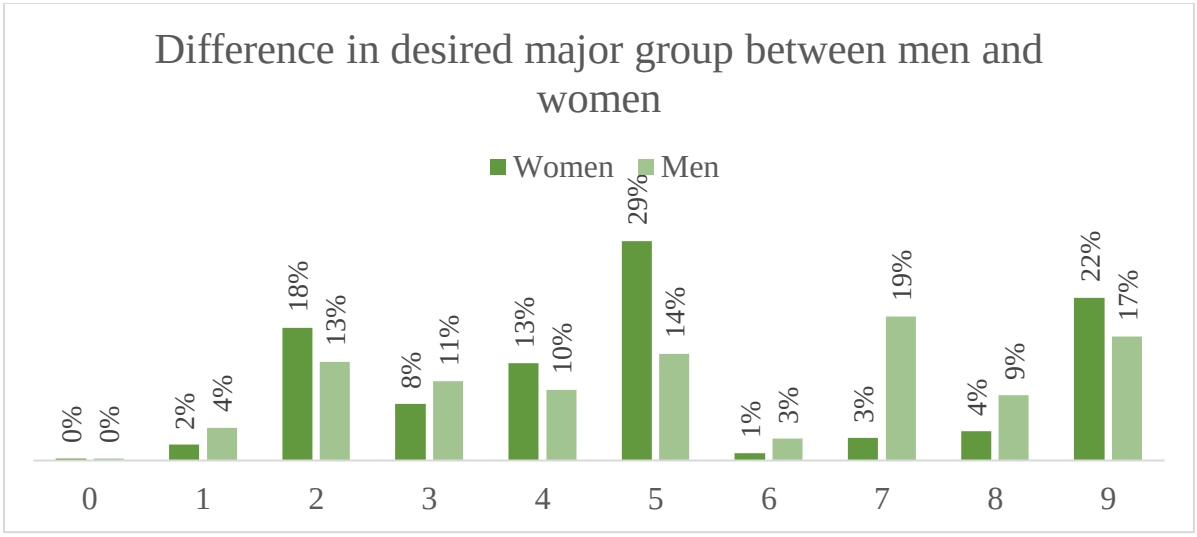

Figure 4. Diagram illustrating the difference in desired major group between men and women.

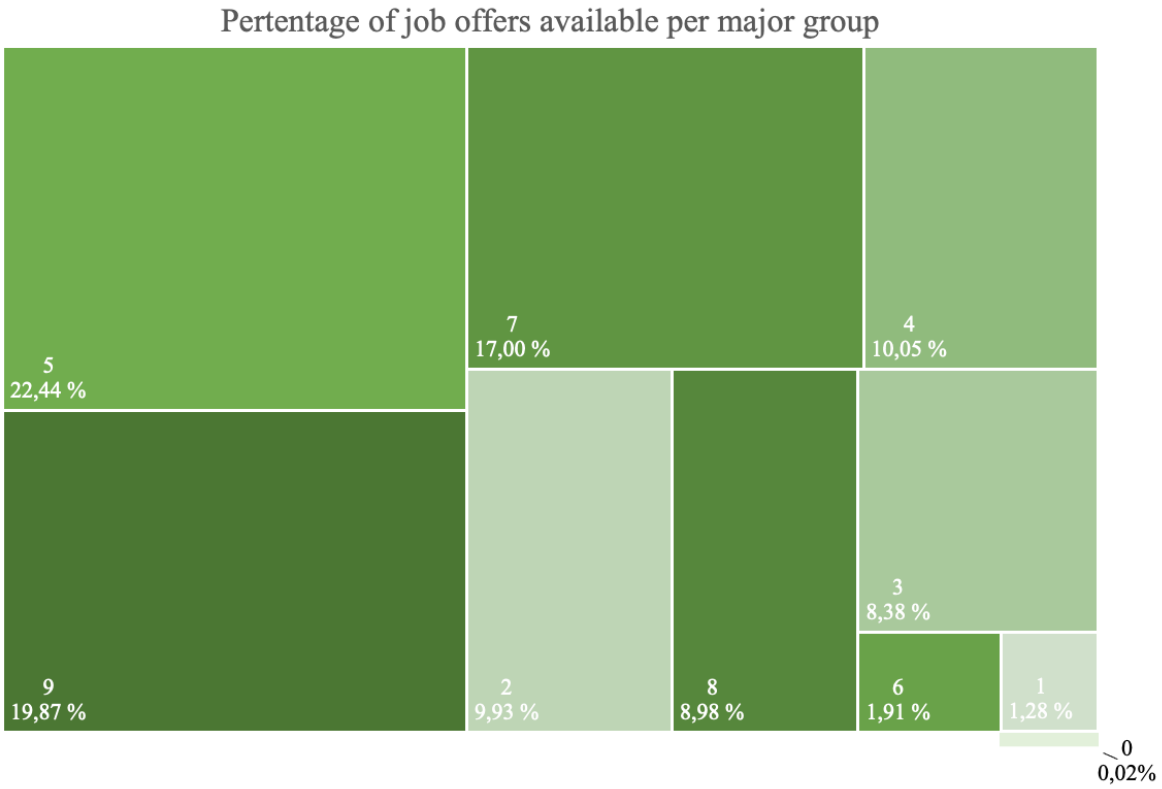

Figure 5. Treemap illustrating the percentage of job offers available per major group.

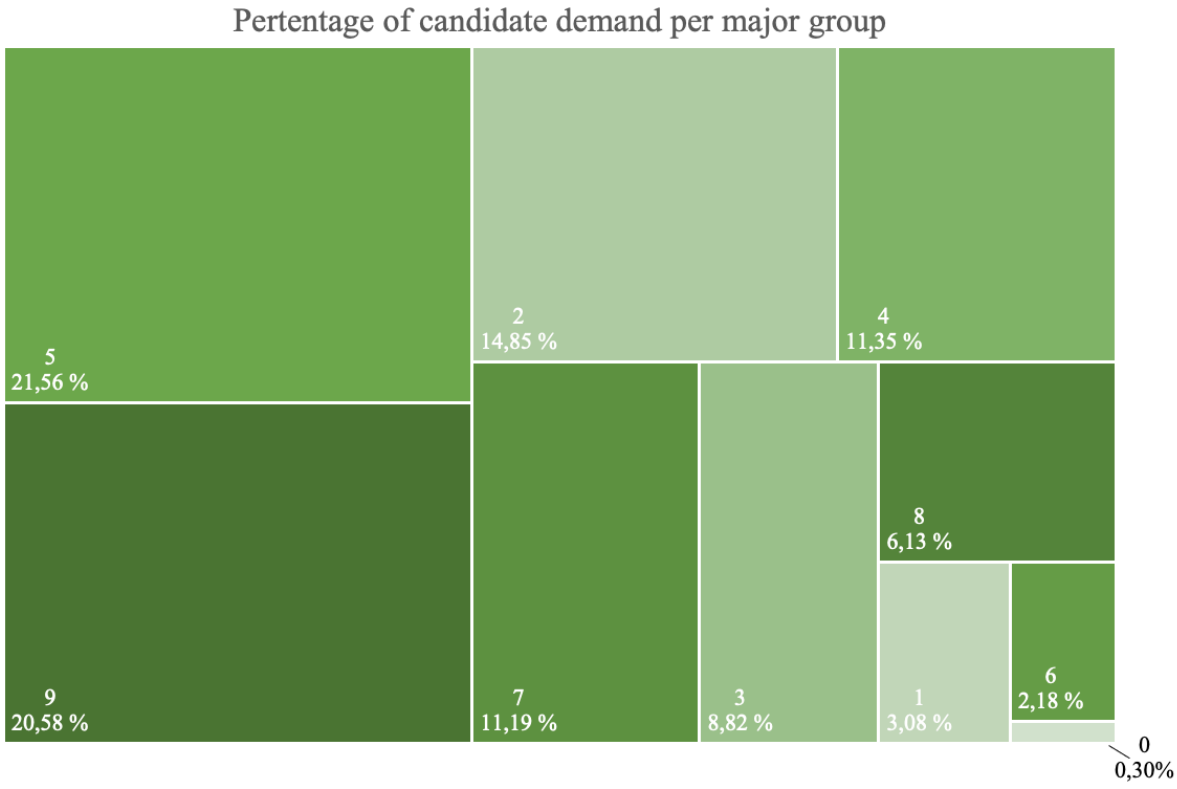

Figure 6. Treemap illustrating the percentage of candidate demand per major group.