Team Control Number

10765

Problem Chosen

A

2020
HiMCM
Summary Sheet

With job recruiting services becoming more popular and numerous, the amount of options for summer jobs is increasing every year. High school students are faced with tons of options and might have a difficult time deciding what services to use. Furthermore, no service lists all available summer jobs so students might miss a job that could have been a perfect match for them. On top of the difficulty of even choosing where to find a job, high school students also have to choose what job to apply for. There are many factors to take into account. High school students need money while still having some free time, and some may prefer a physical job over a sedentary one. We developed a model that finds the most suitable summer jobs for the students.

To even know what summer jobs to recommend, our model needed a reference of summer jobs. Hence, we made a database containing a collection of common summer jobs. We chose various factors we thought were relevant in terms of summer job preference and listed them for the jobs in the database. We also created fictional persons to test our model on and stored them in another database.

To determine which summer job would be the best fit for a certain person, we needed to rank the jobs based on the person's interests. The ranking is best done by assigning each job a score, that is calculated from its factors and the factors of the person. We calculated the score with a distance function weighted with inputs given by the person as factors. We also used the application programming interface (API) of our local public transport service to determine commute times, which were also passed into the aforementioned score function.

Our results showed that there was great variety in how much the jobs were recommended. In fact, the average ranking on recommendation lists showed a linear correspondence to the overall popularity of the job. The effect of location on the average score given to a job appeared to be noteworthy. The jobs located near key traffic areas were naturally more accessible, which increased their popularity.

The accuracy of our model had some flaws due to our data not being based on real summer jobs. If our database were to be reconstructed after getting access to proper data, more accurate and representative results could be obtained.

0. Table of contents

1. Introduction

1.1. Background

Nowadays, there are many different kinds of summer jobs to choose from. Young people often have trouble finding the best one for themselves. In addition to that, not everyone might be aware of all possible summer jobs in their area. Students look for various features and value different things. For example, some might appreciate high wage while others prefer flexible working hours.

To help high school students choose the best summer job for them, we need to create a model that recommends summer jobs based on the students' preferences, such as location and flexibility of working hours.

1.2. Problem restatement

Question 1: What factors should be taken into account when looking for a summer job? Identify all of the factors chosen, and list them with appropriate units.

Question 2. Develop a model or an algorithm that calculates what alternatives high school students have for summer jobs, using their preferences and situation as input.

Question 3. Come up with at least ten fictional persons that you use to test your model on. Explain the formulation of these persons and the selection of the data. Analyze the results of the application of your model on these persons.

Question 4. Explain how you would present the model to the high school students and make it usable for them.

2. Problem Analysis

Question 1. We need to consider the features possessed by summer jobs in general and make our choice of factors based on them. Additionally, we need to contemplate what factors have an impact on choosing a job. To make it more realistic, we need to take into account the importance of each factor for each individual.

Question 2. We need to find many different kinds of summer jobs and create a database including all of these options and their features. After that, we need an algorithm that determines the best job for the applicants.

Question 3. After developing our model, we need to create fictional persons to test the model on. In order to get accurate results, the group of these fictional persons needs to be as diverse as possible for us to be able to test the practicability of our model.

Question 4. We need to consider different alternatives for presenting our model. For example, an app or a website could be a suitable form. In addition to that, we could advertise our app or website on our schools' social media or newspaper.

3. Assumptions

Assumption 1: We limit our research to our home city. Our home city is the most familiar to us, so selecting it allows us to focus our resources better. Our home city also has areas with different levels of urbanity, so it has appropriate places for many different summer jobs. The local public transport service also has a public application programming interface, which we can use to give more proper summer job recommendations based on travel time.

Assumption 2: High school students look for jobs inside their home city. In order to adhere to the first assumption, we will also have to limit the job seeking scope of the students to our home city. This assumption does not harm the truthfulness of our model, since our city is of large area and hosts many different kinds of environments for summer jobs.

Assumption 3: High school students look for summer jobs that match their interests. Given that the summer break is only a limited time period, students want to spend it as pleasantly as possible. Since working at a job that is not interesting would probably lead to less pleasant time, the students will avoid that.

Assumption 4: High school students will be content with a job that does not perfectly match their desires. The students want to get the best possible job, but finding a job with exact matches to one's interests is nearly impossible. The students acknowledge this and will be content with the job that best suits them.

Assumption 5: High school students are not hesitant to take a job that matches their interests better than they specify. Although students want to get a job that is optimal for them, they will often be satisfied with a less suitable job. When given the opportunity to take a job better than they specified, they will take it as gladly as a job that barely matches the specification.

Assumption 6: The summer jobs and seekers are spread evenly throughout the city. Since it is November, we don't have real data of the locations of different summer jobs. To give a representation as close to reality as feasible, we assume that the jobs are distributed uniformly in the given area. We could use a different distribution that would closer approximate reality, but this assumption does not harm the representation because we use the same distribution for the job seekers as well.

Assumption 7. High school students use the model to search for a certain type of job. To avoid overcomplicating our model, it only searches for a job that matches the specified criteria. If, for example, a student would be happy with a job with a good pay and few working hours, or a job with lesser pay but more hours, they would use the model twice, specifying the different criteria in separate runs.

Assumption 8. High school students only use the mean of transport they specify. As with the previous assumption, this assumption is made to keep the model simple and elegant. If multiple transportation methods are applicable to a student's situation, they will be instructed to run the model twice with the different transportation methods on separate runs.

4. Model

4.1 Overview

Our model uses data from both the summer jobs and the persons looking for them. All of the jobs and persons are listed in separate tables with the values of all factors for each entry. Our model then reads the data and scores the summer jobs for each person based on how closely the jobs' factors and the persons' inputs match.

4.2 Choice of factors

For the quantitative factors of the jobs, there were some obvious choices: namely **hourly wage** (USD/h), **weekly working hours** (h/week), and the **frequency** of the occupation in the city. We also decided to include the **position** of the job, ie. the coordinates (latitude and longitude) measured in degrees.

We chose six qualitative factors based on our own experience and common interests shared by many high school students: **physicality**, **sociality**, **work time flexibility**, **concentration**, **monotony**, and **on-call**. In light of the current COVID-19 pandemic, we also decided to include whether a job could be worked **remotely** or not. On-call and remote are boolean values. A job either is or isn't on-call work or worked remotely. All of the other qualitative factors are measured on a scale from zero to three, with zero being the least and three the greatest plausible value for each of the factors. We also gave each job that is not remote a longitude and latitude, which tell where the jobs are located.

For the persons, we used all of the same factors except for frequency, which does not apply to people. Furthermore, we introduced a new quantitative and a new qualitative factor: **travel duration** and **vehicle**. Travel duration represents the amount of time the person is willing to spend on their commute in minutes, and vehicle is the mode of the transport the person uses, either walking, cycling, using public transport or driving. For all of the factors besides vehicle and position, we also included a weight factor, which is a number between one and zero. It represents how much the person wants to emphasize the factor in question.

A comprehensive list of the factors is included in Appendix A.

4.3 Database composition

Our model requires a database to get the information of the jobs from. We weren't able to find a complete database with all the factors we needed, so we created our own. We used various sources to collect the information for our database. Since it's November, we could barely find any open summer jobs. We thus looked for lists of common summer jobs and gathered them into our database.[2,3]Then, we searched for all of the factors for the jobs one by one using our sources and devised reasonable values where necessary.[1,5,6,7] We made a number of instances of each occupation equal to the frequency of the occupation in question. For each separate job, we used a uniform random number generation algorithm to generate an even distribution of locations for the jobs in the city. We created 272 different jobs with unique locations in total.

4.4 Ranking the jobs by suitability

In order to give meaningful information to the user, our model has to sort the available summer jobs based on how well they fit the user. For each of the factors, it calculates the square of the difference between the job's factor and the person's input and multiplies the result by the appropriate weight given by the person. Before that, all of the values are normalized, i.e. the reduced to numbers between one and zero, with one being the maximum and zero the minimum for the factor in question. This is done to give each factor equal influence on the score. This method is called the weighted mean squared error, and is given by the following formula.

$$E = \frac{1}{n}\sum_{i=1}^{n} w_i(Y_i - \hat{Y}_i)^2 \qquad (1)$$

Here, n is the number of factors, w_i are the factor weights specified by the person, Y_i are the comparable factor values specified by the user, and $\widehat{Y}_i$ are the comparable job factor values. The less the score, the better fit the job is for the person. All but frequency, vehicle, coordinates are comparable factors.

For the persons' distances, we retrieve travel data from our city's public transport service's online routing API. Using the coordinates of the job and person, we get an optimal route from the person's location to the job using the vehicle the person specified. For remote jobs, we don't use the location and just leave the travel time difference out of the calculation.

If the job's salary exceeds that of the person's wish, we evaluate it as the amount as the person wished. The same is done with travel time, but with shorter commutes evaluated as the person's wish. This is justified by Assumption 5.

4.5 Formulation of fictional persons

To create the fictional persons for testing our model, we decided to use nearly orthogonal Latin hypercube (NOLH) sampling, utilising a free spreadsheet published by the Naval Postgraduate School.[4]

In orthogonal sampling, the parameters for the samples are chosen so that any two samples have as little correlation as possible.[8] For models where small changes in parameters lead to a small change in the result, this sampling method works very well, because one can represent the entire population (in this case, all possible students) with a very small sample size.

A small sample size was critical for our model, because the runtime was proportional to sample size. Accessing the routing API, which needed to be called hundreds of times for each student, proved very slow. It would have been impractical to reach the same variety with random sampling, but with NOLH we only needed 129 students for our 22 factors.

4.6 Further development

Our model has some inadequacies in various places. The largest problem is indisputably the absence of real job data. Our model merely gives suggestions on what kinds of jobs could the

user search for and no guarantees that the jobs actually exist. If real data were used, all of the factors could be made more accurate and representative of the jobs.

Another inaccuracy is that we assume the jobs and students are distributed evenly throughout our city. In reality, the population density is greater near the city center and the distribution of the jobs probably follows the same trend.

5. Results

5.1 Summer jobs by popularity

Figure 1 shows how many times each job appeared as the first recommendation. There is a clear outlier in the data: graphic designer stands out with its large score. Surprisingly, it was recommended as the first option for half of the fictional persons.

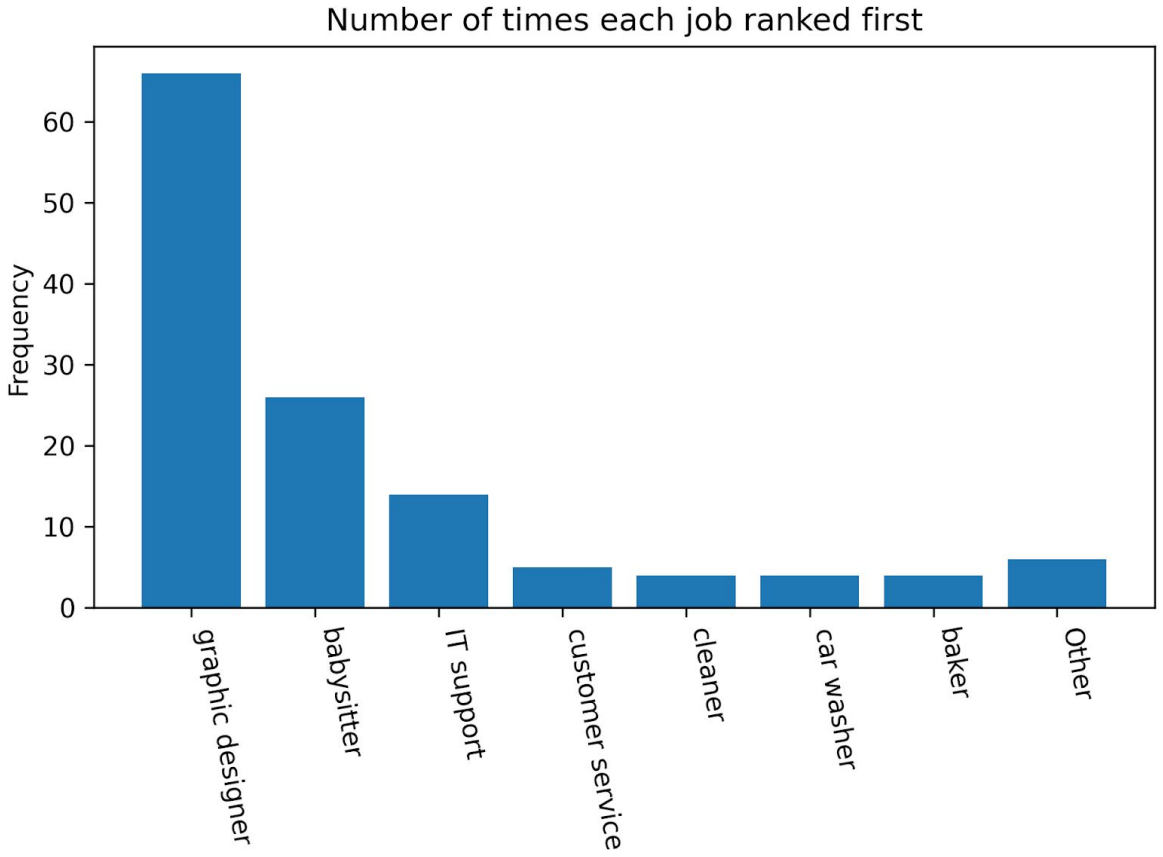

Figure 1: *How many times each job was ranked first in a person's recommendations.*

After analysing our job database, we noticed that graphic designer had very average values for every factor, which caused our model to recommend it over other jobs with more extreme values. In addition, graphic designer was marked as a remote job, and therefore it completely ignored commute time, which explains the other remote jobs scoring high.

We also calculated the average popularity of the jobs. Our results showed an approximately linear progression from most popular through to least popular. Figure 2 shows the average

rank (e.g. first, second, etc.) of each job. Note that smaller bars mean better ranking, since the smaller a rank is the closer it is to the first place.

In this ranking, babysitter surpassed graphic designer. Babysitter, too, had very average attributes, but alongside that, there were 20 instances of it. This meant that often when it was ranked higher than graphic designer, all or most of the other babysitter jobs would be recommended right after it. This dropped graphic designer's average ranking significantly. A fix to this problem could be adding more probabilistic factors to the model to increase deviation within the job titles.

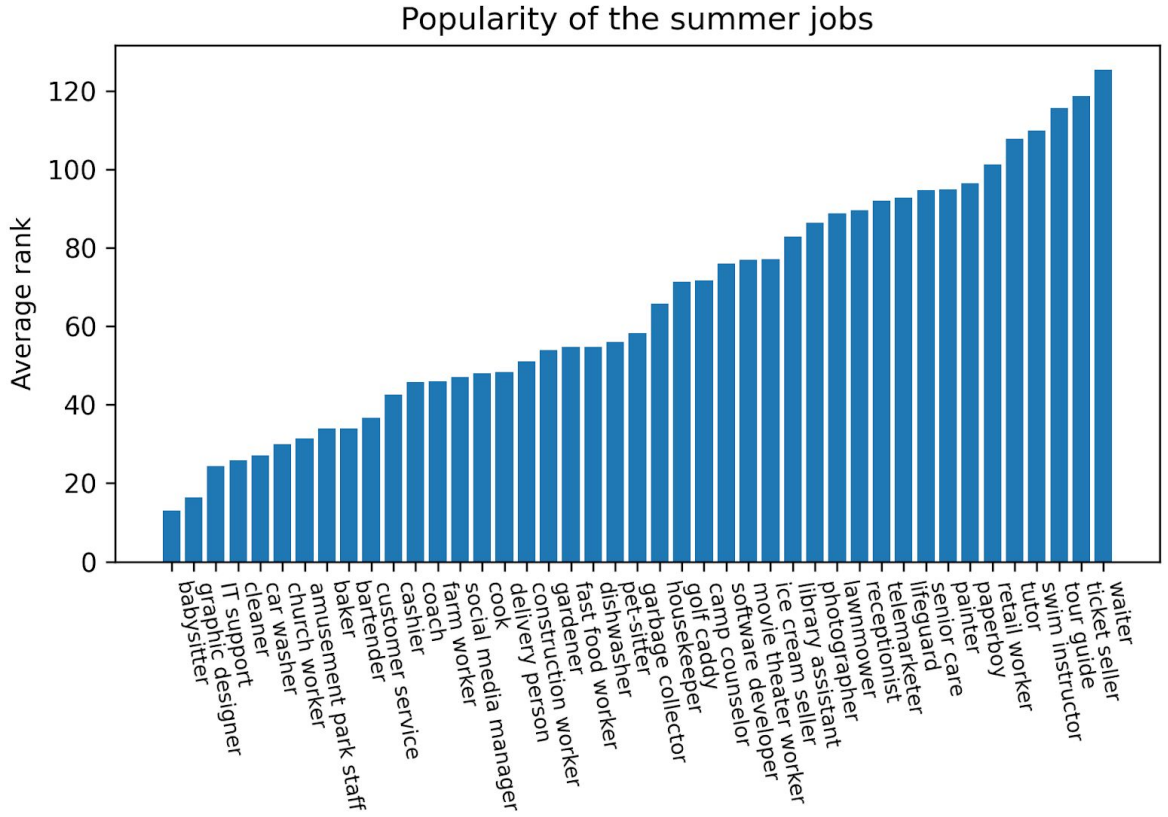

Figure 2: *Occupations ranked by popularity calculated by the described expression.*

This average rank is calculated with the following formula.

$$S(w) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{f_w}\sum_{j=1}^{f_w} R_{ij} \tag{2}$$

Here, $S(w)$ is the average rank for the occupation w, n is the number of persons, f_j is the frequency for the occupation j, and R_{ij} is the ranking of job j for person i. In other words, we take the mean of each job's ranking on each person's recommendation list.

5.2 Popularity distribution

By plotting the popularity of each job in a figure where the x and y axes are the longitude and latitude of the job (Fig 3), we observe interesting patterns.

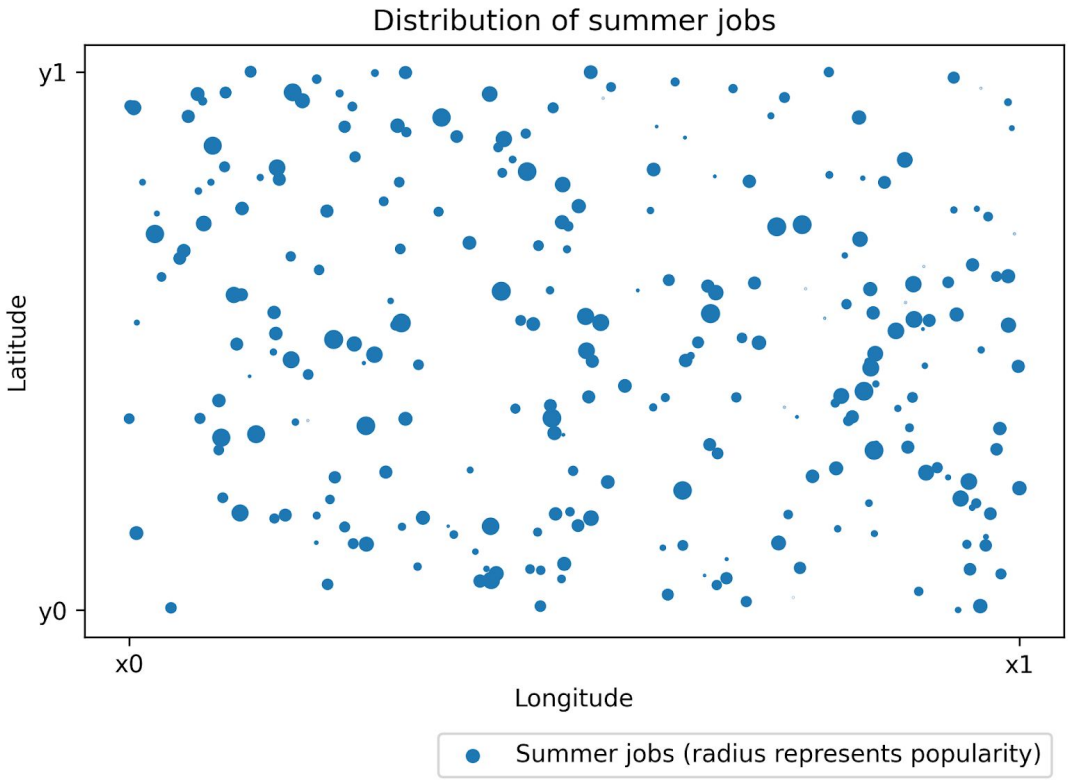

Figure 3: *Summer jobs in the city. Bigger circles represent a bigger population.*

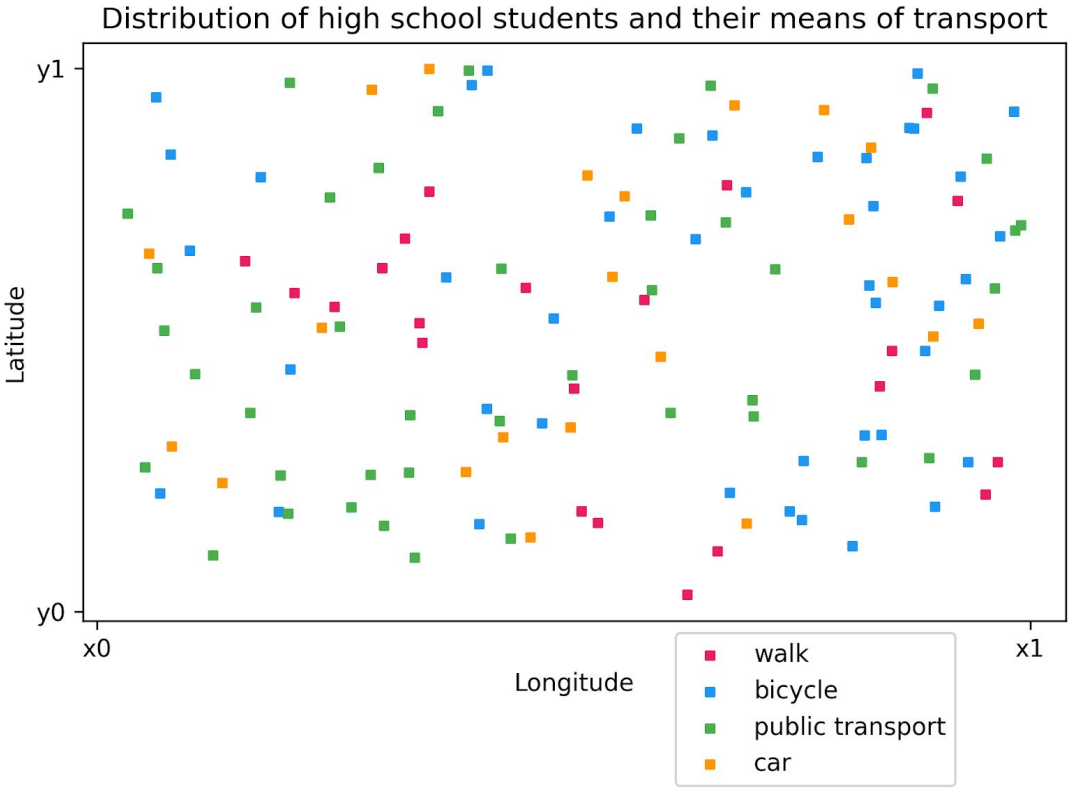

Figure 4: *The students' locations in the city. Color indicates mode of transport.*

The popular jobs form string-like clusters in various areas. By comparing these clusters to the map of our city, we notice that they match active streets and central areas of our city. Especially public transport and car traffic benefits from having good traffic routes, so the jobs that lie in these places are naturally more popular on average. For reference, Figure 4 shows the distribution of the students and the mode of transport they prefer to use.

5.3 Accuracy

Considering that for half of the test persons our model recommended graphic designer, our model seems to favor median factor values. With real data, the accuracy could be improved as outliers would be less frequent.

Our model still has some room for improvement. Our results suggest that our distance function (eq. 1) may not be the best method to rank the summer jobs, as it clearly favored average jobs. Moreover, the factors could be reconsidered since they do not differentiate jobs very well.

The accuracy could also be increased by adding more probabilistic factors. Currently, the only non-deterministic attribute is the location of the job, which was randomly chosen from a uniform distribution. This would add the much-needed variety.

6. Presentation to the public

6.1 Getting proper data

A crucial part of the algorithm is the database. If our program is to be used in real life, it will need a proper database to get the information from. There are a couple of ways to carry this out.

We could make a web service where summer job employers could give information about their open spots. This would require the employers to be aware of our service and willing to use it. In order for our service to gain popularity, extensive advertising would be necessary. Furthermore, our service would probably only get a fraction of the jobs open for application.

On the other hand, we could create a bot that finds summer jobs from the internet and fetches information about them to our program. The problem with this approach is that all of the information needed might not be available. Furthermore, scraping data from websites would be unreliable and vulnerable to false information.

Both of these options have many shortcomings. Thus, the best alternative is to take the best of both worlds and make it possible for employers to add their summer job positions to our website, but also have a bot that looks for summer jobs from various other websites. The jobs reported by the employers would have official and confirmed data, while the jobs fetched from the web would have a warning that the data may be incorrect. This would be a compromise between reliability and broadness, with clear communication to the user.

6.2 Using our algorithm

In order to get our algorithm to the public, we would create a website for people to use it. The website would ask the same factors as our algorithm, although in a more user-friendly manner, and return the most suitable summer jobs. After having our website running for a while, we would develop a mobile application. The app would have the same features as the website but for example saving your results would be easier. For jobs submitted by real employers, we would have a way of applying to the job right from the website or application itself. This would make the process of looking for and applying for a summer job much easier and simpler.

6.3 Advertising

Advertising is an important part of getting users and employers for our website and app. Advertising would be the most efficient if we did it in high schools and on social media platforms used by high school students. In addition, to get employers to submit their job information to our platform, we would also advertise on social media that employers use, such as LinkedIn and Twitter. If we really wanted to get our service going, we could found a startup in order to attract investors and employers.

In high schools, the advertising could be done by putting posters on the walls. They easily catch people's attention. The posters could include a QR code to the app or the website. We would also advertise our program on our school's official communication platform and social media. We could, for example, get our school's Instagram page to promote us and spread our program to our friends and peers.

We designed a poster for our service, and included it in Appendix B.

7. References

1. AllAboutGrads. "Careers | AllAboutCareers." Accessed November 11, 2020. https://www.allaboutcareers.com/careers/.
2. Indeed. "Job Search | Indeed." Accessed November 11, 2020. https://www.indeed.com/.
3. PayScale. "Job Salaries | PayScale." Accessed November 11, 2020. https://www.payscale.com/research/US/Job.
4. Sanchez, S. M. NOLHdesigns spreadsheet. 2011. http://harvest.nps.edu/.
5. The Skills Development Scotland. "Job Profiles | My World of Work." Accessed November 11, 2020. https://www.myworldofwork.co.uk/my-career-options/job-profiles.
6. Totaljobs Group. "Salary checker - A salary calculator tool for different industries, skill sets, job roles and locations." Accessed November 11, 2020. https://www.totaljobs.com/salary-checker/salary-calculator.

7. Trendler, Alex. “What time does a bartender's shift start?” Quora, August 12, 2016. https://www.quora.com/What-time-does-a-bartenders-shift-start.

8. Wikipedia. “Latin hypercube sampling.” Accessed November 15, 2020. https://en.wikipedia.org/wiki/Latin_hypercube_sampling.

8. Appendices

8.1 Appendix A: Factor tables

Job

factor	type	units	description
occupation	string	-	what occupation the job is
frequency	integer	-	how many jobs of this occupation are created
hourly wage	integer	USD/h	hourly wage of the job
weekly work hours	integer	h/week	how many hours of work per week
physicality	integer (0 to 3)	-	how much physical strain the job involves
sociality	integer (0 to 3)	-	how much social interaction the job involves
flexibility	integer (0 to 3)	-	how flexible the working time is
concentration	integer (0 to 3)	-	how much concentration the job requires
monotony	integer (0 to 3)	-	how monotonous the job is
on-call	boolean	-	is the job on-call work
remote	boolean	-	is the job remote
coordinates	coordinates	degrees	coordinates of the job

Person

factor	type	units	description
vehicle	string	-	“walk”, “bicycle”, “public”, or “car”
travel time*	integer	minutes	the preferred duration for the commute
hourly wage*	integer	USD/h	hourly wage of the job
weekly work hours*	integer	h/week	the preferred hours of work per week
physicality*	integer (0 to 3)	-	the preferred amount of physical strain

sociality*	integer (0 to 3)	-	the preferred amount of social interaction
flexibility*	integer (0 to 3)	-	the preferred working time flexibility
concentration*	integer (0 to 3)	-	the preferred amount of concentration needed
monotony*	integer (0 to 3)	-	the preferred monotony of the job
on-call*	boolean	-	is on-call work preferred?
remote*	boolean	-	is remote job preferred?
coordinates	coordinates	degrees	location of the person's home

*For each factor marked with an asterisk, there is a weight factor, which is described below.

factor	**type**	**units**	**description**
weight	floating point	-	weight for the appropriate factor

8.2 Appendix B: Poster

Having trouble finding a summer job?

We have the perfect solution for you!

Facts

- About 30% of 16-19-year-olds are working during the summer[1]
- The amount of teenagers working during the summer is decreasing[2]

About our app

- It suggests the most suitable jobs for you based on your personal preferences
- It has an extensive range of various summer jobs
- You can emphasize the features you value the most in a job

Benefits of a summer job

- You gain valuable work experience that will help you in the future
- You earn money
- Meeting amazing new people
- Getting the hang of work life

Insert QR-code

Try it out!

References

1. Ranzetta, Tim. "Chart: What Percentage of Teens Have Summer Jobs?" Next Gen Personal Finance. Jul 12, 2016. https://www.ngpf.org/blog/career/chart-percentage-teens-summer-jobs/
2. Desilver, Drew, "In the U.S., teen summer jobs aren't what they used to be." Pew Research Center, June 27, 2019. https://www.pewresearch.org/fact-tank/2019/06/27/teen-summer-jobs-in-us/