

It must be stated first that we still have not gotten the dataset, even after contacting the project owner as well as teachers.

Business understanding

Our business goals

Origin of the project

Nowadays, students often struggle to identify suitable scholarships mainly because information is scattered across different platforms, eligibility criteria are complex and sometimes the scholarships are not advertised enough to reach some students.. As a result, many eligible students miss out on financial support.

This project aims at building a recommendation tool that automatically matches students to the most suitable scholarships based on the profiles of the student and the scholarship.

Business Goals

As this project is meant to be used by students for the sake of obtaining scholarship money, in the event where the project is distributed to students by schools then it wouldn't make up for a business.

But in the event where the project would have to be a standalone business providing its services to students, then it would have a market.

The goal of the business is inherently to increase access to financial aid by providing a service that simplifies the scholarship search process for students. In turns, making them gain a better information funnel and gain time as well in exchange for money.

A second goal is to enhance the base of users would be to provide marketing regarding the personalization and adaptability of the solution. For example by making an emphasis on the fact recommendations are given according to the best options for their profile.

The business should also gather data of its users and scholarships in order to be able to potentially sell information regarding student profiles, scholarship trends and how they correlate. Which could be sold to schools or institutions.

As such, the key goals of such a company would be to spread their service to a wider audience while also being able to keep track of their audience (data gathering) not only to understand the trends but to maybe sell key information to institutions that may need it.

Business Success Criteria

Evaluating the success of such a business can be done in a multitude of ways but mainly through performance indexes:

The first performance index that should be implemented is the accuracy of scholarship matching to the student profiles as it is the core function of the service and should be performing at its highest at all times. To evaluate the accuracy we could implement the computation on different levels, on a basic level it could simply be, “did the student register to the scholarship that was recommended” with a simple yes or no answer. Then, one would be able to make a more in depth indicator by comparing characteristics between students and scholarships.

Other criterias of success would be the use rate of the platform, it could be measured by having students confirm when they get a scholarship with the help of our platform and then compare the trends of users getting scholarships through our platform and through manual search.

One should also look if students using the tool apply to more scholarships on average.

A final success criteria to cement the business in its market would be partnerships with schools which could be set up as us exchanging the data of their students so they can study the trends of their students while we get an enlarged user base and database.

Assessment of our situation

Inventory of Resources

Our available Resources:

Though we didn't get access to actual data we still got a hold of a website containing examples of data. Thus, we currently have a rough idea of the type of data we will use. We have a clear idea and definition of the different criterias that will define a student profile and a scholarship profile. We can always use pre-existing similar features from other same concept websites to refine our approach.

Human Resources:

Our human resources consist of us. As such we are 3 Robotics and Computer engineering students with knowledge of entrepreneurship, economics, machine learning and coding in general.

Requirements, Assumptions, and Constraints

Requirements

The requirements of the projects would be that we must allow students to create profiles and have a database to store the information while also allowing them to update said information.

Data from scholarship and students must be standardized, for example scholarships would need standards in criterias, deadlines, and eligibility requirements. While students would require standards with their financial situation, field of studies.

As we store data we also need compliance towards the data and privacy regulations.

Assumptions

We assume students are willing to input accurate personal and academic data. We also assume that scholarship providers will maintain updated information. Finally, we assume to have enough resources or be provided enough resources to maintain the service.

Constraints

Our constraints regarding the project will be that we still haven't received the data as of the submission deadline of HW10. Another constraint would be that information provided by third-parties, regarding students or scholarship may be incomplete or inconsistent.

Risks and Contingencies

Risks of our projects would be that scholarship data could become outdated which would directly cause issues with the accuracy of recommendation and which will require automatic and periodic updates of the data to ensure the most recent information is acquired and shared to users.

Another risk would be third-parties incorrect inputs regarding student information. In fact, it would directly influence the model in creating wrong matches. But it can be contained by input validation processes and potential data validation by involved schools.

A risk would be that the service is unattractive to students, thus if priced it should be reasonably priced for students, if distributed by schools then we make the deals with schools directly. For the students to adopt our platform we need user-friendly UI and a potential integration with student portals while also relying on marketing with advertisement campaigns on campus.

Terminology

- **Recommendation Engine:** The AI model that ranks scholarships based on a student's profile.
- **Eligibility Match:** A measure of how well a student fits the scholarship criteria.
- **User Profile:** Student-entered data including academics, skills, background, and interests.
- **Scholarship Dataset:** A structured collection of scholarship opportunities.
- **Feature Engineering:** Transforming raw data into usable inputs for the machine learning model.

Costs and Benefits

Costs

The main costs of the business will be:

The development and maintenance of the platform.

The potential hiring of data science and engineering personnel in the future.

Benefits

The benefits the business brings:

Time savings for students and advisors.

Increased scholarship subscription rates.

Competitive advantage for institutions offering the tool.

Potential for long-term expansion into grants, internships, or career pathways.

Our Data-Mining Goals

Data-Mining Goals

Since we don't have access to the proper data, we will reason from the scholarships that are able to be seen from the provided website :

<https://sihtasutus.ut.ee/scholarship-competitions>

Thus, we will first gather all the scholarships and classify them by categories (e.g., merit-based, need-based, field-specific and location). Then we can rank them by comparing how much they fit to a student's profile. Thus, the goal of the data-mining is to identify the best categories to use for both scholarship and student's profiles to ensure the best matches possible..

Data-Mining Success Criteria

We can consider our data-mining successful if our matching model achieves a good enough accuracy baseline of above 80% in identifying scholarships for which a student is truly eligible.

Also if our recommendation rankings align with real user preferences (validated through feedback or A/B testing).

Data understanding

At this stage of the project, we have not yet received the dataset that was initially expected from our stakeholder. As a result, our data understanding is currently based on assumptions derived from the project brief rather than on concrete analysis of real data.

From the information provided so far, we expect the final dataset to describe scholarships and related opportunities in a structured way like the Scholarship Competitions part of the University of Tartu website . In particular, it is likely to include:

- **Scholarship Fields** — broad thematic categories such as *Interdisciplinary Scholarships, STEM, Arts, Social Sciences, International Mobility, Research Grants*, etc.
- **Form of Study / Profession** — descriptions of the academic level or professional activity targeted by each scholarship, such as *Bachelor's Studies, Master's Studies, Doctoral Studies, Postdoctoral Research, Professional Training*, or *Continuing Education*.

These attributes will be central to the development of our tool, as they allow filtering, matching and ranking scholarships based on user profiles. However, until the full dataset is delivered, we are unable to explore data and analyse, assess data quality, or determine the presence of additional relevant attributes (e.g., eligibility criteria, deadlines, location, funding amount, or required documents).

Once the data becomes available, we will revisit this stage to:

- Explore and validate the structure of the dataset

- Evaluate data completeness and consistency
- Identify noise, missing values, or unstructured fields
- Define the preprocessing steps required for feature extraction and filtering

For now, our understanding remains high-level and will evolve as soon as the actual dataset is provided.

Planning the project

Task	Time Allocated
Extract the data, clean it, and scrape other if needed	Loic: 1.5h Guillaume: 0.5h Mattis: 2h
Defining categories and features of each scholarship	Loic: 1h Guillaume: 2.5h Mattis: 0.5h
Creating lists of characteristic features for student's profiles to create cross references between profiles and features linked	Loic: 3h Guillaume: 2h Mattis: 5h
Creating or editing a search tool to make use of key-words related to features/categories for better search results	Loic: 5h Guillaume: 2.5h Mattis: 3.5h
Create Poster	Loic: 2h Guillaume: 2h Mattis: 2h

Github Repository Link:

<https://github.com/SocraticJunkieWannabe/F2-UTF-SCHOLARSHIPS>