AI FOR ACADEMIC GUIDANCE

AMRI Tarik[1], IRGUI Ilyas[1], LOTFI Firdaouss[1], BAALLA Yasser[1]
BELKADI Manal[1], YAZIDI Ikrame[1], BAHNASSE Ayoub.[1]

[1]ENSAM CASABLANCA, KAAS Club

April 1st, 2021

1. Refined Project Description

Decision-making is an important process for the development of the individual. The former is not always obvious, since hasty decisions can have disastrous results on the person's future, hence the importance of guidance in general, and in education in particular. It is from this perspective that we are proposing a new AI-based solution that will ensure that undergraduate students get the best recommendation regarding which course of study to follow based on their grades, preferences, and expectations. Our solution is easy to use and to implement, allowing universities to adapt it according to their curricula and students to benefit from its features online. Through this approach, we bring socio-economic benefits. Indeed, recommending the right course of study to a student will help them have more motivation, and will ensure a better productivity in the job market.

2. Refined Project Plan

Our team is composed of the supervisor Dr. Bahnasse Ayoub and 6 members including the team leader Mr. Amri Tarik (A.T). The team members are Mr. Baalla Yasser (B.Y), Ms. Belkadi Manal (B.M), Mr. Irgui Ilyas (I.I), Ms. Lotfi Firdaouss (L.F) and Ms. Yazidi Ikrame (Y.I.). Our work is divided into four main parts, namely Research and Training, Preparation of the Environnement, Design and Creation, and finally Implementation. Regarding Research and Training, we started by studying and comparing the different machine learning algorithms as well as following training courses in the necessary disciplines, for Example Python for Data science, AI and Development, Git and Github, Js, nodeJS, ReactJS as well as the AGILE Scrum training. Undoubtedly, all team members were involved in the research and training stage. About the Environnement setup phase, we opted for the installation of the Anaconda Navigator and the creation of a github repository, then, the implementation of the necessary frameworks, and finally the creation of a virtual environment and the management of dependencies between the team members. As for the Design and Creation, we dispatched the tasks to team members, including the front-end design using Adobe Xd by (I.I), (B.M) and (A.T), as well as the data flow diagram by the same group of people accompanied by (L.F) and (B.Y). Moreover, the creation of the survey was carried out by (B.M) and (Y.I). On the other hand, the choice of questions was carefully studied with coaches and guidance experts. Lastly, we covered the part of the implementation starting with the front-end implementation using JavaScript by (I.I) and (A.M), then the implementation of an API using nodeJS by (B.Y), (Y.I) and (B.M), that were joined by (L.F)

during their work on the database and its indexing. Figure 1 illustrates the project's Gantt chart.

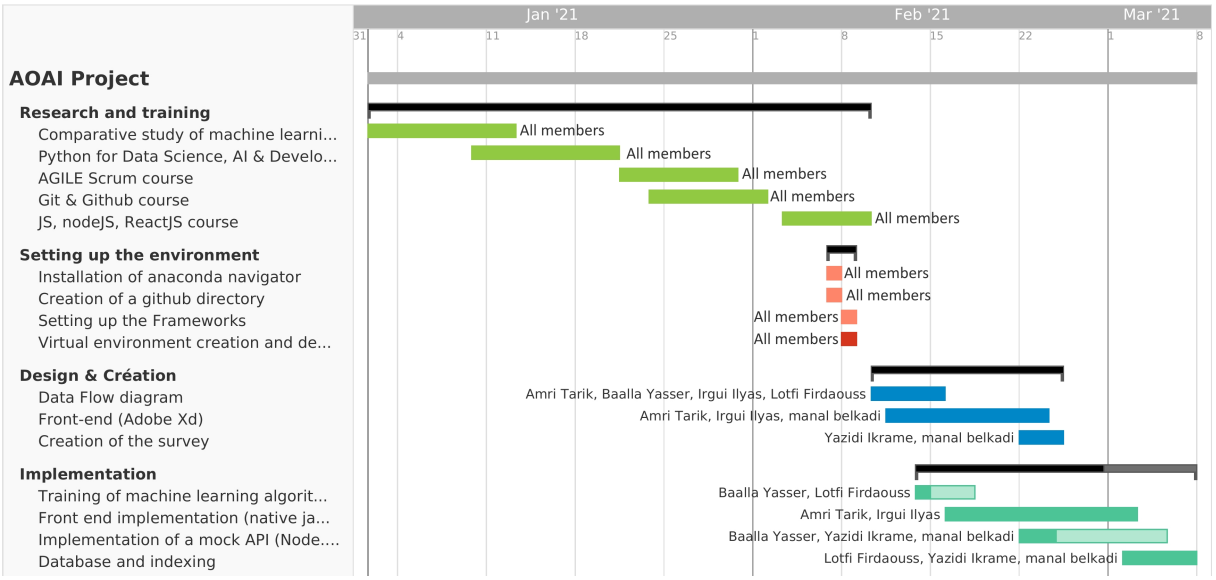

Figure 1: Gantt chart

It is obvious that the team faces several challenges related to productivity and communication during the project execution. For this reason, effective communication is ensured between team members using the SCRUM method and the Trello platform. We organize regular meetings to discuss the progress of each member as well as to manage the urgent obstacles the members face. And in order to maintain the progress of the project, we have made sure that each role is assigned to at least two people. This way, if one person finds it difficult to accomplish their task for one reason or another, their colleague will take over.

3. System Requirements

With respect to the collection of data needed for decision-making, much effort has been made because this type of information is not publicly available. In fact, the data is mainly information on the various courses of the school, the students' grades from the administration of the university, information on the laureates and their professional integration. Thus, many constraints were imposed regarding the data confidentiality. Consequently, the gathering of this data was only possible after long meetings with the head of the university and the general secretary. The second constraint was related to the

architecture of the information system, since a direct access to the database constituted a violation of the security policy. Therefore, sessions must be opened with the role of consultation only, in a way that connections will be made from an administrator's station with an account created specifically for this project. Apart from these imposed limitations, establishing a survey that can detect contradictions in answers, based on the weights of each question, was indispensable. Furthermore, making it accessible to a large number of laureates was only feasible by contacting the partnership affairs department in order to get their contact information. Indeed sharing these surveys in social networks brings a risk of collecting biased data. Obtaining this information was only a preliminary phase, analyzing the obtained data was another challenge. Without any doubt, it was essential to understand how the calculation of the semester averages is carried out for each major. In fact, each one uses a specific weight, in other words, the coefficient of a module declaring its importance relative to the concerned area. A meeting with the deputy director in charge of pedagogy was held to understand the flow of the process. During this experience, we learned to respect the hierarchy of a system in order to access its assets, while assuming all the involved responsibilities. This condition was met by signing a non-disclosure commitment for university information. Furthermore , this experience has shown us that data is the key success factor for any project, and that convincing stakeholders of the importance of the project greatly facilitates its success, especially when the latter is linked to students' future. We also learned from our discussions with the experts that the data collected from the laureates allows a margin of subjectivity either due to lack of concentration when filling out the survey or misunderstanding of certain questions.Therefore meticulous processing of the inputs is essential. The functional specifications/requirements of our project are the ability to classify the taught modules into specialties, the analysis of the responses of the laureates, and the decision of the most suitable program. The constraints of the functional specifications depend mainly on the volume of training data, the objectivity of the responses as well as the performance of the operating machine. Regarding non-functional requirements, we should mention portability or the possibility to run our solution in any deployment environment. Scalability is another fairly important non-functional requirement, the AOAI solution must support scenarios of universities where the number of students, courses and subjects evolve significantly, so that the relevance of the results isn't impacted by this evolution.

4. System Design

We use a micro-services architecture composed of two services coupled with a Single Page App (SPA). The latter was implemented using lightweight native javascript in order to easily integrate it into the websites of universities. The first microservice is an API that sends data to the SPA while the second contains the model that handles the processing of students' grades and other relevant data. The database, managed by the university for security and confidentiality reasons, will contain data collected from students and laureates and processed during the training phase. Figure 2 illustrates this architecture.

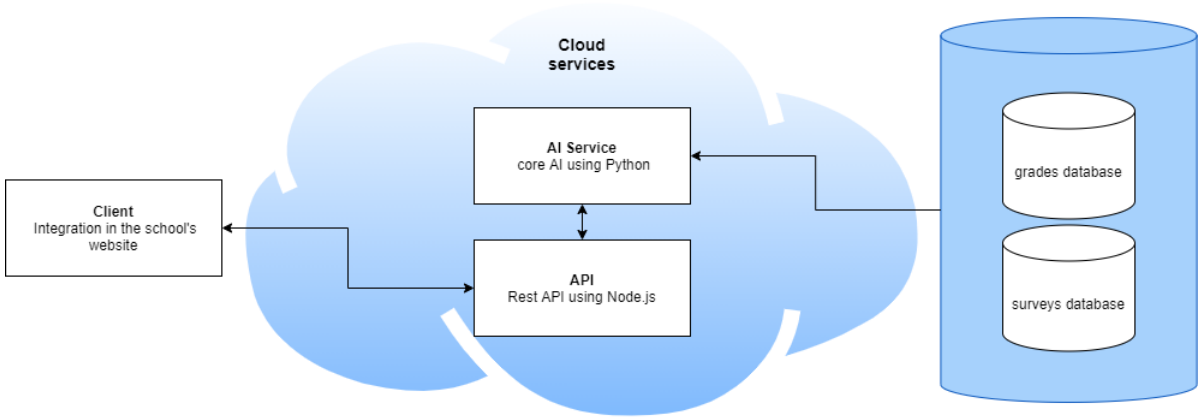

Figure 2: System architecture

Regarding our interfaces, the student can benefit from the AOAI solution via its university's portal where our solution will be integrated. First, the user starts by entering his/her username. Then, the platform displays the dashboard containing all possible majors along with the predicted percentage of the student's academic success in each one of them. After the user selects a major, a dynamic survey is provided to take into account their preferences. Once the survey is completed, the data in the dashboard will be updated with new recommendations and statistics. The student can then download a detailed report that contains the previously mentioned statistics along with explanations of the proposed results. Figure 3 illustrates the data flow diagram :

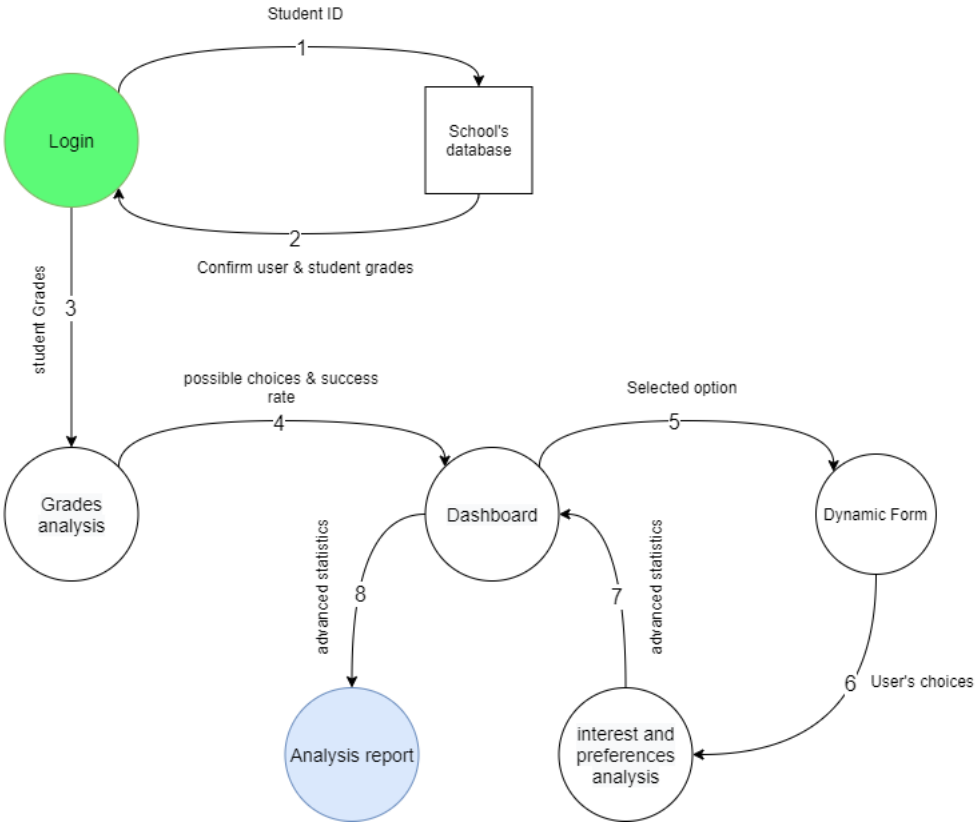

Figure 3: Data flow diagram

The difference between traditional orientation techniques and AOAI is that the latter takes into account not only the students' grades but also their professional expectations in order to recommend the most suitable course for their profile.

5. System Implementation

During the development phase, we relied mainly on two languages: Python and Javascript. Javascript allowed us to quickly develop the front-end as well as the back-end using the Node.js runtime environment as well as the Express.Js library. We served the front-end with the nginx CDN software for simplicity. As for Python, we used the Anaconda development environment which allowed us to have several powerful tools like jupyter, VS code and even the creation of virtual environments. Our solution is also based

on several popular standard open-source machine learning libraries such as sklearn, skmultilearn, numpy, scipy, pandas as well as other complementary libraries. In order to guarantee the isolation of our AOAI solution from the deployment environment's host OS, we suggest the Docker containerization solution.The choice of Docker will allow us to meet the non-functional requirements related to portability, implementation speed, and scalability.

6. Other Relevant Issues and Challenges

Undoubtedly, we faced many challenges, specifically those related to hardware. One of the more prominent ones involved the limited performance of our personal machines, hence the need to use either our university's server or the public cloud. However both solutions have limitations for long-term use. As a matter of fact, our university imposes strict policies that restricts the use of the servers remotely, allowing us a limited time span during breaks to take advantage of them. The use of public clouds, on the other hand, is quite expensive. Aside from the free balance offered during the trial period, it is difficult to sustain without funding given the considerable computational capacity that training requires. Everything considered, access to a cloud service will help us tremendously in implementing our project. Other constraints are manifested in the patenting as well as the commercialization of our solution. This matter arose when we considered exploiting our idea to found a company, as a certain expertise is required in the field of patents and the protection of intellectual property. For example, support by experts from a renowned brand such as DELL in this direction will be extremely beneficial to us.

List of Figures