esprit
Se former autrement
HONORIS UNITED UNIVERSITIES

Career Center Platform

Supervised by:

Ms. Dorra Trabelsi

Submitted by:

Firas Chkoundali

Dhouha Hmem

Rima Reggui

Chahine Jday

Feryel Dridi

Eya Guirat

Prepared at:

ESPRIT in 2022/2023

Table of contents

Table of Figures

Abstract:

The rapid development of data science and related technologies has brought about a significant transformation in a variety of industries and domains, including employability. With the increasing amount of data available and the advancements in data analysis and utilization, data science has the potential to provide organizations with deeper insights into the job market and the skills required by various industries.

This project aims to demonstrate the impact of data science on the domain of employability and how it can help organizations like the Employability Pole of the Esprit Group to overcome the challenges they face and improve their services. By utilizing data science methods and techniques, the PE will be able to manage the job and internship offers it receives more effectively, gain a deeper understanding of the job market, and provide students with valuable guidance on their career paths.

Introduction :

The Employability Pole (PE) of the Esprit Group is facing challenges in managing the vast number of job and internship opportunities it receives, as well as gaining visibility into the skills required by the job market and the geographical distribution of job and internship opportunities. However, the application of data science techniques, such as data analysis, machine learning, and predictive modeling, can play a crucial role in overcoming these challenges and enhancing the PE's services.

In the first part of this report , we are going to state the project overview in which we will be presenting our project’s problems , the solution and the plan . In the next part, we will present the project's business domain, then the analytic approach moving to the functional and non-functional requirements.

Chapter 0: Project Overview

Chapter introduction :

This chapter is dedicated to presenting our project with its problems ,the objectives that we're pursuing and finally the plan that we're going to use.

1. Problematic:

The Employability Pole (PE) of the Esprit Group is responsible for managing the vast number of job and internship opportunities that it receives for its students.

However, the current process of managing these offers is becoming increasingly complex and challenging, given the variety of formats in which they are received, including images, Word documents, PDFs, and emails. The sheer volume of these job and internship offers also creates difficulties in terms of organization and efficient management.

Furthermore, the PE is facing limitations in terms of gaining a comprehensive understanding of the skills required by the job market, as well as the geographical distribution of job and internship opportunities. This lack of visibility has a significant impact on the PE's ability to match students with suitable job and internship opportunities and provide them with comprehensive guidance on their career paths.

2. Solution :

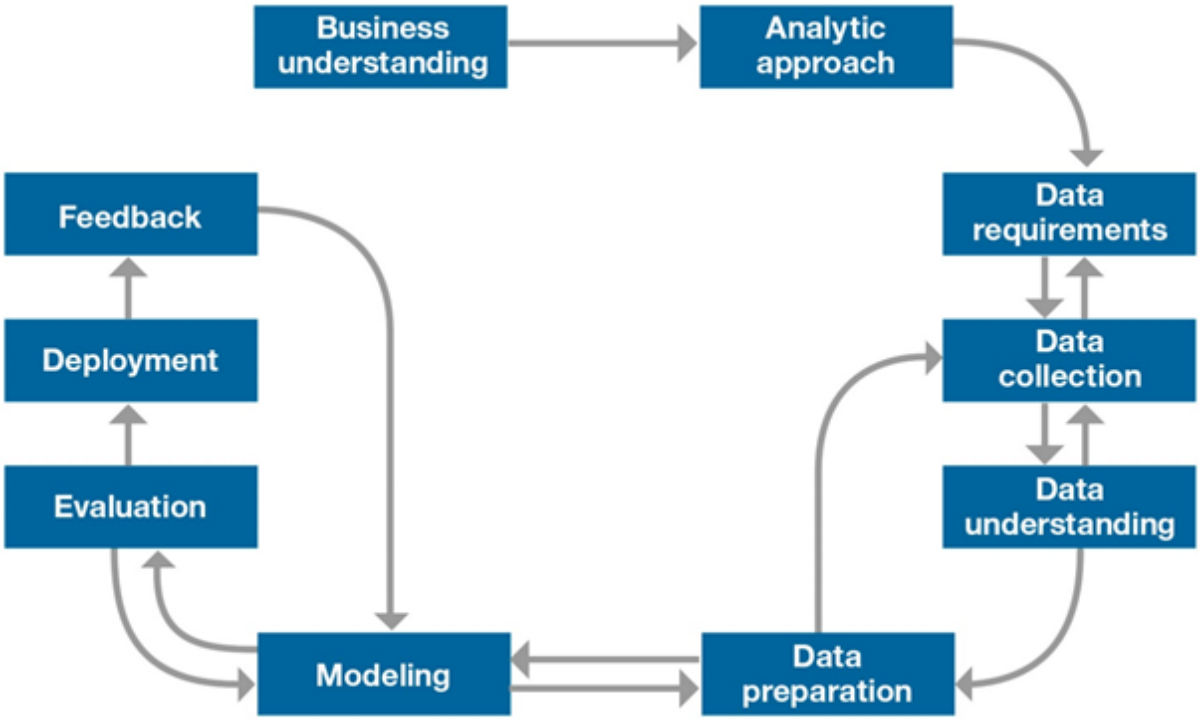
In order to address these challenges, it is essential that the PE implements a comprehensive solution for managing the job and internship offers it receives. This solution should not only allow for efficient management and organization of the offers, but it should also provide the PE with the necessary visibility into the job market, including information on the skills required by various industries and the geographical distribution of job and internship opportunities.

By addressing these challenges, the PE will be better equipped to provide students with valuable guidance on their career paths and match them with suitable job and internship opportunities. This will not only benefit the students, but it will also enhance the reputation of the Esprit Group as a leading provider of education and career development services.

3. IBM Master plan :

A methodology is a general strategy that guides the processes and activities within a given domain. Methodology does not depend on particular technologies or tools, nor is it a set of techniques or recipes. Rather, a methodology provides the data scientist with a framework for how to proceed with whatever methods, processes and heuristics will be used to obtain answers or results.

IBM Master Plan is a methodology based on a problem-solving logic and it's schemed as follows :



Stage 1: Business understanding

Having an understanding for the problem is placed at the beginning of the methodology because getting clarity allows you to determine which data will be used to answer the core question. Two things must be set: The goal and the objectives.

Stage 2: Analytic approach

The analytic approach phase helps limit the algorithm(s) that will be used later. If the question is to determine probabilities of an action, one must use the predictive model. If the question is to show relationships, one must use a descriptive model (clusters). If the question requires a Yes/No answer, one must use a classification model.

Stage 3: Data requirements

The chosen analytic approach determines the data requirements. Specifically, the analytic methods to be used require certain data content, formats and representations, guided by domain knowledge.

Stage 4: Data collection

In the initial data collection stage, data scientists identify and gather the available data resources—structured, unstructured and semi-structured—relevant to the problem domain. If there are gaps in data collection, the data scientist may have to revise the data requirements accordingly and collect new and/or more data.

Stage 5: Data understanding

After the original data collection, data scientists typically use descriptive statistics and visualization techniques to understand the data content, assess data quality and discover initial insights about the data.

Stage 6: Data preparation

Data preparation activities include data cleaning (dealing with missing or invalid values, eliminating duplicates, formatting properly), combining data from multiple sources (files, tables, platforms) and transforming data into more useful variables. This step is usually the most time-consuming step in a data science project.

Stage 7: Modeling

Starting with the first version of the prepared data set, the modeling stage focuses on developing predictive or descriptive models according to the previously defined analytic approach. The modeling process is typically highly iterative to gain insights, leading to refinements in data preparation and model specification. For a given technique, data scientists may try multiple algorithms with their respective parameters to find the best model for the available variables.

Stage 8: Evaluation

Evaluating the accuracy of a model is an essential part of the project. It's the step in which we check if the model we have already generated answers the initial request or not. If the model is a predictive model, a decision tree can be used to evaluate if the modeling output is aligned to the initial design. It can be used to see where there are areas that require adjustments.

Stage 9: Deployment

Once a satisfactory model has been developed and is approved by the business sponsors, it is deployed into the production environment or a comparable test environment. Usually it is deployed in a limited way until its performance has been fully evaluated. Deploying a model into an operational business process usually involves additional groups, skills and technologies from within the enterprise.

Stage 10: Feedback

By collecting results from the implemented model, the organization gets feedback on the model's performance and its impact on the environment in which it was deployed. For example, feedback could take the form of response rates to a promotional campaign targeting a group of customers identified by the model as high-potential responders. Analyzing this feedback enables data scientists to refine the model to improve its accuracy and usefulness.

Conclusion:

This chapter serves as an introduction to outline the problem that the solution aims to solve, the development approach, and the methodology to be followed. The subsequent chapter will delve into the theoretical study required to implement the solution and initiate the first phase of the methodology: business understanding.

Chapter 1: Business Understanding

Chapter introduction :

This chapter provides a detailed overview of the business understanding phase and its key objectives , which is the first step in the IBM Master Methodology to build data-driven solutions.

1-Business domain:

In a more demanding employment market now, more than ever, the need for mentorship and guidance for students and alums to choose their career wisely is a crucial and challenging process for them so they are able to reenter the workforce.

Understanding the wide range of careers available to you can help you choose a professional path that aligns with your skills and interests. If you're deciding on a career path, searching for a job or hoping to improve your professional skills, a career center might help you achieve your goals.

1.1 Definition of a career center:

A career center is an office dedicated to helping people find jobs and develop professional skills. These offices are often part of a school, college or nonprofit organization, but might also be an independent business or government agency. In career centers, trained career development professionals can help you choose a career path, identifying and working toward career goals, finding suitable careers for graduate school programs ,creating a resume and learning skills that employers might prefer in candidates.

1.2 Types of career centers

Different career centers serve people of different ages and experience levels. Here are some types of career centers:

1.2.1 High school career centers

High schools often have college and career centers, which may be independent offices or part of the guidance department. The counselors in charge of these offices might meet with students about different educational paths, coordinate internship programs with local organizations and host college and job fairs. The college and career center might also be the place students can go for information on military service or vocational training.

1.2.2 College and university career centers

College and university career centers help students build professional skills and prepare to enter the job market or graduate education. Career counselors can help students choose a major and minor based on their career goals and advise them on when to take certain courses. While most college and career centers are free to use for current students, some provide services for alumni and employees as well.

1.2.3 Government career centers

The federal government manages many career centers across the nation for job seekers of all ages and experience levels. Staffed by government employees, these centers might offer career guidance, skills workshops and other resources for job seekers. Depending on their location and focus, they may also offer unemployment benefits and services for veterans.

1.2.4 Nonprofit career centers

Some community organizations have career centers for job seekers in their area, staffed by volunteers or paid counselors. These centers might have part-time hours or move from one location to another so they can serve a wide range of people in the community. For example, a local library might host career center services every other Saturday, where job seekers can make an appointment or walk in for counseling.

1.2.5 Private career centers

Private career centers offer counseling and training for a fee and may also provide referrals to tutoring companies or other individuals to help students succeed in school. These firms might offer both virtual and in-person services and often charge by the hour. They can help you find a job, change careers or develop skills to advance in your organization.

1.3 National statistics:

This figure shows that by the first year the career center platform has hit more than 6000 of users that are registered with 73% of them willing to help .

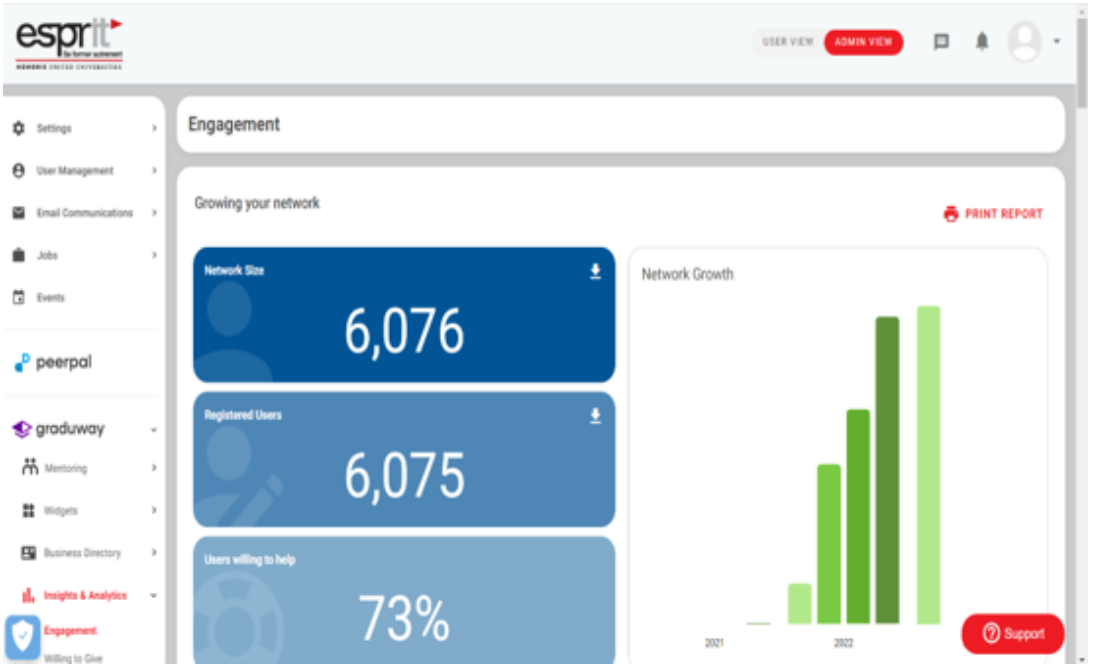


Figure 1: Network Growth.

Here in this figure we can see the big success that this platform has created in a short time within the integration and the helping of students and alums in the employment field to make the job offers very easy to access with plus than 2500 of job applications ..

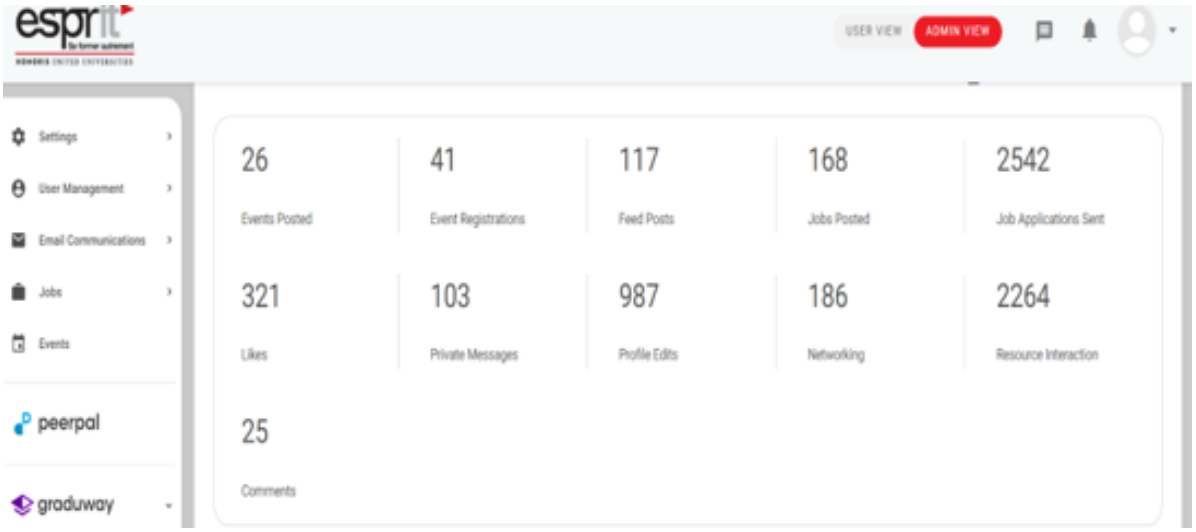


Figure 2 Amount of job applications

In this figure we can have a look at the top companies providing job offers for example NeoXam Tunisia provided 20 job offers furthermore in this statistical chart we can see the number of offers posted every month .

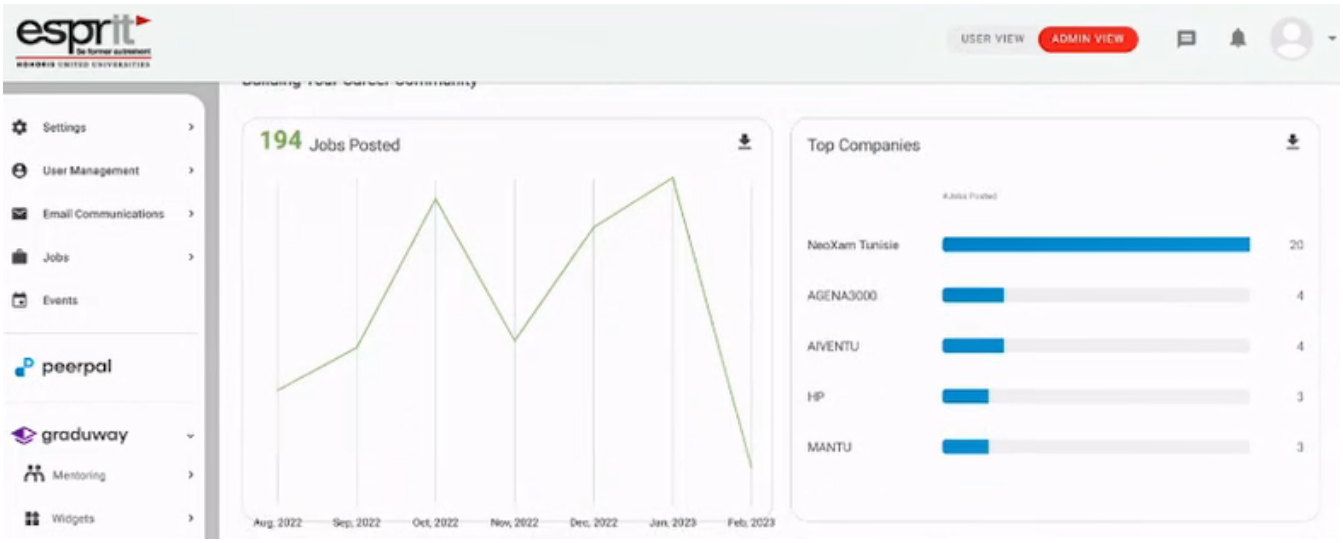


Figure 3 Top companies and Jobs posted

Figure 4 Geographical distribution of students and alumni

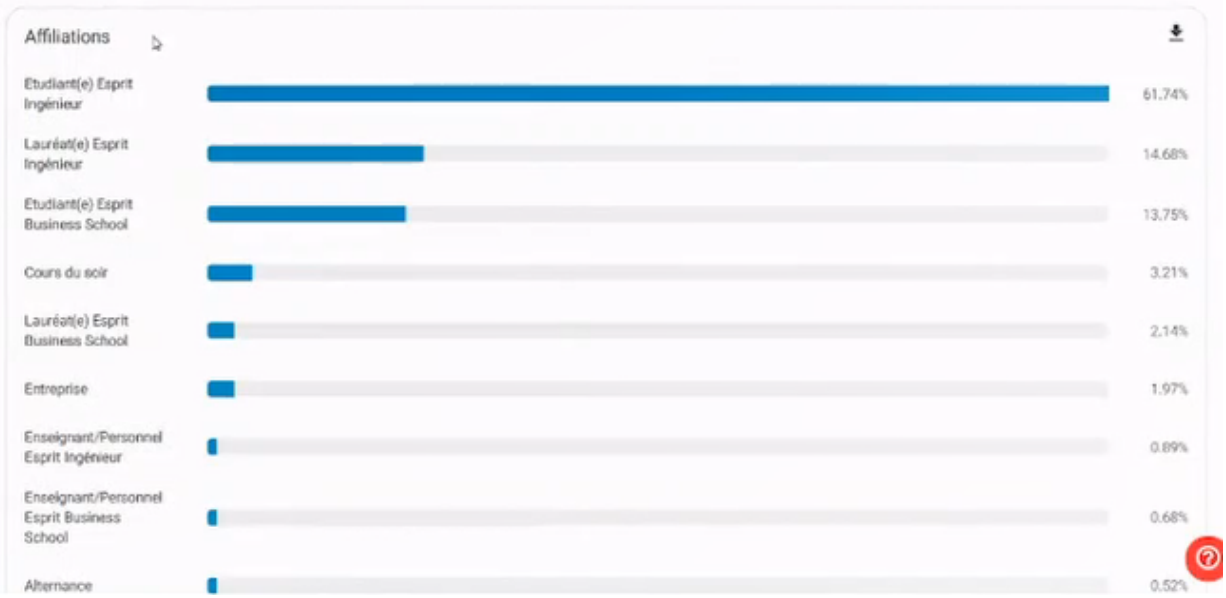


Figure 5 Number of students and alumni

Conclusion:

The business understanding phase is critical to the success of the project because it ensures that the project team has a clear understanding of the business needs and objectives, and can align the solution with those needs.

Chapter 2: Analytic approach

Chapter introduction :

The analytic approach phase involves the use of various statistical and machine learning techniques to analyze and interpret the collected data and generate insights that can inform decision-making in the subsequent phases of the IBM Master Methodology.

1. Define the problem and objectives:

In this phase, the project team would define the problem they are trying to solve (e.g., inefficient job matching), and establish clear objectives for the project (e.g., increasing the number of successful job placements).

2. Gathering data:

The project team would gather data on the platform's current performance, user behavior, and job market trends. This might involve collecting data on user profiles, job listings, job applications, and job placement rates.

3. Analyze the data:

Using statistical analysis and data visualization techniques, the project team would analyze the data to gain insights into user behavior, job market trends, and platform performance. They might identify patterns in the data that indicate which types of candidates are most successful in finding jobs, or which job listings are most popular with users.

4. Develop hypotheses:

Based on the insights gained from the data analysis, the project team would develop hypotheses about which factors are driving successful job matching. These hypotheses might involve user behavior, platform functionality, or other factors that could be affecting the platform's performance.

5. Test the hypotheses:

The project team would then test these hypotheses through experiments or interventions. This might involve making changes to the platform, such as adding new features or improving existing ones, or conducting A/B testing to compare the effectiveness of different job matching algorithms.

6. Evaluate the results :

Once the tests have been conducted, the project team would evaluate the results to determine whether the changes have had a positive impact on the platform's performance. This might involve analyzing user engagement metrics, job placement rates, or other KPIs.

7. Refine the approach :

Based on the results of the evaluation, the project team would refine the approach and iterate the process to continue improving the platform over time. This might involve making further changes or interventions, or developing new hypotheses to test.

Conclusion :

This analytic approach would help the project team to continuously refine and improve the platform's job matching capabilities, ultimately helping to ensure that the best candidate is matched with the best job offer, and vice versa.

Chapter 3: Data Requirements

Chapter introduction :

This chapter will present the data science requirements which includes functional requirements where a function is described as a specification of behavior between inputs and outputs, and nonfunctional requirements being a set of specifications that describe the system's operation capabilities and constraints to improve its functionality .

1. Functional requirements :

1.1 Business Objectives:

1.1.1-Job Offer Management:

To develop a solution that can efficiently and effectively manage the large volume of job and internship offers received in different formats such as images, pdf, word, emails, etc.

1.1.2-Data Extraction:

To extract valuable insights and trends from the raw data that can provide a better understanding of the skills required by the job market and the geographical distribution of the job offers.

1.1.3-Candidate Matching and Recommendation:

To develop a system that can match candidates with suitable job offers with a certain percentage based on their skills and qualifications. The solution should be capable of providing personalized recommendations to each candidate.

1.1.4-Real-Time Data Analytics and Reporting:

To provide real-time data analytics and reporting capabilities based on key performance indicators (KPIs) such as the number of offers per period (month, quarter, year), the

number of offers by technologies, trades or fields, and the geographical distribution of the job offers nationally and internationally.

1.2 Data Science Objectives :

1.2.1-Descriptive Analytics:

In order to provide a comprehensive overview of the job and internship offers, descriptive analytics will play a crucial role in your project.

1.2.2-Web Scraping:

By using web scraping techniques, you can collect relevant CVs from LinkedIn and other professional networking platforms to have a better understanding of the available pool of candidates.

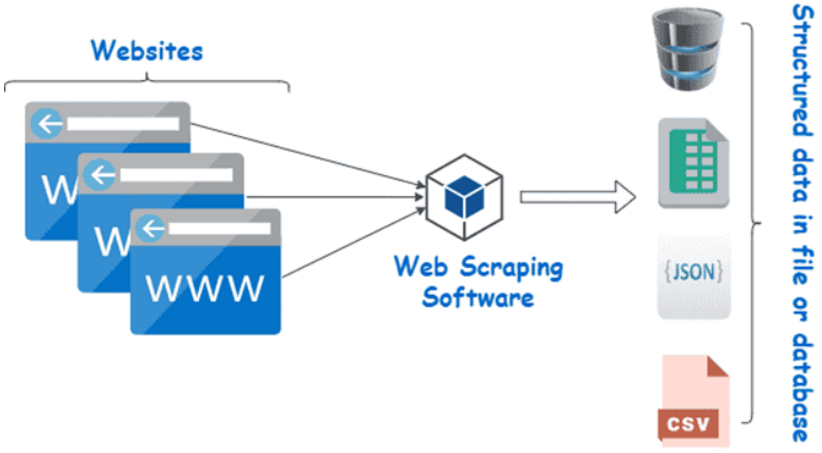


figure 6**:web scraping figure**

1.2.3-OCR and NLP:

To extract and summarize information from PDF documents, Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques can be used to process the unstructured data into a format that can be analyzed.

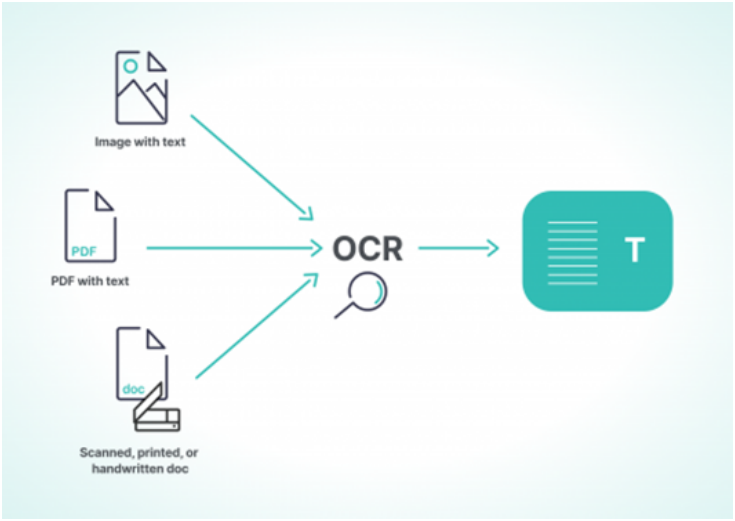


figure 8: Optical character recognition

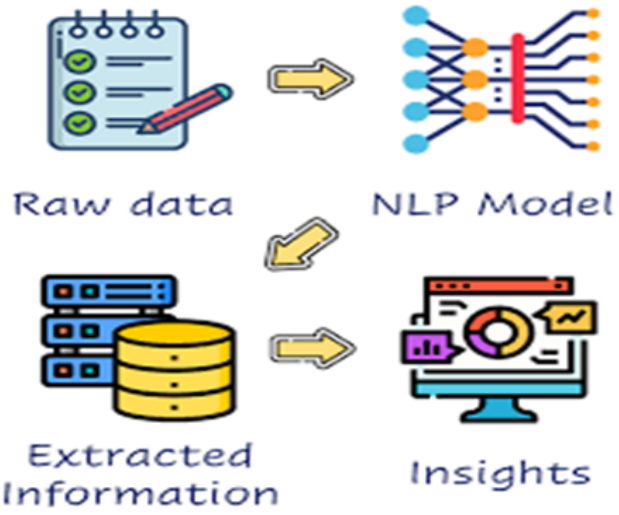


figure 7: Natural Language Processing

1.2.4-Clustering:

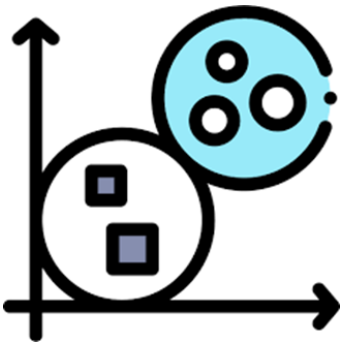


figure 9: **Clustering figure**

Clustering algorithms can be used to categorize the job and internship offers based on their attributes and help understand the distribution of job offers across different industries and fields.

1.2.5-Scatter Plots:

Visualizing the data using scatter plots can provide insights into the geographical distribution of job offers, and help identify areas with a high concentration of job opportunities.

1.2.6-Predictive Analytics:

By using predictive analytics techniques such as regression, you can forecast the number of job offers in the future, and understand the fields that are in high demand in different regions.

1.2.7-Optimization Algorithms:

By utilizing optimization algorithms, you can match job offers with suitable candidates in an efficient manner, based on their skills and qualifications, to help the PE of the Esprit group find the right candidate within a short timeline.

2. Non Functional requirements :

Non-functional requirements or NFRs are a set of specifications that describe the system's operation capabilities and constraints and attempt to improve its functionality.

Moreover they are the constraints or the requirements imposed on the system. They specify the quality attribute of the software and deal with issues. They address vital issues of quality for software systems. These requirements are important because they can have a significant impact on the overall quality and success of a software system. They are often closely related to the system's performance, security, and usability. They also help to ensure that the system is maintainable, portable, and compliant with relevant laws and regulations.

In this project, these are the NFR we will work on:

2.1 Performance and scalability:

Performance refers to the speed and effectiveness of a system under a given workload within a given time frame. In general terms, it measures how well software can perform its intended task. Achieving an acceptable level of performance is mandatory when developing software.While Scalability refers to the characteristic of a system to increase performance by adding additional resources.Both are two of the key aspects of software design.So we can say that The Platform should be able to handle a large volume of data and users without experiencing significant lag or downtime.

2.2 Security:

Security is a non-functional requirement assuring all data inside the system or its part will be protected against malware attacks or unauthorized access ,examples could be authentication, authorization, backup, server-clustering, etc. This requirement artifact can

be derived from best practices, policies, and regulations.The solution should protect the data from unauthorized access and protect the system from potential cyber threats.

2.3 Integration with other systems:

System integration is the process of uniting all virtual and physical components into a single cohesive infrastructure to ensure that all the individual pieces of an organization work as a whole.

The Platform should be able to integrate with other systems and platforms used by the Esprit group, such as student information systems, recruitment systems, and performance management systems.

2.4 Customization and configurability

Customization enhances the software's capabilities so it can offer a tailored user experience. Unlike configuration, customization requires coding, which results in fundamental changes to your baseline system.

It is wise to discuss any desired customizations with your software consultant first. He/she will review your business processes and help you decide which customizations will ultimately be helpful, and which ones may hinder you more than they help you.That's why The Platform should be customizable and configurable to meet the specific needs of the Esprit group and its stakeholders.

2.5 Continuous improvement and maintenance:

The Platform should be designed with the ability to continuously improve and evolve over time based on feedback and new data. It should also be designed with the ability to maintain and update the solution as needed.

Conclusion :

Both functional and non-functional requirements are critical to the success of the project and need to be carefully identified and documented in the early stages of the IBM Master Methodology to ensure that we can design and build a solution that meets the business needs and user requirements.

Chapter 4: Data Collection

Chapter introduction :

In this chapter we discuss the various sources of data required for this project and how we managed it and stored it .

1. Student data Collection from linkedin :

In this part we used web scraping techniques and implemented them in a python script along with selenium library and web driver to get access to the linkedin profiles and to scrape all important data such as the name , job title, company , location ,skills and lastly the experience to store them later in a data set .

Name	job_title	company	location	URL:	skills:	experiences:
omar mhiri	Web and mobile developer	Freelance | Self-Employ	Gouvernorat Tunis, Tunisie	https://tn.linkedin.com/in/om	HTML5, PHP, Shell Scripti	Web Designer, Web and mobile developer, Web and Mobile Developer, Web Develo
Oussema ARC	Fullstack Mobile Developer | | Freelancer	Orange Tunisie	Tunis, Gouvernorat Tunis, T	https://tn.linkedin.com/in/ous	No results	
Mariem Ben F	Mobile Developer (Android - Flutter)	StreamWIDE	Gouvernorat Ben Arous, Tur	https://tn.linkedin.com/in/ma	Mobile development, Applications mobiles, Scrum, user experience, Java, XML, Eclipse, SQL, Unix, Android, U	
Benabdessale	Senior iOS Developer Engineer at eXo plat	eXo Platform	Gouvernorat Tunis, Tunisie	https://tn.linkedin.com/in/ber	Swift, Xcode, Gestion de p	Formateur, Stage Ouvrier
Ghassen Ayec	iOS Mobile Developer Engineer	Avempace	Gouvernorat Nabeul, Tunisi	https://tn.linkedin.com/in/gha	Xcode, Jira, Subversion, B	END OF STUDIES TRAINING, TECHNICIAN INTERNSHIP, SALES COLLABORATOR
Yasmine Baga	Dev Web et Mobile chez Educanet Tunisie	Educanet Tunisia	Gouvernorat Sousse, Tunisi	https://tn.linkedin.com/in/yas	Web Development, Computer Science, Angular, Scrum, SQL, User Requirements, Interfaces, Mobile Applicatic	
Anis BEJAOUI	Flutter advocate | Mobile & Software Engi	bulldozer	Grand Tunis Metropolitan A	https://tn.linkedin.com/in/ani	rxDart, Test Driven Develo	Mobile Developer, iOS Engineer, Flutter developer, Web Developer
Oussama HEN	Web and Mobile Developer , Flutter/React	Manitoo	Gouvernorat Ariana, Tunisie	https://tn.linkedin.com/in/ous	Angular, Scrum, TypeScrip	Full Stack Engineer
Mohamed BE	Web and Mobile Developer chez Continuc	Continuous Net	Akouda, Gouvernorat Souss	https://tn.linkedin.com/in/mo	Framework Symfony, Wo	Web developer, Web developer

figure 10**: Student Dataset from LinkedIn**

2. Job data Collection from Emails :

In this part we used IMAP (internet Message access protocol) which allows us to connect to the Gmail server and retrieve email . After we fetched all the desired emails we stripped them from the unwanted details and we applied NLP (natural language processing) to gather the important data such as the subject , the skills ,the type of the offer , where this email came from and finally the link .

Out[652]:

	type	from	sujet	skills	link
0	job	<pole-employabilite-esprit@esprit.tn>	Offres d'emploi-Banque de Tunisie	[net, core, entity, framework, vmc, javascript]	[https://docsgooglecom/forms/d/e/, https://esp...
1	Unknown	<pole-employabilite-esprit@esprit.tn>	erreur	[]	[https://wwwcofomocom/f, https://docsgooglecom...
2	job	<pole-employabilite-esprit@esprit.tn>	Offres d'emploi-Groupe Lesaffre-Data Scientist...	[htmlapplication, pdfapplication, pdf]	[https://docsgooglecom/forms/d/e/, https://esp...
3	job	<pole-employabilite-esprit@esprit.tn>	Offres d'emploi-ODDO BHF	[recruited, developers, dotnet, engineers, dat...	[https://docsgooglecom/forms/d/e/, https://esp...
4	Unknown	<pole-employabilite-esprit@esprit.tn>	Hewlett Packard Enterprise recruits a graduate...	[plainfor, encourage, line, manager, basically...	[https://espritconnectcom/, https://careershpe...
...	...	...	...	...	...
147	Unknown	<pole-employabilite-esprit@esprit.tn>	erreur	[plainfor, forwarded, message, bousbia, mar, f...	[https://formsgle/, https://formsgle/, https:/...
148	Intership	<pole-employabilite-	Virtual Internships	[minute, ce, formu, lar, jfe,	[https:///jfe/form/, https://wwwvirtualinterns...

figure 11**: Job Offer dataset from Emails**

3. Job data Collection from Esprit connect :

For this part we used beautiful soup in a python script which is a library made for web scraping to help extract important data from HTML or XML pages using the tags (<p>, ,<a>,.....) .In addition to that we used Chrome web driver to help get access and scroll between the pages to extract the wanted data .

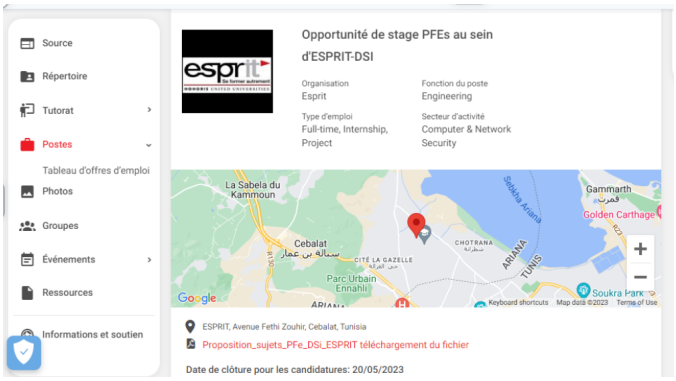


figure 13**: the data from esprit connect connect**

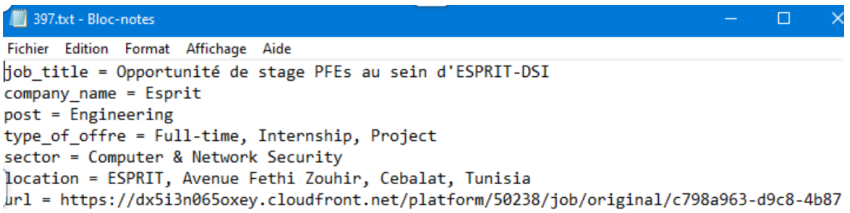


figure 12**: the extracted data from esprit**

4. Job data Collection from PDF :

We first checked if the PDF is scanned by using the is_scanned_pdf() function. If the PDF is scanned, it converts the PDF file to images using convert_pdf() and then extracts the text from each image using the ocr_extract_txt() function where we used the Pytesseract library to perform the OCR (Optical Character Recognition) . The text from all the images is concatenated to form the final text output.

If the PDF is not scanned, it directly extracts the text from the PDF using the extract_text_from_pdf() function.

Then we used the OpenAI API to create a chatbot that can extract job offers and their required skills from a given text. It uses a list of texts and iterates through each text to extract the relevant information using the chatbot. The resulting information is stored in a list of JSON objects with three columns: job title, required skill, and location

After that, the list of JSON responses obtained from OpenAI API is converted into a Pandas dataframe. Each JSON object in the list contains information about job titles, required skills,

and location. A new column 'link' is added to the dataframe that contains the file path of the original PDF from which the job information was extracted. Finally, all data frames are concatenated into a single dataframe.

job_title	skill_required	location	link
DevSecOps	software engineering, data development solutio	SANTANDERTECHHUB-PROFILES	2022-11-11 Job description junior positions-page0.pdf
Technology Talent	Broad spectrum technologies (including Blockch	Poland, Portugal, Spain, UK, Mexico, Bra	2022-11-11 Job description junior positions-page1.pdf
DevSecOps Engineer	Collaboration, architecture, design, deployment	SANTANDERTECHHUB	2022-11-11 Job description junior positions-page2.pdf
New technologies specialist	Kafka, Java8 (JDK11/17), Spring Cloud Stream, S	Not specified	2022-11-11 Job description junior positions-page4.pdf
Data Management Solution Delivery	Cloudera/Spark/SnowFlake/Databricks/Stratio/I	SantanderTechHub	2022-11-11 Job description junior positions-page5.pdf
Data Pipeline Developer	Bachelor/Degree/Master in STEM (Science, Tech	SantanderTechHub	2022-11-11 Job description junior positions-page5.pdf
Cybersecurity Analyst	Internal and external vulnerability management,	Santander Tech Hub	2022-11-11 Job description junior positions-page6.pdf
Required qualifications	Bachelor's/Master's degree in Science, Technolo	Santander Tech Hub	2022-11-11 Job description junior positions-page6.pdf
Technology Specialist	In-depth knowledge technology, cybersecurity sc	Santander Tech Hub	2022-11-11 Job description junior positions-page7.pdf
Execution Management Project Analyst	Consulting, Project Management, Business Analy	Málaga, Spain	2022-page0.pdf
Execution Management Project Analyst	Project management skills, Financial knowledge,	Malaga	2022-page2.pdf
Analista BMC-SBGM Consulting	Operational management skills, Consulting skills	Malaga	2022-page2.pdf

figure 14**: Job Offer dataset from PDF**

5. Data Storage:

To store the collected Data, we used MongoDB which is a document-oriented NoSQL database that offers several benefits for data storage in a project. Some of these benefits include:

- Flexible data model
- High availability and easy scaling
- Fast indexing and efficient search
- Easy integration with data science tools
- Enhanced security

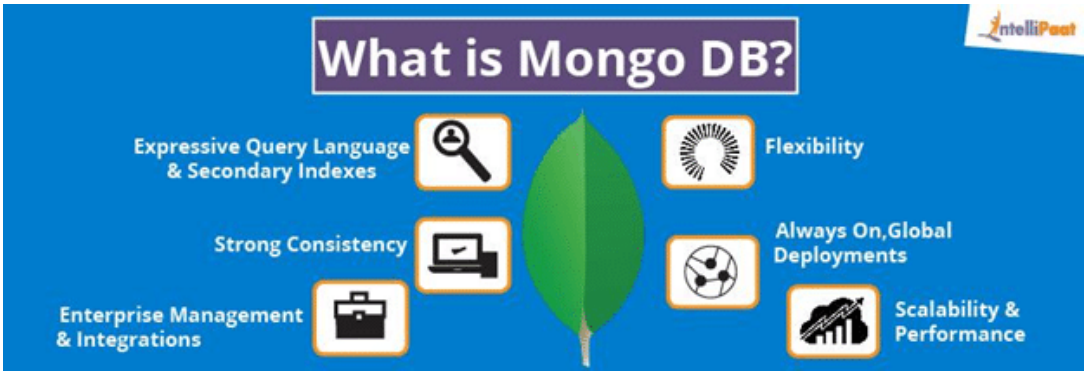


figure 15 **: What is mongo DB**

In MongoDB, data is stored in flexible, JSON-like documents with dynamic schemas. These documents are stored in collections, which are similar to tables in a traditional database. MongoDB uses a binary format called BSON (Binary JSON) to store the documents, which allows for efficient data processing.

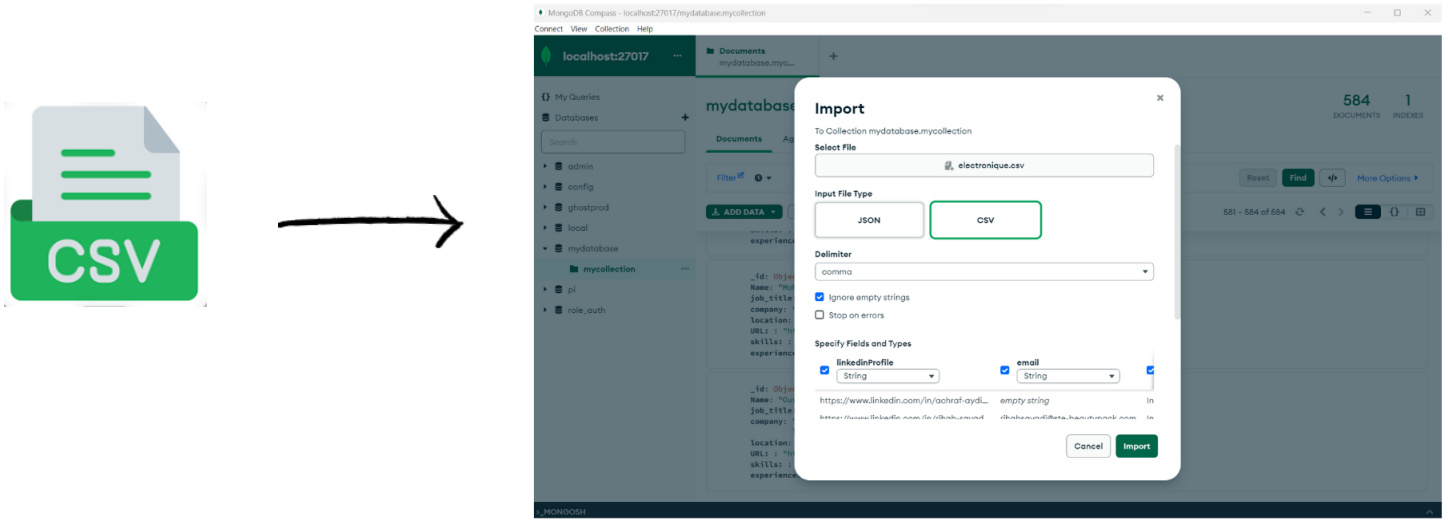


figure 16: Importing data into mongo

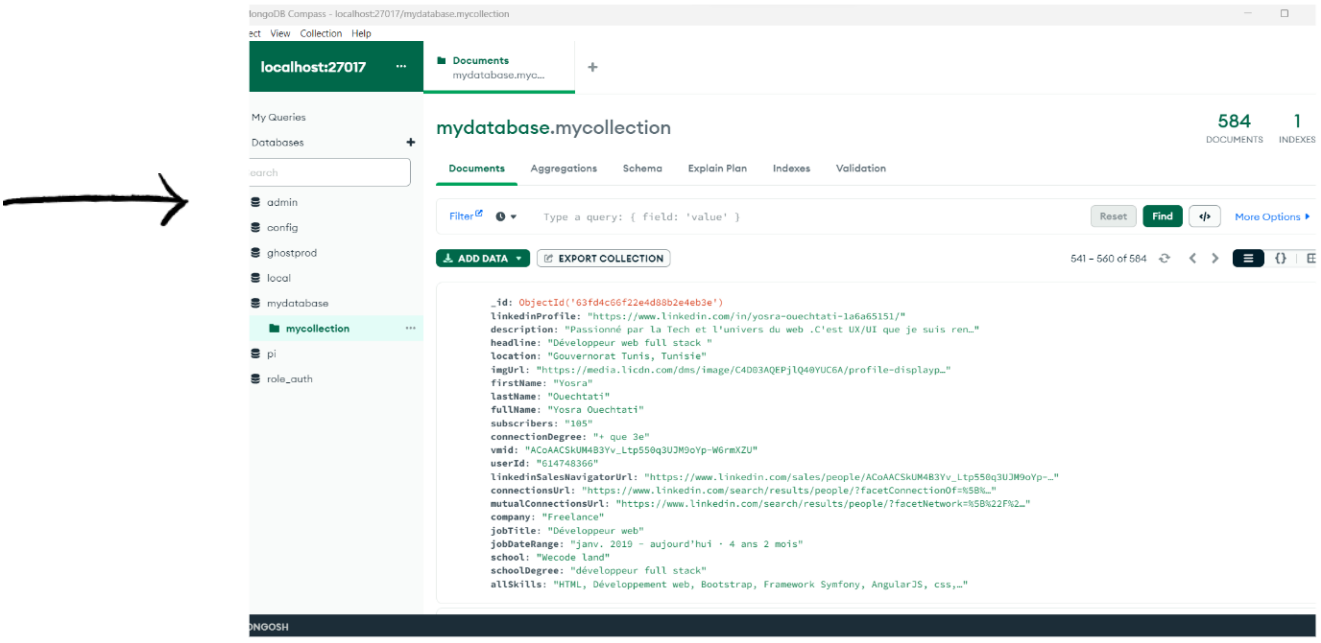


figure 17 **: Example of imported data**

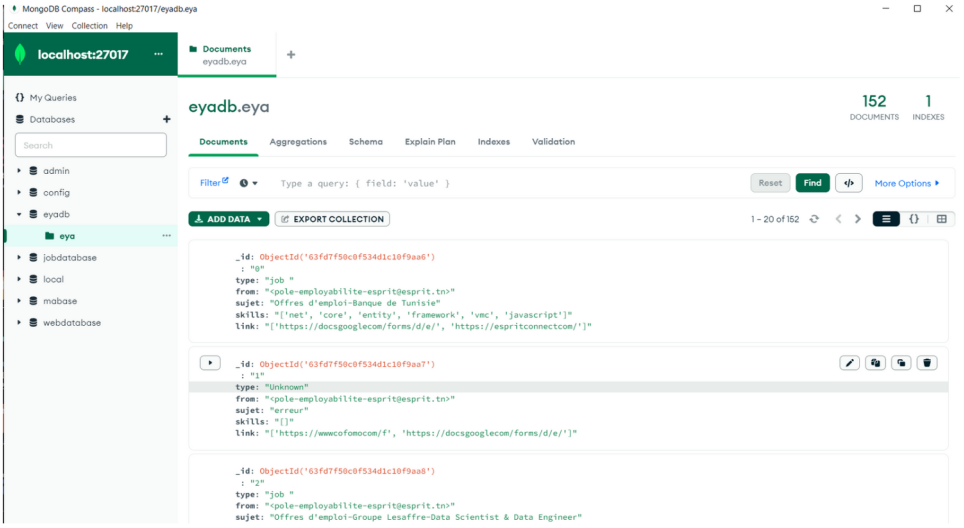


figure 18**: Job Offer Database**

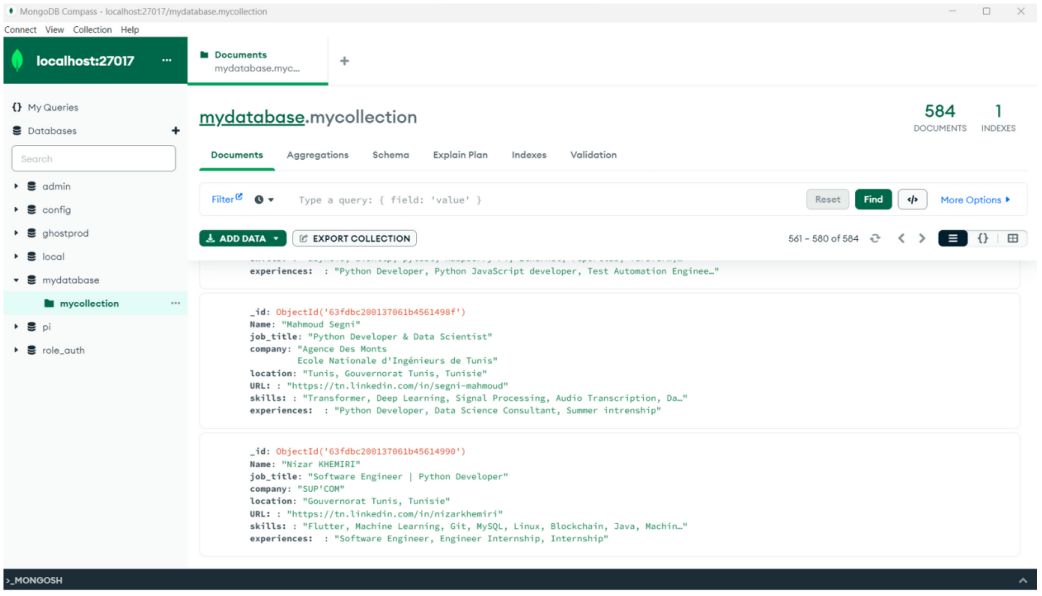


figure 19**: Students Database**

Conclusion :

Collecting and managing data from various sources, can uncover various perspectives, opinions, and insights into the problem. This can help to ensure that the project's results are reliable, accurate, and relevant to the business problem being addressed.

Chapter 5: Data Understanding

Chapter introduction :

In this chapter we will get to know the features of every dataset and their meaning .

1. Student data from linkedIn:

Name: The name of the candidate

Job title: The job title or current position of the candidate

company: The name of the company or school the candidate has worked for or attended

Location: The candidate's location

URL: The profile URL is a unique identifier that links directly to the candidate's online profile.

Skills: The skills of the candidate

Experiences: career history, level of seniority, industry experience, and areas of expertise.

2. Job data from Emails:

Type : the type of position offered(PFE,internship..)

From:the email address of the sender

Subject:the subject of the offer

Skills: the skills required for the offer

Links:The link of the job offer on Esprit connect

3. Job data from Esprit Connect :

Job_title: the title of the offer for exemple :Opportunité de stage PFEs au sein d'ESPRIT-DSI.

company_name : the name of the company presenting the offer for exemple NeoXam Tunisia ,Esprit.

Type_offre: the type of the offre presented for exemple full time internship contract

Post: the post required for the offer for example Engineering.

sector :the global sector that the offer is going to be included in for exemple computer and network security or Information Technology and Services.

Location :the place that the offer is going to be in for exemple tunisia,,france,spain,

url : the link of the attachment posted with the job offer.

4. Job data from Pdf:

Job_title: the title of the offer presented by the company .

Skills_required: the skills required for the offer.

Location :the place that the offer is going to be in .

Link: the position of the offer in the pdf for example :value-page7.

Conclusion :

Data understanding is an important step in the data analysis process that involves examining and exploring the data to gain a better understanding of its characteristics, structure and quality .

Chapter 6: Data preparation

Chapter introduction :

In this chapter, we will prepare our datasets to use them in the modeling phase later on using several techniques that we will be discussing in this chapter.

1. Data preparation for Student data:

Our data contains 464 rows and 7 columns being the name , job_title , company , location ,skills and experience, also we may say that all of our data is 100% categorical .

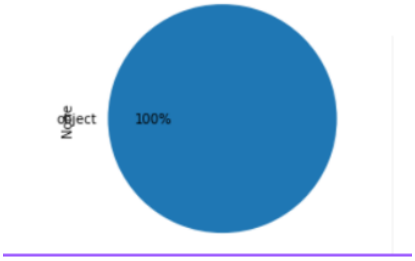


figure 21**: Type of Student data**

```
Entrée [395]: df.shape
Out[395]: (464, 7)
```

figure 20**: Shape of student data**

1.1 Data visualization:

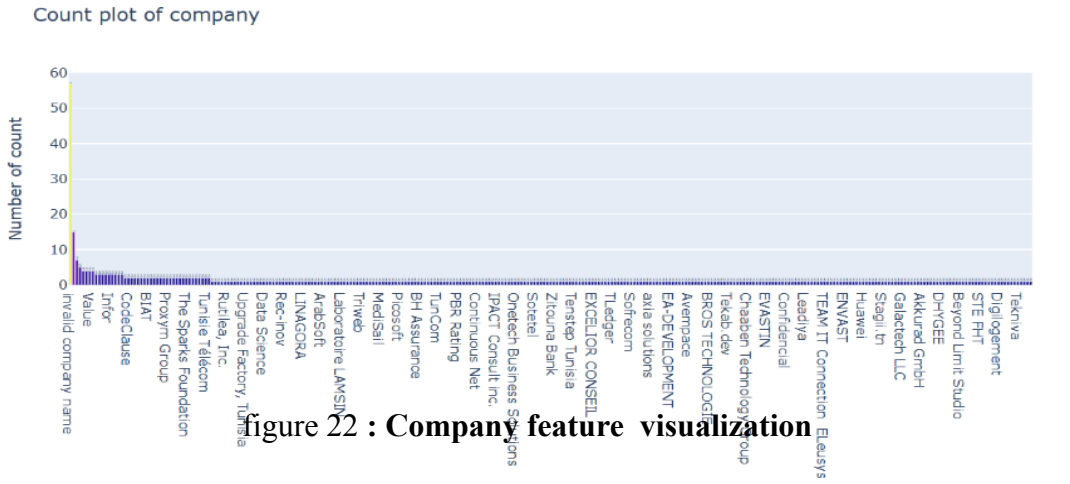


figure 22 **: Company feature visualization**

figure 23 **: Job title feature visualization**

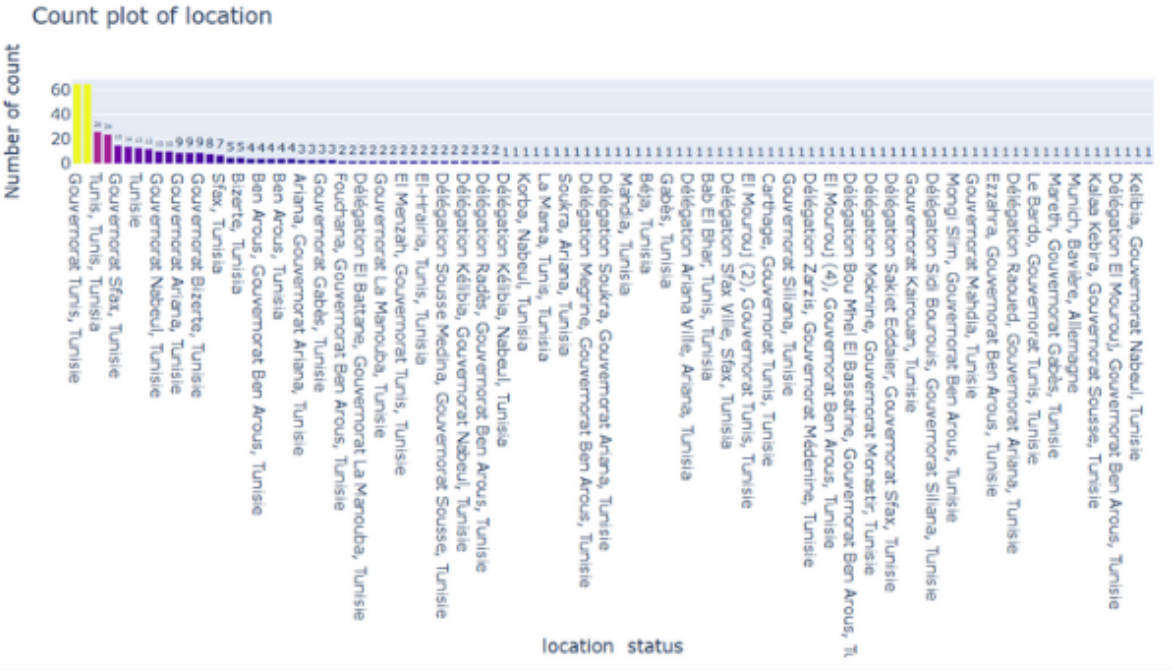


figure 24 **: Location feature visualization**

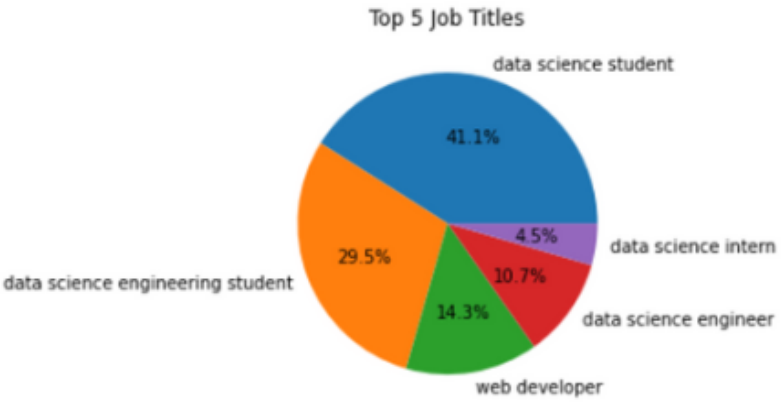


figure 26**: Job title before grouping**

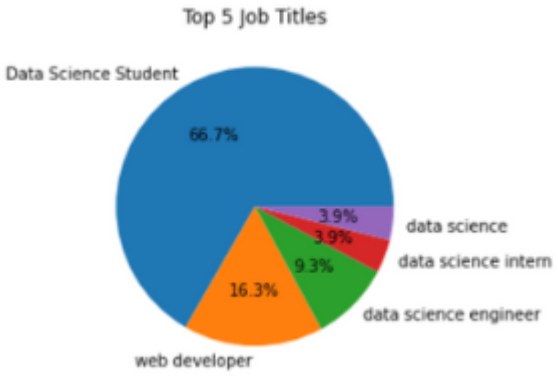


figure 25**: Job title after grouping**

Here for these two pie plots we can see that there is a similarity between data science student with 41.1% and data science engineering students with 29.5% so we decided to combine them since they're the same as Data science student with 66.7% .

1.2 Missing values :

As we can see in the figure we have 109 missing values distributed is all of the data that need to be handled .

```
Entrée [593]: df.isnull().sum()

   Out[593]: URL             1
             job_title       2
             location        1
             Name            1
             company        57
             skills         47
             experiences     0
             dtype: int64

Entrée [594]: str(df.isna().sum().sum())#total

   Out[594]: '109'
```

figure 27 **: Missing values before data prep in Students data**

Starting with the feature Company that contains 57 missing values we replaced all the missing values with "invalid company name " since there is none .

For the feature URL there is only one missing value which is a row that is completely full with missing values so the one missing value of the feature Location and name will go away since we decided to drop that row .

Moving forward to the job_title column we converted all the letters to lowercase and we removed all the punctuation characters in between ,these are some NLP techniques , also we dropped the only missing value that is left in there.

And lastly for the skills column we applied NLP techniques

1.2.1 NLP Cleaning text :

In this part, we have eliminated the special characters, the html tags, the white spaces... Cleaning text is an important step in the NLP process, as it helps to remove noise and irrelevant information from the text, and ensures that the data is in a consistent and standardized format. Here are some common techniques for cleaning text that we used for NLP:

- ☐ Lowercasing: we converted all text to lowercase to avoid issues with case sensitivity.
- ☐ Removing punctuation and special characters: we removed any characters that are not letters or numbers, such as commas, periods, and hashtags.
- ☐ Spell checking and correction: we checked the text for spelling errors and corrected them to improve accuracy.

1.2.2 NLP Removing stop words:

In this part, we removed common words that are unlikely to be useful for analysis, such as "the", "a", and "and".

1.2.3 NLP Word cloud:

A word cloud is a simple yet powerful visual representation object for text processing, which shows the most frequent word with bigger and bolder letters, and with different colors. The smaller the size of the word the lesser it is.

figure 28 **: Word Cloud**

Finally after all the applied techniques we finally have a dataset free of missing values .

```
Entrée [286]: df.isnull().sum()

   Out[286]: URL            0
             job_title      0
             location       0
             Name           0
             company        0
             skills         0
             experiences    0
             dtype: int64

Entrée [288]: str(df.isna().sum().sum())#total

   Out[288]: '0'
```

figure 29 **: Missing values after data prep in students data**

After the data preparation and all the used techniques we finally have a dataset that contains 434 rows and 5 columns.

t[1098]:

	URL	job_title	Name	skills	experiences
0	https://www.linkedin.com/in/talel-kb/	web developer software engineering student	Talel Kbaier	analyse donnees competences analytiques framew...	Développeur web:janv. 2023 - aujourd'hui · 2 m...
1	https://www.linkedin.com/in/omar-talbi-sfax/	freelance web developer	Omar Talbi	laravel react javascript php microsoft bot fra...	Freelance Web Developer:août 2019 - aujourd'hu...
2	https://www.linkedin.com/in/hassen-knani-21991...	web developer chez zetabox	Hassen Knani	management service client microsoft office ven...	React Js Instructor:mars 2022 - aujourd'hui · ...
3	https://www.linkedin.com/in/safwendammak/	freelance web developer	Safwen Dammak	angular phpsymfony developpement web developpe...	Web Developer:oct. 2020 - févr. 2022 · 1 an 5 ...
4	https://www.linkedin.com/in/oumayma-b%C3%A9hi-...	web mobile developer 💻	Oumayma Béhi	mentioned	Projet de fin d'études:févr. 2022 - juin 2022 ...
...	...	...	...	...	...
459	https://www.linkedin.com/in/khaoula-ben-othman...	data science engineer	khaoula Ben othman	mongodb rstudio business intelligence bi pyspa...	nan;nan,nan;nan
460	https://www.linkedin.com/in/farah-abid-989683191/	data science enthusiast	Farah Abid	analyse donnees nlp extraction donnees python ...	Data Science Intern:Oct 2022 - Present · 5 mos...
461	https://www.linkedin.com/in/achref-cherif-data...	data scientist founder ceo of data science t...	Achref Cherif	extract transform load etl analytical skills d...	Data Scientist:Sep 2020 - Present · 2 yrs 6 mo...
462	https://www.linkedin.com/in/hamza-samaiy/	data science master student works at vermeg	Hamza Samaiy	developpement web javascript php laravel feuil...	Software Programmer:Jan 2023 - Present · 2 mos...
463	https://www.linkedin.com/in/zitouni-amal-286b9...	ingénieur en informatique spécialité data sci...	Zitouni Amal	microsoft azure machine learning apprentissage...	Ingénieur BI:Oct 2022 - Present · 5 mos,Member...

434 rows × 5 columns

figure 30**: Student dataset after data preparation**

2. Data preparation for Jobs data:

Our data has 1598 rows and 5 columns which are our features ; job_title, skill_required, location and link. Also all of our data is 100% categorical

```
data.shape
```

```
(1593, 5)
```

figure 32 Shape of job data

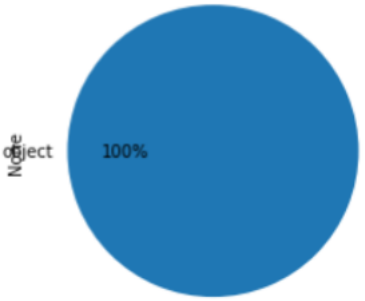


figure 31: Type of Job data

2.1 Data visualization:

figure 33: **job_title feature visualization**

figure 34 **skill_required feature visualization**

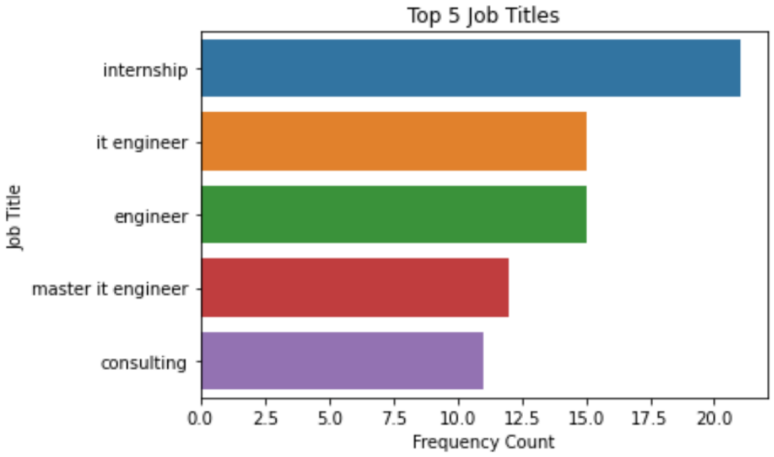


figure 35**: the top 5 job titles**

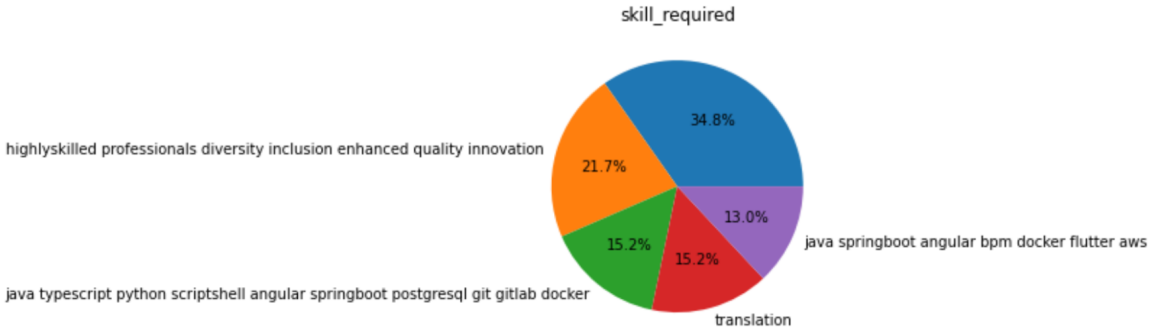


figure 36**: the percentage of skill_required column**

2.2 Translation:

When working with data from various sources, it's common to encounter data that is in different languages. However, to perform analysis on this data, we need to have it in a common language. In the case of our project, we needed to translate the data into English to facilitate analysis.

To achieve this, we used the Googletrans library, which is a Python package that uses the Google Translate API to translate text between languages. This library makes it easy to translate text from one language to another programmatically.

Using Googletrans, we were able to write Python code to automatically translate text from different languages to English. This allowed us to extract insights from the data regardless of the language it was originally written in.

Overall, using Googletrans to translate the data into English was a key step in our data processing pipeline and enabled us to move forward with the analysis phase of our project.

Développement outil monitoring risques	implémentation chatbot intelligent, outil préd...
utilisant différentes données textuelles colle...	Les techniques de webscrapping ou/et parsing d...
Réflexion stratégique cadre conception platefo...	NaN
Evaluation bilan carbone d'une entreprise	méthodologie, questionnaire (questions clés)

figure 37**: Data before translation**

```
                                                job_title  \
                                                DevSecOps
                                          Technology Talent
                                          DevSecOps Engineer
                                    New technologies specialist
                              Data Management Solution Delivery
                                                       ...
                                    Analysis of application logs
                                      Build a reference database
                             Development of risk monitoring tool
                          using different collected textual data
                                  Carbon assessment of a company

                                                skill_required
                          software engineering, data development solutio...
                          Broad spectrum technologies (including Blockch...
                          Collaboration, architecture, design, deploymen...
                          Kafka, Java8 (JDK11/17), Spring Cloud Stream, ...
                          Cloudera/Spark/SnowFlake/Databricks/Stratio/Hi...
                                                                       ...
                          to identify operational anomalies, performance...
                          on the IS, document and produce the common dat...
                          intelligent chatbot implementation, predictive...
                          Techniques for webscrapping and/or parsing pdf...
                                 methodology, questionnaire (key questions)
```

figure 38 **Data after translation**

2.3 Missing values:

In the job offer dataset we have 630 missing values that need to be handled .

```
#nb de valeurs manquantes
str(df.isna().sum().sum())#total
```

```
'630'
```

Figure 40**: Total missing values in job data**

```
# Vérifier les valeurs manquantes
df.isnull().sum()
```

```
job_title          21
skill_required    195
location          414
link                0
dtype: int64
```

figure 39 **: Missing values in job dataset**

In any data analysis project, it's common to encounter missing data values, which can create problems when performing calculations and analyses. Missing values can be caused by a variety of reasons, such as data entry errors or incomplete data.

To handle missing values in our project, we decided to replace them with the "NaN" value. NaN stands for "not a number" and is a special value used in computing to represent undefined or unrepresentable values.

Furthermore we Dropped rows with '-' and '1' values. This step in data cleaning involves removing rows that have invalid or irrelevant data. In our dataset, missing values are also denoted by a '-' or '1' instead of the standard 'NaN' value, which can cause issues during analysis.

By dropping rows with these values, we can ensure that our data is accurate and relevant to the analysis we want to perform. This step is especially important when working with large datasets, as even a few incorrect or irrelevant rows can skew the results and lead to inaccurate conclusions.

In our case, these rows that contain missing values do not provide much information and can potentially affect the accuracy of our analysis. Therefore, dropping these rows might be a better option to ensure the quality of our data.

2.4 NLP Techniques:

In this part, we applied NLP techniques to clean and preprocess the text data in the 'skill_required' column.

- Stop word removal: removing common words like "the," "and," and "of" that do not provide much meaning to the text.
- Punctuation removal: removing punctuation marks like commas and periods that do not provide much meaning to the text.
- Newline character removal
- Dash character removal
- Accent mark removal: removing accent marks from characters using the unidecode library.

By applying these techniques, the text data is transformed into a format that is more suitable for natural language processing tasks such as text classification or topic modeling. And the objective of these techniques is to standardize the text, make it easier to analyze, and remove unnecessary noise.

Finally after all the applied techniques we finally have a dataset free of missing values that contains 1274 rows

```
df.isnull().sum()
```

```
job_title         0
skill_required    0
location          0
link              0
dtype: int64
```

```
str(df.isna().sum().sum())
```

```
'0'
```

figure 41**: Missing values after preparation in job offer data**

```
df
```

	job_title	skill_required	location	link
0	DevSecOps	software engineering, data development solutio...	SANTANDERTECHHUB-PROFILES	2022-11-11 Job description junior positions-pa...
1	Technology Talent	Broad spectrum technologies (including Blockch...	Poland, Portugal, Spain, UK, Mexico, Brazil an...	2022-11-11 Job description junior positions-pa...
2	DevSecOps Engineer	Collaboration, architecture, design, deploymen...	SANTANDERTECHHUB	2022-11-11 Job description junior positions-pa...
3	New technologies specialist	Kafka, Java8 (JDK11/17), Spring Cloud Stream, ...	not specified	2022-11-11 Job description junior positions-pa...
4	Data Management Solution Delivery	Cloudera/Spark/SnowFlake/Databricks/Stratio/Hi...	SantanderTechHub	2022-11-11 Job description junior positions-pa...
...	...	...	...	...
1586	Analyse des logs applicatifs	pour identifier des anomalies de fonctionnemen...	PFE-11	Value-page7.pdf
1587	Construire une base de données de référence	globale sur le SI, documenter et produire le m...	PFE-12	Value-page7.pdf
1588	Développement outil monitoring risques	implémentation chatbot intelligent, outil préd...	not specified	Value-page8.pdf
1589	utilisant différentes données textuelles colle...	Les techniques de webscrapping ou/et parsing d...	not specified	Value-page8.pdf
1591	Evaluation bilan carbone d'une entreprise	méthodologie, questionnaire (questions clés)	not specified	Value-page8.pdf

1274 rows × 4 columns

figure 42**: job offer dataset after preparation in job offer data**

Conclusion :

In this chapter, we applied various techniques, such as data cleaning, data integration, and data transformation. All these techniques are aimed at ensuring that the data is of high quality and ready for the next phase, which is data modeling.

Chapter 7: Modeling

Chapter introduction :

In this chapter, we will explore the process of building a matching and recommendation model using IBM Master Methodology. Here, we created 2 models for every type of recommendation. In every model we used different techniques and libraries .

1. First Model :

In this part we are going to match and recommend the best profiles for a given job offer .

1.1Matching and recommending candidates :

In this model we used the library FuzzyWuzzy which is a Python library used for string matching. Fuzzy string matching is the process of measuring the similarity between two character strings. Basically it uses Levenshtein Distance to calculate the differences between sequences which gives a similarity score ranging from 0 (no similarity) to 100 (perfect match).

We have the dataset of the job offers which contains the job title and the skills required of every job offer and also we have all the skills of the candidates from the student dataset .

So here we created a function that takes the job title and calculates the matching score between the skills required of that given job title and all the skills of the candidates using the function fuzz.token_set_ratio() of the library FuzzyWuzzy .

Then we sorted all the recommended candidates using a descending similarity rank then finally we took the best five candidates with the higher matching score.

```
[80]: find_recommendations1("DevSecOps Engineer")

Pour le 'Job Title': DevSecOps Engineer
Mr/Mme: Oussama HENI
 Avec Experiences: Full Stack Engineer
Son Lien Linkedin est le suivant: https://tn.linkedin.com/in/oussama-heni
Sa Location: Gouvernorat Ariana, Tunisie
Avec un Matching Score : 41

Pour le 'Job Title': DevSecOps Engineer
Mr/Mme: Hamdi Fhal
 Avec Experiences: nan
Son Lien Linkedin est le suivant: https://tn.linkedin.com/in/hamdi-fhal-a6b5121b2/en?trk=people-guest_people_search-card
Sa Location: Gouvernorat Tunis, Tunisie
Avec un Matching Score : 38

Pour le 'Job Title': DevSecOps Engineer
Mr/Mme: Hamza Arfaoui
 Avec Experiences: nan
Son Lien Linkedin est le suivant: https://tn.linkedin.com/in/hamza-arfaoui-91970b124
Sa Location: Gouvernorat Ariana, Tunisie
Avec un Matching Score : 35

Pour le 'Job Title': DevSecOps Engineer
Mr/Mme: Taha Touzri
 Avec Experiences: Python Developer, Python JavaScript developer, Test Automation Engineer, Python Developer, Research Assistan
t, Software Test Engineer
Son Lien Linkedin est le suivant: https://tn.linkedin.com/in/taha-touzri-274793230
```


figure 43 **:Matching and recommending candidates in model 1**

1.2 Matching and recommending job offers :

We implemented a matching and recommendation system for job seekers based on their skills and the required skills for job positions.

First, we checked if the job titles in data1 (which represents the candidate profiles) are valid by comparing them to the job titles in data2 (which represents the available job positions). Then, we created the function find_recommendations() which is defined to find job recommendations for each candidate.

The function iterates over each candidate in data1 and compares their skills with the required skills for each job position in data2 using the fuzzy string matching library called fuzzywuzzy. If the matching score is greater than 80, the job position is considered a potential recommendation and added to the list of recommended jobs for the candidate. The function returns a DataFrame with the candidate ID, name, link to their profile, recommended job titles, and matching score.

Finally, the top recommendation for each candidate is selected and displayed in the output.

```
                     _id                Name             link  \
642b57b2338e229c8c9bb4d8     Mohamed BOURAOUI  Value-page8.pdf
642b57b2338e229c8c9bb4d9   Abdelkader Dhouibi  Value-page8.pdf
642b57b2338e229c8c9bb4da        Mahmoud Segni  Value-page8.pdf
642b57b2338e229c8c9bb4db          Taha Touzri  Value-page8.pdf
642b57b2338e229c8c9bb4dc      Mahmoud Aloulou  Value-page8.pdf
                        ...                ...              ...
642b722e7624402b2227812e        Bassem Gharbi  Value-page8.pdf
642b722e7624402b2227812f         Oussama HENI  Value-page8.pdf
642b722e7624402b22278130        Hamza Arfaoui  Value-page8.pdf
642b722e7624402b22278131          Bacem Smiri  Value-page8.pdf
642b722e7624402b22278132  Mohamed Amine SALAH  Value-page8.pdf

                                    job_title matching_score
[Implementation process management APP Interns...            100
[Integration Internship, Azure Devops Internsh...            100
[Mobile application designer and developer, IA...            100
                  [IA, DataScience, Kotlin, Java]            100
[Buisness Analyst, QA, Data Scientist, IT Busi...            100
                                             ...            ...
[Implementation process management APP Interns...            100
[Flutter Mobile Developers, Développeur Front-...            100
[Implementation process management APP Interns...            100
[Consultant Engineer, Business analyst, C# Pro...             82
[Mobile application designer and developer, IA...            100
```

figure 44 **:Recommending one job for each candidate**

The list of matching scores is then sorted in descending order based on the matching score and converted into a list of dictionaries with keys 'job_title' and 'link' for each recommended job. The function then returns this list of recommended jobs.

```
[59]: recommended_jobs = get_recommended_job_titles1('642b57b2338e229c8c9bb4d8')
      for job in recommended_jobs:
        print('Le job title recommendé:', job['job_title'], ' Son LIEN:', job['link'])

Le job title recommendé: Implementation process management APP Internship  Son LIEN: DATAHORIZON-page17.pdf
Le job title recommendé: Ingénieur Data Scientist  Son LIEN: EY-page86.pdf
Le job title recommendé: Développeur de Reconnaissance de recherche  Son LIEN: GENITECH-page7.pdf
Le job title recommendé: Notification System Administrator  Son LIEN: Keyrus-page12.pdf
Le job title recommendé: équipe  Son LIEN: pfe-book-Aymax (1)-page17.pdf
Le job title recommendé: Data Scientist Engineer  Son LIEN: Value-page20.pdf
Le job title recommendé: Engineer  Son LIEN: Value-page21.pdf
```

figure 45 **:Recommending the list of best jobs for one candidate**

2. Second Model :

In this part we created 2 models and we compared between them to choose the best model to be deployed .

First of all we are going to work only on all skills required from the job dataset and the skills of the students from the student dataset . We did some lemmatization and stemming .

After that we converted all the skills in a list separated by commas so then we can apply word vectorizing and word embedding techniques. Then we created our training data for the model by mapping all the skills required and skills of the candidates after cleaning them with the library Gensim which is a well-known open-source Python library used in NLP and Topic Modeling. Its ability to handle vast quantities of text data and its speed in training vector embeddings set it apart from the other NLP libraries.

Moving forward we concatenated the two skill datasets and finally our training data is complete.

Here we created a word-to-vector model and we trained it on the training data that we created.

Then we created two functions to calculate the similarity and the matching score between two word vectors, one being the skills required and the other being the skills of the student using the given model .

One of the functions is using the cosine matrix and the other using the average cosine matrix .

Then to go a little further we also used The BERT pretrained model to see which model is better to calculate the matching score between the skills and we finally found that the word-2-vector model that we created made better results then the BERT model since the model that we created is trained on our data .

2.1Matching and recommending candidates :

```
match_job(df1.iloc[3],df2,emb_vec)
```

```
C:\Users\MSI GF63\AppData\Local\Temp\ipykernel_15404\3122937574.p
  return dot(sen_vec1,sen_vec2)/(norm(sen_vec1)*norm(sen_vec2))
```

```
(URL                          https://www.linkedin.com/in/nourkobbi/
 job_title                      data science and engineering student
 Name                                                      Nour Kobbi
 skills          [angularjs net framework developpement logicie...
 experiences                                          nan:nan,nan:nan
 Name: 247, dtype: object,
 0.7105570850909376,
 ['Kafka',
  'Java8 ( JDK11',
  '17 )',
  'Spring Cloud Stream',
  'Spring Boot',
  'Spring Cloud ( Web',
  'Data',
  'JPA )',
  'Spring Framework Reactive',
  'WebFlux',
  'Microservices',
  'Event-Driven architecture',
  'MongoDB',
  'Neo4J',
  'SaaS software level',
  'Openshift',
  'K8S',
  'Flink'],
 [1.0,
  1.0000000000000002,
  1.0000000000000002,
  0.36318970446000404,
  0.3722733426152549,
```

figure 46 **:Matching and recommending candidates in model 2**

2.2 Matching and recommending job offers :

figure 47**:Matching and recommending jobs in model 2**

	URL	Name	job_matched	job skills	total_score	score_list
2	https://www.linkedin.com/in/talel-	Talel Kbaier	Implementation process management APP Internsh	['Java']	1	[1.0000000000000002]
3	https://www.linkedin.com/in/oma	Omar Talbi	Implementation process management APP Internsh	['Java']	1	[1.0000000000000002]
4	https://www.linkedin.com/in/hass	Hassen Knani	Master Ingénieur réseau et sécurité	['Projet']	1	[1.0000000000000002]
5	https://www.linkedin.com/in/safw	Safwen Dammak	CONCEPTION ET DÉVELOPPEMENT D'UN SIMULAT	['C++', 'QT', 'LINUX', 'SOME', 'IP', '']	1	[1.0000000000000002]
6	https://www.linkedin.com/in/oum	Oumayma Béhi	innovation specialist	['curiosity', 'creativity', 'adaptability']	0,26507317	[0.2622474085608724, 0.2650820348991975, 0.2678
7	https://www.linkedin.com/in/med-	Med Yassine Ben Romdhane	Trusted Firmware Implementation	['C', 'C++', 'STM32', 'UI Development']	1	[1.0000000000000002]
8	https://www.linkedin.com/in/hous	Houssem Derouich	Développeur Front-End	['Javascript', 'CSS', 'HTML']	0,976186893	[1.0000000000000002, 0.9523737865925803]
9	https://www.linkedin.com/in/abde	Omar Abdelkefi	Implementation process management APP Internsh	['Java']	1	[1.0000000000000002]
10	https://www.linkedin.com/in/med-	Med.Ridha Harhira	candidature	['COMMUNICATION']	1	[1.0000000000000002]

Conclusion :

Since the model word-2-vector that we created brought the best results we decided to deploy it in our platform .

Chapter 8: Deployment

Django is a high-level Python web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel. It's free and open source.

- Ridiculously fast.

  Django was designed to help developers take applications from concept to completion as quickly as possible.
- Reassuringly secure.

  Django takes security seriously and helps developers avoid many common security mistakes.

1. MVT Architecture :

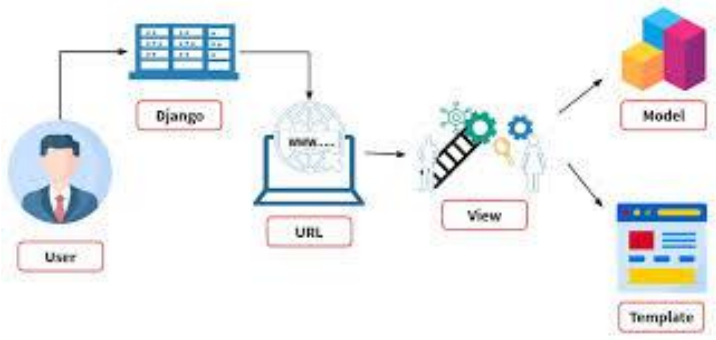


figure 48 **:MVT Architecture**

Django is based on MVT (Model-View-Template) architecture. MVT is a software design pattern for developing a web application.

MVT Structure has the following three parts :

- Model: The model is going to act as the interface of your data.
- View: The View is the user interface — what you see in your browser when you render a website. It is represented by HTML/CSS/Javascript .
- Template: A template consists of static parts of the desired HTML output as well as some special syntax describing how dynamic content will be inserted.

We used also pipeline

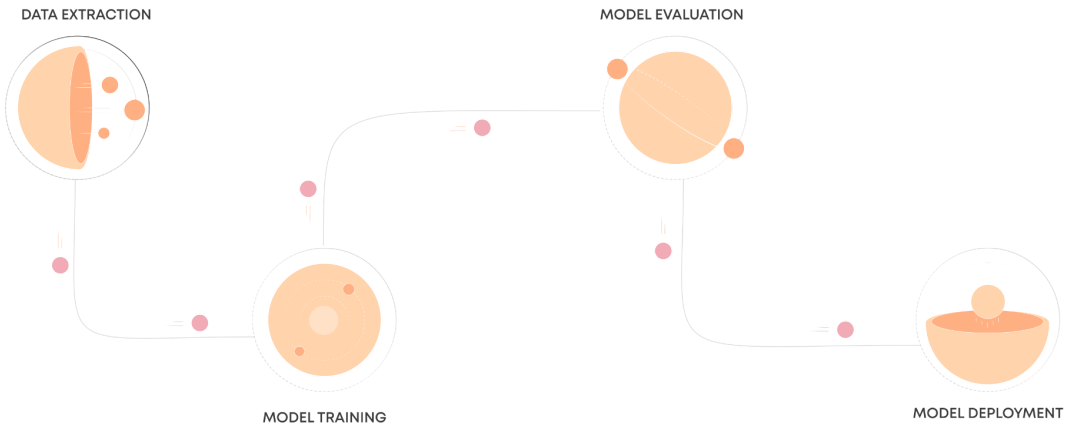


figure 49 **:Pipeline figure**

Pipelines are important in deployment because they provide a consistent and reproducible way to build, train, and deploy models. A pipeline typically consists of a series of steps or stages that are designed to prepare and preprocess data, train a model, and make predictions. By using a pipeline, you can ensure that the same steps are followed every time you deploy a model, which can help to reduce the risk of errors and improve the reliability of the results. In addition, pipelines can make it easier to automate the deployment process, so you can deploy new models or update existing ones more quickly and efficiently. Finally, pipelines can help to improve the performance and scalability of a model by allowing different steps to be run in parallel, which can help to reduce the overall time required to make predictions.

General Conclusion :

In conclusion, this project demonstrates the potential impact of data science on employability and the challenges faced by organizations like the Employability Pole of the Esprit Group. By utilizing data science methods and techniques, the PE can manage job and internship opportunities more effectively, gain a deeper understanding of the job market, and provide valuable guidance to students. This report outlines the project's overview, business domain, analytic approach, requirements, data collection, preparation, modeling, and deployment. By implementing data science techniques, organizations can improve their services and overcome the challenges they face in the modern job market.

References :

https://www.mongodb.com/

https://www.geeksforgeeks.org/fuzzywuzzy-python-library/

https://www.analyticsvidhya.com/blog/2022/03/learn-basics-of-natural-language-processing-nlp-using-gensim-part-1/

https://huggingface.co/models

https://support.microsoft.com/en-us/office/what-are-imap-and-pop-ca2c5799-49f9-4079-aefe-ddca85d5b1c9

https://www.analyticsvidhya.com/blog/2020/10/word-cloud-or-tag-cloud-in-python/

https://www.ibm.com/docs/fr/spss-modeler/saas?topic=dm-crisp-help-overview

https://www.djangoproject.com/

https://chat.openai.com/

https://www.djangoproject.com/?fbclid=IwAR3vfB6yLNzhI3S4cXctKwhizZbRNSeCwR9dlYZNdyl42OfEsmewjbP0

https://scrapeops.io/python-scrapy-playbook/pythonscrapy-linkedin-people-scraper/

https://platform.openai.com/docs/apireference

ESPRIT SCHOOL OF ENGINEERING

www.esprit.tn - E-mail : contact@esprit.tn

Siége Social : 18 rue de l'Usine - Charguia II - 2035 - Tél. : +216 71 941 541 - Fax. : +216 71 941 889

Annexe : 1-2 rue André Ampère - 2083 - Pôle Technologique - El Ghazala - Tél +216 70 250 000 - Fax +216 70 685454