



Competency Monitoring System
application

CRISP-DM

**Jason Tomeij, Baha
Kucuk, Fatima
Memon, Yara Mousa
& Zoë González**



Version management

Version Number	Date	Author	Change
0.1	17-09	Jason	Started on working on the Current situation.
0.2	18-09	Jason + Baha	Change of the documentation + Lay-out of the document.
0.2	18-09	Jason + Baha	Changed the SIPOC + Making the BPMN Model.
0.3	18-09	Zoë	Changed front page, put business goals in table.
0.4	22-09	Zoë	Added intros and small changes to balanced scorecard.
0.4	22-09	Fatima	Onderbouwing + verbetering SWOT-analyse, Vijf porter's krachtenmodel.
0.4	22-09	yara	Visie&scope Stakholders analyse Ist-soll-Gap.
0.5	24-09	Jason	Looking and changing what is needed for a better outcome.
0.6	25-0	Jason	Making the document in English.
0.7	25-09	Zoë	Translated my parts to English.
0.8	26-09	Zoë	Added target levels to balanced scorecard that the client provided.

0.9	25-9-2025	Fatima	Translated the parts to english.
1.0	26-9	Baha	Translated Sipoc to English
1.1	30-09-2025	Jason	Starting the Project plan.
1.2	01-10-2025	Jason	Started on Stakeholder analysis + Worked on new BPMN Model for the expected end result.
1.3	02-10-2025	Jason	Added new BPMN Model +.
1.3	02-10-2025	Baha	Azure database uitgelegd.
1.3	2-10-2025	Fatima	Scraper + deadlines and planning opgesteld.
1.4	03-10-2025	Baha + Jason	Making new document to make things less chaotic. Was needed after feedback.
1.5	06-10-2025	Baha + Jason	Making the new Risk analysis and added and cost and benefits.
1.6	07-10-2025	Jason	<p>Started working on new BPMN Model after review with Hend.</p> <p>Also made a MoSCoW Table for business success criteria.</p> <p>Added SWOT Analysis.</p> <p>Started on Project plan</p>

			Added Stakeholder analysis + Added Project plan Timeline and Project Overview.
1.6	07-10-2025	Zoë+ yara	Inventory of resources and a part of requirements, assumptions and constraints.
1.6	07-10-2025	Fatima	Initial assessment of tools and techniques, data mining success criteria.
1.6	07-10-2025	Baha	Keep document in order and made a part of project plan.
1.6	07-10-2025	Yara	Terminology and a part of requirements, assumptions and constraints.
1.6	10-10-2025	Fatima	Data mining success criteria verbeterd.

Introduction

For Utrecht University of Applied Sciences, we will develop a website where the lectorate and institutes can upload curricula. In the backend, a program will run that matches these curricula and the associated soft skills with the top 100 jobs from LinkedIn.

The choice for a website was made because it is more user-friendly and more accessible for students. Once the system is fully developed, the website can be used everywhere, making the program accessible to anyone who wishes to use it.

By scraping LinkedIn, we will collect job postings and store them in our own database, allowing us to create a reliable match between a student's curriculum and the requirements of companies.



Table of contents

Versiebeheer	Fout! Bladwijzer niet gedefinieerd.
Introduction.....	5
Table of contents	6
Determine business objectives	8
Background	8
Business objectives.....	12
Business success criteria	15
Assess situation.....	17
Inventory of resources	Fout! Bladwijzer niet gedefinieerd.
To carry out the project successfully, it is important to have a clear overview of all available resources. This includes not only the people involved, but also the data sources, the technical tools, and the software needed to build the system. The following table provides an overview of these resources and their specific role within the project.....	17
Requirements, assumptions and constraints	18
Risk and contingencies.....	21
Terminology.....	23
Cost and benefits.....	23
Determine data mining goals.....	24
Data mining goals	24
Data mining success criteria	25
Produce project-plan.....	26
Project-plan.....	26
Initial assessment of tools and techniques	30
Bronnen	Fout! Bladwijzer niet gedefinieerd.
Introduction.....	1
Table of contents	1
Determine business objectives	2
Background	3
Vision & Scope.....	3
Business objectives	3

Business succes criteria	3
Assess situation	3
Inventory of recources	3
Requirements, assumptions and constraints	3
Risk and contingencies.....	3
Terminology.....	3
Cost and benefits.....	3
Determine data mining goals.....	3
Data mining goals	3
Data mining sucess criteria.....	3
Produce project-plan.....	3
Project-plan.....	3
Initial assessement of tools and techniques.....	3
Bronnen	3

Determine business objectives

Background

Vision & Scope

We are starting this project to demonstrably reduce the gap between education and the labour market. Without a clear vision and a well-defined scope, it remains unclear which analyses should be carried out and which solutions should be developed.

Mission: Our mission is to develop a sustainable, reliable, and data-driven Competency Monitoring System that helps the Hogeschool of Utrecht and other Hogeschoolen in the Netherlands. We analyse marketing job vacancies and link them to curricula, so that schools, students, and employers can better anticipate current and future competency needs.

Vision: CMS will develop into a national standard that enables universities of applied sciences to identify which skills are important in marketing, both now and in the future.

Strategy:

- Automatic data collection (APIs, open data, partners)
- Matching algorithm (study programs ↔ labor market roles)
- Annual updates with new analysis
- Interactive dashboard for stakeholders
- GDPR/ToS compliant processes

Scope

Category	Description
In Scope	Collect job postings (APIs, open data, company websites)Extract job details (title, skills, location, sector, salary, language, experience)Compare HU programs with job profilesDevelop dashboard (skills, skill gaps, regional trends)
Out of Scope	Real-time scraping/analysis International vacancies Forecasts beyond five years
Future Scope	Extend to other universities (e.g. HvA)Add scenario analysis (VSM)Early warning system for labor market changes

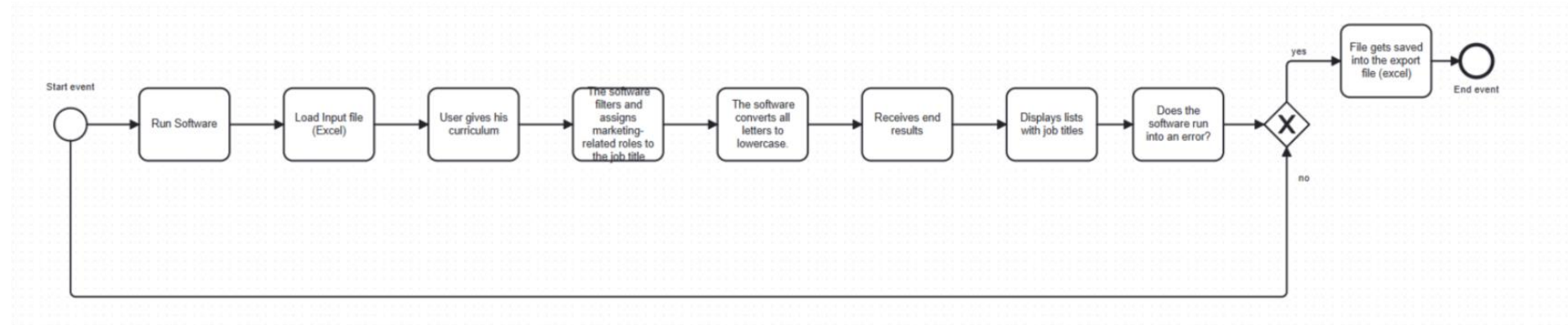
SIPOC:

To gain a proper understanding of the current situation, we use a SIPOC analysis to create insight into the existing processes, systems and stakeholders involved. In addition, by applying a BPMN Model, we can clearly visualize how the current process operates. This provides us with a solid understanding of the present workflow and highlights where improvements can be made in the new website.

Sipoc Analyse				
Supplier	Input	Process	Output	Customer
<ul style="list-style-type: none">- LinkedIn- Indeed- Dennis.H	<ul style="list-style-type: none">- Data from LinkedIn 2021<ul style="list-style-type: none">- Hard code- Entered manually- No documentation- CSV File	<ul style="list-style-type: none">- Manually retrieve data from current vacancies on LinkedIn- Loading the data into code- Code runs software and searches for keywords to connect vacancies- Generating response- Answer comes back through the code	<ul style="list-style-type: none">- Simple answer via code- CSV bestand- Documentatie	<ul style="list-style-type: none">- Dennis.H- Hogeschool Utrecht - Lectoraat Marketing & Customer Experience

With the help of a SIPOC analysis, we mapped out the process and made it as transparent as possible. This made it clear which parties are involved, what data enters the process, how the steps are carried out, and which results are ultimately delivered to the end user.

BPMN (Current situation)

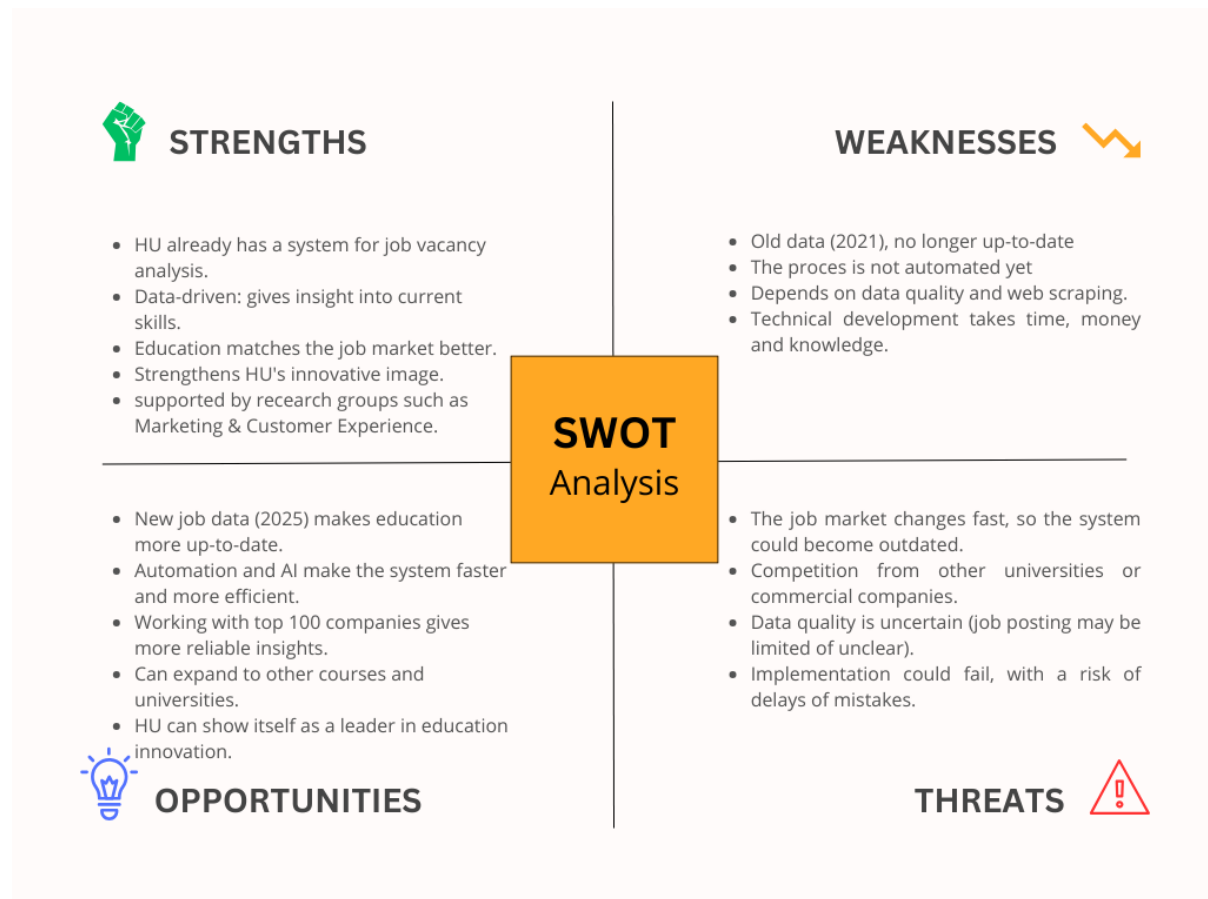


By modelling the process in a BPMN diagram, we now have a clear view of how the process functions. This provides a better understanding of the current situation and forms a foundation for optimization.

The process begins with starting the software. Once initialized, the input file is uploaded. Next, the curriculum is entered, after which the software assigns marketing-related roles to the position titles. The text is then automatically converted to lowercase. The software subsequently generates statistics and displays a list of position titles. A check is performed to determine whether the software produces an error message. If an error occurs, the process restarts. If no error is detected, the file is saved as an output file and the process is completed. Now we have an output file with the jobs that match the curriculum

SWOT-analysis

To understand the organisation and system better, we use a SWOT analysis to look at the system's strengths, weakness, opportunities, and threats. This helps to make the system strong and reliable for the future.



Hogeschool Utrecht has a good base with the competency monitoring system to connect education and jobs. But the system is still old data, takes a lot of work and depends on good job data. There are big chances to improve it with new data, AI and working with companies. Fast changes in the job market, competition and technical problems are big threats.

Business objectives

Business Goals

To provide direction to the project and clarify where we want to go with it, we have defined business goals. These goals explain why we are doing this project and what we aim to achieve. They form the foundation of the project and were established based on discussions with the client.

Business goal	Description	Project goal
Building a bridge between education and the labour market	Contribute to reducing the competency gap and help educational programs train student with the right knowledge and skills.	The system should support programs in evaluating and adjusting their curricula, thereby building a bridge between education and the labor market.
Improving and updating curricula	Evaluate and adapt educational programs so that graduates' competencies better align with the needs and requirements of the labour market.	Develop a matching algorithm that links competencies from job vacancies to the HU curriculum.
Increasing labor market relevance	Reduce the gap between education and the professional field by having real-time insights into required competencies.	Deliver a user-friendly, interactive dashboard that provides programs and researchers with real-time insights into competencies and trends.
Strengthening research and education with up-to-date insights	Provide teachers, researchers and policymakers with reliable and current data to support decision-making.	Use recent and representative data (2025) from the top 100 Dutch companies.
Developing future-proof educational programs	Look ahead and anticipate the future competency needs of the top 100 companies in the Netherlands.	Develop a predictive model that can forecast competency trends based on historical and current data.
Increasing efficiency and sustainability	Automate the process of data collection and analysis so it can be repeated annually and remain up to date without high additional costs or manual effort.	Within 6 months, realize a fully automated pipeline that collects and processes job vacancies, and set up a system that can run at least once a year and be easily updated.

Balanced score card

To view the business goals from different perspectives, a balanced scorecard has been created. This provides a more complete overview, as the goals are distributed across multiple perspectives. This ensures that no aspect of the project gets overlooked or forgotten. This card was discussed with and reviewed by our client, who also helped us determine the target levels.

Perspective	Goals	KPI's	Target Level	Initiatives / Actions
Financial	Cost-efficient system that can be deployed sustainably	<ul style="list-style-type: none"> - Annual operational costs - Savings through automation 	<ul style="list-style-type: none"> - €500 per year in maintenance costs - At least 80%-time savings compared to manual processing 	Automate data collection and processing
Customer (HU professors, training & education, managers, researchers)	Satisfied stakeholders and usability of insights	<ul style="list-style-type: none"> - Satisfaction score of programs/researchers - Number of curricula that have been adjusted 	<ul style="list-style-type: none"> - ≥80% satisfaction - ≥2-10 programs use the dashboard structurally 	Develop an interactive dashboard with real-time competence insights
Internal Processes	Efficient and reliable collection/processing of data from up-to-date sources	<ul style="list-style-type: none"> - % of automated tasks - Lead time of data processing 	<ul style="list-style-type: none"> - ≥95% automated 	Build a scraping pipeline for LinkedIn, indeed including top 100 Dutch companies
Learning & Growth	Future-proof insights and scalability to other disciplines	<ul style="list-style-type: none"> - Prediction accuracy - Number of competencies aligned with the market 	<ul style="list-style-type: none"> - ≥70% prediction accuracy for 2-3 years ahead - At least 2 new programs added/adjusted per year 	Develop a predictive model and matching algorithm (VSM)

By using a MoSCoW table we can decide as team what is essential versus optional. Instead of trying to do everything at once, we can prioritize the core features that are necessary for the system to succeed. It also keeps the project realistic and on schedule.

MoSCoW Table Business Success Criteria

Must have <ul style="list-style-type: none">1. Website2. Data scraping module (LinkedIn / Indeed)3. Matching algorithm4. Interactive dashboard5. Database	Should Have <ul style="list-style-type: none">1. Historical data storage2. Integration with educational systems3. Data quality checks4. Error logging and monitoring tools5. Basic reporting export (PDF, Excel)
Could Have <ul style="list-style-type: none">1. Custom filters and visualization (Dashboard)2. AI-based prediction module3. Data visualization templates	Won't Have <ul style="list-style-type: none">1. Mobile Responsiveness2. Full LinkedIn scraping automation3. Predict trends and change

Business success criteria

What is the end product?

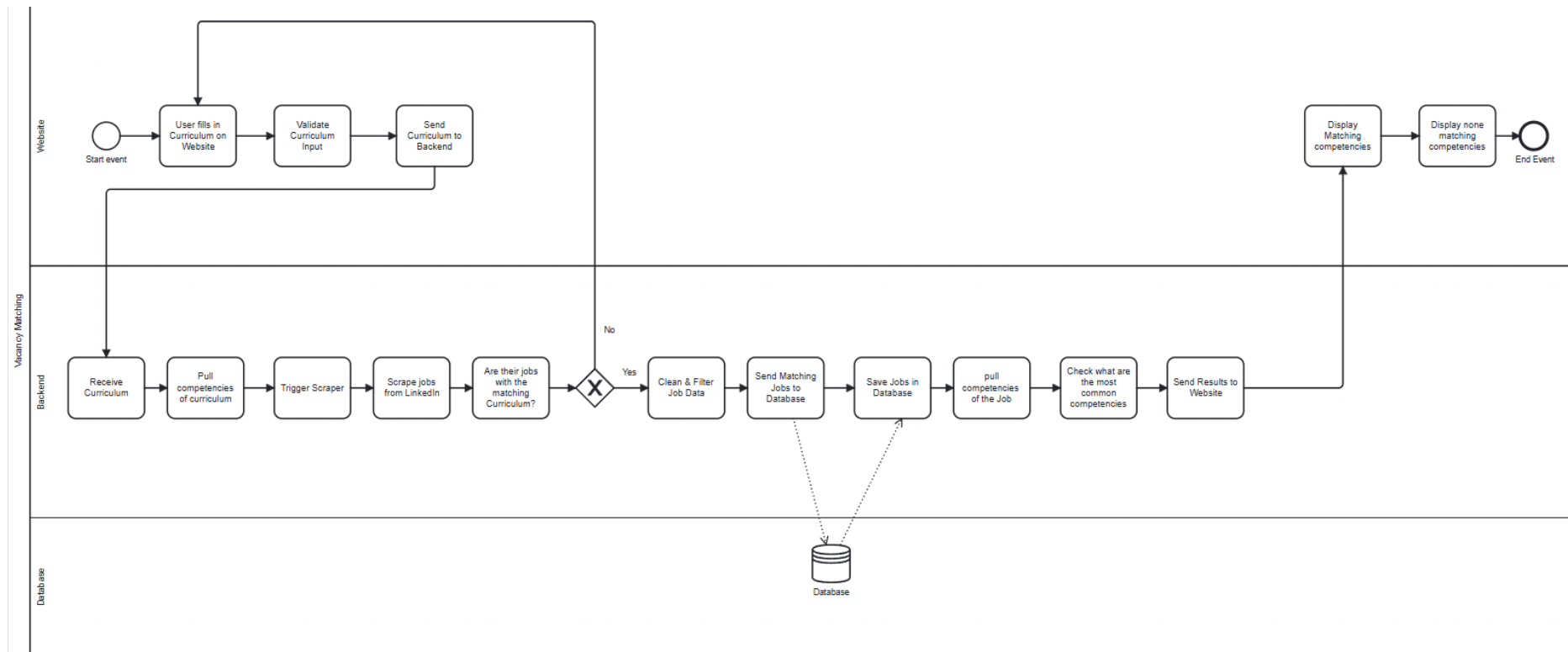
For Utrecht University of Applied Sciences, we will develop a website where the lectorate and institutes can make an input with the name of the curricula. In the backend, a program will run that matches these curricula and the associated soft skills with the top 100 jobs from LinkedIn.

The choice for a website was made because it is more user-friendly and more accessible for students. Once the system is fully developed, the website can be used everywhere, making the program accessible to anyone who wishes to use it.

By scraping LinkedIn, we will collect job postings and store them in our own database, allowing us to create a reliable match between a student's curriculum and the requirements of companies.

To ensure that the process remains clear, structured, and up to date, we created a BPMN model of the desired situation. This model provides a visual representation of the target process, making it easier to identify improvements.

BPMN Model (Wanted situation)



As you can see in this BPMN model, many aspects have been adjusted, especially the steps required to reach the final result. With these changes, we aim to improve the process and ensure that everything runs more smoothly without pitfall

Assess situation

Inventory of resources

To carry out the project successfully, it is important to have a clear overview of all available resources. This includes not only the people involved, but also the data sources, the technical tools, and the software needed to build the system. The following table provides an overview of these resources and their specific role within the project.

Category	Resource	Description
Personnel	Projectteam (Students)	Responsible for the overall project execution, including research, development of the website, data collection and dashboard design
	Business expert (HU lectorare/client)	Provides insight into the project and validates the project goals
	Project supervisor	Provides guidance and feedback throughout the project.
Data	Indeed/LinkedIn job postings (Top 100 Dutch companies)	Used to identify and analyze competencies and skills that are required in the Dutch labour market.
	HU curriculum and study programs	Provides information on current competencies taught within the school's programs
Computing Resources	Azure cloud environment	Hosts the web application, stores datasets and runs the data processing pipeline
	Student workstations/laptops	Used by project team for testing, training and development
	Database	Stores scraped job data and curriculum information securely
Software	Python	For data scraping, cleaning, processing
	Power BI	For visualization and reporting the results in an interactive dashboard
	HTML, CSS, JavaScript and PHP	For developing the website's interface and ensuring accessibility
	Azure/GitHub	For version control, collaboration and hosting the backend system
	Microsoft Office	For documentation, reporting, and communication within the team

Requirements, assumptions and constraints

Requirements *

Functional

Requirement Website	Description
Website	A website where lectorates and programs can upload their curricula. The backend matches these curricula (including soft skills) with job vacancies from LinkedIn and Indeed.
Data storage	Each job posting must include the following data fields: title, skills, competencies, location & sector
Dashboard C reporting	Results are shown in an interactive dashboard with top skills, skill gaps per program, and regional trends.
Curriculum comparison	The system compares curricula with job profiles to identify gaps and trends in competencies.
Data collection	Job data is automatically scraped from LinkedIn and Indeed, focusing on the top 100 Dutch companies
Automation C scheduling	The pipeline must run automatically with error logging and minimal manual input.
Annual update	The data collection and matching process must be repeatable every year with new data.

Non-Functional Requirements

The system must be easy to use for non-data experts

It should operate at least 95% automatically, requiring minimal manual work.

Annual maintenance costs must stay below €500 to ensure sustainability.

All data processing must follow GDPR and the Terms of Service of LinkedIn and Indeed. The system must be stable, reliable, and scalable,

We are building a website where programs and professorships can upload their curricula. A program runs in the backend that links these curricula (including soft skills) to current Dutch.

Assumptions

During the project this far, several assumptions were made that form the foundation of the project's planning and execution. These assumptions describe the conditions that are expected to be true but cannot yet be guaranteed. By defining them, we can ensure that the project remains realistic and that potential risks can be identified early on.

Assumption	Description
Data availability	We assume that LinkedIn and Indeed will continue to have publicly available job postings that can be legally scraped
Data quality	We assume that job postings contain enough structured information to extract useful data
Stakeholder participation	We assume that lecturers and institutes will cooperate by providing curricula to test and validate the system
Tool stability	We assume that Python libraries used for scraping and Azure services will remain available and functional during the project
Planning	We assume that the project can be completed within one semester
Teamwork	We assume that all team members remain available for the duration of the project and are capable of completing their assigned tasks

Constraints

Within the project there are several constraints that affect the scope, timeline and technical possibilities. These limitations are mainly related to the available time, resources and tools provided.

By identifying these limitations, we ensure that the business goals remain achievable, and the project stays manageable throughout the entire process.

Constraint	Description
Project timeline	The system must be developed, tested, and presented within one semester. This limits the time available for testing and optimization.
Resource availability	This project is carried out by students, which means available time, expertise, and computing resources are limited. Development and testing are therefore restricted to what can be achieved within the team's skill set.
Technical limitations	The project depends on the available Azure environment provided by HU. Database storage, computing capacity, and access rights are limited to the resources accessible within this environment.
Data access	The system relies on public job postings from LinkedIn and Indeed. If access is restricted, data scraping may be partially or temporarily unavailable.
Legal and ethical boundaries	All data collection and processing must comply with GDPR regulations and the ToS of the platforms used (LinkedIn, Indeed).
Scope limitation	The CMS focuses solely on Dutch job postings from the top 100 companies and the HU marketing programs. International vacancies and other disciplines are currently outside the project scope

Risk and contingencies

A risk analysis is needed to find and manage problems with privacy, technology, and using the competency monitoring system early. This helps take action and make sure the project succeeds.

Risk	Chance	Impact	Weight	Consequence	Measure
Data format is wrong	High 2/5	2/5	Low	To much data received and its not clear	Filtering list with selected data
LinkedIn not scrapeable	4/5	4/5	Low	Switch to indeed for data scraping	Scrape part by part
LinkedIn partly scrapeable	2/5	2/5	Low	Switch to indeed for scraping data	Scrape part by part
Indeed not scrapeable	1/5	5/5	High	No data available for project	Scrape part by part
Indeed partly scrapeable	1/5	5/5	High	Not enough data for project	Scrape part by part
Sick or personal situation of staff	3/5	3/5	Medium	Part of the sick person can't be worked	Report on time/ Other team member picks it up
LinkedIn under maintanance /Downtime	2/5	5/5	High	Scraping not available at that moment	Waiting till its up
Website response error	3/5	4/5	Medium	System crashes or slow response time	Conduct stress testing: optimized back end performance and surface scaling
HU site has downtime	2/5	4/5	Medium	HU course data temporarily unavailable	Retry connection

Format courses are wrong on the HU website	3/5	3/5	Medium	Wrong or incomplete curriculum in system	Validate and clean course data automatically
Over budget with Azure	2/5	5/5	High	Unexpectedly high hosting or processing caused	Monitor resource usage: set budget alerts, optimized data processing
Leak in data privacy	1/5	5/5	High	Violation of GDPR	Encrypt all data
Deadline is not met	3/5	4/5	Medium	Project deliverables delayed	Use agile sprint
Insufficient alignment between project goals and organizational goals.	2/5	3/5	Medium	Final product doesn't meet HU or lectorate expectations	Regular stakeholder meetings

This table gives us a clear understanding of what could go wrong during the project and what we can do to address it. In this way, we ensure that when we encounter any issues, we can deal with them effectively and continue working toward our final goal.

Terminology

Word	Meaning
Scrapen	Automatically collecting job postings from websites (e.g., LinkedIn/company sites) and storing the relevant info in our database for analysis.
Curriculum	The study plan of a degree program: courses + learning outcomes. We upload this to compare it with the labor market.
Matching-algoritme	The (comparison) between two words: skills in the curriculum vs. skills in job postings. The algorithm finds matches and gaps and shows them in the dashboard.
Skill gap	The difference between the skills in the curriculum and the skills employers ask for (missing or weak areas).
Pipeline	The fixed steps data goes through ingest → clean/transform → store → analyze → show in the dashboard (as automated as possible).
GDPR/AVG	The EU privacy law. We work in compliance with GDPR and also follow each website's Terms of Service—meaning we handle access, rate limits, and what is/isn't allowed properly.

Determine data mining goals

Data mining goals

Organizations have access to increasingly large amounts of data, but leveraging this information is crucial for making better decisions. Data Mining helps uncover hidden patterns and trends, while methods such as Business Case Model, SWOT-analysis, and the five forces model link these insights to strategy and market positioning. This enables companies to seize opportunities, reduce risk and strengthen their competitive position.

Business Case Model:

Organizations need to justify investments carefully before committing time, money and resources. The Business Case Model helps make the cost, benefits and risk of a project clear. This allows for better-informed decisions, more efficient allocation of resources and a higher likelihood of project success.

The Business Goals of the Competency Monitoring System:

The Business Goals are outlined on the Business Goals page: number seven.

The Data Mining Goals of the Competency Monitoring System:

1. From semi-Manual to a fully automated pipeline.
2. Scraping information/data from websites (such as LinkedIn and Indeed)
3. Developing a matching algorithm: building a tool that compares what companies require with what students learn.
4. Reporting/dashboarding (preferably an interactive dashboard)
5. Annual update frequency
6. The ability to use trend reports and historical data to make predictions/forecast.
7. Building a system that can easily incorporate new data sources and courses.

Data mining success criteria

Task

The goal of this step is to describe what the data mining part of the project should achieve in technical terms. These goals support the business aim of connecting education with the labour market, improving study programs, and using data to make better decisions.

Data mining goals and success criteria

Data Mining Goals	Data Mining Success Criteria
From semi-Manuel to a fully automated pipeline.	<ul style="list-style-type: none">- At least 60% of the data proces (input, cleaning, storage) runs automatically.- The route makes the work 50% faster than before.
Scraping information/data from websites (such as LinkedIn)	<ul style="list-style-type: none">- At least 80% of collected data is saved correctly.- At least 70% of the expected fields are collected.- collected data follows the correct format (like CSV, JSON) in 80% of the cases.
Developing a matching algorithm: building a tool that compares what companies require with what students learn.	<ul style="list-style-type: none">-The algorithm matches students and companies correctly 70% of the time.- A user can create a match report within 5 clicks.
Reporting/dashboarding (preferably an interactive dashboard)	<ul style="list-style-type: none">- The dashboard loads quickly (in less than 6 seconds)- All charts can be clicked or filtered.
Annual update frequency	<ul style="list-style-type: none">-Trend charts are 65% correct of the dataset.
Building s system that can easily incorporate new data sources and courses.	<ul style="list-style-type: none">-A new source can be added in less than 6 hour.

The Competency Monitoring System wil automatically collect and analyse LinkedIn data about marketing jobs and skills. It will compare this information with what students learn in the Marketing study program. The results will be shown in an interactive dashboard, helping the university understand skills gaps.

Produce project-plan

Project-plan

Project Overview

For Utrecht University of Applied Sciences, we will develop a website where the lectorate and institutes can make an input with the name of the curricula. In the backend, a program will run that matches these curricula and the associated soft skills with the top 100 jobs from LinkedIn.

The choice for a website was made because it is more user-friendly and more accessible for students. Once the system is fully developed, the website can be used everywhere, making the program accessible to anyone who wishes to use it.

By scraping LinkedIn, we will collect job postings and store them in our own database, allowing us to create a reliable match between a student's curriculum and the requirements of companies.

It's also important to make sure that the project is clear which stakeholders are involved. To make sure that everything is clear we made a Stakeholder Analysis.

Stakeholder	Role	Internal/ External
Hend Elsayed	Client, Final Responsible	Internal
Gerrita van de Ven	lector in the Lectoraat marketing and consumer experience	Internal
HU IT Department	Technical support	Internal
HU Lecturers	Main end users	Internal
Top 100 Companies	Provide job posting data	External
Other Universities	Potential future users	External
Students	Secondary End Users	Internal

Timeline & phases

To create a clear overview of the project flow we are going to define phases. With the phases. The project has a total of 8 phases. These eight phases are each two weeks long. This gives us enough time to make sure that every phase is completed.

Sprint	Week	Goal	Tasks	Deliverables
Sprint 1	1-2	Introduction	Kick-off Meeting + Team agreements	Kick-off document. First version of Business understanding
Sprint 2	3-4	Business Understanding	Making the business understanding assignment	Business understanding document
Sprint 3	5-6	Process feedback business understanding + Starting on Data Understanding	Process feedback of teachers into our document	Final product of Business Understanding + start of Data understanding
Sprint 4	7-8	Finalizing Data Understanding	Finalizing document of Data understanding and getting to know how scraping works	Data Understanding document + Start of Data preparation
Sprint 5	8-10	Finalizing Data preparation + Start on Modelling	Finalizing document of Data Preparation + Start on coding	Data preparation document + First work of Code.
Sprint 6	11-12	Working on Modelling fase	Making the product	Tests of product
Sprint 7	13-14	Evaluation & Start Deployment	Finishing the product + making it ready for deployment	Reviewing the product and evaluate it. Starting the deployment of the product

Sprint 8	15-16	Extra time	Resolve delays + Implement additional client requests	Final presentation + Final product
-----------------	-------	------------	---	------------------------------------

Resources & Roles

The resources and roles within the project have been clearly defined. The resources were previously outlined in the *Inventory of Resources* section on page 15.

The roles have been clarified by creating a table that specifies each team member's responsibilities.

Naam	Rol
Jason	Leading the project. Responsible for project plan in data understanding with Baha.
Baha	Keeping all project documents organized. Responsible for Project plan in data understanding with Jason.
Zoë	Responsible for inventory of resources and a part of requirements, assumptions and constraints in data understanding.
Yara	Responsible for a part of requirements, assumptions and constraints and terminology in data understanding.
Fatima	Responsible for Initial assessments of tools and techniques and data mining success criteria in data understanding.

This is, of course, a summary of how the process works. The specific task assignments are discussed via WhatsApp, where the corresponding deadlines are also communicated.

Documentation & Review

All project documents are managed by **Baha**. This ensures that the documentation remains well-organized and forms a coherent narrative, rather than having separate, uncoordinated sections written by different team members. The completed work is reviewed by **Baha** and **Jason** to ensure that all parts align properly and maintain consistency throughout the project.

Every two weeks, we hold a **sprint review** where we present our project progress and upcoming plans. During these sessions, we receive feedback which we then apply to improve our workflow and the overall project outcome.

Initial assessment of tools and techniques

At the end of the first phase, we reviewed possible tools and techniques for building a prototype of a sustainable, automated competency monitoring system.

The current system predicts which skills are needed in the professional field based on job vacancies. This helps Dutch universities of applied sciences align their study programs with labour market needs. However, it uses outdated data from 2021 and requires manual work.

Given the short timeframe of about four months, the focus is on building a working prototype that:

- Runs the full data process automatically (collecting, cleaning, matching and reporting)
- Matches job skills with student competencies
- Shows result in a simple interactive dashboard
- can be scaled in the future to include new data sources and programs.

Purpose of the assignment.

The assessment helps select tools and techniques that best support the project's data mining goals. These tools and techniques will help collect, clean, and analyze job and curriculum data, match skills and show results in a dashboard. The data mining goals and success criteria, which explain what the system should achieve, are listed on the data mining goals page. This assessment makes sure the chosen methods are practical, efficient and suitable for the project's short timeline.

Assessment of tools

Tool	Function	Strengths	Weaknesses	Evaluation
Python (pandas, beautifulSoup, scikit-learn, spaCy and Selenium)	Data collection, cleaning, matching and automation.	Open-source, large community, supports automation and scraping.	Requires coding skills	Highly recommended; ideal for automated pipeline and webscraping.
MSQL / Azure	Data storage and management	Reliable, scalable, intergrates with python and other tools.	Requires setup and maintenance	Recommended: stores data for historical tracking and updates.
Power BI	Data visualization and dashboards	User friendly interface, interactive visuals, intergrates with multiple data sources.	Limited advances analytics: less flexible than coding sloutions.	Recommended: perfect for reporting and interactive dashboards
Power BI / RapidMiner	Visual data mining workflows	No-code interface, quick to prototype models, easy to test matching or transformation logic.	Limited customization and scalability for large or complex data.	Optional: useful for early experminentatio n or testing matching algorithms.

This project we will primarily use Python for data collection, cleaning, matching and automation. Azure for structured data storage. And we use Power BI for dashboards. This combination is practical, scalable, and open-source.

Assessment of techniques

Technique	Purpose	Strengths	Weaknesses	Evaluation
ETL (extract, transform, load)	Automates collecting, cleaning and storing data.	Saves time, reduces manual work.	Needs some setup	essential: needed for automation and yearly updates.
Natural Language Processing (NLP)	Finds and compares skills from job posts and study programs	Can read and understand text automatically	Needs text cleaning to work well.	Highly recommended: very important for skill matching
Keyword matching / similarity check	Compares skills between jobs and what students learn.	Simple, quick and easy to understand.	Not as accurate as advanced methods.	Recommended: good for prototype
Viable system modelling (VSM)	Helps organize and explain the system and curriculum links	Gives a clear structure or framework	It is hard to understand and use.	Keep as theory: supports background understanding
Predictive modelling / trend analysis	Predicts future skill needs	Uses old data to make simple forecast	Not always accurate with small data	Optional: use if there is time

This project will mainly use NLP, keyword matching and ETL. These techniques help the system read text, compare skills and automate data collection and cleaning. Optional methods like predictive modelling and VSM can be used for forecasting and curriculum guidance in the future.

This assessment shows the best tools and methods for building a sustainable and fully automated Competency Monitoring System within four months.

The system will use Python for automation and data handling, Azure for data storing, and Power BI for dashboards and reports.

Main techniques include Natural language processing (NLP), keyword matching and ETL (Extract, transform, load). These make it possible to collect, clean, match and report data automatically.

Overall, this setup will deliver a working prototype that needs little manual work, gives clear visual result and forms a strong base for future growth and automation.

Sources

[Artificial intelligence in het onderwijs: dit zijn kansen en risico's - Kennisnet](#)

[Privacy in education | Privacy First](#)

[Cyberveiligheid | Onderwerp | Inspectie van het onderwijs](#)

[Meer kennis over AI in het onderwijs nodig, anders risico op discriminatie | PO-Raad](#)

https://www.google.com/aclk?sa=L&ai=DChsSEwj19Lfz3O6PAxVikIMHHfGTNVQYACICCAEQABoCZWY&co=1&gclid=CjwKCAjwisnGBhAXEiwA0zEORzyYBHWiJRQnT7l7LZE5E-sh7ZLsecORnXf-ebppw2rvrqClv5vysRoCA2wQAvD_BwE&cce=2&sig=AOD64_33JlXu4NVM0lQFJ_oyZGfE0q2tGA&q&adurl&ved=2ahUKEwj3ybDz3O6PAxV9-QIHHWhAKkMQ0Qx6BAgYEA

Bronnen SWOT-analyse:

Hogeschool Utrecht. (z.d). Marketing & Customer experience.

https://www.internationalhu.com/research/marketing-and-customer-experience?utm_source=chatgpt.com

Hogeschool Utrecht (z.d). onderwijs en onderzoek.

https://www.hu.nl/werkenbij/onderwijs-en-onderzoek?utm_source=chatgpt.com

Hogeschool Utrecht (z.d). creative business.

https://www.hu.nl/voltijd-opleidingen/creative-business/during-the-programme?utm_source=chatgpt.com

Hogeschool Utrecht (z.d). Studiegids.

https://studiegids.hu.nl/028c0c1c-9805-4519-93a3-a91528dbdbfed?utm_source=chatgpt.com

Bronnen vijf porter's krachtenmodel:

Mckinsey & Company. (z.d). Education technology: five trends to watch in the EdTech industry.

https://www.mckinsey.com/industries/education/our-insights/five-trends-to-watch-in-the-edtech-industry?utm_source=chatgpt.com

Investopedia.(z.d). Porter's five forces explained and how to use the model.

https://www.investopedia.com/terms/p/porter.asp?utm_source=chatgpt.com

Sanghvi, S, & Westhoff, M (2022, 14 november). Five trends to watch in the edtech.

https://www.mckinsey.com/industries/education/our-insights/five-trends-to-watch-in-the-edtech-industry?utm_source=chatgpt.com

Research and markets. (z.d). EdTech market size competitors, trends & forecast to 2029.

https://www.researchandmarkets.com/report/education-software?srsId=AfmBOoplRwE5Azlr1Qp0buVrKTMu27C5AsUIBT1Jg91_g3JCcTsvFUXl&utm_source=chatgpt.com

GlobalData. (z.d). Netherlands EdTech market summary, analysis & forecast 2028.

https://www.globaldata.com/store/report/netherlands-edtech-market-analysis/?utm_source=chatgpt.com

Jermore Knoot. (z.d). Strategische Marketingplan.

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwilmrPjnu2PAxV09LsIHfm7DWYQFnoECBcQAQ&url=https%3A%2F%2Fwww.strategischmarketingplan.com%2Fmarketingmodellen%2Fvijf-krachten-model-porter%2F&usg=AOvVaw0Ly1QpOSovIT2OoJEFOnQl&opi=89978449>

Hogeschool Utrecht (z.d.). *Visie op onderzoek*. Geraadpleegd via <https://www.hu.nl/onderzoek/visie-op-onderzoek>

Agile Scrum Group (z.d.). *Stakeholder management model (Salience Model)*. Geraadpleegd via <https://agilescrumgroup.nl/stakeholder-management-model/>

Master Challenge (z.d.). *Challenge space registratie*. Geraadpleegd via <https://masterchallenge.me/account/challenge-spaces/f0517013-d2fe-411a-b9f9-23235e57d3ec/registration>