



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	09-10-2025	Jason	Started on the first steps of the document. With Lay-out. Also started working on the Collect initial data

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

Version Management	2
Introduction	3
Table of Contents	4
Collect Initial Data	5
Describe Data	6
Explore Data	7
Verify Data Quality	8



Collect Initial Data

Doel van de dataverzameling

Het doel van deze stap is om data te verzamelen over:

1. Die competenties en soft skills die worden aangeboden in de eind curriculum van de hoge school van Utrecht.
2. De vaardigheden en soft skills die gevraagd worden in actuele vacatures op LinkedIn voor relevante banen

Deze data zal later gebruikt worden om te analyseren in hoeverre de vaardigheden uit de opleidingen aansluiten op de top 100 banen van Nederland op LinkedIn

What are our datasets?

Datasets	Source	Type/ Format	Location	Purpose
Curriculum of the HU	Website of the Hogeschool of Utrecht	CSV / Excel (After extraction)		Holds information about Curriculum, description, softskills and competencies
LinkedIn Vacancies	LinkedIn (Webscraping API)	Excel (Python - export)		Includes top 100 vacancies matched with curriculum

These are the datasets we have collected and can work with.

How are we going to collect it?

HU Curriculum data

- Method: Collecting through Python-script web scraping
- Contents: Curriculum-name: Curriculum-URL, Description, Soft skills/competencies.
- Export: Export to Excel File

LinkedIn Vacancies data

- Method: Python-script that scrapes vacancies matching with the curriculum name
- Process:
 - o Searches relevant jobs matching with curriculum name
 - o Selects top 100 vacancies
 - o Extracts soft skills and competencies out of vacancy
- Python tools: Requests, BeautifulSoup & Pandas
- Export: Exports to Excel File

Problems & Solutions

While using these data we can encounter different kind of problems. To make it clear what we can encounter we put the problems into a table with 2 different rows. One marking **Problem** and the second marking **The solution**.

Problem	Solution
LinkedIn limited access or needs an API	Using code to slow down the process so LinkedIn doesn't block us while trying to attempt to collect data. And if LinkedIn completely blocks us we are going to scrape Indeed instead of LinkedIn
Hu-Curriculum pages have unstructured texts	Making a program that runs text analysis and can find keywords on the website
Inconsistency of keywords on the website	Making a mapping-table

Describe Data

Explore Data

Verify Data Quality