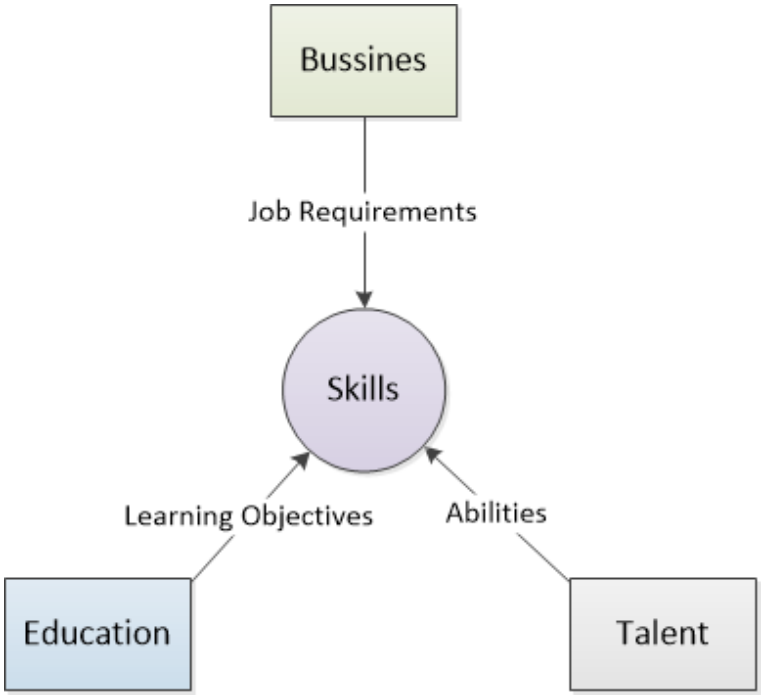
Skill Scanner CV Report - Data Scientist

Skill Scanner uses state of the art Natural Language Processing techniques to compare skills:

* We compared your CV to employer demands for Data Scientists
* We meassured how well your CV covers the skills of peer Data Scientists
* We reccomend courses that fit the skill gaps in your CV

For an in depth explanation of our technique please refer to the last page.



Skill Scanner uses AI to extract and compare skills from three sources

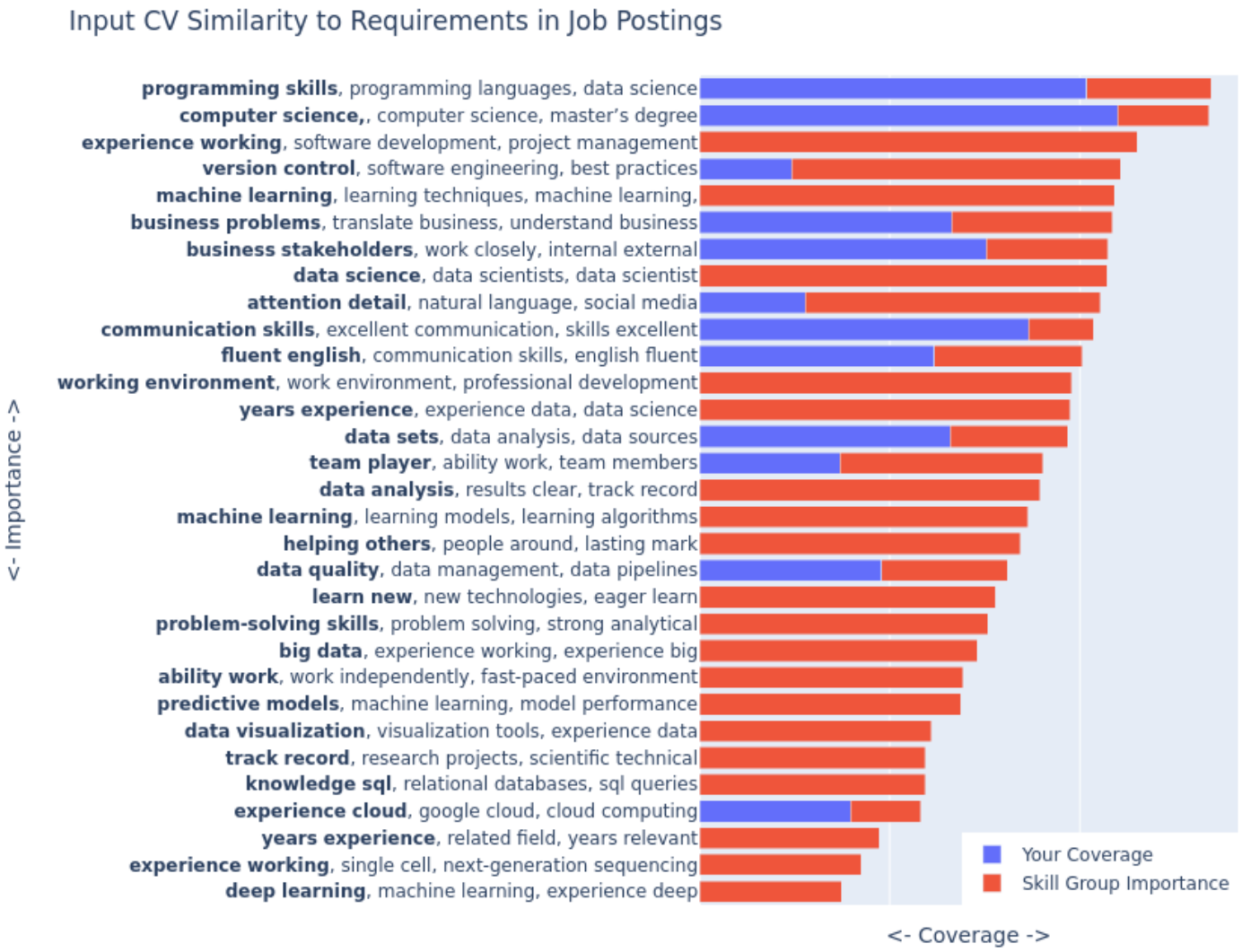
Your CV Coverage Score: 27.98%

Your total score is 27.98 This is below average for a Data Scientist CV:

Your CV Coverage [%]	Data Scientist Top 10% [%]	Data Scientist Top 25% [%]	Data Scientist Top 50% [%]	Data Scientist Average [%]
27.98	49	72	67	61

Comparison Plot

Skill Scanner has extracted the 30 most important skill groups as required in Data Scientist job postings.
Below you find how well your CV covers these skill groups.



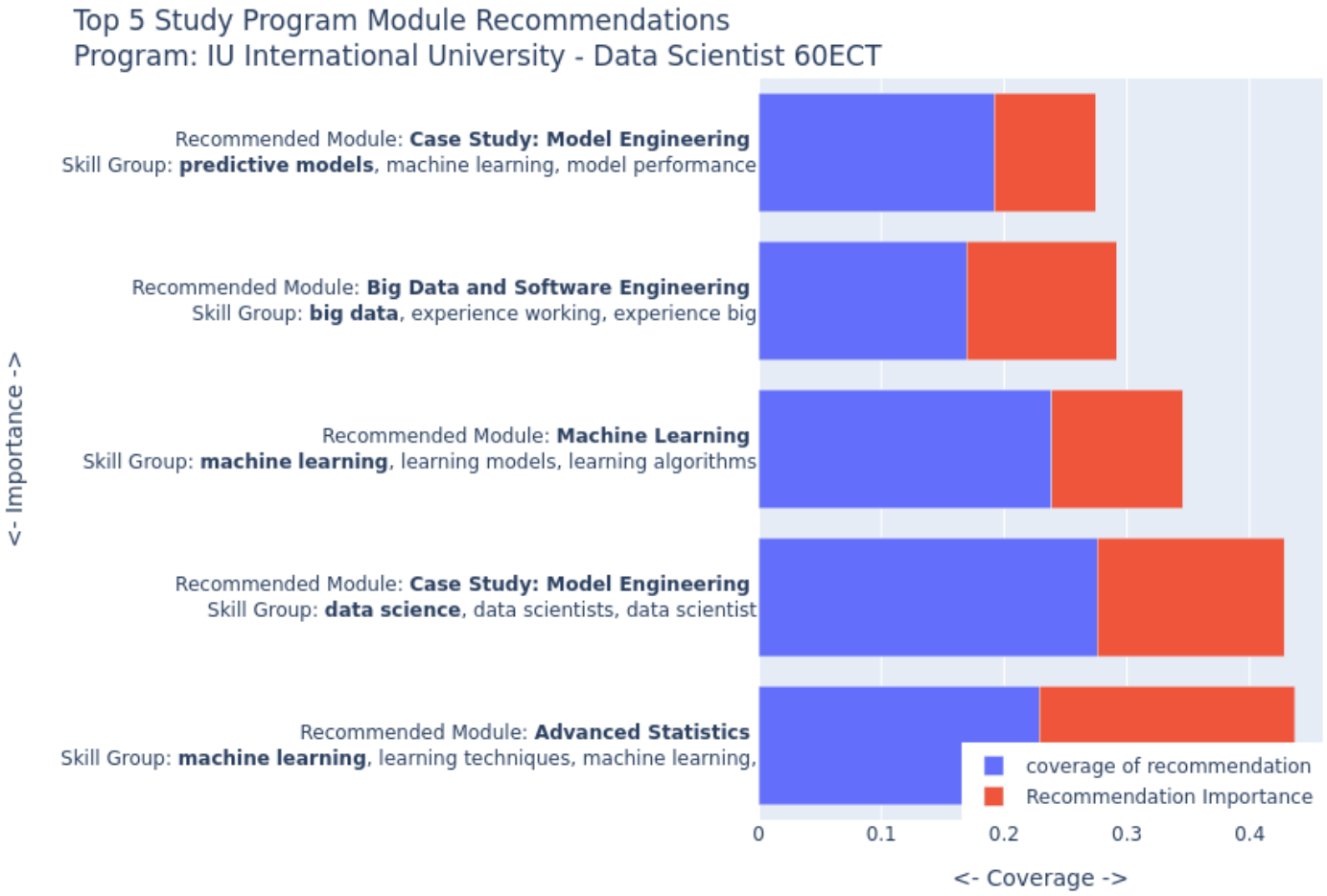
*** Please Note: A coverage of 100% is impossible to attain, a coverage of over 70% can be considered excellent.*

Learning Program Recommendation

Are you considering formal education to upskill yourself? Skill Scanner helps you select the most appropriate programs and modules:

* we analyzed the curriculum of course "MSc - Data Science 60ECT" of IU International University of Applied Sciences.
* We analyzed which gaps in you CV can be filled by which modules
* We show the top 5 of module recommendations in the visual below.

** In the future more learning content will be analyzed so that Skill Scanner can help choose a fitting study program



Your score compared to Data Scientist CV's

The table below shows an analysis of how you compare to other Data Scientists.

Skill Cluster	Your Score	Average	Top 10%	Top 25%	Top 50%
team player, ability work, team members	0.41	0.61	0.74	0.74	0.73
knowledge sql, relational databases, sql queries	0	0.55	0.7	0.68	0.64
ability work, work independently, fast-paced environment	0	0.48	0.62	0.53	0.5
years experience, related field, years relevant	0	0.61	null	null	0.68
learn new, new technologies, eager learn	0	0.63	0.79	0.79	0.74
track record, research projects, scientific technical	0	0.45	null	0.55	0.51
machine learning, learning models, learning algorithms	0	0.53	0.68	0.65	0.61
communication skills, excellent communication, skills excellent	0.83	0.73	0.85	0.83	0.78
big data, experience working, experience big	0	0.57	0.76	0.71	0.66
attention detail, natural language, social media	0.26	0.21	0.41	0.35	0.3
experience working, software development, project management	0	0.57	0.71	0.67	0.63
data sets, data analysis, data sources	0.68	0.51	0.69	0.64	0.59
machine learning, learning techniques, machine learning,	0	0.65	0.77	0.75	0.72
data science, data scientists, data scientist	0	0.62	0.73	0.71	0.68
data visualization, visualization tools, experience data	0	0.6	0.78	0.75	0.7
years experience, experience data, data science	0	0.7	0.82	0.79	0.76
business stakeholders, work closely, internal external	0.7	0.57	0.71	0.68	0.64
problem-solving skills, problem solving, strong analytical	0	0.7	null	null	0.74
programming skills, programming languages, data science	0.76	0.57	0.72	0.69	0.65
predictive models, machine learning, model performance	0	0.5	0.68	0.64	0.58
experience working, single cell, next-generation sequencing	0	0.44	0.62	0.58	0.54
helping others, people around, lasting mark	0	0.5	null	0.68	0.55
version control, software engineering, best practices	0.22	0.45	0.63	0.59	0.54
data analysis, results clear, track record	0	0.46	0.65	0.6	0.54
deep learning, machine learning, experience deep	0	0.52	0.72	0.68	0.62
business problems, translate business, understand business	0.61	0.51	0.68	0.64	0.59
working environment, work environment, professional development	0	0.51	0.67	0.62	0.59
experience cloud, google cloud, cloud computing	0.68	0.54	0.73	0.67	0.62
fluent english, communication skills, english fluent	0.61	0.61	null	null	0.61
data quality, data management, data pipelines	0.59	0.55	0.72	0.67	0.63
computer science,, computer science, master's degree	0.82	0.59	0.69	0.68	0.66

Appendix A: Analysis per input skill

Please find below the result of the analysis of each of your input skills.

Input Skill 1: "Proficient in Python and R"
Was clustered in skill group "programming skills, programming languages, data science". Your similarity score for this skill is 0.76. this is above the average score among Data Scientist CV's which is 0.57

Input Skill 2: "Creating data pipelines "
Was clustered in skill group "data quality, data management, data pipelines". Your similarity score for this skill is 0.59. this is above the average score among Data Scientist CV's which is 0.55

Input Skill 3: "app development using flask"
Was clustered in skill group "attention detail, natural language, social media". Your similarity score for this skill is 0.16.This score is quite low, the average score among Data Scientist CV's is 0.21 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Input Skill 4: "excellent communication skills"
Was clustered in skill group "communication skills, excellent communication, skills excellent". Your similarity score for this skill is 0.83. this is above the average score among Data Scientist CV's which is 0.73

Input Skill 5: "english and dutch"
Was clustered in skill group "fluent english, communication skills, english fluent". Your similarity score for this skill is 0.61. this is above the average score among Data Scientist CV's which is 0.61

Input Skill 6: "have saved my company money by data driven approach"
Was clustered in skill group "business problems, translate business, understand business". Your similarity score for this skill is 0.61. this is above the average score among Data Scientist CV's which is 0.51

Input Skill 7: "Data Analysis "
Was clustered in skill group "data sets, data analysis, data sources". Your similarity score for this skill is 0.68. this is above the average score among Data Scientist CV's which is 0.51

Input Skill 8: "Power BI (Expert)"
Was clustered in skill group "data sets, data analysis, data sources". Your similarity score for this skill is 0.43.This score is quite low, the average score among Data Scientist CV's is 0.51 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Input Skill 9: "Master's degree in Data Science"
Was clustered in skill group "computer science,, computer science, master?s degree". Your similarity score for this skill is 0.82. this is above the average score among Data Scientist CV's which is 0.59

Input Skill 10: "version control with git and github"
Was clustered in skill group "version control, software engineering, best practices". Your similarity score for this skill is 0.22.This score is quite low, the average score among Data Scientist CV's is 0.45 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Input Skill 11: "Presentation skills to present my work"
Was clustered in skill group "communication skills, excellent communication, skills excellent". Your similarity score for this skill is 0.57.This score is quite low, the average score among Data Scientist CV's is 0.73 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Input Skill 12: "Working together with various stakeholders to maximize value"
Was clustered in skill group "business stakeholders, work closely, internal external". Your similarity score for this skill is 0.7. this is

above the average score among Data Scientist CV's which is 0.57

Input Skill 13: "3 scientific publications in the field of natural language processing"

Was clustered in skill group "attention detail, natural language, social media". Your similarity score for this skill is 0.26. this is above the average score among Data Scientist CV's which is 0.21

Input Skill 14: "Experience with Google Cloud Platform"

Was clustered in skill group "experience cloud, google cloud, cloud computing". Your similarity score for this skill is 0.68. this is above the average score among Data Scientist CV's which is 0.54

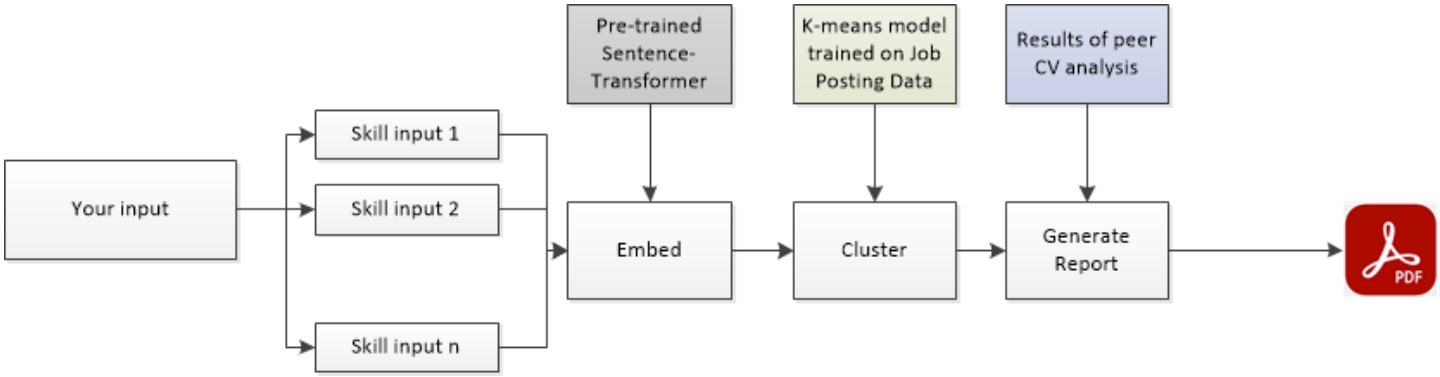
Input Skill 15: "Deployed various web-apps using Docker"

Was clustered in skill group "experience cloud, google cloud, cloud computing". Your similarity score for this skill is 0.32.This score is quite low, the average score among Data Scientist CV's is 0.54 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Input Skill 16: "Acted as a mentor for several junior colleagues"

Was clustered in skill group "team player, ability work, team members". Your similarity score for this skill is 0.41.This score is quite low, the average score among Data Scientist CV's is 0.61 this may be due to a misclassification of our model but this could also indicate an opportunity to further clarify your CV.

Appendix B: Methodology Explanation



Functionality sketch of Skill Scanner backend

Skill Scanner processes your input skills in three steps:
1. Embedding: We use Sentence-Transformers, a state-of-the-art framework for sentence embeddings. In simple terms, a sentence embedding algorithm turns a sentence into a series of numbers from which a computer can infer differences and similarities.

2. K-means modelling: The sentence embeddings from step 1 enable us to compute the similarity of various embeddings. A K-means model was used to cluster skills in groups of similar meaning.

2.1 Training of K-means model: The training data is a dataset of 21.500 job requirements extracted from Data Scientist vacancy retrieved from various sources.

2.2 Evaluation of K-means model: To evaluate how the model generalizes to different data sources we inferred clusters from skillsets found in a sample of Data Scientist CV's. The model was able to infer the correct cluster with an accuracy of 82%.

2.3 Use of the K-means model by Skill Scanner: Skill Scanner uses the K-means model to infer clusters from you input data. It uses cosine similarity to compute the distance from the cluster centroid. This metric is reported as similarity score.

3. Peer CV Analysis: We used Skill Scanner to analyze a dataset of 65 Data Scientist CV's retrieved from Kaggle.com. The results of this analysis allows us to compare your scores to these of your peers (other Data Scientists).