SILICON SURFER

iu INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Learning Ideas Conference 2022

KOEN BOTHMER & TIM SCHLIPPE

SKILL SCANNER: CONNECTING AND SUPPORTING EMPLOYERS, JOB SEEKERS AND EDUCATIONAL INSTITUTIONS WITH AN AI-BASED RECOMMENDATION SYSTEM

New York, USA
June 17, 2022

AGENDA

MOTIVATION

MOTIVATION: UN Sustainable Development Goal 4

Image Sources: United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021); OpenClipart-Vectors/154119/Pixabay.

MOTIVATION: UN Sustainable Development Goal 4

JOB MARKET

FIT DEMAND

Image Sources: United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021); OpenClipart-Vectors/154119/Pixabay.
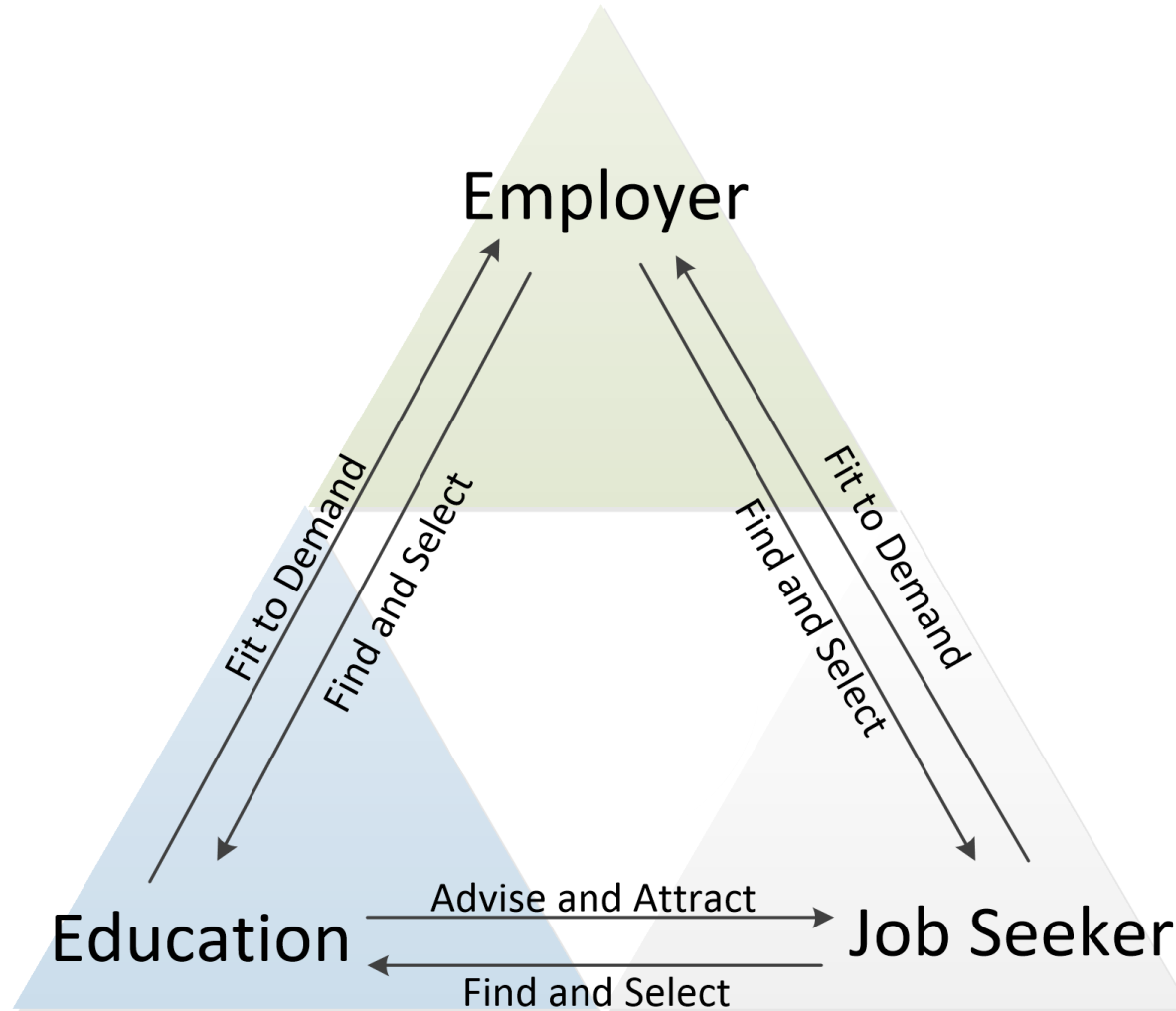
RELATED WORK

SKILL SKANNER: Related Work
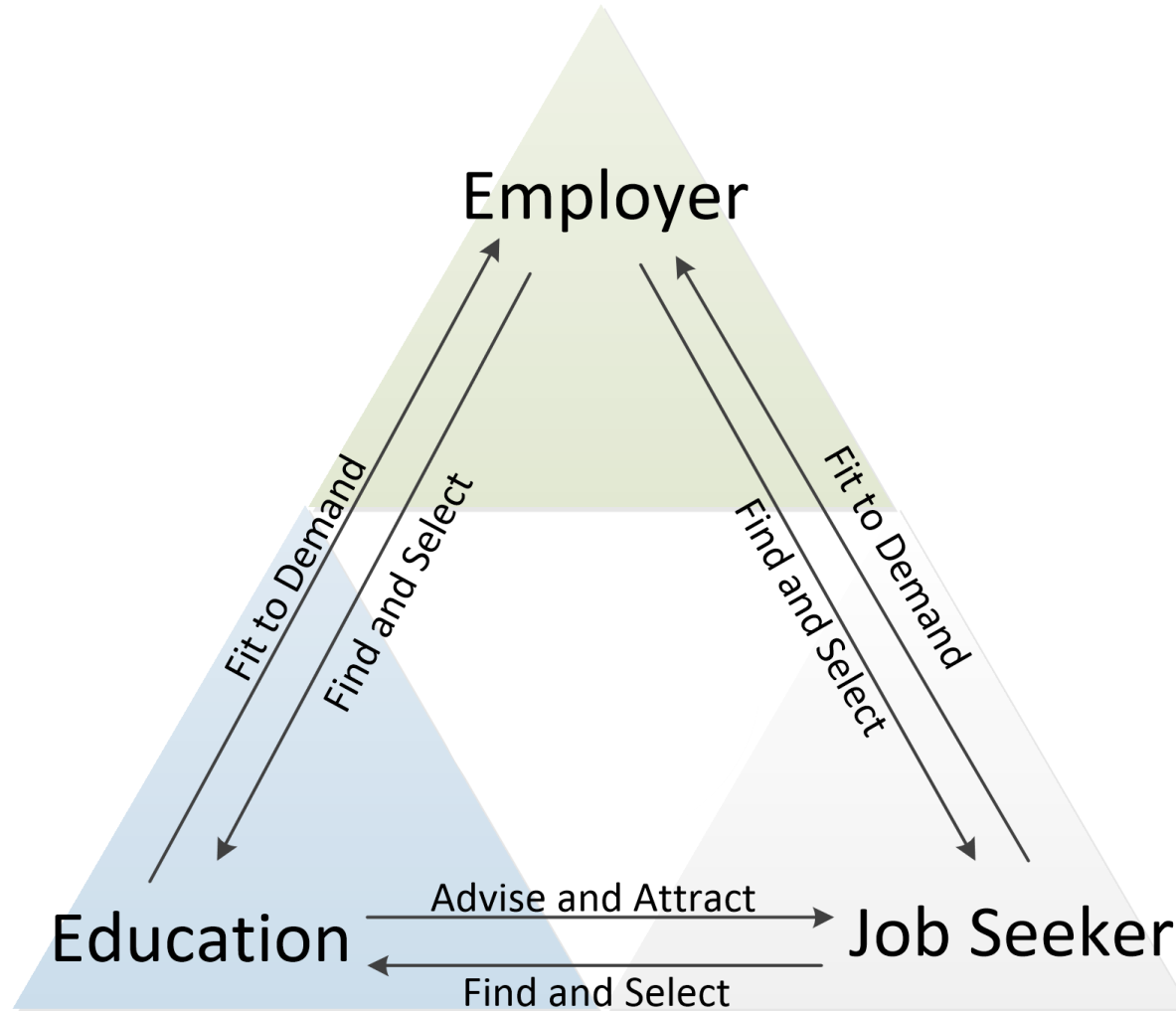
SKILL SKANNER: Related Work
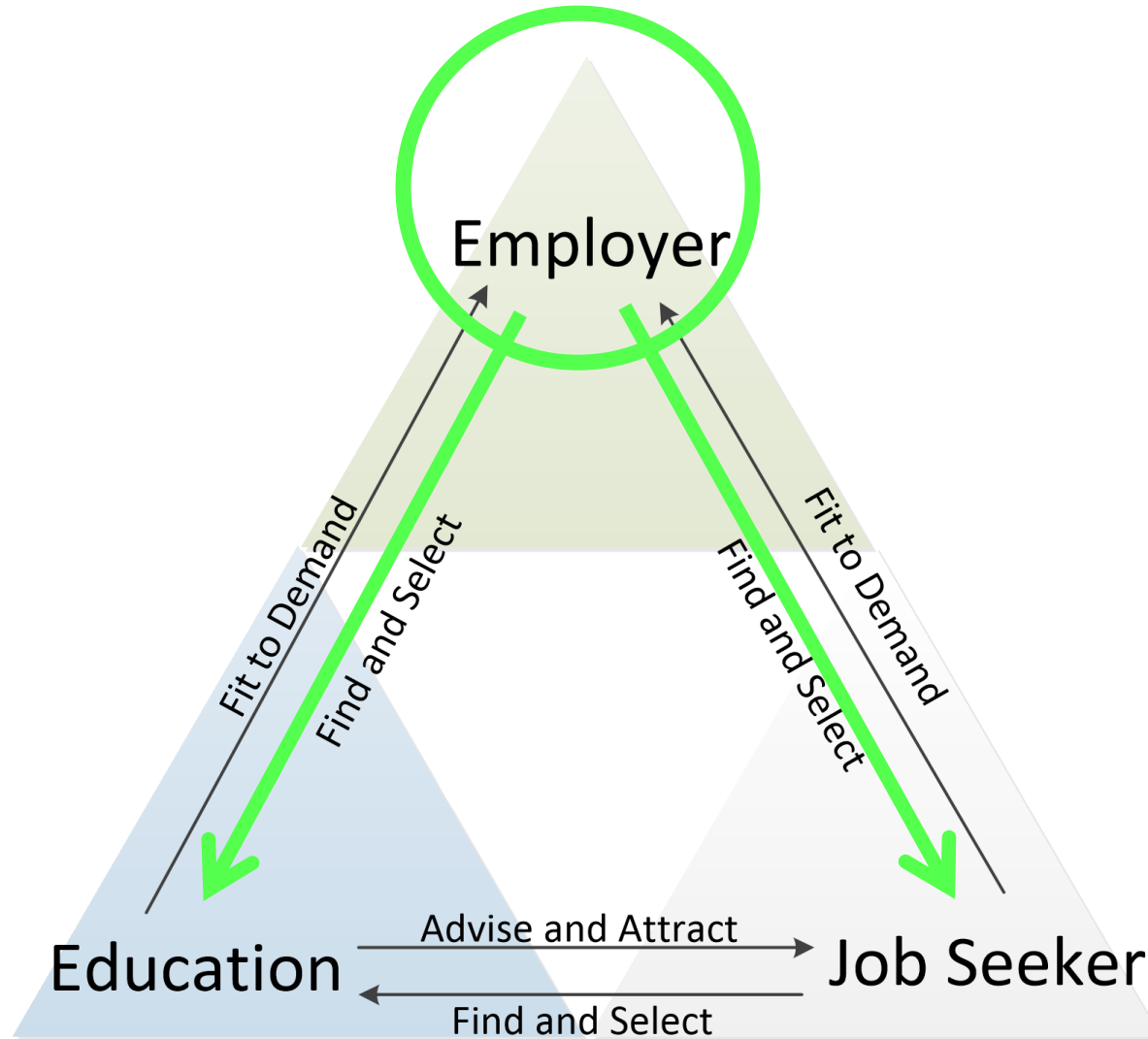
SKILL SKANNER: Related Work
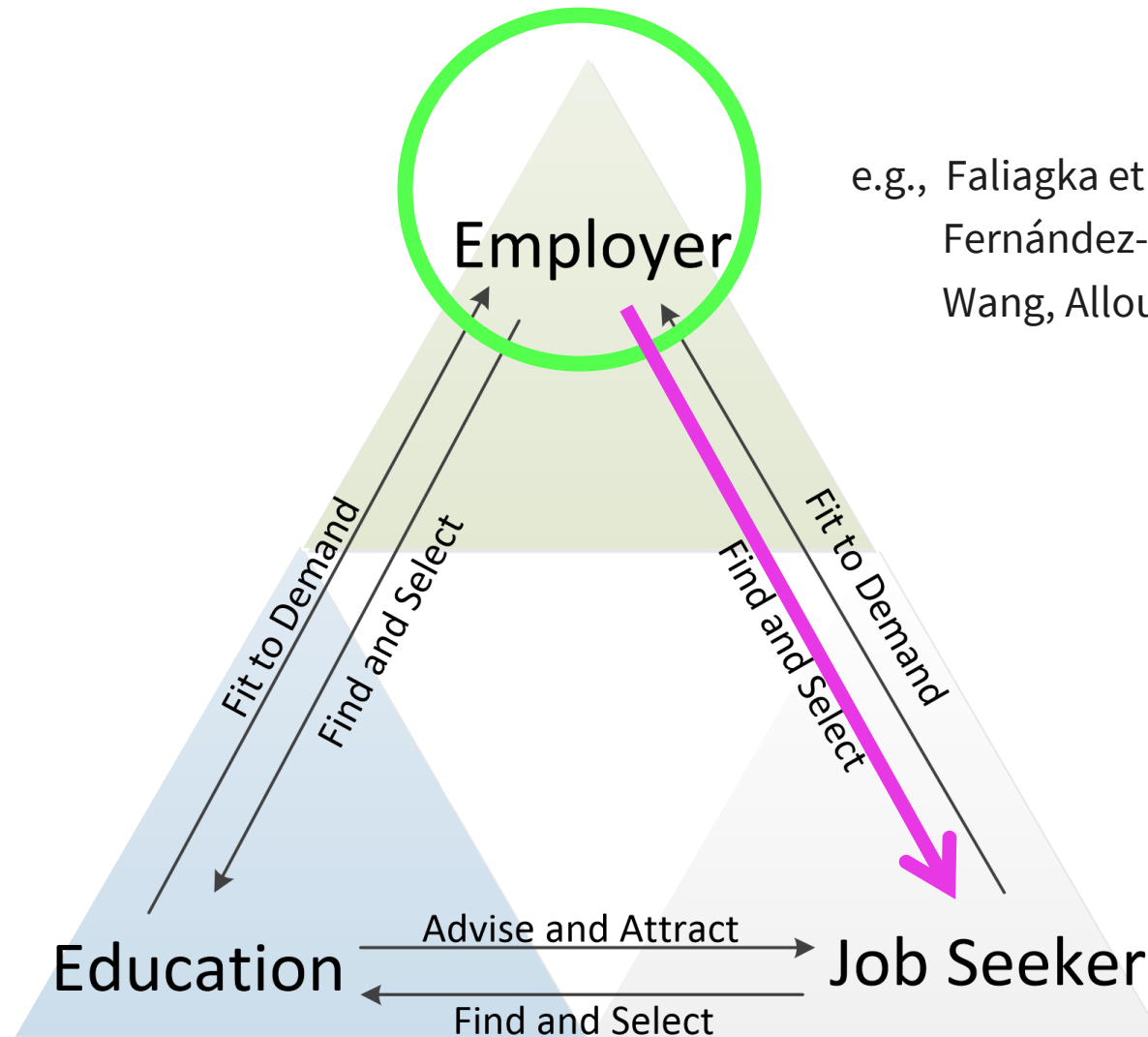
e.g., Faliagka et al. (2014)   `TAXONOMY`

Fernández-Reyes & Shinde (2019)   `WORD2VEC`

Wang, Allouache & Joubert (2021)   `BERT`

SKILL SKANNER: Related Work
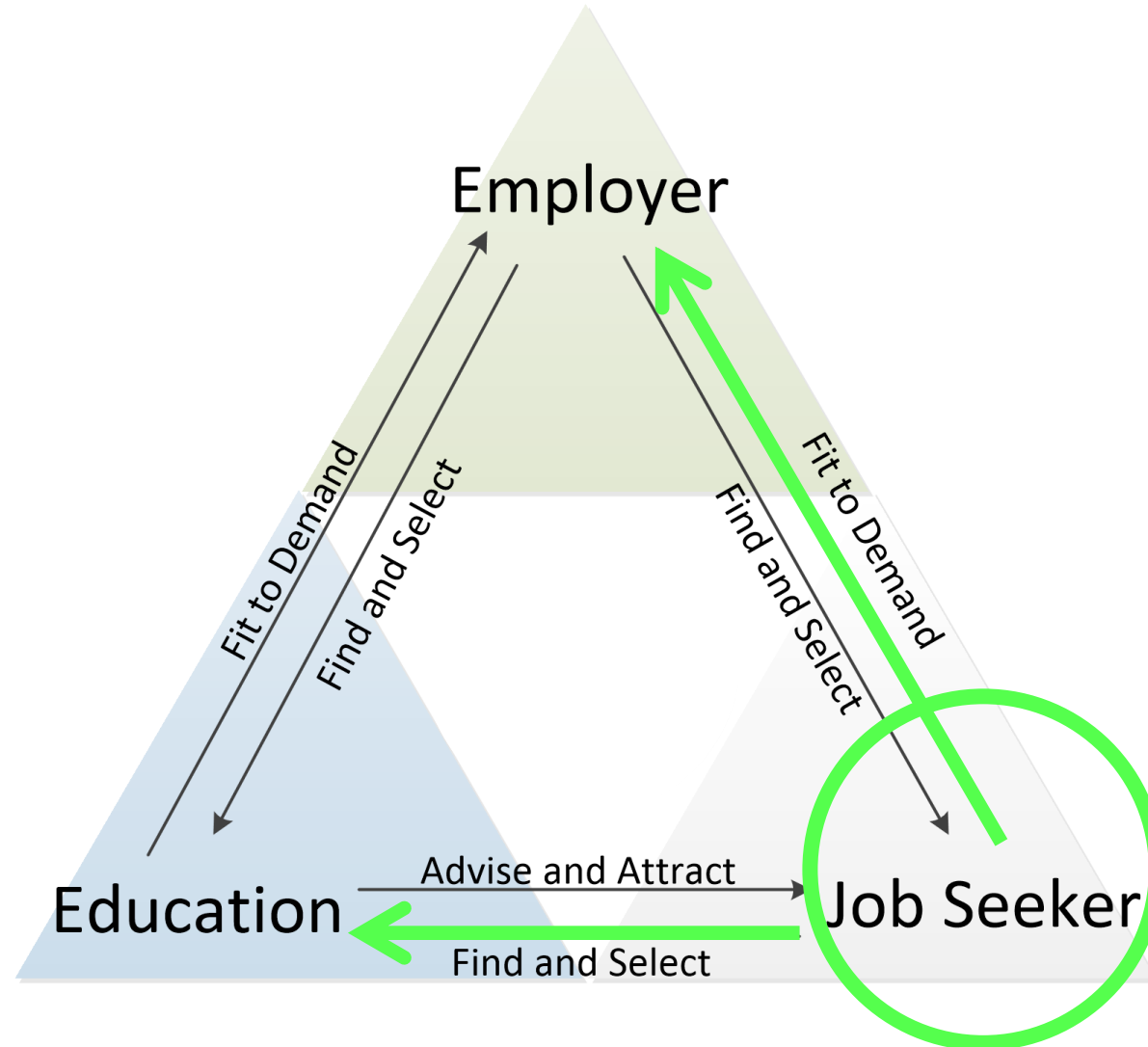
SKILL SKANNER: Related Work
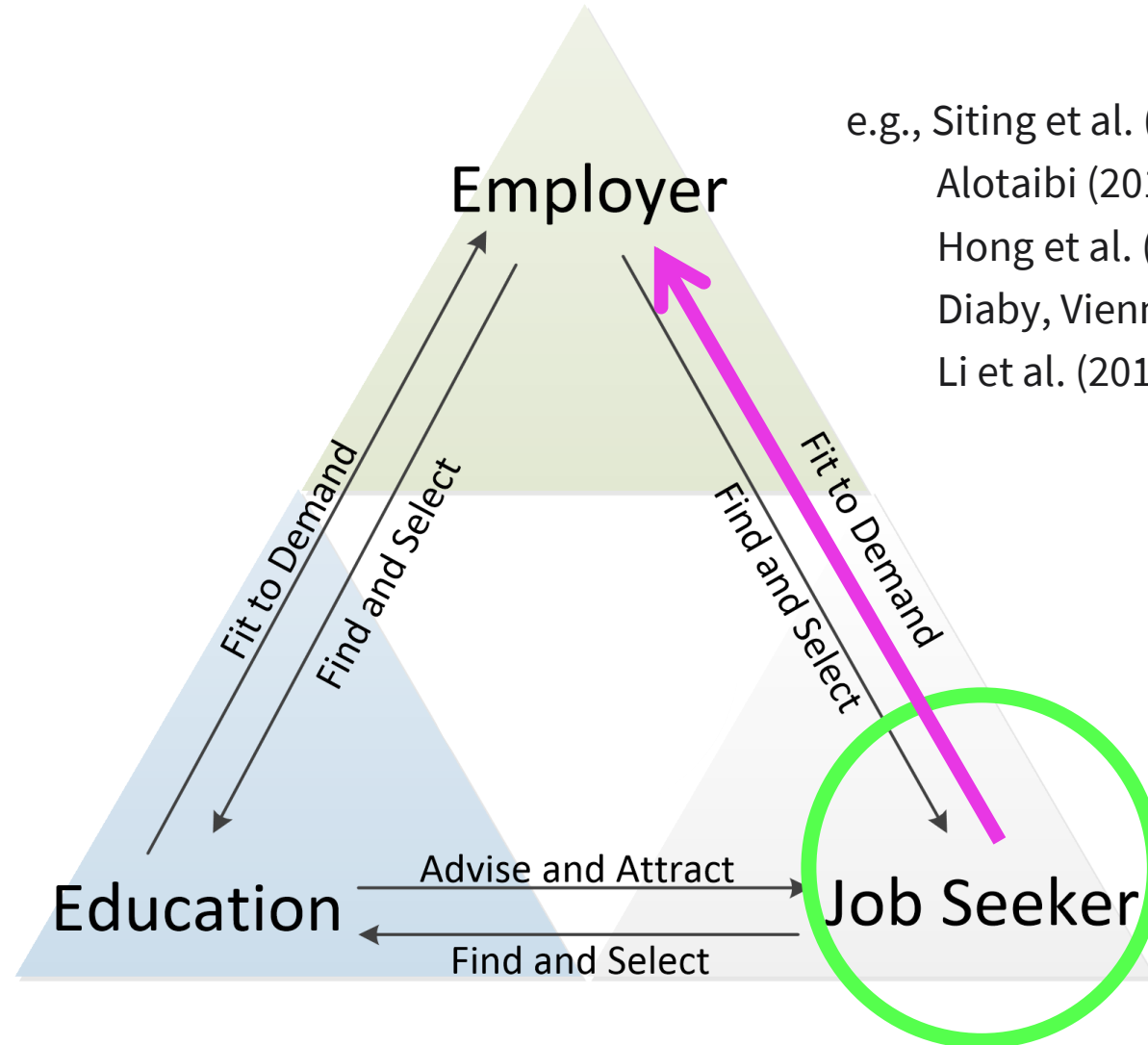
e.g., Siting et al. (2012)
   Alotaibi (2012)
   Hong et al. (2013)
   Diaby, Viennet & Launay (2013)
   Li et al. (2016)

SURVEY PAPERS

CLUSTERING

TAXONOMY

SKILL SKANNER: Related Work
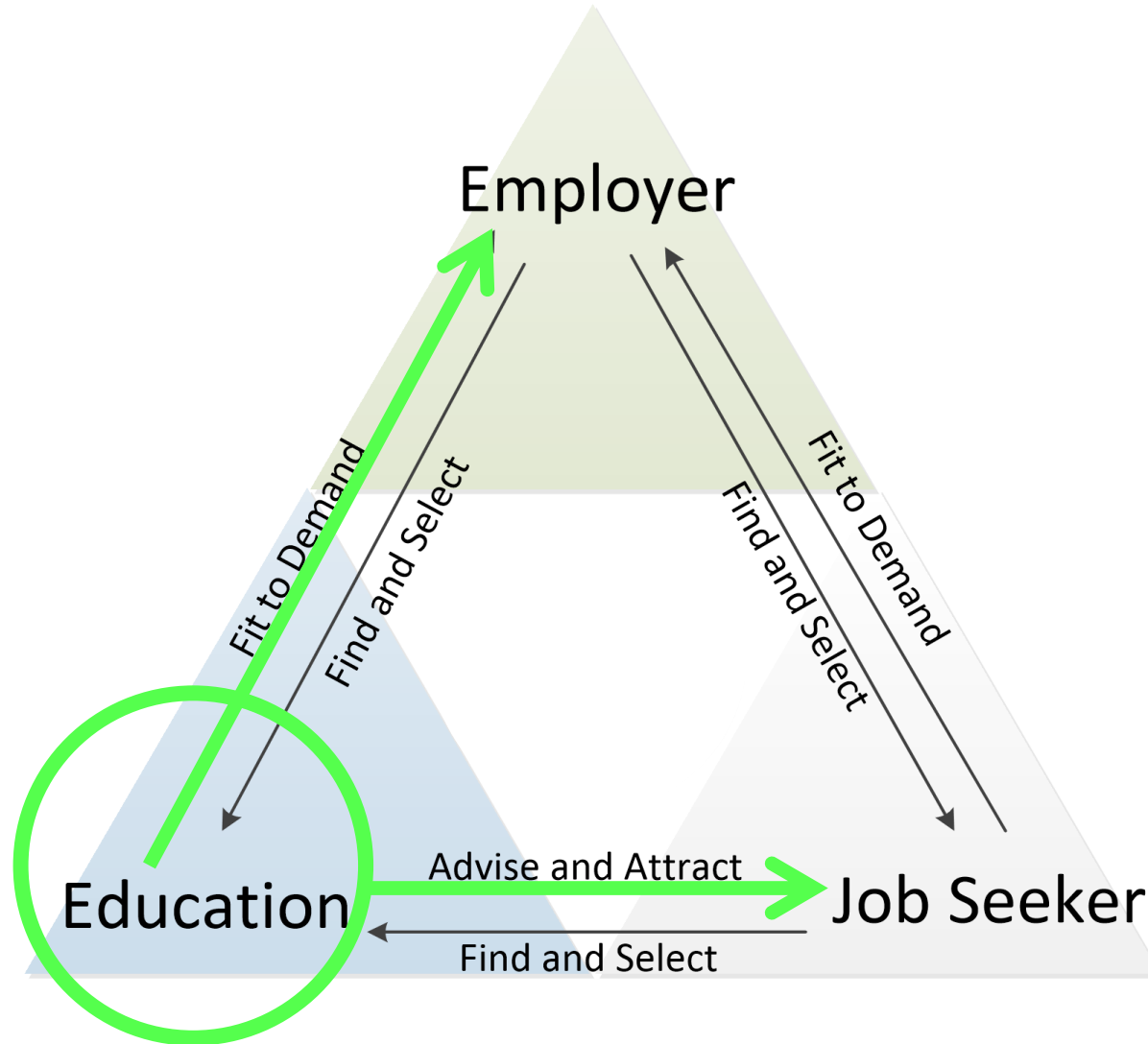
SKILL SKANNER: Related Work

e.g., Guruge, Kadel & Halder (2021)   CLUSTERING

Employer

Fit to Demand

Find and Select

Fit to Demand

Find and Select

Education

Advise and Attract

Find and Select

Job Seeker

SKILL SKANNER: Related Work

3

SKILL SCANNER:
EXTRACTING, VECTORIZING, CLUSTERING
AND COMPARING SKILLS

SKILL SKANNER: Extract & Compare Skills

Data from employer's perspective: Job postings	Data from job seeker's perspective: CV	Data from educational institution's perspective: Learning Curriculum

Data Scientist

Solliciteer ♡

What You'll Do:

- Design, develop and test data science pipelines for a variety of projects to help make informed decisions impacting the business.
- Producing reliable predictive insights, based on statistical modeling and/or machine learning methodologies.
- Approaching data with the aim to increase and maximize performance marketing KPIs such as retention acquisition.
- Analyzing internal datasets, to better understand our customer behavior to support business decisions.
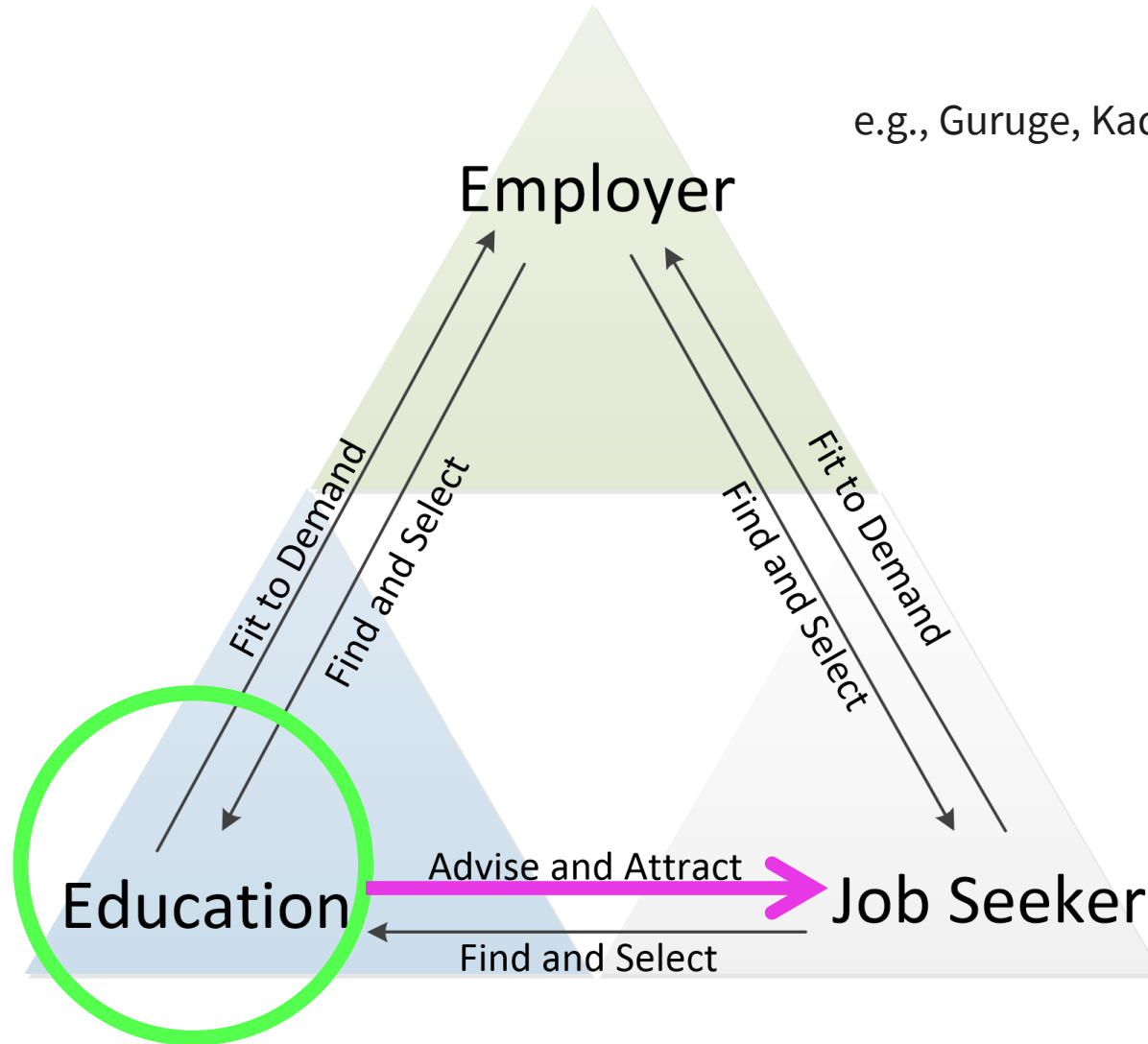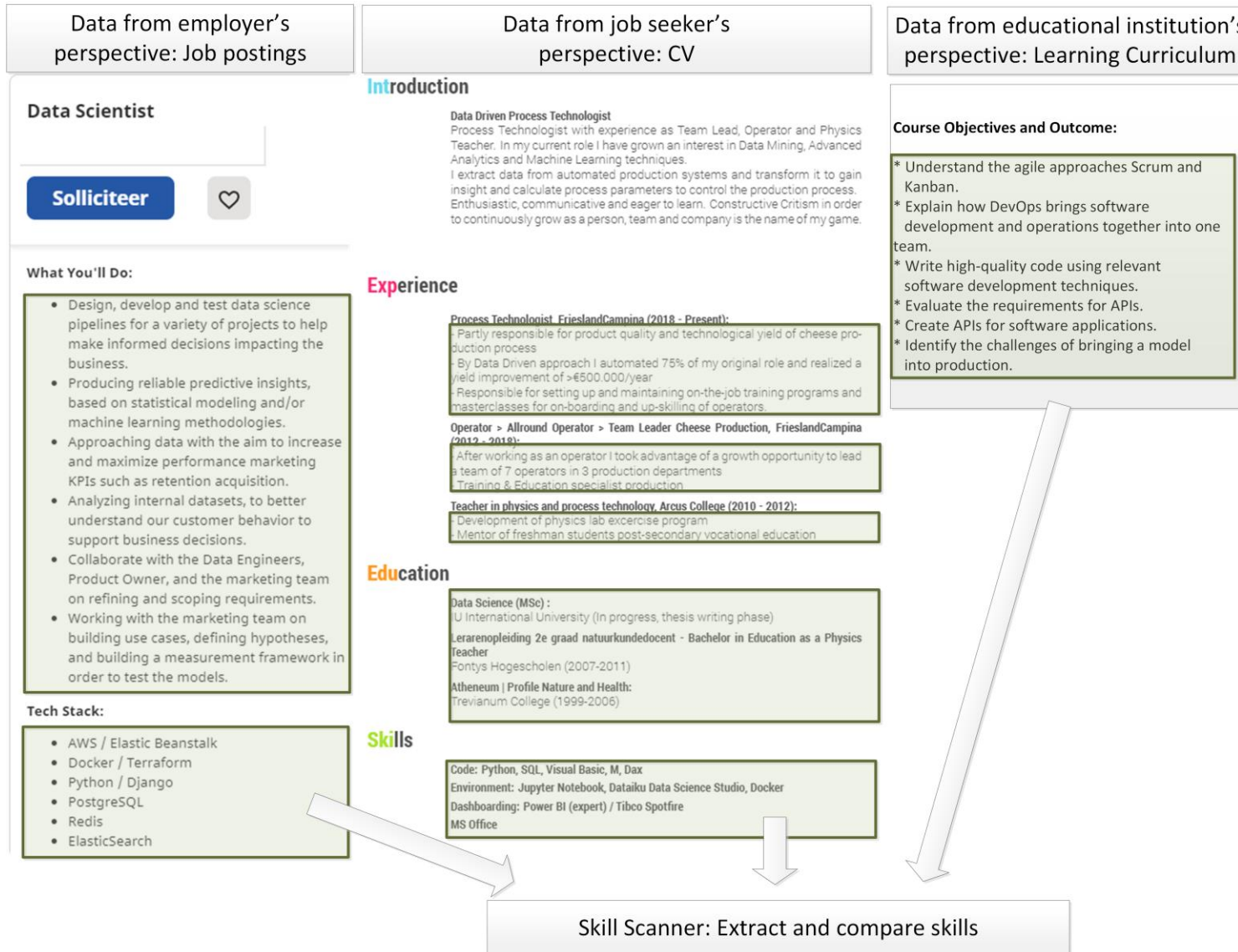- Collaborate with the Data Engineers, Product Owner, and the marketing team on refining and scoping requirements.
- Working with the marketing team on building use cases, defining hypotheses, and building a measurement framework in order to test the models.

Tech Stack:

- AWS / Elastic Beanstalk
- Docker / Terraform
- Python / Django
- PostgreSQL
- Redis
- ElasticSearch

Introduction

Data Driven Process Technologist
Process Technologist with experience as Team Lead, Operator and Physics Teacher. In my current role I have grown an interest in Data Mining, Advanced Analytics and Machine Learning techniques.
I extract data from automated production systems and transform it to gain insight and calculate process parameters to control the production process. Enthusiastic, communicative and eager to learn. Constructive Critism in order to continuously grow as a person, team and company is the name of my game.

Experience

Process Technologist, FrieslandCampina (2018 - Present):
- Partly responsible for product quality and technological yield of cheese production process
- By Data Driven approach I automated 75% of my original role and realized a yield improvement of >€500.000/year
- Responsible for setting up and maintaining on-the-job training programs and masterclasses for on-boarding and up-skilling of operators.

Operator > Allround Operator > Team Leader Cheese Production, FrieslandCampina (2012 - 2018):
- After working as an operator I took advantage of a growth opportunity to lead a team of 7 operators in 3 production departments
- Training & Education specialist production

Teacher in physics and process technology, Arcus College (2010 - 2012):
- Development of physics lab excercise program
- Mentor of freshman students post-secondary vocational education

Education

Data Science (MSc) :
IU International University (In progress, thesis writing phase)

Lerarenopleiding 2e graad natuurkundedocent - Bachelor in Education as a Physics Teacher
Fontys Hogescholen (2007-2011)

Atheneum | Profile Nature and Health:
Trevianum College (1999-2006)

Skills

Code: Python, SQL, Visual Basic, M, Dax
Environment: Jupyter Notebook, Dataiku Data Science Studio, Docker
Dashboarding: Power BI (expert) / Tibco Spotfire
MS Office

Course Objectives and Outcome:

* Understand the agile approaches Scrum and Kanban.
* Explain how DevOps brings software development and operations together into one team.
* Write high-quality code using relevant software development techniques.
* Evaluate the requirements for APIs.
* Create APIs for software applications.
* Identify the challenges of bringing a model into production.

Skill Scanner: Extract and compare skills

SKILL SKANNER: Extract & Compare Skills

Data from employer's perspective: Job postings

Data from job seeker's perspective: CV

Data from educational institution's perspective: Learning Curriculum

Skill Scanner: Extract and compare skills

Job Postings

Learning Curricula

CVs

SKILL SKANNER: Extract & Compare Skills

Retrieving Skill Sets → Pre-Processing Skill Sets → Removing Outliers → Clustering Skill Sets → Application: Skill Scanner

SKILL SKANNER: Extract & Compare Skills

Retrieving Skill Sets	→	Pre-Processing Skill Sets	→	Removing Outliers	→	Clustering Skill Sets	→	Application: Skill Scanner

— **2.6k** online **data scientist job postings** collected

— **21.5k** extracted **bullet points** likely to be skills

Excerpt from job posting

What You'll Do:

- Design, develop and test data science pipelines for a variety of projects to help make informed decisions impacting the business.
- Producing reliable predictive insights, based on statistical modeling and/or machine learning methodologies.
- Approaching data with the aim to increase and maximize performance marketing KPIs such as retention acquisition.
- Analyzing internal datasets, to better understand our customer behavior to support business decisions.
- Collaborate with the Data Engineers, Product Owner, and the marketing team on refining and scoping requirements.
- Working with the marketing team on building use cases, defining hypotheses, and building a measurement framework in order to test the models.

Tech Stack:

- AWS / Elastic Beanstalk
- Docker / Terraform
- Python / Django
- PostgreSQL
- Redis
- ElasticSearch

SKILL SKANNER: Extract & Compare Skills

| Retrieving Skill Sets | → | Pre-Processing Skill Sets | → | Removing Outliers | → | Clustering Skill Sets | → | Application: Skill Scanner |

Bag-of-Words
Word2Vec
GloVe
BERT
S-BERT

— **Vectorize skills**: various methods evaluated

— **Winner**: **Sentence-BERT**

(Reimers & Gurevych, 2020)

with pre-trained model

'*all_distilroberta_v1*'

● You are skilled in the communication with stakeholders

● At least 2 years of relevant experience coding in Python and SQL

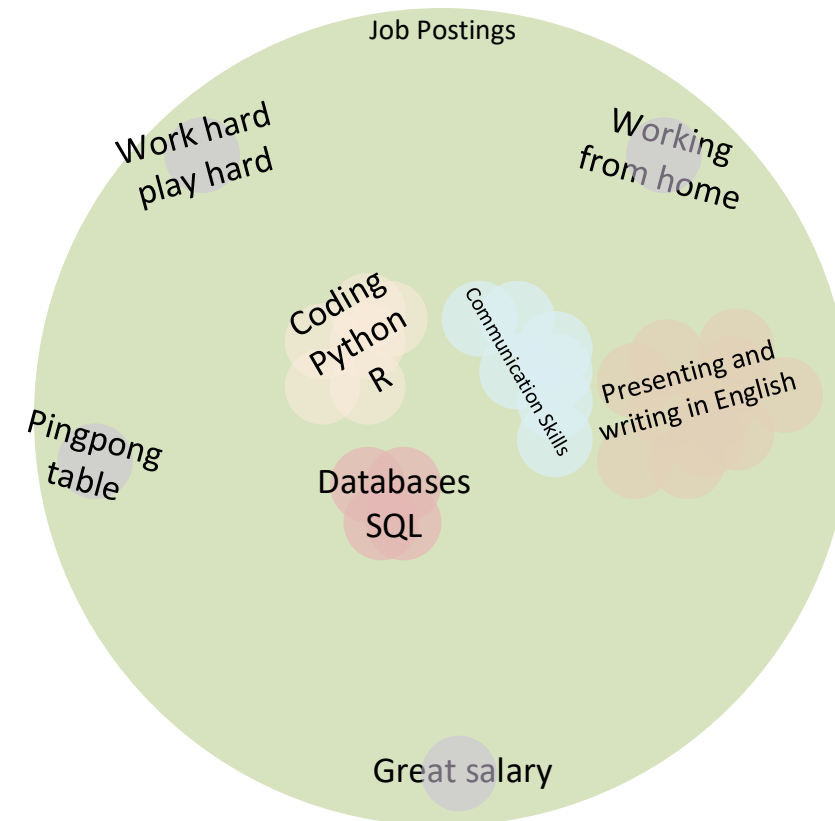● You have scripting experience with Python, R and SQL
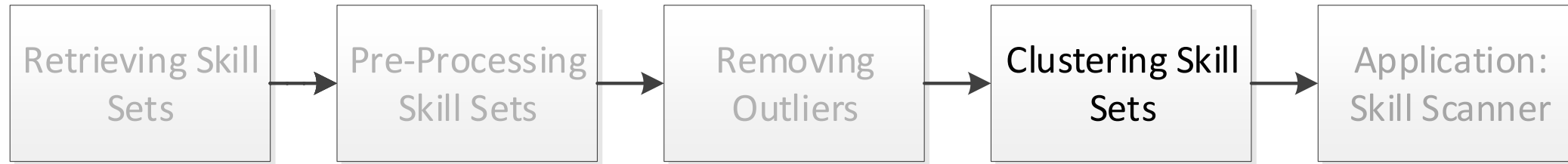
SKILL SKANNER: Extract & Compare Skills

| Retrieving Skill Sets | → | Pre-Processing Skill Sets | → | Removing Outliers | → | Clustering Skill Sets | → | Application: Skill Scanner |

Pre-Processing Skill Sets:
Bag-of-Words
Word2Vec
GloVe
BERT
S-BERT

Removing Outliers:
UMAP
DBSCAN

— **Dim reduction**: **UMAP** (McInnes et al., 2020)

— **Outlier identification**: **DBSCAN** (Ester et al., 1996)
on dim-reduced Sentence-BERT vectors

— **21.5k** extracted bullet points ➔ **18.8k** skills

Job Postings

Work hard play hard

Working from home

Coding Python R

Communication Skills

Presenting and writing in English

Pingpong table

Databases SQL

Great salary

Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-based Recommendation System. Koen Bothmer, Tim Schlippe. The Learning Ideas Conference 2022

SKILL SKANNER: Extract & Compare Skills

Retrieving Skill Sets	→	Pre-Processing Skill Sets	→	Removing Outliers	→	Clustering Skill Sets	→	Application: Skill Scanner

Pre-Processing Skill Sets:
Bag-of-Words
Word2Vec
GloVe
BERT
S-BERT

Removing Outliers:
UMAP
DBSCAN

Clustering Skill Sets:
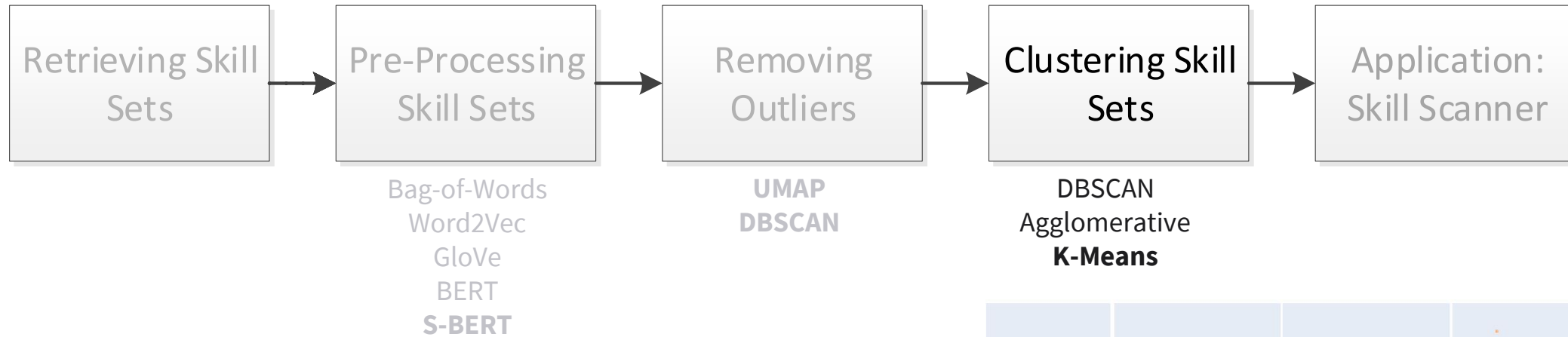DBSCAN
Agglomerative
K-Means

— **Clustering**: **K-Means** (MacQueen et al., 1967)
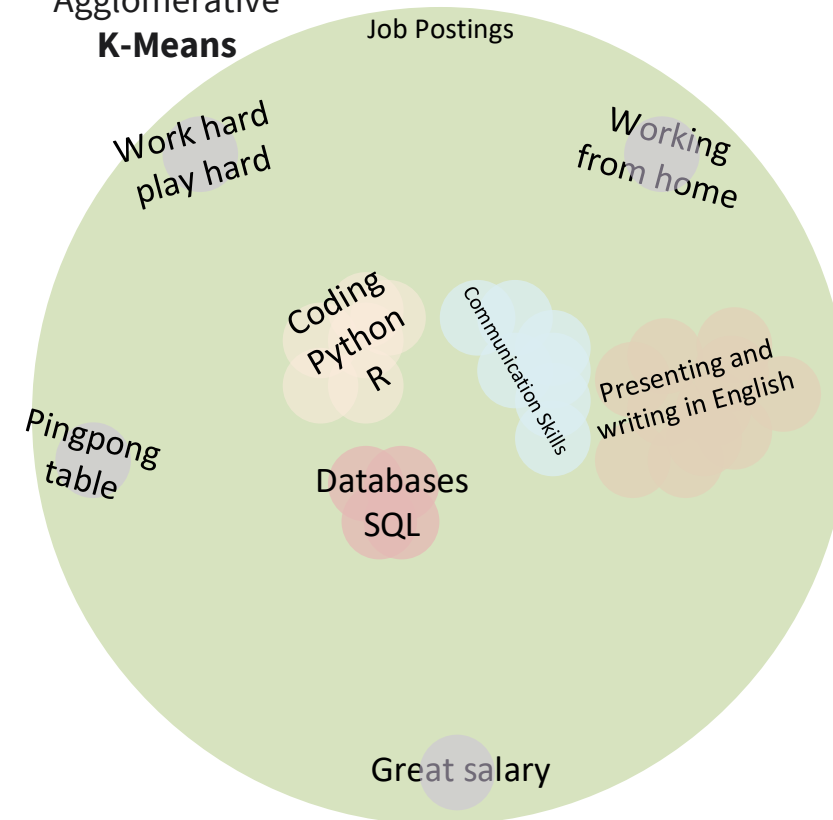
— **31 clusters**: optimal number by silhouette score

Job Postings

- Work hard play hard
- Working from home
- Coding Python R
- Communication Skills
- Presenting and writing in English
- Pingpong table
- Databases SQL
- Great salary

SKILL SKANNER: Extract & Compare Skills

```
Retrieving Skill  →  Pre-Processing  →  Removing  →  Clustering Skill  →  Application:
     Sets              Skill Sets        Outliers         Sets            Skill Scanner
```

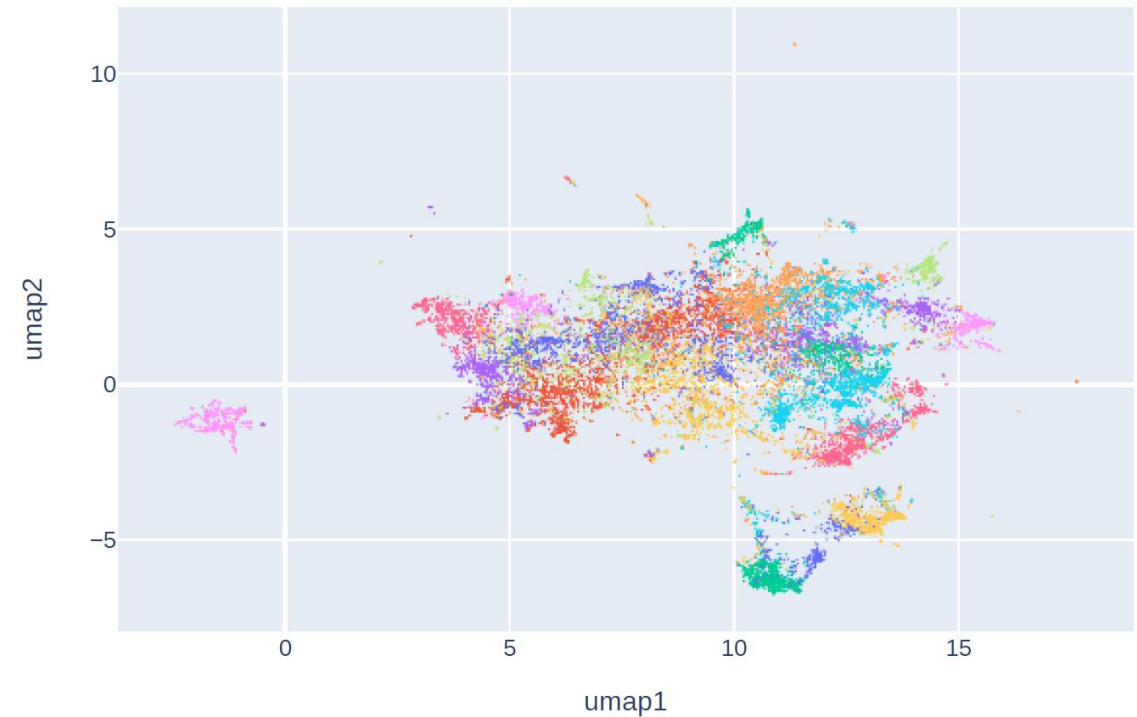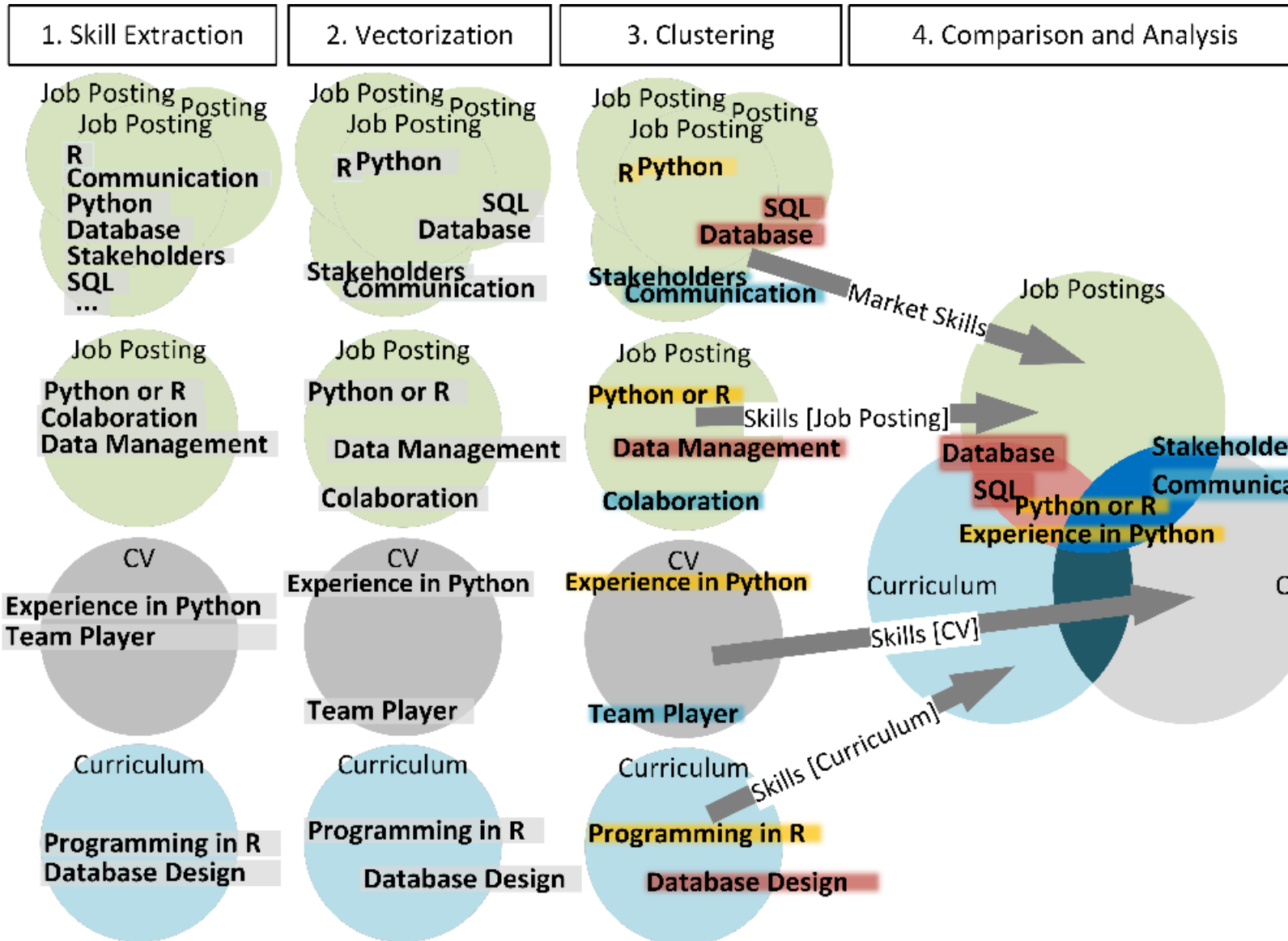| | Bag-of-Words | UMAP | DBSCAN | |
	Word2Vec	DBSCAN	Agglomerative	
	GloVe		**K-Means**	
	BERT			
	S-BERT			

— **Clustering**: **K-Means** (MacQueen et al., 1967)

— **31 clusters**: optimal number by silhouette score

SKILL SKANNER: Extract & Compare Skills

Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-based Recommendation System. Koen Bothmer, Tim Schlippe. The Learning Ideas Conference 2022

4

REPORTS

1. Comparison to Competition

See how you rank up against other JOB SEEKERS' CV's.

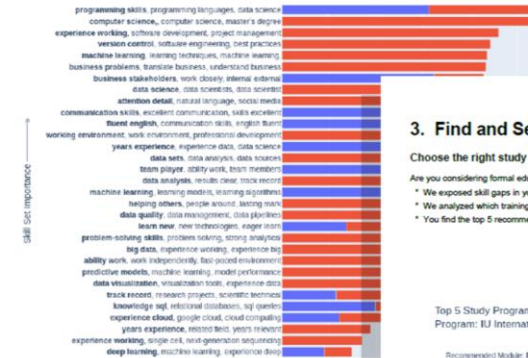Your total CV Coverage Score:

Your Score:

Needs Imp

What do the scores mean?

Outperformed by over 50% of JOB SEEKERS

** Note: In this example you can only input 5 s

2. Fit to Demand

Gain insight in your skill gaps: See how well your CV covers the most important skill sets demanded in Job Postings.

programming skills, programming languages, data science
computer science, computer science, master's degree
experience working, software development, project management
version control, software engineering, best practices
machine learning, learning techniques, machine learning
business problems, translate business, understand business
business stakeholders, work closely, internal external
data science, data scientists, data scientist
attention detail, natural language, social media
communication skills, excellent communication, skills excellent
fluent english, communication skills, english fluent
working environment, work environment, professional development
years experience, experience data, data science
data sets, data analysis, data sources
team player, ability work, team members
data analysis, results clear, track record
machine learning, learning models, learning algorithms
helping others, people around, testing mark
data quality, data management, data pipelines
learn new, new technologies, eager learn
problem-solving skills, problem solving, strong analytics
big data, experience working, experience big
ability work, work independently, fast-paced environment
predictive models, machine learning, model performance
data visualization, visualization tools, experience data
track record, research projects, scientific technical
knowledge sql, relational databases, sql queries
experience cloud, google cloud, cloud computing
years experience, related field, years relevant
experience working, single cell, next-generation sequencing
deep learning, machine learning, experience deep

** Note: A coverage of 100% is impossible to attain, a coverage of over 70% can be co
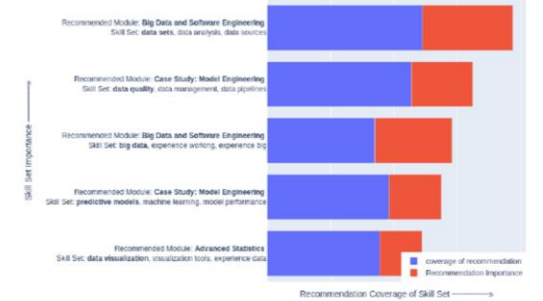
3. Find and Select

Choose the right study program for YOU

Are you considering formal education to upskill yourself? Skill Scanner helps you find the right program:
* We exposed skill gaps in your CV.
* We analyzed which training courses complement your CV.
* You find the top 5 recommendations in the figure below.

Top 5 Study Program Module Recommendations
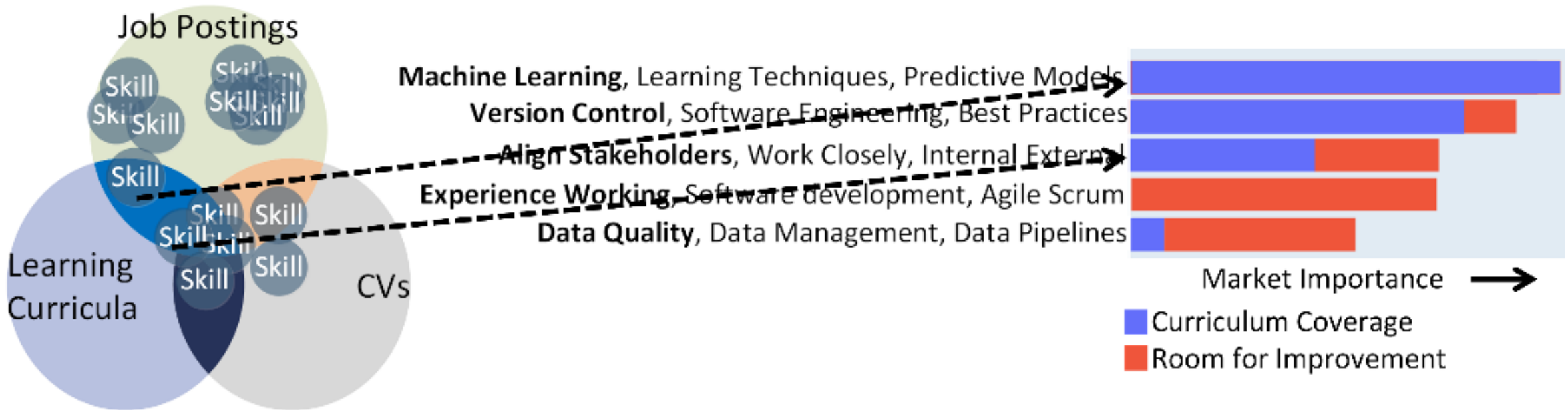Program: IU International University - Data Scientist 60ECT

Recommended Module: Big Data and Software Engineering
Skill Set: data sets, data analysis, data sources

Recommended Module: Case Study: Model Engineering
Skill Set: data quality, data management, data pipelines

Recommended Module: Big Data and Software Engineering
Skill Set: big data, experience working, experience big

Recommended Module: Case Study: Model Engineering
Skill Set: predictive models, machine learning, model performance

Recommended Module: Advanced Statistics
Skill Set: data visualization, visualization tools, experience data

coverage of recommendation
Recommendation Importance

Recommendation Coverage of Skill Set

** Note: In the future more learning content will be analyzed.

Thank you for reading, please remember to fill out our questionnaire at https://forms.gle/ct4DSno6UxN4qofu8

SKILL SKANNER: Reports

Curriculum-Market report

Job Postings

Learning Curricula

CVs

Machine Learning, Learning Techniques, Predictive Models
Version Control, Software Engineering, Best Practices
Align Stakeholders, Work Closely, Internal External
Experience Working, Software development, Agile Scrum
Data Quality, Data Management, Data Pipelines

Market Importance →

Curriculum Coverage
Room for Improvement

Skill Scanner: Connecting and Supporting Employers, Job Seekers and Educational Institutions with an AI-based Recommendation System. Koen Bothmer, Tim Schlippe. The Learning Ideas Conference 2022

5

FEEDBACK

STUDY

Skill Scanner: Improve **Effectivity** from Each Perspective

Skill Scanner: **Accelerate** processes from Each Perspective

33 job seekers, 36 employers, 39 educational institutions

✓ effectivity, ✓ speed, ✓ fairness, ✓ explainability, ✓ autonomy, ✓ support

5

CONCLUSION AND FUTURE WORK

CONCLUSION AND FUTURE WORK

Conclusion

– Unique approach to extract and compare skills from *job postings, learning curricula* and *CVs.*

– Applied with our pipeline

– We introduce *Skill Scanner*

– Feedback of potential users is positive

CONCLUSION AND FUTURE WORK

Conclusion

- Unique approach to extract and compare skills from *job postings, learning curricula* and *CVs.*
- Applied with our pipeline
- We introduce *Skill Scanner*
- Feedback of potential users is positive

Furture Work

- Expand to other domains
- Develop report for *study program design.*
- Develop report for supporting *job coaches* and *study advisors*

THANK YOU

Koen Bothmer

✈ koenbothmer@gmail.com

Tim Schlippe

tim.schlippe@iu.org

REFERENCES

Literature

— **United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021)**: https://www.un.org/sustainabledevelopment/sustainable-development-goals

— **Faliagka, E., Iliadis, L. Karydis, I., Rigou, M., Sioutas, S., Tsakalidis, A. Tzimas, G. (2014)**: *On-line Consistent Ranking on E-Recruitment: Seeking the Truth Behind a Well-Formed CV*. Artif Intell Rev 42, pp. 515–528.

— **Fernández-Reyes, F.C., Shinde, S. (2019)**: *CV Retrieval System Based on Job Description Matching Using Hybrid Word Embeddings*, Computer Speech & Language, vol 56.

— **Wang, Y., Allouache, Y., Joubert, C. (2021):** *Analysing CV Corpus for Finding Suitable Candi-dates using Knowledge Graph and BERT*. DBKDA.

— **Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019)**: *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. In: NAACL-HLT, Minneapolis, Minnesota.

— **Siting, Z., Wenxing, H., Ning, Z., Fan, Yang (2012)**: *Job Recommender Systems: A Survey*. ICCSE (2012)

— **Hong, W., Zheng, S., Wang, H., & Shi, J. (2013)**: *A Job Recommender System Based on User Clustering*. J. Comput., 8, 1960-1967.

Images

— **Images provided by OpenClipart-Vectors/154119/Pixabay**. (https://pixabay.com/vectors/international-project-world-154119 [last access: 16.16.2021])

— **Coursebook (NLP and Computer Vision)**: IU International University of Applied Sciences. Version No.: 001-2020-1211.

REFERENCES

Literature

— **Alotaibi, S. (2012)**: **A Survey of Job Recommender Systems**. Int. J. Phys. Sci.

— **Diaby, M., Viennet, E., Launay, T. (2013)**: *Toward the Next Generation of Recruitment Tools: An Online Social Network-Based Job Recommender System*. ASONAM.

— **Li, J., Arya, D., Ha-Thuc, V., Sinha, S. (2016)**: How to Get Them a Dream Job? Entity-Aware Features for Personalized Job Search Ranking. SIGKDD.

Images

— **Images provided by OpenClipart-Vectors/154119/Pixabay**. (https://pixabay.com/vectors/international-project-world-154119 [last access: 16.16.2021])

— **Coursebook (NLP and Computer Vision)**: IU International University of Applied Sciences. Version No.: 001-2020-1211.