

SENIOR PROJECT · BACHELOR OF DATA SCIENCE

A Hybrid University Recommendation System

for European Higher Education

AI-powered semantic matching across 3,362 European universities — Sentence-BERT embeddings, K-Means clustering, and hybrid rule-based filtering.

Mariam Shreif · Leen Shreif

Lebanese University — Data Science Department

Supervisor: Dr. Kassem Danash · 2025–2026



Choosing a university has become an information-overload problem

Information is scattered

Rankings, tuition, scholarships, and program details live across dozens of separate sites and databases.

Too many factors at once

Country, field of study, budget, language, and reputation all must be weighed together, not one at a time.

Keyword search falls short

Two programs described with different words can be a great match — but keyword search won't see it.

3,362

universities.

36 countries · 26 fields
101 QS-ranked institutions

Build a system that recommends — not just lists — universities

Core Objectives

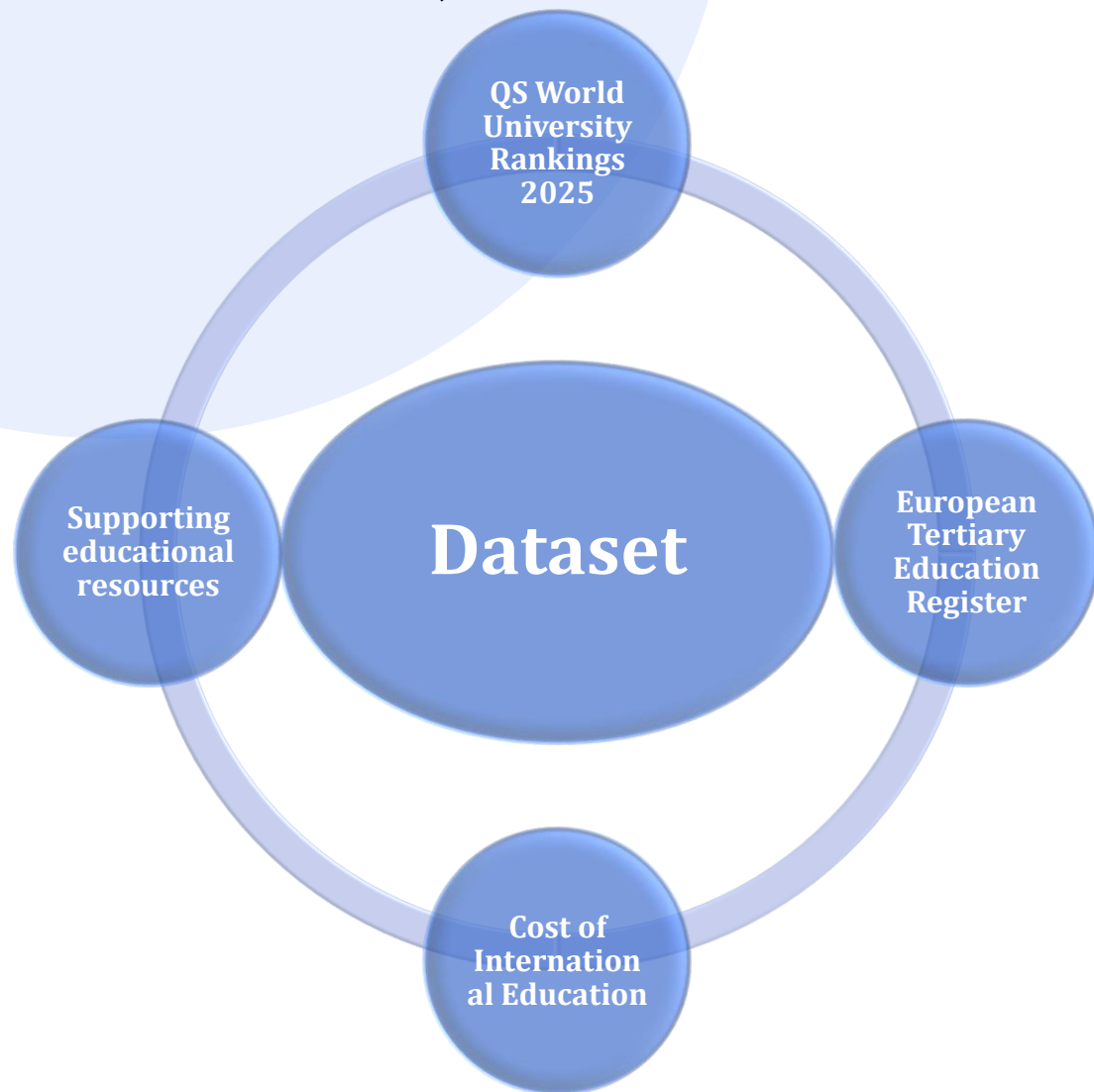
- ▶ Integrate 4 educational data sources into one enriched dataset
- ▶ Generate semantic embeddings of university profiles (Sentence-BERT)
- ▶ Cluster and rank institutions by similarity (K-Means + cosine similarity)
- ▶ Apply rule-based filters for practical, real-world constraints
- ▶ Deliver it as a live, interactive web application



Related systems target pieces of this problem — not the whole thing

Study	Approach	Gap
Alves da Silva & Ferreira (2025)	Transformer embeddings for academic activities	Courses & extracurriculars, not universities
Sangeetha et al. (2025)	Hybrid BERT model for academic collaboration	Finds collaborators, not institutions
Zineb (2025)	Semantic similarity + knowledge graph	Recommends degree programs, not full profiles
Su (2025)	BERT + knowledge-graph recommendation	Built for reading platforms, not education
This project	SBERT + K-Means + cosine similarity + rules	Full university profiles, no interaction history needed

One dataset, stitched from four sources



3,362

UNIVERSITIES

36

COUNTRIES

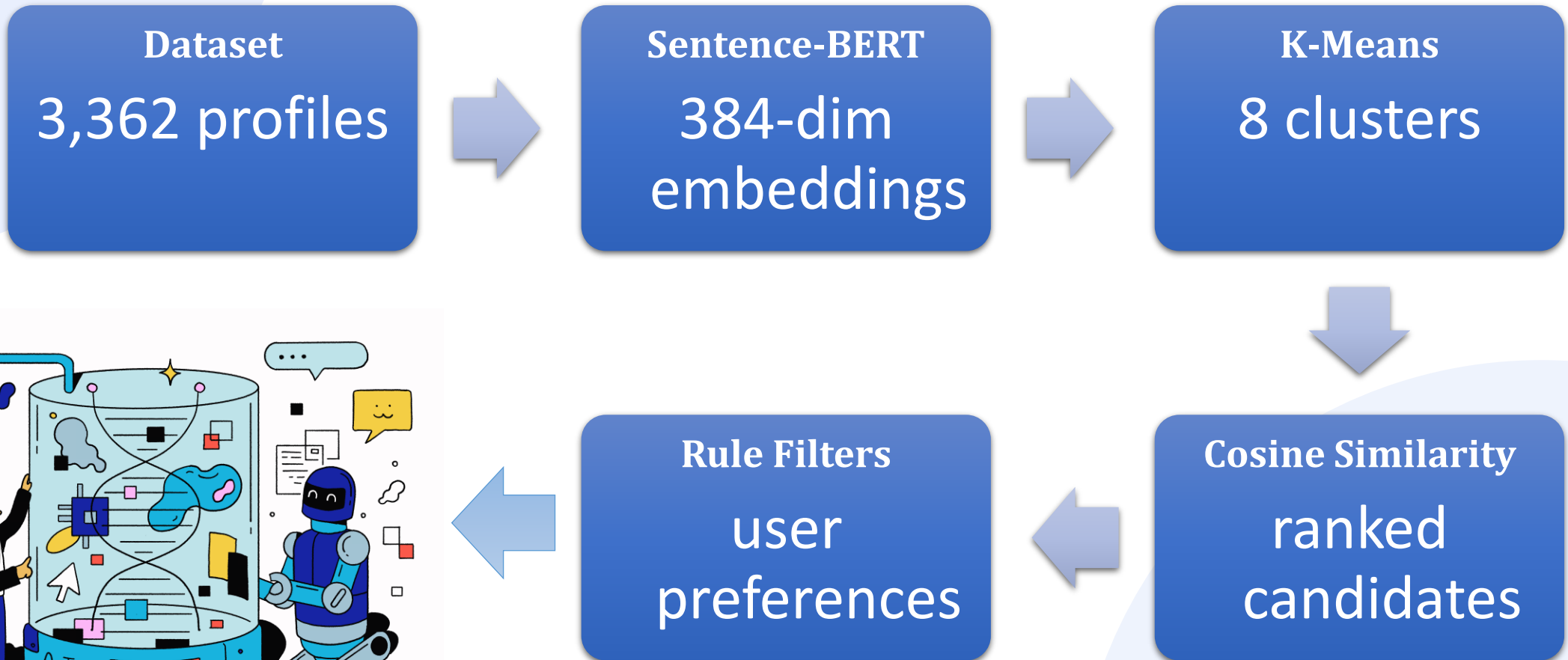
25

FIELDS OF STUDY

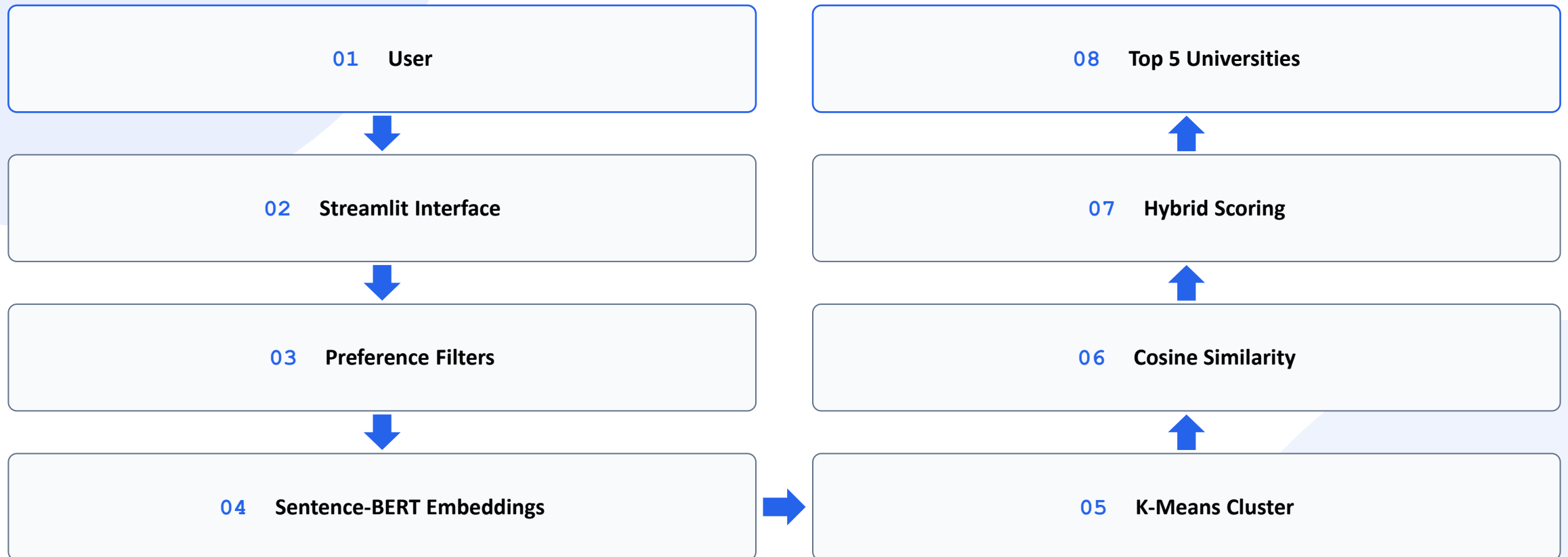
101

QS RANKED

From raw text to a ranked shortlist



How a search request flows through the system



Sentence-BERT turns each university into a 384-number fingerprint

Name, fields of study, and description are merged into one `Combined Text` field, then passed through the pre-trained `all-MiniLM-L6-v2` Sentence-BERT model.

Universities with similar academic profiles land close together in this 384-dimensional space — even when their program descriptions use completely different wording.

```
"University: TU Munich. Fields: Computer Science..."
```

↓ Sentence-BERT (all-MiniLM-L6-v2) ↓

```
[ 0.0412, -0.1187, 0.0733, ... ]
```

384-dimensional embedding vector

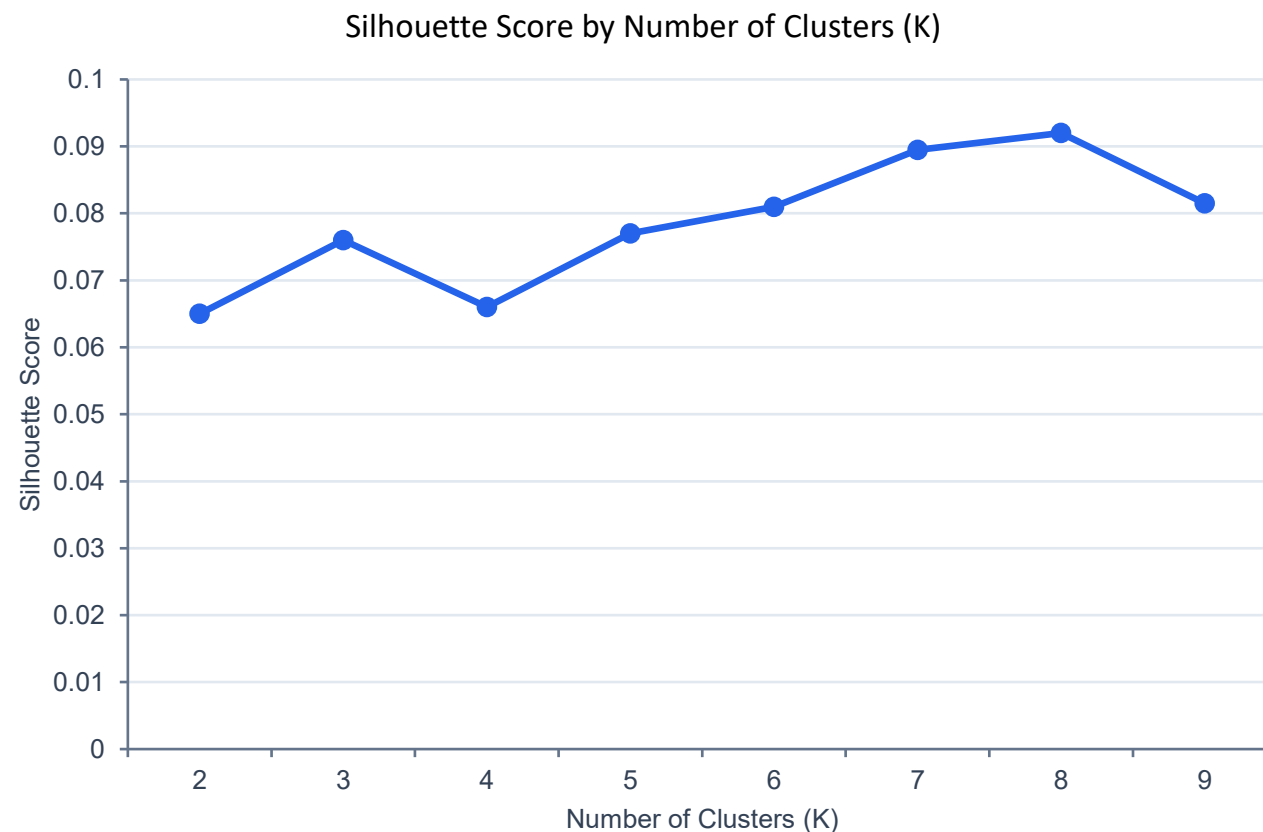
one per university, reused for clustering and similarity

K-Means groups similar universities before ranking begins

Clustering narrows the search space: instead of comparing a university against all 3,362 others, similarity is only computed within its own cluster.

K = 8

clusters selected —
highest silhouette score



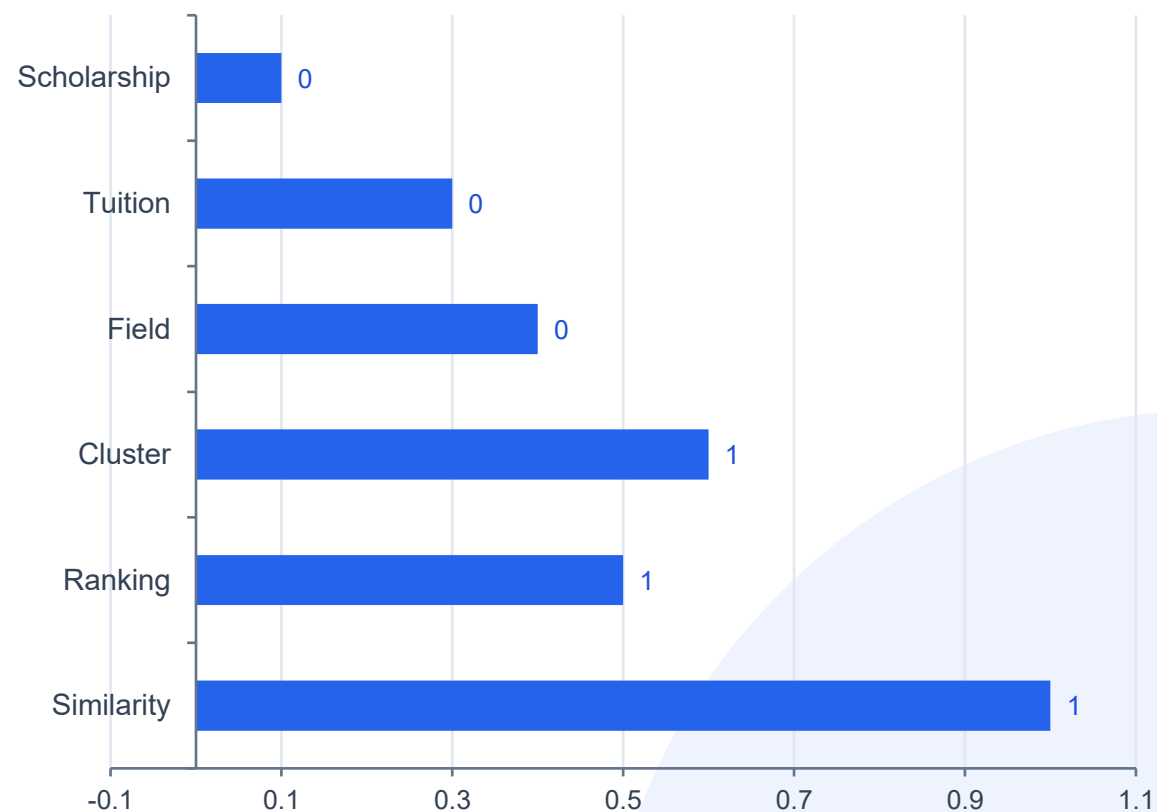
Similarity finds what's alike. Rules make it usable.

Final Score =

```
1.0 × similarity + 0.6 × same cluster  
+ 0.4 × field match + 0.3 × tuition  
match  
+ 0.1 × scholarship + 0.5 × ranking  
match
```

Preferences act as hard filters first — a university that doesn't fit your budget or country is never scored at all. Semantic similarity then ranks what's left.

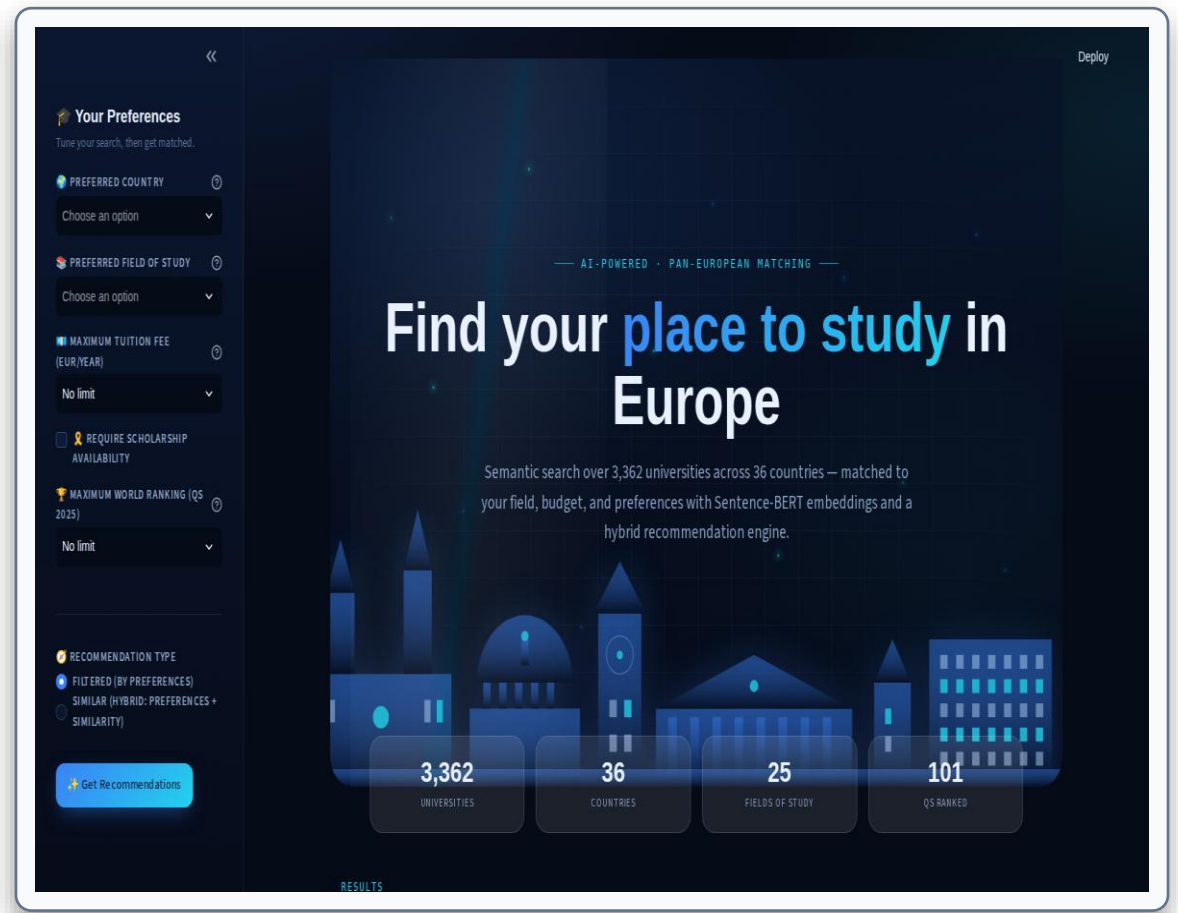
Signal Weights in the Hybrid Score



A live app, not just a notebook

- ▶ Real-time filtering by country, field, tuition, scholarship & ranking
- ▶ Hybrid “similar to this university” mode using live embeddings
- ▶ Explainable results — every recommendation shows why it matched
- ▶ Deployed free on Streamlit Community Cloud

`hybrid-european-university-recommender.streamlit.app`



From five preferences to five ranked universities

User Preferences

COUNTRY

Germany

FIELD

Computer Science

MAX TUITION

€5,000 / yr

SCHOLARSHIP

Required

MAX QS RANK

Top 350

#	University	QS	Tuition	Scholarship
1	Technical University of Munich	28	€0	Yes
2	RWTH Aachen University	99	€0	Yes
3	Karlsruhe Institute of Technology	102	€1,500	Yes
4	University of Stuttgart	314	€1,500	Yes
5	University of Bremen	Not Ranked	€0	Yes

Top match — Technical University of Munich — scored 95% semantic similarity to the reference profile.

Where the system stands — and where it's going



- Some attributes (tuition, fields) are estimated from country-level data when official figures weren't available
- Only 101 of 3,362 universities carry a QS rank — treated as optional, not primary
- Covers Europe only; no admission requirements or acceptance odds

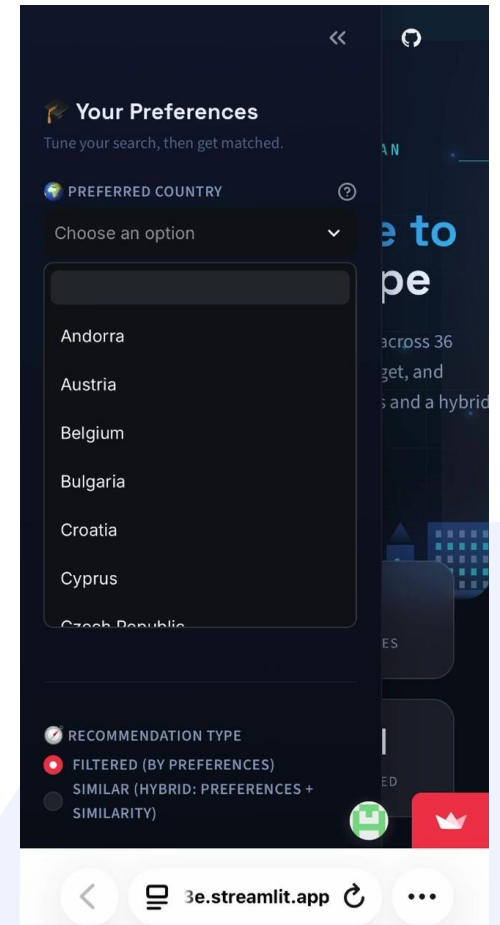


- Add admission requirements, employability, and cost-of-living signals
- Learn from real user interactions to blend in collaborative filtering
- Expand coverage beyond Europe with a persistent, auto-updating database

See the recommender in action

- Live walkthrough of the deployed Streamlit app
- Setting preferences and generating ranked results in real time
- Explainable matches — see why each university was recommended

hybrid-european-university-recommender.streamlit.app





Thank You

Questions & Discussion

Mariam Shreif · Leen Shreif

`hybrid-european-university-recommender.streamlit.app`