Job Recommendation System: An NLP Based Implementation

Presented By

Alessandro Gentili - Manon Lainaud - Marc Martinez - Nigel Tao

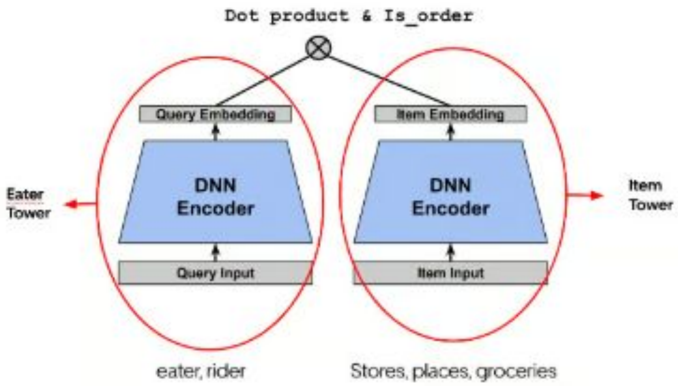
1) Introduction

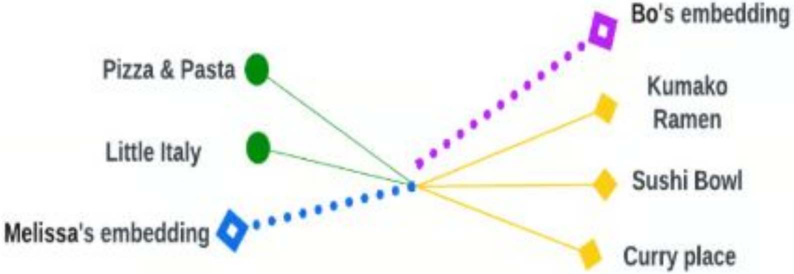
- Objective to this Milestone :Develop a recommendation system
- N
- Focus on three primary fields
  - Job focus
  - Relevant skills
  - Full-descriptive text

2) Related work

- Two tower architecture.proposed by UBER recommendation system
- Confit.A project which uses data augmentation and contrastive learning
- Entity extraction.Furthermore, Das et al. (2018)

Two tower architecture





CONFIT

- Data augmentation for reducing data sparsity
- Contrastive learning for obtaining a dense embedding space
- A simple transformer architecture is applied and cosine similarity is used

NER for resume extraction

- Implement NER techniques to extract relevant information
- Techniques such as POS tagging or tokenization are used
- The author deals with the problem of having different contexts for the same word.

3) Solution Concept And Approach

- Classical **cleaning pipeline**:
    - Named Entity Recognition (NER) for extracting job-relevant entities
- **Two-Tower embedding** model:
    - Learning semantic similarity through contrastive learning
    - Twin Tower Architecture
    - Matching Engine

Contrastive Learning

- How contrastive learning works:
    - A **self-supervised learning technique** that trains a model by comparing pairs of data points
    - The goal is to bring similar points **closer** while pushing dissimilar pairs apart
    - The main idea is that the model learns meaningful representation **without labels**
- How contrastive learning is applied:
    - Each JD and CV is **embedded** in a vector space
    - The model receives **positive and negative pairs**

Twin Tower Architecture

- Two independent neural networks:

    - **Job Tower**: encodes job description
    - **Resume Tower**: encodes resume content

- The embeddings are shared in the **same vector space**

Matching Engine

1. Indexing Job description:

    - Job embeddings are stored in FAISS index for **fast lookup**

2. Querying with a resume:

    - A **resume is encoded** as an embedded vector
    - FAISS **retrieve the top-k** closest nearest job embeddings based on cosine similarity

3. Ranking and final selection:

    - Matches are ranked by cosine **similarity scores**
    - The **top result** represent the best job position for the candidate

4) Experiments

- 2 datasets :
  - Linkedin Job Postings : focus on **job requirements**
  - Resume : **contains “Category”** columns essential for evaluating the model’s performance and analysing its ability to generalize

- Data Preprocessing:
  - **Text cleaning** : lowercasing, tokenization
  - **Embedding generation:** Sentence-BERT (SBERT), capture the semantics meaning and moving beyond simple keyword matching
  - **Additional features :** TF-IDF
  - Data was transformed into share embedding space
  - **Category** : Enables sector-specific analysis -> help identify potential biases and enable visualizations like t-SNE to observe clustering patterns

Model Architecture

- Two-Tower Neural Network

  - **One tower:** Job description
  - **Second tower:** Resume
  - **SBERT embeddings:** follow by dense layers and normalization to ensure compatibility
  - **Training:** Contrastive loss to minimize distance for matching pairs
  -> This ensured that the model learned meaningful semantic relationships

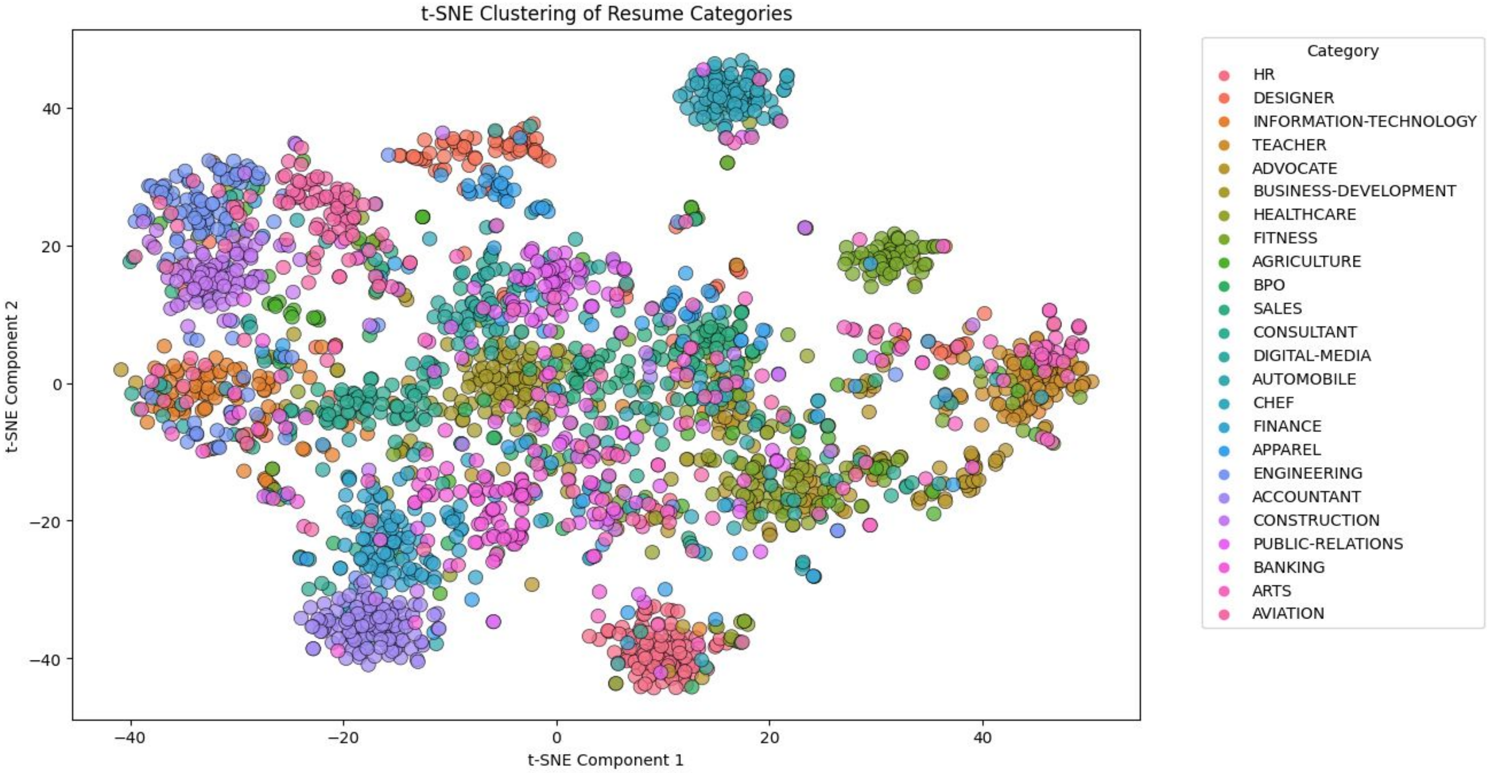
- Matching Process :
  - **FAISS indexing:** Ensuring fast retrieval of relevant matches
  - Input: Job description -> vector representation
  - Matching: **Nearest-neighbour** search in FAISS retrieves the closest matching resumes
  - Refined : **Cosine similarity,** prioritizing resumes with the highest semantic alignment to the job description

- Ethical Considerations:
  - **Data privacy** : No sensitive information
  - **Anonymized resume templates:** provided diverse and structured data without compromising privacy
  - **Compliance :** GDPR, maintaining realistic evaluation framework

Key results

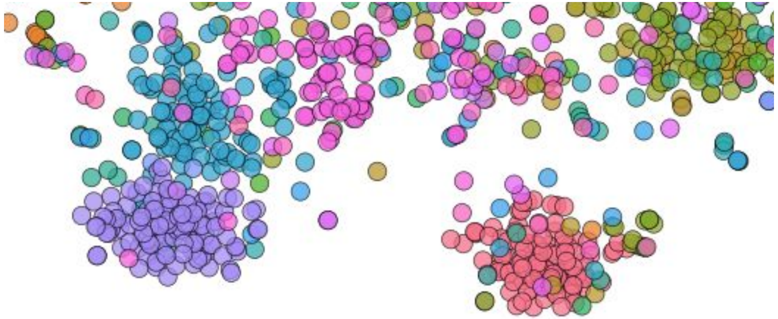
- Systematically assess our model’s performance across industries
- Leveraging structured embedding
- Fast and accurate retrieval
- Sector-specific evaluations -> deliver accurate and meaningful matches

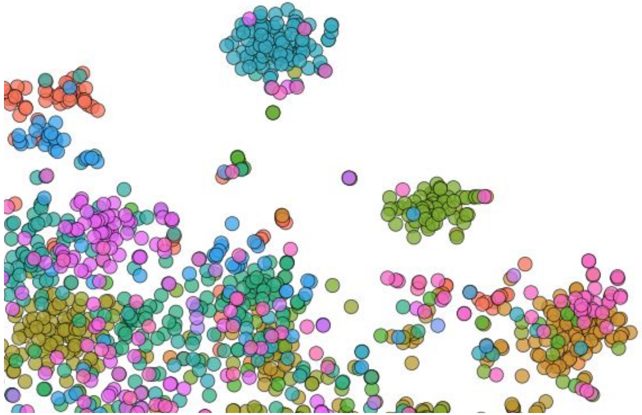
5) Analysis


t-SNE Clustering of Resume Categories
t-SNE Component 1
t-SNE Component 2
−40
−20
0
20
40
Category
HR
DESIGNER
INFORMATION-TECHNOLOGY
TEACHER
ADVOCATE
BUSINESS-DEVELOPMENT
HEALTHCARE
FITNESS
AGRICULTURE
BPO
SALES
CONSULTANT
DIGITAL-MEDIA
AUTOMOBILE
CHEF
FINANCE
APPAREL
ENGINEERING
ACCOUNTANT
CONSTRUCTION
PUBLIC-RELATIONS
BANKING
ARTS
AVIATION

t-SNE Clustering of Resume Categories

- t-SNE plot reveals distinct clusters for certain job categories
- Clusters can be observed with isolated clusters highlighted
  - Indicate niche job roles or unique skill sets





Cosine Similarity vs FAISS

- Facebook AI Similarity Search proves to be of no use, as its results are often entirely mismatched
- The model was trained using cosine similarity rather than relying on FAISS
- FAISS should still provide sensible matches, yet it fails to do so, further indicating that it does not contribute meaningfully

Chef	- lead chef food truck manager (Match: 79.12%) - executive chef (Match: 78.24%) - chef owner (Match: 77.89%) - chef de cuisine (Match: 77.23%) - food preparation workers grill chef (Match: 77.12%)	- noc engineer (Match: 8.81%) - finance accountant (Match: 4.96%) - assistant store manager operations human resources (Match: 3.43%) - hr manager (Match: 0.08%) - graphic designer (Match: 0.00%)
Construction Worker	- construction worker (Match: 73.17%) - construction manager (Match: 72.06%) - construction manager ii (Match: 71.56%) - construction foreman (Match: 71.44%) - senior construction manager (Match: 71.18%)	- sales representative (Match: 9.57%) - executive chef (Match: 4.14%) - director client services films operations technical services west coast (Match: 3.74%) - sales manager (Match: 2.88%) - sales associate (Match: 0.00%)

Identifying Specialities

Type of Resume	Cosine Similarity Evaluation	FAISS Evaluation
Nurse Practitioner	- practicum experience (Match: 74.69%) - charge nurse (Match: 74.24%) - licensed practical nurse step unit (Match: 73.90%) - registered nurse (Match: 73.76%) - field nurse (Match: 73.51%)	- volunteer hr ivolunteer (Match: 10.60%) - accountant (Match: 9.00%) - digital marketing manager (Match: 7.32%) - associate manager desig (Match: 6.41%) - partner business development (Match: 0.00%)
Physical Therapist Assistant	- physical therapist technician (Match: 73.97%) - rehabilitation specialist massage therapist (Match: 70.72%) - certified personal trainer (Match: 70.25%) - dance educator (Match: 69.88%) - physical therapy aide (Match: 69.73%)	- digital strategy manager (Match: 17.15%)) - online coaching personal training (Match: 6.81%) - substitute para professional (Match: 4.14%) - digital marketing account manager (Match: 1.84%) - accountant helper (Match: 0.00%)

Identifying Skill Sets

- Differences in skill sets despite both being in legal

Legal Assistant	- cashier (Match: 66.93%) - customer service manager (Match: 66.44%) - sales associate cashier (Match: 66.42%) - customer service advocate (Match: 66.04%) - sales associate (Match: 65.93%)	- graphic designer (Match: 1.74%) - senior accountant (Match: 1.70%) - healthcare consultant (Match: 0.56%) - chef (Match: 0.25%) - digital marketing director (Match: 0.00%)
Legal Secretary	- legal assistant (Match: 72.51%)) - consultant (Match: 69.45%) - assistant company secretary (Match: 68.85%) - owner attorney mediator (Match: 68.47%) - consultant (Match: 67.97%)	- business development specialist (Match: 15.15%) - administrative assistant (Match: 5.70%) - volunteer advocate (Match: 3.85%) - information technology specialist (Match: 1.12%) - engineering manager (Match: 0.00%)

Identifying Specialized Roles

Type of Resume	Cosine Similarity Evaluation	FAISS Evaluation
iOS Developer	- manager (Match: 66.48%) - information technology intern test automation engineer (Match: 64.49%) - corporate engineering support technician (Match: 63.51%) - director engineering (Match: 63.35%) - platform architect healthcare incubation lab hil (Match: 63.28%)	- mineralogy engineering intern (Match: 3.64%) - director quality improvement network facilitation (Match: 1.00%) - digital marketing associate (Match: 0.35%) - accountant (Match: 0.00%)
Java Developer	- java intern (Match: 72.23%) - information technology intern test automation engineer (Match: 71.77%) - technical designer (Match: 70.03%) - construction worker (Match: 69.86%) - consultant (Match: 69.14%)	- sales representative (Match: 12.63%) - mineralogy engineering intern (Match: 11.18%) - associate teacher (Match: 5.70%) - executive chef (Match: 5.15%) - accountant (Match: 0.00%)

6) Conclusion

- Methodology & Implementation
    - Leveraged NER, TF-IDF, and contrastive learning for feature extraction.
    - Used SpaCy, TF-IDF, and dense embeddings to match CVs and JDs effectively.
    - Improved semantic relevance and ranking for job recommendations.
- Impact & Benefits
    - Reduces manual effort in recruitment for HR teams and candidates.
    - Enhances fairness & efficiency in job matching.
    - Provides a scalable, modular solution for large datasets.

Thank you for your attention