

# Ali Najafi

LinkedIn: [linkedin.com/in/najafi-ali1998](https://www.linkedin.com/in/najafi-ali1998)

Github: [github.com/alinajafi1998](https://github.com/alinajafi1998)

Email: [ali.najafi@sabanciuniv.edu](mailto:ali.najafi@sabanciuniv.edu)

Mobile: +90-543-167-1440

## EDUCATION

- **Sabancı University** Istanbul, Turkey  
*M.Sc. - Computer Science; GPA: 4.00/4.00* *Feb. 2022 - [Expected] Feb 2024*  
*Awarded full tuition waiver, housing, monthly stipend*  
*Courses: Big Data Processing, Data Mining*
- **University of Tabriz** Tabriz, Iran  
*B.Sc. - Computer Engineering; GPA: 3.34/4.00* *Sept. 2017 - Sept. 2021*

## SKILLS SUMMARY

- **Languages:** Python, Java, SQL, Bash, HTML/CSS, C++
- **Frameworks:** HuggingFace, PyTorch, Tensorflow, Keras, Pandas, NumPy, Scikit-Learn, Spark, ElasticSearch
- **Tools:** Docker, GIT, VSCode
- **Platforms:** Linux, Google Cloud

## EXPERIENCE

- **Sabancı University NLP Group** Still Collaborating  
*NLP/Deep Learning Researcher* *Feb 2022 - Present*
  - **Communicative Multi-agent System for clustering Tweeter Stream Data:**
    - \* Clustered and extracted trending topics from millions of tweets on streaming data.
    - \* Worked with a team of five engineers and researchers, used parallelization to assign the load on different cores to speed up the process, and later published the code and results.
    - \* Designed the system, developed the Python code, evaluated, and conducted statistical analysis.
  - **Document/Passage Ranking:**
    - \* Developed a multi-stage pipeline to retrieve the relevant documents/passages for the input queries (MSMARCO).
    - \* Utilized encoder-decoder model DocT5Query to expand the documents with the newly generated queries.
    - \* Used ElasticSearch (BM25) to retrieve candidates for ranking.
    - \* Applied T5 to filter out retrieved documents from the BM25 algorithm.
  - **Youtube Comment Sentiment Analyzer:**
    - \* Collected Youtube comments using Youtube-API.
    - \* Preprocessed the informal text data.
    - \* Implemented a sentiment classifier for youtube tutorial comments using RoBERTa-Base model.
- **Sabancı University** *Feb. 2022 - Present*  
*Teaching Assistant*
  - **Teaching Assistantship at courses:**
    - \* Machine Learning
    - \* Introduction to Computing
- **University of Tabriz** *Sept. 2017 - Sept. 2021*  
*Machine Learning Researcher*
  - **Machine Translation:**
    - \* Built an AI model for translating English to Italian.
    - \* Implemented a custom-made Transformer-based model from scratch using Tensorflow. In spite of the limited resources and the small size of the model, it achieved a BLEU score of 21.5.
- **NOAY** *June. 2021 - Sept. 2021*  
*Software Developer*
  - **Extending Pandas API:**
    - \* Built automatic and reusable data processing modules for general use of scientists from different backgrounds.
    - \* Developed modules by extending Pandas API with team of five developers.
- **RAyin Samaneh Arta (RASA)** *June. 2020 - Sept. 2020*  
*Machine Learning Engineer*
  - **Dental LandMark Detection:**
    - \* Created system to detect specific landmarks on dental images beneficial for dentists.
    - \* Worked with a team of three developers and used VGG-face as a pre-trained model to improve system accuracy.

## PUBLICATIONS

- **Najafi, A.,** Gholipour-Shilabin, A., Dehkharghani, R. et al. ComStreamClust: a Communicative Multi-Agent Approach to Text Clustering in Streaming Data. Ann. Data. Sci. (2022). DOI: 10.1007/s40745-022-00426-4