Ali Najafi

ali.najafi@sabanciuniv.edu

najafi-ali.com · github.com/alinajafi1998

RESEARCH INTERESTS

My research interests are in natural language processing (NLP), particularly in its intersections with machine learning and Social Computations.

EDUCATION

Sabanci University, Istanbul, Turkey
M.Sc., Computer Science and Engineering
Thesis: Developing Turkish Language Models for Detecting Manipulations on Social Media
Advisors: Onur Varol
GPA: 4.0/4.0

University Of Tabriz, Tabriz, Iran
B.Sc., Computer Science and Engineering, (July 2021)
CGPA: 16.58/20

PUBLICATIONS
GOOGLE SCHOLAR

Journal Articles
J.1 **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)

Preprints and Under Review Articles
J.2 Ali Najafi, Nihat Mugurtay, Ege Demirci, Serhat Demirkiran, Huseyin Alper Karadeniz, and Onur Varol. Secim2023: First Public Dataset for Studying Turkish General Election. arXiv preprint arXiv:2211.13121, 2022.

CONDUCTED RESEARCH AND ACADEMIC EXPERIENCE

VRLLAB, Istanbul, Turkey
Turkish-Twitter Roberta LLM
Trained Roberta Model from scratch using HuggingFace Transformers
Used Google TPUs for Training the Language Model
Utilized DeepSpeed accelerator for training the Model
Trained BPE tokenizer from scratch on 900M Turkish Twitter Data

SUNLP, Istanbul, Turkey
TREC 2020 DeepLearningTask - Document/Passage Ranking
Developed a pipeline to retrieve the relevant documents/passages for the input queries (MSMARCO).
Utilized DocT5Query to expand the documents with the newly generated queries.
The expanded documents were retrieved using ElasticSearch(BM25).
T5 was applied to filter out retrieved documents from the BM25 algorithm.

Teaching Experience
Sabanci University, Istanbul, Turkey
Machine Learning (Teaching Assistance, Summer 2022, Fall 2023)
Introduction to Computing (Teaching Assistance, Spring 2022)

WORK EXPERIENCE

NOAY (June. 2021 - Sept. 2021)
Built automatic and reusable data processing modules for general use of scientists from different backgrounds and developed modules by extending Pandas API with team of five developers.

RAyin Samaneh Arta (RASA) (June. 2020 - Sept. 2020)
Created system to detect specific landmarks on dental images beneficial for dentists and worked with a team of three developers and used VGG-face as a pre-trained model to improve system accuracy.

Honors and Awards	Awarded full tuition waiver, housing, monthly stipend from Sabanci University (2022-2024)
	Awarded full tuition waiver from University of Tabriz (2017-2021)

Computer and Language Skills

Programming Languages and Skills:
Frequent user of Python for data analysis using Matplotlib, Pandas, Scikit-learn, etc.
Experience in LaTeX, C / C++, C#, Java
Familiar with HTML, CSS, JS for frontend, Flask, and Django for larger applications
Used MySQL, NoSQL (Redis), Map-Reduce
Languages: English (Fluent), Persian (Native), Azerbaijani (Native), Turkish (Fluent)