Ali Najafi

najafi-ali.com

GitHub: @alinajafi1998
Email: ali.najafi@sabanciuniv.edu

RESEARCH INTERESTS

My research spans the interdisciplinary fields of machine learning and natural language processing, with a particular focus on addressing societal challenges through computational approaches. I aim to develop and apply tools to uncover complex patterns and leverage these insights to address pressing social issues.

EDUCATION

Sabanci University, Istanbul, Turkey *Feb. 2022 - Present*
M.Sc., Computer Science and Engineering
Thesis: Developing Turkish Language Models on Social Media
Advisor: Assistant Prof. Onur Varol
CGPA: 3.71/4.0

University Of Tabriz, Tabriz, Iran *Sept. 2017 - Sept. 2021*
B.Sc., Computer Science and Engineering
CGPA: 16.58/20

PUBLICATIONS
GOOGLE SCHOLAR
CITATIONS: 16
H-INDEX: 3

Journal Articles

J.3 **Najafi, A.**, and Onur Varol. **"Turkishbertweet: Fast and reliable large language model for social media analysis."** Expert Systems with Applications 255 (2024): 124737.

J.2 **Najafi, A.**, et al. **"First public dataset to study 2023 Turkish general election."** Scientific Reports 14 (2024).

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)

Refereed Conference Proceedings

C.2 Yener, Alp Önder, **Ali Najafi**, and Onur Varol. **"City-level User Location Detection from Social Media and Analysis of Political Participation."** 2024 32nd Signal Processing and Communications Applications Conference (SIU). IEEE, 2024.

C.1 **Najafi, Ali**, and Onur Varol. **"VRLLab at HSD-2Lang 2024: Turkish Hate Speech Detection Online with TurkishBERTweet."** Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024). 2024.

CONDUCTED RESEARCH AND ACADEMIC EXPERIENCE

VRL LAB, Istanbul, Turkey *Sept, 2022 - Present*

- **TurkishBERTweet**
  * Achieved almost 10% significant improvement in the F1-Score of Text Classification Tasks
  * Developed classification pipelines for Sentiment Analysis and Hate Speech Detection Tasks
  * Trained TurkishBERTweet From Scratch
  * Used Google TPUs for Training the Language Model
  * Trained BPE tokenizer from scratch on 900M Turkish Twitter Data
- **Turkish Election Dataset**
  * Analyzed the Retweet Speed of specific Turkish Politicians for different time spans of election day
  * Conducted network analysis of politicians on Twitter using Networkx and Gephi

* Analyzed the followers of politicians to identify social bots using Botometer
  – **CON-NET** under CHIST-ERA call funded by EU and TUBITAK
    * Integrating CONtent and NETwork structure for detecting, understanding, and mitigating online misbehaviour
    * Building multimodal machine learning systems for estimating automated activities

---

INVITED LECTURER AND TALKS

1. Presented the paper "VRLLab at HSD-2Lang 2024: Turkish Hate Speech Detection Online with TurkishBERTweet" at EACL"2024 📄 , Malta, March 2024.

2. Presented the paper "First Public Dataset to Study the 2023 Turkish General Election" at the Turkey Computational Social Sciences Conference (TCSS) 2023 📄 , Koc University, Turkey.

3. Invited lecturer for the talk "Applications of Large Language Models for Social Scientists" in the Political Science course at Sabanci University, Turkey, under the supervision of Dr. Fatih Serkant Adiguzel, Fall 2023.

4. Invited lecturer for the talk "Natural Language Processing for Social Scientists" at the Summer Institutes in Computational Social Science (SICSS) 📄 , Ozyegin University, Turkey, Summer 2023.

TEACHING EXPERIENCE

Sabancı University - Faculty of Engineering and Natural Sciences

CS210: Introduction to Data Science	*Spring 2023, Fall 2023*
CS412: Machine Learning	*Spring 2024, Fall 2022*
CS210: Introduction to Computing	*Spring 2022*

Held weekly office hours and recitation sessions to address course-related questions.
Graded exams and assignments

COMPUTER AND LANGUAGE SKILLS

Skills:
    Programming Languages: Python, C++, Java
    Frameworks: Pytorch, Tensorflow, Keras, Flax, Django
    Librares: Matplotlib, Seaborn, Pandas, Numpy, Gephi, NetworkX, OpenCV, Jax, Docker
    DataBase: SQL, NoSQL (Redis)
    Familiar with GCP, AWS, LaTeX, HTML, CSS, JS

Languages: English (Fluent), Persian (Native), Azerbaijani (Native), Turkish (Fluent)

HONORS AND AWARDS

Hate Speech Detection in Turkish - EACL"2024   *Jan. 2024*
Earned **3rd** Rank among 11 participated teams

TUBITAK Scholarship   *Aug. 2023 - Present*
The Scientific and Technological Research Council of Turkey, Turkey
Awarded to full tuition waiver and a monthly stipend as a researcher in a project funded by TUBITAK

Sabancı University Scholarship   *Feb. 2022 - Aug 2023*
Sabancı University, Istanbul, Turkey
Awarded a 100% tuition scholarship and a monthly stipend as a Teaching Assistant