Suleyman Onur Dogan

sonur.dogan4@gmail.com | sonurdogan.github.io

EDUCATION

- **Sabancı University - Master of Science in Computer Science & Engineering** *2024-2026*
  Full Scholarship | %100 English
  - Courses: Machine Learning, NLP, Scalable Learning Systems, IOT
  - TA: Programming Languages (200+ students)
- **Antalya Bilim University - Bachelor of Science in Computer Engineering** *2019-2024*
  GPA: 3.71/4 | %100 English

EXPERIENCE

- **Junior Data Scientist** Sept 2023 – March 2024
  John Snow Labs, USA (Remote)
  - Taken parts in Spark-NLP for Healthcare projects that includes building Retrieval-Augmented Generation (RAG) - LLM pipelines, creating NLP pipelines and training NER, Classification models, integrating and testing models on AWS and Databricks cloud services.
  - Contributed the development of **open source NLU library** which allows applying latest NLP models from Spark-NLP and HugginFace, directly on any dataframe, with a single line of code.
- **Junior Machine Learning Engineer** March – June 2023
  GrainFox, Canada (Remote)
  - Improved and deployed an API which automatically extracts the information from contract document images using Computer Vision techniques.
  - Developed Large Language Model (LLMs) agent for the extraction of information from documents using LangChain and OpenAI API.
  - Integrated the service API into production pipeline and contributed the implementation of MLOps using FastAPI, Django, Docker, RabbitMQ, Celery, Traefik.
- **Undergraduate Research Assistant** March 2022 – Nov 2023
  Computational Biology Group (Professor Hilal Kazan), Antalya Bilim University
  - Worked on Tubitak (Scientific and Technological Research Council of Turkey) Funded Project – "Neuroimmune guidance cues, MicroRNAs & Inflammatory responses: Sex differences in Cardiovascular Diseases" under supervision of Professor Hilal Kazan.
  - Focused on the differential gene co-expression network with pairwise and cluster-wise analysis, via ML based clustering algorithms and statistical methods in bulk RNA-Seq datasets.
  - Contributed to the development of SciTuna, scRNA-seq dataset Integration Tool which uses Network Alignment algorithm to integrate multiple scRNA-seq dataset into single dataset.
- **Research Intern** July – Sept 2022
  Tastan Lab (Asst. Prof. Oznur Tastan), Sabancı University
  - Developed an open-source project called TLMSA for detection of possible SUMOylating sites that emerges through mutations in protein sequence data via fetching data from National Cancer Institute GDC database and UniProt, manipulating data and using deep sequential neural network model.
- **Undergraduate Research Assistant** Sept 2020 – July 2022
  Statistical Signal Processing Lab, Antalya Bilim University
  - Worked on Tubitak Funded Project - "Investigation of the interactions between the neuronal noise and the neuronal network using information-theoretical approaches" under supervision of Asst. Prof. Deniz Gencaga and Assoc. Prof. Sevgi Sengul Ayan.
  - Conditional information transfer between neurons and neuronal network inference on time series action-potential data investigated with Conditional Transfer Entropy from Information Theory via implementation of IDTxl library.

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C#, C++, C, R, Matlab, HTML, CSS, JS, SQL
- **ML Frameworks**: PyTorch, Tensorflow, Keras, Scikit-Learn, LangChain, OpenCV, Spark-NLP, IDTxl
- **Frameworks**: Django, FastAPI, Java Spring, .NET, Streamlit, RabbitMQ, Celery, Traefik, GitHub, Docker
- **Cloud**: Databricks, AWS EMR, AWS Glue, AWS S3

External Projects

- **Domain Generalization for Surface Anomaly Detection**
  - The motivation of the project is to perform domain generalization using common anomalies in common material types used in the industry. Project is currently active.
- **Healthcare Tourism App – Thesis Project**
  - Healthcare tourism app is developed as thesis project where Patient can be able to plan the trip and make reservation on hospital and hotel in one app.
  - I developed a healthcare tourism chatbot that works based on our SQL database, utilizing RAG with LLM using LangChain, FAISS, Gemini API, FastAPI.
- **The use of Deep Learning in Biomedical Imaging**
  - The application of CNN and transfer learning-based models to biomedical images is investigated using examples from brain tumor classification and Covid-19 diagnosis as part of Biomedical Imaging course of **Prof. Umit Demirbas.**
- **SLAM Simulation of Indoor Robot**
  - In Freshman year, Simulations performed with a BreezySLAM algorithm by using existing Odometry and Lidar Data for SLAM indoor robot. Throughout the project, Kalman Filtering and Particle Filtering in SLAM algorithms were the main emphasis.
- **Matrix Factorization method for Movie Recommendation System with Parallel Computing**
  - In sophomore year, I focused on application of Matrix Factorization technique with ALS on Collobrative Filtering-based recommendation system using PySparkML.

Publications & Conference Presentations

- Mehmet Uğur KAHRAMAN, **Süleyman Onur DOĞAN**, Sevgi ŞENGÜL AYAN, Yaren ŞEKERCİ, Ferhat KOYUNCU, Hakan Bal. Harmony in Diversity: Exploring Eclectic Chair Designs Through Deep Neural Networks Analysis in Interior Design Education. **Under Review**
- **Süleyman Onur Dogan**, Burak Onur Erten, Cesim Erten, Aissa Houdjedj, Hilal Kazan, Yacine Marouf, Mekan Myradov and Oznur Tastan. SciTuna: Single Cell RNA-seq data integration tool using network alignment. **Under Review**
- Yacine Marouf, **Onur Dogan**, Ernest Diez Benavente, Gerard Pasterkamp, Hester M. den Ruijter, Janine van Gils, Katey Rayner, Hilal Kazan. Sex-biased Expression of Neuroimmune Guidance Cues in Cardiovascular Diseases. **Ongoing Publication Process, Presented at** (ERA-CVD'22, Latvia, 2022) & (HIBIT'22, Turkey, 2022)
- **O.Dogan,** S.Ş. Ayan, D.Gencaga. Analysis of Neuronal Interaction Using Information-Theoretical Models. **Presented at** The 8'th International Congress on Fundamental an Applied Science (ICFAS'21, Oct 19-21, 2021)

Volunteer Experience

- **Google Developers Group Antalya** Sept 2020 – Nov 2021
  Co-Organizer
  - Aim is to gather people interested in technology, to organize activities and to follow the technological agenda.
  - Designed and leaded developers.june robotics event where we hosted five experts in different fields of robotics.

Certificates

- **Machine Learning,** Stanford University Online
- **Spark NLP for Data Scientist,** John Snow Labs