Mesut Şafak Bilici

Email: m.safak.bilici@gmail.com
Mobile: +90-553-385-3245
Date Of Birth: 08/1999
Address: Kagithane, Istanbul/Turkey
Website: safakkbilici.github.io
Github: github.com/safakkbilici
LinkedIn: linkedin.com/in/msafakbilici

Education

- **Yildiz Technical University** — Istanbul, Turkey
  Bachelor of Science - Computer Engineering; GPA: 3.48/4.00 — *2017 - 2022*
  Graduation Thesis: *Deep Multimodal Learning with Vision-and-Language Transformers*

Experience

- **Machine Learning Engineer at Insider**
  Full-time — *Sep 2023 - Present*
  - Wrangling complex search modules
- **Research Engineer at Huawei**
  Full-time — *Jan 2022 - Sep 2023*
  - Developed Contrastive Learning based fast multi-lingual approximate vector search algorithm for AppGallery and integrated the algorithm into Elasticsearch.
  - Adapted a large language model based few-shot re-ranking model for app search domain.
  - Implemented low-latency (∼10 ms.) multilingual (Arab countries, Russia, Latin America, all Asia countries excl. China) spelling correction module.
  - Conducting A/B Testing methodologies to evaluate and optimize NLP products for enhanced user experience.
  - Leveraged distributed data processing and model deployment techniques for enabling scalability and accelerating model inference.
  - Presented technical documentations to overseas.
- **Machine Learning Researcher at YTU NOVA Lab.**
  Full-time — *Dec 2020 - Feb 2022*
  - Research lab under the supervision of Prof. Dr. Mehmet Fatih Amasyali.
  - Research on intersection of Variational Inference and Low-Resource Language Models.
  - Published two papers about a novel method on data augmentation for NLP.
- **Data Science Intern at Scoutium**
  Internship — *July 2020 - Sep 2020*
  - Implemented efficient big data pre-processing scripts to reduce the online module's runtime.
  - Developed a flexible REST API package with AWS S3 integration using boto3 to automate the update procedure the company's MySQL database and reduce the throughput.
  - Wrote a pipeline for database server management using AWS S3 and Athena.

Open Source Projects

- **x-tagger**: A Natural Language Processing toolkit for part of speech tagging and named entity recognition with various computational linguistics and deep learning methods. Has more than 4k installation on pip.
- **bayesmedaug**: A Python library that optimizes your data augmentation hyperparameters for medical image segmentation tasks by using Gaussian Process and Bayesian Optimization.

Skills Summary

- **Languages**: Python, C/C++, Java, Bash.
- **Tools**: Elasticsearch, Docker, Redis, git, PostgreSQL, MySQL, MongoDB, Tableau, Airflow.
- **Platforms**: Linux, AWS (S3, Athena, Lambda, EC2).
- **Soft Skills**: Leadership, Event Management, Writing, Public Speaking, Time Management.

Volunteer Experience

- **AI Program Specialist at inzva**
  Volunteer — *Aug 2021 - April 2023*
  - Part of the team that organizes the syllabus of the AI programs.
  - Technical mentoring for the project groups.
  - GPU server administration.
- **Machine Learning Bootcamp Guide at Google**
  Volunteer — *Sep 2022*
  - Guidance for two weeks at Google ML Bootcamp. I taught Recurrent Neural Networks (LSTMs, GRUs) and Word Embeddings (word2vec, GloVe, fasttext, multilinguality, cross-linguality, multimodality, search engines).

Publications

- Can Özbey, Talha Çolakoğlu, **M. Şafak Bilici**, Ekin Can Erkuş. "A Unified Formulation for the Frequency Distribution of Word Frequencies using the Inverse Zipf's Law", in Special Interest Group on Information Retrieval (SIGIR), 2023.
- **M. Şafak Bilici**, Mehmet Fatih Amasyali. "Transformers as Neural Augmenters: Class Conditional Sentence Generation via Variational Bayes", 2022. arXiv: 2205.09391.
- E. Sadi Uysal, **M. Şafak Bilici**, B. Selin Zaza, M. Yiğit Özgenç, Onur Boyar, "Exploring The Limits Of Data Augmentation For Retinal Vessel Segmentation", 2021. arXiv: 2105.09365.
- **M. Şafak Bilici**, Mehmet Fatih Amasyali. "Variational Sentence Augmentation For Masked Language Modeling", in Innovations in Intelligent Systems and Applications Conference, 2021.