

# SOBHAN MOOSAVI

sobhan.mehr84@gmail.com · smoosavi.org · LinkedIn · Google Scholar · GitHub · Kaggle

## Summary

Senior Machine Learning & Applied Scientist with 6+ years of industry experience building and deploying scalable, interpretable ML systems. Expertise spans autonomous driving, large-scale routing, and financial fraud detection — with a consistent focus on production-grade AI that drives measurable real-world impact. Track record of technical leadership and cross-functional collaboration at Lyft, Zoox, and FIS.

## Work Experience

**FIS Global — Bellevue, WA** 2025 – 2026  
*Senior Data Scientist, Machine Learning*

- Researching and architecting a **card fraud detection system** integrated into FIS's production framework; designed a two-tier architecture comprising a **generic fraud model** and a **real-time residual model** that adapts to missed detections and evolving fraud patterns.
- Operates on **multi-billion daily transaction volumes**, leveraging distributed training and inference for scalability; strong emphasis on **interpretability and explainability** for broad stakeholder audiences.

**Zoox (an Amazon subsidiary) — San Mateo, CA** 2023 – 2025  
*Senior Data / ML Scientist*

- Contributed to the development and validation of Zoox's **Collision Avoidance System (CAS)**, driving measurable safety and reliability improvements through data-driven validation and cross-functional collaboration.
- **Safety Validation:** Defined safety metrics, led large-scale validation cycles, and delivered reproducible insights guiding CAS development and leadership decisions.
- **ML Integration:** Enhanced kinematic and behavioral models (KEM) for multi-agent scenarios; applied **multi-modal transformer foundational models** to improve dataset quality, scene representation, and risk estimation.
- **Methodology:** Advanced simulation-based validation frameworks to better capture real-world behavior while optimizing cost and data efficiency.
- **Cross-functional Leadership:** Represented CAS V&V in company-wide forums, aligning autonomy, safety, and data science teams.

**Lyft — San Francisco, CA** 2018 – 2023  
*Senior Data Scientist (2021–2023) · Data Scientist (2020–2021) · Research Intern (2018)*

- **Senior Data Scientist:** Led science and engineering vision and roadmap for a **15+ person team** (ML, SWE, DE, PM) focused on in-house optimized, safe routing; drove A/B experimentation cycles and mentored junior DSs, ML engineers, and interns.
- **Data Scientist:** Designed and implemented a **ML-based routing cost model** that improved route accuracy by **15%** and driver compliance by **10%** — a significant departure from suboptimal or unsafe routes, impacting LyftMap, ETA prediction, pricing, and dispatch.
- **Research Intern:** Developed a **deep neural network for ETA estimation** using spatiotemporal data; co-authored a paper on routing-aware supervised travel time estimation.

Tools & Technologies: Python, TensorFlow, PyTorch, Scikit-learn, SQL, Spark / PySpark, AWS S3, Databricks, Java (familiar)

## Education

**Ph.D., Computer Science** — Ohio State University 2014 – 2019

- Dissertation: *Telematics and Contextual Data Analysis and Driving Risk Prediction*
- Advisors: Prof. Rajiv Ramnath & Prof. Srinivasan Parthasarathy | Research Areas: Data Mining, Machine Learning

**M.S., Computer Software Engineering** — University of Tehran 2009 – 2012

**B.S., Computer Science** — Shahid Beheshti University 2005 – 2009

## Selected Publications & Patent

- Context-aware driver risk prediction with telematics data — *Elsevier Accident Analysis & Prevention*, 2023
- Predicting road constructions based on heterogeneous spatiotemporal data — *ACM SIGSPATIAL*, 2022
- LocationTrails: A federated approach to learning location embeddings — *IEEE/ACM ASONAM*, 2021
- Driving style representation in convolutional recurrent neural network model of driver identification — *Preprint*, 2021
- System and method for analyzing vehicle data — *US Patent 10740990*, 2020
- Short and long-term pattern discovery over large-scale geo-spatiotemporal data — *ACM SIGKDD*, 2019
- Accident risk prediction based on heterogeneous sparse data: New dataset and insights — *ACM SIGSPATIAL*, 2019
- QDEE: Question difficulty and expertise estimation in community Q&A — *AAAI ICWSM*, 2018
- Characterizing driving context from driver behavior — *ACM SIGSPATIAL*, 2017

## Published Datasets

---

- **US Accidents** (2016–2023) — ~180K downloads; among Kaggle’s top 40 most popular datasets
- **US Weather Events** (2016–2022)
- **US Road Construction & Closures** (2016–2021)
- **US Traffic Congestions** (2016–2022)
- **CityTrek-14K** — Large-scale dataset of 14,000 car trajectories