Mahmoud Jahanshahi

Email: mj@mahmoudjahanshahi.com   LinkedIn: [mahmoudjahanshahi](#)
Personal Website: https://mahmoudjahanshahi.com   GitHub: [mahmoudjahanshahi](#)

Profile

Highly reliable and experienced **data scientist** with a proven track record in mining open-source software repositories. Extensive experience in business intelligence consulting, financial and investment analysis, and project management across various industries, particularly telecommunications. Familiar with applying machine learning, deep learning, and natural language processing techniques, with proficiency in statistical analysis, to extract insights and drive data-informed decisions. Recognized for key competencies including:

- Advanced analytical and problem-solving capabilities
- Ability to tackle complex, intellectually demanding problems with minimal guidance
- Rapid learning and adaptability in dynamic environments
- Leadership in cross-functional teams, with strong communication and organizational skills

Professional Experience

University of Tennessee Knoxville
Graduate Research Assistant   **May 2021 – Present**
Department of Electrical Engineering & Computer Science

- Leveraged Natural Language Processing (NLP) and Machine Learning (ML) to research Open Source Software supply chains, focusing on copy-based code reuse and its implications for software quality, reliability, and development efficiency.
- Designed an automated source code autocuration technique to improve the quality of pre-training datasets for Large Language Models (LLMs), addressing issues related to bugs, vulnerabilities, and license violations in generated code.
- Served as a key contributor to the [World of Code](#) infrastructure, enabling scalable data curation, augmentation, and repository analysis for over 130M software projects.

Mobile Communications Company of Iran (Hamrahe Aval)
Senior Data Scientist   **May 2019 – Apr 2020**
ERP Department – Data Warehousing & Business Intelligence Office

- Acted as a liaison between business units and data warehouse teams, translating business needs into technical requirements for the design, testing, and development of warehouse solutions.
- Developed and maintained reports, dashboards, and analyses using Oracle Business Intelligence Enterprise.

Mobile Communications Company of Iran (Hamrahe Aval)
Strategic Investments Lead   **Feb 2018 – May 2019**
Investment Department (M&A) – Feasibility Studies Office

- Led a team of five professionals in managing complex investment projects, including business plan development, feasibility analysis, valuation of acquisition targets and due diligence.
- Collaborated with C-level management and the board of directors to negotiate contract terms and cooperation models with potential counterparts, such as strategic partners or business sellers.

Mobile Communications Company of Iran (Hamrahe Aval)
International Investment Analyst   **Feb 2016 – Feb 2018**
Investment Department (M&A) – International Investments Office

- Developed financial models to project earnings and assess the viability of acquisition opportunities in international markets.
- Screened markets for potential investment targets, interpreting financial statements and other data to evaluate risk and profitability.

Skills

- **Programming Languages:** Bash, R, Python, C
- **Data Analysis and Visualization:** SQL, MongoDB, Tableau, Power BI, MATLAB
- **Other Tools:** MS Project, COMFAR
- **Languages:** English (Fluent), German (Working Knowledge), Persian (Native)

Mahmoud Jahanshahi

Email: mj@mahmoudjahanshahi.com   LinkedIn: mahmoudjahanshahi
Personal Website: https://mahmoudjahanshahi.com   GitHub: mahmoudjahanshahi

---

Education

PhD - Computer Science   **May 2021 - May 2025 (Expected)**
University of Tennessee, Knoxville, USA

- Dissertation: Copy-Based Reuse and its Implications in Open Source Software Supply Chains
- Supervisor: Dr. Audris Mockus

MSc – Industrial Engineering – Industrial   **Sep 2011 – Sep 2013**
Sharif University of Technology, Tehran, Iran

- Ranked 6th nationwide in industrial engineering Graduate University Entrance Exam
- Thesis: The Influence of Information Presentation and Risk Attitude on Asset Allocation in Financial Markets
- Supervisor: Dr. S. T. Akhavan Niaki

BSc – Industrial Engineering – System Planning and Analysis   **Jan 2007 – Jul 2011**
Mazandaran Institute of Technology, Babol, Iran

Certificates

- **Deep Learning Specialization** by DeepLearning.AI on Coursera   **Dec 2024**
  - Structuring Machine Learning Projects
  - Neural Networks and Deep Learning
  - Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization
  - Convolutional Neural Networks
  - Sequence Models

Publications

- **Jahanshahi, M**., Reid, D., McDaniel, A., & Mockus, A. "OSS License Identification at Scale: A Comprehensive Dataset Using World of Code". *Accepted in 2025 IEEE/ACM 22st International Conference on Mining Software Repositories (MSR)*. IEEE, 2025.

- **Jahanshahi M.**, Mockus A. "Cracks in The Stack: Hidden Vulnerabilities and Licensing Risks in LLM Pre-Training Datasets". *Accepted in the Second International Workshop on Large Language Models for Code (LLM4Code)*. 2025.

- Miller, C., **Jahanshahi, M.**, Mockus, A., Vasilescu, B., & Kästner, C. "Understanding the Response to Open-Source Dependency Abandonment in the npm Ecosystem". *Accepted in the 47th International Conference on Software Engineering (ICSE)*. 2025.

- **Jahanshahi, M.**, Reid, D., & Mockus, A. "Beyond Dependencies: The Role of Copy-Based Reuse in Open Source Software Development". *arXiv preprint arXiv:2409.04830*. 2024.

- **Jahanshahi, M.** & Mockus, A. "Dataset: Copy-based Reuse in Open Source Software". *2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR)* (pp. 42-47). IEEE, 2024.

- Reid, D., **Jahanshahi, M.**, & Mockus, A. "The extent of orphan vulnerabilities from code reuse in open source software". *Proceedings of the 44th International Conference on Software Engineering (ICSE)* (pp. 2104-2115). 2022. **Nominated for ACM SIGSOFT Distinguished Paper Award.**

- Lyulina, E., & **Jahanshahi, M.** "Building the collaboration graph of open-source software ecosystem". *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)* (pp. 618-620). IEEE, 2021.