Nima Sarajpoor

Ontario, Canada

Github: https://github.com/NimaSarajpoor

nimasarajpoor@gmail.com

LinkedIn: https://www.linkedin.com/in/nimasarajpoor1991/

+1 (587) 284 8089

About me

With a PhD degree in Engineering, specializing in data mining, I possess a robust foundation in mathematics and statistics. Currently serving as a full-time Data Scientist at Financial Corporation Manulife, I contribute my expertise to drive insightful data-driven solutions, maintain projects on production, and do code review. In addition, I am one of the maintainers of STUMPY, a open-source data analysis tool using matrix profile algorithm at its core. My versatile skill set encompasses proficiency in Python, SQL, Shell scripting, and Git, coupled with a profound understanding of Machine Learning, both in supervised and unsupervised contexts. I am experienced in refactoring and unit testing. You can read more about me on my LinkedIn / GitHub profiles.

Highlights (hyperlinks included)

- **Data Science and System design (Fraud Detection Section)**

  - **Developed business-driven classifier for risk management**, added unit tests and data quality checks, worked on new Cloud platform, maintaining relationships with the business, conducted several sessions to gain a thorough understanding of the problem before delving into the solution. Optimized the algorithm to reduce the number of times that a tabular data needs to be read with help of a clustering trick

  - **Enhanced and maintained 3+2 projects on Fraud Detection**, Took responsibility for maintaining and enhancing three main projects and provided support for two additional projects on an intermittent basis. Addressed low data quality issues by redesigning the analytic process. Proposed a new system design to enhance the fault tolerance of the project and conducted a thorough review of the code base to identify potential opportunities for reducing false positives.

  - **Cloud platforms**, Supported both the Machine Learning Engineers and the consulting company in migrating projects from one Cloud platform to another. Provided code review and clear suggestion and guidelines on how to improve the readability of the code for better maintainability.

  - **Worked on new ideas**, On the side, worked on setting up the first GitHub Actions workflow to run tests with each push, ensuring prompt detection of any faulty code. Experimented with fetching metadata from the project's GitHub repository to explore opportunities for establishing a unified process for reporting and tracking issues.

  - **Maintained a good relationship with my teammates and the business users**, Known for flexibility, I adapted quickly to new environments and fostered constructive relationships with my colleagues, contributing to an enhanced sense of teamwork.

- **Machine Learning**

  - **Experienced with several libraries in Python**: TensorFlow, Scikit-learn, Numpy, Pandas, Numba, and spark.

  - **Experienced with supervised and unsupervised machine learning models** (see my journal and conference papers in My Google Scholar

- **Developed support for feature group in the machine learning tool MLXTEND** to enhance feature selection module, particularly for categorical data (see CHANGELOG)
- **Implemented arbitrary-length anomaly detection algorithm** while reducing its computing time by providing a warm start (see MERLIN)
- **Developed a major feature in STUMPY tool to support top-k matching patterns in data** (see PR#595 and PR#586), rolled out in v1.12.0
- **Created** RandomForestExplorer package to extract frequent rules from decision paths of RF
- **Created tutorials in machine learning** (see POSTS)

- **Programming**

  - **Proficient in Python**, Invited to join the maintainer community of STUMPY tool, led lectures for a crash course on python with more than 50 participants
  - **Strong understanding of data structure and algorithms**, ranked 5th in Programming Contest (CCPC).
  - **Experienced with refactoring** (see PR#656), **unit testing** (see PR#657), and **profiling** (see Fast Fourier Transform)
  - **Experienced with version control Git, GitHub, and GitHub Copilot**
  - **Experienced with Shell scripting (see Note)**
  - **Experienced with writing SQL scripts**
  - **Familiar with CSS and HTML** (see my blog project)

- **Math**

  - **Strong quantitative reasoning**, verified an equation after fixing the typo (see the long proof here: VALMOD notebook)
  - **GPA 4.0/4.0 on university-level courses on mathematics**
  - **Took advanced courses:** "data mining [algorithms]", "applied data science", "probability and stochastic process", and "computational statistics." (GPA: 3.85)

Experiences (in Academia and Industry)

- **Data Scientist in Manulife** — Toronto, ON
  Data Scientist in Group Advanced Analytics — *Jan 2023 - Current*
- **Maintainer Collaborator in Stumpy, a data analysis tool** — OSS @ GitHub
  Maintaining the tool by adding new features and fixing bugs — *Dec 2023 - Current*
- **[Top] Contributor in Stumpy, a data analysis tool** — OSS @ GitHub
  Developing / Enhancing features — *2021 - Dec 2023*
- **Python Lecture Lead at Schulich Ignite** — Alberta
  Python instructor for both beginner and advanced level — *2022*
- **Research in data mining area at University of Calgary** — UofC, Alberta
  Developed unsupervised learning techniques for categorizing patterns in energy data — *2017 - 2022*

Education

- **University of Calgary** — Alberta, Canada
  PhD, Dept. of Electrical and Software Engineering (Analyzing energy data) — Grad: Dec 2022
- **Sharif University of Technology (SUT)** — Tehran, Iran
  M.Sc., SUT is ranked 1st among Iranian Universities in Science and Engineering — 2013 - 2015
- **University of Tehran (UT)** — Tehran, Iran
  B.Sc, UT is ranked second among Iranian Universities in Science and Engineering — 2009 - 2013

Selected Achievements and Awards (hyperlinks included)

- **Full-Stack Machine learning Certificate** — Uplimit
  Hands-on project for end-to-end process and tracking experiments — 2023
- **Presented a workshop in Toronto Data Workshop** — Toronto, ON
  Presented a 1-hr workshop on pattern analysis at UofT — 2023
- **Programming Contest (CCPC)** — Alberta, Canada
  Ranked 5th in Calgary Collegiate Programming Contest — 2022
- **Software Development** — OSS
  Top contributor in two open-source tools for data analysis STUMPY and MLXTEND — 2022
- **Gordon Lewis Hedberg Doctoral Scholarship** — Alberta, Canada
  Student with Excellent Grades — 2021
- **Deep Learning Certificates** — Deeplearning.ai
  Cetificates: DL basics, hyperparameter tuning, CNN, ML project structure in DL — 2021
- **Ph.D. National Competition Award** — Tehran, Iran
  Ranked 1st among more than 1000 participants of PhD National Exam — 2016
  (Afterwards, I applied to Canada.)
- **M.Sc. National Competition Award** — Tehran, Iran
  Top 10 among more than 5000 participants of M.Sc. National Exam — 2013

Publications

See my publications here: My Google Scholar)

Voluntary Experiences

- **Involved in evaluating candidates in interview** — Toronto, Canada
  Manulife — 2023
- **Moderator of Employees Engagement Sessions** — Toronto, Canada
  Manulife — 2023
- **President / Vice President of Grad Students** — Alberta, Canada
  Department of Electrical and Software Engineering — 2017-2020
- **Evaluator** — Alberta, Canada
  IEEE award committee, Graduate Student Association — 2017-2018