

Shiva Soleimany

📍 Toronto, Ontario, Canada
🏛️ University of Alberta, Borealis AI
☎️ +17802206534

✉️ soleiman@ualberta.ca
🔗 <https://shivasoleimany.github.io>
🌐 [linkedin.com/in/shiva-soleimany](https://www.linkedin.com/in/shiva-soleimany)

Highlights

I am passionate about data in all forms. I have hands-on experience working on and maintaining data pipelines and implementing predictive modeling and carrying out data science exploration in challenging environments. Furthermore, I have implemented multiple NLP, Computer Vision, and RL projects.

Work Experience

March 2023 MARCH 2022	Data Engineer at BOREALIS AI, Toronto, Canada <i>Xenon (PySpark, OpenShift, Jenkins, Delta Lake, S3, PureStorage, Superset, Grafana)</i> <ul style="list-style-type: none">• Run, maintain, and support an ETL pipeline with three stages of clean and partition, normalization parameters calculation, and indexing in order to process the received data from RBC with more than 15 million records and providing it to the research team• Adding new features to the pipeline such as appending new features (in a distributed manner and without requiring additional memory), and data profiling• Leading the migration of Xenon and its Jenkins integration tests from OCP3 to OCP4• Checking data quality with automated quality scripts cron jobs and monitoring the results on Superset <i>Anahit (SQL, OpenShift, Jenkins, Airflow, Kubernetes)</i> <ul style="list-style-type: none">• Expanding Anapipes to pull data for additional datasets from data warehouse• Migrating Anaflow to OCP4, and optimizing it by reducing the number of pvcs <i>RBC Brain (AWS SageMaker, Jenkins, S3, Docker)</i> <ul style="list-style-type: none">• Working on a POC project in collaboration with RBC Brain to develop an ML platform that enables researchers to develop their models faster by reducing the needed time for environment setup. Improving the platform by adding features such as running in local mode, training in FastFile mode, and upgrading poetry packages.
Jan 2020 JUN 2021	Research Scientist Intern at DELPHI TECHNOLOGY CORP, Calgary, Canada Worked on training an agent to perform straight and level and climb tasks in the X-plane flight simulator, analyzing (OCR,...) data of a real pilot's flight, showing that the trained agent achieves comparable results to a pilot, and comparing the performance of students with the performance of the trained agent. <i>(Tensorflow, openAI-gym, OpenCV, Seaborn)</i>
Aug 2018 MAY 2018	Data Scientist at SADRA EYE CLINIC, Isfahan, Iran Worked on data mining and predictive analysis for Retinopathy of Prematurity on a real-world dataset of over 18000 records. Using bagging methods(Random Forest) to tackle the problem of small dataset size and achieving 84% accuracy. <i>(Matplotlib, scikit-learn)</i>
May 2016 MAR 2017	Backend Developer at ITORBIT, Tehran, Iran Back-end developer for the 'NAAD' project, the first and only ERP(enterprise resource planning) system in Iran, designed to manage the information of students and professors in universities. Worked with the real-world dataset of over 4 million records: Decrypting, replacing missing data using imputation methods, compressing, and encrypting. <i>(Java EE, SQL, Wampserver)</i>

Projects

- Telegram anonymous connector: Working with pyTelegramBotAPI, an API for Telegram application, to let two people talk to each other anonymously.
- Difficult words of movies: Parsing subtitle files of movies and finding the difficult words in them using the WordDifficulty dataset.
- Comparing q-sigma and SARSA algorithms in grid world: the goal of the agent is to reach the goal state in grid world, once trained using the SARSA algorithm and once trained with q-sigma algorithm.
- Forest Cover Type Prediction: A data mining classification problem. The goal of this project was to train a classifier that is able to classify(predict) the forest cover types for areas that were not included in its training data.
- Email Network Analysis: Analyzing the internal email communication network between employees of a company (strongly connected components, centers, diameter, radius, etc)
- Emojifier: Adding Emojis to text messages can make them more expressive and the goal of this project is to add appropriate emojis to text messages automatically.
- Artificial Ant with GP: I wrote the code for solving this problem and visualized the of finding the best path. You can see its video on the project's page.
- Failed Servers Detection: Using Gaussian distribution for anomaly detection in order to detect servers with unnatural behaviors in a network.

Education

SEPT 2019 Master of Science in COMPUTING SCIENCE,
DEC 2021 **University of Alberta**, Edmonton, Canada
Thesis: Improving Transfer Learning using Similarity Metrics
Related courses: Machine Learning, Deep Learning for NLP, Reinforcement Learning I, Reinforcement Learning II

SEPT 2013 Bachelor Degree in SOFTWARE ENGINEERING
SEPT 2018 **University of Isfahan**, Isfahan, Iran
Thesis: Facial Expressions Recognition
Related courses: Data Mining, Data Structures, Algorithm Design, Database Design, Object-Oriented programming and design, Statistics and Probability, Evolutionary Computation, Advanced Programming, Software Engineering

Skills

Programming and Scripting: Python, PySpark, C++, Java, HTML5, CSS3, Bootstrap, JavaScript, YAML
Libraries: Sklearn, PyTorch, Tensorflow, Numpy, Scipy, Pandas, NLTK
Tools: Airflow, Delta Lake, Jenkins, Docker, CUDA, MySQL, Git, AWS (SageMaker, S3, Glue), OpenShift, SQLite, Linux, Tableau, Great Expectations, Grafana, PureStorage

Publications

- **S. Soleimany Dizicheh**, K.Bagheri: *A Note on the Euclidean Algorithm*. Journal of Mathematics and System Science, 2018.