AMIRHOSSEIN RAZAVI

Email: razaviah77@gmail.com, Phone: +1 416-722-4080, Homepage: razaviah.github.io

RESEARCH INTERESTS

Data Science, Applied Machine Learning, Deep Learning, Machine Learning Operations, Natural Language Processing, Reinforcement Learning, Knowledge Representation and Reasoning, Information Retrieval

EDUCATION

York University *2018 - 2022*
Bachelor of Science, Honours Computer Science - Lassonde School of Engineering
GPA: **7.12/9 (B+)**
Concentrated on NLP, Artificial Intelligence, and Machine Learning

Notable Courses:

Computer Science Capstone Project: A+
Artificial Intelligence: A
Design and Analysis of Algorithms: A
Introduction to Applied Statistics: A+
Machine Learning and Pattern Recognition: A+
Introduction to AI and Logic Programming : A
Advanced Object Oriented Programming: A
Elementary Probability: A+

International College of Manitoba *2017 - 2018*
Bachelor of Science, Computer Engineering (Transferred to York University)
GPA: **(4.13/4.5) (A)**

Notable Courses:

Computer Programming for Scientists and Engineers: A+
Calculus 2: A+

PROFESSIONAL TRAINING

Toronto Institute of Data Science & Technology *Sep 2022 - Mar 2023*
Applied Data Engineering and Architect Diploma - WeCloudData

Notable Courses:

AWS Services
Data Warehouses
Apache Airflow
SQL and NoSQL
CI/CD Pipelines
Data Integration, and Data Pipelines
Big Data and Spark
Docker
Data Ingestion, Streaming, and Lakehouse
Linux

RESEARCH AND ACADEMIC EXPERIENCES

Dataset Generation, and Evaluation for Natural Question Generation

Supervisor: Prof. An
Type: Capstone Project, Natural Language Processing (NLP)

This project aimed to create and evaluate a new dataset for natural question generation, emphasizing the practical use of natural-sounding questions in real-world applications. Utilizing state-of-the-art NLP models like BART and T5, the project focused on generating and validating a dataset that effectively aids in natural question generation tasks.

Enhancing Customer Insights in Wholesale Distribution

Supervisor: Prof. Garg
Type: Data Analysis, Machine Learning

Focusing on a wholesale distributor's client data, this project employs advanced techniques like EDA, KMeans clustering, and PCA for deep customer behavior analysis. It explores feature scaling methods and RFECV for feature optimization, and leverages XGBoost for predictive modeling, offering a comprehensive view of customer segmentation in wholesale distribution.

Exploring Red Wine Quality through Data Science

Supervisor: Prof. Garg
Type: Data Science, Machine Learning

This project applies regression and classification models on the Red Wine Quality Dataset to uncover the key chemical factors that indicate quality. It aims to bridge the gap between subjective human tasting and objective chemical analysis, providing insights into wine quality assessment.

Advanced Four in a Row Game Simulation

Supervisor: Prof. Garg
Type: Artificial Intelligence, AI Agents

This project combines GUI development with AI strategy implementation for 'Four in a Row'. It features an intuitive graphical interface and robust AI agents, designed using algorithms like minimax and alpha-beta pruning, to simulate an immersive and strategic board game environment.

USA Map Navigation using Advanced Algorithms

Supervisor: Prof. Garg
Type: Artificial Intelligence

This project showcases the implementation of Uniform Cost Search, Breadth-First Search, Depth-First Search, and A-Star Search to navigate a USA road network. It demonstrates the effectiveness of these algorithms in optimizing routes between cities, offering insights into algorithmic efficiency and geographic problem-solving.

Evaluating Clingo ASP Solver

Supervisor: Prof. Lesperance
Type: Artificial Intelligence

This project delves into the capabilities of the Clingo Answer Set Programming system, exploring its effectiveness in handling and solving complex combinatorial and logic-based problems. The focus is on assessing Clingo's modelling language and problem-solving approach, demonstrating its application in various computational scenarios.

Social Network Influencer Ranking Based On Link Analysis

Supervisor: Prof. Papagelis
Type: Network Analysis

This project explores the application of link analysis methods, such as PageRank and weighted PageRank, to rank influencers on social networking services. Focusing on the burgeoning field of digital marketing and online communication, it aims to provide a systematic approach to measuring the influence and reach of key SNS personalities.

PERSONAL PROJECTS

Real-Time Time Series Data Prediction (Under Development)

Supervisor: N/A
Type: Data Engineering, Deep Learning, Real-Time Data Processing

This project, currently in progress, utilizes Apache Flink, Kafka, and deep learning to analyze and predict trends in time series data. In the model research phase, its goal is to dynamically integrate live data for immediate, accurate forecasting, advancing the field of real-time data analysis.

TTC Real-Time Bus Tracking System

Supervisor: N/A
Type: Data Engineering, Real-Time Data Processing

This project develops a real-time application for tracking TTC buses using GPS data from IoT devices. It harnesses Apache NiFi, Kafka, Spark Structured Streaming, and other technologies to process live data, aiming to optimize TTC bus routes.

Automated Data Pipeline for Transaction Analysis

Supervisor: N/A
Type: Data Engineering, Batch Data Processing

This project automates the extraction of transaction data into a cloud-based storage system, followed by its transformation using a scalable compute service. Post-transformation, the data is stored for further analysis and visualized using a business intelligence tool, enabling detailed reporting and analysis of transactional data.

WORK EXPERIENCE

Data Engineer Consultant @ Beam Data Sep 2022 - Present
Toronto, Canada

Healthcare Data Migration

Migrated a midwifery practice's data to private cloud platform using Apache Airflow, automating data ingestion and establishing advanced transformation rules. This project streamlined healthcare data management and resulted in a scalable pipeline for future migrations, underscored by stringent data quality checks to ensure data integrity.

Automated Data Pipeline for SEO Product

Developed a comprehensive website auditing tool for SEO performance tracking, integrating a robust Apache Airflow data pipeline for daily updates and historical records. This included creating a Metabase dashboard and optimizing data processes with advanced SQL in PostgreSQL, culminating in recognition from the client and a leadership role in training juniors.

TECHNICAL SKILLS

Programming	Python, SQL, PySpark, Pandas, Shell Script, Git
Data Engineering	Airflow, Docker, Kafka, Apache Spark (PySpark), Databricks, Nifi, Spark Streaming, Jenkins, Hadoop, Snowflake, Flink
Cloud Computing	AWS, Compute Canada, GCP, Azure
Data Visualization & BI	Matplotlib, Seaborn, Metabase, Tableau
Machine Learning Tools	Scikit-learn, PyTorch, Tensorflow, Keras, NLTK

LANGUAGES

English, Persian

AWARDS, HONOURS, AND SCHOLARSHIPS

- Member of Dean's Honour Roll at York University 2022
- York University Undergraduate Bursary 2021
- Winner of The Merit Scholarship at International College of Manitoba 2017
- Ranked top 2% between 163000 applicants in National University Entrance Exam 2016

EXTRA-CURRICULAR AND VOLUNTEER WORK

- Volunteered at the "Iranian Stuttering School" association, actively participating in initiatives, creating awareness, and providing peer support for individuals with stuttering (2020 - Present)

- Fundraising for back-to-school equipment and clothing for low-income students (2018 - 2020)

- Taught first year Math and Computer courses to students with learning challenges (2017 - 2018)

REFERENCES

Aijun An
Professor , EECS Department, York University *aan@yorku.ca*

Jeff Edmonds
Professor , EECS Department, York University *jeff@eecs.yorku.ca*

Others available upon request