

Keivan Mokhtarpour, MSc

(438) 722 5044 | m.keivan73@gmail.com | Toronto, ON
<https://www.linkedin.com/in/keivan-mokhtarpour/>

PROFESSIONAL SUMMARY

Dedicated and Detail-oriented professional with hands-on experience in creating, testing, and maintaining ETL pipelines and building automated pipelines. Demonstrated expertise in identifying process improvement opportunities and executing projects within timely and budgetary constraints. Proven ability to quickly learn, offer innovative solutions, and collaborate with senior data scientists/engineers and clients.

TECHNICAL SKILLS

- **Programming:** Python, SQL, Linux, C#, .NET, PySpark
- **Big Data:** Elasticsearch, OpenSearch, Snowflake, Spark (Databricks)
- **CI/CD:** Git, Airflow, Docker, Kubeflow
- **Cloud Services:** AWS Services (EC2, S3, Kinesis, EMR, Lambda, Athena, CloudWatch), Informatica IICS
- **Data Visualization and Business Intelligence:** Kibana, Tableau, Matplotlib, Seaborn, Quick Sight
- **Machine Learning:** Scikit-learn, Pandas, NumPy, Classification, Regression, Computer Vision, Neural Network
- **Project Management:** Azure DevOps, Trello

EXPERIENCE

Data Engineer

May 2022 - Current

Mind in a Box, Montreal, QC – Remote

- Spearheaded the design and implementation of a high-performance, real-time data pipeline using Python, Logstash, and .NET. Integrated and processed over +100 million raw records from +30 diverse data sources.
- Employed SQL querying techniques, Google Trends API integration, and the OpenSearch .NET package for efficient data ingestion and harmonization from disparate sources.
- Migrated data from Elasticsearch to Snowflake using Informatica IICS, ensuring seamless migration, data integrity, and optimal performance. Leveraged Snowflake as a scalable cloud data warehouse for advanced analytics and reporting.
- Engineered sophisticated data views in Elasticsearch Kibana and Power BI, enabling stakeholders to derive valuable insights for informed decision-making.
- Orchestrated the deployment of the data pipeline using Kubeflow, ensuring scalability, reliability, and efficient resource management.
- Led automation initiatives by implementing cron jobs and leveraging Kubeflow to streamline ETL processes, reducing manual workload and enhancing operational efficiency.

Software Developer

Aug 2021 – May 2022

PolyControls Technologies Inc., Montreal, QC

- Demonstrated expertise in debugging and resolving critical issues within C# .NET desktop applications, specifically utilizing XAML UI frameworks.
- Leveraged advanced debugging tools and techniques to efficiently identify and address complex issues, resulting in a remarkable 40% reduction in error resolution time.
- Employed comprehensive knowledge of C# .NET and XAML to troubleshoot and rectify application defects, ensuring optimal performance and user experience.
- Collaborated closely with cross-functional teams to investigate and diagnose software malfunctions, applying analytical thinking and problem-solving skills to deliver timely resolutions.
- Proactively implemented preventive measures and performed code reviews to mitigate future issues and enhance application stability and reliability.

Keivan Mokhtarpour, MSc

Data Science Consultant

Nov 2020 – Aug 2021

Beam Data, Toronto – Remote

- Developed data streaming pipelines using Twitter API and Amazon Kinesis to process real-time data at scale.
- Designed and implemented complex ETL pipelines using PySpark on Databricks for processing large-scale datasets.
- Developed Python Selenium web scraping scripts to extract and parse data from multiple websites, increasing data acquisition efficiency by over 50%.
- Defined task dependencies and concurrency for Dags within Airflow to improve data injection efficiency.
- Collaborated with cross-functional teams to gather requirements, analyze data sources, and design optimal pipeline architectures.
- Conducted testing and performance optimization to ensure reliable and efficient data processing.

DATA PROJECTS

Twitter Sentiment Analysis | Self-Guided Project

- Streamed data with Twitter API using Kinesis firehose (over 3 million rows of data) to analyze sentiment of tweets regarding Afghanistan situation,
- Built an ETL pipeline using PySpark in Databricks for data preprocessing, modelling preparation and save clean data to S3 bucket,
- Designed and trained an NLP model to predict the sentiment of tweets using the text blob library for labeling and dumped data with tags back into S3,
- Implemented data analytics pipeline with Athena using SQL queries to connect database with QuickSight, converted data into actionable insights and prepare visualization and dashboard reports that interpreting volume of twitter posts, sentiment in different countries.

Web Scraping Booking.com & Data Analysis | Self-Guided Project

- Utilized web scraping techniques (Beautiful Soup and Selenium libraries) to extract Hotel information in Canada for the period of (Dec 22~29),
- Implemented ETL pipeline injecting web data directly to MySQL database for data storage and wrangling,
- Conducted data cleaning and data visualization with Matplotlib, Seaborn and Tableau. Compared the prices and services in different provinces and cities.

Instacart Dataset Analysis & Machine Learning | Self-Guided Project

- Trained, tested, and evaluated a model for anonymized data on customer orders over time to predict which previously purchased products will be in a user's next order,
- Imported raw data from 6 sources into MySQL, and extracted, merged, preprocessed using Python, and Pandas,
- Performed feature engineering, created machine learning pipeline using TensorFlow for predication,
- Increased F1 score by 18% after engineering appropriate features using NumPy and SciPy.

EDUCATION

Data Science and Big Data Diploma, 2021 | Toronto Institute of Data Science, Toronto, ON

Master of Science Mechanical Engineering, 2020 | Concordia University, Montreal, QC

Bachelor of Science Mechanical Engineering, 2016 | Tehran Polytechnic, Tehran, IRAN