Keivan Mokhtarpour, MSc

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

PROFESSIONAL SUMMARY

Committed data engineer with a wealth of hands-on expertise in crafting, validating, and upholding ETL pipelines, while also constructing automated systems. Displaying adeptness in recognizing avenues for process enhancement and effectively executing projects within defined time and financial parameters. Established track record of rapid learning, inventive problem-solving, and productive teamwork with senior data scientists, engineers, team leads and clients.

TECHNICAL SKILLS

- **Programing**: Python, SQL (MySQL, MS SQL Server), Linux, C#, .NET, PySpark
- **Data Engineering**: Elasticsearch, Airflow, Kubeflow, Docker, Git, Apache Spark, Databricks
- **Cloud Services:** AWS Services (EC2, S3, Kinesis, EMR, Lambda, Athena, Quicksight, CloudWatch)
- **Data Visualization and Business Intelligence**: Kibana, Tableau, Matplotlib, Seaborn, Quick Sight
- **Machine Learning**: Scikit-learn, Pandas, NumPy, Classification, Regression, Computer Vision, NLP, LLM, Neural Networks
- **Project Management**: Azure DevOps, Trello

EXPERIENCE

Data Engineer   **May 2022 - Current**
Inmind Technologies Inc., Montreal, QC – Remote

- Designed and implemented a real-time data pipeline leveraging Python, Logstash and .NET. Effectively ingested client data into Elasticsearch/AWS Opensearch data lake indices with appropriate shards and replicas allocation,
- Operationalized a linkage between Elasticsearch and Snowflake to amplify the potential for enriched data analysis,
- Orchestrated the deployment of data pipelines using Kubeflow/Airflow and integrated all alerts to Slack,
- Managed data workflows through Docker containers, orchestrating seamless deployment of pipelines across environments,
- Troubleshooted data quality errors in a production environment,
- Engineered dashboards in Kibana and Power BI, enabling stakeholders to derive valuable insights,
- Developed a Streamlit application utilizing PrivateGPT, localGPT and OpenAI LLM APIs, designed to function as a chatbot for customer interactions,
- Crafted machine learning models, such as Light GBM, strategically to facilitate precise and scalable time-series predictions of inventory levels.

Data Science and Engineering Consultant   **May 2021 – May 2022**
Beam Data, Toronto, ON

- Developed data streaming pipelines using Twitter API and AWS Kinesis to process real-time data at scale,
- Designed and implemented complex ETL pipelines using PySpark on Databricks for processing large-scale datasets and stored the outputs in Parquet format,
- 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,
- Documented work progress on a weekly basis using tools such as Jira and Confluence,
- Conducted testing and performance optimization to ensure reliable and efficient data processing.

Keivan Mokhtarpour, MSc

DATA PROJECTS
Twitter Sentiment Analysis | Self-Guided Project

- Streamlined data acquisition through the Twitter API using Kinesis Firehose, collecting over 3 million rows of data for sentiment analysis related to Afghanistan war
- Developed a robust ETL pipeline utilizing PySpark within Databricks to preprocess and prepare data,
- Designed, trained, and implemented an NLP model for sentiment prediction using the Text Blob library for labeling, data with associated tags were then stored in AWS S3,
- Constructed a comprehensive data analytics pipeline employing Athena's SQL queries, connecting the database to QuickSight. This facilitated the conversion of data into actionable insights and the creation of visualizations and dashboard reports, elucidating tweet volume and sentiment across different countries.

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

- Employed advanced web scraping techniques using libraries like Beautiful Soup and Selenium to extract intricate hotel data in Canada
- Established an efficient ETL pipeline for seamless integration of web data directly into a MySQL database, thereby enabling streamlined data storage and manipulation.
- Conducted thorough data cleaning and employed Matplotlib, Seaborn, and Tableau for insightful data visualization. Facilitated comparisons of prices and services across various provinces and cities.

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