

TECHNICAL SKILLS

Data Management: AWS (S3/Kinesis/RDS/DynamoDB/Redshift/Lambda/Glue/API Gateway)
GCP (GCS/BigQuery/Cloud SQL/Compute), Snowflake (warehousing)
PostgreSQL, PySpark, Cassandra, MongoDB, Tecton (feature store for ML)

Programming: Python, Pandas, Airflow, NumPy, SciPy, Jupyter, scikit-learn
Docker, Kubernetes, Git, Jenkins, Javascript

Visualization: Looker, Streamlit, Seaborn, Plotly, matplotlib, D3.js

EXPERIENCE

Data Engineer

Tempo Software

Mar. 2021 - Present

Montréal, QC

- ETL event data (~200 GB/day) from our Atlassian web platform and serve to internal stakeholders.
- Setup multi-cloud (AWS/GCP/Snowflake) pipelines to ingest data from internal and external sources.
- Enhanced existing pipelines' monitoring and orchestration leveraging Airflow and Great Expectations.
- Designed and implemented a feature store using Tecton to support development of ML models.
- Built and optimized ML prediction endpoints to reduce API response time from ~5.5s → <1s.
- Rewrote analytics sales pipeline in SQL (from Python) to improve performance (23m → 6s).
- Created processes to maintain data parity between DWH and CRM tools (Salesforce, HubSpot, etc.).

Data Specialist

AltaML • AlphaLayer

May. 2019 - Feb. 2021

Edmonton, AB

- Built and maintained (CI/CD) batch and streaming ETL pipelines for high-resolution financial data.
- Scaled and improved cloud infrastructure to support the ongoing needs of a fast-paced team.
- Performed EDA on financial and regulatory data acquired via scraping and API sourced pipelines.
- Deployed and monitored ML models with MLFlow; generated reports & dashboards from predictions.
- Developed and maintained microservice REST APIs using AWS API Gateway to meet strict SLAs.
- Gathered data requirements from stakeholders, created detailed data documentation for internal use.

Backend / Platform Developer

DiiTalk Communications

May. 2018 - Sep. 2018

Saskatoon, SK

- Optimized app notification delivery by leveraging ML classification models based on recent user engagement activity to improve interaction and retention by over 40%.
- Increased accessibility to and usability of customer data for internal teams by creating and maintaining databases and creating visualizations in Python.
- Developed a backend service for sending push notifications (message queuing, delivery time optimization, and response handling) reducing third-party service expenses by \$8,000 per year.
- Developed unit tests in Python, Bash, and PHPUnit for Docker containers on AWS EC2 instances.

EDUCATION

University of Saskatchewan

B.Sc. Computer Science; cGPA 91%

Saskatoon, SK, Canada

Sept. 2016 - Apr. 2019

Purdue University

Completed 28 courses towards a Computer Engineering degree; cGPA 3.8

West Lafayette, IN, USA

Sept. 2013 - Apr. 2016

INDEPENDENT PROJECTS

Stock Price Predictor 📈

- Trained models (using Naive Bayes, SVM, k-NN, LSTM, and Linear Regression) for portfolio optimization along with a backtesting engine achieving a high stock price prediction accuracy.

REFERENCES

Navaneeth Kamballur Kottayil

Ph.D., Computer Vision

Senior Machine Learning Developer, AltaML

navaneeth@altaml.com

Chad Langager

B.Comm., Finance

General Manager, AlphaLayer (AltaML-AIMCo Joint Venture)

chad@altaml.com

Rudabeh Meskarin

Ph.D., Statistics & Finance

Data Science Team Lead, AlphaLayer

rudabeh@altaml.com

Further references available upon request.