Yash Jagirdar

✉ jagirdaryash@gmail.com | 📞 647-405-1164 | 📍 Waterloo, ON | 🔗 yash073

Skills

Languages: Python, SQL, R, VBA
Frameworks/Libraries: Pandas, PyTorch NumPy, Scikit-Learn, Matplotlib, Seaborn, SciPy, NLTK, Apache Airflow, Hadoop
Tools/OS: Tableau, SAS, TensorFlow, Power BI, Microsoft Excel, AWS, PostgreSQL

Education

University of Waterloo — Waterloo, Ontario
Management Engineering, BASc. Candidate — *2022 – Present*
- Relevant Courses: Data Structures and Algorithms, Databases and Software Design, Probability and Statistics

Experience

Data Engineer — Jan 2024 - Apr 2024
Kenna Technologies — *Toronto, Ontario*
- Engineered automated data pipelines using SQL to efficiently transfer **100,000+ rows** daily into Tableau dashboards
- Designed data warehouses on AWS Redshift, enabling real-time analytics, resulting in a **75%** improvement in reporting efficiency
- Analyzed over **3TB** of complex datasets with SQL and Python, uncovering patterns that drove data-driven strategies
- Created data quality frameworks with Python and Apache Airflow, reducing errors by **8%** across reporting

Business Intelligence Developer — May 2023 - Aug 2023
Ontario Ministry of Public and Business Service Delivery — *Toronto, Ontario*
- Integrated external systems and APIs to ensure reliable data flow across government operations, minimizing downtime by **28%**
- Optimized ETL pipelines on AWS to convert raw data into actionable insights, driving data-driven decision-making
- Developed scalable, low-code web applications improving performance and enhancing user experience for **5000+ users**

Data Analyst — May 2022 - Aug 2022
Sunlife Financial — *Waterloo, Ontario*
- Designed interactive dashboards in Power BI to track data ingestion trends, reducing analysis time by **67%**
- Developed a cloud-based enterprise solution integrating IoT and AI technologies, increasing operational efficiency for clients
- Collaborated with cross-functional teams to gather feedback and iteratively refine the estimation sheet
- Automated data pipelines by creating Excel VBA macros to transfer datasets to a PostgreSQL database, resulting in a **40%** improvement in process efficiency and reducing manual errors

Projects

Distributed Data Pipeline for Real-Time Analytics
Apache Kafka, Apache Spark, AWS
- Built a distributed data pipeline with Apache Kafka and Spark, processing **500,000+ events** per minute in real-time
- Integrated the pipeline with AWS data lakes and warehouses, optimizing data flow and reducing processing time by **35%**
- Developed AWS QuickSight dashboards for real-time metric visualization

Customer Churn Classification Model for Subscription Services
Python
- Developed a churn prediction model with **90%** accuracy using a dataset of **100,000+** customer records
- Deployed the model in production using cloud services, enabling real-time predictions and reducing churn by **15%** in six months
- Implemented and compared algorithms, selecting the best model based on precision, recall, F1-score, and AUC-ROC

Computer Vision Model
Python, RoboFlow
- Developed a computer vision model using Python and Roboflow to achieve **1st place** in the Toyota Innovation Challenge
- Preprocessed image data using Python by applying adaptive equalization and tripling the dataset through data augmentation
- Trained Ultralytics' YOLOv8 model on Roboflow, leveraging transfer learning for efficient training

Achievements

1st Place in Toyota Innovation Challenge hosted by the University of Waterloo Engineering Faculty
1st Place in Scotiabank Data Science Challenge hosted by the University of Waterloo Engineering Faculty
STEM Fellowship: CCUNESCO Scholarly Communication Award **($1000)**