Sahej Singh Sodhi

647-457-3953 | sssodhi@uwaterloo.ca | linkedin.com/in/sahejsinghsodhi | github.com/Sodhi-S

Technical Skills

Programming & Query Languages: Python, SQL, R, C++
Data Analysis & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Plotly, Power BI
ML & AI: Scikit-Learn, TensorFlow, Keras, NLP, OpenaiAPI
Big Data & Databases: Apache Spark, MySQL, MongoDB, Google Cloud Platform, Google Big Query

Experience

Data Science Intern Sep. 2024 - Dec. 2024
Epoch *San Francisco, CA*

- Reduced client report creation time by **10 hours a week** by building an internal web app in **Python** (**Streamlit**, **Pandas**) to automate data preprocessing tasks
- Achieved **98% accuracy** with a **logistic regression model** (**Scikit-Learn**) to classify roles into departments, streamlining HR analytics and reducing manual classification
- Cut client feedback analysis time by **3 hours a week** by integrating the **OpenAI API** for sentiment analysis, enabling faster insight generation for decision-making
- Delivered accurate and engaging client reports (**weekly, monthly, quarterly**) by applying **SQL**-based data aggregation and visual design (**Figma**, **Excel**)

Software Engineering Intern May. 2025 – Aug. 2025
Epoch *San Francisco, CA*

- Improved analytics performance by **40%** by designing **ETL pipelines** to ingest event data into **SQL databases**
- Refactored backend services from **Flask** to **FastAPI**, enabling asynchronous data ingestion and cutting average data pipeline latency by **35%**
- Reduced debugging time by **30 minutes per issue** by enforcing schema consistency and automated data quality checks in **SQLAlchemy**, improving reliability of analytics insights
- Cut cloud storage costs by **15% ($30K/year)** by building **Python pipelines** to detect and manage unused data, boosting efficiency and database performance.

Data and Development Intern Jan. 2024 - Apr. 2024
Brain Racers *Toronto, ON*

- Built a monitoring platform for **1500+ students** across **39+ teachers** using **Python** (**Streamlit**, **Plotly**) to visualize performance, enabling real-time insights
- Saved **15 hours** every instance by automating performance measurement workflows with **SQL** and **Pandas**, enabling faster re-performance
- Identified key engagement trends by performing **product data analysis**, driving improvements in student learning

Projects

Go-Phish | React, TypeScript, Django, TensorFlow, MySQL, GroqAPI Nov. 2024 – Nov. 2024

- Achieved **97.8% accuracy** in phishing detection by deploying a **TensorFlow Random Forest**, building the platform with **React** and **Django REST**, and integrating **GroqAPI** and **VirusTotal** for link protection

Speech Emotion Recognition | Python, Numpy, Keras, Librosa, Seaborn, NLP Sep. 2024 – Oct. 2024

- Achieved **85% accuracy** in classifying speech emotions from **1,000+ audio samples** using a **Keras neural network** with features extracted via **Librosa**

Education

University of Waterloo Waterloo, ON
Systems Design Engineering (BASc) *Sep. 2023 – Apr. 2028 (expected)*

Relevant Courses: Data Structures and Algorithms, Digital Computation (C++), Digital Systems