Faiz Khwaja

• Boston, MA • khwajafaiz06@gmail.com • github.com/faiz2399 • linkedin.com/in/faizmkhwaja • +1 (857)- 437-9745

EXPERIENCE

Staples Inc., Associate Data Scientist –Framingham, MA. **June 2023 – Present**

- Developed and implemented data-driven solutions for industrial engineers, resulting in a **25% improvement** in fulfillment center design efficiency and a **15% reduction** in operational costs.
- Engineered a resilient data pipeline using Azure Data factory, enabling **hourly integration** of warehouse data into Snowflake for improved operations.

Staples Inc., Data Scientist – Supply Chain Co-op, Framingham, MA. **July 2022 – January 2023**

- Performed Data Wrangling and developed a **Power BI** dashboard that resulted in a **40% improvement** in inventory utilization.
- Built an application using Django and SQL that provided strategic recommendations which helped **free up 15%** space in the inventory.
- Held Bi-weekly meetings with Fulfillment Centers to help managers utilize the above applications to achieve maximum benefit.
- Conducted Data Analysis to analyze the inaccuracies in the forecasting model which predicts the daily outbound demand.
- Implemented a Time Series Model (replacing the above model) which helped in increasing prediction accuracy by **3%.**
- Wrote a Python script to promptly update data in PkMS, which is a real-time inventory tracking system.
- Handled Ad-hoc requests which involved writing Advanced SQL queries using **window functions** and **CTEs.**

Cere Labs, Pvt, Ltd., Machine Learning Intern, Mumbai, India. **December 2019 – January 2020**

- Developed a Deep Learning model using TensorFlow to automate text extraction from National ID images of clients.
- Improved the speed of data extraction from **160 seconds to 10 seconds** for a batch of 200 ID cards.
- Deployed the model into the existing Machine Learning pipeline using TensorFlow Serving by working with the DevOps team.

ACADEMIC PROJECTS

Spark Streaming with Twitter API, *Northeastern University* **July 2023**

- Utilized **Apache Spark** Streaming to ingest and process a continuous stream of tweets from the Twitter API in real-time.
- Leveraged lambda functions and **RDD** transformations to cleanse tweets and extract relevant hashtags GitHub: *Spark Streaming.*

Book Recommendation System, *Northeastern University* **October 2022**

- Scraped data using Google Books API.
- Implemented User Based, KNN, Item-Item, Content based, Popularity based models.
- User - Based Neural Network model predicted user's ratings **30% more accurately** than the other models.
- Developed a UI using **Django** that takes in users' information and recommends books. GitHub: *Book Recommendation.*

Movie Reservation Database Management System, *Northeastern University* **October 2021**

- Designed and developed an efficient and scalable database in **MySQL** to store and manage movie reservations.
- Leveraged Flask API of Python to build a user-friendly interface.
- Provided analytics by performing Linear Regression to predict sales during weekdays and weekends. GitHub: *Movie Reservation.*

Drug Classification, *Northeastern University* **October 2021**

- Implemented Decision Tree Classification model to recommend drugs based on users' information with 95% accuracy.
- Designed an end-to-end web application using **Flask API** and deployed it to **Microsoft Azure**. GitHub: *Drug Classification.*

EDUCATION

Master's degree, Data Science **September 2021 – May 2023**

Northeastern University**, Khoury College of Computer Sciences**, Boston, MA **GPA: 3.83**

B. Tech, Mechanical Engineering **June 2017 – June 2021**

University of Mumbai**, Sardar Patel College of Engineering**, Mumbai, India **GPA: 3.8**

TECHNICAL SKILLS

- **Languages**: Python, R, SQL, C, C++, Spark, JavaScript, HTML.
- **Databases:** MongoDB, Microsoft SQL Server, Snowflake, PostgreSQL, Oracle.
- **Tools and Technologies:** Git, Power BI, Tableau, AWS (Lambda, S3), Azure, Postman, Hadoop (MapReduce), Docker Container.
- **Machine Learning Algorithms:** Linear and Logistic Regression, Ensemble Learning, Bagging and Boosting algorithms, SVM.
- **Coursework:** Machine Learning, Database Management System, Algorithms.

PUBLICATIONS

- **Optimization** of Warehouse Management system **using AI and ML** (JETIR) **August 2021**
- **Statistical model** for selection of electrode parameters (Key Engineering Materials) **September 2020**