Saima Ahmed

North York, Ontario | sasaima@gmail.com | 437-556-7897 | [LinkedIn](#) | [GitHub](#)

Driven Data Scientist with a strong foundation in data science and engineering, specializing in Python, SQL, and advanced data management. Proficient in predictive modeling and crafting compelling data visualizations. Experienced in applying machine learning techniques to solve complex problems, I am committed to delivering impactful, data-driven insights that drive strategic decision-making and create measurable value for the organization.

Technical Skills

Programming: Python, SQL, Flask, HTML, JavaScript, Relational Databases, Document-based Databases
Machine Learning: Regression, Classification, NLP, Model evaluation, Clustering, Forecasting
Visualization: Matplotlib, Plotly, PowerBI, Seaborn
Tools & Libraries: Pandas, NumPy, Scikit-learn, Beautiful Soup, Jupyter Notebook
Additional Skills: AWS, Git/GitHub, Microsoft Excel, Web Scraping, Deep Learning

Projects

Harmful Additive Detection in Food Products | *Python, Scikit-Learn, Pandas, NumPy, Streamlit*

- **Developed a classification model with 93% accuracy** to predict health labels based on the presence of harmful additives.
- **Designed an interactive app** that helps users identify harmful chemical additives in everyday food products.
- **Built a data-driven tool** to raise awareness about foodborne risks, promoting safer dietary choices.
- **Enabled proactive identification of unhealthy food products,** supporting public health initiatives and promoting informed consumer choices.

NLP-Based Reddit Post | *Python, Reddit API, NLP, Scikit-Learn, Pandas, NumPy*

- **Scraped 1,000 daily Reddit posts** from various categories using an automated script and the Reddit API.
- **Applied Natural Language Processing techniques** to analyze and preprocess text data.
- **Developed classification models** to predict the subreddit of a post, achieving **85% accuracy** on validation data.
- **Optimized model performance** by experimenting with different classification techniques.

Identifying Key Drivers of Youth Crime | *Python, Scikit-Learn, Pandas, Data Visualization*

- **Built and evaluated predictive models** to analyze factors influencing youth crime rates.
- **Identified key predictors** such as **lack of school enrollment, foster care status, housing value index, and labor force participation** using the **Gradient Boosting Model**, which provided the most accurate results.
- **Uncovered socioeconomic and educational drivers** of youth crime, highlighting areas for potential policy intervention.
- **Developed data visualizations** to communicate findings effectively and support data-driven decision-making.

Professional Experience

Data Science Bootcamp Fellow   Aug 2024 – Dec 2024
General Assembly (Remote)

- A 12-week immersive training focused on mastering industry best practices in data modeling, visual and statistical analysis, and test-driven development, using a Python-driven approach to data science.
- Emphasized the ability to analyze, interpret, and communicate data-driven insights from large datasets while predicting future trends through predictive modeling and pattern recognition.
- Built a portfolio of individual and collaborative projects, including Harmful Additive Detection in Food Products and recommending healthier food alternatives within the same category.
- Applied extensive EDA techniques, Logistic Regression, Gradient Boosting, and Neural Network models, along with Power BI for visualization and data-driven insights.

Career Break   Nov 2022 – Mar 2024

- Relocated to Canada and took a career break to help family settle in.

React Native Developer   Nov 2021 – Oct 2022

Hurdle Health (contract)

- Company acquired by Backpack Health, which integrated its own IT team.
- Improved data handling and system efficiency by working with MongoDB and Node.js for backend development.
- Enhanced app functionality across Android and iOS, ensuring better performance and user experience.
- Redesigned the user interface with new features such as journaling and Zoom integration to improve accessibility and engagement.

Education

FAST-National University of Computer and Emerging Sciences

Master of Computer Science | Sep 2018 – Jun 2021

- Focused on **Natural Language Processing (NLP), Multi-Agent Systems, and Theory of Automata**.
- Thesis: Context-Aware Urdu Transcription Using Transfer Learning with DeepSpeech.

General Assembly | Data Science Bootcamp

Aug 2024 – Dec 2024

- Intensive training in **machine learning, statistical analysis, and data visualization**.
- Developed projects in **classification models, NLP, and recommender systems**.