

Mohd Faizan Siddiqui

Data Analyst | Data Scientist / ML | Python Backend Developer

Lucknow, Uttar Pradesh, India | +91-9517122461 | mohdfaizansiddiqui8090@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Computer Science undergraduate with hands-on experience across the full data-to-product pipeline: analyzing and visualizing data (SQL, Power BI, Tableau, Excel), building and evaluating machine learning models (scikit-learn, NLP, up to 99% accuracy), and shipping backend systems with Python/FastAPI. Comfortable owning work end-to-end — from raw data and schema design through to trained models, dashboards, and tested APIs.

TECHNICAL SKILLS

Programming: Python, JavaScript, Data Structures & Algorithms

Data Analysis & BI: SQL, MySQL, Power BI, Tableau, Excel, Pandas, NumPy, Data Cleaning, EDA, Data Visualization

Machine Learning / AI: scikit-learn, Logistic Regression, Naive Bayes, SVM, Random Forest, XGBoost, NLP (TF-IDF, CountVectorizer), Model Evaluation, LLM/AI tooling basics

Backend & APIs: FastAPI, SQLAlchemy, Alembic (migrations), REST API design, Pytest

Databases: PostgreSQL, MySQL, MongoDB, SQLite

Tools: Git, GitHub, Docker (basics), Postman, VS Code, Jupyter Notebook

INTERNSHIP EXPERIENCE

AI/LLM Intern — Sipher Web Academy (in association with Integral University)

Jun 2025 – Jul 2025

- Cleaned and preprocessed a text dataset of 5,575 messages, engineering features with CountVectorizer and TF-IDF.
- Trained and compared Naive Bayes, Logistic Regression, and SVM classifiers, selecting the best-performing model.
- Delivered a reusable spam-classification pipeline achieving 98% accuracy and 95% precision, designed for integration into email/SMS filtering applications.

Data Science Intern — IBM

Jun 2024 – Jul 2024

- Analyzed customer behavior data (~1,300 records) to identify patterns driving churn, using Python and Pandas for preprocessing and feature engineering.
- Built and evaluated logistic regression and random forest models, reaching 99% accuracy, and translated results into retention-focused recommendations.

PROJECTS

School ERP System — Backend (FastAPI, PostgreSQL, Alembic, Pytest)

2025

- Built a full school-management backend (students, staff, attendance, audit-logged records) with an API-first architecture, migrated PostgreSQL schema via Alembic, and covered core flows with a Pytest test suite.

SMS Spam Detection using AI/LLM (Python, scikit-learn, NLP)

Jun 2025 – Jul 2025

- Benchmarked SVM, Naive Bayes, Random Forest, and XGBoost on 5,575 messages with only 2% spam class balance, achieving 98% accuracy and 95% precision.

Customer Churn Prediction (Python, Pandas, scikit-learn)

Jun 2024 – Jul 2024

- Developed a Logistic Regression churn model on ~1,300 records, achieving 99% accuracy through feature selection and preprocessing.

Netflix Data Analysis (Python, Pandas, Data Visualization)

- Cleaned and explored ~10,000 Netflix records to uncover genre, content-type, and release-trend patterns using Pandas and visualization libraries.

Personal Portfolio Website (HTML, CSS, JavaScript)

- Designed and deployed a responsive developer portfolio, deployed on GitHub Pages.

CERTIFICATIONS

- Machine Learning with Python — IBM · Artificial Intelligence and LLM · Python for Data Science: From Basics to Projects
- SQL Analytics and BI on Databricks · Power BI for Beginners · Excel Dashboard for Beginners — Microsoft
- Data Analytics Job Simulation — Deloitte (via Forage) · Data Science Job Simulation — British Airways (via Forage)
- Data Visualisation: Empowering Business with Effective Insights — Tata (via Forage) · Create Your Own AI Tools Using AI

EDUCATION

B.Tech, Computer Science and Engineering — Integral University, Lucknow

CGPA: 80%

Class XII, CBSE Board (English Medium)

2023 | 63%

Class X, CBSE Board (English Medium)

2021 | 65%

LANGUAGES

English (Read, Write, Speak) · Hindi (Read, Write, Speak) · Urdu (Read, Write, Speak)