Siddharth Prakash Singh

siddharthsingh290601@gmail.com   9838101122   New Delhi, 110025

PROFILE

Data science and analytics professional with hands-on research experience at IIT Delhi, contributing to a major Government of India project involving multi-state fieldwork, data verification, and stakeholder coordination. Skilled in Python, SQL, and machine learning, with a strong background in statistics, predictive modeling, and data-driven insight generation. Adept at managing cross-functional projects and committed to applying analytical expertise to solve complex, real-world problems.

EDUCATION

Master of Science in Banking and Financial Analytics   **2023 − 2025**
Jamia Millia Islamia   *New Delhi, India*
GGPA-9.12
Core Interest: Python, Machine Learning, Data Analytics, Risk Analysis, SQL

Bachelor of Science in Applied Mathematics   **2020 − 2023**
Jamia Millia Islamia   *New Delhi, India*
GGPA-8.67
Core Interest: Linear Algebra, Real Analysis, Modeling and Simulation, C-Programming

SKILLS

Programming Languages
Python, C Programming, SQL

Data analysis & Machine Learning
Pandas, Matplotlib, Jupyter Notebook, Scikit-learn, Excel

Soft skills
Organization, Initiative (Self-starter), Willingness to learn, Design Thinking, Stakeholder Coordination, Cross-functional Collaboration, Field Research, Policy Analysis, Data Verification

PROFESSIONAL EXPERIENCE

Research Intern   **02/2025 − 07/2025**
IIT Delhi (Under Prof. Sanjay Dhir & Prof. Sonali Jain, Strategic Management Division)   *Delhi, India*

- Conducting data collection, reporting, and qualitative analysis for a major Government of India project, developing methodologies for strategic insights.
- Engaged in extensive client-facing fieldwork, coordinating with multiple stakeholders, including state government officials, to contribute to policy research and recommendations.
- Engaged with 6 state Waqf boards to verify and update records of 350+ properties using the WAMSI portal, ensuring data accuracy and compliance.

Project Manager   **11/2021 − 07/2022**
Global Opine Research   *Lucknow, India*

- Collaborate with cross-functional teams to align research objectives with organizational strategies.
- Ensure timely completion of research projects while maintaining quality and accuracy.
- Ensure the accuracy and reliability of collected data and analysis.

PROJECTS

Comparative Analysis of Machine Learning Models for Spam Detection
Technologies Used: SVM, Naïve Bayes, Decision Tree, Random Forest, scikit-learn

- Cleaned and pre-processed text data, including tokenization, stop-word removal, stemming, and vectorization using **CountVectorizer**.
- Implemented machine learning models such as **Support Vector Machines (SVM), Naïve Bayes, Decision Trees**, and **Random Forests** using **scikit-learn**.
- Trained models using preprocessed data, applying **train-test split** for proper validation and evaluation.
- Conducted a comparative analysis of model performance to identify the most effective algorithm for spam.
- **Results:** SVM outperformed all other models with the highest accuracy and F1-score in classifying spam.

Comparative Analysis of ML vs. Statistical Models for Stock Prediction
Technologies Used: Python, Pandas, NumPy, yfinance, scikit-learn, XGBoost, statsmodels, arch (GARCH), Matplotlib, Seaborn

- Collected historical stock price data for Nifty 50, Bank Nifty, ICICI Bank, HDFC Bank, IndusInd Bank, Axis Bank, and SBI using **yfinance**.
- Cleaned and pre-processed the financial time-series data, including handling missing values, feature engineering (lag features, returns, volatility), and scaling for model readiness.
- Implemented forecasting models such as **Random Forest**, **XGBoost**, and a statistical approach combining **Auto-ARIMA with GARCH(1,1)**.
- Trained all models on the processed datasets and evaluated predictive performance using multiple forecast horizons, including **1-day**, **10-day**, and **100-day predictions**.
- Conducted a comprehensive comparative analysis to determine which model performed best across short-term, medium-term, and long-term stock price forecasting.
- **Results:** The ARIMA+GARCH model consistently outperformed Random Forest and XGBoost in 1-day forecasts, while Random Forest delivered superior accuracy for 10-day and 100-day prediction horizons.

Predicting and Classifying Air Quality Index (AQI) Levels Using Linear Regression Model
Technologies Used: Python, pandas, NumPy, scikit-learn, matplotlib, Linear Regression

- Developed **linear regression models** to predict AQI levels based on environmental parameters (e.g., temperature, humidity, pollution levels).
- Evaluated model performance using **Root Mean Squared Error (RMSE)** for regression model.
- Cleaned and pre-processed raw data, handling missing values, outliers, and inconsistencies to ensure data quality for analysis.
- **Results:** The linear regression model achieved an R^2 score of 87%, demonstrating strong explanatory power and reliability in AQI prediction.

CERTIFICATES
IBM SkillsBuild Mastering Data with Machine Learning
Certificate Internship Program