Saqib Alam Ansari

Integrated PhD (IIIT-B)

saqibdatascience@gmail.com | Bangalore | GitHub | LinkedIn

SUMMARY

Data Science professional with 2+ years of experience specializing in Python, Machine Learning, LLMs, and LangChain-based AI applications. Delivered impactful ML solutions, including fraud detection models and Generative AI tools. Demonstrated strong ownership in end-to-end project delivery and collaborated cross-functionally to align technical outputs with business goals. Pursuing Integrated PhD at IIIT Bangalore after cracking GATE 2025, with proven ability to engineer solutions and drive innovation through teamwork, adaptability, and a deep understanding.

EDUCATION

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, BANGALORE (IIITB) **Aug' 2025 - Present**
Integrated PhD (DSAI)
Coursework: Machine Learning, Deep Learning, Computer Graphics

MAHARISHI UNIVERSITY OF INFORMATION TECHNOLOGY **Sep '20 – June'24**
B.Tech. (Computer and Data Science) **CGPA: 9**

PROFESSIONAL EXPERIENCE

Data Scientist Intern
***Zuno General Insurance*, Mumbai** **March 2024 – June 2024**

- Developed a Motor Fraud Detection model with 80 % recall and 40 % F1-score to identify fraudulent users and cases pre-claim, reducing potential fraud costs.
- Developed a custom risk-based encoding technique for fraud detection features, boosting model performance by 30 % compared to the previous benchmark.
- Conducted renewal analytics by evaluating renewal rates across city-level, vehicle model-level, and state-level dimensions, pinpointing top-performing segments for targeted retention strategies.
- Computed loss ratios as the ratio of claims paid to premiums collected, analyzed key loss drivers, and recommended factor-specific solutions to improve underwriting profitability.
- Engineered end-to-end ETL pipelines in Python to ingest and harmonize multi-view insurance data from SAP HANA, underpinning robust feature engineering.
- Automated impurity and anomaly treatment workflows using Python, standardizing data cleansing and eliminating manual preprocessing bottlenecks.
- Designed Power BI dashboards comparing financial-year metrics by state, product, and service category, enabling data-driven allocation of resources and strategic growth decisions.

Data Scientist Intern
***Aspire Pay Later*, Bangalore** **Aug 2022 – Oct 2023**

- Processed and preprocessed bureau data for 120,000 users, implementing automated data-cleaning pipelines that reduced manual intervention and strengthened fraud risk controls.
- Engineered 1,000+ new features from OTP, SMS, payday-loan and fintech logs using NumPy, Pandas and regex, elevating model predictive power and feature diversity.
- Built and validated targeted ML models on a subset of 6,000+ rejected/delinquent users, leveraging Recursive Feature Elimination, Stability Selection and Boruta for optimal feature selection with k-fold cross-validation.
- Converted multi-format datasets (PDF, JSON) into CSV on an AWS compute cluster, accelerating ETL throughput and enabling scalable downstream analysis.
- Augmented data schema with 600+ entities and 4,300+ headers, reorganizing complex data structures to improve query performance and maintainability.
- Conducted in-depth exploratory data analysis, pinpointing high-risk delinquency patterns and informing targeted mitigation strategies that directly reduced fraud exposure.

- Executed end-to-end ML workflows across six distinct growth and performance-optimization projects, ensuring robust

model deployment and continuous monitoring.
- Partnered closely with cross-functional teams, aligning technical solutions with business goals to drive data-driven decision-making and measurable improvements in operational efficiency.

Data Scientist Intern
***iNeuron*, Bangalore** **Feb 2022 – July 2022**

- Analyzed $186.19 M in Amazon sales data to derive actionable insights by region, product, and time period.
- Prepared comprehensive project documentation to standardize workflows and ensure alignment across stakeholders.
- Identified highest- and lowest-performing products to inform strategic planning and inventory decisions.
- Executed end-to-end data preprocessing and analysis in Tableau, streamlining visualization and reporting.

SKILLS

- **Machine Learning & AI:** Scikit-learn, Model Evaluation, TensorFlow, Deep Learning, Hugging Face, LLMs, Ollama,
- **Data Analysis & Visualization**: Statistics, Exploratory Data Analysis, Seaborn, Power BI, Tableau
- **Libraries/Framework:** NumPy, Pandas, Matplotlib, TensorFlow, LangChain
- **Programming Languages:** Python, SQL
- **Database Management:** MySQL, Vector DB
- **Tools/Platform:** Git, GitHub
- **Soft Skills:** Communication, Problem-Solving, Teamwork, Critical Thinking, Adaptability, Time Management

PROJECTS

Cold E-Mail / Cover Letter Generator Using LLMs
Project Link

- Developed AI tool to generate personalized cover letters and cold emails, reducing efforts by 90%.
- Integrated Groq API with DeepSeek R1, LLaMA 3.3, and Gemma2-9B for variety of text generation
- Flexibility of providing job description via URL or direct copy-paste.
- Deployed a fully functional, real-time system with template-based output generation.

End-to-End ML Project for Water Potability

- Performed Preprocessing, EDA, Feature Engineering on data for making it model building ready
- Utilized multiple data preprocessing techniques in imputation, imbalance handling, feature creation etc.
- Performed null values imputation using Linear Regression with non-null columns to predicting null values
- Built ML models like Logistic Regression, Decision Tree, KNN, Random Forest, SVM, Artificial Neural Network
- Performed hyperparameter tuning using Grid Search CV resulting 72% F1 score
- Published a conference paper on this project in IEEE 10th International Conference

PROFESSIONAL DEVELOPMENT / CERTIFICATIONS

- Generative AI with LangChain and HuggingFace **– Udemy** **May 2025**
- Data Science with AI and Machine learning **– upGrad** **Sep'2020 –June'2024**
- Data Science Specialization **– Internshala** **Sept 2022**

ACHIEVEMENTS

- Qualified GATE 2025 examination in the Data Science Branch, secured a position in top 15% of the competition.
- Mentored & Anchored Multiple Fests in College.
- **Published Paper in IEEE** named 'Mean and Prediction Imputation-Based Approach for Predicting Water Potability Using Machine Learning'. LINK