

# Md Aamir Ansari

Ghaziabad, Uttar Pradesh

☎ 8409427781 ✉ amirbxr2003@gmail.com 🔗 linkedin.com/in/aamirasnari072/ 🌐 github.com/aamiransari072

## Professional Summary

**Aspiring Data Scientist** with expertise in predictive modeling, statistical analysis, and scalable machine learning pipelines. Skilled in Python, TensorFlow, PyTorch, SQL, Spark, and Hadoop, with a proven ability to analyze and solve business problems at the root, identify predictors for forecasting, and guide data-driven decision-making to improve customer experience.

## Technical Skills

**Languages:** Python, C++, SQL  
**ML Frameworks:** TensorFlow, PyTorch, MXNet  
**Data Tools:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn  
**Big Data:** Apache Spark, Hadoop  
**Deployment:** FastAPI, Flask, Docker, MLflow, AWS, GCP

## Education

### KIET Group of Institution

B.Tech in Computer Science and Engineering (Artificial Intelligence)- **CGPA** - 8.0/10

**November 2022 – June 2026**

Ghaziabad, Uttar Pradesh

### P.C College

Class 12th (BSEB) - 85.4%

**April 2019 – April 2020**

Buxar, Bihar

## Work Experience

### Machine Learning Intern

Remote

**Yellow Paper**

February 2025 – Present

- Automated document verification using **Google Document AI**, processing 500+ applications/day and cutting manual effort by 60%.
- Built and deployed **ML pipelines** with **FastAPI** for real-time text extraction and validation, boosting accuracy by 35%.
- Designed **data validation protocols** aligned with government standards, improving efficiency by 40% and reducing false positives by 25%.

### AI Developer

New Delhi, India

**TechCurators**

December 2024 – February 2025

- Developed and integrated 4+ AI agents, accelerating **model deployment** by 35% and enabling production-ready solutions.
- Led **data preprocessing and feature engineering** for sentiment analysis across 50+ brands, improving insight extraction by 20%.
- Deployed ML insights via **Flask APIs** to backend systems, enhancing analytics dashboards and cross-team collaboration.

## Projects

### Customer Churn Prediction | Python, Scikit-learn, Pandas

**January 2025 – March 2025**

- Achieved **92% accuracy** with a classification model to predict customer churn across **10,000+ user records**.
- Engineered key features and applied **cross-validation**, improving model **generalizability by 18%**.
- Delivered real-time **churn insights** via dashboard integration, aiding **business retention strategy**.

### Sales Forecasting Pipeline | Python, TensorFlow, Flask, AWS

**January 2025 – March 2025**

- Built an **LSTM-based model** with **90% accuracy** for predicting daily sales in 3 retail regions.
- Deployed on **AWS EC2** using Flask APIs to handle **500+ daily inferences** in production.
- Tuned hyperparameters using **k-fold validation**, reducing forecast error (MAPE) by **22%**.

### Scalable Data Processing with Spark | PySpark, Hadoop

**November 2024**

- Designed distributed **ETL pipelines** using PySpark to process **10M+ records** from raw customer data.
- Reduced processing time by **60%** and enhanced pipeline **scalability and fault-tolerance**.
- Built **validation and logging layers** to ensure data integrity and maintainability in production.

## Awards & Achievements

- Gold Medalist, Turbo AI Challenge, Dr. APJ Abdul Kalam Technical University, 2024 (1st Place)
- Solved 350+ problems on LeetCode, demonstrating strong algorithmic and data structure skills.