Risheek *Bajaj*

Data Scientist | Predictive Modeling & AI

Contact
Phone: +91-9315938758
Email: bajajrisheek012@gmail.com
DOB: 25 February 2000
Gender: Male

Professional Summary

Forward-thinking Data Scientist with ~2 years of experience engineering predictive ecosystems and scalable ML solutions. Specializes in Statistical Learning, Time-Series Forecasting, and Generative AI pipelines. Adept at translating complex mathematical models into business value—optimizing resource allocation and risk management strategies at organizations like Kritko and Wunderman Thompson. Committed to building Responsible AI systems that drive 2026 innovation goals.

Technical Skills

- **Advanced Modeling:** Predictive Analytics, Random Forest/XGBoost, Causal Inference, Time-Series (ARMA/Prophet).
- **AI & Deep Learning:** Computer Vision (CNNs), Generative AI (LLMs/RAG), NLP, Model Evaluation (ROC-AUC).
- **Statistical Analysis:** Hypothesis Testing (A/B), PCA (Dimensionality Reduction), Bayesian Inference.
- **Tech Stack:** Python (Scikit-Learn, TensorFlow, PyTorch), SQL, Streamlit, Docker, Git.

Work Experience

Data Analyst | Kritko April 2024 – Nov 2025 | Remote

- **Predictive Modeling:** Developed production-grade predictive models to forecast business trends, shifting the organization from reactive reporting to proactive, data-driven strategy.
- **ML Pipelines:** Managed end-to-end Machine Learning pipelines, overseeing data ingestion, feature engineering, and model validation to ensure high-fidelity analytical outputs.
- **Exploratory Analysis:** Conducted Exploratory Data Analysis (EDA) on high-dimensional datasets, uncovering latent patterns that redefined client resource allocation and operational planning.

Analyst | Wunderman Thompson Studios July 2023 – Dec 2023 | Gurgaon

- **Statistical Experimentation:** Designed and executed rigorous A/B Tests and statistical validation, providing confidence-scored recommendations that optimized marketing campaign performance.
- **Clustering & Segmentation:** Applied advanced clustering algorithms (K-Means) to segment large-scale customer bases, identifying "High-Propensity" cohorts for targeted intervention strategies.
- **Data Storytelling:** Translated complex statistical findings into interpretable business narratives, empowering stakeholders to make probability-backed decisions.

Certification

- Diploma in Applied Data Science Lab | WorldQuant University
  - Mastered the full data science lifecycle, from Experimental Design and A/B Testing to deploying predictive models via Flask and Streamlit.
- Published Research: "Currency Recognition System" | Springer
  - Lead researcher for a lightweight Android Computer Vision app, optimizing TensorFlow Lite models for real-time accessibility on mobile devices.
- Published Research: "Health Monitoring & Prediction" | Springer
  - Designed a multi-parameter ensemble learning system to predict individual health risks, focusing on maximizing recall for critical medical alerts.

Education

Diploma in AI Lab | Aug 2025 - Dec 2025
Worldquant University

- **Generative AI Focus:** Specialized in implementing Generative Adversarial Networks (GANs) and Stable Diffusion pipelines for synthetic data generation (e.g., medical image synthesis).
- **Computer Vision Architecture:** Designed and trained custom Convolutional Neural Networks (CNNs) and utilized Transfer Learning (ResNet/YOLO) for real-world object detection challenges (e.g., traffic analysis).
- **Deployment:** Built and deployed interactive AI applications using PyTorch, Streamlit, and Flask, bridging the gap between theoretical models and user-facing products

Diploma in Applied Data Science Lab | Mar 2025 - Jun 2025
Worldquant University

- **Statistical Modeling:** Mastered Time-Series analysis using ARMA/GARCH models and conducted rigorous A/B Testing (Chi-square) to validate business hypotheses.
- **Data Engineering:** Engineered end-to-end ETL pipelines to wrangle unstructured data from SQL, NoSQL (MongoDB), and Web APIs into clean analytical datasets.
- **Visualization:** Developed PCA-powered visualizations and interactive Plotly Dash interfaces to communicate complex high-dimensional data insights to stakeholders.

Master of Computer Application (MCA) | Aug 2021 - Sep 2023
Vivekananda Institute of Professional Studies, GGSIPU | CGPA: 9.5/10

- **Core Specialization:** Focused on Algorithm Design, Advanced Machine Learning, and Big Data Technologies, translating theoretical concepts into scalable software solutions.
- **Research & Leadership:** Led multiple capstone research projects and coordinated institutional analytics for NAAC/NBA accreditation, demonstrating a blend of technical research capability and project leadership

Bachelors's in Computer Application (BCA) | Aug 2018 - Sep 2021
Trinity Institute of Professional Studies, GGSIPU | CGPA: 8.7/10

- **Technical Foundation:** Established a robust bedrock in Object-Oriented Programming (Java/C++), Relational Database Management (DBMS), and Discrete Mathematics.
- **Applied Logic:** honed statistical analysis and logic-building skills through simulation projects, laying the groundwork for advanced data science

Projects

1. Financial Transaction Risk Scoring Engine

- Tools: Python, Scikit-Learn, Random Forest, SMOTE, Docker
- Action: Engineered a real-time Anomaly Detection Classifier using Random Forest to identify fraudulent transaction patterns. Implemented SMOTE to synthesize minority samples, effectively resolving class imbalance issues.
- Impact: Achieved a Precision-Recall AUC of 0.92, reducing false positives by 15% and minimizing financial risk exposure through millisecond-level threat flagging.

2. Market Demand & Environmental Trend Forecasting

- Tools: Python, Time-Series (ARMA/GARCH), Statsmodels, NumPy
- Action: Developed a stochastic ARMA Time-Series Model to predict volatile environmental and market demand trends. Conducted rigorous stationarity tests (ADF) and hyperparameter tuning to capture seasonality.
- Impact: Minimized forecast error (RMSE) by 18% compared to baseline averages, enabling stakeholders to shift from reactive mitigation to proactive, probability-backed resource allocation.

3. Patient Risk Stratification & Early Warning System

- Tools: Python, Logistic Regression, Feature Engineering, Healthcare Analytics
- Action: Implemented a Multivariate Logistic Regression Model to calculate personalized risk probability scores for stroke patients. Performed extensive feature engineering on patient vitals to identify key health determinants.
- Impact: Delivered a calibrated risk scoring system that enabled a Data-Driven Triage Strategy, prioritizing high-risk patients for early medical intervention.