YASH MODI

Data Scientist | AI/ML Engineer

Bharuch, Gujarat, India | ydmodi94@gmail.com | +91 87350 45707 | LinkedIn | Portfolio | GitHub

PROFESSIONAL SUMMARY

Data Science professional with a Post Graduate Diploma in Data Science (First Rank) and an M.Sc. in Physics, specializing in end-to-end Machine Learning pipelines, predictive risk Modelling, and statistical analytics. Proficient in Python, SQL, Power BI, Tableau, and Scikit-learn, with hands-on experience building Explainable AI systems (SHAP), Generative AI integrations (Gemini API), and full-stack analytics platforms (React). Delivered self-directed data science projects spanning banking risk analytics, sports intelligence, and manufacturing quality improvement, combining strong quantitative rigor with clear data storytelling and stakeholder communication.

Tech Stack: Python | SQL | Scikit-learn | Power BI | Tableau | SHAP (Explainable AI) | Gemini API (Generative AI) | React.js | Statistical Modelling | Google Cloud Run (GCP)

DATA SCIENCE PROJECTS

Predictive Churn Risk Intelligence System — Retail Banking

Independent Project | 2026

Python | Scikit-learn | SHAP | Gemini API | ETL Pipeline Design | Financial Risk Analytics

- Engineered a full-stack churn-prediction system with an automated ETL (Extract, Transform, Load) pipeline generating 18 engineered features, then benchmarked four classification models — Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting (GBDT) — across five evaluation metrics (Accuracy, Precision, Recall, F1-Score, ROC-AUC) to select the top performer.
- Applied Explainable AI (SHAP – SHapley Additive exPlanations) waterfall visualizations to make model predictions interpretable, built a Capital Flight Risk module flagging deposit accounts at elevated churn risk, and developed an interactive What-If Simulator for real-time client-level churn scenario testing.
- Integrated the Gemini API to auto-generate personalized retention-strategy recommendations per customer risk profile, built a supervisory-style report generator modeled on Chief Risk Officer (CRO) reporting formats used in regulatory-grade banking (e.g., European Central Bank/ECB-style disclosures), and designed a Macro-Liquidity Simulator to estimate potential savings from targeted retention campaigns.

CricAnalytics — Sports Intelligence & Monte Carlo Prediction Platform

Independent Project | 2026

Python | React 18 | Monte Carlo Simulation | Z-Test | ELO Rating | Recharts | Tailwind CSS

- Architected a stochastic match-outcome simulation engine running 5,000 parallel Monte Carlo trials per prediction, using the Box-Muller transform to generate normally distributed outcome samples for probabilistic match forecasting.
- Built a real-time two-tailed Z-Test statistical module analyzing 1,000+ historical Indian Premier League (IPL) matches to quantify the significance of coin-toss advantage, and a self-correcting ELO rating system tracking franchise strength across 19 IPL seasons.
- Designed a points-table forecasting model using Exponential Moving Average (EMA) time-series techniques, and shipped a responsive React 18 / Vite / Tailwind CSS single-page application (SPA) — using AI-assisted "vibe coding" workflows to speed up iteration on physics-based micro-animations and hit sub-100ms UI responsiveness.

Industrial Defect Reduction Analysis

Solo Project | Gold Medal, 1st Rank | 2025

Atlanta Electricals Pvt. Ltd. (Manufacturing) | Lean Six Sigma (DMAIC) | EDA | Pareto Analysis | Root Cause Analysis

- Conducted exploratory data analysis (EDA) and Pareto analysis on real manufacturing defect data across multiple shifts, machines, and raw-material batches to identify and rank the root causes driving the majority of production quality failures.
- Applied the Lean Six Sigma DMAIC (Define-Measure-Analyze-Improve-Control) framework to design targeted, data-backed process-improvement interventions and translate statistical findings into actionable quality-control recommendations.
- Presented a complete data-driven findings report to faculty as sole contributor — awarded Gold Medal and 1st Rank in class for analytical depth, business-impact framing, and presentation quality.

Hazardous Asteroid Detection Model — NASA Open Dataset

Academic ML Project

Python | Pandas | Scikit-learn | Logistic Regression | Matplotlib | Seaborn

- Built an end-to-end binary classification pipeline on NASA's near-Earth object dataset — including preprocessing, feature engineering, and StandardScaler normalization — to predict hazardous asteroid classification using Logistic Regression.
- Prioritized recall as the primary evaluation metric alongside precision, F1-score, and confusion-matrix analysis, to minimize the risk of misclassifying a truly hazardous asteroid as safe.

Also building: *Commercial Real Estate Market Analytics — a property valuation and financial forecasting platform in Python and Excel (in progress, 2026).*

TECHNICAL SKILLS

Programming Languages & Tools: Python, SQL, React.js (JavaScript, Vite, Tailwind CSS), Jupyter Notebook, VS Code, Spyder

Machine Learning & AI Frameworks: Scikit-learn, Logistic Regression, Decision Trees, Random Forest, Gradient Boosting (GBDT), Predictive Modelling, Explainable AI (SHAP), Generative AI (Gemini API), IBM Generative AI, Prompt Engineering

Data Engineering & ETL: ETL (Extract, Transform, Load) Pipeline Design, Feature Engineering, Data Preprocessing, Data Validation & Quality Assurance, Anomaly Detection

Data Visualization & BI: Power BI, Tableau, Matplotlib, Seaborn, Recharts, Interactive SVG Dashboards, Advanced Excel (Microsoft Certified)

Statistical Analysis: Exploratory Data Analysis (EDA), Hypothesis Testing, Z-Test, ROC-AUC Analysis, Partial Dependence Plots (PDP), Monte Carlo Simulation, Box-Muller Transform, Probability Distributions, Statistical Modelling, SPSS, Econometrics

Financial Analytics: Predictive Risk Scoring, Churn Modelling, Financial Modelling, Time-Series Forecasting (EMA), Capital Flight Risk Analysis, Real Estate Modelling & Valuation, ELO Rating Systems, Financial Markets

MLOps & Deployment: Google Cloud Platform (GCP) – Cloud Run, Model Benchmarking & Evaluation, Full-Stack Application Deployment, AI-Assisted Development ("Vibe Coding"), Lean Six Sigma (DMAIC), Agile Analytics, Responsible AI, Data Governance

Soft Skills: Analytical Thinking, Problem Solving, Data Storytelling, Stakeholder Communication, Attention to Detail, Team Collaboration

EDUCATION

Post Graduate Diploma in Data Science *2025 – 2026*

Sardar Patel University, Vallabh Vidyanagar

GPA: 7.7 / 10 — First Rank in Class

Relevant coursework: Statistical Modelling, Machine Learning, Exploratory Data Analysis, Data Visualization, Econometrics, Predictive Analytics

M.Sc. in Physics (Spectroscopy) *2023 – 2025*

Maharaja Sayajirao University, Vadodara

GPA: 6.0 / 10 — Built a rigorous foundation in statistical analysis, quantitative reasoning, and experimental data interpretation.

B.Sc. in Physics *2020 – 2023*

Sardar Patel University, Vallabh Vidyanagar

GPA: 7.35 / 10 — Developed core competencies in calculus, mathematics, data collection, and laboratory analysis.

CERTIFICATIONS

AI & Data Science: IBM Generative AI Engineering Professional Certificate — IBM/Coursera (In Progress); Google Data Analytics Certificate — Google/Coursera; Python for Data Science — IBM/Coursera; Introduction to Artificial Intelligence (AI) — IBM; Google Prompting Essentials Specialization (4 Courses) — Google/Coursera; Python, SQL, Tableau for Data Science — Coursera (In Progress)

Financial & Business: Career in Finance — CFI; Financial Analyst: AI, Excel, and Power BI Skills — CFI/Coursera (In Progress); Commercial Real Estate Modelling — CFI/Coursera (In Progress); Financial Markets — Yale University/Coursera (In Progress)

Productivity & Tools: Work Smarter with Microsoft Excel — Microsoft

ACHIEVEMENTS

- Gold Medal — 1st Rank in class for Industrial Defect Reduction Project, Sardar Patel University (2025)
- First Rank in Class — Post Graduate Diploma in Data Science, GPA 8.2 / 10 (First Semester)
- Built and deployed two self-directed full-stack analytics platforms: a Bank Churn Risk Intelligence System (SHAP Explainable AI + Gemini API) and a Monte Carlo Sports Intelligence Platform (5,000-trial simulations, Z-Test validation)
- Deployed a live portfolio website on Google Cloud Run (GCP), demonstrating hands-on cloud deployment experience
- Completed 6 industry-recognized certifications from Google, IBM, and Microsoft, with 5 additional specializations in progress across Generative AI Engineering, Financial Analysis, and Real Estate Modelling

LANGUAGES

Gujarati (Native) | Hindi (Fluent) | English (Professional)