

Rahul Kushwaha

+91 8209812650 — rahulkushwaha.quant@gmail.com — <https://linkedin.com/in/rahul098>

Experience

Data Science Analyst — UGRO Capital (Mumbai)

Jul 2023 – Present

Model Validation:

- **Performed Comprehensive Model Validation** – Conducted **back-testing and statistical assessments** to evaluate calibration, stability, sensitivity, and discriminatory power of GST and bureau **risk models (i.e. logistic regression)**.
- **Applied Advanced Statistical Techniques** – Utilized metrics such as **KS, Gini, ROC-AUC, and PSI** to measure model performance, assessing its ability to differentiate and maintain stability across varying conditions.
- **Developed Interactive Validation Dashboard** – Built a dynamic **dashboard** to visualize individual test results, providing clear insights into model performance, enabling real-time monitoring, and data-driven decision-making.

Model Development:

- **Feature Selection and data analysis** – Engineered **100+ features** and conducted comprehensive Exploratory Data Analysis (EDA) to transform features, identify key data patterns, and mitigate anomalies through outlier detection.
- **Feature Engineering and model development** – Performed feature selection based on **optimal binning technique, VIF** to reduce **multicollinearity** and **IVs** to ensure optimal model performance. Then, developed **Logistic model** to identify high-risk loan-seeking customers using bank transaction data, achieving a **14%** improvement in Gini score.

Business Analysis:

- **Authored and Published MSME Sampark Analytical Magazine** – Led the **creation and publication** of 3-volumes of MSME Sampark, providing in-depth analysis of the MSME sector, sectoral insights, financial and delinquency benchmarks, and identifying growth opportunities across various industries.
- **Implemented Analytical Methods to Optimize Loan Underwriting** – Developed and applied **analytical methods**, including repayment cycle analysis and supply chain network positioning, to enhance loan disbursement processes for small businesses, resulting in a more than **30Cr** increase in loan disbursement.

Data Science Intern — United Nations (New Delhi)

Jun 2022 – Aug 2022

- **Data Preparation and Quality Assurance** – Utilized Census-2011 and **World Bank data**, performing data cleaning and testing for multicollinearity, heterogeneity, and omitted variable bias to ensure a high-quality dataset for training.
- **Model Implementation and Optimization** – Developed **linear and ridge regression** models to predict the impact of STEM education on youth productivity. Enhanced model performance by 20% using **Gradient Boosting and XGBoost**.
- **Insight Delivery and Collaboration** – Delivered concise, data-driven insights and findings while coordinating with the World Bank for additional analyses and documentation, ensuring alignment with project goals and requirements.

Education

MSc. Quantitative Economics— Indian Statistical Institute, New Delhi

June 2021 - July 2023

Coursework: Statistics, Econometrics, Probability Theory, Machine Learning, Time Series, Game Theory

B.A. Hons (Economics)— University of Delhi, New Delhi

June 2017 - July 2020

Coursework: Microeconomics, Macroeconomics, Financial Markets, Linear Algebra, Real Analysis

Projects

Applied Data Science Lab — Worldquant University

Jan 2025

- **Data Extraction and Analysis** – Utilized **SQLite** to extract and analyze data, performing extensive exploratory data analysis (EDA) involving the treatment of missing values, outliers, **multicollinearity**, and leaky features.
- **Data Pre-processing and Model Optimization** – Applied **resampling techniques** to balance datasets and employed transformers such as **SimpleImputer, OneHotEncoder, and OrdinalEncoder** to enhance the performance of logistic regression and decision tree models.
- **Predictive and Classification Modelling** – Developed and evaluated various predictive models, including linear/ridge regression, decision trees, and classification models (Random Forest, Gradient Boosting) for performance assessment.
- **Hyperparameter Tuning and Model Accuracy Improvement** – Engaged in **hyperparameter tuning**, adjusting max-depth and number of estimators in decision tree and **random forest** models using validation sets and **K-fold cross-validation** to enhance model accuracy.
- **Unsupervised Learning and Data Visualization** – Standardized variables using **StandardScaler** and built **unsupervised** K-means clustering models on U.S. consumer datasets, employing **Principal Component Analysis (PCA)** for enhanced data visualization.

Additional

- **Technical Skills** – Python, R, SQL, STATA, LATEX, MS Office, Numpy, Pandas, Matplotlib, Seaborn, SKlearn.
- **Scholastic Achievements** – Secured **AIR 16** in ISI, **AIR 25** DSE and **AIR 11** IN IIT-Delhi Masters Entrance Exam.
- **Extracurriculars** – Team captain for **M.P.I.C cricket team**, won **3** trophies and **4** inter-school tournament in the district.