
Harsh Thakur

Data Analyst

harshthakurofficial28@gmail.com | +91 80978 73792 | [LinkedIn](#) | [GitHub](#)

Professional Summary

Data Analyst with expertise in **Machine Learning, Advanced Analytics, and Data Engineering**, delivering production-grade analytics solutions across **banking, FMCG, and manufacturing** domains. Experienced in building **multi-stage ML systems, Spark-based data pipelines, automated ETL workflows, campaign attribution frameworks, and executive dashboards**, with exposure to Deep Learning through research work. Adept at translating complex, multi-source data into actionable business insights while ensuring model interpretability and operational reliability.

Professional Experience

CRG Solutions
Data Analyst

Apr 2024 – Present

- Delivered end-to-end analytics and machine learning solutions across **Banking, FMCG, Supply Chain, and Manufacturing** clients as a consulting data analyst.
- Designed and deployed **ML-driven decision systems**, automated pipelines, and BI dashboards used by business and operations teams.

Client: Diageo (via CRG Solutions) — Machine Learning for Beverage Production

- Designed a **multi-stage machine learning system** to predict malt discovery time, a critical manufacturing KPI impacting product quality and production timelines.
- Integrated data from **four interdependent production stages**, ensuring accurate propagation of upstream outputs into downstream models.
- Engineered advanced features including **interaction terms, lag variables, rolling stability metrics, and derived chemical indicators**.
- Built stage-wise predictive models using **linear regression, decision trees, RFE, and hyperparameter tuning**, prioritizing interpretability so engineers could act directly on feature importance.
- Automated the end-to-end modeling workflow using **Alteryx**, enabling low-maintenance, production-ready deployment.

Client: IDFC First Bank (via CRG Solutions) — Advanced Analytics, Automation & BI

- Built a **fully automated analytics ecosystem** for digital loan acquisition, integrating marketing platforms with internal banking systems.
- Developed **Spark-based ETL pipelines** processing large-scale acquisition data on **AWS S3 and Athena**, orchestrated via **Apache Airflow**.
- Designed **Tableau dashboards** tracking end-to-end funnels from impression to loan disbursal, and implemented **customer journey analytics** to identify bottlenecks and drop-offs.
- Built a **loan bucketing and segmentation framework** for targeted remarketing and prioritization of high-intent customers.
- Developed **anomaly detection logic** spanning business KPIs and ETL pipeline health; reduced data processing time by **70%** and eliminated manual intervention.
- Built **DRIVE**, a multi-touch attribution framework using touch-level journey data (**CleverTap**) to measure incremental campaign performance beyond last-click attribution – decomposing channel impact into **Contribution, Assistance, and Closure Share** to inform budget reallocation.
- Built **LAL Playbook**, a reusable customer cohort framework (behavior, value, and product-affinity based) that replaced manual seed-base creation with on-demand audience pulls for Look-Alike campaigns, reducing campaign launch turnaround time.

Technical Skills

- **Programming:** Python, SQL, R

-
- **Machine Learning:** Predictive Modeling, Feature Engineering, Anomaly Detection, Time Series Forecasting, Neural Networks, Deep Learning
 - **Data Engineering:** Apache Spark, Airflow, ETL Pipelines, Alteryx, PySpark, Tableau Prep
 - **Cloud & Analytics Stack:** AWS S3, Athena, EMR, SageMaker
 - **BI & Visualization:** Tableau, Power BI, TabPy, CleverTap
 - **Libraries & Frameworks:** Scikit-learn, Pandas, NumPy, TensorFlow, Keras, PyTorch

Education

M.Sc. Data Science – Symbiosis International University

GPA: 8.44/10

B.Sc. Statistics – Mumbai University

GPA: 7.69/10

Research & Publications

Analysis and Forecasting of Industrial Production of Non-Durable Food Items

Published in IEEE

ieeexplore.ieee.org/document/10430253

- Applied statistical methods and **deep learning-based time series models** to forecast industrial production trends.