SOURAV DHAR

Senior Analyst
dharsourav03@gmail.com | 8787562928
GitHub | Linkedin | HackerRank

Education

National Institute of Technology — Agartala
Mechanical Engineering BTech — 2009 - 2013
CGPA: 7.7

Great Learnings — University of Austin
Artificial Intelligence & Machine Learning Post Graduate Program — 2022 - 2023

Experience

India Post Payment Bank | Senior Data analyst — July 2021 - Present

- **Credit Risk Scorecard Development**: Led the development of a new scorecard using XGBoost as part of the lead generation for lending partners of IPPB, targeting customers for top-up and renewal of loans, which significantly enhanced model stability and AUC performance.
- **Feature Selection & Data Processing**: Employed **CSI, IV and Correlation techniques** for critical feature identification.
- **Model Monitoring & Integration**: Established monitoring pipelines to track performance metrics like **PSI, AUC, KS, Gini, and bad rates**, ensuring robust model assessment.
- **Cross-Functional Collaboration**: Worked closely with various teams to facilitate the smooth integration of the models into the operational framework, aligning technical solutions with business objectives.

India Post | Business Analyst — 2015 - Present

- **Role**: As a pivotal member of the **Business Development and Analytics team**, I expertly utilized **SQL, PowerBI, Python, and Excel** to optimize lead generation, refine customer targeting strategies, and boost conversion rates.
- Pioneered various **data-driven customer segmentation** strategies to **categorize** customers based on their likelihood to enroll in various financial products of **India Post**.
- Took complete ownership and worked closely with business teams to understand the requirements & developed **PowerBI dashboards** to monitor campaign effectiveness, achieving a **25%** increase in policy adoption.
- **Project**: Designed and developed a **Lead Propensity Model** for DOP and IPPB products including SME Loan Uptake. The execution of the model yielded around 90% disbursal by intelligently targeting 50% of leads, enhancing operational efficiency.
- **Model Improvement**: Participated in crafting a logistic regression model, achieving a 70% AUC.
- **Impact**: The model led to a 15% rise in lead conversions and a 20% increase in targeted marketing efficiency.

Skills

Programming Languages:	Python, SQL
Libraries/Frameworks:	Pandas, Matplotlib, Scikit - Learn, NumPy, Streamlit, XGBoost, Flask
Tools / Platforms:	Statistical Modeling, Supervised Machine Learning, Unsupervised Machine Learning, Predictive Modelling, PowerBi, MLflow, DVC, Docker
Databases:	MySQL, MongoDb

Projects / Open-Source

HomeHorizon: Predictive Analytics and Recommendation Suite for Real Estate *Python, SQL, Pandas, Numpy, Streamlit, Scikit-Learn, XGBoost, AWS*

- **GeoSpatial Insight Engine**: Detailed analysis of spatial data, pricing, and area; highlights key property features.
- **Market Value Navigator**: Utilizes regression for precise market value assessment.
- **Advanced Geo-Synced Property Recommender**: Delivers geo-targeted, user-centric property options.

- **Real Estate Pulse Tracker**: Analyzes market trends and customer preferences succinctly.
- **Deployment**: Provides a streamlined, intuitive interface for real-time analytics.
- **Business Impact**: Streamlines decision-making with precise market analysis and price predictions; augments customer engagement and business efficiency through location-specific recommendations and advanced analytics integration.

SegmentPro: Strategic Customer Navigator *Python, Pandas, Scikit-learn, Matplotlib, Streamlit, MLflow, DVC.*

- **Role**: Key contributor in partnership with Study Table, an analytics startup, focusing on advanced customer segmentation.
- **Methodology**: Executed data preprocessing, **RFM analysis**, and applied **KMeans** and **Gaussian-Mixture Models** for segmentation.Devised targeted marketing strategies based on segment insights.
- **Execution**: Implemented the project's full cycle, from initial **data ingestion** through to **model deployment**, employing a structured, **modular coding** approach for seamless integration.
- **Impact**: Enhanced marketing strategies, leading to ~20% increase in customer retention and improved business growth.

Predictive Analysis of Loan Defaults for NBFC *Python, XGBoost, NumPy, Pandas, MLflow, Streamlit,DVC*

- **Project Focus**: Developed a high-accuracy predictive model for loan default risk, tailored to the Non-Banking Financial Company (NBFC) sector.
- **Scenario Inspiration**: Based on a hypothetical scenario of DHDL Ltd., a fictional NBFC in India, reflecting real-world challenges.
- **Data Analysis Scale**: Analyzed a comprehensive dataset of over 90,000 clients, including detailed loan and default information.
- **Model Accuracy**: Successfully achieved an 88% accuracy rate in predicting loan defaults, demonstrating robust model performance.
- **Strategic Relevance**: Model designed to aid in risk assessment and decision-making, enhancing loan approval processes in the NBFC sector.

Exhibition Art Shipment Cost Predictor *Python, Pandas, Scikit-learn, Matplotlib, Flask, MLflow, DVC.*

- **Project Focus**: Developed a **Cost Predictor application** focusing on rigorous data analysis to identify key cost-driving features for art logistics.
- **Project walkthrough**: Extensive Python-based analysis with **Pandas** was conducted to extract insights from features like the artist's reputation, sculpture dimensions, and shipping details using data visualization libraries like **Seaborn** and **Matplotlib**.
- **UI framework**: Utilized **Flask** for dynamic demonstration of these findings, enhancing stakeholder understanding.
- **Monitoring**: The project also integrated **MLflow** and **DVC** for meticulous tracking and performance monitoring of the machine learning model, significantly improving logistics planning.

Certifications

- SQL (Intermediate) - **HackerRank**

Honors & Awards

Secured a place among the top 10 finishers in the prestigious NBFC Loan Default Hackathon, hosted by Great Learning.