Arnav Jain

Data Scientist

✉ hdan.arnav@gmail.com 📞 9013232094 📍 Delhi- NCR 📅 29 Mar 2000

🎓 EDUCATION

Msc. Statistics & Data Science, *Narsee Monjee Institute of Management Studies*
Grade : 3.90/4

2021 – 2023
Mumbai

Bsc.(H) Statistics, *University of Delhi*
Grade : 9.13/10

2018 – 2021
New Delhi

HSC - CBSE, *DPS Vasundhara*
Grade : 92.8%

2018
Ghaziabad

💼 PROFESSIONAL EXPERIENCE

DATA SCIENTIST - HDFC BANK

Jul 2023 – present
Mumbai

- Money Mule - Transaction monitoring engine:
  - Established a customer-level one-view interface, leveraging daily transactional data to provide detailed insights into inflows, outflows, transaction types, and channels.
  - Identification of suspicious accounts through meticulous monitoring of transactions and historical data. Created 10+ rules with a precision of 80%.
  - Led the migration of the bank's money-mule identification engine from Python to the cloud (Azure Databricks), leveraging Pyspark to improve operational efficiency. Orchestrated and integrated the entire engine on entire bank base flagging around 15000 accounts monthly.
- Scorecard Building:
  - Used Machine Learning algorithms like XGBoost to create cross-selling scorecards.
  - Achieved an 85% capture rate in the top 2 deciles, effectively predicting standalone customers across the bank likely to open an account.
- Embedded Analytics and Automation :
  - Utilized Flask to design applications for streamlining audit and fraud-related processes, drastically reducing daily processing time from 20 hours to a mere 10 minutes.

DATA SCIENCE INTERN- HDFC BANK

Jan 2023 – Jun 2023
Mumbai

- Credit Classification Engine:
  - Developed an end-to-end transaction classification engine that classifies incoming credits on a daily basis into appropriate tiers-sub-tiers achieving a precision rate exceeding 90%.
  - Leveraging advance NLP techniques like Jaccard similarity, Levenshtein distance and word matching to classify transactions as related party transfers.
- Campaign Analytics:
  - Development of micro-segments for marketing campaigns. Using Hypothesis Testing for validation.
  - Performing strategic analysis and developing visualization to communicate business insights (story telling) to make an actionable recommendations.

DATA SCIENCE INTERN- KOTAK MAHINDRA BANK

May 2022 – Jul 2022
Mumbai

- Onboarding Profile Analysis : Led NLP-driven project within HR Analytics ,BIU. Leveraged NLP and cosine similarity for precise Resume Extraction and Ranking, culminating in a comprehensive database of extracted parameters.

RESEARCH AND ANAYSIS INTERN - NATIONAL MISSION ON HIMALAYAN STUDIES

Oct 2019 – Mar 2020
Delhi

- MOEF&CC, GOI funded project entitled "Enhancement of Livelihood Opportunities through Sustainable Development using Multifaceted Support Plans for Village Cluster around Patiya, Almora District, Uttarakhand".
- Role: Pre-processing of socio – economic data based on villages & Statistical Analysis using SPSS

🧠 SKILLS

- Python, Pyspark, Azure (Databricks)
- R , SQL, SPSS , SAS

🪢 CERTIFICATES

- Descriptive Statistics with R, IIT Kanpur

🧩 EXTRA CURRICULAR ACTIVITIES

- Student Placement Coordinator - Placement Cell, NMIMS
- Enactus - Executive Member, RLA - DU
- NSS- Executive Member, RLA- DU