

Yashvi Thakkar

yaashviiphthakkarr@gmail.com | [GitHub](#) | [LinkedIn](#) | [Website](#) | 412-618-9420

EDUCATION

CARNEGIE MELLON UNIVERSITY, Pittsburgh, PA

Jan'21-Dec'22

Master of Information Systems Management, Business Intelligence & Data Analytics

GPA: 3.88

Coursework: Machine Learning, Unstructured Data Analytics, NLP, Statistics, A/B Testing, Marketing Analytics, Database Mgt, Python

Leadership: Heinz College Ambassador, Graduate Student Assembly - Social Committee, Girls Who Code | **Teaching Asst.:** Python, Big Data

VELLORE INSTITUTE OF TECHNOLOGY, India

Jul'16-May'20

Bachelor of Technology – Computer Science and Engineering

GPA 3.97

Leadership: Marketing head for two clubs, Primary Education Program Coordinator for an NGO

WORK EXPERIENCE

HERSHEY

Pittsburgh, PA

Student Data Science Consultant, Hershey Enterprise Data Science Team

Aug'22-Dec'22

- **Shelf Space Optimization:** Created a strategic tool for Hershey to assess quality of retail execution by leveraging POS, market, shelf-space, and SKU data to optimize product spacing by shelf, store, and occasion using **ML models** (OLS, DT, RF etc.) and carried **linear optimization**.
- Designed **A/B** testing experiment to identify average treatment effect of recommended strategy, resulting in 10% lift in revenues.

FEDERATED HERMES

Pittsburgh, PA

Data Science Intern, Advanced Analytics and Business Intelligence

May'22-Aug'22

- Analyzed the behavior of bonds from 2014 to 2022, based on Morning Star categories and forecasted future values using **Multivariate ML Models** (Vector Auto Regression, etc.) to identify high-performing ETFs for clients. Built interactive dashboards using PowerBI.
- Increased Sales Team's competitiveness by constructing a **propensity model** using Logistic Regression to identify financial advisors & banks highly likely to recommend Federated ETFs to investors, resulting in a **10%** quarter-over-quarter lift in **customer acquisition**.
- Won the "**Hackathon MVP Intern**" award among 150 interns for outstanding contribution to Client Engagement Centre by developing **ML-driven apps** to automate performance management for sales reps.

LINDE ENGINEERING INDIA

Gujarat, India

Junior Data Scientist

Nov'20-Apr'21

- Devised a **real-time analytical dashboard** from ground up with ETL pipeline ingesting data from S3 using Python, SQL and Tableau drawing insights for client server load management and eliminated need for searching and merging the data to fetch current license users.
- Enhanced employee productivity by **10+ hours** per day, saving \$1000 per month.

BOTREE TECHNOLOGIES

Gujarat, India

Data Analyst

Jun'20-Oct'20

- Constructed 10+ time-series forecasting models including Exponential Smoothing, ARIMA and Prophet to predict windmill parts **demand** for **inventory optimization**, enhancing status-quo forecasting accuracy by 12%.
- Uncovered key drivers of demand fluctuation using **explainable AI** methods like LIME, generating **\$100k** in annual savings for SCM teams.

SALESFORCE.COM

Gurgaon, India

Solution Engineer Intern

Dec'19-Apr'20

- Developed an XGBoost classification model to identify high-risk consumer based on transactional, demographic & financial data, thereby ascertaining consumers' repayment abilities & **reduced quarterly spending by ~\$300k** for Consumer Risk Management team.
- Resampled the data using SMOTE and built an XGBoost model to achieve best F1 score of 0.71

ACADEMIC PROJECTS

Ford Ka Clustering Analysis [python | exploratory analysis | segmentation | marketing strategy]

Spring'22

- Implemented a **segmentation** scheme using **k-means clustering** on psychographic and demographic data to develop a small-car marketing strategy and advertising message according to the segmented target audience.

Location Intelligence for Music Concerts [python | clustering | data visualization]

Spring'21

- Created a **Python application** to assist music artists with campaign launch strategy by **recommending** most profitable concert locations based on regional audiences' genre preferences, population density, purchasing power etc.
- **Web scraped** Ticketmaster (concert details), Numbeo (US states), and Spotify (music data) and generated **dynamic visualizations** using seaborn, matplotlib and geo-mapped all 50 US states according to filters.

SKILLS

Languages: Python(numpy, pandas, nltk, spacy, scikit-learn, pytorch, tensorflow, scipy), R, Java, C/C++, SQL, HTML, PHP, CSS, MATLAB

Data Analysis Tools: Databricks, Spark, Hadoop, RStudio, Stata, MS SQL Server, Power BI, Tableau, Advanced MS Excel, Azure, GCP

Functional: Predictive Modelling, Regression, Classification & Clustering, A/B Testing, Statistical Modelling, NLP, Data Visualization