VINAYAK KHARE

+1-878-834-9240 | Email | Portfolio | LinkedIn | GitHub | Pittsburgh, PA

EDUCATION

Carnegie Mellon University (CMU), Heinz College **August 2025**
Master of Information Systems Management.
- **Coursework:** Introduction to Deep Learning, Advanced Natural Language Processing, Machine Learning in Production, Data Science for Product Managers, Statistics for IT Manager, Database Management, Agile Methods

Manipal Institute of Technology, Manipal (MIT), India **Jul 2016 - Jun 2020**
Bachelor of Technology, Electrical & Electronics Engineering

WORK EXPERIENCE

PricewaterhouseCoopers (PwC) AC, India, *Senior Data Analyst* **Jul 2021 – Apr 2024**

- **Led the development** of scalable data pipelines in **Python, PySpark, and SQL** to ingest and transform large datasets from **APIs, Azure SQL, and Blob Storage**, reducing manual reporting by **500+ hours annually** and enabling near-real-time **Power BI** dashboards used by **10+ stakeholders.**
- **Collaborated with cross-functional teams** to design **Power BI performance dashboards** tracking key operational metrics (e.g., resolution times, workload distribution, SLA compliance), leveraging **SQL-based data models** to improve resource planning efficiency by **25%** and support data-driven prioritization decisions.
- **Developed CI/CD workflows** in **Azure DevOps** to automate deployment and testing of **SQL and PySpark** data pipelines, reducing pipeline failures by **50%** and ensuring **99% on-time delivery** of critical **Power BI** analytics reports.
- **Integrated diverse data sources** with **SQL transformations** to create unified datasets powering **KPI reporting** and **cohort analyses** in **Power BI**, eliminating **18 hours of manual processing weekly** and accelerating insight generation for leadership reviews.
- **Implemented a Medallion architecture** with **Delta Live Tables and SQL-based aggregation layers** to improve data processing speeds by **3×** and enable self-service **Power BI analytics**, resulting in a **20% increase in dashboard adoption** and improved visibility into performance trends.

Accenture, India, *Data Analyst* **Jan 2021 - Jun 2021**

- **Automated data workflows and reporting processes** using **Python (Pandas, requests)** and **SQL**, consolidating data from **10+ sources**, including APIs and relational databases. Reduced manual preparation time by **600+ hours annually** and improved data reliability for recurring analytics deliverables.
- **Developed Python scripts** to perform **exploratory data analysis** and generate trend visualizations in **Matplotlib**, delivering actionable insights that supported business performance tracking and enhanced reporting efficiency for client stakeholders.

ACADEMIC PROJECTS

Movie Recommendation System (Python, Surprise, Flask, MLflow, Docker, Kubernetes) [Github] **Jan 2025 – Mar 2025**
- **Built a collaborative filtering recommendation engine** using SVD and GridSearchCV, achieving **0.68 precision** and **0.63 recall**, improving suggestion accuracy by **30%** in simulated user scenarios.
- **Deployed a scalable Flask API with MLflow model tracking** and containerized the system via **Docker and Kubernetes**, enabling real-time delivery of recommendations for **A/B testing** and **experimentation workflows.**

Telecom Customer Churn Analysis (Python, Scikit-learn, SHAP) [Github] **Feb 2025 – Mar 2025**
- **Performed EDA and clustering analysis (K-means)** on ~8,000 customer records, identifying a **50% churn risk cohort** characterized by low tenure and minimal premium service adoption.
- **Built and evaluated classification models (Logistic Regression, Random Forest, Gradient Boosting)** with **~88% accuracy**, and generated actionable retention recommendations using **SHAP feature importance.**

Customer Lifetime Value Modeling (Python, Pandas, Matplotlib) [Github] **Dec 2024 – Jan 2025**
- **Developed a cohort-based CLV model** analyzing **8,000+ transactions**, segmenting customers by acquisition year and transaction age to estimate lifetime value and guide acquisition and retention investments.
- **Created visualizations of cumulative spend trends** and performed volume-weighted CLV analysis, enabling prioritization of high-LTV customer segments for targeted marketing.

A/B Testing Simulation on E-commerce Conversion (Python, Pandas, SciPy) **April 2025 – May 2025**
- **Designed and executed a randomized controlled experiment** simulating the impact of a **product recommendation widget** on purchase conversion rates, analyzing **50,000+ user sessions** and calculating uplift using **confidence intervals.**
- **Measured a statistically significant +4.2% uplift ($p < 0.05$)** in conversion, visualized results with **bar charts and confidence bands**, and developed recommendations to inform **product rollout** and future experimentation strategy.

SKILLS

- **Machine Learning & Analytics:** Regression, Gradient Boosting, Random Forest, SVM, Decision Trees, K-Means Clustering, PCA, Hypothesis Testing, A/B Testing, Customer Segmentation
- **NLP**: Text Processing, LSTM, Transformers, Word Embeddings (Word2Vec, BERT), Sentiment Analysis, Topic Modeling (LDA)
- **Data Engineering & Deployment:** PySpark, Databricks, Azure (SQL Database, DevOps), Docker, Jenkins (CI/CD), MLflow, Power BI
- **Libraries & Frameworks:** NumPy, Pandas, PyTorch, Hugging Face, Scikit-learn, LangChain
- **Programming Languages**: Python, SQL