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, ***Data Scientist*** **Jul 2021 – Apr 2024**

- Led development of customer-centric analytics platforms for **10+ Fortune 500 clients**, integrating behavioral, transactional, and system-level data **(~10M+ records)** across **SOC, SIEM, EDR, and threat intelligence sources** using **Python, SQL, and PySpark**.
- **Designed and implemented KPIs** to track **user behavior and operational efficiency** (e.g., **false positive rate, alert resolution time),** enabling consistent insight delivery across business units and increasing stakeholder trust in analytics outputs.
- Developed and maintained **role-specific Power BI dashboards** for different **customer segments (e.g., SOC managers, analysts), driving a 25% improvement in operational efficiency** through better shift planning and workload balancing.
- **Built scalable pipelines (Azure SQL, Blob, APIs) for end-to-end data ingestion**, transforming disparate source systems into a unified analytics layer. Integrated CI/CD using Azure DevOps to maintain **>99%** SLA adherence for reporting cadences.
- Measured adoption and iterated on product experience: Analyzed Power BI usage logs, **applied quasi-experimental methods (pre-post analysis, matched cohort analysis)** to test UI/UX changes across clients. Insights led to design improvements and **a 20% increase in weekly active users.**

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

- Automated recurring sales and performance analytics using Python and SQL, **reducing manual effort by 600+ hours annually** and improving reporting consistency across executive dashboards.
- Conducted customer journey and behavioral trend analysis using Matplotlib and Seaborn, surfacing insights that optimized cross-functional reporting and improved data-driven decisions in client-facing teams.

ACADEMIC PROJECTS

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.

NLP Analysis of Diabetes CGM Consumer Feedback (RoBERTa, Topic Modeling, K-means) [Github]
- Analyzed 37K+ user posts on CGM devices using **RoBERTa sentiment modeling, topic extraction, and K-means clustering** to uncover brand-level concerns (e.g., **accuracy, cost**).
- Segmented users by condition and device preference, delivering **30% more targeted insights** to inform **product messaging and prioritization.**

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] **Oct 2024 – Dec 2024**
- **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.**

SKILLS

- **Machine Learning & Analytics:** Regression, Gradient Boosting, Random Forest, SVM, Decision Trees, K-Means Clustering, PCA, Hypothesis Testing, A/B Testing, Customer Segmentation**,** CLV Modeling, Retention
- **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