# ROHIT MENON

rvmenon@berkeley.edu | www.linkedin.com/in/rohit-menon-a88b65165 | 408-832-3233

## EDUCATION

**University of California-Berkeley** — **Berkeley, California**
*MS in Information and Data Science* — *December 2024*

**University of Wisconsin-Madison** — **Madison, Wisconsin**
*BS in Data Science, Minor in Computer Science* — *May 2023*

### SKILLS

- Statistical Analysis
- Data Analysis
- Technical Analysis
- Python, R, SQL, Java. Git
- Tableau, SPSS
- Pandas, NumPy, Sklearn, TensorFlow, PyTorch, Pyspark, Matplotlib, Seaborn, Pygame
- Strong Presentation Skills
- Analytical Writing

## EXPERIENCE

**VMware** — **Palo Alto, California**
*Data Scientist Intern* — *May 2022 - August 2022*

Developed and deployed a multi-cloud chargeback model (Azure, AWS), processing 5M+ lines of usage and billing data, achieving 98% accuracy in discount allocation across business units.

- Enhanced business decision-making by reducing manual data processing time by 30% through Pandas and NumPy solutions in Jupyter notebooks.
- Improved data integrity by identifying and removing outliers, leading to a 10% increase in analysis accuracy.

*Data Engineer Intern* — *May 2021 - August 2021*

Conducted a critical analysis of agile sprint planning using real-time JIRA data, resulting in a 15% improvement in sprint planning accuracy.

- Provided project leaders with actionable insights by processing structured and unstructured data in Oracle Data Warehouse and generating time-series analyses using ETL techniques.
- Cleaned and standardized datasets, improving data quality and analysis accuracy by 10% through SQL, Python, and Jupyter notebooks, and identifying key trends via EDA graphs.
- Generated a Tableau Dashboard that displayed a detailed breakdown of historic planning of JIRA issues into 6 actionable categories for any project within any team over a variety of dimensions.

## COMPLETED PROJECTS

**University of California-Berkeley** — **Berkeley, California**
*Student* — *August 2023 - Present*

- Architected, implemented and deployed in AWS a 4-tier internet-facing vehicle appraisal system which generates automated car valuation and LLM-powered improvement recommendations.
  - The fronted is implemented using React running on AWS Amplify infrastructure which connects with a backend API-Gateway based service. API-Gateway integrates with two Lambda functions which perform VIN detection and vehicle damage detection. This system uses 2 models - AWS Textract and YOLOv8 along with MarketValue API.
- Designed and implemented an advanced NLP framework using BERT and RoBERTa models, achieving 97% accuracy in writing proficiency classification and integrated explainability methods such as LIME and large language models increasing system transparency.
- Developed a comprehensive PySpark pipeline for predicting flight delays using advanced classification models that handled class imbalance, time-series cross-validation and hyper-parameter tuning achieving an F1 score of 0.35+ on large-scale aviation data.
- Collaborated on a deep learning project using a custom CNN and ResNet-based models, achieving 40% accuracy in predicting major city locations from photos and created a Gradio interface for real-time image predictions.
- Conducted A/B testing to assess the impact of roadside reflectors on nighttime driving speeds, achieving statistical significance and demonstrated a 7% reduction in average driving speeds, providing actionable insights for traffic safety improvements.
- Crafted an interactive Tableau dashboard analyzing performance metrics for the top 200 PGA Tour players and course statistics, offering insights across 4 key metrics (driving, approach, recovery, putting). It was used by 80+ users for testing as well as informing on player and course patterns.

**University of Wisconsin** — **Madison, Wisconsin**
*Student* — *Oct 2019 - May 2023*

- Designed and optimized machine learning models, including a PyTorch MLP for dog breed classification (65% accuracy), football play prediction using various ML classifiers (73% accuracy), and ensemble methods that improved performance by 12%, while streamlining the processing of 50K rows by 20%.