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
- Tableau, SPSS
- Pandas, NumPy, Sklearn, TensorFlow, PyTorch, PyGame, Matplotlib, Seaborn, Pyspark
- Strong Presentation Skills
- Analytical Writing

EXPERIENCE

VMware — **Palo Alto, California**
Data Scientist Intern — *May 2022 - August 2022*
Developed a multi-cloud equitable chargeback model (Azure, AWS) and end to end pipeline that processes several millions of lines of usage and billing data accounting for near perfect discount allocation across business units.

- Worked with structured near real-time cloud data in Jupyter and performed data manipulation and math functions using Pandas and NumPy.
- Removed outliers for an accurate analysis.
- Tested data to ensure data integrity and accuracy

Data Engineer Intern — *May 2021 - August 2021*
Developed critical analysis on the effectiveness of agile sprint planning using real-time JIRA data.

- Used data engineering practices to provide project leaders insights into their agile development use cases; worked with near real-time structured and unstructured data in Oracle Data Warehouse; utilized ETL techniques to generate time series analysis.
- Cleaned the dataset to be consistent and complete for an accurate analysis and created EDA graphs using a combination of SQL, Python and Jupyter notebooks, to find patterns or trends within the data.
- Generated Tableau Dashboards that display the breakdown of historic planning of JIRA issues for any project within any team over a variety of dimensions.

COMPLETED PROJECTS

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

- Designed and executed an advanced NLP framework utilizing BERT and RoBERTa models for classifying writing proficiency, achieving up to 97% accuracy, and integrating explainability methods such as LIME and large language models to create a more accurate and transparent system for assessing writing skill levels.
- Developed a comprehensive pipeline utilizing PySpark for predicting flight delays using advanced classification models along with hyper-parameter tuning and time-series cross-validation to enhance model accuracy and efficiency in processing large-scale aviation data.
- Collaborated on a deep learning project focused on predicting location based on major city photos.
  - Developed a CNN architecture, including a custom model and ResNet-based variants finally creating a Gradio interface for seamless interaction with the model, enabling a real-time prediction of city locations from images with a 40% accuracy.
- Utilized A/B testing methodology to investigate the impact of roadside reflectors on nighttime driving speeds achieving statistical significance and providing robust evidence on whether roadside reflectors induce significant changes in nighttime driving speeds.
- Crafted an interactive Tableau dashboard investigating performance metrics of the top 200 players on the PGA Tour, offering insights into driving, approach, recovery, and putting

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

- Utilized Stanford University's curated Dog Image Data, comprising 120 common dog breeds, to construct a PyTorch-based MultiLayer Perceptron Model for accurate breed classification.
  - Engineered a robust classification framework, including feature transformation, image augmentation, and parameter optimization to enhance consistency and accuracy.
- Predicted football play types during drives based on game features, utilizing ML classifier models on historical play-by-play football data.
  - Streamlined training and testing processes by filtering, cleaning, and transforming necessary columns using Pandas.
  - Implemented a baseline decision tree (based on Down and Yards to Go) to validate play prediction intuition.
  - Expanded model accuracy using sklearn library with k-nearest-neighbors, RandomForest, Bagging, Adaptive Boosting, and a Stacking classifier.