# Rithvik Vukka

# **Machine Learning Engineer**

**(**951)-425-3784

rithvik.v@jobtechsmail.com

in LinkedIn

( ) GitHub

#### **SUMMARY**

Data Scientist with 3 years of experience specializing in advanced machine learning techniques. Proficient in developing predictive models, and driving strategic insights. Expert in Python-based data science tools, and end-to-end model development from preprocessing to deployment. Consistently delivers measurable improvements in operational efficiency and decision-making processes.

## **EXPERINCE**

#### Data Scientist (AI/ML) | KT2i, California

February 2025 - Current

- Developed enterprise Al assistant with hybrid RAG architecture combining vector search and SQL agents, enabling natural language queries across
   HR, finance, and administrative data sources with intelligent routing and automated fallback mechanisms for operational decision-making.
- Architected scalable RAG system integrating Pinecone vector database with Airtable/SQL data sources, delivering real-time analytics through
  conversational AI interfaces and achieving faster query response times while processing 1000+ documents across multiple enterprise domains.
- Built production-ready MLOps pipeline with real-time health monitoring and automated performance validation across document processing, embedding generation, and query routing components, ensuring enterprise security compliance.

#### Data Scientist | BNY Mellon, California

October 2024 - December 2024

- Collaborated with senior data scientists and risk analysts to develop Random Forest and XGBoost models for credit risk assessment, resulting in 15% improved prediction accuracy and 22% reduction in false positives
- Implemented end-to-end MLOps pipelines using MLflow and Evidently AI, incorporating CI/CD, model drift monitoring, and automated performance validation to ensure long-term model stability.
- Engineered an NLP-driven solution using fine-tuned BERT and rule-based algorithms to automate extraction of critical financial risk indicators from unstructured client documents like loan agreements, emails and reduced manual review time by 30% and improving risk flagging accuracy.

#### Data Scientist | AAPI Data, California

October 2023 - September 2024

- Developed and maintained interactive Plotly visualizations comparing 2020 vs 2010 Census data across multiple Asian and Native Hawaiian/Pacific Islander origin communities, with dynamic options for customizable demographic analysis.
- Developed and evaluated advanced name-race classification models by integrating Bayesian Smoothing, Inference Group methodologies, and fine-tuned BERT Transformers to uncover ethnic representation patterns in state legislatures.
- Implemented heuristic bootstrapping techniques to augment limited training datasets, executed robust statistical comparative analyses to identify optimal model performance, and deployed validated models on public FEC datasets.
- Designed and deployed interactive Plotly dashboards for dynamic visualization and comparative analysis of racial and ethnic campaign contribution patterns, highlighting key trends across the 2020 and 2024 election cycles.

#### Data Scientist | Vivma Software Inc, India

January 2022 - June 2023

- Applied large language models (LLMs) like GPT-3.5 and BERT to automate clinical documentation summarization, extracting actionable insights
  from EHRs and reducing manual review time by 35% compared to legacy workflows.
- Collaborated with cross-functional teams in an Agile setting to integrate insights into Power BI dashboards, enhancing clinical decision-making; implemented CI/CD pipelines via GitHub Actions, reducing claim denial rates.
- Executed a patient readmission prediction model using TensorFlow and Neural Networks on AWS SageMaker, enabling proactive interventions and significantly reducing 30-day readmission rates through accurate risk assessment.

### **SKILLS**

Methodologies: Agile, SDLC, Waterfall Programming Languages: Python, R, SQL Databases: MySQL, PostgreSQL, SQL Server

**Big Data Technologies:** Apache Spark, Hadoop **Data Visualization Tools:** Tableau, Power BI, Excel

Machine Learning Algorithms and Library: Pandas, NumPy, SciPy, Scikit-learn, Keras, TensorFlow, PyTorch, XGBoost, Random Forests, Decision

Trees, SVM, Neural Networks, NLP, Large Language Models

Cloud Technologies: AWS (S3, Lambda, SageMaker), Azure (Databricks, Data Factory)

Version Control & Collaboration Tools: Git, GitHub

Soft Skills: Problem-Solving, Critical Thinking, Decision Making, Strong Communication Skills, Teamwork, Attention to Detail

## **EDUCATION**

Master of Science in Computer Science | University of California Riverside, California Bachelor of Technology in Computer Science and Engineering | Amrita School of Computing, India

December 2024

June 2023

#### **PROJECTS**

**EdgeAvtar:** Optimized TTS for real-time speech on Nvidia Jetson Nano, achieving 500-700ms latency. Integrated real-time STT via OpenAI API, improving reliability, reducing cloud dependency, and enhancing TTS efficiency for interactive applications.

Yelp Data Analysis for Strategic Business Insights: Analyzed Yelp datasets using SQL++ on AsterixDB, optimizing spatial queries, data unnesting, and grouping for better performance. Identified customer behavior and geographic trends to support data driven decision-making.

**Crime Data Analysis:** Analyzed Chicago Crime Dataset with Spark SQL and QGIS, integrating geospatial data for crime pattern visualization. Processed 14M+ data points, optimizing analysis time by 30% to aid law enforcement resource allocation.

# **PUBLICATIONS**

Influential Node Identification on a Multilayered Attributed Network  $\square$ 

Applied K-Means and DBSCAN clustering to identify key nodes in a multilayered attributed network using attributes like H-Index, P-Index, Unequal P-Index, and affiliations. Visualized citation networks, enhancing research accessibility and revealing cohesive node groups.