Vijay Verma Rudraraju

Permanent Address:
Chantilly, VA 20151

Email: vijayvr@seas.upenn.edu
Mobile: **(571) 383-1931**
Work Authorization: **U.S. Citizen**

Education

- **University of Pennsylvania - Penn Engineering** 2024 - 2026
  Master of Computer Science and Information Technology
- **University of Michigan, Ann Arbor - Stephen M. Ross School of Business** 2020 - 2024
  Bachelor of Business Administration - Information Technology, Business Analytics, & Finance
  Computer Science Minor (College of Engineering)
  Relevant Coursework: AI and ML in Investment Strategies, Generative AI, Database Management Systems, Data Structures and Algorithms, Discrete Mathematics, Linear Algebra, Multi-variable Calculus, Programming and Intro Data Structures, Business Analytics and Statistics, Business Information Systems

Skills

- **Languages:** Python, C/C++, Java, R, SQL, JavaScript, TypeScript, HTML5, CSS, jQuery, XML, Visual Basic (VBA)
- **Data Science Libraries:** TensorFlow/Keras, PyTorch, NumPy, Pandas, Matplotlib, Scikit-Learn, SciPy, Seaborn, MLflow
- **AI Tools:** LlamaIndex, LangChain, Ollama, Hugging Face, RAG, LoRA, Chroma, FAISS, OpenAI Tools
- **Back-end:** Django, FastAPI, Flask, Spring Boot, Oracle, MongoDB, JDBC API, PostgreSQL, MySQL, MS SQL Server, Redis
- **Developer Tools:** VS Code, Eclipse, Angular, React, Node.js, Next.js, Gradle, Git, GitLab CI/CD, Docker (Desktop/Engine), CUDA toolkit, cuDNN, JIRA, Power BI/Apps, Tableau, Jupyter Notebook, PyCharm, Pytest/Unittest, JUnit, MS Office Suite
- **Cloud:** AWS (EC2, IAM, VPC, S3, EFS, Lake Formation, Lambda, SageMaker, SNS, SQS, RDS, QuickSight, ECS)
- **Hardware/OS:** Nvidia GPUs, Ubuntu, RedHat Linux, Debian, Windows, MacOS
- **Finance Tools:** FactSet, S&P Capital IQ, Bloomberg
- **Leadership:** Alternative Investments Club - VP of Investments, Nexecon Consulting Group - Project Manager, Phi Beta Lambda - Business Fraternity

Certifications

- AWS Certified Cloud Practitioner [Credential]

Experience

- **COLSA Corporation** **Huntsville, AL**
  Data Science Intern - DACS Lab (Data Science Division) *May 2024 - Current*
  - Built ML signal processing system to identify sources, separate, and categorize geostationary satellite signals
    - Performed data pre-processing, visualization, feature extraction, labeling, and time-series analysis on complex petabyte-scale datasets sourced from satellite receivers and generated using Python Math and Scipy libraries
    - Developed and trained ML models for signal source separation, electromagnetic interference (EMI) detection, and modulation type classification using TensorFlow/Keras, Scikit-learn, NumPy, Pandas, and Matplotlib
    - Designed and implemented deep learning architectures with CNN, RNN, LSTM, U-Net, and ResNet frameworks while evaluating and comparing model performance by creating custom accuracy metrics for signal data types
    - Constructed custom loss functions and applied hyperparameter tuning to enhance model performance
    - Generated Docker containers for model training scripts, integrating with MLflow to conduct A/B testing and GitLab CI/CD pipelines for streamlined version control, development, and deployment
    - Utilized MongoDB collections to store signal data, developed web application using TypeScript/React/Next.js, and created RESTful API built through FastAPI framework to deploy interactive user interface
  - Created LLM-based knowledge management system to manage and retrieve information, simplifying user queries
    - Built robust generative AI assistant for internal signal processing project, leveraging Meta Llama 3.1 LLM fine-tuned using LoRA and finalized with full fine-tuning for enhanced performance using Nvidia GPU cluster
    - Integrated Retrieval-Augmented Generation (RAG) using LlamaIndex and Langchain, embedding internal project technical and compliance documentation to improve context-based query response accuracy
    - Implemented document embedding and retrieval pipeline, utilizing Chroma and FAISS to store/index vector embeddings, enabling efficient similarity-based retrieval and 20% improvement in Prod support resolution times
    - Experimented with Agentic AI using LlamaIndex SQL Query Engine/Agent to interact with multiple internal relational databases, leading to deployment of tool for No-Code user queries on processed signal data
  - Created NLP-based navigation platform that translates drone mission plan in text/voice format into API request
    - Constructed custom API framework for object avoidance/detection using LiDAR sensors, Python TensorFlow/Keras ML libraries, fine-tuned YOLOv8 model, and translating reinforcement learning algorithms to code

  - Explored NLP solutions using LoRA to fine-tune Llama 3.1 LLM, employing prompt engineering, adjusting temperature/distribution parameters, and building Langchain agent to output JSON formatted API requests
- Collaborated with data scientists, developers, and devops teams in sprint-based and pair programming workflows
- Wrote 4,500+ lines of peer-reviewed production level code tested using unittest/pytest and integration testing
- Prepared detailed technical documentation, data validation reports, and user guides for ML and generative AI applications while presenting product development briefings to clients, managers, and company executives

- **Solomon Partners** **New York City, NY**
  Software Engineering (Data Science) - Intern *June 2023 - August 2023*
  - Implemented web scraper API to collect market data, leveraging Pandas library for pre-processing and analysis
  - Managed data storage using AWS S3 buckets and executed SQL queries to retrieve scraped data to provide insights
  - Utilized DAX functions to deliver a PowerBI dashboard with detailed charts showcasing key market trends
  - Streamlined workflow by using Python TensorFlow/Keras to auto-populate buyer contact tracker by processing natural language context from email communications, leading to seamless M&A process for sale of company

- **Bloomfield Capital** **Detroit, MI**
  Data Analyst - Intern *May 2022 - August 2022*
  - Analyzed economic growth and correlation to senior living industry by scraping 55,000+ data points and executing Python scripts with Pandas, NumPy, Matplotlib, and Seaborn to predict demand trends via regression analysis
  - Constructed summary of analysis on Tableau dashboard, presenting findings to company executives

- **Blackford Capital** **Grand Rapids, MI**
  Private Equity Junior Analyst - Intern *September 2021 - April 2022*
  - Initiated budget reconciliation process using statistical analysis, Excel pivot tables, and historical company data to identify discrepancies between appropriated budgets and company performance, realizing cost-reduction potential
  - Spearheaded research into e-commerce pool supply business by constructing valuation model with intricate synergy/leverage/returns analysis using FactSet/Capital IQ data and Excel, leading to transaction savings

Academic Projects

Projects Portfolio (Github)

- **University of Michigan - Maize and Blue Endowment Fund** (Quantitative Investment Analyst) *Sept 2022 - April 2024*
  - Directed 6 member pitch team to propose $20k investments in both Joint Chiropractic (JYNT) and Danaos Corporation (DAC), culminating in unanimous votes into endowment portfolio and over 60% returns to date
    - Implemented AI/GPT-based model and Bloomberg/FactSet quant screens to locate market inefficiencies
    - Utilized Python to statistically model Riemann Sum Approximation of normally distributed potential acquisition outcomes for JYNT to measure market sentiment and irrationality towards sale of company
  - Compiled detailed analysis and presentation to recommend investment in in-flight internet company Gogo Inc. (GOGO) for annual stock pitch competition, placing 2nd overall and receiving university acknowledgement
- **Machine Learning for Portfolio Construction** (Python) *January 2024 - April 2024*
  - Collected 18,250+ training data points from notable industry databases and developed ML algorithm by using Pandas, NumPy, and MatPlotLib libraries to perform time-series analysis and predict portfolio returns
  - Adjusted performance by toggling between OLS/LASSO regression, using Decision Tree/Random Forest, exploring Gradient Boosting, and implementing feature scaling/engineering, improving model accuracy by 2.27%
- **Mock Facebook Databases** (SQL/JavaScript) *August 2023 - December 2023*
  - Generated data tables using SQL DDL statements per desired schema requirements including constraints, sequences, and triggers within Oracle DBMS while populating database and creating external views to display loaded data
  - Utilized Java JDBC API to execute SQL queries against created relational database to gather key insights into stored tabular data while exporting information from query results into JSON array format to assess accuracy
  - Stored data within MongoDB collections and ran further required queries within DBMS using JavaScript
- **NLP for Forum Post Classification** (C++) *February 2021 - April 2021*
  - Implemented Natural Language Processing (NLP) to train neural network for categorizing online forum posts
  - Achieved a predictive accuracy of 87.08% on the testing sample by utilizing the Multi-Variate Bernoulli Naive Bayes Classifier and conditional probabilities to understand forum post context and derive appropriate labels
- **University of Michigan - Nexecon Consulting Group** (Project Manager/Business Analyst) *Sept 2020 - April 2024*
  - Led team of 5 analysts to provide recommendations to F500 auto manufacturer detailing expansion strategies
  - Designed statistical model incorporating 14 datasets comprised of socioeconomic metrics and utilized weighting system for model attributes to rank regional demand for company products at zip-code level of granularity
  - Provided client with structured dashboard using Tableau visualization software by creating overlapping heat maps highlighting key under-served geographic markets and employing charting tools to synthesize supporting analyses