ARIVIRKU INIYAN
Data Scientist II

Ph: +16176532287 LinkedIn: *uvira007*
Email: tarivirkuiniyan007@gmail.com GitHub: *uvira007*

Career Summary:

Data driven analytical data scientist with extensive experience in building, training, evaluating and deploying Machine Learning models. Adept in performing statistical inference tests and assessing 'goodness of fit' for models for varying business needs.

Achievements:

- Implemented RAG pipeline with GPT-4o foundational model for condition generation, thus reducing the effort of care providers by over 35%.
- Created fingerprinting model using XGBoost that helped to identify 9% of potential conditions based on comparable dataset.
- Developed an End-to-End ML Pipeline with pre and post processing steps that reduced the manual effort of the data science team by 30%
- Utilized pydantic modelling from langchain to format the LLM response into OMOP data model for information extraction that enabled the medical providers for pre-emptive action.
- Ameliorated the existing prompt by implementing a reflection agent to improve the interpretability of LLM predictions.
- Generated particle archiver that archives blobs older than 5 years and thus helped to reduce the cost by 40% for the search index.

Awards:

- *Feather in My Cap* from multiple clients in Wipro technologies for being the best performer in the team.
- *Thanks a zillion* for going above and beyond to exceed the expectations of the client.
- *VM AllStars Annual Award* for being the best team of the year and making a difference across the organization.
- *VM AllStars Spot Award* for being an encouraging leader and leading the team to success by keeping them motivated.
- *Vibrancy Recognition* for helping create the best customer experience with the V28 conditions project.

Skills meter:

Languages: Python| R| Java| VBScript| Groovy| SQL| VBA| Shell| Bash| Javascript
Tools: Azure Databricks| Vertex AI| BigQuery| PyCharm| Jupyter| R studio| Pega| ReadyAPI| SQL Developer| AWS| Azure| GCP| Jenkins| Postman| Data studio| Visual Studio| IntelliJ| Grafana
Packages/Framework: pandas| numpy| openai| langchain| pydantic| pyspark| openai| keras| scikit-learn| matplotlib| seaborn| rank-bm25| RAG| Plotly| sweetviz| Bancs| Temenos| Finacle| beautifulsoap4| SAFe Agile| REST| SOAP
DataBase/Storage: DB2| SQL| putty| DynamoDB| CouchDB| Microsoft Access| RDS| Aurora| MongoDb| Unity Catalog| BigQuery| S3| Azure blob storage| Google Cloud storage
Domain: Banking – 7 Years| Insurance – 1.4 years| Finance – 3 Years| HeathCare – 1.8 Years

Certifications:

11/2025| AWS Certified Machine Learning specialty | **Amazon**
03/2025| Predictive modelling with BigQuery| **Google**
05/2024| Azure AI fundamentals | **Microsoft**
06/2023| AWS Certified Developer associate | **Amazon**
12/2022| SSM Certified SAFe® 5 Scrum Master| **Scaled Agile.**

05/2022| PCAP: Certified Associate in Python Programming| **openEDG.**

Educational Qualification:

Mastering Deep Neural Networks| GPA-4.0| 2025| Massachusetts Institute of Technology| Massachusetts.
Master's in Data Science| GPA-4.0| 2023| Rochester Institute of Technology| New York.
Bachelor's in Electrical Engineering| GPA-4.0| 2011| Kongu Engineering College| Tamil Nadu.

Relevant Experience:

- **Oak Street Health**
  Data scientist II *Chicago, USA |11/2023 –*
  - Architected a Retrieval Augmented Generation pipeline with OpenAI gpt-4o foundational model to identify potential conditions from a plethora of documents through batch processing.
  - Instrumented a hybrid search using reciprocal rank fusion on the results from bm25 and vector search to retrieve the top N documents for the retrieval pipeline.
  - Iteratively enhanced the prompt from rating-based system to a Boolean based system to improve the model performance.
  - Conducted experiments to include particle medical records to the existing model and validated the improvement in the positivity rate of the model for specific conditions.
  - Developed a serving pipeline to generate the results for the given set of patients stored as .csv in the blob container.
  - Improved the Data Ingestion pipeline by utilizing fitz to skip OCR for pages, if the text density is greater than 80%
  - Designed a find my doc model that uses sentence_transformers to calculate the embeddings in a CUDA compute to identify the documents specific to the conditions provided.
  - Fine-tuned the post processing steps to eliminate the false positives and transform the output to ingest into the Data Warehouse.
  - Implemented SHAP for feature selection and designed a fingerprinting model using XGBoost for predicting potential conditions for the patient.
  - Integrated reflection agent with the RAG pipeline to improve the explainability of the model's response.

Projects:

teamUFO: Executed unsupervised learning (K-means clustering) to identify patterns among the trove of documents released by "FBI: The vault". Embedded a vector space model behind a UI developed in streamlit to retrieve the top N documents based on the query term provided.

DuplicatesExtractor: Replicated the duplicates extractor plugin for IntelliJ that uses a CNN model with approximately 80 metrics as features to identify and provide suggestions for duplicate blocks of code whenever it's injected into the project workspace.

Racing Nightmare: Utilized basic feature selection techniques such as step-wise selection and Lasso regression. Identified potential causes for a horse or horse/jockey combination for their finish position in the race through visualization using plotly. Created an interactive dashboard using tableau that showcases the racing of horses/jockey on different race conditions.

Electoral College Dropout: Implemented basic feature engineering techniques such as one-hot Encoding, discretization and normalization. Applied zeroR and oneR classifiers to create a baseline for performance. Generated basic classification models such as naïve Bayes, Logistic regression and Random Forest to evaluate against the baseline for 'goodness of fit' and predict the voting turn-out based on the survey questions(features).

AI resume filtering: Natural language processing techniques such as lemmatization and stemming were explored. Evaluated performances of basic information retrieval methods such as Boolean Retrieval, Term Frequency, tf-idf and cosine similarity. Created a vector space model with weighted vectorization. Developed a UI to provide the user with the ability to add weights for individual query terms and define the number of profiles to be retrieved.

Previous Experience

- **ValueMomentum**
  Senior Lead Automation Engineer *Salt Lake City, USA |08/2019 –11/2023*
  - Architected an automation framework in R and established CI/CD pipeline using Azure DevOps for one of the leading banks in Utah.
  - Driven the automation of FutureCore project to success by leading a 12-member team covering multiple agile teams.
  - Amplified automation by implementing a unittest Framework for UI based applications using pySelenium.
  - Improved automation process, by orchestrating the frameworks between APIs, GUI and desktop applications.
  - Decreased the manual effort involved in test case design and execution by 70% by developing utility tools using VBMacro and Java.
  - Created a dynamic extent report by using html and CSS bootstrap that is centralized for testing projects across the organization.

- **BlueCross BlueShield of NC**
  Business Solutions Developer *Durham, USA |04/2019 – 08/2019*
  - Improved sanity testing, by integrating bots from PEGA with UFT.
  - Developed API scripts using UFT to reject the incorrectly created service forms.
  - Achieved Continuous testing by integrating the UFT scripts with Jenkins.
  - Integrated PEGA Robotics with UFT to clear claims backlog generated at the end of the day.

- **Wipro Technologies**
  Senior Automation Engineer *Boston, USA |05/2016 – 04/2019*
  - Demonstrated the efficiency and cost-cutting of using Selenium over HP UFT for one of the largest asset management companies by showcasing a Proof of Concept on the Application Under Test.
  - Ameliorated the testing by making API calls from java and thus by-passing the UI. This process proved to significantly reduce the testing time and help performance testing teams.
  - Integrated the in-house built hybrid driven framework with Ranorex to identify the flex objects for the web application developed using Cloud Development Toolkit 3.
  - Implemented Selenium grid with Jenkins to perform cross-browser testing for the application by utilizing the remote machines.

  Automation Engineer *Chennai, India |07/2011 – 05/2016*
  - Worked with the automation team to develop a hybrid driven framework using HP QTP and implemented it for the world's largest issuer of credit cards.
  - Exposed to multiple STLC models and different testing techniques and how to transform client requirements into usable product.
  - Performed load and stress testing using Apache JMeter for both API and web testing to report on the metrics.
  - Implemented Orthogonal Array Testing technique to maximize the coverage by using the minimum number of test cases.