Veena Ramesh

veenaramesh.github.io | github.com/veenaramesh | linkedin.com/in/veenarameshh
+1 (703) 626 3378 | veena.ramesh.1@gmail.com

EDUCATION

University of Virginia **Charlottesville, VA**
Bachelor of Arts in Computer Science, Cognitive Science (Linguistics concentration) May 2021
Coursework: Machine Learning, Artificial Intelligence, Statistical Learning & Graphical Models, Natural Language Processing, Information Retrieval

PROFESSIONAL EXPERIENCE

Lovelytics Data, Inc. **Washington, D.C.**
Senior Machine Learning Engineer April 2022 – Present

- Lovelytics is a data science boutique consulting company that works largely with Databricks, a unifying Data and AI platform.
- Conceptualized LLM process frameworks, guidelines, and LLMOps architecture, and built and presented LLM demos to clients.
- Leading client-facing projects, communicating non-technically, and optimizing results while managing stakeholder expectations.
- Productionized custom machine learning models using PySpark and MLFlow in accordance with the Unity Catalog governance structure in Databricks. Automated testing and deployment of code to different environments through Github Actions.

Itek Infomatic Inc. **Washington, D.C.**
Data Scientist June 2019 – May 2022

- Imputed missing data points using a Random Forest classification model with PCA, achieving a model testing accuracy of 84%.
- Built a SQL pipeline to ingest and standardize feature-rich data by generating unique weight specifications for downstream users.
- Analyzed differences between company and public data. Optimized queries by vectorizing and query tuning SQL statements.

Red Oak Strategic **Washington, D.C.**
Data Science Intern June 2019 – August 2019

- Red Oak is a start-up data science consulting company. Specialized in predictive modeling, business analytics, and AWS services.
- Engineered a time-series forecasting model (ARIMA) using Sagemaker to help predict the likelihood of wildfires in Southern CA.

Chimera Securities **Arlington, VA**
Data Science Intern June 2018 – Feb 2019

- Developed a streaming web app in Python Dash to visualize live SPDR S&P 500 ETFs (SPY) data for investor decision-making.
- Forecasted aggregated stock price data to estimate future S&P performance using a predictive time-series model (ARIMA).

Columbia University Center of Justice **New York City, NY**
Justice Through Code Mentor. Feb 2023 – June 2023

- Selected as a volunteer for JTC, which provided programming education and career advice to those who have conviction histories.

RELEVANT OPEN SOURCE PROJECTS

Databricks Migration Tool (github.com/databrickslabs/migrate) May 2022 - Present

- Key contributor to the repository, building and optimizing tools to assist customer migrations between Databricks workspaces.
- Led the development of a custom library forked from the repository (lovelytics/migrate) to move from one to many workspaces.
- Worked directly with customers using the repo to provide advisory support and troubleshoot issues posted on Github.

arXiv Search Engine May 2021

- Built an end-to-end arXiv search engine by scraping the arXiv site, embedding the data using spaCy's built-in embedding model, chunking and vectorizing documents using the Gensim FastText model, and indexing data using the BM25 algorithm. The heaviest model weights were abstracts and time. The front end application was built using Python Dash and deployed on Flask.
- Across approximately 60 surveyed students, the search engine performed qualitatively better than the existing arXiv search engine; consistent feedback showed that the engine retrieved more 'recent and relevant' research.

Analyzing Celebrity Language on Twitter May 2021

- Analyzed behavioral differences in political language between celebrities (defined as verified users on Twitter) and non-celebrities (identified as unverified users on Twitter) using transformer models and rule-based sentiment models (like VADER) across a variety of topics, including COVID-19 and the ongoing primary presidential campaign during the 2020 election year.
- Demonstrated a statistically significant difference between the two populations through the Welch's unequal variances t-test. Verified users use more neutral sentiment than unverified users, who use more quantitatively extreme language.

LANGUAGES, SKILLS, AND INTERESTS

- **Languages**: Python (advanced; functional and object-oriented frameworks), SQL (mySQL, DBSQL), C, C#, Java, Scala, R.
- **Skills**: Python libraries (spark, scikit-learn, tensorflow, pytorch), natural language processing (transformers), AWS, Databricks
- **Technical Interests:** Multimodal language models, open-source technologies, video game development
- **General Interests:** Oil painting, learning languages, films, Fran Lebowitz