Naila Fatima

US Citizen | (331)-806-0716 | nfatima5596@gmail.com | Glendale Heights, IL 60139

Experience

Software Engineer April 2022 – Present

Andromeda 360 (spinoff of Hypergiant)

- Developed Terraform pipelines to orchestrate AWS Lambda functions and Glue jobs for data ingestion from external APIs, data preprocessing, and upserts into Apache Iceberg tables and MongoDB collections; managed separate dev/prod pipelines to ensure reliable deployments and smooth data workflows.
- Enhanced an ARIMA-based anomaly detection pipeline to forecast product subcategory sales and flag anomalies, enabling earlier detection.
- Developed and fine-tuned NL2SQL functionality to convert user questions into SQL queries, achieving 64% accuracy, and designed a caching system to reduce response times for similar questions.
- Spearheaded the maintenance and augmentation of the model development kit (MDK) by authoring comprehensive unit and regression tests (∼20% of the test suite), updating the Docker containers with new releases and improving documentation.
- Incorporated retraining functionality to the MDK, resulting in a 12% increase in model performance by preserving trained model artifacts in their native format, thereby allowing incremental updates and reducing deployment time.
- Implemented model versioning to enable users to create multiple models with the same name and different versions, simplifying model management.

Software Engineer Aug 2021 – Jan 2022

Hypergiant

- Built datasources with SQLAlchemy to connect the MDK to SQL (MySQL, PostgreSQL, SQLite) and Snowflake databases, enabling expedient querying.
- Automated hyperparameter tuning for Hyperdrive experiments with Optuna, reducing manual effort, and enabling users to achieve optimal performance by specifying the hyperparameters to be tuned.
- Developed a dynamic scheduler utilizing Papermill to execute Jupyter notebooks hourly, with the capacity to append, remove, and modify scheduled tasks.

Computer Vision Researcher May 2017 – May 2019

IIIT Hyderabad

- Created video blur detection models with Python and OpenCV which used variations in video frame intensities and a neural network to achieve 90.13% accuracy; project done in collaboration with Qualcomm.
- Co-led the development of a film shot classification technique which utilized pose estimation (via OpenPose) and a rule-based approach in Python; could distinguish between close-ups, medium shots and long shots with 77.5% accuracy.
- Implemented video stabilization techniques using L1 optimal camera paths and content preserving warps in MATLAB to reduce the impact of camera motion on video.

Education

Georgia Institute of Technology, Atlanta, GA Aug 2019 – Dec 2020

MS in Computer Science, Specialization: Machine Learning GPA: 3.9/4.0

International Institute of Information Technology (IIIT) Hyderabad, India Aug 2015 – May 2019

B. Tech (Honors) in Electronics and Communications Engineering, Dean's List GPA: 8.65/10.0

Skills

Languages: Python, C, C++, Java, SQL, HTML, CSS, MATLAB
Technologies & Tools: AWS, Terraform, Docker, Git, Flask, Bash, MySQL, PostgreSQL, MongoDB, Apache Spark
Libraries: OpenAI, TensorFlow, PyTorch, HuggingFace, OpenCV, Keras, scikit-learn, SQLAlchemy, OpenPose, Optuna

Projects

Automatic Essay Scoring (AES) with Bias Prediction (Python, PyTorch, scikit-learn) Link

- Developed an AES system using machine learning models (Bayes classifier, LSTMs, BiLSTMs) with a 97% agreement among scorers on the ASAP-AES dataset.
- Processed data for feature extraction and trained models to predict essay author demographics (age, gender) to analyze the possibility of bias in essay scoring; indicated a possible gender bias in AES.

GenSketch (Python, PyTorch, scikit-learn) Link

- Programmed unsupervised models (VAE, DCGAN) to generate doodles of specific objects which looked similar to the doodles in the Google QuickDraw dataset and could be correctly classified around 80 % of the time by a trained model.
- Trained a CNN classifier to detect object class with 92.33% accuracy to test the performance of the generative models.