(343)-549-4606
darasupriya99@gmail.com

Supriya Dara

Master of Computer Science Graduate

LinkedIn Portfolio
GitHub LeetCode

Education

Master of Computer Science in Applied Artificial Intelligence , University of Ottawa, GPA: 9.30/10.00 **Sept 2022 — May 2024**
Bachelor of Engineering in Computer Science, Chaitanya Bharathi Institute of Technology, GPA: 8.30/10.00 **Sept 2018 — May 2022**

Experience

Machine Learning Engineer, Aarth Software **Nov 2021 — Jul 2022**

- Led the implementation of an **NLP** project using **BERT** for automated content extraction from product listings, achieving **80% accuracy** in extracting and categorizing product features for improved inventory management.
- Built a scalable **web scraper** with **BeautifulSoup** and **Selenium** to collect data from e-commerce websites, automating metadata extraction and ensuring data consistency for product analysis across over **100** websites.
- Developed and maintained **ETL pipelines** for efficient extraction, transformation, and loading of large datasets.
- Designed and implemented a machine learning pipeline using **AWS S3** for data storage, **AWS Glue** for data transformation, and **AWS EC2** for model training, reducing model training time by **30%**.
- Worked on **A/B testing** to validate model performance against baseline solutions, providing actionable insights for model optimization and business strategy.

Data Science Intern, Exposys Data Labs **May 2021 — Aug 2021**

- Collaborated with senior data scientists to perform **exploratory data analysis (EDA)**, identifying key trends and correlations that led to more accurate model development.
- Developed a predictive model using the PIMA Diabetes dataset, achieving **88% accuracy** with **K-Nearest Neighbors (KNN)** to predict diabetes risk, contributing to health insights for medical decision-making.
- Built interactive dashboards using **Tableau** to track model predictions and present results to stakeholders, improving reporting efficiency.

Machine Learning Intern, National Instruments **May 2020 — Aug 2020**

- Conducted data preprocessing and feature engineering on large sensor datasets using **Pandas, NumPy,** and **Scikit-learn**, reducing training time by **30%**.
- Built a computer vision model with **OpenCV** and **TensorFlow** to automate defect detection in hardware components, improving quality control accuracy by **15%**.
- Assisted in developing **MLOps** workflows using **Docker, Jenkins,** and **Kubernetes**, streamlining the deployment and monitoring of ML models.

Skills

General Programming: Python | Java | C++ | R
Database: MySQL | Oracle | PostgreSQL | NoSQL (MongoDB, DynamoDB, Snowflake)
Big Data & Cloud: Apache Spark | Airflow | Kafka | AWS (EC2, S3, Lambda, SageMaker) | Azure | GCP
DevOps & Deployment: Docker | Kubernetes | GitHub | CI/CD Pipelines | REST APIs
Data Visualization & Analytics Tools: Tableau | Power BI | Matplotlib | Seaborn | Excel
Machine Learning: MLOps | NumPy | Pandas | Scikit-learn | PyTorch | TensorFlow | Keras | NLTK | spaCy | SHAP
Concepts: Statistical Modeling | Deep Learning | Large Language Models (LLMs) | Generative AI

Projects

Log-based Anomaly Detection (Python, PyTorch, Scikit-learn, Transformers)

- Designed a RoBERTa-based method for transforming unstructured log data and analyzing **time series patterns**, achieving an **F1-score of 0.99** in identifying anomalous logs, surpassing traditional **TF-IDF** methods.

Phishing URL Detection (Python, Scikit-learn, PyTorch, Transformers, SHAP)

- Led a RoBERTa-based project to detect phishing URLs, balanced the dataset with different **sampling methods**, and used SHAP for interpretability, achieving **98.34% accuracy**.

Automated Code Review with NLP (Python, Scikit-learn, NLTK, PyTorch, Transformers)

- Created a RoBERTa model for optimizing code review in software development, achieving **97.7% accuracy** in analyzing GitHub comments for both accuracy and efficiency.

Automated Essay Scoring System using LSTM and NLP (Python, TensorFlow, Keras, NLTK, LSTM)

- Constructed a 2-layer LSTM network and used NLP techniques to evaluate and rate essays, achieving a **QWK score of 0.92**.