# Supriya Dara

343-549-4606  ✉ darasupriya99@gmail.com  in LinkedIn  Portfolio  Github

## Education

**University of Ottawa** | Sept 2022 - May 2024
*Master of Computer Science Concentration in Applied Artificial Intelligence (GPA: 9.30 / 10)*

**Chaitanya Bharathi Institute of Technology** | Sept 2018 - May 2022
*Bachelor of Engineering in Computer Science (GPA: 8.30 / 10)*

## Experience

**Machine Learning Engineer** | Nov 2021 – July 2022
*Aarth Software*

- Built a Python web scraper with **BeautifulSoup** for automating metadata extraction from URLs, demonstrating data engineering, achieving a **98% success rate** in scraping product schema from the top 5 Google search results.
- Implemented **NLP project** to extract and process recipes from 100+ websites, achieving **80% accuracy** using **BERT** for ingredient parsing, text classification, and entity recognition.
- Utilized **AWS S3** for storing and managing data during model development, optimizing data processing pipelines.
- Created interactive dashboards using **Power BI** to visualize project outcomes and key performance indicators for management reviews.

**Data Science Intern** | May 2021 – Aug 2021
*Exposys Data Labs*

- Engineered a Diabetes Prediction model in Python, applying **KNN** algorithm and achieving **88%** accuracy.
- Built performance dashboards using **Tableau** to highlight trends and insights in diabetes prediction outcomes, aiding decision-making processes.
- Deployed model workflows using **AWS Lambda** and monitored performance with **CloudWatch**, ensuring scalability and reliability.

**Machine Learning Intern** | May 2020 – Aug 2020
*National Instruments*

- Developed an ensemble ML model using **regression algorithms**, achieving **93% accuracy** in predicting hospital stay lengths for diabetic patients using **SVM**.
- Implemented a sentiment analysis model in Python, achieving **95%** accuracy through different ML algorithms.
- Utilized **Excel** for comprehensive data analysis and visualization, employing pivot tables, charts, and advanced formulas.

## Projects

**Log-based Anomaly Detection** | *Python, PyTorch, Scikit-learn, Transformers*  Link

- 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*  Link

- 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*  Link

- 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*  Link

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

## Technical Skills

- **General Programming:** Python, Java, C++, R.
- **Database:** MySQL, Oracle, PostgreSQL, Snowflake (learning).
- **Web Technologies & Cloud:** HTML, CSS, JavaScript, React, AWS, Azure.
- **Data Visualization  Analytics Tools:** Tableau, Power BI, Matplotlib, Seaborn, Excel.
- **Machine Learning:** NumPy, Pandas, Scikit-learn, PyTorch, TensorFlow, Keras, NLTK, spaCy, HuggingFace, SHAP.
- **Concepts:** Statistical Modeling, Deep Learning, Large Language Models (LLMs).