Dominic Parosh Yamarthi

136 Heath St, Upper Unit, Buffalo, New York - 14214 | dominicparoshyamarthi@gmail.com | 716-247-3129 | Portfolio | LinkedIn | GitHub

Data Science graduate with 4 years of experience in Document Extraction, NLP, Data Analysis, Machine Learning, Big Data, Generative AI, and Large Language Models. Proven ability to develop and deploy advanced models and solutions to optimize data processing and analysis. Seeking a full-time position and available for immediate joining.

PROFESSIONAL EXPERIENCE

Graduate Teaching Assistant | University at Buffalo, Computer Science Department | Buffalo, USA. January 2024 – May 2024

- Appointed as a Graduate Teaching Assistant for the Data Intensive Computing (CS 487/587) course, managing a cohort of 200 students. Designed and developed projects, 3 assignments, 11 quizzes, and 2 term papers, ensuring alignment with course objectives. Graded and provided detailed feedback on over 200 student submissions, contributing to a 20% improvement in overall student performance. Organized and administered 2 term examinations, maintaining a high level of academic integrity.

Skillset – Python, Java, Eclipse, Hadoop, Spark, Hive, Jupyter notebook, Machine Learning.

Machine Learning Engineer | TechNVision | Hyderabad, India. August 2019–December 2022

- Implemented a codebase enabling the cleaning, extraction, and analysis of details from Bank Statements of Bank of Santander, resulting in a 20% reduction in processing time.
- Conducted Image Annotations, Preprocessing, and Augmentation for over 50000 images.
- Implemented code leveraging Beautiful Soup and PyTorch to detect and extract multiple languages from financial forms, leading to a 30% improvement in accuracy compared to previous methods.
- Successfully executed Multilabel Image Classification using Faster R-CNN and YOLO v3 architectures, achieving an average accuracy of 85%.
- Engineered an OCR Pipeline to process and extract meaningful information from identified regions, resulting in a 25% increase in extraction accuracy.
- Designed and deployed Models and Restful APIs using Flask on IIS Web Server, handling over 1000 requests per day with a response time of under 100 milliseconds.
- Enhanced ML models tailored to the company's requirements, resulting in a 15% improvement in forecast accuracy compared to previous methods.
- Implemented a rule-based algorithm using OCR and ML models to extract data from various file formats, processing over 10,000 files per month.
- Utilized ML models to categorize text in files, determining types and grouping them into appropriate folders, enabling efficient extraction code execution on over 95% of files.
- Extracted handwritten data from remittance documents, bank statements, etc., processing over 1000 documents per week.
- Contributed to Solix, a sister company of TechNVision, using Java to provide a large data storage platform and analytics platform, leveraging Apache Spark and Spring Boot frameworks to handle 100 terabytes of data daily.

Skillset – Python, Java, Tableau, Power BI, Postman, Rocket Lane, SQL, NLP, AI, CV, Deep Learning, Reinforcement Learning, IR, AWS, GCS, Azure, NoSQL, Redshift, Oracle, Cuda, Generative Ai and Large Language Models.

EDUCATION

University at Buffalo, The State University of New York–MS in Engineering Science Data Science May 2024

Coursework: Numerical Mathematics, Introduction to Probability, Statistical Data Mining (Supervised and Unsupervised Learning), Introduction to Python, Introduction to Machine Learning, Computer Vision, Data Intensive Computing, Data Model Query Language

Gayatri Vidya Parishad College of Engineering(A)–B.Tech in Computer Science and Engineering May 2018

Coursework: Data Structures, AI, Design and Analysis of Algorithms, OS, Computer Architecture, Software Engineering, Digital Image Processing.

TECHNICAL SKILLS

Programming and Scripting Languages: SQL, Python, R, Java, Spark, C, C++, JavaScript, HTML, CSS, Swift

Databases: MySQL, MongoDB, SQLite, SQL Server, PostgreSQL, NoSQL, Oracle

Data Visualization: Tableau, PowerBI, Matplotlib, Seaborn, Plotly, ggplot2

Frameworks and Libraries: Node.js, Flask, Django, Streamlit, XCode, NumPy, Pandas, Scikit-learn, Keras, PyTorch, TensorFlow, XGBoost

Machine Learning and AI: Machine Learning, Artificial Intelligence, Deep Learning, Reinforcement Learning, Natural Language Processing, Computer Vision, Data Modeling (Regression, Classification), Neural Networks, AutoML, Gen AI, Large Language Models, Responsible AI, Generative AI, Large Language Models, Transformers.

Big Data and Cloud Technologies: Apache Hadoop, Spark Streaming, MapReduce, AWS S3

Software and Tools: VS Code, RStudio, Anaconda, JIRA, Bitbucket, Postman, Rocket Lane, Microsoft Office (Excel, Word, PowerPoint), Jupyter, Eclipse, Vertex AI

Other Skills: Data Science, Data Analysis, Data Visualizations, Image Preprocessing

Version Control: Git

PROJECTS

- **Predicting Nationwide traffic Accidents, the US using Random Forest Regression:** This research project aims to analyze an extensive US traffic dataset, revealing patterns that can inform evidence-based policies and interventions, with the goal of quantifiable mitigating risks and improving nationwide road safety. Predicted the accidents with 80% accuracy and the **research paper is being published by August.**
- **Estimation of Vehicle Weight for Structural Load Analysis**: Deployed advanced computer vision algorithms for instantaneous vehicle detection, classification, and weight estimation, seamlessly integrating them into conventional tracking systems to provide a live video feed featuring real-time data on vehicle class, count, total weight, and alerts, thereby significantly enhancing monitoring capabilities. Detected and estimated with an accuracy of 95%.
- **Automated Machine Learning for automated hyperparameter tuning for Synthetic Data Generation**: To streamline neural network development, a model was created to automatically select optimal hyperparameters, reducing errors associated with manual tuning during development. Achieved 10% increase in accuracy as compared to manual hyper-parameter tuning.
- **Creating Music using Generative AI (RNN):** Generated new music sequences after training an RNN on a dataset of musical pieces. This can be evolved to create background scores for various movie characters based on the attributes of the said character. Generated new music sequences after training an RNN on a dataset of musical pieces. This can be evolved to create background scores for various movie characters based on the attributes of the said character.
- **Gender Prediction Using RNN**: With an accuracy rate of 72%, the objective entails predicting the gender (Male/Female) from provided sonogram images.

AWARDS AND CERTIFICATIONS

- Stood second in a startup pitching competition | SUNY Buffalo, Blackstone.
- Competed in CS Day for one of the projects in SUNY Buffalo, Computer Science
- Introduction to TensorFlow for Artificial Intelligence, Machine Learning, and Deep Learning.
- Generative AI | Responsible AI | LLM certification from Google.