Dominic Parosh Yamarthi

(716) 247-3129 | dominicparoshyamarthi@gmail.com | LinkedIn

EDUCATION

- **University at Buffalo, The State University of New York** Jan 2023 – May 2024
  Master of Science (MS) - Data Science
- **Jawaharlal Nehru Technological University, Kakinada** Aug 2014 – Jul 2018
  Bachelor of Technology (B.Tech) – Computer Science and Engineering

PROJECTS

- **Prompt-to-Song Generation using Large Language Models:** Fine-tuned LLMs to generate complete songs from textual descriptions. This involved generating lyrics with GPT-2, classifying genres with BERT and DistilBERT, and creating chord progressions conditioned on lyrics and genre using GPT-2.
- **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.
- **NovelConvo AI, Intelligent Conversations with Literary Companions**: Integrated Chatterbot with diverse datasets, employed DeBERTa-v3 for classification, and developed an efficient Information Retrieval system with DAAT Boolean query processing. Enhanced responses using OpenAI's Retrieval Augmented Generation.
- **Inverted index, Boolean query**: Implemented and optimized an Inverted Index in Python3, encompassing a preprocessing pipeline for text normalization and tokenization, a Document-at-a-Time (DAAT) search strategy for Boolean queries, and enhanced posting lists with skip pointers for efficient query processing. Additionally, integrated tf-idf scoring for relevance sorting.
- **Adaptive Deep learning for Environment-Agnostic Human Action Recognition**: Developed an advanced deep learning system for human action recognition using a subset of 10 classes from the UCF101 dataset, implementing AlexNet3D and VGG3D models. Achieved significant metrics with 39% accuracy and 37.51% precision in diverse settings.
- **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%.
- **Gender Prediction Using RNN:** With an accuracy rate of 72%, the objective entails predicting the gender (Male/Female) from provided sonogram images.
- **Predicting Nationwide traffic Accidents in 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*.
- **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.
- **Airline Baggage Complaint Analysis**: Conducted time series analysis on airline baggage complaint data using classical decomposition, ETS(AAA), and ARIMA models to forecast trends and identify patterns. Implemented data preprocessing and visualization techniques, and evaluated model performance using metrics like AIC, BIC, and p-values to determine the best-fit model
- **Database Management System with Tableau Visualization**: Designed and maintained a comprehensive Fifa23 Players database using advanced database design theory, Python for data preparation, and E/R modeling with normalization techniques. Crafted efficient SQL queries for retail analytics and developed an interactive Tableau dashboard to provide dynamic visual insights into player statistics.
- **Database Normalization and Analytics**: Designed and implemented a relational database with normalized tables, enforced referential integrity using foreign keys, and developed complex SQL-based data processing functions, including advanced joins and subqueries, to enable detailed customer-focused analysis and sales trend identification.

WORK EXPERIENCE

- **Graduate Teaching Assistant | University at Buffalo** Jan 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
  - **Technologies**: Python, Java, Eclipse, Hadoop, Apache Spark, Spark Streaming, Hive, Jupyter notebook, Machine Learning, Deep Learning, Streamlit.

- **Machine Learning Engineer | Emagia** Mar 2020 – Dec 2022
  - Maintained data reporting for PNG, focusing on D2C and D2B marketing & E-commerce, handling raw data from PNG shopping pages across various countries via Google Analytics streams.
  - Conducted detailed analysis and reporting on key metrics such as NOS penetration growth and IYA categorization, using SQL and data visualization tools, segmented by business units and countries.
  - Processed and analyzed raw data from PNG shopping pages for Male Grooming, Mother and Child Care, and Sleep Respiratory Business clusters using advanced data mining techniques and visualized insights with Power BI, enabling strategic decision-making and improved business performance.
  - **Technologies**: Python, Anaconda, Tableau, Power BI, Postman, Rocket Lane, SQL, NLP, AI, CV, Deep Learning, Reinforcement Learning, IR, AWS, GCS, Azure, NoSQL, Redshift, Oracle, Cuda, GenAI and Large Language Models.
- **Junior Software Engineer | Solix** Aug 2018 – Mar 2020
  - Developed and maintained dashboards using Tableau for in-depth analysis and insight derivation, supporting process tracking, performance, and quality metrics for A/B testing on calling experiments in the India-SA region.
  - Conducted a comprehensive analysis of the Hyderabad COE - India Scaled Ops Team using advanced data analytics, resulting in a 44% reduction in processing costs and a 200% increase in agent efficiency.
  - Enhanced an existing Python script for operational efficiency, resulting in a potential savings of approximately 240-man hours annually.
  - **Technologies**: Java, Spring Boot, RESTful Services, Microservices, Hibernate, JPA, Maven, Jenkins, Docker, Kubernetes, AWS (EC2, S3, RDS), Apache Kafka, RabbitMQ, MySQL, PostgreSQL, Redis, MongoDB, Elasticsearch, Git, CI/CD pipelines, Unit Testing (JUnit, Mockito), Logging (Log4j, SLF4J), API Documentation (Swagger), JSON, XML, OAuth2, JWT, WebSockets.
- **Intern | Singareni Collieries Company** May 2017 – May 2018
  - Developed and maintained advanced dashboards using Tableau to monitor key operational metrics, providing actionable insights that supported data-driven decision-making across various departments.
  - Implemented predictive models to forecast equipment failure rates, improving maintenance schedules and reducing downtime by 15%. Utilized Python and scikit-learn for model development, and optimized hyperparameters to enhance prediction accuracy.
  - Streamlined data collection and preprocessing pipelines using Python and SQL, automating data ingestion from multiple sources, which resulted in a 30% improvement in data processing efficiency.
  - Conducted a detailed analysis of operational data using machine learning algorithms, leading to a 44% reduction in processing costs and a 200% increase in task efficiency within the Hyderabad COE - India Scaled Ops Team.
  - Enhanced existing scripts with natural language processing (NLP) capabilities to automate report generation, reducing manual effort and saving approximately 240 man-hours annually.
  - **Technologies**: Python, SQL, Tableau, scikit-learn, Pandas, NumPy, Jupyter Notebook, Git, Data Visualization, Predictive Analytics, Machine Learning, Data Pipeline Automation.

SKILLS

- **Programming Languages:** Python, Java, SQL, R, Bash, Shell Scripting.
- **Web Development:** JavaScript, HTML, CSS, Spring Boot, RESTful Services, Microservices, JSP, Servlets, Bootstrap, jQuery, Flask, Django.
- **Database Management:** MySQL, PostgreSQL, SQLite, BigQuery, Oracle, MongoDB, Redis, NoSQL, SQLAlchemy, psycopg2, Database Design, ER Modeling, SQL Queries, Database Normalization, Referential Integrity.
- **Business Intelligence & Visualization:** Tableau, PowerBI, Excel, SQL, Data Mining, A/B Testing, ETL, Reporting, Data Visualization, Dashboard Development, Matplotlib, Seaborn, Plotly.
- **Machine Learning & AI:** Model Development, Data Cleaning, Time Series Analysis, Predictive Analytics, Neural Networks, Deep Learning, Reinforcement Learning, Natural Language Processing (NLP), Transformers, Hugging Face, RNNs, LLMs, BERT, GPT-2, DeBERTa-v3, Vector Databases, RAG, Text Classification, Sentiment Analysis, Hyperparameter Tuning, scikit-learn, TensorFlow, Keras, PyTorch, OpenCV, NLTK, SpaCy, Gensim.
- **Big Data Technologies:** Hadoop, Apache Spark, Spark Streaming, Hive, Data Intensive Computing, Distributed Systems, Data Pipelines, Parallel Computing, HDFS, Pig, Sqoop, Flink, Kafka Streams.
- **Cloud & DevOps:** AWS (EC2, S3, RDS), Google Cloud Platform (GCP), Azure, Docker, Kubernetes, Jenkins, CI/CD, Terraform, Ansible, CloudFormation, Lambda, Fargate.
- **Version Control & Collaboration:** Git, GitHub, GitLab, Bitbucket, CI/CD Pipelines, Branching Strategies, Code Reviews, Agile Methodologies, JIRA, Confluence, Trello.
- **APIs & Microservices:** RESTful APIs, Swagger, JSON, XML, OAuth2, JWT, WebSockets, API Documentation, Postman, RabbitMQ, Apache Kafka, Maven, JPA, Hibernate, Jenkins, Logging (Log4j, SLF4J).
- **Data Analysis & Statistics:** SQL, R, Excel, Exploratory Data Analysis (EDA), User Behavior Analysis, A/B Testing, Data Wrangling, Statistical Modeling, Hypothesis Testing, ANOVA, Pandas, NumPy, SciPy, Statsmodels.
- **Operating Systems:** Linux, Windows, Unix, Ubuntu, CentOS, Red Hat.
- **Software Development Tools:** Eclipse, Jupyter Notebook, Anaconda, PyCharm, IntelliJ IDEA, Visual Studio Code, Streamlit, Docker Compose, Vagrant.