Tarunvel Venugopal Santhamani
+1 413-472-9657 • tarunvel.vs@outlook.com • LinkedIn • GitHub • Portfolio • Tableau Portfolio

TECHNICAL SKILLS

Programming: Python, SQL (Oracle, Postgres, Snowflake)
Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, TensorFlow, NLTK, PySpark (SparkSQL, SparkML)
Tools: Tableau, MS PowerBI, MS Excel, Git, GitHub, Atlassian Jira, AWS (SageMaker)
Techniques: ML Algorithms (Supervised & Unsupervised), Natural Language Processing, Large Language Models, Neural Networks, Statistics, ETL Pipelines, Data Warehouse, Agile, Scrum, MLOps

WORK EXPERIENCE

University of Massachusetts, Amherst (Computational Social Science Institute) *Amherst, MA, US*
Researcher (Data Science) *Feb 2024 - Present*

- Conducted an extensive synthesis of scholarly articles, distilling complex research methodologies to establish a methodological framework for the project; Developed an algorithm using Scikit-Learn and TensorFlow to predict the Food Consumption Score.

Tata Consultancy Services Limited | Client: Citi Bank; CoreLogic *Chennai, TN, India*
Software Development Engineer *Sep 2020 – Aug 2022*

- Developed programming scripts to generate tax forms (1099 and 1098) in the rental property management software for tax filing.
- Researched File Handling System & its workflows in the software and presented it to the stakeholders; Developed & Implemented a library file to optimize the process, resulting in a 10% decrease in the processing time.
- Researched the data processing of unstructured data (JSON, XML, Documents) in the software to improve workflow efficiency.
- Implemented customized report template where user can select the data fields required in the report, created programming scripts to fetch data from different tables in the Database.
- Developed and implemented a secure process for fetching check image files from banking servers using RESTful APIs, ensuring compliance, and handling authentication, authorization, and encrypted data transmission.
- Designed and deployed the "Mail Management" feature in the software to provide enhanced features & recommendations for sending mails and alters from the software.
- Designed and created test scripts for implementing Progress 11.7 Migration and software version up-gradation; Awarded "Star Team" for successful deployment of the upgraded version to all SAAS customers in Azure.
- Created Excel Macros to develop a report template to automate the manual parts in the daily status reports; Refined & prioritized over 200 product backlogs in agile sprint using JIRA; Maintained software version using git & SVN; Developed & executed test scripts for regression, unit & functional test case & managed those using Zephyr.

PROJECTS

[ML] Diabetes Prediction [GitHub Link] [Presentation Link] | Logistic Regression, Classification, Decision Tree, Clustering

- Data Cleaning & transformation are executed on 100K healthcare records; Outliers & Patterns are detected through Exploratory Data Analysis along with statistical analysis; Fine-tuned Logistic Regression model by varying Threshold-Value
- Implemented GridsearchCV and Cross-Validation to prune the decision tree and tune the KNN Classifier; In comparison, Logistic Regression has fewer False Negative cases and a better recall percentage of 80%

[ML, Neural Network] Customer Churn Prediction Model Using Artificial Neural Networks [GitHub Link]

- Developed predictive model to identify customer churn using ANN. Data Preprocessing, & exploratory data analysis (EDA) are done using Pandas, Matplotlib, & Seaborn. ANN model is built using TensorFlow

[RAG, LLM] Chat with PDF [GitHub Link] | Sentence Transformers (all-mpnet-base-v2) , HuggingFace

- Developed an LLM model that would develop output based on the context of the PDF document and query. The model is created using pre-trained LLM from HuggingFace.

[ML, NLP] Sentiment Analysis (Tweets & Reviews) [GitHub Link]

- Text Preprocessing, tokenization, and stemming are carried out to obtain vocabulary from corpus followed by vectorization; Classifier models are created and achieved 85% accuracy in the random-forest model

EDUCATION

University of Massachusetts, Amherst *GPA: 3.9/4.0*
Master of Science in Business Analytics *Jan 2024*

- Coursework (Focus: Data Science): Machine Learning, Data Mining - Neural Networks, Application Development, Statistics & Probability, Business Intelligence & Big Data System, Accounting & Finance, Marketing - Strategy, CRM, Supply Chain Analytics

CERTIFICATIONS

"Advanced Data Analytics" – Google; "Advanced SQL" – Hackerrank; "Python for Data Science" – IBM; "Data Analysis" – Tableau

ACHIEVEMENTS

"SEMI-FINALIST" in TCS CodeVita 2K19 (Global Coding Contest) | RESEARCH ASSISTANT: Product Designing & Analysis