

Ajith Kumar Ethirajulu

Buffalo, New York | 716-256-4115 | aethiraj@buffalo.edu

LinkedIn: www.linkedin.com/in/aje2907 | GitHub: <https://github.com/aje2907> | Portfolio: <https://aje2907.github.io>



EDUCATION

University at Buffalo, The State University of New York

Buffalo, NY

Master of Science – Data Science

EXPERIENCE

Research Project - 3D Protein Structures prediction for Huntington's Disease leveraging AlphaFold

Under the guidance of Professor Dr. Murali Ramanathan, University at Buffalo
2023

May 2023 – September

- Evaluated Google's DeepMind AlphaFold AI model for healthcare and pharmaceutical applications, comparing predicted and established protein structures, with a focus on Huntington's disease research.
- Led over 80 experiments using AlphaFold to predict HTT protein structures, analyzing regions with low confidence scores (<50 pLDDT) and forecasting potential protein aggregation applying the Multimer model.

Fractal Analytics

Bengaluru, KA, India

Data Engineer

July 2021 – April 2022

- Merged extensive logistics data from multiple file sources, executed data cleaning, and employed Python for data manipulation. Implemented outlier handling for numerical data, addressed null values, and standardized the data to enhance data quality.
- Designed data preprocessing pipeline on Azure Machine Learning workspace, composed of data loading, transformation, and data storage components as subtasks. Automated the pipeline through scheduled execution.
- Reduced dimensionality by creating derived attributes from high dimensional raw data during transformation stage by utilizing several features, thus reducing number of features by nearly 20%.

Infosys Limited

Chennai, TN, India

Senior Systems Engineer

April 2018 – July 2021

- Collaborated with a banking client to gather requirements, implemented rule-based scenarios using Python and SQL, and analyzed datasets from 50,000 to 30 million records, employing DQL queries and SQL analytical functions.
- Integrated MySQL with Python for seamless data transformation, engaged in comprehensive exploratory data analysis using Pandas, NumPy, Matplotlib, and Seaborn; Devised modular Python scripts with 20+ functions for database tasks, including connectivity, data manipulation, and standardization, improving operational efficiency.
- Applied customer profile gating, executed enterprise rules to enhance customer penetration, and conducted segmentation, resulting in a 1.2% boost in lead-to-deal conversion and customer penetration. Established a robust MySQL-Power BI connection, leveraged Power Query editor to carry out data transformations, and crafted 10+ calculated fields, measures, aggregations, and drill-down features.
- Engineered a procedural SQL script to address a critical data persistence issue, automating manual tasks and reducing issue resolution time by 50% to avoid SLA breaches. Developed SQL scripts for DB archiving and cleanup.

PROJECTS

Time Series Forecasting on Retail Store Sales in R

February 2023 – May 2023

- Led an empirical study forecasting retail sales applying time series methods like ETS, ARIMA, and Prophet in R, on a dataset of 3 million records from 54 stores, identifying Prophet as the best-performing model.
- Interpreted sales components, trends, and seasonality, conducting rigorous model training and testing. Attained best Root Mean Squared Error of 474.02 and Mean Absolute Error of 269.9.

Skin Melanoma Prediction

September 2022 – December 2022

- Leveraged Convolutional Neural Networks to build a binary classification model, trained and validated with a dataset containing about 33000 images to estimate the probability of a given image being skin cancer (melanoma).
- The highest accuracy achieved was 94.1% and 94.6% AUC with the help of EfficientNetB4 transfer learning model. Deployed a web app using Python Flask and HTML to predict skin melanoma on the uploaded images.

Credit Defaulter Analysis and Prediction

September 2022 – December 2022

- Integrated SQLite and Python to store, perform exploratory data analysis, and feature extraction on the vast data containing 120 features and 1.2 million records.
- Built binary classification machine learning model using Support Vector Machines, RandomForest, and Tensorflow (Keras). The highest accuracy obtained was 91% by the optimized Tensorflow model.

Online Retail Sales Analysis using Power BI

June 2021 – August 2021

- Analyzed sales data using Power BI, transformed Excel spreadsheet data into a detailed report, utilizing DAX for calculated measures and custom columns; studied trends and patterns to identify the Key Performance Indicators.
- Interesting observations from the data are that the store's predominant sales are from North America, with customers over 40 years of age contributing to more than 65% of total sales, indicating a significant sales increase in 2013.

Analysis of Loan Defaulters – Tableau

February 2021 – May 2021

- Engineered a dynamic Tableau dashboard featuring comprehensive visualizations showcasing KPIs pertaining to loan defaulters. The resulting insights assist in making informed communication to the business decision-makers.
- Deeper analysis helped to arrive at some key insights like 43% of defaulters had changed their phone numbers in the last 1 year, 80% have less than 3 years of work experience, and 41% have changed their jobs in the last 3 years.

SKILLS

Programming: Python, R, SQL, Machine Learning, Deep Learning, Reinforcement Learning, Git, GitHub
Databases: MySQL, SQLite, Postgres SQL, Data Warehousing
Cloud: Azure Data Factory, Azure Data lakes, Azure Machine Learning
Visualization: Tableau, Power BI, Seaborn, Matplotlib

CERTIFICATIONS

- **AI for Medical Diagnosis** – DeepLearning.ai September 2023
Credential: <https://www.coursera.org/account/accomplishments/certificate/QMZVST438MCX>
Skills obtained: Deep Learning, Image Segmentation, Healthcare data, Clinical Data Analysis, Model Evaluation.
- **Neural Networks and Deep Learning** – DeepLearning.ai September 2023
Credential: <https://www.coursera.org/account/accomplishments/certificate/V5PS7PJAMGFE>
Skills obtained: Deep Learning, Artificial Neural Network, Backpropagation, Python, Neural Network Architecture.
- **Microsoft Certified: Azure Data Engineer Associate** – Microsoft September 2021
Credential: [Microsoft credential](#)
Skills obtained: Azure Data Factory, Azure Data lakes, Azure SQL database
- **Microsoft Certified: Azure Data Fundamentals** – Microsoft February 2021
Credential: [Microsoft credential](#)
Skills obtained: Relational/ Non-relational data, Transactional/ Analytical workloads

ACHIEVEMENTS

INSTA Award - Infosys Limited

July 2020

For having consistently demonstrated outstanding teamwork skills and proactively stepping up in the absence of the team lead, effectively guiding the team and supporting the applications through critical deployment stages without any SLA violations and achieving a progressive reduction in the number of issues by 80%.