Sai Yeshwanth Mekala

7206594782| yesh20@icloud.com | LinkedIn | Portfolio | GitHub | Boulder, CO.

PROFESSIONAL EXPERIENCE

JOHNS MANVILLE   **Littleton, CO, United States**
Data Scientist Intern   **May 2024-Present**

- Architected an efficient ETL pipeline, integrating large data from sources including Microsoft SQL, Salesforce, SAP, Seeq API, and Historian servers into Jupyter Lab, resulting in a 75% improvement in data collection and preprocessing time.
- Spearheaded stationary check and data transformation utilizing ADF and KPSS tests, achieving 100% stability of time series.
- Analyzed seasonality using ACF, PACF, Seasonal Decomposition graphs and statistical analysis with One-Way ANOVA and Kruskal-Wallis rank sum non-parametric Hypothesis Testing, achieving 100% evidence of periodic patterns every 6 months.
- Performed Sequential and Supervised Deep Learning (TensorFlow, Keras) Regression predictive modeling including Long Short-Term Memory (LSTM), Bidirectional LSTM and Gated Recurrent Unit (GRU) on scaled time series, gaining 70% R-Squared.
- Developed state-of-the-art Transformers architecture with attention mechanism in Python (Pytorch), gaining 75% R-Squared.
- Revamped current process with 31% improvement utilizing TFT with SHAP Explainer by reporting Key Performance Indicators (KPI) and waterfall charts enhancing Anomaly detection yielding clear and actionable insights for Ad-hoc Business Efficiency.
- Improved Neural Network training with Adam optimizer cost function, leading 40% faster convergence; researched with Sigmoid, ReLU, and Tanh activation functions across hidden and dense layers to enhance performance.
- Augmented the Neural Network model robustness by integrating 20% dropout regularization, minimizing Overfitting.
- Enhanced reusability by 100% by employing Early Stopping callback, monitoring validation loss with patience of 10 and Deep Learning Model Checkpoint to store the best model for retrieval utilizing Keras and Pickle in Python.
- Led the Data Analysis in R through Correlation analysis and Hypothesis Testing with Randomization at 5% significance level.
- Leveraged Machine Learning algorithms and Deep Learning libraries in Python including TabNet (Pytorch) for tabular data, XGBoost for gradient boosting, LightGBM, and Random Forest for Ensemble Learning, achieving a 33% R-Squared.

TATA CONSULTANCY SERVICES   **Hyderabad, Telangana, India**
Systems Engineer   **July 2021-June 2023**

- Developed SQL, PL/SQL scripts and test plans in Oracle as a Database Administrator and gained 100% resolution time.
- Expedited KT process by 50% by documenting cross-functional Standard Operating Procedures (SOP) in Word with SQL scripts.
- Created Kanban board and Dashboards in Excel to showcase effectiveness to SME, resulting in a 96% Client Satisfaction Index.
- Earned On the Spot award in 2022 for exceptional client feedback for Data Analysis and automation and 15% boost in CSI.

PROJECTS

Heart Disease Identification from Patients' Social Posts, Machine Learning Solution on Spark

- Executed RF classifier in Pyspark, achieving an accuracy of 94.9% by Hyperparameter tuning and K-fold cross validation.
- Streamlined real-time big data ingestion by integrating Apache Spark and Apache Kafka framework, saved 50% processing time.

Data-Driven Meteorology: Unveiling Insights from Weather Data

- Applied K-Means and Hierarchical clustering for 100% optimal clusters (5) leveraging Elbow and Silhouette Method.
- Generated Top 10 rules by Support, Lift and Confidence through Unsupervised learning using Association Rule Mining (Apriori).
- Executed Supervised Machine Learning algorithms including Gaussian and Multinomial Naïve Bayes, Support Vector Machine with linear, RBF and Polynomial kernel and Decision Tree and recorded the highest accuracy of 67.98%.

PUBLICATIONS

Breast Cancer Detection Using Machine Learning (ijatcse12922020.pdf)   **April 2020**

- Performed Exploratory Data Analysis, implemented a high performing Logistic Regression model with an accuracy of 99.3%.
- Collaborated with a professor and published the research work and gained 24 citations highlighting impactful research.

EDUCATION

UNIVERSITY OF COLORADO BOULDER   **Boulder, CO, United States**
Master of Science in Data Science   **August 2023-May 2025**

SASTRA DEEMED UNIVERISTY   **Thanjavur, Tamil Nadu, India**
Bachelor of Technology in Information and Communication Technology   **June 2017-July 2021**

SKILLS

Programming: Python, R, Java, C++, C, Pandas, NumPy, Mathematics, Statistical Modeling, Hypothesis Testing, Predictive Analytics, Software Development, Natural Language Processing (NLP), Relational Databases, MySQL, MongoDB, Data Structures, Scikit-Learn, Git, IBM Watson Studio, Google Cloud, AWS, Web Scrapping, Azure ML, Databricks, CI/CD, ML Ops
Tools: Jupyter Notebook, R Studio, Eclipse, VS Code, Spyder, Minitab, SQL Server Management Studio, Hadoop
Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly, ggplot2, Excel
Interests: Statistics, Data Mining, Transformers, LLM, Multimodal Models, Artificial Intelligence (AI), Big Data Analytics
Languages: English, Telugu, Hindi