Praveen Kumar Arul

Fremont, Ca • +1-510-320-8160 • a.praveenarul@gmail.com • Praveen Kumar Arul | LinkedIn

Data Scientist / Machine Learning Engineer with hands on experience in statistical analysis, data visualization, building ML pipelines, machine learning package development, Integration, project management and client engagement. Dedicated to pioneering innovative and reliable data-driven solutions for business problems.

PROFESSIONAL EXPERIENCE

Accenture Solutions, Chennai, India   **October 2022– January 2023**
Application Development Associate
- Utilized predictive analytics on financial data to identify potential risks proactively. This led to 30% reduction in the occurrence of financial anomalies, safeguarding assets and improving the overall risk management strategy.
- Applied data analytics tools including Python and SQL, to inform strategic decision-making. Conducted in-depth analysis of large datasets, employing regression and clustering techniques to identify key trends, patterns, and opportunities.
- Engaged Excel functions for data manipulation and trend analysis to extract valuable insights, while also creating dynamic reports and visualizations in Tableau. These actions empowered data-driven decision-making processes.

SKILLS

Techniques: Regression, Classification, Clustering, Statistical Modeling, Quantitative Analysis, Statistical Inference, Hypothesis Testing, Feature Engineering, Feature Selection, Time Series Analysis and Forecasting, Sequential Pattern Mining, Survival Analysis, Data Mining, A/B Testing, Natural Language Processing, Object Oriented Programming
Programming languages: Python, R, Java, C, C++
Packages and Tools: Numpy, Pandas, Scikit- learn, OpenCV, Plotly, Scipy, Keras, TensorFlow, Pytorch, Docker, Tableau
Big Data/DBMS: SQL, Spark, BigQuery
Cloud Platforms: GCP, Azure, AWS
Version Control: Git, MLflow

EDUCATION

Arizona State University   Master of Science in Information Technology (GPA:4.0)   **January 2023 – May 2024**
Data Visualization and Reporting, Advanced DBMS, Info Systems Development, Analyzing Big Data, Cloud Architecture
Anna University   Bachelor of Technology in Information Technology   **July 2018 – April 2022**

ACADEMIC PROJECTS

Disease Risk Prediction Using Healthcare Data:
- Developed a predictive model to assess the risk of cardiovascular disease using clinical data, achieving 90% accuracy. Conducted feature selection and engineering to identify the most influential risk factors.
- Applied supervised learning algorithms such as logistic regression, decision trees, and ensemble methods like random forest to build and validate the predictive model. Used techniques such as cross-validation and hyperparameter tuning to optimize model performance.
- Employed the model to provide personalized risk assessments, leading to a 30% reduction in cardiovascular-related hospitalizations and healthcare costs.

Insurance Claim Segmentation and Analysis:
- Analyzed customer data and applied data transformations to handle skewness and outliers, improving model accuracy. Conducted statistical analysis and hypothesis testing to derive actionable insights from the data.
- Implemented and compared multiple clustering algorithms, including K-means and hierarchical clustering, to segment customers into distinct groups, optimizing model performance through iterative tuning and validation.
- Created interactive dashboards using Tableau to present segmentation results and insights to stakeholders, facilitating data-driven decisions for targeted marketing strategies and personalized customer engagement.

Climate Forecasting and Dashboard Development:
- Leveraged ARIMA and SARIMA time series models to forecast climate variables, enhancing the predictive accuracy through iterative evaluation and improvement.
- Designed and deployed dynamic, user-friendly dashboard incorporating interactive visualizations and real-time updates to support strategic decision-making for stakeholders across various sectors
- Orchestrated end-to-end machine learning pipelines for climate forecasting, incorporating data preprocessing, model selection, hyperparameter tuning, and deployment of predictive models, ensuring efficient forecasting solutions

Large-Scale Data Processing and Analysis Using Apache Spark:
- Spearheaded the development of a scalable data processing pipeline using Apache Spark to analyze and derive insights from large volumes of data in real-time.
- Utilized Apache Spark for scalable preprocessing, transformation, and analysis of large datasets. Utilized Spark SQL for querying and Spark MLlib for machine learning tasks on distributed data. Achieved efficient data processing with reduced processing times by 50%, enabling near real-time analytics and decision-making.