Thejus Kannoth

Chicago, IL | +91 9074837971 | kannoththejus@gmail.com | www.linkedin.com/in/thejuskannoth

EDUCATION

Depaul University, Chicago *Sep 2023- Nov 2025*

Master of Science in Data Science *GPA : 3.93/4.0*

Courses: Programming Machine Learning Applications, Mining Big Data, Advanced Machine Learning, Data Visualization, Fundamentals of Data Science, Data Analysis and Regression, Python Programming

APJ Abdul Kalam Technological University, MACE, Kothamangalam, India *Jun 2017 - May 2021*

B.Tech in Electronics and Communication *GPA : 8.2/10*

TECHNICAL SKILLS

Languages: Python, R, SQL

Machine Learning Algorithms: Decision Tree, KNN, SVM, Linear Regression, K-Mean, Random Forest, XGBoost, Logistic Regression

Tools & Frame Works: Tableau, Excel, MySQL, AWS, Oracle, Jupyter Notebook, PySpark, Hadoop, Visual Studio, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

PROFESSIONAL EXPERIENCE

Programmer Analyst | Cognizant Technology Solutions | *Bengaluru, India* *Nov 2021- Aug 2023*

- Coordinated with eight members for the development of CRM-based enterprise applications using Microsoft Dynamics 365, Power Apps, JavaScript.
- Contributing to the implementation of modifications, testing and maintenance of the sales module.
- Handled testing and debugging customisations and creating custom plugins, web resources, and JavaScript functions.

PROJECTS

Predicting Injury Risk in NBA Players *Jan 2025 - March 2025*

- Utilized machine learning models (**XGBoost, Random Forest, Logistic Regression**) to predict NBA player injury risk, achieving 97.97% accuracy and optimizing performance using **SMOTE** and **PCA**.
- Evaluated model performance using precision, recall, F1-score, and confusion matrices to ensure reliability in injury prediction.
- Developed a scalable data pipeline, applied hyperparameter tuning, and implemented effective techniques for class balancing and dimensionality reduction.

E-Commerce Fraud Detection *sep 2024- Nov 2024*

- Built a scalable data processing pipeline using PySpark on AWS EMR to clean, preprocess, and transform a 1.5M-row e-commerce transaction dataset stored in AWS S3.
- Performed advanced data cleaning, aggregation, and exploratory data analysis using PySpark and staged trans formed data in AWS Athena for efficient querying.
- Built a machine learning model using Spark MLlib for fraud detection, leveraging algorithms like Logistic Regression and Random Forest, achieving 95% accuracy.
- Automated pipeline execution and model inference workflows using AWS SageMaker, with end-to-end monitoring and reporting.

Insurance Cost Prediction *Apr 2024- Jun 2024*

- Developed a machine learning pipeline to predict medical insurance costs, achieving an accuracy of 88%.
- Optimized Random Forest, Gradient Boosting, and Stacking Regressor models using hyperparameter tuning and ensemble methods.
- Preprocessed data by encoding categorical variables, normalizing numerical features, and handling missing values.
- Conducted exploratory data analysis with visualizations to identify key factors influencing insurance costs.

CERTIFICATES

- Generative AI with Large Language Models, DeepLearning.AI & AWS
- ChatGPT Prompt Engineering for Developers, DeepLearning.AI & OpenAI

EXTRA CURRICULAR AND ACHIEVEMENTS

- Led the Planning and execution of student led events under the international student office, managing timelines, resources.
- Elected as a Senate Member during undergraduate studies and actively worked on initiatives to enhance student well-being and campus life.