P SAIRAM KUMAR

Data Scientist

psairam12p@gmail.com 9391674255 Yanam, Puducherry Github

PROFESSIONAL SUMMARY

Experienced educator with 5 years of teaching SQL, Python, C-Language, and Advanced Excel, now transitioning into a Data Scientist role. Recently completed a **Master's in Data Science** with hands-on experience in **statistical modeling**, **machine learning**, **data cleaning**, and **visualization**. Skilled in turning raw data into actionable insights through real-world projects using tools like **Pandas**, **Scikit-learn**, and **Power BI**. Eager to use data skills to solve business problems and drive decisions.

SKILLS

- **Tools**: MS SQL Server, Power BI, Excel (Advanced), MongoDB (Basic)
- **Languages**: SQL, Python, JavaScript, C, HTML, CSS
- **Python Libraries**: Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib, Plotly, NLTK
- **ML/NLP**: Machine Learning, NLP, Text Classification, Model Building
- **Development**: Jupyter, VS Code, GitHub, Streamlit
- **Data Concepts**: ETL/ELT, Data Warehousing, Star Schema
- **Soft Skills**: Communication, Problem-Solving, Analytical Thinking

PROJECTS

Project Name: Horse Race Outcome Prediction — GitHub: Link

Tools: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), Jupyter Notebook

- Built a **predictive classification model** to forecast winning horses based on historical race data.
- Cleaned and encoded complex race attributes like **track condition**, **horse rating**, and **jockey details**.
- Performed **EDA** to identify key performance indicators and trends in horse win rates.
- Trained **Logistic Regression**, **Decision Tree**, and **Random Forest** models.
- Achieved **accuracy of ~87%** with **Random Forest**, outperforming baseline models.
- Evaluated model using **confusion matrix**, **classification report**, and **ROC-AUC curve**.

Project Name: Singapore Resale Flat Price Prediction — GitHub: Link

Tools: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), Jupyter Notebook

- Built a **regression model** to predict **HDB resale flat prices** using historical housing data.
- Performed detailed **EDA** and **feature engineering** (flat age, town encoding, floor area).
- Applied and compared models: **Linear Regression**, **Decision Tree**, and **Random Forest Regressor**.
- Achieved **R^2 score of 0.89** with **Random Forest**, indicating strong model fit.
- Evaluated with **RMSE**, **residual plots**, and **train-test split** validation.
- Created insightful **visualizations** for predicted vs. actual prices and price trends.

EXPERIENCE

Computer Teaching & System Administrator — **11/2022 - 05/2025 (Full Time)**
Nallam Residential Concept High School

Computer Faculty — **01/2018 - 01/2022 (Full Time)**
Ct&t Computer Institution — 11/2022 - Present (Part Time)

EDUCATION

Professional Master Data Science Program GUVI July 2025

B.Sc Mathematics Dr. S.R.K Govt Arts College May 2014 - May 2017

CERTIFICATIONS

JavaScript (GUVI) **Chatgpt for Everyone (**GUVI)