

SAYANA M C

Data Scientist

Kannur, Kerala | +91 81298 79427 | sayanamanojmc2003@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Results-driven Data Scientist with an M.Sc. in Data Science and Analytics and a strong foundation in Mathematics. Experienced in developing machine learning and deep learning models, statistical analysis, predictive analytics, and data preprocessing using Python, R, SQL, TensorFlow, and Scikit-learn. Hands-on internship experience in building end-to-end ML solutions, feature engineering, model evaluation, and real-world AI applications across finance, agriculture, and IoT domains. Passionate about leveraging data-driven solutions to solve complex business problems and deliver impactful insights.

TECHNICAL SKILLS

Programming Languages: Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, OOP), SQL, R, SAS

Machine Learning: Regression, Classification, Clustering, Decision Trees, Random Forest, Logistic Regression, Linear Regression, K-Means, PCA, Feature Engineering, Hyperparameter Tuning, Cross Validation, Model Evaluation

Deep Learning & AI: CNN, Neural Networks, TensorFlow, Keras, PyTorch, Image Classification, Transfer Learning, OpenCV

Data Science: Data Cleaning, Data Wrangling, Data Preprocessing, Exploratory Data Analysis (EDA), Statistical Analysis, Predictive Analytics, Time Series Forecasting, Hypothesis Testing

Time Series: ARIMA, Holt-Winters, Forecasting, Stationarity Testing (ADF, HEGY)

Data Visualization: Power BI, Tableau, Matplotlib, Seaborn, Microsoft Excel, Dashboard Development

Databases: MySQL, MongoDB, SQL Querying

Tools & Platforms: Jupyter Notebook, Google Colab, Git, GitHub, VS Code, Anaconda

PROFESSIONAL EXPERIENCE

Data Science & Machine Learning Intern • Kites Software Pvt. Ltd., Kochi, Kerala Mar 2026 – Apr 2026

- Cleaned, preprocessed, and analyzed large-scale business datasets, enabling data-driven decision-making across multiple departmental workflows.
- Designed and deployed interactive Power BI dashboards and visual reports to communicate KPIs, business trends, and analytical insights to stakeholders.
- Applied supervised and unsupervised machine learning algorithms (regression, classification, clustering) to real-world problems, improving predictive model performance.
- Collaborated with cross-functional teams across a 2-month project sprint to deliver data science deliverables on time and to specification.

Data Analysis Intern • Cognifyz Technologies, Bangalore Jun 2025 – Jul 2025

- Performed end-to-end Exploratory Data Analysis (EDA) on structured datasets using Python (Pandas, NumPy, Matplotlib, Seaborn), uncovering critical patterns, outliers, and statistical trends.
- Built interactive Power BI dashboards to translate raw datasets into clear, actionable business insights, supporting informed strategic decisions.
- Developed and validated predictive models using Scikit-learn (regression, classification), applying cross-validation and evaluation metrics (accuracy, precision, recall, F1-score).
- Prepared structured analytical reports summarising data-driven findings and recommendations for cross-functional stakeholder review.

PROJECTS

Online Course Pricing Analysis & Prediction — Python | Power BI | Scikit-learn | Machine Learning

- Conducted market research and survey-based data collection to identify the key features influencing online course pricing across categories.
- Built interactive Power BI dashboards visualising pricing trends, category comparisons, and cost-driver correlations for stakeholder decision support.

- Performed feature engineering and correlation analysis to isolate the top pricing variables, reducing model noise and improving feature relevance.
- Trained ML regression models (Linear Regression, Random Forest) to predict optimal course pricing, delivering evidence-based pricing strategy recommendations.

Gold Price Inflation Forecasting — R | Time Series Analysis | ARIMA | Holt-Winters

- Collected, cleaned, and transformed monthly gold price data into a structured time-series format, followed by exploratory trend and seasonality analysis.
- Validated data stationarity using ADF and HEGY tests to determine the need for differencing before model construction.
- Built and benchmarked ARIMA and Holt-Winters forecasting models, selecting the best fit based on RMSE, MAE, and MAPE performance metrics.
- Engineered a hybrid forecasting model (Holt trend + ARIMA residuals) to generate reliable 12-month gold price forecasts with improved predictive accuracy.

IoT-Based Crop Disease Detection System — Python | TensorFlow | Keras | CNN | ESP32-CAM

- Implemented an end-to-end IoT data pipeline using ESP32-CAM to capture and stream real-time crop leaf images over Wi-Fi for automated disease analysis.
- Built and trained a Convolutional Neural Network (CNN) in Google Colab for multi-class classification of healthy and diseased crop leaf categories.
- Curated and labelled a custom image dataset, ensuring class balance and data quality to support robust model training and evaluation.
- Integrated IoT image acquisition with the deep learning inference pipeline, enabling real-time, automated crop health monitoring at field level.

Drought Impact Detection in Croplands — Python | TensorFlow | Keras | Remote Sensing | CNN

- Developed a CNN-based image classification model to detect drought-stressed croplands from Kaggle remote-sensing satellite datasets using TensorFlow/Keras.
- Applied comprehensive data augmentation (rotation, zoom, shift, flip) and preprocessing (resize, normalise) to enhance model robustness and minimise overfitting.
- Automated the end-to-end dataset pipeline—directory creation, image copying, and class-wise organisation—using Python (os, shutil) for reproducibility.
- Evaluated model accuracy and loss curves to validate performance, achieving reliable drought-stress classification for precision agriculture applications.

EDUCATION

M.Sc – Data Science and Analytics • Jain University, Bangalore, India	2024 – 2026
B.Sc (Hons.) – Mathematics • Kannur University, Kannur, India	2021 – 2024

CERTIFICATIONS

- Probability Foundations for Data Science and AI [↗](#)
- Data Visualization with R [↗](#)
- Principal Component Analysis with NumPy [↗](#)
- Foundations: Data, Data, Everywhere [↗](#)

LANGUAGES

English | Malayalam | Hindi | Tamil