

SETHU S

Data Scientist | Statistician | Data Analyst

sethus4791@gmail.com | [Streamlit-Portfolio](#) | [github.com/itsmesethus](#) | [linkedin.com/in/sethus4791](#)

Villupuram, Tamil Nadu-605103 | +91-8122815260 | [kaggle.com/sethu123123](#)

Summary

Dedicated Data Scientist and Analyst with a strong record of extracting actionable insights from complex data. Proficient in Statistical Analysis, Hypothesis Testing, and Data Modelling, utilizing Python, SQL and R. Demonstrated leadership through project delivery and a commitment to innovation. Pursuing further expertise in data-driven solutions to contribute to team success.

Experience

EVOSCIEN UK, Statistician – Bangalore, India

May 2023 – Present

- Enhanced decision accuracy by 25% through data acquisition, consolidation, and advanced statistical methods and analysis.
- Developed and implemented statistical models, algorithms, and techniques, yielding insights that led to 15% improvements in decision Making processes.
- Executed appropriate statistical tests (t-tests, ANOVA, chi-square tests, and non-parametric tests) to analyze data and uncovered meaningful patterns, trends, and relationships with a 95% confidence interval.
- Collaborated with the Entomology & Engineering departments using a data-driven approach, improving operational efficiency by 15%.
- Produced 20+ detailed reports and interactive data visualizations, clearly communicating analytical findings to both technical and non-technical audiences.

Education

• **8.61/10 M.Sc in Statistics**, Bharathiar University | Coimbatore, Tamil Nadu

Sep 2021 – May 2023

• **9.39/10 B.Sc in Statistics**, Arignar Anna Govt Arts College | Villupuram, Tamil Nadu

Jun 2018 – May 2021

Courses: Descriptive Statistics | Sampling Theory | Probability Theory | Statistical Estimation Theory | Statistical Quality Control | Multivariate Statistical Analysis | Econometrics | Statistical Inference (Hypothesis Testing) | Programming in R | Design of Experiments | Stochastic Processes.

Skills

Programming: Python, R, SQL.

Tools: Microsoft Power BI, IBM SPSS, MINITAB, STATISTICA, Microsoft Excel, MYSQL, Jupyter Notebook.

Libraries: Pandas, Numpy, Matplotlib, Seaborn, Plotly, Scipy, Scikit-Learn, TensorFlow, Keras, Statsmodels, Pingouin.

Data Preprocessing & Others: Data Cleaning, Data Analytics, Data Analysis, Exploratory Data Analysis (EDA), Data Mining, Feature Engineering, Feature Selection Techniques, Data Visualization, Outlier Detection, Correlation Analysis, A/B Testing, Statistical Modelling.

ML Expertise: Regression, Classification, Clustering, Predictive Modelling, Quantitative Analysis, Statistical Modelling, Deep Learning, CNN, RNN, LSTM, GRU, Bidirectional Models, Time Series Analysis, Forecasting.

Projects

Chrono-Power Forecast: Electricity Demand Forecasting for UK Year-2024 - [link](#)

Dec 2023 - Jan 2024

- Implemented diverse forecasting models including Exponential Smoothing, ARIMA, SARIMA, Fb-Prophet models in Phase 1.
- Investigated deep learning architectures like RNN, LSTM, GRU, and hybrid Convolutional-RNN models for 3 output forecasting for Phase 2.
- Enhanced the precision of deep learning models, achieving 20% better results for ND, TSD, EWD.

Cognizant AI Virtual Internship - Machine Learning for Gala Grocery Retail Pricing - [link](#)

Sep 2023 - Oct 2023

- Performed in-depth EDA, establishing a strong foundation for modelling and optimization across 10+ product categories.
- Applied Ensemble Regressors (Adaboost, Bagging) and ANN-based Multiple Linear Regression, achieving over 94% R^2 for predictions.
- Validated forecast reliability using 3+ evaluation metrics, refining ML models with RandomizedSearchCV.

BCG Virtual Internship - Churn Predictive Modelling to Customer Retention - [link](#)

Aug 2023 - Sep 2023

- Employed various predictive models including Ensemble classifiers and KNeighbors Classifiers, achieving an F1 score exceeding 95%.
- Integrated advanced techniques such as SMOTE and Yeo-Johnson transformation, leading to a 20% reduction in prediction errors.
- Interpreted ML insights for customer retention and crafted an impactful Power BI dashboard report with 3 sections for visualization.

Multiclass Disease Classification of Medical Image Data Using CNN - [link](#)

Jan 2023 - Apr 2023

- Assembled CNNs using TensorFlow and Keras, achieving 94.88% accuracy in Pneumonia X-ray classification.
- Exhibited expertise in multi-class classification with accuracies of 95.67%, 95.18%, and 90.87% for various diseases.
- Adopted VGG19 architecture with strategic layers for superior 8 class classification to Gastrointestinal diseases performance.

Certifications

- IBM Data Science Professional Certificate – (IBM-Coursera, Jul 2023).
- Machine Learning Specialization (Supervised Machine Learning- Regression & Classification, Advanced Learning Algorithms, Unsupervised Machine Learning & Recommender Systems) – (DeepLearning.AI - Coursera, Oct 2023).
- Ensemble Methods In Python – (DataCamp, Sep 2023)
- Sequences, Time Series And Prediction – (DeepLearning.AI- Coursera, Nov 2023).
- Feature Engineering For Machine Learning In Python – (DataCamp, Sep 2023)
- Data Analytics With Python – (NPTEL, Swayam, Apr 2022).
- Databases And SQL For Data Science With Python - (IBM - Coursera, Nov 2022)