

# Sri Harsha Peetha

Data Scientist | Data Analyst | Business Analyst

Hatfield, Hertfordshire • +44 0776 064 7729 • sriharsha2698@gmail.com • linkedin.com/in/sri-harsha-peetha • Right to Work in the UK • [sriharsha2698.github.io/Portfolio/](https://sriharsha2698.github.io/Portfolio/)

## PROFESSIONAL SUMMARY

MSc Data Science and Analytics graduate (Distinction) with experience in machine learning, statistical modelling, and business intelligence. Proficient in Python, SQL, Power BI, R, and Excel across the full data lifecycle. Demonstrated ability to build, evaluate, and deploy ML models end-to-end on Azure. Experienced in hypothesis testing, feature engineering, and KPI reporting. Adept at communicating findings to both technical and non-technical audiences.

## CORE COMPETENCIES

- Data Analysis & Insights
- Machine Learning & Modelling
- Business Intelligence
- Statistical Analysis
- XGBoost / Random Forest / LR
- Power BI & DAX
- Feature Engineering
- Model Evaluation & Tuning
- KPI Reporting & Dashboards
- SQL & Database Querying
- Azure App Service
- Stakeholder Communication
- Hypothesis Testing
- Python (Pandas, Scikit-learn)
- Data-Driven Decision Making

## TECHNICAL SKILLS

**Languages:** Python, SQL, R, Java

**Python Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, XGBoost, LightGBM

**BI & Reporting:** Power BI, DAX, Power Query, Excel (PivotTables, VLOOKUP, Power Pivot)

**ML & Statistics:** Regression, Classification, Clustering, Hypothesis Testing, A/B Testing, TF-IDF, ETL, Data Pipeline

**Cloud & MLOps:** Azure App Service, Azure ML, MLflow, Flask API deployment, Gunicorn

**Databases:** MySQL, SQL Server

**Tools & Other:** RStudio, WEKA, TensorFlow, PyTorch, Git, Jupyter Notebook, Google Colab, Tableau (familiar), Agile (Scrum/Kanban)

## EDUCATION

**MSc Data Science and Analytics — University of Hertfordshire, UK** Distinction | 2023–2025

*Modules: Machine Learning, Data Mining, Applied Data Science, Statistical Methods, Team Research & Development Project*

**BTech Mechanical Engineering — Dhanekula Institute of Engineering & Technology, India** 2018–2022

## PROJECTS & EXPERIENCE

**Customer Churn Prediction — End-to-End ML Deployment on Azure** | Personal / Portfolio Project

*Python · Logistic Regression · Random Forest · XGBoost · Scikit-learn · MLflow · Flask · Azure App Service · Gunicorn · Git*

- Built an end-to-end ML pipeline on the Kaggle Telco Churn dataset; trained and benchmarked three classifiers — Logistic Regression (Acc 76%, Recall 81%, F1 0.77), Random Forest (Acc 75%, Recall 82%, F1 0.77), and XGBoost (Acc 72%, Recall 76%, F1 0.73) — selecting Logistic Regression for deployment based on best overall accuracy and F1.
- Tracked all experiments, parameters, and evaluation metrics using MLflow for reproducible model comparison and selection.
- Exposed predictions via a RESTful Flask API and deployed to Azure App Service (Linux, Python 3.10) with Gunicorn; resolved Kudu environment dependency issues via a custom startup command.
- Managed end-to-end project versioning with Git, demonstrating full MLOps awareness from data processing through to live cloud-hosted API serving.

**Flood Prediction Using Ensemble Machine Learning Models** | MSc Dissertation

*Python · XGBoost · LightGBM · Random Forest · Scikit-learn · Pandas · Matplotlib · Google Colab · Agile*

- **Key result:** XGBoost achieved  $R^2$  95.51% and RMSE 0.0106 after hyperparameter tuning, outperforming LightGBM ( $R^2$  94.98%) and Random Forest ( $R^2$  78.51%), demonstrating gradient boosting superiority for non-linear prediction tasks.
- Performed feature importance analysis identifying top predictors (Environmental Pressure, Preparedness Deficiency, Population Environmental Stress), delivering actionable insights for disaster management stakeholders.
- Produced detailed visualisations and a written report translating statistical outputs into clear non-technical recommendations; managed milestones with Trello using Agile-style task tracking.

### Heart Disease Risk Analysis — Proportion Hypothesis Testing | Group Academic Project

*R · RStudio · Statistical Analysis · Hypothesis Testing · Data Cleaning*

- Investigated the statistical association between EKG results and heart disease prevalence using chi-square proportion hypothesis testing in RStudio on the Heart Disease Prediction dataset (n=303).
- Cleaned and prepared structured clinical data and presented findings with clear statistical interpretation for mixed technical and non-technical audiences.

### Text Sentiment Classification — Dimensionality Reduction Study | Academic Project

*WEKA · J48 Decision Tree · LibSVM · TF-IDF · Random Projection · Data Mining*

- Preprocessed an 800-instance sentiment dataset using TF-IDF and Lovins stemming (1,199 features); applied Random Projection to 900 features and evaluated the accuracy/efficiency trade-off across J48 and LibSVM classifiers.
- Random Projection improved LibSVM accuracy to 76.1% while J48 fell from 67% to 59.6% — demonstrating that dimensionality reduction impact is algorithm-dependent.

### Superstore Sales Interactive BI Dashboard | Personal Project

*Power BI · DAX · Power Query · Data Modelling · KPI Reporting*

- Built an interactive Power BI dashboard on the Kaggle Superstore Sales dataset with DAX measures, drill-through filters, and dynamic slicers for self-service exploration by region, product category, and time period — replicating a production BI reporting environment.

## CERTIFICATIONS

---

Software Testing Certification — QSpiders, Bangalore | Manual Testing · Selenium · Java · Python · MySQL

Google Analytics Certificate — Google | In Progress

Microsoft Power BI Data Analyst (PL-300) — Microsoft | In Progress

## ADDITIONAL INFORMATION

---

Right to work in the United Kingdom • Open to hybrid and on-site roles • Strong analytical, problem-solving, and stakeholder communication skills • Comfortable working independently and in cross-functional teams