

Data Science Services

Data Science as a Service

Shreeminfotech Limited's Data Science Services help UK public sector organisations unlock the value in their data through advanced analytics, machine learning, and statistical modelling. We transform raw data into actionable insights that improve decision-making, optimise operations, and deliver better outcomes for citizens.

Data Science combines statistics, mathematics, programming, and domain expertise to extract knowledge and insights from structured and unstructured data. Our services span the entire data science lifecycle—from problem framing and data exploration through model development, deployment, and monitoring. We align with the **Government Data Quality Framework**, **Data Ethics Framework**, and **Algorithmic Transparency Standard** to ensure responsible, ethical, and explainable analytics.

G-Cloud 15 Framework

Service Category: Cloud Software

Supplier: Shreeminfotech Limited

Machine Learning | Predictive Analytics | NLP | Computer Vision | MLOps | Statistical Modelling

Service Overview

Shreeminfotech Limited's Data Science Services provide end-to-end data science capabilities—from initial problem definition and data exploration through advanced modelling, deployment, and ongoing optimisation. We help organisations make evidence-based decisions and automate complex analytical processes.

Our Data Science services cover fifteen core domains:

- **Data Science Strategy & Assessment:** Maturity assessment, use case identification, roadmap development
- **Exploratory Data Analysis (EDA):** Data profiling, pattern discovery, hypothesis generation
- **Statistical Analysis & Modelling:** Hypothesis testing, regression, Bayesian inference
- **Machine Learning:** Supervised, unsupervised, and reinforcement learning models
- **Predictive Analytics:** Forecasting, propensity modelling, risk scoring
- **Natural Language Processing (NLP):** Text analytics, sentiment analysis, entity extraction
- **Computer Vision:** Image classification, object detection, document processing
- **Time Series Analysis:** Forecasting, anomaly detection, seasonal decomposition
- **Recommendation Systems:** Personalisation, content recommendation, next-best-action
- **A/B Testing & Experimentation:** Experiment design, statistical significance, causal inference
- **Data Visualisation:** Dashboards, interactive reports, storytelling with data
- **Feature Engineering:** Feature creation, selection, and optimisation
- **MLOps & Model Deployment:** Production ML pipelines, monitoring, and governance
- **Data Engineering for Data Science:** Data pipelines, feature stores, data quality
- **Data Science Training:** Capability building, mentoring, and knowledge transfer

We work with modern data science platforms including **Python, R, Azure Machine Learning, AWS SageMaker, Databricks, and Google Vertex AI.**

Why Data Science Matters

Public sector organisations hold vast amounts of data but often struggle to extract actionable insights. Data science transforms this data into evidence for better policy, improved services, and more efficient operations—ultimately delivering better outcomes for citizens.

Public Sector Data Science Opportunities

- Predict demand for services to optimise resource allocation
- Identify fraud, error, and overpayments in benefits and grants
- Personalise citizen services based on needs and preferences
- Forecast trends to inform policy development
- Automate document processing and classification
- Detect anomalies in health, safety, or compliance data
- Optimise scheduling, routing, and logistics
- Improve recruitment and workforce planning
- Analyse citizen feedback and sentiment at scale
- Support evidence-based decision making

The Data Science Advantage

- **Evidence-Based Decisions:** Replace intuition with data-driven insights
- **Predictive Power:** Anticipate future trends and outcomes
- **Automation:** Scale analytical processes beyond human capacity
- **Personalisation:** Tailor services to individual needs
- **Efficiency:** Optimise operations and reduce waste
- **Risk Reduction:** Identify and mitigate risks earlier
- **Innovation:** Discover new insights and opportunities
- **Accountability:** Transparent, reproducible analysis

Government Data Standards

- Government Data Quality Framework
- Data Ethics Framework
- Algorithmic Transparency Standard
- CDDO Data Standards
- Open Standards for Government
- UK Statistics Authority Code of Practice

Data Science Lifecycle

Our data science methodology follows a structured lifecycle that ensures rigorous, reproducible, and business-aligned analytics. We use CRISP-DM (Cross-Industry Standard Process for Data Mining) adapted for modern machine learning workflows.

1. Business Understanding

- Problem framing and success criteria definition
- Stakeholder requirements gathering
- Use case prioritisation and scoping
- ROI and feasibility assessment
- Ethical and bias risk assessment

2. Data Understanding

- Data source identification and access
- Data profiling and quality assessment
- Exploratory data analysis (EDA)
- Data dictionary and documentation
- Initial hypothesis generation

3. Data Preparation

- Data cleaning and validation
- Missing value treatment
- Outlier detection and handling
- Feature engineering and transformation
- Data integration and enrichment
- Train/validation/test split design

4. Modelling

- Algorithm selection and benchmarking
- Model training and hyperparameter tuning
- Cross-validation and performance evaluation
- Ensemble methods and model stacking
- Interpretability and explainability analysis

5. Evaluation

- Model performance assessment against success criteria
- Business impact validation
- Fairness and bias testing
- Robustness and edge case testing
- Stakeholder review and sign-off

6. Deployment & Monitoring

- Production deployment and integration
- Model monitoring and alerting
- Performance tracking and drift detection
- Retraining pipelines and model updates
- Documentation and knowledge transfer

Exploratory Data Analysis (EDA)

Exploratory Data Analysis is the foundation of any data science project. We help organisations understand their data, identify patterns, and generate hypotheses before building models.

EDA Capabilities

- Data profiling and summary statistics
- Distribution analysis and visualisation
- Correlation and relationship analysis
- Missing data patterns and impact assessment
- Outlier and anomaly identification
- Temporal and seasonal pattern detection
- Segmentation and clustering exploration
- Feature importance and relevance analysis

Data Quality Assessment

- Completeness, accuracy, and consistency checks
- Data lineage and provenance tracking
- Schema validation and data type verification
- Referential integrity and constraint checking
- Timeliness and freshness assessment
- Recommendations for data quality improvement

EDA Deliverables

- Data quality report with findings and recommendations
- Interactive EDA notebooks (Jupyter/Databricks)
- Visualisation dashboards for data exploration
- Data dictionary and metadata documentation
- Hypothesis document for modelling phase
- Feature engineering recommendations

Statistical Analysis & Modelling

Statistical analysis provides the rigorous foundation for evidence-based decision making. We apply classical and modern statistical methods to answer questions, test hypotheses, and quantify uncertainty.

Descriptive Statistics

- Summary statistics and measures of central tendency
- Variability and dispersion analysis
- Frequency distributions and cross-tabulations
- Cohort analysis and segmentation
- Geospatial analysis and mapping

Inferential Statistics

- Hypothesis testing (t-tests, chi-square, ANOVA)
- Confidence intervals and statistical significance
- Power analysis and sample size determination
- Multiple comparison corrections
- Non-parametric methods

Regression Analysis

- Linear and multiple regression
- Logistic regression for classification
- Generalised linear models (GLM)
- Mixed effects and hierarchical models
- Regularisation (Ridge, Lasso, Elastic Net)
- Quantile regression

Bayesian Analysis

- Bayesian inference and posterior estimation
- Prior specification and sensitivity analysis
- Markov Chain Monte Carlo (MCMC) methods
- Bayesian model comparison
- Probabilistic programming (PyMC, Stan)

Causal Inference

- Propensity score matching
- Difference-in-differences analysis
- Regression discontinuity design
- Instrumental variables
- Synthetic control methods

- Causal discovery algorithms

Machine Learning

Machine learning enables systems to learn from data and make predictions or decisions without being explicitly programmed. We build ML models that solve real business problems with measurable impact.

Supervised Learning

- **Classification:** Logistic regression, decision trees, random forests, gradient boosting (XGBoost, LightGBM), support vector machines, neural networks
- **Regression:** Linear regression, tree-based methods, neural networks, ensemble methods
- **Applications:** Fraud detection, churn prediction, risk scoring, document classification, demand forecasting

Unsupervised Learning

- **Clustering:** K-means, hierarchical clustering, DBSCAN, Gaussian mixture models
- **Dimensionality Reduction:** PCA, t-SNE, UMAP, autoencoders
- **Anomaly Detection:** Isolation forest, one-class SVM, autoencoders
- **Applications:** Customer segmentation, anomaly detection, pattern discovery, data compression

Ensemble Methods

- Random forests and extra trees
- Gradient boosting (XGBoost, LightGBM, CatBoost)
- Model stacking and blending
- Voting classifiers and regressors
- Hyperparameter optimisation (Optuna, Hyperopt)

Deep Learning

- Feedforward neural networks (MLPs)
- Convolutional neural networks (CNNs) for images
- Recurrent neural networks (RNNs, LSTMs) for sequences
- Transformer architectures
- Transfer learning and fine-tuning
- Frameworks: TensorFlow, PyTorch, Keras

AutoML

- Automated feature engineering
- Algorithm selection and hyperparameter tuning
- Neural architecture search
- Tools: Azure AutoML, AWS Autopilot, H2O, Auto-sklearn

Predictive Analytics

Predictive analytics uses historical data to forecast future outcomes. We build models that help organisations anticipate demand, identify risks, and make proactive decisions.

Forecasting

- Demand forecasting for resource planning
- Revenue and budget forecasting
- Workload and capacity planning
- Service demand prediction
- Population and demographic projections

Propensity Modelling

- Propensity to respond/engage
- Propensity to churn or disengage
- Propensity to commit fraud or error
- Propensity to benefit from intervention
- Next-best-action recommendation

Risk Scoring

- Credit and financial risk scoring
- Fraud risk assessment
- Compliance risk identification
- Health and safety risk prediction
- Safeguarding risk indicators

Public Sector Predictive Use Cases

- Predicting hospital admissions and A&E; demand
- Forecasting benefits and grant applications
- Identifying children at risk of poor outcomes
- Predicting infrastructure maintenance needs
- Forecasting crime and anti-social behaviour hotspots
- Predicting tax compliance risk
- Identifying fraud and error in payments

Natural Language Processing (NLP)

Natural Language Processing enables machines to understand, interpret, and generate human language. We apply NLP to extract insights from unstructured text data—documents, emails, surveys, social media, and more.

Text Analytics

- Text classification and categorisation
- Sentiment analysis and opinion mining
- Topic modelling and theme extraction
- Keyword and keyphrase extraction
- Text summarisation (extractive and abstractive)
- Language detection and translation

Information Extraction

- Named Entity Recognition (NER) - people, organisations, locations, dates
- Relation extraction between entities
- Event extraction and timeline construction
- Custom entity extraction for domain-specific concepts
- Knowledge graph construction

Document Intelligence

- Document classification and routing
- Form and invoice data extraction
- Contract analysis and clause extraction
- Redaction and PII detection
- Document similarity and duplicate detection

NLP Technologies

- **Traditional NLP:** spaCy, NLTK, Gensim
- **Transformers:** BERT, RoBERTa, DistilBERT, DeBERTa
- **Large Language Models:** GPT-4, Claude, Llama, Mistral
- **Cloud Services:** Azure Cognitive Services, AWS Comprehend, GCP Natural Language
- **Embeddings:** Word2Vec, FastText, Sentence Transformers

Public Sector NLP Applications

- Analysing citizen correspondence and complaints
- Processing Freedom of Information requests
- Summarising consultation responses
- Monitoring social media for public sentiment
- Extracting information from legacy documents

- Automating casework triage and routing

Computer Vision

Computer vision enables machines to interpret and analyse visual information from images and video. We build computer vision solutions for document processing, asset inspection, and more.

Image Analysis

- Image classification and tagging
- Object detection and localisation
- Image segmentation (semantic, instance)
- Optical Character Recognition (OCR)
- Scene understanding and context analysis

Document Processing

- Intelligent document processing (IDP)
- Form recognition and data extraction
- Handwriting recognition
- Document layout analysis
- Table extraction from documents
- Invoice and receipt processing

Video Analytics

- Object tracking and counting
- Activity recognition
- Anomaly detection in video
- Video summarisation

Computer Vision Technologies

- **Frameworks:** OpenCV, PyTorch, TensorFlow
- **Pre-trained Models:** ResNet, EfficientNet, YOLO, Detectron2
- **Cloud Services:** Azure Computer Vision, AWS Rekognition, GCP Vision AI
- **Document AI:** Azure Form Recognizer, AWS Textract, GCP Document AI

Time Series Analysis & Forecasting

Time series analysis examines data points collected over time to identify trends, seasonality, and patterns. We build forecasting models that predict future values with quantified uncertainty.

Time Series Methods

- **Classical Methods:** ARIMA, SARIMA, Exponential Smoothing (ETS), Holt-Winters
- **Machine Learning:** XGBoost, LightGBM with time features, Random Forest
- **Deep Learning:** LSTM, GRU, Temporal Convolutional Networks, Transformers
- **Probabilistic:** Prophet, NeuralProphet, Bayesian structural time series
- **Hierarchical:** Hierarchical forecasting, reconciliation methods

Time Series Capabilities

- Trend and seasonality decomposition
- Anomaly and changepoint detection
- Multi-step ahead forecasting
- Forecast uncertainty quantification
- Multiple time series modelling
- Causal impact analysis

Forecasting Applications

- Service demand forecasting
- Resource and capacity planning
- Budget and revenue forecasting
- Workload prediction
- Inventory and supply planning
- Energy and utilities forecasting

A/B Testing & Experimentation

Experimentation provides causal evidence of what works. We help organisations design, execute, and analyse experiments to make data-driven decisions about policies and services.

Experiment Design

- Hypothesis formulation and success metrics
- Sample size and power calculations
- Randomisation strategies
- Control group design
- Multi-arm and factorial experiments
- Stratification and blocking

Statistical Analysis

- Statistical significance testing
- Confidence intervals and effect sizes
- Multiple testing corrections
- Bayesian A/B testing
- Sequential testing and early stopping
- Heterogeneous treatment effects

Advanced Experimentation

- Multi-armed bandits for adaptive allocation
- Quasi-experimental methods (when randomisation isn't possible)
- Synthetic control methods
- Causal inference from observational data
- Long-term impact measurement

Data Visualisation & Dashboards

Effective visualisation transforms complex data into clear, actionable insights. We create interactive dashboards and reports that enable stakeholders to explore and understand data.

Visualisation Capabilities

- Interactive dashboards and reports
- Executive KPI dashboards
- Self-service analytics interfaces
- Geospatial mapping and visualisation
- Network and relationship visualisation
- Real-time streaming dashboards
- Embedded analytics in applications

Storytelling with Data

- Narrative-driven analysis reports
- Insight annotation and commentary
- Automated insight generation
- Presentation-ready visualisations
- Data journalism and public reporting

Visualisation Tools

- **BI Platforms:** Power BI, Tableau, Looker, Qlik
- **Python:** Matplotlib, Seaborn, Plotly, Altair, Bokeh
- **R:** ggplot2, Shiny
- **JavaScript:** D3.js, Observable
- **Geospatial:** Kepler.gl, Folium, Mapbox

MLOps & Model Deployment

MLOps brings DevOps practices to machine learning, enabling reliable, scalable, and governed deployment of ML models in production. We help organisations operationalise data science.

ML Pipeline Automation

- Automated data ingestion and validation
- Feature engineering pipelines
- Model training automation
- Hyperparameter optimisation
- Model versioning and registry
- Automated model evaluation and comparison

Model Deployment

- Real-time inference endpoints (REST APIs)
- Batch prediction pipelines
- Edge deployment for IoT scenarios
- A/B testing and canary deployments
- Blue-green deployments for ML
- Model serving infrastructure

Model Monitoring & Governance

- Model performance monitoring
- Data drift and concept drift detection
- Feature drift monitoring
- Prediction quality tracking
- Model fairness and bias monitoring
- Automated retraining triggers
- Model lineage and audit trails

MLOps Platforms

- **Cloud:** Azure Machine Learning, AWS SageMaker, GCP Vertex AI
- **Open Source:** MLflow, Kubeflow, Airflow, DVC
- **Feature Stores:** Feast, Tecton, Databricks Feature Store
- **Experiment Tracking:** MLflow, Weights & Biases, Neptune

Responsible AI & Ethics

Responsible AI ensures that data science solutions are fair, transparent, and accountable. We embed ethical considerations throughout the data science lifecycle, aligned with government standards.

Fairness & Bias

- Bias detection in training data
- Algorithmic fairness assessment
- Disparate impact analysis across protected characteristics
- Bias mitigation techniques
- Fairness metrics and monitoring

Explainability & Transparency

- Model interpretability techniques (SHAP, LIME)
- Feature importance and contribution analysis
- Decision explanations for stakeholders
- Algorithmic Transparency Standard compliance
- Model cards and documentation

Data Ethics Framework Alignment

- Public benefit assessment
- Privacy and data protection
- Transparency and explainability
- Accountability and governance
- Proportionality and necessity
- Human oversight and intervention

Features

- End-to-end data science lifecycle support
- Exploratory data analysis and data quality assessment
- Statistical analysis and hypothesis testing
- Machine learning model development and optimisation
- Predictive analytics and forecasting
- Natural Language Processing and text analytics
- Computer vision and document processing
- Time series analysis and forecasting
- A/B testing and experimentation design
- Data visualisation and dashboard development
- MLOps and production model deployment
- Model monitoring and drift detection
- Responsible AI and bias assessment
- Data science training and capability building
- SC/DV cleared data scientists available

Benefits

- Evidence-based decision making across the organisation
- Predictive insights to anticipate future needs
- Automated analysis at scale beyond human capacity
- Personalised services tailored to citizen needs
- Operational efficiency through optimisation
- Early identification of risks and anomalies
- Faster time to insight with automated pipelines
- Reproducible, auditable analytical processes
- Ethical AI aligned with government standards
- Knowledge transfer and internal capability building
- Reduced cost through automation and efficiency
- Better outcomes for citizens through data-driven services
- Compliance with Data Ethics Framework
- Transparent, explainable models for accountability

Team

Shreeminfotech Limited's Data Science team combines statistical rigour with software engineering best practices to deliver production-ready analytical solutions.

- PhD and Masters-level data scientists
- Statisticians with government experience
- Machine learning engineers
- MLOps and data engineering specialists
- NLP and computer vision experts
- Domain specialists in public sector
- Certified in Azure, AWS, GCP data platforms
- Experience with government data ethics
- SC/DV cleared data scientists

Roles

- Data Science Lead
- Senior Data Scientist
- Data Scientist
- Machine Learning Engineer
- MLOps Engineer
- Statistician
- NLP Specialist
- Computer Vision Engineer
- Data Engineer
- BI Developer / Data Analyst
- Data Visualisation Specialist
- Experimentation Analyst
- AI Ethics Specialist
- Data Science Trainer

Public Sector Experience

Shreeminfotech Limited has extensive experience delivering data science projects across UK government and public sector organisations.

- Central Government - Fraud and error detection models
- HMRC - Tax compliance risk scoring
- DWP - Benefits demand forecasting
- Home Office - Document processing and NLP
- NHS - Patient demand and capacity modelling
- MOJ - Offender risk assessment
- DEFRA - Environmental monitoring and prediction
- Cabinet Office - Cross-government analytics
- Local Government - Service demand prediction
- Police Forces - Crime pattern analysis
- Transport Authorities - Demand forecasting
- Education Sector - Student outcome prediction

Why Choose Shreeminfotech Limited?

- **Statistical Rigour:** Proper statistical methods, not just ML hype
- **Production Focus:** Models that work in the real world, not just notebooks
- **Domain Expertise:** Deep understanding of public sector challenges
- **Ethical AI:** Responsible AI aligned with government frameworks
- **End-to-End:** From problem framing through deployment and monitoring
- **Knowledge Transfer:** Building your internal capability
- **Platform Agnostic:** Azure, AWS, GCP, and open source expertise
- **Government Experience:** Track record in UK public sector
- **Security Cleared:** SC and DV cleared data scientists
- **Explainable:** Models you can understand and trust

Engagement Models

Data Science Discovery

Rapid assessment of data science opportunities—use case identification, data assessment, feasibility analysis, and prioritised roadmap.

Data Science Project

End-to-end delivery of a data science solution—from problem framing through EDA, modelling, evaluation, and deployment.

Embedded Data Science Team

Data scientists embedded in your organisation to work alongside your teams, deliver projects, and build internal capability.

MLOps Implementation

Design and implement ML operations—automated pipelines, model deployment, monitoring, and governance.

Data Science Training

Capability building programmes—from data literacy for business users to advanced ML for technical teams.

Analytics Centre of Excellence

Design and establish an internal data science function—operating model, tools, processes, and governance.

Quality & Compliance

Certifications

- ISO 9001:2015 Quality Management
- ISO 27001:2017 Information Security Management
- Cyber Essentials Plus
- Crown Commercial Service Supplier

Standards & Frameworks

- Government Data Quality Framework
- Data Ethics Framework
- Algorithmic Transparency Standard
- CDDO Data Standards
- UK Statistics Authority Code of Practice
- GDPR and Data Protection Act 2018
- ICO AI Auditing Framework

Social Value

Shreeminfotech Limited is committed to delivering social value through ethical, impactful data science.

Theme 1: Better Public Services

Data science enables more effective, efficient, and personalised public services. Predictive analytics and optimisation improve outcomes for citizens.

Theme 2: Tackling Economic Inequality

Skills Development (MAC2): We develop data science skills through training, mentoring, and apprenticeships, creating pathways into data careers. **Diverse Supply Chains (MAC3):** We partner with UK data science and analytics SMEs.

Theme 3: Fighting Climate Change (MAC4)

Data science supports environmental monitoring, resource optimisation, and sustainability initiatives. We help organisations use data to reduce their environmental impact.

Theme 4: Equal Opportunity (MAC5)

We ensure AI and analytics are fair and do not discriminate. We promote diversity in data science through inclusive hiring and training practices.

Service Summary

Data Science Domain	Key Capabilities
EDA & Data Quality	Data profiling, Quality assessment, Pattern discovery
Statistical Analysis	Hypothesis testing, Regression, Bayesian, Causal inference
Machine Learning	Classification, Regression, Clustering, Deep learning
Predictive Analytics	Forecasting, Propensity, Risk scoring
NLP	Text classification, Sentiment, NER, Document processing
Computer Vision	Image classification, Object detection, OCR
Time Series	Forecasting, Anomaly detection, Trend analysis
Experimentation	A/B testing, Experiment design, Causal analysis
Visualisation	Dashboards, Reports, Storytelling
MLOps	Pipelines, Deployment, Monitoring, Governance
Responsible AI	Fairness, Explainability, Ethics compliance

Technology Stack

Category	Technologies
Languages	Python, R, SQL, Scala
ML Frameworks	Scikit-learn, XGBoost, LightGBM, TensorFlow, PyTorch
NLP	spaCy, Hugging Face Transformers, NLTK, Gensim
Computer Vision	OpenCV, TensorFlow, PyTorch, YOLO
Time Series	Prophet, statsmodels, pmdarima, Darts
Data Processing	Pandas, NumPy, Spark, Dask, Polars
Visualisation	Power BI, Tableau, Matplotlib, Plotly, Seaborn
Cloud ML	Azure ML, SageMaker, Vertex AI, Databricks
MLOps	MLflow, Kubeflow, Airflow, DVC
Notebooks	Jupyter, Databricks, Azure ML Studio

Shreeminfotech Limited is platform and tool agnostic. We recommend the best technologies for your specific context, existing investments, and requirements.

Contact Information

Shreeminfotech Limited

For enquiries about Data Science Services, please contact us through the G-Cloud Digital Marketplace.

This service definition was prepared for G-Cloud 15 submission.

Document version: 1.0 | Date: January 2026