# INFOSOFT IT SOLUTIONS

Training | Projects | Placements

Revathi Apartments, Ameerpet, 1st Floor, Opposite Annapurna Block,

Infosoft It solutions, Software Training & Development Institute, 9059683947|9182540872

## ***Data Science***

### Introduction to Data Science

- Overview of Data Science: Definition, importance, and applications
- Data Science Life Cycle: From data collection to deployment
- Roles and Responsibilities in Data Science Teams

### Programming Fundamentals for Data Science

- Introduction to Python and R: Basics of programming languages
- Data Structures and Algorithms: Arrays, lists, dictionaries, and algorithms for data manipulation
- Libraries and Packages: NumPy, Pandas, Matplotlib (Python), ggplot2 (R)

### Data Wrangling and Preprocessing

- Data Cleaning Techniques: Handling missing data, outliers, and duplicates
- Data Transformation: Normalization, scaling, and encoding categorical variables
- Feature Engineering: Creating new features from existing data

### Exploratory Data Analysis (EDA)

- Statistical Analysis: Descriptive statistics, inferential statistics
- Data Visualization: Plotting with Matplotlib, Seaborn, ggplot2
- EDA Techniques: Univariate, bivariate, and multivariate analysis

## Machine Learning Fundamentals

- Introduction to Machine Learning: Supervised vs. unsupervised learning
- Model Selection and Evaluation: Cross-validation, bias-variance tradeoff
- Performance Metrics: Accuracy, precision, recall, F1-score, ROC curve

## Supervised Learning Algorithms

- Linear Regression and Logistic Regression
- Decision Trees and Random Forests
- Support Vector Machines (SVM)
- Naive Bayes Classifier

## Unsupervised Learning Algorithms

- K-means Clustering
- Hierarchical Clustering
- Principal Component Analysis (PCA)
- Association Rule Mining: Apriori algorithm

## Advanced Machine Learning Techniques

- Ensemble Learning: Bagging, boosting (AdaBoost, Gradient Boosting), stacking
- Neural Networks and Deep Learning: Introduction to artificial neural networks (ANN)
- Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN)
- Natural Language Processing (NLP) and Text Mining

## Model Deployment and Optimization

- Model Deployment Strategies: Flask (Python), Shiny (R)
- Model Optimization Techniques: Hyperparameter tuning, feature selection
- Monitoring and Maintaining Models: Model performance tracking and retraining

## Big Data and Data Science

- Introduction to Big Data: Characteristics, challenges, and opportunities
- Distributed Computing: Hadoop, Spark
- Scalable Machine Learning: Machine learning on big data platforms

## Data Science Tools and Platforms

- Data Science Tools Overview: Jupyter Notebook, RStudio, Anaconda
- Cloud Platforms for Data Science: AWS, Google Cloud Platform, Azure
- Version Control and Collaboration Tools: Git, GitHub, GitLab

## Ethics and Privacy in Data Science

- Data Privacy and Security: GDPR, CCPA, data anonymization techniques
- Ethical Considerations: Bias, fairness, interpretability in machine learning models
- Responsible AI: Guidelines for ethical AI development and deployment

## Capstone Project in Data Science

- Real-world Data Science Project: From problem formulation to model deployment
- Presentation of Findings: Communicating results to stakeholders
- Peer Review and Feedback: Evaluation and improvement based on feedback