DATA SCIENCE | ADVANCED | 5 DAYS

Data Mining for Business

Build a strong foundation in data mining by understanding end to end processes, techniques, and algorithms. Learn how to apply classification, clustering, association, and text mining methods using R or Python, and effectively interpret and communicate insights for real world business problems.

Why Data Mining?

Extracting Actionable Insights

Extract valuable insights and patterns from large, complex datasets that provide actionable information for making informed decisions and driving business growth.

Enhanced Marketing and Sales Efforts

Personalise and optimse your marketing and sales efforts leadingto improved marketing ROI, increased sales, and better customer engagement.

Improved Decision Making

Make evidence-based decisions, resulting in improved business outcomes.

What you'll be able to do

- Gain a solid foundation in the principles and concepts of data mining.
- Understand the different stages of the data mining process, including data preprocessing, exploration, modeling, and evaluation.
- Learn various data mining techniques such as classification, clustering, association rule mining, and text mining.
- Explore data mining algorithms and their applications in different domains.
- Acquire hands-on experience with popular data mining tools and programming languages like R or Python.
- Develop the ability to select appropriate data mining techniques and algorithms based on specific problem requirements.
- Interpret and communicate the results of data mining analyses effectively.

Who this is for

- Data Scientists
- Business Analysts
- Data Analysts
- IT Professionals

Curriculum

01 Data Exploration and Preprocessing

- Profile datasets for structure, content, and quality issues
- Clean, transform, and normalize data for modeling readiness
- Handle missing values, duplicates, and inconsistent types
- Establish a repeatable preprocessing checklist for mining projects

02 Classification Methods

- Frame business problems as classification tasks
- Train and compare common classifiers on labeled data
- Interpret class predictions and confidence for decision support
- **Hands-on:** Build a baseline classifier and review misclassifications

03 Regression Analysis

- Model continuous outcomes from historical patterns
- Validate assumptions and residual behavior
- Use regression outputs for forecasting and what-if analysis
- Compare simple vs multivariate regression trade-offs

04 Clustering and Association Rules

- Group similar records with clustering techniques
- Choose distance metrics and interpret cluster profiles
- Mine association rules for co-occurrence insights
- Apply results to segmentation and basket-style recommendations

05 Text Mining and Sentiment Analysis

- Prepare unstructured text for mining pipelines
- Extract themes and keywords from documents and feedback
- Score sentiment for customer or social text
- Connect text insights to business actions

06 Time Series and Anomaly Detection

- Analyze trends, seasonality, and change over time
- Detect unusual points and behavioral anomalies
- Choose monitoring thresholds for operational alerts
- **Hands-on:** Flag anomalies in a sample time series

07 Feature Selection and Dimensionality Reduction

- Identify informative features for modeling
- Reduce dimensionality while preserving signal
- Balance interpretability vs predictive power
- Document feature choices for stakeholder review

08 Tree Models and Ensembles

- Build decision trees for transparent decision rules
- Improve accuracy with random forests and ensembles
- Tune depth and ensemble size for generalization
- Interpret feature importance from tree ensembles

09 SVM and Neural Networks

- Apply SVMs for classification and regression tasks
- Introduce neural network patterns for complex signals
- Compare when linear/kernel methods beat deep models
- Evaluate compute and data requirements for each approach

10 Evaluation Metrics and Visualization

- Select metrics aligned to business cost of errors
- Compare models with consistent validation design
- Visualize insights with charts that support decisions
- Communicate model quality to non-technical stakeholders

11 Big Data and Real-Time Mining

- Scale mining techniques to large and complex datasets
- Design near-real-time mining flows for immediate decisions
- Identify tooling patterns for batch vs streaming analysis
- Assess latency, freshness, and reliability trade-offs

12 Privacy, Ethics, and Applied Mining (web, social, geo)

- Apply privacy and ethics checks to mining workflows
- Mine web, social, and geographic signals responsibly
- Map network/spatial patterns to business questions
- Document risk, consent, and governance considerations