Data Mining – Roadmap to AI&ML

Duration: 40 hours

Overview

This extensive course on Data Mining spans foundational concepts to expert-level applications, empowering learners to master the entire spectrum of data mining techniques. It begins with Fundamentals of Data Mining, covering essential preprocessing, exploratory analysis, and feature engineering, complemented by hands-on exercises like data cleaning with pandas and SQL queries. Progressing to Core Data Mining Techniques, it dives into clustering, association rule mining, classification, and text mining, enriched with real-world exercises like spam classification. The Advanced Applications module explores regression, deep learning, graph mining, web scraping, and big data tools such as Hadoop and Spark, with applied tasks like time series forecasting. Finally, the course culminates in Expert-Level Integration, combining AI and business intelligence, ethical considerations, and rigorous model evaluation, supported by real-world case studies in domains like marketing, NLP, and telecom. By the end, learners acquire both theoretical knowledge and hands-on expertise to solve complex data problems in various industries.

Audience Profile

- AI/ML Enthusiasts & Data Scientists
- Data Engineers & Analysts
- Business Intelligence Professionals

Tools and Lab Platform

- Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels)
- Jupyter Notebook/Colab

Course Agenda

Module 1: Fundamentals of Data Mining (Beginner Level)

- Introduction to Data Mining and its Role in AI/ML
- Data Types and Data Sources
- Data Preprocessing: Cleaning, Normalization, Handling Missing Values

- Exploratory Data Analysis (EDA)
- Feature Engineering and Selection
- Introduction to Database & SQL for Data Mining

Practice exercises

- Data Cleaning with pandas
- SQL Queries for Data Extraction
- EDA using Titanic/Iris datasets

Module 2: Core Data Mining Techniques (Intermediate Level)

- Supervised vs Unsupervised Learning
- Clustering: K-Means, DBSCAN, Hierarchical
- Association Rule Mining: Apriori, FP-Growth
- Classification: Decision Trees, Naïve Bayes
- Anomaly Detection & Outlier Analysis
- Intro to Text Mining & NLP

Practice exercises

- Clustering with scikit-learn
- Association Rules using mlxtend
- Spam Classifier with Naïve Bayes

Module 3: Advanced Data Mining Applications (Pro-Level)

- Regression Techniques (Linear, Ridge, Lasso)
- Deep Learning for Data Mining (Neural Networks)
- Graph Mining & Social Network Analysis
- Web Mining & Web Scraping
- Big Data Mining (Hadoop, Spark)
- Time Series Data Mining

Practice exercises

- Time Series Forecasting using ARIMA
- Web Scraping with BeautifulSoup, Scrapy
- Graph Analysis using network
- Spark DataFrame operations on large datasets

Module 4: Expert Level – Real-World Integration & Evaluation

- Data Mining in AI & Business Intelligence
- Ethical Issues in Data Mining & Privacy
- Model Evaluation: Accuracy, Precision, Recall, ROC, AUC
- Cross-validation Techniques

Practice exercises

- Evaluation of ML Models using scikit-learn
- Cross-validation with different models
- Ethical Decision Scenarios in Data Handling

Case Studies

- Sales Data Cleaning & EDA for Retail Store
- Customer Segmentation for Marketing Campaigns
- Fake News Detection using NLP
- Churn Prediction in Telecom using Big Data Tools