Daffodil International University

Department of CSE

Program: MSCSE and MIS

Course Code: CSE506, Course Title: Advanced Data Analytics Course Code: MSMIS510, Course Title: Business Data Processing

Course Overview

This course is designed to provide students with advanced techniques in data analysis, focusing on practical applications rather than heavy programming. It is suitable for both Computer Science (MSCSE) and Management Information Systems (MIS) students, including those without a strong programming background. The course covers data exploration, statistical modelling, predictive analytics, and visualization using user-friendly tools like Excel, Tableau, and Python (with minimal coding).

Course Objectives:

- By the end of this course, students will be able to:
- Understand key concepts in advanced data analysis.
- Apply statistical and machine learning techniques to real-world datasets.
- Use data visualization tools to interpret and present findings.
- Perform predictive analytics using accessible software.
- Analyze case studies from business, healthcare, and social sciences.

Detailed Course Outline:

Module 1: Introduction to Advanced Data Analytics

- 1.1 Overview of Data Analytics
 - Evolution of data analytics
 - o Types of analytics: Descriptive, Diagnostic, Predictive, and Prescriptive
 - Applications of advanced analytics in various industries

• 1.2 Data Analytics Lifecycle

- Problem definition
- o Data collection and preparation
- Model building and validation

• Deployment and monitoring

• 1.3 Tools and Technologies

- Overview of tools: Python, R, SQL, Tableau, Power BI, Spark, etc.
- Introduction to cloud platforms: AWS, Google Cloud, Azure

Module 2: Data Preprocessing and Cleaning

- 2.1 Data Collection and Integration
 - o Data sources: Structured, semi-structured, and unstructured data
 - APIs, web scraping, and database integration

• 2.2 Data Cleaning

- Handling missing data
- o Outlier detection and treatment
- Data normalization and transformation

• 2.3 Feature Engineering

- Feature selection and extraction
- Dimensionality reduction techniques (PCA, t-SNE)
- Encoding categorical variables

Module 3: Statistical Modelling for Decision Making

- Hypothesis testing (t-tests, ANOVA, Chi-square)
- Regression analysis (linear & logistic regression)
- Time-series forecasting (moving averages, trend analysis)
- Hands-on: Predicting business trends using Excel/Tableau

Module 4: Machine Learning for Advanced Analytics

- 4.1 Supervised Learning
 - Regression models (Linear, Polynomial, Ridge, Lasso)
 - Classification models (Decision Trees, Random Forest, SVM, k-NN)
 - Model evaluation metrics (Accuracy, Precision, Recall, F1 Score, ROC-AUC)
- 4.2 Unsupervised Learning
 - Clustering techniques (k-Means, Hierarchical, DBSCAN)

- Association rule mining (Apriori, FP-Growth)
- 4.3 Advanced Machine Learning Techniques
 - Ensemble methods (Bagging, Boosting, Stacking)
 - Neural networks and deep learning basics
 - Time series analysis and forecasting (ARIMA, LSTM)

Module 6: Big Data & AI in Modern Analytics

- Introduction to Big Data (Hadoop, Spark conceptual)
- AI & automation in analytics (ChatGPT for data insights)
- Limitations and challenges in real-world data analysis

Module 6: Data Visualization and Storytelling

- 6.1 Principles of Data Visualization
 - Choosing the right chart type
 - Design principles for effective visualizations
- 6.2 Advanced Visualization Tools
 - Tableau, Power BI, and Python libraries (Matplotlib, Seaborn, Plotly)
 - Interactive dashboards and reports
- 6.3 Data Storytelling
 - Communicating insights effectively
 - o Creating narratives with data

Module 7: Capstone Project

• 9.1 Project Planning

- Defining the problem statement
- Data collection and preparation
- 9.2 Implementation
 - Applying advanced analytics techniques
 - o Building and validating models
- 9.3 Presentation
 - o Visualizing results
 - Delivering insights and recommendations

Assessment and Evaluation

- Quizzes (15%)
- Assignment (5%)
- Presentation (8%)
- Mid-term exam (25%)
- Attendance (7%)
- Final exam (40%)

Recommended Textbooks and Resources

- "Python for Data Analysis" by Wes McKinney
- "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron
- "Data Science for Business" by Foster Provost and Tom Fawcett
- Online resources: Kaggle, Towards Data Science, Coursera, and edX