JSPM's
RAJARSHI SHAHU COLLEGE OF ENGINEERING
TATHAWADE, PUNE-33
(An Autonomous Institute Affiliated to Savitribai Phule Pune University, Pune)
Department of Computer Engineering
Elective III: [CS3110A]: Data Mining

Unit-Wise Assignments

Deadline: 31/03/2025

🔷 Task 1: Introduction to Data Mining (Unit I)

📌 **Title:** Comparative Study of Data Mining Techniques
✅ **Objective:** Evaluate different data mining techniques and justify their effectiveness in various applications.
✅ **Task:**

- Research and compare data mining techniques (Clustering, Classification, Regression, Association Rule Mining).
- Identify real-world applications where each technique is best suited.
- Write a report critically evaluating their advantages and limitations.

🎯 **Deliverable:** A report with a comparative analysis of at least 4 techniques.

🔷 Task 2: Data Preprocessing & Knowledge Representation (Unit II)

📌 **Title:** Data Cleaning and Transformation – Impact on Model Accuracy
✅ **Objective:** Evaluate how data preprocessing techniques affect the accuracy of machine learning models.
✅ **Task:**

- Take a real-world dataset (e.g., UCI Machine Learning Repository).
- Apply **data cleaning, transformation, reduction, and discretization** techniques using Weka or Python.
- Compare model performance before and after preprocessing.

🎯 **Deliverable:** A documented case study with before/after accuracy comparisons.

🔷 Task 3: Data Mining Algorithms (Unit III)

📌 **Title:** Performance Evaluation of Classification & Association Rule Mining
✅ **Objective:** Evaluate the efficiency of different classification algorithms and association rule

mining techniques.

✅ **Task:**

- Use a dataset (e.g., Weather Data, Retail Transactions).
- Implement **Decision Trees, Naïve Bayes, k-NN, and Association Rule Mining** algorithms.
- Compare precision, recall, and execution time.

🎯 **Deliverable:** A detailed performance analysis report with justification.

🔷 **Task 4: Descriptive Analytics & Forecasting (Unit IV)**

📌 **Title:** Forecasting Trends Using Predictive Models

✅ **Objective:** Analyze trends and evaluate forecasting accuracy of predictive models.

✅ **Task:**

- Select a dataset (e.g., Stock Market, Sales Data).
- Apply **Linear Regression, Logistic Regression, and Time Series Forecasting.**
- Compare forecasting accuracy using RMSE, MAE, and R^2.

🎯 **Deliverable:** A comparative study with graphs and error analysis.

🔷 **Task 5: Linear & Non-Linear Models (Unit V)**

📌 **Title:** Comparison of Linear & Nonlinear Models in Real-World Data

✅ **Objective:** Justify the choice of linear vs. nonlinear models for specific datasets.

✅ **Task:**

- Choose a dataset (e.g., COVID-19 cases, Weather Data).
- Apply **Generalized Linear Models (GLM) and Nonlinear Regression.**
- Evaluate which model best fits the data based on residual analysis and prediction accuracy.

🎯 **Deliverable:** A report explaining the limitations of linear models and when nonlinear models are required.

🔷 **Task 6: Time Series Models & Prescriptive Analytics (Unit VI)**

📌 **Title:** Time Series Forecasting & Optimization

✅ **Objective:** Evaluate different time series models and implement prescriptive analytics for

decision-making.

✅ **Task:**

Select a **time series dataset** (e.g., Energy Consumption, Temperature Trends).

- Apply **ARIMA, Holt-Winters, and Moving Average models** for forecasting.
- Implement **Decision Trees** for prescriptive analytics in risk assessment.

🎯 **Deliverable:** A final report and forecasting dashboard with model comparison.

💡 **Final Submission Checklist**

✅ **Project Report (8-10 pages) with evaluation, justifications & results**
✅ **Code Files (Python/Weka) with proper documentation**
✅ **Presentation Slides (8-12 slides) summarizing findings**

Evaluation Criteria

Criterion	**Weightage**
Depth of Analysis & Justification	30%
Model Implementation & Accuracy	30%
Clarity of Report & Presentation	20%
Practical Application & Insights	20%

Subject Incharge

Dr. Nisha Ashish Wandile