

Project Title: Exploring Customer Behavior Patterns Using Data Analytics

25 marks

Objective:

The objective of this project is to analyze customer behavior data to identify patterns and trends using various data analytics techniques covered in the course.

Project Overview:

The project will involve analyzing a dataset containing customer-related data, such as demographic information, purchase history, website interactions, and feedback. The dataset will be used to perform a variety of data analysis tasks and techniques, including supervised and unsupervised learning methods, regression, classification, clustering, anomaly detection, principal component analysis (PCA), data summarization, and visualization.

Project Tasks:

Within your team (4 students), the following tasks must be accomplished:

1. Data Acquisition and Preparation (Week1):

- Acquire a dataset containing customer-related data from multiple sources, such as CRM systems, transaction records, and website logs. Here are some resources where you can find datasets for your project:
 - o <https://www.kaggle.com/>
 - o <https://datahub.io/>
 - o <https://archive.ics.uci.edu/>
 - o <https://datasetsearch.research.google.com/>
- Preprocess the data by handling missing values, encoding categorical variables, and normalizing numerical features.

2. Exploratory Data Analysis (EDA) (Week 2):

- Perform exploratory data analysis to gain insights into the dataset's characteristics, distributions, and correlations.
- Visualize key features using histograms, scatter plots, and correlation matrices.

3. Linear and Nonlinear Regression (Week 3):

- Implement linear regression models to predict customer spending based on demographic and behavioral factors.
- Explore the use of nonlinear regression techniques, such as polynomial regression, to capture complex relationships in the data.
- Evaluate the effectiveness of the chosen model using appropriate metrics and suggest improvement actions.

4. Logistic Regression and Classification (Week 4):

- Apply logistic regression to predict binary outcomes, such as customer churn or purchase likelihood.
- Utilize classification algorithms, such as decision trees or random forests, to categorize customers into segments based on their characteristics.
- Evaluate the effectiveness of the chosen algorithms using appropriate metrics and suggest improvement actions.

5. Clustering (Week 5):

- Employ clustering techniques, such as K-means or hierarchical clustering, to group customers with similar attributes or behaviors.
- Evaluate the effectiveness of different clustering algorithms using appropriate metrics and suggest improvement actions.

6. Anomaly Detection (Week 6):

- Identify unusual patterns or outliers in customer behavior data using anomaly detection methods, such as isolation forests or density-based clustering.

7. Principal Component Analysis (PCA) (Week 7):

- Perform PCA to reduce the dimensionality of the dataset while preserving the most important features.
- Visualize the principal components and analyze their contributions to the variance in the data.

8. Data Summarization and Visualization (Week 8):

- Summarize key insights and findings from the analysis, including significant trends, correlations, and clusters.
- Create visualizations, such as heatmaps, scatter plots, and bar charts, to communicate the results effectively.

Deliverables (Week 9):

1. Final project report detailing the dataset, methodology, analysis results **and discussion**, and conclusions.
2. Presentation slides summarizing the project findings and insights.
3. Code repository containing scripts, notebooks, and documentation used for data analysis.

Project Evaluation:

The project will be evaluated based on the following criteria:

1. Effectiveness of data preprocessing and feature engineering (5%)
2. Rigor and appropriateness of the chosen data analysis techniques (5%)
3. Clarity and depth of insights derived from the analysis (5%)
4. Quality of visualizations and presentation of results (5%)
5. Creativity and innovation in addressing the project objectives (5%)