**Objective:**

The primary objective of this study is to utilize data analytics techniques to enhance business decision-making processes. The study aims to identify patterns and insights from **Predict Online Course Engagement Dataset** to support strategic planning, optimize operations, and improve the quality of participation from students.

**Rationale:**

In today’s education system, online learning has become an inevitable part for the students as their learning platforms other than physical tuitions. So there are more and more need of the online course platform. Data analytics offers powerful tools and methodologies to extract valuable information, enabling the online course platforms to make informed decisions. This study aims to demonstrate the **Prediction of a candidate can complete the course or not.**

**Methodology:**

**1. Data Collection and Preprocessing:**

Data Sources: The data is collected from Kaggle website. [https://www.kaggle.com/datasets/rabieelkharoua/predict-online-course-engagement-dataset/data]

Data Cleaning: The data was almost perfect to begin with analyzation.

Data Integration: The data was already found in an integrated csv file.

**2. Data Analysis:**

**Predictive Analytics:** This methodology was chosen for this dataset and the problem statement as to employ machine learning models to forecast future outcomes based on historical data.

**3. Model Development:**

**Algorithm Selection:-**

LogisticRegression, DecisionTree, GradiantBoost, RandomForest, KNeighborsClassifier

**Training and Validation:-**

Split the dataset into training and validation sets. Train models on the training set and evaluate their performance on the validation set.

Following were the train and test set splits:

X\_train shape: (7200, 7)

X\_test shape: (1800, 7)

Y\_train shape: (7200,)

Y\_test shape: (1800,)

Note: “May vary during the run of code.”

**Model Tuning:** Optimize model parameters to enhance accuracy and robustness.

**4. Implementation and Testing:**

**Pilot Study:** A set of data was entered in the model after the evaluation of the algorithms and the result was fairly acceptable.

**Dataset:**

This dataset captures user engagement metrics from an online course platform, facilitating analyses on factors influencing course completion. It includes user demographics, course-specific data, and engagement metrics.

**Features:**

* **UserID:** Unique identifier for each user
* **CourseCategory:** Category of the course taken by the user (e.g., Programming, Business, Arts)
* **TimeSpentOnCourse:** Total time spent by the user on the course in hours
* **NumberOfVideosWatched:** Total number of videos watched by the user
* **NumberOfQuizzesTaken:** Total number of quizzes taken by the user
* **QuizScores:** Average scores achieved by the user in quizzes (percentage)
* **CompletionRate:** Percentage of course content completed by the user
* **DeviceType:** Type of device used by the user (Device Type: Desktop (0) or Mobile (1))
* **CourseCompletion (Target Variable):** Course completion status (0: Not Completed, 1: Completed)

**Target:**

Distribution of the Target Variable (CourseCompletion)

**Expected Outcomes:**

* The important factors needed for the course completion.
* Whether a certain candidate can complete the course or not.
* What features gives the rise to more involvement in the online course learning.

**Conclusion:**

This study aims to show the tool, techniques and methodologies of Data Analytics in upgrading the quality and visibility of the online course learning platforms to rise above it’s peak and provide the best features to enhance the learning experience to the students.