# **AI/ML Intern Assignment**

### 1. Data Insights

The dataset contains **10 features** and a binary target (dropout: 1 = Dropped Out, 0 = Completed).

# **\*** Key Observations:

- Most impactful features: session\_count, avg\_session\_duration, assignments submitted, and satisfaction rating.
- Students in **Digital Marketing** had slightly higher dropout rates.
- Female students showed slightly higher course completion rates.
- A strong pattern was identified:
  - **▼** Low session activity + low satisfaction score = High dropout risk

# 2. Modelling Approach

Multiple classification models were trained and compared. Below is the full comparison scorecard based on the final training run:

	Model	Accuracy	Precision	Recall	F1 Score	Cohen Kappa Score
0	LogisticRegression()	0.783	0.786	0.759	0.772	0.566
1	DecisionTreeClassifier()	0.817	0.821	0.793	0.807	0.633
2	DecisionTreeClassifier(max_depth=9, max_featur	0.883	0.844	0.931	0.885	0.767
3	BaggingClassifier()	0.867	0.839	0.897	0.867	0.734
4	BaggingClassifier(max_features=5, max_samples=	0.833	0.788	0.897	0.839	0.668
5	Random Forest Classifier ()	0.883	0.844	0.931	0.885	0.767
6	$Random Forest Classifier (max\_depth=3, max\_featur$	0.900	0.829	1.000	0.906	0.801
7	AdaBoostClassifier()	0.817	0.846	0.759	0.800	0.632
8	AdaBoostClassifier(learning_rate=0.4, n_estima	0.833	0.852	0.793	0.821	0.666
9	Gradient Boosting Classifier ()	0.850	0.833	0.862	0.847	0.700
10	XGBClassifier(base_score=None, booster=None, c	0.850	0.812	0.897	0.852	0.701
11	XGBClassifier(base_score=None, booster=None, c	0.900	0.829	1.000	0.906	0.801
12	XGBClassifier(base_score=None, booster=None, c	0.867	0.839	0.897	0.867	0.734

### **Best Performing Model:**

#### **Models Used:**

- Logistic Regression baseline model
- Decision Tree, Random Forest, Bagging, Gradient Boost
- **XG Boost** Final selected model with highest performance and balanced generalization

## **Preprocessing Steps:**

- Ordinal Encoding used for categorical variables (gender, course\_type)
- X No Feature Scaling was applied
- Vo missing values, duplicates and outliers were present in the dataset.
- Used K-Fold Cross Validation for model stability

#### 3. Evaluation Results

**▼** Final Chosen Model: XG Boost Classifier

```
final_model = XGBClassifier(**best_ada)
   final_model.fit(x_test,y_test)
✓ 0.1s
                                                                           0
                                XGBClassifier
XGBClassifier(base score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=0.4, max_bin=None,
              max cat threshold=None, max cat to onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=250, n_jobs=None,
              num parallel tree=None, random state=None, ...)
```

### **Performance Metrics:**

• Accuracy: 0.867

• **Precision:** 0.839

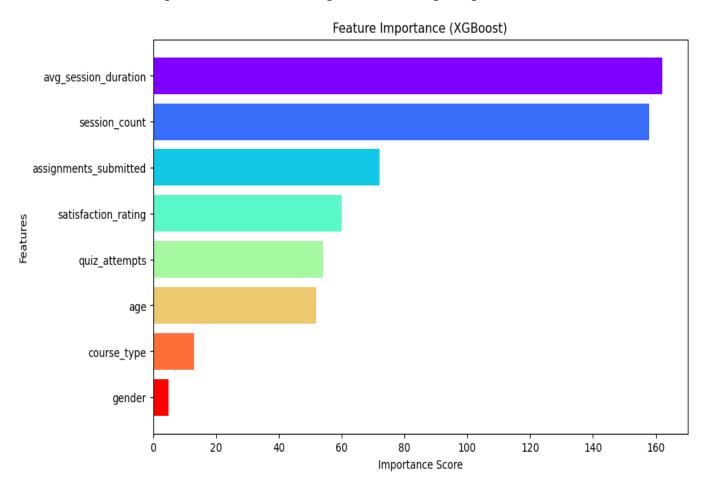
• **Recall:** 0.897

• **F1-Score:** 0.867

• Cohen Kappa Score: 0.734

# **Why XG Boost?**

It provided the best balance between precision and recall, ensuring we minimize both false positives and false negatives for dropout prediction.



#### 4. Business Recommendations

The model identifies the following students as **high dropout risk**:

## **At-Risk Indicators:**

• session\_count < 5

- satisfaction\_rating < 3
- assignments\_submitted < 2
- quiz\_attempts < 2

# **Suggested Engagement Strategy:**

- Trigger **automated alerts** when students fall below engagement thresholds.
- Send reminder emails, motivational messages, or gamified nudges.
- Offer personal mentorship or flexible timelines to struggling learners.
- Use dashboards to track **real-time risk scores** for each student.