# Lead Conversion Analysis for X Education

Logistic Regression

# Data-Driven Approach to Improve Lead Conversion

- **Problem Statement**

- X Education aims to improve lead conversion rates by identifying the most influential factors.
- Goal: Build a predictive model to score leads and optimize conversion efforts.

# Data Overview

**Dataset** : Lead data with demographic, behavioral, and interaction features.

Key Features:

- Numerical: Total Visits, Page Views Per Visit
- Categorical: Lead Source, Specialization, Last Activity
- Target Variable: Lead Converted (0/1)

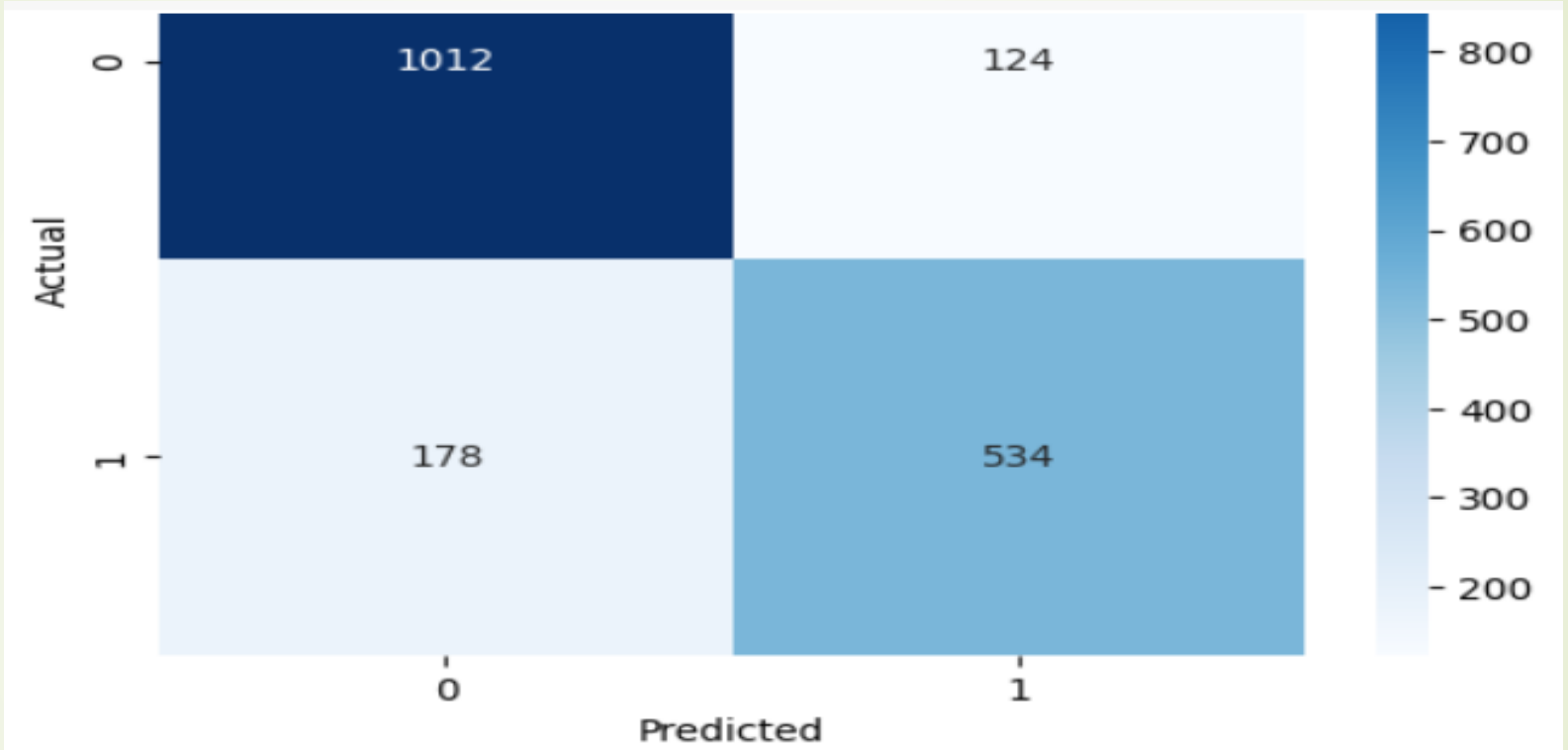
# Data Preprocessing & Feature Engineering

- **Handled Missing Values** via imputation and categorical mode filling.
- **Encoded Categorical Variables** (One-Hot Encoding & Frequency Encoding).
- **Standardized Numerical Features** using **Standard Scaler**.
- Removed high-cardinality & redundant features.

# Model Training and Evaluation

- Logistic Regression Model trained for interpretability.
- Feature Selection via Recursive Feature Elimination (RFE).
- Hyperparameter tuning: max_iter =500, solver= 'liblinear'.
- Accuracy: 84%
- Precision & Recall: Balanced for effective lead prioritization.
- ROC-AUC Score: 0.8849 (Good discriminative ability).
- Confusion Matrix & ROC Curve (Visuals)
- Hh  data split looks correct, with 7,392 samples in the training set and 1,848 in the test set, each having 108 features. The target variable (y_train and y_test) is correctly shaped as a one-dimensional array.

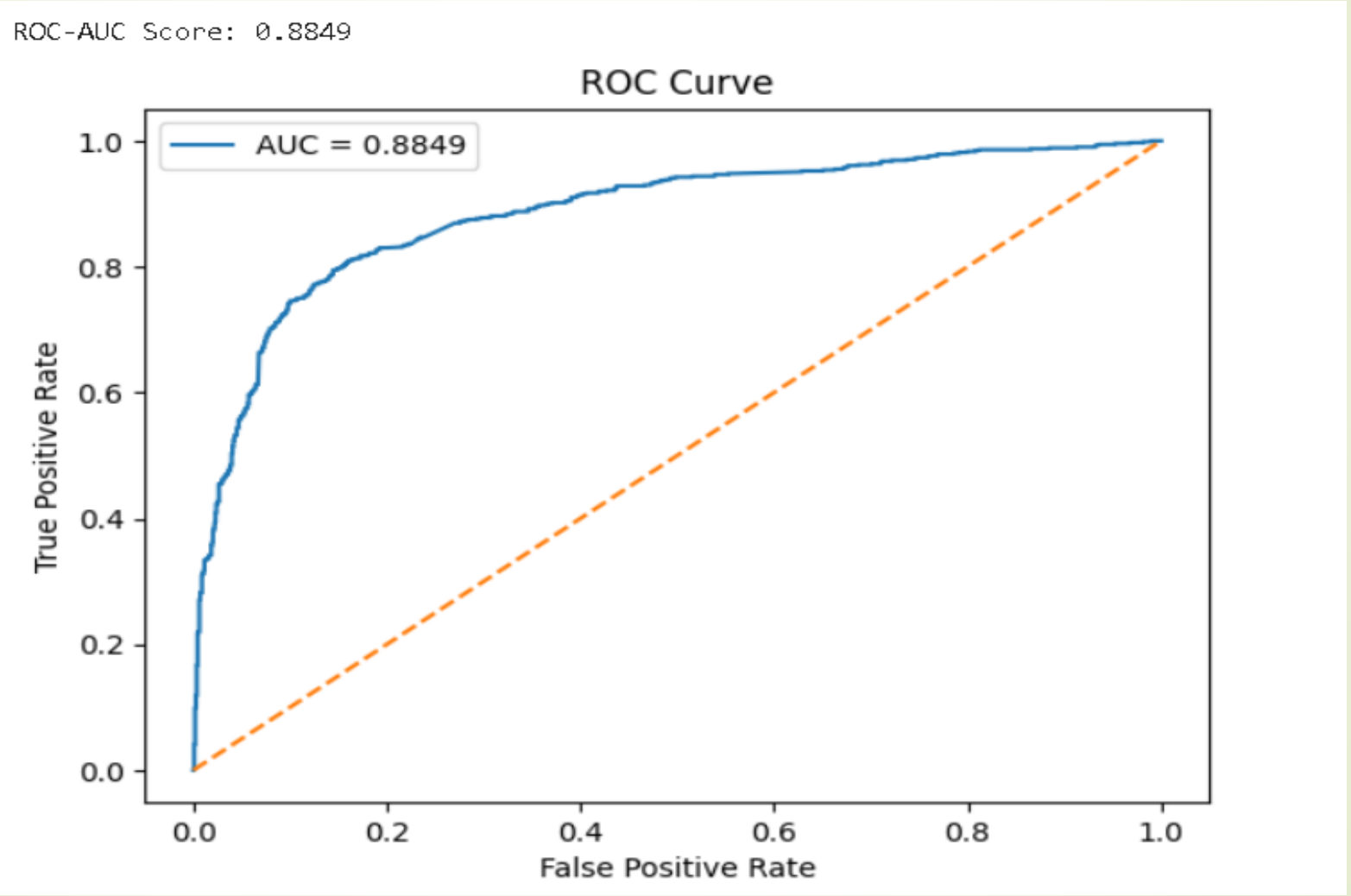
# Actual vs Predicted Values




| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.85 | 0.89 | 0.87 | 1136 |
| 1 | 0.81 | 0.75 | 0.78 | 712 |
| accuracy | | | 0.84 | 1848 |
| macro avg | 0.83 | 0.82 | 0.82 | 1848 |
| weighted avg | 0.84 | 0.84 | 0.84 | 1848 |

# ROC Curve

ROC-AUC Score: 0.8849



# Model Analysis

- Accuracy: 84% → Model correctly classifies 84% of leads.
- Confusion Matrix:
- 1012 true negatives (correctly predicted non-conversions).
- 534 true positives (correctly predicted conversions).
- 124 false positives (incorrectly predicted as conversions).
- 178 false negatives (missed actual conversions).
- Precision:
- 85% for non-converting leads.
- 81% for converting leads (important for minimizing marketing waste).
- Recall (Sensitivity):
- 89% for non-converting leads.
- 75% for converting leads (misses some true conversions).

# Key Insights and Business Recommendations

- **Key Insights from Model**

  **Top 3 Most Influential Features:**

  - **Tags (3.30)** → Highly impacts conversion.
  - **Lead Origin Lead Add Form (2.81)** → Form submissions drive conversions.
  - **Lead Source Welingak Website (1.88)** → Strong lead source.

- **Business Recommendations**

1. **Prioritize High-Scoring Leads** for proactive engagement.
2. **Focus on ‘Lead Add Form’ & ‘Welingak Website’ channels.**
3. **Increase personalization for 'Tags' (highly influential leads).**

# Conclusion

- **Successful Model Deployment** to optimize lead conversion.

- **Future Scope:**
  Implement **automated lead scoring** in CRM.
  Further optimize calling & engagement strategies.

- **Final Takeaway:**
  Implement **automated lead scoring** in CRM.
  Further optimize calling & engagement strategies.