LEAD SCORING PRESENTATION

GROUP MEMBERS: APARNA ACHARYA, ANKIT KUMAR, KASHIF KHALID AHMED

AGENDA

Problem statement

Chosen Methodology

Data Preparation

Feature selection

Model building

Model evaluation

Conclusion

PROBLEM STATEMENT

- **Objective:** Develop a model to identify and prioritize leads most likely to convert into paying customers.

- **Target Conversion Rate:** 80% as specified by the CEO.

- **Goal:** Create a scoring system to reflect the likelihood of conversion.

CHOSEN METHODOLOGY

- **Model Used:** Logistic Regression

- **Reason for Choice:**
- Effective for binary classification.
- Provides probability scores for lead conversion.

DATA PREPARATION

Data Cleaning:

- **Removed Columns:** Excluded columns with more than 3,400 null values.
- **Removed Rows:** Eliminated rows with missing values in remaining columns.
- **Purpose:** Ensured dataset quality and completeness.

Data Feature Engineering:

- Transformed categorical variables into dummy variables.
- Created binary columns for each category.

Data Scaling:

- Technique Used: MinMaxScaler

Data Splitting:

- Split Ratio: 70% training, 30% testing

FEATURE SELECTION

- **Method:** Recursive Feature Elimination (RFE)
- **Outcome:** Selected top 15 features for the model.

MODEL BUILDING

•**Initial Model:** Built with selected features.
•**Refinement:**
•Removed variables with high Variance Inflation Factor (VIF) and insignificant p-values.
•Reduced multicollinearity and improved model interpretability.

MODEL VALIDATION

- **Initial Validation:** Used a random cutoff value.
- **Refined Validation:**
- Employed Receiver Operating Characteristic (ROC) curve.
- Determined an optimal cutoff value for classification.

PERFORMANCE EVALUATION

Performance Evaluation

- **Plot Analysis:**Assessed precision and recall.
- Satisfactory results indicated effective model performance.

Metrics on Test Data

- **Metrics Evaluated:**Sensitivity
- Specificity
- Precision
- Recall
- Accuracy

CONCLUSION

- **Approach:** Structured methodology of data cleaning, feature engineering, model building, and validation.
- **Outcome:** Developed a logistic regression model that meets X Education’s conversion goals.
- **Impact:** High scoring leads are more likely to convert, achieving the target conversion rate of 80%.

THANK YOU