

Presentation and Recommendations

Lead Conversion Prediction

By Umar Farooq Bhat

Project Objective

- Predict which leads are likely to convert.
- Help marketing & sales prioritize high-potential leads.
- Improve efficiency and conversion rates using data-driven insights.

Dataset Overview

- Total Rows: 9,240
- Columns: 37 (categorical, numerical, and text features)
- Target Variable: Converted (1 = Converted, 0 = Not Converted)
- Challenges: Missing values, multicollinearity, irrelevant features

Key Insights from EDA

- Top Lead Sources: Olark Chat, Google, and Direct Traffic lead to the highest conversion rates.
- Tag Influence: Tags such as “Interested in Next Batch” and “Will revert after email” are strong predictors of conversion.
- Behavioral Indicators: More time spent on the website and more page views strongly correlate with conversions.
- High-Performing Specializations: HR and Marketing specializations convert at higher rates.

Feature Engineering

- Addressed missing data via dropping or imputation.
- Removed irrelevant columns and duplicate entries.

- Encoded categorical variables using one-hot encoding (dummies).
- Removed multicollinearity using Variance Inflation Factor (VIF).
- Scaled numerical features before applying Logistic Regression.

Model Building

- Models Used: Logistic Regression and Random Forest
- Data Split: 80% Training / 20% Testing (stratified)
- Logistic Regression: Used for interpretability, selected features based on p-values
- Random Forest: Used for robustness and feature importance visualization

Model Evaluation

- Accuracy: 96%
- Precision: 96.3%
- Recall: 94.1%
- F1 Score: 95.2%
- ROC AUC Score: 0.986 (Logistic Regression)
- Random Forest achieved comparable performance.

Business Recommendations

1. Prioritize leads with engagement tags like “Interested in Next Batch” or “Will revert”.
2. Invest more in high-performing lead sources like Google, Olark Chat, and Direct Traffic.
3. Improve website experience to encourage longer time-on-site and more page views.
4. Automate lead scoring by integrating the model into CRM systems.

Assumptions

- All leads belong to the same course/product line.
- Tags and lead scores are available at prediction time.

- Lead behavior patterns will remain stable over time.
- No major changes in marketing strategy occurred during data collection.

Conclusion

- The model is highly accurate (AUC = 0.986), interpretable, and suitable for deployment.
- Helps marketing and sales teams focus on the right leads.
- Can improve ROI, conversion rates, and decision-making through automation and insights