Improving Lead Conversion Efficiency at X Education

Anish Gupta

Aachal Deshmukh
Ninad Kalambe

Problem Statement & Business Goals

- **Business Problem**:
- X Education's lead conversion rate is only **30%**.
- Sales team struggles to focus on potential leads.
- **Goal**:
- Build a **Lead Scoring Model** to prioritize leads most likely to convert.
- Target: Achieve **80% lead conversion**.
- Optimize sales efforts.
- Improve ROI on marketing and sales outreach.

Data Understanding & Preprocessing

Dataset:

•~9000 leads, features include:

- **Categorical**: Lead Source, Last Activity
- **Numerical**: Total Time Spent, Total Visits
- Target: **Converted** (1 = Converted, 0 = Not Converted)

Preprocessing Steps:

1.Handle Missing Values:

1. Replace invalid "Select" with NaN and drop rows.

2.Encoding Categorical Features:

1. Used **Label Encoding** for Lead Source, Last Activity.

Continue

2. Encoding Categorical Features:
    1. Used **Label Encoding** for Lead Source, Last Activity.

Python Code:

```
df.replace('Select', np.nan, inplace=True)
 df.dropna(inplace=True)
from sklearn.preprocessing import LabelEncoder
 for col in ['Lead Source', 'Last Activity']:
            df[col] = LabelEncoder().fit_transform(df[col])
```

Exploratory Data Analysis (EDA)

Key Insights:

1.Leads spending **more time on the website** have a higher chance of conversion.

2.Some **lead sources** (e.g., referrals, search engines) outperform others.

Visualization:

```
import seaborn as sns
sns.boxplot(x='Converted', y='Total Time Spent on Website', data=df)
```

Observation:

•Higher time on the website → Higher conversion probability.

Model Building & Results

Model: Logistic Regression

•Chosen for its simplicity, interpretability, and effectiveness for binary classification.

Model Performance: Recommendations

•Achieved ~80% accuracy.

•Successfully prioritized leads for sales efforts.

Model Building & Results

Model: Logistic Regression

•Chosen for its simplicity, interpretability, and effectiveness for binary classification.

Model Performance: Recommendations

•Achieved ~80% accuracy.

•Successfully prioritized leads for sales efforts.

Recommendations

1. Prioritize Hot Leads:

•Use lead scores (70–100) to focus sales efforts.

•Categories:

- Hot Leads: High Priority
- Warm Leads: Medium Priority
- Cold Leads: Low Priority

2. Increase Engagement:

•Encourage more time spent on the website for Warm Leads.

•Invest in high-performing lead sources (e.g., referrals).

Business Impact & Conclusion

Projected Impact:

- Conversion rate improves from **30% to ~80%**.
- Optimized sales team efficiency → More conversions with fewer resources.
- Higher ROI on marketing and sales.

Conclusion:

- Built a **Logistic Regression Lead Scoring Model**.
- Identified key predictors (Total Time Spent, Lead Source).
- Provided actionable strategies to improve lead nurturing and conversions.

Thank you