Lead Scoring Model for X Education

1. Business Problem

X Education, an online learning platform, aims to optimize its sales process by prioritizing high-potential leads. The goal is to develop a predictive model to estimate the probability of lead conversion, allowing the sales team to focus on leads with the highest likelihood of enrollment.

2. Methodology

- **Data Collection & Preprocessing:**
    - The dataset contains multiple categorical and numerical features related to lead interactions, demographics, and engagement.
    - Unnecessary columns such as 'Prospect ID' and 'Lead Number' were removed.
    - Missing values were handled through imputation and removal of redundant data.
- **Feature Engineering & Encoding:**
    - Categorical variables were transformed using **One-Hot Encoding** to facilitate model learning.
    - Numerical features were standardized using **StandardScaler** for better model performance.
- **Model Selection:**
    - **Logistic Regression** was chosen for its interpretability and efficiency in binary classification.
    - The dataset was split into **80% training and 20% testing** to evaluate generalization performance.

3. Model Performance

The Logistic Regression model demonstrated excellent performance:

- **Accuracy:** 96.4%
- **Precision:** 96.2%
- **Recall:** 97.1%
- **F1 Score:** 96.7%
- **ROC-AUC Score:** 98.7%

These metrics indicate a strong balance between identifying actual conversions and minimizing false positives.

4. Key Insights

Top Contributing Variables

- **Numerical Variables:**
    1. Total Time Spent on Website
    2. Lead Source - Google Search
    3. Number of Page Views per Visit
- **Categorical Variables (One-Hot Encoded):**
    1. Tags - "Will revert after reading the email"
    2. Tags - "Closed by Horizzon"
    3. Tags - "Lost to EINS"

5. Business Recommendations

Increasing Lead Conversions:

- Prioritize leads with high scores (**Lead Score > 90**) for immediate follow-up.
- Allocate additional resources to optimize marketing efforts around high-converting sources.
- Focus on refining the follow-up process for leads with specific tags like **'Will revert after reading the email'.**

Intern Hiring Phase Strategy:

- During peak hiring seasons, sales interns should focus on **all high-scoring leads (Lead Score > 80).**
- Implement an **automated call scheduling system** to maximize contact rates.
- Increase touchpoints via email and SMS before direct sales engagement.

Quarterly Target Achievement Strategy:

- When targets are met early, **reduce call volume** to low-priority leads (Lead Score < 50).
- Focus the team's efforts on **upselling to existing customers** or engaging potential B2B clients.

- Implement a **lead nurturing strategy** for low-scoring leads rather than immediate calls.

6. Conclusion

The lead scoring model effectively identifies high-value prospects, allowing the sales team to prioritize efforts efficiently. By leveraging **high-impact variables** and adjusting sales strategies based on lead scores, X Education can maximize conversions while optimizing resource allocation.

Further improvements can be made by integrating real-time lead scoring into CRM systems and continuously updating the model with new data insights.