

Lead Scoring Model for X Education

IMPROVING LEAD CONVERSION RATES

Team Members:

Manjit Singh, Kavita Vishnu Purohit, Shubham Jaiswal

Problem Statement

- ▶ X Education, an online course provider, faces low lead conversion rates (~30%).
- ▶ The company generates leads through various marketing channels and past referrals.
- ▶ Leads are classified when potential customers fill out a form on the website.
- ▶ The goal is to improve the lead conversion rate to approximately 80%.
- ▶ Identify "Hot Leads" that are more likely to convert, optimizing the sales team's efforts.
- ▶ Build a logistic regression model to assign lead scores (0-100) to prioritize potential leads.
- ▶ Dataset includes ~9000 data points with features like Lead Source, Time Spent on Website, and Last Activity.

Analysis Approach

- ▶ **Data Prep and Exploration:** Imported dataset, handled categorical variables and missing data ('Select' values), addressed outliers, checked for duplicates.
- ▶ **Visualization:** Utilized plots to explore data distributions and correlations.
- ▶ **Model Development:** Engineered features, created dummy variables, split data into training and testing sets, applied feature scaling, initialized logistic regression model.
- ▶ **Model Building:** Used RFE for feature selection, evaluated models using metrics like Accuracy and ROC-AUC.
- ▶ **Refinement:** Tuned models for performance improvement, determined optimal cutoff points for predictions.

Results and Findings

► Model Capability Overview

- The Final logistic regression model accurately predicts lead conversions by analyzing key features like lead origin, source, activity, and tags. This capability enables efficient resource allocation and targeted sales strategies.

► Cutoff Point Selection

- Using a cutoff point of 0.3, the model categorizes leads into "Hot Leads" with high conversion potential and "Cold Leads" with lower potential. This approach optimizes lead prioritization and minimizes false predictions.

► Key Features Influencing Conversion

- Identified 13 key features that significantly impact conversion likelihood. Positive influencers include Tags_Lost to EINS and Tags_Closed by Horizzon, while negative influencers include Lead Quality_Worst and Tags_switched off.

Model Performance and Implementation

► Model Performance Metrics

- Achieved 90% Accuracy, indicating correct classification of 90% of cases. Sensitivity of 87% ensures identification of 87% of converting customers, with a Precision of 87% for accurately identifying Hot Leads.

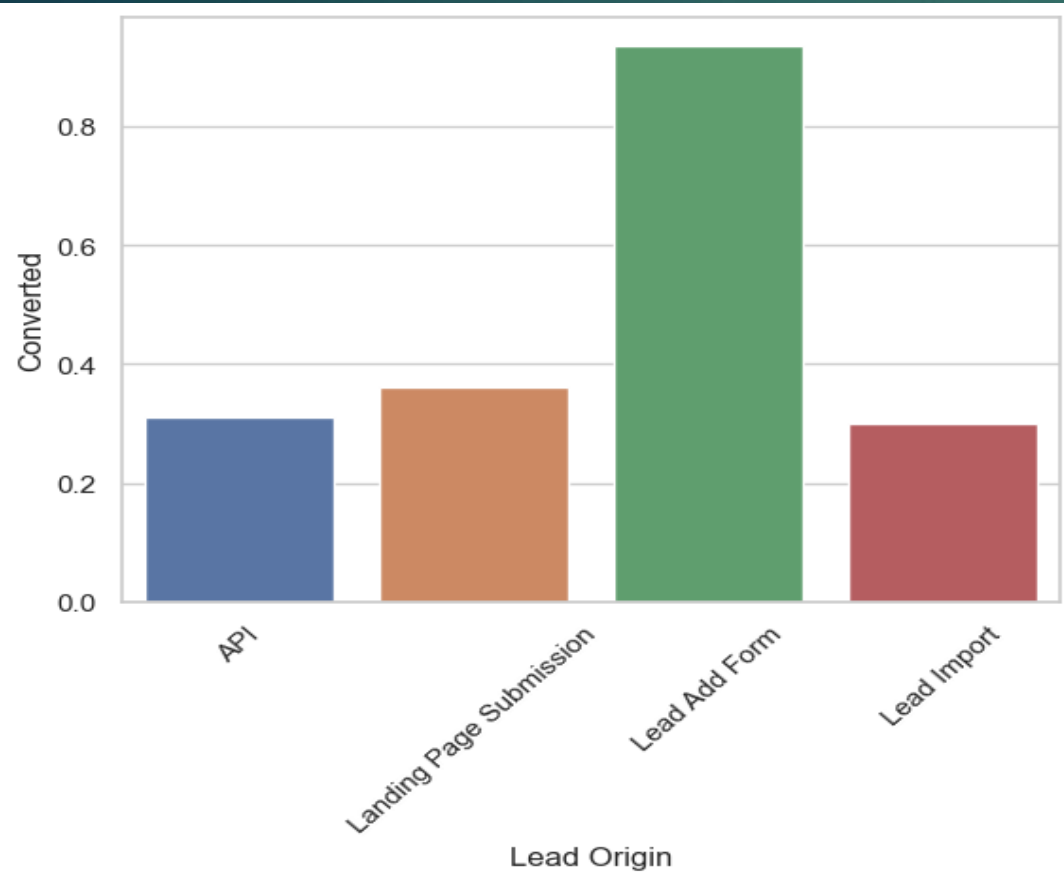
► Operational Implementation

- Developed a reusable prediction code block for conversion probabilities and lead scores. Adjustable cutoff points optimize strategies based on specific performance metrics like sensitivity or precision, enhancing decision-making capabilities.

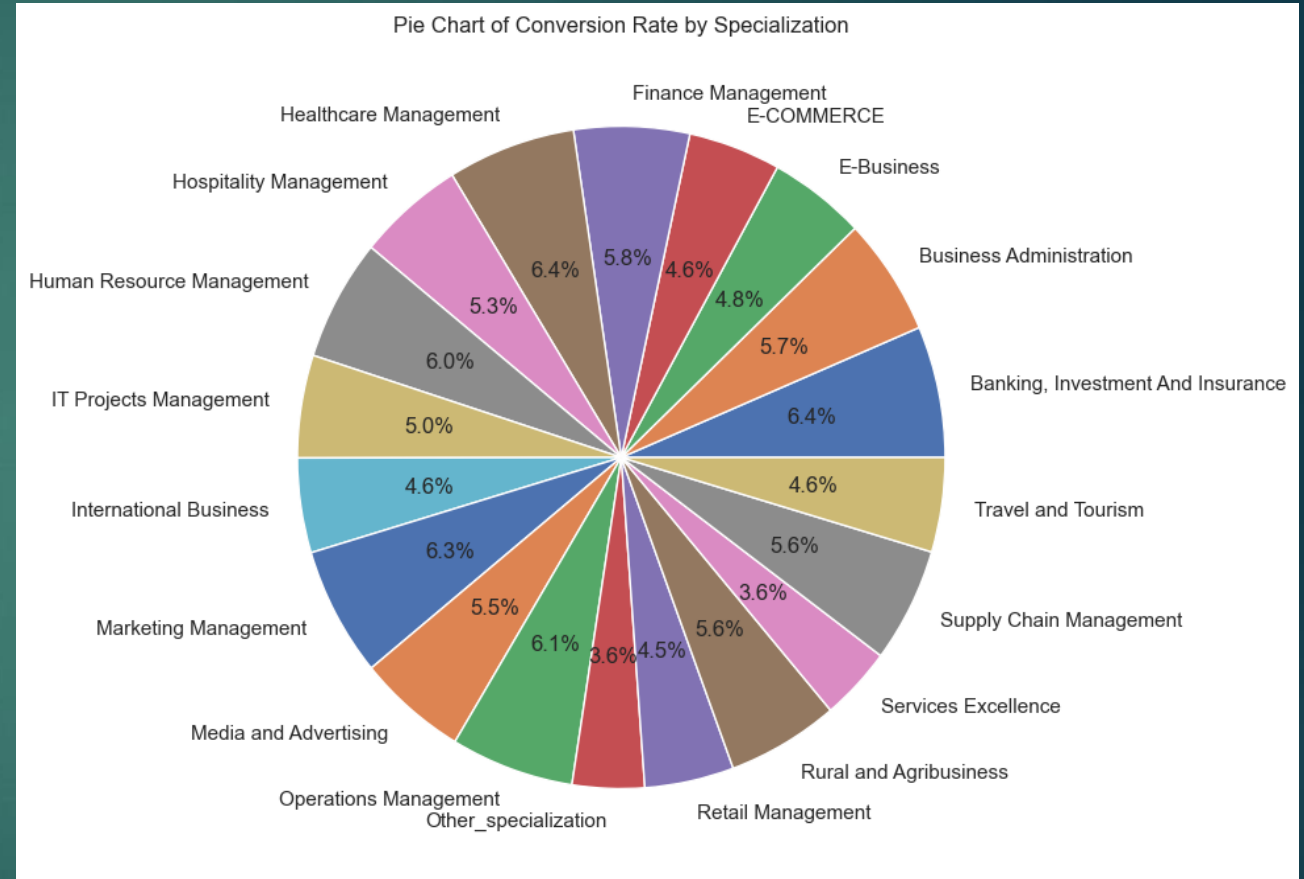
Business Impact

- ▶ **Enhanced Lead Prioritization:** By accurately identifying Hot Leads, the model allows sales teams to focus efforts on prospects with higher conversion potential, improving overall sales efficiency.
- ▶ **Strategic Resource Allocation:** Insights from key features enable targeted marketing campaigns and personalized engagement strategies, maximizing ROI on marketing investments.
- ▶ **Improved Decision-Making:** Operational tools for predicting conversion probabilities and adjusting cutoff points empower strategic decision-making, facilitating continuous improvement in lead conversion strategies.

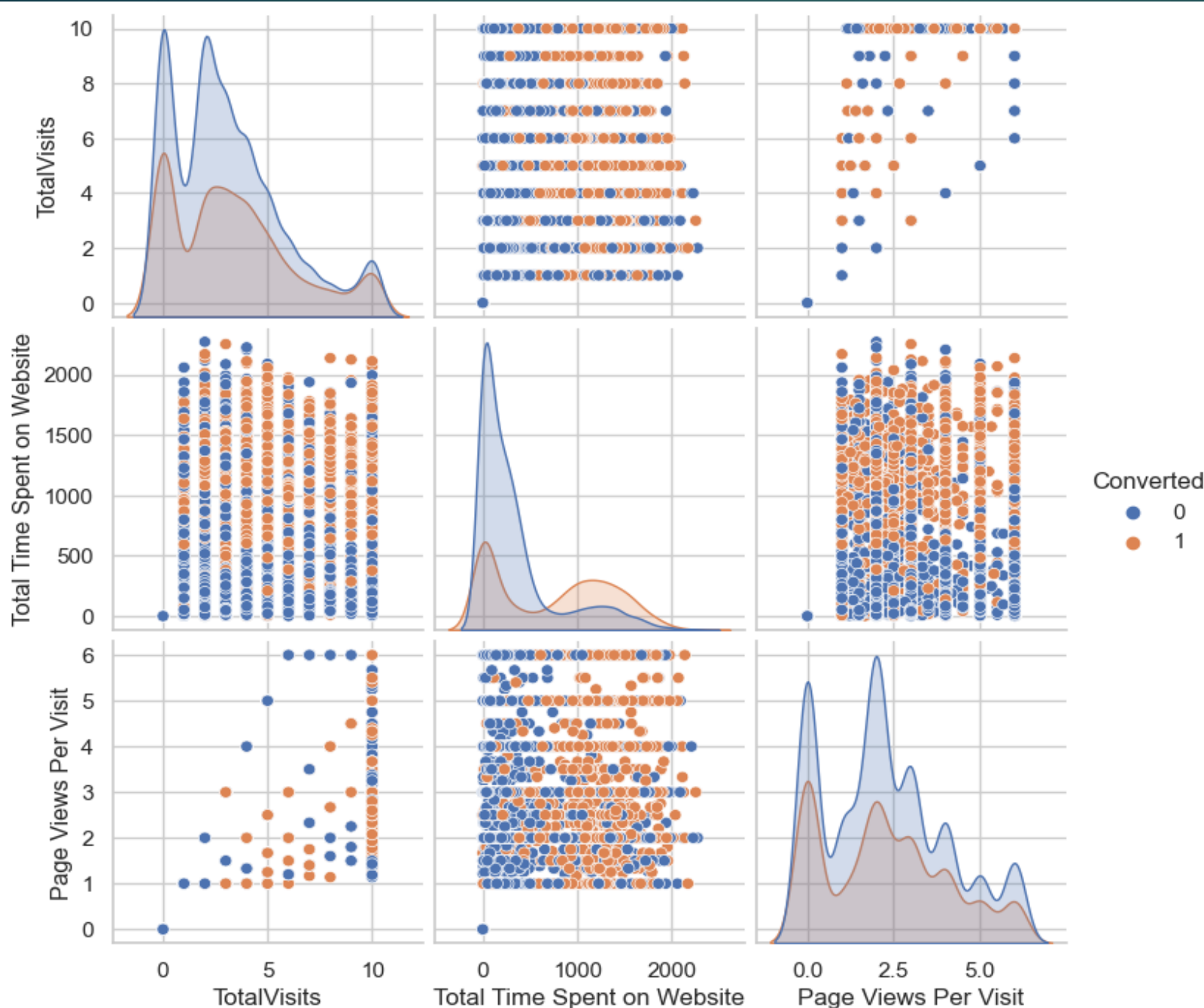
Visualizations and Key Insights



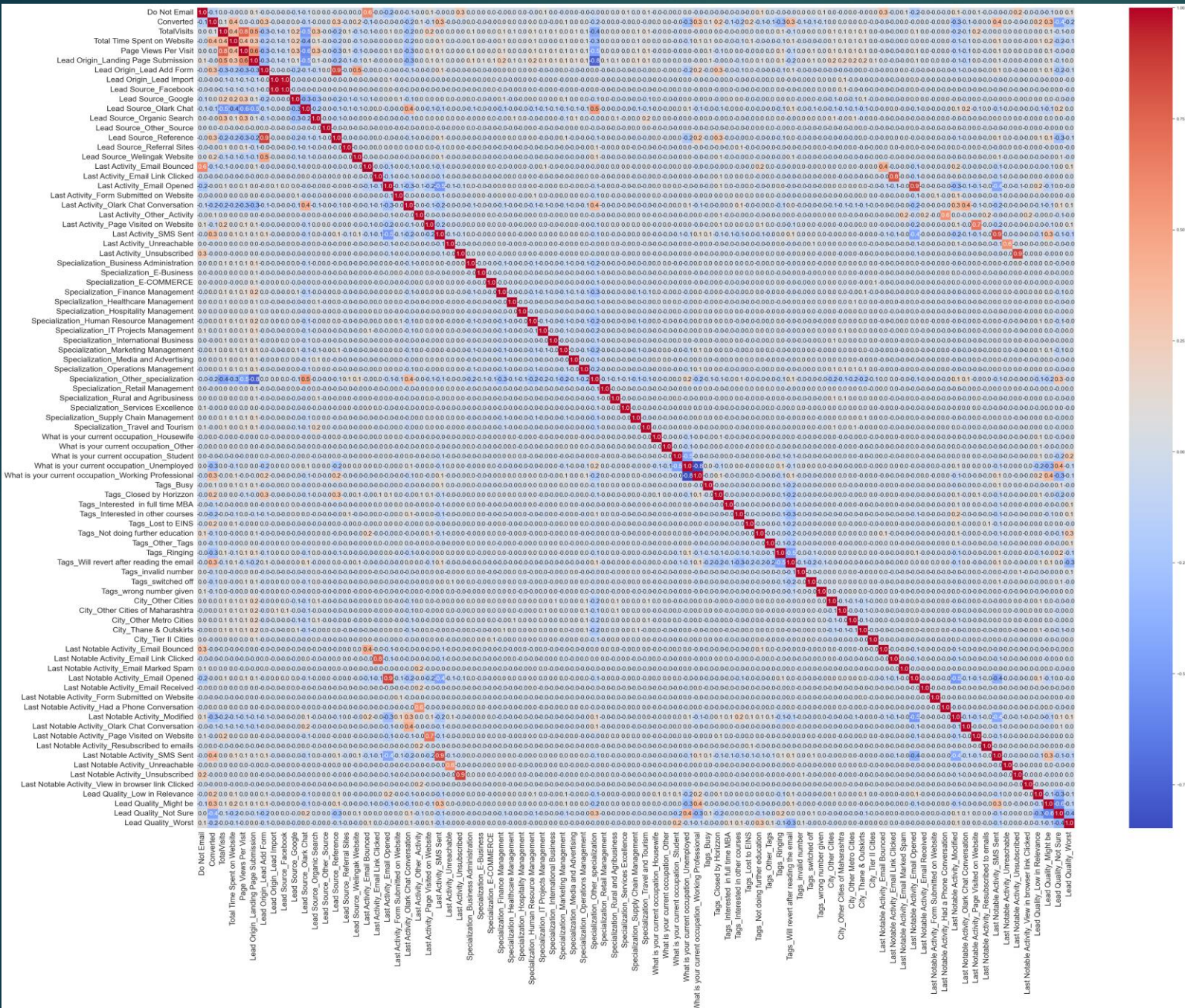
The bar graph shows that '**Lead Add Form**' has the highest conversion rate at over **0.7**, indicating it's the most effective lead origin for conversions.



The pie chart indicates that **E-Business** specialization leads have the highest conversion rate at **6.9%**, suggesting a targeted approach in this area could enhance overall lead conversions.

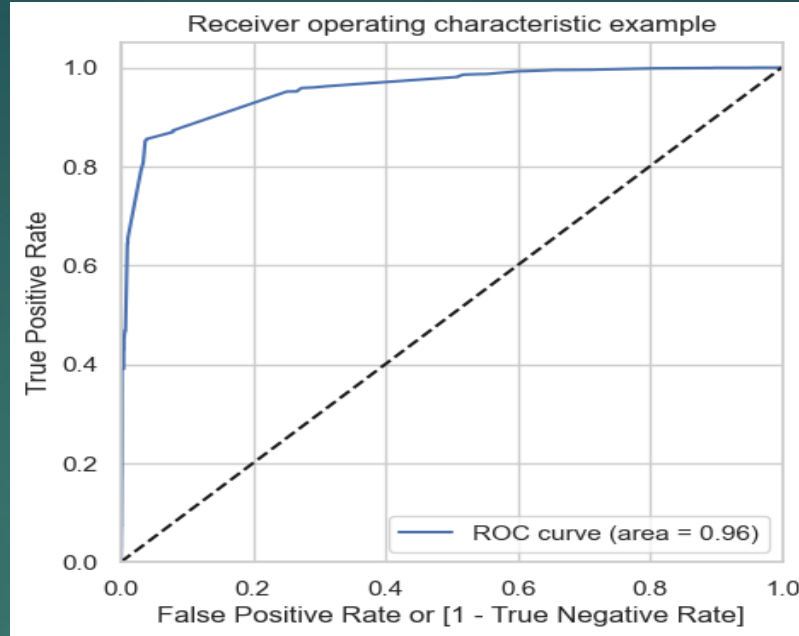


- **Total Visits:** There's a positive correlation between the number of visits and conversions, suggesting frequent visits increase the likelihood of conversion.
- **Time on Website:** Longer engagement on the website is strongly associated with higher conversion rates, indicating the value of an immersive user experience.
- **Page Views Per Visit:** A moderate number of page views per visit seems to be optimal for conversion, as too few may indicate disinterest and too many could suggest difficulty in finding relevant content.
- **Conversion Clusters:** There are distinct clusters of conversions at lower total visits but higher page views and time spent, which could imply that effectively targeted content can lead to quicker conversions.

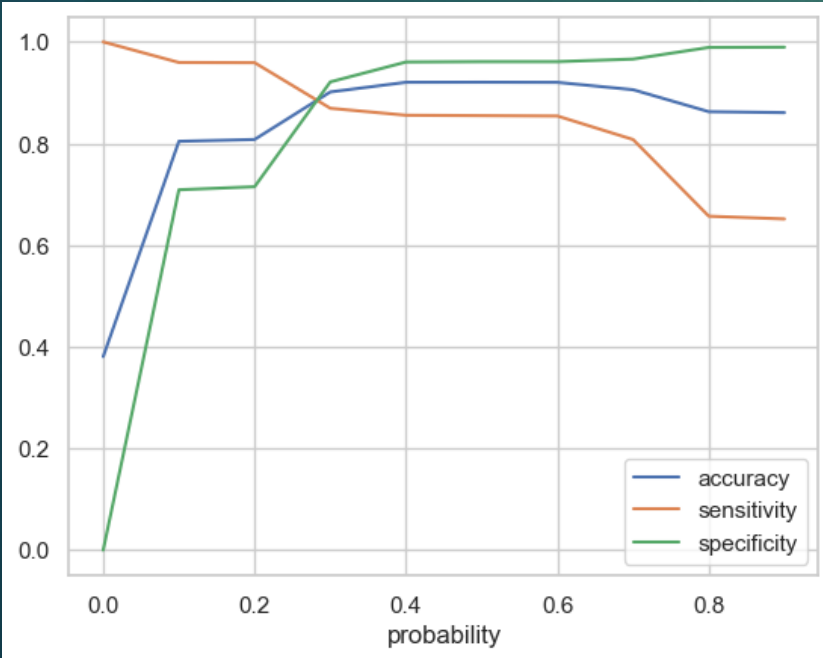


The heatmap indicates strong correlations along the diagonal, suggesting each gene has a perfect positive correlation with itself. Clusters of high correlation between different genes may imply potential functional relationships, while areas of low correlation could indicate independent operation or involvement in different biological pathways.

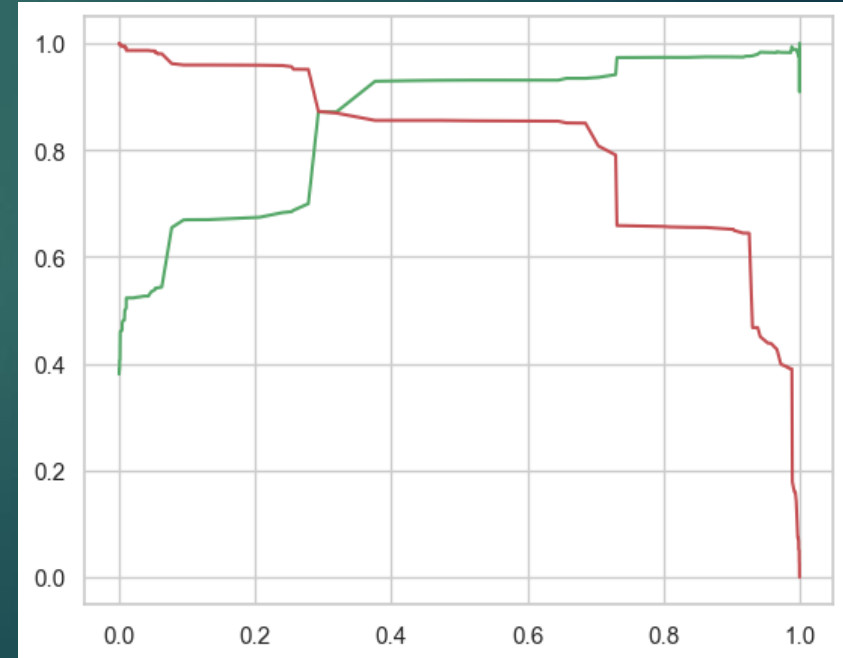
The line graph indicates that as the probability threshold for predicting lead conversion increases, accuracy and specificity improve, suggesting a more precise targeting of potential leads, while sensitivity slightly decreases, pointing to a trade-off in identifying true positives.



The graph illustrates that by strategically adjusting our criteria for lead qualification, we can balance and enhance lead engagement and conversion efforts, ultimately impacting our sales performance positively.



The ROC curve shows that our predictive model is highly effective at identifying potential leads, ensuring that we target the right individuals with a high chance of conversion, which can optimize our marketing efforts and resources.



Conclusion

Through comprehensive analysis and strategic insights, our study highlights key factors influencing lead conversions for X Education:

- ▶ **Lead Origin and Specialization:** Leads from the 'Lead Add Form' and those interested in E-Business specialization show the highest conversion rates, highlighting the importance of targeted marketing in these areas.
- ▶ **User Engagement:** Increased time spent on the website and moderate page views per visit positively influence conversion rates, emphasizing the need for engaging user experiences and relevant content.
- ▶ **Model Effectiveness:** Using logistic regression, we developed a predictive model with a 0.3 cutoff point to identify "Hot Leads" effectively. This approach optimizes resource allocation by focusing efforts on leads with the highest conversion potential.
- ▶ **Operational Impact:** Our implementation includes a flexible prediction framework with adjustable cutoff points, facilitating adaptive strategies to maximize conversion rates and optimize resource utilization.

In conclusion, by leveraging these insights, X Education can enhance its lead conversion strategies, improve sales efficiency, and achieve sustainable business growth through targeted and data-driven decision-making.



Thank You