Lead Scoring Model for X Education

IMPROVING LEAD CONVERSION RATES

Team Members : Amrit Saikia, Divya Reddy, Geo Abraham

Problem Statement

- X Education, an online learning platform, is currently facing a challenge with lead conversion rates hovering around 30%.
- The organization acquires leads through a mix of marketing channels and referrals from previous customers.
- Leads are categorized once prospective clients fill out a form on the company’s website.
- The primary objective is to boost the lead conversion rate to about 80%.
- The initiative aims to identify "Hot Leads" that demonstrate a higher likelihood of converting, thereby streamlining the sales team's efforts.
- To support this goal, a logistic regression model will be developed to generate lead scores (from 0 to 100) for effective prioritization of potential leads.
- The available dataset contains approximately 9,000 data entries, featuring variables such as Lead Source, Time Spent on Website, and Last Activity.

Analysis Approach

- Data Preparation and Exploration: The dataset was imported and preprocessed by managing categorical variables and addressing missing values represented as 'Select.' Outliers were examined and handled, and duplicate entries were checked and removed.
- Visualization: Various plots were utilized to investigate data distributions and examine relationships among features.
- Model Development: Features were engineered, and dummy variables were created. The data was then divided into training and testing subsets, followed by the application of feature scaling and the initialization of a logistic regression model.
- Model Building: Recursive Feature Elimination (RFE) was employed for feature selection, and models were assessed using metrics such as Accuracy and ROC-AUC.
- Refinement: The models were fine-tuned to enhance performance, and optimal cutoff points for making predictions were determined.

Results and Findings

- Model Capability Overview
  - The finalized logistic regression model effectively predicts lead conversions by examining critical features such as lead origin, source, activity, and tags. This capability facilitates efficient resource allocation and the implementation of targeted sales strategies.
- Cutoff Point Selection
  - By setting a cutoff point of 0.3, the model classifies leads into two categories: "Hot Leads," which have a high likelihood of conversion, and "Cold Leads," which exhibit lower potential. This strategy enhances lead prioritization and reduces the occurrence of false predictions.
- Key Features Influencing Conversion
  - A total of 13 significant features were identified as having a substantial impact on the likelihood of conversion. Notable positive influencers include Tags_Lost to EINS, Tags_Closed by Horizzon, Lead Source_Welingak Website, whereas negative influencers consist of Lead Quality_Worst, Lead Quality_Not Sure, and Tags_switched off,

Model Performance and Implementation

- Model Performance Metrics
  - The model attained an accuracy of 90%, signifying that 90% of cases were classified correctly. It has a sensitivity of 87%, which ensures that 87% of converting customers are accurately identified. Additionally, it maintains a precision of 88% for correctly identifying Hot Leads.

- Operational Implementation
  - A reusable code block was developed to facilitate predictions of conversion probabilities and lead scores. This code includes adjustable cutoff points, allowing for the optimization of strategies based on specific performance metrics such as sensitivity and precision, thereby enhancing the organization's decision-making capabilities.

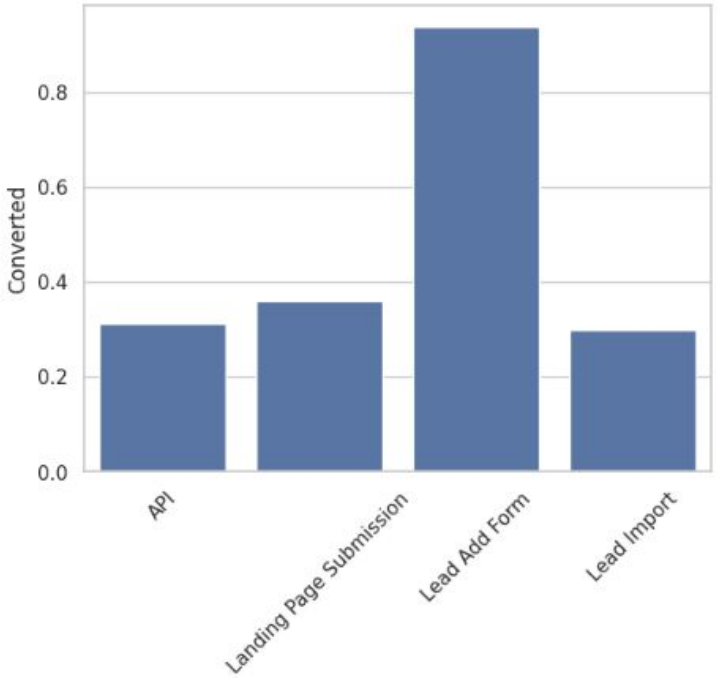
Business Impact

- Enhanced Lead Prioritization: The model enables accurate identification of Hot Leads, allowing sales teams to concentrate their efforts on prospects with greater conversion potential, thereby enhancing overall sales efficiency.

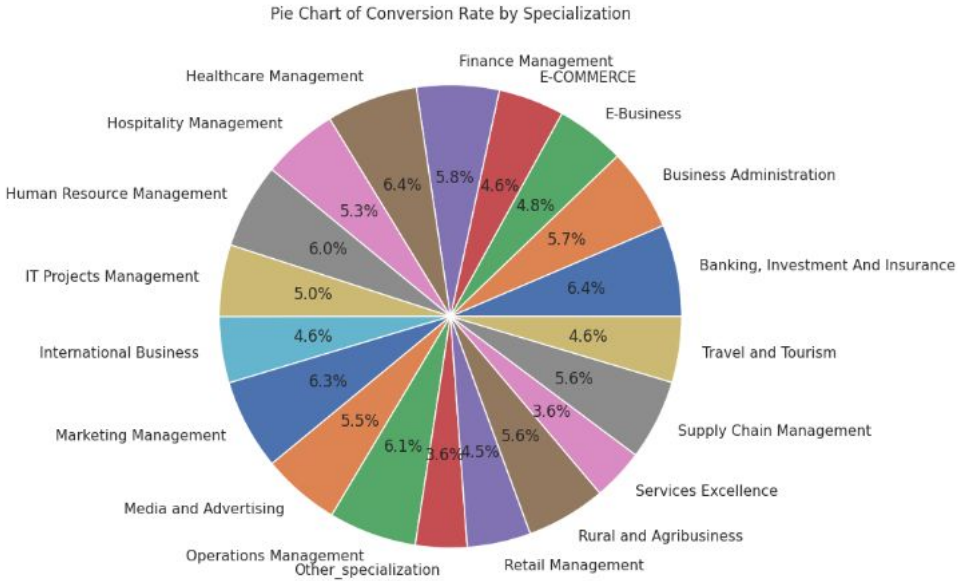
- Strategic Resource Allocation: Insights derived from key features support targeted marketing campaigns and personalized engagement strategies, optimizing the return on investment (ROI) for marketing efforts.

- Improved Decision-Making: The operational tools designed for predicting conversion probabilities and adjusting cutoff points provide a foundation for strategic decision-making, promoting ongoing improvements in lead conversion strategies.

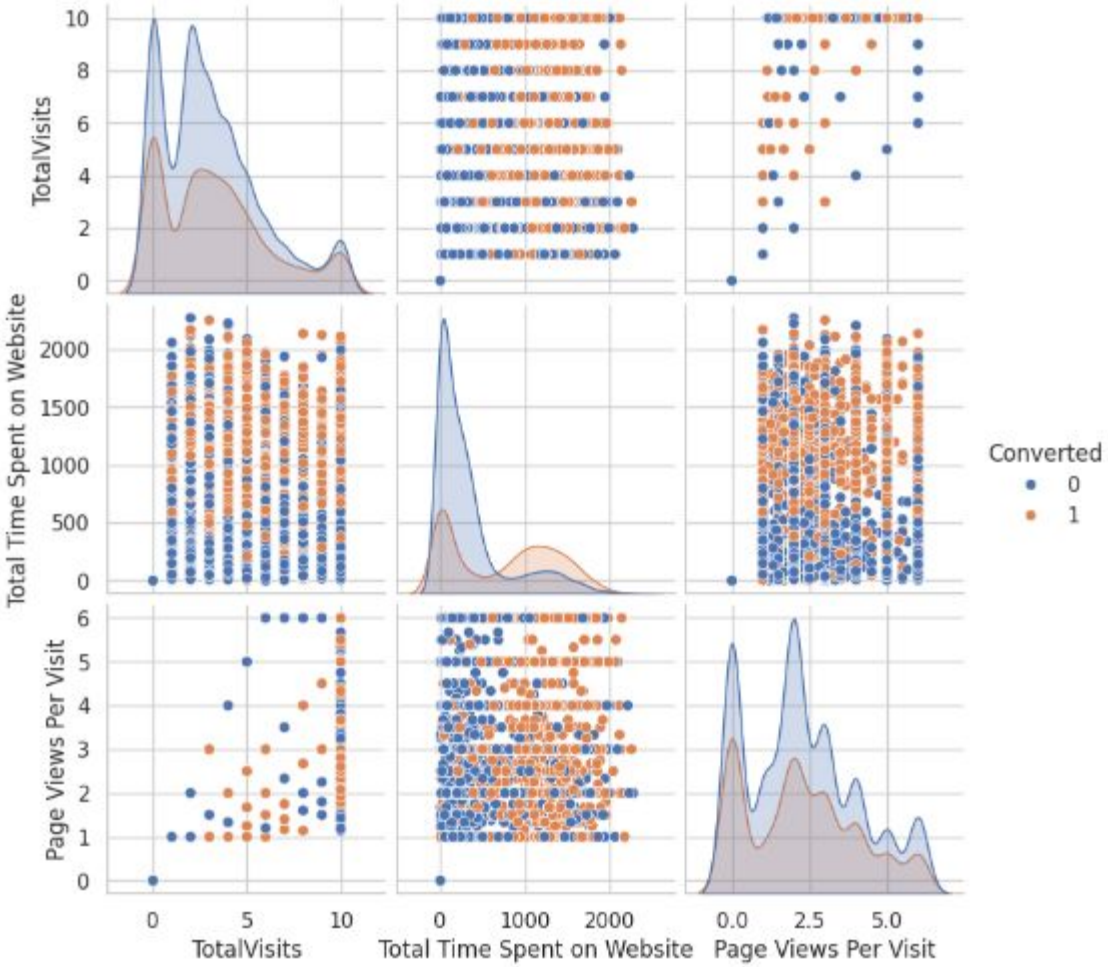
Key Insights



‘Lead Add Form’ has the highest conversion rate, exceeding 0.7, indicating that it is the most effective source of leads for conversions.



Leads specializing in E-Business have the highest conversion rate at 6.9%, suggesting that a focused strategy in this domain could improve overall lead conversions.

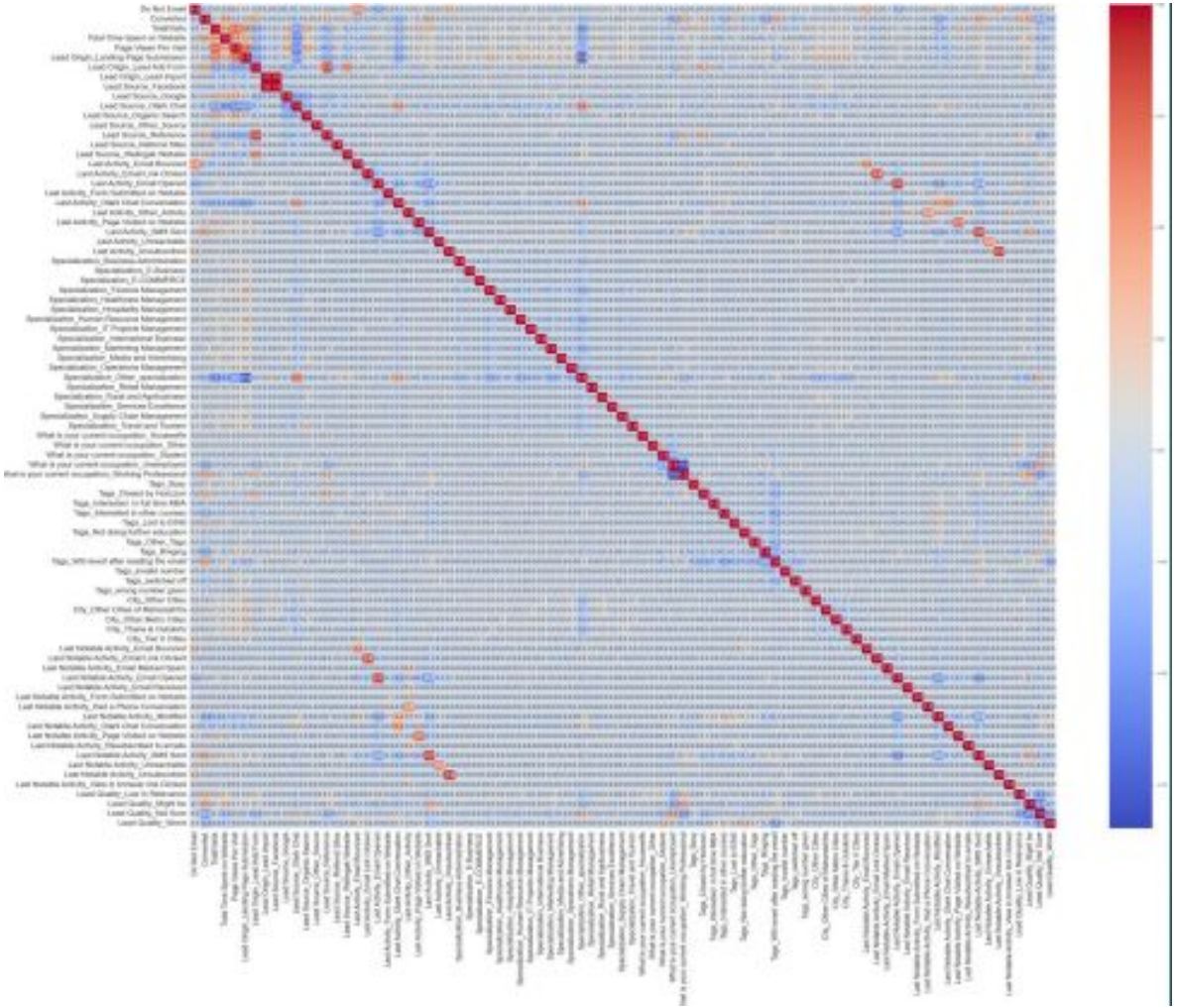


Total Visits: A positive correlation suggests that increased visits enhance conversion likelihood.

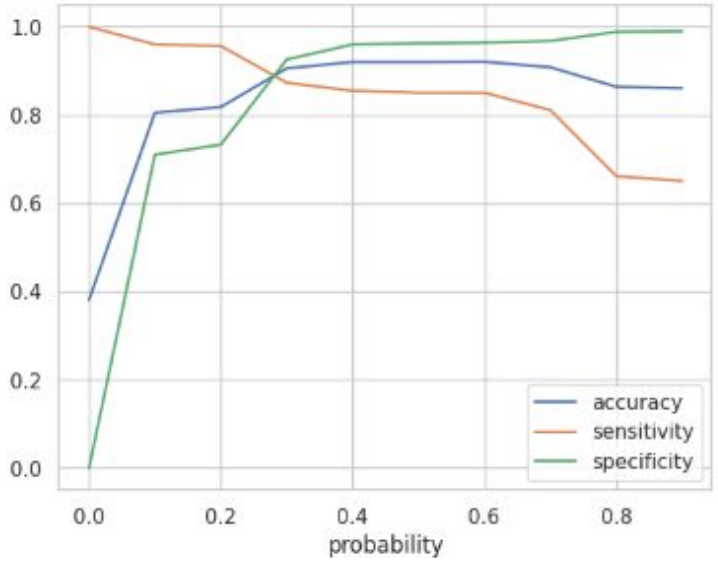
Time on Website: Longer engagement is strongly associated with higher conversion rates, emphasizing the need for an immersive user experience.

Page Views Per Visit: An optimal number of page views facilitates conversions; too few may indicate disinterest, while too many could reflect difficulty finding relevant content.

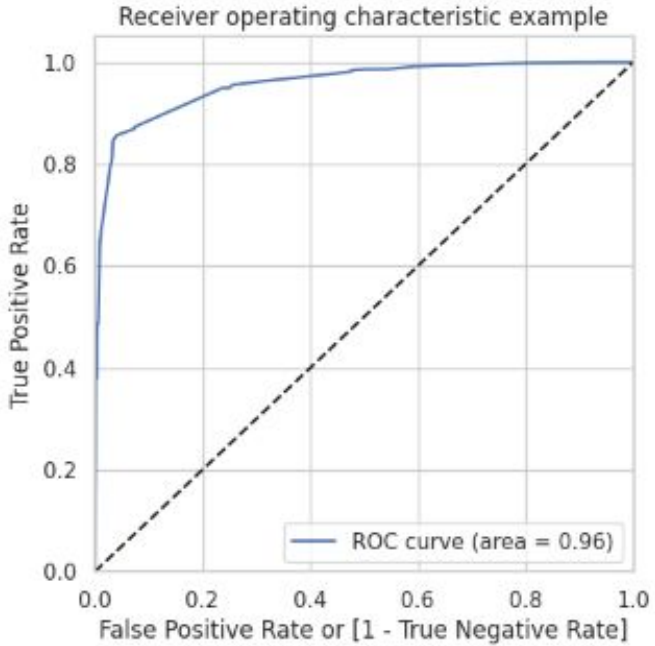
Conversion Clusters: Notable clusters show that lower total visits combined with higher page views and time spent imply that targeted content can accelerate conversions.



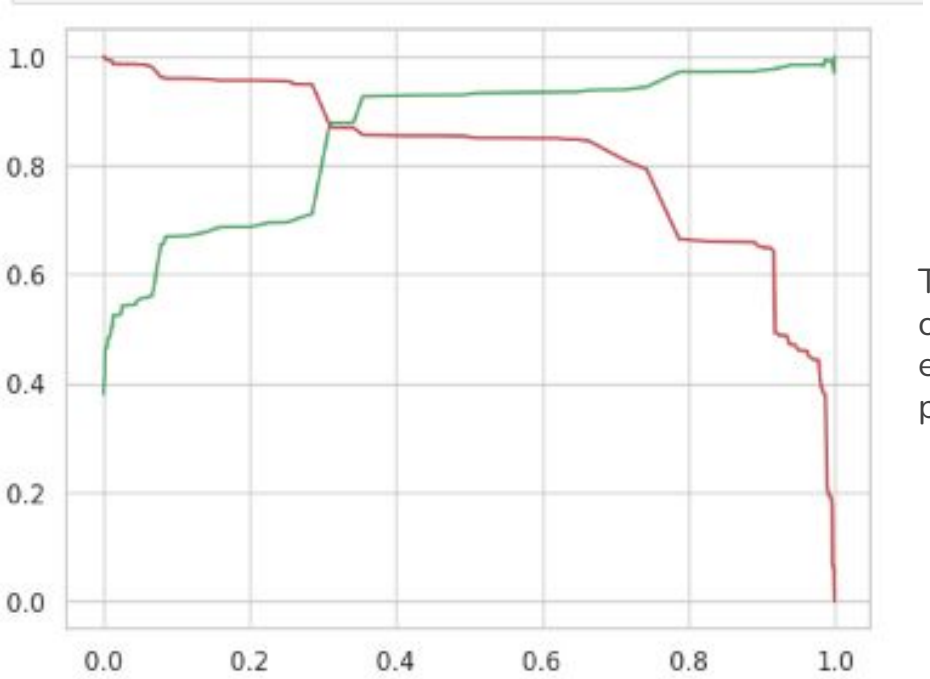
The heatmap displays strong diagonal correlations, indicating that each gene is perfectly correlated with itself. Clusters of high correlation among different genes may point to potential functional connections, while regions of low correlation could suggest independent roles or involvement in separate biological pathways.



The line graph demonstrates that as the probability threshold for predicting lead conversions rises, both accuracy and specificity enhance, indicating more precise targeting of potential leads. However, sensitivity shows a slight decline, highlighting a trade-off in identifying true positives.



The ROC curve indicates that our predictive model is very effective in identifying potential leads, ensuring that we focus on individuals with a high likelihood of conversion. This can enhance the efficiency of our marketing efforts and resource allocation.



The graph demonstrates that by carefully refining our lead qualification criteria, we can optimize both engagement and conversion efforts, leading to a positive impact on overall sales performance.

Conclusion

Our study identifies critical factors that influence lead conversions for X Education:

- Lead Origin and Specialization: Leads generated through the 'Lead Add Form' and those expressing interest in E-Business specialization exhibit the highest conversion rates, underscoring the significance of targeted marketing efforts in these areas.
- User Engagement: Longer time spent on the website and a moderate number of page views per visit are positively correlated with conversion rates, highlighting the importance of providing engaging user experiences and relevant content.
- Model Effectiveness: We developed a predictive model using logistic regression, with a cutoff point of 0.3 to effectively identify "Hot Leads." This method allows for optimized resource allocation by concentrating on leads with the greatest potential for conversion.
- Operational Impact: Our implementation features a flexible prediction framework with adjustable cutoff points, enabling adaptive strategies to maximize conversion rates and enhance resource utilization.

In summary, by utilizing these insights, X Education can refine its lead conversion strategies, boost sales efficiency, and foster sustainable business growth through targeted and data-driven decision-making.