

Lead-Scoring Case Study

Introduction:

Our group tackled the challenge of optimizing lead conversion for X Education, focusing on identifying 'Hot Leads' through a precision-oriented machine learning model. The report outlines our strategic approach, data preprocessing decisions, key model metrics, and the learnings garnered from the assignment.

Data Preprocessing:

We initiated the process by addressing data quality. Columns with significant null values, such as 'Lead Quality' and 'Asymmetrique' attributes, were removed to maintain model integrity. Imputing these records could introduce biases. Additionally, we dropped columns like 'Country' and 'What matters most to you in choosing a course' due to imbalance and lack of relevance. 'Tags' was excluded to prevent potential misalignment with real-world lead actions.

Machine Learning Approach:

Our machine learning strategy centered on simplicity and interpretability. Logistic regression emerged as the primary model, supported by Recursive Feature Elimination (RFE) for effective feature selection. The model was trained and evaluated, with the threshold selection identified at 0.35. This threshold, where precision, recall, and F1-score intersected, became pivotal for our precision-oriented strategy.

Model Evaluation:

The evaluation metrics provided a comprehensive snapshot of model performance. While accuracy reached 71%, precision soared to 94%, indicating low false positives. However, recall stood at 27%, highlighting the precision-recall trade-off. The F1-score, a balanced measure, settled at 0.42. The chosen threshold of 0.35 aligned with our precision-focused strategy, emphasizing targeted outreach.

Recommendations and Implementation:

Our recommendations emphasized actionable insights for the sales team. We advocated for the strategic targeting of 'Hot Leads' based on model predictions, supported by continuous monitoring for dynamic adjustments. The implementation plan included training sessions, effective communication, and documentation to ensure seamless integration into existing processes.

Challenges and Mitigations:

Addressing challenges played a crucial role in our approach. Threshold sensitivity and model interpretability were acknowledged, with mitigations involving dynamic threshold optimization and effective communication of complex model outcomes. The collaborative approach between data science and sales teams emerged as a key success factor.

Conclusion and Future Steps:

In conclusion, our journey encapsulated data-driven decision-making and strategic model development. The precision-focused strategy aligned seamlessly with business goals, and interdisciplinary collaboration played a crucial role in overcoming challenges. Future steps

involve iterative optimization, scaling impact, and embracing an agile mindset for continuous improvement.

Key Learnings:

Our learnings encompassed understanding the precision-recall trade-off, the significance of dynamic threshold adjustments, the importance of interdisciplinary collaboration, the need for continuous monitoring, and the art of strategic model development in alignment with business objectives.

In summary, the assignment provided valuable insights into the practical application of machine learning in a business context, emphasizing precision, collaboration, and continuous improvement

Bhagabati Prasad Mishra
Swamy Bore
Krishna Prasanth