

# Lead Scoring Assignment

## Summary Report



**upGrad**  
RAHO AMBITIOUS



**Executive  
PG Program  
in Data Science**



**Team Members:**

Lavan Verma

Varun Seth

---

## Project Overview and Problem Statement

X Education faced a key issue with lead conversion inefficiency. We aimed to tackle this by creating a predictive model to score leads, pinpointing 'Hot Leads' with greater conversion potential. Our goal: achieve an 80% conversion rate, closing the gap between lead generation and conversion significantly.

## Methodology and Execution

Our approach was data-centric, involving several crucial steps:

- **Cleaning and preparing data:** Handled various kind of missing values, and clubbed some categorical values, removed low-variation columns, handled outliers.
- **Conducted exploratory data analysis (EDA):** Visualized the distributions of each feature, compared distribution across conversion, examined conversion rates in different categories, and performed correlation analyses.
- **Model Building:** Created dummy variables, scaled the data, split between test/train (80%, 20%), used RFE (20 columns), iteratively removed non-significant columns ( $pvalue > 0.05$ ) and reduced multi collinearity (high VIF  $> 5$ ). Further tuned the threshold using ROC analysis.
- **Model Evaluation:** Training accuracy is 94% and test accuracy 91%. Same values for Sensitivity and specificity.
- **Predictive modeling:** Predicted conversion likelihood and analyzed feature importance.

## Key Data-Driven Insights and Learnings

- Engagement as a Conversion Indicator:
  - Analysis indicates: Longer website visits correlate with higher conversion odds. Implying, compelling content and seamless website navigation are vital for boosting conversions.
  - Many outliers observed: Potentially interested prospects who didn't convert, perhaps opting for a competitor's solution. Emphasizing the need for prompt, impactful engagement with promising leads.
- Predictive Modeling Efficacy:
  - We constructed a logistic regression model to forecast the likelihood of lead conversion with remarkable accuracy, reflected by an AUC of 0.98 on the ROC curve, signifying outstanding discriminatory ability between converted and non-converted leads. Moreover, both precision and recall metrics surpassed 0.90, affirming the model's dependability in identifying true positives and reducing false negatives.

- Optimal Lead Scoring:
  - The model generates a lead score indicating conversion likelihood, aiding lead prioritization for targeted engagement. Establishing a probability threshold categorizes leads as 'converted' or 'not converted,' facilitating resource allocation to high-potential leads efficiently.
- Conversion Rate Improvement:
  - The primary goal was to enhance the conversion rate, aiming for 80%. Model insights guide the company toward this goal by pinpointing and prioritizing high-potential leads.
- Impact of Last Notable Activity:
  - The analysis indicates that actions like 'SMS Sent' and 'Email Opened' strongly correlate with conversion rates. SMS, due to its direct and personalized nature, elicits swift responses, emerging as the most influential factor driving conversions.

## Strategic Recommendations

- Maximize Engagement Focus: Utilize the lead scores generated by the model to prioritize engagement with leads that have a higher probability of conversion.
- Continual Monitoring and Model Refinement: Maintain model performance by continually assessing and refining it in response to new data and market changes.
- Expand Data Exploration: Explore more data points and external variables to enhance model precision and adjust marketing strategies accordingly.
- Deploy A/B Testing: Conduct A/B tests to measure the effectiveness of lead prioritization strategies in real-world scenarios.
- Improve Data Integrity: Invest in improving data collection processes to ensure high-quality inputs for the predictive model.

---

## Conclusion

The Lead Scoring Assignment represents a significant advancement in our quest to optimize lead conversion. Through a rigorous, data-driven approach, we have uncovered valuable insights that pave the way for more targeted and efficient marketing strategies, ensuring that our resources are invested in leads with the highest likelihood of conversion, thereby driving business growth and operational efficiency.