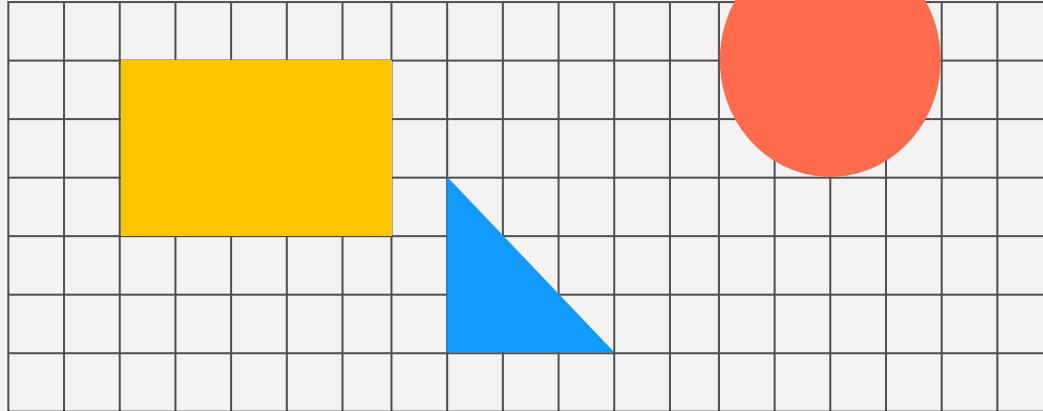# LEAD SCORING

- By : Ashwini H V, Deepanjali Bhatt, Shreya Singh
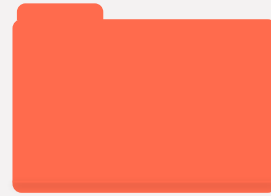
Problem Statement

Summary
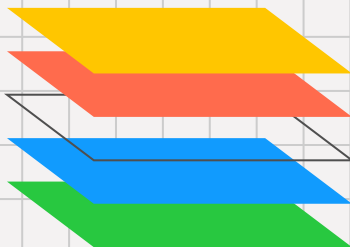
Approach and Methodology

Expectations , Inference, Insights and Recommendations

Problem Statement

Overview:

An X Education need help to select the most promising leads, i.e. the leads that are most likely to convert into paying customers. The company requires us to build a model wherein you need to assign a lead score to each of the leads such that the customers with higher lead score have a higher conversion chance and the customers with lower lead score have a lower conversion chance. The CEO, in particular, has given a ballpark of the target lead conversion rate to be around 80%.

Goals

Build a logistic regression model to assign a lead score between 0 and 100 to each of the leads which can be used by the company to target potential leads. A higher score would mean that the lead is hot, i.e. is most likely to convert whereas a lower score would mean that the lead is cold and will mostly not get converted.

Learning Outcome

Summary

- We have evaluated both Sensitivity-Specificity and Precision-Recall metrics, but the final prediction was based on the optimal cutoff determined by Sensitivity and Specificity.
- The Accuracy, Sensitivity, and Specificity values for the test set are approximately 77%, 83%, and 74%, which are close to the respective values obtained from the training set.
- Additionally, the lead score calculated from the training set indicates that the conversion rate for the final predicted model is around 80%.
- Therefore, overall, this model appears to be effective.

Key features contributing to a high conversion rate or those that influence the probability of a lead converting are:

- Lead Origin_Lead Add Form
- What is your current occupation_Working Professional
- Total Time Spent on Website

Approach and Methodology

Data Acquisition

• Data Sourcing
• Data Description

Data Preparation

• Data Cleaning
• Dropping rows/columns that are completely null
• Missing values treatment
• Visualizing
• Checking outliers
• Filtering data
• Creating dummy variables
• Test-train split

Model Building using RFE

• Dropping model with high p value
• Visualizing summary of each model
• Checking VIF values for feature variables
• Creating new data frame containing feature variables and their respective VIFs
• Getting predicted values on train dataset
• Confusion matrix
• Checking overall accuracy
• Calculating specificity
• Plotting ROC curve & finding optimal cutoff
• Calculating and plotting sensitivity
• Prediction on the train set
• Assigning lead score
• Inference and conclusion

Expectations

Do's

- **Prioritize High-Scoring Leads:** Focus on leads from the "Lead Add Form" and working professionals, as they have a higher chance of conversion.
- **Enhance Sales & Marketing Strategies:** Use automated follow-ups, A/B testing on landing pages, and remarketing to improve engagement and conversion rates.
- **Continuously Improve the Lead Scoring Model:** Regularly update the model with new data, fine-tune probability thresholds, and monitor key performance metrics.
- **Leverage Behavioral Insights:** Leads who spend more time on the website, watch videos, or engage with content should be given priority in follow-ups.

Don'ts

- **Don't Treat All Leads Equally:** Avoid wasting time on low-quality leads; prioritize based on data-driven insights.
- **Don't Ignore Data Quality Issues:** Handle missing values, clean categorical variables (e.g., "Select"), and ensure data consistency for accurate predictions.
- **Don't Rely Solely on the Model:** While the lead score is valuable, human judgment and personalized follow-ups are essential for improving conversions.
- **Don't Assume Model Performance is Static:** Regularly retrain the model and adjust strategies to adapt to changes in customer behavior and market trends.

Inference & Insights

Topic	Inference	Insight
Model Performance	Accuracy: **77.52%**, Sensitivity: **83.01%**, Specificity: **74.13%** indicate a well-performing model that can effectively identify potential conversions.	The model significantly improves conversion efficiency over the **baseline rate of ~30%**, making it a valuable tool for lead prioritization.
Sensitivity & Specificity Trade-off	The model identifies **83.01%** of actual conversions but still misclassifies **25.87%** of non-converting leads.	Prioritizing sensitivity ensures that fewer potential leads are missed, but sales teams should verify borderline cases manually to optimize efforts.
Lead Origin -"Lead Add Form"	Leads who fill out the **Lead Add Form** have a **high conversion probability**, showing a strong intent to purchase.	Immediate follow-ups should be prioritized for these leads to capitalize on their high engagement.
Current Occupation - Working Professionals	Leads who are **working professionals** are more likely to convert, as they seek career advancement opportunities.	Marketing efforts should highlight **career benefits, skill enhancement, and certifications** to appeal to this audience.

Topic	**Inference**	**Insight**
Total Time Spent on Website	The more time a lead spends on the website, the higher the likelihood of conversion.	Implement **chatbots, real-time engagement tools, or time-based pop-ups** to capture interest and drive conversions.
Lead Conversion Funnel Efficiency	Although many leads enter the funnel, **only a small fraction converts** without additional engagement.	Use the model to **narrow down high-potential leads**, allowing the sales team to focus their efforts more efficiently.
Drop-off Points in User Journey	Many users visit the website but leave without taking action.	mplement **retargeting strategies, email follow-ups, and exit-intent pop-ups** to re-engage these users.
Effectiveness of Lead Sources	Not all marketing sources generate high-converting leads—**organic referrals and past customer recommendations perform better than paid ads**.	Reallocate marketing budgets towards **high-converting sources**, reducing spending on low-quality lead generation channels.

Recommendations

Recommendations for Improving Lead Conversion at X Education. Based on the model's performance and insights, here are some actionable recommendations

1. Prioritize Leads Based on Key Features

- Leads generated from **"Lead Origin: Lead Add Form"** have a higher probability of conversion. The sales team should focus more on these leads.
- Working professionals are more likely to convert. Personalize marketing efforts and follow-ups for this segment.
- Leads spending **more time on the website** show higher intent. Implement retargeting strategies for these visitors.

- Immediate Follow-Up
- Personalized Communication
- Segmented Email Campaign
- Retargeting Ads

Tips

2. Optimize Lead Scoring & Cutoff Selection

- The current model achieves an accuracy of **77.52%** with **83.01% sensitivity**. Fine-tuning the cutoff threshold can help prioritize leads more effectively.
- Since the goal is **80% conversion**, focus on precision-recall trade-offs to maximize the number of converted leads.

3. Improve Engagement Strategies

- **Follow-Up Priority**: Leads with high engagement (e.g., those who watch videos or browse multiple pages) should receive **immediate follow-ups**.
- **Personalized Communication**: Emails and calls should emphasize the benefits specific to **working professionals**, as they are a key converting segment.
- **Chatbots and Live Assistance**: Implement real-time support to guide high-intent visitors towards form submissions.

- Experiment with Different Cutoff Values
- Use Business Metrics for Decision-Making
- Segmented Email Campaign
- Limited-Time Offers

Tips

4. Refine Marketing & Lead Acquisition Strategy

- Invest more in channels that generate high-converting leads (e.g., improve targeting on Google Ads).
- Reduce marketing spend on **low-quality leads** (e.g., those who fill out forms but spend minimal time on the website).
- A/B test landing pages and calls-to-action (CTAs) to encourage longer website engagement.

5. Address Data & Operational Improvements

- **Handle "Select" values in categorical fields**, as they might introduce bias or reduce model performance.
- Conduct **regular model retraining** to ensure the lead-scoring system stays relevant with changing user behavior.
- Provide **real-time lead scores** to the sales team so they can prioritize outreach efforts dynamically.

- Optimize Targeting
- Exclude Negative Keywords
- Focus on Quality over Quantity
- Automate Alerts
- Review Data Quality Regularly

Tips

Thank You!

Thank you for reviewing our presentation. We appreciate the opportunity to share our work and insights. If you have any questions or require further clarification, please feel free to reach out to us.

We look forward to your feedback and hope you find our analysis valuable.

Ashwini H V

ashwinipushkarrr@gmail.com

Deepanjali Bhatt

bhattdeepanjali123@gmail.com

Shreya Singh

bisen11shreya@gmail.com