

Lead Score Assignment Analysis

Problem Statement:

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses.

The company markets its courses on several websites, search engines, and even social media sometimes. Once these people land on the website, they might browse the courses, fill out a form for the course, or watch some videos. When these people fill out a form with their email address or phone number, they are classified as leads. Moreover, the company also gets leads through past referrals. Once these leads are acquired, employees from the sales team start making calls, writing emails, etc. Through this process, some of the leads get converted into successful sales, while most of the leads do not. The typical lead to successful sale conversion rate at X education is around 30%.

Analysis Approach

- Data Collection: Gathering relevant lead data and understanding the data.
- Data Preprocessing: Cleaning, transforming, and preparing data for analysis.
- Feature Selection: Identifying key variables influencing lead conversion.
- Model Development: Building a predictive model for lead scoring.
- Validation & Deployment: Testing the model's performance and integrating it into business processes.

Results Overview:

Let's evaluate the metrics as well

TP = confusion[1,1] # true positive

TN = confusion[0,0] # true negatives

FP = confusion[0,1] # false positives

FN = confusion[1,0] # false negatives

Results :

1) Comparing the values obtained for Train & Test:

Train Data:

Sensitivity : 88.8 %

Specificity : 89.4 %

Test Data:

Sensitivity : 86.9 %

Specificity : 88.8 %

Precision-Recall

- **Precision**

$$TP / TP + FP$$

$$\text{confusion}[1,1]/(\text{confusion}[0,1]+\text{confusion}[1,1]) = 88.5\%$$

- **Recall**

$$TP / TP + FN$$

$$\text{confusion}[1,1]/(\text{confusion}[1,0]+\text{confusion}[1,1]) = 84.5\%$$

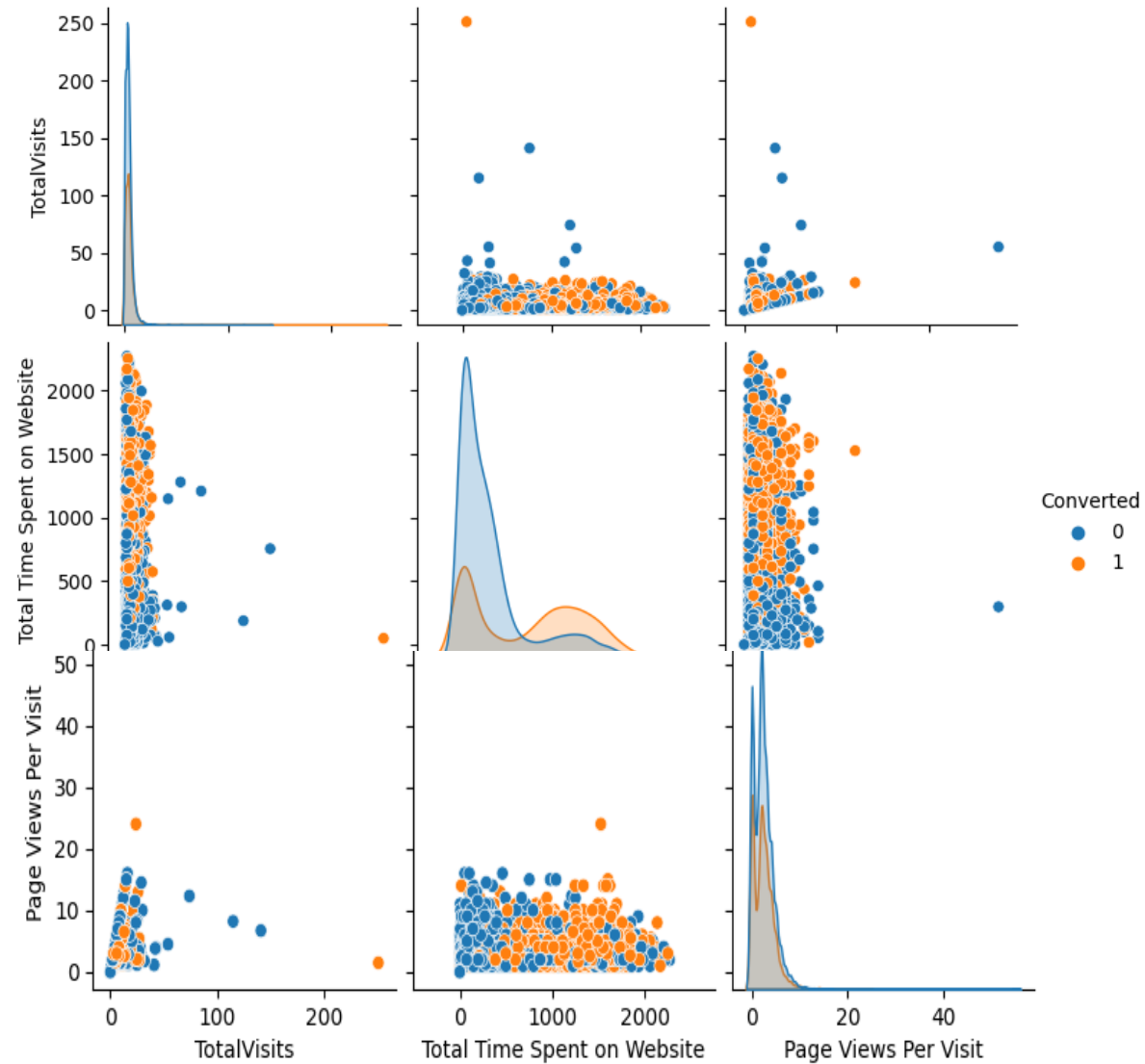
Precision and recall tradeoff

Precision: 87.3%

Recall: 87.2%

Visualization:

```
leads = lead[['TotalVisits', 'Total Time Spent on Website', 'Page Views Per Visit', 'Converted']]
sns.pairplot(leads, diag_kind='kde', hue='Converted')
plt.show()
```



Implications for Lead Scoring

The business insights derived from a scenario where precision is greater than recall and specificity is greater than sensitivity in lead score assignment, the following implications and strategic considerations can be highlighted:

- **Implications:**
- **High Confidence in Positive Predictions:** A higher precision suggests that when the model predicts a lead as "high-quality" or likely to convert, it is more likely to be accurate. This can instill greater confidence in sales and marketing teams when prioritizing leads for engagement.
- **Potential Missed Opportunities:** However, the higher precision at the expense of recall implies that the model may miss out on identifying some leads that are actually of high quality. This could result in missed sales opportunities and revenue potential.
- **Effective Resource Allocation:** With a higher specificity, the model is more reliable in filtering out leads that are unlikely to convert. This can enable sales and marketing teams to allocate their resources more efficiently, focusing on leads that are more likely to result in conversions.
- **Risk of Overlooking Potential Leads:** Conversely, the higher specificity may also mean that some potentially valuable leads are overlooked or not adequately prioritized, leading to missed opportunities for customer acquisition and business growth.