Lead Score Case Study

Content

- Objective
- Summary
- Project introduction.
- Design Workflow
- Model building and evaluation
- Results
- Conclusion

Objective

- 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.
- Company has collected information from various sources of people who in one way or other connected X-eduction.
- Our objective is to create a model that predicts the convertibility of lead based on certain factors to help the X-education convert most of the leads they generate.

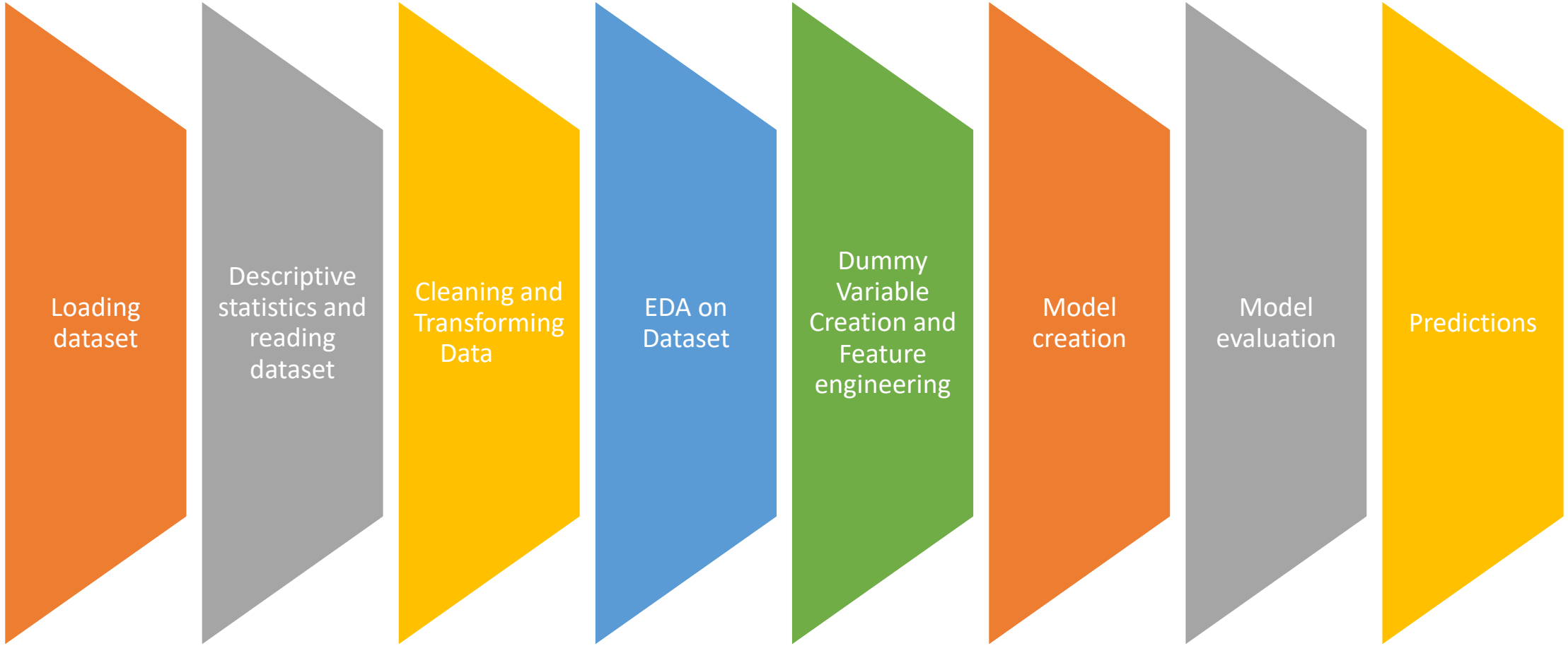
Summary

- Features such as source_Welingak Website, last_activity_Had a Phone Conversation, source_Olark Chat, occupation_Working Professional, source_Olark Chat, and last_activity_SMS Sent are a few of the most important attribute that decides the convertibility of the lead into a customer.
- The model we have created is able to predict approximately 78% of conversions correctly.

Introduction to Project

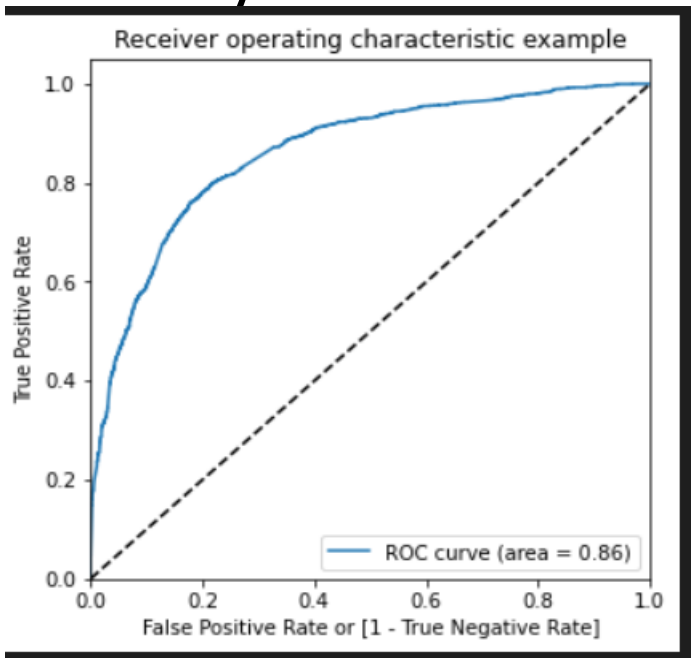
- The dataset contains more than 9000 rows and 23 columns.
- Several of them contained null values.
- It had many single variable features.
- Over 60% of columns contained some null values.
- Many two variable features columns are included in the study such as do not email, get a free copy of mastering interview book etc.
- Over all the dataset contained many outliers, null values, single feature columns etc.

Design Workflow.



Model Creation and Evaluation

- We have used sklearn for model creation and stats model for reading the statistical description of the data.
- We have used RFE and VIF to reduce the number of independent features participating in the study.
-



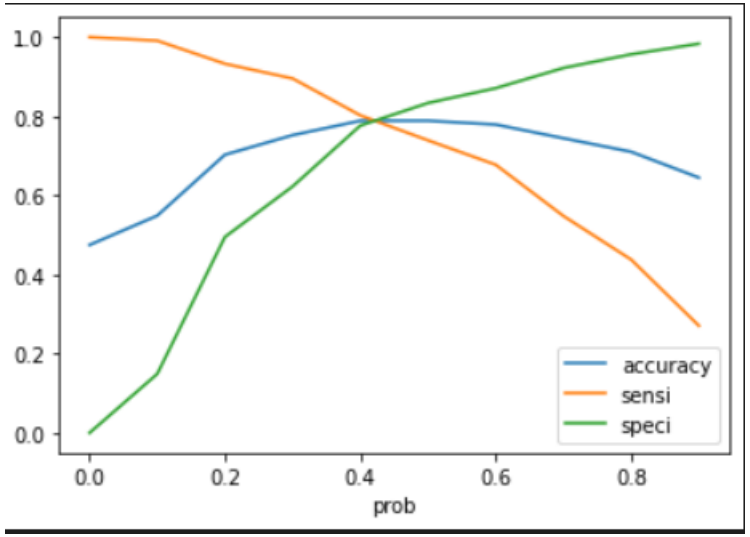
- Roc and AUC curves are used to find the best optimal cutoff.
- Our model gave around 80% accuracy of prediction over train dataset.

```
# overall accuracy of the model is: -
print(metrics.accuracy_score(pred_df.Converted_leads, pred_df.Predicted))
```

```
0.7887323943661971
```

Results

- Optimal cutoff for our model is around 4.2 which is derived from the following graph: -



- Our model has accuracy of 75% which can help in predicting the possible leads which can convert into customer.

Conclusion

- Dataset contained many rows and columns we have used sklearn and stat model to model and evaluate the classification model.
- Further this model can be deployed into the application to use for future prediction based on the available details from the lead.
- This is model useful for decide the aggressiveness which we wants to show towards converting lead into potential lead to customer of our company.

Thank you

Created by: -

Yash Shukla