Lead Scoring Case Study

upGrad

By

Shailendra Sharma

Punith Vadla

Alice Soni

Problem Statement

upGrad

- *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 and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead. 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 while most do not. The typical lead conversion rate at X education is around 30%.*

Business Goal

upGrad

- *X Education company needs to build a model wherein you need to assign a lead score to each of the leads such that the customers with a higher lead score have a higher conversion chance and the customers with a 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%.*

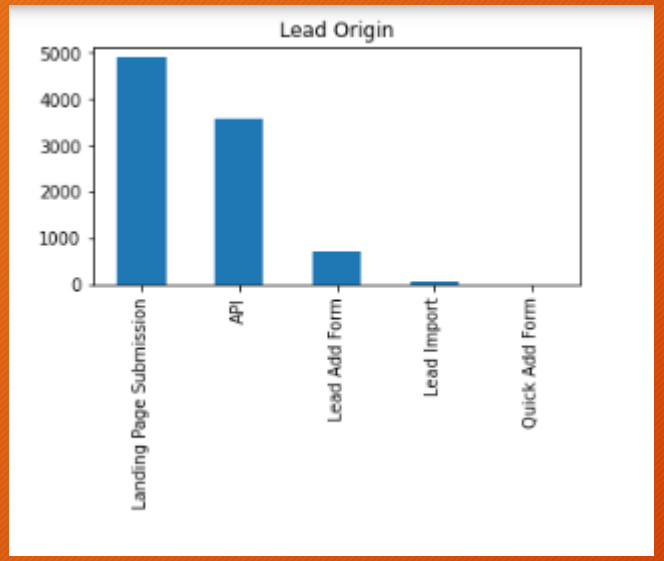
Methodology

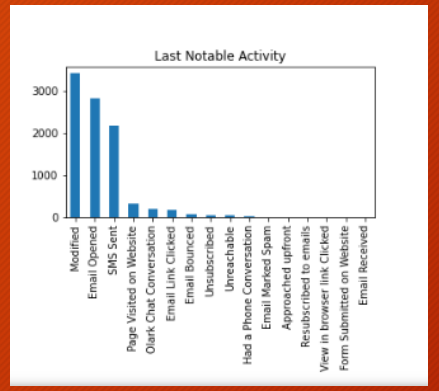
upGrad

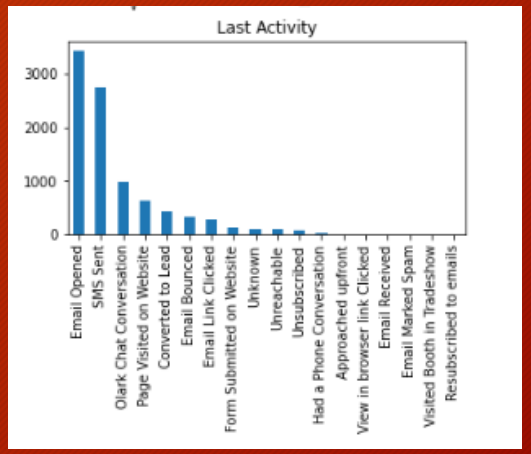
- Import data.
- Clean and prepare the acquired data for further analysis
- Exploratory data analysis for figuring out most helpful attributes for conversion.
- Scaling features
- Prepare the data for model building
- Build a logistic regression model
- Assign a lead score for each leads
- Test the model on train set
- Evaluate model by different measures and metrics
- Test the model on test set
- Measure the accuracy of the model and other metrics for evaluation

Exploratory Data Analysis

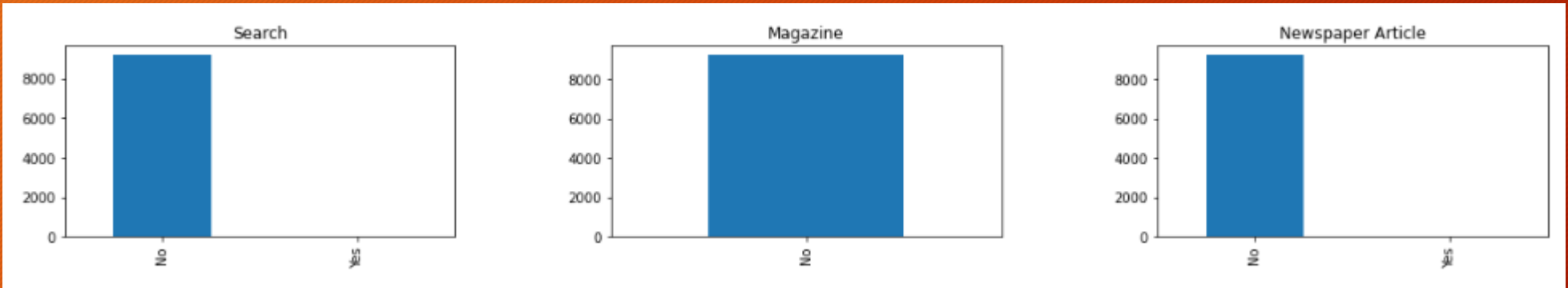
upGrad

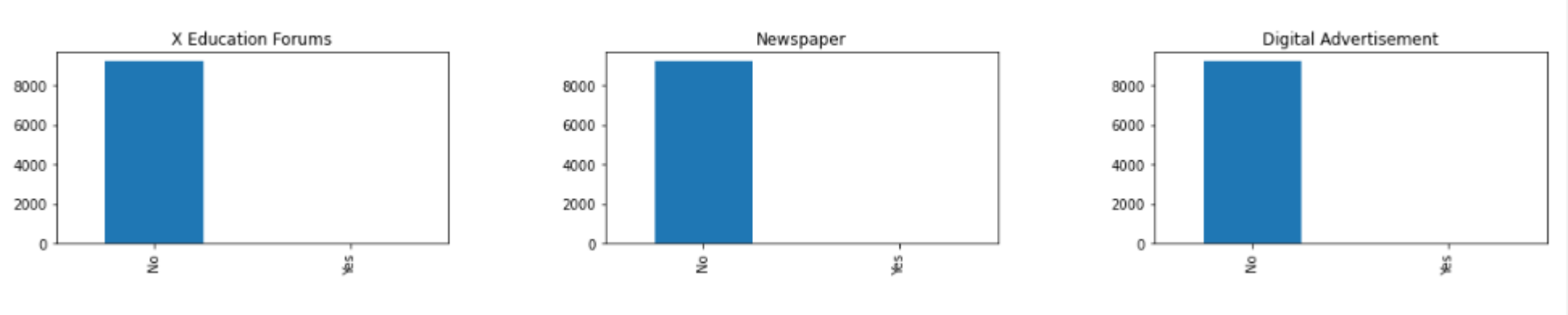




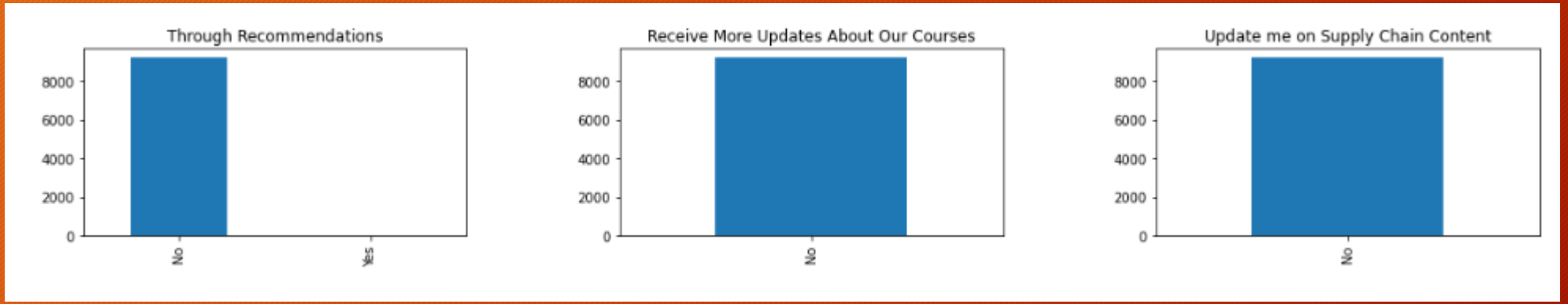


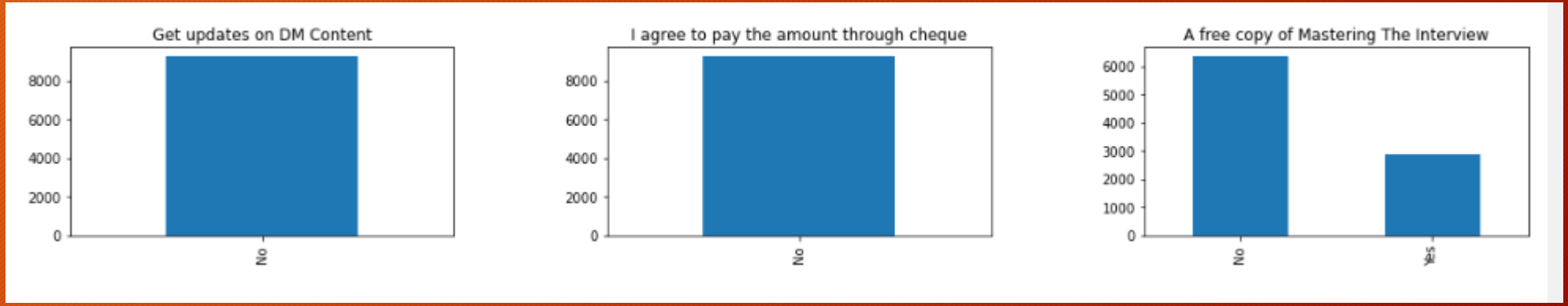
upGrad


Search
8000
6000
4000
2000
0
No
Yes
Magazine
8000
6000
4000
2000
0
No
Newspaper Article
8000
6000
4000
2000
0
No
Yes


X Education Forums
8000
6000
4000
2000
0
No
Yes
Newspaper
8000
6000
4000
2000
0
No
Yes
Digital Advertisement
8000
6000
4000
2000
0
No
Yes

upGrad


Through Recommendations
8000
6000
4000
2000
0
No
Yes
Receive More Updates About Our Courses
8000
6000
4000
2000
0
No
Update me on Supply Chain Content
8000
6000
4000
2000
0
No


Get updates on DM Content
8000
6000
4000
2000
0
No
I agree to pay the amount through cheque
8000
6000
4000
2000
0
No
A free copy of Mastering The Interview
6000
5000
4000
3000
2000
1000
0
No
Yes

Model Building

upGrad

- Splitting into train and test set
- Scale variables in train set
- Build the first model
- Use RFE to eliminate less relevant variables
- Build the next model
- Eliminate variables based on high p-values
- Check VIF value for all the existing columns
- Predict using train set
- Evaluate accuracy and other metric
- Predict using test set
- Precision and recall analysis on test predictions

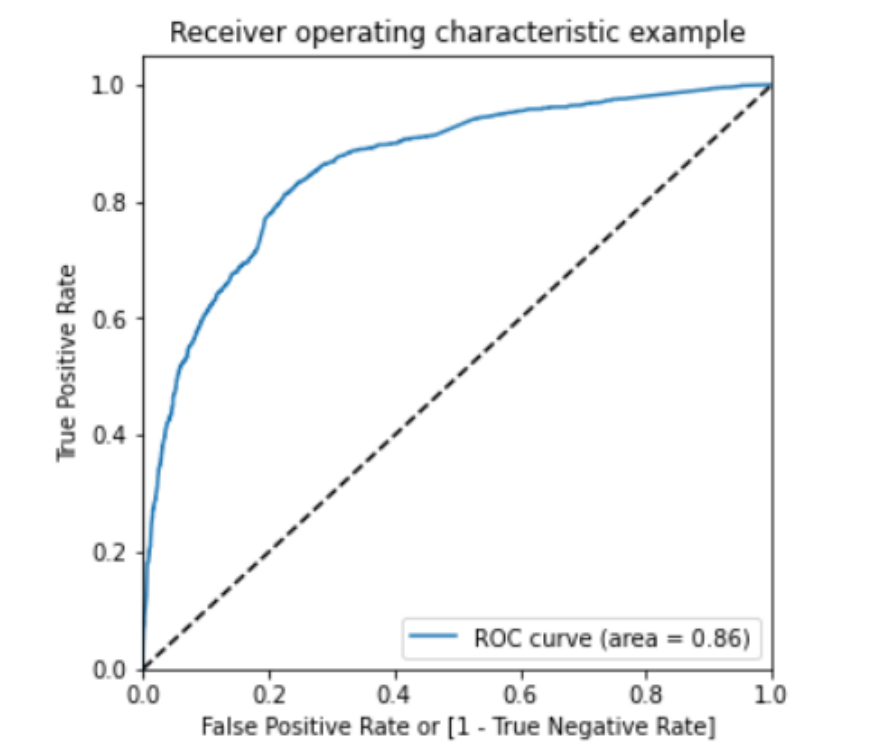
VIF Values

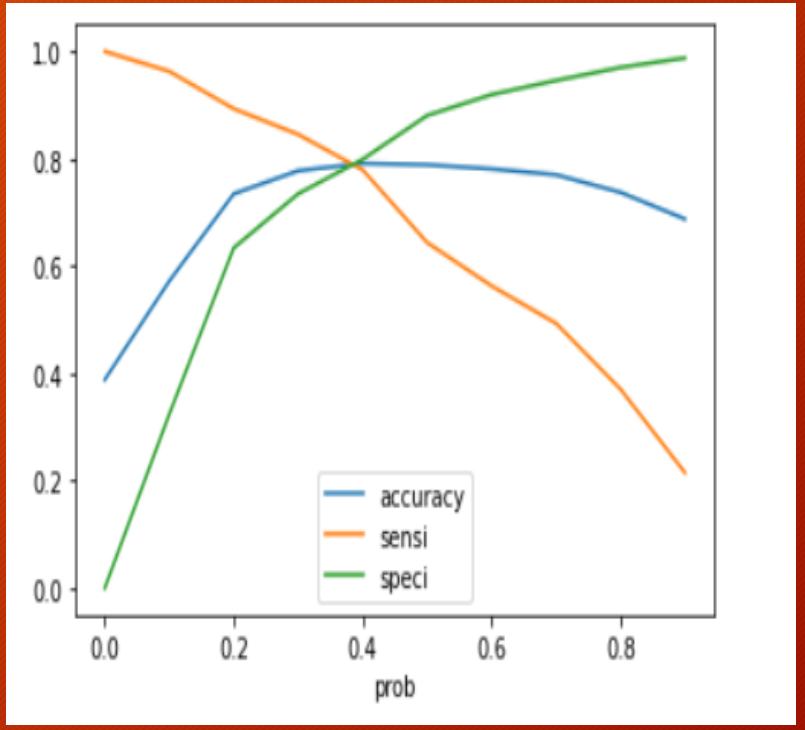
Out[112]:

	Features	VIF
6	Last Notable Activity_Modified	1.84
1	Lead Origin_Landing Page Submission	1.73
3	Specialization_Unknown	1.69
8	Last Notable Activity_SMS Sent	1.57
2	Lead Origin_Lead Add Form	1.13
0	Total Time Spent on Website	1.12
7	Last Notable Activity_Olark Chat Conversation	1.06
4	Last Notable Activity_Email Bounced	1.02
9	Last Notable Activity_Unreachable	1.01
5	Last Notable Activity_Had a Phone Conversation	1.00

ROC Curve and Sensitivity, Specificity cutoff

upGrad





Model evaluation (Train)

upGrad

ACCURACY = 79.13%

SENSITIVITY/RECALL = 64.38

SPECIFICITY = 88.06

PRECISION = 77.39

CONFUSION MATRIX

# Predicted	not_converted	converted
# Actual		
# not_converted	3482	472
# converted	894	1616

Model evaluation (Test)

upGrad

ACCURACY = 79.82%

SENSITIVITY/ RECALL = 78.07%

SPECIFICITY = 80.89%

PRECISION = 71.34

CONFUSION MATRIX

# Predicted	not_converted	converted
# Actual		
# not_converted	1393	329
# converted	230	819

Conclusion

upGrad

- People spending higher than average time on the website are promising leads, so targeting them and approaching them can be helpful in conversions
- Leads whose last known activity is ‘Had a phone conversation’ are potential conversions
- Lead Add Form and Quick Add forms can help find out more leads
- References and offers for referring a lead can be good source for higher conversions
- An alert messages or information could be sent to have high lead conversion rate

Conclusion ‘logistic regression model’

upGrad

- The model shows high close to 79% accuracy
- The threshold has been selected from Accuracy, Sensitivity, specificity measures and precision, recall curves.
- The model shows 76% sensitivity and 83% specificity
- The model finds correct promising leads and leads that have less chances of getting converted
- Overall this model proves to be accurate