Conclusions

After trying out several models, our final model has following characteristics:

1. All p-values are very close to zero.
2. VIFs for all features are very low. There is hardly any multicollinearity present.
3. The overall testing accuracy of **90.78%** at a probability threshold of 0.05 is also very good.

Dataset	Accuracy	Sensitivity	Specificity	False Positive Rate	Positive Predictive Value	Negative Predictive Value	AUC
Train	0.9111	0.8573	0.9449	0.0550	0.9070	0.9135	0.9488
Test	0.9078	0.8412	0.9457	0.0542	0.8984	0.9126	0.9388

The **optimal threshold** for the model is **0.20** which is calculated based on tradeoff between sensitivity, specificity and accuracy. According to business needs, this threshold can be changed to increase or decrease a specific metric.

High sensitivity ensures that most of the leads who are likely to convert are correctly predicted, while high specificity ensures that most of the leads who are not likely to convert are correctly predicted.

Twelve features were selected as the most significant in predicting the conversion:

Features having **positive impact** on conversion probability in **decreasing order** of impact:

Features with Positive Coefficient Values
Tags_Lost to EINS
Tags_Closed by Horizzon
Tags_Will revert after reading the email
Tags_Busy
Lead Source_Welingak Website
Last Notable Activity_SMS Sent

Features with Positive Coefficient Values
Lead Origin_Lead Add Form

Features having **negative impact** on conversion probability in **decreasing order** of impact:

Features with Negative Coefficient Values
Lead Quality_Worst
Lead Quality_Not Sure
Tags_switched off
Tags_Ringing
Do Not Email