Recommended Model for Good Prediction:

Model Compare on Test Dataset

	Model	Accuracy	Precision	Recall	F1 score	ROC AUC Score
0	Random Forest with PCA,with Hyperparameter Tuned,with Adjusted Threshold (0.4) and with SMOTE	0.72	0.20	0.85	0.32	0.86
0	XGboost Classifier with PCA, with Hyperparameter Tuned and with SMOTE, with Adjusted Threshold (0.4) and with SMOTE	0.76	0.22	0.84	0.35	0.88
0	Decision Tree with PCA,with Hyperparameter Tuned,with Adjusted Threshold (0.4) and with SMOTE	0.68	0.17	0.82	0.29	0.81
0	Logistic regression with PCA,without Hyperparameter Tuned and with SMOTE	0.81	0.26	0.81	0.40	0.88
0	Logistic regression with PCA, with Hyperparameter Tuned and with SMOTE	0.81	0.26	0.81	0.40	0.88
0	XGboost Classifier with PCA, with Hyperparameter Tuned and with SMOTE	0.82	0.28	0.80	0.41	0.88
0	XGboost Classifier with PCA, with Hyperparameter Tuned and with SMOTE, with Optimul Threshold (0.53) and with SMOTE	0.84	0.30	0.78	0.43	0.88
0	XGboost Classifier with PCA, without Hyperparameter Tuned and with SMOTE	0.84	0.30	0.77	0.43	0.88
0	Random Forest with PCA,with Hyperparameter Tuned and with SMOTE	0.83	0.28	0.76	0.41	0.86
0	Random Forest with PCA,with Hyperparameter Tuned,with Optimul Threshold (0.51) and with SMOTE	0.84	0.29	0.75	0.42	0.86
0	Decision Tree with PCA,with Hyperparameter Tuned and with SMOTE	0.78	0.22	0.72	0.34	0.81
0	Decision Tree with PCA,with Hyperparameter Tuned,with Optimul Threshold (0.53) and with SMOTE	0.78	0.22	0.72	0.34	0.81
0	Random Forest with PCA , without Hyperparameter Tune and with SMOTE	0.91	0.44	0.57	0.50	0.88
0	Decision Tree with PCA, without Hyperparameter Tune and with SMOTE	0.84	0.24	0.48	0.32	0.68

Recall is the most important business metric for the telecom churn problem. The company would like to identify most customers at risk of churning, even if there are many customers that are misclassified as churn. The cost to the company of churning is much higher than having a few false positives.

Regression Model:

- Overall, the **Logistic Regression** model without any threshold adjustment, performs best. It achieved the **best recall of 81.5%** for test data. Also, the overall accuracy is consistent for Test and train data, thus avoiding overfitting. The precision is compromised in this effort but the business objective to predict Churn customers is most accurately captured by it.

Boosting Model:

- In Boosting Methods, Extreme Gradient Boosting Classifier (XGBoost) achieved 84.5% recall rate with adjusted threshold (cutoff = 0.4) on test data.

Tree Model:

- From the Tree Family, the Decision Tree overfitted the data slightly while obtaining 82.3% recall with adjusted threshold (cutoff = 0.4) on test data.

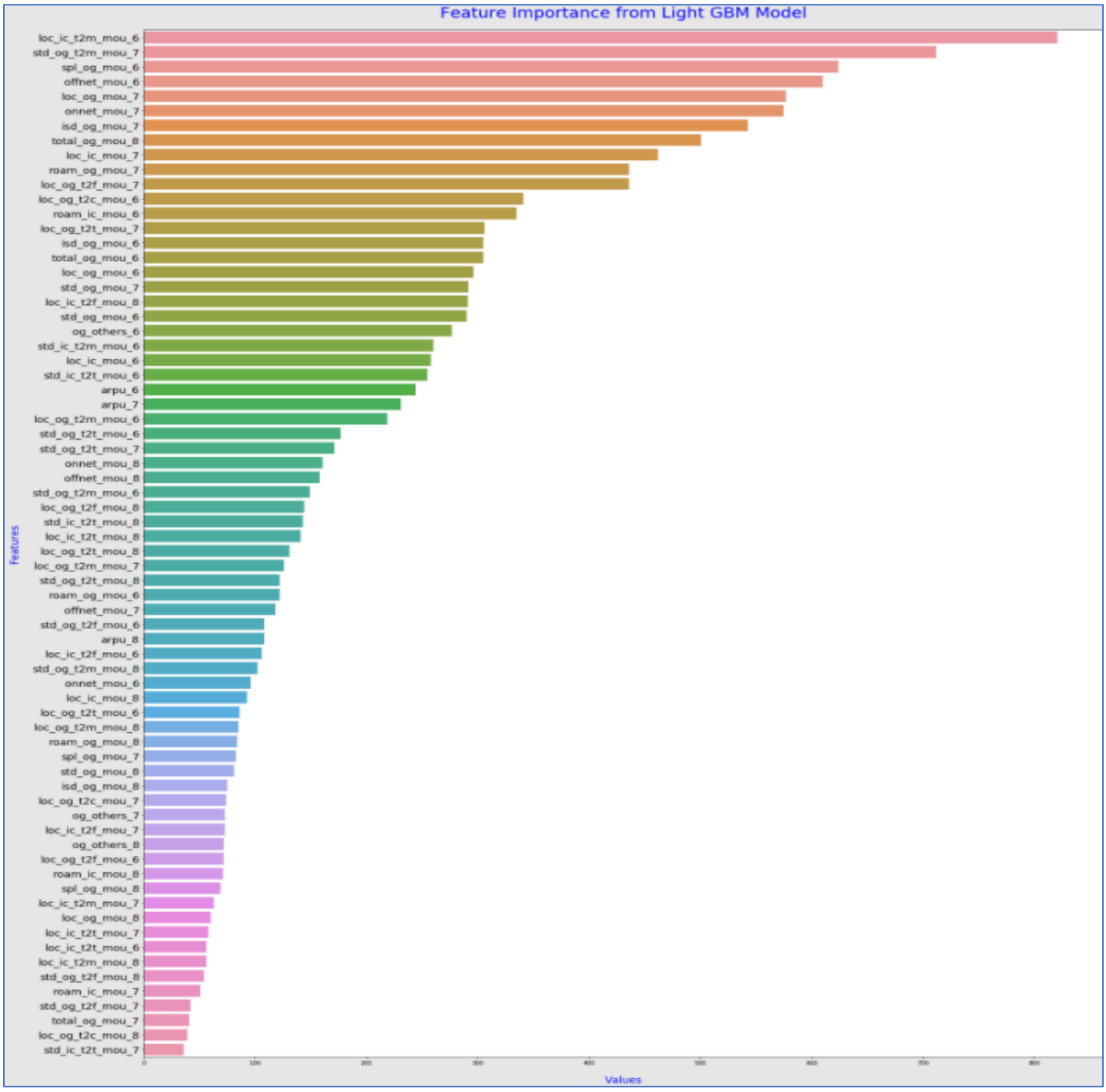
Bagging Model:

- The Random Forest is also overfitted the data slightly while obtaining 85% recall with adjusted threshold (cutoff = 0.4) on test data.

Recommendation:

- Out of these models, the recommended model would be **Logistic Regression** as a good predictor model as it is simpler than other models and having good accuracy and sensitivity/recall without overfitting.

Model for Good Interpretation



Recommended Model for Good Interpretability:

- From the above models, we will choose **Light GBM model** as a good interpretable model because this model has selected many features which belong to month 8 i.e., the action month. Hence, it is clear that what happens in the action phase has a direct impact on the customer churn of high value customers.

The important features given by Light GBM model are as follows:

1. **arpu_8**
2. **isd_og_mou_8**
3. **loc_ic_mou_8**
4. **loc_ic_t2m_mou_8**
5. **loc_og_mou_8**
6. **loc_og_t2c_mou_8**
7. **loc_og_t2m_mou_8**
8. **og_others_8**
9. **roam_ic_mou_8**
10. **roam_og_mou_8**
11. **spl_og_mou_8**
12. **std_og_mou_8**
13. **std_og_t2f_mou_8**
14. **std_og_t2m_mou_8**
15. **loc_ic_t2f_mou_6**
16. **loc_ic_t2f_mou_7**
17. **loc_ic_t2m_mou_7**
18. **loc_ic_t2t_mou_6**
19. **loc_ic_t2t_mou_7**
20. **loc_og_t2c_mou_7**
21. **loc_og_t2f_mou_6**
22. **loc_og_t2t_mou_6**
23. **og_others_7**
24. **onnet_mou_6**
25. **roam_ic_mou_7**
26. **spl_og_mou_7**
27. **std_ic_t2t_mou_7**
28. **std_og_t2f_mou_6**
29. **std_og_t2f_mou_7**
30. **total_og_mou_7**

Local calls Mou's be it incoming or outgoing have a very important role for churn predictions. Reduction in these KPI's forms a clear indicator of churn.

Overall, drop in any of the above indicator KPI is a signal that the customer is not actively engaging in the services offered by the Network operator and thus may choose to churn in the near future.

Business Recommendation:

Based on top 30 features, 14 features are from action phase, whereas rest features are from good phase. We can propose following business decision which will help retain the high value customer who may be planning to leave:

Incoming Calls: Total incoming minutes, local & std incoming minutes are in top features. It shows that more incoming calls have high impact on retaining the customer. Incoming calls (local & std) should be made free for all customer

Outgoing Calls: Local, STD and Total outgoing call minutes are among the top feature. It shows that higher the outgoing calls, the better chances of the customer to stay with the telecom company. To ensure that, telecom company should launch various discounted offer on outgoing calls. Lower price/minute or unlimited outgoing calls to same telecom company provided connections will also benefit in retaining customer group.

Roaming: Incoming and Outgoing calls during roaming are critical factor as these are part of top features. Incoming call roaming charges should be made free within the country. STD outgoing call should be same and shouldn't attract extra cost if the roaming option is within the country. For outside country incoming and outgoing roaming as well as data connection, company should launch special country wise roaming packages. These benefits will attract customer to stay with the telecom company for longer period.

The telecom company should pay close attention to drop in MoU, ARPU and data usage (2g and 3g) month over month. If feasible, the company should track these numbers week over week. Since billing cycles are typically monthly, a drop in usage numbers will give the company time to react when tracked at weekly level.

In general, when the above features have a decline trend from the good phase to the action phase, then the customer care executives can reach them out and try to understand if there are any issues.

If the goal is to engage and talk to the customers to prevent them from churning, its ok to engage with those who are mistakenly tagged as not churned, (False Positives) as it does not cause any negative problem. It could potentially make them even happier for the extra attention they are getting.