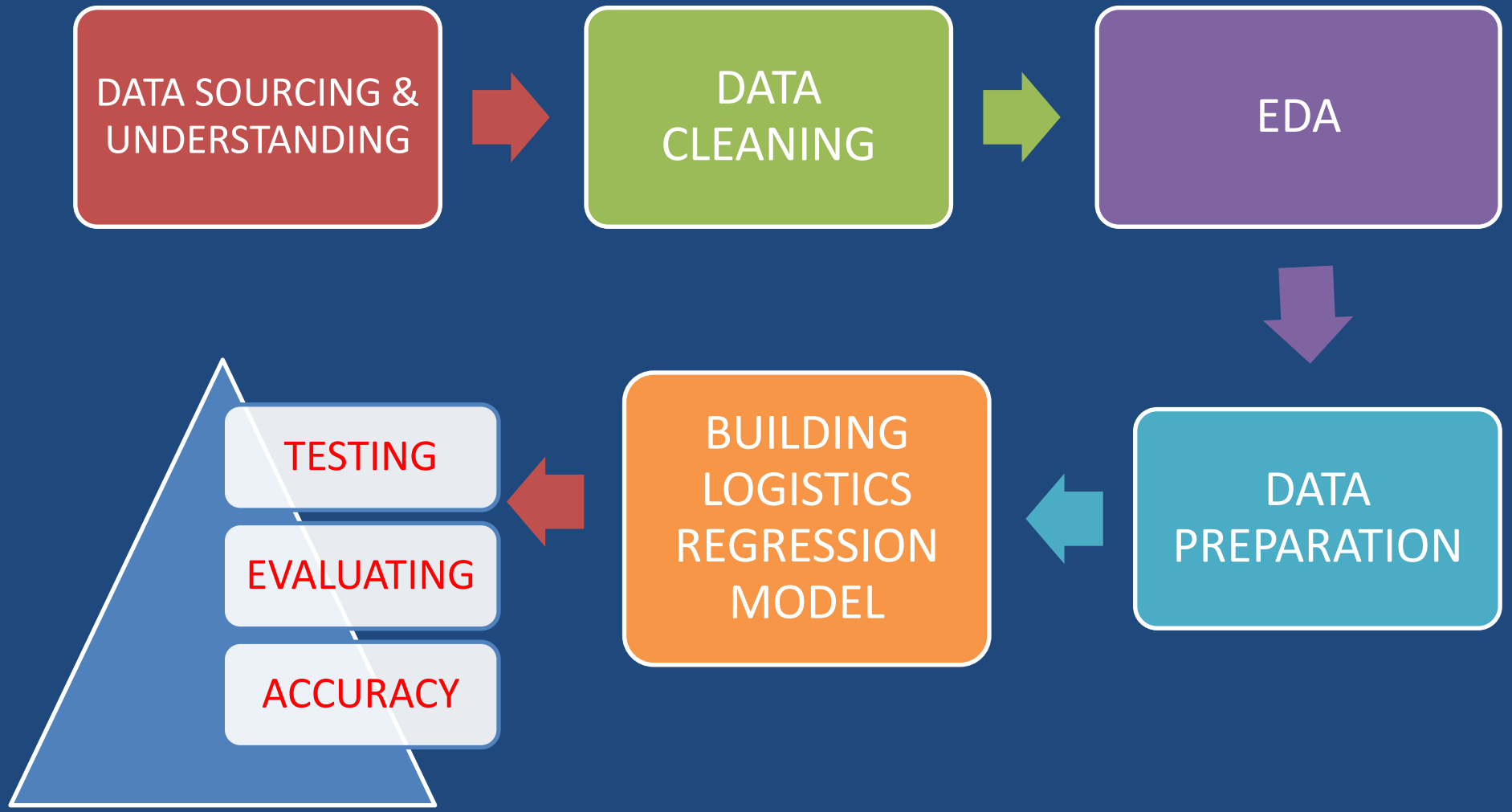
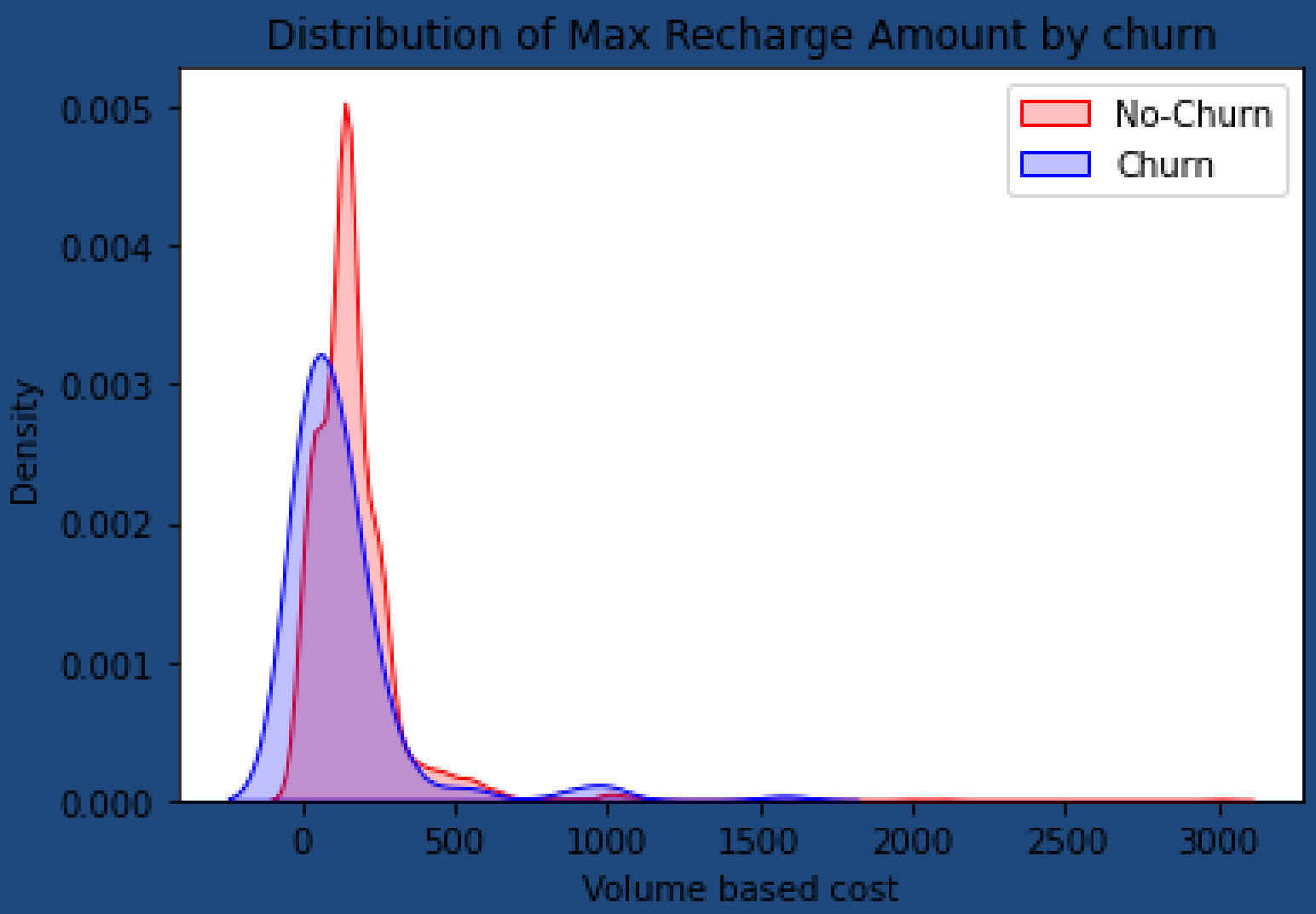
TELECOM CHURN CASE STUDY

PROBLEM STATEMENT

- To analyse customer-level data
- To build predictive models to identify customers at high risk of churn
- To identify the main indicators of churn

APPROACH


DATA SOURCING & UNDERSTANDING
DATA CLEANING
EDA
DATA PREPARATION
BUILDING LOGISTICS REGRESSION MODEL
TESTING
EVALUATING
ACCURACY

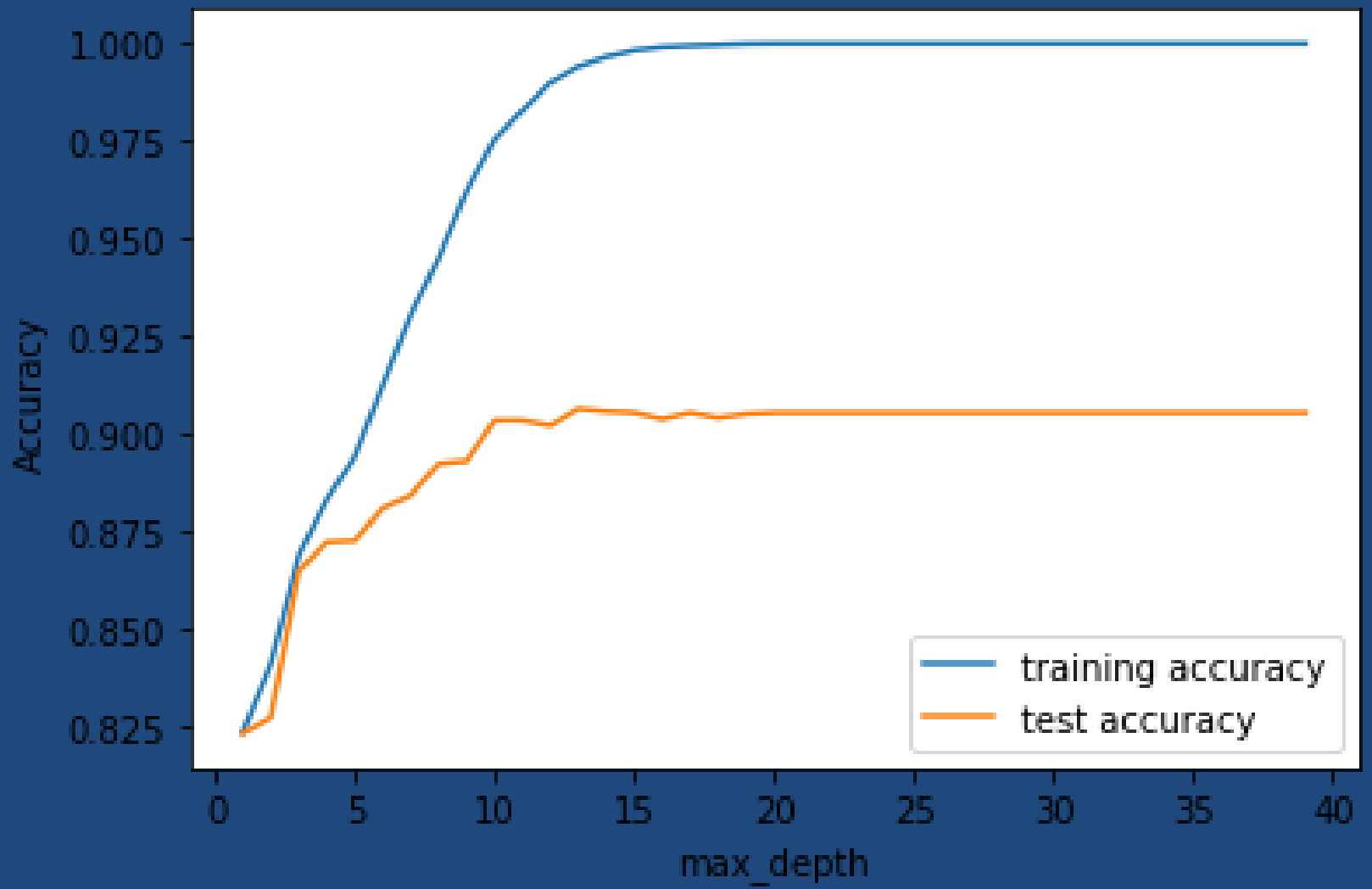

Distribution of Max Recharge Amount by churn
No-Churn
Churn
Density
0.005
0.004
0.003
0.002
0.001
0.000
0
500
1000
1500
2000
2500
3000
Volume based cost

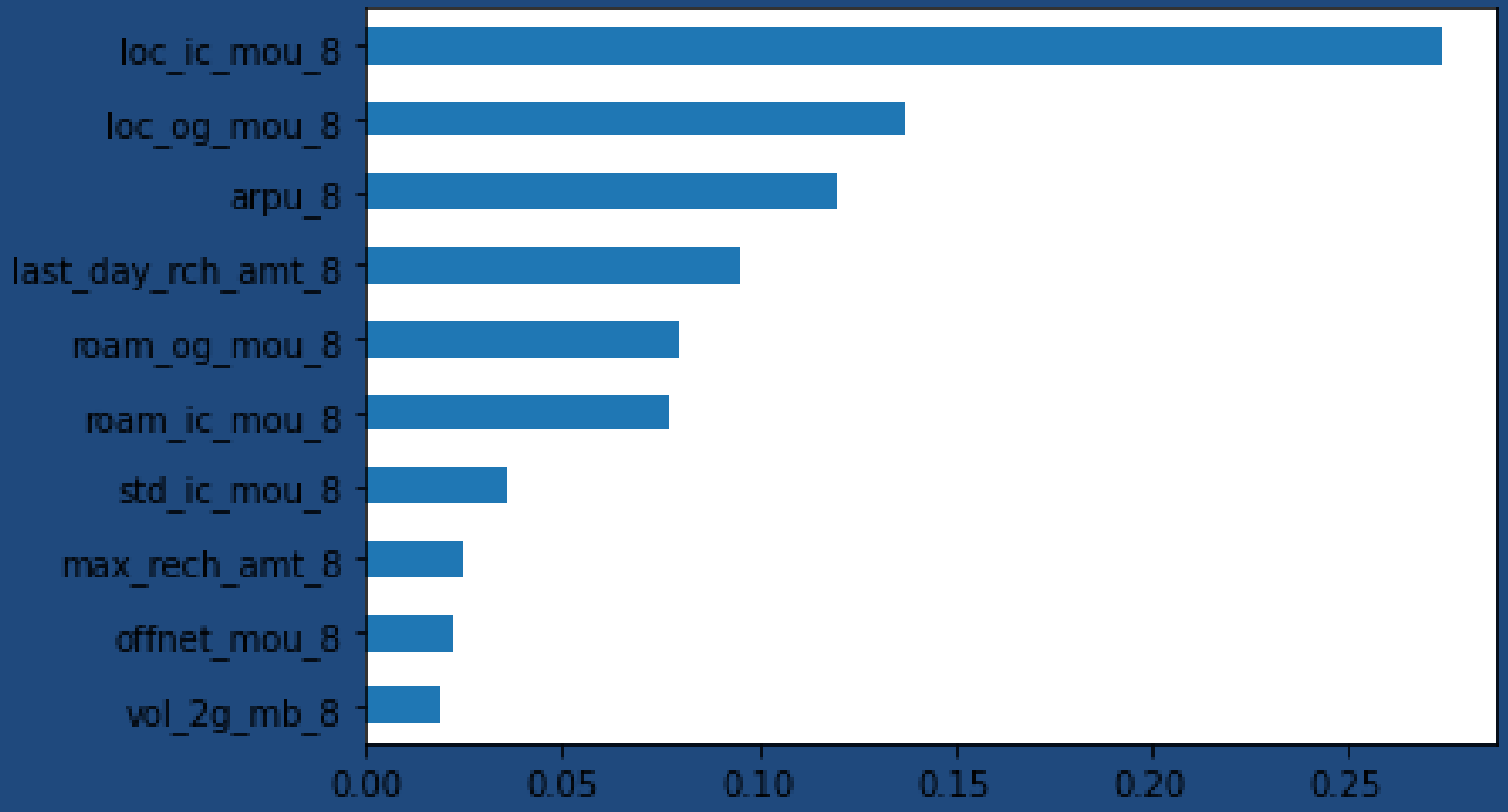
ANALYSIS

- SVM gave an accuracy of 94%

15 most important features selected by RFE

['loc_og_t2c_mou_8', 'std_og_mou_8', 'loc_ic_mou_8',
'std_ic_t2t_mou_8', 'spl_ic_mou_8',
'max_rech_amt_8', 'last_day_rch_amt_8',
'vol_2g_mb_8', 'monthly_2g_8', 'sachet_2g_8',
'monthly_3g_8', 'sep_vbc_3g',
'avg_std_og_mou_av67', 'avg_loc_ic_mou_av67',
'avg_std_ic_t2t_mou_av67


1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825
Accuracy
0
5
10
15
20
25
30
35
40
max_depth
training accuracy
test accuracy


loc_ic_mou_8
loc_og_mou_8
arpu_8
last_day_rch_amt_8
roam_og_mou_8
roam_ic_mou_8
std_ic_mou_8
max_rech_amt_8
offnet_mou_8
vol_2g_mb_8
0.00
0.05
0.10
0.15
0.20
0.25

Conclusions from Random Forest¶

Local Incoming for Month 8, Average Revenue Per Customer for Month 8 and Max Recharge Amount for Month 8 are the most important predictor variables to predict churn.

CONCLUSION

- **Std Outgoing Calls and Revenue Per Customer are strong indicators of Churn.**
- **Local Incoming and Outgoing Calls for 8th Month and avg revenue in 8th Month are the most important columns to predict churn.**
- **cutomers with tenure less than 4 yrs are more likely to churn.**
- **Max Recharge Amount is a strong feature to predict churn.**
- **Random Forest produced the best prediction results followed by SVM.**