EXPONEA

EXPERIENCE CLOUD

DATA MINING ASSIGMENT

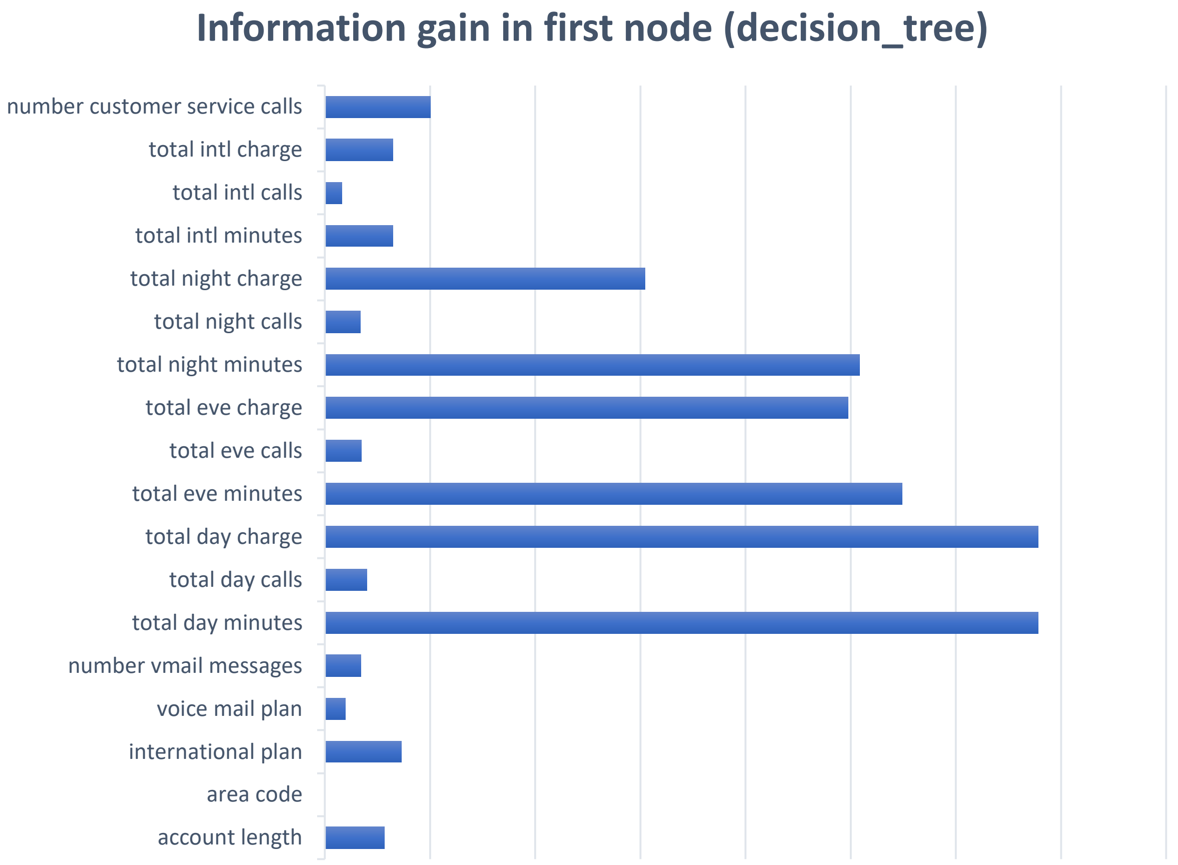
Models

Implemented are these models:

- Logistic regression - ~ 86,64 % succes on data for testing (UAC – 0.5719)
- Decision tree - ~ 92,24 % succes on data for testing (UAC – 0.8259)
- Random forest - ~ 95,24 % succes on data for testing (UAC – 0.8395)

Affects on the churn

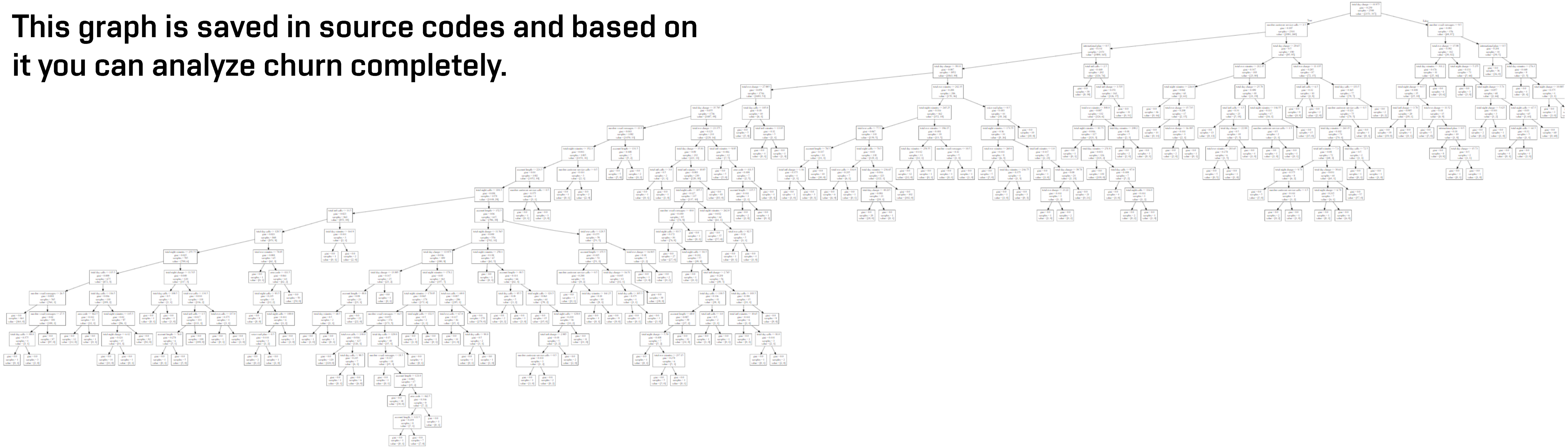
- Decision Tree model is the least black box model out of all three, so I will describe the most important parameters connected to churn on it.
- On the left you can see the graph showing importance of parameters for its decision making process.
- Model takes the most important parameter and based on it predicts probable churn, then takes the second most important parameter and predicts more accurate churn and so on.
- Importance of parametrs is changing during decision making process, graph on the left showing importance of parametrs in first step.



Affects on the churn

- This decision making process can be visualized by graph, on which you can predict even without computer.

- This graph is saved in source codes and based on it you can analyze churn completely.



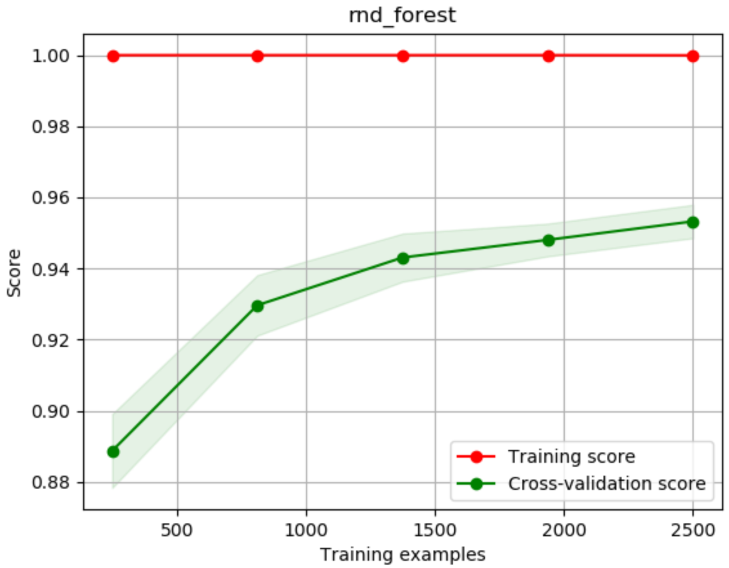
Affects on the churn and recommendations

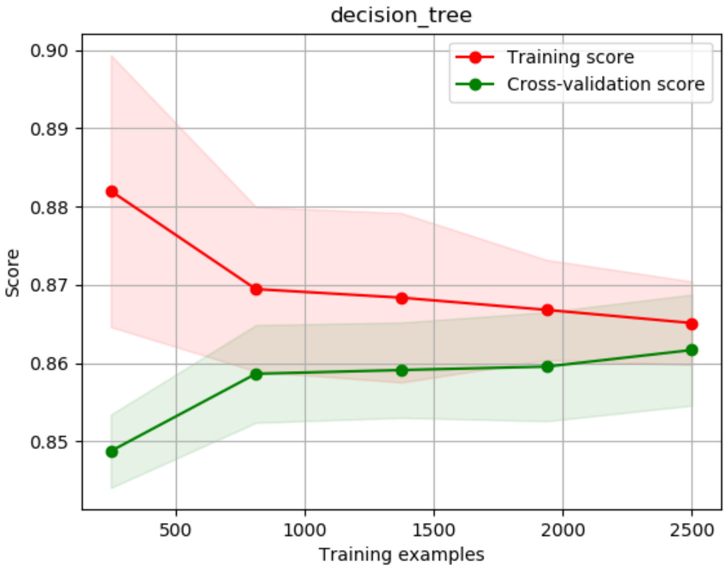
- Based on my models most important parametrs are in general charges, so my recommendation is to offer more offten plans (voice mail, international)

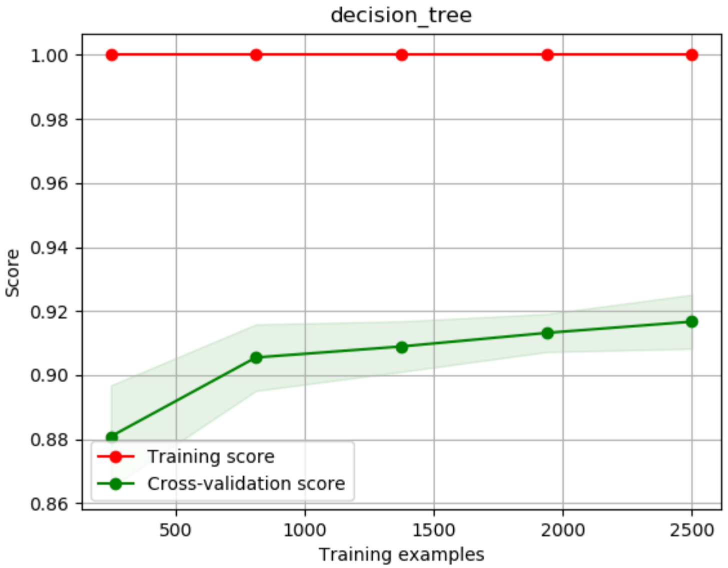
Quality of models

Quality of models was affected by dataset length.

This is based on learning graphs, where you can see, that models were still learning when they reached the end of the learning dataset.


rnd_forest
Score
Training examples
Training score
Cross-validation score


decision_tree
Score
Training examples
Training score
Cross-validation score


decision_tree
Score
Training examples
Training score
Cross-validation score