AIDA-Project
Telekom Customer Churn Prediction

M. Sauer
GIT, data models, feature selection

J. E. Dietert
EDA, classic fit, NN

T. Habermann
NN models, KNN

Definition

Churn rate (sometimes called **attrition** rate), in its broadest sense, is a measure of the number of individuals or items moving out of a collective group over a specific period. It is one of two primary factors that determine the steady-state level of customers a business will support.

source: wikipedia.org

Project And Objectives

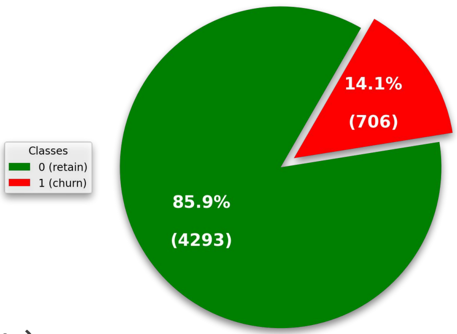
Knowing the intent of a customer to leave, is a highly valuable information. It can be used for prevention, by providing specific offers to these.

Objective of this project, is to create a machine learning algorithm, which is able to provide this knowledge based on everyday data of a telecommunication company.

Knowing the data

- **Set of 5000 customers**
- **21 features** about each of them (including a binary value if they do churn or not)
- 1/7 churn customers, 6/7 no churn customers

Raw data class distribution



class (target)	state	account_length	total_day_minutes	total_eve_minutes	total_night_minutes	total_intl_minutes
international_plan	area_code	number_vmail_messages	total_day_calls	total_eve_calls	total_night_calls	total_intl_calls
voice_mail_plan	phone_number	number_customer_service_calls	total_day_charge	total_eve_charge	total_night_charge	total_intl_charge

Metrics...Metrics...Metrics

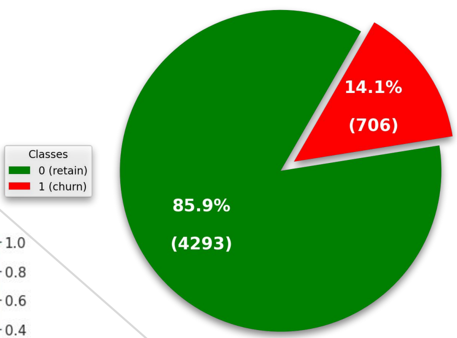
The given task is a ***binary classification***

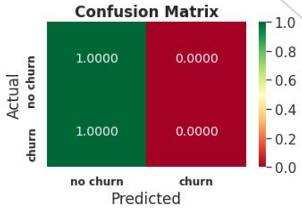
➢ Since we have an ***imbalanced class*** distribution accuracy is not the optimal metric.
E.g. if a model would predict always 0 (no churn), we get approx. 0.86 accuracy but would totally fail to predict churns

➢ It is necessary to have a ***high recall value*** i.e. to have nearly all churners in the set of customers identified for potential churn

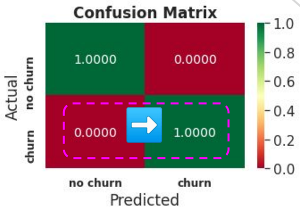
➢ But increasing number of retaining customers seen as churners by mistake makes it necessary to ***monitor precision***

Raw data class distribution

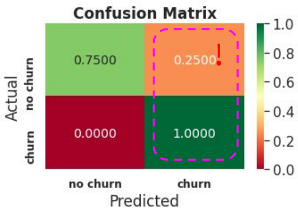




Constant 0 prediction but already 0.86 accuracy

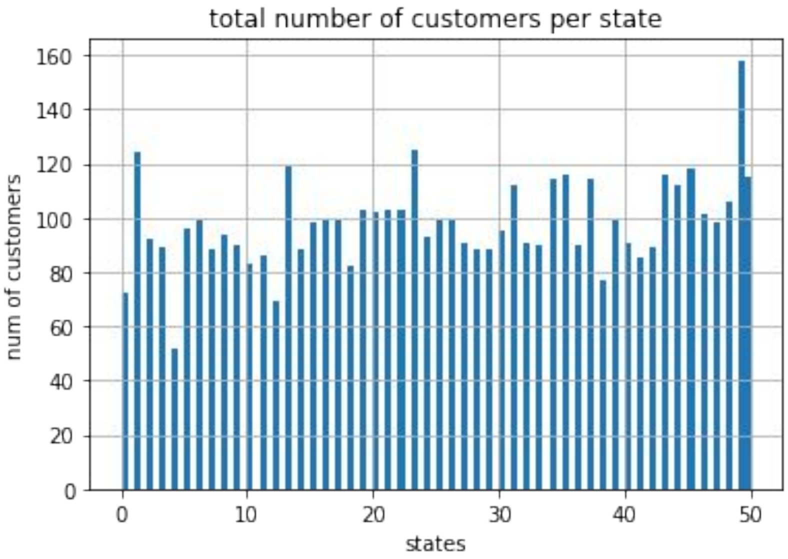


Raise recall

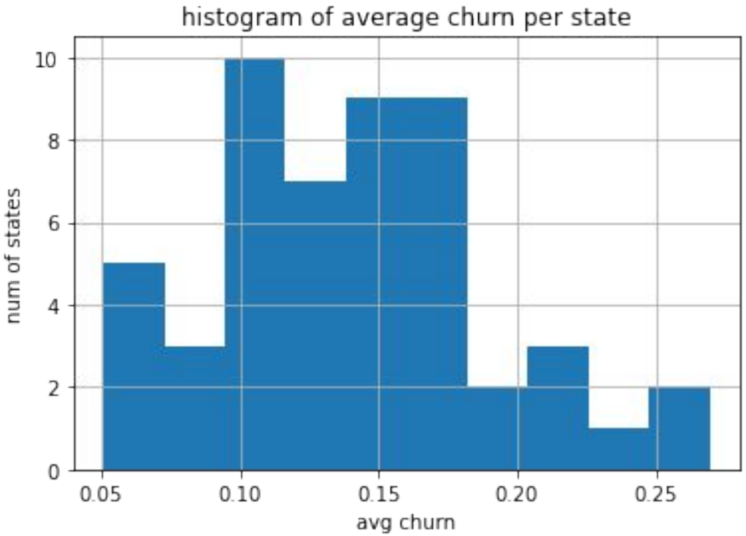


Monitor precision

Data Inspection



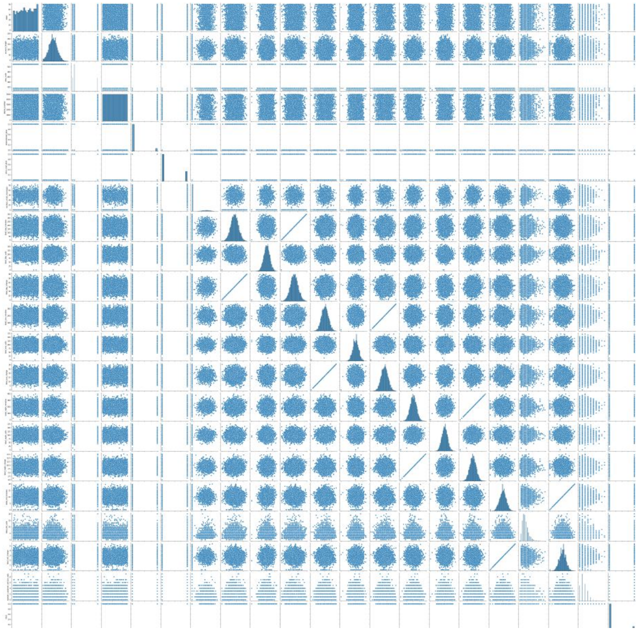
distribution of #customers per (US)state: 50...125



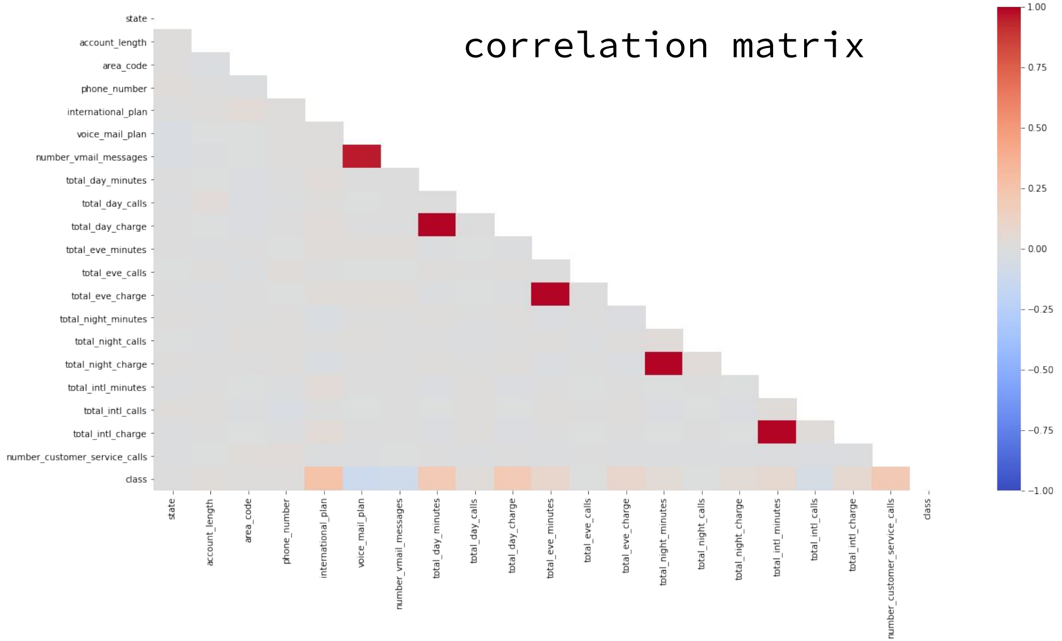
average churn rate per (US) state: 5%...27%

Data Inspection: eliminate correlations

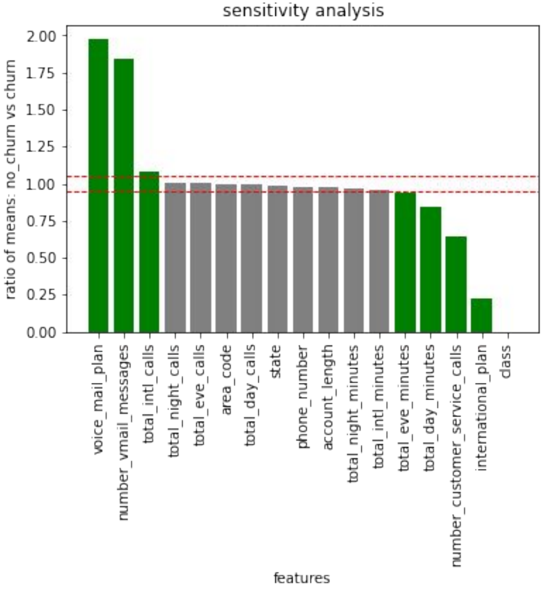
pairplot of dependencies



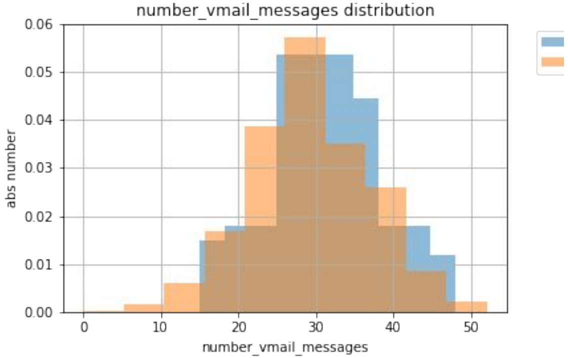
correlation matrix



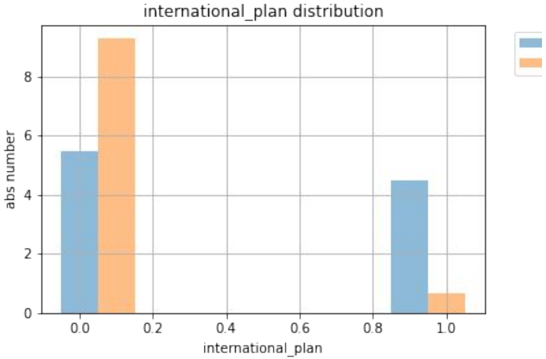
Data Preprocessing



grouping according to class churn
build the ratio of feature means:
mean (no_churn) / mean (churn)



#vmail_messages <= 30: true: no churn
#vmail_messages <= 30: false: churn



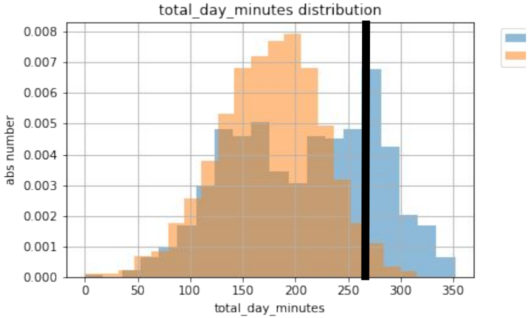
customers without international plan churn less, which might come from some misalignment of international charges.

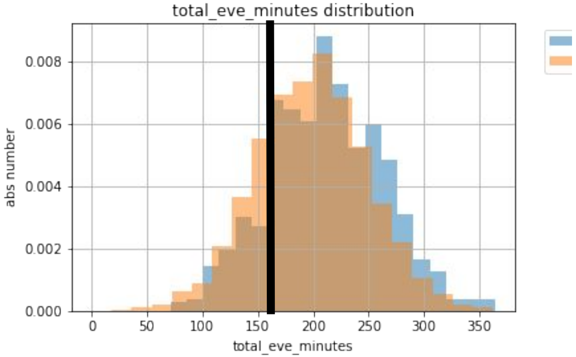
International charges per minute (27ct/min) do not differ for customers with or without international plan

Dietert

features with high effects > 1.05 or < 0.95 (as green), while other features have similar averages in their distributions

Data Preprocessing:



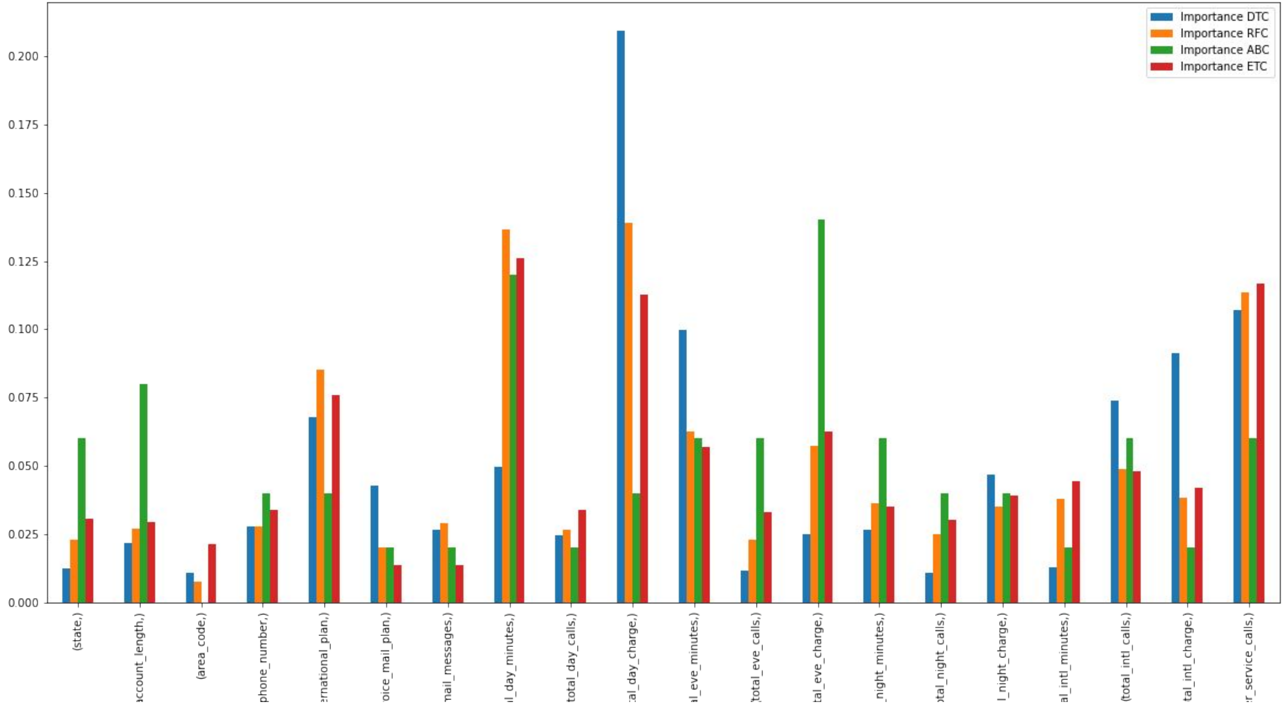


total day minutes <= 265.4
This decision is valid since also seen in the statistics of no_churn vs churn customers. There seem two sub-sets of churners in their day-minutes.

total eve minutes <= 168.35
This decision is valid since also seen in the statistics of no_churn vs churn customers.

number customer calls <= 3.5
This decision is valid since also seen in the statistics of no_churn vs churn customers. If a customer calls less than 3.5 times the hotline to complain, they could be identified as “happy”, so they stay.

Feature Selection:



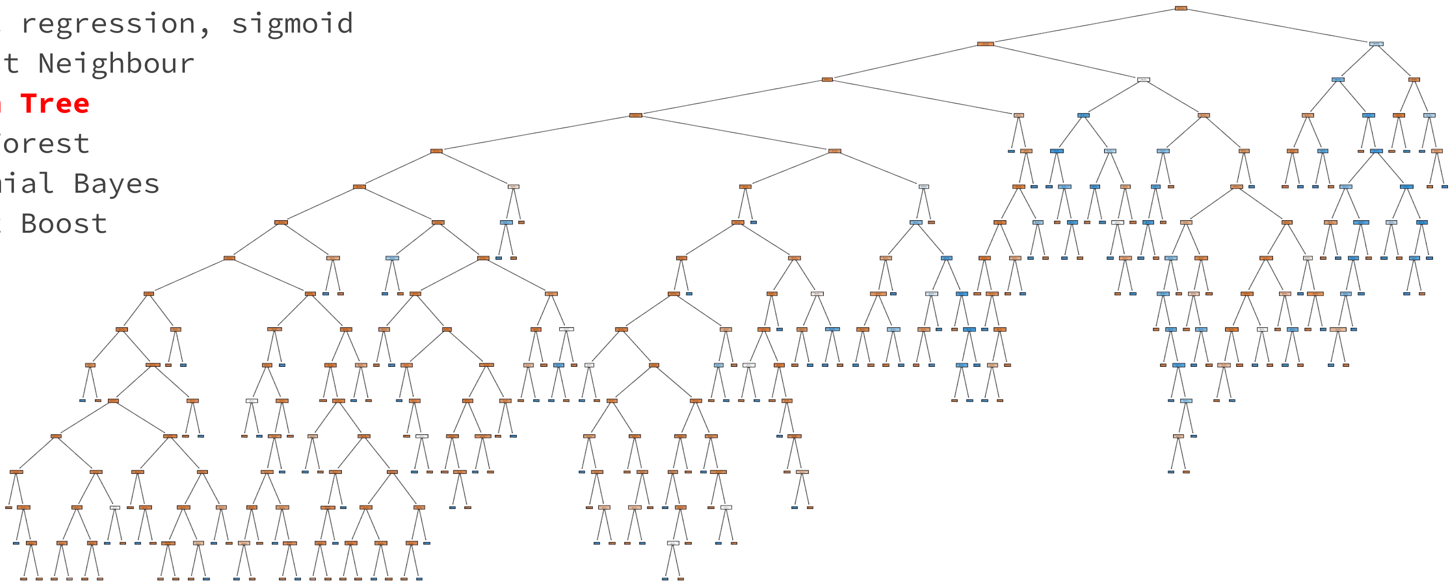
Feature Importance Comparison for different Classifiers

Decision Tree Classifier (DTC)	Random Forest Classifier (RFC)
Extra Trees Classifier (ETC)	Ada Boost Classifier (ABC)

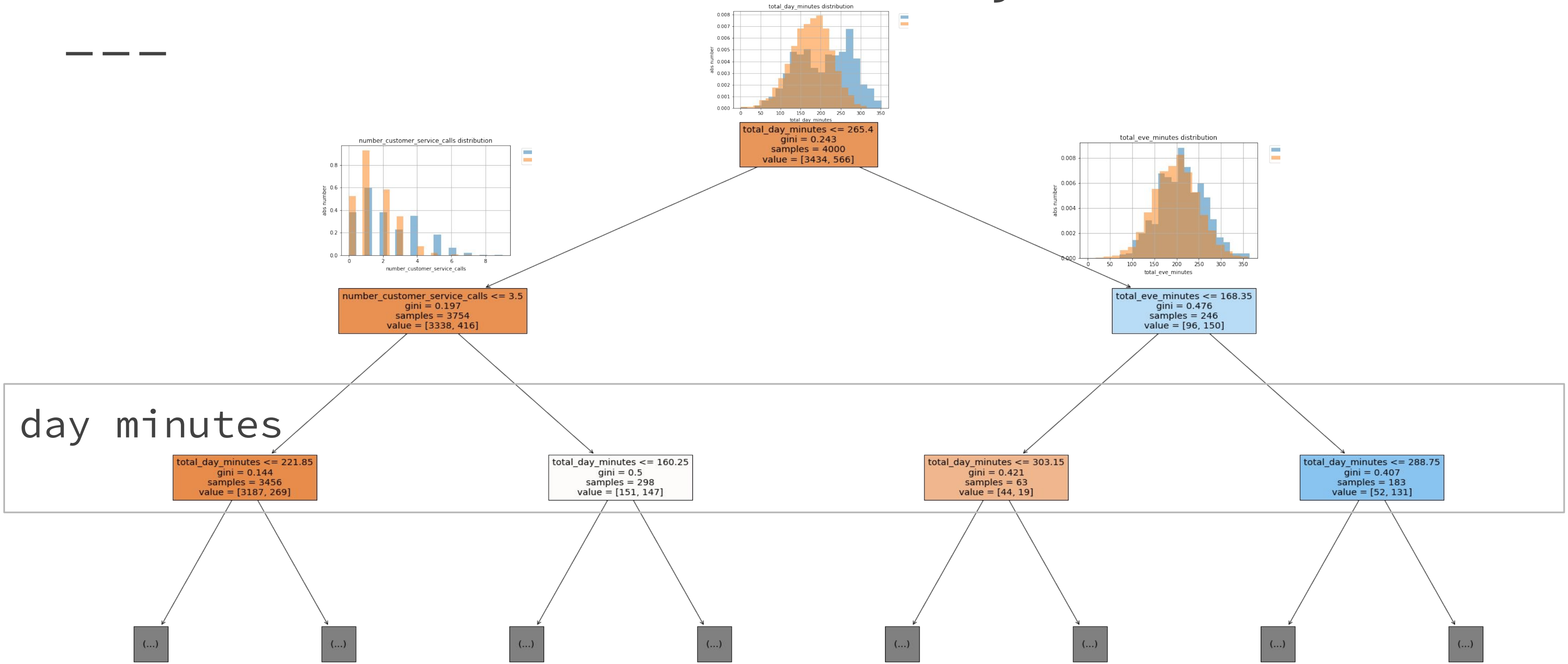
Standard classifiers: Decision Tree with best results

Standard classifier with individual parameterization:

- logistic regression, sigmoid
- k-Nearest Neighbour
- **Decision Tree**
- Random Forest
- Multinomial Bayes
- Gradient Boost
- XGBoost
- SVC

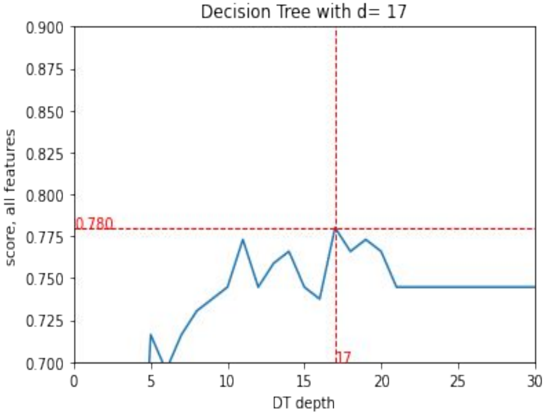


DT with reasonable decisions backed by statistics

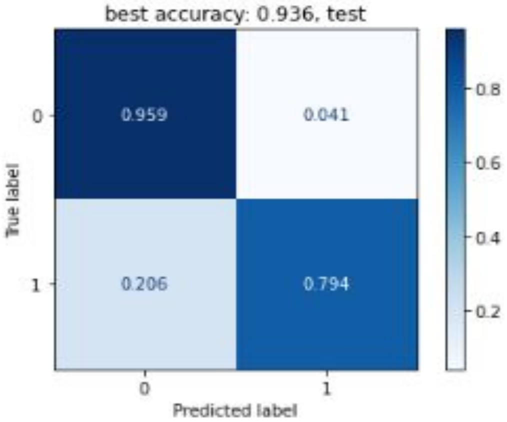


Dietert

Classification result

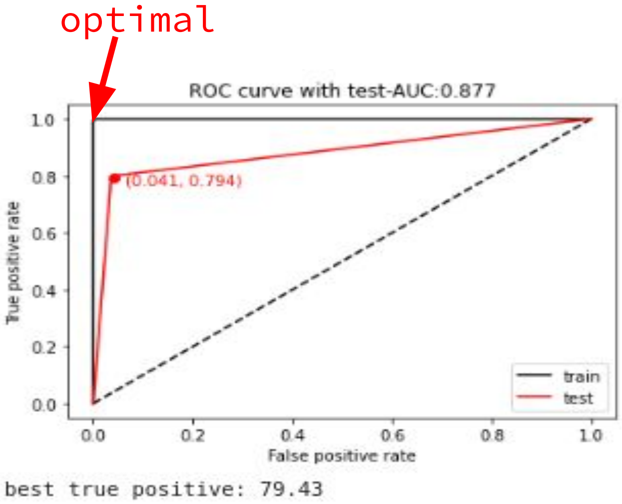


The hyperparameter tuning of the DT classifier comes up with a degree of 17, slightly higher than the input feature dimension of 13.



The classifier on test set leads to

- **accuracy of 93.6%**
- **true positive rate of 79.4%**
- ~20% (28) were classified as no-churners despite they were churned customers.
- 869* 4%= 35 customers are misclassified as potential churners



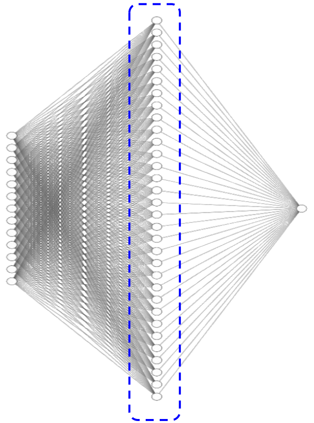
Alternative visualization as ROC plot with the ideal target prediction as black line with the top-most left corner as the ideal point.

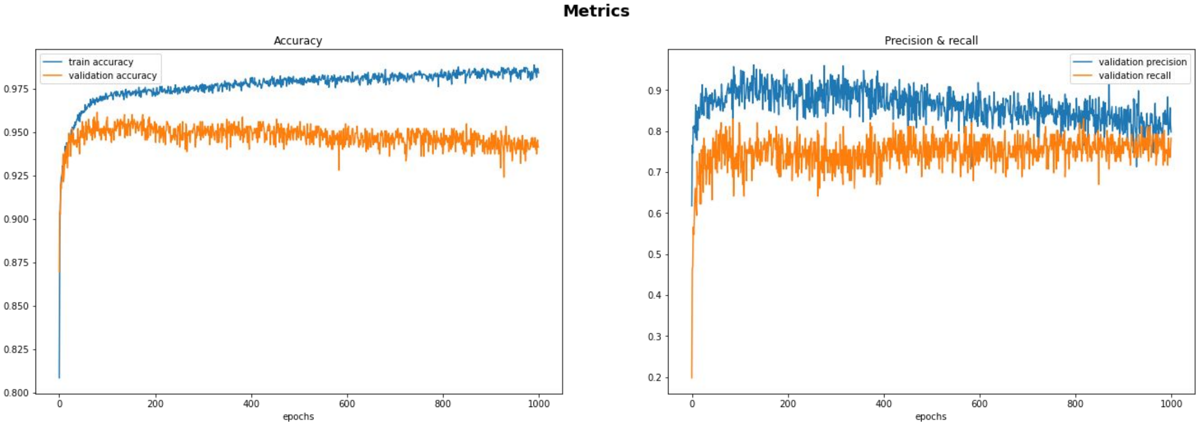
The area below the curve is 0.877 which is close to optimal value 1.0, compared to the poor 0.5 as dice-solution

Simple Neural Network

Feed forward neural network with one hidden layer of 32 neurons (activation relu) and a single output (activation sigmoid)

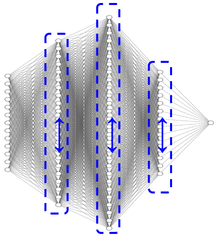
- Accuracy reaches up to 95% and recall and precision converge at approx. 80%





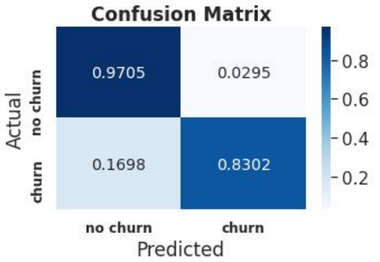
Multiple Hidden Layer Neural Network

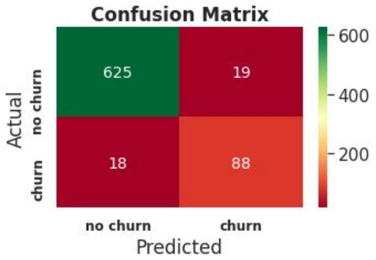
Performance for different hidden layer sizes and more hidden layers with models chosen by different metrics:



hidden layer neurons	Model taken with best parameter	Accuracy	0 Precision	0 Recall	1 Precision	1 Recall
[16]	Loss	0.952000	0.963415	0.981366	0.872340	0.773585
[16]	Accuracy	0.954667	0.966361	0.981366	0.875000	0.792453
[32]	Accuracy	0.953333	0.963470	0.982919	0.881720	0.773585
[128]	Accuracy	0.956000	0.963581	0.986025	0.901099	0.773585
[256]	Precision	0.956000	0.963581	0.986025	0.901099	0.773585
(128, 128)	Recall	0.937333	0.971564	0.954969	0.752137	0.830189
(128, 32, 64)	Accuracy	0.953333	0.964885	0.981366	0.873684	0.783019
(128, 32, 128)	**Recall**	**0.957333**	**0.972222**	**0.978261**	**0.862745**	**0.830189**
(64, 64, 32, 32)	Accuracy	0.956000	0.969278	0.979814	0.868687	0.811321

- An **input-128-32-128-1** fully connected neural network with best recall outperforms the decision tree classifier for churn prediction





Conclusions

1. The classical classifiers lead to broad results/performances.
2. The decision tree with advantage of clear guidance for action and good results: high accuracy 93.6% and high true positives 79.4%
3. Neural Networks with low complexity outperform classical classifiers.
4. results: higher accuracy >95% and high true positives >80% possible

➢ **Thus a neural network is recommended for predictions!**

In general it is possible to further reduce the false positives and reach a better precision for churn predictions but this also reduces the recall rate.

This decision would be done by the business owner.