Tuning your model

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Mark Peterson
Director of Data Science, Infoblox

datacamp

Refresher

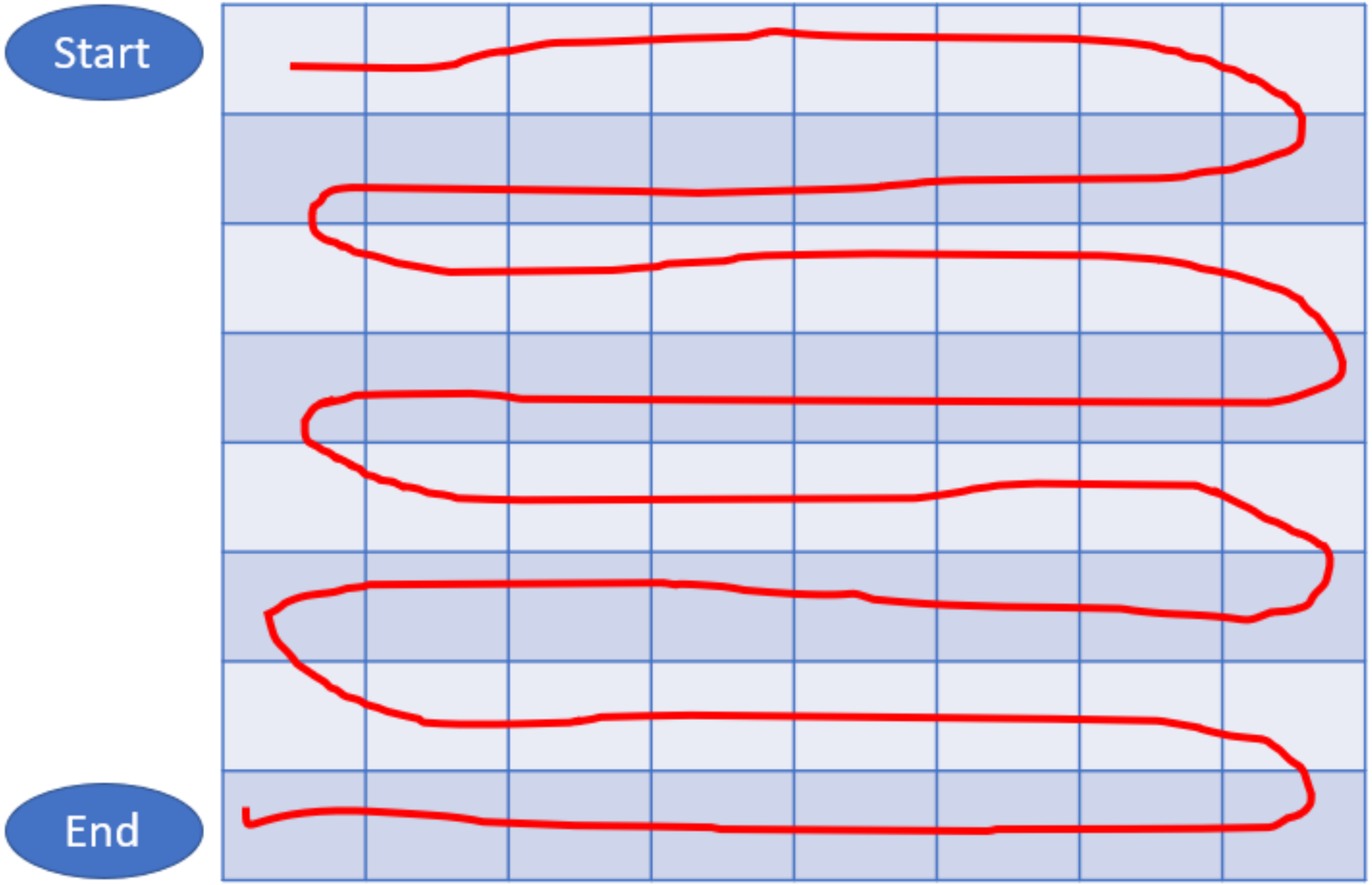
```
from sklearn.svm import SVC

svc = SVC()

svc.fit(telco['data'], telco['target'])
```

```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma='scale', kernel='rbf',
  max_iter=-1, probability=False, random_state=None, shrinking=True,
  tol=0.001, verbose=False)
```

Random forest hyperparameters

Parameter	Purpose
n_estimators	Number of trees
criterion	Quality of Split
max_features	Number of features for best split
max_depth	Max depth of tree
min_sample_splits	Minimum samples to split node
bootstrap	Whether Bootstrap samples are used

Grid search



Grid search in sklearn

```
from sklearn.model_selection import GridSearchCV
param_grid = {'n_estimators': np.arange(10, 51)}
clf_cv = GridSearchCV(RandomForestClassifier(), param_grid)
clf_cv.fit(X, y)
clf_cv.best_params_
```

```
{'n_estimators': 43}
```

```
clf_cv.best_score_
```

```
0.9237923792379238
```

Happy tuning!

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Feature importances

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Mark Peterson
Director of Data Science, Infoblox

datacamp

Feature importances

- Scores representing how much each feature contributes to a prediction
- Effective way to communicate results to stakeholders

- Which features are important drivers of churn?
- Which features can be removed from the model?

Interpretability vs accuracy

- Different models have different strengths
- Need to balance prediction accuracy vs. interpretability

Random forest feature importances

```
random_forest = RandomForestClassifier()


random_forest.fit(X_train, y_train)
```

```
random_forest.feature_importances_
```

```
array([0.02716028 0.02960326 0.13820047 0.0700711  0.03798189 0.04722036
 0.13258838 0.07066312 0.02333901 0.02982636 0.13711014 0.0300661
 0.06351097 0.02860241 0.0382932  0.05200864 0.04375429])
```

Let's practice!

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

datacamp

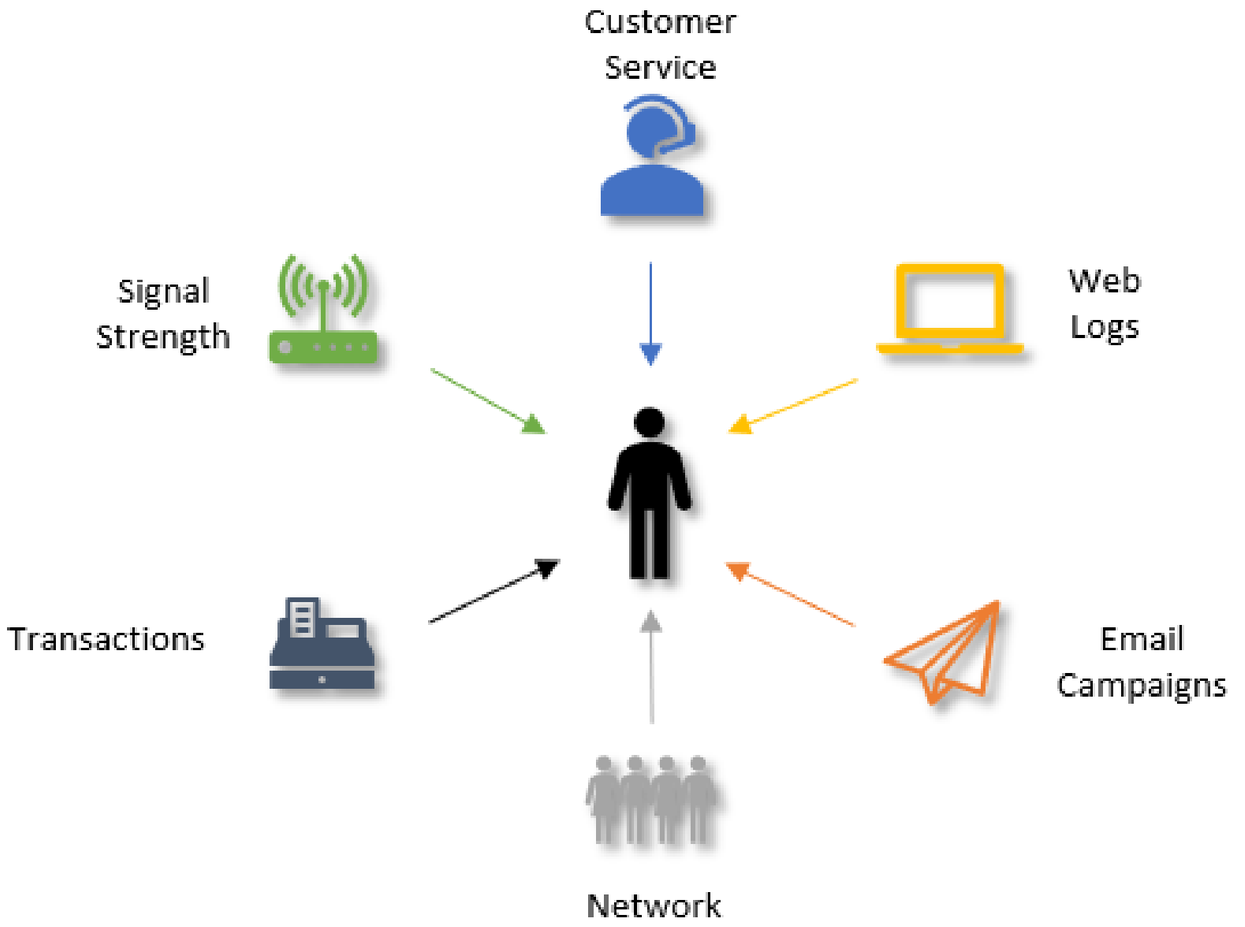
Adding New Features

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Mark Peterson
Director of Data Science, Infoblox

datacamp

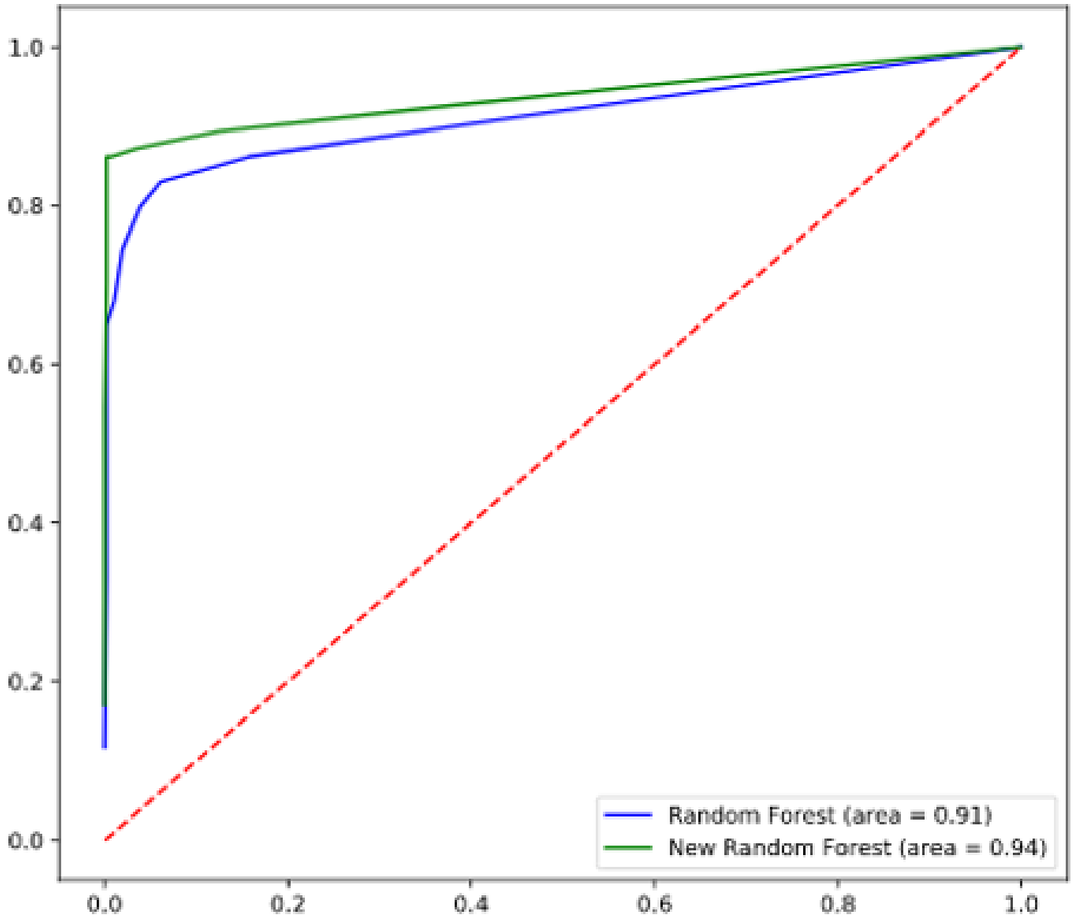
Additional Data Sources



Churn Features

- Region Code
- Total Charges
- Total Minutes
- Minutes Per Call
- Cost Per Call
- Total Calls

Model Improvement



Benefits | Costs

- Benefits
  - Improved Return on Investment
  - Decreased Costs
  - Increased Performance
- Costs
  - Increased Complexity
  - Increased Resources
  - Increased Time to Operationalizing

Let's practice!

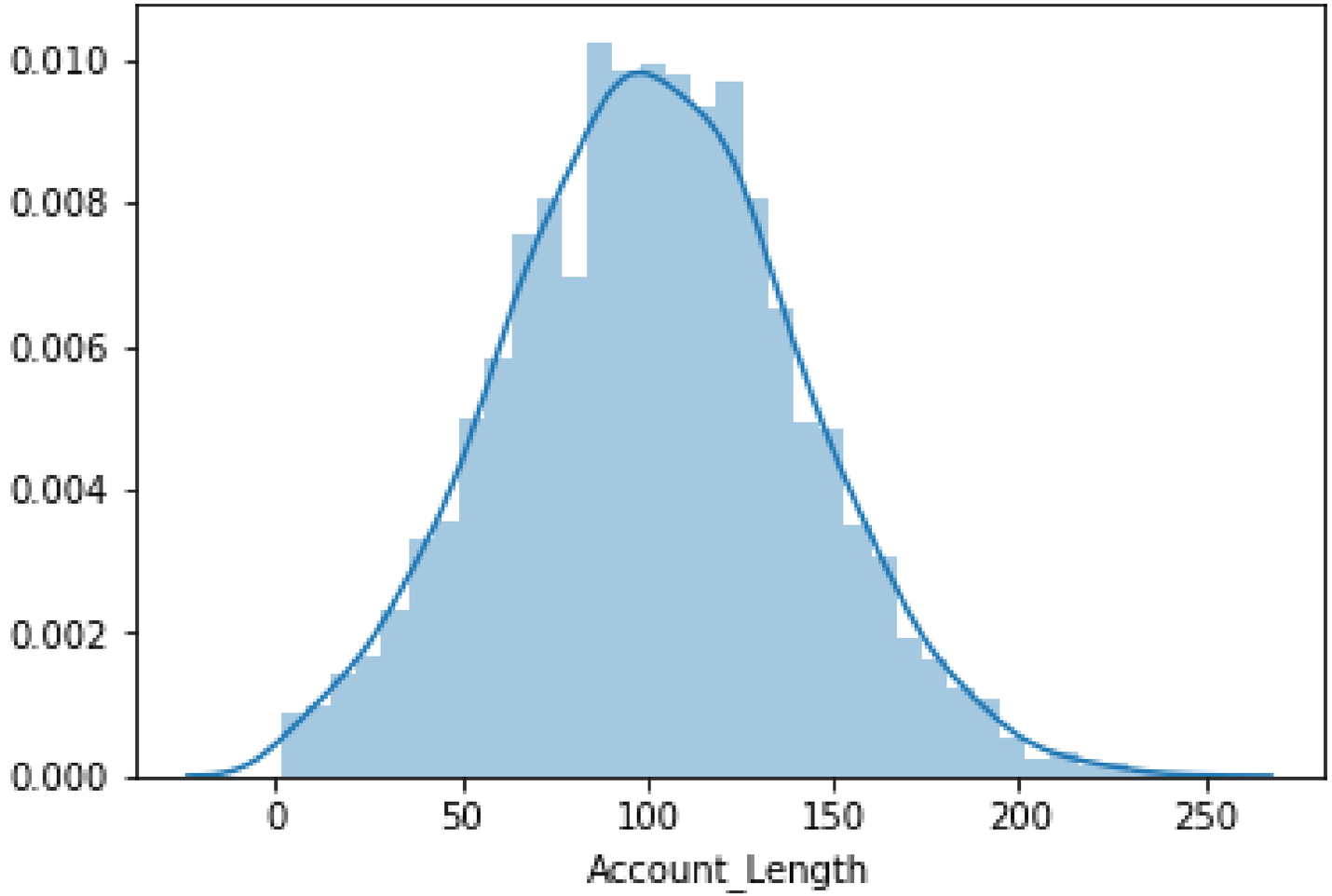
MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Final thoughts

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

Mark Peterson
Director of Data Science, Infoblox

- Defined customer churn
- Exploratory data analysis



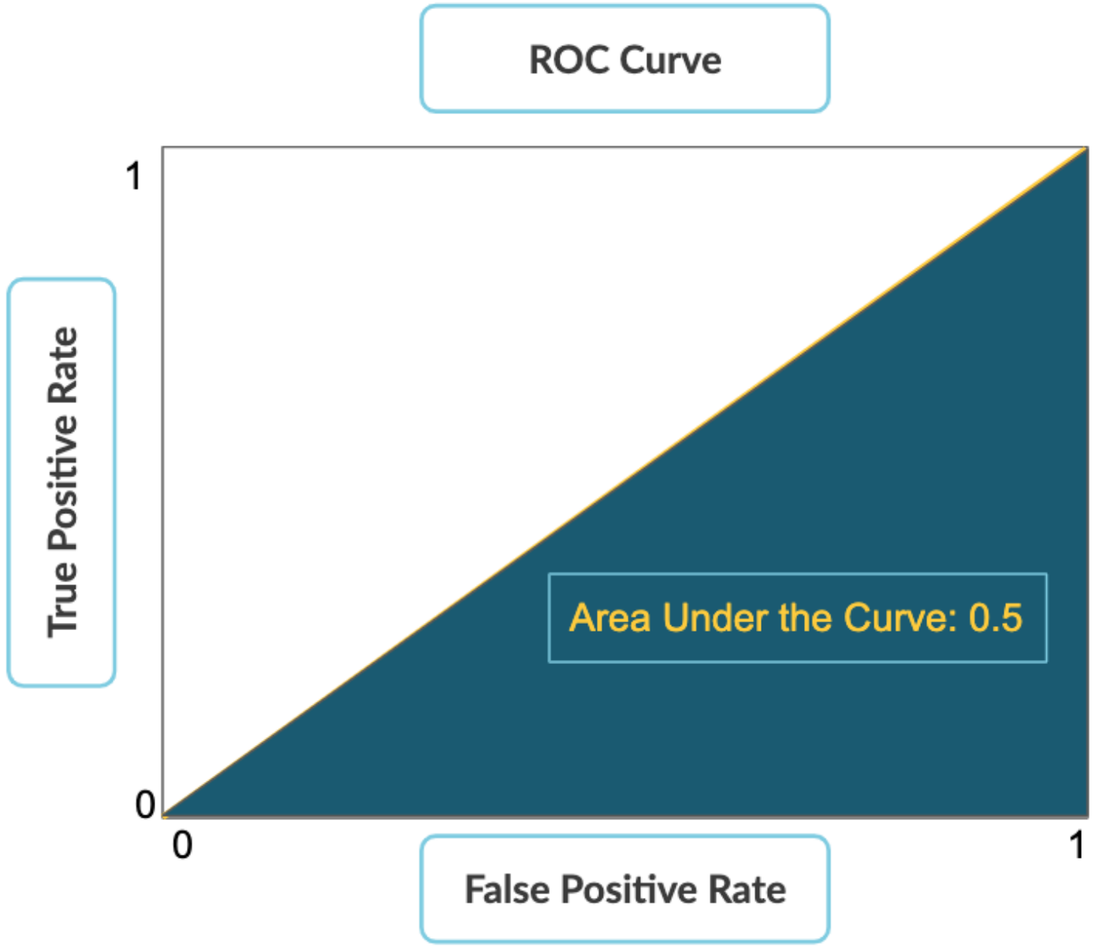
- Drop unnecessary features

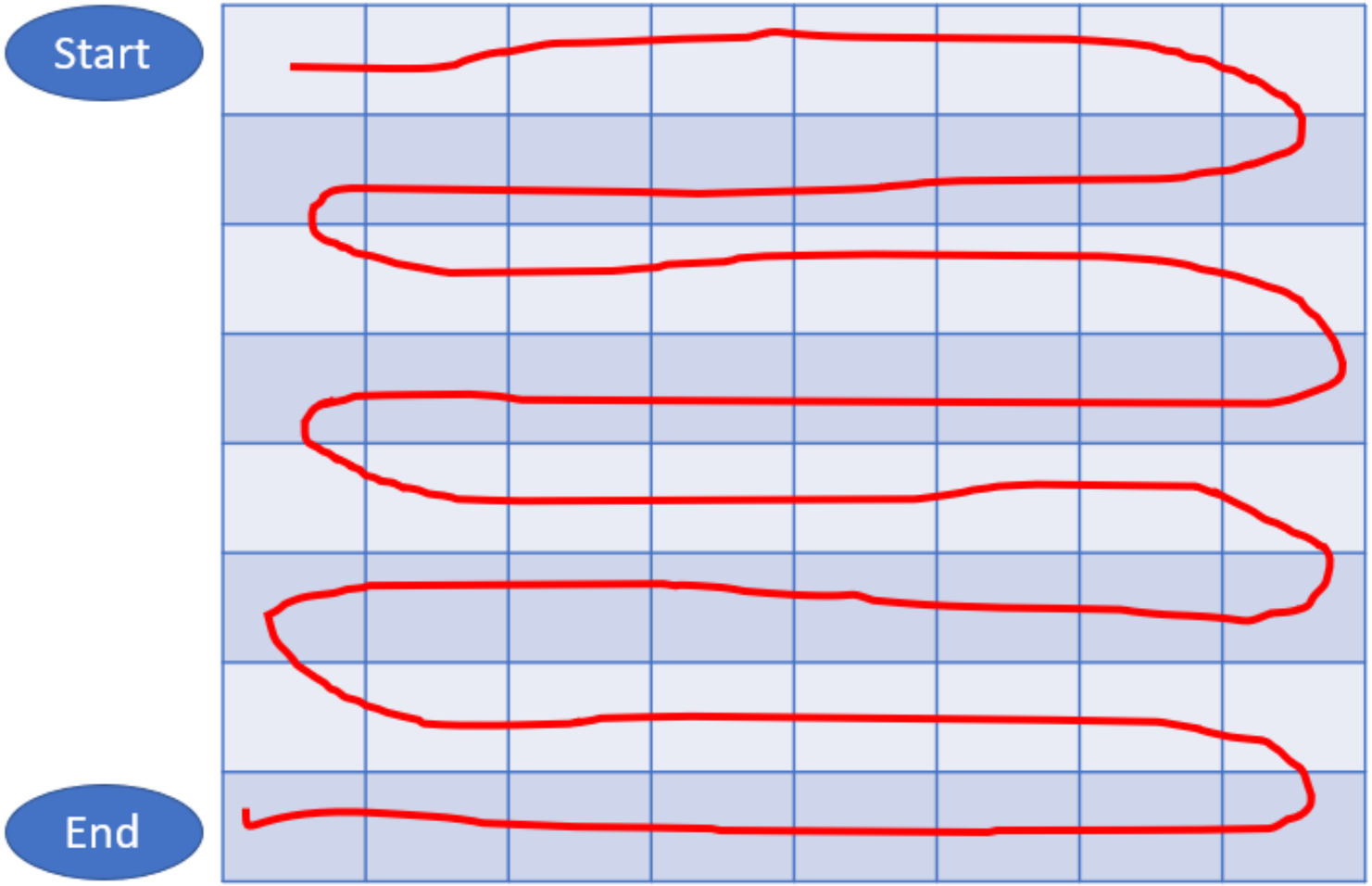
State
KS
OH
NJ
OH

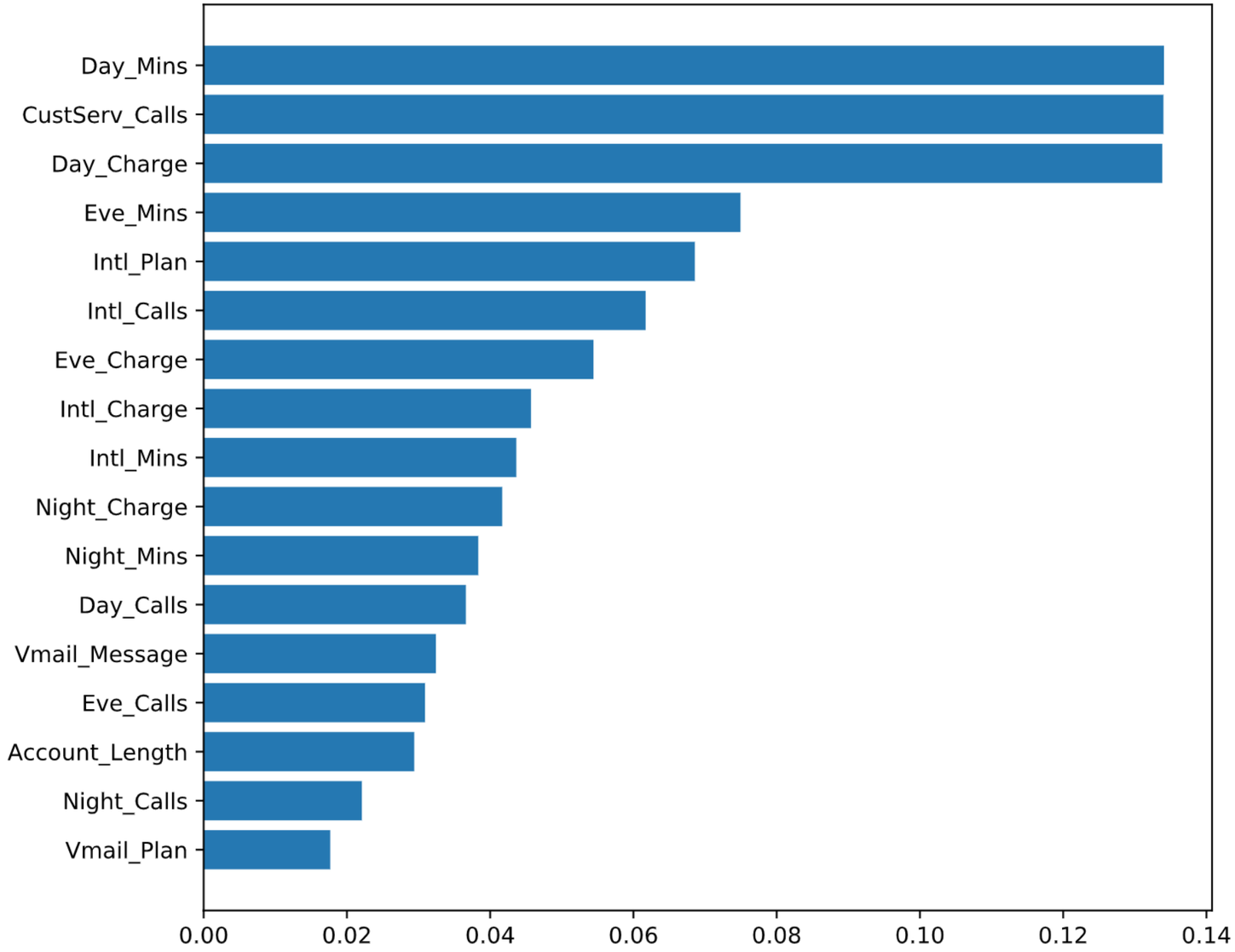
State_KS	State_OH	State_NJ
1	0	0
0	1	0
0	0	1
0	1	0

- Feature scaling

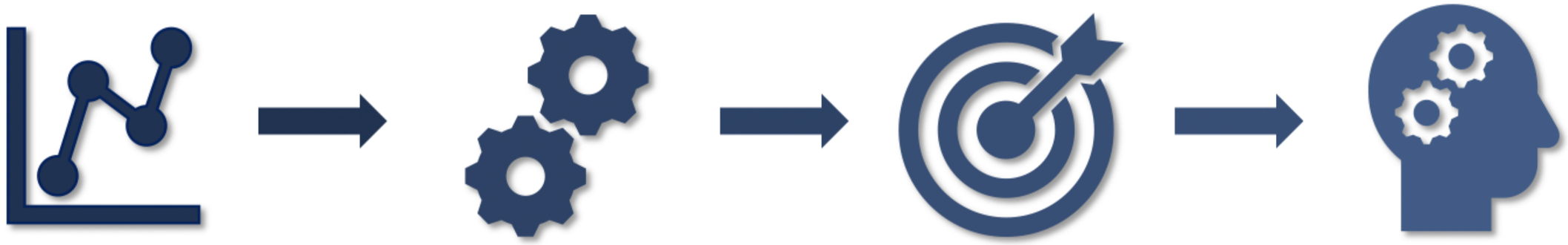
- Making predictions
- Training and testing sets




Start
End


Day_Mins
CustServ_Calls
Day_Charge
Eve_Mins
Intl_Plan
Intl_Calls
Eve_Charge
Intl_Charge
Intl_Mins
Night_Charge
Night_Mins
Day_Calls
Vmail_Message
Eve_Calls
Account_Length
Night_Calls
Vmail_Plan
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14

Churn Workflow



Recommended next steps

- Exploratory Data Analysis in Python
- Designing Machine Learning Workflows in Python

Additional resources

- Kaggle competitions
- Coursera advanced business analytics specialization

Great Work!

MARKETING ANALYTICS: PREDICTING CUSTOMER CHURN IN PYTHON

datacamp