Employee Churn

Wahrscheinlichkeit der Mitarbeiterfluktuation im BASF Service Hub Berlin

Eric Beier, Lucas Oldenburg, Martin König

30.01.2022

Agenda

1. Problemstellung
2. Datengrundlage
3. Deskriptive Analyse
4. Modellauswahl
5. Modellgüte
6. Blick in den Code
7. Modellvergleich
8. Fazit

Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences

BASF
We create chemistry

1. Problemstellung

12-13 % Turnover jährlich

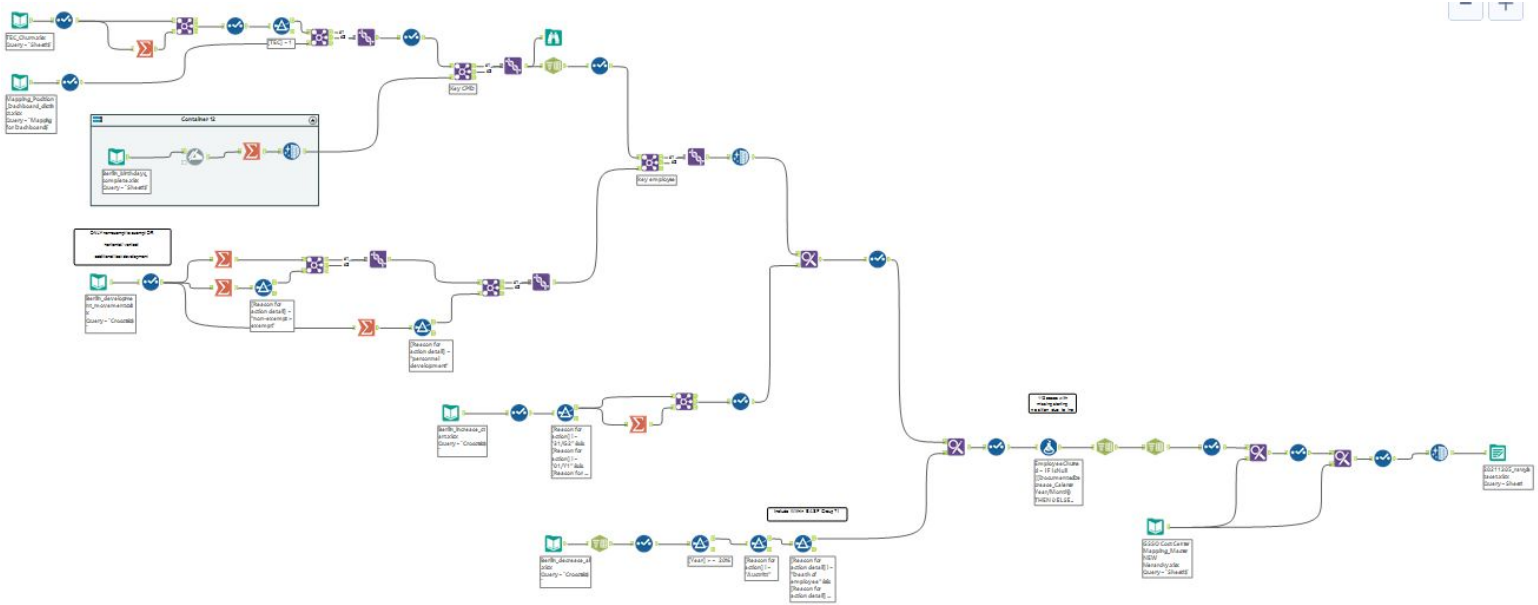
➔ Hohe Fluktuation im BASF Service Hub Berlin

➔ **rechtzeitig** Maßnahmen einleiten

➔ Nicht im Scope: Sabbatical, Elternzeit, etc.

2. Datengrundlage

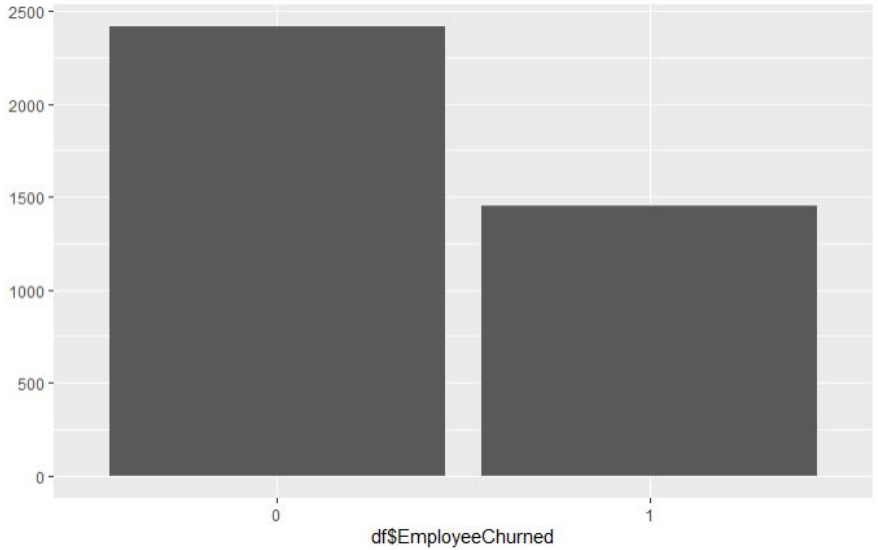
Daten verbinden und erste Features entwickeln in **Alteryx**



2. Datengrundlage

- circa 4000 Datenpunkte
- 41 Merkmale
  - Numerisch: 5
  - Kategorien: 32
  - Daten: 4
- Systemeinführung 2006
- betrachten letzten 6 Jahre
- personalisierte Daten herausgenommen

aktiv: 0 | gegangen: 1



3. Deskriptive Analyse

- bereinigt: 14
  - Kostenstelle, Position, Stelle Employee Sub-category...
- verändert: 5
  - Nationality, Last_Development_Calendar_Year...
- angelegt: 5
  - Nationality_Classification, Month of Service...
- geclustert: 3
  - Nationality_Classification, Development...

-> final 23 Merkmale

4. Modellauswahl

- Supervised Learning - Daten mit Label (Arbeitskraft gegangen/ noch aktiv)

- Decision Tree (Martin)
- Logistic Regression (Lucas)
- Support Vector Machine (Eric)

5. Modellgüte

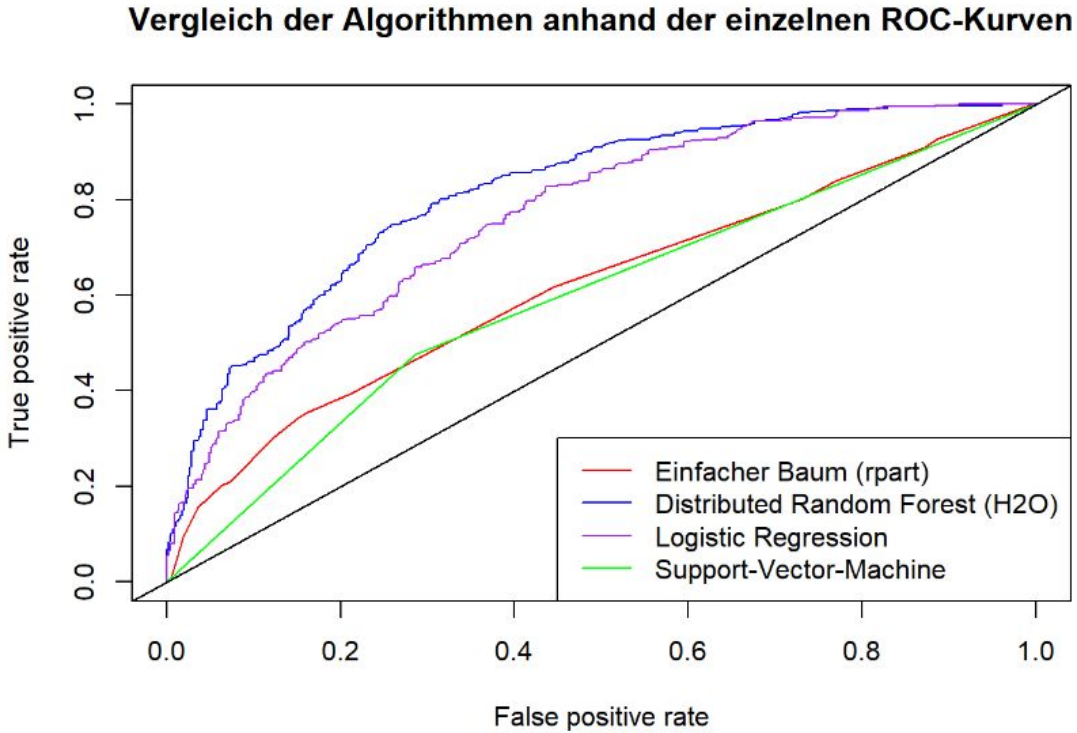
Güte: Confusion Matrix und Kostenfunktion

- False Positive - Kosten durch Aufwand für Manager und Arbeitskraft sowie mögliche “vermeidbare” Maßnahmen zum Halten der Arbeitskraft
- False Negative - Kosten durch Rekrutierung und Belastung für das Team

	Mitarbeiter:in geht	Mitarbeiter:in bleibt
Vorhersage Mitarbeiter:in geht	0€	500€
Vorhersage Mitarbeiter:in bleibt	3000€	0€

6. Blick in den Code

7. Modellvergleich

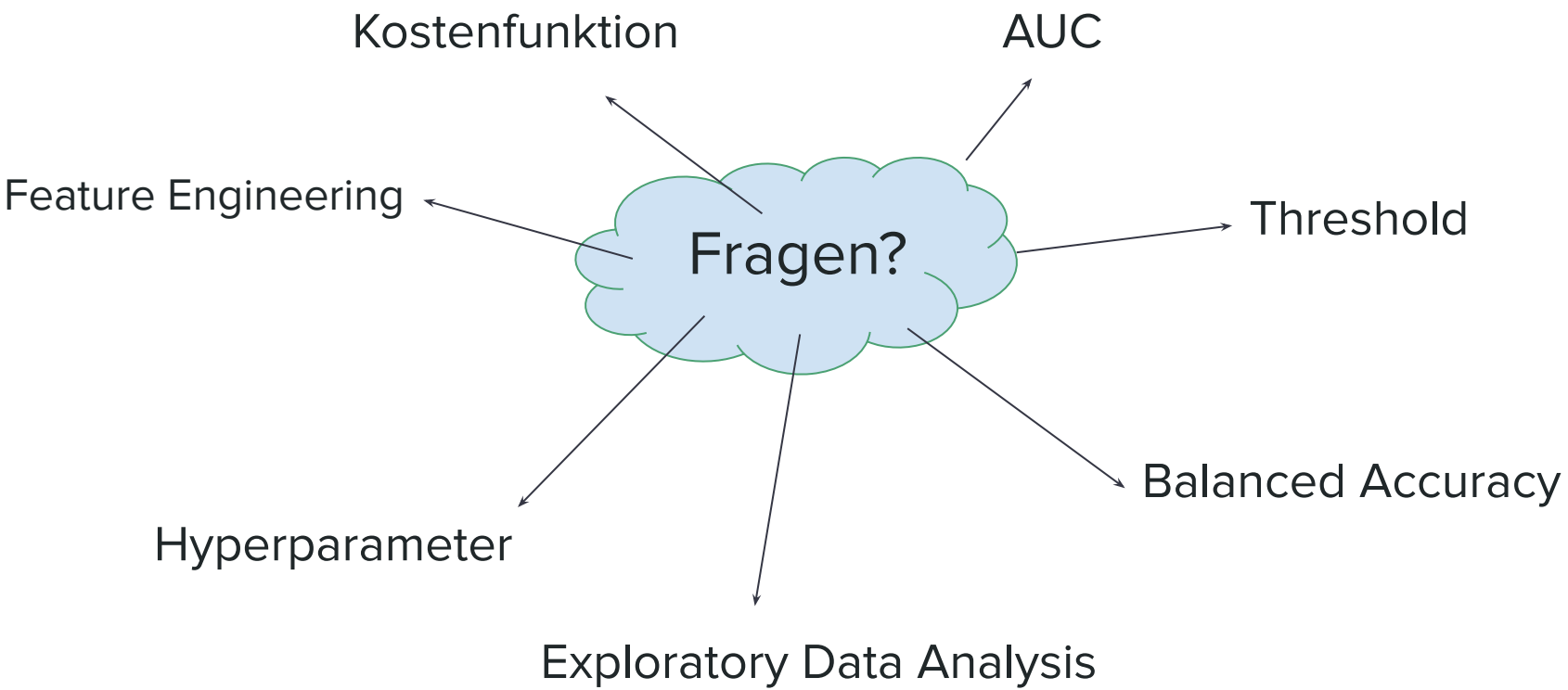


7. Modellvergleich

	False-Positive-Rate	Acurracy
Distributed Random Forest	10	65
Logistic Regression	7	59
Support Vector Machine	29	62

	Cost
Distributed Random Forest	240.000
Logistic Regression	233.500

8. Fazit und Ausblick


Kostenfunktion
AUC
Feature Engineering
Fragen?
Threshold
Balanced Accuracy
Hyperparameter
Exploratory Data Analysis

Präsentationsverteilung

Anmoderation - Eric

1. Problemstellung - Martin
2. Lösungsansatz - Martin
3. Datengrundlage - 1. Martin ab Datenbearbeitung Eric
4. Deskriptive Analyse - 1. Eric dann Lucas
5. Modellauswahl und -güte - Lucas
6. Ausblick - Lucas

Fazit: Datenstruktur (Format) - 80 % data wrangling

Erkenntnis

Nutzen -> optimieren () -> Hyperparameter