

Projekt-Bericht und Dokumentation im Kurs
Ausgewählte Kapitel sozialer Webtechnologien (SoSe15)
Machine Learning

Claus Holland Richard Remus

25. Juli 2015

Contents

1	Aufgabenstellung	3
2	Explorative Datenanalyse	3
3	Data Wrangling	4
4	Erstellung der Modelle	4
4.1	Logistic Regression	4
4.1.1	Cross Validation	4
4.2	Support Vector Machine	4
4.2.1	Grid Search	4
4.3	Random Forest	4
5	Performancevergleich der Modelle	4
6	Ergebnisse	4
7	Fazit	4

1 Aufgabenstellung

2 Explorative Datenanalyse

Die Zielklasse ist *SeriousDlqin2yrs*, welche Aussage darüber trifft, ob der Kreditnehmer innerhalb zweier Jahre nach Vertragsschluss in ernsthafte Zahlungsschwierigkeiten kam. Zunächst betrachten wir die Verteilung dieser Klasse auf den Trainingsdaten.

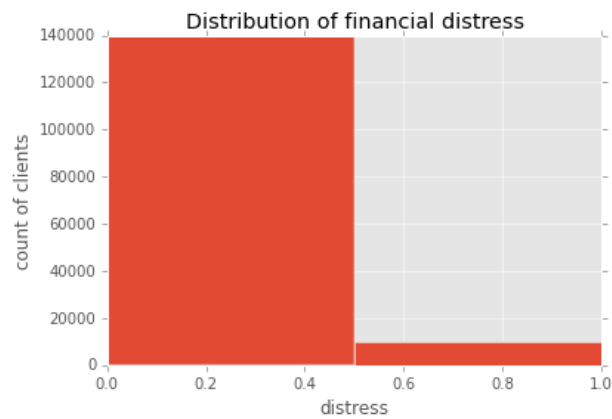


Figure 1: Verteilung von SeriousDlqin2yrs

3 Data Wrangling

4 Erstellung der Modelle

4.1 Logistic Regression

4.1.1 Cross Validation

4.2 Support Vector Machine

4.2.1 Grid Search

4.3 Random Forest

5 Performancevergleich der Modelle

6 Ergebnisse

7 Fazit