

# Dokumentation der Modelle

## Contents

Training .....	1
Vergleichsmaß .....	1
Die Modelle .....	2
<i>Lineare Regression</i> .....	2
<i>Lasso Regression</i> .....	2
<i>Ridge Regression</i> .....	2
Zusammenfassung .....	2

## Training

Die Daten, welche für das Training verwendet wurden, sind die folgenden:

- Flughafengesellschaft
- Flugnummer
- Abflugsflughafen
- Flughafennamen
- Stadt, Region, Land und Kontinent des Flughafens
- Erste Schätzung des TOWs

Die Vorhersage soll das ATOW möglichst genau vorhersagen.

Die Daten setzten sich aus allen Flügen zusammen, von denen diese Daten vorhanden sind. Das sind 7.100 Flüge. Es wurden diese Daten ausgewählt, da sie zu Beginn eines Fluges immer vorliegen.

## Vergleichsmaß

Um die Genauigkeit der Modelle beurteilen zu können, wird ein Vergleichsmaß gebraucht. Als solches wurde die erste Schätzung verwendet, die im aktuellen Planungsprozess erstellt wird.

Für diese Schätzung gelten folgende Daten:

Erste Schätzung:

Mean Squared Error	62752796443.978
R <sup>2</sup> Score	-0.06404224017588755

Das Ziel der Modelle ist es, den Mean Squared Error möglichst gering zu halten, da das bedeutet, dass die Werte genauer an den wirklichen Werten liegt.

Der R<sup>2</sup> Score hingegen sollte möglichst nahe der 1 sein. Diese Metrik misst, wie exakt auch besonders große und kleine Werte vorhergesagt werden und verhindert, dass immer nur der

Durchschnittswert vorhergesagt wird. Dieser hätte zwar potenziell einen geringen Mean Squared Error, jedoch sind gerade die Ausreißer für die Planung der Flüge interessant und sollten deswegen genau vorhergesagt werden.

## Die Modelle

### *Lineare Regression*

Mean Squared Error	58963409644.03798
R <sup>2</sup> Score	0.00021095407180948822

### *Lasso Regression*

Mean Squared Error	58963045701.38076
R <sup>2</sup> Score	0.0002171251172804478

### *Ridge Regression*

Mean Squared Error	58947820780.84952
R <sup>2</sup> Score	0.00047528028273224887

Der Mean Squared ist bei der Ridge Regression etwas niedriger als bei den anderen beiden Modellen. Auch ist der R<sup>2</sup> Score mit rund 0.00048 zwar sehr gering, jedoch besser als der der anderen Modelle. Somit ist das Modell mit der Ridge Regression das beste Modell, um das ATOW vorherzusagen. Die Vergleichsmaße sind zu der ersten Schätzung für alle Modelle besser (Mean Squared Error geringer und R<sup>2</sup> Score höher).

## Zusammenfassung

Keines der Modelle kann das ATOW besonders gut vorhersagen. Die Vorhersagen sind zwar besser als die aktuelle erste Schätzung, jedoch ist die Abweichung von den realen Werten immer noch sehr hoch. Um die Modelle zu verbessern, könnte einerseits eine Hyperparametertuning gemacht werden. Andererseits könnten auch weitere Daten für das Training verwendet werden. Diese könnten sein:

- Flugzeugtyp des Fluges und dessen Basisgewicht
- Größe der Flugzeuge an den jeweiligen Flughäfen
- Verschiedene Klassifizierung von Flügen (z.B. nach Flugstrecke)

Mit diesen Daten könnten die Modelle potenziell besser werden.