

# Report on Regression Project for Machine Learning 2

Maluna Menke, Ari (Sara) Wahl

2023-11-17

## Inhaltsverzeichnis

2. Descriptive Analysis of the Dataset . . . . .	1
3. Modellierung der Parteiwahl durch ein geeignetes Regressionsmodell . . . . .	1
3.1 Hypothesen, die dem Regressionsmodell zugrunde liegen . . . . .	1
3.1.1 Hypothesen und Signifikanzniveau . . . . .	1
3.1.2 Hypothesen bezüglich der Variablen . . . . .	1
3.1.3 Hypothesen bezüglich der Interaktion zwischen den Variablen . . . . .	1
3.1.2 Kodierung . . . . .	1
4. Interpretation . . . . .	1
5. Final Discussion . . . . .	1
6. References . . . . .	1

---

*PetFinder.my Adoption Prediction* challenge. “PetFinder.my has been Malaysia’s leading animal welfare platform since 2008, with a database of more than 150,000 animals”[1].

## 2. Descriptive Analysis of the Dataset

Summary:

## 3. Modellierung der Parteiwahl durch ein geeignetes Regressionsmodell

### 3.1 Hypothesen, die dem Regressionsmodell zugrunde liegen

#### 3.1.1 Hypothesen und Signifikanzniveau

#### 3.1.2 Hypothesen bezüglich der Variablen

In unser Modell lassen wir nun folgende, mithilfe sachlogischer Zusammenhänge erarbeiteter und in der deskriptiven Analyse erarbeiteten, Hypothesen einfließen. Indem wir die entgegengesetzt lautenden Nullhypothesen in unserem Regressionsmodell prüfen und, falls unsere Hypothesen zutreffen, die Nullhypothesen gegebenenfalls zum gewählten 5% Signifikanzniveau verwerfen können. Unsere Hypothesen lauten:

#### 3.1.3 Hypothesen bezüglich der Interaktion zwischen den Variablen

Die deskriptive Analyse lässt außerdem folgende **Interaktionsterme** vermuten, da es einen (systematischen) Zusammenhang zwischen diesen Merkmalen zu geben scheint. Hier lauten die diesbezüglichen Hypothesen:

#### 3.1.2 Kodierung

In unserem multinomialen Modell ist die Referenzkategorie der Responsevariablen *vote* die konservative Wahlentscheidung.

Das optimierte Modell weist nochmal ein etwas besseres AIC auf

Mithilfe eines Likelihood-Ratio Tests wird überprüft welches Modell das bessere ist.

Das optimierte Modell ist signifikant besser als das erarbeitete zum festgelegten Signifikanzniveau, die Nullhypothese des LR-Tests, dass das optimierte Modell nicht besser ist als das komplexere, kann verworfen werden. Wir stellen außerdem fest: Beide Modelle weisen einen signifikanten Erklärungsgehalt auf, sie sind besser als das Nullmodell.

## 4. Interpretation

Das Ergebnis des Pseudo-R-Quadrats liegt in einem Bereich mittlerer Modellgüte.

## 5. Final Discussion

## 6. References

[1] <https://www.kaggle.com/competitions/petfinder-adoption-prediction/overview>