

Abgabe der Challenge zu “Anwendungen der KI”

TEAMNAME:

Die Mannschaft

TEAMMITGLIEDER:

Nr.:	Name:	Matrikelnummer:	E-Mail:
1	Julian Herrmann	1955722	ueesu@student.kit.edu
2	Felix Schüber	1927759	felix.schueber@web.de
3	Marco Spraul	2055452	uuqmj@student.kit.edu
4	Oliver Becker	1940529	uketr@student.kit.edu
5	Martin Thoma	1954967	uxebx@student.kit.edu

KURZBESCHREIBUNG VORGEHEN:

Um die gegebene Challenge, das Klassifizieren eines “Review”-Text-Datensatzes in positive und negative Reviews, mit Hilfe der Anwendung von existierenden Methoden aus dem Text Mining und Maschinellen Lernen, zu lösen, sind wir wie folgt vorgegangen. Nach der Beschaffung der Trainingsdaten wurden diese eingelesen und mit den noch fehlenden Labels 1 (positive Review) und 0 (negative Review) versehen und anschließend zusammengeführt. Für alle weiteren Schritte wurde auf die weit bekannte “Sci Kit Learn” Python-Bibliothek zurückgegriffen. Im ersten Schritt wurde das Gesamtset in Training und Test im Verhältnis 2 zu 1 (66.7% Training, 33.3% Test) aufgeteilt. Aufgrund guter Klassifizierungserfahrungen von Reviews, wurde der Test Satz großzügig bestimmt, auch um das Overfitting minimal zu halten. Im Anschluss wurden mithilfe des sklearn packages in einer Pipeline sowohl unerwünschte Satzzeichen gelöscht, als auch eine SVM auf das Model gefittet. Mit dieser Methode konnte eine accuracy von 0.86 erreicht werden. Die Konfusionsmatrix lautet wie folgt:

```
[[444  82 ]  
 [56  408]]
```