

Gemeinderanking Part 2: AI Ranking

1) Idee

In Assignment 1 haben Sie eine Rangliste der Gemeinden ihres Kantons erstellt. Der Kantonsrat war begeistert von Ihrer Arbeit und möchte mit den Daten weitere Forschung betreiben. Das Ziel ist es, ein einfaches Machine Learning Modell zu erstellen, welches die von Ihnen aufgestellte Rangliste möglichst gut abbilden kann, ohne dass nochmals die gleichen Features verwendet werden müssen.

2) Aufgabenstellung

Wählen Sie beliebig eine der 3 Ranglisten aus Aufgabe 1 aus. Der Score der einzelnen Gemeinden (zwischen 0 und 100) ist nun die Ground Truth, die es nachzubilden gilt mit **neuen Features**. In dem zur Verfügung gestellten Jupyter Notebook **assignment2_template.ipynb** können Sie die Daten einlesen und haben zwei Machine Learning Modelle zu Ihrer Verfügung.

Der zur Verfügung gestellte Code erstellt Ihnen eine **Baseline** mit zufällig generierten Daten. Nun geht es darum, dass Sie neue Features erstellen und diese dem Machine Learning Modell übergeben, um die beiden Metriken zu verbessern. Wichtig: Die Parameter des Modells und der Metriken sollen dabei nicht angepasst werden! Für gute Scores werden aber für die beiden Modelle unterschiedliche Feature Engineering Schritte nötig sein.

Als Grundlage für die Features können Sie alles verwenden, was Sie in Assignment 1 gesammelt haben, ausgenommen alle Kriterien, welche für das Erstellen der gewählten Rangliste verwendet wurden, oder direkt von diesen abhängig sind. Wenn z.B. eines der ursprünglichen Kriterien die Anzahl Hühner in der Gemeinde war, so ist jetzt sowohl Hühner pro m² als auch Anzahl Hühnerställe verboten. Weitere Datenquellen wie z.B. das beiliegende Dokument GemeindePortraits2021.xlsx dürfen ebenfalls verwendet werden.

Zusätzlich müssen Sie ein neues Feature von **geographischer Art** ihrem Datensatz hinzufügen. Dies kann beispielsweise eine Auswertung von OpenStreetMap oder Geo.Admin sein, oder vergleichbare Quellen.

Erstellen Sie einen **kurzen Bericht** (ca. 1 Seite A4) welcher die Resultate zusammenfasst. Im Gegensatz zu dem Bericht bei Assignment 1 richtet sich dieser Bericht nun an ein wissenschaftliches Publikum und soll aufzeigen, wie ihr Feature Engineering zu einem verbesserten Machine Learning Modell für das Schätzen von Gemeindequalität geführt haben.

3) Abgabe

Packen Sie Ihre Daten in ein .zip File mit dem Namen «**dawr_2_vorname_nachname.zip**». Dieses zip File soll enthalten: Ein Pdf mit dem Bericht (1 bis 2 Seiten), ein Jupyter Notebook mit dem Code sowie benötigte Files mit Daten falls nötig. Maximum 5 MB, wenn die Daten grösser sind, können Sie auch einen Link im zip oder im Notebook (NICHT im Mail) hinterlegen. Das Jupyter Notebook muss vollständig ausführbar sein. Verwenden Sie keine absoluten Pfade im Code.

Senden Sie das .zip bis am **7. Mai, 23:59:59** an lucas.broennimann@fhnw.ch

Code, welcher nicht von Ihnen selbst ist, muss entsprechend markiert sein. Plagiate werden mit der Note 1 bewertet. Nichteinhalten der Termine und Abgabebedingungen wird mit mindestens einer Note Abzug bestraft.