

# Optimierung von Convolutional Neural Network Modellen zur Prediction von Wasserversorgung, Nahrungsmittelversorgung oder Wohlstandsindizes

Abschlussbericht im Modul DIS22 - ML Projekt  
Studiengang Data and Information Science (B. Sc.)  
an der Fakultät für Informations- und Kommunikationswissenschaften  
der Technischen Hochschule Köln

vorgelegt von:	Felix Krause und Roman Poledniok
Semester:	6. Semester
Prüfer:	Prof. Dr. Gernot Heisenberg
Zweitprüfer:	Sven Wöhrle, MSc.

Köln, den 21.07.2022

## Inhaltsverzeichnis

<b>Inhaltsverzeichnis .....</b>	<b>I</b>
<b>1. Testroutinen .....</b>	<b>II</b>
1.1 Learning-Rate, Decay und Autoadjustment .....	II
1.2 Ergebnis .....	II
<b>2. DHS-API .....</b>	<b>III</b>
2.1 API Basics und Advanced Queries .....	III
2.2 Ergebnis .....	III
<b>3. GEO-Visualisierung .....</b>	<b>IV</b>
3.1 GEO-Visualisierung .....	IV
<b>4. Lessons Learned.....</b>	<b>V</b>
4.1 Felix Krause:.....	V
4.2 Roman Poledniok: .....	V

## 1. Testroutinen

### 1.1 Learning-Rate, Decay und Autoadjustment

Wir haben einen Learning Rate Decay eingebaut, da somit das lokale Minimum beim Trainieren des Neuronalen Netzes besser gefunden werden kann. Wenn man kein Learning Rate Decay benutzt, kann es dazu kommen, dass das lokale Minimum verpasst wird und man eine schlechtere Accuracy erhält. Durch das stetige Senken der Learning Rate, wird sich somit besser an das lokale Minimum angenähert.

Ebenso hilft die Auto Adjust Funktion dabei, das Lokale Minimum zu finden. Die Auto Adjust Funktion ist ein vordefiniertes Framework, das die Learning Rate in jeder Epoche anpasst. Damit soll das Lokale Minimum besser gefunden werden.

### 1.2 Ergebnis

Durch den Learning Rate Decay und die Auto Adjust Funktion, konnten wir eine bessere Accurcay beim Training des Neuronalen Netzes bekommen. Die folgenden Testdurchläufe zeigen, dass bei unveränderten anderen Hyperparametern, ein Learning Rate Decay von 0.001 und eingeschaltetem Auto Adjust am besten ist.

Epoch	LR	Test-Accuracy
5	1e-4	0.3694
5	0.01	0.6662
5	0.1	0.6651
10	0.01	0.6662
10	0.1	0.6595
10	0.3	0.6584
20	0.001	0.6595
50	0.001	0.6987

## 2. DHS-API

### 2.1 API Basics und Advanced Queries

Im Rahmen unserer zweiten Aufgabe, der Erschließung der DHS-API zum automatisierten Auslesen der Metadaten der Indikatoren auf Household Level, haben wir die Herangehensweise zum Benutzen dieser API getestet.

Folgende Informationen waren dabei für uns relevant:

- Metadaten auf Household Level
- GPS-Daten

Hierfür wurde ein API-Key beantragt über welchen wir zusätzliche Vorteile als registrierter API-Partner bekommen haben. So ist ein frühzeitiger Zugang zu Beta-Versionen der API möglich und die Möglichkeit größere Anzahl von Ergebnissen von der API anzufordern.

Die Daten konnten in vordefinierten Basics der DHS-API oder anhand einer Advanced-Query abgefragt werden. Wir haben beide Abfragemethoden getestet und diese nach Survey-art, Landesebene und anhand der „List of Indicators“ gefiltert und getestet.

### 2.2 Ergebnis

Leider konnten wir keine Ergebnisse mit der API verzeichnen. Dies lag daran, dass die Daten nicht pro Ort vorhanden sind, sondern nur als Polygon pro Provinz. Demensprechend ergab die Abfrage nur Daten auf Provinzebene, welche für uns unbrauchbar war, da die Umfragen auf Stadtebene passiert sind.

Ein weiteres Problem war, dass es eine sehr große Anzahl an Label gab, die nicht gebündelt waren. Diese müssten in einem vorherigen Schritt noch zusammengefasst werden, damit man eine allgemeine Abfrage schreiben kann.

Sollte die API kleingranulare Daten in Zukunft aufweisen, könnten diese für unsere Analyse relevant sein.

## 3. GEO-Visualisierung

### 3.1 GEO-Visualisierung

Ziel dieser Visualisierung ist es, die einzelnen Umfragewerte zu "elementary access", "improved access", "limited access" und "other" auf einer Karte darzustellen. Die Umfragewerte stellen die Label dar, welche wir im Endeffekt über matplotlib visualisieren möchten.

Die nötigen GPS-Daten wurden uns über Teams bereitgestellt. Diese befanden sich in einem Shapefile und konnten mithilfe von geopandas in Python ausgelesen und in einem Dataframe gespeichert werden.

Als nächster Schritt konnten die Neuen Survey Label, welche innerhalb einer anderen Gruppe festgelegt werden über eine Excel-Tabelle ebenfalls in ein Dataframe geladen und mit den GPS-Daten gemerged werden.

	DHSClust	DHSREGNA	LATNUM	LONGNUM	geometry	elementary_access	improved_access	limited_access	other
0	101	Niassa	-13.285618	35.231138	POINT (35.23114 -13.28562)	15.0	2.0	11.0	NaN
1	102	Niassa	-13.322655	35.255217	POINT (35.25522 -13.32266)	16.0	8.0	3.0	NaN
2	103	Niassa	-13.271620	35.196597	POINT (35.19660 -13.27162)	27.0	NaN	1.0	NaN
3	104	Niassa	-14.811983	36.563687	POINT (36.56369 -14.81198)	10.0	18.0	NaN	NaN
4	105	Niassa	-14.788214	36.514077	POINT (36.51408 -14.78821)	16.0	12.0	NaN	NaN

Abbildung 1: Auszug der gemerged Daten

Nachdem alle Daten erfolgreich gemerged, gefiltert und sortiert werden konnten, wurden diese über einen Plot auf einer Karte von Mosambik visualisiert werden.

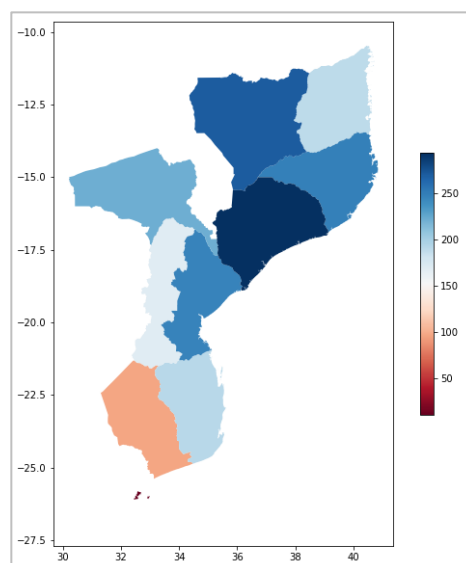


Abbildung 2: Darstellung der elementary\_access Surveys in Mosambik

## 4. Lessons Learned

### 4.1 Felix Krause:

Innerhalb des Moduls DIS22, für welches ich mich aufgrund des Themas Neuronale Netzwerke entschieden habe, konnte ich über diese und die praktische Anwendung zur Prediction von z.B. Wasserversorgungen vieles lernen. So hat sowohl unsere erste Aufgabe im Rahmen der Erstellung einer Testroutinen als auch das spätere Auseinandersetzen mit der Learning-Rate sowie dem Decay dazu beigetragen mich erneut tiefer mit dem Thema CNNs auseinander zu setzen.

Nachdem unsere Aufgabe dann um die API erweitert wurde, konnte ich mein Wissen über REST-APIs, welches ich innerhalb meines Praxissemesters bereits vertieft habe, anwenden. Das Arbeiten mit Geometry-, POINT- und GPS-Daten war jedoch in unserem kleinen Team ein neues und hat viel Zeit gekostet, da das Arbeiten mit solchen Daten für uns ein neues war.

Dadurch dass innerhalb der ersten Wochen kleine Teilteams erstellt wurden und wir uns zu zweit in Aufgaben und Themengebiete einarbeiten mussten, war sowohl die Absprache als auch das Planen von Terminen unkompliziert und gut realisierbar.

Insgesamt war das Modul trotz eines schwierigen Starts sehr lehrreich und eine passende und sogar meiner Meinung nach notwendige Erweiterung des Moduls DIS21 Big Data. Das Arbeiten mit neuen Programmen wie Py-Charm, das Anwenden von neuen Libraries und lernen von neuen Analysen wie der PCA-Analyse werden zukünftig sowohl im Studium als auch Berufsleben sehr hilfreich sein.

### 4.2 Roman Poledniok:

In diesem Modul hatte ich wieder die Möglichkeiten zu lernen, wie Neuronale Netze funktionieren und wie man bestimmte Hyperparameter Tunen kann, um eine bessere Accuracy zu erhalten. Dabei fand ich nicht nur das Auseinandersetzen mit dem bereits vorhandenen Code, sondern auch die

## Abschlussbericht im Modul DIS22

Technische Implementierung in Python spannend. Ich konnte meine Fähigkeiten im Umgang mit der Sprache verbessern und neue Funktionen und Libraries kennenlernen.

Besonders beim Visualisieren der Umfragedaten hatte ich die Möglichkeit viele neue Libraries und Herangehensweisen zu entdecken, wodurch sich mein Wissen in diese Richtung stark steigern konnte.

Beim Arbeiten mit der API musste ich feststellen, dass der Frust, dass etwas nicht funktioniert, häufig Teil eines Projektes ist und es einfach Dead-Ends gibt, bei denen man nicht weiterkommt.

Überfachlich konnte ich auch meine Kompetenzen im Zeitmanagement und in der Teamarbeit verbessern. Dies sind beides Ziele, die ich mir am Anfang des Studiums im Modul DIS20a-ASC gesetzt habe. Durch die wöchentlich wiederkehrenden Meetings und das ständige Arbeiten mit einem Partner, der auf die Verlässlichkeit von mir angewiesen war, konnte ich meine Kompetenzen in diesem Rahmen verbessern.

Im Allgemeinen fand ich noch spannend, dass das Projekt auf echten Daten und einem Real World Beispiel basiert. Bis jetzt konnte ich diesen Ansatz nur bei der Arbeit kennenlernen, nicht jedoch in der UNI. Im Hinblick auf mein weiteres Studium und meine Berufslaufbahn war es Hilfreich Einblicke in die Hintergründe beim Ausarbeiten eines solchen Themas zu bekommen.