**HTBLA Grieskirchen**

4710 Grieskirchen, Parzer Schulstraße 1

**Fachrichtung Medizininformatik**

**Schuljahr 2020**

**Betreuer/Betreuerin:**

Mag. Dr. Bernhard Mayr, MBA

Marchtrenk, am 15.04.2020

**Ausgeführt von:**

Tobias Brzezowsky, 5AHGBM-3

**Bio-Inf**

Abschlussprojekt

**Bildklassifizierer**

Inhaltsverzeichnis

[1 Allgemeines 3](#_Toc37927558)

[2 Dataset 3](#_Toc37927559)

[2.1 Zusammenstellung 3](#_Toc37927560)

[2.2 Beispiele 3](#_Toc37927561)

[2.2.1 Brezel 3](#_Toc37927562)

[2.2.2 Kornspitz 4](#_Toc37927563)

[2.2.3 Mohnflesserl 4](#_Toc37927564)

[2.2.4 Semmel 4](#_Toc37927565)

[3 Modell trainieren 5](#_Toc37927566)

[4 Deployment auf Heroku 6](#_Toc37927567)

[5 Probleme 7](#_Toc37927568)

[5.1.1 Runtime type 7](#_Toc37927569)

[5.1.2 Pop-up Blocker 7](#_Toc37927570)

[6 To Do 7](#_Toc37927571)

Anhang A

Abbildungsverzeichnis

1. Allgemeines

Bei meinem Projekt handelt es sich um einen Bildklassifizierer der vier verschiedene Gebäcksorten unterscheidet. Die vier Kategorien setzen sich aus Brezel, Kornspitz, Mohnflesserl und Semmel zusammen.

1. Dataset
   1. Zusammenstellung

Um ein Datenset für jede Kategorie zu bekomme setzte ich die Variante ein, mit der man eine Liste von URLs aus Google Bildern erhält und diese anschließend auf Google Drive herunterlädt. Zu Beginn, bevor die Testdaten aussortiert wurden befanden sich 200 Bilder pro Kategorie in meinem Verzeichnis.

Danach wurden alle Bilder die nicht geöffnet werden konnten entfernt.

Zum Schluss wertet ich mein Datenset per Hand aus und entfernte all jene die meiner Klassifizierung im Weg stehen könnten.

Das Ergebnis sieht wie folgt aus:

|  |  |
| --- | --- |
| **Name des Gebäcks** | **Anzahl der Bilder** |
| Brezel | 107 |
| Kornspitz | 126 |
| Mohnflesserl | 119 |
| Semmel | 79 |
| Gesamt | 431 |

Besonders oft habe ich Bilder entfernt, bei denen das Gebäck nicht im Fokus lag. Damit meine ich Bilder von Gerichten (v. a. Sandwiches) oder Rabattanzeigen aus Zeitschriften. Darüber hinaus waren besonders viele Bilder des „Teig-Zustands“ oder eine Anleitung im Datenset „Mohnflesserl“ zu finden.

* 1. Beispiele
     1. Brezel



* + 1. Kornspitz



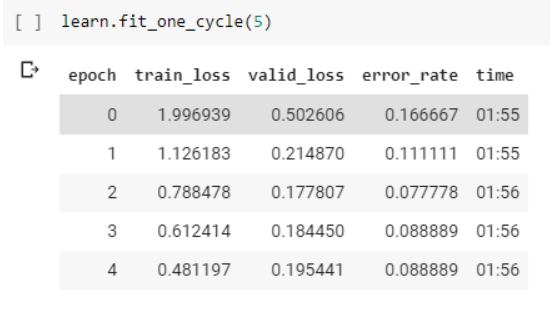
* + 1. Mohnflesserl

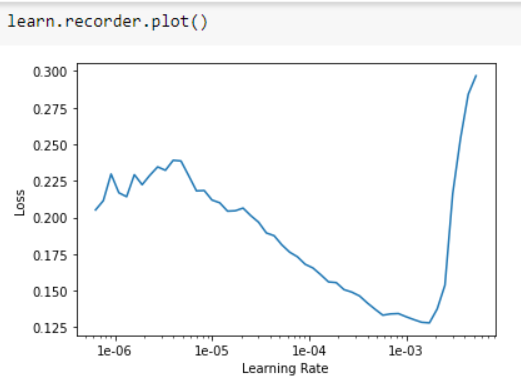


* + 1. Semmel

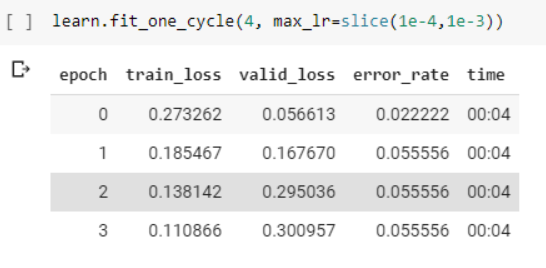
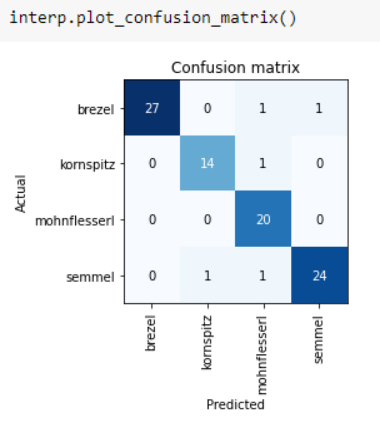


1. Modell trainieren

Stage 1:

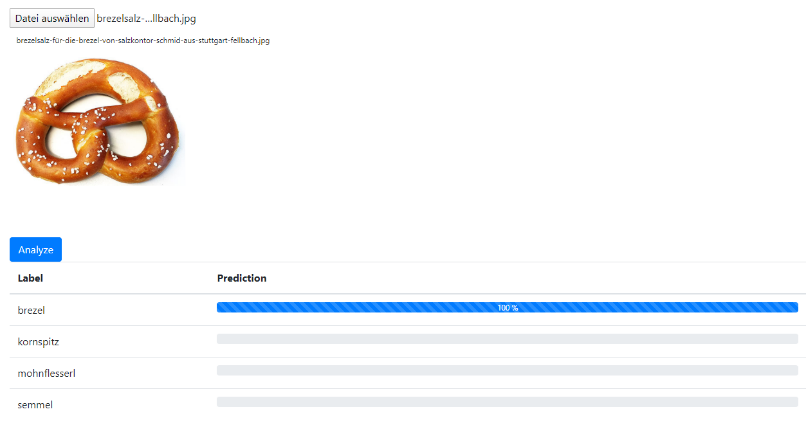
Learning Rate

Stage 2:

Confusion Matrix

1. Deployment auf Heroku

Nachdem ich einen Heroku-Account erstellt habe und meine App mit dem zugehörigen GitHub-Repository verknüpft habe, habe ich den Verweis auf mein „export.pkl“ und die Klassen geändert. Das Folgende Bild steht die derzeitige App mit einem Bild aus dem Datenset dar.



1. Design

Im Folgenden sieht man das angepasste Design der Heroku-Applikation



1. Probleme

Im Folgenden werden Probleme angesprochen die im Laufe des Projektes aufgetreten sind.

* + 1. Runtime type

Während des Projektes habe ich den Rechner geändert, beim Wechseln habe ich vergessen den Runtime type wieder auf GPU zu stellen weshalb sich die Zeit eines Durchlaufes drastisch erhöht hat. (s. Stage 1 – Modell trainieren)

* + 1. Pop-up Blocker

Während des Hochladens auf Heroku musste man einen Hardlink erstellen, damit die An-wendung den exportierten learner herunterladen konnte. Jedoch hatte ich eine Pop-up Blo-cker aktiviert welcher es aussehen ließ als würde die Datei korrekt freigegeben werden – dem war aber nicht so.

Jedes Mal, wenn also die Anwendung das „export.pkl“ File herunterladen wollte erwartete es einen Login mit meinem Google-Account den ich selbstverständlich nicht durchführen konnte.

* + 1. Verbindung zu Github

Die Verbindung zum GitHub-Repository war fehlerhaft. Jedes Mal, wenn ich deployen wollte schien die Fehlermeldung "your account has reached its concurrent builds limit" und der Support weist daraufhin eine Kreditkarte im Account festzulegen damit die fehlende Verifizierung stattfinden kann. Dies ist jedoch nicht das Problem die Verbindung zu GitHub musste lediglich wiederholt werden.