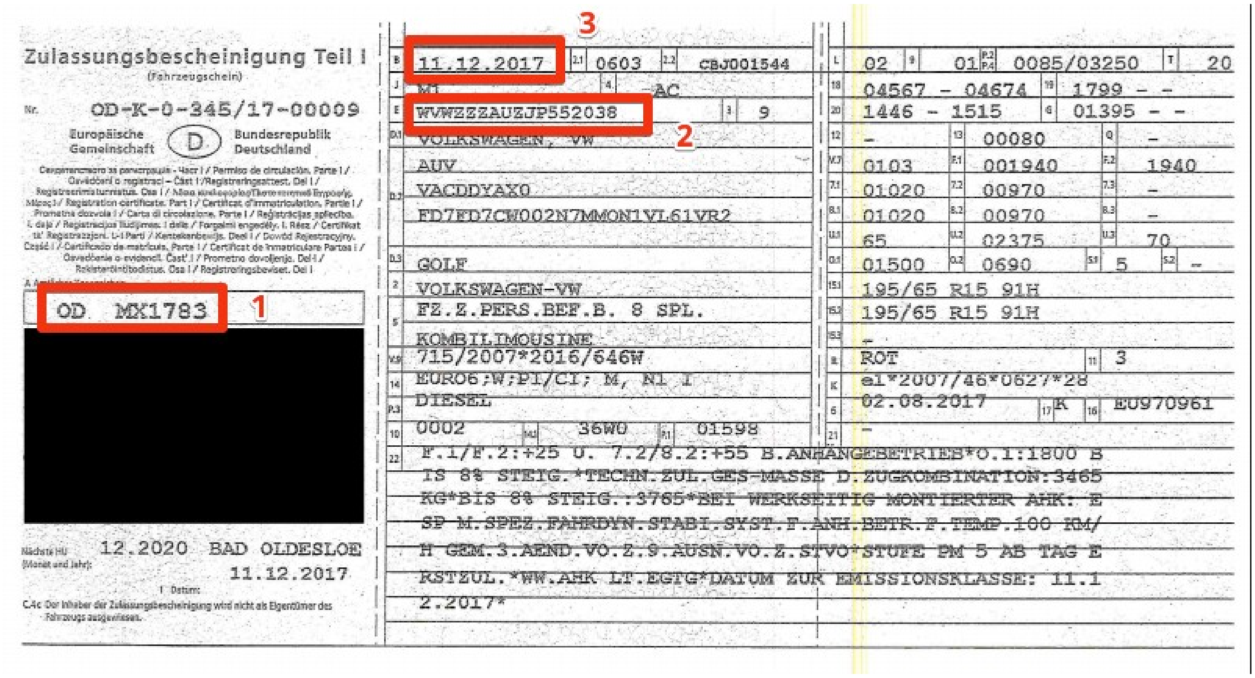
*** - Registration certificate readout

Background on the problem

- Customers upload registration certificates as a scan into ***
- Currently, these documents are read out with a lot of manual work
- They are used as a part of the vehicle procurement process and a faster read-out leads to a faster wow-moment of the customer because the vehicle is ready earlier and there is no suspense.
- We aim to read out three values from german registration certificates:
  - ○ Registration date (3)
  - ○ VIN (Vehicle identification number) (2)
  - ○ License plate (1)



Zulassungsbescheinigung Teil I
(Fahrzeugschein)

Nr. OD-K-0-345/17-00009

Europäische Gemeinschaft (D) Bundesrepublik Deutschland

Свидетелство за регистрация - Част I / Permiso de circulación, Parte I /
Osvědčení o registraci – Část I /Registreringsattest, Del I /
Registreerimistunnistus. Osa I / Άδεια κυκλοφορίας/Τεύχος εγγραφής,
Μέρος I / Registration certificate. Part I / Certificat d'immatriculation, Partie I /
Prometna dozvola I / Carta di circolazione, Parte I / Reģistrācijas apliecība,
I. daļa / Registracijos liudijimas. I dalis / Forgalmi engedély. I. Rész / Certifikat
ta' Registrazzjoni. L-I Parti / Kentekenbewijs. Deel I / Dowód Rejestracyjny.
Część I / Certificado de matrícula, Parte I / Certificat de înmatriculare Partea I /
Osvedčenie o evidencii. Časť I / Prometno dovoljenje. Del I /
Rekisteröintitodistus. Osa I / Registreringsbeviset. Del I

A Amtliches Kennzeichen

OD MX1783 1

[redacted]

Nächste HU (Monat und Jahr): 12.2020 BAD OLDESLOE
11.12.2017
I Datum:

C.4c Der Inhaber der Zulassungsbescheinigung wird nicht als Eigentümer des Fahrzeugs ausgewiesen.

3

B 11.12.2017 | 2.1 0603 | 2.2 CBJ001544

J M1 | 4 AC

E WVWZZZAUZJP552038 | 3 9

2

D.1 VOLKSWAGEN, VW

D.2 AUV
VACDDYAX0
FD7FD7CW002N7MMON1VL61VR2

D.3 GOLF

2 VOLKSWAGEN-VW

5 FZ.Z.PERS.BEF.B. 8 SPL.
KOMBILIMOUSINE

V.9 715/2007*2016/646W

14 EURO6;W;P1/CI; M, N1 I

P.3 DIESEL

10 0002 | 14.1 36W0 | P.1 01598

L	02	9	01	P.2 P.4	0085/03250	T	20
18	04567 - 04674	19	1799 - -				
20	1446 - 1515	G	01395 - -				
12	–	13	00080	Q	–		
V.7	0103	F.1	001940	F.2	1940		
7.1	01020	7.2	00970	7.3	–		
8.1	01020	8.2	00970	8.3	–		
U.1	65	U.2	02375	U.3	70		
O.1	01500	O.2	0690	S.1	5	S.2	–
15.1	195/65 R15 91H						
15.2	195/65 R15 91H						
15.3	–						
R	ROT	11	3				
K	e1*2007/46*0627*28						
6	02.08.2017	17	K	16	E0970961		
21	–						

22 F.1/F.2:+25 U. 7.2/8.2:+55 B.ANHÄNGEBETRIEB*O.1:1800 B
IS 8% STEIG.*TECHN.ZUL.GES-MASSE D.ZUGKOMBINATION:3465
KG*BIS 8% STEIG.:3765*BEI WERKSEITIG MONTIERTER AHK: E
SP M.SPEZ.FAHRDYN.STABI.SYST.F.ANH.BETR.F.TEMP.100 KM/
H GEM.3.AEND.VO.Z.9.AUSN.VO.Z.STVO*STUFE PM 5 AB TAG E
RSTZUL.*WW.AHK LT.EGTG*DATUM ZUR EMISSIONSKLASSE: 11.1
2.2017*

Goal

The goal is to evaluate if registration certificates can be read out automatically with good confidence: It is ok if not all the files can be processed fully automatically for all fields, but it is important that in this case we can identify which files and fields need further (manual) processing. For this test task we would like you to create a PoC that

1.) demonstrates the feasibility of the above goal
2.) gets as good accuracy as possible given the time constraints that this is "just" a test

Materials & Constraints

You get the following materials from us:

- Training data set of 284 registration certificate scans
- CSV file with filename, license plate, VIN and registration date for each sample in the data set

Since these scans were obtained from actual customers some of the scans are of very similar style and quality, this is a problem you will also find in reality. It is up to you to create a training, validation and test data set from the materials. Please ensure that there is sufficient diversity in your test set, ie. don't use a test set which contains scans from only 2 different customers/styles.

To keep this task simpler we have already separated the registration certificates into individual images for you and have only included german registration certificates in the data set. We hope this helps you focus on the data extraction and accuracy prediction problem.

You are allowed to use any tools of your choosing, including cloud machine learning APIs such as Google OCR or Amazon TextExtract. However, if you do use such tools we ask you to keep the data set on private storages in these services and delete all copies of it upon completion of the project.

Deliverables

- All code used to create your PoC, preferably with install & run instructions (no need to get fancy, we would just like to try to run it locally)
- Your training, validation and test data set
- Field level accuracy of your best performing model for the test set
- If your model produces it, information on the confidence of the results achieved on the test set (e.g. probability of correctness for each field)
- Any additional materials which you think would be helpful for us to understand how you approached the problem and how you would proceed from here

For evaluating the project we will of course look at the accuracy of your model in processing the test set as well as the accuracy of the confidence prediction. We prefer a slightly less accurate solution with very accurate confidence over a highly accurate, but on individual samples erratic solution (aka we would rather your models knows when it is unsure than a very bold model who happens to get it right on this test set).

We will also look at the code you created (knowing that this is a test project, so we don't expect things like extensive documentation, refactored & cleaned code etc.) and how you approached the problem. In the end we are interested both in the result that you have achieved as well as your path to get there.