

21.04.2022

- Daten sortieren
- „unnötige“ Daten löschen, Jahre auf den gleichen Stand bringen (Spalten Namen) (IstGkfz -> 2017, Istsonstig -> 2019/2020)
- Code umschreiben, sodass die Anzahl der Unfälle für jedes Jahr ausgegeben werden kann

27.04.2022

- fehlende Daten für NRW raussuchen (16 -18) -> Problem
- Daten zu Wetter raussuchen
 - wetter stationen nicht genau für AGS-Nummer -> Problem
 - > Daten evtl neu sortieren und benennen? Wie ist es möglich auf die richtige Daten zuzugreifen?
 - einige Daten nicht von 2016 bis 2021

30.04.2022

- Datensatz auf eine Stadt/Gemeinde reduzieren zur Vereinfachung (Mainz)
- > -Unfall daten verringern (Zeitpunkt, Unfall ja/nein) und um „nicht“ unfalldaten erweitern Jahr evtl unnötig?
- Trend bezüglich Wochentagen?
- 12 monate, 7 tage, 24 stunden -> ca. 2016 daten ($12*7*24$) -> Schaltjahr/Uhrumstellung?
- > wenn in einer stunde mehrere Unfälle dann eher output Anzahl der Unfälle und nicht binär?
- Allgemeine veranschaulichung der Daten um evtl. mehr Ideen zu bekommen

IDEEN

- ➔ Knn (Input: Zeitpunkt -> Output: Unfallanzahl/Unfall (ja/nein))
- ➔ Knn (Input: Zeitpunkt und Straßen Zustand -> Output: Unfallanzahl/Unfall (ja/nein))
- ➔ Entscheidungsbaum?
- ➔ Regression (Wochentag – Trend)

06.05.2022

- Code kommentieren und aufräumen
 - > # Kommentare erlaubt? Ja aber oft unnötig
- Funktionen ordentlich aufteilen
 - > Funktionen im Dokument Unfälle
- Alle benötigten Daten generieren und speichern
 - > Alle Daten für Mainz von 16 bis 20 -> 2016x5 Matrix (Jahr, Monat, Tag, Stunde, Anzahlunfälle)

09.05.2022

- Tex Dokument sortieren und bereits geschriebenes in Word einfügen
- in Overleaf einlesen/registrieren/arbeiten
- KNN fertigstellen
 - > läuft, Ausgabe allerdings unbrauchbar (für jeden Zeitpunkt in 2021 gleiche Vorhersage)
- Literatur durchsuchen (Deep Learning with Python/Youtubevideo)

14.05.2022

- Hinweis in Moodle zur eingabe Oberfläche einlesen und informieren
- Plots „verschönern“ -> auflösung, Achsen, für Mainz
- Lineare Regression/Entscheidungsbaum

16.05.2022

- Ändern des Optimizers zu Adam verbessert die Vorhersage

23.05.2022

- KNN -> Input gewichten nach Tensorflow Angaben (auf ganzes Dataset oder nur train?)
 - > Output wieder in allen Einträgen gleich
 - > Änderung der Optimizer etc. bewirkt nichts

28.05.2022

- Wetter Daten anpassen
 - Jahre auf 2016 bis 2021 reduzieren
 - Dataframe auf Datum und Wetter beschränken (durchschnitte berechnen)
 - Durchschnitte für alle Montage eines Monats in einem Jahr als Annäherung berechnen
 - `datetime.datetime -> x.weekday()` -> `mo = 0, di = 1, ... so = 6`
- KNN mit neuem Datensatz (Datum, Anzahlunfälle, Wetter) trainieren und testen
 - > Vorhersage nicht mehr zu allen Stunden gleich

- Abweichung zu eigentlichen Werten prüfen

01.06.2022

- warum nie mehr als 75% accuracy? Accuracy Berechnung anpassen
 - > unbalancierte Daten
 - > somit anderes Bewertungsmaß (Mail f1-Score)
- Code aufräumen und kommentare nachholen

09.06.2022

- Daten für Wiesbaden generieren um Knn zu testen (und sonst auch alles was für Mainz gemacht wurde)

 - > eventuell auch andere richtung da Wiesbaden mehr Daten hat

- Code interaktiv gestalten

18.06.2022

- Code interaktiv gestalten

- Ausprobieren Kategorie vorhersage -> klappt nicht

06.07.2022

- F1-Score berechnen

- Plots final für Arbeit Erstellen

 - Testen der Anzahl an Neuronen

 - f1-Score für Labels

 - f1-Score insgesamt

 - Histogramm der Differenzen

MEETING 20.05.22 13 Uhr

- Welches Problem wird gelöst (d.h. welche Vorhersagen werden getroffen)?

 - > Vorherage für die Wahrscheinlichkeit der Anzahl der Unfälle (Niederschlag)

- Wer ist die Zielgruppe der interaktiven Anwendung?

 - > Ottonormalverbraucher

- Welche Methoden werden eingesetzt

 - > Künstliches neuronales Netz (Deep Learning)

 - > Polynomiale Regression

- Welcher zusätzliche Datensatz wird verwendet?

 - > Wetter bzw Niederschlag