

# Stromerzeugung bei unterschiedlichen Wetterlagen – Datenaufbereitung und Analyse zur Korrelation Mithilfe von Regressionsalgorithmen in Weka

Korbinian Eller, Kay Gietenbruch,

## Abstract

In dieser Ausarbeitung wird das Vorgehen beim Sammeln von Daten, deren Aufbereitung und die Analyse mithilfe der Regressionsalgorithmen in Weka erläutert und die Ergebnisse dokumentiert.

Ziel ist es durch Wetterdaten des Deutschen Wetter Dienstes (DWD) auf Stromerzeugungszahlen der erneuerbaren Energien Solar und Wind (größte Beeinflussung durch Wetter) zu schließen

## Idee

Zur gegebenen Themenstellung "Einen Themenbereich der KI vertiefen" war eine erste Idee die Klassifikation von Datensätzen. Dies war sehr ähnlich einer der letzten Aufgabenstellungen der Übungsstunden des Fachs. Es war vor allem aufgrund der Implementierung in der Programmiersprache C und damit dem greifbar machen des Algorithmus interessant.

Der Gedanke der Klassifikation war schnell gefestigt nun fehlten noch die Daten mit denen gearbeitet werden soll. Über Daten wie Covid-19 Erkrankungszahlen, Bußgeldbescheide, Denkmalstandorte oder Amazon Personen Daten ist eine Idee herausgestochen.

"Es wäre doch interessant, Wetterdaten und Stromdaten in Korrelation miteinander zu bringen und so die Stromerzeugung anhand des vorherrschenden Wetters deuten zu können". Und das war dann das Noema mit dem fortgefahren werden sollte. Ziel ist eine repräsentative Menge der Daten zu sammeln um eine Klassifikation sinnvoll ausführen zu können. Trotz allem war die Beschränkung der Daten auch ausschlaggebend. Festgelegt wurde sich dann auf den Zeitraum eines Jahres und wegen der Zeit in der die Daten gesammelt wurden (Ende 2023) war das Jahr 2022 passend für die Aufgabe.

Das Projekt konnte nun in mehrere Schritte unterteilt werden: die Daten sammeln, die Daten aufbereiten, den Klassifikator programmieren, Testen und verbessern und die Arbeit zu dokumentieren.

Bei einer Besprechung mit dem Betreuer des Projekts Herrn Prof. Dr. Baumann das Projekt besprochen wurde wurde klar, dass eine Klassifizierung der Daten nicht das

geeignetste Modell für die Analyse der Korrelation ist. So wurde der Plan neu geschrieben und eine Analyse mithilfe von Regressionsalgorithmen in dem Machine-Learning Programm Weka stand ab dem Zeitpunkt im Vordergrund. Dafür sollten die Daten noch angepasst und dann mit Weka und den Regressionsalgorithmen experimentiert werden. In dem Sinne, dass das am besten geeignete Modell zu Regression gefunden wird!

## Daten sammeln - Korbi

Siehe Doku in den MD files. Woher die Daten runtergenommen? Wie waren diese Daten unterteilt? Einmal Wetter und einmal Strom! Auswahlverfahren bei den Wetterdaten, welche waren interessant für uns? Wie geil waren die Stromdaten – Epoch time und so! Wetterdaten waren teilweise nicht vorhanden, also musste man da einen Kompromiss von Historical und recent finden.

## Aufbereitung - Kobi

Siehe Doku in den MD files.

Wetter: Von den Zip dateien in die CSV Dateien. Von den CSV In eine Große CSV. Fehlwerte, fehlende Csvs, erst einmal eine Große mit allen und dann eine kompaktere mit nur den 157 vollständigen. Dann einen noch kompakteren ohne RR-2 i think. Dann eine noch kompaktere meaned csv. -> ab in eine arff file für weka über den Weka explorer.

Power: Viele einzelne Json Dateien die nicht über genau das Jahr auch gehen. epoch in normale time codes. Json in CSV umwandeln. Die Paar csvs dann in eine größere mit den Epoch timecodes. Dann auf das Jahr 2022 beschneiden mit den Epoch Codes

## Regressionanalyse mit Weka - Kay

Vorgehen beschreiben. Welche Modelle? Weka Version? Welche Files? Gab es Probleme? Auffällige Zeiten? Kompatibilität mit fehlwerten? Hat was gepackt uws. Sachen aus dem Weka Buch hilfreich?

## Vergleich von Modellen - Kay

Was war denn eigentlich das Beste und warum. Was war das Fehlermaß? Wie haben sie sich zeitlich geschlagen? Weka gut/schlecht? Andere Ideen oder so. Wie war eigentlich Klassifikation - hätte das doch Sinn ergeben?

## **Fazit**

Was konnten wir aus dieser Arbeit schließen? Erfolg der Bearbeitung der Fragestellung? Aufwand/Zeit Verhältnis. Ergebnis der Modelle. Hätte man was besser machen können und wie?

## **References**

Witten, I. H.; Frank, E.; and Hall, M. A., eds. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington, MA: Morgan Kaufmann, third edition.