

# Klassifikation und Analyse aus Stromdaten im Haushalt mit neuronalem Netzwerk

Tim Schrodi

01.01.2018

# Inhaltsverzeichnis

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Einleitung</b>                                    | <b>1</b>  |
| <b>2</b> | <b>Grundlagen</b>                                    | <b>2</b>  |
| 2.1      | Maschinelles Lernen (Machine Learning) . . . . .     | 2         |
| 2.2      | Physikalische Grundlagen zur Netzaktivität . . . . . | 2         |
| 2.3      | Erhebung der Messdaten . . . . .                     | 2         |
| 2.3.1    | Klassifikation der Messdaten . . . . .               | 3         |
| 2.4      | Visualisierung . . . . .                             | 4         |
| <b>3</b> | <b>Ausführung</b>                                    | <b>6</b>  |
| 3.1      | Manuelle Analyse . . . . .                           | 6         |
| 3.2      | Neuronales Netz . . . . .                            | 10        |
| 3.3      | Vorbereiten der Daten . . . . .                      | 10        |
| 3.4      | Trainingsprozess . . . . .                           | 11        |
| 3.5      | Auswertung der Ergebnisse . . . . .                  | 11        |
| <b>4</b> | <b>Ergebnis</b>                                      | <b>12</b> |
| <b>5</b> | <b>ProduktivBetrieb</b>                              | <b>13</b> |
| <b>6</b> | <b>Wirtschaftlichkeit</b>                            | <b>14</b> |
| <b>7</b> | <b>Ausblick</b>                                      | <b>15</b> |

## **Zusammenfassung**

Das Ziel dieser Studienarbeit war es mit Hilfe von verschiedenen Machine Learning-Methoden aus Stromdaten eines Haushalts verschiedene Geräte zu klassifizieren. Dazu wurden mit einem Messgeräte die allgemeine Spannung, Frequenz und verschiedene Oberwellen eines üblichen Stromnetzwerks eines privat Haushalts über mehrere Monate erfasst. Hinzu wurden manuell verschiedene Geräte wie eine Kaffemaschine oder eine Mikrowelle klassifiziert. Anhand der Stromverläufe und den dazu klassifizierten Geräten wurden verschiedene neuronale Netze trainiert und miteinander verglichen. Außerdem wurde versucht mit unüberwachten Lernen Gemeinsamkeiten innerhalb des Stromverlaufs ohne vorherige Klassifizierung zu erkennen.

Auch wurde die Wirtschaftlichkeit sowie der produktive Einsatz der Ergebnisse beachtet. Die neuronalen Netze, welche die besten Ergebnisse erzielten werden außerdem im produktiv Betrieb getestet und eingesetzt.

# Kapitel 1

## Einleitung

Was bis vor kurzer Zeit nur ein wissenschaftlicher Teil der Informatik war, erhält nun immer größere Bedeutung und Einfluss in vielen weiteren wirtschaftlichen und wissenschaftlichen Themen. Seit Google, Facebook, etc. große Fortschritte in diesem Bereich erzielen und Maschinelles Lernen auch in produktivem Betrieb Einsatz findet, liegt es Nahe dies auch auf bisher unberührten Branchen auszuweiten.

In dieser Arbeit wird Data-Mining auf elektrotechnische Größen angewendet um weiterführende semantische Aussagen über diese Werte zu erhalten. Es werden verschiedene Netzdaten und dazugehörige Verläufe aufgezeichnet und mithilfe von verschiedenen Methoden von Maschinellern analysiert. Hauptbestandteil ist die Mustererkennung in den aufgenommenen Verläufen und deren Zuordnung zu verschiedenen Geräten in einem Haushalt. Hierbei sollen verschiedene normale Haushaltgeräte, wie Kaffeemaschinen oder Fernseher, welche in einem Stromnetzes eines privaten Haushalts erkannt werden um somit Aussagen über Laufleistungen zu treffen. Außerdem wird auf die wirtschaftliche Nutzung des resultierenden Modells, sowie deren Produktivbetrieb eingegangen.

Für die maschinelle Analyse wird die Keras-API<sup>1</sup> mit Tensorflow im Backend verwendet. Die Datenverarbeitung, Visualisierung sowie die Trainings- und Testphasen werden mit Python<sup>2</sup> umgesetzt.

---

<sup>1</sup><https://keras.io/>

<sup>2</sup><https://www.python.org/>

# Kapitel 2

## Grundlagen

### 2.1 Maschinelles Lernen (Machine Learning)

### 2.2 Physikalische Grundlagen zur Netzaktivität

### 2.3 Erhebung der Messdaten

Um aussagekräftige Analysen und Klassifikationen über ein Stromnetz bzw. die Geräte in einem Stromnetz mit Maschinellern Lernen machen zu können, werden viele Trainings- und Testdaten benötigt. Die Daten bestehen aus verschiedenen physikalischen Größen, die zu einem bestimmten Zeitpunkt in einem Stromnetz auftreten. Zu diesen Größen gehört die allgemeine Netzspannung, die Netzfrequenz sowie sieben harmonischen Oberwellen (vgl. 2.2). Um einen allgemeinen Überblick über den Verlauf der Netzaktivität zu erhalten sowie verschiedene Zeiten und Geräte vergleichen zu können, müssen Daten über lange Zeiträume erhoben werden.

Zur Erhebung der Werte zur Netzaktivität wurde ein WeSense-Messgerät<sup>1</sup> verwendet. Dieses Gerät misst alle benötigten Werte und sendet diese über einen MQTT-Broker<sup>2</sup> an einen Service von WeSense, welcher dann die Daten aufbereitet und in einer MSSQL Datenbank abspeichert. Die Werte werden sekundlich gemessen und in die Datenbank gespeichert, weshalb zunächst in eine row-based Datenbank gespeichert wird und später dann die Daten in eine column-based Datenbank zur schnellen Abfrage überführt werden.

---

<sup>1</sup><http://www.wesense-app.com/home-en/>

<sup>2</sup>Message Queuing Telemetry Transport

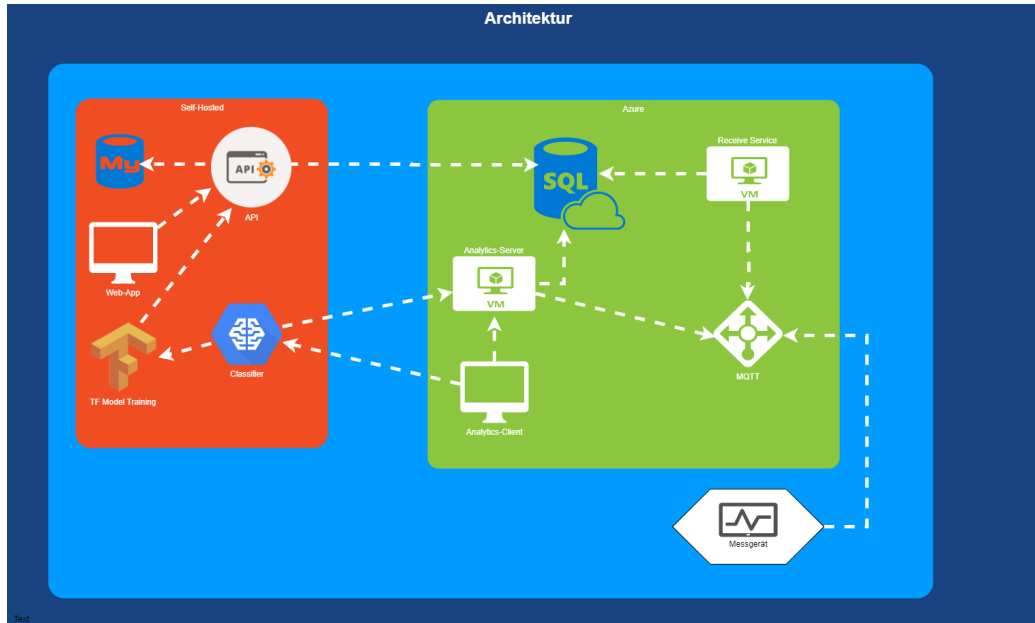


Abbildung 2.1: Complete Architecture

### 2.3.1 Klassifikation der Messdaten

Durch die oben beschriebene Erhebung sind die physikalischen Werte zu bestimmten Zeitpunkten bestimmt worden. Zusätzlich wird nun zur Identifikation der Geräte sowie zum Maschinellen Lernen, genau definierte Zeiträume benötigt in denen bestimmte Geräte aktiv waren. Dies bedeutet, dass jedem Zeitpunkt ein oder mehrere Geräte zugewiesen werden.

Um diese gelabelten Daten zu erheben gibt es verschiedene Möglichkeiten. Die Daten können entweder durch eine Person, welche Zeiten zu denen sicher Geräte aktiv waren manuell erfasst, oder durch eine Maschine automatisch erhoben werden. Jedoch wird zur automatischen Erhebung ein weiteres Gerät benötigt, welches zwischen dem zu messenden Gerät und dem Stromnetz zwischengeschaltet wird und sobald Strom fließt Daten erfasst. Somit werden die manuell durch eine Person erfasst.

Hierzu wurde eine progressiv Web-App(vgl. Abbildung 2.2) mit einer einfachen MySQL-Datenbank erstellt, mit der die Daten sehr einfach erfasst und abgespeichert werden können.

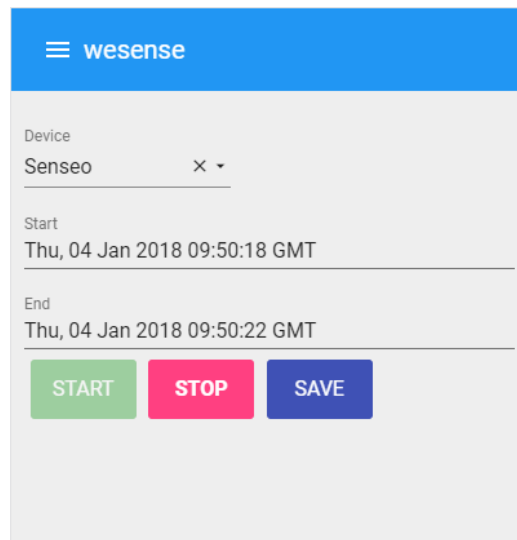


Abbildung 2.2: Screenshot der progressive Web-App

## 2.4 Visualisierung

Zusätzlich zur manuellen Erhebung der Daten wurden zur besseren Analyse der Daten verschiedene Visualisierungsmöglichkeiten implementiert. Zum einen können die verschiedenen Physikalischen Größen eines gelabelten Gerätes zu einem bestimmten Zeitpunkt miteinander verglichen werden. Außerdem können bestimmte Größen zu verschiedenen gelabelten Zeiträumen eines Gerätes verglichen und analysiert werden. Durch diese Visualisierung können sehr gut und genau Gemeinsamkeiten in verschiedenen Größen oder Zeiten erkannt werden.

Es werden verschiedene Diagramme sowie Normalisierungen der Daten zur Analyse bereitgestellt. Es besteht die Möglichkeit die Daten in einem Liniendiagramm sekundlich oder in frei wählbaren zusammengefassten Datenpunkten, sogenannten Klassen, anzuzeigen. Des weiteren können Histogramme mit verschiedenen Klassen gewählt werden.



Abbildung 2.3: Screenshot eines gelabelten Zeitraumes aus der Web-App



# Kapitel 3

## Ausführung

Dieses Kapitel beschreibt die Vorgehensweise von der ersten manuellen Analyse der Daten bis zum fertigen Produkt. Als Beispielprodukt für die Klassifizierung wurde eine Senseo Kaffeemaschine gewählt, da durch häufige Benutzung viele Daten erhoben werden können. Außerdem wurde eine Mikrowelle als zweite Gerät mit wenigen Daten gewählt um verschiedene Parameter und Kennzahlen zu vergleichen. Hierbei soll die Anzahl der Daten sowie die Klassifikation von mehreren Geräten innerhalb eines Neuronalen Netzes untersucht werden.

### 3.1 Manuelle Analyse

Zur manuellen Analyse der Daten wird das in 2.4 beschriebene Tool verwendet. Zunächst werden sehr aussagenkräftige Größen wie die Spannung oder die Frequenz der Kaffeemaschine verwendet und mit anderen aktiven Zeiträumen der Kaffeemaschine verglichen. Hierbei kann ein sehr spezifischer Verlauf der Spannung erkannt werden.

Wie in Schaubild reffig:Spannungsverlauf zu sehen ist, ist die Kurve zu Beginn start fallend und verbleibt dann eine gewissen Zeit auf diesem Tief. nach einem längeren steigenden Abschnitt fällt die Kurve wieder bis der Zubereitungsvorgang beendet wurde und wieder steigt.

Diesem Verlauf können nach mehreren Beobachtungen bestimmte Vorgänge einer Kaffeezubereitung zugeordnet werden. Zu Beginn der Kaffeezubereitung wird die Kaffeemaschine manuell eingeschaltet. Dies führt automatisch zum erwärmen des Brühwassers, welches dem ersten Fallen der Kurve zugeordnet werden kann. Da dort viel Energie benötigt um das Wasser zu erhitzen steigt der Stromverbrauch der Kaffeemaschine stark an und somit fällt die

Netzspannung stark ab. Die Netzspannung bleibt solange auf einem gewissen Tiefpunkt mit minimaler Netzschwankung bis das Wasser erwärmt wurde und ein weiterer manueller Schritt zum Fortfahren des Prozesses notwendig ist. Nach Wahl der Tassengröße wird dann der Brühvorgang gestartet. Dabei wird das erhitzte Wasser mit einem gewissen Druck durch einen Kaffeepad gepresst. Da dieser Druck bei der Senseo Kaffeemaschine durch eine elektronische Pumpe erzeugt wird, sinkt demnach die Netzspannung ab bis der komplette Kaffee durchgelaufen ist. Somit kann die Zweite Tiefpunktphase dem "Pressvorgang" der Kaffeemaschine zugeordnet werden.

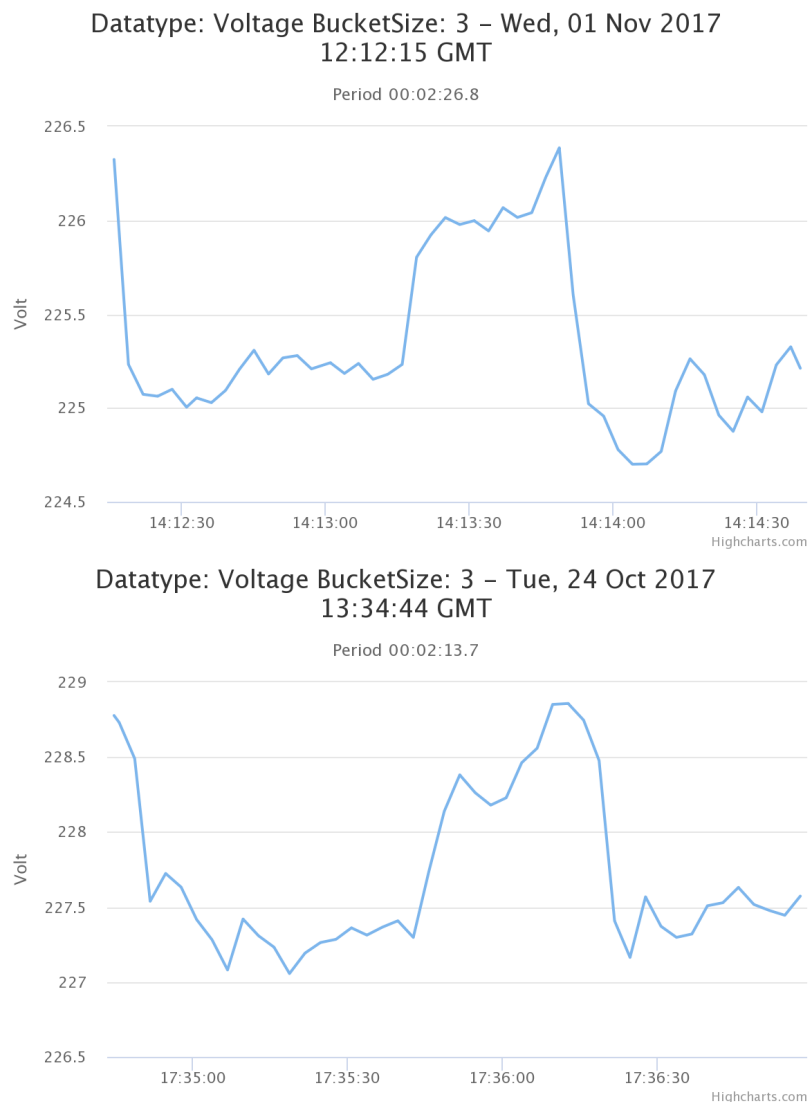


Abbildung 3.1: Spannungsverlauf der Senseo Kaffeemaschine mit einer Klasse von je 3 Datenpunkten zu 2 verschiedenen Zeitpunkten

Bei der manuellen Analyse der Mikrowelle können bei Analyse des Spannungs- oder Frequenzverlaufs leider keine hervorstechenden Merkmale oder Gemeinsamkeiten erkannt werden. Somit wird die Mikrowelle demnach einen anderen Einfluss auf die Netzaktivität ausüben. Wie bei dem Vergleich der verschiedenen harmonischen Oberwellen zu sehen ist (siehe Schaubild 3.2), besteht eine große Ähnlichkeit der Kurven der dritten harmonischen Welle.

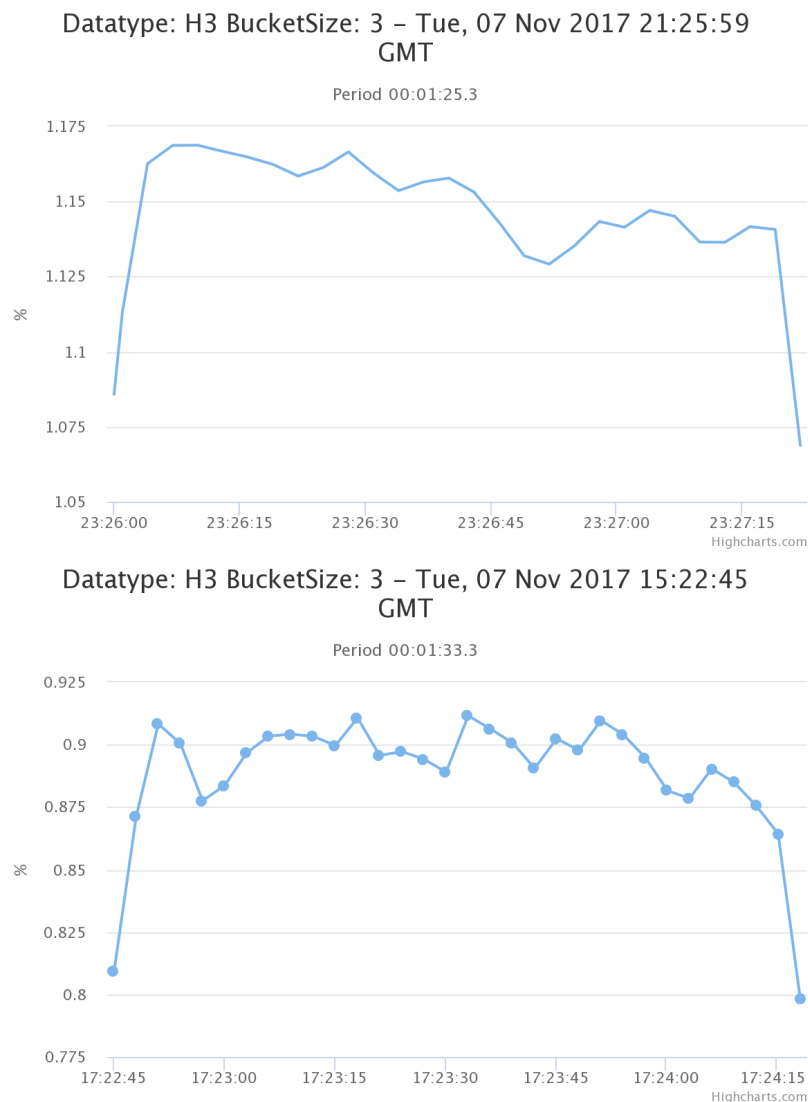


Abbildung 3.2: Verlauf der 3. harmonischen Oberwelle einer Mikrowelle mit einer Klasse von je 3 Datenpunkten zu 2 verschiedenen Zeitpunkten

Wie dieser Abschnitt zeigt können schon mit bloßen Auge bestimmte Auswirkungen der verschiedenen Geräte erkannt werden. Auch wenn einige Störungen auftreten und den Verlauf verfälschen sollte es dennoch möglich sein diese Geräte herauszufiltern. Weiterführend wird dieser Prozess automatisiert und zu verbessert um auch trotz großer Störungen, Geräte präzise Klassifizieren zu können.

## 3.2 Neuronales Netz

## 3.3 Vorbereiten der Daten

Um Klassifizierung der Geräte auf neuronalen Netzen abzubilden müssen die vorhandenen Daten in ein geeignetes Format transformiert werden. Dabei gibt es verschiedene Vorraussetzungen zu beachten, sowie verschiedene Möglichkeiten die Daten zu transformieren.

Bei konventionellen Neuronalen Netzen werden Eigenschaften Objekten zugewiesen, wobei alle Objekte mit denselben Eigenschaften betitelt werden. Ein Beispiel dafür sind Tiere, bei denen anhand von Reproduktion oder Atmung in Säugetiere, Vögel, Reptilien, etc. eingeteilt wird. So können spezielle Tiere wie ein Hund in eine dieser Klassen zugewiesen werden. Alle klassifizierten Tiere werden hierbei dieselben Eigenschaftsklassen mit unterschiedlichen Werten zugewiesen. Das heißt, dass einem Hund oder einer Taube die Anzahl der Beine und die Reproduktion zugewiesen werden jedoch mit jeweils unterschiedlichen Werten.

Durch die in Kapitel 2.3 beschriebene Erhebung der Daten, werden die Stromdaten in einem bestimmten Format von der API bereitgestellt. Es werden 2 verschiedene Listen zurückgegeben wobei eine die reinen Messdaten enthält und die andere die Labels für diese Daten.

Die reinen Messdaten werden zusammengetragen, indem zu den gelabelten Zeiträumen jeweils alle sekundlich gemessenen Werte des Stromnetzes als eine Matrix eingefügt werden. Somit ergibt sich eine drei dimensionale Matrix welche inder ersten Dimension die Zeiträume abbildet, in der zweiten Dimension die sekundliche Zeitreihe und die dritte die vorhandenen physikalischen Größen.

Listing 3.1: Datenstruktur der gelabelten Messdaten, die von der API bereit gestellt wird

```
[
  [
    [u, f, h3, h5, h7, h9, h11, h13, h15],
    [u, f, h3, h5, h7, h9, h11, h13, h15],
    [u, f, h3, h5, h7, h9, h11, h13, h15],
    ...
  ]
]
```

### **3.4 Trainingsprozess**

### **3.5 Auswertung der Ergebnisse**

# Kapitel 4

## Ergebnis

## Kapitel 5

## ProduktivBetrieb



# Kapitel 6

## Wirtschaftlichkeit

# Kapitel 7

## Ausblick

|                  |   |
|------------------|---|
| <b>MSSQL</b>     | Microsoft SQL-Server  |
| <b>CI</b>        | Continuous Integration  |
| <b>CD</b>        | Continuous Delivery   |
| <b>E2E</b>       | end to end  |
| <b>TDD</b>       | Test-Driven development   |
| <b>JaaS</b>      | Jenkins as a Service  |
| <b>SaaS</b>      | Software as a Service   |
| <b>PaaS</b>      | Platform as a Service   |
| <b>MTA</b>       | Multi-Target Application  |
| <b>VM</b>        | Virtual Machine   |
| <b>PPMS</b>      | Product and Production Management System  |
| <b>E2E tests</b> | End to End tests, is a testing process which simulates user interactions like clicks or text input to test the whole application in a complete workflow |

# Literaturverzeichnis