

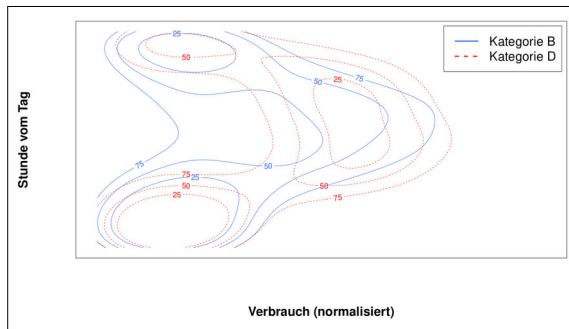
Alexander
van Schie



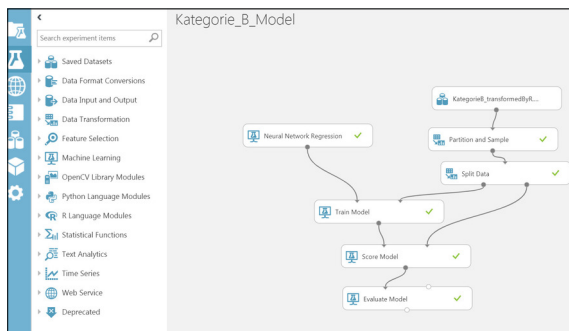
Nicolas
Ganz

Studenten	Alexander van Schie, Nicolas Ganz
Examinator	Prof. Stefan Richter
Themengebiet	Internet-Technologien und -Anwendungen
Projektpartner	EWZ, Zürich

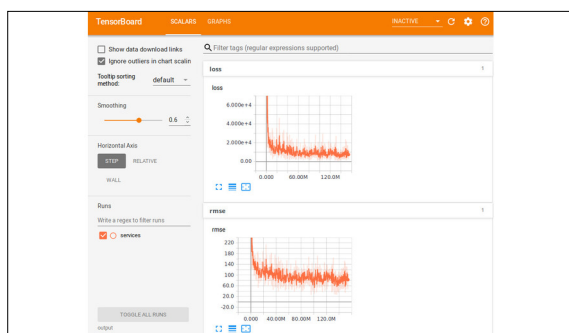
EWZ und Machine Learning



Analyse der Verteilung der Daten



Experiment im Microsoft ML Studio



Auswertung im TensorBoard

Ausgangslage: Um Ausgleichszahlungen zu reduzieren muss ewz möglichst genaue Prognosen des Stromverbrauchs berechnen. In dieser Studienarbeit werden mittels Machine Learning folgende drei Ziele verfolgt:

1. Analyse von möglichen Optimierungen des Prognoseverfahrens von Stromwerten
2. Prüfung, ob die Klassifizierung der Kunden verbessert werden kann
3. Optional: Erkennung von Eigenstromproduzenten

Die Lastgangdaten wurden aus den Jahren 2015 bis 2018 alle 15 Minuten gesammelt. Diese kommen von 164 verschiedenen Messanschlüssen.

Vorgehen / Technologien: Zu Beginn wurde mit einer Analyse der Daten ein Überblick verschafft, um Patterns und Zusammenhänge zu erkennen. Die Ergebnisse der Analyse zeigen den Einfluss verschiedener Faktoren wie Wetter und Zeit und werden in Grafiken veranschaulicht.

Mit dem gewonnen Einblick wurden diverse Algorithmen getestet, welche aus den Daten lernen. Zuerst wurden simplere Algorithmen angewandt. Mit den daraus folgenden Resultaten wurden Algorithmen gewählt, welche geeigneter und optimaler sind. Die Untersuchung wurden mit diesen fortgeführt.

Für die Analyse der Daten und die Erstellung der meisten Grafiken kam R zum Einsatz. Bei der Entwicklung der Models wurde Microsoft Azure ML Studio, Microsoft Azure ML services, TensorFlow und Keras verwendet.

Ergebnis: Die simplen Algorithmen, wie lineare Models und der Decision Tree, erbrachten bei der Prognostizierung keine sinnvollen Resultate. Grund dafür ist die Komplexität der Daten.

Mit den komplexeren Algorithmen konnten je nach Konfiguration gute Ergebnisse erzielt werden. Nach mehreren Versuchen wurden mit Artificial Neural Networks Fortschritte erzielt. Die besten Resultate jedoch entstanden unter Anwendung des Boosted Decision Tree Algorithmus.

Bei der Segmentierung von Kundengruppen brachte der verwendete Algorithmus keine signifikante Verbesserung hervor. Die Daten sollen pro Anschluss klassifiziert werden, haben jedoch pro Anschluss und Zeiteinheit einen Eintrag. Die nächste Idee war die Umstellung der Datenstruktur, sodass es nur noch einen Eintrag pro Anschluss gibt. Leider lieferte diese Umstellung keine besseren Resultate.