

6 Diskussion und Fazit

Die größte Herausforderung bei der Prognose von Verbraucherleistung ist, dass als Eingangsdaten für die Vorhersage fast ausschließlich vergangene Leistungsdaten genutzt werden können. Für eine erfolgreiche Vorhersage ist es also notwendig, dass hohe Korrelationen zwischen den vergangenen und zukünftigen Daten vorliegen. Zudem kann dadurch schlecht auf plötzliche Veränderungen im Verbrauch geschlossen werden. Dies beeinträchtigt stark das Ergebnis der Prognosen. Es hat sich herausgestellt, dass manche Modelle, welche in dieser Arbeit erstellt wurden, sich nicht oder kaum für das Vorhersagen von Stromlastprofilen eignen.

ARIMA wurde in dieser Arbeit nur mit einer täglichen Saison trainiert. Deswegen schneidet das Modell in allen hier betrachteten Szenarien schlechter ab als die anderen Modelle. Das größte Problem ist hierbei, dass durch das Fehlen der wöchentlichen Saison, das Modell nicht in der Lage ist, zwischen den Wochenenden und Wochentagen in seiner Prognose zu unterscheiden. Es wurde nicht mit einer wöchentlichen Saison trainiert, da dadurch eine zu hohe Rechenkapazität erforderlich wäre. Auch das Auswählen der Hyperparameter war durch die verfügbare Rechenkapazität begrenzt. Zudem war auffällig, dass die letzten zwei Tage der Trainingsdaten großen Einfluss auf die Prognose haben. ARIMA erzeugte die besten Ergebnisse an Prognosen für Donnerstag und Freitag, da die vorherigen drei Tage alle Wochentage waren. Durch das Fehlen der wöchentlichen Saison haben Datenlücken oder Feiertage im Zeitraum der Trainingsdaten kaum Einfluss auf das Ergebnis, da so keine für das Modell relevante Regelmäßigkeit durchbrochen wird. Das Modell hatte jedoch eine geringe Verlässlichkeit und zeigte meistens einen relativ hohen RMSE.

TBATS prognostiziert Saisonalität durch die Erstellung von Fourier-Reihen. Dies war ein entscheidender Vorteil des Modells, denn so konnte mit einer wöchentlichen und täglichen Saison trainiert werden. Dies ermöglichte dem Modell zwischen den Wochenenden und den Wochentagen meist zuverlässig zu unterscheiden. Jedoch war dies auch die einzige Möglichkeit auf das Modell Einfluss zu nehmen, da die restlichen Parameter von der TBATS Funktion des sktime Modell automatisiert erstellt werden. Der RMSE in den behandelten Prognosen war meist im Vergleich zu SVR hoch. Zudem ist TBATS sehr anfällig für Datenlücken oder Feiertage in den Trainingsdaten, hier produzierte das Modell keine brauchbaren Ergebnisse.

SVR hat sich als zuverlässigstes Modell herausgestellt. Da SVR mehrere Features als Eingangsdaten nutzt, sind hier die Anpassungsmöglichkeiten am größten. So ist es zum Beispiel möglich ein Feature für Feiertage einzuführen, welches das Modell sehr robust macht und Feiertage auch korrekt vorhersagt. Dies ist in den anderen beiden Modellen nicht möglich.

Durch die Features „Tag der Woche“ und „Stunde des Tages“ ist die Vorhersage des SVR Modells zudem nicht alleinig abhängig von vergangenen Leistungsdaten. Dies macht es robuster für Fehler in den Trainingsdaten. Außerdem gelingt so die korrekte Darstellung der Saisonalität. Die Ermittlung der Hyperparameter erfolgt zudem zuverlässig durch eine Bayes'sche Optimierung und muss nicht empirisch durchgeführt werden. SVR erzielt meist bessere Ergebnisse im Vergleich zu den beiden anderen Machine Learning Modellen. Dabei ist die iterative Vorhersage von SVR zu bevorzugen, da der Rechenaufwand hierfür wesentlich geringer ist.

Alle Machine Learning Modelle sind jedoch sehr fehleranfällig. Zudem ist der Programmieraufwand wesentlich größer als bei den statistischen Algorithmen. Diese funktionieren mit höherer Zuverlässigkeit und Genauigkeit als die Machine Learning Modelle. Vor allem die Algorithmen, welche auf Mittelwertbildungen basieren, produzieren zuverlässige Ergebnisse. Der erste Algorithmus, welcher als Vorhersage dieselben Tage der letzten Woche nutzt, ist sehr anfällig für Fehler an diesen Tagen, da diese übernommen werden. Durch die Bildung von Mittelwerten haben die Fehler in einzelnen Tagen in den Trainingsdaten weniger Einfluss auf das Ergebnis und es wird ein für die jeweiligen Standorte repräsentatives Lastprofil erstellt. Dabei ist der Unterschied zwischen den Ergebnissen von Algorithmus 2, also der Mittelwertbildung aus den Tagen an denen vorhergesagt wird, und Algorithmus 3, also der Mittelwertbildung für alle Wochentage und Wochenenden, geringfügig. Beide dieser Algorithmen eignen sich zur Prognose von Lastprofilen.

Das Ergebnis der Algorithmen ließ sich noch verbessern indem die Leistung des letzten Zeitpunktes der Trainingsdaten mit der Prognose der Algorithmen linear verbunden wurde. Dadurch ist man in der Lage auf plötzlich angestiegene Verbräuche unmittelbar vor der Prognose zu reagieren und diese in die Prognose zu implementieren.

Dem Machine Learning sind für den Anwendungsfall dieser Arbeit Grenzen gesetzt und die Vorhersage mithilfe von statistischen Algorithmen basierend auf Mittelwertbildungen eignet sich eher für die Verwendung im Projekt eMobiGrid. Dies hat einige Vorteile, wie zum Beispiel die einfachere Umsetzung und der geringere Rechenaufwand. Dennoch weisen die Prognosen der statistischen Modelle Abweichungen zum tatsächlichen Lastprofil auf. Ein weiterer Nachteil ist, dass sie nicht in der Lage sind auf veränderte Gegebenheiten zu reagieren. Beispielsweise würden die Algorithmen einen wesentlich höheren Verbrauch durch eine neu installierte Anlage nicht beachten. Deswegen wird in einer weiteren Arbeit versucht eine Lösung mithilfe von maschinellem Lernen zu finden.