

# Generierung synthetischer Testdaten

Studiengang: MAS Data Science

Beim Testen von Software in einem zum produktiven Einsatz äquivalenten System bedarf es eine grosse Menge an Testdaten. Datengeneratoren unterstützen bei der Erzeugung synthetischer Testdaten. Diese Arbeit prüft den Einsatz von Machine Learning zur Generierung synthetischer Arbeitszeitbuchungen.

Heutige Datenschutzgesetze erschweren die Verwendung von produktiven Daten in Testsystemen. Dies hat zur Folge, dass der Einsatz von Datengeneratoren zur Erzeugung synthetischer Testdaten nicht mehr wegzudenken ist. Bei komplexeren Datenstrukturen stossen diese jedoch oft an ihre Grenzen. Mit Hilfe von spezifisch für diese Strukturen erstellten Generatoren kann diese Lücke geschlossen werden.

Arbeitszeitbuchungen geben Auskunft über geleistete Arbeitsstunden. Sie beinhalten zusätzlich zum Kalendertag eine Start- und Endzeit und dadurch die Arbeitsdauer. Wie viele Stunden pro Tag gearbeitet wird, hängt vom Beschäftigungsgrad der einzelnen Person ab. Zusätzlich können auch saisonale Abweichungen auftreten. Neben der Verbuchung des Ferienbezugs ist auch die Anzahl an erstellten Buchungen pro Tag nicht einheitlich und es kommt, beispielsweise durch die Mittagspause, zu Arbeitspausen.

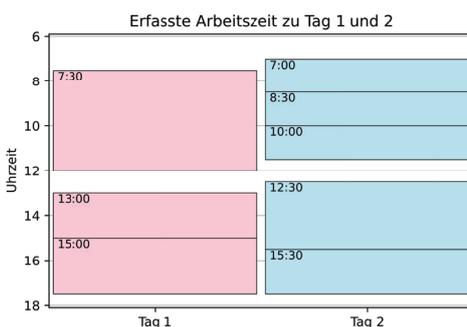
In dieser Arbeit werden Arbeitszeitbuchungen mit Hilfe von Machine Learning generiert. Die erzeugten Buchungen sollen zu einem späteren Zeitpunkt als Testdaten eingesetzt werden. Die Qualität der Buchungen wird geprüft und mit den Referenzdaten verglichen. Ein Augenmerk liegt dabei auf der Erzeugung grosser Mengen an Daten.

Da die Anzahl an Buchungen pro Tag variiert, ist die Zahl der resultierenden Datensätze zu Beginn nicht bekannt. Dies führt dazu, dass der Prozess in zwei Schritte aufgeteilt wird. In einem ersten Schritt wird die pro Tag geleistete Arbeitszeit und die Anzahl an gewünschten Buchungen erzeugt. Dies definiert die Gesamtzahl an resultierenden Buchungen. In einem zweiten Schritt werden die effektiven Arbeitszeitbuchungen generiert. Die einzelnen Buchungen beinhalten hierbei auch Informationen zu den Folgebuchungen. Einerseits kann die Startzeit einer Buchung erst nach der Endzeit der vorhergehenden Buchung liegen. Andererseits wird anhand der Buchungsdauer die maximale Buchungsdauer der Folgebuchungen innerhalb eines Tages definiert.

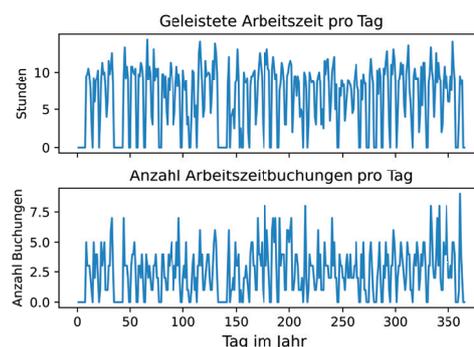
Die Arbeitszeit stellt nur einen kleinen Teil der Informationen dar, welche bei der Arbeitszeiterfassung hinterlegt werden können. Zukünftige Masterarbeiten könnten hier ansetzen und die Bewirtschaftung von Zeitsparkonten, Abwesenheiten und Überzeit- oder Geldspesen integrieren.



Christoph Wüthrich



Erfasste Arbeitszeit als Kalender



Geleistete Arbeitszeit und Anzahl Buchungen pro Tag