

Bewertung von Unsupervised und Supervised Learning-Algorithmen für eine Fluktuationsprognose

Studiengang: MAS Data Science

Das Hauptziel dieser Arbeit besteht darin zu überprüfen, ob die Erstellung einer praxistauglichen Fluktuationsprognose auf Basis von Personaldaten mittels Machine Learning-Modellen möglich ist. Dazu werden verschiedene Algorithmen aus dem Supervised und Unsupervised Learning-Bereich auf deren Anwendungsmöglichkeiten und Eignung für dieses Vorhaben untersucht.

Ausgangslage

Die frühzeitige Erkennung von ungewollten Abgängen und ein zeitnaher Start der Rekrutierung sind entscheidend, um finanzielle Folgen und Wissensverlust so gering wie möglich zu halten. Erkenntnisse aus der Analyse von Mitarbeitendendaten können helfen, Strategien zur Reduzierung der Fluktuation und Verbesserung der Mitarbeitendenbindung zu entwickeln.

Zielsetzung

Mit Hilfe von Machine Learning-Methoden soll eine Identifikation potenzieller Mitarbeitendenabgänge innerhalb der nächsten 12 Monate erreicht werden. Machine Learning-Algorithmen sollen Muster und Zusammenhänge identifizieren, die traditionellen Methoden entgehen, und präzise Vorhersagen für zukünftige Entwicklungen ermöglichen. Dabei ist es zwingend erforderlich, dass sämtliche Vorgaben des Datenschutzgesetzes für die Verarbeitung von personenbezogenen Daten berücksichtigt und eingehalten werden.

Umsetzung und Resultate

Es wird ein explorativer Ansatz verfolgt, der eine vertiefte Datenanalyse und das Ausloten verschiedener Machine Learning-Algorithmen umfasst. Mithilfe der Python Bibliothek PyCaret wurden 14 verschiedene Klassifikationsalgorithmen getestet und anhand diverser Metriken, insbesondere des F1-Scores und der Confusion Matrix, evaluiert. Die mit PyCaret evaluierten Klassifikationsmodelle zeigten schwache Signale. Die Resultate erwiesen sich als nicht robust und leistungsfähig genug, um ein Modell zu identifizieren, das in der Praxis eingesetzt werden könnte.

Durch die Kombination unterschiedlicher Features wurde mittels k-Means versucht, homogene Cluster zu bilden, um daraus Muster und Gesetzmässigkeiten zu identifizieren. Die Evaluation mittels Silhouette Plot und t-SNE deutete auf geringe Modelleistungen hin, ebenso wie der Ansatz aus dem Bereich Collaborative Filtering (Similarity Matrix).

Fazit

Das verwendete Datenset weist eine zu hohe Varianz auf, sodass computergestützte Methoden Schwierigkeiten haben, wertvolle Muster und Gesetzmässigkeiten zu erkennen oder Modelle zu entwickeln, die die erforderliche Konsistenz und Robustheit aufweisen. Eine zuverlässige Anwendung in der Unternehmenspraxis ist damit nicht gewährleistet.

Die Ergebnisse unterstreichen die Herausforderungen und Limitationen bei der Anwendung von Machine Learning-Algorithmen auf Personendaten. Soziale Aspekte und persönliche Veränderungen, die häufig zu einem Arbeitsplatzwechsel führen, sind in den vorliegenden Daten nur eingeschränkt zugänglich. In zukünftigen Untersuchungen sollten deshalb vermehrt qualitative Daten einfließen, wie persönliche Meinungen und Einschätzungen aus Interviews und Umfragen oder Verhaltensdaten, die auf Rückzug und wachsende Negativität gegenüber der Unternehmung hinweisen. Datenerhebungen, die in Richtung Profiling einer Person gehen und für ein solches Vorhaben notwendig erscheinen, sind in der Schweiz und dem europäischen Raum mit den aktuellen Datenschutzgesetzen jedoch nicht möglich.



Laura von Rohr
MAS Data Science



Wer wird uns verlassen?

Geschlecht
Alter
Arbeitsweg
Job
Lohn
Mitarbeiterzufriedenheit
Überstunden
Absenzen
Leistungsbewertung