



Anfertigen eines CAT bzw. PIKE

Author: Andreas Schau - AS

Kann ... // Algorithmus x // exact ... // (Kunden)-Problem ... berechnen / lösen?

Ja, ein Machine-Learning-Ensemble kann die Regression mit einer Genauigkeit von 80-85% durchführen.

Data Science Kernaussage:

(P) roblem):

[Welcher Frage hat für die Lösung des Kunden / Auftraggeber die größte Bedeutung?]:

Kann mit Hilfe eines Neuronalen Netzes zuverlässig eine Vorhersage zur Schlafqualität eines Menschen angefertigt werden, wenn dessen Rahmenbedingungen, wie Schlafdauer, Alkoholkonsum, REM-Schlaf-Anteil und weitere bekannt sind?

(I) ntervention:

(Bibliotheken- und Algorithmen-Auswahl, ... z.B. pandas für Finanzdaten ...)

[Welche Berechnung erwäge ich vornehmlich?]:

Für das Ensemble-Learning wurden einige Bibliotheken verwendet, darunter sind „numpy“, „sklearn“ und „xgboost“ zum Aufbereiten der Daten und zur Erstellung des Lern-Netzwerkes, sowie „bayes_opt“ zur Optimierung des besten Ensemble-Learners. Die „pandas“ Bibliothek wurde zum Laden der CSV-Daten verwendet und „matplotlib“, „seaborn“, „shap“ und „json“ um die Ergebnisse ansprechend zu visualisieren.

(K) ontrollintervention

(falls erforderlich: Bliotheken- und Algorithmen-Auswahl ... z.B. scikit-learn für Finanzdaten ...)

[Was ist die andere Möglichkeit?]:

Man hätte auch ein Ensemble aus Neuronalen Netzen via der „tensorflow“ oder „pytorch“ Bibliothek anfertigen können. Dies wäre jedoch mit deutlich höherem Aufwand verbunden und da die Scikit-Learn Implementierungen schon sehr gute Vorhersagen erlauben ist dafür kein Grund gegeben.

(E) rgebnismaß (Zielgröße(n)) – Die Evidence

[Was möchte ich / der Kunde erreichen? Z.B. Prädiktor oder Klassifikator erstellen ...]:

Es soll eine Regression mit hoher Genauigkeit von 80% und besser erreicht werden, die dann verwendet werden kann um Aussagen über die mögliche Schlaf-Effizienz von weiteren Probanden machen zu können. Zum Beispiel um diesen zu ermöglichen an Ihren Schlafbedingungen oder Ernährungsweisen zielgerichtete Veränderungen vorzunehmen, die Ihnen zu einem erholsameren Schlaf verhelfen sollten.

Anmerkungen

- [Link zum verwendeten Datensatz](#)

Literaturhinweise

Die Suche nach der besten Evidenz

1. Problem

Eine Bestimmung der wahrscheinlichen Schlaf-Effizienz anhand der gegebenen Schlafbedingungen soll mit Hilfe eines Ensemble-Learning Netzwerkes durchgeführt werden.

2. Definition einer wichtigen suchbaren Frage

Kann man mit einer Genauigkeit von über 80% mittels eines Ensemble-Learners bestimmen, wie Effizient ein Mensch unter seinen Schlafbedingungen geschlafen hat?

3. Auswahl der wahrscheinlichsten Quelle für diese Evidenz

Nach der Auswahl eines geeigneten Ensemble-Aufbaus wird dieses Modell trainiert und mit einem Test-Datensatz, der ca. 20% des Gesamt-Datensatzes groß ist, evaluiert.

4. Erstellung einer Suchstrategie

Um einen geeigneten Ensemble-Aufbau zu finden wurden viele verschiedene Ensemble-Learner erstellt, welche durch Cross-Validation auf dem gesamten Datensatz trainiert und bewertet wurden. Unter diesen gab es größere Ensembles, welche noch etwas besser als die Kleineren waren, jedoch nicht besonders viel. Daher wurde der gewählt, der die höchste Genauigkeit hatte, bei noch relativ kleiner Rechenzeit und dieser wurde dann via Hyperparameter-Optimierung noch etwas verbessert, wodurch dieser dann das beste größere Ensemble in der Vorhersagegenauigkeit überholt hat.

5.0 Zusammenstellung der Evidenzausbeute

5.1 Falle Ausbeute schlecht

Auswahl der zweit-wahrscheinlichsten
Quelle für diese Evidenz
Erstellung einer Suchstrategie
Zusammenfassung der Evidenz
Anwendung der Evidenz

Das gewählte Ensemble-Modell ist ein Gradient-Boosting-Regressor mit fein abgestimmten Parametern, welcher auf dem Trainings-Datensatz eine Vorhersagegenauigkeit via R-Squared-Metrik von 86% und auf dem Test-Datensatz von 82% erreicht hat.

6. Anwendung der Evidenz

Das oben beschriebene Modell wurde via seiner Performance in der Cross-Validierung verschieden komplexer Modelle ausgewählt, über Hyperparameter-Tuning optimiert und dann auf 80% der Gesamtdaten trainiert und auf den restlichen 20% getestet.

Das gewonnene Modell kann nun mit 82% Genauigkeit vorhersagen, wie effizient ein Proband unter seinen Gegebenheiten schlafen müsste.

Dadurch erlaubt es Menschen mit schlechtem Schlafverhalten, einen Einblick darin, an welchen Umgebungsgegebenheiten Sie etwas tun können, um sich eine möglichst gute Schlafumgebung zu schaffen.