

Klassifikation Challenge

Aufgabe

Sie arbeiten in der ML-Abteilung einer Versicherung. Ende letzten Jahres hat Ihre Abteilung einen neuen Head of ML. Dieser möchte nun gerne ein Projekt initiieren, bei dem es darum geht, eine Vorhersage zu treffen, ob ein Kunde demnächst seine KfZ-Versicherung kündigt. Ziel ist es, frühzeitig mit Kunden in Kontakt zu treten, die mit hoher Wahrscheinlichkeit demnächst kündigen und diese von einem Verbleib bei der Versicherung zu überzeugen (z.B. mit einem neuen Angebot). Sie haben nun die Aufgabe, ein ML-Modell zu entwickeln, das vorhersagt, ob ein Kunde seinen Versicherungsvertrag weiterführt oder auflöst.

Dazu wird Ihnen die Datei „records.csv“ zur Verfügung gestellt. Diese beinhaltet Daten von 164210 Kunden und hat die folgenden Spalten:

- *CustomerId*: Eine einzigartige Kennung für jeden Kunden
- *Surname*: Der Nachname des Kunden
- *CreditScore*: Ein numerischer Wert, der die Kreditwürdigkeit des Kunden darstellt
- *Geography*: Das Land, in dem der Kunde wohnt (Frankreich, Spanien oder Deutschland)
- *Gender*: Das Geschlecht des Kunden (Männlich oder Weiblich)
- *Age*: Das Alter des Kunden
- *Tenure*: Die Anzahl der Jahre, die der Kunde bei der Versicherung ist
- *NumOfProducts*: Die Anzahl anderer Versicherungsprodukte, die der Kunde nutzt
- *HasCrCard*: Ob der Kunde eine Kreditkarte besitzt (1 = ja, 0 = nein)
- *EstimatedSalary*: Das geschätzte Gehalt des Kunden
- *Exited*: Ob der Kunde die Versicherung gekündigt hat (1 = ja, 0 = nein)

Evaluation

- Die Prognosegüte wird anhand der Fläche unter der ROC-Kurve (vgl. auch AUC-Kriterium) gemessen.
- Entwickeln Sie ein Modell zur Vorhersage des Kundenverhaltens, welches die Fläche unter der ROC-Kurve optimiert.
- Machen Sie anschließend eine Prognose für den Datensatz in der csv-Datei *test.csv* und exportieren Sie die Vorhersageergebnisse als csv-Datei. Es sollte für jede CustomerId in *test.csv* eine Prognose vorliegen.

Abgabe

- Bearbeiten Sie die Aufgabenstellung in einem Colab-Notebook.
- Verschaffen Sie sich einen ausreichenden Überblick zu den Daten, z.B. durch visuelle Darstellungen oder Kennzahlen.
- Testen Sie verschiedene Modellvarianten – z.B. Modelle nach Land oder Geschlecht neben einem globalen Modell.
- Achten Sie darauf, dass Ihr Code gut lesbar und verständlich ist.
- Ein **Link auf Ihr Colab-Notebook** ist bis zum **23.03.25, 20 Uhr** an alexander.kressner@dhbw-stuttgart.de zu senden. Achten Sie bitte darauf, dass das Notebook nach laden der Daten fehlerfrei durchläuft und eine Prognose erstellt.

- Weiterhin erstellen Sie bitte eine **15-minütige Präsentation**, die Ihre Arbeitsergebnisse dokumentiert (Problemstellung, Lösungsansatz, Ergebnisse). Diese schicken Sie bitte als pdf-Datei ebenfalls bis zum **23.03.25, 20 Uhr** an alexander.kressner@dhbw-stuttgart.de
- Bei der **Abschlusspräsentation am 26.03.25** übernimmt bitte jedes Gruppenmitglied einen Präsentationsteil. Anschließend findet eine ca. 35–40-minütige Diskussion zur Fallstudie statt.