

CODING CHALLENGE – DATA SCIENTIST MARKETING

Aufgabe

Zusammen mit dieser Challenge wurde Ihnen ein Datensatz bestehend aus 3 .csv Dateien zur Verfügung gestellt. Diese beinhalten Merkmale von Kfz-Versicherungskunden (eine Datensatzbeschreibung finden Sie am Ende dieses Textes). Ihre Aufgabe besteht in der Modellierung der Kundenaffinität zum Abschluss einer Kfz-Versicherung.

Die Ergebnisse ihrer Auswertungen stellen Sie bitte im Rahmen des 60-minütigen technischen Interviews vor. Dabei haben Sie zunächst 15 Minuten Zeit, um die aus ihrer Sicht wesentlichen Ergebnisse vorzustellen. Die Form der Präsentation (frei, Folien, Jupyter-Notebook, R-Markdown...) wählen Sie dabei selbst. Anschließend findet eine Diskussion statt, in deren Rahmen Ihnen Fragen zu ihrer Vorgehensweise sowie zu Ihrem Programmcode gestellt werden. Es sollte anhand ihrer Arbeit gut nachvollziehbar sein, wie sie vorgegangen sind (beispielsweise anhand eines Jupyter-Notebooks oder R-Markdowns).

Gehen Sie dabei in folgenden Teilschritten vor:

- **Explorative Datenanalyse:** Machen Sie sich mit dem Datensatz vertraut. Identifizieren Sie dabei mögliche Probleme sowie grundlegende statistische Zusammenhänge, welche für die anschließende Modellierung wichtig sein könnten.
- **Feature Engineering:** Bereiten Sie, soweit für ihre Modellierung nötig, die Variablen geeignet auf.
- **Modellvergleich:** Entscheiden Sie sich für ein geeignetes Modell zur Prognose der Kundenaffinität zum Abschluss einer Kfz-Versicherung. Erläutern Sie wie Sie dabei vorgehen und begründen Sie ihre Entscheidung.
- **Modellbuilding:** Trainieren Sie unter Berücksichtigung der vorangegangenen Schritte das von Ihnen gewählte Modell für die Kundenaffinität zum Abschluss einer Kfz-Versicherung. Wählen Sie geeignete Metriken um die Güte des finalen Modells zu beurteilen. Zeigen Sie, welche Variablen und Zusammenhänge für Ihr finales Modell relevant sind. Überlegen Sie sich (ohne dies umzusetzen) wie Sie Ihr Modell weiter optimieren könnten.

Diese Coding Challenge ist in etwa 5h gut zu erledigen. Wir erwarten nicht, dass Sie mehr Zeit investieren. Bitte teilen Sie sich ihre Zeit entsprechend ein und konzentrieren Sie sich auf das – aus ihrer Sicht - Wesentliche. Eine erschöpfende Tiefenanalyse der Daten und oder eine aufwändige Optimierung vieler verschiedener Modelle werden dementsprechend nicht von Ihnen erwartet.

Datensatzbeschreibung

`alter_geschlecht.csv`:

- ID: ID des Kunden
- Geschlecht: Geschlecht des Kunden
- Alter: Alter des Kunden in Jahren

rest.csv:

- ID: ID des Kunden
- Fahrerlaubnis: Kunde hat Fahrerlaubnis (Ja=1)
- Regional_Code: Eindeutiger Code für Region des Wohnorts des Kunden
- Vorversicherung: Kunde hat bereits eine Kfz-Versicherung (Ja=1)
- Alter_Fzg: Alter des Fahrzeugs
- Vorschaden: Kunde hatte bereits einen Schaden am Fzg (Ja=Yes)
- Jahresbeitrag: Erwarteter Versicherungsbeitrag bei Abschluss in €
- Vertriebskanal: Code für den Vertriebskanal
- Kundentreue: Anzahl der Tage seit dem Zeitpunkt, ab dem eine Kundenbeziehung mit dem Kunden besteht

interesse.csv:

- ID: ID des Kunden
- Interesse: Kunde ist an einem Angebot interessiert (Ja=1) [Zielvariable]