

Prädiktive Modellierung - Mini-Challenge Kreditkarten

Studiengang Data Science (FS2023)

Daniel Perruchoud, Institut für Data Science

AML-Leistungsnachweis

AML-Leistungsnachweis

Beurteilung: Praktisches Verständnis via Mini-Challenge und theoretisches Verständnis via schriftliche MSP. Beide Elemente werden zu je 50% gewichtet, die MSP kann erst nach Abgabe der Mini-Challenge absolviert werden.

Kriterien Mini-Challenges:

- Analysen sind inhaltlich vollständig und korrekt, ausführlich diskutiert und sauber visualisiert,
- Notebooks sind sinnvoll strukturiert und lauffähig,
- Code verwendet, wo sinnvoll, eigene Funktionen, enthält Kommentare und berücksichtigt Best-practices in Bezug auf Namensgebung / Formatierung.

Inhalte, Lernziele und Rahmenbedingungen

Mini-Challenge Beschreibung

Aufgabe: Entwickle und evaluiere Affinitätsmodelle für Kreditkarten auf Basis von transaktionellen Kundeninformationen mittels binärer Klassifikation in Hinsicht auf personalisierte Werbekampagnen.

Inhalte:

- Aufbereitung eines Modellierungsdatensatz,
- Modellentwicklung und systematischer Performance-Vergleich,
- Vergleich der Haupteinflussfaktoren und Top-N Listen der Modelle,
- Modell-Selektion und systematische Hyperparameter-Optimierung,
- Modellvereinfachung und -beschreibung für Non-Data Scientist.

Mini-Challenge Lernziele

Lernziele: Kenntnis von Vor- und Nachteilen verschiedener

- Ansätze zur Behandlung unbalancierter Daten,
- Modellperformance-Metriken,
- Methoden zur Hyperparameter-Optimierung,
- analytischer Verfahren zur Erklärung prädiktiver Modelle,
- praxis-relevanter Ansätze zur Beschreibung prädiktiver Modelle,
- baum-basierter Vorhersagemodelle.

Mini-Challenge Grundlagen

Software: R, RStudio

R Pakete: `tidyverse`, `tidymodels`, `DataExplorer` und weitere.

Daten: Daten aus der SG DS Challenge [Cross-Selling in Banking](#).

Quellen:

- [Tidy Modeling with R](#) von M. Kuhn und J. Silge.
- [Applied Predictive Modeling](#) von M. Kuhn und K. Johnson.
- [A Gentle Introduction to tidymodels](#) von E. Ruiz.

Mini-Challenge Abgabebedingungen

Form: Die Mini-Challenge ist als Einzelarbeit einzureichen.

Abgabetermin: 9. Juni 2023

Lieferobjekte: Notebook, ein .HTML-File und ein .RMD-File.

Erarbeitung: Für die Erarbeitung der Inhalte der Mini-Challenge darf zusammengearbeitet werden. Diese Zusammenarbeit ist jedoch ausschliesslich auf algorithmische Fragen und Verständnisaspekte beschränkt. Es darf kein zusammenhängender Code oder Text von anderen oder vom Internet kopiert werden.

Product Affinity Modeling - Spezifikation und Anleitung

Mini-Challenge Vorgehen

- Laden, transformieren und überprüfen der Datenqualität mittels explorativer Datenanalyse; entfernen von Junior-Kreditkarten Kunden.
- Kombinieren der Informationen zu Kunden und Bankdienstleistungen.
- Identifizieren bestehender Kreditkartenkäufer inkl. Bestimmung des Kaufdatums und Rollup-Fensters, definiert durch 1 Monat Lag und 12 Monate History vor Kreditkartenkauf.
- Bestimmen der Nicht-Käufer zum Vergleich (inkl. Rollup-Fenster).
- Erzeugen event-bezogener Kundeninformationen für 12 Monate vor Kreditkartenkauf (analog für Nicht-Käufer).
- Konstruieren der Vermögen und Umsätze im Rollup-Fenster auf Basis der Transaktionshistorie.

Mini-Challenge Vorgehen

- Herleiten Kunden-spezifischer, statistischer Kennzahlen für Vermögen und Umsätze im Rollup-Fenster mittels Funktionen.
- Kombinieren event-bezogener Informationen von Kreditkarten-Käufern und Nicht-Käufern.
- Bereinigen unnötiger Informationen (z.B. IDs) und Überprüfen der Struktur der Modellierungsdaten mittels explorativer Datenanalyse.
- Partitionieren der Daten in Trainings- und Testdaten.
- Erstellen eines Baseline Modelles mittels logistischer Regression und den Informationen “Alter”, “Geschlecht”, “Domizilregion”, “Vermögen” und “Umsatz” vor Kreditkartenkauf.

Mini-Challenge Vorgehen

- Systematisches Explorieren von Verbesserungsmöglichkeiten des Baseline Modelles durch Erweiterung erklärender Variablen, Verwendung anderer Algorithmen und Optimierung.
- Vergleichen der Kandidatenmodelle und identifizieren des bzgl. Performance “besten” Modelles mit ROC, AUC und Precision.
- Quantitatives Untersuchen der Unterschiede von Top-N Kunden-Listen verschiedener Modelle.
- Untersuchen der globalen Wichtigkeit der Einflussfaktoren des “besten” Modelles und Reduzieren des “besten” Modelles mittels Balancieren von globaler Wichtigkeit und Modellperformance.
- Beschreiben des Mehrwerts des “finalen” Modelles in der Praxis.

Mini-Challenge Lieferobjekte

Die minimalen Lieferobjekte im Notebook umfassen:

- Anzahl Kreditkarten-Käufer und Nicht-Käufer mit kompletter 12 Monate-Rollup Information (Barplot).
- Verteilung der Kaufzeitpunkte der Kreditkarten-Käufer bzw. Vergleichszeitpunkte der Nicht-Käufer (Densityplot).
- Beschreibung der konstruierten, eigenen Variablen.
- Beschreibung von Baseline und mind. 3 Kandidaten-Modellen.
- Performance-Vergleich von Baseline und mind. 3 Kandidaten-Modellen für 10-fache Kreuzvalidierung (Metrik: ROC Kurve und AuC).

Mini-Challenge Lieferobjekte

- Performance-Vergleich von Baseline und mind. 3 Kandidaten-Modellen für Testdaten (Metriken: Konfusionsmatrix, Accuracy, Kohen's Kappa, Matthew's Korrelation, Precision, Recall und F1 als Tabelle, ROC Kurve und AuC als Abbildung).
- Globale Wichtigkeit der Einflussfaktoren von Baseline und mind. 3 Kandidaten-Modellen (Variable Importance Plot).
- Quantifizierung der Unterschiede von Top-5%, Top-10% Kunden-Listen für Baseline und mind. 3 Kandidaten-Modellen (eigene Ideen).
- Modellgüte (Lift Kurve) und Einfluss zentraler Variablen des finalen Modelles (eigene Ideen, um Non-Data Scientists zu überzeugen).