

Machine Learning im Direct-Mail-Marketing

Anwendungsfelder Business Analytics
Prof. Dr. Adem Alparslan

Lern-Ziele	Vermittlung von Anwendungswissen: Lösen eines konkreten betriebswirtschaftlichen Problems durch den Einsatz von Descriptive, Predictive und Prescriptive Analytics
Methoden	Explorative Datenanalyse, Datenvorbereitung, Entwicklung/ Optimierung und Bewertung eines Machine-Learning-Modells beispielhaft an einem ausgewählten Algorithmus, Ableiten von Handlungsempfehlungen, Präsentation der Ergebnisse
Sprache	Python
Datensatz	1/10 (sehr leicht)
Aufgabenstellung	3/10 (leicht bis mittel)

1 Betriebswirtschaftliche Problemstellung

Der (Online-) Handel zeichnet sich durch hohen Konkurrenzdruck aus. Den Problemen wechselbereiter Kunden und anderer negativer Entwicklungen versucht man verstärkt mit Marketingmaßnahmen entgegenzuwirken. Solche Marketingmaßnahmen können sich auf die Verkaufsförderung durch die direkte Ansprache von potenziellen Kunden via Mail/ Post (Direct-Mail-Marketing) beziehen. Dadurch lässt sich beispielsweise die Marge je Bestandskunde erhöhen. Andererseits verursachen die dabei eingeleiteten Maßnahmen jeweils Kosten.

In der Projektarbeit geht es im Besonderen um die Entwicklung eines Machine-Learning-Modells für Zwecke des Direct-Mail-Marketings. Als Datengrundlage dient ein Extrakt aus dem CRM-System einer Bekleidungskette aus New England mit Daten über 21.740 Kunden (vgl. Larose/ Larose (2015)). Das Ziel ist es, den Gewinn der Direct-Mail-Marketing-Maßnahmen zu maximieren, indem potenzielle Kunden mit hoher Kaufwahrscheinlichkeit identifiziert werden. Das Ergebnis des Machine-Learning-Modells sollte immer besser als der Gewinn sein, der durch die Baseline-Strategie erreicht wird.

Baseline-Strategie: Es werden immer alle Bestandskunden angeschrieben

Bei der Entwicklung und Bewertung des Machine-Learning-Modells ist nachfolgende Gewinn-/ Kosten-Matrix zu berücksichtigen.

	Kunde wird NICHT angeschrieben	Kunde wird ange- schrieben
Kunde wird NICHT kaufen	0	-2,00 EUR
Kunde wird kaufen	-16,00 EUR	16,00 EUR

Wenn ein Kunde angeschrieben wird und anschließend ein Produkt erwirbt (True Positive), dann werden 16,00 EUR (nach Abzug der Kosten für Mail/ Post) erwirtschaftet. Wenn hingegen ein Kunde angeschrieben wird, der kein Produkt erwirbt (False Positive), dann entstehen Kosten für Mail/ Post in Höhe von 2,00 EUR. Schließlich entsteht ein entgangener Gewinn (= Kosten) in Höhe von 16,00 EUR, wenn ein Kunde, der ein Produkt erworben hätte, nicht angeschrieben wird (False Negative).

2 Aufgabenstellung

1. Trainieren Sie ein Machine-Learning-Modell zur Vorhersage der Kaufwahrscheinlichkeit von Bestandskunden als Reaktion auf Werbung via Mail/ Post. Das Machine-Learning-Modell sollte in der Lage sein, jene Kunden zu identifizieren, bei denen ein Kauf sehr wahrscheinlich ist,

sodass ein maximaler Gewinn aus der Direct-Mail-Kampagne erreicht wird.

2. Realisieren Sie das Machine-Learning-Modell in Python.
3. Verwenden Sie als Datengrundlage die Datei: `Case_Mailing_2022.csv`.
4. Entwickeln Sie das Machine-Learning-Modell auf Basis des Random-Forest-Algorithmus und weiterer zwei Algorithmen (z.B. Gradient Boosting, Neuronales Netz, Nächste-Nachbarn-Klassifikation, Naive Bayes und Support Vector Machines).
5. Sofern Sie ein Seed setzen müssen, verwenden Sie: `random_state = 4711`.
6. Trainieren Sie das Machine-Learning-Modell nur mit einem Teil der Daten. Zwecks Vergleichbarkeit: Splitten Sie die Daten in Trainingsdaten, Testdaten 1 und Testdaten 2 in einem im einem Verhältnis von ca. 75-20-5:

```
X_train, X_test_2,  
y_train, y_test_2 = train_test_split(X, y,  
test_size = 0.05, ...)
```

```
X_train, X_test_1,  
y_train, y_test_1 = train_test_split(X_train,  
y_train, test_size = 0.2, ...)
```

Die Trainingsdaten (`X_train`, `y_train`) verwenden Sie, um das Machine-Learning-Modell zu trainieren. Die Testdaten 1 (`X_test_1`, `y_test_1`) sollten genutzt werden, wenn eventuelle Adjustierungen erforderlich sind. Die Testdaten 2 (`X_test_2`, `y_test_2`) dienen dazu, das trainierte Machine-Learning-Modell abschließend zu evaluieren.

7. Bewerten Sie die Güte des Machine-Learning-Modells anhand des generierten Gewinns unter Berücksichtigung der oben angegebenen Gewinn-/ Kosten-Matrix. Grundlage für die Bewertung sind die Testdaten 2.
8. Die folgenden Phasen eines Machine-Learning-Projekts (CRISP-DM) sollten aus der Projektarbeit und dem Python-Skript ersichtlich sein:

- Data Understanding,
- Data Preprocessing,
- Modeling und
- Evaluation.

Business Understanding und Deployment sind also **obsolet**.

3 Bewertung

3.1 Projektarbeit

1. Erfüllung/ Einhaltung der formalen Anforderungen an das wissenschaftliche Arbeiten
2. Darstellung/ Diskussion der theoretischen/ konzeptionellen Grundlagen
3. Hinreichend gute Abdeckung aller oben genannten Phasen eines Machine-Learning-Projekts und methodisch einwandfreie Entwicklung/ Optimierung des Machine-Learning-Modells/ Ableitung von Handlungsempfehlungen
4. Güte des entwickelten Random-Forest-Modells gemessen in generiertem Gewinn
5. Kritische Diskussion über mögliche Anwendungsgrenzen des entwickelten Machine-Learning-Modells
6. Extra-Punkte gibt es für den Einsatz von Methoden, die die Nachvollziehbarkeit/ Transparenz des trainierten Machine-Learning-Modells erhöhen (Feature Importance, XAI etc.)
7. Fügen Sie das Python-Skript als Anhang (**im** Dokument und **nicht zusätzlich** zum Dokument) der Projektarbeit bei. Versehen Sie das Coding mit Inline-Kommentaren
8. Sofern die Projektarbeit als Teamleistung erbracht wird: Nachweis darüber, wer welche Abschnitte verfasst hat

3.2 Sonstige Beteiligung: Präsentation der Ergebnisse

1. Sie werden abschließend das Machine-Learning-Projekt präsentieren; bereiten Sie hierzu eine Präsentation vor (im Idealfall PPT-Datei).
2. Im Rahmen der Präsentation sind auf wesentliche Ergebnisse und Erkenntnisse des Projekts einzugehen.
3. Es werden meinerseits Rückfragen u.a. zu den Grundlagen und zur Implementierung gestellt.

4 Literatur

Larose, D.T./ Larose, C.D.: Data Mining and Predictive Analytics, 2. Aufl., Hoboken, New Jersey 2015