

Business Analytics B

Gruppenprojekt

Studiengang: Business Analytics
Semester: WiSe 2022 / 2023
Dozent: Prof. Dr. Jan Schemm
Datum: 16.11.2022

Hintergrund

Merkantil ist ein US-amerikanischer Einzelhändler, der ein erweitertes Supermarkt-Sortiment über eine Menge von Filialen in verschiedenen US-Bundesstaaten vertreibt.

Das Unternehmen setzt Nachfrageprognosen zur Entscheidungsunterstützung in verschiedenen Funktions- und Prozessbereichen ein. Bisher arbeitet Merkantil hierbei zum Großteil mit qualitativen Prognosen und einfachen quantitativen Modellen auf Basis von Tabellenkalkulationsprogrammen.

Das Unternehmen möchte prüfen, ob durch den Einsatz erweiterter quantitativer Verfahren die Prognosequalität und damit die Entscheidungsgrundlage in den verschiedenen Einsatzbereichen signifikant verbessert werden kann.

Ausgangssituation

Inspiziert vom Netflix-Prize¹ führt Merkantil eine Ausschreibung in Form eines Kaggle-Wettbewerbs durch, um qualifizierte Dienstleister für die Entwicklung von Nachfrageprognosemodellen zu bewerten.

Ihre Gruppe hat sich für die Teilnahme an der Ausschreibung entschieden.

¹ https://en.wikipedia.org/wiki/Netflix_Prize

Daten

Im Rahmen der Ausschreibung stellt Merkantil eine Stichprobe historischer Nachfragedaten einer ausgewählten Filiale zur Verfügung.

Die Daten umfassen drei Datensätze:

1. Abverkaufsdaten („sales.csv“) einer Auswahl von Produkten in verschiedenen Warengruppen / Kategorien und Abteilungen:
 - *date*: das Datum im Format "y-m-d"
 - *item_id*: die ID des Produkts
 - *dept_id*: die ID der Abteilung, zu der das Produkt gehört
 - *cat_id*: die ID der Kategorie, zu der das Produkt gehört
 - *sales*: der Abverkauf in Stück
2. Kalenderdaten („calendar.csv“) zu den Daten, an denen die Produkte verkauft wurden:
 - *date*: das Datum im Format "y-m-d"
 - *week_id*: die ID der Woche
 - *event_name_1*: falls auf dem Datum ein Ereignis liegt, der Name dieses Ereignisses
 - *event_type_1*: falls auf dem Datum ein Ereignis liegt, der Typ dieses Ereignisses
 - *event_name_2*: falls auf dem Datum ein zweites Ereignis liegt, der Name dieses Ereignisses
 - *event_type_2*: falls auf dem Datum ein zweites Ereignis liegt, die Art dieses Ereignisses
 - *snap*: eine binäre Variable (0 oder 1), die angibt, ob die Filiale am entsprechenden Datum SNAP-Einkäufe² erlaubt; der Wert 1 zeigt an, dass SNAP-Einkäufe erlaubt sind
3. Preisinformationen („prices.csv“) über den Preis der verkauften Produkte pro Datum:
 - *item_id*: die ID des Produkts
 - *week_id*: die ID der Woche
 - *price*: der durchschnittliche Verkaufspreis des Produkts für die angegebene Woche

Die Daten decken einen Nachfragezeitraum von insgesamt 5 Jahren vom 01.03.2011 bis 29.02.2016 ab.

Die Ausschreibungsteilnehmer sollen den tagesgenauen Gesamtabatz der Filiale für den Monat März 2016 unter Nutzung eines quantitativen Modells prognostizieren.

Die durch Merkantil auf Basis eines Evaluationsdatensatzes für März 2016 bewertete Prognosegenauigkeit ist ein Element der Dienstleister-Bewertung und -Auswahl.

² Die US-Bundesregierung bietet eine Ernährungshilfe an, die als *Supplement Nutrition Assistance Program* (SNAP) bezeichnet wird. SNAP bietet einkommensschwachen Familien und Einzelpersonen regelmäßige Geldleistungen zum Kauf von Lebensmitteln auf einer Debitkarte.

Aufgabenstellung

Die konkrete Aufgabenstellung umfasst die folgenden Bestandteile:

1. Visualisieren und analysieren Sie die historischen Daten:
 - a. Importieren Sie die Daten in passende Datenstrukturen.
 - b. Verschaffen Sie sich einen Überblick über die Daten im Hinblick auf die Modellbildung mithilfe von passenden Visualisierungen. Betrachten Sie dabei den zeitlichen Verlauf der Nachfrage, die Zerlegung der Zeitreihe sowie (Auto-)Korrelationen.
2. Bereiten Sie die Daten für Ihre Modellbildung vor:
 - a. Kodieren Sie Datenattribute in geeignete Repräsentationsformen.
 - b. Generieren Sie aus den Daten ggf. weitere Attribute / Merkmale als Basis für Ihre Modelle.
 - c. Wählen Sie geeignete Merkmale als Grundlage für Ihre Modelle aus.
3. Entwickeln Sie Modelle für die Nachfrageprognose:
 - a. Entwickeln Sie ein einfaches Modell als Maßstab / Benchmark für Ihre Prognose.
 - b. Entwickeln Sie mindestens zwei weitere Prognosemodelle und evaluieren Sie die Modell-Leistung
 - c. Wählen Sie ein aus Ihrer Perspektive optimales Modell aus.
4. Stellen Sie Ihr Ergebnis für Merkantil zusammen:
 - a. Prognostizieren Sie den tagesgenauen Gesamtabsatz der Filiale für den Zeitraum vom 01.03.2016 – 31.03.2016.
 - b. Speichern Sie Ihre Prognose als CSV-Datei mit folgenden Attributen:
 - *date*: Prognosedatum
 - *sales*: prognostizierter Absatz am Prognosedatum

Bewertung

Einzureichende Ergebnisse

Merkantil erwartet folgende Ergebnisse für eine Teilnahme an der Ausschreibung:

- Die Implementierung der Aufgaben in Python inkl. Dokumentation des Lösungswegs in Form eines Jupyter-Notebooks: „Forecast.ipynb“.
- Die in Aufgabe 4 beschriebenen Prognosewerte in einer CSV-Datei: „Forecast.csv“
- Die ausgefüllte und von allen Gruppenmitgliedern unterschriebene Erklärung auf der folgenden Seite im PDF-Format: „Erklärung.pdf“

Bewertungskriterien

Merkantil bewertet die Einreichungen anhand der folgenden Kriterien:

- Dokumentiertes Vorgehen
 - Nachvollziehbare, methodische Struktur des Jupyter-Notebooks
 - Ausführliche Dokumentation des Vorgehens in Form von Code-Kommentierung und/oder separater Kommentierung in ergänzenden Jupyter-Zellen
- Qualität der Modellbeurteilung und -auswahl
 - Auswahl geeigneter Merkmale und Modellbildung
 - Anzahl der getesteten und bewerteten Modelle
 - 1 einfaches Modell und
 - mindestens 2 weitere Modelle
 - Modellbewertung und -auswahl anhand geeigneter Verfahren
- Prognosegenauigkeit des Modells anhand der Prognosen für den Evaluationsdatensatz auf Basis der Metrik „mean absolute error“ (MAE)
 - Final bewertet wird die Prognosegenauigkeit für die letzten 24 Tage des Evaluationsdatensatzes.
 - Die ersten 7 Tage des Evaluationsdatensatzes können vorab zur Prognosevalidierung auf der Kaggle-Plattform verwendet werden.

Erklärung

Erklärung

Hiermit bestätigen wir, dass wir die Aufgabenstellung eigenhändig und/oder in Zusammenarbeit mit den u.g. Gruppenmitgliedern bearbeitet haben und keine Lösungsbestandteile von anderen Gruppen übernommen haben.

Bewertung

Für die Bewertung wählen wir eine der folgenden beiden Alternativen (bitte ankreuzen):

- ☐ Wir bitten um einheitliche Bewertung des Gruppenprojekts als Gruppenleistung. In diesem Fall erhält jedes Gruppenmitglied die gleiche Bewertung.

oder

- ☐ Wir bitten um eine individuelle Bewertung jedes Gruppenmitglieds. Bitte geben Sie in diesem Fall an, welches Gruppenmitglied für welchen der u.g. Teile der Aufgabenstellung verantwortlich ist:

Teil der Aufgabenstellung	Verantwortliches Gruppenmitglied
1.	
2.	
3. a, b (mind. 1 Modell)	
3. b (mind. 1 Modell), c 4.	

Unterschriften

Name	Datum	Unterschrift