

FALLSTUDIE

Zusatzinformationen zur Erstellung der Fallstudie im
Kurs: DLMDWME01 – Model Engineering

INHALTSVERZEICHNIS

1. Erstellen eines Prognosemodells des Kreditkartenzahlungsverkehrs für Online-Einkäufe	2
1.1. Dateiname	2
1.2. Liste der Zahlungsdienstleister (PSP=payments service providers) und Servicegebühren	2
1.3. Spaltenbeschreibung	2
1.4. Weiterführende Informationen von geschäftlicher Seite	2
2. Automatisierung des Bereitschaftsdienstplans für Einsatzfahrer mithilfe eines Vorhersagemodells ..	2
2.1. Dateiname	2
2.2. Spaltenbeschreibung	2
2.3. Weiterführende Information von geschäftlicher Seite	3
3. Effizienter Flugbetrieb bei East Carmen Airlines mittels eines Machine-Learning Modells zur Ankunftszeit-Vorhersage	3
3.1. Dateiname	3
3.2. Spaltenbeschreibung – Fluginformation	3
3.3. Spaltenbeschreibung – Bodenprozesse	4
3.4. Zusätzliche Information von geschäftlicher Seite	4

1. ERSTELLEN EINES PROGNOSEMODELLS DES KREDITKARTENZAHLUNGSVERKEHRS FÜR ONLINE-EINKÄUFE

1.1. Dateiname

- **zip-Ordner:** use_case_1.zip
- *PSP_Jan_Feb_2019.xlsx*: Liste an Kreditkartentransaktionen der DACH Länder (Deutschland, Österreich, Schweiz)

1.2. Liste der Zahlungsdienstleister (PSP=payments service providers) und Servicegebühren

Name	Gebühr für erfolgreiche Transaktionen	Gebühr für fehlgeschlagene Transaktionen
Moneycard	5 Euro	2 Euro
Goldcard	10 Euro	5 Euro
UK_Card	3 Euro	1 Euro
Simplecard	1 Euro	0,5 Euro

1.3. Spaltenbeschreibung

- **tmstp:** Zeitstempel der Überweisung/Transaktion
- **country:** Land der Überweisung
- **amount:** Überweisungsbetrag
- **success:** wenn "1", dann ist die Überweisung erfolgreich
- **PSP:** Name des Zahlungsdienstleisters (PSP = payments service provider)
- **3D_secured:** wenn "1", dann ist der Kunde 3D-identifiziert (dadurch eine noch sicherere Online-Kreditkartenzahlung)
- **card:** Kreditkartenanbieter (Master, Visa, Diners)

1.4. Weiterführende Informationen von geschäftlicher Seite

Oftmals scheitern Überweisungen beim ersten Mal. Deshalb versuchen viele Kunden, ein und dieselbe Überweisung öfters zu tätigen. Wenn zwei Überweisungen in derselben Minute, aus demselben Land und mit demselben Überweisungsbetrag stattfinden, dann ist (für eine angemessene Anzahl an Überweisungsversuchen) davon auszugehen, dass es sich um denselben Zahlungsveruch für einen Einkauf handelt. Berücksichtige beim Entwickeln eines Machine-Learning Modells diesen Fall von mehreren Zahlungsveruchen für denselben Einkauf!

2. AUTOMATISIERUNG DES BEREITSCHAFTSDIENSTPLANS FÜR EINSATZFAHRER MITHILFE EINES VORHERSAGEMODELLS

2.1. Dateiname

- **zip-folder:** use_case_2.zip
- *sickness_table.csv*: Anzahl an Krankenständen, Notrufen, Ersatzfahrern im Bereitschaftsdienst und Zahl der zusätzlichen Einsatzfahrer, die aktiviert werden müssen (auf Tagesbasis)

2.2. Spaltenbeschreibung

- **date:** Datum
- **n_sick:** Anzahl der Einsatzfahrer, die einen Krankenstand angemeldet haben
- **calls:** Anzahl der Notrufe

- **n_duty**: Anzahl der Einsatzfahrer im Dienst
- **n_sby**: Anzahl der verfügbaren Ersatzfahrer
- **sby_need**: Anzahl der Ersatzfahrer, die aktiviert werden
- **dafted**: Anzahl der zusätzlichen Einsatzfahrer, die aktiviert werden müssen, wenn die Anzahl der Einsatzfahrer im Bereitschaftsdienst nicht ausreicht

2.3. Weiterführende Information von geschäftlicher Seite

- Die Kollegen aus dem Planungsbereich behaupten, dass eine fixe Anzahl an Bereitschaftsfahrern ($n_{sby} = 90$) keiner effizienten Planung entspricht, weil es Tage mit zu wenigen, aber auch Tage mit zu vielen Bereitschaftsfahrern gibt. Das Ziel der Kollegen aus dem Planungsbereich ist es, den Bereitschaftsdienst dynamischer zu planen, d.h. dass saisonale Muster abgebildet werden und nicht eine fixe Anzahl an Fahrern vorgehalten werden.
- Tage mit zu wenigen Bereitschaftsfahrern sollen unbedingt vermieden werden!

3. EFFIZIENTER FLUGBETRIEB BEI EAST CARMEN AIRLINES MITTELS EINES MACHINE-LEARNING MODELLS ZUR ANKUNFTSZEIT-VORHERSAGE

3.1. Dateiname

- **zip-Ordner**: use_case_3.zip
- **flight_information.csv** : Fluginformation in einem bestimmten Zeitfenster, wo jeder Datenpunkt ein sogenanntes leg beschreibt, d.h. ein Flugereignis vom Abflughafen zum Ankunftsflughafen mit allen relevanten Flug- und Crew-Informationen
- **ground_information.csv** Informationen über relevante Bodenprozesse auf unserem Heimat-Flughafen (dem Hub) "East Carmen". Jeder Datenpunkt beschreibt für ein Ankunfts- und Abflugs-Paar die Prozesse zwischen zwei Flugereignissen (inbound = Landung, outbound = Abflug).

3.2. Spaltenbeschreibung – Fluginformation

- **leg_no**: eindeutige Identifizierung eines Flugereignisses an einem bestimmten Tag zu einer bestimmten Uhrzeit mit einer bestimmten Flugnummer
- **fn_carrier**: Name der Airline
- **fn_number**: Flugnummer, die an einem Betriebstag immer eindeutig sein soll
- **dep_ap_sched**: geplanter Abflughafen
- **arr_ap_sched**: geplanter Ankunftsflughafen
- **dep_sched_date**: geplantes Abflugdatum
- **dep_sched_time**: geplante Abfluguhrzeit
- **arr_sched_date**: geplantes Ankunftsdatum
- **arr_sched_time**: geplante Ankunftsuhrzeit
- **m_offblockdt**: Zeitstempel des Abflugs
- **m_onblockdt**: Zeitstempel der Ankunft
- **ac_registration**: Flugzeug-Registrierungsnummer (das "Nummernschild" des Flugzeugs)
- **change_reason_code**: Grund der Verspätung (wird erst nach einem Flugereignis zugeordnet)
- **dep_delay**: Abflugverspätung
- **Ac Type Code**: Flugzeugtyp (Beispiel: 320 = Airbus A320)

- *trans_time*: tatsächliche Umsteigzeit jenes Crewmitglieds mit der minimalsten Umsteigzeit
- *sched_trans_time*: geplante Umsteigzeit jenes Crewmitglieds mit der minimalsten Umsteigzeit
- *Crew Group*: Zuweisung, was mit einer Crew zwischen zwei Flugereignissen passiert
 - Start : erster Flug des Tages
 - A : die gesamte Crew bleibt am Flugzeug für den nächsten Flug
 - B, B2: alle Crewmitglieder wechseln gemeinsam das Flugzeug
 - C: zumindest ein Crew-Mitglied wechselt das Flugzeug für den nächsten Flug
- *TLC_trans*: Name der Crew-Mitglieder am Flug und eine weitere Zusatzinformation zu jedem Crewmitglied. Die Kollegen des Fachbereichs können leider nicht genau zuordnen, was diese Zusatzinformation besagt.
- *crew_type_change*: Rang der Crew-Mitglieder (cp = Pilot, ca = Kabinenbesatzung), die neu auf das Flugzeug kommen
- *Sched Groundtime*: geplante Bodenzeit zwischen zwei Flügen
- *Act Groundtime*: tatsächliche Bodenzeit zwischen zwei Flügen

3.3. Spaltenbeschreibung – Bodenprozesse

- Jeder Datenpunkt besteht aus einem “inbound” (=ankommenden) Flug an unser Hub „East Carmen“ und einem “outbound” (=abhebenden) Flug von unserem Hub „East Carmen“
- Dauer des Caterings, Dauer der Reinigung des Flugzeugs und Dauer des Passagierboardings auf unserem Hub “East Carmen” zwischen einem inbound und einem outbound leg
- *day_of_origin*: Tag des Flugbetriebs (auch im Datensatz Fluginformation angegeben)
- *ac_type*: Flugzeugtyp (auch im Datensatz Fluginformation angegeben)
- *fn_number*: Flugnummer (auch im Datensatz Fluginformation angegeben)
- *ac_registration*: Kennzahl des Flugzeugs (auch im Datensatz Fluginformation angegeben)
- *mingt*: minimal mögliche geplante Bodenzeit für einen Flugzeugtyp
- *dep_leg_inbound*: Name des Abflughafens des ankommenden Fluges
- *arr_leg_inbound*: Name des Ankunftsflughafens des ankommenden Fluges
- *arr_leg_outbound*: Name des Ankunftsflughafens des abhebenden (=outbound) Fluges
- *sched_inbound_dep*: geplante Abflugzeit des ankommenden Fluges
- *sched_inbound_arr*: geplante Ankunftszeit des ankommenden Fluges
- *sched_outbound_dep*: geplante Abflugzeit des abhebenden (=outbound) Fluges
- *sched_outbound_arr*: geplante Ankunftszeit des abhebenden (=outbound) Fluges
- *sched_turnaround*: geplante Bodenzeit zwischen zwei Flügen
- *leg_inbound*: Leg-ID des ankommenden Fluges
- *leg_outbound*: Leg-ID des abhebenden Fluges
- *catering_duration*: Dauer des Caterings zwischen zwei Flügen in Minuten
- *cleaning_duration*: Dauer der Reinigung zwischen zwei Flügen in Minuten
- *pax_boarding_duration*: Dauer des Passagierboardings zwischen zwei Flügen in Minuten

3.4. Zusätzliche Information von geschäftlicher Seite

BI (=business intelligence)-Kollegen haben Informationen über die Flugereignisse aus unterschiedlichen Fachbereichen gewonnen: Flugbewegung, Crew, und Bodenprozesse. Jeder Fachbereich verwendet jedoch

eine unabhängige Datenbank. Die BI-Kollegen weisen deshalb darauf hin, dass sich in einigen Spalten der Datensätze Informationen wiederholen, aber sich auch Daten mit schlechter Qualität sowie nicht konsistente Datenpunkte befinden. Deshalb ist es vor allem für diese Fallstudie von zentraler Bedeutung, den finale Datensatz zu säubern, zu transformieren und Konsistenzchecks einzuführen, bevor ein Machine-Learning Modell auf den Datensatz angewendet wird.