

ENTWICKLUNG EINES ATFM-SLOT-VORHERSAGEMODELLS FÜR DIE FLUGPLANUNG AUF BASIS VON MACHINE-LEARNING ALGORITHMEN

F. Ebus, Deutschland

F. Knabe, Deutsches Zentrum für Luft- und Raumfahrt e.V. - Institut für Flugführung,
Deutschland

M. Pak, Austrian Airlines AG, Österreich

Zusammenfassung

Die zuverlässige Vorhersage potenzieller ATFM-Slots ermöglicht Airlines eine angemessene Reaktion durch Umplanungen, um die Auswirkungen auf den Flugbetrieb möglichst gering zu halten. Das hier vorgestellte Vorhersagemodell wurde zur Prognose für das Auftreten von ATFM-Slots für den Flugbetrieb von Austrian Airlines (AUA) entwickelt. Das Modell beruht auf Machine-Learning Algorithmen und berechnet zu jedem Flug die Wahrscheinlichkeit für das Eintreten eines ATFM-Slots bis zu 24 Stunden im Voraus. Diese Vorhersagen ermöglichen der Flugplanung frühzeitiger zu agieren, um Kostenbelastungen zu vermeiden. Es werden die verwendeten Datensätze, die ausgetesteten Machine-Learning Modelle sowie die wichtigsten Ergebnisse und Einschränkungen erläutert.

1. EINLEITUNG

Im europäischen Luftraum werden bei Kapazitätsproblemen an Flughäfen oder in Lufträumen Maßnahmen zur Verkehrsflussregelung umgesetzt. Hierbei handelt es sich vor allem um regulierte Startzeiten, sogenannte „Air Traffic Flow Management“ (ATFM)-Slots [1]. Diese stellen ein großes Problem hauptsächlich für die Flugbetriebe der Airlines dar. So waren im Jahr 2019 mehr als 12% aller Flüge im europäischen Luftraum auf Grund von ATFM-Slots verspätet [2], was eine erhebliche Kostenbelastung für Fluglinien darstellt [3]. Nach der für die Luftfahrt einschneidenden Corona-Pandemie [4] und den damit verbundenen Einsparungsmaßnahmen wird es für die einzelnen Airlines noch wichtiger sein, effizient zu steuern und operative Kostenbelastungen zu minimieren [5].

In diesem Beitrag wird ein Vorhersagemodell für das Auftreten von ATFM-Slots für den Flugbetrieb von Austrian Airlines vorgestellt, das im Rahmen einer Masterarbeit an der TU Braunschweig in Zusammenarbeit mit Austrian Airlines und dem DLR Institut für Flugführung entstanden ist. Dieses Modell erlernt mithilfe von Machine-Learning Algorithmen (Logistische Regression und Random Forest) historische Muster für das Eintreten von ATFM-Slots [6]. Für jeden Flug wird die Wahrscheinlichkeit für das Eintreten eines ATFM-Slots, beginnend 24 Stunden vor dem jeweiligen Flugereignis, berechnet. Die Vorhersagen werden mit den aktuellen Wetterprognosen und Flugbewegungen stündlich aktualisiert, sodass eine Anpassung an die aktuellen Gegebenheiten möglich ist. Dies ermöglicht der Flugplanung sowohl frühzeitig zu agieren als auch kurzfristig Maßnahmen einzuleiten, um Kostenbelastungen zu vermeiden.

Nach einem Einblick in die Problemstellung und der Vorstellung der Methodik wird der Algorithmus beschrieben, auf welchem das Machine-Learning Modell beruht. Darüber hinaus werden die verwendeten Datensätze und ihre Relevanz erläutert. Hierzu gehören im Wesentlichen Wetterinformationen, sowie Daten aus dem Flugbetrieb der Austrian Airlines und der EUROCONTROL. Vor allem

wird detailliert darauf eingegangen, wie das Tool für die Flugplanung genutzt werden kann. Hierfür wird ein Konzept für eine grafische Benutzeroberfläche vorgestellt. Zudem werden Weiterentwicklungsmöglichkeiten und Einschränkungen des Vorhersagetools erläutert. Programmiert wurde das Modell mit der Sprache „Python“, die vor allem im Umfeld von Data Science zur Anwendung kommt [6].

2. PROBLEMATIK VON ATFM-SLOTS FÜR DIE FLUGPLANUNG

Ein wesentliches Problem der Flugplanung ist, dass ATFM-Slots maximal zwei Stunden vor der geplanten Startzeit vergeben werden [1]. Häufig ist die Bekanntgabe der ATFM-Slots noch kurzfristiger. Der Flugplan ist zu diesem Zeitpunkt schon mindestens seit einer Stunde aufgegeben. In diesem Zeitraum Veränderungen an bereits geplanten Flügen vorzunehmen, Flugzeuge oder Flugbesatzungen zu tauschen, um somit auftretende Verspätungen zu minimieren, bedarf eines erhöhten Koordinierungsaufwands.

In BILD 1. ist dargestellt, wie die geplante Vorhersage der ATFM-Slots die Situation verbessern könnte. Die Vorhersagen stehen ab 24 Stunden im Voraus zur Verfügung. Eine Umplanung einzelner Flugrouten, Rotationen oder die Umplanung von Flugbesatzungen wären noch realistisch. Zudem würde die Flugplanung eine erste Einschätzung bekommen, ob an einem Tag Flüge mit ATFM-Slots auftreten oder nicht. Dies führt zu einer besseren Einschätzung der gesamten Verspätungen und der anstehenden Arbeitsbelastung der Flugplanung. ATFM-Slots sind für den Flugbetrieb von Nachteil, wenn diese eine Verspätung generieren [3]. Deswegen werden in der Berechnung der Wahrscheinlichkeiten für ATFM-Slots nur ATFM-Slots berücksichtigt, welche eine Verspätung von 15 min oder mehr gemessen von der ursprünglichen Startzeit zur Folge haben. Diese werden im Nachfolgenden „Delay-Slot“ (D-Slot) genannt. 15 Minuten ist ein Schwellenwert, ab dem Verspätungen signifikante Auswirkungen auf die operationelle Stabilität haben [7, 8].

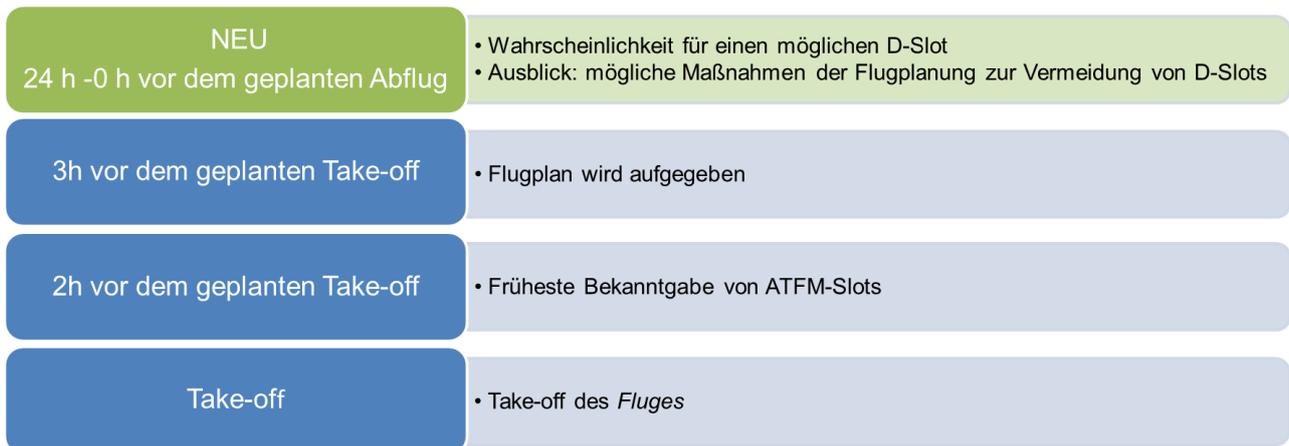


BILD 1. Flugplanung in Bezug auf ATFM-Slots

3. ALGORITHMEN DES MASCHINELLEN LERNENS

Bei der vorliegenden Fragestellung handelt es sich um ein Klassifikationsproblem [6]: D-Slot oder kein D-Slot. Es liegen historische Daten vor, auf denen das Modell trainiert wird und überprüfen kann, ob die Vorhersage zutrifft oder nicht. Somit handelt es sich um einen überwachten Lernprozess. Grundlagen zu Fragestellungen des maschinellen Lernens sind beispielsweise in Cook et al [6] beschrieben.

Es wird zwischen folgenden Ausgängen eines Vorhergesageereignisses unterschieden:

- das Eintreten des Ereignisses wurde vorausgesagt und es ist eingetreten (RP: richtig positiv)
- das Eintreten des Ereignisses wurde negiert und es ist nicht eingetreten (RN: richtig negativ)
- das Eintreten des Ereignisses wurde vorhergesagt, es ist aber nicht eingetreten (FP: falsch positiv)
- das Eintreten des Ereignisses wurde negiert, aber es ist eingetreten (FN: falsch negativ)

Hier werden zwei Algorithmen des maschinellen Lernens zu Prognosezwecken angewendet – die Logistische Regression und das Entscheidungsbaumverfahren „Random Forest“ [9]. Diese werden verwendet, da sie vielversprechende Ergebnisse auf einer Vielzahl von Datensätzen liefern [6] und bei ähnlichen Fragestellungen bereits zur Anwendung gekommen sind [9–11]. Beide Algorithmen lassen sich anhand der gleichen Metriken auswerten und somit vergleichen [6].

Besonders anschaulich lassen sich die berechneten Ergebnisse des Algorithmus in einer Konfusionsmatrix darstellen, wie in TAB 1. zu sehen ist. Die eingetroffenen Vorhersagen liegen auf der Hauptdiagonalen (blau hervorgehoben). Wichtige Metriken zur Auswertung der Ergebnisse lassen sich zudem aus der Konfusionsmatrix gewinnen [6].

Drei dieser wichtigen Metriken sind die Relevanz, die Sensitivität und das f1-Maß. Bei allen drei Metriken ist ein Wert von 1 optimal und 0 am schlechtesten. Die Relevanz – im englischen precision - ist das Verhältnis zwischen den Richtig Positiven und der Summe der Richtig und Falsch Positiven. Es sagt aus, wie viel Prozent der positiv vorhergesagten Ergebnisse tatsächlich richtig positiv sind. Die Sensitivität dagegen ist das Verhältnis zwischen Richtig Positiven und der Summe aus Richtig Positiven und Falsch Negativen. Es sagt aus, wie viele der insgesamt richtig detektierten Ergebnisse, richtig positive sind. Das f1-Maß ist ein Maß für das Verhältnis aus Relevanz und Sensitivität [6]. Mithilfe dieses Wertes kann das Verhältnis zwischen den beiden Metriken Sensitivität und Relevanz optimiert werden.

$$(1) \text{ F1 Maß} = 2 * \frac{\text{Relevanz} + \text{Sensitivität}}{\text{Relevanz} + \text{Sensitivität}}$$

Die Kalibrierungskurve ist eine weitere grafische Aufarbeitung der vorhergesagten Ergebnisse [6]. Diese verdeutlicht den Zusammenhang zwischen der berechneten Wahrscheinlichkeit und den tatsächlichen Beobachtungen. Im Anhang sind verschiedene in Kapitel 5 genauer beschriebene Kalibrierungskurven dargestellt. Die berechnete Wahrscheinlichkeit (x-Achse) ist gegen tatsächlich eingetroffene Ereignisse (y-Achse) aufgetragen (blaue Linie). Ein perfekt kalibriertes Modell würde auf der gestrichelten Linie liegen. Es werden jeweils die minimale und maximale berechnete Wahrscheinlichkeit, sowie die tatsächliche Häufigkeit ausgegeben. Zudem wird angegeben, wie viele Datenpunkte jeweils den verschiedenen Bereichen zugeordnet werden können. In den nachfolgenden Kapiteln wird genauer auf die Interpretation dieser Kalibrierungskurve eingegangen.

		Vorhergesagtes Ereignis	
		Negativ	Positiv
Wahres Ereignis	Negative Kategorie	Richtig Negativ	Falsch Positiv
	Positive Kategorie	Falsch Negativ	Richtig Positiv

TAB 1. Konfusionsmatrizen nach Müller [6]

3.1. Logistische Regression

Bei der Logistischen Regression wird berechnet, ob das Ereignis eintritt (1) oder nicht (0) [10, 12]. Zudem können die Wahrscheinlichkeiten für das Eintreten des Ereignisses berechnet werden.

Die zugrundeliegende Gleichung der Logistischen Regression (2) besteht aus Koeffizienten (ω) und Eingabedaten bzw. Merkmalen (x). Damit wird die Hilfsvariable z berechnet. Der Algorithmus berechnet für jedes Merkmal (x) einen dazugehörigen Koeffizienten (ω), welcher aussagt, wie das bestimmte Merkmal das Eintreten des Ereignisses beeinflusst: ein positiver Wert bestärkt das Eintreten des Ereignisses und ein negativer beeinflusst das Eintreten des Ereignisses negativ. Der Vorteil dieses Algorithmus ist, dass durch die Berechnung und Ausgabe der Koeffizienten ω die Ergebnisse nachvollziehbar und anschaulich sind. Zudem lässt sich, wie aus Gleichung (3) ersichtlich wird, die jeweilige Wahrscheinlichkeit einfach berechnen. Dazu wird nur e , die Euler'sche Zahl, und die Hilfsvariable z benötigt.

$$(2) z = \omega_0 + \omega_1 * x_1 + \dots + \omega_n * x_n$$

$$(3) y \approx 1 / (1 + e^{-(z)})$$

3.2. Entscheidungsbaumverfahren

Entscheidungsbaumverfahren klassifizieren die vorliegenden Daten anhand von Ja/Nein Fragen [9], wie in BILD 2. beispielhaft dargestellt. Ziel des Algorithmus ist es, die vorliegenden Daten mit möglichst wenig Fragen zu strukturieren und daraus zu lernen. Ein einzelner Entscheidungsbaum ist jedoch nicht gut geeignet, um Vorhersagen zu treffen. Zum Beispiel leidet ein einzelner Entscheidungsbaum am Problem der Überanpassung von Daten. Deshalb wird hier der Random-Forest-Algorithmus verwendet, welcher aus einer Vielzahl von Entscheidungsbäumen besteht und dadurch das Überanpassen der Daten verhindert und aufgrund seiner positiven Eigenschaften in vielen Anwendungen zum Einsatz kommt [9, 11].

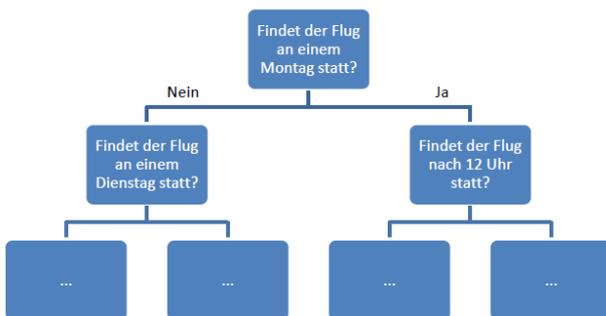


BILD 2. Beispiel für ein Entscheidungsbaumverfahren

4. DATENAUSWAHL

Die Datenauswahl und die damit verbundene Merkmalsauswahl ist ein wichtiger Schritt, bevor das Machine-Learning Modell trainiert werden kann [6]. Die Daten müssen die Realität, welche vorhergesagt werden soll,

möglichst genau abbilden. Im Nachfolgenden werden die einzelnen Datenquellen beschrieben, welche zur Verfügung stehen und welche Merkmale jeweils von ihnen bezogen werden. Zusätzlich werden aus den vorhandenen Daten neue Merkmale generiert. Merkmale, welche nicht numerisch vorliegen, werden in Dummy-Variablen überführt, um eine Auswertung zu ermöglichen. Dummy-Variablen sind Hilfsmerkmale, um nicht binäre Merkmale in binäre Merkmale zu überführen [6]. Alle Merkmale eines Fluges werden schlussendlich kombiniert und in einem Datensatz für einen vorher definierten Zeitraum abgespeichert. Dieser Datensatz wird zum Trainieren des Modells verwendet. In BILD 3. ist eine grafische Übersicht über die verwendeten Daten dargestellt.

Die Flugoperationsdaten, sowie Wetterdaten liegen auf einem firmeninternen SQL-Server vor. Die EUROCONTROL Daten werden täglich von Internetservern der EUROCONTROL abgerufen und ebenfalls gespeichert. Sämtliche Daten werden von dem Vorhersagemodell automatisiert abgerufen und eingelesen. Um zu gewährleisten, dass keine fehlerhaften oder falschen Daten verarbeitet werden, werden die Daten automatisiert erstellt und bereinigt. Dafür werden eigens programmierte Algorithmen [13] angewendet. Diese stellen auch ein stringentes Datenformat sicher und ermöglichen somit eine automatisierte Erstellung und Bereinigung der Daten.

Die Auswahl der Daten wurde mit Hilfe statistischer Verfahren, wie beispielsweise t-test [14, 15] oder ANOVA [16] durchgeführt. Außerdem wurde auf Expertenwissen zurückgegriffen, um gezielt Merkmale zu generieren wie beispielsweise das Merkmal „Stunde des Tages * Städtepaar * Wochentag“ oder die Vorauswahl für die fünf Wettermerkmale, welche in TAB 2. mit aufgeführt werden.

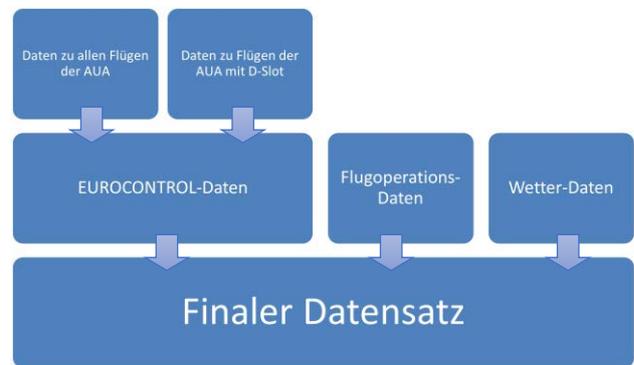


BILD 3. Übersicht über die Erstellung des finalen Datensatzes

4.1. Daten der EUROCONTROL

Die Datensätze der EUROCONTROL beinhalten Informationen zu jedem einzelnen Flug der AUA mit Verspätungen und ATFM-Slot- Informationen.

Aus den EUROCONTROL Datensätzen lassen sich zudem temporäre Merkmale generieren. Ein Beispiel hierfür ist „Stunden des Tages“, welches in BILD 4. [13] dargestellt wird. Eine Korrelation zwischen der Vergabe eines ATFM-Slots und der Uhrzeit ist deutlich erkennbar. Dies wird zusätzlich mit statistischen Verfahren belegt [8].

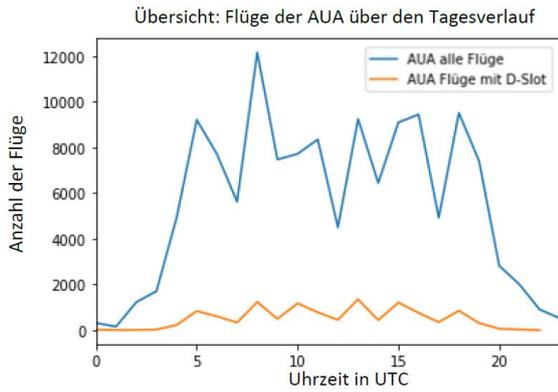


BILD 4. Korrelation zwischen dem Merkmal „Stunde des Tages“ und der Vergabe von D-Slots [13]

4.2. Wetterdaten der Firma UBIMET

Seit November 2019 stellt die Firma UBIMET historische Wetterdaten sowie Vorhersagen zur Verfügung. Aus den vorhandenen Datensätzen werden die 5 Merkmale Nebel, Gewitter, Sichtweite, Windgeschwindigkeit und Böen ausgewählt – siehe TAB 2. Ab einem vorher definierten Schwellenwert wird für jedes dieser Merkmale ein binär kodiertes Merkmal erstellt. Der Schwellenwert wurde anhand der Bestimmung eines t-Testes und dem dazugehörigen p-Wert ermittelt [13, 14]. Aus dem Datenpunkt Windgeschwindigkeit wird beispielsweise das Merkmal „Starker Wind“ ab einer Geschwindigkeit von 16 m/s generiert. Dies geschieht jeweils für den Abflugflughafen (ADEP) und den Ankunftsflughafen (ADES).

4.3. Flugoperationsdaten der Austrian Airlines

Anhand der Flugoperationsdaten werden den einzelnen Flügen ihre Flugnummern zugeordnet, um eine einfachere Auswertung für die Flugplanung zu gewährleisten. Zudem stehen Informationen zur geplanten Ankunftszeit zur Verfügung.

4.4. Auswahl der finalen Datensätze

Hier werden exemplarisch zwei Datensätze betrachtet. Ein Sommerdatensatz mit Daten zwischen dem 1.4.2019 und dem 1.10.2019 und ein Winterdatensatz zwischen dem 1.11.2019 und dem 1.3.2020.

Der Sommerdatensatz umfasst den Sommerflugplan des Jahres 2019. Es sind keine Wetterdaten vorhanden. 12% der Flüge der AUA hatten einen ATFM-Slot.

Der Winterdatensatz ist so gewählt, dass alle Daten Wetterdaten beinhalten und zum anderen die Auswirkung der Corona Pandemie noch nicht zu verzerrten Ergebnissen führt.

In TAB 2. sind alle berücksichtigten Merkmale aufgeführt, welche dem Algorithmus zum Trainieren zur Verfügung gestellt werden. In einer weiteren Spalte ist aufgeführt, wie die einzelnen Merkmale kodiert sind.

Merkmal	Kodierung
Städtepaar	Dummy
Stunde des Tages	0...23
Wochentag	Dummy
Woche des Jahres	0...52
Tag des Jahres	1...365 bzw. 366
Monat	1...12
Flugnummer	Dummy
Stunde des Tages * Städtepaar	Dummy
Stunde des Tages * Städtepaar * Wochentag	Dummy
Starker Nebel ADES	[0,1]
Starker Nebel ADEP	[0,1]
Starke Gewitter ADES	[0,1]
Starke Gewitter ADEP	[0,1]
Schlechte Sicht ADES	[0,1]
Schlechte Sicht ADEP	[0,1]
Starker Wind ADES	[0,1]
Starker Wind ADEP	[0,1]
Starke Böen ADES	[0,1]
Starke Böen ADEP	[0,1]

TAB 2. Verwendete Merkmale

5. ERSTELLUNG DES D-SLOTVORHERSAGE-MODELLS

Den Algorithmen wird ein Teil des generierten Datensatzes zum Trainieren und Validieren während der Trainingsphase zur Verfügung gestellt – der Trainingsdatensatz. Der restliche Datensatz – der Testdatensatz - wird zur Überprüfung des Algorithmus benötigt. Die Aufteilung des Datensatzes in Trainings- und Testdatensatz erfolgt in einem Verhältnis 4:1 [6]. Im Nachfolgenden werden nur die Ergebnisse der Algorithmen auf dem Testdatensatz erläutert, da diese die Vorhersage von zukünftigen, unbekanntem Flügen am ehesten simulieren.

5.1. Logistische Regression

Der Algorithmus der Logistischen Regression wird anhand verschiedener Metriken optimiert - Genauigkeit, f1, Sensitivität und Relevanz [13]. Die Metrik „f1“ liefert hierbei die für die Anwendung besten Ergebnisse. Die dazugehörigen Kalibrierungskurven sind im Anhang in BILD 10. für den Sommer-Testdatensatz und in BILD 11. für den Winter-Testdatensatz dargestellt und werden in den nachfolgenden Kapiteln näher erläutert. Zudem werden verschiedene Hyperparameter getestet und analysiert. Dazu gehören die verschiedenen Regularisierungen (L1/L2) und ihre Gewichtung (C).

Wie in Kapitel 3 beschrieben, ist ein Vorteil der Logistischen Regression die anschauliche Interpretation der ermittelten Koeffizienten. In der nachfolgenden Tabelle TAB 3. ist ein Ausschnitt der Koeffizienten aufgelistet. Die Koeffizienten befinden sich in einem Wertebereich zwischen +1.996 und -1,967 sowohl für den Sommer- als auch den Winter-Testdatensatz. Anhand dieser Koeffizienten wer-

den die einzelnen Merkmale gewichtet. Ein Koeffizient von -1.967 des Städtepaars von Wien nach Graz spricht für eine starke Tendenz gegen einen D-Slot auf dieser Strecke. Somit können aus den einzelnen Koeffizienten Rückschlüsse gezogen werden.

Koeffizient	Gewichtung
Flug von Wien nach Graz	-1.967
Flug von Wien nach Klagenfurt	-1.655
Flug von Wien nach Salzburg	-1.555
Flugnummer 2682	-1.480
...	...
Flugnummer 9116	1.918
Flugnummer 801	1.884
Flugnummer 9037	1.905
Flugnummer 3067	1.996

TAB 3. Exemplarische Koeffizienten der Logistischen Regression

5.1.1. Sommerdatensatz

Die Ergebnisse des Sommer-Testdatensatzes werden im nachfolgenden erläutert. BILD 10. im Anhang zeigt die Kalibrierungskurve der Vorhersage. Die berechnete Wahrscheinlichkeit (blaue Linie) liegt bis zu einem Wert von 33% auf der Linie der perfekten Kalibrierung (gestrichelte Linie). Danach verläuft sie mit leichten Abweichungen unter der perfekten Kalibrierung. Das heißt, der Algorithmus überschätzt die Wahrscheinlichkeit für das Eintreten eines D-Slots. Positiv hervorzuheben ist, neben der fast perfekten Kalibrierung, dass auch in den Bereichen von hohen Wahrscheinlichkeiten noch eine große Anzahl an Datenpunkten zugeordnet wird (38 Flüge bei 85% und 24 bei 96%).

Wird die Konfusionsmatrix (BILD 5) zur Auswertung hinzugezogen, lässt sich erkennen, dass nahezu alle „Richtig Negative“ Ereignisse vorhergesagt werden – 98%. Für die Vorhersage der D-Slots ist eine möglichst hohe Prozentzahl von „Richtig Positiven“ erstrebenswert. Dieser Wert liegt bei 18%, das heißt fast jeder fünfte der wahren D-Slots wird richtig vorhergesagt. Die Anzahl der „Falsch Positiven“ im Verhältnis zu den „Richtig Positiven“ liegt bei 56%.

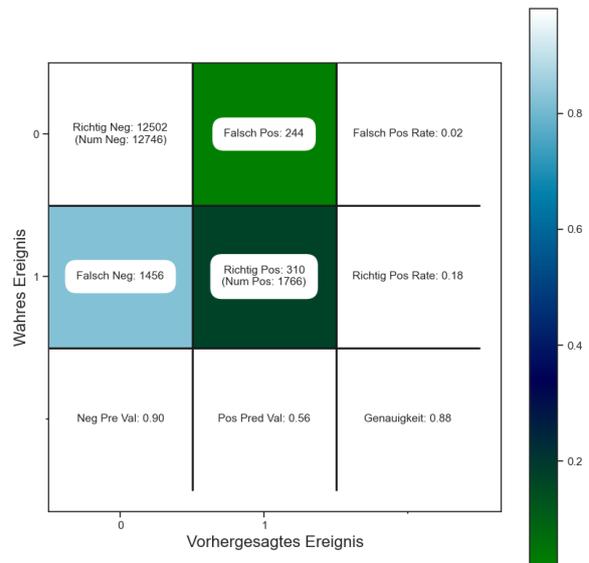


BILD 5. Konfusionsmatrix der Logistischen Regression (Sommer-Testdatensatz)

5.1.2. Winterdaten

Der Winterdatensatz weist im Vergleich zum Sommerdatensatz deutlich weniger Datenpunkte und D-Slots auf. Die niedrigere Anzahl an Datenpunkten lässt sich durch die geringere Anzahl an betrachteten Monaten erklären. Zudem sind in den Wintermonaten im Flugplan, sowohl der AUA als auch anderer Fluggesellschaften, weniger Flugbewegungen, welches in einer generell niedrigeren Auslastung der Kapazitäten in den Wintermonaten resultiert.

Die Kalibrierungskurve in BILD 11. zeigt einen nahezu perfekten Verlauf auf der Linie der perfekten Kalibrierung mit Abweichungen von 4% bis zu einer Wahrscheinlichkeit von 60%. Besonders auffällig ist der unterhalb der perfekten Kalibrierung verlaufende Teil ab 60%. Dies spricht für eine Überschätzung der Wahrscheinlichkeiten. Das Maximum erreicht die Abweichung von 23% bei einer tatsächlichen Häufigkeit von 75%. Dies lässt sich durch zu wenige Datenpunkte erklären. Dem Bereich zwischen 80% und 100% werden lediglich noch 34 Datenpunkte zugeordnet. Dies führt zu Ungenauigkeiten und der beobachteten Schwankung.

Die Konfusionsmatrix des Winterdatensatzes ist in BILD 6. dargestellt. Aus dieser lässt sich ablesen, dass 12% aller D-Slots erkannt werden. Das Verhältnis zwischen falsch und richtig erkannten D-Slots beträgt 45%. Die Genauigkeit mit 97% ist sehr hoch.

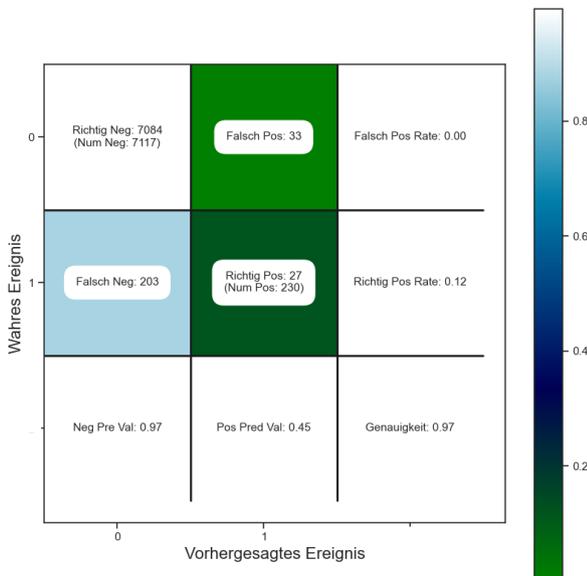


BILD 6. Konfusionsmatrix der Logistischen Regression (Winter-Testdatensatz)

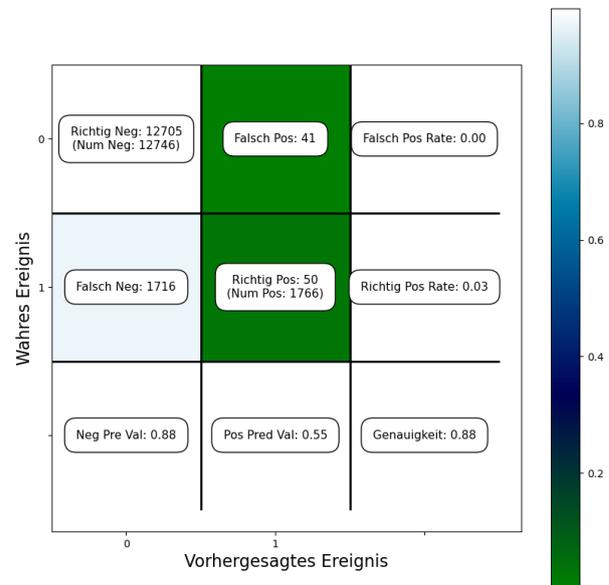


BILD 7. Konfusionsmatrix des Random Forests (Sommer-Testdatensatz)

5.2. Random Forest

Die Vorhersage des Random Forest Algorithmus für den Winterdatensatz ergibt keine aussagekräftigen Ergebnisse auf Grund der geringen Anzahl an vorliegenden Daten. Es werden keine D-Slots vorhergesagt. Deswegen wird auf die Auswertung an dieser Stelle verzichtet. Die Ergebnisse des Sommerdatensatzes ergeben die Konfusionsmatrix in BILD 7., sowie die Kalibrierungskurve in BILD 12. Aus der Kalibrierungskurve lässt sich eine nahezu perfekte Kalibrierung bis 50% ablesen. Danach kommt es zu einer Überanpassung, die tatsächliche Häufigkeit ist geringer als die vorhergesagte. Es ergeben sich Abweichungen bis zu 12%. Zudem werden nur Wahrscheinlichkeiten bis zu einer Wahrscheinlichkeit von 75% vorhergesagt. Es werden in den letzten beiden Wahrscheinlichkeitsbereichen insgesamt 32 Flüge zugeordnet - 19 Flüge bei 65% und 13 Flüge bei 75%.

Anhand der Konfusionsmatrix wird deutlich, dass 3% aller D-Slots richtig erkannt werden. Außerdem ist das Verhältnis von „Richtig Positiven“ zu „Falsch Positiven“ 55%.

6. DISKUSSION

Bedingt durch den Einbruch der Flugbewegungen in Europa seit März 2020 ist ein Trainieren des Algorithmus mit Wetterdaten der Sommermonate nicht möglich. Zudem kann der Algorithmus mit aktuellen Flugbewegungen nicht validiert werden. Jedoch ist es möglich, die berechneten Wahrscheinlichkeiten mit den Testdatensätzen abzugleichen.

Es wird ersichtlich, dass die Vorhersagen der Wahrscheinlichkeiten für den Sommer-Testdatensatz 2019 mit der Logistischen Regression sehr gute Ergebnisse liefern. Die Kalibrierungskurve verläuft entlang der Linie der perfekten Kalibrierung mit vereinzelt Abweichungen. Hervorzuheben ist, dass Wahrscheinlichkeiten über den ganzen Wertebereich von 0% bis 100 % berechnet werden. Bei der Kalibrierungskurve der Logistischen Regression für den Winterdatensatz ist vor allem der Bereich ab einer Wahrscheinlichkeit von 70 % noch sehr ungenau und es werden keine zufriedenstellenden Ergebnisse erzielt. Dies liegt vor allem an einer zu geringen Anzahl an Datenpunkten in den entsprechenden Wahrscheinlichkeitsbereichen. Durch Ergänzung weiterer Merkmale oder Datenpunkte sollte eine Verbesserung zu erzielen sein. Die Ergebnisse des Random Forest Algorithmus sind ebenfalls für den Sommerdatensatz besser als für den Winterdatensatz aus denselben Gründen wie bei der Logistischen Regression. Durch die höhere Komplexität des Algorithmus werden jedoch insgesamt schlechtere Ergebnisse erzielt. Durch weitere Verfeinerungen und Ergänzungen der Merkmale und Algorithmen, sowie die Wahl von anderen Entscheidungsbaumverfahren ist auch hier mit einer Verbesserung der Ergebnisse zu rechnen.

Wahrscheinlichkeitsbereich	Anzahl der gesamten Flüge	Anzahl der vorhergesagten D-Slots	Anzahl der tatsächlichen D-Slots
0-0.09	9046	284	380
0.1-0.19	2485	356	366
0.2-0.29	1322	324	331
0.3-0.39	716	247	238
0.4-0.49	389	173	141
0.5-0.59	253	138	124
0.6-0.69	138	90	79
0.7-0.79	101	76	64
0.8-0.89	38	32	27
0.9-0.99	24	24	16
	Summe: 14512	Summe: 1745	Summe: 1766

TAB 4. Auswertung der Vorhersagen verglichen mit den wahren D-Slots [13]

7. OPERATIONALISIERUNG DER ERGEBNISSE

Die Ergebnisse der Logistischen Regression für den Sommer-Testdatensatz – optimiert nach f1-Maß - sind bereits genau genug, um eine Operationalisierung zu realisieren, welche im nachfolgenden genauer erläutert wird.

Die Multiplikation der berechneten Wahrscheinlichkeiten mit der Anzahl der jeweiligen zugeordneten Flüge ergibt die Anzahl der vorhergesagten Flüge mit einem D-Slot (TAB 4. – Spalte 3). Wird ebenso mit den tatsächlichen Wahrscheinlichkeiten verfahren, ergibt sich die Anzahl der Flüge mit einem D-Slot (TAB 4. -Spalte 4). Zur besseren Übersicht ist in TAB 4. der jeweilige Wahrscheinlichkeitsbereich sowie die Anzahl aller zugeordneten Flüge aufgeführt (TAB 4. – Spalte 2).

Werden die Flüge in dem Wahrscheinlichkeitsbereich von 90%- 99% betrachtet, ergibt sich, dass für alle 24 Flüge ein D-Slot vorhergesagt wird, jedoch nur 16 Flüge tatsächlich einen D-Slot haben. Daraus folgt, dass die Flugplanung 8 Flüge umplanen würde, die gar keinen D-Slot bekommen würden. Dies ergibt eine Fehlerquote von 33%, welche vertretbar wäre. Werden alle Flüge bis zu einer Wahrscheinlichkeit bis 50% betrachtet, ergibt sich eine Quote von vergebens umgeplanten Flügen von 35% (359 zu 554 Flügen). Für eine erste Operationalisierung ist das eine akzeptable Fehlerquote. Niedrige Wahrscheinlichkeiten unter 50% haben eine zu hohe Fehlerquote zur Folge und werden deswegen nicht mehr der Flugplanung zur Betrachtung vorgelegt – eine 56% Fehlerquote bei Wahrscheinlichkeiten von 40%-49% und eine 66% Fehlerquote bei Wahrscheinlichkeiten von 30%-39%.

Ein Ansatz wäre, dass die Flugplanung die Flüge mit der höchsten Wahrscheinlichkeit für einen D-Slot zuerst umplant und sich dann sukzessiv zu niedrigeren Wahrscheinlichkeiten vorarbeitet.

Ein erstes Konzept für eine mögliche grafische Oberfläche wird im BILD 9. dargestellt. Alle geplanten Flüge, identifizierbar anhand ihrer Flugnummer für die nächsten 24 h, werden nach Stunden sortiert aufgelistet. Falls die Wahrscheinlichkeit eines ATFM-Slots einen vorher definierten Schwellenwert erreicht hat, ändert sich die Farbe in Orange. Die Flugplanung hätte dann die Möglichkeit, die Flugnummer auszuwählen und somit ein Fenster mit weiteren Informationen zu öffnen. Diese würden unter anderem die berechnete Wahrscheinlichkeit für einen D-Slot beinhalten, sowie den Luftraum auf der Strecke, der historisch betrachtet am häufigsten einen D-Slot ausgelöst hat. Dadurch soll der Flugplanung eine gezielte Umplanung der Flugroute ermöglicht werden. Weitere Daten könnten ergänzt werden.

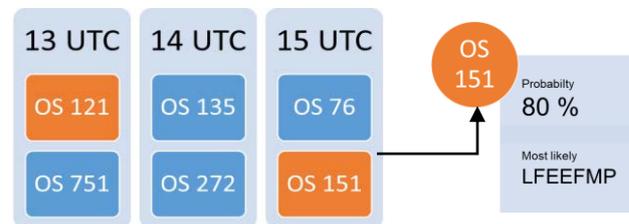


BILD 9. Konzept für eine Benutzeroberfläche

8. FAZIT UND AUSBLICK

Die hier vorgestellten Ergebnisse können bereits operationalisiert werden und die Flugplanung unterstützen. Zudem sind sie eine gute Basis für Erweiterungen des Modells und für die Ausarbeitung eines vollständigen Programms für die Endnutzer. Im weiteren Verlauf ist die Umsetzung einer grafischen Oberfläche für die Flugplanung geplant, damit die gewonnenen Erkenntnisse in der Praxis Anwendung finden können. Als weitere Unterstützung soll das Tool auch Auskunft über die Lufträume, die am häufigsten einen ATFM-Slot entlang der jeweiligen Route ausgelöst haben, geben.

Durch die Ergänzung weiterer aussagekräftiger Merkmale wäre es möglich, das Modell weiter zu verfeinern und zu verbessern. Mögliche Merkmale wären Informationen über geplante Landebahnschließungen oder Baustellen auf dem Vorfeld oder Merkmale, die das Wetter im Flugverlauf beschreiben. Außerdem ist es denkbar, das Modell dahingehend zu erweitern, dass vorhergesagt wird, wie viele Minuten der voraussichtliche D-Slot nach der ursprünglich geplanten Abflugzeit liegt.

LITERATURVERZEICHNIS

- [1] EUROCONTROL, *ATFM Regulation: a power for good: Understanding how it works*. [Online]. Available: <https://www.eurocontrol.int/sites/default/files/2020-01/eurocontrol-think-paper-2-atfm-regulation.pdf> (accessed: Jul. 29 2021).
- [2] EUROCONTROL, *EUROCONTROL Data*. [Online]. Available: <https://ansperformance.eu/data/> (accessed: Jul. 29 2021).
- [3] A. Cook, G. Turner, and A. Lawes, "The hidden cost of airline unpunctuality," *Journal of transport economics and policy*, vol. 46, no. 2, pp. 157–173, 2012.
- [4] EUROCONTROL, *What COVID-19 did to European Aviation in 2020, and Outlook 2021*. [Online]. Available: <https://www.eurocontrol.int/sites/default/files/2021-02/eurocontrol-think-paper-8-impact-of-covid-19-on-european-aviation-in-2020-and-outlook-2021.pdf> (accessed: Jul. 29 2021).
- [5] H. N. Kariyawasam, "Cost Economies for an Airline: An Analysis of Airlines' Operating Costs," *International Journal of Management Excellence*, vol. 12, no. 3, p. 1856, 2019, doi: 10.17722/ijme.v12i3.562.
- [6] A. C. Müller and S. Guido, *Einführung in Machine Learning mit Python: Praxiswissen Data Science*, 1st ed. Heidelberg: O'Reilly, 2017. [Online]. Available: <http://proquestcombo.safaribooksonline.com/9781492064619>
- [7] I. Mohammadian, B. Abbasi, A. Abareshi, and M. Goh, "Antecedents of flight delays in the Australian domestic aviation market," *Transportation Research Interdisciplinary Perspectives*, vol. 1, no. 6, p. 100007, 2019, doi: 10.1016/j.trip.2019.100007.
- [8] M. Efthymiou, E. T. Njoya, P. L. Lo, A. Papatheodorou, and D. Randall, "The Impact of Delays on Customers' Satisfaction: An Empirical Analysis of the British Airways On-Time Performance at Heathrow Airport," *J.Aerosp. Technol. Manag.*, vol. 11, 2019, doi: 10.5028/jatm.v11.977.
- [9] H. Lee *et al.*, "Critical Parameter Identification for Safety Events in Commercial Aviation Using Machine Learning," *Aerospace*, vol. 7, no. 6, p. 73, 2020, doi: 10.3390/aerospace7060073.
- [10] R. Nigam and K. Govinda, "Cloud based flight delay prediction using logistic regression," in *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, Palladam, 2017, pp. 662–667.
- [11] M. Zoutendijk and M. Mitici, "Probabilistic Flight Delay Predictions Using Machine Learning and Applications to the Flight-to-Gate Assignment Problem," *Aerospace*, vol. 8, no. 6, p. 152, 2021, doi: 10.3390/aerospace8060152.
- [12] *2017 International Conference on Intelligent Sustainable Systems (ICISS)*: IEEE, 2017.
- [13] Friederike Ebus, "Entwicklung und Validierung eines ATFM-Slot-Vorhersagemodells für die Flugplanung," TU Braunschweig, Braunschweig, 2020.
- [14] I. Frost, *Statistische Testverfahren, Signifikanz und p-Werte: Allgemeine Prinzipien verstehen und Ergebnisse angemessen interpretieren*. Wiesbaden: Springer VS, 2017. [Online]. Available: <http://dx.doi.org/10.1007/978-3-658-16258-0>
- [15] G. Christoph and H. Hackel, *Starthilfe Stochastik: Studium*, 1st ed. Stuttgart: Vieweg + Teubner, 2010.
- [16] K. Backhaus, B. Erichson, W. Plinke, and R. Weiber, *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, 14th ed. Berlin, Heidelberg: Springer Gabler, 2016. [Online]. Available: <http://www.springer.com/>

ANHANG

Logistische Regression – Sommer-Testdatensatz

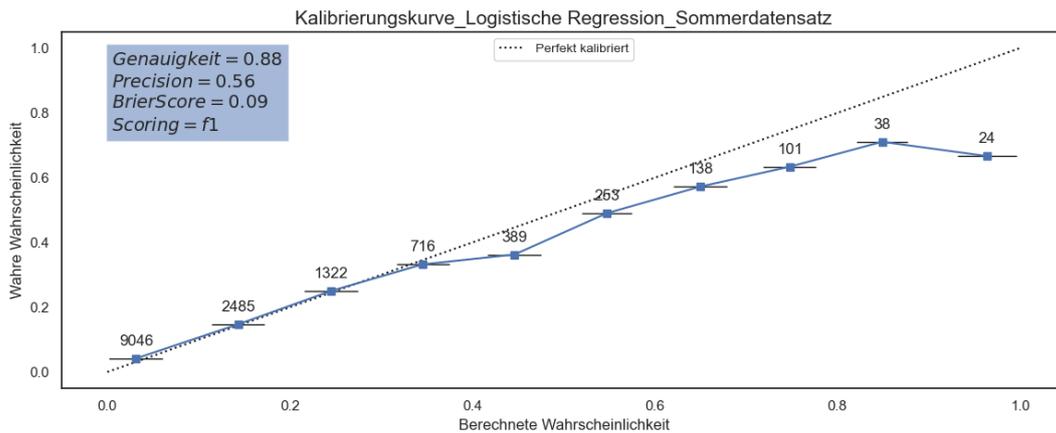


BILD 10. Kalibrierungskurve der Logistischen Regression optimiert nach f1 (Sommer-Testdatensatz) und Metriken (a) Genauigkeit (b) Precision (c) Brier Score. Als Zielmetrik wurde hier f1 -Maß verwendet. Die horizontalen Linien des Plots geben die Standardabweichung der berechneten Wahrscheinlichkeit an.

Logistische Regression – Winter-Testdatensatz

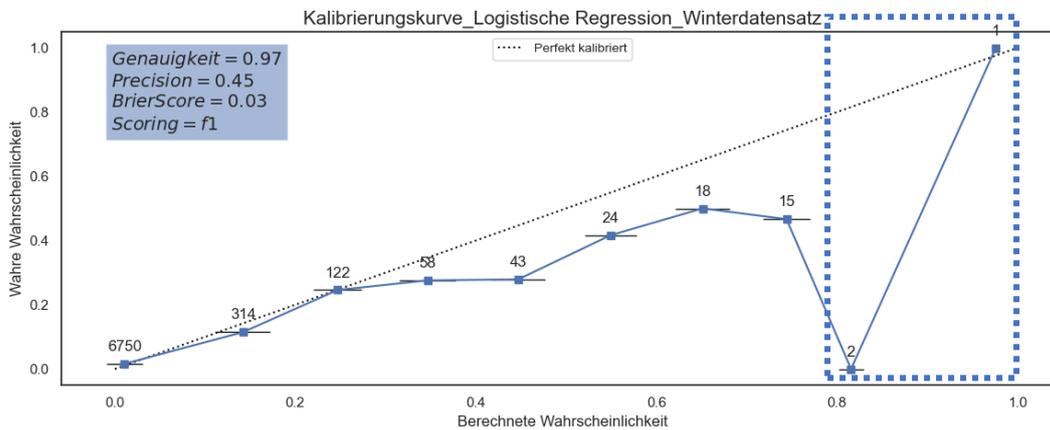


BILD 11. Kalibrierungskurve der Logistischen Regression optimiert nach f1 (Winter-Testdatensatz) und Metriken (a) Genauigkeit, (b) Precision (c) Brier Score. Als Zielmetrik wurde hier das f1-Maß verwendet. Der blau markierte Bereich zeigt den Bereich für berechnete Wahrscheinlichkeiten größer als 80% für den im Wintermodell nur drei Datenpunkte vorliegen. Die horizontalen Linien des Plots geben die Standardabweichung der berechneten Wahrscheinlichkeit an.

Random Forest – Sommer-Testdatensatz

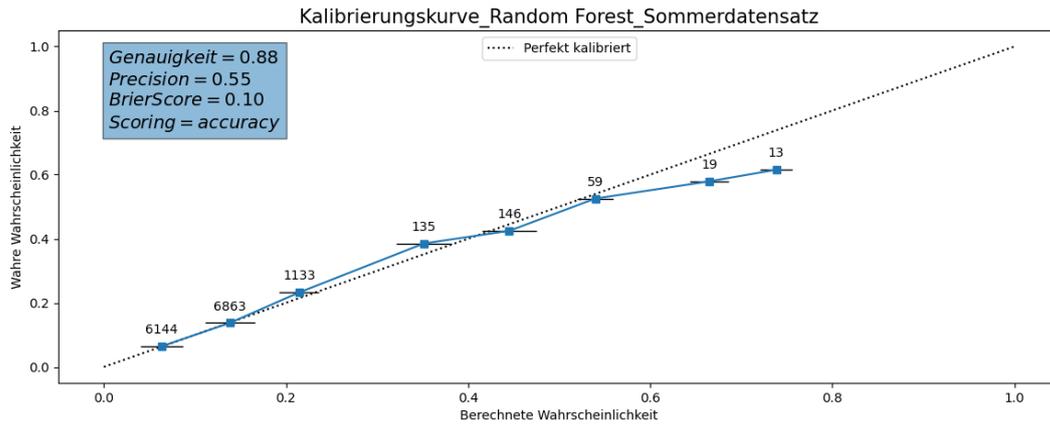


BILD 12. Kalibrierungskurve des Random Forest (Sommer-Testdatensatz) und Metriken (a) Genauigkeit, (b) Precision, (c) Brier Score. Als Zielmetrik wurde hier die Genauigkeit verwendet, da diese - obwohl nur Wahrscheinlichkeiten bis 75% vorhergesagt werden – die aussagekräftigsten Ergebnisse durch Vermeidung von Überanpassungen liefert. Die horizontalen Linien des Plots geben die Standardabweichung der berechneten Wahrscheinlichkeit an.