# Unternehmensbeschreibung

Im Fokus dieser Fallstudie steht das fiktive Unternehmen "StayAwhile Hotels". StayAwhile Hotels ist eine international agierende Hotelkette mit Hauptsitz in Bonn, Deutschland. Mit über 500 Hotels in 50 Ländern bietet StayAwhile Hotels eine Vielzahl von Dienstleistungen an, darunter Unterkünfte, Catering, Veranstaltungsplanung und Reisedienstleistungen.

StayAwhile Hotels hat sich einen Namen gemacht durch seinen Fokus auf Kundenzufriedenheit und Qualitätsservice. Das Unternehmen hat eine starke Marke aufgebaut, die auf den Prinzipien von Gastfreundschaft, Integrität und Exzellenz basiert. Trotz ihres Erfolgs in der Branche sieht sich StayAwhile Hotels mit einer Reihe von Herausforderungen konfrontiert, insbesondere im Zusammenhang mit Stornierungen.

Im Rahmen ihrer digitalen Transformation hat StayAwhile Hotels erkannt, dass Technologie und Datenanalyse Schlüsselkomponenten zur Bewältigung dieser Herausforderungen sind. Das Unternehmen hat daher in die Entwicklung von Machine-Learning-Modellen investiert, um seine Betriebsabläufe zu optimieren und eine datengesteuerte Entscheidungsfindung zu ermöglichen.

# Problem und Fragestellung

Stornierungen sind in der Hotelbranche ein weit verbreitetes Problem, das erhebliche Auswirkungen auf den Betrieb und die Rentabilität eines Unternehmens haben kann. Für StayAwhile Hotels ist das Problem der Stornierungen besonders akut. Unvorhergesehene Stornierungen führen zu unbesetzten Zimmern, die wiederum zu verlorenen Einnahmen führen. Darüber hinaus können sie die Planung und Ressourcenallokation des Unternehmens erheblich beeinträchtigen.

Laut einer Studie von D-EDGE Hospitality Solutions aus dem Jahr 2019 lag die durchschnittliche Stornierungsrate in der europäischen Hotellerie bei etwa 40%. Diese hohe Rate unterstreicht die Dringlichkeit des Problems und die Notwendigkeit, effektive Lösungen zu finden.

Die zentrale Fragestellung dieser Fallstudie ist daher: Wie kann StayAwhile Hotels mithilfe von Machine Learning die Wahrscheinlichkeit von Stornierungen vorhersagen und so ihre Betriebsplanung und Ressourcenallokation verbessern?

Um diese Frage zu beantworten, wurde ein Machine-Learning-Modell entwickelt, das auf historischen Buchungsdaten basiert und in der Lage ist, die Wahrscheinlichkeit zukünftiger Stornierungen vorherzusagen. Durch die Verbesserung der Vorhersagegenauigkeit von Stornierungen kann StayAwhile Hotels seine Betriebsplanung optimieren, seine Ressourcen effizienter zuweisen und letztendlich seine Rentabilität steigern.

# Einordnung in den Bereich Machine Learning

Machine Learning (ML) ist ein Teilbereich der künstlichen Intelligenz, der sich darauf konzentriert, Computern das "Lernen" aus Daten zu ermöglichen, um Muster zu erkennen und Vorhersagen zu treffen. Es hat sich als ein mächtiges Werkzeug in vielen Branchen erwiesen, einschließlich der Hotellerie.

In der Hotellerie kann ML dazu beitragen, eine Vielzahl von Herausforderungen zu bewältigen, von der Verbesserung der Kundenzufriedenheit bis hin zur Optimierung der Betriebsabläufe. Laut einer Studie von McKinsey aus dem Jahr 2020 nutzen bereits 47% der Unternehmen in der Reise- und Logistikbranche ML in irgendeiner Form, und weitere 21% planen, es in den nächsten drei Jahren einzuführen.

In dieser Fallstudie wird ML verwendet, um das Problem der Stornierungen bei StayAwhile Hotels anzugehen. Das entwickelte Modell ist ein Beispiel für überwachtes Lernen, eine Art von ML, bei der ein Modell aus einem Satz von Eingabe- und Ausgabedaten "lernt". In diesem Fall sind die Eingabedaten die historischen Buchungsdaten und die Ausgabedaten sind, ob eine Buchung storniert wurde oder nicht.

Das Modell verwendet verschiedene Merkmale aus den Buchungsdaten, wie z.B. den Buchungszeitpunkt, den Aufenthaltszeitraum und den Buchungstyp, um Muster zu erkennen und Vorhersagen über zukünftige Stornierungen zu treffen. Durch die Verbesserung der Vorhersagegenauigkeit von Stornierungen kann StayAwhile Hotels seine Betriebsplanung optimieren, seine Ressourcen effizienter zuweisen und letztendlich seine Rentabilität steigern.

Zum Angehen des Problems wird sich auf das Framework des CRISP- DM Models bezogen. Hierbei werden Datenanalyse Probleme strukturiert angegangen und können einer Logik folgen. Um so das Businessproblem zu lösen. Das Business Understanding wurde in den ersten beiden Kapiteln behandelt. Zum Data Understanding wurde Jupyter Notebooks hinzugezogen, aus welchem auch Code Snippets mit eingebracht werden.

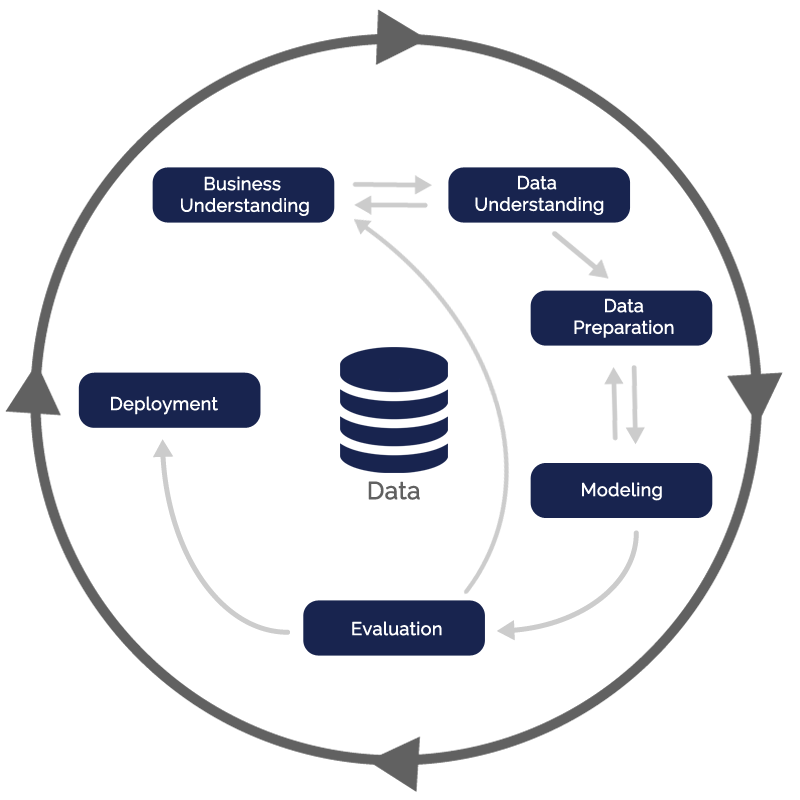


Abbildung 1: Crisp DM Model (https://www.otaris.de/beratung-data-science/)

# Data Understanding

# Data Preparation

# Modeling

# Evaluation

# Deployement

# Datenbasierte Services

# Relevante Effekte für StayAwhile

# Datensatz und ML- Modell

Die Daten, welche zur Entwicklung des Modells benutzt, werden kommen aus dem Dataset [Hotel Reservations Dataset](https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset), auf Kaggle. Das Dataset ist aus 19 Features und 36.000 Datensätzen aufgebaut. Zu jeder stornierten Buchung kommen 2 nicht stornierte Buchungen. Also ist der Datensatz relativ ausbalanciert.  
  
**Feature 1:** Booking\_ID  
dient nur der Zuordnung zu einem Datensatz und ist nicht relevant für das Modell

**Feature 2-18:** Nominal: Booking\_ID, type\_of\_meal\_plan, required\_car\_parking\_space, room\_type\_reserved, market\_segment\_type, repeated\_guest, booking\_status

Kardinal: no\_of\_adults, no\_of\_children, no\_of\_weekend\_nights, no\_of\_week\_nights, lead\_time, no\_of\_previous\_cancellations, no\_of\_previous\_bookings\_not\_canceled, avg\_price\_per\_room, no\_of\_special\_requests

Ordinal: arrival\_year, arrival\_month, arrival\_date

**Feature 19:** Booking\_status – dieses Feature ist die Zielvariable und wird in binär angegeben.

# Methodik

Für die Vorhersage von Stornierungen im Hotelgewerbe bieten sich verschiedene Algorithmen an, darunter Entscheidungsbäume, Random Forest, XGBoost und K-Nearest Neighbors (KNN).

Entscheidungsbäume ermöglichen die Erstellung von Entscheidungsregeln basierend auf den Merkmalen der Buchungsdaten. Durch die Kombination mehrerer Entscheidungsbäume im Random Forest-Algorithmus können robuste und genaue Vorhersagen erzielt werden. Random Forest berücksichtigt die Vorhersagen aller Bäume und aggregiert sie zu einem kohärenten Ergebnis.

XGBoost ist eine Weiterentwicklung des Gradient Boosting-Algorithmus, der speziell auf Effizienz und Genauigkeit ausgelegt ist. Es verwendet Entscheidungsbäume und optimiert die Modellvorhersagen durch das Minimieren einer Kostenfunktion. XGBoost bietet erweiterte Funktionen wie Regularisierung und paralleles Training, um präzise Vorhersagen zu erzielen.

K-Nearest Neighbors (KNN) basiert auf der Idee, dass ähnliche Beispiele tendenziell die gleiche Klassenzugehörigkeit haben. KNN sucht nach den k nächsten Nachbarn eines Datenpunkts und bestimmt die Mehrheitsklasse dieser Nachbarn als Vorhersage. KNN ist besonders nützlich, um Muster und Ähnlichkeiten in den Buchungsdaten zu erkennen und Vorhersagen zu treffen.

# Ausblick

Die erfolgreiche Implementierung eines prädiktiven Modells zur Vorhersage von Stornierungen im Hotelgewerbe bietet vielfältige Auswirkungen und Möglichkeiten für das Unternehmen. Durch die präzise Vorhersage von Stornierungen kann das Unternehmen seine Betriebsplanung optimieren, die Kapazitätsauslastung verbessern und finanzielle Verluste durch leere Zimmer minimieren.

# Quellen

D-EDGE Hospitality Solutions, "European Hotel Distribution Study: The Rise of Online Travel Agents", Zugriff am 2. Juli 2023, <https://www.d-edge.com/download-european-hotel-distribution-study-the-rise-of-online-travel-agents/>

McKinsey, "The state of AI in 2020", Zugriff am 2. Juli 2023, https://www.mckinsey.com/featured-insights/artificial-intelligence/the-state-of-ai-in-2020