

Entwicklung eines Maschine Learning Modells ohne die Verwendung von Hotel spezifischen Vergangenheitsdaten

William Mendat

MASTERARBEIT

zur Erlangung des akademischen Grades Master of Science (M.Sc.)

Studiengang Informatik Master

Fakultät Elektrotechnik, Medizintechnik und Informatik
Hochschule für Technik, Wirtschaft und Medien Offenburg

30.03.2024

Durchgeführt bei der Firma happyhotel

Betreuer

Prof. Dr.-Ing. Janis Keuper, Hochschule Offenburg
Prof. Dr. rer. nat. Klaus Dorer, Hochschule Offenburg

Mendat, William:

Entwicklung eines Maschine Learning Modells ohne die Verwendung von Hotel spezifischen Vergangenheitsdaten / William Mendat. –

MASTERARBEIT, Offenburg: Hochschule für Technik, Wirtschaft und Medien Offenburg, 2024. 9 Seiten.

Mendat, William:

Development of a machine learning model without using hotel-specific historical data / William Mendat. –

MASTER THESIS, Offenburg: Offenburg University, 2024. 9 pages.

Vorwort

Die Entwicklung eines Maschine Learning Modells ohne die Verwendung von Vergangenheitsdaten ist ein umfangreiches und sehr interessantes Thema. Es ist ein sehr weitreichendes Thema, an dem sehr lange geforscht werden könnte und an sich auch eine Wissenschaft für sich ist. Nicht alle Facetten und Möglichkeiten, dieses doch recht Komplexe Thema zu bewältigen, werden in der folgenden Arbeit dargestellt aber es wird einblick gegeben, wie an dieses Thema herangegangen werden kann.

Ich möchte mich an dieser Stelle bei allen Personen bedanken, die mich während dieser Bachelor-Thesis unterstützt haben. Ein besonderer Dank geht dabei an Prof. Dr.-Ing. Janis Keuper für die Betreuung während der Arbeit und die Möglichkeit, dieses Thema überhaupt bearbeiten zu können. Des Weiteren möchte ich mich bei Kai Schmidt und Marius Müller bedanken, die mich während der Bearbeitung tatkräftig Unterstütz haben. Als letztes möchte die happyhotel an sich danken, dass sie mir erlaubt haben bei Ihnen in der Firma diese Thesis zu schreiben.

Eidesstattliche Erklärung

Hiermit versichere ich eidesstattlich, dass die vorliegende Bachelor-Thesis von mir selbstständig und ohne unerlaubte fremde Hilfe angefertigt worden ist, insbesondere, dass ich alle Stellen, die wörtlich oder annähernd wörtlich oder dem Gedanken nach aus Veröffentlichungen, unveröffentlichten Unterlagen und Gesprächen entnommen worden sind, als solche an den entsprechenden Stellen innerhalb der Arbeit durch Zitate kenntlich gemacht habe, wobei in den Zitaten jeweils der Umfang der entnommenen Originalzitate kenntlich gemacht wurde. Ich bin mir bewusst, dass eine falsche Versicherung rechtliche Folgen haben wird.

Offenburg, 30.03.2024

William Mendat

Sperrvermerk

Die vorliegende Abschlussarbeit beinhaltet vertrauliche Informationen und interne Daten des Unternehmens happyhotel. Sie darf aus diesem Grund nur zu Prüfungszwecken verwendet und ohne ausdrückliche Genehmigung durch die happyhotelweder Dritten zugänglich gemacht, noch ganz oder in Auszügen veröffentlicht werden. Die Sperrfrist endet 5 Jahre Jahre nach dem Einreichen der Arbeit bei der Hochschule Offenburg. Unbeschadet hiervon bleibt die Weitergabe der Arbeit und Einsicht in die Arbeit an die mit der Prüfung befassten Mitarbeiter der Hochschule und Prüfer möglich, die ihrerseits zur Geheimhaltung verpflichtet sind, sowie die Verwendung der Arbeit in eventuellen prüfungsrechtlichen Rechtsschutzverfahren nach Maßgabe der geltenden verwaltungsprozessualen Regeln.

Zusammenfassung

Entwicklung eines Maschine Learning Modells ohne die Verwendung von Hotel spezifischen Vergangenheitsdaten

Test Abstarct

Abstract

Development of a machine learning model without using hotel-specific historical data

Test Abstract

Inhaltsverzeichnis

1. Einleitung	1
1.1. happyhotel	2
1.1.1. Das Unternehmen	3
1.1.2. Dynamische Preisgenerierung	3
1.1.3. Momentane Problematik	4
1.2. Vorgehensweise	5
1.2.1. Vorgehen in der Datenwissenschaft	5
1.2.2. Benchmark Hotels	7
2. Konzepte	9
Tabellenverzeichnis	
Abbildungsverzeichnis	i
Quellcodeverzeichnis	ii
Literatur	iii
A. Anhang	iv

1. Einleitung

In einer Welt, die sich mit rasanter Geschwindigkeit digitalisiert, suchen die Menschen stets nach Wegen, um die Komplexität des modernen Lebens zu bewältigen. Diese Digitalisierung hat eine stetig wachsende Sehnsucht nach der Vorhersage zukünftiger Ereignisse hervorgebracht - sei es in der Wirtschaft, der Gesundheitsbranche oder auch im Bereich des Dienstleistungssektors wie dem Hotelgewerbe. Es ist ein Streben nach Präzision, ein Bestreben, aus Daten und Mustern eine art Kristallkugel zu formen, um die Zukunft vorhersagen zu können.

Albert Einstein hat einst mit einem Buchtitel von Ihm gesagt:

If you want to know the future, look at the past. [1]

Dieser Gedanke illustriert die gängige Annahme, dass die Vergangenheit Hinweise auf die Zukunft liefern kann. Es ist interessant anzumerken, dass dieses Zitat auch als Titel eines Buches von Einstein dient, welches seine philosophischen Ansichten zur Zeit, Raum und Vorhersage behandelt.

Doch was passiert, wenn diese Vergangenheitsdaten nicht verfügbar sind oder nicht genutzt werden können? In Branchen wie der Hotelindustrie, die oft noch auf traditionelle, statische Preisstrategien zurückgreifen, stellt sich die Frage, wie eine effektive Vorhersage ohne spezifische historische Daten möglich ist.

Es wird immer deutlicher, dass ein dynamischerer Ansatz im Hotelwesen erforderlich ist, um die Umsatzoptimierung durch Revenue Management zu steigern. Dies erfordert die Anpassung von Preismodellen an sich ändernde Nachfrage und andere Einflussfaktoren. Eine mögliche Lösung liegt in der Verlagerung der traditionellen Rolle des Revenue Managements auf Modelle, die auf breiteren Datenquellen und fortgeschrittenen Methoden des maschinellen Lernens basieren.

Die Suche nach einem solchen Modell, das ohne die spezifischen Vergangenheitsdaten eines bestimmten Hotels auskommt, bildet das Herzstück dieser Forschungsarbeit. Der Fokus liegt darauf, alternative Datenquellen zu erkunden und innovative Ansätze zu entwickeln, um Prognosen und Entscheidungsgrundlagen für das Revenue Management in der Hotellerie zu schaffen. Ziel ist es, dass diese nicht ausschließlich auf vergangenen Daten eines spezifischen Hotels basieren, sondern auf einer Vielzahl von allgemeinen, zugänglichen Informationen und fortschrittlichen Analysemethoden beruhen. Es geht darum, einen Weg zu finden, wie Hotels, selbst ohne ihre spezifischen vergangenen Daten, zukünftige Entscheidungen im Bereich des Revenue Managements treffen können, um ihre Leistung zu optimieren und ihre Wettbewerbsfähigkeit zu stärken.

1.1. happyhotel

Die vorliegende Masterthesis fängt mit einem umfassenden Überblick über die Firma happyhotel an. In diesem ersten Kapitel wird eingehend auf die fundamentale Idee und das herausragende Produkt von happyhotel eingegangen, welches einen signifikanten Beitrag zur Weiterentwicklung der Hotelbranche leistet.

Im Anschluss wird der Fokus auf das gegenwärtige Vorgehen des Unternehmens, der Dynamischen Preisgenerierung, gelegt. Diese fortschrittliche Methode, die auf einer Künstlichen Intelligenz und umfassender Datenanalyse basiert, ermöglicht es happyhotel, in Echtzeit auf den momentanen Markt zu reagieren und optimale Preise für Hotels zu generieren.

Abschließend wird in diesem einführenden Kapitel die zugrunde liegende Problematik hervorgehoben, die den Ausgangspunkt dieser Arbeit bildet. Dabei wird der Fokus auf eine Herausforderung gerichtet, die mit der Dynamischen Preisgenerierung einhergeht. Diese Problematik dient als Basis für die nachfolgende Analyse und Forschung, die darauf abzielt, innovative Lösungsansätze und Optimierungen im Rahmen der Preisstrategie von happyhotel zu entwickeln.

1.1.1. Das Unternehmen

Die Firma happyhotel wurde im Jahr 2019 von den drei Gründern Sebastian Kuhnhardt, Marius Müller und Rafael Weißmüller gegründet. Sie wollten wie der Name schon vermuten lässt Hotels glücklicher machen. Angefangen hat es mit der Erkenntnis von Sebastian, dass sich viele Hoteliers nicht mit der Dynamischen Preisgestaltung beschäftigen. Meist vertrauen diese Hoteliers einfach auf ihr Bauchgefühl, welcher Preis zur momentanen Situation passen könnte oder passen ihre Preise gar nicht an.

Somit stellten sie sich die Fragen:

- Wie können die Preise für die Übernachtung in einem Zimmer besser vorausgesagt werden?
- Wonach sollten sich die Preise richten und wie kann man sie bestimmen?

Eine Lösung musste her um mehr Dynamik in die Preisgestaltung zu bringen. Sie erschufen die Software happyhotel ein Revenue Management System. Mit happyhotel kann der Hotelier sein gesamtes Hotel analysieren und Ihm werden Preisvorschläge für seine Zimmerkategorien generiert um mehr Umsatz zu erzeugen.



Abbildung 1.1: happyhotel

1.1.2. Dynamische Preisgenerierung

Wie im vorherigen Kapitel erwähnt fokussiert sich happyhotel auf die Dynamische Preisgenerierung. Um diese Preise zu generieren braucht es vor allem zwei Sachen:

- Die Daten des Hotels wie zum Beispiel Buchungen
- Ein Vorgehen um aus den gesammelten Daten Preise zu generieren

Die Daten bekommen sie aus den verschiedensten Quellen. Sogenannte Property Management Systeme kurz PMS sind Systeme um ein Hotel zu verwalten. In diesem Property Management Systeme können Hotels zum Beispiel ihre Zimmer verwalten oder aber auch Buchungen anlegen und pflegen. Mit den Herstellern dieser

Property Management Systeme arbeitet happyhotel zusammen um an die Daten des Hotels zu gelangen.

Da die Daten vorhanden sind, braucht es ein Vorgehen um aus den Daten einen Preis zu generieren. Dazu ist happyhotel auf die folgenden zwei Ideen gekommen:

- Buchungskurvenmodell
- Kombination aus RevPAR und Buchungskurvenmodell

Das Buchungskurvenmodell war die erste Idee von happyhotel. Bei dem Buchungskurvenmodell wird sich die Vergangenheit angeschaut um das zukünftige Buchungsverhalten vorherzusagen. Ziel dabei ist es die Auslastung für einen Tag vorherzusagen um anhand dessen einen akkuraten Preis zu bestimmen.

Da bei dem Buchungskurvenmodell sich einige Schwächen aufgezeigt haben, wurde ein neues Modell erschaffen um die Schwächen entgegen zu wirken. Es sollte nun der RevPAR wert vorhergesagt werden und mit dem Buchungskurvenmodell angepasst werden. RevPAR steht dabei für Revenue per Available Room. Auch bei diesem Modell wird sich die Vergangenheit des Hotels angeschaut um den zukünftigen Umsatz pro verfügbarem Zimmer vorherzusagen und basierend darauf den endgültigen Preis zu ermitteln.

1.1.3. Momentane Problematik

Wie es Albert Einstein schon sagte: Soll die Zukunft vorhersagen gesagt werden, so sollte sich die Vergangenheit angeschaut werden. Auf diesem Grundprinzip ist happyhotel auch vorgegangen, sie schauen sich bei beiden Ansätzen die Vergangenheit an um Vorhersagen über die Zukunft zu tätigen.

Doch was passiert, wenn die Daten der Vergangenheit nicht vorhanden sind? Dies kann zum Beispiel passieren wenn ein Hotel die Software nutzen möchte, welches erst in der Zukunft eröffnet. Auch dieses Hotel soll mit adäquaten Preisvorschlägen gefüttert werden. Die soeben beschriebene Situation ist ein generelles Problem im Maschine Learning Bereich. Es kommt nicht allzu selten vor, dass keine Vergangenheitsdaten aus den verschiedensten Gründen vorliegen. Dieser Problematik soll in dieser Arbeit auf dem Grund gegangen werden.

Ziel dieser Arbeit ist es, ein Modell zu entwickeln welches Preisempfehlungen für ein Hotel liefert, für das es bisher noch keine Vergangenheitsdaten gibt. Dieses Modell soll dann für folgende zwei Szenarien genutzt werden können:

- Neue happyhotel Kunden ohne Daten
- Nachfrageeinschätzung für bestimmte Märkte

1.2. Vorgehensweise

Vor dem Eintauchen in die Lösung eines Problems ist es entscheidend, einen klaren Weg dorthin festzulegen. Antoine de Saint-Exupéry hat mit den Worten *Ein Ziel ohne Plan ist nur ein Wunsch* treffend darauf hingewiesen, dass ein bloßes Ziel ohne einen durchdachten Plan lediglich eine vage Vorstellung bleibt. Das Verständnis und die Festlegung einer angemessenen Vorgehensweise sind daher der Schlüssel, um ein Problem effektiv anzugehen. Aufgrund dessen wird in dem folgenden Abschnitt dieser Arbeit auf die Vorgehensweise eingegangen. Zudem werden innerhalb dieser Sektion die Benchmark-Hotels ermittelt an welchen getestet werden kann, ob die Vorgehensweise ein Erfolg war.

1.2.1. Vorgehen in der Datenwissenschaft

Ein Projekt welches in der Datenwissenschaft (engl. Data Science) angesiedelt ist, beginnt in der Regel mit einem geschäftlichen Problem, so wie es auch in dieser Thesis der Fall ist. Sobald das Problem klar definiert ist, sollten ein oder mehrere Konzepte ausgearbeitet werden, wie das Problem gelöst werden kann. Diese Konzepte sollen als Leitfaden dienen um das angestrebte Ziel zu erreichen. In der Regel ist es ratsam mehr als ein Konzept auszuarbeiten, da so ausweichmöglichkeiten festgelegt werden können, sollte ein Konzept nicht funktionieren. So werden auch in dieser Arbeit, in dem Kapitel *Konzepte*, für das vorliegende Problem Konzepte ausgearbeitet und Evaluert.

Mit der klaren Definition des Problem und der darauffolgenden Konzeptionieren, kann der Datenwissenschaftler mit Hilfe des *OSEMN*-Vorgangs die Problematik angehen [2].

Der *OSEMN*-Vorgang besteht aus den folgenden Elementen:

- Obtain data (Erhalten von Daten)
- Scrub data (Daten reinigen)
- Explore data (Untersuchen von Daten)
- Model data (Modelldaten)
- Interpret results (Interpretieren von Ergebnissen)

Obtain data

Die wichtigste Ressource des 21. Jahrhunderts besteht in den Daten, die zur Verfügung stehen. Dies erläuterte Klaus Schwab, der Gründer des Weltwirtschaftsforums, mit seinen Worten: *Die wertvollste Ressource des 21. Jahrhunderts sind nicht mehr Öl, sondern Daten.* Aufgrund dessen besteht der erste Schritt, nach der Evaluierung der Konzepte, in der Beschaffung der Daten. Es muss zunächst ein Überblick geschaffen werden. Zum Überblick gehören Informationen wie:

- Welche Daten bereits vorhanden sind.
- Welche Daten eventuell noch intern neu erworben werden müssen.
- Welche Daten aus dem Internet gezogen werden können

Sobald die Daten beschaffen worden sind, kann mit dem nächsten Schritt fortgefahren werden.

Scrub data

Der nächste Schritt besteht darin, die Daten zu bereinigen. Zum bereinigen der Daten gehört der Vorgang mit dem die Daten in ein standardisiertes Format gebracht werden. Dazu gehört der Umgang mit fehlenden Daten, sowie die Korrektur von Fehlern und das Entfernen von sogenannten *outlier* [2]. Outlier sind Daten, welche im Verhältnis zu der gesamten Datenmenge aus der Reihe tanzen.

Explore data

Die Datenuntersuchung oder auch Datenanalyse dient dazu, um mit den Daten vertraut zu werden und ein besseres Verständnis für die Daten zu gewinnen. Dies ist ein sehr wichtiger Schritt bei einem *Data Science* Projekt, da nur dann ein gutes Ergebnis erzielt werden kann, wenn die Daten, mit denen gearbeitet werden kann, verstanden sind. Das Verständnis über die Daten trägt zu dem auch maßgeblich dazu bei, den richtigen Ansatz für die Modellierung zu finden.

Model

Nach dem Erforschen der Daten, kann ein *Maschine Learning* Modell eingesetzt werden. Es existieren viele verschiedene Modelle, die meist einen für einen speziellen Fall implementiert worden sind. Die Auswahl des Modells hängt ganz davon ab, welche Art von Problem gelöst werden soll und welche Daten vorhanden sind, um dieses Problem zu lösen.

Interpret results

Der letzte Schritt besteht darin, die gesammelten Ergebnisse zu interpretieren. Dazu gehörend ist die Entscheidung, ob die erzielten Ergebnisse gut oder schlecht sind. Meistens werden zur Hilfe der Entscheidung ob die Ergebnisse gut oder schlecht sind, Diagramme, Grafiken und Tabellen erstellt. Es soll hier zudem auch entschieden werden, ob die Ergebnisse verwendet werden sollten oder ob nicht weiter geforscht werden muss. So entsteht ein Kreislauf, der mit dem erstellen weiteren Konzepte startet und mit dem Interpretieren der Ergebnisse endet.

1.2.2. Benchmark Hotels

Noch vor der Konzeptionierung zur eigentlichen Lösung der Problematik, sollten schon vorhandene Hotels in unserer Datenbank als *Benchmark-Hotels* ausgesucht werden. Die Idee dahinter ist es, Hotels auszusuchen, bei denen viele Daten vorhanden sind und so zu tun als wären gar keine Daten von diesen Hotels vorhanden. Mithilfe dieser Hotels sollen denn die ausgearbeiteten Konzepte von Kapitel *Konzepte* validiert werden. Ein Konzept wird dann als gut empfunden, wenn es in der

Lage ist, gute Preisvorschläge für die Benchmark-Hotels zu erzeugen.

Da nicht jedes Hotel, welches in der Datenbank vorhanden ist in Frage kommt, wurden einige Kriterien aufgestellt, die ein Hotel erfüllen muss, um als Benchmark-Hotel gelten zu können. Diese Kriterien sehen wie folgt aus:

- Haben Daten von mehr als zwei Jahren
- Haben einen Benutzer mit der Role *Revenue Manager*
- Haben oft Preise geändert
- Haben oft vorgeschlagene Preise von happyhotel nicht angenommen

Der Hintergrund warum die letzten drei Kriterien dazu kamen, liegt darin, dass diese Hotels mit den Preisvorschlägen von happyhotel nicht zufrieden sind und vermutlich auch manuell *Revenue Management* betreiben. Die Hoffnung besteht darin, dass das Konzept nicht nur für Hotels gute Preisvorschläge liefert, sondern, dass das auch verwendet werden kann, als Alternative zu den vorhandenen zwei Modellen, für Hotels die noch manuell dynamische Preise gestalten.

Es wurde zur Analyse der Hotels eine *json*-Datei erstellt. Diese *json*-Datei besteht aus einer Liste von einzelnen Objekten. Jedes Objekt innerhalb der Liste repräsentiert ein Hotel in der Datenbank. Ein Beispiel für diese *json*-Datei könnte wie folgt aussehen:

```
1  [  
2    {  
3      "company_id": "11111111111111",  
4      "not_accepted_recs": 10,  
5      "price_change_average": 5  
6    }  
7  ]
```

Listing 1.1: Beispielhafte json-Datei

Das Feld *not_accepted_recs* beschreibt dabei wie viele Preisvorschläge das Hotel nicht angenommen hat beziehungsweise ignoriert oder abgelehnt hat und das Feld *price_change_average* ist die durchschnittliche Preisänderung pro Tag. Anhand von diesen Informationen konnten zwei Hotels als Benchmark-Hotels ausgesucht werden.

2. Konzepte

Tabellenverzeichnis

Abbildungsverzeichnis

1.1. happyhotel	3
---------------------------	---

Listings

1.1. Beispielhafte json-Datei	8
---	---

Literatur

- [1] M. KARATAŞ und A. Einstein, *If You Want to Know the Future, Look at the Past*. Amazon Digital Services LLC - KDP Print US, 2017, ISBN: 9781976758331. Adresse:
https://books.google.de/books?id=u_14twEACAAJ.
- [2] Amazon Web Services, Inc., *Was ist Datenwissenschaft? – Datenwissenschaft erklärt – AWS*, 15.11.2023. Adresse:
<https://aws.amazon.com/de/what-is/data-science/>.

A. Anhang

Test