



## Projektliste

Zeitraum:	10/2019 - 03/2020
Auftrag:	<b>Modul 13: Masterarbeit Prognose von Marktpreisen mittels ML-Methoden am Beispiel des Gebrauchtwagenmarktes</b>
Technologien	Python, Azure Cloud, PostgreSQL, Oracle, Shiny R
Bibliotheken	Selenium, Pandas, Sklearn
Projektbeschreibung:	<p>Um dem immer weiter steigenden Druck in der Automobilhandelsbranche standzuhalten ist es wichtig, seinen Markt und die Preise zu kennen, um sein Produkt zum optimalsten Preis anbieten zu können. Somit kann die höchste Rentabilität erzielt werden. Besonders das Bestandsmanagement und die Einpreisung der Fahrzeuge ist ein Bereich, auf das ein besonderes Augenmerk gelegt werden sollte. In dieser Arbeit wird untersucht, welches Machine Learning Modell am besten geeignet ist, um den Marktpreis eines Gebrauchtwagens vorherzusagen. Hierzu ist es im ersten Schritt notwendig die Marktdaten aus den Autobörsen zu extrahieren und in einer geeigneten Form zu speichern. Des Weiteren soll der Prozess für die Ermittlung des Risikos im aktuellen Fahrzeugbestand optimiert werden. Die Ergebnisse sollen anschließend in einer geeigneten Lösung visualisiert werden, um zum einen den Verkaufsleiter bei der Preisfindung / Einpreisung des Fahrzeugs zu unterstützen und um den Prozess in der Controllingabteilung zu verbessern, um die Risikobewertung des Gebrauchtwagenbestandes und die damit verbundenen Rückstellungen in der Buchhaltung zu optimieren.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Webscraping der führenden Automobilbörsen mittels BeautifulSoup und Selenium</li> <li>• Speicherung der Daten in die Azure Postgre SQL-Cloud</li> <li>• Selektion der relevanten Daten aus dem Produktivsystem (Oracledatenbank)</li> <li>• Preprocessing und Zusammenführung der Daten</li> <li>• Evaluation einer geeigneten ML Methode zur Preisvorhersage in Python</li> <li>• Entwicklung einer interaktiven Dashboardlösung zur Kommunikation der Ergebnisse mittels Shiny R / Shiny Dashboard</li> <li>• Diskussion der Ergebnisse</li> </ul>

Zeitraum:	07/2019 – 09/2019
Auftrag:	<b>Modul 12 - Applied Statistics</b> Vorhersage von Filmbewertungen auf Basis von TMBD Filmreviews
Technologien	R
Bibliotheken	Readr, Tidyverse, Dplyr, Ggplot2, Plotly, Jsonlite, RSentiment, Psych
Projektbeschreibung:	<p>Ein Filmstudio möchte in der „Filmpitch Phase“ (Ideenfindung) wissen, wie unsere Filme bei dem Publikum ankommen werden und mit welcher Filmkritik wir rechnen können (Bewertungen). Das Ziel ist es, ein Modell zu entwickeln, dass es möglich macht, die Bewertungen von Filmtiteln vorherzusagen.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Zusammenführen und Aufbereiten der Datenbestände</li> <li>• Sentimentanalyse der Reviews</li> <li>• Korrelationsanalysen</li> <li>• Explorative Datenanalyse</li> <li>• Evaluation einer geeigneten ML Methode zur Vorhersage von Filmbewertungen in "R"</li> <li>• Decision Tree, K-Nearest Neighbor , Random Forrest,</li> </ul>
Zeitraum:	05/2019 – 07/2019
Auftrag:	<b>Modul 11: Web and Social Media Analytics</b> <b>Früherkennung von negativen Bewertungen auf der Plattform Kununu am Beispiel von Premium Automobilmarken</b>
Technologien	Python, Amazon Web Service (AWS), Postgre SQL, Shiny R
Bibliotheken	psycopg2, json, pandas, BeautifulSoup, re
Projektbeschreibung:	<p>Den größten Unmut äußern Benutzer von Karriereportalen darüber, dass Unternehmen zu den Kommentaren / negativen Kommentaren keine Stellung beziehen. (In der "Studie Saatkorn 2018" erachten 66,4 % der Befragten dies als sehr schlecht oder schlecht). Diese Problemstellung soll die Data Science Lösung unterstützen. Für die Unternehmen soll ein Dashboard erstellt werden, das die negativen Kommentare anzeigt und bei Überschreitung eines bestimmten negativen Sentiment Schwellenwerts, automatisch eine Benachrichtigung (E-Mail Notification) an den jeweiligen Personalverantwortlichen raussendet. Somit kann auf negative Kommentare schneller und effektiver reagiert werden.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Webscraping der Kommentare der Premium-Automobilmarken mittels Beautiful Soup und Selenium</li> <li>• Speicherung der Daten (Postgre SQL)</li> <li>• Datenzusammenführung und Bereinigung</li> <li>• Evaluation einer geeigneten Methode zur Sentimentanalyse der Bewertungen</li> <li>• Entwicklung einer interaktiven Dashboardlösung zur Kommunikation der Ergebnisse mittels Shiny R / Shiny Dashboard</li> </ul>

Zeitraum:	04/2019-06/2019
Auftrag:	<b>Modul 10: Data-Mining-Process: Algorithms and Implementation</b> <b>Erstellen eines Prototyps für einen TOP N Film Recommender</b>
Technologien	Python
Bibliotheken	Pandas, Numpy, Matplotlib, Surprise, Sklearn, Math
Projektbeschreibung:	<p>Das Ziel ist es, einen Top N Film Recommender zu erstellen. Dieser gibt auf Basis von Benutzerdaten (Mitglieder) die Besten personalisierten Empfehlungen im gesamten Katalog für jedes Mitglied aus. Nach Eingabe des Benutzernamens (ID) sollen dem Benutzer Filme vorgeschlagen werden, die für ihn interessant sind. Im Laufe des Projekts werden verschiedene Verfahren zur Konzeption eines Recommenders angewendet und die Vor- und Nachteile diskutiert</p> <ul style="list-style-type: none"> <li>• Explorative Datenanalyse / Preprocessing</li> <li>• Random-Based Recommender</li> <li>• Popularity-Based Recommender</li> <li>• Score-Based Recommender</li> <li>• Content-Based Recommender (Similarity Based)</li> <li>• Collaborative Filtering Recommender</li> </ul>
Zeitraum:	02/2019 – 04/2019
Auftrag:	<b>M 9 – BI and Big Data-Architecture</b> <b>Empfehlungsvorhersagen von Börsenkursen</b>
Technologien	Google Cloud Plattform (GCP), Python, Power BI
Bibliotheken	
Projektbeschreibung:	<p>Ziel des Prototyps ist es, einem Investor eine Kaufempfehlung zu geben, indem die Auswirkungen einer Nachricht auf die Börsenkurse vorhergesagt werden soll.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Extraktion der Texte und Stockpreise mittels Webservice Tiingo (API)</li> <li>• Design des ETL Prozesses mittels Google Dataflow</li> <li>• Datenstreaming via PubSub / Lambda Architektur</li> <li>• Vorhersage der Stockpreise mittels Google ML und Auto ML</li> <li>• Anlage der Cronjobs in der Google App Engine zur Automatisierung des Systems</li> </ul>

Zeitraum:	12/18 - 02/19
Auftrag:	<b>Modul 8 - Business- and CRM-Analytics</b> <b>Optimierung der E-Mailkampagnen für das Marketingforum der Hochschule der Medien</b>
Technologien	Power BI, Microsoft Dynamics CRM, Azure ML Studio
Bibliotheken	
Projektbeschreibung:	<p>Entwicklung eines Konzepts für den Marketingmanager zur Optimierung der Marketingmaßnahmen für das Marketingforum auf Basis von Marketingkampagnendaten aus Rapidmail.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Analyse der Kampagnen und Untersuchung der aufgestellten Hypothesen</li> <li>• Analyse der Fachinteressen und Entwicklung eines geeigneten Kundenscores</li> <li>• Vorhersage der Teilnahme eines Kunden an der Veranstaltung</li> <li>• Erstellung eines interaktiven Power BI Dashbaords</li> <li>• Integration der Ergebnisse in Microsoft Dynamics CRM (Feedbackloop)</li> </ul>
Zeitraum:	10/2018 - 12/2018
Auftrag:	<b>Modul 7 – Programming for Data Science</b> <b>Analyse der Daten der Olympischen Spiele</b>
Technologien	Python
Bibliotheken	
Projektbeschreibung:	<p>Analyse der Daten der Olympischen Spiele aus 118 Jahren (1896 - 2014) zur Beantwortung der Forschungsfragen</p> <ul style="list-style-type: none"> <li>• Bringen die Länder mit den meisten Einwohnern verhältnismäßig die meisten Teilnehmer?</li> <li>• Kann man davon ausgehen, dass mehr Teilnehmer auch mehr Medaillen bedeuten?</li> <li>• Gibt es einen Zusammenhang zwischen dem BIP eines Landes und dessen Erfolg?</li> </ul> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Webcrawling der Daten für die Information zum Bruttoinlandsprodukt und Ländergröße (TraidingEconomics / Worldometers)</li> <li>• Zusammenführen und Aufbereiten der Datenbestände</li> <li>• Explorative Datenanalyse</li> <li>• Datenanalyse zur Beantwortung der Forschungsfragen</li> </ul>

Zeitraum:	08/2018 – 09/2018
Auftrag:	<b>M6 - BI- and Big-Data-Architectures Music Lyrics Analyzer</b>
Technologien	Rapidminer, Tableau
Bibliotheken	
Projektbeschreibung:	<p>Für den Music Lyrics Analyzer wurden insgesamt 250 Lieder aus 5 Genres mit jeweils 5 Interpreten ausgewählt. Basierend auf den Lyrics sollen zwei Modelle zur Vorhersage des Interpreten und des Künstlers erstellt werden.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Datenextraktion der Songtexte und Metdaten aus MLDB</li> <li>• Textanalyse der Lyriks nach Worthäufigkeiten, TOP 10 Wörter (Künstler, Genre) , TOP 10 - 2 grams</li> <li>• Feature Selection (Wordmatrizen, Worttags-Stimmung)</li> <li>• Modell Building (Naive Bayes, SVM, KNN, Decision Tree, Random Forrest)</li> <li>• Ergebnisevaluation und Visualisierung mittels Tableau</li> </ul>
Zeitraum:	06/2018 – 09/2018
Auftrag:	<b>M5 - Data-Warehouse-Workshop Entwicklung eines Prototyps eines Clinical Trial Management System mittels SAP BW ON HANA</b>
Technologien	SAP Hana Studio, SAP Logon, BW on HANA, MS Analysis, MS Visio
Bibliotheken	
Projektbeschreibung:	<p>In der klinischen Forschung möchte der Sponsor zu jeder Zeit die Übersicht über die Studien haben, um im Projektmanagement schnell und risikobasiert entscheiden und reagieren zu können. Reporting ist hierbei das A und O und ohne eine gute Datengrundlage unmöglich.</p> <p>Als Lösungsansatz bietet sich hier ein CTMS (Clinical Trial Management System) mit einem soliden DWH-Fundament an.</p> <ul style="list-style-type: none"> <li>• Bottom UP Konzeption - Analyse der Datenfelder und Attribute im Quellsystem (csv-Dateien)</li> <li>• Erstellen eines ERM Modells / DIM FACT Modells</li> <li>• Erstellen einer KPI Dokumentation</li> <li>• Prototyping in Power Pivot. Testen des DIM FACT Modells und der KPI's</li> <li>• Konzeptionelle Ausarbeitung der einzelnen SAP BW Elemente (PSA, ADSO, Composit Provider, Info Objects, Queries)</li> <li>• Umsetzung im SAP BW System und anschließender Aufbau von Dashboards mittels Analysis for Office</li> </ul>

Zeitraum:	04/2018 – 06/2018
Auftrag:	<b>M4 - Ethics and Law Recht und Eigentum an Daten</b>
Technologien	Power Point
Bibliotheken	
Projektbeschreibung:	<p>Präsentation zum Themengebiet Recht und Eigentum an Daten am Beispiel der Automobilindustrie</p> <ul style="list-style-type: none"> <li>• Einleitung Recht</li> <li>• Fallstudien – KFZ-Instandhaltung &amp; Wartung / Skoda Connect</li> <li>• Rechtliche Erfassung des Dateneigentums im geltenden Recht (de lege lata)</li> <li>• Entwicklung eines neuen Datengesetzes bzw. sukzessive Vorgehensweise (de lege ferenda)</li> </ul>
Zeitraum:	02/2018 – 04/2018
Auftrag:	<b>M3 - Introduction to Data Science Sentimentanalyse der Tweets auf Twitter im Zeitraum des Superbowls 2018</b>
Technologien	Rapidminer
Bibliotheken	Tweepy
Projektbeschreibung:	<p>Sentimentanalyse der Tweets auf Twitter im Zeitraum des Superbowls 2018 (03.02.2018 bis 15.02.2018) für die Erfolgsmessung der Werbeschaltungen im Bereich Automobil</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Extrahieren der Tweets von den gängigen Accounts (@SuperBowl,...) mittels Tweepy API und Rapidminer (2,8 Mio Tweets)</li> <li>• Datenexploration des Datenbestandes und Analyse der Verteilung nach Uhrzeiten, Tweets, Retweets und Hashtags</li> <li>• Datenaufbereitung mit Filterung des Keywords, Entfernen der Retweets und Bereinigung der Twitertexte auf Sonderzeichen</li> <li>• Methodenauswahl und Durchführung der Sentiment Analyse mittels Rapidminer</li> <li>• Anschließend erfolgten eine Auswertung und Evaluation der Ergebnisse</li> </ul>

Zeitraum:	12/2017 – 01/2018
Auftrag:	<b>M2 - Introduction to Business Analytics Simulation eines Verkaufsszenarios am Beispiel eines Fahrradhändlers</b>
Technologien	SQL-Server Integration Services (SSIS), Power BI, Power Pivot
Bibliotheken	
Projektbeschreibung:	<p>Simulation von Geschäftsereignissen eines Fahrradhändlers auf der Plattform bi-academy.eu</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Datenmodellierung / DIM FACT Modell</li> <li>• Erstellen und Laden des Datamarts</li> <li>• OLAP Reporting</li> <li>• Erstellen von Cubes</li> <li>• Modellierung ETL Prozess</li> <li>• Data Exploration</li> </ul>
Zeitraum:	10/2017-12/2017
Technologien	Power Point, Proto.io
Bibliotheken	
Auftrag:	<b>Modul 01: New Business Models and Strategies Entwicklung einer Geschäftsmodellinnovation für Chefkoch.de</b>
Projektbeschreibung:	<p>Für das Unternehmen Chefkoch soll eine Geschäftsmodellinnovation entwickelt werden. Dies umfasste in der ersten Projektphase die Analyse der Prozesse von Chefkoch, um ein Verständnis des Geschäftsmodells zu generieren. In der zweiten Projektphase wurde anschließend eine Geschäftsmodellinnovation für das Unternehmen erarbeitet.</p> <p>Vorgehensweise:</p> <ul style="list-style-type: none"> <li>• Analyse zum Unternehmen Chefkoch</li> <li>• Analyse des Geschäftsmodells nach dem Lean Canvas- / Business Modell Canvas Modell</li> <li>• Entwickeln eines ER Modells auf Basis von Mockupdaten aus den Analysen des Webauftritts</li> <li>• Ableitung von relevanten KPI's für das Unternehmen</li> <li>• Entwicklung der Geschäftsmodellinnovation „TreuKoch“</li> <li>• Entwicklung einer User-Interface App für TreuKoch (Proto.io)</li> </ul>