**Einreichung Projektvorschlag**

**für**

**Data Analytics**

**(28.07.25 – 22.08.25)**

**Projektersteller/Gruppe:**

**Wu, Guangfa**

**Daten:**  [**https://www.kaggle.com/datasets/hellbuoy/car-price-prediction**](https://www.kaggle.com/datasets/hellbuoy/car-price-prediction)

**Projektidee, Beschreibung:**

**Problembeschreibung:**

Ein chinesisches Automobilunternehmen, Geely Auto, plant den Eintritt in den US-amerikanischen Markt, indem es dort eine Produktionsstätte errichtet und Autos lokal herstellt, um mit amerikanischen und europäischen Konkurrenten zu konkurrieren. Dazu hat Geely ein Beratungsunternehmen für die Automobilbranche beauftragt, um die Faktoren zu analysieren, die den Preis von Autos beeinflussen. Insbesondere möchten sie verstehen, welche Faktoren den Preis von Autos auf dem amerikanischen Markt bestimmen, da diese sich deutlich vom chinesischen Markt unterscheiden könnten.

**Das Unternehmen möchte wissen:**

* Welche Variablen sind signifikant für die Vorhersage des Autopreises?
* Wie gut erklären diese Variablen den Preis eines Autos?

Basierend auf verschiedenen Marktforschungen hat das Beratungsunternehmen einen umfangreichen Datensatz mit unterschiedlichen Fahrzeugtypen auf dem US-amerikanischen Markt zusammengetragen.

**Lösung der Aufgabestellung:**

Ein Modell soll entwickelt werden, das den Preis von Autos basierend auf den verfügbaren unabhängigen Variablen am besten vorhersagt. Da ein Auto aus verschieden Komponenten besteht, ist es nur sinnvoll, alle Komponenten zusammen zu betrachten und ein Preismodell zu erstellen. Gleichzeitig kann auch untersucht werden welche einzelnen Faktoren den Preis beeinflussen.

**Anwendung:**

Das Modell soll dem Management dabei helfen zu verstehen:

* Wie sich der Preis je nach Veränderung einzelner Merkmale verändert
* Welche Stellschrauben (Design, Strategie etc.) genutzt werden können, um bestimmte Preisniveaus zu erreichen.
* Zusätzlich ist das Modell ein hilfreiches Werkzeug, um die Preisdynamik in einem neuen Markt wie den USA zu verstehen.

**Projektschritte, Aufgaben zum Projekt (falls schon vorhanden):**

Folgende Schritte sind zu unternehmen, um das definierte Ziel zu erreichen

Es werden Arbeitsschritte nach CRISP-DM vorgesehen.

1. Business Understanding (Aufgabendefinition)
2. Data Understanding (Auswahl der relevanten Datenbestände)

2.1 Deskriptive Analyse

2.2 Explorative Datenanalyse (EDA)

1. Data Preparation (Datenaufbereitung)

3.1. Datenaufbereitung

3.2. Umwandlung von kategorischen Daten in numerischen Daten

1. Modeling (ML-Maschinelles Lernen)

4.1 Ein lineares Modell mit One-Hot Encoding für kategorische Spalten

4.2 Ein lineares Modell mit ordinal Encoding für kategorische Spalten

4.3 Ein lineares Modell mit ordinal Encoding und **Ausreißer behandeln**

4.4 Ein lineares Modell durch Transformation mit sqrt() oder x^2

1. Evaluation (Bewertung und Interpretation der Ereignisse)

5.1 Modellbewertung

**5.2 Transformation von kategorialen Merkmale in numerische Werte**

**5.3. Modellauswahl für die Anwendung(Deployment)**

1. Deployment (Anwendung der Ergebnisse)

6.1. Dashoard für Modelltraining für ein ausgewähltes Preissegment

6.2. Dashoard für Preis-Vorhersage