**Der Inhalt**

[Einleitung 3](#_Toc167972877)

Kapitel [1. Analyse des Automobilmarktes 5](#_Toc167972878)

[1.1 Trends auf dem Automobilmarkt 5](#_Toc167972879)

[1.2. Analyse von Angebot und Nachfrage auf dem Automobilmarkt 5](#_Toc167972880)

[1.3. Wettbewerbsanalyse von Automobilmarken in Kirgisistan 5](#_Toc167972881)

Kapitel [2. Methoden zur Vorhersage von Autopreisen 6](#_Toc167972882)

[2.1. Statistische Methoden 6](#_Toc167972883)

[2.2. Methoden des maschinellen Lernens 7](#_Toc167972884)

Kapitel [3. Software-Implementierung 9](#_Toc167972885)

[3.1. Beschreibung der Merkmale 9](#_Toc167972886)

[3.2. Datenbereinigung 11](#_Toc167972887)

[3.3. Erstellen eines Modells 14](#_Toc167972888)

[3.4. Auswahl des besten Modells 21](#_Toc167972889)

[Schlussfolgerung 23](#_Toc167972890)

[Liste der verwendeten Quellen 25](#_Toc167972891)

# **Einleitung**

Die Vorhersage von Autopreisen spielt sowohl für Käufer als auch für Autohändler in der Automobilindustrie eine wichtige Rolle. Die Stabilität und Genauigkeit solcher Prognosen ermöglicht es, Angebot und Nachfrage auf dem Markt vorherzusagen, optimale Preisstrategien zu bestimmen und rationale Entscheidungen über den Kauf und Verkauf von Autos zu treffen.

Der Automobilmarkt ist dynamisch und unterliegt verschiedenen externen und internen Faktoren, die die Preisdynamik beeinflussen können. Das wirtschaftliche Umfeld, technologische Innovationen, Änderungen der Regulierung sowie Verbraucherpräferenzen haben einen erheblichen Einfluss auf die Autopreise.

Die Anwendung von Modellen des maschinellen Lernens zur Vorhersage von Autopreisen bietet viele Vorteile. Erstens können Sie damit eine Vielzahl von Faktoren berücksichtigen, die sich auf die Preisgestaltung auswirken, wie z. B. Marke und Modell des Autos, seine technischen Eigenschaften, Alter, Kilometerstand, Zustand sowie externe Faktoren wie die wirtschaftliche Situation, saisonale Schwankungen und Verbraucherpräferenzen. Zweitens können solche Modelle kontinuierlich aktualisiert und verbessert werden, wenn neue Daten verfügbar sind, um sicherzustellen, dass die Prognosen aktuell und genau sind. Drittens kann die Anwendung von maschinellem Lernen die Subjektivität und menschliche Fehler bei Preisentscheidungen erheblich reduzieren.

**Gegenstand der Studie** ist der Automobilmarkt, zu dem auch Gebrauchtwagen gehören. **Gegenstand der Forschung** sind die Methoden zur Prognose von Preisen für Autos, einschließlich ihrer Anwendung und Effizienzanalyse.

**Ziel des Kursprojekts** ist es, ein Modell zur Vorhersage von Autopreisen mit modernen Methoden der Datenanalyse und des maschinellen Lernens zu entwickeln. Dieses Modell wird zu einer genaueren Bestimmung der Preisdynamik im Automobilmarkt beitragen und nützliche Erkenntnisse für die Entscheidungsfindung sowohl für Verbraucher als auch für Marktteilnehmer liefern.

Um dieses Ziel zu erreichen, wurden folgende Aufgaben identifiziert:

* Untersuchung des aktuellen Zustands des Automobilmarktes und der Hauptfaktoren, die die Autopreise beeinflussen.
* Analysieren Sie bestehende Methoden der Preisprognose und wählen Sie die am besten geeignete aus.
* Sammeln und Vorbereiten von Daten für die Erstellung eines Modells.
* Entwickeln und testen Sie ein Prognosemodell.
* Bewerten Sie die Genauigkeit des Modells und ziehen Sie Schlussfolgerungen über seine Anwendbarkeit.

**Strukturiert** besteht das Kursprojekt aus: einer Einleitung, drei Kapiteln, einer Schlussfolgerung und einer Liste der verwendeten Quellen.

An diesem Kursprojekt arbeitete folgendes Team:

1. **Zarlykov Kelsinbek** war verantwortlich für die Leitung des Teams und die Koordination der Arbeit sowie für wichtige Entscheidungen über das Projekt. Zu seinem Beitrag gehört auch die Unterstützung bei der Erstellung von Dokumentationen und deren Einhaltung der Standards. Seine Arbeit spiegelt sich in den Abschnitten 1,2,3 wider.

2. **Erdovlatov Ismadiyar** beschäftigte sich mit der Entwicklung und Wartung von Code sowie mit der Implementierung von Algorithmen zur Datenvisualisierung und zum Testen von Modellen des maschinellen Lernens. Sein Beitrag wird in den Abschnitten 1,2,3 vorgestellt.

3. **Mirlanbek Daribayev** entwickelte den Frontend-Teil des Projekts, erstellte die Benutzeroberfläche und sorgte für die Integration mit dem Backend. Seine Arbeit wird in den Abschnitten 1,2,3 beschrieben.

# **Kapitel 1. Analyse des Automobilmarktes**

**1.1 Trends auf dem Automobilmarkt**

Insbesondere werden Veränderungen der Verbraucherpräferenzen in Bezug auf die Modelle Mercedes-Benz W210, Honda FIT, KIA K5 und Toyota Camry 70 [[1]](#footnote-1)analysiert. In Kirgisistan gibt es auch ein erhöhtes Interesse an selbstfahrenden Technologien und dem Internet der Dinge, was sich auf die veränderten Verbraucherpräferenzen und -anforderungen an Autos auswirkt. Die Analyse dieser Trends hilft zu verstehen, welche Veränderungen im Design, in der Funktionalität und im Marketing von Autos auf dem kirgisischen Markt gefragt sein könnten.

## **1.2. Analyse von Angebot und Nachfrage auf dem Automobilmarkt**

Die Analyse von Angebot und Nachfrage auf dem Automarkt in Kirgisistan untersucht die Verkaufsdynamik, die Marktanteile verschiedener Automarken sowie die Auswirkungen wirtschaftlicher Faktoren wie Einkommenshöhe der Bevölkerung, Kreditzinsen und Steuerpolitik auf Preise und Verkaufsmengen von Autos.

## **1.3. Wettbewerbsanalyse von Automobilmarken in Kirgisistan**

Analyse des Wettbewerbsumfelds auf dem Pkw-Markt in Kirgisistan unter Berücksichtigung von Mercedes-Benz W210, Honda FIT, KIA K5 und Toyota Camry 70. Die wichtigsten Marktteilnehmer, ihre Vertriebs- und Marketingstrategien, Marktanteile sowie die wichtigsten Wettbewerbsvor- und -nachteile werden untersucht. Der Einfluss globaler und lokaler Faktoren auf die Position von Marken auf dem kirgisischen Markt wird ebenfalls berücksichtigt.

# **Kapitel 2. Methoden zur Vorhersage von Autopreisen**

## **2.1. Statistische Methoden**

Statistische Methoden sind ein wichtiges Werkzeug zur Analyse und Prognose von Autopreisen. Sie basieren auf der Analyse statistischer Daten und Muster, die in historischen und aktuellen Preisdaten identifiziert wurden.

Zeitreihenanalyse:

Einer der wichtigsten Ansätze in der statistischen Prognose ist die Analyse von Zeitreihen von Autopreisen. Wir untersuchen Daten über einen bestimmten Zeitraum und identifizieren wichtige Trends, saisonale Schwankungen und zyklische Preisänderungen. Dies ermöglicht es uns, zukünftige Veränderungen im Automobilmarkt vorherzusagen und entsprechende Entscheidungen zu treffen.

Anti-Aliasing- und Extrapolationsmethoden:

Glättungs- und Extrapolationsmethoden werden verwendet, um zukünftige Autopreise vorherzusagen. Wir wenden verschiedene Techniken wie gleitende Durchschnitte und exponentielle Glättung an, um Rauschen und Ausreißer in den Daten zu eliminieren und zukünftige Preiswerte auf der Grundlage der verfügbaren Daten vorherzusagen.

Methoden der Regressionsanalyse:

Eine weitere wichtige statistische Methode ist die Regressionsanalyse. Wir erstellen Regressionsmodelle, die die Beziehung zwischen Autopreisen und verschiedenen Faktoren wie Autoleistung, Wirtschaftsleistung usw. beschreiben. Diese Modelle ermöglichen es uns, die Auswirkungen jedes Faktors auf die Preise abzuschätzen und zukünftige Werte auf der Grundlage dieser Schätzungen vorherzusagen.

Vorteile statistischer Verfahren:

Statistische Methoden bieten eine Reihe von Vorteilen, darunter die Fähigkeit, komplexe Beziehungen zwischen Variablen zu berücksichtigen, sowie die Fähigkeit, große Datenmengen zu verarbeiten. Sie ermöglichen es uns, genaue und zuverlässige Modelle zur Vorhersage von Fahrzeugpreisen zu erstellen, die für strategische Entscheidungen in der Automobilindustrie verwendet werden können.[[2]](#footnote-2)

## **2.2. Methoden des maschinellen Lernens**

Methoden des maschinellen Lernens sind ein leistungsstarkes Werkzeug zur Vorhersage von Autopreisen. Sie bieten die Möglichkeit, automatisch Muster aus riesigen Datenmengen zu extrahieren und Modelle zu erstellen, die auf der Grundlage dieser Informationen genaue Vorhersagen treffen können.

Verwenden von Algorithmen für maschinelles Lernen:

Wir wenden häufig verschiedene Algorithmen für maschinelles Lernen an, wie z. B. Random Forest, Entscheidungsbäume, lineare Regression und viele andere. Dies ermöglicht es uns, Modelle zu erstellen, die die komplexen Beziehungen zwischen verschiedenen Parametern berücksichtigen und in der Lage sind, Autopreise mit hoher Genauigkeit vorherzusagen.

**Datenaufbereitung:**

Damit Algorithmen für maschinelles Lernen erfolgreich sind, ist jedoch eine sorgfältige Datenvorverarbeitung erforderlich. Dieser Prozess umfasst mehrere Phasen:

* Datenbereinigung: Wir entfernen Ausreißer, ergänzen fehlende Werte und behandeln Fehler in den Daten.
* Datennormalisierung: Wir bringen Funktionen auf eine einzige Skala für eine effizientere Nutzung durch Algorithmen.
* Auswahl informativer Schilder: Wir wählen die wichtigsten Anzeichen aus, die sich auf die Autopreise auswirken können, um sie weiter zu analysieren.

**Modellauswahl:**

Nach der Vorverarbeitung der Daten stehen wir vor der Aufgabe, das am besten geeignete Modell für maschinelles Lernen auszuwählen, um ein bestimmtes Vorhersageproblem zu lösen. Diese Wahl hängt von mehreren Faktoren ab, darunter das Volumen und die Struktur der Daten, die Art des Problems (Regression oder Klassifizierung), die erforderliche Trainings- und Vorhersagerate und die Interpretierbarkeit des Modells.[[3]](#footnote-3)

# **Kapitel 3. Software-Implementierung**

## **3.1. Beschreibung der Merkmale**

Unser Datensatz enthält Informationen über Autos nach folgenden Kriterien:

**name**: Der Name des Automodells.

**Unternehmen:** Automobilhersteller.

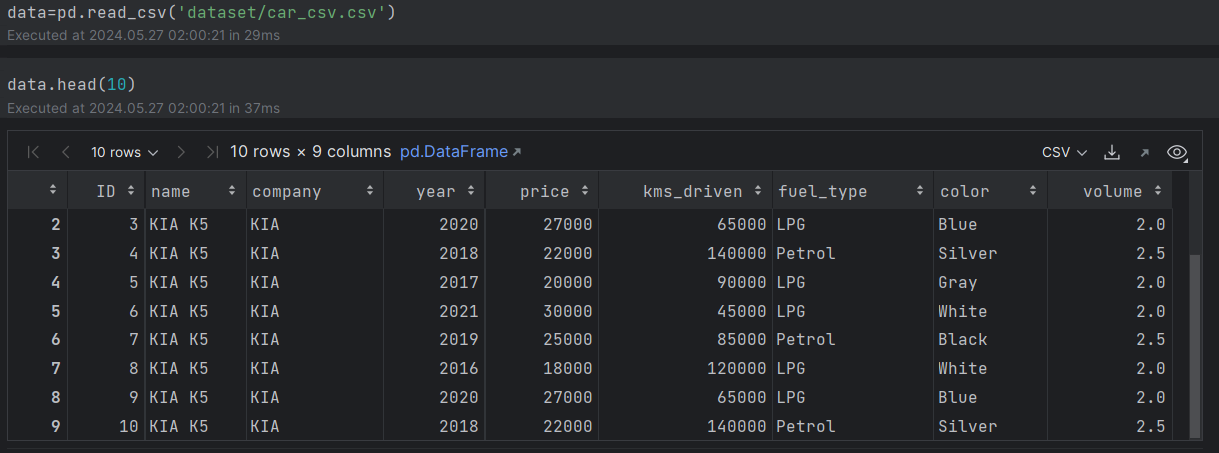
**Jahr:** Das Baujahr des Autos.

**price:** Der Preis des Autos.

**kms\_driven:** Kilometerstand des Autos in Kilometern.

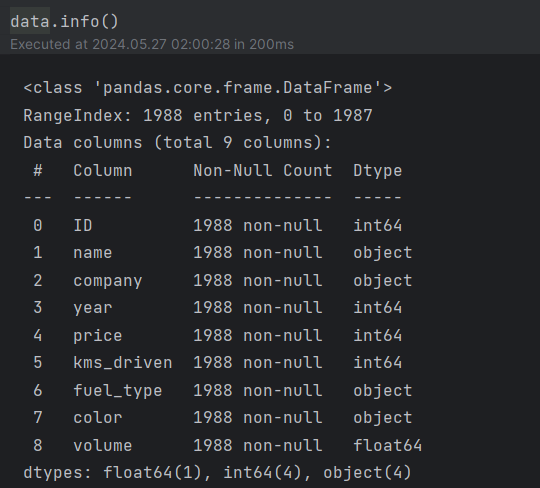
**fuel\_type**: Die Art des vom Fahrzeug verwendeten Kraftstoffs (z. B. Benzin, Diesel, Elektro).

**color:** Die Farbe des Autos. volume: Der Hubraum des Autos.

****

*Abb.1 Lesen von Dateien aus einer CSV-Datei*

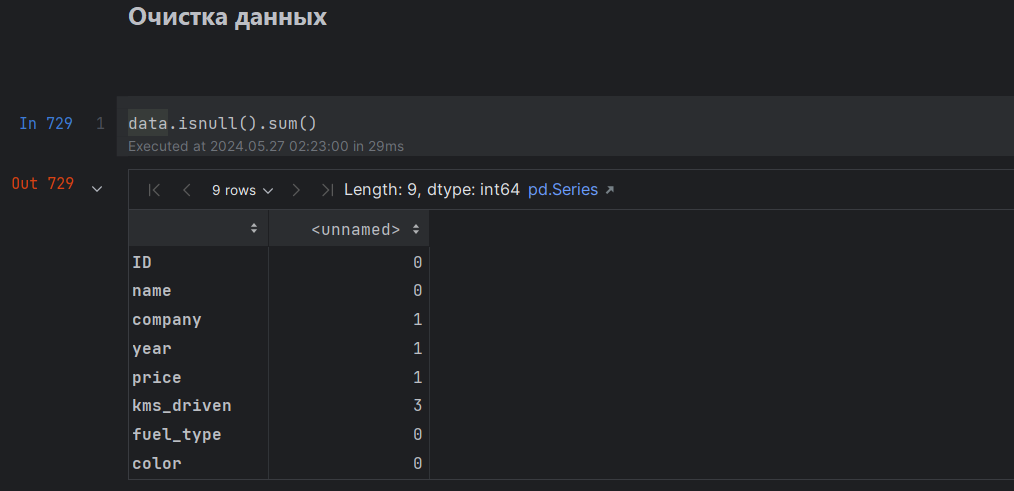
Dieser Code (Abb.1) lädt Daten aus einer CSV-Datei und gibt 10 Zeilen des Datensatzes aus.

****

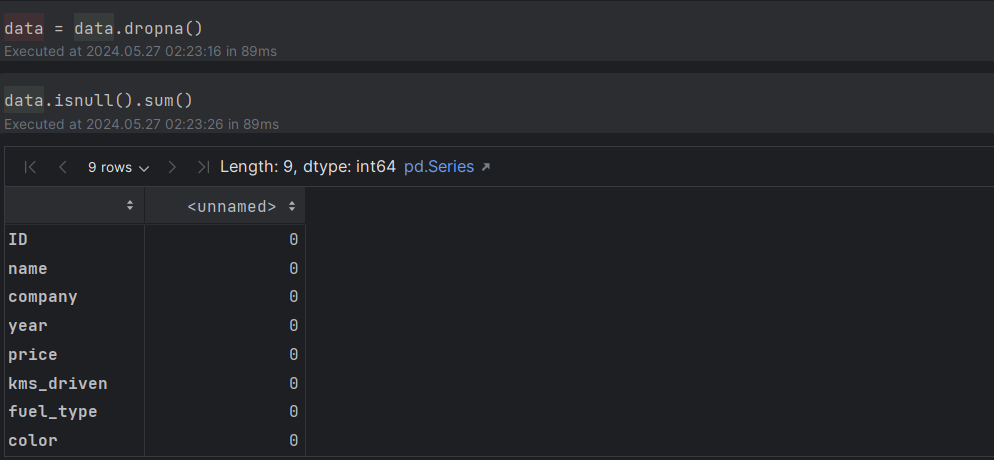
*Abb.2 Aufbau eines Datenrahmens*

Die Methode data.info() aus der Pandas-Bibliothek in Python bietet eine Zusammenfassung der Struktur des Datenrahmens

## **3.2. Datenbereinigung**

*****Abb.3 Prüfung auf fehlende Werte*

Mit diesem Code (Abb. 3) verwenden wir die Methode isnull(), um nach fehlenden Werten im Datensatz zu suchen.

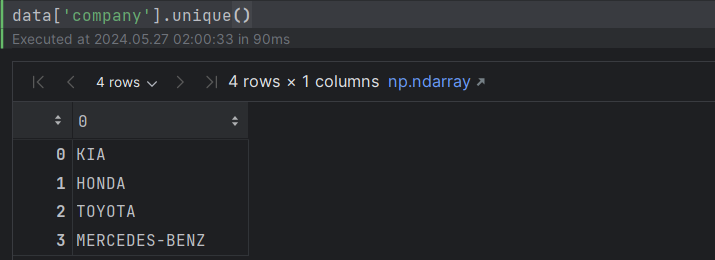
****

*Abb.4 Entfernen fehlender Werte*

Dieser Code (Abbildung 4) entfernt die fehlenden Werte mit der Methode dropna() und prüft erneut, ob die fehlenden Werte erhalten bleiben.

****

*Abb.5 Prüfung auf Duplikate*

****

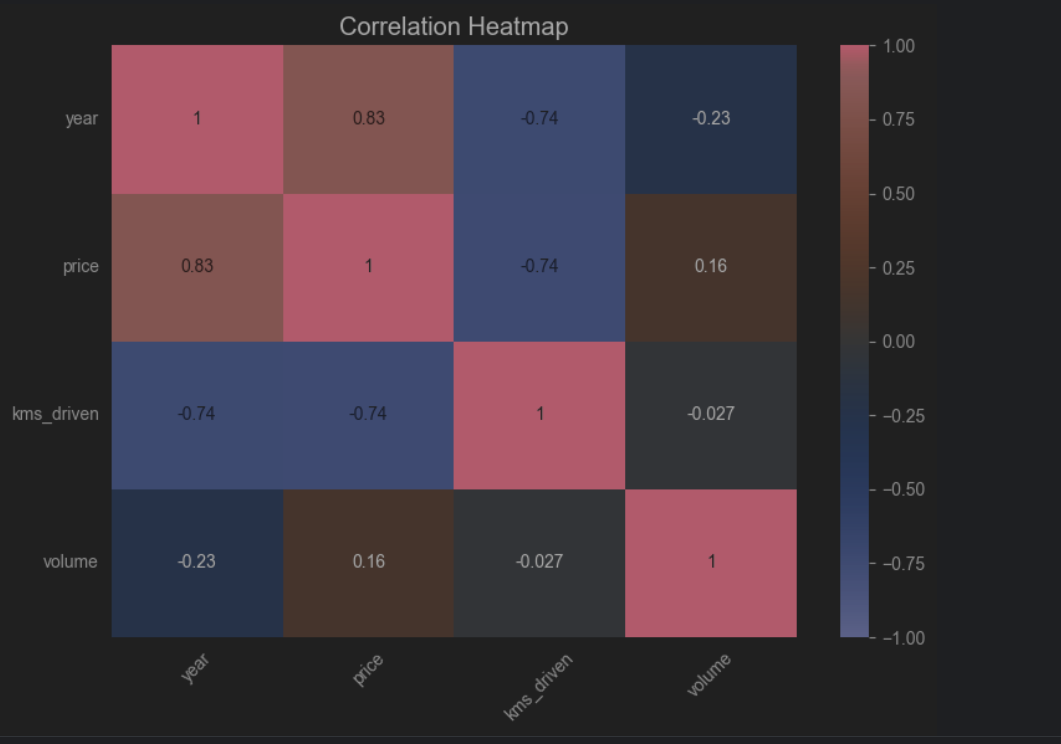
*Abb.6 Eindeutige Bedeutung*

Dieses Code-Snippet extrahiert eindeutige Werte aus der Spalte "company" im Datenrahmen.

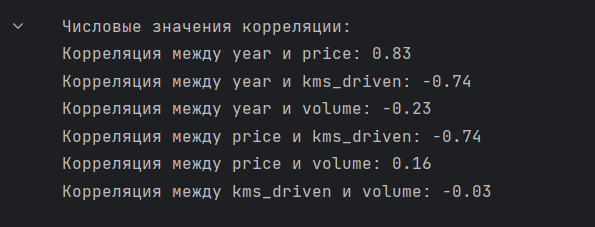
Dies ist notwendig, um alle im Datensatz vertretenen eindeutigen Automobilhersteller zu identifizieren.

Einzelwerte können nützlich sein, um die Verteilung von Autos nach Unternehmen zu analysieren

oder um weitere unternehmensbasierte Filterungen und Aggregationen durchzuführen.

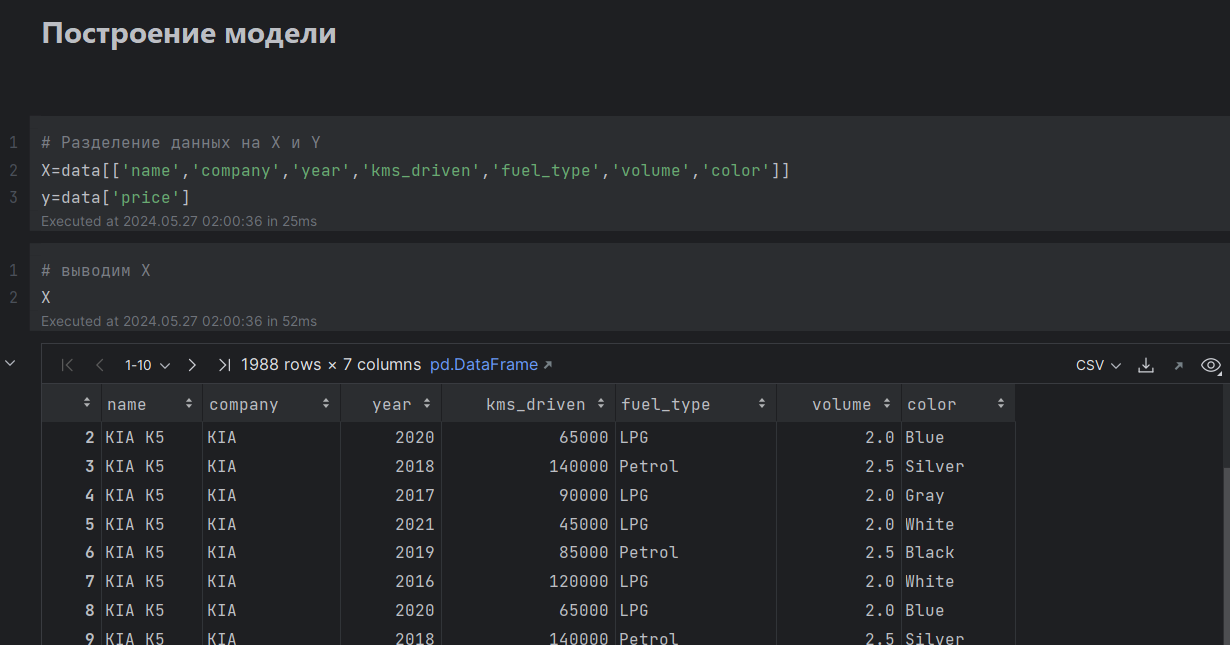
****

*Abb.7 Visualisierung der Korrelation zwischen Merkmalen*

****

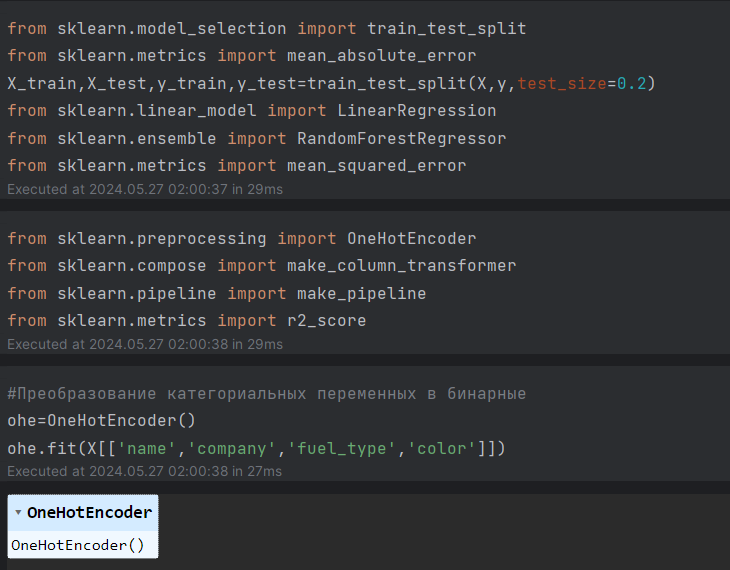
*Abb.8 Korrelation von Merkmalen*

## **3.3. Erstellen eines Modells**

****

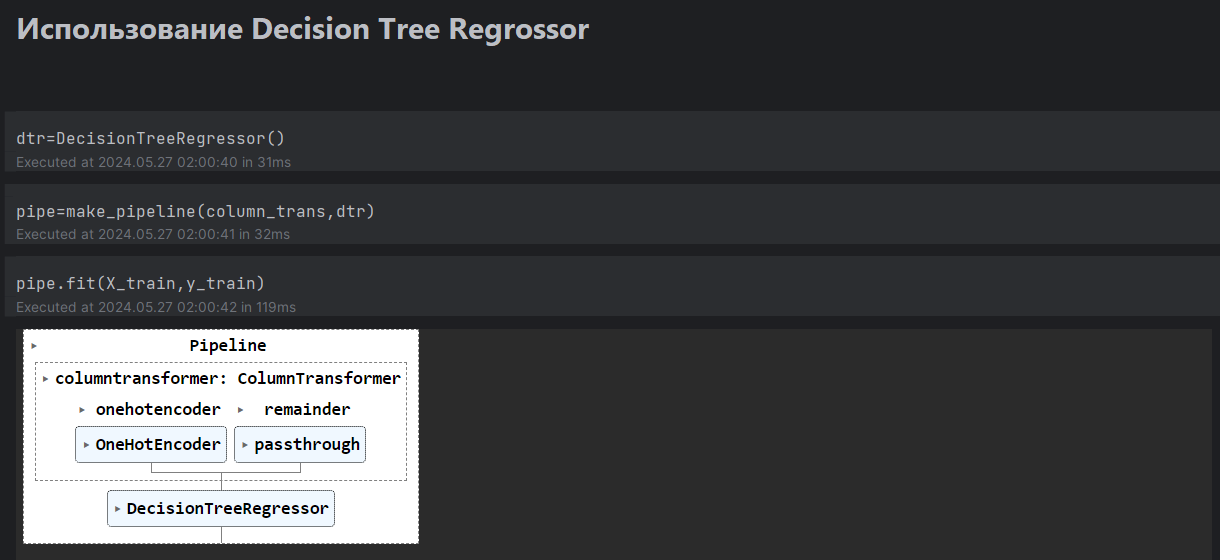
*Abb.9 Aufteilen von Daten in X und Y*

Wir unterteilen die Daten in Merkmale (X) und die Zielvariable (Y). X enthält Informationen über Modell, Hersteller, Baujahr, Kilometerstand, Kraftstoffart, Volumen und Farbe der Autos. Y enthält die Preise der Autos, die wir vorhersagen möchten.

*****Abb.10 Transformation kategorialer Daten*

Dieser Codeblock enthält Importe der erforderlichen Module und Funktionen, um die folgenden Aktionen auszuführen:

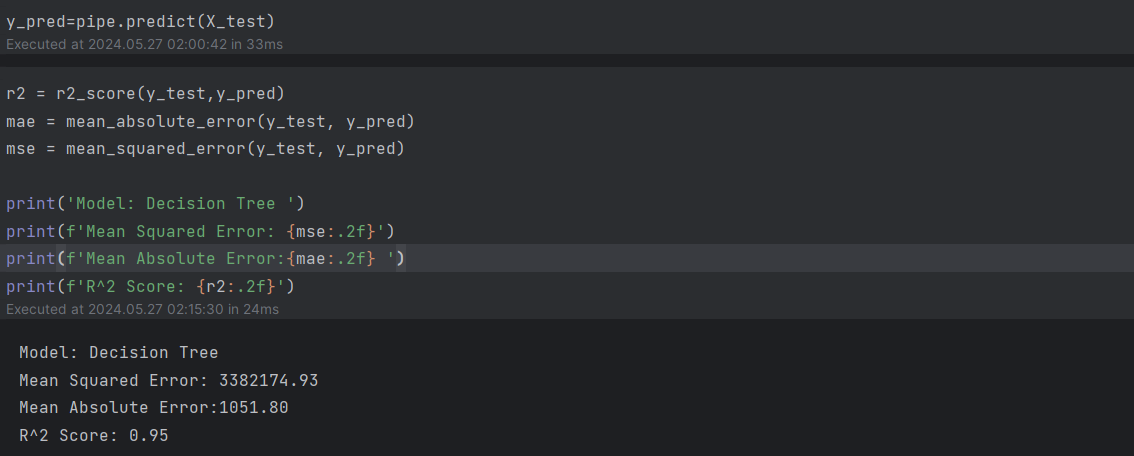
* Importieren Sie die Funktion "train\_test\_split" aus dem Modul "sklearn.model\_selection", um die Daten in Trainings- und Testsätze aufzuteilen.
* Importieren Sie die Funktion "mean\_absolute\_error" aus dem Modul "sklearn.metrics", um den durchschnittlichen absoluten Fehler zu berechnen.
* Importieren Sie die Klasse "LinearRegression" aus dem Modul "sklearn.linear\_model", um ein lineares Regressionsmodell zu erstellen.
* Importieren Sie die Klasse "Random Forest Regressor" aus dem Modul "sklearn.ensemble", um ein Random-Forest-Modell zu erstellen.
* Importieren Sie die Funktion "mean\_squared\_error" aus dem Modul "sklearn.metrics", um den mittleren quadratischen Fehler zu berechnen.
* Sowie die Umwandlung kategorialer Variablen in Binärwerte.

****

*Abb. 11 Verwenden der Entscheidungsbaumregression*

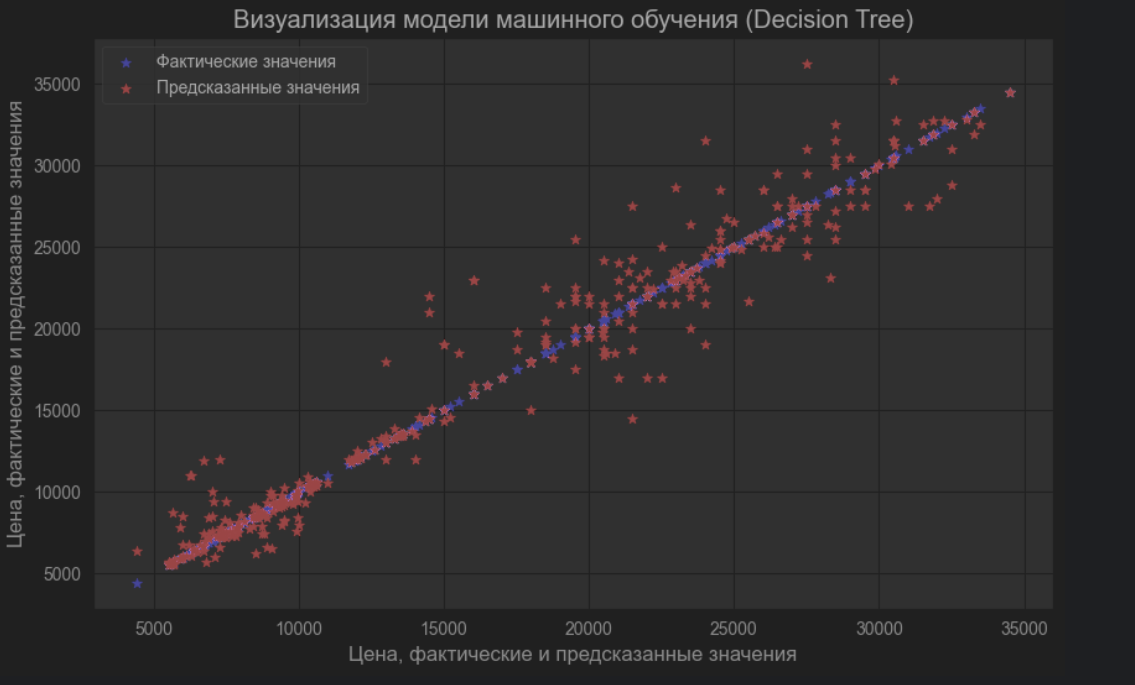
Dieser Code (Abb. 11) führt die folgenden Aktionen aus:

* Erstellt eine Instanz eines Regressionsmodells basierend auf der Entscheidungsstruktur mithilfe der Klasse "DecisionTreeRegressor".
* Importiert die Funktion "make\_pipeline" aus dem Modul "sklearn.pipeline", um eine Pipeline für die sequenzielle Datentransformation und Modellanwendung zu erstellen.
* Erstellt eine Pipeline mithilfe von Spaltentransformationen (column\_trans) und einem Entscheidungsbaum-Regressionsmodell (dtr).
* Trainiert die Pipeline mit dem Trainingsdataset ('X\_train' und 'y\_train') mit der Methode 'fit()'.

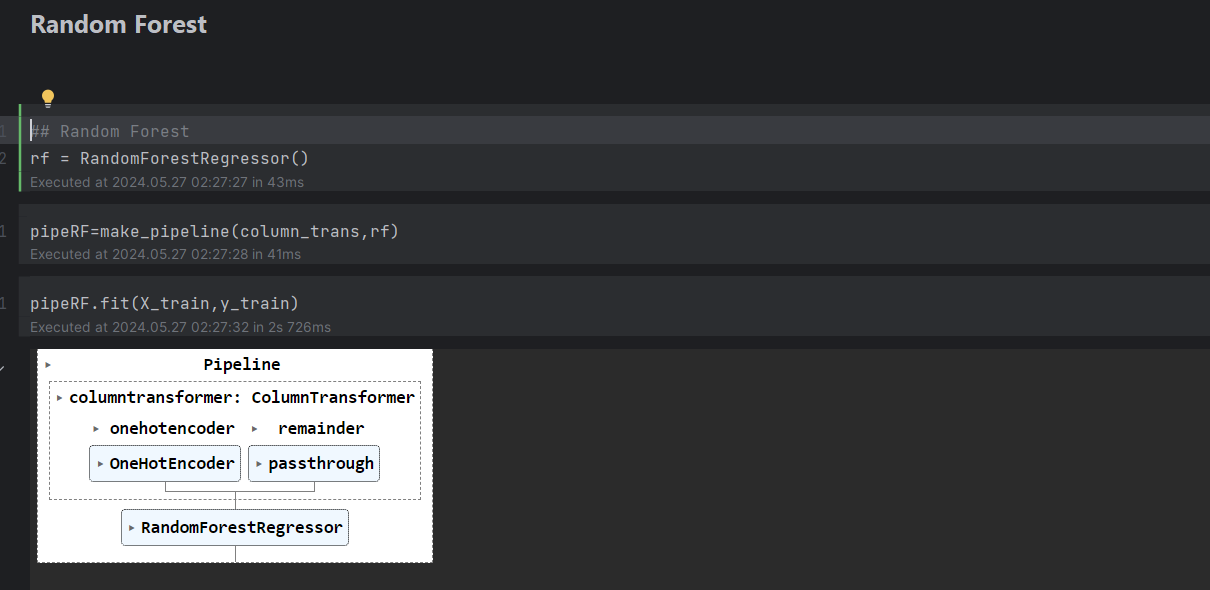
*****Abb.12 Vorhersage auf Basis von Testdaten*

Dieser Codeblock (Abb. 12) führt die folgenden Aktionen aus:

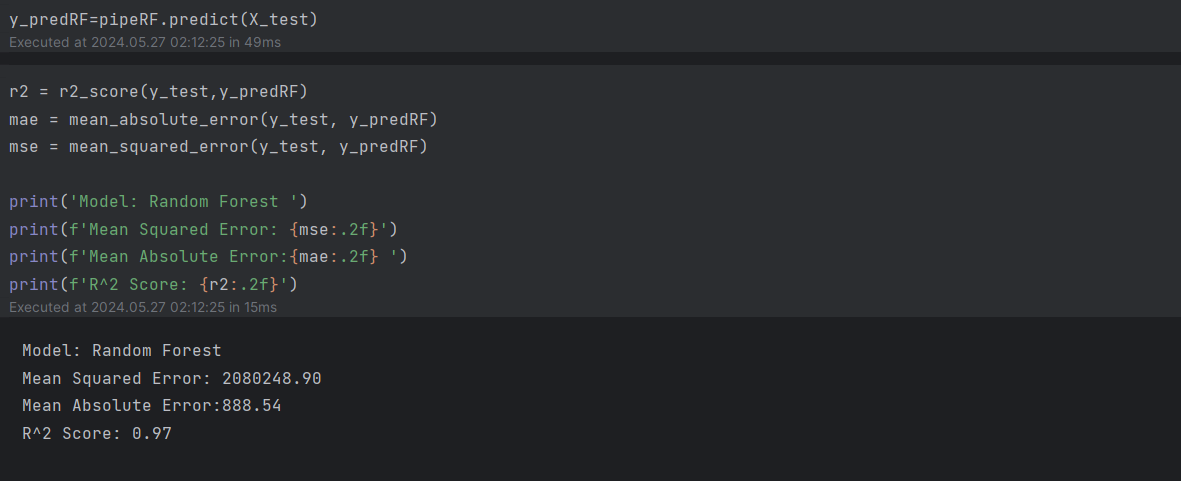
* Sagt die Zielvariable 'y\_pred' für das Testdataset 'X\_test' mit der Methode 'predict()' der 'pipe'-Pipeline voraus.
* Importiert die Funktion 'r2\_score' aus dem Modul 'sklearn.metrics', um den Bestimmtheitskoeffizienten (R-Quadrat) zu berechnen.
* Berechnet den Bestimmtheitskoeffizienten (R-Quadrat), den mittleren absoluten Fehler (MAE) und den mittleren quadratischen Fehler () mit den Funktionen "r2\_score", "mean\_absolute\_error" bzw. "mean\_squared\_error".
* Zeigt Ergebnisse wie Modellname,, MAE und R-Quadrat-Werte an.

****

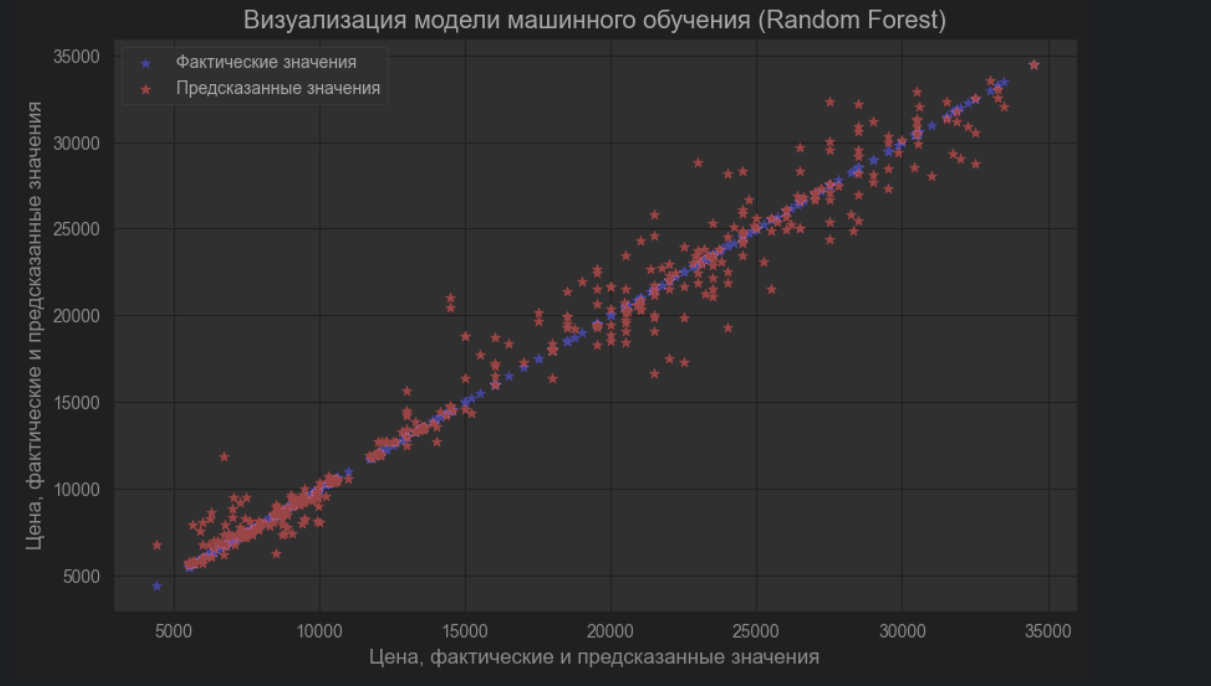
*Abb.13 Visualisierung des maschinellen Lernmodells des Entscheidungsbaums*

****

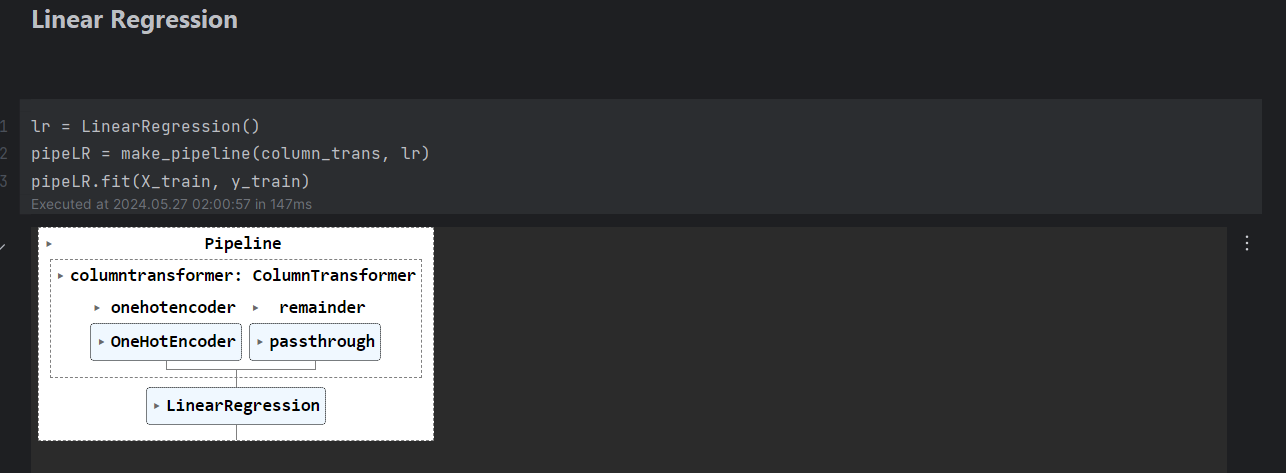
*Abb. 14 Verwenden von Random Forest*

****

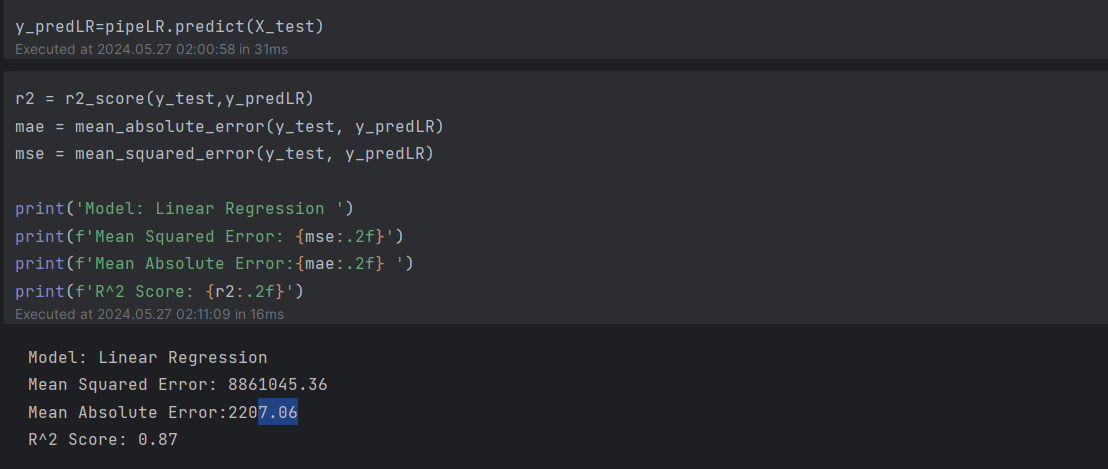
*Abb. 15 Leistung des Machine Learning-Modells*

****

*Abb.16 Visualisierung des maschinellen Lernmodells Random Forest*

****

*Abb. 17 Lineare Regression verwenden*

****

*Abb. 18 Leistung des Machine Learning-Modells*

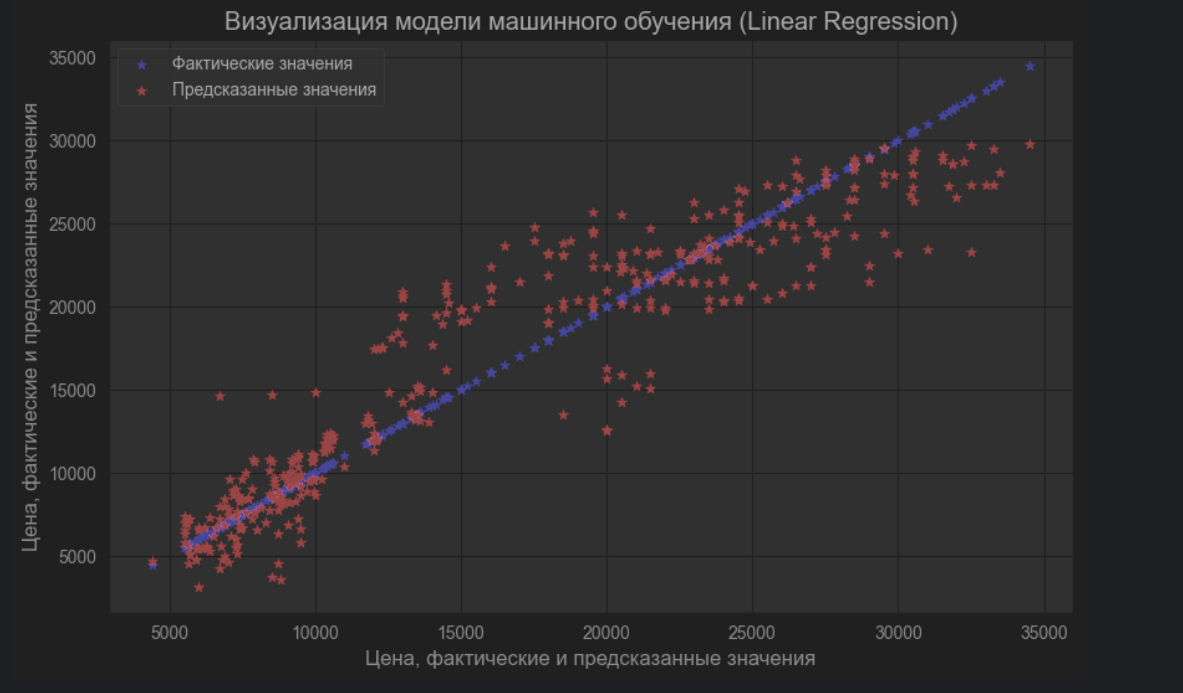
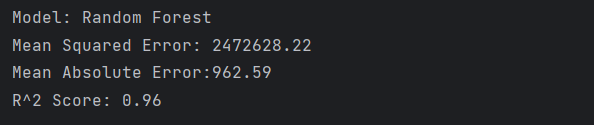
****

Abb.19 *Visualisierung des maschinellen Lernmodells der linearen Regression*

## **3.4. Auswahl des besten Modells**

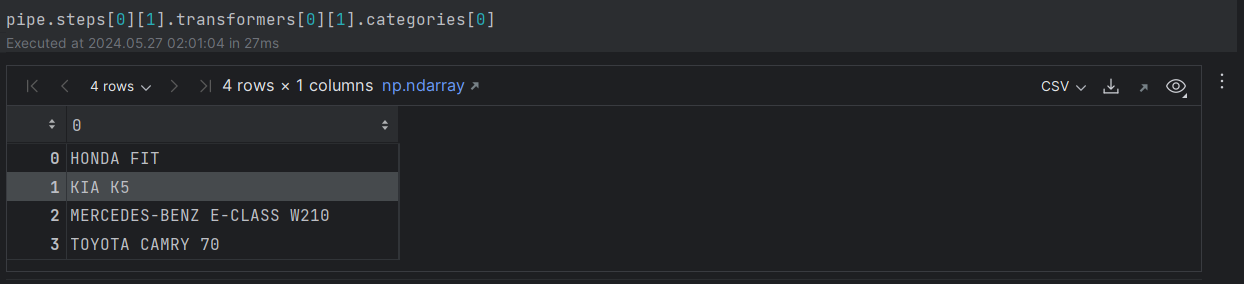
****

*Abb.20 Auswahl des besten Modells*

Nachdem wir unsere Daten mit drei Machine-Learning-Modellen trainiert hatten, zeigte Random Forest das höchste Ergebnis.

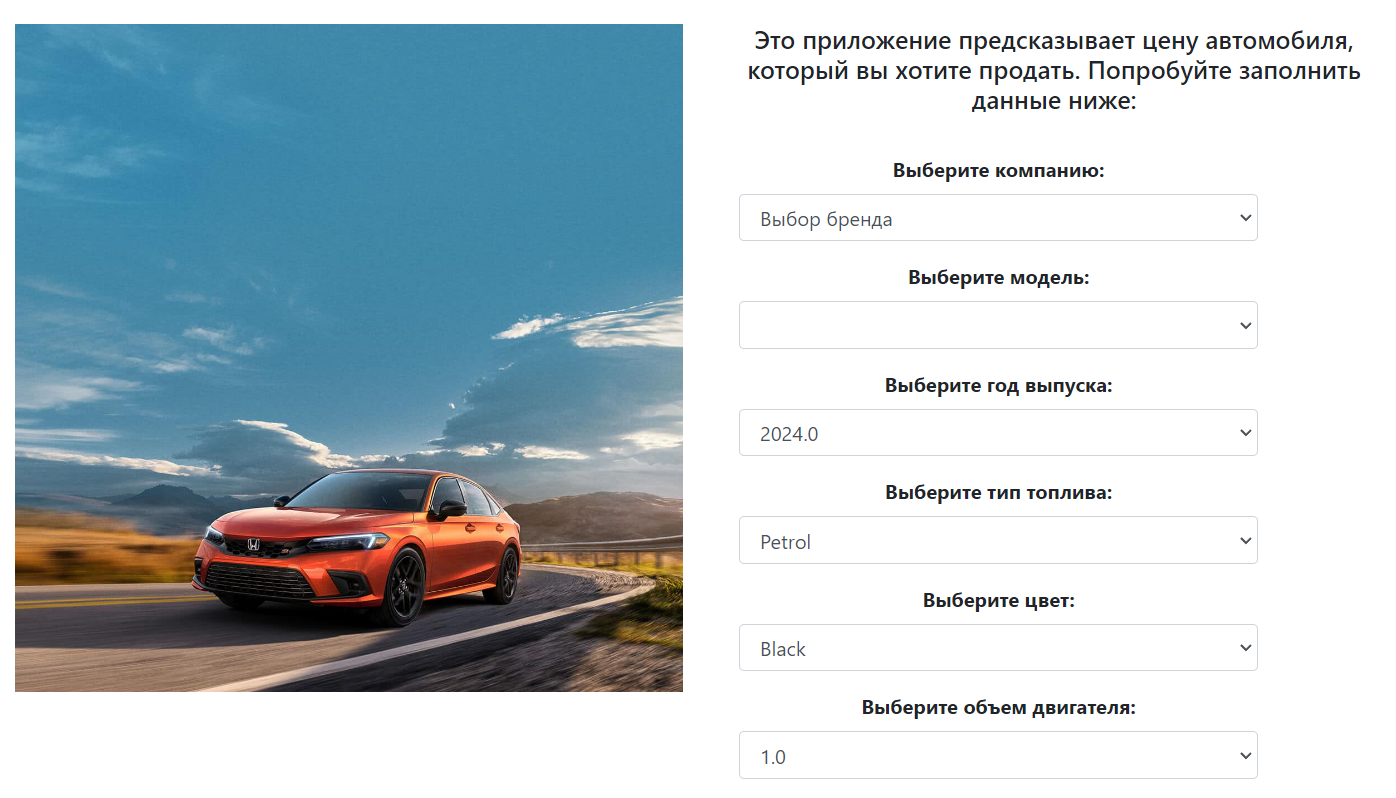
****

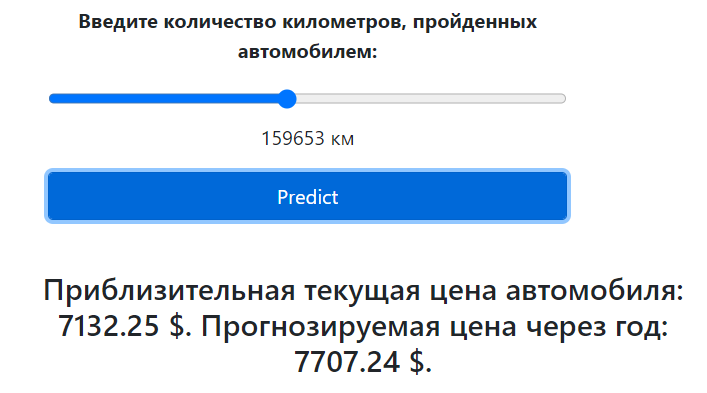
*Abb.21 Serialisierung von vorhergesagten Daten in pkl*

****

*Abb.22 Ausgabe einzigartiger Automarken*

Dieser Code (Abb.22) zeigt die Marken unserer Autos an.

****

****

*Abb.23 Projekt-Website*

Abbildung 23 ist ein Screenshot eines Programms, das verschiedene Eigenschaften eines Autos verwendet, um seinen Preis vorherzusagen. Dies ist eine gute Möglichkeit, Ihr Programm zu visualisieren und seine Funktionalität zu zeigen.

# **Schlussfolgerung**

Das Ergebnis unserer Entwicklung eines Kursprojekts zur Prognose von Autopreisen war die Entwicklung eines Modells, das in der Lage ist, die Preisdynamik auf dem Automobilmarkt genau vorherzusagen. Während der gesamten Studie durchliefen wir mehrere wichtige Phasen, darunter die Datenerfassung, -verarbeitung und -analyse sowie die Verwendung verschiedener Methoden des maschinellen Lernens zur Erstellung eines Vorhersagemodells.

Im Zuge der Arbeit wurden Daten zu verschiedenen Merkmalen von Autos wie Marke, Modell, Baujahr, Kilometerstand sowie externen preisbeeinflussenden Faktoren gesammelt. Diese Daten wurden gründlich bereinigt und aufbereitet, einschließlich der Behandlung fehlender Werte, der Normalisierung und der Codierung kategorialer Merkmale, um ihre Eignung für die Analyse und Modellierung sicherzustellen.

In der Phase der explorativen Datenanalyse (EDA) wurden verschiedene Visualisierungstechniken eingesetzt, um die Verteilung und Dynamik von Preisen zu verstehen. Wir erstellten Histogramme, Heatmaps von Korrelationen, die halfen, wichtige Muster und Faktoren zu identifizieren, die die Autopreise beeinflussen. Die Visualisierung ermöglichte es uns, ein tieferes Verständnis der Datenstruktur zu erlangen und wichtige Trends wie saisonale Schwankungen und langfristige Veränderungen zu erkennen.

Zur Vorhersage von Autopreisen wurden verschiedene Techniken des maschinellen Lernens angewendet, darunter lineare Regression, Random Forests und ein Entscheidungsbaum. Jedes Modell durchlief ein Hyperparameter-Tuning und wurde auf der Grundlage von Genauigkeitsmetriken wie dem mittleren quadratischen Fehler (), dem mittleren absoluten Fehler (MAE) und dem Bestimmtheitskoeffizienten (R²) bewertet. Dies ermöglichte es, das genaueste und stabilste Modell für die Preisprognose zu wählen.

Unser Modell berücksichtigt eine Vielzahl von Faktoren wie Marke, Modell, Baujahr, Kilometerstand, Zustand des Autos und andere Merkmale. Die Ergebnisse der Studie zeigten, dass der Einsatz moderner Technologien und analytischer Ansätze die Genauigkeit von Autopreisprognosen deutlich verbessern kann. Das entwickelte Modell weist eine hohe Genauigkeit und Stabilität auf, was es zu einem wertvollen Werkzeug für verschiedene Teilnehmer des Automobilmarktes macht.

Verbraucher können das Modell nutzen, um fundierte Entscheidungen über den Kauf oder Verkauf von Autos zu treffen und zukünftige Preistrends einzuschätzen. Autohändler können Prognosen nutzen, um Preisstrategien zu optimieren

Insgesamt trägt unser Modell zu einem transparenteren und effizienteren Funktionieren des Automobilmarktes bei, von dem alle Beteiligten profitieren.

Daher unterstreicht unsere Studie die Bedeutung der Anwendung von Datenanalyse- und maschinellen Lerntechniken bei der Vorhersage von Autopreisen. Die Ergebnisse bestätigen, dass moderne Technologien die Qualität von Prognosen deutlich verbessern können und nützliche Erkenntnisse für strategische und taktische Entscheidungen im Automobilmarkt liefern.

# **Liste der verwendeten Quellen**

Studienmaterialien:

1. Rasmussen, Karen. "Datenvisualisierung: Ein praktischer Leitfaden". Mir Verlag, 2017.

2. McKinley, Will. "Python für die Datenanalyse". O'Reilly Verlag, 2021.

3. Johnson, Emma. "Einführung in die Datenanalyse". O'Reilly Verlag, 2020.

4. Smith, John. "Maschinelles Lernen für Anfänger". Alpina Verlag, 2019.

Ressourcen im Internet:

5. Random Forest (Proglib) -<https://proglib.io/p/mashinnoe-obuchenie-dlya-nachinayushchih-algoritm-sluchaynogo-lesa-random-forest-2021-08-12>

6. Grundlagen der linearen Regression (habr) - <https://habr.com/ru/articles/514818/>

7.Was ist sklearn? (Datenstart) -<https://datastart.ru/blog/read/chto-takoe-scikit-learn-gayd-po-populyarnoy-biblioteke-python-dlya-nachinayuschih>

8. Fahrzeugdaten -<https://m.mashina.kg/>

9. Einführung in Pandas -<https://khashtamov.com/ru/pandas-introduction/>

10.Numpy Was ist das für eine Bibliothek -<https://blog.skillfactory.ru/glossary/numpy/>

1. Die Auswirkungen von Umweltvorschriften auf Verbraucherpräferenzen", Smith J., 2022, Umweltvorschriften und Verbraucherpräferenzen [↑](#footnote-ref-1)
2. *Johnson, Emma. "Einführung in die Datenanalyse". O'Reilly Verlag, 2020* [↑](#footnote-ref-2)
3. *Smith, John. "Maschinelles Lernen für Anfänger". Alpina Verlag, 2019* [↑](#footnote-ref-3)