

# 中古車価格のプレミア性に対する新車購入時仕様が与える影響分析

2019SC036 水谷風稀

指導教員：河野浩之

## 1 中古車市場規模の拡大

2021年の中古車市場の規模は過去最大であった。主な要因は新型コロナウイルスの世界的流行である。特に半導体不足で自動車生産に大きな打撃を与えたことで新車納期が長期化し、すぐに納車可能な中古車の需要が増すことで本来より割高に取引されていることが市場規模拡大の要因のひとつである。本研究では中古車価格予測モデルを構築し、特徴量重要度を出力し中古車価格に影響を与える仕様について妥当性を検証する。

## 2 中古品または中古車価格予測の先行研究

中古車価格予測を行った先行研究について次の表1に示す。工藤ら [1] は日本オートオークション 78 会場で行われた中古車取引の落札価格データを用い中古車価格の割安評価を行った。線形的なモデルであるヘドニック法を用いた分析は不動産価格の予測に用いられていたが、非線形性への対策から分析を行った。成果として決定係数の向上、中古車価格には地域性や季節性があることを示した。また龍田ら [3] は3つのGBDTモデルの中からLightGBMの計算速度が最も早いことを示し、SHAPを用いて説明変数の影響度の可視化をおこなった。

表1 中古車または中古品価格予測に関する先行研究

研究者	手法・成果
工藤ら [1]	落札データで中古車価格割安評価 手法：XGBoost, SHAP 値ボジネガ 成果：高価格帯で安定的価格推定
龍田ら [3]	GBDTを用いた橋梁損傷原因の影響分析 手法：XGBoost, LightGBM, Catboost 成果：モデル別精度・時間、説明変数の影響分析

## 3 価格予測および特徴量の影響分析のための分析手順・提案手法

### 3.1 中古車価格の影響分析における分析手順

本研究の分析手順としてフローチャートを図1へ示す。工藤ら [1] ではXGBoostを用いていたが本研究ではその後発モデルであるLightGBMを用いて精度の向上を狙う。また工藤らはオークション落札データを用いており一般的に我々は観ることができないデータである。そこで本研究では初めにBeautiful Soupで閲覧制限のないインターネットの中古車サイトからデータをスクレイピングする。次に前処理を行いLightGBMによる中古車価格予測モデルを構築し、特徴量重要度をもとに中古車価格に影響の

ある特徴量を選定して新車購入時に設定するべき仕様とする。

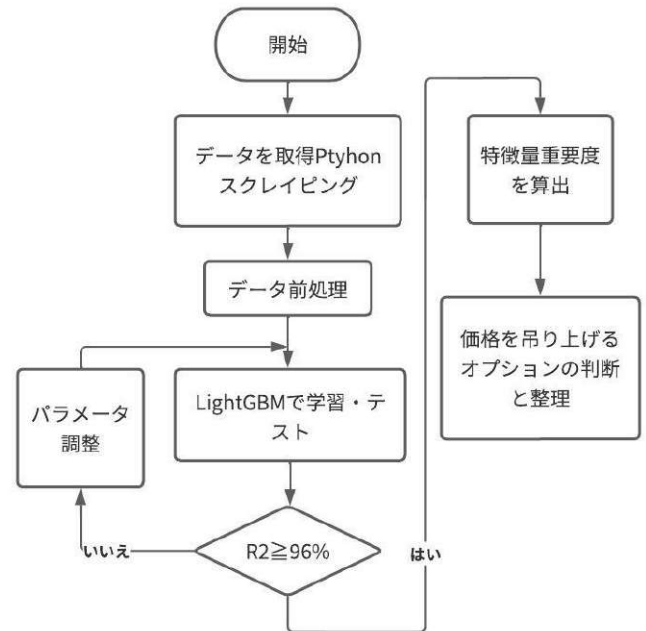


図1 分析フロー

### 3.2 WEBスクレイピングからデータ前処理と価格予測モデルの構築

実験に使用するデータの収集をスクレイピングツールを用いておこなう。条件検索の簡便さや車体色表記、車検日表記などの統一性からデータのの前処理が容易となるため本研究では価格ドットコムを用いる。

取得したcsvファイルをPythonを用いて車体色、エンジンサイズなどはOne-hot表現に変更しグレード情報の抽出などの処理を行う。XGBoostの後継であるLightGBMを用いて中古車価格予測モデルを構築し、特徴量重要度から価格に与える影響の大きい特徴量を特定する。

## 4 実データを用いた中古車価格予測と影響分析

### 4.1 実験環境とデータ準備

実験環境を次の表2へ示す。データの取得はBeautiful Soupを用いる。前節の図1で提示した分析手順に沿って実験を行う。まずは使用するデータの準備である。本研究は価格ドットコムの中古車サイトよりスクレイピングソフトを用いて任意のデータを取得する。

表 2 本研究の実験環境

項目	環境
スクレイピングツール	BeautifulSoup
プログラムツール	Jupyter Notebook
使用言語	Python3
データセット	価格ドットコム
アプリケーション	Excel
ライブラリ	pandas numpy seaborn requests optuna

#### 4.2 データの前処理と LightGBM を用いた中古車価格予測モデルの構築

取得した 2 車種のデータに前処理を行う。名称列からグレード情報を取得し新たな特徴量として加える。色表記のずれについても修正をおこない整数値で表した。

LightGBM を用いた中古車価格予測モデルにおいてパラメータチューニングを行った結果を次の表 3 へ示す

表 3 パラメータ最適化後の LightGBM の結果

車種	RSME	r2_train	MAE_train
CX-5	15.73	0.94	8.39
ロードスター	24.64	0.88	12.01
	r2_valid	MAE_valid	
CX-5	0.88	11.67	
ロードスター	0.79	17.82	

CX-5 の決定係数が 0.8833 と一番高い結果となったが先行研究では XGBoost で 0.96 の決定係数を出しており、それを上回ることができなかった。

#### 4.3 中古車価格への影響分析

中古車価格予測モデルの特徴量重要度を算出し都合上省略したものを次の表 4 へ示す。表 4 は CX-5 のグレードに

表 4 グレード情報についての特徴量重要度

特徴量名	特徴量重要度
エクスクルーシブモード	1002179
スマートエディション	190220
プロアクティブ	13466

ついでの特徴量重要度を出力した様子で、エクスクルーシブモードが一番影響度がある。これらを車両情報、車体色、オプション装備品、グレードについて出力し次の表 5 のよ

うに CX-5 の仕様として示す。

表 5 特徴量重要度をもとにした最適な仕様

項目	仕様
車体色	ロジウムホワイト プレミアムメタリック
グレード	XD Exclusive Mode
駆動方式	2WD AT
OP 装備	電動ガラスサンルーフ ナビ用 SD カード ETC

表 5 は特徴量重要度をもとにして選択した仕様である。この仕様の新車時の車両額は約 419 万円であり、2022 年式とし LightGBM による価格予測モデルで同様の条件にて予測を行った結果約 382 万円となり実際の中古車に近い金額を算出することができ、寝落ち幅も低い様子が見れる。車体色は 2 車種とも一般的に寝落ちの少ない白・黒色系とブランドの代表色であるソウルレッドが上位に位置し、市場と同様な結果から妥当性があると感じる。

## 5 結び

本研究では先行研究で用いた XGBoost より計算量の少ない LightGBM で中古車価格予測モデルを構築したが精度を上回ることができなかった。また工藤ら [1] が用いたデータとの違いや中古車の価格の決定方法は販売店によって差があり、そのすべての属性を予測に使用することができなかった。

しかし龍田ら [3] のモデル別の計算時間から LightGBM は XGBoost を改良し計算量が減少、学習速度が向上している。精度の面で上回ることができなかったが LightGBM を用いることで中古車価格の予測においても大きなデータを持ちいることや、比較的性能の低い CPU などでも扱えることが期待される。

## 参考文献

- [1] 工藤大輝, 福西亮介, 黛広樹, 鈴木智也, “機械学習による中古車価格の要因分析および割安評価,” 情報処理学会論文誌, Vol.14, No.3, pp.39-50, Aug.2021.
- [2] Muhammad Asghar, “Used Cars Price Prediction using Machine Learning with Optimal Features,” Pakistan Journal of Engineering and Technology, vol.4, No.2, pp.113-119, Jun.2021.
- [3] 龍田 齊, 横山 広, 永見 武司, 榎谷 浩, 近田 康夫, 山田 宗明, “勾配ブースティング決定木を用いた橋梁損傷原因および補修工法の推定と分析,” AI・データサイエンス論文集, vol.J1, No.1, pp63-70, Jul.2020.