

BIP Machine Learning for Data Science

-Documentation Machine Learning for Data Science-

group b - ML4DS

January 6, 2025

Felipe lastname (MNr.,degree, Uni), Guilherme lastname (MNr.,degree, Uni), Jost Nickel (5196032, DSI7, HSB), Néstor lastname (MNr.,degree, Uni)

Hochschule Bremen | Faculty 4: School of Electrical Engineering and Computer Science

Examiner: Prof. Dr. Uta Bohnebeck



HSB

Hochschule Bremen
City University
of Applied Sciences

Contents

List of Tables	2
Abbreviations	2
1 Introduction	3
2 Problem Description	3
3 Data Description	3
4 Methodology	3
5 Results and Discussion	3
6 Challenges and Limitations	3
7 Conclusions and Future Work	3
References	3

List of Tables

Abbreviations

EDA	exploratory data analysis
RMSE	Root Mean Squared Error

1 Introduction

The automotive industry is highly competitive, and car dealerships face significant challenges in optimizing their operations to maintain profitability and customer satisfaction. One critical aspect of dealership efficiency is minimizing the time vehicles spend unsold at the dealership, referred to as the 'laid-up time'. A long laid-up time can lead to increased inventory costs, reduced profitability, and inefficiencies in dealership operations. This project aims to address this challenge by developing a predictive model to estimate the laid-up time of vehicles at the time of purchase. By leveraging historical sales data and advanced machine learning techniques, the project not only seeks to provide accurate predictions but also to identify key features that influence laid-up time. Additionally, it aims to offer actionable insights into which vehicle specifications are associated with shorter laid-up times, thereby assisting dealerships in making data-driven purchasing decisions.

The dataset used in this study, provided by Emil Frey, contains over 140,000 records of vehicle sales spanning a 10-year period, with more than 100 features. A training dataset of approximately 100,000 records includes the laid-up time, while a test dataset of 40,000 records lacks this target variable and is used for evaluation purposes. The performance of the model will be assessed using the Root Mean Squared Error (RMSE) metric, ensuring a robust evaluation of predictive accuracy.

Predictive modeling has become a crucial tool in the automotive industry, helping businesses make informed decisions and streamline operations. As noted by Doe et al. (2023), data-driven approaches provide valuable insights for inventory management and demand forecasting [1].

This report presents a comprehensive end-to-end approach, from data preprocessing and exploratory data analysis (EDA) to model development, validation, and deployment. Additionally, it explores the key attributes driving laid-up time and provides recommendations for dealership inventory optimization. The outcomes of this project are expected to support dealerships in improving operational efficiency and maximizing profitability through predictive insights and data-driven strategies.

2 Problem Description

3 Data Description

4 Methodology

5 Results and Discussion

6 Challenges and Limitations

7 Conclusions and Future Work

References

- [1] Doe, John and Smith, Jane. 'Predictive Modeling in Automotive Industry: Applications and Insights'. In: *Journal of Machine Learning Applications* 15.3 (2023), pp. 45–62.