MARMARA UNIVERSITY
FACULTY OF ENGINEERING

CSE4062 – S25

Data Science Project Final Delivery Contribution Report

Car Price Prediction System

Group 8:

Muhammed Furkan Kahyaoğlu (ME) – 150420058, furkankahyaoglu@marun.edu.tr

Özlem Demirtaş (IE) – 150320006, demirtasozlem444@gmail.com

Niyazi Ozan Ateş (CSE) – 150121991, niyaziozanates@gmail.com

Doğukan Onmaz (CSE) – 150120071, dogukanonmaz@marun.edu.tr

Şükrü Can Mayda (CSE) – 150120031, canmayda@marun.edu.tr

Instructor: Dr. Murat Can Ganiz

Muhammed Furkan Kahyaoğlu:

- Initiated and managed the GitHub repository to ensure smooth collaboration and version control from the very beginning.
- Conducted a thorough research to identify and select the most suitable dataset for the project objectives.
- Provided consistent and meaningful input to various reports and documentation throughout the project process
- Took a role in designing and preparing the final project presentation to effectively communicate our findings.

Özlem Demirtaş:

- Participated in the initial evaluation of sample datasets and contributed to selecting the most suitable one for the project's goals.
- Supported the team during coding phases by contributing to idea exchange and providing feedback when needed.
- Contributed to the writing and editing of project reports, ensuring clarity and consistency in content.
- Assisted in the preparation of the final presentation, including content organization and visual alignment.
- Took initiative in maintaining team communication and cohesion, helping the team stay organized and aligned throughout the project.
- Kanban created for the follow up.

Niyazi Ozan Ateş:

- Help to find and analyze the dataset. Checking what the data looks like by analyzing the types.
- As the preprocessing steps I have dropped the id, offer_description, and registration_date columns because they are irrelevant, and some other columns look very similar. By doing this I have reduced the training time without affecting the data much. Filled in the empty data for the column color with unknown. Truncated some columns from strings to strings that contain only numbers. After that I have converted them to numeric values. Here I replace some empty data and NaN data for these columns with their mean and median values. Some columns that are still in the form of string are categorized by their top 5 most common strings and the rest are set to others. After that I converted them to numeric values by encoding them.
- Plotted price distributions, created a scatter plot for price, and a boxplot. I also have created a correlation matrix. Checked the z-score.
- For clustering I worked on the elbow method for k-means. Also, I have plotted hierarchical clustering. Printed the support, confidence, lift, Jaccard, etc.
- For classification, I focused on Naïve Bayes, K-NN, and Random Forest. Here for each model, I tried to predict the Price, Year, and Milage. Resulting in 9 predictions and 3 models. While doing this I calculated the mutual information, ANOVA F-value, and RF importance. As results I calculated the accuracy, F1-macro, F1-micro and AUC. I plotted the ROC and Confusion Matrix for each model and prediction. At last, I calculated the t-test.
- I have contributed to writing all reports.

Doğukan Onmaz:

- Led the regression analysis efforts by implementing and comparing Artificial Neural Network (ANN) and Support Vector Regression (SVR) models for predicting price_in_euro.
- Applied preprocessing techniques including logarithmic transformation and dataset sampling to improve model performance and training efficiency.
- Evaluated model performances using MAE, MSE, MAPE, and R^2 metrics; carried out a detailed interpretation of each metric to assess model quality.
- Conducted a paired t-test on cross-validated MAE results to assess the statistical significance of the performance difference between ANN and SVR.
- Took part in the DBSCAN clustering process by selecting optimal parameters (eps and min_samples), interpreting clustering output including noise points, and summarizing cluster characteristics.
- Helped design and interpret Association Rule Mining (ARM) results; reviewed key association rules by analyzing support, confidence, and lift values with respect to real-world relevance.
- Contributed to organizing the final presentation content, ensuring that statistical insights and model outcomes were clearly communicated.
- Supported report writing by contributing sections on regression, clustering, and statistical evaluation, maintaining clarity and consistency throughout.

Şükrü Can Mayda:

- Searched and reviewed various sources and helped to choose the most suitable dataset for the project.
- Explained dataset statistics with showing feature distributions using histograms, boxplots, and scatter plots; also with pure dataset values without changing in the beginning.
- On descriptive analysis, applied 3 clustering methods such that: z-score normalization, determined optimal K via the elbow method, and performed both K-Means and hierarchical clustering, also made another apriori method implementation to ensure a clean analysis.
- Implemented and compared key metrics and patterns in the data through summary tables and charts.
- Identified the most informative variables with feature selection using mutual information, Lasso and Information gain.
- Helped with data preprocessing steps and saved the preprocessed dataset to a new CSV file.
- On predictive analysis, ran Naive Bayes, K-NN, and Random Forest models to predict Price, Year, and Mileage; compared accuracy, F1-score, and AUC.
- Contributed to all reports.