****

**Kaunas University of Technology**

Faculty of Informatics

**Lab work 1. Input & output analysis. Decision tree. Random forest. KNN REPORT**

|  |
| --- |
|  |
| **ADNANE HROU**  Project author |
|  |
| **assoc. prof. dr. Dalia Čalnerytė**  Supervisor |
|  |

**2024**

**Lab work 1. Input & output analysis. Decision tree. Random forest. KNN**

**Name and family name: ADNANE HROU**

**Lab Work 1: Real Estate Price Prediction**

**Introduction:** This lab focused on predicting real estate prices using historical transaction data. The main goal was to analyze and preprocess data, apply multiple machine learning models (KNN, Decision Tree, Random Forest, and ANN), and evaluate their performance using metrics such as MAE and MAPE.

**Data Analysis and Preprocessing:** The dataset contained various features like the number of bedrooms, bathrooms, and property size. Initial steps included cleaning the data by dropping columns like address and pets\_allowed, which had over 60% missing values, and filling missing amenities with "Unknown". Continuous columns, such as price and square\_feet, were standardized, and outliers were handled through IQR capping. One-hot encoding was applied to the categorical feature cityname.

**Model Implementation:**

1. K-Nearest Neighbors (KNN): KNN was run with values of kkk ranging from 3 to 11. The best result was obtained with k=3k=3k=3, yielding an MAE of 0.3460 and R² of 0.731.
2. Decision Tree: The Decision Tree was tested with various max\_depth and min\_samples\_leaf configurations. The optimal result was achieved with max\_depth=10 and min\_samples\_leaf=2, giving an MAE of 0.3903 and R² of 0.7204.
3. Random Forest: Random Forest performed best with 150 estimators and a maximum depth of 15, achieving the lowest MAE of 0.2904 and the highest R² of 0.827.
4. Artificial Neural Networks (ANN): Different architectures were tested. The best-performing configuration was [64, 32], resulting in an MAE of 0.4164 and R² of 0.668.

**Hyperparameter Optimization:** All models were tuned with various hyperparameters. Random Forest, with 150 trees and max depth of 15, provided the best balance between error and model complexity, making it the top-performing model overall.

**Benchmark Comparison:** A simple average prediction was used as a benchmark, yielding an MAE of 0.812 and R² close to zero. The machine learning models, particularly Random Forest, significantly outperformed the benchmark.

**Conclusion:** Random Forest proved to be the most effective model for real estate price prediction, providing the most accurate results. Future work could include further tuning of the ANN model and experimenting with additional features to improve performance.