## 1. Abstract

Accurate housing price prediction is a critical tool for real estate investors, urban planners, policy makers, and financial institutions (Chen et al., 2017; Glaeser and Nathanson, 2017). This project explores the use of machine learning models to forecast Q3 2024 housing prices across European Union (EU) regions, using a structured dataset comprising quarterly housing indices, macroeconomic indicators, and demographic attributes. The aim is to evaluate and compare various regression models to identify the most accurate and generalizable approach for predicting property values.

The dataset includes housing price indices from Q1 2020 to Q3 2024, along with relevant features such as GDP growth rate, average income, unemployment rate, population growth, and climate zone (Eurostat, 2023). After extensive exploratory data analysis (EDA), several engineered features were introduced to enhance predictive power, including interaction terms and ratio-based indicators. Six machine learning models were implemented: Linear Regression, Random Forest, Gradient Boosting, XGBoost, Support Vector Regression (SVR), and K-Nearest Neighbors (KNN) (Pedregosa et al., 2011; Chen and Guestrin, 2016).

Hyperparameter tuning and cross-validation were applied to optimize model performance, with metrics such as R², Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) used for evaluation. Results show that KNN and SVR outperformed traditional models, achieving R² scores of 0.9970 and 0.9932 respectively, with minimal prediction error. Feature importance analysis highlighted the impact of income, urbanization, and past housing trends on pricing (Lundberg and Lee, 2017).

The findings demonstrate that incorporating both economic indicators and historical price trends significantly improves model accuracy, and that non-linear models offer superior performance in this predictive context.

## 2. Introduction

The real estate sector plays a vital role in economic development, financial markets, and social well-being. Understanding and forecasting housing prices has become increasingly important in the context of market volatility, urban expansion, and post-pandemic economic shifts across the European Union (Gyourko et al., 2021; European Commission, 2022). Accurate prediction of housing prices supports informed decision-making for investors, developers, policy makers, and homebuyers alike. With the growing availability of structured data and the evolution of predictive modeling techniques, machine learning (ML) offers a powerful solution for addressing this complex challenge (Kaufmann and Steinmetz, 2020).

Traditional statistical models, such as linear regression, have been widely used in housing price prediction but often fall short in capturing non-linear relationships and interactions among features (Selim, 2009). Housing prices are influenced by a broad set of factors, including economic indicators (GDP growth, unemployment), demographic trends (urbanization, population growth), and regional attributes such as climate zone (Eurostat, 2023). Moreover, temporal trends embedded in historical housing prices provide additional predictive value, which can be overlooked by simplistic models.