**Documentation**

**README.md**

|  |
| --- |
| #Project 1-House Price Prediction  ## Overview   * The main objective of this project (e.g., House Price Prediction based on historical data). * A brief explanation of the approaches used (Random Forest, Multiple Linear Rergression, Laso, Ridge, Adaboost, and XGBoost).   ## Dataset   * Data source. * Description of key features ('MSSubClass', 'MSZoning', 'LotFrontage', 'LotArea', 'Street', 'Alley', 'LotShape', 'LandContour', 'Utilities', 'LotConfig', 'LandSlope', 'Neighborhood', 'Condition1', 'Condition2', 'BldgType', 'HouseStyle', 'OverallQual', 'OverallCond', 'YearBuilt', 'YearRemodAdd', 'RoofStyle', 'RoofMatl', 'Exterior1st', 'Exterior2nd', 'MasVnrType', 'MasVnrArea', 'ExterQual', 'ExterCond', 'Foundation', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinSF1', 'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'TotalBsmtSF', 'Heating', 'HeatingQC', 'CentralAir', 'Electrical', '1stFlrSF', '2ndFlrSF', 'LowQualFinSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'FullBath', 'HalfBath', 'BedroomAbvGr', 'KitchenAbvGr', 'KitchenQual', 'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'FireplaceQu', 'GarageType', 'GarageYrBlt', 'GarageFinish', 'GarageCars', 'GarageArea', 'GarageQual', 'GarageCond', 'PavedDrive', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch', '3SsnPorch', 'ScreenPorch', 'PoolArea', 'PoolQC', 'Fence', 'MiscFeature', 'MiscVal', 'MoSold', 'YrSold', 'SaleType', 'SaleCondition', 'SalePrice').   ## Tools & Technologies Used   * **Environment:** Google Colab / Jupyter Notebook / Local Python Environment. * **Modeling Tools:** * Random Forest using from sklearn.ensemble import RandomForestRegressor * Multiple Linear Rergression using from sklearn.linear\_model import LinearRegression * Laso using from sklearn.linear\_model import Lasso * Ridge using from sklearn.linear\_model import Ridge * AdaBoost using from sklearn.ensemble import AdaBoostRegressor * XGBoost using from xgboost import XGBRegressor   ## Data Preprocessing   * Feature normalization / standardization. * Encode variable * Splitting data into train and test sets.   ## Model Development   * Random Forest * Multiple Linear Rergression * Laso * Ridge * AdaBoost * XGBoost   ## Model Evaluation & Comparison   * Evaluation metrics: MAE, R2, MSE, RMSE, MAPE * R-squared (R2): Nilai R2 yang lebih tinggi menunjukkan model yang lebih baik dalam menyesuaikan data. Kita menginginkan model dengan nilai R2 yang tinggi pada set pelatihan dan pengujian, yang menunjukkan generalisasi yang baik. * Mean Absolute Error (MAE): Nilai MAE yang lebih rendah lebih baik, karena merepresentasikan rata-rata perbedaan absolut antara nilai prediksi dan nilai aktual. * Mean Squared Error (MSE): Mirip dengan MAE, nilai MSE yang lebih rendah lebih disukai. MSE memberikan penalti yang lebih berat pada error yang lebih besar. * Root Mean Squared Error (RMSE): RMSE adalah akar kuadrat dari MSE dan memiliki satuan yang sama dengan variabel target. Nilai RMSE yang lebih rendah lebih baik. * Mean Absolute Percentage Error (MAPE): MAPE merepresentasikan rata-rata perbedaan persentase antara nilai prediksi dan nilai aktual. Nilai MAPE yang lebih rendah lebih diinginkan. Berdasarkan metrik ini dan hasil yang dihasilkan, berikut analisis yang mungkin: * XGBoost dan Random Forest umumnya berkinerja sangat baik di semua metrik untuk set pelatihan dan pengujian. Mereka cenderung memiliki nilai R2 yang lebih tinggi dan metrik error yang lebih rendah dibandingkan dengan model lain. Model-model ini memiliki kekuatan prediksi yang tinggi dan menggeneralisasi dengan baik ke data yang tidak terlihat. * Ridge dan Lasso memberikan kinerja yang layak tetapi mungkin tidak sekuat XGBoost atau Random Forest. Kinerja mereka dapat bervariasi tergantung pada dataset dan penyetelan hyperparameter. Model-model ini memberikan beberapa regularisasi untuk berpotensi meningkatkan generalisasi. * Linear Regression dan AdaBoost dapat menunjukkan kinerja yang lebih rendah dibandingkan dengan model lain. Linear Regression mungkin terlalu sederhana untuk dataset ini, sementara AdaBoost dapat sensitif terhadap outlier. Model-model ini mungkin memerlukan rekayasa fitur atau penyetelan parameter yang lebih cermat untuk meningkatkan kinerjanya. * Performance comparison of Importance Score. |