Novelty of the Project

The dataset for my project 'Airbnb Price Predictor' was taken from Kaggle. The original dataset contained 29 columns relating to different details such as id, zipcode, thumbnail etc.
In order to ensure correct fitting of the model I reduced the number of columns to remove unnecessary data such as 'first_review', 'last_review', 'id' etc. since they are not of use for predicting the price.
The dataset is also transformed and new columns are generated since certain columns need encoding. The project makes use of label encoding to convert object dtype to int dtype and one hot encoding to convert ~~boolean~~ city names to 1/0 depending on whether city is present or not.
I have used model.feature_importances_ to get the list of important features. Based on the results obtained the two least important features are dropped. In this way using feature importances methodology we are able to remove unnecessary features.

For model training using RandomForest Regressor I employed hyperparameter tuning. I have fit 5 folds for each of 15 candidates totalling 75 fits. Randomised search CV is used to define the hyperparameter values. A grid is defined for n_estimators, max_features, max_depth, min_samples_split and min_samples_leaf to search for best hyperparameters. This grid is then used in param_distributions parameter of Randomised Search CV and then fitting is done using best parameters. In this way hyperparameter tuning

helps improve R2 score and give good results.