Flag the junk property listings

As the real estate becomes a scarce commodity in big cities of the world; renovating old properties for sale is becoming a thriving business. However, not every property becomes an attractive portfolio after tons of time, money and effort goes into giving it a second life.

Reasons can be its location , condition being too dilapidated or simply the neighbourhood having lost it lustre of the olden days. These things are not very difficult to figure out after a thorough assessment. But by the time that assessment is over; a lot of time , manpower and money is lost, which makes the overall endeavour at times resulting in crippling loss for this upcoming branch of alternative business opportunity for real estate industry.

Flagging a property as possibly junk before hand can help businesses prioritise their efforts and focus them on the more probable successes rather than bogged down with the weight of unsaleable portfolio .

In this project we will be making use of data from such past operations where many properties were found to be junk after their purchase for renovations. Our task is to predict if a property is going to be unfit using its listing details and other features which are part of first preliminary assessment .

Data Files

Train Dataset =Property_train.csv

Test Dataset = Property_test_share.csv

Formal Problem Statement

Variable names are self explanatory and there is no formal data dictionary provided by the client .

Your task here is to build a predictive model for predicting whether a property should be marked as junk on the basis of listing details and preliminary assessment details. You need to build your model on the train dataset. **Test dataset does not have a response column; you need to predict those values and submit it in a csv format.**

target column = Junk

Evaluation Criterion Part 1:

You will first attempt Part 1 of this project which is a quiz. You can access it through LMS. This quiz needs to be answered based on exploration of the dataset given and some generic questions about algorithms discussed in the course. Consider only the training dataset for data cleaning and exploration to answer the quiz questions. There will be 10 questions of which you need to get at least 7 correct in order to pass the project.

Part 2:

Here you work on creating the machine learning models and choosing the one which gives the best performance. **You can refer to the Project Process Guides provided in LMS to understand how to approach and work on a project.**

For this project, score will be calculated as:

roc_auc score

This score will be calculated using your predictions for the test file. **You need to score more than 0.65 in order to pass the project submission for this particular project**. You need to submit predicted probabilities [**not the** hard classes] for this project.

Submission:

Please give appropriate names to submission file. Preferable containing your name and attempt number [this is for you to keep track of your different experiment performance] . Submission needs to be a csv file . Any other format like excel , pdf etc will not be graded.

Number of rows in the submission csv should be exactly the same as test data. If this is not taken care of, your submission will not be graded.

You can make as many submissions you want . [We might ask you to submit the script which was used to generate the submission at any time].

In order to clear this project, you are required to clear both, Part 1 as well as Part 2 of this assignment. Wish you all the best!