Apartment prices forecasting

The questions to outline project.

1 Planning the industrial research project

1. Goal of the project. (**Expected development result.**) — A model that makes a forecast on apartment prices based on various characteristics of real estate.

2. Applied problem solved in the project. (**How will the result be used?**) — Project will help people buy an apartment at a time when its price will be minimal and it will also help sellers and real estate agencies in making informed decisions.

3. Description of historical measured data. (**Formats and timing.**) — Data can be collected from various sources, such as databases of real estate agencies and classifieds sites (for example, CIAN). Time frame — data for the last 5-10 years. The data will be presented as a DataFrame, where the rows represent individual apartments and the columns represent characteristics: price, area in square meters, number of bedrooms, number of bathrooms, district, year of construction, condition of the apartment. The risks of an increase in the central bank's key rate, standard discount periods from the developer and the possibility of buying on a mortgage should also be taken into account.

4. Quality criteria. (**How is the quality of the obtained result measured, what is in the report?**) — A loss function mean square error (MSE) will be used for optimization:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{1}$$

where y_i - groung truth, $\hat{y}_i$ - prediction.

5. Project feasibility. (**How to show that the project is feasible, list of possible risks.**) — The possible risks are: lack of data or their lack of representativeness, high correlation of features, changing market conditions, which may lead to the obsolescence of the model. To identify and eliminate systematic errors after training the model, it is necessary to perform cross-validation and run on test data.

6. Conditions necessary for successful project implementation. (**Organization of work.**) — The data set must be complete, relevant and diverse to cover different market segments. The team of the project should consist of data scientists, analysts, developer and real estate experts.

7. Solution methods. (**Procedure libraries.**) — The method can be implemented using machine learning models Random Forest or XGBoost. Also using neural networks. The project can be implemented in Python3 using NumPy, Pandas, Scikit-learn, Statsmodels.

2 Research or development?

In other words, novelty or technological advancement?

Expert: (**How long will the model be used? What will replace it in the future?**)

Apartment prices are growing rapidly and the market is constantly changing, so the model will be relevant for no more than 5 years, then it will need to be trained on new data. Also, new more powerful models are likely to be invented, which in a trained form will give much more accurate predictions. But the idea of the project itself, it seems to me, will not lose its relevance.