UDACITY

‹ Return to "Machine Learning Engineer Nanodegree" in the classroom

DISCUSS ON STUDENT HUB

Capstone Proposal

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Hello José María,

This is a really interesting project, both because it tackles an important real-world problem and also because it requires regression and time-series. Your proposal is very well organized and you demonstrated that you have familiarity with the strategy that you are going to use in the proposed problem.

I hope the suggestions are useful for the development of your final project.

Best regards,

Project Proposal

Student clearly describes the problem that is to be solved. The problem is well defined and has at least one relevant potential solution. Additionally, the problem is quantifiable, measurable, and replicable.

Awesome
The problem is clearly defined as a forecasting task, which will be modeled as a regression and time-series problem based on data from a Wind Turbine to predict the energy generation.

Student briefly details background information of the domain from which the project is proposed.

Historical information relevant to the project should be included. It should be clear how or why a problem in the domain can or should be solved. Related academic research should be appropriately cited. A discussion of the student's personal motivation for investigating a particular problem in the domain is encouraged but not required.

Awesome
You provided an excellent overview of the proposed project:

- It mentioned the challenges related to forecasting models.
- You clearly presented the relevance of the problem domain and its applicability to ML.
- Wind power forecasting is a very interesting topic.

Well done!

Student proposes at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model presented. The evaluation metric(s) proposed are appropriate given the context of the data, the problem statement, and the intended solution.

Awesome
The metrics to evaluate the benchmark and the final models were presented:

• R2 score
• Mean Squared Error

A benchmark model is provided that relates to the domain, problem statement, and intended solution. Ideally, the student's benchmark model provides context for existing methods or known information in the domain and problem given, which can then be objectively compared to the student's solution. The benchmark model is clearly defined and measurable.

Awesome
You presented two benchmarks for your project based on Kaggle competition:

- XGBRegressor: R2_score in testing set: 0.8823
- LGBMRegressor: R2_score in testing set: 0.9108

Student clearly describes a solution to the problem. The solution is applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, the solution is quantifiable, measurable, and replicable.

Suggestion
In the "solution statement," you presented your solution through the application of different ML models (in general) and the necessity of some steps of preprocessing data. However, in this section, it is important to present the models that you are going to use (as you did in "problem statement"). You could move the following passage to this section:

> "In summary, the main objectives are as follows:
> • Develop different statistical and AI models to predict energy forecasting using meteorological data. Some of these models could be:
> o LinearRegression
> o MultilinearRegression o XGBoost
>
> o RegressionForest
> o SVR
> o DeepNeuralNetwork
> • Optional: Predict the wind energy 3-hour-ahead using: o ARIMA
> o LSTM
> o CNN-LSTM"

You are proposing a quite challenging task to your project, modeling the problem first as regression and then as time-series. It may be interesting to set the problem as a time-series (only). Knowing that the wind generation is influenced by seasonal aspects, it can presents better results.

The dataset(s) and/or input(s) to be used in the project are thoroughly described. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

Awesome
The origin of the dataset was presented, as well as the description of relevant inputs.

Student summarizes a theoretical workflow for approaching a solution given the problem. A discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

Suggestion
You summarized the theoretical workflow for approaching a solution to the problem. However, it is important to discuss each step of your solution. For instance:

- Exploratory Data Analysis - Which aspects you intend to verify?
- Feature Engineering - What is the strategy behind?
- Develop Different Models - Which models? How?

It is also interesting to add a Refinement phase:

- What are you going to attempt to improve your final model? One approach is to try to optimize some of the hyperparameters.

The proposal follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling

mistakes are present. All resources used and referenced are properly cited.

Awesome

You followed the template and the proposal is well written.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review