☰   **G** **Great**
         Learning   🔍   👤

← Go Back to Model Tuning

⁝☰ Course Content

ReneWind : Problem Statement

Submission type	:	File Upload
Due Date	:	Feb 10, 1:30 AM CET
Total Marks	:	60
Available from	:	Jan 25, 3:30 PM

Description   ⌃

Context

Renewable energy sources play an increasingly important role in the global energy mix, as the effort to reduce the environmental impact of energy production increases.

Out of all the renewable energy alternatives, wind energy is one of the most developed technologies worldwide. The U.S Department of Energy has put together a guide to achieving operational efficiency using predictive maintenance practices.

Predictive maintenance uses sensor information and analysis methods to measure and predict degradation and future component capability. The idea behind predictive maintenance is that failure patterns are predictable and if component failure can be predicted accurately and the component is replaced before it fails, the costs of operation and maintenance will be much lower.

The sensors fitted across different machines involved in the process of energy generation collect data related to various environmental factors (temperature, humidity, wind speed, etc.) and additional features related to various parts of the wind turbine (gearbox, tower, blades, break, etc.).

Objective

"ReneWind" is a company working on improving the machinery/processes involved in the production of wind energy using machine learning and has collected data of generator failure of wind turbines using sensors. They have shared a ciphered version of the data, as the data collected through sensors is confidential (the type of data collected varies with companies). Data has 40 predictors, 20000 observations in the training set, and 5000 in the test set.

The objective is to build various classification models, tune them, and find the best one that will help identify failures so that the generators could be repaired before failing/breaking to reduce the overall maintenance cost. The nature of predictions made by the classification model will translate as follows:

  True positives (TP) are failures correctly predicted by the model. These will result in repair costs.
  False negatives (FN) are real failures where there is no detection by the model. These will result in replacement costs.
  False positives (FP) are detections where there is no failure. These will result in inspection costs.

It is given that the cost of repairing a generator is much less than the cost of replacing it, and the cost of the inspection is less than the cost of repair.

"1" in the target variables should be considered as "failure" and "0" represents "No failure".

Data Description

The data provided is a transformed version of original data which was collected using sensors.

Train.csv - To be used for training and tuning of models.

Test.csv - To be used only for testing the performance of the final best model.

Both the datasets consist of 40 predictor variables and 1 target variable

Submission Guidelines

1. There are two ways to work on this project:

i. Full-code way: The full code way is to write the solution code from scratch and only submit a final Jupyter notebook with all the insights and observations.

ii. Low-code way. The low-code way is to use an existing solution notebook template to build the solution and then submit a business presentation with insights and recommendations.

The primary purpose of providing these two options is to allow learners to opt for the approach that aligns with their individual learning aspirations and outcomes. The below table elaborates on these two options.

Submission type	Who should choose	What is the same across the two	What is different across the two	Final submission file [IMP]	Submission Format
Full-code	Learners who aspire to be in hands-on coding roles in the future focussed on building solution codes from scratch	Perform exploratory data analysis to identify insights and recommendat	Focus on code writing: 10-20% grading on the quality of the final code submitted	Solution notebook from the full-code template	.html

			ions for the problem		submitted in .html format
Lo w- cod e	Learners who aspire to be in managerial roles in the future-focussed on solution review, interpretation, recommendations, and communicating with business		Focus on business presentation: 10-20% grading on the quality of the final business presentation submitted	Business presentation in .pdf format with problem definition, insights, and recommendat ions	.pdf

Please follow the below steps to complete the assessment. Kindly note that if you submit a presentation, ONLY the presentation will be evaluated. Please make sure that all the sections mentioned in the rubric have been covered in your submissions.

i. Full-code version

Download the full-code version of the learner notebook.
Follow the instructions provided in the notebook to complete the project.
Clearly write down insights and recommendations for the business problems in the comments.
Submit only the solution notebook prepared from the learner notebook [format: .html]

ii. Low-code version

Download the low-code version of the learner notebook.
Follow the instructions provided in the notebook to complete the project.
Prepare a business presentation with insights and recommendations to the business problem.
Submit only the presentation [format: .pdf]

2. Any assignment found copied/plagiarized with other submissions will not be graded and awarded zero marks.

3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation.

4. Submission will not be evaluated if

it is submitted post-deadline, or,

more than 1 file is submitted.

Best Practices for Full-code submissions

The final notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
The notebook should be run from start to finish in a sequential manner before submission.
It is important to remove all warnings and errors before submission.
The notebook should be submitted as an HTML file (.html) and NOT as a notebook file (.ipynb).
Please refer to the FAQ page for common project-related queries.

Best Practices for Low-code submissions

The presentation should be made keeping in mind that the audience will be the Data Science lead of a company.
The key points in the presentation should be the following:
  Business Overview of the problem and solution approach
  Key findings and insights which can drive business decisions
  Business recommendations
  Focus on explaining the key takeaways in an easy-to-understand manner.
  The inclusion of the potential benefits of implementing the solution will give you the edge.
Copying and pasting from the notebook is not a good idea, and it is better to avoid showing codes unless they are the focal point of your presentation.
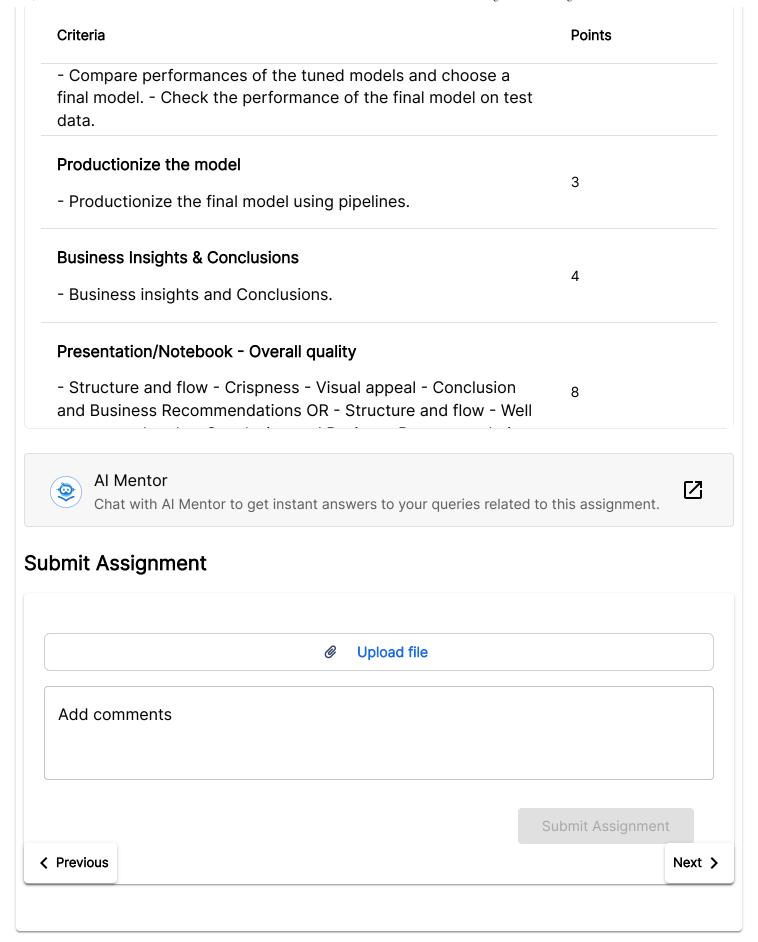The presentation should be submitted as a PDF file (.pdf) and NOT as a .pptx file.
Please refer to the FAQ page for common project-related queries.

Happy Learning!

Scoring guide (Rubric) - ReneWind   ∧

Criteria	Points
Exploratory Data Analysis and Insights - Overview of the data - Univariate analysis	4
Data pre-processing - Prepare the data for analysis - Missing value Treatment - Ensure no data leakage	4
Model building - Original data - Build atleast 6 classification models (Using logistic regression, decision trees, random forest, bagging classifier and boosting methods) - You can choose not to build XGBoost if you are facing issues with installation	6
Model building - Oversampled data - Build atleast 6 classification models using oversampled train data (Using logistic regression, decision trees, random forest, bagging classifier and boosting methods) - You can choose not to build XGBoost if you are facing issues with the installation	7
Model building - Undersampled data - Build atleast 6 classification models using undersampled train data (Using logistic regression, decision trees, random forest, bagging classifier and boosting methods) - You can choose not to build XGBoost if you are facing issues with the installation	7
Hyperparameter tuning - Choose at least 3 best performing models among all the models built previously (Mention the reason for the choices made) - Tune the chosen models. - Check the performance of the tuned models.	12
Model Performances	5

Criteria	Points
- Compare performances of the tuned models and choose a final model. - Check the performance of the final model on test data.	
Productionize the model - Productionize the final model using pipelines.	3
Business Insights & Conclusions - Business insights and Conclusions.	4
Presentation/Notebook - Overall quality - Structure and flow - Crispness - Visual appeal - Conclusion and Business Recommendations OR - Structure and flow - Well	8

AI Mentor

Chat with AI Mentor to get instant answers to your queries related to this assignment. ⬈

Submit Assignment

📎 **Upload file**

Add comments

Submit Assignment

‹ Previous   Next ›

© 2024 All rights reserved

Privacy Terms of service Help