greatlearning
Power Ahead

AIML | MODULE PROJECT

greatlearning
Power Ahead

5 Takeaways

1. AIML module projects are designed to have a detailed hands on to integrate theoretical knowledge with actual practical implementations.

2. AIML module projects are designed to enable you as a learner to work on realtime industry scenarios, problems and datasets.

3. AIML module projects are designed to enable you simulating the designed solution using AIML techniques onto python technology platform.

4. AIML module projects are designed to be scored using a predefined rubric based system.

5. AIML module projects are designed to enhance your learning above and beyond. Hence, it might require you to experiment, research, self learn and implement.

AIML | MODULE PROJECT

greatlearning
Power Ahead

FEATURISATION & MODEL TUNING

AIML module project consists of industry based problems framed as detailed questions which can be solved using all ML modelling techniques.

TOTAL SCORE 60

PROJECT BASED

TOTAL SCORE 60

- **DOMAIN:** Semiconductor manufacturing process
- **CONTEXT:** A complex modern semiconductor manufacturing process is normally under constant surveillance via the monitoring of signals/ variables collected from sensors and or process measurement points. However, not all of these signals are equally valuable in a specific monitoring system. The measured signals contain a combination of useful information, irrelevant information as well as noise. Engineers typically have a much larger number of signals than are actually required. If we consider each type of signal as a feature, then feature selection may be applied to identify the most relevant signals. The Process Engineers may then use these signals to determine key factors contributing to yield excursions downstream in the process. This will enable an increase in process throughput, decreased time to learning and reduce the per unit production costs. These signals can be used as features to predict the yield type. And by analysing and trying out different combinations of features, essential signals that are impacting the yield type can be identified.
- **DATA DESCRIPTION:** sensor-data.csv : (1567, 592)

  The data consists of 1567 examples each with 591 features.

  The dataset presented in this case represents a selection of such features where each example represents a single production entity with associated measured features and the labels represent a simple pass/fail yield for in house line testing. Target column " –1" corresponds to a pass and "1" corresponds to a fail and the data time stamp is for that specific test point.
- **PROJECT OBJECTIVE:** We will build a classifier to predict the Pass/Fail yield of a particular process entity and analyse whether all the features are required to build the model or not.

Steps and tasks: [Total Score: 60 points]

1. Import and explore the data.
2. Data cleansing:
   - Missing value treatment.
   - Drop attribute/s if required using relevant functional knowledge.
   - Make all relevant modifications on the data using both functional/logical reasoning/assumptions.
3. Data analysis & visualisation:
   - Perform detailed relevant statistical analysis on the data.
   - Perform a detailed univariate, bivariate and multivariate analysis with appropriate detailed comments after each analysis.
4. Data pre-processing:
   - Segregate predictors vs target attributes
   - Check for target balancing and fix it if found imbalanced.
   - Perform train-test split and standardise the data or vice versa if required.
   - Check if the train and test data have similar statistical characteristics when compared with original data.
5. Model training, testing and tuning:
   - Model training:
     - Pick up a supervised learning model.
     - Train the model.
     - Use cross validation techniques.

       Hint: Use all CV techniques that you have learnt in the course.
     - Apply hyper-parameter tuning techniques to get the best accuracy.

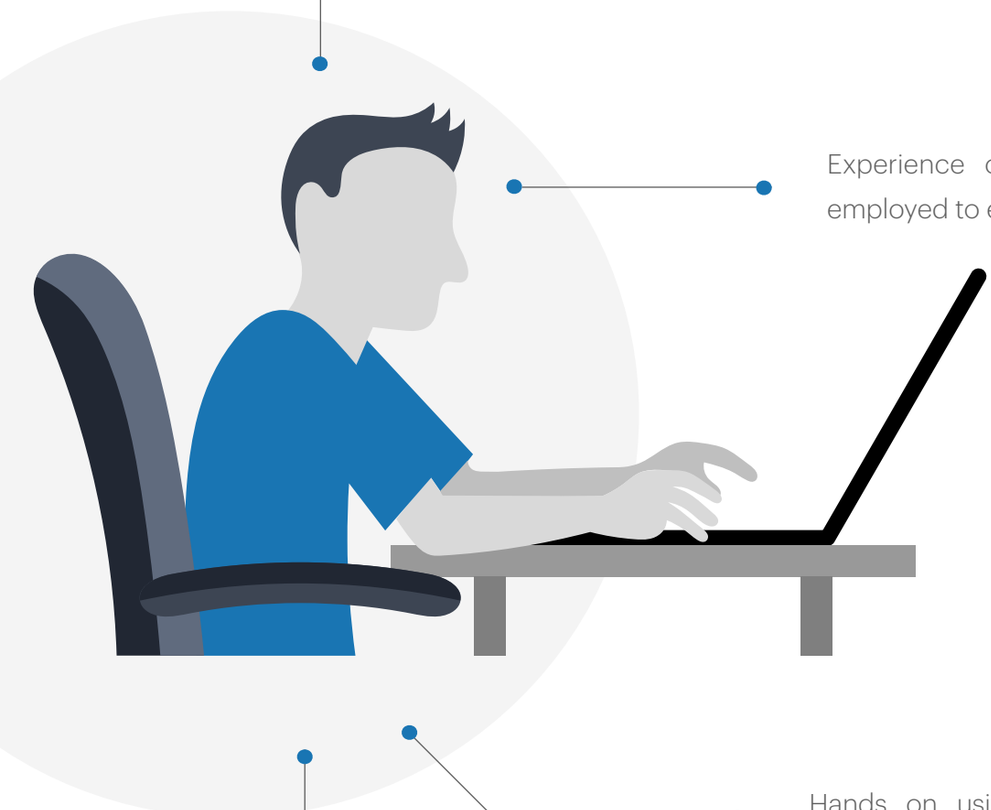
       Suggestion: Use all possible hyper parameter combinations to extract the best accuracies.
     - Use any other technique/method which can enhance the model performance.

       Hint: Dimensionality reduction, attribute removal, standardisation/normalisation, target balancing etc.
     - Display and explain the classification report in detail.
     - Design a method of your own to check if the achieved train and test accuracies might change if a different sample population can lead to new train and test accuracies.

       Hint: You can use your concepts learnt under Applied Statistics module.
     - Apply the above steps for all possible models that you have learnt so far.
   - Display and compare all the models designed with their train and test accuracies.
   - Select the final best trained model along with your detailed comments for selecting this model.
   - Pickle the selected model for future use.
   - Import the future data file. Use the same to perform the prediction using the best chosen model from above. Display the prediction results.
6. Conclusion and improvisation:
   - Write your conclusion on the results.

LEARNING OUTCOME

Experience the complete process involved in ML model designing, starting from data cleansing, manipulation, data validation, model training, model tuning, model validation, model baselining/pickling and data prediction using future data.



Experience on using all possible techniques employed to enhance the model performance.

Hands on using supervised and unsupervised learning models together for model development.

Deep dive in understanding that ML models are data dependent and the performance varies as per the data and it's hidden patterns. Enabling you to understand under better model selection depending on the situation.

" Put yourself in the shoes of an actual "

DATA SCIENTIST

THAT's YOU

Assume that you are working at the company which has received the above problem statement from internal/external client. Finding the best solution for the problem statement will enhance the business/ operations for your organisation/project. You are responsible for the complete delivery. Put your best analytical thinking hat to squeeze the raw data into relevant insights and later into an AIML working model.

PLEASE NOTE

Designing a data driven decision product typically traces the following process:

1. Data and insights:
   Warehouse the relevant data. Clean and validate the data as per the the functional requirements of the problem statement. Capture and validate all possible insights from the data as per the the functional requirements of the problem statement. Please remember there will be numerous ways to achieve this. Sticking to relevance is of utmost importance. Pre-process the data which can be used for relevant AIML model.
2. AIML training:
   Use the data to train and test a relevant AIML model. Tune the model to achieve the best possible learnings out of the data. This is an iterative process where your knowledge on the above data can help to debug and improvise. Different AIML models react differently and perform depending on quality of the data. Baseline your best performing model and store the learnings for future usage.
3. AIML end product:
   Design a trigger or user interface for the business to use the designed AIML model for future usage. Maintain, support and keep the model/ product updated by continuous improvement/training. These are generally triggered by time, business or change in data.

IMPORTANT POINTERS

Project should be submitted as a single ".html" and ".ipynb" file. Follow the below best practices where your submission should be:

- ".html" and ".ipynb" files should be an exact match.
- Pre-run codes with all outputs intact.
- Error free & machine independent i.e. run on any machine without adding any extra code.
- Well commented for clarity on code designed, assumptions made, approach taken, insights found and results obtained.

Project should be submitted on or before the deadline given by the program office.

Project submission should be an original work from you as a learner. If any percentage of plagiarism found in the submission, the project will not be evaluated and no score will be given.

greatlearning
Power Ahead

HAPPY LEARNING