



AIML

MODULE PROJECT



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

ENSEMBLE TECHNIQUES

AIML module project consists of industry based dataset and problem statement which can be solved using ensemble supervised learning techniques



TOTAL
SCORE

60

PROJECT BASED

TOTAL
SCORE

60

- **DOMAIN:** Telecom
- **CONTEXT:** A telecom company wants to use their historical customer data to predict behaviour to retain customers. You can analyse all relevant customer data and develop focused customer retention programs.
- **DATA DESCRIPTION:** Each row represents a customer, each column contains customer’s attributes described on the column Metadata. The data set includes information about:
 - Customers who left within the last month – the column is called Churn
 - Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
 - Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
 - Demographic info about customers – gender, age range, and if they have partners and dependents
- **PROJECT OBJECTIVE:** Build a model that will help to identify the potential customers who have a higher probability to churn. This help the company to understand the pinpoints and patterns of customer churn and will increase the focus on strategising customer retention.
- Steps to the project: [Total score: 60 points]
 1. **Import and warehouse data:** [Score: 5 point]
 - Import all the given datasets. Explore shape and size.
 - Merge all datasets onto one and explore final shape and size.
 2. **Data cleansing:** [Score: 10 point]
 - Missing value treatment
 - Convert categorical attributes to continuous using relevant functional knowledge
 - Drop attribute/s if required using relevant functional knowledge
 - Automate all the above steps
 3. **Data analysis & visualisation:** [Score: 10 point]
 - Perform detailed statistical analysis on the data.
 - Perform a detailed univariate, bivariate and multivariate analysis with appropriate detailed comments after each analysis.
 4. **Data pre-processing:** [Score: 5 point]
 - Segregate predictors vs target attributes
 - Check for target balancing and fix it if found imbalanced.
 - Perform train-test split.
 - Check if the train and test data have similar statistical characteristics when compared with original data.
 5. **Model training, testing and tuning:** [Score: 25 point]
 - Train and test all ensemble models taught in the learning module.
 - Suggestion: Use standard ensembles available. Also you can design your own ensemble technique using weak classifiers.
 - Display the classification accuracies for train and test data.
 - Apply all the possible tuning techniques to train the best model for the given data.
 - Suggestion: Use all possible hyper parameter combinations to extract the best accuracies.
 - 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.
 6. **Conclusion and improvisation:** [Score: 5 point]
 - Write your conclusion on the results.
 - Detailed suggestions or improvements or on quality, quantity, variety, velocity, veracity etc. on the data points collected by the telecom operator to perform a better data analysis in future.

“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.

LEARNING OUTCOME

Hands on experience on data design, data cleansing, data warehousing and data manipulation required to drive an AIML model.

Hands on working knowledge on all traditional and customised ensemble techniques and algorithms.

Deep dive onto designing, training, testing and tuning a AIML model.

Hands-on experience on converting the designed model into a sharable pickled file to be used for production.

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.