



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

SUPERVISED LEARNING



PART I

AIML module project part I consists of industry based dataset and problem statement which can be solved using KNN supervised learning technique.

PART II

AIML module project part II consists of industry based dataset and problem statement which can be solved using all supervised learning techniques.

TOTAL
SCORE

60

PART
ONE

PROJECT BASED

TOTAL
SCORE

30

- **DOMAIN:** Healthcare
- **CONTEXT:** Medical research university X is undergoing a deep research on patients with certain conditions. University has an internal AI team. Due to confidentiality the patient’s details and the conditions are masked by the client by providing different datasets to the AI team for developing a AIML model which can predict the condition of the patient depending on the received test results.
- **DATA DESCRIPTION:** The data consists of biomechanics features of the patients according to their current conditions. Each patient is represented in the data set by six biomechanics attributes derived from the shape and orientation of the condition to their body part.
 1. P_incidence
 2. P_tilt
 3. L_angle
 4. S_slope
 5. P_radius
 6. S_degree
 7. Class
- **PROJECT OBJECTIVE:** Demonstrate the ability to fetch, process and leverage data to generate useful predictions by training Supervised Learning algorithms.

Steps and tasks: [Total score: 30 points]

1. **Import and warehouse data:**
 - Import all the given datasets and explore shape and size of each.
 - Merge all datasets onto one and explore final shape and size.
2. **Data cleansing:**
 - Explore and if required correct the datatypes of each attribute
 - Explore for null values in the attributes and if required drop or impute values.
3. **Data analysis & visualisation:**
 - 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:**
 - Segregate predictors vs target attributes
 - Perform normalisation or scaling if required.
 - Check for target balancing. Add your comments.
 - Perform train-test split.
5. **Model training, testing and tuning:**
 - Design and train a KNN classifier.
 - Display the classification accuracies for train and test data.
 - Display and explain the classification report in detail.
 - Automate the task of finding best values of K for KNN.
 - Apply all the possible tuning techniques to train the best model for the given data. Select the final best trained model with your comments for selecting this model.
6. **Conclusion and improvisation:**
 - 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 research team to perform a better data analysis in future.

PART
TWO

PROJECT BASED

TOTAL
SCORE

30

- **DOMAIN:** Banking and finance
- **CONTEXT:** A bank X is on a massive digital transformation for all its departments. Bank has a growing customer base whee majority of them are liability customers (depositors) vs borrowers (asset customers). The bank is interested in expanding the borrowers base rapidly to bring in more business via loan interests. A campaign that the bank ran in last quarter showed an average single digit conversion rate. Digital transformation being the core strength of the business strategy, marketing department wants to devise effective campaigns with better target marketing to increase the conversion ratio to double digit with same budget as per last campaign.
- **DATA DESCRIPTION:** The data consists of the following attributes:
 1. **ID:** Customer ID
 2. **Age** Customer's approximate age.
 3. **CustomerSince:** Customer of the bank since. [unit is masked]
 4. **HighestSpend:** Customer's highest spend so far in one transaction. [unit is masked]
 5. **ZipCode:** Customer's zip code.
 6. **HiddenScore:** A score associated to the customer which is masked by the bank as an IP.
 7. **MonthlyAverageSpend:** Customer's monthly average spend so far. [unit is masked]
 8. **Level:** A level associated to the customer which is masked by the bank as an IP.
 9. **Mortgage:** Customer's mortgage. [unit is masked]
 10. **Security:** Customer's security asset with the bank. [unit is masked]
 11. **FixedDepositAccount:** Customer's fixed deposit account with the bank. [unit is masked]
 12. **InternetBanking:** if the customer uses internet banking.
 13. **CreditCard:** if the customer uses bank's credit card.
 14. **LoanOnCard:** if the customer has a loan on credit card.
- **PROJECT OBJECTIVE:** Build an AIML model to perform focused marketing by predicting the potential customers who will convert using the historical dataset.

Steps and tasks: [Total Score: 30 points]

 1. Import and warehouse data:
 - Import all the given datasets and explore shape and size of each.
 - Merge all datasets onto one and explore final shape and size.
 2. Data cleansing:
 - Explore and if required correct the datatypes of each attribute
 - Explore for null values in the attributes and if required drop or impute values.
 3. Data analysis & visualisation:
 - 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:
 - Segregate predictors vs target attributes
 - Check for target balancing and fix it if found imbalanced.
 - Perform train-test split.
 5. Model training, testing and tuning:
 - Design and train a Logistic regression and Naive Bayes classifiers.
 - Display the classification accuracies for train and test data.
 - Display and explain the classification report in detail.
 - Apply all the possible tuning techniques to train the best model for the given data. Select the final best trained model with your comments for selecting this model.
 6. Conclusion and improvisation:
 - 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 bank to perform a better data analysis in future.

LEARNING OUTCOME

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

Balancing a highly imbalanced data by creating a synthetic data.

Detailed design using visualisations and EDA techniques for relevant insights

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

“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.