NGEE ANN
P O L Y T E C H N I C

School of InfoComm Technology

Predictive Modeling

Diploma in Financial Informatics (FI)
Diploma in Information Technology (IT)
YR 2/3 (2020/21), Semester 4/6

INDIVIDUAL ASSIGNMENT 2

(40% of Predictive Analytics Module)

Deadline for Submission:

Presentation: 31st Jan 2021 (Sunday), 2359 Hours
Report: 14th Feb 2021 (Sunday), 2359 Hours

Student Name :	
Student Number :	

Penalty for late submission:
10% of the marks will be deducted every day after the deadline.
NO submission will be accepted after 21st Feb 2021, 23:59.

PREDICTIVE ANALYTICS ASSIGNMENT 2

1. OBJECTIVES

In this assignment we will utilize the Predictive Models to solve both classification and regression problems using Python:

- To build Predictive Models to solve HR Analytics problem and Airbnb Singapore problem.
- To train the models by adjusting different hyperparameters.
- To evaluate the model performance and document findings

2. DATASETS

Please refer to assignment 1 document for the problem statements and data dictionaries for HR Analytics problem and Airbnb Singapore problem. You have already explored and cleansed the datasets in the previous assignment, so here in this assignment 2, we will simply use the cleansed data for Predictive Modeling.

2.1. HR ANALYTICS (CLASSIFICATION PROBLEM): hr_data_new.csv

2.2. AIRBNB SINGAPORE (REGRESSION PROBLEM): listings_new.csv

3. SUGGESTED TASKS

3.1. HR ANALYTICS (CLASSIFICATION PROBLEM)

You are suggested to tackle this problem in the below FOUR steps.

Step 1: **Load and Sample data**

Load the cleansed dataset (hr_data_new.csv). Stratified Sampling the data to make sure you have balanced samples (if you haven't done this in Assignment 1). Sample the data into training data & testing data.

Step 2: **Build the Classification Model(s)**

Build predictive model(s) using training data to predict whether the employees will be promoted or not. You are required to build at least one model. But it will be highly appreciated if you can build several different models (e.g. logistic regression, decision tree, artificial neural network, ensemble models and etc.) to solve this problem and compare them later.

Step 3: **Evaluate and Improve the Model(s) Performance**

Evaluate the model(s) performance (e.g. accuracy) using testing data and see whether you can further improve the model performance by:
- Tuning the model hyperparameters
- Selecting different input features
- Adjusting the input data
- Other effective techniques

Step 4: **Summarize the findings**

Summarize your findings. If you build several different models, please recommend the best model and explain why this model performs better than the other models.

3.2. AIRBNB SINGAPORE (REGRESSION PROBLEM)

You are suggested to tackle this problem in the below FOUR steps.

Step 1: **Load and Sample the data**

Load the cleansed dataset (listings_new.csv). Sample the data into training data & testing data.

Step 2: **Build the Regression Model(s)**

Build predictive model(s) using training data to estimate the listing price. You are required to build at least one model. But it will be highly appreciated if you can build several different models (e.g. linear regression, decision tree, artificial neural network, ensemble models and etc.) to solve this problem and compare them later.

Step 3: **Evaluate and Improve the Model(s) Performance**

Evaluate the model(s) performance (e.g. MSE, MAE, R Squared Value) using testing data and see whether you can further improve the model performance through:
- Tuning the model hyperparameters
- Selecting different input features
- Adjusting the input data
- Other effective techniques

Step 4: **Summarize the findings**

Summarize your findings. If you build several different models, please recommend the best model and explain why this model performs better than the other models.

4. SUGGESTED REPORT FORMAT & CONTENT GUIDELINES

Write an **INDIVIDUAL** report with the following sections (see Table below). Sample content description is provided for each section. You are free to include other relevant information you deem necessary in the sections. You are strongly encouraged to include screen shots in your explanation, description and analysis.

*(Note: For a page with 1 inch margins, 11 point Calibri font, and minimal spacing elements, a good rule of thumb is **500 words** for a single spaced page)*

	Suggested Report Sections & Content Guidelines	Word Count
1.	Table of Contents	NA
2.	Introduction • A brief introduction on solving classification and regression problems using Predictive Models	Min: 200 words Max: 500 words
3.	HR Analytics • Problem understanding and the approaches • Build the model(s) • Evaluate and Improve the model(s) • Summary	Min: 1000 words Max: 3000 words
4.	Airbnb Singapore • Problem understanding and the approaches • Build the model(s) • Evaluate and Improve the model(s) • Summary	Min: 1000 words Max: 3000 words
5.	Conclusion and Further Improvements • Summarize your work on these two problems • Explain the possible further improvements	Min: 500 words Max: 1000 words

5. DELIVERABLES

Presentation and demonstration

- Each student will be required to submit a video recorded presentation to showcase and demo the work. The video recorded presentation should be not exceed 10 minutes. Video recorded presentations which exceed the allotted time will be penalized.

- Students to submit the presentation slides in MeL. Deadline for slides submission is **Sunday 31st Jan 2021, 2359 hours.**

- The video presentation must be recorded using Microsoft Teams. After completion of the video recorded presentation, submit the link to the video (from Microsoft Stream). Submit the link to your video recorded presentation using the link below:
  - Assignment 2 Video Presentation Submission Link

    (Login using only your NP student account)

- Deadline for the video submission is **Sunday 31st Jan 2021, 2359 hours.**

Assignment report

- Submit the **softcopy** of the report via **SafeAssign** in MeL. Deadline for softcopy submission is **Sunday 14th Feb 2021, 2359 hours**.
- Submit the Jupyter Notebook file (PA_Assignment_2.ipynb) in MeL. Deadline for softcopy submission is **Sunday 14th Feb 2021, 2359 hours**.

Note: DO NOT PLAGIARIZE (please refer to MeL, Ngee Ann Polytechnic Plagiarism Policy webpage for more information)

6. GRADING CRITERIA

	Grading Criteria	Component Weightage
Presentation	a) Quality of work b) Flow of presentation based on content guidelines (see section 4) c) Quality of presentation slides d) Presentation and articulation skills	**50%**
Final Report	a) Quality of work b) Completeness of report based on suggested report sections and content guidelines (see section 4) c) Clarity of report, Quality of analysis and discussions d) Use of proper visual aids and Use of proper grammar	**50%**