

Data Science Project

1501410 Data Science and Artificial Intelligence

Semester 1 Year 2020, Assoc.Prof.Dr.Tossapon Boongoen (Instructor)

Objectives

- Be able to apply general concept and practice of data science to a real-world data problem,
- Be able to identify possible flaws in the raw data and clean it such that it carries less errors and ready for analysis,
- Be able to suggest a possible step of feature engineering to create useful or informative features, and
- Be able to select and apply appropriate machine learning models to the aforementioned data, with performance evaluation and discussion.

Instructions

1. This is a group project of 4 students (group members: <https://tinyurl.com/y4qc26v7>),
2. The raw data is provided as a separate file named 'tcas.xlsx',
3. Each group has to specify the type of machine learning model that is appropriate and to be used with this data (e.g., clustering),
4. The common steps shown below should be follows:
 - a. EDA
 - b. Data preprocessing and feature engineering
 - c. Model development
 - d. Performance evaluation and discussion
5. Produce a Jupyter notebook that records all the following details
 - a. Names and IDs of group members
 - b. Descriptions of design and coding for all the steps in 4
6. Create web application for your project and deploy it on Heroku
7. Prepare a set of slides for the final presentation.

Submission

1. Presentation slides
2. Jupyter notebook(s) and data files
3. Files for web application

Timeline

- | | |
|--------------------|---|
| • 13 November 2020 | Project assignment |
| • 19 November 2020 | Project consultation (Online via appointment) |
| • 26 November 2020 | Project consultation |
| • 3 December 2020 | Lab Test + Project consultation |
| • 10 December 2020 | Project presentation |
| • 12 December 2020 | Final Exam |