In Assignment 1, we trained the titanic data. In our Final Project, we are going to have a similar data set called Spaceship Titanic.

For the details of data set and other information, please visit the link:

<https://www.kaggle.com/competitions/spaceship-titanic/overview>

Project Description:

* Each team will form 3 students. The team is already assigned. See the group in Canvas.
* There are two competitions your team will play, Kaggle – against professional and armature data scientists and CS559-WS – against 20 different teams in class.

**Kaggle Competition (15 pts):**

* Make your team’s name and join the Competition.
* Use the given train.csv file from Kaggle to train the models.
* Submit your final result (see the submission sample file) by 5/5th and report the rank.
* The submission sample file has the same ID as test.csv.
* Any late submission will not be counted.
* Report the score and rank to the shared CS559\_S23\_FinalProject\_Ranking\_Report/Kaggle\_Competition spreadsheet ([here](https://docs.google.com/spreadsheets/d/1a15R5E7iSRT6iMpKUAKB1Z-dVwU5N0tnE17NhL3ASso/edit?usp=sharing)).

**CS559-WS Competition (85 pts):**

* Use the provided train\_.csv, validation\_.csv, and test\_.csv.
* **Preprocessing (20 pts)** – Any libraries/packages are allowed. All members must be involved.
* **Modeling (15 pts):**
  + Each member will train at least one parametric and non-parametric model using GridSearchCV. **Algorithms must not be duplicated** between team members.
  + Each team must train the model using the stacking method for the final model.

where is the weight of -model, and is the -trained model predicted value.

* **Model Evaluation (5 pts):**
  + For each model, report the accuracy of the test data set to the shared CS559\_S23\_FinalProject\_Ranking\_Report/CS559\_Competition spreadsheet ([here](https://docs.google.com/spreadsheets/d/1a15R5E7iSRT6iMpKUAKB1Z-dVwU5N0tnE17NhL3ASso/edit?usp=sharing)).
* **Ranking Score (15 pts):**
  + See the table below. The overall rank score will be marked.
  + If the accuracy is the same between the same algorithm, then a model with fewer features will be advanced.
* **Presentation (15 pts):**
  + Each team will record a 10-15-minute-long presentation video and submit it.
* **Report (15 pts):**
  + Each team will write a 3 to 5 pages long report (including visualizations) that discusses the workflow summary, the final rank in the Kaggle competition on 5/5th (provide a screenshot of your rank with the team name), result on each model (accuracy and features), and each member’s role and portion of work.

Ranking Board

|  |  |  |  |
| --- | --- | --- | --- |
| Rank | Ensembled Model Accuracy (pts) | Individual Model Accuracy (pts) | Overall Rank (average) (pts) |
| 1 | 15 | 15 | 15 |
| 2 | 14.3 | 14.3 | 14.3 |
| 3 | 13.6 | 13.6 | 13.6 |
| 4 | 12.9 | 12.9 | 12.9 |
| 5 | 12.2 | 12.2 | 12.2 |
| 6 | 11.5 | 11.5 | 11.5 |
| 7 | 10.8 | 10.8 | 10.8 |
| 8 | 10.1 | 10.1 | 10.1 |
| 9 | 9.4 | 9.4 | 9.4 |
| 10 | 8.7 | 8.7 | 8.7 |
| 11 | 8 | 8 | 8 |
| 12 | 7.3 | 7.3 | 7.3 |
| 13 | 6.6 | 6.6 | 6.6 |
| 14 | 5.9 | 5.9 | 5.9 |
| 15 | 5.2 | 5.2 | 5.2 |
| 16 | 4.5 | 4.5 | 4.5 |
| 17 | 3.8 | 3.8 | 3.8 |
| 18 | 3.1 | 3.1 | 3.1 |
| 19 | 2.4 | 2.4 | 2.4 |
| 20 | 1.7 | 1.7 | 1.7 |