Competition 1

Deadline: 2018.03.20 (Tuesday) 23:59

* Competition System: http://140.116.247.120:8001/

Problem: Kobe Bryant's Shot Type Prediction

Kobe Bryant marked his retirement from the NBA by scoring 60 points in his final game as a Los Angeles Laker on Wednesday, April 12, 2016. Drafted into the NBA at the age of 17, Kobe earned the sport's highest accolades throughout his long career. Using 20 years of data on Kobe's swishes and misses, can you predict the action types of his shots. The exampling action types of shots include "Jump Shot", "Driving Dunk Shot", "Layup Shot", "Running Jump Shot", "Reverse Layup Shot", "Slam Dunk Shot", "Turnaround Jump Shot", etc. This competition is well suited for practicing classification basics, and feature engineering. Practice got Kobe an eight-figure contract and 5 championship rings. What will it get you? In the following, we provide you the settings for the competition.

Column Name	Meaning
shot_id	Unique ID of a shot.
action_type	The action type of a shot.
lat	Latitude of a shot.
lon	Longitude of a shot.
loc_x, log_x	The coordinates on the court where the shot took place. NOTE: -250 to 250 for loc_x (negatives to differentiate which side of the basket) and loc_y goes -50-890 (most shots are from less than 400 since that would be forty feet in longitude away from the basket; negatives to account for the space "behind" the basket.
minutes_remaining	Minutes remaining in that quarter.
period	Period = Quarter = 1, 2, 3, 4 (第幾節) NOTE: Period 5 is the first overtime (OT), period 6 the second OT, and period 7 the third OT.
playoffs	It is in the playoff (季後賽) 0=yes, 1=no
season	Season of the game.
seconds_remaining	Seconds remaining in that quarter
shot_distance	Distance to the basket.

shot_made_flag	0=missed, 1=made. There could be some missing values.
shot_zone_area	Right Side(R), Center(C), Right Side(R)
game_date	Date of the game.
opponent	Opponent of Lakers.

- **Training Data (train.csv).** We provide you a training data with 24,557 shots of Kobe Bryant. You can download the network data in Moodle. The data contains a number of columns, as explained in the above table. The "action_type" is the target of prediction.
- **Competition System with Rules.** To upload your prediction results, you should follow the instructions of provided by TA to see the format of the uploading file. The uploaded file is he pure text format. In the file, each line is a pair of "shot_id" and "action_type" separate by comma. For example, "17, Jump Shot" means shot_id 17 is predicted as "Jump Shot". It is also important to let you know that each team has only 20 times for the submission within a day. So please cherish your submission times and do not intend to perform try-and-error tests that may waste your submission times.
- **Performance Evaluation.** We also provide you the test data (**test.csv**), in which the "action type" column is hidden. Since we have the answer of shot action types in test data, we can evaluate any methods by computing some accuracy measures.
- **Competing Methods.** Your competing methods include four TA's approaches, which could be some basic machine learning methods. In addition, two baselines are also provided, including "most frequent" and "random". You are strongly encouraged to beat such two baselines.

You are asked to submit prediction results to the competition website, submit your source codes in a "ipynb" file with clear comments, and describe the details of your methods in a report containing the following justification, and submit your code and the report in Moodle. The maximum length of the report is 10 pages in A4. Note that you are encouraged to submit your report in PDF.

- **Description.** What are your methods for the prediction of action type of shots? What are the main ideas, intuitions, and physical meanings of your methods? You are asked to write down the detailed setting and procedure implemented in your code.
- **Analysis.** You need to analyze why your methods lead to high or low accuracy scores by varying some parameters if any. You might want (highly recommended but not necessary) to answer questions like: when does your method work better? For all the methods you have tried, which is better and which is worse and why? Which parameters (if any) significantly affect the accuracy of your methods? What are the strong and weak points of your methods? Have you combined the prediction results of several classification methods to produce the final prediction? If so, how do you make the combination? Note that if none of the methods that you had tried and developed lead to high accuracy, it's fine. Then reporting which methods

you developed and tried will become important. The grade of your homework will highly depend on both of your description and the analysis in your report. Therefore, it would be better for you to write down the details about all the methods you have tried, show the abovementioned items, and analyze why they lead to better or worse results in your report. With your report, we are able to understand which methods cannot work even though they seem reasonable.

- **[Optional] Visualization.** Seeing is believing. You are suggested to plot some figures in the ipynb file using matplotlib, seaborn, and pandas to investigate the potential relationships between different features and action types, together with some textual description, to explain which features are more important in prediction action types of shots.
- **[Optional] References.** If you methods are using any other packages or external resources, you are asked to clearly write down which packages (e.g. which Github with webpage link) and which machine learning methods you used.

Note that in your codes for these problems,
you need to write some comments to describe the meaning of each part.

Final Note This competition is modified from the competition of "Kobe Bryant Shot Selection" in Kaggle: https://www.kaggle.com/c/kobe-bryant-shot-selection . This task is to predict whether each of Kobe's shot will be made (=1) or missed (=0) (binary classification). You may want to (you are suggested) see how people use various machine learning in scikit-learn to do such shot prediction via https://www.kaggle.com/c/kobe-bryant-shot-selection/kernels .

How to Submit Your Homework?

Submission in NCKU Moodle. Before submitting your homework, please zip the files of your ipynb file (**cp1.ipynb**) and your report (**cp1.pdf**), and name the zip file as "學號 1_學號 2_cp1.zip". For example, if your 學號 of your team are H12345678 and H87654321, then your file name is:

"H12345678_H87654321_cp**1**.zip" or "H12345678_H87654321_cp**1**.rar"

When you zip your files, please follow the instructions provided by TA's slides to submit your file using NCKU Moodle platform http://moodle.ncku.edu.tw .

Have Questions about This Homework?

Please feel free to visit TAs, and ask/discuss any questions in their office hours. We will be more than happy to help you.