



Data Science and Machine Learning Bootcamp

IRON KAGGLE

Iron Kaggle

LET'S COMPETE! Before you jump further, we want you to have an end-to-end experience of solving a Machine Learning problem.

So let's try and predict the sales of shops, shall we?

You have a dataset containing information on shops' sales per day (and several characteristics of them) - your goal is to predict the sales!

Iron Kaggle

Training data (640841 entries): we will share with you a training set of store sales per day, with bits of information of what happened in that day in that store.

Real-Life Data (+70k entries): we will also share with you entries without the sales. This will be used (on the teachers side) to verify how good your model really is!

Iron Kaggle

Expected Deliver:

- “Real-life data set” with an extra column called “sales”, with your predictions (in .csv)
- An expected value of R^2 of performance of your model
- A 5' presentation on the choices you did and the road you took

Iron Kaggle

Schedule:

- 8am: Introduction, Group organization, framing the problem
- 8.30am: Development time
- 1pm: Lunch break
- 2pm: You will receive the “real-life” data
- 2pm: Development time + presentation creation
- 4.30pm: Delivery of the real life data
- 430pm-5pm: Finish your presentation
- 5pm: presentations + winner announcement

Iron Kaggle

DELIVERABLE:

- A .csv file called with your groupd name (e.g, 'G1.csv', 'G2.csv')
- The value of R2 you are expecting to get
- Send this in a .zip file containing two elements:
 - the csv file
 - a txt file with the R2 score inside

Iron Kaggle

Winners:

- Winner 1 -> Group with the highest R2 score
- Winner 2 -> Group with the most correct prediction of their own R2 score

BOTH CAN BE THE SAME GROUP

Iron Kaggle - Results

	Group 1:	Group 2:	Group 3:	Group 4:	Group 5:
R2 Prediction (how much your thought your model would perform)					
R2 Score (how your model performed with real life data provided by Joao)					
R2 Difference					



Good luck

Let's get started!