kaggle DAYS MEETUP

Introduction to kaggle

[where you get to solve "real-world" problems]

| Date : 15th June , 2019 | Venue : Grofers , Gurugram | Speaker : Akarsh Raj |

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

WHO AM I?

- **Akarsh Raj**
  - Data Scientist @ ZS Associates
  - Former Data Scientist @ Xceedance
  - B.Tech | CSE | University of Petroleum & Energy Studies
  - Kaggle Competitions Expert - Top 2 percent
    - 5 Competition medals – 1 Silver, 4 Bronze (Still waiting for my first gold)
    - 12 Discussion medals – 2 Gold, 10 Bronze

- Hobbies – Kaggle, Cryptography, Hacking, Soccer, Poker

- Where can you find me? (Click on the thumbnails below)

POLL

How much do you know about Kaggle?

- **Category 1** : I'm just curious about it
- **Category 2** : I registered to have a look at data, kernels and discussions
- **Category 3** : I have joined a real world data competition
- **Category 4** : It has already turned to an addiction

Why Kaggle?

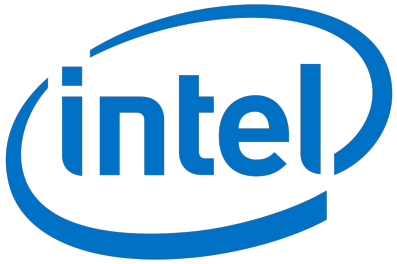
kaggle DAYS MEETUP

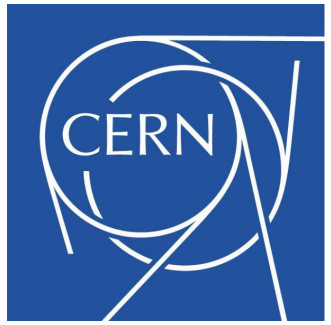
- Leading platform for ML related competitions, discussions since 2010
- Top companies post their real world problems which can be solved using ML, DL or some magic maybe.
- Google acquired Kaggle in 2017
- One can find implementations of almost all the algorithms, in most creative ways possible (Refer to the awesome kernel by Shivam Bansal who tried to collate implementations of various algorithms)
- Gets to meet the best data scientists from all over the world

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

FEW TOP SPONSORS

kaggle DAYS MEETUP





kaggle DAYS MEETUP

WHAT ALL YOU CAN DO AT KAGGLE ?

COMPETE

(Competition Tier)

SHARE

(Kernels Tier)

DISCUSS

(Discussions Tier)

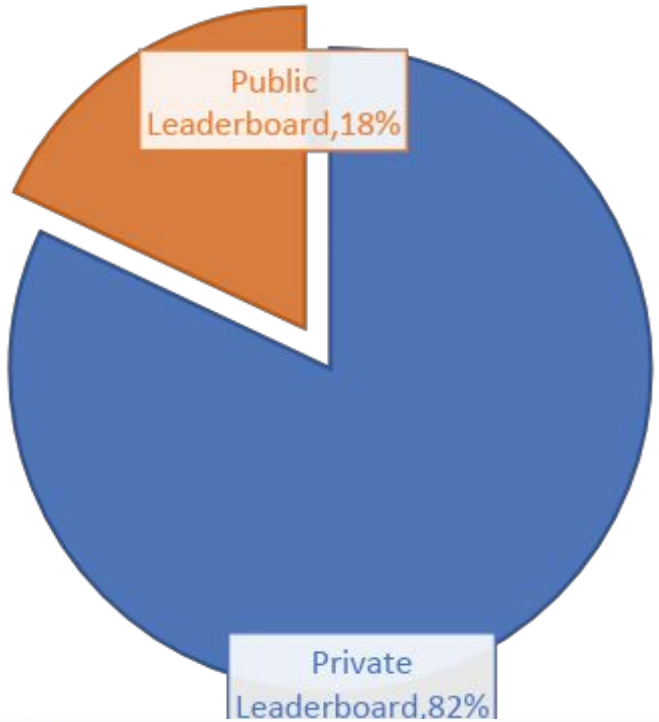
GROFERS Google LOGICMI kaggle CODING NINJAS www.codingninjas.in

COMPETITION

kaggle DAYS MEETUP

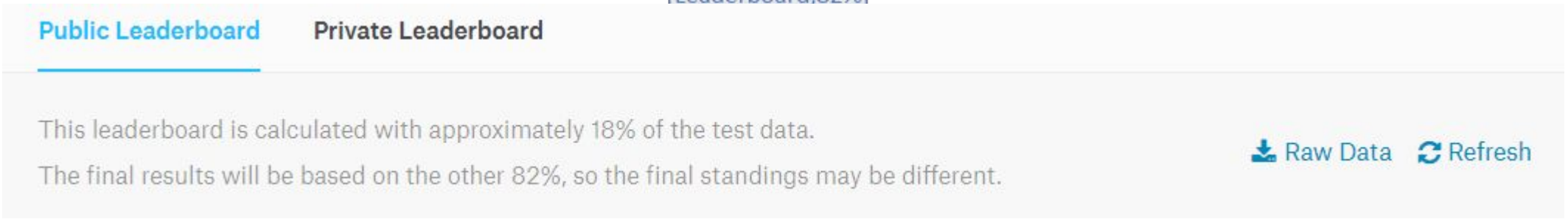
Public Leaderboard – **Test Data**

Private leaderboard – **Unseen Data**



NOTE: *We don't know which data points are in private or public.*

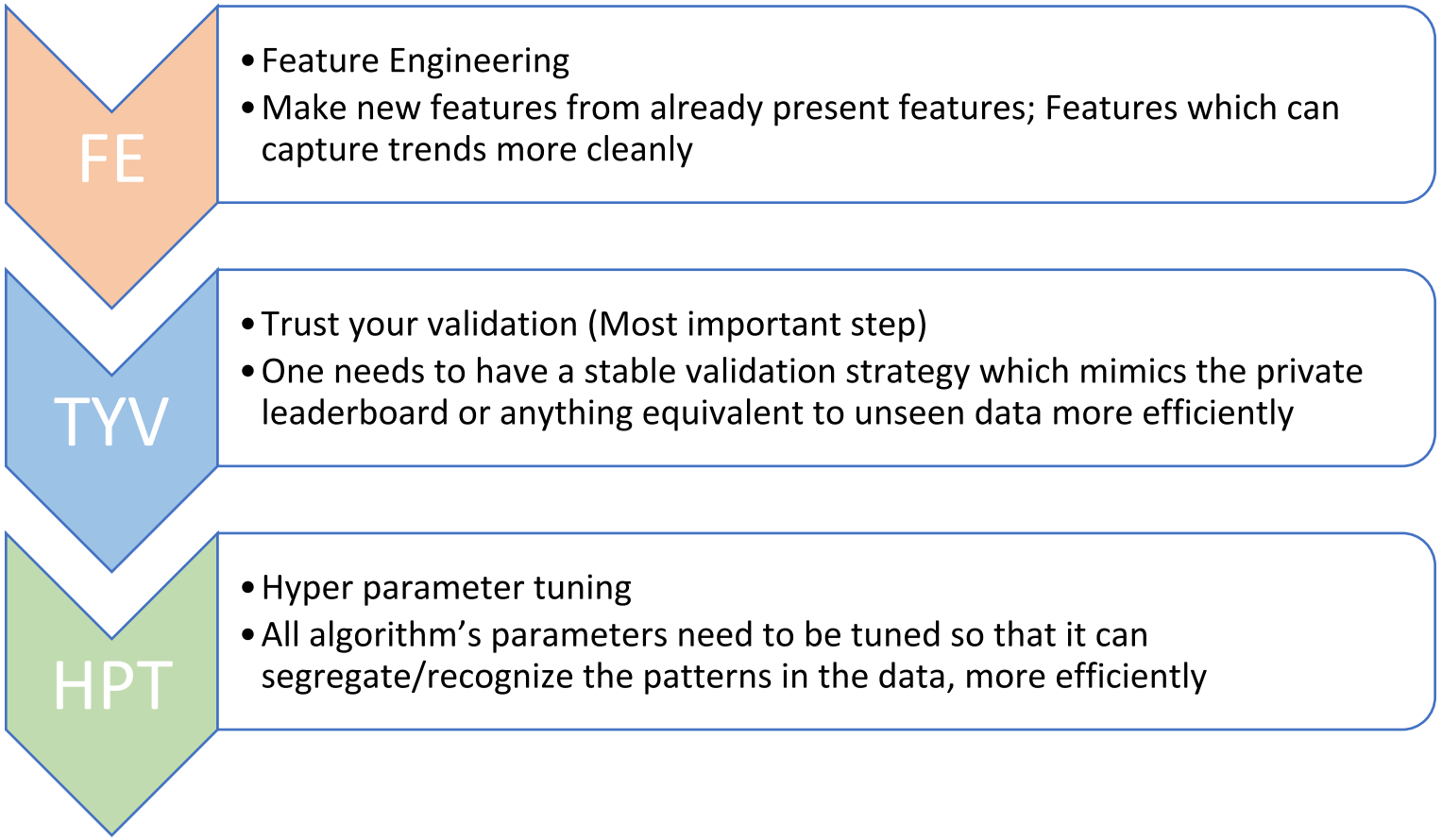
No submission of code/model, until one is in the money zone.



GROFERS Google LOGICMI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

MOST IMPORTANT THINGS TO BE DONE IN ANY COMPETITION



GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

HOW I APPROACH A KAGGLE COMPETITION ?

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

Data Description

Data fields

Each row of the training data contains a click record, with the following features.

- `ip` : ip address of click.
- `app` : app id for marketing.
- `device` : device **type** id of user mobile phone (e.g., iphone 6 plus, iphone 7, huawei mate 7, etc.)
- `os` : os version id of user mobile phone
- `channel` : channel id of mobile ad publisher
- `click_time` : timestamp of click (UTC)
- `attributed_time` : if user download the app for after clicking an ad, this is the time of the app download
- `is_attributed` : the target that is to be predicted, indicating the app was downloaded

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

BIGGEST CHALLENGES

- **Data Size** – More than 175 millions rows which corresponds to 175 million clicks
- It was a **binary classification problem** where number of 1s i.e. positive samples (identified frauds) constitute of **less than 0.5 percent** of the total training data.

WHAT WORKED FOR ME

- **Magic feature –** Any feature which give more than average boost to the model is mostly termed as magic feature on Kaggle. (In this case, delta of timestamps was the one)
- **Down-sampling** worked wonders for me (Maybe we got lucky, we need that luck sometimes, on Kaggle)

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

- 200 columns/features
- All column names were redacted as var0, var1, var2, var3, …….
- No context or background of any columns
- More of a **HIT AND TRIAL** at the start
- Hunt for the Magic feature began again

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

WHAT WORKED FOR ME ?

- *Numerous experiments* – Number of experiments carried out in this competition just to let the data talk was huge
- Peaks in the distribution of few of the variables
- Value counts turned out to be the magic feature

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

CAN YOU LEVERAGE KAGGLE EXPERIENCE INTO YOUR DAY JOB ?

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

- ✔ **YES, Definitely!**
- ✔ You get to the know about the best practices in your field. For e.g. *how to avoid overfitting*
- ✔ Habit of doing the *feature engineering* on every dataset is very healthy
- ✔ You're updated on latest developments in the field of Data Science
- ✔ You get to learn from the best data scientists across the globe, on how to approach
  - ✔ A classical ML problem
  - ✔ Computer vision problem
  - ✔ NLP problem

kaggle DAYS MEETUP

WHAT CHANGES FROM KAGGLE TO DAY JOB WORK?

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

- You **never get cleaned & prepared dataset** in day job, the way you get on Kaggle (If you do, tell me your company name :P)
  - One needs to spend major amount of time preparing or cleaning the dataset

- If you look at the bigger picture, you don't have to fight for **minor increase in performance** but you have to fight for it on Kaggle

- On kaggle, you don't always have to have a model which can be explained, whereas business mostly requires a model which has explainability factors attached to it. For e.g. using Shapley values to explain the contribution of each feature in the prediction at local level as well as global level

kaggle DAYS MEETUP

WILL KAGGLE HELP YOU GET A JOB?

GROFERS Google LOGICNI kaggle CODING NINJAS www.codingninjas.in

- There is a job board on Kaggle too, though no jobs in India were posted on that portal yet, as far as i know
- Kaggle on its own won't help you get a job directly, but the experience, the peer group you get on kaggle is unmatched
- If you happen to win a medal, that's a plus for your resume
- You get to meet people from the community who might guide you to a good opportunity

kaggle DAYS MEETUP

AWESOME RESOURCES (FREE)

GROFERS Google LOGICMI kaggle CODING NINJAS www.codingninjas.in

kaggle DAYS MEETUP

CLICK ON THE PICTURES

fast.ai

Making neural nets uncool again

GROFERS Google LOGICMI kaggle CODING NINJAS www.codingninjas.in

Thank you !!

Q/A Session