# DOTA2



## **Dataset and Context**

- 50.000 games of Dota2
- 18 different .csv files
- Two types of data:
  - Match summaries
  - In-game data

• Looking at a single game (sampled):

#rows/#columns

-	"match.csv":	1/13
-	"players.csv":	10/73
-	<pre>"player_time.csv":</pre>	56/32
-	"objectives.csv":	28/9
-	"ability_upgrades.csv":	207/5
-	"teamfights.csv":	13/5
-	"teamfights_players.csv":	130/8
-	"purchase_log.csv":	434/4
-	"chat.csv":	35/5

- Total: 914/~150

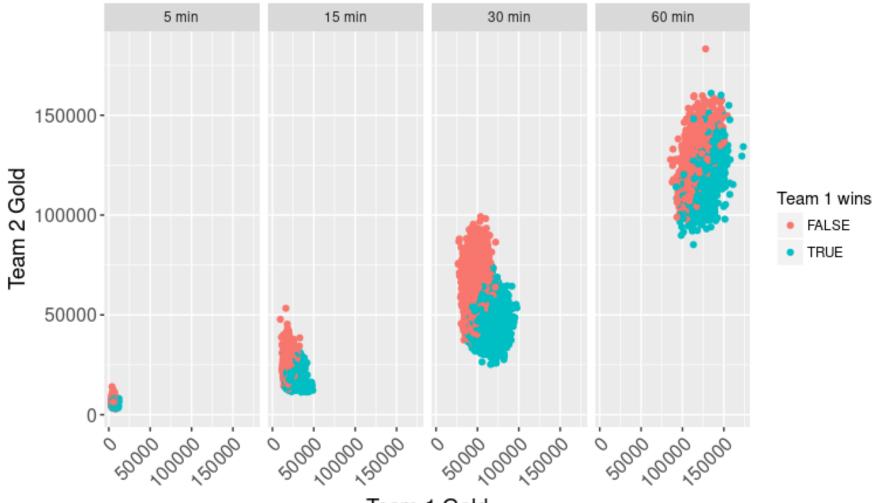


# Ideas & Challenges

- General win-prediction:
  - Use player history and team composition to predict wins
- On-line win-prediction:
  - Use all available in-game data to predict
  - Update prediction each time new information is available
  - Challenge:
    - Very different types of information
    - Many variables
    - Information is not equispaced
- On-line prediction for events in general:
  - Similar approach as above, while we do not limit us to the binary outcome "win"



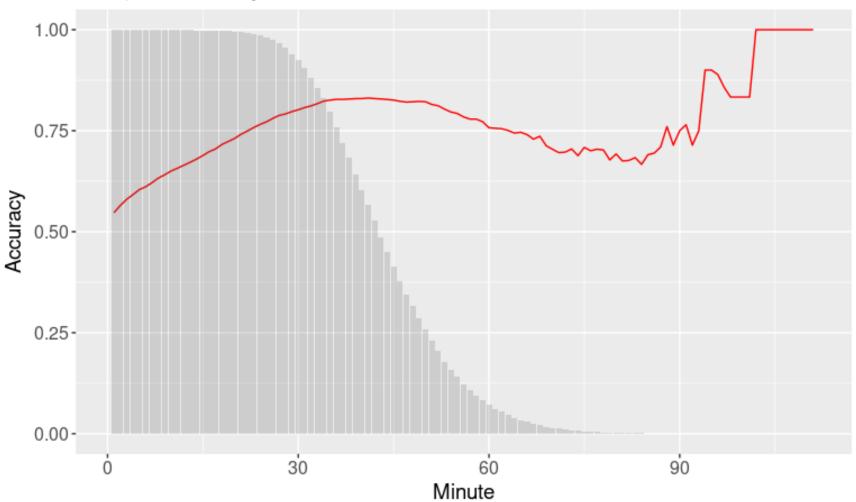
#### Gold per team



Team 1 Gold



## Win-prediction by gold lead

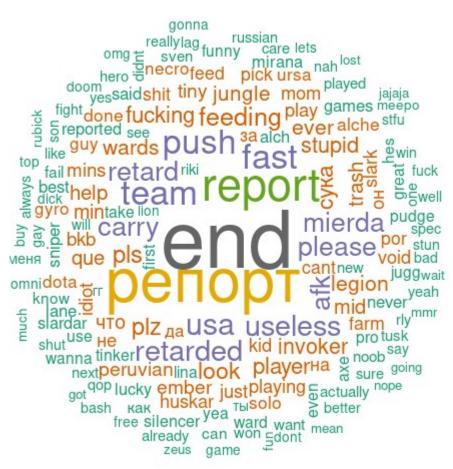




# How relevant is the chat?

```
help fucking peruvian
                             cant alche lucky playing
                      tiny on use noob "
               he wannacare nahriki game hes pudge useless better piz void
       mierdagreat rubickokay glhf
                                      sorry sad noobs actually
                            tho die def life hatewirdont lane
gameswin done of dude stop of luck
                                                           Imao got can 5
      y just 💆 🎅 saneed jaja run nice bot friend 🚅 🕏 funny
       push going dazzle love coming mean sven report legion
               best omni thats = ты still,
                                  still son o next que doom one russian o sinana farmy o huskar please
                 necro didnt fight
                 player sniper
                    retard ember mirana
```

Winning teams



Losing teams