

# PREDICTING DOTA 2 MATCHES USING MACHINE LEARNING

KYLE POMERLEAU 8/12/2015

# PROJECT QUESTION

- IS IT POSSIBLE TO PREDICT THE MATCH OUTCOME BASED ON HERO SELECTION?



# WHAT IS DOTA 2?

- “DEFENSE OF THE ANCIENTS”
- DOTA 2 IS MULTIPLAYER ONLINE BATTLE ARENA (MOBA) PC GAME
- TEN PLAYERS PLAY AT ONCE, FIVE ON A TEAM. EACH TEAM BASED IN A CORNER OF THE MAP





# WHAT IS DOTA 2?

- EACH TEAM OF 5 PLAYERS CHOOSES FROM 110 “HEROES” THAT HAVE UNIQUE ABILITIES.





# WHAT IS DOTA 2?

- EACH HERO BECOMES MORE POWERFUL OVER TIME BY GAINING EXPERIENCE. PLAYERS ALSO EARN GOLD AND PURCHASE ITEMS THAT INCREASE SKILL OF HEROES





# WHAT IS DOTA 2?



- A TEAM WINS WHEN THEY DESTROY THE OPPOSING TEAM'S ANCIENT



# DATA COLLECTION

- STEAM API
  - ACCESS TO DETAILS ON EVERY PUBLIC MATCH
- LIMITATIONS
  - COULD NOT FILTER BY SKILL
  - FILTER BY TYPE OF GAME INSTEAD (RANKED)

MATCH ID  
169177907

DIRE VICTORY

22:50

The Radiant

PLAYER	LEVEL	HERO	K	D	A	ITEMS	GOLD	LAST HITS	DENIES	GOLD/MIN	XP/MIN
LGD.Taobao.longdd	9	Shadow Shaman	0	2	7		1665	32	5	217	219
LGD.Taobao.Sylar	14	Phantom Lancer	1	0	6		3474	173	25	522	478
LGD.Taobao.Yao	12	Lifestealer	1	0	3		2932	70	12	319	358
LGD.Taobao.ddc	12	Brewmaster	5	0	3		3651	79	6	374	391
LGD.Taobao.xiao8	10	Keeper of the Light	2	0	2		1494	65	2	273	248

The Dire

PLAYER	LEVEL	HERO	K	D	A	ITEMS	GOLD	LAST HITS	DENIES	GOLD/MIN	XP/MIN
zEn th` xFreedom	8	Juggernaut	2	0	0		357	34	0	162	152
? zEn th` Ice	8	Nyx Assassin	0	3	1		524	13	4	123	151
? zEn th` N.W.P	12	Storm Spirit	0	3	0		667	129	13	365	379
zEn th` iceiceice	11	Lone Druid	0	2	1		517	80	4	276	307
standin.Meracle	12	Mirana	0	1	0		1168	145	8	376	370

CLOSE

# DATA COLLECTION

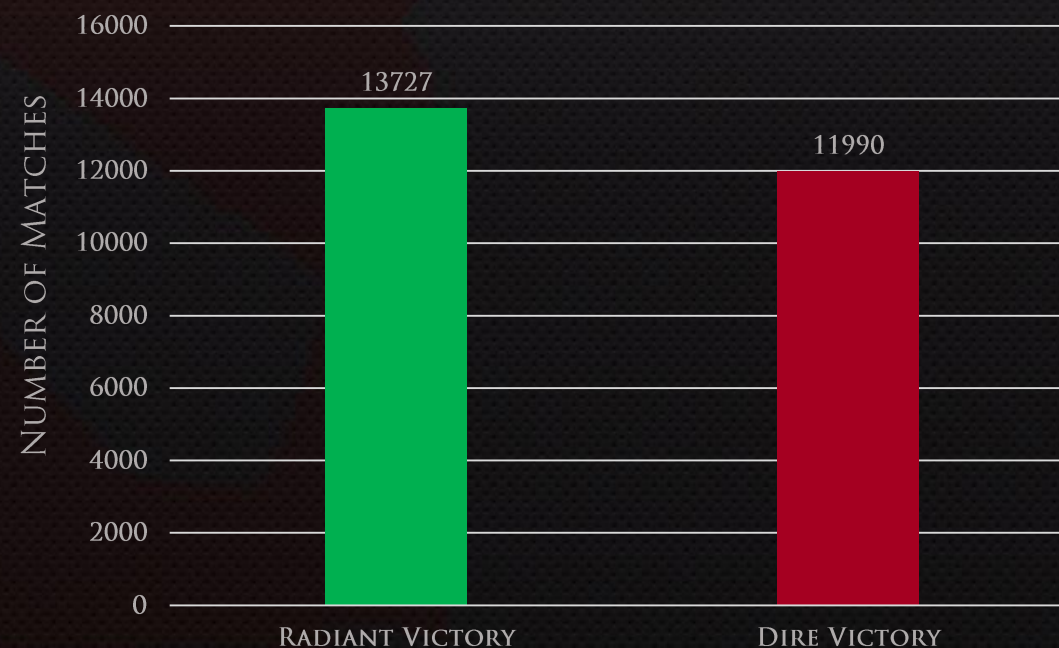
- DATA SET:
  - N = 25717 MATCHES
  - 110 HEROES REPRESENTED
  - 220 FEATURES
    - 2 FOR EACH HERO
    - ONE ON RADIANT, ONE ON DIRE
- Y VARIABLE:
  - 1 = RADIANT VICTORY



# DATA COLLECTION

- DATA SET:
  - N = 25717 MATCHES
  - 110 HEROES REPRESENTED
  - 220 FEATURES
    - 2 FOR EACH HERO
    - ONE ON RADIANT, ONE ON DIRE
  - Y VARIABLE:
    - 1 = RADIANT VICTORY
- SLIGHTLY IMBALANCED
  - 53.3% FOR 1

RADIANT TEAM HAS A  
SLIGHT EDGE





# MODELING DOTA 2 MATCHES

- FIRST MODEL USED:
  - LOGISTIC REGRESSION



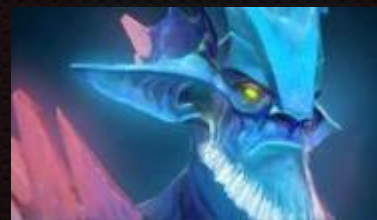
# MODELING DOTA 2 MATCHES

- FIRST MODEL USED:
  - LOGISTIC REGRESSION
- RESULTS:
  - ACCURACY: 61.3%
  - AUC: 64.7%



# MODELING DOTA 2 MATCHES

- LIMITATIONS OF LOGIT:
  - DOESN'T AUTOMATICALLY ACCOUNT FOR INTERACTIONS





# MODELING DOTA 2 MATCHES

- FIRST MODEL USED:
  - LOGISTIC REGRESSION
- SECOND MODEL USED:
  - RANDOM FOREST
- RESULTS:
  - ACCURACY: 61.3%
  - AUC: 64.7%



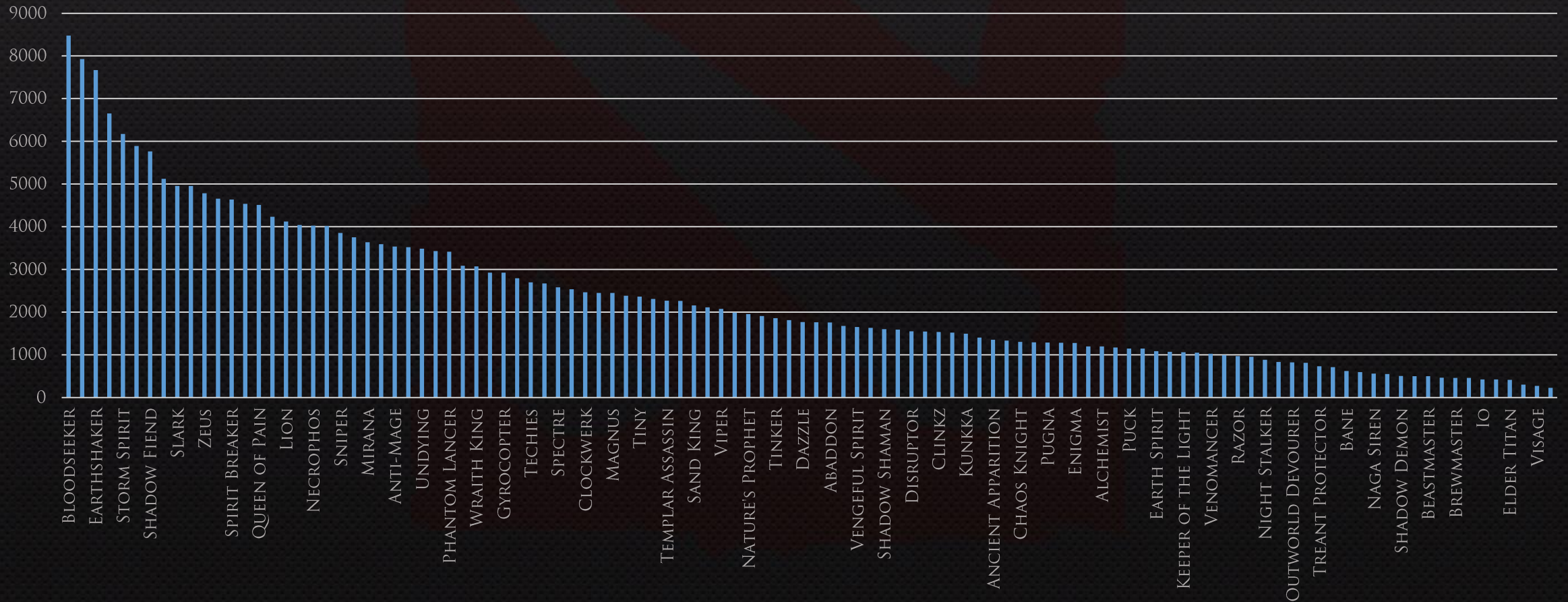
# MODELING DOTA 2 MATCHES

- FIRST MODEL USED:
  - LOGISTIC REGRESSION
- RESULTS:
  - ACCURACY: 61.3%
  - AUC: 64.7%
- SECOND MODEL USED:
  - RANDOM FOREST
- RESULTS:
  - ACCURACY: 60%
  - 450 – 500 ESTIMATORS



# MODELING DOTA 2 MATCHES

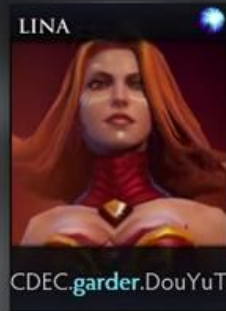
HERO USE IS VERY SKEWED





# IS MY MODEL USEFUL?

CDEC Gaming



VS.



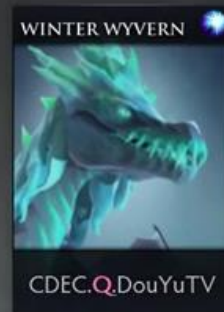
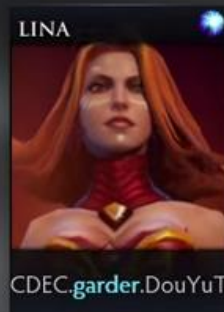
Evil Geniuses





# IS MY MODEL USEFUL?

CDEC Gaming



VS.



Evil Geniuses



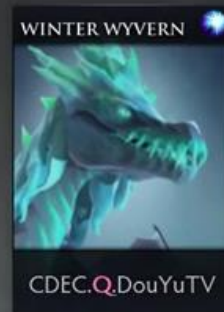
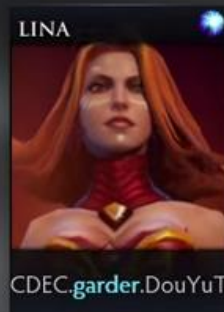
37%

63%



# IS MY MODEL USEFUL?

CDEC Gaming



VS.



Evil Geniuses



37%

63% ✓

# ISSUES/FUTURE WORK

- DATA COLLECTION WAS LESS THAN IDEAL
  - WAS NOT ABLE TO FILTER BY SKILL LEVEL
  - LOW-SKILLED GAMES ARE PROBABLY RANDOM, BUT SYSTEMATICALLY BIASED TOWARDS SPECIFIC HEROES
- COMMON SUPPORT
  - SOME HEROES ARE NOT USED VERY MUCH
  - SOME HERO COMBINATIONS MAY NOT EXIST IN DATASET
- BETTER DATA COLLECTION WOULD HELP





QUESTIONS?