Defense of the MOBA Leaderboard

An examination of the popularity of MOBA games through text classification

Tan Song Kai
DSIF-SG-9

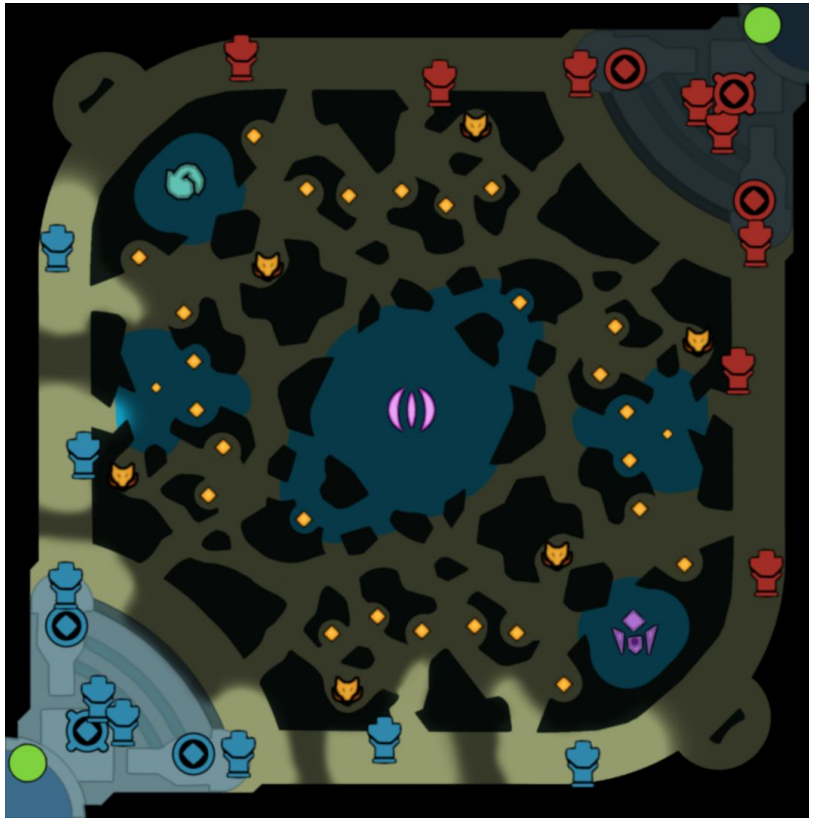
Agenda

- Background & Problem Statement
- Methodology
- Exploratory Data Analysis
- Modelling
  - ❏ Results
  - ❏ Insights
- Implementation
- Recommendations
- Future Work

What is MOBA?

Defense of the Ancients (Dota 2)

League of Legends (LoL)



Background

MOBA Games Leaderboard

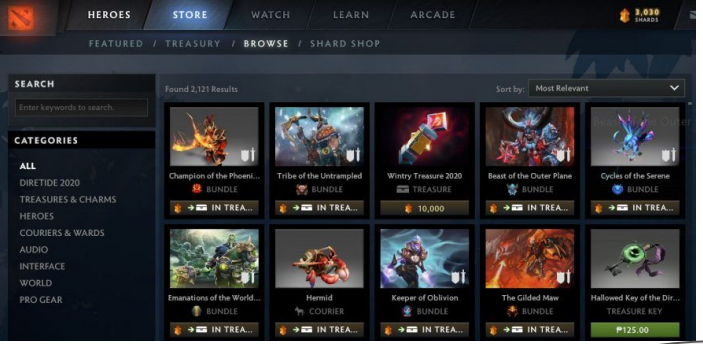
#	Name	Viewer Hours	Hours Streamed	ACV	Creators	Streams
1	League of Legends	34,013,220	626,620	202,459	101,431	296,173
2	VALORANT	26,740,605	1,161,834	159,170	202,578	581,062
3	Dota 2	10,041,905	193,946	59,773	25,974	86,452

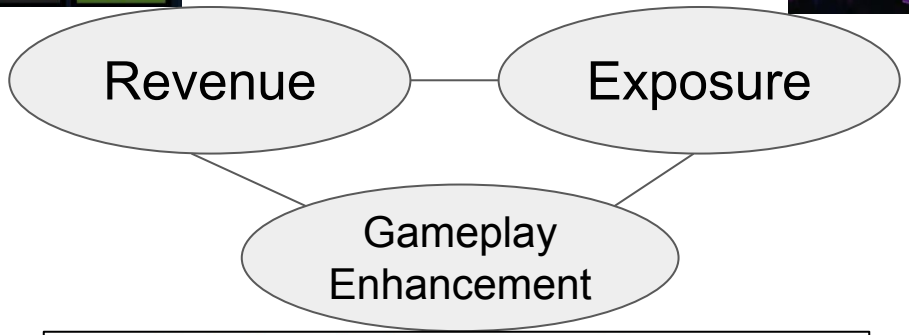
Source: https://gamesight.io/

Problem Statement

What can we learn from Dota 2 and LoL posts to attract existing and new gamers to play Dota 2?

Motivation



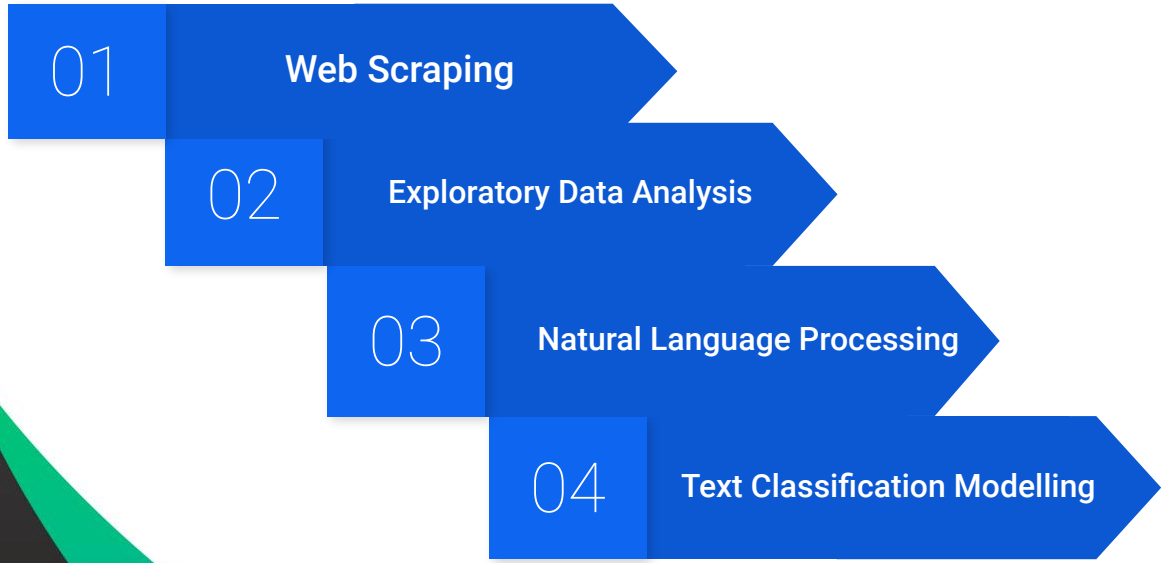


OpenAI's Dota 2 AI steamrolls world champion e-sports team with back-to-back victories

/ The International 2018 champion OG loses to OpenAI's bots in a stunning defeat

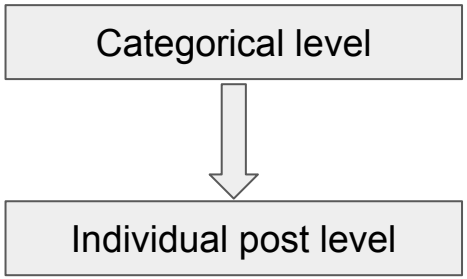
By NICK STATT / @nickstatt
Apr 14, 2019, 5:05 AM GMT-8 | 0 Comments

Methodology

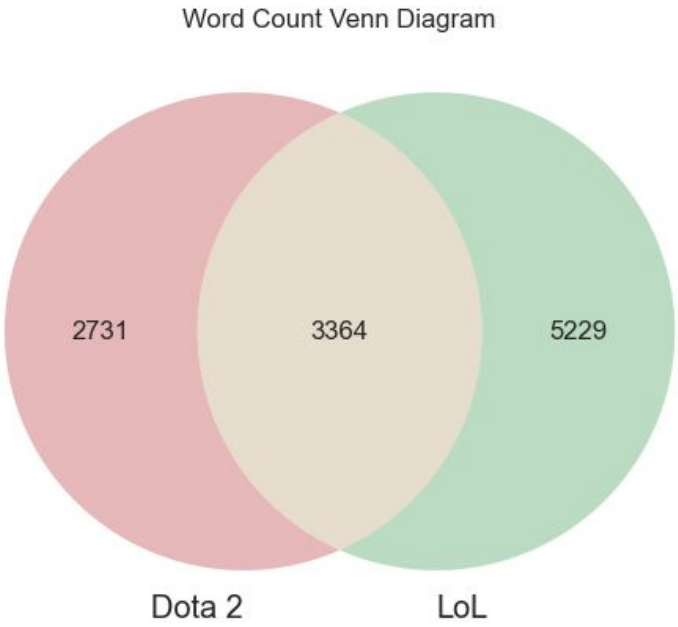


Web Scraping

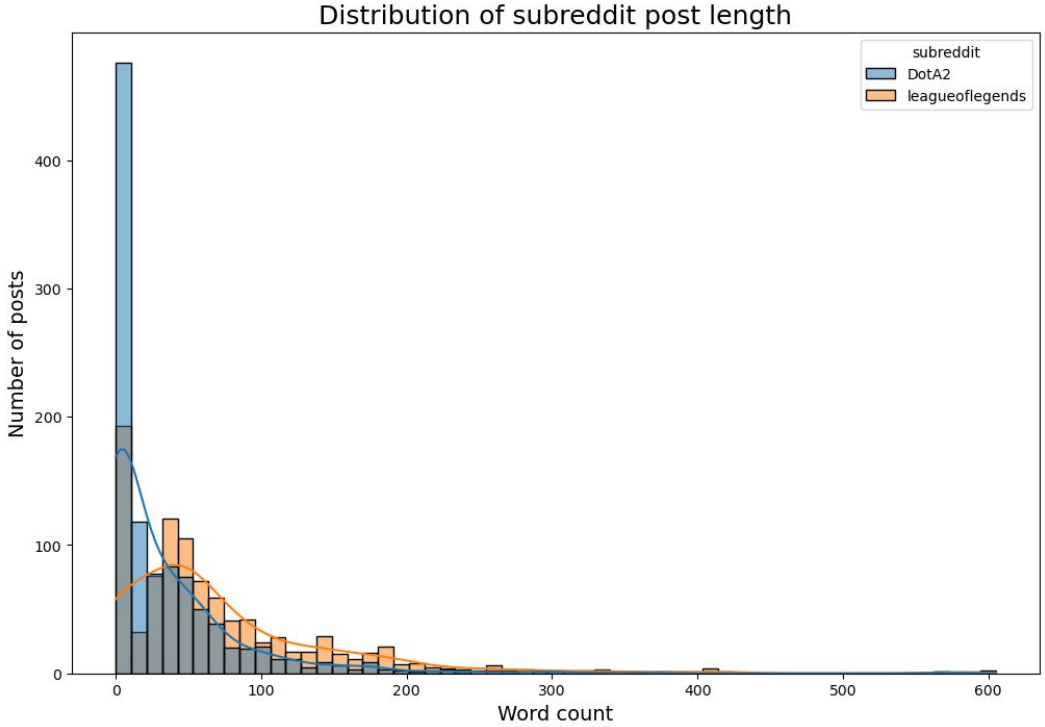
- r/DotA2 and r/leagueoflegends



Exploratory Data Analysis

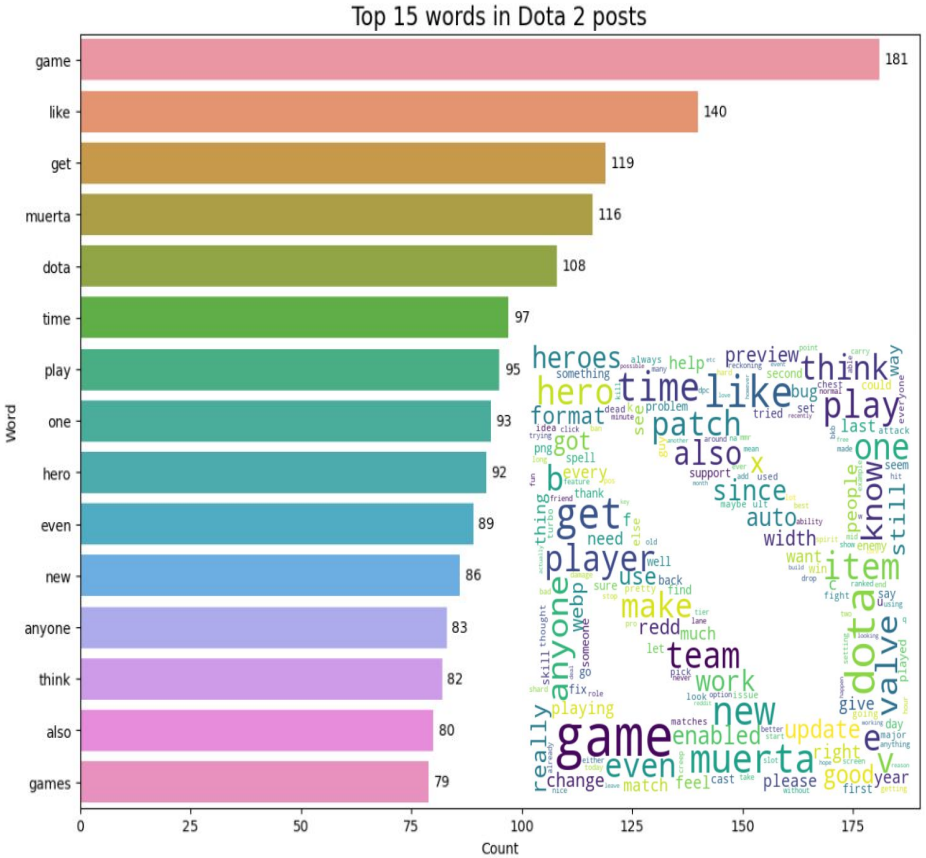


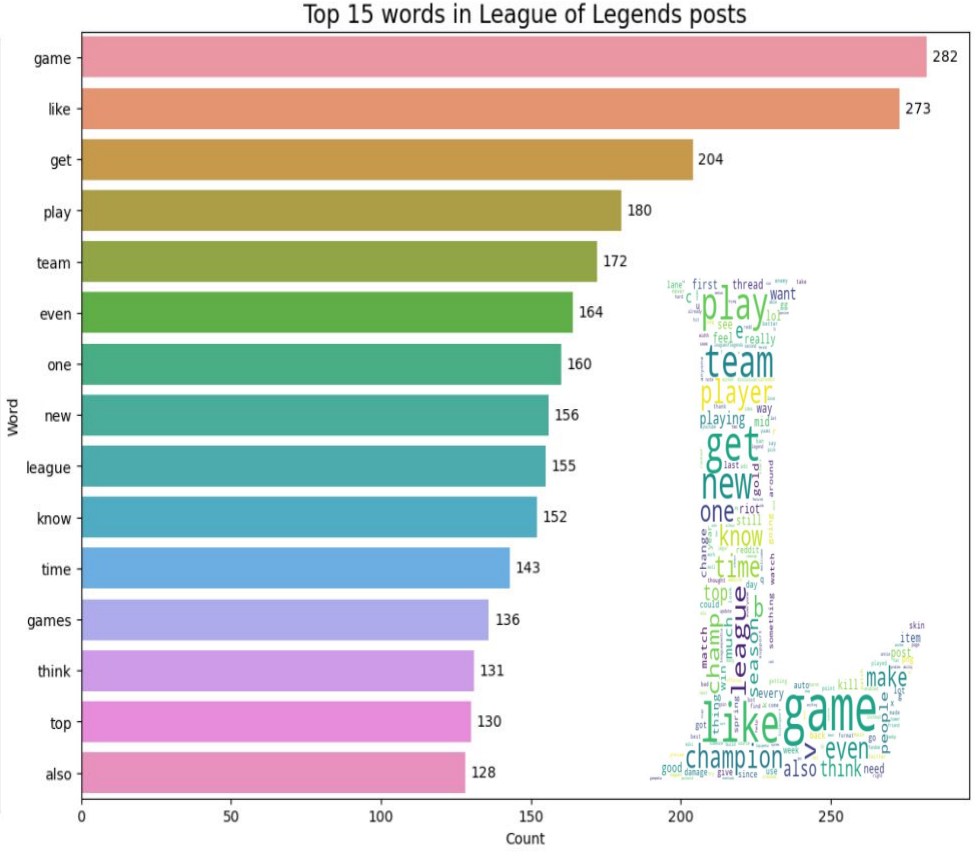
strat	streak	guldborg
alliance	find	crab
tping	always	alice
arcwarden	rid	transition
nisha	price	oneshotting



subreddit	count	mean	min	25%	50%	75%	max
DotA2	1057	34.2	0.0	0.0	16.0	48.0	584.0
leagueoflegends	983	72.1	0.0	25.0	50.0	93.5	605.0

Exploratory Data Analysis





Modelling Approach

- Punctuation
- Lowercasing
- Customized stopword removal
- Tokenization
- Lemmatization and stemming

Input

```
hand big lina thank creator new set move good
creep muerta showcas abil
durat solut long game buyback mechan
affect coupl pro also realli envi quinn mate a...
screen ever ui rememb feel exist useless
```

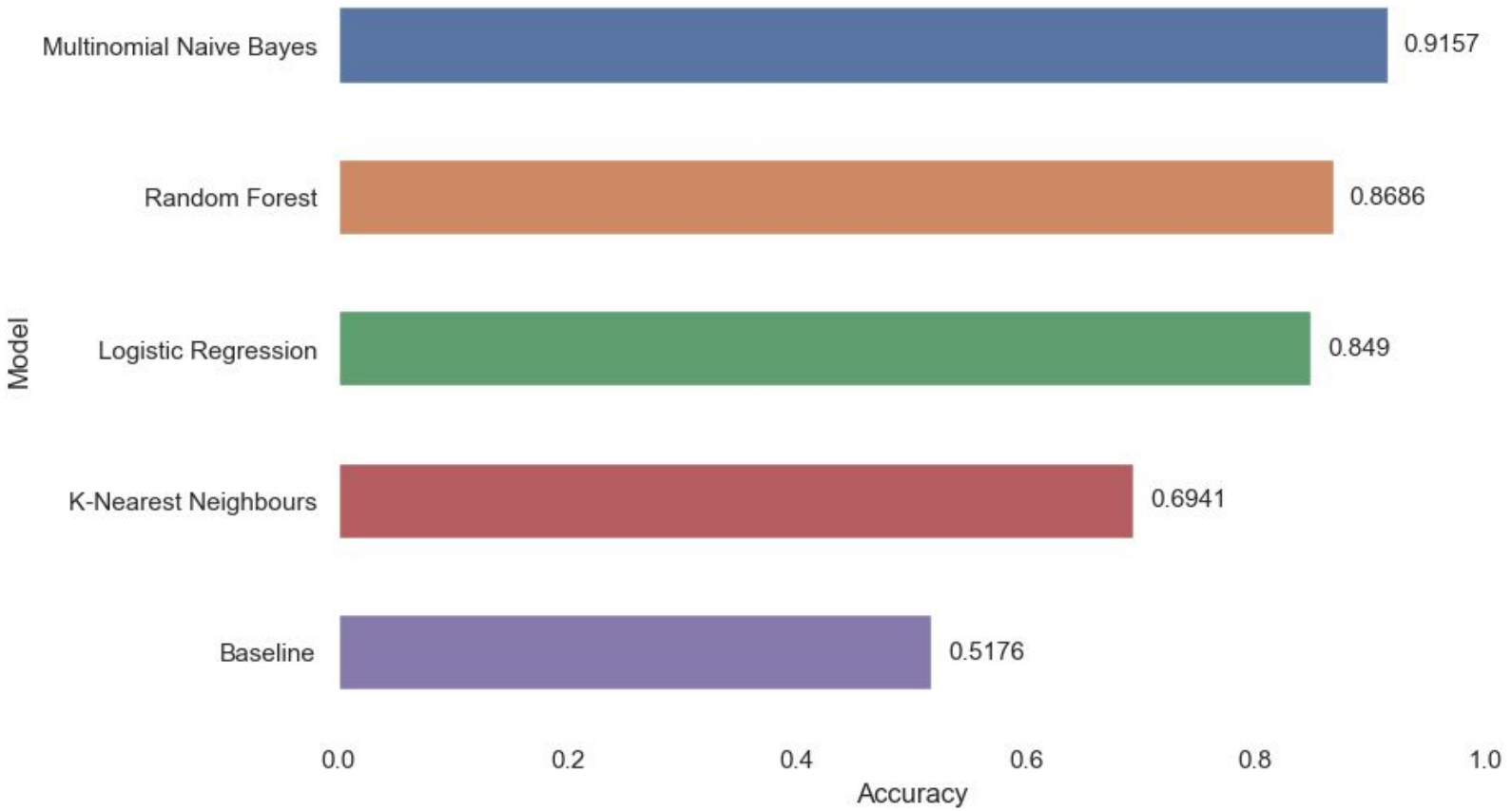
Output

Transformer

Classifier

aa	aatrox	ab	abaddon	abandon	abil	abl	absolut	absurd
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0

Model Accuracy

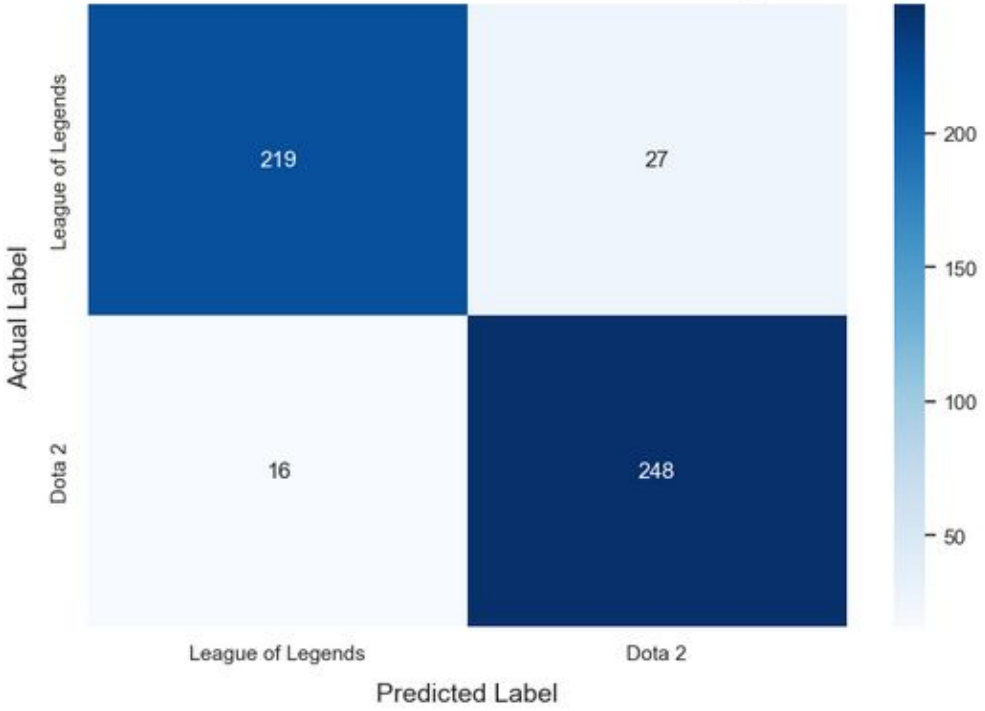


Modelling Results

Baseline score: 0.5176

Model	Training Score	Test Score	Sensitivity	Specificity	ROC-AUC Score	Hyperparameters
Multinomial Naive Bayes	0.9542	0.9157	0.9394	0.8902	0.9717	'cvec__max_df': 0.8, 'cvec__max_features': 3200, 'cvec__min_df': 2, 'cvec__ngram_range': (1, 1)
Logistic Regression	0.9876	0.8843	0.9091	0.8577	0.9634	'cvec__max_df': 0.8, 'cvec__max_features': 3800, 'cvec__min_df': 2, 'cvec__ngram_range': (1, 1), 'logreg__C': 10, 'logreg__penalty': 'l1'
Random Forest	0.9974	0.8706	0.9091	0.8293	0.9385	'cvec__max_df': 0.8, 'cvec__max_features': 3500, 'cvec__min_df': 4, 'cvec__ngram_range': (1, 1)

Multinomial Naive Bayes



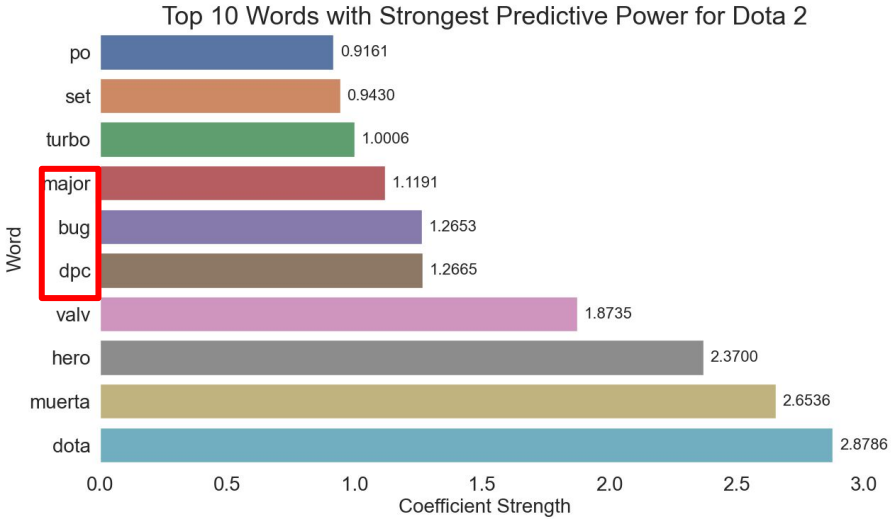
LoL	
champion	0.988042
champ	0.984953
riot	0.980955
adc	0.969872
yuumi	0.963357
elo	0.958456
leagueoflegend	0.955479
ash	0.950776
baron	0.948031
anni	0.948031

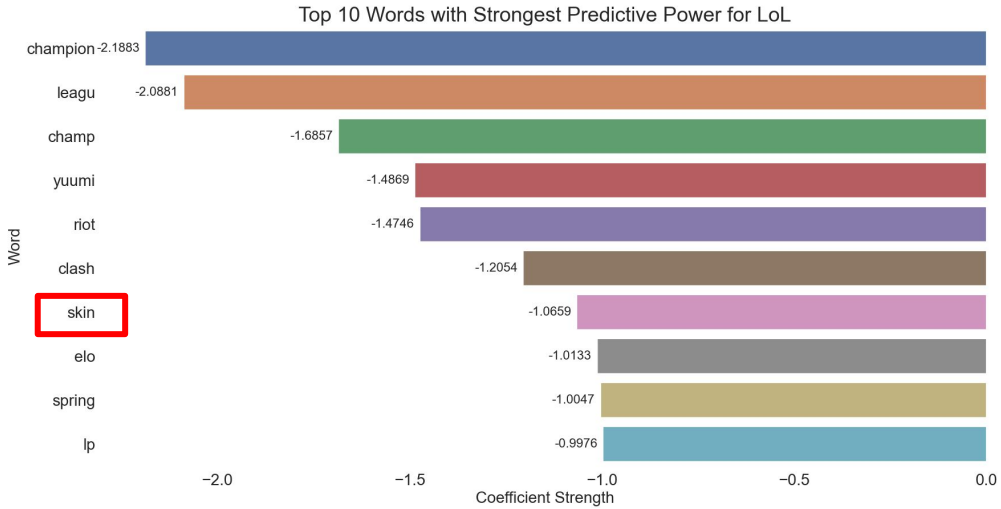
Dota 2	
valv	0.991936
muerta	0.985706
bkb	0.979781
dpc	0.977204
dota	0.975653
fp	0.971055
quickcast	0.969405
hero	0.967771
offlan	0.965466
hammer	0.965466

Game characters
Game items
Game events

Game ranking
Game issues

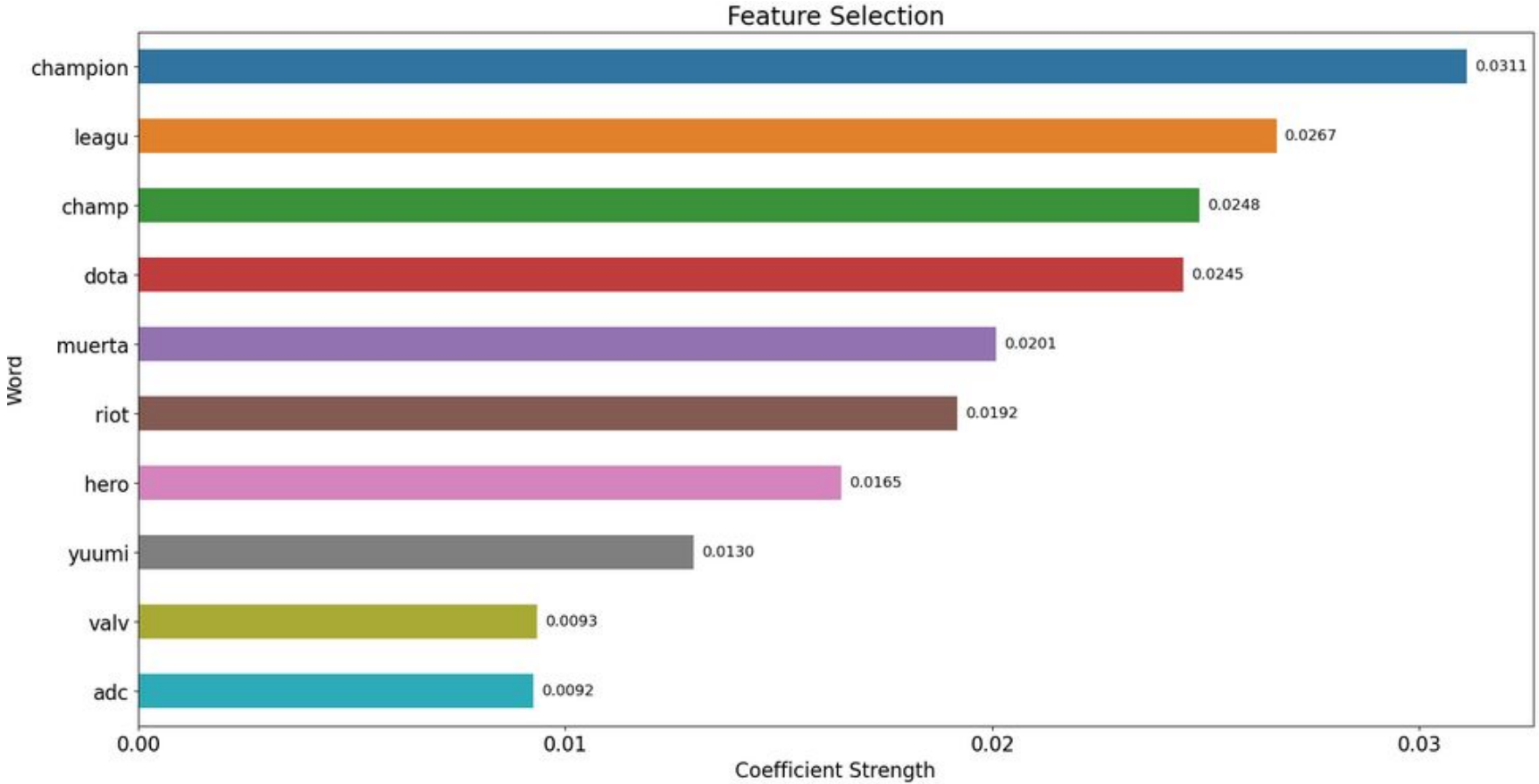
Logistic Regression





Game bugs
Character skins

Random Forest



Implementation of MNB:

$$P\left(Subreddit_j \mid \bigcup_{k=1}^{n} Word\_status_k\right) = \frac{P(Word\_status_1 \mid Subreddit_j) * P(Word\_status_2 \mid Subreddit_j) * \ldots * P(Word\_status_n \mid Subreddit_j)}{P(\bigcup_{k=1}^{n} Word\_status_k)}$$

$Subreddit_j$: *Dota 2 or LoL post*

$\bigcup_{k=1}^{n} Word\_status_k$: *combination of words, eg. Word A exists, Word B doesnt exist, Word C exists …*

$P(Word\_status_k \mid Subreddit_j)$: *probability that word k exists/does not exist given that subreddit type is known*

Probability formula can be set up in Excel

Recommendations

Model	Possible Opportunities
Multinomial Naive Bayes	● Revisit MMR system ● FPS issue troubleshooting
Logistic Regression	● Reward system for character skin design and bug testing ● Branding of tournaments (commonly known as dpc)
-	● Picks/bans rework ● Map changes

Future Work

- Text data collection
- Stopword optimization
- Model evaluation on different transformers/classifiers
- Hyperparameter optimization

THANK YOU

Q&A