CLASSIFICATION MODEL FOR BOARDGAMES AND MOBILEGAMES

Jasper

Contents

Problem Statement

As a Data Scientist in a new company which aims to develop both mobile games and board games, building a classification model that can correctly classify these 2 categories based on posts/comments on social media platform

This project aims to make use of the posts by the users in the 2 subreddit groups namely 'boardgames' and 'MobileGaming' for the modeling of this classification model

Data Cleaning

	subreddit	selftext	title	created_utc	author
0	boardgames	\- 4 players in 2h \n\- plays nice with 2 pla...	Best strategy game that doesn't take more than...	1633032711	radicalrj
1	boardgames	I am a fifth grade teacher who in the long ter...	Looking to use board games at school:	1633032651	concernedgumbo
2	boardgames	Parents are retired and only play the traditio...	Recommendations: quick (20-30 minutes) games f...	1633031701	Turbodong
3	boardgames	Got to play Res Arcana last night (I know, I'...	I really liked Res Arcana!	1633031451	kryzak123
4	boardgames		I really liked Res Arcana!	1633031095	kryzak123

empty string

	subreddit	selftext	title	created_utc	author
29	boardgames	[removed]	God of War Ragnorak: Recent Delay Of Sequel Ex...	1633003464	Wppppp002
49	boardgames	[removed]	TGG Bishops Arts left today	1632958467	Coffeelatte4567
162	boardgames	[removed]	The pro-democracy movement has never stopped	1632834716	Sad_Anxiety_6033
174	boardgames	[removed]	Photoghasts : A Haunted Card Game : Kickstarte...	1632823425	manoghosts

removed post

moderator post

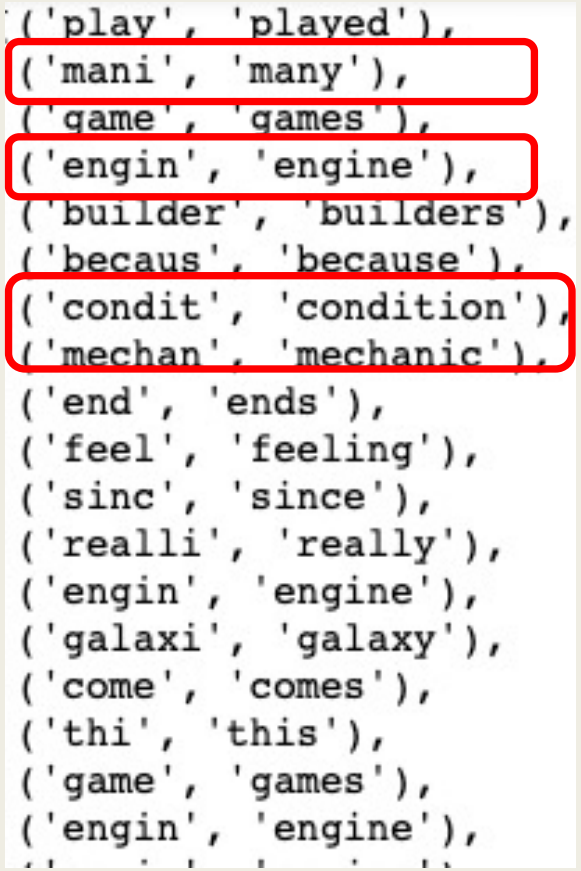
	subreddit	selftext	title	created_utc	author
39	boardgames	**Welcome to /r/boardgames's Daily Discussion ...	Daily Discussion and Game Recommendations Thre...	1632978073	AutoModerator
40	boardgames	The BGG database is enormous and getting bigge...	Forgotten Favorites & Hidden Gems - (Septe...	1632978017	AutoModerator
117	boardgames	**Welcome to /r/boardgames's Daily Discussion ...	Daily Discussion and Game Recommendations Thre...	1632891677	AutoModerator
118	boardgames	What are your favourites when you're playing s...	One-Player Wednesday - (September 29, 2021)	1632891615	AutoModerator
187	boardgames	**Welcome to /r/boardgames's Daily Discussion ...	Daily Discussion and Game Recommendations Thre...	1632805276	AutoModerator

	subreddit	selftext	title	created_utc	author
820	MobileGaming	I play on my phone a lot and I wanted to try s...	Good singleplayer reverse horror games?	1628439629	I-hate-spicy-foods
821	MobileGaming	I was playing some old mobile games for nostal...	Does anyone remember playing age of war2? (Imp...	1628438567	Barrelrider420
822	MobileGaming	[deleted]	Netflix Gaming Service | Techno Tunes	1628436609	[deleted]
823	MobileGaming		Mobile Gaming News Weekly EP 30 - We saw 17 an...	1628434995	Tousif_03

deleted post

Data Cleaning – Lemmatizing vs Stemming

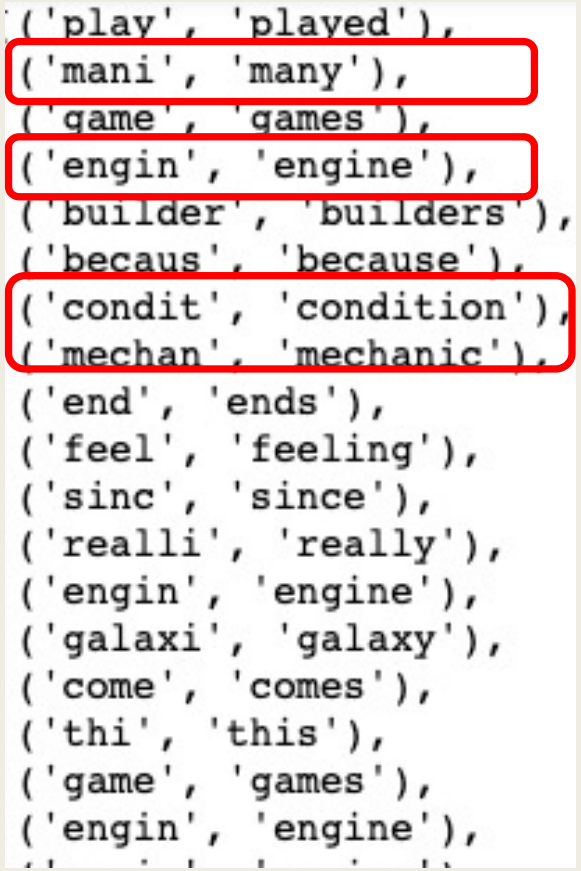
stemming



lemmatizing

Data Cleaning – Lemmatizing vs Stemming

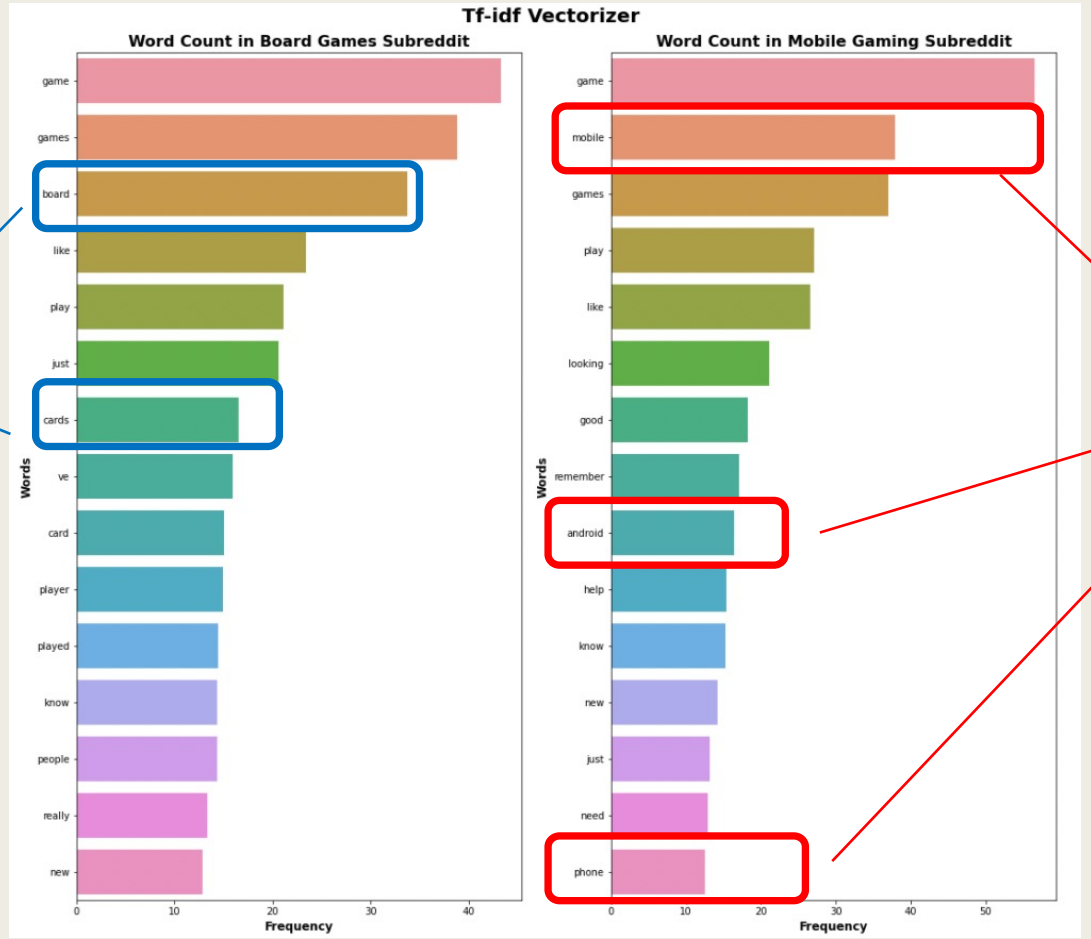
stemming



lemmatizing

Lemmatizer CHOSEN

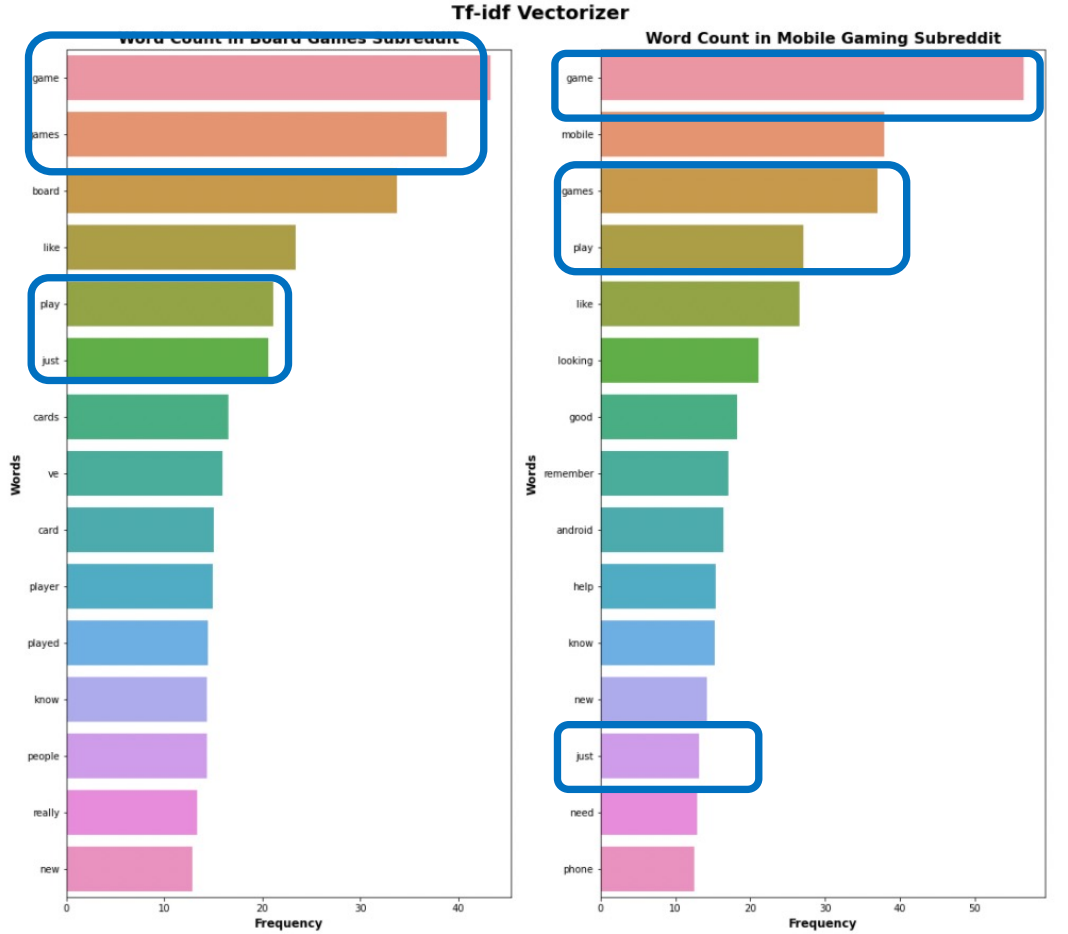
Data Cleaning - Stopwords



Boardgames

Mobile games

Data Cleaning - Stopwords



Other words like 'games', 'play', 'new' , 'like', 'just' might not be helpful for classification

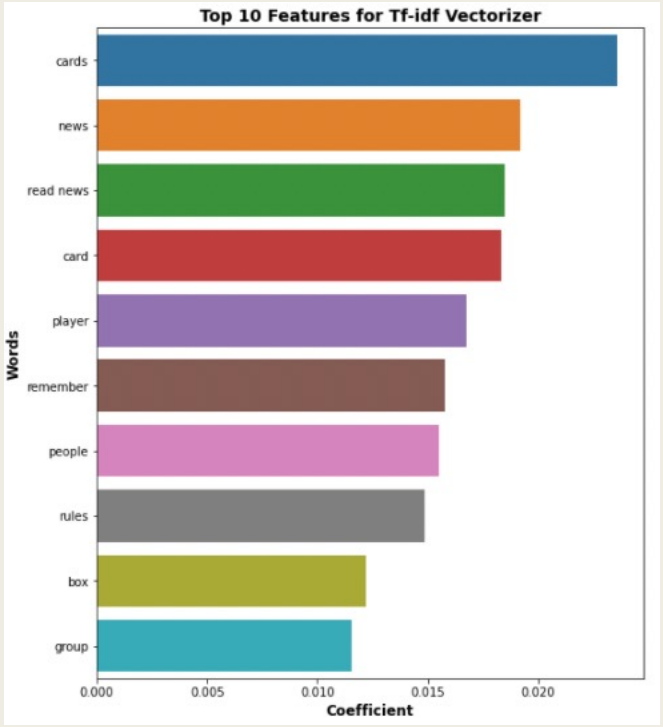
Model Comparison - Logistic Regression vs Random Forest

	Train CV	Accuracy	Sensitivity	Specificity	Precision	F1_score
Logistic Regression	88.6%	85.7%	85.3%	86.1%	85.7%	85.5%
Random Forest	85.8%	85.5%	82.8%	88%	87.1%	84.9%

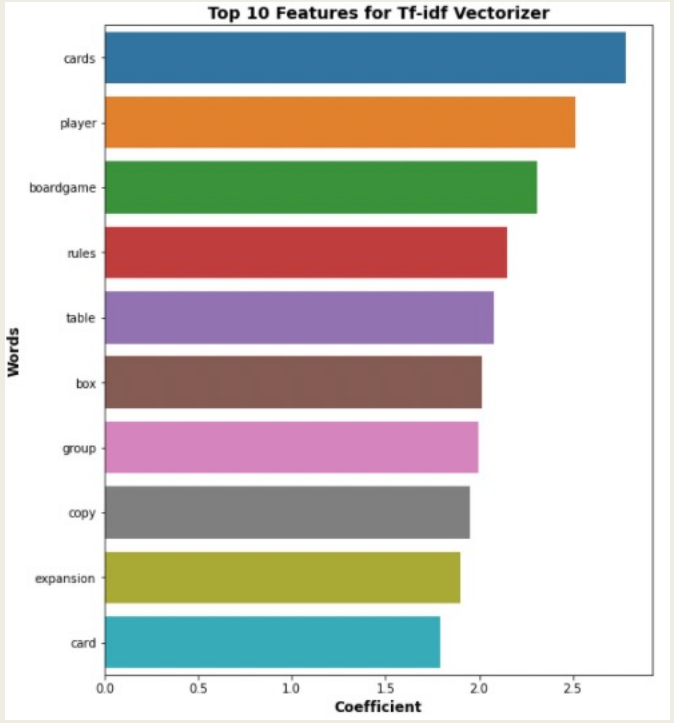
Sensitivity & Specificity IMPORTANT METRICS

Model Comparison – Feature Importance

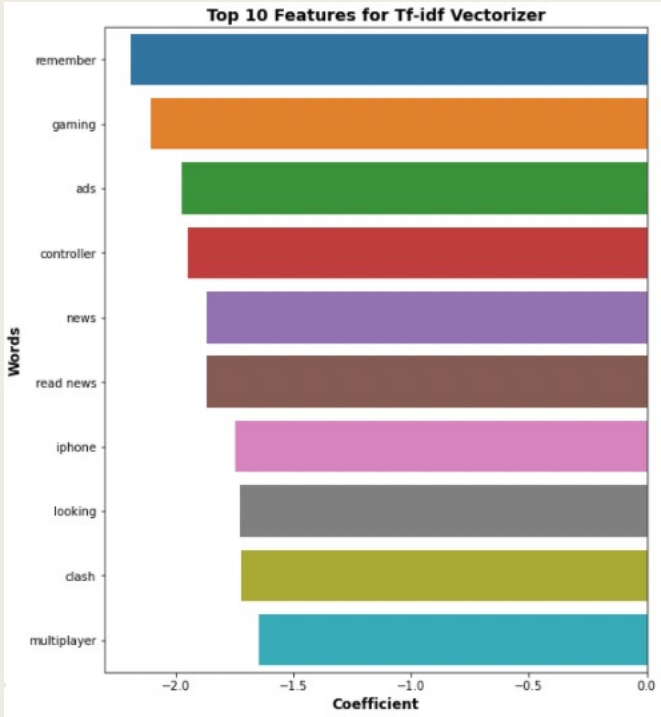
Random Forest



Logistic Regression



Boardgames



Mobile games

Causes of Misclassification

- Content of the post too short
- Lack of key words to classify
- Key word that belong to another category

text	comment
what popular game in what popular game in	Content too short
I like it i have played it alot on my own can t wait to play it with my friends but i don t get the sorcerer he is the only one that does the least amount of damage if you face the cage spider simple lvl monster it is basicly game over since it has defence weakness skilled and health it takes stamina cards and soul arrows cards out of to kill him provided that you drew the right cards so if you are lucky it will still take cards damage card since it takes turns to kill him that is almost half your deck and you lost weapon cards and that is just enemy other classes can keep their weapons while dealing damage so they make atleast some progress against defence enemies can someone tell me how that is balanced dark souls the card game question	'cards' is key word in 'boardgames , hence wrongly classified as 'boardgames

Possible improvement on Misclassification

- Only scrape in data with enough content
- Add in more stopwords

Conclusion

- Logistic Regression is chosen as the model for the classification
- Achieve highest accuracy of 85.7% as well as high sensitivity and specificity of 85.3% and 86.1% respectively
- Features in logistic regression are more informative
- Misclassification can be due to limited content or words that might be associated with the other subreddit