

Mobile Safety Paper

-Summary Slides

Jenny Gong

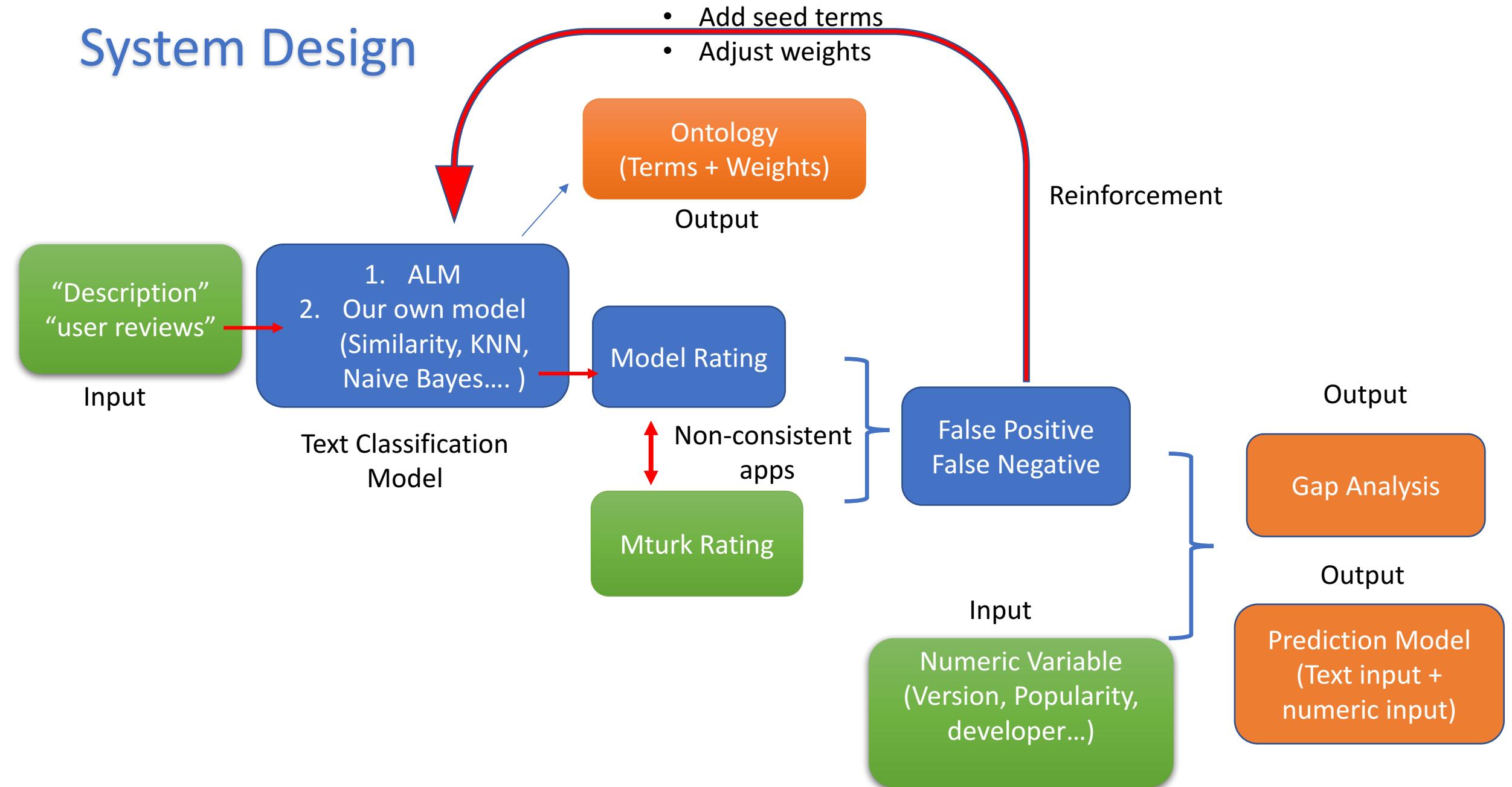
Problem Statement

“Everyone” is one of the four maturity rating levels on Android platform, which represents that the app is safe and can be played for everyone, including kids. However, is the “Everyone” app actually contains no adult content at all? Is the “Everyone” App really safe for your kids?

Actually, our study finds that it is not necessarily the case. Some of the “Everyone” app contains adult content such as violence or sexual themes, which is apparently not suitable for little kids. The Android platform uses self-report system to determine the app maturity rating. Therefore, it gives developers opportunities to underrate their games, so that they can gain more downloads for their game apps.

This study is to detect those “Everyone” game apps which should be categorized as “Everyone 10+”, “Teen”, or even “Maturity 17+”, due to the adult content they contain.

System Design



Algorithm 1 – Strong Vocabulary

Step 1: Identify reason for maturity rating

Content Rating

Teen

Suggestive Themes,
Partial Nudity

→ “Sex” feature

Step 2: Tag each app based on feature,
generate app dataset for each feature

Feature	Violence	Fantasy Violence	Alcohol Drug Tobacco	Sex	Language	Gambling
# Apps	2347	2045	184	221	262	1811

Step 3. Generate key word frequency
distribution for each feature

Frequency, Normalized Frequency

Step 4. Build lexicon

Ratio

Step 5. Human screening, human adding

Step 6. Testing Lexicon -> suspicious apps

Step 3. Generate key word frequency distribution for each feature

1. Tokenize, stop word removal
2. Stemming: PorterStemmer()
3. Frequency distribution for each feature: FreqDist()
4. Aggregate word frequency of each feature for every token, calculate total frequency of each token

Step 3. Generate key word frequency distribution for each feature

5. Normalization

Feature	Violence	Fantasy Violence	Alcohol Drug Tobacco	Sex	Language	Gambling
# Apps	2347	2045	184	221	262	1811

“Sex”, “Language”, “ADT” has much fewer apps than the other three features.

	Total_freq	V_freq	%V	FV_freq	%FV	A_freq	%A	S_freq	%S	G_freq	%G	L_freq	%L
game	25889	8446	33%	7192	28%	515	2%	596	2%	8364	32%	776	3%
play	10981	2509	23%	2414	22%	231	2%	318	3%	5164	47%	345	3%

 Normalization

	V_norm	%V_norm	FV_norm	%FV_norm	A_norm	%A_norm	S_norm	%S_norm	G_norm	%G_norm	L_norm	%L_norm
game	26632	17%	27318	18%	22810	15%	20768	13%	34627	22%	23134	15%
play	7911	11%	9169	13%	10231	15%	11081	16%	21379	31%	10285	15%

Final Distribution: tokens.xlsx

Step 4: Build Lexicon

1) Calculate “Ratio”

For each word:

$$\%V = V_freq / Total_freq$$

$$\%FV = FV_freq / Total_freq$$

$$\%A = A_freq / Total_freq$$

$$\%S = S_freq / Total_freq$$

$$\%G = G_freq / Total_freq$$

$$\%L = L_freq / Total_freq$$

Equally Distributed words -> Stop words

	V_norm	%V_norm	FV_norm	%FV_norm	A_norm	%A_norm	S_norm	%S_norm	G_norm	%G_norm	L_norm	%L_norm
game	26632	17%	27318	18%	22810	15%	20768	13%	34627	22%	23134	15%
play	7911	11%	9169	13%	10231	15%	11081	16%	21379	31%	10285	15%
thi	11973	18%	11745	18%	9965	15%	10105	15%	10781	16%	12193	18%
free	4001	9%	3977	9%	6954	16%	5157	12%	19396	44%	4204	10%
get	4834	14%	4877	14%	6156	17%	7840	22%	7022	20%	4710	13%
featur	5023	18%	4626	17%	3720	14%	4495	16%	5854	21%	3548	13%
new	4459	13%	5120	15%	7352	22%	5471	17%	3581	11%	7095	21%
use	4922	18%	5424	20%	4518	16%	5053	18%	3184	11%	4681	17%
world	4654	18%	5591	22%	4252	17%	4216	17%	3213	13%	3577	14%

Unequally Distributed words -> Lexicon

	V_norm	%V_norm	FV_norm	%FV_norm	A_norm	%A_norm	S_norm	%S_norm	G_norm	%G_norm	L_norm	%L_norm
slot	158	0%	53	0%	3056	8%	3659	9%	32292	82%	30	0%
win	1041	5%	1064	5%	1816	8%	2474	11%	15260	67%	984	4%
enemi	7079	39%	5504	31%	1683	9%	1115	6%	178	1%	2474	14%
casino	79	0%	95	1%	1462	8%	801	4%	15302	86%	89	1%
machin	801	5%	718	4%	1240	7%	1150	7%	12466	75%	328	2%
shoot	6981	60%	2594	22%	133	1%	732	6%	203	2%	984	8%

Step 4: Build Lexicon

2) Lexicon select rules:

1. Big head: 1 feature distribution > 80%
2. Big head: 2 features distribution > 85%
 $V + FV > 85\%$ (they are similar)
3. Small tail: 3 features distribution < 5%

Step 5. Human screening, human adding

1) Human expert screening all the lexicon

- Categorize each feature into “**Strong list**” and “**Weak list**”
- Delete unreasonable words
- Delete “Language” feature
 - Top frequency words doesn’t make sense, hard to extract “strong language” content in app description
 - “Language” feature always come with other feature

2) Human expert add some words into the lexicon

- “Sex”, “Alcohol, Drug, Tobacco” features have too few words, so add some words
- Add “Direct warning” feature: “for adults”, “21 years older”

Final Lexicon: Lexicon.xlsx

Step 6: Testing Lexicon -> Suspicious apps

1. Combine all “Strong list” into “Strong word list”

Strong word list = Violence_Strong + Sex_Strong + ADT_Strong +
Gambling_Strong + DirectWarning_Strong

2. Identify strong words in each app description

3. Look at “Everyone” level

It shouldn't contain any Strong Words!!

Suspicious app 1



Full Moon Wolf Slot machines

Moriam Akter Casino

★★★★★ 39

E Everyone

Contains ads · Offers in-app purchases

⚠ You don't have any devices

Add to Wishlist

Install

8 Strong words!

[u'jackpot', u'casino', u'slot', u'vega', u'payout', u'reel', u'gambl', u'fortun']



Suspicious app 2

Android:



Guns & Cowboys: Bounty Hunter

Play Cool Zombie Sport Games Action

★★★★★ 266

Everyone

Contains ads

⚠ You don't have any devices

Add to Wishlist

Install



iOS:

Guns & Cowboys: Bounty Hunter

By Play Cool Zombie Sport Games sp. z o.o.

Open iTunes to buy and download apps.



[View in iTunes](#)

+ This app is designed for both iPhone and iPad

Free

Category: Games

Released: 13 May 2016

Version: 1.0

Size: 148 MB

Language: English

Seller: PLAY COOL ZOMBIE

SPORT GAMES SP Z O O

© PLAY COOL ZOMBIE SPORT GAMES SP Z O O

Rated 12+ for the following:

Infrequent/Mild Realistic Violence

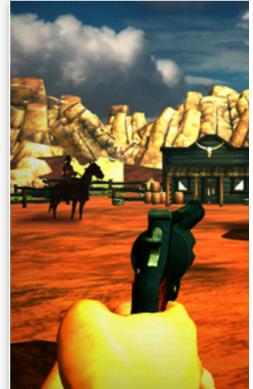
Description

There's a new sheriff in town! Wild West! Mission of the ga

gun fights. Shoot precisely, |

[Play Cool Zombie Sport Ga](#)

Screenshots



9 Strong words!

[u'gun', u'gunsling', u'shoot', u'shooter', u'robberi', u'shot', u'shotgun', u'weapon', u'bullet']

Suspicious apps in “Everyone” level

	A	B	C	H	I
	name	sub_categ	maturity	strong_tokens	#strong_tokens
1	Real Shooting Gun Simulator	Simulation	Everyone	[u'sniper', u'ak47', u'gun', u'rifl', u'assault', u'shoot', u'shooter']	16
2	Traffic Sniper Shooter	Action	Everyone	[u'sniper', u'headshot', u'terrorist', u'marksman', u'rifl', u'assas	16
3	Custom Gun Simulator FREE	Simulation	Everyone	[u'sniper', u'ak47', u'gun', u'rifl', u'assault', u'ammunit', u'shoot	15
4	3D Drone Attack Simulator Game	Simulation	Everyone	[u'antiterrorist', u'terrorist', u'ak47', u'rifl', u'assault', u'kill', u'd	14
5	Real Gun Simulator	Simulation	Everyone	[u'sniper', u'ak47', u'gun', u'rifl', u'assault', u'm16', u'shoot', u's	14
6	Revolver Shooting Range Magnum	Simulation	Everyone	[u'sniper', u'gun', u'rifl', u'assassin', u'shoot', u'shooter', u'pisto	14
7	Water Bottle Flip Shooter Game	Action	Everyone	[u'sniper', u'ak47', u'gun', u'rifl', u'assassin', u'm4', u'assault', u	14
8	Black Ops Shooting Range 3D	Simulation	Everyone	[u'sniper', u'frontlin', u'gun', u'marksman', u'rifl', u'assault', u'k	13
9	Gun Club 2	Action	Everyone	[u'sniper', u'ak47', u'gun', u'marksman', u'rifl', u'm4', u'assault'	13
10	FRONTLINE COMMANDO 2017	Action	Everyone	[u'sniper', u'headshot', u'frontlin', u'terrorist', u'assassin', u'kill	12
11	Gun Builder Simulator 2 FREE	Simulation	Everyone	[u'gun', u'rifl', u'assault', u'ammunit', u'shoot', u'shot', u'grenad	12
12	Real Shooting Army Training	Action	Everyone	[u'terrorist', u'gun', u'assassin', u'kill', u'shoot', u'shooter', u'sh	12
13	Shoot All Bottles	Arcade	Everyone	[u'sniper', u'gun', u'marksman', u'rifl', u'assault', u'shoot', u'sho	12

Suspicious_app_0314.xlsx

28,974 “Everyone” apps

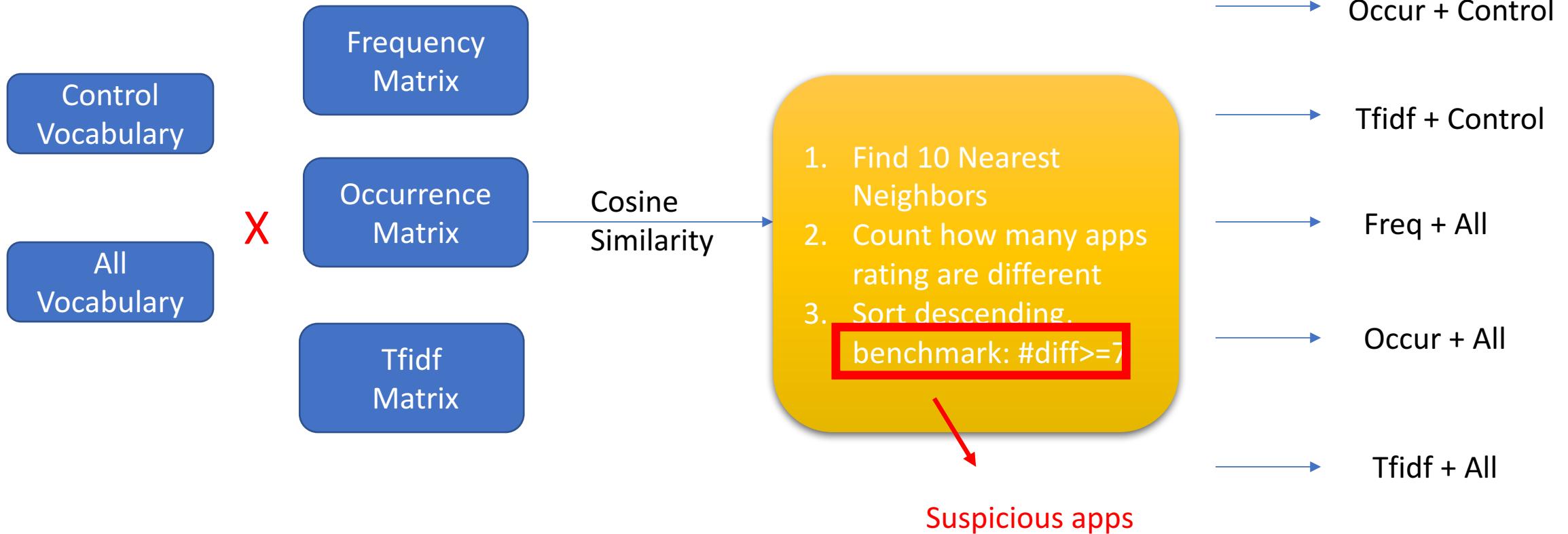
Num of strong words in the app	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Grand Total
Num of apps	5528	1595	632	200	67	39	23	17	10	3	7	8	2	4	1	2	8138

54 apps: has more than 8 strong words in their description

1015 apps: has more than 3 strong words in their description

Algorithm 2 – Similarity

Similarity Steps





Jungle Lion Hunting Sniper 3D

Best apps 2017 Action

Teen

Contains ads

⚠ You don't have any devices

Teen



Lion Hunting 3D

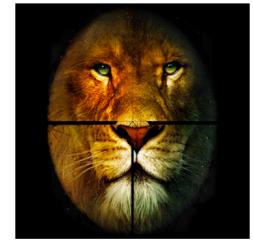
Seena Technocratics Action

Teen

Contains ads · Offers in-app purchases

⚠ You don't have any devices

Teen



Jungle Lion Hunting 2015

CryGameStudio Action

Teen

Contains ads

⚠ You don't have any devices

Teen

Add to Wishlist



True Lion Hunter

Hunt The Hunter Action

Everyone 10+

Contains ads · Offers in-app purchases

⚠ You don't have any devices

Everyone 10+



Lion Sniper Hunter 3D

Game Canvas Action

Teen

Contains ads

⚠ You don't have any devices

Teen

Add to V



Lion Hunting Challenge 3D

BOSPHORUS APPS & GAMES Action

Teen

Contains ads

⚠ You don't have any devices

Teen

Add to Wishlist

Angry Lion Hunting Challenge

SoloTech Action

Everyone

Contains ads

⚠ You don't have any devices

Add to Wishlist

kill them or be killed



Wild Lion Hunting-2017

Game Town Studio Action

Teen

Contains ads · Offers in-app purchases

⚠ You don't have any devices

Teen

Add to W



Real Lion Hunting Challenge

MARTIL Games Action

Teen

Contains ads

⚠ You don't have any devices

Teen

Add to Wishlist



LION HUNTING: MASSACRE

Ladik Apps & Games Adventure

Teen

Contains ads

⚠ You don't have any devices

Teen

Add to Wishlist

Performance

Result_alltoken.xlsx

Number of different apps within 10 nearest neighbors ≥ 7

Only look at 'Everyone' level

	Freq + Control	Occur + Control	Tfidf + Control	Freq + All	Occur + All	Tfidf + All
Suspicious Apps Count	182	116	190	323	270	321
False Positive	39	19	71	70	115	76
FP%	21.4%	16.4%	37.3%	21.6%	42.5%	23.6%



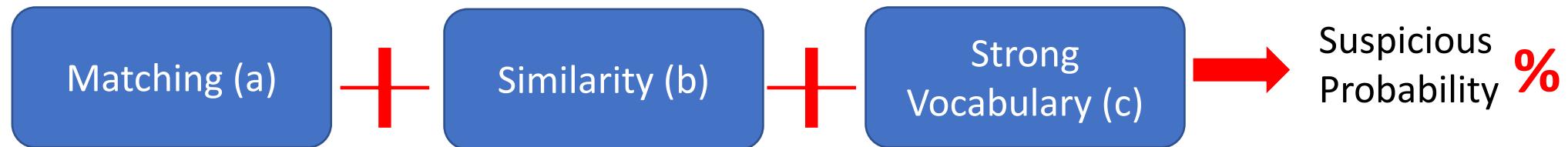
False Positive: algorithm detect as suspicious app, but it actually should belong to 'Everyone' level.

Improvement

- Control vocabulary: strong list * 2
- In neighbors: ‘Maturity 17+’ *4, ‘Teen’ *3, ‘Everyone 10+’ *2

Ensemble

III. Ensemble 3 Algorithms



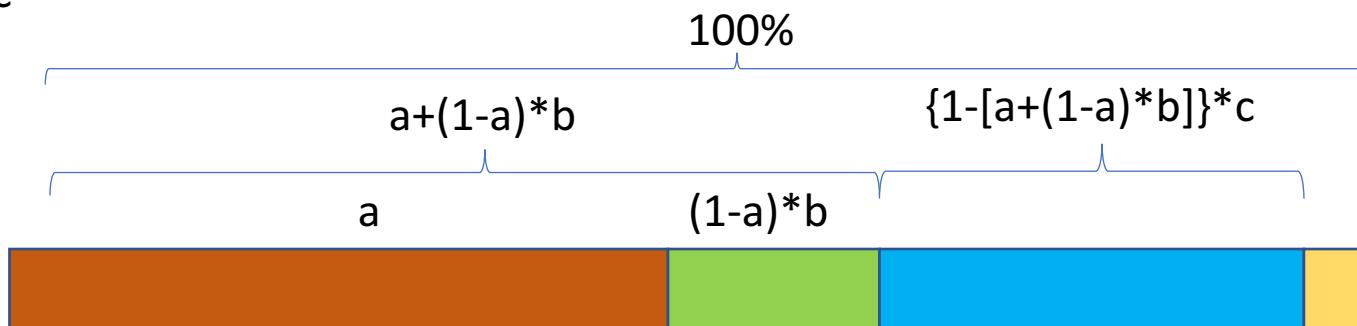
Match with 3,4 : 0.5
Match with 2 : 0.3

9-7 : 0.5
6-5 : 0.3

16-8: 0.8
7-5 : 0.5
4-1 : 0.3

$$a + (1-a)*b + \{1 - [a + (1-a)*b]\}*c = \text{Suspicious Probability (\%)}$$

Impact of a Impact of b Impact of c



For example:
 $a = 0.5$
 $b = 0.2$
 $c = 0.8$

0.5

$(1-0.5)*0.2 = 0.1$

$(1-0.6)*0.8 = 0.32$

92%

Compared with an Average approach

1. $(a+b+c)/3$

	a	b	c	
1)	0.9	0	0	0.3
2)	0.3	0.3	0.3	0.3

2. $(a+b+c)/(number\ of\ non-zeros)$

	a	b	c	
1)	0.9	0.1	0	0.5
2)	0.5	0	0	0.5

Will not show the strong suspicious probability

Suspicious Apps

C	D	E	F	G	H	I	J	K	L
name	sub_category	maturity	maturity	description	a value	b value	c value	probability	
Ultimate Gun Simulator	Simulation	Everyone		Ultimate Gun Simulator perfect gun app with	0.5	0.3	0.8	93.0%	
Weapon simulator	Simulation	Everyone		Weapons Simulator is the best game about g	0.5	0.3	0.8	93.0%	
Black Ops Shooting Range 3D	Simulation	Everyone		Discover the hidden marksman inside you ge	0	0.5	0.8	90.0%	
FRONTLINE COMMANDO 2017	Action	Everyone	Mild Fantasy	Brace yourself for the best first person shoot	0	0.5	0.8	90.0%	
Full Moon Wolf Slot machines	Casino	Everyone		We launched for people who love to live on t	0	0.5	0.8	90.0%	
Gun Club 2	Action	Everyone		GUN SPECIAL ALL THEME PACKS 50% OFF! W	0.5	0	0.8	90.0%	
Modern Little Hunter	Arcade	Everyone		Step into the action as the latest installment	0	0.5	0.8	90.0%	
Real Shooting Army Training	Action	Everyone		Aim and Shoot! Modern Army Commando W	0	0.5	0.8	90.0%	
Shooting Targets	Sports	Everyone		Are you crazy about guns and snipers? Then I	0	0.5	0.8	90.0%	
Sniper Action School	Action	Everyone	Mild Fantasy	Sniper Action School is an interactive shooter	0	0.5	0.8	90.0%	
Traffic Sniper Shooter	Action	Everyone		Aim and Shoot! Realistic graphics and amazin	0	0.5	0.8	90.0%	
Wild Hunter	Adventure	Everyone		Wild Hunter is a complete action pack shooti	0.5	0.5	0.5	87.5%	
Air Fighter: Airplane Shooting	Action	Everyone		Fight the wicked goblin invaders and beat the	0	0.3	0.8	86.0%	
Gun Builder Custom Guns Pro	Simulation	Everyone		Start your journey through the world of weap	0	0.3	0.8	86.0%	
Gun Camera: Augmented Reality	Action	Everyone	Mild Fantasy	Augmented Shooter in Augmented Reality is	0	0.3	0.8	86.0%	
Gun Sim Weapons	Simulation	Everyone		Gun Sim Weapons perfect gun app with big a	0	0.3	0.8	86.0%	
Guns & Cowboys: Bounty Hunter	Action	Everyone		Theres a new sheriff in town! Outlaws and ba	0	0.3	0.8	86.0%	
Guns simulator	Simulation	Everyone		Do you like guns sound? Realistic Guns sound	0	0.3	0.8	86.0%	
Guns: Shooting Range	Action	Everyone		Guns: Shooting Range is an awesome weapon	0	0.3	0.8	86.0%	
Gunslinger Ghostrider Bullseye	Role Playing	Everyone		What happens when an western gunslinger d	0	0.3	0.8	86.0%	
Metal Shooter Super Soldiers	Action	Everyone		Soldiers Gun contra classic is a game take yo	0	0.3	0.8	86.0%	
Modern Army Strike	Action	Everyone	Mild Fantasy	Modern Army strike throws a shadow of airp	0	0.3	0.8	86.0%	
Prison Escape Jail Breakout 3D	Role Playing	Everyone		Prison Escape Jail Breakout 3D simulator, is a	0	0.3	0.8	86.0%	
Real Shooting Gun Simulator	Simulation	Everyone		Welcome to the world of Weapons, for your	0	0.3	0.8	86.0%	
Shark Hunter 2017	Action	Everyone		Get ready hunting the sharks in the Deep oce	0	0.3	0.8	86.0%	
SimGun2 Custom	Simulation	Everyone		Total 91 guns is free now. Welcome to custo	0	0.3	0.8	86.0%	
Sky Runner Balloons Fight	Arcade	Everyone		If you are one of a favorite retro funny game	0	0.3	0.8	86.0%	
Supermarket Prisoner Escape 3D	Role Playing	Everyone		Some dangerous gangsters and prisoners spe	0	0.3	0.8	86.0%	
Survival ONLINE: War at Island	Action	Everyone		Welcome to online battlefield! Grab your gur	0	0.3	0.8	86.0%	
Backgammon	Board	Everyone		Backgammon is one of the oldest board game	0.5	0.5	0.3	82.5%	

Suspicious Probability

0%	1%-49%	50%-59%	60%-69%	70%-79%	80%-89%	90%-100%	total
20105	7804	697	246	51	60	11	28974

