

# Mobile Safety

-Android

03.16.2017

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

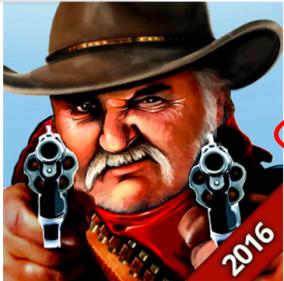
8 Strong words!

[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

Everyone

Contains ads

⚠ You don't have any devices

★★★★★ 266

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

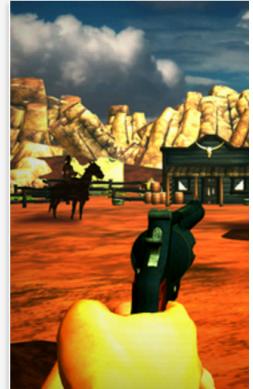
## Description

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

# Next step:

- Strong word list – Absolute method
- Similarity – Relative method