

Détection de compte suspects sur Twitter

*CECILLE Aurélien, DELLA VEDOVA Thomas, GILLOT Francois, JOBARD
Timothée, ZRIG Hazem*

Sommaire

- Introduction
- Data exploration et traitement des données
- Approche non supervisée
- Approche supervisé
- Comparaison des deux méthodes
- Conclusion



Introduction



Data exploration et traitement des données

Traitement sous MongoDB

```
_id: ObjectId("5ee1fb796ebd3103e39bde36")
created_date: "2018-06-14 04:14:25"
current_time: 1528942465150
> quoted_status: Object
  in_reply_to_status_id_str: null
  in_reply_to_status_id: null
  created_at: "Thu Jun 14 02:14:24 +0000 2018"
  in_reply_to_user_id_str: null
  source: "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter fo..."
  quoted_status_id: 1007083141671026689
  retweet_count: 0
  retweeted: false
  geo: null
  filter_level: "low"
  in_reply_to_screen_name: null
  is_quote_status: true
  id_str: "1007083806053068801"
  in_reply_to_user_id: null
  favorite_count: 0
  id: 1007083806053068801
  text: "Aula! Nós somos a história! https://t.co/x0r46v3F0S"
  place: null
> quoted_status_permalink: Object
  lang: "pt"
  quote_count: 0
```

Exploration de critères pertinents

|||

▼

\$group

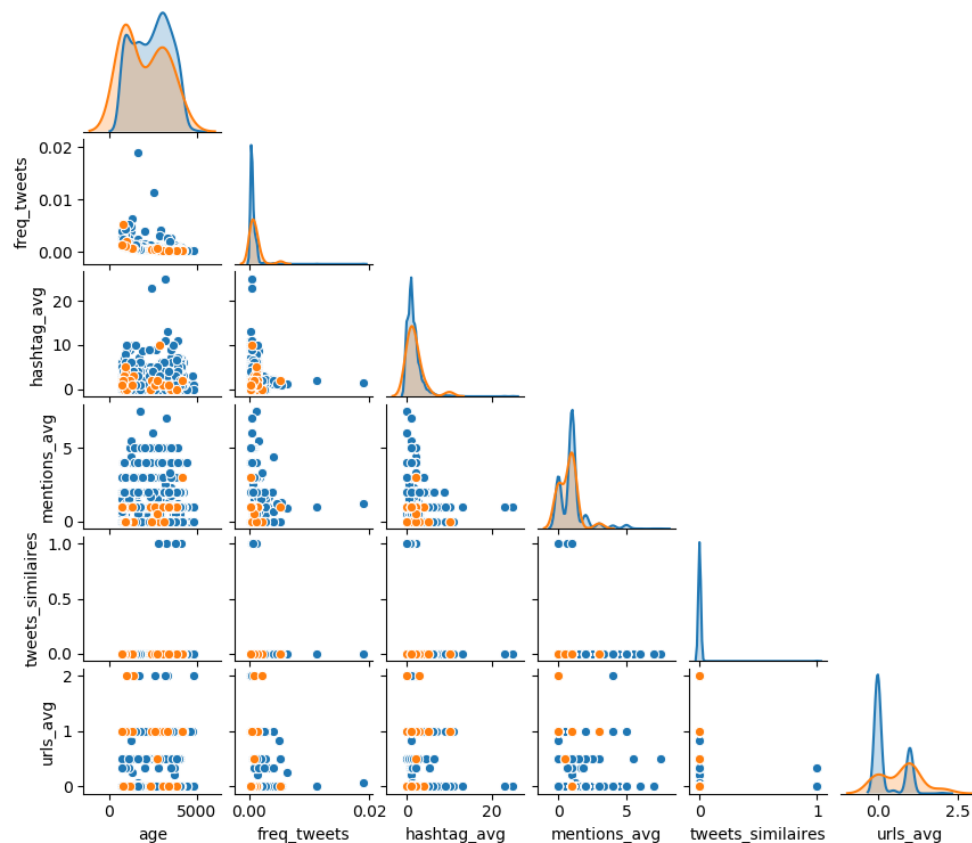
Output after `$group` stage ⓘ (Sample of 20 documents)

```
1 {
2   _id: '$id',
3   nom: {
4     $max: '$name'
5   },
6   age: {
7     $min: '$age'
8   },
9   mentions_avg: {
10    $avg: {
11      $size: '$entities.user_mentions'
12    },
13  },
14  hashtag_avg: {
15    $avg: {
16      $size: '$entities.hashtags'
17    },
18  },
19  tweets_urls: {
20    $sum: '$urls_count'
21  },
22  tweets_count: {
23    $sum: 1
24  },
25
26  tweets_differents: {
27    $addToSet: '$text'
28  },
29  all_tweets: {
30    $push: '$text'
31  },
32  friends_count: {
33    $max: '$friends count'
```

```
_id: 519131046
nom: "JoseLikesGatos"
age: 72827
mentions_avg: 1
hashtag_avg: 1
tweets_urls: 0
tweets_count: 1
▼ tweets_differents: Array
  0: "RT @goldenxgen: i'm so fucking ready for
    tomorrow!!!! let's fucking go..."
  ► all_tweets: Array
    friends_count: 419
```

```
_id: 39598451
nom: "iktripz"
age: 97577
mentions_avg: 1
hashtag_avg: 6
tweets_urls: 0
tweets_count: 2
▼ tweets_differents: Array
  0: "RT @ONTV_NIGERIA: Who is #Mourinho tipping for
    glory in #Russia? See t..."
  1: "RT @ONTV_NIGERIA: #Argentina or #Portugal? -
    Jose #Mourinho predicts #..."
  ► all_tweets: Array
    friends_count: 2182
```

Graphes de relations des premiers critères choisis



Recherche de nouveaux indicateurs plus pertinents

$$Visibility = \frac{\sum_{E \in \{ @, \# \}} Avg(E) \cdot C(E)}{140}$$

avec $C(E)$: cout moyen de caractère nécessaire pour une référence de @ ou # (=11,5)

$$Aggressiveness = \frac{f_{tweets} + f_{friends}}{350}$$

avec f_{tweets} : fréquence des tweets publiés par heure

$f_{friends}$: fréquence du nombre d'amis ajoutés par heure

350 : nombre d'actions maximum possibles imposé par l'API

2 nouveaux indicateurs calculés

Output after `$addFields` stage (Sample of 20 documents)

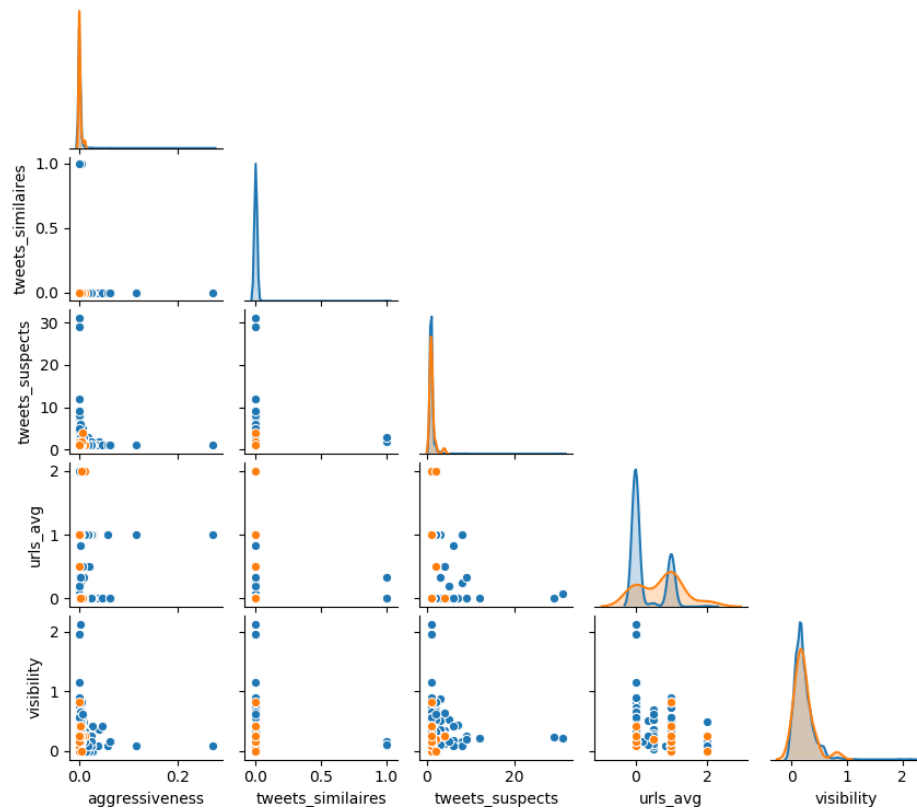
```
1 {
2   aggressiveness: {
3     $divide: [
4       {
5         $sum: [
6           {
7             $divide: [
8               '$tweets_count',
9               '$age'
10            ]
11           },
12          {
13            $divide: [
14              '$friends_count',
15              '$age'
16            ]
17          }
18        ],
19        350
20      ]
21    },
22    visibility: {
23      $divide: [
24        {
25          $sum: [
26            '$hashtag_avg',
27            '$mentions_avg'
28          ]
29        },
30        12.174
31      ]
32    }
33  },
```

```
_id: 954793774483431424
nom: "Saddiq825"
age: 21451
mentions_avg: 0
hashtag_avg: 0
tweets_urls: 0
tweets_count: 1
tweets_suspects: 1
tweets_differents: Array
all_tweets: Array
friends_count: 1881
aggressiveness: 0.0002506709643906045
visibility: 0
tweets_similaires: 0
danger: 1
```

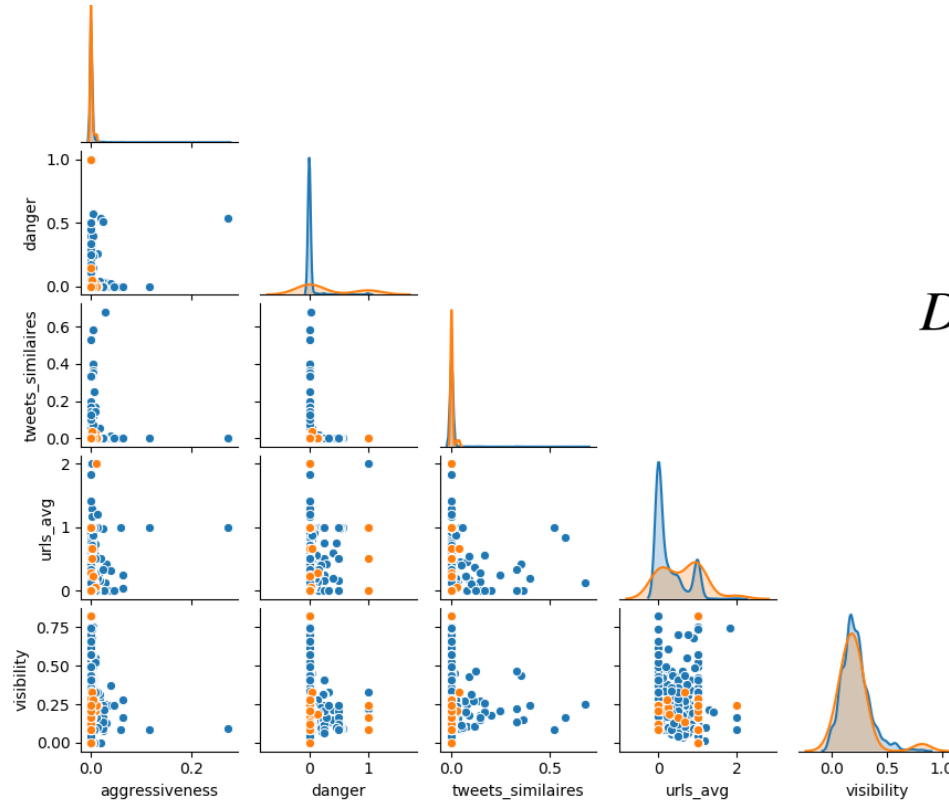
```
_id: 905223162400567296
nom: "ameisaber"
age: 24733
mentions_avg: 0
hashtag_avg: 6
tweets_urls: 0
tweets_count: 1
tweets_suspects: 1
tweets_differents: Array
all_tweets: Array
friends_count: 164
aggressiveness: 0.000019060711253328404
visibility: 0.49285362247412523
tweets_similaires: 0
danger: 1
```

```
_id:
nom:
age:
mentions_avg: 0
hashtag_avg: 6
tweets_urls: 0
tweets_count: 1
tweets_suspects: 1
tweets_differents: Array
all_tweets: Array
friends_count: 164
aggressiveness: 0.000019060711253328404
visibility: 0.49285362247412523
tweets_similaires: 0
danger: 1
```

Graphes de relations avec ces nouveaux critères



Graphes de relations en rajoutant le critère *danger*



$$Danger = \frac{tweets\ suspects}{tweets\ count}$$

Finalisation du traitement des données pertinentes sous MongoDB

The screenshot displays the MongoDB Compass interface. On the left, the query editor shows a pipeline with a single stage: `$project`. The document structure being projected is:

```
1 {
2   all_tweets: 0,
3   tweets_count: 0,
4   friends_count: 0,
5   age: 0,
6   mentions_avg: 0,
7   hashtag_avg: 0,
8   tweets_urls: 0,
9   tweets_suspects: 0
10 }
```

On the right, the output of the `$project` stage is shown as a sample of 20 documents. The first two documents are visible:

```
{
  "_id": 1719604994,
  "nom": "m7md1o16",
  "tweets_differeents": Array,
  "tweets_similaires": 0,
  "danger": 1,
  "aggressiveness": 0.000007780708139909543,
  "visibility": 0.3285690816494168
}
```

```
{
  "_id": 957249187350851584,
  "nom": "6JtnUc2wZKVDyam",
  "tweets_differeents": Array,
  "tweets_similaires": 0,
  "danger": 1,
  "aggressiveness": 0.00008984895153805803,
  "visibility": 0.1642845408247084
}
```

Indicateurs finaux

Pour conclure voici nos 4 indicateurs :

$$Aggressiveness = \frac{f_{tweets} + f_{friends}}{350}$$

avec f_{tweets} : fréquence des tweets publiés par heure

$f_{friends}$: fréquence du nombre d'amis ajoutés par heure

350 : nombre d'actions maximum possibles imposé par l'API

$$Danger = \frac{tweets\ suspects + tweets\ urls}{tweets\ count}$$

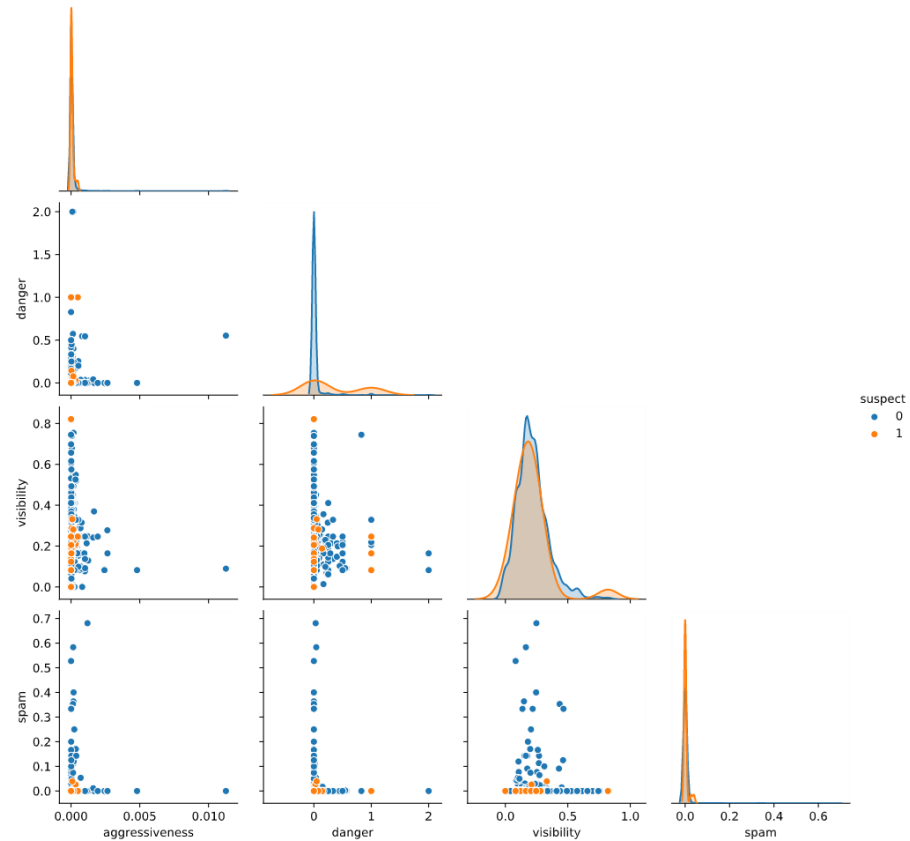
modification de **url_avg** en **tweets_url** = nombre de tweets contenant au moins 2 urls ou plus, sur un même compte

$$Visibility = \frac{\sum_{E \in \{ @, \# \}} Avg(E) \cdot C(E)}{140}$$

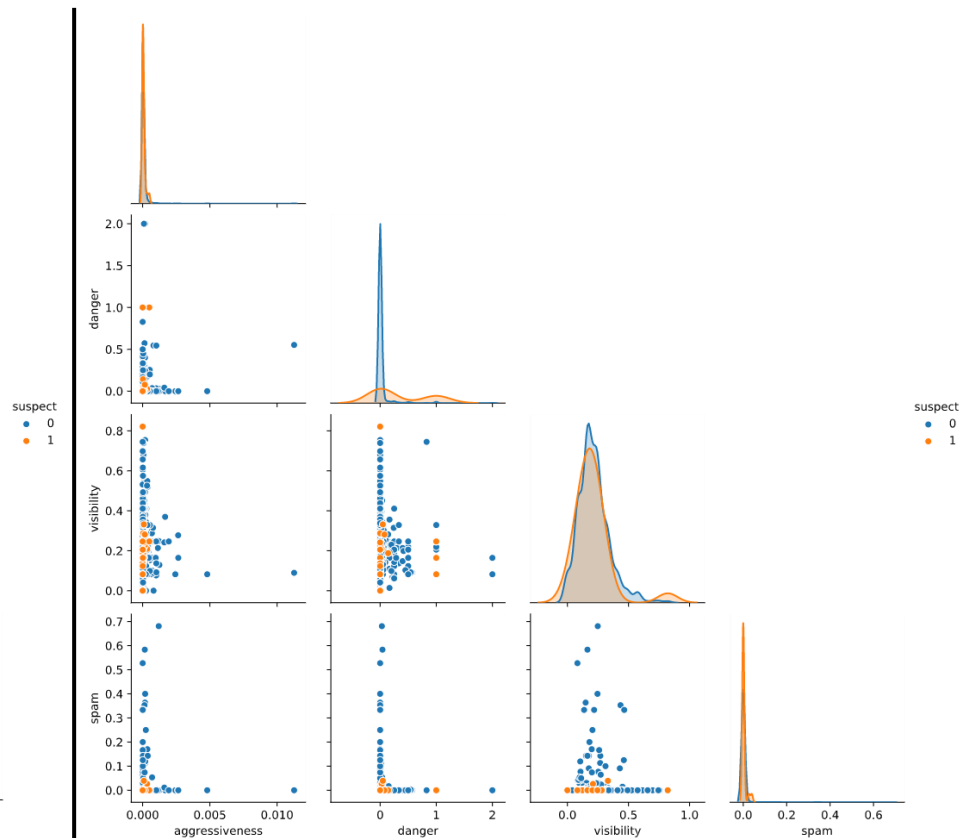
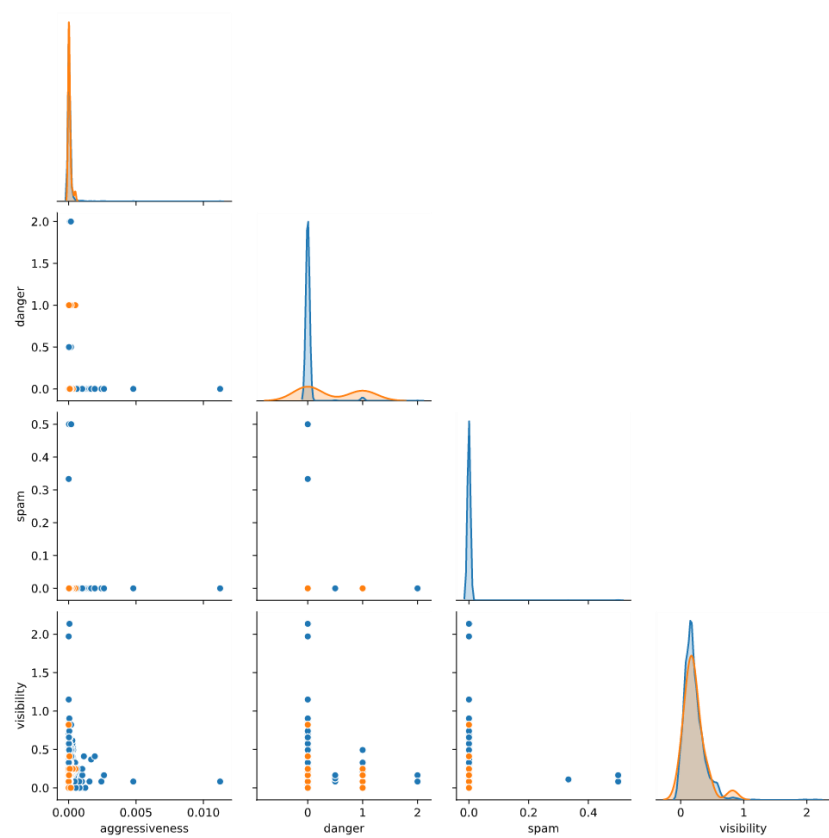
avec $C(E)$: cout moyen de caractère nécessaire pour une référence de @ ou # (= 11,5)

$$Spam = \frac{tweets\ similaires}{tweets\ count}$$


Graphes de relation des indicateurs finaux



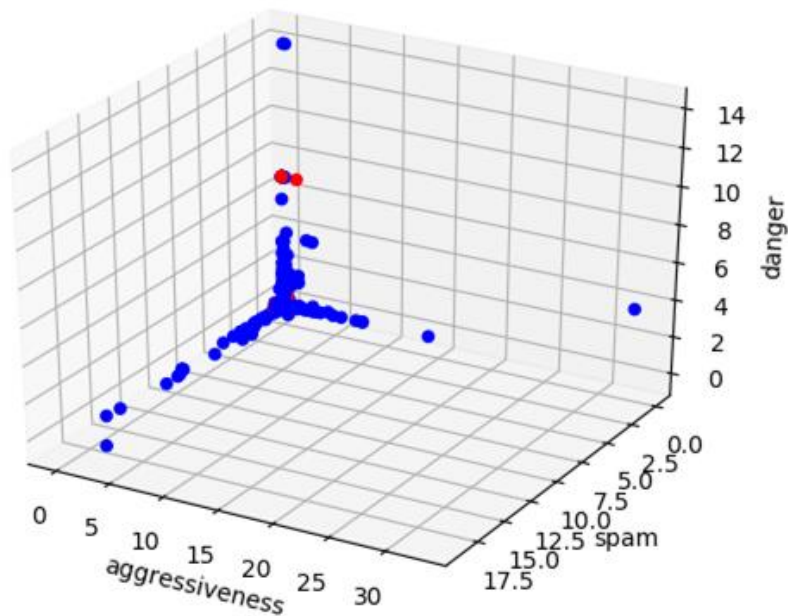
Augmentation des données



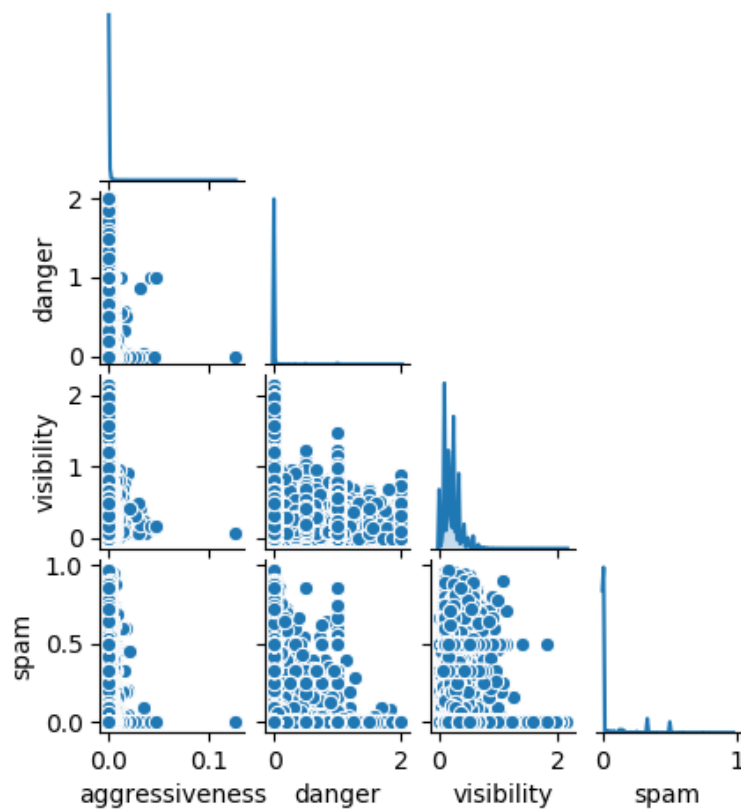
Intérêt de l'ACP ?

Variable	1	2	3	4
Pourcentage expliqué	27%	25%	24%	22% 

Visualisation des suspects trouvés à la main



Visualisation sur les 1.8 millions de compte du dataset total de tweets

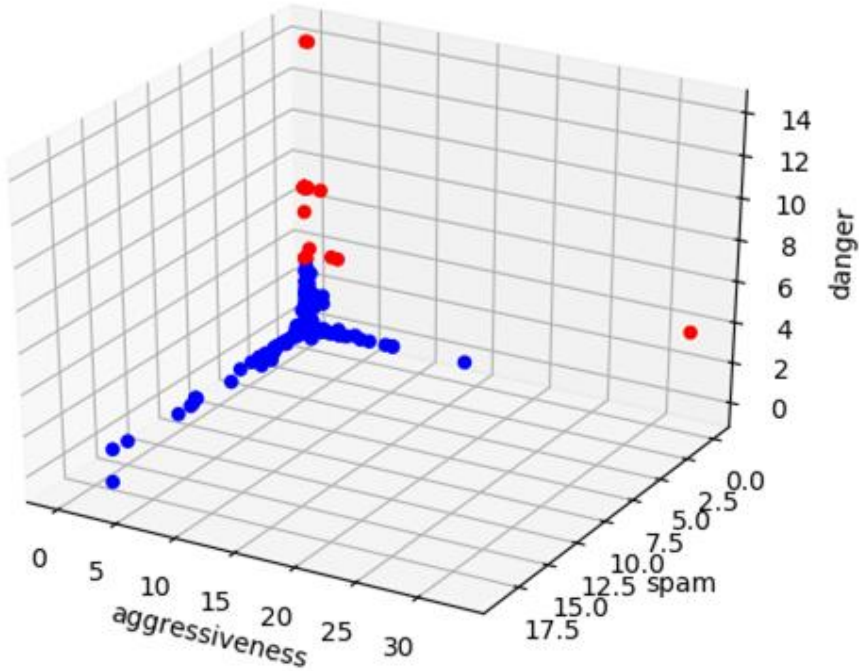


Modèle Non-Supervisé

Matrice de confusion

Valeur à la main Prédiction	Positif	Négatif
Positif	Comptes non-suspects correctement prédit (Vrai positif)	Comptes non-suspects incorrectement prédit (Faux positif)
Négatif	Comptes suspects incorrectement prédit (Faux négatif)	Comptes suspects correctement prédit (Vrai négatif)

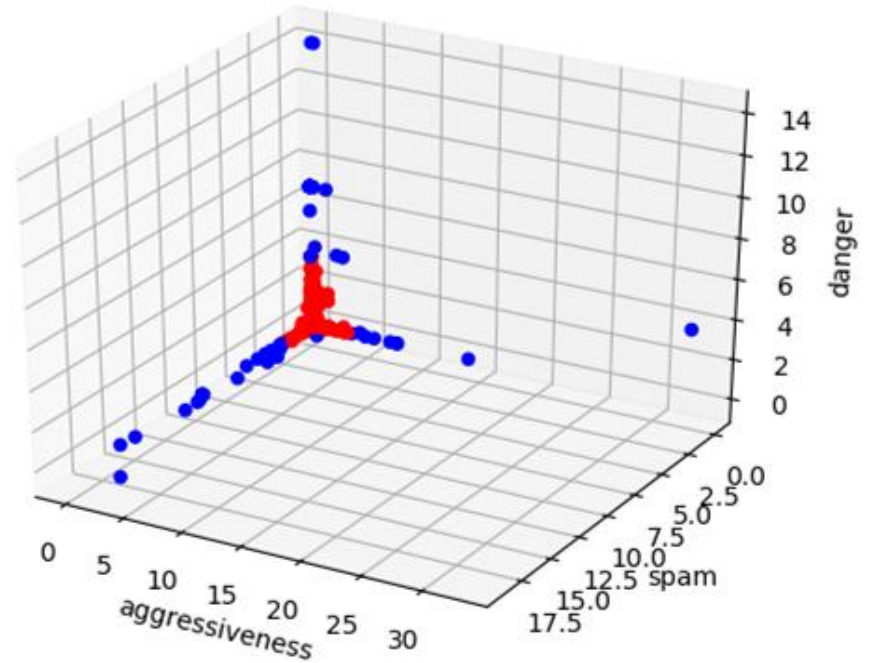
Algorithme de clustering



KMeans

:

1649	27
16	8

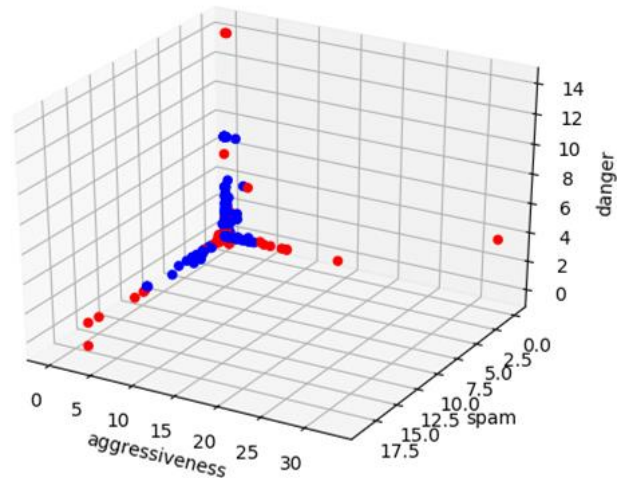


DBSCAN:

1624	52
16	8

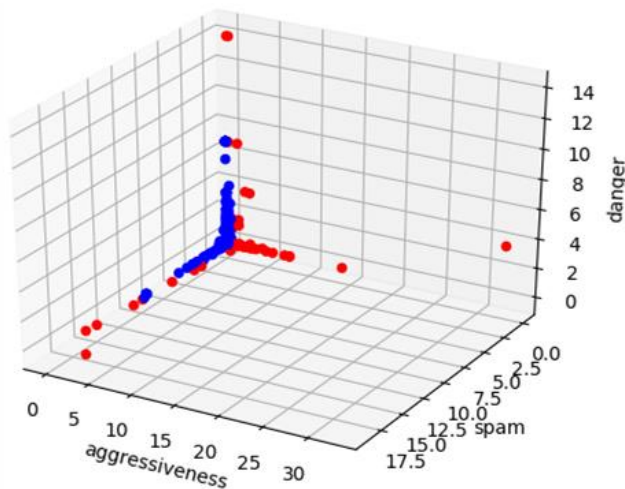
Algorithme de détection d'anomalie

Local Outlier Factor



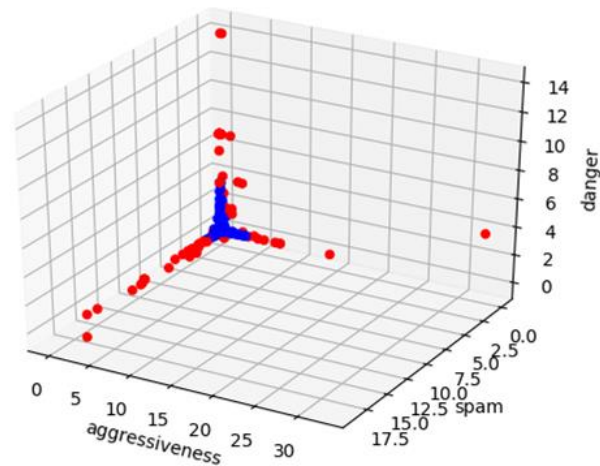
1509	167
21	3

Elliptic Envelope



1594	82
21	3




Isolation Forest





1633	43
16	8

Exemples de Faux-Positifs

RT @casinobitcoin: Bet with confidence and the best odds for the #WorldCup with #bitcoin and #adk only at <https://t.co/s5Z1cFLQTC>

 WORLD CUP SPECIALS  \n\nGet price boosts, free bets and tons more ahead of Saturday's games in Russia ! ... <https://t.co/nEZ2kc8Ddp>

 Science seed\n<https://t.co/RGDgpW9iPs>\n\n#HealthXPh #Jobs #life #love #animals #startup #music #photography #poetry... <https://t.co/icKkbYICvN>

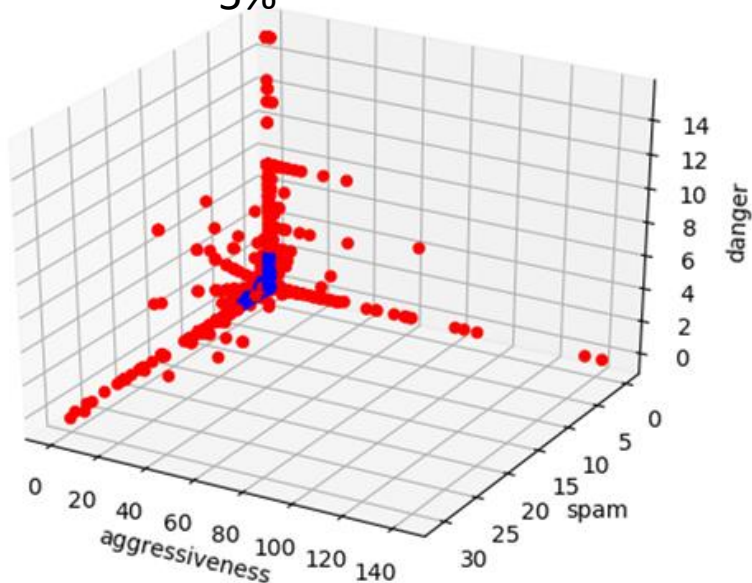
Get 100% up to £50 when you sign up for an account during the #WorldCup  \n\n18+. T&Cs apply.... <https://t.co/JwjYDQGxmr>

We are doing a #WorldCup2018 Challenge Pool. Sign up for free and win some \$\$:
<https://t.co/zGJoJgMAYd>
<https://t.co/2BvJjOJXCn>

Optimisation sur l'ensemble des données

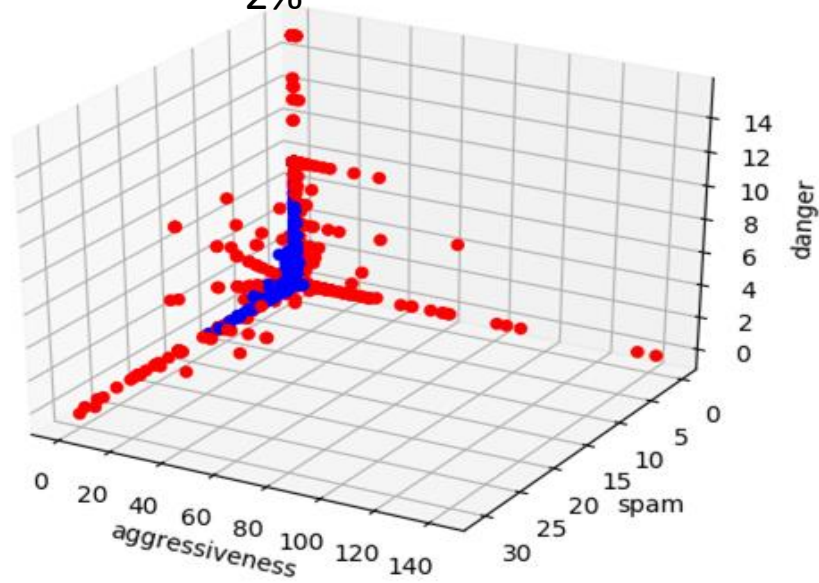
Contamination:

5%



Contamination:

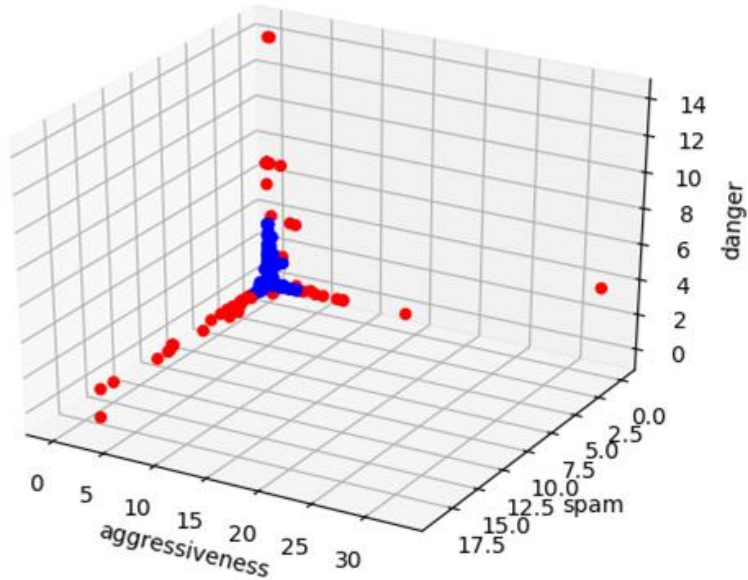
2%





Modèle Supervisé

SVM sur 1700 compte



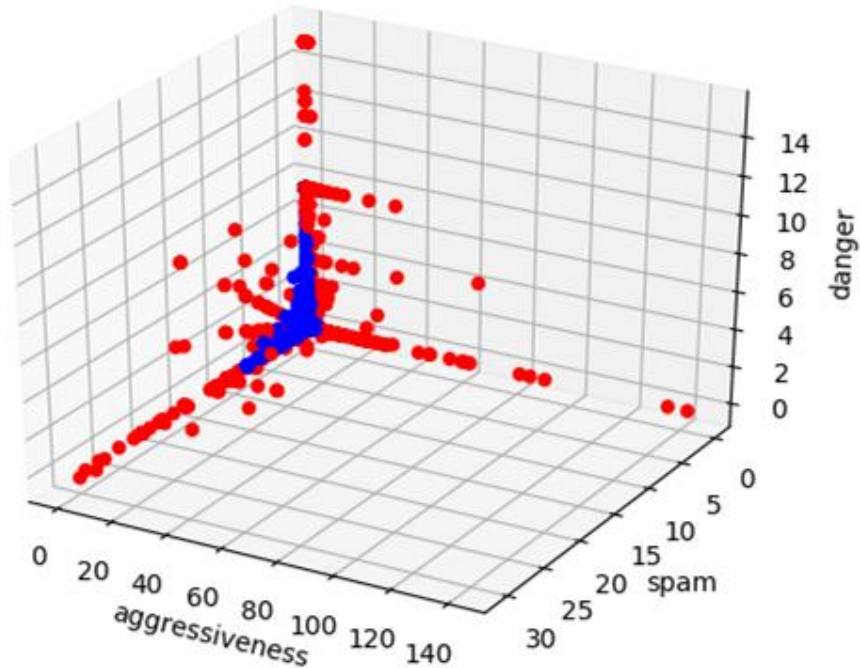
Données de

test:	
482	0
7	21

Données à la

main:	
1614	62
15	9

SVM sur l'ensemble des données

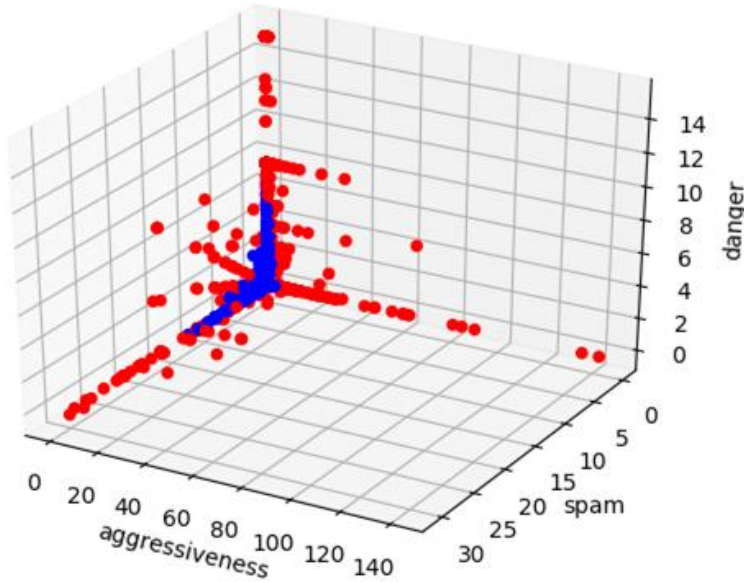


Données de

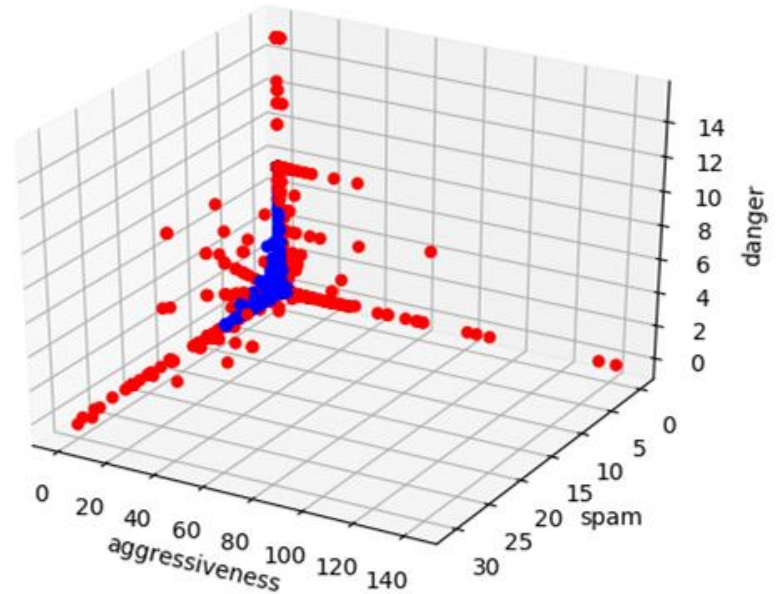
test:

36147	22
70	645

Comparaison des deux méthodes



Non-
supervisé



Supervisé



Conclusion