

# Création d'un classifieur de tickets de bugs


IA pour le GL - 8 Novembre 2021

---

Quentin Perez, Christelle Urtado et Sylvain Vauttier

EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Ales, Ales, France

- Objectif principal du TP : créer un classifieur de tickets de bug
- Approche supervisée avec le dataset de Herzig *et al.* [?] (2013)

 HttpComponents HttpClient / HTTPCLIENT-1598

35 of 53

### Native Windows Negotiate/NTLM via JNA + 407 Proxy Authentication Required

**Details**

Type: Bug  
Priority: Major  
Affects Version/s: 4.4 Beta1  
Component/s: HttpClient (classic)  
Labels: stuck, volunteers-wanted  
Environment: Windows 8

**Description**

I'm trying to use the native Windows NTLM negotiation as described at <http://svn.apache.org/repos/asf/httpcomponents/httpclient/trunk/httpclient-win/src/examples/org/apache/http/examples/client/win/ClientWinAuth.java> but I need to explicitly set a proxy.

```
if (!WinHttpClient.isWinAuthAvailable()) {
    System.out.println("Integrated Win auth is not supported!!!");
}

HttpClientBuilder httpClientBuilder = WinHttpClient.custom();

Host httpProxy = new Host("proxysrvr.example.com", 3128);
httpClientBuilder.setProxy(httpProxy);
CloseableHttpClient httpClient = httpClientBuilder.build();

try {
    HttpGet httpget = new HttpGet("http://www.google.it");
    System.out.println("Executing request " + httpget.getRequestLine());
    CloseableHttpResponse response = httpClient.execute(httpget);
    try {
        System.out.println("-----");
        System.out.println(response.getStatusLine());
        EntityUtils.consume(response.getEntity());
    } finally {
        response.close();
    }
}
```

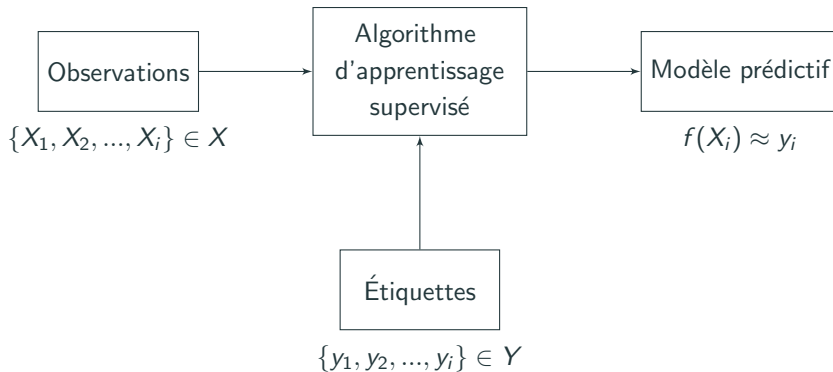
**People**

Assignee: Unassigned  
Reporter: Giacomo Boccardo  
Votes: 1 Vote for this issue  
Watchers: 3 Start watching this issue

**Dates**

Created: 14/jan/15 16:29  
Updated: 02/May/17 12:44

# Apprentissage supervisé



# Objectifs du TP

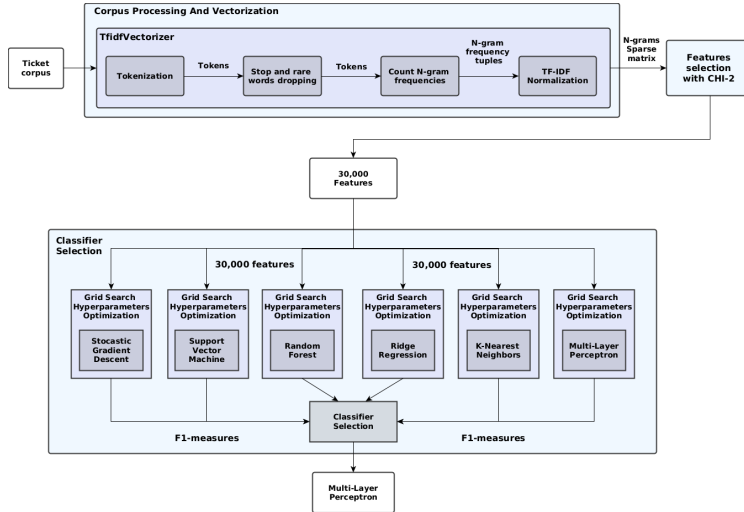
1. Sélectionner les features représentatives avec un Chi-deux
2. Optimiser et sélectionner un classifieur parmi un panel de 5 classifieurs
3. Échantillonner le nombre de features afin de trouver un nombre optimal
4. Créer la matrice de confusion du classifieur
5. Expliquer la classification de faux positifs et faux négatifs

## Contexte et jeu de données

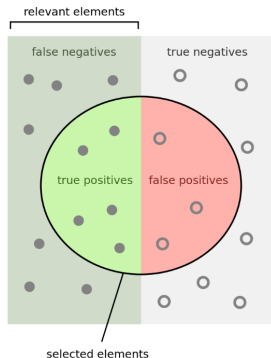
Dataset créé et utilisé par Herzig *et al.* [?] (2013) contenant 5591 tickets issus de 3 projet open-source écrits en Java. Ce jeu de données a été manuellement annoté et nettoyé.

	Maintainer	Tracker type	#Reports	#Labelled "BUG"	#Labelled "NON-BUG"
<b>HTTPClient</b>	Apache	JIRA	746	305	441
<b>Jackrabbit</b>	Apache	JIRA	2402	697	1705
<b>Lucene-Java</b>	Apache	JIRA	2443	938	1505
<b>Total :</b>			<b>5591</b>	<b>1940</b>	<b>3651</b>

# Corpus processing and classifier selection



# Mesures utilisées et matrice de confusion



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$F1 = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{(\text{Précision} + \text{Rappel})}$$

(Walber, Wikipedia, Licence CC BY-SA 4.0)

# Matrice de confusion

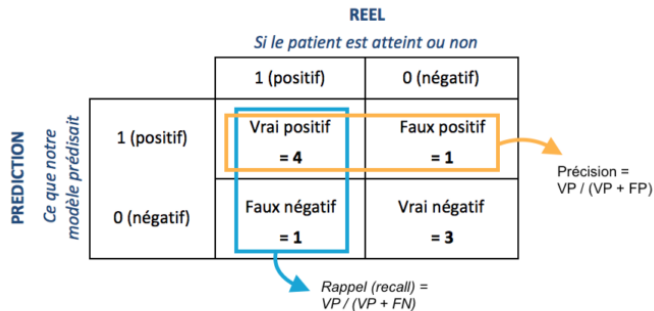
**PREDICTION**  
*Ce que notre modèle prédisait*

**REEL**  
*Si le patient est atteint ou non*

	1 (positif)	0 (négatif)
1 (positif)	Vrai positif = 4	Faux positif = 1
0 (négatif)	Faux négatif = 1	Vrai négatif = 3

Précision =  $VP / (VP + FP)$

Rappel (recall) =  $VP / (VP + FN)$

The diagram shows a 2x2 confusion matrix. The columns are labeled 'REEL' (Actual) with values '1 (positif)' and '0 (négatif)'. The rows are labeled 'PREDICTION' (Predicted) with values '1 (positif)' and '0 (négatif)'. The cells contain: Top-left: 'Vrai positif = 4' (highlighted with a blue box); Top-right: 'Faux positif = 1' (highlighted with an orange box); Bottom-left: 'Faux négatif = 1' (highlighted with a blue box); Bottom-right: 'Vrai négatif = 3'. An orange arrow points from the 'Faux positif' cell to the 'Précision' formula. A blue arrow points from the 'Faux négatif' cell to the 'Rappel (recall)' formula.



## Definition

L'explicabilité se définit comme le fait de pouvoir comprendre les mécanismes internes du classifieur qui fondent une ou plusieurs prédictions. Cette explicabilité peut se faire de manière globale (mécanismes internes du classifieur qui conduisent à la classification) ou de manière locale (mécanisme qui conduisent à la classification d'une instance).

2 types d'interprétabilité :

- Interprétabilité locale : expliquer la classification d'une instance par le classifieur
- Interprétabilité globale : expliquer le fonctionnement interne du classifieur

## *Local Interpretable Model-agnostic Explanations (LIME)*

y=BUG (probability 0.751, score 1.103) top features

Contribution?	Feature
+1.243	Highlighted in text (sum)
-0.141	<BIAS>

anonymous user fail to add to cart "attempting to save a customer with an id that already exists" anonymous user fail to add to cart "attempting to save a customer with an id that already exists" anonymous user fail to add to cart "attempting to save a customer with an id that already exists" \*\*sometimes when multiple anonymous customer trying to add to cart, i get the following exception:\*\* attempting to save a customer with an id that already exists in the database. this can occur when legacy customers have been migrated to broadleaf customers, but the batchstart setting has not been declared for id generation. in such a case, the defaultbatchstart property of idgenerationdaoimpl (spring id of blidgenerationdao) should be set to the appropriate start value \*\*but i actually don't have any migrated data and already started from clean database\*\* many thanks for support