



Projet Data Mining

Coraline Parguey



Contexte

- Lilly = Industrie pharmaceutique
 - > 90% de la production est automatisée
- Equipes de support IT
- Incidents = arrêts de production => perte économique

Problématique et Objectif du projet

- **Problématique** : comment réduire les temps d'arrêts de production en utilisant les logs des applications MES ?
- **Objectif** : développer un modèle mathématique qui prédit l'apparition des incidents (=maintenance prédictive).

Données



- **Logs :**

```
0 12/21/20 13:03:10 [31952:main]{0} H:/core/PDCore\pdos.cpp-72 Try to set process affinity mask = 1
1 12/21/20 13:03:10 [31952:main]{0} H:/core/PDCore\pdos.cpp-76 Systems affinity is now: process=1; system=15
```

- **Fichier d'incidents**

Heure de création, titre, description de l'incident, personne en charge,...

Combinaison => classier les logs

Objectifs étudiant CSMI

Précédemment...

2 modèles de classification testés : ACP et SVM

Précision : 20% / Rappel : 66%

Objectif étudiant CSMI :

Tester et comparer d'autres modèles de classification (Arbre de décision, Réseaux de neurones, Sélection de variables, ...)

Outils :

Python (Pandas, Scikit learn, ...) GitHub

Informations supplémentaires

- Données confidentielles
- Stage possible