



# MÉMOIRE

pour obtenir le grade de Master délivré par

Université Paris 8 Vincennes à Saint-Denis

Mention *Informatique*

*Parcours MIASHS Big data et fouille de données*

*présenté et soutenu publiquement par*

**Bahia BOUAADA**

le 19 septembre 2021

**Prédiction de données : application sur des  
données de matériels informatiques**

Encadrant universitaire : Prenom Nom

Tuteur de stage : Prenom Nom

Stage effectué à : ECONOCOM  
Adresse  
Suite Adresse

Université Paris 8  
Laboratoire d'Informatique Avancée de Saint Denis  
EA n° 4383 Saint Denis, France



# Sommaire

<b>I</b>	<b>Présentation de l'entreprise</b>	<b>5</b>
1	Présentation de l'entreprise	7
<b>II</b>	<b>Problématique</b>	<b>11</b>
2	Le contexte de résolution du problème	13
3	Le problème à résoudre	15
<b>III</b>	<b>État de l'art</b>	<b>17</b>
4	État de l'art des techniques	19
<b>IV</b>	<b>Système réalisé</b>	<b>25</b>
5	Implémentation du système	27



## partie I

# Présentation de l'entreprise



# Sommaire

---

<b>1</b>	<b>Présentation de l'entreprise</b>	<b>7</b>
1.1	Présentation de l'Entreprise ECONOCOM . . . . .	8
1.2	Les domaines d'expertise chez ECONOCOM . . . . .	9

---





# Chapitre 1

## Présentation de l'entreprise

### Sommaire

---

1.1	Présentation de l'Entreprise ECONOCOM . . . . .	8
1.2	Les domaines d'expertise chez ECONOCOM . . . . .	9

---

## 1.1 Présentation de l'Entreprise ECONOCOM

Econocom est une entreprise française spécialisée dans la conception, le financement et l'accompagnement de la transformation digitale des systèmes d'information de ses clients. Soutenue par plus de 9 000 collaborateurs répartis dans 19 pays, et un chiffre d'affaires qui dépasse les 3 milliards d'euros pour l'année 2018, Econocom dispose de plusieurs moyens technologiques et financiers ainsi qu'un niveau élevé d'expertise pour la réussite des grands projets digitaux, services aux infrastructures, conseil applicatifs et solutions métiers. Depuis 2015, Econocom a acquis le statut de Société Européenne. Cotée sur Euronext à Bruxelles depuis 1986, l'action Econocom Group fait partie des indices Bel Mid et Tech40.

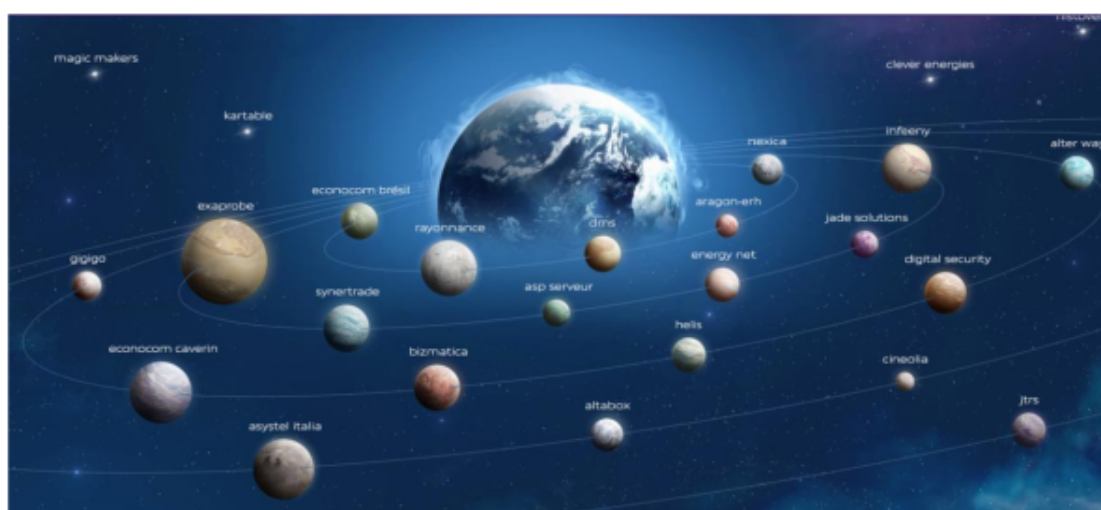


FIG. 1.1: Organigramme du groupe

Comme le montre le schéma de la Figure 1, Econocom présente une organisation innovante inspirée de la voie lactée, la Galaxie, comme on l'appelle au sein de la société. La planète Econocom comporte toutes les activités historiques, en constante expansion et qui restent entièrement possédées par le groupe. Des petites et moyennes entreprises représentées par des satellites, dans plusieurs domaines principalement rattachés au monde du digital. Les entrepreneurs dirigeants et généralement fondateurs conservent une part significative dans le capital social de l'entreprise, ce qui leur permet une meilleure maîtrise et une totale autonomie. Econocom compte aujourd'hui 19 sociétés, positionnées dans les domaines stratégiques de la sécurité, du développement d'applications web et mobile, des solutions digitales en mode cloud, et du conseil en transformation digitale.

Les tailles de ces sociétés sont en proportion avec leur contribution au chiffre d'affaires du groupe. On peut citer les plus performantes en terme de chiffre d'affaire :

- Exaprobe : C'est l'intégrateur du groupe Econocom. Avec plus de 15 années d'expérience, son métier est de définir, intégrer et exploiter, chez le client ou dans le Cloud, des infrastructures et solutions Réseaux, Sécurité, Communications Unifiées et Audiovisuelles.
- Infeeny : Elle offre à ses clients un accompagnement et des solutions Microsoft adaptées pour répondre aux enjeux de la transformation digitale. Infeeny regroupe plus de 330 collaborateurs et se positionne dans le top 3 des spécialistes Microsoft en France.

## 1.2 Les domaines d'expertise chez ECONOCOM

Econocom met en œuvre pour et avec ses clients, du conseil, des projets d'approvisionnement et de gestion administrative des actifs numériques, des services aux infrastructures, des applicatifs et des solutions métiers et le financement de ces projets. Nous détaillons ci-dessous, les différents domaines d'expertises, mais aussi les entités (BU) qui les prennent en charge :

- La Cyber Sécurité est assurée par Digital Security et Exaprobe. Ils font du conseil, de l'audit et de l'expertise en sécurité numérique et notamment celle des IOT (Internet Of Things) ou des objets connectés.
- Microsoft est assurée par Infeeny.
- Web Apps, SaaS et Cloud : ces services sont assurés par Alter Way, Aragon-ERH, ASP Serveur, Econocom Brésil, Nexica, Synertrade Infrastructure et réseaux : Asystel Italia, ASP Serveur, Exaprobe et Nexica offrent une panoplie de services et solutions comme les services d'hébergement critique On-Demand, réseau et infogérance.
- Mobilité : Bizmatica, DMS, Econocom Brésil, GIGIGO, Jade Solutions, JTRS et Rayonnance mettent en œuvre des solutions qui vont accélérer le processus de transformation numérique des entreprises en se focalisant et en mettant en avant l'expérience utilisateur, qui devient essentielles pour l'adoption d'un nouveau produit / services.
- Digital Signage et Multimedia : assuré par Altabox, Caverin, Cineolia, Energy Net, cette entité permet la conception et le déploiement de solutions d'affichage dynamique, du marketing sensoriel et d'analyse de trafic de données avec pour objectif l'amélioration de l'expérience client et l'accélération des ventes.
- Conseil : Helis, spécialiste en conseil stratégique, AMOA et Direction de projet, offrant ainsi à Econocom une brique supplémentaires qui lui permet de mieux

PRÉDICTION DE DONNÉES : APPLICATION SUR DES DONNÉES DE  
MATÉRIELS INFORMATIQUES

répondre aux besoins de ses clients.

<b>CYBER SÉCURITÉ</b>		DIGITAL SECURITY • EXAPROBE
<b>MICROSOFT</b>		INFEEENY
<b>WEB APPS, SAAS &amp; CLOUD</b>		ALTER WAY • ARAGON-ERH • ASP SERVEUR • ECONOCOM BRÉSIL • NEXICA • SYNERTRADE
<b>INFRASTRUCTURE &amp; RÉSEAUX</b>		ASYSTEL ITALIA • ASP SERVEUR • EXAPROBE • NEXICA
<b>MOBILITÉ</b>		BIZMATICA • DMS • ECONOCOM BRÉSIL • GIGIGO • JADE SOLUTIONS • JTRS • RAYONNANCE
<b>DIGITAL SIGNAGE &amp; MULTIMEDIA</b>		ALTABOX • CAVERIN • CINEOLIA • ENERGY NET
<b>CONSEIL</b>		HELIS

FIG. 1.2: Les domaines d'expertise chez ECONOCOM

**partie II**

**Problématique**



# Sommaire

---

<b>2</b>	<b>Le contexte de résolution du problème</b>	<b>13</b>
2.1	Introduction . . . . .	14
2.2	Conclusion . . . . .	14
<b>3</b>	<b>Le problème à résoudre</b>	<b>15</b>
3.1	Introduction . . . . .	16
3.2	Conclusion . . . . .	16

---





# Chapitre 2

## Le contexte de résolution du problème

### Sommaire

---

2.1	Introduction . . . . .	14
2.2	Conclusion . . . . .	14

---

## 2.1 Introduction

La société ECONOCOM propose des solutions de transformation digitale comme la location des matériels informatiques pour faciliter les besoins des clients. Les données sont collectées et renseignées sur une plateforme cloud qui s'appelle ServiceNow. Nous avons décidé de prendre pour notre projet, des données de matériels informatiques (des ordinateurs). Comme dit plus haut ces données concerne la location d'ordinateurs par les clients d'ECONOCOM. Une fois les données collectées sur le ServiceNow, elles sont consolidées, nettoyées, préparées et enrichies. La sortie finale est un fichier au format csv de 28 colonnes et 130 000 lignes.

Ci-dessous, une description des contenus de ces colonnes :

- \* ComputerSystemId : Identifiant du système d'exploitation de l'ordinateur
- \* Model : Modèle de l'ordinateur
- \* DateProd : Date de fabrication de l'ordinateur
- \* IdleSeconde : Température quand le processeur n'est pas exploité par seconde
- \* ActiveSeconde : Le temps en seconde pendant lequel le processeur est exploité
- \* CPU : Microprocesseur initial d'un ordinateur
- \* StandBySeconde : Temps d'attente du redémarrage du processus par seconde
- \* HibernateSeconde : L'état de veille du système d'exploitation de l'ordinateur par seconde
- \* PhysicalMemory : Taille Mémoire physique de l'ordinateur
- \* MaxPhysicalMemory : Taille maximale de la mémoire physique de l'ordinateur
- \* NbNormReboot : Nombre de redémarrage du système
- \* NbAbnReboot : Nombre de redémarrage du système
- \* TotalDataSentMo : Taille totale des données envoyées par le système
- \* TotalDataReceivedMo : Taille totale des données reçues par le système
- \* TotalDiskReadMo : Taille totale des données lues par le disque dur interne de l'ordinateur
- \* TotalDiskWriteMo : Taille totale des données écrites sur le disque dur interne de l'ordinateur
- \* AvgDiskActivity : Temps moyen d'activité du disque interne de l'ordinateur
- \* MaxDiskActivity : Temps maximum d'activité du disque dur interne de l'ordinateur
- \* SLAComplianceTimeSeconde : Mesure de compilation du système par seconde

## 2.2 Conclusion

En conclusion, l'objectif de notre sujet de mémoire est de faire de la prédiction sur des matériels informatiques afin de détecter ceux qui sont susceptibles de tomber en panne.

# Chapitre 3

## Le problème à résoudre

### Sommaire

---

<b>3.1 Introduction . . . . .</b>	<b>16</b>
3.1.1 Prédiction de pannes sur des matériels informatiques . .	16
<b>3.2 Conclusion . . . . .</b>	<b>16</b>

---

## 3.1 Introduction

La société ECONOCOM propose à ses clients la possibilité d'externaliser toute la gestion des infrastructures informatiques. Pour ce faire, il nous a été confié notamment comme missions de mettre en place des solutions pour l'analyse, le traitement et la restitution des données collectées de plusieurs sources de la part des différents équipements informatiques installés sur les sites des clients.

### 3.1.1 Prédiction de pannes sur des matériels informatiques

Une fois les données collectées, il est nécessaire pour l'entreprise d'informer les clients de façon automatique, les équipements informatiques qui sont susceptibles de tomber en panne. Ce qui nous amène à notre problématique qui est la prédiction d'équipements ou matériels informatiques qui vont tomber en panne.

## 3.2 Conclusion

Ainsi la problématique de notre sujet de mémoire s'articule autour de la prédiction de données. La suite de notre travail consistera à étudier les méthodes et algorithmes existants nous permettant de résoudre cette problématique

partie III

État de l'art



# Sommaire

---

<b>4</b>	<b>État de l'art des techniques</b>	<b>19</b>
4.1	Introduction . . . . .	20
4.2	Comparaison entre Random Forest et Decision Tree . . . . .	20
4.3	Prédiction de pannes DSL par mesure passive sur des passerelles domestiques . . . . .	23
4.4	Conclusion . . . . .	24

---





# Chapitre 4

## État de l’art des techniques

### Sommaire

---

4.1	Introduction . . . . .	20
4.2	Comparaison entre Random Forest et Decision Tree . .	20
4.3	Prédiction de pannes DSL par mesure passive sur des passerelles domestiques . . . . .	23
4.4	Conclusion . . . . .	24

---

## 4.1 Introduction

Notre état de l'art va donc tourner autour des algorithmes d'apprentissage supervisé qui peuvent nous permettre de faire de la prédiction de panne encore appelée maintenance prédictive. Nous avons trouvé deux articles qui traitent de ce genre de problématique.

## 4.2 Comparaison entre Random Forest et Decision Tree

Nous allons utiliser cet article [2] dans lequel ont été comparés entre eux les efficacités des algorithmes de Random Forest et Arbre de Décision. Pour faire ce test, dans l'article on a utilisé les données du cancer du sein se trouvant dans la base de données UCI Machine Learning repository [3].

Serial Number	Dataset name	number of instances	number of attributes
1	Lymph	148	19
2	Autos	205	26
3	Sonar	208	61
4	Heart-h	270	14
5	Breast cancer	286	10
6	Heart-c	303	14
7	Ionosphere	351	35
8	colic	368	23
9	Colic.org	368	28
10	Primary tumor	399	18
11	Balance Scale	625	25
12	Soybean	683	36
13	Credit a	690	16
14	Breast W	699	10
15	Vehicle	846	19
16	vowel	990	14
17	Credit g	1000	21
18	Segment	2310	20
19	Waveform	5000	41
20	Letter	20,000	17

FIG. 4.1: Caractéristiques et détails sur le jeu de données

Serial NO	Data Set	No. of instances	No. of attributes	Random Forest		J48 Results	
				Correctly classified instances	Incorrectly classified instances	Correctly classified instances	Incorrectly classified instances
1	Lymph	148	19	81.08%	18.91%	77.02%	22.97%
2	Autos	205	26	83.41%	16.58%	80.95%	18.04%
3	Sonar	208	61	80.77%	19.23%	71.15%	28.84%
4	Heart-h	270	14	77.89%	22.10%	80.95%	19.04%
5	Breast cancer	286	10	69.23%	30.76%	75.52%	24.47%
6	Heart-c	303	14	81.51%	18.48%	77.56%	22.44%
7	Ionosphere	351	35	92.88%	7.12%	91.45%	8.54%
8	colic	368	23	86.14%	13.85%	85.32%	14.67%
9	Colic.org	368	28	68.47%	31.52%	66.30%	33.69%
10	Primary tumor	399	18	42.48%	57.52%	39.82%	60.17%
11	Balance Scale	625	25	80.48%	19.52%	76.64%	23.36%
12	Soybean	683	36	91.65%	8.34%	91.50%	8.49%
13	Credit a	690	16	85.07%	14.92%	86.09%	13.91%
14	Breast W	699	10	96.13%	3.68%	94.56%	5.43%
15	Vehicle	846	19	77.06%	22.93%	72.45%	27.54%
16	vowel	990	14	96.06%	3.03%	81.51%	18.48%
17	Credit g	1000	21	72.50%	27.50%	70.50%	29.50%
18	Segment	2310	20	97.66%	2.33%	96.92%	3.07%
19	Wave form	5000	41	81.94%	18.06%	75.30%	24.70%
20	Letter	20,000	17	94.71%	5.29%	87.98%	12.02%

FIG. 4.2: Comparaison du F1 score d'apprentissage de Random Forest (J48) et Arbre de décision sur 20 jeu de données

Dans les résultats présents dans le tableau ci-dessous, nous remarquons que dans les deux cas les algorithmes ont un pourcentage d'apprentissage assez élevé. Si nous comparons les scores des deux algorithmes d'apprentissage nous remarquons que le random forest donne de meilleurs résultats pour des grands ensemble de données c'est à dire des données avec des grands nombre d'instance alors que les Arbres de décision sont meilleurs sur des petits jeux de données avec de nombre d'instance assez faible. Les résultats des tests sur les données du cancer du sein montrent que lorsque le nombre d'instance est passé de 286 à 699, le nombre des instances correctement classés est passé de 69,23% à 96,13% pour les forêts aléatoires.

Dataset	No. of Instances	No. of attributes	Random forest			J-48		
			Precision	Recall	F-measure	Precision	Recall	F-measure
Breast cancer	286	10	0.667	0.692	0.674	0.752	0.755	0.713
Breast W	699	10	0.962	0.961	0.961	0.946	0.946	0.946
Credit a	690	16	0.851	0.851	0.851	0.861	0.861	0.861
Credit g	1000	21	0.705	0.725	0.707	0.687	0.705	0.692
colic	368	23	0.854	0.853	0.85	0.86	0.861	0.861
Colic.org	368	28	0.662	0.685	0.63	0.44	0.663	0.529
Heart-h	270	14	0.775	0.779	0.774	0.807	0.81	0.806
Heart-c	303	14	0.819	0.815	0.813	0.776	0.776	0.774
vowel	990	14	0.961	0.961	0.961	0.816	0.815	0.815
Ionosphere	351	35	0.929	0.929	0.929	0.915	0.915	0.913
Soyben	683	36	0.926	0.917	0.918	0.917	0.915	0.913
Vehicle	846	19	0.764	0.771	0.767	0.722	0.725	0.722
Sonar	208	61	0.813	0.808	0.808	0.713	0.712	0.712
Autos	205	26	0.836	0.834	0.834	0.833	0.822	0.82
Balance Scale	625	25	0.817	0.805	0.81	0.732	0.766	0.749
Lymph	148	19	0.804	0.811	0.8	0.776	0.77	0.772
<b>Segment</b>	<b>2310</b>	<b>20</b>	<b>0.977</b>	<b>0.977</b>	<b>0.977</b>	<b>0.969</b>	<b>0.969</b>	<b>0.969</b>
Primary tumor	399	18	0.394	0.425	0.406	0.333	0.398	0.704
Waveform	5000	41	0.82	0.819	0.82	0.753	0.753	0.753
Letter	20,000	17	0.948	0.947	0.947	0.881	0.88	0.88

FIG. 4.3: Comparaison des scores Precision, Recall et F-measure de Random Forest (J48) et Arbre de décision sur 20 jeu de données

Le tableau ci-dessous présente des valeurs de validation des algorithmes dans le tableau pour Random Forest et Arbre de décision pour les 20 jeux de données. Dans la table, on peut voir que pour le même ensemble de données avec une plus grande nombre d'instances, c'est-à-dire lorsque le nombre d'instances passe de 286 à 699 tout en conservant les attributs constante Precision augmente de 0,667 à 0,962, F-measure de 0,674 à 0,961 et Recall de 0,692 à 0.961 pour le classificateur Random Forest. De même pour le J48, Precision est passée de 0,752 à 0,946, F-measure de 0,713 à 0,946 et Recall de 0,755 à 0.946. La valeur de précision la plus élevée obtenue est pour le Random Forest, c'est-à-dire 0,977 pour l'ensemble de données. D'après les résultats, on peut conclure que le Random Forest atteint des performances de classification accrues et donne des résultats exacts et précis dans les cas de grand nombre d'instances. Ces scénarios couvrent également les problème de valeurs manquantes dans les jeux de données, il surmonte également le problème de sur-ajustement généré en raison de valeurs manquantes dans les ensembles de données. Par conséquent, pour les problèmes de classification, si l'on

doit choisir un classificateur parmi l'ensemble des classificateurs basés sur l'arbre, il est fortement recommandé d'utiliser le Random Forest en toute confiance pour divers problèmes de classification.

### 4.3 Prédiction de pannes DSL par mesure passive sur des passerelles domestiques

Dans cet article [1], l'objectif visé est de faire de la détection proactive des problèmes de ligne DSL. Pour ce faire les données qui ont été exploitées, sont des données collectées des informations liées au DSL à partir d'un ensemble de réseaux domestiques une période de deux années dont l'ensemble cumulative comprend 98 maisons d'essai sur une durée cumulée de cette expérience d'essai est de 35455 jours.

- **Expérimentations**  
Les compteurs CNT comptent le nombre d'évènements dans une période d'interrogation (une minute). Les tarifs et les valeurs de bruit sont les valeurs actuelles telles que récupérées par le client au moment de la collecte.

Feature	Description	Feature	Description
cntErrSecs	#secs with CRC err. or LoS	{up down}CurrRate	rate of up/downstream link
cntSevErrSecs	#secs with more than 18 CRC errors or one LoS	{up down}MaxRate	theoretical maximum up/downstream link rate according to line attenuation
cntLoS	loss of signal (LoS)	{up down}NoiseMargin	noise margin on up/downstream link
cntLoF	loss of framing (LoF)	linkStatus	DSL link status. Values : Up, NoLink, Initializing
cntFECErr	number of FEC errors		
cntCRCErr	number of CRC errors		
cntHECErr	number of HEC errors		

FIG. 4.4: Caractéristiques collectées pour chaque lien DSL

- **Evaluations et résultats**  
Les algorithmes qui ont été appliqués sur le jeu de données sont entre autres Random Forest, Naive Bayésien, AdaBoost, Décision Tree. Ci-dessous une figure comparative des résultats obtenus. Au vue de ces expérimentations, nous pouvons conclure que le Random Forest nous permet d'avoir les meilleurs résultats.

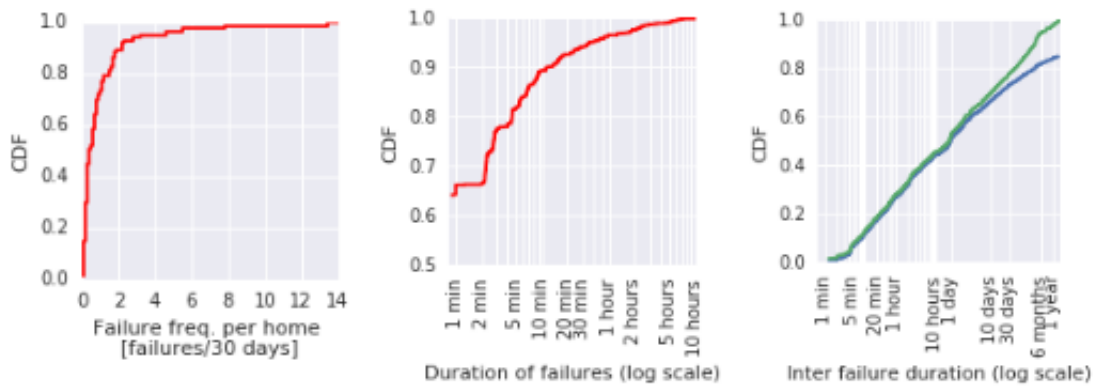


FIG. 4.5: Caractéristiques de défaillance

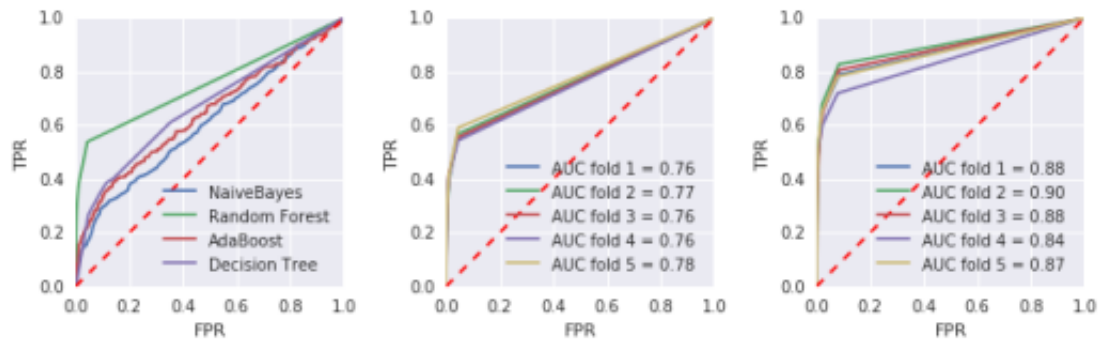


FIG. 4.6: Evaluation et choix du meilleur modèle

## 4.4 Conclusion

En somme ces techniques d'apprentissage supervisé nous permettront de faire de la maintenance prédictive.

**partie IV**

**Systeme réalisé**





# Sommaire

---

<b>5</b>	<b>Implémentation du système</b>	<b>27</b>
5.1	Introduction . . . . .	28
5.2	Expérimentations et résultats . . . . .	28
5.3	Conclusion . . . . .	33

---



# Chapitre 5

## Implémentation du système

### Sommaire

---

<b>5.1</b>	<b>Introduction . . . . .</b>	<b>28</b>
<b>5.2</b>	<b>Expérimentations et résultats . . . . .</b>	<b>28</b>
5.2.1	Expérimentations . . . . .	28
<b>5.3</b>	<b>Conclusion . . . . .</b>	<b>33</b>

---

## 5.1 Introduction

Comme nous l'avons expliqué dans les parties précédentes de notre mémoire, la problématique de notre sujet est de faire de la prédiction de matériels informatiques plus des ordinateurs afin de savoir éventuellement ceux qui sont susceptibles de tomber en panne. Il s'agit de prédiction de données quantitatives et qualitatives.

## 5.2 Expérimentations et résultats

Pour mettre en place notre solution, nous aurons plusieurs étapes notamment les étapes de nettoyage, de préparation de données, et de machine learning. Après étude de l'état de l'art, nous avons donc décidé d'appliquer comme algorithmes, les Arbres de Décision, les Random Forest et la Regression Logistique. Pour rappel, pour obtenir notre jeu de données final nous avons effectué des étapes de consolidation et d'agrégation de plusieurs jeux de données d'historisation sur des matériels informatiques. Ci-dessous, les étapes de notre travail.

### 5.2.1 Expérimentations

- Chargement du jeu de données brutes

	ComputerSystemId	Model	DateProd	IdleSeconde	ActiveSeconde	CPU	StandBySeconde	HibernateSeconde	PhysicalMemory
4	1756	10AU003DFR	07/01/2019 00:00	31840	23678	8	0	0	
5	1756	10AU003DFR	08/01/2019 00:00	65033	21367	4	0	0	
6	1756	10AU003DFR	09/01/2019 00:00	62836	23564	5	0	0	
7	1756	10AU003DFR	10/01/2019 00:00	66239	20161	4	0	0	
8	1756	10AU003DFR	14/01/2019 00:00	29568	26028	6	0	0	

5 rows × 28 columns

FIG. 5.1: Jeu de données sur des caractéristiques d'ordinateurs loués

La Figure nous montre quelques colonnes de notre jeu de données après y avoir fait les étapes de consolidation de table. Il s'agit d'un jeu de 28 colonnes avec une colonne de type date (DateProd) et une colonne de type chaîne de caractères (Model). C'est cette dernière colonne que nous allons prédire.

- Répartition des modèles d'ordinateur.

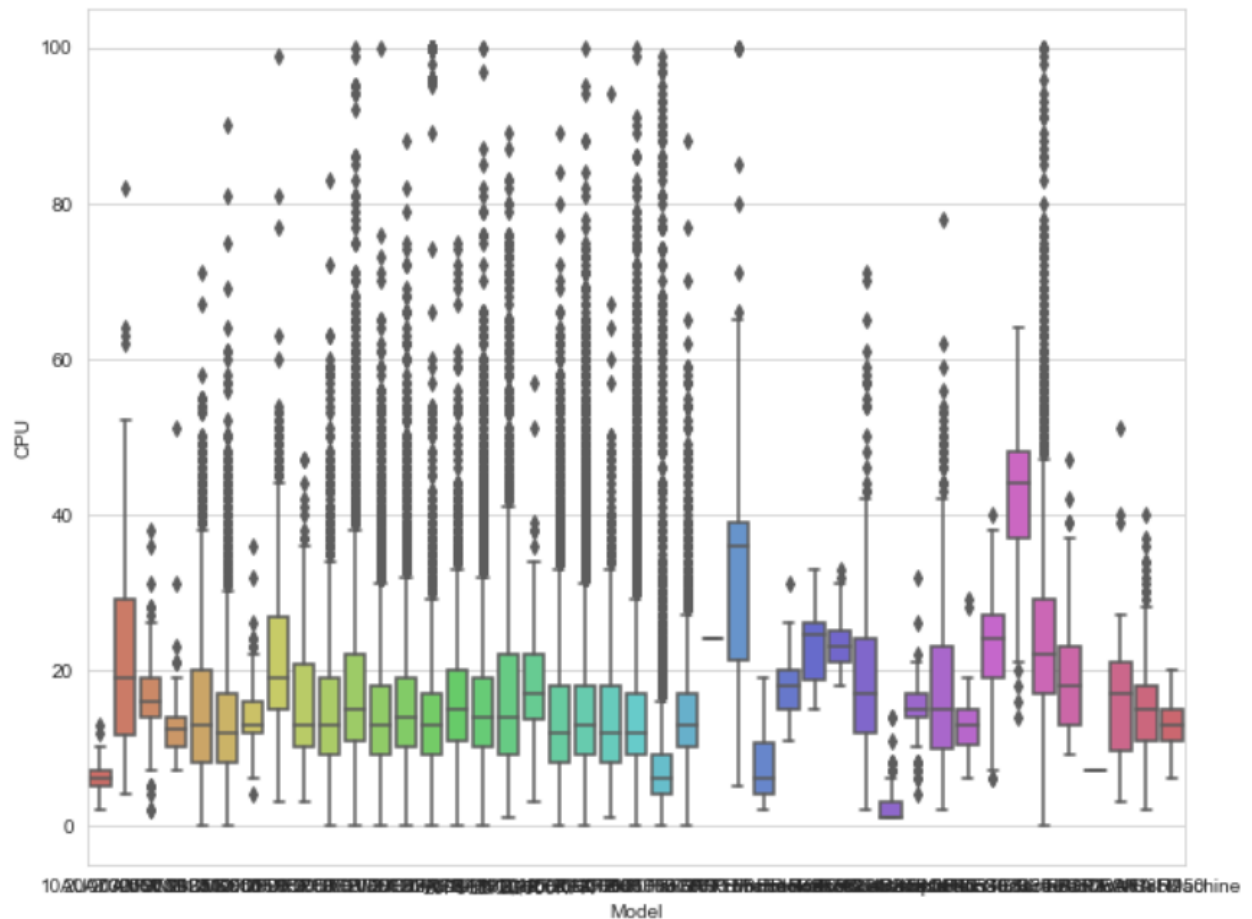
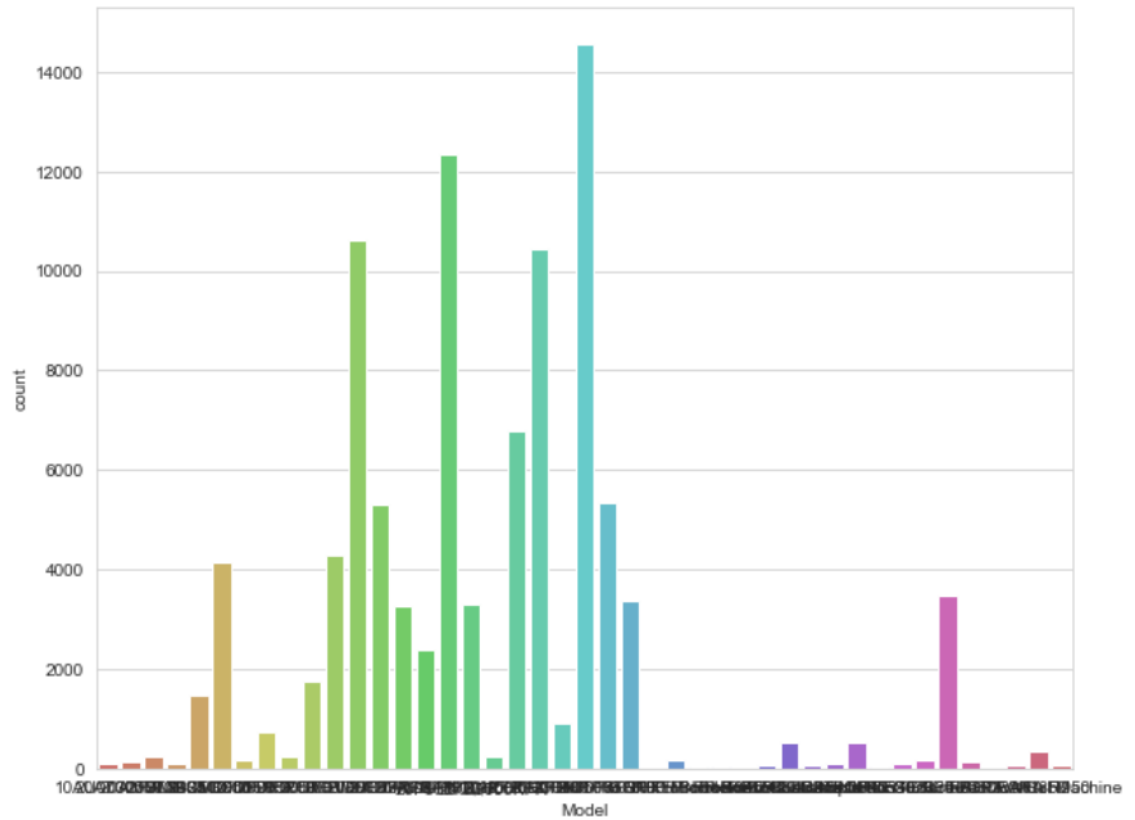


FIG. 5.2: Boite à moustache

Cette figure montre une répartition des modèles d'ordinateurs en fonction du processeur

- Histogramme sur les nombres d'ordinateurs par modèle



- Description du jeu de données préparées et nettoyées pour l'apprentissage  
Les caractéristiques du jeu de données nous montre types des colonnes. Pour ce jeu de données, nous avons eu à supprimer les lignes de données manquantes, à faire des conversion des formats, par exemple des type float en interger comme c'est le cas de plusieurs colonnes. Nous avons également enrichi nos données avec deux colonnes supplémentaires, une colonne **Size\_Model** qui transforme et affecte des numéros de classe à chaque valeur de la colonne Model, une seconde colonne NumberDays qui constitue le nombre de jours d'inactivité de l'ordinateur. Cette dernière colonne importante nous permettra donc avec les autres colonnes caractéristiques de l'ordinateur de prédire les modèles d'ordinateur qui vont tomber en panne.

```

Int64Index: 91281 entries, 92 to 98158
Data columns (total 28 columns):
ComputerSystemId      91281 non-null int64
IdleSeconde           91281 non-null int64
ActiveSeconde         91281 non-null int64
CPU                   91281 non-null int64
StandBySeconde        91281 non-null int64
HibernateSeconde      91281 non-null int64
PhysicalMemory         91281 non-null int64
MaxPhysicalMemory     91281 non-null int64
NbNormReboot          91281 non-null int64
NbAbnReboot           91281 non-null int64
TotalDataSentMo       91281 non-null int64
TotalDataReceivedMo   91281 non-null int64
TotalDiskReadMo       91281 non-null int64
TotalDiskWriteMo      91281 non-null int64
AvgDiskActivity       91281 non-null int64
MaxDiskActivity       91281 non-null int64
SLAComplianceTimeSeconde 91281 non-null int64
SLAStressTimeSeconde  91281 non-null int64
SLAHangTimeSeconde    91281 non-null int64
CpuStressTimeSeconde  91281 non-null int64
CpuHangTimeSeconde    91281 non-null int64
MemStressTimeSeconde  91281 non-null int64
MemHangTimeSeconde    91281 non-null int64
TotalPagesFault       91281 non-null int64
PacketsOutboundErrors 91281 non-null int64
PacketsReceivedErrors 91281 non-null int64
NumberDays            91281 non-null int64
Size_Model            91281 non-null int32
dtypes: int32(1), int64(27)
memory usage: 19.8 MB

```

FIG. 5.4: Description du jeu de données préparées

- Matrice de corrélation entre les variables  
La matrice ci-dessus présente les relations de lien encore appelées corrélations entre les différentes variables de notre jeu de données.

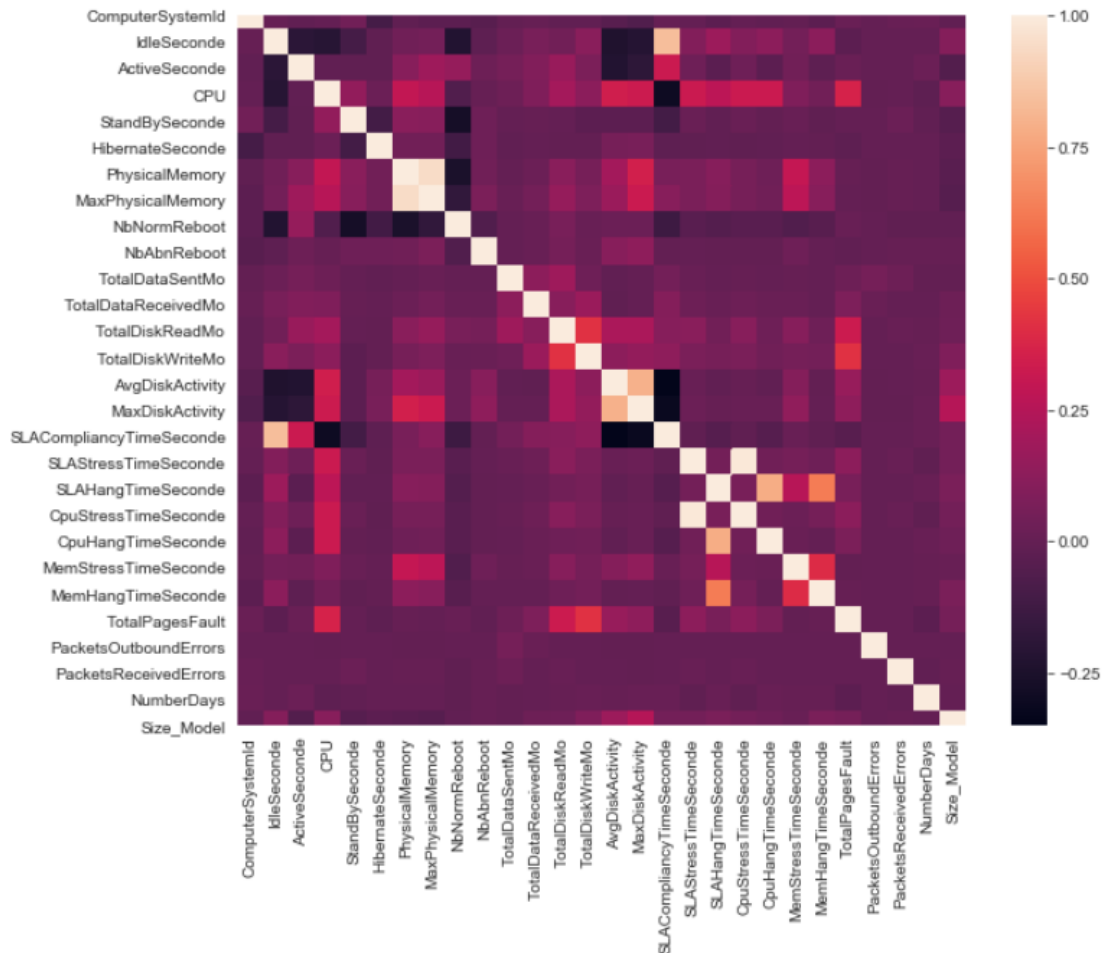


FIG. 5.5: Matrice de corrélation entre les variables

La matrice de corrélation nous a permis de vérifier si nécessaire que nous a permis de mieux analyser nos données afin de voir si éventuellement il y a des variables à forte corrélation avec d'autres variables. Si c'est le cas ces variables ne seront pas gardées pour l'apprentissage car elle ne permettront pas de donner de bons résultats.



- Comparaison des algorithmes

La figure ci-dessus compare les scores d'apprentissage des trois algorithmes que nous avons appliqué sur notre jeu de données notamment le Random Forest, les Arbres de Décision et La Regression Logistique.

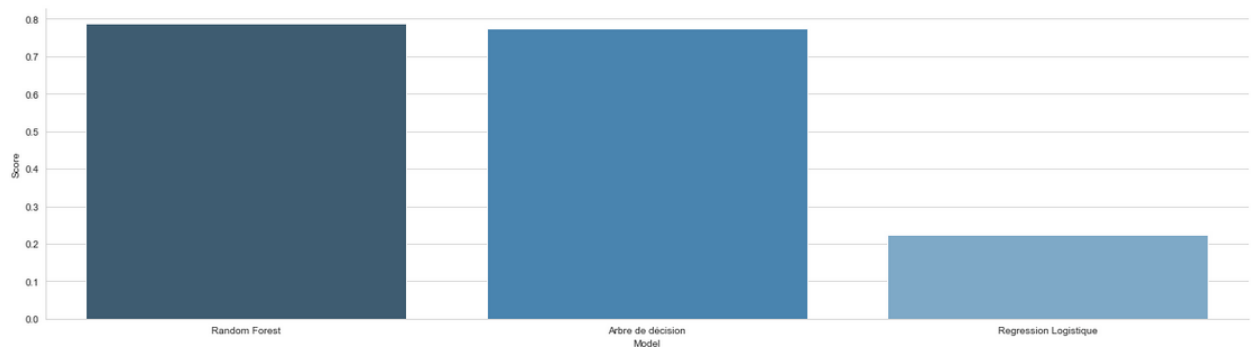


FIG. 5.6: Comparaison des algorithmes

Nous pouvons conclure que le Random Forest donne des résultats meilleurs par rapport aux deux autres algorithmes que sont les Arbres de Décision et la Regression Logistique.

## 5.3 Conclusion

Au cours de ces expérimentations, nous avons donc eu à manipuler et traiter nos données de différentes manières. Toutefois les étapes de pré-traitement nous ont permis notamment d'enrichir nos données, de les nettoyer afin d'obtenir des résultats relativement bonnes pour la phase d'apprentissage.



# Table des figures

1.1	Organigramme du groupe . . . . .	8
1.2	Les domaines d'expertise chez ECONOCOM . . . . .	10
4.1	Caractéristiques et détails sur le jeu de données . . . . .	20
4.2	Comparaison du F1 score de Random Forest (J48) et Arbre de décision sur 20 jeu de données . . . . .	21
4.3	Comparaison des scores Precision, Recall et F-measure de Random Forest (J48) et Arbre de décision sur 20 jeu de données . . . . .	22
4.4	Caractéristiques collectées pour chaque lien DSL . . . . .	23
4.5	Caractéristiques de défaillance . . . . .	24
4.6	Evaluation et choix du meilleur modèle . . . . .	24
5.1	Jeu de données sur des caractéristiques d'ordinateurs loués . . . . .	28
5.2	Boîte à moustache . . . . .	29
5.3	Histogramme sur les nombres d'ordinateur spar modèle . . . . .	30
5.4	Description du jeu de données préparées . . . . .	31
5.5	Matrice de corrélation entre les variables . . . . .	32
5.6	Comparaison des algorithmes . . . . .	33



## Liste des tableaux



# Table des matières

<b>I</b>	<b>Présentation de l'entreprise</b>	<b>5</b>
<b>1</b>	<b>Présentation de l'entreprise</b>	<b>7</b>
1.1	Présentation de l'Entreprise ECONOCOM . . . . .	8
1.2	Les domaines d'expertise chez ECONOCOM . . . . .	9
<b>II</b>	<b>Problématique</b>	<b>11</b>
<b>2</b>	<b>Le contexte de résolution du problème</b>	<b>13</b>
2.1	Introduction . . . . .	14
2.2	Conclusion . . . . .	14
<b>3</b>	<b>Le problème à résoudre</b>	<b>15</b>
3.1	Introduction . . . . .	16
3.1.1	Prédiction de pannes sur des matériels informatiques . . . . .	16
3.2	Conclusion . . . . .	16
<b>III</b>	<b>État de l'art</b>	<b>17</b>
<b>4</b>	<b>État de l'art des techniques</b>	<b>19</b>
4.1	Introduction . . . . .	20
4.2	Comparaison entre Random Forest et Decision Tree . . . . .	20
4.3	Prédiction de pannes DSL par mesure passive sur des passerelles domestiques . . . . .	23
4.4	Conclusion . . . . .	24

<b>IV</b>	<b>Système réalisé</b>	<b>25</b>
<b>5</b>	<b>Implémentation du système</b>	<b>27</b>
5.1	Introduction . . . . .	28
5.2	Expérimentations et résultats . . . . .	28
5.2.1	Expérimentations . . . . .	28
5.3	Conclusion . . . . .	33