

N° d'ordre :

N° de série :

République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique

UNIVERSITÉ ECHAHID HAMMA LAKHDAR D'EL OUED

FACULTÉ DES SCIENCES EXACTES

DEPARTEMENT D'INFORMATIQUE

Mémoire de fin d'étude

**MASTER ACADEMIQUE**

Domaine: Mathématiques et Informatique

Filière: Informatique

Spécialité: Systèmes Distribués et Intelligence Artificiel (SDIA)

**Thème**

**Vers un système de prédiction et des solutions  
relatives à la perte de boue de forage en utilisant  
les techniques de la fouille de données.**

Présenté par: - Ghenabzia Ahmed

- Kaid Hicham

Soutenu devant le jury composé de

M. Naoui Med Anouar	MA(B)	Rapporteur	Univ. d'El Oued
M.Gherbi kaddour	MAA	Président	Univ. d'El Oued
M.Yaakoub Amine	MAA	Examineur	Univ. d'El Oued

Année universitaire 2015-2016.

## Remerciement

*Nous remercions tout d'abord notre Dieu qui nous a donné la force et la volonté pour élaborer ce travail.*

*Nous adressons nos vifs remerciements à notre encadreur “**Naoui Med Anouar** “, qui n’a pas cessé de nous donner les conseils et les bonnes orientations et nous prive pas de son temps et aussi on remercie “**Mr Merouane Zoubeidi** “, qui nous écoute avec grande patience pendant toute la période de préparation malgré ses contraintes d’enseignement.*

*Sans son aide, notre travail n'aurait pas vu la lumière. Notre reconnaissance va aussi à tous ceux qui ont collaboré à notre formation en particulier les enseignants du département d'Informatique, Universitaire Echahide Hamma Lakhdar El-Oued*

*Aussi à nos collègues de la promotion 2015-2016 On remercie également tous ceux qui ont participé de près ou de loin à élaborer ce travail*



## **Résumé**

La perte de circulation est un problème habituel dans l'industrie de forage pétrolier, ce qui nécessite une très grande dépense pécuniaire pour les entreprises de forage. Aujourd'hui, la minimisation des pertes de boue de forage reste l'objectif principal et la préférence pour les superintendants de forage. Les techniques de la fouille de données sont une nouvelle méthode de résolution de problèmes d'ingénierie, car elles ont la capacité d'examiner tous les paramètres efficaces simultanément. En outre, ces techniques augmentent et facilitent la capacité d'apprentissage directement à partir de données réelle du terrain.

L'objet de notre étude est d'anticiper en premier lieu les causes de survenue de ces problèmes afin d'éviter les risques et les coûts de forage excessifs car si ces risques sont identifiés à l'avance, de meilleures solutions peuvent être préconisées pour réduire les conséquences négatives. Ensuite, en deuxième lieu, notre sujet consiste aussi à essayer de mettre en place un système de prédiction de la perte de boue de forage en utilisant les algorithmes de la fouille des données Datamining.

### **Mots clés :**

Forage pétrolier, perte de boue de forage, prédiction, Datamining, perte de circulation.

## **Abstract**

The loss circulation is a big problem in the oil drilling industry, which requires a very large financial expense for drilling companies. Today, the minimization of drilling mud losses remains the primary objective and preference for drilling superintendents. data mining techniques are a new approach to solving engineering problems, because they have the ability to examine all effective parameters simultaneously. In addition, these techniques increase and facilitate the ability to learn directly from actual field data.

The purpose of our study is anticipated first occurrence of the causes of these problems in order to avoid excessive risks and drilling costs as if such risks are identified in advance, better solutions can be recommended for reduce the negative consequences. Then, secondly, our subject is also to try to set up a drilling mud loss prediction system using the algorithms of the search Data mining of data.

### **Keywords :**

Oil drilling, Drilling mud loss, Prediction, Data Mining, Lost circulation.

## المخلص

تعمل الشركات النفطية المتخصصة في مجال التنقيب عن النفط إلى إنقاص تكلفة الحفر وتسريعه.

فقدان رغبة الحفر هو من أكبر المشاكل المتكررة التي تواجه الشركات الأمر الذي يترتب عليها عدة خسائر مالية أو بشرية. الهدف الأساسي لمدراء هذه الشركات هو التقليل من فقدان رغبة الحفر.

تنقيب واستخراج البيانات هو مجال جديد من مجالات الذكاء الاصطناعي يستعمل في حل مثل هذه المشاكل من خلال القدرة على فحص البيانات وإيجاد علاقات بينها والتعلم الآلة مباشرة من خلال البيانات الميدانية الفعلية.

الهدف من هذه الدراسة محاولة اكتشاف الخصائص المؤثرة في فقدان رغبة الحفر والبحث عن وسيلة لتجنبه وإنقاص التكاليف والحد من مخاطره وإنشاء نظام للتنبؤ بفقدان الرغبة باستعمال خوارزميات تنقيب واستخراج البيانات.

### الكلمات المفتاحية :

التنقيب على النفط ، فقدان رغبة الحفر ، التنبؤ ، تنقيب واستخراج البيانات ، ضياع الرغبة

## ***Liste de tableaux***

***Tableau III.1 : Les ranges des paramètres utilisés dans la modélisation.....46***

***Tableau III.2 : Tableau comparatif entre les différentes méthodes étudiées.....49***

## ***Listes des figures***

<b><i>Figure I.1 :L'équipement de forage</i></b> .....	3
<b><i>Figure I.2 : L'équipement de forge</i></b> .....	4
<b><i>Figure I.3 : l'installation de forage et la garniture</i></b> .....	7
<b><i>Figure I.4: Perte partiel et Perte totale</i></b> .....	16
<b><i>Figure II.1: Processus de data-mining</i></b> .....	20
<b><i>Figure II.2 : Le neurone biologique</i></b> .....	29
<b><i>Figure III.1:Le diagramme de séquence.</i></b> .....	44
<b><i>Figure III.2 : la validation croisée de support vecteur machine sous WEKA</i></b> .....	46
<b><i>Figure III.3 : la validation croisée de réseau de neurone sous WEKA</i></b> .....	47
<b><i>Figure III.4 : Architecture de RN utilisé pour la prédiction de la perte</i></b> .....	47
<b><i>Figure III.5 : la validation croisée de l'arbre de décisions sous WEKA</i></b> .....	48
<b><i>Figure III.6 : l'arbre de décision de la modèle utilisé sous WEKA</i></b> .....	48
<b><i>Figure III.7 : l la validation croisée du k plus proche voisin sous WEKA</i></b> .....	49
<b><i>Figure III.8: L'architecture de système de prédiction</i></b> .....	51
<b><i>Figure IV.1: Environnement de développement</i></b> .....	54
<b><i>Figure IV.2: L'architecture de JVM et JDK</i></b> .....	55
<b><i>Figure IV.3 : L'espace de travail Weka</i></b> .....	55
<b><i>Figure IV.4: L'interface de ANN</i></b> .....	57
<b><i>Figure IV.5: L'interface de DT</i></b> .....	57
<b><i>Figure IV.6: L'interface de KNN</i></b> .....	58
<b><i>Figure IV.7 : L'interface de SVM</i></b> .....	58

# *Sommaire*

<b>Introduction général .....</b>	<b>1</b>
-----------------------------------	----------

## **Chapitre I: Introduction au forage pétrolier**

<b>I – Introduction.....</b>	<b>2</b>
<b>II – Le Forage.....</b>	<b>3</b>
<b>II .1 Définition.....</b>	<b>3</b>
<b>II .2 Type de forage.....</b>	<b>5</b>
1- Le forage d'exploration.....	5
2- Le forage d'expertise .....	5
3- Le forage de développement.....	5
<b>II .3 Processus de forage.....</b>	<b>5</b>
<b>II .4 La Cimentation.....</b>	<b>6</b>
<b>II .5 Les paramètres de forage.....</b>	<b>8</b>
1 - Les paramètres mécaniques.....	8
2 - Les Paramètres hydrauliques .....	9
<b>II .6 Les Problèmes de forages.....</b>	<b>13</b>
<b>III – Conclusion.....</b>	<b>17</b>

## **Chapitre II: Data-mining**

<b>I - Introduction.....</b>	<b>19</b>
<b>II - Définition.....</b>	<b>19</b>
<b>III - Le processus de data-mining.....</b>	<b>19</b>
<b>IV - Les taches de data-mining.....</b>	<b>21</b>
1 -La classification.....	21
2 - La régression.....	22
3 - La prédiction.....	22

4 - Le groupement par similitude.....	22
5 - L'analyse des clusters.....	22
<b>V - Les algorithmes de data-mining :</b> .....	<b>23</b>
1- SVM (Support Vector Machine) .....	23
2- Le Réseau de neurones artificiel .....	29
3 - K-Means .....	33
4 - K Nearest Neighbors.....	34
5 - Naïve Bayésien.....	36
6 - Arbre de décision.....	36
<b>VI - Conclusion</b> .....	<b>41</b>

### **Chapitre III: Conception**

<b>I - Introduction</b> .....	<b>42</b>
<b>II - Architecture Globale de système</b> .....	<b>42</b>
<b>III - Fonctionnement globale de système</b> .....	<b>43</b>
<b>IV - Choix de la méthode</b> .....	<b>45</b>
<b>V - L'architecture de système de prédiction</b> .....	<b>50</b>
1- Partie d'apprentissage automatique.....	51
2- Partie de prédiction.....	51
3- Le réapprentissage.....	51
<b>VI - Conclusion</b> .....	<b>52</b>

### **Chapitre IV: Implémentation et réalisation**

<b>I - Introduction</b> .....	<b>53</b>
<b>II - Environnement du développement</b> .....	<b>53</b>
1 - Java.....	53

1.1 - Environnement de programmation (Eclipse) .....	53
1.2 - JDK .....	54
1.3 - JVM.....	54
2 - Weka (Waikato environment for knowledge analysis) .....	55
<b>III - Résultat obtenues .....</b>	<b>56</b>
<b>IV - Conclusion.....</b>	<b>58</b>
<b>Conclusion Générale.....</b>	<b>59</b>
<b>Bibliographie.....</b>	<b>60</b>

Le forage représente l'ensemble des opérations permettant de creuser pour atteindre -en profondeur- de nouvelles zones (réservoirs) susceptibles de contenir des hydrocarbures.

La finalité d'un forage est d'atteindre l'objectif, qui est le réservoir, dans les plus brefs délais et avec un prix de revient minimum, sans l'endommager ni compromettre son exploitation (production, injection). Et pour y parvenir, plusieurs techniques ainsi que plusieurs équipements ont été développés selon les besoins et les difficultés rencontrées durant le forage.

### **Problématique :**

La perte de circulation est un problème habituel dans l'industrie de forage pétrolière, ce qui nécessite une très grande dépense pécuniaire pour les entreprises de forage. Aujourd'hui, la minimisation des pertes de boue de forage reste l'objectif principal et la préférence pour les superintendants de forage. Les techniques de la fouille de données sont une nouvelle méthode de résolution de problèmes d'ingénierie, car elles ont la capacité d'examiner tous les paramètres efficaces simultanément. En outre, ces techniques augmentent et facilitent la capacité d'apprentissage directement à partir de données réelle du terrain.

### **L'objectif :**

L'objet de notre étude est d'anticiper en premier lieu les causes de survenue de ces problèmes afin d'éviter les risques et les coûts de forage excessifs car si ces risques sont identifiés à l'avance, de meilleures solutions peuvent être préconisées pour réduire les conséquences négatives. Ensuite, en deuxième lieu, notre sujet consiste aussi à essayer de mettre en place un système de prédiction de la perte de boue de forage en utilisant les algorithmes de la fouille des données Datamining.

Ce mémoire s'articule autour de quatre chapitres :

- Le premier chapitre présente un état de l'art contenant les notions de base du forage pétrolier et ces divers problèmes dont la perte de la boue de forage qui représente notre principale problématique
- Le second chapitre illustre et expose les différentes approches, méthodes et technique de Datamining.
- Le troisième chapitre présente les différentes approches existantes de régression et compare entre ces approches.
- Le dernier chapitre présente la conception et l'implémentation de système réalisé.

Enfin, nous concluons ce projet par une conclusion générale.

***Chapitre I:***

***Introduction au forage pétrolier***

## **I - Introduction :**

Le pétrole est une roche liquide d'origine naturelle, une huile minérale composée d'une multitude de composés organiques, essentiellement des hydrocarbures, piégée dans des formations géologiques particulières

Sa formation est le fruit de la transformation lente de matière organique, souvent végétale, qui se dépose sur le fond des océans et qui, sous l'effet de la pression des autres couches qui s'accumulent et d'une augmentation de température se transforme peu à peu en kérogène, puis en pétrole. Ce pétrole se retrouve ensuite piégé par des roches non perméables. Ce sont ces sites (gisements) que les compagnies pétrolières traquent afin d'effectuer des forages pour enfin, l'exploiter.

L'exploration de ce pétrole ou gaz, que ce soit en mer (offshore) ou sur terre (onshore), nécessite un processus de forage pour extraire la matière et parvenir aux réservoirs d'hydrocarbures. L'acheminement des hydrocarbures impose de connecter les réservoirs à la surface par des canaux reliés à des systèmes de pompage.

Si le forage aboutit à la découverte d'hydrocarbures exploitables, cet investissement considérable est remboursé par la production de pétrole et de gaz. Mais s'il ne révèle aucune réserve pétrolière ou gazière exploitable, la compagnie perd l'argent investi. C'est pourquoi la décision de forer nécessite une évaluation précise tenant compte des incertitudes et des risques liés à l'opération, et une étude économique complète pour vérifier la rentabilité du gisement si les prévisions attachées au prospect se réalisent.

Ajouter à cela, souvent pendant ces opérations de forage on rencontre divers problèmes qui font alourdir la note. Parmi ces incidents le perte de circulation, celui-ci peut se produire durant le forage, la connexion, le logging, les tests, ou durant n'importe quel opération qui implique de le perte de boue de forage; Les problèmes de perte de circulation sont généralement divisés en deux catégories: perte partiel, et perte totale.

Avant d'en dire plus sur les perte de circulation nous allons d'abord expliquer les bases du forage, afin de mieux assimiler ces notions. Nous allons donc commencer par définir le forage et ses processus, ensuite nous parlerons de l'équipement utilisé, enfin expliquer les causes des pertes de boue de forage.

## II - LE FORAGE

### II.1 Définition :

Un forage est un ensemble de techniques permettant de creuser un puits jusqu'à des profondeurs parfois très élevées.

Le forage est un procédé de découpe qui utilise un foret de couper ou agrandir un trou de section circulaire en matériaux solides. Le trépan de forage “, en anglais Drill bit” est un outil de coupe rotatif, souvent multipoint. Le bit est pressée contre la pièce et tourner à des vitesses de plusieurs centaines à plusieurs milliers de tours par minute. [16]

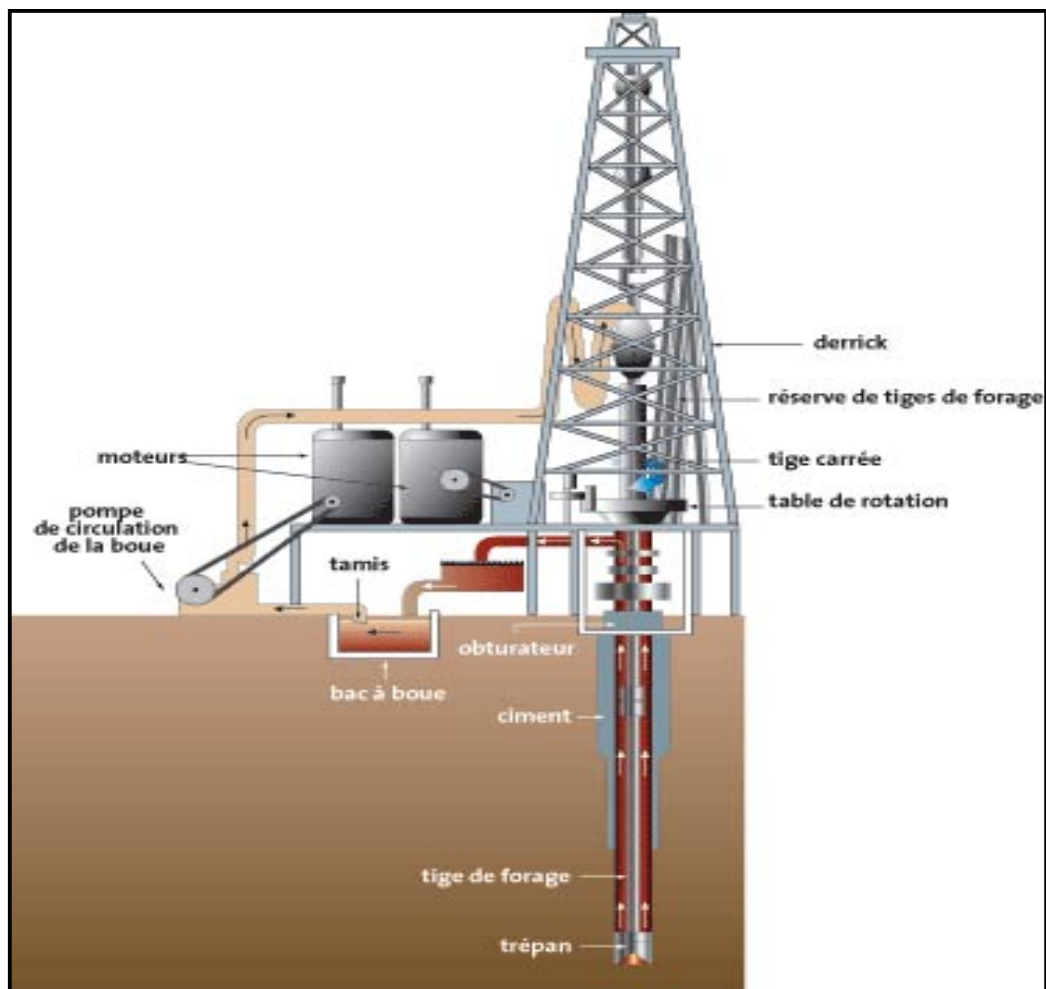
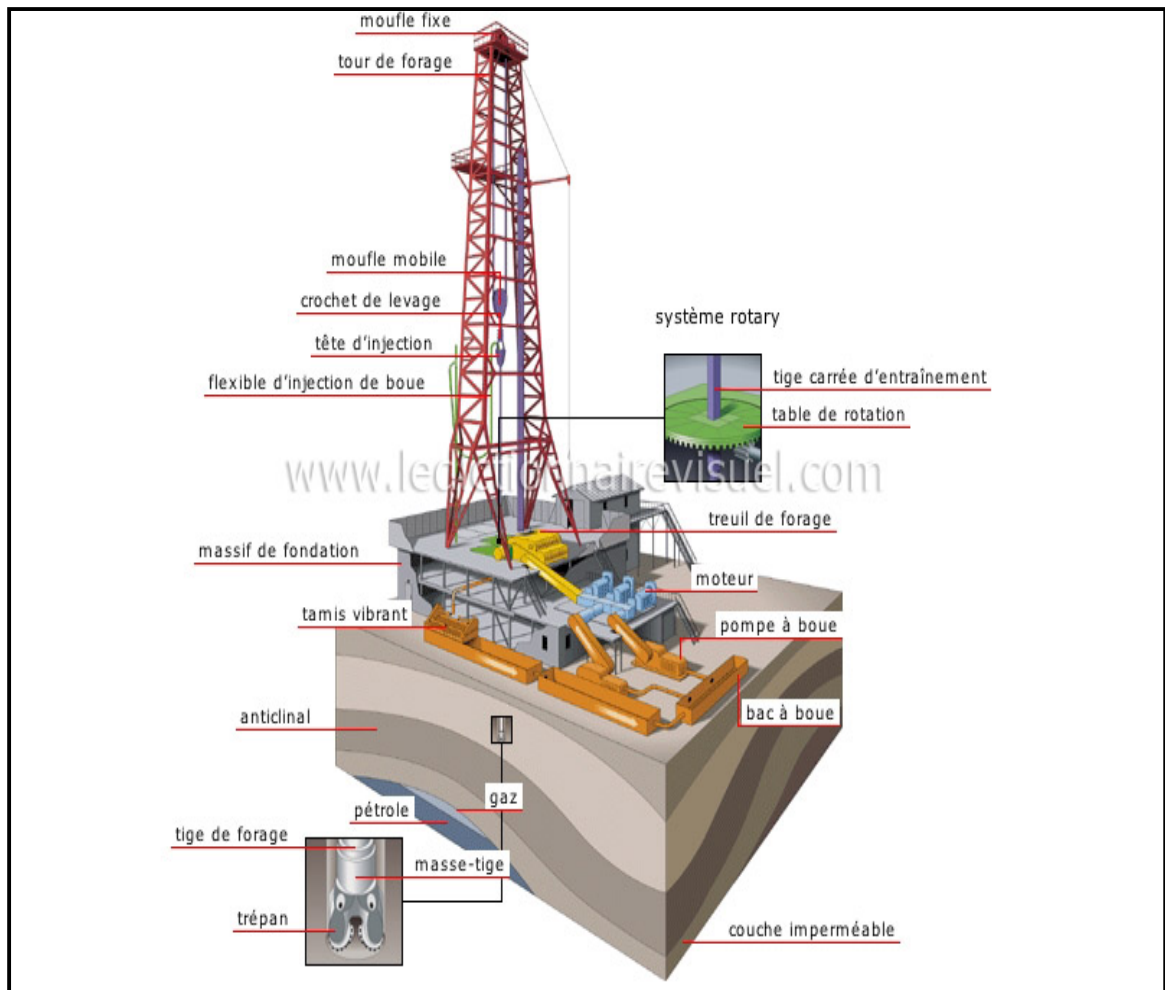


Figure I.1 :L'équipement de forage



**Figure I.2 : L'équipement de forage**

Le forage pétrolier est une cavité approximativement cylindrique, son diamètre varie en fonction de l'outil de forage "trépan", en anglais Drill bit, et la diminution de la profondeur atteint 36" (~ 91 cm) à 6,5" (~ 16,5 cm) d'un forage peut être exploration pour explorer une zone et vérifier les résultats de l'étude sismique pour confirmer la présence d'hydrocarbures, ou de développement afin d'en extraire le pétrole et le gaz présent dans le sous-sol, car il peut être tout à fait vertical (vertical) ou dirigé (diagonale à l'horizontale) il permet de spécifier les dimensions du champ et la portée des réserves qui ne sont pas directement sous le forage de l'installation (offshore par exemple).

## **II .2 Type de forage**

### **1) Le forage d'exploration :**

Le premier forage à effectuer pour confirmer l'existence des fluides dans la roche réservoir, puis définir la nature de ce dernier. [20]

### **2) Le forage d'expertise :**

L'appréciation, l'objectif de ce dernier est de préciser ou de compléter les informations fournies par le forage précédent, en obtenant des renseignements sur le gisement et déterminant ses caractéristiques lointaines, ses limites, sa rentabilité, afin de prendre la décision sur son développement. [6]

### **3) Le forage de développement :**

L'objectif principal de ce type de forage est d'effectuer des puits en service ou en production.

## **II .3 Processus de forage**

Les puits de pétrole et de gaz sont caractérisés par leur Extrémité tranchant sur le côté de l'entrée et de la présence de bavures sur le côté de sortie (sauf si elles ont été supprimées). En outre, l'intérieur du puits a généralement repères de margeur hélicoïdales.

Le forage peut affecter les propriétés mécaniques de la pièce en créant de faibles contraintes résiduelles autour de l'ouverture du trou et une très fine couche de matériau fortement sollicités et perturbé sur la surface nouvellement formée. Cela provoque la pièce pour devenir plus sensibles à la corrosion et à la propagation des fissures à la surface souligné. Une opération de finition peut être fait pour éviter ces conditions néfastes.

Pour forets cannelés, les copeaux sont évacués par les flûtes. Puces peuvent former de longues spirales (indésirables) ou de petits flocons, en fonction du matériau et des paramètres du procédé. Le type de puces formés peuvent être un indicateur de l'usinabilité du matériau, avec de longs copeaux suggérant une mauvaise usinabilité du matériau. [7]

Lorsque les trous forés possibles doivent être situés perpendiculairement à la surface de la pièce. Ceci minimise la Forets tendance à "marche", qui est, pour détourner de l'axe prévu de l'alésage, ce qui provoque le trou pour être déplacée. Plus le rapport de la longueur au diamètre du foret, plus la tendance à marcher. La tendance à marcher est également préemptée de diverses autres façons, notamment:

- Établir une marque de centrage ou une fonction avant de percer, comme par: Coulage, moulage, le forgeage ou une marque dans la pièce Pointage
- forage Spot (à savoir, un centre de forage) Lamage, qui fait face à une certaine zone sur un crépissage ou forgeage d'établir, pour l'essentiel, une île de la surface connue avec précision dans une mer de surface imprécise connu
- Contraindre la position du trépan de forage en utilisant un gabarit de perçage avec des douilles de forage
- Finition produit par le forage peut aller de 32 à 500 micro-pouces. coupes de finition généreront surfaces près de 32 micro-pouces et ébauche seront près de 500 micro-pouces.
- fluide de coupe est couramment utilisé pour refroidir le foret, augmenter la vie de l'outil, la vitesse d'augmentation et RSS, augmenter la finition de surface, et de l'aide dans les puces d'éjection. L'application de ces fluides est habituellement effectuée par inondation de la pièce de liquide de refroidissement et de lubrification ou par application d'un brouillard de pulvérisation.

En décidant qui forage (s) à utiliser, il est important de considérer la tâche à accomplir et évaluer quels forage serait mieux accomplir la tâche. Il existe une variété de styles de forage qui servent chacun un but différent. La perceuse sur blond est capable de forer plus d'un diamètre. Le foret plat est utilisé pour percer des trous plus grands formats. Le foret indéxable est utile dans la gestion de jetons. [8]

## **II .4 La Cémentation**

La cimentation est réalisé par circulation d'un laitier de ciment à travers l'intérieur du tubage et à l'extérieur dans l'espace annulaire à travers le sabot de tubage à la partie inférieure de la colonne de tubage. Afin de placer avec précision le coulis de ciment dans un espace de temps désiré à l'extérieur du tubage, un bouchon est pompé avec un fluide de déplacement derrière la colonne de ciment en suspension, qui "bosses" dans le sabot de tubage et empêche la poursuite de l'écoulement de fluide à travers le sabot. Cette bosse peut être vue à la surface comme un pic de pression à la pompe de ciment. Pour éviter que le ciment de refluer à

l'intérieur du tubage, un collier flottant au-dessus des actes de sabot de tubage d'un clapet anti-retour et empêche le fluide de circuler à travers la chaussure de l'espace annulaire. [9]

Un système de forage appelé aussi RIG est composé de deux parties : l'installation de forage et la garniture, comme le montre le schéma simplifié de la figure

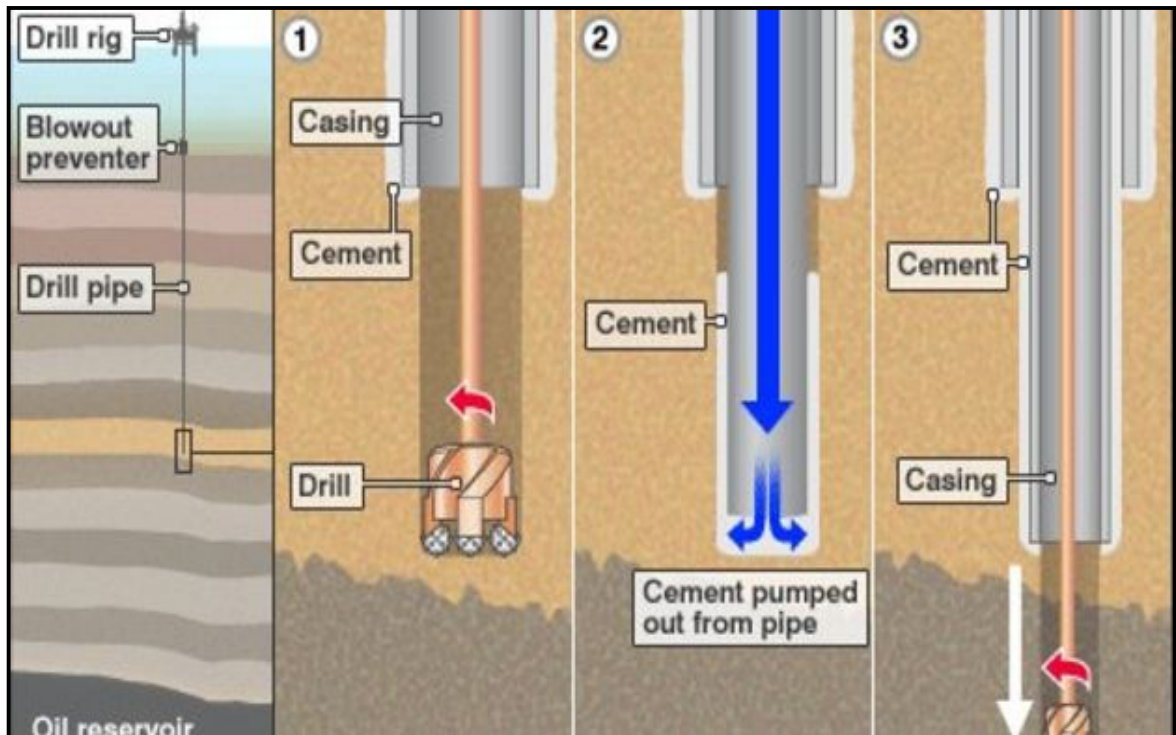


Figure I. 3 : l'installation de forage et la garniture

L'installation de forage est la partie située à la surface. Elle comprend une tour, appelée « derrick » ou « mât », sur laquelle sont situés les équipements de levage et de plancher qui permettent la manutention, le visage et dévissage des tiges et le changement de l'outil (trépan). [11] Le matériel employé pour le levage sert aussi à suspendre la garniture de forage dans le puits, à faire monter et descendre cette garniture dans le puits et finalement il permet d'ajuster le poids sur l'outil (trépan) qui est nécessaire pour que ce dernier puisse forer. Ce système inclus :

- un système de suspension
- un système rotary
- un système de production d'énergie
- et enfin un système de contrôle du puits.

La garniture désigne la partie souterraine du système de forage. Elle est constituée, de deux parties principales :

- La partie haute formée du train des tiges (Drill Strings)
- La partie basse formée de l'assemblage de fond que l'on a coutume de désigner par BHA « Bottom Hole Assembly ». (Plus de détaille sur le rôle de chaque équipement dans l'annexe 01- Equipement de forage) [10]

## **II .5 Les paramètres de forage**

On appelle paramètres de forage les différents facteurs qui conditionnent la vitesse d'avancement d'un outil de forage. Ils peuvent être classés en deux catégories :

- 1- Les paramètres mécaniques
- 2- Les paramètres hydrauliques

### **1 - Les paramètres mécaniques**

Les terrains traversés présentent une grande variété de dureté. Les avancements peuvent varier de 80 mètres/heure à quelques dizaines de centimètres par heure. Il est évident que pour forer des roches de différentes duretés on utilise des outils bien spécifiques selon la nature géologique.

#### **1.1-Poids sur l'outil**

A une vitesse de rotation constante, l'augmentation du poids sur l'outil contribue une augmentation de la vitesse d'avancement, à peu près en proportion directe si le débit de Circulation est suffisant. Le poids sur l'outil(WOB) c'est un paramètre calculer selon un principe physique, prenons on considération le poids de la garniture libre(WOHL) et le poids de la garniture durant le forage(WOHF). [12]

$$\text{WOB} = \text{WOHL} - \text{WOHF}$$

La charge de tous la garniture est supportée par le corps du cône qui s'applique contre la formation et il en résulte une diminution de la vie de l'outil. Cette charge est d'autant plus importante dans les terrains de plus en plus durs. le range maximale et minimale du poids sur l'outil (WOB) est limité par le constructeur.

**1.2-Vitesse de rotation**

La vitesse de rotation (RPM) c'est un paramètre physique qui indique le nombre de tour de la table dans le rig-floor par minute, identifié a l'aide d'un capteur.

La vitesse de rotation croît en fonction de la dureté de la formation et le paramètre torque. Et peut être limitée également par Les vibration qui affect la garniture de forage (phénomène de résonance qui peut être la cause de fatigues et de ruptures) .

**1.3-torque**

C'est un paramètre physique enregistré au cour du forage par un capteur, a vrai dire le torque et la force opposite de la rotation tous dépend aussi de la nature géologique des roche traversés comme le poids, le couple en surface n'est pas transmis intégralement sur l'outil de forage, mais la mesure de surface est la seule possible actuellement.

Un capteur à effet Hall est installé autour du conducteur d'amenée du courant (mesure de la consommation de courant électrique par le moteur de la table de rotation).

Le torque indique :

- l'état de l'outil et surtout de ses roulements
- les changements de lithologie
- molette coincée
- éboulement sur l'outil
- transmission du couple lors d'un back-off y tentative de décroincement

**2 - Les Paramètres hydrauliques****2.1-Débit FLOW IN**

C'est un volume de boue diverse par les pompes de forage a travers le stand pipe et le drill string à l'intérieur du trou de forage.

La quantification du débit est baser sur le débit unitaire de des pompes(POP) et le nombre de coup par minute contribue par les pompes(SPM), le SPM il est enregistré à l'aide d'un capteur de fréquence similaire à celle de la rotation, L'unité de débit est le (L/mn)

Les pompes de forage sont caractérisées par:

- Le nombre de pistons (02 duplex, 03 triplex)
- Le diamètre de la chemise (piston)
- La Course (10 '' ou 12'')
- Le débit unitaire et son efficacité

### **2.1.1-Procédure pour le comptage de l'efficacité des pompes**

On pompe à partir d'un bac actif environ 250 (à 300) coups de pompes et Le retour dans un autre bac

On mesure la variation de volume en centimètre dans le bac actif duquel on a pompé connaissant le volume unitaire de la pompe (exemple 19.57 l/ coup) et le volume linéaire du bac actif

Alors

$$250 \times 19.57 \rightarrow 100 \%$$

$$\text{Var Vol actif} \rightarrow X$$

$$X = \text{Var Vol actif} \times 100 / 250 \times 19.57 (\%)$$

### **2.1.2-Notion de pertes de charge**

#### **2.1.2.1 -Définition**

C'est la résistance à l'écoulement des fluides Les pertes de charge dépendent:

- ✓ de l'écoulement
- ✓ du circuit
- ✓ du fluide

Elles sont essentiellement dues aux frottements fluides avec fluide et fluide avec parois, il faut noter aussi qu'En l'absence d'écoulement:

### 2.1.2.2 -les pertes de charge sont nulles

Les pertes de charge dépendent:

- ✓ du débit
- ✓ de la longueur du circuit
- ✓ du diamètre du circuit
- ✓ de la densité et la rhéologie du fluide

Elles sont données par la formule approchée:

$$P_c = K \times (d \times L \times Q^2) / D^5$$

- K : Constante
- d : densité du fluide
- L : longueur du circuit
- Q : débit dans le circuit
- D : diamètre du circuit

### 2.1.3-Choix des duses et débit de forage pour l'outil

Les duses de l'outil doivent être choisies dans une simulation hydraulique de façon à transmettre le maximum de puissance pour l'outil. L'idéal est d'avoir le HHP ~ 65 %. Selon la disponibilité des duses sur chantier, choisir le maximum de HHP pour l'outil.

## 2.2 -Notion de pression

### 2.2 .1 -Pression hydrostatique :

$$P_h = h \times d / 10.2 \text{ (bar)}$$

- h est la hauteur de fluide (m)
- d est la densité (kg/l)

**2.2.2-Pression exercée sur le fond d'un puits :**

$$P_{\text{fond}} = P_h + P_c \text{ annulaire}$$

Ou :

- $P_h$  est la pression hydrostatique
- $P_c \text{ annulaire}$  = pertes de charge annulaire

**2.2.3-A l'état statique: (pas de circulation)**

$$P_{\text{fond}} = P_h \text{ (} P_c \text{ annulaires} = 0 \text{)}$$

- En circulation:  $P_{\text{fond}} = P_h + P_c \text{ annulaires} > P_h$

$P_c \text{ annulaires}$

**2.2.4-Pression de pore et de fracturation ( $P_{\text{pore}}$ ) / ( $P_{\text{frac}}$ ) :**

Chaque formation est caractérisée par:

- La pression de pore
- la pression de fracturation La pression de pore d'une formation est la pression de l'effluent qu'elle contient, La pression de fracturation est déterminée par un leak off test, c'est la pression à exercer sur la formation jusqu'à initier l'injection du fluide dans la formation cette pression est la pression admissible qu'on ne doit pas atteindre.

**A- Densité équivalente :**

C'est la densité correspondant à la pression exercée sur le fond du puits En statique (pas de circulation):  $d_{eq} = d$  (densité de la boue) En dynamique (avec circulation)  $d_{eq} = ECD$  (densité équivalente en circulation)

**B- Densité requise :**

C'est la densité nécessaire pour équilibrer la pression de pore de la formation Surge et Swab en manœuvre:

**1 - Remontée**  $P_{\text{fond}} = P_h - \text{Depression}$

**2 - Descente**  $P_{\text{fond}} = P_h + \text{Surpression}$

### **2.2 .5-La dépression et la surpression dépendent:**

- la profondeur
- la taille de l'espace annulaire BHA / trou
- la rhéologie de la boue
- la vitesse de remontée ou descente

La boue doit être alourdie afin d'avoir une marge de sécurité  $h = P_{\text{pore}} + S$  (ou  $S$  est la marge de sécurité).

La pression de la boue est mesurée à l'aide de capteurs sur le manifold de plancher pour obtenir la valeur d'entrée (Stand Pipe Pressure) et sur choke manifold ( $p$ ) pour obtenir la valeur de sortie (CASINGPRESSURE).

Le capteur à l'injection doit être compatible avec la pression maximum de fonctionnement du système de refoulement (400 bars).

Le capteur annulaire doit être aussi compatible avec la série de la tête de puits pour permettre des mesures correctes.

Les jauges utilisées transforment la pression en signal électrique. Les capteurs de pression nous permettent de savoir :

- perte ou bouchage d'une Duse
- sifflage ou rupture  $\dot{\gamma}$  surveillance de déplacement des bouchons de densité différente
- manœuvre hydraulique d'équipements de fond
- contrôle de kick  $\dot{\gamma}$  déroulement des opérations lorsque le puits est mis sous pression (étanchéité d'un packer)

### **II .6 Les Problèmes de forages**

Il est presque certain que les problèmes se produisent pendant le forage d'un puits, même dans des puits très soigneusement planifiées. Par exemple, dans les domaines dans lesquels les pratiques de forage similaires sont utilisées, des problèmes de trous peuvent avoir été

signalés où de tels problèmes existent déjà, car les formations sont non homogène. Par conséquent, deux puits près de l'autre peuvent avoir totalement différentes conditions géologiques.

Lors de la planification ainsi, la clé pour atteindre les objectifs avec succès est de concevoir des programmes de forage sur la base de l'anticipation des problèmes potentiels de trous plutôt que sur la prudence et de confinement. Problèmes de forage peuvent être très coûteux.

Les problèmes de forage les plus courantes :

- Stuck pipe
- Loss of circulation
- Hole deviation
- Pipe failures
- Borehole instability
- Mud contamination
- Formation damage
- Hole cleaning
- H<sub>2</sub>S-bearing zones
- Shallow gas zones
- Equipment and personnel-related problems
- kick

Comprendre et anticiper les problèmes de forage, la compréhension de leurs causes, et les solutions de planification sont nécessaires pour le contrôle du puits coût global et pour atteindre avec succès la zone cible. Ce chapitre traite de ces problèmes, les solutions possibles, et, dans certains cas, des mesures préventives. [7]

### **II .6 .1 - Stuck pipe**

Au cours des opérations de forage, un tuyau est considéré comme bloqué si elle ne peut être libérée du trou sans endommager le tuyau, et sans dépasser le maximum de tours de forage charge au crochet permis. Tuyau collage peut être classé en deux catégories: différentiel et mécaniques. [12]

**II .6 .2 - Mud Contamination**

La boue est dite contaminée, lorsque un matériau étranger pénètre dans le système de boue et provoque des changements indésirables dans les propriétés de la boue, comme la densité, la viscosité et la filtration. En général, les systèmes de boue à base d'eau sont les plus sensibles à la contamination. Boue de contamination peut résulter d'un traitement excessif du système de boue avec des additifs ou de matières entrant dans la boue en cours de forage.

**II .6 3 - Kick (Venue)**

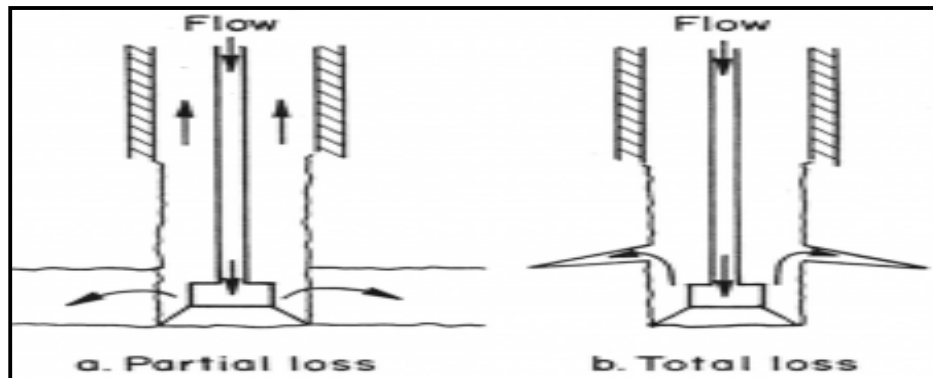
Une venue est l'entrée de fluide de formation dans le puits pendant les opérations de forage. Elle se produit en raison de la pression insuffisante exercée par la boue de forage pour maîtriser la pression exercée par les fluides de la formation forée.

L'essence même du contrôle de puits (Well control) est d'empêcher la venue et si elle arrive, de l'empêcher de se développer en éruption. Une venue incontrôlée résulte habituellement du mauvais déploiement de l'équipement approprié, en utilisant de mauvaises pratiques, ou un manque de formation des équipes de forage. La perte de contrôle du puits peut conduire en éruption, ce qui représente l'une des menaces les plus graves liées à l'exploration des ressources pétrolières impliquant le risque de vie et les conséquences environnementales et économiques

**II .6 .4 - Loss Circulation (Perte)**

Perte de circulation est le flux incontrôlé de boue ensemble dans une formation, parfois appelé une «zone de voleur." Cet article décrit les causes, la prévention, les mesures correctives à perte de circulation.

### 1 -zones Lost-circulation



**Figure I.4 : Perte partielle et Perte totale**

La figure montre les zones partielles et totales perdu-circulation. Dans partielle perte de circulation, de la boue continue de couler à la surface avec une certaine perte de la formation. Totale perte de circulation, cependant, se produit lorsque toutes les coulées de boue dans une formation sans retour à la surface. Si le forage se poursuit durant totale perte de circulation, il est appelé forage aveugle. Cela ne veut pas d'une pratique courante dans le domaine, sauf si toutes les conditions suivantes sont réunies:

La formation au-dessus de la zone de voleur est mécaniquement stable.

Il n'y a pas de production.

Le fluide est de l'eau claire.

Il est économiquement faisable et sûr.

### 2 -Les causes de zones perdu-circulation

Il y a plusieurs situations qui peuvent entraîner la perte de circulation:

- 1- Formations qui sont intrinsèquement fracturées, cavernes, ou ont une perméabilité élevée
- 2- conditions de forage inappropriées
- 3- fractures induites provoquées par les pressions de fond excessifs et la mise carter intermédiaire trop élevées

**3 -fractures induites**

Fractures Induits ou inhérentes peuvent être horizontaux à faible profondeur ou verticale à des profondeurs supérieures à environ 2500 pieds. Pressions de puits de forage excessives sont dues à des débits élevés (haute annulaire friction perte de pression) ou de déclenchement trop rapide (hypertension hausse), ce qui peut conduire à la densité de circulation de la boue de l'équivalent (ECD). Fractures induites peuvent également être causées par:

- 1- nettoyage annulaire de trou incorrect
- 2- poids de la boue excessive
- 3- Arrêt dans un puits de gaz peu profonde à haute pression

**4 -formations caverneux**

Formations caverneux sont des calcaires souvent avec de grandes cavernes. Ce type de perte de circulation est rapide, totale, et le plus difficile à sceller. Formations de haute perméabilité qui sont potentielles zones perdu tirage sont ceux de sable peu profonde avec une perméabilité de plus de 10 Darcies. En général, le sable profond a une faible perméabilité et ne présente pas de problèmes de perte de circulation. Dans les zones de voleur non caverneux, le niveau de boue dans les réservoirs de boue diminue progressivement et, si la poursuite du forage, la perte totale de la circulation peut se produire.

**Conclusion**

Nous avons vu dans ce chapitre les principales notions de base de ce domaine très vaste qui est le forage pétrolier, ce qui nous a permis d'acquérir les concepts et processus importants pour son déroulement. Cependant nous avons constaté que malgré les moyens déboursés dans ce sens les foreurs peuvent se heurter à plusieurs obstacles entre autres les coincements de la garniture.

A vrai dire il y a seulement quelques cas où il n'était pas possible d'éviter le problème, il est important d'avoir une réaction rapide et bien adaptée car l'homme est la clé pour prohiber, mais il n'est pas le seul ; C'est dans ce sens qu'est orientée notre réalisation nous allons tenter de réaliser un système capable de prévenir à l'avance certains de ces cas de pertes, et cela à l'aide des algorithmes de fouilles de données (datamining), nous allons voir dans le second chapitre des descriptions de ce moyen de calcul.

# ***Chapitre II:***

## ***Data-mining***



## **I - Introduction**

L'exploration de données est une nouvelle technologie puissante avec un grand potentiel pour aider les entreprises se concentrent sur les informations les plus importantes dans leurs entrepôts de données.

Dans ce chapitre, nous allons examiner l'exploration de données, que nous allons choisir parmi eux le choix idéal pour les données de forage pétrolier.

## **II - Définition**

Le data-mining (fouille de données) est le processus d'analyse des données provenant de différentes perspectives et résumant en informations utiles - information qui peut être utilisé pour augmenter les revenus, réduit les coûts, ou les deux.

Le data-mining est d'expliquer le passé et prédire l'avenir par le biais de l'analyse de données. L'exploration de données est un domaine multidisciplinaire qui combine les statistiques, l'apprentissage machine, l'intelligence artificielle et de la technologie de base de données. La valeur des applications d'exploration de données est estimée souvent très élevé. De nombreuses entreprises ont stocké de grandes quantités de données sur des années d'exploitation, et l'exploration de données est capable d'extraire des connaissances de grande valeur à partir de ces données. Les entreprises sont alors en mesure de tirer parti de la connaissance extraite en plus de clients, plus de ventes, et de plus grands profits. Ceci est également vrai dans le domaine de l'ingénierie et médicales. [2]

## **III - Le processus de data-mining**

Le data-mining a été populairement considéré comme un synonyme de découverte de connaissances dans les bases de données, bien que certains chercheurs considèrent l'exploration de données comme une étape essentielle de la découverte de connaissances. En général, un processus de découverte de connaissances constituée d'une séquence itérative des étapes suivantes :

**1 - Le nettoyage des données** : Qui gère les données bruitées, erronées, manquantes ou non pertinentes.

**2 - L'intégration de données** : Où de multiples sources de données hétérogènes, peuvent être intégrées dans un seul.

**3 - Sélection de données** : Où les données relatives à la tâche d'analyse sont récupérées à partir de bases de données.

**4 - Transformation de données** : Où les données sont transformées ou regroupées dans des formes appropriées pour l'exploitation minière en effectuant des opérations de synthèse ou agrégation.

**5 - L'exploration de données** : Qui est processus essentiel où les méthodes intelligentes sont appliquées afin d'en extraire les modèles de données.

**6 - L'évaluation de modèle** : Ce qui est d'identifier les motifs vraiment intéressants représentant connaissances basées sur des mesures de qualité.

**7 - Présentation de la connaissance** : Où les techniques de visualisation et de représentation des connaissances sont utilisées pour présenter les connaissances extraites à l'utilisateur.

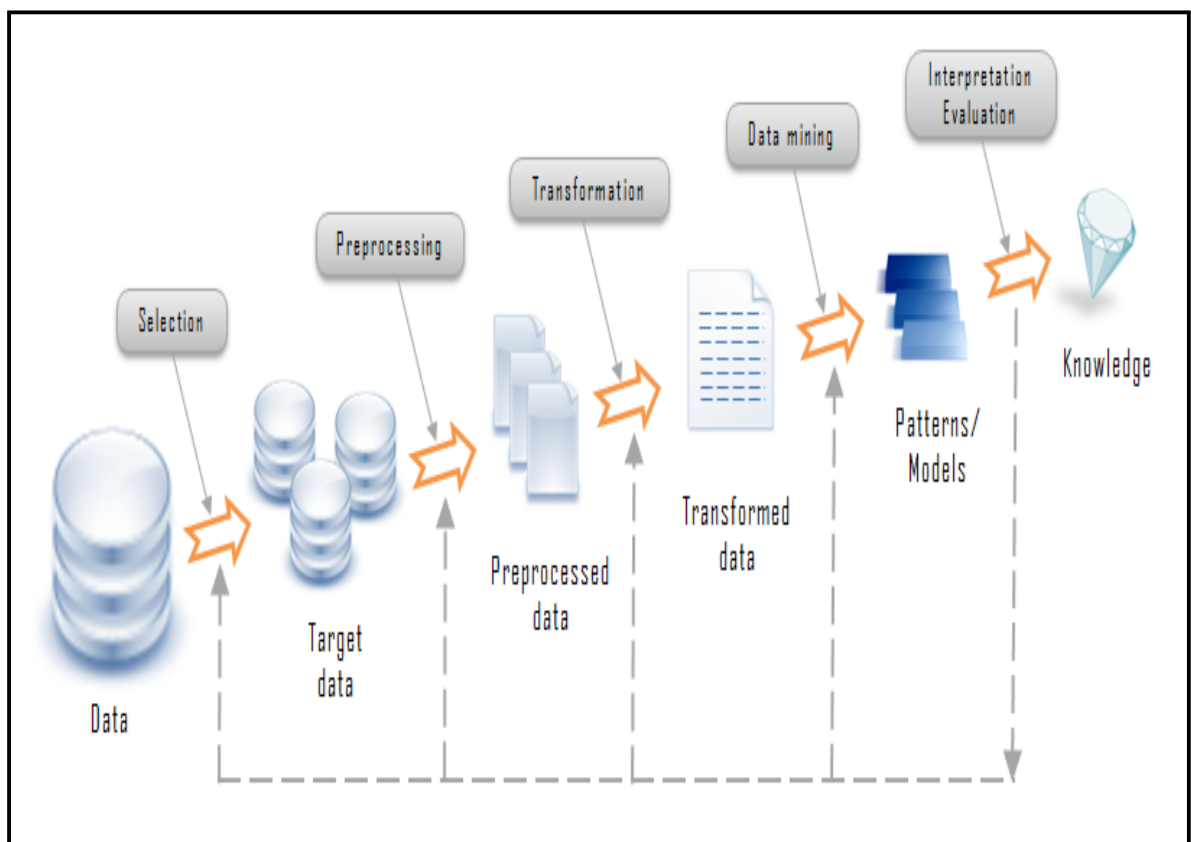


Figure II.1: Processus de data-mining

## **IV - Les taches de data-mining**

### **1 -La classification**

La classification est la dérivation d'une fonction ou d'un modèle qui détermine la classe d'un objet sur la base de ses attributs. Un ensemble d'objets est donnée à l'ensemble d'apprentissage dans lequel chaque objet est représenté par un vecteur d'attributs avec sa classe. Une fonction de classification ou modèle est construit en analysant la relation entre les attributs et les classes des objets dans la formation pour établir une telle fonction de classification ou d'un modèle peuvent être des objets et de développer une meilleure compréhension des classes utilisées pour classer les futures des objets dans le base de données.

Par exemple, à partir d'un ensemble de patients diagnostiqués, qui servent l'ensemble de la formation, un modèle de classification peut être construit, ce qui met un terme à la maladie d'un patient de son / ses données de diagnostic. Le modèle de classification peut être utilisée pour diagnostiquer la maladie d'un nouveau patient sur la base des données de diagnostic du patient comme l'âge, le sexe, le poids, la température, la tension artérielle...etc.  
[2]

### **2 - La régression**

La régression est une fonction d'exploration de données qui prédit un certain nombre. Âge, le poids, la distance, la température, le revenu ou les ventes pourraient tous être prédits en utilisant des techniques de régression. Par exemple, un modèle de régression peut être utilisé pour prédire la taille des enfants, étant donné leur âge, le poids et d'autres facteurs.

Une tâche de régression commence avec un ensemble de données dans laquelle les valeurs cibles sont connues. Par exemple, un modèle de régression qui prédit la taille de l'enfant peut être élaboré à partir des données observées pour de nombreux enfants, sur une période de temps. Les données pourraient suivre âge, la taille, le poids, les étapes du développement, de l'histoire de la famille, et ainsi de suite. Hauteur serait la cible, les autres attributs seraient les prédicteurs, et les données pour chaque enfant constitueraient un cas.

Dans le processus d'apprentissage (formation), l'algorithme de régression estime la valeur de la cible en fonction des prédicteurs pour chaque cas dans les données de

construction. Ces relations entre les prédicteurs et cible sont résumés dans un modèle, qui peut ensuite être appliqué à un ensemble de données différent dans lequel les valeurs cibles sont inconnus. [2]

### **3 - La prédiction**

La prédiction est semblable à la classification et la régression, sauf que pour la prévision, les résultats se situent dans l'avenir. Exemples de tâches de prévision appliquée au marketing : « Prédire le prix d'un stock de trois mois dans le futur » [2]

### **4 - Le groupement par similitude**

La recherche de règles d'association est la tâche la plus intéressante du data-mining. C'est également celle qui est la plus répandue dans le monde des affaires, notamment en marketing pour l'analyse du panier de consommation. La recherche de règles d'association cherche à découvrir les règles de quantification ou de relation entre deux ou plusieurs attributs. Les règles d'association sont de la forme «Si antécédent, puis conséquente », avec une mesure confiance associée à la règle. La recherche de règles d'associations dans une grande base de données permet de découvrir des règles cachées utiles pour la prise de décision.

Exemple de règle célèbre : lorsqu'un homme achète des couches pour bébés, il achète 2 packs de bières dans 65% des cas. Il serait alors intéressant pour le gestionnaire d'adapter ses promotions à ces nouvelles règles. [2]

### **5 - L'analyse des clusters**

Le clustering est l'identification des classes, également appelés clusters ou groupes, pour un ensemble d'objets dont les classes ne sont pas connues. Les objets sont donc regroupés que les similitudes intra classe sont maximisés et les similitudes entre classes sont réduites au minimum sur la base de certains critères définis sur les attributs des objets. Une fois que les grappes sont décidées, les objets sont étiquetés avec leurs groupes correspondants, et les caractéristiques communes des objets dans un cluster sont résumées pour former la description de classe. Par exemple, une banque peut regrouper ses clients en plusieurs groupes en fonction des similitudes de leurs revenus de l'âge, de résidence, etc., et les caractéristiques communes des clients dans un groupe peuvent être utilisés pour décrire ce groupe de clients.

Les grappes aideront la banque à comprendre ses costumiers mieux et donc de fournir des produits et des services personnalisés plus appropriés. [2]

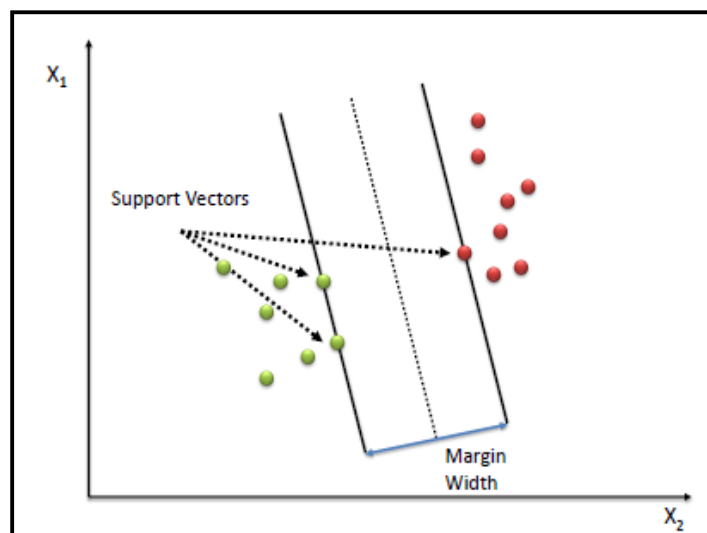
## V - Les algorithmes de data-mining

### 1- SVM (Support Vector Machine)

#### 1-1 SVM – Classification

Les machines à vecteur support se situent sur l'axe de développement de la recherche humaine des techniques d'apprentissage. Les SVMs sont une classe de techniques d'apprentissage introduite par Vladimir Vapnik au début des années 90, elles reposent sur une théorie mathématique solide à l'inverse des méthodes de réseaux de neurones. Elles ont été développées au sens inverse du développement des réseaux de neurones : ces derniers ont suivi un chemin heuristique de l'application et l'expérimentation vers la théorie ; alors que les SVMs sont venues de la théorie du son vers l'application.

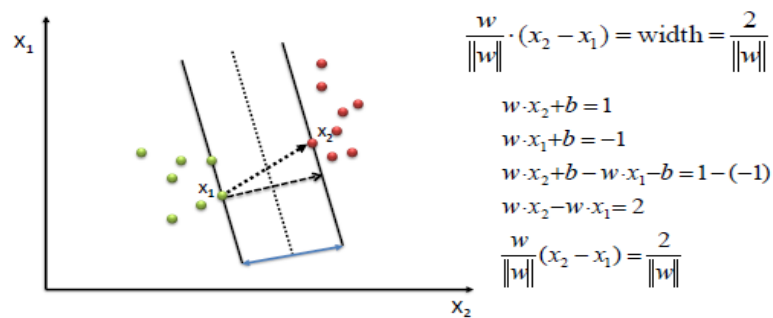
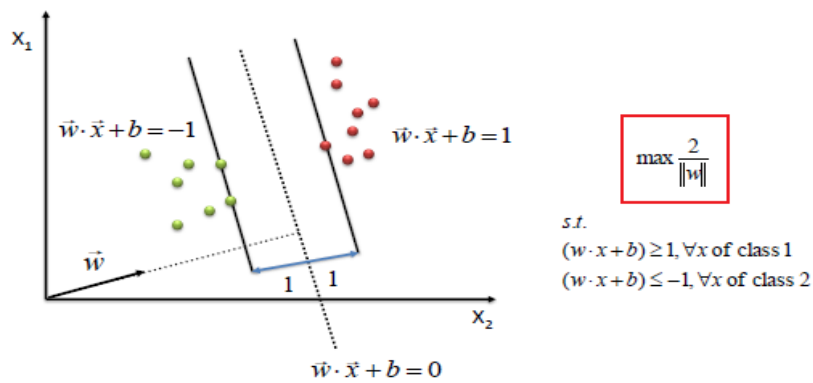
SVM effectue la classification en trouvant l'hyperplan qui maximise la marge entre les deux classes. Les vecteurs qui définissent l'hyperplan sont les vecteurs de support.



**1-2 Algorithme :**

- 1- Définir un hyperplan optimal : maximiser la marge
- 2- Étendre la définition ci-dessus pour des problèmes non linéairement séparables : un terme de pénalité pour les erreurs de classification.
- 3- Les données cartographiques à l'espace de grande dimension où il est plus facile de classer avec des surfaces de décision linéaires : formuler problème afin que les données sont mappé implicitement à cet espace.

Pour définir un hyperplan optimal, nous devons maximiser la largeur de la marge (w).

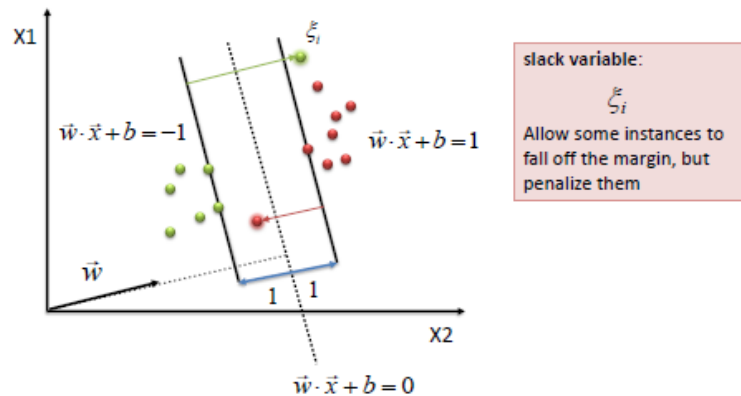


Nous trouvons w et b en résolvant la fonction objective suivante en utilisant la programmation quadratique.

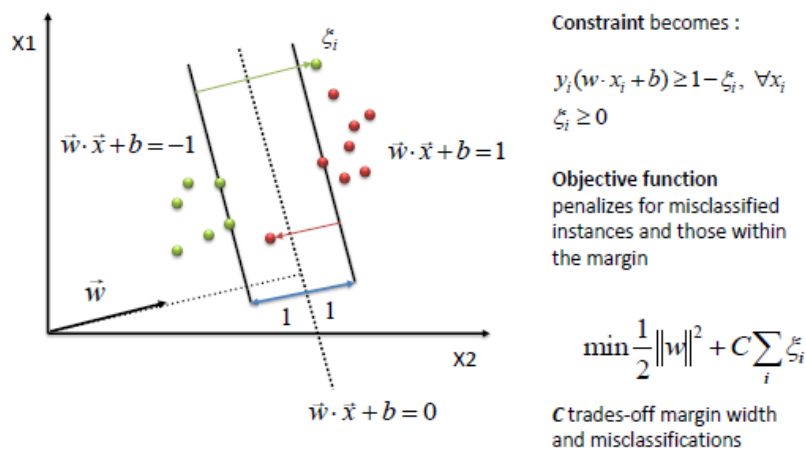
$$\min \frac{1}{2} \|w\|^2$$

$$s.t. y_i (w \cdot x_i + b) \geq 1, \forall x_i$$

La beauté de SVM est que si les données sont linéairement séparables, il y a une valeur minimale globale unique. Une analyse de SVM idéal devrait produire un hyperplan qui sépare complètement les vecteurs (cas) en deux classes ne se chevauchent pas. Cependant, la séparation parfaite peut ne pas être possible, ou il peut en résulter un modèle avec autant de cas que le modèle ne classe pas correctement. Dans cette situation SVM trouve l'hyperplan qui maximise la marge et minimise les erreurs de classification.

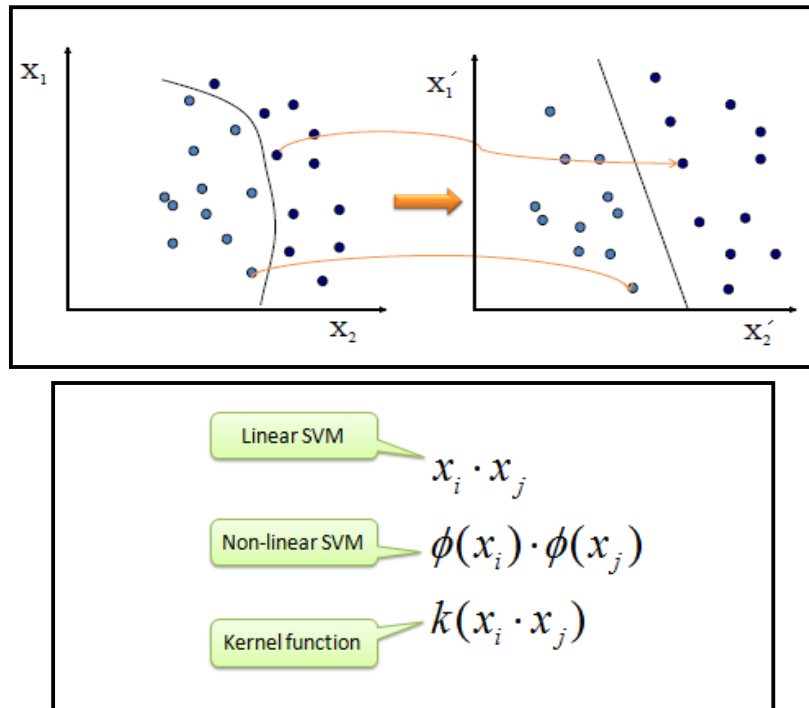


L'algorithme essaie de maintenir la variable d'écart à zéro tout en maximisant la marge. Cependant, il ne minimise pas le nombre d'erreurs de classification (problème NP-complet), mais la somme des distances des hyperplans de marge.



La façon la plus simple de séparer deux groupes de données est par une ligne droite (1 dimension), surface plane (2 dimensions) ou un hyperplan de dimension n. Cependant, il y a des situations où une région non linéaire peut séparer les groupes de manière plus efficace. SVM gère cela en utilisant une fonction du noyau (non linéaire) pour cartographier les

données dans un espace différent où un hyperplan (linéaire) ne peut pas être utilisé pour faire la séparation. Cela signifie une fonction non-linéaire est appris par une machine d'apprentissage linéaire dans un espace de grande dimension caractéristique alors que la capacité du système est contrôlé par un paramètre qui ne dépend pas de la dimension de l'espace. Ceci est appelé noyau truc qui signifie que la fonction du noyau transformer les données dans un espace de caractéristiques de plus grande dimension pour permettre d'effectuer la séparation linéaire.

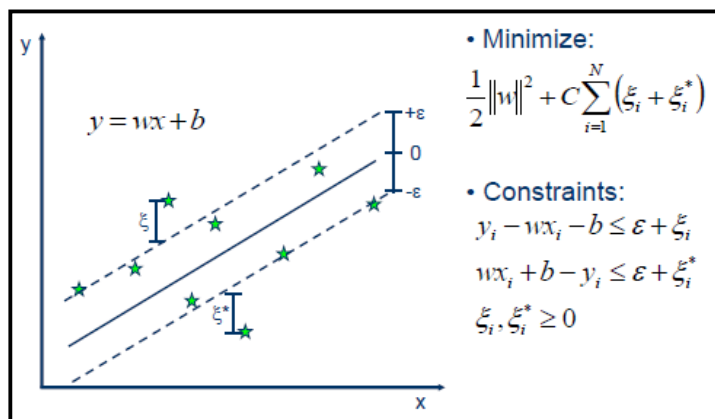
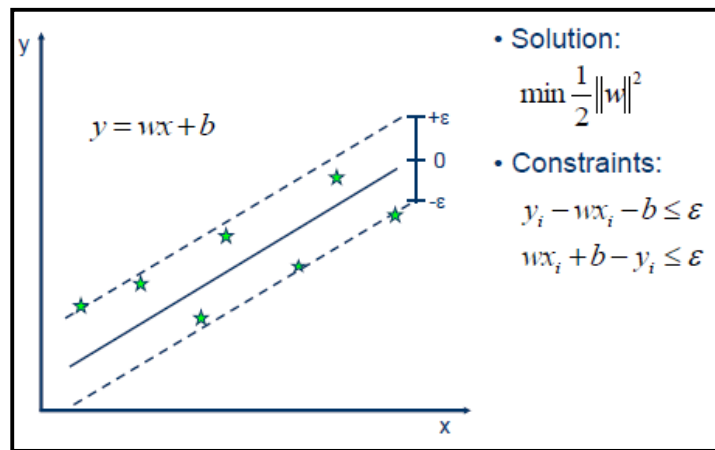


Les données cartographiques dans un nouvel espace, puis prendre le produit scalaire des nouveaux vecteurs. L'image du produit scalaire des données est le produit scalaire des images des données. Deux fonctions du noyau sont présentés ci-dessous.

Polynomial	$k(x_i, x_j) = (x_i \cdot x_j)^d$
Gaussian Radial Basis function	$k(x_i, x_j) = \exp\left(-\frac{\ x_i - x_j\ ^2}{2\sigma^2}\right)$

1-2 SVM – Régression (SVR)

Support Vector Machine peut également être utilisé comme une méthode de régression, en conservant toutes les principales fonctions qui caractérisent l’algorithme (marge maximale). Le Support Vector Régression (SVR) utilise les mêmes principes que le SVM pour la classification, avec seulement quelques différences mineures. Tout d’abord, parce que la production est un nombre réel, il devient très difficile de prévoir les informations à portée de main, ce qui a des possibilités infinies. Dans le cas de la régression, une marge de tolérance (epsilon) est située dans approximation de la SVM qui aurait déjà demandé du problème. Mais à côté de ce fait, il y a aussi une raison plus compliqué, l’algorithme est plus complexe donc être pris en considération. Cependant, l’idée principale est toujours la même : pour minimiser l’erreur, l’individualisation de l’hyperplan qui maximise la marge, en gardant à l’esprit qu’une partie de l’erreur est tolérée.



1-2.1- SVR linéaire

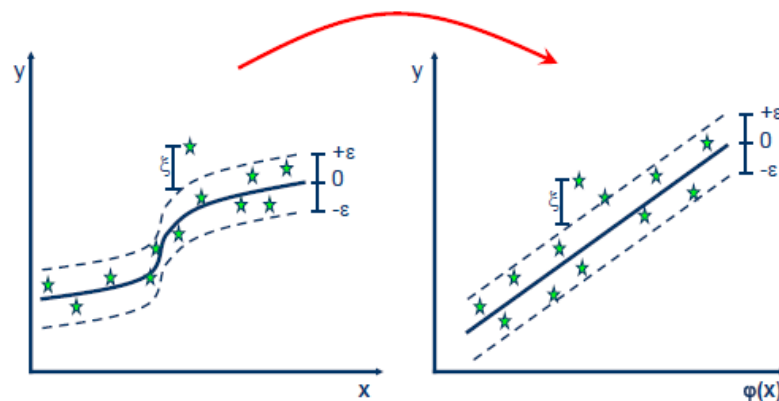
$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot \langle x_i, x \rangle + b$$

1-2.2- SVR non linéaire

Les fonctions du noyau de transformer les données dans un espace de caractéristiques de plus grande dimension pour permettre d'effectuer la séparation linéaire.

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot \langle \varphi(x_i), \varphi(x) \rangle + b$$

$$y = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b$$



1-2.3- les fonctions kernel

Polynomial

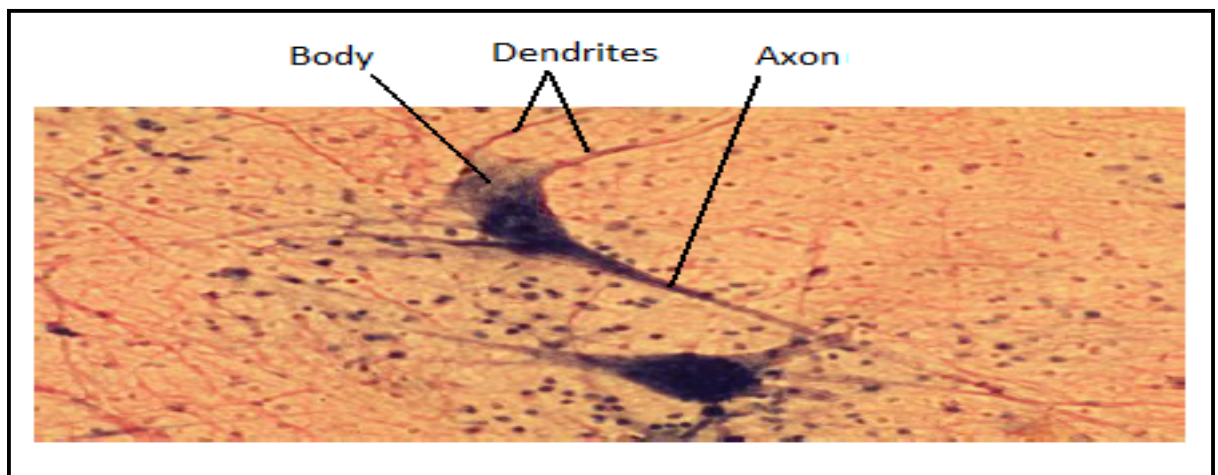
$$k(x_i, x_j) = (x_i \cdot x_j)^d$$

Gaussian Radial Basis function

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

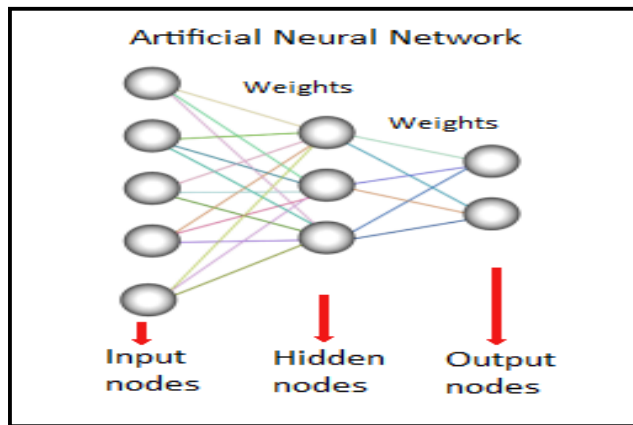
### 3- Le Réseau de neurones artificiel

Un réseau de neurones artificiel (RNA) est un système qui est basé sur le réseau neuronal biologique, tel que le cerveau. Le cerveau a environ 100 milliards de neurones, qui communiquent par des signaux électrochimiques. Les neurones sont reliés par des jonctions appelées synapses. Chaque neurone reçoit des milliers de connexions avec d'autres neurones, recevant en permanence les signaux entrants pour atteindre le corps de la cellule. Si la somme résultante des signaux dépasse un certain seuil, une réponse est envoyée par l'intermédiaire de l'axone. Le RNA tente de recréer le miroir de calcul du réseau neuronal biologique, même si elles ne sont pas comparables car le nombre et la complexité des neurones et utilisé dans un réseau neuronal biologique est beaucoup plus que ceux dans un réseau neutre artificiel.

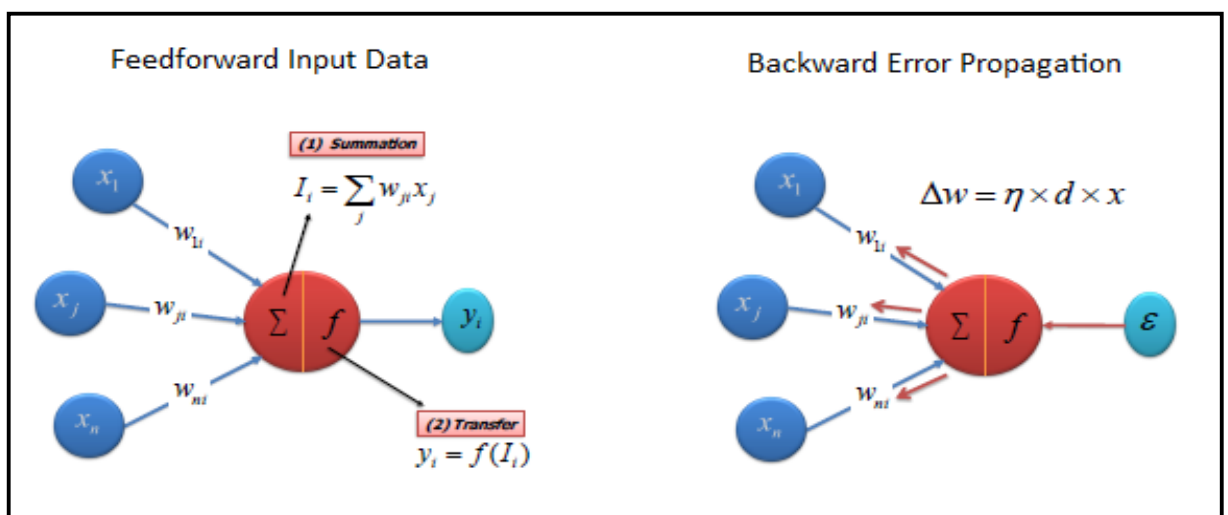


**Figure II.2 : Le neurone biologique**

Un RNA est constitué d'un réseau de neurones artificiels (également appelé «nœuds»). Ces nœuds sont reliés les uns aux autres, et la force de leurs connexions à l'autre est affecté une valeur sur la base de leur force d'inhibition (maximum étant de -1,0) ou d'excitation (maximum étant 1,0). Si la valeur de la connexion est élevée, cela indique qu'il y a un lien très fort. Au sein de la conception de chaque nœud, une fonction de transfert est intégrée. Il existe trois types de neurones dans un RNA, nœuds d'entrée (input nodes), nœuds cachés (hidden nodes) et des nœuds de sortie (output nodes).



Les nœuds d'entrée prennent en informations, sous la forme qui peut être numériquement exprimé. L'information est présentée sous forme de valeurs d'activation, où chaque nœud est donné un certain nombre, plus le nombre est élevé, plus l'activation. Cette information est ensuite transmise à travers le réseau. Sur la base des forces de connexion (poids), l'inhibition ou d'excitation, et des fonctions de transfert, la valeur d'activation est passé de nœud à nœud. Chacun des nœuds résume les valeurs d'activation qu'il reçoit ; il modifie alors la valeur en fonction de sa fonction de transfert. L'activation circule à travers le réseau, à travers les couches cachées, jusqu'à ce qu'il atteigne les nœuds de sortie. Les nœuds de sortie reflètent alors l'entrée d'une manière significative sur le monde extérieur.

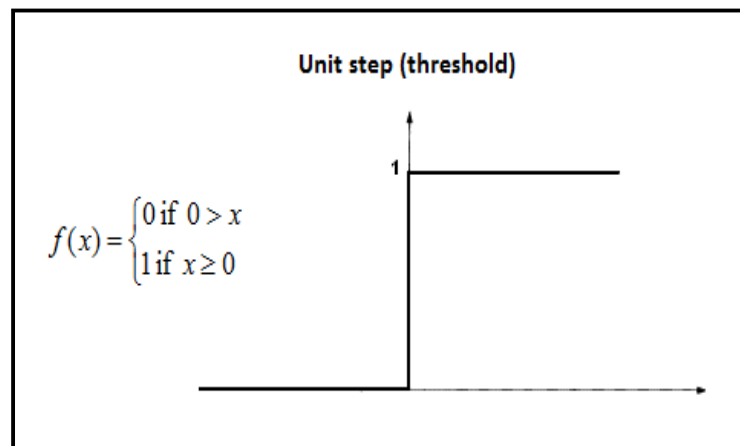


2-1 - Les Fonctions de Transfert (activation)

La fonction de transfert convertit les signaux d'entrée en signaux de sortie. Quatre types de fonctions de transfert sont couramment utilisés, Unité étape (le seuil), sigmoïde, linéaire par morceaux, et gaussien.

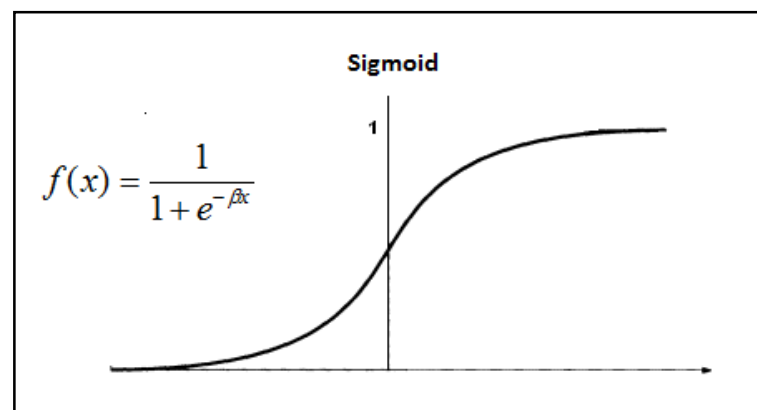
### 2-1.1 -Unité étape (le seuil)

La sortie est fixée à une de deux niveaux, selon que la puissance totale est supérieure ou inférieure à une certaine valeur de seuil.



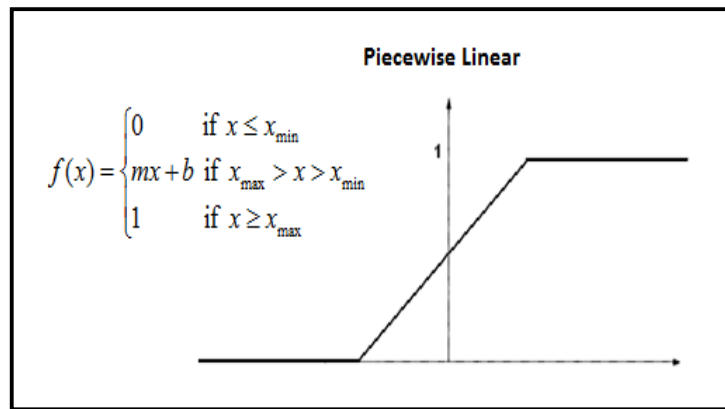
### 2-1.2 -Sigmoïde :

La fonction sigmoïde se compose de 2 fonctions, logistique et tangentielle. Les valeurs de la logistique de toute les fonctions de 0 et 1 et -1 à +1 pour fonction tangente.



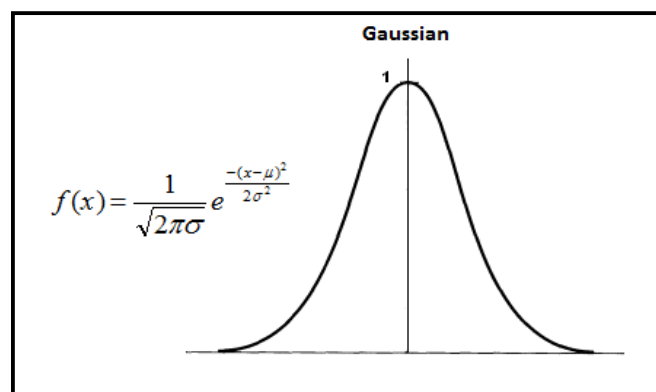
### 2-1.3 -Linéaire par morceaux :

La sortie est proportionnelle à la puissance totale pondérée.



### 2-1.4 - Gaussien :

Gaussiennes sont des courbes en forme de cloche qui sont continues. La sortie de nœud (haut / bas) est interprété en termes d'appartenance à une classe (1/0), en fonction de la proximité de l'entrée nette est à une valeur choisie de la moyenne.



### 2-2 -Algorithme :

Il existe différents types de réseaux de neurones, mais ils sont généralement classés dans les réseaux de feed-forward et feed-back.

Un réseau de feed-forward est un réseau non récurrent qui contient les entrées, les sorties et les calques masqués ; les signaux ne peuvent voyager dans une direction. Les données d'entrée est passé sur une couche d'éléments de traitement où il effectue des calculs. Chaque élément de traitement effectue son calcul sur la base d'une somme pondérée de ses entrées. Les nouvelles valeurs calculées deviennent alors les nouvelles valeurs d'entrée qui se nourrissent de la couche suivante. Ce processus se poursuit jusqu'à ce qu'il ait passé par toutes les couches et détermine la sortie. Une fonction de transfert de seuil est parfois utilisée pour quantifier la

sortie d'un neurone dans la couche de sortie. Réseaux de feed-forward comprennent Perceptron (linéaire et non-linéaire) et les réseaux de fonctions de base radiale. Réseaux de feed-forward sont souvent utilisés dans l'extraction de données.

Un réseau feed-back a des pistes de feed-back qui signifie qu'ils peuvent avoir des signaux voyageant dans les deux directions à l'aide de boucles. Toutes les connexions possibles entre les neurones sont autorisées. Depuis boucles sont présents dans ce type de réseau, il devient un système dynamique non linéaire qui change continuellement jusqu'à ce qu'il atteigne un état d'équilibre. Réseaux de feed-back sont souvent utilisés dans les mémoires associatives et des problèmes d'optimisation où le réseau recherche le meilleur arrangement de facteurs interdépendants.

### 3 - K-Means :

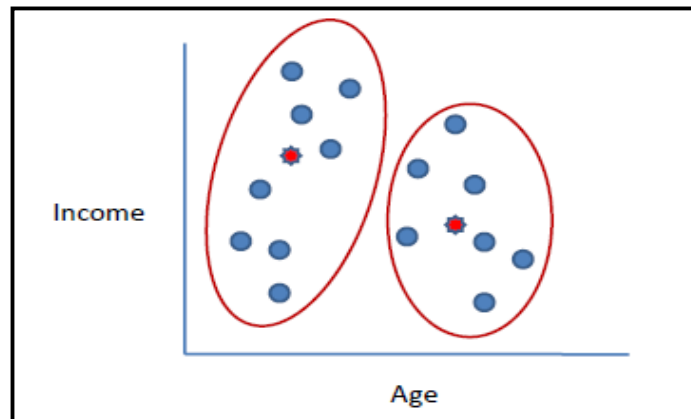
Le regroupement K-Means a l'intention de diviser n objets en k grappes dans lequel chaque objet appartient à la grappe avec la moyenne la plus proche. Cette méthode produit exactement k différents groupes de plus grande distinction possible. Le meilleur nombre de clusters k menant à la plus grande séparation (distance) ne sont pas connus en tant que priori et doit être calculée à partir des données. L'objectif du regroupement K-Means est de minimiser la variance totale intra-cluster, ou, la fonction d'erreur au carré :

The diagram shows the objective function formula for K-Means clustering:  $J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$ . Annotations include: 'number of clusters' pointing to 'k', 'number of cases' pointing to 'n', 'case i' pointing to 'x\_i^{(j)}', 'centroid for cluster j' pointing to 'c\_j', and 'Distance function' pointing to the squared norm term. An arrow labeled 'objective function' points to the entire equation.

#### 3-1 - Algorithme :

- 1 - Clusters les données en K groupes où k est prédéfinie.
- 2 - sélectionner des points de k au hasard en tant que centres de cluster.
- 3 - Affecter des objets à leur centre de pôle le plus proche conformément à la fonction de distance euclidienne.
- 4 - Calculer le centre de gravité vs moyenne de tous les objets dans chaque grappe.

5 - Répéter les étapes 2, 3 et 4 jusqu'à ce que les mêmes points sont attribués à chaque groupe en tours consécutifs.



K-Means est relativement une méthode efficace. Cependant, nous avons besoin de spécifier le nombre de grappes, à l'avance et les résultats finaux sont sensibles à l'initialisation et se termine souvent à un optimum local. Malheureusement, il n'y a pas de méthode théorique globale pour trouver le nombre optimal de clusters. Une approche pratique consiste à comparer les résultats de plusieurs essais avec  $k$  différent et choisir le meilleur sur la base d'un critère prédéfini. D'une manière générale, un grand  $k$  diminue probablement l'erreur, mais augmente le risque de sur ajustement.

#### 4 - K Nearest Neighbors

K Nearest Neighbors est un algorithme simple qui stocke tous les cas disponibles et classe nouveaux cas basé sur une mesure de similarité (par exemple, des fonctions de distance). KNN a été utilisé dans la reconnaissance et le motif d'estimation statistique déjà au début de 1970, comme une technique non paramétrique.

##### 4-1 - Algorithme :

Un cas est classé par un vote à la majorité de ses voisins, avec le cas étant affecté à la classe la plus commune parmi ses plus proches voisins  $K$  mesurées par une fonction de distance. Si  $K = 1$ , alors l'affaire est tout simplement affectée à la classe de son voisin le plus proche.

**Distance functions**

<b>Euclidean</b>	$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$
<b>Manhattan</b>	$\sum_{i=1}^k  x_i - y_i $
<b>Minkowski</b>	$\left( \sum_{i=1}^k ( x_i - y_i ^q) \right)^{1/q}$

Il convient également de noter que les trois mesures de distance ne sont valables que pour les variables continues. Dans le cas de variables catégoriques la distance de Hamming doit être utilisée. Il soulève aussi la question de la normalisation des variables numériques entre 0 et 1 quand il y a un mélange de variables numériques et catégorielles dans l'ensemble de données.

**Hamming Distance**

$$D_H = \sum_{i=1}^k |x_i - y_i|$$

$x = y \Rightarrow D = 0$

$x \neq y \Rightarrow D = 1$

X	Y	Distance
Male	Male	0
Male	Female	1

Le choix de la valeur optimale de K est préférable de faire d'abord inspecter les données. En général, une grande valeur de K est plus précise car il réduit le bruit dans l'ensemble mais il n'y a aucune garantie. La validation croisée est une autre façon de déterminer rétrospectivement une bonne valeur de K en utilisant un ensemble de données indépendante pour valider la valeur K. Historiquement, le K optimal pour la plupart des ensembles de données a été entre 3-10. Qui produit de bien meilleurs résultats que KNN.

## 5 - Naïve Bayésien

Le classificateur bayésien naïf est basé sur le théorème de Bayes avec des hypothèses d'indépendance entre les prédicteurs. Un modèle bayésien naïf est facile à construire, sans estimation des paramètres itératifs complexes qui le rend particulièrement utile pour les très grands ensembles de données. Malgré sa simplicité, le classificateur bayésien naïf fait souvent étonnamment bien et est largement utilisé car il surpasse souvent des méthodes de classification plus sophistiquées.

### 5-1 - Algorithme

Théorème de Bayes fournit un moyen de calcul de la probabilité a posteriori,  $P(c | x)$ , de  $P(c)$ ,  $P(x)$  et  $P(x | c)$ . Classification naïve bayésienne suppose que l'effet de la valeur d'un prédicteur ( $x$ ) sur une classe donnée ( $c$ ) est indépendant de la valeur des autres prédicteurs. Cette hypothèse est appelée classe indépendance conditionnelle.

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

↓
↓

Posterior Probability
Predictor Prior Probability

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

- $P(c | x)$  est la probabilité a posteriori de la classe (cible) prédicteur donnée (attribut).
- $P(c)$  est la probabilité a priori de la classe.
- $P(x | C)$  est la probabilité qui est la probabilité de classe donnée prédicteur.
- $P(x)$  est la probabilité a priori de prédiction.

## 6 - Arbre de décision

Arbre de décision construit des modèles de classification ou de régression sous la forme d'une structure arborescente. Il se décompose un ensemble de données en sous-ensembles plus en plus petits, Le résultat final est un arbre avec des nœuds de décision et des nœuds terminaux. Un nœud de décision comprend deux ou plusieurs branches. Nœud feuille

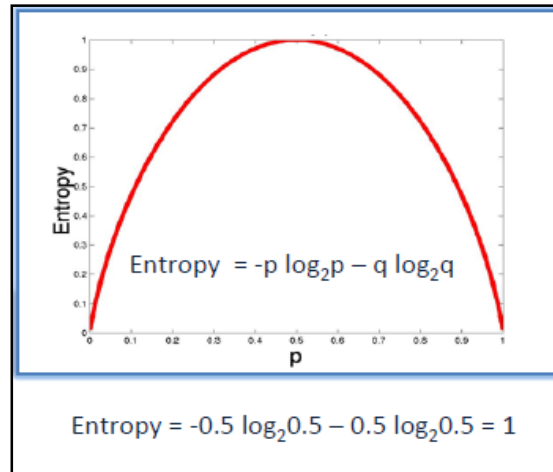
représente une classification ou de la décision. Le nœud le plus haut de la prise dans un arbre qui correspond à la meilleure prédiction appelé nœud racine. Les arbres de décision peuvent traiter des données à la fois qualitatives et numériques

### 6-1 - Algorithme

L'algorithme de base pour la construction d'arbres de décision appelés ID3 par J. R. Quinlan qui emploie un top-down, recherche gloutonne dans l'espace des branches possibles sans retour en arrière. ID3 utilise Entropie et information Gain de construire un arbre de décision.

### 6-2 - Entropie

Un arbre de décision est construit de haut en bas à partir d'un nœud racine et consiste à diviser les données en sous-ensembles qui contiennent des instances avec des valeurs similaires (homogènes). ID3 algorithme utilise l'entropie pour calculer l'homogénéité d'un échantillon. Si l'échantillon est complètement homogène l'entropie est nulle et si l'échantillon est tout aussi divisée n'a entropie d'un.



Pour construire un arbre de décision, il faut calculer deux types d'entropie en utilisant des tableaux de fréquence comme suit :

a) L'entropie en utilisant la table de fréquence d'un attribut :

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

Play Golf	
Yes	No
9	5

Entropy(PlayGolf) = Entropy(5,9)  
 = Entropy(0.36, 0.64)  
 = -(0.36 log<sub>2</sub> 0.36) - (0.64 log<sub>2</sub> 0.64)  
 = 0.94

b) L'entropie en utilisant la table de fréquence de deux attributs :

$$E(T, X) = \sum_{c \in X} P(c)E(c)$$

		Play Golf		
		Yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(PlayGolf, Outlook) = P(Sunny)\*E(3,2) + P(Overcast)\*E(4,0) + P(Rainy)\*E(2,3)  
 = (5/14)\*0.971 + (4/14)\*0.0 + (5/14)\*0.971  
 = 0.693

### 6-3 - Gain informations

Le gain d'information est basé sur la diminution de l'entropie après un ensemble de données est divisé sur un attribut. La construction d'un arbre de décision est de trouver attribut qui renvoie le gain d'information le plus élevé (à savoir, les branches les plus homogènes).

❖ **Etape 1 :**

Calculer l'entropie de la cible.

Entropy(PlayGolf) = Entropy(5,9)  
 = Entropy(0.36, 0.64)  
 = -(0.36 log<sub>2</sub> 0.36) - (0.64 log<sub>2</sub> 0.64)  
 = 0.94

❖ Étape 2 :

L'ensemble de données est ensuite divisé sur les différents attributs. L'entropie pour chaque branche est calculée. Ensuite, il est ajouté proportionnellement, pour obtenir l'entropie totale de la scission. L'entropie résultante est soustraite de l'entropie avant la scission. Le résultat est le gain de l'information, ou la diminution de l'entropie.

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
		Gain = 0.247	

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
		Gain = 0.029	

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1
		Gain = 0.152	

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3
		Gain = 0.048	

$$Gain(T, X) = Entropy(T) - Entropy(T, X)$$

$$G(\text{PlayGolf}, \text{Outlook}) = E(\text{PlayGolf}) - E(\text{PlayGolf}, \text{Outlook})$$

$$= 0.940 - 0.693 = 0.247$$

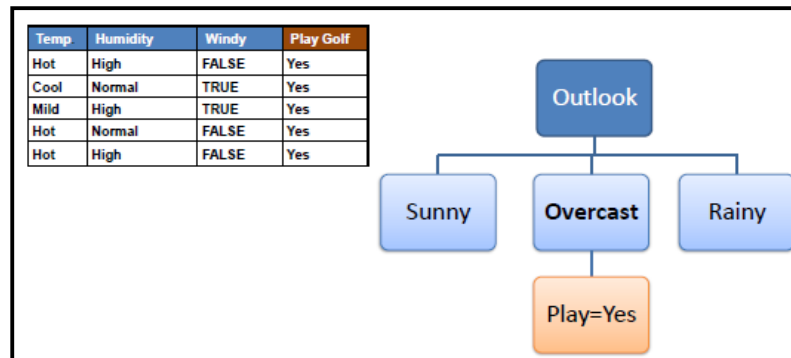
❖ Étape 3 :

Choisissez un attribut avec la plus forte hausse de l'information que le nœud de décision.

★		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
		Gain = 0.247	

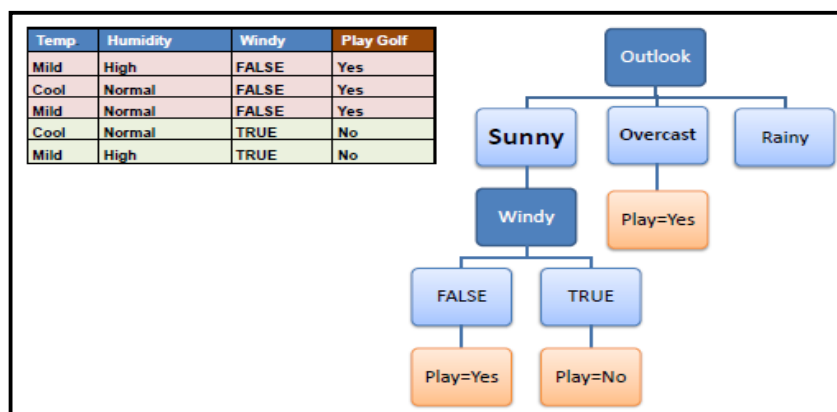
❖ **Étape 4 :**

a) Une branche d'entropie de 0 est un nœud feuille.



❖ **Étape 5 :**

b) une branche avec entropie plus de 0 doit encore fractionnement.

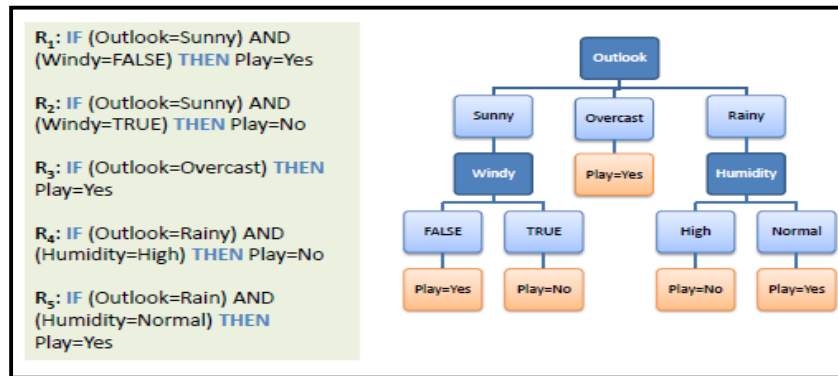


❖ **Étape 6 :**

L'algorithme ID3 est géré de manière récursive sur les branches non-feuilles, jusqu'à ce que toutes les données sont classées.

**6-4 - Arbre de décision et règles de décision**

Un arbre de décision peut être facilement transformé en un ensemble de règles de mappage à partir du nœud racine aux nœuds feuilles une par une.



### 6-5 - Les Avantages

- Applicables a des variables quantitatives ou qualitatives.
- Intelligibilité de la procédure de classification (avec une traduction possible sous forme de règles).
- Rapidité du processus de classification (après construction de l'arbre).
- Invariance a toute transformation monotone des données.
- Sélectionne automatiquement les variables les plus importantes.

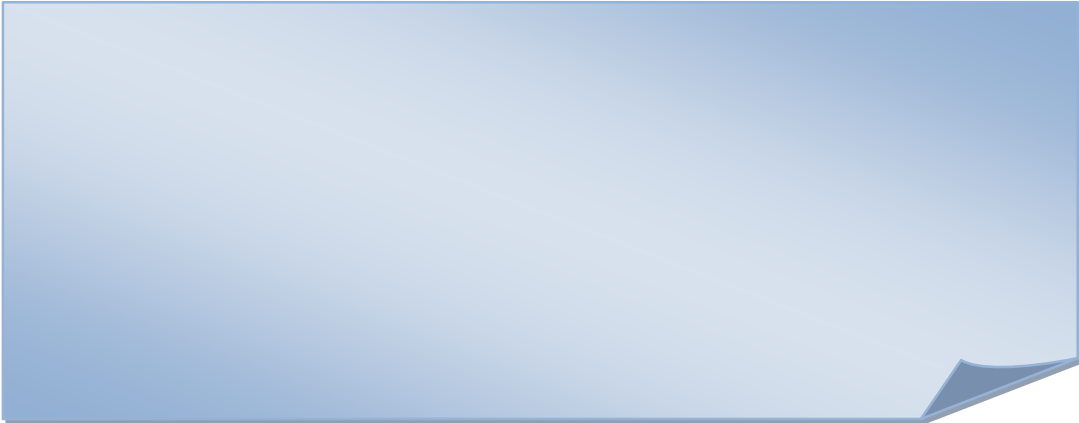
### 6-6 - Les Inconvénients

- Performance sur de grands volumes de données ?
- Choix des variables : arbres multi variés potentiellement plus performant, mais au détriment de l'intelligibilité des solutions proposées.

## VI - Conclusion

Dans ce chapitre, nous avons obtenu le processus, les taches et les différent techniques de data-mining Dans le prochain chapitre, nous allons expliquer la conception de notre système et son fonctionnement, ensuite nous choisissons entre les différentes techniques de datamining.

***Chapitre III:***  
***Conception***



## **I - Introduction**

Après avoir présenté l'aspect théorique de notre projet, et pour comprendre son fonctionnement, on a besoin d'expliquer la conception à l'aide des outils et des méthodes qui permettent la réalisation de ce genre du système.

Dans ce chapitre nous allons présenter l'architecture globale du système en détaillant son fonctionnement pour simplifier la compréhension de notre travail. Ensuite, on va choisir entre les différentes méthodes de travail, nous terminons ce chapitre par une conclusion.

## **II - Architecture Globale de système**

Notre système est basé sur les techniques des fouilles des données (Datamining), en prenant le composant de formalisation de données qui est de nature des données à travers des fichiers WITSMLs.

WITSML est une norme de transmission de données techniques entre les organisations dans l'industrie du pétrole. Il continue d'être mis au point par un groupe d'intérêt spécial énergétiques facilité d'élaborer des normes XML pour le forage, l'achèvement, et l'échange de données sur les interventions.

Le système prend comme entrée un ensemble des données instantané, qui contiennent les informations sur notre étude de cas, qui est les pertes de boue de forage pour prédire la perte de la boue avant qu'elle ne survienne.

A partir de ces données reçues, et en les interprétant les OSEs (Operations Support Engineer) -décideur- peuvent détecter grâce à leur expérience et observation certains signaux indiquant la possibilité d'avoir un problème notamment la perte, ce qui n'est pas tâche aisée vu le nombre et la complexité des données reçues, en dépit de la présentation explicite de ces données en forme de graphe et courbe et valeur de paramètre, ce travail nécessite une bonne analyse et donne rarement le résultat escompté.

### **1 - Mud Logging**

Le mud logging implique le suivi rig site et à l'évaluation de l'information qui vient à la surface pendant le forage, à l'exclusion des données provenant de capteurs de fond. La diagraphie de boue terme est considéré, par certains, d'être dépassée et pas suffisamment

descriptif. En raison de la relativement large gamme de services fournis par les géologues, les ingénieurs et les techniciens traditionnellement appelés mud loggers, le terme «surface logging» est parfois utilisé, et le personnel assurant les services peut être appelé surface-logging specialists. [16]

## **2 - Le projet RTOM :**

Le projet « Real Time Operating Management » est un projet de la société SONATRACH, en collaboration avec le segment SIS de la société SCHLUMBERGER, débuté en 2010, le projet se charge du suivi en temps réel des opérations de forage grâce à une installation réseau qui transmet les données à partir de tous les puits connectés (42 et prochainement 84) au centre de contrôle des opérations ROC «Remote Operation Center », situé à la base de SONATRACH 24 Février à HASSI MESSAOUD wilaya d'OURAGLA .

L'équipe présente dans le projet est divisée en deux groupes « Front office » et « Back office »

**Front office** : ce groupe se charge de tout ce qui est technique : la liaison entre les sites de Forage et le ROC, l'émission et la réception des données, ainsi que la résolution de tous les problèmes relatifs aux matériels et logiciels. Leur mission, assurée est que les données soient toujours disponibles.

**Back office** : ce groupe se charge du monitoring des opérations de forage et l'élaboration des rapports de suivi, leur mission est de détecter d'éventuels problèmes afin de les signaler aux superviseurs sur le RIG dans le but de les éviter. [3]

## **3 - Base de données**

Le rôle d'une base de données est de stocker les informations disponibles sous forme de fichiers WITSML.

## **III - Fonctionnement globale de système**

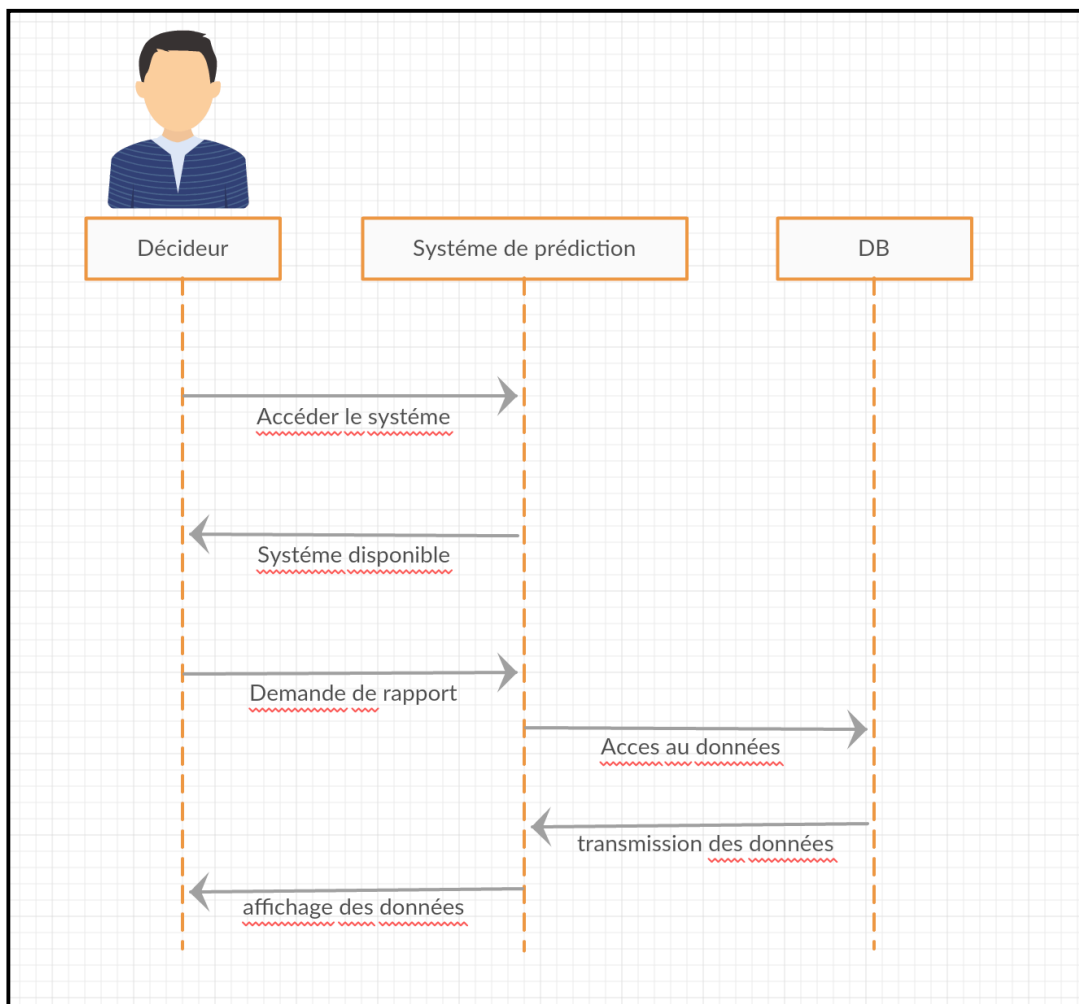
Le système est pratiquement divisé en deux parties, la première pour la source de données en temps réel, le mud logging reçoit les informations à partir des capteurs dans les différents composants de rig, les informations sont enregistrées au niveau de la base de données sous forme de fichiers WITSML. La deuxième partie est pour le décideur (RTOM) qui décide à partir des indicateurs ou des paramètres qui ont été affichés dans le Dashboard.

**1- Identification des acteurs** : notre système contient deux types d'acteurs qui sont :

- Surface logging specialist : c'est la personne qui suit et valide les données de mud logging.
- Décideur/Analyseur : c'est la personne qui contrôle le forage a partir de Dashboard.

**2- Diagramme de séquence :**

Diagramme de séquence permet de représenter les interactions entre les objets en prenant en compte l'ordre de l'échange des messages. Le diagramme de séquence qui représente l'interaction entre le décideur et le système est illustré dans la figure suivante



**Figure III.1:Le diagramme de séquence.**

**IV - Choix de la méthode :**

Pour avoir un modèle complet pour prédire la circulation de perte, il est nécessaire d'examiner les paramètres efficaces. Dans cette modélisation, les entrées sont :

- Depth
- Flow In
- Flow paddle
- Mud Loses
- Mud Température
- Pressure
- Rate of pénétration
- Volume totale

Parmi les ensembles 14973 de données (entrée et sortie) après l'élimination des données illogiques qui sont une indication de l'erreur humaine et le dispositif, les 14973 ensembles de données de 3 puits ont été utilisés dans la modélisation.

paramètre	range
Depth TMD	12-3680
Depth TVD	12-3679.52
Flow In	240-3741
Flow Paddle	0-84
Mud losses Rate	27-6273
Mud losses Vol	0-999
Mud :Lag Time Shakers	0-99999
Mud :MWI AVG	0.642-2.066
MUD : MWO AVG	0-2.057
MUD : Temperature In Avg	15.9-55.2
MUD : Temperature Out Avg	-273-138.8
Active volume	-25-55.9
Total Active volume	27-72.2
Total All Volume	50.1-154.9
Pressure Casing Max	0-491

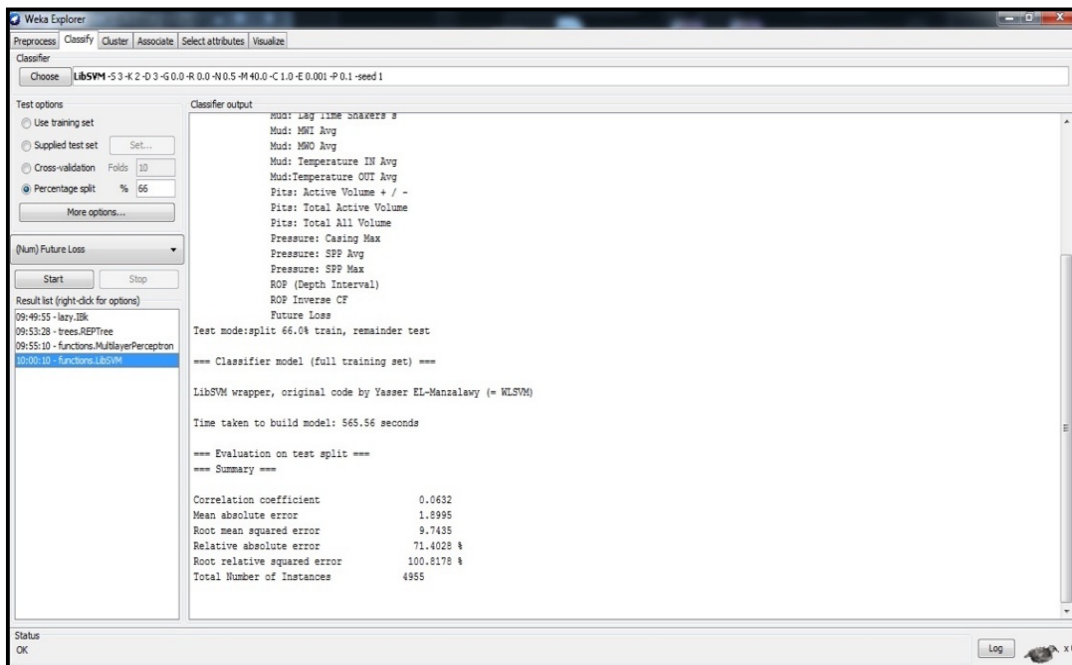
Pressure SPP Avg	0-3963
Pressure SPP Max	0-5903
ROP (Depth Interval)	0-500
ROP Inverse CF	0-4838.9

**Tableau III.1 : Les ranges des paramètres utilisés dans la modélisation**

Après la normalisation des données, nous utilisons 10 fois la validation croisée. La validation croisée (« cross-validation ») est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage, on divise l'échantillon original en k échantillons, puis on sélectionne un des k échantillons comme ensemble de validation et les (k-1) autres échantillons constitueront l'ensemble d'apprentissage. On calcule comme dans la première méthode l'erreur quadratique moyenne. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les (k-1) échantillons qui n'ont pas encore été utilisés pour la validation du modèle. L'opération se répète ainsi k fois pour qu'en fin de compte chaque sous-échantillon ait été utilisé exactement une fois comme ensemble de validation. La moyenne des k erreurs quadratiques moyennes est enfin calculée pour estimer l'erreur de prédiction . [19]

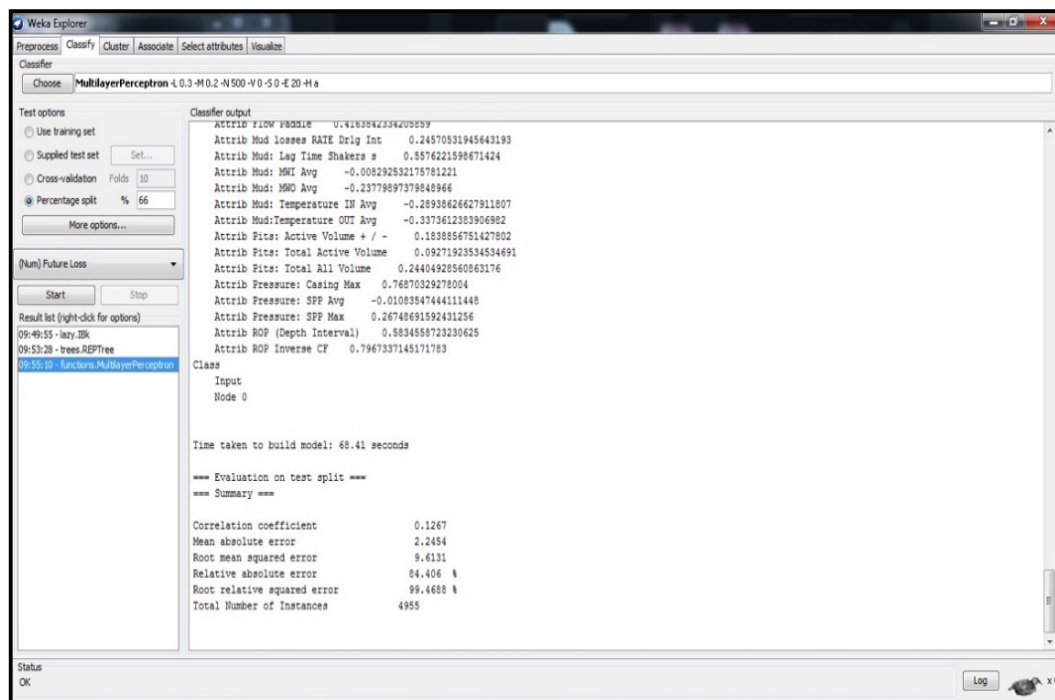
Après l'expérimentation avec le mode de la validation croisée.

**SVM :**

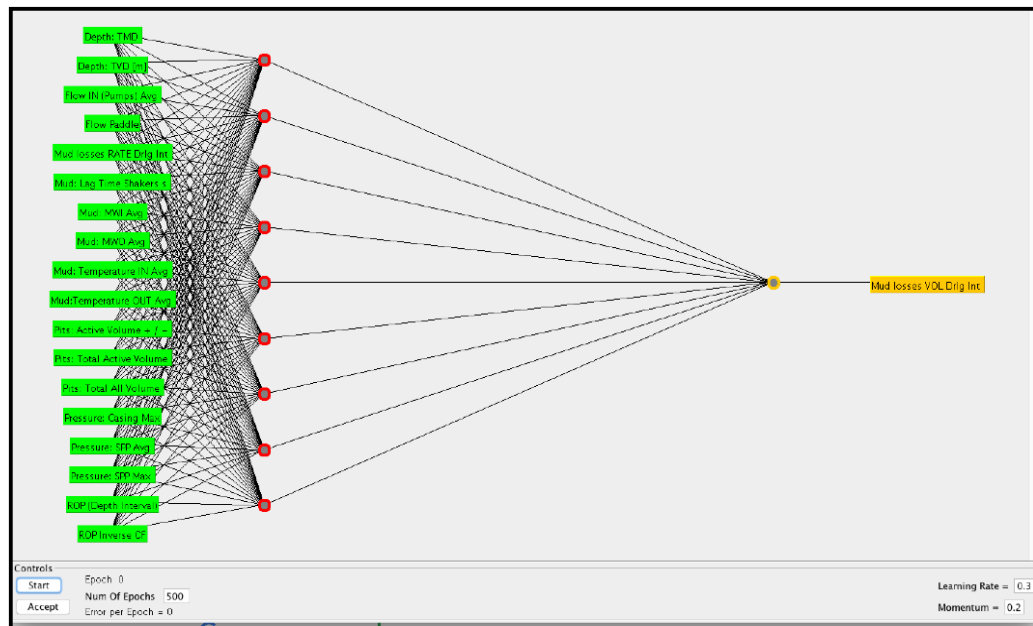


**Figure III.2: la validation croisée de support vecteur machine sous WEKA.**

**Réseau de neurone :**



**Figure III.3 : la validation croisée de réseau de neurone sous WEKA.**



**Figure III.4 : Architecture de RN utilisé pour la prédiction de la perte.**

**Arbre de décision :**

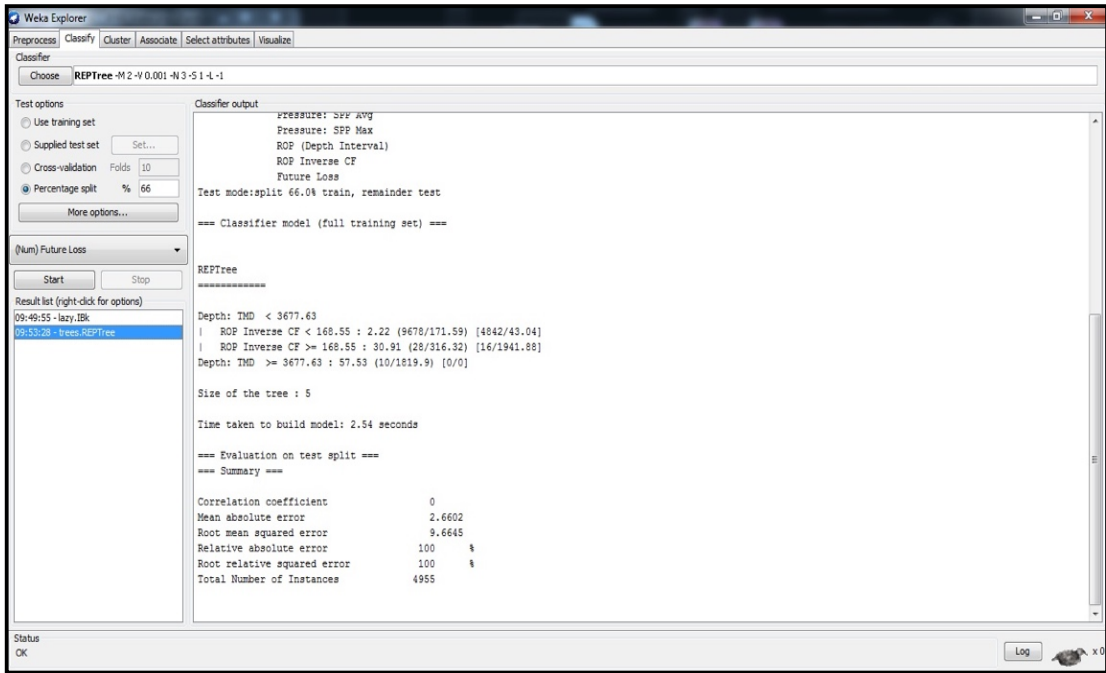
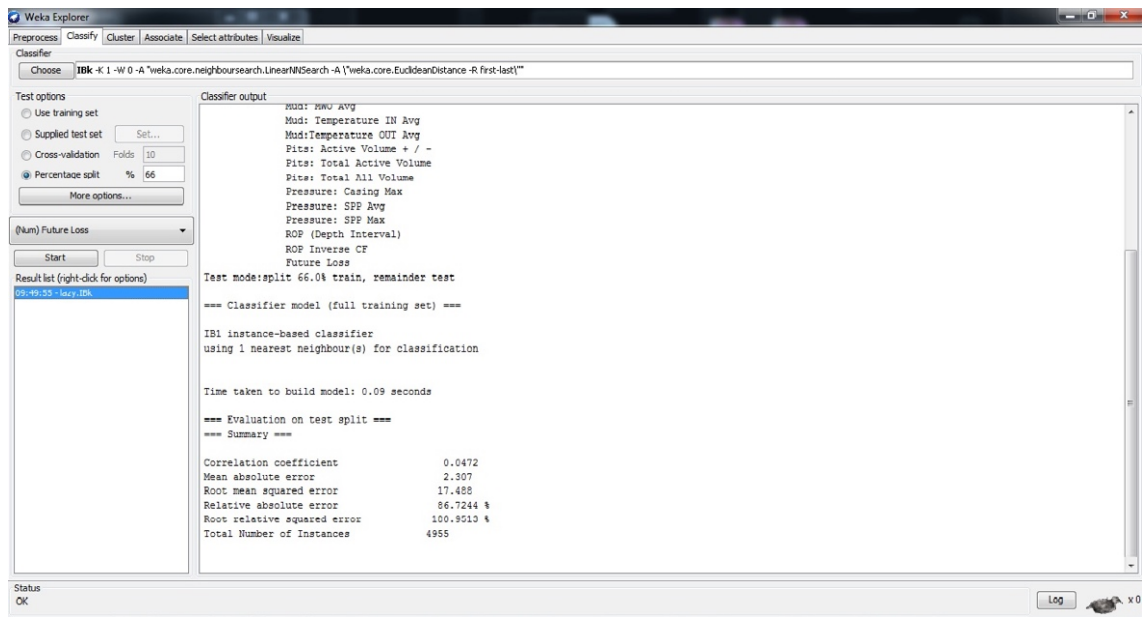


Figure III.5 : la validation croisée de l'arbre de décision sous WEKA.



Figure III.6 : l'arbre de décision de la modèle utilisé sous WEKA.

**K plus proche voisin :**



**Figure III.7 : la validation croisée du k plus proche voisin sous WEKA.**

Nous avons obtenu les résultats présentés dans le tableau suivant (Tableau III.2)

	SVM	ANN	DECISION TREE	KNN
Instances number	14973	14973	14973	14973
Correlation coefficient	0.0816	0.5486	0.3803	0.5032
Mean absolute error	1.8631	1.8241	1.6631	0.8925
Root mean squared error	11.7044	9.8755	10.8064	10.0726
Relative absolute error	68.78%	67.34%	61.40%	32.95%
Root relative squared error	100.44%	84.74%	92.73%	86.44%

**Tableau III.2 : Tableau comparatif entre les différentes méthodes étudiées**

**Corrélation coefficient**

Est la variance de vos données est expliquée par votre modèle. Une valeur faible est pas mal si elle est vraiment le meilleur modèle disponible. [17]

**Mean absolute error**

Est la distance moyenne des modèles prédictions sont des points de données réelles. Absolue dans le titre indique que les prévisions ci-dessous des points de données mangèrent pas traités comme des distances négatives. [17]

### Root mean squared error

Est une autre façon de calculer l'erreur absolue moyenne les deux derniers lignes dans la table suivantes échelle l'erreur à la moyenne. Cela vous permet de comparer entre les modèles construits avec des données / petites valeur plus grandes.

D'après les résultats de nos expériences, l'algorithme K-plus proche voisin (KNN) a donné le meilleur taux de détection et le temps de construction (temps qu'il faut pour construire le modèle sur la formation du réseau de données) plus rapide à 0.04 secondes tout en ayant un intervalle d'erreur 0.5032 de détection comme indiqué dans Table .

L'arbre de décision a eu le deuxième meilleur temps de construction à 1,9 secondes, mais un intervalle d'erreur 1.6631 de détection.

La performance est particulièrement important lorsque l'on considère la classification en temps réel de potentiellement des milliers de trafic de réseaux simultanés. De nos expériences, le KNN semble être le mieux adapté pour les tâches de classification en temps réel en raison de sa vitesse de classification relativement rapide et l'intervalle d'erreur bas.

À condition que le but est que la détermination du type de circulation de la perte et la quantité précise de circulation de perte ne soit pas nécessaire, la sortie du modèle est la prédiction du taux de perte qualitativement dans les fourchettes suivantes :

Le numéro zéro de la perte de  $\leq 1,95$  m<sup>3</sup> / h (perte d'infiltration).

Le nombre de 0,5 pour la gamme de de 1,59 à 7,95 m<sup>3</sup> / h (perte sévère).

Le numéro 1 pour la perte de plus de 7,95m<sup>3</sup> / h (perte complète). [1]

## V - L'architecture de système de prédiction

L'architecture globale de notre système est base sur la prédiction de la perte de circulation, et principalement compose par trois parties principales:

- ❖ Partie d'apprentissage automatique.
- ❖ Partie de prédiction.
- ❖ Partie de réapprentissage.

La figure suivante présente l'architecture de système de prédiction de notre étude :

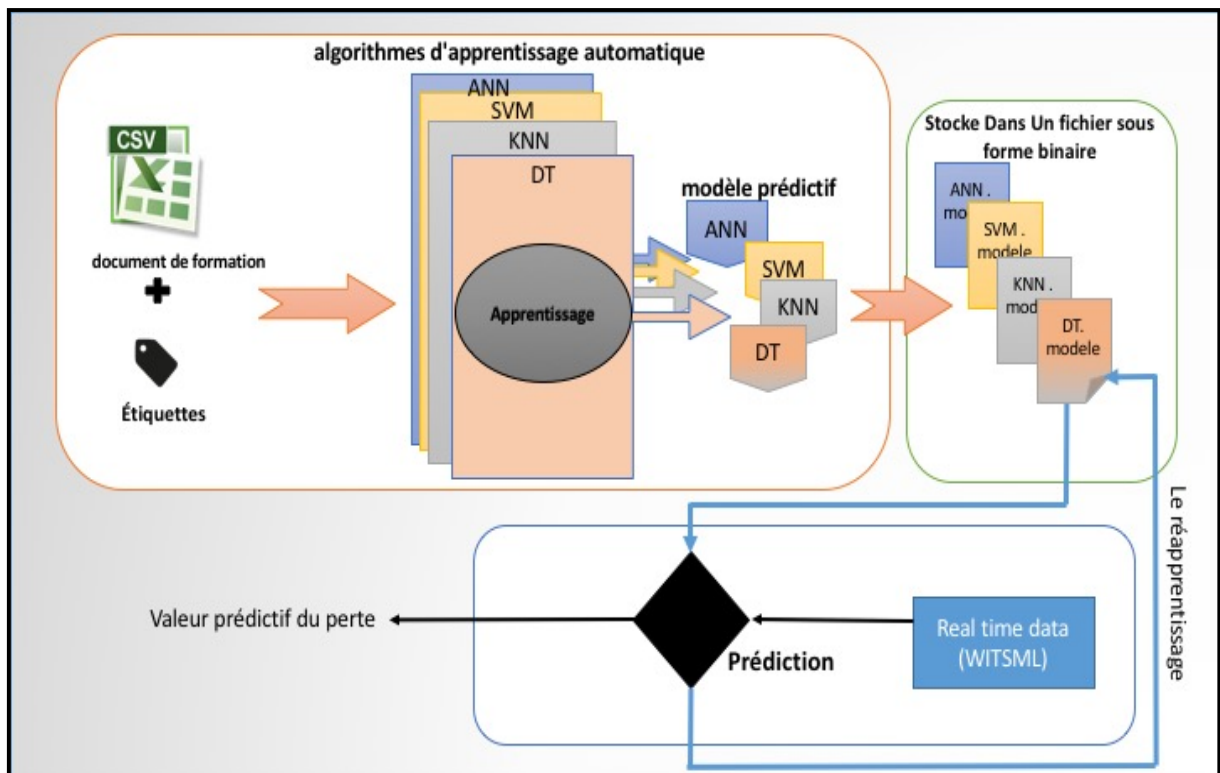


Figure III.8: L'architecture de système de prédiction.

**A- Partie d'apprentissage automatique**

Le système prend comme entrée un ensemble des données ce forme des fichiers Excel, et dans cette phase consiste à utiliser des méthodes de régression (dans notre étude: SVM, ANN, KNN, DT). Cette phase consiste à initialiser ou créer la base des modèles prédictifs, considérant que le système prend les deux tiers de fiches pour apprend a le problème de la perte de circulation et le reste pour la décision.

**B- Partie de prédiction**

Le but de cette partie est donner une solution, grâce à prédictif une valeur du perte de boue de forage. Et a travers d'insérer et de lier les donnes qui sont sous la forme d'une fichier WITSML avec un modèle prédictif.

**C- Le réapprentissage**

Cette phase assure le développement continu de notre système de prédiction et de réduire la taille des erreurs avec le temps. Ce processus se fait en collaboration avec la prédiction en temps réel.

## **VI - Conclusion**

Dans ce chapitre nous avons donné une spécification sur notre conception du système, cette phase représente l'une des phases les plus importantes dans le processus de développement d'un logiciel. Elle décrit le système de point de vue général et détaillée pour comprendre et assister la phase de programmation.

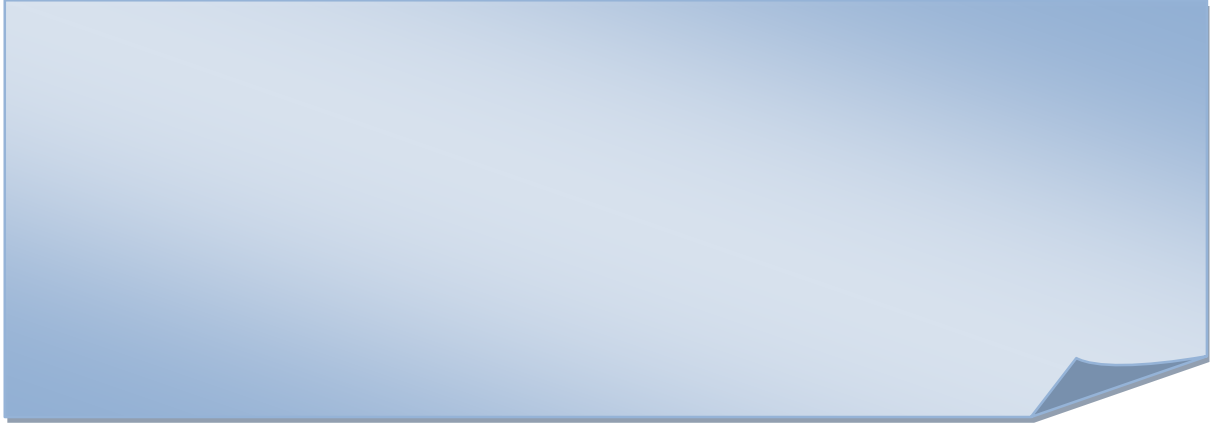
Nous avons vu dans ce chapitre l'architecture globale du système et le fonctionnement du système.

Ensuite, nous avons choisi la meilleure méthode répondre aux besoins de l'utilisateur, ainsi que les méthodes de prévision utilisée et nous avons détaillé l'architecture de ce système de prédiction de perte de la boue de forage.

Dans le prochain chapitre, nous allons illustrer la réalisation de notre système en représentant quelques interfaces et quelques résultats obtenus, avec les structures des données qui sont choisis pour implémenter ce système.

# ***Chapitre IV:***

## ***Implémentation et réalisation***



## **I - Introduction**

L'implémentation d'un logiciel devient après un enchaînement de plusieurs étapes dans le processus de développement, et son but principal est de réaliser un produit capable de résoudre les problèmes posés en utilisant des outils et des algorithmes.

Dans le précédent chapitre nous avons choisi la méthode utilisée pour résoudre notre problème, le présent chapitre donne une vue sur le système et les outils utilisées afin d'arriver à un système fiable.

Dans ce chapitre nous allons présenter en premier temps, l'environnement du développement avec les différentes bibliothèques utilisées, nous terminons ce chapitre par une conclusion.

## **II - Environnement du développement**

Notre système était développé sur un ordinateur de processeur Core I7 avec une RAM de 16 Go sous Mac OS X El Capitan Version 10.11.3, mais on peut implémenter ce système sur n'importe quel machine grâce au virtuel machine de Java.

### **1 - Java**

Le langage Java est un langage de programmation orienté objet développé par la société Sun Microsystems.

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et Frameworks associés visent à guider, sinon garantir, cette portabilité des applications développées en Java.

#### **1.1 - Environnement de programmation (Eclipse)**

Eclipse est un environnement de développement intégré (IDE) pour développer des applications utilisant le langage de programmation Java et d'autres langages de programmation tels que C / C ++, Python, PERL, Ruby, etc.

La plate-forme Eclipse, qui constitue le fondement de l'IDE Eclipse est composé de plug-ins et est conçu pour être extensible à l'aide de plug-ins supplémentaires. Développé en Java, la plate-forme Eclipse peut s'être utilisée pour développer des applications client riche, des environnements de développement intégrés et d'autres outils. Eclipse peut être utilisé comme un IDE pour tout langage de programmation pour lequel un plug-in est disponible. [18]



**Figure IV. 1: Environnement de développement.**

## **1.2 - JDK**

**JDK « Java Développement Kit »** de Sun Microsystems, qui est une mise en œuvre de référence puisqu'elle est fournie par le créateur du Java. Le JDK est utilisable sur les plates- formes Linux, Solaris, Mac OS, et Windows

**JVM « Java Virtual machine »** est une machine virtuelle permettant d'interpréter et d'exécuter le byte-code Java. Ce programme est spécifique à chaque plate-forme ou couple (machine/système d'exploitation) et permet aux applications Java compilées en byte-code de produire les mêmes résultats quelle que soit la plate-forme

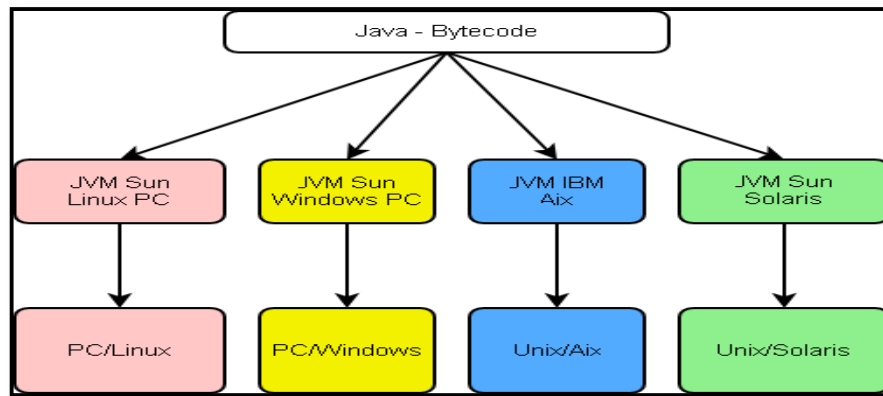


Figure IV.2: L'architecture de JVM et JDK

## 2 - Weka (Waikato environment for knowledge analysis)

Weka, en français : « environnement Waikato pour l'analyse de connaissances ») est une suite de logiciels d'apprentissage automatique. Écrite en Java, développée à l'université de Waikato en New-Zélande. Weka est un logiciel libre disponible sous la Licence publique générale GNU (GPL).

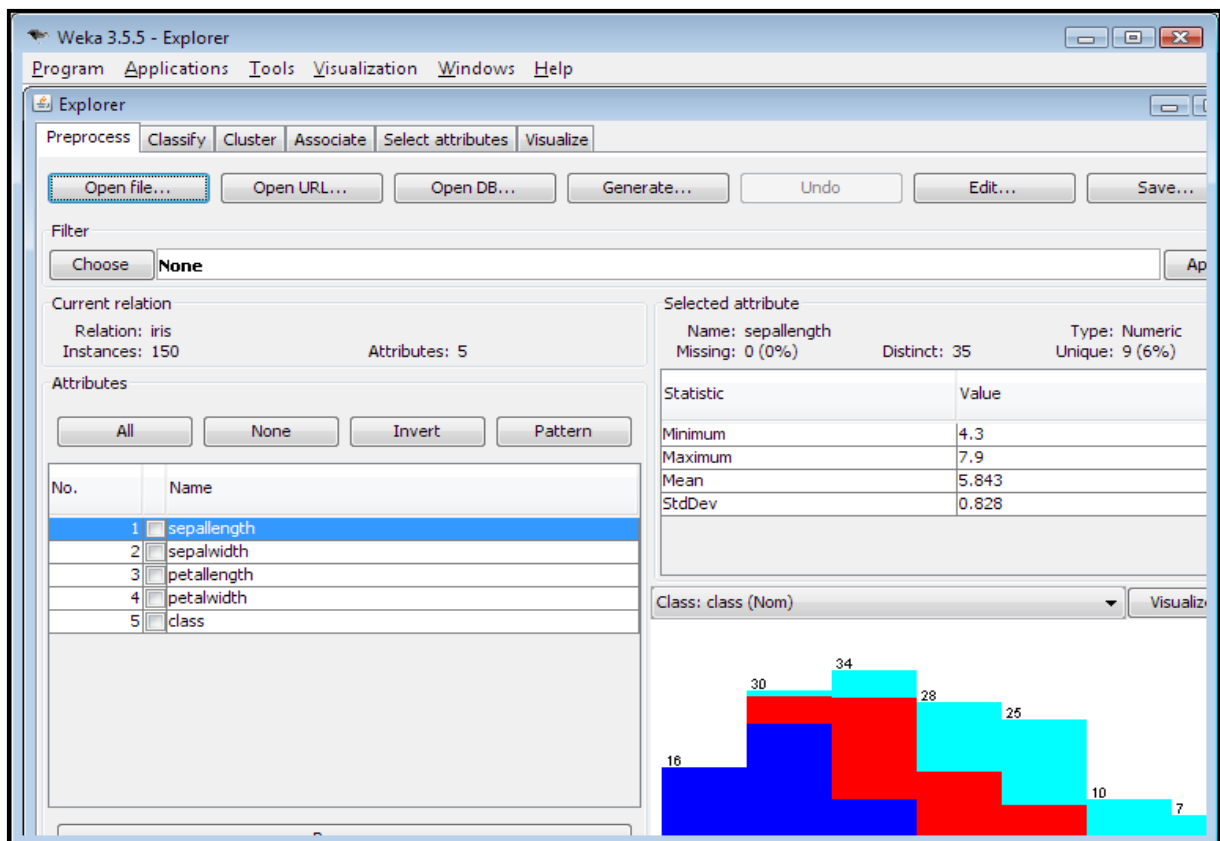


Figure IV.3 : L'espace de travail Weka.

L'espace de travail Weka contient une collection d'outils de visualisation et d'algorithmes pour l'analyse des données et la modélisation prédictive, allié à une interface graphique pour un accès facile de ses fonctionnalités. La version « non-Java » originale de Weka était un front end en **Tcl/Tk** pour des algorithmes de modélisation (essentiellement tierces) implémentés dans d'autres langages de programmation, complété par un des utilitaires de préprocesseur de données en C [13] , et un système à base de **makefile** pour lancer les expériences d'apprentissage automatique. Cette version originale était avant tout conçue comme un outil pour analyser des données agricoles<sup>2,3</sup>, mais la version plus récente entièrement basée sur Java (Weka 3) [14] , pour laquelle le développement a débuté en 1997, est désormais utilisée dans beaucoup de domaines d'application différents, en particulier pour l'éducation et la recherche. Les principaux points forts de Weka sont qu'il :

- est libre et gratuit, distribué selon les termes de la licence publique générale GNU
- est portable car il est entièrement implémenté en Java et donc fonctionne sur quasiment toutes les plateformes modernes, et en particulier sur quasiment tous les systèmes d'exploitation actuels
- contient une collection complète de préprocesseurs de données et de techniques de modélisation ;
- est facile à utiliser par un novice en raison de l'interface graphique qu'il contient. [15]

### **III - Résultat obtenues**

Notre système contient 4 différentes interfaces, Chaque interface représente l'un des algorithmes étudiés : L'utilisateur peut basculer entre les algorithmes et la sélection des meilleurs d'entre eux Les interfaces suivantes représentent les interfaces de notre système

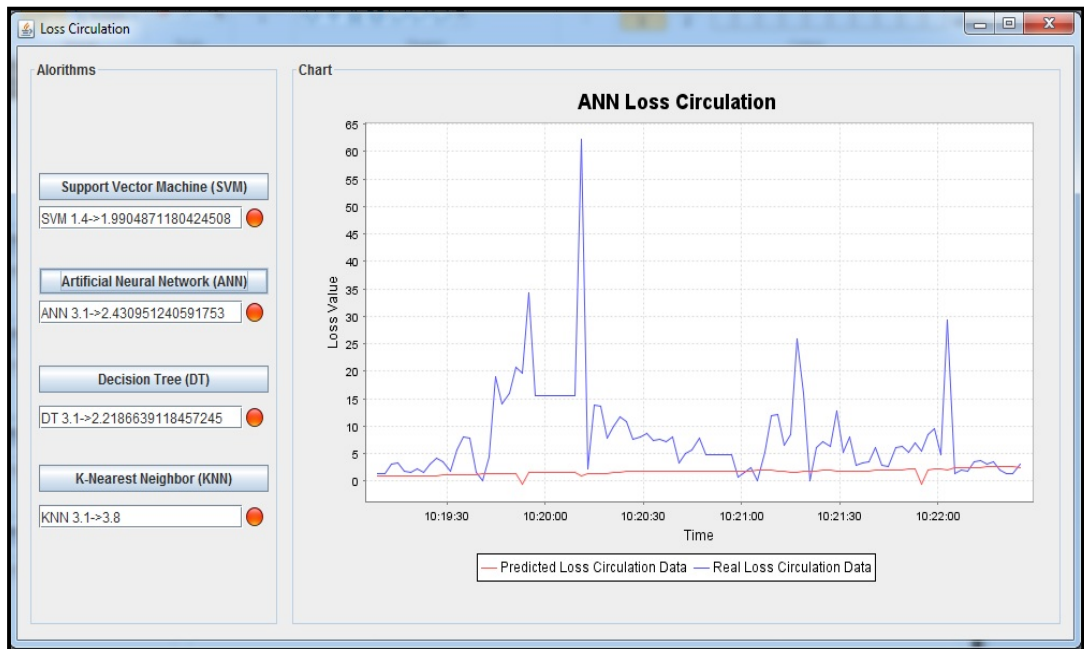


Figure IV.4: L'interface de ANN.

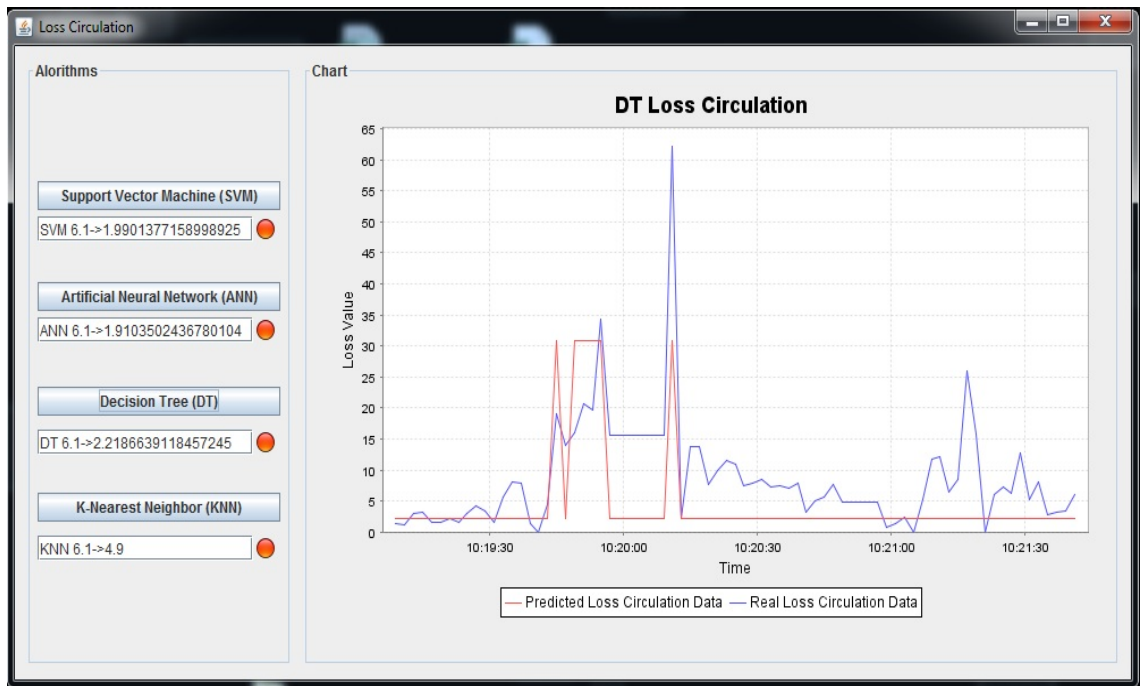


Figure IV.5: L'interface de DT.

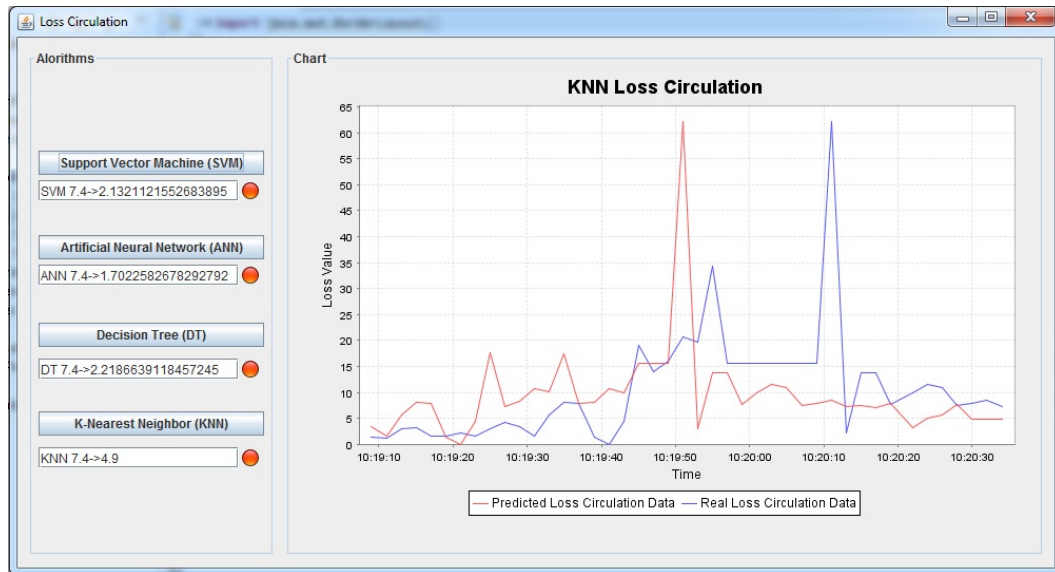


Figure IV.6: L'interface de KNN.

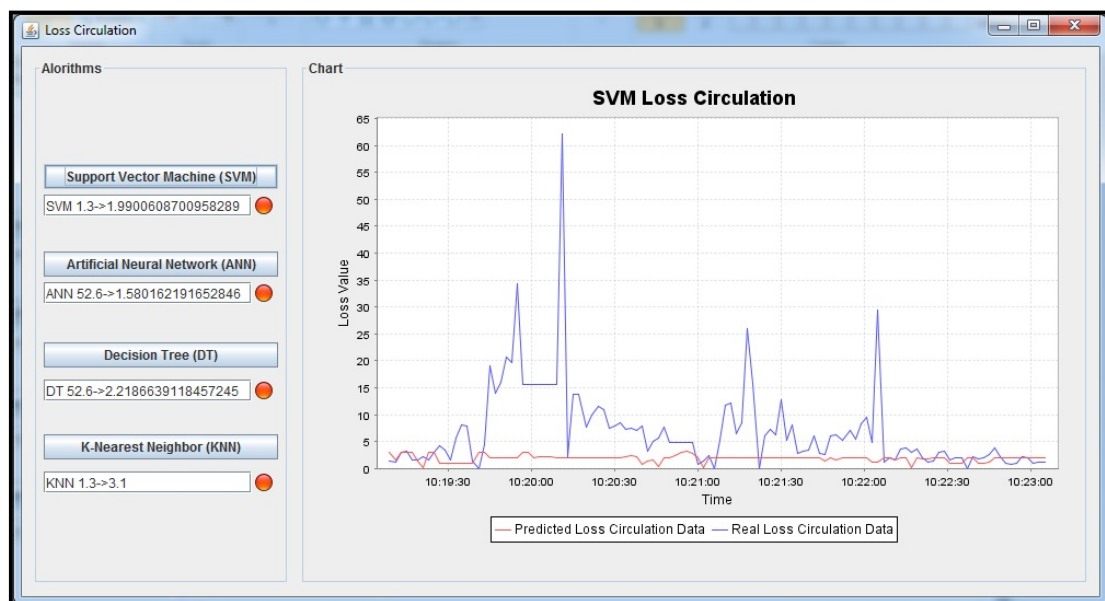


Figure IV.7 : L'interface de SVM.

## IV - Conclusion

Dans ce chapitre, nous avons présenté l'implémentation de notre système, pour cela, nous avons décrit les outils de réalisation du système de suivi des forages, en commençant par la présentation de l'environnement du travail, ainsi les outils utilisés, enfin nous avons montré quelques interfaces qui représentent les résultats de notre application.

La perte de circulation est un problème commun dans l'industrie de forage qui provoque une forte dépense sur les entreprises de forage. Aujourd'hui, la minimisation de la perte de circulation est un objectif principal pour les ingénieurs de forage

Dans ce travail ont été intéresser à présenter un système de prédiction et solutions relatives à la perte de boue de forage en utilisant les techniques de la fouille de données, En noyant d'anticiper en premier lieu les causes de survenue de ces problèmes afin d'éviter les risques et les coûts de forage excessifs car si ces risques sont identifiés à l'avance.

Nous avons vu en premier temps les notions de base de ce domaine très vaste qui est le forage pétrolier, ce qui nous a permis d'acquérir les concepts et processus importants pour son déroulement. Cependant nous avons constaté que malgré les moyens déboursés dans ce sens les foreurs peuvent se heurter à plusieurs obstacles entre autres la perte de boue de forage.

Ensuite, nous avons obtenu Le processus, les taches et les différent techniques de data Mining.

Dans la phase de conception nous avons proposé une architecture pour implémenter ce genre de système, nous avons donné l'architecture globale et son fonctionnement de façon générale, ensuite Nous avons choisi La meilleure méthode Parmi les modèles que nous avons étudiés.

Le dernier chapitre donne une vue de fond de notre système, à travers les outils de programmation ainsi que les structures de données et les méthodes utilisés.

Les données utilisées dans le cadre de l'étude de la prédiction de la perte de boue de forage (mud losses) appartiennent à des puits forés en Sud de l'Algérie (Hassi Messouad). la perte de boue de forage est prédite à l'aide de l'algorithme KNN (K Nearest Neighbors ) pour régression qui est plus performant que les autres algorithmes de régression à savoir les RNA(réseau de neurone artificiel) et les SVM (support vecteur machine)et AD (arbre de décision) d'après les comparaisons que nous avons effectuées dans l'outil Learning-machine WEKA présentées sur les tableaux (Tableau III.2) Les résultats ont indiqué que la perte de boue de forage peut être prédite avec une haute précision.

Notre application peut être utilisée comme aide à décision pour élaborer les programmes de forage. Elle prédit la perte de boue a l'avance et elle fait le réapprentissage à partir de données en temps réel. Elle peut servir comme un système d'aide à la décision pour minimiser la perte de boue de forage en temps réel.

- [1] Hojjat Toreifi • Habib Rostami • Abbas Khaksar manshad :”New method for prediction and solving the problem of drilling fluid loss using modular neural network and particle swarm optimization algorithm “, Article: 1 September 2013 / Accepted: 11 January 2014 / Published online: 20 February 2014 at Springer.
- [2] Dr. Abdelhamid DJEFFAL : “Cours Fouille de données avancée “ , cours de Master 2 IDM .
- [3] MILI Saoussen, KAOUANE Housseyn : Conception et réalisation d’un système de prévention des « Stuck Pipe» HASSI MESSAOUD, Mémoire de Fin d’étude, esi 2014.
- [4] TutorialPoint : ” Data mining data patern evolution “ cours de totutorialPoint
- [5] assem shale shaker comitee : “Driling fluids processing “ .Handbook
- [6] Jean-Paul NGUYEN: “ techniques d’exploitation pétrolière le forage”. Cour de l’institut fançais de Pétrole . paris
- [7] M. DADDOU: “Introduction à l’industrie pétrolière Et Initiations au forage “,FORMATION D’INGENIEURS FORAGE , Sonatrach
- [8] A. SLIMANI , M. DADDOU: “ Formation JPF Module M1 “. Division Forage . Département Formation. Sonatrach
- [9] F. GRONDIN : “ Formation JPF Module M2 “. Division Forage .Département Formation. Sonatrach
- [10] A. SLIMANI : “ Formation JPF Module M3 “. Division Forage . Département Formation. Sonatrach
- [11] A. SLIMANI : “ Formation JPF Module M4 “. Division Forage . Département Formation. Sonatrach
- [12] Zehri Ilyes , Slimani Sami :“Optimisation des paramètres de forage dans les Réservoirs cambro- ordoviciens”, Mémoire fin d’études MASTER PROFESSIONNEL , UNIVERSITE KASDI MERBAH OUARGLA , 2011/2012
- [13] Svetlana S. Aksenova , WEKA Explorer Tutorial for WEKA Version 3.4.3 , California State University, Sacramento California, 2004
- [14] Remco R. Bouckaert ,Eibe Frank ,Mark Hall ,Richard Kirkby ,Peter Reutemann ,Alex Seewald ,David Scuse : WEKA Manual for Version 3-7-8 ,The Waikato University , 2013
- [15] Charalampos Mavroforakis, Data mining with WEKA A use-case to help you get started,Boston University CS105, Fall 2011
- [16] PetoWiki: <http://petrowiki.org/PetroWiki>, site web pour les information sur le domaine pétrolier

- [17] Ajayi Adebawale, Idowu S.A, Anyaehie Amarachi A:” Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection ” , IJSCE , 2013
- [18] MERIZIG ABDELHAK ,” Conception et Réalisation d’un système d’information pour le suivi des dattes ”, memoire de master Université Mohamed Khider – BISKRA 2013
- [19] Djamil Rezki :” Système intelligent d’aide à la décision pour le pilotage d’un processus de forage pétrolier ”, memoire de magister Université BATNA 2 ,2016
- [20] MILPARK-CKS: “Manuel technique fluids de forage”, Tome I et Tome II, edition 1990