

**République Algérienne Démocratique et Populaire**



**Ministère de l'Enseignement Supérieur et de la Recherche Scientifique**

**UNIVERSITÉ ECHAÏD 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**

**Prédiction de coincement de la garniture de forage par les techniques des fouilles des Données**

Présenté par: M<sup>elle</sup> Bebboukha Meriem  
M<sup>elle</sup> Bouta Hadjer

Soutenu devant le jury composé de

Mr. Merouane Zoubeydi	MA (B)	Rapporteur	Univ. d'El Oued
Mr. Saci Medilah	MA(A)	President	Univ. d'El Oued
Mr. Ibrahim Lejdel	MA(A)	Examinatrice	Univ. d'El Oued

Année universitaire 2015-2016.



## Remerciements



*Avant tous, nous remercions dieu le tout puissant de nos avoir donné le courage et la patience pour réaliser ce travail malgré toutes les difficultés rencontrées .*

*Nous remercions infiniment tous qui nous a aidé de près et de loin d'avoir compléter ce travail et dépasser tous les obstacles surtout notre enseignant*

*" **Mr. Merouane Zoubeydi** "qui n'a pas cessé de nous donner les conseils et les bonnes orientations et nous prive pas de son temps et aussi, Nos remerciements, les plus vifs " **Mr.Ghenabzia Ahmed**" & "**Mr. Hicham kaid**" dont nous écoute avec grande patience pendant toute la période de préparation malgré ses contraintes d'enseignement  
Nous remercions vivement l'enseignante "**Melle. Settou Tarablesse**"*

*Nous remercions vivement a tous*

*"**Mr.Ibrahim Belhadi**" INGENIEUR SYSTEME ELECTRONIQUE*

*"**Mr. djamil REZKI**" CHEF DE SERVICE MAINTENANCE*

*INFORMATIQUE à Direction Pilotage Stratégique*

*"**Mr. hadj FEDLAOUI**" CHEF DE SERVICE CONTRÔLE QUALITÉ à la  
DIRECTION MUD LOGGING.*

*"**Mr. Ahmed BENHAMIDA**" CHEF de Projet Informatique PPL à la  
Sonatrach Direction des Opérations Hassi Messaoud*

*"**Mr. Mourad SELAMI**" Drilling Manager at Schlumberger*

*Notre reconnaissance va aussi à tous ceux qui ont collaboré à notre Formation en particulier les enseignants du département D'Informatique, de l'université Hama Lakhdar d'El-Oued.*

*Nous remercions également tous ceux qui ont participé de près Ou de loin à élaborer ce travail.*



# RÉSUMÉ

Le coincement de la garniture de forage (*stuck pipe*) est un problème commun avec impact énorme sur l'efficacité et les coûts dans l'industrie du forage pétrolier.

Généralement, les troubles de la tige coincée sont résolus après leurs occurrences en utilisant des techniques standard; ici, nous essayons de prévoir les causes de survenue de ces problèmes à éviter les risques et les coûts de forage excessives. Si ces risques sont identifiés à l'avance, de meilleures solutions peuvent être fournies pour réduire les conséquences associées.

L'objet de notre étude est d'essayer de mettre en place un modèle de prédiction du coincement de la garniture de forage en utilisant les algorithmes de la fouilles des données data-mining.

## **Mots clés :**

Forage pétrolier, coincement de la garniture de forage, prédiction, data-mining.

## **ABSTRACT**

The jamming of the drill string (stuck pipe) is a common problem with enormous impact on the effectiveness and costs in the oil drilling industry. Generally, disorders of stuck pipe are fixed after their occurrences using standard techniques; Here, we try to predict the causes of occurrence of these problems to avoid risks and excessive drilling costs. If these risks are identified in advance, better solutions can be provided to reduce the associated consequences.

The purpose of our study is to try to develop a predictive model of wedging the drill string using the algorithms of data mining ( data searches )

### **Keywords:**

oil drilling, stuck pipe, prediction, data mining

## الملخص

التشويش على سلسلة الحفر (الأنابيب العالقة) هو مشكلة مشتركة مع تأثير هائل على الفعالية والتكاليف في صناعة التنقيب عن النفط.

عموما، يتم إصلاح اضطرابات الأنابيب العالقة بعد الحوادث وذلك باستخدام التقنيات القياسية. هنا، ونحن نحاول التنبؤ بأسباب حدوث هذه المشاكل لتجنب المخاطر وتكاليف الحفر المفرطة. إذا تم تحديد هذه المخاطر في وقت مبكر، ويمكن توفير أفضل الحلول للحد من العواقب المرتبطة بها.

والغرض من دراستنا هو محاولة اعداد نموذج تنبئي من تشويش منصة الحفر باستخدام خوارزميات

التنقيب عن البيانات

الكلمات المفتاحية:

التنقيب عن النفط، التصاق أنبوب حفر الآبار، التنبؤ، استخراج البيانات

## *Liste des figures*

---

### **I. Générale au Forage pétrolier**

Figure I.1 : Système de forage .....	4
Figure I.2: Le processus de forage.....	5
Figure I.3 : Principe du forage Rotary.....	6
Figure I.4: Sonde de forage rotary.....	7
Figure I. 5: L’outil de forage.....	9
Figure I.6 : Stabilisateur.....	10
Figure I.7 : Quelques éléments de la garniture de forage.....	10
Figure I .8: Exemple de Stuck mécanique.....	14
Figure I.9: Exemple de Stuck différentiel.....	15

### **II. Data-mining**

Figure II. 1: Processus de data-mining.....	18
Figure II.2: phase d’apprentissage pour les SVM.....	23
Figure II.3 : Etape de test du modèle.....	24
Figure II. 4 : Schéma globale de fonctionnement de l’SVM.....	24

### **III. Conception**

Figure III. 1 :Description de la transmission de données RTOM.....	27
Figure III.2: l'architecture global du système.....	28
Figure III.3:Normalisation des données dans WEKA.....	30
Figure III.4: l’expérimentation et l’évaluation en utilise l'algorithme de SVM.....	32
Figure III.5: l’expérimentation et l’évaluation en utilise l'algorithme de RN.....	32
Figure III.6: Architecture de RN utilisé pour la prédiction .....	34
Figure III.7 : l’expérimentation et l’évaluation en utilise l'algorithme de KNN.....	34

## **IV. Implémentation et réalisation**

Figure IV.1:La langage de développement.....	36
Figure IV.2: Environnement de développement-NetBeans.....	37
Figure IV.3: L'outil WEKA.....	38
Figure IV.4:L'interface principale de l'application.....	38
Figure IV.5:L'interface principale de prototype version zéro.....	39
Figure IV.6:L'interface du comparaison entre les méthodes d'application.....	39
Figure IV.7:L'interface de résultat avec le meilleur algorithme (KNN) no coincement ...	40
Figure IV. 8: L'interface de résultat avec le meilleur algorithme (KNN) coincement .....	40

## *Liste de tableau*

---

Table.III.1 : Les paramètres de notre modèle .....30

Table.III.2 : comparaison entre les algorithmes de classification .....34

# *Sommaire*

---

<b>Liste des figures .....</b>	<b>I</b>
<b>Liste des tableaux .....</b>	<b>III</b>
<b>Sommaire .....</b>	<b>IV</b>
<b>Liste d'abréviation .....</b>	<b>VI</b>
<b>Introduction générale.....</b>	<b>1</b>

## **Chapitre I : Générale au Forage pétrolier**

Introduction .....	3
1- Forage pétrolier.....	3
2- Principe de forage rotary.....	6
3- Description généralement de forage.....	6
4- Equipements de lavage.....	8
4-1. Le mat de forage.....	8
4-2. La substructure.....	8
4-3. Le mouflage.....	8
4-4. Le moufle fixe.....	8
4-5. Le moufle mobile.....	8
4-6. Le crochet.....	8
4-7. Le câble de forage.....	9
4-8. Le treuil.....	9
5- Garniture de forage.....	9
5-1. L'outil de forage.....	9
5-2. Stabilisateurs.....	10
5-3. Les masses-tiges ( Drill Collars , DC).....	10
5-4. Les trains de tiges (Drill Pipes, DP).....	10
6- Parameters de forage.....	11
6-1. Les paramètres mécaniques.....	11
6-2. Les paramètres hydraulique.....	12
7- Types des problèmes de forage.....	13
7-1. Mud Contamination.....	13
7-2. Loss Circulation (Perte).....	13
7-3. Venue (Kick).....	13
7-4. Coincements de la garniture de forage (Stuck Pipe).....	14
Conclusion.....	16

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

Introduction.....	17
1- Qu'est ce que la fouille de données ?.....	17
2- Le processus de data-mining.....	17
3- Les taches de data-mining.....	18
3-1. La classification.....	19
3-2. L'estimation.....	19
3-3. La prediction.....	20
4- Techniques du data-mining.....	20
4-1. Les réseaux de neurons.....	20
4-1-1. Avantages et inconvénients.....	20
4-2. L'algorithme des k-Plus proches voisins.....	21
4-2-1. Comment cela marche-t-il ?.....	21
4.2.2 Avantages et inconvénients.....	21
4-3. Les machines à support vecteurs (SVM).....	22
Conclusion.....	25

## **Chapitre III : Conception**

Introduction.....	26
1- Description du processus de prévision des coincements.....	26
1-1. Le projet RTOM.....	26
2- Architecture global de système.....	27
3- Fonctionnement global .....	28
4- Description des données.....	29
5- Comparaisons entre les méthodes.....	31
Conclusion.....	34

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

Introduction.....	35
1. Environnement et outils de mise en œuvre.....	35
1-1. Java.....	35
1-2. Netbeans.....	36
1-3. Weka (machine Learning).....	36
2. L'application développée.....	37
Conclusion.....	39
Concision general.....	40
Bibliographies.....	41

## *List des privations*

---

ANN: **A**rtificiel **N**eural **N**etwork (réseaux de neurone )

KNN : **K** Nearest Neighbor (k-plus proche voisin)

RPM : Rotation per minute

RTOM: **R**eal **T**ime **O**perating **M**anagement

SVM: **S**upport **V**ector **M**achine

WEKA : **W**aikato **E**nvironment for **K**nowledge **A**nalysis

WOB: **W**eight **O**n **B**it (Poids sur l'outil)

### **1- Introduction**

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

La compétitivité accrue entre les compagnies pétrolières internationales incite à aller vite pour creuser des puits devenus de plus en plus coûteux. Aujourd'hui le forage, réalisé généralement par une société de service spécialisée, est devenu une activité très technique. La tendance actuelle est au développement de méthodes ou de moyens d'aide pour assister le foreur dans le choix des différents paramètres du forage pour aller vite et diminuer le prix de revient du mètre foré.

### **2- Problématique**

Pendant les différentes étapes du processus de forage, le superviseur des opérations peut rencontrer divers obstacles qui ralentissent l'avancement du forage, par conséquent, l'augmentation du prix de revient. Parmi ces obstacles nous avons identifié le problème de coincement de la garniture de forage «*Stuck Pipe*», ce phénomène bloque l'avancement du forage car le foreur ne peut plus ni avancer ni retirer la garniture du puits, il peut être détecté grâce à certains paramètres hydrauliques et mécaniques, néanmoins ce qui serait intéressant est de pouvoir le prévoir bien avant son apparition, et le représenter d'une façon explicite, afin d'en faire bénéficier les responsables. Le coincement est un problème très courant qui représente des pertes de centaines de millions de dollars chaque année pour l'industrie pétrolière.

### **3- Objectifs**

Pour pallier cette problème cités ci-dessus notre système a pour objectifs de permettre à ces exploitants de :

- ❖ Notre sujet consiste à essayer de mettre en place un système de prédiction du coincement de garniture de forage en utilisant les algorithmes de la fouille des données Datamining.

Notre travail est divisé en quatre chapitres principaux:

- Le premier chapitre consiste à état de l'art qui donne les notions de base du forage pétrolier, et ces divers problèmes dont le coincement de la garniture qui représente notre principale problématique.

- Le deuxième chapitre aborde les méthodes de data-mining qui sont utilisées pour résoudre le problème de ce travail.
- Dans le troisième chapitre nous présentons l'architecture globale du système, et nous décrivons d'une façon détaillée l'interaction entre les entités du système, les outils et les méthodes utilisées de développement du notre système.
- Dans le quatrième chapitre nous présentons l'environnement de développement, avec quelques interfaces de notre logiciel, en terminant par une discussion sur les résultats obtenus.

Dans la conclusion générale nous dressons la synthèse du travail réalisé dans ce mémoire suivi des perspectives envisagés.

*Chapitre I :*

---

*Généralité au Forage pétrolier*

## **Introduction**

Après avoir fait de nombreuses études géologiques et géophysiques, les experts de forage ont décidé de l'emplacement de l'huile.

Nous devons maintenant creuser cela signifie forage profond, pour confirmer la présence de pétrole et de gaz. Dans cette étude, nous nous intéressons uniquement au sol de forage vertical, conscient que d'autres types de puits, et l'existence d'un tel forage horizontal sur terre et diverses autres techniques d'extraction.

Dans ce chapitre, on s'intéresse d'une part à introduire et présenter un état de l'art le domaine au forage pétrolier, et d'autre part à exposer le principe de forage, les problèmes que nous rencontrons pendant le forage.

## **1- Forage pétrolier**

Le forage pétrolier est l'ensemble des opérations permettant d'atteindre les roches poreuses et perméables du sous-sol, susceptibles de contenir des hydrocarbures liquides ou gazeux.

L'implantation d'un forage pétrolier est décidée suite a des études géologiques et géophysiques effectuées sur un bassin sédimentaire. Ces études permettent de se faire une thèse de la constitution du sous-sol et des possibilités de gisements, mais elles ne peuvent préciser ou non la présence d'hydrocarbures. Seuls les forages pourront confirmer les hypothèses faites et mettre en évidence la nature des fluides contenus dans les roches.

L'incertitude quant à la nature des fluides piégés dans le sous-sol et la complexité des dépôts sédimentaires expliquent le nombre encore élevé de forages d'exploration négatifs (un sondage d'exploration productif pour une dizaine négatifs) et la part importante des forages dans le coût de la recherche.[1]

Donc, les objectifs des forages pétroliers sont toujours les mêmes a savoir :

- Atteindre la profondeur souhaitée par les géologues.
- Avoir la possibilité de tester ou de mettre en production le (ou les) réservoir (s).

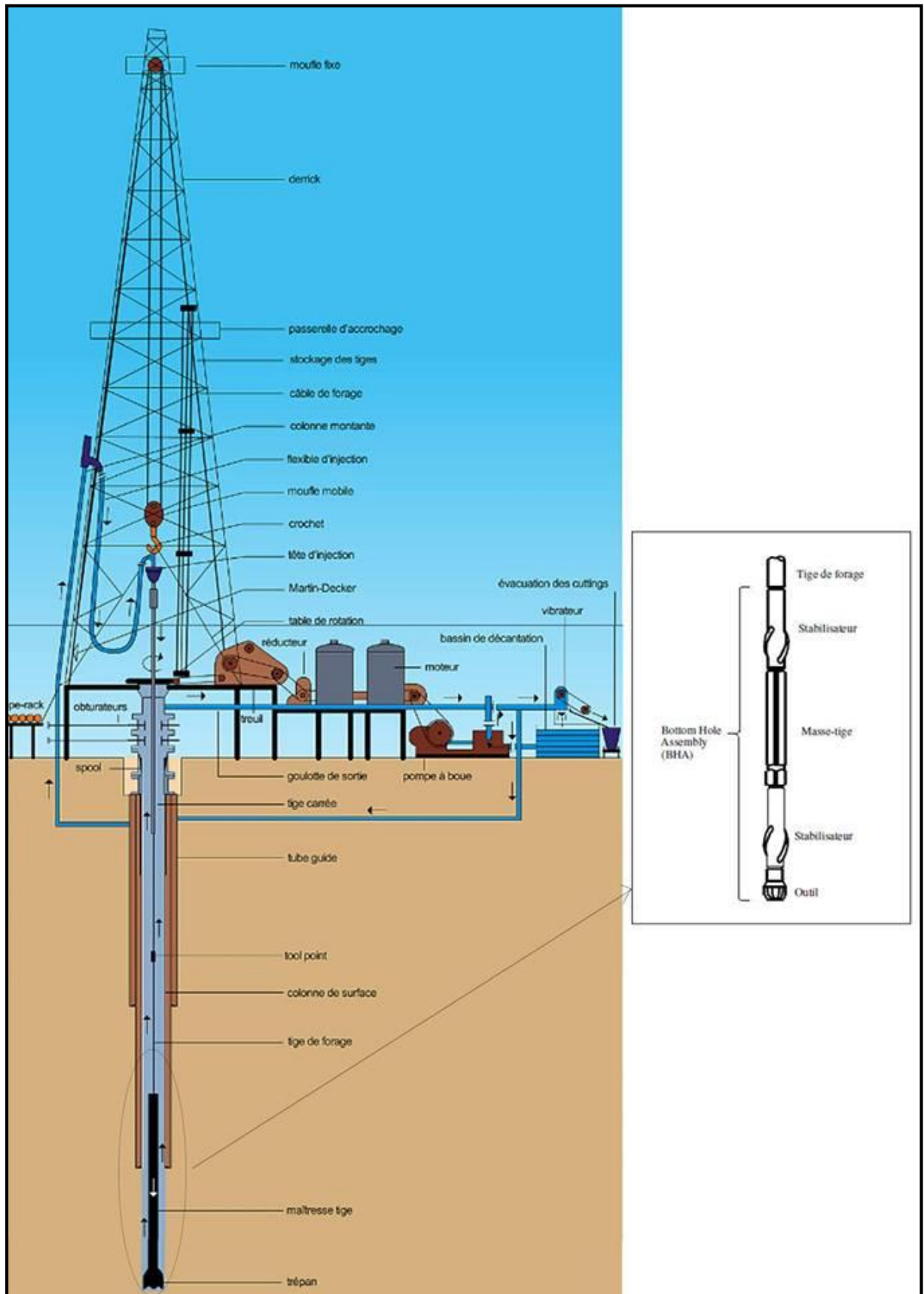


Figure I.1 : Système de forage

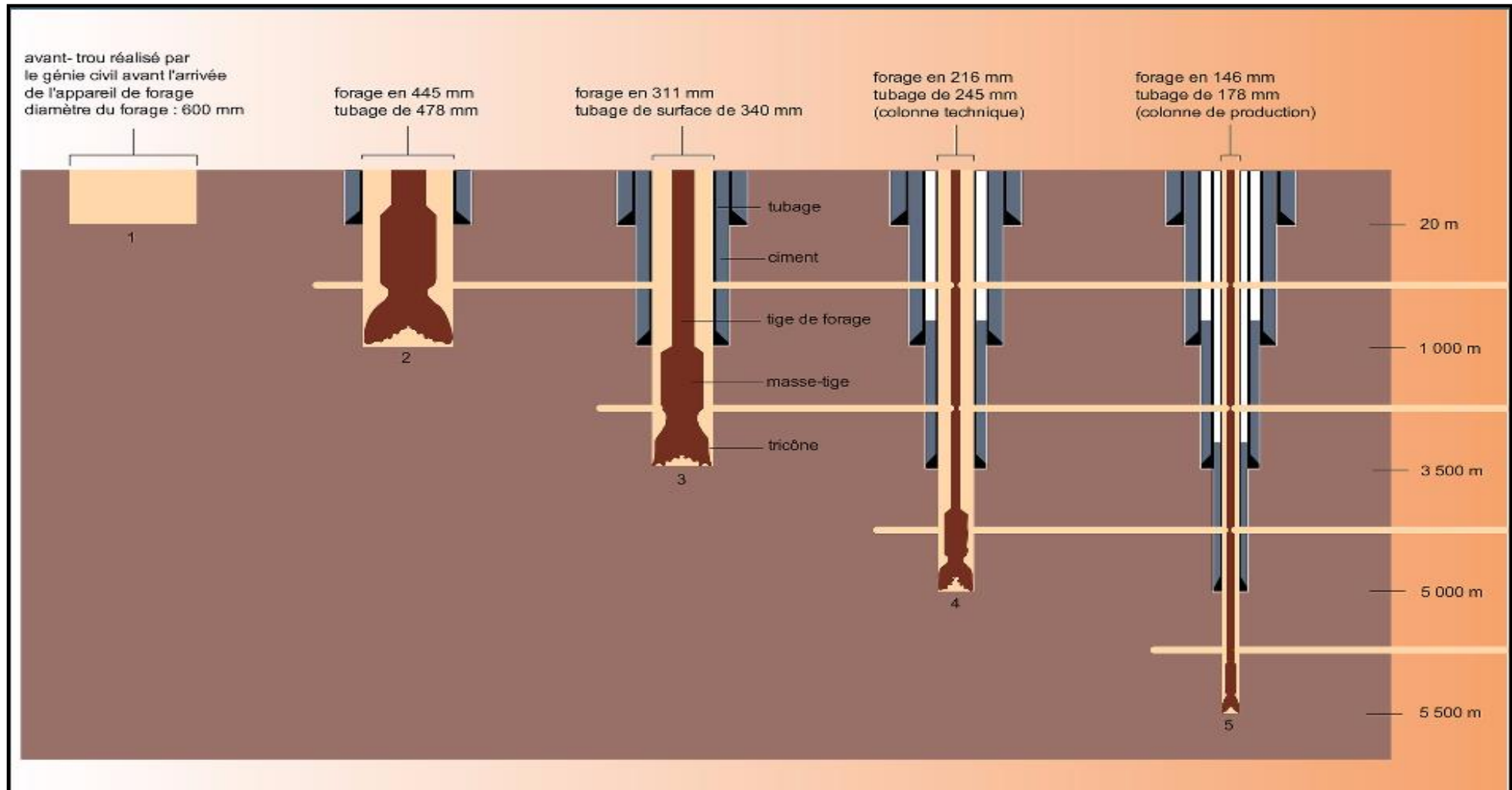


Figure I.2: Le processus de forage

## 2- Principe de forage rotary

La technique du forage rotary consiste à utiliser un outil, permettant de détruire la roche sous l'effet du poids et de la rotation. Le poids est assuré par un ensemble de tiges lourdes et creuses, assemblées en un train, qui achemine sous pression les boues de forage. Celles-ci refroidissent l'outil et assurent le déblaiement du trou. Autour des parois du forage, des tubes sont descendus et du ciment est injecté afin de garantir la tenue des parois du puits, ainsi qu'une protection contre la corrosion mais aussi la protection des nappes d'eau souterraines et leur isolation thermique.[1]

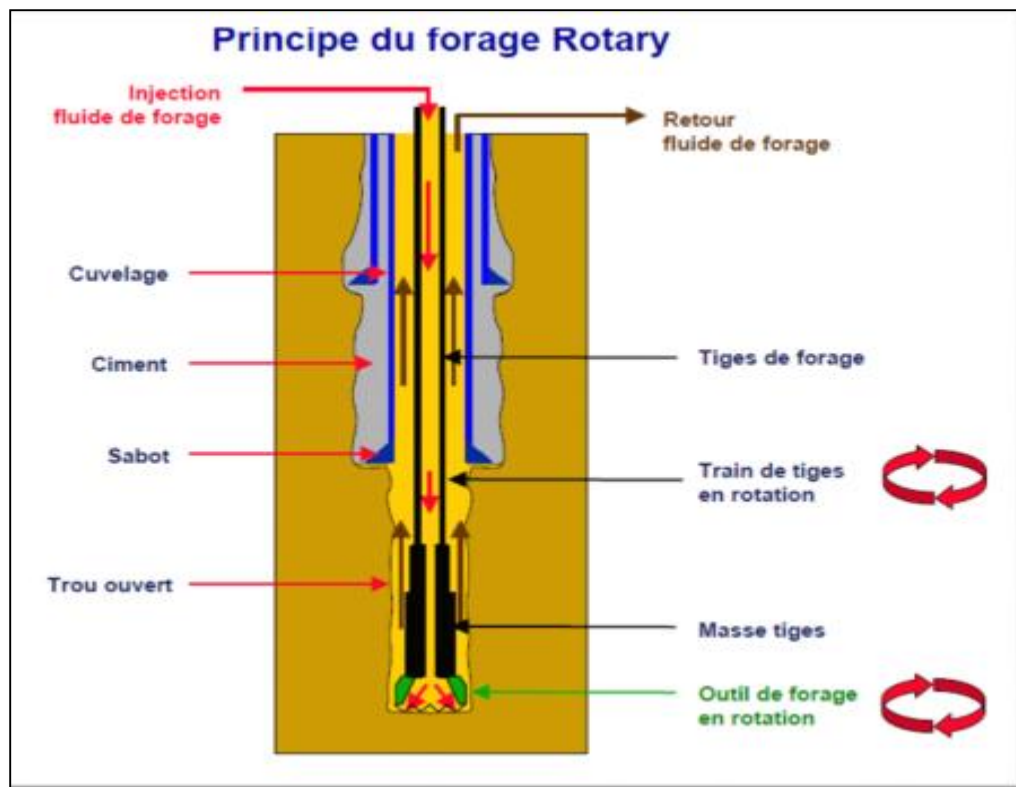


Figure I.3 : Principe du forage Rotary

## 3- Description généralement de forage

Un appareil de forage est constitué d'un ensemble d'équipements, permettant les fonctions de levage, rotation et pompage afin de réaliser un sondage, ainsi que des équipements de sécurité, notamment ceux de mise en sécurité du puits en cas de venue . [1]

La classification d'un appareil de forage se fait en premier lieu par une capacité ou bien une profondeur de forage maximale :

- **Appareil léger** : pour des profondeurs inférieure à 1200 m.
- **Appareil moyen** : pour des profondeurs comprises entre 1200-2500 m.
- **Appareil lourd** : pour des profondeurs comprises entre 2500-4000 m.
- **Appareil ultra- lourd** : pour des profondeur supérieur à 4000 m.

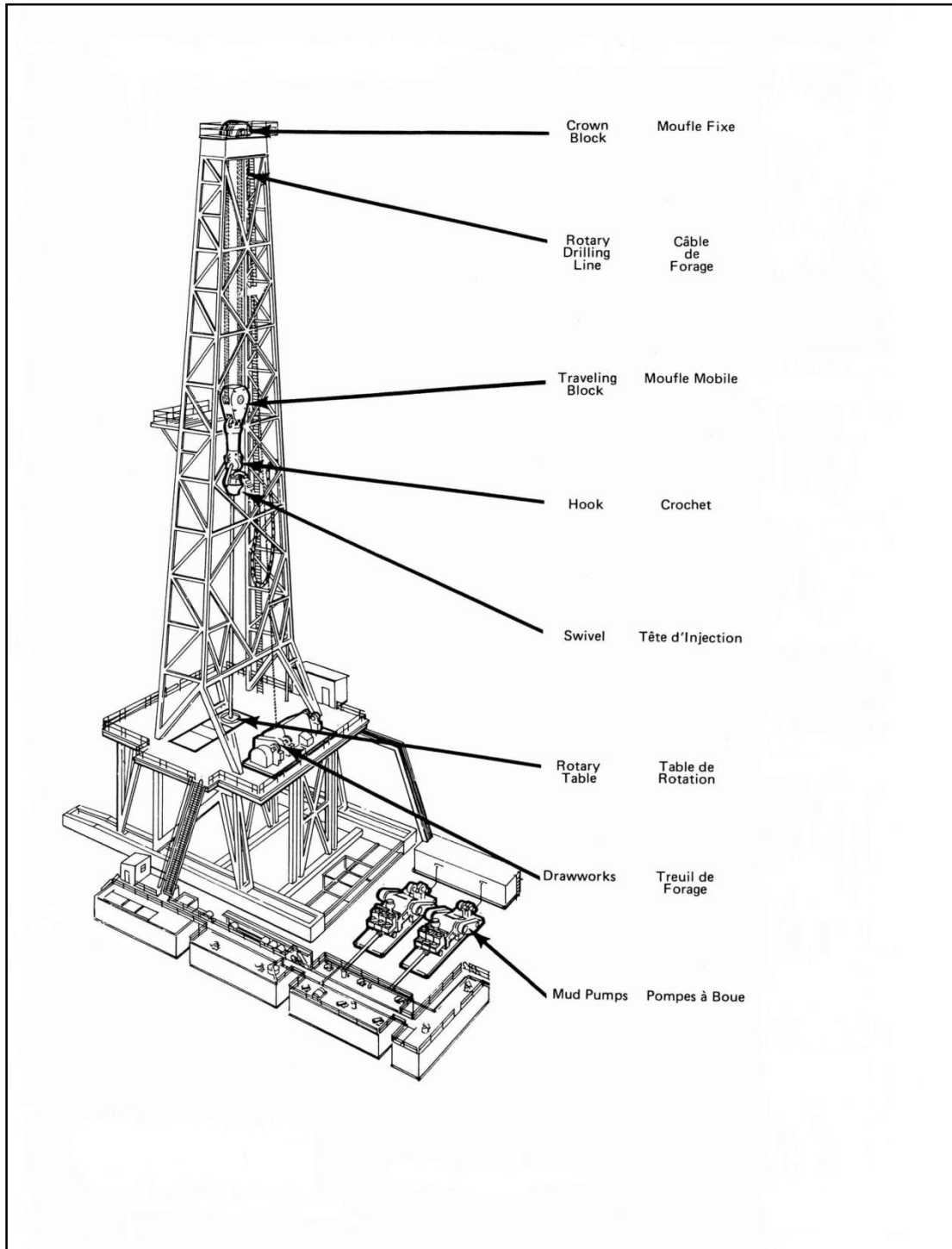


Figure I.4: Sonde de forage rotary

## 4- Equipements de levage

Ils servent à suspendre la garniture de forage et la faire monter et descendre dans le puits, ils permettent aussi d'ajuster le poids sur le trépan qui est nécessaire pour que celui-ci puisse forer. [1]

**4-1. Le mat de forage :** c'est le « trépied » qui supporte le palan. Il a remplacé la tour pour la rapidité de son montage et démontage. [1]

À son sommet est placé le moufle fixe. Une passerelle d'accrochage est placée à son milieu ; elle sert de lieu de travail pour l'accrocheur, qui accroche ou décroche les « longueurs » de tiges lors de la remontée ou la descente de l'outil dans le puits. Une autre passerelle de hauteur ajustable, placée plus bas, sert à guider le tubage pour le visser et le descendre dans le puits. Un plancher de travail est aménagé aux pieds du mât. Il sert d'aire de travail pour l'équipe. Une cabine est aménagée sur ce plancher pour permettre aux ouvriers de se reposer. [1]

**4-2. La substructure :** le plancher est surélevé de quelques mètres au-dessus du sol, pour permettre l'introduction des éléments de la tête de puits et des obturateurs. [1]

**4-3. Le mouflage :** est l'enroulement du câble de forage entre les poulies des moufles fixe et mobile en plusieurs brins (jusqu'à 14 brins). Le mouflage permet de démultiplier le poids de la garniture de forage et diminuer la vitesse de son déplacement. [1]

En négligeant les frottements, la tension sur le brin actif est divisée par le nombre des brins, et sa vitesse est multipliée par ce nombre. [1]

**4-4. Le moufle fixe :** est formé d'un certain nombre de poulies et placé au sommet du mât, il possède une poulie de plus que le moufle mobile. [1]

**4-5. Le moufle mobile :** est formé également d'un certain nombre de poulies par lesquelles passe le câble de forage, il se déplace sur une certaine hauteur entre le plancher de travail et le moufle fixe. [1]

**4-6. Le crochet :** sert à la suspension de la garniture pendant le forage. Des bras sont accrochés de part et d'autre de ce crochet servent à supporter l'élévateur, utilisé pour la manœuvre de la garniture. [1]

**4-7. Le câble de forage :** les câbles des appareils de forage sont ensemble de qualité métallique sur laquelle sont entourés de six torons constitués en fil d'acier. [1]

**4-8. Le treuil :** c'est un fût (tambour) autour duquel s'enroule le câble de forage. Aux bouts de ce fût sont fixées des jantes qui servent au freinage, au contact de bandes en acier contenant des patins en ferodo, actionnées par un levier. Le treuil est le cœur de l'appareil de forage, donc c'est la capacité du treuil qui caractérise un appareil de forage et indique la classe de profondeur des forages que l'on pourra effectuer. [1]

## **5- Garniture de forage**

C'est cet élément qui nous intéresse le plus dans notre étude et nous tenterons de modéliser son comportement statique à l'intérieur du puits de forage. La connaissance des différents constituants de la garniture s'avère donc nécessaire pour la modélisation. [1]

De manière générale, une garniture se compose, de bas en haut, des éléments suivants :

**5-1. L'outil de forage :** Extrémité basse de la garniture, l'outil est entraîné en rotation et permet la progression de la garniture par son action directe sur le front de taille. On distingue deux types d'outils, à molettes (tricônes) et monoblocs (diamant, PDC). [1]



**Figure I. 5: L'outil de forage**

### 5-2. Stabilisateurs :

Ils ont un diamètre presque égal à celui de l'outil. Intercalés entre les masse-tiges, ils les maintiennent droites dans le puits et évitent leur flexion et la déviation du puits. [1]

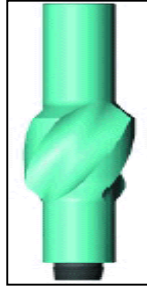


Figure I.6 : Stabilisateur

**5-3. Les masses-tiges ( Drill Collars , DC) :** il s'agit d'une série de tubes massifs en acier dont le rôle est d'assurer une poussée suffisante sur l'outil. Par conséquent, pendant la phase de forage, le train de masses-tiges est en partie, voire en totalité, en état de compression. [1]



Figure I.7 : Quelques éléments de la garniture de forage

### 5-4. Les trains de tiges (Drill Pipes, DP)

c'est la partie la plus longue de la garniture ; elle est constituée d'une série importante de tubes minces d'acier qui s'étendent jusqu'à la surface et dont le rôle est de transmettre le couple vers l'outil et de soutenir plus ou moins le train de masses tiges. Chaque tige est munie à ses extrémités de renflements appelés *tool-joint* (TJ) (figure 1) servant pour le vissage des tiges entre elles et comme points d'appui du train de tiges sur la paroi de puits. En conditions normales, on s'arrange pour que le train de tiges soit en traction afin d'éviter son flambement car celui-ci favorise l'apparition de forces de contacts très élevées entre tiges et parfois de puits. Ce train de tiges comporte parfois (cas du forage dirigé ou du forage horizontal) une série supplémentaire de tiges lourdes (*Heavy weight Drill Pipe, HWDP*) plus épaisses placées

souvent juste après les masses tiges. Ces tiges lourdes ont pour but de réduire les risques de flambement des tiges. [1]

## 6- Paramètres de forage

**Définition:** Les paramètres de forage consistent en plusieurs grandeurs physiques mesurées et enregistrées pendant le forage à l'aide de capteurs placés un peu partout dans le RIG et dont les variations dépendent du terrain traversé, ces paramètres peuvent être classés en deux catégories mécanique et hydraulique .[2]

**6-1. Les paramètres mécaniques:** Selon la dureté des couches géologiques traversées la valeur de ces paramètres mécaniques diffère.

- ❖ **Poids sur l'outil (WOB) :** Le poids sur l'outil « Weight on Bit » représente le poids appliqué sur le trépan, calculé selon la formule suivante :

$$\text{WOB} = \text{WOH L} - \text{WOH F}$$

Ou :

WOH L : C'est le poids de la garniture

WOH F : C'est le poids de la garniture durant le forage.

Ce poids croit de plus en plus dans les formations les plus dures et est de l'ordre de 250 à 1 000 daN par centimètre de diamètre de trépan par exemple un trépan pour terrains durs de 311 mm de diamètre pourra être 30 000 daN. Cette augmentation de poids résulte sur la diminution de la durée de vie du trépan .[2]

- ❖ **Vitesse de rotation (RPM) :** La vitesse de rotation « Révolution per minute » elle indique le nombre de tour par minute de la table de rotation sur le rig-floor, est en général comprise entre 50 et 300 tours par minute selon la dureté de la formation, par exemple pour le trépan cité précédemment pour terrains durs de 311 mm de diamètre la vitesse de rotation pourra être utilisée à 100 tours par minute. Elle est déduite via un capteur placé à la table de rotation.[2]
- ❖ **Torque :** Le torque n'est pas un paramètre à optimiser, c'est juste la force opposée de la rotation, il est principalement fonction du poids sur l'outil (WOB), la vitesse de rotation (RPM), et la dureté de la formation traversée. La couple en surface n'est pas transmis intégralement sur l'outil de forage, mais la mesure de surface est la seule possible actuellement .[2]

❖ **Vitesse d'avancement (ROP)** : La vitesse d'avancement des trépan est fonction de la nature des terrains et varie de 1 à 2 m par heure dans les terrains durs à 30 m par heure et plus dans les terrains tendres. La durée de vie des tricônes à molettes dentées est de l'ordre de 40 heures ; le travail du trépan a pour conséquences: une usure des dents qui peut aller jusqu'à disparition complète. Une usure des roulements qui peut causer la perte d'une molette au fond du trou ; une perte de diamètre dans les terrains abrasifs. Pour améliorer la durée de vie des tricônes, les paliers des molettes sont lubrifiés et éventuellement lisses. Dans ce cas, la durée de vie de l'outil peut dépasser 150 heures .[2]

## 6-2. Les paramètres hydraulique

- ❖ **Le débit (Flow in)** : C'est le volume de boue déversé par les pompes de forage à l'intérieur du trou de forage, la quantification du débit est basée sur le débit unitaire des pompes (POP) selon le nombre de coup par minute contribué par les pompes (SPM), le SPM est enregistré à l'aide d'un capteur de fréquence similaire à celle de la rotation. L'unité du débit est le L/mn .[2]
- ❖ **Type de boue** : Le type de boue est choisi en fonction des performances recherchées et désigne les propriétés physico-chimiques du fluide de forage. Trois types de boues sont souvent employés : la boue à base d'eau (Water Based Mud, WBM), la boue à base d'huile (Oil Based Mud, OBM) et la boue synthétique (Synthetic Based Mud, SBM). Une boue synthétique est constituée d'un mélange d'eau et d'additifs chimiques.[2]
- ❖ **Densité de la boue**: L'obtention des informations relatives au puits et particulièrement le contrôle de la pression dans le puits s'effectue à travers la densité de la boue. La boue de forage ramène à la surface les déblais, mais aussi du gaz contenu dans les roches. Cela fournit des indications sur la nature des fluides se situant dans le réservoir et représente un élément important dans le pilotage de la garniture.[2]
- ❖ **Pression hydraulique** : Le débit et la pression hydraulique représentent les variables physiques qui doivent favoriser une bonne évacuation des déblais et éviter des problèmes d'encrassement du trépan ou du puits.[2]

## **7- Types de problèmes de forage**

Lors de la planification, 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. Nous présentons les problèmes les plus courant a savoir. [3]

### **7-1. 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.[2]

### **7-2. 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 .[2]

### **7-3. Venue (Kick) :**

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.[2]

#### 7-4. Coincements de la garniture de forage (Stuck Pipe) :

**Définition :** Le coincement de la garniture de forage représente 40% du cout total de forage d'un puits et indique le faite que la garniture soit prise au piège dans le puits, on ne peut plus ni avancer ni la retirer du puits ceci peut avoir plusieurs cause selon l'état de la garniture qu'elle soit statique ou en mouvement, ces coincement sont classé en deux catégorie [2] :

- les coincements mécanique.
- les coincements différentiel.

##### A- Les coincements mécanique

Le coincement mécanique se produit lorsque la garniture est en mouvement dans le puits et est causé par des obstructions ou restrictions physiques. Cette catégorie est à son tour divisée en deux sous-groupes [2]:

- ✓ Coincement liés à l'instabilité du puits ou aux déblais dans le puits.
- ✓ Coincement liés aux conditions géométriques du puits.

Les causes majeures des coincements mécaniques sont l'instabilité de la formation et le mauvais nettoyage du puits, l'instabilité de la formation est liée à des couches de schiste due à un gonflement et élargissements du puits résultant du défaut de compression en raison de la pression du puits trop faible, un nettoyage adéquat du puits d'un autre coté est une partie essentielle des opérations de forage, si les déblais ne sont pas proprement retirés du puits ils risquent de s'installer autour de la garniture de forage et causer le coincement, le risque du mauvais nettoyage du puits augmente dans les puits déviés, les puits ayant un inclinaison de 30 à 60 degré sont les pires à nettoyer .[2]

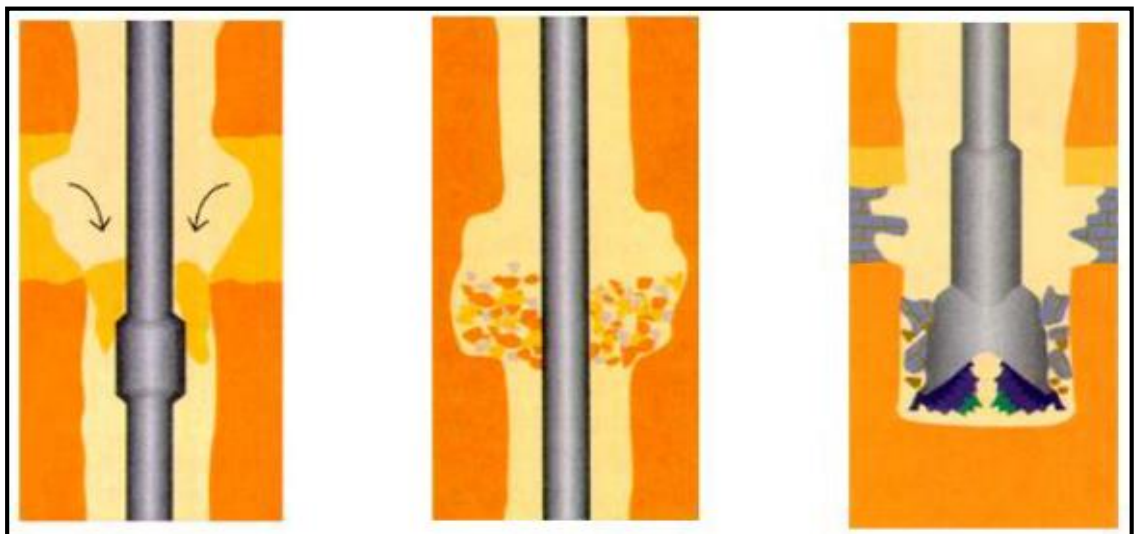
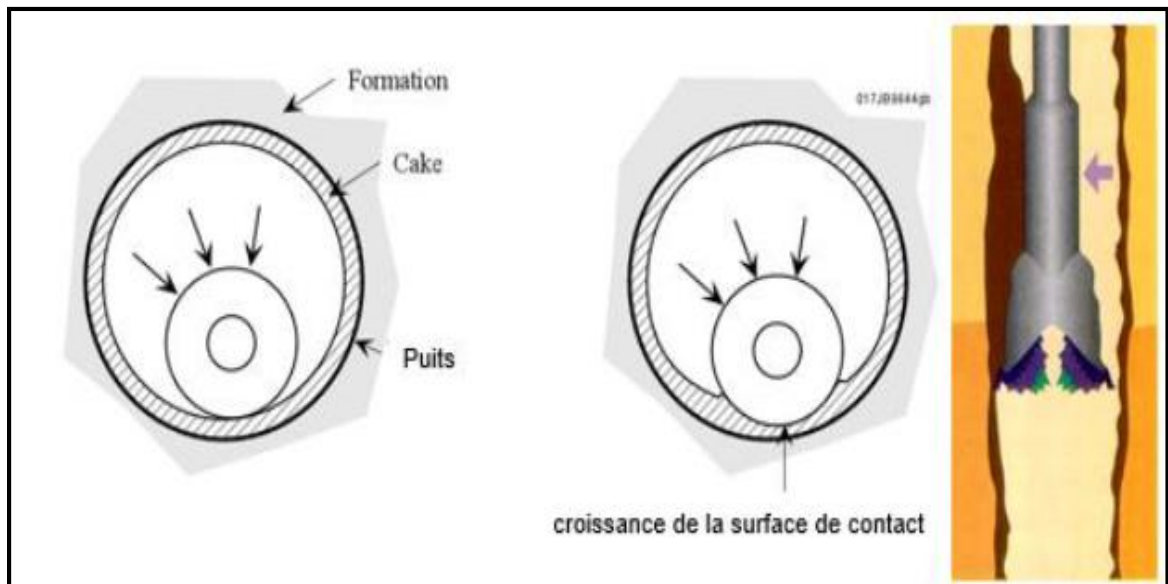


Figure I .8: Exemple de Stuck mécanique

## B- Les coincements différentiel

les coincements différentiels due à une force causée par la pression surpassée de la boue qui agit sur la garniture de forage et la coince dans une formation perméable. La partie de la garniture de forage coincée dans ce qu'on appelle un gâteau de boue « Mud cake » à une pression égale à celle de la formation qui agit sur elle, alors que l'autre partie de la garniture à une pression égale à la pression hydrostatique de la boue de forage. Lorsque cette pression hydrostatique ( $P_h$ ) dans le puits est supérieure à celle de la formation ( $P_f$ ) elle engendre une force qui colle la garniture sur les parois du puits .[2]

Ce type de coincement ne se produit pas dans les schistes et autre formation non perméable où les gâteaux de boue ne se forment pas habituellement, les coincements différentiels se produisent généralement quand la garniture ou l'outil de forage est stationnaire ou en mouvement mais lentement .[2]



**FigureI.9: Exemple de Stuck différentiel**

Une fois la garniture coincée tous les efforts devrais être fait pour la libérer le plus rapidement possible car la probabilité de la libérer diminue rapidement avec le temps, l'identification prématurée de la cause la plus probable d'un problème de coincement est cruciale, puisque chaque cause doit être remédiée avec des mesures différentes. Une réaction incorrecte à ce problème pourrait facilement le rendre plus graves. Une évaluation des événements menant jusqu'au coincement indique fréquemment la cause la plus probable et peut mener aux mesures correctives appropriées.[2]

**Conclusion**

Nous avons vu durant ce chapitre les principales notions de base du domaine de rétre étude . 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 étude et réalisation, nous allons tenter de réaliser un système capable de prévenir à l'avance certains de ces cas de coincements.

*Chapitre II :*

---

*Data-Mining*

## Introduction

Dans le chapitre précédent, nous avons vu les principales notions de forage pétrolier ,les paramètres et le problème "le coincement de garniture de forage ".

Durant ce chapitre, nous allons présent la technique de data-mining, les méthodes de classification (réseaux de neurone -RN-,Le machine à support vecteurs SVM, k-plus proche voisin K-PPV ) qui sont utilisées pour résoudre le problème de ce travail.

### 1- Qu'est ce que la fouille de données (Data-mining)?

Le data-mining est un domaine interdisciplinaire utilisant dans le même temps des techniques d'apprentissage automatiques, de reconnaissance des formes, des statistiques, des bases de données et de visualisation pour déterminer les manières d'extraction des informations de très grandes bases de données. [4]

Aussi qui, est un processus inductif, itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, utiles et compréhensibles dans de larges Bases de Données .[5]

### 2- Le processus de data-mining

le processus de data-mining est 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 .[6]

❑ Plus précisément, le datamining peut se décomposer en 7 étapes [6]:

- ❖ **le nettoyage des données** : qui gère les données bruitées, erronées, manquantes ou non pertinentes.
- ❖ **l'intégration de données** : où de multiples sources de données hétérogènes, peuvent être intégrés dans un seul.
- ❖ **sélection de données** : où les données relatives à la tâche d'analyse sont récupérées à partir de bases de données.
- ❖ **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.
- ❖ **l'exploration de données** : qui est processus essentiel où les méthodes intelligentes sont appliquées afin d'en extraire les modèles de données.

- ❖ **l'évaluation de modèle** : ce qui est d'identifier les motifs vraiment intéressants représentant les connaissances basées sur des mesures de qualité.
- ❖ **présentation de la connaissance** : où les techniques de visualisation et de représentation des connaissances sont utilisées pour présenter les connaissances extraites à l'utilisateur.

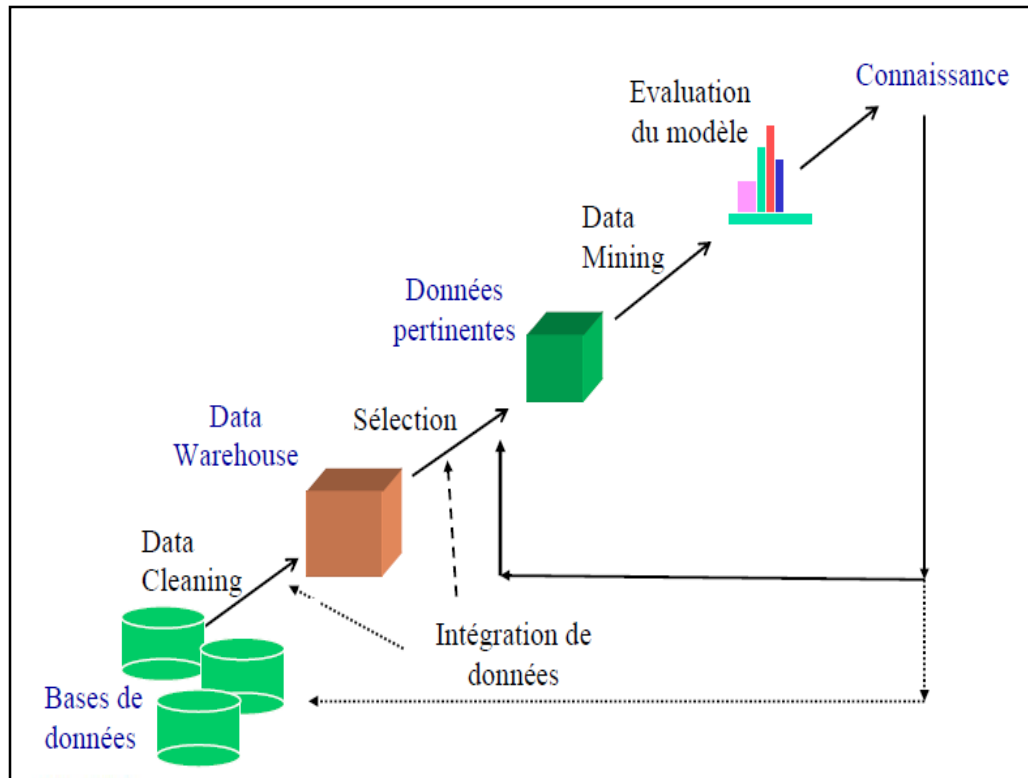


Figure II. 1: Processus de data-mining

### 3- Les taches de data-mining

Beaucoup de problèmes intellectuels, économiques ou même commerciaux peuvent être exprimés en termes des six tâches suivantes [7]:

- ❖ La classification.
- ❖ L'estimation.
- ❖ La prédiction.
- ❖ Le groupement par similitude.
- ❖ L'analyse des clusters.
- ❖ La description .

Les trois premières tâches sont des exemples du data-mining supervisé dont le but est d'utiliser les données disponibles pour créer un modèle décrivant une variable particulière prise comme but en termes de ces données. Le groupement par similitude et l'analyse des clusters sont des tâches non-supervisées où le but est d'établir un certain rapport entre toutes les variables [8].

La description appartient à ces deux catégories de tâche, elle est vue comme une tâche supervisée et non-supervisée en même temps [7].

### **3-1. La classification**

La classification est la tâche la plus commune du data-mining et qui semble être une obligation humaine. Afin de comprendre notre vie quotidienne, nous sommes constamment classifiés, catégorisés et évalués [7].

La classification consiste à étudier les caractéristiques d'un nouvel objet pour lui attribuer une classe prédéfinie. Les objets à classifiés sont généralement des enregistrements d'une base de données, la classification consiste à mettre à jour chaque enregistrement en déterminant un champ de classe. La tâche de classification est caractérisée par une définition de classes bien précise et un ensemble d'exemples classés auparavant. L'objectif est de créer un modèle qui peut être appliqué aux données non classifiées dans le but de les classifiées [8].

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.[8]

### **3-2. L'estimation**

L'estimation est similaire à la classification à part que la variable de sortie est numérique plutôt que catégorique. En fonction des autres champs de l'enregistrement l'estimation consiste à compléter une valeur manquante dans un champ particulier. Par exemple on cherche à estimer La lecture de tension systolique d'un patient dans un hôpital, en se basant sur l'âge du patient, son genre, son indice de masse corporelle et le niveau de sodium dans son sang. La relation entre la tension systolique et les autres données vont fournir un modèle d'estimation. Et par la suite nous pouvons appliquer ce modèle dans d'autres cas [7] [9].

### **3-3. La prédiction**

La prédiction est la même que la classification et l'estimation, à part que dans la prédiction les enregistrements sont classés suivant des critères (ou des valeurs) prédites (estimées). La principale raison qui différencie la prédiction de la classification et l'estimation est que dans la création du modèle prédictif on prend en charge la relation temporelle entre les variables d'entrée et les variables de sortie . [7]

## **4- Techniques du data-mining**

Pour effectuer les tâches du data-mining il existe plusieurs techniques issues de disciplines scientifiques diverses (statistiques, intelligence artificielle, base de données) afin de faire apparaître des corrélations cachées dans des gisements de données pour construire des modèles à partir de ces données. Dans cette adresse, nous proposons des techniques d'extraction de données qui sont utilisées dans notre étude . [6]

### **4-1. Les réseaux de neurones**

Un réseau de neurones est un modèle de calcul dont le fonctionnement vise à simuler le fonctionnement des neurones biologiques, il est constitué d'un grand nombre d'unités (neurones) ayant chacune une petite mémoire locale et interconnectées par des canaux de communication qui transportent des données numériques. Ces unités peuvent uniquement agir sur leurs données locales et sur les entrées qu'elles reçoivent par leurs connections. Les réseaux de neurones sont capables de prédire de nouvelles observations (sur des variables spécifiques) à partir d'autres observations (soit les mêmes ou d'autres variables) après avoir exécuté un processus d'apprentissage sur des données existantes.[10]

La phase d'apprentissage d'un réseau de neurones est un processus itératif permettant de régler les poids du réseau pour optimiser la prédiction des échantillons de données sur lesquelles l'apprentissage a été fait. Après la phase d'apprentissage le réseau de neurones devient capable de généraliser .[10]

#### **4-1-1. Avantages et inconvénients**

- **Avantages**

Les réseaux de neurones sont théoriquement capables d'approximer n'importe quelle fonction continue et ainsi le chercheur n'a pas besoin d'avoir aucune hypothèse du modèle sous-jacent. [10]

**Inconvénients**

- Généralement les réseaux de neurones ne sont pas souvent utilisées dans les tâches du data-mining parce qu'ils produisent des modèles souvent incompréhensibles et demande un longtemps d'apprentissage .[11]

**4-2. L'algorithme des k-Plus proches voisins**

L'algorithme des k plus proches voisins (K-PPV, k nearest neighbor en anglais ou KNN) est un algorithme de raisonnement à partir de cas qui est dédié à la classification qui peut être étendu à des tâches d'estimation. Le but de cet algorithme est de prendre des décisions en se basant sur un ou plusieurs cas similaires déjà résolus en mémoire. Dans ce cadre, et Contrairement aux autres méthodes de classification (arbres de décision, réseaux de neurones, algorithmes génétiques, ...etc.) l'algorithme de KNN ne construit pas de modèle à partir d'un échantillon d'apprentissage, mais c'est l'échantillon d'apprentissage, la fonction de distance et la fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constituent le modèle .[14]

**4-2-1. Comment cela marche-t-il ?**

Nous supposons avoir une base de données d'apprentissage constituée de  $N$  couples «entrée-sortie». Pour estimer la valeur de sortie d'une nouvelle entrée  $x$ , la méthode des  $k$  plus proches voisins consiste à prendre en compte (de façon identique) les  $k$  échantillons d'apprentissage dont l'entrée est la plus proche de la nouvelle entrée  $x$ , selon une distance à définir.[12]

Si nous prenons une base d'apprentissage de 100 éléments, Dès que nous recevons un nouvel élément que nous souhaitons classifier, l'algorithme calcule sa distance à tous les éléments de la base. Si cette base comporte 100 éléments, alors il va calculer 100 distances et donc obtenir 100 nombres réels. Si  $k = 25$  par exemple, il cherche alors les 25 plus petits nombres parmi ces 100 nombres qui correspondent donc aux 25 éléments de la base qui sont les plus proches de l'élément que nous souhaitons classifier. La classe attribuée à l'élément à classifier est la classe majoritaire parmi ces 25 éléments .[13]

**4.2.2 Avantages et inconvénients****· Avantages**

- La qualité de la méthode s'améliore en introduisant de nouvelles données sans nécessiter la reconstruction d'un modèle. Ce qui représente une différence majeure avec des méthodes telles que les arbres de décision et les réseaux de neurones.

- La clarté des résultats : la classe attribuée à un objet peut être expliquée en exhibant les plus proches voisins qui ont amené à ce choix.
- La méthode peut s'appliquer à tout type de données même les données complexes tels que des informations géographiques, des textes, des images, du son. C'est parfois un critère de choix de la méthode K-PPV car les autres méthodes traitent difficilement les données complexes. Nous pouvons noter, également, que la méthode est robuste au bruit.
- Facile à mettre en œuvre. [14]

### Inconvénients

- temps de classification : la méthode ne nécessite pas d'apprentissage ce qui implique que tous les calculs sont effectués lors de la classification. Contrairement aux autres méthodes qui nécessitent un apprentissage (éventuellement long) mais qui sont rapides en classification.
- méthode donnera de mauvais résultats Si le nombre d'attributs pertinents est faible relativement au nombre total d'attributs, car la proximité sur les attributs pertinents sera noyée par les distances sur les attributs non pertinents.
- Les performances de la méthode dépendent du choix de la distance, du nombre de voisins et du mode de combinaison des réponses des voisins.[15]

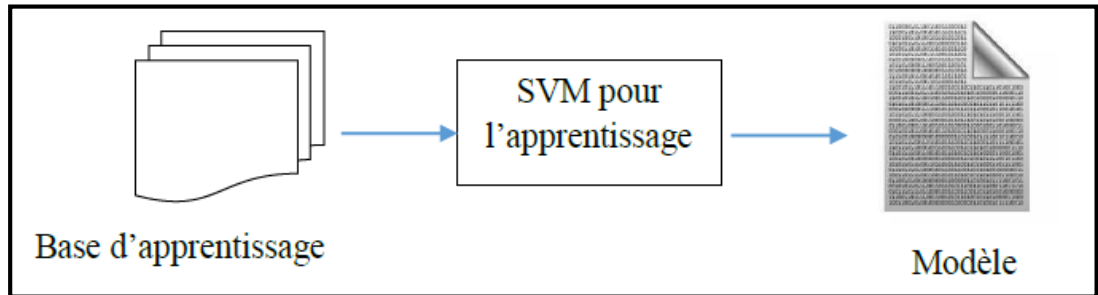
### 4-3. Les machines à support vecteurs (SVM)

**Définition :** La machine à support vecteur (*support vector machine SVM*) est une méthode très efficace pour l'apprentissage automatique, elle est basée sur la théorie statistique de l'apprentissage. Les SVMs sont conçus pour résoudre les problèmes de classification et de régression, elles évitent le processus traditionnel de l'induction de la déduction, afin de simplifier les problèmes de la classification ainsi les problèmes de la régression.[16]

#### □ Principe des SVM

Le principe des SVM est basé sur deux étapes la première pour l'apprentissage et la deuxième pour l'utilisation tel que les méthodes traditionnelles d'aide à la décision (K-means, les algorithmes génétiques, data-mining...), cela a pour but de trouver un modèle performant afin de donner des résultats efficaces.[17]

- ✓ Phase d'apprentissage et validation du modèle : l'apprentissage suit le principe du schéma représenté dans la figure suivante :



**Figure II.2: phase d'apprentissage pour les SVM**

A l'entrée de l'SVM nous avons besoin de présenter un ensemble de données sous forme des fichiers, ces derniers contenant les caractéristiques de phénomène à traiter, dans le cas de notre étude le coincement de garniture du forage. Ces données doivent être normalisées pour que la méthode fonctionne bien. [17]

- ❑ Les paramètres des SVM sont [17] :
  - le type de traitement, soit régression ou classification.
  - le type de noyau utilisé (RBF, Sigmoïde, Linéaire, Polynomial...).
  - Un constant C.
  - Gamma, epsilon ...etc. sont des paramètres choisis selon le type de noyau.

Pour valider le modèle obtenu nous avons besoin de faire des essais plusieurs fois sur la base de test afin de minimiser l'erreur (**Figure II.3**), et pour calculer l'erreur il faut choisir une équation parmi les membres qui existent. Dans notre travail, on a choisi le pourcentage de l'erreur moyenne absolue et l'erreur quadratique qui appartiennent à l'analyse de l'erreur absolue [40], qui sont définies comme suit.[17] :

- ✓ Erreur quadratique moyenne :

$$E = \sqrt{1/N \sum_{t=1}^N [X_t - \hat{X}_t]^2}$$

N : Les nombre des instances du test

X<sub>t</sub> : Représente l'observation

X<sup>^</sup><sub>t</sub> : Représente la prévision

- ✓ Pourcentage de l'erreur moyenne absolue :

$$E = \left( \frac{\text{valeur approchée} - \text{valeur réelle}}{\text{valeur réelle}} \right) \times 100$$

- ✓ L'erreur absolue peut prendre des valeurs négatifs, si la valeur approchée moins que la valeur réelle.

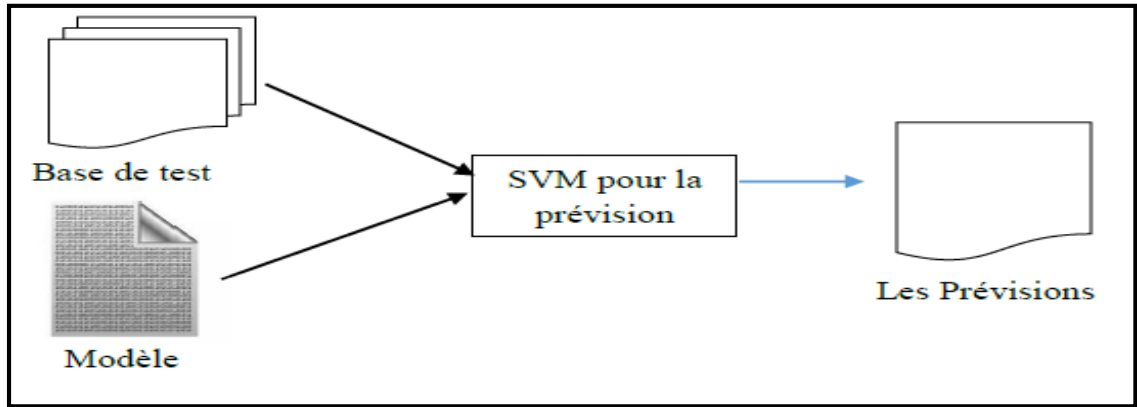


Figure II.3 : Etape de test du modèle.

La base de test contient une partie de la base de données pour comparer les données avec les résultats obtenus et de calculer les erreurs afin de trouver un modèle valide. Le processus général de l’SVM suit le schéma présenté dans la figure suivante (Figure II.4) [17]:

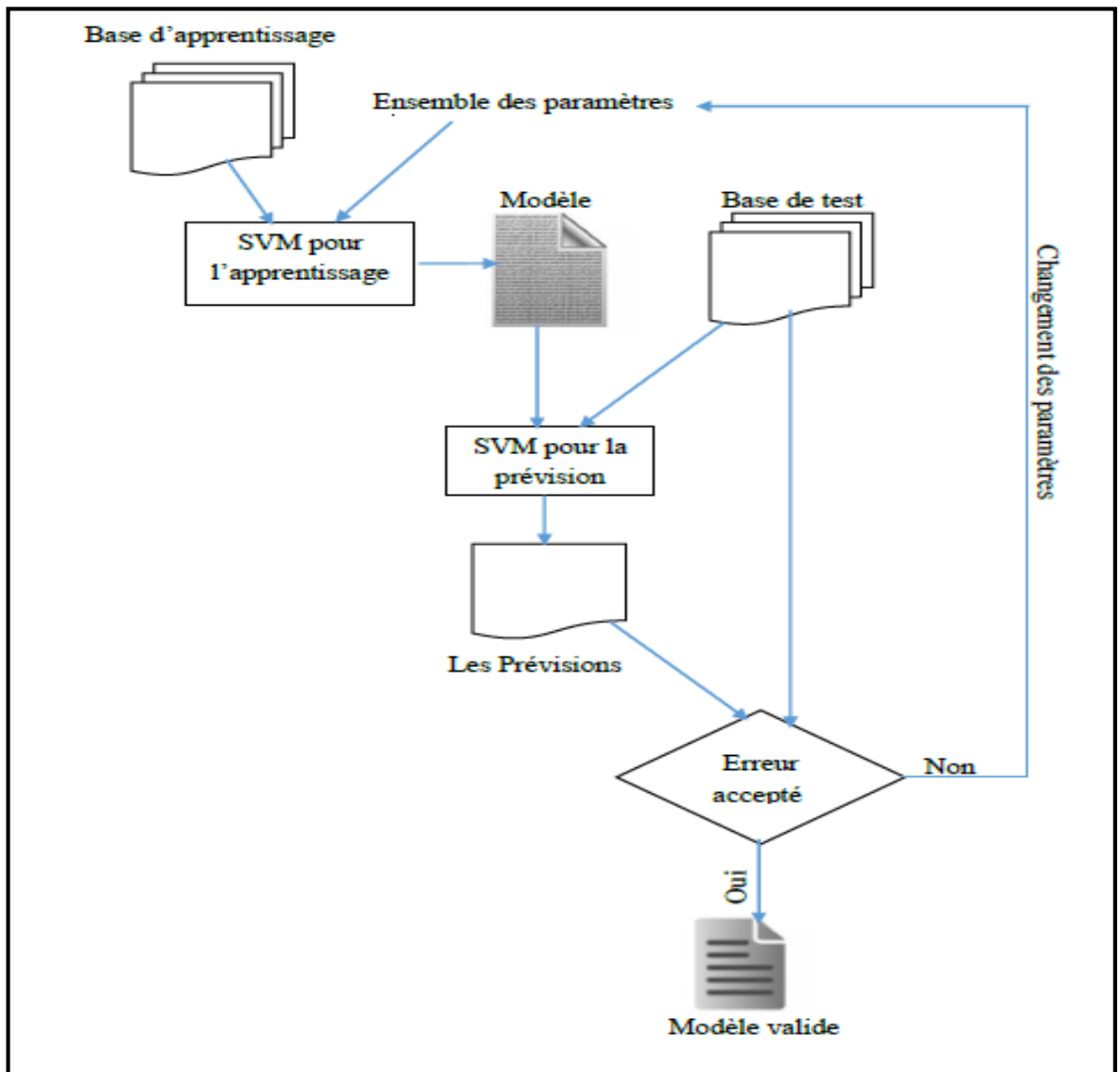


Figure II. 4 : Schéma globale de fonctionnement de l’SVM.

## **Conclusion**

Dans ce chapitre nous avons présenté, un aperçu sur les techniques de fouille de données, le processus et les tâches (classification, régression ...) de fouille de données, et leur différents algorithmes (RN,SVM,K-PPV) que nous avons utilisé, afin de résoudre le problème de notre étude.

Cette première étape sert comme une base permettant d'entamer la prochaine phase de l'étude conceptuelle.

***Chapitre III:***

---

***Conception***

## Introduction

Dans les chapitres précédent nous avons présentés les déférentes problèmes que nous rencontrons pendant le forage, et ensuit une explication détaillée de certaines des techniques utilisées pour détecter cette problèmes (RNA, SVM,K-PPV).

Dans ce chapitre nous allons proposer une conception par affinement successif du système en donnant son architecture générale, puis nous détaillons en étudiant séparément chacun de ses composants, Ensuite, nous présentons les résultats obtenus.

## 1- Description du processus de prévision des coincements

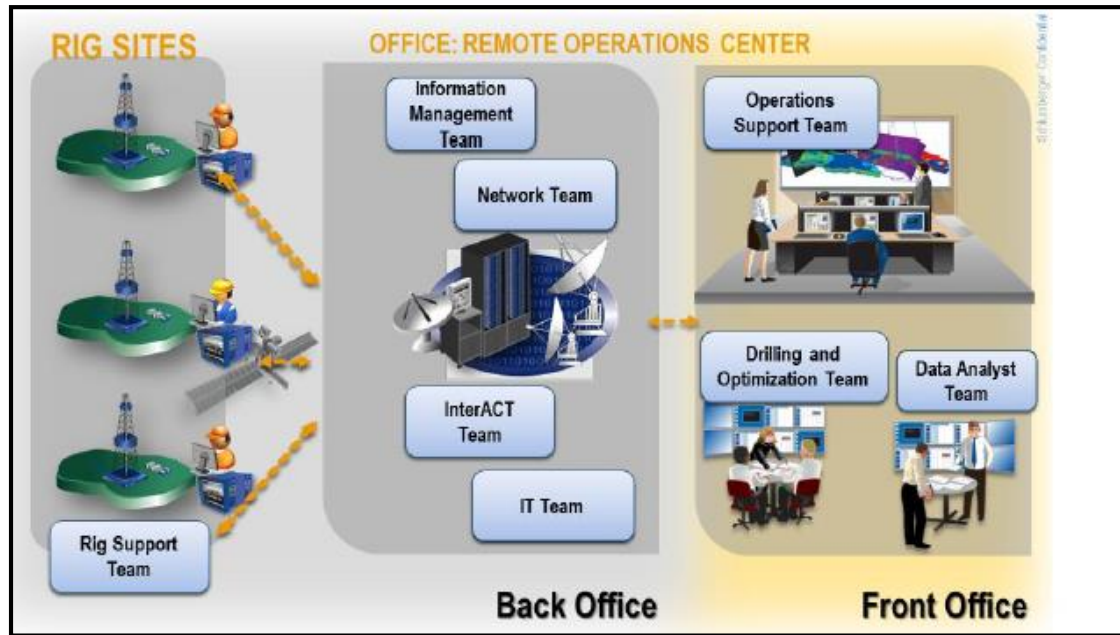
### 1-1. Le projet RTOM

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 suivie 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 opération ROC «Remote Operation Center», situé à la base de SONATRACH 24 Février à HASSI MESSAOUD wilaya d'OURAGLA. L'équipe présente dans le projet et divisé en deux groupe

« **Front office** » et « **Back office** ». [2]

- ❖ **Front office** : ce groupe se charge de tous 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 relatives aux matériels et logiciels. Leur mission, assurée à ce que les données soient toujours disponibles.[2]
- ❖ **Back office** : ce groupe se charge du monitoring des opérations de forage et l'élaboration des rapports de suivi, leur mission détecté d'éventuelles problèmes afin de les signalés aux superviseurs sur le RIG dans le but de les éviter.
- L'objectif principal du projet RTOM est d'améliorer le déroulement des opérations de forage grâce au suivi de leur avancement en temps réel.[2]

Ce suivi révèle les résultats interprétés à partir des données reçues de la base de données chargées depuis le site de forage grâce à l'architecture décrite dans le schéma ci-dessous :



**Figure III. 1 :Description de la transmission de données RTOM**

A partir de ces données reçues, et en les interprétant les OSEs (Operations Support Engineer ) peuvent détecter grâce à leur expérience et observation certains signaux indiquant la possibilité d'avoir un problème notamment le coincement de la garniture, 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, ce travail nécessite une bonne analyse et donne rarement le résultat escompté. [2]

## 2- Architecture global de système

L'architecture globale de notre système est base sur la prédiction du possibilité de l'existence un coincement ou non, et principalement compose par deux partie principale:

- ❖ Partie d'apprentissage automatique.
- ❖ Partie de test.

- Le figure suivante présente l'architecture global de notre étude :

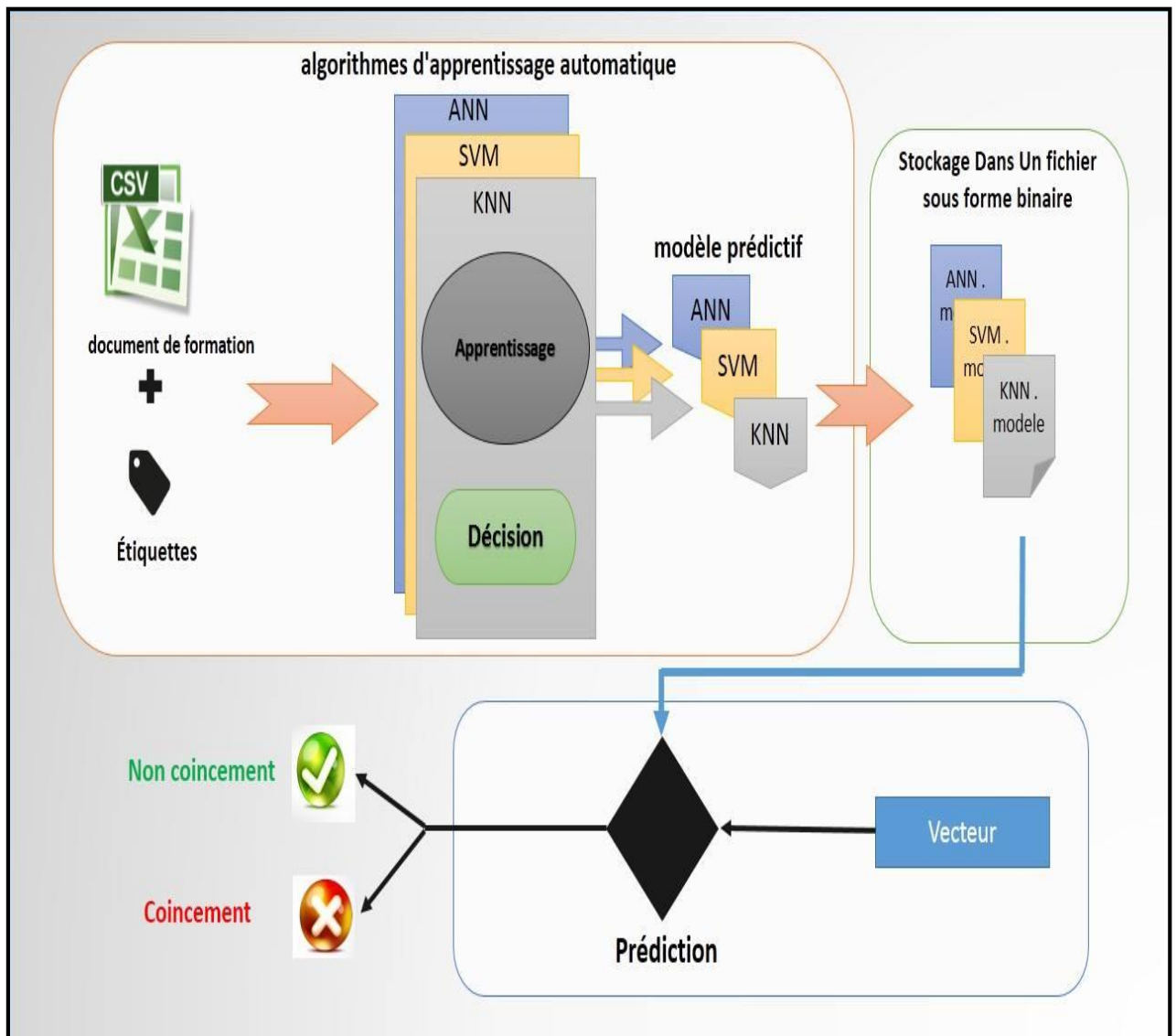


Figure III.2: l'architecture global du système

### 3- Fonctionnement global

Notre système pratiquement contient deux parties selon l'architecture globale. Maintenant nous allons voir comment ces composants interagissent entre eux afin de donner la meilleure efficacité des résultats.

### A- Partie d'apprentissage automatique

Le système prend comme entrée un ensemble des données ce forme de fiches Excel, et dans cette phase consiste à utiliser des méthodes de classification (dans notre étude: RN,SVM,K-PPV). Elle se divise en deux sous phases: Apprentissage et Test ou Décision. La première consiste à initialiser la base des modèles, autant que la deuxième consiste à assigner une classe pour chaque nouveau exemple donnée (vecteur ).

- ❖ **Apprentissage** : 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 coincement et le reste pour la décision.
- ❖ **Décision** : Elle consiste à utiliser les caractéristiques extraites dans la phase précédente pour attribuer une classe en se basant sur les données de la base des modèles prédictifs
  - ✓ Le objectif de cette phase est obtenue des modèles prédictifs, qui stocke dans un base de connaissance .

### B- Partie de test

le but de cette partie est donner une solution, grâce à prédictif un coincement ou non. Et a travers d'insérer et de lier les donnes qui sont sous la forme d'une vecteur avec un modèle prédictif qui choisie a le premier partie. Et aussi, pour renforcer et clarifier le meilleur méthode qui choisie dans le premier partie pour résoudre ce type de problème.

## 4- Description des données

Les données de terrain obtenues à partir d'un puits (la phase 6 pouces pour notre application) on shore verticaux forés dans le champs de TELL OFFSHORE au sud de l'Algérie ont été utilisés dans cette étude.

Notre modèle de prédiction du coincement est basé sur les paramètres (mud logging) suivants (Table.III.1) :

Type de données	Variable d'entrée		Unite
Les parameters de forage	RPM	Rotation per minute	c/min
	STOR	avg surface torque	1000 ft.lbf
	DEPT	bit depth	m
	BPOS	block position	m
	HKLX	Hklx	kkgf
	HDTH	hole depth	m
	TVDE	hole true vertical depth	m
	HOOKLD	Hookld	t
	HKLD	Hookload	kkgf
	PD_PADS_HY	pressure drop over power drive pads	psi
	SPM1	pump 1 stroke rate	1/min
	SPM2	pump 1 stroke rate	1/min
	SPM3	pump 1 stroke rate	1/min
	SPPA	stand pipe pressure	psi
	TQX	Rotary Torque X	1000 ft.lbf
	WOB	Weight on bit	t
WOBX	Weight on bit X	kkgf	

Table.III.1 : Les paramètres de notre modèle .

Avant applique les données pour donne le modèle prédictif, nous avons normalise les données, pour assurer le filtrage des données, et libre de toute qualités indésirables ,ce qui peut conduire à la perte de validité des données.

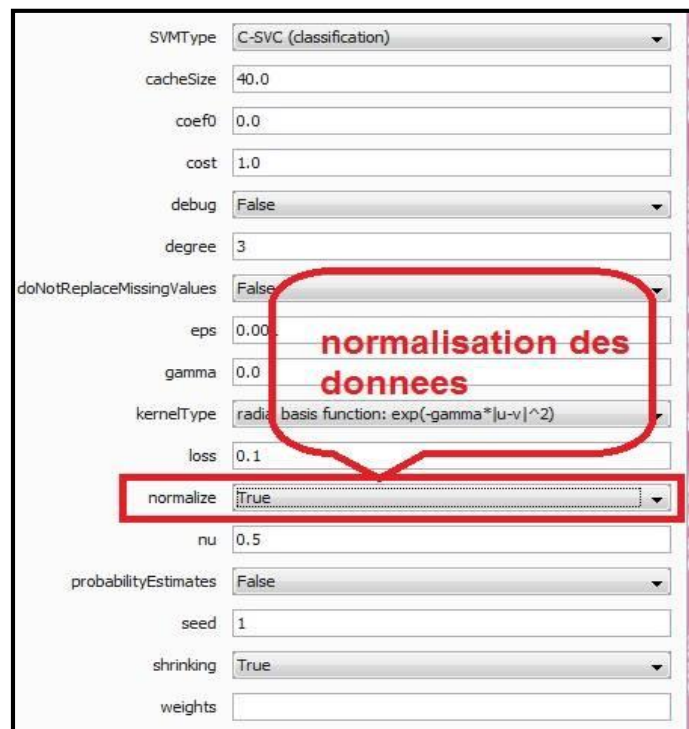


Figure III.3:Normalisation des données dans WEKA

## 5- Comparaison entre les méthodes

La prédiction du coïncement qui est une valeur réelle (continue) est considéré comme un problème de classification, l'outil WEKA (explique dans le chapitre suivant) est doté de plusieurs algorithmes pour prendre en charge de ce type de problème, parmi eux nous avons choisi les plus performants d'après les résultats des expérimentations dans l'environnement WEKA et ce que nous avons trouvé dans la littérature. Citons les RNA, les SVM et les K-PPV qui feront l'objet de notre expérimentation et comparaison.

- Le critère de comparaison est le taux de corrélation, il existe deux mode d'évaluation :

### A- Division en ensemble d'apprentissage et test

Ce mode consiste à diviser la base de données en deux parties, la première (66%) pour l'apprentissage et le reste pour le test. Le modèle est bâti sur l'échantillon d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant un test, une mesure ou un score de performance du modèle sur l'échantillon de test, par exemple l'erreur quadratique moyenne . [18]

### B- Validation croisée (cross-validation)

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 .[18]

- Dans cette évaluation nous avons soins a les critères de comparaisons suivantes:
  - ❖ *Correctly classified instances.*
  - ❖ *Incorrectly classified instances.*
  - ❖ *Mean absolute error*
  - ❖ *Time taken to build model.*

Le figure suivant montre bien les paramètres de notre système

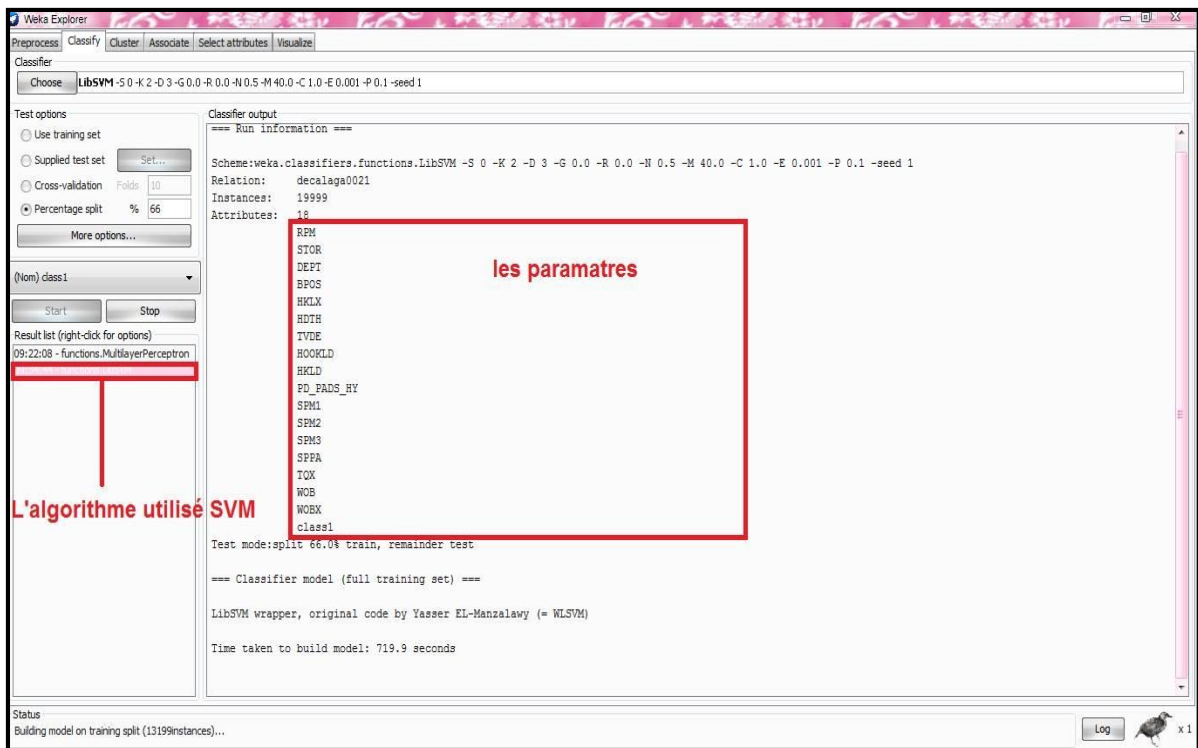


Figure III.4: l'expérimentation et l'évaluation en utilise l'algorithm de SVM

Le figure suivant présente le pourcentage de l'apprentissage, et le temps nécessaire pour construire le model par l'algorithm de (ANN).

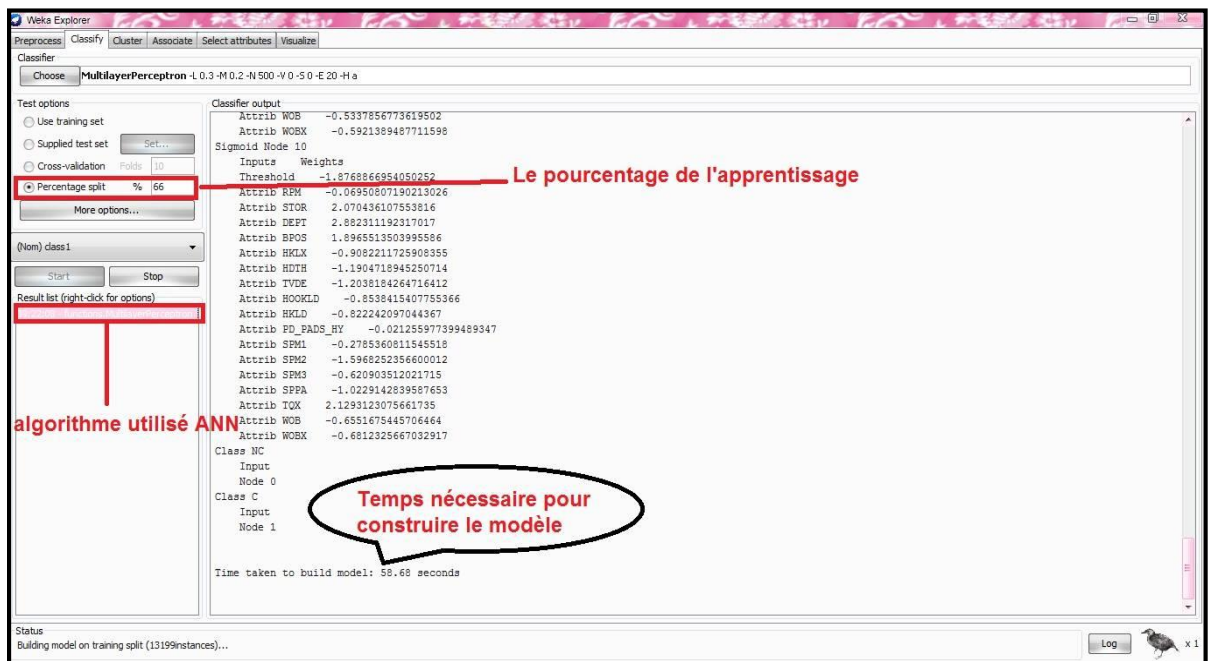


Figure III.5: l'expérimentation et l'évaluation en utilise l'algorithm de ANN

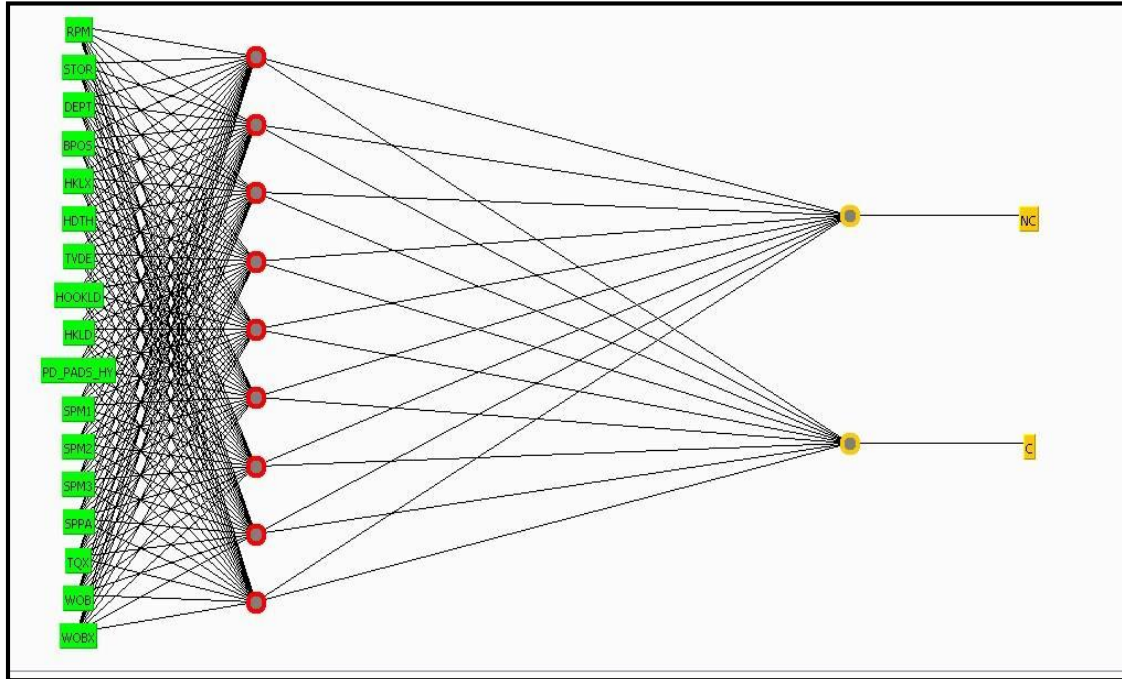


Figure III.6 : Architecture de ANN utilisé pour la prédiction .

Le figure suivant présente les résultats d'apprentissage ,dans cette cas nous avons utilise l'algorithme k-ppv (KNN) , et les critères de cette évaluation

**L'algorithme qui est utilisé KNN**

**résultats**

```

=== Summary ===
Correctly Classified Instances 6695      98.4559 %
Incorrectly Classified Instances 105      1.5441 %
Kappa statistic 0.7633
Mean absolute error 0.0178
Root mean squared error 0.124
Relative absolute error 25.6254 %
Root relative squared error 67.6046 %
Total Number of Instances 6800

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.993	0.253	0.991	0.993	0.992	0.891	NC
	0.747	0.007	0.797	0.747	0.771	0.891	C
Weighted Avg.	0.985	0.245	0.984	0.985	0.984	0.891	

```

=== Confusion Matrix ===
 a  b  <- classified as
6510 45 | a = NC
 60 177 | b = C

```

Figure III.7: l'expérimentation et l'évaluation en utilise l'algorithme de kNN

Le tableau (Tab.III.2 ) ci-dessous obtenu les résultats de comparaison entre les algorithmes de classification pour donner le modèle prédiction

les méthodes les critères	<i>K-PPV</i>	<i>SVM</i>	<i>RN</i>
<i>Correctly classified instances</i>	6685 – 98.45%	6563 - 96.51%	6610 - 97.20 %
<i>Incorrectly classified instances</i>	105 – 1.5%	273 – 3.48%	190 – 2.79%
<i>Kappa statistic</i>	0.76	0.88	0.34
<i>Mean absolute error</i>	0.01	0.03	0.03
<i>Root mean squared error</i>	0.12	0.18	0.15
<i>Relative absolute error</i>	25.62%	50.07%	49.03%
<i>Root relative squared error</i>	67.60%	101.78%	80.39%
<i>Total number of Instances</i>	6800	6800	6800
<i>Time taken to build model</i>	0.01s	719.9s	58.68s

**Table.III.2 : comparaison entre les algorithmes de classification**

Les résultats des tableaux (Table.III.2) montrent que l’algorithme K-plus proche voisin donne meilleur taux de construire le modèle prédiction (*Time taken to build model =0.01 s*), et le pourcentage d’erreur (*Relative absolute error =25.62%*) . Alors l’algorithme k-ppv il est le meilleur choix pour prendre en charge notre problème de prédiction.

## Conclusion

Nous avons vu dans ce chapitre l’architecture globale du système, et détaillé chaque composant, ainsi que le structure de données qui est entre dan notre système. En fin, la comparaison entre les méthodes des classifications pour choisie le meilleur méthode qui utilise a la fin de l’étude.

Les résultats de ce chapitre seront enrichis par de détails d’implémentation dans le chapitre suivant pour réaliser notre système.

*Chapitre IV :*

---

*Implémentation et Réalisation*

## 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.

Nous avons donné dans le précédent chapitre la conception de façon globale et détaillée avec les méthodes utilisées, 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, ainsi que les structures de données choisies pour implémenter ce type de système. Ensuite, on va présenter les algorithmes utilisés illustré par quelque résultat obtenu, nous terminons ce chapitre par une conclusion.

### 1. Environnement et outils de mise en œuvre

Notre système était développé sur un ordinateur de processeur Core I3 avec une RAM de 4 Go sous Windows 7, mais on peut implémenter ce système sur n'importe quel machine grâce au virtuel machine de Java.

#### 1-1. Java

Java est un langage orienté objet développé par la société Sun. La syntaxe générale du langage java est très proche de celle du langage C, et parmi ces avantages [19] :

- ❖ Le byte-code assure à Java une portabilité complète vers d'autres systèmes.
- ❖ L'importance de l'API de base qui offre tous les services de base, notamment pour la construction des interfaces graphiques.
- ❖ La troisième force du Java, c'est son adaptabilité dans des nombreux domaines, autant pour le web que pour les systèmes embarqués.



Figure IV.1: La langage de développement.

## 1-2. Netbeans

Concernant l'environnement de développement java, notre choix s'est porté sur NETBEANS qui est un environnement de développement intégré (EDI). Il inclut une gamme complète d'outils, de concepteurs et d'experts pour aider à concevoir, développer, tester, déployer et gérer les applications Java .[19]



Figure IV.2: Environnement de développement-NetBeans

## 1-3. Weka (machine Learning)

**Waikato Environment for Knowledge Analysis (Weka)** est un ensemble d'outils permettant de manipuler et d'analyser des fichiers de données, implémentant la plupart des algorithmes d'intelligence artificielle, entre autres, les arbres de décision et les réseaux de neurones .[20]

Il est écrit en java, disponible sur le web

- Il se compose principalement [18]:
  - ❖ De classes Java permettant de charger et de manipuler les données.
  - ❖ De classes pour les principaux algorithmes de classification supervisée ou non supervisée.
  - ❖ D'outils de sélection d'attributs, de statistiques sur ces attributs.
  - ❖ De classes permettant de visualiser les résultats. On peut l'utiliser `a trois niveaux :
    - Via l'interface graphique, pour charger un fichier de données, lui appliquer un algorithme, vérifier son efficacité.
    - Invoquer un algorithme sur la ligne de commande.
    - Utiliser les classes définies dans ses propres programmes pour créer d'autres méthodes, implémenter d'autres algorithmes, comparer ou combiner plusieurs méthodes. C'est cette troisième possibilité qui sera utilisée en travaux pratiques.

Dans notre approche nous avons utilisé l'outil logiciel open source WEKA pour la comparaison des algorithmes de prédiction afin de bien justifier notre choix de l'algorithme K-plus proche voisin, pour l'implémentation le langage JAVA a été choisi avec l'environnement NETBEANS



**Figure IV.3: L'outil WEKA**

## **2- L'application développée**

Nous présentons ici des captures d'écran de l'application développée.



**Figure IV.4:L'interface principale de l'application.**

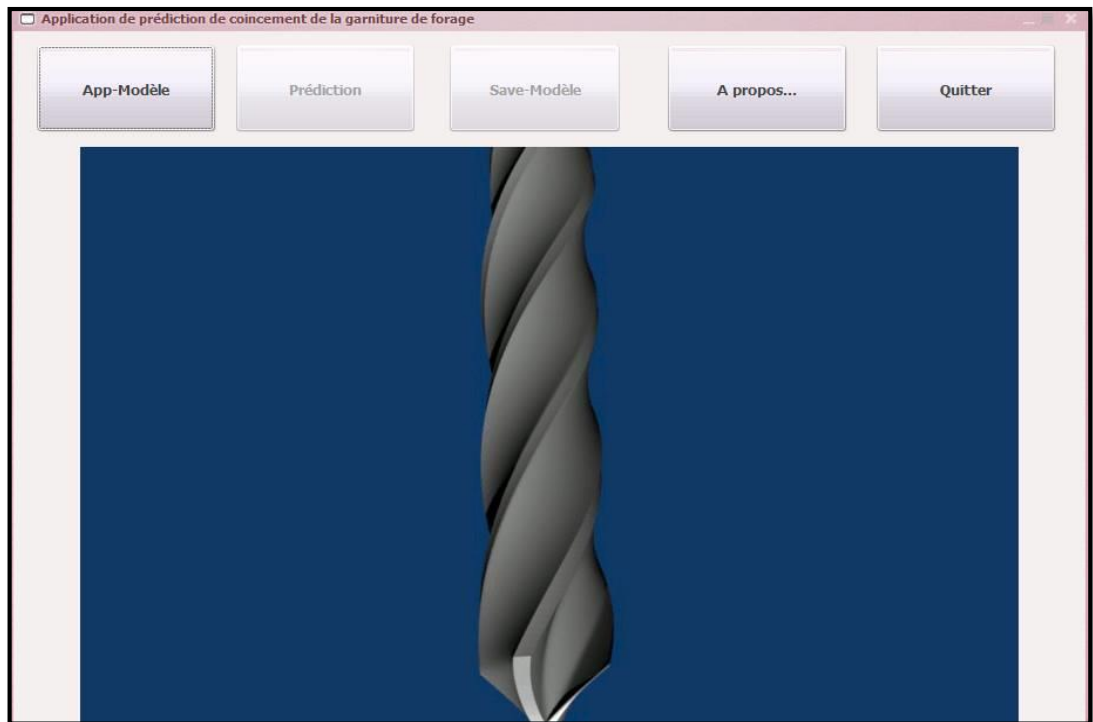


Figure IV.5:L'interface principale de prototype version zéro.

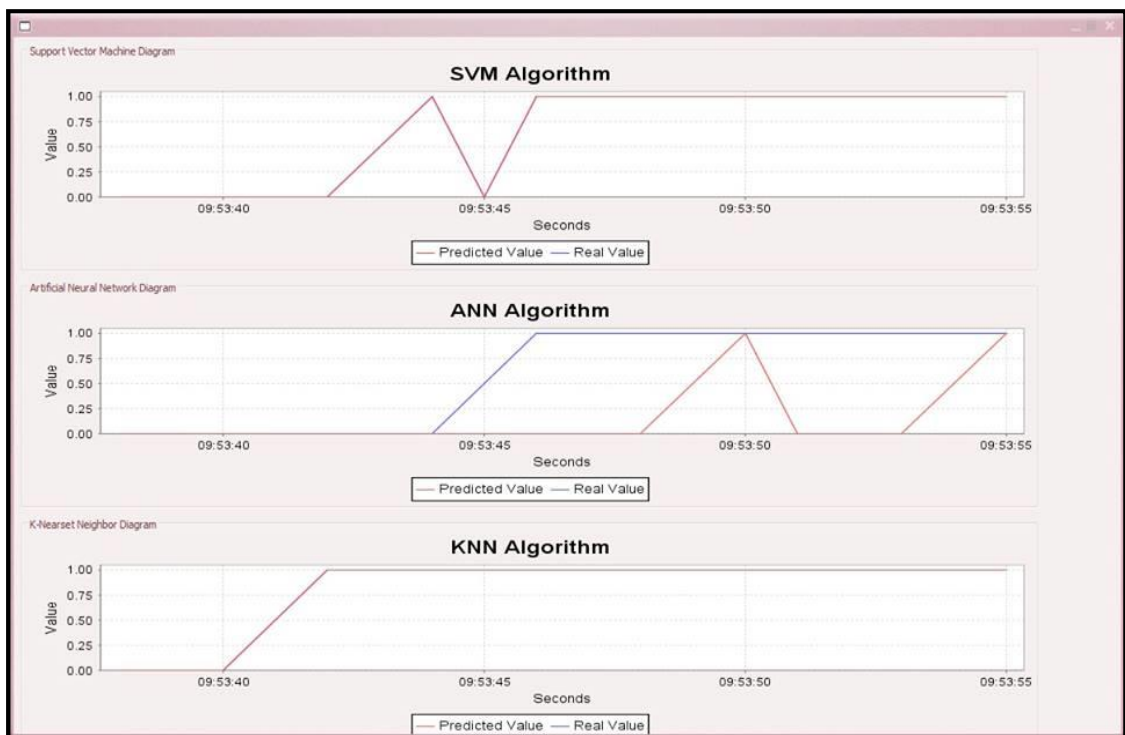


Figure IV.6:L'interface du comparaison entre les méthodes d'application

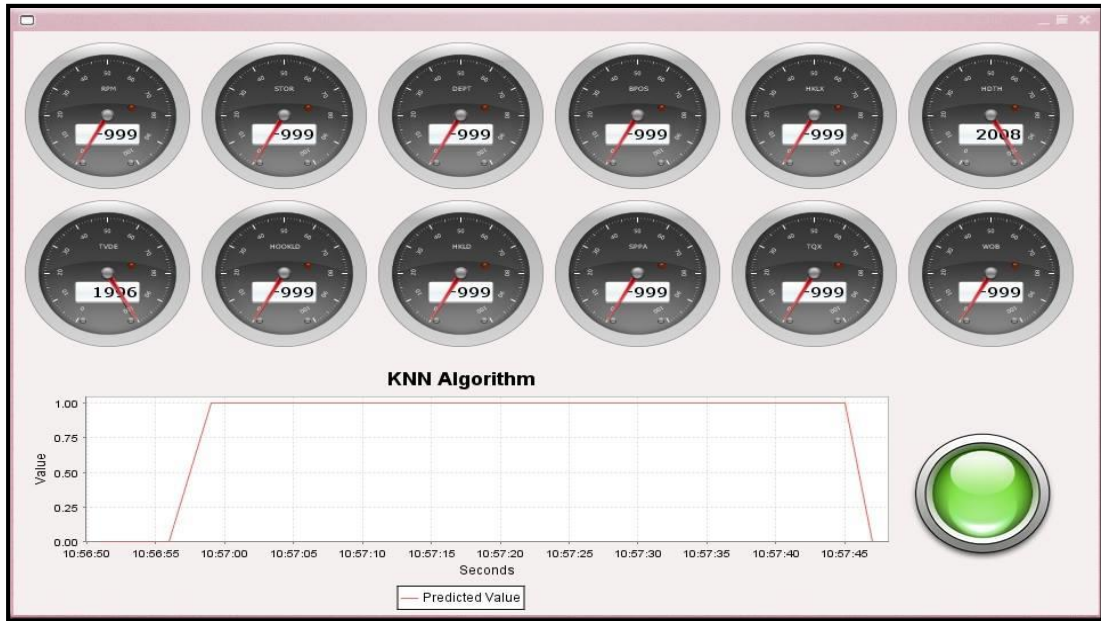


Figure IV.7:L'interface de résultat avec le meilleur algorithmme (KNN) no coincidence

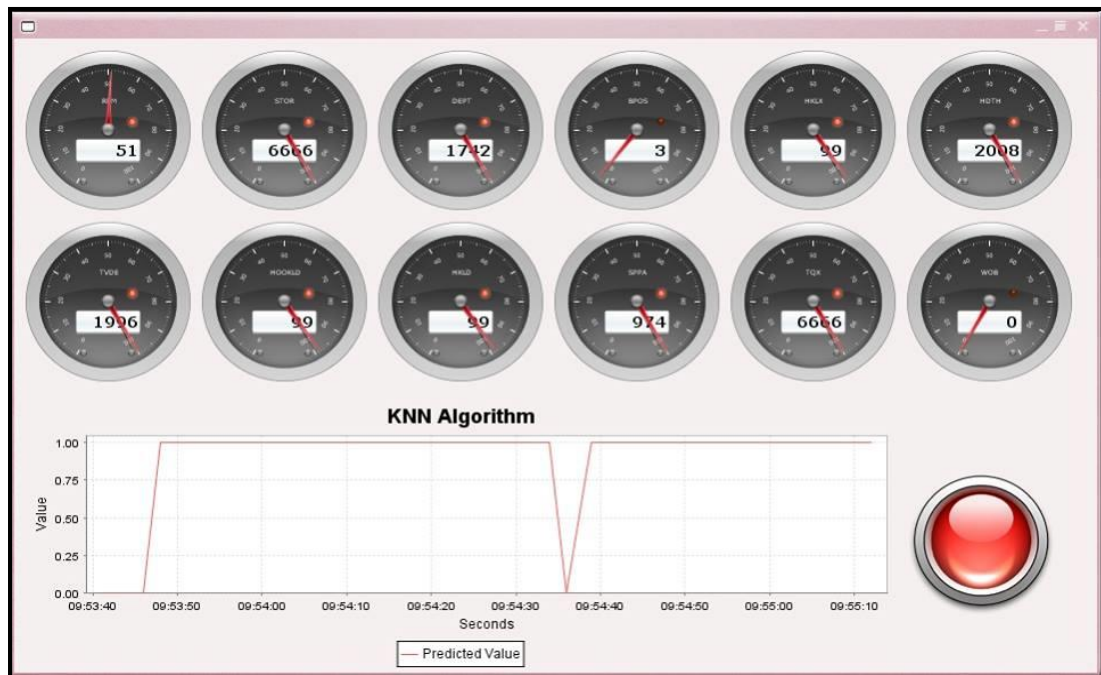


Figure IV. 8:L'interface de résultat avec le meilleur algorithmme (KNN) coincidence

**Conclusion :**

Arrivé à la fin de ce chapitre, nous avons décrit les outils et environnements utilisés ainsi que l’aspect technique en relation avec le déploiement de la solution, nous avons donc codé et mit en place notre application en suivant les descriptions fournies tout au long du chapitre de conception.

Notre projet de fin d'étude s'inscrit dans le cadre d'un projet de suivi des opérations de forage pétrolier en temps réel. Notre travail consistait à concevoir et réaliser un système de prévention des coincements de la garniture de forage.

Une méthodologie de prédiction et d'optimisation de le coincement de la garniture de forage (stuck pipe) a été développée, démontrée et appliquée afin atteindre des paramètres de forage contrôlables optimales. La première tâche de cette étude est la prédiction de le coincement de la garniture de forage. Les données utilisées dans le cadre de cette étude appartiennent à des puits forés en Sud de l'Algérie.

Le coincement de la garniture de forage (stuck pipe) est prédite à l'aide les algorithmes k plus proche voisin pour classification qui est plus performant que les autres algorithmes de classification à savoir les RNA et les SVM d'après les comparaisons que nous avons effectuées dans l'outil Learning-machine WEKA présentées sur les tableaux III.2 . Les résultats ont indiqué que Le coincement de la garniture de forage peut être prédit avec une haute précision, (*Relative absolute error =25.62%*) après le choix des meilleurs paramètres de notre algorithme k plus proche voisin.

Travailler sur ce projet durant, nous a permis de capitaliser des concepts techniques en complément aux enseignements théoriques acquis. Cela nous a permis de nous rapprocher et de nous familiariser avec le monde du travail.

En perspective nous pensons exploiter des approches d'intelligence artificielle avancée comme l'apprentissage, Elle peut servir comme un système d'aide à la décision pour optimiser les opérations de forage en temps réel.

- [1] : Abdesslem BELAID : Modélisation Tridimensionnelle Du Comportement Mécanique De La Garniture De Forage Dans Les Puits A Trajectoires Complexes : APPLICATION A LA PREDICTION DES FROTTEMENTS GARNITURE-PUIT, These pour obtenir le grade de DOCTEUR DE L'ECOLE NATIONALE SUPERIEURE DES MINES DE PARIS, Anne 2005.
- [2]: MILI Saoussen, KAOUANE Housseyn : Conception et réalisation d'un système de prévention des « Stuck Pipe» HASSI MESSAOUD, Mémoire de Fin d'étude, esi 2014.
- [3] : PetoWiki: <http://petrowiki.org/PetroWiki>, site web pour les information sur le domaine pétrolier .
- [4]: P. CABENA, P. HADJINIAN, R. STADLER, J. VERHEES et A. ZANASI, Discovering Data Mining: From Concept to Implementation, Prentice Hall, Upper Saddle River, NJ, 1998.
- [5]: E-G. TALBI, Fouille de données (Data Mining) : Un tour d'horizon, Laboratoire d'Informatique Fondamentale de Lille.
- [6] : CHAMI Djazia : Une plate forme orientée agent pour le data mining, Mémoire pour l'obtention du diplôme de Magister en informatique , Université HADJ LAKHDAR – BATNA, Anne 2010
- [7] : M. J. BERRY, G. S. LINOFF, Data Mining Techniques For Marketing, Sales, and Customer Relationship, Management, Second Edition, 2004.
- [8] : M. J. BERRY, G. S. LINOFF, Mastering Data Mining: The Art and Science of Customer Relationship Management, 2000.
- [9] : D.T. LAROSE, Discovering Knowledge In Data: An Introduction to Data Mining, Central Connecticut State University, 2005.
- [10] : S. PRABHU, N. VENKATESAN, Data Mining and Warehousing, New Age International (P) Ltd., Publishers, New Delhi, 2007.
- [11]: G. CALAS, Études des principaux algorithmes de data mining, Spécialisation Sciences Cognitives et Informatique Avancée, France.
- [12] : L'encyclopédie en ligne WIKIPEDIA 2009.
- [13] : Rapporté de [http://interstices.info/encart.jsp?id=c\\_41867&encart=3&size=600,500](http://interstices.info/encart.jsp?id=c_41867&encart=3&size=600,500).
- [14] : R.GILLERON, M. TOMMASI, Découverte de connaissances à partir de données, 2000.
- [15]: C. GROUIN, Les techniques de la fouille de données, INaLCO, 2009/2010.
- [16] : T. Yang, L. Cuixia, « The Study on Livestock Production Prediction in Heilongjiang Province Based on Support Vector Machine », Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE 2013), Published by Atlantis Press, Paris, France.

[17] : MERIZIG ABDELHAK :conception Conception et Réalisation d'un système d'information pour le suivi des dattes, mémoire pour obtenir le diplôme de master académique en informatique, Université Mohamed Khider – BISKRA , Anne 2013

[18] : 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

[19]: T. Ahmed « Une approche basée agents mobile pour la télésurveillance industrielle», Master en informatique, intelligence artificielle, 2011.

[20] : weka :<http://www.fil.univ-lille1.fr/~decomite/ue/APE/tp/tp1/weka2009.pdf>, site web pour les information sur l'environnement weka