



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université AMO de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

Mémoire de Master

en Informatique

Spécialité : ISIL

Thème

Une Approche Sémantique Pour Le Data Mining Et
Le Machine Learning

Encadré par

— DR.AID Aicha

Réalisé par

— LARBAOUI Racha

— NACERI Siham

2020/2021

Remerciements

Nous tenons à remercier Dieu, le tout puissant et miséricordieux, qui nous a donné la force, l'intelligence et la patience d'accomplir ce modeste travail.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de **Mm Aid Aicha** , nous la remercions du fin fond de notre cœur pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire. Nous n'oublierons jamais ses conseils .

Nous adressons nos remerciements au président et aux membres de jury d'avoir accepté le jugement de notre travail.

À tous nos enseignants du département informatique, un remerciement particulier et sincère pour tous les efforts fournis durant notre encadrement tout au long de ces cinq années, où nous avons acquis les connaissances nécessaires pour notre progression et amélioration durant le cursus qu'est le nôtre.

Nos remerciements à nos parents pour leurs aides, soutien moral et leurs encouragements tout au long de notre carrière universitaire.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours soutenues et encouragés au cours de la réalisation de ce mémoire.

Dédicaces

Je dédie cet événement marquant de ma vie à la mémoire de mon papa disparu trop tôt. J'espère que, du monde qui s'en va maintenant, il apprécie cet humble geste comme preuve de reconnaissance de la part de sa fille qui a toujours prié pour le salut de son âme. Puisse dieu, le tout puissant, l'avoir en sa sainte miséricorde.

À ma maman qui m'a soutenu et encouragé durant ces années d'études. Qu'elle trouve ici le témoignage de ma profonde reconnaissance.

À ma grande soeur Manel, mon frère Lamara , et ceux qui ont partagé avec moi tous les moments d'émotions lors de la réalisation de ce travail. Ils m'ont chaleureusement supporté et encouragé tout au long de mon parcours.

À ma famille, mes proches et à ceux qui me donnent de l'amour et de la vivacité.

À ma collègue Siham, qui a partagé avec moi ce mémoire .

À tous mes amis qui m'ont toujours encouragé, et à qui je souhaite plus de succès.

À tous ceux que j'aime

Larbaoui Racha

Dédicaces

Je ne peux commencer sans évoquer le nom d'ALLAH notre Dieu le tout puissant qui m'a donné la patience, la santé et sa bien vaillance tout au long de ma vie, grâce à lui j'ai pu réaliser ce travail, que je dédie :

À mes très chers parents « Nacéri Abd El Kader et Moulai Nadjia » qu'ils trouvent dans ce travail l'expression de ma profonde reconnaissance pour tous leurs sacrifices, leur encouragement et leur amour. A toute ma famille spécialement :

À mes très chers grands parents.

À mes très chers tantes Hayet et Madani Zoubida.

À toute mes très chers oncles .

À mes chers frères Bilel, Mohamed et Abd el Hak.

À ma merveilleuse sœur Hadil qui n'a cessé de m'encourager tout au long du projet.

À mes adorables Ritadj et Hanaa.

À ma chère binôme Larbaoui Racha que j'aime tant, Je tiens à la remercier pour sa grande patience, sa disponibilité et son sérieux. C'est un honneur pour moi d'avoir partagé ce travail avec toi.

À mes chères amies Sakina , Soraya , Oussama , Nada , Lynda et Mousteghanemi Soraya, Je tiens à les remercier pour leurs incessants encouragements ainsi que leur aide précieuse.

Nacéri Siham

Résumé

La sensibilité au contexte est née d'un terme de l'informatique omniprésente ou d'une informatique dite omniprésente qui cherchait à relier les changements de l'environnement aux systèmes informatiques, qui sont autrement statiques.

A travers ce mémoire, nous avons proposé un système sensible au contexte qui ne prend pas compte un seul modèle du Machine Learning, mais il prendra plusieurs modèles. Chaque modèle sera utilisé pour un contexte.

Dans notre proposition, nous avons comparé l'utilisation d'un modèle qui a été formé en utilisant tous les Contextes à un système composé d'un ensemble de modèles spécialisés qui ont été formés pour chaque Contexte dans le but d'améliorer la performance et la précision d'un Modèle du Machine Learning.

Pour illustrer notre approche proposée, nous reproduisons trois expériences qui utilisent un modèle basé sur la classification/Prédiction, elles ont été testées par l'algorithme Forêts Aléatoires, étant donné qu'il a donné des meilleurs résultats en terme d'Accuracy "précision" de la classe.

Nous montrons que notre approche Contextuelle peut améliorer les résultats avec différents Algorithmes du Machine Learning.

Mot clés : Apprentissage Automatique Supervisé , Contexte, Système Sensible au Contexte, Précision, Prédiction, Forêts Aléatoires . . .

Abstract

Context Sensitivity arose out of a term of ubiquitous computing or so-called ubiquitous computing that sought to relate changes in the environment to computer systems, which are otherwise static.

Through this thesis, we have proposed a Context sensitive system that does not take into account a single machine learning model, but it will take several models. Each model will be used for a Context.

In our proposal, we compared the use of a model which was trained using all Contexts to a system composed of a set of specialized models which were trained for each Context with the aim of improving performance and the Accuracy of a Machine Learning Model.

To illustrate our proposed approach, we reproduced three experiments which use a model based on Classification / Prediction, they were tested by the Random Forests algorithm, since it gave better results in terms of Accuracy .

We show that our Contextual approach can improve results with different Machine Learning Algorithms.

Key words : Machine Learning , Context, Context Sensitivity , Accuracy, Prediction, Random Forests ...

Table des matières

Table des matières	i
Table des figures	iv
Liste des tableaux	vi
Liste des abréviations	vii
Introduction générale	1
1 Système Sensible au Contexte et Le Web Sémantique	4
1.1 Introduction	4
1.2 Le Contexte	5
1.2.1 Notions Sur le Contexte	5
1.2.2 Sensibilité au Contexte et Systèmes Sensibles au Contexte	9
1.2.3 Modélisation du Contexte	11
1.3 Le Web Sémantique	13
1.3.1 Définition du Web Sémantique	13
1.3.2 Les Composants du Web Sémantique	14
1.3.3 Les Langages du Web Sémantique	15
1.4 Conclusion	18
2 Machine Learning	19
2.1 Introduction	19
2.2 Machine Learning	20

2.2.1	Définition du Machine Learning	20
2.2.2	Types du Machine Learning	20
2.2.3	Les Algorithmes De Machine Learning	24
2.3	Méthode Dévaluation d’Un Modèle du Machine Learning	32
2.3.1	La Matrice De Confusion	32
2.4	Conclusion	34
3	Solution Proposée	35
3.1	Introduction	35
3.2	Problématique	35
3.3	Travaux Connexes	36
3.3.1	De vito et Coll	36
3.3.2	Nascimento et Lucena	37
3.3.3	Hernandez et Coll	38
3.3.4	Mmezouar et Coll	38
3.4	La Différence entre Les Travaux Connexes et notre Solution	39
3.5	Approche et Solution Proposée	39
3.5.1	La première étape	42
3.5.2	La 2ème étape	42
3.5.3	La 3ème étape	42
3.5.4	La 4ème étape	43
3.5.5	La 5ème étape	43
3.5.6	La 6ème étape	43
3.6	Analyse et Exploration des Données (Dataset)	44
3.6.1	E-Commerce Shipping	44
3.6.2	CardioVasculaire	45
3.6.3	Rain In Australia	45
3.7	Identification du Contexte pour chaque DataSet choisi	48
3.7.1	Identification du Contexte pour ”Shipment Tracking”	48
3.7.2	Identification du Contexte pour ”CardioVascular Diseases”	49
3.7.3	Identification du Contexte pour ” Rain In Australia ”	50
3.8	Modélisation du Contexte (Représentation)	51
3.9	Conclusion	51

4 Résultats et évaluation.	52
4.1 Introduction	52
4.2 Environnement de développement	52
4.2.1 Plateforme (Environnement) De développement	52
4.2.2 Langage de Développement	53
4.2.3 Bibliothèques Utilisées	54
4.3 Discussion et Résultats Obtenu Pour les Trois Expériences (DataSets) . . .	54
4.3.1 La Première Expérience (Rain In Australia)	54
4.3.2 La Deuxième Expérience (CardioVascular Diseases)	66
4.3.3 La Troisième Expérience "Shipment Tracking"	74
4.4 Conclusion	87
Conclusion Générale	88

Table des figures

1.1	Les Composants du Web Sémantiques [41]	15
1.2	Exemple d'un Graphe RDF [42]	16
2.1	Schéma d'Apprentissage Supervisé [50]	21
2.2	Hyperplan Optimal Avec Une Marge Maximale [56]	28
2.3	Exemple d'Un Réseau Neurone [59]	30
2.4	Architecture d'un Réseau Neurone [60]	31
2.5	Fonctionnement d'un Réseau Neurone [61]	31
3.1	Architecture Générale.	41
3.2	Information sur le Dataset CardioVascular Disease	45
3.3	Ensemble de Donnée (Rain in Australia)	46
3.4	Former un Modèle du Machine Learning	47
4.1	Caractéristiques du Modèle Rain in Australia	56
4.2	Modélisation des Contextes " Humidité , Saison "-1 Graphe RDF	63
4.3	Modélisation des Contextes " Humidité , Saison "-2 Graphe RDF	64
4.4	Modélisation des Contextes " Humidité , Saison "-3 Graphe RDF	64
4.5	Exemple Pour la Prédiction en Temps Réel	65
4.6	Résultat de la Prédiction en Temps Réel Pour le Modèle Rain In Australia	66
4.7	Modélisation des Contextes Age-1 Graphe RDF	72
4.8	Modélisation des Contextes Age-2 Graphe RDF	73
4.9	Exemple du Dataset2	74
4.10	Prédire si le Patient a la Maladie	74

4.11	Statiques et Caractéristiques du Dernier DataSet (Shipment Tracking) . . .	78
4.12	Modélisation des Contextes Shipment Tracking .1 Graphe RDF	85
4.13	Modélisation des Contextes Shipment Tracking .2 Graphe RDF	86
4.14	Modélisation des Contextes Shipment Tracking .3 Graphe RDF	86
4.15	Exemple pour la prédiction d'un Produit	87
4.16	Résultat Obtenu pour la Prédiction d'un Produit	87

Liste des tableaux

1.1	Caractéristiques du Contexte [18]	7
1.2	Comparaison entre la Catégorisation Conceptuelle et Opérationnelle	9
1.3	Catégories d'Applications Sensibles au Contexte d'après Schilit et al	10
2.1	Différence Entre Les 4 Types de Machine Learning	23
2.2	Les Avantages et les Inconvénients Des Arbres De Décision	24
2.3	Les Avantages et Les Inconvénients des Forêts Aléatoires	27
2.4	Avantages et Inconvénients Du SVM	28
2.5	Avantages et Inconvénients de la Classification Naïve Bayes	29
2.6	Avantages et Inconvénients des Réseaux de Neurones	32
2.7	Matrice de Confusion	33
4.1	Amélioration des Résultats Pour le Modèle Général (Rain in Australia)	55
4.2	Amélioration des Résultats du Modèle Conrextuel (Saison)	55
4.3	Optimisation des Résultats du Deuxième Contexte (Humidité)	57
4.4	Accuracy Du Modèle Général Après l'Optimisation	66
4.5	Résultats obtenus pour la colonne âge	67
4.6	Résultat du Modèle Général en Utilisant la Validation Croisée	75
4.7	Résultat Après Optimisation	76
4.8	Résultats du Contexte " Prix "	76
4.9	Optimisation des Résultats pour la Colonne Weight	78

Liste des abréviations

CCML :	Centaurus Capability Markup Language
CSCP :	Comprehensive Structured Context Profiles
CDF :	Context Description Framework
UML :	Unified Modeling Language
ORM :	Object-Role Modeling
CML :	Context Modeling Language
RDF :	Resource Description Framework
RDFS :	Resource Description Framework Schema
OWL :	Ontology Web Language
URIs :	Universal Resource Identifier
DL :	Logique Description
CG :	Graphe Conceptuel
XML :	Extensible Markup Language
IA :	L'Intelligence Artificielle
ML :	Machine Learning
RL :	Reinforcement Learning
Deep RL :	Deep Reinforcement Learning
SVM :	Support Machine Vector

Introduction générale

Avec l'avènement, des appareils mobiles et des réseaux mobiles et sans fil, divers paradigmes de l'informatique tels que l'informatique ubiquitaire, ambiante, sensible au contexte et autonome ont pris naissance. Ces paradigmes partagent toute la notion de contexte qui joue un rôle important et révèlent des comportements intelligents qui entourent les utilisateurs de ces environnements. La nature du contexte peut prendre plusieurs formes, à savoir, spatial (localisation, orientation, vitesse), temporel (date, temps, saison), environnemental (lumière, température ambiante, bruit) et social (utilisateurs, ses préférences et activités). En conséquence, l'implication du contexte entraîne l'apparition d'applications sensibles au contexte.

En tant que technologie clé dans le domaine étendu de l'intelligence artificielle un champ de recherche en pleine croissance appelé Context awareness a pour but de fournir de l'information pour construire des services plus intelligents et accessibles.

De nombreux efforts de recherche dans le domaine des applications sensibles au contexte sont axés sur la modélisation, le raisonnement et la gestion d'informations contextuelles, afin de réduire la complexité de l'ingénierie de tels systèmes. Ces efforts ont permis d'accomplir des progrès considérables dans la formulation de modèles de contexte qui sont une condition préalable importante pour la construction de systèmes capables de répondre aux changements dans leur environnement. Les contextes utilisés par des modèles, des architectures logicielles et plates-formes regroupant des applications sensibles au contexte sont essentiellement de nature physique (emplacement, temps, activité, etc.).

Au cours de la dernière décennie, les chercheurs ont travaillé sur le problème comment créer des modèles généraux du Machine Learning (ML) pour permettre aux systèmes de gérer des tâches complexes. Par exemple, une tâche de classification qui utilise un modèle de Machine Learning pour prédire le Climat ("La météo"). D'après cet exemple nous déduisons qu'un seul modèle pour cette tâche sera formé et que nous les appelons un modèle général.

Afin qu'il soit testé et validé en utilisant plusieurs algorithmes du Machine Learning. Les résultats de ce dernier ne peuvent être pas efficaces. C'est-à-dire lorsque nous allons faire une prédiction en temps réel pour notre modèle et nous allons voir s'il fonctionne parfaitement, nous pouvons tomber dans l'erreur, car nous avons obtenu un seul modèle. D'après cette déduction, notre idée (proposition) vient qui consiste à créer un système sensible au contexte dans le but de gérer plusieurs modèles du Machine Learning. Chaque modèle généré sera utilisé pour un Contexte bien défini dont le Contexte peut être une ou plusieurs colonnes du DataSets.

Nous avons proposé notre proposition(solution) dans le but d'améliorer la performance et la précision d'un modèle du Machine Learning en utilisant le contexte d'où notre système devient sensible au contexte.

Objectifs :

Notre travail a été réalisé en suivant les objectifs :

- Proposer un Système sensible au Contexte.
- Améliorer la performance et la précision d'un modèle de Machine Learning en utilisant le Contexte.
- Le système ML devient sensible au Contexte dont le Contexte peut être une ou plusieurs colonnes.
- Le Contexte sera modélisé en RDF en utilisant des ontologies de domaine pour apporter plus de sémantiques.

Pour ce qui est de la présentation du plan de ce mémoire, le travail que nous avons mené s'organise comme suit :

le premier chapitre dresse la notion du Contexte en présentant et en discutant les différentes définitions, catégorisation, dimension et approche de modélisation proposée dans la littérature. Ensuite, nous avons parlé sur le web sémantique ainsi les différents langages du Web sémantique.

Dans le deuxième chapitre, nous avons présenté le Machine Learning en appliquant ses différents types ainsi ses algorithmes qui nous ont permis de modéliser un modèle du Machine Learning en se basant sur leurs avantages et limites.

Le troisième chapitre représente notre proposition (solution) qui a été expliquée et détaillée en passant par des étapes essentielles afin d'arriver à son implémentation finale.

Le dernier chapitre présente la discussion des résultats obtenus et l'implémentation de notre proposition.

Nous terminerons ce mémoire par une conclusion générale qui résume le travail effectué.

Systeme Sensible au Contexte et Le Web Sémantique

1.1 Introduction

La notion du Contexte est assez universelle, elle a initialement émergé dans divers domaines comme la recherche psychologie et philosophie et c'est à partir des années 90, que les chercheurs dans le domaine de l'informatique ont commencé à s'y intéresser. Elle désigne l'ensemble des éléments qui peuvent influencer la compréhension d'une situation particulière. Cette description est ensuite utilisée pour mieux comprendre l'environnement. Y en a plusieurs définitions pour la notion Contexte qui ont été proposées[1]. Nous proposons un état d'art sur la notion du Contexte y compris le Web Sémantique et l'Ontologie, nous commençons d'abord par définir le Contexte.

Ce chapitre est organisé en deux sections, la première, nous nous concentrons sur la notion de Contexte en présentant et en discutant les différentes définitions, catégorisations, dimensions et approches de modélisation proposée dans la littérature. En revanche dans la deuxième section, nous parlerons sur le Web Sémantique et les Ontologies et en incluant les différents langages du Web Sémantiques et différents formalismes d'Ontologie.

1.2 Le Contexte

1.2.1 Notions Sur le Contexte

Comme nous l'avons introduit, le Contexte a été utilisé dans plusieurs disciplines comme la psychologie, la philosophie et le langage naturel. Dans le domaine de l'informatique, de nombreuses définitions du Contexte ont été proposées par plusieurs communautés : Intelligence Artificielle, Informatique Diffuse, Recherche d'Informations, Bases de Données, etc. Avant de présenter les définitions du Contexte proposées dans la littérature scientifique, nous commençons par introduire et analyser des définitions proposées dans différents dictionnaires [2] [3][4][5][6][7] :

Grand Dictionnaire Technologique (GDT) :¹ « Texte entourant un terme, ou situation dans laquelle ce terme est employé ». En informatique « ensemble d'informations concernant l'action du stylet, en rapport principalement avec sa localisation à l'écran, qui permet au système d'exploitation de l'ordinateur a stylet de différencier les commandes et l'entrée des données, et de fonctionner en conséquence ».

Le dictionnaire Larousse :² Le Contexte comme « un ensemble de circonstances dans lesquelles se produit un évènement, se situe une action »

Techno-Science :³, **encyclopédie scientifique en ligne :** « Le contexte d'un évènement inclut les circonstances et conditions qui l'entourent ». En informatique, « le Contexte est l'ensemble des conditions sous lesquelles un dispositif est en train d'être utilisé par exemple l'occupation actuelle de l'utilisateur ».

Centre National de Ressources Textuelles et Lexicales :⁴

« Ensemble de circonstances liées, situation ou un phénomène apparaît, un événement se produit ». En informatique, « le Contexte est l'ensemble d'informations concernant un événement en rapport principalement avec sa localisation, qui permet au système

1. <http://gdt.oqlf.gouv.qc.ca/>

2. <http://www.larousse.fr/dictionnaires/francais/contexte/18593?q=contexte18491>

3. [texthttps://www.techno-science.net/](https://www.techno-science.net/)

4. Centre National de Ressources Textuelles et Lexicales (cnrtl.fr)

informatique de fonctionner et de s'adapter en conséquence ».

D'après les définitions, nous remarquons qu'elles décrivent tout le Contexte comme un ensemble d'information associé à quelque chose. Maintenant, nous allons comprendre la notion du Contexte dans le domaine informatique. Dans l'intelligence artificielle, **McCarthy** l'a défini comme une entité mathématique abstraite avec des propriétés [8].

Tandis que **Brézillon**, l'a définit comme un ensemble de conditions pertinentes et d'influences qui font qu'une situation est unique et compréhensible [9].

Dans le domaine de l'Informatique Ubiquitaire, ils ont dit que le Contexte se réfère à la localisation et l'identité des personnes et des objets à proximité ainsi les modifications pouvant intervenir sur ces objets (définition par énumération d'exemples [10][11][12][13].

Dans un autre travail, **Schilit** reprend sa **première définition** (Qui dit que : le Contexte se réfère à la localisation et à l'identité des personnes et des objets à proximité ainsi que les modifications pouvant intervenir sur ces objets) et propose une méthode pour déterminer le Contexte en répondant a trois questions : où êtes-vous ? Avec qui êtes-vous ? Et quelles sont les ressources à proximité de vous ?. Quelques années plus tard, Brown et al introduit l'heure, la saison, la température, l'identité et la localisation de l'utilisateur comme nouveaux éléments du Contexte [14].

Abowd et al ont constaté que :

- Une entité dans le domaine centré application logicielle est une fonction, un événement dans le logiciel ou le logiciel lui-même [15].
- Une entité peut être un dispositif. Dans ce cas, le Contexte est un environnement physique et social dans lequel ce dispositif est utilisé, « physical and social situation in which computational devices are embedded » [16].
- Dans un système centré utilisateur, les entités impliquées dans des interactions sont divisées en trois catégories : les lieux (bâtiments, pièces, etc.), les personnes (individus ou groupes d'individus) et les objets (objets physiques, ordinateurs, etc.) [17].

Au final, ils sont arrivés à une définition plus générique qui encapsule toutes les autres définitions : « toute information pouvant être utilisée pour caractériser la situation d'une

entité. Une entité est une personne, un lieu, ou un objet considéré comme pertinent pour l'interaction entre l'utilisateur et l'application, y compris l'utilisateur et l'application eux-mêmes » [18].

1.2.1.1 Les Caractéristiques du Contexte

Le tableau (1.1)[18] suivant, nous montre par quoi le Contexte est caractérisé

Caractéristiques du Contexte			
L'imprévisibilité	L'hétérogénéité	La variabilité	La continuité
Les changements de l'environnement et les interactions de l'utilisateur sont imprévisibles, L'une des principales questions qui se pose est de savoir comment un système doit prendre en compte ces nouveaux éléments et les mettre en œuvre dans le système	C'est-à-dire les données du Contexte sont d'origine de source variée ce qui apporte une grande hétérogénéité en termes de modélisation.	Dans le Contexte dynamique, le fonctionnement des systèmes des mobiles peut être perturbé par la modification du Contexte, c'est pour cette raison, il faut prendre en compte l'évolution du Contexte dans le but de maintenir la qualité de service et satisfaire le client.	Pendant le déplacement d'un utilisateur d'une localisation à une autre, le Contexte évolue, cette continuité doit s'accompagner d'un fonctionnement continu.

TABLE 1.1 – Caractéristiques du Contexte [18]

1.2.1.2 Catégorisation du Contexte

Le travail de **Van Bunningen et al.** classe les différentes catégorisations du Contexte en deux grandes familles : catégorisation conceptuelle et catégorisation opérationnelle[11].

1.2.1.2.1 Catégorisation Conceptuelle :

Elle permet de fournir des méthodes assez claires et structurées pour organiser le Contexte. Elle regroupe les catégorisations qui sont basées sur le sens du Contexte et qui distinguent les différents types de Contexte au niveau conceptuel. Les travaux qui appartiennent à cette famille ont d'abord restreint la notion de Contexte aux deux paramètres : localisation et temps. Ensuite, cette notion a été enrichie par la considération d'un ensemble de nouvelles dimensions .Comme nous le voyons dans (le tableau 1.3) .

1.2.1.2.2 Catégorisation Opérationnelle :

Elle est basée sur la façon dont le Contexte a été modélisé, acquis et traité, les travaux qui appartiennent à cette famille sont les suivants :

1. **Physique (externe)** : comme la localisation GPS.
2. **Logique (interne)** : comme les noms des rues.

Pour mieux comprendre les catégorisations du Contexte, nous présentons un tableau comparatif : « voir Table 1.2 [13] [19] [20] [21] [22] [23] [24] [25] [26] »

	Catégorisation conceptuelle	Catégorisation opérationnelle
Schilit et al	Où, Qui, Quel(les)	-
Ryan et al	Identité, Localisation, Temps, Environnement	-
Abowd et al	Identité, Localisation, Temps, Activité	-
Chen et al	Utilisateur, Physique, Calcul, Temps	-
Wang et al	-	Bas niveau, Haut niveau
Henricksen	-	Capturé, Statique, Dynamique, Dérivé
Prekop et al et Hofer et al	-	Physique(externe), Logique (interne)
Chang Xu	-	Physique, Logique
Guan et al	-	Bas niveau, Haut niveau
Chong et al	Calcul, Physique, Historique, Identité, Temps	
Zhong	Utilisateur, Système, Environnement, Social, Temps	-
Rizou et al	-	Bas niveau, Haut niveau

TABLE 1.2 – Comparaison entre la Catégorisation Conceptuelle et Opérationnelle

1.2.2 Sensibilité au Contexte et Systèmes Sensibles au Contexte

Le terme ‘*Context-Awareness*’ (Sensibilité au Contexte) vient des travaux sur l’informatique pervasive, ou informatique ubiquitaire. Ces systèmes utilisent les modifications de l’environnement dans les systèmes informatiques. Même si c’est un terme d’origine informatique, il a également été appliqué à la théorie des entreprises dans le domaine de la gestion des processus métiers. Il est introduit donc pour la première fois par **Schilit** and **Theimer** en 1994 au travers de la conception d’un système d’active map service, proposant des informations à un utilisateur en fonction de sa localisation. Ils définissent

alors la sensibilité au Contexte comme la capacité des applications d'un utilisateur mobile à découvrir et à réagir aux changements survenant dans l'environnement où ils sont situés[60].

Schilit et al définit ensuite 4 catégories d'applications Sensible au Contexte, selon que la tâche à réaliser consiste uniquement à obtenir des informations ou à réaliser des actions, et que cette tâche soit réalisée manuellement ou automatiquement (voir Table 1.3)[63].

	Manuelle	Automatique
Information	Sélection à proximité	Reconfiguration Contextuelle automatique
Commande	Commande Contextuelle	Action Contextuellement déclenchées

TABLE 1.3 – Catégories d'Applications Sensibles au Contexte d'après Schilit et al

Salbert et al Définissent la Sensibilité au Contexte comme étant la meilleure capacité d'un système à agir en temps réel avec des données provenant du Contexte [55].

D'autres définitions sont plus orientées vers l'adaptation au Contexte : Brown dit d'une application Sensible au Contexte qu'elle doit automatiquement extraire de l'information ou effectuer des actions en fonctions du Contexte utilisateur détecté par les capteurs[58]. Autour des années 2000, Chen et Kotz ont également donné deux définitions pour la Sensibilité au Contexte :

La sensibilité active : toute application ayant la capacité de s'adapter automatiquement au Contexte découvert par le changement du comportement de l'application.

La sensibilité passive : toute application ayant la capacité de rendre un Contexte pertinent pour une utilisation ultérieure.

Enfin Dey et Abowd proposent qu'un système soit Sensible au Contexte s'il utilise des informations du Contexte pour mettre à disposition des informations ou des services utiles à l'utilisateur, l'utilité dépendant de la tache de l'utilisateur. Cette définition a été adoptée par plusieurs chercheurs dans le domaine de l'informatique ubiquitaire [59].

Miraoui considère un Système Sensible au Contexte s'il possède la capacité de changer automatiquement les formes de ses services ou de déclencher un service à la suite d'un changement au niveau d'une information ou un ensemble d'informations caractérisant le service [56].

1.2.3 Modélisation du Contexte

Un modèle général du Contexte a une signification pratique importante pour le développement de systèmes Sensibles au Contexte. Il peut réduire la charge de travail des développeurs et augmenter l'efficacité du développement.

Dans la partie suivante, les méthodes principales de modélisation du Contexte vont être introduites.

1.2.3.1 Modélisation par Paire Attribut / Valeur (Key-value Pair Model)

Un des premiers travaux sur la modélisation du Contexte est publié par **Schilit** Les informations du Contexte sont modélisées par une paire attribut/valeur. L'attribut représente le nom d'une information Contextuelle et la valeur représente la valeur actuelle de cette information Par exemple, nous pouvons décrire un Contexte Context1 comme suit : (Name=Context1, User=x, Localisation=y, Time=t). L'avantage de cette méthode est la facilité d'implantation. En effet, la gestion du Contexte revient à parcourir la liste des Contextes disponibles. Cependant, ce modèle manque d'expressivité et de complétude [27] [28] [29].

1.2.3.2 Modélisation Par Mots Clés

Il s'agit de la structure de données la plus simple pour modéliser le Contexte. Elle utilise des paires de type 'clé-valeur' pour représenter l'information Contextuelle (par exemple 'temps-10h00'). Elle est simple d'utilisation mais très limitée en termes d'expressivité et d'évolution du modèle .

1.2.3.3 Modélisation par Langages à Balise

Le Contexte est modélisé par une structure de données hiérarchique constituée de balises avec des attributs et des contenus , les langages utilisés ici sont SGML et en

particulier XML . L'avantage du langage de balises est qu'il permet la mise en forme et l'échange de modèles entre plusieurs plateformes différentes facilement. Cependant, cette méthode de modélisation reste non intuitive et difficile à utiliser pour décrire des informations plus complexes. Parmi ces modèles, nous pouvons citer à titre d'exemples : Centaurus Capability Markup Language (CCML) , Comprehensive Structured Context Profiles (CSCP) et Context Description Framework (CDF) [30] [31] [32] .

1.2.3.4 Modélisation Par Graphe

Le modèle graphique utilise les outils orientés vers les graphiques pour modéliser les informations Contextuelles. IL se concentre sur les descriptions des structures des informations Contextuelles. UML (Unified Modeling Language) peut être utilisé comme un langage de modélisation graphique grâce à sa forte capacité de représentation graphique. Le modèle typique est ORM (Object-Role Modeling) proposé par Henrichen et al, ORM est basé sur des faits, selon la longueur de la durée du fait, les faits peuvent être divisés en deux types, un type de fait stable et un type de fait dynamique. Bien que le modèle graphique puisse décrire les relations entre les entités Contextuelles de manière graphique, il présente des limitations dans l'application pratique.

Tout d'abord, il n'y a pas un logiciel universel ni un outil qui peut supporter toutes les représentations du Contexte, et ensuite il manque une norme uniforme, qui résout le problème des différences importantes entre les notions Contextuelles et les expressions. Donc, ce modèle ne favorise pas l'inférence des informations Contextuelles d'une manière plus poussée [33].

1.2.3.5 Modélisation Orienté Objet

La conception orientée objet peut être introduite pour la modélisation des informations Contextuelles, c'est-à-dire, les informations peuvent être encapsulées et réutilisées suivant leur classe, les informations qui ont des attributs et les méthodes similaires peuvent être encapsulées comme une classe. Comme la classe a la capacité d'encapsulation et d'héritage, non seulement les duplications des définitions peuvent être évitées facilement, mais aussi les descriptions de la qualité des attributs des informations Contextuelles et les descriptions de complétude peuvent être envisagées.

Un modèle typique est le CML (Context Modeling Language) proposé dans la littérature. Ses avantages sont : l'encapsulation et la réutilisabilité. Cependant, ces modèles souffrent de deux inconvénients majeurs : le manque de capacités de raisonnement ainsi que la difficulté de validation des conceptions orientées objet en raison de l'absence de normes et de standards dédiés [34].

1.2.3.6 Modélisation Basée Sur la Logique

Cette méthode est une méthode de modélisation d'information Contextuelle à un niveau abstrait, il est proposé par (McCarthy et Buvac, 1997). Il gère les faits, les expressions et les règles pour définir le modèle Contextuel. Ce type de modèles est plus expressif comparé aux modèles précédents, mais il est normalement couplé à l'application [35][36].

1.2.3.7 Modélisation Basée Sur Les Ontologies

Cette approche utilise les Ontologies pour modéliser le Contexte, en utilisant une sémantique telle que : RDF , RDF'S et OWL , elle offre plusieurs avantages comme [37] :

- Elle offre la capacité du raisonnement
- Divers outils de développement et moteur de raisonnement sont disponibles

D'une autre part, l'inconvénient principal dans cette approche, c'est la représentation peut être complexe et la récupération d'informations peut être soumis à forte intensité de calcul si le nombre de données est élevé.

1.3 Le Web Sémantique

1.3.1 Définition du Web Sémantique

L'expression Web Sémantique, due à Tim Berners-Lee au sein du W3C, fait d'abord référence à la vision du Web de demain comme un vaste espace d'échange de ressources entre êtres humains et machines permettant une exploitation, qualitativement supérieure, de grands volumes d'informations et de services variés. Le Web actuel est essentiellement syntaxique, dans le sens que la structure des documents (ou ressources au sens large) est bien défini, mais que son contenu reste quasi-inaccessible aux traitements machines. Seuls les humains peuvent interpréter leurs contenus. La nouvelle génération de Web – Le Web Sémantique – a pour ambition de lever cette difficulté. Les ressources du Web seront plus

aisément accessibles aussi bien par l'homme que par la machine, grâce à la représentation sémantique de leurs contenus[38].

Le Web Sémantique, concrètement, est d'abord une infrastructure pour permettre l'utilisation de connaissances formalisées en plus du contenu informel actuel du Web, même si aucun consensus n'existe sur jusqu'où cette formalisation doit aller. Cette infrastructure doit permettre d'abord de localiser, d'identifier et de transformer des ressources de manière robuste et saine tout en renforçant l'esprit d'ouverture du Web avec sa diversité d'utilisateurs.

1.3.2 Les Composants du Web Sémantique

Le Web Sémantique admet deux composantes distinctes. La première composante est l'effort de lier le contenu existant aux significations sémantique et, ultérieurement, la création d'un ensemble d'applications qui utilisent les métadonnées créées. Le mécanisme d'ajout du contenu sémantique est accompli à travers le modèle standard de description des ressources appelé RDF 1 , RDF est une spécification qui définit le modèle de données pour la représentation des objets du monde "ressources" et les relations entre eux. Ce modèle permet aux développeurs de lier les "mots" dans les sites Web, les documents et les bases de données avec un vocabulaire décentralisé qui définit les significations de ces mots. Ce vocabulaire décentralisé est le cœur du WS et il est référé fréquemment aux Ontologies. Les métadonnées RDF concernant une ressource Web peuvent être considérées comme des annotations sémantiques et le schéma RDF comme une Ontologie. Les métadonnées peuvent être disponibles sous d'autres formes telles que des définitions de type de documents (DTD), Schéma XML, etc[39][40].

La figure 1.1 suivante représente la pile du Web Sémantique .

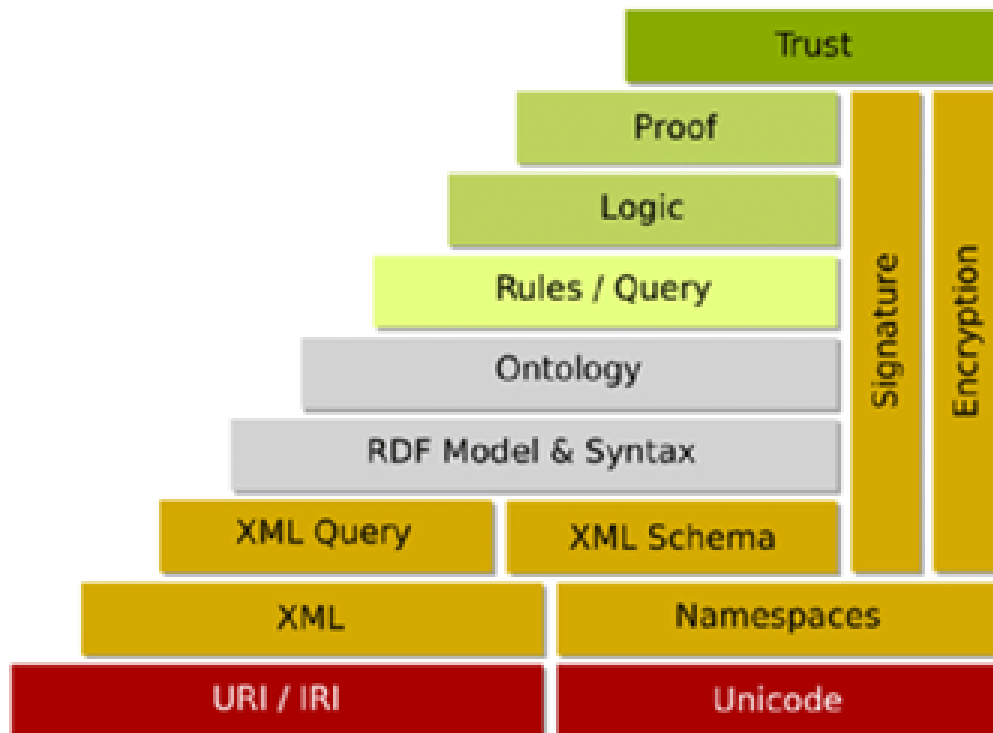


FIGURE 1.1 – Les Composants du Web Sémantiques [41]

1.3.3 Les Langages du Web Sémantique

Les assertions affirment l'existence de relations entre des objets. Elles sont donc adaptées à l'expression des annotations que l'on veut associer aux ressources du Web. On évoquera principalement RDF ici, car il nous semble présenter des avantages déterminants pour la manipulation informatique.

1.3.3.1 RDF (Resource Description Framework)

RDF C'est un langage formel qui permet d'affirmer des relations entre des « ressources ». Il sera utilisé pour annoter des documents écrits dans des langages non structurés, ou comme une interface pour des documents écrits dans des langages ayant une sémantique équivalente (des bases de données, par exemple). Un document RDF est un ensemble de triplets de la forme. Les éléments de ces triplets peuvent être des URIs (Universal Resource Identifiers des littéraux ou des variables). Cet ensemble de triplets peut être représenté

de façon naturelle par un graphe (plus précisément un multi-graphe orienté étiqueté), où les éléments apparaissant comme sujet ou objet sont les sommets, et chaque triplet est représenté par un arc dont l'origine est son sujet et la destination son objet. Ce document sera codé en machine par un document RDF/XML, mais est souvent représenté sous une forme graphique [38][39] .

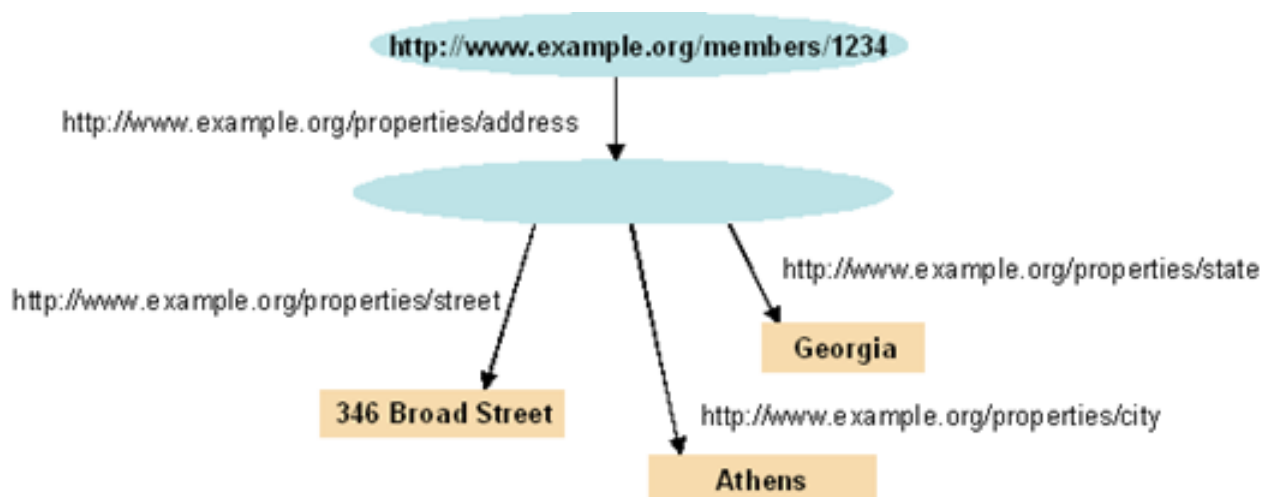


FIGURE 1.2 – Exemple d'un Graphe RDF [42]

Dans la « figure 1.2 » précédente, nous avons montré un simple exemple d'un graphe RDF , nous remarquons qu'il contient un objet vide au même temps un sujet vide avec des objets de type « chaîne de caractère »

L'objectif du RDF est de donner des contraintes sur les mondes qui peuvent être décrits par un document RDF. L'utilisation de la théorie des ensembles pour décrire ces modèles a deux intérêts : la généralité de la notion d'ensemble (fondement des mathématiques) et son universalité (culture commune pour ceux qui vont s'intéresser à cette sémantique).

1.3.3.2 RDFS (Resource Description Framework Schema)

RDFS a pour but d'étendre le langage en décrivant plus précisément les ressources utilisées pour étiqueter les graphes. Pour cela, il fournit un mécanisme permettant de spécifier les classes dont les ressources seront des instances, comme les propriétés. RDFS s'écrit toujours à l'aide de triplets RDF, en définissant la sémantique de nouveaux mots-clés comme [43] :

«`ex :Vehicule rdf:type rdfs:Class`» la ressource `ex : Vehicule` a pour type `rdfs : Class`, et est donc une classe

Ce besoin de spécifier davantage les classes est à l'origine du langage dédié aux définitions de classes : OWL.

1.3.3.3 DAML + OIL

DAML+OIL est la fusion de deux langages de représentation des connaissances DAML7 et OIL basés essentiellement sur les logiques de descriptions et qui ont été proposés comme langage de description d'Ontologies. Le but de DAML+OIL est d'étendre RDFS en lui ajoutant des primitives, plus expressives pour la définition des classes et des propriétés d'une Ontologie [44] [45] [46] .

1.3.3.4 OWL (Ontology Web Language)

RDF, langage dédié à l'expression d'assertions sur les relations entre objets, s'est heurté à la nécessité de définir les propriétés des classes dont ces objets sont instances. Cependant, l'extension à RDFS ne fournit que des mécanismes primitifs pour spécifier ces classes. Le langage, quant à lui, est dédié aux définitions de classes et de types de propriétés, et donc à la définition d'Ontologies. Inspiré des logiques de descriptions (et successeur de DAML+OIL, il fournit un grand nombre de constructeurs permettant d'exprimer de façon très fine les propriétés des classes définies. La rançon de cette expressivité est l'indécidabilité du langage obtenu en considérant l'ensemble de ces constructeurs. C'est pour cela que OWL a été fractionné en trois langages distincts [47] :

- **OWL Lite** : c'est la version légère de OWL qui reprend RDFS et l'enrichit avec de nouvelles primitives.
- **OWL DL** : contient toutes les primitives de OWL (y compris OWL Lite) avec des contraintes particulières sur leur utilisation qui assurent la décidabilité du langage.
- **OWL Full** : plus flexible que OWL DL ce qui le rend vraisemblablement indécidable.

OWL est basé essentiellement sur le formalisme des logiques de descriptions et tire profit des inférences et des mécanismes de raisonnements associés à ces formalismes.

1.4 Conclusion

Dans ce premier chapitre, nous avons d'abord abordé, de manière un peu détaillée la notion de Contexte, nous l'avons vu qu'elle a été étudiée par plusieurs communautés de recherche en informatique, ensuite nous avons également présenté les caractéristiques et les différentes catégorisations ainsi les différentes approches de modélisation du Contexte.

Au final, nous avons présenté la définition du Web Sémantique et ses différents langages.

Dans le chapitre suivant , nous allons aborder le Machine Learning en général, en représentant les différentes méthodes d'apprentissage automatique ainsi ses Algorithmes populaires.

Machine Learning

2.1 Introduction

L'intelligence artificielle (IA) est de nos jours, l'une des évolutions technologiques les plus médiatisées, elle est définie comme un domaine de recherche scientifique destiné à reproduire le raisonnement humain, en se basant sur la notion d'apprentissage. L'accès à des données massives et aux calculateurs super puissant a permis, l'émergence de plusieurs formes de l'IA, comme l'apprentissage artificiel (Machine Learning) puis, l'apprentissage profond (deep Learning ou DL).L'apprentissage artificiel se résume en un ensemble d'algorithmes automatisés qui apprennent à modéliser des fonctions et à prévoir les actions futures à partir de données structurées, alors que l'apprentissage profond s'apparente à un réseau de neurones artificiels tels que les réseaux neuronaux d'un cerveau humain.

Dans ce chapitre, nous proposons un état de l'art sur la machine Learning, tout d'abord nous allons établir plus clairement ce qui relève de l'apprentissage automatique (Machine Learning) en commençant par définir l'apprentissage automatique, puis nous allons expliquer ses approches. (Types).

En outre, nous allons expliquer les différents et populaires algorithmes pour modéliser un modèle du machine Learning en basant sur leurs principaux avantages ainsi leurs limites.

Finalement, nous avons cité les différentes méthodes d'évaluation d'un modèle du

machine Learning ainsi les différentes métriques.

2.2 Machine Learning

2.2.1 Définition du Machine Learning

Le terme « Machine Learning » ou ML en bref, a été inventé en 1959 par Arthur Samuel dans le Contexte de la résolution de jeux de dames par machine. Le terme fait référence à un programme informatique qui peut apprendre à produire un comportement qui n'est pas explicitement programmé par l'auteur du programme. Il est plutôt capable de montrer un comportement dont l'auteur peut ne pas être complètement conscient. Ce comportement est appris basé sur trois facteurs :

- Les données qui sont consommées par le programme.
- Une métrique qui quantifie l'erreur ou une certaine forme de distance entre le comportement actuel et le comportement idéal.
- Un mécanisme de rétroaction qui utilise l'erreur quantifiée pour guider le programme pour produire un meilleur comportement lors des événements ultérieurs.

Comme nous pouvons le voir, le deuxième et troisième facteur rendent rapidement le concept abstrait et en soulignent les racines mathématiques profondes. Les méthodes de la théorie de l'apprentissage automatique sont essentielles pour construire des Systèmes artificiellement intelligents [48] .

2.2.2 Types du Machine Learning

Les approches d'apprentissage en profondeur peuvent être classées comme suit : supervisées, semi- supervisées ou partiellement supervisé et non supervisé. De plus, il existe une autre catégorie d'approche d'apprentissage appelé Renforcement Learning (RL) ou Deep RL (DRL) qui sont souvent abordés dans le cadre de semi-supervisé ou parfois dans le cadre d'approches d'apprentissage non supervisées.

2.2.2.1 L'Apprentissage Supervisé

Dans ce type d'apprentissage, les entrées et les sorties sont fournies au préalable. Ensuite, le réseau traite les entrées et compare ses résultats aux sorties souhaitées. Les

pois sont ensuite ajustés grâce aux erreurs propagées à travers le Système. Ce processus se produit à plusieurs reprises tant que les poids sont continuellement améliorés. L'ensemble de données qui permet l'apprentissage est appelé l'ensemble d'apprentissage [49]. La figure 2.1 montre comment l'apprentissage automatique est défini.

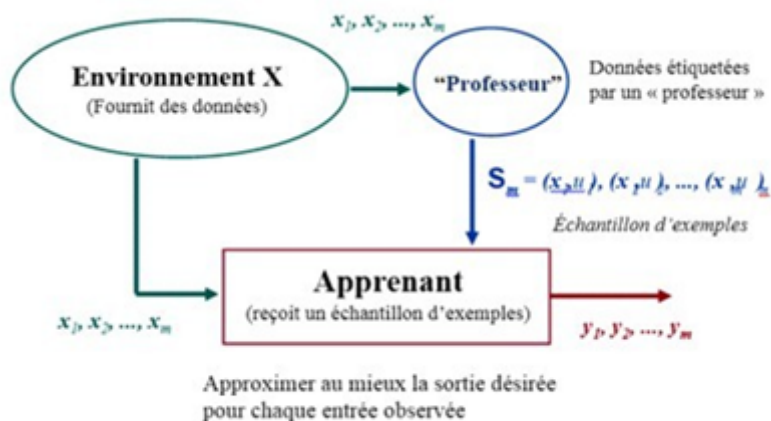


FIGURE 2.1 – Schéma d'Apprentissage Supervisé [50]

2.2.2.2 Apprentissage Semi-Supervisé profond

L'apprentissage Semi-Supervisé est un Apprentissage qui se produit sur la base d'ensembles de données partiellement étiquetés. Dans certains cas, les DRL et les Réseaux Adversaires Génératifs (GAN) sont utilisés comme techniques d'apprentissage Semi-Supervisé [51].

2.2.2.3 Apprentissage Non Supervisé

Dans l'apprentissage non supervisé, le réseau est fourni avec des entrées, mais pas avec les sorties souhaitées. Le Système lui-même doit alors décider quelles fonctionnalités il utilisera pour regrouper les données d'entrée. C'est ce qu'on appelle souvent l'auto-organisation ou l'adaptation [49].

2.2.2.4 Apprentissage Par Renforcement

Le Système est invité à entreprendre n'importe quelle action et reçoit une récompense ou une pénalité en fonction de l'adéquation de cette action à la situation donnée. Le Système doit apprendre quelles actions rapportent le plus de récompenses dans des situations données au fil du temps. Avec ces techniques, le domaine de l'apprentissage automatique s'est épanoui. Ils ont particulièrement bien réussi dans les domaines de la vision par ordinateur et de l'analyse de texte. Au fil des ans, de nombreux modèles ont été introduits pour mettre en œuvre des techniques d'apprentissage automatique, tels que les réseaux de neurones artificiels (modèles inspirés du fonctionnement des neurones du cerveau), les arbres de décision (modèles qui utilisent des structures en forme d'arbre pour modéliser les décisions et les résultats). Modèles de régression (modèles qui utilisent des méthodes statistiques pour mapper les variables d'entrée et de sortie), etc [48] .

Vers 2010, certaines choses se sont produites qui ont influencé la recherche sur l'apprentissage automatique :

- La puissance de calcul est devenue plus disponible et l'évaluation de plus les modèles complexes sont devenus plus faciles.
- Le traitement et le stockage des données sont devenus moins chers. Plus de données sont devenues disponibles à la consommation.
- Notre compréhension du fonctionnement du cerveau naturel a augmenté, nous permettant de modéliser de nouveaux algorithmes d'apprentissage automatique autour d'eux [48] .

Le tableau 2.1 [52] montre la différence entre les 4 types de Machine Learning :

Apprentissage Supervisé	Apprentissage Non Supervisé	Apprentissage Semi-Supervisé	Apprentissage Par Renforcement
Les experts de données (data) fournissent des entrées, des sorties et des commentaires pour construire le modèle (comme définition)	utiliser l'apprentissage profond pour arriver à des conclusions et à des modèles grâce à données de formation étiquetées.	Construit un modèle à travers un mélange de données étiquetées et non étiquetées, un ensemble de catégories, suggestions et exemples d'étiquettes	Auto-interprétation, mais basé sur un Système de récompenses et de punitions apprises par essais et erreurs, recherche d'une récompense maximale.
EXEMPLE D'ALGORITHMES	EXEMPLE D'ALGORITHMES	EXEMPLE D'ALGORITHMES	EXEMPLE D'ALGORITHMES
<p>Régressions linéaires</p> <ul style="list-style-type: none"> — prévisions des ventes — évaluation des risques <p>Machine à vecteurs de support</p> <ul style="list-style-type: none"> — classification des images — Comparaison des performances financières <p>Arbre de décision</p> <ul style="list-style-type: none"> — Analyse prédictive — Tarification 	<p>Apriori</p> <ul style="list-style-type: none"> — fonctions de vente — associations de mots — searcher <p>Clustering K-means</p> <ul style="list-style-type: none"> — Surveillance des performances — intention du chercheur 	<p>création de données</p> <ul style="list-style-type: none"> — Réseaux antagonistes génératifs — manipulation audio et vidéo <p>Naive Bayes autodidacte classificateur</p> <ul style="list-style-type: none"> — traitement du langage naturel 	<p>Q-Learning</p> <ul style="list-style-type: none"> — politique création — diminution de la consommation <p>Valeur basée sur un modèle estimation</p> <ul style="list-style-type: none"> — tâches linéaires — estimation des paramètres

TABLE 2.1 – Différence Entre Les 4 Types de Machine Learning

2.2.3 Les Algorithmes De Machine Learning

2.2.3.1 Les Arbres De Décision

À la fin des années 70s et au début des années 80s, J. Ross Quinlan, un chercheur a développé un algorithme d'arbre de décision connu sous le nom d'ID3 (Iterative dichotomiser). Le modèle d'arbres de décision fait partie des modèles d'apprentissage statistique automatique basée sur une approche géométrique pour segmenter l'espace de caractéristiques. L'algorithme consiste à construire un ensemble de classeurs, appelés classeurs faibles, qui sont séquentiellement introduits avec l'intention de faire une segmentation de l'espace de caractéristiques permettant de classer les données. Cette séquence de classeurs est ensuite représentée comme une série de nœuds connectés et placés hiérarchiquement. Le résultat final est un arbre qui représente la séquence de classeurs en plaçant dans la partie supérieure le nœud correspondant au classeur dont la capacité de classification est la plus grande entre tous les classeurs et, en allant vers le bas des autres nœuds, en représentant d'autres classeurs faibles, en ordre d'importance décroissante. Le résultat final est un arbre renversé [53].

2.2.3.1.1 Les Avantages Et Les Inconvénients Des Arbres De Décision [53]

Arbre de décision	
Avantages	Inconvénients
Ils permettent de faire une représentation graphique du processus de décision	Ils ont une performance plus mauvaise que d'autres techniques d'apprentissage statistique.
Facile à interpréter pour toute personne	Ils ont été peu utilisées jusqu'à l'apparition d'une nouvelle techniques (forêts aléatoires)
Il est considéré comme un miroir du processus humain de prise de décision	Son apprentissage peu amener des arbres de décisions plus complexes qui généralise mal l'ensemble d'apprentissage

TABLE 2.2 – Les Avantages et les Inconvénients Des Arbres De Décision

2.2.3.1.2 Les Limites d'Arbres De Décisions

- Ils peuvent être aussi complexes, ils ne généralisent pas bien (overfitting :surapprentissage) [53]
- Ils existe des concepts qui sont un peu difficile à apprendre par les arbres de décision. Ils ne sont pas faciles à exprimer, par exemple : XOR [53]

2.2.3.2 Random Forest (Forêts Aléatoires)

Les arbres de décisions produisent des classeurs avec un petit biais, mais avec une grande variance ceci fait que les arbres tendent à sur adapter les données d'entraînement , Leo.Breiman en 2001 a résolu ce problème en introduisant une nouvelle technique qui est "les forêts aléatoires (Random forest)" [54] .

C'est un algorithme qui se base sur l'assemblage d'arbres de décision. Il est assez intuitif à comprendre, rapide à entraîner et il produit des résultats généralisables.

2.2.3.2.1 Principe Fonctionnement Du Random Forest

- Un random forest est constitué d'un ensemble d'arbres de décision indépendants.
- Chaque arbre dispose d'une vision parcellaire du problème du fait d'un double tirage aléatoire :
 - Un tirage aléatoire avec remplacement sur les observations (les lignes de votre base de données). Ce processus s'appelle le tree bagging [54] .
 - Un tirage aléatoire sur les variables (les colonnes de votre base de données). Ce processus s'appelle le feature sampling [54] .

D'une manière plus claire et simple le RandomForest =tree bagging + feature sampling. Le Random forest fonctionne sur ce même principe : plutôt que d'avoir un estimateur complexe capable de tout faire, le random forest utilise plusieurs estimateurs simples (de moins bonnes qualités individuelles). Chaque estimateur a une vision parcellaire du problème. Ensuite, l'ensemble de ces Estimateurs est réuni pour obtenir la vision globale du problème. C'est l'assemblage de tous ces estimateurs qui rend extrêmement performante la prédiction [54] .

2.2.3.2.2 Tree Bagging le bagging signifie "bootstrap aggregation" :c'est un processus de tirage aléatoire sur les observations (lignes de données) ,son fonctionnement est

déterminé par trois étapes clés :

- Construction de n arbres de décisions en tirant aléatoirement n échantillons d'observations,
- Entraînement de chaque arbre de décision.
- Pour faire une prévision sur de nouvelles données, il faut appliquer chacun de n arbres et prendre la majorité parmi les n prévisions [54] .

2.2.3.2.3 Feature Sampling C'est un processus de tirage aléatoire sur les variables (les colonnes de données). Par défaut, nous tirons une variable pour un problème à n variables au total, il permet de baisser la corrélation entre les arbres qui pourraient perturber la qualité des résultats. Il permet aussi de réduire la variance de l'ensemble créé [54].

2.2.3.2.4 Critère de Division/Split Critère d'un arbre de décisions construit des sous-populations par séparations successives des feuilles d'un arbre. Il existe différents critères de séparation pour construire un arbre :

— **Critère de Gini**

Cet indice mesure l'impureté, qui est un concept très utile dans la construction des arbres de décision : la qualité d'un nœud et son pouvoir discriminant peuvent être évalués par son impureté. L'indice Gini est donné par la relation suivante :

$$Gini(T) = 1 - \sum_{j=1}^m \left(\frac{T_j}{T}\right)^2 \quad (2.1)$$

— **le critère d'entropie**

Soit une distribution de probabilité $P = (p_1, p_2, \dots, p_n)$. L'entropie de la distribution P est la quantité d'information qu'elle peut apporter. Elle est donnée par l'équation

$$E(P) = - \sum_{i=1}^n p_i \log p_i \quad (2.2)$$

Après avoir expliqué le fonctionnement et le principe de l'algorithme random forest, maintenant nous allons présenter ses avantages et ses inconvénients.

2.2.3.2.5 Avantages et Inconvénients des Forêts Aléatoires [53]

Forêts Aléatoires	
Avantages	Inconvénients
Pas de sur-apprentissage	Boite noire : difficilement interprétable, difficilement améliorable
En général : meilleure performance que les arbres de décision, calcul de l'erreur "Out-of- Bag" direct	Entraînement plus lent.
Effet 2 en 1 : validation croisée non-nécessaire grâce aux échantillons OOB(Out Of Bage)	un outil de modélisation prédictive et non un outil descriptif
Paramètres faciles à calibrer.	
Parallélisation possible souvent utilisée comme benchmark dans les compétitions de Machine Learning .	

TABLE 2.3 – Les Avantages et Les Inconvénients des Forêts Aléatoires

2.2.3.3 SVM (Support Vector Machine)

Support Vector Machine ou SVM est l'un des algorithmes d'apprentissage supervisé les plus populaires, utilisé pour les problèmes de classification et de régression. Cependant, il est principalement utilisé pour les problèmes de classification dans le Machine Learning. Le but de l'algorithme SVM est de créer la meilleure ligne ou frontière de décision qui puisse séparer l'espace à n dimensions en classes afin que nous puissions facilement placer le nouveau point de données dans la catégorie correcte à l'avenir. Cette meilleure frontière de décision est appelée un hyperplan [55] .

SVM choisit les points/vecteurs extrêmes qui aident à créer l'hyperplan. Ces cas extrêmes sont appelés vecteurs de support, et par conséquent, l'algorithme est appelé Support Vector Machine. Considérez le diagramme (figure 2.2) ci-dessous dans lequel il existe deux catégories différentes qui sont classées à l'aide d'une frontière de décision ou d'un hyperplan :

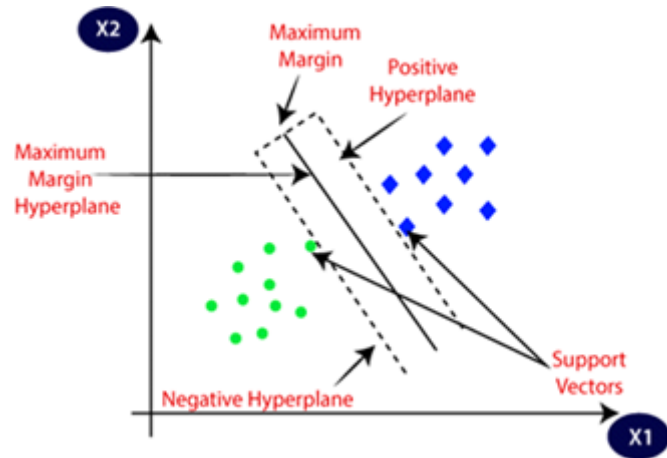


FIGURE 2.2 – Hyperplan Optimal Avec Une Marge Maximale [56]

Comme nous l'avons vu précédemment, nous choisirons l'hyperplan qui maximise la **marge**, c'est-à-dire la distance minimale entre les vecteurs d'entraînement et l'hyperplan. De tels vecteurs situés à la distance minimale sont appelés **vecteurs supports**.

2.2.3.3.1 Avantages et Inconvénients Du SVM [57]

Support Vecteur Machine	
Avantages	Inconvénients
Capacité à traiter de grandes dimensionnalités (variables élevées)	Problème lorsque les classes sont bruitées (multiplication des points supports)
Non paramétrique	Pas de modèle explicite pour les noyaux non linéaires (utilisation des points supports)
Traitement des problèmes non linéaires avec le choix des noyaux	Difficulté d'interprétations (ex. pertinence des variables)
Traitement des problèmes non linéaires avec le choix des noyaux	Le traitement des problèmes multi-classes reste une question ouverte

TABLE 2.4 – Avantages et Inconvénients Du SVM

2.2.3.4 Classification Naïve Bayésienne

Naïve Bayes est un algorithme d'apprentissage simple qui utilise le théorème de Bayes avec une forte supposition que les attributs sont conditionnellement indépendants, compte tenu de la classe. Bien que cette hypothèse d'indépendance soit souvent violée dans la pratique, les Bayes naïves offrent néanmoins souvent une Précision de classification compétitive. Associé à son efficacité de calcul et à de nombreuses autres caractéristiques souhaitables, cela conduit à des Bayes naïves largement appliquées dans la pratique [53]. Naïve Bayes fournit un mécanisme permettant d'utiliser les informations des données d'échantillon pour estimer la probabilité postérieure $P(y/x)$ de chaque classe y , étant donné un objet x . Une fois que nous avons ces estimations, nous pouvons les utiliser pour la classification ou d'autres applications d'aide à la décision [54].

Principe La construction d'un classifieur naïf de bayes est comme suite :

- On estime la probabilité jointe $P(X, Y)$.
- On calcule la probabilité conditionnelle de la classe c .

$$P(X/Y) = P(X)P(Y/X)/P(Y) \quad (2.3)$$

2.2.3.4.1 Avantages et Inconvénients de la Classification Naïve Bayes [58]

Classification Naïve Bayes	
Avantages	Inconvénients
Facilité et simplicité de leurs implémentations	Il implique que chaque fonctionnalité soit indépendante, ce qui n'est pas toujours le cas.
Leurs rapidités	ce type d'algorithmes permet de faire le même travail de classification que les autres algorithmes qui existent déjà (cause de l'hypothèse d'indépendance des mots)
La classification Naïve bayes donne un bon résultat	

TABLE 2.5 – Avantages et Inconvénients de la Classification Naïve Bayes

2.2.3.4.2 Parmi ses Limites Ces performances sont limitées quand il s'agit d'une grande quantité de lexiques à traiter. En effet, si le nombre de lexiques(classification textuelle) augmente, alors le nombre des dépendances entre l' ensemble des mots augmentent, et donc, la vérification de l'hypothèse de Naïve Bayes diminue [58] .

2.2.3.5 RNA (Réseaux De Neurones)

Un réseau de neurones est une méthode de calcul à base de fonctions mathématiques. En effet, les réseaux de neurones sont une méthode de l'intelligence artificielle optimisée par des méthodes d'apprentissage de type probabiliste. Nous accompagnons cette définition par un schéma (figure 2.3).

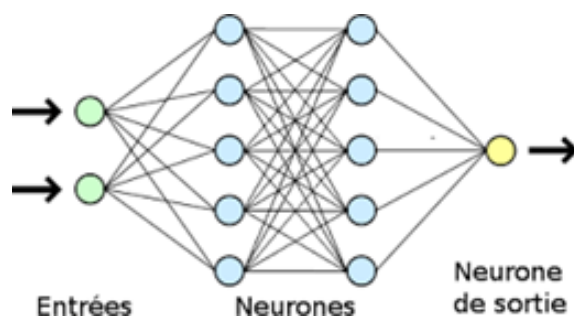


FIGURE 2.3 – Exemple d'Un Réseau Neurone [59]

2.2.3.5.1 L'Architecture d'un Réseau Neurone :

Les architectures ont leurs forces et faiblesses et peuvent être combinées pour optimiser les résultats. Le choix de l'architecture s'avère ainsi crucial et il est déterminé principalement par l'objectif. Un réseau de neurones peut prendre des formes différentes selon l'objet de la donnée qu'il traite et selon sa complexité et la méthode de traitement de la donnée[60].(figure 2.4 ,nous montre son architecture).

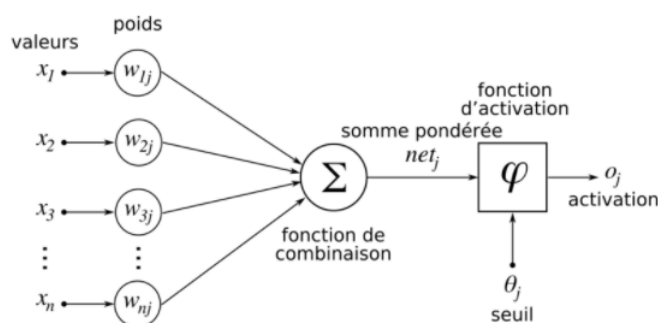


FIGURE 2.4 – Architecture d'un Réseau Neurone [60]

2.2.3.5.2 Fonctionnement d'un Réseau de Neurone :

En règle générale, un réseau de neurones repose sur un grand nombre de processeurs opérant en parallèle et organisés en tiers. Le premier tiers reçoit les entrées d'informations brutes, un peu comme les nerfs optiques de l'être humain lorsqu'il traite des signaux visuels. Par la suite, chaque tiers reçoit les sorties d'informations du tiers précédent. On retrouve le même processus chez l'Homme, lorsque les neurones reçoivent des signaux en provenance des neurones proches du nerf optique. Le dernier tiers, quant à lui, produit les résultats du Système « voir la figure 2.5 ».

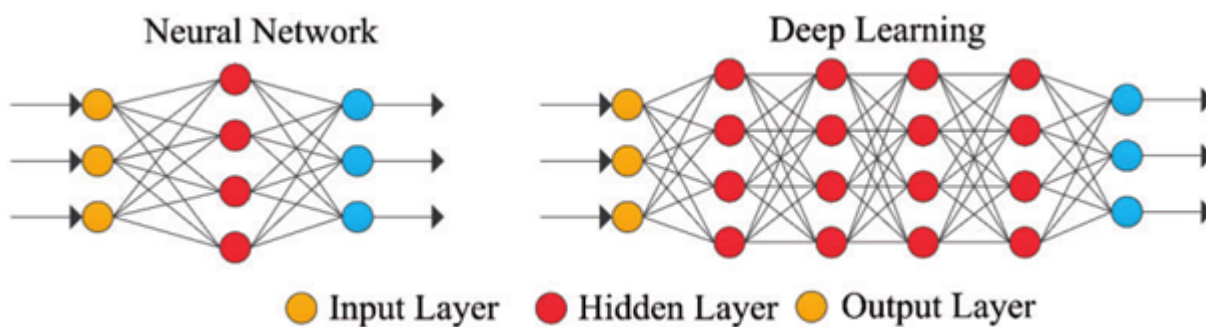


FIGURE 2.5 – Fonctionnement d'un Réseau Neurone [61]

2.2.3.5.3 Avantages et Inconvénients des Réseaux de Neurones [58]

Réseaux de Neurones	
Avantages	Inconvénients
La rapidité et l'efficacité de traitement des grands corpus	La lenteur d'apprentissage.
La possibilité de combiner ce type d'algorithmes avec d'autres méthodes de classification	Les résultats obtenus par la classification des réseaux de neurones ne sont pas interprétables
un taux d'erreur très faible par rapport aux autres méthodes de classification	Il est considéré comme une boîte noire
Les réseaux de neurones ne nécessitent pas l'utilisation de modèles mathématiques très complexes pour leur fonctionnement	La convergence des résultats des réseaux de neurones est incertaine

TABLE 2.6 – Avantages et Inconvénients des Réseaux de Neurones

2.3 Méthode Dévaluation d'Un Modèle du Machine Learning

L'évaluation du modèle joue un rôle crucial lors du développement d'un modèle d'apprentissage automatique prédictive. Construire simplement un modèle prédictif sans vérification ne compte pas comme un modèle d'ajustement, mais un modèle qui donne une Précision maximale en compte sûrement un bon. Nous allons citer quelques-unes.

2.3.1 La Matrice De Confusion

Est également appelée matrice d'erreur et est représentée par un tableau « voir Table 2.7 [62] » qui décrit les performances d'un modèle de classification sur un ensemble de données de test dans l'apprentissage automatique. Dans le tableau ci-dessus, la classe 1 est représentée comme le tableau positif et la classe 2 comme le tableau négatif.

Toute prédiction unique peut appartenir à l'un des quatre compartiments suivants [62] :

- * **Vrai positif (TP)** : le vrai label de l'instance donnée est positif, et le classificateur le prédit également comme positif [62] .
- * **Vrai négatif (TN)** : la vraie étiquette est négative et le classificateur prédit également un négatif [62] .
- * **Faux positif (FP)** : le vrai label est négatif, mais le classificateur le prédit à tort comme positif [62] .
- * **Faux négatif (FN)** : le vrai label est positif, mais le classificateur le prédit à tort comme négatif [62] .

Matrice de confusion		
	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

TABLE 2.7 – Matrice de Confusion

L'évaluation de notre algorithme d'apprentissage automatique est une partie essentielle de notre projet. Notre modèle peut nous donner des résultats satisfaisants lorsqu'il est évalué à l'aide d'une métrique. Dans notre cas, nous avons utilisé la précision de la classification (Accuracy) pour mesurer les performances de notre modèle.

2.3.1.1 l'Accuracy

Est la mesure de performance la plus intuitive et il s'agit simplement d'un rapport entre les observations correctement prédites et les observations totales

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \text{ [63].}$$

2.4 Conclusion

Nous avons présenté dans ce chapitre, l'apprentissage automatique d'une manière générale, puis nous avons présenté une vue détaillée de ses méthodes qui existent.

Ensuite, nous avons présenté les différents algorithmes de modélisation d'un modèle de Machine Learning ainsi ses méthodes de modélisations et ses différentes métriques.

Dans le chapitre suivant, nous allons expliquer en détails notre propre solution en donnant quelques expériences que nous avons fait.

Solution Proposée

3.1 Introduction

La sensibilité au Contexte est la meilleure capacité d'un système à agir en temps réel avec des données provenant du Contexte, elle permet au système d'interagir avec les utilisateurs par le biais de dispositifs sensibles au Contexte. Son principal objectif est de maximiser l'efficacité et la convivialité des services. Nous pouvons dire, qu'un système est considéré comme Sensible au Contexte lorsqu'il se base sur les différentes informations Contextuelles détectées ou déduites de l'environnement.

Dans ce chapitre, nous allons expliquer la problématique de notre travail en incluant notre principal objectif, ensuite nous allons parler sur les travaux connexes qui ont été utilisés pour la Sensibilité au Contexte.

En revanche, nous allons décrire et détailler notre proposition en parlant d'une manière générale sur les DataSets (expériences) avec qui nous avons testé et validé notre solution.

Finalement, nous terminerons de proposer une Architecture générale qui résume toutes les étapes que nous avons passées pour arriver à notre solution.

3.2 Problématique

L'intelligence artificielle a fait des progrès impressionnants, elle a permis au monde de la technologie de faire un grand pas en avant, surtout ces dernières années, grâce aux

grandes quantités de données collectées, et après avoir jeté un coup d'œil à ce que les ingénieurs en IA font avec ces données, nous avons constaté que certaines d'entre elles ne sont pas pleinement utilisées, et de là vient l'idée d'optimiser ce qu'ils font déjà en trouvant une approche qui nous permet d'utiliser pleinement ces données.

En général, lorsqu'un ingénieur en IA tente de résoudre un problème en Machine Learning ou Deep Learning, le résultat de son travail obtenu consiste un seul modèle (que nous l'appelons également « le modèle général ») qui résout le problème, ce modèle unique peut avoir une grande précision de ce qu'il fait, mais personne ne peut dire que sa précision est maximale.

Notre idée est qu'au lieu de créer un seul et unique modèle, nous allons créer plusieurs modèles (donc notre système Machine Learning devient Sensible au Contexte.), chacun d'eux est spécialisé dans une colonne et par cela, il sera Sensible de cette colonne. Après avoir créé ces modèles, nous les combinerons en espérant qu'ils nous donneront une meilleure précision que le modèle général.

Théoriquement, ils donneront toujours de meilleurs résultats parce que chaque modèle traite moins de données que le modèle général, et ces données traitées seront ordonnées par une ou plusieurs colonnes contextuelles que nous choisissons. En construisant ces modèles et en les combinant, au lieu d'avoir un seul et unique modèle, nous aurons un système complet Sensible au Contexte.

3.3 Travaux Connexes

Avant d'arriver à notre solution, nous avons vu ce que d'autres chercheurs ont fait pour la sensibilité au Contexte. Donc, nous allons présenter leurs principaux travaux.

3.3.1 De vito et Coll

De vito et Coll ont proposé une approche pour basé sur la régression de réseau neuronal pour effectuer la prédiction de la qualité de l'air en tenant compte la Contextualisation des informations. Ils ont déterminé les Contextes que cette application qui pourra prendre,

leurs objectifs était de calibrer le capteur pour mesurer le gaz benzène, et pour le prédire (gaz de benzène), les chercheurs ont proposé un schéma de régression basé sur une architecture de réseau neuronal [64] .

Afin d'arriver à leurs but, ils ont implémenté un réseau de neurone avec 5 neurones, ils ont observé que la MAE(Erreur Absolue Moyenne) était plus élevée pendant l'hiver spécialement en novembre , par conséquent ils ont expliqué que la concentration et la dynamique de gaz de benzène d'octobre à mars sont différents d'avril à septembre à cause de la basse de température . Pour cela ils ont divisé l'ensemble de données en deux groupes : (Avril septembre, octobre mars) (Avec 70% à l'ensemble de formation ,10% ensemble de validation ,20% ensemble de test). .Puis ils ont formé 3 groupes en considérant uniquement la collecte des données :(Avril septembre, Octobre mars, Compte tenu toutes les données) après avoir fait ça. Les chercheurs ont testé quel est le modèle qui fournit une MAE la plus faible avec le modèle qui représente le meilleur résultat pour chaque Contexte[64].

Au final, pour fournir une comparaison avec le modèle spécifique, ils ont sélectionné le modèle qui a présenté le meilleur résultat pour l'ensemble de validation, ils ont déduit que le modèle d'octobre à mars présente le meilleur résultat que le modèle d'hiver, Après 2000 itération le modèle générique commence a présenté un meilleur résultat que le modèle spécifique, il fournira un meilleur résultat pendant l'été et l'hiver. Les chercheurs ont conclu que la meilleure configuration système est celle qui utilise le modèle générique pendant toute l'année [64].

3.3.2 Nascimento et Lucena

Nascimento et Lucena ont décrit un ensemble de lampadaire pour prise de décision en fonction des données collectées dans l'environnement en décrivant un ensemble de lampadaire avec des réseaux neurones. Leurs objectif étaient de comparer l'ensemble des lampadaires pouvant utiliser plus d'un modèle au moment de l'exécution contre l'ensemble de lampadaires qui n'utilise qu'un seul modèle générique [64] .

Pour cela ils ont simulé 4 Catégories en se basant uniquement sur la configuration de la lumière (la prise de décision sera faite par un algorithme générique pour former un réseau neuronal), ils ont trouvé [64] :

- i. Toujours brillant
- ii. Toujours sombre
- iii. Toujours en fin d'après midi
- iv. Un qui modifier dynamiquement la lumière d'arrière-plan au moment de l'exécution.

Tout d'abord ils ont formé un réseau de neurones qui se compose de cinq unités et qui utilise une fonction d'activation binaire (ils ont mis la lumière de la rue toujours lumineuse pour interagir avec le premier scénario), ensuite ils ont formé un RN sigmoïdes trois fois en utilisant les variantes (ii ,iii ,iv) séparément qui va leur permettre de générer 3 modèles formés avec valeurs de poids particulières plus le modèle qui a été généré pendant que le jeu de lampadaire interagissait ,les modèles ont obtenu des résultats différents pour chaque partie de la journée ils ont trouvé que le modèle générique fourni les meilleurs résultats pendant la nuit et le matin Après avoir augmenté le nombre d'interactions de 20 à 100 pour vérifier si les résultats pourraient être amélioré malheureusement ils ne le pourraient pas sauf s'ils vont former un modèle unique pour traiter toutes les variantes d'éclairages la performance moyenne d'ensemble de lampadaire sera -59,39 .

Pour eux la meilleure solution est de créer un système capable d'utiliser le modèle générique la nuit et le matin et de passer au modèle spécifique en Fin d'après-midi [64].

3.3.3 Hernandez et Coll

Hernandez et Coll ont discuté de l'importance de prendre le Contexte en compte lors de la formation d'un modèle d'apprentissage automatique et le besoin d'approches pour anticiper l'analyse des changements de Contexte. En conséquence, ils encouragent l'utilisation de l'apprentissage progressif par recadrage, une démarche qui consiste à créer un premier modèle dans un Contexte spécifique, puis en ajustant ce modèle aux nouveaux Contextes. Les auteurs déclarent que leur approche peut être généralisée à de nombreux problèmes et domaines de l'apprentissage automatique [65].

3.3.4 Mmezouar et Coll

Mmezouar et Coll ont formé sur l'ensemble de données (local, global) des modèles de prédiction des défauts tenant compte de l'effort utilisant 15 projets. Ils ont observé que, selon la façon dont le modèle local est construit (c'est-à-dire la façon dont le sous-ensemble des données d'apprentissage d'origine est sélectionné), le modèle local ne fonc-

tionne généralement pas aussi bien que le modèle global pour la tâche de prédiction des défauts. Leurs résultats montrent que la séparation des données d'entraînement en fonction des informations Contextuelles est prometteur [66].

Dans deux expériences différentes, leur approche Contextuelle atténue les biais envers le modèle global. Cependant, ils doivent effectuer plus d'expériences, afin d'étudier l'utilisation de cette approche dans plus de domaines d'application, l'utilisation d'autres algorithmes et si l'analyse des changements de Contexte peut être anticipée [66].

3.4 La Différence entre Les Travaux Connexes et notre Solution

Après avoir parlé sur les travaux connexes qui ont été utilisés pour le système Sensible au Contexte, maintenant nous allons expliquer et détailler la différence entre les travaux précédents et notre solution proposée :

- L'une des premières différences, c'est la partie sémantique . Nous avons modélisé le Contexte en utilisant la méthode de modélisation par ontologie , dans le but d'apporter à notre système Machine Learning plus de sémantique .
- Notre solution est basée sur la classification/prédiction en utilisant l'algorithme RandomForest tandis que la proposition de De Vito et Coll[64] et Nascimento et Lucena [65] était basée sur la régression de réseau neurone en utilisant une Architecture de réseau neuronal et le neurone sigmoïde.
- Dans notre proposition, nous avons utilisé la précision de la classe (Accuracy) comme métriques d'évaluations des résultats, par contre les chercheurs ont travaillé avec La MAE (Erreur Absolue Moyenne).

3.5 Approche et Solution Proposée

Après avoir abordé notre problématique dans la section précédente et présenté les travaux connexes qui ont été utilisés pour le système Sensible au Contexte. Maintenant, nous allons présenter notre proposition .

Comme nous l'avons déjà cité notre objectif principal est d'améliorer la performance et la précision d'un modèle du machine Learning en utilisant le Contexte. D'où le système

Machine Learning devient Sensible au Contexte.

Pour cette raison, nous proposons un système Sensible au Contexte qui ne prend pas en compte un seul modèle du Machine Learning , mais il prendra plusieurs modèles.

Chaque modèle sera utilisé pour un Contexte bien défini dont le Contexte peut être une ou plusieurs colonnes du DataSets .

Notre proposition a été testée et validée sur plusieurs DataSets tirés et choisis au hasard, elle est basée sur la prise de la sensibilité au Contexte cela implique que le modèle que nous avons choisi sera donc divisé en sous-ensemble, et un modèle sera généré pour chaque valeur du Contexte (nous pourrions trouver deux colonnes contextuelles dans un seul modèle et c'était notre cas) dans le but de trouver des modèles qui améliorent l'Accuracy globale que nous avons trouvé avant avec le dataset entier.

Pour mieux comprendre comment notre système fonctionne, nous avons fait une Architecture générale qui résume notre proposition (Solution) .« voir la figure 3.1 ».

Architecture Générale

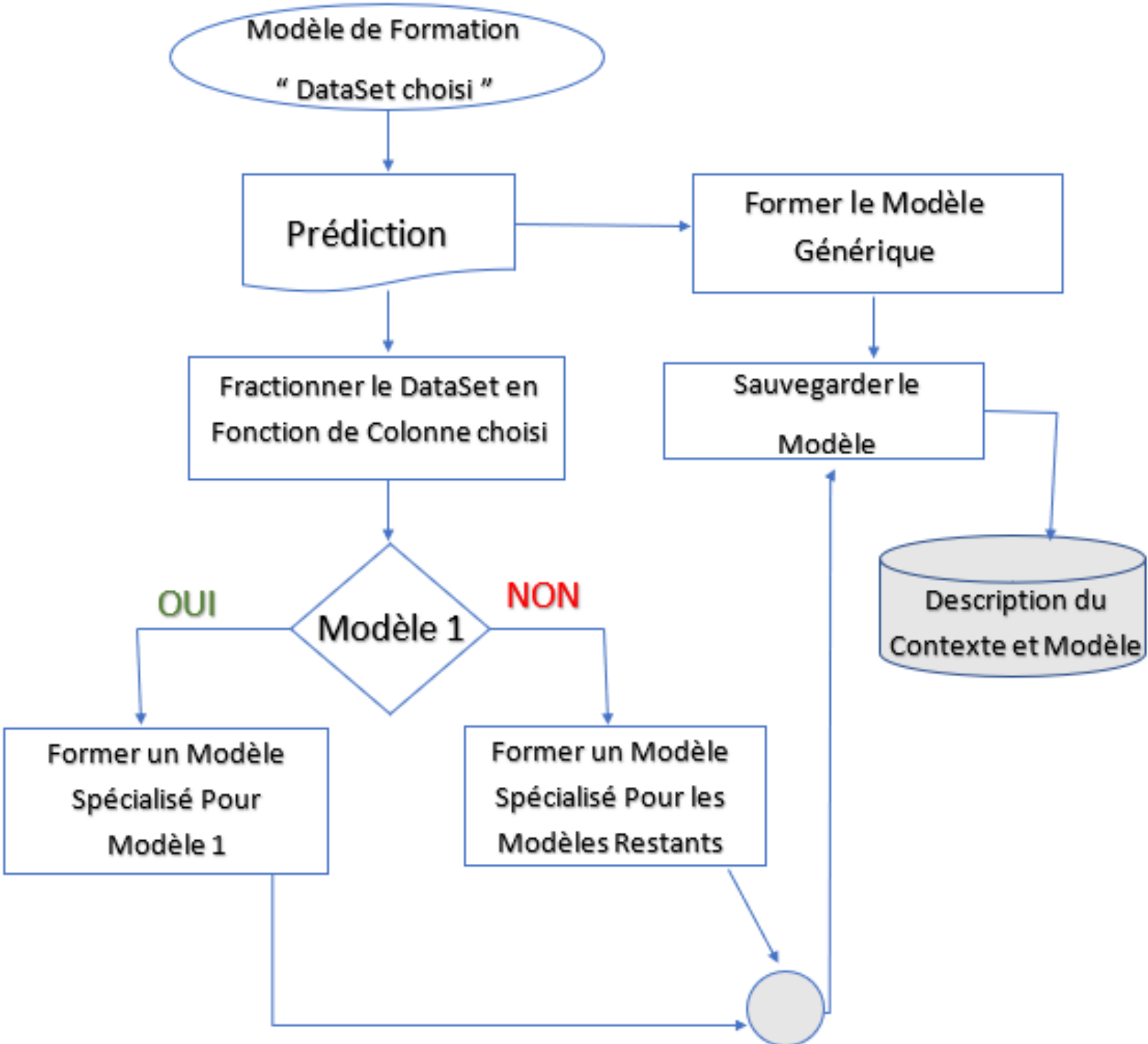


FIGURE 3.1 – Architecture Générale.

Notre proposition (Solution) s'est concrétisée en suivant les étapes ci-dessous :

3.5.1 La première étape

Nous avons créé un modèle d'apprentissage automatique sur 3 ensembles de données distinct (Shipment Tracking , CardioVascular Disease , Rain in Australia) en utilisant les différents algorithmes du Machine Learning tels que RandomForest(RF), (Support Vecteur Machine(SVM), Desicion Tree(DT), k-nearest neighbors(KNN).

Ensuite, pour chaque jeu de données, nous avons pris l'algorithme qui atteint la précision maximale. De plus, pour chaque algorithme, nous avons essayé d'augmenter la précision (Accuracy) en appliquant la méthode de validation croisée pour avoir les meilleurs paramètres .

Le modèle que nous obtiendrons dans cette étape s'appelle le modèle général.

3.5.2 La 2ème étape

Cette deuxième étape est la première véritable étape de notre démarche. Elle consiste à sélectionner une ou plusieurs colonnes Contextuelles .

3.5.3 La 3ème étape

Après avoir choisi la colonne de Contexte, nous sommes maintenant prêts à diviser l'ensemble de données en catégories logiques selon la colonne de Contexte, et cette division dépend davantage de l'ensemble de Données (ex : mois de l'année, divisé en 4 catégories qui sont les saisons), après la division, nous aurons de nombreux sous-ensembles pour chaque ensemble de données cela implique notre ensemble sera divisé et un modèle sera généré pour chaque valeur du Contexte.

Le but de cette étape est de trouver des modèles qui améliorent l'Accuracy globale que nous avons trouvée avec le modèle entier (Général) .

3.5.4 La 4ème étape

Nous allons créer un modèle Machine Learning pour chaque sous-ensemble avec l'algorithme qui donne la Précision (Accuracy) maximale. Et augmentez-le avec une validation croisée comme nous l'avons fait pour le modèle général.

3.5.5 La 5ème étape

Après avoir créé les modèles pour chaque Contexte, nous voyons quels modèles surpassent le modèle général .

3.5.6 La 6ème étape

Cette étape consiste à modéliser le Contexte par les méthodes de modélisations que nous avons citées dans le chapitre 1 . Dans notre cas nous avons utilisé la méthode (modélisation par Ontologie), nous déduisons que le Contexte sera modélisé sous format graphe RDF (dans le but d'apporter plus de sémantique) ou les comparaisons des valeurs du Contexte se feront à base de similarité conceptuelle , quand celles-ci ont comme valeur une URI(Uniform Ressource Identifier).

Nous avons utilisé une ontologie de domaine qui est « **OntoDM** ».

OntoDM : C'est une ontologie générale pour l'exploration de données dans le but de fournir un cadre unifié pour la recherche en data mining.

Elle tente de couvrir toute la largeur de la recherche sur l'exploration de données, contenant des classes de haut niveau, telles que les tâches et les algorithmes d'exploration de données, et des classes plus spécifiques liées à certains sous-domaines, telles que les contraintes pour l'exploration de données basée sur les contraintes [74].

La sémantique a pour but de nous permettre d'être plus flexibles lorsque nous faisons cette comparaison , elle ne permet pas de garder juste la partie syntaxique mais de partir dans la sémantique pour englober plusieurs choses à la fois dans le but d'avoir un résultat de haut charge qui est est varié .

3.6 Analyse et Exploration des Données (Dataset)

Dans la section précédente (problématique), nous avons mentionné que notre solution sera testée et validée sur différents DataSets tirés et choisis au hasard.

- nous avons choisi 3 DataSets en consultant le site Kaggle (Shipment Tracking , CardioVascular Disease , Rain in Australia).

3.6.1 E-Commerce Shipping

Ce Dataset est proposé par une société de commerce électronique internationale, ils souhaitent utiliser certaines des techniques d'apprentissage automatique les plus avancées pour étudier leurs clients. Il contient 10 999 observations de 12 variables [67] .

Les données contiennent les informations suivantes :

- **ID** : Numéro d'identification des clients.
- **Warehouse block** : la société a un grand entrepôt qui est divisé en bloc tel que A, B, C, D, E.
- **Mode of shipment** : la société expédie les produits de plusieurs manières telles que bateau, vol et route.
- **Customer care calls** : le nombre d'appels effectués à partir d'une demande de renseignements sur l'envoi.
- **Customer rating** : la société a évalué de chaque client. 1 est le plus bas (pire), 5 sont le plus élevé (meilleur).
- **Cost of the product** : Coût du produit en dollars américains.
- **Prior purchases** : le nombre d'achats précédents.
- **Product importance** : La société a classé le produit dans les différents paramètres tels que faible, moyen, élevé.
- **Gender** : homme et femme.
- **Discount offered** : Remise offerte sur ce produit spécifique.
- **Weight in gms** : C'est le poids en grammes.
- **Reached on time** : il s'agit de la variable cible, où 1 indique que le produit n'est PAS arrivé à l'heure et 0 indique qu'il a atteint l'heure.

Ces données de suivi des expéditions de produits, répondent instantanément à la question

suivante : qu'est-ce que l'évaluation des clients ? Et le produit a-t-il été livré à temps.

3.6.2 CardioVasculaire

Cet ensemble de données est fait pour la prédiction de la maladie Cardiovasculaire en tenant compte (L'âge , le poids, le taux du glucose ..etc) [68] . Il comporte 68 564 entrées avec 12 colonnes + la classe de prédiction. « Voir la figure 3.2 ».

Nous avons testé 4 algorithmes « **SVM** , **Décision Tree**, **RandomForest** et **KNN** », et nous avons trouvé que l'algorithme RandomForest donne de très bons résultats .

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 68564 entries, 0 to 68563
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   AGE                   68564 non-null  int64
1   GENDER                68564 non-null  int64
2   HEIGHT                68564 non-null  int64
3   WEIGHT                68564 non-null  int64
4   AP_HIGH               68564 non-null  int64
5   AP_LOW                68564 non-null  int64
6   CHOLESTEROL           68564 non-null  int64
7   GLUCOSE                68564 non-null  int64
8   SMOKE                 68564 non-null  int64
9   ALCOHOL               68564 non-null  int64
10  PHYSICAL_ACTIVITY     68564 non-null  int64
11  CARDIO_DISEASE        68564 non-null  int64
dtypes: int64(12)
memory usage: 6.3 MB
```

FIGURE 3.2 – Information sur le Dataset CardioVascular Disease

3.6.3 Rain In Australia

Cet ensemble de données contient environ 10 ans d'observations météorologiques quotidiennes à partir de nombreux endroits à travers l'Australie [69] .

RainTomorrow est la variable cible à prédire. Cela signifie - a-t-il plu le lendemain, oui ou non ? Cette colonne serait Oui si la pluie pour ce jour-là était de 1 mm ou plus. Le data contient 23 colonnes et environ 145 460 entrées « voir la figure 3.3 »

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 145460 entries, 0 to 145459
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Date                  145460 non-null object
1   Location              145460 non-null object
2   MinTemp               143975 non-null float64
3   MaxTemp               144199 non-null float64
4   Rainfall              142199 non-null float64
5   Evaporation           82670 non-null float64
6   Sunshine              75625 non-null float64
7   WindGustDir           135134 non-null object
8   WindGustSpeed         135197 non-null float64
9   WindDir9am            134894 non-null object
10  WindDir3pm            141232 non-null object
11  WindSpeed9am          143693 non-null float64
12  WindSpeed3pm          142398 non-null float64
13  Humidity9am           142806 non-null float64
14  Humidity3pm           140953 non-null float64
15  Pressure9am           130395 non-null float64
16  Pressure3pm           130432 non-null float64
17  Cloud9am              89572 non-null float64
18  Cloud3pm              86102 non-null float64
19  Temp9am               143693 non-null float64
20  Temp3pm               141851 non-null float64
21  RainToday             142199 non-null object
22  RainTomorrow          142193 non-null object
dtypes: float64(16), object(7)
memory usage: 25.5+ MB

```

FIGURE 3.3 – Ensemble de Donnée (Rain in Australia)

Dans ce DataSet nous avons Catégorisé la colonne de date en 4 saisons (hiver, printemps, été, automne) (nous aurions besoin de cette étape dans notre approche proposée).

Après avoir bien défini chaque DataSets que nous avons choisi ,nous allons former leur modèle général.(la figure 3.4 montre l'architecture de comment former un modèle du Machine Learning) .

Architecture

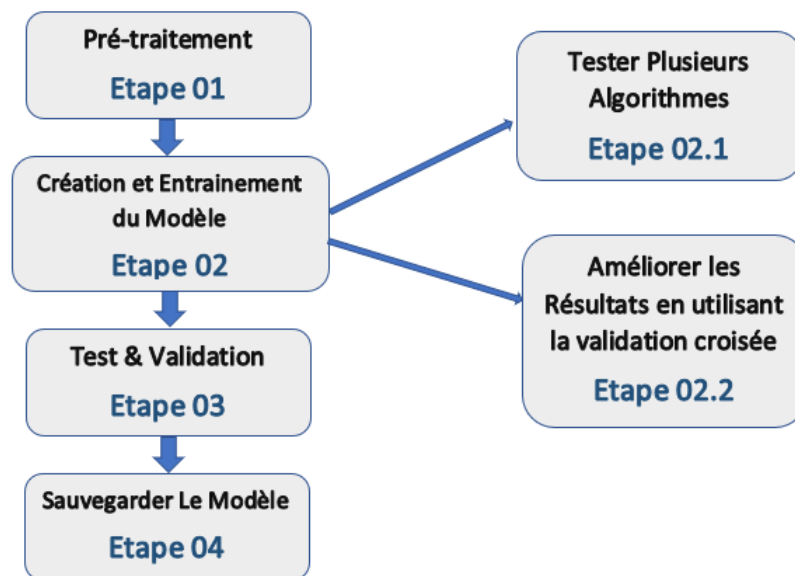


FIGURE 3.4 – Former un Modèle du Machine Learning

Chaque DataSet sera formé de la même manière en suivant la procédure au-dessous :

- Afin que nos DataSets seront bien formés, nous allons réaliser une lecture approfondie de nos données afin de comprendre leur rôle et les impacts qu'elles peuvent avoir dans l'objectif de prédiction que nous nous sommes fixés. L'étude des données passe notamment par leur description (nom, type...), ainsi que par divers processus de traitement tels que le nettoyage (suppression des données inutiles, recherche des données manquantes) et enfin la combinaison entre elles, aussi appelée agrégation, dans le but de disposer d'un jeu de connaissances (observations) utilisables et appropriées à l'apprentissage et à l'atteinte de notre objectif.

- Après avoir préparé nos données pour chaque datasets, nous allons les entraîner en utilisant les algorithmes du Machine Learning. Nous avons simulé plusieurs algorithmes (**Arbre de Décision, RandomForest, SVM, KNN**) en jouant sur plusieurs paramètres jusqu'à ce que nos modèles atteignent le niveau de précision souhaité.

Après avoir bien formé les dataSets, nous passons par implémenter notre proposition. Elle consiste les étapes suivantes :

3.7 Identification du Contexte pour chaque DataSet choisi

Dans chaque DataSet, nous allons identifier une ou plusieurs colonnes contextuelles. Après avoir les identifiées nos modèles de Machine Learning seront donc re entraînés en se basant seulement sur chaque valeur du Contexte/colonne choisi. Donc un modèle sera généré pour chaque valeur du Contexte. Nous commençons par le premier DataSet .

3.7.1 Identification du Contexte pour "Shipment Tracking"

D'après notre analyse, nous remarquons que le DataSet que nous avons choisi ne contient pas de colonne contextuelle, mais nous voulons implémenter notre solution et voir si elle fonctionne avec tous les dataSets. Pour cela, nous avons identifié trois colonnes (**Mode Of Shipment** , **Prix** , **Weight**) .

L'identification des colonnes sera comme suit :

3.7.1.1 Contexte Mode Of Shipment

Nous avons divisé cette colonne en trois catégories :

- Catégorie 1 :Ship
- Catégorie 2 :Flight
- Catégorie 3 : Road

3.7.1.2 Contexte Prix

En fonction du coût du produit, nous diviserons nos 4 sous-ensembles comme suit :

1. sous-ensemble 1 : [96,150]
2. sous-ensemble 2 : [151,200]

3. sous-ensemble 3 : [201,250]
4. sous-ensemble 4 : [251,310]

3.7.1.3 Contexte Weight

Nous divisons la colonne Weight (poids) en sous-ensemble (catégorie) comme suit :

1. Sous-ensemble : $\text{Weight} \leq 2000$
2. Sous-ensemble : $\text{Weight} > 2000 \ \& \ \text{Weight} \leq 4000$
3. Sous-ensemble : $\text{Weight} > 4000 \ \& \ \text{Weight} \leq 5000$
4. Sous-ensemble : $\text{Weight} > 5000$

De la même manière, nous continuerons l'identification du Contexte pour le reste des DataSets .

3.7.2 Identification du Contexte pour "CardioVascular Diseases"

Nous avons identifié la colonne âge et nous la diviserons en six catégories :

1. Sous-ensemble 1 : [30,40]
2. Sous-ensemble 2 : [41,45]
3. Sous-ensemble 3 : [46,50]
4. Sous-ensemble 4 : [51,55]
5. Sous-ensemble 5 : [56,60]
6. Sous-ensemble 6 : [60,65]

3.7.3 Identification du Contexte pour ” Rain In Australia ”

Dans ce DataSet , nous remarquons qu’il contient 4 colonnes contextuelles, nous avons testé avec deux (**Humidité, saison**) .

3.7.3.1 Contexte ’Saison’

1. Sous-ensemble 1 : été
2. Sous-ensemble 2 : automne
3. Sous-ensemble 3 : hiver
4. Sous-ensemble 4 : printemps

3.7.3.2 Contexte ’ Humidité ’

1. Sous-ensemble 1 : [0,20]
2. Sous-ensemble 2 : [21,40]
3. Sous-ensemble 3 : [41,60]
4. Sous-ensemble 4 : [61,80]
5. Sous-ensemble 5 : [81,100]

Après avoir bien défini les colonnes contextuelles pour chaque Datasets , nous passons à l’étape qui suit

3.8 Modélisation du Contexte (Représentation)

Cette étape consiste tous les Contextes qui ont été identifiés à partir de nos trois Datasets choisis.

Après avoir identifié le Contexte dans ce Modelé et l'utilisé dans le but d'améliorer les performances, maintenant, nous allons le modéliser pour augmenter l'efficacité du développement et réduire la charge de travail des développeurs. Nous avons modélisé les Contextes des DataSets en utilisant la méthode de modélisation par ontologie dans le but d'apporter plus de sémantiques à notre système. Donc nous aurions obtenu un profil Contextuel sous format Graphe RDF.

3.9 Conclusion

Dans ce chapitre, nous avons présenté notre problématique en incluant les travaux connexes pour la prise de conscience du Contexte. Ensuite, nous avons défini les ensembles de données que nous avons utilisés.

Au final, nous avons proposé notre approche pour optimiser et améliorer les résultats d'un modèle du Machine Learning.

Dans le chapitre suivant, nous discuterons l'implémentation de notre solution ainsi les résultats obtenus.

Chapitre 4

Résultats et évaluation.

4.1 Introduction

Après avoir détaillé notre approche dans le chapitre précédent, dans ce chapitre, nous allons parler d'une manière plus claire sur l'environnement de développement, le langage ainsi les bibliothèques utilisées qui nous ont aidé à implémenter notre solution en l'accompagnant avec une architecture globale pour l'implémentation.

En outre, d'une autre part, nous allons discuter en détails sur ce que nous avons obtenu comme résultats.

Nous terminerons par une comparaison pour chaque résultat obtenu.

4.2 Environnement de développement

Dans cette partie, nous allons parler sur l'environnement ou nous avons développé notre approche ainsi le langage et les bibliothèques utilisés.

4.2.1 Plateforme (Environnement) De développement

Google Colaboratory :¹il est offert par Google qui veut dire ,il est gratuit ,il est basé sur une application Web Open source (Jupyter Notebook) dans le but de permettre de créer et de partager des documents qui contiennent des codes exécutables directement dans le document. Cette application permet de faire le traitement de données, les modéliser

1. <https://colab.research.google.com>

et les visualiser .>

Google Colab a une caractéristique très importante qui est de traiter les données d'un modèle du Machine Learning directement du cloud cela implique que le programmeur n'a pas besoin d'installer quoi que ce soit sur son Ordinateur. [70]

4.2.2 Langage de Développement

Python est le langage de programmation le plus utilisé dans le domaine du Machine Learning, du Big Data et de la Data Science.

4.2.2.1 Python

Il a été créé en 1991 par Guido van Rossum, aux Pays-Bas, il tire son nom de l'émission Monty Python's Flying Circus. Il a été apparu à l'époque comme une façon d'automatiser les éléments les plus ennuyeux de l'écriture de scripts. Il s'agit d'un langage de programmation interprété, qui ne nécessite donc pas d'être compilé pour fonctionner. Un programme " interpréteur " permet d'exécuter le code Python sur n'importe quel ordinateur. Ceci permet de voir rapidement les résultats d'un changement dans le code [71].

Le langage **Python** doit sa popularité à plusieurs avantages qui profitent aussi bien aux débutants qu'aux experts. Tout d'abord :

- Il est facile à apprendre et à utiliser. Ses caractéristiques sont peu nombreuses, ce qui permet de créer des programmes rapidement et avec peu d'efforts [71].
- Un autre avantage du Python est sa popularité. Ce langage fonctionne sur tous les principaux systèmes d'exploitation et plateformes informatiques. De plus, même s'il ne s'agit clairement pas du langage le plus rapide, il compense sa lenteur par sa versatilité [71].
- Enfin, même s'il est principalement utilisé pour le scripting et l'automatisation, ce langage est aussi utilisé pour créer des logiciels de qualité professionnelle [71].

4.2.3 Bibliothèques Utilisées

4.2.3.1 Sklearn

C'est un module Python pour l'apprentissage automatique construit sur SciPy et est distribué sous la licence BSD à 3 clauses [72].

4.2.3.2 Rdfib

C'est un package Python pur pour travailler avec un modèle de graphe RDF (Resource Description Framework). Cette bibliothèque contient des analyseurs/sérialiseurs pour presque toutes les sérialisations RDF connues, telles que RDF/XML, Turtle, N-Triples et JSON-LD. [73]

4.3 Discussion et Résultats Obtenus Pour les Trois Expériences (DataSets)

Après avoir bien détaillé notre proposition (Solution) , nous passons par donner les résultats que nous avons testés avec les 3 DataSets discutés dans la section précédente. Nous commençons d'abord par le premier DataSet

4.3.1 La Première Expérience (Rain In Australia)

Comme nous l'avons déjà mentionné, nous avons modélisé cet ensemble de données avec deux Algorithmes du Machine Learning, (**RandomForest, Arbre de Décision**). Nous avons essayé d'optimiser les résultats qui nous a donné ce dernier en utilisant la validation croisée dans le but d'extraire les meilleurs paramètres.

Le tableau 4.1, au-dessous, montre les résultats du premier Modèle général.

Modèle Général	
Accuracy en Utilisant la Validation Croisée	
RandomForest	Arbre de Décision
0.8449 «84,49 %»	0.83 « 83 % »

TABLE 4.1 – Amélioration des Résultats Pour le Modèle Général (Rain in Australia)

4.3.1.1 Identification et Modélisation (Représentation) Du Contexte Pour le Modèle Rain in Australia

4.3.1.1.1 Identification du Contexte :

Nous avons déjà expliqué en détails cette étape dans le chapitre précédent, donc nous allons directement mettre les résultats pour chaque contexte choisi.

1. Le Premier Contexte ” Saison ”

Pour voir quel résultat nous a mené ses 4 modèles, nous avons les représentés dans le tableau (4.2) suivant :

Modèle Contextuel (Saison)			
Modèle 1 (Hiver)	Modèle 2 (Printemps)	Modèle 3 (Eté)	Modèle 4 (Automne)
Accuracy (RandomForest) en Utilisant la Validation Croisée			
0.8597 « 85.97 % »	0.8568 « 85.68 %»	0.8414 « 84.14 %»	0.8487« 84.87 % »

TABLE 4.2 – Amélioration des Résultats du Modèle Conrextuel (Saison)

D'après les dernières expériences que nous avons vues, la saison n'affecte pas tellement la précision, même si nous l'avons prise comme une caractéristique de Contexte, Donc nous avons fait un graphe pour montrer l'importance des caractéristiques de cet ensemble de données. (voir la figure 4.1)

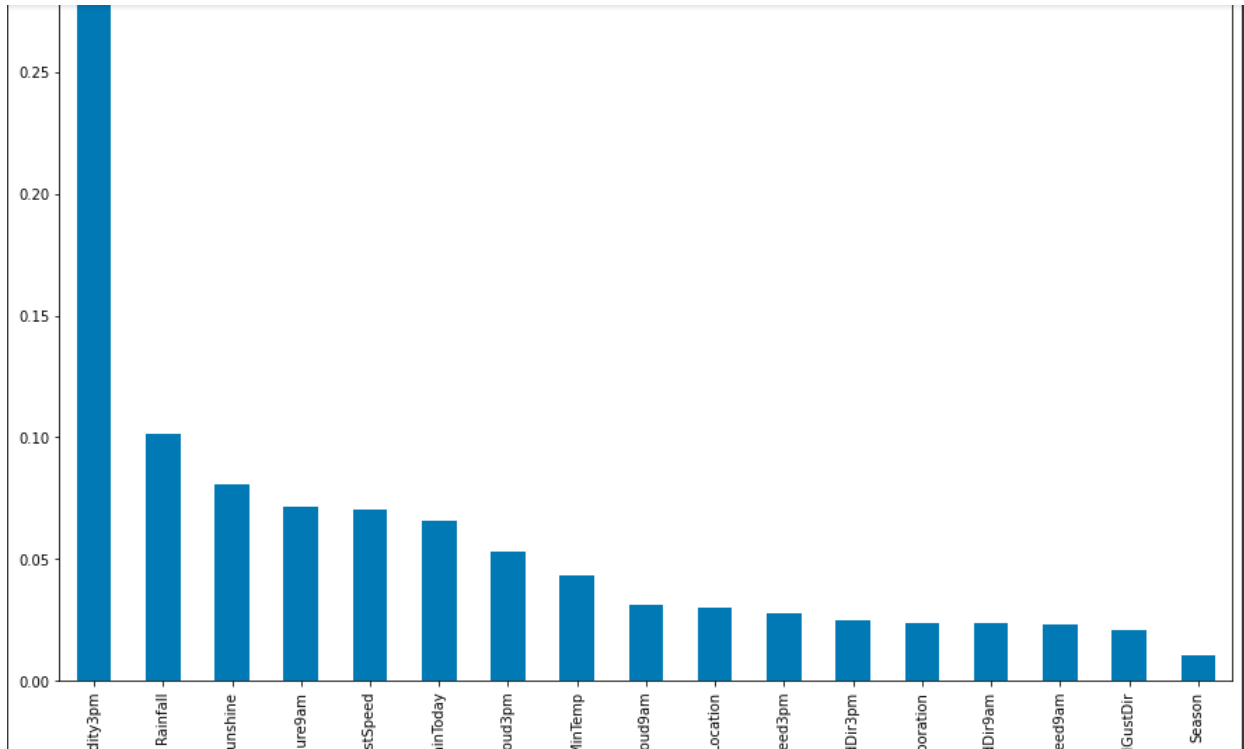


FIGURE 4.1 – Caractéristiques du Modèle Rain in Australia

À partir de ce graphe, nous remarquons que la colonne humidité influence beaucoup sur les résultats du modèle général, donc c'est pour cette raison, nous l'avons choisie comme une deuxième colonne contextuelle.

2. Le Deuxième Contexte ' Humidité '

Le tableau 4.3 suivant montre les résultats obtenus de cette colonne

Modèle Contextuel (Humidité)				
Modèle 1 : Humidité [0 ,20]	Modèle 2 : Humidité [21,40]	Modèle 3 : Humidité [41,60]	Modèle 4 : Humidité [61,80]	Modèle 5 : Humidité [81,100]
Accuracy en Utilisant la Validation Croisée (RandomForest)				
0.9707 « 97.07 % »	0.9329 « 93.29 %»	0.8711 « 87.11 %»	0.7406 « 74.06 % »	0.7988 « 79.88 %»

TABLE 4.3 – Optimisation des Résultats du Deuxième Contexte (Humidité)

- À partir des expérimentations ci-dessus, nous voyons que la précision des modèles de perception du Contexte dépasse le modèle général lorsque l'humidité est comprise entre 0 et 60 ([0,60]) mais la précision générale du modèle l'emporte lorsque l'humidité est comprise entre 61 et 100 ([61,100])).
- Nous allons essayer de tester nos modèles de prise de conscience des Contextes avec le modèle général sur le même ensemble de tests que nous avons testé dans le modèle général et voir si la précision augmentera.
- La séparation sera la suivante :
 1. Humidité [0,20] : modèle de prise de conscience du Contexte 1.
 2. Humidité [21,40] : modèle de prise de conscience du Contexte 2.
 3. Humidité [41,60] : modèle de prise de conscience du Contexte 3.
 4. Humidité [61,100] : modèle général.
- Nous obtenons les résultats suivants :
 - * La Précision du premier Modèle = 0.9931818181818182
 - * La Précision du Deuxième Modèle = 0.9664586583463338
 - * La Précision du Troisième Modèle = 0.9455626715462031
 - * La Précision du Modèle Général = 0.7578451882845189
- À partir des dernières expériences, nous avons constaté des améliorations de la

précision lorsque nous avons utilisé une approche de sensibilisation au Contexte de 84,49 % à 91,57 %. Et cela prouve l'efficacité de l'utilisation de la conscience du Contexte dans l'apprentissage automatique et l'intelligence artificielle en général.

4.3.1.1.2 Modélisation du Contexte (Représentation) du "Rain In Australia"

:

Après avoir identifié le Contexte dans ce Modèle et l'utilisé dans le but d'améliorer les performances, maintenant, nous allons le modéliser pour augmenter l'efficacité du développement et réduire la charge de travail des développeurs.

Nous avons modélisé les Contextes de ce modèle en utilisant la méthode de modélisation par ontologie dans le but d'apporter plus de sémantiques à notre système. Donc nous aurions obtenu un profil Contextuel sous format Graphe RDF.

Après avoir extrait le profil contextuelle, il nous apparut long et illisible. Pour cette raison, nous l'avons divisé en sous-graphe pour qu'il soit bien compréhensible. « voir les figures 4.2; 4.3; 4.4 »

ET nous avons sauvegardé la figure du Graphe dans notre Drive et voici son URI :

<https://drive.google.com/file/d/1zHGOiMNA2sinAMGHQoVyX4Hrwy8Aw8U5/view?usp=sharing>

Nous avons fait la syntaxe RDF/XML pour la modélisation du Contexte, elle sera comme suite.

Modélisation des Contextes

” Humidité , Saison ” Syntaxe Rdf/XML

```

<rdf :RDF
xmlns :rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns :dbo="http://dbpedia.org/ontology/"
xmlns :OntoDM="http://kt.ijs.si/panovp/OntoDM/"
xmlns :ex="http://kt.ijs.si/panovp/OntoDM/"
xmlns :rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdf :Description rdf :about="http://www.dbpedia.org/resource/DataSET">
    <ex :name xml :lang="an">Rain In Australia</ex :name>
    <rdf :type rdf :resource="http://fr.dbpedia.org/resource/DataSet pour la classification/prédiction"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Date"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Localisation"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/MinTemp"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/MaXtemps"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/RainFall"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Evaporation"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/SunSHine"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WindGustDir"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinGustSpeed"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinDir9am"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinDir3pm"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinSpeed9am"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinSpeed3pm"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Humidity9am"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Humidity3pm"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Pressure9am"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Pressure3pm"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Cloud9am"/>

```

```

<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Cloud3pm"/>
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Temp9am"/>
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Temp3pm"/>
<ex :iscontext rdf :resource="http://www.dbpedia.org/resource/Saison"/>
<ex :iscontext rdf :resource="http://www.dbpedia.org/resource/Humidité"/>
</rdf :Description>

```

```

<rdf :Description rdf :about="http://www.dbpedia.org/resource/Humidité" >
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle1"/>
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle2"/>
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle3"/>
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle4"/>
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle5"/>
</rdf :Description>

```

```

<rdf :Description rdf :about ="http://https://dbpedia.org/ontology/Saison" >
  <ex :Name rdf :resource="http://kt.ijs.si/panovp/OntoDM/Eté" ></ex :Name>
  <ex :Name rdf :resource="http://kt.ijs.si/panovp/OntoDM/Automne" ></ex :Name>
  <ex :Name rdf :resource="http://kt.ijs.si/panovp/OntoDM/Hiver" ></ex :Name>
  <ex :Name rdf :resource="http://kt.ijs.si/panovp/OntoDM/Printemps" ></ex :Name>
</rdf :Description>

```

```

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle1" >
  <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/Model1"/>
  <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
    0 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
    20 ^^ xsd :intger </rdfs :end>
</rdf :Description>

```

```

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle2" >
  <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/Model2"/>

```

```

<rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    21 ^^ xsd :integer </rdfs :start>
<rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    40 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle3" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/MOdelé3" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        41 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        60 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle4" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/GeneralModel" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        61 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        80 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle5" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/GeneralModel" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        81 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        100 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/#Eté" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model1" />

```

```

<rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    21.06 ^^ xsd :integer </rdfs :start>
<rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    21.09 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/Automne" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model2" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.09 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.12 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/Hiver" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/ModèleGénéral" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.12 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.03 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/Printemps" >
    <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model3" />
    <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.03 ^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
        21.06 ^^ xsd :integer </rdfs :end>
</rdf :Description>
</rdf :RDF>

```

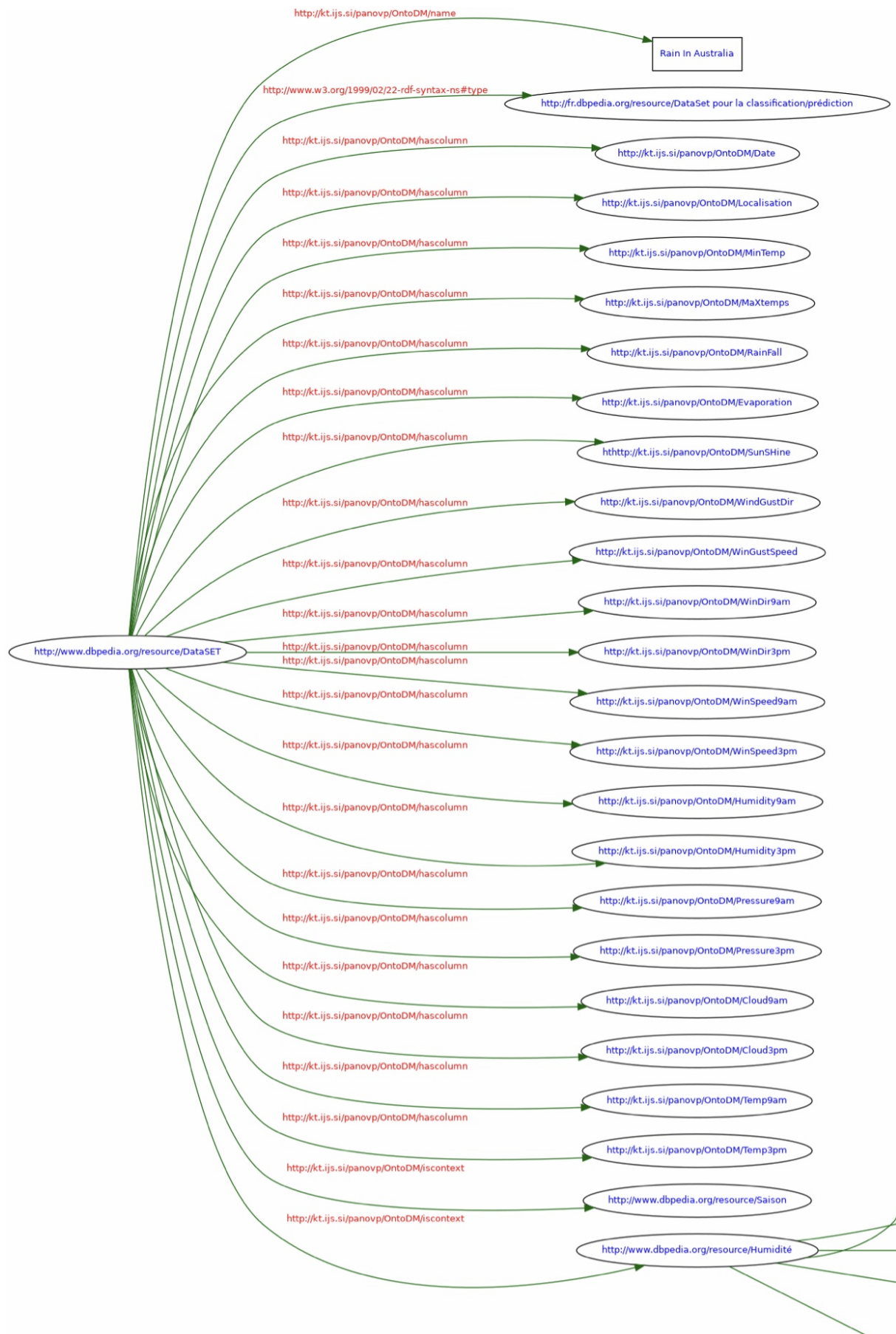


FIGURE 4.2 – Modélisation des Contextes ” Humidité , Saison ”-1 Graphe RDF



FIGURE 4.3 – Modélisation des Contextes ” Humidité , Saison ”-2 Graphe RDF

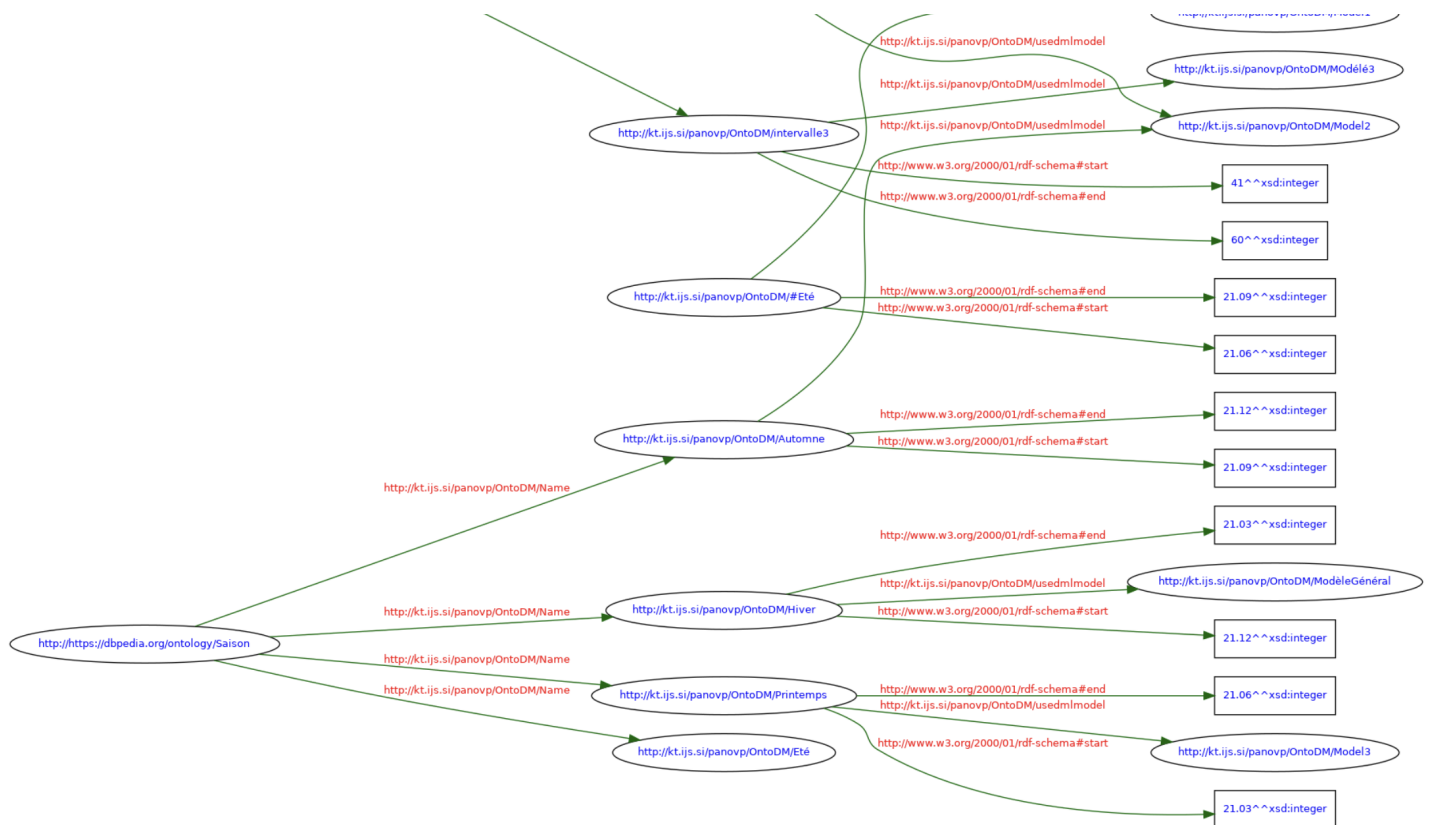


FIGURE 4.4 – Modélisation des Contextes ” Humidité , Saison ”-3 Graphe RDF

4.3.1.1.3 Classification Sensible au Contexte "Rain In Australia" :

Maintenant ce qu'il nous reste à faire, c'est d'implémenter notre système complet qui automatisera en temps réel et nous aurions obtenu un système sensible au contexte :

- Après les expériences précédentes, nous avons vu que la précision des modèles de perception sur texte dépasse le modèle général lorsque l'humidité est comprise entre 0 et 60 ([0,60]) mais la précision générale du modèle l'emporte lorsque l'humidité est comprise entre 61 et 100 ([61,100]) .
- Nous allons donc créer une combinaison des trois premiers modèles d'approche de la conscience du Contexte et du modèle général et classer par l'un des modèles en fonction de l'humidité.
- Les étapes à suivre sont donc les suivantes :
 1. importer les quatre modèles et le jeu de données + préparer les données à intégrer dans nos modèles.
 2. choisissez au hasard une ligne à partir de celui-ci (au lieu de saisir les données manuellement)
 3. retirez la colonne de pluie demain et extrayez l'humidité de la rangée.
 4. classer le point de données (ligne) dans l'un des modèles en fonction de l'humidité (savoir s'il va pleuvoir demain ou non).
 5. imprimez s'il va pleuvoir demain ou non.

En appliquant les étapes précédentes, nous sommes sorties par un exemple choisi et extrait du DataSet (voir la figure 4.5), pour tester notre système et voir l'importance que notre solution a mené.

Season	Location	MinTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity3p	
82590	3	12	5.5	3.8	1.8	4.1	3	48.0	3	14	22.0	15.0	95

FIGURE 4.5 – Exemple Pour la Prédiction en Temps Réel

Nous avons obtenu le résultat suivant (figure 4.6)

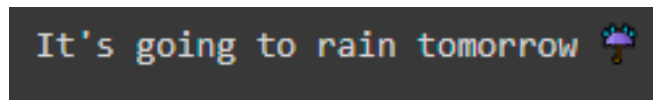


FIGURE 4.6 – Résultat de la Prédiction en Temps Réel Pour le Modèle Rain In Australia

4.3.2 La Deuxième Expérience (CardioVascular Diseases)

4.3.2.1 Modélisation/Classification (Prédiction) Des DataSets

Dans cet ensemble de donnée, nous l'avons modélisé avec quatres Algorithmes du Machine Learning ,(RandomForest,Arbre de Décision ,SVM ,KNN).

Le Tableau 4.4 ci-dessous montre les résultats que nous avons obtenus

Modèle Général			
Accuracy en Utilisant la Validation Croisée			
RandomForest	SVM	KNN	Arbre De Décision
0.7339<<73.39 % >>	0.7349<<73.49 % >>	0.7314<<73.14 % >>	0.7334 < 73.34 % >>

TABLE 4.4 – Accuracy Du Modèle Général Après l'Optimisation

D'après les évaluations précédentes, nous remarquons que l'algorithme RandomForest donne un bon résultat par rapport aux autres algorithmes. Donc nous l'avons utilisé dans l'étape qui suivie.

4.3.2.2 Identification et Modélisation (Représentation) du Contexte Pour le Modèle CardioVascular Diseases

4.3.2.2.1 Identification du Contexte :

Nous remarquons que le modèle contient une colonne contextuelle qui est l'âge, nous allons travailler sur cette dernière.

1.Contexte 'âge' :

Pour voir quel résultat nous a mené ses 6 modèles (Sous-ensembles), nous avons les représentés dans le tableau 4.5 suivant :

Modèle Contextuel (AGE)					
Modèle 1	Modèle 2	Modèle 3	Modèle 4	Modèle 5	Modèle 6
Accuracy (RandomForest) en utilisant la validation croisée					
0.8601	0.8046	0.7544	0.7259	0.6849	0.7105
<< 86.01 % >>	<< 80.46 % >>	<< 75.44 % >>	<< 72.59 % >>	<< 68.49 % >>	<< 71.05 % >>

TABLE 4.5 – Résultats obtenus pour la colonne âge

- Les tests montrent que la précision moyenne des modèles de sensibilité du Contexte est de : **75,821%**, ce qui est généralement mieux que la précision générale du modèle.

- À partir des expérimentations ci-dessus, nous voyons que la précision des modèles prend en compte le Contexte surpasse le modèle général lorsque l'âge est compris entre 30 et 50 ans ([30,50]) mais que la précision générale du modèle l'emporte lorsque l'âge est compris entre 51 et 65 ans ([51, 65]).

- D'où, nous allons essayer de tester nos modèles de prise de conscience des Contextes parallèlement au modèle général sur le même ensemble de test que nous avons testé dans le modèle général et voir si la précision augmentera.

La séparation sera la suivante :

- Age [30 ,40]
- Age [40,45]
- Age [45,50]
- Age > 50 et Age < 30

Nous obtenons les résultats suivants :

Précision du premier modèle = 0,8524590163934426

Précision du deuxième modèle = 0,8106312292358804

Précision du troisième modèle = 0,7656826568265682

Précision du quatrième modèle = 0.6925238898257448

À partir des dernières expériences, nous avons constaté une augmentation notable de la précision lorsque nous avons utilisé une approche de sensibilisation au Contexte de **73%** à **78%**

4.3.2.2 Modélisation du Contexte (Représentation) :

Comme nous l'avons fait avec le premier modèle, c'est la même chose, après avoir identifié le Contexte et l'entraîner. Maintenant, nous allons essayer de créer un profil contextuel en utilisant le RDF.

Après avoir extrait le profil contextuel . Il nous apparut long et illisible. Pour cette raison, nous l'avons divisé en sous-graphe pour qu'il soit bien compréhensible. « voir les figures 4.7; 4.8 »

ET nous avons sauvegardé la figure du Graphe dans notre Drive et voici son URI :

<https://drive.google.com/file/d/10-QhDGm8x4hzDWejFJPGX2yeBx4nBm0V/view?usp=sharing>

Nous avons fait la syntaxe RDF/XML pour la modélisation du Contexte, elle sera comme suite.

Modélisation des Contextes "Age"

Syntaxe Rdf/XML

```

<rdf :RDF
xmlns :rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns :dbo="http://dbpedia.org/ontology/"
xmlns :OntoDM="http://kt.ijs.si/panovp/OntoDM/"
xmlns :ex="http://kt.ijs.si/panovp/OntoDM/"
xmlns :rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdf :Description rdf :about="http://www.dbpedia.org/resource/DataSET">
    <ex :name xml :lang="fr">Maladie CardioVasculaire</ex :name>
    <rdf :type rdf :resource="http://fr.dbpedia.org/resource/DataSet pour la classification/prédiction"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Gender"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Height"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Weight"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/AP_High"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/AP_Low"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Cholesterol"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Glucose"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Smoke"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Alcool"/>
    <ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Physical-Activity"/>
    <ex :iscontext rdf :resource="http://www.dbpedia.org/resource/AGE"/>
  </rdf :Description>

  <rdf :Description rdf :about="http://www.dbpedia.org/resource/AGE">
    <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle1"/>
    <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle2"/>
    <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle3"/>
    <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle4"/>

```

```

    <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/intervalle5" />
</rdf :Description>

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle1" >
    <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/Model2" />
    <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        41^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        45^^ xsd : </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle2" >
    <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/Model3" />
    <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        45^^ xds :integer </rdfs :start>
    <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        50^^ sxd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle3" >
    <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/GeneralModel" />
    <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        51^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        55^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle4" >
    <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        56^^ xsd :integer </rdfs :start>
    <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
        60^^ xsd :integer </rdfs :end>

```

```
<ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/Model1"/>
</rdf :Description>

<rdf :Description rdf :about="http://kt.ijs.si/panovp/OntoDM/intervalle5" >
  <ex :usedmlmodel rdf :resource="http://kt.ijs.si/panovp/OntoDM/GeneralModel"/>
  <rdfs :start rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
    61^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http://www.w3.org/2000/01/rdf-schema#integer" >
    65^^ xsd :integer </rdfs :end>

</rdf :Description>
</rdf :RDF>
```

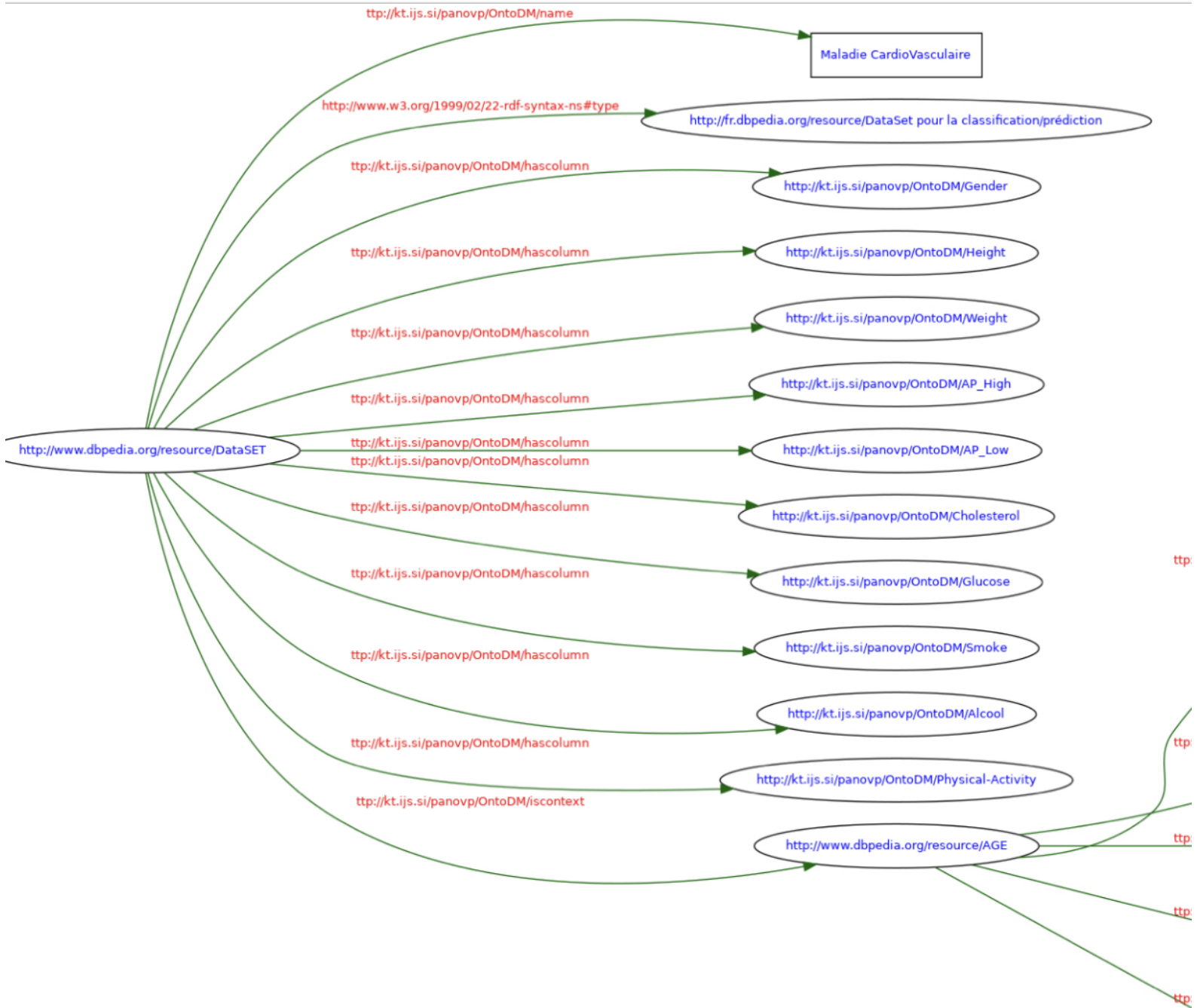


FIGURE 4.7 – Modélisation des Contextes Age-1 Graphe RDF



FIGURE 4.8 – Modélisation des Contextes Age-2 Graphe RDF

4.3.2.2.3 Classification Sensible au Contexte "CardioVascular Disease" :

- Nous avons vu que la précision des modèles de perception textuelle dépasse le modèle général lorsque l'âge est compris entre 30 et 50 ans ([30,50]) mais que la précision générale du modèle l'emporte lorsque l'âge est compris entre 51 et 65 ans ([51,65]).
- Nous allons donc créer une combinaison des trois premiers modèles d'approche de la conscience du Contexte et du modèle général et classer par l'un des modèles en fonction de l'âge du patient.
- Les étapes à suivre sont donc les mêmes que nous avons suivi dans le modèle précédent.
- Nous avons choisi un exemple (figure 4.9) et testé avec le modèle qui a donné un meilleur résultat.

	AGE	GENDER	HEIGHT	WEIGHT	AP_HIGH	AP_LOW	CHOLESTEROL	GLUCOSE	SMOKE	ALCOHOL	PHYSICAL_ACTIVITY	CARDIO_DISEASE
6773	61	1	155	53	120	80	1	1	0	0	1	1

FIGURE 4.9 – Exemple du Dataset2

- Prédire si le patient a une maladie cardiaque ou non. (Figure 4.10)

THIS patient have CARDIO DISEASE

FIGURE 4.10 – Prédire si le Patient a la Maladie

4.3.3 La Troisième Expérience "Shipment Tracking"

En appliquant les étapes que nous faites dans les expériences précédentes .
Le tableau (4.6) suivant montre les résultats obtenus.

Modèle Général		
Accuracy En Utilisant la Validation Croisée		
RandomForest	SVM	Arbre de Décision
0.73875 « 73,875 % »	0.7375 « 73.75 % »	0.7393 « 73.93 % »

TABLE 4.6 – Résultat du Modèle Général en Utilisant la Validation Croisée

4.3.3.1 Identification et Modélisation (Représentation) du Contexte Pour (Shipment Tracking)

4.3.3.1.1 Identification du Contexte :

Comme nous l'avons déjà mentionné en haut, notre modèle ne contient pas de colonne contextuelle, nous étions obligés de tester plusieurs colonnes afin d'atteindre notre but . Les deux premières colonnes (ModeOfShipment et Prix) que nous avons choisies ont donné un résultat meilleur que celui du modèle général mais leurs augmentations étaient trop minable, par contre la dernière colonne (Weight) que nous avons choisie a donné de très bon résultat .

Nous allons détailler notre identification du Contexte comme suit :

1.Contexte ” Mode of shipment ”

Nous avons divisé cette colonne en trois catégories :

- Catégorie 1 :Ship
- Catégorie 2 :Flight
- Catégorie 3 : Road

Après avoir fait la catégorisation, nous allons entraîner les catégories en utilisant l'algorithme Decision Tree .

Le tableau 4.7 montre les résultats obtenus de chaque catégorie

Modèle Contextuel (ModeOfShipment)		
Modèle 1 (Ship)	Modèle 2 (Flight)	Modèle 3 (Road)
Accuracy (Decision Tree) en Utilisant la Validation Croisée		
0.6338 « 63.38 % »	0.6822 « 68.22 % »	0.6307 « 63.07 % »

TABLE 4.7 – Résultat Après Optimisation

• Ce que nous remarquons d'après le tableau ci-dessus , il n y a pas d'amélioration .
Donc nous étions obligées de tester avec une autre colonne.

2.Contexte ” prix ”

En utilisant cette colonne (prix) , nous constatons que nous avons obtenu des résultats un peu meilleurs que celui du modèle général.

Le tableau 4.8 ci-dessous montre ce que nous avons obtenu, dans cette colonne, nous avons testé avec l'algorithme Arbres de Décision.

Modèle Contextuel (Price)			
Modèle 1	Modèle 2	Modèle 3	Modèle 4
Prix Compris Entre 96 ET 150	Prix Compris Entre 151 ET 200	Prix Compris Entre 201 ET 250	Prix Compris Entre 251 ET 310
Accuracy (RandomForest)			
0.7685 « 76.85 % »	0.75 « 75 % »	0.7181 « 71.81 % »	0.6996 « 69.96 % »

TABLE 4.8 – Résultats du Contexte ” Prix ”

• À partir des expérimentations ci-dessus, nous voyons que le Contexte comprend que la précision des modèles dépasse le modèle général où le coût du produit est compris entre 96 et 200 ([96,200]) mais que la précision générale du modèle l'emporte lorsque le coût du produit est compris entre 201 et 310 ([201 310]).

- Nous allons essayer de tester nos modèles de prise de conscience des Contextes avec le modèle général sur le même ensemble de tests que nous avons testé dans le modèle général et voir si la précision augmentera.

La division sera la suivante :

- Prix ≥ 150
- Prix : [150, 200]
- Prix > 200

Les résultats obtenus :

Précision du premier modèle = 0.7619047619047619

Précision du deuxième modèle = 0,7717391304347826

Précision du troisième modèle = 0,6808510638297872

D'où la précision globale des modèles combinés = 0,7381649853897773 .

- D'après toutes les évaluations précédentes, nous remarquons qu'il y a une amélioration de 72% à 73 %, mais c'est toujours peu. Donc d'après les statistiques de cet ensemble (voir le figure 4.11).

Nous remarquons que la colonne **Weight** a une grande influence sur ce data néanmoins la colonne **Discount_Offred** classée première, nous l'avons testée. Mais le résultat n'était toujours pas performant.

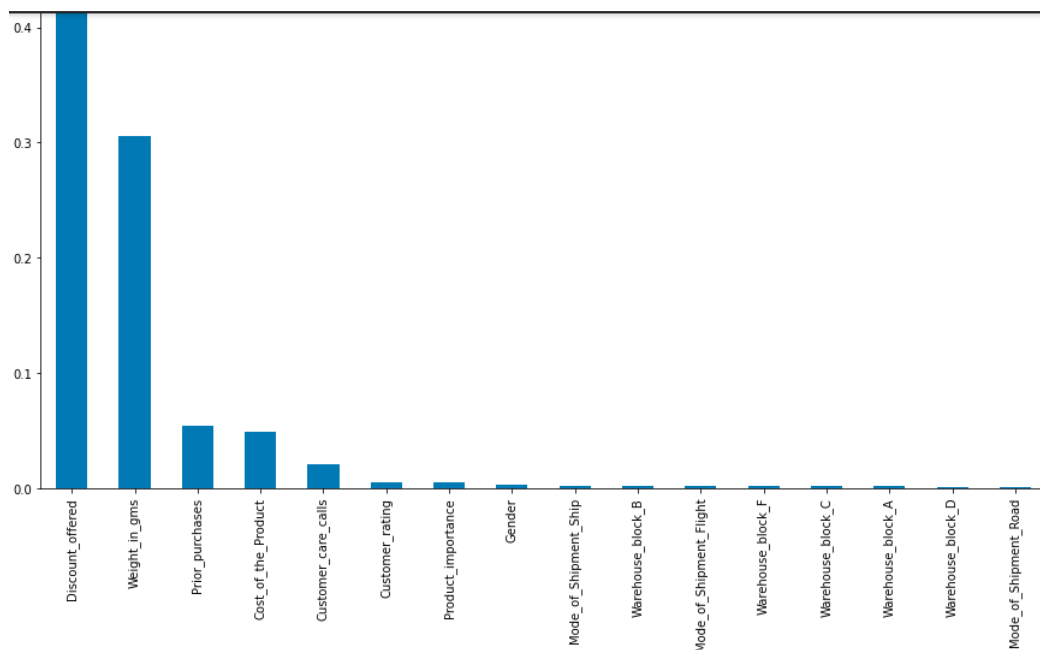


FIGURE 4.11 – Statiques et Caractéristiques du Dernier DataSet (Shipment Tracking)

c'est pour cette raison, nous l'avons choisie comme autant qu'une colonne contextuelle en refaisant les mêmes étapes .

3. Contexte ” Weight ”

- Nous avons testé l'entraînement des sous-ensembles avec l'algorithme **RandomForest**. Le tableau (4.9) suivant nous montre les résultats pour cette colonne .

Modèle Contextuel (Weight)			
Modèle 1	Modèle 2	Modèle 3	Modèle 4
Poids \leq 2000	Poids Compris Entre 2000 ET 4000	Poids Compris Entre 4000 ET 5000	Poids $>$ 5000
Accuracy (RandomForest) En Utilisant la Validation Croisée			
0.7444 « 74.44 % »	1 « 100 % »	0.64 « 64 % »	0.66 « 66 % »

TABLE 4.9 – Optimisation des Résultats pour la Colonne Weight

- À partir des expérimentations ci-dessus, nous voyons que le Contexte comprend que la précision des modèles dépasse le modèle général où le poids (Weight) du produit est ≤ 2000 et lorsque le poids est compris entre ($2000 > \text{poids} \leq 4000$) mais que la précision générale du modèle l'emporte lorsque le poids du produit est compris entre 4000 et 5000 ($[4000 - 5000]$).

- Nous allons essayer de tester nos modèles de prise de conscience des Contextes avec le modèle général sur le même ensemble de tests que nous avons testé dans le modèle général et voir si la précision augmentera.

La division sera la suivante :

- Weight ≤ 2000
- Weight $[2000,4000]$
- Weight > 4000

Les résultats obtenus :

La précision de la classe(Précision) Du premier modèle = 0.7802197802197802

La précision de la classe(Précision) Du deuxième modèle = 1.0

La précision de la classe(Précision) Du modèle général= 0.6868131868131868

D'où La précision globale des modèles combinés = 0.8223443223443224, maintenant nous pouvons de dire que nous sommes arrivées à notre but et qu'avec cette colonne (Weight) , nous avons eu une bonne Amélioration des résultats.

4.3.3.1.2 Modélisation du Contexte (Représentation) :

Comme nous l'avons fait avec le premier et le 2 ème modèle, c'est la même chose, après avoir identifié le Contexte et l'entraîner. Maintenant, nous allons essayer de créer un profil contextuel en utilisant le RDF.

Après avoir extrait Le profil contextuel . Il nous apparut long et illisible. Pour cette

raison, nous l'avons divisé en sous-graphe pour qu'il soit bien compréhensible. « Voir les figures 4.12 ; 4.13 ; 4.14 »

ET nous avons sauvegardé la figure du Graphe dans notre Drive et voici son URI :

<https://drive.google.com/file/d/1TKDZhmEDEolYMIHqnUoU-Cgi3p72IzpL/view?usp=sharing>

Nous avons fait la syntaxe RDF/XML pour la modélisation du Contexte, elle sera comme suite.

Modélisation des Contextes

” Shipment Tracking ” Syntaxe Rdf/XML

```
<rdf :RDF
```

```
xmlns :rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```

```
xmlns :dbo="http://dbpedia.org/ontology/"
```

```
xmlns :OntoDM="http://kt.ijs.si/panovp/OntoDM/"
```

```
xmlns :ex="http://kt.ijs.si/panovp/OntoDM/"
```

```
xmlns :rdfs="http://www.w3.org/2000/01/rdf-schema#">
```

```
<rdf :Description rdf :about="http://www.dbpedia.org/resource/DataSET">
```

```
<ex :name xml :lang="an">Shipment Tracking</ex :name>
```

```
<rdf :type rdf :resource="http://fr.dbpedia.org/resource/DataSet pour la classification/prédiction"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/ID"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WarehousBlock"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Mode Of Shipment"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Customer care calls"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/RainFall"/>
```

```
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Costumor rating"/>
```

```

<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Prior purchase" />
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Cost of product" />
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Gender" />
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Discount offered" />
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/Weight in gms" />
<ex :hascolumn rdf :resource="http://kt.ijs.si/panovp/OntoDM/WinSpeed9am" />
<ex :iscontext rdf :resource="http://www.dbpedia.org/resource/ModeOfShipment" />
<ex :iscontext rdf :resource="http://www.dbpedia.org/resource/Prix" />
<ex :iscontext rdf :resource="http://www.dbpedia.org/resource/Weight" />
</rdf :Description>

<rdf :Description rdf :about="http://www.dbpedia.org/resource/Prix">
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter1" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter2" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter3" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter4" />
</rdf :Description>

<rdf :Description rdf :about="http://www.dbpedia.org/resource/ModeOfShipment">
  <ex :Name>Flight</ex :Name>
  <ex :Name>Road</ex :Name>
  <ex :Name>Ship</ex :Name>
</rdf :Description>

<rdf :Description rdf :about="http://www.dbpedia.org/resource/Weight">
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter1" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter2" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter3" />
  <ex :value rdf :resource="http://kt.ijs.si/panovp/OntoDM/inter4" />
</rdf :Description>

```

```

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle1" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model1" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    96 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    150 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle2" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model2" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    151 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    200 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle3" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/ModèleGénéral" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    200 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    250 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle4" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/GeneralModel" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    251 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    310 ^^ xsd :integer </rdfs :end>
</rdf :Description>

```

```

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/intervalle5" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/GeneralModel" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    81 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    100 ^^ xsd :intger </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/inter1" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model1" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    1000 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    2000 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/inter2" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/Model2" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    2000 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    3000 ^^ xsd :integer </rdfs :end>
</rdf :Description>

<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/inter3" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/ModèleGénéral" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    3000 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    4000 ^^ xsd :integer </rdfs :end>
</rdf :Description>

```

```
<rdf :Description rdf :about="http ://kt.ijs.si/panovp/OntoDM/inter4" >
  <ex :usedmlmodel rdf :resource="http ://kt.ijs.si/panovp/OntoDM/ModèleGénéral" />
  <rdfs :start rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    4000 ^^ xsd :integer </rdfs :start>
  <rdfs :end rdf :datatype="http ://www.w3.org/2000/01/rdf-schema#integer" >
    5000 ^^ xdd :integer </rdfs :end>
</rdf :Description>
</rdf :RDF>
```

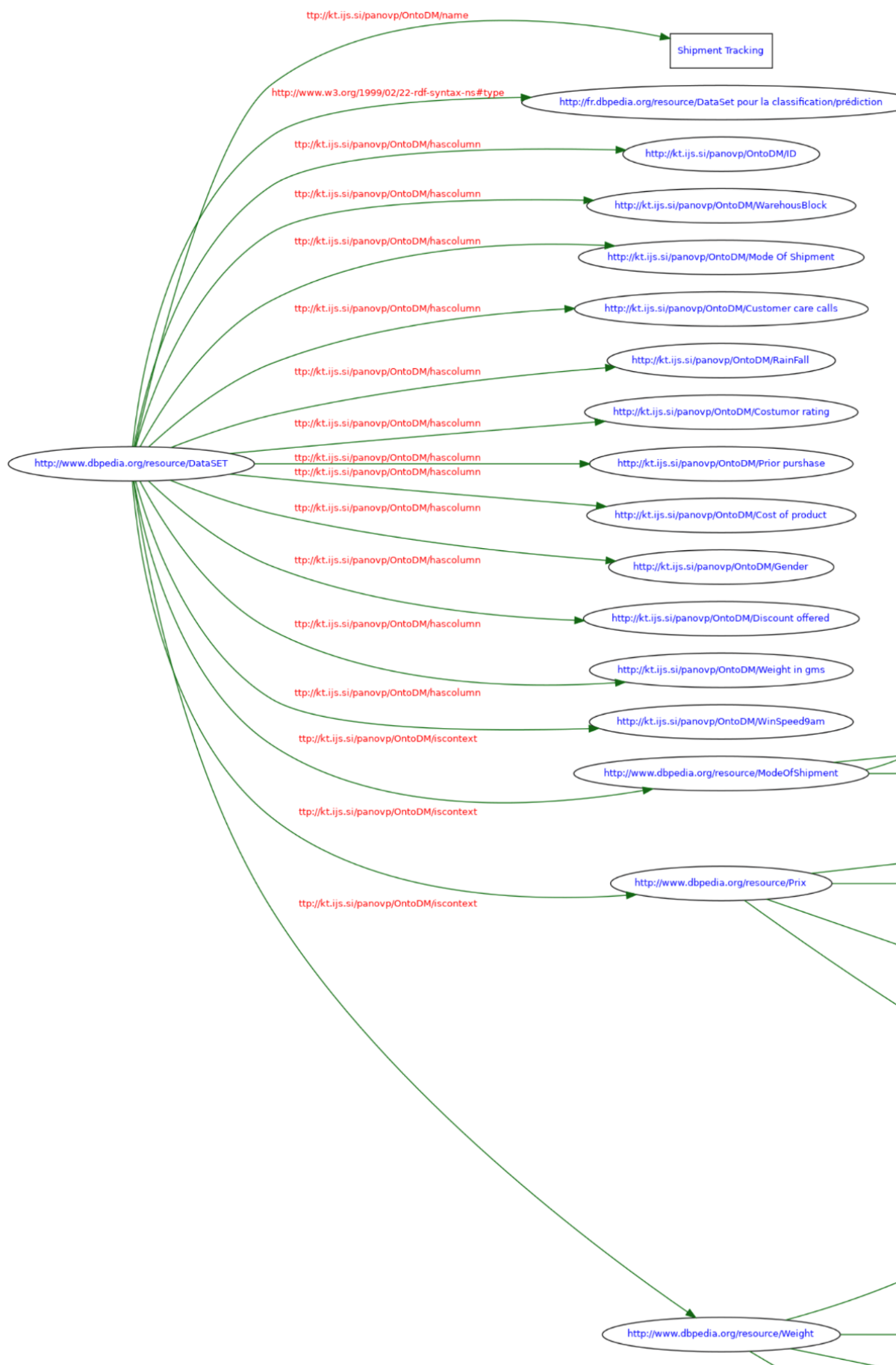


FIGURE 4.12 – Modélisation des Contextes Shipment Tracking .1 Graphe RDF

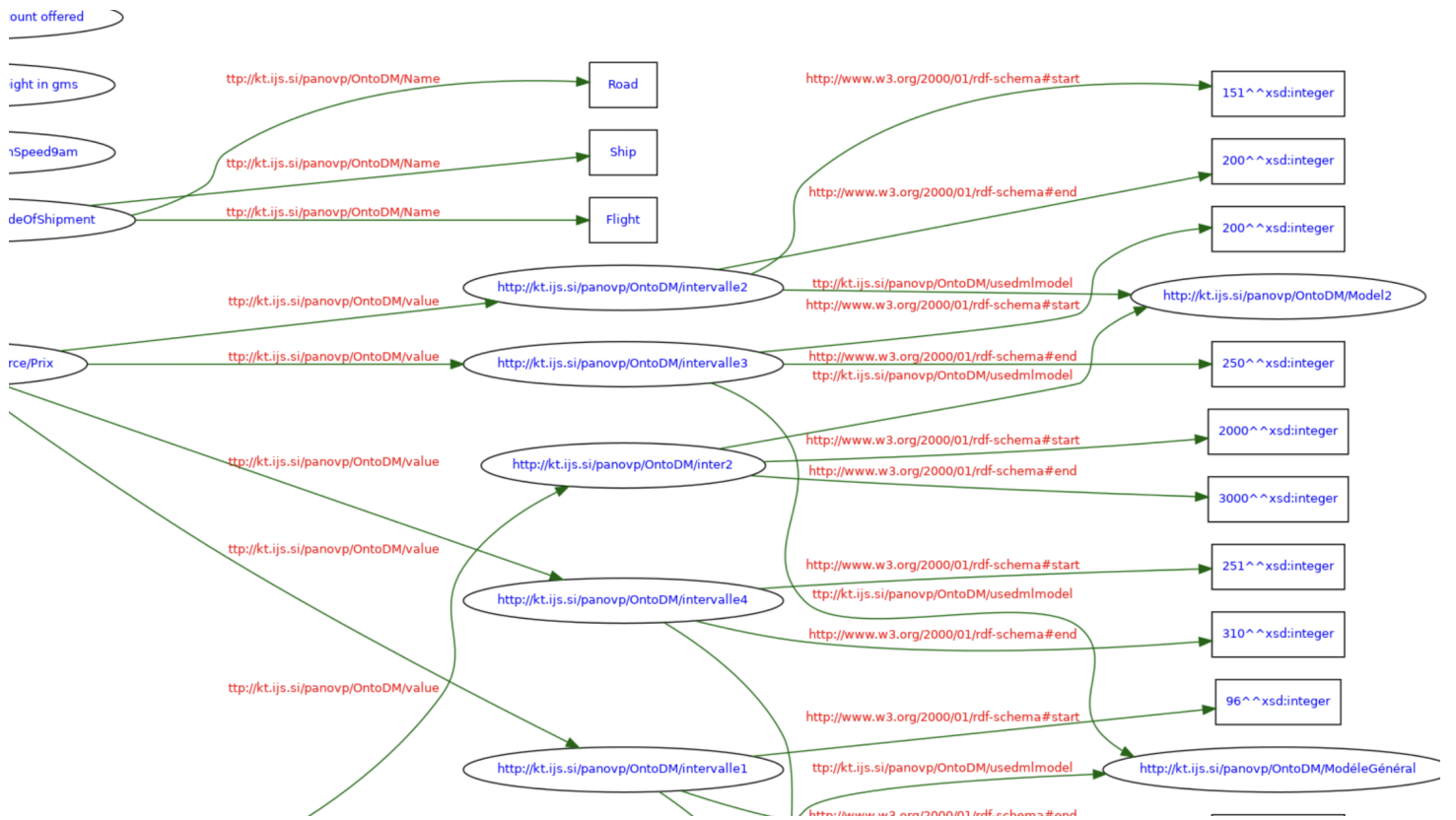


FIGURE 4.13 – Modélisation des Contextes Shipment Tracking .2 Graphe RDF



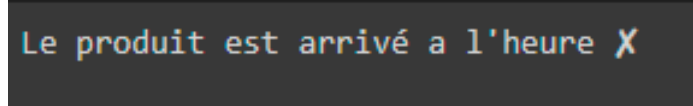
FIGURE 4.14 – Modélisation des Contextes Shipment Tracking .3 Graphe RDF

4.3.3.1.3 Classification Sensible au Contexte :

Comme nous avons l'habitude de faire la prédiction en temps réel de cet ensemble. Nous avons choisi un exemple du data (figure 4.15), et le testé avec le meilleur modèle obtenu. (figure 4.16)

Customer_care_calls	Customer_rating	Cost_of_the_Product	Prior_purchases	Product_importance	Gender	Discount_offered	Weight_in_gms	Warehouse_block
288	5	5	156	2	1	1	44	1561

FIGURE 4.15 – Exemple pour la prédiction d'un Produit



Le produit est arrivé a l'heure X

FIGURE 4.16 – Résultat Obtenu pour la Prédiction d'un Produit .

4.4 Conclusion

Dans ce chapitre, nous avons présenté les différents langages et outils de développement qui nous ont menés d'implémenter notre proposition.

Ensuite, nous avons décrit l'architecture de notre modèle d'une manière générale.

En outre, nous avons discuté sur ce que nous avons obtenu comme résultats en implémentant notre proposition (Solution). Nous clôturons notre chapitre par une comparaison entre notre modèle et celui des autres.

Conclusion Générale et Perspectives

Durant le travail de notre mémoire, nous avons présenté les différentes étapes de la réalisation de notre modèle optimisé pour la prédiction/classification d'un dataSet en utilisant une approche Contextuelle pour le Machine Learning.

Notre principal objectif était d'améliorer la performance et la précision d'un modèle de Machine Learning en utilisant le Contexte (il peut être une ou plusieurs colonnes)), le modélisé en RDF pour apporter plus de sémantique ce qui implique notre modèle devient sensible au Contexte.

Nous avons proposé un système sensible au Contexte cela signifie que plusieurs modèles du dataSets seront générés chaque modèle sera utilisé pour un Contexte bien défini dont le Contexte peut être une ou plusieurs colonnes du DataSet.

Après toutes les expériences que nous avons faites durant ce travail, nous constatons que l'algorithme des forêts aléatoires (RandomForest) donne toujours d'excellents résultats, il nous permet de gagner le temps en entraînant notre modèle. C'est pour cette raison, nous l'avons utilisé afin d'implémenter notre proposition.

Pour arriver à notre objectif, nous avons testé et validé notre proposition (Solution) sur différents Datastets tirés et choisis au hasard en utilisant plusieurs algorithmes du Machine Learning.

Perspectif :

- Sur le côté sémantique : la collecte des informations sera utilisée par des réseaux de capteurs, en se basant sur le concept d'ontologie pour que la comparaison d'une valeur à une autre ne soit pas directe, mais qu'elle soit en utilisant les concepts de similarité conceptuels avec un certain seuil de comparaisons.

Bibliographie

- [1] Matthew Chalmers. A historical view of context. *Computer Supported Cooperative Work (CSCW)*, 13(3-4) :223–247, 2004.
- [2] C Bastien. Contexte et situation. Houdé, O., Kayser, D., Koenig, O., Proust, J. et Rastier, F., *Dictionnaire des Sciences Cognitives*. Paris : PUF, 1998.
- [3] Thomas K Landauer, Peter W Foltz, and Darrell Laham. An introduction to latent semantic analysis. *Discourse processes*, 25(2-3) :259–284, 1998.
- [4] atrick Brézillon. Context in problem solving : a survey. *The Knowledge Engineering Review*, 14(01) :47–80, 1999
- [5] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *First Workshop on Mobile Computing Systems and Applications. WMCSA 1994*, pages 85–90. IEEE, 1994.
- [6] Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. Placing search in context : The concept revisited. In *Proceedings of the 10th international conference on World Wide Web*, pages 406–414. ACM, 2001
- [7] ristiana Bolchini, CA Curino, Giorgio Orsi, Elisa Quintarelli, Rosalba Rossato, Fabio A Schreiber, and Letizia Tanca. And what can context do for data ? *Communications of the ACM*, 52(11) :136–140, 2009.
- [8] John McCarthy. Notes on formalizing context. In *Proceedings of the 13th international joint conference on Artificial intelligence*, pages 555–560. Mor
- [9] Patrick Brézillon. Context dynamic and explanation in contextual graphs. In *Modeling and Using Context*, pages 94– 106. Springer, 2003.

- [10] Disseminating active map information to mobile hosts. *IEEE Network : The Magazine of Global Information Exchange*, 8(5) :22– 32.
- [11] Arthur H Van Bunningen, Ling Feng, and Peter MG Apers. Context for ubiquitous data management. In *International Workshop on Ubiquitous Data Management, 2005. UDM 2005.*, pages 17–24. IEEE, 2005
- [12] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *First Workshop on Mobile Computing Systems and Applications. WMCSA 1994*, pages 85–90. IEEE, 1994
- [13] Nick Ryan, Jason Pascoe, and David Morse. Enhanced reality fieldwork : the context aware archaeological assistant. *Bar International Series*, 750 :269–274, 1999
- [14] Brown, P., Bovey, J., and Chen, X. (1997). Context-aware applications : from the laboratory to the marketplace. *IEEE Personal Communications*, 4(5) :182–196.
- [15] Abowd, G., A.K., Brown, P., Davies, N., Smith, M., and Steggles, P. (1999). Towards a better understanding of context and context-awareness. In Gellersen, H.-W., editor, *Handheld and Ubiquitous Computing*, volume 1707, pages 304– 307. Springer Berlin Heidelberg
- [16] Moran, T. and Dourish, P. (2001). Introduction to this special issue on context aware computing. *Human-Computer Interaction*, 16(2) :87–9
- [17] dey, A., Abowd, G., and Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *HumanComputer Interaction*, 16(2) :97–166.
- [18] Dey, A. (2000). *Providing Architectural Support for Building Context-aware Applications*. PhD thesis, Atlanta, GA, USA.
- [19] Gregory D Abowd, Anind K Dey, Peter J Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and contextawareness. In *Handheld and ubiquitous computing*, pages 304–307. Springer, 1999
- [20] Guanling Chen, David Kotz, et al. A survey of context-aware mobile computing research. Technical report, Technical Report TR2000-381, Dept. of Computer Science, Dartmouth College, 2000.

- [21] Suan Khai Chong, Ian McCauley, Seng Wai Loke, and Shonali Krishnaswamy. Context-aware sensors and data muffling. Context awareness for selfmanaging systems (devices, applications and networks) proceeding, pages 103–117, 2007.
- [22] gu jun zhong. Context aware computing. Journal of East China Normal University, 5 :1–20, 2009.
- [23] Xiao Hang Wang, Da Qing Zhang, Tao Gu, and Hung Keng Pung. Ontology based context modeling and reasoning using owl. In Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops, 2004., pages 18–22. IEEE, 2004
- [24] onghai Guan, Weiwei Yuan, Sungyoung Lee, and Young-Koo Lee. Context selection and reasoning in ubiquitous computing. In Intelligent Pervasive Computing, 2007. IPC.
- [25] Stamatia Rizou, Kai Haussermann, Frank Durr, Nazario Cipriani, and Kurt Rothermel. A system for distributed context reasoning. In Sixth International Conference on Autonomic and Autonomous Systems (ICAS), pages 84–89. IEEE, 2010.
- [26] Karen Henriksen. A framework for context-aware pervasive computing applications. University of Queensland Queensland, Australia, 2003
- [27] Schilit, B. and Theimer, M. (1994). Disseminating active map information to mobile hosts. IEEE Network : The Magazine of Global Information Exchange, 8(5) :22– 32.
- [28] Makssoud, K. A. (2008). Systeme d’Acces Personnalise a l’Information : Application au Domaine Medical . PhD thesis, Institut National des Sciences Appliquees de Lyon.
- [29] Chaari, T. (2007). Adaptation d’applications pervasives dans des environnements multi-contextes. PhD thesis, INSA de Lyon.
- [30] Lalana Kagal, Vladimir Korolev, Sasikanth Avancha, Anupam Joshi, Tim Finin, and Yelena Yesha. Centaurus : an infrastructure for service management in ubiquitous computing environments. Wireless Networks, 8(6) :619–635, 2002.
- [31] Albert Held, Sven Buchholz, and Alexander Schill. Modeling of context information for pervasive computing applications. In Proceeding of the World Multiconference on Systemics, Cybernetics and Informatics, 2002.

- [32] Oleksiy Khriyenko and Vagan Terziyan. Context description framework for the semantic web. In Proceedings Context 2005 Context representation and reasoning workshop, Paris (FR), 2005
- [33] Henricksen, Karen, Jadwiga Indulska, and Andry Rakotonirainy. "Modeling context information in pervasive computing systems." Pervasive Computing. Springer Berlin Heidelberg, 2002. 167-180.
- [34] Henricksen, Karen, and Jadwiga Indulska. "Developing context-aware pervasive computing applications : Models and approach." Pervasive and mobile computing 2.1 (2006) : 37-64.
- [35] McCarthy, J. et Buvac, S. (1997). Formalizing Context (Expanded Notes). Computing Natural Language, 13-50. Récupéré de [http ://wwwformal.stanford.edu/jmc/mccarthy -buvac-98/ context. pdf](http://wwwformal.stanford.edu/jmc/mccarthy-buvac-98/context.pdf)
- [36] Perora, C., Zaslavsky, A., Christen, P. et Georgakopoulos, D. (2014). Context aware computing for the internet of things : A survey. IEEE Communications Surveys and Tutorials, 16(1), 414-454.
- [37] Wang, X. H., Zhang, D. Q., Gu, T. et Pung, H. K. (2004). Ontology based context modeling and reasoning using OWL. In IEEE Annual Conference on Pervasive.
- [38] BERNERS-LEE Tim, HENDLER James et LASILLA Ora (2001). The Semantic Web, Scientific American,.
- [39] Ora Lassila et Ralph Swick, "Resource Description Framework (RDF) Model and Syntax Specification", W3C Recommendation 22 February 1999, World Wide Web Consortium, Cambridge (MA) ; available on-line as.
- [40] Dan Brickley et R.V.Guha, "Resource Description Framework (RDF) Schema Specification 1.0", W3C Candidate Recommendation 27 March 2000, World Wide Web Consortium, Cambridge (MA) ; available on-line as.
- [41] Mahdi Gueffaz, Sylvain Rampacek, Christophe Nicolle, Des outils pour la vérification des graphes sémantiques , France , Septembre 2011
- [42] [http ://www.uqac.ca/](http://www.uqac.ca/)
- [43] McBride B. (2004), RDF Vocabulary Description Language 1.0 : RDF Schema, W3C Recommendation, [http ://www.w3.org/TR/rdf-schema/](http://www.w3.org/TR/rdf-schema/).

- [44] Van Harmelen F., Horrocks I., Peter F. (2001) : A model theoretic semantics for DAML+OIL. W3C Note, 18 December 2001. <http://www.w3.org/TR/2001/NOTE-daml+oil-model-20011218>
- [45] Fensel D., Horrocks I., Van Harmelen F., Decker S., Erdmann M., et Klein M. (2000) : Oil in a nutshell. In 12th International Conference on Knowledge Engineering and Knowledge Management EKAW2000, Juan les-Pins, France.
- [46] A. Napoli, "une introduction aux logiques de descriptions" N 3314, Décembre 1997
- [47] McGuinness D.L. et Van Harmelen F. (2004) : OWL Web Ontology Language Overview, <http://www.w3.org/TR/owl-features/>
- [48] Thimira Amaratunga , « Deep Learning on Windows ,Building Deep Learning Computer Vision Systems on Microsoft Windows »
- [49] Labiad ALI, SÉLECTION DES MOTS CLÉS BASÉE SUR LA CLASSIFICATION ET L'EXTRACTION DES RÈGLES D'ASSOCIATION , juin 2017.
- [50] [researchgate.net/Diagramme-dun-apprentissage-supervise](https://www.researchgate.net/publication/312214444-Diagramme-dun-apprentissage-supervise)
- [51] A State-of-the-Art Survey on Deep Learning Theory and Architectures
- [52] <https://searchentrepriseai.techtarget.com/definition/supervised-learning?amp=1> et <https://www.facebook.com/techtarget/?fbclid=IwAR3C0st1E7WWLM-qthHaLRIb0AJq4zB6e5j11wBFEEZPz9DudCTBZXdyO2U>
- [53] 52. JOSE ALEJANDRO LOPEZ GONZALEZ,EXPLORATION DES ARBRES DE DÉCISION ET DES SUPPORT VECTORMACHINES EN VUE D'APPLICATIONS DANS L'ANALYSE DE TEXTE, JUIN 2016.
- [54] <https://www.datascience.com>
- [55] <https://dataanalyticspost.com>
- [56] <https://ichi.pro/fr/tout-ce-qu-il-faut-savoir-support-vector-machines-svm-27379677884233?fbclid=IwAR19CzUsjE-AJRHHf-W3eGYuEkfUcoHF27Py-k-YOn0rFElxEI9jWzIM-bo>
- [57] <https://eric.univ-lyon2.fr>
- [58] Hassan hilal ,APPLICATION DE LA CLASSIFICATION TEXTUELLE POUR L'EXTRACTION DES RÈGLES D'ASSOCIATION MAXIMALES, avril 2009.
- [59] [Juripredis.com/Démystifier le Machine Learning, Partie 2 : les Réseaux de Neurones artificiels](https://www.juripredis.com/actualites/2016/07/20/demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels/)

- [60] <https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels>
- [61] <http://www-igm.univ-mlv.fr/dr/XPOSE2001/seguin/final/ReseauNeuro.html>
- [62] <https://analyticsindiamag.com/10-model-evaluation-techniques-every-machine-learning-enthusiast-must-know/>, avril 2019
- [63] <https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures>, 2018
- [64] Nathalia Nascimento , Carlos Lucena , Paulo Alencar , Donald Cowan , A context-Aware Machine Learning-Based Approach , October 2018
- [65] José Hernández-Orallo, Adolfo Martínez-Usó, Ricardo BC Prudêncio, Meelis Kull, Peter Flach, Chowdhury Farhan Ahmed, and Nicolas Lachiche. 2016. Reframing in context : A systematic approach for model reuse in machine learning. *AI Communications* 29, 5 (2016), 551–566
- [66] Mariam El Mezouar, Feng Zhang, and Ying Zou. 2016. Local versus global models for effort-aware defect prediction. In *Proceedings of the 26th Annual International Conference on Computer Science and Software Engineering*. IBM Corp., 178–187
- [67] <https://www.kaggle.com/prachi13/customer-analytics>
- [68] <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>
- [69] <https://www.kaggle.com/jsphyg/weather-dataset-rattle-package>
- [70] <https://ledatascientist.com/google-colab-le-guide-ultime/>, consulté le 05-06-2021.
- [71] <https://www.lebigdata.fr/python-langage-definition> , consulté le 25/02/2021.
- [72] <https://github.com/scikit-learn/scikit-learn> , consulté le 05/06/2021.
- [73] <https://github.com/RDFLib/rdfliib>, consulté le 05/06/2021.
- [74] anov, P., Soldatova, L., Dzeroski, S. : Towards an ontology of data mining investigations. *Lecture Notes in Artificial Intelligence* 5808, 257–271 (200