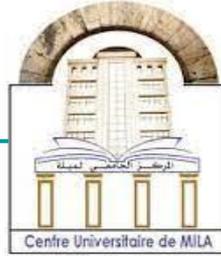


الجمهورية الجزائرية الديمقراطية الشعبية

République Algérien démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et De la Recherche Scientifique



Centre Universitaire Abdelhafid Boussouf – Mila

Institut des Mathématiques et informatique

Département d'Informatique

Mémoire préparé pour obtenir le diplôme de Master en

Informatique

Spécialité : Intelligence artificielle et ses applications (I2A)

Thème

Intégration des Techniques d'Apprentissage
Automatique dans les Marchés Publics

Soutenue devant le jury :

Réalisé par :

- Hettab Meryem

- Meghzili Asma

Président : Dr. Djaaboub Salim.

Examineur : Dr. Guettiche Mourad.

Encadré par Dr. Meghzili Said.

L'année scolaire

2024/2025

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Remerciement

Avant tout, nous exprimons notre profonde gratitude à Dieu Tout-Puissant pour la force, la santé, la persévérance et la patience qu'Il nous a accordées tout au long de la réalisation de ce travail.

Nous adressons nos remerciements les plus sincères à notre directeur de mémoire, **Dr. Meghzili Said**, pour son accompagnement bienveillant, ses conseils éclairés et son soutien constant. Sa disponibilité et son expertise ont été d'une grande richesse et ont largement contribué à la qualité de ce travail.

Nos remerciements vont également aux membres du jury, **Dr. Guettiche Mourad** et **Dr. Djaboub Salim**, pour avoir accepté d'évaluer ce mémoire et pour leurs remarques constructives qui ne manqueront pas d'enrichir notre démarche scientifique et de nourrir nos travaux futurs.

Nous exprimons aussi notre reconnaissance à l'ensemble de nos enseignants, qui ont su nous transmettre leur savoir avec passion et rigueur tout au long de notre parcours universitaire.

Enfin, nous prions Allah d'agréer ce modeste effort, de le rendre utile et bénéfique, et de nous accorder à tous réussite et épanouissement. Que ce travail puisse, si Dieu le veut, inspirer et guider les étudiants des générations à venir.

Dédicace

Avant tout, je rends grâce à Allah de m'avoir donné le courage, la santé, la force pour mener à bien ce travail.

Ce travail m'a enseigné que les efforts constants et la patience sincère sont les clés de toute réussite durable.

Je dédie cette réussite à moi-même, ainsi qu'à ma chère mère, à mon père et à ma précieuse famille, dont la tendresse a été ma lumière dans l'obscurité et dont l'amour a nourri ma persévérance face à chaque difficulté.

Je remercie chaleureusement mon encadreur, **Dr. Meghzili Said**, ainsi que tous mes professeurs respectés, qui ont été de véritables phares de savoir et de guidance.

À tous ceux qui m'ont soutenu, qui se sont tenus à mes côtés par une parole, un conseil, ou une prière sincère, qu'Allah vous récompense tous en bien.

Meghzili Asma

Dédicace

Dédicace

Tout d'abord, je loue Allah le Très-Haut, qui m'a accompagné de Sa miséricorde infinie tout au long de ce parcours. Grâce à Sa bénédiction et à Sa lumière, j'ai pu accomplir chaque étape et surmonter chaque obstacle.

Je souhaite exprimer ma profonde gratitude à mon directeur de recherche, le **Docteur Said Meghzili**, pour ses précieux conseils, son orientation constante et sa bienveillance tout au long de ce projet. Je remercie également chaleureusement les membres du jury, le **Docteur Djaaboub Salim** et le **Docteur Guettiche Mourad**, pour l'intérêt qu'ils ont porté à cette thèse et pour leurs commentaires précieux.

À ma chère famille, source intarissable d'amour, de soutien et de motivation, je dédie cette réussite. Grâce à vos encouragements, je suis arrivé(e) jusqu'ici aujourd'hui.

À mon oncle, le **Professeur Hettab Abdelkamel**, dont le soutien moral, les encouragements constants et la présence affectueuse ont été pour moi une source de réconfort et de force. Ce travail est le fruit de votre confiance en moi. Merci du fond du cœur.

J'ai obtenu mon diplôme parce que les rêves ne se donnent pas ; ils s'arrachent par la force de la volonté, l'obstination de l'ambition et à une poursuite sincère.

Hettab Meryem

Résumé

Face à la complexité croissante des procédures de passation des marchés publics et au besoin d'améliorer la transparence et l'efficacité dans le choix des offres, cette recherche repose sur un ensemble de données extraites des archives des marchés publics de la commune de Mila et d'autres sources. Elle vise à développer un système prédictif automatisé capable d'anticiper le candidat gagnant et le nombre de candidats, ainsi que les caractéristiques des cahiers de charges, notamment les offres techniques.

Pour cela, une approche combinant l'apprentissage automatique, l'apprentissage profond et le traitement du langage naturel (NLP) a été adoptée. Plusieurs algorithmes performants comme Random Forest, Extra Trees et KNN ont été comparés après transformation des données textuelles via TF-IDF, et renforcés par des techniques d'augmentation de données.

Des modèles profonds tels que DNN et GRU ont été implémentés pour améliorer la précision, tandis que des modèles basés sur la génération augmentée par récupération (RAG) ont permis des prédictions contextuelles plus intelligentes. L'ensemble a été intégré dans une interface web interactive facilitant des prédictions en temps réel et une gestion efficace des processus de marchés publics.

Mots-clés : marchés publics, apprentissage automatique, apprentissage profond, traitement du langage naturel, TF-IDF.

Abstract

Given the increasing complexity of public procurement procedures and the need to enhance transparency and efficiency in offer selection, this research is based on a dataset extracted from the public procurement archives of the municipality of Mila and other sources. It aims to develop an automated predictive system capable of anticipating the winning candidate and the number of candidates, as well as the characteristics of the specifications, including technical offers.

To achieve this, an approach combining machine learning, deep learning, and natural language processing (NLP) was adopted. Several high-performing algorithms such as Random Forest, Extra Trees, and KNN were compared after transforming textual data using TF-IDF and enhanced with data augmentation techniques.

Deep models such as DNN and GRU were implemented to improve prediction accuracy, while retrieval-augmented generation (RAG) models enabled more intelligent contextual predictions. All these solutions were integrated into an interactive web interface allowing real-time predictions and efficient management of public procurement processes.

Keywords: public procurement, artificial intelligence, machine learning, deep learning, natural language processing, TF-IDF.

ملخص

نظرا للتعقيد المتزايد في اجراء ابرام الصفقات العمومية، وأهمية ضمان الشفافية والفعالية وقابلية التنبؤ في اختيار العروض، تعتمد هذه الدراسة على مجموعة بيانات مستخرجة من أرشيف الصفقات العمومية لبلدية ميله ومصادر أخرى. وتهدف إلى تطوير نظام تنبؤي آلي قادر على توقع المرشح الناجح وعدد المرشحين، وكذلك خصائص المواصفات، بما في ذلك العروض الفنية.

لتحقيق هذا الهدف، تم اعتماد منهجية تجمع بين التعلم الآلي، والتعلم العميق، ومعالجة اللغة الطبيعية (NLP). وقد تم مقارنة عدة خوارزميات فعالة مثل Random Forest و Extra Trees و KNN بعد تحويل النصوص باستخدام تقنية TF-IDF، وتعزيزها بتقنية توليد البيانات .

كما تم استخدام نماذج عميقة مثل الشبكة العصبية العميقة (DNN) ووحدات GRU لتحسين دقة التنبؤ، الى جانب نماذج تعتمد على التوليد المعزز بالاسترجاع (RAG) لتقديم تنبؤات سياقية أكثر ذكاء.

وتم دمج هذه الحلول ضمن واجهة ويب تفاعلية تتيح اجراء التنبؤات في الوقت الفعلي وتساهم في تحسين إدارة الصفقات العمومية.

الكلمات المفتاحية: الصفقات العمومية، الذكاء الاصطناعي، التعلم الآلي، التعلم العميق، معالجة اللغة الطبيعية،

.TF-IDF

Table des matières

Remerciement	I
Dédicace.....	II
Dédicace.....	III
Résumé.....	VIII
Abstract	IX
ملخص.....	X
Table des matières	VIII
Liste des figures	XI
Liste des tableaux	XIV
Liste Des Abréviations	XV
Introduction Générale.....	1
Chapitre 01 : Présentation de l'organisme du lieu de stage et service des marchés public	2
1.1-Introduction	3
1.2-Presentation de la commune de Mila :	3
1.2.1-Les composantes de la commune de Mila :	3
1.3-Service des marchés publics :.....	5
1.3.1-Les unités de service des marchés publics :.....	6
1.3.1.1-Bureau des équipements :.....	6
1.3.1.2-Bureau de Suivi :.....	6
1.3.2- la loi 15-247 :	7
1.4-Application de l'Intelligence Artificielle dans les Marchés Publics :	7
1.5-Conclusion.....	8
Chapitre 02 : Techniques d'IA.....	9

Table des matières

2.1-Introduction	10
2.2- Intelligence artificielle :.....	10
2.3-Apprentissage automatique :	10
2.3.1-Apprentissage supervisé :	11
2.3.2-Apprentissage non supervisé :	14
2.3.3-Apprentissage par renforcement :	14
2.3.4-Les phases de l'apprentissage automatique :	14
2.4-Apprentissage profond (Deep Learning - DL) :	15
2.5-Principe et fonctionnement du Deep Learning :.....	15
2.6-Méthodes du Deep Learning :	16
2.6.1-Réseau de neurones profonds (Deep Neural Network « DNN ») :.....	16
2.6.2-Réseau de neurones à convolution (Convolutional Neural Network « CNN ») :	16
2.6.3- Réseau de Neurones Récurrent (Recurrent Neural Network « RNN ») :	17
2.6.4- L'unité récurrente à portes (Gated Recurrent Unit « GRU ») :	18
2.7-Traitement automatique du langage naturel (NLP) :	19
2.7.1-Transformer :	19
2.7.2-BERT (Bidirectional Encoder Representations from Transformers) :	19
2.7.3- génération augmentée de récupération (retrieval augmented generation « RAG ») :.....	20
2.7.4 - Fusion dans le décodeur (Fusion-in-Decoder (FiD)) :	20
2.8-Techniques de prétraitement :	21
2.8.1- La tokenisation :.....	21
2.8.2- La lemmatisation :.....	21
2.8.3- TF-IDF :	21
2.8.4- L'encodage d'étiquettes « Label Encoding » :.....	22
2.8.5- Transformateur de phrases (Sentence Transformer) :.....	22

Table des matières

2.8.5.1-Modèle Distiluse-base-multilingual-cased-v1 :	22
2.8.5.2-Modèle Paraphrase-multilingual-mpnet-base-v2 :	22
2.8.5.3-Modèle paraphrase-MiniLM-L6-v2 :	23
2.8.6-Modèle de Flan T5 Base :	23
2.9- Indicateurs de performance :	23
2.9.1- Précision « Accuracy» :	23
2.9.2-matrice de confusion :	23
2.10-Conclusion	24
Chapitre 03 : Validation et évaluation des performances des modèles proposés	25
3.1-Introduction	26
3.2- Logiciels et outils d'apprentissage profond :	26
3.4-Dataset : « _dataset 1»	28
3.5-Préparation de dataset :	31
3.5.1-Nettoyage et réorganisation de dataset :	31
3.5.2-Augmentation de dataset :	31
3.5.3-Encodage :	32
3.6-Les approches proposées :	32
3.6.1-Prédire le candidat gagnant et le nombre de candidats :	32
3.6.1.1-Modèles basés sur l'apprentissage automatique :	33
3.6.1.1.1- Modèle RandomForestClassifier :	35
3.6.1.1.2-Modèle ExtraTreesClassifier :	36
3.6.1.1.3-Modèle KNeighborsClassifier :	37
3.6.1.2-Modèles basés sur Deep Learning :	39
3.6.1.2.1- Modèle GRU :	41
3.6.1.2.2-Modèle DNN :	42
3.6.1.3-Modèles basés sur NLP :	44
3.6.1.3.1- Modèle de RAG :	46

Table des matières

3.6.1.3.2-Modele de FID :.....	47
3.6.2-Prédire l'ensemble des conditions du cahier des charges :.....	48
3.6.2.1-Modèles basés sur l'apprentissage automatique :	49
3.6.2.1.1- Modèle RandomForestClassifier :.....	51
3.6.2.1.2- Modèle ExtraTreesClassifier :.....	51
3.6.2.1.3- Modèle KNeighborsClassifier :.....	52
3.6.2.2-Modèle basé sur Deep Learning :.....	53
3.6.2.2.1-DNN :	53
3.6.2.3-Modèles basés sur NLP :.....	56
3.6.2.3.1-Modèle RAG :	58
3.6.2.3.2-Modèle de FID :.....	58
3.7-Conclusion.....	59
Chapitre 04 : Application Web de système de prédiction des marchés publics	60
4.1-Introduction	61
4.2-Présentation des outils et langages utilisés :.....	61
4.3-Interfaces graphiques du Système de prédiction des marchés publics :.....	62
4.3.1-Page principale :.....	62
4.3.2- L'interface de description des modèles utilisés :.....	63
4.3.3- Les domaines utilisés pour la prédiction des marchés publics :	65
4.3.4-Prédire le candidat :	66
4.3.5- Prédire les conditions de cahier des charges :.....	68
4.4-Conclusion.....	70
Conclusion générale	71
Bibliographies	73

Liste des figures

Figure 1.1: La commune de Mila.	3
Figure 1.2: Organigramme des directions de la commune de la wilaya de Mila.	4
Figure 1.3: Organigramme de la direction de l'administration générale et des finances.....	5
Figure 2.1: Exemple des chiffres manuscrits [2]......	11
Figure 2.2: Ensemble de données étiquetées. Chaque image d'entrée est associée avec la prédiction voulue [2].	12
Figure 2.3: Visualisation du modèle de Random Forest [5]......	12
Figure 2.4: Architecture du Extra Tree Classifier [7]......	13
Figure 2.5: Illustration du processus d'apprentissage K-NN (K plus proches voisins) avec classification binaire [8].	14
Figure 2.6: Ensemble de données non-étiquetées les images d'entrées [2].	14
Figure 2.7: Architecture d'un modèle Deep Learning [11].	16
Figure 2.8: Architecture d'un réseau de neurones profond [12].	16
Figure 2.9: Concept de réseau de neurones à convolution [13].	17
Figure 2.10: Concept de réseau de neurones récurrent [13]......	18
Figure 2.11: L'architecture de l'unité récurrente fermée de base (GRU) [15]......	19
Figure 2.12: Fonctionnement d'un Modèle RAG [20]......	20
Figure 2.13: Le modèle Fusion-In-Decoder (FID) [22].	20
Figure 3.1: Classement TIOBE du May 2025 des langages de programmation [32]......	27
Figure 3.2: Dataset utilise.....	28
Figure 3.3: entrées / sorties pour la prédiction du candidat gagnant et du nombre de candidats.	33
Figure 3.4: Architecture du modèle basé sur l'apprentissage automatique de prédiction du candidat gagnant et du nombre de candidats.	34
Figure 3.5: Précision par cible pour le modèle Random forest.	35
Figure 3.6: Matrices de confusion - nombre de candidats - pour le modèle RandomForestClassifier.	36
Figure 3.7: Précision par cible pour le modèle Extra tree.	36
Figure 3 8: Matrices de confusion - nombre de candidats - pour le modèle ExtraTreesClassifier.	37

Liste des Figures

Figure 3.9: Précision par cible pour le modèle KNN.	37
Figure 3.10: Matrices de confusion - nombre de candidats - pour le modèle KNeighborsClassifier.	38
Figure 3.11: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	39
Figure 3.12: Évolution des valeurs de la métrique de performance à l'aide du modèle GRU.	41
Figure 3.13: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	42
Figure 3.14: Évolution des valeurs de la métrique de performance à l'aide du modèle DNN.	42
Figure 3.15: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	44
Figure 3.16: Analyse de similarité documentaire et distribution des candidats pour le modèle RAG.	47
Figure 3.17: Analyse de similarité documentaire et distribution des candidats pour le modèle FID.	47
Figure 3.18: entrées / sorties pour la prédiction des conditions à partir des caractéristiques du projet.	49
Figure 3.19: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	50
Figure 3.20: Courbe de précision pour RandomForest.	51
Figure 3.21: Courbe de précision pour ExtraTreesClassifier.	52
Figure 3.22: Courbe de précision pour KNeighborsClassifier.	52
Figure 3.23: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	53
Figure 3.24: Évolution des valeurs de la métrique de performance à l'aide du modèle DNN.	55
Figure 3.25: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.	56
Figure 3.26: Courbe de RAG.	58
Figure 3.27: Courbe de FID.	58

Liste des Figures

Figure 4.1: Logo de Framework Visual Studio Code.....	61
Figure 4 2: Page principale.....	63
Figure 4.3: Modèles utilisés dans le système.	65
Figure 4.4: Domaines utilisés dans la prédiction des Marchés Publics.....	66
Figure 4.5: la page des informations de projet.	67
Figure 4.6: Résultats de la prédiction.	67
Figure 4.7: Message d’alerte.	68
Figure 4.8: L’interface des informations de projet.	69
Figure 4.9: Résultats de la prédiction.	69

Liste des tableaux

Tableau 3.1: La liste des colonnes de dataset.	30
Tableau 3.2: Les modèles utilisés basés sur l'apprentissage automatique avec leurs différences.	34
Tableau 3.3: Évaluation et comparaison des modèles basés sur l'apprentissage automatique.	39
Tableau 3.4: Entraînement des hyperparamètres des modèles basés sur Deep Learning. .	40
Tableau 3.5: Évaluation et comparaison des modèles DNN et GRU.	43
Tableau 3.6: Hyperparamètres des modèles RAG et FID.	45
Tableau 3.7: Comparaison des hyperparamètres des modèles RAG et FID.	46
Tableau 3.8: Évaluation et comparaison des deux modèles.	48
Tableau 3.9: Les modèles basés sur l'apprentissage automatique utilisés avec leurs différences.	50
Tableau 3.10: Hyperparamètres du modèle DNN.	54
Tableau 3.11: Division de dataset.	54
Tableau 3.12: Évaluation du modèle DNN.	55
Tableau 3.13: Hyperparamètres des modèles RAG et FID.	57

Liste Des Abréviations

IA	Intelligence Artificielle.
NLP	Natural language processing.
ML	Machine Learning.
KNN	K-Nearest Neighbour.
DL	Deep Learning.
DNN	Deep Neural Network.
CNN	Convolutional Neural Network.
RNN	Recurrent Neural Network.
GRU	Gated Recurrent Unit.
BERT	Bidirectional Encoder Representations from Transformers.
RAG	Retrieval Augmented Generation.
FID	Fusion-In-Decoder.
TF-IDF	Term Frequency- Inverse Document Frequency.

Introduction Générale

Dans un contexte de modernisation des administrations publiques, plusieurs structures administratives souffrent encore d'un déficit de numérisation, notamment dans la gestion des marchés publics. L'absence de systèmes numériques performants, associée à une exploitation essentiellement manuelle des données, rend les procédures souvent longues, opaques et peu efficaces. Cela limite la capacité des administrations à assurer une gestion optimale des appels d'offres, en particulier en l'absence d'outils intelligents permettant d'anticiper les résultats et d'analyser les aspects techniques des offres.

Face à ces défis, l'optimisation des processus liés aux marchés publics constitue un enjeu majeur pour garantir la transparence, l'efficacité et l'équité dans l'allocation des projets. La complexité croissante des appels d'offres et la nécessité d'améliorer la prévisibilité des résultats ont conduit à l'émergence de solutions innovantes basées sur l'intelligence artificielle. L'intégration de technologies avancées telles que l'apprentissage automatique, l'apprentissage profond et le traitement du langage naturel révolutionne désormais la capacité d'analyser, de comprendre et de prédire les dynamiques complexes des systèmes de marchés publics, permettant une automatisation intelligente de tâches critiques comme la prédiction des candidats gagnant, et l'optimisation des processus décisionnels.

Cette recherche s'inscrit dans cette démarche d'innovation en développant un système prédictif sophistiqué visant à améliorer la précision et l'efficacité des processus de marchés publics. L'étude se concentre spécifiquement sur l'application de méthodologies avancées d'intelligence artificielle pour anticiper les résultats des principaux marchés publics, en prenant pour terrain d'investigation les processus d'appel d'offres de la commune de Mila. L'arsenal technologique mobilisé comprend des algorithmes éprouvés tels que Random Forest et Extra Trees, des approches de proximité comme KNN, des architectures d'apprentissage profond incluant les réseaux de neurones profonds (DNN) et les unités récurrentes à portes (GRU), ainsi que des techniques de pointe en traitement du langage naturel comme la génération augmentée par récupération (RAG).

La structuration de ce mémoire s'articule autour de quatre axes complémentaires qui progressent de l'analyse contextuelle vers l'implémentation pratique.

Introduction Général

Le premier chapitre établit le cadre organisationnel de la commune de Mila, offrant une analyse détaillée de sa structure administrative et du rôle stratégique du Département des marchés publics. Cette exploration du contexte opérationnel, incluant les unités administratives et les processus de gestion des appels d'offres, constitue le socle indispensable pour appréhender l'environnement pratique d'application des techniques d'intelligence artificielle.

Le deuxième chapitre développe les fondements théoriques et méthodologiques en présentant de manière approfondie les techniques d'intelligence artificielle déployées dans cette étude. L'analyse porte sur les paradigmes de l'apprentissage automatique, de l'apprentissage profond et du traitement du langage naturel, avec un focus particulier sur les algorithmes Random Forest, Extra Trees, KNN, DNN et GRU, ainsi que sur les méthodes avancées de NLP comme RAG. Cette section examine les principes théoriques, les avantages spécifiques et les applications optimales de chaque approche dans le contexte du traitement des données textuelles et catégorielles caractéristiques des marchés publics.

Le troisième chapitre constitue le cœur empirique de la recherche en présentant le cadre expérimental et les résultats de l'application des modèles d'IA pour la prédiction des résultats de marchés publics. Après avoir formalisé la problématique de prédiction des candidats gagnant, du nombre de candidat et les offres techniques de cahier de charge, nous détaillons la constitution d'un ensemble de données original collecté à partir des archives des marchés publics de la commune de Mila. Les méthodologies de prétraitement et de modélisation sont rigoureusement décrites, suivies d'une évaluation exhaustive des performances des modèles utilisant des métriques de précision et de perte, étayée par des analyses visuelles incluant tableaux, matrices de confusion et courbes d'évolution des performances.

Le quatrième chapitre finalise l'approche par la présentation de l'implémentation pratique, détaillant les outils technologiques utilisés dans le développement de l'application web. Cette section décrit l'architecture et les fonctionnalités de l'application conçue pour l'exécution des modèles prédictifs, en analysant chaque interface utilisateur et en mettant en évidence les capacités opérationnelles du système développé.

Chapitre 01 :
Présentation de
l'organisme du lieu de
stage et service des
marchés publics

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

1.1-Introduction

Au cours de notre stage à la commune de Mila, nous avons exploré la gestion des projets au sein du service des marchés publics. Cette expérience nous a permis de comprendre les défis spécifiques et les pratiques de gestion de marchés publics au sein de l'organisation.

Dans ce chapitre, nous présentons d'abord la commune de Mila, ses composants. Ensuite, nous décrivons le service de marché public et ses unités, enfin l'application de l'IA dans les marchés publics.

1.2-Présentation de la commune de Mila :

La commune de la wilaya de Mila est une unité administrative locale rendant des services essentiels visant à améliorer la qualité de vie des citoyens dans les zones qui lui sont attribuées. La commune de Mila se charge de la gestion des affaires locales et de la mise en œuvre de projets de développement répondant aux besoins des habitants des quartiers et des zones rurales ou à faible densité de population dans la commune de Mila. la **Figure 1.1** représente l'emplacement de la commune de la wilaya de Mila.



Figure 1.1: La commune de Mila.

La commune de la wilaya de Mila est située dans la rue Bounamous.

1.2.1-Les composantes de la commune de Mila :

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

Dans ce qui suit, nous allons présenter les composants de la commune de la wilaya de Mila : *Direction de l'organisation et des affaires sociales, culturelles et sportives, Direction de l'administration générale et des finances, Direction des services de l'état civil, Direction des services techniques et Direction de la propreté, de l'esthétique publique et de la protection de l'environnement.*

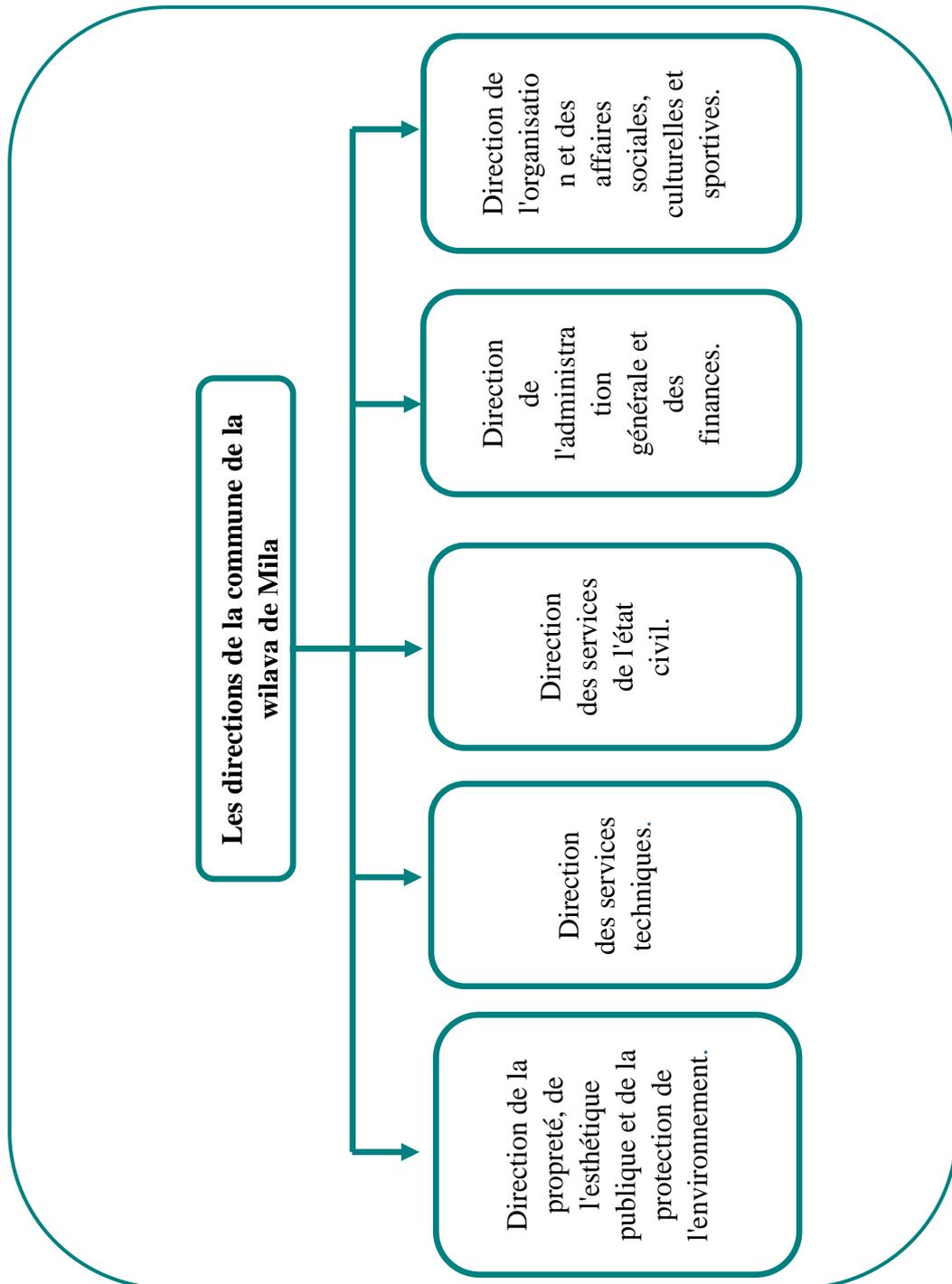


Figure 1.2: Organigramme des directions de la commune de la wilaya de Mila.

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

La **Figure 1.2** représente les différentes directions de la commune de la wilaya de Mila.

Notre étude porte sur les marchés publics relevant de la direction de l'administration générale et des finances, dont les services sont : *Le Service de gestion des ressources humaines, Le Service du budget et des comptes, Le Service des marchés publics et du suivi des programmes, composé de deux bureaux et le Service des biens municipaux.*

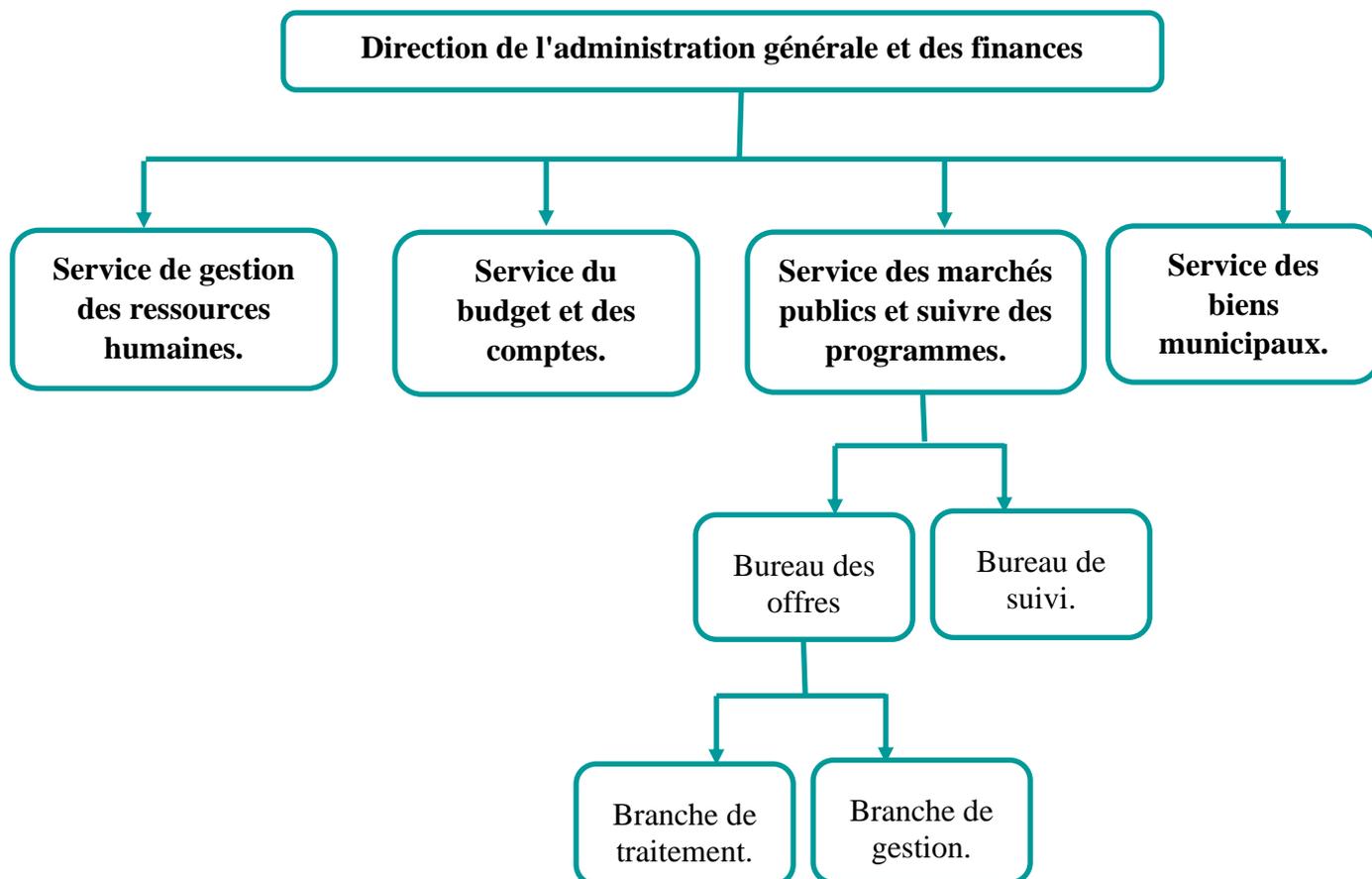


Figure 1.3: Organigramme de la direction de l'administration générale et des finances.

La **Figure 1.3** représente un organigramme de la Direction de l'administration générale et des finances.

1.3-Service des marchés publics :

Le service des marchés publics de la commune de Mila est responsable de l'organisation des marchés et des contrats publics pour la réalisation des projets au sein de la commune. Ce service s'occupe de résoudre les problèmes qui peuvent survenir lors du processus de passation des marchés, ainsi que de la sélection des prestataires, de la gestion et de l'exécution des contrats. Elle veille à garantir une utilisation transparente et efficace

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

des ressources publiques afin de fournir les meilleures infrastructures et services municipaux.

Le service des marchés publics et du suivi des programmes, dans une commune comme celle de Mila, est une structure clé qui gère les processus liés à la passation des marchés publics et au suivi de l'exécution des projets.

1.3.1-Les unités de service des marchés publics :

Dans cette partie, nous allons présenter *le Bureau des équipements* et *le Bureau de Suivi*.

1.3.1.1-Bureau des équipements :

Ce bureau est responsable de la gestion des processus liés aux offres soumises par les prestataires pour les marchés publics. Il est divisé en deux branches principales :

a. Branche de traitement :

Cette branche se charge principalement de l'enregistrement et de l'examen administratif des offres reçues.

Elle vérifie que les soumissions respectent les exigences des appels d'offres, notamment la conformité des documents requis, le respect des délais, ainsi que la validité des offres techniques et financières.

Elle prépare des rapports détaillés destinés au comité d'évaluation des offres.

b. Branche de gestion :

Cette branche se concentre sur la gestion opérationnelle des processus d'appel d'offres.

Elle assure la coordination avec les prestataires, communique les décisions et transmet les notifications officielles (par exemple : les avis d'attribution des marchés).

Elle veille à ce que les contrats soient correctement rédigés et signés conformément aux réglementations en vigueur.

1.3.1.2-Bureau de Suivi :

Ce bureau est responsable de la supervision et du contrôle de l'exécution des projets attribués dans le cadre des marchés publics.

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

Il surveille l'avancement des travaux pour s'assurer qu'ils respectent les clauses des contrats, les budgets alloués et les délais impartis.

En cas de problème (retards, dépassements de coûts, qualité non conforme), ce bureau agit en coordination avec les prestataires et trouve des solutions appropriées.

Élabore des rapports périodiques pour informer les autorités de l'état d'avancement des programmes.

Les marchés publics sont soumis à la loi *15-247*.

1.3.2- la loi 15-247 :

La loi *15-247* du 16 septembre 2015 régit les marchés publics et les délégations de service public en Algérie. Elle encadre les conditions dans lesquelles les administrations (comme les communes, ministères, établissements publics...) passent des contrats avec des entreprises pour : la réalisation de travaux, étude, la fourniture, ou la prestation de services. Cette loi vise à : garantir la transparence, assurer la concurrence loyale, veiller à une bonne gestion des fonds publics, et mettre en place une surveillance efficace.

Contient plusieurs étapes réglementées :

1. Lancement de l'appel d'offres.
2. Dépôt des offres.
3. Ouverture et évaluation des plis.
4. Classement et choix du soumissionnaire.
5. Attribution provisoire.
6. Attribution définitive.

1.4-Application de l'Intelligence Artificielle dans les Marchés Publics :

L'utilisation de l'intelligence artificielle dans le domaine des marchés publics se justifie par la complexité croissante des procédures, le volume important de données à traiter, ainsi que la nécessité de garantir l'objectivité, la rapidité et la transparence dans le processus décisionnel. Les méthodes traditionnelles, basées sur des analyses manuelles, montrent leurs limites face à l'augmentation du nombre d'appels d'offres, la diversité des critères techniques et la sensibilité des choix opérés.

L'IA offre ici des avantages significatifs :

Chapitre 01 : Présentation de l'organisme du lieu de stage et le service de marchés publics

- Elle permet une évaluation standardisée et impartiale des offres techniques et financières.
- Elle améliore la capacité de prédiction, en anticipant par exemple le nombre de soumissionnaires ou les chances de succès d'un candidat.
- Elle limite les erreurs humaines et renforce l'intégrité du processus.

En somme, l'intégration de l'IA dans ce contexte n'est pas seulement une option technologique, mais une nécessité stratégique pour répondre aux exigences de performance, de transparence et de gouvernance dans les achats publics.

1.5-Conclusion

Dans ce chapitre, nous avons introduit la commune de Mila. Nous avons ensuite décrit les différentes directions de la commune au sein de la wilaya. Par la suite, nous avons détaillé les services des marchés publics ainsi que leurs unités. Enfin, nous avons exploré le rôle de l'intelligence artificielle dans la gestion des marchés publics.

Chapitre 02 :

Techniques d'IA

Chapitre02: Techniques d'IA

2.1-Introduction

L'intelligence artificielle connaît une évolution spectaculaire, transformant profondément des secteurs variés tels que la santé, la finance et les services publics. Son impact croissant dans le domaine de la gestion des marchés publics ouvre la voie à de nouvelles méthodes d'analyse, de prédiction et d'automatisation, offrant des gains considérables en précision, en rapidité et en transparence.

Ce chapitre vise à présenter les principales techniques d'IA mobilisées dans notre projet, en commençant par l'apprentissage automatique, puis en explorant l'apprentissage profond et ses architectures spécifiques telles que les réseaux convolutifs ou récurrents. Nous aborderons également les techniques de traitement du langage naturel, particulièrement utiles pour l'analyse des documents textuels des marchés publics.

2.2- Intelligence artificielle :

L'intelligence artificielle (IA) est un processus d'imitation de l'intelligence humaine qui repose sur la création et l'application d'algorithmes exécutés dans un environnement informatique dynamique. Son but est de permettre à des ordinateurs de penser et d'agir comme des êtres humains [1].

2.3-Apprentissage automatique :

L'apprentissage automatique (Machine Learning « ML ») est un champ de l'IA porté sur l'analyse statistique de données d'apprentissage. Historiquement, cette branche est définie comme le développement de machines capables d'apprendre sans avoir été explicitement programmées à apprendre une tâche [2].

Exemple :

La **Figure 2.1** représente un exemple de chiffres manuscrits extraits de la base de données MNIST, couramment utilisée en apprentissage automatique.

Chapitre02: Techniques d'IA



Figure 2.1: Exemple des chiffres manuscrits [2].

L'exemple répandu d'apprentissage automatique est la reconnaissance automatique des chiffres manuscrits. Des exemples de ces chiffres sont présentés dans la **Figure 2.1**. Ces caractères présentent une grande diversité en forme, orientation, épaisseur de traits... Il est difficile dans ce cas d'écrire une liste explicite de règles permettant de discriminer les chiffres. On va donc chercher à mettre en œuvre un algorithme permettant d'extraire des règles implicites en se basant sur des données d'exemples. Ces règles peuvent par la suite être appliquées sur de nouveaux chiffres afin de les reconnaître [2].

Dans ce qui suit, nous présentons les types d'apprentissage automatique et ses phases.

2.3.1-Apprentissage supervisé :

C'est un type d'apprentissage automatique dans lequel un algorithme s'entraîne à une tâche en utilisant un jeu de données annotées. Chaque annotation indique le résultat attendu de l'algorithme. On considère que l'entraînement est réussi lorsque l'écart entre les prédictions de l'algorithme et les annotations est minimal [3].la **Figure 2.2** représente un ensemble de données étiquetées. Chaque image d'entrée est associée avec la prédiction voulue.)

Chapitre02: Techniques d'IA

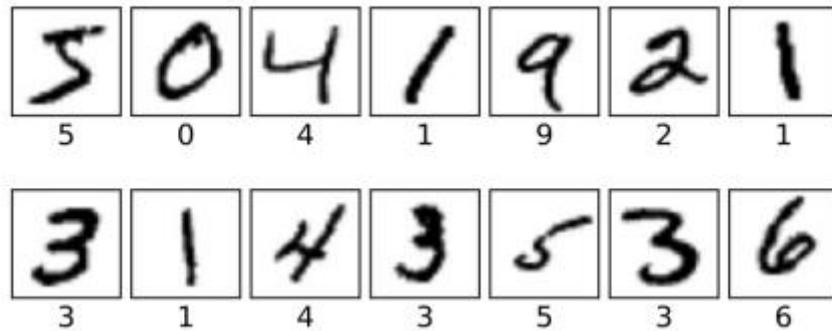


Figure 2.2: Ensemble de données étiquetées. Chaque image d'entrée est associée avec la prédiction voulue [2].

Parmi les algorithmes d'apprentissage supervisé les plus utilisés, on trouve *Random Forest*, *Extra Trees*, *les réseaux de neurones*, *Machines à vecteurs de support* et *K-Nearest Neighbors*.

•Random Forest :

C'est une méthode d'apprentissage automatique, reposant sur un ensemble d'arbres de décision (classification ou régression) qui, individuellement, sont simples mais peuvent manquer de précision sur des données complexes. Pour améliorer la performance, Random Forest génère plusieurs arbres à partir d'échantillons aléatoires de données et de variables, puis combine leurs résultats pour faire une prédiction finale. Cette approche permet une meilleure précision tout en conservant la lisibilité des relations entre variables et résultats. Elle est reconnue pour offrir une très bonne performance en classification, souvent supérieure à celle d'autres modèles [4]. la **Figure 2.3** représente une visualisation du modèle de Random Forest.

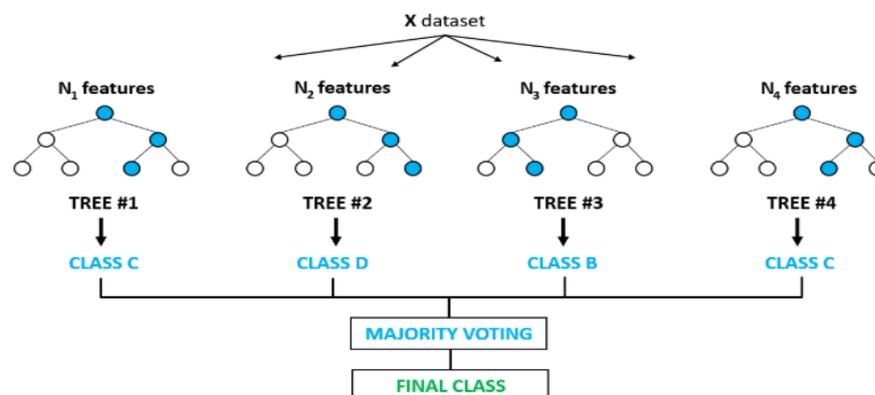


Figure 2.3: Visualisation du modèle de Random Forest [5].

Chapitre02: Techniques d'IA

•Extra Trees (arbres extrêmement randomisés) :

C'est un modèle d'apprentissage automatique par ensemble qui combine plusieurs arbres de décision, similaire à la forêt aléatoire (Random Forest), mais avec une randomisation supplémentaire.

Alors que les deux utilisent plusieurs arbres, Extra Trees génère les séparations de manière totalement aléatoire, sans chercher les seuils optimaux.

Au lieu d'utiliser un échantillonnage bootstrap, il utilise généralement l'ensemble complet du jeu de données pour chaque arbre, tout en maintenant la randomisation en sélectionnant des points de séparation aléatoires [6]. La **Figure 2.4** représente une architecture de l'Extra Tree Classifier.

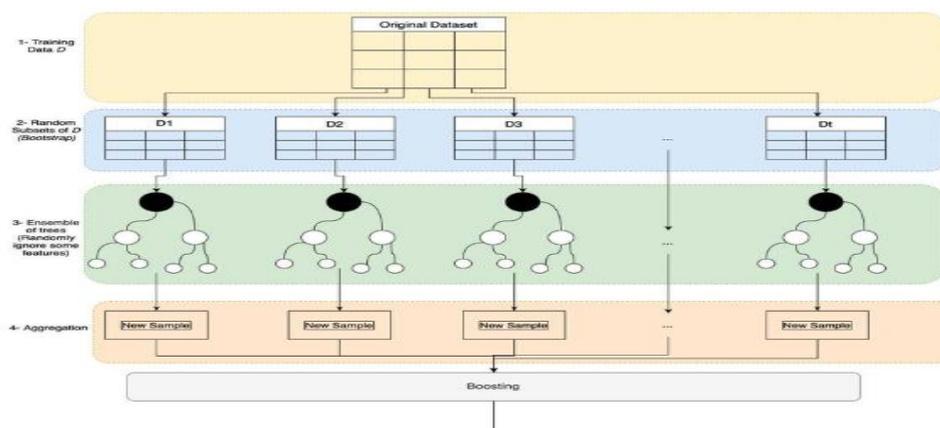


Figure 2.4: Architecture du Extra Tree Classifier [7].

•K-Voisins les plus proches (K-Nearest Neighbors (KNN)) :

L'algorithme des k plus proches voisins (KNN) est un classificateur d'apprentissage supervisé non paramétrique qui utilise la proximité pour effectuer des classifications ou des prédictions sur le regroupement d'un point de données individuel. Il s'agit de l'un des classificateurs de classification et de régression les plus populaires et les plus simples utilisés aujourd'hui en apprentissage automatique. [8].la **Figure 2.5** Il apparaît une illustration du processus d'apprentissage K-NN (K plus proches voisins) avec classification binaire.

Chapitre02: Techniques d'IA

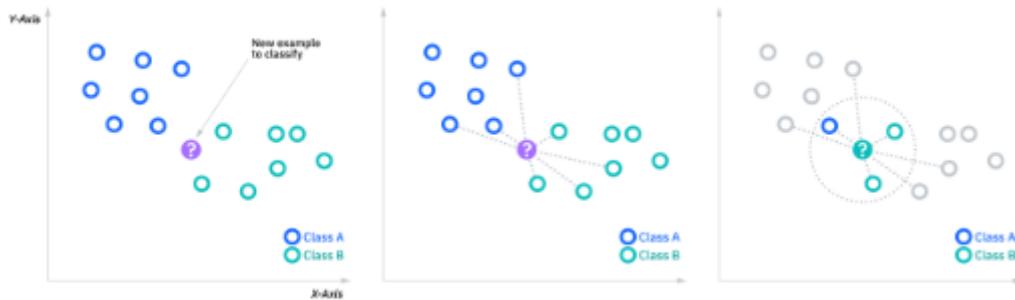


Figure 2.5: Illustration du processus d'apprentissage K-NN (K plus proches voisins) avec classification binaire [8].

2.3.2-Apprentissage non supervisé :

La principale différence entre cette catégorie d'apprentissage et la précédente réside dans l'absence de données annotées. Ainsi, les algorithmes de cette famille cherchent à comprendre les données et à proposer une structuration [9].la **Figure 2.6** représente un ensemble de données non étiquetées, les images d'entrées.



Figure 2.6: Ensemble de données non-étiquetées les images d'entrées [2].

2.3.3-Apprentissage par renforcement :

Il s'agit d'une autre famille de méthodes, peu exploitée, mais qui commence à attirer l'attention des chercheurs. Cette approche simule la capacité humaine à apprendre à partir des expériences et des erreurs, en renforçant les actions qui permettent d'atteindre un objectif donné [2].

2.3.4-Les phases de l'apprentissage automatique :

De manière générale, les algorithmes d'apprentissage automatique se séparent en deux phases.

Chapitre02: Techniques d'IA

- **Phase d'entraînement (ou d'apprentissage)** : Le modèle choisi est soumis à un grand nombre d'exemples significatifs. Le système cherche alors à apprendre des règles implicites en se basant sur ces données (appelées données d'entraînement). Cette phase d'entraînement précède généralement l'utilisation du modèle, bien que certains systèmes continuent d'apprendre indéfiniment s'ils disposent d'un retour sur les résultats (on appelle cela de l'apprentissage en ligne).

- **Phase d'inférence** : Le modèle entraîné peut être utilisé sur de nouvelles entrées. Les entrées fournies lors de la phase d'inférence peuvent être traitées même si elles n'ont pas été vues par le modèle lors de la phase d'apprentissage. En effet, grâce à l'extraction de règles implicites, le modèle peut se généraliser à des entrées inconnues [2].

2.4-Apprentissage profond (Deep Learning - DL) :

L'apprentissage profond est une branche de l'apprentissage automatique qui repose sur l'utilisation de réseaux de neurones artificiels comportant plusieurs couches cachées. Ces architectures permettent aux machines d'apprendre des représentations hiérarchiques des données, facilitant ainsi la résolution de tâches complexes telles que la reconnaissance d'images, la traduction automatique. [10].

2.5-Principe et fonctionnement du Deep Learning :

La **Figure 2.7** offre un aperçu standard, compréhensible et simple de l'architecture globale du Deep Learning. Comme mentionné auparavant, le Deep Learning est essentiellement un réseau de neurones. Cette figure montre qu'il comprend trois types de couches de neurones (nœuds) : une couche d'entrée (verte), des couches cachées (bleues) et une couche de sortie (orange). Les couches adjacentes sont reliées par des connexions pondérées (lignes noires) [11].

Chapitre02: Techniques d'IA

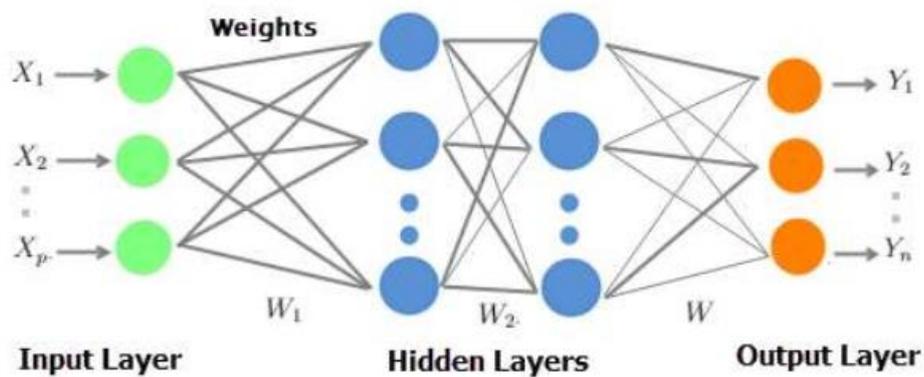


Figure 2.7: Architecture d'un modèle Deep Learning [11].

2.6-Méthodes du Deep Learning :

Dans ce qui suit, nous présentons les modèles suivants : *DNN*, *CNN*, *RNN* et *GRU*.

2.6.1-Réseau de neurones profonds (Deep Neural Network « DNN ») :

L'objectif de cette technologie est de simuler l'activité du cerveau humain, et plus particulièrement la reconnaissance de schémas et la transmission d'informations entre les différentes couches de connexions neuronales. Un réseau neuronal profond présente une particularité, il est composé d'au moins deux couches. Cela lui permet de traiter les données de manière complexe, à l'aide de modèles mathématiques avancés [12].la **Figure 2.8** représente l'architecture d'un réseau de neurones profond.

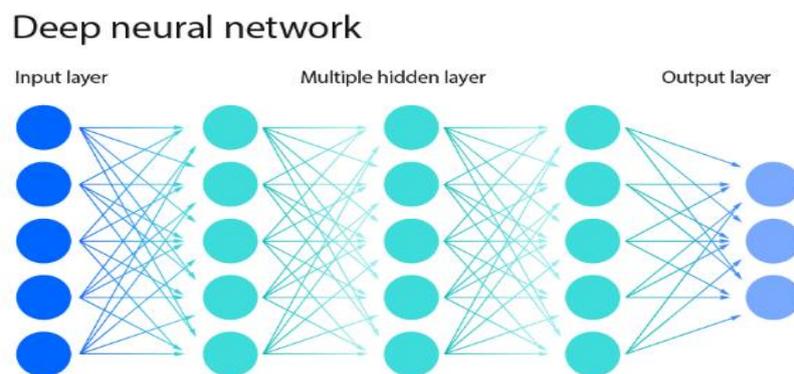


Figure 2.8: Architecture d'un réseau de neurones profond [12].

2.6.2-Réseau de neurones à convolution (Convolutional Neural Network « CNN ») :

Un réseau de neurones convolutifs (CNN, ou ConvNet) est une autre classe de réseaux de neurones profonds. On les retrouve le plus souvent en vision par ordinateur. À partir d'une série d'images ou de vidéos du monde réel, grâce à un CNN, le système d'IA apprend à

Chapitre02: Techniques d'IA

extraire automatiquement des caractéristiques de ces entrées pour réaliser une tâche spécifique, par exemple la classification d'images, l'authentification faciale et la segmentation sémantique d'images [13].la **Figure 2.9** représente le concept de réseau de neurones à convolution.

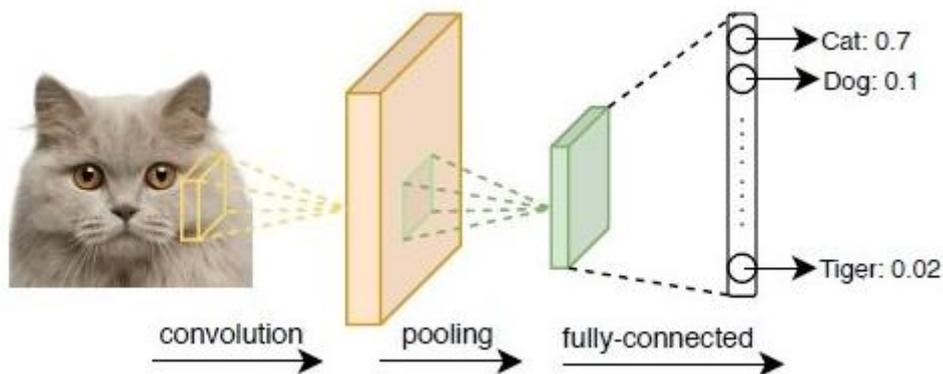


Figure 2.9: Concept de réseau de neurones à convolution [13].

2.6.3- Réseau de Neurones Récurrent (Recurrent Neural Network « RNN ») :

Les RNN ont été développés pour résoudre le problème des séries temporelles de données d'entrée séquentielles.

L'entrée d'un RNN est constituée de l'entrée courante et des échantillons précédents. Par conséquent, les connexions entre les nœuds forment un graphe orienté le long d'une séquence temporelle. De plus, chaque neurone d'un RNN possède une mémoire interne qui conserve les informations de calcul des échantillons précédents [13].la **Figure 2.10** représente le concept de réseau de neurones récurrent.

Chapitre02: Techniques d'IA

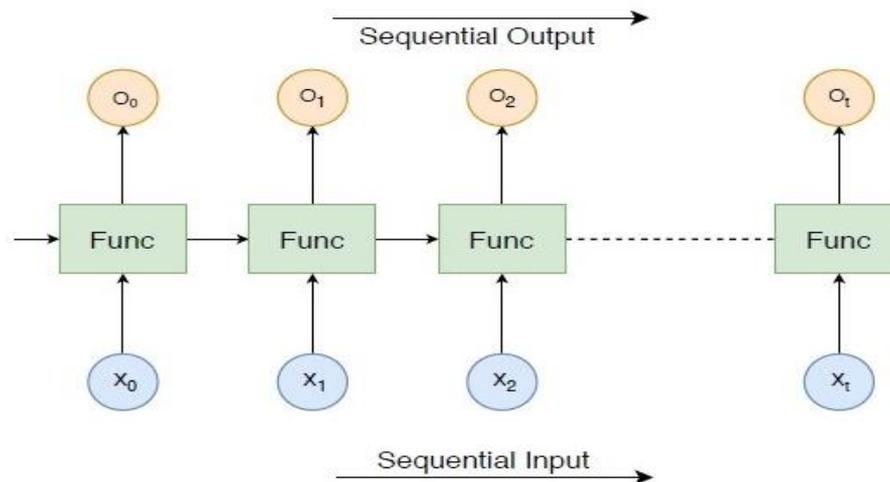


Figure 2.10: Concept de réseau de neurones récurrent [13].

2.6.4- L'unité récurrente à portes (Gated Recurrent Unit « GRU ») :

GRU est un type de réseau neuronal récurrent (RNN) qui, dans certains cas, présente des avantages par rapport à la mémoire à long terme (LSTM). La GRU utilise moins de mémoire et est plus rapide que la LSTM. Cependant, cette dernière est plus précise lors de l'utilisation d'ensembles de données contenant des séquences plus longues.

Si une couche d'un réseau neuronal ne peut pas apprendre, les RNN peuvent « oublier » les séquences plus longues.

Les GRU résolvent ce problème grâce à deux portes : la porte de mise à jour et la porte de réinitialisation. Ces portes déterminent les informations autorisées à passer en sortie et peuvent être entraînées à conserver les informations plus anciennes. Cela leur permet de transmettre des informations pertinentes tout au long d'une chaîne d'événements afin de réaliser de meilleures prédictions [14]. La **Figure 2.11** représente l'architecture de l'unité récurrente fermée de base (GRU).

Chapitre02: Techniques d'IA

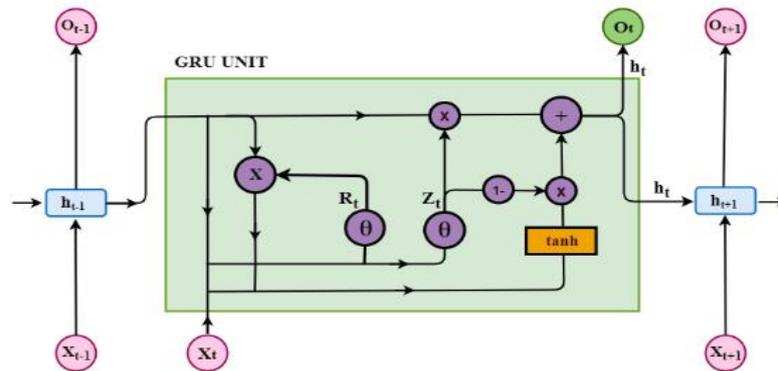


Figure 2.11: L'architecture de l'unité récurrente fermée de base (GRU) [15].

2.7-Traitement automatique du langage naturel (NLP) :

Le traitement automatique du langage naturel (NLP) est un sous-domaine de l'informatique et de l'intelligence artificielle (IA) qui utilise la machine learning pour permettre aux ordinateurs de comprendre et de communiquer en langage humain.

Le NLP permet aux ordinateurs et aux appareils numériques de reconnaître, comprendre et générer du texte et de la parole en combinant la linguistique computationnelle (la modélisation du langage humain basée sur des règles) avec la modélisation statistique, le machine learning et l'apprentissage profond [16].

Ces dernières années, des architectures avancées telles que les Transformers ont révolutionné le domaine du NLP en permettant une compréhension plus fine du langage, notamment grâce à des modèles comme BERT, qui utilisent une approche bidirectionnelle pour capter le contexte complet des phrases.

Dans ce qui suit, nous présentons *Transformer*, *BERT*, *RAG*, *DPR* et *FID*.

2.7.1-Transformer :

Les transformateurs ont été initialement développés pour résoudre le problème de la transduction de séquences, ou traduction automatique neuronale. Ils sont donc destinés à résoudre toute tâche transformant une séquence d'entrée en séquence de sortie. C'est pourquoi on les appelle « Transformateurs » [17].

2.7.2-BERT (Bidirectional Encoder Representations from Transformers) :

C'est un modèle de traitement du langage naturel développé par Google en 2018. Il utilise une architecture de transformer pour comprendre le contexte des mots dans une phrase en considérant les mots qui les précèdent et les suivent, d'où le terme "bidirectionnel".

Chapitre02: Techniques d'IA

BERT est principalement utilisé pour des tâches de compréhension du langage telles que la classification de texte, la réponse aux questions et l'analyse de sentiment. Sa capacité à capturer le contexte bidirectionnel le rend particulièrement efficace pour comprendre les nuances et les subtilités du langage [18].

2.7.3- génération augmentée de récupération (retrieval augmented generation « RAG »):

Le modèle RAG est une technique de traitement du langage naturel (NLP) qui combine les forces des modèles d'intelligence artificielle (IA) basés sur la récupération et sur la génération. RAG peut fournir des résultats précis qui tirent le meilleur parti des connaissances préexistantes. Elle peut également traiter et consolider ces connaissances pour créer des réponses, des instructions ou des explications uniques, adaptées au contexte et rédigées dans un langage proche de celui des humains, au lieu de se contenter de résumer les données récupérées [19]. La **Figure 2.12** représente le fonctionnement d'un Modèle RAG.

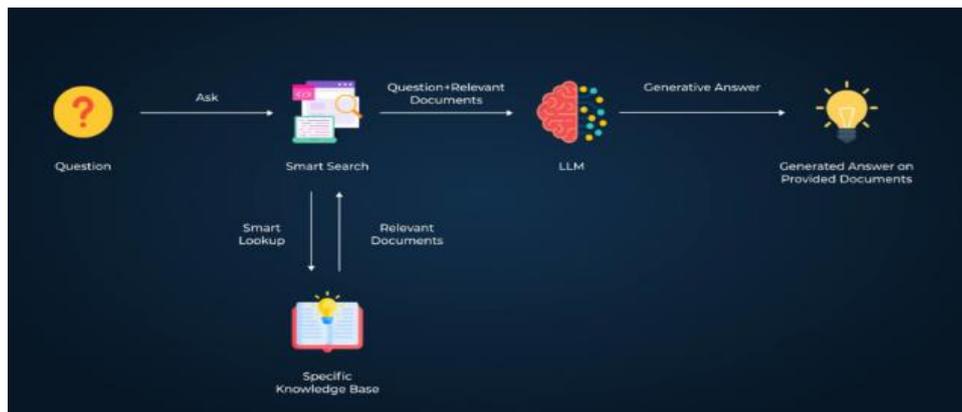


Figure 2.12: Fonctionnement d'un Modèle RAG [20].

2.7.4 - Fusion dans le décodeur (Fusion-in-Decoder (FiD)) :

FiD est un puissant modèle de langage à récupération augmentée, à la pointe de la technologie pour de nombreuses tâches de traitement du langage naturel (TALN) gourmandes en connaissances [21]. La **Figure 2.13** représente le modèle Fusion-In-Decoder (FiD).

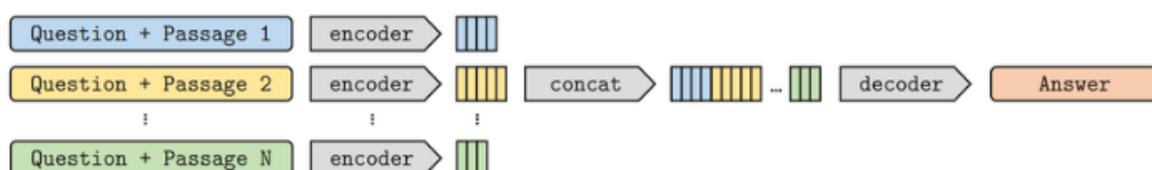


Figure 2.13: Le modèle Fusion-In-Decoder (FiD) [22].

Chapitre02: Techniques d'IA

2.8-Techniques de prétraitement :

L'étape de prétraitement du texte consiste simplement à amener le document dans un format facilement compréhensible, prévisible et analysable par la machine grâce aux différents algorithmes d'apprentissage automatique. Certaines des techniques de prétraitement les plus utilisées sont : *La tokenisation*, *La lemmatisation*, *TF-IDF*, *L'encodage d'étiquettes « Label Encoding »* et *Transformateur de phrases (Sentence Transformer)* et *Modèle de Flan T5 Base*.

2.8.1- La tokenisation :

Consiste à décomposer les phrases en unités individuelles telles que des mots, des caractères et des signes de ponctuation, chacun étant appelé un jeton. Le processus de division se produit généralement au niveau des espaces ou des signes de ponctuation. Cette étape initiale permet de supprimer les mots inutiles lors des étapes de traitement ultérieures [23].

2.8.2- La lemmatisation :

Consiste à supprimer ou à remplacer le suffixe d'un mot pour le ramener à sa forme de base, appelée lemme. Contrairement à la lemmatisation, le lemme obtenu est toujours un mot significatif. La lemmatisation est une étape de prétraitement de texte largement utilisée en traitement automatique du langage naturel (TALN) et s'est avérée très efficace [23].

2.8.3- TF-IDF :

La méthode TF-IDF est utilisée pour réduire l'influence des mots fréquents mais moins significatifs dans le contexte d'un document. Elle se compose de deux éléments : la fréquence des termes (TF) et la fréquence inverse des documents (IDF). La TF mesure la fréquence d'un mot dans un document et la normalise par sa longueur, garantissant ainsi une précision satisfaisante pour les ensembles de données de petite et de grande taille. Cette technique de normalisation divise la fréquence par la longueur du document. De son côté, l'IDF détermine l'importance d'un mot dans le document en réduisant les mots moins importants. Pour ce faire, elle calcule le logarithme du rapport entre le nombre total de documents et le nombre total d'occurrences du mot dans ce document [23].

$$tfidf_{i,j} = tf_{i,j} \log \left(\frac{N}{df_i} \right)$$

Chapitre02: Techniques d'IA

Où :

$tf_{i,j}$ = nombre total d'occurrences de i dans j .

df_i = nombre total de documents (discours) contenant i .

N = nombre total de documents (discours).

2.8.4- L'encodage d'étiquettes « Label Encoding » :

L'encodage d'étiquettes est une technique utilisée en apprentissage automatique et en analyse de données pour convertir des variables catégorielles en données numériques. Il est particulièrement utile pour les algorithmes nécessitant des données numériques, car la plupart des modèles d'apprentissage automatique ne peuvent fonctionner qu'avec des données numériques [24].

2.8.5- Transformateur de phrases (Sentence Transformer) :

C'est une classe de la bibliothèque Python sentence-transformers, conçue pour générer des représentations vectorielles (embeddings) de phrases ou de paragraphes. Ces embeddings permettent de comparer des textes en termes de similarité sémantique, facilitant des tâches telles que la recherche sémantique, le regroupement de documents ou la recommandation de contenu [25].

Nous allons présenter les trois modèles de transformateur de phrase : *Modèle distiluse-base-multilingual-cased-v1*, *Modèle paraphrase-multilingual-mpnet-base-v2* et *Modèle paraphrase-MiniLM-L6-v2*.

2.8.5.1-Modèle Distiluse-base-multilingual-cased-v1 :

Version simplifiée de l'encodeur de phrases universel multilingue. Prise en charge de 15 langues [26].

Il mappe les phrases et les paragraphes dans un espace vectoriel dense de 512 dimensions et peut être utilisé pour des tâches telles que le clustering ou la recherche sémantique [27].

2.8.5.2-Modèle Paraphrase-multilingual-mpnet-base-v2 :

Version multilingue de paraphrase-mpnet-base-v2, formée sur des données parallèles pour plus de 50 langues [26].

Chapitre02: Techniques d'IA

Il mappe les phrases et les paragraphes dans un espace vectoriel dense de 768 dimensions et peut être utilisé pour des tâches telles que le clustering ou la recherche sémantique [28].

2.8.5.3-Modèle paraphrase-MiniLM-L6-v2 :

C'est un outil puissant pour les tâches de traitement du langage naturel. Il s'agit d'un type de modèle *sentence-transformers*, ce qui signifie qu'il peut transformer des phrases et des paragraphes en un type particulier de vecteur compréhensible par les ordinateurs [29].

2.8.6-Modèle de Flan T5 Base :

Il s'agit d'un modèle de langage de pointe développé par Google, affiné sur plus de 1000 tâches supplémentaires couvrant plusieurs langues. Il offre d'excellentes performances en *few-shot learning*, même comparé à des modèles beaucoup plus grands, et se montre particulièrement efficace dans des tâches telles que le raisonnement et les questions-réponses. Grâce à son architecture de type *transformer* et sa prise en charge de plusieurs langues, il est bien adapté à la recherche sur les modèles de langage ainsi qu'à diverses tâches de traitement du langage naturel [30].

2.9- Indicateurs de performance :

Après avoir développé des modèles d'apprentissage profond, l'étape suivante consiste à évaluer leurs performances prédictives. Cela implique de diviser les données en ensembles de formation, de validation et de test pour calculer les métriques pertinentes. Ces mesures comprennent : *Précision* « *Accuracy* » et *matrice de confusion*.

2.9.1- Précision « Accuracy » :

Cette mesure quantifie le pourcentage de prédictions correctes effectuées par le modèle [23].

2.9.2-matrice de confusion :

La matrice de confusion (ou matrice d'erreur) est une méthode qui permet de visualiser les résultats d'un algorithme de classification. Plus précisément, il s'agit d'une table qui décompose le nombre d'instances de vérité terrain d'une classe donnée par rapport au nombre d'instances de classes prédites. La matrice de confusion est l'un des nombreux indicateurs qui permettent d'évaluer la performance des modèles de classification. On peut

Chapitre02: Techniques d'IA

s'en servir pour calculer un certain nombre d'autres indicateurs de performance des modèles, tels que la précision et le rappel [31].

2.10-Conclusion

Dans ce chapitre, nous avons exposé les principales techniques d'intelligence artificielle qui sous-tendent notre démarche méthodologique. De l'apprentissage automatique classique à l'apprentissage profond, en passant par les outils de traitement du langage naturel, nous avons également mis en lumière les algorithmes et modèles les plus pertinents, tels que les réseaux de neurones, les modèles transformés et les techniques d'extraction de caractéristiques. Cette base théorique constitue une étape indispensable avant de passer à l'implémentation pratique et aux expérimentations détaillées dans le prochain chapitre.

Chapitre 03 :
Validation et évaluation
des performances des
modèles proposés

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.1-Introduction

L'évolution rapide des technologies d'intelligence artificielle a transformé la gestion des données et des processus dans de nombreux secteurs, y compris celui des marchés publics. Dans le cadre de notre projet, l'exploitation de ces technologies permet d'optimiser la prévision des résultats des appels d'offres, renforçant ainsi la transparence et l'efficacité des processus administratifs.

Ce chapitre se consacre à l'évaluation des performances des modèles proposés pour la prédiction des projets de marchés publics. Nous commençons par présenter les outils et logiciels utilisés, avant de détailler notre méthodologie, incluant l'objectif de travail, le dataset exploité, les approches testées, ainsi que le processus d'entraînement des modèles. Enfin, nous analysons les expériences réalisées et les résultats obtenus.

3.2- Logiciels et outils d'apprentissage profond :

Nous avons exploité une variété d'outils logiciels afin de mener à bien notre projet. Ces divers outils nous ont facilité la création de l'ensemble des diagrammes présentés dans ce document et l'implémentation des différents programmes informatiques.

•Python :

C'est un langage de programmation de haut niveau reconnu pour sa lisibilité et sa simplicité. Il utilise l'indentation pour définir les blocs de code, améliorant ainsi la lisibilité et facilitant la maintenance. Compatible avec la programmation procédurale, orientée objet et fonctionnelle, Python est polyvalent pour le développement web, le calcul scientifique et l'IA.

Nous avons choisi Python en raison de sa large adoption dans le domaine de l'apprentissage automatique. Selon le classement TIOBE4 « **Figure 3.1** ».

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

May 2025	May 2024	Change	Programming Language	Rating	Change
1	1		Python	25.35%	+9.02%
2	3	▲	C++	9.94%	+0.41%
3	2	▼	C	9.71%	-0.27%
4	4		Java	9.31%	+0.62%
5	5		C#	4.22%	-2.27%
6	6		JavaScript	3.68%	+0.66%
7	8	▲	Go	2.70%	+1.10%
8	7	▼	Visual Basic	2.62%	+0.61%
9	11	▲	Delphi/Object Pascal	2.29%	+1.05%
10	9	▼	SQL	1.90%	+0.45%

Figure 3.1: Classement TIOBE du May 2025 des langages de programmation [32].

•Google Colaboratory :

Google Colaboratory Colab est une plateforme cloud utilisant la technologie Jupyter Notebook pour la recherche en apprentissage automatique et en apprentissage profond. Elle s'intègre à Google Drive pour faciliter le partage et la collaboration. Colab fournit un GPU puissant et gratuit pour les tâches intensives, notamment avec TensorFlow et Pytorch, bien que limité en termes de temps de calcul et de temps d'exécution. Malgré cela, Colab est un outil précieux et sans maintenance pour l'enseignement et les études préliminaires en IA [33].

3.3-Les Objectifs de travail :

À ce jour, peu de travaux se sont penchés sur l'utilisation de l'intelligence artificielle pour la prédiction appliquée aux projets de marchés publics, notamment en ce qui concerne l'estimation des offres et l'analyse des cahiers des charges. Cette étude s'inscrit donc dans une perspective innovante, en développant deux modèles prédictifs complémentaires.

Le premier modèle vise à estimer le nombre de candidats ainsi que le candidat gagnant pour chaque appel d'offres. L'objectif est d'optimiser l'évaluation en amont, d'améliorer la planification, de réduire les délais de traitement et de garantir une meilleure adéquation entre les offres reçues et les soumissionnaires. Ce modèle offre également aux entreprises la possibilité d'anticiper plus efficacement les opportunités de marché avant de s'y engager.

Le second modèle se concentre sur la prédiction des éléments du cahier des charges, notamment les conditions d'éligibilité, les moyens humains et matériels requis, ainsi que les offres de prestation attendues. En facilitant la préparation de propositions conformes aux exigences techniques, ce modèle contribue à améliorer la qualité et la pertinence des offres

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

soumises. L'approche adoptée, basée sur des techniques d'apprentissage profond, permet d'affiner les prédictions, de renforcer la prise de décision et de réduire significativement le temps et les efforts nécessaires à la préparation des offres.

3.4-Dataset : « dataset 1»

L'ensemble de données a été collecté manuellement au moyen de recherches sur le terrain, de formations pratiques et d'observations directes au sein des services administratifs de la wilaya de Mila. Contrairement aux ensembles de données publics ou en ligne, les informations de cet ensemble de données ont été collectées par nos soins, à travers l'utilisation de documents administratifs (formulaire d'appel d'offres, formulaires, etc.) et d'entretiens que nous avons menés lors de nos stages dans des institutions locales (par exemple, communes, centre universitaire de Mila...).

Cet ensemble de données rassemble diverses informations issues de projets dans divers domaines. Chaque ligne représente un projet donné.

Nous avons organisé ces informations dans un fichier Excel en structurant les colonnes pour représenter différents aspects des marchés publics. Ce travail nous a permis de disposer d'une base de données personnalisée et adaptée à notre étude de cas.

Nous avons donc collecté 98 exemples de projets différents dans de nombreux domaines, La **Figure 3.2** représente le dataset utilisé.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	de public	Description	Localite	leur d'offr	anels (jour	service contractant	Type de projet	taire de chdidat	gagr	de candit	ion d'elighique	(moinique	(mçon	(realisaine	Projet	
2	#####	ecole : ziani said constr	Mila	14500501	14800400	90 jours	Municipalite de mila	Realisation des tr	moins dis:gerich ane		2	calificator	1	betonnik	1	ingenieur en genie Education
3	#####	ecole : shabi mouhamm	Mila	15300201	15400010	90 jours	Municipalite de mila	Realisation des tr	moins dis:lgrani abd		2	calificator	1	betonnik	1	ingenieur en genie Education
4	#####	ecole : sdrati said constr	Mila	17113401	17800081	90 jours	Municipalite de mila	Realisation des tr	moins dis:bida tarek		3	calificator	1	betonnik	1	ingenieur en genie Education
5	#####	ecole : birch mekhtar cc	Mila	17533300	17550050	90 jours	Municipalite de mila	Realisation des tr	moins dis:hamouda		2	calificator	1	betonnik	1	ingenieur en genie Education
6	#####	ecole : ahmed hemla cor	Mila	13200500	13400000	90 jours	Municipalite de mila	Realisation des tr	moins dis:benabdla		3	calificator	1	betonnik	1	ingenieur en genie Education
7	#####	ecole : boutebdja aissa	Mila	18500040	18900801	90 jours	Municipalite de mila	Realisation des tr	moins dis:latamna cl		2	calificator	1	betonnik	1	ingenieur en genie Education
8	#####	ecole : boukhareba mou	Mila	15550000	16000001	90 jours	Municipalite de mila	Realisation des tr	moins dis:zouaghi al		2	calificator	1	betonnik	1	ingenieur en genie Education
9	#####	ecole : ben dakir salah c	Mila	13200200	13900901	90 jours	Municipalite de mila	Realisation des tr	moins dis:nassim be		2	calificator	1	betonnik	1	ingenieur en genie Education
10	#####	ecole : rabeh lahmer cor	Mila	14001001	14500010	90 jours	Municipalite de mila	Realisation des tr	moins dis:anes haro		4	calificator	1	betonnik	1	ingenieur en genie Education
11	#####	ecole: ben al-chiheb abd	Mila	14110000	14122000	90 jours	Municipalite de mila	Realisation des tr	moins dis:gerich ane		2	calificator	1	betonnik	1	ingenieur en genie Education
12	#####	ecole: sidani ali construi	Mila	15444000	15900001	90 jours	Municipalite de mila	Realisation des tr	moins dis:mesiaad r		2	calificator	1	betonnik	1	ingenieur en genie Education
13	#####	ecole: sidi abd elah cor	Mila	16500000	16600000	90 jours	Municipalite de mila	Realisation des tr	moins dis:mesiaad r		2	calificator	1	betonnik	1	ingenieur en genie Education

Figure 3.2: Dataset utilise.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Voici un exemple :

-Date de publication : 19/03/2024.

-Description : Finalisation du réseau d'assainissement de Machta Al Bir afin d'améliorer les conditions sanitaires et de vie des habitants en garantissant une gestion efficace des eaux usées.

-Localité : Grarem Gouga.

-Valeur de l'offre : 22 204 390 DA.

-AE (Autorisation d'engagement) : 22 304 440,4 DA.

-Délais d'exécution : 60 jours.

-Service contractant : Municipalité de Grarem Gouga.

-Type de projet : Réalisation des travaux.

-Critère de choix : Moins-disant.

-Candidat gagnant : Mouhammed Hakim.

-Nombre de candidats : 1.

-Conditions d'éligibilité : Expérience en gestion des eaux usées, Certification en génie sanitaire.

-Critères techniques – Moyens matériels : 1 Canalisation et raccords, 2 Stations d'épuration, 3 Matériel de curage.

-Critères techniques – Moyens humains : 1 Ingénieur en assainissement, 2 Techniciens en traitement des eaux, 3 Ouvriers spécialisés.

-Offre de prestation (réalisation des études) : (non précisé).

-Domaine du projet : Hydraulique.

Notre ensemble de données contient 16 colonnes détaillées ci-dessous :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Tableau 3.1: La liste des colonnes de dataset.

Nom de colonne	Description
Date de publication	Date de lancement officiel de l'appel d'offres.
Description	Nature du projet (ex : construction d'un restaurant scolaire).
Localite	Lieu de réalisation (commune, daïra, wilaya).
valeur d'offre	Montant proposé par le candidat sélectionné.
AE (autorisation d'engagement)	Budget prévisionnel autorisé pour la réalisation du projet.
delais (jours)	Durée estimée des travaux.
service contractant	Organisme public responsable de l'appel d'offre.
Type de projet	Catégorie du projet (ex : réalisation de travaux).
créataire de choix	Critère principal de sélection (ex : "moins disant").
Candidat gagnant	Nom de l'entreprise retenue à l'issue de la procédure.
Nbr de candidat	Nombre de candidats ayant soumissionné.
condition d'éligibilite	Conditions imposées aux soumissionnaires (qualification, chiffre d'affaires...).
Criterer Technique (moyen matériel)	Moyens matériels requis pour le projet (bétonnière, camions, etc.).
Criterer Technique (moyen humain)	Moyens humains exigés (ingénieur, chef de chantier, etc.).
offre de prestation (réalisation des études)	Informations sur les études préalables (souvent non remplies).
Domaine_Projet	Extraire à partir de la description.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.5-Préparation de dataset :

N'importe quel modèle d'apprentissage automatique nécessite une phase préalable de prétraitement. C'est une phase primordiale, car elle a une influence sur les résultats du modèle.

Nous allons présenter les différentes étapes de la préparation de dataset : *Nettoyage et réorganisation de dataset, Augmentation de dataset et Encodage.*

3.5.1-Nettoyage et réorganisation de dataset :

•L'objectif du nettoyage est d'optimiser les champs textuels (notamment la Description) pour qu'ils soient exploitables par des modèles de machine learning, tout en conservant les informations essentielles. Voici les opérations réalisées :

-Nettoyage de texte brut (colonne "Description").

-Suppression des caractères spéciaux et ponctuations inutiles.

-Mise en minuscules de tout le texte pour uniformiser.

-Lemmatization : les mots sont ramenés à leur forme de base (ex : "construire", pas "construira").

-Suppression des stopwords (mots courants comme "le", "de", "et" qui n'apportent pas de valeur informative).

•La réorganisation vise à rendre la structure du fichier plus claire, plus logique et directement exploitable par les modèles d'apprentissage automatique.

3.5.2-Augmentation de dataset :

L'augmentation de dataset dans ce contexte consiste à enrichir artificiellement les données disponibles en générant des variations légères des descriptions de projets. Plus précisément, cela se fait en réorganisant partiellement les mots au début de chaque description, ce qui introduit une diversité syntaxique tout en conservant le sens général du texte. Cette méthode simple permet de créer plusieurs versions plausibles des descriptions originales, contribuant ainsi à améliorer la robustesse du modèle prédictif. Elle aide le modèle à mieux apprendre les différentes structures possibles des phrases, ce qui est particulièrement utile lorsque le jeu de données initial est limité.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Grâce à cette stratégie, le nombre de descriptions exploitables est passé de 98 à 1568, offrant ainsi un corpus d'entraînement plus riche et mieux adapté aux exigences de l'apprentissage profond.

3.5.3-Encodage :

L'encodage est une étape cruciale pour transformer les données textuelles et catégorielles en formats numériques exploitables par les modèles de machine learning. Dans ce projet, plusieurs techniques ont été utilisées pour capturer les informations de manière optimale. Les colonnes catégorielles, telles que « Service contractuel » ou « Type de projet », ont été encodés à l'aide de LabelEncoder afin de les convertir en entiers algorithmiquement compréhensibles. Pour la colonne textuelle « Description », des méthodes de fréquence simples, comme TF-IDF, et des méthodes avancées, comme Sentence Transformer, ont été utilisées pour créer des représentations vectorielles denses et contextuelles. Cette combinaison de codages permet d'intégrer efficacement la structure catégorielle et la signification sémantique du texte au processus d'apprentissage, améliorant ainsi les performances et la généralisation du modèle.

3.6-Les approches proposées :

Dans cette partie, nous allons présenter l'approche utilisée pour prédire le candidat gagnant et le nombre de candidats, puis celle permettant de prédire les conditions du cahier des charges.

3.6.1-Prédire le candidat gagnant et le nombre de candidats :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

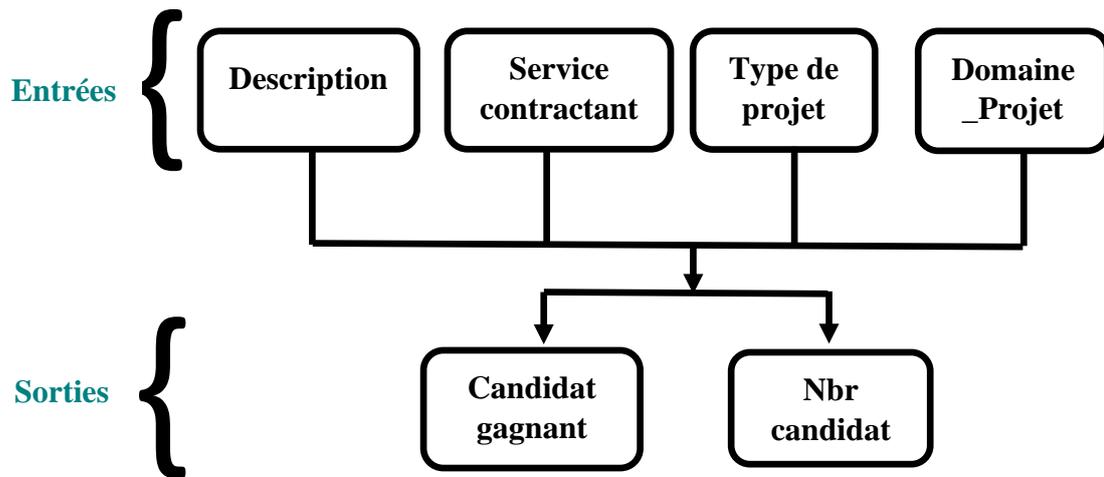


Figure 3.3: entrées / sorties pour la prédiction du candidat gagnant et du nombre de candidats.

La **Figure 3.3** présente l'architecture du modèle de prédiction du candidat gagnant et du nombre de candidats.

Dans ce qui suit, on va présenter les modèles basés sur *l'apprentissage automatique*, *Deep learning* et *NLP*.

3.6.1.1-Modèles basés sur l'apprentissage automatique :

Nous avons utilisé trois modèles de classification différents pour développer un système de classification multi-sorties capable de prédire des variables cibles à partir de données textuelles et catégoriques. Le premier modèle : *ExtraTreesClassifier*, le deuxième modèle : *RandomForestClassifier* et le troisième modèle : *KNeighborsClassifier*. La **Figure 3.4** représente l'architecture du modèle basé sur l'apprentissage automatique de prédiction du candidat gagnant et du nombre de candidats.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

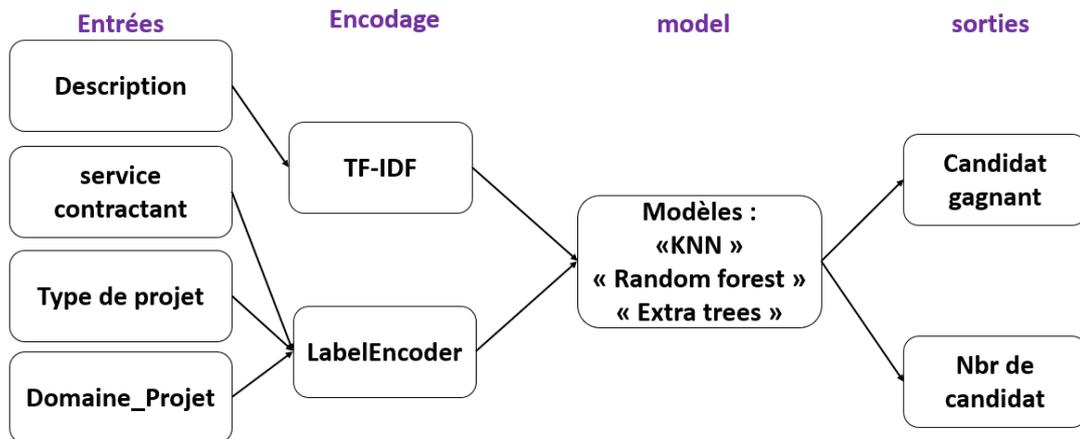


Figure 3.4: Architecture du modèle basé sur l'apprentissage automatique de prédiction du candidat gagnant et du nombre de candidats.

Le **Tableau 3.2** représente les modèles utilisés avec les différences entre eux :

Tableau 3.2: Les modèles utilisés basés sur l'apprentissage automatique avec leurs différences.

Modèle	Approche	Hyperparamètres	Particularité
RandomForest Classifier	Arbres avec divisions optimales.	n_estimators=250, max_depth=15, random_state=42	Plus précis, mais plus lent.
ExtraTreesClassifier	Arbres avec randomisation accrue.	n_estimators=250, max_depth=15, random_state=42	Rapide, moins précis sur données complexes.
KNeighbors Classifier	Basé sur les distances entre points.	n_neighbors=15, weights='distance'	Nécessite normalisation, sensible à k.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.6.1.1.1- Modèle RandomForestClassifier :

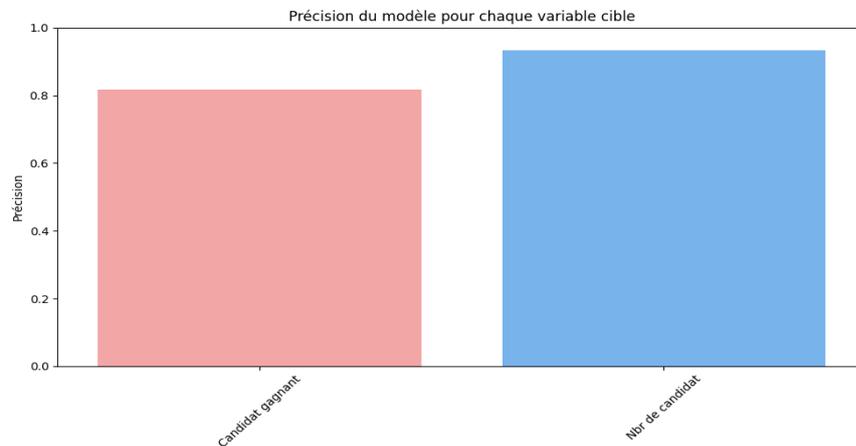


Figure 3.5: Précision par cible pour le modèle Random forest.

La **Figure 3.5** représente la précision du modèle *RandomForest* pour le nombre de candidats et les résultats du candidat gagnant. Nous constatons que la précision est d'environ 0.82 pour le candidat gagnant, tandis que le nombre de candidats est d'environ 0.93.

Cela indique que le modèle possède de solides capacités de classification et de prédiction numériques, ce qui en fait un choix fiable pour les applications de prévision de projets publics.

➤ **matrice de confusion :**

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

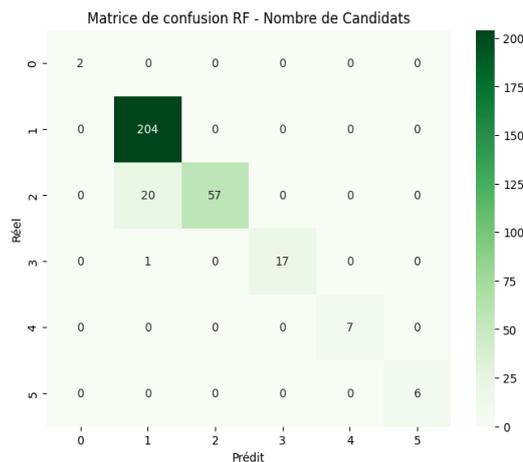


Figure 3.6: Matrices de confusion - nombre de candidats - pour le modèle RandomForestClassifier.

RandomForestClassifier affiche une précision globale élevée (93 %), avec une excellente prédiction de la classe dominante (1 candidat), atteignant 204 prédictions correctes. Cependant, des erreurs sont observées, notamment 20 cas où la classe réelle est 2 mais prédite comme 1, ce qui indique une certaine confusion entre ces deux classes voisines.

3.6.1.1.2-Modèle ExtraTreesClassifier :

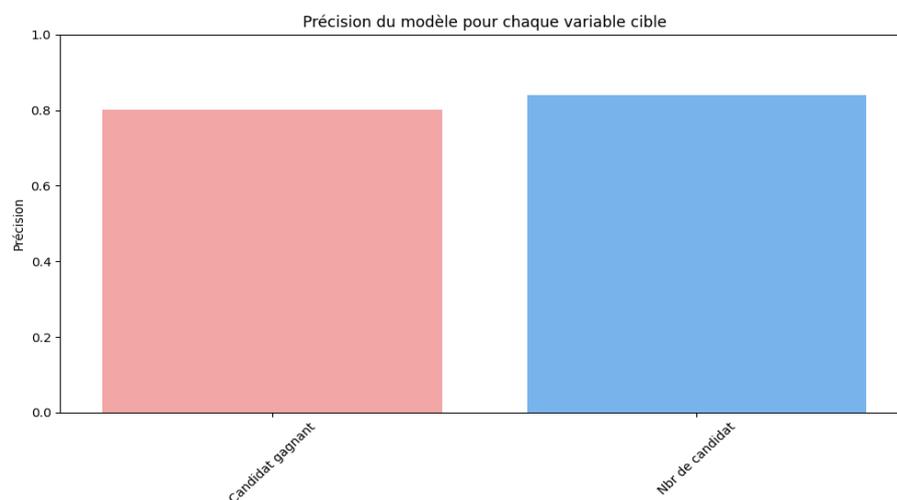


Figure 3.7: Précision par cible pour le modèle Extra tree.

La **Figure 3.7** représente la précision du modèle *ExtraTree* pour le nombre de candidats et les résultats du candidat gagnant. Nous constatons que la précision est d'environ 0.80 pour le candidat gagnant, tandis que le nombre de candidats est d'environ 0.84.

Ces performances reflètent des capacités acceptables, mais pas idéales.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

➤ matrice de confusion :

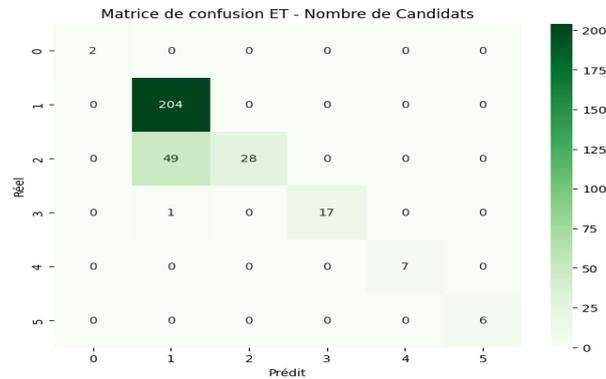


Figure 3 8: Matrices de confusion - nombre de candidats - pour le modèle *ExtraTreesClassifier*.

ExtraTreesClassifier, avec une précision légèrement inférieure (84 %), maintient aussi 204 prédictions correctes pour la classe 1. Néanmoins, il montre une plus grande confusion avec la classe 2, avec 49 cas de mauvaise classification de la classe 2 comme 1, ce qui dégrade sa performance globale sur les classes minoritaires.

3.6.1.1.3-Modèle *KNeighborsClassifier* :

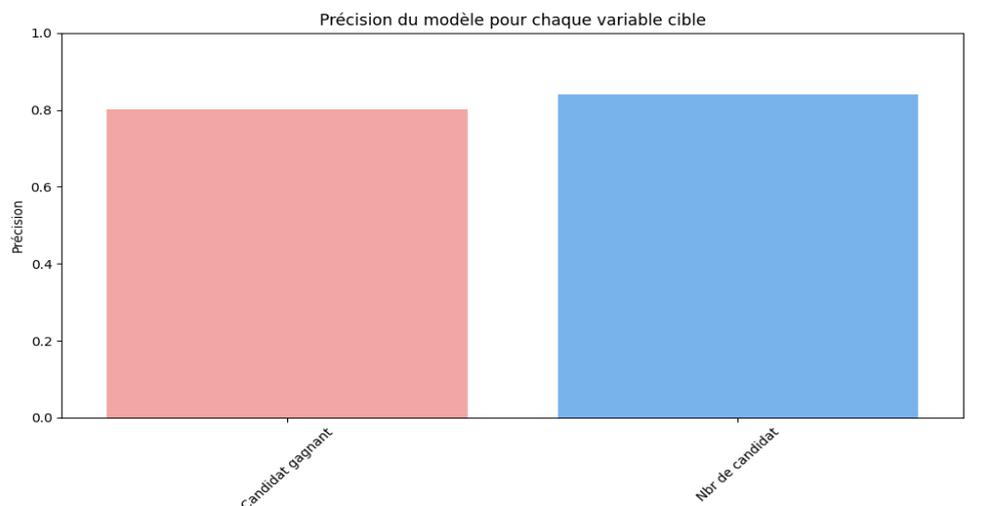


Figure 3.9: Précision par cible pour le modèle *KNN*.

La **Figure 3.9** représente la précision du modèle *KNN* pour le nombre de candidats et les résultats du candidat gagnant. Nous constatons que la précision est d'environ 0.83 pour le candidat gagnant, tandis que le nombre de candidats est d'environ 0.95.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Le modèle KNN a nettement surpassé les autres modèles, avec une précision élevée de 0.83 pour le candidat gagnant et de 0.95 pour le nombre de candidats.

Ces résultats démontrent son excellente capacité à gérer à la fois la classification et la prédiction numériques avec une grande précision, ce qui en fait un choix idéal pour les applications de prévision qui exigent les plus hauts niveaux de précision dans le contexte actuel.

➤ matrice de confusion :

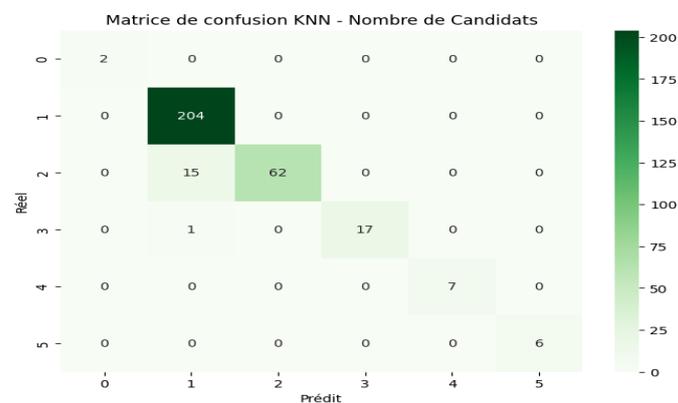


Figure 3.10: Matrices de confusion - nombre de candidats - pour le modèle `KNeighborsClassifier`.

`KNeighborsClassifier` se distingue avec une précision encore plus élevée (95 %). Il prédit également correctement 204 instances de la classe 1, mais avec moins d'erreurs sur la classe 2 comparé aux autres modèles : seulement 15 cas où la classe 2 est prise pour 1. De plus, il conserve une bonne précision sur les autres classes, ce qui souligne une meilleure capacité de généralisation.

✚ Évaluation et comparaison des résultats des modèles basés sur l'apprentissage automatique :

Le **Tableau 3.3** représente la précision des trois modèles (`RandomForestClassifier`, `ExtraTreesClassifier`, `KNeighborsClassifier`).

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Tableau 3.3: Évaluation et comparaison des modèles basés sur l'apprentissage automatique.

Modèle	Précision Candidat Gagnant	Précision Nombre de Candidats
RandomForestClassifier	0.82	0.93
ExtraTreesClassifier	0.80	0.84
KNeighborsClassifier	0.83	0.95

Les résultats obtenus démontrent clairement la supériorité du modèle *KNN* par rapport aux autres, notamment *RandomForestClassifier* et *ExtraTreesClassifier*, en particulier pour la prédiction à long terme du candidat gagnant et du nombre de candidats, avec des précisions élevées de 0.83 et 0.95 respectivement, ce qui en fait le choix idéal ; cette différence de performance, surtout marquée dans la prédiction du nombre de candidats, est nettement visible dans les matrices de confusion.

3.6.1.2-Modèles basés sur Deep Learning :

Dans cette section, nous présenterons deux approches différentes : le modèle de *DNN* et le modèle de *GRU*. La **Figure 3.11** représente l'architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

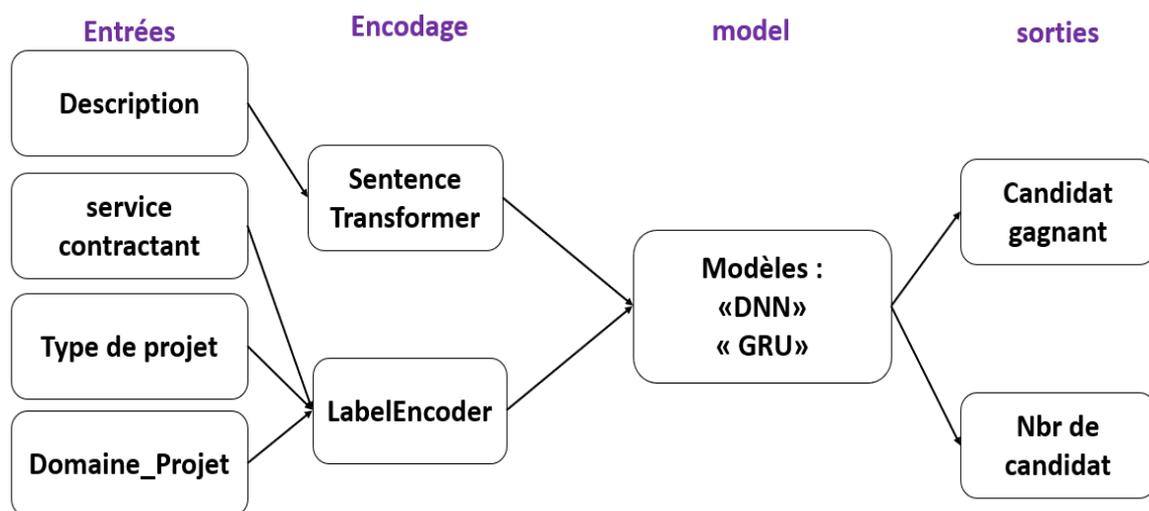


Figure 3.11: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Ces deux approches partagent de nombreux hyperparamètres représentés dans le tableau ci-dessous :

Tableau 3.4: Entraînement des hyperparamètres des modèles basés sur Deep Learning.

Hyperparamètre	Valeur	Description
Hidden Size	256	Nombre d'unités dans les couches cachées pour les deux modèles.
Embedding Size	32	Taille d'incrustation pour les données catégorielles (service, type, domaine).
Learning Rate	0.001	Taux d'apprentissage utilisé avec l'optimiseur Adam pour les deux modèles.
Epochs	100	Nombre d'itérations d'entraînement pour les deux modèles.
Batch Size	32	Nombre d'échantillons par lot pendant l'entraînement et la validation.
Test Size	0.2	Proportion des données utilisées pour la validation (20 %) par rapport à l'entraînement.
Weight Decay	1e-4	Valeur de régularisation utilisée dans Adam pour éviter le surapprentissage.
Candidate Count Head	Linear (128, 1)	Structure de la tête de prédiction pour le nombre de candidats avec une couche intermédiaire de taille 128.
Activation Function in Head	ReLU	Utilisation de ReLU comme fonction d'activation avant les couches finales.
Nombre of Hidden layer	4 DNN / 2 GRU	Nombre de couches cachées des modèles.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Les indicateurs de performance, tels que la perte et la précision, sont présentés dans le **Tableau 3.4**. Lors de l'entraînement, nous avons fixé les hyperparamètres afin de pouvoir placer tous les modèles dans le même environnement d'entraînement et les évaluer ensemble.

3.6.1.2.1- Modèle GRU :

Dans les figures ci-dessous, nous présentons l'évolution des métriques de performance au cours des premières étapes de l'entraînement et de la validation de notre modèle (GRU).

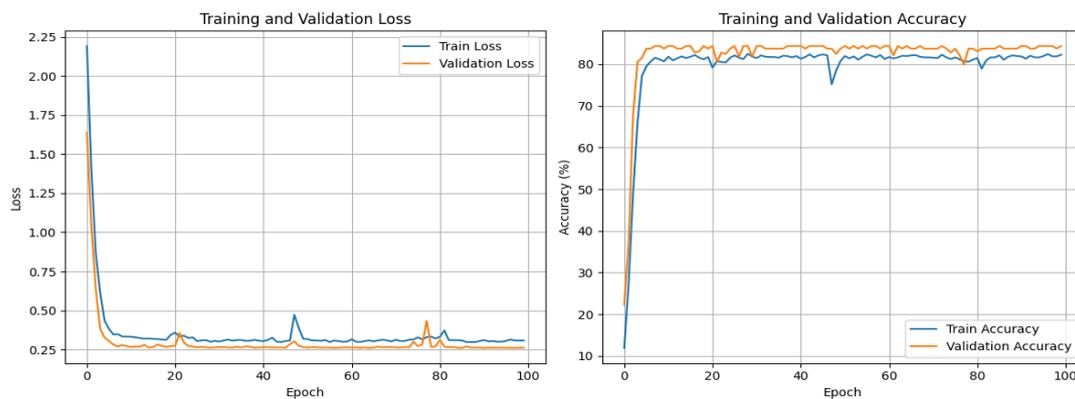


Figure 3.12: Évolution des valeurs de la métrique de performance à l'aide du modèle GRU.

D'après **Figure 3.12**, nous remarquons que la courbe de perte d'entraînement (ligne bleue) démarre à une valeur élevée d'environ 2.25 et diminue rapidement au cours des premières époques, pour se stabiliser autour de 0.3 après l'époque 20. La perte de validation (ligne orange) suit une évolution similaire, débutant à un niveau élevé puis diminuant rapidement pour se maintenir entre 0.25 et 0.3. On note cependant de légères augmentations de la perte d'entraînement autour des époques 40 et 80, suggérant une certaine instabilité durant l'apprentissage. Parallèlement, la précision d'entraînement (ligne bleue) grimpe rapidement à partir de valeurs faibles pour atteindre environ 80 % après l'époque 10, tandis que la précision de validation (ligne orange) augmente encore plus vite, dépassant celle de l'entraînement et se stabilisant autour de 82-83 %. Une chute brutale de la précision d'entraînement vers l'époque 40 coïncide avec un pic de perte, illustrant une phase d'instabilité temporaire. Malgré cela, les performances globales du modèle restent bonnes, la précision de validation étant généralement supérieure à celle de l'entraînement, ce qui témoigne d'une bonne capacité de généralisation.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.6.1.2.2-Modèle DNN :

Dans la **Figure 3.13** ci-dessous, nous présentons l'architecture de DNN utilisée :

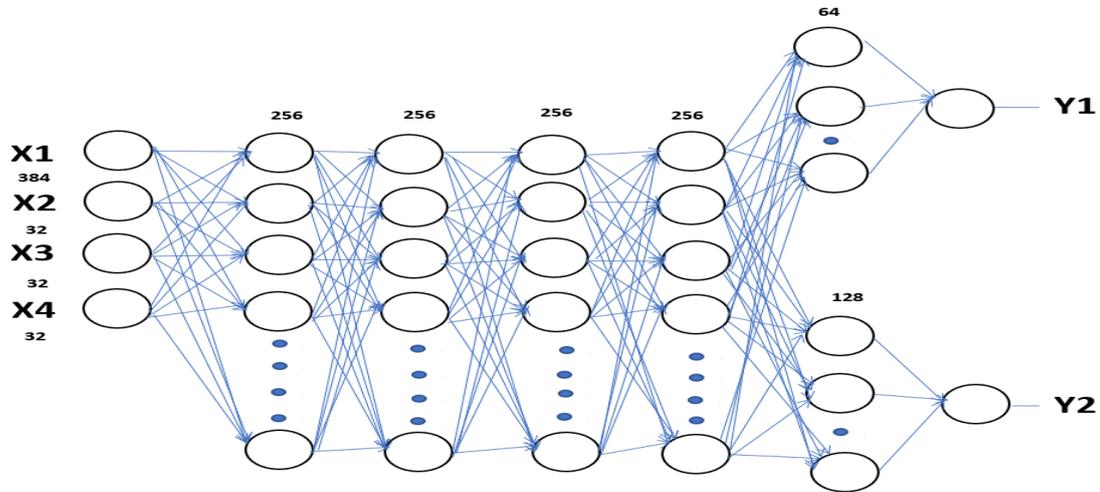


Figure 3.13: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

X1 : Description ; X2 : Service contractant ; X3 : Type de projet ; X4 : Domaine du projet ;

Y1 : Candidat gagnant ; Y2 : Nombre de candida.

Dans la **Figure 3.14** ci-dessous, nous présentons l'évolution des métriques de performance au cours des premières étapes de l'entraînement et de la validation de notre modèle (DNN).

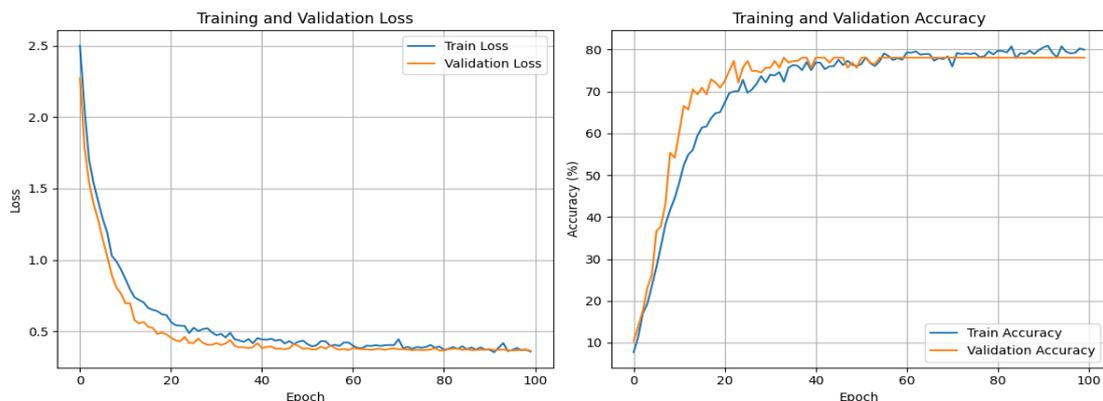


Figure 3.14: Évolution des valeurs de la métrique de performance à l'aide du modèle DNN.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

D'après **Figure 3.14**, nous remarquons que La perte d'entraînement et de validation a commencé à des valeurs élevées, autour de 2.5.

Elles ont toutes deux diminué de manière similaire jusqu'à l'époque 20.

Par la suite, elles ont continué à diminuer progressivement, avec une perte d'entraînement se stabilisant autour de 0.4 et une perte de validation autour de 0.35 après l'époque 60.

La perte de validation étant inférieure à la perte d'entraînement dans la plupart des cas, cela suggère une bonne capacité de généralisation.

La précision d'entraînement et de validation a commencé à des valeurs faibles et a augmenté progressivement.

La précision de validation (ligne orange) a augmenté plus rapidement que celle de l'entraînement entre les époques 10 et 30.

Après l'époque 40, les deux précisions se sont rapprochées, avec un léger avantage pour la précision d'entraînement, et se sont stabilisées autour de 78-80 %.

Le modèle montre une amélioration continue jusqu'à environ l'époque 60, après quoi les performances se stabilisent.

Bonnes performances sans surapprentissage. Le point d'arrêt optimal se situe autour de 60 époques. La précision finale acceptable se situe entre 78% et 80 %.

Évaluation et comparaison des résultats des modèles basés sur Deep learning :

Tableau 3.5: Évaluation et comparaison des modèles DNN et GRU.

Architecture	Train_loss	Val_loss	Train_Acc	Val_Acc
DNN	0.3521	0.3640	81.10%	78.09%
GRU	0.2986	0.2638	82.54%	84.39%

D'après le **Tableau 3.5**, le modèle GRU se distingue clairement comme le plus performant parmi ceux testés. En effet, il affiche une précision (Accuracy) stable d'environ 0.82 en entraînement et 0.84 en validation, surpassant ainsi le modèle DNN, dont la précision atteint 0.81 en entraînement et 0.78 en validation. En termes de perte (loss), le GRU confirme

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

sa supériorité avec des valeurs de 0.2986 pour l'entraînement et 0.2638 pour la validation, contre 0.3521 et 0.3640 respectivement pour le DNN. Ces résultats montrent que le GRU offre de meilleures performances globales, tant en précision qu'en minimisation de la perte. Les figures illustrant l'évolution des métriques de performance au cours des premières étapes d'entraînement et de validation pour les deux modèles (DNN et GRU) révèlent une progression régulière de l'apprentissage, suivie d'une stabilisation des pertes et des précisions, traduisant une convergence effective.

3.6.1.3-Modèles basés sur NLP :

La **Figure 3.15** représente l'architecture du modèle basé sur NLP de prédiction du candidat gagnant et du nombre de candidats.

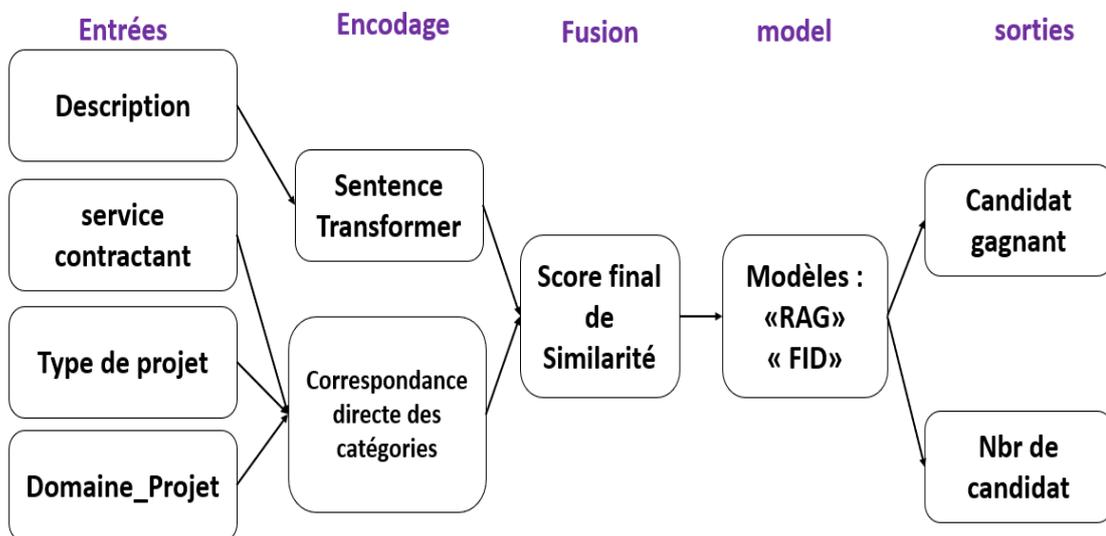


Figure 3.15: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

Nous avons utilisé deux méthodes basées sur NLP, à savoir **RAG** et **FID**. Nous avons soumis les deux approches à des conditions similaires pour mesurer leurs performances et déterminer ainsi les hyperparamètres utilisés dans les deux systèmes (FID et RAG) lors des étapes de récupération et de génération :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Tableau 3.6: Hyperparamètres des modèles RAG et FID.

Hyperparamètre	Valeur	Description
top_k	5	Nombre de documents récupérés depuis la base de connaissances.
text_weight	0.7	Poids relatif de la similarité textuelle dans le calcul du score final.
metadata_weight	0.3	Poids relatif de la similarité des métadonnées dans le calcul du score final.
max_length	512	Longueur maximale des entrées lors de l'encodage (documents ou contexte).
early_stopping	True	Activation de l'arrêt anticipé lors de la génération pour éviter les répétitions ou les sorties trop longues.

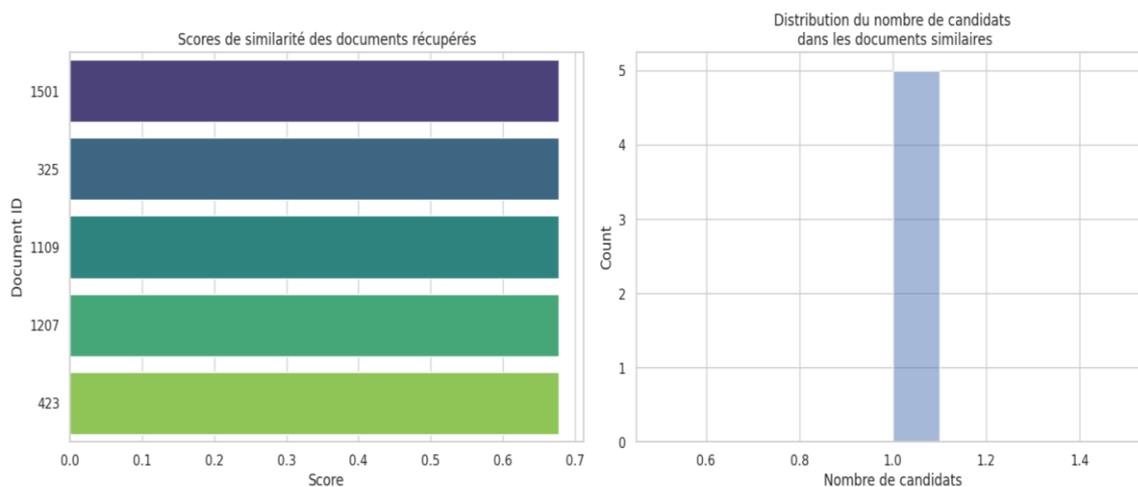
Il existe des différences entre les deux modèles (*RAG*, *FID*), que nous représentons dans le **Tableau 3.7** :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Tableau 3.7: Comparaison des hyperparamètres des modèles RAG et FID.

Hyperparamètre	FID	RAG	Description
num_beams	4	5	Nombre de faisceaux dans la recherche en faisceau, influence la qualité et la vitesse de génération.
max_new_tokens	128	200	Longueur maximale du texte généré (utilisé explicitement seulement dans RAG).
do_sample	False	True	Activation de l'échantillonnage aléatoire lors de la génération (rend RAG plus diversifié).
no_repeat_ngram_size	Non spécifié	2	Longueur maximale des entrées lors de l'encodage (documents ou contexte).
temperature	Non spécifié (par défaut : 1.0)	0.7	Contrôle de l'aléa dans la génération (RAG est moins aléatoire grâce à une température plus basse).

3.6.1.3.1- Modèle de RAG :



Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Figure 3.16: Analyse de similarité documentaire et distribution des candidats pour le modèle RAG.

La **Figure 3.16** se compose de deux sections : la première affiche les scores de similarité des documents récupérés pour cinq documents (423, 1 207, 1 109, 325 et 1 501), avec des valeurs comprises entre 0.6 et 0.7 environ. La seconde présente la distribution du nombre de candidats dans les documents similaires, avec une nette concentration autour de 1.0 candidat.

Scores de similarité (RAG) : Le premier graphique contient des barres horizontales représentant les scores de similarité des documents récupérés à l'aide de la technique RAG (Retrieval-Augmented Generation). Tous les documents ont un score (autour de 0.7), en raison de la similarité des données. Répartition du nombre de candidats : Le deuxième graphique est un histogramme montrant la répartition du nombre de candidats dans des documents similaires. La concentration maximale est de 1.0 candidat, ce qui signifie que la plupart des documents récupérés ne contiennent qu'un seul candidat.

3.6.1.3.2-Modele de FID :

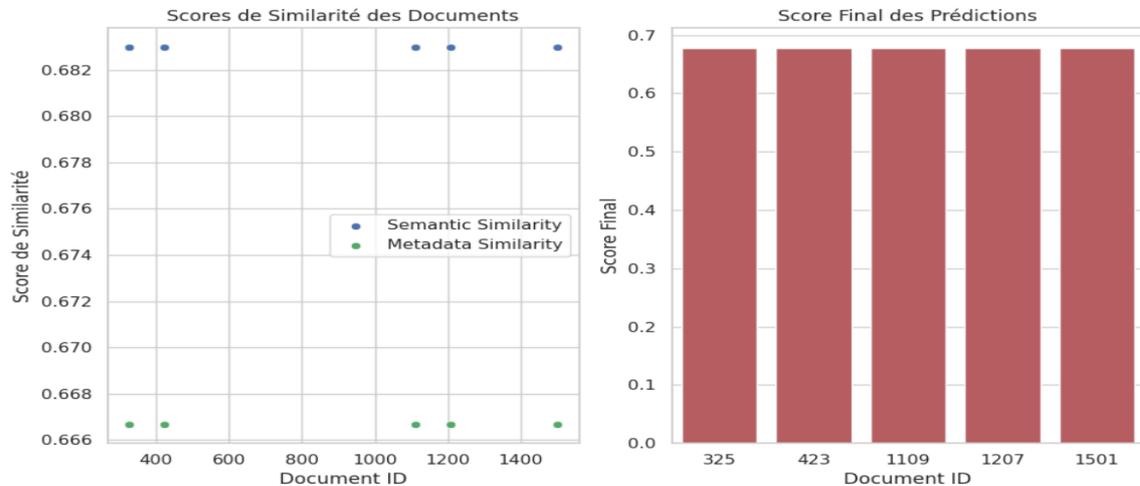


Figure 3.17: Analyse de similarité documentaire et distribution des candidats pour le modèle FID.

La **Figure 3.17** comporte deux sections : la première affiche les scores de similarité des documents basés sur la similarité sémantique et la similarité des métadonnées, avec des valeurs comprises entre 0.666 et 0.682. La deuxième section affiche les scores de prédiction finaux pour les documents sélectionnés (325, 423, 1 109, 1 207 et 1 501), avec des valeurs comprises entre 0.5 et 0.7.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Scores de similarité des documents : Le premier graphique présente des points dispersés représentant les scores de similarité sémantique (en bleu) et les scores de similarité descriptive (en vert) sur une large plage de numéros de documents (environ 400 à 1 400). Les valeurs sont relativement proches, indiquant une concordance modérée entre les deux similarités, avec un léger écart d'environ 0.016 (de 0.666 à 0.682).

Scores de prédiction finaux : Le deuxième graphique affiche les colonnes de cinq documents sélectionnés, avec des valeurs comprises entre 0.6 et 0.7. Cela peut refléter une moyenne pondérée ou un score final dérivé de la combinaison des scores de similarité sémantique et descriptive, avec une nette tendance vers les valeurs élevées (0.6-0.7) dans la plupart des cas.

✚ Évaluation et comparaison des résultats des modèles basés sur NLP :

Nous avons utilisé l'exemple présenté dans le **Tableau 3.8** ci-dessous, et les résultats sont les suivants :

Tableau 3.8: Évaluation et comparaison des deux modèles.

Description	Entrée	Réalisation d'un projet d'assainissement
Service contractant	Entrée	Municipalite de mila
Type projet	Entrée	Realisation des travaux
Domaine projet	Entrée	Hydraulique
Candidat gagnant	Sortie	mouhammed hakim
Nbr de candidat	Sortie	1

À partir des résultats obtenus, **Tableau 3.8**, nous constatons que les deux modèles (RAG, FID) ont produit la même prédiction. De plus, en les comparant aux ensembles de données, leurs résultats se sont révélés très proches.

3.6.2-Prédire l'ensemble des conditions du cahier des charges :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

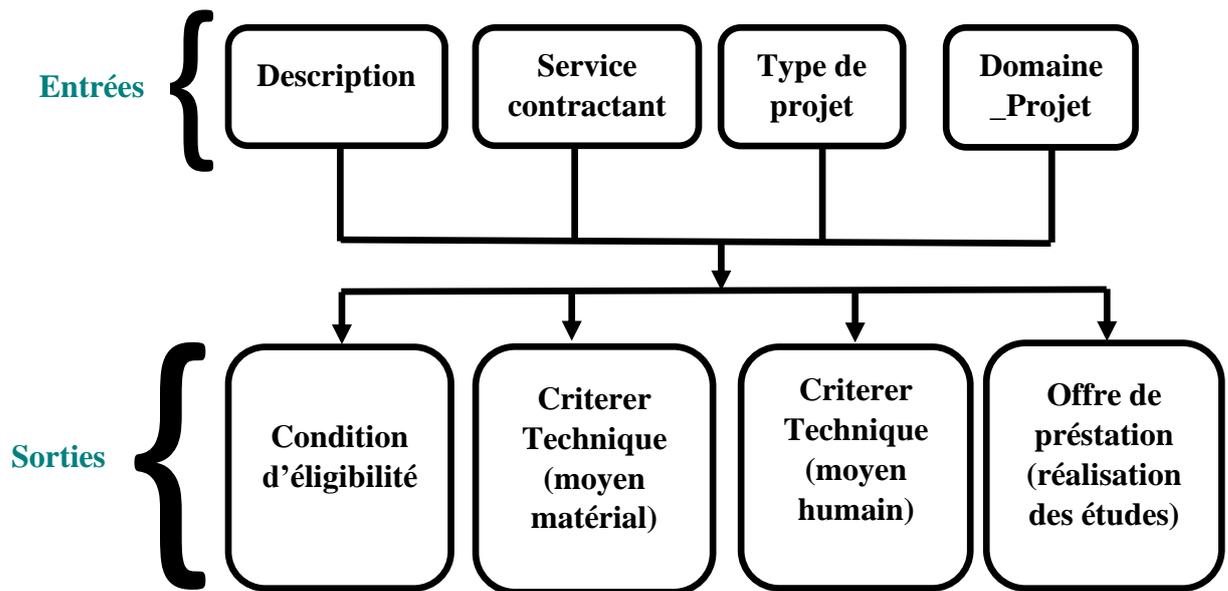


Figure 3.18: entrées / sorties pour la prédiction des conditions à partir des caractéristiques du projet.

3.6.2.1-Modèles basés sur l'apprentissage automatique :

Nous avons utilisé trois modèles de classification différents pour développer un système de classification multi-sorties capable de prédire des variables cibles à partir de données textuelles et catégoriques. Le premier modèle : *RandomForestClassifier*, le deuxième modèle : *ExtraTreesClassifier* et le troisième modèle : *KNeighborsClassifier*. la **Figure 3.19** représente l'architecture du modèle basé sur l'apprentissage automatique de prédiction des conditions d'un cahier des charges.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

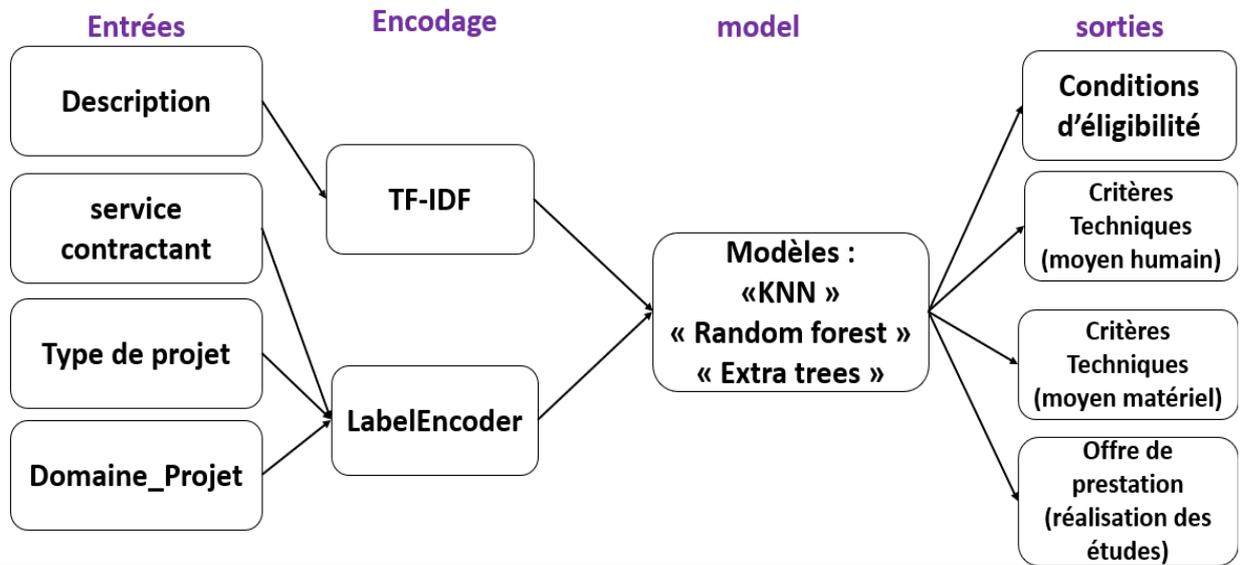


Figure 3.19: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

Le **Tableau 3.9** représente les modèles utilisés avec les différences entre eux :

Tableau 3.9: Les modèles basés sur l'apprentissage automatique utilisés avec leurs différences.

Modèle	Approche	Hyperparamètres	Précision	Particularité
RandomForestClassifier	Arbres avec divisions optimales.	n_estimators=100, max_depth=10, random_state=42	1.00 (toutes variables)	Plus précis, mais plus lent.
ExtraTreesClassifier	Arbres avec randomisation accrue.	n_estimators=250, max_depth=15, random_state=42	0.93 (majorité), 1.00 (une variable)	Rapide, moins précis sur données complexes.
KNeighborsClassifier	Basé sur les distances entre points.	n_neighbors=15, weights='distance'	1.00 (toutes variables)	Nécessite normalisation, sensible à k.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.6.2.1.1- Modèle RandomForestClassifier :

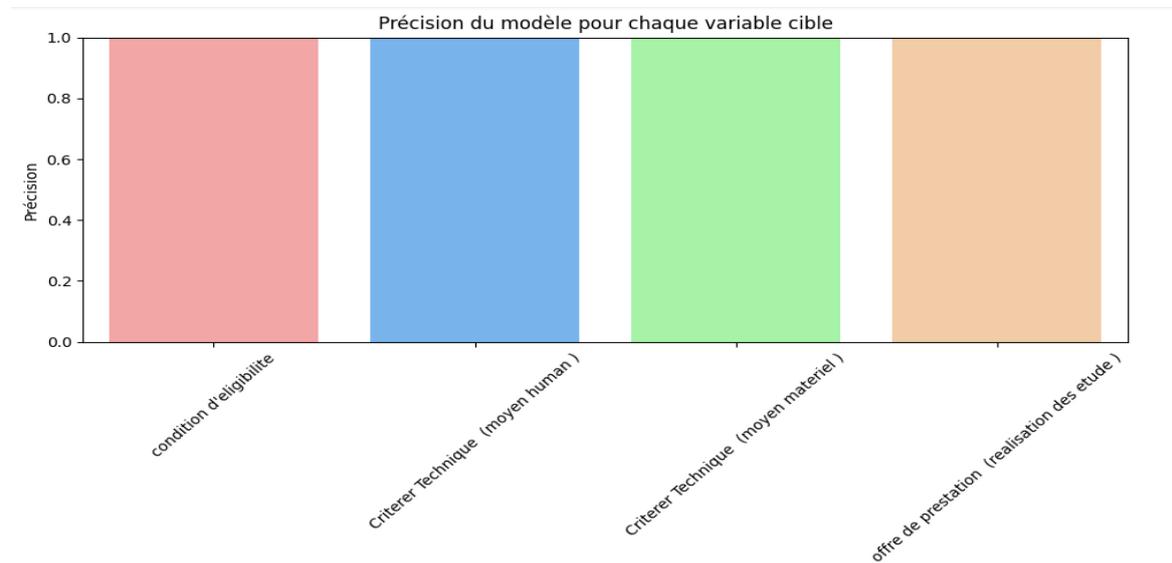


Figure 3.20: Courbe de précision pour RandomForest.

À partir de la **Figure 3.20**, nous remarquons que le modèle RandomForest a atteint une excellente précision de 1.00 pour toutes les variables ciblées, démontrant une forte capacité à gérer des relations complexes dans les données.

3.6.2.1.2- Modèle ExtraTreesClassifier :

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

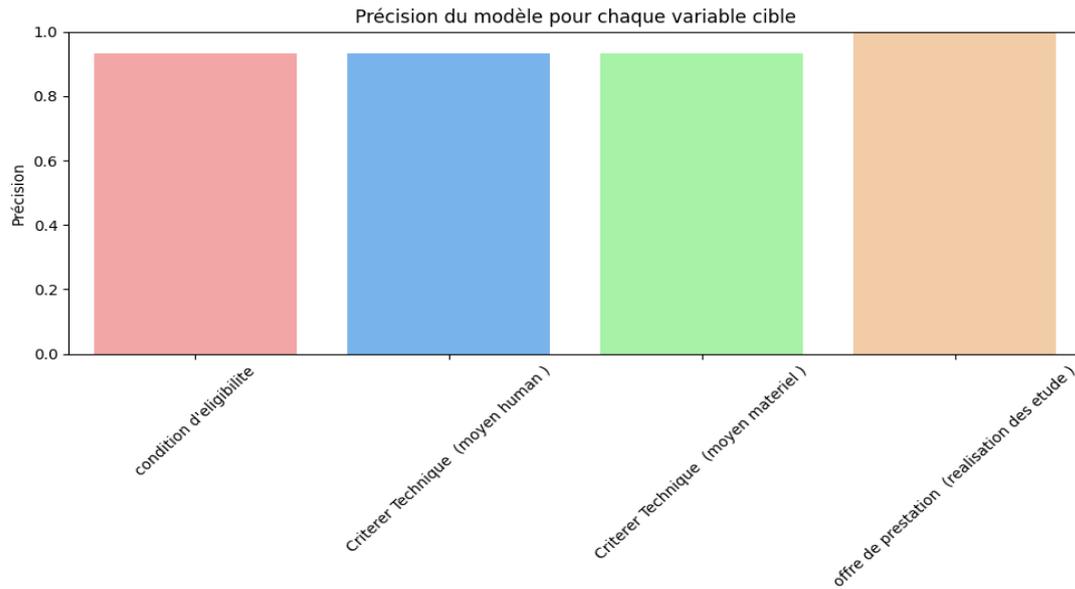


Figure 3.21: Courbe de précision pour ExtraTreesClassifier.

À partir de la **Figure 3.21**, nous voyons que le modèle ExtraTreesClassifier classe la majorité des variables avec une très grande précision de 0.93. La précision la plus élevée semble être pour la variable "offre de prestation (réalisation des études)" qui atteint 1.0. Tandis que les autres variables ont des précisions légèrement inférieures mais toujours très bonnes.

3.6.2.1.3- Modèle KNeighborsClassifier :

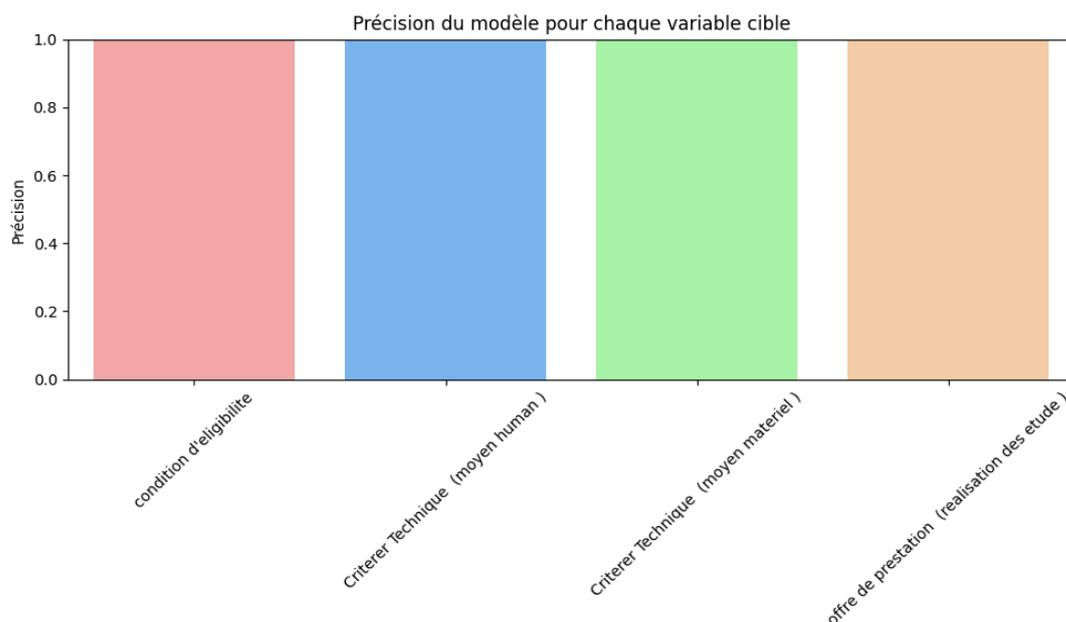


Figure 3.22: Courbe de précision pour KNeighborsClassifier.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

À partir de la **Figure 3.22**, nous voyons que le modèle *KNeighborsClassifier* a également atteint une précision élevée de 1.00 pour toutes les variables cibles, ce qui est comparable aux autres modèles testés.

3.6.2.2-Modèle basé sur Deep Learning :

Dans ce qui suit, nous allons présenter le modèle DNN. la **Figure 3.23** représente l'architecture du modèle basé sur l'apprentissage profond de prédiction des conditions d'un cahier des charges.

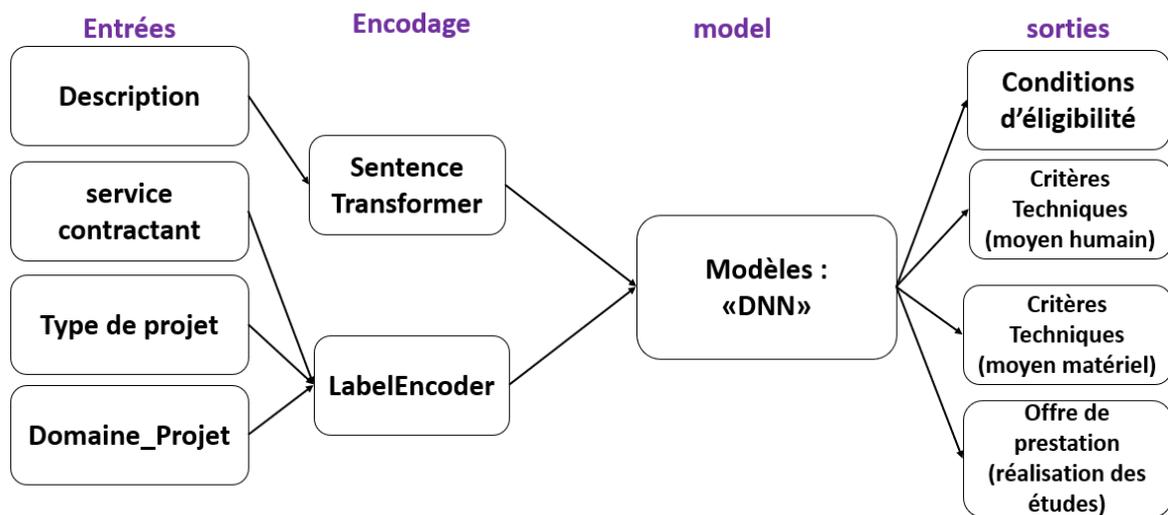


Figure 3.23: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

3.6.2.2.1-DNN :

Le modèle utilisé est le réseau de neurones profond (*DNN*) permettant de prédire quatre sorties catégorielles (conditions d'éligibilité, ressources humaines, ressources physiques et offres de services) à partir de représentations textuelles (SentenceTransformer : paraphrase-MiniLM-L6-v2) et de données catégorielles (service, type, domaine). Il comprend des couches d'intégration, des couches partagées (trois couches cachées de taille 128) et des sommets de prédiction discrets. Il est entraîné avec Adam et CrossEntropyLoss avec arrêt anticipé.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Tableau 3.10: Hyperparamètres du modèle DNN.

Hyperparamètre	Valeur	Description
Text Embedding Size	384	Taille du vecteur texte (MiniLM).
Categorical Embedding Size	32	Taille du vecteur pour les catégories.
Hidden Layer Size	128	Nombre d'unités dans les couches partagées.
Number of Hidden Layers	3	Couches partagées avec ReLU et Dropout.
Dropout Rate	0.3, 0.2	Dropout pour réduire le surajustement.
Prediction Head Layer Size	64	Unités dans la couche de tête de prédiction.
Number of Epochs	100	Maximum de cycles d'entraînement.
Learning Rate	0.001	Taux de mise à jour des poids (Adam).
Batch Size	32	Nombre d'échantillons par lot.
Patience	10	Époques avant arrêt anticipé.
Loss Function	CrossEntropyLoss	Perte par sortie, somme des pertes.
Optimizer	Adam	Algorithme de mise à jour des poids.

Les données ont été divisées comme suit :

Tableau 3.11: Division de dataset.

Division de dataset	Pourcentages	N-projets
Ensemble d'entraînement	59.9%	940
Ensemble de validation	20%	314
Ensemble de tests	20%	314
Tout	100%	1568

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Les mesures de performance telles que 'loss', 'accuracy' sont présentées dans le (tableau).

Tableau 3.12: Évaluation du modèle DNN.

Architecture	Train_loss	Val_loss	Train_Acc	Val_Acc
DNN	0.1606	0.0032	0.9851	1.0000

À partir du **Tableau 3.12**, on peut observer que le modèle DNN obtient de meilleurs résultats. Cette approche donne des performances remarquables, et on peut constater la même tendance avec l'évaluation utilisant la métrique de précision, où le modèle a atteint une valeur maximale de 0.9915 en termes de précision.

Dans **Figure 3.24**, nous présentons l'évolution des métriques de performance lors des premières étapes de l'entraînement et de la validation de notre modèle (DNN), là où l'entraînement était à son apogée.

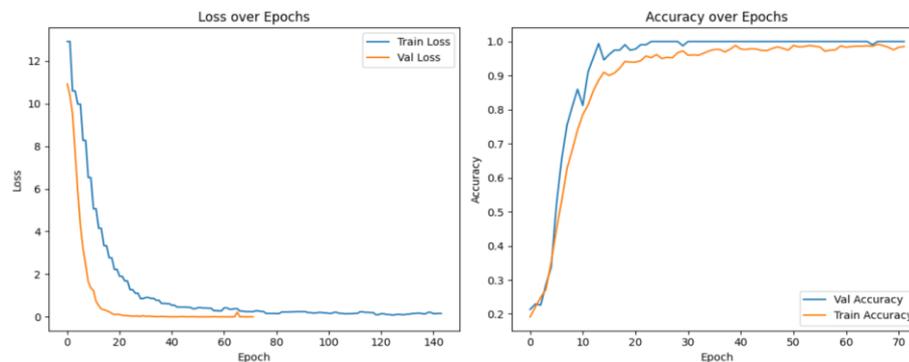


Figure 3.24: Évolution des valeurs de la métrique de performance à l'aide du modèle DNN.

La **Figure 3.24** montre la précision de l'entraînement (Train Accuracy) et la précision de la validation (Val Accuracy) au fil des époques.

La précision de l'entraînement augmente de manière stable, atteignant finalement un plateau autour de 0.94 après environ 20 époques.

La précision de la validation augmente rapidement au début, atteignant environ 1 après environ 15 époques.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Le modèle montre une bonne capacité de généralisation. Toutefois, il reste une inquiétude concernant la ressemblance excessive des données.

Ce graphique illustre la perte d'entraînement (Train Loss) et la perte de validation (Val Loss) au fil des époques.

La perte d'entraînement diminue continuellement, se stabilisant autour de 0.1 après environ 40 époques.

La perte de validation diminue fortement au début, puis oscille légèrement avant de se stabiliser autour de 0.01 également après environ 40 époques.

La diminution de la perte d'entraînement indique que le modèle apprend efficacement à partir des données d'entraînement.

3.6.2.3-Modèles basés sur NLP :

On a conçu deux modèles pour la récupération d'informations et la génération de réponses : *RAG* et *FID*, la **Figure 3.25** Représente l'architecture du modèle basé sur l'apprentissage profond de prédiction des conditions d'un cahier des charges.

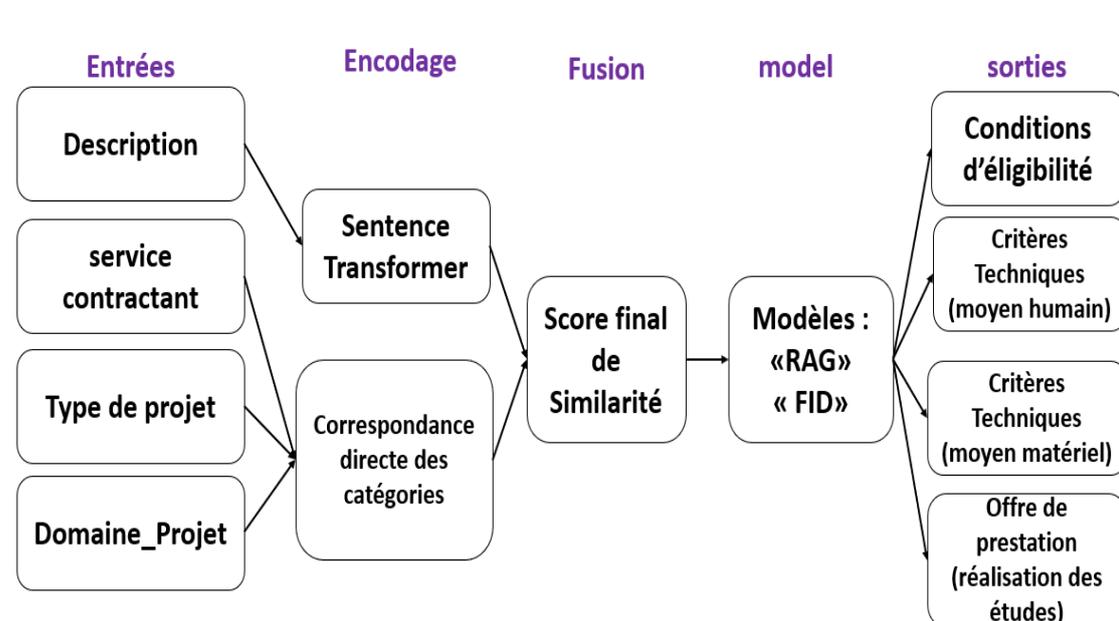


Figure 3.25: Architecture du modèle basé sur l'apprentissage profond de prédiction du candidat gagnant et du nombre de candidats.

On a choisi d'utiliser le modèle Sentence Transformer ('paraphrase-MiniLM-L6-v2') pour la récupération dans les deux codes.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

Pour la génération de texte, on s'est basé sur le modèle google/flan-t5-base.

On a utilisé la similarité sémantique (cosine similarity) en intégrant les métadonnées.

Voici **Tableau 3.13** des hyperparamètres utilisés :

Tableau 3.13: Hyperparamètres des modèles RAG et FID.

Hyperparamètres	Valeur	Description
top_k	5	Nombre de documents récupérés
text_weight	0.7	Poids de la similarité du texte dans la récupération
metadata_weight	0.3	Poids de la similarité des métadonnées dans la récupération
max_new_tokens	200 (RAG), 256 (FID)	Longueur maximale du texte généré
do_sample	True (RAG), - (FID)	Utilisation d'échantillons aléatoires dans la génération
num_beams	5 (RAG), 4 (FID)	Nombre de faisceaux (beam) dans la recherche
temperature	0.7 (RAG), - (FID)	Température de génération (diversité)
no_repeat_ngram_size	2 (RAG), - (FID)	Prévention de la répétition des n-grams

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.6.2.3.1-Modèle RAG :

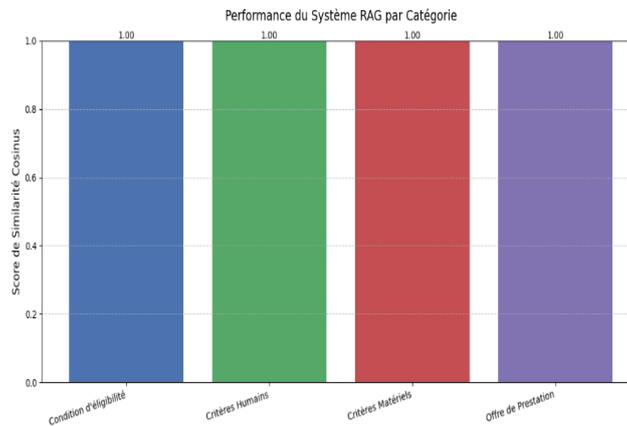


Figure 3.26: Courbe de RAG.

D'après la **Figure 3.26** représentant l'évaluation du modèle RAG (à gauche), on observe que la précision atteint la valeur maximale de 1.00 pour l'ensemble des sorties. Ce résultat indique une correspondance parfaite entre les prédictions du modèle et les valeurs réelles pour toutes les catégories, reflétant ainsi une performance optimale avec une précision de 100 %.

3.6.2.3.2-Modèle de FID :

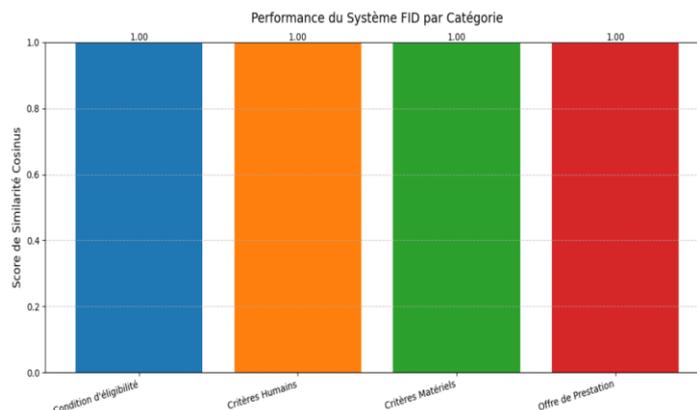


Figure 3.27: Courbe de FID.

D'après la **Figure 3.27** illustrant l'évaluation du modèle FID (à droite), on constate que la précision atteint la valeur maximale de 1.00 pour toutes les sorties. Le système FID a ainsi démontré une performance remarquable, avec des scores de similarité cosinus parfaits dans l'ensemble des catégories évaluées, ce qui témoigne de son efficacité tant dans la récupération de documents pertinents que dans la génération de prédictions précises.

Chapitre 03 : Validation et Évaluation des performances des modèles proposés

3.7-Conclusion

Dans ce chapitre, nous avons exposé les outils utilisés, décrit l'objectif de l'étude et le jeu de données utilisé, puis détaillé les modèles développés, comprenant des approches d'apprentissage automatique (Random Forest, ExtraTrees, KNN), d'apprentissage profond (DNN, GRU) et de retrieval et génération (RAG, FiD).

Le chapitre suivant présente une application web élaborée à partir de ces modèles.

Chapitre 04 :
Application Web de
système de prédiction des
marchés publics

Chapitre 04 : Application web de système de prédiction des marchés publics

4.1-Introduction

Dans ce chapitre, nous avons développé une application web visant à prédire automatiquement certaines informations issues du cahier des charges (offre technique), et prédire aussi le nombre de candidats et le candidat gagnant, à partir de caractéristiques textuelles et catégorielles.

Ce chapitre est consacré à la présentation des outils et langages utilisés pour le développement, ainsi qu'à la description des différentes interfaces de l'application.

4.2-Présentation des outils et langages utilisés :

•Visual Studio Code :

Visual Studio Code est un éditeur de code source léger mais puissant qui s'exécute sur votre bureau et est disponible pour Windows, MacOS et Linux. Il est livré avec une prise en charge intégrée de JavaScript, TypeScript et Node.js et dispose d'un riche écosystème d'extensions pour d'autres langages et environnements d'exécution (tels que C++, C#, Java, Python, PHP, Go,.NET) [34].



Figure 4.1: Logo de Framework Visual Studio Code.

•Outils de développement web :

→ Le HTML permet de structurer le contenu d'une page web, en définissant des éléments tels que les titres, les paragraphes et les images.

→ Le CSS est utilisé pour le style et la mise en page, permettant de contrôler l'apparence des éléments HTML, tels que les couleurs, les polices et l'espacement.

→ JavaScript est un langage de programmation qui permet d'intégrer des éléments interactifs sur les pages web [23].

•Framework Flask :

Flask est un framework web léger pour Python qui vous permet de créer des applications web facilement et rapidement. Il fournit les outils et fonctionnalités nécessaires

Chapitre 04 : Application web de système de prédiction des marchés publics

pour créer un serveur web, gérer les requêtes et générer des réponses avec un minimum de code [23].

4.3-Interfaces graphiques du Système de prédiction des marchés publics :

Dans cette section, nous présentons nos interfaces d'application. Ce qui permet aux utilisateurs d'interagir facilement avec différentes fonctionnalités, telles que la sélection et la prédiction de modèles. Chaque fenêtre est conçue pour être simple et claire.

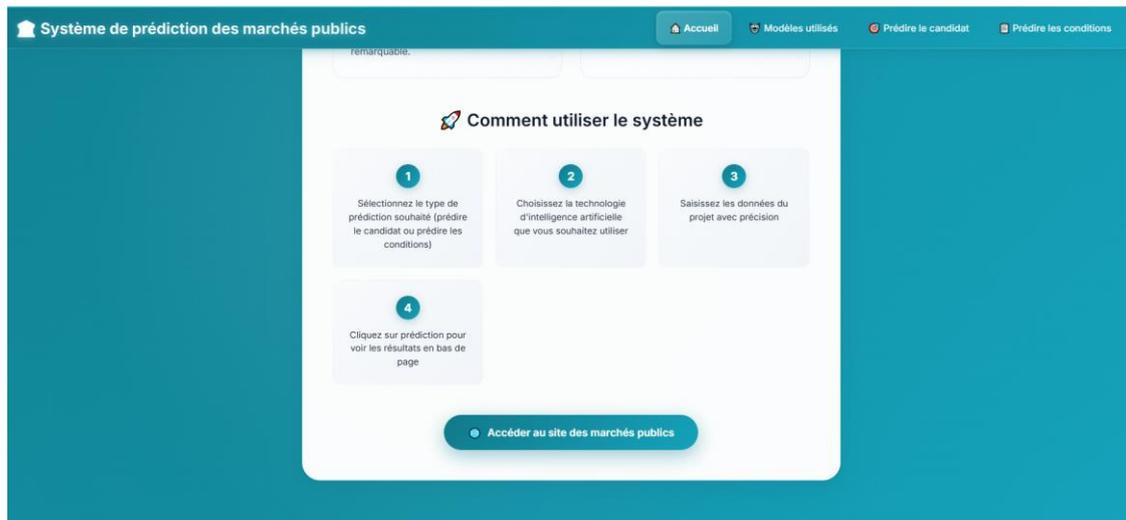
4.3.1-Page principale :

Cette interface représente la page d'accueil de notre application web. Elle permet à l'utilisateur de comprendre le rôle du système, d'accéder aux prédictions, et de suivre les étapes d'utilisation. Le design est simple, clair et orienté vers une interaction facile avec les modèles d'IA proposés.



(a) Définition du système ainsi que des marchés publics.

Chapitre 04 : Application web de système de prédiction des marchés publics



(b) Comment utiliser le système.

Figure 4 2: Page principale.

La **Figure 4.2** représente la page principale de l'application web du système de prédiction des marchés publics.

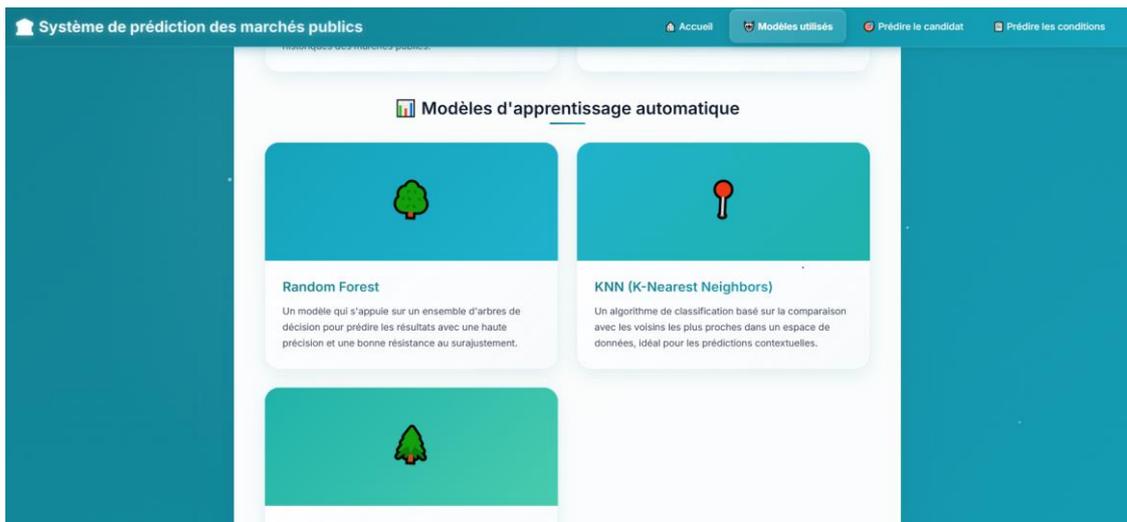
4.3.2- L'interface de description des modèles utilisés :

Cette interface présente un système de prévision des marchés publics qui utilise une variété de modèles d'intelligence artificielle et d'apprentissage automatique. Le système est organisé en trois catégories principales : les modèles d'apprentissage profond (GRU et DNN), les modèles d'apprentissage automatique (Random Forest, KNN et Extra tree) et les modèles de traitement du langage naturel (RAG et FiD). Accompagné d'une description expliquant son rôle spécifique dans la prédiction des résultats des transactions générales avec une précision remarquable.

Chapitre 04 : Application web de système de prédiction des marchés publics

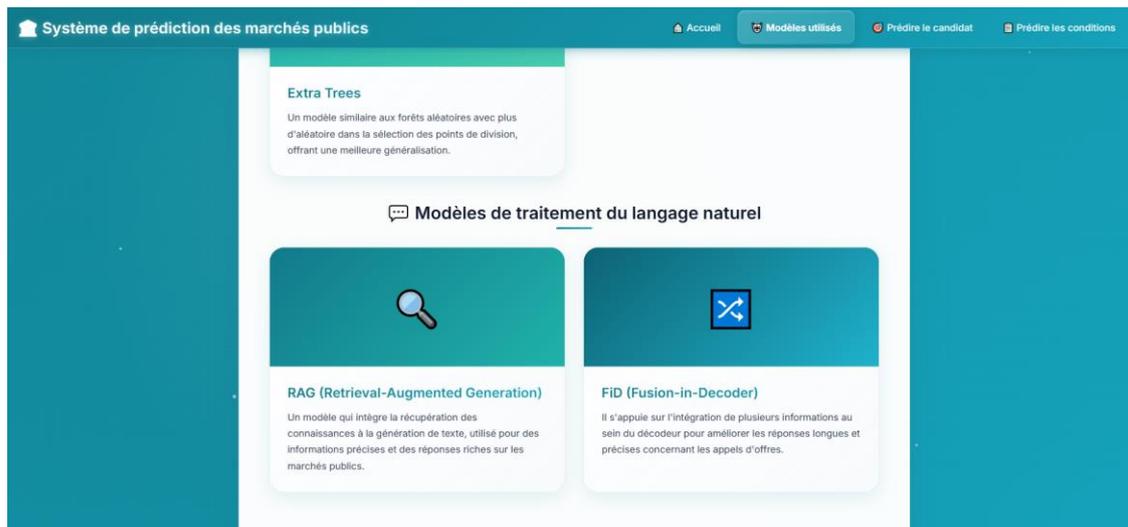


(a) Modèles d'apprentissage profond.



(b) Modèles d'apprentissage automatique.

Chapitre 04 : Application web de système de prédiction des marchés publics



(c) Modèles de NLP.

Figure 4.3: Modèles utilisés dans le système.

La **Figure 4.3** représente les modèles utilisés dans le système (DNN, GRU, Randomforest, Extra tree, KNN, RAG, FID).

3.4.3- Les domaines utilisés pour la prédiction des marchés publics :

Dans les deux cases, les champs qui seront utilisés pour les prévisions sont affichés (Deep learning, Machine learning et NLP).



(a) Prédire le candidat.

Chapitre 04 : Application web de système de prédiction des marchés publics



(b)Prédire les conditions.

Figure 4.4: Domaines utilisés dans la prédiction des Marchés Publics.

La **Figure 4.4** illustre les champs utilisés pour prédire le candidat et les conditions de spécification. Lorsque vous cliquez sur un champ (par exemple : Deep Learning), vous serez dirigé vers la page de prédiction.

4.3.4-Prédire le candidat :

Après avoir sélectionné le modèle dans le champ "Prédire le candidat", cette interface s'affiche. Elle contient les champs suivants : Description du projet, Type de projet, Service contractant, Domaine du projet, et Modèle de prédiction.

Une fois ces champs remplis avec les informations du projet souhaité, cliquez sur "Générer les prédictions".

Le résultat s'affiche en bas de la page, indiquant le nombre de candidats et le candidat gagnant.

Si un champ est laissé vide, un message d'erreur apparaît : "**Veillez renseigner ce champ**".

Chapitre 04 : Application web de système de prédiction des marchés publics

Prédiction de candidat GRU/DNN

Description du projet

réalisation d'un projet d'assainissement

Type de projet

Réalisation des travaux

Service contractant

Municipalité de Mita

Domaine du projet

Hydraulique

Modèle de prédiction

Réseau de Neurones Profond (DNN)

Générer les prédictions

Figure 4.5: la page des informations de projet.

La **Figure 4.5** représente la page des informations de projet.

Générer les prédictions

Résultats de la prédiction

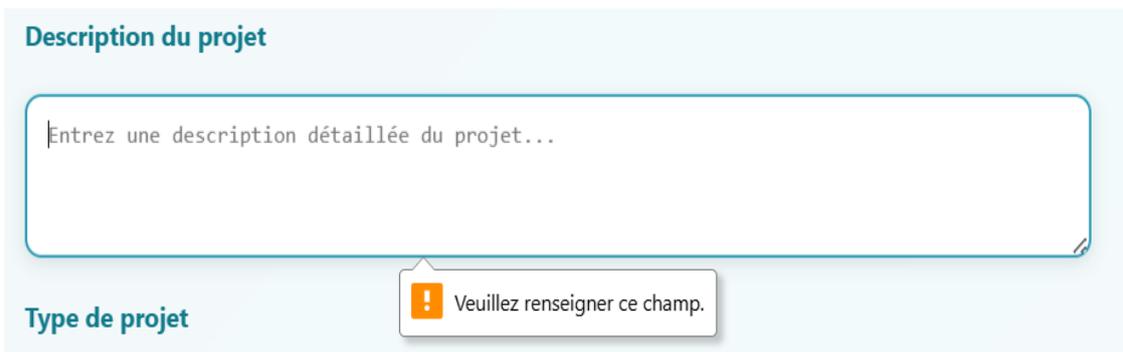
Nombre de candidats: 2 candidats

Candidat gagnant: mouhammed hakim

Figure 4.6: Résultats de la prédiction.

La **Figure 4.6** représente le résultat de la prédiction.

Chapitre 04 : Application web de système de prédiction des marchés publics



The screenshot shows a web form with two main sections: "Description du projet" and "Type de projet". The "Description du projet" section contains a large text input field with the placeholder text "Entrez une description détaillée du projet...". Below this field, an error message box is displayed, containing an orange exclamation mark icon and the text "Veuillez renseigner ce champ.".

Figure 4.7: Message d'alerte.

La **Figure 4.7** représente le cas où le champ n'est pas rempli, le message d'erreur suivant apparaît : " Veuillez renseigner ce champ".

4.3.5- Prédire les conditions de cahier des charges :

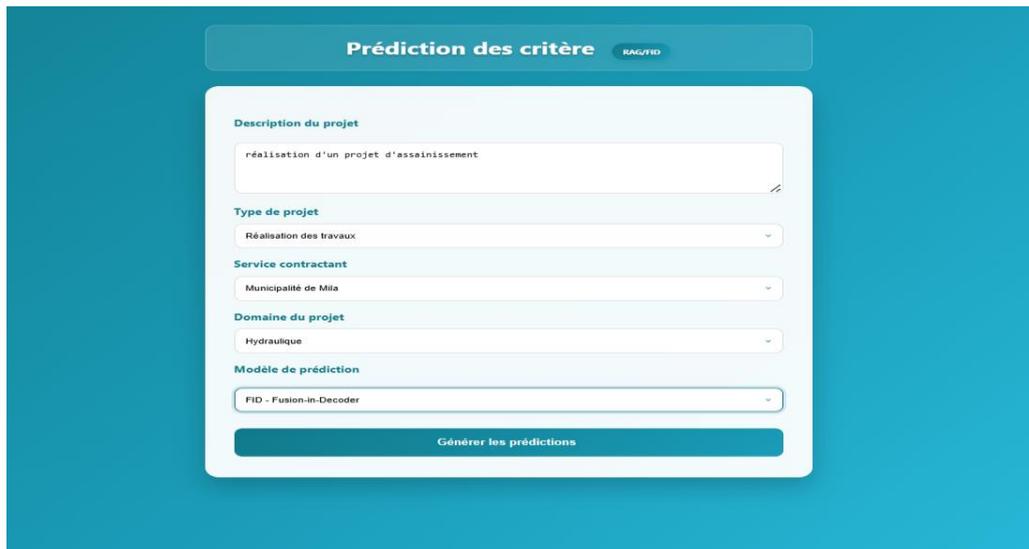
Après avoir sélectionné le modèle dans le champ "**Prédire les conditions**", cette interface s'affiche. Elle contient les champs suivants : Description du projet, Type de projet, Service contractant, Domaine du projet, et Modèle de prédiction.

Une fois ces champs remplis avec les informations du projet souhaité, cliquez sur "Générer les prédictions".

Le résultat s'affiche en bas de la page, indiquant les Conditions d'éligibilité, les Critères techniques (moyens humains), les Critères techniques (moyens matériels) et Offre de prestation (réalisation des études).

Si un champ est laissé vide, un message d'erreur apparaît : "**Veillez renseigner ce champ.**"

Chapitre 04 : Application web de système de prédiction des marchés publics



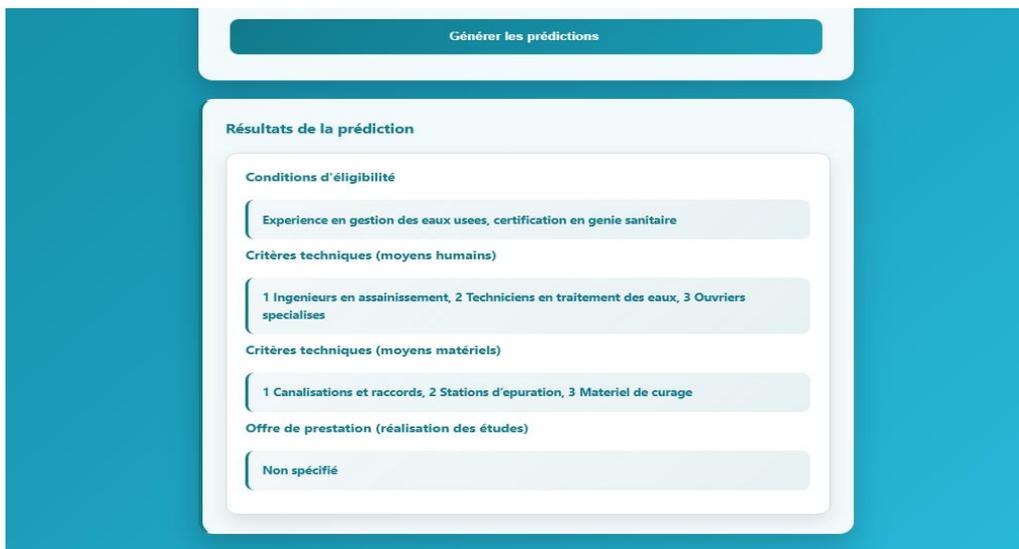
The screenshot shows a web interface titled "Prédiction des critères" with a "RAG/VID" button. Below the title is a form with the following fields:

- Description du projet:** A text input field containing "réalisation d'un projet d'assainissement".
- Type de projet:** A dropdown menu with "Réalisation des travaux" selected.
- Service contractant:** A dropdown menu with "Municipalité de Mla" selected.
- Domaine du projet:** A dropdown menu with "Hydraulique" selected.
- Modèle de prédiction:** A dropdown menu with "FID - Fusion-in-Decoder" selected.

At the bottom of the form is a button labeled "Générer les prédictions".

Figure 4.8: L'interface des informations de projet.

La **Figure 4.8** représente la page des informations de projet.



The screenshot shows a web interface titled "Résultats de la prédiction" with a "Générer les prédictions" button at the top. Below the title is a list of criteria:

- Conditions d'éligibilité:** "Expérience en gestion des eaux usées, certification en génie sanitaire"
- Critères techniques (moyens humains):** "1 Ingénieurs en assainissement, 2 Techniciens en traitement des eaux, 3 Ouvriers spécialisés"
- Critères techniques (moyens matériels):** "1 Canalisations et raccords, 2 Stations d'épuration, 3 Matériel de curage"
- Offre de prestation (réalisation des études):** "Non spécifié"

Figure 4.9: Résultats de la prédiction.

La **Figure 4.9** représente le résultat de la prédiction.

Ces pages ont été spécialement conçues pour les utilisateurs de l'application des achats publics (l'administration et l'entrepreneur), dans le but de faciliter l'utilisation des modèles prédictifs dans le cadre des projets publiés. Elles peuvent être intégrées ultérieurement dans l'application principale et être améliorées progressivement afin de mieux répondre aux besoins des utilisateurs.

Chapitre 04 : Application web de système de prédiction des marchés publics

4.4-Conclusion

Le développement et l'intégration des interfaces utilisateur au sein de cette application web ont représenté une avancée significative dans l'application de l'apprentissage automatique à l'analyse des marchés publics.

Dans ce chapitre, nous avons présenté les outils et les langages qui ont été utilisés, ainsi que les interfaces utilisateur graphiques du système de prédiction des marchés publics.

Conclusion générale

En conclusion, cette recherche a exploré le potentiel transformateur de l'intelligence artificielle dans l'amélioration des processus de marchés publics, en prenant pour terrain d'étude la commune de Mila. L'ambition était de concevoir un écosystème prédictif automatisé capable d'anticiper les résultats des marchés publics à travers l'identification du candidat gagnant, l'estimation du nombre des candidats et la définition des offres techniques de cahier des charges optimales.

L'investigation a mobilisé un arsenal technologique diversifié, combinant l'apprentissage automatique classique (Random Forest, Extra Trees, KNN), apprentissage profond (DNN, GRU) et techniques émergentes de traitement du langage naturel (RAG), permettant une évaluation comparative rigoureuse des performances algorithmiques face aux spécificités des données textuelles et catégorielles.

Les résultats obtenus valident l'hypothèse centrale selon laquelle l'IA peut révolutionner la gestion des appels d'offres en transformant des processus traditionnellement opaques et chronophages en systèmes transparents et efficaces, ouvrant des perspectives concrètes d'optimisation de l'allocation des ressources publiques et d'accélération des cycles décisionnels. Néanmoins, cette recherche ouvre des horizons prometteurs pour des travaux futurs qui pourraient enrichir considérablement les résultats obtenus.

Plusieurs perspectives d'amélioration se dessinent naturellement, notamment l'expansion de la base de données actuelle pour inclure des informations provenant de projets et de marchés publics issus de différentes communes et régions, ce qui permettrait d'améliorer significativement la capacité de généralisation des modèles et d'augmenter la précision des prédictions, l'amélioration des modèles utilisés est également cruciale pour améliorer la réactivité. L'exploitation de techniques d'intelligence artificielle plus avancées, telles que les modèles basés sur les Transformers ou les approches hybrides, offrirait également de meilleures performances en matière d'analyse et de prévision.

Par ailleurs, il serait particulièrement pertinent d'intégrer le module de prédiction intelligent au sein de l'application web dédiée à la gestion des marchés publics, permettant ainsi aux utilisateurs d'accéder aux analyses et recommandations directement dans leur environnement de travail quotidien.

Cette recherche s'inscrit dans une démarche de modernisation des pratiques administratives publiques, contribuant à l'émergence d'un nouveau paradigme de

Conclusion générale

gouvernance basé sur la donnée et l'intelligence artificielle, posant les fondements d'une transformation structurelle des processus d'approvisionnement public orientée vers une gestion plus rationnelle, transparente et efficiente des ressources collectives au service de l'intérêt général.

Bibliographies

Bibliographies

- [1] NetApp. What is Artificial Intelligence? [en ligne]. Disponible sur : <https://www.netapp.com/fr/artificial-intelligence/what-is-artificial-intelligence/> (Consulté le 27/04/2025).
- [2] DE MATTEIS, Ludovic, et al. "Introduction à l'apprentissage automatique." Culture Sciences de l'ingénieur (2022).
- [3] DataFranca. Apprentissage supervisé [en ligne]. Disponible sur : http://www.datafranca.org/wiki/Apprentissage_supervis%C3%A9 (Consulté le 27/04/2025).
- [4] Speiser, Jaime Lynn, et al. "A comparison of random forest variable selection methods for classification prediction modeling." Expert systems with applications 134 (2019): 93-101.
- [5] TOUZOT, Benoit. Machine Learning [en ligne]. Disponible sur : https://rpubs.com/Touzot_benoit/MachineLearning (Consulté le 02/05/2025).
- [6] ESRI. Fonctionnement de la classification et de la régression ExtraTrees [en ligne]. ArcGIS Pro. Disponible sur : <https://pro.arcgis.com/fr/pro-app/latest/tool-reference/geoai/how-extra-tree-classification-and-regression-works.htm#:~:text=L'algorithm%20ExtraTrees%2C%20comme%20,arbre%20avec%20des%20C3%A9chantillons%20uniques.> (Consulté le 27.04.2025).
- [7] AL-MOMANI, Bashar et al. Extra Tree Classifier Architecture [en ligne]. ResearchGate, 2021. Disponible sur : https://www.researchgate.net/figure/Extra-Tree-Classifer-Architecture_fig4_370242370 (Consulté le 02/05/2025).
- [8] IBM. What is the KNN algorithm? [en ligne]. Disponible sur : <https://www.ibm.com/think/topics/knn#:~:text=the%20KNN%20algorithm,What%20is%20the%20KNN%20algorithm%3F,of%20an%20individual%20data%20point.> (Consulté le 27.04.2025).
- [9] StudySmarter. Apprentissage non supervisé [en ligne]. Disponible sur : <https://www.studysmarter.fr/resumes/informatique/big-data/apprentissage-non-supervise/> (Consulté le 27/04/2025).

Bibliographies

- [10] Krebs, Julian. Le recalage robuste d'images médicales et la modélisation du mouvement basée sur l'apprentissage profond. Diss. Université Côte d'Azur, 2020.
- [11] Hamouda Djallel, « Intrusion Detection System for Cyber Security », Thèse de Master, Université 8 Mai 1945, Guelma, 2020.
- [12] DATASCIENTEST. Deep Neural Network: What Is It and How Is It Working? [en ligne]. Disponible sur : <https://datascientest.com/en/deep-neural-network-what-is-it-and-how-is-it-working> (Consulté le 27.04.2025).
- [13] VISO.AI. Deep Neural Network (DNN): 3 Popular Types [en ligne] Disponible sur : [https://viso.ai/deep-learning/deep-neural-network-three-popular-types/#:~:text=determines%20the%20weights,-,3%20Types%20of%20Deep%20Neural%20Networks%20\(DNN\),Recurrent%20Neural%20Networks%20\(RNN\)](https://viso.ai/deep-learning/deep-neural-network-three-popular-types/#:~:text=determines%20the%20weights,-,3%20Types%20of%20Deep%20Neural%20Networks%20(DNN),Recurrent%20Neural%20Networks%20(RNN)) (Consulté le 27.04.2025).
- [14] MARKETMUSE. Gated Recurrent Unit (GRU) Definition [en ligne]. MarketMuse Blog, Glossary. Disponible sur : <https://blog.marketmuse.com/glossary/gated-recurrent-unit-gru-definition/> (Consulté le 02/05/2025).
- [15] RAGHAVAN, Vyshnavi et al. The Architecture of basic Gated Recurrent Unit (GRU) [en ligne]. ResearchGate, 2020. Disponible sur : https://www.researchgate.net/figure/The-Architecture-of-basic-Gated-Recurrent-Unit-GRU_fig1_343002752 (Consulté le 02/05/2025).
- [16] IBM France. Think : Natural Language Processing [en ligne]. (sans date). Disponible sur : <https://www.ibm.com/fr-fr/think/topics/natural-language-processing> (consultée le 27.04.2025).
- [17] DATACAMP. How Transformers Work [en ligne], s.d., s.d. Disponible sur : <https://www.datacamp.com/tutorial/how-transformers-work> (Consulté le 02/05/2025).
- [18] LOUD TECHNOLOGY. BERT (Bidirectional Encoder Representations from Transformers) [en ligne]. Disponible sur : <https://loud-technology.com/programmation/definitions/bert-bidirectional-encoder-representations-from-transformers/> (Consulté le 27.04.2025).
- [19] COHESITY. Retrieval Augmented Generation (RAG) [en ligne]. Disponible sur : <https://www.cohesity.com/fr/glossary/retrieval-augmented-generation->

Bibliographies

[rag/#:~:text=RAG%20est%20l'abr%C3%A9viation%20de,des%20mod%C3%A8les%20d'IA%20g%C3%A9n%C3%A9rative.](#) (Consulté le 27.04.2025).

[20] SOLULAB. What is Retrieval Augmented Generation (RAG): Architectural Overview [en ligne]. Disponible sur : <https://www.solulab.com/what-is-retrieval-augmented-generation/> (Consulté le 27.04.2025).

[21] YANG, Y., HUANG, K., LI, C., ZHAO, W., ZHANG, R., JIANG, Y., WANG, Y. & ZHANG, L. Dense X Retrieval: What Retrieval Granularity Should We Use? [en ligne]. arXiv, 2022, arXiv:2212.08153. Disponible sur : <https://arxiv.org/abs/2212.08153> (Consulté le 27.04.2025).

[22] YANG, Joyce. The Fusion-In-Decoder (FID) model [en ligne]. ResearchGate, 2021. Figure issue de: Izacard and Grave, 2020. Disponible sur : https://www.researchgate.net/figure/The-Fusion-In-Decoder-FID-model-Figure-credit-Izacard-and-Grave-2020_fig24_357875438 (Consulté le 27.04.2025).

[23] KERKATOU, Raid Anis & SELLAHI, Abdelkadir. Disaster Vision: a computer vision system for disaster management. Mémoire de Master, Artificial Intelligence and its Applications (I2A). Mila : University Center Abd Elhafid Boussof, 2024.

[24] GREAT LEARNING. Label Encoding in Python: Simple Implementation Example [en ligne]. Disponible sur : <https://www.mygreatlearning.com/blog/label-encoding-in-python/#:~:text=Label%20encoding%20is%20a%20technique,only%20operate%20on%20numerical%20data.> (Consulté le 27.04.2025).

[25] MARQO AI. Introduction to Sentence Transformers [en ligne]. Disponible sur : <https://www.marqo.ai/course/introduction-to-sentence-transformers#:~:text=Sentence%20transformers%20address%20this%20limitation,accuracy%20and%20meaningful%20text%20analysis.> (Consulté le 27.04.2025).

[26] REIMERS, Nils & GUREVYCH, Iryna. SentenceTransformers : Pretrained Models [en ligne]. Disponible sur : https://www.sbert.net/docs/sentence_transformer/pretrained_models.html?utm_source (Consulté le 27.04.2025).

[27] HUGGING FACE. distiluse-base-multilingual-cased-v1 [en ligne]. SentenceTransformers. Disponible sur : https://huggingface.co/sentence-transformers/distiluse-base-multilingual-cased-v1?utm_source (Consulté le 27.04.2025).

Bibliographies

- [28] HUGGING FACE. paraphrase-mpnet-base-v2 [en ligne]. Sentence-Transformers. Disponible sur : <https://huggingface.co/sentence-transformers/paraphrase-mpnet-base-v2#sentence-transformersparaphrase-mpnet-base-v2> (Consulté le 27.04.2025).
- [29] DATALOOP AI. craig_paraphrase-minilm-l6-v2 [en ligne]. Model Library. Disponible sur : https://dataloop.ai/library/model/craig_paraphrase-minilm-l6-v2/#:~:text=Format,like%20clustering%20or%20semantic%20search. (Consulté le 27.04.2025).
- [30] DATALOOP AI. google_flan-t5-base [en ligne]. Model Library. Disponible sur : https://dataloop.ai/library/model/google_flan-t5-base/ (Consulté le 27.04.2025).
- [31] IBM France. Matrice de confusion [en ligne]. Think. Disponible sur : <https://www.ibm.com/fr-fr/think/topics/confusion-matrix> (Consulté le 27.04.2025).
- [32] TIOBE Organisation, « TIOBE Index », TIOBE, www.tiobe.com/tiobe-index/, juin 2024.
- [33] Peng, Liang, et al. "Answer again: Improving VQA with cascaded-answering model." IEEE Transactions on Knowledge and Data Engineering 34.4 (2020): 1644-1655.
- [34] Visual Studio Code: <https://code.visualstudio.com/docs>.