

République Algérienne Démocratique et Populaire

Université A. MIRA de Béjaia
Faculté des Sciences Exactes
Département de Recherche Opérationnelle



Mémoire présenté pour l'obtention du diplôme de master
Spécialité : Sciences de Données et Aide à la Décision

Implémentation d'un modèle machine learning pour la prévision des ventes de Cevital



Présenté par :
SADOUN Salim
SADOUNE Hacene

Sous la direction de :
Mme S. AMROUN
Mme L. DJERROUD

Défendu le 02/07/2024, devant le jury composé de :

Mr N. ZOUGAB	Professeur	Président de jury	UAMB - Bejaia
Mme Z. AOUDIA	M.A. classe/A	Examinatrice	UAMB - Bejaia
Mr A. LAOUR	M.A. classe/A	Examineur	UAMB - Bejaia
Mr A. IDIR	Data Analyst	Invité	Cevital

Année Universitaire 2023 2024

Remerciements

Avant tout, nous tenons à exprimer notre profonde gratitude au Bon Dieu, pour sa guidance et ses bénédictions tout au long de ce projet. Sans sa grâce, rien de tout cela n'aurait été possible.

Nous souhaitons également remercier chaleureusement nos deux encadrantes, Mme Sonia AMROUNE et Mme Louiza DJERROUD. Leur soutien constant, leurs conseils avisés et leur expertise ont été essentiels à la réalisation de ce travail. Leur patience et leur disponibilité nous ont beaucoup aidés tout au long de ce projet.

Nous tenons aussi à exprimer notre reconnaissance à M. OUNCERE Farouk, directeur du département des systèmes d'information chez Cevital, pour son accueil chaleureux et son soutien. Merci de nous avoir offert un environnement propice à la réalisation de ce projet.

Nous remercions sincèrement notre maître de stage, M. IDIR Anis, pour son encadrement attentif, ses encouragements constants et ses orientations précieuses. Votre expérience et vos conseils ont été d'une grande aide et ont grandement contribué à la réussite de ce projet.

Nous adressons également nos plus sincères remerciements au département de Recherche Opérationnelle ainsi qu'à nos enseignants pour leur soutien et leurs précieux enseignements qui ont enrichi notre parcours académique et professionnel.

Enfin, nous adressons nos plus profonds remerciements à nos parents, pour leur amour, leur soutien inconditionnel et leurs encouragements tout au long de notre parcours. Leur présence et leurs sacrifices ont été une source inestimable de motivation et d'inspiration.

Salim & Hacene

Table des matières

Liste des figures	5
Liste des abréviations	8
Introduction générale	11
1 Présentation de l'entreprise d'accueil "CEVITAL"	12
1.1 Introduction	12
1.2 Historique de Cevital	12
1.3 Présentation de Cevital	13
1.4 Valeurs de l'Entreprise	13
1.5 Présentation de L'Entreprise Cevital Agro-Industrie	14
1.5.1 La répartition de Cevital Agro-industrie	14
1.6 Les produits de cevital	16
1.7 Les Partenaires et Clients de Cevital	19
1.8 Engagement de Cevital Food envers la Qualité et l'Innovation	19
1.9 L'Organisation Générale des Composantes	20
1.10 Les Missions des Directions	20
1.11 L'organigramme de l'entreprise Cevital Agro-Industrie	22
1.12 Répartition des Effectifs	23
1.13 Impact Économique et Social	24
1.14 Protection de l'environnement	24
1.15 Conclusion	24

2	Fondements et Applications d'apprentissage automatique dans la Prédiction des Ventes	25
	I. Introduction à l'apprentissage automatique	25
2.1	Introduction	25
2.2	Définition de l'apprentissage automatique	25
2.3	Objectifs de l'apprentissage automatique	26
2.4	Types de l'apprentissage automatique	26
2.4.1	Apprentissage Supervisé	27
2.4.1.1	Classification	27
2.4.1.1.1	Modèles de classification	28
	Naïf Bayésien	28
	k-plus proches voisins	28
	Arbres de décision	29
	Machine à support de vecteur (SVM)	29
2.4.1.1.2	Métriques d'évaluation	30
2.4.1.2	Régression	31
2.4.1.2.1	Modèles de régression	31
	Régression linéaire simple	31
	Régression linéaire multiple	32
	Régression polynomial	32
	Forêts aléatoires	33
	Régression à support de vecteur(SVR)	34
	Réseaux de Neurones artificiels	34
2.4.1.2.2	Métriques d'évaluation	35
	Racine carré d'erreur quadratique moyenne (RMSE)	35
	R2 score	35
2.4.2	Apprentissage non supervisé	36
2.4.2.1	Clustering	36
2.4.3	Apprentissage par renforcement	37
2.5	Domaines d'application	37
	II. Prédiction des Ventes en Utilisant l'apprentissage automatique	37
2.6	Qu'est-ce que la prédiction des ventes ?	38
2.7	L'importance de la prédiction des ventes	38
2.8	Défis de la prédiction des ventes	38
2.8.1	Fiabilité des données	38
2.8.2	Volatilité du marché	38
2.9	Utilisation du machine learning pour prévoir les ventes	39
2.9.1	Apprentissage automatique comme force de prédiction	39
2.9.2	Avantages de l'apprentissage automatique dans la prédiction des ventes	39

TABLE DES MATIÈRES

2.10	Etapes pour mettre en place un modèle d'apprentissage automatique	40
2.10.1	Collecte et préparation des données	40
2.10.1.1	Sources de données	40
2.10.1.2	Nettoyage et prétraitement	40
2.10.1.3	Analyse exploratoire des données (AED)	40
2.10.2	Ingénierie des fonctionnalités	40
2.10.3	Choix du modèle de machine learning	41
2.10.4	Entraînement et évaluation de modèle	41
2.10.5	Déploiement	41
2.11	Problématique :	41
2.12	Conclusion	42
3	Implémentation	43
	I. Prévission des ventes pour l'année 2024	43
3.1	Outils et environnements de développement	43
3.1.1	Langage de programmation : "Python"	43
3.1.2	Environnement de développement	44
3.1.3	Bibliothèques utilisées	44
3.2	Collecte des données	44
3.3	Préparation des données	45
3.3.1	Nettoyage des données	47
3.3.2	Encodage et Standarisation des variables	50
3.4	Analyse exploratoire des données (AED)	51
3.5	Modélisation	57
3.5.1	Séparation des données	57
3.5.2	Choix de modèle	57
3.5.3	Entraînement de modèle	57
3.5.4	Évaluation et optimisation du modèle avec GridSearchCV	58
3.5.5	Évaluation du modèle sur l'ensemble de test	59
3.5.6	Prévisions des ventes pour l'année 2024	59
3.5.7	Etude comparative entre les prédictions et les quantités réelles 2024	61
	II. Création d'un tableau de bord itératif : Prévisions des ventes 2024	62
3.6	Outil de Visualisation : "Power BI"	62
3.7	Construction de tableau de bord	63
	Conclusion générale	70
	Résumé	72

Table des figures

1.1	Organigramme de l'entreprise Cevital	22
1.2	L'effectif de cevital et sa répartition	23
2.1	La relation entre ML, IA et deep learning	26
2.2	Types d'apprentissage automatique	27
2.3	Types d'apprentissage supervisé	27
2.4	Exemple de l'algorithme KNN	28
2.5	Schéma d'un arbre de décision	29
2.6	Exemple de classification par svm	29
2.7	Matrice de confusion	30
2.8	Exemple de régression linéaire	32
2.9	Exemple de régression polynomiale	33
2.10	Algorithme de la forêt aléatoire	34
2.11	L'approche des ANN	35
2.12	Méthode de clustering	36
2.13	Processus d'apprentissage par renforcement	37
3.1	Importation des bibliothèques nécessaires	44
3.2	Extrait de quelques lignes de notre ensemble de données	45
3.3	Résultats des tests statistiques appliqués sur notre variables	46
3.4	Les valeurs négatives des colonnes avant et après la suppression.	47
3.5	Comparaison des données avant et après la gestion des valeurs manquantes	48
3.6	Test statistiques préliminaires	49
3.7	Boxplot de la distribution des variables numériques avant et après le traitement des valeurs aberrantes	49
3.8	Résultats de l'encodage et de la standarisation des variables	50
3.9	Analyse des corrélations	51
3.10	Statistiques descriptives	52

TABLE DES FIGURES

3.11 Répartition des produits les plus vendus	53
3.12 Tendances historiques des ventes pour chaque produit	54
3.13 Évaluation mensuelle des performances de vente 2023	55
3.14 Répartition des Quantités vendues par Ville	56
3.15 Les clients les plus importants en termes de quantités vendues	56
3.16 Entraînement de modèle	58
3.17 Les résultats d'évaluation avec la validation croisée	58
3.18 Les résultats d'évaluation du Modèle sur l'Ensemble de Test	59
3.19 Prévisions des ventes mensuelles pour l'année 2024	60
3.20 Etude comparative entre les prédictions et les quantités réelles 2024	61
3.21 La source de données (Excel)	62
3.22 Table de données	63
3.23 Affichage des données	63
3.24 Table des visualisations	64
3.25 Evaluation des prévisions mensuelles	65
3.26 Répartition des quantités prédites pour chaque produit pour l'année 2024	65
3.27 Répartition des quantités prédites par famille de produits pour l'année 2024"	66
3.28 Répartition des quantités prédites par Sous-famille de produits pour l'année 2024	67
3.29 Répartition géographique des quantités prédites par ville	68
3.30 Tableau de bord itératifs des prévisions de ventes 2024	69

Liste des abréviations

ML Machine Learning

IA Intelligence artificiel

KNN Le k plus proche voisin (*K Nearest Neighbors*)

SVM Machine à support vecteur

TP Vrais Positifs (*True Positives*)

TN Vrais Négatifs (*True Negatives*)

FP Faux Positifs (*False Positives*)

FN Faux Négatifs (*False Negatives*)

SVR Régression à support de vecteur

ANN Réseaux de neurones artificiels

MSE Erreur quadratique moyenne

RMSE Racine carré d'erreur quadratique moyenne

MAE Erreur Absolue Moyenne

DB Indice de Davies Bouldin

CH L'indice Calinski-Harabasz

ARI Indice de Rand ajust

MCTS Recherche d'arbres à Monte-Carlo

AED Analyse Exploratoire des Données

QT Tonne Quantité en tonnes

CA TTC Chiffre d'Affaires Toutes Taxes Comprises

CA HT Chiffre d'Affaires Hors Taxes

CA HT DZD Chiffre d'Affaires Hors Taxes en Dinars Algériens

TABLE DES FIGURES

QTE_PLT Quantité en Palettes

QT US Quantité en Unités de Stock

Num_Annee Numéro d'année

NOM_MOIS Nom de mois

BI Business Intelligence

Introduction Générale

Dans un secteur commercial, la capacité à prédire avec précision les ventes futures est devenue un atout stratégique crucial. Cette compétence permet aux entreprises la gestion efficace de la production et la prise de décision stratégique. Traditionnellement, les entreprises font appel à des méthodes statistiques classique telles que la moyenne mobile et les modèles de régression, bien que ces méthode aient été utiles, elles ont souvent des limites.

Avec l'émergence des technologies de d'information et l'accès croissant à des données massives, l'intelligence artificielle et notamment l'apprentissage automatique s'est imposé comme une solution efficace pour améliorer la précision des prévisions des ventes.

L'apprentissage automatique est une branche de l'intelligence artificielle permet aux ordinateurs d'apprendre automatiquement à partir des données sans être explicitement programmés pour chaque tâche. Les algorithmes d'apprentissage automatique, comme les machines à vecteurs de support et les forêts aléatoires, ont prouvé leur efficacité dans différents domaines, y compris la prévision des ventes. Ces techniques offrent la possibilité de traiter des quantités massive de données, en prenant compte de multiple facteurs influencent les ventes, tels que les tendances saisonnières, les promotions et les évolutions du comportement des consommateurs.

La mise en place de modèles de l'apprentissage automatique pour prédire les ventes nécessite plusieurs étapes essentielles. En premier lieu, la collecte et la préparation des données, où les données historiques sont regroupées et analysées. Puis, la sélection et l'entraînement du modèle, où divers algorithmes du machine learning sont évalués et ajustés afin d'améliorer leurs performances. Enfin, l'évaluation et la validation du modèle sont effectuées en évaluant les performances du modèle à partir de données de test afin de garantir sa précision et sa robustesse.

TABLE DES FIGURES

L'objectif principale est de prédire les ventes future de l'entreprise Cevital. Ce mémoire est organisé de la manière suivante :

Le premier chapitre, offre une vue d'ensemble de Cevital. Nous examinerons son histoire, sa structure organisationnelle, ses principales activité et ses accomplissement.

Dans le deuxième chapitre, nous allons abordé deux aspects essentiels : Introduction à l'apprentissage automatique et leur utilisation dans la prévision des ventes. Dont, nous aborderons les concepts fondamentaux de l'apprentissage automatique, ses différents types d'algorithmes (supervisé, non supervisé et apprentissage par renforcement) et ses objectifs. Nous étudierons l'importance de la prévision des ventes dans la gestion d'une entreprise, les défis courants ainsi que l'utilisation de machine learning pour la prévision.

Enfin, le dernier chapitre décrira le processus d'implémentation d'un modèle d'apprentissage automatique pour la prévision des ventes de l'année 2024 de Cevital. Noux exposerons en détail les différentes étapes, depuis la collecte et la préparation des données jusqu'à l'évaluation du modèle. L'utilisation de Python pour le développement d'un modèle approprié et Power BI pour la visualisation des résultats sera également abordée.

Ce mémoire se termine par une conclusion générale qui résume les principaux résultats obtenus concernant l'application de l'apprentissage automatique dans la prévision des ventes.

Chapitre 1

Présentation de l'entreprise d'accueil "CEVITAL"

1.1 Introduction

Cevital est une entreprise privée algérienne fondée en 1998 par Issad Rebrab. En l'espace de quelques décennies, elle s'est imposée comme un acteur incontournable de l'économie algérienne et internationale grâce à une stratégie de diversification et d'innovation constante. Ce chapitre se propose de présenter de manière exhaustive l'entreprise Cevital, en mettant en lumière son histoire, ses secteurs d'activité, ses valeurs, son organisation interne, et son impact économique et social. Nous aborderons d'abord l'historique de Cevital pour comprendre les étapes clés de son développement. Ensuite, nous détaillerons les différents secteurs dans lesquels l'entreprise opère, en insistant sur ses principales unités de production et leur capacité. Nous discuterons également des valeurs et de la philosophie qui guident les actions de Cevital, ainsi que de son engagement en matière d'éthique professionnelle. Par ailleurs, nous présenterons l'organisation interne de l'entreprise, incluant la répartition des effectifs et la structure des différentes directions. Enfin, nous examinerons l'impact économique et social de Cevital en Algérie, en soulignant ses contributions significatives au développement national. Ce chapitre vise non seulement à informer le lecteur sur la structure et les activités de Cevital, mais aussi à illustrer comment une entreprise peut se développer et prospérer tout en respectant des principes éthiques et en contribuant positivement à la société.

1.2 Historique de Cevital

Depuis sa création en 1998, Cevital a connu une croissance remarquable, débutant avec des investissements dans l'agroalimentaire avant de diversifier ses activités dans d'autres

secteurs tels que la grande distribution, l'électronique et la sidérurgie. Chaque étape de son développement a été marquée par des acquisitions stratégiques et des expansions industrielles.

- 1998 : Lancement de l'huile alimentaire Elio ;
- 2001 : Lancement de la margarine Fleurial et Matina et du beurre tendre Gourmand ;
- 2002 : Lancement de Skor 1kg ;
- 2006 : Lancement des boissons aux jus Tchina ;
- 2007 : Lancement de l'eau minérale Lalla Khedidja ;
- 2010 : Lancement du sucre liquide ;
- 2013 : Lancement de la chaux calcique ;
- 2015 : Lancement du gaz CO2 alimentaire ;
- 2016 : Lancement du sucre roux ;
- 2018 : Lancement des sauces et condiments ;
- 2020 : Lancement de la production plasturgie ;
- 2023 : La trituration des graines oléagineuses.

Ces initiatives ont contribué à renforcer la présence de Cevital sur le marché et à diversifier ses produits et services [1].

1.3 Présentation de Cevital

Cevital est une Société par Actions au capital privé de 68,760 milliards de DA. Elle a été créée en mai 1998. Elle est implantée à l'extrême Est du port de Béjaia. Elle est l'un des fleurons de l'industrie agroalimentaire en Algérie, constituée de plusieurs unités de production équipées de la dernière technologie et poursuit son développement par divers projets en cours de réalisation. Son expansion et son développement durant les 5 dernières années, font d'elle un important pourvoyeur d'emplois et de richesses [1].

1.4 Valeurs de l'Entreprise

Un référentiel des valeurs a été mis en œuvre. Leurs applications doivent être le reflet d'un comportement exemplaire.

1. **Écoute et Respect ;**
2. **Intégrité et Transparence ;**
3. **Solidarité et Esprit d'équipe ;**

4. **Initiative et Persévérance** ;

5. **Courage et Engagement de performance.**

Ces valeurs sont fondées sur le respect mutuel et sont essentielles pour maintenir un environnement de travail exemplaire [1].

1.5 Présentation de L'Entreprise Cevital Agro-Industrie



Créée en 1998 et implantée au sein du port de Bejaia, Cevital Agro-industrie dispose de plusieurs unités de production ultramodernes de sucre, corps gras, eau minérale, boissons et sauces. Elle couvre les besoins nationaux et a permis de faire passer l'Algérie du stade d'importateur à celui d'exportateur pour les huiles, les margarines et le sucre. Ses produits se vendent dans plusieurs pays, notamment en Europe, au Maghreb, au Moyen Orient et en Afrique de l'Ouest.

1.5.1 La répartition de Cevital Agro-industrie

À l'arrière-port de Béjaia à 200 ML du quai : Ce terrain à l'origine marécageux et inconstructible a été récupéré en partie d'une décharge publique, viabilisé avec la dernière

CHAPITRE 1. PRÉSENTATION DE L'ENTREPRISE D'ACCEUIL "CEVITAL"

technologie de consolidation des sols par le système de colonnes ballastées (337 KM de colonnes ballastées de 18 ML chacune ont été réalisées) ainsi qu'une partie à gagner sur la mer.

Les raffineries et unités de production :

- 2 raffineries de sucre 3000 et 3500 Tn ;
- 1 unité de sucre liquide et une unité de production de sucre roux ;
- 2 unité de conditionnement de sucre ;
- 1 raffinerie d'huile ;
- 1 unité de conditionnement de huile ;
- 1 margarinerie ;
- 1 unité d'eau minérale et gazéifiée ;
- 1 unité de fabrication et de conditionnement de boissons fruitées et de production de conserves et confitures ;
- 1 unité de production des sauces ;
- 1 unité de fabrication de chaux calcinée et CO₂ ; Silos portuaires ainsi qu'un terminal de déchargement portuaire d'une capacité de 2000 tonnes/jour.

À Béjaia

Nous avons entrepris la construction des installations suivantes :

- Raffinerie Huile ;
- Margarinerie ;
- Silos portuaires ;
- Raffinerie de sucre.

1. Pôle corps gras : Solide

Unité	Capacité	Unité	Observations
Silos de stockage Huile Brut	61 900	Tn	17 bacs de stockages
Raffinerie D'huile	2 100	Tn/jour	
Silos de stockage Huile Raffiné	6 620	Tn	19 bacs de stockages
Conditionnement d'huile	2 500	Tn/jour	

1. Pôle corps gras : Solide

Unité	Capacité	Unité
Margarinerie	600	Tn/jour

1. Pôle de sucre

CHAPITRE 1. PRÉSENTATION DE L'ENTREPRISE D'ACCEUIL "CEVITAL"

Unité	Capacité	Unité	Observations
Hangar de stockage MP sucre	200 000	Tn	02 Hangars, (50 ktn, 150 ktn)
Raffinerie sucre	6 500	Tn/jour	02 Raffinerie (3000 tn, 3500 tn)
Silos de stockage PF	80 000	Tn	01 silo
	11 800	Tn	04 silos de 2950 tn
Conditionnement de sucre	10 000	Tn/jour	02 unités de conditionnements
Raffinerie sucre Liquide	600	Tn/jour	01 unité
Silos de stockage PF	1 500	Tn	05 bacs de 300 tn

À El Kseur

Une unité de production de jus de fruits Cojek a été rachetée par le groupe Cevital dans le cadre de la privatisation des entreprises publiques algériennes en novembre 2006. Un immense plan d'investissement a été consenti visant à moderniser l'outil de production de jus de fruits Cojek. Sa capacité de production est de 14 400 T par an. Le plan de développement de cette unité portera à 150 000 T/an en 2010.

1. Usine de production jus de fruits Tchina & de sauces

Unité	Capacité	Unité
Boissons	1 130	Tn/jour
CAT (confitures)	66	Tn/jour
Sauce	100	Tn/jour

À Tizi Ouzou

À Agouni Gueghrane : au cœur du massif montagneux du Djurdjura qui culmine à plus de 2300 mètres :

— L'Unité d'Eau Minérale Lalla Khedidja a été inaugurée en juin 2007 [1].

Unité	Capacité	Unité
Eau	2 688	Tn/jour

1.6 Les produits de cevital

Le Complexe Agro-alimentaire est composé de plusieurs unités de production :

1. Huiles Végétales ;
2. Margarinerie et graisses végétales ;
3. Sucre blanc ;

4. Sucre liquide ;
5. Silos portuaires ;
6. Boissons.

Huiles Végétales

Les huiles de table : elles sont connues sous les appellations suivantes :
Fleurial plus : 100% tournesol sans cholestérol, riche en vitamine (A, D, E)

(Elio et Fridor) : se sont des huiles 100% végétales sans cholestérol, contiennent de la vitamine E. Elles sont issues essentiellement de la graine de tournesol, soja et de palme, conditionnées dans des bouteilles de diverses contenances allant de 1 à 5 litres, après qu'elles aient subi plusieurs étapes de raffinage et d'analyse.

- Capacité de production : 570 000 tonnes/an ;
- Part du marché national : 70% ;
- Exportations vers le Maghreb et le Moyen-Orient, en projet pour l'Europe.



Margarinerie et graisses végétales

Cevital produit une gamme variée de margarine riche en vitamines A, D, E. Certaines margarines sont destinées à la consommation directe telle que Matina, Rania, le beurre gourmand et Fleurial, d'autres sont spécialement produites pour les besoins de la pâtisserie moderne ou traditionnelle, à l'exemple de la Parisienne et MEDINA « SMEN ».

- Capacité de production : 180 000 tonnes/an ;
- Part du marché national : 30% ;
- Exportations : Europe, Maghreb et Moyen-Orient. [1]

Sucre Blanc

Le sucre blanc est issu du raffinage du sucre roux de canne riche en saccharose. Le sucre raffiné est conditionné dans des sachets de 50 kg et aussi commercialisé en morceaux dans des boîtes de 1 kg. Cevital produit aussi du sucre liquide pour les besoins de l'industrie agroalimentaire et plus précisément pour les producteurs des boissons gazeuses.



- Entrée en production : 2ème semestre 2009 ;
- Capacité de production : 1 800 000 tonnes/an ;
- Part du marché national : 30% ;
- Exportations vers le Maghreb et le Moyen-Orient.

Silose portuaires

La manutention du sucre est assurée par des installations performantes :

- 02 déchargeurs automatiques de bateaux de 300 T/h chacun ;
- 07 trémies alimentées par gravité à partir des silos ;
- 02 lignes d'ensachage automatiques pour la mise en sac de 50 Kg ;
- 01 installation d'ensachage en sacs de 1 Kg ;
- 10 cellules de stockage d'une capacité totale de 110 000 T de sucre en vrac ;
- 01 dépôt couvert de 10 000 T pour le stockage de sucre en sac de 50 Kg.



Boissons

L'unité de production Cojek, spécialisée dans les jus de fruits et boissons, a bénéficié d'importants investissements pour moderniser ses équipements et augmenter sa capacité de production à 150 000 tonnes par an en 2010. Elle produit les jus de fruits sous la marque Vitajus.

- Capacité de production : 150 000 T/an en 2010 ;
- Part du marché national : 30% ;
- Exportations : en projet.



[1]

1.7 Les Partenaires et Clients de Cevital

Cevital Food sert une clientèle variée, incluant entreprises et particuliers, tant au niveau national qu'international. Ses principaux clients sont :

- Les grandes surfaces et supermarchés ;
- Les industries agroalimentaires ;
- Les petites et moyennes entreprises ;
- Les collectivités et institutions ;
- Les consommateurs finaux.

L'entreprise s'engage à fournir des produits de haute qualité et un service après-vente performant, répondant aux besoins et attentes spécifiques de chaque client.

1.8 Engagement de Cevital Food envers la Qualité et l'Innovation

La devise de Cevital Food est : **Qualité, Innovation, Respect de l'Environnement**. Ces valeurs sont intégrées dans chaque aspect de l'activité, de la production à la commercialisation et au service client. Cevital Food accorde une grande importance au contrôle de qualité, illustrée par l'engagement de la direction dans le processus de certification ISO 22000 version 2005. Toutes les unités de production disposent de laboratoires de microbiologie et de contrôle de qualité, équipés d'outils d'analyse performants.



1.9 L'Organisation Générale des Composantes

Cevital Food est organisée en différentes composantes pour assurer une gestion efficace et une production optimale. Les principales composantes sont :

- Direction Générale ;
- Direction des Ressources Humaines ;
- Direction de la Production ;
- Direction de la Qualité ;
- Direction Commerciale ;
- Direction Logistique ;
- Direction Financière.

Chaque direction joue un rôle crucial dans le fonctionnement global de l'entreprise et travaille en étroite collaboration avec les autres départements pour atteindre les objectifs de l'entreprise.

1.10 Les Missions des Directions

Direction Générale

La Direction Générale est responsable de la stratégie globale de l'entreprise. Elle définit les grandes orientations, prend les décisions stratégiques et supervise l'ensemble des activités de l'entreprise.

- Élaboration de la stratégie d'entreprise ;
- Prise de décisions stratégiques ;
- Supervision des opérations de l'entreprise ;
- Représentation de l'entreprise auprès des parties prenantes.

Direction des Ressources Humaines

La Direction des Ressources Humaines gère les relations humaines et les aspects administratifs des employés. Elle veille au recrutement, à la formation et au bien-être des salariés.

- Recrutement et intégration des nouveaux employés ;
- Gestion des carrières et des compétences ;
- Formation et développement des compétences ;
- Gestion des relations sociales et du climat de travail.

Direction de la Production

La Direction de la Production est responsable de la planification, de l'organisation et du contrôle des processus de production. Elle veille à l'optimisation des ressources et à la qualité des produits fabriqués.

- Planification de la production ;
- Gestion des unités de production ;
- Contrôle de la qualité des produits ;
- Optimisation des processus de production.

Direction de la Qualité

La Direction de la Qualité assure que les produits et services de l'entreprise répondent aux standards de qualité. Elle met en place des procédures de contrôle et d'amélioration continue.

- Définition des standards de qualité ;
- Mise en place de procédures de contrôle qualité ;
- Suivi et amélioration continue de la qualité ;
- Gestion des certifications et des audits qualité.

Direction Commerciale

La Direction Commerciale gère les relations avec les clients et les activités de vente. Elle développe les stratégies commerciales et assure la satisfaction des clients.

- Développement de la stratégie commerciale ;
- Gestion des relations clients ;
- Suivi des ventes et des performances commerciales ;
- Analyse du marché et des opportunités de croissance.

Direction Logistique

La Direction Logistique est en charge de la gestion des flux de matières et de produits. Elle veille à l'efficacité des approvisionnements, des stocks et des livraisons.

- Gestion des approvisionnements ;
- Gestion des stocks ;
- Organisation des livraisons ;
- Optimisation des coûts logistiques.

Direction Financière

La Direction Financière gère les aspects financiers de l'entreprise. Elle s'assure de la santé financière et de la rentabilité de l'entreprise.

- Élaboration des budgets et des prévisions financières ;
- Suivi de la trésorerie et des dépenses ;
- Analyse financière et gestion des risques ;
- Gestion des relations avec les investisseurs et les banques [1].

1.11 L'organigramme de l'entreprise Cevital Agro-Industrie

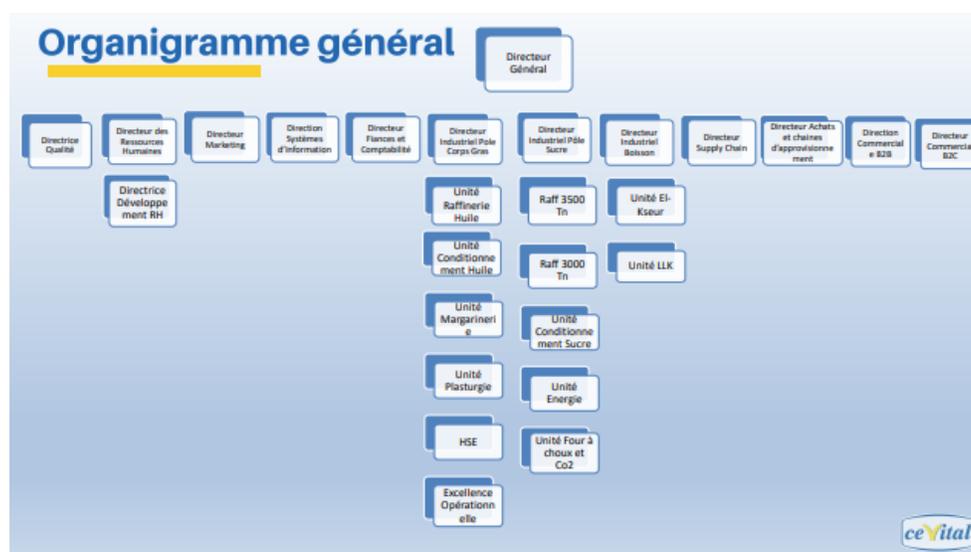


FIGURE 1.1 – Organigramme de l'entreprise Cevital

La représentation graphique montre une structure organisationnelle bien définie, avec la Direction Générale au sommet, responsable de la vision globale et des décisions stratégiques. En dessous se trouvent des directions clés : Finances & Comptabilité, Marketing, Stratégie & Planification, Ressources Humaines, Qualité, Hygiène, Sécurité & Environnement (QHSE), et Exploitations, chacune jouant un rôle essentiel dans le succès de l'entreprise.

L'organigramme inclut également notre position au sein de la Direction des systèmes d'informations pendant notre stage chez Cevital, direction cruciale pour le développement technologique et l'innovation numérique de l'entreprise.

1.12 Répartition des Effectifs

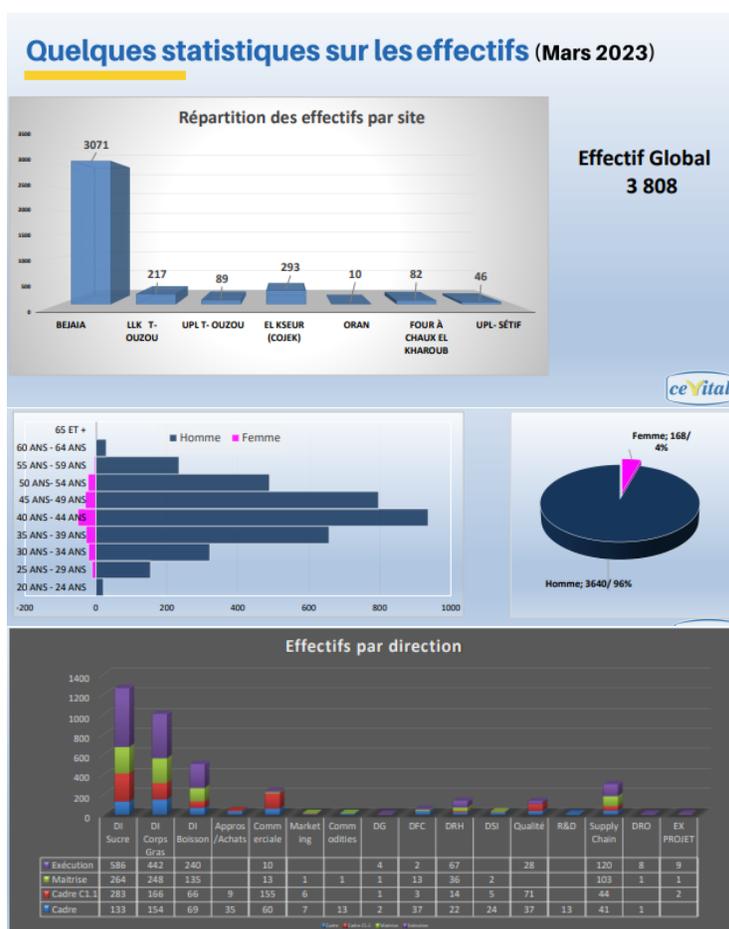


FIGURE 1.2 – L’effectif de cevital et sa répartition

Les effectifs de Cevital sont répartis de manière équilibrée entre les sites et les directions de l’entreprise. Le site de Bejaia compte 3071 employés et celui de Tizi Ouzou en a 217. La majorité des employés sont des hommes (96%) avec une distribution homogène à travers les différentes tranches d’âge. L’effectif total de Cevital Agro-industrie est de 3808 employés [1].

1.13 Impact Économique et Social

Cevital joue un rôle crucial dans le développement économique de l'Algérie, créant des emplois et des opportunités de croissance. L'entreprise est également engagée dans plusieurs initiatives de responsabilité sociale et environnementale, contribuant ainsi au bien-être de la communauté et à la protection de l'environnement.

1.14 Protection de l'environnement

Cevital Food s'engage à protéger l'environnement et à promouvoir des pratiques durables. Les actions mises en place comprennent :

- Réduction des déchets et recyclage ;
- Utilisation de technologies propres ;
- Sensibilisation des employés à l'environnement ;
- Participation à des projets écologiques locaux [1].

1.15 Conclusion

En résumé, Cevital est un acteur économique majeur en Algérie, avec une forte présence dans plusieurs secteurs d'activité. Son engagement envers l'innovation, la création de richesse, et la responsabilité sociale en fait une entreprise exemplaire et un pilier de l'économie algérienne. Les perspectives futures de Cevital sont prometteuses, avec des opportunités de croissance et d'expansion continue.

Chapitre 2

Fondements et Applications d'apprentissage automatique dans la Prévision des Ventes

I. Introduction à l'apprentissage automatique

2.1 Introduction

L'apprentissage automatique, une innovation en science des données, transforme la façon en que les entreprises anticipent leurs ventes, permettant une prévision précise de la demande, un ajustement efficace, une amélioration de la satisfaction client.

2.2 Définition de l'apprentissage automatique

Apprentissage automatique (en anglais : machine learning (ML)), apprentissage artificiel ou apprentissage statistique est un domaine de l'intelligence artificielle qui utilise des méthodes mathématiques et statistiques pour donner aux ordinateurs la capacité d'apprendre à partir des données.

En voici une définition plus technique :

«Étant donné une tâche T et une mesure de performance P , on dit qu'un programme informatique apprend à partir d'une expérience E si les résultats obtenus sur T , mesurés par P , s'améliorent avec l'expérience E .[2]

Les algorithmes d'apprentissage automatique sont utilisés pour automatiser et exécuter n'importe quelle tâche basée sur des règles ou des modèles définis par des données. L'ap-

l'apprentissage automatique peut être mis à l'échelle pour gérer des problèmes techniques plus étendus tels que la détection d'images pour les voitures autonomes, la prévision des ventes et la résolution de nombreux autres problèmes.

La figure(2.1) montre la relation entre ML, IA et deep learning :

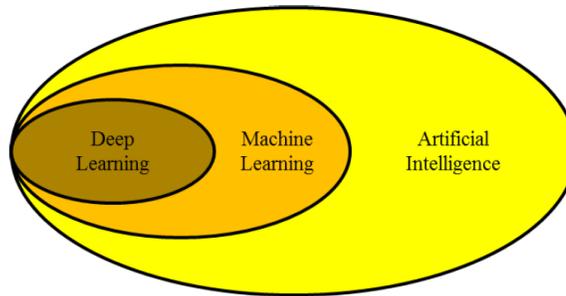


FIGURE 2.1 – La relation entre ML, IA et deep learning

2.3 Objectifs de l'apprentissage automatique

Les objectifs d'apprentissage automatique sont variés et dépendent des applications spécifiques, mais on peut identifier plusieurs objectifs :

- La prédiction : Créer des patterns (modèles) capables d'établir des prédictions précises sur de nouvelles données.
- La reconnaissance de forme : identifier des motifs et des structures complexes dans les données. Nous pouvons citer la reconnaissance d'images, de la parole, etc.
- L'optimisation : Permettre aux systèmes de prendre des décisions plus intelligentes.
- La recommandation : L'analyse des préférences et des comportements des utilisateurs pour leur suggérer des produits, des services ou des contenus pertinents.

Les divers objectifs des différents modèles de Machine Learning illustrent la polyvalence de cette technologie.

2.4 Types de l'apprentissage automatique

Il existe plusieurs catégories de problèmes distinctes dans le domaine d'apprentissage automatique. Ils varient selon le type de données fournies et le type de conclusions à en tirer. Trois types standard sont décrits ci-dessous, afin d'établir une certaine notation et terminologie :

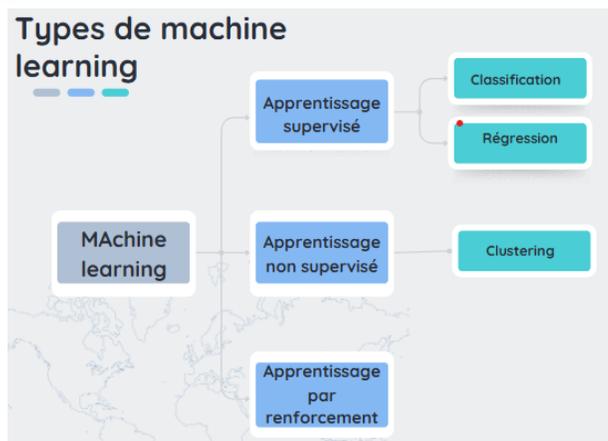


FIGURE 2.2 – Types d'apprentissage automatique

2.4.1 Apprentissage Supervisé

L'apprentissage supervisé consiste à utiliser un ensemble de données historiques pour prédire des événements futurs statistiquement probables, c'est-à-dire à partir d'événements déjà prédits. L'objectif est d'entraîner le modèle choisi pour qu'il puisse prédire correctement la sortie pour des entrées étiquetées. L'apprentissage supervisé est de deux types ; classification et régression.

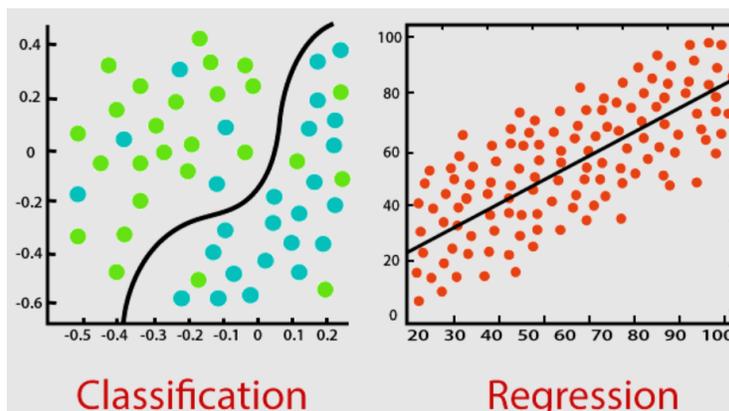


FIGURE 2.3 – Types d'apprentissage supervisé

2.4.1.1 Classification

Elle consiste à attribuer une classe (étiquette ou label) à une observation en fonction de ses caractéristiques. Les méthodes de classification peuvent être basées sur des hypothèses probabilistes (le naïf bayésien), des notions de proximité (les k plus proches voisins) ou des

recherches dans des espaces d'hypothèses (l'arbre de décision). Cet algorithme est utilisé pour prédire les valeurs discrètes telles que masculine|féminine, vrai|faux, spam|pas de spam, etc.

2.4.1.1.1 Modèles de classification

a- Bayésien naïf

Est la méthode de classification qui classe un ensemble d'observations selon les règles déterminées par l'algorithme lui-même. Le classificateur naïf bayésien est basé sur le théorème de "Bayes" avec une forte hypothèse de dépendance conditionnelle entre les caractéristiques. Dans la phase d'apprentissage, l'algorithme développe ses règles de classification sur cet ensemble de données et les applique à la classification de l'ensemble de données prédit dans la deuxième étape [3].

Le théorème est donné par la formule suivante :

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (2.1)$$

Où :

A et B sont des événements ;

$P(A)$: est la probabilité d'observer l'événement A ;

$P(B)$: est la probabilité d'observer l'événement B ;

$P(A|B)$: est la probabilité conditionnelle d'observer A, sachant qu'un autre événement B de probabilité non nulle s'est réalisé.

b- k-plus proches voisins

La méthode du k-plus proche voisin en anglais(K-nearest neighbors (kNN)) est une méthode non paramétrique où une nouvelle observation est classée dans la classe d'appartenance de l'observation de l'échantillon d'apprentissage qui lui est la plus proche en fonction des covariables utilisées.

La figure (2.4) montre un exemple de l'algorithme KNN, où un échantillon est assigné à la classe la plus fréquente parmi ses k voisins les plus proches :

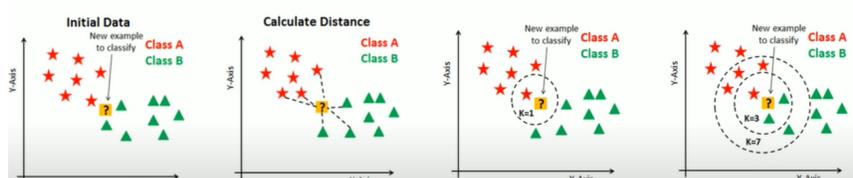


FIGURE 2.4 – Exemple de l'algorithme KNN

c- Arbres de décision

Un arbre de décision est un outil visuel permettant d'expliquer une variable cible à partir d'autres variables dites explicatives.

Les arbres de décision peuvent gérer à la fois des variables continues et discrètes, fournissant des indications claires pour la prédiction et la classification sans trop de calculs. Ils sont faciles à comprendre et à interpréter, ce qui permet d'assurer de meilleures approximations, quelle que soit la complexité des données. Ils ont une structure hiérarchique arborescente comprenant un nœud racine, des branches, des nœuds internes et des nœuds floraux [4].

Considérons l'exemple simple représenté dans la figure (2.5) :

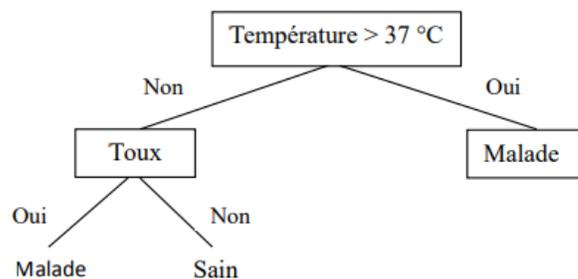


FIGURE 2.5 – Schéma d'un arbre de décision

d- Machine à support de vecteur (SVM)

Le SVM est un modèle d'apprentissage automatique supervisé qui est principalement utilisé pour les classifications (mais il peut aussi être utilisé pour la régression!). Il vise à déterminer l'hyperplan optimal dans un espace à N dimensions qui peut séparer les points de données de différentes classes dans l'espace des caractéristiques [5].

La figure (2.6) illustre un exemple de classification par svm :

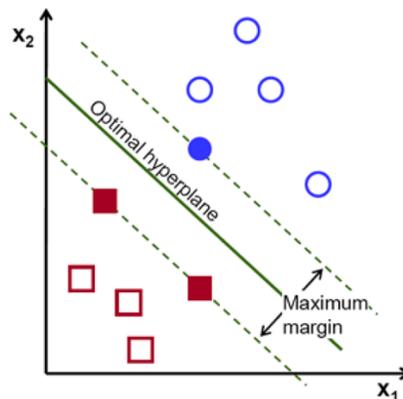


FIGURE 2.6 – Exemple de classification par svm

2.4.1.1.2 Métriques d'évaluation

a- Matrice de confusion

Est une évaluation des performances pour les classificateurs binaires ou multi-classes qui s'appuient sur machine learning ou deep learning. Elle est présentée sous forme de tableau, chaque cellule représente le nombre de fois où le modèle peut prédire correctement ou mal la classe. La figure (2.7) montre la matrice de confusion qui représente le calcul de TP, TN, FP et FN [6] :

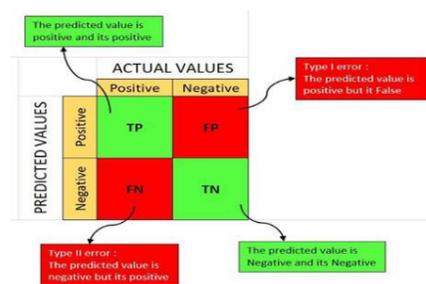


FIGURE 2.7 – Matrice de confusion

Le TP (True Positives) indique le nombre de prédictions où le modèle a correctement prédit une classe prédéterminée positive, le TN (True Negatives) désigne le nombre de prédictions où le modèle a correctement prédit une classe prédéterminée négative, le FP (False Positives) représente le nombre de prédictions où le modèle a prédit incorrectement une classe prédéterminée positive, alors qu'elle était négative, et le FN (False Negatives) indique le nombre de prédictions pour lesquelles le modèle a prédit incorrectement une classe prédéterminée négative, alors qu'elle était positive [6].

b- Précision, rappel et score F1

Grâce à la matrice de confusion, la précision, le rappel ainsi que le score F1 peuvent être calculés.

La précision représente la proportion d'échantillons prédits qui appartiennent réellement à cette classe par rapport à tous les échantillons prédits de cette classe ; elle est exprimée de la manière suivante [6] :

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

Le rappel représente la proportion d'échantillons prévus dans une classe qui appar-

tiennent réellement à cette classe par rapport à tous les échantillons réels de cette classe ; il peut être exprimée de la manière suivante [6] :

$$Rappel = \frac{TP}{TP + FN} \quad (2.3)$$

Le score F1 est une évaluation unique qui combine précision et rappel, il est représenté comme suit [6] :

$$ScoreF1 = 2 * \frac{precision * rappel}{precision + rappel} \quad (2.4)$$

2.4.1.2 Régression

La régression est une technique couramment utilisée en apprentissage automatique et en statistique pour modéliser et analyser la relation entre une variable dépendante (cible) et une ou plusieurs variables indépendantes (caractéristiques). Cet algorithme est utilisé pour prédire les valeurs continues telles que les prix, les températures, ou d'autres mesures quantitatives.

2.4.1.2.1 Modèles de régression

a- Régression linéaire simple

La régression linéaire est un modèle de régression simple qui calcule la relation linéaire entre la variable dépendante (Y) et une variable indépendante (X) en ajustant une équation linéaire aux données observées. L'équation est donnée de cette forme :

$$y = a + bx + \epsilon \quad (2.5)$$

Où :

y : variable dépendante(cible) ;

x : Variable indépendante (caractéristique) ;

b : Coefficient de régression ;

a : est l'ordonnée à l'origine (représentant la valeur de y lorsque x=0) ;

ϵ : est l'erreur aléatoire.

La formule de b pour la régression linéaire simple est :

$$b = \frac{n \sum (xy) - \sum x \sum y}{n \sum (x^2) - (\sum x)^2} \quad (2.6)$$

La formule de a pour la régression linéaire simple est :

$$a = \frac{\sum y - b \sum x}{n} \quad (2.7)$$

Un exemple de régression linéaire est illustré dans la figure (2.8) :

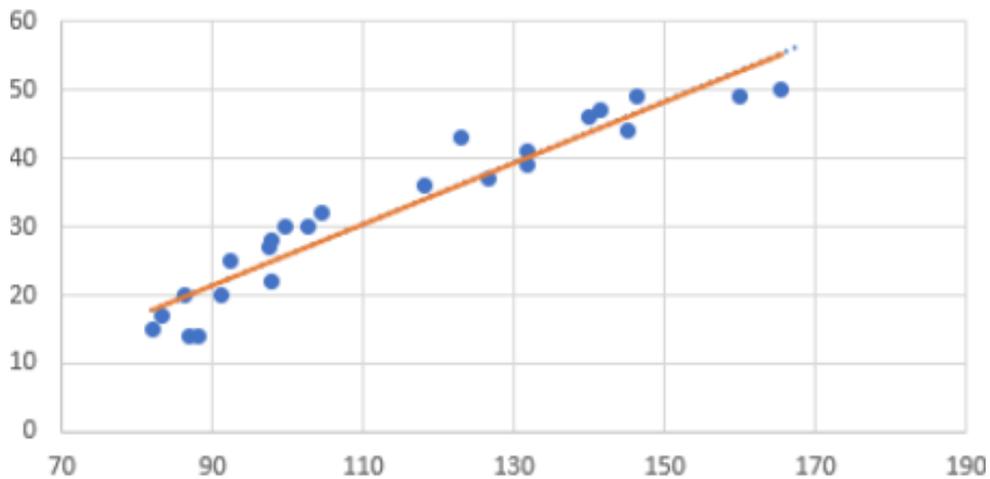


FIGURE 2.8 – Exemple de régression linéaire

b- Régression linéaire multiple

Cela est une extension de la régression linéaire simple qui permet de modéliser la relation entre la variable dépendante (y) et plusieurs variables indépendantes. L'équation de la régression linéaire multiple est la suivante :

$$y = a + b_1x_1 + b_2x_2 + \dots + b_kx_k + \epsilon \quad (2.8)$$

Où :

x_1, x_2, \dots, x_k : Variables indépendantes (caractéristiques) ;

b_1, b_2, \dots, b_k : Coefficients de régression associés à chaque variable indépendante.

c- Régression polynomiale

La régression polynomiale est une technique d'analyse de régression dans laquelle la relation entre la variable indépendante X et la variable dépendante y est modélisée comme un polynôme de degré n . La formule générale pour une régression de degré polynomiale est :

$$y = a + b_1x + b_2x^2 + b_3x^3 \dots + b_kx^n + \epsilon \quad (2.9)$$

Dans la figure (2.9), on peut observer un exemple de régression polynomiale :

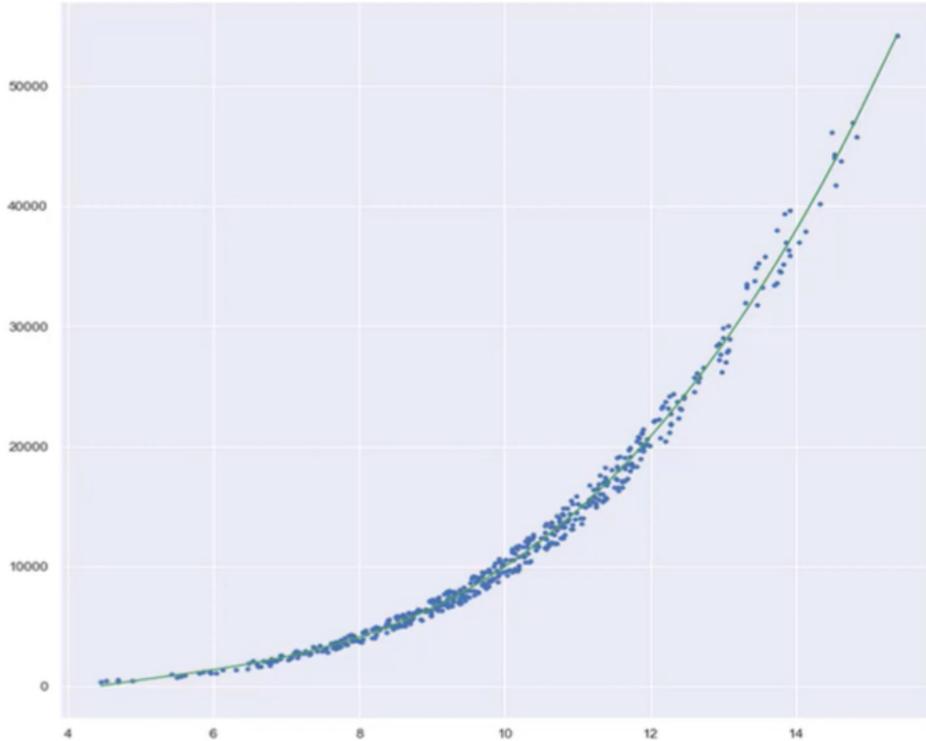


FIGURE 2.9 – Exemple de régression polynomiale

d- Forêts aléatoires

Les forêts aléatoires sont une méthode d'ensemble qui combine plusieurs arbres de décision afin de prédire la valeur cible. Elles consistent en une combinaison de plusieurs arbres de décision pour améliorer la précision et la robustesse par rapport aux arbres individuels. Elles sont utilisées à la fois pour des tâches de régression et de classification.

Les principaux paramètres de l'algorithme de la forêt aléatoire incluent :

- `n_estimators` : le nombre d'arbres dans la forêt ;
- `max_features` : le nombre maximal de caractéristiques à considérer pour diviser chaque nœud ;
- `max_depth` : la profondeur maximale des arbres. Limiter la profondeur peut aider à prévenir le surapprentissage (overfitting) ;
- `min_samples_split` : le nombre minimum d'échantillons requis pour diviser un nœud.

L'algorithme de la forêt aléatoire est représenté dans la figure (2.10), démontrant comment l'ensemble des arbres de décision contribue à la prédiction finale :

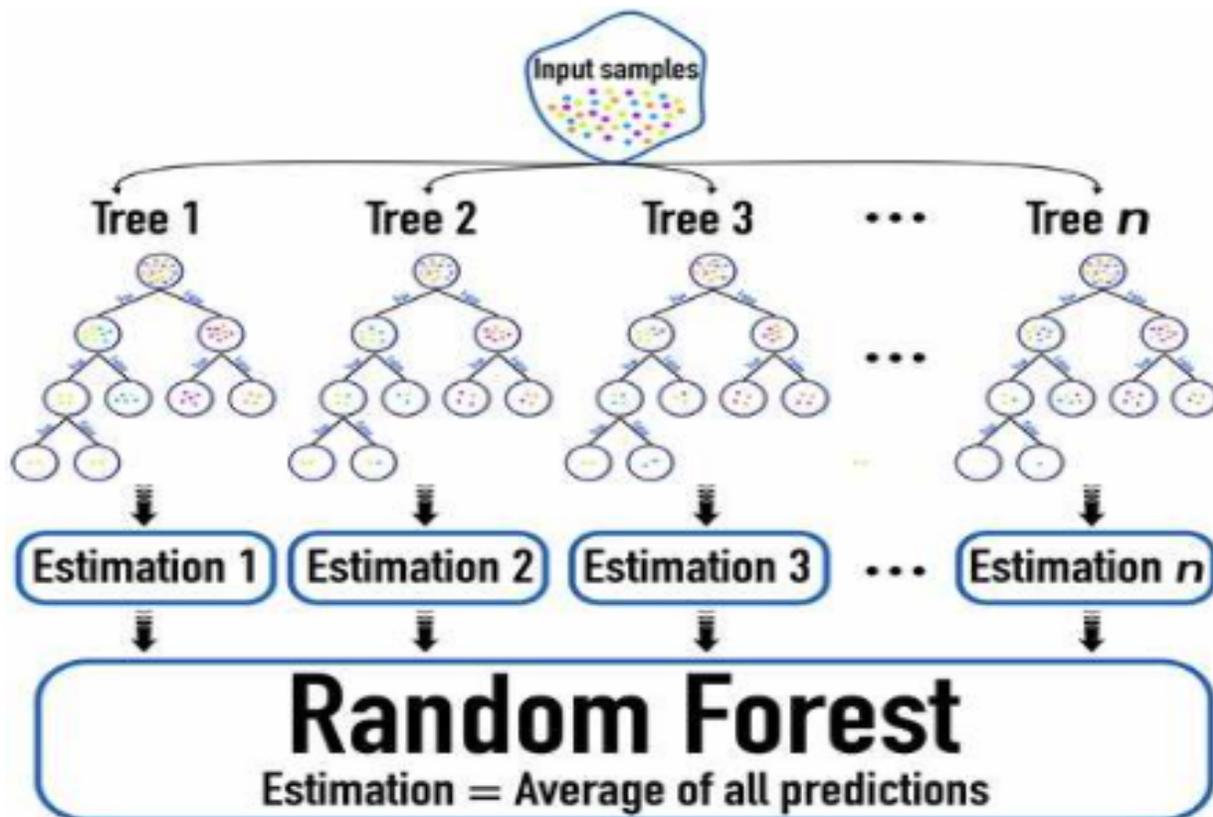


FIGURE 2.10 – Algorithme de la forêt aléatoire

e- Régression à support de vecteur(SVR)

La régression à support de vecteur est une technique basée sur la méthode des Machines à Vecteurs de Support adaptée aux tâches de régression. Son objectif est de trouver une fonction qui prédit de manière optimale les valeurs de la variable dépendante y en fonction des variables indépendantes x .

f- Réseaux de neurones artificiels

Un réseau neuronal artificiel (ANN), souvent simplement appelé réseau neuronal, est un modèle d'apprentissage supervisé utilisé pour prédire à la fois les valeurs discrètes et continues. L'architecture du réseau est basée sur des neurones connectés dans une couche d'entrée, une ou plusieurs couches cachées et une couche de sortie. Ce réseau est utilisé pour apprendre des modèles complexes à partir de données grâce à un processus appelé entraînement. Pendant ce processus, le réseau ajuste ses paramètres (poids et biais) en fonction des données d'entrée et de la sortie attendue [7].

La figure (2.11) montre l'approche des ANN :

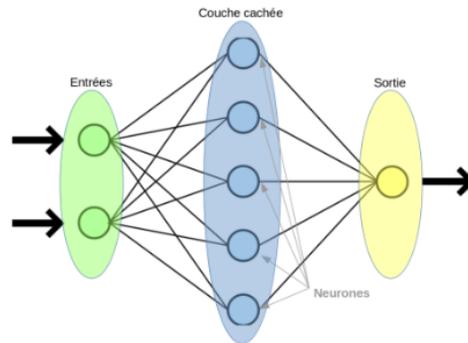


FIGURE 2.11 – L'approche des ANN

2.4.1.2.2 Métriques d'évaluation

a- Erreur moyenne quadratique (MSE)

La MSE (en anglais Mean Square Error) est une mesure très simple qui calcule la différence quadratique entre les valeurs réelles et prédites. Elle est définie comme suit :

$$MSE = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n} \quad (2.10)$$

Où :

y_i : est la valeur réelle pour l'observation i ;

\hat{y}_i : est la valeur prédite pour l'observation i .

b- Racine d'erreur moyenne quadratique (RMSE)

RMSE (Root Mean Square Error) est la racine carrée de la moyenne des carrés des erreurs(MSE), elle est définie de la manière suivante :

$$RMSE = \sqrt{MSE} = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (2.11)$$

c- Erreur Moyenne Absolue (MAE)

La MAE (Mean Absolute Error) est une mesure très simple qui calcule la différence absolue entre les valeurs réelles et prédites. Elle est définie comme suit :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.12)$$

d- R^2 score

Le R^2 score, aussi appelé R-squared ou coefficient de détermination, est une mesure qui indique les performances de votre modèle, calcule la proportion de la variance de la variable

cible expliquée par le modèle.. Des valeurs plus élevées indiquent un meilleur ajustement, allant de 0 à 1. Il est défini par la formule suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.13)$$

Où :

\bar{y} : est la moyenne des valeurs réelles.

2.4.2 Apprentissage non supervisé

Est une branche d'apprentissage automatique qui implique l'extraction de modèles, de relations et de structures à partir de données non étiquetées. Cela signifie qu'aucune donnée d'entraînement ne peut être fournie et que la machine est amenée à apprendre par elle-même. Il est essentiel que la machine puisse classer les données sans avoir besoin d'informations préalables sur celles-ci [8].

2.4.2.1 Clustering

Le clustering, ou le regroupement, est une technique fondamentale d'exploration de données pour découvrir la structure de regroupement cachée des données. Le clustering implique l'organisation des données non étiquetées en groupes similaires, appelés clusters, où chaque cluster désigne un groupe d'éléments de données similaires.

Parmi les algorithmes d'apprentissage non-supervisé les plus courants, on peut citer l'algorithme des k-moyennes (K-means).

La méthode clusternig est illustrée dans la figure (2.12) :

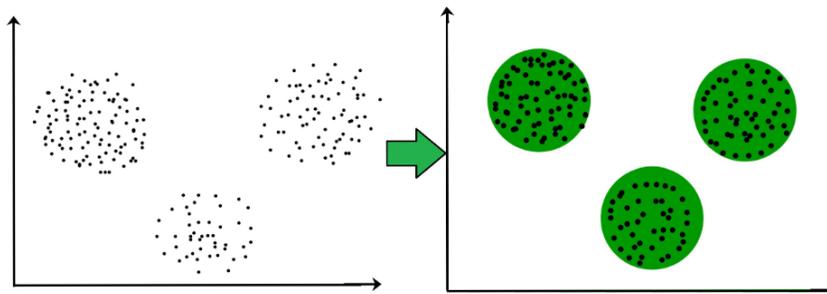


FIGURE 2.12 – Méthode de clustering

2.4.3 Apprentissage par renforcement

L'apprentissage par renforcement est un type d'algorithme de l'apprentissage automatique dans lequel un agent apprend à prendre des décisions successives en interagissant avec son environnement. Il repose sur le principe de découvrir des tactiques optimales qui maximisent les récompenses cumulées au fil du temps par essais et erreurs [9]

. La figure (3.13) présente le processus d'algorithme par renforcement :

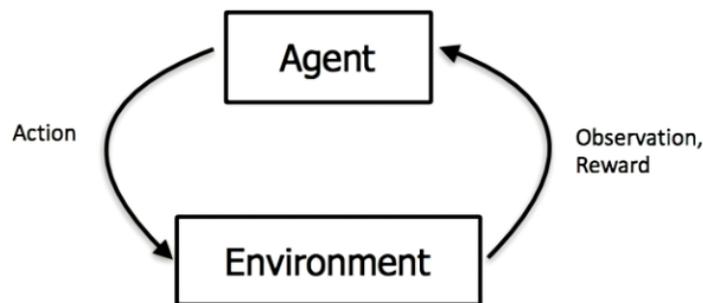


FIGURE 2.13 – Processus d'apprentissage par renforcement

Parmi les algorithmes d'apprentissage par renforcement les plus courants, on peut citer l'algorithme Recherche d'arbres à Monte-Carlo et Q-learning (pour plus d'informations voir [9]).

2.5 Domaines d'application

L'apprentissage automatique trouve des applications dans divers domaines, notamment dans la médecine pour le diagnostic médical automatisé et l'analyse d'images, la technologie pour la reconnaissance d'images, et le commerce pour la prévision des ventes.

II. Prévision des Ventes en Utilisant l'apprentissage automatique

La prévision des ventes est un processus complexe mais indispensable pour la gestion stratégique d'une entreprise suite à les défis qu'elle rencontre mais grâce à l'apprentissage automatique qui est un outil précieux pour améliorer la précision et l'efficacité des prévisions de ventes, en adaptant des méthodes robustes de collecte, de nettoyage et d'analyse des données.

2.6 Qu'est-ce que la prévision des ventes ?

La prévision des ventes, autrement dit demand planning, est une méthode dont l'objectif est de prévoir les ventes à venir d'une entreprise consiste à l'aide des données historiques et d'autres variables pertinentes. La prévision englobe également un ensemble de méthodes statistiques très diverses qui visent à réduire l'incertitude liée à la non connaissance du futur. Elle constitue de ce fait un maillon essentiel de la chaîne logistique et du supply chain (ou chaîne d'approvisionnement) d'une entreprise [10].

2.7 L'importance de la prévision des ventes

La prévision des ventes est cruciale pour toute entreprise, une prévision réaliste des ventes futures permet :

- . Une gestion des stocks optimale ;
- . Une meilleure anticipation et une planification optimisée de la production ;
- . Une optimisation de la supply chain ;
- . Une planification stratégique précise ;
- . Une réduction de coût.

2.8 Défis de la prévision des ventes

La prévision des ventes fait face à divers obstacles, mais l'un des plus grands défis de la prévision des ventes est la fiabilité des données de vente :

2.8.1 Fiabilité des données

La précision des prévisions dépend en grande partie de la qualité et de la fiabilité des données historiques, des données incomplètes, incorrectes ou obsolètes peuvent conduire à des prévisions imprécises et des décisions stratégiques erronées.

2.8.2 Volatilité du marché

Les évolutions rapides causées par des éléments économiques, technologiques ou sociaux rendent les prévisions de ventes délicats. En outre, les consommateurs ont un comportement imprévisible et peuvent changer rapidement en fonction des tendances, des nouvelles technologies ou des événements inattendus.

Pour surmonter ces défis, de nombreuses entreprises utilisent désormais des techniques de machine learning.

2.9 Utilisation du machine learning pour prévoir les ventes

Grâce au machine learning, les entreprises peuvent mieux comprendre et prévoir les tendances de consommation, intégrant une multitude de facteurs internes et externes pour une prise de décision plus éclairée et réactive.

2.9.1 Apprentissage automatique comme force de prédiction

Le machine learning est une puissante méthode d'analyse des données qui met la lumière sur les tendances et les comportements. Grâce à des algorithmes sophistiqués, il explore et analyse des ensembles de données internes (prix, canaux de distribution, etc.) ainsi que des données externes (saisonnalité, concurrence, etc.) afin de découvrir des modèles, des corrélations et des relations cachées. Ces informations permettent de générer des prédictions précises sur des événements futurs, offrant ainsi aux entreprises des insights précieux pour prendre des décisions éclairées [?].

2.9.2 Avantages de l'apprentissage automatique dans la prévision des ventes

La prédiction des ventes à l'aide de machine learning offre de nombreux avantages à l'entreprise, notamment [11] :

- . L'apprentissage automatique peut gérer de grandes quantités de données (big data) provenant de diverses sources ;
- . Les modèles d'apprentissage automatique offrent une meilleure prévision en détectant des tendances et des relations qui échappent aux méthodes classiques, ce qui permet ainsi de générer des prévisions plus fiables ;
- . Les algorithmes d'apprentissage automatique ont la capacité de s'adapter aux évolutions de l'environnement, ce qui est crucial pour les entreprises qui opèrent dans des secteurs en constante évolution ;
- . Une fois que le modèle d'apprentissage automatique est entraîné, il est capable de produire des prévisions rapidement et avec peu d'intervention humaine, ce qui permet de libérer des

ressources pour d'autres tâches

2.10 Etapes pour mettre en place un modèle d'apprentissage automatique

La construction d'un modèle de prédiction des ventes à l'aide d'apprentissage automatique implique plusieurs étapes [12] :

2.10.1 Collecte et préparation des données

2.10.1.1 Sources de données

Cette étape permet l'identification et l'intégration des sources de données internes et externes telle que les données historiques des ventes, les campagnes marketing, les saisons, etc.

2.10.1.2 Nettoyage et prétraitement

Le nettoyage et le prétraitement des données sont des étapes cruciales dans le processus de prévision des ventes. Il est essentiel de gérer les valeurs manquantes (méthode moyennes, médianes, etc), de détecter et corriger les valeurs aberrantes qui pourraient fausser les prédictions, de transformer les données en normalisant ou standardisant les fonctionnalités, et de convertir les variables catégorielles (comme les catégories de produits et les régions) en représentations numériques par encodage à chaud ou encodage d'étiquettes.

2.10.1.3 Analyse exploratoire des données (AED)

Cette étape, vise à comprendre la structure et les caractéristiques des données disponibles. Grâce à des techniques de visualisation telles que les graphiques, les histogrammes et les matrices de corrélation, elle permet d'identifier non seulement les variables directement liées aux ventes, mais aussi les interrelations subtiles entre différentes caractéristiques. De plus, l'AED permet la réalisation des test statistiques (tels que tests de stationnarité, d'autocorrélation) qui permettent de valider les hypothèses.

2.10.2 Ingénierie des fonctionnalités

Cette étape consiste à sélectionner et créer des caractéristiques pertinentes à partir des données disponibles. La sélection des fonctionnalités comprend l'analyse de corrélation afin

d'identifier les variables influentes et de supprimer de manière itérative les fonctionnalités moins importantes.

2.10.3 Choix du modèle de machine learning

Après avoir préparé les données, des algorithmes de machine learning, telles que les forêts aléatoires (comme notre cas) peuvent ensuite être appliquées pour construire le modèle prédictif en fonction des caractéristiques des données et des objectifs de prédiction et des ressources disponibles.

2.10.4 Entraînement et évaluation de modèle

Cette étape implique le partage des données en ensembles d'entraînement et de test, d'entraîner le modèle sur l'ensemble d'entraînement, puis d'évaluer les performances de modèle sur l'ensemble de test à l'aide des métriques d'évaluation appropriées telles que MAE, RMSE et MSE. Les prédictions sont ensuite comparées aux données réelles afin d'assurer leur précision et leur fiabilité.

Après avoir divisé les données en ensembles d'entraînement et de test, l'utilisation de la validation croisée permet d'évaluer la performance de modèle de manière plus robuste. La validation croisée K-fold est couramment utilisée pour trouver les meilleurs hyperparamètres et éviter le surapprentissage.

2.10.5 Déploiement

Le déploiement de l'algorithme de prévision des ventes consiste à intégrer un modèle de machine learning, une fois qu'il a été développé et validé, dans un environnement de production afin de pouvoir réaliser des prévisions en temps réel ou à intervalles réguliers. Plusieurs étapes cruciales sont prises dans ce processus, telles que l'intégration avec les systèmes déjà en place, l'automatisation des prévisions et la mise en place de la surveillance continue afin d'assurer la performance et l'efficacité du modèle.

2.11 Problématique :

La prédiction des ventes est un aspect crucial de la gestion commerciale des entreprises. Cependant, les méthodes traditionnelles utilisées présentent souvent des limites en termes de précision et de fiabilité. Dans ce contexte, comment mettre en place un modèle de prédiction des ventes efficace et précis pour l'entreprise Cevital, en utilisant des techniques avancées d'apprentissage automatique et d'analyse de données ?

Questions centrales :

1. Quelles sont les méthodes traditionnelles actuellement utilisées par Cevital pour prédire les ventes et quelles sont leurs limites ?
2. Quelles sont les techniques d'apprentissage automatique les plus adaptées pour améliorer la précision des prévisions de ventes ?
3. Quels sont les principaux indicateurs et données à prendre en compte dans le modèle de prédiction des ventes de Cevital ?
4. Comment intégrer efficacement le modèle de prédiction des ventes dans les processus décisionnels de l'entreprise ?
5. Quels sont les défis potentiels liés à la mise en œuvre et à la maintenance d'un tel modèle au sein de Cevital, et comment peuvent-ils être surmontés ?

2.12 Conclusion

Dans ce chapitre, nous avons présenté les bases nécessaires d'apprentissage automatique en introduisant, en décrivant et en classant divers types d'algorithmes. Nous avons saisi l'essence même de l'application de l'apprentissage automatique en explorant ses objectifs, guidés par la recherche de modèles et de prédictions dans les données. Nous avons ainsi illustré comment l'apprentissage automatique révolutionne la prédiction des ventes grâce à des algorithmes sophistiqués.

Chapitre 3

Implémentation

I. Prévision des ventes pour l'année 2024

Introduction

Dans cette partie, nous exposons le processus d'implémentation d'un modèle de prévision des ventes pour cevital pour l'année 2024. L'objectif principal, consiste à créer un modèle d'apprentissage automatique capable de prédire avec précision les ventes futures en se basant sur les données historiques de l'année 2023. Suite à une analyse approfondie de différentes méthodes, nous avons choisi l'algorithme des forêts aléatoires en raison de sa solidité et de sa capacité à gérer des données complexes et volumineuses.

3.1 Outils et environnements de développement

3.1.1 Langage de programmation : "Python"

Même si d'autres langages proposent une multitude d'options, on a opté pour "Python" en raison de son écosystème complet en machine learning, de sa syntaxe claire et de sa facilité d'installation. C'est quoi Python ?

Python est le langage de programmation informatique le plus populaire et le plus utilisé, en particulier dans les domaines de la Science des Données et de Machine Learning. De plus, Python est un langage multiplateforme qui fonctionne sur divers systèmes d'exploitation, tels que Windows, macOS et Linux, ce qui en fait une option parfaite pour les développeurs qui travaillent dans différents environnements.

3.1.2 Environnement de développement

Nous avons opté pour "Jupyter Notebook" comme environnement de développement, car c'est une application web open-source qui permet de partager du code source. Ce cadre est utilisé pour le développement interactif, la visualisation des données et la documentation.

3.1.3 Bibliothèques utilisées

On a utilisé pas mal de bibliothèques ;

- **Pandas** : Pour la manipulation et l'analyse des données.
- **Numpy** : Pour les opérations mathématiques et statistiques sur des tableaux de grandes dimensions.
- **Matplotlib et Seaborn** : Pour la visualisation des données. Ces bibliothèques offrent la possibilité de concevoir des graphiques et des visualisations de haute qualité afin de comprendre les tendances et les répartitions des données.
- **Scikit-learn** : Pour le développement des modèles de machine learning. Elle comprend notamment des fonctions pour estimer des forêts aléatoires, des régressions et des algorithmes de classification.

Dans la figure (3.1), on trouve toutes les bibliothèques requises pour effectuer une analyse :

```
import pandas as pd
import numpy as np
from scipy.stats import f_oneway #pour réaliser un test d'analyse de la variance (ANOVA) à des variables catégorielles
from scipy.stats import wilcoxon
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV, cross_val_score # 'train_test_split' pour diviser les données en ensembles
#                                     d'entraînement et de test,
#                                     # 'GridSearchCV' pour optimiser les hyperparamètres d'un modèle
#                                     en utilisant la validation croisée,
#                                     # 'cross_val_score' pour évaluer le modèle en utilisant validation cro
from sklearn.ensemble import RandomForestRegressor #pour créer et entraîner un modèle utilisant une forêt aléatoire
from sklearn.preprocessing import StandardScaler, OneHotEncoder # 'StandardScaler' Pour standardiser les caractéristiques en supprimant la moyenne
#                                     et en mettant à l'échelle par la variance.
#                                     # 'OneHotEncoder' Pour convertir les variables catégorielles en variables indicatrices.
from sklearn.compose import ColumnTransformer # Pour appliquer différentes transformations à des colonnes spécifiques d'un DataFrame.
from sklearn.pipeline import Pipeline # Pour créer un pipeline séquentiel d'étapes de transformation et de modélisation.
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score # pour calculer les métriques d'évaluation (régression)
```

FIGURE 3.1 – Importation des bibliothèques nécessaires

3.2 Collecte des données

Pour la prévision des ventes de Cevital, nous disposons d'un jeu de données complet pour l'année 2023 qui contient 728520 lignes et 46 colonnes au format excel. Ce jeu de données comprend une variété de variables, à la fois numériques telles que NumLigne, QT Tonne, etc., et catégorielles comme Produit, Famille, Sousfamille, etc.

CHAPITRE 3. IMPLÉMENTATION

Voici un extrait de quelques lignes de notre ensemble de données :

NumLigne	NumeroFacture	NO_BC	NO_BL	RéférenceFacture	CodeProduit	Produit	Famille	Sousfamille	CodeDepot	...	Immatricul
4000	FCE2023000043	COMCEV0223020687	BLNDLP2723000006	SCMIX23002/1	B9812Y8515	BOISSON COCKTAIL MANGUE 125CL PET	BOISSONS	EAU FRUITEE PET 125CL	DLP27	...	
5000	FCE2023000043	COMCEV0223020687	BLNDLP2723000006	SCMIX23002/1	B9812Y8516	BOISSON FRAISE POMME BANANE 125CL PET	BOISSONS	EAU FRUITEE PET 125CL	DLP27	...	
1000	FCE2023000043	COMCEV0223020687	BLNDLP2723000006	SCMIX23002/1	B9812Y8512	BOISSON ORANGE 125CL PET	BOISSONS	EAU FRUITEE PET 125CL	DLP27	...	
2000	FCE2023000043	COMCEV0223020687	BLNDLP2723000006	SCMIX23002/1	B9812Y8513	BOISSON ORANGE ABRICOT 125CL PET	BOISSONS	EAU FRUITEE PET 125CL	DLP27	...	
3000	FCE2023000043	COMCEV0223020687	BLNDLP2723000006	SCMIX23002/1	B9812Y8514	BOISSON ORANGE PECHE 125CL PET	BOISSONS	EAU FRUITEE PET 125CL	DLP27	...	
...

FIGURE 3.2 – Extrait de quelques lignes de notre ensemble de données

3.3 Préparation des données

Avant de passer au processus de nettoyage, nous avons effectué quelques tests tels que test ANOVA, wilcoxon pour les variables catégorielles et la corrélation pour les variables numériques afin d'extraire les variables pertinentes pour notre variable cible (QT Tonne). Les résultats des tests statistiques appliqués sur notre variables sont présentés ci-dessus :

```
Test de Wilcoxon pour la variable 'NOM_MOIS' et 'QT Tonne':  
Statistique de test : 1314108.0  
p-value : 0.0  
Test de Wilcoxon pour la variable 'statutFacturation' et 'QT Tonne':  
Statistique de test : 1314108.0  
p-value : 0.0
```

CHAPITRE 3. IMPLÉMENTATION

```
Test ANOVA pour la variable 'NumeroFacture' et 'QT Tonne':
Statistic F : 20.040495615024152
p-value : 0.0
Test ANOVA pour la variable 'CodeProduit' et 'QT Tonne':
Statistic F : 39.112660115933615
p-value : 0.0
Test ANOVA pour la variable 'Produit' et 'QT Tonne':
Statistic F : 39.41264625874963
p-value : 0.0
Test ANOVA pour la variable 'Famille' et 'QT Tonne':
Statistic F : 265.88448382715944
p-value : 0.0
Test ANOVA pour la variable 'Sousfamille' et 'QT Tonne':
Statistic F : 106.66960451175656
p-value : 0.0
Test ANOVA pour la variable 'TypePieceFacture' et 'QT Tonne':
Statistic F : 556.880476409535
p-value : 1.5978022698673324e-122
Test ANOVA pour la variable 'National_Etranger' et 'QT Tonne':
Statistic F : 3463.221228425844
p-value : 0.0
Test ANOVA pour la variable 'UV_Facture' et 'QT Tonne':
Statistic F : 168.1777685219072
p-value : 0.0
```

QT US	0.434678
QT UV	0.014647
Coef_UV_US	0.058945
QTE_PLT	0.939552
QT Tonne	1.000000
CA HT	0.227724
CA HT DZD	0.823763
CA TTC	0.819647
PrixNet	0.026858
PrixNetUnitaire	0.004431
QT Tonne Liquide	0.999585
TauxDeChange	-0.504518

FIGURE 3.3 – Résultats des tests statistiques appliqués sur notre variables

Interprétation :

- Pour chacune des variables catégorielles testées, la p-valeur < 0.05 indiquant que les différences observées entre les groupes sont statistiquement significatives.
- Les variables numériques sont sélectionnées en fonction de leurs coefficients de corrélation élevés (>0.20) avec 'QT Tonne'.

Ces variables 'NumeroFacture', 'CodeProduit', 'Produit', 'Famille', 'Sousfamille', 'TypePieceFacture', 'National_Etranger', 'UV_Facture', 'NOM_MOIS', 'statutFacturation', 'DateFacture', 'QTE_PLT', 'CA HT DZD', et 'CA TTC', 'QT US', 'CA HT', seront donc retenues pour la suite de l'analyse et incluses dans le modèle de prévision des ventes.

3.3.1 Nettoyage des données

a- Suppression des valeurs négatives :

Nous avons supprimé les valeurs négatives, car elles sont considérées comme des retours de consignation. Les valeurs négatives avant et après la suppression sont visualisées ci-dessus :

<p>Colonnes avec des valeurs négatives :</p> <pre> NUM_ANNEE 0 QT US 20362 QTE_PLT 84 QT Tonne 84 CA HT 20354 CA HT DZD 20354 CA TTC 20354 CodeClient 0 PrixNetUnitaire 0 dtype: int64 </pre>	<p>→</p>	<p>Colonnes sans des valeurs négatives :</p> <pre> NUM_ANNEE 0 QT US 0 QTE_PLT 0 QT Tonne 0 CA HT 0 CA HT DZD 0 CA TTC 0 CodeClient 0 PrixNetUnitaire 0 dtype: int64 </pre>
---	----------	---

FIGURE 3.4 – Les valeurs négatives des colonnes avant et après la suppression.

On peut observer dans la figure (3.4) à droite que certaines colonnes comme CA TTC, CA HT, CA HT DZD, QT US contiennent un grand nombre de valeurs négatives. Par conséquent, toutes les valeurs inférieures à 0 ont été supprimées, ce qui est visible dans la figure (3.4) à gauche.

b- Gestion des valeurs manquantes :

Nous avons détecté les valeurs manquantes de chaque variable en utilisant la commande `'isnull().sum()'`. Les valeurs manquantes sont visibles dans la figure (3.5) :

NumeroFacture	0		NumeroFacture	0
CodeProduit	0		CodeProduit	0
Produit	0		Produit	0
Famille	3		Famille	0
Sousfamille	3		Sousfamille	0
TypePieceFacture	0		TypePieceFacture	0
National_Etranger	0		National_Etranger	0
DateFacture	0		DateFacture	0
UV_Facture	0		UV_Facture	0
NOM_MOIS	0		NOM_MOIS	0
NUM_ANNEE	0		NUM_ANNEE	0
QT US	0	→	QT US	0
QTE_PLT	0		QTE_PLT	0
QT Tonne	0		QT Tonne	0
CA HT	0		CA HT	0
CA HT DZD	0		CA HT DZD	0
CA TTC	0		CA TTC	0
CodeClient	0		CodeClient	0
RaisonSociale	0		RaisonSociale	0
Ville	0		Ville	0
PrixNetUnitaire	0		PrixNetUnitaire	0
statutFacturation	0		statutFacturation	0
statutLivraison	0		statutLivraison	0
dtype: int64			dtype: int64	

FIGURE 3.5 – Comparaison des données avant et après la gestion des valeurs manquantes

Les données sont globalement complètes avec seulement quelques valeurs manquantes dans les deux colonnes 'Famille' et 'Sousfamille' comme illustré dans la figure (3.5) à droite, tandis que dans la figure (3.5) à gauche, on visualise l'absence de valeurs manquantes dans l'ensemble de données est extrêmement favorable pour l'analyse et la modélisation.

c- Traitement des valeurs aberrantes :

Pour la détection et la suppression des valeurs aberrantes, nous avons opté pour utiliser le boxplot pour visualiser la distribution de la variable "QT Tonne" et appliqué la technique IQR (Interquartile Range) pour la suppression. En calculant la différence entre le troisième quartile (Q3) et le premier quartile (Q1), les valeurs aberrantes sont souvent définies comme celles qui se situent à l'extérieur des borne définier par :

$$Borneinf = Q1 - 1.5 * IQR$$

$$Bornesup = Q3 + 1.5 * IQR$$

La figure (3.6) présente le resultat de la méthode IQR :

```

Valeurs supprimées :
      Column  Removed_Value
6119  QT Tonne      40.000
6119  QT Tonne       3.128
6119  QT Tonne       3.312
6119  QT Tonne       4.692
6119  QT Tonne       1.564
...
20523  CA TTC  1980000.000
20523  CA TTC  1980000.000
20523  CA TTC  195600.000
20523  CA TTC   94858.470
20523  CA TTC  6201404.160

[21372 rows x 2 columns]
Nombre de lignes avant suppression des valeurs aberrantes : 257162
Nombre de lignes après suppression des valeurs aberrantes : 255403
    
```

FIGURE 3.6 – Test statistiques préliminaires

On observe que la méthode IQR a efficacement identifiée et supprimée 1 759 lignes contenant des valeurs aberrantes dans "QT Tonne". Cela assure que les données restantes sont plus fiables.

La figure (3.7), représente la distribution des variables numériques avant et après le traitement. Il offre une comparaison visuelle entre les deux distributions :

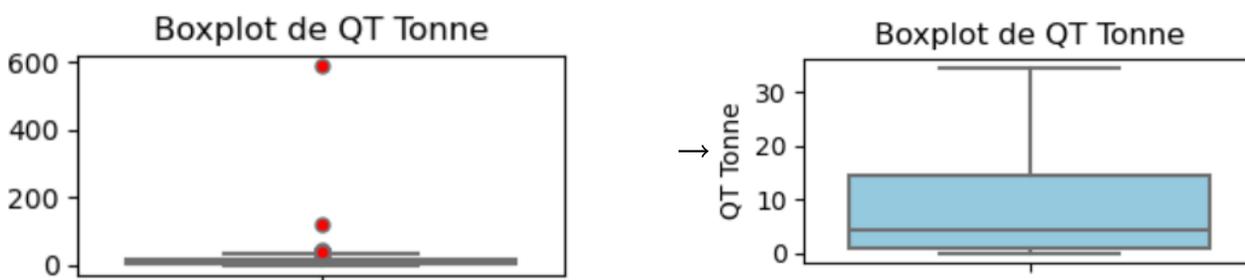


FIGURE 3.7 – Boxplot de la distribution des variables numériques avant et après le traitement des valeurs aberrantes

Le boxplot à droite illustre la distribution des données, où les points en rouge sont considérés comme des valeurs aberrantes. Le boxplot à gauche montre la distribution des données après la suppression de ces valeurs aberrantes, indiquant ainsi une distribution nettoyée.

3.3.2 Encodage et Standarisation des variables

Dans cette étape, nous avons divisé les variables en deux catégories, numériques et catégorielles.

- Pour les variables catégorielles avec un grand nombre de catégories sont encodées à l'aide de 'LabelEncoder', et pour celles avec un nombre limité de catégories sont encodées en utilisant l'encodage one-hot qui crée une colonne binaire pour chaque catégorie.

- Les variables numériques sont standarisées en soustrayant la moyenne et en divisant par l'écart type, afin de mettre toutes les variables sur la même échelle.

Après l'encodage et la standarisation, les transformations sont appliquées au données à l'aide de la méthode fit_transform() du ColumnTransformer, qui combine les différentes étapes de prétraitement. Les résultats sont converti en DataFrame avec les noms des colonnes apropiées.

Dans la figure (3.8), on peut observer les résultats de l'encodage et de la standarisation des variables :

```

NUM_ANNEE    QT US    QTE_PLT    QT Tonne    CA HT    CA HT DZD    CA TTC \
0      0.0    0.454330    0.082542   -0.070960    0.294297    0.294297    0.244654
1      0.0   -0.515191   -0.674957   -0.846360   -0.811891   -0.811891   -0.818272
2      0.0   -0.192017   -0.422457   -0.587894   -0.443162   -0.443162   -0.463963
3      0.0   -0.299742   -0.506624   -0.674049   -0.566072   -0.566072   -0.582066
4      0.0   -0.192017   -0.422457   -0.587894   -0.443162   -0.443162   -0.463963

CodeClient    PrixNetUnitaire    TypePieceFacture_Facture    ...    NumeroFacture \
0   -0.998528           -0.223518                    1.0    ...           51.0
1   -0.326457           -0.223518                    1.0    ...           56.0
2   -0.326457           -0.223518                    1.0    ...           60.0
3   -1.073282           -0.223518                    1.0    ...           26.0
4   -1.073282           -0.223518                    1.0    ...           58.0

CodeProduit    Produit    Famille    Sousfamille    DateFacture    UV_Facture \
0      10.0     73.0     7.0           15.0           0.0           19.0
1      10.0     73.0     7.0           15.0           0.0           19.0
2      10.0     73.0     7.0           15.0           0.0           19.0
3      10.0     73.0     7.0           15.0           0.0           19.0
4      10.0     73.0     7.0           15.0           0.0           19.0

NOM_MOIS    RaisonSociale    Ville
0      4.0           200.0     82.0
1      4.0           205.0     6.0
2      4.0           205.0     6.0
3      4.0           298.0     58.0
4      4.0           298.0     58.0

```

FIGURE 3.8 – Résultats de l'encodage et de la standarisation des variables

On observe que les variables numériques comme 'CA TTC', 'QT Tonne', 'CA HT', etc, ont été standardisées avec des valeurs telles que 0.244654, -0.818272, etc, qui représentent le nombre d'écart-types par rapport à la moyenne. Tandis que, les variables catégorielles avec un grand nombre de catégories comme 'Produit', 'Famille', 'Sousfamille', etc, ont été encodées par des entiers, tels que 'Ville', qui a des valeurs comme 82.0, 6.0, etc. Les colonnes binaires ont été utilisées pour encoder les variables avec un nombre restreint telles que 'TypePieceFacture'.

3.4 Analyse exploratoire des données (AED)

a- Analyse des corrélations

Nous avons Examiné des corrélations entre les variables quantitatives pour identifier les relations linéaires potentielles.

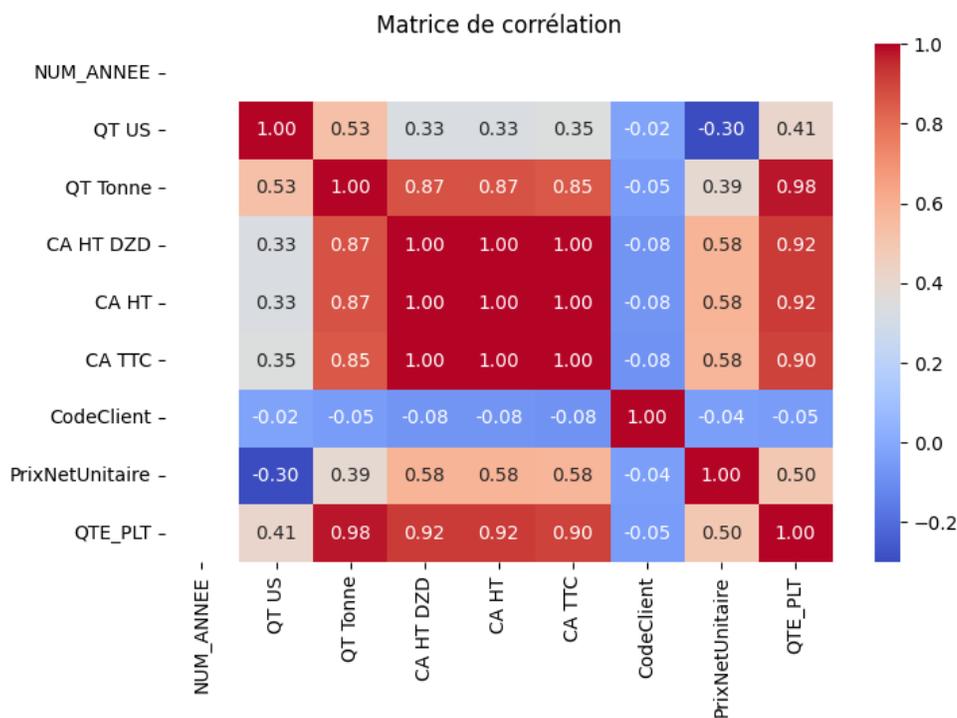


FIGURE 3.9 – Analyse des corrélations

En observant la matrice, les cases rouges témoignent de fortes corrélations, tandis que les cases orangées témoignent également de corrélations moins fortes par rapport à la case rouge. En revanche, les cases en bleu témoignent de faibles corrélations.

b- Statistiques descriptives

Nous avons effectuer une analyse descriptive des données à l'aide de la commande 'describe()', ce qui nous permet d'avoir un résumé statistique des variables. Les résultats sont présentés dans la figure (3.10) :

	NUM_ANNEE	QT US	QTE_PLT	QT Tonne	CA HT	CA HT DZD	CA TTC	CodeClient	PrixNetUnitaire
count	255403.0	255403.000000	255403.000000	255403.000000	2.554030e+05	2.554030e+05	2.554030e+05	2.554030e+05	255403.000000
mean	2023.0	4809.119559	8.847392	8.412018	7.229180e+05	7.229180e+05	7.500027e+05	2.697205e+07	3936.267837
std	0.0	7429.868911	9.134275	9.074506	7.518870e+05	7.518870e+05	7.643832e+05	1.488291e+07	17078.740815
min	2023.0	0.020000	0.000313	0.000235	1.813000e+01	1.813000e+01	2.157000e+01	1.862000e+03	3.000000
25%	2023.0	672.000000	1.000000	0.942300	1.277280e+05	1.277280e+05	1.461615e+05	2.101018e+07	82.500000
50%	2023.0	1848.000000	4.000000	4.140000	4.059000e+05	4.059000e+05	4.059000e+05	2.213608e+07	150.000000
75%	2023.0	4500.000000	16.000000	14.400000	1.169750e+06	1.169750e+06	1.188000e+06	3.200819e+07	565.000000
max	2023.0	480000.000000	251.500000	34.340000	2.797452e+06	2.797452e+06	2.822391e+06	8.209100e+07	113300.000000

FIGURE 3.10 – Statistiques descriptives

Selon les résultats, la variable 'NUM_ANNEE' ne prend en compte que des valeurs pour l'année 2023. La variable 'QT US' présente une moyenne de 4809.11 et un écart type de 7429.86, avec une valeur minimale de 0.02 et une valeur maximale de 480000. La variable 'QTE_PLT', a une moyenne de 8.85 et une médiane de 4.0. La variable 'QT Tonne' montre une moyenne de 8.41, une médiane de 4.14 et un écart type de 9.07 qui suggère que les valeurs de l'échantillon sont relativement dispersées autour de la moyenne. La variables 'PrixNetUnitaire' a une moyenne de 3936.27 et une médiane de 150. Alors que, les autres variables telles que 'CA HT', 'CA HT DZD', 'CA TTC' sont toutes exprimées en millions et présentent des valeurs variées avec des écarts-types significatifs.

c- Les produits les plus vendus

Nous avons identifié les produits les plus vendus avec leurs quantités vendus en tonnes. La figure (3.11) illustre le résultat :

Les 10 produits les plus vendus sont :

	Produit	QT Tonne
141	SUCRE SKOR 1Kg	473364.51800
74	HUILE ELIO 5 LITRES	374908.32140
55	EAU MINERALE 1.5 LITRE	316181.81700
139	SUCRE LIQUIDE SACCAROSE	166939.13200
130	SUCRE BLANC 25 Kg	138253.25000
73	HUILE ELIO 2 LITRES	129305.55104
132	SUCRE BLANC BIG BAG LOCAL 1.1T	117579.00000
143	SUCRE SKOR 2Kg	72482.22000
72	HUILE ELIO 1 LITRE	58323.92072
131	SUCRE BLANC 50 Kg	49686.70000

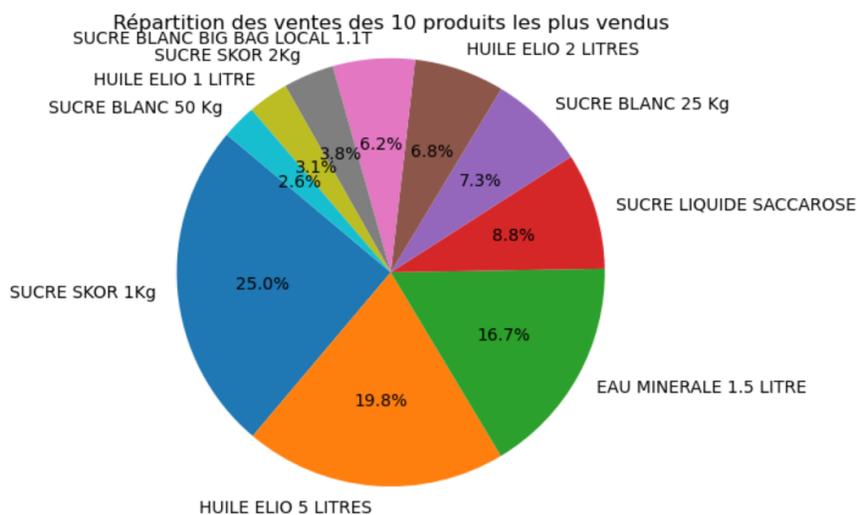


FIGURE 3.11 – Répartition des produits les plus vendus

D'après les résultats, le SUCRE SKOR 1Kg prend la première place avec une quantité de 473364.52 tonnes, suivi près par l'HUILE ELIO 5 LITRES avec 374908.32 tonnes vendus. L'EAU MINERALE 1.5 LITRE a connu une forte demande avec 316181.82 tonnes vendus. Les ventes de SUCRE LIQUIDE SACCAROSE sont élevées à 166939.13 tonnes, tandis que le SUCRE BLANC 25 Kg a atteint 138253.25 tonnes de vente. En outre, les autres produits du top 10 témoignent également d'une demande significative sur le marché. Le diagramme à camembre, montre que le "SUCRE SKOR 1KG" occupe une part significative de 25%. Cela signifie que ce produit constitue un quart de l'ensemble des ventes parmi les produits analysés.

d- Tendances historiques des ventes pour chaque produit

Nous allons analyser l'évolution des ventes de chaque produit au cours de l'année 2023.

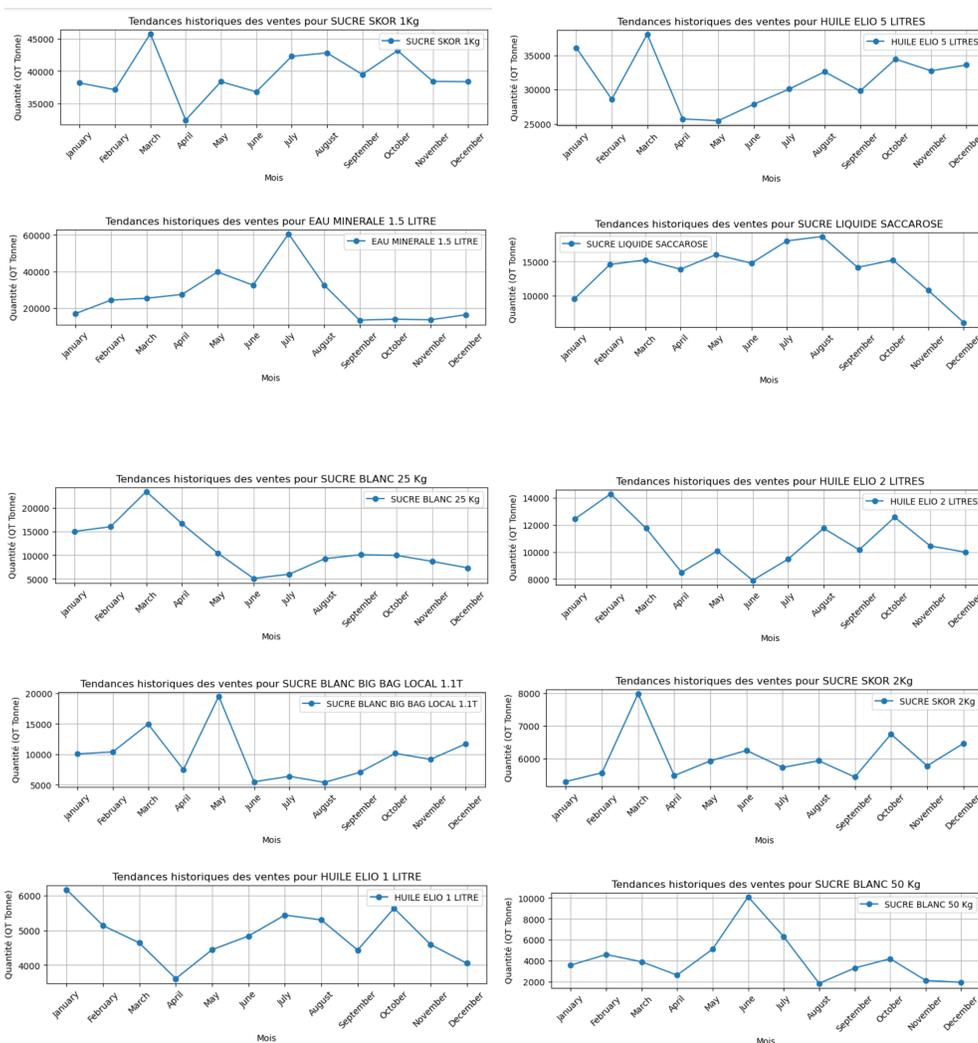


FIGURE 3.12 – Tendances historiques des ventes pour chaque produit

On a repéré que le mois de mars a connu des pics de ventes importants, tels que le SUCRE SKOR 1KG, HUILE HELIO 5L, SUCRE BLANC 25KG et SUCRE SKOR 2KG. En revanche, certains produits tels que EAU MINIRALE 1.5LITRES montrent des ventes réduites en novembre et décembre.

e- Évaluation mensuelle des performances de vente

Nous avons effectué une évaluation mensuelle des performances de vente, en mettant en évidence le chiffre d'affaires total (CA TTC) et la quantité vendue en tonnes (QT Tonne) pour chaque mois de l'année.

Dans la figure (3.13), un graphique qui illustre les résultats :

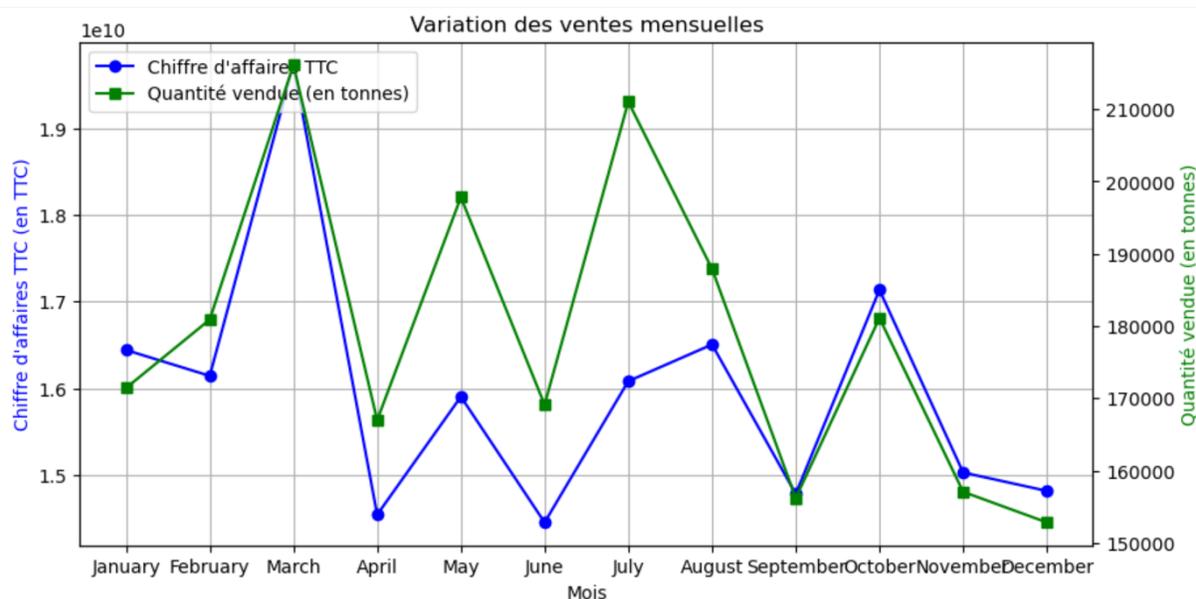


FIGURE 3.13 – Évaluation mensuelle des performances de vente 2023

Les résultats montrent des variations importantes au cours de l'année. Le chiffre d'affaires varie chaque mois, avec des montants allant de 14.45 milliards en juin à un pic notable de 19.73 milliards en mars. De même, les quantités vendues fluctuent au fil des mois, avec des niveaux plus élevés en mars (216128 tonnes) et en juillet (211023 tonnes), et des niveaux plus bas en décembre (152827 tonnes). En résumant, le mois de mars est particulièrement rentable en terme de chiffre d'affaires total et de quantité vendue.

f- Répartition des quantités vendues par ville

Nous avons regroupé les données par ville et calculé les quantités totales en tonnes pour chaque ville. Le diagramme à barres présenté par la figure (3.14) montre comment les quantités vendues sont réparties par ville :

CHAPITRE 3. IMPLÉMENTATION

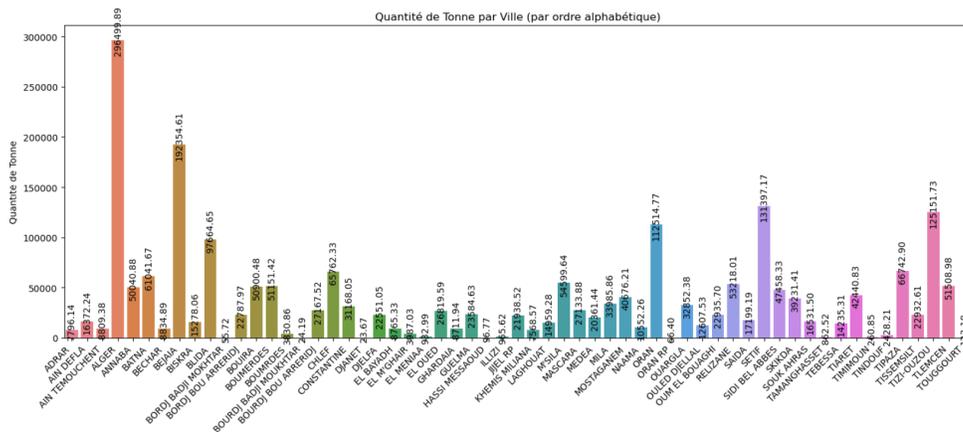


FIGURE 3.14 – Répartition des Quantités vendues par Ville

En visualisant le diagramme, On remarque une augmentation significatives du volume de vente dans les grandes villes telles qu’Alger, Oran et Béjaia par rapport à d’autres villes. Le volume de vente d’Alger est particulièrement remarquable avec 296499 tonnes, suivi de Béjaia avec 192354 tonnes et de sétif avec 131397 tonnes. Ces résultats indiquent une forte demande dans ces grandes agglomérations, probablement en raison de leur densité de population et de leur pouvoir d’achat élevé.

g- Les clients les plus importants en termes de quantités vendues

Nous allons effectuer une étude du volume d’achat par client afin d’identifier les clients les plus performants. Les résultats sont illustrés dans l’histogramme présenté par la figure (3.15) :

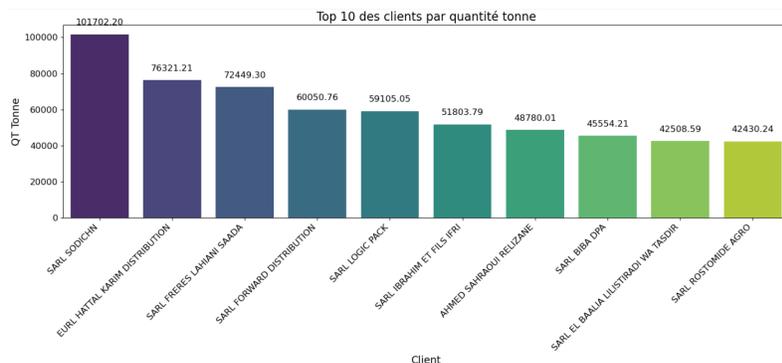


FIGURE 3.15 – Les clients les plus importants en termes de quantités vendues

Les résultats montrent que SARL SODICHN est le principal acheteur avec un volume d'achat de 101702 tonnes, suivi par SARL HATTAL KARIM DISTRIBUTION avec un volume de 76321 tonnes. Tandis que, les autres clients dans le top 10 ont des volumes d'achat variant de 42430 tonnes à 72449 tonnes.

3.5 Modélisation

Dans cette phase, l'objectif est de développer un modèle prédictif précis pour les ventes de l'année 2024 en utilisant les données historiques de l'année 2023.

3.5.1 Séparation des données

Nous avons divisé nos données en ensembles d'entraînement et de test. Cette séparation a alloué 80% des données à l'ensemble d'entraînement et 20% à l'ensemble de test, conformément à la spécification du paramètre `'test_size=0.2'` dans la fonction `'train_test_split'`. Les variables indépendantes incluent les caractéristiques telles que `'CA TTC'`, `'CA HT'`, `'CA HT DZD'`, `'QTE_PLT'`, `'Produit'`, `'NOM_MOIS'`, `'NUM_ANNEE'`, `'National_Etranger'`, `'Ville'`, `'Famille'`, `'Sousfamille'`, `'QT US'`, `'statutLivraison'`, `'statutFacturation'`, `'DaysSinceReference'` tandis que la variable cible est `'QT Tonne'`.

3.5.2 Choix de modèle

On note que notre problème s'agit d'un problème de régression, alors nous avons opté à utiliser les forêts aléatoires car elles sont destinées à gérer les grands ensembles de données et le surajustement n'est pas un problème dans cet algorithme.

3.5.3 Entraînement de modèle

Dans cette étape, nous avons créé un pipeline à l'aide de la classe `'Pipeline'` de `scikit-learn`. Ce pipeline comprenant un préprocesseur (inclure des transformation des données) et un modèle `RandomForestRegressor`. Ensuite, le modèle est entraîné sur l'ensemble d'entraînement en utilisant la méthode `'fit'`.

Voici la figure (3.16) présente les résultats :

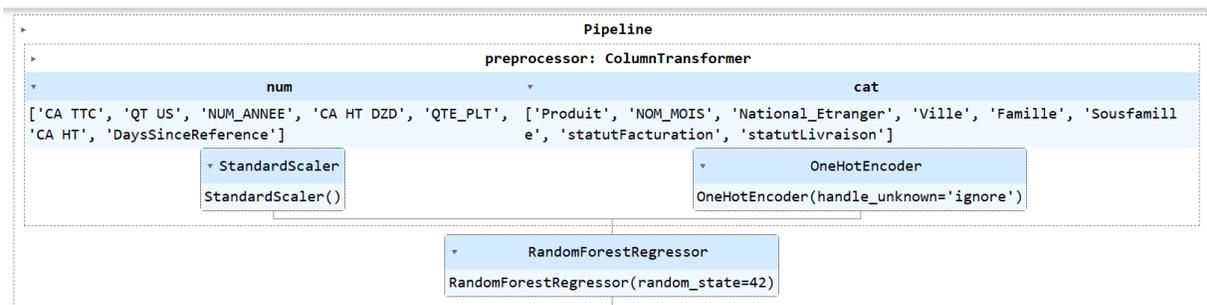


FIGURE 3.16 – Entraînement de modèle

Les résultats du pipeline montrent qu’au cours de l’étape de Préprocesseur, les variables numériques telles que ‘CA TTC’, ‘QT US’, ‘NUM_ANNEE’, ‘CA HT DZD’, ‘QTE_PLT’, ‘CA HT’, ‘DaysSinceReference’ sont standardisées à l’aide de la fonction StandardScaler(). Tandis que, les variables catégorielles comme ‘Produit’, ‘NOM_MOIS’, ‘National_Etranger’, ‘Ville’, ‘Famille’, ‘Sousfamille’, ‘statutFacturation’, ‘statutLivraison’ sont encodées à l’aide de la fonction ‘OneHotEncoder(handle_unknown=‘ignore’)’. Ensuite, le modèle RandomForestRegressor est entraîné sur les données prétraitées.

3.5.4 Évaluation et optimisation du modèle avec GridSearchCV

Dans cette partie, nous avons évalué les performances du modèle RandomForestRegressor avec une validation croisée à 5 folds en utilisant cross_val_score de scikit-learn. Ensuite, nous définissons une grille de paramètres à explorer lors de l’optimisation du modèle. Les hyperparamètres ‘n_estimators’, ‘max_depth’, et ‘min_samples_split’ sont définis avec différentes valeurs pour rechercher la meilleure combinaison en utilisant ‘GridSearchCV’. Enfin, le modèle est évalué à l’aide de la validation croisée pour évaluer ses performances sur différentes métriques de performance telles que MSE, MAE et le coefficient de détermination (R^2). Les résultats de la validation croisée sont ensuite affichés dans la figure (3.17) :

```

Best parameters: {'model__n_estimators': 300, 'model__max_depth': 30, 'model__min_samples_split': 2}
Mean Squared Error (MSE) with Cross-Validation: 0.004987675983564734
Mean Absolute Error (MAE) with Cross-Validation: 0.0030667320780194693
Coefficient de détermination ( $R^2$ ) with Cross-Validation: 0.999939445151707
    
```

FIGURE 3.17 – Les résultats d’évaluation avec la validation croisée

Les résultats indiquent que les meilleurs hyperparamètres trouvés sont Nombre d’estimateurs (n_estimators) : 300, profondeur maximale (max_depth) : 30 et nombre minimal

d'échantillons pour diviser un nœud (`min_samples_split`) : 2. Les métriques de performance obtenues sont les suivantes :

- Erreur quadratique moyenne (MSE) : 0.004987675983564734 ;
- Erreur absolue moyenne (MAE) : 0.0030667320780194693 ;
- Coefficient de détermination (R^2) : 0.999939445151707.

D'après les résultats, le modèle ajusté avec les meilleurs hyperparamètres obtenus indique une performance excellente.

3.5.5 Évaluation du modèle sur l'ensemble de test

Cette partie elle est considérée comme la dernière étape où le modèle est entraîné sur l'ensemble de l'entraînement complet. Par la suite, nous employons le modèle entraîné afin de faire des prédictions sur l'ensemble des tests. Les résultats du modèle sur l'ensemble des tests sont évalués en calculant les mesures suivantes :

```
Mean Squared Error (MSE) on Test Set: 0.0035857497876937336
Mean Absolute Error (MAE) on Test Set: 0.002380634239872585
Coefficient de détermination ( $R^2$ ) on Test Set: 0.9999565343367927
```

FIGURE 3.18 – Les résultats d'évaluation du Modèle sur l'Ensemble de Test

Les résultats indiquent que le modèle est très performant avec un erreur quadratique moyenne (MSE) et un erreur absolue moyenne (MAE) très faible ce qui signifie que les prédictions du modèle sont très proches des valeurs réelles. Le coefficient de détermination (R^2) est proche de 1, ce qui suggère que le modèle explique très bien la variance des données et est donc un ajustement très précis aux données d'entraînement.

3.5.6 Prévisions des ventes pour l'année 2024

Nous présentons les résultats de nos prédictions pour les ventes de l'année 2024, obtenus par le modèle dans la figure (3.19) :

NOM_MOIS	QT Tonne prédit pour 2024
January	171462.251516
February	192291.188170
March	217691.633836
April	181759.263794
May	200327.944108
June	171374.340724
July	213749.934724
August	190731.022531
September	159250.224975
October	183940.458615
November	159882.745459
December	155710.165481

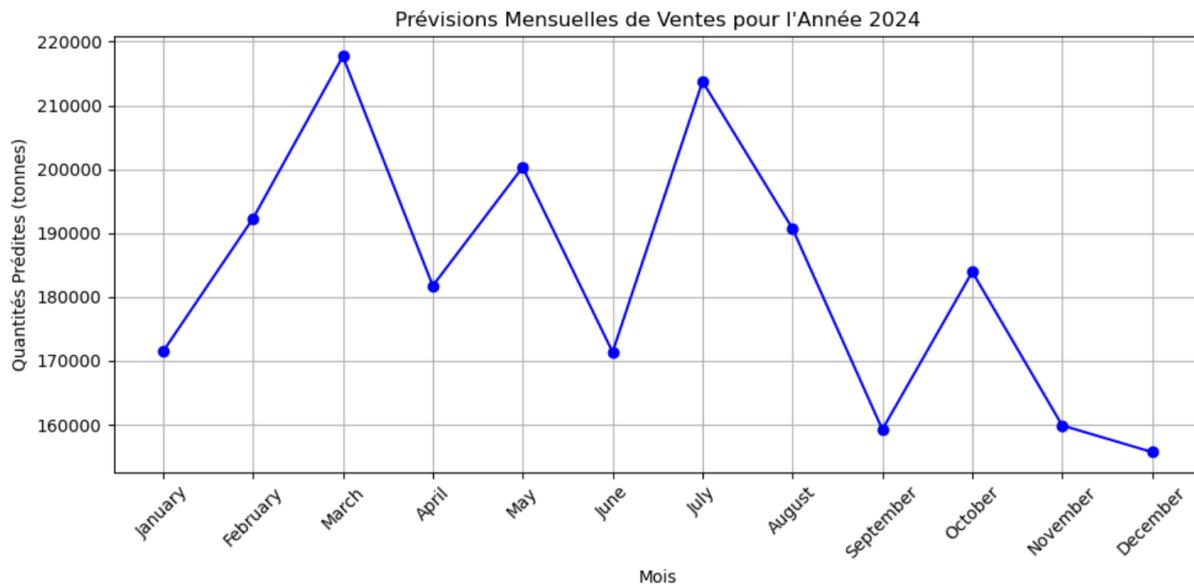


FIGURE 3.19 – Prévisions des ventes menseulles pour l'année 2024

Ces prévision fournissent une estimation des quantités de produits vendus pour chaque mois de l'année 2024. En obsarvant les chiffres, le mois de mars et juillet affichent les quantités les plus élevées, dépassant les 200 000 tonnes. Sache que, les mois de septembre, novembre et décembre ont les quantités les plus basses, toutes en dessous de 160 000 tonnes.

3.5.7 Etude comparative entre les prédictions et les quantités réelles 2024

Pour légitimer notre étude, nous comparons les prévisions obtenues avec les résultats concrets de 2024 (nous ne disposons que de janvier, février, mars et avril). Dans le rapport ci-dessus, on peut observer la comparaison :

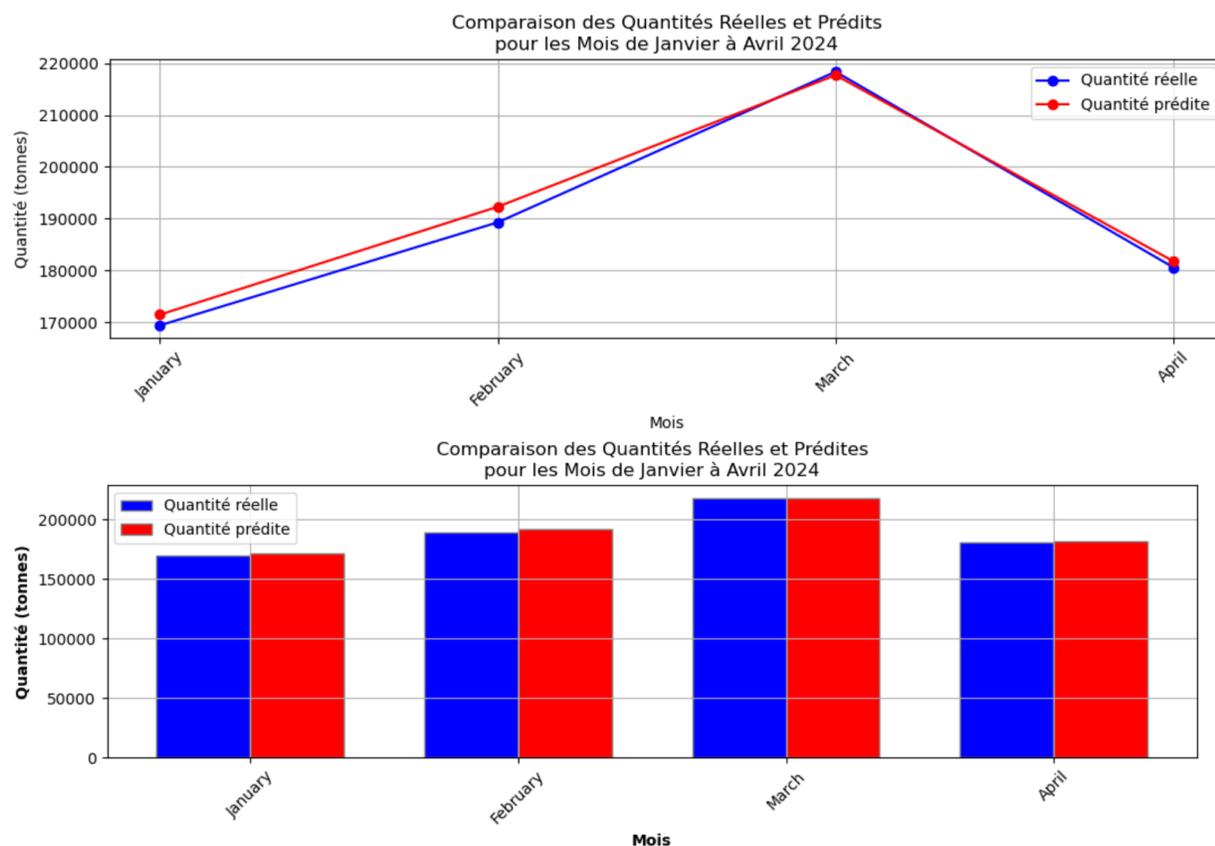


FIGURE 3.20 – Etude comparative entre les prédictions et les quantités réelles 2024

En visualisant les résultats, Les prédictions du modèle sont très proches des valeurs réelles, montrant une excellente précision. Les petites différences indiquent que le modèle est bien calibré et capable de fournir des prévisions fiables pour les ventes mensuelles.

II. Création d'un tableau de bord itératif : Prévisions des ventes 2024

Introduction

Dans cette partie dédiée à la création d'un tableau de bord itératif pour les prédictions de ventes de l'année 2024 de Cevital, nous avons utilisé plusieurs types de graphiques pour représenter les prédictions de la QT Tonne par rapport aux variables NOM_MOIS, Produit, Famille, Sousfamille et Ville, dont l'objectif est de fournir une plateforme dynamique permettant d'explorer les tendances de ventes, ce qui facilite la prise de décision stratégique.

3.6 Outil de Visualisation : "Power BI"

Power BI est l'outil utilisé pour la visualisation des résultats et la création de rapports interactifs pour visualiser les tendances des ventes prévues, les écarts par rapport aux ventes réelles, et d'autres métriques clés. C'est une plateforme unifiée pour le décisionnel (BI) en entreprise et libre-service.

Intégration des Données

Power BI Desktop nous permet de se connecter aux données de nombreuses sources différentes. En cliquant sur le bouton "Obtenir des données", une nouvelle fenêtre apparaîtra avec diverses options de sources de données, on sélectionne "Excel".

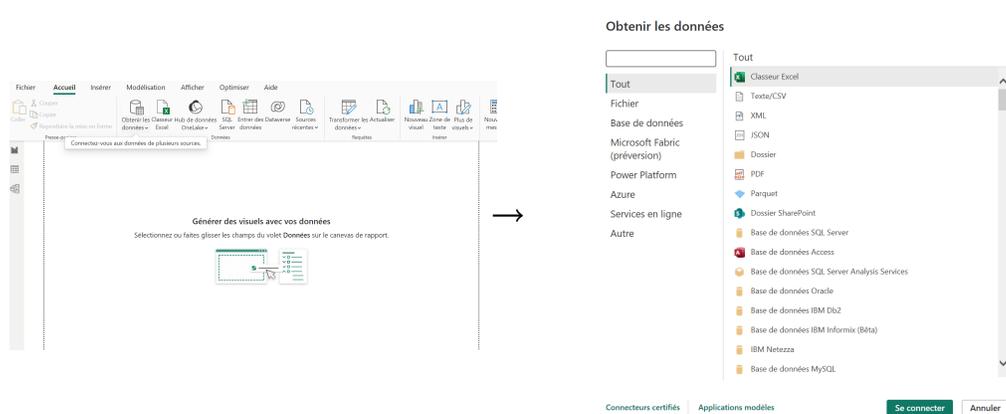


FIGURE 3.21 – La source de données (Excel)

CHAPITRE 3. IMPLÉMENTATION

Après avoir connecté , dans la boîte de dialogue Navigateur nous allons sélectionner la tables que nous utiliserons pour les importer dans Power BI Desktop puis nous cliquons sur le bouton "Charger".

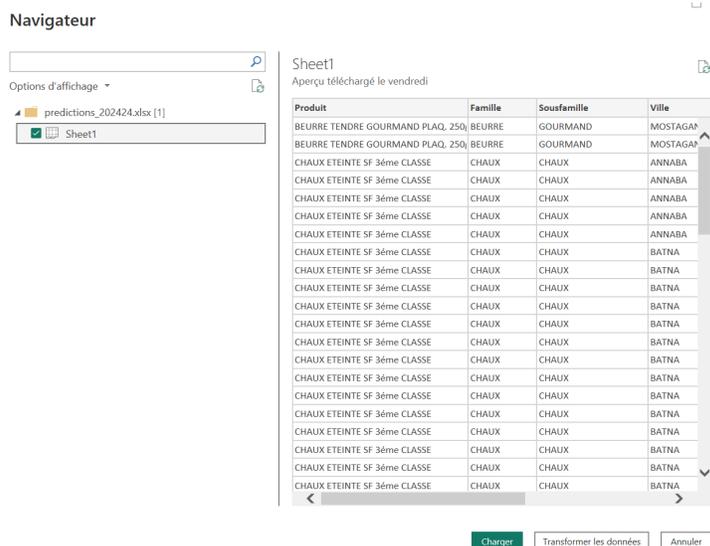


FIGURE 3.22 – Table de données

Une fois la table chargée, le volet données montre les informations.

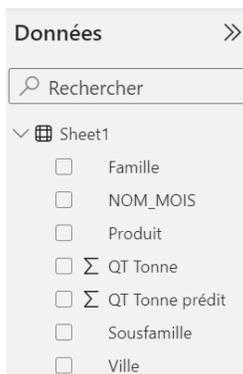


FIGURE 3.23 – Affichage des données

Maintenant, nous pouvons créer des visuels pour notre tableau de bord.

3.7 Construction de tableau de bord

Au cours de cette partie, nous allons concevoir le tableau de bord en incluant des illustrations de données sur la page du rapport.

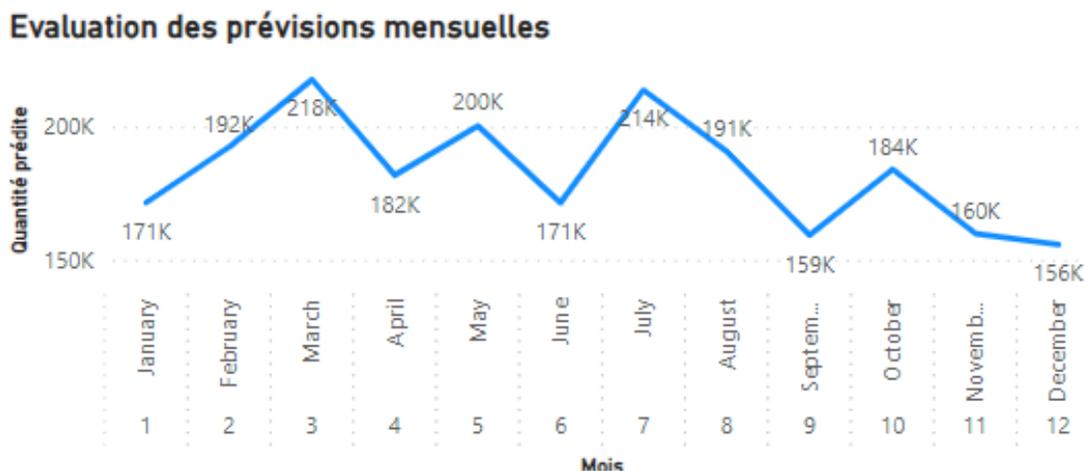


FIGURE 3.25 – Evaluation des prévisions mensuelles

En observant le graphique, on constate que le pic est enregistré en mars avec 217691 tonnes, suivi de juillet avec 213749 tonnes, après avoir été très faible en décembre.

b. Graphique en secteur : Répartition des quantités prédites par produit

Le graphique en secteur nous permet d'illustrer la répartition des quantités prédites pour chaque produit pour l'année 2024.

Dans la page du rapport, on sélectionne le graphique en secteur, pour la légende, on fait glisser "Produit", et pour la rubrique valeurs, on fait glisser "Somme de QT Tonne prédit" de volet de données.

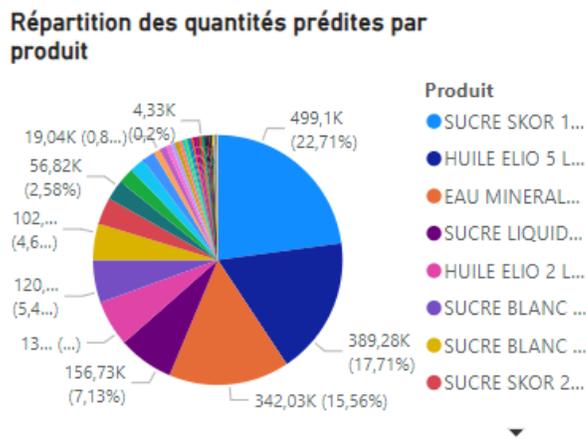


FIGURE 3.26 – Répartition des quantités prédites pour chaque produit pour l'année 2024

Chaque secteur, représente un pourcentage total de chaque produit. On remarque le SUCRE SKOR 1KG atteint une quantité maximale de 499.1K tonnes avec un pourcentage de 22.71%, l'HUILE ELIO 5 LITRES atteint un quantité assez notable de 389.28K tonnes avec un pourcentage de 17.71%, aussi l'EAU MINIRALE 1.5LITRES atteint une quantité de 342.03K avec un pourcentage de 15.56%. Tandis que, les autres produits varient entre 7.13% et 0.2%.

c. Histogramme groupé : Quantités prédites par famille

Nous avons utilisé l'histogramme groupé afin de représenter les quantités prédites par famille de produits pour l'année 2024.

Dans la page du rapport, on sélectionne l'histogramme groupé , pour l'Axe X, on fait glisser "Famille", et pour l'Axe Y, on fait glisser "Somme de QT Tonne prédit" de volet de données.

Répartition des quantités prédites par Famille

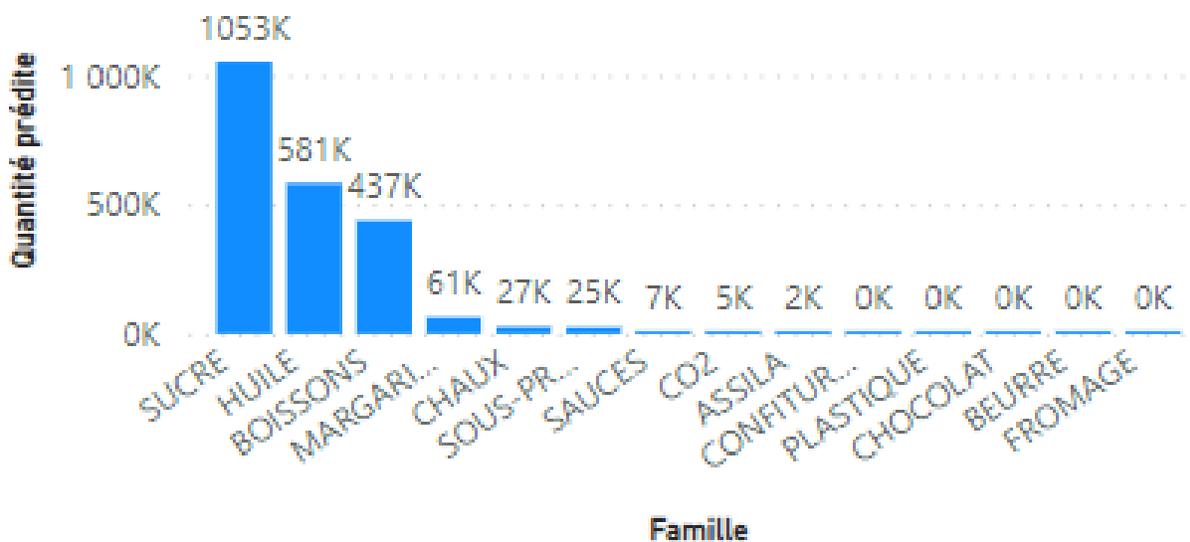


FIGURE 3.27 – Répartition des quantités prédites par famille de produits pour l'année 2024"

En observant l'histogramme, on remarque la famille de SUCRE prend la première place avec un total de 1 million de tonnes, suivi de l'HUILE avec une quantité de 580.56K tonnes, la BOISSON avec une quantité remarquable de 436.52K tonnes, tandis que, les autres familles varient de 19 tonnes jusqu'au 61K tonnes.

d. Graphique à barre groupé : Répartition des prévisions de ventes par Sous-famille

Le graphique à barre visualise la répartition des prévisions de ventes par Sous-famille de produits pour l'Année 2024.

On sélectionne le graphique à barre groupé, pour l'Axe Y, on fait glisser "Sousfamille", et pour l'Axe X, on fait glisser "Somme de QT Tonne prédit" de volet de données.

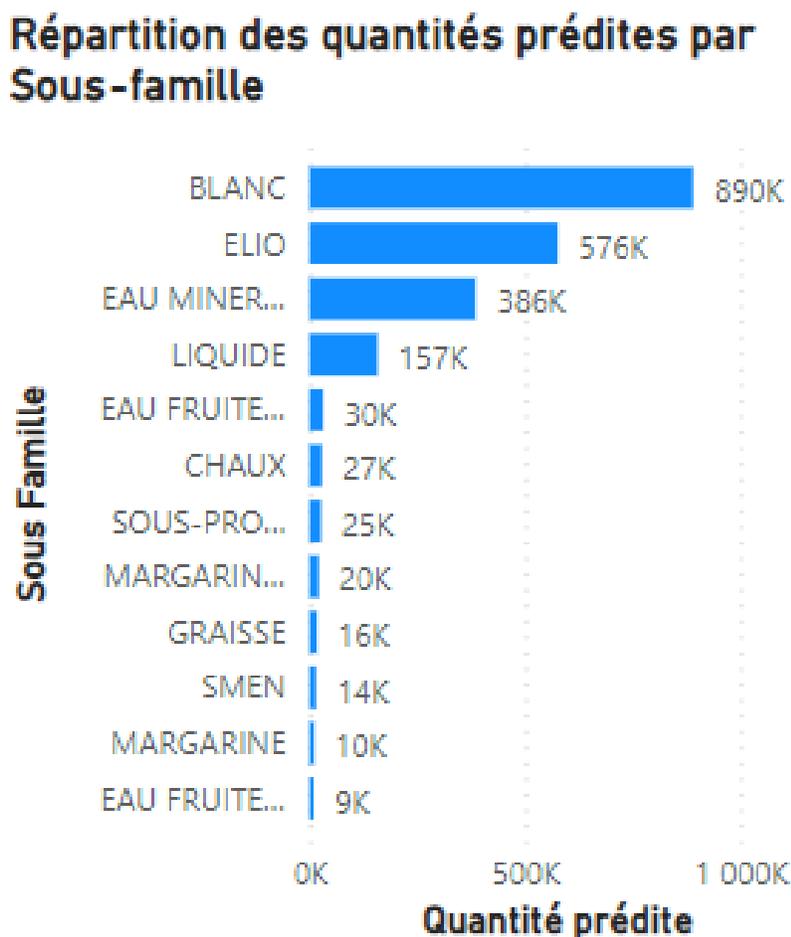


FIGURE 3.28 – Répartition des quantités prédites par Sous-famille de produits pour l'année 2024

Le résultat démontre que la sous-famille "BLANC" produit la quantité la plus importante de 889.89k tonnes, suivie de "ELIO" avec 576,32k tonnes et de l'EAU MINIRALE avec 385.84k tonnes.

e. Carte géographique : Répartition des prévisions de ventes par ville

La figure (3.30), représente la répartition géographique des prévisions de ventes par ville pour l'Année 2024.

On sélectionne la carte géographique, pour l'Axe Y, on fait glisser "Sousfamille", et pour l'Axe X, on fait glisser "Somme de QT Tonne prédit" de volet de données.

Répartition géographique des quantités prédites par ville

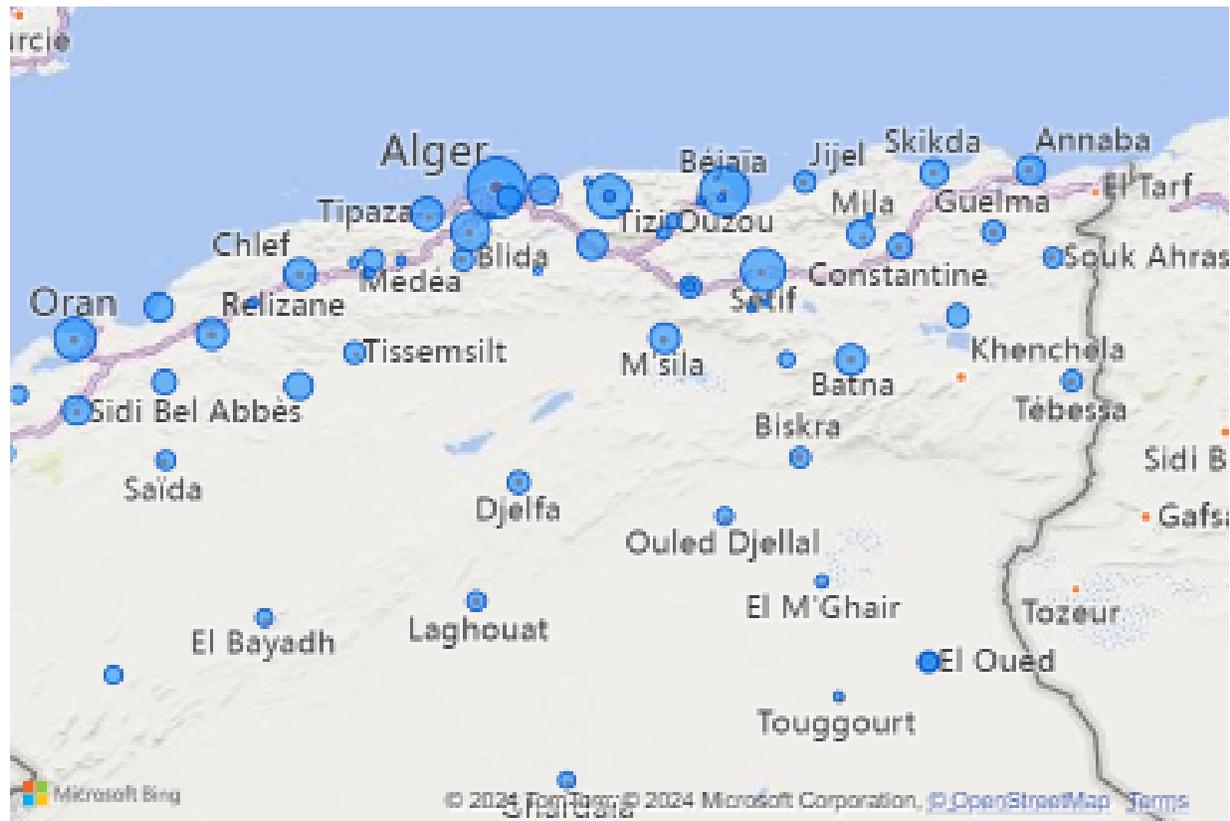


FIGURE 3.29 – Répartition géographique des quantités prédites par ville

La carte montre que les grands volume de quantité sont générés dans les grandes ville telles que Alger, Bejaia, Oran, Sétif, Tizi Ouzou.

CHAPITRE 3. IMPLÉMENTATION

3. Tableau de bord itératifs des prévisions de ventes 2024 :

La figure (3.31), présente notre tableau de bord final :

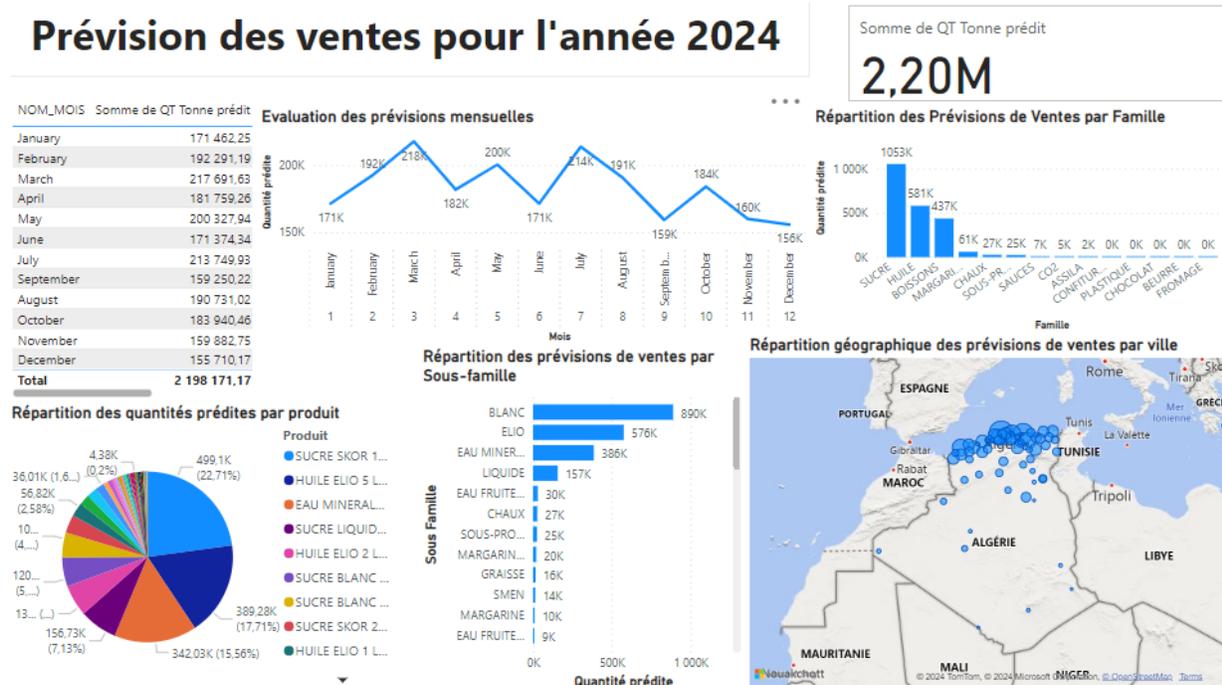


FIGURE 3.30 – Tableau de bord itératifs des prévisions de ventes 2024

Le tableau de bord inclut toutes les visualisations effectuées, ainsi que la quantité totale qui s'élève à 2,2 millions de tonnes.

Conclusion

Au cours de ce chapitre, nous avons abordé deux aspects essentiels : prévision des ventes pour l'année 2024 et création d'un tableau de bord itératif des ventes prévisionnelles 2024. Nous avons présenté en détails l'utilisation de modèle des forêts aléatoires en python afin de réaliser des prédictions précise en basant sur des données historiques. Nous avons aussi construit un tableau de bord itératif en utilisant "Power BI", ce tableau intègre les ventes prévisionnelles et offre une visualisation claire, facilitant ainsi la prise de décision.

Conclusion générale

En conclusion, nous avons exploré en profondeur l'implémentation d'un modèle d'apprentissage automatique pour la prévision des ventes de Cevital pour l'année 2024. Nous avons débuté par une analyse sur l'entreprise d'accueil, en soulignant son histoire, sa structure organisationnelle et ses activités principales.

Ensuite, nous avons plongé dans le monde complexe de l'apprentissage automatique, explorant ses principes fondamentaux, ses différents types d'algorithmes et son processus de développement. Nous avons également étudié son utilisation dans la prévision des ventes, en examinant l'importance et les avantages de cette dernière.

Dans la partie pratique, nous avons décrit en détail l'implémentation pratique du modèle d'apprentissage automatique pour la prévision des ventes de Cevital pour l'année 2024 en se basant sur des données historiques de 2023. Nous avons souligné l'utilisation de Python pour la création du modèle et de Power BI pour générer un tableau de bord itératif, en exposant en détail les étapes essentielles du processus.

En somme, ce mémoire a démontré comment l'intégration des techniques de l'apprentissage automatique peut améliorer de manière significative la précision des prévisions des ventes, offrant ainsi à Cevital et à d'autres entreprises des insights pour optimiser leur gestion des stocks, planifier leurs opérations et maximiser leur rentabilité.

Bibliographie

- [1] <https://www.cevital.com/>
- [2] A. Géron. *Machine Learning avec scikit-learn*. 2017.
- [3] G. Hackeling. *Mastering Machine Learning with scikit-learn Second Edition*. Packt Publishing Ltd, Livery Place, UK, Juin 2017.
- [4] M. Taffar. *Initialisation à l'apprentissage automatique*. Université de Jijel, Département Informatique, Algérie, 2014.
- [5] G. Battineni, N. Chintalapudi, F. Amenta. *Apprentissage automatique en médecine : calcul des performances de la prédiction de la démence par les machines à vecteurs de support (SVM)*. L'informatique en médecine déverrouillée, vol. 16, 2019, pp. 100200.
- [6] M. Comlan, M. Alokpo. *Classification of Gestational Diabetes Using Machine Learning Techniques*. Applied Mathematics, Modeling and Computer Simulation, IOS Press, 2023, pp. 1224-1234.
- [7] J. B. O. Mitchell. *Méthodes d'apprentissage automatique en chimioinformatique*. Revues interdisciplinaires Wiley : Computational Molecular Science, vol. 4, no. 5, 2014, pp. 468-481.
- [8] L. Candillier. *Contextualisation, visualisation et évaluation en apprentissage non supervisé*. Thèse de doctorat, Université Charles de Gaulle-Lille III, 2006.
- [9] M. Zimmer. *Apprentissage par renforcement développemental*. Thèse de doctorat, Université de Lorraine, 2018.
- [10] J.-C. Usunier, R. Bourbonnais. *Prévision des Ventes : Théorie et pratique*. No. hal-01507734, 2017.
- [11] P. Hajek, L. Hikkerova. *L'avantage de l'intelligence artificielle et de la prise en compte du sentiment des investisseurs dans la prévision des prix du pétrole en période de crise*. Management & Avenir, 5 (2023)
- [12] I. El Naqa, M. J. Murphy. *Qu'est-ce que l'apprentissage automatique ?*. Éditions Springer International, 2015.

Résumé

Résumé

L'accélération spectaculaire de l'intelligence artificielle ces dernières années, notamment le Machine Learning (ML), incite les entreprises à incorporer ces nouvelles technologies afin d'améliorer la gestion, de répondre aux diverses problématiques et de prendre des décisions stratégiques. C'est dans ce contexte que s'introduit le travail présenté dans ce mémoire qui consiste à l'implémentation d'un modèle machine learning pour la prévision des ventes au sein de Cevital, Béjaïa. Notre objectif est de fournir à Cevital un outil précieux pour prévoir les demandes futures, optimiser les stocks et affiner leurs stratégies commerciales. Plus spécifiquement, nous avons opté pour le modèle des forêts aléatoires, réputé pour sa capacité à gérer des ensembles de données complexes tout en offrant des prédictions précises. En complément, Power BI a été intégrée pour fournir des visualisations interactives et intuitives, permettant ainsi une analyse approfondie des données et des prises de décision stratégiques informées. Les prédictions du modèle sont très proches des valeurs réelles, montrant une excellente précision.

Mots clés :

Apprentissage automatique, Forêts aléatoires, Prévision des ventes, Python, Analyse des données, Power BI, Prise de décision, Visualisations interactives.

الملخص

يشجع التسارع المذهل للذكاء الاصطناعي في السنوات الأخيرة، ولا سيما التعلم الآلي، الشركات على دمج هذه التقنيات الجديدة من أجل تحسين الإدارة والاستجابة لمختلف المشاكل واتخاذ القرارات الاستراتيجية. وفي هذا السياق تم تقديم العمل المقدم في هذه الأطروحة، والذي يتمثل في تطبيق نموذج التعلم الآلي للتنبؤ بالمبيعات في شركة سيفيتال بجاية. هدفنا هو تزويد شركة سيفيتال بأداة قيمة للتنبؤ بالطلب المستقبلي وتحسين المخزون وتحسين استراتيجياتها التجارية. وبشكل أكثر تحديداً، اخترنا نموذج الغابة العشوائية، المعروف بقدرته على التعامل مع مجموعات البيانات المعقدة مع تقديم تنبؤات دقيقة. بالإضافة إلى ذلك، تم دمج برنامج Power BI لتوفير تصورات تفاعلية وبدئية، مما يتيح إجراء تحليل متعمق للبيانات واتخاذ قرارات استراتيجية مستنيرة. تنبؤات النموذج قريبة جداً من القيم الفعلية، مما يدل على دقة ممتازة.

الكلمات المفتاحية:

التعلم الآلي، الغابات العشوائية، التنبؤ بالمبيعات، بايثون، تحليل البيانات، Power BI، اتخاذ القرارات، التصورات التفاعلية.

Abstract

The spectacular acceleration of artificial intelligence in recent years, particularly Machine Learning (ML), is prompting companies to incorporate these new technologies to improve management, respond to various issues and make strategic decisions. This is the background to the work presented in this thesis, which consists in implementing a machine learning model for sales forecasting at Cevital, Béjaïa. Our aim is to provide Cevital with a valuable tool for forecasting future demand, optimizing inventories and refining their sales strategies. More specifically, we opted for the random forest model, renowned for its ability to handle complex data sets while offering accurate predictions. In addition, Power BI was integrated to provide intuitive, interactive visualizations, enabling in-depth data analysis and informed strategic decision-making. Model predictions are very close to actual values, demonstrating excellent accuracy.

Keywords :

Machine learning, Random forests, Sales forecasting, Python, Data analysis, Power BI, Decision-making, Interactive visualizations.
