

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 système centralisé intelligent pour l'amélioration du traitement des réclamations chez BMT

Présenté par :

Achouri Agnes

Lazizi Melissa

Sous la direction de : **Dr.M Moktefi**

Défendu le 01/07/2024, devant le jury composé de :

M ^{rs} K. Hassaini	M.A.A	Président de jury	UAMB - Bejaia.
M ^r L. Asli	M.C.A	Examineur	UAMB - Bejaia.
M ^{rs} Z. Bouzeria	Doctorante	Examineur	UAMB - Bejaia.
M ^r A. Boudra	Encadrant	Invité	Entreprise BMT.

Année Universitaire 2023 – 2024

Remerciements

Nous tenons à exprimer notre gratitude la plus sincère à Dieu Tout-Puissant pour nous avoir accordé la force et la persévérance nécessaires à l'accomplissement de ce travail.

Nous souhaitons exprimer notre profonde reconnaissance à **Mr. BOUDRA Adel**, responsable de la section recouvrement au service commercial de Bejaia Mediterranean Terminal (BMT). Son dévouement et ses conseils avisés ont été déterminants à chaque étape, et sa compréhension des enjeux de l'entreprise ainsi que sa disponibilité ont été essentielles à la réussite de ce projet.

Nous adressons nos remerciements les plus chaleureux à **Mme. KOURTA Hadjila**, chef du service commercial, pour son accueil bienveillant et son soutien constant, ainsi qu'à tous les membres du service commercial pour leur précieuse collaboration.

Nos sincères remerciements vont à **Mr. ASLI Larbi**, responsable de la Spécialité Sciences de Données et Aide à la Décision, pour son soutien inébranlable et ses encouragements.

Nous tenons également à exprimer notre gratitude à nos encadreurs, à l'université **Mr. MOKTEFI Mohand** et en stage **Mr. BOUMERZOUG Moussa**, pour leur accompagnement et leurs conseils avisés tout au long de ce parcours.

Nos remerciements vont également à tous nos enseignants de l'université pour leurs efforts inlassables à nous fournir une formation de qualité et pour leur sérieux tout au long de notre parcours universitaire.

Dédicace

À ma chère maman et mon cher papa,

Ce travail est dédié à vous, pour tout ce que vous avez sacrifié et offert pour mon bonheur et ma réussite. Votre amour et votre soutien ont été ma lumière et ma motivation.

À mon petit frère Amine,

Merci pour ton soutien inconditionnel et ton inspiration constante.

À mes amies Sarah et Manel,

Vous êtes mes sœurs de cœur, merci pour votre amitié sincère.

À ma binôme,

Merci d'avoir été mon bras droit et notre réussite commune.

À vous tous,

Merci d'avoir rendu ce chemin si enrichissant et mémorable.

ACHOURI Agnes.

dédicace

A mes chers parents, dont le soutien inébranlable et les sacrifices ont rendu ce parcours possible. Votre amour et vos encouragements m'ont donné la force de persévérer et de poursuivre mes rêves.

A mes amis, compagnons de tant d'aventures, qui ont été des piliers d'encouragement et de soutien tout au long de cette période intense. Vos conseils et votre amitié ont rendu ce voyage non seulement possible, mais aussi enrichissant.

À ma binôme, Merci d'avoir été mon bras droit et notre réussite commune.

Enfin, à ma famille élargie et à tous ceux qui, de près ou de loin, ont contribué à mon éducation et à mon développement personnel. Vous avez tous joué un rôle crucial dans cette réussite.

Ce travail est dédié à chacun de vous, pour votre soutien inestimable et votre amour indéfectible.

LZIZI Melissa.

Table des matières

Liste des figures	IX
Liste des tables	X
Liste d'abréviations et notations	XI
Introduction générale	1
1 Généralités sur la Business Intelligence Et Machine Learning	4
1.1 Introduction	4
1.2 Business Intelligence	5
1.2.1 Historique	5
1.2.2 Définition de la BI :	5
1.2.3 Fonctionnement du processus de la BI	5
1.2.4 Notions de base en Informatique Décisionnelle(BI)	7
1.2.5 Enjeux de la BI pour les entreprises	8
1.2.6 Importance de la business intelligence pour l'entreprise	8
1.2.7 Étapes clés pour réussir un projet de BI	9
1.2.8 Domaines d'application de la BI	10
1.2.9 Utilité des Données Collectées dans un Projet de BI	10
1.2.10 Raisons de Choisir un Logiciel ou un Outil de Business Intelligence (BI)	11
1.2.11 Avantages de la BI	11
1.2.12 Inconvénients de la BI	11
1.3 Machine Learning	12
1.3.1 Définition des notions	12
1.3.2 Types d'algorithmes de Machine Learning	13
1.3.3 Algorithmes des machines Learning	13
1.3.4 Algorithmes de classification	14
1.3.5 Évaluation de la performance des algorithmes de classification	18
1.4 Conclusion	19
2 Contexte de l'Entreprise et Approche Méthodologique de la Collecte des Données	21
2.1 Introduction	21
2.2 Présentation de l'entreprise d'Accueil	22
2.2.1 Historique	22
2.2.2 Situation Géographique	22
2.2.3 Présentation de BMT-Spa	23
2.2.4 Organigramme de BMT	24
2.2.5 Les valeurs de la BMT (Bureau des Messageries Maritimes)	24

2.2.6	Présentation détaillée du champ d'étude	25
2.3	Procédure actuelle de traitement des réclamations à BMT	26
2.3.1	Procédure	26
2.3.2	Défis potentiels dans la procédure actuelle de traitement des réclamations clients à BMT	27
2.4	Collecte des Données	28
2.4.1	Définition de la Collecte de Données	28
2.4.2	Objectifs Principaux de la Collecte de Données	28
2.4.3	Processus de Collecte de Données	28
2.5	Description du Dataset Collecté	30
2.6	choix de l'outil Power BI	32
2.6.1	Définition	32
2.6.2	Fonctionnement de Power BI	33
2.6.3	Comparaison entre Power BI et Excel	33
2.7	Conclusion	34
3	Analyse et conception	35
3.1	Introduction	35
3.2	Analyse des besoins	36
3.2.1	Présentation d'UML	36
3.2.2	Quelques définitions de base	36
3.2.3	Contexte du projet	36
3.2.4	Identification des acteurs	36
3.3	Identification des besoins	37
3.3.1	Besoins fonctionnels	37
3.3.2	Besoins non fonctionnels	38
3.4	Diagramme de contexte statique	38
3.5	Description des cas d'utilisation	39
3.5.1	Description du cas d'utilisation « MES RÉCLAMATIONS »	39
3.5.2	Description du cas d'utilisation « GESTION DES RÉCLAMATIONS »	40
3.6	Diagrammes de cas d'utilisation	41
3.6.1	Diagramme de cas d'utilisation pour la Gestion Mes Réclamations	41
3.6.2	Diagramme de cas d'utilisation pour la Gestion des Réclamations	42
3.6.3	Diagramme de cas d'utilisation totale	43
3.7	Diagrammes d'activités	44
3.7.1	Diagramme d'activité pour le cas d'utilisation « Authentification » :	44
3.7.2	Diagramme d'activité pour le cas « Ajouter Réclamations »	45
3.7.3	Diagramme d'activité pour le cas « Gérer Réclamations »	45
3.7.4	Diagramme d'activité pour le cas « Prédire Réclamations »	46
3.8	Modèle de domaine	47
3.9	conception	47
3.9.1	Diagramme de séquence	47
3.9.2	Diagramme de classes	50
3.9.3	Conception de la base de données	51
3.9.4	Modèle Relationnel	53
3.10	Conclusion	54

4	Implémentation de la Solution Analytique et Prédicative	55
4.1	Introduction	55
4.2	Élaboration des indicateurs de performance essentiels intégrés dans notre solution BI	56
4.3	Implémentation de Solutions Power BI	57
4.3.1	Installation et Configuration de Power BI	57
4.3.2	Collecte et préparation de données	58
4.3.3	Modélisation des données	65
4.3.4	Conception de rapport	66
4.4	Implémentation de solution de prediction	77
4.4.1	Étapes de Réalisation du Code	77
4.4.2	Résultats des classificateurs	80
4.4.3	Analyse des Résultats et Sélection du Meilleur Modèle	81
4.5	Réalisation d'une application de Prédiction avec Streamlit	82
4.5.1	Définition de Streamlit	82
4.5.2	Avantages de Streamlit	82
4.5.3	Inconvénients de Streamlit	83
4.5.4	Commandes essentielles pour l'installation et l'exécution	83
4.5.5	Archetecture de l'application	83
4.5.6	Explication de l'Application Web de Prédiction avec Streamlit	83
4.6	Conclusion	86
5	Développement et Intégration du système centralisé intelligent	87
5.1	Introduction	87
5.2	Présentation des applications web	88
5.2.1	Définition	88
5.2.2	Fonctionnement d'une application web	88
5.3	Langages et environnements de développement	88
5.3.1	HTML	88
5.3.2	CSS	88
5.3.3	JavaScript	88
5.3.4	SQL	89
5.3.5	Python	89
5.3.6	PHP	89
5.4	Outils de développement	89
5.4.1	XAMPP	89
5.4.2	MYSQL	89
5.4.3	Serveur Apache	89
5.4.4	Visual Studio Code	90
5.4.5	Git	90
5.4.6	GitHub	90
5.5	Frameworks utilisés	90
5.5.1	Bootstrap	90
5.5.2	Symfony	90
5.6	Sécurité dans le système de gestion des réclamations à BMT-Spa	90
5.6.1	Importance de la Sécurité	91

5.7	Architecture de notre application	91
5.7.1	Phase 1 : Interface Client	91
5.7.2	Phase 2 : Serveur Web	91
5.7.3	Phase 3 : Interface Administrateur	91
5.8	Présentation de l'application	92
5.8.1	Interfaces Communes	92
5.8.2	Interfaces Dédiées aux Clients	92
5.8.3	Interfaces Dédiées à l'Administration	94
5.9	Conclusion	96
6	Conclusion générale et travaux futures	97
	Bibliographie	102
	Annexes	107
	Résumé	109

Table des figures

1.1	L'architecture du processus de la BI	6
1.2	Représentation des différents types d'intelligence artificielle	12
1.3	Structure d'un arbre de décision	15
1.4	GBRT avec Early Stopping	16
1.5	Fonctionnement de l'algorithme AdaBoost	17
1.6	L'Evolution de XGBoost	18
2.1	Historique de BMT	22
2.2	Situation Géographique de BMT	23
2.3	Organigramme de BMT	24
2.4	Champ d'étude	26
2.5	procédure actuelle de BMT	27
2.6	Processus de Collecte de Données	30
3.1	Diagramme de contexte	38
3.2	Diagramme cas d'utilisation « Mes Réclamations »	41
3.3	Diagramme cas d'utilisation « Réclamations »	42
3.4	Diagramme cas d'utilisation totale	43
3.5	Diagramme d'activité pour le cas d'utilisation Authentification	44
3.6	Diagramme d'activité ajouter réclamation	45
3.7	Diagramme d'activité pour le cas Gérer Réclamations	46
3.8	Diagramme d'activité prédire réclamation	46
3.9	Modèle de Domaine	47
3.10	Diagramme de séquence « authentification »	48
3.11	Diagramme de séquence (Mes Réclamations)	49
3.12	Diagramme de séquence « Gérer les Réclamations »	50
3.13	Diagramme de classe	51
4.1	Installation de Power BI Desktop	57
4.2	Écran de connexion de Power BI Desktop	58
4.3	Écran d'accueil de Power BI Desktop	58
4.4	Sélection de source de données dans Power BI Desktop	59
4.5	Importation des Données de Réclamations dans Power BI	59
4.6	Interface utilisateur de Power BI	60
4.7	Affichage des données dans le volet Données	61
4.8	Transformation des Données dans Power BI	61
4.9	Renommer l'entête de la table	62

4.10 Filtrage des lignes vides	62
4.11 Suppression des lignes vides	62
4.12 Affichage des réclamations en double	63
4.13 suppression des doublons de la colonne « Réclamation »	63
4.14 Renommer les noms des colonnes	64
4.15 Modification du Type de Données dans Power Query	64
4.16 Fermeture de l'Éditeur de requêtes	65
4.17 Taux de résolution en un mois	65
4.18 Taux de résolution en un trimestre	65
4.19 Temps Moyen de Traitement Total	66
4.20 Filtrage des années	66
4.21 Nombre total des réclamations	66
4.22 Classement des 10 clients les plus actifs	67
4.23 Nombre de réclamations par mois	68
4.24 Nombre de réclamations par classe de réclamation	68
4.25 Vue générale des réclamations	70
4.26 Nombre de réclamations par type de réclamation	71
4.27 Nombre de réclamations par type de prestation	71
4.28 Nombre de réclamations par commission	72
4.29 Nombre de réclamations par décision	73
4.30 Analyse Détaillée des Réclamations	73
4.31 Évaluation de la performance du service	74
4.32 Nombre de délais de traitement par mois	75
4.33 Suivi de la Performance de Traitement des Réclamations	75
4.34 Matrice de Corrélation	78
4.35 Matrice de Confusion	79
4.36 Courbe de ROC	80
4.37 Analyse comparative des performances des divers classificateurs	81
4.38 Architecture de L'Application de prédiction	83
4.39 Interface de Prédiction des Réclamations Passant en Commission	84
4.40 Interface de Prédiction des Réclamations Ne Passant Pas par Commission	85
4.41 Documentation des Types de Données pour l'Application de Prédiction	85
5.1 Architecture de notre système centralisé intelligent	92
5.2 Interface d'authentification	92
5.3 Interface permet aux clients de consulter la liste de leurs réclamations	93
5.4 Formulaire de Saisie de Réclamation	93
5.5 Interface de la liste des réclamations des clients	94
5.6 Interface permet aux administrateurs de créer de nouveaux comptes	95
5.7 Dashboard statistique	95
5.8 Interface permet aux administrateurs de prédire	96

Liste des tableaux

1.1	Matrice de confusion pour la classification binaire	18
2.1	Description du Dataset	31
2.2	Comparaison des fonctionnalités entre Excel et Power BI	34
3.1	Description du cas d'utilisation « Mes Réclamations »	39
3.2	Description du processus de « Gestion des Réclamations »	40
3.3	Description des attributs des clients finaux	51
3.4	Description des attributs de la table des factures	52
3.5	Description des attributs de la table des conteneurs	52
3.6	Description des attributs de la table des prestations	52
3.7	Description des attributs de la table des réclamations	52
3.9	Description des attributs de la table des réponses aux réclamations	53
3.11	Description des attributs de la table des utilisateurs	53
4.1	Principaux KPI pour l'analyse des réclamations	56
4.2	Total de réclamations par année après filtrage	67
4.3	Réclamations par client pour les années 2021 à 2023	67
4.4	Mois avec le plus de réclamations par année après filtrage	68
4.5	Pourcentage des classes ayant le plus de réclamations au fil des années	69
4.6	Pourcentage des types de réclamation les plus fréquentes au fil des années	71
4.7	Pourcentage des types de prestation les plus réclamés au fil des années	72
4.8	Nombre de réclamations par commission au fil des années	72
4.9	Décisions des réclamations au fil des années	73
4.10	Évaluation de la performance du service	74
4.11	Réclamations mensuelles les plus élevées	75
4.12	Résultats des classificateurs	80

Liste d'abréviations et notations

- BI** : Business Intelligence
- BMT** : Béjaïa Mediterranean Terminal
- Cart** : Classification And Regression Tree
- CSS** : Cascading Style Sheet
- CSRF** : Cross Side Request Forgery
- DT** : Decision Tree (Arbre de Décision)
- ETL** : Extract, Transform, Load
- GBRT** : Gradient Boosted Regression Trees
- GPS** : Global Positioning System
- HTML** : HyperText Mark-up Language
- IA** : Intelligence Artificielle
- KNN** : K-Nearest Neighbors (K plus proches voisins)
- KPI** : Key Performance Indicator (Indicateur Clé de Performance)
- LOLO** : Lift-On (chargement) Lift-Off (déchargement)
- MAX(DT)** : Maximum de Délais de Traitement
- MIN(DT)** : Minimum de Délai de Traitement
- ML** : Machine Learning
- OLAP** : On-Line Analytical Processing
- SPA** : Société par Actions
- SQL** : Structured Query Language
- SVM** : Support Vector Machine
- TMTT** : Temps Moyen de Traitement Total
- TRM** : Taux de Résolution Mensuel
- TRT** : Taux de Résolution Trimestriel
- XGBOOST** : eXtreme Gradient Boosting

Introduction générale

Dans le paysage actuel des affaires, l'utilisation stratégique des données est essentielle pour maintenir un avantage concurrentiel. La science des données, en extrayant des insights à partir de vastes ensembles de données, joue un rôle central dans la prise de décision et l'innovation des entreprises. Un data scientist utilise des techniques statistiques, d'apprentissage automatique et de visualisation des données pour transformer des données brutes en informations exploitables, aidant ainsi les entreprises à optimiser leurs opérations et à prendre des décisions éclairées.

Les méthodes de business intelligence (BI) et de machine learning sont cruciales dans ce processus. La BI permet de visualiser et d'interpréter les données, facilitant l'analyse des performances et l'identification des opportunités d'amélioration. Le machine learning automatise l'extraction de patterns complexes, permettant de faire des prédictions précises et d'optimiser les processus en temps réel.

Face à la dynamique incessante de Bejaia Mediterranean Terminal (BMT), la gestion optimale des réclamations clients est cruciale pour garantir la satisfaction des clients et améliorer l'efficacité opérationnelle. Avec l'augmentation de la complexité des demandes et l'importance des retours clients, BMT doit trouver des solutions innovantes pour relever ces défis.

Ce mémoire se concentre sur la conception et l'implémentation d'un système centralisé intelligent pour l'amélioration du traitement des réclamations chez BMT, en exploitant la science des données. L'objectif est de développer une plateforme robuste et intelligente qui centralise, analyse et gère les réclamations de manière efficace, tout en améliorant la réactivité et la qualité du service client. Cette plateforme intégrera des outils de BI pour une visualisation claire des données de réclamation et des algorithmes de machine learning pour automatiser le traitement et la classification des réclamations, garantissant une réponse rapide et appropriée à chaque situation.

Problématique

Après une analyse approfondie du domaine d'étude, plusieurs défis critiques ont été identifiés dans la gestion des réclamations au sein de BMT Spa. Ces défis ont un impact significatif non seulement sur la performance opérationnelle de l'entreprise, mais également sur la satisfaction des clients. Parmi les principales problématiques rencontrées, on peut citer :

- **Gestion non centralisée et dispersion des réclamations** : Les réclamations issues de divers canaux ne sont pas centralisées efficacement, entraînant des doublons et des incohérences.
- **Retards dans le traitement et inefficacités opérationnelles** : Les processus manuels actuels entraînent des délais importants, impactant la trésorerie et la satisfaction client.

- **Impact financier et risque de pertes** : Les inefficiences dans le traitement des réclamations peuvent conduire à des pertes financières, notamment par le biais de factures impayées.
- **Retards dans le reporting et décisions non informées** : L'absence d'un système efficace de reporting retarde la production de rapports essentiels, compromettant la prise de décisions opportunes.
- **Manque d'outils d'aide à la décision** : L'absence d'outils pour prioriser et gérer efficacement les réclamations limite la réactivité de l'entreprise.
- **Données mal exploitées** : Des données précieuses ne sont pas analysées de manière adéquate, ce qui empêche l'extraction d'insights utiles et la prise de décisions basées sur des informations concrètes.

Objectifs de l'étude

Cette étude vise à analyser et à comprendre les défis critiques rencontrés par BMT Spa dans sa gestion des réclamations clients. Les objectifs spécifiques sont :

- Optimiser les processus de gestion des réclamations pour accroître l'efficacité et traiter un plus grand nombre de réclamations dans des délais réduits.
- Développer un modèle prédictif permettant d'identifier les réclamations nécessitant une intervention spécifique, afin de prioriser et de rationaliser les décisions.
- Mettre en place un système centralisé et sécurisé de gestion des réclamations, garantissant la protection des données et facilitant une gestion plus cohérente et efficiente.
- Concevoir et implémenter une solution automatisée pour le traitement des réclamations, augmentant ainsi la rapidité et l'efficacité du processus.
- Élaborer des tableaux de bord pour visualiser les données et faciliter une analyse approfondie du traitement des réclamations clients, soutenant ainsi la prise de décisions stratégiques.

Méthodologie

Pour atteindre ces objectifs, une méthodologie rigoureuse a été définie, incluant les étapes suivantes :

- Collecte et structuration des données existantes pour former un ensemble de données complet et cohérent.
- Développement d'une application web sécurisée pour la gestion centralisée des réclamations, garantissant la sécurité des données.
- Utilisation d'outils de Business Intelligence pour la visualisation des données de réclamation et l'analyse approfondie du processus de traitement.
- Implémentation d'un modèle prédictif basé sur le machine learning pour l'identification des réclamations nécessitant une intervention spécifique.

Plan du mémoire

Pour mener à terme notre travail, nous le répartissons de la manière suivante :

- Le premier chapitre explore les généralités sur la Business Intelligence et le Machine Learning. Nous commencerons par définir la BI, son fonctionnement, ses enjeux, et son importance dans les décisions stratégiques des entreprises. Enfin, nous introduirons les notions de base du Machine Learning, en abordant ses types et principaux algorithmes.
- Le deuxième chapitre décrit l'entreprise BMT-Spa, ses défis dans le traitement des réclamations, la méthodologie de collecte des données, le dataset utilisé, et le choix de Power BI comme outil d'analyse.
- Le troisième chapitre détaille les besoins du projet à travers l'utilisation d'UML, en identifiant les besoins fonctionnels et non fonctionnels. Il propose des indicateurs de performance intégrés à la solution BI et présente des diagrammes comme le diagramme de contexte statique et des cas d'utilisation spécifiques pour la gestion des réclamations. La conception est approfondie avec des diagrammes d'activités pour les processus clés et des modèles comme les diagrammes de séquence, de classes, et un modèle relationnel pour la base de données.
- Le quatrième chapitre se concentre sur l'application pratique des outils de Business Intelligence (BI) pour analyser et visualiser les données de réclamations chez BMT-Spa. Nous explorerons comment Power BI est utilisé pour prétraiter les données, créer des tableaux de bord interactifs et fournir des insights stratégiques. De plus, nous aborderons la prédiction des réclamations à l'aide de techniques avancées de Machine Learning, visant à déterminer si une réclamation nécessite une commission.
- Le dernier chapitre détaille le développement d'un système web centralisé pour la gestion des réclamations chez BMT, utilisant des technologies telles que Symfony, Bootstrap, et MySQL. Il met l'accent sur la sécurité avec une authentification Symfony personnalisée et l'intégration de mesures contre les attaques CSRF. L'architecture inclut des interfaces utilisateur intuitives et un tableau de bord statistique pour optimiser les processus internes et améliorer la satisfaction client.

1

Généralités sur la Business Intelligence Et Machine Learning

1.1 Introduction

La transformation numérique a révolutionné la gestion des données et la prise de décisions stratégiques. Ce premier chapitre introduit les concepts de la Business Intelligence (BI) et du Machine Learning (ML), essentiels pour le traitement des réclamations clients chez Bejaia Mediterranean Terminal (BMT).

La Business Intelligence utilise des technologies avancées pour fournir des informations cruciales, tandis que le Machine Learning crée des systèmes autonomes capables d'apprentissage et d'amélioration continue. Ensemble, ces outils permettent d'optimiser les processus et d'anticiper les évolutions futures.

Ce chapitre établit les bases pour comprendre l'application de la BI et du ML dans la gestion des réclamations clients chez BMT, offrant des solutions innovantes pour améliorer la satisfaction et les performances organisationnelles.

Sommaire

1.1	Introduction	4
1.2	Business Intelligence	5
1.3	Machine Learning	12
1.4	Conclusion	19

1.2 Business Intelligence

1.2.1 Historique

Début des Années 70 à Années 90 : Les années 70 marquent le début de la prise de conscience sur l'importance stratégique de l'information. Avec l'avènement des ordinateurs, les entreprises s'informatisent, permettant aux dirigeants d'accéder aux résultats de leurs activités de manière plus automatisée. Les années 80 voient l'émergence des Data Centers pour gérer la croissance massive des données. Dans les années 90, le Reporting, bien que soulageant partiellement le service informatique, introduit des problèmes de surcharge des systèmes et d'insatisfaction des décideurs.

Évolution vers la BI Moderne (Années 90 à années 2000) : Le concept de BI prend forme dans les années 90, centré sur la livraison d'informations factuelles pour améliorer la prise de décision. Géré par les départements informatiques via des entrepôts de données, il se concentre sur la technologie de l'information dans les grandes entreprises.

Changement de Perspective (Années 2000 et Au-delà) : À partir des années 2000, la compréhension de la BI évolue. Elle cesse d'être simplement une technologie de l'information pour devenir la capacité d'une organisation à définir des objectifs et à suivre leur réalisation de manière intelligente. La BI adopte une vision plus holistique de son rôle et de son potentiel dans le contexte moderne des entreprises.[13]

1.2.2 Définition de la BI :

La BI permet aux entreprises de convertir de grandes quantités de données en informations exploitables. En utilisant des modèles dimensionnels bien conçus, les entreprises peuvent effectuer des analyses plus rapides et plus précises, améliorant ainsi la qualité des décisions stratégiques et opérationnelles. La capacité à comprendre les données de manière plus approfondie et à agir en conséquence est un atout essentiel dans le paysage commercial moderne [34].

Selon Alain Fernandez, expert en management et auteur, la Business Intelligence (BI), également appelée «Informatique Décisionnelle», représente la clé pour construire une entreprise proactive de manière authentique. La BI englobe un ensemble d'outils technologiques soigneusement agencés et déployés de manière cohérente avec la stratégie d'entreprise préalablement définie. Sa finalité principale est de fournir des informations pertinentes à chaque responsable, lui permettant de prendre des décisions optimales en fonction de son contexte opérationnel, de ses responsabilités, et de ses objectifs tant tactiques que stratégiques[22].

1.2.3 Fonctionnement du processus de la BI

Le processus de la Business Intelligence comporte quatre fonctions principales : collecte, stockage, distribution et exploitation des données. La chaîne décisionnelle se présente selon l'architecture suivante [27] :

Phase de collecte :

Il s'agit d'extraire les données provenant des différentes sources de l'entreprise, de les transformer, puis de les charger dans la base de données en utilisant l'outil ETL (Extract, Transform, Load), qui permet d'adapter les données à un usage décisionnel.

Phase de stockage :

Appelée également modélisation des données, cette phase consiste à centraliser les données structurées et traitées afin qu'elles soient disponibles pour l'usage décisionnel, et faciles à analyser. Pour ce faire, on entrepose les données dans un entrepôt de données (Data Warehouse) ou un magasin de données (Data Mart).

Phase de distribution :

Cette phase consiste à utiliser différents outils afin de restituer l'information sous une forme exploitable pour la prise de décision. En utilisant des outils de reporting, des portails d'accès à des tableaux de bord, ou des outils de statistiques. Des portails décisionnels comme l'EIP (Enterprise Information Portal) permettent aussi de distribuer l'information à l'ensemble des partenaires.

Phase de restitution ou d'exploitation :

Une fois que les données ont été nettoyées, consolidées, rendues accessibles et stockées, elles sont prêtes à être analysées par les utilisateurs finaux ou les experts en analyse. Selon les besoins, différents types d'outils d'extraction et d'exploitation seront utilisés tels que :

- OLAP pour les analyses multidimensionnelles, notamment analyser les données.
- Le data mining pour rechercher des corrélations.
- Les tableaux de bord présentant les indicateurs clés de l'activité.
- Le Reporting pour communiquer la performance.

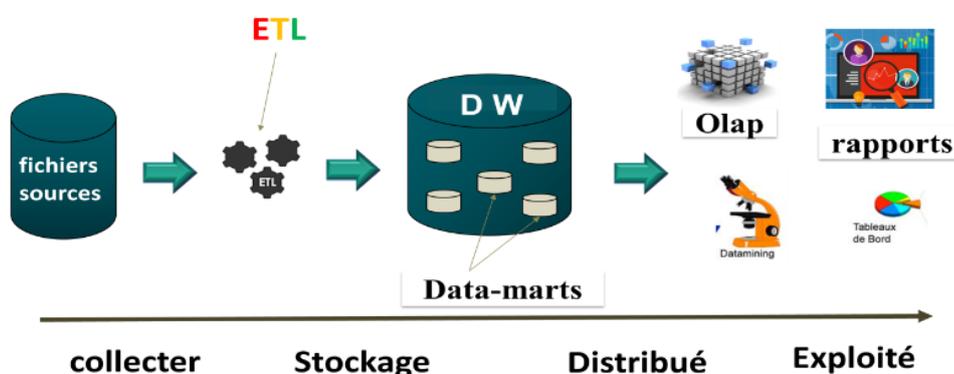


FIGURE 1.1 – L'architecture du processus de la BI

1.2.4 Notions de base en Informatique Décisionnelle(BI)

source de données

Une source de données est un lieu ou un système fournissant les données utilisées pour l'analyse et la prise de décision, comprenant les bases de données relationnelles (MySQL, Oracle, SQL Server), les entrepôts de données, les fichiers CSV/Excel, les CRM et les sources externes.

ETL

ETL (Extract, Transform, Load) est un processus visant à collecter des données de diverses sources, les transformer pour assurer leur cohérence et les charger dans une destination centrale, facilitant ainsi leur utilisation pour la prise de décision, l'analyse et la création de rapports[29] :

Extraction : Consiste à collecter les données à partir de diverses sources. Ces sources peuvent inclure des bases de données relationnelles, des fichiers (CSV, Excel), des applications web et cloud.

Transformation : implique le nettoyage, la modification et l'organisation des données extraites afin qu'elles soient dans un format approprié pour l'analyse ou d'autres besoins de l'entreprise.

Chargement : Consiste à transférer les données transformées vers la destination finale. L'objectif est de stocker les données de manière à ce qu'elles soient facilement accessibles pour des analyses futures, des rapports et d'autres applications métiers.

Data warehouse

Selon Bill Inmon dans son ouvrage «Building the Data Warehouse», l'entrepôt de données, ou data warehouse, est défini comme suite [29] : « Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d'un processus d'aide à la décision».

Data mart

Un Datamart est un dépôt de données spécialisé, un sous-ensemble du Datawarehouse qui stocke uniquement les données d'un domaine spécifique de l'entreprise, contrairement au Datawarehouse qui englobe toutes les données de l'entreprise pour tous les domaines. L'avantage de sa mise en œuvre réside dans sa simplicité d'utilisation, du fait de sa portée et de son étendue réduites au sein de l'organisation[53].

Reporting

Le reporting facilite la création, la distribution et la gestion des rapports en utilisant des requêtes préétablies pour extraire des informations dans un format spécifique, et ce de manière régulière[18].

Data minig

Est un ensemble de méthodes visant à extraire des modèles à partir de bases de données historiques, de Datawarehouses, etc. Ces modèles permettent de décrire le comportement actuel ou de prédire le comportement futur d'un processus [16].

OLAP

Un système d'analyse en ligne des données (OLAP : On-line Analytical Processing) est décrit comme « une catégorie de logiciels conçus pour l'exploration et l'analyse rapides des données selon une approche multidimensionnelle avec plusieurs niveaux d'agrégation » [50]. Il s'agit également d'un système décisionnel où les entrepôts de données sont organisés de manière multidimensionnelle pour fournir un support efficace aux analyses OLAP [57].

Tableaux de bord

Est un ensemble d'indicateurs peu nombreux conçus pour permettre aux gestionnaires de prendre connaissance de l'état et de l'évolution des systèmes qu'ils pilotent et identifier les tendances qui les influenceront sur un horizon cohérent avec la nature et leurs fonctions[21].

1.2.5 Enjeux de la BI pour les entreprises

Les évolutions technologiques actuelles imposent à l'entreprise de relever les défis de réactivité, d'agilité et de flexibilité informatique. Les approches traditionnelles de stockage, de restitution et de modification séparées des données provenant des différents services opérationnels (logistique, gestion de la qualité, marketing, finance via l'outil comptable) ne sont plus adaptées. Il est désormais essentiel que tous les services de l'entreprise adoptent une structuration et une codification de données uniformes, avec des périmètres similaires, afin de permettre une vision globale et synthétique.

Cette transition relève avant tout d'une démarche stratégique reposant sur l'exploitation d'un système d'information décisionnel alimenté par l'extraction de données variées concernant l'entreprise ou son environnement. En outre, la BI offre l'exploitation des données en fournissant une aide à la prise de décision aux professionnels, aboutissant à la création de rapports et de tableaux de bord analytiques et prospectifs pour suivre les activités de l'entreprise[12].

1.2.6 Importance de la business intelligence pour l'entreprise

la Business Intelligence joue un rôle crucial dans la réussite des entreprises modernes en leur permettant de [34] :

- **Prendre des décisions basées sur des données** : La BI fournit aux décideurs des informations précises et à jour, issues de l'analyse des données historiques et actuelles de l'entreprise. Cela permet d'éclairer les choix stratégiques et opérationnels, réduisant ainsi la dépendance aux intuitions ou aux estimations.

- **Améliorer l'efficacité opérationnelle** : En centralisant les données provenant de diverses sources dans un entrepôt de données, la BI permet une analyse plus cohérente et complète des processus de l'entreprise. Cela aide à identifier les inefficacités, à optimiser les opérations et à réduire les coûts.
- **Identifier les tendances et les opportunités** : Grâce à des outils de visualisation et d'analyse avancés, la BI aide les entreprises à détecter des tendances, des modèles et des opportunités de marché qui pourraient autrement passer inaperçus. Cela permet aux entreprises d'anticiper les changements et de réagir proactivement.
- **Améliorer la satisfaction client** : En analysant les données clients, les entreprises peuvent mieux comprendre les besoins et les préférences de leur clientèle. Cela permet de personnaliser les offres, d'améliorer les services et, in fine, de renforcer la satisfaction et la fidélité des clients.
- **Renforcer la compétitivité** : Dans un marché de plus en plus concurrentiel, la capacité à exploiter les données de manière efficace peut donner un avantage compétitif significatif. Les entreprises utilisant la BI peuvent réagir plus rapidement aux conditions changeantes du marché, ajuster leurs stratégies et surpasser leurs concurrents.
- **Faciliter la conformité réglementaire** : De nombreuses industries sont soumises à des réglementations strictes concernant la gestion et la protection des données. La BI aide à centraliser et à auditer les données, facilitant ainsi la conformité avec les exigences légales et réglementaires.

1.2.7 Étapes clés pour réussir un projet de BI

Étapes Clés pour Réussir un Projet de Business Intelligence (BI)[20] :

- Définition des Objectifs du Projet
- Collecte des Données
- Préparation des Données
- Conception et Développement de l'Entrepôt de Données
- Développement des Tableaux de Bord et des Rapports
- Analyse des Données
- Mise en Œuvre des Modèles de Machine Learning (si applicable)
- Test et Validation
- Déploiement de la Solution
- Formation des Utilisateurs

1.2.8 Domaines d'application de la BI

- Finance et assurance : analyse des risques pour les prêts.
- Service client : suivi des interactions et évaluation de la satisfaction.
- Marketing : segmentation des clients et analyses comportementales.
- Logistique : optimisation de la gestion de stock et suivi des livraisons.
- Gestion de projet : suivi, allocation des ressources et évaluation des performances.
- Qualité : contrôle qualité, suivi des défauts et analyse pour l'amélioration des processus.

1.2.9 Utilité des Données Collectées dans un Projet de BI

Les données sont le fondement de toute initiative de Business Intelligence. Elles permettent de transformer des informations brutes en insights exploitables, facilitant ainsi la prise de décision stratégique. L'intégration et l'analyse des données collectées offrent aux entreprises une compréhension approfondie de leurs opérations, de leurs clients et de leur marché, ce qui est essentiel pour rester compétitif et innovant[52] :

1. **Prise de Décision Informée** Les données collectées fournissent une base solide pour prendre des décisions éclairées. Elles permettent aux décideurs de baser leurs choix sur des faits et des analyses plutôt que sur des intuitions ou des estimations. Cela réduit le risque d'erreurs et augmente la probabilité de succès des initiatives stratégiques.
2. **Amélioration de l'Efficacité Opérationnelle**
En analysant les données opérationnelles, les entreprises peuvent identifier des inefficacités et des goulots d'étranglement dans leurs processus. Cela permet de mettre en place des améliorations pour optimiser les opérations, réduire les coûts et augmenter la productivité.
3. **Identification des Tendances et des Modèles**
Les données collectées permettent d'identifier des tendances et des modèles qui peuvent révéler des opportunités ou des menaces pour l'entreprise. Par exemple, une analyse des ventes peut montrer des produits en croissance ou en déclin, permettant ainsi de réagir de manière proactive.
4. **Segmentation et Personnalisation**
Les données clients permettent de segmenter le marché et de personnaliser les offres. En comprenant mieux les préférences et les comportements des clients, les entreprises peuvent créer des campagnes marketing ciblées, améliorer les produits et services, et augmenter la satisfaction client.
5. **Suivi des Performances**
Les données collectées permettent de suivre les performances à travers des indicateurs clés de performance (KPIs). Les tableaux de bord et les rapports fournissent

une vue d'ensemble de la santé de l'entreprise et permettent de surveiller les progrès vers les objectifs stratégiques.

6. Prévisions et Planification

Les analyses prédictives basées sur les données historiques permettent de prévoir les tendances futures. Cela aide les entreprises à planifier efficacement les ressources, à anticiper les besoins du marché et à préparer des stratégies de croissance.

1.2.10 Raisons de Choisir un Logiciel ou un Outil de Business Intelligence (BI)

Les outils de Business Intelligence jouent un rôle crucial dans la transformation des données brutes en informations exploitables. Ils permettent aux entreprises de naviguer dans un paysage commercial complexe avec une clarté et une précision accrues, améliorant ainsi leur capacité à prendre des décisions stratégiques et opérationnelles éclairées[52].

1.2.11 Avantages de la BI

- Rapidité des réponses en temps réel.
- Décisions éclairées par la transformation des données.
- Efficacité opérationnelle et résolution rapide des problèmes.
- Retour sur investissement accéléré et augmentation de la productivité.
- Reporting instantané pour résoudre des problèmes complexes.
- Précision stratégique dans l'identification des tendances clés.
- Satisfaction client grâce à l'amélioration des produits et services[1].

1.2.12 Inconvénients de la BI

- L'investissement dans des outils de veille économique peut représenter un coût élevé.
- La mise en œuvre de la Business Intelligence requiert souvent des équipes spécialisées.
- La création d'un système d'entreposage de données peut prendre du temps.
- La centralisation des données peut augmenter le risque de violations de sécurité et de confidentialité.
- Les évolutions rapides des besoins commerciaux peuvent rendre rapidement obsolètes les solutions de BI[1].

1.3 Machine Learning

1.3.1 Définition des notions

Intelligence artificielle

L'intelligence artificielle est définie comme « L'intelligence artificielle (IA) fait référence à des systèmes qui présentent un comportement intelligent en analysant leur environnement et en prenant des mesures avec un certain degré d'autonomie pour atteindre des objectifs spécifiques » [11].

Il s'agit d'ordinateurs ou de machines équipés de programmes capables de performances similaires à l'intelligence humaine, voire amplifiées par la technologie. Ces machines peuvent raisonner, traiter de grandes quantités de données, discerner des modèles indétectables par l'œil humain, comprendre et analyser ces modèles, interagir avec les humains, apprendre progressivement et améliorer continuellement leurs performances. Les utilisations actuelles de l'IA se répartissent en trois catégories principales : l'identification, la prédiction et la génération de données. Ces applications de l'IA incluent notamment les concepts de machine learning et de deep learning.

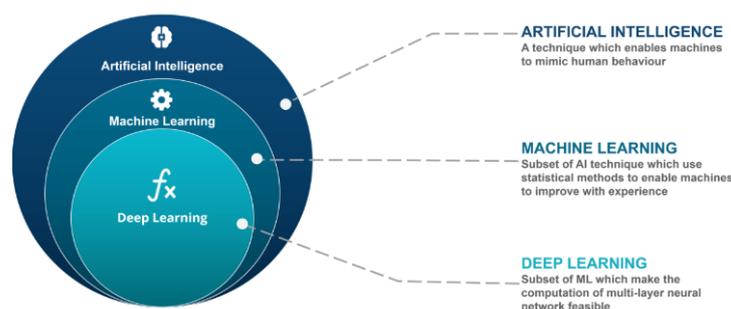


FIGURE 1.2 – Représentation des différents types d'intelligence artificielle

Machine Learning

L'apprentissage automatique, ou "machine learning" en anglais, est un domaine de l'intelligence artificielle (IA) qui se concentre sur le développement d'algorithmes et de techniques permettant aux ordinateurs d'apprendre et de s'améliorer automatiquement à partir de données. Plutôt que d'être explicitement programmés pour effectuer une tâche spécifique, les systèmes d'apprentissage automatique utilisent des modèles statistiques pour identifier des motifs dans les données et faire des prédictions ou des décisions basées sur ces informations [38].

Deep Learning

L'apprentissage profond (Deep Learning) est une technique d'apprentissage automatique utilisant des réseaux de neurones artificiels à plusieurs couches pour analyser des données

et améliorer des domaines tels que la reconnaissance vocale et visuelle, la détection d'objets, le développement de médicaments et la génomique [51].

1.3.2 Types d'algorithmes de Machine Learning

L'apprentissage automatique est un domaine de l'intelligence artificielle qui cherche à comprendre la structure des données pour les modéliser de manière accessible à l'ensemble des utilisateurs [5].

L'apprentissage supervisé : L'apprentissage supervisé utilise des données étiquetées pour entraîner un algorithme à faire des prédictions ou à prendre des décisions. Ces données comprennent à la fois des entrées et des sorties attendues pour chaque exemple. Les tâches courantes incluent la classification et la régression. L'objectif est de créer un modèle capable de généraliser à de nouvelles données et de produire des résultats précis.

L'apprentissage non supervisé : L'apprentissage non supervisé utilise des données non étiquetées pour découvrir des structures ou motifs cachés sans supervision externe, en appliquant des techniques comme le clustering, la réduction de dimensionnalité et l'apprentissage des règles d'association.

1.3.3 Algorithmes des machines Learning

a. Algorithmes de régression :

C'est un processus qui vise à trouver un modèle pour distinguer les données en valeurs réelles continues plutôt qu'en classes discrètes. Il existe deux types de ces algorithmes.

La régression Linéaire : La régression linéaire est une approche très simple pour l'apprentissage supervisé, particulièrement utile pour prédire des réponses quantitatives. Bien qu'elle puisse paraître moins intéressante comparée à des méthodes plus modernes d'apprentissage statistique, elle demeure une technique précieuse et largement employée. De plus, elle sert souvent de point de départ pour des approches plus avancées, qui en sont des généralisations ou des extensions.

La régression logistique : La régression logistique est essentielle en apprentissage automatique pour classer les données en fonction des données historiques. Plus les données d'entrée sont pertinentes, plus les prédictions de l'algorithme sont précises. Ce modèle de régression binomiale associe un vecteur de variables aléatoires à une variable binomiale y , constituant un cas particulier de modèle linéaire généralisé.

b. Algorithme de Classification :

La classification automatique est le processus qui consiste à analyser et organiser un ensemble de données en classes de similarité en fonction de leurs caractéristiques. Elle repose principalement sur des représentations classiques des données, dont les limites de traitement sont bien connues, et requiert souvent un temps de calcul considérable.

c. Clustering :

Une méthode d'apprentissage non supervisée est une méthode dans laquelle nous tirons des références d'ensembles de données constitués de données d'entrée sans réponses étiquetées. En général, elle est utilisée comme un processus pour trouver une structure significative, des processus sous-jacents explicatifs, des caractéristiques génératives et des regroupements inhérents à un ensemble d'exemples [54].

1.3.4 Algorithmes de classification

Il existe plusieurs algorithmes et techniques utilisés pour la classification supervisée. Ces derniers sont de différents types mais ayant tous la même intention d'avoir une bonne performance tôt en étant efficace[8].

K plus proches voisins(KNN) :

En intelligence artificielle, spécifiquement dans le domaine de l'apprentissage automatique, la méthode des k plus proches voisins (k-NN ou KNN) est une approche de l'apprentissage supervisé. Elle repose sur une base de données d'apprentissage composée de N paires d'entrées et de sorties. Pour prédire la sortie associée à une nouvelle entrée x, l'algorithme identifie les k échantillons d'apprentissage les plus proches de x en termes de distance définie. Par exemple, dans le cas de la classification, la classe la plus fréquente parmi les k sorties associées aux k entrées les plus proches de x est retenue comme prédiction[26].

Algorithme arbre de décision (DT)

Un arbre de décision est une structure hiérarchique utilisée en apprentissage machine, composée de nœuds et de feuilles reliés par des branches. La racine est en haut, les feuilles en bas. Les nœuds internes, ou nœuds de décision, contiennent des règles ou conditions. Les valeurs des variables sont appelées instances ou attributs. Les nœuds terminaux contiennent la classe à prédire. Après construction, un arbre de décision peut être converti en un ensemble de règles de décision. Si la classe à prédire est qualitative, on parle d'arbre de classification; si elle est quantitative, d'arbre de régression. Ici, nous nous intéressons aux arbres de classification [37].

Exemple : Cet exemple simple permet d'expliquer comment un arbre de décision peut être traduit sous la forme d'un ensemble de règles de décision. Soit l'arbre de décision de la figure suivante qui permet de décider si une institution bancaire peut accorder ou non un prêt à un demandeur de prêt. La décision de la variable cible Prêt se base sur les variables moyenne du salaire (MoySal), âge (Âge) et possession d'autres comptes (Autres comptes). Les attributs des feuilles de l'arbre, c'est-à-dire Prêt, sont des valeurs booléennes qui correspondent à une classification dans l'ensemble {Oui, Non}.

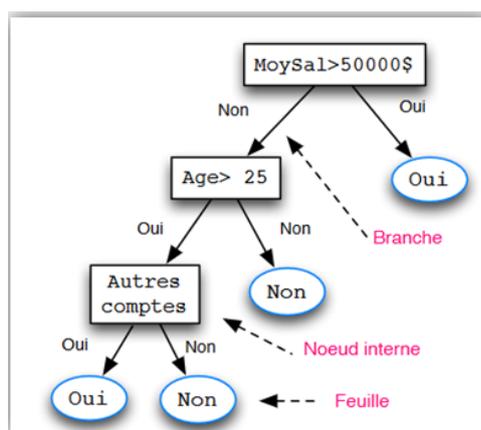


FIGURE 1.3 – Structure d’un arbre de décision

Algorithme Random Forest :

Random Forest est un algorithme d’apprentissage automatique supervisé qui crée et combine plusieurs arbres de décision pour former une ”forêt”. Il est utilisé pour les problèmes de classification et de régression, améliorant la précision des prédictions en réduisant le surapprentissage. Chaque arbre est construit à partir de sous-ensembles aléatoires des données d’entraînement et des caractéristiques, ce qui confère robustesse et stabilité au modèle. L’algorithme est populaire pour sa capacité à gérer de grands ensembles de données, sa rapidité d’entraînement et sa facilité d’interprétation[49].

Bagging :

Bagging, ou Bootstrap Agrégation également appelé ensachage, est un méta-algorithme d’apprentissage automatique conçu pour améliorer la stabilité et la précision des algorithmes d’apprentissage automatique utilisés dans la classification statistique et la régression. Il réduit également la variance et aide à éviter l’overfitting. Bien qu’il soit généralement appliqué aux méthodes des arbres de décision, il peut être utilisé avec n’importe quel type de méthode. Le bagging est un cas particulier de l’approche de moyenne du modèle[49].

Algorithme Naïve Bayes :

Le classificateur Naïve Bayes est une méthode d’apprentissage bayésienne très pratique. Il s’applique aux tâches d’apprentissage où chaque instance x est décrite par une conjonction de valeurs d’attributs et la fonction cible $f(x)$ peut prendre n’importe quelle valeur d’un ensemble fini. Un ensemble d’exemples d’apprentissage de la fonction cible est fourni, et une nouvelle instance est présentée, décrite par un tuple des valeurs d’attributs. L’apprenant est invité à prédire la valeur cible, ou la classification, pour cette nouvelle instance.

L’algorithme est basé sur le théorème de Bayes[59], tel que :

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (1.1)$$

Gradient Boosting :

Le Boosting désigne toute méthode d'ensemble qui combine plusieurs modèles faibles en un modèle fort. La plupart des méthodes de boosting entraînent les modèles de manière séquentielle, chaque modèle essayant de corriger les erreurs de son prédécesseur, ce qui aboutit à des modèles complémentaires. Parmi les algorithmes de Boosting, le Gradient Boosting ajoute des modèles successifs, chacun corrigeant les erreurs résiduelles du précédent.

Le Gradient Boosting est particulièrement efficace pour les tâches de régression, d'où son appellation Gradient Boosted Regression Trees (GBRT). Pour déterminer le nombre optimal d'arbres, on peut utiliser l'early stopping afin d'éviter le sur-apprentissage. La bibliothèque XGBoost permet d'implémenter des méthodes de Gradient Boosting. La figure suivante illustre un modèle GBRT utilisant l'early stopping pour trouver le nombre optimal d'arbres [23].

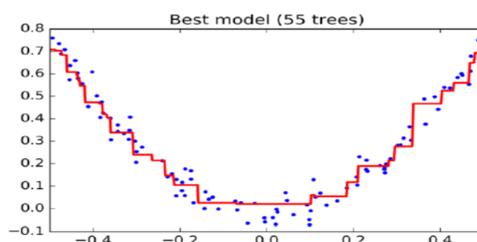


FIGURE 1.4 – GBRT avec Early Stopping

La Régression Logistique

La régression logistique est un algorithme de classification couramment utilisé pour modéliser la relation entre une variable dépendante binaire et une ou plusieurs variables indépendantes. Elle permet d'estimer la probabilité de survenue d'un événement en fonction des valeurs des variables explicatives [7].

Cette technique transforme une combinaison linéaire des variables indépendantes en une probabilité à l'aide de la fonction logistique (sigmoïde). Cette fonction contraint les prédictions à être comprises entre 0 et 1, rendant ainsi le modèle particulièrement adapté aux problèmes de classification binaire.

Le modèle de régression logistique est exprimé comme suit :

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1.2)$$

où :

- $P(Y = 1|X)$ est la probabilité que la variable dépendante Y prenne la valeur 1 (événement se produisant) conditionnellement aux valeurs des variables indépendantes X_1, X_2, \dots, X_n .

- $\beta_0, \beta_1, \dots, \beta_n$ sont les coefficients du modèle estimés par la méthode de maximisation de la vraisemblance.
- X_1, X_2, \dots, X_n sont les variables indépendantes.

Algorithme AdaBoost

L'algorithme AdaBoost est maintenant une méthode bien connue et profondément étudiée pour construire un ensemble de classificateurs avec de très bonnes performances. Il fonctionne en apprenant une série de classificateurs faibles, et en combinant leurs prédictions pondérées pour former un classificateur final fort. Le processus d'apprentissage utilise un poids [36] :

$$H(x) = \text{sign} \left(\sum_{i=1}^M \alpha_i h_i(x) \right) \quad (1.3)$$

où :

- $h_i(x)$ est le i -ème classificateur faible,
- α_i est le poids associé à ce classificateur,
- M est le nombre total de classificateurs faibles.

Les poids α_i sont calculés de manière à refléter la précision des classificateurs faibles, et la somme pondérée des prédictions des classificateurs faibles forme la décision finale.

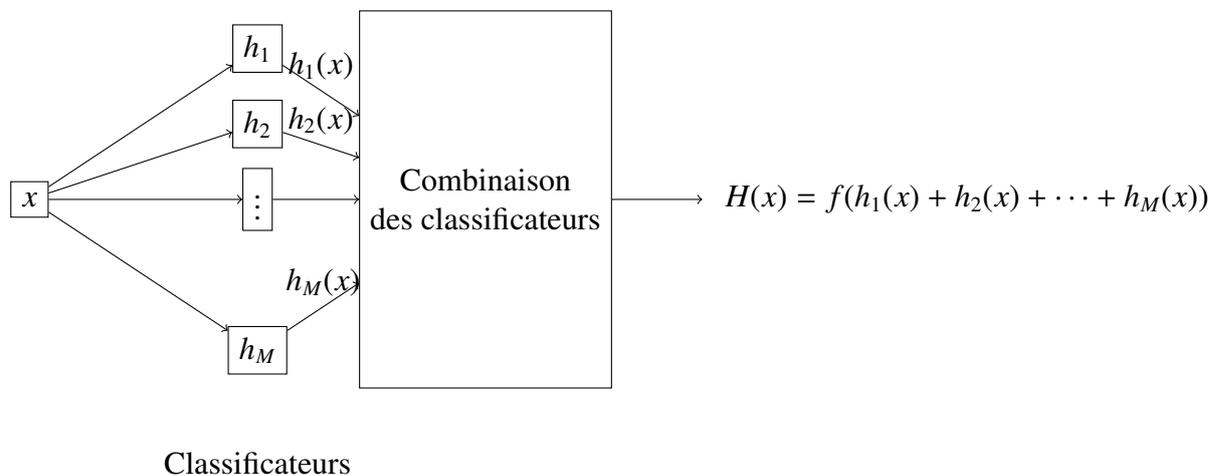


FIGURE 1.5 – Fonctionnement de l'algorithme AdaBoost

L'algorithme XGBoost (eXtreme Gradient Boosting)

XGBoost a été proposé par Chen et Guestrin en 2014 et s'est imposé comme un outil majeur dans le domaine de l'apprentissage automatique ces dernières années. Illustré dans la Figure, XGBoost repose sur l'arbre de boosting par gradient [35].

Il représente une version améliorée et scalable des machines à gradient boosting (GBM), ayant remporté plusieurs compétitions de machine learning grâce à sa facilité d'utilisation, sa capacité à fonctionner efficacement en parallèle, et sa précision prédictive remarquable [4].

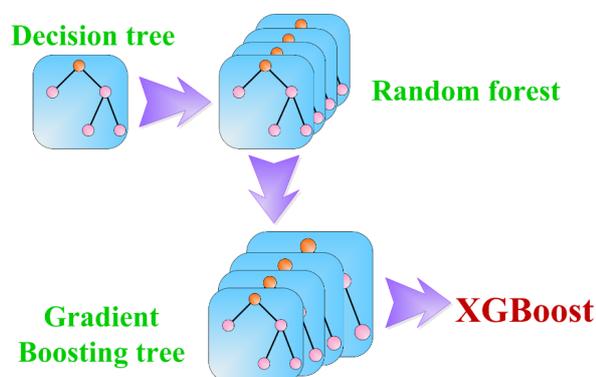


FIGURE 1.6 – L'Evolution de XGBoost

1.3.5 Évaluation de la performance des algorithmes de classification

L'évaluation des algorithmes de classification binaire, à travers des métriques telles que la précision et le rappel, permet aux praticiens de comprendre les performances des modèles. Cette compréhension guide l'ajustement des algorithmes pour des décisions précises en apprentissage automatique, sélectionnant ainsi le meilleur modèle pour des applications réelles [8].

Matrice de confusion

La matrice de confusion est un outil qui mesure les performances d'un modèle de classification à deux classes ou plus. Elle permet de confronter les valeurs référencées avec celles de la prédiction, vérifiant ainsi le bon classement. Pour une classification binaire

		Classe prédiction	
		Classe 1	Classe 2
Classe réelles	Classe 1	VP	FN
	Classe 2	FP	VN

TABLE 1.1 – Matrice de confusion pour la classification binaire

Tel que :

- VP : vrai positif (*true positif*) : les cas où les prédictions sont positives, et la valeur réelle est effectivement positive.
- VN : vrai négatif (*true négatif*) : les cas où les prédictions sont négatives, et la valeur réelle est effectivement négative.
- FN : faux négatif (*false négatif*) : les cas où les prédictions sont négatives, et la valeur réelle est effectivement positive.

- FP : faux positif (*false positif*) : les cas où les prédictions sont positives, et où la valeur réelle est effectivement négative.

Accuracy

La précision (Accuracy) est une mesure de performance de classification qui indique le nombre d'observations correctement classées parmi toutes les observations.

$$\text{Accuracy} = \frac{VP + VN}{VP + FP + VN + FN} \quad (1.4)$$

Précision

La précision, également connue sous le nom de valeur prédictive positive ou taux de prédictions correctes, représente le rapport entre les observations positives prédites et le nombre total d'observations prédites comme positives. Elle évalue la capacité du modèle à minimiser les erreurs lors de la prédiction des positives.

$$\text{Précision} = \frac{VP}{VP + FP} \quad (1.5)$$

Rappel

Également appelé le taux de vrai positif (sensibilité ou sensibilité), le rappel est le rapport entre les observations positives prédites et toutes les observations de la classe réelle.

$$\text{Rappel} = \frac{VP}{VP + FN} \quad (1.6)$$

F1-score

est une mesure qui combine la précision et le rappel en une seule métrique, prenant en compte les faux positifs et les faux négatifs. Il est souvent préféré à la précision seule lorsque la distribution des classes est inégale.

$$\text{F1-score} = \frac{2 \times \text{precision} \times \text{rappel}}{\text{precision} + \text{rappel}} \quad (1.7)$$

1.4 Conclusion

Ce premier chapitre a fourni une vue d'ensemble essentielle des concepts de la Business Intelligence et du Machine Learning, soulignant leur rôle crucial dans l'écosystème numérique actuel. Nous avons retracé l'évolution historique de la BI, défini ses principales composantes, et exploré son fonctionnement et ses avantages, ainsi que ses défis. De même, nous avons introduit les fondements du Machine Learning, en détaillant les différents types d'apprentissage et les algorithmes les plus couramment utilisés.

En comprenant les principes de la BI et du ML, les entreprises peuvent mieux exploiter leurs données pour prendre des décisions éclairées, anticiper les tendances futures et automatiser des

tâches complexes. Ce chapitre pose les bases pour une exploration plus approfondie des applications pratiques et des innovations dans ces domaines, soulignant leur potentiel pour transformer les données en un avantage concurrentiel tangible. Les connaissances acquises ici serviront de tremplin pour les chapitres suivants, où nous examinerons en détail les outils, techniques et stratégies permettant de maximiser l'impact de la BI et du ML au sein des organisations.

2

Contexte de l'Entreprise et Approche Méthodologique de la Collecte des Données

2.1 Introduction

Ce chapitre vise à offrir une présentation détaillée de Bejaia Mediterranean Terminal (BMT), notre entreprise d'accueil, ainsi qu'une analyse approfondie de notre démarche de collecte de données. Nous débuterons par une introduction à BMT, en décrivant son histoire, sa position stratégique dans le paysage portuaire algérien, et son rôle clé dans le commerce maritime.

Nous examinerons ensuite la structure organisationnelle de BMT et les procédures actuelles de traitement des réclamations clients, mettant en exergue les défis liés à l'absence de centralisation des réclamations et aux dysfonctionnements hiérarchiques.

Pour conclure, nous détaillerons la méthodologie employée pour la collecte de données, incluant la définition des besoins, les étapes d'extraction, de transformation et de chargement (ETL) des données, ainsi que les outils de visualisation utilisés. Cette approche méthodologique servira de base pour les analyses et recommandations présentées dans les chapitres ultérieurs.

Sommaire

2.1	Introduction	21
2.2	Présentation de l'entreprise d'Accueil	22
2.3	Procédure actuelle de traitement des réclamations à BMT	26
2.4	Collecte des Données	28
2.5	Description du Dataset Collecté	30
2.6	choix de l'outil Power BI	32
2.7	Conclusion	34

2.2 Présentation de l'entreprise d'Accueil

2.2.1 Historique

BMT SPA est né d'un partenariat entre l'Entreprise Portuaire de Bejaia (EPB) et le groupe PORTEK, spécialisé dans la gestion des terminaux à conteneurs.

Approuvé par le gouvernement en 2004, BMT SPA s'est engagé à fournir des services de qualité supérieure dans la gestion et l'exploitation du terminal à conteneurs.

Avec l'acquisition de PORTEK Systems and Equipment par le groupe MITSUI en 2011, BMT a consolidé sa position en tant que fournisseur de solutions logistiques de premier plan.

Guidé par un engagement envers l'excellence opérationnelle et la satisfaction client, BMT s'efforce de rester un partenaire de confiance pour ses clients commerciaux, offrant des services innovants et adaptés aux besoins du marché mondial.



FIGURE 2.1 – Historique de BMT

2.2.2 Situation Géographique

Situé au cœur de la Méditerranée, dans le nord de l'Afrique, le Port de Bejaia occupe une position géographique centrale et stratégique. Grâce à cette localisation privilégiée, il dessert un vaste hinterland. La ville de Bejaia, son port et son terminal à conteneurs bénéficient ainsi d'un réseau de communication étendu, comprenant des routes nationales, des voies ferrées et la proximité d'un aéroport international. Cette centralité offre à Bejaia une plateforme moderne et accessible, propice aux échanges commerciaux et au développement régional.

Position GPS : l'attitude nord : 36° 45' 24 "

Longitude est : 05° 05' 50 "

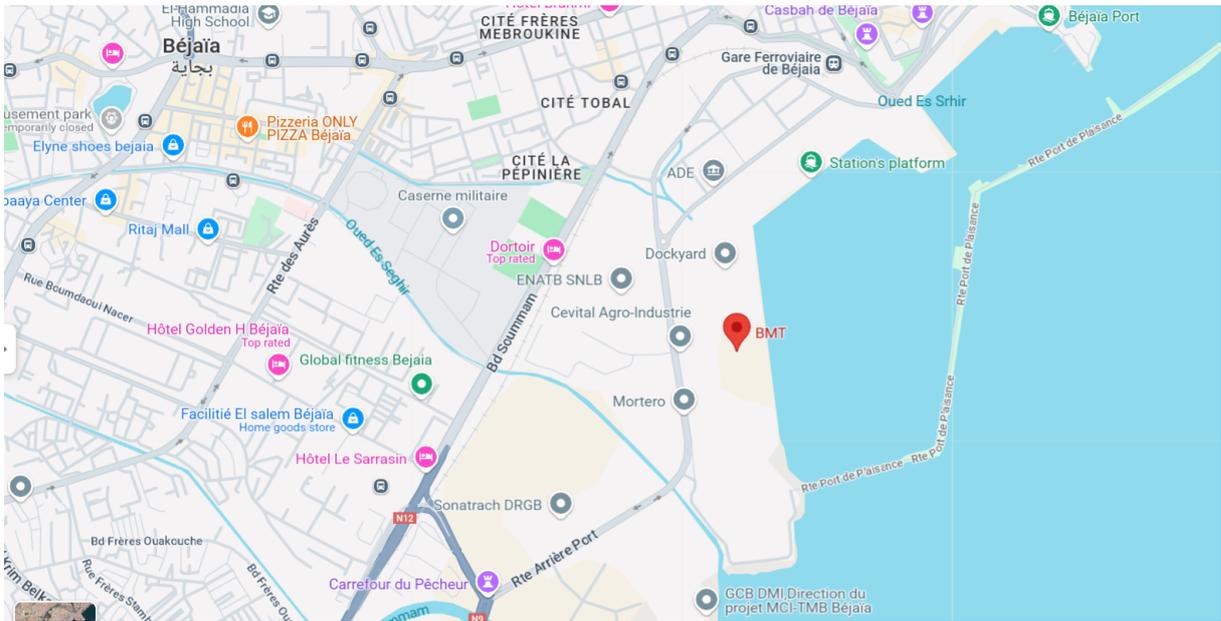


FIGURE 2.2 – Situation Géographique de BMT

2.2.3 Présentation de BMT-Spa

BMT Spa est une société par actions qui se spécialise dans les services liés à l'exploitation et à la gestion de terminaux à conteneurs. Son objectif est d'offrir des prestations de qualité, d'efficacité et de fiabilité grâce à un personnel hautement compétent et spécialement formé pour la gestion de ces terminaux. Dotée d'équipements de maintenance et d'acconage de pointe, BMT Spa garantit des opérations rapides et économiquement compétitives. Ses services sont disponibles 24 heures sur 24, 7 jours sur 7. Le port de Bejaia et BMT Spa sont désormais reconnus comme le principal terminal moderne d'Algérie, offrant une plate-forme portuaire hautement performante.

2.2.4 Organigramme de BMT

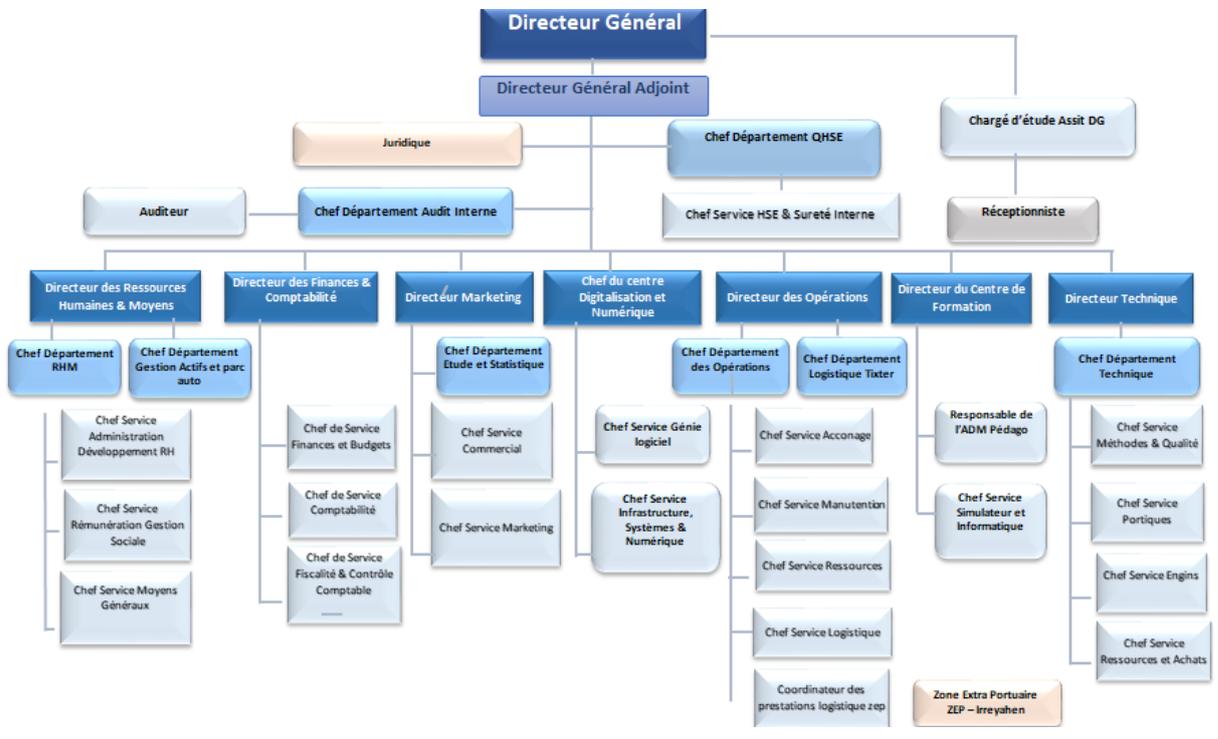


FIGURE 2.3 – Organigramme de BMT

2.2.5 Les valeurs de la BMT (Bureau des Messageries Maritimes)

- **Intégrité** : La BMT s’engage à agir avec intégrité et honnêteté dans toutes ses interactions. Mentir, voler, tromper ou favoriser certains individus vont à l’encontre de cette valeur fondamentale.
- **Innovation** : La BMT encourage la curiosité, la créativité et la recherche de nouvelles opportunités. Remettre en question les conventions établies et explorer de nouveaux domaines font partie de cette valeur, tout en gérant les risques de manière responsable.
- **Ténacité** : Fixer des objectifs ambitieux, respecter les engagements pris, prendre des décisions et les mettre en œuvre avec détermination. Travailler en équipe, surmonter les obstacles et persévérer jusqu’à la réalisation des objectifs sont des aspects clés de cette valeur.
- **Sécurité** : Contribuer à la protection de la santé et à l’amélioration des conditions de travail, ainsi qu’à la sécurité des employés, des clients et des visiteurs. Respecter les règles de protection de l’environnement et assurer la sécurité des biens des clients font également partie de cette valeur.
- **Courtoisie** : Placer le client au centre des préoccupations, respecter les services, l’autorité, la hiérarchie et les règlements établis. Respecter les normes de professionnalisme et de décence sociale, ainsi que ses collègues en tout temps.

En plus de ces valeurs, la BMT a une mission et une vision claires :

- **Mission :** Faire du Terminal à Conteneur de BMT un terminal conforme aux normes internationales, compétitif sur le plan de la productivité et de la rentabilité, assurant ainsi son succès à long terme.
- **Vision :** Développer et gérer le meilleur Terminal à Conteneurs d'Algérie, où l'intégrité, la productivité, l'innovation, la courtoisie et la sécurité sont des priorités. Cette vision vise à offrir aux clients les meilleures prestations de service à des coûts appropriés, tout en maintenant un engagement envers l'excellence opérationnelle

2.2.6 Présentation détaillée du champ d'étude

Pour mieux appréhender notre domaine d'étude, nous examinerons en détail l'organigramme, en mettant l'accent sur le département Marketing, où notre problématique a émergé. Cette approche nous permettra de situer notre étude dans la section Recouvrement du Service Commercial, où nous avons été accueillis et où nous avons identifié le besoin de traiter efficacement les réclamations pour réduire les créances.

Direction Marketing

Assure l'élaboration des plans marketing (analyse de marché, détermination des cibles, plan d'actions, choix des axes publicitaires...) intervient sur différents domaines fonctionnels du marketing tels que : la stratégie, les services, l'opérationnel. Assure la coordination avec l'ensemble des intervenants portuaires en adéquation avec la politique marketing.

- 1. Service Marketing :** Assure la promotion de l'image de marque de l'entreprise et la mise en œuvre du plan marketing et commercial.
- 2. Service Commercial :** Procède à la facturation des prestations fournies et au recouvrement des créances.
 - a. Section facturation :**
 - Le chargé du secteur de la facturation assure la validation, la distribution et la rectification des factures.
 - b. Section recouvrement :**
 - assure la remise des factures aux clients dans les délais.
 - Effectuer des relances pour le paiement.
 - Traitement des chèques impayés.

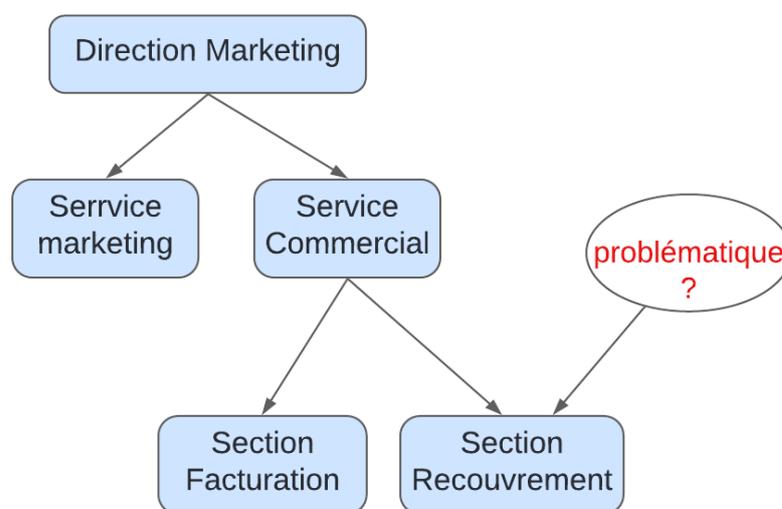


FIGURE 2.4 – Champ d'étude

2.3 Procédure actuelle de traitement des réclamations à BMT

La gestion efficace des réclamations client est essentielle pour Bejaia Mediterranean Terminal (BMT), non seulement pour réduire les créances et améliorer la satisfaction client, mais aussi pour optimiser ses opérations. Les réclamations fournissent une opportunité précieuse de rectifier les erreurs, d'identifier les lacunes dans les processus internes et d'améliorer continuellement les services. Cette étude examine de près la procédure actuelle de gestion des réclamations à BMT, mettant en lumière les défis tels que le manque de centralisation et les processus hiérarchiques. L'objectif est de proposer des solutions pour renforcer l'efficacité et la réactivité de BMT face aux réclamations, augmentant ainsi la satisfaction client et optimisant les performances opérationnelles de l'entreprise.

2.3.1 Procédure

L'objectif de cette procédure est de définir les étapes actuelles pour le traitement des réclamations des clients au sein de la Direction Marketing et de la Direction des Opérations, ainsi que dans les cas de la Commission de Traitement des Doléances :

1. Réception de la réclamation par le Chef de Service Commercial ou le Directeur Marketing.
2. Enregistrement sur le portail intranet de la Direction Marketing.
3. Étude préliminaire par le service Marketing ou réponse immédiate pour les questions administratives.
4. Réclamation complexe examinée par la commission sur demande du Directeur Marketing.
5. Action corrective prise si la réclamation est validée, avec communication transparente au client.

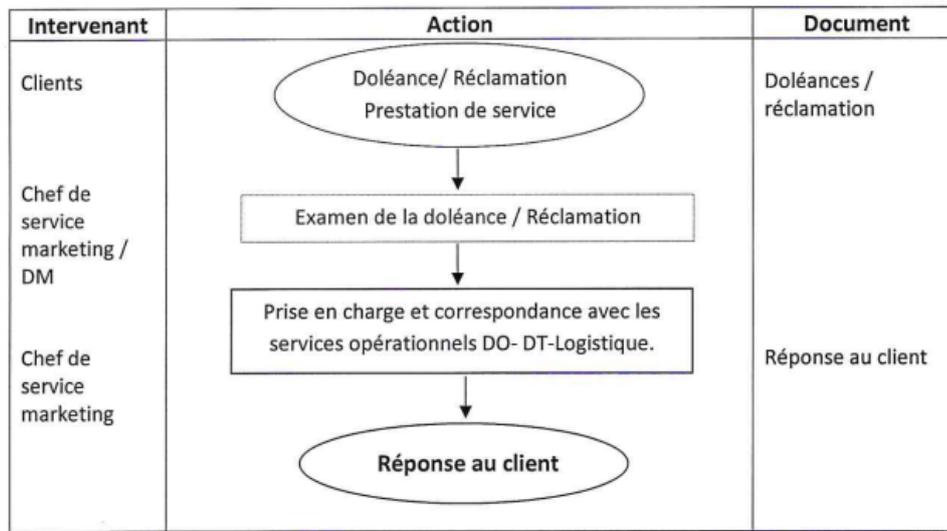


FIGURE 2.5 – procédure actuelle de BMT

2.3.2 Défis potentiels dans la procédure actuelle de traitement des réclamations clients à BMT

— **Manque de Centralisation :**

La procédure mentionne l’enregistrement sur un portail intranet de la Direction Marketing, mais il n’est pas spécifié si ce portail est facilement accessible à tous les départements concernés ou s’il permet une vue centralisée et intégrée des réclamations.

— **Complexité Administrative :**

Les réclamations nécessitant une étude approfondie sont renvoyées à plusieurs services, ce qui peut entraîner des retards dus à la coordination interne et à la communication entre les différents départements.

— **Absence de Délais Définis :**

La procédure ne précise pas de délais standards pour chaque étape du traitement des réclamations. Cela peut conduire à des délais prolongés et à une insatisfaction client accrue.

— **Manque de Transparence :**

Bien que la communication avec le client soit mentionnée, il n’est pas spécifié comment et à quelle fréquence les clients sont tenus informés de l’état d’avancement de leur réclamation. Un manque de transparence peut engendrer des frustrations et une perception négative.

— **Besoins en Formation et Gestion du Personnel :**

Il n’y a pas de plan régulier de formation pour le personnel qui traite les réclamations à BMT, ni de stratégie pour développer leurs compétences afin d’améliorer la rapidité et l’efficacité dans la résolution des réclamations des clients.

— **Évaluation et Amélioration Continue :**

Il n’est pas spécifié comment les données collectées à partir des réclamations sont analysées et utilisées pour améliorer les processus. Une rétroaction limitée sur les résultats

des réclamations peut compromettre l'apprentissage organisationnel et la capacité à prévenir les réclamations récurrentes.

2.4 Collecte des Données

2.4.1 Définition de la Collecte de Données

La collecte de données est définie comme le processus de rassemblement systématique d'informations provenant de diverses sources pour les analyser et en extraire des connaissances utiles. Ce processus est crucial pour les projets de data science car la qualité des données collectées influence directement la qualité des analyses et des modèles prédictifs qui en découlent[48].

2.4.2 Objectifs Principaux de la Collecte de Données

Garantir la qualité des données : Les données doivent être précises, complètes, cohérentes, et actuelles pour garantir des résultats fiables.

Faciliter l'analyse : Les données bien collectées permettent une analyse efficace et l'application de techniques avancées telles que le machine learning.

Soutenir la prise de décision : Les données collectées et analysées correctement fournissent des insights précieux pour les décisions stratégiques et opérationnelles.

2.4.3 Processus de Collecte de Données

Méthodologie Appliquée dans notre Étude

La collecte de données s'est déroulée en plusieurs étapes clés, chacune essentielle pour garantir l'intégrité et la pertinence des informations recueillies.

1. Définition des Objectifs et des Besoins en Données

Avant de commencer la collecte, il était essentiel de définir clairement les objectifs de l'étude et les types de données nécessaires. Pour notre projet, il s'agissait de collecter des données sur les réclamations clients, les transactions facturées, et les suivis des traitements des réclamations.

2. Identification des Sources de Données

Les principales sources de données utilisées dans cette étude incluent :

- **Factures :** Documents comptables émis par un fournisseur pour établir une créance envers son client suite à la fourniture d'un bien ou à la prestation d'un service (voir Annexe 2).
- **Demandes de Réclamation :** Documents fournissant des détails spécifiques sur les problèmes rencontrés par les clients (voir Annexe 3).
- **Fiches de Suivi :** Documents émis par la Direction des opérations détaillant le type, la quantité et la durée de séjour des conteneurs traités pour un client donné, ainsi que la nature des services fournis (voir Annexe 4).

3. Planification de la Collecte de Données

La collecte de données a été planifiée pour se dérouler sur plusieurs mois afin de garantir l'exhaustivité des informations recueillies. Un calendrier détaillé a été établi, précisant les étapes et les responsables de chaque phase de collecte.

4. Extraction des Données

Les données ont été extraites manuellement des documents disponibles. Cette extraction manuelle a impliqué la numérisation et la saisie des informations pertinentes dans un format structuré, garantissant ainsi l'exactitude et la cohérence des données.

5. Transformation et Normalisation

Les données extraites ont ensuite été transformées pour assurer leur cohérence et leur normalisation. Par exemple, les dates ont été uniformisées, et les types de réclamation ont été catégorisés de manière standardisée.

6. Calculs Nécessaires pour l'Analyse des Réclamations

Les calculs sont une étape cruciale dans l'analyse des réclamations. Deux formules principales sont utilisées pour évaluer les performances logistiques et de traitement des réclamations :

- **Calcul des Pénalités** : Cette formule prend en compte les dates clés des opérations logistiques pour déterminer les jours de pénalité.

$$\text{Nbre_jours_de_pénalité} = (\text{Date_Livraison} - \text{Date_Bon_à_enlever}) - (\text{Date_Bon_à_enlever} - \text{Date_Accostage}) \quad (2.1)$$

- **Calcul des Délais de Traitement** : Cette formule permet de calculer le délai de traitement des réclamations en se basant sur la différence entre la date de réponse et la date de réclamation.

$$\text{Délais_de_traitement} = (\text{Date_réponse} - \text{Date_réclamation}) \quad (2.2)$$

7. Vérification et Nettoyage des Données

Une fois les données collectées, une étape de vérification et de nettoyage a été réalisée pour corriger les erreurs et s'assurer de la qualité des informations. Les incohérences ont été identifiées et corrigées, et les données manquantes ont été complétées autant que possible.

8. Documentation du Processus de Collecte

Chaque étape du processus de collecte a été soigneusement documentée. Cette documentation inclut les méthodes utilisées, les sources de données, les outils employés, et les difficultés rencontrées. Elle sert de référence pour les futures collectes de données et garantit la reproductibilité de l'étude.

9. Chargement des Données

Une fois transformées, les données ont été chargées dans un fichier Excel pour faciliter leur analyse. L'outil Excel a été choisi pour sa capacité à gérer efficacement de grandes quantités de données et à fournir des fonctionnalités d'analyse et de visualisation robustes.

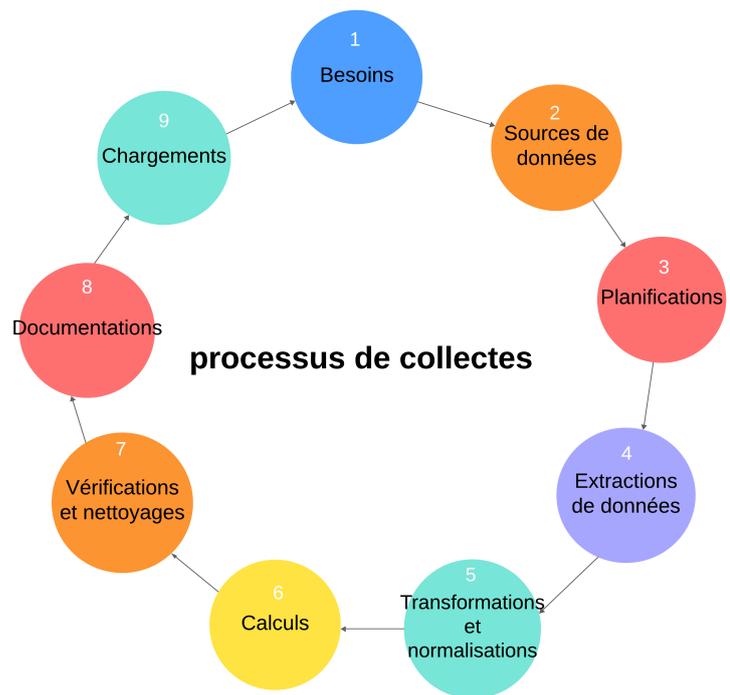


FIGURE 2.6 – Processus de Collecte de Données

Cette collecte de données structurée et rigoureuse a permis de constituer une base solide pour l’analyse des réclamations et l’amélioration des processus internes de l’entreprise. En centralisant les données dans un fichier Excel, nous avons pu non seulement faciliter l’analyse, mais aussi identifier les tendances et les problèmes récurrents, offrant ainsi un outil précieux pour les prises de décision futures et l’amélioration de la satisfaction client.

2.5 Description du Dataset Collecté

Les données collectées sont décrites dans le tableau suivant, qui présente les différentes colonnes et leur signification :

Nom de la colonne	Description
Reclamation	Un identifiant unique pour chaque réclamation
Code_client	Le code unique attribué au client
Nom_client_finale	Le nom du client ayant soumis la réclamation
N_Facture	Le numéro de la facture associée à la réclamation
N_Conteneur	Le numéro du conteneur concerné par la réclamation
Type_Conteneur	Le type de conteneur parmi les valeurs possibles : 20 EVP, 40 EVP, 45 EVP
Classe_Reclamation	La classe de réclamation parmi les valeurs possibles : Erreur opérationnelle, Problèmes de tarification, Réclamations de paiement, Pénalité de retard causée par le client, Erreur de facturation, Problème de conventions, Autres réclamations
Type_Reclamation	Le type de réclamation parmi les valeurs possibles : Les pannes d'engin, Mauvais positionnement du conteneur, Autre erreur opérationnelle, Erreur d'application de la tarification, Demande de réduction, Manque de camions, Problème de douane, Demande de paiement par tranche, Erreur de calcul, Erreur de saisie, Non application des tarifs réduits de 50 % spécifiquement pour ce client, Non application des tarifs réduits de 30 % spécifiquement pour ce client, Autres réclamations, Problème FIFO, Non application des tarifs réduits de 10 % spécifiquement pour ce client, Demande d'annulation de la facture, Surcharge de travail, Didier a des raisons personnelles liées au client, Non application des tarifs réduits de 5 % spécifiquement pour ce client, Non application des tarifs réduits de 20 % spécifiquement pour ce client
Type_Prestation	Le type de prestation concerné parmi les valeurs possibles : Over time, LOLO pour Livraison, Embarquement, Débarquement, Empotage, Shifting à bord, Entreposage, Préparation et nettoyage de TC, Location d'engins, Rapprochement de TC, Traitement de TC Frigorifique, LOLO pour Pesée, Dépotage, Entreposage de TC vide, LOLO pour Restitution, LOLO pour visite (voir annexe 1)
Date_Accostage	La date à laquelle le conteneur a accosté
Date_Bon_a_enlevé	La date à laquelle le bon à enlever a été délivré
Date_Livraison	La date de livraison du conteneur
Nbre_jours_de_pénalité	Le nombre de jours de pénalité associés à la réclamation, calculé à partir des dates d'accostage, de bon à enlever et de livraison
Date_Reclamation	La date à laquelle la réclamation a été soumise
Date_Réponse	La date à laquelle une réponse a été fournie au client
Decision	La décision prise concernant la réclamation parmi les valeurs possibles : Favorable, Défavorable
Détail_de_la_décision	Les détails concernant la décision prise
Justification_du_client	La justification fournie par le client pour la réclamation : Fondé, Non fondé
Procédure	La procédure suivie pour traiter la réclamation : Import, Export
Commision	Indique si la réclamation nécessite de passer par une commission (1) ou non (0)
Delai_traitement	Le délai de traitement de la réclamation en jours, calculé comme la différence entre la Date_Réponse et la Date_Reclamation

TABLE 2.1 – Description du Dataset

2.6 choix de l'outil Power BI

2.6.1 Définition

Power BI

Power BI est une plateforme évolutive et unifiée de Business Intelligence (BI), développée par Microsoft, qui permet aux utilisateurs en entreprise et en libre-service de se connecter à diverses sources de données, de les transformer, de les visualiser sous forme de rapports et de tableaux de bord interactifs. Elle facilite l'intégration transparente des visualisations dans les applications quotidiennes, offrant ainsi une solution complète pour la prise de décision basée sur les données[39].

Power BI distop

Power BI Desktop est une application de bureau gratuite conçue pour la création de rapports interactifs et de visualisations à partir de diverses sources de données. Elle permet aux utilisateurs de développer des modèles de données, de réaliser des requêtes via Power Query, et d'écrire des calculs avec DAX, facilitant ainsi la modélisation et la création de rapports partageables via le service Power BI[41].

Power BI Mobile Apps :

Ce sont les applications mobiles pour Android et iOS de Power BI, conçues pour visualiser les rapports sur les appareils mobiles[12].

Power BI Service :

Il s'agit de la composante basée sur le cloud de Power BI, principalement dédiée à la visualisation des rapports, ainsi qu'à la gestion du partage et de la diffusion de ces rapports au sein et en dehors de l'organisation[12].

Langage M :

est un langage fonctionnel qui s'exécute en arrière-plan de Power Query, est un langage de requête en mashup. Il permet d'interroger diverses sources de données, de les transformer à l'aide de commandes dédiées, et de transférer les résultats vers un modèle de données Power BI[2].

Langage DAX :

Le langage DAX, « Data Analysis eXpressions » en anglais, est un langage de programmation de formules largement utilisé dans les vues Données et Rapport de Power BI. Il se compose

d'une variété de fonctions, d'opérateurs et de constantes qui peuvent être intégrés dans une formule ou une expression. Ces éléments permettent de calculer et de retourner une ou plusieurs valeurs, qui peuvent ensuite être exploitées dans la création de visuels[2].

Power Query :

Power Query, également connu sous le nom d'éditeur de requêtes, est l'outil d'extraction, de transformation et de chargement (ETL) intégré à Power BI Desktop. Il offre aux utilisateurs la possibilité d'extraire des données à partir de multiples sources, de les transformer selon leurs besoins spécifiques, puis de charger ces données finales dans Power BI Desktop[2].

2.6.2 Fonctionnement de Power BI

- Connecter des données, notamment à partir de plusieurs sources de données.
- Mettre en forme les données avec des requêtes qui génèrent des modèles de données pertinents et convaincants.
- Utiliser les modèles de données pour créer des visualisations et des rapports.
- Partager les fichiers de rapport permet à d'autres utilisateurs de les exploiter, de les développer et de les partager. Un fichier Power BI Desktop (.pbix) peut être partagé comme n'importe quel autre fichier, mais il est préférable de le charger sur le service Power BI.

2.6.3 Comparaison entre Power BI et Excel

Ce tableau compare les capacités d'Excel et de Power BI dans différents critères liés à la gestion et à l'analyse des données. Power BI se distingue par ses fonctionnalités avancées de visualisation et de gestion des données en temps réel, tandis qu'Excel est plus traditionnel avec des fonctionnalités robustes mais moins interactives.

Critère	Excel	Power BI
Connexions de modèles de données	Certains graphiques ne peuvent pas se connecter au modèle de données	Idéal pour les tableaux de bord, les KPI et les visualisations en temps réel
Qualité des rapports	Rapports ordinaires	Offre des rapports de qualité supérieure par rapport à Excel.
Filtrage croisé	Pas de filtrage croisé avancé entre les graphiques	Fonctionnalités avancées de filtrage croisé entre les graphiques
Niveau d'analyse	Analyses simples	Analyses de haut niveau
Modélisation des données	Modèles de données structurés et simples	Ingestion efficace et création simplifiée de modèles de données complexes
Association de tables	Difficile de relier les tableaux séparés	Facile d'associer des tables séparées
Gestion des données	Gère un ensemble de données limité	Capable de gérer un ensemble de données volumineux
Création de rapports	Fonctionnalités limitées pour créer des rapports	Création de tableaux de bord personnalisés
Vitesse de traitement	Vitesse de traitement lente avec tableaux, graphiques et filtres	Traitement plus rapide avec tableaux, graphiques et filtres
Usage principal	Principalement utilisé pour l'analyse approfondie	Principalement utilisé pour la visualisation et le partage de tableaux de bord
Langage de requête	Langage MDX pour interroger le modèle de données	Langage DAX pour interroger le modèle de données
Facilité d'utilisation	Applications un peu difficiles à utiliser	Outil puissant et facile à utiliser

TABLE 2.2 – Comparaison des fonctionnalités entre Excel et Power BI

2.7 Conclusion

Ce chapitre a fourni une vue d'ensemble de BMT-SPA, en détaillant son historique, sa situation géographique, ses valeurs et sa structure organisationnelle. Nous avons également examiné la procédure actuelle de traitement des réclamations, identifiant les principaux défis comme le manque de centralisation et la complexité administrative. Enfin, nous avons décrit le processus de collecte de données, soulignant l'importance de la qualité des données pour soutenir une analyse efficace et une prise de décision éclairée. Ces éléments sont cruciaux pour améliorer la gestion des réclamations et optimiser les opérations de BMT.

3

Analyse et conception

3.1 Introduction

Nous présentons dans ce chapitre « analyse et conception » notre travail qui consiste à développer un système centralisé intelligent pour l'amélioration du traitement des réclamations chez BMT. Nous commencerons par présenter les fonctionnalités de notre projet. Ensuite nous consacrerons une large place pour la modélisation de notre application en orienté objet à l'aide du langage UML (Unified Modeling Language) qui permet de bien représenter les aspects statiques et dynamiques de notre projet par la série des diagrammes qu'il offre. Nous terminons par la conception de la base de données utile à notre application en spécifiant les tables à utiliser et leurs structures.

Sommaire

3.1	Introduction	35
3.2	Analyse des besoins	36
3.3	Identification des besoins	37
3.4	Diagramme de contexte statique	38
3.5	Description des cas d'utilisation	39
3.6	Diagrammes de cas d'utilisation	41
3.7	Diagrammes d'activités	44
3.8	Modèle de domaine	47
3.9	conception	47
3.10	Conclusion	54

3.2 Analyse des besoins

3.2.1 Présentation d'UML

Définition UML

UML (Unified Modeling Language) est un langage de modélisation standardisé employé pour visualiser, spécifier, construire et documenter les composants d'un système logiciel. Il permet de créer des diagrammes représentant la structure et le comportement du système, facilitant ainsi la compréhension et la communication entre les développeurs, les analystes et les parties prenantes. UML comprend divers types de diagrammes, tels que les diagrammes de classes, de séquence, de cas d'utilisation et d'activités, pour couvrir différents aspects de la conception et du développement logiciel[9].

3.2.2 Quelques définitions de base

Acteur : Est une entité externe (personne, système ou organisation) qui interagit avec un système logiciel en participant aux cas d'utilisation. Ils définissent les interactions fonctionnelles et les exigences du système.

Scénario : une liste d'actions qui décrivent une interaction entre un acteur et le système.

Interaction : un comportement qui comprend un ensemble de messages échangés par un ensemble d'objets dans un certain contexte pour accomplir une certaine tâche.

Message : Un message représente une communication unidirectionnelle entre objets qui transporte de l'information avec l'intention de déclencher une réaction chez le récepteur.

3.2.3 Contexte du projet

Dans le cadre de ce projet, notre objectif principal est de répondre aux attentes des décideurs et des responsables de BMT en matière d'analyse des données. Pour ce faire, une collecte minutieuse d'informations et une évaluation précise des besoins sont indispensables. Nous portons une attention particulière à la direction marketing, en mettant spécifiquement l'accent sur les réclamations liées au service commercial. Ces réclamations constituent un ensemble crucial à identifier et à traiter efficacement grâce à la mise en place d'un système décisionnel performant.

3.2.4 Identification des acteurs

Dans notre analyse, un acteur représente un rôle assumé par une entité externe telle qu'un utilisateur humain, un dispositif matériel ou un autre système, qui interagit directement avec le système étudié. En d'autres termes, un acteur peut consulter et/ou modifier directement l'état du système.

Dans le cadre de notre étude, nous avons identifié deux acteurs principaux qui interagissent directement avec le système :

1. Clients : Il s'agit de clients individuels transitaires ou d'une compagnie maritimes qui possèdent un compte sur notre application. Ils ont la possibilité de consulter et de modifier les informations pertinentes.

2. Administrateur : Ce rôle détient les droits d'accès à l'espace administrateur. L'administrateur est responsable du maintien de la dynamique du site et veille aux mises à jour des réclamations, assurant ainsi le bon fonctionnement global du système.

3.3 Identification des besoins

3.3.1 Besoins fonctionnels

— Pour le client :

1. Authentification : Permettre aux clients (transitaires ou compagnies maritimes) de se connecter au système en utilisant leur adresse e-mail et un mot de passe géré par l'administration.

2. Consultation des réclamations : Offrir la possibilité aux clients de visualiser leur liste de réclamations existantes.

3. Recherche de réclamations : Permettre aux clients de rechercher des réclamations spécifiques .

4. Gestion des réclamations : Permettre aux clients d'ajouter, supprimer, modifier et valider leurs réclamations.

5. Suivi des réclamations : Permettre aux clients de suivre l'état de leurs réclamations ainsi que les réponses fournies par l'administration.

— Pour l'administrateur :

1. Consultation de la liste des réclamations : Offrir à l'administrateur la possibilité de consulter la liste complète des réclamations enregistrées dans le système ainsi que de l'imprimer.

2. Gestion des réponses : Permettre à l'administrateur de valider ou refuser les réclamations et de fournir une réponse adéquate aux clients.

3. Gestion des comptes clients : Autoriser l'administrateur à créer de nouveaux comptes clients au besoin et à gérer les informations d'authentification associées (e-mail et mot de passe).

4. Consultation du tableau de bord : Fournir à l'administrateur un accès au tableau de bord pour surveiller diverses métriques et indicateurs de performance.

5. Analyse prédictive : Intégrer une fonctionnalité permettant à l'administrateur de consulter une analyse prédictive pour déterminer si une réclamation pourrait entraîner des commissions ou non.

6. Authentification : Assurer également un mécanisme d'authentification sécurisé pour l'accès de l'administrateur au système.

3.3.2 Besoins non fonctionnels

Les besoins non fonctionnels de notre application se concentrent sur les points suivants :

- **Rapidité de traitement** : Avec un grand nombre de transactions chaque jour, il est crucial d'avoir des temps de réponse rapides.
- **Performance** : L'application doit être optimisée pour répondre efficacement à toutes les exigences des utilisateurs, offrant ainsi une expérience fluide et sans faille.
- **Sécurité** : Étant donné la sensibilité des données personnelles impliquées, l'application doit respecter les normes de sécurité les plus strictes pour protéger la vie privée des utilisateurs.
- **Disponibilité** : L'accessibilité à tout moment est essentielle pour répondre aux besoins des utilisateurs, garantissant ainsi une disponibilité continue de l'application.

3.4 Diagramme de contexte statique

Le diagramme de contexte est l'un des premiers diagrammes qui peuvent être réalisés. Il s'agit d'une représentation visuelle montrant les interactions entre un système et les acteurs externes qui y sont connectés [63].

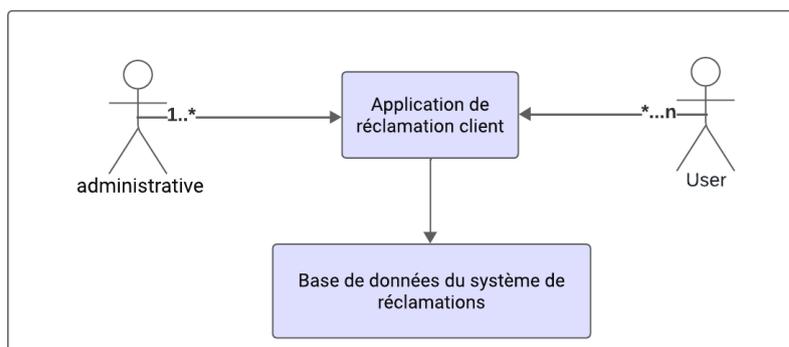


FIGURE 3.1 – Diagramme de contexte

3.5 Description des cas d'utilisation

Afin de mieux comprendre les différents cas d'utilisation de notre système, nous allons fournir une description textuelle pour chaque cas.

3.5.1 Description du cas d'utilisation « MES RÉCLAMATIONS »

une description textuelle du cas d'utilisation MES RÉCLAMATIONS est illustrée dans le Tableaux suivant :

Titre	Mes réclamations
Acteurs	Client : Le client (transitaires ou compagnies maritimes) est un individu qui souhaite déposer une réclamation.(voir annexe 4)
Précondition	L'utilisateur doit s'authentifier
Scénario normal	<ol style="list-style-type: none"> 1. L'utilisateur S'authentifie 2. L'utilisateur sélectionne l'option « Ajouter une réclamation » 3. Le système capture les détails de la réclamation 4. Le client soumet la réclamation
Scénarios alternatifs	<ul style="list-style-type: none"> – Annuler : Le client peut annuler le dépôt d'une réclamation avant de la soumettre – Modifier : Le client peut modifier les détails de la réclamation avant de la soumettre. Cela étend le cas d'utilisation « Ajouter une réclamation » – Consulter l'état de la réclamation : Le client peut consulter l'état d'une réclamation qu'il a déposée. Ceci est un cas d'utilisation distinct

TABLE 3.1 – Description du cas d'utilisation « Mes Réclamations »

3.5.2 Description du cas d'utilisation « GESTION DES RÉCLAMATIONS »

»

une description textuelle du cas d'utilisation GESTION DES RÉCLAMATIONS est illustrée dans le Tableaux suivant :

Titre	Gestion des réclamations
Acteurs	Administrateur : La personne qui gère les réclamations et les comptes clients sur le système.
Précondition	L'administrateur doit s'authentifier
Scénario normal	<ol style="list-style-type: none"> 1. Consultation de la liste des réclamations : Offrir à l'administrateur la possibilité de consulter la liste complète des réclamations enregistrées dans le système ainsi que de l'imprimer. 2. Gestion des réponses : Permettre à l'administrateur de valider ou refuser les réclamations et de fournir une réponse adéquate aux clients. 3. Gestion des comptes clients : Autoriser l'administrateur à créer de nouveaux comptes clients au besoin et à gérer les informations d'authentification associées (e-mail et mot de passe). 4. Consultation du tableau de bord : Fournir à l'administrateur un accès au tableau de bord pour surveiller diverses métriques et indicateurs de performance. 5. Authentification : Assurer également un mécanisme d'authentification sécurisé pour l'accès de l'administrateur au système.
Scénarios alternatifs	<ul style="list-style-type: none"> – Consultation des réclamations : L'administrateur peut rechercher des réclamations spécifiques par filtre ou mot-clé. – Modification des réponses : L'administrateur peut modifier une réponse avant qu'elle ne soit envoyée au client. – Génération de rapports : L'administrateur peut générer des rapports périodiques sur les réclamations pour analyse interne.

TABLE 3.2 – Description du processus de « Gestion des Réclamations »

3.6 Diagrammes de cas d'utilisation

Un diagramme de cas d'utilisation représente les acteurs et les cas d'utilisation ainsi que leurs relations. Il montre en détail les interactions fonctionnelles entre les acteurs et le système, fournissant une vue d'ensemble dynamique et globale du système [63].

3.6.1 Diagramme de cas d'utilisation pour la Gestion Mes Réclamations

Ce diagramme de cas d'utilisation montre les interactions entre un client et le système de gestion de ses réclamations.

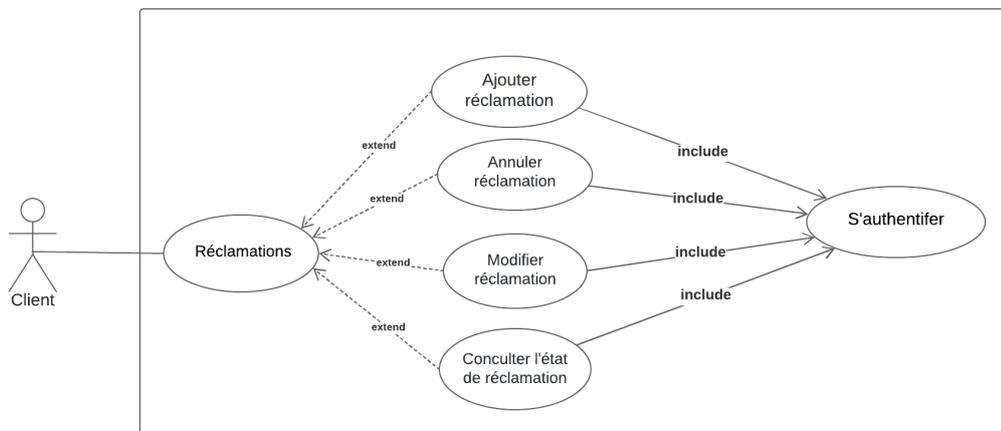


FIGURE 3.2 – Diagramme cas d'utilisation « Mes Réclamations »

3.6.2 Diagramme de cas d'utilisation pour la Gestion des Réclamations

Ce diagramme de cas d'utilisation décrit les interactions entre un administrateur et le système de gestion des réclamations.

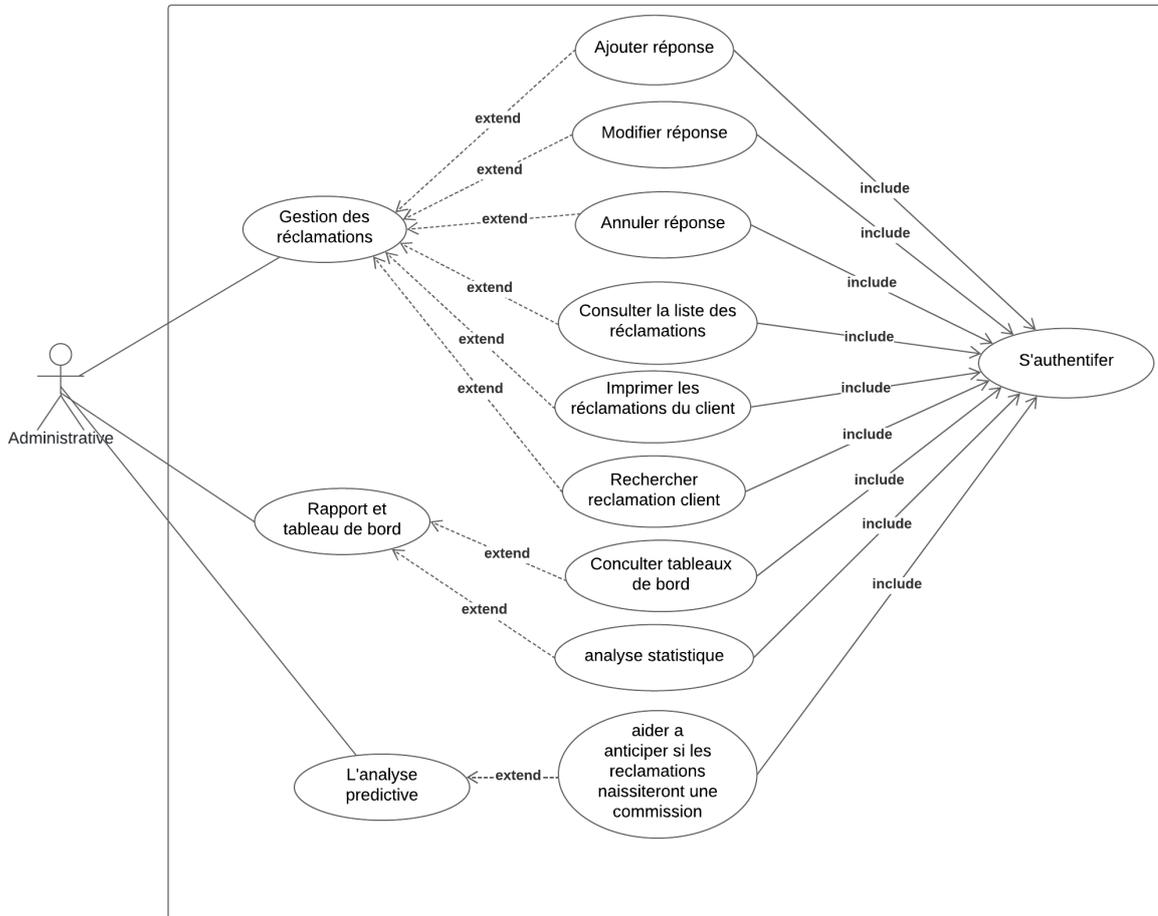


FIGURE 3.3 – Diagramme cas d'utilisation « Réclamations »

3.6.3 Diagramme de cas d'utilisation totale

Ce diagramme de cas d'utilisation illustre les interactions entre les clients et les administrateurs de notre application.



FIGURE 3.4 – Diagramme cas d'utilisation totale

3.7 Diagrammes d'activités

Les diagrammes d'activités servent de représentations graphiques des flux de travail ou des processus au sein d'un système. Ils permettent de mettre en évidence les séquences d'activités, les actions entreprises et les décisions prises tout au long d'un processus donné. Leur utilité première réside dans la capacité à visualiser la logique des flux de travail, à identifier les dépendances entre les différentes tâches, et à modéliser les comportements dynamiques [63].

Ses principaux composants sont :

- **Activités** : Actions ou tâches réalisées dans le cadre du processus.
- **Transitions** : Flux de contrôle entre les différentes activités, illustrant les chemins empruntés.
- **Nœuds de décision** : Points clés où des choix sont effectués, influençant la suite du processus.
- **Début et fin** : Points de départ et d'achèvement du processus, délimitant ainsi son début et sa conclusion.

3.7.1 Diagramme d'activité pour le cas d'utilisation « Authentification » :

Ce diagramme illustre le processus d'authentification dans notre système, offrant une représentation visuelle des actions et des flux d'activités associés à ce scénario spécifique.

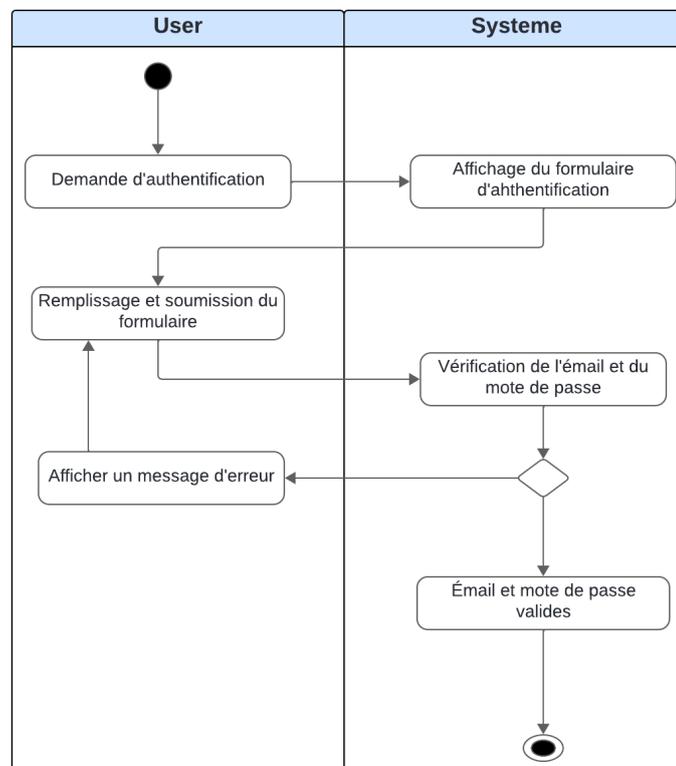


FIGURE 3.5 – Diagramme d'activité pour le cas d'utilisation Authentification

3.7.2 Diagramme d'activité pour le cas « Ajouter Réclamations »

Ce diagramme visualise le processus d'ajout de réclamations, détaillant les actions et les flux associés.

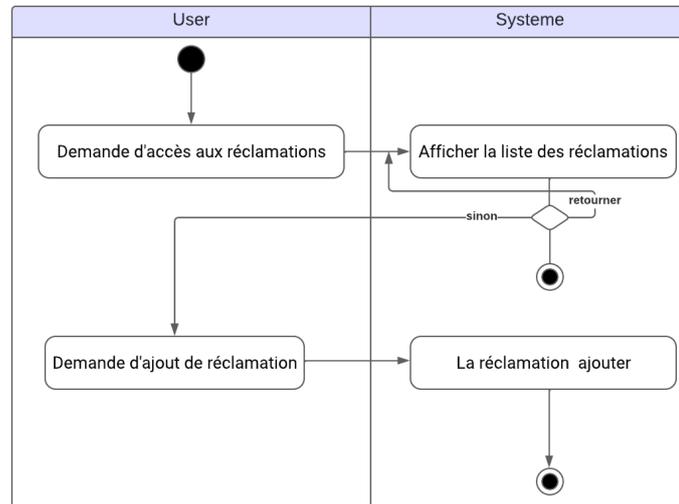


FIGURE 3.6 – Diagramme d'activité ajouter reclamation

3.7.3 Diagramme d'activité pour le cas « Gérer Réclamations »

Ce diagramme illustre le processus de gestion des réclamations, montrant les étapes et interactions impliquées.

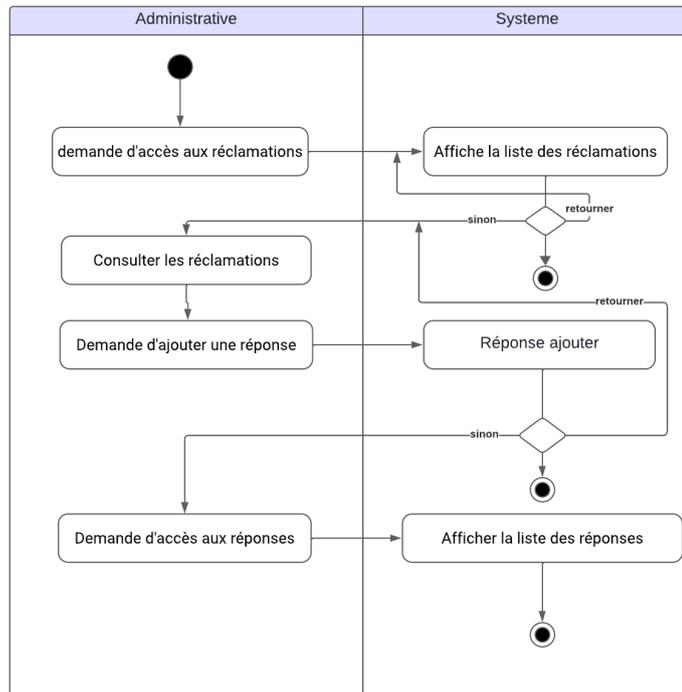


FIGURE 3.7 – Diagramme d'activité pour le cas Gérer Réclamations

3.7.4 Diagramme d'activité pour le cas « Prédire Réclamations »

Ce diagramme illustre le processus de prédiction des réclamations, détaillant les actions et les flux associés.

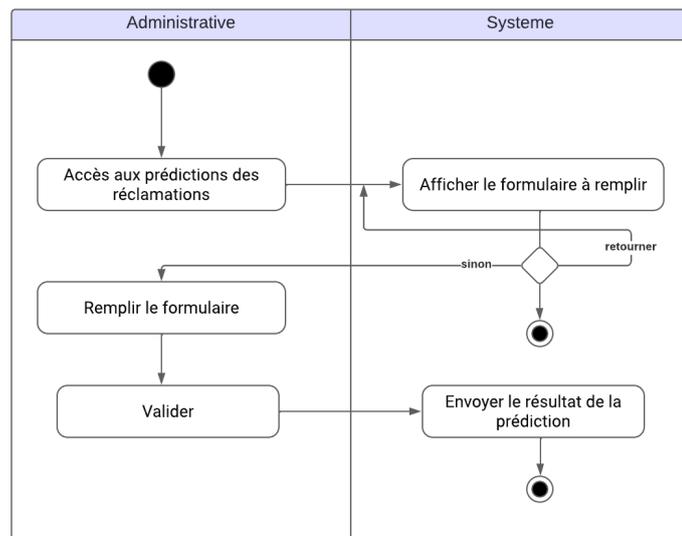


FIGURE 3.8 – Diagramme d'activité prédire reclamation

3.8 Modèle de domaine

Le modèle de domaine représente abstraitement la structure statique des principaux éléments et leurs relations dans le système. Il capture les concepts essentiels, leurs attributs, et associations, facilitant la compréhension et la conception du système [63].

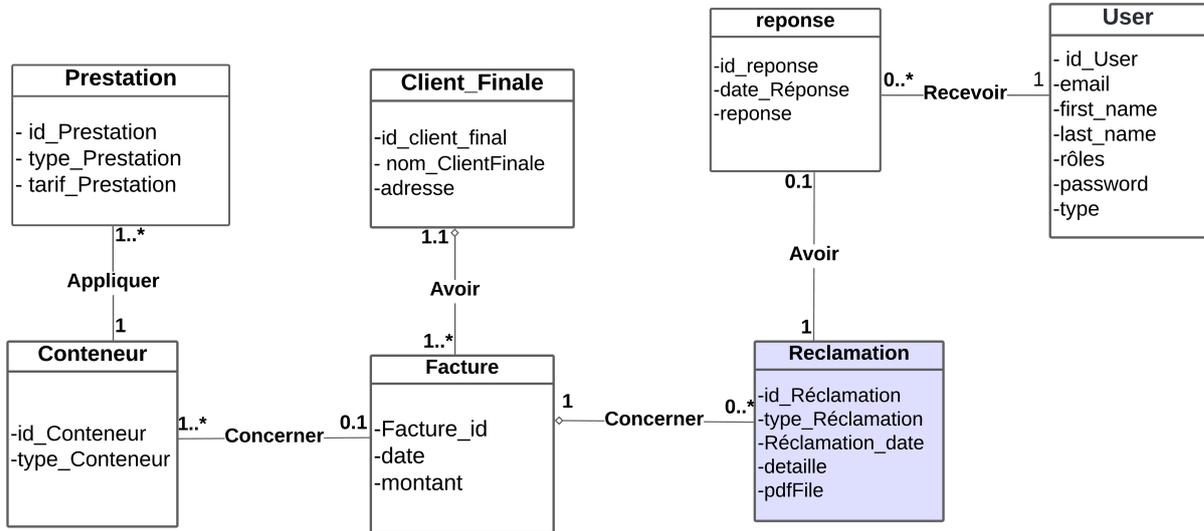


FIGURE 3.9 – Modèle de Domaine

3.9 conception

3.9.1 Diagramme de séquence

Le diagramme de séquence complète le diagramme de cas d'utilisation en fournissant une description détaillée des scénarios. Il illustre le déroulement des interactions entre les éléments du système et les acteurs pour chaque cas d'utilisation, en mettant l'accent sur la chronologie des opérations. Il montre également les objets impliqués dans l'interaction à travers leurs « lignes de vie » et les messages échangés, présentés de manière séquentielle[6].

Diagramme de séquence « authentification »

La figure suivante représente le diagramme de séquence pour l'authentification.

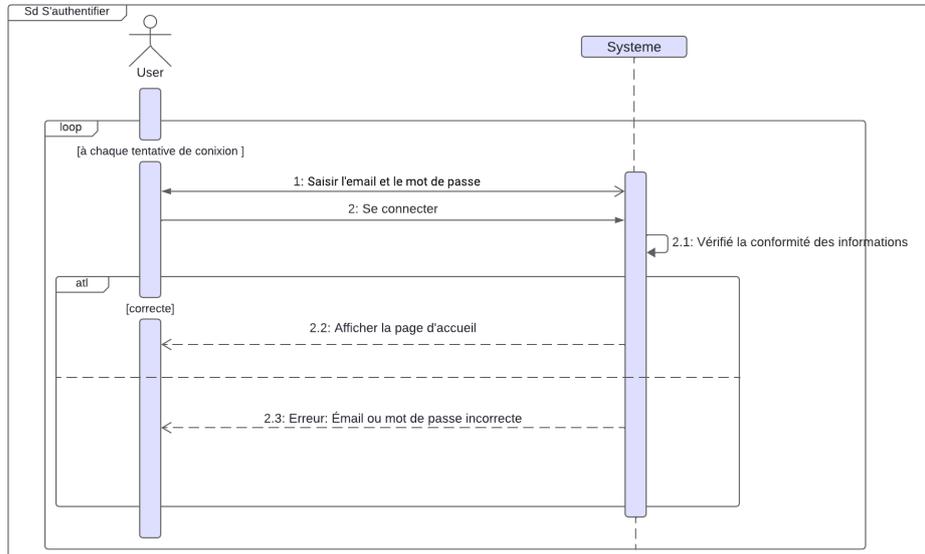


FIGURE 3.10 – Diagramme de séquence « authentification »

Diagramme de séquence « Mes Réclamations »

La figure suivante représente le diagramme de séquence pour « Mes réclamations »

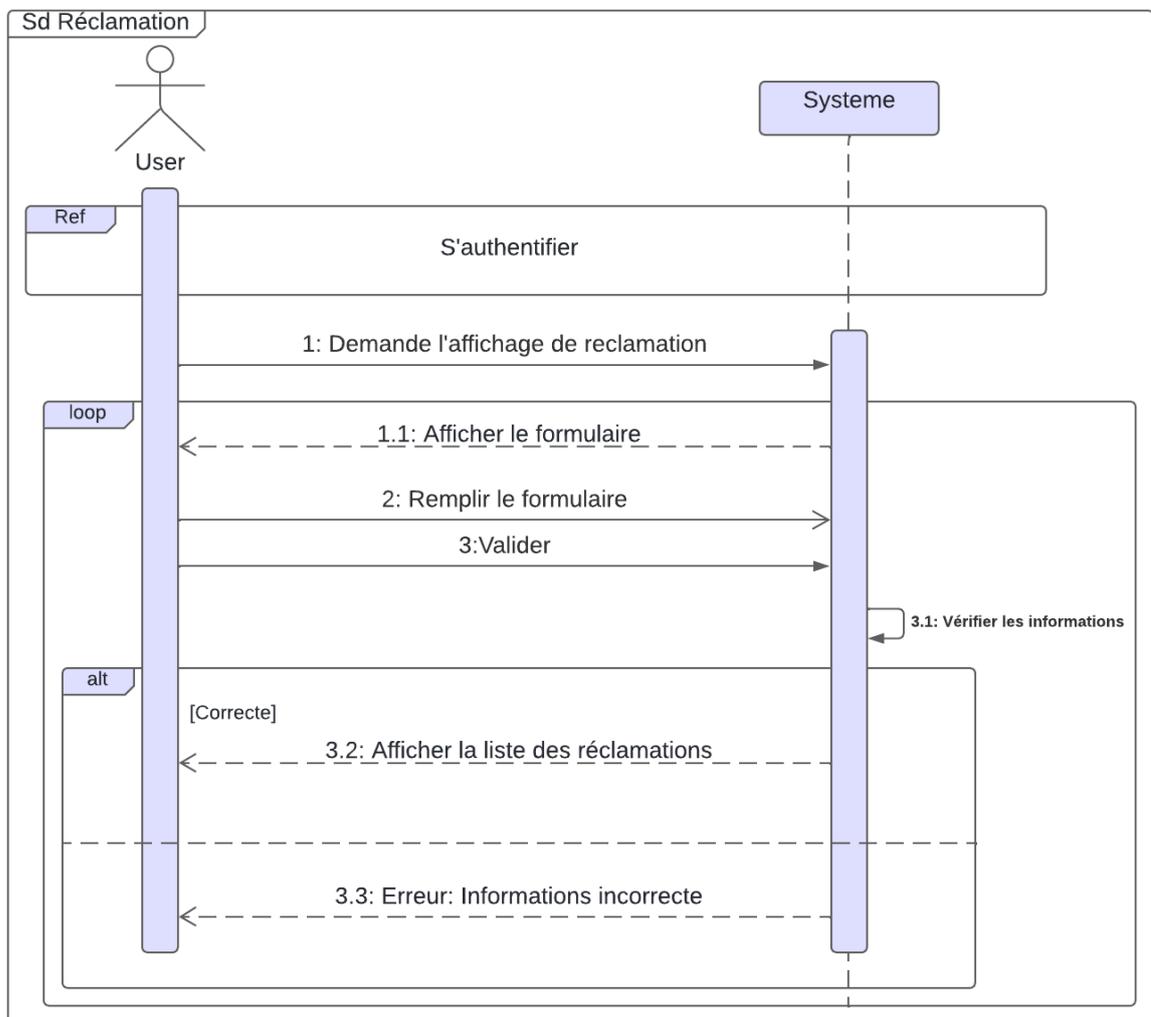


FIGURE 3.11 – Diagramme de séquence (Mes Réclamations)

Diagramme de séquence « Gérer les Réclamations »

La figure suivante représente le diagramme de séquence pour gérer les réclamations.

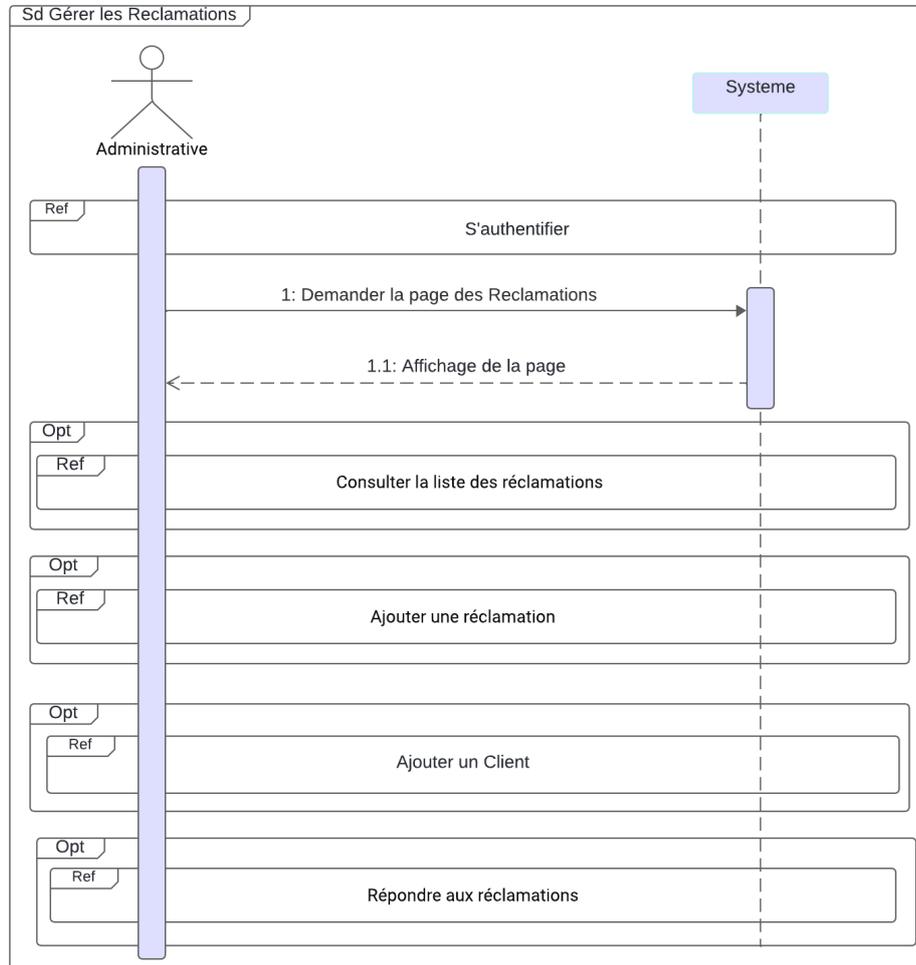


FIGURE 3.12 – Diagramme de séquence « Gérer les Réclamations »

3.9.2 Diagramme de classes

Un diagramme de classes constitue une représentation graphique détaillée de la structure statique d'un système logiciel. Il permet de visualiser les différentes classes qui composent le système, leurs attributs, leurs méthodes et les relations qui les lient entre elles [63].

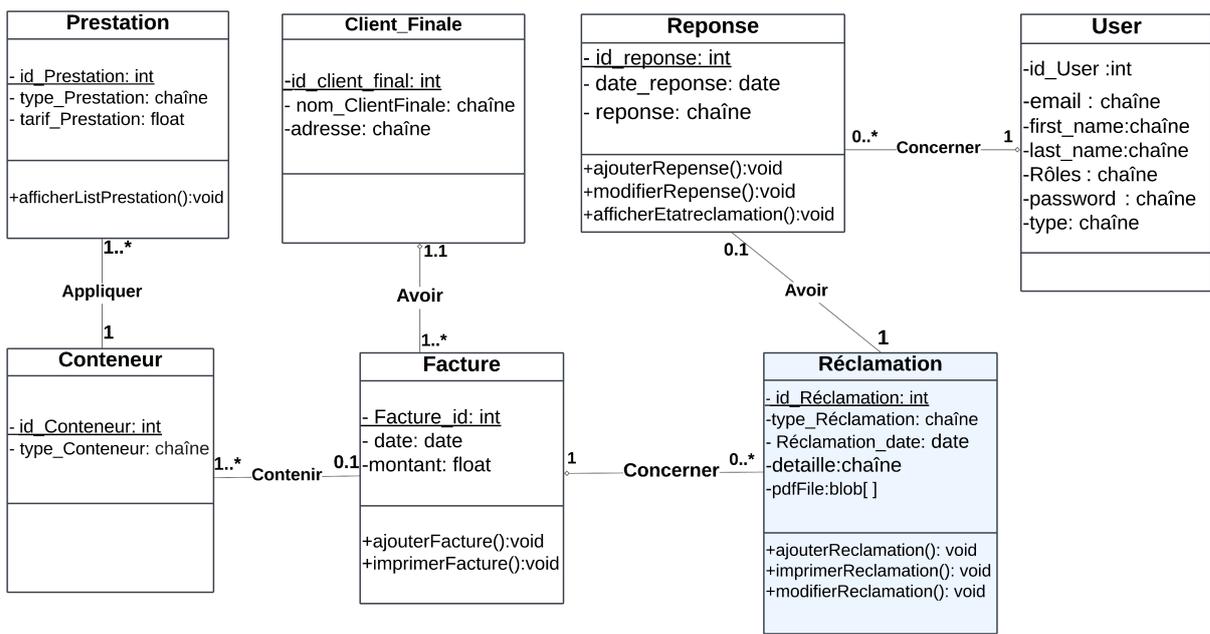


FIGURE 3.13 – Diagramme de classe

3.9.3 Conception de la base de données

Après avoir modélisé notre application à l’aide des différents diagrammes fournis par le langage de modélisation UML, il est essentiel de mettre en œuvre une base de données sur un serveur web. Cela permet d’étendre les possibilités d’interaction avec les utilisateurs, leur offrant ainsi la possibilité de consulter, saisir ou mettre à jour des données, tout en garantissant les droits d’accès qui leur sont accordés. Pour l’implémentation de cette base de données, l’élaboration d’un modèle relationnel de données est nécessaire[15].

subsubsection Les tables de la base de données :

1. Table : Client Finale

Attribut	Type	Description
id_client_final	int	Identifiant unique du client final
nom_client_final	chaîne	Nom du client final
adresse	chaîne	Adresse du client final

TABLE 3.3 – Description des attributs des clients finaux

2. Table : Facture

Attribut	Type	Description
Facture_id	int	Identifiant unique de la facture
id_client_final	int	Référence à l'identifiant du client final (clé étrangère)
date	date	Date de création de la facture
montant	float	Montant total de la facture

TABLE 3.4: Description des attributs de la table des factures

3. Table : Conteneur

Attribut	Type	Description
id_Conteneur	int	Identifiant unique du conteneur
type_Conteneur	chaîne	Type de conteneur
Facture_id	int	Référence à l'identifiant de la facture associée (clé étrangère)

TABLE 3.5: Description des attributs de la table des conteneurs

4. Table : Prestation

Attribut	Type	Description
id_Prestation	int	Identifiant unique de la prestation
type_Prestation	chaîne	Type de prestation
tarif_Prestation	float	Tarif de la prestation
id_Conteneur	int	Référence à l'identifiant du conteneur (clé étrangère)

TABLE 3.6: Description des attributs de la table des prestations

5. Table : Reclamation

Attribut	Type	Description
id_Reclamation	int	Identifiant unique de la réclamation.
type_Reclamation	chaîne	Type de réclamation.
Reclamation_date	date	Date de la réclamation.
detaille	chaîne	Détail de la réclamation.
pdfFile	blob	Fichier PDF associé à la réclamation.

TABLE 3.7: Description des attributs de la table des réclamations

6. Table : Reponse

Attribut	Type	Description
id_reponse	int	Identifiant unique du suivi de réclamation
date_Réponse	date	Date de la réponse
reponse	chaîne	Réponse donnée pour la réclamation
id_Réclamation	int	Référence à l'identifiant de la réclamation (clé étrangère)
id_User	int	Référence à l'identifiant de l'utilisateur (clé étrangère)

TABLE 3.9 – Description des attributs de la table des réponses aux réclamations

7. Table : User

Attribut	Type	Description
id_User	int	Identifiant unique de l'utilisateur
email	chaîne	Adresse e-mail de l'utilisateur
first_name	chaîne	Prénom de l'utilisateur
last_name	chaîne	Nom de famille de l'utilisateur
Rôles	chaîne	Rôles de l'utilisateur
password	chaîne	Mot de passe de l'utilisateur
type	chaîne	Type d'utilisateur

TABLE 3.11 – Description des attributs de la table des utilisateurs

3.9.4 Modèle Relationnel

- **Client Finale** (id_client_final, nom_ClientFinale, adresse)
- **Facture** (Facture_id, **id_client_final**, date, montant)
- **Conteneur** (id_Conteneur, type_Conteneur, **Facture_id**)
- **Prestation** (id_Prestation, type_Prestation, tarif_Prestation, **id_Conteneur**)
- **Reclamation** (id_Réclamation, type_Réclamation, Réclamation_date, detaille, pdfFile)
- **reponse** (id_reponse, date_Reponse, reponse, **id_Réclamation**, **id_User**)
- **User** (id_User, email, first_name, last_name, Rôles, password, type)

3.10 Conclusion

Au cours de la phase d'analyse et de conception, les besoins auxquels le système doit répondre ont été décrits et analysés à l'aide des divers diagrammes proposés par le langage UML. Cela inclut la représentation des centres d'intérêts du profil utilisateur sous forme d'ontologie, utilisée dans le module de recherche, avec une description détaillée de la construction de cette ontologie. Par la suite, la conception du système a permis de définir l'ensemble des fonctionnalités nécessaires pour répondre à ces besoins. Le chapitre suivant sera consacré à la réalisation du système, en explicitant les plateformes utilisées ainsi que les moyens mis en œuvre pour sa conception.

4

Implémentation de la Solution Analytique et Prédictive

4.1 Introduction

Ce chapitre se concentre sur l'intégration stratégique des outils de Business Intelligence (BI) et des modèles de Machine Learning pour optimiser la gestion des réclamations chez BMT-Spa. En mettant l'accent sur l'élaboration d'indicateurs de performance clés et l'utilisation de Power BI pour l'analyse proactive des données, nous visons à améliorer la réactivité et la qualité des décisions organisationnelles. De plus, nous explorerons comment les modèles prédictifs en Python enrichissent cette approche, offrant ainsi une vision prédictive précieuse pour anticiper les besoins d'intervention des commissions.

Sommaire

4.1 Introduction	55
4.2 Élaboration des indicateurs de performance essentiels intégrés dans notre solution BI	56
4.3 Implémentation de Solutions Power BI	57
4.4 Implimentation de solution de prediction	77
4.5 Réalisation d'une application de Prédiction avec Streamlit	82
4.6 Conclusion	86

4.2 Élaboration des indicateurs de performance essentiels intégrés dans notre solution BI

Un indicateur clé de performance (KPI) est une mesure quantifiable utilisée pour évaluer le succès d'une organisation, d'une équipe ou d'un individu dans la réalisation d'objectifs spécifiques. Les KPI fournissent une indication claire de la performance par rapport aux objectifs stratégiques définis, facilitant ainsi la prise de décision basée sur les données[17].

KPI	Description
Volume global des réclamations	Affiche le nombre total de réclamations enregistrées, mesurant ainsi la charge de travail et les problèmes rencontrés par les clients
Distribution des réclamations par classe	Identifie les types de problèmes les plus fréquents en répartissant les réclamations par différentes classes
Suivi mensuel des réclamations	Observe l'évolution au fil du temps et repère les variations saisonnières ou les périodes de forte activité, ce qui permet d'ajuster les ressources et les stratégies de gestion
Répartition des réclamations par type	Permet de comprendre les problèmes spécifiques rencontrés par les clients et d'adapter les processus pour les résoudre efficacement
Répartition des réclamations par type de prestation	Identifie les services les plus problématiques en classant les réclamations selon les différents types de prestation
Répartition des réclamations par commission	Évalue l'efficacité des différentes commissions dans la résolution des réclamations
Répartition des décisions pour les réclamations	Comprendre les résultats et les tendances des décisions prises pour chaque réclamation (favorable, défavorable)
Taux de résolution en 1 mois	Pourcentage de réclamations résolues dans un délai d'un mois, mesurant l'efficacité et la rapidité de la résolution
Temps moyen de traitement mensuel	Mesure le temps moyen de traitement mensuel des réclamations pour suivre les performances et identifier les périodes de pic d'activité
Temps moyen de traitement total	Mesure le temps moyen de traitement de toutes les réclamations, permettant une évaluation globale de l'efficacité du processus de traitement
Délai maximum de traitement	Délai maximum pris pour traiter une réclamation, identifiant ainsi les cas extrêmes nécessitant une analyse approfondie
Délai minimum de traitement	Délai minimum pris pour traiter une réclamation, mettant en lumière les bonnes pratiques et les cas de résolution rapide
Classement des clients par nombre de réclamations	Identifier les clients les plus fréquents en réclamations, mettant en évidence ceux nécessitant une gestion particulière de leur relation client

TABLE 4.1 – Principaux KPI pour l'analyse des réclamations

4.3 Implémentation de Solutions Power BI

4.3.1 Installation et Configuration de Power BI

Installation de Power BI Desktop

Nous pouvons obtenir Power BI Desktop de deux façons : soit en l'installant en tant qu'application depuis le Microsoft Store, soit en téléchargeant directement le programme d'installation de Power BI Desktop. Les deux approches permettent d'obtenir la dernière version de Power BI Desktop sur notre ordinateur.

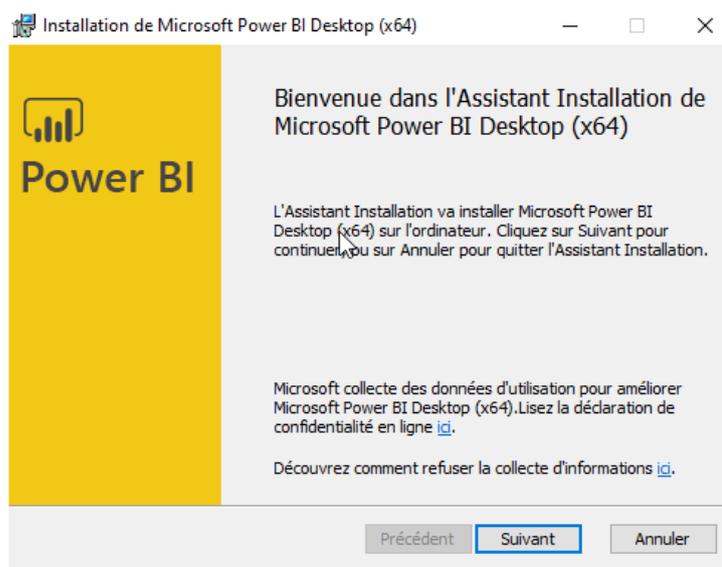


FIGURE 4.1 – Installation de Power BI Desktop

Configuration de Power BI Desktop

Après l'installation, nous lançons Power BI Desktop à partir du menu Démarrer ou du raccourci sur le bureau. Lors du premier lancement, nous pouvons nous connecter avec notre compte Microsoft. Cliquez sur « Se connecter » et entrez vos informations de connexion. Cette étape est facultative mais recommandée pour accéder aux services en ligne de Power BI.

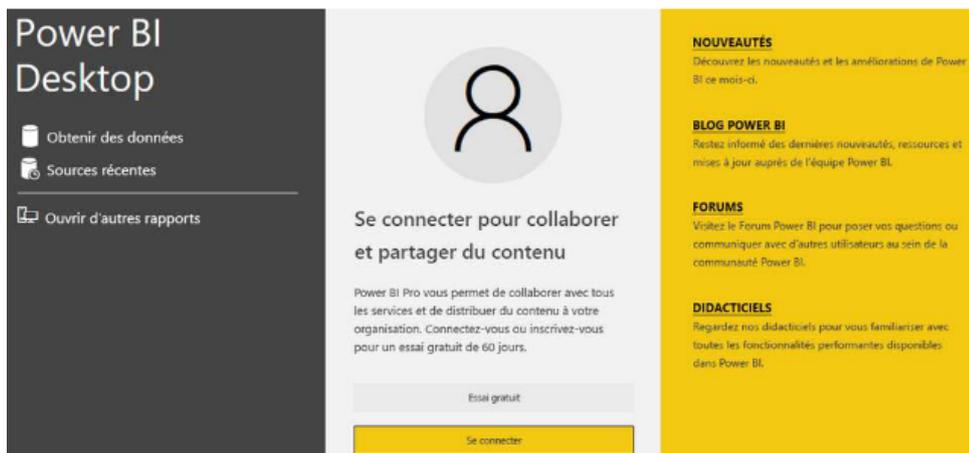


FIGURE 4.2 – Écran de connexion de Power BI Desktop

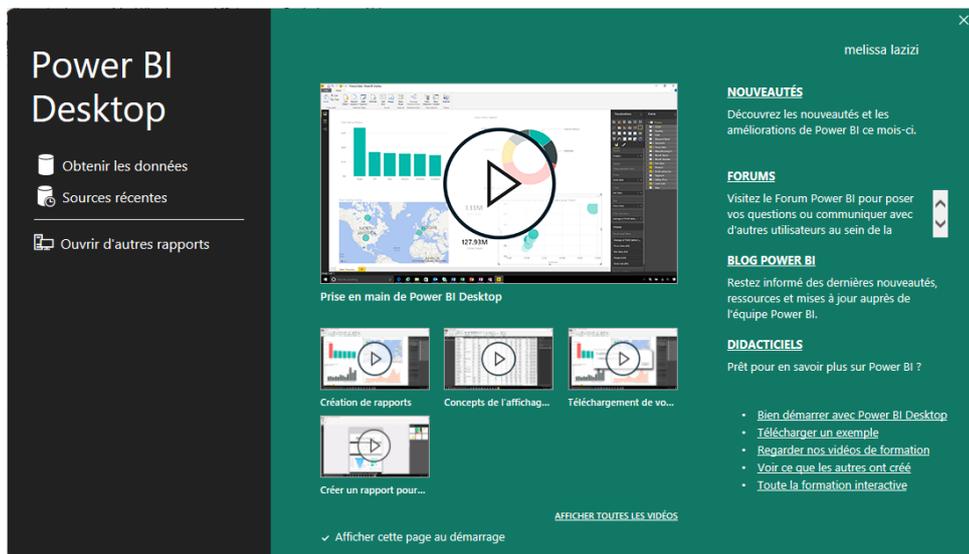


FIGURE 4.3 – Écran d'accueil de Power BI Desktop

4.3.2 Collecte et préparation de données

Identifier la source de données

Power BI Desktop nous permet de nous connecter à des données provenant de nombreuses sources différentes. Pour nous connecter aux données Excel, nous utilisons le ruban Accueil pour afficher le menu des types de données les plus courants. Ensuite, nous sélectionnons l'étiquette du bouton Obtenir des données, puis nous choisissons Classeur Excel.

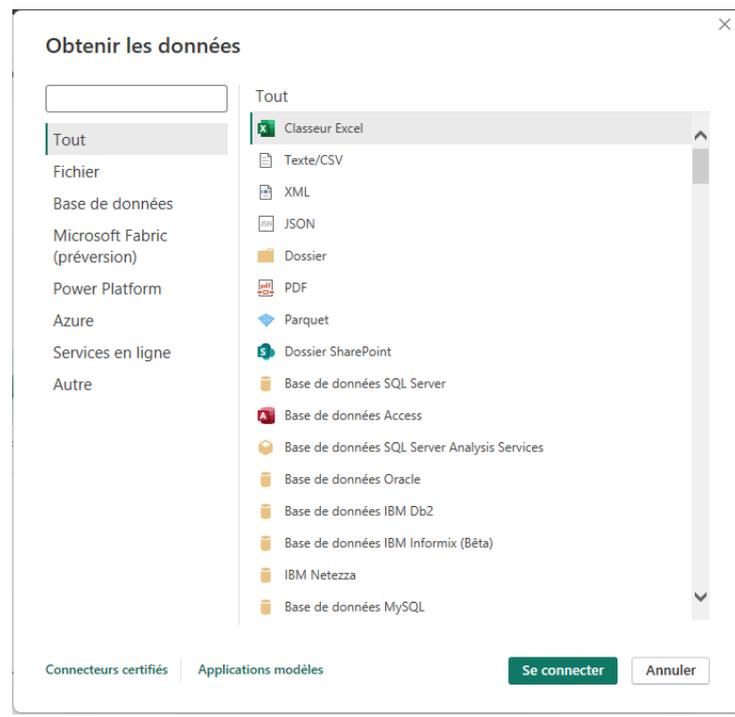


FIGURE 4.4 – Sélection de source de données dans Power BI Desktop

Charger les données

Après avoir sélectionné notre source de données et fourni les informations requises, nous cliquons sur « Charger » pour importer les données dans Power BI Desktop.

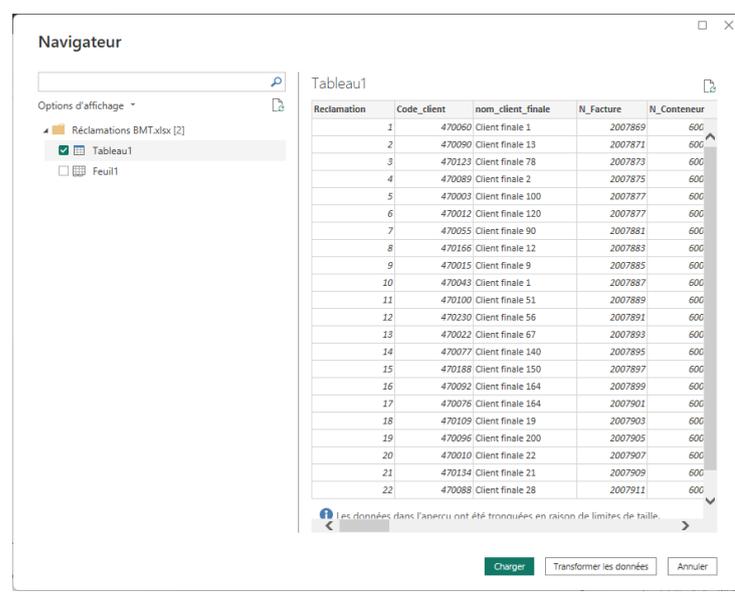


FIGURE 4.5 – Importation des Données de Réclamations dans Power BI

Après avoir chargé nos données, cette procédure nous donne accès à l’interface utilisateur de Power BI. Nous pouvons alors mettre en forme les données selon nos besoins pour créer des

rapports visuels localement.

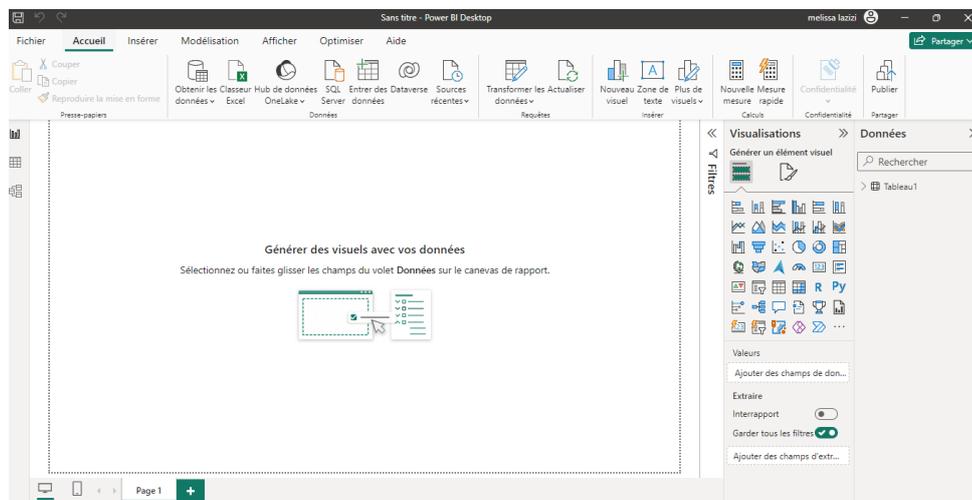


FIGURE 4.6 – Interface utilisateur de Power BI

Une fois les tables chargées, le volet Données affiche les informations. Nous pouvons développer notre table en cliquant sur la flèche à côté de son nom. Dans l'image suivante, la table « Tableaux 1 » est déployée, montrant chacun de ses champs.

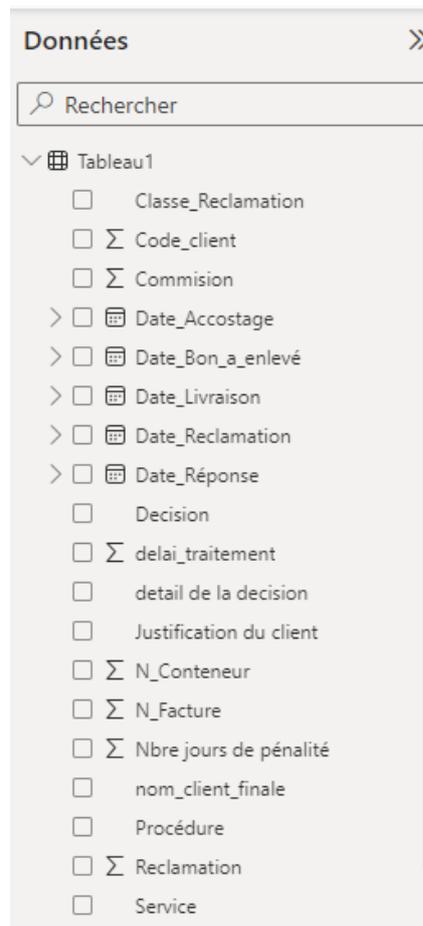


FIGURE 4.7 – Affichage des données dans le volet Données

Intégration des données à travers Power BI

Une fois connectés aux données, Power BI Desktop nous permet de créer des visuels et des rapports. Les données importées nécessitent parfois une transformation pour une présentation adéquate. En cliquant sur la requête "Transformer les données », nous accédons à Power Query, où nous pouvons nettoyer et transformer les données avant de les transférer dans notre modèle. Les étapes de transformation sont mémorisées, permettant des retours en arrière si nécessaire.

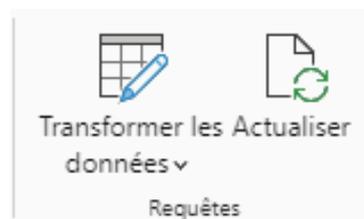


FIGURE 4.8 – Transformation des Données dans Power BI

a. Renommer l'entête de la table

Renommer l'entête de la table « Tableau 1 » par « Réclamations Client ». Cela permettra de

mieux organiser nos données et de rendre nos rapports plus lisibles et compréhensibles pour les utilisateurs finaux.

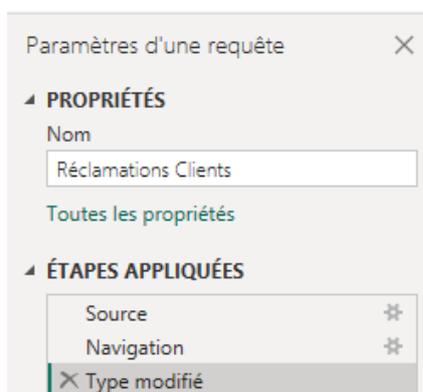


FIGURE 4.9 – Renommer l'entête de la table

b. Suppression des lignes vides

Nous souhaitons supprimer les lignes vides de la requête « Réclamations Clients ». Nous avons remarqué que plusieurs lignes sont complètement vides, ce qui peut nuire à la qualité et à la lisibilité de nos données.

Pour cela, nous allons utiliser l'outil de filtrage dans Power Query afin de supprimer toutes les lignes qui ne contiennent aucune information utile.

FIGURE 4.10 – Filtrage des lignes vides

Pour supprimer les lignes vides, dans le ruban de l'éditeur de requête, nous cliquons sur l'option « Réduire les lignes », puis sélectionnons « Supprimer les lignes » et ensuite « Supprimer les lignes vides ». Cela nous permettra d'éliminer toutes les lignes qui ne contiennent aucune donnée.

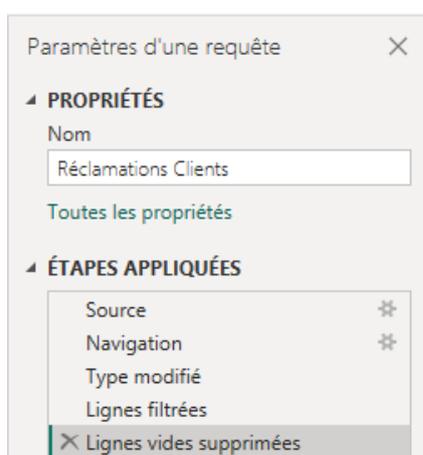


FIGURE 4.11 – Suppression des lignes vides

c. Suppression des doublons

Pour afficher les réclamations en double, la distribution des valeurs et les statistiques des colonnes de réclamation nous montrent les IDs des réclamations en doublon.

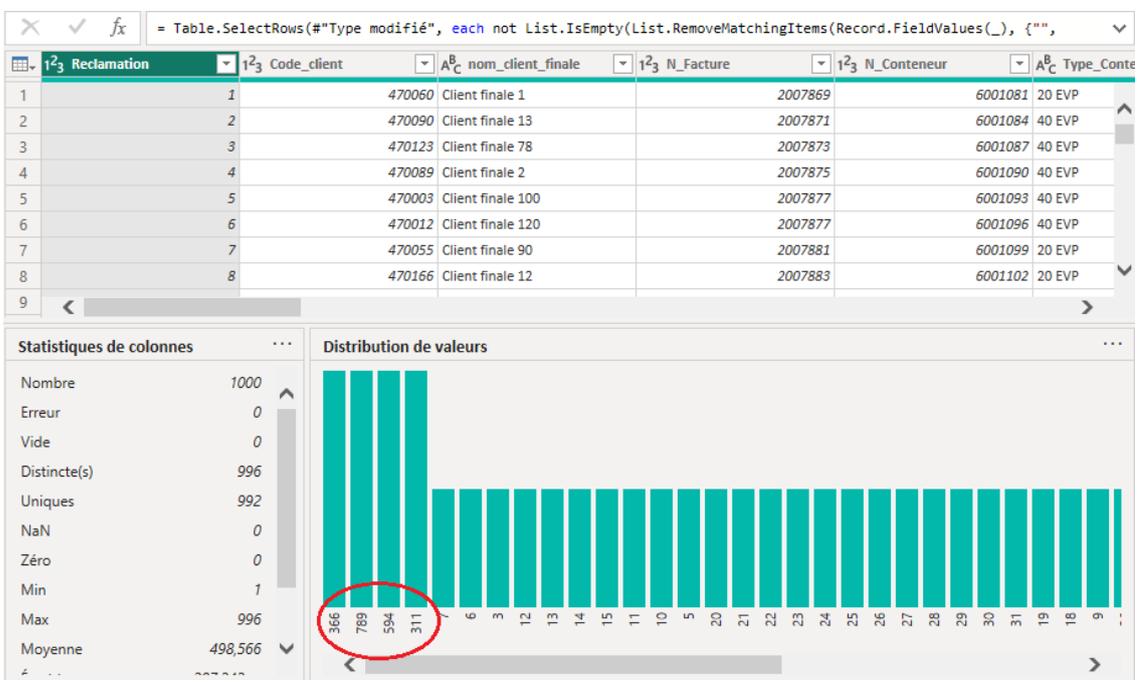


FIGURE 4.12 – Affichage des réclamations en double

Pour supprimer les lignes vides, dans le ruban de l’éditeur de requête, nous cliquons sur l’option « Réduire les lignes », puis sélectionnons « Supprimer les lignes » et ensuite « Supprimer les lignes vides ». Cela nous permettra d’éliminer toutes les réclamations en double.

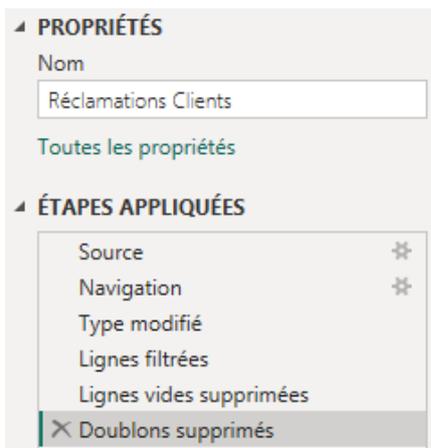


FIGURE 4.13 – suppression des doublons de la colonne « Réclamation »

d. Renommer les noms des colonnes

Pour modifier l'en-tête d'une colonne, nous cliquons sur le nom de la colonne, puis nous choisissons « Renommer » dans le menu contextuel, nous saisissons le nouveau nom, puis appuyons sur Entrée.
On refait les mêmes étapes pour les autres colonnes.

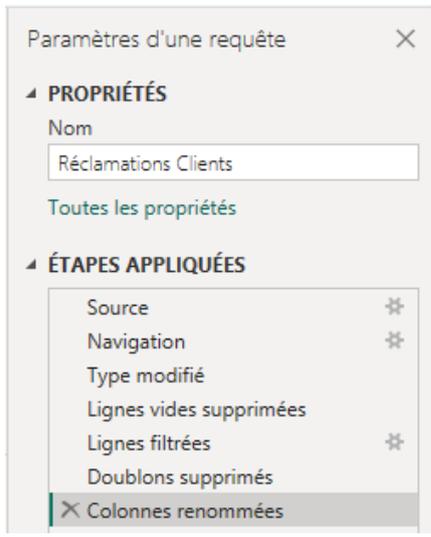


FIGURE 4.14 – Renommer les noms des colonnes

e. Modification du Type de Données

Pour modifier le type de donnée « Date.livraison » en « Date », cliquons sur le bouton droit sur l'en-tête de « Date.livraison », puis sélectionnons « Modifier le type », et choisissons le type « Date ».

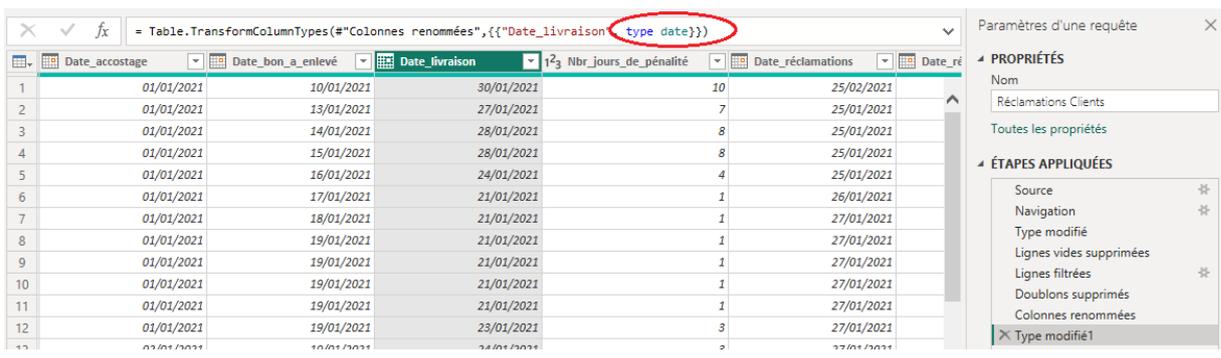


FIGURE 4.15 – Modification du Type de Données dans Power Query

Après avoir effectué toutes les modifications sur la requête « Réclamations Clients », nous sélectionnons « Fermer et appliquer » depuis le Menu Fichier.



FIGURE 4.16 – Fermeture de l’Éditeur de requêtes

Actuellement toutes les requêtes sont transférées dans l’environnement Power BI et nous pouvons créer des visuels pour notre rapport.

4.3.3 Modélisation des données

L’ajout de mesures consiste à créer des calculs dynamiques personnalisés en utilisant le langage DAX (Data Analysis Expressions). Ces mesures permettent d’effectuer des analyses spécifiques et complexes selon les besoins de l’utilisateur. Elles sont utilisées dans des visuels, des rapports et des tableaux de bord pour fournir des insights approfondis sur les données.

Taux de résolution en un mois :

Cette mesure nous permet de calculer le pourcentage de réclamations clients résolues dans un délai de 30 jours, mesurant ainsi l’efficacité et la rapidité du service client sur une période mensuelle :

```

1 Taux de Résolution dans le 1 mois =
2 DIVIDE(
3     CALCULATE(
4         COUNTROWS('Réclamations Clients'),
5         'Réclamations Clients'[Délai_traitement] <= 30
6     ),
7     COUNTROWS('Réclamations Clients')
8 ) * 100

```

FIGURE 4.17 – Taux de résolution en un mois

Taux de résolution en un trimestre :

Cette mesure nous permet de calculer le pourcentage de réclamations clients résolues dans un délai de 90 jours, mesurant ainsi l’efficacité et la rapidité du service client sur une période trimestrielle :

```

1 Taux de résolution en un trimestre =
2 DIVIDE(
3     CALCULATE(
4         COUNTROWS('Réclamations Clients'),
5         'Réclamations Clients'[Délai_traitement] <= 90
6     ),
7     COUNTROWS('Réclamations Clients')
8 ) * 100
9

```

FIGURE 4.18 – Taux de résolution en un trimestre

Temps Moyen de Traitement Total :

Cette mesure nous permet de déterminer la durée moyenne nécessaire pour traiter les réclamations clients, évaluant ainsi l’efficacité et la rapidité du service client au cours des trois dernières années :

```

1 Temps Moyen de Traitement Total =
2 FORMAT(
3     ROUND(
4         AVERAGE('Réclamations Clients'[Délai_traitement]),
5         0
6     ),
7     "0 jours"
8 )

```

FIGURE 4.19 – Temps Moyen de Traitement Total

4.3.4 Conception de rapport

La conception de Power BI présente toutes les visualisations et tableaux de bord améliorés, essentiels pour une analyse efficace des données. Ces outils permettent de visualiser, filtrer et interpréter les informations de manière intuitive, facilitant ainsi la prise de décisions éclairées.

En utilisant le visuel « segment », nous pouvons filtrer les données des trois dernières années, ce qui permet de gérer l’ensemble de notre rapport Power BI.

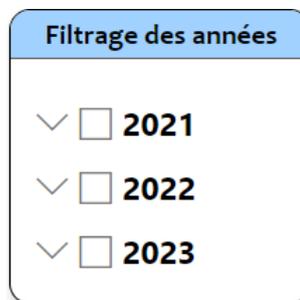


FIGURE 4.20 – Filtrage des années

Vue générale des réclamations

Nombre total des réclamations

En utilisant le visuel « cart », cette visualisation met en évidence le total des réclamations enregistrées durant les trois dernières années (2021, 2022, 2023)



FIGURE 4.21 – Nombre total des réclamations

Après avoir filtré notre rapport, ce tableau présente le total de réclamations pour chaque année :

Année	Total de réclamations
2021	535
2022	238
2023	727

TABLE 4.2 – Total de réclamations par année après filtrage

Classement des 10 clients les plus actifs

En utilisant le visual « graphique en aire » cette visualisation présente les 10 clients ayant généré le plus de réclamations sur une période donnée.

Durant les trois dernières années, le client le plus actif a été le client 164, avec un total de 91 réclamations enregistrées.

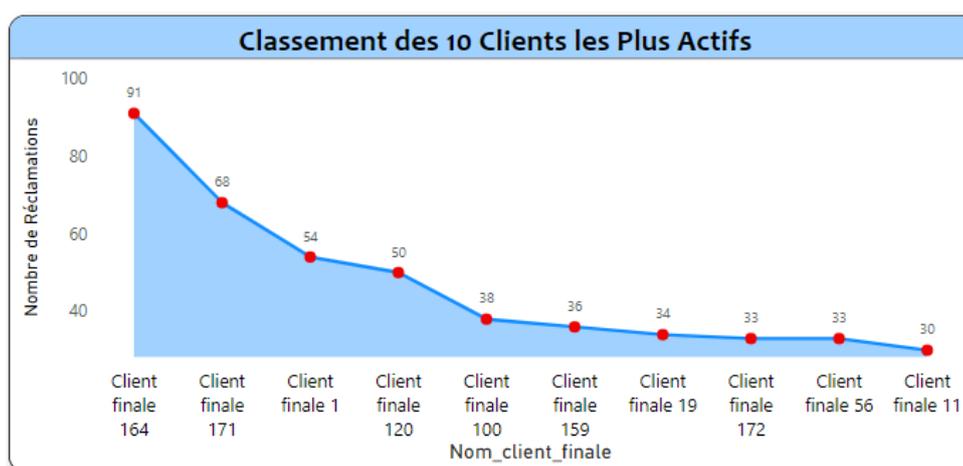


FIGURE 4.22 – Classement des 10 clients les plus actifs

Après avoir filtré notre rapport, ce tableau présente le client ayant généré le plus de réclamations au cours des trois dernières années :

Année	Client	Nombre de réclamations
2021	164	39
2022	171	14
2023	164	45

TABLE 4.3 – Réclamations par client pour les années 2021 à 2023

Nombre de réclamations par mois

En utilisant le visual « l'histogramme empilé », cette visualisation nous permet d'afficher le total des réclamations pour chaque mois au cours des trois dernières années.

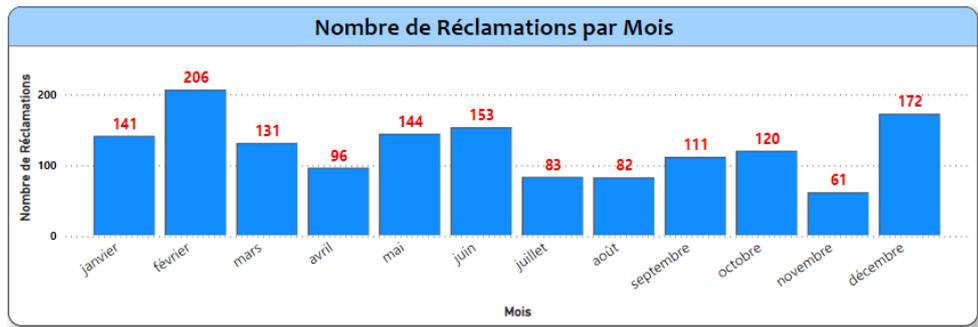


FIGURE 4.23 – Nombre de réclamations par mois

Après avoir filtré notre rapport, ce tableau présente le mois ayant le plus de réclamations pour chaque année :

Année	Mois avec le plus de réclamations	Nombre de réclamations
2021	Février	142
2022	Juin	35
2023	Décembre	129

TABLE 4.4 – Mois avec le plus de réclamations par année après filtrage

Nombre de réclamations par classe de réclamation

En utilisant le visuel « graphique en secteur », cette visualisation nous permet d’afficher le pourcentage total pour chaque classe de réclamation au cours des trois dernières années.

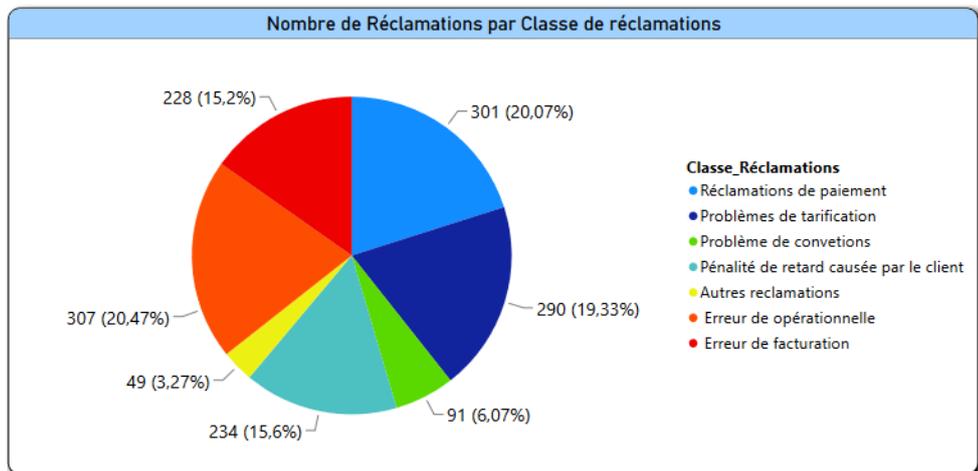


FIGURE 4.24 – Nombre de réclamations par classe de réclamation

Après avoir filtré notre rapport, ce tableau présente le pourcentage de la classe de réclamation ayant le plus de réclamations pour chaque année :

Année	Classe de réclamation	Pourcentage (%)
2021	Erreur Opérationnelle	22,24
2022	Errur de facturation	27,73
2023	Erreur Opérationnelle	22,15

TABLE 4.5 – Pourcentage des classes ayant le plus de réclamations au fil des années

Tableau de bord : Vue générale des réclamations

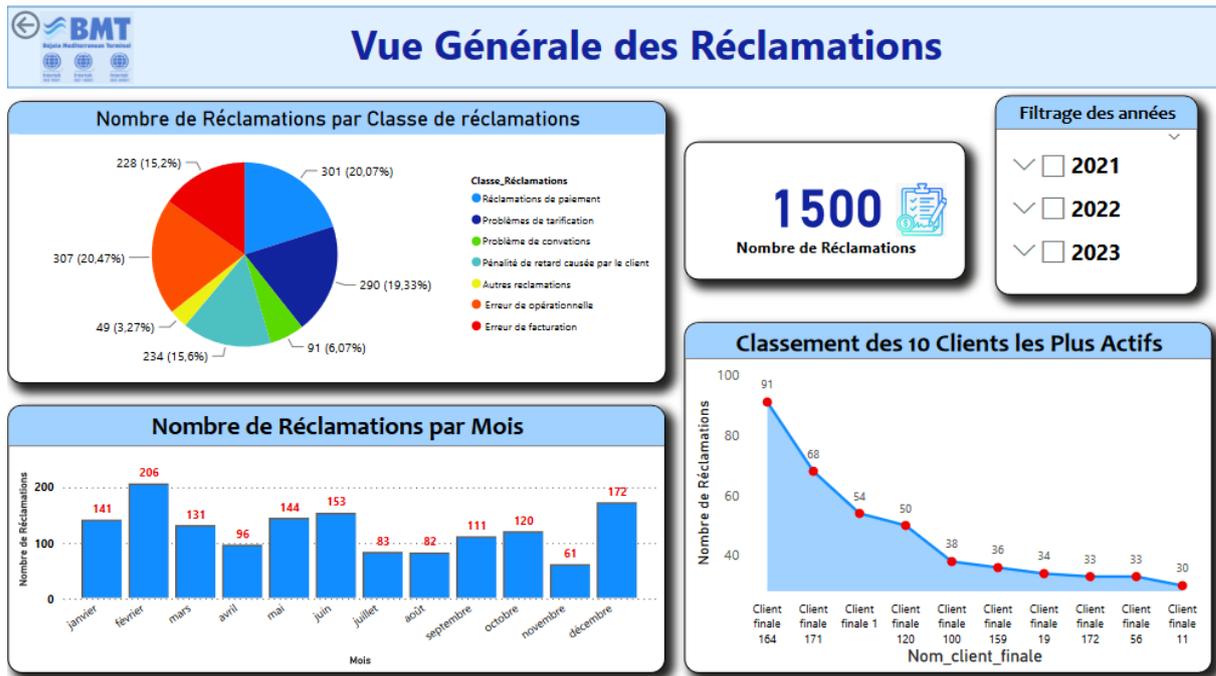


FIGURE 4.25 – Vue générale des réclamations

Ce tableau de bord répond aux questions suivantes :

1. Quel est le nombre total de réclamations enregistrées et comment ce nombre a-t-il évolué au cours des trois dernières années ?
2. Comment se répartissent les réclamations par catégorie (classe) sur les trois dernières années ?
3. Quels sont les dix clients ayant déposé le plus de réclamations au cours des trois dernières années ?
4. Quelle est la répartition mensuelle des réclamations sur les trois dernières années ?

Analyse Détaillée des Réclamations

Nombre de réclamations par type de réclamation

En utilisant le visuel « graphique en secteur », cette visualisation permet de représenter la répartition en pourcentage de chaque type de réclamation au cours des trois dernières années.

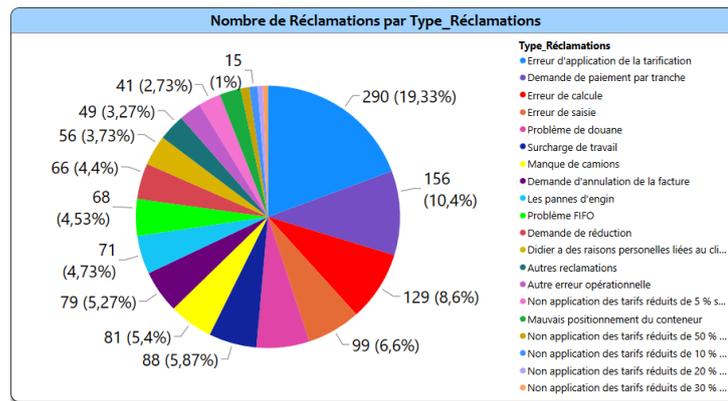


FIGURE 4.26 – Nombre de réclamations par type de réclamation

Après avoir filtré notre rapport, ce tableau présente le pourcentage des types de réclamation les plus fréquents au fil des années :

Année	Type de réclamation	Pourcentage (%)
2021	Erreur d'application de la tarification	21,68
2022	Erreur d'application de la tarification	27,73
2023	Erreur d'application de la tarification	14,86

TABLE 4.6 – Pourcentage des types de réclamation les plus fréquentes au fil des années

Nombre de réclamations par type de prestation

En utilisant le visuel« graphique en secteur », cette visualisation représente la répartition en pourcentage de chaque type de prestation au cours des trois dernières années.

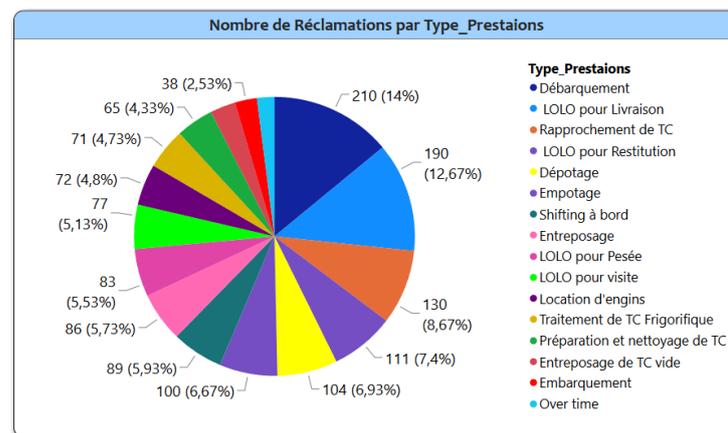


FIGURE 4.27 – Nombre de réclamations par type de prestation

Après avoir filtré notre rapport, ce tableau présente le pourcentage des types de prestation les plus fréquents au fil des années :

Année	Type de prestation	Pourcentage (%)
2021	Débarquement	15,14
2022	Débarquement	17,65
2023	Débarquement	13,62

TABLE 4.7 – Pourcentage des types de prestation les plus réclamés au fil des années

Nombre de réclamations par commission

En utilisant le visuel « le graphique en secteur », cette visualisation représente la répartition en pourcentage du nombre de réclamations par commission au cours des trois dernières années.

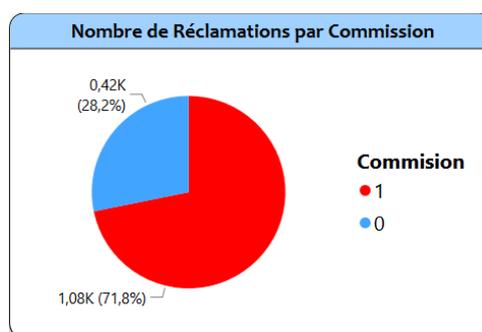


FIGURE 4.28 – Nombre de réclamations par commission

Après avoir filtré notre rapport, ce tableau présente le nombre de réclamations qui passent par la commission pour les trois dernières années :

Année	Commission (%)	Non Commission (%)
2021	97,76	2,24
2022	98,82	1,68
2023	44,02	55,98

TABLE 4.8 – Nombre de réclamations par commission au fil des années

Nombre de réclamations par décision

En utilisant le visuel du graphique en secteur, cette visualisation représente la répartition en pourcentage du nombre de réclamations par décision (favorable, défavorable) au cours des trois dernières années.

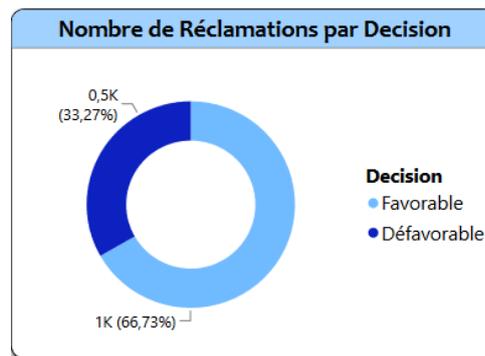


FIGURE 4.29 – Nombre de réclamations par décision

Après avoir filtré notre rapport, ce tableau présente le pourcentage des décisions les plus fréquentes au fil des années :

Année	Favorable (%)	Défavorable (%)
2021	71,03	28,97
2022	51,68	48,32
2023	68,5	31,5

TABLE 4.9 – Décisions des réclamations au fil des années

Tableau de bord :Analyse Détaillée des Réclamations

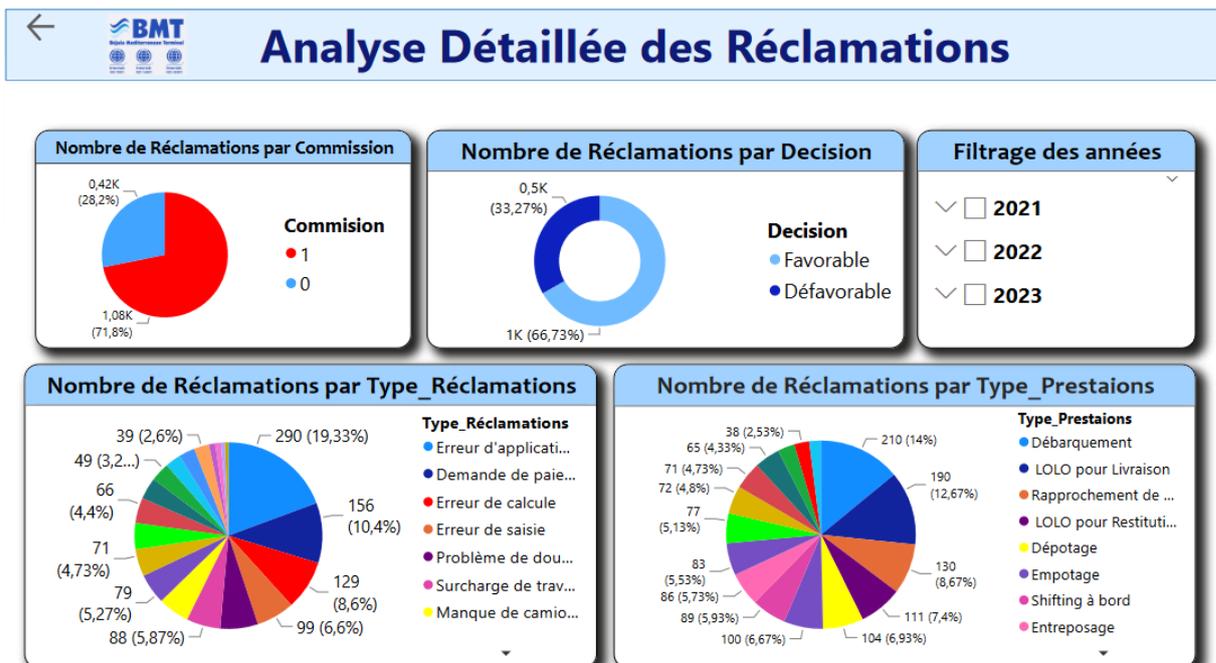


FIGURE 4.30 – Analyse Détaillée des Réclamations

Ce tableau de bord répond aux questions suivantes :

1. Quelle est la répartition des réclamations par type de réclamation au cours des trois dernières années ?
2. Comment se répartissent les réclamations par type de prestation sur les trois dernières années ?
3. Quel est le nombre de réclamations ayant nécessité une commission et celles traitées directement au cours des trois dernières années ?
4. Quelle est la répartition des réclamations par décision prise au cours des trois dernières années ?

Suivi de la Performance de Traitement des Réclamations

Évaluation de la performance du service Cette partie présente une analyse des principaux indicateurs de performance du service sur les trois dernières années. Les KPI tels que le Temps Moyen de Traitement Total (TMTT), le Taux de Résolution Mensuel (TRM), le Taux de Résolution Trimestriel (TRT), le maximum de délai de traitement (MAX(DT)), et le minimum de délai de traitement (MIN(DT)) sont examinés pour évaluer l'efficacité et l'amélioration du service.



FIGURE 4.31 – Évaluation de la performance du service

KPI	Années		
	2021	2022	2023
TMTT	230 jours	210 jours	125 jours
TRM	0,37 %	8,40 %	21,87%
TRT	11,78%	11,34%	41,40%
MAX(DT)	506 jours	414 jours	334 jours
MIN(DT)	9 jours	2 jours	0 jours

TABLE 4.10 – Évaluation de la performance du service

Nombre de délais de traitement par mois

Cette visualisation présente une vue générale du nombre de réclamations mensuelles.

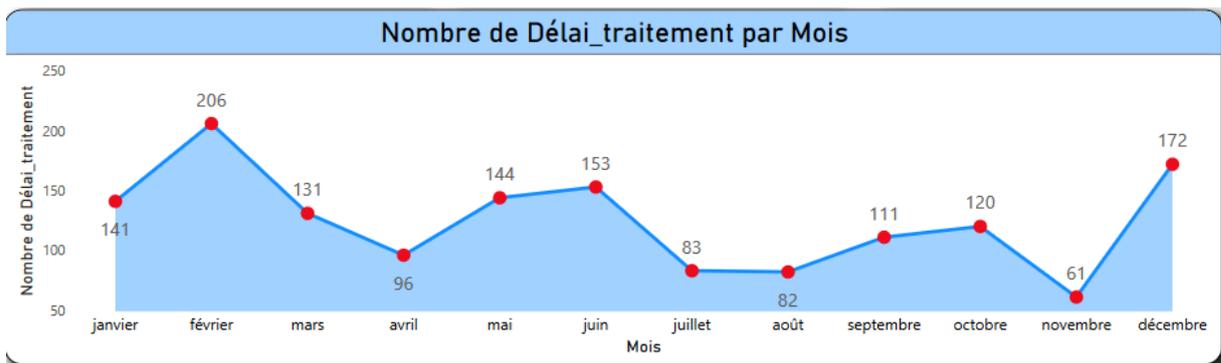


FIGURE 4.32 – Nombre de délais de traitement par mois

Après avoir filtré notre rapport, ce tableau montre le nombre de réclamations mensuelles au cours des trois dernières années, mettant en évidence les mois avec le plus grand nombre de réclamations.

Année	Mois	Nombre de Réclamations
2021	Février	142
2022	Juin	35
2023	Décembre	129

TABLE 4.11 – Réclamations mensuelles les plus élevées

Tableau de bord :Suivi de la Performance de Traitement des Réclamations

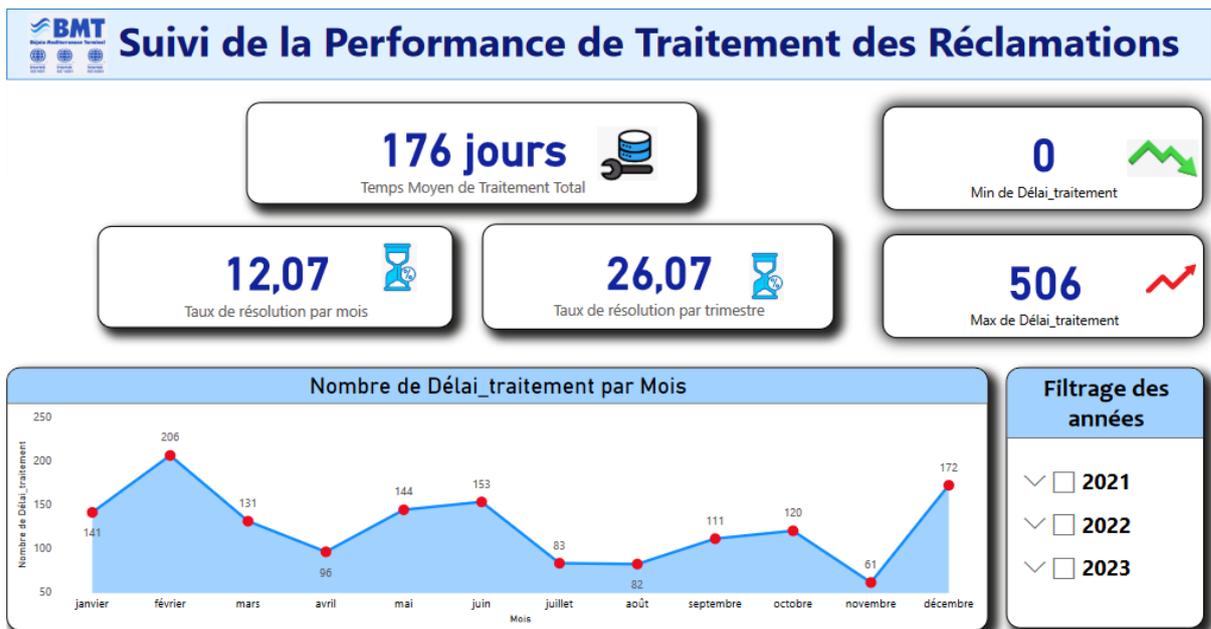


FIGURE 4.33 – Suivi de la Performance de Traitement des Réclamations

Ce tableau de bord répond aux questions suivantes :

1. Quel est le nombre de délais de traitement par mois au cours des trois dernières années ?
2. Quels sont les délais de traitement minimum et maximum des réclamations au cours des trois dernières années ?
3. Quels sont les taux de résolution mensuels et trimestriels des réclamations au cours des trois dernières années ?
4. Quel est le temps moyen de traitement total des réclamations au cours des trois dernières années ?

4.4 Implimentation de solution de prediction

Dans ce projet, nous avons développé un modèle de prédiction visant à déterminer si une réclamation nécessite une commission ou non. À partir d'un ensemble de données initial comprenant diverses caractéristiques des réclamations, nous avons suivi un processus structuré. Tout d'abord, nous avons chargé et prétraité les données, en éliminant les valeurs manquantes, les doublons, et en encodant les variables catégorielles pour les rendre exploitables par les algorithmes de machine learning. Ensuite, nous avons utilisé des techniques d'analyse de corrélation pour sélectionner les caractéristiques les plus pertinentes, minimisant ainsi la redondance et améliorant la pertinence des prédictions. À l'aide d'algorithmes tels que Random Forest, AdaBoost, et XGBoost, nous avons entraîné plusieurs modèles en optimisant leurs paramètres via une recherche aléatoire et une validation croisée. Les performances de chaque modèle ont été évaluées selon des métriques telles que l'accuracy, le F1-score, et d'autres mesures spécifiques à la régression pour garantir leur fiabilité et leur précision. Enfin, nous avons visualisé les résultats à travers des graphiques comme les courbes ROC pour évaluer et comparer la capacité prédictive de chaque modèle. Ce travail a abouti à la création d'un modèle robuste capable de prédire avec précision si une réclamation nécessite une commission, fournissant ainsi un outil décisionnel précieux pour l'organisation.

4.4.1 Étapes de Réalisation du Code

1. Importation des Bibliothèques :

- Importe toutes les bibliothèques nécessaires pour la manipulation des données, la visualisation et le traitement des modèles de machine learning. Utilisation de pandas pour la manipulation de données, numpy pour les calculs numériques, matplotlib et seaborn pour la visualisation des données, ainsi que plusieurs modules de scikit-learn pour la construction et l'évaluation des modèles. Importe également xgboost pour le classificateur XGBoost et imblearn pour le traitement du déséquilibre des classes avec SMOTE.

2. Chargement des Données :

- Charge les données à partir d'un fichier Excel en utilisant `pd.read_excel` de pandas, essentiel pour préparer les données brutes pour l'analyse.

3. Prétraitement des Données :

- Nettoie les données en supprimant les valeurs manquantes (`dropna()`) et les doublons (`drop_duplicates()`) pour assurer la qualité des données utilisées.

4. Encodage des Variables Catégorielles :

- Transforme les variables catégorielles en valeurs numériques à l'aide du `LabelEncoder` de scikit-learn, nécessaire car les algorithmes de machine learning ne traitent pas directement les variables catégorielles.

5. Sélection des Fonctionnalités Basée sur la Corrélation :

- Identifie et supprime les fonctionnalités fortement corrélées pour éviter la redondance et améliorer la performance du modèle.

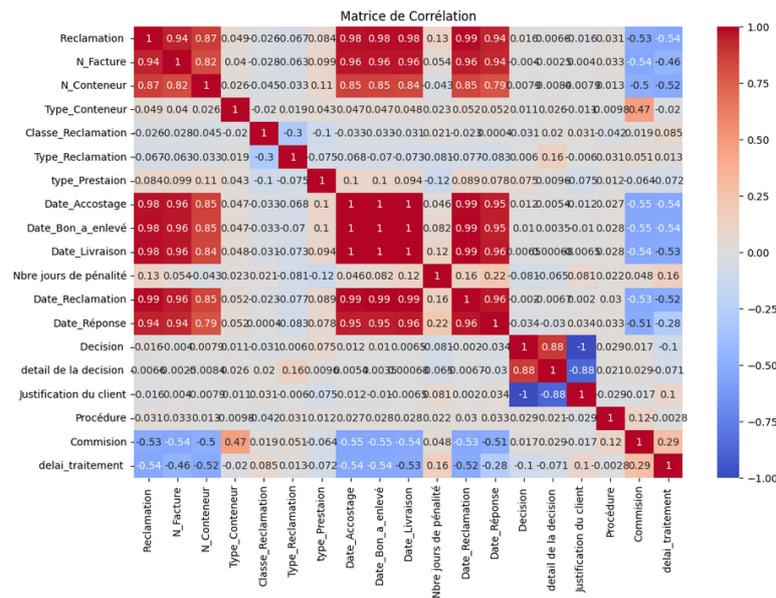


FIGURE 4.34 – Matrice de Corrélation

6. **Suppression des Colonnes Non Pertinentes :**

- Élimine les colonnes jugées non pertinentes (comme les identifiants ou descriptions textuelles) du jeu de données pour concentrer l’analyse sur les variables pertinentes.

7. **Sélection des Fonctionnalités à l’Aide d’ExtraTreesClassifier :**

- Utilise ExtraTreesClassifier pour sélectionner les variables les plus importantes, réduisant ainsi la dimensionnalité des données et améliorant la performance du modèle.

8. **Division des Données en Ensembles d’Entraînement et de Test :**

- Divise les données en ensembles d’entraînement (80%) et de test (20%) à l’aide de train_test_split, avec une stratification basée sur la variable cible pour garantir une répartition équilibrée des classes. Cela permet de former et de tester les modèles de manière équilibrée.

9. **Définition des Classificateurs :**

- Définit une liste de différents classificateurs de machine learning à utiliser pour l’analyse, incluant des modèles populaires comme Random Forest, AdaBoost, Bagging, Decision Tree, Logistic Regression, KNN, Naive Bayes, Gradient Boosting et XGBoost.

10. **Définition des Distributions de Paramètres pour la Recherche Aléatoire :**

- Définit les distributions de paramètres pour chaque classificateur afin d’optimiser les hyperparamètres avec RandomizedSearchCV, essentiel pour trouver la meilleure combinaison de paramètres pour chaque modèle.

11. **Entraînement et Évaluation des Classificateurs :**

- Entraîne les modèles sur les données d’entraînement et évalue leur performance sur les données de test, en utilisant diverses métriques telles que l’accuracy, la précision, le rappel, le score F1, le MSE, le MAE et le R².

12. Conversion des Résultats en DataFrame pour Affichage :

- Convertit les résultats des différents modèles en un DataFrame pour faciliter l’affichage et la comparaison des performances.

13. Création des Visualisations :

- Crée des graphiques pour visualiser la précision et le temps d’exécution de chaque modèle, utilisant seaborn pour des barplots montrant les performances des modèles.

14. Initialisation et Ajustement du Meilleur Modèle (XGBoost) :

- Ré-entraîne et ajuste le modèle XGBoost identifié comme le meilleur sur l’ensemble d’entraînement complet, puis fait des prédictions sur l’ensemble de test pour une évaluation finale.

15. Matrice de Confusion :

- Génère une matrice de confusion pour le modèle XGBoost afin de visualiser ses performances de classification en termes de vrais positifs, faux positifs, vrais négatifs et faux négatifs.

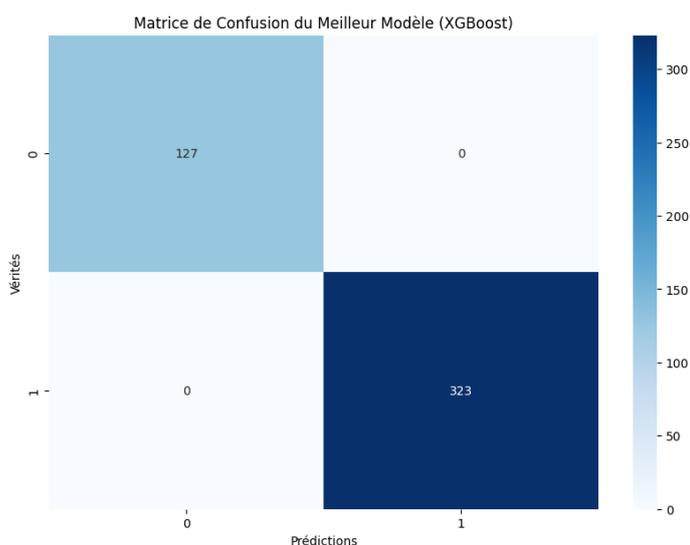


FIGURE 4.35 – Matrice de Confusion

16. Courbe ROC :

- Trace la courbe ROC et calcule l’Aire sous la Courbe (AUC) pour évaluer la performance du modèle XGBoost en termes de compromis entre le taux de vrais positifs et le taux de faux positifs.

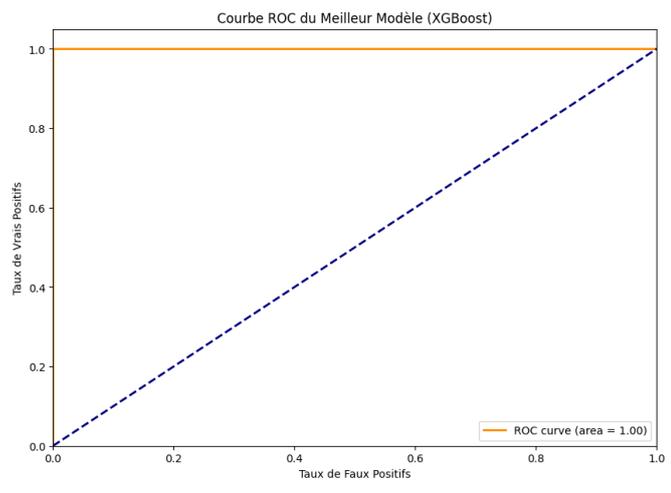


FIGURE 4.36 – Courbe de ROC

4.4.2 Résultats des classificateurs

Cette partie présente une comparaison des performances de différents classificateurs utilisés dans notre étude. Le tableau ci-dessous résume les résultats obtenus en termes de métriques d'évaluation standard :

Modèle	Accuracy	Precision	Recall	F1-score	MSE	MAE	R ²	Temps d'exécution (s)
Random Forest	0.9933	0.9938	0.9969	0.9954	0.0067	0.0067	0.9671	30.156
AdaBoost	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000	31.727
Bagging	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000	18.381
Decision Tree	0.9622	0.9811	0.9659	0.9735	0.0378	0.0378	0.8135	0.762
Logistic Regression	0.9778	0.9757	0.9938	0.9847	0.0222	0.0222	0.8903	7.299
KNN	0.7644	0.9303	0.7523	0.8209	0.2356	0.2356	- 0.1628	1.678
Naive Bayes	0.9244	0.9339	0.9628	0.9482	0.0756	0.0756	0.6270	0.006
Gradient Boosting	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000	16.872
XGBoost	0.9956	0.9938	1.0000	0.9969	0.0044	0.0044	0.9781	1.991

TABLE 4.12 – Résultats des classificateurs

Tandis que la figure suivante offre une visualisation comparative de ces performances.

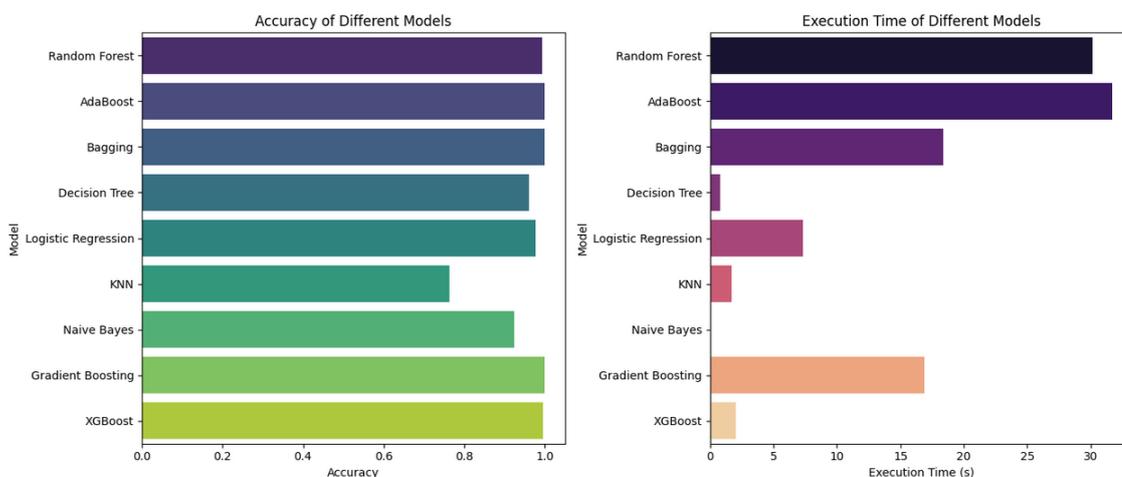


FIGURE 4.37 – Analyse comparative des performances des divers classificateurs

4.4.3 Analyse des Résultats et Sélection du Meilleur Modèle

Pour conclure sur la sélection du meilleur modèle basé sur nos résultats, il est essentiel de prendre en compte plusieurs aspects en fonction de nos besoins spécifiques, tels que l'accuracy, la précision, le rappel, le F1-score et le temps d'exécution. Voici quelques approches que nous avons envisagées :

Meilleure Accuracy et Performances Globales

- **Accuracy** : Le critère principal est souvent l'accuracy, qui mesure la proportion de prédictions correctes. Dans notre analyse, les modèles **AdaBoost**, **Bagging**, **Gradient Boosting**, et **XGBoost** ont tous atteint une accuracy de 100%. Cependant, il est crucial de considérer également d'autres aspects pour s'assurer que le modèle n'est pas simplement bien ajusté à nos données spécifiques mais qu'il est également généralisable.
- **Autres métriques** : Outre l'accuracy, il est pertinent d'examiner des métriques complémentaires telles que la précision, le rappel et le F1-score. Ces métriques offrent une vision plus équilibrée des performances prédictives du modèle. Par exemple, un modèle avec une haute précision mais un faible rappel pourrait ne pas identifier correctement toutes les instances positives, ce qui est crucial dans certains contextes.

Équilibre entre Performance et Temps d'Exécution

- **Temps d'exécution** : Si la rapidité d'exécution est un facteur clé, il est important de choisir un modèle qui offre un bon compromis entre performance (accuracy, précision, rappel) et temps d'exécution. Dans ce contexte, **XGBoost** semble être un candidat de choix avec une très bonne performance et un temps d'exécution raisonnable de 1.97 secondes.

Considération des Autres Métriques

- **Erreurs de prédiction** : Il est également important d'examiner les métriques d'erreur telles que la MSE (Mean Squared Error) et la MAE (Mean Absolute Error). Ces métriques fournissent des informations sur la robustesse du modèle et sur la gestion des erreurs de prédiction, ce qui peut être crucial pour certains cas d'utilisation.

Synthèse et Recommandations

En résumé, basé sur les résultats obtenus, **XGBoost** se distingue comme le meilleur choix pour plusieurs raisons :

- **Excellente Accuracy (99.56%)** : Cela indique que le modèle est très précis dans ses prédictions.
- **Précision Élevée (99.38%)** : Le modèle est capable de prédire correctement les instances positives.
- **Rappel Parfait (100%)** : Le modèle identifie toutes les instances positives sans omission.
- **Bon F1-score (99.69%)** : Cette métrique montre un bon équilibre entre précision et rappel.
- **Temps d'exécution raisonnable (1.97 secondes)** : Le modèle est suffisamment rapide pour être utilisé en pratique sans compromettre les performances.

4.5 Réalisation d'une application de Prédiction avec Streamlit

Après avoir développé et amélioré un modèle de machine learning à l'aide de Python, nous avons créé une application de prédiction avec Streamlit.

4.5.1 Définition de Streamlit

Streamlit est un framework open-source pour la création rapide d'applications web de data science et de machine learning en Python. Il permet aux développeurs de construire des interfaces utilisateur interactives sans nécessiter une expertise approfondie en développement web[55].

4.5.2 Avantages de Streamlit

- **Facilité d'utilisation** : Streamlit offre une prise en main rapide grâce à sa syntaxe simple, permettant aux développeurs de se concentrer sur la logique métier plutôt que sur les détails techniques.
- **Interactivité** : Il permet de créer des applications web où les utilisateurs peuvent ajuster les paramètres et visualiser instantanément les résultats, favorisant ainsi l'exploration interactive des données.

- **Intégration avec Python** : En étant intégré nativement avec Python, Streamlit facilite l'utilisation de bibliothèques populaires comme Pandas, Matplotlib et TensorFlow pour l'analyse de données et le machine learning.
- **Déploiement simplifié** : Streamlit propose des outils intégrés pour le déploiement rapide des applications web, réduisant ainsi la complexité liée à la mise en production[55].

4.5.3 Inconvénients de Streamlit

- **Personnalisation limitée** : Comparé à d'autres frameworks plus flexibles, Streamlit peut offrir moins de possibilités de personnalisation avancée de l'interface utilisateur.
- **Scalabilité** : Bien qu'efficace pour les projets de taille moyenne, Streamlit peut montrer des limites pour les applications nécessitant une grande échelle ou des capacités de traitement intensif[55].

4.5.4 Commandes essentielles pour l'installation et l'exécution

Pour commencer à utiliser Streamlit [55] :

1. **Installation** : `pip install streamlit`
2. **Vérification** : `streamlit --version`
3. **Exécution** : `streamlit run nom_de_votre_fichier.py`

4.5.5 Architecture de l'application

L'architecture de l'application de prédiction ci-dessous illustre la structure et le flux de données de notre solution. Cette représentation visuelle offre un aperçu clair de la manière dont les différentes composantes interagissent pour fournir des prédictions précises sur le traitement des réclamations

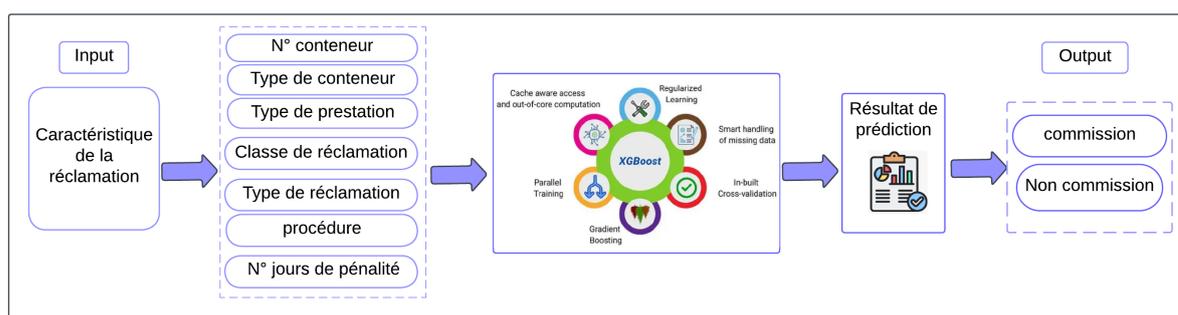


FIGURE 4.38 – Architecture de L'Application de prédiction

4.5.6 Explication de l'Application Web de Prédiction avec Streamlit

L'application web développée avec Streamlit a été conçue pour prédire si une réclamation doit passer en commission ou non, en se basant sur plusieurs caractéristiques spécifiques à

chaque réclamation. Cette application s'inscrit dans un projet plus large visant à améliorer la gestion des réclamations chez BMT grâce à des outils de machine learning .

Capture1 : Réclamation Passant en Commission

Cette capture d'écran montre un exemple où la réclamation saisie doit passer en commission



FIGURE 4.39 – Interface de Prédiction des Réclamations Passant en Commission

Capture2 : Réclamation Ne Passant Pas en Commission

Cette capture d'écran illustre un exemple où la réclamation saisie ne nécessite pas de passage en commission.



FIGURE 4.40 – Interface de Prédiction des Réclamations Ne Passant Pas par Commission

Capture3 : Documentation des Types de Données pour l'Application de Prédiction

Cette dernière capture d'écran présente la section de documentation et d'aide de l'application.



FIGURE 4.41 – Documentation des Types de Données pour l'Application de Prédiction

4.6 Conclusion

En conclusion, ce chapitre a démontré comment l'intégration harmonieuse de la Business Intelligence et du Machine Learning transforme la gestion des réclamations chez BMT-Spa. À travers l'élaboration et l'intégration d'indicateurs de performance stratégiques, ainsi que l'implémentation de solutions analytiques avancées avec Power BI et Python, nous avons créé un environnement décisionnel robuste et adaptatif. Cette approche permet à BMT-Spa non seulement d'améliorer sa capacité à traiter efficacement les réclamations, mais aussi d'anticiper proactivement les besoins futurs, renforçant ainsi sa compétitivité et sa satisfaction client. En continuant à investir dans ces technologies innovantes, BMT-Spa se positionne à la pointe de l'excellence opérationnelle et de la gestion proactive des relations clients.

5

Développement et Intégration du système centralisé intelligent

5.1 Introduction

Ce chapitre se concentre sur le développement et l'intégration d'un système centralisé intelligent visant à améliorer le traitement des réclamations au sein de BMT. Nous débutons par une exploration détaillée des applications web, en examinant leurs différents types, leurs mécanismes de fonctionnement, ainsi que leurs architectures. De plus, nous passons en revue les langages et environnements de développement cruciaux pour ce projet, tels que HTML, CSS, JavaScript, SQL, Python et PHP. Ensuite, nous présentons les outils essentiels utilisés, tels que XAMPP, MySQL, Apache, Visual Studio Code, Git et GitHub, et leur rôle dans la réalisation de notre application. Enfin, nous décrivons en détail les frameworks utilisés, notamment Bootstrap et Symfony, soulignant leur contribution à la création d'une application web robuste et responsive. Nous examinons également les interfaces de l'application destinées aux clients et aux administrateurs, illustrant ainsi l'étendue des fonctionnalités mises en place.

Sommaire

5.1	Introduction	87
5.2	Présentation des applications web	88
5.3	Langages et environnements de développement	88
5.4	Outils de développement	89
5.5	Frameworks utilisés	90
5.6	Sécurité dans le système de gestion des réclamations à BMT-Spa	90
5.7	Architecture de notre application	91
5.8	Présentation de l'application	92
5.9	Conclusion	96

5.2 Présentation des applications web

5.2.1 Définition

Application web : Une application web est un logiciel accessible et utilisable directement depuis un navigateur web, sans nécessiter de téléchargement ou d'installation préalable sur l'appareil de l'utilisateur. Elle est hébergée sur des serveurs distants, ce qui permet aux entreprises de gérer et de mettre à jour le logiciel de manière centralisée[14].

Application web dynamique : Une application web dynamique est un logiciel manipulable via un navigateur web, hébergé sur un serveur distant. Son contenu est généré en temps réel en réponse aux actions de l'utilisateur, souvent à partir de données stockées dans une base de données. L'application inclut généralement un panneau d'administration permettant aux administrateurs de gérer et de mettre à jour le contenu[45].

5.2.2 Fonctionnement d'une application web

Une application web dynamique fonctionne par la communication entre machines en réseau utilisant un langage commun. Les serveurs offrent des ressources, tandis que les clients les utilisent via le protocole HTTP (avec des commandes comme GET et POST). Les serveurs web transmettent directement les fichiers statiques aux navigateurs. Cependant, pour les pages dynamiques, un serveur d'application intervient, exécutant le code pour générer une page statique qui est ensuite renvoyée au serveur web et finalement transmise au navigateur sous forme de code HTML pur [45].

5.3 Langages et environnements de développement

5.3.1 HTML

Le langage HTML est utilisé pour structurer les pages web et les documents en utilisant des balises. Ces balises définissent la présentation et les liens entre les documents. Grâce au protocole http, les documents HTML peuvent être lus sur Internet à partir de différentes machines en utilisant des adresses URL uniques[62].

5.3.2 CSS

Les feuilles de styles(Cascading Style Sheets) est un langage de programmation qui contrôle l'apparence des pages Web en définissant des règles pour le positionnement, l'alignement, les polices, les couleurs, les bordures, etc. Son objectif principal est de séparer la structure HTML du contenu de sa présentation, améliorant ainsi la clarté du code[61].

5.3.3 JavaScript

Est un langage de script orienté objet largement employé en informatique, surtout pour animer et embellir les pages web (côté front-end). Son utilisation s'est étendue au développement côté serveur avec l'émergence de Node.js, permettant ainsi la création d'applications web complètes[?].

5.3.4 SQL

SQL (Structured Query Language) est un langage informatique standard utilisé pour gérer les bases de données relationnelles et manipuler les données. Il permet d'effectuer des opérations telles que l'interrogation, l'insertion, la mise à jour et la modification des données. Étant pris en charge par la plupart des bases de données relationnelles, SQL facilite la gestion des bases de données sur différentes plates-formes pour les administrateurs de bases de données[31].

5.3.5 Python

Est un langage de programmation polyvalent et interprété, offre une expérience de développement agile. Sans besoin de compilation, il peut être utilisé de manière interactive depuis une console, accélérant ainsi le processus de création. Grâce à sa disponibilité gratuite et à sa compatibilité avec divers systèmes d'exploitation, Python est largement adopté, notamment dans les domaines scientifiques et mathématiques. De plus, il est relativement facile à apprendre par rapport aux autres langages[32].

5.3.6 PHP

PHP, qui signifie "PHP : Hypertext Preprocessor," est un langage de script généraliste et open source, spécialement conçu pour le développement d'applications web. Il est rapide, flexible et pragmatique, et alimente tout, du simple blog aux sites web les plus populaires du monde.[46]

5.4 Outils de développement

5.4.1 XAMPP

XAMPP signifie Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) et Perl (P). C'est un ensemble de logiciels permettant de mettre en place facilement un serveur web et un serveur FTP. Il s'agit d'une distribution de logiciels libres facile à installer, offrant une bonne souplesse d'utilisation et permettant l'exploitation d'un serveur Apache, d'un SGBD MariaDB et d'un interpréteur PHP. XAMPP est également multiplateforme[3].

5.4.2 MYSQL

MySQL est un système de gestion de bases de données relationnelles (SGBDR) open source, développé, distribué et soutenu par Oracle Corporation. Utilisé par une vaste gamme d'applications, des solutions web aux systèmes d'entreprise, MySQL permet la gestion, le stockage et la manipulation des données structurées grâce à un modèle de base de données relationnel[44].

5.4.3 Serveur Apache

Apache HTTP Server est un serveur web open source de qualité commerciale, développé par la Apache Software Foundation, permettant de servir des pages web avec fiabilité et flexibilité, grâce à la contribution de bénévoles du monde entier[58].

5.4.4 Visual Studio Code

Visual Studio Code (VS Code) est un éditeur de code source gratuit, open source et multi-plateforme, conçu pour le développement moderne avec des fonctionnalités intégrées telles que Git, le débogage et les extensions[40].

5.4.5 Git

Git est un système de gestion de version décentralisée, créé par Linus Torvalds en 2005, permettant de suivre les modifications du code source et de collaborer efficacement sur des projets de toutes tailles[24].

5.4.6 GitHub

GitHub est un service basé sur le cloud qui permet d'héberger des dépôts Git. C'est le plus grand hébergeur de dépôts Git du marché. De plus, n'importe qui peut s'inscrire et héberger gratuitement un dépôt de code, ce qui rend GitHub particulièrement populaire auprès des projets open source [25].

5.5 Frameworks utilisés

5.5.1 Bootstrap

est un framework open source qui accélère le développement de sites web responsifs en étendant les capacités de CSS avec des fonctionnalités et composants intégrés. Il propose des styles prédéfinis pour des éléments tels que les boutons et les barres de navigation, facilitant la création de designs modernes adaptés à tous les écrans[10].

5.5.2 Symfony

est un framework PHP open-source pour le développement rapide d'applications web, offrant une structure modulaire et flexible. Il permet de construire des applications robustes et évolutives en suivant les meilleures pratiques, réputé pour sa stabilité, sa sécurité et sa capacité à être étendu grâce à des composants réutilisables.[56].

5.6 Sécurité dans le système de gestion des réclamations à BMT-Spa

- Utilisation d'un authenticator Symfony personnalisé pour l'authentification des utilisateurs.
- Stockage sécurisé des identifiants avec Symfony Security.
- Intégration de jetons CSRF dans les formulaires Symfony pour prévenir les attaques CSRF.

- Redirection après authentification basée sur le chemin de navigation précédent de l'utilisateur.
- Mesures avancées pour sécuriser les données sensibles des utilisateurs et des réclamations.
- Application des bonnes pratiques recommandées par Symfony pour un développement sécurisé.

5.6.1 Importance de la Sécurité

La sécurité est cruciale dans le projet de gestion des réclamations à BMT-Spa, garantissant non seulement la protection des données sensibles mais aussi la confiance des utilisateurs et la conformité aux normes de sécurité élevées.

5.7 Architecture de notre application

La figure ci-dessus illustre l'architecture de notre application de gestion des réclamations, organisée en trois phases principales : Interface Client, Serveur Web, et Interface Administrateur.

5.7.1 Phase 1 : Interface Client

Cette phase permet aux utilisateurs de soumettre des réclamations via une interface conviviale. Elle utilise HTML et CSS pour créer des formulaires et des pages web interactives. L'objectif est de garantir une expérience utilisateur fluide et intuitive, facilitant la collecte des informations nécessaires pour chaque réclamation.

5.7.2 Phase 2 : Serveur Web

Le serveur web est responsable de la gestion des requêtes des clients. Cette phase comprend des fichiers PHP qui traitent la logique métier de l'application et interagissent avec la base de données pour stocker et récupérer les informations sur les réclamations. Le serveur web joue un rôle crucial en assurant le bon déroulement des opérations et la gestion efficace des données soumises par les utilisateurs.

5.7.3 Phase 3 : Interface Administrateur

Cette phase est destinée aux administrateurs et inclut les outils nécessaires pour la gestion et le suivi des réclamations.

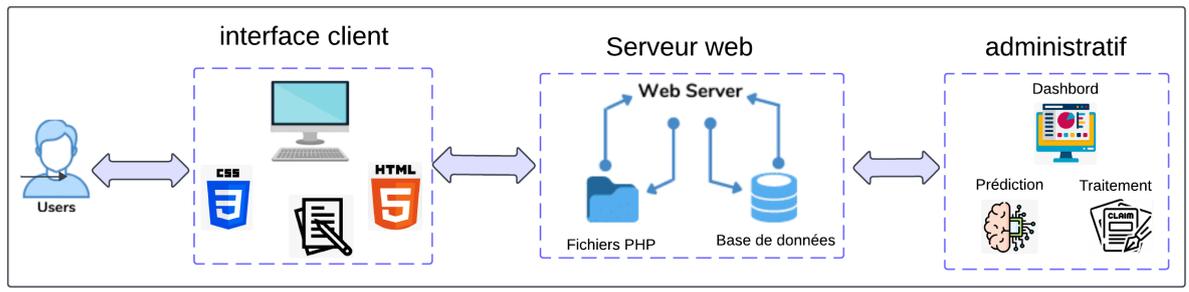


FIGURE 5.1 – Architecture de notre système centralisé intelligent

5.8 Présentation de l'application

5.8.1 Interfaces Communes

Authentification

L'interface d'authentification permet aux utilisateurs de se connecter à l'application en entrant leurs identifiants (son adresse et mot de passe). Cette interface assure la sécurité de l'accès et l'authentification des utilisateurs pour garantir que seules les personnes autorisées peuvent accéder aux fonctionnalités de l'application.

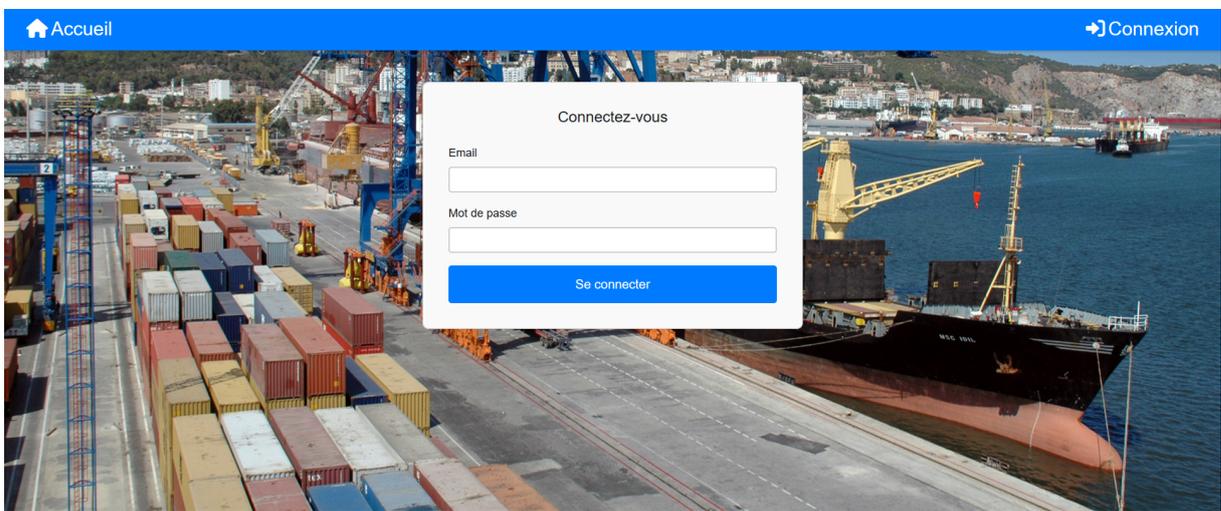


FIGURE 5.2 – Interface d'authentification

5.8.2 Interfaces Dédiées aux Clients

Liste de Ses Réclamations

Cette interface permet aux clients de consulter la liste de toutes leurs réclamations. Les clients peuvent imprimer cette liste, vérifier le statut de chaque réclamation, voir les réponses

fournies, et ont également la possibilité de modifier, ajouter ou supprimer des réclamations.

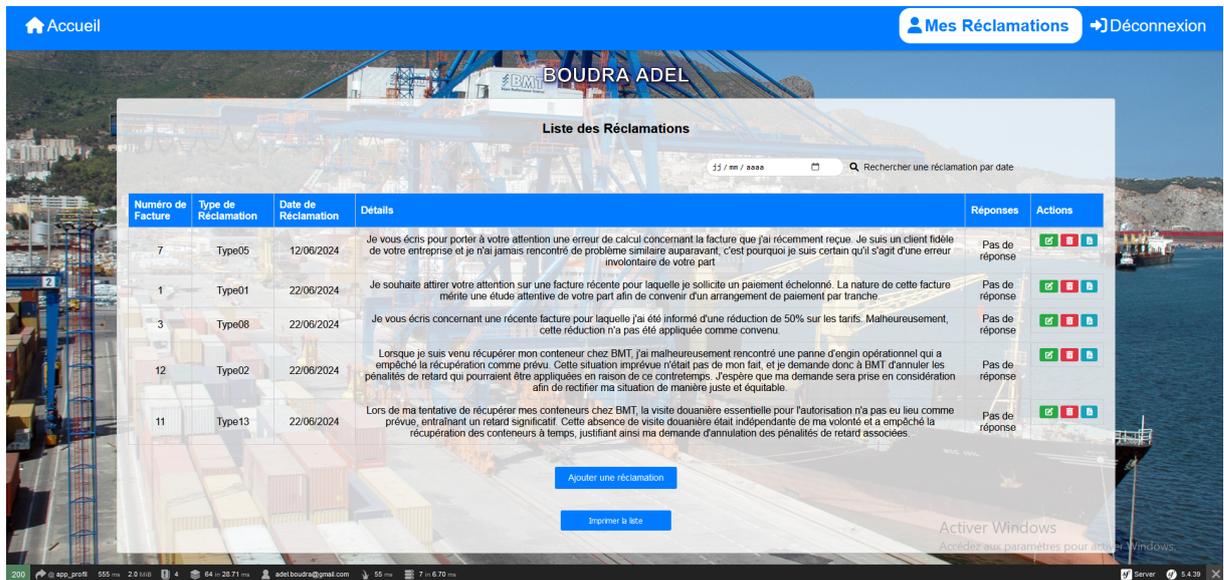


FIGURE 5.3 – Interface permet aux clients de consulter la liste de leurs réclamations

Formulaire de Saisie de Réclamation

e formulaire de saisie de réclamation permet aux clients de soumettre de nouvelles réclamations. Ils doivent entrer le numéro de la facture concernée, sélectionner le type de réclamation, détailler le problème rencontré, et joindre un fichier PDF contenant la demande formelle de réclamation.

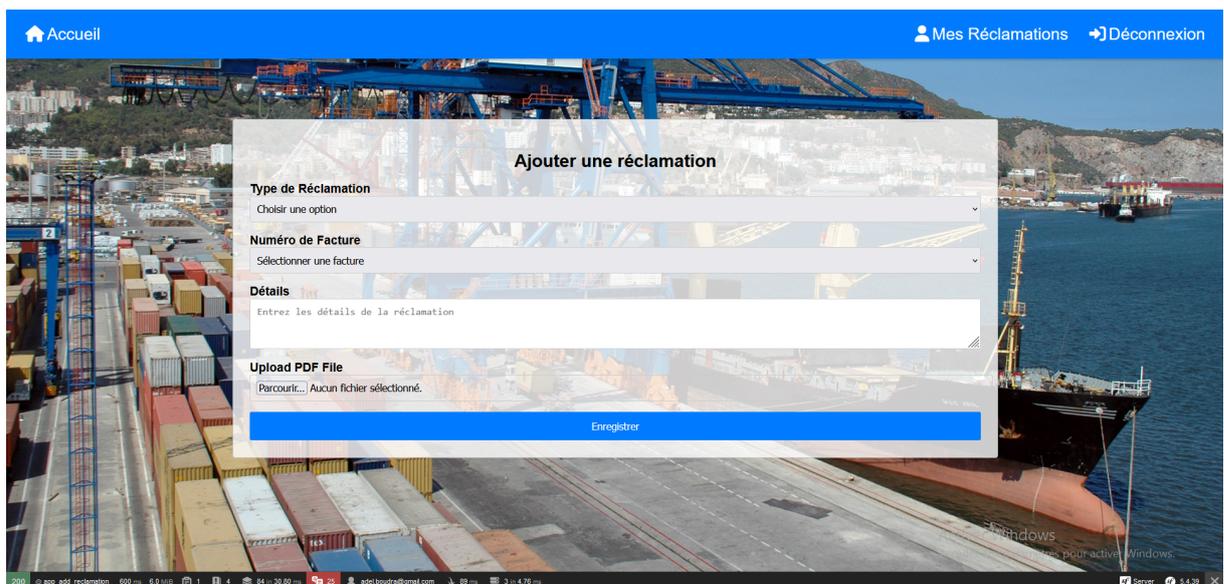


FIGURE 5.4 – Formulaire de Saisie de Réclamation

5.8.3 Interfaces Dédiées à l'Administration

Liste des Réclamations des Clients

L'interface de la liste des réclamations des clients permet aux administrateurs de consulter toutes les réclamations soumises par les clients. Ils peuvent effectuer des recherches par date, répondre aux réclamations, et imprimer la liste complète des réclamations.

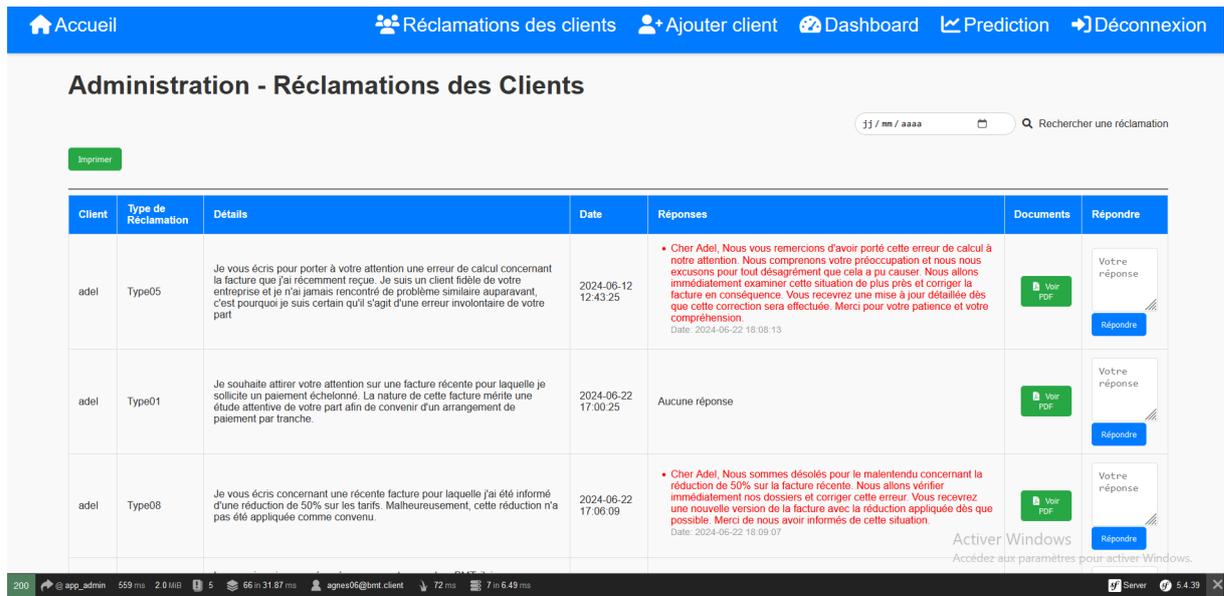


FIGURE 5.5 – Interface de la liste des réclamations des clients

Ajout de Nouveaux Clients

Cette interface permet aux administrateurs de créer de nouveaux comptes clients en entrant les informations nécessaires. Cela facilite la gestion des nouveaux clients et l'intégration de leurs réclamations dans le système.

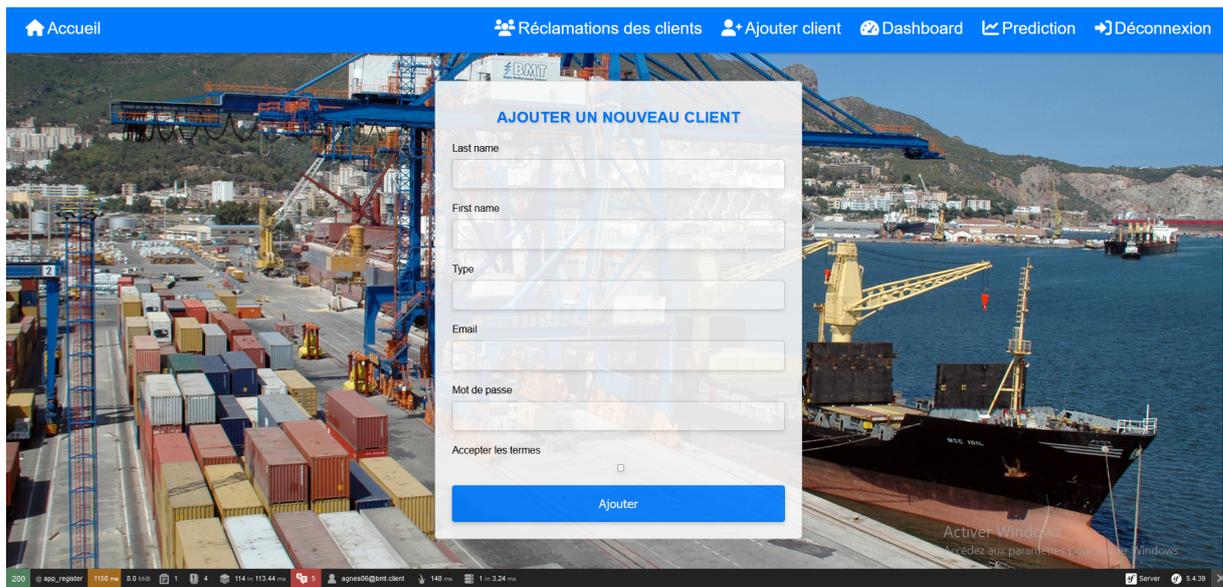


FIGURE 5.6 – Interface permet aux administrateurs de créer de nouveaux comptes

Dashboard Statistique

Le dashboard statistique permet aux administrateurs de visualiser et d'étudier les données des réclamations des trois dernières années. Cette interface offre des graphiques et des analyses pour aider à comprendre les tendances et améliorer le processus de gestion des réclamations.

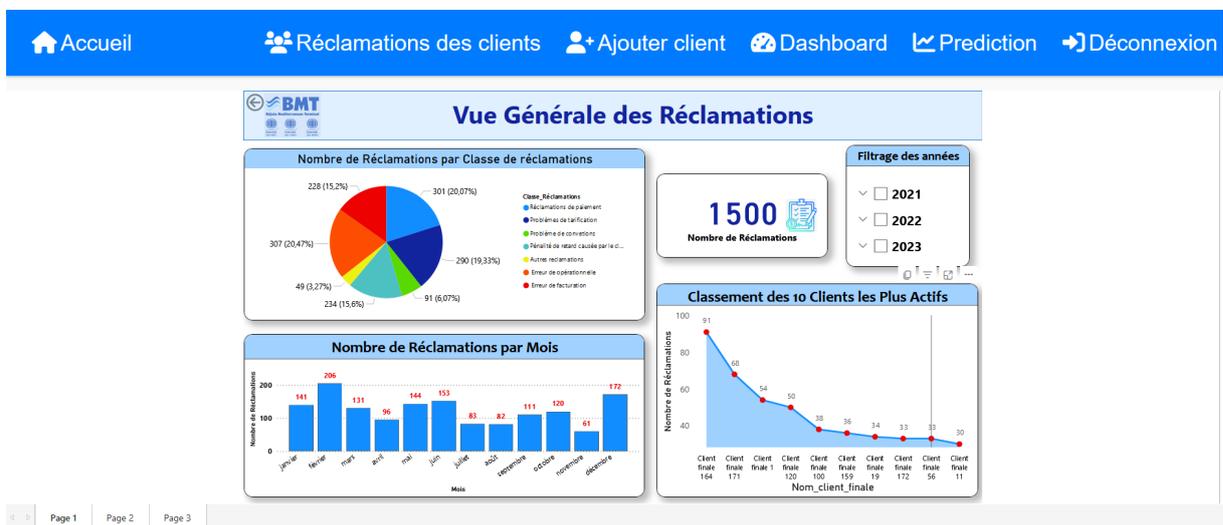


FIGURE 5.7 – Dashboard statistique

Prédiction des Réclamations Nécessitant une Commission

Cette interface permet aux administrateurs de prédire si une réclamation nécessite l'intervention d'une commission. En entrant les caractéristiques de la réclamation, le système utilise

un modèle de prédiction pour déterminer la nécessité d'une commission.

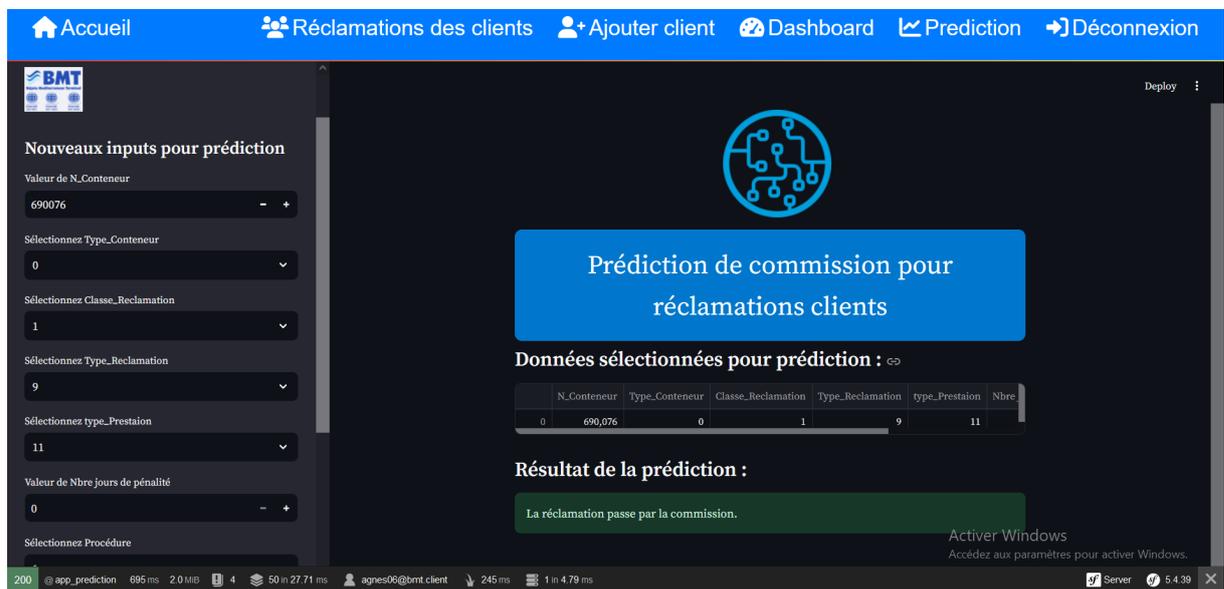


FIGURE 5.8 – Interface permet aux administrateurs de prédire

5.9 Conclusion

La mise en place de ce système centralisé et intelligent pour améliorer la gestion des réclamations chez BMT représente une avancée significative dans la modernisation des processus de l'entreprise. Grâce à une architecture web robuste et à l'adoption de technologies de pointe, nous avons créé une solution intégrée qui facilite non seulement la gestion des réclamations mais également leur analyse prédictive. Cette application démontre l'importance croissante d'une approche basée sur les données, où le rôle crucial du data scientist est de transformer les données en stratégies opérationnelles efficaces. En intégrant des techniques de Business Intelligence et de Machine Learning, nous avons non seulement augmenté l'efficacité opérationnelle, mais aussi renforcé la capacité de BMT à répondre proactivement aux besoins et aux attentes de ses clients.

6

Conclusion générale et travaux futures

Dans le paysage dynamique des affaires actuelles, l'exploitation stratégique des données est devenue impérative pour maintenir un avantage concurrentiel. La combinaison de la Business Intelligence (BI) et du Machine Learning (ML) offre un potentiel énorme pour transformer les données en informations exploitables, permettant ainsi aux entreprises de prendre des décisions éclairées et d'innover dans leurs opérations.

Notre étude s'est concentrée sur la conception et l'implémentation d'un système centralisé intelligent pour améliorer le traitement des réclamations chez BMT. En exploitant les principes de la BI et du ML, notre objectif était de développer une plateforme robuste qui centralise, analyse et gère efficacement les réclamations clients. Nous avons identifié plusieurs défis critiques dans le processus actuel de gestion des réclamations, notamment la dispersion des réclamations, les retards dans le traitement, et le manque d'outils d'aide à la décision.

À travers une méthodologie rigoureuse, nous avons réussi à relever ces défis en optimisant les processus de gestion des réclamations, en développant un modèle prédictif pour prioriser les interventions, et en mettant en place un système centralisé de gestion des réclamations. Notre solution intègre des outils de BI pour une visualisation claire des données et des algorithmes de ML pour automatiser le traitement des réclamations.

En implémentant cette solution, nous anticipons plusieurs avantages pour BMT, notamment une amélioration de l'efficacité opérationnelle, une augmentation de la satisfaction client, et une réduction des risques financiers. De plus, notre système fournira à BMT des outils puissants pour une analyse approfondie du traitement des réclamations, soutenant ainsi la prise de décisions stratégiques.

Perspectives Futures

Alors que notre solution offre des avantages significatifs à court terme, plusieurs perspectives futures peuvent être explorées pour renforcer davantage les capacités de gestion des réclamations chez BMT :

- L'intégration de technologies émergentes telles que l'intelligence artificielle conversationnelle (chatbots) pour une assistance client automatisée.
- L'exploration de techniques avancées de ML telles que le deep learning pour des prédictions encore plus précises.
- Le développement d'une interface utilisateur plus intuitive et conviviale pour une adoption maximale par les utilisateurs finaux.
- La mise en place d'un système de feedback continu pour une amélioration continue de la solution en fonction des besoins changeants de l'entreprise.

En conclusion, notre travail représente une étape significative dans l'évolution de la gestion des réclamations chez BMT. En combinant la puissance de la BI et du ML, nous avons créé une solution innovante qui positionnera BMT à l'avant-garde de l'excellence opérationnelle et de la satisfaction client. Ce projet démontre notre engagement à résoudre les défis organisationnels complexes tout en exploitant les dernières avancées technologiques pour transformer les processus métier.

Bibliographie

- [1] AKIF, L ET BERKATI, S. Conception et réalisation d'une application web pour la gestion de stock (cas d'étude : S.p.a -general emballage- taharacht akbou wilaya de bejaia). Mémoire de fin de cycle, Université Abderrahmane Mira, Béjaïa, 2022.
- [2] ANAYAT, N ET FAID, D. Mise en œuvre d'une solution de bi pour le suivi des résultats des étudiants cas d'étude : département informatique del'université de béjaia. Mémoire de master, Université Abderrahmane Mira, Béjaïa, 2020.
- [3] APACHE FRIENDS. Xampp. <https://www.apachefriends.org/index.html>. Consulté le : 2024-07-09.
- [4] BABAJIDE MUSTAPHA, I., AND SAEED, F. Bioactive molecule prediction using extreme gradient boosting. *Molecules* 21, 8 (2016), 983.
- [5] BELLAHMER, H. Implémentation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande de propriétés immobilières. Mémoire de fin d'études, Université Mouloud Mammeri, 2020.
- [6] BELLAKHDAR, A. gestion de patient de l'hopital ophtalmologie de l'amitié algero-cubain. Mémoire de fin d'études, Université Kasdi Merbah Ouargla, 2015.
- [7] BENALI, A. C., AND GHEBRIOUT, M. Vers une détection des sentiments et des phrases subjectives dans les réseaux sociaux. Mémoire de fin d'études, Université Saad DAHLAB - Blida 1, oct 2021.
- [8] BENSSAADA, A. Sélection des termes co-occurents avec entropie minimale pour la classification des textes. Mémoire de fin d'études, université de guelma, juin 2022.
- [9] BOOCH, G. *The unified modeling language user guide*. Pearson Education India, 2005.
- [10] BOOTSTRAP. Bootstrap : The world's most popular front-end open source toolkit. <https://getbootstrap.com>. Consulté le : 2024-06-24.
- [11] BRIGANTI, G. Intelligence artificielle : une introduction pour les cliniciens. *Revue des Maladies Respiratoires* 40, 4 (2023), 308–313.
- [12] BURQUIER, B. *Business intelligence avec SQL Server 2005 : mise en oeuvre d'un projet décisionnel*. Dunod, 2007.
- [13] CARLIER, A. *Business Intelligence et Management*. AFNOR, 2013.
- [14] CONALLEN, J. *Concevoir des applications Web avec UML*, vol. 1. Eyrolles Paris, France, 2000.
- [15] CONNOLLY, T. M., AND BEGG, C. E. *Database systems : a practical approach to design, implementation, and management*. Pearson Education, 2005.

- [16] DATA, L. B. Data mining : Définition et exemples. <https://www.lebigdata.fr/data-mining-definition-exemples>. Consulté le : 2024-07-09.
- [17] DAVID, P. Key performance indicators (kpi)-developing, implementing, and using winning kpis, 2010.
- [18] DEDIĆ, N., AND STANIER, C. Measuring the success of changes to business intelligence solutions to improve business intelligence reporting. *Journal of Management Analytics* 4, 2 (2017), 130–144.
- [19] DRESNER, H. *Profiles in Performance : Business Intelligence Journeys and Roadmap for Change*. John Wiley & Fils, Octobre 2009.
- [20] FAWCETT, T., AND PROVOST, F. *Data science pour l'entreprise : Principes fondamentaux pour développer son activité*. Editions Eyrolles, 2018.
- [21] FERNANDEZ, A. Les nouveaux tableaux de bord des décideurs. *Les éditions de l'organisation*, (2000).
- [22] FERNANDEZ, A. *Les nouveaux tableaux de bord des managers : le projet Business Intelligence clés en main*. Eyrolles, Mai 2013.
- [23] GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc, September 2019.
- [24] GIT SCM. Git. <https://git-scm.com/>. Consulté le : 2024-07-09.
- [25] GITHUB. Github : Where the world builds software. <https://github.com>. Consulté le : 2024-06-24.
- [26] GUEMIDI, A., AND KAZITANI, N. Détection et classification des émotions des personnes. Mémoire de fin d'études, Université Abdelhamid Ibn Badis de Mostaganem, 2020.
- [27] HADJI, S ET RAMDANI, N. Mise en place d'une solution de business intelligence cas : société de vente de matériels informatique. Mémoire de fin d'études, Université Mouloud Mammeri, 2020.
- [28] HAMDI, N. Conception et réalisation d'une solution business intelligence cas d'étude spa tchin-lait. Mémoire de fin d'études, Université Abderrahmane Mira, Béjaïa, 2020.
- [29] INMON, W. H. *Building the data warehouse*, 4 ed. John Wiley & Sons, octobre 2005.
- [30] INMON, W. H., TERDEMAN, R. H., AND IMHOFF, C. *Exploration Warehousing :Turning Business Information into Business Opportunity*, 1 ed. Juin 2000.
- [31] ISO/IEC. SQL - structured query language. <https://www.iso.org/standard/63555.html>. Consulté le : 2024-07-09.
- [32] JULIEN, D. *Designing Machine Learning Systems with Python*. Packt Publishing Ltd, 2016.
- [33] KESAVARAJ, G., AND SUKUMARAN, S. A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)* (2013), IEEE, pp. 1–7.
- [34] KIMBALL, R., AND ROSS, M. *The data warehouse toolkit : The definitive guide to dimensional modeling*, ed. wiley.
- [35] LU, H., CHENG, F., MA, X., AND HU, G. Short-term prediction of building energy consumption employing an improved extreme gradient boosting model : A case study of an intake tower. *Energy* 203 (2020), 117756.

- [36] MAHI, A. Détection de visage par l’algorithme de boosting. Mémoire de fin d’études, Université Aboubakr Belkaid–Tlemcen, 2018.
- [37] MEZGHANI, N. Module 7 arbres de décision, 2019. Tous droits réservés.
- [38] MEZILI, H. Vers une amélioration de la détection d’intrusion par les méthodes de sélection des fonctionnalités à l’aide des arbres de décision. Mémoire de fin d’études, Université Ibn Khaldoun-Tiaret, 2021.
- [39] MICROSOFT. Power bi. <https://www.microsoft.com/fr-fr/power-platform/products/power-bi/#Resources>. Consulté le 14 juin 2024.
- [40] MICROSOFT. Visual studio code. <https://code.visualstudio.com/docs>. Consulté le : 2024-06-02.
- [41] MICROSOFT CORPORATION. Power bi desktop : Qu’est-ce que power bi desktop? <https://learn.microsoft.com/fr-fr/power-bi/fundamentals/desktop-what-is-desktop>. Consulté le 14 juin 2024.
- [42] MOUSSOUS, H ET RARRBO, A. Projet bi : de l’extraction à l’analyse de données pour le ciblage des campagnes marketing. Mémoire de fin de cycle, Ecole Nationale Polytechnique, 2019.
- [43] MOZILLA DEVELOPER NETWORK. Javascript. <https://developer.mozilla.org/en-US/docs/Web/JavaScript>. Consulté le : 2024-07-09.
- [44] ORACLE CORPORATION. Mysql. <https://www.mysql.com/>. Consulté le : 2024-07-09.
- [45] OURDIA, K., AND DYHIA, S. Conception et réalisation d’une application web jee pour la gestion de centre médecine de travail. Mémoire de fin d’études, Université Mouloud Mammeri, 2019.
- [46] PHP DOCUMENTATION TEAM. Php : The basics - manual. <https://www.php.net/manual/fr/intro-what-is.php>. Consulté le 14 juin 2024.
- [47] POWER, D. J. *Decision support systems : concepts and resources for managers*, vol. 13. Quorum Books Westport, 2002.
- [48] PROVOST, F., AND FAWCETT, T. *Data Science for Business : What you need to know about data mining and data-analytic thinking*. O’Reilly Media, Inc., 2013.
- [49] RANIA, H. Amélioration des forêts aléatoires pour la classification des données médicales. Mémoire de fin d’études, Université Badji Mokhtar - Annaba, 2021.
- [50] RIVEST, S. *Investigation des modes d’intégration physique entre un serveur de base de données multidimensionnelle et un SIG*. PhD thesis, Faculté de foresterie de géomatique, Université Laval, Canada, 2000.
- [51] SAHLI, S. Prédiabète : Un système de détection et prédiction de diabète. Mémoire de fin de cycle, université de guelma, 2022.
- [52] SHERMAN, R. *Business intelligence guidebook : From data integration to analytics*. Newnes, 2014.
- [53] SIMON, A. R. *90 days to the data mart*. Wiley Publishing, 1998.
- [54] SINGH, G. A. P., AND GUPTA, P. Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications* 31, 10 (2019), 6863–6877.

- [55] STREAMLIT INC. Streamlit, official website. <https://streamlit.io>. Consulté le : 2024-06-24.
- [56] SYMFONY. Symfony - high performance php framework for web development. <https://symfony.com>. Consulté le : 2024-06-20.
- [57] TESTE, O. *Modélisation et manipulation des systèmes OLAP : de l'intégration des documents à l'usager*. PhD thesis, Université Paul Sabatier - Toulouse III, 2009.
- [58] THE APACHE SOFTWARE FOUNDATION. About apache http server project. https://httpd.apache.org/ABOUT_APACHE.html. Consulté le : 2024-06-02.
- [59] VAIDYA, J., KANTARCIOĞLU, M., AND CLIFTON, C. Privacy-preserving naive bayes classification. *The VLDB Journal* 17, 4 (2008), 879–898.
- [60] VASILEV, I., SLATER, D., SPACAGNA, G., ROELANTS, P., AND ZOCCA, V. *Python Deep Learning : Exploring deep learning techniques and neural network architectures with Pytorch, Keras, and TensorFlow*. Packt Publishing Ltd, 2019.
- [61] W3C. CSS - cascading style sheets. <https://www.w3.org/Style/CSS/>. Consulté le : 2024-07-09.
- [62] W3C. HTML - hypertext markup language. <https://www.w3.org/TR/html52/>. Consulté le : 2024-07-09.
- [63] YAHIAOUI, L., AND KESSIRA, D. Conception et réalisation d'une application web de commerce électronique et de business intelligence. Mémoire de fin d'études, Université de Béjaïa, 2022.

Annexes

Annexe 1 : Type de prestations

Embarquements : désigne le processus de chargement des conteneurs depuis le quai à bord d'un navire.

Débarquements : signifie le processus de déchargement des conteneurs d'un navire à quai.

Shifting à bord : désigne le déplacement ou le changement de position d'un conteneur à l'intérieur d'un navire pendant les opérations de manutention.

Rapprochement de TC : Rapprochement des conteneurs vides depuis la zone extra-portuaire d'Ireyaheh vers le terminal à conteneurs.

Entreposage de TC vide : fait référence au stockage de conteneurs vides dans les zones extra-portuaires.

Over time : une majoration tarifaire de 50% sera appliquée sur les prestations de manutention les jours fériés.

LOLO pour Visite : Les tarifs sont appliqués pour effectuer le contrôle des marchandises transportées dans les conteneurs.

LOLO pour Pesée : La prestation de pesée implique de contrôler le poids d'un conteneur afin de s'assurer qu'il respecte le poids autorisé.

LOLO pour Livraison : Il s'agit du service de transport des conteneurs du parc jusqu'au destinataire final.

LOLO pour Restitution : le renvoie les conteneurs vides aux sociétés de transport pour rééquilibrer l'offre et la demande.

Location d'engins : permet de disposer des équipements nécessaires pour effectuer ces opérations de manière efficace et sécurisée

Empotage et dépotage : ce sont les opérations de chargement et de déchargement des marchandises à l'intérieur d'un conteneur.

Préparation et nettoyage de TC : comporte le nettoyage des déchets, ainsi que l'inspection de l'état du conteneur pour qu'il soit prêt à être transporté à nouveau.

Traitement des conteneurs frigorifiques : comprend le branchement, le gardiennage, la surveillance et le maintien de la température.

Entreposage : signifie le stockage des TCs dans le terminal à conteneurs.

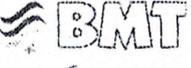
LOLO empotage : concerne le chargement des conteneurs sur les navires à l'aide de grues.

LOLO dépotage : concerne le déchargement des conteneurs des navires à l'aide de grues.

Annexe 2 : Facture

الجمهورية الجزائرية الديمقراطية الشعبية

MINISTRE DES TRANSPORTS
GROUPE SERVICES PORTUAIRES
SERPORT SPA
BEJAIA MEDITERRANEAN TERMINAL SPA



وزارة النقل
مجمع الخدمات المينائية
SERPORT SPA
ش.د.أ. بجاية مدينيرانيين ترمينال

SPA au Capital de 500 000 000 Da
N° du registre de commerce: 048 0184739 1°
N° Immatriculation Fiscale: 000406018473045
N° d'Article: 06010108961
N° RZA BADR: 00 380 357 000 083 430 086

NIS: 0 004 0601 09003 73

Facture N°: 29851/23

Code Client: 472001 Client Final:

Adresse: I
N° Immatriculation Fiscale: 1206010229930
N° du Registre de commerce: 10182543
N° d'Article fiscal: 122.03.12

Nom Bateau: N° Voyage: YB337A/23*
ATB: 28/09/2023 N° Gros: 1330/23
ATD: 01/10/2023 N° Article: 122
N° Dossier: 0

Code	Désignation	Quantité	Nbr J	Prbx U	Montant
100.	PENALITE DE RETARD D'ENLEVEMENT	3	3	5 000,00	45 000,00
Total Hors Taxe					45 000,00
Total TVA (19%)					8 550,00
Total TTC					53 550,00

Facture arrêtée à la somme de :
Cinquante-trois mille cinq cent cinquante dinar .

Faita à Bejaia le 23/10/2023
Cachet Facturation



intertek
Total Quality. Assured.

ISO 9001 : 2015 - ISO 14001 : 2015 - ISO 45001 : 2018

IMP_PR1/01

Le client doit s'adresser au service commercial de BMT au plus tard le 5eme jour ouvrable après la fin des opérations pour récupérer la facture.
Le délai de paiement est fixé à (15) jours à compter de la date de réception de la facture.
Le délai de la contestation des factures est fixé à (08) jours à compter de la date de réception des factures.

Nouveau Quay - port de Bejaia BP 640 RP Bejaia 06000
Tél: +213 (0) 982 401 952 - Fax: +213 (0) 34 10 36 47
Site web: www.bejaiamed.com

1/1

Annexe 3 : Demande de reclamation

Monsieur le Directeur général DE BMT

Bejaia le 30/10/2023

'DI/2023.

Objet : Demande d'avoir.

Copie :

Concernant : V/ Factures /23 du 23/10/23

du 28/09/23 gros /23 art 122

Monsieur le Directeur général,

Nous venons par la présente vous demander de bien vouloir revoir votre facture en raison d'une erreur portant sur le calcul des pénalités de retard de notre clients en effet nous avons obtenu le bon à enlever le 16/10/23 et la livraison a été faite le jour même et non le 12/10/23 comme c'est mentionné sur votre fiche de suivi.

Nous vous demandons de bien vouloir vérifier et rectifier cette erreur.

Veillez agréer, Monsieur le directeur, nos salutations distinguées.

Copie : Bon à enlever

Annexe 4 : Fiche de suivi



DEPARTEMENT ETUDES STATISTIQUES

Réf :

/2023

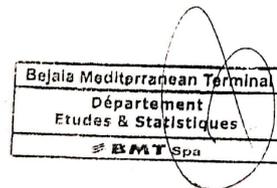
date: 22/10/2023

FICHE DE SUIVI PENALITES DE RETARD D'ENLEVEMENT

BILL LADING :	Client :	Client final :	GROS N° :	Article N° :
				122

Item	N° TC	date accostage ATB	Date Bon à enlevé	Date CMR	Date Livraison	Nbre jours de pénalité	Montant de pénalité
1	TGBU2493220	28/09/2023	12/10/2023	12/10/2023	16/10/2023	3	15 000,00
2	MSMU2322949	28/09/2023	12/10/2023	12/10/2023	16/10/2023	3	15 000,00
3	MEDU5168849	28/09/2023	12/10/2023	12/10/2023	16/10/2023	3	15 000,00
TOTAL							45 000,00

Arrêté la présente fiche à la somme de :
quarante-cinq mille Dinars



SPA au Capital Social de 500 000 000 DA - RC N° 05/00-0184730804 - NIF N° 000406018473048
Siège Social : Nouveau Quai - Port de Béjaia - BP 549 BP Béjaia 06000 - ALGERIE
☎ : 213 (0)24 103 346 / 213 982 481 950 🌐 : 213 982 400 799 Site web : bejaiaamdt.com

Annexe 5 : Définition dans le contexte maritime

transitaire :

Un transitaire est un professionnel ou une entreprise spécialisée dans l'organisation et la gestion du transport de marchandises. Il facilite le déplacement des biens d'un point d'origine à une destination, en prenant en charge tous les aspects logistiques et administratifs du transport.

Consignataire :

Un consignataire est un agent ou une entreprise qui représente les intérêts d'un navire et de son propriétaire (l'armateur) dans un port. Il joue un rôle essentiel dans la coordination et la gestion des opérations portuaires. Ce métier se trouve dans le domaine maritime.

Compagnie maritime :

Une compagnie maritime est une entreprise spécialisée dans le transport maritime de marchandises et/ou de passagers. Elle opère en possédant ses propres navires (armateur) ou en louant ceux d'autres (affréteur). Ces compagnies jouent un rôle crucial dans le commerce international en assurant la mobilité des biens et des personnes. Parmi les compagnies maritimes présentes à Béjaïa Mediterranean Terminal (BMT), on trouve :



Résumés

Bejaia Mediterranean Terminal (BMT) est une entreprise dynamique opérant dans le secteur maritime, où une gestion efficace des réclamations des clients est essentielle pour garantir leur satisfaction et optimiser l'efficacité opérationnelle. Face à la complexité croissante des demandes et à l'importance des retours clients, BMT a besoin de solutions innovantes pour relever ces défis.

Ce mémoire se concentre sur la conception et la mise en œuvre de tableaux de bord utilisant des indicateurs de performance pour soutenir la prise de décision chez BMT, en employant des techniques avancées de la science des données. L'objectif principal est de développer une plateforme robuste qui centralise, analyse et gère efficacement les réclamations, améliorant ainsi la réactivité et la qualité du service client. Cette plateforme intègre des outils de Business Intelligence (BI) pour visualiser clairement les données de réclamation et des algorithmes de Machine Learning (ML) pour automatiser le traitement et la classification des réclamations, assurant une réponse rapide et appropriée à chaque situation.

La réalisation de ce projet s'est articulée autour de plusieurs étapes clés. Tout d'abord, nous avons procédé au prétraitement et à l'analyse des données de réclamation, en définissant précisément les besoins et en structurant les données pour assurer leur qualité et leur pertinence. Ensuite, nous avons développé des tableaux de bord intégrant des indicateurs de performance pour analyser les données sous divers angles, fournissant ainsi des informations pertinentes aux décideurs. Nous avons ensuite mis en place des algorithmes de Machine Learning pour prédire les réclamations nécessitant une intervention spécifique, améliorant ainsi la rapidité et l'efficacité du traitement.

Les résultats obtenus montrent une amélioration significative de l'efficacité opérationnelle, une augmentation de la satisfaction client et une réduction des risques financiers. Notre système offre à BMT des outils puissants pour une analyse approfondie du traitement des réclamations, soutenant ainsi la prise de décisions stratégiques et l'amélioration continue des processus internes.

À travers ce projet, notre objectif est de transformer la gestion des réclamations chez BMT en utilisant les technologies avancées de la science des données, assurant la fiabilité et la pertinence des informations présentées aux décideurs.

Mots clés : tableaux de bord, Business Intelligence, Machine Learning, traitement des réclamations, web, visualisation des données, collecte, analyse de données, décision.

Abstract

Bejaia Mediterranean Terminal (BMT) is a dynamic company operating in the maritime sector, where effective management of customer complaints is crucial for ensuring customer satisfaction and optimizing operational efficiency. Faced with the increasing complexity of requests and the importance of customer feedback, BMT requires innovative solutions to address these challenges.

This thesis focuses on the design and implementation of dashboards using performance indicators to support decision-making at BMT, leveraging advanced data science techniques. The primary objective is to develop a robust platform that centralizes, analyzes, and efficiently manages complaints, thereby enhancing responsiveness and service quality. This platform integrates Business Intelligence (BI) tools to clearly visualize complaint data and Machine Learning (ML) algorithms to automate the processing and classification of complaints, ensuring a prompt and appropriate response to each situation.

The execution of this project was structured around several key steps. We first carried out the preprocessing and analysis of complaint data, precisely defining the requirements and structuring the data to ensure its quality and relevance. Next, we developed dashboards incorporating performance indicators to analyze the data from multiple perspectives, thereby providing relevant insights to decision-makers. We then implemented Machine Learning algorithms to predict complaints requiring specific interventions, thereby improving the speed and efficiency of processing.

The results demonstrate a significant improvement in operational efficiency, an increase in customer satisfaction, and a reduction in financial risks. Our system provides BMT with powerful tools for in-depth analysis of complaint processing, supporting strategic decision-making and continuous improvement of internal processes.

Through this project, our goal is to transform complaint management at BMT by utilizing advanced data science technologies, ensuring the reliability and relevance of the information presented to decision-makers.

Keywords : dashboards, Business Intelligence, Machine Learning, complaint processing, web, data visualization, data collection, data analysis, decision-making.