



Rapport de stage d'initiation sous thème

Modélisation des Paramètres de Production des Phosphates en vue de Prédire les Prix et les Quantités de Production, avec une Interface Flask utilisant le Machine Learning

Encadrant de l'entreprise :

Mr touijer mohamed

Réaliser par :

NESSRINE LAFHAL

KAOUTAR SARSARI

Khadija ABDEL WASSAA

Anass EL-HABTI

Encadré à l'école par :

Ghazdali Abdelghani

Encadré à l'école par :

OMAR ALBANNAY

Encadré à l'école par :

El Habib Nfaoui

Encadré de l'école :

Tawfik rachad

Spécialité:

**Informatique et ingénierie
des données
(IID)**

Spécialité:

GENIE INFORMATIQUE (GI)

Spécialité:

**Master Big Data
Analytics & Smart Systems
(BDSaS)**

Spécialité :

**Data science
&
IoT**

Dédicaces :

Nous souhaitons exprimer nos sentiments les plus profonds à travers ces mots envers tous ceux qui nous ont aidés dans l'élaboration de notre projet ainsi que de ce rapport.

Nous dédions ce modeste travail à nos parents qui nous ont soutenus pendant la période d'études, nos professeurs de l'école nationale des sciences appliquées Khouribga (ENSAk) , de l'École Nationale Supérieure d'Informatique et d'Analyse des Systèmes (ENSIAS) et la faculté des sciences à Fès .

Ainsi que notre cher superviseur Mr. [Mr touijer mohamed](#) et toute l'équipe du service réseau informatique qui nous ont accompagnés et soutenus dans ce projet et tout au long de la période du stage

Remerciement :

Nous tenons à vous exprimer toute notre gratitude pour votre accompagnement tout au long de ce projet de fin d'études. Votre soutien et vos conseils avisés nous ont permis de mener à bien ce travail et d'acquérir de nouvelles compétences.

Nous tenons également à remercier Mr touijer mohamed et toute l'équipe du service réseau informatique pour nous avoir offert l'opportunité de réaliser ce projet de stage sous leur supervision.

Enfin, nous aimerions adresser nos remerciements à nos professeurs, nos familles, nos amis et nos proches pour leur soutien indéfectible durant cette période de travail acharné. Merci infiniment !

Abstract

Ce rapport présente une étude approfondie sur la prédiction du chiffre d'affaires de l'Office Chérifien des Phosphates (OCP) en utilisant des techniques de Machine Learning. L'objectif principal de cette recherche est de développer un modèle de prédiction précis pour estimer le chiffre d'affaires futur de l'OCP, basé sur des données historiques pertinentes. Pour faciliter la compréhension et l'interaction avec les résultats de la prédiction, une interface de visualisation interactive a été créée à l'aide du framework Flask.

Le rapport aborde tout d'abord le contexte et les enjeux de la prédiction du chiffre d'affaires dans le secteur de l'industrie minière et chimique. Il expose ensuite la méthodologie de collecte et de préparation des données, ainsi que le processus de sélection des caractéristiques pertinentes pour la modélisation.

Différents algorithmes de Machine Learning ont été évalués pour la construction du modèle de prédiction, en mettant l'accent sur leur performance en termes de précision et de stabilité. Le modèle final a été sélectionné en fonction de ses performances supérieures sur un ensemble de données de validation.

En parallèle à la modélisation, une interface web a été développée à l'aide de Flask, permettant aux utilisateurs d'interagir avec le modèle de prédiction. Cette interface offre une visualisation intuitive des résultats de prédiction, ainsi que des fonctionnalités d'exploration et de personnalisation des entrées.

Les résultats de ce projet démontrent l'efficacité du modèle de prédiction développé pour estimer le chiffre d'affaires futur de l'OCP. De plus, l'interface de visualisation en temps réel offre aux utilisateurs un outil convivial pour explorer et comprendre les prédictions générées par le modèle.

En conclusion, cette étude combine la puissance du Machine Learning avec la convivialité d'une interface web interactive pour réaliser des prédictions de chiffre d'affaires de manière efficace et accessible. Ces résultats peuvent avoir des implications significatives pour la prise de décision stratégique au sein de l'OCP et d'autres entreprises similaires.

Résumée

Ce rapport présente une étude exhaustive axée sur la prédiction du chiffre d'affaires de l'Office Chérifien des Phosphates (OCP) en utilisant des techniques avancées de Machine Learning. L'objectif central de cette recherche consiste à développer un modèle de prédiction précis et fiable pour anticiper le chiffre d'affaires futur de l'OCP, en se basant sur des données historiques pertinentes.

La méthodologie adoptée pour atteindre cet objectif repose sur plusieurs étapes clés. Tout d'abord, une exploration minutieuse des données est entreprise, suivie d'une préparation soigneuse pour assurer la qualité et la pertinence des informations utilisées dans le processus de modélisation. Ensuite, une sélection rigoureuse des caractéristiques est effectuée pour identifier les variables ayant le plus grand impact sur les prédictions.

Différents algorithmes de Machine Learning sont évalués et comparés pour développer le modèle de prédiction optimal. L'algorithme retenu est choisi en fonction de ses performances de prédiction et de sa capacité à gérer les variations temporelles des données.

Un aspect essentiel de ce projet est la création d'une interface de visualisation interactive utilisant le framework Flask. Cette interface permet aux utilisateurs d'interagir facilement avec le modèle de prédiction et d'explorer les résultats générés. Les graphiques dynamiques et les fonctionnalités de personnalisation offrent une expérience utilisateur enrichissante pour comprendre les prédictions et prendre des décisions éclairées.

Les résultats obtenus révèlent l'efficacité du modèle de prédiction développé, avec une précision prometteuse dans l'estimation du chiffre d'affaires futur de l'OCP. L'interface de visualisation ajoute une couche d'accessibilité en permettant aux utilisateurs de prendre en main les prédictions de manière intuitive.

En somme, cette étude démontre comment l'utilisation judicieuse du Machine Learning et la conception d'une interface interactive peuvent contribuer à améliorer la prédiction du chiffre d'affaires dans un contexte complexe comme celui de l'OCP. Les résultats obtenus ouvrent des perspectives intéressantes pour l'optimisation des décisions stratégiques et la planification financière au sein de l'entreprise.

Liste d'abréviations

OCP : Office Chérifien des Phosphates

ML : Machine Learning

CA : Chiffre d'Affaires

GUI : Interface Graphique Utilisateur

API : Interface de Programmation Applicative

UI : Interface Utilisateur

LSTM : Long Short-Term Memory (un type de réseau de neurones récurrents)

RF : Random Forest (un algorithme d'apprentissage automatique)

SVM : Support Vector Machine (un algorithme de classification)

FLASK : Framework Web Flask (pour le développement d'applications web en Python)

RMSE : Root Mean Squared Error (une métrique d'évaluation des modèles de prédiction)

MAE : Mean Absolute Error (une autre métrique d'évaluation des modèles)

API : Application Programming Interface (Interface de Programmation)

GUI : Graphical User Interface (Interface Graphique Utilisateur)

CSV : Comma-Separated Values (un format de fichier pour les données tabulaires)

HTML : HyperText Markup Language (langage de balisage pour la création de pages web)

CSS : Cascading Style Sheets (feuilles de style pour la mise en forme des pages web)

JSON : JavaScript Object Notation (un format de données)

RNN : Recurrent Neural Network (un type de réseau de neurones récurrents)

API : Application Programming Interface (Interface de Programmation Applicative)

ML : Méthodes de Machine Learning

STAT : Statistiques et Analyse

EDA : Analyse Exploratoire des Données

PCA : Analyse en Composantes Principales

ROI : Retour sur Investissement

BD : Base de données

Liste figure

Contents

INTRODUCTION GENERALE OCP	12
INTRODUCTION	12
But de stage	12
1 Présentation globale de l'Office Chérifien des Phosphates	13
1.1 INTRODUCTION:	13
1.2 L'historique :	13
1.3 Chronologie de l'entreprise:	14
1.4 Fiche technique :	18
1.5 Statut juridique :	18
1.6 ORGANIGRAMME JURIDIQUE:	19
1.7 Filiales du groupe OCP	19
1.8 Organigramme de l'OCP :	20
1.9 Les Activités du groupe OCP :	20
1.9.1 La prospection	20
1.9.2 La production	21
1.9.3 La Valorisation	22
1.9.4 La commercialisation	22
1.10 Le Phosphate et la phosphatogenèse :	22
1.10.1 Définition :	22
1.10.2 L'origine du phosphate :	22
1.10.3 Différents types de phosphates :	23
1.10.4 Théories de la phosphatogenèse :	24
1.11 L'emplacement géographique d'OCP	25
1.11.1 Présence au Maroc	25
1.11.2 Présence à l'international	26
1.11.3 Importance de l'Industrie Minière des Phosphates	27
1.11.4 Enjeux et Défis de l'Industrie Minière des Phosphates :	28
1.12 Conclusion	29
2 PhosphoForecast: Integrated Prediction of Price, Quantity in Phosphate Industry	30
2.1 Introduction	30
2.2 Problématique	31
2.2.1 Facteur 1 : Problèmes géopolitiques et dynamique du marché	31
2.3 Facteur 2 : perturbations de la chaîne d'approvisionnement et défis de production	31
2.3.1 Facteur 3 : Modification des pratiques agricoles et préoccupations environnementales	

2.4	Conséquences de prédictions inexactes	32
2.4.1	Allocation inefficace des ressources :	32
2.4.2	Pertes financières :	32
2.4.3	Impacts environnementaux :	33
2.5	Environnement de travail	33
2.5.1	Environnement de travail Python :	34
2.5.2	Google Colab :	34
2.6	Jeux de données de l'étude	35
2.6.1	Description du jeu de données de prédiction des prix (Price prediction):	35
2.6.2	Description du jeu de données de prédiction quantité de phosphate nutritif :	37
2.7	Conclusion :	39
3	Préparation de données	40
3.1	Lecture le fichier CSV :	40
3.2	Nettoyage des données :	40
3.2.1	Identification des données manquantes	41
3.2.2	Nettoyage des données textuelles :	41
3.3	Transformation :	43
3.4	Exploration de données	46
Conclusion		53
3.5	Séparation des données	53
3.6	conclusion	54
4	Random Forest Regression model :	55
4.1	Définition :	55
4.2	Fonctionnement de Random Forest de régression	56
4.3	Implémentation en python	58
4.4	Caractéristiques de La régression avec Random Forest	58
4.5	Avantages Clés :	59
4.6	Applications Pratiques :	59
4.7	Conclusion :	60
5	Evaluation de model de Régression avec Random Forest	60
5.1	Introduction :	61
5.2	Utilisation de la Fonction `score()` :	61
5.3	Interprétation du R^2 :	61
5.4	Exemple de Code :	62
5.5	Conclusion :	62
5.6	Conclusion	63

6	Mise en production :	Erreur ! Signet non défini.
6.1	Environnement de travail :	64
6.2	Partie Frontend (Interface Utilisateur) :	66
6.2.1	Conception de l'Interface Utilisateur :	66
6.2.2	Implémentation de l'Interface Utilisateur :	70
6.3	le résultat finale :	74
6.3.1	1- home page :	74
6.3.2	2- page :	76
6.3.3	3- les Formulaires:	77
6.3.4	4-page des résultats est téléchargement de rapport :	79
6.3.5	Conclusion :	80
6.4	Partie backend :	81
6.4.1	Configuration du Projet Flask :	81
6.4.2	Création et Gestion des Sessions:	84
6.4.3	Stockage et Récupération des Données Utilisateur:	84
6.4.4	Génération de Rapports avec ReportLab :	88
6.4.5	Conclusion :	91

INTRODUCTION GENERALE

INTRODUCTION

L'industrie des phosphates, en tant que socle fondamental de l'agriculture et de multiples domaines associés, repose sur une dynamique complexe en constante évolution. Les fluctuations des prix mondiaux du phosphate exercent un impact majeur sur l'économie mondiale, la disponibilité d'engrais et la sécurité alimentaire. Dans cet environnement exigeant, l'aptitude à anticiper les variations futures des prix et de la production de phosphate revêt une importance stratégique cruciale.

C'est dans ce contexte que s'inscrit le projet "PhosPredictAI". Cette initiative révolutionnaire ambitionne de créer une application reposant sur des modèles d'apprentissage automatique avancés, en vue de prédire avec une précision accrue les prix mondiaux du phosphate et d'envisager les fluctuations à venir. En parallèle, cette application novatrice intègre la capacité de projeter les volumes de production de phosphate pour divers pays en prenant en compte un éventail de facteurs, englobant la production d'autres éléments, l'évolution démographique ainsi que les habitudes de consommation.

But de stage

Le but de ce stage au sein de l'Office Chérifien des Phosphates (OCP) est de concevoir, développer et mettre en œuvre l'application "PhosPredictAI", en combinant l'expertise de l'industrie des phosphates avec les capacités de l'apprentissage automatique pour prédire avec précision les prix mondiaux du phosphate et anticiper les variations futures, tout en estimant les quantités de production de phosphate dans divers pays. Les objectifs clés incluent la collecte et l'analyse de données, la conception et le développement de modèles de prédiction avancés, l'intégration de facteurs multiples pour renforcer la précision, l'évaluation des performances et la documentation complète du processus.

Ce stage offre une opportunité de développement personnel en acquérant des compétences en apprentissage automatique, en développement d'applications et en analyse de données, tout en contribuant de manière significative à l'industrie des phosphates par le biais d'une solution innovante pour la prise de décision éclairée.

1 Présentation globale de l'Office Chérifien des Phosphates

1.1 INTRODUCTION:

Ce chapitre débute en introduisant l'entreprise OCP en tant qu'acteur global, en mettant en évidence sa configuration et sa structure organisationnelle. Par la suite, il se focalise sur un aspect spécifique, à savoir l'importance de la Division Minière de Khouribga au sein du groupe OCP. Ultérieurement, une analyse consacrée aux défis et aux besoins de l'ensemble d'OCP.

1.2 L'historique :

Établi en 1920 sous l'appellation "Office Chérifien des Phosphates" (OCP), l'entreprise a initié ses opérations en inaugurant sa première mine à Khouribga. Depuis lors, l'envergure des activités s'est étendue à travers cinq continents, couvrant l'intégralité de la chaîne de valeur des phosphates. Cette couverture englobe diverses phases, allant de l'extraction minière à la transformation industrielle, tout en englobant des initiatives visant l'éducation et le développement des communautés. Cette trajectoire historique illustre l'évolution constante d'OCP au fil des décennies, marquée par une expansion stratégique de nos activités et un engagement soutenu dans divers secteurs au sein de l'industrie des phosphates.

1.3 Chronologie de l'entreprise:

Voici une chronologie de l'entreprise OCP, le premier exportateur mondial de phosphate et de ses dérivés :

1920	<ul style="list-style-type: none">▪ Création, le 7 août, de l'Office Chérifien des Phosphates (OCP).
1921	<ul style="list-style-type: none">▪ Début de l'exploitation en souterrain sur le gisement des Oulad Abdoun, le 1er mars.▪ Première exportation de phosphate à partir du port de Casablanca le 23 juillet.
1931	<ul style="list-style-type: none">▪ Début de l'extraction en souterrain à Youssoufia (ex-Louis Gentil).
1951	<ul style="list-style-type: none">▪ Démarrage de l'extraction à ciel ouvert à Sidi-Daoui (Khouribga). Début du▪ développement des installations de séchage et de calcination à Khouribga.
1954	<ul style="list-style-type: none">▪ Démarrage des premières installations de séchage à Youssoufia.
1961	<ul style="list-style-type: none">▪ Mise en service de la première laverie à Khouribga
1965	<ul style="list-style-type: none">▪ Création de la société Maroc Chimie.▪ Début de la valorisation des phosphates avec le démarrage des installations de l'usine de Maroc Chimie, à Safi.
1974	<ul style="list-style-type: none">▪ Lancement des travaux pour la réalisation du centre minier de Benguérir, en mai. Le Groupe OCP prend le contrôle de la Société Marocaine des Fertilisants (Fertima), créée en 1972.▪ Naissance de l'Institut de Promotion Socio-Éducative (IPSE), en août.
1975	<ul style="list-style-type: none">▪ Création du Groupe OCP (décision de création en juillet 1974 et mise en place en janvier 1975).▪ Intégration des industries chimiques aux structures internes du Groupe OCP, en janvier.

	<ul style="list-style-type: none"> ▪ Création du Centre d'Études et de Recherches des Phosphates Minéraux (Cerphos), en octobre.
1976	<ul style="list-style-type: none"> ▪ Démarrage à Safi de Maroc Phosphore I et Maroc Chimie, en novembre. ▪ Prise de participation dans le capital de Phosboucraa à hauteur de 65%
1979	<ul style="list-style-type: none"> ▪ Transfert des bureaux de la Direction Générale au nouveau siège à Casablanca.
1980	<ul style="list-style-type: none"> ▪ Démarrage du site minier de Benguéir
1981	<ul style="list-style-type: none"> ▪ Démarrage à Jorf Lasfar de MP II. Le Groupe OCP entre dans le capital de Prayon (Belgique).
1982	<ul style="list-style-type: none"> ▪ Début des travaux de construction du complexe chimique MP III-IV à Jorf Lasfar.
1986	<ul style="list-style-type: none"> ▪ Démarrage des différentes lignes d'acide sulfurique et d'acide phosphorique de Maroc Phosphore III-IV.
1987	<ul style="list-style-type: none"> ▪ Démarrage des lignes d'engrais de Maroc Phosphore III-IV (octobre-décembre).
1996	<ul style="list-style-type: none"> ▪ Création de la société Euro-Maroc Phosphore (Emaphos). ▪ Lancement des travaux de construction de l'usine d'acide phosphorique purifié de Emaphos, à Jorf Lasfar (janvier-février). ▪ Regroupement des activités des deux sociétés Maroc Chimie et Maroc Phosphore au sein de Maroc Phosphore (janvier). ▪ Introduction de Fertima à la Bourse des Valeurs de Casablanca (30% du capital) dans le cadre du projet de privatisation de la société (octobre).
1997	<ul style="list-style-type: none"> ▪ Création de la société Indo-Maroc Phosphore (Imacid) en joint-venture avec le Groupe Birla.
1998	<ul style="list-style-type: none"> ▪ Démarrage de la production d'acide phosphorique purifié (Emaphos, Jorf Lasfar), le 31 janvier.
1999	<ul style="list-style-type: none"> ▪ Démarrage de la production d'acide phosphorique de l'usine d'Imacid à Jorf Lasfar, le 1er novembre.
2002	<ul style="list-style-type: none"> ▪ Prise de participation dans la société indienne PPL en Joint-venture avec le Groupe Zuari.

	<ul style="list-style-type: none"> ▪ Acquisition de l'intégralité du capital de Phosboucraa.
2004	<ul style="list-style-type: none"> ▪ Création de la Société « Pakistan Maroc Phosphore SA. » en Joint-venture entre le Groupe OCP et Fauji Fertilizer Bin Qasim Limited (Pakistan).
2005	<ul style="list-style-type: none"> ▪ Démarrage de l'usine de Lavage/Flottation à Youssoufia.
2008	<ul style="list-style-type: none"> ▪ Transformation de l'Office Chérifien de Phosphates en société anonyme OCP SA le 28 février. ▪ Démarrage de Pakistan Maroc Phosphore à Jorf Lasfar (PMP).
2009	<ul style="list-style-type: none"> ▪ Démarrage de Bunge Maroc Phosphore à Jorf Lasfar. ▪ Augmentation de capital d'OCP SA, le 13 janvier 2009, d'un montant de 5 milliards de dirhams entièrement réservée à la Banque Centrale Populaire (BCP).
2010	<ul style="list-style-type: none"> ▪ Création d'une JV avec Jacobs engineering (JESA). ▪ Ouverture d'un bureau de représentation au Brésil. ▪ Ouverture d'un bureau de représentation en Argentine. ▪ Création d'un fonds d'investissement Agricole. ▪ Création d'une société de valorisation du patrimoine immobilier (SADV).
2011	<ul style="list-style-type: none"> ▪ Lancement des travaux de l'usine de dessalement à Jorf Lasfar. ▪ Emission d'un emprunt obligataire inaugural sur le marché local.
2012	<ul style="list-style-type: none"> ▪ Fusion entre OCP S.A. et Maroc Phosphore S.A.
2013	<ul style="list-style-type: none"> ▪ Lancement de la construction de quatre nouvelles usines de granulation, ayant chacune une capacité de production annuelle de 1 million de tonnes, à Jorf Lasfar. ▪ Création de DuPont OCP Operations Consulting S.A. (DuPont OCP) ▪ Bunge cède sa participation dans Bunge Maroc Phosphore à OCP SA qui devient actionnaire à 100%. L'entreprise change de dénomination pour devenir Jorf Fertiliser Company V (JFC V).
2014	<ul style="list-style-type: none"> ▪ Emission obligataire inaugurale sur le marché international. ▪ Achèvement de plusieurs projets de développement : slurry pipeline de transport de pulpe de roche de Khouribga à Jorf Lasfar, station terminale du slurry pipeline, nouvelle unité de production d'acide phosphorique et centre de compétences industrielles basé à Jorf Lasfar. ▪ Ouverture d'un bureau de représentation à Singapour.

2015	<ul style="list-style-type: none"> ▪ Ouverture d'un bureau de représentation en Côte d'Ivoire. ▪ Acquisition, à travers la filiale OCP International cooperative, d'environ 10 % des actions de Fertilizantes Heringer S.A. ▪ Lancement d'OCP Africa.
2016	<ul style="list-style-type: none"> ▪ Ouverture d'OCP Research à Washington. ▪ Signature d'un partenariat avec l'Etat éthiopien visant la construction d'un complexe industriel d'engrais en Ethiopie. ▪ Signature avec le Groupe Dangote au Nigeria d'un MOU pour étudier l'opportunité de création d'une JV amenée à porter des actifs industriels à Jorf et un complexe chimique au Nigeria. ▪ Emission obligataire perpétuelle subordonnée sur le marché national.
2017	<ul style="list-style-type: none"> ▪ Création d'une JV entre la Société et IBM visant à accélérer la transformation digitale des entreprises marocaines et africaines. ▪ Signature d'un partenariat de long terme avec ADNOC pour l'approvisionnement en soufre. ▪ Mise en service du Downstream et réalisation de la première exportation de la roche séchée au niveau du Downstream à Jorf Lasfar.
2018	<ul style="list-style-type: none"> ▪ JFC4, le dernier projet de la première vague du programme d'investissement est désormais opérationnel. ▪ Création de deux écoles de codage : 1337 à Khouribga, et Youcode à Youssoufia. ▪ Emission obligataire perpétuelle subordonnée sur le marché national ▪ Acquisition de 20% des actions de Fertinagro Biotech SL. ▪ Le Groupe reçoit la médaille d'or de l'International Fertilizer Association (IFA)
2019	<ul style="list-style-type: none"> ▪ Création d'une JV entre la Société et Fertinagro Biotech SL : OCP Fertinagro Advanced Solutions pour la production d'engrais enrichis. ▪ Nomination de Mostafa TERRAB en tant que président de l'IFA ▪ Le Groupe rejoint le WBCSD (Conseil mondial des entreprises pour le développement durable).
2020	<ul style="list-style-type: none"> ▪ Acquisition de 30% restants du capital social de BSFT, OCP International Cooperative devient actionnaire unique de BSFT.
2021	<ul style="list-style-type: none"> ▪ Création d'une JV entre la Société et Hubei Forbon Technology. ▪ OCP SA a procédé avec succès à une nouvelle émission obligataire internationale pour 1,5 milliard USD ainsi qu'au rachat partiel des obligations 2024-2025

2022	<ul style="list-style-type: none"> ▪ Cession de 50 % de JFC III à Koch Ag & Energy Solutions finalisée le 30 juin 2022.
-------------	--

L'OCP a connu une croissance constante au cours de son histoire, passant d'une petite entreprise minière à un groupe industriel mondial. L'entreprise est aujourd'hui un acteur majeur de l'industrie des phosphates et joue un rôle important dans l'approvisionnement mondial en engrais.

1.4 Fiche technique :

Le tableau suivant regroupe les différentes informations de la fiche signalétique de l'OCP :

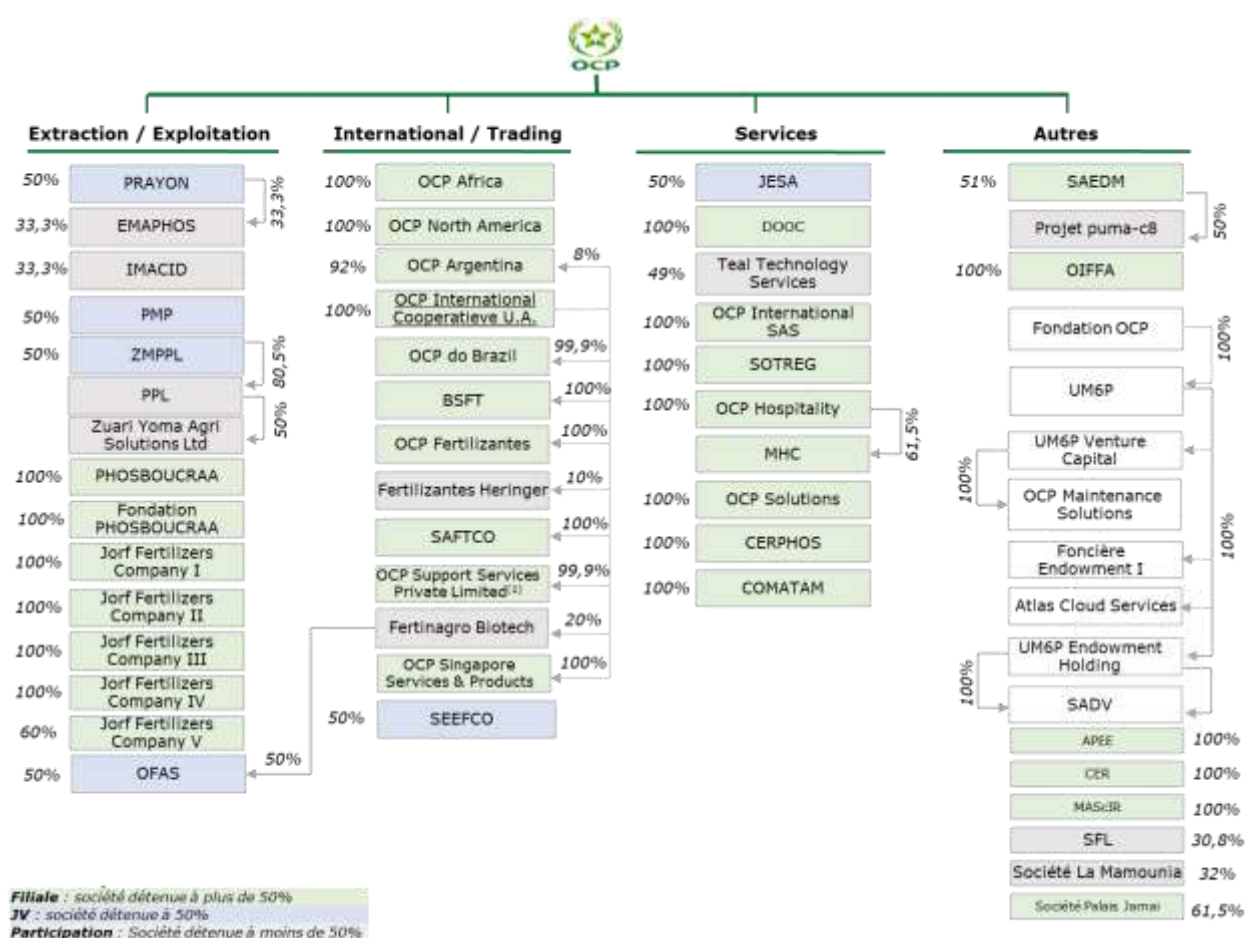
Raison sociale	Office Chérifien des Phosphates OCP
Numéro du registre de commerce	Casablanca 40.327
Date de création	Dahir du 07/08/1920
Mise en place de la structure d'un groupe	Juillet 1975
Siège social	Angle Route d'El Jadida et BD de la Grande Ceinture, B.P 5196 Casa Maarif, Casablanca Tél. : 02-23-(00-01-20)-25
Directeur Général	M. Mostafa TERRAB
Produits commercialisés	Phosphate, Acide Phosphorique, et Engrais.
Effectif	22 677, dont 725 ingénieurs

1.5 Statut juridique :

Le groupe OCP est une entreprise semi-publique sous contrôle de l'état, mais elle agit avec le même dynamisme et la même souplesse qu'une grande entreprise privée servant à l'état marocain tous les droits de recherche et d'exploitation des phosphates, gérée par un directeur est contrôlée par un conseil d'administration présidé par le Premier ministre.

La gestion financière est séparée de celle de l'état. Le groupe OCP est inscrit au registre de commerce et soumis sous le plan fiscal à la même obligation que n'importe quelle entreprise privée (impôt sur les salaires, sur les bénéfices, taxes sur l'exportation...), et chaque année, le groupe OCP participe au budget de l'état par versement de ses dividendes.

1.6 ORGANIGRAMME JURIDIQUE:

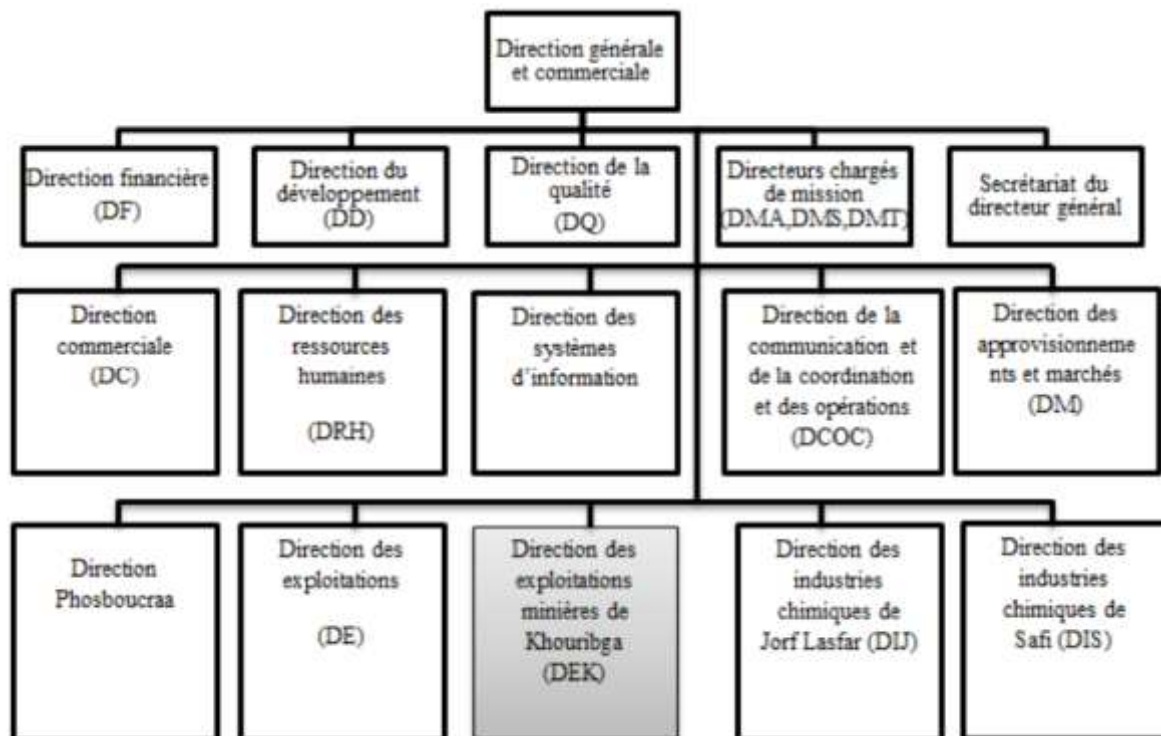


1.7 Filiales du groupe OCP :

Dans le but d'amélioration la gestion de la richesse que l'O.C.P a l'obligation de fructifier pour l'intérêt public, il a créé plusieurs filiales:

- SOTREG: Société des Transports Régionaux; pour le transport des agents O.C.P.
- MARPHOCEAN: Société chargée de transport maritime des chimiques
- I.P.S.I: Institut de Promotion Socio-éducative; dispense un enseignement fondamental de qualité pour les fils des agents du groupe.
- MAROC PHOSPHORE: Chargé du traitement industriel du phosphate

1.8 Organigramme de l'OCP :



1.9 Les Activités du groupe OCP :

1.9.1 La prospection

La prospection est une activité commerciale qui consiste à identifier et à contacter des prospects potentiels. L'OCP prospecte de nouveaux clients

dans le monde entier, en particulier dans les pays en développement, où la demande en engrais est forte.

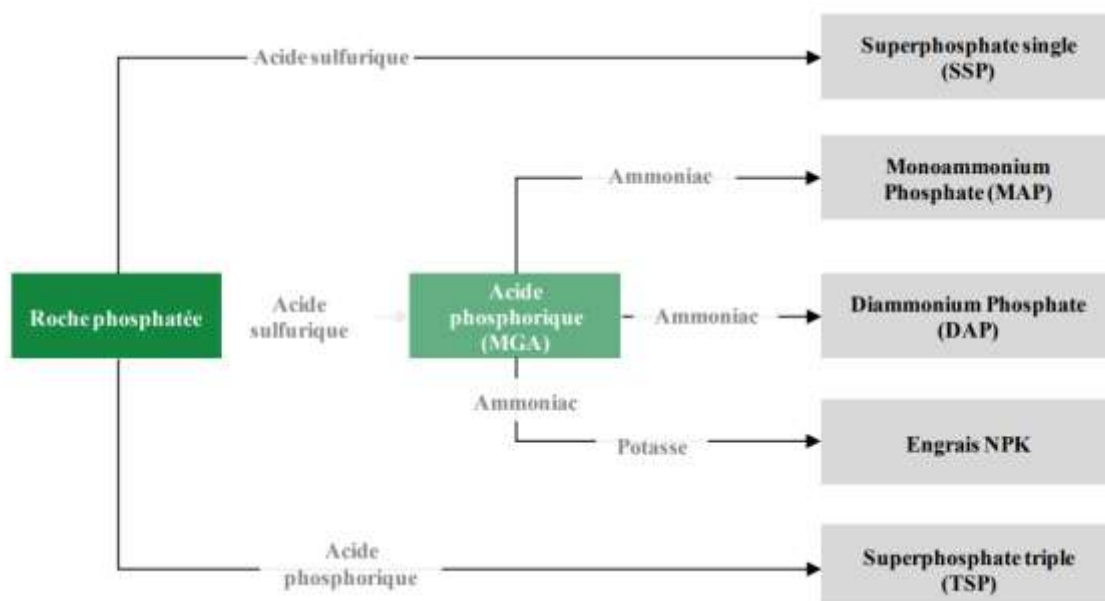
L'OCP utilise une variété de méthodes de prospection, notamment :

- La participation à des salons professionnels
- La publication de publicités
- L'envoi de mailings
- Le développement de partenariats avec des distributeurs locaux

1.9.2 La production :

L'OCP est le premier producteur mondial de phosphate. L'entreprise exploite six mines de phosphate au Maroc, dont la plus grande est la mine de Khouribga. Le phosphate extrait est transporté par pipeline jusqu'au site de traitement de Jorf Lasfar, où il est transformé en roche de phosphate.

Le phosphate est une ressource non renouvelable. L'OCP est conscient de la nécessité de gérer durablement ses ressources et d'investir dans des technologies d'extraction et de traitement plus respectueuses de l'environnement.



1.9.3 La Valorisation :

La roche de phosphate est transformée en engrais, en acide phosphorique et en produits dérivés. L'OCP dispose de trois complexes chimiques à Jorf Lasfar, Safi et Khouribga. Ces complexes transforment la roche de phosphate en engrais, en acide phosphorique et en produits dérivés.

Les engrais sont utilisés pour fertiliser les sols et améliorer la production agricole. L'acide phosphorique est utilisé dans la fabrication de produits chimiques, de plastiques et de détergents.

1.9.4 La commercialisation :

L'OCP commercialise ses produits à travers le monde via un réseau de distributeurs. L'entreprise est présente dans plus de 100 pays et exporte ses produits vers toutes les régions du monde.

L'OCP travaille en étroite collaboration avec ses distributeurs pour leur fournir des produits de qualité et un service client de premier ordre.

En conclusion, le groupe OCP est une entreprise intégrée qui couvre l'ensemble de la chaîne de valeur du phosphate. L'entreprise est un acteur majeur de l'économie mondiale et joue un rôle important dans la sécurité alimentaire.

1.10 Le Phosphate et la phosphatogenèse :

1.10.1 Définition :

Le phosphate est un élément chimique de symbole P et de numéro atomique 15. C'est un non-métal présent dans les roches, les sols, les eaux et les organismes vivants. Le phosphate est un élément essentiel à la vie, car il est impliqué dans de nombreux processus biologiques, notamment la croissance, le développement et la reproduction.

1.10.2 L'origine du phosphate :

Le phosphate est un élément abondant dans l'univers. Il est présent dans les météorites, les comètes et les étoiles. Sur Terre, le phosphate est principalement présent dans les roches phosphatées, qui se forment à partir de la décomposition de la matière organique.

Les principales sources de phosphates sont :

Les roches phosphatées, qui sont des roches sédimentaires constituées de phosphate de calcium, de phosphate de fer et de phosphate de magnésium.

Les engrais phosphatés, qui sont utilisés pour fertiliser les sols.

Les eaux usées, qui contiennent des phosphates provenant des engrais et des produits d'hygiène

1.10.3 Différents types de phosphates :

En fonction de sa teneur en BPL, le phosphate est classé en quatre catégories :

-Phosphate 75% en BPL qui ne subit en général aucune opération d'enrichissement sauf le criblage pour éliminer les grosses pierres, les stériles, et séchage pour éliminer l'humidité

-Phosphate 72% en BPL qui subit lui aussi le criblage et le séchage.

Phosphate pauvre en BPL et contient beaucoup d'impuretés, pour les enrichir il subit différents modes de traitements, à savoir :

- Le criblage pour éliminer les stériles.
- Le lavage à l'eau pour éliminer les argiles.
- Le séchage pour éliminer l'eau de lavage.

Phosphate qui contient beaucoup de matières organiques et de calcaire, est traité par calcination à 950°C pour éliminer les matières organiques et le CO₂.

Types de Phosphates	Signification	Tranche de BPL	Code
SHT	Super haut teneur	Supérieur à 75	1
THT	Très haut teneur	73-75	2

HTN	Haute teneur normale	71,5-73	3
HTM	Haute teneur moyen	69.5-71.5	4
MT	Moyen teneur	68.5-69.5	5
BTR	Basse teneur riche	65-68	6
BTN	Basse teneur normale	63-65	7
BTP	Basse teneur pauvre	61-63	8
TBT	Très basse teneur	56-61	9

1.10.4 Théories de la phosphatogenèse :

La phosphatogenèse est le processus de formation des phosphates. Il existe deux principales théories de la phosphatogenèse :

La théorie géochimique suggère que les phosphates se forment à partir de la décomposition de la matière organique. La matière organique est riche en phosphore, qui est libéré sous forme de phosphates lors de sa décomposition.

La théorie biologique suggère que les phosphates se forment à partir de l'activité des bactéries. Certaines bactéries sont capables de fixer le phosphore de l'eau et de le transformer en phosphates.

La phosphatogenèse est un processus complexe qui est encore mal compris. Les deux théories mentionnées ci-dessus sont complémentaires et il est probable que les deux processus interviennent dans la formation des phosphates.

1.11 L'emplacement géographique d'OCP :

1.11.1 Présence au Maroc

Trois régions au Maroc concentrent les activités minières du Groupe avec 4 sites miniers à Khouribga (Sidi Daoui, Merah El Ahrach, Sidi Chennae et Béni Amri), 2 à Gantour (Benguérir, Bouchane et Mzinda) et 1 à Boucraâ. Les activités de transformation du phosphate en acidephosphorique et en engrais phosphatés sont essentiellement concentrées au niveau des sites de Jorf Lasfar et de Safi. Un projet d'envergure lié au développement industriel de Phosboucraa est également déployé sur la période 2014-2020 en vue de renforcer les activités industrielles du site de Boucraâ, diversifier le portefeuille produits, développer l'écosystème régional et contribuer au développement socio-économique des régions du Sud du Maroc : Guelmim-Oued Noun, Laâyoune-Sakia El Hamra et Dakhla – Oued Ed-Dahab.



1.11.2 Présence à l'international

Avec plus des 160 clients à travers les 5 continents, le Groupe OCP consolide davantage ses positions sur les produits finis tout en renforçant sa présence en particulier en Afrique, en Amérique du Nord et en Amérique Latine. La flexibilité industrielle et l'agilité commerciale du Groupe se reflètent sur ses portefeuilles produits et régions de plus en plus diversifiés.



1.11.3 Importance de l'Industrie Minière des Phosphates

Approvisionnement en engrais et produits chimiques : Les phosphates extraits des gisements miniers sont transformés en engrais phosphatés, qui sont essentiels pour soutenir la croissance des plantes et l'agriculture. De plus, les phosphates sont utilisés dans la production d'une variété de produits chimiques industriels.

Sécurité alimentaire et agriculture : Les engrais phosphatés sont un élément clé pour augmenter les rendements agricoles et assurer la sécurité alimentaire mondiale en fournissant aux cultures les nutriments nécessaires.

Croissance économique et emplois : L'industrie minière des phosphates génère des emplois directs et indirects, contribuant ainsi au développement économique des régions où elle opère.

Exportations et revenus nationaux : De nombreux pays dépendent des exportations de phosphates pour générer des revenus et soutenir leur économie.

Rôle environnemental : Les entreprises minières sont de plus en plus conscientes de l'importance de l'exploitation durable et responsable des ressources, afin de minimiser l'impact environnemental et de préserver les écosystèmes locaux.

1.11.4 Enjeux et Défis de l'Industrie Minière des Phosphates :

Gestion durable des ressources : Les phosphates sont des ressources non renouvelables, ce qui rend essentielle leur gestion responsable pour assurer leur disponibilité à long terme.

Impact environnemental : L'extraction minière peut causer des dommages environnementaux, tels que la dégradation des sols, la contamination de l'eau et la destruction de la biodiversité. Les entreprises doivent mettre en œuvre des pratiques d'extraction responsable pour minimiser ces impacts.

Santé et sécurité des travailleurs : Les opérations minières comportent des risques pour la santé et la sécurité des travailleurs. Il est crucial de garantir un environnement de travail sûr et d'assurer la formation adéquate des employés.

Innovation et technologie : Pour améliorer l'efficacité de l'extraction, réduire l'impact environnemental et maintenir la compétitivité, l'industrie minière des phosphates doit continuer à innover et à adopter de nouvelles technologies.

Responsabilité sociale : Les entreprises minières ont un rôle dans le développement local des communautés où elles opèrent. Elles doivent s'engager dans des pratiques socialement responsables pour contribuer au bien-être des populations locales.

1.12 Conclusion

Ce chapitre a jeté les bases pour une compréhension approfondie de l'Office Chérifien des Phosphates (OCP) et de son rôle majeur dans l'industrie des phosphates. Nous avons examiné l'évolution historique de l'entreprise, sa structure organisationnelle, ses activités clés allant de la prospection à la commercialisation, ainsi que sa présence géographique tant au Maroc qu'à l'international. De plus, nous avons exploré la signification et les enjeux de l'industrie minière des phosphates, en mettant l'accent sur la gestion responsable des ressources et la contribution à la sécurité alimentaire mondiale. Cette vue d'ensemble nous prépare à approfondir nos connaissances sur les aspects spécifiques de l'OCP et de son rôle au sein du secteur des phosphates.

2 – PHOS-PREDECT-AI: Integrated Prediction of Price, Quantity in Phosphate Industry

2.1 Introduction

PHOS-PREDECT-AI est un système de prévision intégré de pointe conçu pour prévoir les tendances des prix et des quantités au sein de l'industrie du phosphate. En tirant parti de techniques avancées d'analyse de données, d'apprentissage automatique et de modélisation prédictive, PHOS-PREDECT-AI vise à fournir des informations précieuses aux entreprises, aux investisseurs et aux parties prenantes opérant dans le secteur des phosphates.

L'industrie du phosphate joue un rôle crucial dans divers secteurs, notamment l'agriculture, la production alimentaire et les processus industriels. La dynamique de l'industrie est influencée par une myriade de facteurs tels que l'offre et la demande mondiales, les événements géopolitiques, les progrès technologiques, les réglementations environnementales et les tendances économiques. PHOS-PREDECT-AI aborde la complexité de ces variables en analysant les données historiques, les tendances du marché et les influences externes pour générer des prévisions précises.

En intégrant des capacités de prévision des prix et des quantités, PHOS-PREDECT-AI offre une vue complète du paysage du marché du phosphate. Cette double approche de prédiction permet aux parties prenantes de prendre des décisions éclairées concernant la production, l'investissement, l'approvisionnement et l'allocation des ressources. Par exemple, les sociétés minières de phosphate peuvent utiliser les prévisions pour optimiser les niveaux de production en fonction de la demande projetée et des conditions du marché. De même, les fabricants d'engrais agricoles peuvent ajuster leurs stratégies en fonction des fluctuations de prix anticipées.

Les techniques de modélisation avancées de PHOS-PREDECT-AI prennent en compte l'interaction complexe entre l'offre, la demande et les facteurs externes. Le système adapte et affine continuellement ses prévisions à mesure que de nouvelles données deviennent disponibles, garantissant ainsi que les prévisions restent pertinentes et précises même dans un environnement de marché dynamique.

En résumé, PHOS-PREDECT-AI représente une avancée significative dans le domaine de l'analyse prédictive pour l'industrie du phosphate. En proposant des prévisions intégrées sur les tendances des prix et des quantités, il permet aux entreprises de naviguer en toute confiance dans les complexités du marché, d'optimiser leurs opérations et de prendre des décisions stratégiques alignées sur les réalités du marché.

2.2 Problématique

L'industrie du phosphate joue un rôle central dans l'agriculture et les économies mondiales, fournissant l'ingrédient essentiel des engrais qui soutiennent la production alimentaire. Cependant, cette industrie est entachée de complexités et d'incertitudes inhérentes qui donnent lieu à une volatilité des prix et des quantités de phosphate. Plusieurs facteurs clés contribuent à cette volatilité, créant un paysage difficile qui nécessite des solutions innovantes comme le projet PHOS-PREDECT-AI.

Facteur 1 : Problèmes géopolitiques et dynamique du marché

Les prix mondiaux du phosphate sont fortement influencés par les tensions géopolitiques, les accords commerciaux et les changements réglementaires. Les changements dans le climat politique et dans les relations internationales peuvent entraîner des perturbations soudaines dans les chaînes d'approvisionnement du phosphate, entraînant des fluctuations considérables des prix. En outre, la dynamique du marché, telle que l'arrivée de nouveaux concurrents ou l'évolution des préférences des consommateurs, peut avoir un impact significatif sur la demande et l'offre, contribuant ainsi à l'instabilité des prix et des quantités.

Facteur 2 : perturbations de la chaîne d'approvisionnement et défis de production

Les perturbations de la chaîne d'approvisionnement peuvent provenir de diverses sources, telles que des catastrophes naturelles, des grèves ou des goulots d'étranglement dans les transports. Ces perturbations peuvent paralyser le transport opportun du phosphate depuis les installations de production vers les utilisateurs finaux, entraînant des pénuries ou des excédents qui perturbent l'équilibre du marché. Les problèmes de production, tels que les pannes d'équipement

ou les pénuries de matières premières, peuvent également amplifier ces perturbations et exacerber les incertitudes sur les prix et les quantités.

Facteur 3 : Modification des pratiques agricoles et préoccupations environnementales

Les pratiques agricoles modernes évoluent en réponse aux préoccupations environnementales et aux objectifs de durabilité. À mesure que les agriculteurs adoptent des pratiques minimisant l'utilisation d'engrais chimiques ou se tournent vers des alternatives biologiques, la demande d'engrais traditionnels à base de phosphate connaît des fluctuations. Ce changement affecte non seulement la quantité de phosphate requise, mais influence également la structure globale de la demande, ce qui a un impact sur les prix.

2.3 Conséquences de prédictions inexactes

Des prévisions inexactes dans l'industrie du phosphate ont de profondes répercussions qui se répercutent sur divers secteurs :

2.3.1 Allocation inefficace des ressources :

Les agriculteurs, les fournisseurs et les fabricants s'appuient largement sur des prévisions précises pour planifier leurs activités, y compris les calendriers de plantation, les décisions d'achat et les plans de production. Des prévisions inexactes peuvent conduire à une mauvaise allocation des ressources, provoquant une surproduction ou une sous-production qui met à rude épreuve les chaînes d'approvisionnement et entraîne des coûts inutiles.

2.3.2 Pertes financières :

Les entreprises de l'industrie du phosphate fonctionnent dans un équilibre délicat, où les marges bénéficiaires sont influencées à la fois par les coûts de production et les prix du marché. De mauvaises prévisions peuvent conduire à des décisions de tarification qui érodent

les marges bénéficiaires, ou à une production excessive qui entraîne des stocks invendus et des pertes financières.

2.3.3 Impacts environnementaux :

La durabilité de l'industrie du phosphate dépend de l'utilisation responsable des ressources. Des prévisions inexactes peuvent conduire à une application excessive d'engrais, entraînant un ruissellement de nutriments nuisible aux écosystèmes aquatiques. À l'inverse, une sous-application peut compromettre les rendements des cultures et la sécurité alimentaire.

La nécessité d'une approche intégrée :

Relever ces défis nécessite une approche intégrée qui prend simultanément en compte les prévisions de prix et de quantités. Les prévisions isolées ne parviennent pas à saisir l'interaction complexe entre la dynamique du marché, les perturbations de la chaîne d'approvisionnement et l'évolution des pratiques agricoles. Un modèle intégré permet aux parties prenantes de prendre des décisions plus éclairées, garantissant une utilisation optimale des ressources, une gestion améliorée des risques et une planification stratégique améliorée. En réunissant ces prévisions, le projet PHOS-PREDECT-AI vise à doter l'industrie du phosphate des outils nécessaires pour faire face à sa volatilité inhérente et contribuer à un avenir plus durable et plus résilient.

2.4 Environnement de travail

L'environnement de travail Python et Google Colab sont deux outils extrêmement utiles pour le développement, l'analyse de données et la création de modèles d'apprentissage automatique. Ils offrent une plateforme flexible et conviviale pour les développeurs, les chercheurs et les data scientists. Voici un aperçu de chacun de ces environnements :

2.4.1 Environnement de travail Python :

Python est un langage de programmation polyvalent et largement utilisé dans le domaine de l'analyse de données, de l'intelligence artificielle et du développement logiciel en général. Il est apprécié pour sa syntaxe simple et lisible, ce qui en fait un excellent choix pour les débutants tout en offrant une grande puissance pour les tâches avancées.

Dans un environnement Python, vous pouvez utiliser des éditeurs de code tels que Visual Studio Code, PyCharm, Jupyter Notebook, ou simplement un éditeur de texte comme Notepad++. Vous pouvez installer des bibliothèques externes telles que NumPy, pandas, Matplotlib et scikit-learn pour effectuer des calculs, manipuler des données et créer des visualisations.



2.4.2 Google Colab :

Google Colab (abrégé de Colaboratory) est une plateforme basée sur le cloud qui offre un environnement Jupyter Notebook (Notebook interactif) pour exécuter du code Python. L'un des avantages majeurs de Colab est qu'il est entièrement hébergé par Google, ce qui signifie que vous n'avez pas besoin de configurer un environnement Python sur votre propre machine. Il vous permet d'exécuter du code, d'afficher des visualisations et d'analyser des données directement depuis votre navigateur.

Colab propose également des ressources gratuites en termes de puissance de calcul (GPU et TPU) pour accélérer l'exécution de tâches intensives en calcul, comme la formation de modèles d'apprentissage automatique.



En résumé, Python offre une grande flexibilité et est idéal pour le développement général, tandis que Google Colab est un excellent choix pour exécuter des notebooks interactifs basés sur Jupyter dans le cloud, ce qui est particulièrement utile pour le développement d'applications d'apprentissage automatique. Chacun de ces environnements a ses avantages spécifiques, et le choix entre les deux dépendra des besoins de votre projet et de votre préférence personnelle.

2.5 Jeux de données de l'étude

Dans notre projet, nous devons travailler avec deux ensembles de données pour notre objectif

2.5.1 Description du jeu de données de prédiction des prix (Price prediction):

Le jeu de données que nous avons à disposition pour la prédiction des prix avec une taille de (206, 6) est composé de plusieurs colonnes importantes qui fournissent des informations cruciales pour comprendre et anticiper les tendances des prix. Voici un aperçu des différentes colonnes :

1. Mois :

Cette colonne indique le mois auquel les données se rapportent. Elle permet de saisir l'aspect temporel de l'ensemble de données et de visualiser comment les prix évoluent tout au long de l'année.

2. Phosphate Price (Dollars américains par tonne métrique)

Cette colonne enregistre les prix du phosphate, un élément crucial dans l'industrie. Les données sont exprimées en dollars américains par tonne métrique, fournissant ainsi une mesure standard pour les prix.

3. Diesel Price (Dollars US par gallon):

Le prix du diesel est une autre variable essentielle qui peut influencer les prix des matières premières comme le phosphate. Les valeurs sont enregistrées en dollars américains par gallon, offrant une perspective sur les coûts énergétiques.

4. Phosphate ROC :

Cette colonne représente le taux de variation du prix du phosphate par rapport à une période précédente. Le taux de variation offre des informations sur la direction et l'intensité des fluctuations de prix.

5. Diesel ROC:

Semblable au taux de variation du phosphate, cette colonne présente le taux de variation du prix du diesel par rapport à une période antérieure. Cela permet de capturer les changements dans les coûts énergétiques.

6. Phosphate/Diesel Price Ratio:

Le ratio entre le prix du phosphate et le prix du diesel est une métrique importante pour comprendre la relation entre ces deux facteurs. Il peut avoir des implications sur les coûts de production et les marges bénéficiaires.

Ce jeu de données est essentiel pour entraîner des modèles de prédiction de prix, car il contient des informations temporelles et plusieurs variables importantes qui influencent les prix du phosphate. L'utilisation de ces caractéristiques permettra de créer des modèles prédictifs plus précis pour anticiper les tendances et prendre des décisions éclairées dans l'industrie.

	Mois	Phosphate Price (Dollars américains par tonne métrique)	Diesel Price (Dollars US par gallon)	Phosphate ROC	Diesel ROC	Phosphate/Diesel Price Ratio
0	juin 2006	44,00	2,09	-	-	210,426
1	juil. 2006	44,00	2,22	0,00 %	6,03 %	19,8468
2	août 2006	44,00	2,25	0,00 %	1,35 %	19,5817
3	sept. 2006	44,00	1,81	0,00 %	-19,45 %	24,3094
4	oct. 2006	44,00	1,79	0,00 %	-0,88 %	24,5262
5	nov. 2006	44,00	1,81	0,00 %	0,95 %	24,2960
6	déc. 2006	44,00	1,83	0,00 %	1,21 %	24,0044

2.5.2 Description du jeu de données de prédiction quantité de phosphate nutritif :

Cet ensemble de données fournit des informations précieuses sur les quantités de phosphates nutritifs avec une taille de (7670, 6) ainsi que d'autres nutriments clés, au cours de différentes années et entités. L'ensemble de données comprend plusieurs colonnes, chacune contenant des informations spécifiques liées aux quantités de phosphate nutritif et à leurs tendances.

1. Entité :

La colonne « Entity » représente les entités ou régions géographiques pour lesquelles les données sur la quantité de phosphate nutritif sont enregistrées. Cela peut faire référence à des pays, des régions ou à toute autre entité définie pertinente au contexte de la consommation et de la production de nutriments.

2. Année :

La colonne « Year » indique l'année spécifique pour laquelle les données sur la quantité de phosphate nutritif sont enregistrées. Il sert de référence temporelle, permettant l'analyse des tendances et des changements au fil du temps.

3. Nutrient potash :

Cette colonne « Nutrient potash » contient des données concernant la quantité de potasse nutritive. La potasse est un nutriment important dans les engrais et est essentielle à la croissance et au développement des plantes.

4. Nutrient phosphate

La colonne «Nutrient phosphate» fournit des informations sur la quantité de phosphate nutritif. Le phosphate est un nutriment essentiel pour les plantes et se trouve couramment dans les engrais pour favoriser le développement, la floraison et la fructification sains des racines.

5. Nutrient nitrogen:

Cette colonne « Nutrient nitrogen» enregistre la quantité d'azote nutritif. L'azote est un nutriment fondamental qui joue un rôle central dans la synthèse des protéines végétales, la croissance des feuilles et la vitalité globale des plantes.

Ce jeu de données est essentiel pour entraîner des modèles de prédiction de quantité, car il contient des informations temporelles et plusieurs variables importantes qui influencent les quantités du phosphate. L'utilisation de ces caractéristiques permettra de créer des modèles prédictifs plus précis pour anticiper les tendances et prendre des décisions éclairées dans l'industrie.

	Entity	Code	Year	Nutrient potash K20 (total)		Nutrient phosphate P205 (total)		Nutrient nitrogen N (total)	
				00003104 Production	005510 tonnes	00003103 Production	005510 tonnes	00003102 Production	005510 tonnes
0	Afghanistan	AFG	2002		0.0		0.0		23414.0
1	Afghanistan	AFG	2003		0.0		0.0		16206.0
2	Afghanistan	AFG	2013		0.0		1539.0		15694.0
3	Afghanistan	AFG	2014		0.0		1541.0		14416.0
4	Afghanistan	AFG	2015		0.0		1541.0		14416.0
5	Afghanistan	AFG	2016		0.0		1541.0		14416.0
6	Afghanistan	AFG	2017		0.0		1541.0		14416.0
7	Afghanistan	AFG	2018		0.0		1541.0		14416.0
8	Afghanistan	AFG	2019		0.0		1541.0		14416.0
9	Afghanistan	AFG	2020		0.0		1541.0		14416.0
10	Afghanistan	AFG	2011		NaN		1563.0		17555.0

2.6 Conclusion :

En résumé, le jeu de données constitue le fondement de toute analyse et modélisation dans l'analyse de données et l'apprentissage automatique. Sa qualité et préparation sont essentielles pour des résultats précis. Nous avons préparé les données en les nettoyant, les transformant en formats compatibles, et en les séparant en ensembles d'entraînement et de test pour évaluer les modèles de manière impartiale. Chaque étape a contribué à mieux comprendre les données, identifier des tendances et prendre des décisions éclairées. La rigueur dans la préparation du jeu de données est cruciale pour créer des modèles puissants et réaliser des analyses pertinentes.

3 - Préparation de données

La préparation de données est une étape cruciale et incontournable dans le domaine de l'analyse de données, de l'apprentissage automatique et de l'intelligence artificielle. Elle englobe un ensemble de processus visant à collecter, nettoyer, transformer et organiser les données brutes afin de les rendre utilisables et pertinentes pour des analyses approfondies ou des modèles prédictifs.

La préparation de données vise à résoudre ces problèmes de prédiction des prix et de quantités de phosphate en les traitant de la manière suivante:

3.1 Lecture le fichier CSV :

Pour démarrer notre processus de préparation de données, nous avons commencé par importer les données à partir d'un fichier CSV pour initier le processus de préparation. Cette étape fondamentale a permis de charger les informations tabulaires, posant ainsi les bases de l'exploration, du nettoyage et de la transformation ultérieure des données. En utilisant la bibliothèque pandas en python de programmation dédiées au traitement de données.

3.2 Nettoyage des données :

Le nettoyage des données est une étape cruciale et incontournable dans le processus d'analyse de données et de préparation pour des tâches telles que l'apprentissage automatique. Il s'agit d'un processus méticuleux visant à détecter, corriger et éliminer les incohérences, les erreurs et les valeurs aberrantes présentes dans un ensemble de données, afin de garantir leur qualité et leur fiabilité.

Lorsque des données sont collectées à partir de diverses sources, il est courant qu'elles soient entachées de bruit et d'irrégularités. Le nettoyage des données vise à résoudre ces problèmes pour que les données soient exploitables et fournissent des résultats précis et significatifs. Voici quelques-unes des étapes et des activités typiques impliquées dans le processus de nettoyage des données:

3.2.1 Identification des données manquantes

Ensemble des données de prédiction des prix

Pour vérifier que on a des valeurs manquant en utilise la commande suivante :

```
df.isnull().sum()
Mois 0
Phosphate Price (Dollars américains par tonne métrique) 0
Diesel Price (Dollars US par gallon) 0
Phosphate ROC 0
Diesel ROC 0
Phosphate/Diesel Price Ratio 0
dtype: int64
```

Qu'après avoir exécuté cette commande, nous n'avons aucune valeur manquante, comme il montre cette figure.

Ensemble des données de prédiction des quantités

Pour vérifier que on a des valeurs manquant en utilise la commande suivante :

```
[6]: df.isnull().sum()
Entity 0
Code 1790
Year 0
Nutrient_potash 3337
Nutrient_phosphate 850
Nutrient_nitrogen 304
dtype: int64
```

Qu'après avoir exécuté cette commande, nous n'avons des valeurs manquantes, comme il montre cette figure. Et la solution proposée pour résoudre ce problème c'est la suppression utilisant la commande suivant :

```
df=df.dropna()
```

3.2.2 Nettoyage des données textuelles :

Lors du travail avec des données textuelles, cela peut impliquer la suppression de caractères spéciaux, la conversion en minuscules, la correction d'orthographe ou même la suppression de mots vides (stop words) inutiles.

Ensemble des données de prédiction des prix

Noter ensemble de données on a identifié plusieurs caractères spéciaux comme '%' dans les colonnes « Diesel ROC » et « phosphat ROC », ','

dans les valeur numérique et '-' dans certaine colonne identifie les nulle valeur, comme il montre cette figure.

	Mois	Phosphate Price (Dollars américains par tonne métrique)	Diesel Price (Dollars US par gallon)	Phosphate ROC	Diesel ROC	Phosphate/Diesel Price Ratio
0	juin 2006	44,00	2,09	-	-	210,426
1	juil. 2006	44,00	2,22	0,00 %	6,03 %	19,8466
2	août 2006	44,00	2,25	0,00 %	1,35 %	19,5817
3	sept. 2006	44,00	1,81	0,00 %	-19,45 %	24,3094
4	oct. 2006	44,00	1,79	0,00 %	-0,88 %	24,5262
5	nov. 2006	44,00	1,81	0,00 %	0,95 %	24,2960
6	déc. 2006	44,00	1,83	0,00 %	1,21 %	24,0044

Pour ressouder ces types des problèmes c'est de supprimer '%' et remplace les valeurs ',' par '.', '-' par '0'.

```
df['Phosphate ROC'] = df['Phosphate ROC'].replace('-', '0')
df['Phosphate ROC'] = df['Phosphate ROC'].str.replace(',', '.')
df['Phosphate ROC'] = df['Phosphate ROC'].str.rstrip('%')

df['Diesel ROC'] = df['Diesel ROC'].replace('-', '0')
df['Diesel ROC'] = df['Diesel ROC'].str.replace(',', '.')
df['Diesel ROC'] = df['Diesel ROC'].str.rstrip('%')

df['Diesel Price (Dollars US par gallon)'] = df['Diesel Price (Dollars US par gallon)'].str.replace(',', '.')

df['Phosphate Price (Dollars américains par tonne métrique)'] = df['Phosphate Price (Dollars américains par tonne métrique)'].astype(str).str.replace(',', '.')

df['Phosphate/Diesel Price Ratio'] = df['Phosphate/Diesel Price Ratio'].str.replace(',', '.')
```

Ensemble des données de prédiction des quantités

Dans cette ensemble des données nous n'avons pas besoin de faire cette procédure.

Remarque :

Tous concernant cette ensemble des données on à faire d'autre nettoyage comme la suppression de colonne « Code » avoir la même signification que le colonne « Entity » et aussi les lignes en colonne « Nutrient phosphate » qui la valeur égale a '0.0'.

3.3 Transformation :

Dans le cadre de la préparation minutieuse de nos données pour une analyse approfondie, nous avons entrepris une étape essentielle de transformation c'est de Conversion les données dans un format approprié pour l'analyse ou la modélisation. Cela peut inclure la normalisation des valeurs, la création de nouvelles caractéristiques à partir des données existantes, ou encore la conversion de données catégorielles en données numériques.

Pour rendre ces variables utilisables dans des calculs numériques, des analyses statistiques et la création de modèles, nous avons entrepris une conversion soigneuse.

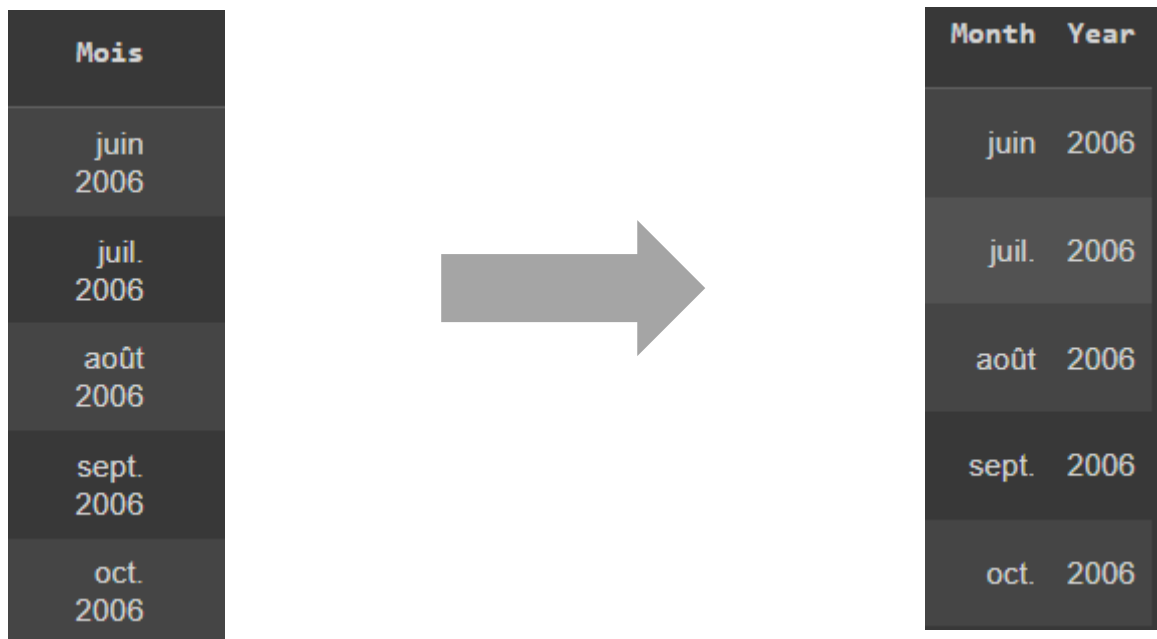
Ensemble des données de prédiction des prix

Tout d'abord, la création de nouvelles caractéristiques ou colonne à partir des données existantes comme le colonne « Mois » a deux valeur, le mois et l'année, cela signifie que on peut créer deux colonnes "Month" et "Year" utilisant cette commande :

```
df[["Month", "Year"]] = df["Mois"].str.split(" ", n=1, expand=True)
```

Après on supprime la colonne « Mois » utilisant la fonction « drop ».

On trouve ce résultat :



Ensuite, nous avons identifié les variables susceptibles de contenir des valeurs numériques, mais qui étaient enregistrées sous forme de chaînes. Par exemple, des colonnes telles que 'Year', ' Phosphate/Diesel Price Ratio ', ' Diesel ROC ', ' Phosphate ROC ', ' Diesel Price (Dollars US par gallon)' et ' Phosphate Price (Dollars américains par tonne métrique)' semblaient représenter des données numériques, mais étaient stockées en tant qu'objets.

En appliquant des méthodes de conversion appropriées pour transformer ces valeurs en nombres à virgule flottante (float) et pour colonne 'Year' conversion en type entier, en prenant en compte la structure décimale et le format spécifique de chaque variable utilisant fonction « `astype()` » dans pandas .

Puis, la conversion de données catégoriel dans le colonne « Month » en valeurs numériques est une étape fréquente dans la préparation de données, en particulier lorsque les mois sont initialement représentés sous forme de chaînes de caractères. Cette conversion facilite l'analyse et le traitement ultérieur des données, car les valeurs numériques permettent une manipulation plus efficace dans diverses opérations, telles que le calcul de moyennes, de tendances saisonnières ou la création de modèles de machine Learning.

Lors de cette conversion, chaque mois est associé à une valeur numérique correspondante. Par exemple, janvier pourrait être converti en 1, février en 2, mars en 3 et ainsi de suite jusqu'à décembre qui serait converti en 12. Voici comment cette conversion pourrait être réalisée :

Nous ayons une colonne appelée 'Month' dans notre ensemble de données, où les valeurs sont actuellement représentées sous forme de chaînes (par exemple, 'janvier', 'février', etc.). Pour convertir ces mois en valeurs numériques, nous suivrions les étapes suivantes :

a- Création d'une correspondance mois-valeur numérique :

Créez une correspondance où chaque mois est associé à une valeur numérique.

- Tout d'abord :
 - janv → Janvier
 - févr → Février
 - mars → Mars
 - ...
 - dec→ Décembre

- Puis :
 - Janvier → 1
 - Février → 2
 - Mars → 3
 - ...
 - Décembre → 12

b- **Application de la conversion :** Appliquez cette correspondance à la colonne 'Mois' de votre ensemble de données. Par exemple, remplacez chaque mention de 'janvier' par le chiffre 1, 'février' par 2, et ainsi de suite.

c- **Vérification de la conversion :** Vérifiez que la conversion a été correctement effectuée en examinant quelques lignes de votre ensemble de données pour vous assurer que les mois sont maintenant représentés par des valeurs numériques.

Ensemble des données de prédiction des quantités

Lorsque nous travaillons avec des données catégorielles, nous devons attribuer une représentation numérique à chaque catégorie unique. Ce processus est appelé "encodage" et il existe plusieurs approches pour le faire, mais nous avons choisi celle-ci

Encodage par numérotation (Label Encoding) :

Chaque catégorie unique est assignée à un nombre unique. Cela fonctionne bien pour les variables ordinales (où il existe un ordre naturel entre les catégories), mais cela peut être problématique pour les variables nominales (sans ordre spécifique) car cela peut fausser l'interprétation des résultats. Noter que pour encoder la colonne « Entity » utilisant la commande « LabelEncoder » dans scikit-learn.

En somme, la conversion de données de mois en valeurs numériques est une étape importante pour rendre vos données plus exploitables et accessibles à une gamme plus large d'analyses et d'applications.

Élimination des doublons :

Identifier et supprimer les lignes en double dans le jeu de données, évitant ainsi toute duplication inutile et potentiellement trompeuse des informations. En notant les deux ensembles de données nous n'avons aucune donnée en double.

3.4 Exploration de données

L'exploration de données implique l'application de diverses techniques d'analyse pour découvrir des tendances, des schémas, des valeurs aberrantes et des caractéristiques intéressantes présentes dans les données. Voici quelques activités courantes dans l'exploration de données

1. **Résumé statistique initial** : Calcul des statistiques descriptives telles que la moyenne, la médiane, l'écart-type, etc., pour chaque variable dans les deux ensembles de données comme il est montré dans les tableaux suivants :

index	Phosphate Price	Diesel Price	Phosphate ROC	Diesel ROC	Phosphate/Diesel Price Ratio	Month	Year
count	206.0	206.0	206.0	206.0	206.0	206.0	206.0
mean	139.27388349514564	2.2959708737864077	1.7296116504854366	0.4989320388349514	64.9061504854369	6.5	2014.5
std	89.17080840540902	0.7495662349725607	13.270281444059373	8.937275124934523	49.12460347033782	3.4439767144619458	4.983631744606904
min	44.0	0.88	-54.55	-26.63	19.5817	1.0	2006.0
25%	87.5	1.8	-0.275	-4.367500000000001	41.671525	4.0	2010.0
50%	110.5	2.12	0.0	0.8500000000000001	54.47525	6.5	2014.5
75%	169.845	2.94	2.5025	5.7625	66.4576	9.0	2019.0
max	450.0	4.65	124.31	31.62	344.5635	12.0	2023.0

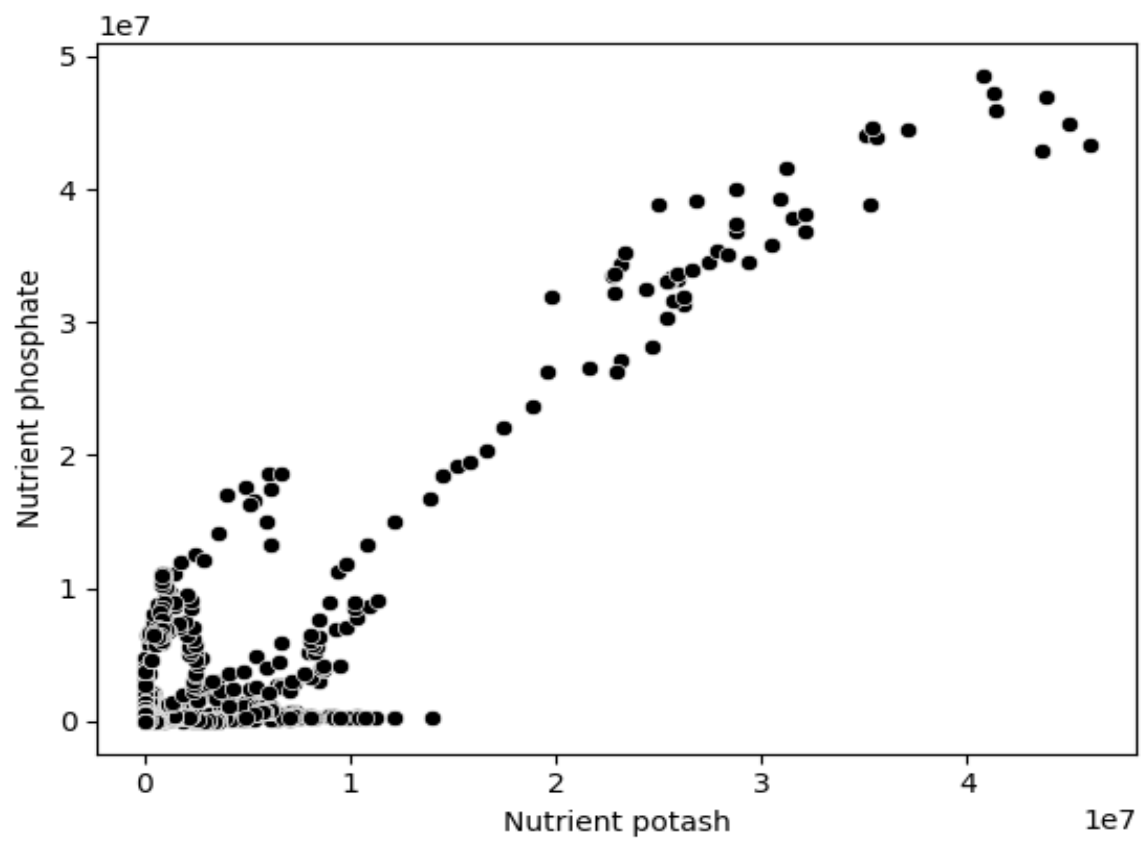
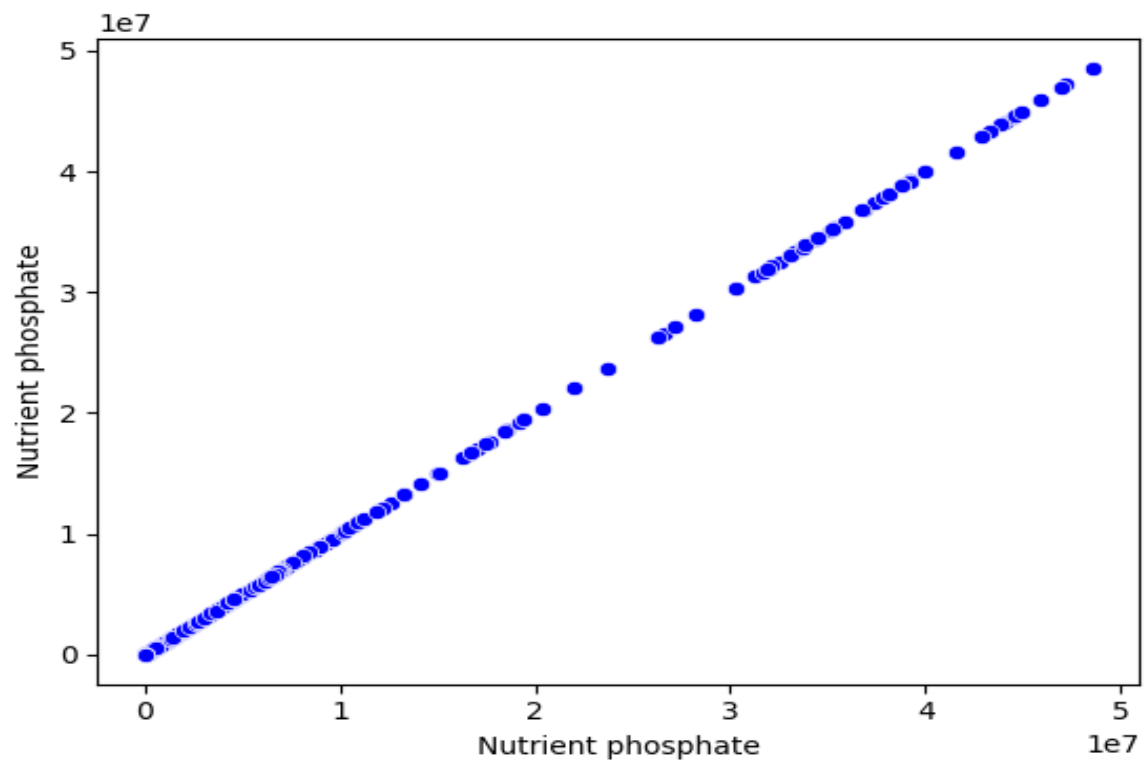
Tableaux : Ensemble des données de prédiction des prix

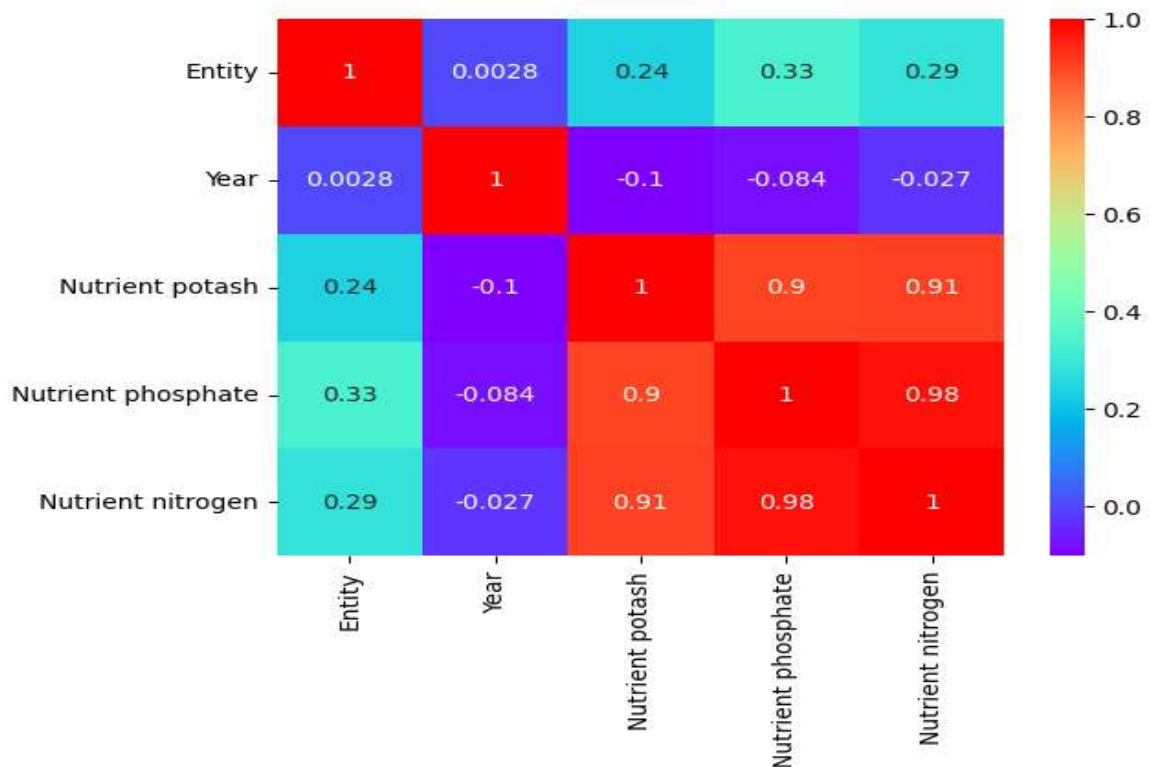
index	Entity	Year	Nutrient potash	Nutrient phosphate	Nutrient nitrogen
count	1608.0	1608.0	1608.0	1608.0	1608.0
mean	34.789800995024876	2000.1797263681592	1984169.341374378	2203458.6357089556	4845103.16960199
std	20.92653093053784	16.549355931714533	5582340.64851379	6711983.274935667	15639392.821112309
min	0.0	1961.0	0.0	0.2	0.0
25%	18.0	1989.0	0.0	75635.26250000001	157000.57749999998
50%	31.0	2005.0	17941.5	272650.0	728524.0
75%	52.0	2013.25	1061205.25	658250.0	1833250.0
max	71.0	2020.0	45925720.0	48530900.0	123987360.0

Tableau : Ensemble des données de prédiction des quantités

2. **Visualisation** : Création de graphiques et de visualisations pour représenter graphiquement les données, tels que des histogrammes, des diagrammes en boîte, des nuages de points, etc.

Par exemple dans **Ensemble des données de prédiction des quantités**



[illegible]

Conclusion :

En résumé, l'exploration de données vous permet de prendre connaissance de vos données, de repérer les problèmes potentiels et d'obtenir des informations préliminaires pour orienter les étapes de préparation. Elle vous aide à mieux comprendre vos données avant de commencer à les nettoyer, les transformer et les adapter à des analyses plus avancées ou à des modèles de machine learning.

3.5 Séparation des données :

La séparation des données en ensembles d'entraînement (train) et de test est une étape cruciale lors de la préparation des données pour la construction et l'évaluation de modèles d'apprentissage automatique. Un ratio communément utilisé pour diviser les données en ensembles de formation et de test est 0.7 pour l'ensemble de formation et 0.26 pour l'ensemble de test. Voici comment vous pourriez expliquer cette séparation avec un ratio de 0.7 pour la formation et 0.26 pour le test :

Pour garantir une évaluation précise et fiable de nos modèles d'apprentissage automatique, nous avons entrepris la séparation méthodique de nos données en deux ensembles distincts : l'ensemble d'entraînement (train) et l'ensemble de test. Cette démarche vise à évaluer la performance réelle du modèle sur des données qu'il n'a jamais rencontrées auparavant, ce qui évite le piège du surapprentissage et fournit une indication plus réaliste de sa capacité à généraliser. Utilisant la commande `train_test_split` dans la bibliothèque `scikit-learn`.

Nous avons choisi un ratio de 0.7 pour l'ensemble d'entraînement et 0.26 pour l'ensemble de test, ce qui signifie que 70 % de nos données seront utilisées pour entraîner le modèle et 30 % seront réservées pour le tester. Cette répartition permet d'avoir un ensemble de test significatif tout en fournissant suffisamment de données pour l'entraînement.

3.6 Conclusion :

En résumé, la préparation de données joue un rôle fondamental dans la création de modèles fiables et précis. Des données bien préparées permettent aux analystes et aux chercheurs de prendre des décisions éclairées et d'entraîner des modèles performants, jetant ainsi les bases d'analyses approfondies et de prises de décision pertinentes.

4 - Random Forest Regression model :

Dans le domaine de l'apprentissage automatique, la Random Forest Regression est une méthode puissante et polyvalente qui trouve de larges applications dans la prédiction et la modélisation de données. Cette approche repose sur le concept d'ensemble learning (apprentissage par ensemble), qui vise à combiner les prédictions de plusieurs modèles pour obtenir une prédiction plus précise et robuste. La Random Forest Regression est une extension de l'algorithme d'Arbres de Décision, qui est l'une des méthodes fondamentales de l'apprentissage automatique.

4.1 Définition :

Random Forest, ou "Forêt aléatoire" en français, est une technique d'apprentissage automatique utilisée pour la classification et la régression. C'est un type d'algorithme ensembliste qui combine les prédictions de multiples modèles d'arbres de décision individuels pour obtenir une prédiction plus robuste et précise.



////////////////////////////////////

Un arbre de décision est une structure hiérarchique qui divise progressivement les données en sous-groupes en fonction des caractéristiques, en prenant des décisions basées sur des règles logiques. Cependant, un seul arbre de décision peut être sujet à un surapprentissage (overfitting), c'est-à-dire qu'il peut être très précis sur les données d'entraînement mais moins performant sur de nouvelles données.

La Random Forest résout ce problème en combinant plusieurs arbres de décision, d'où le terme "forêt". Chaque arbre est construit en utilisant un échantillon aléatoire des données d'entraînement, et chaque arbre vote pour la prédiction finale. Cette agrégation de prédictions réduit le risque de surapprentissage et augmente la stabilité et la précision de la prédiction globale.

Le "random" dans le nom se réfère au fait que chaque arbre est formé avec une sous-section aléatoire des données et des caractéristiques, ce qui diversifie les arbres et augmente la variabilité, conduisant ainsi à un modèle plus généralisable.

4.2 Fonctionnement de Random Forest de régression

Un Random Forest de régression fonctionne en agrégeant les prédictions de plusieurs arbres de régression individuels pour obtenir une prédiction finale plus robuste et précise. Voici comment cela fonctionne en détail :

1. Entraînement des arbres de régression individuels : Pour construire un Random Forest de régression, plusieurs arbres de régression sont créés. Chaque arbre est formé à l'aide d'un sous-ensemble aléatoire des données d'entraînement et des caractéristiques. L'algorithme construit chaque arbre en divisant récursivement les données en fonction des caractéristiques pour minimiser l'erreur de régression. Chaque feuille de l'arbre contient une valeur de prédiction pour la cible.

2. Prédiction individuelle des arbres : Une fois que tous les arbres sont formés, chaque arbre est utilisé pour faire des prédictions individuelles pour chaque observation du jeu de données de test ou de validation. Chaque arbre produit une valeur de prédiction basée sur les caractéristiques de l'observation.

3. Agrégation des prédictions : Les prédictions individuelles de tous les arbres sont agrégées pour obtenir une prédiction finale. Pour les problèmes de régression, la moyenne ou la médiane des prédictions de tous les arbres peut être calculée pour obtenir la prédiction finale du modèle Random Forest. Cette agrégation permet de réduire la variance et d'obtenir une prédiction plus stable et précise.

4.3 Implémentation en python

L'implémentation d'un modèle de Random Forest de régression en Python en utilisant la bibliothèque scikit-learn, l'une des bibliothèques les plus populaires pour l'apprentissage automatique utilisant cette commande :

```
rf_regressor = RandomForestRegressor(n_estimators=100, random_state=42)
```

4.4 Caractéristiques de La régression avec Random Forest

La régression avec Random Forest caractérisées:

- **Type d'algorithme :**

Random Forest est une méthode d'apprentissage ensembliste qui combine plusieurs arbres de décision pour créer une prédiction plus robuste et précise.

- **Flexibilité :**

Elle peut traiter à la fois les relations linéaires et non linéaires dans les données, la rendant adaptée à un large éventail de problèmes de régression.

- **Surapprentissage :**

Random Forest a tendance à être moins sujet au surapprentissage par rapport à un seul arbre de décision.

- **Importance des caractéristiques :**

Elle peut fournir des informations sur l'importance des caractéristiques, ce qui aide à comprendre l'influence des différentes caractéristiques sur la prédiction.

- **Taille des données :**

En général, Random Forest fonctionne bien avec des ensembles de données modérés à grands.

- **Temps d'entraînement :**

L'entraînement peut prendre plus de temps par rapport à la régression linéaire, surtout si la forêt compte un grand nombre d'arbres.

- **Hyperparamètres :**

Elle a plus d'hyperparamètres à régler par rapport à la régression linéaire.

4.5 Avantages Clés :

- **Modélisation Non Linéaire :**

Contrairement à la régression linéaire, Random Forest Regression peut capturer des relations non linéaires complexes entre les caractéristiques et la variable cible, améliorant ainsi la qualité des prédictions.

- **Gestion des Caractéristiques :**

La méthode peut gérer des ensembles de données avec un grand nombre de caractéristiques, évitant ainsi les problèmes de colinéarité et d'overfitting.

- **Réduction du Surapprentissage :**

En utilisant des échantillons aléatoires et des sous-ensembles de caractéristiques, la méthode limite le risque de surapprentissage, offrant une généralisation plus solide sur de nouvelles données.

4.6 Applications Pratiques :

- **Prévisions Financières :**

Random Forest Regression est utilisée pour prédire les prix des actifs financiers, tenant compte des relations complexes entre divers indicateurs économiques.

- **Sciences de la Santé :**

La méthode est employée pour modéliser des variables biomédicales, tels que les taux de croissance de certaines cellules en fonction de diverses conditions.

- **Industrie :**

Random Forest Regression est appliquée pour prédire la demande de produits en fonction de facteurs tels que la saison, les tendances du marché et les conditions économiques.

4.7 Conclusion :

La Régression avec Random Forest ouvre des horizons passionnants pour la modélisation prédictive. Grâce à sa capacité à gérer des données complexes et non linéaires, cette approche offre une alternative puissante à la régression linéaire traditionnelle. Que ce soit dans la finance, les sciences de la santé ou l'industrie, Random Forest Regression s'impose comme un outil indispensable pour anticiper les comportements futurs à partir de données existantes.

5 - Evaluation de model de Régression avec Random Forest :

des données L'évaluation est un pilier fondamental de l'apprentissage automatique, visant à quantifier les performances des modèles par rapport aux données réelles. Lorsqu'il s'agit de la Régression avec Random Forest, une méthode puissante, il est crucial de comprendre comment évaluer la précision et la robustesse des prédictions qu'elle génère. Cette évaluation informe les praticiens sur la fiabilité du modèle et sa capacité à généraliser au-delà d'entraînement.

L'évaluation d'un modèle de Régression avec Random Forest à l'aide de la fonction `score()` offre un moyen rapide et pratique de mesurer sa performance sur des données de test ou de validation. Cependant, il est important de noter que la fonction `score()` peut différer légèrement en fonction de la bibliothèque ou du framework que vous utilisez pour implémenter Random Forest. Dans cet exemple, je vais supposer que

vous utilisez la bibliothèque scikit-learn en Python pour mettre en œuvre Random Forest Regression.

Voici comment vous pourriez expliquer l'évaluation à l'aide de la fonction `score()` :

5.1 Introduction :

Lors de l'évaluation d'un modèle de Régression avec Random Forest, l'objectif principal est de quantifier sa performance prédictive sur des données non vues. Pour ce faire, la fonction `score()` fournie par la bibliothèque scikit-learn en Python est un outil pratique qui permet de mesurer l'adéquation du modèle aux données de test ou de validation.

5.2 Utilisation de la Fonction `score()` :

La fonction `score()` est une méthode intégrée dans les modèles de Régression avec Random Forest de scikit-learn. Elle prend en entrée un ensemble de données de test (ou de validation) et les valeurs cibles correspondantes. Ensuite, elle calcule le coefficient de détermination R^2 , qui mesure la proportion de la variance des valeurs cibles expliquée par les prédictions du modèle. Plus le R^2 est proche de 1, meilleures sont les prédictions.

5.3 Interprétation du R^2 :

Un R^2 proche de 1 indique que le modèle explique la variance de la cible de manière satisfaisante. Cependant, il est important de noter que le R^2 peut être trompeur si le modèle est complexe. Par conséquent, il est recommandé de l'utiliser en conjonction avec d'autres métriques d'évaluation, telles que l'erreur quadratique moyenne (RMSE) ou l'erreur absolue moyenne (MAE), pour obtenir une image complète de la performance du modèle.

5.4 Exemple de Code :

Voici comment vous pourriez utiliser la fonction `score()` dans scikit-learn :

```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(features, target,
test_size=0.3, random_state=42)
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
# Évaluer le modèle en utilisant la fonction score()
r2_score = model.score(X_test, y_test)
print("Coefficient de détermination R² : {:.2f}".format(r2_score))
'''
```

Conclusion :

La fonction `score()` est un outil utile pour évaluer la performance d'un modèle de Régression avec Random Forest en termes de R^2 . Cependant, il est recommandé de considérer plusieurs métriques d'évaluation pour obtenir une image complète de la performance du modèle.

5.5 Conclusion

En somme, Random Forest s'impose comme une méthodologie d'apprentissage puissante. En agrégeant les prédictions de multiples arbres, il surmonte les limites des modèles individuels, résistant au surapprentissage et modélisant des relations non linéaires. Ses applications sont vastes, de la finance à la médecine. Cependant, l'ajustement des paramètres et l'interprétation exigent une expertise. En conclusion, Random Forest offre des prédictions robustes, reliant puissance et interprétabilité.

6 - Partie de mise en production :

6.1 Environnement de travail :

1. Virtual Environments (Environnements virtuels) :

a. Activation de env :

```
PS C:\Users\oki\Desktop\exempleApp> python -m venv env
>>
PS C:\Users\oki\Desktop\exempleApp> .\env\Scripts\activate
(env) PS C:\Users\oki\Desktop\exempleApp> pip install Flask
```

2. Structure de Dossiers :

Structure du dossier app flask :

C:\Users\Phos_Predict_App

```
| main.py
|
|--- Code_model
|    Price_phosphat.ipynb
|    quantityPhosphate.ipynb
|
|--- DataSet
|    |--- Clean
|    |    phosphateQuantite.csv
|    |    phosphate_clean.csv
|    |
|    |--- dataSet
|    |    phosphate36.csv
|    |    quantityPhosphate.ipynb
|
|--- instance
|
|--- model_save
|    modelQuntity_RFR.pkl
|    model_price_RFR.pkl
|
```



```

├── static
│   ├── assets
│   ├── css
│   ├── fonts
│   ├── images
│   └── js
└── templates
    base.html
    home.html
    predictedPrice.html
    predictedQuantity.html
    predictedTurnover.html
    prediction.html
    result_price.html
    result_quantity.html
    result_turnover.html

```

3. Fichier d'Application Principale :

Le fichier **main.py** (ou un autre nom selon votre choix) est le point d'entrée de notre application Flask. C'est là que nous importons et configurons l'application, définissons les routes et les vues, et lançons l'application.

4. Dépendances :

Assurez-vous que toutes les dépendances nécessaires pour votre projet Flask sont installées dans l'environnement virtuel. Vous pouvez le faire en exécutant `pip install` pour chaque package mentionner dans le fichier `requirements.txt`.

Ce fichier est crée en utilisant la commande suivantes

```
(env) PS C:\Users\oki\Desktop\exempleApp> pip freeze > requirements.txt
```

Commande utiliser pour installer flask

requirements.txt:

```
pip install blinker==1.6.2 certifi==2023.7.22 charset-normalizer==3.2.0 click==8.1.7
colorama==0.4.6 contourpy==1.1.0 cycler==0.11.0 Flask==2.3.3 Flask-SQLAlchemy==3.0.5
fonttools==4.42.1 greenlet==2.0.2 idna==3.4 itsdangerous==2.1.2 Jinja2==3.1.2 joblib==1.3.2
kiwisolver==1.4.5 MarkupSafe==2.1.3 matplotlib==3.7.2 numpy==1.25.2 packaging==23.1
pandas==2.0.3 Pillow==10.0.0 pyparsing==3.0.9 python-dateutil==2.8.2 pytz==2023.3
reportlab==4.0.4 requests==2.31.0 scikit-learn==1.3.0 scipy==1.11.2 seaborn==0.12.2
six==1.16.0 SQLAlchemy==2.0.20 threadpoolctl==3.2.0 typing_extensions==4.7.1
tzdata==2023.3 urllib3==2.0.4 Werkzeug==2.3.7
```

6.2 Partie Frontend (Interface Utilisateur) :

6.2.1 Conception de l'Interface Utilisateur :

Pour la conception de l'interface utilisateur, notre équipe a utilisé l'outil de conception Figma pour créer les frameworks visuels initiaux.



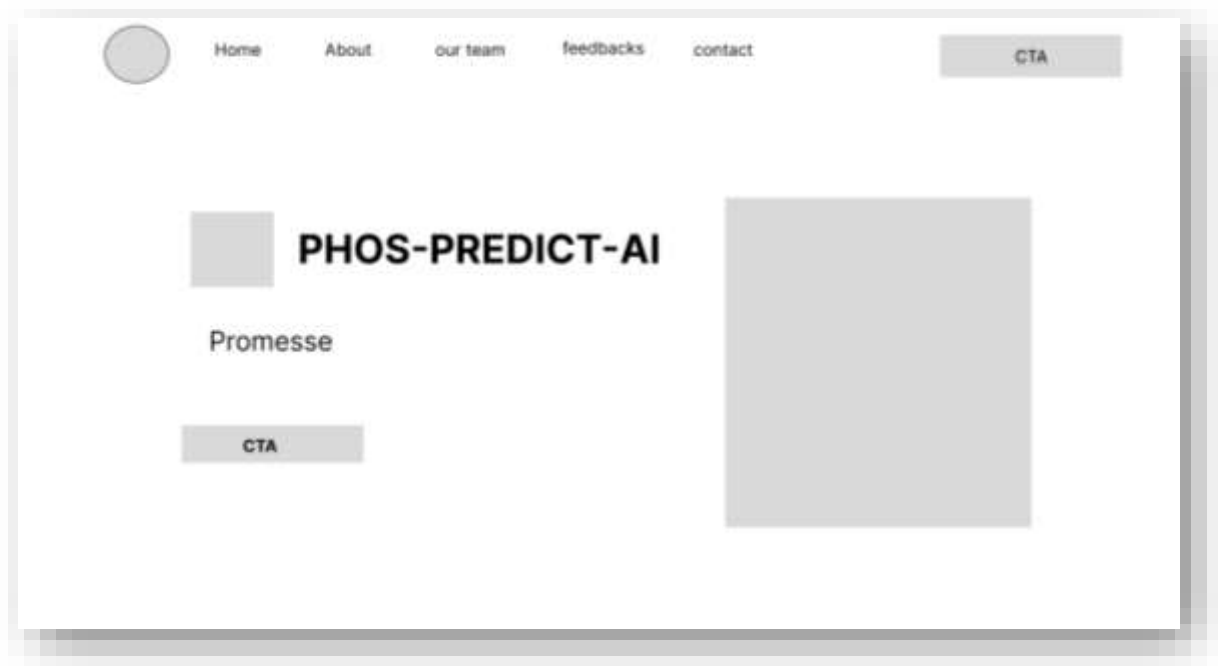
Logo du logiciel Figma

Figma nous a permis de concevoir chaque aspect de l'interface de manière détaillée avant de passer à l'implémentation.

Voici comment nous avons abordé différents aspects de la conception :

a. Disposition :

Nous avons soigneusement planifié la disposition de l'interface en plaçant les éléments de manière logique et intuitive. Nous avons divisé l'interface en sections distinctes pour accueillir les différents éléments, tels que les champs de saisie des paramètres, les boutons de soumission et les zones d'affichage des résultats.



Maquette du Template 'home' créée en utilisant Figma

b. Couleurs :

Nous avons opté pour une palette de couleurs inspirée des nuances de vert et de bleu. Ces couleurs ont été choisies délibérément pour refléter le contexte de l'industrie tout en offrant une expérience visuelle agréable. Les dégradés de vert et de bleu ont été utilisés pour ajouter de la profondeur et de la clarté aux éléments visuels de l'interface.

Shades of #6bb22c



Tints of #6bb22c



Shades of #677b99



Tints of #677b99



Palette de couleurs utiliser dans notre projet

c. Typographie :

Nous avons sélectionné des polices de caractères appropriées pour l'interface. L'objectif était d'assurer une lisibilité optimale tout en maintenant une apparence moderne et professionnelle. Nous avons veillé à utiliser différentes tailles de police pour hiérarchiser les informations et guider les utilisateurs à travers l'interface.

d. Vue d'Ensemble de l'Interface :

L'interface utilisateur se compose de plusieurs éléments clés pour faciliter l'interaction des utilisateurs avec l'application :

- **Page d'Accueil :**

La page d'accueil présente une vue d'ensemble de l'application. Elle explique brièvement son objectif et les avantages de son utilisation. Nous avons utilisé des éléments visuels pour rendre cette page engageante et informative .

- **Formulaire de Saisie des Paramètres :**

Au cœur de l'application se trouve un formulaire interactif où les utilisateurs peuvent saisir les paramètres pertinents. Nous avons conçu ce formulaire pour être convivial et facile à comprendre, en utilisant des libellés clairs et des indicateurs visuels pour guider les utilisateurs.

- **Affichage des Résultats :**

Une fois que les utilisateurs soumettent les paramètres, les résultats des prédictions sont affichés de manière claire et concise. Nous avons utilisé des éléments visuels tels que des graphiques et des tableaux pour présenter les données de manière engageante.

- **Navigation Intuitive :**

Nous avons mis en place une navigation intuitive avec des boutons bien placés pour guider les utilisateurs à travers les différentes étapes. Cela permet aux utilisateurs de passer facilement du formulaire de saisie aux résultats de prédiction.

Conclusion :

En utilisant Figma, nous avons pu visualiser et affiner chaque élément de l'interface avant de passer à la phase d'implémentation. Cela nous a permis de garantir une expérience utilisateur cohérente et convaincante, en mettant en valeur les dégradés de vert et de bleu pour une esthétique visuelle harmonieuse.

6.2.2 Implémentation de l'Interface Utilisateur :

Après avoir finalisé la conception initiale dans **Figma**, notre équipe a mis en œuvre l'interface utilisateur en utilisant une combinaison de **CSS** personnalisé, **Bootstrap**, **HTML** et **JavaScript** (avec **jQuery**) pour obtenir une expérience utilisateur riche et interactive.

Voici comment nous avons abordé la mise en œuvre :



Les langages utilisés pour la création du UI/UX du site.

a. CSS Personnalisé avec Base Bootstrap :

Nous avons commencé avec une base Bootstrap en utilisant un modèle gratuit de <https://startbootstrap.com/>. Nous avons ensuite personnalisé ce modèle de manière approfondie pour répondre aux besoins spécifiques et aux préférences esthétiques de notre projet. Notre CSS personnalisé a été utilisé pour modifier la mise en page, la typographie, les couleurs et d'autres éléments visuels.

b. Structure HTML :

Nous avons utilisé HTML pour structurer chaque composant de l'interface. En travaillant avec des balises HTML sémantiques, nous avons assuré une base solide pour l'affichage et l'interaction des éléments. La structure a été adaptée pour accueillir les modifications que nous avons apportées au modèle.

c. Améliorations JavaScript et jQuery :

JavaScript, avec l'aide de jQuery, a joué un rôle crucial dans l'amélioration de l'interactivité de l'interface. Nous avons écrit des scripts pour gérer la validation des formulaires, la soumission des paramètres et la mise à jour

en temps réel des résultats de prédiction. Ce comportement dynamique a considérablement amélioré l'expérience utilisateur.

d. Création de Logo :

En tant qu'équipe, nous avons conçu un logo unique pour représenter notre projet en utilisant **canva**.



Logiciel utiliser pour la création du logo

Le logo a été élaboré pour résonner avec le thème général de l'application. Il a été intégré avec soin dans l'interface pour créer une expérience de branding cohérente.



PHOS-PREDICT-AI Logo

e. Adaptations et Personnalisations :

Bien que nous ayons commencé avec un modèle, nos modifications et personnalisation approfondies l'ont transformé en une interface unique qui

correspond parfaitement aux objectifs et à l'identité de notre projet.

Cela incluait des modifications des schémas de couleurs, des dispositions de mise en page, de la disposition du contenu et des fonctionnalités supplémentaires.

- **Utilisez la méthode objets disponibles dans le Template :**

Pour que le style de page soit visible dans notre application Flask, nous avons utilisé la fonction `url_for()` pour générer correctement les URL des fichiers statiques tels que les feuilles de style CSS. Cela garantit que les navigateurs peuvent localiser les fichiers statiques et appliquer le style à nos pages web.

La méthode `url_for()` fait partie des objets disponibles dans le template.

Cette methode **génère une URL** en fonction des paramètres transmi

```
<link href="https://fonts.googleapis.com/css2?family=DM+Sans:wght@400;500;700&display=swap" rel="stylesheet">
<link href="{{url_for('static',filename='css/bootstrap-icons.css')}}" rel="stylesheet">
<link href="{{url_for('static', filename='css/bootstrap.min.css')}}" rel="stylesheet">
<link href="{{url_for('static', filename='css/magnific-popup.css')}}" rel="stylesheet">
<link href="{{url_for('static', filename='css/templatemo-first-portfolio-style.css')}}" rel="stylesheet">
```

Les liens pour les feuilles de style

```
<!-- JAVASCRIPT FILES -->
<script src="{{url_for('static', filename='js/jquery.min.js')}}"></script>
<script src="{{url_for('static', filename='js/bootstrap.min.js')}}"></script>
<script src="{{url_for('static', filename='js/jquery.sticky.js')}}"></script>
<script src="{{url_for('static', filename='js/click-scroll.js')}}"></script>
<script src="{{url_for('static', filename='js/jquery.magnific-popup.min.js')}}"></script>
<script src="{{url_for('static', filename='js/magnific-popup-options.js')}}"></script>
<script src="{{url_for('static', filename='js/custom.js')}}"></script>
```

Les liens pour les feuilles de JS

- **Utilisez l'héritage :**

Dans notre application Flask, nous avons adopté une approche basée sur l'héritage pour faciliter la création de modèles de page réutilisables. Nous avons créé un modèle de base, généralement appelé "base.html", qui contient la structure de base de la page, le code HTML commun à toutes les pages, et les liens vers les fichiers CSS et JavaScript nécessaires.

```
templates > <> base.html > ...
1  <!doctype html>
2  <html lang="en">
3      <head>
4          <meta charset="utf-8">
5          <meta name="viewport" content="width=device-width, initial-scale=1">
6          <meta name="description" content="">
7          <meta name="author" content="TemplateMo">
8
9      <title>{% block title %} {% endblock %}</title>
```

Contenu du modèle de base « base.html »

En utilisant l'héritage, nous créons des modèles enfants pour chaque page spécifique en étendant le modèle de base. Dans ces modèles enfants, nous pouvons définir les blocs spécifiques à chaque page, tels que le contenu unique de la page. Cela nous permet de réutiliser le même en-tête, pied de page et mise en page générale sur toutes les pages, tout en personnalisant le contenu de chaque page individuellement.

L'utilisation de l'héritage de templates facilite la maintenance et la mise à jour de l'apparence globale de notre application, car nous pouvons apporter des modifications au modèle de base et voir ces modifications reflétées sur toutes les pages qui en héritent.

```

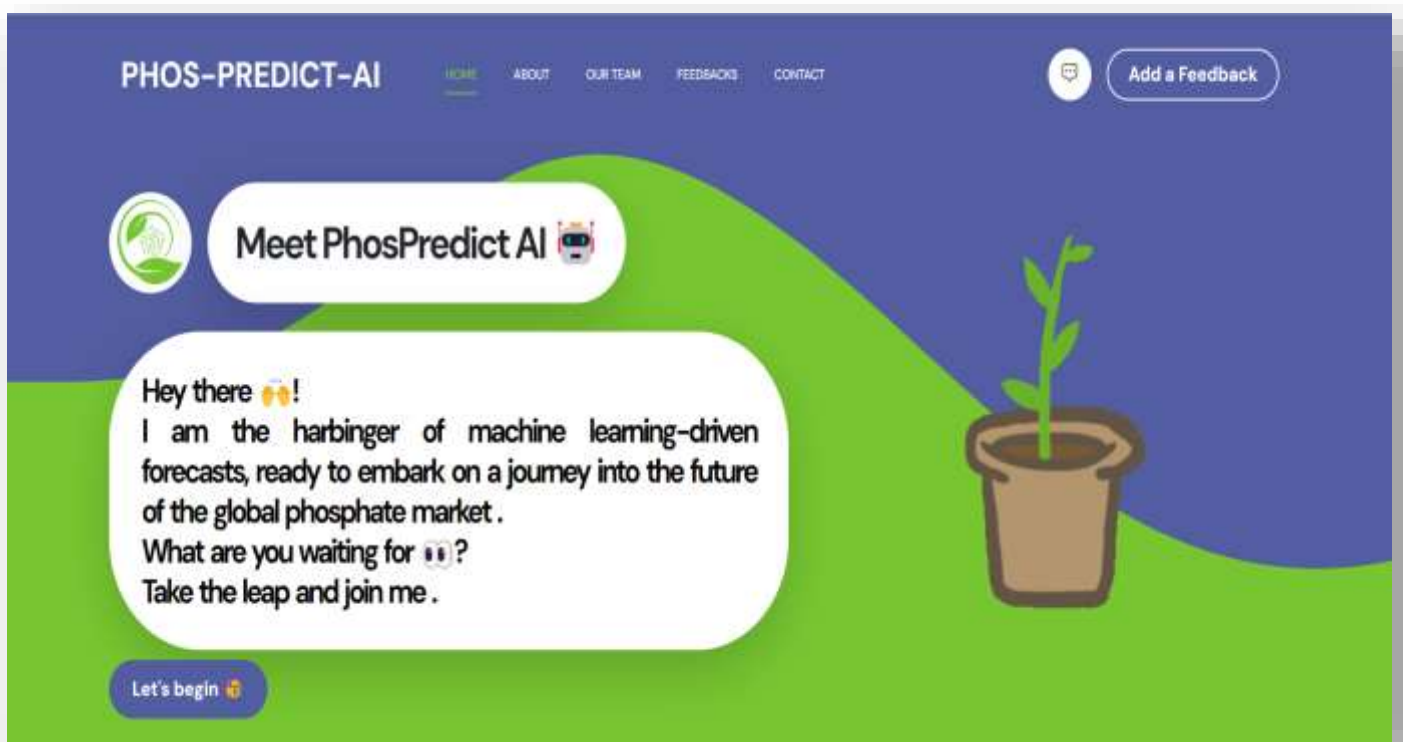
home.html > X
templates > home.html > main > section#section_1.hero.d-flex.justify-
1  {% extends 'base.html' %}
2
3
4  {% block title %} PHOS_PREDICT_AI_APP {% endblock %}
5
6  {% block content %}
7

```

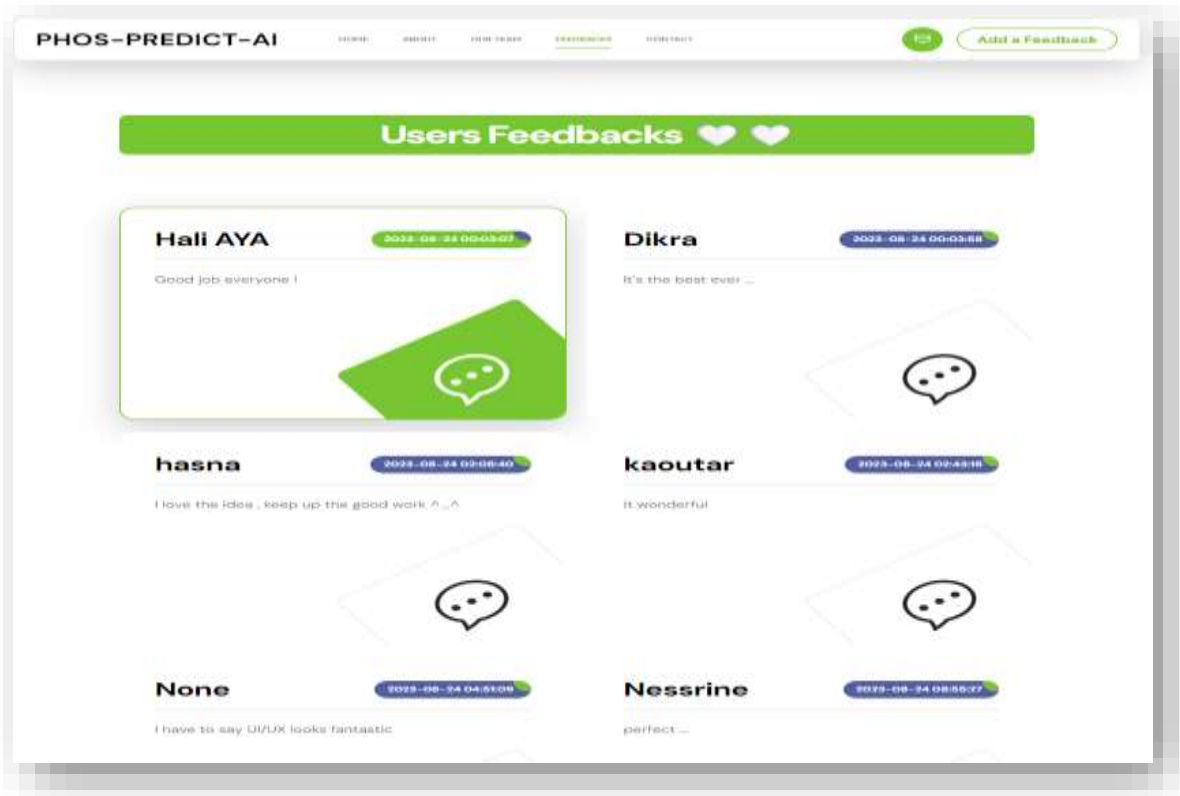
Le contenu de page « home.html » qui hérite de modèle de base

6.3 Le résultat final :

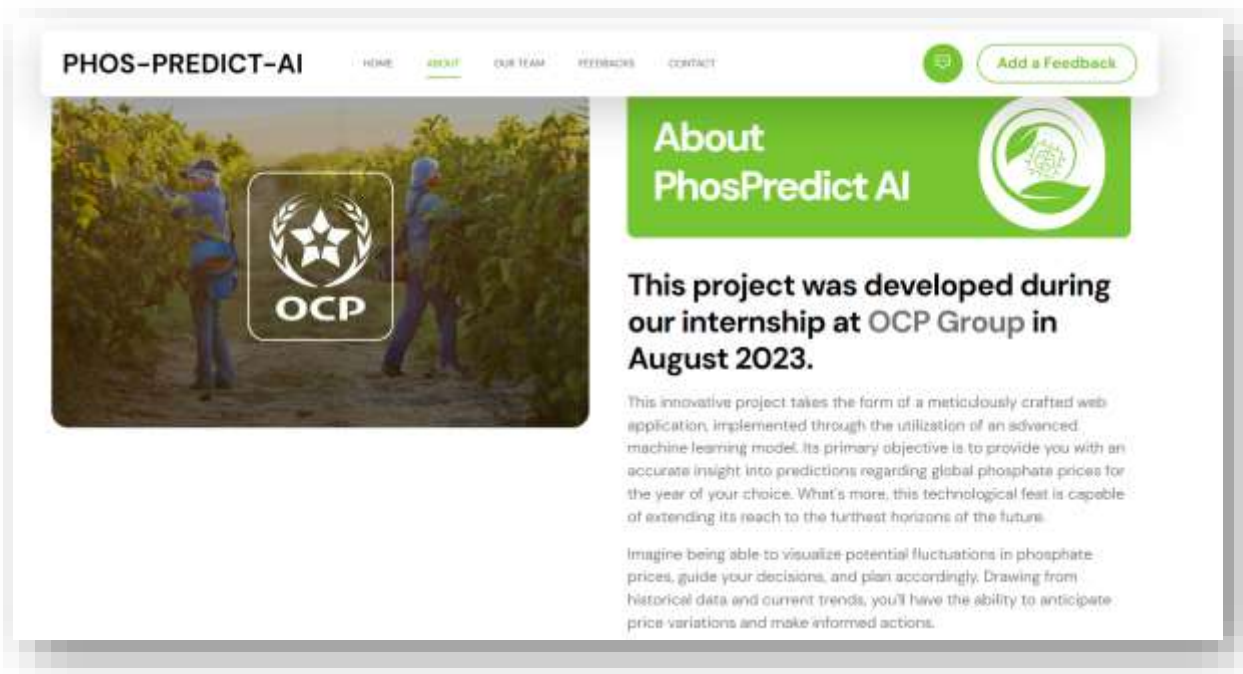
6.3.1 Home page :



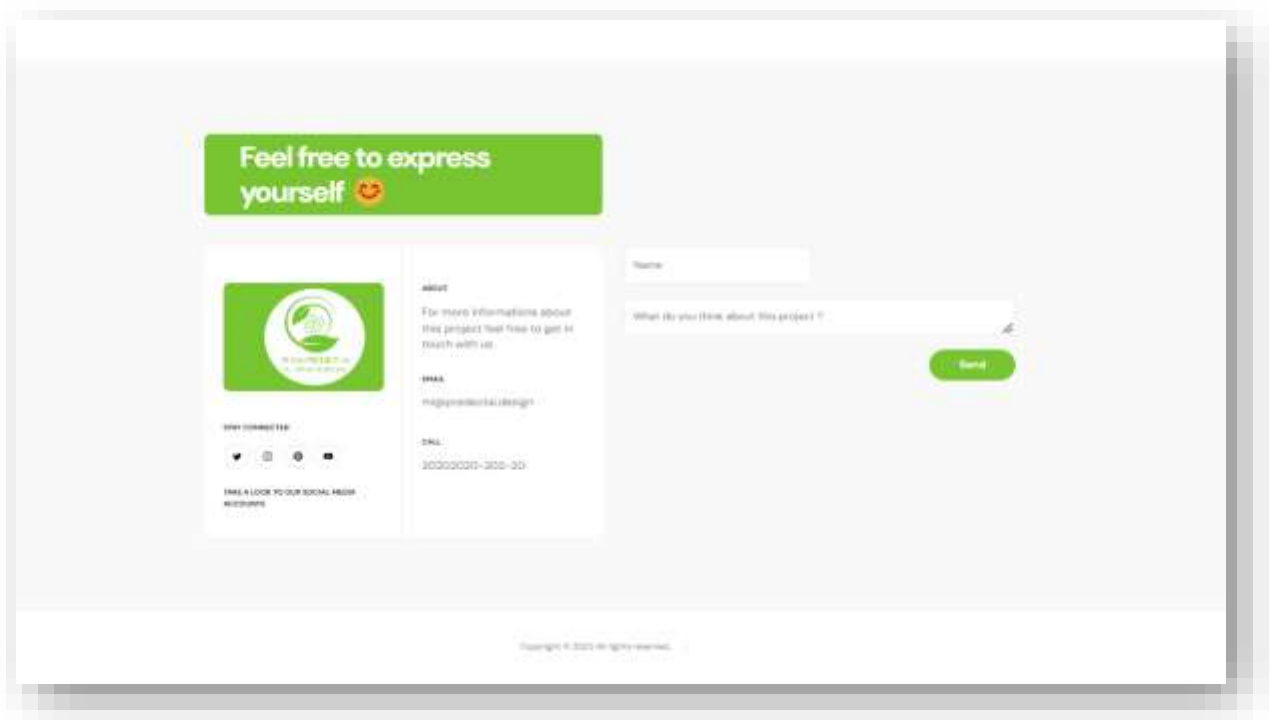
Aperçu de la section "Preloader" de la page d'accueil.



Aperçu de la section "Feedbacks" de la page d'accueil.



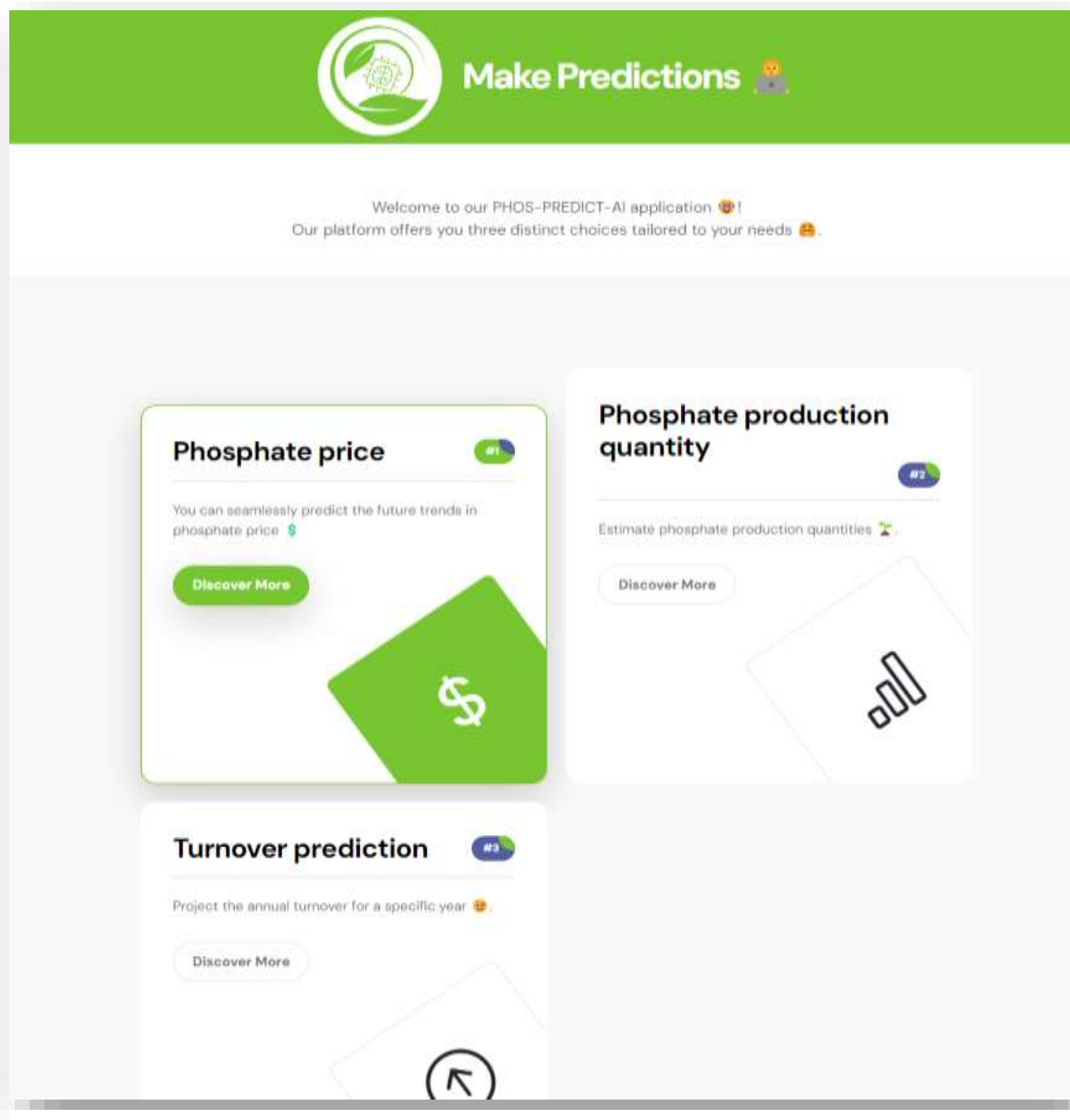
Aperçu de la section "About" de la page d'accueil.



Aperçu de la section "Footer" de la page d'accueil.


6.3.2 Page de choix de valeur à prédire :

Pour choisir entre prédictions de prix de phosphate ou quantité de production ou bien avoir un rapport annuelle :



Aperçu de la page "Predictions".

6.3.3 Les Formulaires :




Predict phosphate price 💰

✦ ✦ Please provide the required input to predict the phosphate price.
 ✓ Fill in the necessary information in the form to proceed with the prediction.

Phosphate Price (US Dollars \$) :

Enter a Year:	Enter a Month: January
Enter the Diesel Price (Dollars US per gallon):	Enter the Diesel ROC :
Enter the Phosphate/Diesel Price Ratio :	Go ✓

Aperçu de la page "result_price.html".



Predict phosphate production quantity 🇫🇷

✦ ✦ Please provide the required input to predict the phosphate production quantity.
 ✓ Fill in the necessary information in the form to proceed with the prediction.

Phosphate production quantity 🌱 :

Select a country : Afghanistan	Enter a Year:
Enter Nutrient potash production quantity (tonnes):	Enter Nutrient nitrogen production quantity (tonnes):
Go ✓	

Copyright © 2023 All rights reserved.

Aperçu de la page "result_quantity.html".

Predict price & production quantity of phosphate

⚠️ Please provide the required input to predict the phosphate price.
 ✅ Fill in the necessary information in the form to proceed with the prediction.

Phosphate production quantity & price :

Select a country : Afghanistan	Enter a year :
Enter Nutrient (potash) production quantity (tonnes) :	Enter Nutrient (nitrogen) production quantity (tonnes) :
Enter a Month : January	Enter the Diesel Price ((liters 120 per gallon) :
Enter the Fertilizer POC :	Enter the Phosphate/Diesel Price Ratio :

Go

Copyright © 2023 All rights reserved.

Aperçu de la page "result_turnover.html".

6.3.4 Page des résultats & Téléchargement de rapport :

Predicted phosphate production quantity

Here are the results based on your inputs, check it out !
 ⚠️ If you want the results for the entire year, please download the report below !

Predicted Phosphate Production quantity for Turkey in 2025 (tonnes) :

8100.0

Predicted Phosphate Price for April in \$:

372.3392857142857

Download PDF Report

Copyright © 2023 All rights reserved.

Aperçu de la page "predictedTurnover.html".

6.3.5 Conclusion :

En combinant le modèle Bootstrap comme base avec notre CSS personnalisé, la structuration HTML et les améliorations JavaScript, nous avons créé une interface qui marie harmonieusement l'attrait esthétique à l'efficacité fonctionnelle. Le résultat est une interface engageante et conviviale qui reflète à la fois la structure du modèle original et les contributions créatives de notre équipe.

6.4 Partie backend :

Le développement backend de notre projet a été une étape cruciale pour la création d'une application web fonctionnelle et performante à l'aide du Framework de **python Flask**. Cette partie du projet a été responsable de la mise en œuvre de la logique métier, de la gestion des données, de la sécurisation des interactions et de l'établissement d'une communication fluide entre le frontend et la base de données.



Figure 1 : Logo du langage python

6.4.1 Configuration du Projet Flask



Figure 2 : Logo du Framework "flask"

Nous avons choisi le framework Flask en raison de son approche légère et flexible du développement web. Notre application Flask suit le modèle MVT (Modèle-Vue-Template), qui nous a permis d'organiser notre code en différents composants pour une maintenance et une évolutivité simplifiées.

Le Contrôleur, ou les "vues" dans le contexte de Flask, a joué un rôle crucial dans la gestion des requêtes HTTP entrantes. Nous avons défini des vues pour chaque fonctionnalité, en indiquant les méthodes HTTP correspondantes (GET, POST, etc.).

Ces vues ont ensuite dirigé les requêtes vers les fonctions appropriées dans les vues.

a. Interaction avec la Base de Données en utilisant un ORM (*Object Relational Mapping*) :



Figure 3 : Logo du ORM SQLAlchemy

Pour faciliter la communication avec la base de données, nous avons fait usage de Flask-SQLAlchemy.

```

db=SQLAlchemy(app)

class feedback(db.Model):
    __id = db.Column("id", db.Integer, primary_key=True)
    name = db.Column(db.String(100))
    message = db.Column(db.String(100))
    date = db.Column(db.DateTime, default=datetime.utcnow)

    def __init__(self, name, message):
        self.name = name
        self.message = message
        self.date = datetime.now()

```

Figure 4: Exemple d'un modèle feedback dans la BD



Figure 5: Logo du SQLite

Bien que **SQLite** soit le **SGBD** utilisé dans notre projet, **SQLAlchemy** nous a fourni une couche d'abstraction qui nous a permis de travailler avec différents SGBD de manière transparente.

Les modèles **SQLAlchemy** ont été utilisés pour représenter la structure des tables de la base de données ainsi que les relations entre elles. Ces modèles ont été conçus en fonction de nos besoins pour stocker et extraire les données.

Nous avons implémenté les opérations **CRUD** (Create, Read, Update, Delete) pour interagir avec la base de données. Cela nous a permis d'ajouter de nouvelles données, de récupérer des données existantes, de les mettre à jour et de les supprimer de manière cohérente.

Gestion des Données Utilisateur avec des Sessions

Lors de la mise en place du développement backend de notre application, nous avons opté pour l'utilisation de sessions pour gérer les données utilisateur de manière sécurisée et efficace. Les sessions sont un mécanisme essentiel pour conserver et suivre l'état des interactions entre les utilisateurs et l'application, offrant ainsi une expérience utilisateur cohérente.

6.4.2 Création et Gestion des Sessions

Lorsqu'un utilisateur accède à notre application, une session est créée pour lui. Cette session est stockée côté serveur et associée à un identifiant unique stocké dans un cookie sur le navigateur de l'utilisateur.

Cela permet à l'application de reconnaître l'utilisateur lors de ses interactions ultérieures.

Nous avons utilisé la bibliothèque Flask-Session pour gérer ces sessions de manière transparente. Cette bibliothèque nous a permis de stocker et de gérer les données spécifiques à chaque utilisateur de manière sécurisée.

6.4.3 Stockage et Récupération des Données Utilisateur

L'une des principales utilisations des sessions était la possibilité de stocker temporairement les données utilisateurs spécifiques pour chaque session.

Par exemple, après avoir entré des informations dans un formulaire, les données étaient stockées dans la session, ce qui nous permettait de les récupérer et de les utiliser ultérieurement, par exemple pour générer des rapports personnalisés.

Cela nous a également permis de suivre l'état de l'utilisateur à mesure qu'il naviguait à travers différentes parties de l'application, en conservant les informations dont nous avons besoin pour fournir des fonctionnalités cohérentes et personnalisées.

```
# Create data for the table
table_data = [['Month', 'Predicted Price']]

# Define month names as strings
month_names = ['January', 'February', 'March', 'April', 'May', 'Ju

for month, values in session.values.items():
    predicted_price = f"{values['predicted_price']:.6f}$"
    month_str = month_names[month - 1] # Convert month to string
    table_data.append([month_str, predicted_price])
```

Figure 6: Exemple d'utilisation de session

a- Passez des variables au Template :

Dans notre application Flask, nous avons utilisé la fonction `render_template()` pour passer des variables depuis notre code Python vers les templates HTML. Cela nous permet d'afficher dynamiquement des données sur nos pages web en utilisant les valeurs que nous avons calculées ou récupérées dans notre application.

Lorsque nous utilisons **render_template()**, nous pouvons passer des variables en tant qu'arguments dans la fonction.

Par exemple :

```
@app.route("/submit_feedback",methods=["POST","GET"])
def submit_feedback():
    if request.method=="POST":
        session.permanent=True
        nm=request.form["name"]
        msg=request.form["message"]
        #store the feedback on the database
        feedbk = feedback(name=nm,message=msg)
        db.session.add(feedbk)
        db.session.commit()

    #display all of the feedbacks stored in database
    Feedbacks = feedback.query.all()
    return render_template('home.html', feedbacks=Feedbacks)
```

Figure 7: Utilisation de render_template()

Dans cet exemple, nous avons passé la variable feedbacks à notre template "home.html".

b- Pour accéder aux variables passées depuis notre code Python :

À l'intérieur du template, nous pouvons accéder à cette variable en utilisant des balises **Jinja2**, comme ceci :

```

{% for feedback in feedbacks %}
  <div class="col-lg-6 col-12">
    <div class="services-thumb">
      <div class="d-flex flex-wrap align-items-center border-bottom mb-4 pb-3">
        <h3 class="mb-0">{{ feedback.name }} </h3>

        <div class="services-price-wrap ms-auto">
          <p class="services-price-text mb-0">{{ feedback.date.strftime('%Y-%m-%d %H:%M:%S') }}</p>
          <div class="services-price-overlay"></div>
        </div>
      </div>

      <p>{{ feedback.message }}</p>

      <div class="services-icon-wrap d-flex justify-content-center align-items-center">
        <i class="services-icon bi-chat-dots"></i>
      </div>
    </div>
  </div>
{% endfor %}

```

Figure 8: Utilisation des balises Jinja2

Lorsque nous accédons à la page générée par notre application, la balise Jinja2 `{{ feedback.name }}` sera remplacée par la valeur de la variable `name` stocker dans la base de données que nous avons passée depuis notre code Python.

Par conséquent, dans cet exemple, nous verrons section des feedbacks affichée sur la page home . :

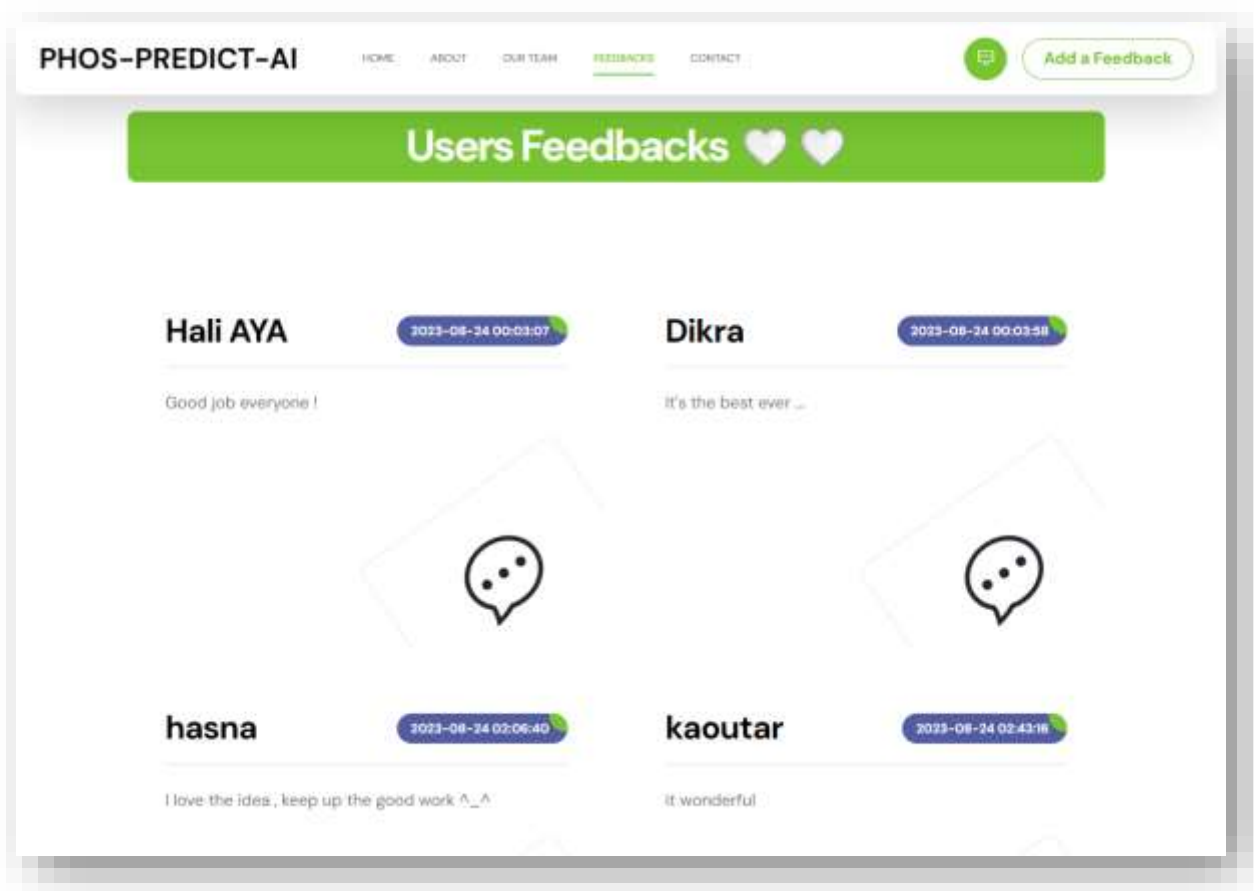


Figure 9: Affichage des variables en utilisant Jinja2

6.4.4 Génération de Rapports avec ReportLab

Pour offrir à nos utilisateurs la possibilité de télécharger des rapports détaillés, nous avons intégré la bibliothèque **ReportLab**.



Figure 10: Logo du bibliothèque ReportLab

Cette bibliothèque nous a permis de créer des rapports au format PDF de manière programmatique.

```
# generate pdf function :
def generate_pdf(session_values, country_name, year, nutrient_potash):
    buffer = BytesIO()
    doc = SimpleDocTemplate(buffer, pagesize=letter)
    elements = []
    # Add logo
    logo_path = r'static\images\1-removebg-preview.png'
    logo = Image(logo_path, width=300, height=300)
    elements.append(logo)
    # Add space below the logo
    elements.append(Spacer(1, 20))
    # Add title
    title = f"Annual Report of {country_name} - Year {year}"
    elements.append(Paragraph(title, getSampleStyleSheet()['Title']))
    elements.append(PageBreak()) # Move to the next page
    # Add nutrient quantities
    # Center the following elements and add more spacing
    style_center = getSampleStyleSheet()['BodyText']
    style_center.alignment = 1 # Center alignment
    date = f>Date: {datetime.now().strftime('%Y-%m-%d %H:%M:%S')}"
    # Add the date with centered style
    style_center = getSampleStyleSheet()['Normal']
    style_center.alignment = 1 # 1 indicates centered alignment

    # Create a new Paragraph with "Date:" in bold
    date_text = "<b>Date:</b> " + date
```

Figure 11: Utilisation de bibliothèque ReportLab dans la fonction "generate_pdf"

Nous avons développé une fonctionnalité qui génère un rapport PDF contenant les résultats de l'année entière. Les données sont formatées et présentées de manière claire dans le rapport. Les utilisateurs peuvent télécharger ce rapport pour une consultation ultérieure.

Exemple d'un rapport :

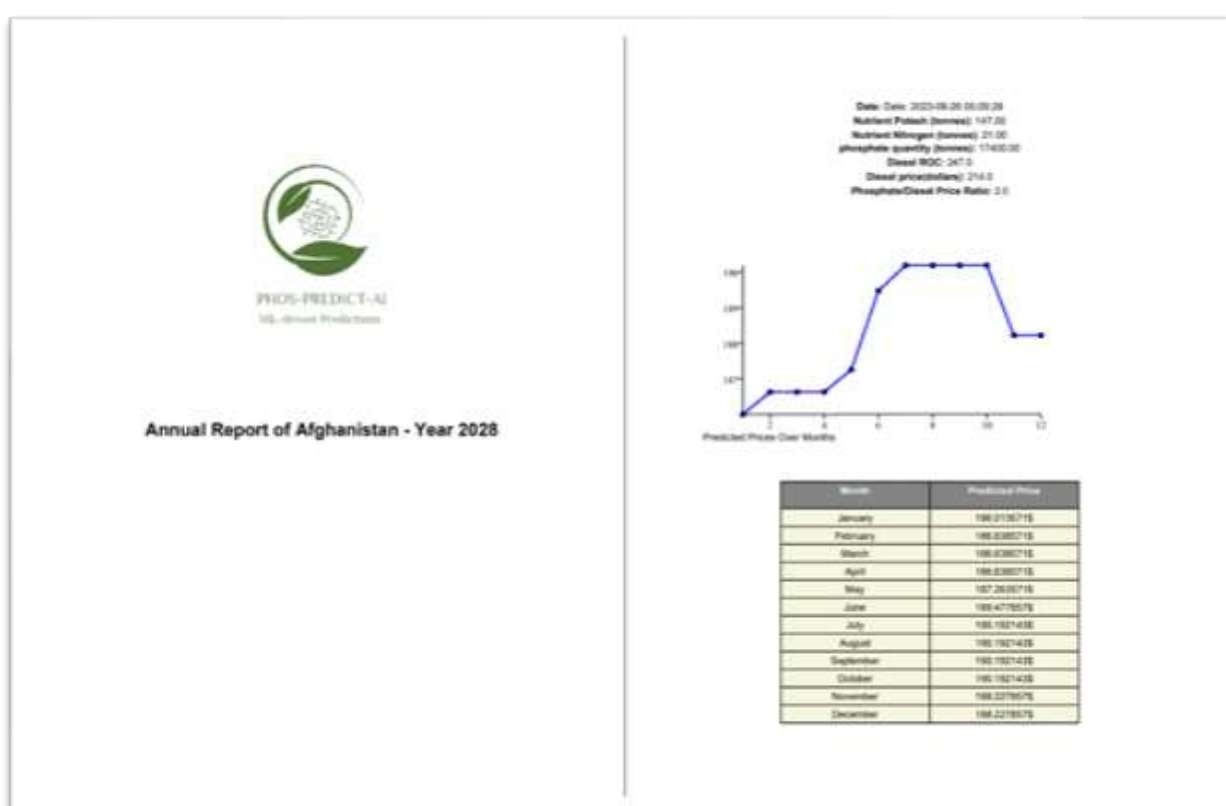


Figure 12: Exemple d'un rapport crée en utilisant ReportLab

6.4.5 Conclusion

En conclusion, le développement backend de notre application avec Flask a été essentiel pour fournir des fonctionnalités robustes, des interactions sécurisées et des performances optimales. Nous avons réussi à créer une architecture bien structurée, à gérer les données avec précision et à garantir la sécurité des utilisateurs. Cette partie du projet a jeté les bases d'une application web réussie, en établissant une connexion harmonieuse entre le frontend et la base de données.