

Analyse des Données de Fabrication de Ciment

Application des méthodes de découverte de connaissances et d'apprentissage automatique

Étudiants :

Horace Mendel AGOSSOU
Charly Kenneth ANANI
Sylvestre HOUNTON

Enseignants :

Dr OULOULADE
Moctard

Table des matières

1	Introduction	2
1.1	Objectifs	2
2	Données et Prétraitement	2
2.1	Description des données	2
2.2	Étapes de prétraitement	2
2.3	Comparaison avant/après traitement	3
3	Analyse exploratoire	3
3.1	Corrélations significatives	3
3.2	Visualisation des distributions	3
4	Modélisation prédictive	3
4.1	Algorithmes utilisés	3
4.2	Performances des algorithmes	4
5	Conclusion et recommandations	4
5.1	Principales découvertes	4
5.2	Recommandations	4
6	Annexes	4

1 Introduction

L'industrie du ciment est un pilier essentiel du développement infrastructurel mondial. Cependant, la complexité de ses processus de fabrication requiert une optimisation constante pour garantir une qualité de produit uniforme et une efficacité énergétique maximale. Cette étude approfondie des données de fabrication de ciment met en lumière l'application de méthodes avancées de découverte de connaissances, de cloud computing et d'exploration de données.

1.1 Objectifs

Nos objectifs principaux sont :

- Maîtriser les techniques d'apprentissage automatique pour le nettoyage, l'analyse et l'interprétation des données.
- Élaborer des recommandations pratiques basées sur les résultats obtenus, afin d'améliorer la prise de décision dans ce secteur industriel.

2 Données et Prétraitement

2.1 Description des données

Notre base de données comprend 74 lignes et 44 colonnes, englobant diverses variables du processus de fabrication du ciment. La majorité des colonnes sont de type `float`, à l'exception de `test_eligibilite` qui est de type `object`. Nous avons identifié 347 valeurs manquantes dans le DataFrame.

2.2 Étapes de prétraitement

- **Gestion des valeurs manquantes** : Identification et traitement pour prévenir toute distorsion d'analyse.
- **Suppression des colonnes inutiles** : Élimination des colonnes à valeurs constantes ou peu informatives.
- **Transformation des variables** : Standardisation des données pour assurer une contribution égale de chaque variable à l'analyse.

2.3 Comparaison avant/après traitement

Statistique	Avant Traitement	Après Traitement
Comptage Valeurs	72 valeurs	72 valeurs (aucune perte)
Moyenne	Élevée (ex : 8.07×10^7)	Proche de 0 (ex : 4.93×10^{-17})
Écart-type	Variable (ex : 8.62×10^7)	Uniforme (≈ 1.007)
Min/Max	Large plage (valeurs aberrantes)	Réduite, centrée autour de 0
Quantiles	Déséquilibrés	Équilibrés autour de 0

TABLE 1 – Comparaison des statistiques avant et après traitement

3 Analyse exploratoire

3.1 Corrélations significatives

L'analyse des corrélations a révélé des liens importants :

- **Pression bas préchauffeur 4 / Pression précalcinateur (0.76)** : Une augmentation de la pression dans le précalcinateur semble influencer la pression dans le préchauffeur.
- **Refroidisseurs secondaire/primaire (0.67)** : Les performances des deux refroidisseurs sont interconnectées.
- **Ventilateur tertiaire / Refroidisseur secondaire (0.67)** : La pression du ventilateur affecte le refroidissement secondaire.

3.2 Visualisation des distributions

Les graphiques de distribution ont mis en évidence :

- **Échelle d'alimentation** : Forte concentration autour de 410.
- **Calcium libre** : Asymétrie avec longue traîne à droite.
- **Test d'éligibilité** : Majorité des valeurs autour de 1.
- **Vitesse du four** : Concentration autour de 3.88.
- **Température du corps** : Distribution quasi-normale autour de 370.
- **Pression précalcinateur** : Multimodale, suggérant différents régimes.

4 Modélisation prédictive

4.1 Algorithmes utilisés

Nous avons évalué plusieurs algorithmes pour prédire des variables clés :

- Arbre de Décision
- SVR (Support Vector Regression)
- Régression Polynomiale
- Régression Linéaire
- CNN (Convolutional Neural Networks)

4.2 Performances des algorithmes

Variable Prédite	Meilleur Algorithme	R ² Test
Test d'éligibilité	Arbre de Décision	0.9982
Temp. préchauffeur 1	Arbre de Décision	0.9780
Calcium libre	SVR	0.9406

TABLE 2 – Performances des algorithmes sur les ensembles de test

5 Conclusion et recommandations

5.1 Principales découvertes

Cette étude a permis de :

- Mettre en lumière des corrélations importantes entre les variables du processus.
- Développer des modèles prédictifs performants pour anticiper certaines variables clés.
- Visualiser les distributions de données pour mieux comprendre les tendances et caractéristiques.

5.2 Recommandations

- **Contrôle en temps réel** : Surveillance continue des pressions, températures et autres variables critiques pour détecter les anomalies et ajuster les paramètres du processus.
- **Optimisation des refroidisseurs** : Gestion synchronisée des refroidisseurs pour maximiser l'efficacité et la qualité du clinker.
- **Outils d'analyse prédictive** : Intégrer les modèles prédictifs pour anticiper la qualité et ajuster les paramètres en conséquence.
- **Recherche et formation** : Investir dans la recherche pour améliorer les modèles et former le personnel à leur utilisation.

6 Annexes

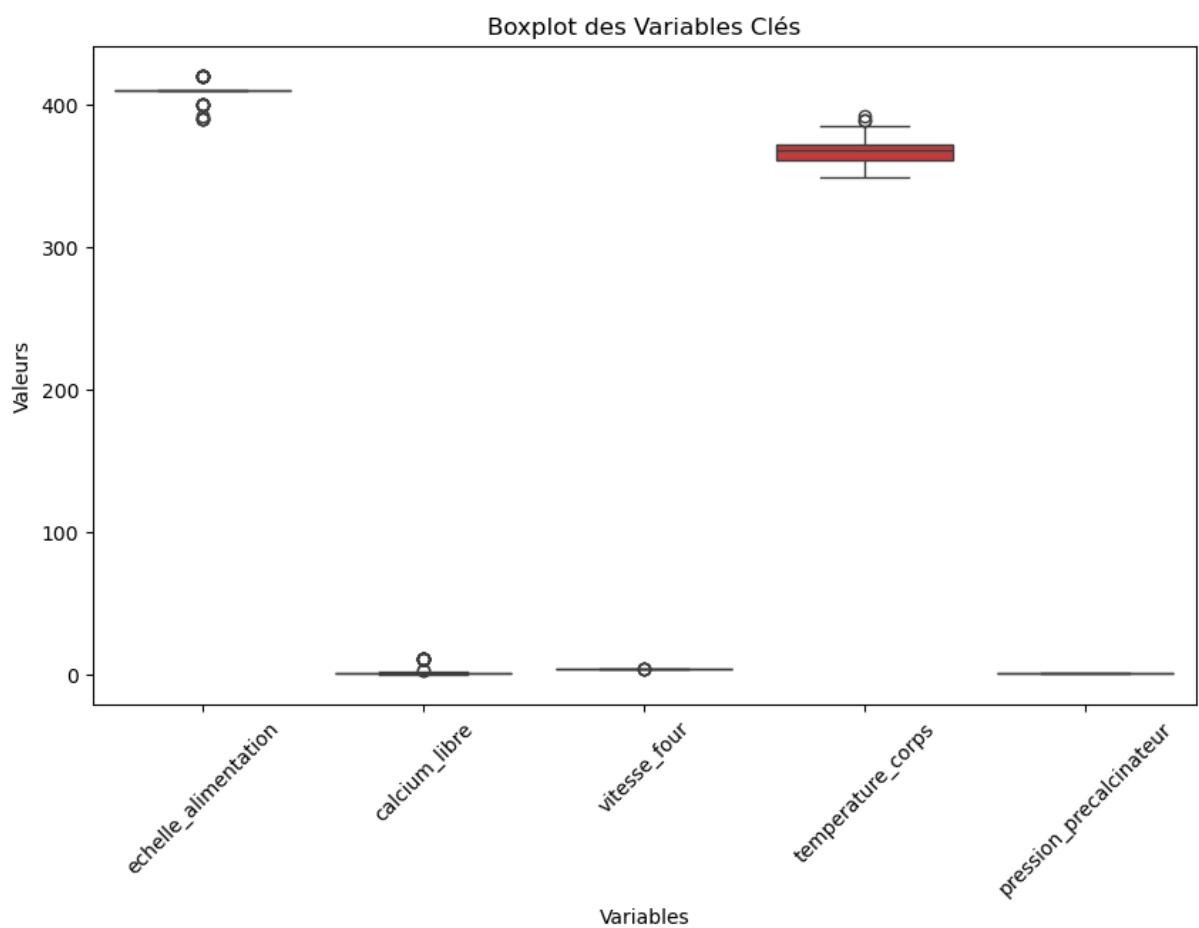


FIGURE 1 – Boxplot des Variables Clés

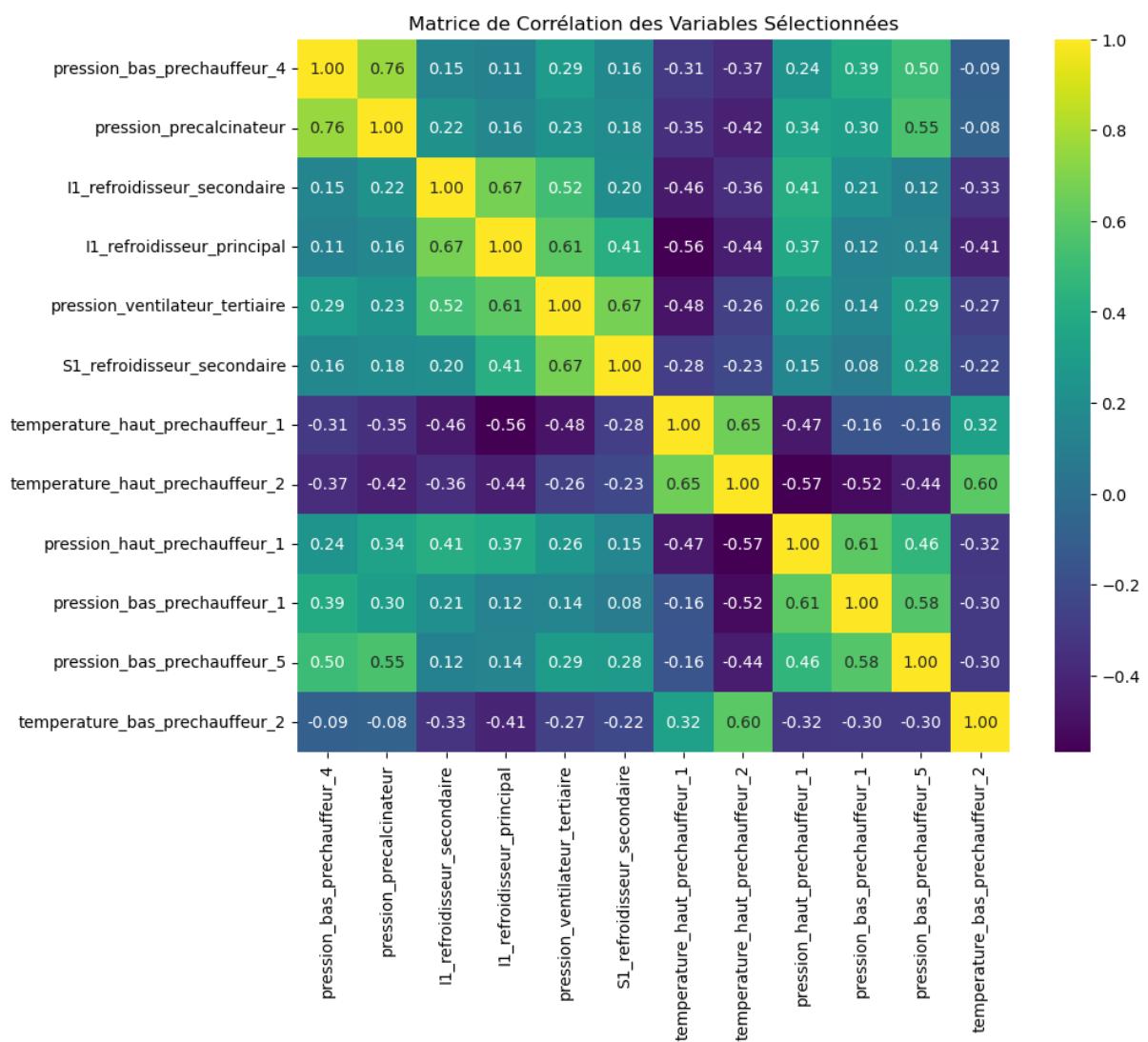


FIGURE 2 – Matrice de Corrélation des Variables Sélectionnées

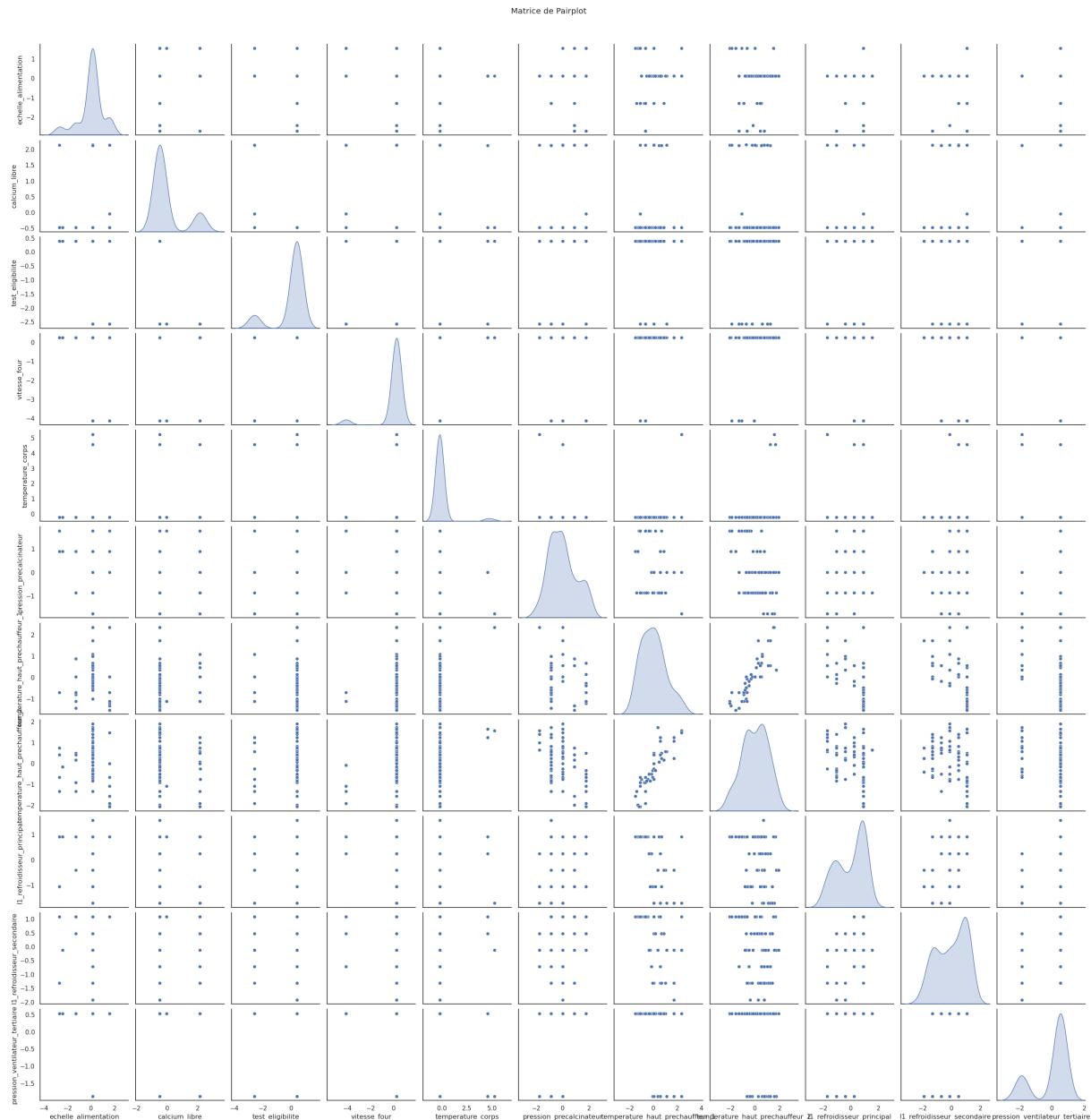


FIGURE 3 – Graphiques de distribution

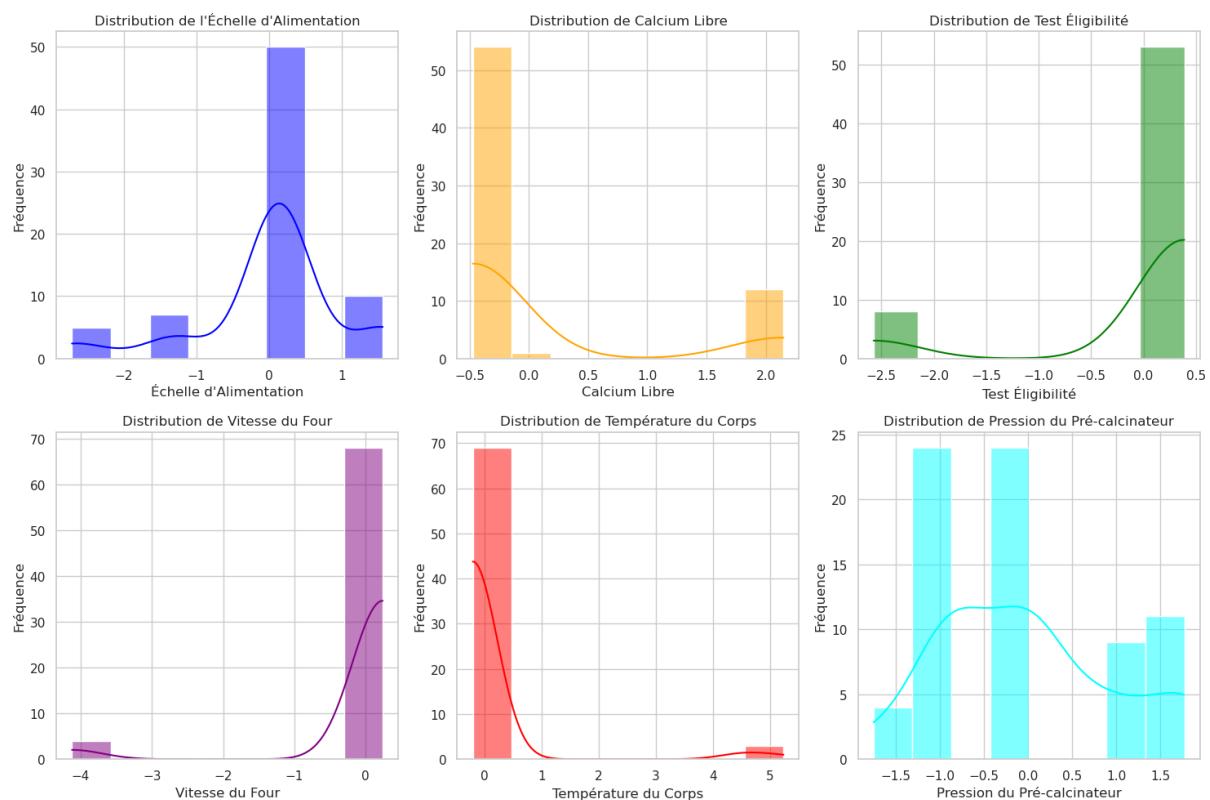


FIGURE 4 – Enter Caption

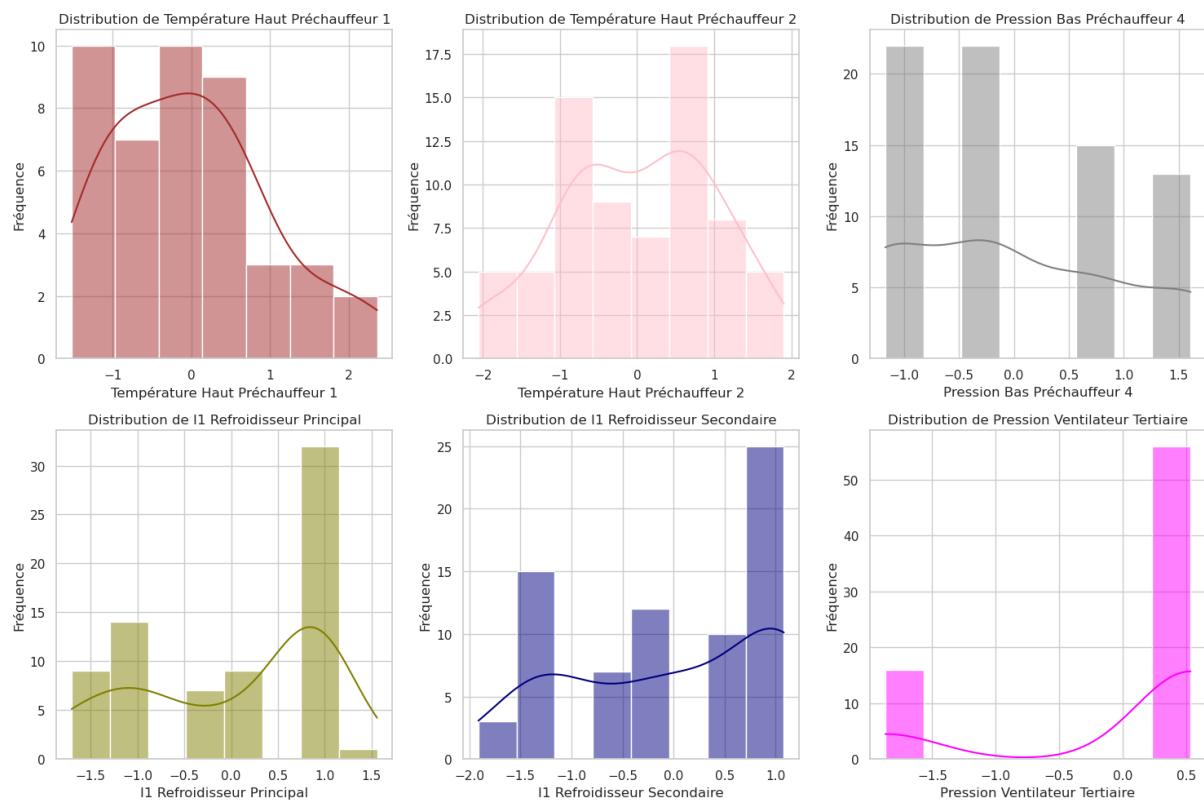


FIGURE 5 – Graphiques de distribution suite