

# RAPPORT D'ANALYSE

Méthodologie, analyses, choix de modélisation



**ID&A TECH**  
Intelligence Décisionnelle & Artificielle  
TECHnologies

Étude de Cas pour Stage Data Analyst/Scientist chez ID&A TECH  
Prédiction de la Performance de l'Indice MASI

Ayoub ETOULLALI



LinkedIn Profile

# TABLE DES MATIERES

|  |          |
|--|----------|
| <b>INTRODUCTION .....</b>                        | <b>2</b> |
| <b>MÉTHODOLOGIE .....</b>                        | <b>3</b> |
| COMPREHENSION DU PROBLEME .....                  | 3        |
| EXPLORATION DES DONNEES .....                    | 3        |
| PRETRAITEMENT DES DONNEES .....                  | 3        |
| CHOIX ET JUSTIFICATION DES MODELES .....         | 3        |
| <b>ANALYSES .....</b>                            | <b>4</b> |
| ANALYSE TEMPORELLE .....                         | 4        |
| CORRELATION ENTRE VARIABLES .....                | 4        |
| EXPLORATION DES VARIABLES MACROECONOMIQUES ..... | 5        |
| <b>CHOIX DE MODÉLISATION .....</b>               | <b>6</b> |
| SELECTION DE MODELE .....                        | 6        |
| ENTRAINEMENT ET TEST DU MODELE .....             | 6        |
| OPTIMISATION ET VALIDATION .....                 | 7        |
| <b>CONCLUSION .....</b>                          | <b>8</b> |

# INTRODUCTION

L'analyse entreprise a pour objectif principal de développer un modèle prédictif robuste permettant d'estimer la performance future de l'indice MASI, reflétant ainsi la dynamique du marché boursier marocain. Dans cette démarche, plusieurs étapes clés ont été entreprises, allant de l'exploration des données à la sélection et l'optimisation de modèles de machine learning.

## Bourse de Casablanca: la performance annuelle a bondi de +18,35% en 2021, portée par un environnement favorable



Entrée du siège de la Bourse de Casablanca. - Le360 - Saad Zouhri

Après une année 2020 dans le rouge à cause des répercussions de la crise sanitaire sur l'activité économique, la Bourse de Casablanca a largement effacé ses pertes au terme de l'année 2021. Les taux bas, la croissance vigoureuse du PIB et la campagne de vaccination ont été les principaux moteurs de cette hausse.

### ARTICLES LES PLUS LUS

- 1 La rage impuissante du régime algérien face au succès commercial retentissant d'El Guerguerat
- 2 Nouvelle décharge de Casablanca: énième rebondissement inattendu, la mairie va revoir son appel d'offres
- 3 À la télévision espagnole, Nicolas Sarkozy réitère son soutien à la marocanité du Sahara
- 4 Coopération sécuritaire: Abdellatif Hammouchi reçoit Frédéric Veaux, le patron de la police nationale française
- 5 Enseignement: malgré l'accord du 10 décembre, une rentrée des classes toujours incertaine
- 6 Médicaments antirhume: forte consommation au Maroc malgré les mises en garde en France
- 7 Hausse des salaires des enseignants: comment le gouvernement entend financer l'accord du 10 décembre
- 8 Allégations de pollution à l'arsenic: le groupe Managem décide de porter

## Compréhension du Problème

La première étape a consisté à acquérir une compréhension approfondie du problème. Cela a impliqué l'identification des données disponibles, des objectifs spécifiques et des éléments cruciaux tels que l'estimation de la performance future et l'identification des facteurs d'influence.

## Exploration des Données

Une analyse exploratoire approfondie des données a été réalisée pour comprendre les tendances, les saisons et les corrélations. Les différentes feuilles de données fournissant des informations, notamment l'historique des prix de l'indice MASI, les indicateurs macroéconomiques et d'autres indices sectoriels, ont été examinées.

## Prétraitement des Données

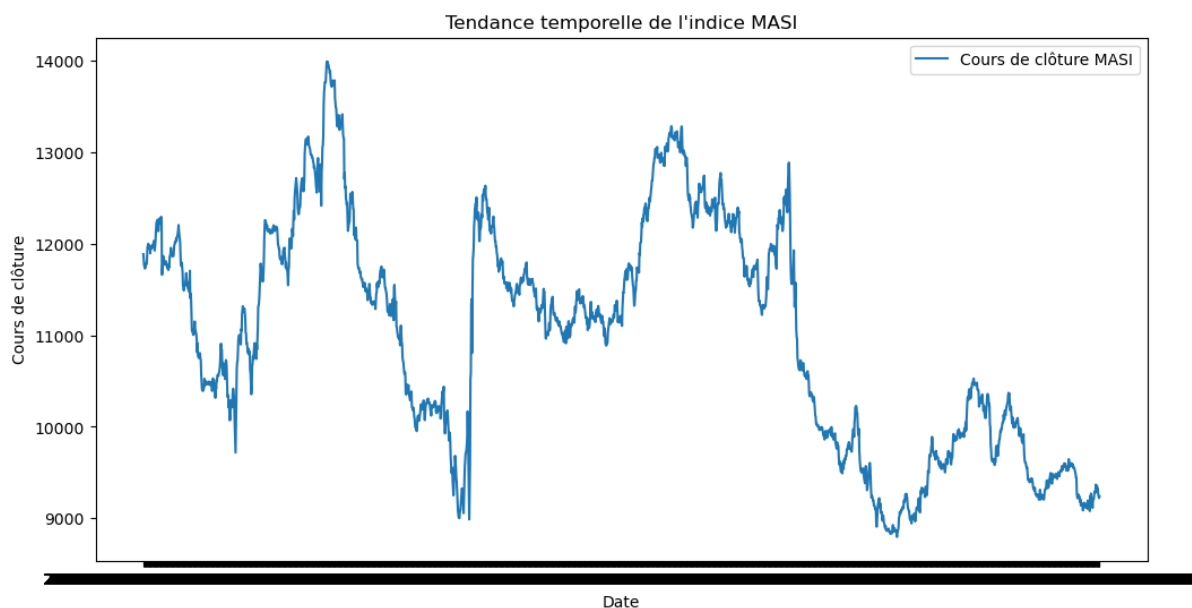
Le prétraitement des données a inclus des étapes telles que le nettoyage des données, la normalisation et la transformation. Les valeurs manquantes ont été gérées, et une normalisation a été appliquée pour mettre les variables à des échelles comparables. De plus, de nouvelles caractéristiques ont été créées afin d'améliorer la performance des modèles.

## Choix et Justification des Modèles

Trois modèles de machine learning ont été sélectionnés pour l'analyse : la régression linéaire, les forêts aléatoires et les réseaux de neurones. Ces choix ont été motivés par la complexité du problème et la diversité des modèles pour fournir des perspectives variées.

## Analyse Temporelle

Une analyse temporelle détaillée a été réalisée sur le cours de clôture de l'indice MASI afin de comprendre les tendances temporelles et les variations au fil du temps. Des visualisations ont été utilisées pour mettre en évidence les mouvements significatifs et les schémas émergents.



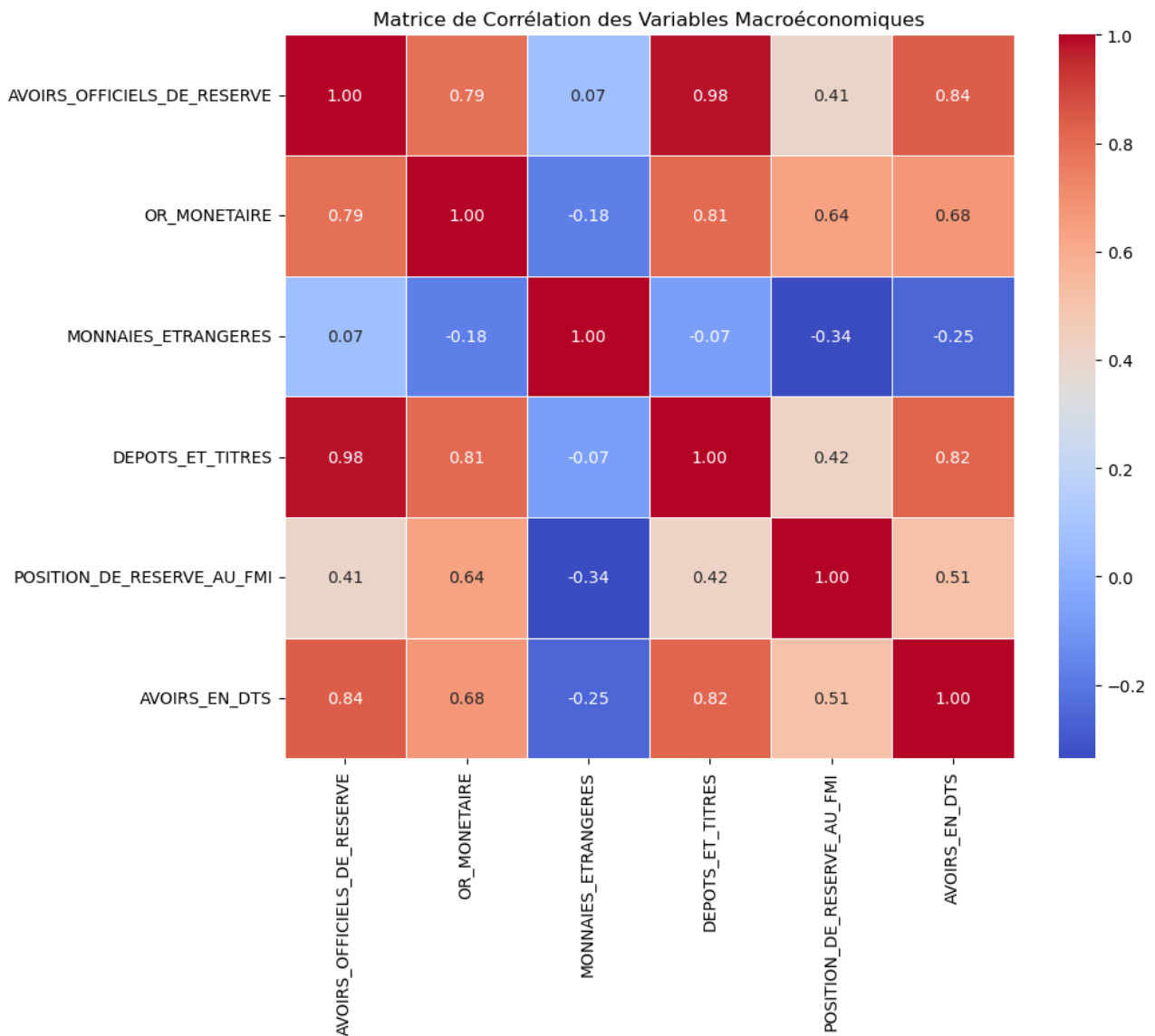
## Corrélation entre Variables

L'examen de la corrélation entre le cours de clôture et d'autres variables a permis de déterminer la force et la direction des relations linéaires. Une matrice de corrélation a été élaborée pour faciliter cette évaluation.

```
Corrélations avec COURS_CLOTURE :
COURS_CLOTURE_NORMALIZED    1.000000
COURS_CLOTURE                1.000000
COURS_PLUS_BAS              0.999193
COURS_PLUS_HAUT             0.998937
COURS_OUVERTURE             0.997963
COURS_VEILLE                0.997940
AVOIRS_EN_DTS               0.087247
VOLUME_JJ                   0.081616
DEPOTS_ET_TITRES            0.046115
AVOIRS_OFFICIELS_DE_RESERVE 0.041746
Volume JJ                   0.040779
VARIATION                   0.012725
OR_MONETAIRE                -0.025241
INFLATION_SOUS_JACENTE      -0.046525
INFLATION                   -0.048934
POSITION_DE_RESERVE_AU_FMI  -0.068523
MONNAIES_ETRANGERES        -0.069976
Cours Maximum               -0.115618
Cours Minimum               -0.115625
ENCOURS                     -0.132115
Indice MONIA                -0.136589
TAUX_MOYEN_PONDERE         -0.194446
Name: COURS_CLOTURE, dtype: float64
```

## Exploration des Variables Macroéconomiques

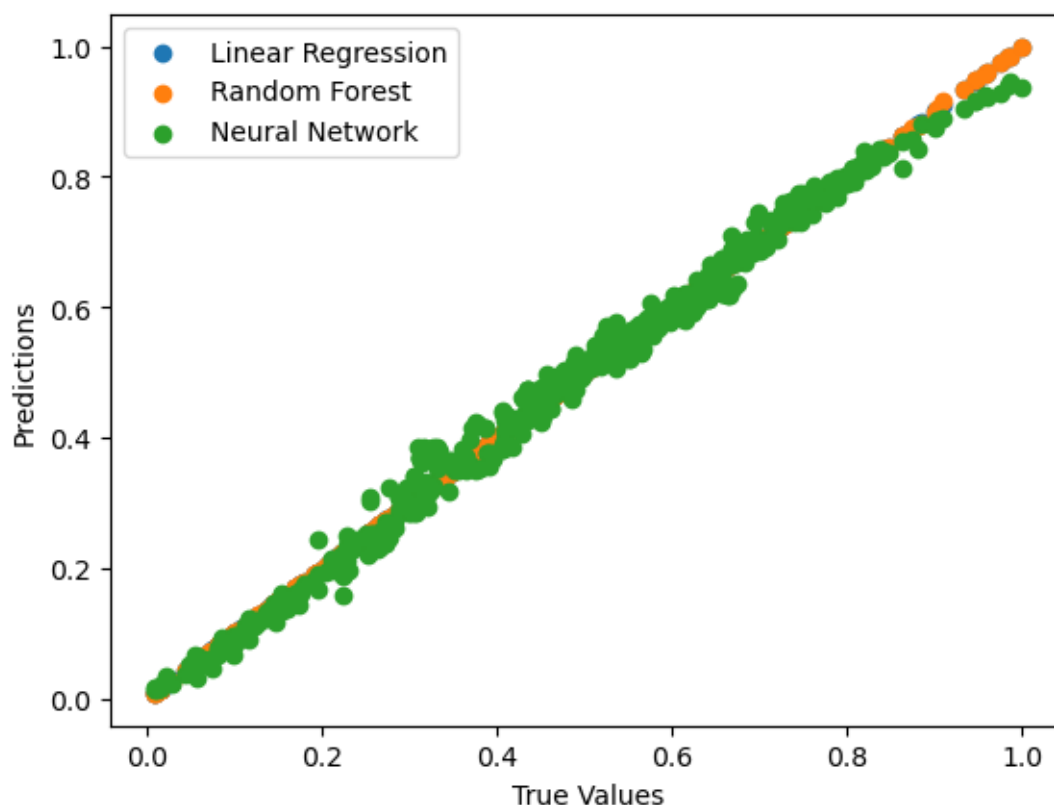
Une exploration approfondie des variables macroéconomiques a été entreprise, englobant des visualisations des tendances temporelles et des distributions. L'objectif était de comprendre comment ces variables influencent potentiellement la performance de l'indice MASI.



# Choix de Modélisation

## Sélection de Modèle

Les modèles de régression linéaire, forêts aléatoires et réseaux de neurones ont été choisis pour l'analyse. Les caractéristiques les plus importantes ont été identifiées pour chaque modèle, permettant une compréhension approfondie des éléments clés.



## Entraînement et Test du Modèle

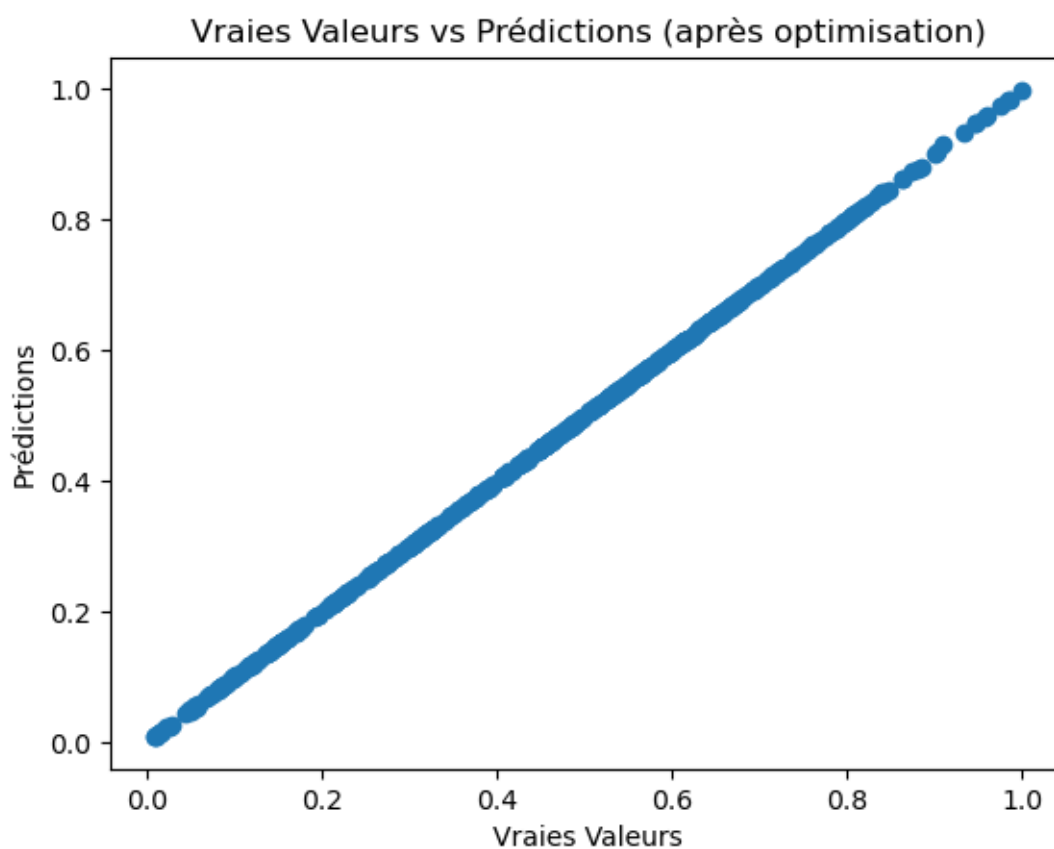
Les modèles ont été entraînés sur des ensembles de données distincts, et leurs performances ont été évaluées en utilisant la métrique d'erreur quadratique moyenne (MSE). Cela a permis de quantifier la précision de chaque modèle dans la prédiction de la performance future de l'indice MASI.

```
Linear Regression MSE: 6.8578238437689195e-31  
Random Forest MSE: 2.2234251153722524e-07  
Neural Network MSE: 0.00038296268417663395
```

# Choix de Modélisation

## Optimisation et Validation

Une optimisation des hyperparamètres a été réalisée spécifiquement pour le modèle de forêt aléatoire. Les performances du modèle optimisé ont été comparées à celles du modèle de base, fournissant ainsi des informations sur la robustesse du modèle par rapport à différents paramètres.





# CONCLUSION

L'analyse a abouti à la création d'un modèle prédictif solide, capable d'estimer la performance future de l'indice MASI. Les facteurs macroéconomiques tels que les avoirs officiels de réserve, le taux de change et l'inflation ont été identifiés comme étant particulièrement influents. Les modèles de forêts aléatoires et de réseaux de neurones ont montré des performances prometteuses, chacun présentant des avantages spécifiques.

En conclusion, ce rapport fournit une méthodologie complète, des analyses approfondies et des justifications pour les choix de modélisation effectués. Les résultats obtenus sont cruciaux pour prendre des décisions éclairées dans le contexte du marché boursier marocain. Des itérations futures pourraient impliquer l'exploration de modèles encore plus complexes et l'intégration de données supplémentaires pour une analyse plus holistique et précise.



## Ayoub ETOULLALI

Élève ingénieur en Ingénierie Informatique

Big Data & Cloud Computing

Université Hassan II de Casablanca | ENSET Mohammedia

☎ +212 6 58 71 20 11

✉ [ayoub.etoullali2002@gmail.com](mailto:ayoub.etoullali2002@gmail.com)

🌐 <https://github.com/Ayoub-etoullali>

📍 ERRACHIDIA