



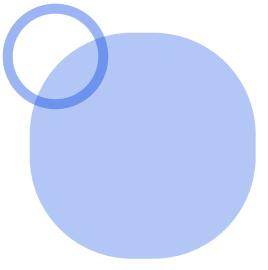
Intelligence Décisionnelle & Artificielle
TECHnologies

Étude de Cas pour Stage Data Analyst/Scientist chez ID&A TECH

PRÉDICTION DE LA PERFORMANCE DE L'INDICE MASI

DAH ABDALLAHI

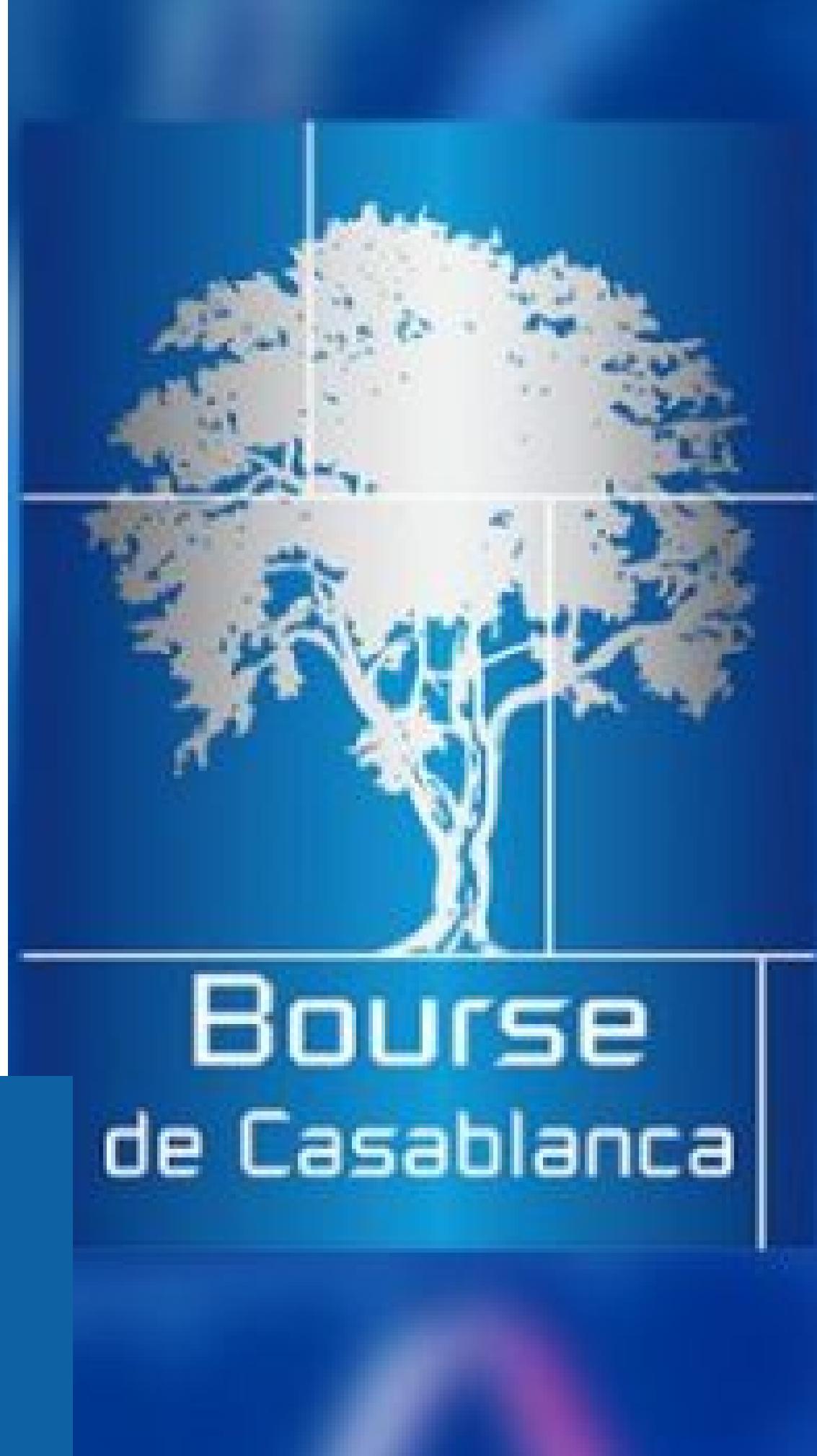




PLAN

Introduction

- 01** Exploration Approfondie des Données
 - 02** Prétraitement des Données
 - 03** Modélisation de Série Temporelle
 - 04** Modélisation avec Variables Exogènes
 - 05** Optimisation et Évaluation des Modèles
 - 06** Résultats et Analyses
- Conclusion



INTRODUCTION

CONTEXTE

Le marché financier est un écosystème complexe influencé par divers facteurs économiques et sectoriels. L'analyse de ces facteurs est cruciale pour anticiper les tendances et prendre des décisions éclairées. Dans ce contexte, notre étude s'est concentrée sur l'Indice MASI, une référence clé pour évaluer la performance boursière au Maroc.

INTRODUCTION

OBJECTIVES



modèle de prédiction

Développer un modèle de prédiction robuste pour anticiper la performance de l'Indice MASI



Identification de facteurs

identifier les facteurs clés influençant les rendements de l'Indice MASI

INTRODUCTION

Méthodologie Adoptée

Pour atteindre nos objectifs, nous avons suivi une méthodologie rigoureuse articulée autour de plusieurs étapes clés

Exploration approfondie des Données (EDA)

Prétraitement des Données

Modélisation de Série Temporelle

Modélisation avec Variables Exogènes

Optimisation et évaluation des modèles

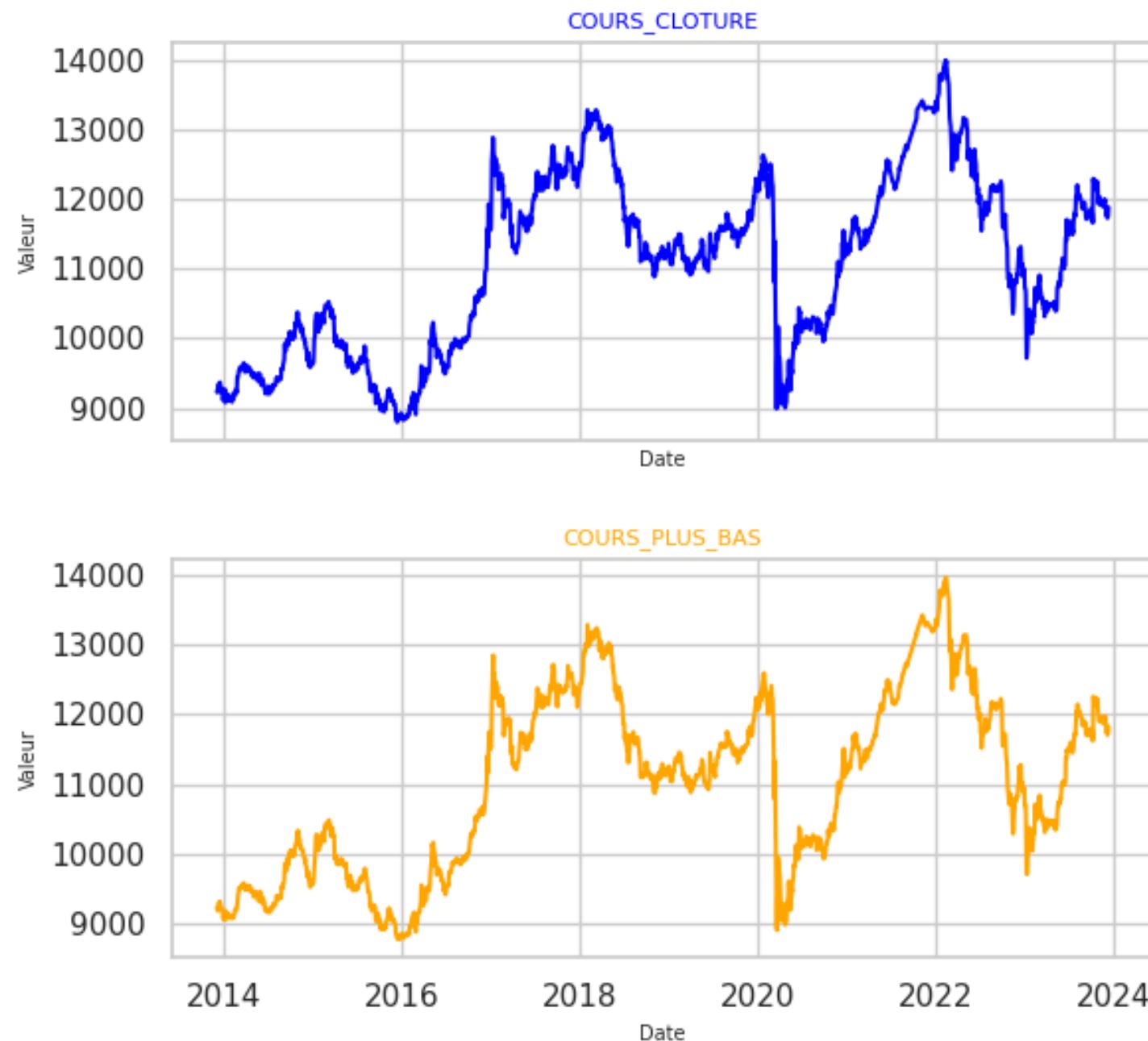
EXPLORATION APPROFONDIE DES DONNÉES (EDA)



EXPLORATION DE L'INDICE GÉNÉRAL

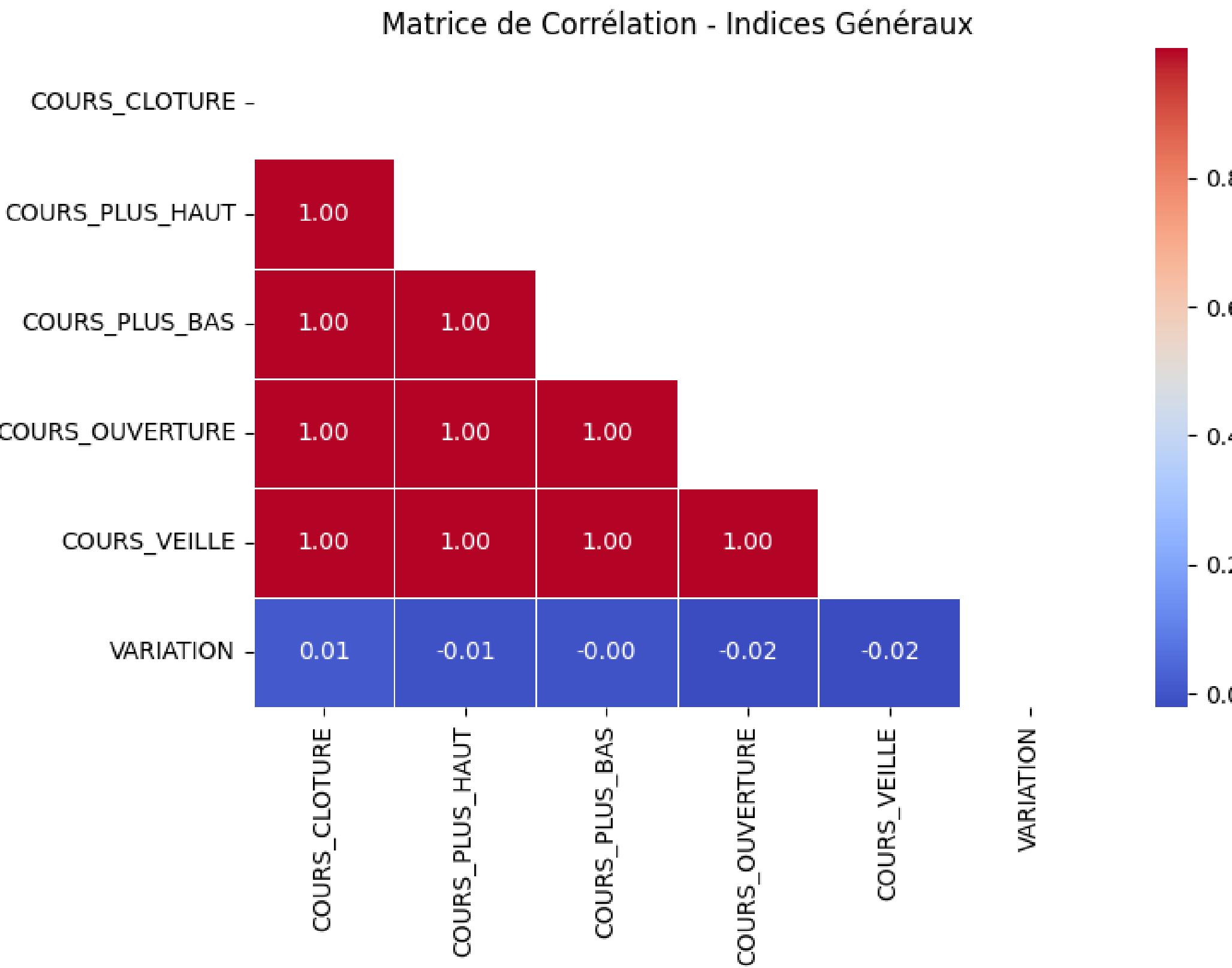
- Exploration de l'Indice Général

Cours des Indices en fonction du temps



EXPLORATION DE L'INDICE GÉNÉRAL

- Exploration de l'Indice Général



EXPLORATION DE L'INDICE GÉNÉRAL

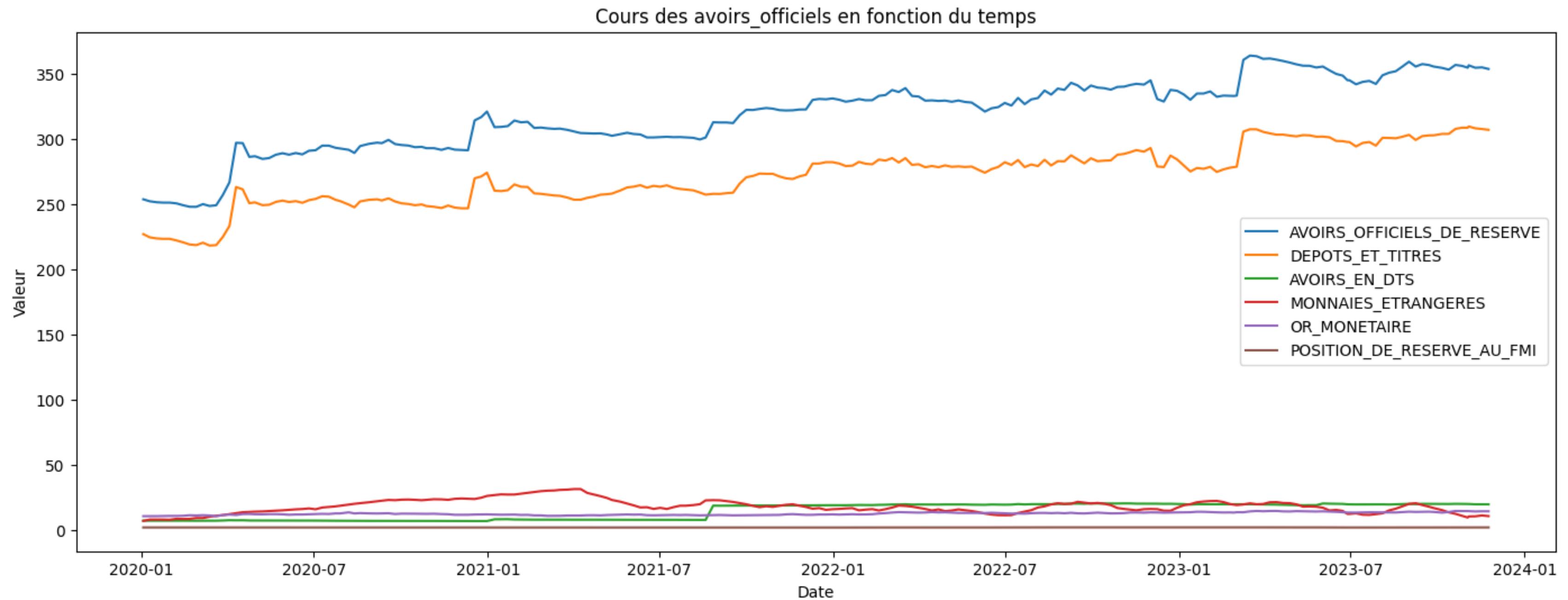
- Exploration de l'Indice Général

	COURS_CLOTURE	COURS_PLUS_HAUT	COURS_PLUS_BAS	COURS_OUVERTURE	COURS_VEILLE	VARIATION
count	2366.000000	2366.000000	2366.000000	2366.000000	2366.000000	2366.000000
mean	11007.492700	11046.029461	10962.958392	11003.195150	11006.688533	-0.001067
std	1228.034432	1234.011316	1225.577153	1230.984464	1228.639022	0.344325
min	8796.650391	8837.969727	8789.889648	8796.650391	8796.650391	-3.820000
25%	9927.517578	9957.692383	9888.077393	9923.909678	9927.104736	-0.004097
50%	11159.635254	11201.135254	11119.455078	11156.564941	11158.260254	0.000210
75%	11969.242559	12012.205000	11918.409902	11971.260000	11971.260000	0.004357
max	13991.470000	14012.640000	13950.560000	13991.470000	13991.470000	5.080000

statistique descriptive

EXPLORATION DE L'INDICE GÉNÉRAL

- Exploration de Avoirs officiels de réserve :



EXPLORATION DE L'INDICE GÉNÉRAL

- Exploration de l'Indice Général

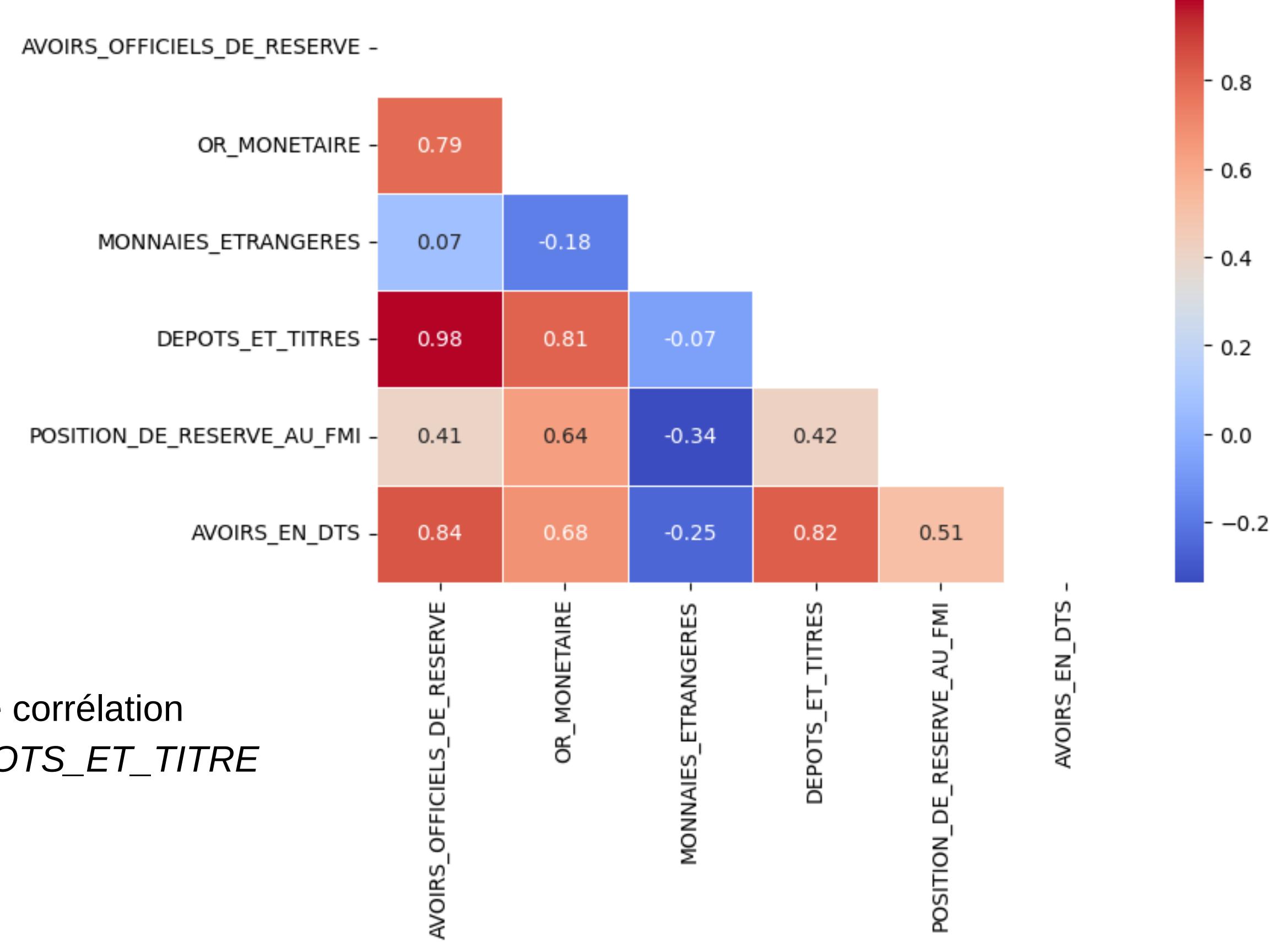
	AVOIRS_OFFICIELS_DE_RESERVE	OR_MONETAIRE	MONNAIES_ETRANGERES	DEPOTS_ET_TITRES	POSITION_DE_RESERVE_AU_FMI	AVOIRS_EN_DTS
count	207.000000	207.000000	207.000000	207.000000	207.000000	207.000000
mean	318.721258	12.583768	18.090473	271.568289	1.957151	14.527534
std	28.083077	1.070565	5.140158	22.118081	0.058049	6.034526
min	247.751000	10.560000	6.992000	218.044000	1.871077	6.852960
25%	300.747593	11.672304	15.006020	255.733000	1.903109	7.510500
50%	324.153140	12.590958	17.863413	274.850257	1.957449	18.945310
75%	338.558395	13.455218	21.087087	284.093164	1.994542	19.695408
max	363.692490	14.600335	31.386899	309.361444	2.076412	20.403429

statistique descriptive

EXPLORATION DE L'INDICE GÉNÉRAL

- Exploration de l'Indice Général

Matrice de Corrélation - Indices Sectoriels



La matrice de corrélation, révèle une corrélation significative entre les variables *DEPOTS_ET_TITRE* et *AVOIR_OFFICE_DE_RESERVE*.

PRÉTRAITEMET DES DONNÉES



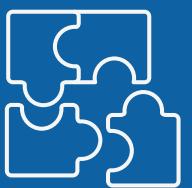
PRÉTRAITEMET DES DONNÉES



➤ Sélection de Variables

- indices_generaux_cleaned : "COURS_CLOTURE" avec "DATE".
- avoirs_officiels : supprimé la variable "DEPOTS_ET_TITRES"
- change : "Change_Minimum" et "Change_Maximum" avec "DATE".
- monia : "val_Indice_MONIA" et "vol_Volume_JJ" avec "DATE".
- tmp : "TMP" avec "DATE".
- inflation : "INFLATION" avec "DATE".
- indices_taux : "NOMINAL 1 AN", "NOMINAL 2 ANS", "NOMINAL 3 ANS", "NOMINAL 10 ANS" et "NOMINAL 20 ANS" avec "DATE".
- indices_sectoriels_cleaned : éliminé les variables fortement corrélées

PRÉTRAITEMENT DES DONNÉES

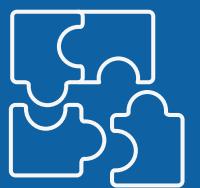


PRÉPARATION DES DONNÉES

Intégrant des variables Exogènes
comme les indices de
macroéconomie et secteielels

considération uniquement la
variation de prix de clôture de
MAIS dans le temps

PRÉTRAITEMENT DES DONNÉES

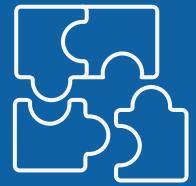


PRÉPARATION DES DONNÉES

considération uniquement la
variation de prix de clôture de
MAIS dans le temps



PRÉTRAITEMET DES DONNÉES



PRÉPARATION DES DONNÉES

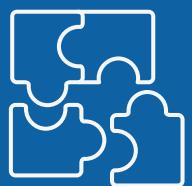
Integrant des variables Exogènes
comme les indices de
macroéconomie et secteielels

Les données de indices taux,
tmp, les indices sectorielle et
monia ...etc, ne sont
disponibles que depuis 2021

2021

nous avons décidé de conserver uniquement les
enregistrements de notre variable cible à partir de l'année 2021

PRÉTRAITEMET DES DONNÉES

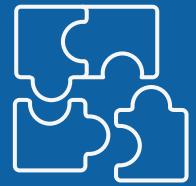


PRÉPARATION DES DONNÉES

Integrant des variables Exogènes
comme les indices de
macroéconomie et secteielels

DATE	0
COURS_CLOTURE	0
AVOIRS_OFFICIELS_DE_RESERVE	4
OR_MONETAIRE	4
MONNAIES_ETRANGERES	4
POSITION_DE_RESERVE_AU_FMI	4
AVOIRS_EN_DTS	4
val_Indice_MONIA	0
vol_Volume_JJ	0
TMP	0
INFLATION	129
NOMINAL 1 AN	0
NOMINAL 2 ANS	0
NOMINAL 3 ANS	0
NOMINAL 10 ANS	0
NOMINAL 20 ANS	0
BANQUES	0

PRÉTRAITEMET DES DONNÉES



PRÉPARATION DES DONNÉES

Integrant des variables Exogènes
comme les indices de
macroéconomie et secteielels

Utiliser la fonction `merge_asof()` de la bibliothèque Pandas afin d'associer les dates aux dates les plus proches.

- Imputation

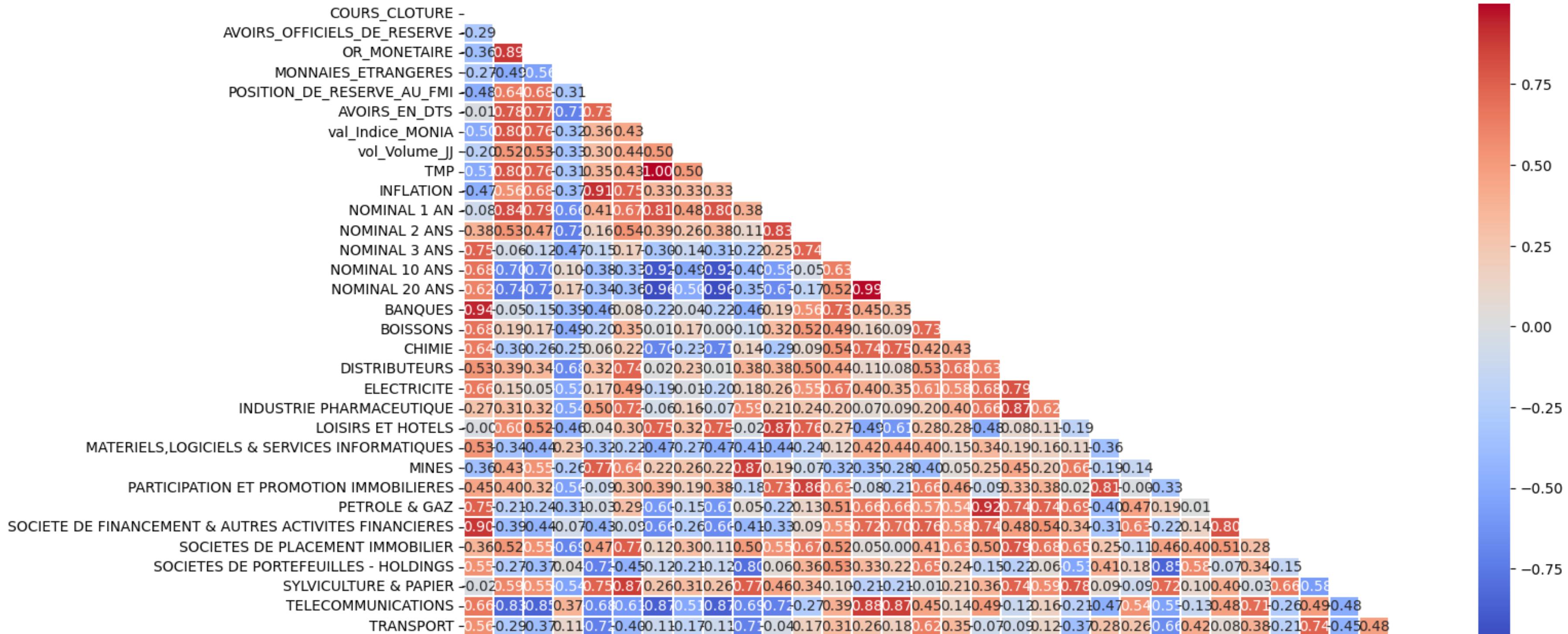
le taux
d'inflation
moyen au
Maroc en 2021

1.4%

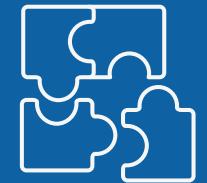
PRÉTRAITEMENT DES DONNÉES

Intégrant des variables Exogènes comme les indices de macroéconomie et secteielels

Matrice de Corrélation



PRÉTRAITEMENT DES DONNÉES



PRÉPARATION DES DONNÉES

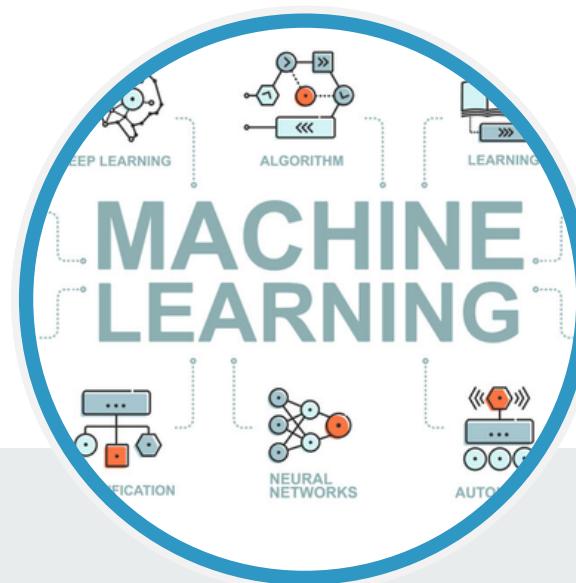
Standardisation des Données

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0, 1]$$

MODÉLISATION DE SÉRIE TEMPORELLE



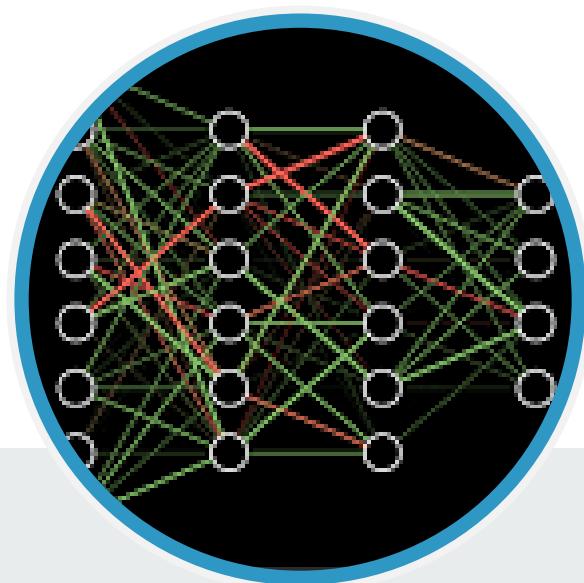
MODÉLISATION DE SÉRIE TEMPORELLE



**Machine
Learning**



**Approche
Statistique**



**Deep
Learning**

MODÉLISATION DE SÉRIE TEMPORELLE

—



**Machine
Learning**

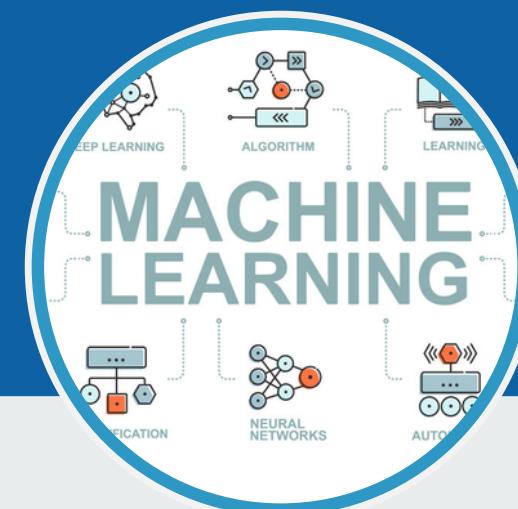
Modèles d'Apprentissage
Automatique (ML)

**Random
Forest**

XGBoost

MODÉLISATION DE SÉRIE TEMPORELLE

T



**Machine
Learning**

Modèles d'Apprentissage Automatique (ML)

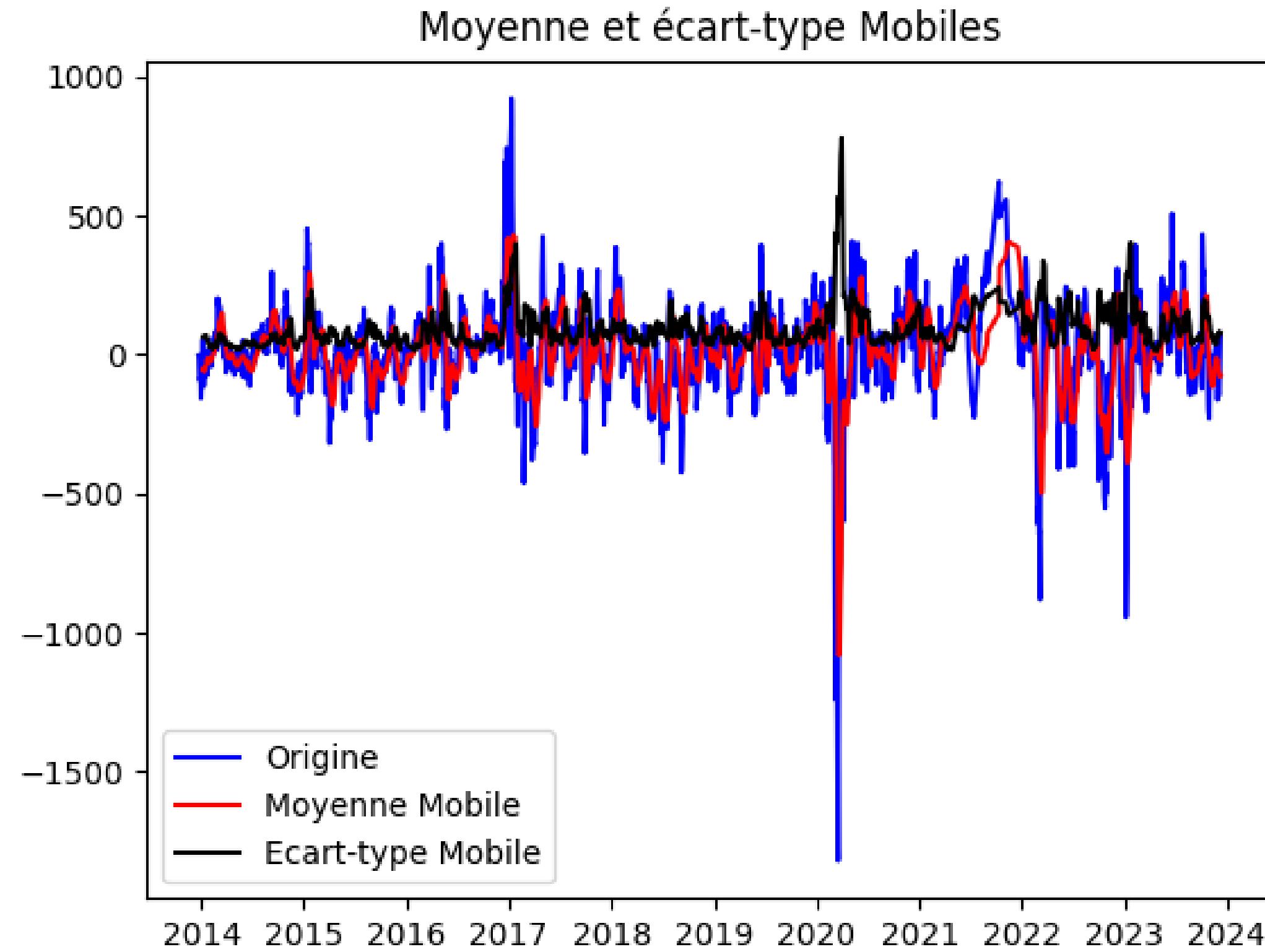


MODÉLISATION DE SÉRIE TEMPORELLE



Approche
Statistique

Modèle Statistique ARIMA

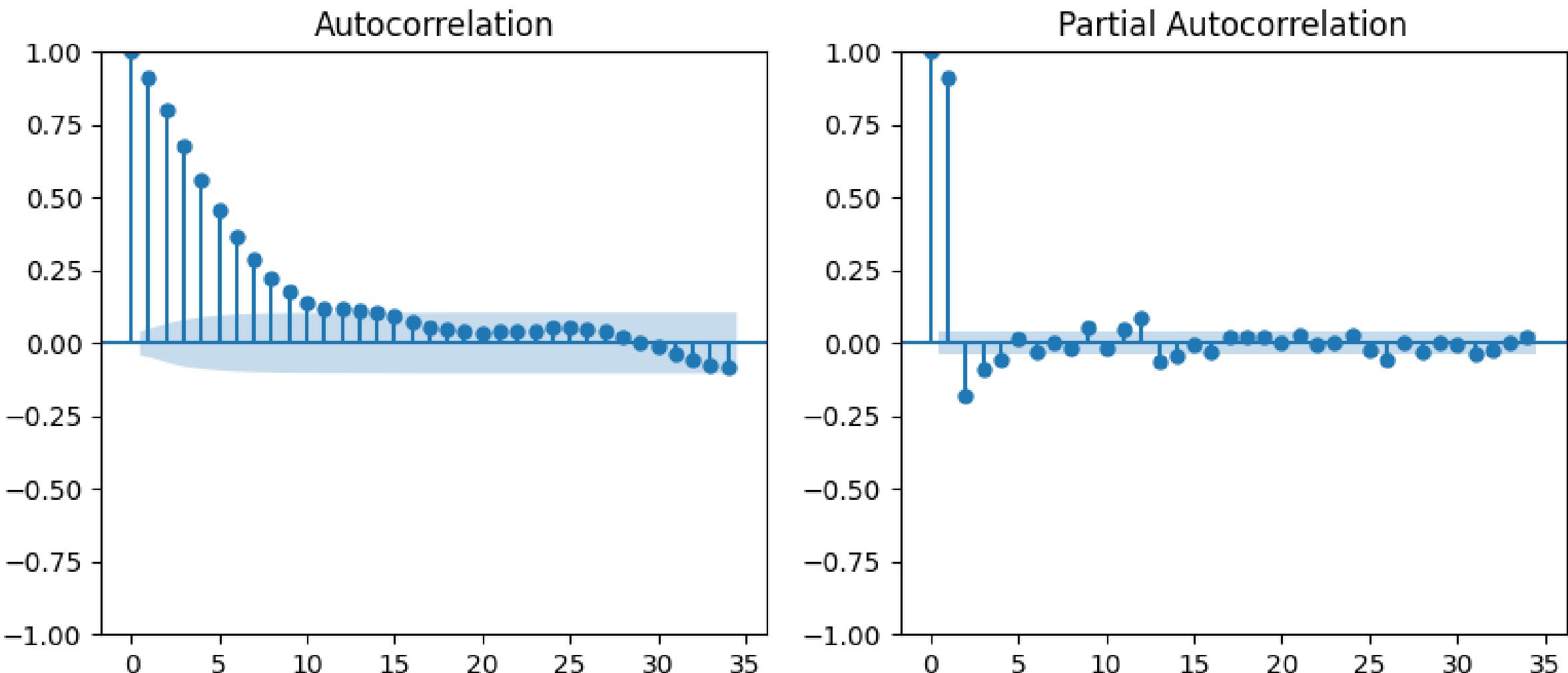


MODÉLISATION DE SÉRIE TEMPORELLE

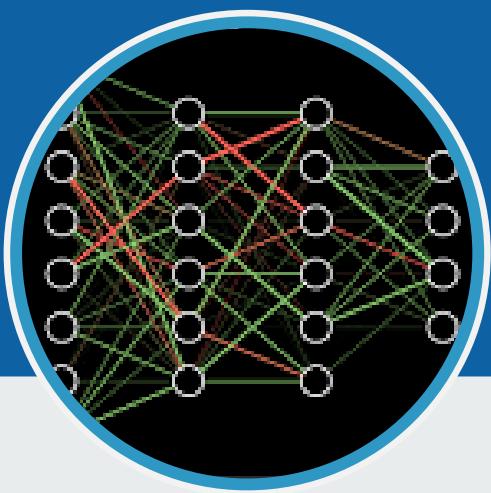


Approche
Statistique

Modèle Statistique ARIMA



MODÉLISATION DE SÉRIE TEMPORELLE



Deep
Learning

Modèle de Deep Learning - LSTM

Model: "sequential_9"

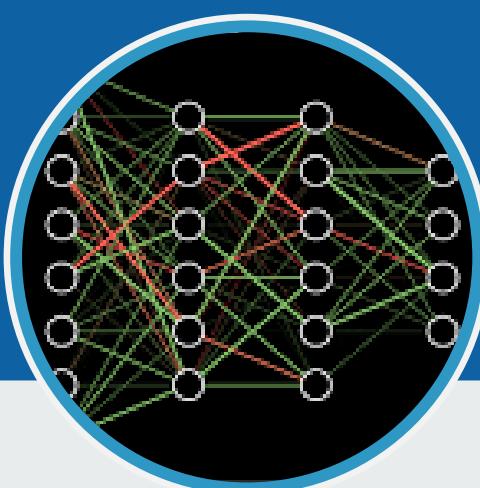
Layer (type)	Output Shape	Param #
=====		
lstm_15 (LSTM)	(None, 7, 50)	10400
lstm_16 (LSTM)	(None, 50)	20200
dense_16 (Dense)	(None, 25)	1275
dense_17 (Dense)	(None, 1)	26
=====		

Total params: 31981 (124.61 KB)

Trainable params: 31981 (124.61 KB)

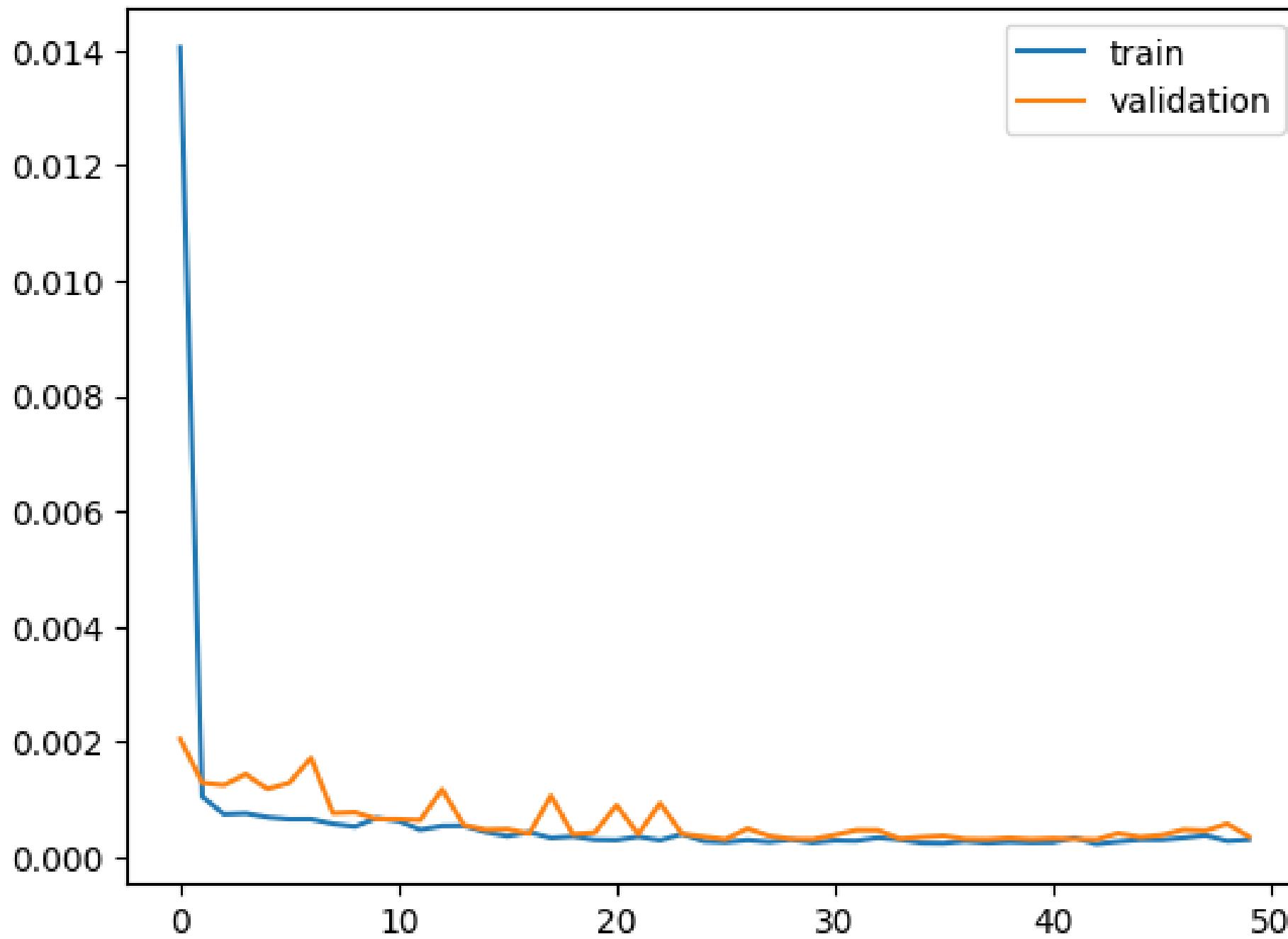
Non-trainable params: 0 (0.00 Byte)

MODÉLISATION DE SÉRIE TEMPORELLE



Deep
Learning

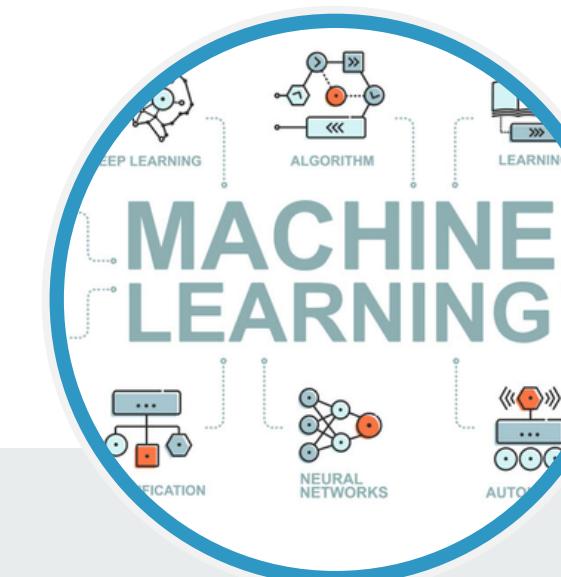
Modèle de Deep Learning - LSTM



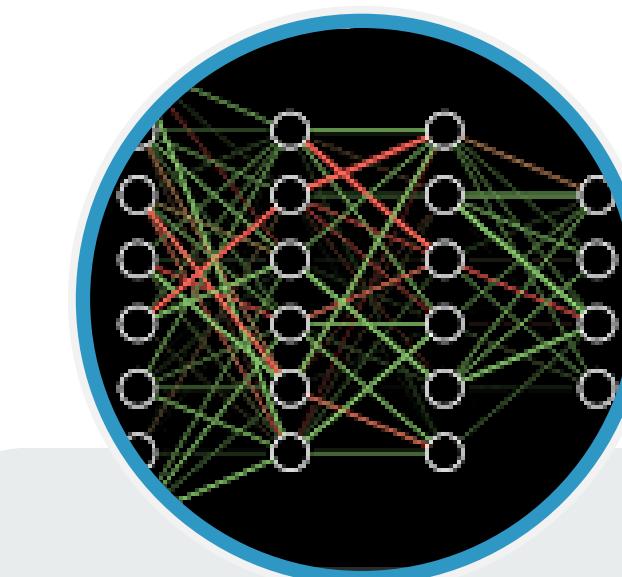
MODÉLISATION AVEC VARIABLES EXOGÈNES



MODÉLISATION AVEC VARIABLES EXPLICATIVES



**Machine
Learning**

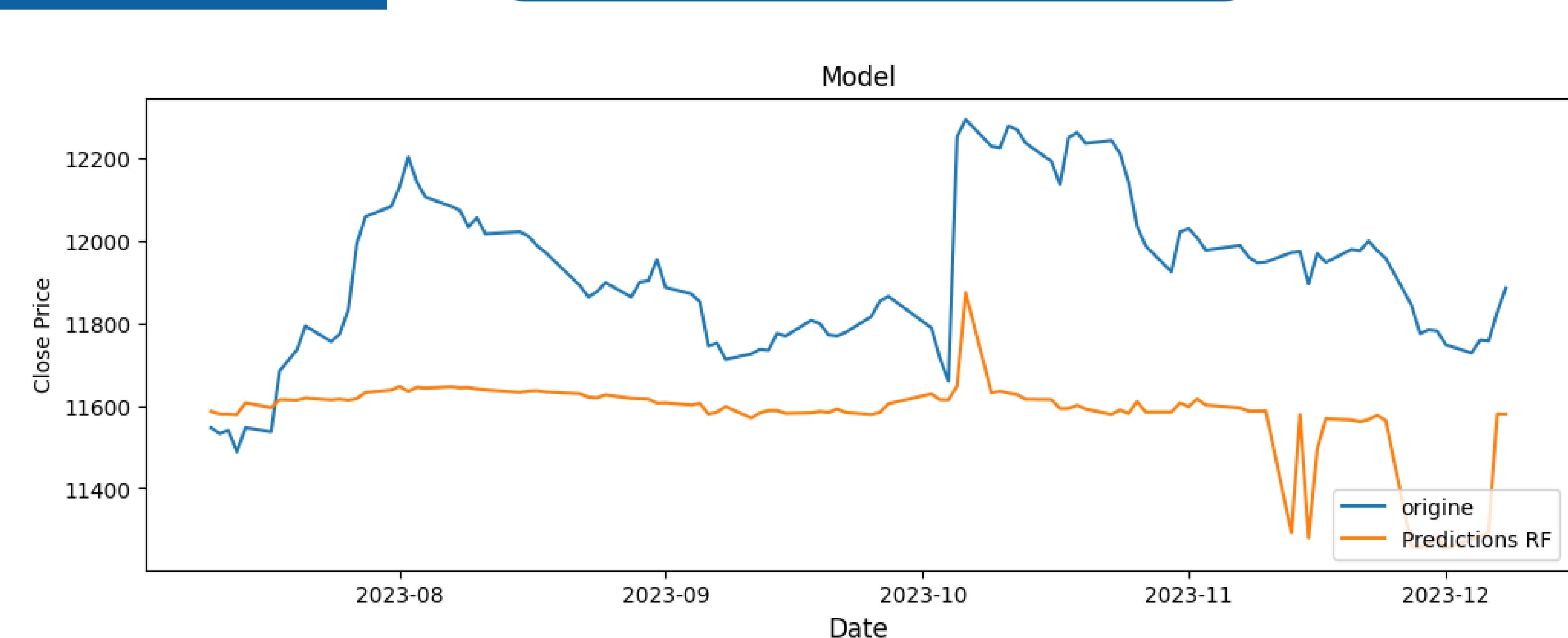


**Deep
Learning**

MODÉLISATION AVEC VARIABLES EXPLICAT

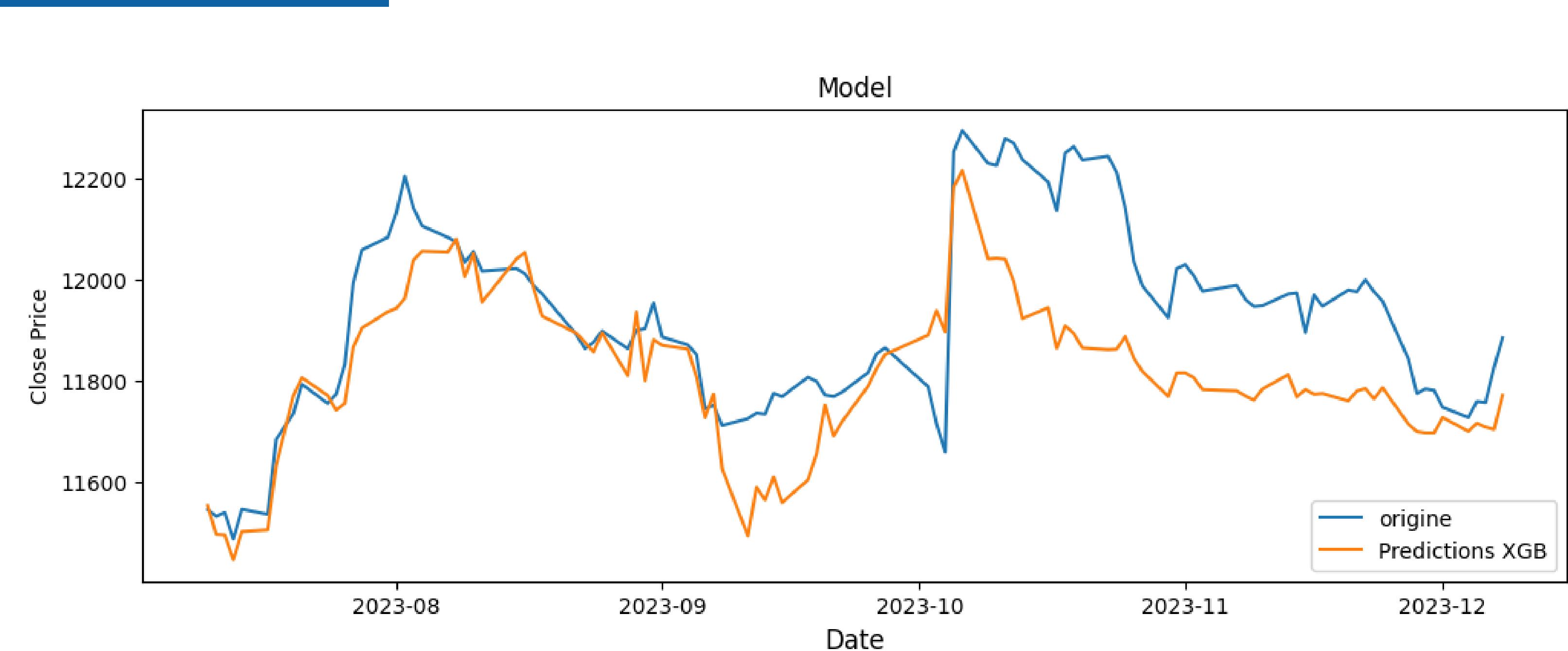


Random Forest Regressor

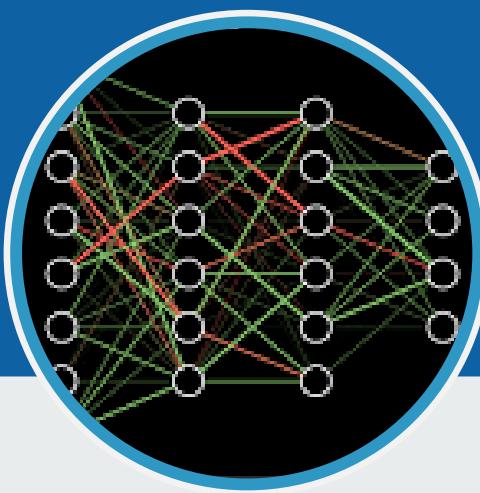


MODÉLISATION AVEC VARIABLES EXPLICATI

extreme gradient boosting (XGB)



MODÉLISATION DE SÉRIE TEMPORELLE



Deep
Learning

Modèle de Deep Learning - LSTM

Model: "sequential_7"

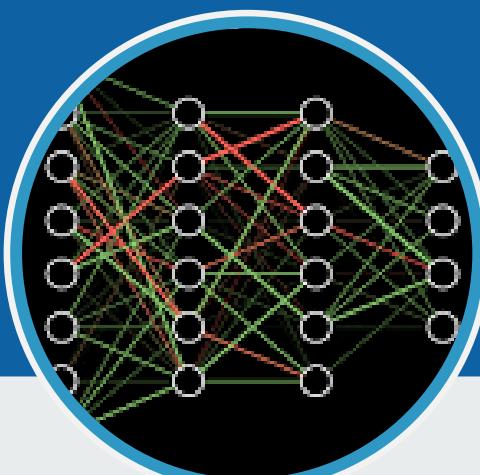
Layer (type)	Output Shape	Param #
lstm_11 (LSTM)	(None, 7, 50)	16800
lstm_12 (LSTM)	(None, 50)	20200
dense_12 (Dense)	(None, 10)	510
dense_13 (Dense)	(None, 1)	11

Total params: 37521 (146.57 KB)

Trainable params: 37521 (146.57 KB)

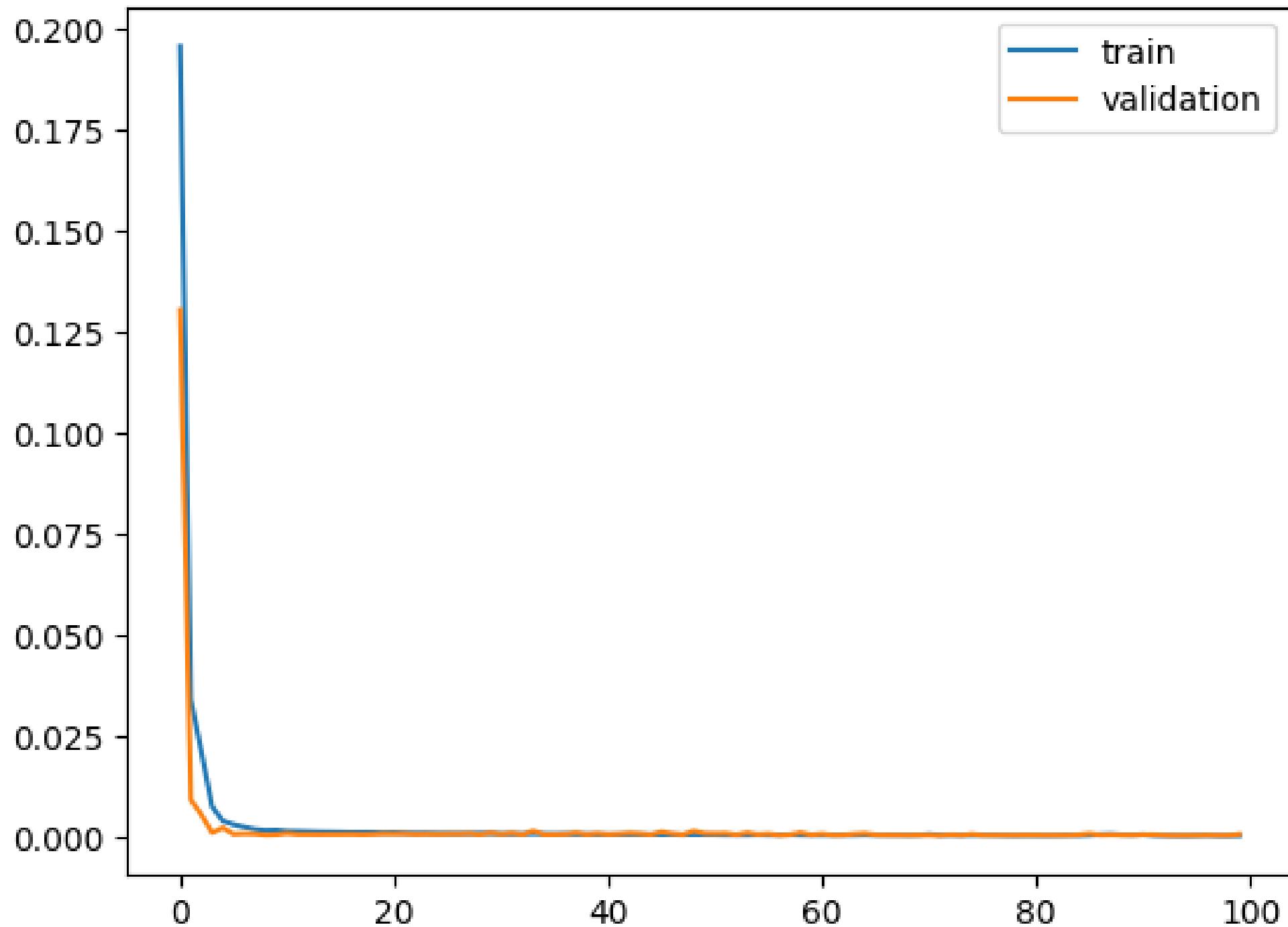
Non-trainable params: 0 (0.00 Byte)

MODÉLISATION DE SÉRIE TEMPORELLE



Deep
Learning

Modèle de Deep Learning - LSTM



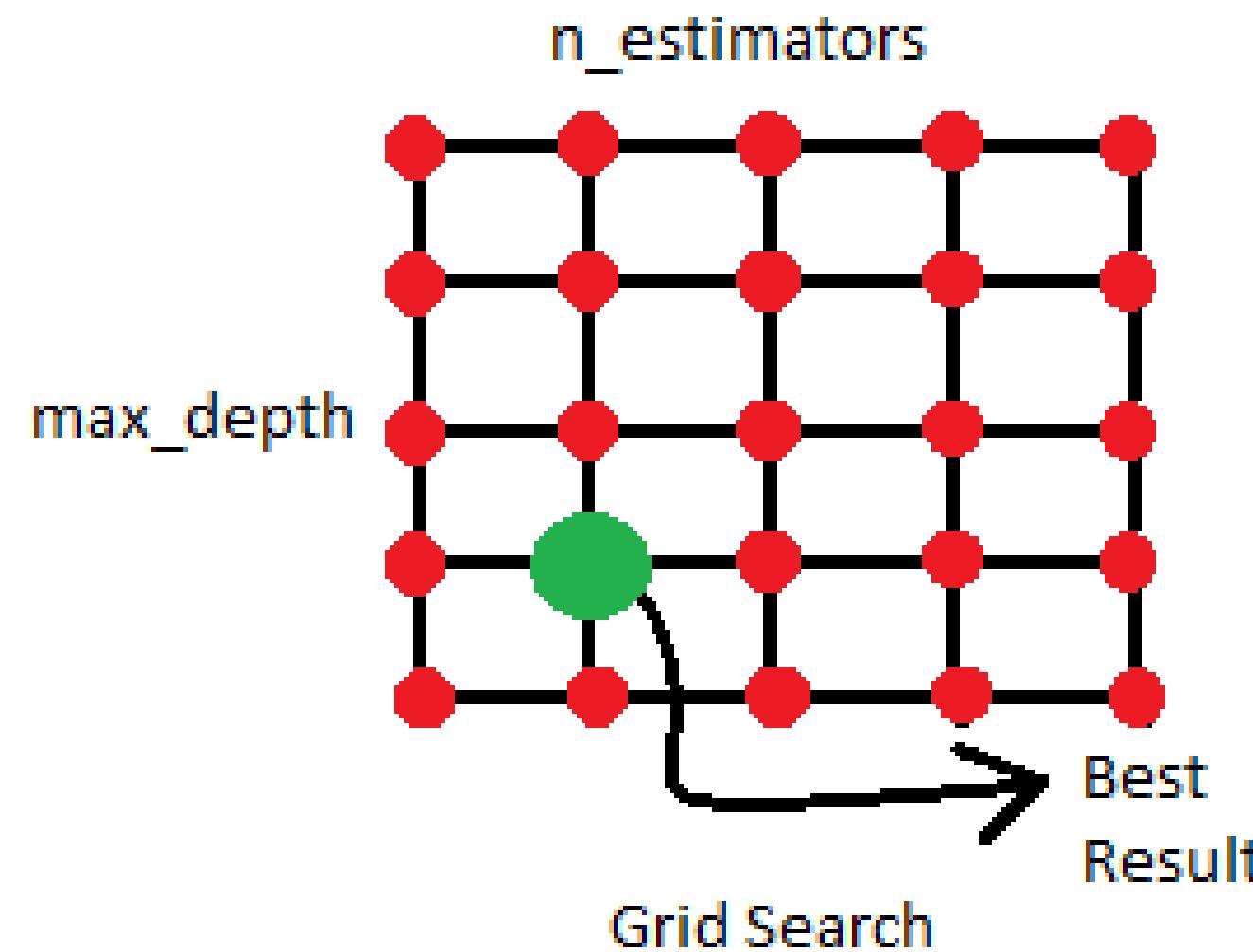
OPTIMISATION ET ÉVALUATION DES MODÈLES



OPTIMISATION ET ÉVALUATION DES MODÈLES :

T

GRID SEARCH



La technique de Grid Search fonctionne en définissant un ensemble de valeurs possibles pour chaque hyperparamètre, puis en évaluant le modèle pour chaque combinaison de ces valeurs. Cela permet d'identifier les paramètres qui conduisent à une performance optimale du modèle.

OPTIMISATION ET ÉVALUATION DES MODÈLES :

T

ÉVALUATION DES MODÈLES

Nous avons choisi le Root Mean Squared Error (RMSE) comme critère d'évaluation principal. Le RMSE mesure la dispersion des erreurs entre les valeurs prédites et les valeurs réelles. Plus spécifiquement, l'équation du RMSE est définie comme suit :

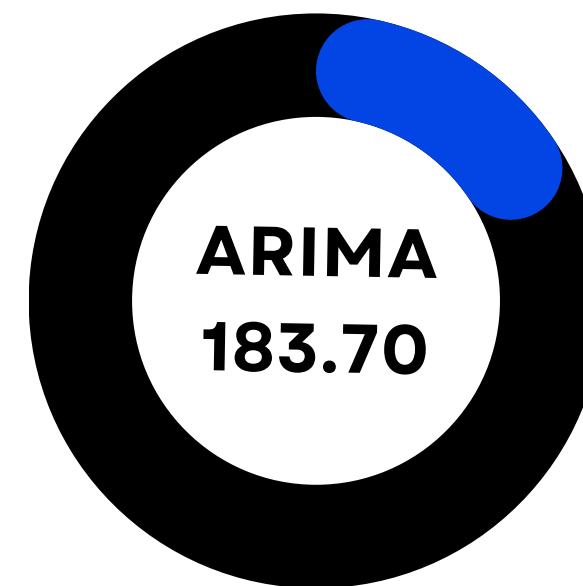
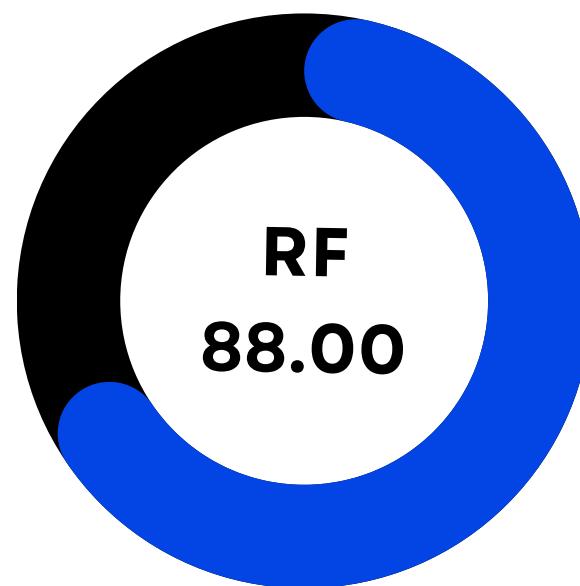
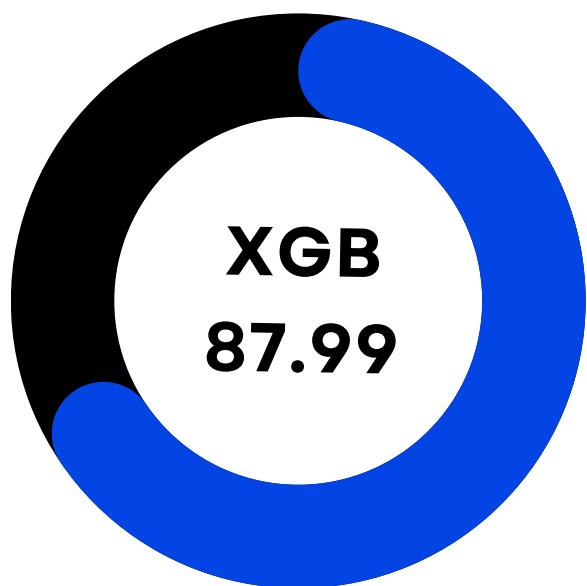
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

RÉSULTATS ET ANALYSES



RÉSULTATS ET ANALYSES

RMSE : MODÉLISATION DE SÉRIE TEMPORELLE



RÉSULTATS ET ANALYSES

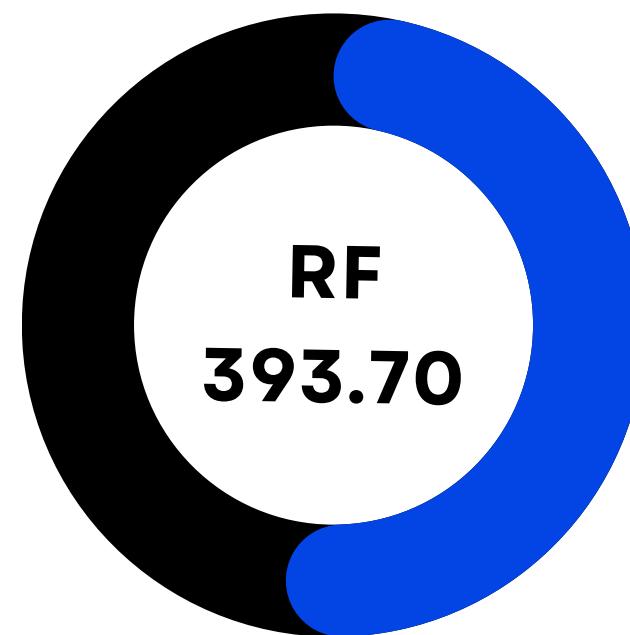
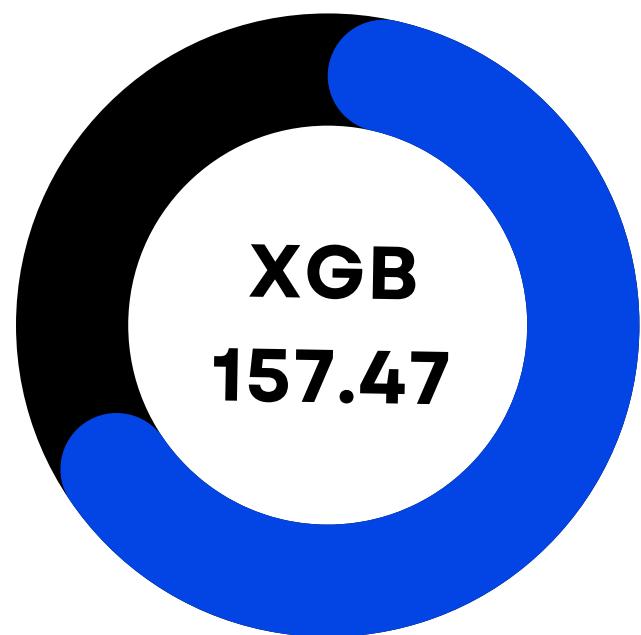
T

RÉSULTATS SUR LA MODÉLISATION DE SÉRIE TEMPORELLE



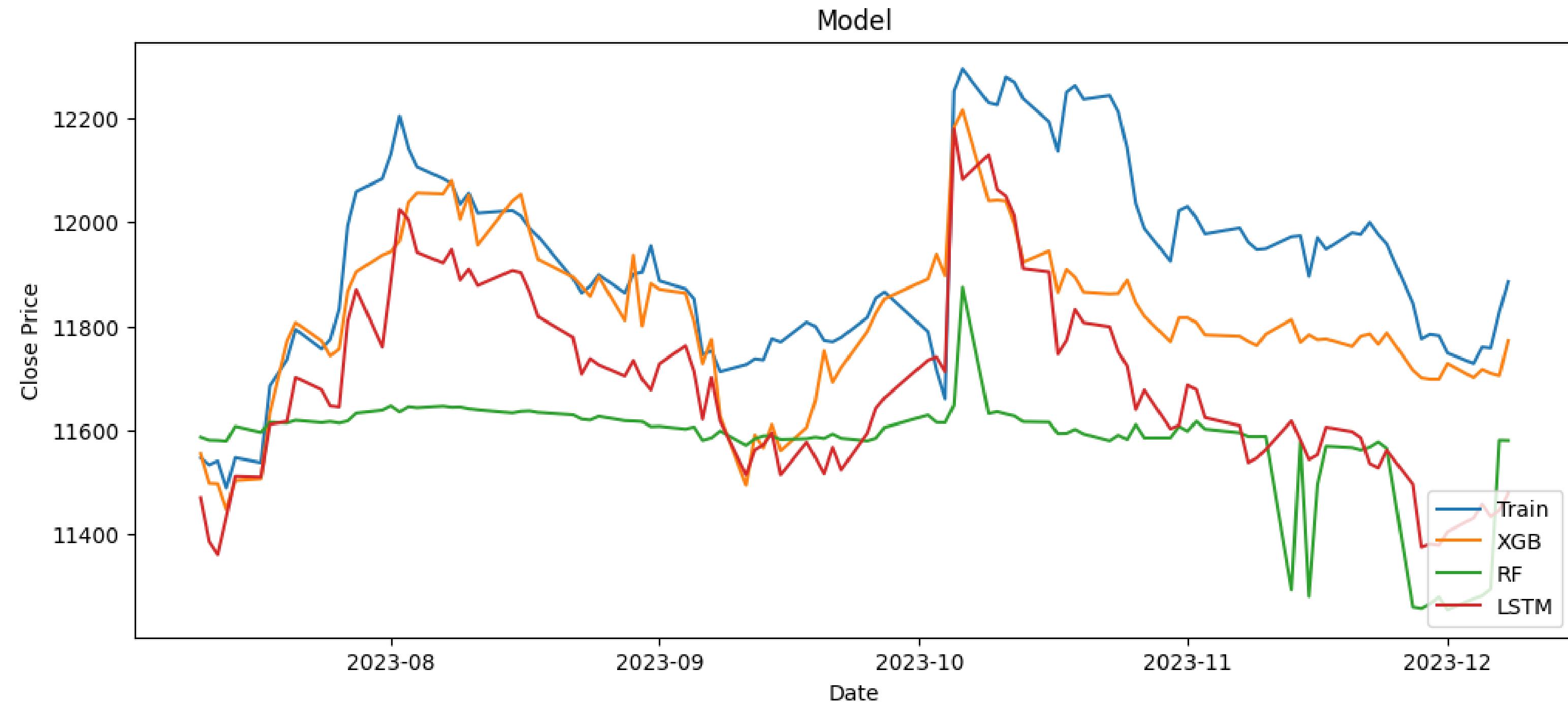
RÉSULTATS ET ANALYSES

RMSE : MODÉLISATION AVEC VARIABLES EXPLICATIVES



RÉSULTATS ET ANALYSES

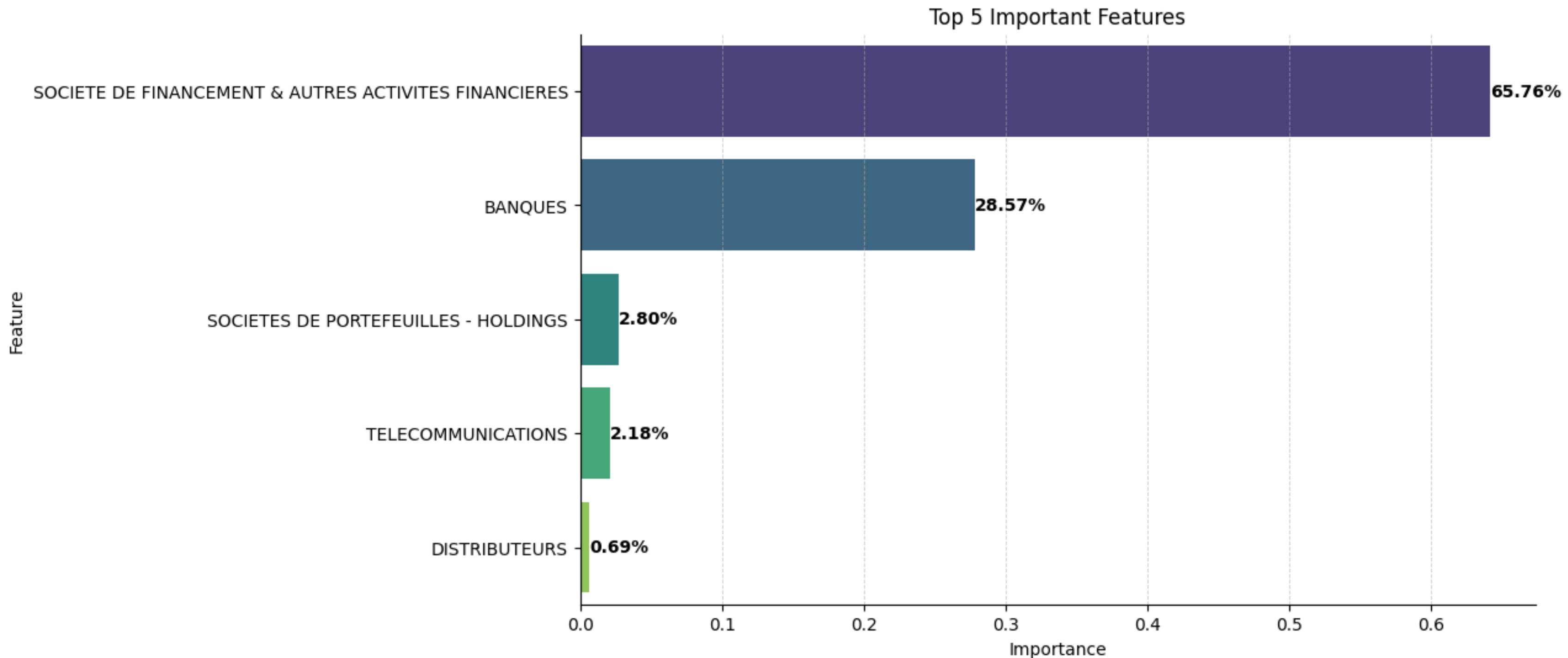
RÉSULTATS SUR LA MODÉLISATION DE SÉRIE TEMPORELLE



RÉSULTATS ET ANALYSES

T

LES FACTEURS LES PLUS IMPORTANTS



CONCLUSION



L'utilisation combinée des modèles LSTM pour la prédiction du prix du MASI sans prendre en compte les variables exogènes, et de XGB en considérant les autres variables explicatives, constitue une approche robuste pour anticiper les performances de l'indice MASI. Ces stratégies se révèlent être des indicateurs de qualité, fournissant des informations essentielles aux investisseurs et aux analystes financiers.

MERCI!

