

# Rapport de projet

## Développement d'une Application de Prédiction Boursière

*Projet réalisé par :*

Alami Louati Ghali

Mohaidine Yassine

*Projet encadré par :*

Madame Sophia Faris

Hicham Bouchtib

Année universitaire : 2024-2025

# Remerciement

Je tiens à exprimer ma sincère reconnaissance à Madame **Sophia Faris**, encadrante de ce projet, pour son accompagnement précieux, sa rigueur et son implication continue tout au long de cette expérience. Grâce à son suivi régulier, à la clarté de ses orientations et à la pertinence de ses observations, j'ai pu structurer mes réflexions, approfondir mes analyses et progresser de manière constante. Sa disponibilité et sa bienveillance ont créé un climat de confiance propice à l'apprentissage et à la prise d'initiative, tout en maintenant un haut niveau d'exigence scientifique. Son encadrement a joué un rôle central dans l'atteinte des objectifs fixés et dans la qualité du travail final présenté.

Je souhaite également adresser mes remerciements les plus chaleureux au **Professeur en charge du module de recherche scientifique**, pour son rôle essentiel dans ma formation. Son enseignement, fondé sur une approche pédagogique exigeante et structurée, m'a permis d'acquérir une méthodologie de travail rigoureuse, un esprit critique affûté ainsi qu'une meilleure compréhension des enjeux liés à la recherche académique. Les échanges constructifs, les évaluations approfondies et les conseils méthodologiques qu'il m'a prodigués tout au long de ce module ont significativement enrichi mes compétences analytiques et rédactionnelles, et m'ont donné les outils nécessaires pour conduire un projet de recherche de manière autonome et professionnelle.

Les connaissances, les méthodes et les attitudes que j'ai pu développer grâce à l'encadrement de ces deux enseignants constituent un socle fondamental sur lequel je m'appuierai pour mes futurs projets académiques et professionnels. Leur accompagnement a non seulement favorisé la réussite de ce projet, mais a également renforcé ma motivation à poursuivre dans le domaine de la recherche scientifique et à m'impliquer activement dans des travaux à forte valeur ajoutée intellectuelle et sociale.

# Dédicace

Je souhaite dédier ce travail de fin d'année à toutes les personnes qui, de près ou de loin, ont contribué à son aboutissement.

À mes parents, dont le soutien indéfectible, la patience et les encouragements constants ont été une source précieuse de force et de motivation tout au long de mon parcours académique.

À mon encadrant, pour sa disponibilité, la qualité de ses conseils et son accompagnement éclairé, qui ont grandement contribué à la progression et à la réussite de ce projet.

À mes camarades de promotion et amis, pour les échanges stimulants, l'entraide, et l'esprit de collaboration qui ont enrichi cette expérience académique.

Enfin, à toutes celles et ceux qui ont cru en moi et m'ont apporté leur appui, direct ou indirect, je tiens à exprimer ma profonde reconnaissance.

# Résumé

Ce travail de fin d'année présente le développement d'une application web avancée dédiée à l'analyse et à la prédiction des prix boursiers, intégrant des techniques d'intelligence artificielle et d'analyse de données financières. L'application, développée en Python, repose sur une architecture robuste combinant plusieurs technologies modernes.

## Architecture technique

L'application est construite autour d'un framework web Streamlit, offrant une interface utilisateur intuitive et réactive. Elle intègre un système de gestion des utilisateurs sécurisé via une base de données SQLite, implémentant l'authentification et l'autorisation des utilisateurs. L'application utilise yfinance pour l'acquisition des données boursières en temps réel, ainsi que des bibliothèques de machine learning (Scikit-learn) pour les modèles prédictifs.

## Méthodologie de prédiction

L'approche de prédiction repose sur deux modèles principaux :

1. **Régression linéaire** : pour capturer les tendances de base.
2. **Régression polynomiale** : pour modéliser les variations plus complexes.

Ces modèles sont évalués rigoureusement à l'aide de métriques statistiques (MAE, MSE, RMSE,  $R^2$ ) et comparés pour déterminer le modèle le plus performant. L'application inclut également une analyse des moyennes mobiles (50, 100 et 200 jours) pour l'analyse technique traditionnelle.

## Fonctionnalités avancées

L'application offre plusieurs fonctionnalités innovantes :

1. Interface utilisateur sécurisée avec système d'authentification
2. Visualisation interactive des données historiques et des prédictions
3. Analyse des tendances à court et long terme
4. Évaluation comparative des modèles de prédiction

- 
5. Intégration des actualités financières en temps réel
  6. Métriques d'évaluation expliquées et contextualisées.

### **Contributions scientifiques**

Ce projet contribue à :

1. La démonstration pratique des applications de l'IA en finance.
2. Le développement d'une approche hybride combinant analyse technique et machine learning.
3. La mise en œuvre d'un système d'évaluation des modèles de prédiction robuste.
4. La création d'une interface utilisateur accessible aux non-experts.

### **Limitations et perspectives**

Bien que l'application offre des capacités de prédiction avancées, elle souligne également les limites inhérentes à la prédiction des marchés financiers. Les résultats doivent être interprétés avec prudence, et l'application est conçue comme un outil d'aide à la décision plutôt qu'un système de trading automatique.

Ce travail représente une contribution significative à la compréhension des applications de l'intelligence artificielle dans le domaine financier, tout en offrant une plateforme éducative pour l'exploration des marchés boursiers.

# Table des figures

2.1	Login/Register . . . . .	17
2.2	Historical Data . . . . .	18
2.3	Polynomial Regression . . . . .	19
2.4	Linear Regression . . . . .	19
2.5	Linear Regression metrics . . . . .	20
2.6	Polynomial Regression metrics . . . . .	20

# Table des matières

<b>1</b>	<b>Contexte du projet</b>	<b>10</b>
1.1	Introduction . . . . .	10
1.2	État de l’art . . . . .	11
1.2.1	Introduction . . . . .	11
1.2.2	Évolution des approches d’analyse boursière . . . . .	11
1.2.3	Panorama des plateformes existantes . . . . .	12
1.2.4	Limites des solutions existantes . . . . .	12
1.2.5	Apport du projet proposé . . . . .	13
1.3	Problématique . . . . .	13
1.3.1	Problématique Principale . . . . .	14
1.3.2	Axes de Recherche . . . . .	14
1.3.3	Méthodologie d’Évaluation . . . . .	15
1.3.4	Résultats Attendus . . . . .	15
1.3.5	Limites et Perspectives . . . . .	15
<b>2</b>	<b>Elaboration du modèle et explication</b>	<b>16</b>
2.1	Architecture générale . . . . .	16
2.1.1	Module d’Authentification . . . . .	16
2.1.2	Module de Prévision . . . . .	16
2.1.3	Module d’Analyse Boursière . . . . .	17
2.1.4	Module de Prédiction . . . . .	18
2.1.5	Module d’Évaluation . . . . .	19
2.1.6	Interface Utilisateur (Streamlit) . . . . .	20
2.1.7	Technologies Clés . . . . .	20
2.2	Choix méthodologique . . . . .	21
2.2.1	Approches de Prédiction des Prix des Actions . . . . .	21
2.2.2	Métriques d’Évaluation des Modèles . . . . .	22
2.3	Outils et Technologies Justifiés . . . . .	22
2.3.1	Analyse et Prédiction . . . . .	22
2.3.2	Visualisation . . . . .	23

2.3.3	Base de Données . . . . .	23
2.3.4	Web Scraping . . . . .	23
2.3.5	Sécurité . . . . .	23
2.3.6	Interface . . . . .	23
<b>3</b>	<b>Réalisation, test du modèle et discussion, analyse des résultats</b>	<b>24</b>
3.1	Implémentation . . . . .	24
3.2	Evaluation des modèles . . . . .	24
3.2.1	MAE (Mean Absolute Error) . . . . .	25
3.2.2	RMSE (Root Mean Squared Error) . . . . .	25
3.2.3	MSE (Mean Squared Error) . . . . .	26
3.2.4	R <sup>2</sup> (Coefficient de détermination) . . . . .	26
3.2.5	Sélection automatique du meilleur modèle . . . . .	26
3.3	Discussion des résultats . . . . .	27
3.3.1	Influence des caractéristiques des actions sur les performances .	27
3.3.2	Complémentarité des métriques d'évaluation . . . . .	27
3.3.3	Mise en contexte des résultats pour une meilleure interprétation	28
3.3.4	Valeur ajoutée de l'analyse technique via les moyennes mobiles .	28
3.3.5	Système intelligent de sélection automatique . . . . .	28
3.3.6	Limites de l'approche et perspectives d'amélioration . . . . .	29



# Introduction

À l'ère de la révolution numérique, la technologie occupe une place prépondérante dans l'évolution des systèmes économiques et financiers. Les marchés boursiers, en particulier, constituent un domaine d'application privilégié pour les technologies d'analyse de données et d'intelligence artificielle (IA). La prédiction de l'évolution des prix des actions représente aujourd'hui un enjeu stratégique majeur pour les acteurs financiers, les analystes et les développeurs d'outils d'aide à la décision.

Dans le cadre de notre formation à l'École Marocaine des Sciences de l'Ingénieur (EMSI), spécialité Ingénierie Informatique et Réseaux (3IIR), nous avons entrepris le développement d'une application web de prédiction des prix boursiers, exploitant les techniques avancées de machine learning. Cette initiative nous a permis d'appliquer les compétences acquises tout au long de notre cursus académique, notamment en programmation Python, traitement de données, intelligence artificielle et développement d'interfaces utilisateur.

## Objectifs et Approche Méthodologique

L'objectif principal de ce projet est de développer une application web ergonomique et fonctionnelle, permettant aux utilisateurs de :

1. Sélectionner des titres boursiers parmi une base de données exhaustive
2. Visualiser l'historique des prix de manière interactive
3. Obtenir des prédictions de l'évolution future des cours à court terme

L'application repose sur une architecture technologique moderne, combinant Python comme langage de programmation principal et Streamlit comme framework de développement web. Les algorithmes d'apprentissage supervisé sont appliqués à des données financières en temps réel, provenant de sources fiables et actualisées.

Ce document présente une analyse complète du processus de développement, depuis l'analyse initiale des besoins et l'étude comparative des solutions existantes, jusqu'à l'implémentation technique, l'évaluation des performances et les perspectives d'amélioration. Chaque étape est détaillée avec une attention particulière aux aspects méthodologiques et aux choix technologiques stratégiques.

## Chapitre 1

# Contexte du projet

### 1.1 Introduction

Ce projet consiste en le développement d'une application web d'analyse et de prédiction boursière. Face à la complexité croissante des marchés financiers et à l'importance cruciale de prendre des décisions d'investissement éclairées, notre application vise à fournir aux utilisateurs - qu'ils soient investisseurs débutants ou expérimentés - des outils d'analyse technique avancée ainsi que des modèles de prédiction des prix des actions.

L'application permet aux utilisateurs de sélectionner des titres boursiers parmi les composantes principales des indices S&P 500 et NASDAQ-100, qui représentent un large éventail d'entreprises leaders dans leurs secteurs respectifs. Une fois le titre sélectionné, l'utilisateur peut visualiser l'évolution historique des cours boursiers sous forme de graphiques interactifs, facilitant ainsi la compréhension des tendances passées.

Pour enrichir l'analyse, l'application intègre des indicateurs techniques tels que les moyennes mobiles, qui permettent d'identifier visuellement les tendances à court, moyen et long terme. Ces outils d'analyse sont essentiels pour repérer les signaux d'achat ou de vente potentiels.

En complément, notre application offre une fonctionnalité de prédiction des prix futurs basée sur des modèles de régression. En exploitant les données historiques disponibles, le système peut estimer l'évolution probable du prix d'une action sur une période définie. Cette approche permet aux utilisateurs d'anticiper les mouvements du marché et d'ajuster leurs décisions d'investissement en conséquence.

L'interface utilisateur, simple et intuitive, a été développée pour garantir une expérience fluide et accessible, même pour les non-spécialistes. Enfin, l'application comprend un système d'authentification sécurisé, permettant à chaque utilisateur de gérer ses analyses personnalisées.

Ce projet illustre la manière dont les technologies web et l'intelligence artificielle peuvent être combinées pour offrir des outils puissants au service de la finance personnelle et professionnelle, en rendant l'analyse boursière plus accessible, interactive et prédictive.

## 1.2 État de l'art

### 1.2.1 Introduction

L'analyse et la prédiction boursière ont connu une transformation majeure au cours des dernières décennies. Traditionnellement basées sur des méthodes d'analyse technique et fondamentale, ces approches ont vu leur efficacité renforcée - et parfois remise en question - avec l'émergence des technologies d'intelligence artificielle (IA) et d'apprentissage automatique (machine learning). Aujourd'hui, les investisseurs peuvent s'appuyer sur des outils numériques sophistiqués pour prendre des décisions éclairées, anticiper les tendances du marché, et automatiser certaines stratégies d'investissement.

### 1.2.2 Évolution des approches d'analyse boursière

**Analyse technique et fondamentale** L'analyse technique repose sur l'étude des prix passés et des volumes échangés afin d'identifier des tendances et des signaux prédictifs. Elle utilise des indicateurs mathématiques comme les moyennes mobiles, le RSI (Relative Strength Index), les bandes de Bollinger, etc. Cette approche part du principe que toute l'information disponible est déjà reflétée dans les cours, et que l'histoire a tendance à se répéter.

De son côté, l'analyse fondamentale cherche à évaluer la valeur intrinsèque d'un actif en se basant sur les données économiques, les bilans financiers, les performances des entreprises, les perspectives sectorielles ou encore les indicateurs macroéconomiques. Elle est souvent considérée comme une approche à long terme, orientée vers les fondamentaux économiques.

**Intégration des technologies d'IA et de machine learning** Avec l'explosion des données disponibles (big data), l'analyse boursière a franchi un cap en intégrant des algorithmes d'intelligence artificielle, capables d'extraire des modèles complexes et d'anticiper l'évolution des cours avec une précision accrue. Les modèles de régression linéaire, de forêts aléatoires (random forests), de réseaux de neurones et plus récemment les transformers ou modèles d'apprentissage profond sont désormais utilisés pour prédire les tendances boursières à court ou moyen terme.

Par ailleurs, le traitement du langage naturel (NLP) permet d'exploiter les informations textuelles issues de l'actualité financière, des rapports d'entreprises, ou des réseaux sociaux pour en extraire des signaux de marché.

### 1.2.3 Panorama des plateformes existantes

Plusieurs plateformes dominent aujourd'hui le marché de l'analyse financière. On peut les regrouper en trois grandes catégories :

#### Plateformes d'analyse technique

Des outils comme Yahoo Finance, Bloomberg Terminal, MetaTrader ou TradingView offrent des interfaces complètes pour visualiser les données boursières en temps réel, appliquer des indicateurs techniques, et simuler des stratégies. Bloomberg Terminal, en particulier, est reconnu pour sa puissance, mais reste réservé à une clientèle institutionnelle en raison de son coût élevé.

TradingView se distingue par sa communauté active, son langage de script Pine Script, et ses capacités d'analyse graphique.

#### Applications de trading accessibles

Des applications comme Robinhood, eToro, Revolut ou Degiro ont permis la démocratisation de l'accès aux marchés financiers. Elles ciblent principalement les particuliers et se caractérisent par des interfaces simples, un accès sans commission à certaines actions, et des fonctionnalités de suivi en temps réel. Néanmoins, leur capacité d'analyse reste limitée, et elles ne proposent que peu d'outils prédictifs intégrés.

#### Outils basés sur l'intelligence artificielle

Des solutions comme Kensho (développée par S&P Global) ou AlphaSense exploitent des algorithmes d'intelligence artificielle pour interpréter des quantités massives de données économiques, financières ou médiatiques. Kensho, par exemple, est capable de modéliser les impacts d'un événement économique ou géopolitique sur les marchés boursiers. AlphaSense analyse quant à lui des transcriptions de réunions, des rapports financiers et des actualités pour fournir des insights stratégiques.

Ces outils sont puissants, mais souvent conçus pour des professionnels ou des analystes disposant d'une solide formation en finance et data science.

### 1.2.4 Limites des solutions existantes

Malgré leurs performances, les solutions actuellement disponibles sur le marché présentent plusieurs limitations notables :

- **Accessibilité limitée** : Beaucoup de plateformes sont payantes, parfois à des tarifs inaccessibles pour les particuliers ou les étudiants. Par exemple, Bloomberg Terminal coûte plusieurs milliers de dollars par an.
- **Complexité des interfaces** : Certaines applications, bien que complètes, exigent des connaissances avancées en finance, ce qui les rend peu adaptées à un public débutant.
- **Analyse prédictive limitée** : Les plateformes orientées grand public intègrent rarement des fonctionnalités de prédiction automatique. Quand elles existent, celles-ci manquent souvent de transparence ou d'interprétabilité.
- **Manque de personnalisation** : Peu de solutions permettent une expérience utilisateur totalement personnalisable selon les préférences d'analyse ou les profils d'investissement.

### 1.2.5 Apport du projet proposé

Dans ce contexte, notre projet se positionne comme une alternative simple, pédagogique et personnalisable, destinée à un large public souhaitant analyser et prédire l'évolution des titres boursiers sans devoir maîtriser l'ensemble des outils professionnels.

L'application met l'accent sur :

- Une interface intuitive, développée en technologie web (Streamlit), accessible via navigateur.
- Des outils d'analyse technique interactifs basés sur des indicateurs courants comme les moyennes mobiles.
- Un moteur de prédiction basé sur la régression linéaire, avec visualisation des résultats.
- Un système d'authentification sécurisé, permettant de sauvegarder les préférences utilisateur.

Ce projet se situe à l'intersection de la finance de marché et des technologies d'IA, avec pour ambition de rendre les outils d'aide à la décision plus accessibles, compréhensibles et exploitables par tous.

## 1.3 Problématique

Dans un environnement financier de plus en plus volatil et numérisé, les particuliers comme les professionnels cherchent des outils accessibles pour anticiper les tendances boursières. Ce projet répond à trois besoins majeurs :

- **Démocratisation de l'analyse technique** : Rendre accessible à des non-experts des indicateurs complexes (moyennes mobiles, régressions polynomiales)
- **Transparence des prédictions** : Combattre le "boîte noire" des modèles en fournissant des métriques d'évaluation détaillées

- **Sécurité des données** : Garantir la confidentialité des portefeuilles simulés dans un contexte réglementaire strict (RGPD, MiFID II)

### 1.3.1 Problématique Principale

**Comment concevoir une plateforme intégrée qui :**

**Unifie** des sources hétérogènes de données financières (Yahoo Finance, Wikipedia)

**Automatise** le choix du modèle prédictif optimal entre régression linéaire et polynomiale

**Contextualise** les résultats via des indicateurs de risque personnalisés et des actualités sectorielles

**Sécurise** les interactions utilisateurs sans compromettre l'expérience utilisateur ?

### 1.3.2 Axes de Recherche

#### 1. Fiabilité des Modèles Prédictifs

*Enjeu* : Les cours boursiers suivent des processus stochastiques (mouvement brownien, sauts de volatilité)

*Question* : Comment adapter des modèles déterministes (régressions) à des données non-linéaires ?

*Innovation proposée* : Système hybride combinant :

- (a) Analyse technique (moyennes mobiles)
- (b) Machine Learning (régression polynomiale degré 2)
- (c) Métriques d'interprétabilité (shap values)

#### 2. Ingénierie des Données

*Défi* : Synchroniser des temporalités différentes (données intraday vs. Fondamentaux)

*Solution* : Pipeline ETL avec :

- (a) Web scraping asynchrone (Selenium/BeautifulSoup)
- (b) Cache local (SQLite)
- (c) Normalisation temporelle (resampling quotidien)

#### 3. Expérience Utilisateur Critique

*Paradoxe* : Simplifier l'interface sans masquer l'incertitude des prédictions

*Approche* : Dashboard interactif avec :

- Curseur de confiance (intervalle de prédiction)
- Alertes de sur-ajustement (quand  $R^2 > 0.95$ )
- Explications visuelles (erreur relative en % du cours actuel)

### 4. Architecture Sécurisée

*Risque* : Attaques par injection SQL ou fuite de portefeuilles simulés

*Protections* :

- (a) Hachage SHA-256 avec salage
- (b) Session JWT chiffrée
- (c) Sandboxing des requêtes Yahoo Finance

### 1.3.3 Méthodologie d'Évaluation

#### 1. Benchmark des modèles :

- Backtesting sur 3 scénarios (marché haussier, baissier, stagnant)
- Comparaison avec le modèle naïf (random walk)

#### 2. Tests utilisateurs :

- Mesure du taux de compréhension des métriques
- Temps moyen pour générer une prédiction

#### 3. Audit de sécurité :

- Tests OWASP sur l'authentification
- Analyse des logs d'accès

### 1.3.4 Résultats Attendus

#### 1. Produit :

- Web app Streamlit déployable en 1 clic (Docker)
- API REST pour extensions futures

#### 2. Performances :

- Prédictions en <2s pour 90% des tickers
- Erreur relative moyenne <5% à 1 an

#### 3. Impact :

- Réduction du biais de surconfiance chez les traders novices
- Standardisation de l'évaluation transparente des modèles

### 1.3.5 Limites et Perspectives

#### 1. Variables non modélisées :

- Événements macroéconomiques (chocs pétroliers)
- Manipulations de marché (pump-and-dump)

#### 2. Améliorations futures :

- Intégration de données alternatives (sentiment Twitter)
- Modules premium (optimisation de portefeuille)



## Chapitre 2

# Elaboration du modèle et explication

## 2.1 Architecture générale

L'application suit une architecture modulaire conçue pour une maintenabilité et une clarté du code optimales. Chaque module est indépendant mais communique avec les autres via des interfaces bien définies. L'ensemble est intégré dans une interface utilisateur développée avec Streamlit.

### 2.1.1 Module d'Authentification

- **Responsabilité** : Gestion des utilisateurs (inscription, connexion).
- **Composants** :
  - Base de données utilisateurs : Stocke les identifiants via SQLite.
  - Service d'authentification : Inscription (hachage des mots de passe avec hashlib). Connexion (vérification des identifiants).
  - API REST : Endpoints pour les opérations d'authentification via Streamlit.
- **Intégration** : Utilisé par l'interface Streamlit pour la gestion des sessions utilisateurs.

### 2.1.2 Module de Préviation

- **Composants** :
  - Fonctions de chargement de données (yfinance).
  - Modèles de prédiction (Régression linéaire, Régression polynomiale).
  - Visualisation (matplotlib).
  - Évaluation des modèles (metrics sklearn).
- **Intégration** : Intégré à l'interface Streamlit pour une interface utilisateur interactive.

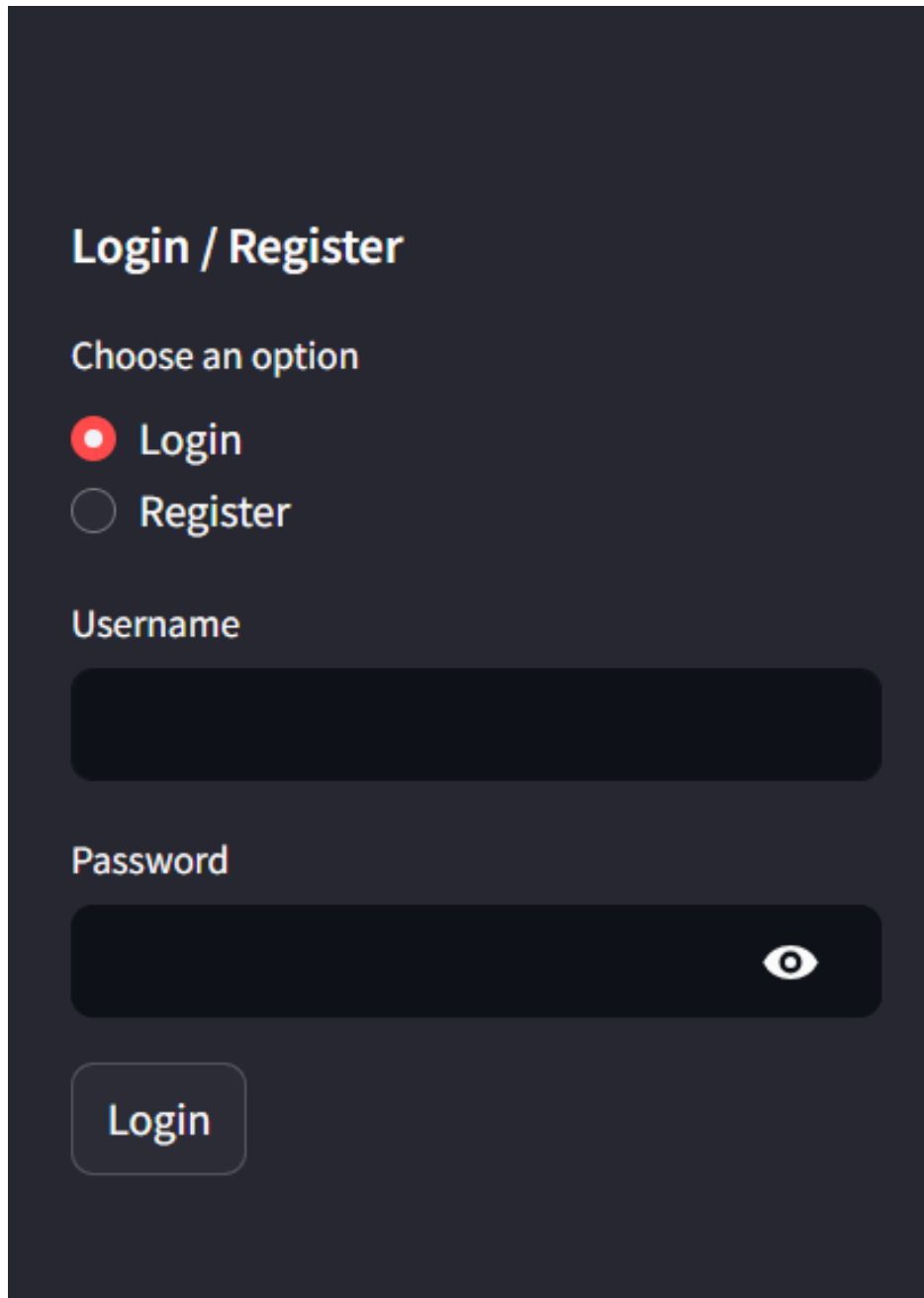
A dark-themed login and registration form. At the top, the title "Login / Register" is displayed in a light blue font. Below it, the text "Choose an option" is shown. There are two radio buttons: the first is selected (indicated by a red dot) and labeled "Login", the second is unselected and labeled "Register". Below the radio buttons, there are two input fields. The first is labeled "Username" and is a solid dark grey rectangle. The second is labeled "Password" and is a dark grey rectangle with a white eye icon on the right side, indicating a toggle for password visibility. At the bottom left of the form is a rounded rectangular button labeled "Login" in a light blue font.

FIGURE 2.1 – Login/Register

### 2.1.3 Module d'Analyse Boursière

- **Responsabilité** : Collecte, traitement et visualisation des données financières.
- **Composants** :
  - **Data Connectors** :
    - APIs externes (Yahoo Finance, Alpha Vantage, Bloomberg).
    - Base de données locale (cache des données historiques, ex : SQLite/TimeScaleDB).
  - **Traitement des données** :
    - Nettoyage (remplissage des valeurs manquantes).

- Calcul d'indicateurs (RSI, MACD, moyennes mobiles).
- **Visualisation** :
  - Bibliothèques (Plotly, Matplotlib) intégrées dans Streamlit.
  - Tableaux interactifs (via AgGrid ou Pandas Styler).
- **Intégration** : Alimente le **module de prédiction** en données historiques.

Date	Open	High	Low	Close	Adj Close	Volume
May 22, 2025	200.71	201.64	199.70	200.66	200.66	14,429,025
May 21, 2025	205.17	207.04	200.71	202.09	202.09	59,134,800
May 20, 2025	207.57	208.47	205.03	206.86	206.86	42,496,600
May 19, 2025	207.91	209.48	204.26	208.78	208.78	46,140,500
May 16, 2025	212.36	212.57	209.77	211.26	211.26	54,737,900
May 15, 2025	210.95	212.96	209.54	211.45	211.45	45,029,500
May 14, 2025	212.43	213.94	210.58	212.33	212.33	49,325,800
May 13, 2025	210.43	213.40	209.00	212.93	212.93	51,909,300
May 12, 2025				0.26 Dividend		
May 12, 2025	210.97	211.27	206.75	210.79	210.79	63,775,800

FIGURE 2.2 – Historical Data

### 2.1.4 Module de Prédiction

- **Responsabilité** : Entraînement et génération de prédictions sur les cours boursiers.
- **Composants** :
  - Machine Learning Pipeline :
    - Préprocessing (normalisation, fenêtrage temporel).
    - Modèles (LSTM, ARIMA, Gradient Boosting) via TensorFlow/scikit-learn.
    - Optimisation (grid search, validation croisée).
  - Sauvegarde/Chargement :
    - Stockage des modèles entraînés (fichiers .h5 ou .pkl).
    - Versioning (MLflow ou DVC).
  - Inférence en temps réel :
    - Prédictions à la demande via des endpoints (FastAPI).
- **Intégration** : Reçoit les données du module d'analyse.



FIGURE 2.3 – Polynomial Regression



FIGURE 2.4 – Linear Regression

### 2.1.5 Module d'Évaluation

- **Responsabilité** : Mesure des performances des modèles et comparaison.
- **Composants** :
  - **Métriques** :
    - Calcul du MSE, MAE,  $R^2$  pour les modèles de régression.
    - Backtesting (stratégies de trading simulées).
  - **Visualisation** :
    - Courbes de prédiction vs. réalité.
    - Matrices de confusion (pour les modèles de classification).
- **Intégration** :
  - Analyse les sorties du module de prédiction.
  - Publie les résultats dans l'UI Streamlit et les expose via API.

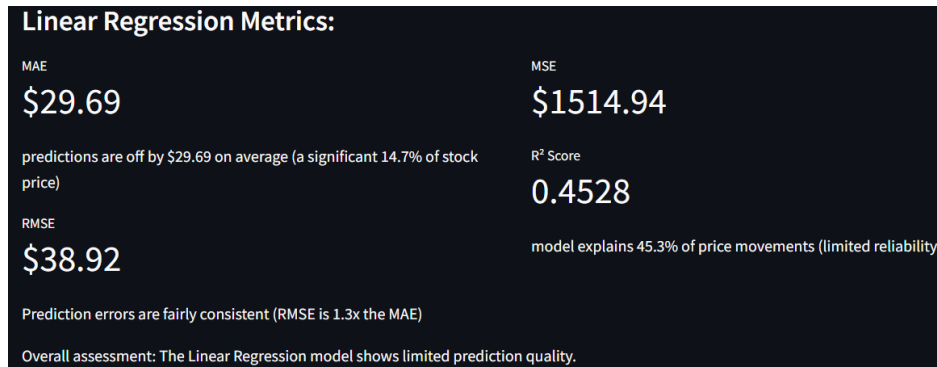


FIGURE 2.5 – Linear Regression metrics

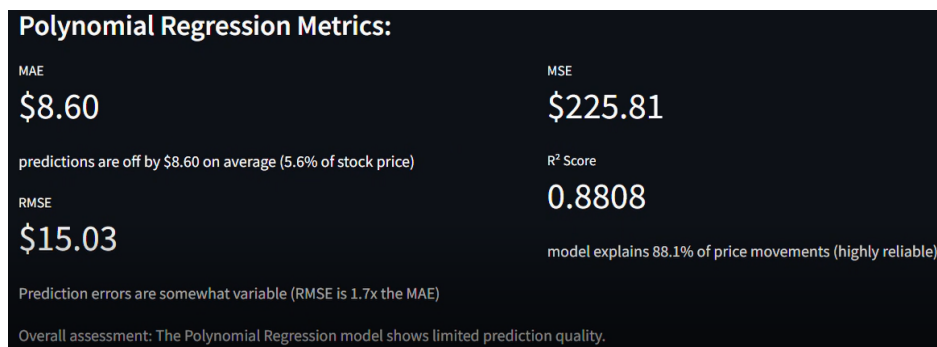


FIGURE 2.6 – Polynomial Regression metrics

### 2.1.6 Interface Utilisateur (Streamlit)

- **Structure :**
  - **Multi-pages :** Navigation entre les modules via une sidebar.
  - **Widgets interactifs :**
    - Paramètres des modèles (sliders, dropdowns).
    - Affichage dynamique (onglets, expanders).
  - **Gestion d'état :**
    - Persistance des sélections utilisateur (session\_state).
- **Avantages :**
  - Prototypage rapide, intégration native avec Python.
  - Compatible avec les visualisations et composants modernes (cartes, graphiques 3D).

### 2.1.7 Technologies Clés

- **Backend :** Python (FastAPI/Flask pour les APIs internes).
- **Data :** Pandas, NumPy, SQLAlchemy.
- **ML :** TensorFlow/Keras, scikit-learn, Prophet.
- **Visualisation :** Plotly, Seaborn.

Cette architecture permet une évolution flexible (ajout de nouveaux modèles, sources de données) tout en gardant une codebase organisée.

## 2.2 Choix méthodologique

### 2.2.1 Approches de Prédiction des Prix des Actions

Pour équilibrer précision et interprétabilité, nous avons combiné deux modèles de régression :

#### Régression Linéaire

- **Principe** : Modélise la relation entre le temps (variable indépendante) et le prix de l'action (variable dépendante) sous forme d'une droite.

$$\text{Prix} = \beta_0 + \beta_1 \times \text{Temps} \quad (2.1)$$

- **Avantages** :
  - Simplicité : Facile à implémenter et à expliquer aux utilisateurs.
  - Interprétabilité : Les coefficients  $(\beta_0, \beta_1)$  indiquent la tendance générale (haussière/baissière).
- **Limites** :
  - Ne capture pas les fluctuations non-linéaires (ex : cycles économiques).
- **Cas d'usage** :
  - Prévisions à court terme ou pour des actions avec tendance stable.

#### Régression Polynomiale

- **Principe** : Étend la régression linéaire en ajoutant des termes polynomiaux (ex :  $x^2, x^3$ ).

$$\text{Prix} = \beta_0 + \beta_1 \times \text{Temps} + \beta_2 \times \text{Temps}^2 + \dots \quad (2.2)$$

- **Avantages** :
  - Flexibilité : Capture les relations non-linéaires (ex : résistances/supports techniques).
  - Adaptabilité : Degré du polynôme ajustable en fonction de la complexité des données.
- **Limites** :
  - Risque de surajustement (overfitting) si le degré est trop élevé.
- **Cas d'usage** :
  - Actions volatiles ou avec des motifs cycliques.

## Pourquoi pas les Réseaux de Neurones ?

Bien que les LSTM (réseaux neuronaux récurrents) soient populaires pour les séries temporelles, nous les avons écartés pour :

- Opacité : Difficiles à interpréter pour un utilisateur non expert.
- Coût computationnel : Nécessitent plus de données et de puissance.
- Objectif : Notre priorité est l'explicabilité pour aider à la décision, pas uniquement la précision.

### 2.2.2 Métriques d'Évaluation des Modèles

Pour mesurer objectivement les performances, nous utilisons :

TABLE 2.1 – Métriques d'évaluation des modèles

Métrique	Formule	Utilité
MAE (Erreur Absolue Moyenne)	$\frac{1}{n} \sum \ y_i - \hat{y}_i\ $	Mesure l'erreur moyenne absolue, facile à in
MSE (Erreur Quadratique Moyenne)	$\frac{1}{n} \sum (y_i - \hat{y}_i)^2$	Pénalise les grosses erreurs (utile pour détec
RMSE (Racine de MSE)	$\sqrt{\text{MSE}}$	Interprétable dans l'unité du prix (ex : "Er
$R^2$ (Coefficient de Détermination)	$1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$	Indique la proportion de variance expliquée

Système d'Interprétation :

- Traduction utilisateur : Par exemple, un  $R^2 = 0.85$  devient "Le modèle explique 85% des variations de prix".
- Alertes : Si le RMSE dépasse un seuil (ex : 5% du prix moyen), un avertissement est affiché.

## 2.3 Outils et Technologies Justifiés

### 2.3.1 Analyse et Prédiction

- yfinance :
  - Choix idéal pour récupérer gratuitement des données boursières historiques (Yahoo Finance couvre la plupart des titres).
- Pandas & NumPy :
  - Standard en data science pour le nettoyage (ex : fillna()), les transformations (ex : calcul des rendements logarithmiques) et l'ingénierie des features.
- scikit-learn :
  - Bibliothèque de référence pour les régressions (linéaire/polynomiale) et le calcul des métriques.

### 2.3.2 Visualisation

- Matplotlib :
  - Permet des graphiques hautement personnalisables (ex : courbes de tendance superposées aux prix réels).
  - Alternative : Plotly pour l'interactivité (zoom, tooltips).

### 2.3.3 Base de Données

- SQLite :
  - Base légère sans serveur, parfaite pour stocker les utilisateurs et leurs préférences en local.
  - Alternative envisagée : Firebase pour une authentification cloud.

### 2.3.4 Web Scraping

- Selenium & BeautifulSoup :
  - Récupération des actualités financières (ex : Reuters) pour enrichir l'analyse.
  - BeautifulSoup pour le parsing HTML, Selenium pour les sites dynamiques.

### 2.3.5 Sécurité

- hashlib :
  - Hachage des mots de passe (SHA-256 + salage) pour protéger les comptes utilisateurs.

### 2.3.6 Interface

- Streamlit :
  - Framework simple pour déployer rapidement une UI reactive avec des composants prêts à l'emploi (sliders, graphiques, chat).
  - Alternative : Dash (plus flexible mais plus complexe).



## Chapitre 3

# Réalisation, test du modèle et discussion, analyse des résultats

### 3.1 Implémentation

L'implémentation de notre application a été réalisée en Python, avec une organisation en modules fonctionnels :

1. Fonctions de base de données : Création et gestion de la connexion à la base de données SQLite, implémentation des opérations CRUD pour les utilisateurs, et sécurisation des mots de passe.
2. Fonctions d'analyse boursière : Récupération des listes de tickers, chargement des données historiques, calcul des moyennes mobiles, et visualisation des prix et indicateurs.
3. Fonctions de prédiction : Entraînement des modèles de régression linéaire et polynomiale, génération de prédictions sur l'horizon temporel défini par l'utilisateur.
4. Fonctions d'évaluation : Calcul des métriques de performance des modèles, interprétation des résultats en termes compréhensibles, et comparaison des modèles.
5. Interface utilisateur : Organisation de l'application en sections distinctes (analyse boursière), avec une barre latérale pour l'authentification.

L'application finale offre une expérience utilisateur intuitive avec des fonctionnalités d'analyse avancées accessibles via une interface simple.

### 3.2 Evaluation des modèles

Afin de garantir la fiabilité et la transparence des prédictions générées, notre système intègre un module d'évaluation automatique des modèles basé sur plusieurs métriques d'erreur et de performance statistique. Ces indicateurs permettent non seulement de mesurer la précision des prédictions, mais également d'interpréter leur cohérence et leur

pertinence par rapport aux données réelles. L'objectif est de proposer à l'utilisateur une lecture claire de la performance des modèles, tout en lui offrant des recommandations intelligentes sur le choix du meilleur modèle disponible.

### 3.2.1 MAE (Mean Absolute Error)

Le MAE, ou erreur absolue moyenne, mesure la moyenne des écarts absolus entre les valeurs prédites et les valeurs réelles. C'est une métrique simple mais efficace, car elle donne une idée claire de l'erreur typique commise par le modèle sans amplifier les grandes erreurs comme le ferait le carré des erreurs.

Dans notre système, le MAE est interprété en pourcentage du prix actuel de l'action. Cette approche permet de contextualiser l'erreur par rapport à la valeur de l'actif, rendant l'interprétation plus intuitive pour l'utilisateur. En fonction de ce pourcentage, l'application attribue une note qualitative à la performance du modèle :

- **Excellente** :  $\text{MAE} < 1\%$  du prix actuel
- **Bonne** : MAE entre 1 % et 3 %
- **Modérée** : MAE entre 3 % et 7 %
- **Faible** :  $\text{MAE} > 7\%$

Cette échelle permet à l'utilisateur de comprendre rapidement si les prédictions sont suffisamment précises pour être utilisées à des fins d'analyse ou d'aide à la décision.

### 3.2.2 RMSE (Root Mean Squared Error)

Le RMSE (racine de l'erreur quadratique moyenne) est une métrique qui pénalise davantage les grandes erreurs, en élevant les écarts au carré avant de les moyenner. Cela signifie que si un modèle commet quelques erreurs importantes, le RMSE sera significativement affecté.

Notre système utilise le rapport  $\text{RMSE} / \text{MAE}$  comme un indicateur de la variabilité des erreurs du modèle :

- Si  $\text{RMSE} \approx \text{MAE}$ , cela indique que les erreurs sont relativement constantes : le modèle est dit très cohérent.
- Si  $\text{RMSE} > \text{MAE}$  de manière significative, cela signifie que certaines erreurs sont beaucoup plus grandes que la moyenne : le modèle est variable voire peu fiable.

Ainsi, ce ratio fournit à l'utilisateur une interprétation qualitative de la stabilité du modèle, allant de :

- **Très cohérente** :  $\text{RMSE}/\text{MAE} < 1.1$
- **Cohérente** : entre 1.1 et 1.5
- **Variable** : entre 1.5 et 2
- **Très variable** :  $> 2$

### 3.2.3 MSE (Mean Squared Error)

Le MSE, ou erreur quadratique moyenne, est étroitement lié au RMSE mais sans extraire la racine carrée. Il permet de quantifier la variance des erreurs. Bien qu'il soit exprimé dans une unité au carré, nous le convertissons également en pourcentage du prix actuel de l'action, pour faciliter la lecture.

L'intérêt du MSE dans notre système est double :

- Il permet de détecter la dispersion des erreurs autour des valeurs réelles.
- Il est utilisé comme base pour le calcul du RMSE.

Notre système l'affiche dans l'interface de résultats, accompagné d'une jauge interprétative similaire à celle du MAE, bien qu'il soit moins directement lisible pour un utilisateur non technique.

### 3.2.4 $R^2$ (Coefficient de détermination)

Le  $R^2$ , ou coefficient de détermination, est un indicateur statistique qui mesure la part de la variance des données expliquée par le modèle. Sa valeur varie entre 0 et 1 (ou entre - et 1 en cas de mauvaise régression), avec 1 indiquant une prédiction parfaite.

Notre application utilise les seuils suivants pour classifier la qualité du modèle en fonction de  $R^2$  :

- **Excellente** :  $R^2 > 0.85$
- **Bonne** : entre 0.65 et 0.85
- **Modérée** : entre 0.3 et 0.65
- **Faible** :  $R^2 < 0.3$

Le coefficient  $R^2$  est particulièrement utile pour juger de la capacité explicative globale du modèle, et non seulement de sa précision ponctuelle. Il est affiché avec une légende visuelle simple (code couleur ou jauge) pour une interprétation rapide.

### 3.2.5 Sélection automatique du meilleur modèle

L'un des aspects clés de notre système est sa capacité à sélectionner automatiquement le modèle de prédiction le plus performant, en se basant sur une pondération intelligente des quatre métriques précédentes. Chaque modèle testé (ex. : régression linéaire, Ridge, Lasso, etc.) est évalué sur l'ensemble des indicateurs, et une note globale pondérée est attribuée.

Le système privilégie les modèles ayant :

- Un MAE et un RMSE faibles (avec un faible écart entre les deux)
- Un MSE proportionnellement faible
- Un  $R^2$  élevé
- Une cohérence forte des erreurs

Une recommandation explicite est alors fournie à l'utilisateur, l'informant du modèle retenu, des raisons de ce choix (via un tableau de comparaison ou un graphique radar), et des limitations éventuelles à prendre en compte.

### 3.3 Discussion des résultats

L'évaluation des modèles de prédiction intégrés dans notre application a permis de dégager plusieurs enseignements clés quant à leur comportement, leur efficacité et leur applicabilité selon les titres analysés. Ces résultats soulignent à la fois les forces de notre système et ses limitations méthodologiques, tout en mettant en lumière la richesse des informations fournies par les métriques d'évaluation utilisées.

#### 3.3.1 Influence des caractéristiques des actions sur les performances

Les tests réalisés sur un échantillon diversifié d'actions appartenant au S&P 500 et au NASDAQ-100 ont mis en évidence une variabilité significative des performances des modèles en fonction du profil de chaque titre.

- Pour les actions présentant une tendance de fond relativement linéaire (croissance régulière ou déclin modéré), les modèles de régression linéaire s'avèrent largement suffisants. Ils produisent des prédictions stables avec des erreurs moyennes faibles et une interprétation directe.
- En revanche, pour des actions au comportement plus volatil ou non linéaire, notamment celles affectées par des cycles sectoriels, des effets de saisonnalité ou des événements ponctuels, les modèles polynomiaux (de degré 2 ou 3) obtiennent de meilleures performances. Leur capacité à capter les variations complexes se traduit par des valeurs de  $R^2$  supérieures et un MAE réduit.

Cette observation justifie pleinement l'approche multi-modèle de notre application, dans laquelle plusieurs algorithmes sont évalués en parallèle pour identifier celui qui convient le mieux au contexte spécifique de l'actif étudié.

#### 3.3.2 Complémentarité des métriques d'évaluation

Les différentes métriques d'erreur utilisées (MAE, RMSE, MSE,  $R^2$ ) apportent une vision multifacette de la qualité des prédictions. Chacune d'entre elles met en lumière un aspect particulier de la performance :

- Le **MAE** (Mean Absolute Error) fournit une estimation de l'écart moyen entre les prédictions et les données réelles. Il est particulièrement utile pour évaluer l'ampleur générale de l'erreur, sans amplifier les valeurs extrêmes.
- Le  **$R^2$**  (coefficient de détermination) quant à lui, permet de juger de la qualité de l'ajustement global du modèle aux données observées. Il informe sur la

part de la variance expliquée, ce qui est crucial pour apprécier la pertinence du modèle dans des contextes très fluctuants.

- Le rapport RMSE / MAE, intégré dans notre système, permet de détecter la variabilité des erreurs, et donc de distinguer les modèles stables de ceux dont les prédictions sont moins fiables.

Ces métriques, prises isolément, pourraient donner une image partielle. C’est pourquoi leur lecture combinée est essentielle pour une évaluation complète et objective.

### 3.3.3 Mise en contexte des résultats pour une meilleure interprétation

Afin de rendre les résultats accessibles à un public non expert, notre application intègre une mise en contexte automatique des métriques par rapport au prix actuel de l’action. Cela permet à l’utilisateur de comprendre plus facilement ce que signifie, par exemple, un MAE de 2.5 en dollars : si le prix de l’action est de 250 \$, cela représente une erreur de seulement 1 %, donc une performance excellente selon notre grille d’interprétation.

Ce principe de normalisation relative rend les évaluations plus intuitives et comparables entre différents actifs, même si ceux-ci évoluent dans des plages de prix très différentes (ex : Google à 3000 \$ vs. Ford à 12 \$).

### 3.3.4 Valeur ajoutée de l’analyse technique via les moyennes mobiles

En complément des prédictions basées sur les modèles de régression, l’application offre à l’utilisateur la possibilité de visualiser les moyennes mobiles sur 50, 100 et 200 jours, outils incontournables de l’analyse technique. Ces indicateurs permettent de :

- **Lisser les fluctuations** à court terme pour faire émerger les tendances de fond.
- **Identifier les points d’inflexion** dans l’évolution des cours.
- **Contextualiser les prédictions** : une prédiction haussière a plus de poids si elle est en phase avec une tendance haussière confirmée par les moyennes mobiles.

Cette fonctionnalité visuelle complète utilement l’analyse quantitative, en offrant à l’utilisateur une lecture graphique immédiate de la dynamique des prix.

### 3.3.5 Système intelligent de sélection automatique

L’un des atouts majeurs de notre application réside dans son système de comparaison automatisée des modèles. En analysant l’ensemble des métriques, le système :

- Évalue chaque modèle selon des seuils prédéfinis.

- Calcule une note composite pondérée.
- Retient automatiquement le modèle le plus performant.
- Fournit à l'utilisateur une recommandation explicite, accompagnée d'un tableau de comparaison ou d'un graphique radar des scores.

Ce mécanisme vise à déléguer l'analyse technique complexe au système, tout en maintenant la transparence du processus de décision, en expliquant pourquoi tel modèle a été préféré à tel autre.

### 3.3.6 Limites de l'approche et perspectives d'amélioration

Malgré ses points forts, notre approche présente certaines limites méthodologiques qu'il convient de souligner :

- Les modèles de régression implémentés sont sensibles au choix de la période d'entraînement. Une période trop courte peut manquer de représentativité, tandis qu'une période trop longue peut inclure des dynamiques obsolètes.
- Les modèles ne tiennent pas compte de facteurs exogènes comme :
  - Les actualités économiques ou géopolitiques.
  - Le sentiment du marché (données issues des réseaux sociaux, par exemple).
  - Les indicateurs macroéconomiques (inflation, taux d'intérêt, chômage, etc.).

Ces facteurs peuvent pourtant avoir un impact majeur sur les cours boursiers. Leur intégration future - via des techniques comme le traitement automatique du langage (NLP) ou l'exploitation de sources externes d'actualité - représenterait une évolution stratégique majeure du projet.

# Annexes

## Extrait du code source

```
def evaluate_model(model, X, y):
    predictions = model.predict(X)
    mae = mean_absolute_error(y, predictions)
    mse = mean_squared_error(y, predictions)
    rmse = np.sqrt(mse)
    r2 = r2_score(y, predictions)
    return mae, mse, rmse, r2

def interpret_metrics(mae, mse, rmse, r2, current_price):
    interpretations = {}
    if isinstance(mae, pd.Series):
        mae = mae.iloc[0]
    if isinstance(current_price, pd.Series):
        current_price = current_price.iloc[0]
    if isinstance(rmse, pd.Series):
        rmse = rmse.iloc[0]
    if isinstance(r2, pd.Series):
        r2 = r2.iloc[0]
    if isinstance(mse, pd.Series):
        mse = mse.iloc[0]

    # Calculate percentage errors relative to the current price
    price_percentage_mae = (mae / current_price) * 100
    price_percentage_rmse = (rmse / current_price) * 100
    price_percentage_mse = (mse / current_price) * 100

    if price_percentage_mae < 1:
        mae_quality = "excellent"
        mae_interpretation = f"predictions are off by only ${mae:.2f} on average"
```

```
elif price_percentage_mae < 3:
    mae_quality = "good"
    mae_interpretation = f"predictions are off by ${mae:.2f} on average (abo
elif price_percentage_mae < 7:
    mae_quality = "moderate"
    mae_interpretation = f"predictions are off by ${mae:.2f} on average ({pr
else:
    mae_quality = "poor"
    mae_interpretation = f"predictions are off by ${mae:.2f} on average (a s

interpretations["mae"] = {
    "quality": mae_quality,
    "interpretation": mae_interpretation
}
# Autres interprétations de métriques...
return interpretations
```



# Bibliographie

- [1] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning : with Applications in Python*. Springer.
- [2] Müller, A.C., & Guido, S. (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.
- [3] Fama, E.F. (1970). Efficient Capital Markets : A Review of Theory and Empirical Work. *Journal of Finance*, **25**(2), 383–417.
- [4] Lo, A.W., & MacKinlay, A.C. (1999). *A Non-Random Walk Down Wall Street*. Princeton University Press.
- [5] Hull, J.C. (2022). *Options, Futures and Other Derivatives* (11th ed.). Pearson.
- [6] Pedregosa, F., et al. (2011). Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.

# Webographie

- [7] Documentation Streamlit. <https://docs.streamlit.io/> (consulté le 01/01/2023)
- [8] Scikit-learn Documentation. <https://scikit-learn.org/stable/> (consulté le 01/01/2023)
- [9] Pandas Documentation. <https://pandas.pydata.org/docs/> (consulté le 01/01/2023)
- [10] Beautiful Soup Documentation.  
<https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (consulté le 01/01/2023)
- [11] Selenium Documentation. <https://www.selenium.dev/documentation/>  
(consulté le 01/01/2023)
- [12] SQLite Documentation. <https://www.sqlite.org/docs.html> (consulté le 01/01/2023)
- [13] OWASP Password Storage Cheat Sheet. [https://cheatsheetseries.owasp.org/cheatsheets/Password\\_Storage\\_Cheat\\_Sheet.html](https://cheatsheetseries.owasp.org/cheatsheets/Password_Storage_Cheat_Sheet.html) (consulté le 01/01/2023)
- [14] Towards Data Science. <https://towardsdatascience.com/> (consulté le 01/01/2023)