



Université de la Manouba  
École Nationale des Sciences de l'Informatique



## RAPPORT DU PROJET DE CONCEPTION ET DE DÉVELOPPEMENT

---

Sujet : Prédiction des cours boursiers  
d'entreprises IT

---

*Auteurs :*

M<sup>me</sup> Nesrine JOUINI      M. Wael KHEMAKHEM  
M. Ihebeddine MARNAOUI

*Encadrante :*

Dr. Rym BESROUR

---

Année Universitaire : 2020/2021

**Résumé** — Ce projet s'inscrit dans le cadre d'un projet de conception et de développement à l'École Nationale des Sciences de l'informatique (ENSI).

Notre projet consiste à réaliser un site web qui expose les résultats d'un modèle d'intelligence artificielle permettant de prédire les cours boursiers de différentes entreprises technologiques cotées dans la bourse NASDAQ.

**Mots clés** : Cours boursiers, Séries Temporelles, Prédiction, ARIMA, Decision Tree, Apprentissage profond récurrent, LSTM, Python, Flask, ReactsJs, Keras, Tensorflow.

---

**abstract**— This project is part of the design and development project at the National School of Computer Sciences (ENSI).

Our project consists of developing a website that exposes the results of an artificial intelligence model making it possible to predict the stock prices of different technology companies listed on the NASDAQ stock exchange.

**Key words** : Stock prices, Time Series, Prediction, ARIMA, Decision Tree, Recurrent deep learning, LSTM, Python, Flask, ReactsJs, Keras, Tensorflow.

---

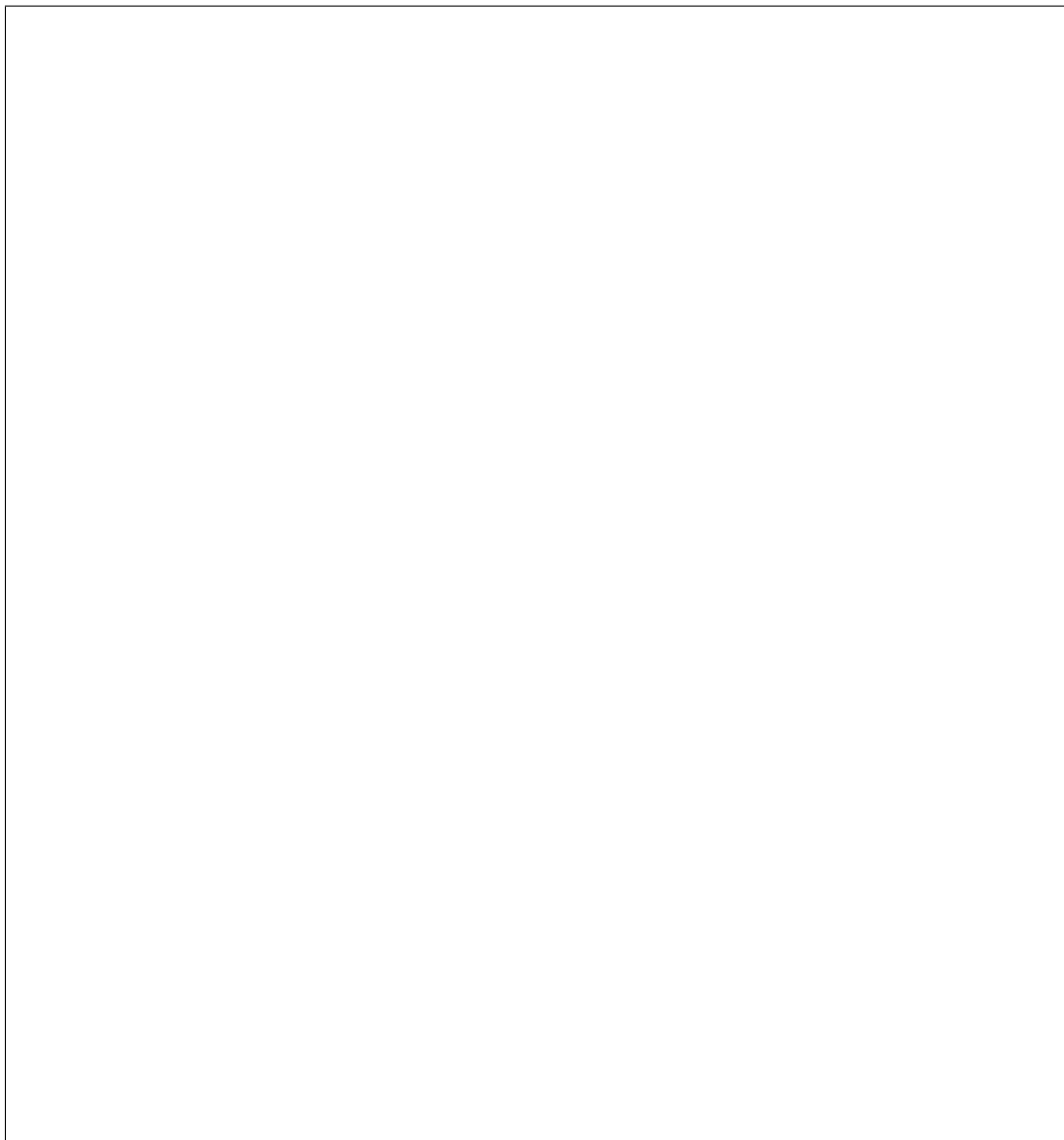
---

**ملخص** — هذا المشروع جزء من مشروع التصميم والتطوير في المدرسة الوطنية لعلوم الإعلامية يتكون مشروعنا من تطوير موقع ويب يعرض نتائج نموذج ذكاء اصطناعي يجعل من الممكن التنبؤ بأسعار أسهم شركات التكنولوجيا المختلفة المدرجة في بورصة ناسداك .

**الكلمة الرئيسية**: أسعار الأسهم ، السلاسل الزمنية ، التنبؤ، شجرة القرار، التعلم العميق المتكرر، LSTM، ARIMA، Flask، ReactsJs، Python .

---

## Appréciations et signature de l'encadrant

A large, empty rectangular box with a thin black border, intended for the appreciations and signature of the supervisor.

# Remerciements

Exprimer sa reconnaissance à ceux qui ont participé à l'élaboration de ce projet est un devoir agréable.

Nous tenons, tout d'abord, à exprimer notre profonde gratitude à notre encadrante **Docteur Rym BESSROUR** pour ses précieux conseils, sa disponibilité et son encadrement qui nous ont permis de mener à bien notre projet.

Nous tenons également à remercier vivement **Docteur Imen SNOUSSI** pour son aide précieuse, ses remarques constructives et sa disponibilité.

Nous remercions également les membres de jury d'avoir eu la bienveillance de nous accorder leur attention et d'évaluer notre présentation.

Nous tenons, enfin, à remercier toute personne ayant contribué dans la réalisation de ce projet.

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 Etat de l'art</b>	<b>3</b>
1.1 Définitions générales pour le projet . . . . .	3
1.2 Présentation générale du projet . . . . .	4
1.2.1 Cadre du projet . . . . .	4
1.2.2 Problématique . . . . .	4
1.2.3 Solutions existantes . . . . .	4
1.2.4 Critique de l'existant . . . . .	6
1.2.5 Solution proposée . . . . .	6
1.3 Étude théorique . . . . .	6
1.3.1 Généralités sur les séries temporelles . . . . .	6
1.3.2 Méthodes de prédiction . . . . .	6
1.4 Méthodologie de développement . . . . .	14
1.4.1 Présentation de SCRUM . . . . .	14
1.4.2 Organisation . . . . .	16
<b>2 Analyse des besoins</b>	<b>17</b>
2.1 Présentation des acteurs . . . . .	17
2.2 Identification des besoins . . . . .	17
2.2.1 Besoins fonctionnels . . . . .	17
2.2.2 Besoins non fonctionnels . . . . .	18
2.2.3 Besoins du domaine . . . . .	18
2.3 Spécification . . . . .	18
2.3.1 Diagramme des cas d'utilisation . . . . .	19
2.3.2 Diagramme de séquences . . . . .	20
<b>3 Conception</b>	<b>24</b>
3.1 Conception architecturale . . . . .	24
3.1.1 Architecture logique . . . . .	24
3.1.2 Architecture physique . . . . .	26

---

3.1.3	Répartition de l'architecture logique sur l'architecture physique . .	26
3.2	Conception détaillée . . . . .	27
<b>4</b>	<b>Réalisation</b>	<b>32</b>
4.1	Environnement de travail . . . . .	32
4.1.1	Environnement matériel . . . . .	32
4.1.2	Environnement logiciel . . . . .	33
4.2	Explication du travail réalisé . . . . .	34
4.2.1	Recherche d'une dataset . . . . .	34
4.2.2	Description de la dataset . . . . .	34
4.2.3	Application de l'approche statistique : ARIMA . . . . .	35
4.2.4	Application de l'approche machine learning : Decision Tree . . . . .	38
4.2.5	Application de l'approche deep learning : RNN-LSTM . . . . .	39
4.3	Résultats du travail réalisé . . . . .	46
4.3.1	Choix des modèles IA . . . . .	46
4.3.2	Interfaces de l'application web . . . . .	48
	<b>Conclusion et perspectives</b>	<b>52</b>
	<b>Netographie</b>	<b>53</b>

# Table des figures

1.1	AIStockFinder.com . . . . .	5
1.2	WalletInvestor.com . . . . .	5
1.3	Exemple de décomposition d'une série temporelle . . . . .	7
1.4	Exemple de décision tree . . . . .	10
1.5	Support Vector Regressor linéair à une dimension . . . . .	11
1.6	Réseau de neurone récurrent . . . . .	12
1.7	Long Short Term Memory . . . . .	13
1.8	Long Short Term Memory - Équations . . . . .	13
1.9	SCRUM . . . . .	15
2.1	Diagramme de cas d'utilisation . . . . .	19
2.2	Diagramme de séquence du CU "S'inscrire" . . . . .	20
2.3	Diagramme de séquence du CU "S'authentifier" . . . . .	21
2.4	Diagramme de séquence du CU "Visualiser la prédiction du cours boursier d'une entreprise" . . . . .	22
2.5	Diagramme de séquence du CU "Visualiser la prédiction du jour j+1 des cours boursiers de toutes les entreprises" . . . . .	22
3.1	Architecture 2 tiers client serveur . . . . .	26
3.2	Diagramme de déploiement . . . . .	27
3.3	Conception de la page d'accueil . . . . .	28
3.4	Conception de l'interface d'inscription . . . . .	28
3.5	Conception de l'interface d'authentification . . . . .	29
3.6	Conception de l'interface de visualisation des prédictions de toutes les entreprises . . . . .	29
3.7	Conception de l'interface de visualisation des prédictions pour une entreprise . . . . .	30
3.8	Diagramme d'activité du système . . . . .	31
4.1	DataSet Yahoo Finance . . . . .	34
4.2	ACF et PACF - IBM . . . . .	36
4.3	ACF et PACF - HPQ . . . . .	37
4.4	ACF et PACF - HPQ . . . . .	38

---

4.5	Évaluation des modèles de Decision Tree . . . . .	39
4.6	Architecture d'un modèle RNN-LSTM . . . . .	40
4.7	Trace d'exécution d'un modèle RNN-LSTM . . . . .	41
4.8	Illustration de l'apprentissage du modèle . . . . .	41
4.9	Évaluation des modèles RNN-LSTM univariés . . . . .	42
4.10	Test de corrélation de features - APPLE . . . . .	44
4.11	Architecture d'un modèle RNN-LSTM . . . . .	45
4.12	Évaluation des modèles RNN-LSTM multivariés . . . . .	46
4.13	Résumé des évaluations des modèles RNN-LSTM . . . . .	47
4.14	Page d'accueil . . . . .	49
4.15	La page d'inscription . . . . .	50
4.16	Tableau des entreprises . . . . .	51



# Liste des sigles et acronymes

<b>IT</b>	<i>Technologies de l'information</i>
<b>ML</b>	<i>Machine learning</i>
<b>DL</b>	<i>Deep learning</i>
<b>RNN</b>	<i>Réseau de neurones récurrent</i>
<b>LSTM</b>	<i>Long short term memory</i>
<b>API</b>	<i>Application Programming Interface</i>
<b>REST</b>	<i>Representational State Transfer</i>

# Introduction générale

La bourse est un marché financier dans lequel sont négociés les actions et les obligations des entreprises cotées en bourse. Ces entreprises sont généralement de grandes firmes qui ont un réel pouvoir sur l'économie nationale, voire internationale. Avoir un aperçu sur le futur des marchés et capturer la tendance de la bourse est donc intéressant, voire fondamental, pour certains corps de métier comme les traders qui sont amenés à prendre des décisions de façon rapide et efficace.

L'analyse des activités boursières et plus particulièrement la prédiction des cours boursiers est un problème qui a suscité l'intérêt de la communauté scientifique, auquel se sont intéressés aussi bien la statistique, l'économétrie que la science des données.

La prédiction des cours boursier est un problème qui repose sur des séries chronologiques puisque les valeurs passées du cours boursiers sont des valeurs numériques qui représentent l'évolution du prix de l'action au cours du temps. Mais ce problème implique plusieurs autres facteurs tel que le comportement rationnel ou irrationnel des agents économiques ou encore les évènements imprévus.

Tous ces aspects combinés rendent les cours boursiers difficile à prévoir de manière précise.

Dans le cadre de ce projet de conception et de développement, nous nous intéressons à cette problématique. Nous nous proposons de concevoir et de développer un modèle d'intelligence artificielle qui prédit la valeur des cours boursier d'entreprises IT et ce, en nous basant sur l'analyse des séries temporelles.

Le présent rapport comprend cinq chapitres, rend compte des diverses étapes que nous avons suivies pour la réalisation de ce projet.

Le premier chapitre « Etat de l'art » introduit le cadre général du projet, décrit l'étude préliminaire réalisée et présente la méthodologie de développement choisie pour développer ce projet.

Le deuxième chapitre « Analyse et spécification des besoins » présente les acteurs, les besoins fonctionnels et non fonctionnels extraits du cahier de charge ainsi que leurs modélisation.

Le troisième chapitre « Conception » est consacré à description de la conception architecturale et de la conception détaillée du projet.

Le dernier chapitre « Réalisation » synthétise le travail réalisé pour aboutir à une solution implémentée et présente des résultats obtenus.

Une conclusion générale clôture ce rapport en récapitulant le travail réalisé et en présentant des perspectives à notre projet.

# Chapitre 1

## Etat de l'art

### Introduction

L'état de l'art est une étape préliminaire à la réalisation de tout projet. Cette étape est essentielle parce qu'elle nous permet de mieux comprendre et mieux cerner notre projet. Ce premier chapitre est donc consacré à la présentation générale et à l'état de l'art du projet.

Il se compose de trois parties. La première partie présente le projet, la deuxième partie décrit l'état de l'art du projet et enfin la troisième partie explique le choix de la méthodologie de développement adoptée.

### 1.1 Définitions générales pour le projet

Dans cette partie, nous abordons quelques définitions qui relèvent du domaine de la finance mais qui sont essentielles à la compréhension de ce projet.

-*La Bourse* est le lieu où les valeurs mobilières (actions et obligations) sont vendues et achetées par des investisseurs. [1]

- *Une entreprise cotée en bourse* est une entreprise dont la totalité ou une partie de son capital est placé sur le marché boursier. Être coté en bourse permet à une société d'accéder au marché de capitaux pour lever des fonds. [2]

-*Une action boursière* est un titre représentant une participation au capital d'une société et donnant droit à divers privilèges. [3]

-*La volatilité d'une action boursière* définit la mesure du risque selon les variations du cours de l'action. [3]

-*Un indice boursier* est un groupe d'actions utilisé pour évaluer un secteur, une Bourse ou une économie. En général, un indice boursier est constitué des actions les plus performantes d'une bourse donnée. [4]

## 1.2 Présentation générale du projet

Dans cette partie, nous abordons le cadre du projet ainsi que le travail demandé.

### 1.2.1 Cadre du projet

L'élaboration de ce projet s'inscrit dans le cadre du module de projet de conception et de développement (PCD) au sein de l'École Nationale des Sciences de l'Informatique (ENSI).

### 1.2.2 Problématique

La bourse est un marché financier complexe et il n'est pas facile d'appréhender son activité et de prévoir son évolution, même parfois pour des traders expérimentés.

Prévoir correctement l'activité boursière consomme beaucoup de temps et des efforts réguliers et qui ne donnent pas toujours des résultats proche de la réalité.

En outre, les traders sont soumis à une pression psychologique forte à cause des montants qu'ils manipulent et des risques liés à la volatilité des marchés financiers.

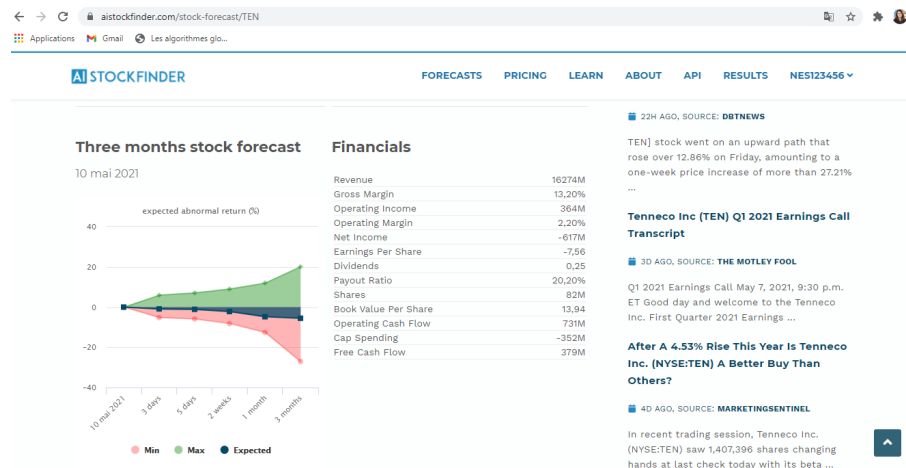
Les bourses américaines sont parmi les plus actives du monde et donc les plus difficiles à prévoir pour les traders. La bourse NASDAQ (National Association of Securities Dealers Automated Quotation system), l'une des deux plus importantes bourses en Amérique avec Wall Street, est fortement influencée par les entreprises IT, la preuve en est que l'indice boursier NASDAQ-100 est principalement composé d'actions du secteur IT à hauteur de 47,25%.

### 1.2.3 Solutions existantes

Dans cette section, nous étudions quelques exemples de sites web spécialisé dans la prédiction des cours boursiers.

- **AIStockFinder.com**

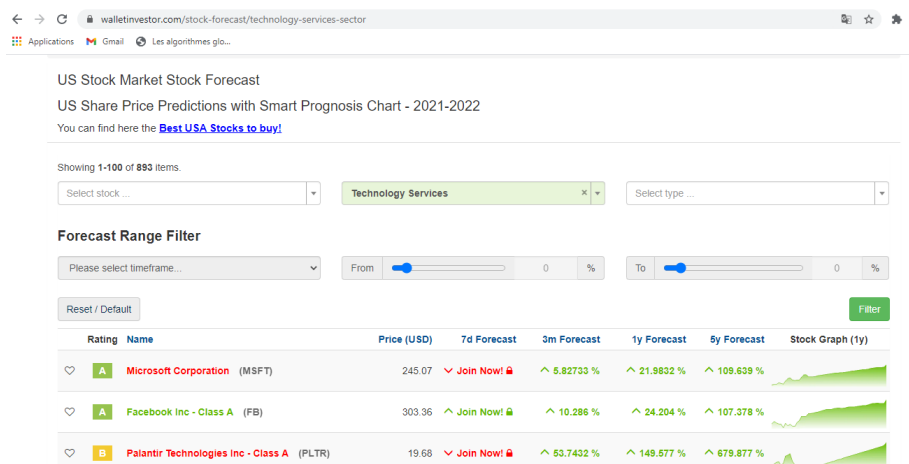
AIStockFinder propose des prédictions des cours boursiers américain sur le long et le court. Ce site propose une prédiction des valeurs au jour  $j+3$ ,  $j+5$ ,  $j+15$ ,  $j+1$ mois et  $j+3$ mois comme le montre la figure ci-après.

FIGURE 1.1 – AIStockFinder.com  
[5]

## ● WalletInvestor.com

WalletInvestor offre des prévisions dans de multiples catégories dont :

- Prévisions de crypto-monnaie
- Prévisions boursières américaines
- Prévisions Forex (EUR à USD, USD à JPY ...)

FIGURE 1.2 – WalletInvestor.com  
[6]

Dans la section Prévisions boursières américaines (celle qui nous intéresse), le menu fourni offre (en pourcentage du prix actuel) :

- 7d Forecast : le prix prévu de l'action dans 7 jours

- 3m Forecast : le prix prévu du stock dans 3 mois
- 6m Forecast : le prix prévu de l'action dans 6 mois
- Prévision 1 an : un prix prévu de l'action dans un an
- Prévisions sur 5 ans : un prix prévu de l' action dans 5 ans [5]

### **1.2.4 Critique de l'existant**

Les deux solutions précédemment étudiées proposent des prédictions à la fois sur le court et sur le long terme, mais n'offrent pas de prédiction jour par jour la valeur des cours boursiers. Or les traders interviennent dans les marchés financiers quotidiennement, et donc une prédiction au jour le jour est plus à même de les aider dans leurs tâches.

### **1.2.5 Solution proposée**

Nous nous proposons, dans ce projet, de réaliser un modèle d'intelligence artificielle qui permet de prédire les cours boursiers de différentes entreprises technologiques cotées dans la bourse américaine NASDAQ en se basant sur l'analyse de leurs cours passés. Le résultat des prédictions obtenues sera visualisé dans une interface web.

## **1.3 Étude théorique**

Cette section aborde la base théorique pour ce projet. Nous détaillons, dans ce qui suit, les diverses techniques et les méthodes applicables à notre problématique.

### **1.3.1 Généralités sur les séries temporelles**

« Une série temporelle, ou série chronologique, est une suite finie de valeurs numériques représentant l'évolution d'une quantité spécifique au cours du temps. » [7]

Une série temporelle univariée a une seule variable qui change au cours du temps alors qu'une série temporelle multivariée en a plusieurs.

L'analyse des séries temporelles a deux objectifs :

- Modéliser pour expliquer un phénomène observé
- prédire la série dans le futur [8]

### **1.3.2 Méthodes de prédiction**

#### **● Approche statistique**

Dans l'approche statistique, il est primordial de comprendre et de bien utiliser le processus statistique pour arriver à prédire de nouvelles valeurs d'une série temporelle.

Afin de mieux comprendre les données d'une série temporelle, une décomposition de cette dernière est souvent réalisée. La série chronologique est décomposée en :

- Une *tendance*  $T(t)$  représente l'évolution de la série temporelle à long terme.
- Une *saisonnalité*  $S(t)$  représente un comportement périodique qui se répète avec une fréquence constante.
- Un *résidu*  $r(t)$  est un bruit statistique qui évolue de façon irrégulière et inexpliquée.

Il existe deux méthodes de décomposition d'une série temporelle :

La méthode additive  $Y(t) = T(t) + S(t) + r(t)$

La méthode multiplicative  $Y(t) = T(t) * S(t) * r(t)$

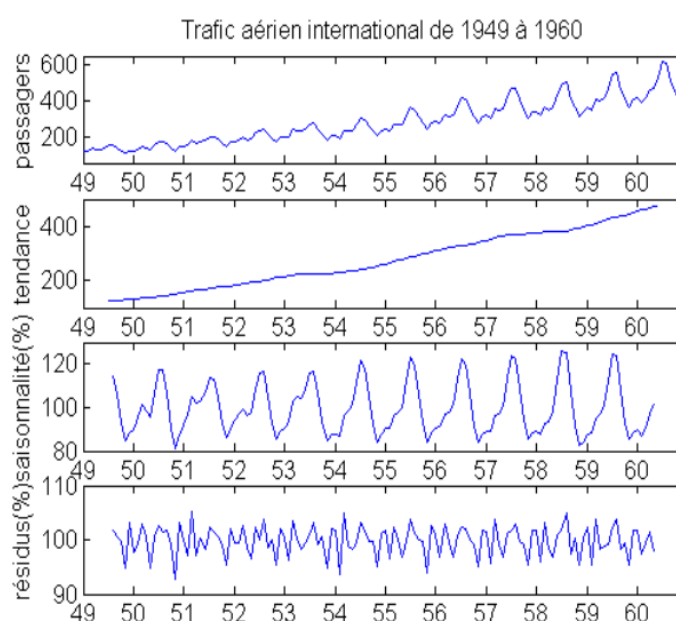


FIGURE 1.3 – Exemple de décomposition d'une série temporelle [9]

Dans un premier temps, la tendance est extraite en appliquant un filtre de convolution aux données originales. Par la suite, la saisonnalité est extraite en calculant la moyenne de cette série dénuée de tendance pour chaque période. Enfin, le résidu est obtenu par extraction du signal d'origine de la tendance et la saisonnalité.

Un autre concept fondamental des séries temporelle est la stationnarité. En effet, par exemple, le problème de la régression fallacieuse stipule d'une régression linéaire sur une série temporelle non stationnaire donne des résultats erronés et trop optimistes.

Un processus est dit stationnaire si sa structure sous-jacente reste la même.

Il existe 2 types de séries temporelles stationnaires :



*Stationnarité au sens fort* : Un processus  $X$  est strictement stationnaire si pour tout  $k > 0$ , toute suite  $t_1, t_2, \dots, t_k$  et tout  $h \in Z$ , les lois de  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  et de  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$  sont identiques. [10]

*Stationnarité au sens faible* : Un processus  $(x_t, t \in Z)$  est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si :

$$-\forall t \in Z, E(x_t^2) < \infty$$

$$-\forall t \in Z, E(x_t) = m, \text{ indépendant de } t$$

$$-\forall (t, h) \in Z^2, \text{cov}(x_t, x_{t+h}) = E[(x_t - m)(x_{t+h} - m)] = \gamma(h), \text{ indépendant de } t \text{ [10]}$$

Il existe divers tests statistiques permettant de déterminer la nature des séries temporelles, nous prenons le Test de Dickey Fuller Augmenté (ADF).

$H_0$  : la série est non stationnaire et  $H_1$  : la série est stationnaire

Règle de décision :

-Si ADF calculé  $<$  ADF théorique alors l'hypothèse  $H_1$  est vérifiée et la variable est stationnaire.

-Si ADF calculé  $>$  ADF théorique alors l'hypothèse  $H_0$  est vérifiée et la variable est non stationnaire.

Dans la réalité, la majorité des séries chronologiques ne sont pas stationnaires, mais il existe une méthode pour les stationnariser.

Il existe divers techniques statistiques qui modélisent et aident dans la prédiction des séries temporelles. Dans ce qui suit, nous présentons quelques modèles des plus connus.

#### \* Moyenne mobile MA

Le modèle de moyenne mobile MA d'ordre  $q$   $MA(q)$  utilise les erreurs de prévision passées dans un modèle de type régression.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

où  $\varepsilon_t$  est le bruit blanc. [11]

Afin de déterminer l'ordre de MA, il faut recourir au test d'auto-corrélation ACF (Auto-Correlation Function). L'autocorrélation représente la relation des valeurs d'une série avec ses propres valeurs décalées  $p$  fois en arrière.

#### \* Autorégression AR

Dans un modèle d'autorégression (AR), la variable d'intérêt est prédite avec une combinaison linéaire des valeurs passées de la variable.

Le terme autorégression indique qu'il s'agit d'une régression de la variable par rapport à

elle-même. Un AR d'ordre  $p$   $AR(p)$  s'écrit comme suit :

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

où  $\varepsilon_t$  est un bruit blanc (suite de variables aléatoires réelles identiquement distribuée).[14]  
AP est semblable à une régression multiple mais avec des valeurs décalées de  $y_t$  comme valeurs de prédiction.[12]

Afin de déterminer l'ordre de AR, il faut recourir au test d'auto-corrélation PACF (Partial Auto-Correlation Function). Cette fonction mesure l'auto-corrélation entre une série et ses valeurs décalées  $q$  fois en arrière et en prenant en considération les valeurs des intervalles intermédiaires.

#### \* ARIMA

Le modèle ARIMA est la combinaison des deux modèles MA et AR présentés précédemment avec ajout de l'intégration.

L'intégration consiste à faire la différence entre les valeurs d'une série temporelle et de ses valeurs décalées en arrière, ce qui permet la rendre stationnaire si elle ne l'est pas.

ARIMA(p,d,q) où  $p$  et  $q$  sont respectivement les paramètres des modèles MA et AR et  $d$  est le nombre d'intégration que la série devra subir.

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

### ● Approche Machine Learning

« Nous pouvons définir le Machine Learning (ML) comme étant une technologie d'intelligence artificielle permettant aux machines d'apprendre sans avoir été préalablement programmées spécifiquement à cet effet. »[13]

Les modèles ML sont de deux types : l'apprentissage supervisé et l'apprentissage non supervisé.

L'apprentissage supervisé : consiste en des variables d'entrée ( $X$ ) et une variable de sortie ( $y$ ). L'algorithme permet d'apprendre la fonction de mapping de l'entrée à la sortie. Le but étant de prédire pour de nouvelles données en entrée ( $X$ ) les valeurs des variables en sortie ( $y$ ).

Les problèmes d'apprentissage supervisé peuvent être divisés en deux catégories : des problèmes de régression et des problèmes de classification.

- Classification : Un problème de classification est un problème où les variables de sortie représentent des catégories. variable qualitative
- Régression : un problème de régression est un problème où la/les variable(s) de sortie est/sont des valeurs réelles. variable quantitative

L'apprentissage non supervisé : consiste à ne disposer que de données d'entrée ( $X$ ) et pas de variables de sortie ( $Y$ ) correspondantes.

L'objectif de l'apprentissage non supervisé est de modéliser la structure ou la distribution sous-jacente dans les données.

En ce qui concerne notre projet, la prévision des cours boursiers peut être considérée comme un problème d'apprentissage supervisé de régression. Nous nous baserons sur les anciennes valeurs en termes de temps comme variables d'entrée et soit la valeur suivante en temps (prédiction One-step) soit un ensemble de valeurs suivantes en temps (prédiction Multi-step) comme variable(s) de sortie.

Il existe divers algorithmes de machine learning adaptés à notre problème, nous présentons dans ce qui suit deux algorithmes parmi les plus utilisés dans la prédiction des séries temporelles, le Decision tree Regressor et le SVR.

#### \* Decision tree Regressor

L'arbre de décision est un algorithme machine learning de classification qui a été adapté aux problèmes de prédiction.

Le principe consiste à construire une structure arborescente en divisant l'ensemble des données en entrée en des ensembles de taille de plus en plus petites.

La figure suivante illustre un exemple d'application de cet algorithme. Le noeud racine

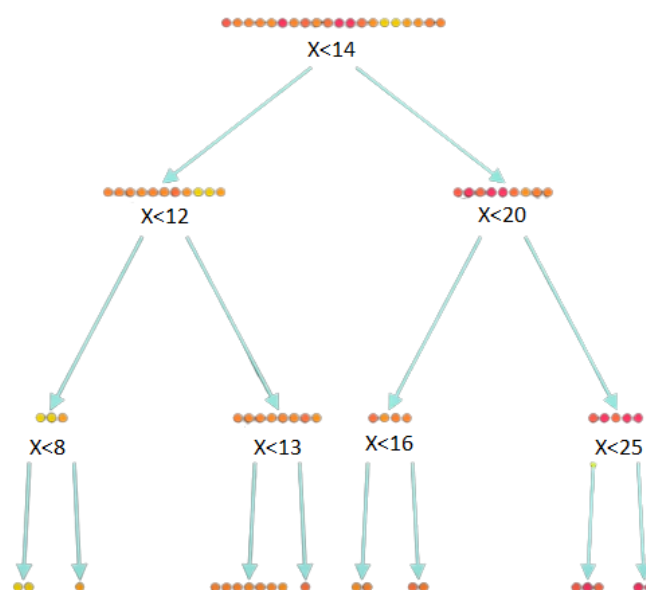


FIGURE 1.4 – Exemple de décision tree  
[14]

contient toutes les observations.

L'arbre est construit par séparation des noeuds. Un noeud sera divisé si son impureté (l'écart type entre les données dans le noeud) est au-dessus d'un seuil donné, sinon c'est une feuille.

Pour mesurer la qualité d'une séparation d'un noeud, il existe divers critères parmi lesquels nous pouvons citer la réduction de la variance entre les valeurs des noeud fils.

$$variance_{reduction} = var(parentnode) - \sum_{i=1}^n w_i * var(childnode_i)$$

Pour ce modèle, la valeur prédite  $Y$  est la moyenne des données de la feuille atteinte après parcours de l'arbre selon une valeur  $X$  en entrée.

#### \* Support Vector Regressor

Le Support Vector Machine est un algorithme machine learning qui résout des problèmes classification binaire. Le SVM ou « séparateurs à vaste marge » sépare les données en classes à l'aide d'une frontière de telle façon que la distance entre les différents groupes de données et la frontière qui les sépare soit maximale. Cette distance est appelée « marge » et les « vecteurs de support » sont les frontières de la marge.[15]

Le SVM a été généralisé pour s'adapter aux problèmes de régression. Cette généralisation s'est faite grâce à l'introduction de la région  $\epsilon$ -deviation autour de l'hyperplan, appelée  $\epsilon$ -tube. Le but du SVR devient la recherche du tube optimal qui est le tube le plus plat possible contenant la majorité des données d'apprentissage.

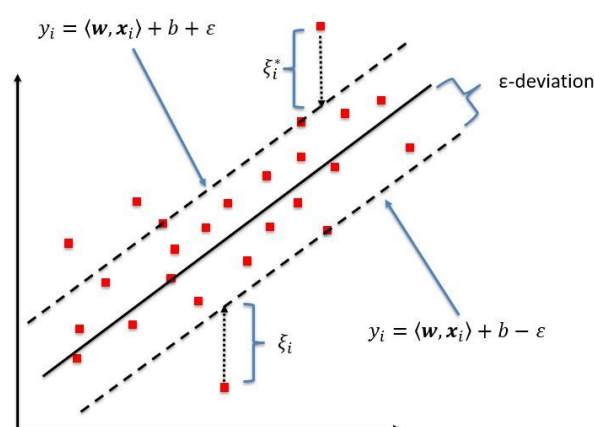


FIGURE 1.5 – Support Vector Regressor linéaire à une dimension  
[15]

Une seconde notion sur laquelle se base les SVM-SVR est le Kernel (noyau). Le noyau est une fonctions mathématiques qui prend des données en entrée et essaye de trouver le  $\epsilon$ -tube optimal. Il existe divers types de noyaux : linéaire, non linéaire, polynomiale, fonction de base radiale (RBF) et sigmoïde.

En ce qui concerne la prédiction, l'hyperplan (la ligne de séparation entre les classes de données dans un SVR unidimensionnel linéaire comme le montre la figure ci-dessus) sera utilisé pour prédire la valeur cible.

### • Approche Deep Learning

Le Deep Learning est un sous-domaine du ML inspiré par les neurones du cerveau humain. Cette approche tente de dégager les caractéristiques principales des données en entrée pour pouvoir prédire un résultat en sortie.

Le deep learning se base sur un réseau de neurones artificiels organisé en couches de neurones liées entre elles.

Les données sont introduites au réseau de neurones à travers la première couche “input layer” et fournit son résultat sur la dernière couche “output layer”. Ces deux couches sont reliées à des couches intermédiaires “hidden layers” par des connexions pondérées comme le montre la figure ci-dessus.

Il existe différents types de réseaux de neurones mais pour notre projet, nous nous intéressons uniquement aux réseaux de neurones récurrents (RNN) car cette catégorie de réseaux de neurones est dédiée au traitement des séquences.

#### \* Réseau de neurone récurrent (RNN)

Un RNN est un modèle de la famille du Deep Learning adapté aux prédictions de données séquentielles grâce à une architecture basée sur la mémoire. C’est un réseau artificiel simple où les informations persistent grâce à une connexion de retour d’information, il devient ainsi possible de relier les informations passée à la tâche courante.

La figure suivante illustre un RNN associée au temps  $t$  où les entrées sont  $x(t)$  et sa propre sortie du temps précédent  $y(t-1)$ .

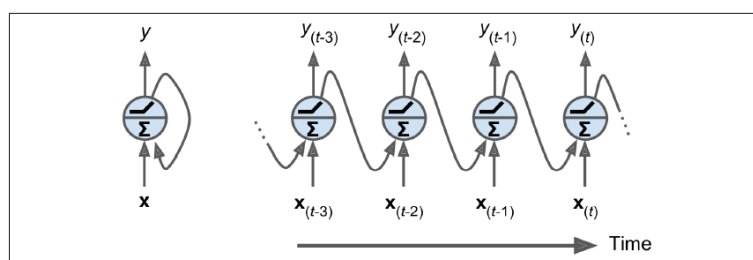


FIGURE 1.6 – Réseau de neurone récurrent  
[16]

#### \* Long Short Term Memory (LSTM)

Le principal inconvénient des RNN est la difficulté à apprendre de longues séquences de données. En effet, avec une architecture RNN profonde (avec un nombre important de couches), le gradient devient de plus en plus petit, les mises à jour des paramètres deviennent insignifiantes et donc aucun véritable apprentissage n’est effectué.

Le LSTM corrige ce problème en adoptant des mécanismes internes appelés gates (portes) qui régulent le flux d'information.

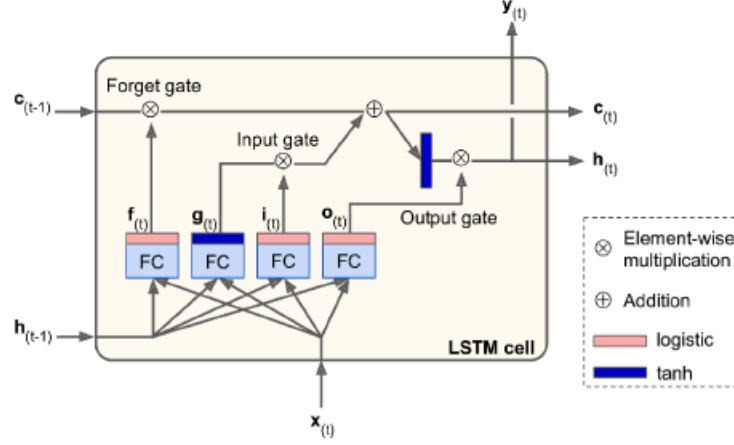


FIGURE 1.7 – Long Short Term Memory  
[16]

Le LSTM possède 3 types de gates :

- *La forget gate* (contrôlée par  $f(t)$ ) contrôle quelles parties de l'état à long terme doit être effacé.
- *L'input gate* (contrôlée par  $i(t)$ ) contrôle quelles parties de  $g(t)$  doivent être ajoutées à l'état à long terme.
- *L'output gate* (contrôlée par  $o(t)$ ) contrôle quelles parties du long terme l'état doit être lu et sorti à ce pas de temps, à la fois vers  $h(t)$  et vers  $y(t)$ .

$$\begin{aligned}
 i_{(t)} &= \sigma(W_{xi}^T x_{(t)} + W_{hi}^T h_{(t-1)} + b_i) \\
 f_{(t)} &= \sigma(W_{xf}^T x_{(t)} + W_{hf}^T h_{(t-1)} + b_f) \\
 o_{(t)} &= \sigma(W_{xo}^T x_{(t)} + W_{ho}^T h_{(t-1)} + b_o) \\
 g_{(t)} &= \tanh(W_{xg}^T x_{(t)} + W_{hg}^T h_{(t-1)} + b_g) \\
 c_{(t)} &= f_{(t)} \otimes c_{(t-1)} + i_{(t)} \otimes g_{(t)} \\
 y_{(t)} &= h_{(t)} = o_{(t)} \otimes \tanh(c_{(t)})
 \end{aligned}$$

FIGURE 1.8 – Long Short Term Memory - Équations  
[16]

## Évaluation des modèles de prédiction

L'Évaluation est une étape cruciale de la construction d'un modèle de régression qui permet de décider un modèle peut être déployé ou non.

Nous présentons dans ce qui suit quelques métriques de régression :

- MAE (Mean absolute error) : représente la moyenne des valeur absolues des différences entre les observations et les prédictions d'un modèle

$$mae = \left(\frac{1}{n}\right) \sum_{i=1}^n |y_i - x_i|$$

- MSE (mean squared error) : représente la moyenne des carrés des différences entre les observations et les prédictions d'un modèle

$$mse = \left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - x_i)^2$$

- RMSE (root mean squared error) : est la racine carrée de MSE

$$rmse = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - x_i)^2}$$

## 1.4 Méthodologie de développement

Une méthodologie de développement est un ensemble de principes et de pratiques cohérents qui permettent d'abstraire et d'organiser les étapes du développement d'un logiciel. Le choix d'une méthodologie adéquate aux caractéristiques d'un projet permet une performance optimale de l'équipe et une bonne qualité du produit obtenu.

Pour le développement de notre projet, nous nous intéressons aux méthodologies agiles qui sont basées sur les principes du manifest agile et qui sont itératives et adaptatives c'est-à-dire que la réalisation du projet se fait sur plusieurs itérations tout en s'adaptant aux changements. Dans cette famille des méthodologies de développement agiles, nous nous intéressons plus précisément au framework SCRUM que nous présentons ci-après.

### 1.4.1 Présentation de SCRUM

Scrum est un framework simple à comprendre, qui vise une collaboration efficace en équipe sur des projets complexes et qui aide les personnes, les équipes et les organisations à générer de la valeur grâce à des solutions adaptatives.

C'est un aussi processus empirique qui encourage les équipes à apprendre par l'expérience, à

s'auto-organiser, et à réfléchir à leurs victoires aussi bien qu'à leurs défaites afin de s'améliorer en continu.[17]

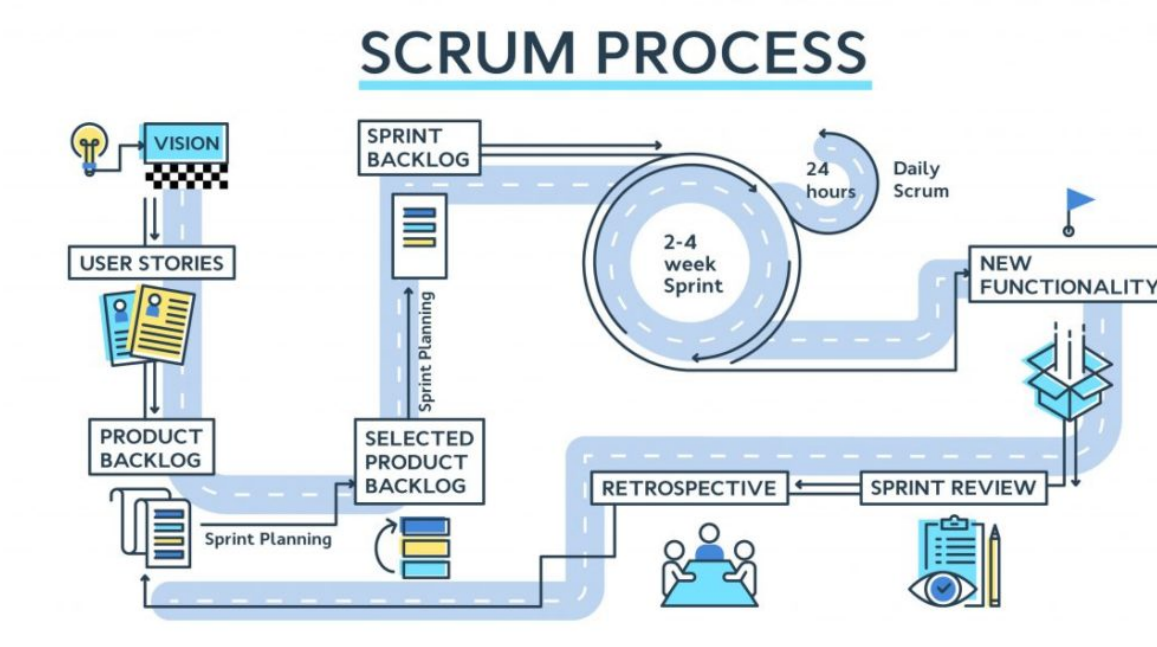


FIGURE 1.9 – SCRUM  
[18]

Au début du projet, le client, le Product Owner, fournit une description des fonctionnalités souhaitées appelées « Users Stories » et l'ensemble de ces User Stories sera regroupé hiérarchiquement dans le « Product Backlog » qui devient un référentiel pour le projet. Le Product Backlog évolue tout au long du projet pour s'adapter aux changements des besoins du client.

Le projet est organisé en un ensemble d'itérations ou « Sprints ». Chaque Sprint dure entre deux et six semaines. Un « Sprint Planning » avec toute l'équipe scrum (Product Owner + Scrum Master, l'animateur de l'équipe scrum + équipe de développement) précédant le début de chaque sprint sera consacrée à la formation du sprint backlog en sélectionnant les Users Stories qui seront réalisées durant le sprint.

Durant d'un sprint, un « Daily Scrum » a lieu quotidiennement avec l'équipe de développement. Cette réunion permet à l'équipe de se synchroniser et de faire le point sur l'avancement du sprint. Chaque membre de l'équipe raconte ce qu'il a fait, les difficultés rencontrées et ce qu'il va faire, il peut demander de l'aide ou bien aider les autres membres du groupe.

A la fin du sprint, un « Sprint Review » est organisé avec toute l'équipe scrum lors de laquelle il y a inspection de l'incrément. Un « Sprint Review » suivra pour relever ce qui



s'est bien déroulé, ce qui s'est moins bien déroulé et proposer des axes d'améliorations.

Tout au long des différentes itérations, l'équipe scrum se perfectionne, s'adapte et améliore sa productivité en continue.

A la fin du dernier sprint, le produit fini est remis au client.

Notre choix a été motivé par les caractéristiques de Scrum. En effet, cette méthodologie est simple à comprendre, légère en documentation et permet un développement itératif et incrémental ce qui permet de se concentrer sur une seule tâche à la fois.

### **1.4.2 Organisation**

Les acteurs SCRUM de notre projet sont :

- Product Owner et Scrum Master : Dr. Rym BESROUR
- Équipe de développement : Mme. Nesrine JOUINI, M. Wael KHEMAKHEM, M. Ihebbeddine MARNAOUI

Nous avons divisé notre projet en 6 sprints :

- Sprint 1 : Recherche et analyse de la dataset à utiliser
- Sprint 2 : Implémentation de l'algorithme statistique
- Sprint 3 : Implémentation des algorithmes machine learning
- Sprint 4 : Implémentation des l'algorithmes deep learning
- Sprint 5 : L'implémentation de la partie Front-end et Back-end et de la liaison entre eux

## **Conclusion**

Dans ce premier chapitre, nous avons commencé par présenter le projet d'une manière générale, ensuite nous avons établi l'état de l'art avec l'étude de l'existant et l'étude de la théorie derrière la prédiction des séries temporelles et enfin nous avons choisi une méthodologie de développement pour la réalisation de ce travail.

Désormais, nous sommes à même d'aborder le développement de notre projet.

# Chapitre 2

## Analyse des besoins

### Introduction

L'analyse des besoins est une étape primordiale dans la réalisation de tout projet parce qu'elle nous aide à comprendre les besoins du client et à définir les fonctionnalités que le système doit remplir. Dans ce deuxième chapitre, nous analysons, dans un premier temps, le cahier de charge pour en extraire les acteurs, les besoins fonctionnels et les besoins non fonctionnels du domaine. Nous réalisons la modélisation des besoins identifiés avec le langage de modélisation adopté.

### 2.1 Présentation des acteurs

Un acteur est une entité humaine, logicielle ou matérielle autonome externe au système qui interagit directement avec lui.[19]

Pour notre application, nous identifions un seul acteur, le trader qui est une personne dont le travail est d'acheter des actions boursières et de les revendre à un prix plus élevé en vue de réaliser un profit.

### 2.2 Identification des besoins

Cette section consiste à identifier les besoins fonctionnels, les besoins non fonctionnels ainsi que les besoins du domaine de notre système.

#### 2.2.1 Besoins fonctionnels

Un besoin fonctionnel exprime une fonctionnalité que le système doit fournir en réponse à un besoin d'un ou de plusieurs utilisateurs[19]

En ce qui concerne notre système, nous identifions les besoins fonctionnels suivants :

- S'inscrire
- S'authentifier
- Visualiser la prédiction à court terme ( $j+1 \rightarrow j+7$ ) du cours boursier d'une entreprise dans une courbe (à court terme).
- Visualiser la prédiction du cours boursier d'une entreprise à long terme.
- Visualiser la prédiction du jour  $j+1$  dans un tableau de toutes les entreprises prises en compte par notre système.

### 2.2.2 Besoins non fonctionnels

Un besoin non fonctionnel représente une contrainte qui pèse sur le système et qui doit être mesurable.[19]

En ce qui concerne notre système, nous identifions les besoins non fonctionnels suivants :

- L'application devra supporter les navigateurs majeurs (Chrome, Firefox, Opera, Edge, Safari).
- Ergonomie : l'application devra fournir une interface agréable à utiliser et ne contenant pas plus de trois couleurs.
- Utilisabilité : l'utilisateur n'aura besoin d'aucun pré-requis pour utiliser l'application.
- Disponibilité : l'application devra être disponible 24h/24 et 7j/7, sauf en période de maintenance.
- Performance : le temps de chargement de la page web ne devra pas dépasser 15 secondes.

### 2.2.3 Besoins du domaine

Un besoin du domaine représente une caractéristique que doit remplir le système et qui dérive directement du domaine d'application[19], la finance dans notre cas.

En ce qui concerne notre système, nous identifions les besoins du domaine suivants :

- Le système devra afficher des courbes conformes aux normes des graphiques boursiers.

## 2.3 Spécification

La spécification décrit la structure, les fonctions ainsi que les changements d'états du système de façon précise en se basant sur le langage de modélisation UML.

### 2.3.1 Diagramme des cas d'utilisation

Le diagramme des cas d'utilisation regroupe l'ensemble des services visibles depuis l'extérieur du système et qui viennent en réponse à une stimulation de l'utilisateur. La figure présente le diagramme des cas d'utilisation de notre système.

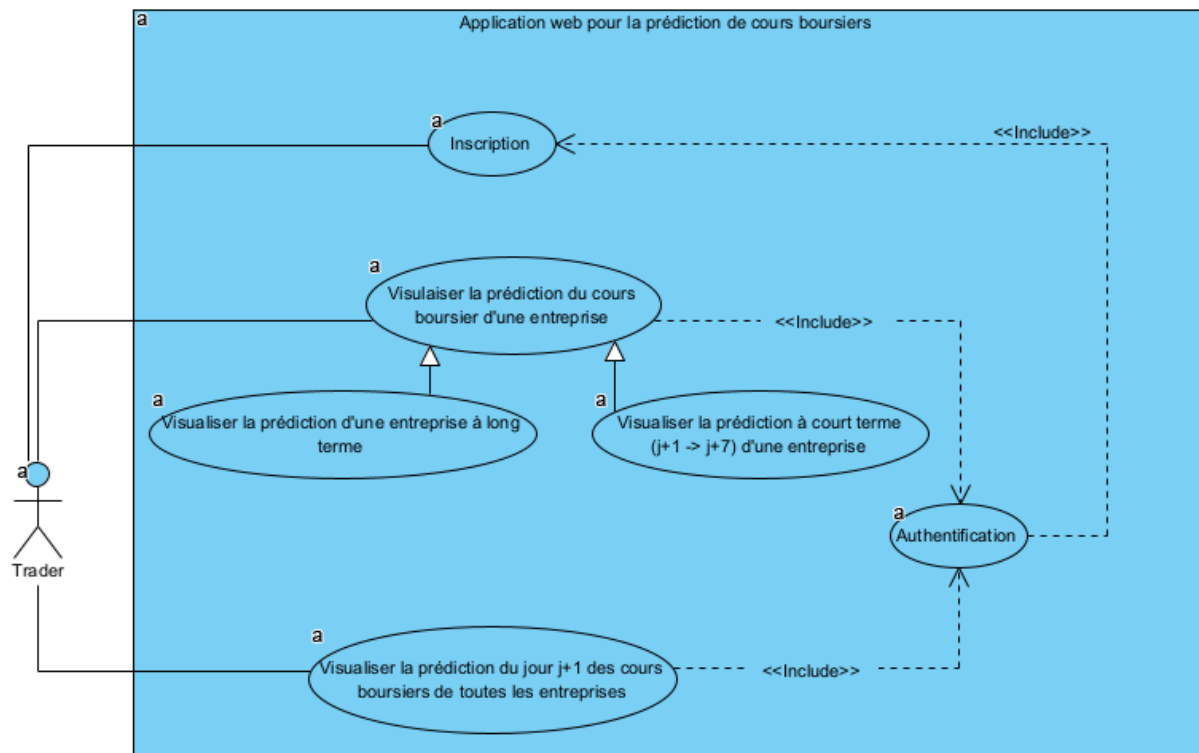


FIGURE 2.1 – Diagramme de cas d'utilisation

### 2.3.2 Diagramme de séquences

Le diagramme de séquences sert à détailler et à expliquer les échanges entre l'utilisateur et le système pour un cas d'utilisation.

- Analyse du cas d'utilisation "S'inscrire"

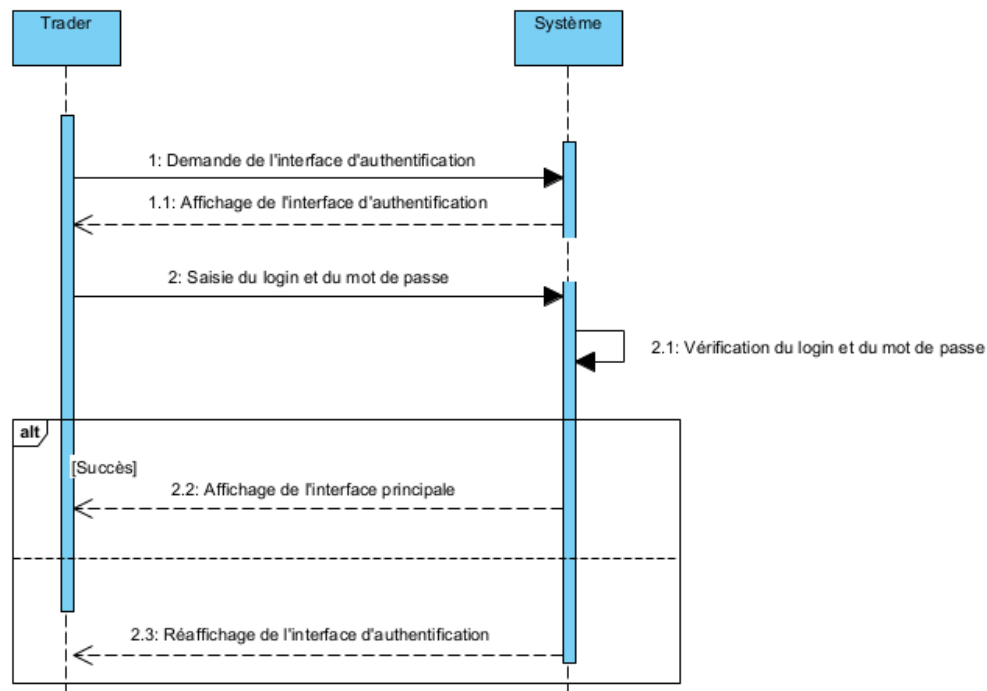


FIGURE 2.2 – Diagramme de séquence du CU "S'inscrire"

- Analyse du cas d'utilisation "S'authentifier"

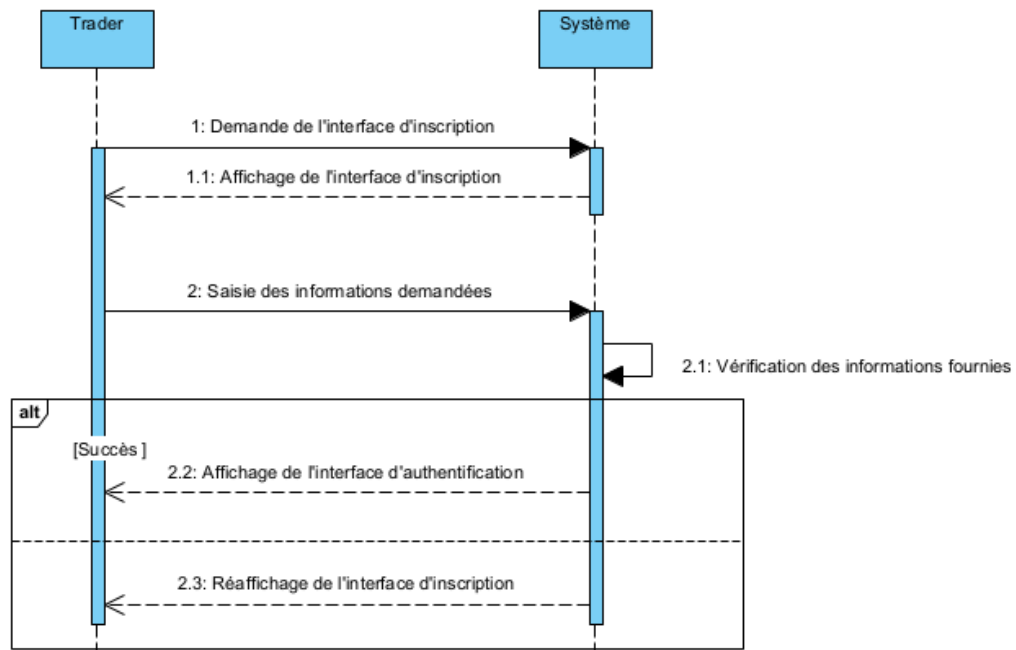


FIGURE 2.3 – Diagramme de séquence du CU "S'authentifier"

- Analyse des cas d'utilisation "Visualiser la prédiction du cours boursier d'une entreprise"

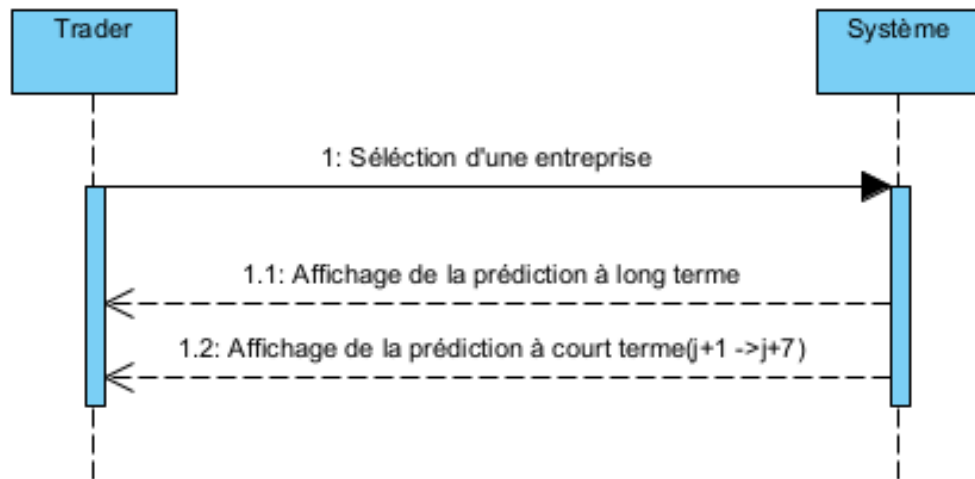


FIGURE 2.4 – Diagramme de séquence du CU "Visualiser la prédiction du cours boursier d'une entreprise"

- Analyse du cas d'utilisation "Visualiser la prédiction du jour j+1 des cours boursiers de toutes les entreprises"

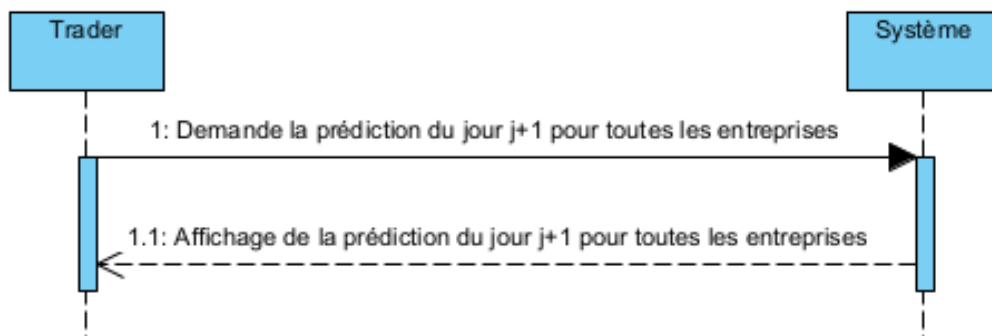


FIGURE 2.5 – Diagramme de séquence du CU "Visualiser la prédiction du jour j+1 des cours boursiers de toutes les entreprises"

## Conclusion

Ce chapitre a été dédié à l'analyse des besoins. Nous avons extrait les besoins fonctionnels et non fonctionnels de notre système et nous les avons modélisé ce qui nous a permis de mieux comprendre notre système.

Maintenant nous sommes en mesure de concevoir notre système.



# Chapitre 3

## Conception

### Introduction

La conception est une étape qui consiste à apporter une solution aux problèmes préalablement identifiés dans la phase d'analyse des besoins.

Dans ce troisième chapitre, nous abordons la phase de conception de notre projet. Nous commençons par présenter l'architecture physique de la solution, puis nous passons à l'architecture logique et enfin nous détaillons ces deux architectures lors de la conception détaillée.

### 3.1 Conception architecturale

La conception architecturale consiste à identifier une organisation générale logique et physique du système.

Dans cette section, nous abordons l'architecture physique et l'architecture logique de la solution que nous proposons.

#### 3.1.1 Architecture logique

L'architecture logique divise le système d'un point de vue logique (c'est-à-dire selon les traitements qu'ils effectuent) en des composants et des relations qui les relient.

- **Choix d'une architecture logique**

Pour réaliser notre système, nous optons pour une architecture orientée service (SOA). Elle consiste à développer des composants logiciels distribués avec une forte cohérence interne et un couplage inter-composants faible, presque inexistant.

Dans notre cas le backend et le frontend seront donc indépendants.

### ● Choix des technologies

#### \* Backend

Pour l'implémentation de cette partie, nous choisissons :

- « Flask » qui est un micro-framework python, open source pour le développement web qui est simple, léger et minimaliste. Bien qu'il a moins de fonctionnalités que d'autres frameworks web, il facilite le développement et garantit une meilleure souplesse.[20]
- « RESTful API » est un style architectural simple à utiliser, qui offre une communication sans état et qui n'impose pas de contraintes côté client. Il utilise le protocole HTTP.

En ce qui concerne notre projet, nous voulons développer un backend simple qui se charge seulement d'exécuter le modèle IA pour prédire les valeurs requises et de les communiquer au front. Ainsi, les caractéristiques de Flask et des restful APIs sont parfaitement adapté à notre besoin.

#### \* Frontend

Pour l'implémentation de cette partie, nous choisissons « React JS » qui est une bibliothèque javascript open source qui facilite la création d'interfaces utilisateur rapide et interactive pour des applications web. La page web est divisée en des composants ayant un état. Elle regénère la page ou une partie de la page HTML à chaque changement d'état. [21]

Ce choix fut motivé par la simplicité d'utilisation de la bibliothèque et sa popularité au près de grandes entreprises internationales (Netflix, Facebook ....).

#### \* Partie IA

Pour la réalisation des modèles de prédiction intelligent, nous optons pour ces bibliothèques python :

- Approche statistique :

*statsmodels* est un module Python qui fournit des outils pour l'exploration et la modélisation de données statistiques ainsi que pour la réalisation de tests statistiques.

- Approche Machine Learning :

*Scikit-learn* est une bibliothèque Python open source pour l'apprentissage automatique. Elle contient divers algorithmes et outils de Machine Learning.

- Deep Learning

*Tensorflow* est une bibliothèque open-source développée par Google. Elle implémente diverses méthodes d'apprentissage profond.

### 3.1.2 Architecture physique

L'architecture physique décrit l'ensemble des composants matériels et leurs relations nécessaires au système.

Pour notre projet, nous choisissons une architecture 2-tiers de type client serveur.

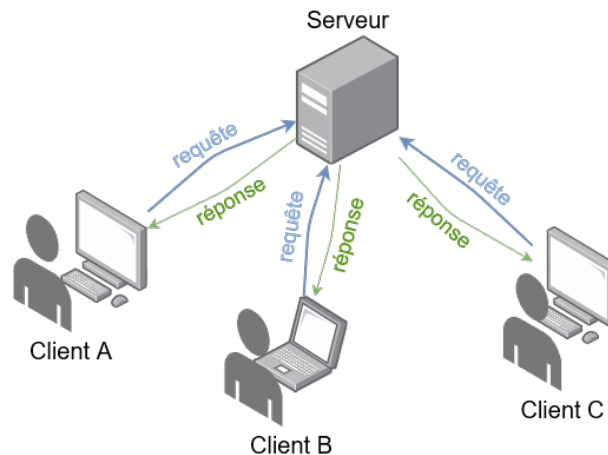


FIGURE 3.1 – Architecture 2 tiers client serveur  
[22]

Cette architecture se compose de deux tiers :

- Le tier client, un client léger, envoie des requêtes au serveur et récupère la réponse.
- Le tier serveur attend les requêtes des clients pour les satisfaire.

Nous optons pour cette architecture car :

- Notre système n'a pas besoin d'une base de données.
- Le traitement étant indépendant des clients, les mises à jour sont faciles à réaliser.

### 3.1.3 Répartition de l'architecture logique sur l'architecture physique

Le diagramme de déploiement est un diagramme structurel qui décrit le déploiement des composants logiciels sur les composants physiques.

La figure suivante illustre le diagramme de déploiement de notre système.

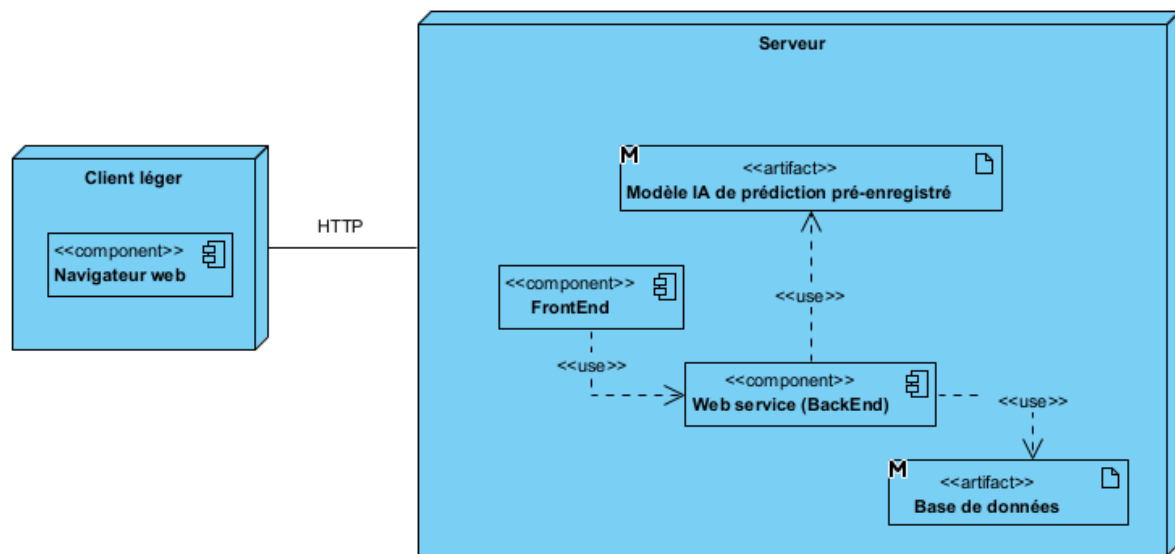


FIGURE 3.2 – Diagramme de déploiement

## 3.2 Conception détaillée

Nous abordons dans cette section la conception détaillée de notre système. Elle permet de détailler la conception architecturale en fournissant tous les détails de l'implémentation.

### • Conception détaillée "frontend"

La création et la conception de site web ou web design est la conception de l'interface web, l'architecture interactionnelle, l'organisation des pages, l'arborescence et la navigation dans un site web. [23]

Dans ce qui suit, nous présentons la conception de nos interfaces web.

La figure 3.4 illustre la conception de la page d'accueil de notre application.



FIGURE 3.3 – Conception de la page d'accueil

La figure 3.5 illustre la conception d'une interface qui offre à l'utilisateur la possibilité d'inscrire.

Cette interface s'affiche suite au clic de l'utilisateur sur le bouton Sign up.

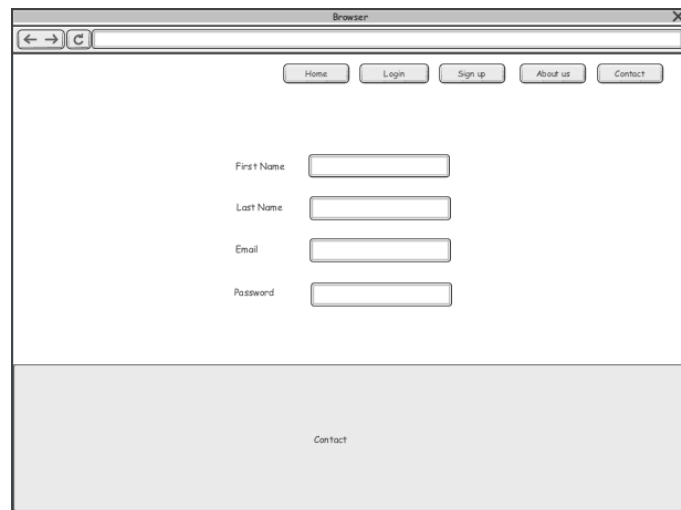
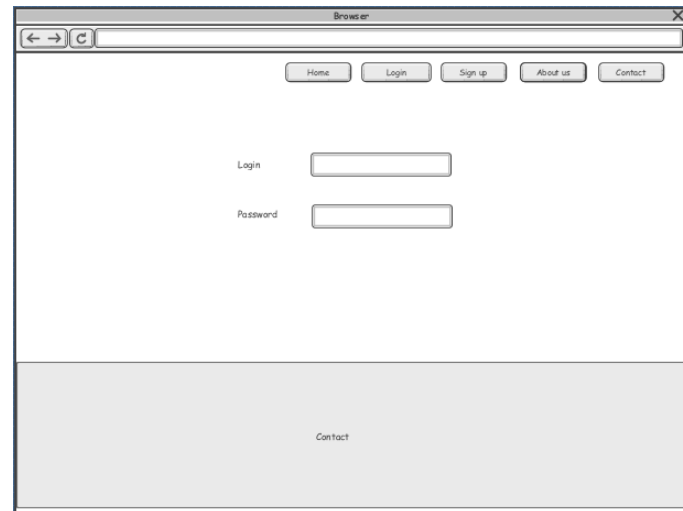


FIGURE 3.4 – Conception de l'interface d'inscription

La figure 3.5 illustre la conception d'une interface qui offre à l'utilisateur la possibilité de s'authentifier.

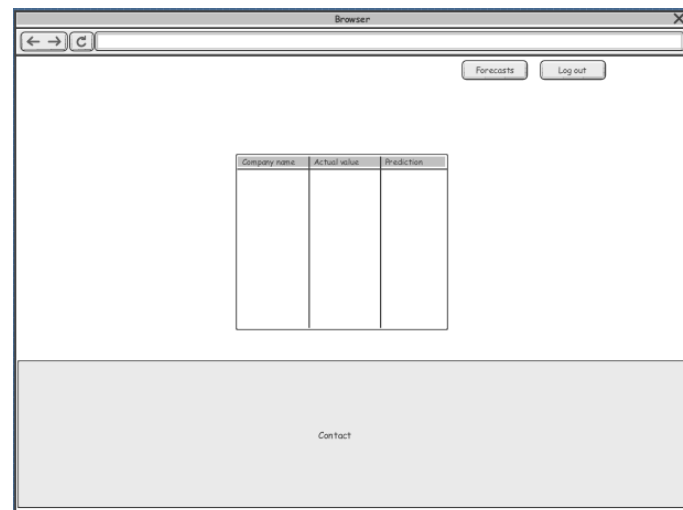
Cette interface s'affiche suite au clic de l'utilisateur sur le bouton log in.



The image shows a web browser window titled "Browser". At the top, there is a navigation bar with five buttons: "Home", "Login", "Sign up", "About us", and "Contact". Below the navigation bar, the main content area contains a "Login" label next to a text input field, and a "Password" label next to another text input field. At the bottom of the browser window, there is a grey footer bar with the text "Contact" centered.

FIGURE 3.5 – Conception de l'interface d'authentification

La figure 3.6 présente la conception de la liste des entreprises prises en compte par notre système avec leurs valeur actuelle et la prédiction de la valeur du cours boursier de demain. Cette interface s'affiche suite au clic sur le lien « Forecast » de la barre de navigation de l'interface précédente.



The image shows a web browser window titled "Browser". At the top, there is a navigation bar with two buttons: "Forecasts" and "Log out". Below the navigation bar, the main content area contains a table with three columns: "Company name", "Actual value", and "Prediction". The table has one empty row below the header. At the bottom of the browser window, there is a grey footer bar with the text "Contact" centered.

Company name	Actual value	Prediction

FIGURE 3.6 – Conception de l'interface de visualisation des prédictions de toutes les entreprises

La figure 3.7 illustre la conception d'une interface qui offre à l'utilisateur la possibilité de consulter la prédiction détaillée (prédiction à court et à long terme) pour une entreprise choisie.

Cette interface s'affiche suite au clic de l'utilisateur sur le nom d'une des compagnies se trouvant dans le tableau de l'interface précédente.

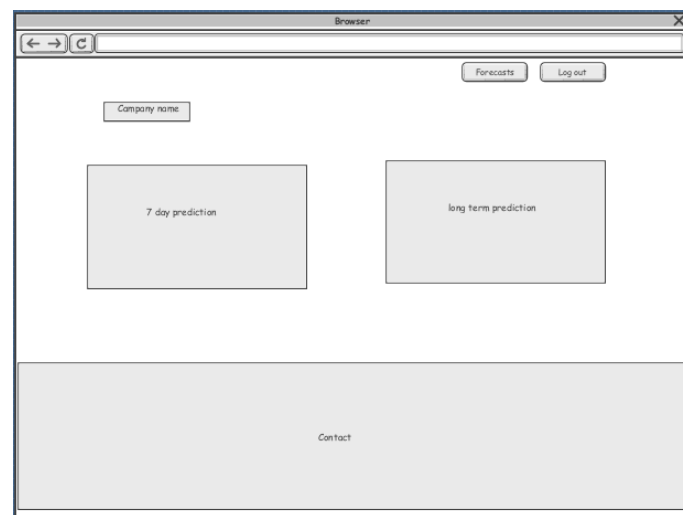


FIGURE 3.7 – Conception de l'interface de visualisation des prédictions pour une entreprise

### ● Conception détaillée "backend"

Pour illustrer la conception détaillée du Backend, nous adoptons le diagramme d'activité permettant de modéliser le comportement du système, dont la séquence des actions et leurs conditions d'exécution.[19]

Le diagramme d'activité dans la figure suivante présente un scénario d'exécution possible de notre système.

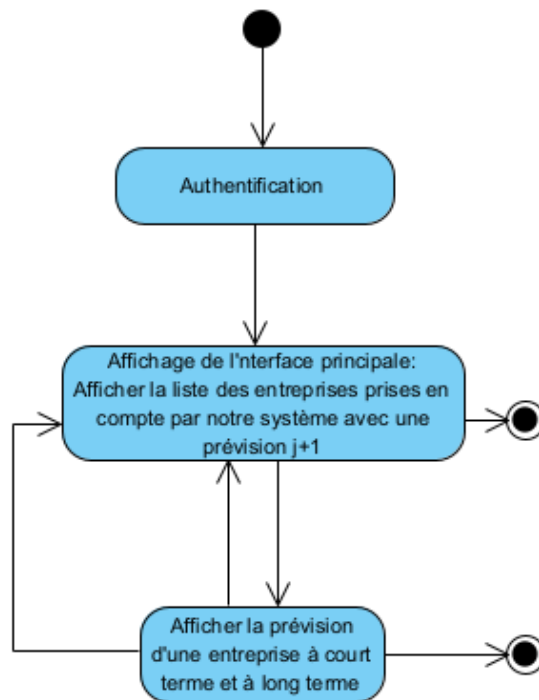


FIGURE 3.8 – Diagramme d'activité du système

## Conclusion

Dans ce chapitre, nous avons présenté l'architecture de notre système puis nous l'avons détaillé.

Nous pouvons ainsi passer au dernier chapitre de ce rapport, à savoir la réalisation.



# Chapitre 4

## Réalisation

### Introduction

Dans ce dernier chapitre, nous consacrons la première partie à la présentation de l'environnement matériel et logiciel que nous utilisons pour la réalisation du système. La deuxième partie expose la démarche suivie dans la réalisation de ce travail. La troisième partie présente les résultats de l'implémentation. Et enfin, la dernière partie décrit le déroulement chronologique du projet.

### 4.1 Environnement de travail

Pour réaliser ce projet, nous avons eu recours à divers environnements matériel et logiciel que nous décrivons dans cette section.

#### 4.1.1 Environnement matériel

- Premier poste de travail :
  - Marque : MSI
  - Processeur : Intel Core i7 8th Generation
  - Mémoire : 16.00 Go
  - Disque Dur : 1 To + 128 Go SSD
  
- Deuxième poste de travail : -Marque : HP
  - Processeur : Intel Core i5 8th Generation
  - Mémoire : 8.00 Go
  - Disque Dur : 1 To

- Troisième poste de travail : -Marque : LENOVO
- Processeur : Intel Core i5 9th Generation
- Mémoire : 8.00 Go
- Disque Dur : 2 Tera + 256 Go SSD

### 4.1.2 Environnement logiciel

#### *-Google Colaboratory - CoLab*

Google Colaboratory ou Colab est un outil gratuit et simple d'utilisation. Il permet d'utiliser l'environnement de développement Jupyter Notebook sur un navigateur sans imposer de configuration particulière à l'utilisateur.

Il offre l'accès à un processeur graphique GPU et supporte le travail collaboratif en temps réel.

C'est aussi un outil adapté à l'analyse des données, au Machine learning et au Deep learning.

#### *-Visual Studio Code*

Visual Studio Code est un éditeur de code léger mais puissant extensible.

Il permet la prise en charge du débogage, la mise en évidence de la syntaxe, la complétion intelligente du code, les snippets, la refactorisation du code et Git intégré.

#### *-Postman*

Postman est un client REST proposé par Google. Il est disponible sous la forme d'une extension Chrome ou bien d'une application stand-alone.

L'outil permet de tester des Restful API en créant des requêtes sous la forme d'url vers le service REST en spécifiant des paramètres d'entête, de chaîne de requête, d'authentification ou encore le verbe HTTP à utiliser.

Ces requêtes peuvent être enregistrées dans des collections et des sous collections.

#### *-Overleaf*

Overleaf est un éditeur LaTeX en ligne, collaboratif en temps réel.

LaTeX est un langage et un système de composition de documents qui permet de rédiger des documents dont la mise en page se fait automatiquement suivant des normes typographiques.

Un des avantages majeurs de LaTeX est sa capacité à prendre en charge des formules complexes.

#### *-Visual Paradigm*

Visual Paradigm est un outil de modélisation UML qui prend en charge tous les diagrammes de UML2.

-Trello

Trello est un outil collaboratif de gestion de projet.

Il permet de renseigner les collaborateurs sur le projets et sur son état d'avancement rapidement grâce à une présentation par tableaux, ce qui accélère les processus de travail.

## 4.2 Explication du travail réalisé

### 4.2.1 Recherche d'une dataset

Nous cherchons l'historique des cours boursiers des entreprises que nous voulons étudier dans notre projet.

Nous sélectionnons YAHOO FINANCE qui en plus d'être un des meilleur site web d'information et de médias, fournit l'historique des valeurs de cours boursiers de plus de 150000 entreprises.

Pour extraire ces données, nous utilisons l'API de YAHOO FINANCE.

### 4.2.2 Description de la dataset

La figure ci-dessous présente les différentes informations que fournit YAHOO FINANCE pour chaque entreprise.

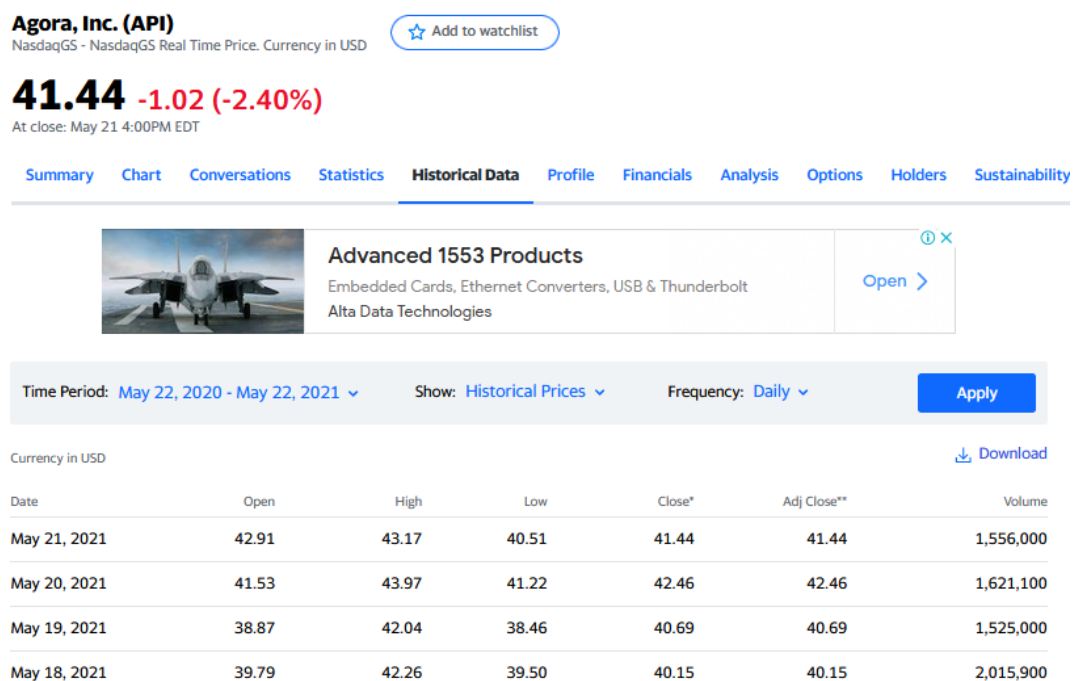


FIGURE 4.1 – DataSet Yahoo Finance

Pour chaque entreprise et chaque date, YAHOO FINANCE renseigne :

- Open : prix du cours boursier d'une entreprise à l'ouverture de la du marché boursier.
- High : prix de l'action d'une entreprise qui a atteint le plus haut niveau au cours de cette journée.
- Low : prix de l'action d'une entreprise qui a atteint le plus bas niveau au cours de cette journée.
- Close : prix de la dernière action vendue d'une entreprise au cours de cette journée.
- Adj. Close : prix de l'action ajusté pour les dividendes.
- Volume : nombre d'actions négociées au cours de cette journée.

Pour notre système, nous voulons prédire une seule valeur du cours boursier pour toute la journée qui fournit au trader une moyenne du prix de l'action au cours d'une journée. Pour se faire, nous nous basons sur la moyenne entre le prix d'ouverture (open) et le prix de fermeture (close).

Pour notre système, nous considérons 18 entreprises technologiques cotées dans la bourse NASDAQ à savoir APPLE (AAPL), AMAZON (AMZN), GOOGLE (GOOGL) , FACEBOOK (FB), INTEL (INTC), MICROSOFT (MSFT), NVIDIA (NVDA), NETFLIX (NFLX), TESLA (TSLA), ZOOM (ZM), PAYPAL (PYPL), HP (HPQ), CISCO (CSCO), IBM (IBM), APPLIED MATERIALS (AMAT), ADVANCED MICRO DEVICES (AMD), QUALCOMM (QCOM), ADOBE (ADBE).

### 4.2.3 Application de l'approche statistique : ARIMA

- Choix de dataset :

Nous prenons une dataset univariée composée des valeurs (la moyenne de l'open et du close) du cours boursier à prédire indexé par les dates correspondantes.

- Test de stationnarité :

Nous effectuons un test de Dicky fuller pour déterminer la stationnarité de notre dataset. Pour les 18 entreprises que nous analysons seules 2 sont stationnaires (IBM et INTEL).

- Détermination du paramètre I du modèle :

- Pour les datasets stationnaires :  $I = 0$
- Pour les datasets non stationnaires : nous avons créé une nouvelle série temporelle telle que  $y(t) = x(t) - x(t-1)$  (différentiation d'ordre 1). La série ainsi obtenue est stationnaire, donc  $I = 1$

- Détermination des paramètres p et q :

Nous avons réalisé les graphes d'auto-corrélation et d'auto-corrélation partielle.

- Pour les datasets stationnaires : dans la figure qui suit, nous présentons les graphes ACF

et PACF de l'entreprise IBM.

Nous remarquons d'après le graphique ACF que les autocorrélations sont significatives pour un grand nombre de décalages. Mais d'après le graphique PACF, nous remarquons la présence de pics significatifs uniquement au décalage 1 et 2, donc toutes les autocorrélations d'ordre supérieur sont expliquées par la propagation de l'autocorrélation au lag-1 et lag-2.

Ainsi, nous pouvons conclure que  $p = 2$  et  $q = 1$ .

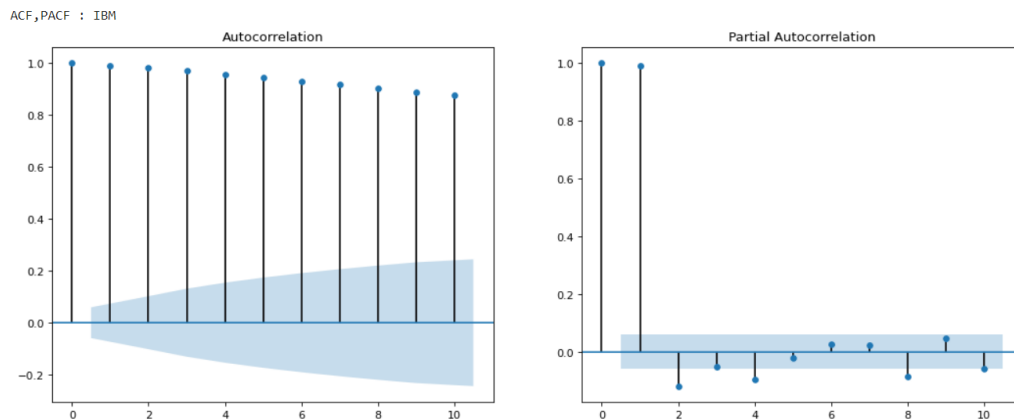


FIGURE 4.2 – ACF et PACF - IBM

- Pour les datasets non stationnaires :

Dans la figure qui suit, nous présentons les graphes ACF et PACF de l'entreprise HPQ. Ces tests doivent être réalisés sur la dataset stationnarisée (obtenue après différentiation). La bande en bleu indique l'erreur standard c'est-à-dire que tout pic n'ayant pas dépassé cette bande ne sera pas pris en considération.

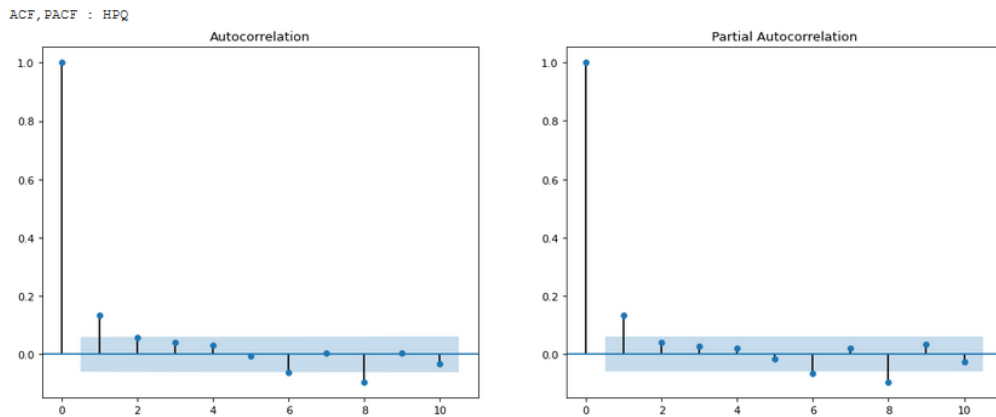


FIGURE 4.3 – ACF et PACF - HPQ

Dans la figure ci-dessous, nous présentons un cas de série non stationnaire, les paramètres  $p$  et  $q$  sont ainsi égaux à 1 et 1 respectivement.

- Entrainement du modèle :

Application du modèle ARIMA sur notre dataset.

- Évaluation du modèle :

Nous comparons les valeurs réelles de la partie test aux valeurs prédites par les modèles ARIMA. Le résultat obtenu est mauvais puisque, comme le montre la figure suivante, le MAE et le RMSE sont de 4 au minimum (CSCO) ce qui signifie que les modèles obtenus se trompent de 4 dollars dans leurs prédictions.

Dans le domaine du trading, une telle erreur est inacceptable.

Mais le modèle ARIMA fournit un intervalle dans lequel la valeur peut se trouver. Nous comptons les valeurs réelles qui appartiennent à cet interval par rapport au nombre total des valeurs de test. Les résultats obtenus sont assez satisfaisants exception faite de AMZN, AMAT.

	mae	rmse	accuracy
AAPL	10.60	12.41	0.9705882352941176
AMZN	176.06	214.96	0.9926470588235294
GOOG	310.82	351.60	0.3088235294117647
FB	16.40	19.79	0.9779411764705882
INTC	9.10	9.88	0.5411764705882353
MSFT	10.27	12.67	0.9926470588235294
HPQ	4.96	5.96	0.7647058823529411
NVDA	28.59	36.22	1.0
NFLX	33.32	41.37	0.9926470588235294
PYPL	14.91	21.15	0.8602941176470589
CSCO	3.79	4.73	1.0
ADBE	40.16	45.48	0.8382352941176471
AMD	6.49	7.76	0.8161764705882353
AMAT	29.31	31.88	0.051470588235294115
IBM	6.15	7.25	1.0
QCOM	18.30	20.58	0.9117647058823529
TSLA	94.54	105.58	0.6029411764705882
ZM	51.10	58.58	1.0

FIGURE 4.4 – ACF et PACF - HPQ

#### 4.2.4 Application de l'approche machine learning : Decision Tree

- Choix de la dataset :

Nous prenons la même dataset que dans le modèle précédent.

- Preprocessing de la dataset :

Nous adoptons la méthode de la fenêtre coulissante qui permet au modèle de prendre en considération plusieurs pas antérieurs dans sa prédiction future.

Il existe deux cas possibles, la prédiction en "one-step" et la prédiction en "multi-step".

La prédiction en "one-step" permet de prédire une seule valeur (celle du jour  $j+1$ ) en se basant sur plusieurs valeurs passées (le nombre de valeurs prises en considération représente le window size).

La prédiction "multi-steps" permet de prédire plusieurs valeurs (celles des jours  $j+1 \dots j+n$ ) en se basant sur plusieurs valeurs passées.

Il faut donc préparer les données en conséquence.

- Entraînement du modèle :

Application de l'algorithme de decision tree sur la train dataset.

- Évaluation du modèle :

L'analyse de la figure suivante montre que les prédictions des modèles de decision tree sont plus ou moins satisfaisants pour HPQ, CISCO et ADM seulement.

	mae	rmse
AAPL	4.97	6.35
AMZN	83.90	113.59
GOOG	125.92	147.97
FB	7.40	9.54
INTC	1.72	2.48
MSFT	7.73	9.28
HPQ	0.64	0.87
NVDA	14.50	18.65
NFLX	19.86	24.93
PYPL	17.45	24.04
CSCO	1.00	1.34
ADBE	14.52	17.56
ADM	0.93	1.26
AMAT	13.40	16.62
IBM	3.45	4.60
QCOM	26.93	29.32
TSLA	99.23	133.57
ZM	47.62	57.17

FIGURE 4.5 – Évaluation des modèles de Decision Tree

#### 4.2.5 Application de l'approche deep learning : RNN-LSTM

Pour cette approche nous appliquons deux types de réseaux de neurones.

##### ● RNN-LSTM univarié

Dans ce type de réseaux, nous considérons uniquement le prix de l'action dans la prédiction. Nous utilisons l'architecture "many-to-one" des réseaux de neurones récurrent pour le modèle RNN univarié one-step. Ce type de réseau prend en paramètre plusieurs valeurs pour en produire une seule.

Par contre, nous utilisons l'architecture "many-to-many" pour les modèles RNN univariés multi-steps qui prennent en paramètre plusieurs valeurs et permettent d'en produire plusieurs aussi.

- Choix de la dataset :

Nous considérons la même dataset que précédemment (decision tree).



- Preprocessing de la dataset :

Nous avons réalisé le même traitement de la dataset que précédemment puisque nous avons le même input et le même output.

- Choix d'une architecture pour le réseau de neurones :

Nous avons testé cinq modèles sur notre dataset avec différentes valeurs pour les hyperparamètres (window size, nombre de neurones par couche ...) et différents nombre de couches. La figure suivante présente un des modèles appliqués.

Model: "sequential"

Layer (type)	Output Shape	Param #
LSTM (LSTM)	(None, 32, 63)	16380
lstm (LSTM)	(None, 10)	2960
Output (Dense)	(None, 1)	11

FIGURE 4.6 – Architecture d'un modèle RNN-LSTM

Ce réseau est composé de 3 couches. La première et la deuxième couche sont des couches de neurones LSTM. La troisième et dernière couche du réseau, la couche Dense est utilisée pour produire des prédictions.

- Compilation du modèle :

Cette phase est importante puisqu'elle permet de spécifier :

- optimiseur du modèle : "ADAM"
- loss : le modèle essaye de minimiser cette fonction d'évaluation des erreurs lors du training.
- une métrique pour la mesure de l'erreur : MAE mean squared error

- Apprentissage du modèle :

La figure suivante présente la trace d'exécution du training d'un de nos modèles RNN-LSTM.

```
Epoch 1/70
122/122 [=====] - 4s 12ms/step - loss: 0.0315 - val_loss: 0.1601
Epoch 2/70
122/122 [=====] - 1s 7ms/step - loss: 0.0390 - val_loss: 0.0408
Epoch 3/70
122/122 [=====] - 1s 7ms/step - loss: 0.0354 - val_loss: 0.0195
.....
122/122 [=====] - 1s 7ms/step - loss: 0.0200 - val_loss: 0.0236
Epoch 36/70
122/122 [=====] - 1s 7ms/step - loss: 0.0122 - val_loss: 0.0293
Epoch 37/70
122/122 [=====] - 1s 7ms/step - loss: 0.0122 - val_loss: 0.0121
Epoch 38/70
122/122 [=====] - 1s 7ms/step - loss: 0.0128 - val_loss: 0.0733
Epoch 39/70
```

FIGURE 4.7 – Trace d'exécution d'un modèle RNN-LSTM

Nous utilisons dans cette phase le principe du « EarlyStopping » qui permet de stopper l'apprentissage du modèle avant qu'il n'atteigne un état d'overfitting et d'accélérer la phase de training des modèles.

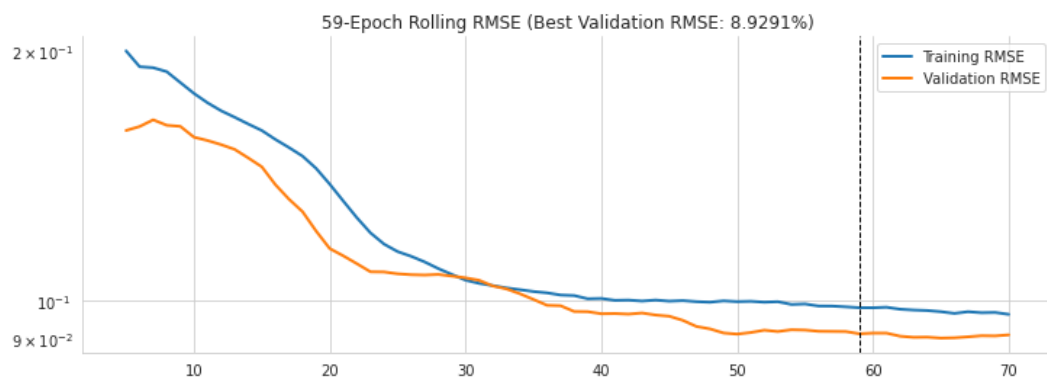


FIGURE 4.8 – Illustration de l'apprentissage du modèle

- Évaluation du modèle :

L'évaluation des modèles RNN-LSTM univariés est résumée dans le tableau suivant.

Dans la figure suivante, les deux premiers tableaux résument les erreurs entre les valeurs réelles du test et les valeurs prédites par les modèles.

Nous remarquons que pour INTEL (INTC), le premier modèle univarié "one-step" donne le meilleur résultat avec une mae de 0.713, par contre pour AMAZON (AMZN) cette métrique est assez élevée, de l'ordre de 24.

**Modèle RNN-LSTM univarié one-step**

	mae1	mae2	mae3	mae4	mae5
AAPL	2.198	3.790	3.454	2.835	1.829
AMZN	74.098	59.394	79.218	54.645	46.773
GOOG	34.079	140.110	33.604	161.606	52.714
FB	6.219	5.811	8.075	7.783	7.029
INTC	0.713	0.763	1.803	4.854	2.307
MSFT	4.965	8.216	16.921	15.843	7.011
HPQ	1.212	2.598	6.313	6.118	1.644
NVDA	17.076	13.592	19.731	16.639	12.641
NFLX	7.530	6.205	8.773	6.919	7.440
PYPL	14.837	5.829	36.941	21.656	5.861
CSCO	0.600	0.423	2.072	2.597	0.650
ADBE	7.148	4.835	9.849	6.205	7.653
ADM	0.724	0.876	2.537	1.675	0.588
AMAT	14.859	11.856	12.894	26.890	3.897
IBM	0.946	0.844	4.894	1.028	1.659
QCOM	1.622	1.840	4.281	2.232	1.708
TSLA	43.728	45.585	261.361	201.478	30.536
ZM	23.366	20.021	19.469	13.540	16.569

**Modèle RNN-LSTM univarié multi-steps**

	mae1	mae2	mae3	mae4	mae5
AAPL	1.667	1.182	3.433	2.080	1.907
AMZN	67.102	21.966	32.388	35.630	41.419
GOOG	15.358	16.822	36.784	35.337	35.953
FB	3.501	3.345	2.892	2.646	3.729
INTC	0.737	1.171	1.250	1.440	2.757
MSFT	2.186	1.639	5.317	2.122	3.649
HPQ	1.001	0.992	6.587	3.298	6.970
NVDA	16.952	6.834	24.274	14.028	10.793
NFLX	5.541	5.066	6.120	5.563	6.648
PYPL	9.293	4.607	20.196	8.273	4.843
CSCO	0.789	0.374	5.708	4.304	0.582
ADBE	7.082	5.639	18.757	7.559	6.272
ADM	0.758	0.500	1.787	1.163	1.408
AMAT	6.987	8.811	5.626	13.168	9.185
IBM	0.795	0.882	0.996	0.802	0.792
QCOM	1.555	1.458	2.602	1.639	4.606
TSLA	23.963	14.309	49.186	159.502	29.455
ZM	15.203	12.726	23.056	16.250	11.673

FIGURE 4.9 – Évaluation des modèles RNN-LSTM univariés

### ● RNN-LSTM multivarié

Dans ce type de réseau, nous considérons plusieurs features dans la prédiction.

- Choix de la dataset :

A partir de notre dataset, nous extrayons de nouveaux features qui pourraient nous aider dans notre tâche :

- Des indicateurs temporelles : jour (Day), mois (Month), année (Year), numéro du jour de l'année (DayOfYear), numéro du jour de la semaine (DayOfWeek), numéro de la semaine de l'année (WeekOfYear).
- Des indicateurs boursiers : Ce sont des formules mathématiques se basent sur le prix et/ou le volume d'un titre sur une période donnée dans le but de prédire l'évolution des cours.[24]

Nous prenons trois types d'indicateurs à savoir :

+Indicateur de volume : permet de mesurer la dynamique du volume d'un cours boursier. Nous adoptons le Volume Weighted Average Price (VWAP) ou le prix moyen pondéré.

+Indicateur de momentum : permet de mesurer la vitesse d'évolution des cours boursiers sur une période donnée c'est à dire la phase d'accélération et de décélération des prix. Nous adoptons le Relative Strength Index (RSI) ou l'indice de force relative.

+Indicateur de volatilité : permet de mesurer le rythme d'évolution des cours boursiers, le mouvement des prix. Nous adoptons le AverageTrueRange(ATR).

Nous cherchons aussi des features que nous récupérons de l'API YAHOO FINANCE aussi :

- NASDAQ index : est un des indices américains le plus important qui regroupe 100 des plus grandes entreprises non financières cotées à la bourse du Nasdaq. Il se compose essentiellement d'entreprises technologiques à hauteur de 54%.
- SP500 index : est un des indices américains les plus importants qui regroupe 500 des plus grandes entreprises cotées dans les bourses NASDAQ ou NYSE.

Afin de déterminer les features que nous pouvons utiliser dans notre prédiction, nous réalisons un test de corrélation.

Ce test permet de mesurer la dépendance d'une variable par rapport à une autre. Il est compris entre -1 et 1. Si le coefficient est proche de 1 ou -1, la dépendance est forte, alors que si elle est proche de 0, la dépendance est faible.

La figure suivante présente le cas APPLE.

Seuls NASDAQ, YEAR, ATR et VWAP sont corrélés avec  $OpenClose_{Average}$  et ces features sont tous les quatre auto-corrélés entre eux. Ainsi nous prenons un seul de ces quatre features avec le prix dans notre modèle multivarié pour APPLE.

Après analyse des différentes Heatmaps, nous obtenons 2 groupes d'entreprises. Un premier auquel nous pouvons considérer en plus du prix le VWAP et pour le deuxième, le

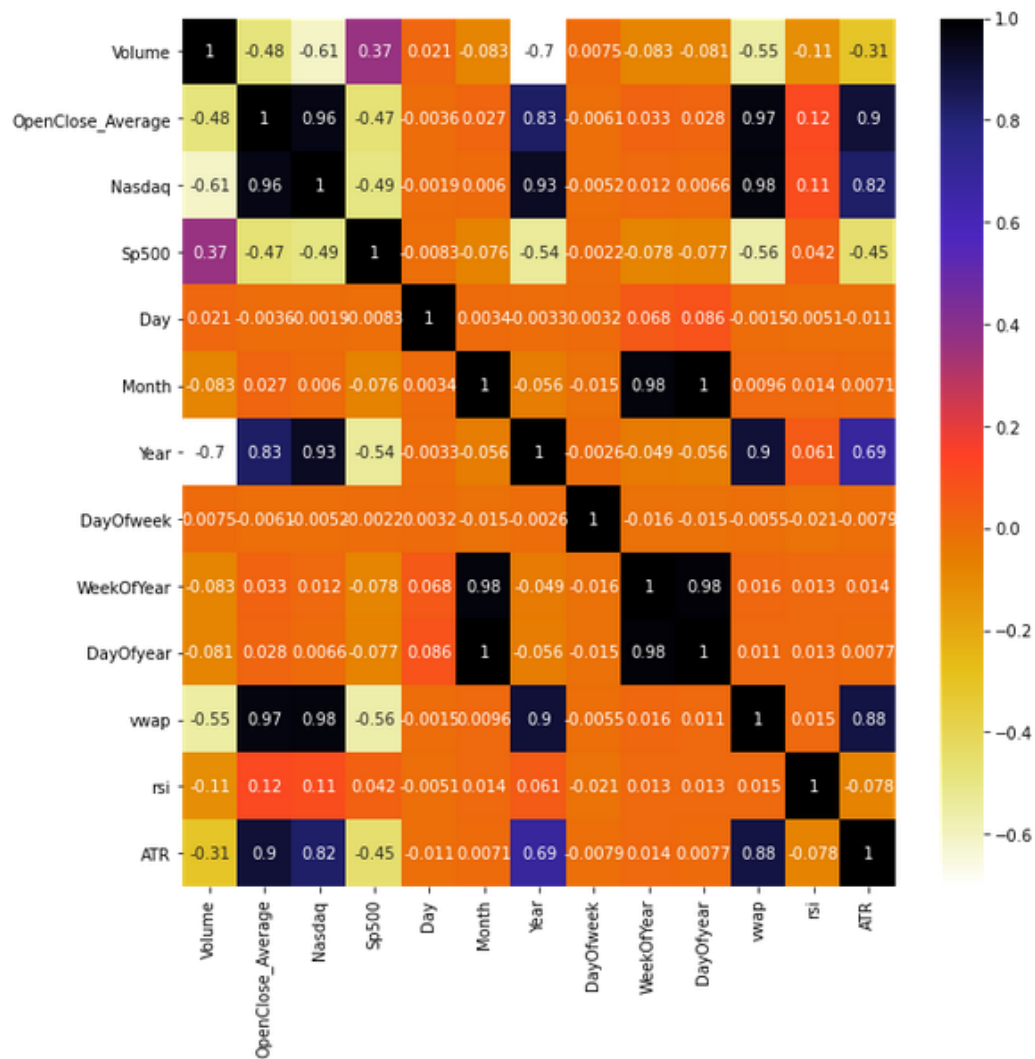


FIGURE 4.10 – Test de corrélation de features - APPLE

VWAP et NASDAQ index dans la prédiction.

- Preprocessing de la dataset :

L'input de notre modèle est une matrice de valeur et l'output est un vecteur de valeurs, il faut donc préparer la dataset en conséquence.

- Choix d'une architecture :

Nous avons testé trois modèles sur notre dataset avec différentes valeurs pour les hyperparamètres (window size, nombre de neurones par couche ...) et différents nombre de couches. La figure suivante présente un des modèles appliqués.

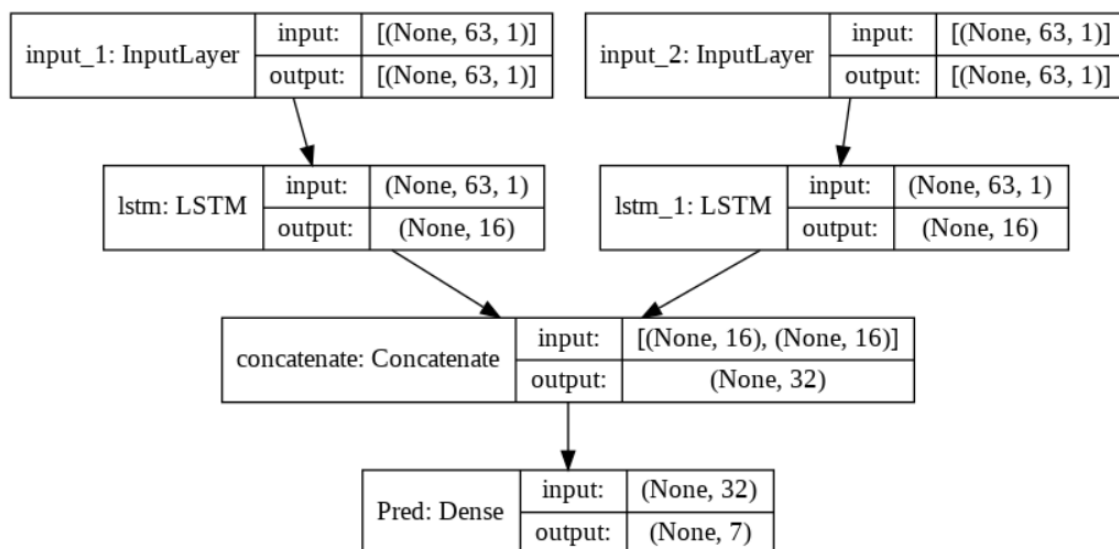


FIGURE 4.11 – Architecture d'un modèle RNN-LSTM

- Évaluation du modèle :

L'évaluation des modèles RNN-LSTM multivariés est présentée dans la figure suivante. Les résultats obtenus ne sont pas satisfaisants dans la majorité des cas car en moyenne le meilleur modèle des trois a une mae de 23.

Groupe 1				Groupe 2			
	mae1	mae2	mae3		mae1	mae2	mae3
AAPL	11.666	6.223	6.590	ADM	2.848	3.241	3.022
AMZN	296.712	155.868	167.132	HPQ	1.485	1.801	1.551
GOOG	181.754	91.092	105.965				
FB	25.820	13.249	14.912				
INTC	4.537	2.167	2.689				
MSFT	20.691	11.002	11.803				
ZM	37.910	11.715	26.773				
NVDA	52.680	27.743	30.735				
NFLX	51.143	26.928	28.633				
TSLA	60.184	36.302	31.580				
PYPL	19.396	11.282	11.381				
ADBE	43.881	22.483	24.877				
AMAT	7.962	4.616	4.309				
QCOM	9.949	5.364	5.781				
CISCO	2.038	1.387	1.494				
IBM	0.887	0.798	0.844				

FIGURE 4.12 – Évaluation des modèles RNN-LSTM multivariés

## 4.3 Résultats du travail réalisé

### 4.3.1 Choix des modèles IA

Dans cette partie, nous sélectionnons les modèles ML prêt à être déployés que nous utiliserons dans notre application.

#### Modèles pour la prédiction à court terme

Le modèle de decision tree est en overfitting, c'est à dire que le modèle prédictif est trop bien appris et que donc il ne peut pas généraliser et faire une prédiction correcte (erreur minimale) sur de nouvelles données, donc nous ne l'utiliserons pas dans notre modèle.

Nous étudions maintenant les résultats des modèles RNN-LSTM. La figure suivante présente un résumé de l'évaluation de ces modèles.

Les tableaux de la première ligne présentent les MAE des modèles et les tableaux de la

deuxième ligne présentent la différence entre l'erreur de la prédiction sur les valeurs du train et l'erreur de la prédiction sur les valeurs du test. Ces tableaux sont importants car ils permettent de savoir si le modèle est en overfitting ou non.

Modèle RNN-LSTM univarié one-step						Modèle RNN-LSTM univarié multi-steps						RNN-LSTM multivarié multisteps			
	mae1	mae2	mae3	mae4	mae5		mae1	mae2	mae3	mae4	mae5		mae1	mae2	mae3
AAPL	2.198	3.790	3.454	2.835	1.829	AAPL	1.667	1.182	3.433	2.080	1.907	AAPL	11.666	6.223	6.590
AMZN	74.098	59.394	79.218	54.645	46.773	AMZN	87.102		32.388	35.630	41.419	AMZN	296.712	155.868	167.132
GOOG	34.079	140.110	33.604	161.606	52.714	GOOG	15.358	16.822	36.784	35.337	35.953	GOOG	181.754	91.092	105.965
FB	6.219	5.811	8.075	7.783	7.029	FB	3.501	3.345	2.892	2.646	3.729	FB	25.820	13.249	14.912
INTC	0.713	0.763	1.803	4.854	2.307	INTC	0.737	1.171	1.250	1.440	2.757	INTC	4.537	2.167	2.689
MSFT	4.965	8.216	16.921	15.843	7.011	MSFT	2.186	1.639	5.317	2.122	3.649	MSFT	20.691	11.002	11.803
HPQ	1.212	2.598	6.313	6.118	1.644	HPQ	1.001	0.992	6.587	3.298	6.970	HPQ	1.485	1.801	1.551
NVDA	17.076	13.592	19.731	16.639	12.641	NVDA	16.952	6.834	24.274	14.028	10.793	NVDA	52.680	27.743	30.735
NFLX	7.530	6.205	8.773	6.919	7.440	NFLX	5.541	5.066	6.120	5.563	6.648	NFLX	51.143	26.928	28.633
PYPL	14.837	5.829	36.941	21.656	5.861	PYPL	9.293	4.607	20.196	8.273	4.843	PYPL	19.396	11.282	11.381
CSCO	0.600	0.423	2.072	2.597	0.650	CSCO	0.789	0.374	5.708	4.304	0.582	CISCO	2.038	1.387	1.494
ADBE	7.148	4.835	9.849	6.205	7.653	ADBE	7.082	5.639	18.757	7.559	6.272	ADBE	43.881	22.483	24.877
AMD	0.724	0.876	2.537	1.675	0.588	AMD	0.758	0.500	1.787	1.163	1.408	AMD	2.848	3.241	3.022
AMAT	14.859	11.856	12.894	26.890	3.897	AMAT	6.987	8.811	5.626	13.168	9.185	AMAT	7.962	4.616	4.309
IBM	0.946	0.844	4.894	1.028	1.659	IBM	0.795	0.882	0.996	0.802	0.792	IBM	0.887	0.798	0.844
QCOM	1.622	1.840	4.281	2.232	1.708	QCOM	1.555	1.458	2.602	1.639	4.606	QCOM	9.949	5.364	5.781
TSLA	43.728	45.585	261.361	201.478	30.536	TSLA	23.963	14.309	49.186	159.502	29.455	TSLA	60.184	36.302	31.580
ZM	23.366	20.021	19.469	13.540	16.569	ZM	15.203	12.726	23.056	16.250	11.673	ZM	37.910	11.715	26.773

	diff_1	diff_2	diff_3	diff_4	diff_5		diff_1	diff_2	diff_3	diff_4	diff_5		diff_1	diff_2	diff_3
AAPL	0.348	0.843	2.688	1.485	2.799	AAPL	1.353	2.040	0.146	0.156	0.454	AAPL	7.931	3.993	5.086
AMZN	55.725	11.453	19.965	20.120	152.550	AMZN	1.384	5.841	107.536	20.726	72.404	AMZN	170.103	71.217	109.405
GOOG	8.570	9.569	20.209	31.930	26.464	GOOG	42.850	7.170	40.929	182.705	59.665	GOOG	113.018	52.989	76.295
FB	1.700	1.160	4.012	0.777	7.940	FB	3.152	1.689	4.082	0.737	19.967	FB	16.373	7.872	11.005
INTC	0.091	0.639	0.505	0.146	1.152	INTC	0.063	0.049	1.263	1.251	0.308	INTC	2.549	1.342	1.587
MSFT	1.250	0.815	2.119	0.168	15.921	MSFT	3.676	1.425	7.406	9.486	3.530	MSFT	11.882	5.009	7.575
HPQ	0.603	0.583	5.414	2.621	3.357	HPQ	0.960	2.150	5.431	5.061	1.186	HPQ	1.200	1.395	1.402
NVDA	2.391	1.541	14.937	3.990	0.063	NVDA	14.066	3.370	17.751	3.674	10.694	NVDA	24.778	8.013	17.441
NFLX	3.931	2.461	1.579	4.149	7.708	NFLX	5.019	3.885	12.867	0.858	2.495	NFLX	30.204	13.190	19.554
PYPL	7.943	2.964	9.575	6.363	6.394	PYPL	12.888	3.093	31.662	20.870	3.540	PYPL	11.605	5.949	8.911
CSCO	1.028	0.790	3.682	1.624	1.046	CSCO	1.076	0.172	2.306	11.531	2.056	CISCO	0.236	0.551	0.446
ADBE	4.961	2.987	5.051	3.202	1.702	ADBE	3.883	2.565	2.944	0.359	23.312	ADBE	24.622	9.256	15.748
AMD	0.435	0.219	1.427	0.736	0.704	AMD	0.378	0.434	1.993	1.285	0.229	AMD	2.173	2.383	2.616
AMAT	0.907	3.057	3.637	9.132	0.592	AMAT	14.031	4.878	11.655	21.442	8.215	AMAT	3.415	1.609	2.730
IBM	0.371	1.253	6.128	6.076	1.958	IBM	1.504	3.987	39.008	5.491	17.148	IBM	4.304	1.516	1.933
QCOM	1.222	1.437	1.888	1.405	2.849	QCOM	1.254	0.818	2.731	0.852	1.676	QCOM	7.610	4.301	5.067
TSLA	19.995	11.623	48.553	140.637	5.593	TSLA	46.382	43.863	240.555	200.295	32.600	TSLA	24.003	10.957	16.233
ZM	7.325	1.190	7.716	1.909	5.750	ZM	5.714	7.852	2.818	0.647	2.056	ZM	19.057	2.517	16.241

FIGURE 4.13 – Résumé des évaluations des modèles RNN-LSTM

Nous choisissons d'appliquer pour :

- APPLE (AAPL) : le modèle univarié multi-steps 4.
- AMAZON (AMZN) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- GOOGLE (GOOGL) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.



- FACEBOOK (FB) : le modèle univarié multi-steps 4.
- INTEL (INTC) : le modèle univarié one-step 1.
- MICROSOFT (MSFT) : le modèle univarié multi-steps 2.
- HP (HPQ) : le modèle univarié multi-steps 1.
- NVIDIA (NVDA) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- NETFLIX (NFLX) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- PAYPAL (PYPL) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- CISCO (CSCO) : le modèle univarié multi-steps 2.
- ADOBE (ADBE) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- ADVANCED MICRO DEVICES (AMD) : le modèle univarié multi-steps 2.
- APPLIED MATERIALS (AMAT) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- IBM (IBM) : le modèle univarié one-step 1.
- QUALCOMM (QCOM) : le modèle univarié multi-steps 2.
- TESLA (TSLA) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.
- ZOOM (ZM) : de ne pas l'intégrer dans notre application web en raison des résultats non satisfaisants obtenus.

### Modèles pour la prédiction à long terme

Les modèles ARIMA donnent une accuracy stisfaisante pour toutes les entreprises sélectionnées lors de l'analyse des modèles Decision Tree et RNN-LSTM sauf pour INTEL (INTC) qui présente une accuracy de 0.5. Nous choisissons de ne pas l'intégrer dans notre site web.

#### 4.3.2 Interfaces de l'application web

Dans cette partie, nous exposons quelques interfaces web de notre application web. La figure suivante illustre la page d'accueil de notre site web. Cette page d'accueil est une interface dédiée à la présentation du service fournie par le site web et de l'équipe.

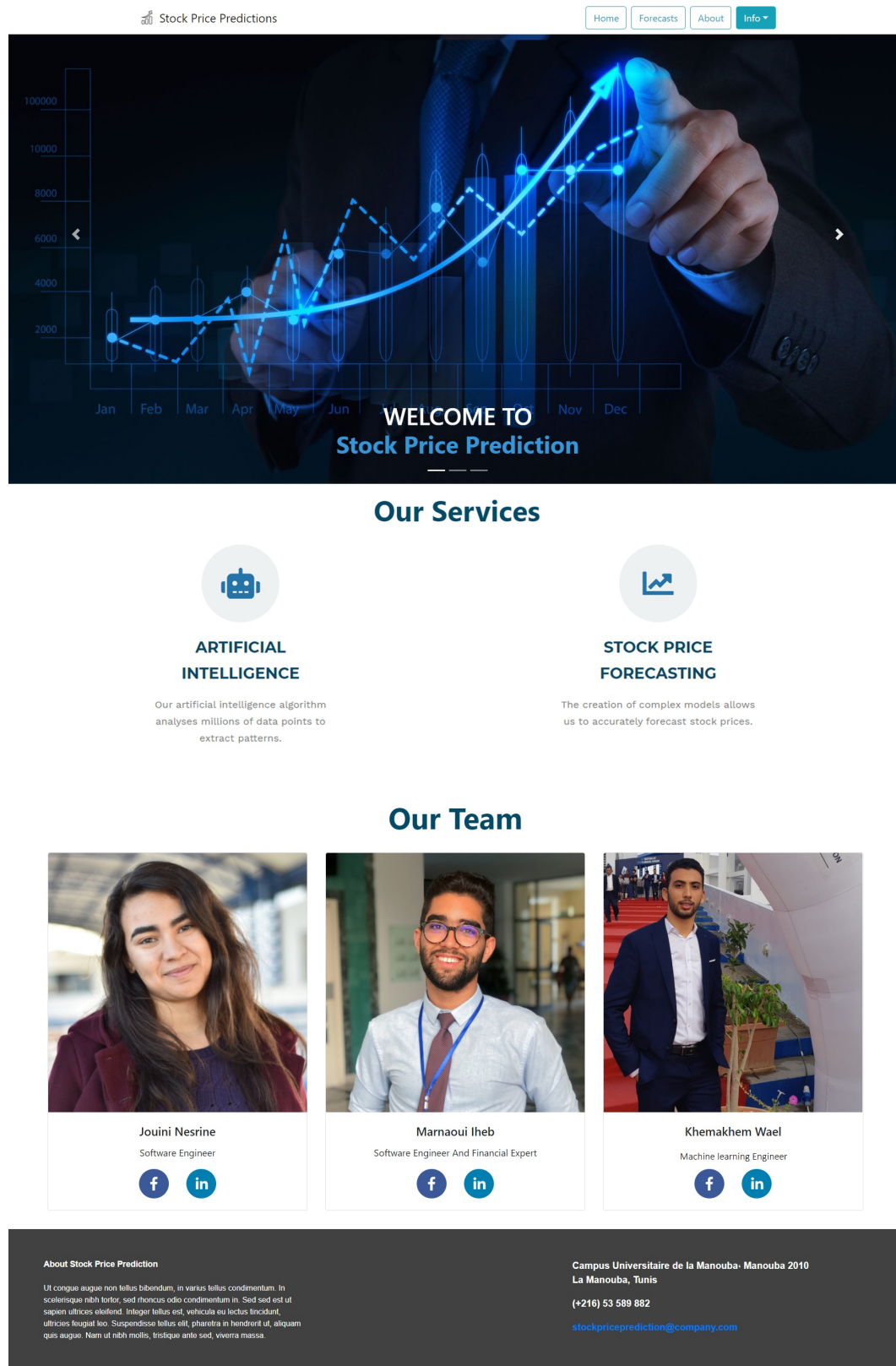
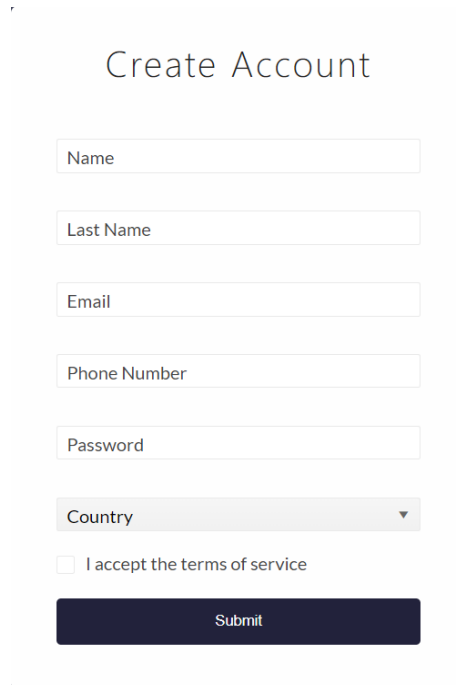


FIGURE 4.14 – Page d'accueil

La figure suivante présente l'interface d'inscription, qui permet à nos utilisateurs de s'inscrire sur notre plateforme.



The image shows a 'Create Account' registration form. It features a title 'Create Account' at the top. Below the title are several input fields: 'Name', 'Last Name', 'Email', 'Phone Number', and 'Password'. There is also a 'Country' dropdown menu. At the bottom of the form, there is a checkbox labeled 'I accept the terms of service' and a dark blue 'Submit' button.

Create Account

Name

Last Name

Email

Phone Number

Password

Country ▼

☐ I accept the terms of service

Submit

FIGURE 4.15 – La page d'inscription

La figure suivante présente l'interface qui permet au trader de visualiser la prédiction du  $j+1$  de toutes les entreprises.

Stock Price Predictions			
		Home	Forecasts
		About	Info
Symbols	Names	Actual Values	Predictions
AAPL	Apple Inc	125.09000015258789	127.4852294921875
FB	Facebook Inc	329.86500549316406	325.40032958984375
MSFT	'Microsoft Corporation'	290.33999633789062	249.23306274414062
HPQ	HP Inc	29.844999313354492	29.4942684173584
CSCO	Cisco Systems Inc	53.040000915527344	51.284183502197266
ADM	Advanced Micro Devices Inc	66.72000122070312	64.90538787841797
IBM	IBM Common	143.97500610351562	143.59764099121094
QCOM	QUALCOMM Inc	134.3449935913086	132.767578125

About Stock Price Prediction

Ut congue augue non tellus bibendum. In varius tellus condimentum. In scelerisque nibh tortor, sed rhoncus odio condimentum in. Sed sed est ut sagittis ultrices eleifend. Integer tellus est, vehicula eu lectus fincidunt, ultrices feugiat leo. Suspendisse tellus elit, pharetra in hendrerit ut, aliquam quis augue. Nam ut nibh mollis, tristique ante sed, viverra massa.

Campus Universitaire de la Manouba: Manouba 2010  
La Manouba, Tunis  
(+216) 53 589 882  
stockpriceprediction@company.com

Stock Price Predictions

© Stock Price Predictions 2021. All Rights Reserved.

FIGURE 4.16 – Tableau des entreprises

## Conclusion

Ce chapitre a été dédié à la description du déroulement du projet et à la présentation des résultats obtenus.

Nous sommes maintenant en mesure de passer à la conclusion générale afin de clôturer notre rapport.

# Conclusion et perspectives

Le présent document rend compte du travail effectué pour notre projet de conception et de développement à l'École Nationale des Sciences de l'Informatique. Le but de ce projet consistait à concevoir et développer une application web qui permet au traders d'avoir une prédiction à court et à long terme des valeurs d'actions boursières d'entreprises technologiques cotées dans la bourse NASDAQ. Ce projet a été proposé suite à une constatation que nous avons faite. Il existe peu d'outils en ligne permettant d'aider les traders dans leurs tâches en leur proposant une prédiction des valeurs de cours boursiers.

La prédiction des cours boursiers est un projet intéressant mais difficile à réaliser pour nous car il requiert des compétences transversales en finance et en science des données. Dans cette perspective, nous avons commencé par nous familiariser avec les principes financiers qui entrent dans le cadre de la prédiction des cours boursiers et dans les marchés boursiers plus généralement et nous avons étudié la tâche de prédiction d'un point de vue théorique, nous avons notamment découvert diverses approches applicables à la prédiction des séries temporelles. Nous avons ensuite analysé et conçu notre système pour arriver enfin à l'implémenter.

Les résultats obtenus pour la prédiction des cours boursiers sont certes satisfaisants pour certaines entreprises mais pour d'autres les divers modèles que nous avons réalisés donnent un taux d'erreur inacceptable. Afin de remédier à ce problème, nous envisageons d'intégrer à notre prédiction l'analyse des sentiments qui se base sur des données fournies par les réseaux sociaux ou des plateformes équivalentes.

# Netographie

- [1] [t.ly/rF21](https://t.ly/rF21) la bourse definition et fonctionnement,consulté le 13/05/2021
- [2] [t.ly/rxqa](https://t.ly/rxqa) Cotation : définition simple, traduction, synonymes ,consulté le 13/05/2021
- [3] [t.ly/JnZt](https://t.ly/JnZt) La volatilité dans la mesure du risque ,consulté le 13/05/2021
- [4] [t.ly/sxF2](https://t.ly/sxF2) Indice boursier - définition ,consulté le 13/05/2021
- [5] <https://www.aistockfinder.com> aistockfinder ,consulté le 13/05/2021
- [6] <https://walletinvestor.com> walletinvestor ,consulté le 13/05/2021
- [7] [http://www.abdelhamid-djeffal.net/web\\_documents/coursseries.pdf](http://www.abdelhamid-djeffal.net/web_documents/coursseries.pdf) Série temporelles,consulté le 13/05/2021
- [8] <https://cel.archives-ouvertes.fr/cel-00762833/document> Séries Temporelles : modèles et statistiques ,consulté le 13/05/2021
- [9] [https://olivierroustantfr.files.wordpress.com/2018/09/polycopie\\_series\\_temporelles\\_2008\\_2009.pdf](https://olivierroustantfr.files.wordpress.com/2018/09/polycopie_series_temporelles_2008_2009.pdf), consulté le 08/02/2021" INTRODUCTION AUX SÉRIES CHRONOLOGIQUES ,consulté le 13/05/2021
- [10] [https://www.univ-orleans.fr/deg/masters/ESA/CH/CoursSeriesTemp\\_Chap2.pdf](https://www.univ-orleans.fr/deg/masters/ESA/CH/CoursSeriesTemp_Chap2.pdf) Econométrie AppliquéeSéries Temporelles ,consulté le 13/05/2021
- [11] <https://otexts.com/fpp2/MA.html> Moving average models,consulté le 13/05/2021
- [12] <https://www.lpsm.paris/pageperso/lemaire/docs/stats-ENPC/series-chrono.pdf> series chronologiques ,consulté le 13/05/2021
- [13] <https://ia-data-analytics.fr/machine-learning/> machine learning ,consulté le 13/05/2021
- [14] [https://www.saedsayad.com/decision\\_tree\\_reg.htm](https://www.saedsayad.com/decision_tree_reg.htm) decision tree ,consulté le 13/05/2021
- [15] [t.ly/Fbfo](https://t.ly/Fbfo) Schematic-of-the-one-dimensional-support-vector-regression-SVR-model ,consulté le 13/05/2021
- [16] <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/> Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition
- [17] <https://www.atlassian.com/fr/agile/scrum> SCRUM ,consulté le 13/05/2021

- 
- [18] <https://weblog.wemacity.com/fr/la-methode-scrum-definition-et-mode-demploi/> SCRUM definition et mode d'emploi ,consulté le 13/05/2021
  - [19] <https://lipn.univ-paris13.fr/~gerard/uml-s2/> Cours de conception orientée objet ,consulté le 13/05/2021
  - [20] <https://arduino103.blogspot.com/2015/06/python-flask-pythonanywhere-partie-2.html> Flask ,consulté le 13/05/2021
  - [21] <https://reactjs.org> reactjs,consulté le 13/05/2021
  - [22] <https://info.blaise-pascal.fr/nsi-sockets-python> Les sockets python ,consulté le 13/05/2021
  - [23] <https://www.50a.fr/0/agence-web-design> Web design ,consulté le 13/05/2021
  - [24] [https://www.zonebourse.com/formation/espace\\_pedagogique/Les-indicateurs-techniques-117/](https://www.zonebourse.com/formation/espace_pedagogique/Les-indicateurs-techniques-117/) Web design,consulté le 13/05/2021