استخدام خوارزمية XGBOOST للتنبؤ بأسعار الإغلاق لسهم شركة Apple

Using the XGBoost algorithm to predict the closing pricesof Apple inc. stock

محـــمـد روابـــــة*

مخبر تطوير المؤسسة الاقتصادية الجزائرية، جامعة ابن خلدون تيارت – الجزائر

mohammed.rouaba@univ-tiaret.dz

تاريخ الإستلام: 2024/07/12 تاريخ القبول: 2024/09/29 تاريخ النشر:2024/10/31

ملخص:

هدفت هذه الدراسة إلى التنبؤ بأسعار الإغلاق اليومية لسهم شركة*Apple*بإستخدام أحد نماذج التعلم الآلي والمتمثل في خوارزمية *XGBoost*لمساعدة المستثمرين على إتخاذ قرار الاستثمار في مجال الأسهم. وذلك خلال الفترة الممتدة من 2020/01/02 -2024/07/11 وبلغ إجمالي عدد المشاهدات (1138) مشاهدة، بواقع (911) مشاهدة (بنسبة 80% من اجمالي المشاهدات) تم استخدامها بغرض التدريب، و(227) مشاهدة (بنسبة 20% من اجمالي المشاهدات) تم استخدامها للإختبار.تم بناء نموذج *XGBoost* باستخدام لغة البرمجة *Python* بالاعتماد على المكتبات التابعة له وأهم هذه المكتبات المستخدمة هيمكتبة *XGBoost*.وأظهرت نتائج هذه الدراسة بأن هناك قدرة تنبؤية عالية لخوارزمية *XGBoost* بناءا على بعض مقاييس الأداء االتي تم الاعتماد عليها في هذهالدراسة.

الكلمات المفتاحية: التنبؤ، أسعار الأسهم، التعلم الآلي، خوارزمية XGBoost، قرار الاستثمار.

تصنيفات JEL:C58، E30.

Abstract:

The aim of this study was to predict the daily closing prices of Apple Inc. stock using a machine learning model, specifically the XGBoost algorithm, to assist investors to making stock market investment decisions. The study covered the period from January 2, 2020, to July 11, 2024, with a total sample size of 1138 observations. Of these, 911 observations (80% of the total) were used for training, and 227 observations (20% of the total) were used for testing. The XGBoost model was built using the Python programming language, relying on its associated libraries, with the XGBoost library being the most important. The results of this study indicated a high predictive ability of the XGBoost algorithm based on certain performance metrics used in this study.

Keywords: Predicting; Stock price; Machine learning; XGBoost algorithm; Investment Decision.

Jel Classification Codes:C58;E30.

* المؤلف المراسل.

1. مقدمة:

تعتبر الأسهم الركيزة الأساسية في الأسواق المالية وأهم الأدوات المتداولة فيها، حيث يحضى الاستثمار في الأسهم النسبة الأكبر في الاستثمارات المالية مقارنة بأنواع الاستثمارات الأخرى، وتتميز الأسهم بحركة دائمة وتقلبات مستمرة في أسعارهامما يجعلها تستجيب بشكل سريع لمختلف الظروف والأحداث وتتأثر كثيرا بالبيئة المحيطة وهذا الأمر يجعل الاستثمار فيها محفوفا بالمخاطر. لذلك تتطلب عملية التداول بالأسهم ترشيدا للقرارات الاستثمارية عن طريق دراسة اتجاه الأسواق المالية والتنبؤ بالقيم المستقبلية للأسعار بما يضمن لنا الحد الأدنى من المخاطر.

اختلفت أساليب التنبؤ بأسعار السهم وتراوحت بين البسيطة التي نعتمد من خلالها على التقديرات الشخصية والأساليب التقليدية كالنماذج الإحصائية والرياضية وصولا إلى الأساليب الحديثة التي تعتمد على الذكاء الاصطناعي وأنظمة المحاكاة، إذ شهدت تطورات كبيرة خلال السنوات القليلة الماضية، وهذا كان نتيجة لمواكبة الثورات التكنولوجية، وتطور وتعدد وسائل الاتصال، وظهور الانترنت الذي بدوره أسهم بشكل فعال في عولمة وانفتاح الأسواق المالية وزيادة الترابط فيما بينها. الأمر الذي نتج عنه زيادة في تعقيد البيئة المالية الدولية، وأسهم بشكل كبير في ظهور وتنامي الأزمات المالية والاقتصادية التي تلقي بظلالها مباشرة على الأسواق المالية، إذ تستجيب أسعار الأسهم بشكل سريع للحالة المؤثرة بحيث يصعب معه التنبؤ باتجاهاتها المستقبلية، ومن ذلك ظهرت نماذج حديثة كخوارزمية XGBoost التي تستخدم لحل مشاكل تتعلق بالتصنيق والانحدار، إذ يمكنها انجاز المهام الموكلة إليها عبر أسلوب التعلم الخاضع للإشراف من خلال اكتشاف الأنماط المميزة في البيانات التاريخية وبناء التوقعات على أساسها. وأثبتت هذه الخوارزمية نجاعتها مؤخرا في العديد من التطبيقات المالية والاقتصادية كالتنبؤ بأسعار الأسهم الذي يعتبر موضوع هذه الدراسة، إذ تحاول اختبار مدى إمكانية التنبؤ بأسعار الأسهم باستخدام خوارزمية XGBoot.

1.1. **الإشكالية:** يمكن عرض مشكلة الدراسة من خلال التساؤل التالي:

ما مدى دقة خوارزمية XGBoost في التنبؤ بأسعار سهم شركة Apple ؟.

2.1. فرضيات الدراسة: للإجابة على إشكالية هذه الدراسة تم صياغة الفرضية التالية: هناك قدرة تنبؤية عالية لخوارزمية XGBoost للتنبؤ بأسعار سهم شركة Apple.

3.1. أهمية الدراسة: تكمن أهمية هذه الدراسة في محاولة المساهمة في ترشيد عملية اتخاذ القرار الاستثماري باستخدام خوارزمية XGBoost من خلال التنبؤ بأسعار الأسهم، وهذا يعتبر مجال خصب، وذلك لبدء انتشار مثل هذه التقنية المستحدثة في العديد من المجالات البحثية.

4.1. أهداف الدراسة: يكمن الهدف الرئيسي لهذه الدراسة في اختبار قدرة خوارزمية XGBoostفي التنبؤ بأسعار سهم شركة Apple.

2.خوارزميةXGBOOST

خوارزمية تعزيز التدرج الشديد XGBOOST تعتبر أحد أهم خوارزميات التعزيز الشائعة والمستخدمة على نطاق واسع، وهي مشابهة لخوارزمية التعزيز الاشتقاقي Gradient Boostلكنها تحتوي على بعض الميزات الإضافية التي تجعل منها أقوى بكثير، كما أنها أيضا تتميز بطريقة تديب سريعة بحيث يمكن موازنتها أو توزيعها عبر المجموعات.

تم تطوير هذه الخوارزمية عام 2016 من طرف Chen and Guestrin(Wang et al., 2021, p. 4)حيث فازت بالمركز الأول في مسابقة كاغلKaggle لتحليل البيانات وتم نشر ورقة بحثية بطريقة عملها حيث أثبتت نجاحا باهرا في فترة وجيزة وقد تم اقتباس هذا البحث في أكثر من 40787 مقالة علمية.

تابع الهدف لخوارزمية XGBoot يعرف بالمعادلة التالية(Chen et al., 2019, p. 13151):

$$Obj = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \qquad (1)$$

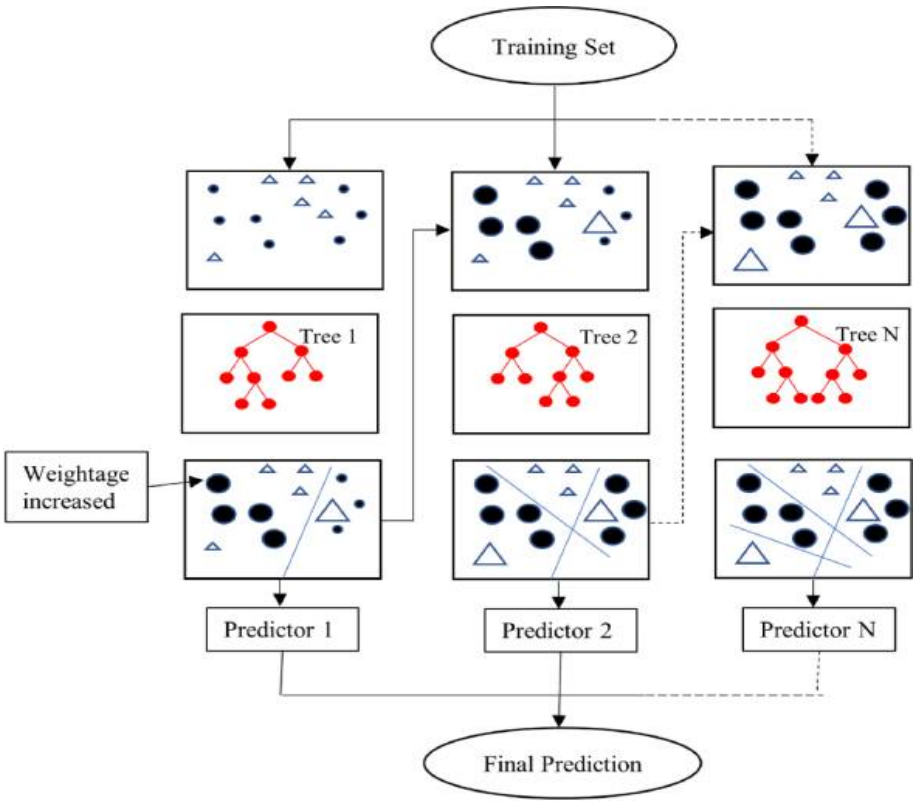
من خلال المعادلة (1) أعلاه نلاحظ أنتابع الهدف يتكون من حدين، الحد الأول يسمى تابع الخسارة ويقيس مقدار الاختلاف بين القيمة المتنبأ بها $\hat{y}_i$ وقيمة الهدفy_i. والحد الثاني يسمى تابع التنظيم يستخدم لضبط تعقيد التعلم مما يساعد على تعميم النموذج ومنع فرط التخصيص في مرحلة التعلم. ويعطى بالصيغة التالية(Deng et al., 2023, p. 3):

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

حيثγ: تمثل معلمة التعقيد،w: قيمة عقد الأوراق، T: عدد الأوراق في الشجرة، λ: معامل التنظيم.

والشكل 1 أدناه ألية عمل خوارزمية XGBoostلحل المشاكل في مجال التصنيف أو الانحدار.

الشكل 1: التمثيل البياني لعملية التنبؤ باستخدام خوارزمية XGBOOST



Source: Pant, A., & Ramana, G. (2022). Prediction of pullout interaction coefficient of geogrids by extreme gradient boosting model. *Geotextiles and Geomembranes*, *50*(6), 1188-1198,p: 1191.

3.التنبؤ بإستخدام منهجية XGBOOST

1.3.استيراد المكتبات اللازمة وتحميل البيانات

من أجل التنبؤ بأسعار الاغلاق اليومية لسهم شركة Apple باستخدامخوارزمية XGBoost ببرنامج Pythonنقوم في البداية بإستدعاء المكتبات الجاهزة المناسبة ومن أشهرها: pandas، mathplotlib، seaborn، numpy، tensorflow، sklearnXgboost. ثم نقوم بعدها بتحميل ملف البيانات التي سنعمل عليها.

2.3. تحميل البيانات على برنامج Python والتمثيل البياني لها:

في هذه الدراسة تم الاعتماد على أسعار الاغلاق اليومية لأسعار سهم شركة Apple والتي تم الحصول عليها بالاعتماد على مكتبة yfinanceالتابعة للغة البرمجة Pythonحيث بلغت عدد المشاهدات 1138 مشاهدة والتي تغطي الفترة 2020/01/02 – 2024/07/11.

جدول 1: بيانات الدراسة

	Date	Open	High	Low	Close	Adj Close	Volume
0	2020-01-02	74.059998	75.150002	73.797501	75.087502	72.960457	135480400
1	2020-01-03	74.287498	75.144997	74.125000	74.357498	72.251160	146322800
2	2020-01-06	73.447502	74.989998	73.187500	74.949997	72.826851	118387200
3	2020-01-07	74.959999	75.224998	74.370003	74.597504	72.484344	108872000
4	2020-01-08	74.290001	76.110001	74.290001	75.797501	73.650337	132079200

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

والشكل البياني رقم 2 الموالي يوضح تطور أسعار الإغلاق لسهم شركة Appleخلال الفترة 2020/01/02 – 2024/07/11.

الشكل 2: تطور أسعار الإغلاق لسهم شركة Apple خلال الفترة 2020/01/02 – 2024/07/11.

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

3.3. تقسيم بيانات الدراسة

في هذه المرحلة تم تقسيم بيانات الدراسة إلى مجموعتين، المجموعة الأولى تتمثل في مجموعة بيانات التدريب وقدرت نسبتها بـ 80% من إجمالي المشاهدات بينما المجموعة الثانية تمثلت في مجموعة بيانات الاختبار بنسبة قدرها 20%. والهدف من تقسيم البيانات هو لاختبار قدرة النموذج المبني بإستخدام خوارزمية XGBoostفي التعميم على البيانات التي لم يتدرب عليها، وقد تم تقسيم بيانات الدراسة كما هو موضح في الجدول الموالي:

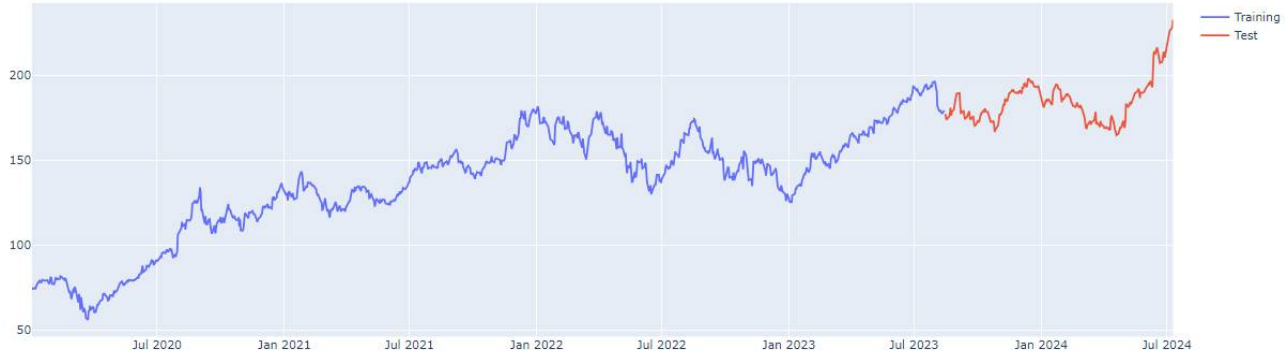
جدول 2: تقسيم بيانات الدراسة

	مجموعة التدريب	مجموعة الاختبار
نسبة التقسيم	%80	%10
الفترة	2022/08/15 – 2020/01/02	2024/07/11 – 2023/08/16
المشاهدات	911	227

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

والشكل البياني رقم03الموالي يعرض كيفية إجراء عملية التقسيم.

الشكل 3: بيانات التدريب وبيانات الاختبار



المصدر: من إعداد الباحث بالاعتماد على برنامج Python

4.3. متغيرات الدراسة:

تسمى المتغيرات التي تتعامل نماذج الذكاء الاصطناعي مع قيمها "مزايا (سمات)Features"، هذه المزايا تقسم إلى مزاياإدخال Input Features ومزايا هدف Target Features، حيث يقوم النموذج بقراءة مزايا الادخال وتحليلها ويقيس العلاقة بينها وبين بعضها وبين ميزة الهدف للتوصل إلى طريقة يتنبأ بها بقيم ميزة الهدف (Bhandari et al., 2022, p. 5).

- **ميزة الهدف:** أخذنا كميزة هدف سعر الاغلاق اليومي لسهم شركة Apple.
- **مزايا الإدخال:** الجدول التالي يوضح المزايا المستخدمة للتنبؤ بميزة الهدف.

جدول 3: مزايا الإدخال

SMA_5 ، SMA_10،SMA_15، SMA_30	المتوسط المتحرك البسيط لمدة (5، 10، 15، 30 يوم)
EMA_9	المتوسط المتحرك الأسي لمدة 9 أيام
RSI	مؤشر القوة النسبية
MACD	مؤشر تقارب /التباعد المتوسطات المتحركة

المصدر: من إعداد الباحث

وفيما يلي شرح لهذه المؤشرات المستخدمة كمزايا للإدخال:

❖ **المتوسط المتحرك البسيط:** يعتبر هذا المؤشر من بين المؤشرات الأكثر استخداما من طرف المحللين، ويحسب وفق العلاقة التالية(Pring, 2014, p. 209):

$$SMA_t = \frac{C_t + C_{t-1} + \cdots C_{t-n}}{\mathrm{n}} = \frac{1}{n}\sum_{i=1}^{n} C_{t-1}$$

حيث:

SMA_t: المتوسط المتحرك البسيط في الفترة t.

C_t , C_{t-1} , ... C_{t-n}: سعر إغلاق السهم خلال الفترات:t,t-1...t-n

n : طول فترة حساب المتوسط المتحرك.

❖ **المتوسط المتحرك الأسي:** لحساب المتوسط المتحرك الأسي تعطى أهمية نسبية مختلفة لكل سعر إغلاق للسهم، حيث تعطى أوزان كبيرة للأسعار الحديثة ثم تتناقص هذه الأوزان كلما اتجهنا نحو الماضي(لطرش، 2019، p. 141). ويحسب وفق الصيغة التالية:

$$EMA_t = \frac{1C_t + aC_{t-1} + a^2C_{t-2} + \cdots a^{n-1}C_{t-n+1}}{1 + \mathrm{a} + \mathrm{a}^n + \cdots a^{n-1}}$$

حيث:

EMA_t: المتوسط المتحرك الأسي في الفترة t.

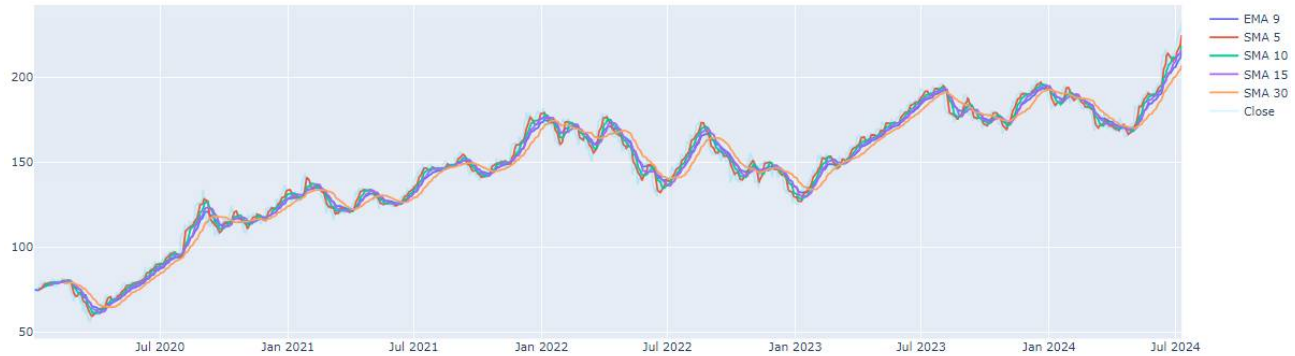
C_t , C_{t-1} , ... C_{t-n}: سعر إغلاق السهم خلال الفترات:t,t-1...t-n

n : طول فترة حساب المتوسط المتحرك.

a: معامل الترجيح، حيث $a = \frac{2}{n+1}$.

والشكل البياني رقم 4الموالي يعرض أسعار الاغلاق اليومية لسهم شركة Apple، والمتوسطات المتحركة البسيطة لمدة (5، 10، 15، 30 يوم) بالإضافة إلى المتوسط المتحرك الأسي لمدة 9 أيام.

الشكل 4:أسعار الاغلاق اليومية لسهم شركة Apple،المتوسطات المتحركة البسيطة والمتوسط المتحرك الأسي



المصدر: من إعداد الباحث بالاعتماد على برنامج Python

❖ **مؤشر القوة النسبية**: هو مؤشر شائع الاستخدام تتراوح قيمه بين (0 و100) يحسب رياضيا من خلال نسبة متوسطات التغيرات الإيجابية إلى التغيرات السلبية.والشكل البياني رقم 5 التالي يعرض مؤشر القوة النسبية لأسعار الاغلاق اليومية لسهم شركة Apple.

الشكل 5: مؤشر القوة النسبية

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

❖ **مؤشر تقارب /التباعد المتوسطات المتحركة:**وضع هذا المؤشر منطرف الباحث Gerald Appelفي سنة 1981 ويعتبر من بين أكثر المؤشرات المستخدمة لسهولته ومصداقيته. ويتم حسابه وفق الصيغة التالية(Vu et al., 2017, p. 1445):

$$MACD = EMA_{12} - EMA_{26}$$

والشكل البياني رقم 6 التالي يعرض مؤشر تقارب /التباعد المتوسطات المتحركة.

الشكل 6: مؤشر تقارب /التباعد المتوسطات المتحركة

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

والجدول رقم 4الموالي يعرض بعض قيم مشاهدات مزايا الإدخال كما تظهر في برنامج Python.

جدول 4: بعض القيم لمزايا الإدخال

	EMA_9	SMA_5	SMA_10	SMA_15	SMA_30	RSI	MACD
30	79.835177	80.662999	79.923500	79.835000	78.647250	62.073647	0.654183
31	79.980972	80.909000	80.309500	79.945667	78.852250	51.031981	0.555134
32	79.957053	80.781500	80.568000	80.113167	79.032000	49.414062	0.556832
33	80.054871	80.982001	80.687250	80.212000	79.230500	47.223514	0.489931
34	80.056941	80.637000	80.658499	80.144667	79.413083	53.095466	0.296309

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

5.3. تحديد معالم الخوارزمية

عند استخدام خوارزمية XGBoost نحتاج إلى ضبط معالم هذه الخوارزمية وفقا لخصائص سعر السهم، هذه المعالم موضحة فيما يلي(Zhang, 2022, p. 598):

n_estimators	[100, 200, 300, 400]
Learning_rate	[0.001, 0.005, 0.01, 0.05]
Max_depth	[3, 5, 8, 10, 12, 15]
Gamma	[0.001, 0.005, 0.01, 0.2]
Random_state	[42, 43, 44, 45]

حيث:

n_estimators : يمثل عدد الأشجار التي سيتم إنشاؤها في النموذج.

Learning_rate: يمثل معدل التعلم للنموذج.

Max_depth:هو أقصى عمق لكل شجرة في النموذج.

Gamma:هي المعلمة التي تتحكم في الحد الأدنى للتحسين المطلوب في تقسيم العقدة.

Random_state:تمثل هذه العملية قيمة التوليد العشوائي، بحيث يضمن إمكانية إعادة النتائج.

والقيم المثلى لهذه المعالم المتحصل عليهابعد تطبيق الخوارزمية موضحة في الجدول رقم 05 الموالي:

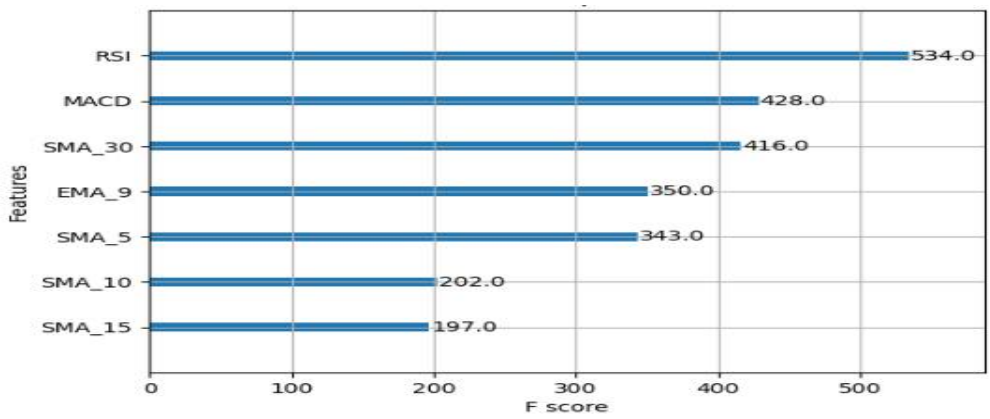
جدول 5: القيم المثلى لمعلمات النموذج

n_estimators	400
Learning_rate	0.05
Max_depth	3
Gamma	0.2
Random_state	42

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

والشكل رقم 7 يعرض لنا أهمية مزايا الإدخال في التنبؤ بسعر الإغلاق اليومي لسهم شركة Apple.

الشكل 7: أهمية مزايا الإدخال في التنبؤ بميزة الهدف



المصدر: من إعداد الباحث بالاعتماد على برنامج Python

6.3. التنبؤ والتحقق من دقة النموذج

بعد الانتهاء من عملية تقدير النموذج يجب اختبار هذا النموذج على بيانات خارج عينة التدريب للتأكد من كفاءته في التعميم بحيث يتم تقدير عينة الاختبار التي لم يسبق للنموذج التدرب عليها. والجدول الموالي يعرض نتائج تقييم النموذج المقترح باستخدام بعض معايير الأداء.

جدول 6: قيمة معايير الأداء للنموذج

المعيار	MAE	MSE	RMSE	MAPE
القيمة	**4.6537194**	**62.417282**	**7.9004607**	**0.02385254**

المصدر: من إعداد الباحث بالاعتماد على برنامج Python

حيث:

MAE: يمثل متوسط الأخطاء المطلقة (MeanAbsoluteError)

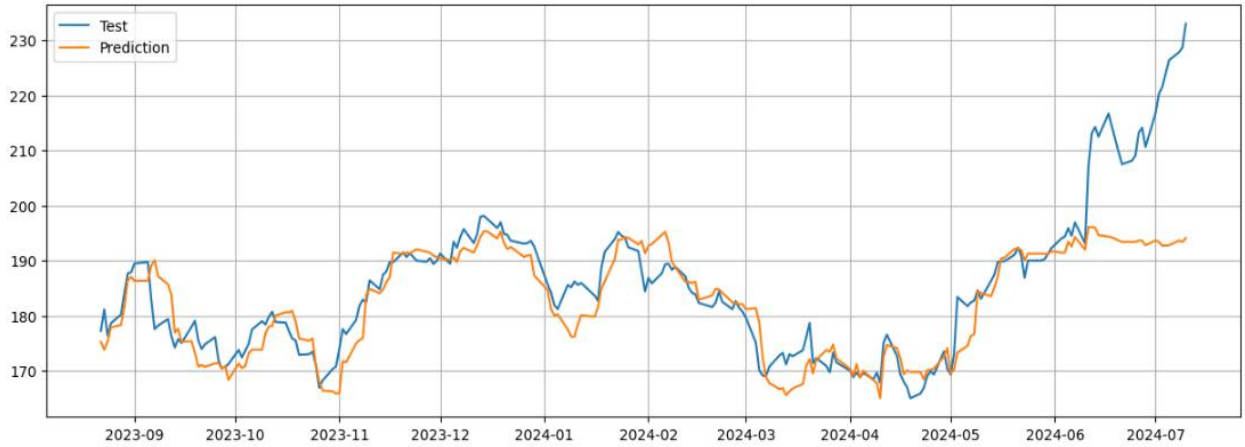
MSE:يمثل متوسط مربعات الأخطاء (MeanSquaredError)

RMSE:يمثل الجذر التربيعي لمتوسط مربعات الأخطاء (RootMeanSquare Error)

MAPE:يمثل متوسط الأخطاء المطلقة المئوية (Mean Absolute Percentage Error)

والشكل رقم 8 أدناه يعرض لنا القيم الفعلية والقيم المتنبأ بها لبيانات الاختبار باستخدام خوارزمية XGBoost

الشكل 8: القيم الفعلية والمتنبأ بها باستخدام خوارزمية XGBoost



المصدر: من إعداد الباحث بالاعتماد على برنامج Python

4. تحليل النتائج:

من خلال الجدول رقم 6 والشكل البياني رقم 8 نستنتج أن هناك قدرة تنبؤية عالية لخوارزمية XGBoost في التنبؤ بأسعار سهم شركة Apple. وهو الأمر الذي يتوافق مع نتائج بعض الدراسات السابقة كدراسة(Zhang, 2022)التي أشارت إلى تفوق خوارزمية XGBoost في التنبؤ بأسعار الأسهم وتوصلت إلى أنها تعتبر أداة علمية قوية لتحليل البيانات الضخمة والتوقعات المالية في عصر البيانات الضخمة.

إلى جانب ذلك تتوافق هذه الدراسة مع نتائج دراسة(Yang et al., 2021)التي توصلت إلى أن دمج خوارزمية XGBoost مع خوارزمية LightGBM يعطي دقة أكبر في توقع أسعار الأسهم مقارنة بالنماذج الفردية وأكدت بأنهما يعتبران أداة فعالة لتحليل الأسواق المالية وتحقيق عوائد استثمارية أفضل.

5. خاتمة:

يمكن أن يساعد التنبؤ الدقيق بإستخدام منهجية XGBoost على ترشيد القرارات لتكون أكثر حكمة، وهذا بدوره يساعد على التقليل من حالة عدم اليقين، لذلك هناك اتجاه في السنوات القليلة الماضية نحو تبني نماذج التنبؤ التي تعتمد على الذكاء الاصطناعيالتي بدورها جعلت عملية التنبؤ أكثر دقة. ويرجع ذلك إلى قدرة هذه النماذج الحديثة على الربط بين المتغيرات ذات العلاقات الخطية وغير الخطية، ويمكن استعراض توصيات هذه الدراسة فيما يلي:

- ضرورة إعطاء أهمية للنماذج التي تعتمد على الذكاء الاصطناعي عند القيام بعملية التنبؤ في السوق المالي لإتخاذ قرارات استثمارية رشيدة.
- تقييم كفاءة وفعالية استخدام الشبكات العصبية في التنبؤ بأسعار الأسهم.

6. قائمة المراجع:

1. Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320.
2. Chen, M., Liu, Q., Chen, S., Liu, Y., Zhang, C.-H., & Liu, R. (2019). XGBoost-based algorithm interpretation and application on post-fault transient stability status prediction of power system. IEEE Access, 7, 13149-13158.
3. Deng, S., Huang, X., Zhu, Y., Su, Z., Fu, Z., & Shimada, T. (2023). Stock index direction forecasting using an explainable eXtreme Gradient Boosting and investor sentiments. The North American Journal of Economics and Finance, 64, 101848.
4. Pant, A., & Ramana, G. (2022). Prediction of pullout interaction coefficient of geogrids by extreme gradient boosting model. Geotextiles and Geomembranes, 50(6), 1188-1198.
5. Pring, M. J. (2014). Technical analysis explained: The successful investor's guide to spotting investment trends and turning points. McGraw-Hill.
6. Vu, V.-H., Mashal, I., & Chung, T.-Y. (2017). A novel bandwidth estimation method based on MACD for DASH. KSII Transactions on Internet and Information Systems (TIIS), 11(3), 1441-1461.
7. Wang, P., Zhang, G., Yu, Z.-G., & Huang, G. (2021). A deep learnING And XGBoost-based method for predicting protein-protein interaction sites. Frontiers in Genetics, 12, 752732.
8. Yang, Y., Wu, Y., Wang, P., & Jiali, X. (2021). Stock price prediction based on xgboost and lightgbm. E3s web of conferences,
9. Zhang, Y. (2022). Stock Price Prediction Method Based on XGboost Algorithm. 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022).
10. بن لطرش خالد، (2019)، اختبار مدى تنبؤية وربحية التحليل الفني للأوراق المالية في الأسواق المالية العربية، أطروحة مقدمة لنيل شهادة الدكتوراه، كلية العلوم الاقتصادية والتجارية وعلوم التسيير، جامعة فرحات عباس سطيف 1، الجزائر.