

المجلة العلمية التجارة والتمويل

<https://caf.journals.ekb.eg>



استخدام طرق تعلم الآلة ونماذج بوكس وجنكينز للتنبؤ بمؤشر البورصة المصرية EGX30

أسماء مسعد السيد السقا^{a*}، نصر إبراهيم رشوان أبو زيد^a، مي محمد كامل مصطفى^a

^a قسم الإحصاء والرياضة والتأمين. كلية التجارة، جامعة طنطا، مصر

تاريخ النشر الإلكتروني: يونيو 2024

للتأصيل المرجعي السقا، أسماء مسعد السيد، أبو زيد، نصر إبراهيم رشوان، مصطفى، مي محمد كامل. استخدام طرق

تعلم الآلة ونماذج بوكس وجنكينز للتنبؤ بمؤشر البورصة المصرية EGX30،

المجلة العلمية للتجارة والتمويل، المجلد 44 (2) 793-818

المعرف الرقمي: caf.2024.371011/10.21608

استخدام طرق تعلم الآلة ونماذج بوكس وجنكينز للتنبؤ بمؤشر البورصة المصرية EGX30

أسماء مسعد السيد السقا

قسم الإحصاء والرياضة والتأمين . كلية التجارة ، جامعة طنطا ، مصر

نصر إبراهيم رشوان أبو زيد

قسم الإحصاء والرياضة والتأمين. كلية التجارة، جامعة طنطا، مصر

مي محمد كامل مصطفى

قسم الإحصاء والرياضة والتأمين. كلية التجارة، جامعة طنطا ، مصر

تاريخ المادة

تم استلامه في 27 إبريل 2024، وتم قبوله في 30 مايو 2024، وهو متاح على الإنترنت يونيو 2024

مستخلص الدراسة:

تهدف الدراسة إلى المقارنة بين نماذج بوكس وجنكينز متمثلة في نموذج الانحدار الذاتي والمتوسطات المتحركة التكاملية (ARIMA) وطرق تعلم الآلة متمثلة في خوارزم المتجه الداعم للانحدار (SVR)، وخوارزم Prophet، وخوارزم الوحدات المتكررة ذات البوابات (GRU)، وذلك للتنبؤ بمؤشر البورصة المصرية EGX30 في الفترة من 2 يناير 2013 إلى 2 يناير 2023 بإجمالي مشاهدات 2438 مشاهدة يومية. تم استخدام 1950 مشاهدة منها كمجموعة تدريبية للسلسلة بنسبة 80% من إجمالي البيانات، بينما يتم استخدام 488 مشاهدة كمجموعة اختبار بنسبة 20% من إجمالي المشاهدات. وتوصلت الدراسة إلى أن أفضل خوارزم لتمثيل السلسلة الزمنية لمؤشر البورصة المصرية EGX30 هو خوارزم المتجه الداعم للانحدار (SVR)، يليه خوارزم الوحدات المتكررة ذات البوابات (GRU) يليه نموذج ARIMA (0,1,1) ثم خوارزم Prophet وذلك لأنه يمتلك أقل قيم لمقاييس ومعايير دقة التنبؤ المتمثلة في متوسط الخطأ المطلق (MAE) والجذر التربيعي لمتوسط مربعات الخطأ (RMSE) ومتوسط الأخطاء النسبية المطلقة (MAPE).

الكلمات المفتاحية: التنبؤ، مؤشر البورصة EGX30، ARIMA، SVR، Prophet Model، GRU Model

Abstract:

The study aims to compare Box-Jenkins models represented by the Autoregressive Integrated Moving Average (ARIMA) model, with machine learning methods represented by the Support Vector Regression (SVR) algorithm, Prophet algorithm and Gated Recurrent Unit (GRU) algorithm for predicting the Egyptian stock market index EGX30. The study covers the period from 2 January 2013 to 2 January 2023 with a total of 2438 daily observations. Out of these, 1950 observations were used as a training set, which accounts for 80% of the total data, and 488 observations were used as a test set, which accounts for 20% of the total observations. The study found that the best algorithm for representing the EGX30 time series is the Support Vector Regression (SVR) algorithm, followed by the Gated Recurrent Unit (GRU) algorithm, then the ARIMA (0,1,1) model and finally the Prophet algorithm. This ranking is based on the algorithms' lower values of accuracy metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

Keywords: Forecasting, EGX30 stock market index, ARIMA, SVR, Prophet Model, GRU Model.

مقدمة Introduction

يعد التنبؤ ببيانات السلاسل الزمنية الاقتصادية والمالية وغيرها من المهام الصعبة بسبب التغيرات غير المسبوقة في الاتجاهات والظروف الاقتصادية من ناحية، والمعلومات غير الكاملة من ناحية أخرى. وقد ظهرت في السنوات الأخيرة أهمية التنبؤ بالسلاسل الزمنية الاقتصادية والمالية ولذلك فإن تقييم دقة التنبؤات أمر ضروري عند استخدام أشكال مختلفة من أساليب التنبؤ. ولقد قدمت تقنيات تعلم الآلة وخصوصاً خوارزميات التعلم العميق أساليب جديدة لمشاكل التنبؤ حيث يتم تصميم العلاقات بين المتغيرات في تسلسل هرمي عميق ومتعدد الطبقات. لقد اكتسبت التقنيات القائمة على تعلم الآلة مثل آلة المتجهات الداعمة (SVM) والغابات العشوائية (RF) والتعلم العميق والخوارزميات المعتمدة على الشبكة العصبية المتكررة (RNN) والذاكرة طويلة المدى (LSTM) الكثير من الاهتمام في السنوات الأخيرة مع تطبيقاتها في العديد من التخصصات بما في ذلك المالية. (Makridakis et al., 2018)

1- الدراسات السابقة:

يوجد العديد من الدراسات النظرية والتطبيقية التي تناولت تحليل السلاسل الزمنية والتنبؤ بها بالاعتماد على نماذج ARIMA وخوارزميات تعلم الآلة في مختلف المجالات نذكر منها التالي:

ركزت دراسة (Yenidogan et al., 2018) على التنبؤ بأسعار البيتكوين باستخدام طرق PROPHET و ARIMA باستخدام منصة R analytics. للوصول إلى نموذج التنبؤ الأكثر دقة وتمت مقارنة مقاييس أداء طرق PROPHET و ARIMA في نفس مجموعة البيانات. حيث بدأت السلسلة محل الدراسة من مايو 2016 وانتهت في مارس 2018 وهي الفترة التي تغيرت فيها قيم البيتكوين بشكل كبير مقابل العملات الأخرى. تم اختيار النموذج لكل من ARIMA و PROPHET باستخدام تقنية تقسيم ثلاثية الأبعاد مع مراعاة خصائص السلاسل الزمنية

لمجموعة البيانات. وتوصلت نتائج الدراسة إلى تفوق نموذج PROPHET على ARIMA حيث بلغت قيمة معامل التحديد (R^2) 94% بدلا من 68%.

وفي دراسة (Abonazel & Abd-Elftah, 2019) تم استخدام منهجية بوكس وجنكينز لبناء نموذج (ARIMA) المناسب لبيانات الناتج المحلي الإجمالي المصري (GDP). وتم الحصول على بيانات الناتج المحلي الإجمالي السنوية لمصر من البنك الدولي للأعوام من 1965 إلى 2016 وتوصلت الدراسة إلى أن النموذج الإحصائي المناسب للناتج المحلي الإجمالي المصري هو (1,2,1) ARIMA وتم استخدامه للتنبؤ بالناتج المحلي الإجمالي لمصر للسنوات العشر التالية على سنوات الدراسة. وركزت دراسة (Samal et al., 2019) على تحديد فائدة نماذج التحليلات لبناء نظام قادر على إعطاء تقدير تقريبي لمستويات التلوث المستقبلية خلال فترة ثقة كبيرة، حيث تم اقتراح طريقتين للتنبؤ بناءً على البيانات التاريخية في بوبانسوار في الهند والتي تحتوي على معلومات من 2005 إلى 2015 وأشار التحليل التجريبي للتنبؤ بمستويات تلوث الهواء إلى فعالية الطريقة المقترحة باستخدام نموذج SARIMA وProphet وتوقع النموذج أيضا قيمة الملوثات لعام 2016 وبمقارنة أداء كلا النموذجين تبين أن كلا من نموذج SARIMA وProphet يوفر جودة عالية من الدقة. وفي دراسة (Gupta et al., 2021) تم استخدام ثلاثة نماذج لتعلم الآلة SVM وProphet Model وLinear Regression Model للتنبؤ بمعدل النشاط ومعدل الوفيات ومعدل الشفاء في الهند من خلال تحليل بيانات COVID-19. وأظهرت النتائج أن نموذج Prophet أفضل طريقة للتنبؤ بمعدل النشاط ومعدل الوفيات والشفاء مقارنة بـ SVM والانحدار الخطي عند مجموعات البيانات الكبيرة غير المؤكدة والصغيرة. وتم العمل في دراسة (Zhang & Yang, 2021) على توفير طريقة ديناميكية للتنبؤ بالإزاحة للانهياريات الأرضية التدريجية في جزيرة بالصين والتي تستند إلى نموذج الوحدة المتكررة ذات البوابات (GRU) مع تحليل السلاسل الزمنية. تم استخدام (SVM) ونموذج GRU كمقارنات. تم إثبات أن أسلوب تنبؤ (GRU) يمكن أن يتنبأ بدقة تامة بإزاحة الانهياريات الأرضية، والتي يمكن الاستفادة منها بنظام إنذار فعال للمخاطر الجيولوجية المبكرة. علاوة على ذلك فإن طريقة (GRU) لديها دقة تنبؤ أعلى من نموذج SVM. وتناولت دراسة (Lee, 2022) نموذجي ARIMA وSVR للتنبؤ السنوي بنفايات الورق المقوى (الكرتون) المنزلية بأمريكا في الفترة الزمنية من عام 1960 إلى عام 2018 وتوصلت الدراسة إلى تفوق نموذج SVR على نموذج ARIMA في التنبؤ بالاعتماد على معايير ومقاييس التنبؤ المختلفة. واهتمت دراسة (Chen et al., 2024) بحالة (SOH) لبطاريات الليثيوم أيون حيث إنها معلمة مهمة لنظام إدارة البطارية، ويمكن أن يؤدي التقدير الدقيق لها إلى تحسين العمر الإجمالي للبطارية. ولذلك اقترحت هذه الدراسة طريقة تقدير SOH على أساس الشبكة العصبية المتكررة للوحدة المتكررة (GRU) حيث تستخرج الارتباط طويل المدى من بيانات المخرجات المرجحة، وتم استخدام البيانات التجريبية لبطاريات الليثيوم أيون من ثلاثة مصادر للتحقق من صحة وفعالية هذه الطريقة. أظهرت النتائج التجريبية أن الطريقة تظهر دقة تقدير جيدة، مع الحد الأقصى لمتوسط الخطأ المطلق أقل من 1.78% والحد الأقصى لجذر متوسط مربع الخطأ أقل من 2.3%.

2- مشكلة البحث:

تتمثل مشكلة البحث في استخدام نماذج تقليدية مثل ARIMA بالإضافة إلى خوارزميات حديثة ونماذج للتعلم الآلي مثل SVR و prophet و GRU والمقارنة بينهم بالاعتماد على معايير المقارنة الإحصائية للتوصل إلى أفضل النماذج للتنبؤ والتي تلائم بيانات الدراسة.

3- أهداف البحث

تتلخص أهداف الدراسة فيما يلي:

- التنبؤ بالسلسلة الزمنية محل الدراسة باستخدام نموذج ARIMA
- التنبؤ بالسلسلة الزمنية محل الدراسة باستخدام أساليب جديدة ومستحدثة وذلك بالاعتماد على تعلم الآلة.
- الوقوف على القدرة التنبؤية والتفسيرية لكل نموذج وخوارزم مستخدم في هذه الدراسة.
- المقارنة بين نموذج ARIMA وأساليب تعلم الآلة في التنبؤ بالسلاسل الزمنية والتوصل إلى أكثر النماذج دقة بالاعتماد على مجموعة من المقاييس والمعايير الإحصائية.
- تقديم نماذج لتحليل بيانات السلاسل الزمنية للظاهرة محل الدراسة والتنبؤ بها وتصورها والكشف عن أنماطها.

4- أهمية البحث

ركزت هذه الدراسة على طرق تعلم الآلة (ML) في التنبؤ بالسلاسل الزمنية ومقارنتها بالطرق الإحصائية التقليدية والوقوف على الدقة التنبؤية والتفسيرية لكل نموذج من هذه النماذج حيث إن أهمية التقييم الموضوعي للأداء النسبي لأساليب تعلم الآلة في التنبؤ واضحة، ولكنها لم تتحقق حتى الآن مما يثير تساؤلات حول قيمتها العملية لتحسين دقة التنبؤ ودفع مجال التنبؤ، بالإضافة إلى التوصل إلى مجموعة من النتائج والتوصيات التي يستند إليها متخذي القرارات في تخطيط وتطوير السياسات المستقبلية وذلك بالاعتماد على أكثر الأساليب دقة في التنبؤ بالظواهر المختلفة.

وهذه الدراسة تسلط الضوء أيضا على تقييم الأساليب الكلاسيكية واستخدام نتائجها كخط أساس عند تقييم أي تعلم آلي وطرق التعلم العميق للتنبؤ بالسلسلة الزمنية من أجل إثبات أن تعقيدها الإضافي يضيف مهارة إلى التنبؤ.

5- حدود البحث

تم تقسيم حدود الدراسة إلى حدود مكانية وحدود زمنية وحدود موضوعية كما يلي:

الحدود المكانية: دراسة مؤشر البورصة EGX30 في مصر

الحدود الزمانية: وهي الاعتماد على بيانات مؤشر البورصة EGX30 في الفترة من 2 يناير 2013 إلى 2 يناير 2023

الحدود الموضوعية: تتمثل الحدود الموضوعية في الاقتصار على استخدام أربعة نماذج محددة وهم (ARIMA والمقارنة بينهم بغرض الوصول إلى أفضل نموذج في التنبؤ، وذلك لمساعدة متخذي القرارات الاقتصادية الخاصة بمؤشر البورصة المصرية EGX30 في التخطيط للسياسات المالية والنقدية في مصر.

6- الأساليب والنماذج الإحصائية المستخدمة في البحث:

يعتمد البحث على استخدام أربعة نماذج محددة وهم (ARIMA Model, Support Vector Regression Model, Prophet Model, Gated Recurrent Model) والمقارنة بينهم بغرض الوصول إلى أفضل نموذج في التنبؤ يمكن إيجازها كما يلي:

6-1 نماذج ARIMA (منهجية بوكس وجنكينز): (شعراوي، 2005)

وصف الباحثان بوكس وجنكينز النموذج المختلط ARMA على نحو أشمل ووضع أسلوباً لفهم عدم الإستقرارية ومعالجتها في حال كون السلسلة الزمنية غير مستقرة عن المتوسط إذ يتم تحويلها إلى سلسلة زمنية مستقرة وذلك بأخذ الفروق اللازمة لها ويرمز لنموذج ARMA بعد أخذ الفروق للسلسلة الزمنية بالرمز ARIMA، ويتم أخذ الفروق بأعداد صحيحة موجبة من الدرجة d فإن y_t هي عملية ARIMA إذا كان:

$$\nabla^d Y_t = (1 - B)^d Y_t \quad (1)$$

ونموذج ARIMA للسلسلة الزمنية Y_t يكون بالشكل الآتي:

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)a_t \quad (2)$$

حيث إن:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3)$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (4)$$

وكلا من $\phi_p(B)$ ، $\theta_q(B)$ هي كثيرة حدود في B من الدرجة p لمعالم نموذج الانحدار الذاتي وكثيرة حدود في B من الدرجة q لمعالم نموذج المتوسطات المتحركة على الترتيب.

ويتم بناء نماذج السلاسل الزمنية طبقاً لمنهجية بوكس وجنكينز بأربع مراحل (Box and Jenkins , 1976) وهي:

1- التعرف على النموذج Identification (فاندال، والتر 1992)

تعد مرحلة التعرف من المراحل الأساسية والهامة التي تساعد على بناء نموذج السلسلة الزمنية، وتعتبر كلا من دالة الارتباط الذاتي ودالة الارتباط الذاتي الجزئي أداتان أساسيتان في التعرف على النموذج المناسب للسلسلة، ويمكن تلخيص سلوك دالتي الارتباط الذاتي والارتباط الذاتي الجزئي في النماذج غير الموسمية كما يلي:

النموذج	دالة الارتباط الذاتي	دالة الارتباط الذاتي الجزئي
AR(p)	تقترب من الصفر تدريجياً وتأخذ صورة أسية	تتقطع بعد الفجوة الزمنية p وتساوي صفر
AM(q)	تتقطع بعد الفجوة الزمنية q وتساوي صفر	تقترب من الصفر تدريجياً وتأخذ صورة أسية
ARMA (p, q)	تقترب من الصفر تدريجياً وتأخذ صورة أسية	تقترب من الصفر تدريجياً وتأخذ صورة أسية

2- مرحلة تقدير معالم النموذج Estimation Stage

بعد التعرف على معالم النموذج نستطيع تقدير معالم السلسلة الزمنية وتحديد أي منها يولد بيانات السلسلة وتعتبر طريقة المربعات الصغرى Least Square Method طريقة أساسية للتقدير، حيث يتم اختيار مقدرات المعالم التي تجعل مجموع مربعات البواقي أقل ما يمكن، وهناك صعوبتان في استخدام طريقة المربعات الصغرى وهما:

1- النموذج بصفة عامة غير خطي في المعالم

2- يتطلب النموذج الحصول على القيم المبدئية للبيانات وأيضا للأخطاء

ويمكن التغلب على هذه المشكلة بعدة طرق:

الطريقة الأولى: حساب مجموع المربعات للقيم بدءاً من القيمة الثانية أي بإهمال المشاهدة الأولى في البيانات.

الطريقة الثانية: طريقة المربعات الصغرى غير الشرطية

الطريقة الثالثة: طريقة المربعات الصغرى الشرطية

3- مرحلة فحص النموذج Diagnostic checking of model stage

يعد تحديد أفضل نموذج نقوم بفحصه وذلك من خلال أربع مجموعات من الاختبارات المستخدمة لتقويم مدى ملائمة النموذج، فإذا تبين عدم ملائمة النموذج فإنه يجب دراسة نموذج آخر لتوفيق البيانات، كما يجب اختبار معنوية معالم النموذج، ويتم ذلك من خلال اختبار t حيث يتم مقارنة قيمة $P.V$ مع مستوى المعنوية المستخدم يساوي 0.05 فإذا كانت $p.v < 0.05$ نستطيع رفض فرض العدم القائل بأن معلمة النموذج لا تختلف عن الصفر، وبالتالي تظل المعلمة في النموذج، أما في حالة قبول فرض العدم لابد من استبعاد أحد معالم النموذج MA, AR

كما يمكن دراسة ملائمة النموذج من خلال أربعة أنواع من الاختبارات وفيما يلي نستعرض الأنواع الأربعة كما يلي:

1- تحليل الاستقرار

2- تحليل البواقي

3- حذف معالم من النموذج

4- إضافة معالم إلى النموذج

4- مرحلة التنبؤ Prediction

تستند عملية التنبؤ على البيانات التاريخية وتحاول إسقاطها على الأحداث المستقبلية من خلال استخدام النماذج الرياضية، فهي عبارة عن مزيج من استخدام النماذج الرياضية بالإضافة إلى الخبرة في قراءة الأحداث المستقبلية. فالتنبؤ Forecast هو عملية توقع قيمة مستقبلية بناء على معلومات ماضية، بينما في التنبؤ Prediction فإن القيمة المتوقعة أو المقدرة ليس من الضروري أن تكون مستقبلية ويمكن أن تكون قيمة مفقودة وبالتالي فإن مصطلح التنبؤ Prediction هو أعلم وأشمل من مصطلح التنبؤ Forecast (Matsuo, 2003)

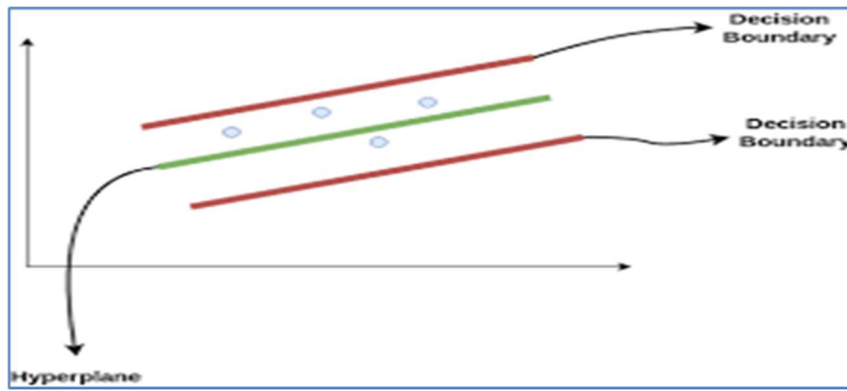
6-2 خوارزمية انحدار المتجه الداعم (SVR) (Gupta et al., 2019)

خوارزمية انحدار المتجه الداعم (SVR) هي خوارزمية تعلم آلي تُستخدم لتحليل الانحدار، والتي تتضمن التنبؤ بالقيم العددية المستمرة. وهي تندرج تحت خوارزمية آلة المتجهات الدعمة (SVM)، والتي تستخدم لمهام التصنيف والتنبؤ. والهدف في خوارزمية انحدار المتجه الداعم هو ملاءمة خط الانحدار الذي يفصل البيانات على أفضل وجه إلى فئتين، حيث تمثل إحدى الفئات المتغير المستهدف بينما تمثل الفئة الأخرى الأخطاء المتبقية. ويمثل خط الانحدار هذا المستوى الفائق (hyperplane) الذي يزيد الهامش (margin) بين المتغير الهدف والأخطاء المتبقية.

يتم تعريف الهامش على أنه المسافة بين المستوى الفائق وأقرب نقاط البيانات، والمعروفة باسم متجهات الدعم. والتعامل مع العلاقات غير الخطية بين المتغيرات المستقلة والتابعة، ويمكنها أيضاً التعامل مع القيم المتطرفة بشكل فعال. هذا يجعل خوارزمية انحدار متجه الدعم مفيدة لمهام التنبؤ المختلفة، مثل التنبؤ بأسعار الأسهم، والتنبؤ بالطقس. واختصاراً تعتبر خوارزمية انحدار متجه الدعم خوارزمية تعلم آلي قوية ومرنة، خاصةً عندما تحتوي البيانات على علاقة غير خطية أو قيم متطرفة.

6-2-1 آلية عمل خوارزمية انحدار المتجه الداعم (Parrella, 2007)

تتمثل مشكلة الانحدار في العثور على دالة تقترب من التعيين من مجال الإدخال إلى الأرقام الحقيقية على أساس عينة تدريبية، ويمكن توضيح فكرة عمل خوارزم المتجه الداعم للانحدار كما في شكل (1) كما يلي:



شكل 1: انحدار المتجه الداعم (Gupta et al., 2019)

في الشكل (1) يمثل الخطان الخارجيان حدود القرار والخط الذي في الوسط هو المستوى الفائق. والهدف من تطبيق خوارزمية SVR بشكل أساسي النظر في النقاط الموجودة داخل خط حدود القرار، وأفضل خط لدينا هو المستوى الفائق الذي يحتوي على أقصى عدد من النقاط.

وحدود القرار تقع على أي مسافة "a" من المستوي الفائق. وهي الخطوط التي نرسمها على مسافة "a+" و "a-" من المستوي الفائق يشار إلى هذه المسافة "a" بشكل أساسي باسم إبسيلون. ومن ثم سنأخذ فقط تلك النقاط التي تقع ضمن حدود القرار ولديها أقل معدل خطأ أو ضمن هامش التسامح فهذا يعطينا أفضل نموذج مناسب.

6-2-2 الصيغة العامة لمعادلات انحدار المتجهات الداعمة

Support Vector Regression Formulation (Ojemakinde, 2006)

نفترض أن بيانات التدريب هي T

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\} \quad (5)$$

بافتراض أن الدالة غير خطية وهي $f(x)$ وتعطى بالمعادلة التالية

$$f(x) = w^T \Phi(x_i) + b \quad (6)$$

حيث إن:

$$W = W_1, W_2, \dots, W_T \quad W^T: \text{تمثل قيمة الأوزان (معاملات الانحدار)}$$

b : تمثل مقدار التحيز

$\Phi(x_i)$: تمثل المساحات عالية الابعاد، والتي يتم تعيينها خطياً من مساحة الإدخال X

بافتراض أيضاً أن الهدف هو ملاءمة البيانات T من خلال إيجاد دالة $f(x)$ التي لديها أكبر انحراف ϵ عن الأهداف الفعلية y لجميع بيانات التدريب T ، وفي نفس الوقت تكون صغيرة قدر الإمكان.

ويتم تحويل المعادلة السابقة إلى مشكلة تحسين مقيدة كالتالي:

$$\begin{aligned} & \text{Minimize } \frac{1}{2} w^T w \\ & \text{Subject to: } \begin{cases} y_i - (w^T \Phi(x_i) + b) \leq \epsilon \\ y_i - (w^T \Phi(x_i) + b) > \epsilon \end{cases} \quad (7) \end{aligned}$$

حيث إن ($\epsilon \geq 0$) يتم تحديده من قبل المستخدم ويمثل الحد الأقصى للانحراف المقبول

$$\begin{aligned} & \text{Minimize } \frac{1}{2} w^T w \\ & \text{Subject to: } \begin{cases} y_i - (w^T \Phi(x_i) - b) \leq \epsilon \\ (w^T \Phi(x_i) + b) - y_i > \epsilon \end{cases} \quad (8) \end{aligned}$$

والهدف من هذه المعادلة هو جعل الدالة مسطحة قدر الإمكان، أي جعل w أصغر ما يمكن مع استيفاء القيود، وكذلك تستخدم من أجل حل المتغيرات الراكدة في المعادلة السابقة وللتعامل مع مشاكل التحسين المحتملة غير الممكنة.

3-6 خوارزمية Prophet:

خوارزمية Prophet هي خوارزمية مفتوحة المصدر متوفرة في برمجية Python و R للتنبؤ ببيانات السلسلة الزمنية. تم نشر هذه الخوارزمية من قبل فريق علم البيانات لشركة Facebook (Yenidogan et al., 2018) ويهدف إلى أن تكون هذه الخوارزمية قادرة على التنبؤ على نطاق واسع وهذا يعني أن خوارزمية Prophet يراد بها أن تكون أداة تنبؤ وفق معالجات مناسبة لطبيعة الخوارزمية.

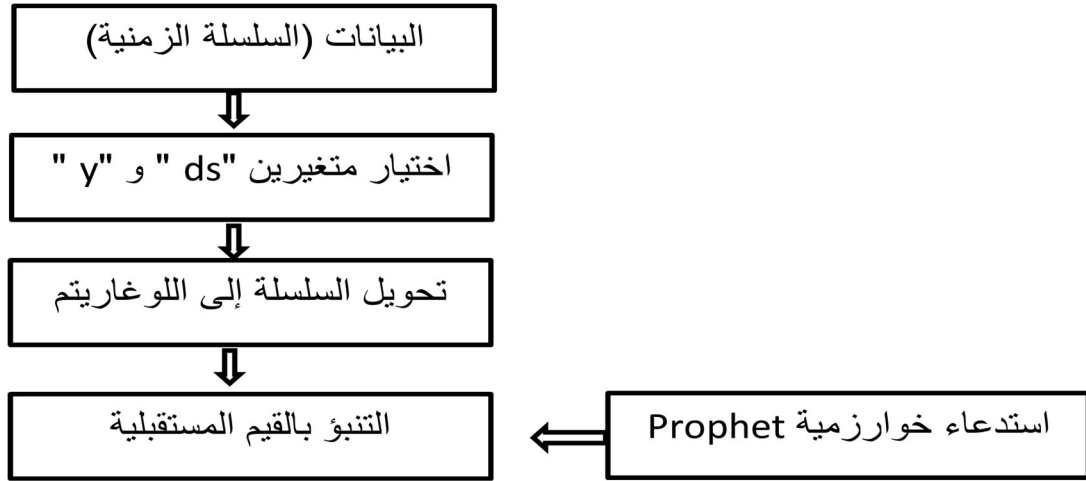
وفقاً لشركة Facebook، فإن Prophet تعمل بشكل أفضل في التنبؤ السلاسل حتى في ظل وجود تأثيرات موسمية قوية، أو العطلات، أو الأحداث، أو ذات بنية غير خطية. فهي خوارزمية تنبؤ مرنة يمكنها التعامل مع

تحديات بيانات العالم الحقيقي وأداة فعالة بشكل خاص في التعامل مع القيم المفقودة والقيم المتطرفة والأنماط المعقدة (Žunić et al., 2020) وعند الاستعانة بهذه الخوارزمية فإننا نحتاج إلى إعداد معالجة البيانات وتحويلها إلى عمودين:

العمود الأول تحت اسم "DS"، والعمود الثاني باسم "y" على أن تكون قيمها رقمية.

فالخوارزمية تمكننا من التنبؤ والحصول على عدة أعمدة، مثل "ds" و "yhat" حيث "yhat" هو عمود يحتوي على النتائج المتنبأ بها لمتغير "y" انطلاقاً من البيانات التاريخية لها. ويمكن رسم "ds" و "yhat" لإظهار ميزات مثل الاتجاه المستقبلي أو الموسمية (Aditya et al., 2021)

ويمكن إيجاز طريقة تنفيذ خوارزمية Prophet وفق الخطوات التالية كما في شكل (2) التالي:



شكل 2: خطوات تطبيق خوارزمية Prophet (Chadalavada et al., 2020)

6-3-1 آلية عمل خوارزمية Prophet (Menculini et al., 2021):

تستخدم خوارزمية Prophet ثلاث مركبات للتنبؤ بالسلسلة الزمنية، المتمثلة في مركبة الاتجاه، والموسمية، والوهمية (العطلات) وتعطى

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (9)$$

حيث إن:

$h(t)$ المركبة الوهمية، $g(t)$ مركبة الاتجاه، $s(t)$ المركبة الموسمية، ε_t الخطأ العشوائي

ومركبة الاتجاه وهي كما يلي:

$$g(t) = \left(k + \sum_{i:t>s_i} \delta_i \right) t + \left(m + \sum_{j:t>s_j} \gamma_j \right) \quad (10)$$

حيث k و m هي معلمات النموذج، δ_i هي معاملات التصحيح، و γ_j هي معاملات تجعل الدالة مستمرة.

4-6 خوارزمية الوحدات المتكررة ذات البوابات (GRU) (Shen et al.,) (2018)

هي نوع من أنواع الشبكات العصبية المتكررة (RNN) وقد تم استخدامها على نطاق واسع في مجال الترجمة الآلية، وأول من قدم شكل البنية الهيكلية الخاصة بها هو Kyunghyun Cho وزملاؤه في عام 2014. وهي تشبه إلى حد كبير الذاكرة طويلة المدى (LSTM) للشبكات، وتم تصميم وحدات GRU لمعالجة قيود RNNs القياسية في تعلم التبعيات طويلة المدى ضمن البيانات المتسلسلة. وتعمل وحدات GRU على تبسيط بنية LSTM من خلال الجمع بين بوابات الإدخال والنسيان في بوابة "تحديث" واحدة، وعن طريق دمج حالة الخلية والحالة الخفية. وينتج عن هذا نموذج أكثر إحكاما وفعالية من الناحية الحسابية فهو أسهل بكثير في الحساب والتنفيذ.

تمتلك GRU بوابتين رئيسيتين:

بوابة التحديث: ويحدد فيها مقدار المعلومات الخفية السابقة التي يجب الاحتفاظ بها ومقدار المدخلات الجديدة التي يجب استخدامها لتحديث الحالة الخفية.

بوابة إعادة الضبط: تتحكم في درجة تعديل دمج المدخلات الجديدة مع الذاكرة السابقة، مما يسمح لـ GRU بالنقاط التبعيات قصيرة المدى. تم اقتراح الوحدة المتكررة ذات البوابات (GRU) من قبل Cho et al. (2014) لحل مشكلة التبعية طويلة المدى للشبكات العصبية المتكررة. يتكون نموذج GRU من بوابتين هما؛ بوابة إعادة الضبط وبوابة التحديث. وتتحكم بوابة التحديث في المعلومات السابقة التي سيتم نقلها إلى الطبقة الحالية بينما تقرر بوابة إعادة الضبط كمية المعلومات لإخفائها. معادلة بوابة التحديث Z_t وبوابة إعادة الضبط r_t على التوالي هي:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (11)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (12)$$

ومتجه الإدخال في الوقت t يعطي x_t وعندما يتم إدخال x_t في وحدة الشبكة يكون مضروبة في الوزن W_z . و h_{t-1} تم توليدها من الطبقة المخفية السابقة وضربها في وزنها U_z . تم إضافة النتيجتين معاً ويتم استخدام الدالة السينية (σ a sigmoid function) كدالة منشطة لتوليد قيمة احتمالية بين 0, 1. يتم تعليم متجهات الوزن تخزين المعلومات ذات الصلة باستخدام بوابة إعادة التعيين r_t و σ هي (sigmoid function) و \otimes هي دالة element-wise لمتجهين. W_r, U_r هي مصفوفات الوزن التي يتم تعلمها من خلال العملية. يتم حساب ناتج الطبقة المخفية h_t في الوقت t عن طريق المعادلة التالية:

$$h_t = (1 - z_t) \otimes h_{t-1} + z_t \otimes \tilde{h}_t \quad (13)$$

حيث إن:

$$\tilde{h}_t = \tanh(W_c x_t + r_t \otimes (U h_{t-1})) \quad (14)$$

هذه المعادلات يتم حسابها تكرارياً من أول مرة إلى شريحة آخر مرة وأخيراً تولد ناتج GRU. ومقارنة بالذاكرة طويلة المدى فإن وحدة GRU أبسط في الحساب والتنفيذ وتتحسن كفاءة تدريب GRU مع تقليل عدد الأوزان ومع ذلك يمكن لـ LSTM تذكر تسلسلات أطول من GRU لأن LSTM بها خلية ذاكرة أكثر تعقيداً. (Zhang et al., 2020)

7- الدراسة التطبيقية:

الوصف الإحصائي لبيانات السلسلة الزمنية محل الدراسة:

سوف نستعرض نتائج الجانب التطبيقي للدراسة، من خلال العرض التفصيلي المسبق لخطوات التحليل الإحصائي لبيانات السلسلة الزمنية التي تتمثل في بيانات مؤشر البورصة المصرية EGX30 (وهو مؤشر سعري يقيس أداء أعلى 30 شركة من حيث السيولة والنشاط في البورصة المصرية) خلال الفترة من 2 يناير 2013 إلى 2 يناير 2023، وبالتالي فإن السلسلة الزمنية محل الدراسة تحتوي على 2438 مشاهدة. تم نمذجتها باستخدام نماذج (ARIMA) وفق منهجية بوكس وجنكينز وذلك بالاعتماد على برنامج Minitab للتحليل الإحصائي، وكذلك باستخدام نماذج تعلم الآلة (Prophet Model, SVR, GRU) بالاعتماد على اللغة البرمجية للتحليل الإحصائي Python ومكتباتها المتنوعة. حيث تم تقسيم البيانات إلى قسمين: بيانات التدريب وتمثل 80% من البيانات بما يعادل 1950 مشاهدة، بيانات الاختبار وتمثل 20% من البيانات بما يعادل 488 مشاهدة. وتم استخدام النماذج التي تم التوصل إليها لتقدير فترات زمنية مستقبلية تقدر بعدد عشر فترات زمنية بداية من 2/1/2023 إلى 19/12/2022 كما تم مقارنة القيم المتنبأ بها بالقيم الفعلية للمؤشر للتأكد من صلاحية النموذج المطبق في التحليل.

جدول 1: إحصائيات السلسلة الزمنية لمؤشر البورصة EGX30 في حدود البيانات المتاحة

فترة البيانات	من 2/1/2013 إلى 2/1/2023
حجم البيانات	2438
الوسط الحسابي للبيانات	10653.17
الانحراف المعياري للبيانات	3120.588
أكبر قيمة	18363.29
أقل قيمة	4523.32
معامل الالتواء	0.047
معامل التفرطح	-0.852

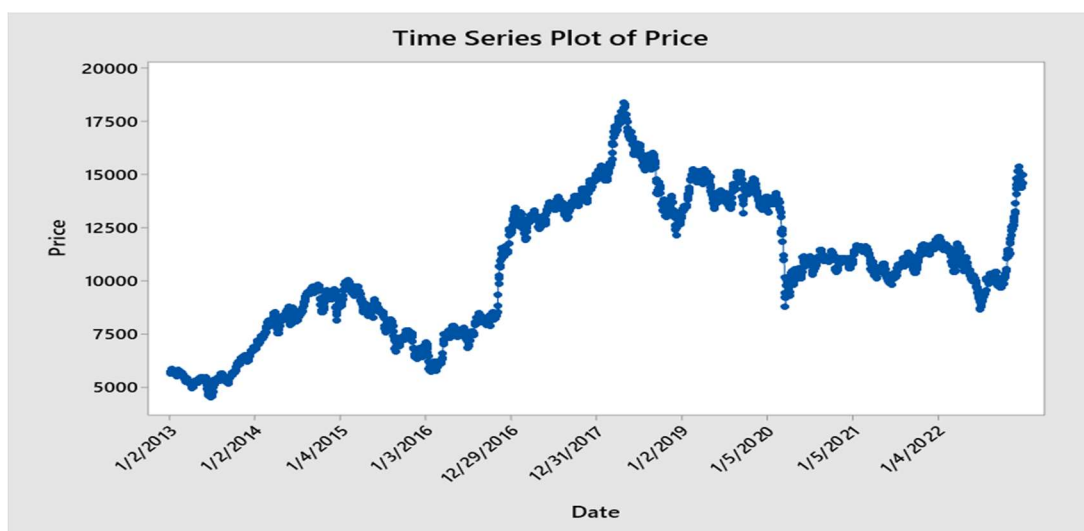
1-7 التحليل باستخدام نماذج ARIMA: يقوم أسلوب تحليل السلاسل الزمنية على استخدام نماذج ARIMA

على عدة خطوات رئيسية وهي كالتالي:

المرحلة الاولى: التعرف على النموذج

تتمثل الخطوة الأولى في التعرف على نموذج السلسلة الزمنية في التوقيع البياني لها وفقا لبياناتها التاريخية مقابل الزمن، ودراسة الشكل البياني الناتج للتعرف على الخصائص المميزة للسلسلة الزمنية من وجود اتجاه عام Trend أي وجود أي تحرك صاعد أو هابط للسلسلة الزمنية أو وجود تغيرات موسمية Seasonal Variations أو عدم ثبات التباين.

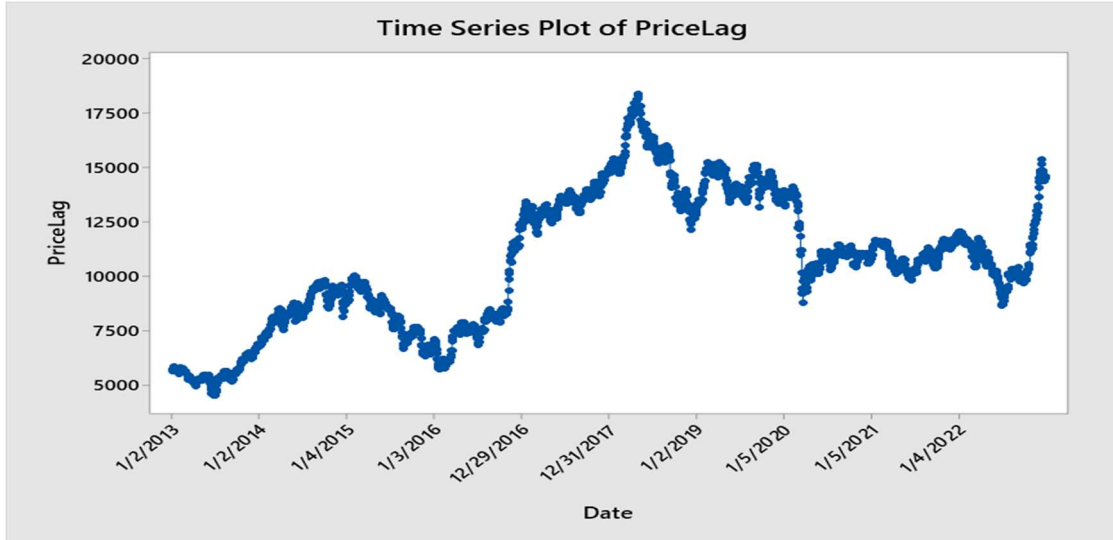
والشكل البياني التالي يوضح التوقيع البياني لبيانات مؤشر EGX30 مع الزمن، ومن خلال دراسة هذا الشكل يتضح لنا عدم ثبات تباين السلسلة الزمنية وكذلك وجود اتجاه عام متزايد عبر الزمن، كما هو موضح بالشكل (3) التالي:



شكل 3: التوقيع البياني للبيانات الاصلية

من الشكل (3) يتضح عدم ثبات كلا من التباين والوسط الحسابي للسلسلة الزمنية مما يستوجب أخذ الإجراءات المناسبة لتثبيت كلا منهما:

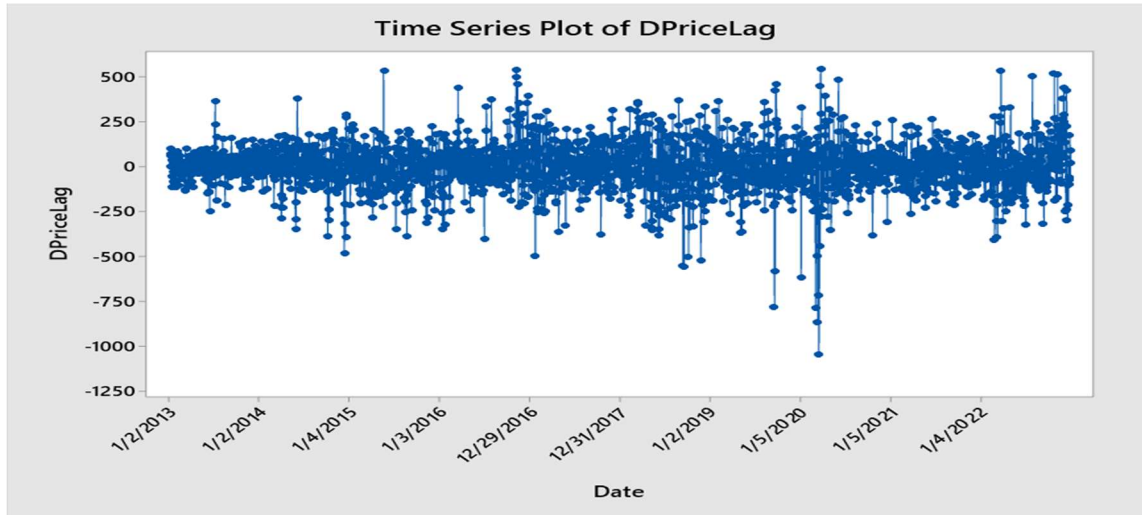
1- **تثبيت التباين:** تم تجربة العديد من التحويلات من أجل تثبيت التباين واستقر الأمر على استخدام تحويلة Log للفروق الأولى للبيانات حيث حققت هذه التحويلة استقرار السلسلة الزمنية كما هو موضح بالشكل التالي:



شكل 4: التوقيع البياني لبيانات السلسلة بعد أخذ تحويلة للفروق الأولى

يلاحظ من خلال رسم السلسلة الزمنية بعد أخذ التحويلة أنه مازال هناك عدم ثبات في الوسط الحسابي لها مما يستلزم أخذ الاجراء المناسب لتثبيته.

2- تثبيت الوسط الحسابي: بعد أخذ التحويلة لبيانات السلسلة الزمنية تم أخذ الفرق الأول للسلسلة الزمنية كما هو موضح بشكل (5) التالي:

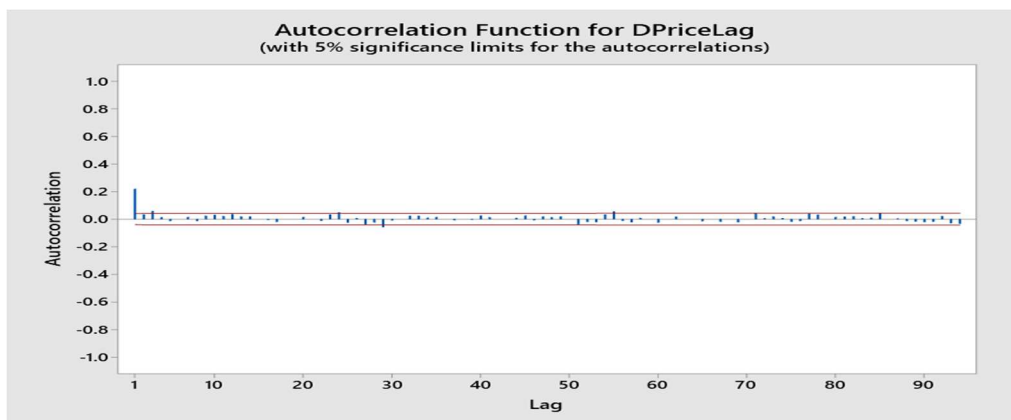


شكل 5: التوقيع البياني للفروق الأولى لبيانات السلسلة الزمنية

نلاحظ أنه بعد تجربة الفرق الثاني والثالث للسلسلة الزمنية لم نحصل على أي نتائج مختلفة عن الفرق الأول فنكتفي بالفرق الأول للسلسلة الزمنية، وبالتالي السلسلة أصبحت مستقرة بعد اخذ الفرق الأول للوغاريتم السلسلة الاصلية ويرسم دالة الارتباط الذاتي ACF ودالة الارتباط الذاتي الجزئي PACF كما يلي:

1- التوقيع البياني لدالتي الارتباط الذاتي والذاتي الجزئي:

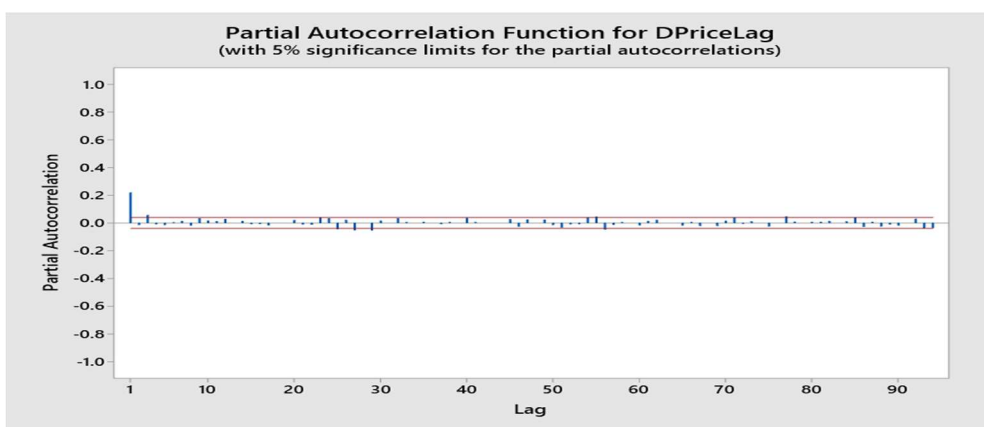
أولاً: دالة الارتباط الذاتي كما في شكل (6) التالي:



شكل 6: التوقيع البياني لمعاملات الارتباط الذاتي للفروق الأولى لقيم بيانات السلسلة

يتضح من خلال التوقيع البياني لدالة الارتباط الذاتي للسلسلة الزمنية ان الدالة تتناقص وتقترب تدريجياً من الصفر وتأخذ صورة أسية.

ثانياً: دالة الارتباط الذاتي الجزئي كما في شكل (7) التالي:



شكل 7: التوقيع البياني لمعاملات الارتباط الذاتي الجزئي للفروق الأولى لقيم بيانات السلسلة

ومن خلال التوقيع البياني لدالة الارتباط الذاتي والذاتي الجزئي نلاحظ أن دالة الارتباط الذاتي تنقطع عند الفجوة الزمنية الأولى مما يدل على أن $q=1$ بينما تتلاشى دالة الارتباط الذاتي الجزئي تدريجياً مما يدل على أن $p=0$ وبالتالي فإن بيانات السلسلة الزمنية من الممكن أن يتم تمثيلها من خلال نموذج $ARIMA(0, 1, 1)$ في صورة المعادلة التالية:

$$y_t = a_t - \theta_1 e_{t-1}$$

المرحلة الثانية: تقدير النموذج

بعد التعرف على النموذج نستطيع تقدير معالمه وتعتبر طريقة المربعات الصغرى من أكثر الطرق التي يعتمد عليها في تقدير معالم النموذج كما في جدول (2) التالي:

جدول 2: تقديرات النموذج المقترح

Model	Estimate	St. Error	t-ratio	p.v
ARIMA (0,1,1)	$a_t = 2.69$	3.73	0.72	0.471
	$\theta_1 = -0.2361$	0.0220	-10.72	0.00

ومن الجدول (2) نجد ان معادلة النموذج المقدر تكون كما يلي:

$$y_t = 2.69 + 0.2361 e_{t-1}$$

نلاحظ من النتائج السابقة ومن خلال قيمة P-Value معنوية قيمة معامل المتوسطات المتحركة والذي بلغت قيمته $\theta_1 = -0.2361$ ، كما يتضح أن التغيرات التي تحدث في النموذج ما هي إلا تغيرات عشوائية بحتة ويتضح من خلال نتائج الجدول (3) التالي:

جدول 3: تغيرات النموذج المقترح

Lag	12	24	36	48
Chi-Square	13.03	23.97	46.59	59.08
DF	10	22	34	46
P-Value	0.222	0.349	0.074	0.090

وبالنظر إلى نتائج إحصاء Chi-Square لاختبار ان بواقي النموذج المقدر ما هي الا تغيرات عشوائية بحتة، حيث نجد أن قيمة P.V أكبر من مستوى المعنوية المستخدم عند 5% مما يدل على أن بواقي النموذج المقدر ما هي إلا تغيرات عشوائية بحتة وذلك بفحص أول 48 معامل ارتباط ذاتي للبواقي.

المرحلة الثالثة: فحص النموذج

بعد التعرف على النموذج الملائم لبيانات السلسلة الزمنية وذلك من خلال الفحوص التشخيصية وتم تقدير معالمه نقوم بوصف ملائمة النموذج للبيانات، ولتقويم مدى ملائمة النموذج المقترح ARIMA (0,1,1) نتبع الفحوص التشخيصية الآتية:

فحص الاستقرار والانعكاس:

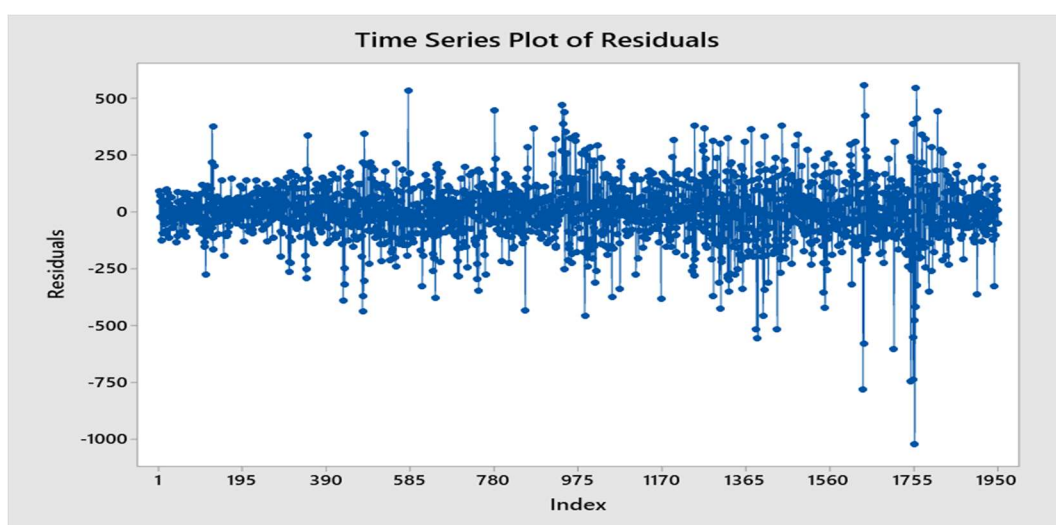
من خلال الجدول (3) يتضح أن معلمة النموذج تحقق شرط الاستقرار حيث أن جذور المعادلة المميزة $(B) = \Phi_0$ تقع كلها خارج دائرة الوحدة، فإذا كانت القيمة المطلقة لكل جذر من جذور المعادلة أكبر من الواحد الصحيح

فهذا يدل على استقرار العملية العشوائية، كما أن شرط الانعكاس تحقق حيث أن قيمة تقدير معلمة نموذج المتوسطات المتحركة المطلقة أقل من الواحد الصحيح.

تحليل البواقي:

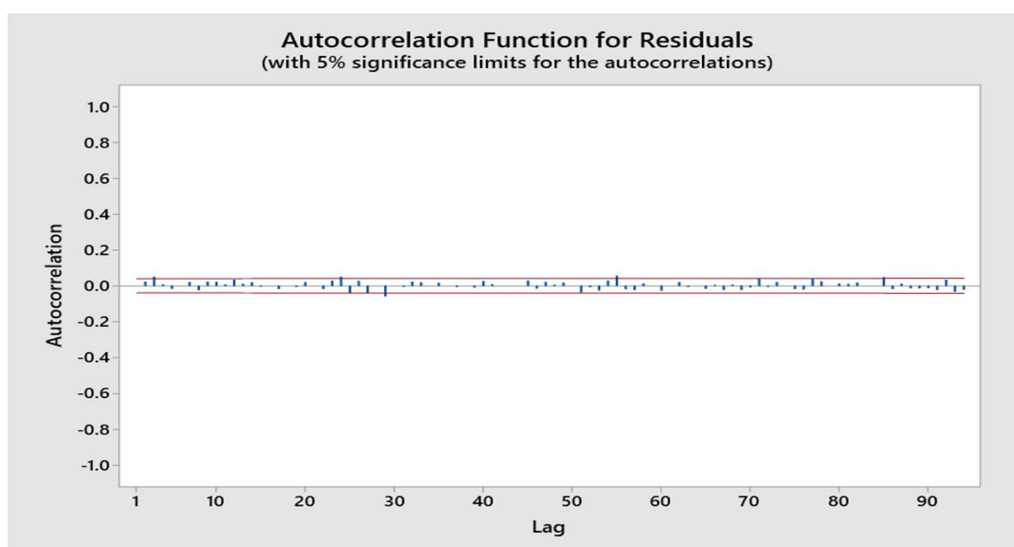
حيث تم اختبار بواقي النموذج المقدر وذلك برسم دالة البواقي للسلسلة الزمنية وكذلك رسم دالتي الارتباط الذاتي والذاتي الجزئي لبواقي السلسلة الزمنية للتأكد من أنها تغيرات عشوائية بحتة وذلك كما هو موضح في الأشكال (8)، (9)، (10) التالية:

1- التوقيع البياني لبواقي النموذج المقدر



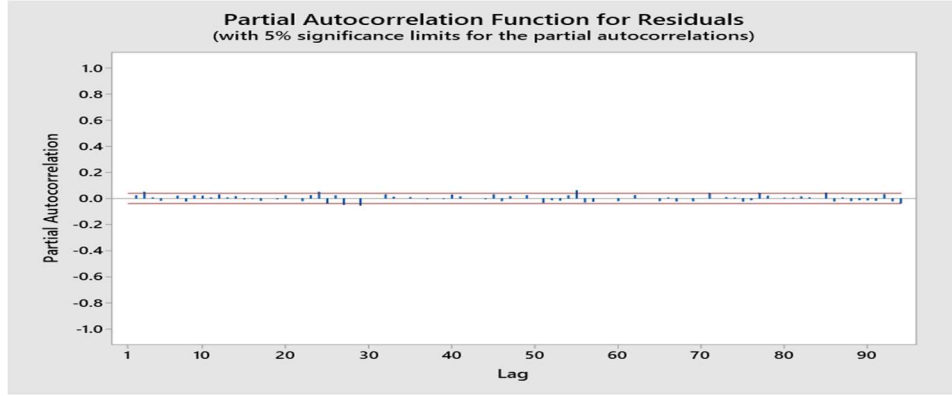
شكل 8: التوقيع البياني لبواقي النموذج المقدر

2- التوقيع البياني لدالة الارتباط الذاتي لبواقي النموذج المقدر:



شكل 9: التوقيع البياني لمعاملات دالة الارتباط الذاتي لبواقي النموذج المقدر

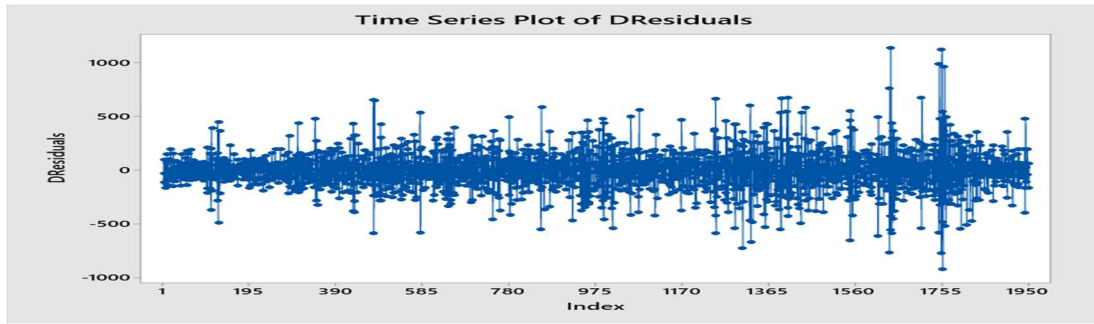
3- التوقع البياني لدالة الارتباط الذاتي الجزئي لبواقي النموذج المقدر



شكل 10: التوقع البياني لمعاملات الارتباط الذاتي الجزئي لبواقي النموذج المقترح

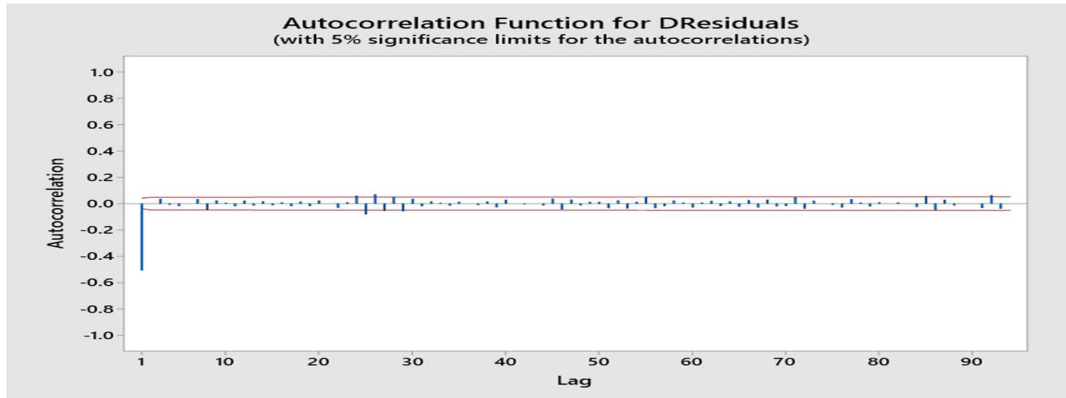
ومن خلال التوقع البياني لدالتي الارتباط الذاتي والارتباط الذاتي الجزئي للبواقي يتضح أن بواقي النموذج المقدر ما هي إلا تغيرات عشوائية بحتة.

1- فروق البواقي:



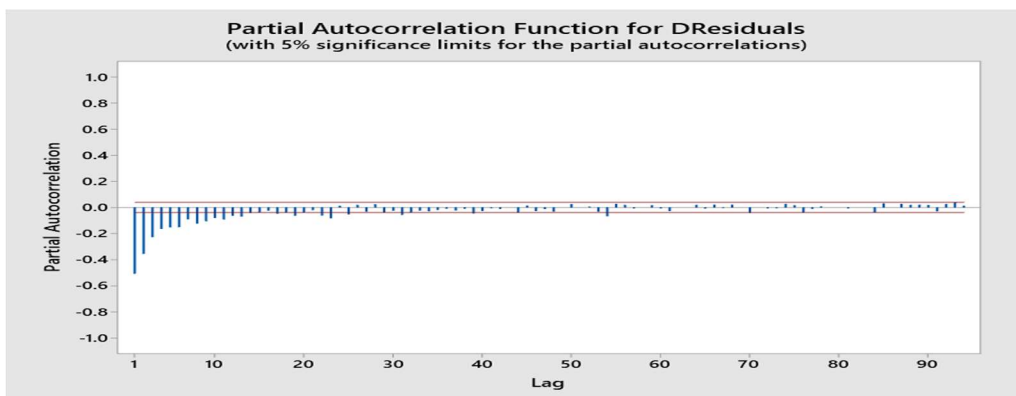
شكل 11: التوقع البياني لفروق البواقي

2- التوقع البياني لدالة الارتباط الذاتي لفروق البواقي:



شكل 12: التوقع البياني لدالة الارتباط الذاتي لفروق البواقي

3- التوقع البياني لدالة الارتباط الذاتي الجزئي لفروق البواقي:



شكل 13: التوقع البياني لدالة الارتباط الذاتي الجزئي لفروق البواقي

ومن خلال التوقع البياني لدالتي الارتباط الذاتي والذاتي الجزئي لفروق البواقي نجد أن:

1- معاملات الارتباط الذاتي قريبا من -0.5

2- معامل الارتباط الذاتي الجزئي يتناقص اسيا ويقترب من الصفر

ومما سبق يتضح صلاحية النموذج المقدر $ARIMA(0,1,1)$ وايضا لتحقق شروط الاستقرار والانعكاس.

المرحلة الرابعة: التنبؤ

بعد تقدير النموذج وفحصه للتأكد من ملائمة لسلسلة بيانات مؤشر البورصة EGX30 ، فقد تم استخدام النموذج للتنبؤ بالقيم المستقبلية للسلسلة ويكون شكل النموذج هو

$$y_t = 2.69 + 0.2361 e_{t-1}$$

2-7 خوارزمية انحدار المتجه الداعم Support Vector Regression

خوارزمية انحدار المتجه الداعم (SVR) هي خوارزمية تعلم آلي تُستخدم لتحليل الانحدار، والتي تتضمن التنبؤ بالقيم العددية المستمرة. وهي تندرج تحت خوارزمية آلة المتجهات الدعمة (SVM)، والتي تستخدم لمهام التصنيف والتنبؤ. والهدف في خوارزمية انحدار المتجه الداعم هو ملاءمة خط الانحدار الذي يفصل البيانات على أفضل وجه إلى فئتين، حيث تمثل إحدى الفئات المتغير المستهدف بينما تمثل الفئة الأخرى الأخطاء المتبقية. ويمثل خط الانحدار هذا المستوى الفائق الذي يزيد الهامش بين المتغير الهدف والأخطاء المتبقية. يتم تعريف الهامش على أنه المسافة بين المستوى الفائق وأقرب نقاط البيانات، والمعروفة باسم متجهات الدعم. والتعامل مع العلاقات غير الخطية بين المتغيرات المستقلة والتابعة، ويمكنها أيضاً التعامل مع القيم المتطرفة بشكل فعال. هذا يجعل خوارزمية انحدار متجه الدعم مفيدة لمهام التنبؤ المختلفة، مثل التنبؤ بأسعار الأسهم، والتنبؤ بالطقس. واختصارا تعتبر

خوارزمية انحدار متجه الدعم خوارزمية تعلم آلي قوية ومرنة، خاصةً عندما تحتوي البيانات على علاقة غير خطية أو قيم متطرفة (Gupta et al., 2019)

وتعتمد دقة التقدير لنموذج انحدار متجه الدعم على ضبط معلماته، وكانت البيانات محل الدراسة ذات مدى كبير جداً لذلك تم استخدام معيار قياسي لتقليل وحصر مدى البيانات بين القيمتين (0:5) وذلك بغرض تعلم الخوارزمية بشكل أسرع.

$$\text{stander scaler} = \frac{x-\mu}{s}$$

حيث إن:

X هي قيمة كل مشاهدة في مجموعة بيانات التدريب

μ : هي قيمة المتوسط الحسابي لبيانات التدريب

S : هي الانحراف المعياري لبيانات التدريب

3-7 خوارزمية Prophet

لبناء نموذج يتنبأ بسعر السهم لمؤشر EGX30 باستخدام تحليل السلسلة الزمنية للمؤشر عن طريق خوارزمية Prophet

هناك عدة خطوات لابد من اتباعها:

- 1- تثبيت مكتبة Pandas في برنامج Python لمعالجة البيانات وبنية البيانات
- 2- تثبيت خوارزمية Prophet في برنامج Python للبدأ في عملية التنبؤ
- 3- عمل استدعاء لكل المكتبات التي تم تثبيتها لاستخدامها في عملية التنبؤ
- 4- تحميل ملفات البيانات المراد استخدامها ولا بد ان تكون بصيغة CSV حتى يتعرف عليها برنامج Python ويبدأ في قراءتها

ونلاحظ أن خوارزمية Prophet تتنبأ بالبيانات فقط عندما تكون بتنسيق معين. ويجب أن يحتوي إطار البيانات DataFrame الذي يحتوي على البيانات على عمود محفوظ ك ds لبيانات السلسلة الزمنية و y للبيانات المراد التنبؤ بها. وبيانات السلسلة الزمنية هنا هي عمود "التاريخ اليومي" والبيانات التي سيتم التنبؤ بها هي عمود "سعر السهم". لذلك لابد من إنشاء DataFrame جديدًا بأسماء الأعمدة الجديدة ونفس البيانات. كما يجب أن يكون ds بتنسيق DateTime.

5- باستخدام الأمر Prophet () يمكننا تهيئة نموذج prophet للتدريب على مجموعة البيانات الخاصة بنا ومن ثم مساعدتنا في تنفيذ التنبؤ بالسلسلة الزمنية.

الجدول ds كما ذكرنا هو بيانات السلسلة الزمنية، yhat هو التنبؤ بالسلسلة الزمنية، و yhat_lower، و yhat_upper هما مستويات عدم اليقين (وهذا يعني في الأساس أن التنبؤ والقيم الفعلية يمكن أن تختلف ضمن حدود مستويات عدم اليقين). بعد ذلك لدينا اتجاه يُظهر النمو على المدى الطويل أو تقلص أو ركود البيانات و Trend_lower، و Trend_upper هي مستويات عدم اليقين.

وبالتالي تم استخدام ثلاثة أعمدة بشكل أساسي طبقاً لدالة النموذج وهم:

الأول: ويمثل التاريخ وتم تسميته ds، الثاني: يمثل السعر الافتتاحي للمؤشر Open، الثالث: يمثل سعر المؤشر وهو ما نريد أن نتنبأ به Price

4-7 خوارزمية (GRU) Gated recurrent unit

تم استخدام برنامج Python للتنبؤ وتحليل السلسلة الزمنية محل الدراسة لمؤشر البورصة EGX30 باستخدام خوارزمية الوحدات المتكررة ذات البوابات Gated Recurrent Unit والتي تمتلك بوابتين رئيسيتين:

بوابة التحديث: ويحدد فيها مقدار المعلومات المخفية السابقة التي يجب الاحتفاظ بها ومقدار المدخلات الجديدة التي يجب استخدامها لتحديث الحالة المخفية.

بوابة إعادة الضبط: تتحكم في درجة تعديل دمج المدخلات الجديدة مع الذاكرة السابقة، مما يسمح لـ GRU بالنقاط المتبعيات قصيرة المدى.

وتتحكم بوابة إعادة الضبط وبوابة التحديث بشكل تكميلي في مقدار ما تتذكره أو تنسى كل وحدة مخفية أثناء قراءة / إنشاء تسلسل. ويسمح هذا للحالة المخفية بتجاهل أي بيانات يتبين أنها غير ذات صلة في المستقبل، وتسمح هذه النتيجة بتمثيل أكثر إحكاماً بينما تتحكم بوابة التحديث في مقدار البيانات من الحالة المخفية السابقة التي سيتم نقلها إلى الحالة المخفية الحالية. تعمل هذه العملية بطريقة مشابهة لخلية الذاكرة في شبكة الذاكرة طويلة المدى وتساعد RNN على تذكر المعلومات طويلة المدى.

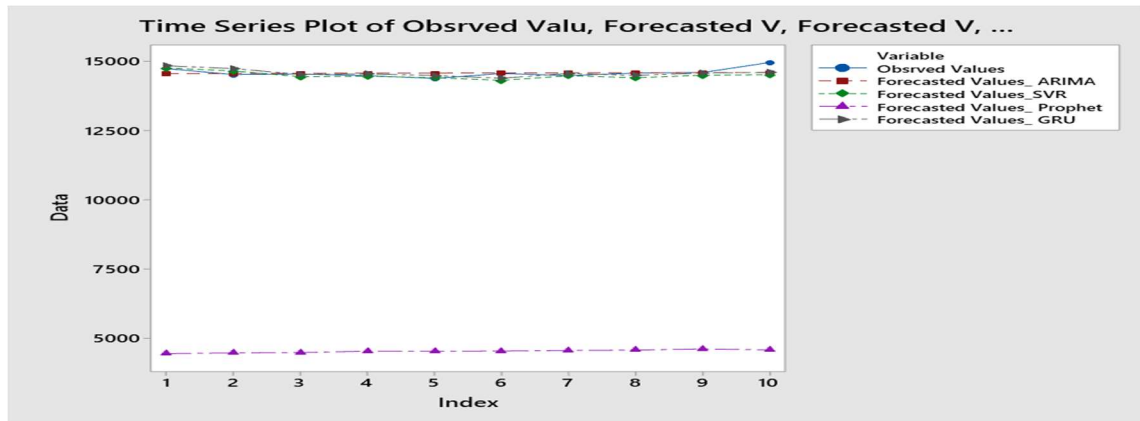
8- الحكم على دقة النماذج المختلفة والمقارنة بينهم:

للمقارنة بين النماذج الأربعة المختلفة تم المقارنة بين القيم الفعلية والقيم المتنبأ بها كما في جدول (4)

جدول 4: مقارنة النتائج الفعلية مع القيم المتنبأ بها للنماذج الأربعة

Obsrved Values	Forecasted Values_ ARIMA	Forecasted Values_SVR	Forecasted Values_ Prophet	Forecasted Values_ GRU
14738.6	14562.1	14757.0	4429.6	14852.0
14523.3	14566.0	14654.8	4454.0	14749.5
14542.0	14569.8	14446.4	4468.2	14538.6
14482.9	14573.7	14464.4	4515.1	14557.0
14385.3	14577.6	14407.7	4512.1	14499.2
14560.4	14581.4	14314.1	4521.7	14400.8
14485.2	14585.3	14482.2	4544.9	14575.0
14578.2	14589.2	14409.8	4555.5	14501.4
14598.5	14593.0	14499.4	4596.5	14592.4
14960.7	14596.9	14519.0	4562.8	14612.3

ويتضح من الجدول (4) التقارب الكبير بين النتائج المتنبأ بها مع القيم الفعلية مع كلا من نماذج (ARIMA,GRU,SVR)، بينما يوجد اختلاف بقدر كبير بين القيم المتنبأ بها والقيم الفعلية لنموذج Prophet مما يؤكد أفضلية النماذج (ARIMA, GRU, SVR) للتنبؤ بالسلسلة الزمنية وذلك في حدود البيانات المتاحة. والشكل التالي يوضح العلاقة بين القيم الفعلية مع القيم المتنبأ بها للنماذج الأربعة محل الدراسة كما في شكل (14) كما يلي:



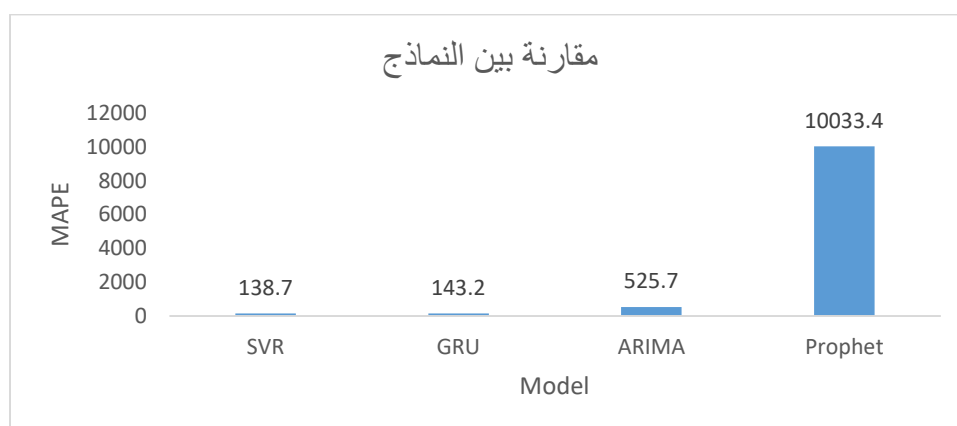
شكل 14: توزيع القيم المتنبأ بها مع القيم الفعلية

ويتضح من الشكل (14) تطابق النتائج المتنبأ بها مع القيم الفعلية مع كلا من نماذج (ARIMA,GRU,SVR) ، بينما يوجد اختلاف بقدر كبير بين القيم المتنبأ بها والقيم الفعلية لنموذج Prophet مما يؤكد أفضلية النماذج (ARIMA, GRU, SVR) للتنبؤ بالسلسلة الزمنية وذلك في حدود البيانات المتاحة.

ولاختيار النموذج الأنسب لتمثيل سلسلة بيانات EGX30 من بين الأربعة نماذج محل الدراسة تم الاعتماد على المعايير الإحصائية التالية (RMSE, MAE, MAPE) للمقارنة بين النماذج كما في الجدول التالي:

جدول 5: مقارنة بين النماذج الأربعة طبقاً للمعايير الإحصائية

النماذج المقترحة	RMSE	MAE	MAPE
SVR	162.6	138.7	0.0093
GRU	162.5	143.2	0.0096
ARIMA	535.2	525.7	0.0358
Prophet	10034	10033.4	0.68



شكل 15: المقارنة بين النماذج الأربعة طبقاً للمعيار الإحصائي (MAPE)

من خلال جدول (5) والشكل (15) نلاحظ أنه طبقاً للمقياس الإحصائية المستخدمة في المقارنة بين النماذج الأربعة فإن أفضل نموذج والذي يناسب البيانات محل الدراسة لأنه يمثل قيم أقل لمعايير دقة التنبؤ هو SVR يليه GRU يليه ARIMA والفروق بينهما بسيطة جداً لا تكاد تذكر، أما النموذج Prophet فهو غير ملائم تماماً لتمثيل سلسلة البيانات محل الدراسة.

9- الاستنتاجات

هدفت هذه الدراسة إلى المقارنة بين طرق وخوارزميات تعلم الآلة متمثلة في SVR و Prophet و GRU ومنهجية بوكس وجنكينز متمثلة في نموذج ARIMA وذلك للتنبؤ بمؤشر البورصة المصرية EGX30 من الفترة 2 يناير 2013 حتى 2 يناير 2023 حيث تم تقسيم السلسلة الزمنية إلى مجموعة تدريب تمثل 80% من إجمالي البيانات بواقع 1950 مشاهدة ومجموعة اختبار تمثل 20% من إجمالي البيانات بواقع 488 مشاهدة. وتوصلت الدراسة إلى أن نموذج SVR هو الأفضل لتمثيل بيانات السلسلة الزمنية محل الدراسة وأن نموذج GRU و ARIMA (0,1,1) أظهر أداءً جيداً في التنبؤ بمؤشر البورصة المصرية EGX30 بينما أظهر نموذج Prophet أداءً ضعيفاً في التنبؤ وذلك طبقاً للمعايير الإحصائية لقياس دقة التنبؤ.

المراجع:

أولاً: المراجع العربية

شعراوي، سمير مصطفى. (2005) "مقدمة في التحليل الحديث للسلاسل الزمنية" مركز النشر العلمي - جامعة الملك عبد العزيز.

فاندال- والتر (1992) "السلاسل الزمنية من الوجهة التطبيقية ونماذج بوكس وجنكينز". تعريب ومراجعة د. عبد المرضي حامد عزام، د. أحمد حسين هارون - دار المريخ للنشر.

ثانياً: المراجع الأجنبية

Abonazel, M. R., & Abd-Elftah, A. I. (2019). Forecasting Egyptian GDP using ARIMA models. *Reports on Economics and Finance*, 5(1), 35–47.

Aditya, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 179, 524–532.

Box, G.E.P. and Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*. 2nd Edition, Holden-Day, S. Francisco.

Chadalavada, R. J., Raghavendra, S., & Rekha, V. (2020). Electricity requirement prediction using time series and Facebooks PROPHET. *Indian Journal of Science and Technology*, 13(47), 4631–4645.

Chen, Z., Xu, Z., Wang, H., Shi, N., & Yang, L. (2024). State-of-health estimation of lithium-ion battery based on convolutional-gated recurrent neural network with self-attention mechanism. *International Journal of Green Energy*, 1–14.

Gupta, D., Pratama, M., Ma, Z., Li, J., & Prasad, M. (2019). Financial time series forecasting using twin support vector regression. *PLoS ONE*, 14(3), 1–27.

Gupta, A. K., Singh, V., Mathur, P., & Travieso-Gonzalez, C. M. (2021). Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *Journal of Interdisciplinary Mathematics*, 24(1), 89–108.

Lee, P.H. (2022) "ARIMA and SVR forecasting in the US paper waste", *Proc. SPIE 12259, 2nd International Conference on Applied Mathematics, Modelling, and Intelligent Computing*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3).

- Matsuo, Y. (2003). Prediction, Forecasting, and Chance Discovery. In Y. Ohsawa & P. McBurney (Eds.), *Chance Discovery* (pp. 30–43). Springer Berlin Heidelberg.
- Ojemakinde, B. T. (2006). Support Vector Regression for Non-Stationary Time Series. [Master's Thesis, University of Tennessee, Knoxville, TN, USA.](#)
- Parrella, F. (2007). Online support vector regression. Available at: <http://onlinesvr.altervista.org/>, accessed: 16/05/2013.
- Samal, K. K. R., Babu, K. S., Das, S. K., & Acharaya, A. (2019). Time series based air pollution forecasting using SARIMA and prophet model. *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, 80–85.
- Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia Computer Science*, 131, 895–903.
- Yenidogan, I., Cayir, A., Kozan, O., Dag, T., & Arslan, C. (2018). Bitcoin Forecasting Using ARIMA and PROPHET. *UBMK 2018 - 3rd International Conference on Computer Science and Engineering*, 621–624.
- Zhang, S., Abdel-Aty, M., Wu, Y., & Zheng, O. (2020). Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU). *Accident Analysis and Prevention*, 148(December), 105844.
- Zhang, Y., & Yang, L. (2021). A novel dynamic predictive method of water inrush from coal floor based on gated recurrent unit model. *Natural Hazards*, 105(2), 2027–2043. _
- Žunić, E., Korjenić, K., Hodžić, K., & Đonko, D. (2020). Application of Facebook's Prophet Algorithm for Successful Sales Forecasting Based on Real-world Data. *International Journal of Computer Science and Information Technology*, 12(2), 23–36.