بناء خوارزمية للنمذجة الاحتمالية لأنماط الحركة في السوق المالي باستخدام التنقيب في البيانات وتحديد الحجم الأمثل لها

(دراسة مقارنة مع النماذج القياسية ARIMA-(G)ARCH)

اد. أيهم الحميد*

(الإيداع: 14 نيسان 2025، القبول: 26 آيار 2025)

الملخص:

في عالم السوق المالي المتقلب، يُقدم هذا البحث خارطة طريق لتحسين دقة التنبؤ باتجاهات السوق المالي باستخدام خوارزمية مقترحة تعتمد على التوزيع الاحتمالي الشرطي لأنماط الحركة المتتابعة والتنقيب في البيانات لتحديد الحجم الأمثل لبيانات النمذجة الاحتمالية، حيث يُظهر البحث كفاءة تطبيق هذه الخوارزمية وإمكانية تعميمها من خلال تطبيقها على عدة أسواق مالية عالمية وعربية (14 مؤشر مالي من مؤشرات أسواق المال العالمية والعربية).

تتمثل النتائج الرئيسية لهذا البحث في تحديد مدى تفوق الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات على كل من الخوارزمية بدون تحديد الحجم الأمثل للبيانات والنماذج القياسية ARIMA (G)ARCH وذلك من خلال تطبيق معيار القيمة المضافة للنموذج التنبؤي والذي يعطي تقييم جوهري لمدى الاستفادة من أي نموذج تنبؤي مطبق، إضافة إلى تحليل الفروق ذات الدلالة المعنوية بين متوسط دقة التنبؤ باتجاه الحركة باستخدام الخوارزمية المقترحة والنماذج القياسية من جهة وكذلك بين الخوارزمية المقترحة مع وبدون تحديد الحجم الأمثل للبيانات من جهة أخرى. وقد خلصت جميع هذه المقارنات والمعايير المطبقة إلى إثبات تفوق الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات ووجود استقرار وثبات في نتائجها وإمكانية تعميمها، هذه الابتكارات تفتح أبوابًا جديدة لتحسين استراتيجيات الاستثمار وتقليل المخاطر، مما يجعلها أداة قيمة للمستثمرين والمحللين الماليين.

الكلمات المفتاحية: النمذجة الاحتمالية، التنقيب في البيانات، التنبؤ باتجاه الحركة في السوق المالي، الحجم الأمثل لبيانات التدريب، القيمة المضافة للنموذج التنبؤي.

* عضو هيئة تدريسية (مدرس) في قسم الاقتصاد في جامعة حماة – اختصاص علوم مالية ومصرفية.

Developing a Probabilistic Modeling Algorithm for Movement Patterns in Financial Markets Using Data Mining and Optimal Data Size Determination

(A Comparative Study with ARIMA-(G)ARCH econometric models)

Dr. Ayham Al- Hamed*

(Received: 14 April 2025. Accepted: 26 April 2025)

Abstract:

In the volatile realm of financial markets, this research outlines a roadmap to enhance the accuracy of financial market trend predictions by proposing an algorithm based on the conditional probability distribution of sequential movement patterns and data mining techniques to determine the optimal size of probabilistic modeling data. The study demonstrates the algorithm's efficiency and generalizability through its application to multiple global and Arab financial markets (14 financial indices from global and Arab stock markets).

The key findings of this research highlight the superiority of the proposed algorithm—which incorporates optimal data size determination—over both the algorithm without optimal data size determination and ARIMA-(G)ARCH econometric models. This conclusion is supported by the application of the *Predictive Model Value-Added criterion*, which provides an intrinsic evaluation of the utility of any applied predictive model. Additionally, a statistically significant difference analysis was conducted between: (1) the average prediction accuracy of the proposed algorithm and econometric models, and (2) the proposed algorithm with and without optimal data size determination. All applied comparisons and criteria conclusively demonstrated the superiority of the proposed algorithm with optimal data size determination, confirming its stability, consistency, and generalizability. These innovations open new avenues for refining investment strategies and mitigating risks, establishing the algorithm as a valuable tool for investors and financial analysts.

Keywords: Probabilistic Modeling, Data Mining, Financial Market Trend Prediction, Optimal Training Data Size, Predictive Model Value-Added criterion.

* Faculty member (lecturer) in the Department of Economics at Hama University – specializing in Financial and Banking Sciences.

1.المقدمة:

تعتبر الأسواق المالية من الأنظمة الديناميكية المعقدة التي تتطلب أدوات تحليل متقدمة للتنبؤ باتجاهاتها المستقبلية. في ظل التغيرات السريعة والمفاجئة في الأسواق، يصبح التنبؤ باتجاه حركة السوق أمرًا حيويًا للمستثمرين ومدراء صناديق الاستثمار. يعتمد الكثير من المستثمرين على نماذج رياضية وإحصائية لتحليل البيانات التاريخية والتنبؤ بالاتجاهات المستقبلية. في هذا البحث، نركز على بناء خوارزمية تعتمد على تقنيات التنقيب في البيانات لتحسين القدرة على التنبؤ باتجاه حركة مؤشر السوق المالي. سيتم استخدام التوزيع الاحتمالي لأنماط الحركة المتتابعة لمؤشر السوق المالي وتحديد حجم بيانات التدريب المثالي، مما يسهم في تعزيز دقة التنبؤ. كما سيتم إجراء مقارنة شاملة مع نماذج ARIMA وARCH، التي تعتبر من النماذج التقليدية المستخدمة في تحليل السلاسل الزمنية المالية.

2. مشكلة البحث:

تتمثل المشكلة الأساسية في كيفية تحسين دقة التنبؤ باتجاه حركة السوق من خلال استخدام خوارزمية فعالة للتنقيب في البيانات. يعتمد العديد من الباحثين والممارسين على نماذج ذات حجم بيانات تدريب غير قد تكون غير مثالية، مما يؤدي إلى نتائج غير دقيقة. يواجه المستثمرون تحديات في تحديد النقاط المثلى لبيانات التدريب، مما يؤثر سلبًا على دقة التنبؤ. لذا، يسعى هذا البحث إلى دراسة هذه المشاكل من خلال تطوير خوارزمية تستند إلى التوزيع الاحتمالي للتنبؤ باتجاه السوق وتحديد الحجم الأمثل لبيانات التدريب، مما يعزز من دقة التنبؤ باتجاهات الحركة.

بناء على ما سبق يمكن التعبير عن مشكلة البحث من خلال مجموعة من التساؤلات الرئيسية وهي:

1. كيف يمكن تحسين دقة التنبؤ باتجاه حركة السوق المالي باستخدام خوارزميات تعتمد على النمذجة الاحتمالية والتنقيب في البيانات؟
2. ما هو تأثير حجم بيانات التدريب غير المثالي على دقة التنبؤ باتجاه الحركة في السوق المالي؟
3. كيف سيتم بناءً على الخوارزمية المقترحة تحديد الحجم الأمثل لبيانات التدريب لتحقيق دقة أفضل للتنبؤ؟
4. إلى أي مدى يساهم استخدام النمذجة الاحتمالية في تعزيز دقة التنبؤ باتجاهات السوق؟

3.أهداف البحث:

يمكن تلخيص أهداف البحث في النقاط الآتية:

- ✓ **تطوير خوارزمية للتنقيب في البيانات:** بناء خوارزمية تستخدم التوزيع الاحتمالي لأنماط الحركة المتتالية لتحسين دقة التنبؤ باتجاه حركة السوق. ستعتمد الخوارزمية على تحليل البيانات التاريخية لاستخراج الأنماط.
- ✓ **تحديد حجم بيانات التدريب المثالي:** دراسة تأثير حجم بيانات التدريب على دقة التنبؤ باستخدام الخوارزمية المعتمدة. سيتم تحديد الحجم الأمثل من بيانات التدريب لتحديد أفضل معاملات للتنبؤ.
- ✓ **مقارنة الأداء:** إجراء مقارنة دقيقة بين دقة الخوارزمية الجديدة ونماذج ARIMA وARCH. حيث سيتم تحليل النتائج لتقييم فعالية الخوارزمية مقارنة بالنماذج القياسية وتطبيق معيار حساب القيمة المضافة لكل أسلوب تنبؤي متبع بهدف تعزيز هذه المقارنة.

4.أهمية البحث:

تتجلى أهمية هذا البحث في تقديم طريقة جديدة لتحسين دقة التنبؤ باتجاه الحركة في الأسواق المالية، مما يساعد المستثمرين على اتخاذ قرارات أكثر استنارة. كما يسهم البحث في تطوير الأدوات الإحصائية المستخدمة في تحليل البيانات المالية، مما يمكن أن يؤدي إلى تحسين استراتيجيات الاستثمار وتقليل المخاطر. وفي ظل التحديات التي تواجه الأسواق المالية، يمكن أن يوفر هذا البحث رؤى جديدة حول كيفية استخدام تقنيات التنقيب في البيانات بشكل فعال.

5. فرضيات البحث:

الفرضية الأولى: وفق معيار القيمة المضافة لنموذج التنبؤ باتجاه الحركة في السوق المالي: لا تعطي الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات قيمة مضافة أفضل من القيمة المضافة التي تعطيها كل من الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات وكذلك التي تعطيها النماذج القياسية ARIMA(G)ARCH.

الفرضية الثانية: لا يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقة التنبؤ بالاعتماد على نماذج ARIMA–ARCH

الفرضية الثالثة: لا يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقتها بالاعتماد على تحديد الحجم الأمثل للبيانات.

6. الدراسات السابقة:

- **دراسة** Imad Bou–Hamad و Ibrahim Jamali (2020)، **بعنوان**

Forecasting financial time–series using data mining models: A simulation study[1]

توقع السلاسل الزمنية المالية باستخدام نماذج تنقيب البيانات– دراسة محاكاة:

تتناول هذه الدراسة قدرة التنبؤ الثابتة والديناميكية لكل من الشبكات العصبية الاصطناعية وغابات القرار في السلاسل الزمنية المالية ضمن سياق محاكاة. تم تصميم المحاكاة لتوليد بيانات من نموذج(1,1)AR(1)–GARCH ، مما يسمح بوجود درجات متعددة من الاستمرارية في معادلة المتوسط لمحاكاة سلوك عوائد الأصول على المدى القصير والطويل.تشير النتائج إلى أن النموذج الحقيقي يتفوق على تقنيات التنقيب في البيانات في التنبؤ الثابت، ولكن الأهم هو أن تقنيات التنقيب في البيانات تتفوق على النموذج الحقيقي في إطار التنبؤ الديناميكي للسلاسل الزمنية ذات الاستمرارية المتوسطة إلى العالية. تم استخدام عوائد يوم واحد وعوائد طويلة الأجل على أسعار صرف معينة (الدولار مقابل الفرنك السويسري) كتطبيق تجريبي لمشاهدات تم تجزئتها إلى عدة عينات (1000.2000.5000.10000)، وأكدت النتائج التجريبية قدرة نماذج التنقيب في البيانات على التنبؤ بشكل أفضل للسلاسل الزمنية ذات الاستمرارية العالية. تناقش الدراسة أهمية هذه النتائج في سياق تخصيص الأصول وإدارة المحافظ، مشيرةً إلى أن مديري المحافظ يمكنهم تحقيق توقعات أكثر دقة لعوائد الأصول باستخدام تقنيات التنقيب في البيانات، مما يقلل الحاجة لإعادة توازن المحافظ ويخفض تكاليف المعاملات.

- **دراسة** Luo, J (2022)، **بعنوان:** Modeling of Data Mining Technology in Financial Data Recognition Mining and Forecasting:[2]_نمذجة تقنية التنقيب في البيانات في استخراج البيانات المالية والتنبؤ بها.

هدفت هذه الدراسة إلى استكشاف كيفية تطبيق تقنيات التنقيب في البيانات في مجال التحليل المالي، مع التركيز على تحسين دقة التنبؤ بالأحداث المالية المستقبلية من خلال تقنيات متقدمة. وقد اعتمدت الدراسة على نماذج التجميع وقواعد الارتباط وأشجار القرار لتحديد العلاقات بين المتغيرات، وقد أظهرت النتائج أن النظام المطور حسّن كفاءة التحليل المالي

[1] : Bou-Hamad, I., & Jamali, I. (2020). Forecasting financial time-series using data mining models: A simulation study. ***Research in International Business and Finance***, 51, 101072

[2]: Luo, J. (2022, January). Modeling of data mining technology in financial data recognition mining and forecasting. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1168-1171). IEEE.

بنسبة 8.2% مقارنة بالأساليب التقليدية. كما تم التأكيد على أهمية التنقيب في البيانات في تحسين دقة التنبؤات المالية، مما يعزز من قدرة المؤسسات المالية على اتخاذ قرارات مستنيرة.

- **دراسة** Asare, C., Asante, D., & Essel, J. F. (2023) **بعنوان:**

Probabilistic LSTM Modeling for Stock Price Prediction with Monte Carlo dropout Long Short-Term Memory Network:[1]

استخدمت هذه الدراسة النمذجة الاحتمالية LSTM لتوقع أسعار الأسهم باستخدام تقنية Monte Carlo للإسقاط في الشبكات العصبية طويلة وقصيرة المدى، حيث تم استخدام نموذج LSTM البايزي (ذاكرة طويلة وقصيرة المدى) لتوقع أسعار الأسهم ودرست أداءه مقارنةً بنموذج LSTM التقليدي. وذلك بالتطبيق في الشبكات العصبونية الاصطناعية، ولتحديد عدد العينات المأخوذة من النموذج البايزي لتقدير عدم اليقين، فقد تم تحديد عدد عينات مونت كارلو إلى 100. وقد تم اتخاذ هذه الخيارات بناءً على الممارسات القياسية والتجارب.

أظهرت النتائج أن نموذج LSTM البايزي ينتج نتائجاً أفضل من نموذج LSTM التقليدي عند النظر في قيم(R-squared) وMAPE(نسبة الخطأ المطلق المتوسطة) و RMSE(الجذر التربيعي لمتوسط الخطأ التربيعي). أظهر نموذج LSTM البايزي وجود عدم اليقين الكامن في البيانات الأساسية وتجنب إظهار ثقة مفرطة في تنبؤاته، حيث كانت نسبة الثقة 48.67% في توقع بيانات S&P 500 للدراسة. توفر هذه الدراسة نهجًا أكثر موثوقية لتوقع أسعار الأسهم لمساعدة المستثمرين والمهنيين الماليين في اتخاذ قرارات مستنيرة.

- **دراسة** Alzahrani, A., & Alshahrani (2023)، بعنوان: Financial Time Series Forecasting: A Data Stream Mining-Based System:[2] التنبؤ بالسلاسل الزمنية المالية: نظام قائم على التنقيب في تدفق البيانات.

تسعى هذه الدراسة إلى اقتراح منهجية جديدة للتنبؤ بسلاسل الزمن المالية باستخدام تقنيات تنقيب البيانات المتدفقة، مما يساعد في التعامل مع التغيرات السريعة في البيانات المالية حيث تؤدي التغيرات في السلاسل الزمنية المالية إلى تغير في توزيعها الاحتمالي بشكل تدريجي ومن ثم فقدان أي نموذج تم تدريبه سابقاً لقدرته على التنبؤ بشكل تدريجي. وللحفاظ على أداء تنبؤ جيد مع نغير توزيع البيانات استخدمت الدراسة خوارزمية الانحدار المتدرج العشوائي (SGD: Stochastic Gradient Descent algorithm) واعتمدت على منهجية DSM(Data Stream Mining) ، مما يسمح بتحسين أداء النموذج بشكل مستمر استنادًا إلى تدفق البيانات الجديد. حيث يتم تحديد أنماط الرسم البياني المتكررة في بيانات الفوركس التاريخية من خلال التنبؤ بقيم صرف اليورو/الدولار الأمريكي المستقبلية. يتم اكتشاف تغيير اتجاه البيانات باستخدام تقنية إحصائية تدرس ما إذا كانت حالات السلسلة الزمنية المستلمة (المتدفقة) ثابتة أم لا، وأهم النتائج: أظهرت النتائج أن التنبؤ باستخدام النظام المقترح يكون أفضل للتكيف وفقًا لتغيرات التوزيع المكتشفة مقارنة بنظام لا يتضمن أي تقنيات للكشف عن التحولات في الأنماط والاستجابة لها.

- **دراسة** Dar, G. F., Padi, T. R., & Rekha, S (2022)، بعنوان: STOCK PRICE PREDICTION USING A MARKOV CHAIN MODEL: A STUDY FOR TCS SHARE VALUES:[1] التنبؤ بسعر السهم باستخدام نموذج سلسلة ماركوف: دراسة لقيم أسهم شركة TCS:

[1] : Asare, C., Asante, D., & Essel, J. F.(2023). Probabilistic LSTM Modeling for Stock Price Prediction with Monte Carlo dropout Long Short-Term Memory Network. International Journal of Innovative Science and Research Technology. Volume 8, Issue 7, July – 2023. ISSN No:-2456-2165.

[2] : Bousbaa, Z., Sanchez-Medina, J., & Bencharef, O. (2023). Financial time series forecasting: a data stream mining-based system. *Electronics*,MDPI, *12*(9), 2039.

الهدف لرئيس من هذه الدراسة هو تطبيق سلسلة ماركوف لنمذجة وتوقع اتجاه أسعار أسهم شركة TCS Ltd للخدمات الاستشارية في سوق الأوراق المالية الهندية. أجريت الدراسة من خلال تصميم دراسة حالة طولية وتجريبية. تم استرجاع البيانات الثانوية لأسعار إغلاق الأسهم اليومية لشركة TCS Ltd. من (www.yahoofinance.com) على مدى الفترة من 1 يناير 2020 إلى 31 يناير 2022. معلمات نموذج سلسلة ماركوف هي مصفوفة احتمالية الانتقال ومتجه احتمالية الحالة الأولية بناءً على البيانات اليومية التاريخية. من أجل تلبية أهداف هذه الدراسة، تم الحصول على تحركات سلوك سعر السهم على المدى الطويل، والعدد المتوقع للزيارات إلى حالة معينة ووقت العودة المتوقع إلى حالات مختلفة.

تكشف نتائج مصفوفة الاحتمالات الثابتة أنه بغض النظر عن السعر الأولي لسهم TCS، في الأمد البعيد، يمكننا التنبؤ بأن سعر سهمها سيحقق مكسبًا مرتفعًا باحتمال 0.420747 ومكسبًا منخفضًا باحتمال 0.1105 وعدم تغير باحتمال 0.0291 وخسارة منخفضة باحتمال 0.0891 وخسارة عالية باحتمال 0.3504. ومن النتائج المذكورة أعلاه، نلاحظ أن احتمالات الخسارة العالية في أسعار أسهم TCS أقل من احتمالات الربح المرتفع في المستقبل ويمكننا أن نستنتج أن الاستثمار في سهم TCS هو خيار استثماري جيد للمستثمرين لتحقيق مكاسب رأسمالية. توضح النتائج أيضاً أنه إذا كانت حالة سعر سهم TCS في مكسب مرتفع، فسوف يصل إلى حالة الربح المرتفع مرة أخرى بعد يومين. ستصل السلسلة إلى حالة عدم التغيير بعد 34 يومًا إذا كانت في حالة عدم التغيير وأخيرًا تستغرق السلسلة 2.85 أو 3 أيام تقريبًا للوصول إلى حالة الخسارة المرتفعة إذا كانت في البداية في تلك الحالة.

- **بناءً على ما تم ذكره يمكن توضيح ما تتميز به هذه الدراسة عن الدراسات سابقة فيما يلي:**

✓ من حيث مجتمع الدراسة: الدراسة الحالية تستهدف دراسة أسواق مال متنوعة عالمية وعربية بهدف إثبات وتعميم الخوارزمية المقترحة.

✓ من حيث النماذج المستخدمة: الدراسة الحالية تسعى إلى بناء خوارزمية جديدة تستند في مضمونها إلى حساب احتمالات أنماط الحركة المتتالية اعتمادا على التنقيب في بيانات اتجاهات الحركة، وهي بذلك تنطلق من مبدأ سلاسل ماركوف، ولكن هذه الخوارزمية تدمج بين علم الاحتمالات والتعلم الآلي في تحديد الحجم المثالي لبيانات التدريب التي سيتم الاعتماد عليها في حساب الاحتمالات والتنبؤ، **وهذا يعد إضافة علمية في هذا المجال.**

✓ من حيث أهداف الدراسة: تتميز هذه الدراسة عن الدراسات السابقة بهدفها الأساسي والذي هو بناء خوارزمية جديدة تستخدم التوزيع الاحتمالي لأنماط الحركة المتتالية. وتحديد حجم بيانات التدريب المثالي لتحديد أفضل معاملات للتنبؤ. إضافة إلى مقارنة الأداء من خلال إجراء مقارنة دقيقة بين دقة الخوارزمية الجديدة ونماذج ARIMA وARCH بهدف جعلها نقطة مرجعية لاختبار أداء الخوارزمية المقترحة، فبدون إجراء هذه المقارنة المرجعية لا يمكننا تقييم كفاءة أداء الخوارزمية المقترحة.

7.المراجعة الأدبية:

1.7. مفهوم النمذجة الاحتمالية:

النمذجة الاحتمالية هي فرع من فروع الرياضيات والإحصاء الذي يهدف إلى تمثيل وتحليل الظواهر العشوائية باستخدام نماذج رياضية. تعتمد هذه النمذجة على مفهوم المتغيرات العشوائية، والتي تُستخدم لوصف النتائج المحتملة للأحداث.

تعتبر النمذجة الاحتمالية أداة من الأدوات التي تستخدم في تحليل البيانات المالية، حيث تُستخدم لوصف وتوقع سلوك الأسواق المالية من خلال تمثيل عدم اليقين. تعتمد هذه النماذج على استخدام المتغيرات العشوائية لتقدير النتائج المحتملة،

[1] : Dar, G. F., Padi, T. R., & Rekha, S. (2022). Stock price prediction using a Markov chain Model: a study for TCS Share Values. *Advances and Applications in Statistics*, *80*, 83-101.

مما يسمح للباحثين والممارسين بفهم الأنماط المحتملة في البيانات المالية واتخاذ قرارات مستنيرة بناءً على تقديرات احتمالية.[1] وفيما يلي توضيح لبعض المفاهيم المرتبطة بالنمذجة الاحتمالية:

المتغير العشوائي :هو كيان رياضي يمثل نتيجة عشوائية يمكن أن تأخذ قيمًا مختلفة بناءً على الظروف المحيطة، ويمكن تصنيفه إلى نوعين: متغيرات عشوائية متقطعة ومستمرّة. [2]

النماذج الشبكية الاحتمالية: تُعتبر أداة قوية لتمثيل العلاقات المعقدة بين المتغيرات العشوائية، مما يساعد في فهم التفاعلات والاعتماديات.[3]

دالة الكثافة الاحتمالية: تستخدم لوصف توزيع الاحتمالات للمتغيرات العشوائية، حيث تعبر عن احتمال وقوع حدث معين ضمن نطاق محدد[4]

وباختصار، النمذجة الاحتمالية هي أداة لا غنى عنها في الأسواق المالية، حيث تمكن من التنبؤات الأفضل وتقييم المخاطر في ظل حالة عدم اليقين. من خلال الاستفادة من التقنيات الإحصائية والتقدم المحرز في التعلم الآلي، يمكن للمحللين الماليين تعزيز فهمهم لسلوكيات السوق وتحسين عمليات صنع القرار.

2.7 . التنقيب في البيانات والحجم الأمثل لها:

1.2.7. مفاهيم أساسية في تنقيب البيانات:[5]

- ✓ تنقيب البيانات: هو عملية اكتشاف الأنماط والمعرفة من كميات كبيرة من البيانات. يشمل ذلك تقنيات مختلفة من الإحصاء وتعلم الآلة.
- ✓ اكتشاف المعرفة في قواعد البيانات (KDD): هو عملية أوسع تشمل تنقيب البيانات كأحد خطواتها، مع التركيز على استخراج المعلومات المفيدة من البيانات.
- ✓ معالجة البيانات الأولية: تعد ضرورية لتحضير البيانات الخام للتحليل، وتشمل التنظيف والتحويل وتقليل البيانات لتحسين جودة وكفاءة عملية التنقيب.

2.2.7. المهام الرئيسية في تنقيب البيانات:[6]

يمكن تلخيص أهم المهام الرئيسية لاستخدام التنقيب في البيانات في النقاط التالية:

- ✓ تنقيب الأنماط المتكررة: يمكن أن تتضمن:

– قواعد الارتباط: تقنيات مثل خوارزمية Apriori تُستخدم للعثور على العلاقات المثيرة للاهتمام بين المتغيرات في قواعد بيانات كبيرة (مثل تحليل سلة السوق).

– المجموعات المتكررة: تحديد مجموعات العناصر التي تظهر معًا بشكل متكرر في المعاملات.

- ✓ التصنيف (Classification): يمكن أن يشمل:

– بناء النماذج: إنشاء نماذج يمكنها التنبؤ بتسميات الفئات بناءً على الميزات المدخلة. تشمل الخوارزميات الشائعة أشجار القرار، آلات الدعم الناقل، والشبكات العصبية.

– تحليل الانحدار: التنبؤ بالقيم المستمرة بناءً على المتغيرات المدخلة.

---

[1] : Ross, S. M. (2010). *Introduction to Probability Models* (10th ed.). Academic Press.PP:3-5.

[2] : Evans, M. J., & Rosenthal, J. S. (2004). *Probability and Statistics: The Science of Uncertainty*. Springer.P12.

[3] : المرجع السابق نفسهزص:78.

[4] : Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to Probability*. Athena Scientific.P:45.

[5] : Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and. *Techniques, Waltham: Morgan Kaufmann Publishers*.(PP:3-12).

[6] : المرجع السابق نفسه.

- ✓ التجميع (Clustering): تجميع مجموعة من الكائنات بطريقة تكون فيها الكائنات في نفس المجموعة (التجمع) أكثر تشابهًا مع بعضها البعض مقارنة بتلك الموجودة في مجموعات أخرى. تشمل التقنيات التجميع باستخدام خوارزمية k– means والتجميع الهرمي.
- ✓ كشف الشذوذ (Anomaly Detection): تحديد العناصر أو الأحداث النادرة التي تختلف بشكل كبير عن غالبية البيانات. هذا مهم للكشف عن الاحتيال ومراقبة الجودة.
- ✓ التعلم العميق (Deep Learning): فرع من تعلم الآلة يتضمن الشبكات العصبية ذات الطبقات العديدة (الشبكات العميقة) التي يمكن أن تتعلم الأنماط المعقدة في مجموعات بيانات كبيرة.

3.2.7. حجم البيانات المثالي:

يتطلب تحسين قدرة التنبؤ فهمًا دقيقًا لحجم البيانات المثالي الذي يمكن استخدامه لتحقيق نتائج دقيقة وموثوقة ويعتمد تحديد حجم البيانات المثالي على عدة عوامل، منها:[1]

- نوع البيانات: البيانات المالية تتطلب دقة عالية، لذا يجب أن تكون مجموعة البيانات شاملة وممثلة للسوق.
- التحليل المطلوب: بعض التحليلات تتطلب بيانات تاريخية طويلة، بينما يمكن أن تكون بعض التحليلات أكثر فعالية مع بيانات أقل.
- القدرة الحاسوبية: يجب أن يتناسب حجم البيانات مع قدرة الأنظمة المستخدمة في التحليل. البيانات الضخمة قد تتطلب موارد حاسوبية أكبر.
- التوازن بين الجودة والكمية: يجب أن تكون البيانات ذات جودة عالية، حيث أن البيانات السيئة يمكن أن تؤدي إلى نتائج مضللة.

إن تحديد الحجم الأمثل للبيانات في النمذجة الاحتمالية هو موضوع معقد يعتمد على عدة عوامل، بما في ذلك تعقيد النموذج، خصائص البيانات، والسياق التطبيقي المحدد. فيما يلي بعض النقاط الرئيسية المستندة إلى دراسات حديثة:

- ✓ تعقيد النماذج: يؤثر تعقيد النموذج الاحتمالي بشكل كبير على كمية البيانات المطلوبة. على سبيل المثال، يمكن أن تؤدي النماذج البسيطة أداءً جيدًا مع مجموعات بيانات أصغر، بينما تتطلب النماذج الأكثر تعقيدًا مثل العمليات الغاوسية (Gaussian Processes) مجموعات بيانات أكبر لالتقاط الأنماط الأساسية بشكل فعال[2].
- ✓ تغير البيانات: يمكن أن يتطلب التغير العالي في البيانات إلى أحجام بيانات أكبر لضمان تقديرات موثوقة. على سبيل المثال، في دراسة حول أوقات الإصلاح لمكونات توزيع الطاقة، وُجد أن الاعتماد فقط على القيم المتوسطة قد يكون مضللاً بسبب التغير العالي الملحوظ في أوقات الإصلاح[3].

باختصار، يتطلب تحديد الحجم الأمثل للبيانات في النمذجة الاحتمالية النظر بعناية في تعقيد النموذج، وتغير البيانات، واحتياجات التطبيق المحددة. بينما لا توجد إرشادات عالمية، فإن الاستفادة من مجموعات بيانات واسعة وعالية الجودة تؤدي عمومًا إلى تحسين أداء النموذج وتوقعاته الأكثر موثوقية.

---

[1] : Provost, F. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking* (Vol. 355). O'Reilly Media, Inc.

[2] : Babichev, D., & Bach, F. (2018). Constant step size stochastic gradient descent for probabilistic modeling. *arXiv preprint arXiv:1804.05567*.

[3] : Rojas, F., Leiva, V., Huerta, M., & Martin-Barreiro, C. (2021). Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data. *Sensors*, *21*(15), 5198.

3.7. النماذج القياسية ARIMA-ARCH: [1] [2] [3]

تُعتبر النماذج القياسية ARIMA–ARCH من الأدوات الأساسية في تحليل السلاسل الزمنية، حيث تؤدي دورًا محوريًا في التنبؤ بالبيانات التي تتسم بالتغيرات المتكررة واللاخطية.

نموذج ARIMA (Autoregressive Integrated Moving Average) يُستخدم بشكل رئيسي للتعامل مع السلاسل الزمنية التي تحتاج إلى تحويلها إلى سلسلة مستقرة عبر الزمن. في المقابل، يُعنى نموذج ARCH (Autoregressive Conditional Heteroskedasticity) بتحليل التباين المتغير بمرور الزمن، مما يجعله مناسبًا للبيانات المالية التي تتميز بالتقلبات العالية.

يمكن دمج هذين النموذجين لإنشاء نموذج هجين يُعرف بـ ARIMA–ARCH، الذي يُستخدم على نطاق واسع في مجالات مثل الاقتصاد والتمويل لتقديم تنبؤات دقيقة حول الأسعار والتقلبات. أظهرت الدراسات أن النماذج الهجينة مثل ARIMA–GARCH تستطيع تحسين دقة التنبؤ بشكل ملحوظ مقارنة باستخدام كل نموذج على حدة، مما يعكس أهمية اختيار النموذج المناسب بناءً على خصائص البيانات المدروسة.

ويمكن التعبير عن نموذج ARIMA (Autoregressive Integrated Moving Average) من خلال ثلاثة مكونات رئيسية:

1. **المكون الذاتي :(AR)** يمثل الاعتماد الذاتي على القيم السابقة للسلسلة الزمنية. يتم التعبير عنه بالمعادلة:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon t \ldots\ldots(1)$$

حيث $\phi i \phi i$ هي معاملات النموذج ϵt: هو الحد العشوائي.

2. **المكون المتكامل :(I)** يُستخدم لجعل السلسلة الزمنية مستقرة عبر الزمن من خلال إجراء عمليات الفروق.

3. **المكون المتحرك :(MA)** يمثل تأثير الأخطاء السابقة في التنبؤات. يُعبر عنه بالمعادلة:

$$Y_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \ldots\ldots(2)$$

كما يمكن التعبير عن **نموذج** ARCH (Autoregressive Conditional Heteroskedasticity) والذي يهدف إلى تحليل التباين المتغير بمرور الزمن، من خلال الصيغة الرياضية التالية:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1^2} + \alpha_2 \epsilon_{t-2^2} + \cdots + \alpha_q \epsilon_{t-q^2} \ldots\ldots(3)$$

حيث σ_t^2هو التباين الشرطي و αiهي معاملات النموذج.

8. القسم العملي:

1.8. الأسواق المدروسة: تم اعتماد العينة بحيث تتيح تعميم النتائج وتطبيقها على نطاق أوسع، مما يزيد من قيمة البحث وأهميته للمستثمرين والمحللين الماليين في مختلف الأسواق. وقد تم اختيار الأسواق بناءً على معايير منهجية تُراعي تنوع

[1] : Monika, P., Ruchjana, B., & Abdulla, A. (2022). The implementation of the ARIMA-ARCH model using data mining for forecasting rainfall in Bandung city. *International Journal of Data and Network Science*, *6*(4), 1309-1318.

[2] : Yahaya, A., Etuk, E., & Emeka, A. (2021). Comparative performance of arima and garch model in forecasting crude oil price data. *Asian Journal of Probability and Statistics*, *15*(4), 251-275.

[3] : Nugraha, E. S., & Alvina, C. (2024). THE APPLICATION OF STANDARD GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (SGARCH) MODEL IN FORECASTING THE STOCK PRICE OF BARITO PACIFIC. *BAREKENG Journal of Mathematics and Its Applications*, *18*(2), 0849-0862.

المؤشرات العالمية والإقليمية، واختلاف الظروف الاقتصادية وتوافر البيانات اللازمة لضمان اختبار شامل لفعالية الخوارزمية في التنبؤ بحركة الأسواق المالية. وفيما يلي توضيح لهذه العينة ومبررات اختيارها:

أولاً مؤشرات أسواق المال العالمية :

– المؤشرات الرئيسية من مثل: S&P500وDow Jonesو NASDAQ (الولايات المتحدة): والهدف هو:

– تمثيل الاقتصاد الأمريكي (أكبر اقتصاد عالمي) مع قطاعات متنوعة (تكنولوجيا، صناعة، خدمات).

– سيولة عالية جدًا وتقلبات معتدلة إلى مرتفعة، مما يتيح قياس دقة الخوارزمية في ظروف سوقية مختلفة.

– توافر بيانات تاريخية مفصلة ومحدثة بشكل فوري.

– Nikkei 225 (اليابان): مؤشر رئيسي لأسواق آسيا، يعكس أداء قطاعات التصنيع والتكنولوجيا اليابانية.

– Hang Seng (هونغ كونغ) وShanghai Composite (الصين): تمثيل الأسواق الصينية الناشئة والمُتحكمة باقتصاد عالمي ضخم ، وتأثرها العالي بالأحداث الجيوسياسية والسياسات الحكومية.

ثانياً: مؤشرات أسواق مال عربية :

– مؤشر دبي (DFM) وأبو ظبي (ADX) (الإمارات): والتي تعتبر مركز مالي إقليمي ذو سيولة عالية، مع تركيز على قطاعات العقارات والخدمات المالية. وكذلك تأثر الأسواق بالتقلبات في أسعار النفط والاستثمارات الأجنبية.

– السعودي (تداول) وقطر (QSE): وهي أسواق ذات رسملة سوقية كبيرة مدعومة باقتصادات نفطية قوية، وتحركاتها تعكس التوجهات الاستثمارية في منطقة الخليج وتأثيرات أسعار الطاقة.

– مصر (EGX) وعمان (MSX):اقتصادات متنوعة (زراعة، سياحة، صناعة) مع تعرضها لتقلبات العملة المحلية والاضطرابات السياسية. وهي مناسبة لاختبار قدرة الخوارزمية على التنبؤ في بيئات ذات مخاطر سيولة متوسطة.

– العراق (ISX) ودمشق (DSE): أسواق ناشئة (أو شبه ناشئة كسوق دمشق للأوراق المالية) ذات سيولة محدودة وتأثر كبير بالأحداث السياسية والأمنية، تُستخدم كحالة دراسة لقياس أداء الخوارزمية في أسواق هامشية عالية المخاطر.

ثالثاً: معايير التقييم المشتركة:

– التنوع الاقتصادي: تغطية أسواق نفطية (السعودية، أبو ظبي)، واقتصادات متنوعة (مصر، اليابان)، وأسواق تكنولوجيا (NASDAQ).

– التقلب التاريخي: مثل تقلبات "شانغ هاي" المرتبطة بالسياسات الصينية، أو تقلبات "نايكي" مع تغير سعر الين.

– العمق الاستراتيجي: اختبار الخوارزمية في أسواق ذات كفاءة عالية (مثل التطبيق على مؤشر S&P500) مقابل أسواق أقل كفاءة (مصر، العراق ، دمشق) لقياس قابليتها للتكيف.

رابعاً: الهدف النهائي: التحقق من قدرة الخوارزمية على:التعامل مع البيانات عالية التردد (في الأسواق العالمية). والتنبؤ في ظل مستويات مختلفة من السيولة (من دبي عالية السيولة إلى دمشق محدودة السيولة).

هذا الاختيار يُعتبر نموذجًا مصغرًا للاقتصاد العالمي، مما يعزز موثوقية النتائج وقابلية تعميمها.

2.8. الخوارزمية المقترحة:

1.2.8. خطوات الخوارزمية المقترحة: تستهدف الخوارزمية المقترحة تحسين دقة التنبؤ باتجاه حركة الأسعار في السوق المالي من خلال تحليل البيانات التاريخية. حيث يتم ذلك عبر عدة خطوات رئيسية، ويمكن تلخيص مبدأ عمل الخوارزمية المقترحة من خلال الخطوات الآتية:

1. تعريف التسلسلات الممكنة: تبدأ الخوارزمية بتعريف مجموعة من التسلسلات المحتملة التي تمثل الاتجاهات المختلفة لحركة الأسعار. فمثلاً يمكن أن يكون كل تسلسل يتكون من ثلاثة عناصر تأخذ القيم (هابط:–1، ثابت: 0، صاعد: 1)، مما يسمح بتغطية جميع السيناريوهات الممكنة.

2. تقسيم البيانات: يتم تحديد نسبة معينة من البيانات لاستخدامها كبيانات اختبار، حيث يتم تقسيم البيانات إلى ثلاث مجموعات، مجموعة التدريب والتي يتم فيها حساب الاحتمالات وجدول التوزيع الاحتمالي لأنماط اتجاهات الحركة، ومجموعة الاختبار الأولي ومجموعة الاختبار النهائي، ونلاحظ هنا أنه تم تقسيم مجموعة الاختبار إلى مجموعتين متساويتين وذلك بهدف اختبار استقرار النتائج من جهة، ومن جهة أخرى استخدام مجموعة الاختبار الأولي كجزء من التقييم والتغذية الراجعة أثناء بناء جدول التوزيع الاحتمالي وتحديد حجم البيانات الأمثل.

3. تحديد نقطة البداية لبيانات التدريب: تعتمد الخوارزمية على اختيار نقاط بداية متدرجة لبيانات التدريب بدءاً من النقطة الأولى (المشاهدة الأولى) للسلسلة الزمنية، مما يعني أنها تختبر عدة فترات زمنية لبيانات التدريب وبشكل متدرج. هذا يساعد في تحديد أفضل نقطة بداية لتحقيق أعلى دقة للتنبؤ، وبالتالي تحديد الحجم الأمثل للبيانات.

4. بناء جدول التوزيع الاحتمالي الشرطي لجميع الحالات المختلفة لمتسلسلات الاتجاه: حيث يتم حساب احتمال تحقق الحالات المختلفة للمتسلسلات في كل مرة يتحدد فيها بداية جديدة للسلسلة الزمنية للبيانات (التي تتم في الخطوة السابقة) بحيث يتم حساب الاحتمال الشرطي لكل حالة من حالات الاتجاه (صعودي – هبوطي – ثابت) مشروط بالحالات السابقة المقابلة لها وذلك بالاعتماد على التعريف الاحصائي للاحتمالات.

5. التنبؤ: يتم بتحديد أنماط الاتجاهات السابقة (أحداث وقعت بالفعل) في كل خطوة يتم التنبؤ بها، وبعدها فإن التنبؤ في الخطوة التالية (الاتجاه التالي) تعتمد على الحالة الأكثر احتمالاً والمحسوبة سابقاً (في الخطوة السابقة) من جدول التوزيع الاحتمالي لجميع الحالات المختلفة للمتسلسلات.

6. حساب الدقة: يتم حساب دقة النموذج باستخدام مجموعة بيانات التدريب ومجموعة بيانات الاختبار الأولي. هنا، يتم استخدام دالة لحساب الدقة بناءً على التنقيب في بيانات التدريب فقط، حيث تُحسب الاحتمالات المرتبطة بالتسلسلات الممكنة استنادًا إلى بيانات التدريب فقط، اما بيانات الاختبار الأولي وبيانات الاختبار النهائي فلا يتم فيها أية عملية تنقيب في البيانات.

7. حساب الدقة المرجَحة: تُحسب الدقة المرجَحة للخوارزمية بناءً على حجم بيانات التدريب والاختبار الأولي. حيث يتم في كل مرة محاكاة بيانات التدريب (والتي تم على أساسها حساب جدول التوزيع الاحتمالي لأنماط الحركة) والتنبؤ ببيانات الاختبار الاولي ومن ثم حساب الدقة النهائية المرجحة لكل من بيانات التدريب والاختبار الأولي معاً، كي يتسنى تقييم فعالية التوزيع الاحتمالي المبني على بيانات التدريب فقط وقدرته على التنبؤ مستقبلاً (الاختبار الأولي) والوصول إلى أفضل سيناريو يعطي أعلى دقة. وهذا يسمح للخوارزمية بتحديد الأداء العام لقدرتها على التنبؤ بشكل أكثر استقراراً.

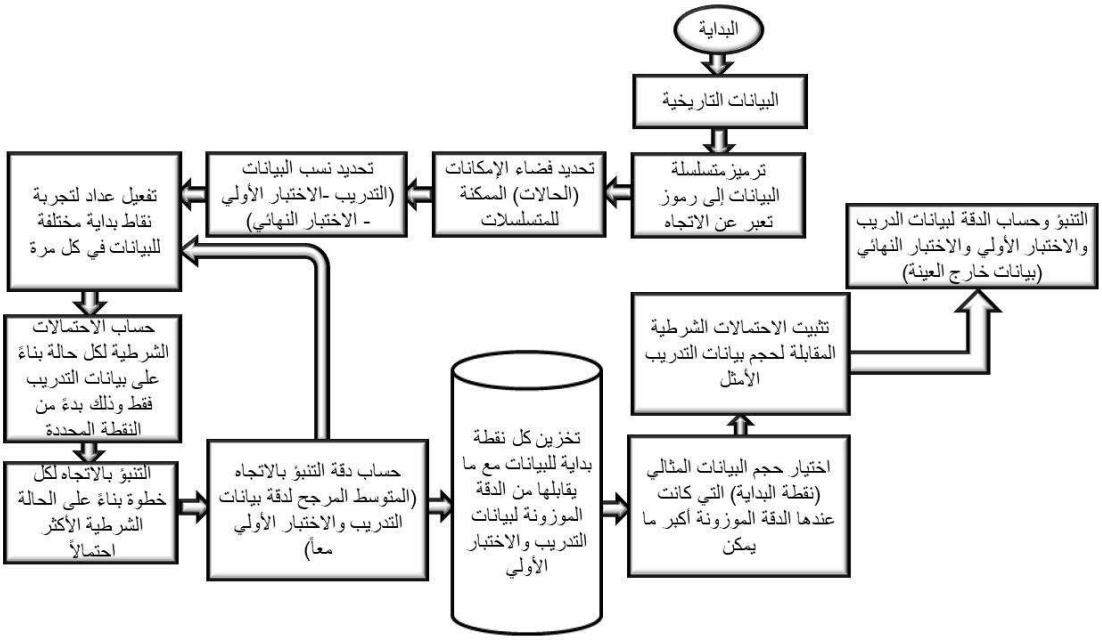
8. الهدف من مجموعة الاختبار الأولي: الهدف الأساسي من حساب الدقة المرجحة هو تحسين أداء النموذج عن طريق مراعاة تأثير حجم كل مجموعة من البيانات (التدريب والاختبار الأولي) على دقة التنبؤات. وبينما تُحسب الاحتمالات فقط بناءً على بيانات التدريب، فإن الدقة وتحديد حجم البيانات المثلى يعتمد على كل من دقة التنبؤ للبيانات الخاصة ببيانات التدريب وبيانات الاختبار الأولي وذلك لضمان أن الخوارزمية تعمل بشكل جيد في سيناريوهات متعددة. بهذه الطريقة، تضمن الخوارزمية تحقيق توازن بين التعلم من البيانات المتاحة وتقييم الأداء بشكل شامل وتحقيق الاستقرار والاستفادة من الاختبار الأولي كتغذية راجعة أثناء التدريب لتحديد الاحتمالات المثلى ونقطة البداية الأمثل للبيانات.

9. تحديد الحجم الأمثل للبيانات واعتماد الاحتمالات المثلى: في كل نقطة بداية جديدة للبيانات تتم المقارنة، فإذا كانت الدقة المحسوبة (المرجحة) عند هذه النقطة أعلى من الدقة المثلى السابقة لها، يتم تحديث النقاط المثلى ونتائج الاحتمالات النهائية وهكذا حتى نصل إلى النقطة المثلى في البيانات والتي عندها تكون دقة التنبؤ في لبيانات التدريب والاختبار الأولي أعلى ما يمكن وعندها يتم تثبيت الاحتمالات المثلى في جدول التوزيع الاحتمالي النهائي وتحديد الحجم الأمثل للبيانات.

10. إجراء التنبؤات النهائية: بعد تحديد أفضل نقطة بداية (تحديد الحجم الأمثل للبيانات)، تُجرى التنبؤات النهائية على مجموعات البيانات المختلفة (التدريب، الاختبار الأولي، والاختبار النهائي).

11. إخراج النتائج: تُجمع النتائج في هيكل بيانات وتُخرج، بما في ذلك دقة الخوارزمية وحجم البيانات المثلى ودقة التنبؤ لبيانات التدريب ولبيانات الاختبار الأولي ولبيانات الاختبار النهائي كل مجموعة على حدة، وهذا يسهم في دراسة استقرار تنبؤ الخوارزمية خلال فترة زمنية محددة.

2.2.8. المخطط العام للخوارزمية المقترحة:



الشكل رقم(1): المخطط العام للخوارزمية المقترحة

المصدر: من إعداد الباحث باستخدام PowerPoint

هذا المخطط يوضح كيفية تفاعل المكونات المختلفة للخوارزمية لتحقيق الهدف النهائي وهو تحسين دقة التنبؤ باتجاه حركة الأسعار في السوق المالي.

3.2.8. خطوات تحديث البيانات بشكل دائم للخوارزمية المقترحة:

من أجل ضمان استمرارية الأداء الجيد للخوارزمية يجب تحديث قاعدة البيانات بشكل دائم لضمان فعالية الخوارزمية في قدرتها على التنبؤ ويكون ذلك من خلال الخطوات التالية:

بدء الخوارزمية ← جمع البيانات الجديدة ← دمج البيانات الجديدة مع القديمة← تحديث مجموعة البيانات (تقسيم مجموعة البيانات للتدريب والاختبار) ← **إعادة التدريب باستخدام البيانات المحدثة ← تقييم الأداء بعد إعادة التدريب.**

3.8. نتائج تطبيق الخوارزمية المقترحة والنماذج القياسية ARIMA–ARCH:

تم تطبيق الخوارزميات المقترحة من خلال كتابة كود برمجي كامل في برنامج MATLAB وتم أيضاً استخدام البرنامج القياسي EViews لبناء النماذج القياسية، حيث تم الأخذ بالاعتبار كافة شروط تطبيقها وتحديد رتب كل نموذج وبناء عدة نماذج واختبار أفضلها وفق معيار Schwarz وأيضاً التحقق من توزيع البواقي وغيرها من الشروط وذلك لكل مؤشر من مؤشرات أسواق المال المدروسة على حدة.

الجدول رقم (1): يلخص نتائج دقة التنبؤ باتجاه الحركة لنماذج ARIMA–(G)ARCH والخوارزمية المقترحة:

نتائج دقة التنبؤ باتجاه الحركة لنماذج ARIMA-(G)ARCH والخوارزمية المقترحة												
مؤشرات أسواق المال	حجم المشاهدات الكلية (السلسلة الزمنية) N	حجم المشاهدات المتنبأ بها (بيانات جديدة مستقبلية خارج بيانات التدريب)	تطبيق نموذج ARIMA-ARCH			تطبيق الخوارزمية المقترحة						
						قبل تحديد الحجم الأمثل للبيانات			بعد تحديد الحجم الأمثل للبيانات			
			أفضل نموذج	دقة محاكاة بيانات التدريب	دقة التنبؤ للبيانات المستقبلية	حجم بيانات التدريب	دقة محاكاة بيانات التدريب	دقة التنبؤ للبيانات المستقبلية	نقطة البداية المثلى لبيانات التدريب	حجم بيانات التدريب + الاختبار الأولي	دقة محاكاة بيانات التدريب	دقة التنبؤ للبيانات المستقبلية
S&P500	1004	28	GARCH(1.1) ARIMA(1.1.1)	48.82%	%57.14	986	53%	64.3%	931	40	%72.34	%78.57
داونجونز	1257	29	ARCH(2) ARIMA(3.1.4)	50.2%	%37.9	1228	54.84%	57.14%	1184	44	68.28%	65.52%
نازداك	1257	29	ARCH(4) ARIMA(1.4.3)	55.59%	68.42%	1228	55.52%	57.89%	1203	19	80.01%	68.42%
نايكي	1222	16	GARCH(2.1) ARIMA(3.1.3)	55.52%	68.65%	1163	54.17%	75%	1176	30	76.78%	68.75%
هانغ سينغ	1229	20	GARCH(1.1) ARIMA(2.1.2)	52.36%	65%	1209	53%	60%	1172	37	70.22%	65%
شانغ هاي	1426	28	ARIMA(4.0.3)	47.88%	28.57%	1378	53.45%	64.29%	1356	42	71.27%	71.42%
دبي	1469	18	ARCH(2) ARIMA(3.1.2)	51.8%	%50	1451	%56.68	%50	1406	45	73.24%	72.22%
أبو ظبي	1477	17	GARCH(1.1) ARIMA(1.1.1)	54.59%	52.94%	1460	56.05%	%52.94	1431	29	68.78%	70.58%
السعودي	1468	24	ARIMA(4.1.2)	56.28%	%41.66	1444	57.51%	37.5%	1358	86	63.93%	62.5%
عمان	1425	23	ARIMA(3.1.2)	55.36%	%47.82	1402	54.58%	34.78%	1356	46	69.56%	73.91%
مصر	1179	25	GARCH(1.2) ARIMA(3.1.2)	54.95%	%52	1154	55.75%	60%	1113	41	65.9%	76%
قطر	1468	18	ARCH(1) ARIMA(3.1.2)	55.46%	44.44%	1450	%55.96	%50	1426	24	80.15%	%72.22
العراق	1093	30	ARCH(1) ARIMA(2.1.3)	48.49%	23.33%	1063	%54.02	%70	998	65	72.27%	86%
دمشق	1235	30	ARCH(2) ARIMA(3.1.3)	63.14%	53.53%	1205	%63.8	%50	966	239	66.5%	73.33%

المصدر: من إعداد الباحث بالاعتماد على مخرجات برنامج MATHLAB وبرنامج Eviews

في الجدول السابق تم حساب دقة التنبؤ باتجاه الحركة من خلال قسمة عدد الأيام التي كان فيها توافق باتجاه الحركة على عدد الأيام الكلي الخاص بكل مرحلة على حدة، ومن خلال الجدول السابق نلاحظ بشكل عام تفوق الخوارزمية المقترحة على النماذج القياسية من حيث القدرة على التنبؤ باتجاه الحركة في السوق المالي، كما يلاحظ أن الخوارزمية المقترحة المحسنة (التي تم من خلالها تحديد الحجم الأمثل للبيانات) قد تفوقت على الخوارزمية المقترحة التي لا يتم فيها تحديد الحجم الأمثل للبيانات. وهذه المقارنة هي مقارنة مبدئية ووصفية وسيتم اثبات ذلك رياضياً واحصائياً فيما يلي. حيث سيتم بداية تطبيق المعيار المقدم من قبل الحميد في أطروحتة الدكتوراة والذي من خلاله يمكن قياس القيمة المضافة التي يعطيها أي نموذج تنبؤي باتجاه الحركة في السوق المالي، ومن خلال ذلك يمكن المقارنة بين النماذج والخوارزميات المقترحة. ومن جهة أخرى سيتم دراسة وجود فروق معنوية احصائياً لمتوسط دقة التنبؤ بالاتجاه بين النماذج والخوارزميات المقترحة.

4.8. تطبيق معيار القيمة المضافة للنموذج التنبؤي:[1]

هذا المعيار تمَّ تقديمه من قبل الحميد في أطروحته الدكتوراة والذي من خلاله يمكن قياس القيمة المضافة التي يعطيها أي نموذج تنبؤي باتجاه الحركة في السوق المالي ومن خلال ذلك يمكن المقارنة بين النماذج والخوارزميات المقترحة، حيث تقوم فكرة المعيار المقترح على الآتي:

في حال كان الهدف الأساسي هو تقييم وقياس فعالية النموذج (أو الأسلوب المتبع) من حيث قدرته على التنبؤ باتجاه الحركة في السوق المالي فيتم ذلك من خلال مقارنة نسبة التوافق باتجاه الحركة لهذا النموذج مع نقطة مرجعية تتمثل في الحد الأدنى لنسبة التوافق باتجاه الحركة الممكن تحققها بدون تطبيق أي نموذج تنبؤي، ويتم ذلك باعتبار أنَّ اتجاه الحركة في اليوم (t) هو نفسه في اليوم (t+1) وهذه الطريقة تتم عموماً في حال كان السوق المالي يتمتع بكفاءة عالية، فعندها يكون من الصعب التنبؤ باتجاه الحركة، لذا فأفضل تنبؤ بالاتجاه هو اعتبار اتجاه الحركة غداً هو نفسه اتجاه الحركة اليوم، وهكذا. وعلى هذا الأساس يتم حساب نسبة التوافق باتجاه الحركة في كل خطوة اعتماداً على الاتجاه الحالي ولنرمز للحد الأدنى لنسبة التوافق باتجاه الحركة الممكن تحققها بالرمز(cr_{Min}) ولنرمز لنسبة التوافق باتجاه الحركة التي يحققها النموذج بالرمز(Cr_{Model}) عندها فإنه يمكن حساب النسبة المئوية للقيمة المضافة للنموذج تحت التقييم في التنبؤ باتجاه الحركة وفق العلاقة (1) الآتية:

$$\text{Percentage Cr added value} = \frac{cr_{Model} - cr_{Min}}{cr_{Min}} * 100 \ldots\ldots (4)$$

ووفقاً للعلاقة (4) السابقة إذا كانت إشارة الناتج (النسبة المئوية للقيمة المضافة لنموذج التنبؤ باتجاه الحركة) موجبة فهذا يشير إلى أن النموذج يمكن الاعتماد عليه في التنبؤ لأنه يعطي قيمة مضافة ويؤدي دور إيجابي في عملية التنبؤ باتجاه الحركة، وقيمة النسبة تعبر عن ثقل الاستفادة من هذا النموذج. أما إذا كانت إشارة الناتج سالبة فهذا مفاده أن النموذج يؤدي دور سلبي في عملية التنبؤ بالاتجاه ولا يمكن الاعتماد عليه في عملية التنبؤ باتجاه الحركة وقيمة النسبة تشير إلى ثقل السلبية فيما لو تمَّ الاعتماد عليه في التنبؤ بالاتجاه. فمثلاً قد نستطيع تحقيق نسبة توافق قد تصل إلى 62% اعتماداً على اتجاه الحركة في الفترة السابقة بدون تطبيق أو اعتماد أي نموذج تنبؤي في كل مرَّة يتم بها التنبؤ باتجاه الحركة (cr_{Min}) فإذا لم يستطع النموذج تحت التقييم تحقيق نسبة توافق باتجاه الحركة أكبر من (cr_{Min}) ولتكن مثلاً (58%) فعندها يمكن رفض هذا النموذج لأنه لم يقدم أي قيمة مضافة بالرغم من أن نسبة التوافق التي حققها أكبر من (50%). وهذا لأن الاعتماد على الاتجاه الحالي للتنبؤ بالاتجاه المستقبلي أفضل من الاعتماد على هذا النموذج.

وفيما يلي نتائج تطبيق هذا المعيار:

[1] : الحميد، أيهم .(2020) .*الكفاءة السعرية والتنبؤ بمؤشر السوق المالي باستخدام النماذج القياسية والشبكات العصبونية الاصطناعية* (دراسة مقارنة بين سوق دمشق للأوراق المالية وبعض أسواق المال العربية)، أطروحة دكتوراه، جامعة حماة، سوريا.

الجدول رقم (2): نتائج المعيار المطبق لقياس القيمة المضافة لنماذج وخوارزميات التنبؤ المقترحة

القيمة المضافة لنموذج وخوارزمية التنبؤ وفق المعيار المطبق								
السوق المالي	Cr min%		النماذج القياسية ARIMA (G)ARCH		القيمة المضافة للخوارزمية المقترحة قبل تحديد الحجم الأمثل للبيانات		القيمة المضافة للخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات	
	التدريب	التنبؤ	التدريب	التنبؤ	التدريب	التنبؤ	التدريب	التنبؤ
S&P500	49%	21%	-0.37%	172.10%	8.16%	206.19%	47.63%	274.14%
داونجونز	51.3%	42.8%	2.45%	-11.45%	6.90%	33.50%	33%	53%
نازداك	50.5%	50.0%	13.45%	36.84%	9.98%	15.78%	58.5%	36.8%
نايكي	48.8%	46.6%	13.31%	47.32%	11.12%	60.94%	57.5%	47.5%
هانغ سينغ	50.9%	47.4%	6.86%	37.25%	4.21%	26.69%	38.1%	37.2%
شانغ هاي	47.6%	29.6%	-2.29%	-3.48%	12.29%	117.20%	49.7%	141.3%
دبي	54.6%	44.4%	5.71%	12.51%	3.90%	12.51%	34.3%	62.5%
أبو ظبي	54.9%	52.9%	11.41%	0.08%	2.09%	0.08%	25.3%	33.4%
السعودي	56.6%	60.8%	14.86%	-31.48%	1.61%	-38.32%	12.95%	2.80%
عمان	54.6%	73.9%	12.98%	-35.29%	-0.04%	-52.94%	27.40%	0.01%
مصر	53.8%	75.0%	12.14%	-30.67%	3.61%	-20.00%	22.47%	1.33%
قطر	55.5%	44.4%	13.18%	0.00%	0.90%	12.51%	44.52%	62.51%
العراق	54.4%	85.8%	-1.04%	-72.81%	-0.74%	-18.41%	32.80%	0.23%
دمشق	63.6%	50%	-1.04%	-72.87%	-0.74%	-18.60%	32.80%	46.66%
المتوسط			**9%**	**9%**	**5%**	**24.08%**	**35%**	**57%**
أعلى قيمة			**29%**	**172%**	**12%**	**206%**	**58%**	**274%**
أقل قيمة			**-2%**	**-73%**	**-1%**	**-53%**	**5%**	**0.01%**
معامل الاختلاف			**67.0%**	**644.1%**	**93.4%**	**271.0%**	**36.7%**	**122.4%**

المصدر: من إعداد الباحث باستخدام برنامج Microsoft Excel

هذا المعيار أعطى بُعد أوسع في المقارنة ووقف فعلياً على ما يمكن الاستفادة منه في حال تمَّ الاعتماد على النموذج التنبؤي وماهي نسبة الاستفادة من هذا النموذج والمعبر عنها بالقيمة المضافة لنموذج التنبؤ باتجاه الحركة في السوق المالي. فمن خلال الجدول السابق رقم (2) نلاحظ أن الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات يمكن الاعتماد عليها وفق كل المقاييس المطبقة، وكانت الأفضل مقارنة بالنماذج القياسية ومقارنة بنظيرتها قبل تحديد الحجم الأمثل للبيانات وذلك للأسباب التالية:

- ✓ نتائج الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات كانت جميعها ذات قيمة مضافة إيجابية وبفارق ملحوظ بالمتوسط 35% كقيمة مضافة إيجابية لمرحلة التدريب و 57% كقيمة مضافة إيجابية في مراحل التنبؤ، مقارنة بالخوارزمية نفسها بدون تحديد الحجم الأمثل للبيانات، وهي أيضاً أفضل من النماذج القياسية ARIMA (G)ARCH في التقاطها لأنماط الحركة التي لم تتجاوز قيمتها المضافة بالمتوسط 9% في كلا المرحلتين (البناء والتنبؤ).

- ✓ كما تشير النتائج على أنه في جميع الأسواق المدروسة فإن أعلى قيمة مضافة حققتها الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل كانت 274% مقارنة بقيمة مضافة لنفس الخوارزمية بدون تحديد الحجم الأمثل للبيانات والتي بلغت اعلى قيمة مضافة لها 206%
- ✓ إن الخوارزمية المقترحة سواء بتحديد الحجم الأمثل أو بدون تحديد الحجم الأمثل للبيانات كانت أفضل من النماذج القياسية وذلك لسببين: الأول أن أعلى قيمة مضافة حققتها الخوارزمية المقترحة بشقيها هي اعلى من أفضل قيمة مضافة تم تحقيقها باستخدام النماذج القياسية ARIMA (G)ARCH . والسبب الثاني أن أقل قيمة مضافة للخوارزمية المقترحة كانت أفضل بكثير من اقل قيمة مضافة للنماذج القياسية ARIMA (G)ARCH، وهذا يدعم قدرة الخوارزمية المقترحة وتفوقها على النماذج القياسية المطبقة في التقاط اتجاه الحركة في السوق المالي.
- ✓ كما أن الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات هي الأفضل بكل المقاييس حيث أقل قيمة مضافة كانت 0.01% وهي موجبة ولم تكن سالبة، وهذا مفاده أن الخوارزمية المقترحة المعتمدة على تحديد حجم البيانات الأمثل لم تعطي نتائج سلبية وإنما في أسوأ الأحول كانت تعطي قيمة مضافة ولم تؤدي دور سلبي كما هو عليه في النماذج القياسية أو الخوارزمية دون تحديد الحجم الأمثل.
- ✓ إن تحديد الحجم الأمثل للبيانات باستخدام هذه الخوارزمية المقترحة أثبت قدرة أعلى للتنبؤ باتجاه الحركة في جميع الأسواق المدروسة وهذا ما يثبت فعالية تحديد الحجم الأمثل للبيانات بهدف النمذجة والتنبؤ، حيث كان معامل الاختلاف هو الأقل سواء بمرحلة التنبؤ أو بمرحلة بناء(تدريب) النماذج والخوارزميات المقترحة.

5.8. اختبار الفرضيات:

<u>الفرضية الأولى:</u> وفق معيار القيمة المضافة لنموذج التنبؤ باتجاه الحركة في السوق المالي: لا تعطي الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات قيمة مضافة أفضل من القيمة المضافة التي تعطيها كل من الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات وكذلك التي تعطيها النماذج القياسية ARIMA(G)ARCH

بناءً على نتائج الجدول رقم (2) والتحليل السابق له يمكن إثبات صحة الفرضية البديلة للفرضية الأولى وبالتالي فإن: **الخوارزمية المقترحة بعد تحديد الحجم الأمثل للبيانات تعطي قيمة مضافة أفضل من القيمة المضافة التي تعطيها كل من الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات وكذلك التي تعطيها النماذج القياسية ARIMA(G)ARCH.**

<u>الفرضية الثانية:</u> لا يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقة التنبؤ بالاعتماد على نماذج –ARIMA ARCH

تم اختبار وجود فروق معنوية بين متوسطات دقة التنبؤ باستخدام تحليل **Independent Samples Test** وذلك كما هو موضح في الجدول (3) التالي:

الجدول رقم (3):

Group Statistics

	العينة	N	Mean	Std. Deviation	Std. Error Mean
التنبؤ	الخوارزمية بعد تحديد الحجم الأمثل	14	.7175	.05992	.01601
	النماذج القياسية	14	.4939	.13602	.03635

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
التنبؤ	Equal variances assumed	6.071	.021	5.629	26	.000	.22360	.03972	.14195	.30525
	Equal variances not assumed			5.629	17.863	.000	.22360	.03972	.14010	.30710

المصدر: من اعداد الباحث اعتماداً على برنامج SPSS

من خلال الجدول السابق رقم (3) نجد أن متوسط دقة التنبؤ للخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات للأربع عشرة سواقاً مختلفة كان 71.75% وهو أكبر من متوسط دقة التنبؤ باتجاه الحركة اعتماداً على النماذج القياسية ARIMA(G)ARCH والتي بلغت 49.39%، ومن خلال اختبار Independent Samples Test نجد أن قيمة SIG=0.000<0.05 وبالتالي فإن هذا الفارق معنوي وذو دلالة إحصائية وهذا مفاده أن الخوارزمية المقترحة كانت أفضل من النماذج القياسية المطبقة وهذا الفارق هو دال احصائياً، وعليه يمكن إثبات صحة الفرضية البديلة للفرضية الثانية والقول بأنه: **يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقة التنبؤ بالاعتماد على نماذج ARIMA–ARCH**

الفرضية الثالثة: لا يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقتها بالاعتماد على تحديد الحجم الأمثل للبيانات.

تم اختبار وجود فروق معنوية بين متوسطات دقة التنبؤ باستخدام تحليل **Independent Samples Test** وذلك كما هو موضح في الجدول (4) التالي:

الجدول رقم (4):

Group Statistics

	العينة	N	Mean	Std. Deviation	Std. Error Mean
التنبؤ	الخوارزمية بعد تحديد الحجم الأمثل	14	.7175	.05992	.01601
	الخوارزمية قبل تحديد الحجم الأمثل	14	.5599	.11237	.03003

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
التنبؤ	Equal variances assumed	4.385	.046	4.630	26	.000	.15757	.03403	.08761	.22753
	Equal variances not assumed			4.630	19.840	.000	.15757	.03403	.08654	.22860

المصدر: من اعداد الباحث اعتماداً على برنامج SPSS

من خلال الجدول السابق رقم (4) نجد أن متوسط دقة التنبؤ للخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات للأربعة عشر سواقاً مختلفاً كان 71.75% وهو أكبر من متوسط دقة التنبؤ باتجاه الحركة اعتماداً على الخوارزمية دون تحديد الحجم الأمثل للبيانات والتي بلغت 55.99%، ومن خلال اختبار Independent Samples Test نجد أن قيمة SIG=0.000<0.05 وبالتالي فإن هذا الفارق معنوي وذو دلالة إحصائية وهذا مفاده أن الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات كانت الأفضل وهذا الفارق هو دال احصائياً، وعليه فإن لتحديد الحجم الأمثل للبيانات أثر معنوي ويعزز من قدرة الخوارزمية المقترحة، وبالتالي يمكن إثبات صحة الفرضية البديلة للفرضية الثالثة والقول بأنه: **يوجد فروق ذات دلالة معنوية بين متوسط دقة التنبؤ باتجاه الحركة بالاعتماد على الخوارزمية المقترحة بدون تحديد الحجم الأمثل للبيانات مقارنة بمتوسط دقتها بالاعتماد على تحديد الحجم الأمثل للبيانات.**

9. النتائج: يمكن تلخيص أهم النتائج التي توصلت إليها هذه الدراسة فيما يلي:

- ✓ إن نتائج الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات كانت جميعها ذات قيمة مضافة إيجابية وبفارق ملحوظ بالمتوسط 35% كقيمة مضافة إيجابية لمرحة التدريب و 57% كقيمة مضافة إيجابية في مراحل التنبؤ، مقارنة بالخوارزمية نفسها بدون تحديد الحجم الأمثل للبيانات، وهي أيضاً أفضل من النماذج القياسية ARIMA (G)ARCH في التقاطها لأنماط الحركة التي لم تتجاوز قيمتها المضافة بالمتوسط 9% في كلا المرحلتين (البناء والتنبؤ).

- ✓ إن الخوارزمية المقترحة سواء بتحديد الحجم الأمثل أو بدون تحديد الحجم الأمثل للبيانات كانت أفضل من النماذج القياسية وذلك لسببين: الأول أن أعلى قيمة مضافة حققتها الخوارزمية المقترحة بشقيها هي اعلى من أفضل قيمة مضافة تم تحقيقها باستخدام النماذج القياسية ARIMA (G)ARCH . والسبب الثاني أن أقل قيمة مضافة للخوارزمية المقترحة كانت أفضل بكثير من اقل قيمة مضافة للنماذج القياسية ARIMA (G)ARCH، وهذا يدعم قدرة الخوارزمية المقترحة وتفوقها على النماذج القياسية المطبقة في التقاط اتجاه الحركة في السوق المالي.
- ✓ كما أن الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات هي الأفضل بكل المقاييس حيث أقل قيمة مضافة كانت 0.01% وهي موجبة ولم تكن سالبة، وهذا مفاده أن الخوارزمية المقترحة المعتمدة على تحديد حجم البيانات الأمثل لم تعطي نتائج سلبية وإنما في أسوأ الأحول كانت تعطي قيمة مضافة ولم تؤدي دور سلبي كما هو عليه في النماذج القياسية أو الخوارزمية دون تحديد الحجم الأمثل.
- ✓ إن تحديد الحجم الأمثل للبيانات باستخدام هذه الخوارزمية المقترحة أثبت قدرة أعلى للتنبؤ باتجاه الحركة في جميع الأسواق المدروسة وهذا ما يثبت فعالية تحديد الحجم الأمثل للبيانات بهدف النمذجة والتنبؤ، حيث كان معامل الاختلاف هو الأقل سواء بمرحلة التنبؤ أو بمرحلة بناء(تدريب) النماذج والخوارزميات المقترحة
- ✓ تم التحقق من وجود فروق ذات دلالة معنوية إحصائية تدعم وتؤكد النتائج السابقة وتثبت تفوق الخوارزمية المقترحة المعتمدة على الحجم الأمثل للبيانات وإمكانية اعتمادها كأسلوب مساعد للتنبؤ باتجاه الحركة في السوق المالي.

10. التوصيات: في ضوء النتائج التي توصل إليها البحث يمكن ذكر مجموعة من التوصيات:

- ✓ يُوصى بتطبيق الخوارزمية المقترحة المعتمدة على تحديد الحجم الأمثل للبيانات في تحليل الأسواق المالية، نظرًا لفعاليتها العالية حيث يمكن استخدامها كأداة مساعدة لاتخاذ القرارات الاستثمارية أو تحسين استراتيجيات التداول.
- ✓ التأكيد على أهمية تحسين حجم البيانات في النمذجة التنبؤية، واقتراح دمج هذه الآلية في الخوارزميات القائمة أو الجديدة، لا سيما في المجالات التي تتسم البيانات فيها بالديناميكية والتعقيد، مثل الأسواق المالية أو سلاسل توريد الطاقة.
- ✓ تصميم أدوات وواجهات برمجية سهلة الاستتعمال تعتمد على الخوارزمية المقترحة، لتسهيل وصول المحللين والمستثمرين غير المتخصصين إلى تنبؤات دقيقة
- ✓ إجراء دراسات متابعة لتقييم استقرار أداء الخوارزمية خلال فترات التقلبات الحادة (كالأزمات الاقتصادية)، وضمان تحديثها باستمرار لمواكبة تغيرات السوق

المراجع:

أولاً: المراجع العربية:

- الحميد، أيهم .(2020) *الكفاءة السعرية والتنبؤ بمؤشر السوق المالي باستخدام النماذج القياسية والشبكات العصبونية الاصطناعية* (دراسة مقارنة بين سوق دمشق للأوراق المالية وبعض أسواق المال العربية)، أطروحة دكتوراه، جامعة حماة، سوريا.

- Asare, C., Asante, D., & Essel, J. F.(2023). Probabilistic LSTM Modeling for Stock Price Prediction with Monte Carlo dropout Long Short-Term Memory Network. International Journal of Innovative Science and Research Technology. Volume 8, Issue 7, July – 2023. ISSN No:-2456-2165.
- Babichev, D., & Bach, F. (2018). Constant step size stochastic gradient descent for probabilistic modeling. *arXiv preprint arXiv:1804.05567*.
- Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to Probability*. Athena Scientific.P:45.

- Bou-Hamad, I., & Jamali, I. (2020). Forecasting financial time-series using data mining models: A simulation study. ***Research in International Business and Finance***, 51, 101072
- Bousbaa, Z., Sanchez-Medina, J., & Bencharef, O. (2023). Financial time series forecasting: a data stream mining-based system. Electronics,MDPI, 12(9), 2039.
- Bousbaa, Z., Sanchez-Medina, J., & Bencharef, O. (2023). Financial time series forecasting: a data stream mining-based system. *Electronics*,MDPI, *12*(9), 2039.
- Dar, G. F., Padi, T. R., & Rekha, S. (2022). Stock price prediction using a Markov chain Model: a study for TCS Share Values. *Advances and Applications in Statistics*, *80*, 83-101.
- Evans, M. J., & Rosenthal, J. S. (2004). *Probability and Statistics: The Science of Uncertainty*. Springer.P12.
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and. *Techniques, Waltham: Morgan Kaufmann Publishers*.(PP:3-12).
- Luo, J. (2022, January). Modeling of data mining technology in financial data recognition mining and forecasting. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1168-1171). IEEE.
- Monika, P., Ruchjana, B., & Abdulla, A. (2022). The implementation of the ARIMA-ARCH model using data mining for forecasting rainfall in Bandung city. *International Journal of Data and Network Science*, *6*(4), 1309-1318.
- Nugraha, E. S., & Alvina, C. (2024). THE APPLICATION OF STANDARD GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSCEDASTICITY (SGARCH) MODEL IN FORECASTING THE STOCK PRICE OF BARITO PACIFIC. *BAREKENG Journal of Mathematics and Its Applications*, *18*(2), 0849-0862.
- Provost, F. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking* (Vol. 355). O'Reilly Media, Inc.
- Rojas, F., Leiva, V., Huerta, M., & Martin-Barreiro, C. (2021). Lot-size models with uncertain demand considering its skewness/kurtosis and stochastic programming applied to hospital pharmacy with sensor-related COVID-19 data. *Sensors*, *21*(15), 5198.
- Ross, S. M. (2010). *Introduction to Probability Models* (10th ed.). Academic Press.PP:3-5.
- Yahaya, A., Etuk, E., & Emeka, A. (2021). Comparative performance of arima and garch model in forecasting crude oil price data. *Asian Journal of Probability and Statistics*, *15*(4), 251-275.