

استخدام خوارزمية XGBoost لتنبؤ بالتضخم وأسعار الغذاء العالمية:

تحليل زمني حسب المنتج السوق والدولة

مجدي رمضان يوسف المتخطري¹ محمد مصطفى عبد العالي عبد الجواد²المعهد العالي للعلوم والتقنية /أ مساعد/ ليبيا¹⁻²Astef_li@yahoo.com²Magdibosif@Gmail.com¹ البريد الالكتروني

المخلص

تهدف هذه الدراسة إلى استخدام خوارزمية XGBoost لتنبؤ بالتضخم وأسعار الغذاء العالمية، بالاعتماد على البيانات التي تم جمعها من قاعدة بيانات Kaggle. تم تقييم النموذج باستخدام مجموعة من المقاييس البصرية والأداء، والتي تم تحليلها بشكل شامل في هذه الدراسة.

تم استخدام مجموعة متنوعة من البيانات لتدريب النموذج وتقييمه، بهدف تحليل البيانات الزمنية وتحديد الأنماط والعلاقات بين المتغيرات المختلفة. النتائج أظهرت دقة عالية للنموذج في التنبؤات، حيث كانت معظم النقاط قريبة من الخط الأحمر المثالي. تم تقييم أداء النموذج باستخدام المعايير التالية:

(MSE) 12.8% خطأ الجذر المتوسط التربيعي 35.7% (RMSE) متوسط الخطأ المطلق

(MAE) 5.4% معامل التحديد 99.5% (R^2)

تعكس هذه النتائج كفاءة النموذج في تقدير أسعار الغذاء والتضخم، مما يساهم في تقديم توصيات موثوقة وفعالة لصناع القرار. يمكن تحسين أداء النموذج بإجراء دراسات إضافية لاستكشاف معماريات مختلفة وتحسين دقة النموذج بشكل أكبر.

الكلمات المفتاحية: XGBoost، تنبؤ بالتضخم، أسعار الغذاء، تحليل البيانات الزمنية، التوصيات الاقتصادية.

"Using the XGBoost Algorithm to Forecast Inflation and Global Food Prices: A Time-Lapse Analysis by Product, Market, and Country"

Abstract

This study aims to utilize the XGboost algorithm to predict global inflation and food prices, based on data collected from the Kaggle database. The model was evaluated using a set of visual and performance metrics, which were comprehensively analyzed in this

study. A diverse dataset was used to train and evaluate the model, with the goal of analyzing temporal data to identify patterns and relationships between different variables. The results demonstrated high accuracy for the model in predictions, with most points being close to the ideal red line. The model's performance was evaluated using the following metrics:

- (MSE): 12.8%
- (RMSE): 35.7%
- (MAE): 5.4%
- R^2 Score: 99.5%

These results reflect the model's efficiency in estimating food prices and inflation, providing reliable and effective recommendations for decision-makers. The model's performance can be further enhanced by conducting additional studies to explore different architectures and improve its accuracy.

Keywords: XGboost, inflation prediction, food prices, time series analysis, economic recommendations.

1. المقدمة

شهدت أسعار الغذاء العالمية تقلبات كبيرة في العقود الأخيرة، مما أثر بشكل كبير على الاقتصاد العالمي ورفاهية الشعوب. تعد تقلبات أسعار الغذاء أحد العوامل الرئيسية التي تؤثر على معدلات التضخم والتوازن الاقتصادي في مختلف الدول. في هذا السياق، أصبح من الضروري توفير أدوات ونماذج تنبؤية دقيقة تساعد صناع القرار في التنبؤ بالتغيرات المستقبلية واتخاذ الإجراءات المناسبة. تعتمد الدراسات التقليدية على الأساليب الإحصائية لتحليل وتنبؤ أسعار الغذاء والتضخم. ومع ذلك، قد تفقر هذه الأساليب إلى الدقة في بعض الأحيان بسبب تعقيد البيانات والتفاعلات المتعددة بين العوامل الاقتصادية المختلفة. هنا تأتي أهمية استخدام خوارزمية XGboost، هي اختصار لـ "Extreme Gradient Boosting" وهي إحدى خوارزميات التعلم الآلي القوية التي تستخدم لتعزيز الأداء في مهام التصنيف والتنبؤ. تُعتبر واحدة من أبرز تقنيات التعلم الآلي المتقدمة، وتتميز بقدرتها العالية على تحليل البيانات المعقدة والتنبؤ بدقة. تُتيح هذه الخوارزمية لصناع القرار تحسين دقة التنبؤات واتخاذ قرارات اقتصادية أكثر فعالية. (Shively, 2001)

تعاني الاقتصادات العالمية في الفترة الحالية من ضغوط كبيرة على صعيد تحقيق النمو والاستقرار المالي، وذلك نتيجة للارتفاع الحاد في أسعار الطاقة وأزمة الغذاء المستمرة. بعد التعافي من الأزمة المالية العالمية، يبدو أننا نواجه أزمة جديدة ناتجة عن الضغوط التضخمية العالمية. فقد شهد سعر النفط ارتفاعاً ملحوظاً إلى مستويات غير متناسبة مع طبيعة المرحلة الحالية، وعادت أسعار الغذاء العالمية إلى مستويات مرتفعة تشابه تلك التي كانت قبل الأزمة المالية العالمية. **FSIN (Food Security Information Network) 2022**

تعاني دول العالم من ارتفاع معدلات التضخم، ولم يقتصر هذا الارتفاع على الدول النامية فقط، بل شمل أيضاً الدول الصناعية ذات البنوك المركزية المتطورة التي تستهدف التضخم كهدف نهائي للسياسة النقدية. يشير هذا إلى أن التضخم أصبح ظاهرة عالمية وليست محلية أو إقليمية، مما يدل على نهاية عصر الغذاء الرخيص مع ارتفاع تكلفة إنتاج السلع الغذائية. اعتاد العالم على أسعار الغذاء المنخفضة في الثمانينيات والتسعينيات، ومنذ عام 2006، زاد تأثير أسعار الغذاء في الضغوط التضخمية العالمية. أصبحت أسعار الغذاء مؤشراً قوياً للضغوط التضخمية، خصوصاً في الدول النامية حيث يمثل الإنفاق على الغذاء نسبة كبيرة من إجمالي الإنفاق الاستهلاكي. يصبح السؤال المهم الآن: إلى أي مدى يستطيع صناع سياسات الأمن الغذائي، وخاصة في الاقتصادات الصاعدة والنامية، الاستمرار في مواجهة التحديات الناتجة عن ارتفاع أسعار الغذاء وتقلبها. (FAOSTAT, 2022).

تواجه هذه الدراسة العديد من التحديات المحتملة، من بينها التحدي في جمع بيانات كاملة ودقيقة عن أسعار الغذاء والتضخم حسب المنتج والسوق والدولة. كما أن التعامل مع البيانات الزمنية والاقتصادية المتعددة المتغيرات يتطلب إجراءات معالجة وتنظيف متقدمة. إلى جانب ذلك، تحديد النموذج الأمثل وضبط المعلمات لتحقيق أعلى دقة في التنبؤ يشكل تحدياً كبيراً. بالإضافة إلى التغيرات غير المتوقعة في الاقتصاد العالمي التي قد تؤثر على دقة التنبؤات وتزيد من مستوى عدم اليقين. وأخيراً، يواجه تطبيق النتائج والنماذج في السياسات الاقتصادية وصنع القرار تحديات تنظيمية وإدارية. (Minot, 2011). ركزت هذه الدراسة على تطوير خوارزمية XGboost لتحليل وتنبؤ بأسعار الغذاء العالمية والتضخم حسب المنتج، السوق، والدولة. تم استخدام خوارزمية XGboost لتقدير أسعار الغذاء والتضخم، وتحليل

البيانات الزمنية لتحديد الأنماط والعلاقات بين المتغيرات المختلفة. كما تم تقديم توصيات تستند إلى نتائج النموذج لمساعدة صناع القرار في اتخاذ الإجراءات المناسبة

2. الادبيات النظرية

2.1 مقدمة إلى نماذج التعلم الآلي (XGboost)

تقنية XGboost، التي تعني تعزيز التدرج الفائق، تمثل التنفيذ الأمثل لتقنية تعزيز التدرج. لكن ماذا يعني ذلك بالضبط؟ في جوهرها، تعد XGboost مكتبة مصممة لزيادة سرعة وكفاءة خوارزميات التعلم الآلي، وتعتبر فعالة بشكل خاص عند التعامل مع مجموعات البيانات الكبيرة والمشكلات المعقدة. تم تطوير هذه الأداة لتكون عالية الكفاءة، مرنة، وقابلة للنقل.

لقد برزت XGboost في العديد من مسابقات تنبؤ البيانات، بما في ذلك المسابقات الشهيرة على منصة Kaggle، حيث حققت النماذج المستندة إلى XGboost نتائج متميزة. العامل الرئيسي الذي يميزها عن تطبيقات التعزيز الأخرى هو قدرتها على الاستفادة الكاملة من أجهزة الحوسبة الحديثة. وهذا يعني أنها تستطيع معالجة البيانات على نطاق واسع بطريقة لا تستطيع الخوارزميات الأخرى مجاراتها، مع الاحتفاظ بسرعة ودقة عالية نسبياً.

"Story and Lessons behind the evolution of XGboost" 2016 "

2.2 أهمية ودور التنبؤ بأسعار الغذاء والتضخم

يُعد التضخم من القضايا ذات التأثير الكبير على حياة وتطور المجتمعات البشرية كافة. ولا يقتصر هذا التأثير على التآكل الذي يلحق بالدخول الحقيقية للأفراد بسبب الزيادات المضطردة في المستوى العام للأسعار، بل يتجاوز ذلك، كما يؤكد Barro (2013)، إلى تثبيط معدلات النمو الاقتصادي على المدى الطويل. وتشير العديد من الدراسات إلى أن الدول الأقل نموًا قد تتأثر بهذه الظاهرة بشكل أكبر وأكثر وضوحًا (Arize & Darrat, 1990).

1.2.2 تأثير تقلبات الأسعار على الاقتصاد العالمي

تُعد الدول المستوردة للغذاء وأدوات وتكنولوجيا الصناعة والسلع الصناعية الأكثر تأثرًا بتقلبات الأسعار العالمية. نظرًا لكون الدولار العملة السائدة في التجارة العالمية، فإن تقلب قيمته مقابل العملات المحلية يمثل تحديًا كبيرًا لسياسة الاستقرار الاقتصادي.

في ظل غياب الاستقرار، يصبح من الصعب اتخاذ القرارات الاقتصادية الهامة، التنبؤ بالمشكلات، التخطيط ووضع الاستراتيجيات. يؤثر هذا الوضع على تدفق الاستثمارات الخارجية، الميزان التجاري، التصنيف الائتماني، ومعدل النمو الاقتصادي وحركة التجارة الخارجية للدول. إذا كانت هذه الدول تسعى لتحقيق التكامل الاقتصادي، فمن الضروري إجراء دراسة موحدة ككيان موحد لتفهم التحديات المشتركة، مما يساهم في وضع استراتيجيات لتحقيق الاستقرار الاقتصادي والتكامل المثمر. (اسماء فهمي، 2023).

2.3 الدراسات السابقة

1.2.3 الدراسات المتعلقة باستخدام نماذج التعلم الآلي للتنبؤ بالأسعار

تعتمد مسألة التنبؤ اعتماداً كبيراً على جودة النماذج المستخدمة في هذه العملية وملاءمتها لطبيعة البيانات الممثلة للظاهرة محل الاهتمام. ومن خلال مراجعة الأدب الاقتصادي التجريبي المتعلق بالتضخم، يلاحظ أن التنبؤ بمستويات الأسعار كان موضوعاً للعديد من الدراسات السابقة التي تنوعت بتنوع الطرائق المستخدمة في القياس والنماذج المستعملة في التنبؤ.

الدراسات السابقة هي كالتالي:

قام Walled & Huselius (2014) بدراسة للتنبؤ بمعدلات التضخم في السويد باستخدام نماذج التنبؤ ARIMA و Exponential Smoothing مع نموذج State Space. وجدت الدراسة أن العديد من النماذج يمكن أن تمثل البيانات المستخدمة تمثيلاً جيداً وتُمكن من إجراء تنبؤات دقيقة. ومع ذلك، تبين أن نماذج State Space تعمل بشكل أفضل.

قام Jere & Siyanga (2016): بدراسة هدفت إلى التنبؤ بمعدلات التضخم في زامبيا باستخدام نماذج Holt's Exponential Smoothing و ARIMA. وجدت الدراسة أن نموذج ARIMA (12,1,0) هو أفضل النماذج لتمثيل البيانات الخاصة بالرقم القياسي لأسعار المستهلك (CPI). كما استخدمت الدراسة طريقة Holt's Double Exponential Smoothing نظراً لعدم وجود مركبة موسمية في السلاسل الزمنية ولتغير معدلات النمو عبر الزمن.

ركزت الدراسة التي قام بها Lidiema (2017) على التنبؤ بمستويات التضخم في كينيا باستخدام نماذج SARIMA و Holt-Winters Triple Exponential Smoothing. وجدت الدراسة أن

نموذج SARIMA كان أفضل من نموذج Holt-Winters Triple Exponential Smoothing

بناءً على نتائج معايير MAPE، MASE، MAE.

دراسة أخرى هدفت إلى تصميم نماذج هجينة من خلال الدمج بين نموذج الانحدار المتعدد ونموذج الشبكات العصبونية للتنبؤ بأثر أسعار حوامل الطاقة على تكاليف المعيشة الغذائية للأسرة السورية. يُعد نموذج الشبكات العصبونية من الأساليب القادرة على التعلم والتكيف مع أي نموذج ولا يحتاج إلى افتراضات لطبيعة السلسلة الزمنية. قام الباحث باستخدام بيانات الأسعار الشهرية للمازوت، الغاز، البنزين، وتكلفة المعيشة الغذائية الشهرية المحسوبة للأسرة السورية للفترة الممتدة من يناير 2020 إلى مايو 2023. من خلال معالجة البيانات باستخدام البرنامج الإحصائي SPSS، توصل الباحث إلى بناء أربعة نماذج هجينة بين الأسلوبين.

عند المقارنة بين النماذج الهجينة باستخدام معايير التقييم MSE، RMSE، MAE، MAPE، توصل الباحث إلى تفوق النموذج الهجين C باستخدام المشاهدات والبواقي السابقة والقيم المقدرة لنموذج الانحدار الخطي المتعدد مقارنة مع النماذج الهجينة الأخرى.

بالاعتماد على النموذج الهجين C، تم التنبؤ بتكاليف المعيشة الغذائية الشهرية للأسرة السورية لمدة 12 شهراً قادمة من يونيو 2023 إلى مايو 2024. أظهرت النتائج ارتفاع تكاليف المعيشة الغذائية الشهرية للأسرة السورية في الفترات المقبلة. Nassra, Z., & Haidar, A. (2024).

Sekine (2001) في اليابان، استخدم الباحث نموذج تصحيح التوازن (EqCM) للتنبؤ بالتضخم. وجد أن عرض النقود وفجوة الناتج (Output Gap) هما أهم محددات التضخم في اليابان على المدى الطويل. كما أظهرت الدراسة أن أداء نموذج EqCM في التنبؤ بالتضخم في اليابان كان جيداً، وتوصلت أيضاً إلى إمكانية تحسين أداء هذا النموذج بدمجه مع النماذج الأخرى المنافسة.

دراسة في تونس عام (2015) قام بها Bilel et al، استخدم الباحثون نموذج العوامل الديناميكية (DFM) للتنبؤ بالتضخم. تم تطبيق النموذج على البيانات الاقتصادية وتوصلت الدراسة إلى نتائج مهمة حول قدرة النموذج على التنبؤ بمعدلات التضخم.

هذه الدراسات تسلط الضوء على أهمية استخدام النماذج المناسبة لتحقيق توقعات دقيقة وفعالة للتضخم في مختلف السياقات الاقتصادية.

3. منهجية الدراسة

تم تصميم هذه الدراسة باستخدام منهجية الدراسة التحليلية بهدف التنبؤ بأسعار الغذاء العالمية والتضخم حسب المنتج والسوق والدولة، باستخدام نماذج التعلم الآلي. تم اختيار هذه المنهجية نظراً لأهميتها في تحليل البيانات واستخلاص الأنماط والعلاقات بين المتغيرات. بدأت الدراسة بجمع البيانات من قاعدة بيانات Kaggle، التي تحتوي على سجلات شاملة لعدد من الدول. تم إخضاع البيانات لعمليات تنقية وتحليل دقيقة لإزالة القيم المفقودة والمتطرفة وأخطاء الإدخال، مما نتج عنها 4798 سجلاً صالحاً للتحليل. ركز التحليل على عدد من المتغيرات وهي: الفتح، أعلى، أدنى، إغلاق، التضخم، الدولة، التاريخ، و ISO3 رمز الدولة.

بالإضافة إلى ذلك، تم استخدام منصة Google Colab ولغة البرمجة بايثون لتحليل البيانات، مما يساهم في تحسين دقة التنبؤات واستخلاص توصيات فعالة. هذه المنهجية تتيح تحليل الأنماط والعلاقات بين المتغيرات بشكل دقيق، مما يمكن من تقديم توصيات تساعد صناع القرار في اتخاذ الإجراءات المناسبة.

3.1 تطبيق الأدوات البرمجية في تحليل بيانات أسعار الغذاء العالمية والتضخم

لإتمام العمل على تحليل بيانات أسعار الغذاء والتضخم وبناء نماذج التنبؤ، تم استخدام عدة مكتبات برمجية في لغة البرمجة بايثون. وفيما يلي تفصيل لكل مكتبة واستخداماتها في هذا النظام: XGboost: مكتبة تُستخدم لإنشاء نماذج التعلم الآلي عالية الأداء باستخدام تقنية تعزيز التدرج. تتيح للمستخدمين تنفيذ نماذج التعلم الآلي بكفاءة وسرعة.

NumPy: مكتبة تُستخدم للعمليات الحسابية والمعالجة الرياضية، خاصة عند التعامل مع المصفوفات والمتجهات.

Pandas: مكتبة تُستخدم لمعالجة وتحليل البيانات. تُسهل العمل مع البيانات المجدولة (data frames) وتتيح إجراء العمليات المختلفة عليها.

Plot Tree: دالة من مكتبة XGboost تُستخدم لتصوير الأشجار النموذجية.

Matplotlib Pyplot: مكتبة تُستخدم لإنشاء الرسوم البيانية والتصورات البيانية.

Sklearn Preprocessing MinMaxScaler: تُستخدم لتقييس البيانات وتحويلها إلى نطاق معين، عادة بين 0 و 1.

XGboost XGB Regressor: نموذج الانحدار من مكتبة XGboost يُستخدم لتدريب النماذج التنبؤية.

Sklearn Model Selection Train Test Split: تُستخدم لتقسيم البيانات إلى مجموعات التدريب والاختبار.

Sklearn model selection Grid Search CV: تُستخدم للبحث عن أفضل مجموعة من المعلمات لتحسين أداء النموذج.

Sklearn Metrics: تُستخدم لتقييم أداء النماذج من خلال مقاييس مختلفة مثل متوسط الخطأ المربع (MSE) ومتوسط الخطأ المطلق (MAE) ومعامل التحديد (R^2).

توفر هذه الأدوات البرمجية إمكانية تحليل البيانات بكفاءة وتطوير نماذج تنبؤية قوية تساهم في تقديم توصيات دقيقة لصناع القرار. إليك الخطوات التي يجب اتباعها لإتمام الجزء العملي لتحليل البيانات باستخدام لغة بايثون:

1. تحميل ومعالجة البيانات

في الجدول التالي تم تحميل مجموعة البيانات باستخدام مكتبة Pandas وقراءة البيانات من ملف CSV يحتوي على معلومات مالية واقتصادية لعدد من الدول عبر سنوات. البيانات تتضمن 4798 سجلاً و 8 أعمدة، وتشمل الأعمدة التالية:

Open: قيمة فتح السوق.

High: أعلى قيمة للسوق خلال الفترة...

Low: أدنى قيمة للسوق خلال الفترة...

Close: قيمة إغلاق السوق.

Inflation: معدل التضخم (مع وجود بعض القيم المفقودة).

Country: اسم البلد.

ISO3: رمز البلد المكون من ثلاثة أحرف.

Date: تاريخ السجل.

عرض البيانات الأولية: تم اختيار مجموعة بيانات عشوائية من إطار البيانات لعرض خمسة صفوف بشكل عشوائي. يهدف هذا العرض إلى إعطاء فكرة عامة عن القيم المخزنة في كل عمود وللتأكد من عرض البيانات بشكل صحيح، كما هو موضح في الجدول أدناه:

جدول رقم 1 عرض للمتغيرات البيانات مستخرجة من مكتبة **Pandas** لتحليل البيانات من مجموعة بيانات

(قاعدة بيانات Kaggle)

التاريخ	IOS3	الدولة	التضخم	أغلاق	أدني	أعلي	فتح
2018-09-01	SOM	Somalia	-5.87	0.97	0.94	0.97	0.95
2017-12-01	BDI	Burundi	3.15	1.05	1.05	1.10	1.08
2017-03-01	CMR	Cameroon	Non	0.98	0.97	0.98	0.98
2011-07-01	GMB	Gambia,	1.43	0.61	0.61	0.61	0.61
2019-09-01	GNB	Guinea	Non	1.12	1.12	1.21	1.17

2. معالجة البيانات المفقودة

في هذا القسم، تمت معالجة القيم المفقودة في بيانات التضخم باستخدام طريقة التعبئة بالصفر. وهذه الخطوة مهمة لضمان عدم وجود قيم ناقصة تؤثر على دقة التنبؤات. نتائج المعالجة بعد معالجة القيم المفقودة، لم يحتوي إطار البيانات على أي قيم مفقودة في الأعمدة المختلفة، مما يضمن جاهزية البيانات للتحليل الإحصائي وبناء النماذج التنبؤية.

تمت معالجة جميع القيم المفقودة في إطار البيانات، وأصبح الإطار جاهزاً للتحليل الإحصائي وبناء النماذج التنبؤية. يضمن هذا النهج أن جميع الأعمدة تحتوي على بيانات كاملة، مما يتيح مزيداً من الدقة في التحليل والنمذجة. كما هو موضح في الجدول رقم 2

جدول رقم 2 معالجة البيانات المفقودة مستخرج من مكتبة (**Pandas** لتحليل البيانات)

DATE	IOS3	COUNTRY	INFLATION	CLOSE	LOW	HIGH	OPEN
0	0	0	0	0	0	0	0

3. فصل الميزات والهدف، إعادة تقييس البيانات

في هذه الخطوة، قمنا بفصل البيانات إلى ميزات مستقلة وهدف، لتجهيز البيانات للتدريب على نموذج التنبؤ. هذه الخطوة أساسية في عملية تحليل البيانات وبناء النماذج التنبؤية، حيث تتيح لنا:

تنظيم البيانات: يساعد في تنظيم البيانات بشكل يسهل على النموذج فهم العلاقات بين المتغيرات المختلفة.

تحسين الدقة: يسمح للنموذج بتحليل كل ميزة بشكل منفصل ومعرفة تأثير كل منها على النتيجة النهائية.

تجهيز البيانات: يوفر إطار عمل واضح لتجهيز البيانات قبل استخدامها في التدريب أو الاختبار. تقيس البيانات تم استخدام مكتبة MinMaxScaler لتقيس البيانات في مجموعة البيانات لدينا، والتي تحتوي على الأعمدة: 'Open' و 'High' و 'Low' و 'Close'. هذه العينة من البيانات المقيسة تظهر أول صف من البيانات بعد التقيس. تقيس البيانات يعني تحويل القيم لتكون في مدى محدد (عادة من 0 إلى 1) لتسهيل عمليات التحليل والنمذجة.

جدول رقم 3 يوضح البيانات بعد التقيس البيانات مستخرجة من مكتبة (MinMaxScaler)

Open	HIGH	LOW	CLOSE
0.00507565	0.00497793	0.00550789	0.07975771

4. تقسيم البيانات إلى تدريب واختبار

عملية تقسيم البيانات إلى مجموعتين، التدريب والاختبار، هي خطوة أساسية في إعداد النماذج التنبؤية. الهدف من هذه العملية هو تدريب النموذج على مجموعة من البيانات (مجموعة التدريب) ومن ثم تقييم أدائه على مجموعة مستقلة من البيانات (مجموعة الاختبار) للتأكد من قدرته على التعميم والتنبؤ بدقة. تقسيم البيانات إلى مجموعات تدريب واختبار يتيح تقييم أداء النموذج بشكل أكثر دقة، حيث يساعد على التحقق من قدرة النموذج على التعميم والتنبؤ بدقة على بيانات لم يشاهدها من قبل. هذا الخطوة الأساسية تعزز من موثوقية النتائج وتقدم رؤى أعمق حول أداء النموذج التنبؤي.

جدول رقم 4 يوضح بيانات التدريب مستخرجة من مكتبة (Sklearn Model)

	0	1	2	3	4
0	0.006540	0.006387	0.006991	0.098462	0.007097
1	0.006540	0.004039	0.004449	0.079758	0.004449
2	0.010054	0.009862	0.010592	0.067263	0.010698
3	0.005466	0.005354	0.005826	0.132296	0.005826
4	0.005857	0.005729	0.006355	0.081430	0.006461

جدول رقم 5 يوضح بيانات الاختبار مستخرجة من مكتبة (Sklearn Model)

	0	1	2	3	4
0	0.007321	0.007232	0.007838	0.052918	0.008050
1	0.008980	0.008641	0.009745	0.081963	0.009745
2	0.009858	0.009768	0.010380	0.057455	0.010698
3	0.008492	0.008547	0.008897	0.066477	0.009639
4	0.008882	0.008735	0.009427	0.147883	0.009427

3.2 إنشاء نموذج XGboost وتدريبه.

لقد قمنا باستخدام نموذج XGboost لتدريب البيانات والتنبؤ بها. نموذج XGboost هو نموذج تعزيز متدرج مشهور يتميز بأدائه العالي وسرعته في التنفيذ.

1. إنشاء نموذج XGboost:

- تم تحديد معدل التعلم ليكون 0.1 للتحكم في مقدار التغيير في وزن كل شجرة جديدة.
- تم تحديد الحد الأقصى لعمق الشجرة ليكون 3 لمنع الإفراط في التعميم والتحكم في التعقيد.
- تم استخدام 100 شجرة في النموذج لتحقيق تنبؤات دقيقة.
- 2. تدريب النموذج: تم استخدام دالة (fit) لتدريب النموذج على مجموعة البيانات التدريبية، حيث تتعلم العملية العلاقات بين الميزات المستقلة والقيم المستهدفة.

4. عرض النتائج

في هذا القسم من الدراسة، قمنا بتطبيق خوارزمية XGboost لتحليل وتنبؤ أسعار الغذاء والتضخم، وتحليل البيانات الزمنية لتحديد الأنماط والعلاقات بين المتغيرات المختلفة. كان الهدف من هذه الدراسة هو تقديم توصيات دقيقة وفعالة لصناع القرار بناءً على نتائج النماذج التنبؤية. لتقييم أداء النموذج، تم استخدام المعايير التالية والنتائج موضحة في الجدول أدناه:

جدول رقم 6 يوضح نتائج المعايير مستخرجة من مكتبة (Sklearn Metrics)

(MSE) مربع متوسط الخطأ	12.8%
(RMSE): خطأ الجذر المتوسط التربيعي	35.7%
(MAE): متوسط الخطأ المطلق	5.4%
(R ²): معامل التحديد	99.5%

1. Mean Squared Error (MSE): يمثل المتوسط الحسابي لمربعات الفروقات بين القيم الفعلية والقيم المتوقعة. كلما كانت هذه القيمة أصغر، كان النموذج أدق.

النتيجة المستخرجة: 0.1276477298779174. هذه القيمة منخفضة، مما يدل على أن النموذج لديه أداء جيد في التنبؤ.

2. Root Mean Squared Error (RMSE): هو الجذر التربيعي يُعتبر مؤشراً على مستوى الخطأ في التنبؤات بوحدة المخرجات الأصلية.

النتيجة المستخرجة: 0.3572782247463696. كونه قريب من 0 يشير إلى أداء جيد للنموذج. 3. Mean Absolute Error (MAE): يمثل المتوسط الحسابي لمطلق الفروقات بين القيم الفعلية والقيم المتوقعة. مثل MSE، كلما كانت القيمة أصغر، كان الأداء أفضل.

النتيجة المستخرجة: 0.05416641249018017. هذه القيمة منخفضة، مما يعني أن النموذج يتنبأ بدقة عالية.

4. R² Score: يمثل نسبة التباين في البيانات التي يمكن تفسيرها بواسطة النموذج. القيمة القصوى هي 1، مما يعني أن النموذج يشرح جميع التباينات في البيانات.

النتيجة المستخرجة: 0.9947063990599773. هذه القيمة مرتفعة جداً وتشير إلى أن النموذج قادر على تفسير معظم التباينات في البيانات.

توضح هذه النتائج دقة النموذج وكفاءته في تقدير أسعار الغذاء والتضخم، مما يساهم في تقديم توصيات موثوقة وفعالة.

4.1 تقييم النموذج باستخدام المعلمات المحسنة

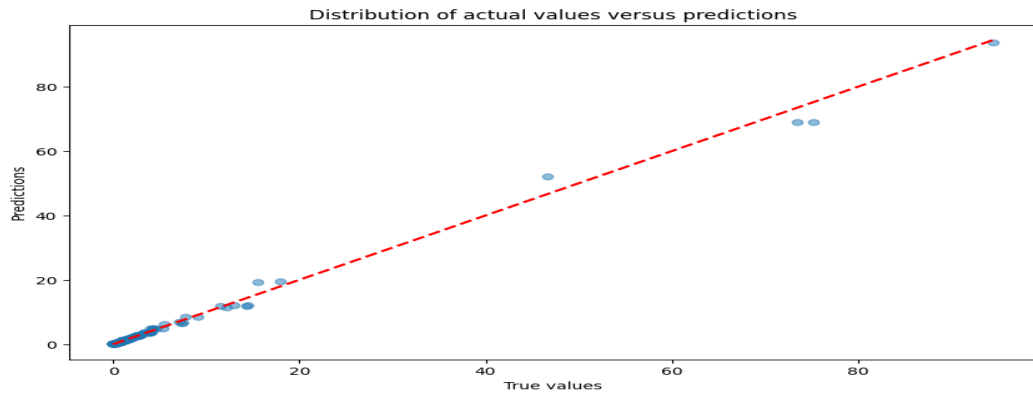
لقد قمنا بتحسين نموذج XGboost باستخدام Grid Search CV لتحديد أفضل المعلمات وتحسين الأداء. حيث أنه أداة قوية تستخدم للبحث عن التوليفات المثلى للمعلمات من خلال إجراء عمليات تدريب متعددة باستخدام مجموعات مختلفة من المعلمات. تم إعداد Grid Search CV باستخدام النموذج (XGb Regressor) ومعلمات البحث المحددة. تم تحديد $cv=3$ لإجراء التحقق المتقاطع بثلاثة أجزاء، وتم استخدام (scoring= 'neg_mean_squared_error') لتقييم الأداء باستخدام متوسط مربع الخطأ العكسي، بعد الانتهاء من تنفيذ Grid Search CV، تم عرض أفضل المعلمات التي تم العثور عليها:

جدول رقم 7 بوضوح تحسين النموذج باستخدام Grid Search CV. مستخرج من مكتبة (Sklearn model)

Learning_rate:	Max_depth	N_estimators
0.1	3	100
Best Mean Squared Error:	0.5691553331782125	

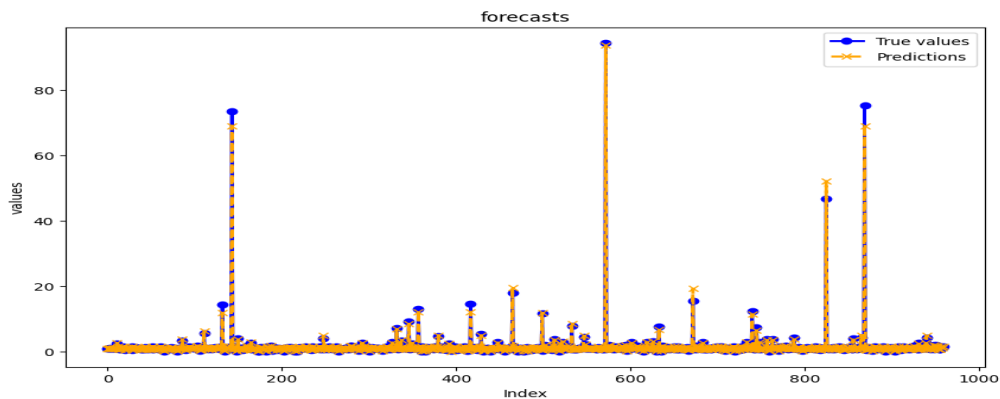
4.2. تقييم أداء النموذج التنبؤي باستخدام المقاييس البصرية

تستخدم الأشكال البصرية لتقييم وتحليل أداء النموذج التنبؤي. توضح الأشكال التالية العلاقة بين القيم الفعلية والتنبؤات، وتُظهر مدى دقة النموذج وتوزيع الأخطاء. توزيع القيم الفعلية مقابل التنبؤات في الشكل 1 "يظهر هذا الشكل العلاقة بين القيم الفعلية والقيم المتوقعة من النموذج. النقاط الزرقاء تمثل القيم الفعلية مقابل القيم المتوقعة، بينما الخط الأحمر يمثل الخط المثالي حيث تكون القيم الفعلية مساوية تمامًا للقيم المتوقعة. النموذج يظهر دقة عالية في التنبؤات كما هو موضح في الشكل ادناه، حيث تكون معظم النقاط قريبة من الخط الأحمر المثالي.



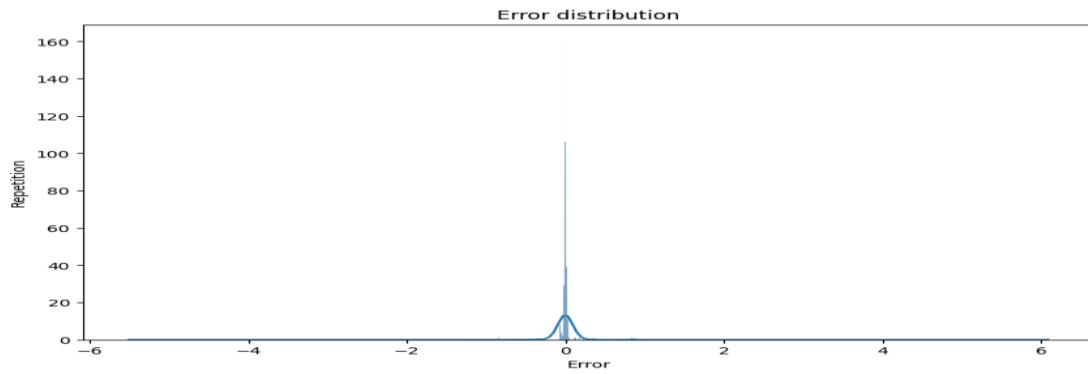
الشكل رقم (1) توزيع القيم الفعلية مقابل التنبؤات

الشكل 2 خط الاتجاه للقيم الفعلية مقابل التنبؤات: يُظهر هذا الشكل خط الاتجاه للعلاقة بين القيم الفعلية والقيم المتوقعة. يُظهر الخط المستمر الأزرق القيم الفعلية، بينما يُظهر الخط المتقطع البرتقالي التنبؤات. يمكن ملاحظة أن الخطين يتبعان اتجاهات متشابهة بشكل كبير، مما يشير إلى قدرة النموذج على التنبؤ بدقة.



الشكل رقم (2) خط الاتجاه للقيم الفعلية مقابل التنبؤات

الشكل 3 توزيع الأخطاء "يظهر هذا الشكل توزيع الأخطاء في التنبؤات التي قام بها النموذج. المحور الأفقي يمثل الأخطاء، بينما المحور العمودي يمثل عدد التكرارات لكل خطأ. يمكن ملاحظة أن معظم الأخطاء تتركز حول الصفر، مما يشير إلى أن النموذج دقيق في التنبؤات حيث أن الأخطاء صغيرة في الغالب. هذا التوزيع الضيق حول الصفر يعكس أداءً جيدًا للنموذج في التنبؤ بالقيم الفعلية."



الشكل رقم (3) يوضح توزيع الأخطاء في التنبؤات

توضح هذه الأشكال الثلاثة معاً مدى دقة وفعالية النموذج التنبؤي المستخدم. يظهر الشكل الأول تماسكاً قوياً بين القيم الفعلية والمتوقعة، مما يشير إلى دقة عالية للنموذج. الشكل الثاني يوضح أن النموذج يتبع الاتجاهات العامة للقيم الفعلية بشكل جيد. أخيراً، يظهر الشكل الثالث أن توزيع الأخطاء مركّز حول الصفر، مما يعزز من الثقة في أداء النموذج. مجتمعة، هذه الأشكال تقدم نظرة شاملة ومتكاملة على كفاءة النموذج التنبؤي.

4.3 الاستنتاجات

بناءً على النتائج والمناقشة أعلاه، يمكننا استخلاص النقاط التالية: دقة النموذج: النموذج المستخدم (XGboost) أظهر دقة عالية في التنبؤ بالقيم، كما يتضح من معايير الأداء والرسومات البيانية.

اتساق الأداء: النموذج أظهر أداءً متسقاً عبر مختلف البيانات، مما يعزز من الثقة في التنبؤات. توزيع الأخطاء: الأخطاء التي تم تحليلها تظهر تركيزاً حول الصفر، مما يعني أن النموذج غير متحيز ويقوم بعمل جيد في التنبؤ.

4.4 توصيات مستقبلية

بالرغم من أن النموذج أظهر أداءً جيداً، يمكن النظر في بعض التحسينات المستقبلية:

1. توجيه أبحاث مستقبلية: يمكن استخدام نتائج هذه الدراسة لتوجيه الأبحاث المستقبلية في مجال التنبؤ ببيانات أسعار الغذاء والتضخم. مثل هذه الدراسات يمكن أن توفر رؤى جديدة وتحسينات في نماذج التنبؤ المستقبلية.
2. زيادة وتنوع البيانات: جمع مجموعة بيانات أكبر وأكثر تنوعًا يمكن أن يساعد في تحسين دقة التنبؤ. يمكن تضمين بيانات من مصادر متعددة لتعزيز شمولية التدريب وتقليل التحيز.
3. تجربة تحسين المعاملات: العمل على تحسين المعاملات الخاصة بنموذج XGboost ، مثل حجم الطبقات، نوع وظيفة التنشيط، ومعامل التعلم. تحسين هذه المعاملات يمكن أن يعزز الأداء ويزيد من دقة التنبؤ.
4. استخدام تقنيات التعلم الآلي الأخرى: دمج نموذج XGboost مع تقنيات التعلم الآلي أو نماذج التعلم العميق الأخرى مثل الشبكات العصبية العميقة (Deep Learning) ، الغابات العشوائية (Random Forest)، أو الدعم التوجيهي (Support Vector Machines) يمكن أن يساهم ذلك في تحسين دقة النموذج.
5. تحسين أداء النموذج: يُنصح بإجراء دراسات إضافية لاستكشاف معماريات مختلفة لنموذج XGboost مثل استخدام طبقات متعددة أو أنواع مختلفة من الخلايا العصبية بهدف تحسين أداء النموذج بشكل أكبر.
6. معالجة عدم طبيعية المتبقيات: يُنصح بإجراء دراسات إضافية لمعالجة مشكلة عدم طبيعية المتبقيات في نموذج XGboost. يمكن استخدام تقنيات مثل تحويل البيانات أو إضافة ميزات جديدة لتحسين توزيع المتبقيات.
7. دعم اتخاذ القرارات الاقتصادية: يمكن استخدام هذا النموذج في أنظمة دعم اتخاذ القرارات الاقتصادية وتطبيقات الذكاء الاصطناعي في تحليل البيانات الاقتصادية. مثل هذه التطبيقات يمكن أن تساهم في تحسين النظام الاقتصادي وتقديم توصيات فعالة.
8. مقارنة مع دراسات أخرى: تنفيذ مقارنات أوسع مع نماذج الذكاء الاصطناعي الأخرى المستخدمة في دراسات سابقة للوصول إلى أفضل أداء. تحليل أداء النماذج المختلفة ومقارنتها يمكن أن يساعد في اختيار النموذج الأمثل.

المراجع العربية

- أسماء، فهمي. 2023. "معاناة اقتصادية: تأثير الأزمات العالمية المتتالية على أسعار السلع الغذائية". المركز المصري للفكر والدراسات الاستراتيجية. تم الاسترجاع من <https://ecss.com.eg/37469/?form=MG0AV3>.
- Nassra, Z., & Haidar (2024). تصميم نماذج هجينة بين الشبكات العصبونية والانحدار الخطي المتعدد للتنبؤ أثر أسعار المشتقات النفطية على تكاليف المعيشة الغذائية للأسرة السورية. مجلة جامعة حماة، 17(7).

References

- Shively, G. E. (2001). Price thresholds, price volatility, and the private costs of investment in a developing country grain market. *Economic Modelling*, 18(3), 399–414.
- FSIN (Food Security Information Network). (2022). *Global Report on Food Crises: Mid-Year Update*. Retrieved from <https://www.fsinplatform.org>.
- FAOSTAT. (2022). Food and Agriculture Data. Retrieved from <https://www.fao.org/faostat/en/#home>. (accessed on 7 June 2021).
- Minot, N. (2011). Transmission of world food price changes to markets in sub-Saharan Africa. Discussion Papers 1059. International Food Policy Research Institute (IFPRI), Washington, DC.
- Story and Lessons behind the evolution of XGBoost. (2016). Archived from the original on 2016-08-07. Retrieved 2016-08-01.
- Darrat, A. F., & Arize, A. C. (1990). Domestic and international sources of inflation in developing countries: Some evidence from the monetary approach. *International Economic Journal*, 4(4), 55-69.
- Huselius, E., & Walled, L. (2014). Forecasting inflation in Sweden: A univariate approach. *Stockholms Universitet, Department of Statistics*, 11.
- Jere, S., & Siyanga, M. (2016). Forecasting inflation rate of Zambia using Holt's exponential smoothing. *Open Journal of Statistics*, 6(2), 363-372.
- Lidiema, C. (2017). Modelling and Forecasting Inflation Rate in Kenya Using SARIMA and Holt-Winters Triple Exponential Smoothing. *American Journal of Theoretical and Applied Statistics*, 6(3), 161-169.
- Sekine, T. (2001). *Modeling and Forecasting Inflation in Japan*. (No. 01-82). International Monetary Fund.

- Bilel, A., Hassen, T., Fakhri, I., & Habib, Z. (2015). Forecasting Inflation in Tunisia into instability: Using Dynamic Factors Model a two-step based on Kalman filtering. MPRA Paper No. 68455. Retrieved from <https://mpra.ub.uni-muenchen.de/68455>.
- Shively, G. E. (2001). Price thresholds, price volatility, and the private costs of investment in a developing country grain market. *Economic Modelling*, 18(3), 399–414.
- FSIN (Food Security Information Network). (2022). Global Report on Food Crises: Mid-Year Update. Retrieved from <https://www.fsinplatform.org>
- FAOSTAT. (2022). Food and Agriculture Data. Retrieved from <https://www.fao.org/faostat/en/#home>. (accessed on 7 June 2021).
- Minot, N. (2011). Transmission of world food price changes to markets in sub-Saharan Africa. Discussion Papers 1059. International Food Policy Research Institute (IFPRI), Washington, DC.
- Story and Lessons behind the evolution of XGboost. (2016). Archived from the original on 2016-08-07. Retrieved 2016-08-01.
- Darrat, A. F., & Arize, A. C. (1990). Domestic and international sources of inflation in developing countries: Some evidence from the monetary approach. *International Economic Journal*, 4(4), 55-69.
- Huselius, E., & Walled, L. (2014). Forecasting inflation in Sweden: A univariate approach. *Stockholms Universitet, Department of Statistics*, 11.
- Jere, S., & Siyanga, M. (2016). Forecasting inflation rate of Zambia using Holt's exponential smoothing. *Open Journal of Statistics*, 6(2), 363-372.
- Lidiema, C. (2017). Modelling and Forecasting Inflation Rate in Kenya Using SARIMA and Holt-Winters Triple Exponential Smoothing. *American Journal of Theoretical and Applied Statistics*, 6(3), 161-169.
- Sekine, T. (2001). Modeling and Forecasting Inflation in Japan. (No. 01-82). International Monetary Fund.
- Bilel, A., Hassen, T., Fakhri, I., & Habib, Z. (2015). Forecasting Inflation in Tunisia into instability: Using Dynamic Factors Model a two-step based on Kalman filtering. MPRA Paper No. 68455. Retrieved from <https://mpra.ub.uni-muenchen.de/68455>.