Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN : 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.45
https://jatuc.uoturath.edu.iq/

المقارنة بين بعض الدوال اللبّية للتنبؤ بالسلاسل الزمنية باستعمال SVR

علي عبدالله صالح [1]*, أحلام احمد جمعة [2]

[2,1] قسم الإحصاء، كلية الإدارة والاقتصاد، جامعة بغداد، بغداد، العراق.

البريد الإلكتروني للتواصل: ali.saleh2201@coadec.uobaghdad.edu.iqq

المستخلص

يُعد التنبؤ من الموضوعات المهمة في تحليل السلاسل الزمنية لما له من دور في دعم اتخاذ القرارات المستقبلية من خلال استعمال نماذج إحصائية للتنبؤ بالأنماط والاتجاهات. هدفت هذه الدراسة إلى توظيف أنموذج الانحدار المتجه الداعم (SVR) للتنبؤ بالسلاسل الزمنية غير الخطية باستعمال خمس دوال لبية هي الخطية(Linear) ، ومتعددة الحدود(Polynomial) ، والكاوسية(RBF) ، والسكمويد (Sigmoid) عن طريق توليد بيانات تجريبية (محاكاة) تحاكي بيانات تلوث الهواء وذلك لان مثل هذه السلاسل الزمنية البيئية غالبًا ما تتسم بعدم الخطية واحتوائها على قيم متطرفة، مما يجعل النماذج التقليدية أقل كفاءة في التنبؤ. لذلك تم مقارنة أداء الدوال اللبية الخمسة ضمن أنموذج SVR باستخدام مقاييس تقييم إحصائية هي(MSE) و(MAE) و(MAPE) و (R^2)أظهرت نتائج المحاكاة تفوق الدالة الكاوسية (RBF) على بقية الدوال بعد توليد أحجام عينات مختلفة من صغيرة الى متوسطة وكبيرة (2000، 4000، 6000، 8000)، مما يدل على قدرتها الأفضل في تمثيل العلاقات غير الخطية وتحسين دقة التنبؤ. وتوصلت الدراسة إلى أن أنموذج SVR قادر على التعامل مع البيانات غير الخطية بكفاءة أعلى من النماذج التقليدية، مع التوصية باستعمال أنموذج SVR مع الدالة RBF في مجال البيئة مثل ملوثات الهواء وإنتاج الطاقة .

الكلمات المفتاحية

التنبؤ؛ السلاسل الزمنية؛ انحدار متجه الدعم (SVR)؛ دوال النواة؛ دالة الأساس الشعاعي (RBF)؛ البيانات غير الخطية.

© 2026.This is an open access article under the CC by licenses http://creativecommons.org/licenses/by/4.0

1. مقدمة

السلسلة الزمنية هي مجموعة من البيانات تقاس بفترات زمنية متتابعة وتدخل في مجالات متعددة منها الاقتصاد والجغرافية والصناعة وغيرها وتعتمد دراستها على أساليب ونماذج رياضية متعددة بالإضافة الى نماذج التعلم الالي وتعتبر السلاسل الزمنية الأداة الضرورية في تحليل البيانات واتخاذ القرارات وبناء توقعات مستقبلية. يتمثل التنبؤ باستعمال البيانات التاريخية مع استعمال التحليلات الإحصائية والتقنيات الحسابية لتقدير القيم المستقبلية لمتغيرات الظاهرة المدروسة منها أسعار العملات وأسعار الأسهم وإنتاج الطاقة والبيئة لغرض توفير معلومات دقيقة و موثوقة تساعد متخذي القرار في اتخاذ قرارات مستنيرة بشأن الاستثمار والطاقة والبيئة وتقليل المخاطر المرتبطة بها بناءً على التوقعات المستقبلية.

1.1 مشكلة البحث :

تعاني الأجواء العراقية في السنوات الأخيرة من ارتفاع في الملوثات والغازات وخاصة في مدينة بغداد التي تؤثر بشكل مباشر على صحة الفرد العراقي وغالبا ما تكون بيانات تلوث الهواء لها أنماط معقدة في تذبذباتها، و ان اهم ما يميز السلاسل الزمنية هي عملية التنبؤ، لكن هناك عدة مشاكل ترافق عملية التنبؤ ومنها عدم الخطية والانماط المعقدة في السلاسل الزمنية ، وهي المشكلة التي غالبا ما تعاني منها بيانات تلوث الهواء والذي من الصعب التنبؤ بها من خلال الأساليب التقليدية في السلاسل الزمنية ، حيث تم في هذه الورقة البحثية توليد بيانات تجريبية تحاكي بيانات تلوث الهواء.

2.1 أهمية البحث:

تبرز أهمية البحث في كونه يقدم مقاربة تجمع بين بعض الدوال اللبية والانحدار المتجه الداعم لتحسين دقة التنبؤ، مما قد يسهم في بناء نماذج دقيقة وموثوقة يمكن تطبيقها في مجالات عديدة منها البيئة.

3.1 هدف البحث:

يهدف البحث إلى بناء أنموذج يجمع بين بعض الدوال اللبية والانحدار المتجه الداعم SVR وأجراء مقارنة بين الدوال من خلال تقييم ادائها وإيجاد أفضل دالة لبية للتنبؤ بالسلاسل الزمنية من خلال بعض معايير التقييم الإحصائية ومدى قدرتها على التنبؤ بالقيم المستقبلية بدقة وموثوقية عالية.

4.1 الاستعراض المرجعي :

- في عام 2004 قدم الباحث Smola [1] واخرون دراسة عن انحدار متجه الدعم SVR وتطبيقاته في تحليل السلاسل الزمنية. هدفت الدراسة إلى توضيح آلية عمل SVR ومميزاته مقارنةً بالأساليب التقليدية، مع استعراض تطبيقاته في مجالات الاقتصاد والطاقة وتضمنت الدراسة شرح نظري شامل لـ SVR من خلاله استعرض الباحثان الأساس الرياضي لـ انحدار متجه الدعم وشرح كيفية تقليل الخطأ من خلال هامش الخطأ ε-insensitive، مما يسمح للأنموذج بتجاهل الأخطاء الصغيرة وزيادة قدرة التعميم ،و وضحوا آلية استعمال الدالة اللبية المناسبة مع البيانات التي تناسبها كما استعرضوا الأساس الرياضي لـ انحدار متجه الدعم ومن تطبيقاته التنبؤ بأسعار الأسهم والأسواق المالية بناءً على البيانات التاريخية وتوقع استهلاك الطاقة من خلال تحليل أنماط استهلاك الطاقة في فترات زمنية مختلفة ومن اشهر معايير

التقييم التي تم استعمالها هي الجذر التربيعي لمتوسط الخطأ RMSE و الخطأ المطلق المتوسط MAE و معامل التحديد R^2 التي أظهر الـ SVR أداءً أفضل في التنبؤ بالسلاسل الزمنية مقارنةً بالنماذج الإحصائية التقليدية مثل الانحدار الخطي ونماذج ARIMA

- في العام 2009 اسهم الباحث Setiawan [2] واخرون في تقديم دراسة هدفت إلى توقعات للطاقة الشمسية بشكل مستمر لمدة 24 ساعة وعلى مدار عام كامل لمحاكاة الظروف العملية ، تم استعمال بيانات درجة الحرارة وسرعة الرياح كمتغيرات ، وتمت دراسة تأثير هذه المتغيرات على دقة التنبؤ خلال فصول السنة المختلفة من خلال التحليل الموسمي ،ولتحسين دقة الانموذج تمت مقارنة أداء الـSVR مع كل من الشبكات العصبية الاصطناعية (ANNs) و الانحدار الخطي المتعدد (MLR) وكانت النتائج ان انموذج الـ SVR اثبت كفاءة أعلى في التنبؤ مقارنة بالنماذج التقليدية الأخرى، خاصة عند إدخال المتغيرات المناخية ، باستعمال معايير إحصائية منها متوسط الخطأ المطلق (MAE) وجذر متوسط مربع الخطأ (RMSE). أظهرت النتائج أن انموذج الـ SVR، عند تضمين متغيرات إضافية كدرجة الحرارة وسرعة الرياح، قدّم توقعات أكثر دقة وموثوقية، مما يعزز فعاليته في تطبيقات التنبؤ بالطاقة.
- عام 2011 درس الباحثان Kramer and Gieseke [3] بحثاً عن "التنبؤ بسرعة الرياح على المدى القصير باستعمال انحدار متجه الدعم SVR مع بيان مدى تأثير البيانات على دقة التنبؤ وكيفية استعمال الانموذج من دراسة تلك التوقعات ومدى قدرته وفعاليته في التنبؤات قصيرة المدى. و ابرز المعايير التي تم استعمالها هي دقة التنبؤ (Prediction Accuracy) و معلمة دالة الخسارة (Loss Function Parameter) و كمية البيانات التاريخية (Historical Data) كونها تحدد مدى أهمية البيانات السابقة لتحسين دقة التنبؤ، اثبت انموذج SVR قدرته العالية على التنبؤ بسرعة الرياح بشكل دقيق وموثوق عند نقاط محددة وعلى المدى القصير من خلال مقارنة مع عينة الاختبار المستخدمة ويمكن استعمال الانموذج لتحسين إدارة الطاقة المتجددة.
- في عام 2013 قدّم الباحثان Meesad and Rasel [4] دراسة هدفت إلى استعمال تقنية انحدار متجه الدعم SVR كأنموذج للتنبؤ بأسعار الأسهم واتجاهات السوق. تم جمع بيانات يومية لمدة أربع سنوات (2009-2012) من بورصة دكا (DSE) لشركة ACI المحدودة، وهي شركة معروفة في السوق البنغلاديشي ، استعمل الباحثان مشغلات التصفية (Windowing Operators) كخطوة تمهيدية لاختيار البيانات المدخلة لنماذج الـSVR. تضمنت هذه المشغلات التصفية المستطيلة (Rectangular Window) و التصفية المسطحة (Flatten Window) و التصفية غير المسطحة (De-Flatten Window) تم تقييم أداء النماذج باستعمال متوسط نسبة الخطأ المطلق (MAPE). أظهرت النتائج أن استعمال مشغلات التصفية المختلفة أدى إلى تحسين دقة التنبؤات ، استنتجا ان تقنية انحدار متجه الدعم SVR ، عند دمجها مع مشغلات التصفية كخطوة تمهيدية، يمكن أن تكون أداة فعّالة للتنبؤ بأسعار الأسهم واتجاهات السوق. يساعد هذا النهج في تحسين دقة التنبؤ من خلال اختيار بيانات مدخلة أكثر موثوقية لنماذج الـSVR .
- عام 2016 قدّم الباحث Ahn [5] واخرون بحثًا عن تحسين دقة توقعات تدفق حركة المرور باستعمال انموذج SVR في سياق البيانات الضخمة. توصل الباحثون إلى عدة نتائج مهمة حول استعمال الـ SVR للتنبؤ بتدفق المرور في بيئة البيانات الضخمة، بعد ان عالجوا تحديات البيانات الضخمة (الحجم، السرعة، التنوع) من خلال دمج تقنيات معالجة موزعة مثل استعمال إطار عمل Hadoop أو Spark، مما سمح بمعالجة كميات هائلة من البيانات بكفاءة. أظهر الانموذج قدرة على التنبؤ بتدفق المرور في فترات زمنية قصيرة (مثل 5-15 دقيقة) بدقة عالية، مما يجعله مناسبًا للتطبيقات العملية مثل إدارة الحركة المرورية الذكية، كما اظهرSVR قدرة عالية على التعامل مع الأنماط غير الخطية والعلاقات المعقدة في بيانات تدفق المرور، من خلال استعمال الدالة اللبية الكاوسية (RBF Kernel) التي سمحت بنمذجة ديناميكيات التدفق بدقة. بالمقارنة حققت SVR دقة تنبؤ أعلى مقارنةً بالطرائق الأخرى مثل (ARIMA) وغيرها، حيث أظهرت فيه النتائج انخفاضًا ملحوظًا في مقاييس المقارنة مثل الخطأ المطلق المتوسط (MAE) وخطأ متوسط المربعات (MSE).
- في عام 2021 درس الباحث Rohmah [6] وأخرون بحثا عن " مقارنة بأربعة دوال لبية مع SVR للتنبؤ بمؤشر أسعار المستهلك من خلال تكبير المستوى الفائق (the hyperplane) والذي بمقدورة التغلب على التجهيز الزائد (the overfitting) للتوصل الى افضل دقة تنبؤية حيث استعملوا الدوال التالية Spline، و Gaussian-RBF، والدالة الخطية L ، ودالة متعددة الحدود P. وتمت المقارنة بمعياري (MAPE) و (MAPE) وبينت النتائج بأن دالة (RBF - Gaussian) هي الأفضل.
- في عام 2022 قدم الباحث Onal [7] دراسة تناول فيها انموذج SVR القائم على الدالة اللبية الكاوسية للتنبؤ بالطاقة الكهروضوئية قصيرة المدىMPP ، في هذه الدراسة، تم اقتراح انموذج تنبؤ باستعمال انحدار المتجه الداعم SVR القائم على الدالة الكاوسية RBF باستعمال متغيرات إدخال متعددة لتقدير الحد الأقصى للطاقة الكهروضوئية قصيرة المدى والتي تم الحصول عليها من استعمال طريقة المراقبة لكل تورب في الإشعاع المختلف ودرجات الحرارة المختلفة على المدى القصير في محول التعزيز DC-DC في نظام الطاقة الكهروضوئية. يعتمد أداء انموذج التنبؤ القائم على النواة على توفر دالة لبية مناسبة تتوافق مع هدف التعلم، حيث أن وظيفة النواة غير المناسبة أو ضبط المعلمات المفرطة يؤدي إلى ضعف الأداء بشكل كبير. وتحديد دقة وتقييم أداء انموذج التنبؤ المقترح باستعمال (RMSE) و (MSE) حيث اعطى SVR أداء تنبؤ افضل مقارنة الشبكة العصبية الاصطناعيةANN وتقنية تحليل السرب والشبكة العصبية الأمامية (SWD-FFNN)
- في عام 2023 قدم الباحث Malang [8] وأخرون مقارنة بين الدوال اللبية المختلفة في انموذج الانحدار المتجه الداعم لأقل مربع لعملية تخمير البنسلين وقد عملت هذه الدراسة حول توسيع استعمال وظائف الدوال اللبية غير التقليدية في انموذج انحدار متجه الدعم بطريقة المربعات الصغرى LSSVR من خلال اخذ درجة الحرارة ومعدل التهوية وعدة مؤشرات تدخل في عملية التخمر، تم استعمال بيانات محاكاة من عملية تخمير البنسلين الصناعية (IndPenSim) من خلال دراسة مقارنة مع نماذج الانحدار الجزئي للمربعات الصغرى (PLS) ونماذج الانحدار للمكونات الأساسية (PCR) واستعمال (RMSE) و (MAE) وخطأ التقريب (Ea) كمعيار لمقارنة للأداء وبينت النتائج تفوق الدالة اللبية MQ LSSVRعلى دالة RBF LSSVR
- في العام 2025 قدم الباحث Hamzeh [9] واخرون دراسة ستة مناهج تنبؤية منها ثلاثة نماذج للتعلم الآلي وهي انحدار المتجه الداعمSVR والغابة العشوائية(RF) وخوارزمية أقرب جار (KNN) وأنموذج التعلم العميق الذاكرة طويلة المدى (LSTM) وإنموذجان للسلاسل الزمنية هما CARMA وGARCH-CARMA من خلال اختيار ثماني محطات قياس هيدرومترية في حوض نهر كاشكان غرب إيران لتقييم أداء النماذج من خلال معيارين هما الخطأ جذر متوسط مربع الخطأ (RMSE)، ومعامل التحديد (R^2).واشارت النتائج بان نماذج الغابات العشوائية حققت افضل النتائج في نماذج التعلم الالي بينما كان انموذج GARCH-CARMAهو الأفضل في نماذج السلاسل الزمنية.

2. مراحل بناء الانموذج

1.2 تشخيص السلسلة الزمنية باستعمال (Support Vector Regression – SVR)

تعد مرحلة تشخيص السلسلة الزمنية من المراحل المهمة في تحليل السلسلة الزمنية، إذ يتم في هذه الخطوة فحص السلسلة الأصلية بصريًا للكشف عن وجود اتجاه عام أو تقلبات غير مستقرة في التباين. كما يتم اختبار استقرارية السلسلة من حيث المتوسط والتباين، لأن نماذج SVR تتأثر بشكل مباشر بتوزيع البيانات. في حال ثبوت عدم الاستقرارية. [10]

2.2 تحديد بنية أنموذج SVR

بعد تشخيص السلسلة ومعالجتها، تنتقل إلى مرحلة تحديد بنية الانموذج في هذه المرحلة يتم اختيار الدالة اللبية التي تمثل العلاقة غير الخطية بين المدخلات والمخرجات، مثل النواة الخطية أو متعددة الحدود أو الكاوسية RBF أو السكمويد أو لابلاس، وذلك اعتمادًا على طبيعة السلسلة الزمنية. بالتوازي مع ذلك يتم تحديد فضاء البحث لمعلمات الأنموذج الأساسية، والمتمثلة بمعامل التنظيم C الذي يتحكم في درجة الموازنة بين تعقيد الأنموذج ودقة التوافق، ومعامل الخطأ ε الذي يحدد عرض أنبوب الخطأ، ومعامل γ في حال استعمال الدالة غير الخطية.

3.2 تقدير معلمات الأنموذج SVR

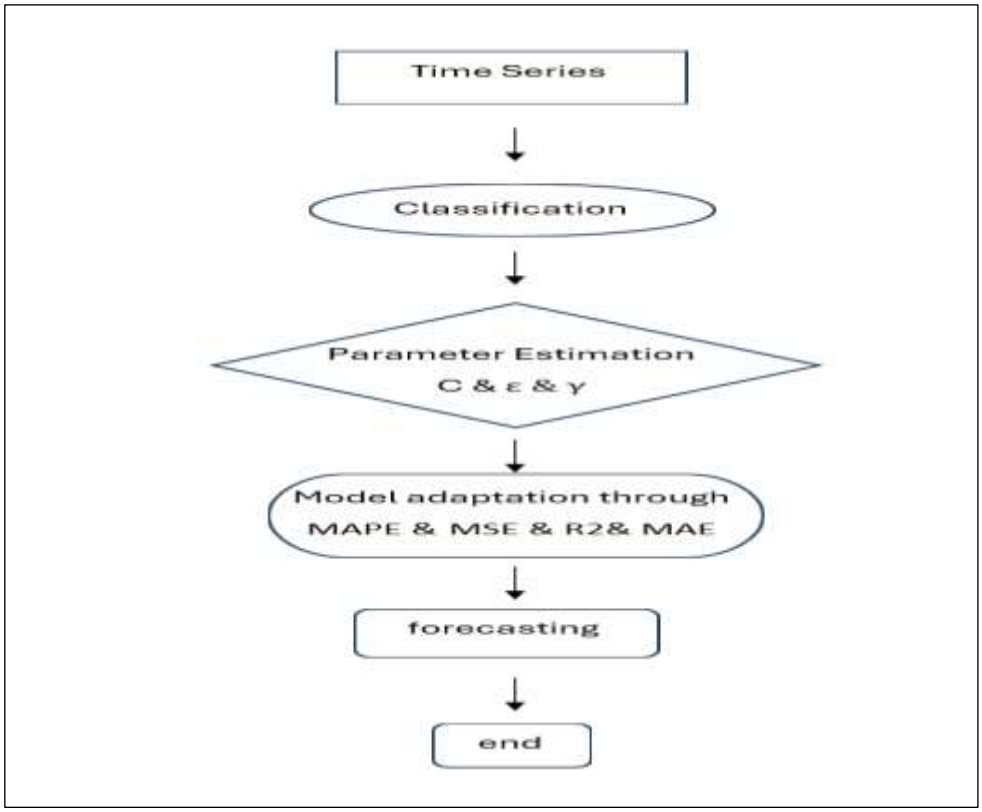
يتم تقدير معلمات أنموذج SVR بصورة مثلى، حيث تُستخدم تقنيات التحقق المتقاطع Cross-validation للحد من مشكلة فرط التوافق، ويُطبَّق البحث الشبكي Grid Searchأو خوارزميات التحسين لاختيار القيم المثلى للمعلمات التي تحقق أقل قيمة لدالة الخطأ. تُعد هذه المرحلة جوهرية لأنها تحدد الأداء التنبؤي للأنموذج بشكل مباشر.

4.2 اختبار مدى ملاءمة الأنموذج

بعد تقدير المعلمات، تأتي مرحلة اختبار مدى ملاءمة الأنموذج ، والتي يتم فيها تقييم كفاءة أنموذج SVR باستعمال مجموعة من المقاييس الإحصائية مثل متوسط مربع الخطأ، ومتوسط الخطأ المطلق، ومتوسط نسبة الخطأ المطلق، ومعامل التحديد. كما يتم تحليل أخطاء التنبؤ للتحقق من عدم وجود أنماط منتظمة أو تحيز زمني، إضافةً إلى مقارنة أداء النماذج الناتجة عند استعمال دوال مختلفة لاختيار الانموذج الأكثر ملاءمة للبيانات[11].

5.2 التنبؤ

يعتبر التنبؤ من اهم ما تتميز به السلاسل الزمنية ، حيث تُجمع فيه المعلومات والملاحظات من البيانات السابقة لنفس المتغير وتُحلل لتطوير انموذج يصف العلاقة الكامنة. وبذلك فهو يتمثل بالمرحلة الأخيرة حيث يستخدم أنموذج SVR المعتمد لإجراء عملية التنبؤ بالقيم المستقبلية للسلسلة الزمنية. وتشمل هذه المرحلة عرض النتائج بيانياً من خلال رسم القيم الحقيقية مقابل القيم المتنبأ بها، فضلاً عن تمثيل مسار التنبؤ المستقبلي، مما يتيح تقييمًا بصريًا واضحًا لجودة الانموذج ودقته التنبؤية [12].
والمخطط الاتي يوضح خوارزمية عملSVR :



الشكل 1. خوارزمية عمل SVR

6.2 خوارزمية التعلم الالي في السلاسل الزمنية

تُعد خوارزميات التعلّم الآلي Machine Learning Algorithm من الأدوات الحديثة والفعّالة في تحليل السلاسل الزمنية والتنبؤ بقيمها المستقبلية، لما تمتلكه من قدرة عالية على تمثيل العلاقات غير الخطية والتعامل مع التعقيد البنيوي وعدم الاستقرارية التي غالبًا ما تميّز هذا النوع من البيانات، لاسيّما في السلاسل المالية والبيئية [13].

7.2 الانحدار المتجه الداعم (Support Vector Regression)

الانحدار بالمتجهات الداعمة SVR هو امتداد لخوارزمية (SVM-Supporting vector machine) التي تُستخدم للتصنيف. الفكرة في SVR هي إيجاد أفضل خط انحدار يمر من خلال البيانات مع محاولة تقليل الخطأ باستعمال مفهوم (ε-tube) [14] لكن في SVR يسمح بهامش خطأ صغيرε ، مما يعني أنه إذا كانت التوقعات $f(x)$ قريبة من القيم الحقيقية y بمقدار ϵ . [15] .

وان المعادلة الرياضية بالنسبة للحالة الخطية له تُعطى بالشكل التالي (Layth & Tasnim, 2024)

$$\hat{y} = w \cdot x + b \qquad (1)$$

اذ ان:
$\hat{y}$: الناتج أو القيمة المتوقعة (predicted output)
x : المدخل أو المتغير المستقل (weights vector)
w : الوزن أو معامل الانحدار، وهو ما يحدد تأثير x على y
b : معامل الانحياز

اما في الحالة غير الخطية تكون المعادلة الرياضية له بالشكل التالي:[16]

$$\hat{y} = \sum_{i=}^{n} a_i k(x_i, x) + b \qquad (2)$$

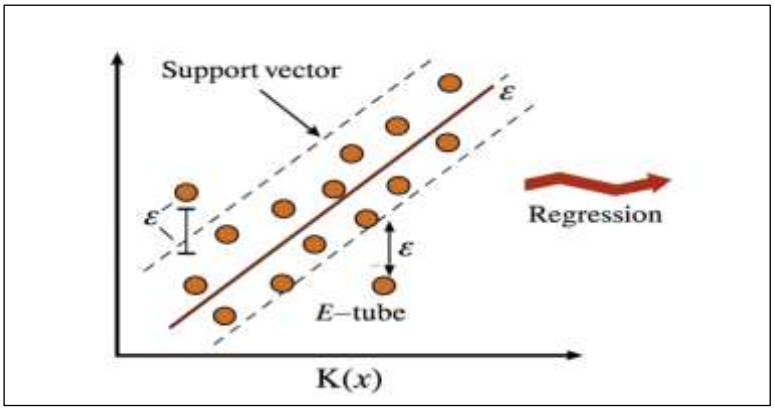
اذ ان:
$\hat{y}$: الناتج أو القيمة المتوقعة (predicted output)
a_i : معاملات لاكرانج لمتجه الدعم(Lagrange coefficients support vector)
$k(x_i, x)$: الدالة اللبية

8.2 خوارزميات انحدار متجه الدعم (Support Vector Regression SVR Algorithm)

خوارزمية انحدار متجه الدعم (Support Vector Regression-SVR) تعتبر امتداداً لخوارزمية متجه الدعم الالي (Support Vector Machine - SVM) وهي خوارزمية تعلم آلي تستعمل لتحليل الظواهر و تهدف إلى نمذجة العلاقات غير الخطية بين المتغيرات عبر الهامش المتاح للخطا (ε-insensitive tube)حول الخط التنبؤي، حيث تُهمل الأخطاء الأصغر والاكبر من الهامش المسموح به ε ، مما يقلل من تأثير الضوضاء ومن ضمن معاييرها معامل التنظيم (C)الذي يتحكم في توازن الانموذج بين تقليل الأخطاء وتجنب الإفراط في التخصيص و (ε) تحدد الاخطاء التي تهمل وترفض الأخطاء التي خارج هامش الخطأ ε-insensitive tube وكما يلي (Layth & Tasnim, 2024)

$$f(x) = w^T \phi(x) + b \qquad (3)$$

اذ ان :
w: هو متجه الأوزان(weights vector)
$\phi(x)$: تحويل غير خطي للبيانات الى فضاء عالي الابعاد(Nonlinear mapping to feature space)
الهدف من استعمال Support Vector Regression ايجاد w و b بحيث تكون قيمة الانحرافات بين القيم الحقيقية والقيم التنبؤية ضمن هامش الخطاء المسموح به (ε) كما في الشكل التالي (Muthiah et al., 2021).



الشكل 2. الحالة الخطية

مع التصغير وتقليل دالة الهدف (Optimization Problem) [18]

$$min \ \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}(\xi_i + \xi_i^*) \qquad (4)$$

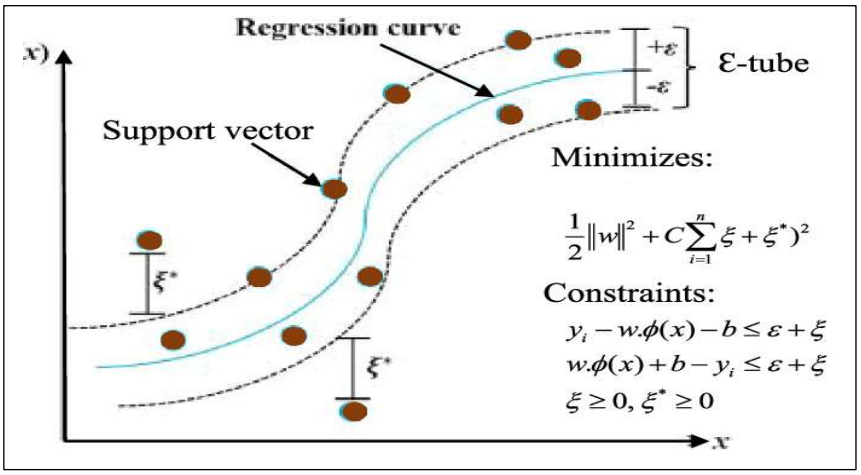
اذ ان:
ξ_i : يقيس الانحراف فوق خط الانحدار Slack variable (upper)
ξ_i^* : يقيس الانحراف تحت خط الانحدار Slack variable (lower)
$\frac{1}{2}\|w\|^2$: يعبر عن تعقيد الانموذج Squared norm of weights
C: معامل التنظيم Regularization parameter
ضمن القيود Constraints

$$\boldsymbol{y_i - (w^T \phi(x_i) + b) \leq \varepsilon + \xi_i} \qquad (5)$$

اذ ان :
b : هو معامل الانحياز الذي يحدد المستوى العمودي لدالة التنبؤ

$$(W^T\phi(x)+b)-y_i \le \varepsilon+\xi_i^* \quad (6)$$

$$\xi_i,\xi_i^* \ge 0 \quad (7)$$



الشكل 3. للحالة غير الخطية [19]

يتم اشتقاق دالة الهدف الاصلية للحصول على دالة التصغير لمضاعفات لاكرانج بالشكل الاتي بناء دالة لاكرانج [1] :

$$min\ \frac{1}{2}\|w\|^2+c\sum_{i=1}^{N}(\xi_i+\xi_i^*) \quad (8)$$

بناء دالة لاكرانج

$$L=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)-\sum_{i=1}^{n}a_i\,(\varepsilon+\xi_i-y_i+w^Tx_i+b)-\sum_{i=1}^{n}a_i^*\,(\varepsilon+\xi_i^*+y_i-w^Tx_i-b)$$
$$-\sum_{i=1}^{n}(n_i\xi_i+n_i^*\xi_i^*) \quad (9)$$

اشتقاق بالنسبة لـ W

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{n}(a_i-a_i^*)\,x_i=0\ \Rightarrow\ w=\sum_{i=1}^{n}(a_i-a_i^*)\,x_i$$

بالنسبة للـ b:

$$\frac{\partial L}{\partial b}=w-\sum_{i=1}^{n}(a_i-a_i^*)\ =0$$

بالنسبة لـ ξ_i و ξ_i^*

$$\frac{\partial L}{\partial \xi_i}=C-a_i-n_i=0\ \Rightarrow a_i\ge C$$

$$\frac{\partial L}{\partial \xi_i^*}=C-a_i^*-n_i^*=0\ \Rightarrow\ a_i^*\ge C$$

وبعد التعويض في دالة لاكرانج نحصل على المعادلة الثنائية بحيث تجعل الانموذج اكثر كفاءة للتعامل مع البيانات الكبيرة وهي الصيغة المستخدمة عند التطبيق.

$$\min_{a_i,a_i^*}\Big(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i-a_i^*)\,(a_j-a_j^*)k(a_i-a_i^*)+\varepsilon\sum_{i=1}^{n}(a_i-a_i^*)-\sum_{i=1}^{n}y_i(a_i-a_i^*)\Big) \quad (10)$$

ضمن القيود

$$\sum_{j=1}^{n}(a_i-a_i^*)=0 \quad (11)$$

$$0<\alpha_i a_i<C \quad (12)$$

اذ ان:

a_i و a_i^* : مضاعفات لاكرانج وان الفرق بينهما يستخدم لاحقًا في دالة التنبؤ
$K(x_i, x_j)$: الدالة اللبية
المعادلة الرياضية لحساب متجه الاوزان في الـ SVR تساوي

$$\boldsymbol{w} = \sum_{i=1}^{n} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)\boldsymbol{\phi}(\boldsymbol{x_i}) \qquad (\mathbf{13})$$

وان المعادلة التنبؤية للـ SVR تستخدم بعد تدريب الانموذج تكون:

$$\widehat{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{i=1}^{n} (\boldsymbol{a_i} - \boldsymbol{a_i^*})\boldsymbol{k}(\boldsymbol{x_i}, \boldsymbol{x_k}) + \boldsymbol{b} \qquad (\mathbf{14})$$

اذ ان :
$\hat{y}(x)$: القيمة التي يراد التنبؤ بها
$k(x_i, x_k)$: الدالة اللبية
a_i و a_i^* : معاملات لاغرانج المستخرجة من التدريب
المعادلة الرياضية لأيجاد قيمة b معامل الانحياز

9.2 الـــــــدوال الـــــلبـــــية (kernel function)

الدوال اللبية او النواة هي أداة رياضية جوهرية في التعلم الآلي، تُمكّن من تحويل البيانات إلى فضاءات ذات أبعاد أعلى دون الحاجة للحساب الصريح للإحداثيات الجديدة، تقوم هذه الدوال بتسهيل نمذجة العلاقات غير الخطية بكفاءة حسابية عالية ، تُستخدم على نطاق واسع في خوارزميات مثل SVM و PCA لمعالجة مشاكل التصنيف والانحدار المعقدة والصيغة العامة لها [20] .

$$k(X, Y) = \phi(X), \phi(Y)$$

اذ ان:
$\phi(.)$: دالة تحويل (Mapping Function) تنقل البيانات من فضاء الإدخال إلى فضاء (Feature Space)

الجدول 1. تحديد قرار استعمال الدالة المناسبة حسب طبيعة البيانات

ت	الدالة	الصيغة الرياضية	المصدر
1	Linear	$\boldsymbol{k(x,y) = x^T y}$	[21]
2	Polynomial	$\boldsymbol{k(x,y) = (\gamma \cdot x^T \cdot y + cof)^d}$	[22]
3	RBF	$\boldsymbol{k(x,y) = exp(-\gamma \|x-y\|^2)}$	[18]
4	Sigmoid	$\boldsymbol{k(x,y) = tanh(\gamma \cdot x^T \cdot y + cof)}$	[23]
5	Laplacian	$\boldsymbol{k(x,y) = exp(-\gamma \|x-y\|)}$	[23]

10.2 طرائق تقدير معلمات انموذج الانحدار المتجه الداعم (SVR Parameter Estimation Methods)

1-10-2 التحقق المتقاطع (Cross-Validation)

يُستخدم أسلوب التحقق المتقاطع (Cross-Validation) لتقييم قدرة الانموذج على التعميم (Generalization) وتقليل الانحياز من خلال تقسيم البيانات الى مجموعة اختبار وتدريب ونظرًا لطبيعة بيانات السلاسل الزمنية تم اعتماد أسلوب (Time Series Cross-Validation) باستخدام عدد طيات k مع الحفاظ على الترتيب الزمني للبيانات،إذ تُقسَّم البيانات إلى نوافذ زمنية متتالية بحيث يتم في كل طية تدريب نموذج Support Vector Regression (SVR)على جميع المشاهدات السابقة حتى نقطة زمنية معينة ثم اختباره على المشاهدات التي تليها مباشرة مع توسيع مجموعة التدريب تدريجيًا (Expanding Window) وقد استُخدم هذا الأسلوب أيضًا في ضبط المعلمات مما يسهم في تقليل مشكلة الإفراط في التخصيص (Overfitting) وتحسين قدرة النموذج على التعميم(Generalization) . [24]

2-10-2 البحث الشبكي (Grid Search)

هناك عدة معلمات مثل ε, γ, C لا يمكن معرفة قيمها المثلى ولا يُعرف مسبقًا اي من القيم هي الأفضل للمشكلة المعطاة لذلك يجب إجراء نوع من اختيار الانموذج أي البحث عن المعلمات، الهدف هو تحديد قيم جيدة للمعلمات بحيث يستطيع الأنموذج أن يتنبأ بدقة بالبيانات غير المعروفة أي بيانات الاختبار [25]

11.2أخطاء التنبؤ Forecasting Errors

عادة ما تدرس السلسلة الزمنية لغرض اكتشاف نمط التطور التاريخي للظاهرة المدروسة واستعمال هذا النمط في التنبؤ بالقيم المستقبلية و يعرف كالآتي :(Shaarawy, 2005)

$$\boldsymbol{e_t = y_t - \widehat{y}_t} \qquad \ldots . (\mathbf{15-1}) \qquad t = 1,2,\ldots, n$$

حيث يرمز n إلى طول السلسلة أي حجم العينة.

12.2 معايير تقييم النتائج Results Evaluation Criteria

يتمثل بعدد من المعايير الإحصائية التي تعمل المفاضلة بين النماذج ، ويؤخذ الانموذج الذي يقابل أقل قيمة لتلك المعايير والذي يتمثل بأعلى دقة تنبؤية فيه ومنها ما يلي:

2-12-1 متوسط مربعات الأخطاء (Mean Squared Error- MSE)

متوسط مربعات الخطأ يعتبر من مقاييس أداء الانموذج و يرمز له بـ MSE ويعرف كما يأتي [27]:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}} \mathrm{e}_{\mathrm{t}}^2 = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}} (\mathrm{y}_{\mathrm{t}} - \hat{\mathrm{y}}_{\mathrm{t}})^2 \qquad \dots\dots (16-2)$$

اذ ان :

n :عدد المشاهدات

y_t : القيمة الحقيقية للملاحظة(Actual Value)

$\hat{y}_t$: القيمة المقدرة للملاحظة(Predicted Value)

2-12-2 متوسط الأخطاء النسبية المطلقة (Mean Absolute Percentage Errors -MAPE)

ان مقياس متوسط الأخطاء النسبية المطلقة (MAPE) يستعمل لتقييم دقة النماذج التنبؤية، ويعبر عن متوسط نسبة الخطأ المطلق بين القيم المتوقعة والقيم الفعلية وكما يلي [28]:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t}\right| \qquad \dots\dots (17-2)$$

ويتميز هذا المقياس عن كل المقاييس بأنه مقياس نسبي أي لا يعتمد على الوحدات المستخدمة وبنفس الوقت يستخدم كثيراً في السلاسل الزمنية لتمتعه بصفات إحصائية معينة .

2-12-3 معيار متوسط الأخطاء المطلقة (Mean Absolute Error - MAE)

يستعمل متوسط الأخطاء المطلقة (MAE) لتقييم دقة النماذج التنبؤية أو الانحدارية، ويُحسب عن طريق أخذ متوسط القيم المطلقة للفرق بين القيم الحقيقية والقيم المتوقعة كلما اقترب MAE من الصفر، زادت دقة الانموذج. الصيغة الرياضية هي:[29]

$$\mathrm{MAE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}} |\mathrm{y}_{\mathrm{t}} - \hat{\mathrm{y}}_{\mathrm{t}}| \qquad \dots\dots (18-2)$$

2-12-4 معامل التحديد(R^2) (Coefficient Of Determination - R^2)

ان معامل التحديدR^2 يمثل مقياس إحصائي يُستعمل لتقييم جودة انموذج الانحدار (Regression Model) في تفسير التباين في المتغير التابع (Dependent Variable)ويُعبِّر عن نسبة التباين المفسَّر من قبل الانموذج مقارنة بالتباين الكلي في البيانات والمعادلة الرياضية هي[30]:

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\sum_{\mathrm{t}=1}^{\mathrm{n}} (\mathrm{y}_{\mathrm{t}} - \hat{\mathrm{y}}_{\mathrm{t}})^2}{\sum_{\mathrm{t}=1}^{\mathrm{n}} (\mathrm{y}_{\mathrm{t}} - \bar{\mathrm{y}}_{\mathrm{t}})^2} \qquad \dots\dots (19-2)$$

3. المحاكاة Simulation

المحاكاة هي تقليد للواقع الفعلي للعملية أو النظام باستعمال نماذج يتم تنفيذها بالحاسوب، وتكون مخرجاتها غالبًا مقابلة وبشكل مباشر للمخرجات المسجلة في الأنظمة الحقيقية. اذ تم في هذه الورقة البحثية توليد اربع حجوم عينة غير خطية تحاكي بيانات تلوث الهواء

3-1 منهجية البرنامج التجريبي

1-توليد بيانات عشوائية غير خطية بحجم (2000-4000-6000-8000) تحاكي بيانات تلوث الهواء بـ 1000 تكرار للتوليد .

2- تقسيم البيانات إلى تدريب واختبار (Train/Test Split) يتم تقسيم البيانات بنسبة 80% تدريب و20% اختبار، وهي ممارسة معيارية لضمان تقييم الأنموذج على بيانات غير مرئية أثناء التدريب

3- تحديد نطاقات المعلمات (Grid Search Ranges) يتم تحديد شبكات بحث للمعلمات الأساسية لنماذج عبر مدى واسع من المعلمات

4- تضمين ازاحات lags للسلسلة الزمنية مثل(y_{t-1}, y_{t-2}) تُضاف كمدخلات للنموذج لتوضيح الاعتماد الزمني .أهميتها أنها تمكّن الأنموذج من التقاط الأنماط الزمنية (اتجاهات، موسمية، دورات) وتحسين دقة التنبؤ بشكل كبير.

3-2 نتائج المحاكاة

3-2-1 عند حجم عينة 2000

تم اخذ معايير المقارنة (MSE Test & Train , MAPE , MAE , R^2) وكما موضح في الجدول (3-1)

جدول 2. معايير المقارنة للدوال عند حجم عينة 2000

Kernel	$\overline{\mathrm{MSETrain}}$	$\overline{\mathrm{MSETest}}$	$\overline{\mathrm{MAE}}$	$\overline{\mathrm{MAPE}}$	$\overline{R^2}$	C	ε	γ	Coeff
Laplacian	13.6465	12.7095	1.6189	18.0132	0.7113	10	0.05	0.1	---
linear	16.8484	14.6827	1.3434	12.4017	0.6685	10	0.01	---	---
poly	14.6847	12.8667	1.4482	13.8075	0.7093	10	0.05	0.001	d=1
rbf	12.4744	11.9007	1.2153	11.3391	0.7317	50	0.01	0.001	---
sigmoid	14.7407	13.4905	1.5989	14.9204	0.6950	100	0.05	0.001	1

تشير النتائج بشكل واضح الى كفاءة الدالة RBF عند حجم العينة 2000 حيث أظهرت أدنى قيم لخطأ التدريب والاختبار مع أعلى قيمة لمعامل التحديد

2-3-3 عند حجم عينة 4000

عند عينة بحجم 4000 النتائج موضحة في جدول (3-2) الاتي :

جدول 3. معايير المقارنة للدوال عند حجم عينة 4000

Kernel	$\overline{\text{MSETraın}}$	$\overline{\text{MSETest}}$	$\overline{\text{MAE}}$	$\overline{\text{MAPE}}$	$\overline{R^2}$	C	ε	γ	Coeff
Laplacian	13.1627	9.1022	1.3106	14.3890	0.7610	10	0.05	0.001	---
linear	16.5514	10.7320	1.1782	11.6160	0.7187	200	0.005	---	---
poly	14.3212	9.3285	1.2107	11.9125	0.7553	10	0.001	0.001	d=1
rbf	12.1234	8.9986	1.0670	10.5344	0.7640	200	0.001	0. 1	---
sigmoid	14.3061	9.8721	1.4010	13.9880	0.7411	50	0.05	0.001	1

مع زيادة حجم العينة تحسن الأداء العام لجميع النماذج، وانخفضت قيم أخطاء الاختبار بصورة ملحوظة، مما يعكس استفادة النماذج من وفرة البيانات.

3-3-3 عند حجم عينة 6000

جدول 4. معايير المقارنة لدوال عند حجم عينة 6000

Kernel	$\overline{\text{MSETraın}}$	$\overline{\text{MSETest}}$	$\overline{\text{MAE}}$	$\overline{\text{MAPE}}$	$\overline{R^2}$	C	ε	γ	Coef
Laplacian	8.9528	7.7372	1.0935	11.5472	0.7987	10	0.05	0.001	--
linear	11.0653	9.6939	0.9748	9.1321	0.7478	300	0.005	0.001	--
poly	9.6374	8.3226	1.0050	9.1634	0.7834	10	0.01	0.001	d =1
rbf	8.2747	7.5963	0.9023	8.4069	0.8023	10	0.005	0. 1	--
sigmoid	9.6856	8.5100	1.2330	13.8092	0.7786	10	0.001	0.001	1

تُظهر النتائج استمرار الاتجاه التنازلي لقيم الخطأ مع تحسن ملحوظ في R^2، ما يدل على استقرار أكبر في التقدير.

4-3-3 عند حجم عينة 8000

جدول 5. معايير المقارنة لدوال عند حجم عينة 8000

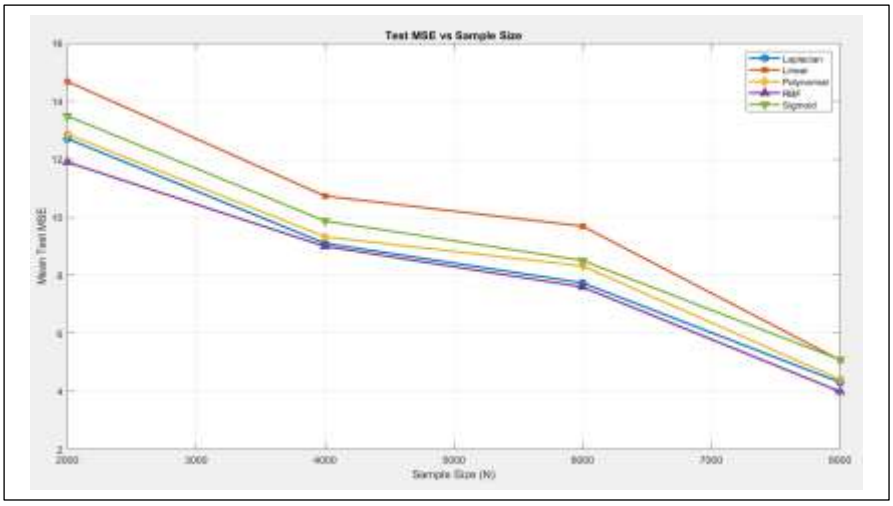
Kernel	$\overline{\text{MSETraın}}$	$\overline{\text{MSETest}}$	$\overline{\text{MAE}}$	$\overline{\text{MAPE}}$	$\overline{R^2}$	C	ε	γ	Coeff
Laplacian	5.4485	4.3123	0.9397	9.4788	0.8774	10	0.05	0.01	----
linear	7.5774	5.0732	0.8387	8.0834	0.8562	10	0.005	----	----
poly	6.4635	4.4099	0.8838	8.6220	0.8748	10	0.001	0.001	d=1
rbf	5.4599	3.9907	0.7882	7.6411	0.8864	50	0.001	0. 1	----
sigmoid	6.8564	5.0908	1.2641	16.0693	0.8548	10	0.05	0.001	1

عند حجم العينة الأكبر بلغت النماذج مستوى مرتفعًا من الدقة، حيث انخفضت قيم MSE وMAE إلى أدنى مستوياتها.

4-3 معايير المقارنة للدوال الخمسة

1-4-3 معيارTest MSE

تم احتساب معيار Test MSE للدوال الخمسة ولجميع احجام العينات والنتائج موضحة بالشكل (3-1)

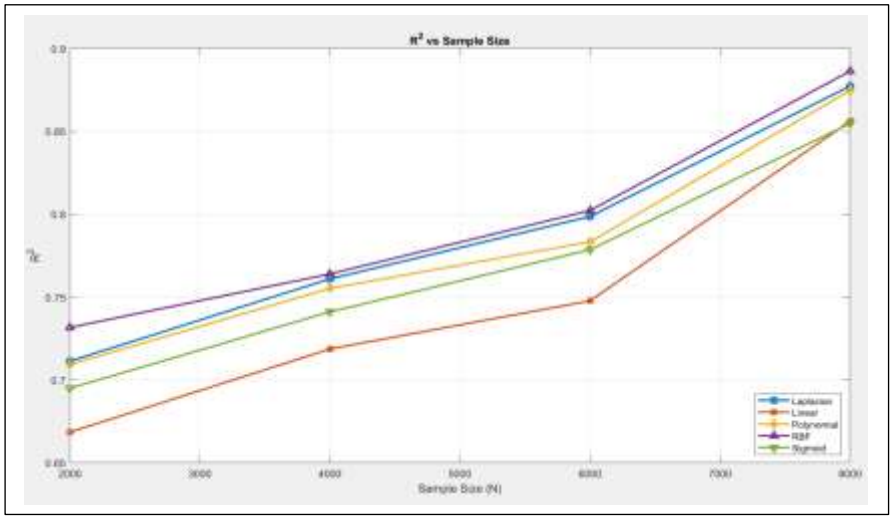


الشكل 4. معيار Test MSE للدوال الخمسة ولجميع حجوم العينات

وضح الشكل العلاقة بين حجم العينة ومتوسط مربعات الخطأ للاختبار (Test MSE) لأنموذج انحدار المتجه الداعم باستعمال خمس دوال لبية مختلفة.

2-4-3 معيار R^2

تم احتساب معامل التحديد R^2 للدوال الخمسة ولجميع حجوم العينات والنتائج موضحة في الشكل (3-2)



الشكل 5. معيار معامل التحديد R^2 للدوال الخمسة ولجميع حجوم العينات

يبين الشكل تطور معامل التحديد R^2 مع زيادة حجم العينة لأنموذج انحدار المتجه الداعم باستعمال دوال لبية مختلفة. يلاحظ وجود تحسن تدريجي في قيمة R^2 لجميع الدوال مع زيادة حجم العينة

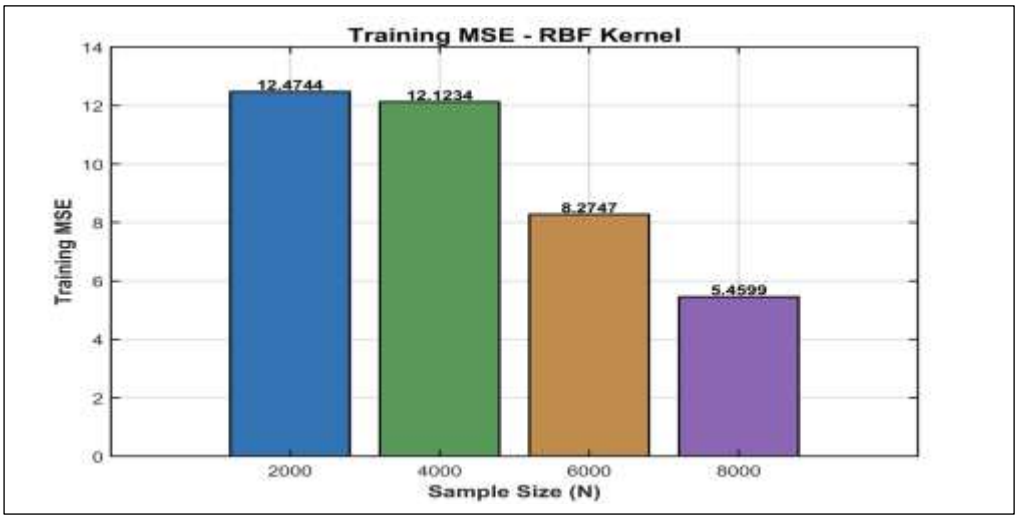
5-3 النتائج بالنسبة للدالة RBF

تم اخذ نتائج دالة RBF وذلك لتفوق هذه الدالة ولجميع حجوم العينات وفقاً للنتائج الموضحة في والجدول(3-5) يوضح ملخص لنتائج المعايير وفقاً لهذه الدالة.

جدول 6. معايير المقارنة لدالة RBF لجميع حجوم العينات

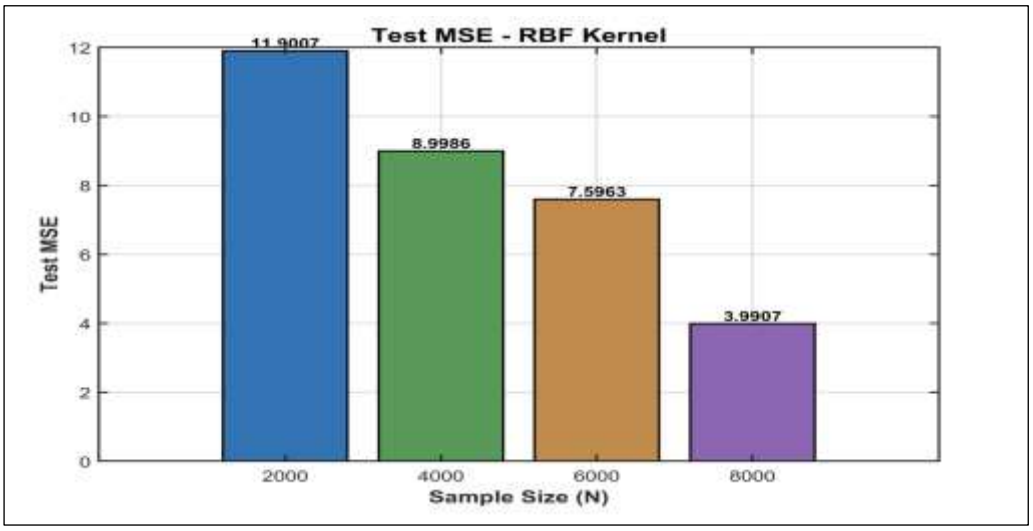
Kernel	N	$\overline{\text{MSETrain}}$	$\overline{\text{MSETest}}$	$\overline{\text{MAE}}$	$\overline{\text{MAPE}}$	$\overline{R^2}$	C	ε	γ
rbf	8000	5.4599	3.9907	0.7882	7.6411	0.8864	50	0.001	0.1
rbf	6000	8.2747	7.5963	0.9023	8.4069	0.8023	10	0.005	0.1
rbf	4000	12.1234	8.9986	1.0670	10.5344	0.7640	200	0.001	0.1
rbf	2000	12.4744	11.9007	1.2153	11.3391	0.7317	50	0.01	0.001

يتبين من نتائج جدول (3-5) أن أداء دالة RBF يتحسن تدريجيًا بزيادة حجم العينة، حيث تنخفض قيم الخطأ وترتفع القدرة التفسيرية بشكل منتظم.



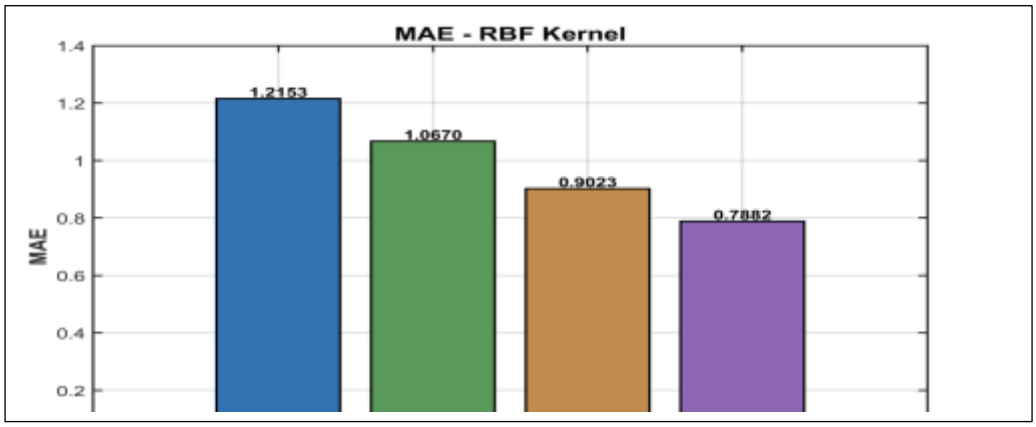
الشكل 6. معيار Training MSE لدالة RBF عند كل حجم عينة

يظهر الشكل(3-3) مجموع مربعات خطا التدريب ينخفض كلما كبر حجم العينة حيث بدا 12.47 عند حجم العينة 2000 وانخفض الى 5.45 عند حجم العينة 8000 وهذا يعزز ان زيادة حجم العينة يزيد من دقة الانموذج.



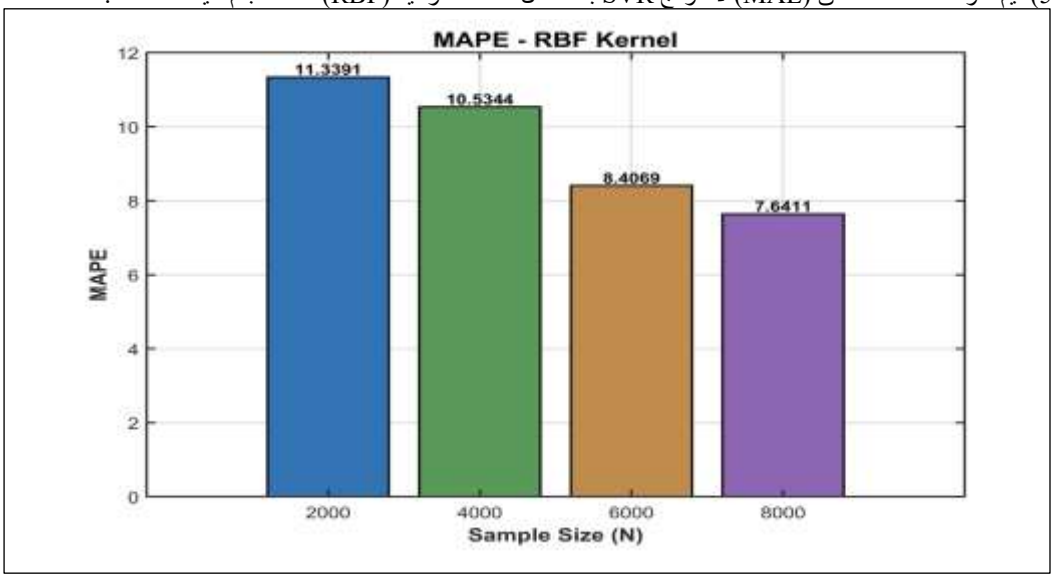
الشكل 7. يمثل معيار Test MSE

يوضح الشكل متوسط مربعات الخطأ لعينة الاختبار (Test MSE) لأنموذج انحدار المتجه الداعم باستعمال الدالة RBF عند أحجام عينات مختلفة.



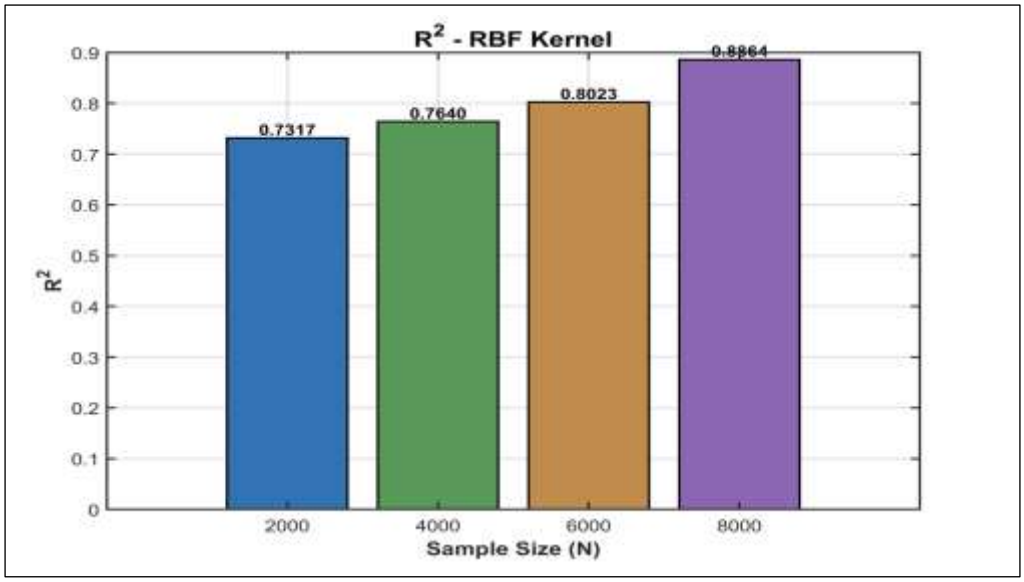
الشكل 8. يمثل معيار المقارنةMAE لدالة RBFعند كل حجم عينة

يوضح الشكل (3-5) قيم متوسط الخطأ المطلق (MAE) لأنموذج SVR باستعمال الدالة الكاوسية (RBF) عند أحجام عينات مختلفة.



الشكل 9. معيار المقارنةMAPE لدالة RBFعند كل حجم عينة

يوضح الشكل قيم متوسط نسبة الخطأ المطلق (MAPE) لنموذج SVR باستعمال الدالة الكاوسية (RBF) عند أحجام عينات مختلفة. نلاحظ وجود اتجاه تنازلي واضح في قيم MAPE مع زيادة حجم العينة، مع زيادة حجم العينة، مما يؤكد أهمية استعمال أحجام عينات كبيرة للحصول على نتائج تنبؤ أكثر دقة بالإضافة الى رسم نتائج معامل التحديد (R^2) للدالة عند كل حجم عينة كما في الشكل (3-7)



الشكل 10. معيار R^2 لدالة RBF عند كل حجم عينة

يوضح الشكل (3-7) قيم معامل التحديد (R^2) لأنموذج SVR باستعمال الدالة الكاوسية (RBF) عند أحجام عينات مختلفة.

4. الاستنتاجات

يمكن تلخيص استنتاجات البحث بما يأتي :

1- أظهرت النتائج التجريبية (المحاكاة) بعد توليد البيانات غير الخطية بأربع حجوم للعينة تحاكي بيانات تلوث الهواء بأحجام 2000- 4000 -6000- 8000 تفوقا للدالة الكاوسية في جميع احجام العينات التجريبية مما يدل على كفاءة الأنموذج الذي تم اختياره وبالتالي أمكانية استعماله للتنبؤ في ملوثات الهواء بشكل جيد و بناءاً على المعايير الإحصائية المهمة مثل MSE_ وMAE وMAPE وR^2 كانت متوافقة على ان الدالة الكاوسية هي الأفضل .

2- اثبتت النتائج انه كلما زاد حجم العينة كانت النتائج اكثر دقة والخطأ كان اقل وان معامل التحديد يتناسب عكسيا مع متوسط مربعات الاخطاء.

3- نستنتج من خلال تطبيق الأنموذج SVR مع الدوال اللبية قادر على التعامل مع البيانات غير الخطية والتنبؤ بها بدقة أفضل من النماذج التقليدية

5. التوصيات

1- يوصي البحث باستعمال الأنموذج SVR في عدة مجالات ومنها البيئية مثل ملوثات الهواء وانتاج الطاقة واسعار الاسهم وغيرها.
2- توصي الورقة البحثية باستعمال الأنموذج SVR مع دالة الأساس الشعاعية RBF في حال احتواء البيانات على قيم متطرفة وفي حالة تأخذ البيانات طابع غير خطي معقد كون ان انموذج انحدار متجه الدعمSVR مع الدالة RBF يتمتع بالمرونة في التعامل مع السلاسل الزمنية التي لا تتسم بالخطية والتي تكون فيها نسبة القيم الشاذة كبيرة نسبياً .

تضارب المصالح

لا يوجد تضارب مصالح

الموافقة على النشر

"لقد قرأنا ووافقنا على النسخة النهائية من المخطوطة للنشر".

توافر البيانات والمواد

جميع البيانات قد تم تضمينها ضمن المخطوطة .

مساهمات المؤلفين

م. ق. تنفيذ الدراسة، وتحليل النتائج، وكتابة المقال

ن.ص. تصميم الفكرة

التمويل

لا يوجد تمويل

الشكر والتقدير

شكر لكل الأشخاص أو المؤسسات التي قدّمت لنا مساعدة علمية أو مادية في إنجاز الدراسة وكتابة المقال

6. المصادر (References)

[1] A. J. Smola, B. Sch¨olkopf, and S. Sch¨olkopf,.(2004). "A tutorial on support vector regression *," *Stat. Comput.*, vol. 14, pp. 199–222

[2] A. Setiawan, I. Koprinska, and V. G. S, (2009). "Very short-term electricity load demand forecasting using support vector regression," in *2009 International Joint Conference on Neural Networks*, IEEE, pp. 2888–2894.

[3] O. Kramer and F. Gieseke, (2011). "Short-Term Wind Energy Forecasting Using Support Vector Regression," in *6th International Conference* , pp. 271–280.

[4] P. Meesad and R. I. Rasel, (2013). "Predicting stock market price using support vector regression," in *2013 International Conference on Informatics, Electronics and Vision, ICIEV 2013*. doi: 10.1109/ICIEV.2013.6572570.

[5] J. Ahn, E. Ko, and E. Y. Kim, (2016). *Highway traffic flow prediction using support vector regression and Bayesian classifier*.

[6] M. F. Rohmah, I. K. G. D. Putra, R. S. Hartati, and L. Ardiantoro, (2021). "Comparison Four Kernels of SVR to Predict Consumer Price Index," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan.. doi: 10.1088/1742-6596/1737/1/012018.

[7] Y. Onal, (2022). "Gaussian kernel based SVR model for short-term photovoltaic MPP power prediction," *Computer Systems Science and Engineering*, vol. 41, no. 1, pp. 141–156, doi: 10.32604/csse.2022.020367.

[8] J. Malang, W. S. Yeo, Z. Y. Chua, J. Nandong, and A. Saptoro, (2023). "A comparison study between different kernel functions in the least square support vector regression model for penicillin fermentation process," *MATEC Web of Conferences*, vol. 377, p. 01025 , doi: 10.1051/matecconf/202337701025.

[9] N. T. Hamzeh, Haghiabi Amir and Zahra, Askari and Mohammad, (2025). "A comparative study between time series and soft computing models for river discharge forecasting," *Appl. Water Sci.*, vol. 15, no. 11, p. 273.

[10] J. N. Hussein and S. N. Dahd, (2004). "Diagnosing efficient time series models using ARIMA with a practical application," *Warith Scientific Journal*, vol. 2, no. 1, pp. 192–202, 2020.

[11] W. W. S. Wei, (2006). *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. Boston, MA: Pearson Addison Wesley,

[12] G. P. Zhang, (2004). "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–179, 2003.

[13] T. Hall and K. Rasheed, (2025). "A survey of machine learning methods for time series prediction," *Applied Sciences*, vol. 15, no. 11, p. 5957,.

[14] N. Cristianini and J. Shawe-Taylor, (2000). "An introduction to support vector machines and other kernel-based learning methods," *Cambridge university press*,.

[15] N. J. Nilsson, (1998). "Artificial Intelligence: A New Synthesis," , vol. 1.

[16] Layth S. and Tasnim H., (2024). "Review of Regression Models and Their Characteristics," Nov.

[17] H. Muthiah, U. Sa'adah, and A. Efendi, (2021). "Support Vector Regression (SVR) Model for Seasonal Time Series Data," Brawijaya University Malang, Indonesia, , pp. 3191–3200.

[18] K. Smets, B. Verdonk, and E. M. Jordaan, (2007). "Evaluation of Performance Measures for SVR Hyperparameter Selection," *IEEE*, pp. 637–642.

[19] W. Gani, H. Taleb, and M. Limam, (2010). "Support vector regression based residual control charts," *J. Appl. Stat.*, vol. 37, no. 2, pp. 309–324, Feb. doi: 10.1080/02664760903002667.

[20] M. Alida and M. Mustikasari, (2020). "Rupiah exchange prediction of US dollar using linear, polynomial, and radial basis function kernel in support vector regression," *Jurnal Online Informatika*, vol. 5, no. 1, pp. 53–60.

[21] B. Scholkopf, C. Schafer, and G. Ratsch, (2006). "Learning with Kernels," *The Journal of Machine Learning Research*, vol. 7, pp. 1531–1565.

[22] T. Hofmann, B. Schölkopf, and A. J. Smola, (2008). "Kernel methods in machine learning," Jun. doi: 10.1214/009053607000000677.

[23] V. Vapnik, (1998). "Statistical learning theory Wiley," *New York*, vol. 1, no. 624.

[24] R. Kohavi, (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Montreal, Canada*, vol. 14, no. 2, pp. 1137–1145, [Online]. Available: http://robotics.stanford.edu/~ronnyk

[25] V. Cherkassky and Y. Ma, (2004). "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural networks*, vol. 17, no. 1, pp. 113–126.

[26] S. M. Shaarawy, (2005). "Introduction to Modern Time Series Analysis," *King Abdulaziz University*, pp. 1–446,.

[27] M. V. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. P. Tyukov, T. A. Janovsky, and V. A. evich Kamaev, (2013). "A survey of forecast error measures," *World Appl. Sci. J.*, vol. 24, no. 24, pp. 171–176, doi: 10.5829/idosi.wasj.2013.24.itmies.80032.

[28] U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, (2017). "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Dec. doi: 10.1088/1742-6596/930/1/012002.

[29] T. O. Hodson, (2022). "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," Jul. 19, 2022, *Copernicus GmbH*. doi: 10.5194/gmd-15-5481-.

[30] J. A. Cornell and R. D. Berger, (1987). "Special Topics Factors that Influence the Value of the Coefficient of Determination in Simple Linear and Nonlinear Regression Models," *Phytopathology*, vol. 77, no. 1, pp. 63–70,.

Journal of Al-Turath University College
Open Access Journal
ISSN : 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.45
https://jatuc.uoturath.edu.iq/

Kernel Comparison for Time Series Forecasting Using SVR

Ali Abdullah Saleh[1*], Ahlam Ahmed Juma[2]

[1,2]Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq.

**Correspondence email*: ali.saleh2201@coadec.uobaghdad.edu.iqq

KEYWORDS

Forecasting; Time Series; Support Vector Regression (SVR); Kernel Functions; Radial Basis Function (RBF); Nonlinear Data.

ABSTRACT

Forecasting is considered a crucial topic in time series analysis due to its role in supporting future decision-making through the use of statistical models to predict patterns and trends. This study aimed to employ the Support Vector Regression (SVR) model to forecast nonlinear time series using five kernel functions: Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid, by generating simulated data that mimic air pollution data. This approach was chosen because such environmental time series are often nonlinear and contain outliers, making traditional models less efficient for prediction. Therefore, the performance of the five kernels within the SVR model was compared using statistical evaluation metrics: MSE, MAE, MAPE, and R^2. Simulation results showed the superiority of the RBF kernel over the other kernels across different sample sizes (2000, 4000, 6000, 8000), indicating its better capability to represent nonlinear relationships and improve prediction accuracy. The study concluded that the SVR model can efficiently handle nonlinear data compared to traditional models, with a recommendation to use SVR with the RBF kernel in environmental applications such as air pollutants and energy production.

© 2026.This is an open access article under the CC by licenses http://creativecommons.org/licenses/by/4.0