AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

تصنيف بيانات الأسهم المالية باستخدام تقنية آلة الموجه الداعم في التعليم الاحصائي

محمد هشام ابراهيم	ا.م.د. اسماء غالب جابر
mohammed.ibrahim1201@coadec.uobaghdad.edu.iq	drasmaa.ghalib@coadec.uobaghdad.edu.iq
قسم الاحصاء - كلية الادارة والاقتصاد - جامعة بغداد، بغداد، العراق	

المستخلص

تعد الاسواق المالية بانها ذات حركة مستمرة وديناميكية ومتسارعة مما يؤدي الحاجة الى طرائق ووسائل التقنية في التحليل واتخاذ القرارات وذلك يدفع المستثمرين والمحللين في الاسواق المالية الى استخدام اساليب التحليل المختلفة للوصول الى اساليب التنبؤ بحركة اتجاه السوق المالي لاتخاذ القرارات في الاستثمارات المختلفة لأجل تحديد اتجاه حركة الأسهم اذ تم استخدام طريقة آلة المتجه الداعم svm لتصنيف بيانات الأسهم المالية لتحديد اتجاه الأسهم فيما اذا كانت ذات اتجاه صاعد او هابط. ان الهدف من البحث هو تصنيف بيانات الأسهم المالية باستخدام خمسة متغيرات حيث تم استخدام بيانات المصرف العراقي الاسلامي للاستثمار والتنمية لتوضيح طريقة آلة المتجه الداعم حيث اظهرت النتائج دقة هذه التقنية وكان اداء تقنية آلة الموجه الداعم جيدا.

معلومات البحث

تواريخ البحث:
تاريخ تقديم البحث: 2021/6/21
تاريخ قبول البحث: 2021/9/13
تاريخ رفع البحث على الموقع: 2022/6/25

الكلمات المفتاحية:
آلة المتجه الداعم، تصنيف، الأسهم المالية

للمراسلة:
محمد هشام ابراهيم
mohammed.ibrahim1201@coadec.uobaghdad.edu.iq

DOI: https://doi.org/10.55562/jrucs.v51i1.526

1. المقدمة

تعد الاسواق المالية بكونها ذات اهمية في الاقتصاد والمالية وذلك لما تمتلكه من امكانية في انعاش الاقتصاد وايضا تكون مصدر استثمارات للأموال وان الاسواق المالية تتميز بانها نظام ديناميكي وتحتوي على كميات كبيرة من البيانات وان عملية تصنيف اتجاه حركة الأسهم والتي تتصف بكثافه البيانات غير منظمة وتحتوي على العلاقات الخفية وان حركة اسعار السوق تكون غير عشوائية ديناميكية وغير خطية. ان تصنيف اتجاه حركة الأسهم تتصف بكونها غير خطية.

يهدف البحث الى استخدام طريقة Support vector machines (Svm) وان آلة المتجه الداعم وهي احدى خوارزميات تعلم الآلة machine learning حيث تتميز بقدرتها على التحكم بدالة القرار Decision function وامكانيتها على استخدام دالة اللب Kernel function.

تعد آلة المتجه الداعم Support vector machines بانها واسعة الانتشار وذلك لإمكانيتها على حل المسائل غير الخطية من خلال تحويل البرمجة التربيعية Quadratic programming وكذلك ان آلة المتجه الداعم يكون حلها فريد والافضل عالميا وتم استخدام طريقة العبور الشرعية Cross-validation وذلك لتجنب مشكله الافراط في المطابقة وكذلك عن طريق تحديد اكبر هامش maximum margin للحد الفصل Hyper plane

هنالك العديد من الدراسات السابقة التي تناولت موضوع آلة الموجه الداعم فيما يلي عرض بعض هذه الدراسات ومنها قام الباحثان (غيداء وسجى) سنة 2014 باكتشاف نسبة التصحر في منطقة الحضر بالاعتماد على محتويات الصور.

نشـر الباحثـان (رنا وغيداء) سنـة 2018 دراسـة مقارنـة لخوارزميـات التنقيـب وتحليـل العواطـف وتطبيقاتهـا كما قدم (Rosillo & fuente) سنة 2014 محاكاة لسوق الأسهم باستخدام آلة المتجه الداعم.

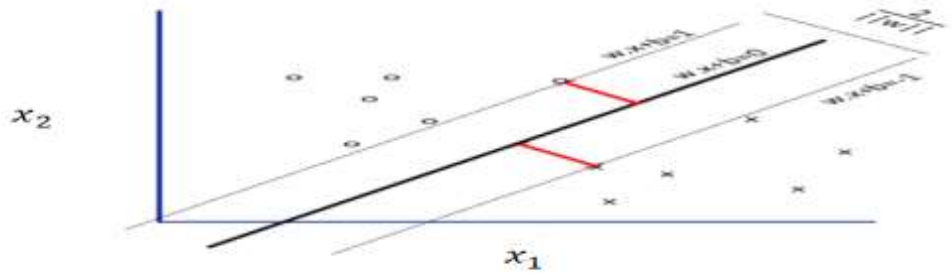
2. الجانب النظري

2.1. آلة المتجه الداعم (Support vector machines)

وهـي مـن تقنيـات التصنيـف فـي تنقيـب البيانـات و احـدى خوارزميـات التعليم المشـرف او الموجه Supervised والتـي تقـوم بتصنيـف البيانـات وذلـك عـن طريـق ايجـاد افضـل مستوى فاصـل hyper plane حيث تـم التوصـل الـى هذه الطريقـة عـام 1992 مـن قبـل Vapnik وتعـد مـن الأنظمـة التـي تستنـد الـى نظريـة التعليـم الاحصائيـة Statistical learning theory وتقـوم طريقـة آلة المتجه الداعـم svm بتصنيـف البيانات الى فئات بواسطة المستوى الفاصل Hyper plane وكذلـك تستخـدم آلة المتجه الداعـم بالإضافة الـى التصنيف في حل مسائل الانحدار Regression. فيما يأتي انواع آلة الموجه الداعم

1. آلة المتجه الداعم الخطي Linear Support vector machines

حيـث يتـم فصـل البيانـات خطيـا بواسطـة الحـد الفاصـل Hyper plane اذا كانت البيانات ثنائية الابعاد D=2. الابعـاد مقسومـة قسمين وكالاتـي (حيث D تمثل عدد الابعاد) اما اذا كانت متعددة الابعاد D>2 فان البيانات تنقسم الى ثلاثة خطوط من الحل الافضل كالاتي:



شكل(1): آلة المتجه الداعم svm الخطي

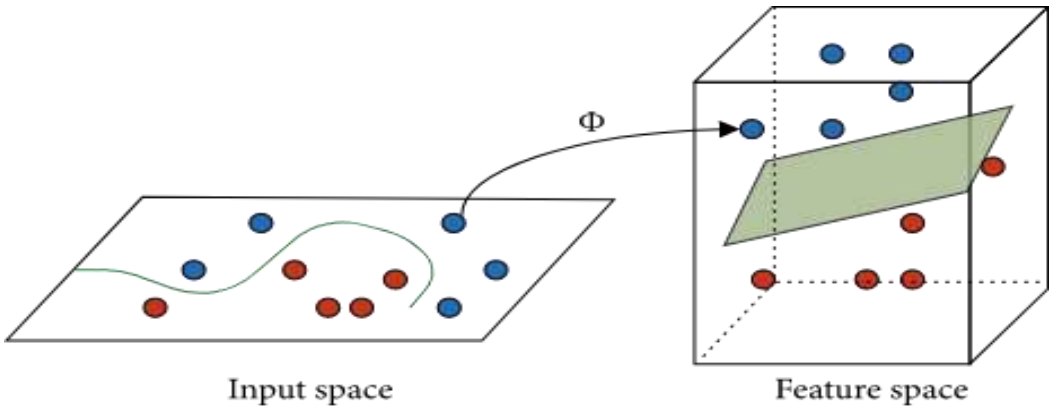
Sourse: Baraithiya, H., & Pateriya, R. K. (2019). Classifiers ensemble for fake review detection. Int. J. Innov. Technol. Explor. Eng, 8(4), 730-736.

x : تمثل متغير بيانات المشاهدات

b : تمثل متغير حد التحيز

2. آلة المتجه الداعم غير الخطي Nonlinear support vector machines

هنالـك حـالات لا يمكـن فصـل البيانـات بصـورة خطيـة لذلـك تستخـدم طريقـة دالة اللـب kernel والتـي تقـوم بتحويل الحآلة الخطية الى حآلة غير خطية وكالاتي:



شكل (2): يمثل آلة المتجه الداعم svm غير الخطي

Source: Hussain, L., Awan, I. A., Aziz, W., Saeed, S., Ali, A., Zeeshan, F., & Kwak, K. S. (2020). Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques. *BioMed research international, 2020*.

ϕ: تمثل تعيين من مساحة الادخال الى مساحة المشاهدات

3. صيغة آلة المتجه الداعم (Formula Support Vector Machines)

ان طريقـة آلة المتجه الداعم Svm تقوم بعملية التصنيف وذلك عن طريق وضع البيانات الى احدى جهتي الحد الفاصل وتهـدف طريقـة آلة المتجه الداعـم الـى بنـاء الحد الفاصل كسطح قرار وذلك ليكـون الهامـش الفاصـل بيـن البيانـات ذات الفئـة الموجبـة والسالبـة لتكـون اكبر ما يمكن (maximized). ان طريقـة آلة المتجه الداعم تسعى الـى ايجاد دالة القرار والغايـة هـي ان نجد دالة القرار التي تقلل من تعميم الخطأ.

تعتمـد آلة المتجه الداعـم علـى عـدة معلمـات التـي تؤثـر بشكل مباشر وغير مباشر فـي ايجـاد الحل النهائي وكذلك تؤثـر فـي دقـة التصنيـف والمستـوى الفاصـل hyper plane ومضاعفـات لاكرانـج Lagrange multipliers

Support Vector: وهي نقاط التي تقع في حدود الخط الفاصل بين الفئتين او صفتين classes و يمكن توضيح الحدود الفاصلة بالمعادلتين الاتيتين:

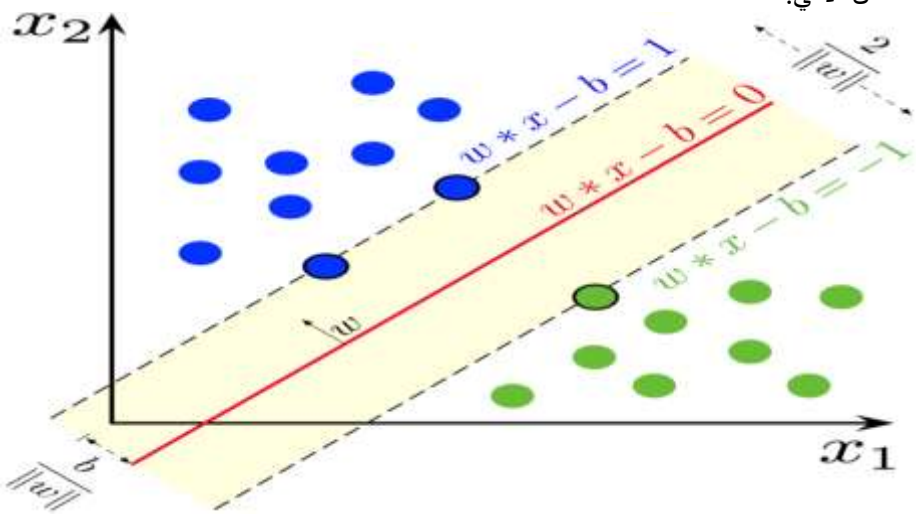
$$w^t x_i + b = +1 \qquad for\ y_i = +1 \qquad i = 1,2 \ldots . N \tag{1}$$

$$w^t x_i + b = -1 \qquad for\ y_i = -1 \qquad i = 1,2 \ldots . N \tag{2}$$

يمكن تمثيل معادلة المستوى الفاصل التي تقع في اقصى كل فئة كالاتي:

$$w^t x_i + b = 0 \tag{3}$$

كما هو موضح في الشكل الاتي:



شكل |(3): مخطط الحدود الفاصلة في Svm

Source: Noyum, V. D., Mofenjou, Y. P., Feudjio, C., Göktug, A., & Fokoué, E. (2021). Boosting the Predictive Accuracy of Singer Identification Using Discrete Wavelet Transform For Feature Extraction.

يمكن تمثيل الحد الفصل الثانوي الاول في آلة المتجه المساند Support victor machine كالاتي:

$$w^t x_i + b = +1 \tag{4}$$

وهي تمثل الحد الفاصل الثانوي الثاني في آلة المتجه المساند وكالاتي:

$$w^t x_i + b = -1 \tag{5}$$

a. معادلة Primal بالصيغة الاتية:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|_2^2 - \sum_{i=1}^{N} \alpha_i (y_i (x^{i^T} w + b) - 1) \tag{6}$$

x^{i^T}: تمثل نقاط بيانات التدريب
y_i: تمثل اشارة نقاط التدريب

وعند فتح المعادلة تكتب كالاتي

$$L(w,b,\alpha) = \frac{1}{2}\, w^T w - \sum_{i=1}^{N} \alpha_i\, y_i\, x^{i^T}\, w + b(\sum_{i=1}^{n} \alpha_i\, y_i\,) + \sum_{i=1}^{n} \alpha_i\, \quad (7)$$

ونطبق الشروط وكالاتي:

.1

$$\frac{\partial L}{\partial \mathrm{w}} = 0 \;\rightarrow\; w^* = \sum_{i=1}^{N} \alpha_i\, y_i\, x^i \quad (8)$$

α_i: تمثل مضاعفات لاكرانج (Lagrange multipliers)

$$0 \le \alpha_i \le c$$

حيث ان

c: وتمثل معيار يقوم بالموازنة بين الخطأ التقديري ومقدار التباعد عن المحور الرئيسي استعمال قيمة 1

.2

$$\frac{\partial L}{\partial \mathrm{b}} = 0 \;\rightarrow\; \sum_{i=1}^{N} y_i\, \alpha_i = 0 \quad (9)$$

وتكون معادلة Dual بالصيغة الاتية:

$$max_{\alpha_1 \ldots\ldots, \alpha_n} L(w^*, b^*, \alpha) \quad (10)$$

Subject to:

$$\alpha_i \ge 0\,, i \in 1\,, \ldots n \;\; and \;\; \sum_{i=1}^{N} \alpha_i\, y_i\, = 0 \quad (11)$$

وعند تعويض المعادلة تكون بالصيغة الاتية

$$max_{\alpha_1 \ldots\ldots, \alpha_n} L(w^*, b^*, \alpha) = max_{\alpha_1 \ldots\ldots, \alpha_n} - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j\, x^{iT} x^j + \sum_{i=1}^{N} \alpha_i \quad (12)$$

Subject to

$$\alpha_i \ge 0\,, i \in 1\,, \ldots N \quad and \;\; \sum_{i=1}^{N} \alpha_i\, y_i\, = 0 \quad (13)$$

$$w^* = \sum_{i=1}^{N} \alpha_i\, y_i\, x^i \quad (14)$$

وتكون مسافة الحد الفاصل hyper plane كالاتي:

$$\frac{2}{\|w^*\|} \quad (15)$$

ولإيجاد حد التحيز وذلك من خلال ما يأتي:

ان نقطة البيانات التي تستوفي المعادلة (9) والتي تاخذ الصيغة الاتية :

$$y(w^t x + b\,) = 1$$

ثم يتم التعويض فيها بالمعادلة (14) لكي نحصل على المعادلة الاتية:

$$y(\sum_{i=1}^{N} \alpha_i\, y_i\, \mathrm{x}_i \mathrm{x} + \mathrm{b}\,) = 1$$

ثم يتم بعد ذلك الضرب بالـ y في المعادلة لكي نحصل على المعادلة الاتية

$$\mathrm{y}^2(\sum_{i=1}^{N} \alpha_i\, y_i\, \mathrm{x}_i \mathrm{x} + \mathrm{b}\,) = y$$

وذلك لان y^2 تمثل واحد 1 كما هو يمثل في المعادلتين الاتية

$$w^t x + b \geq 1 \qquad \text{for} \qquad y_i = +1$$

$$w^t x + b \leq 1 \qquad \text{for} \qquad y_i = -1$$

فإننا نحصل على المعادلة التالية

$$b = y - \sum_i^N \alpha_i \, y_i \, x_i \, . \, x \tag{16}$$

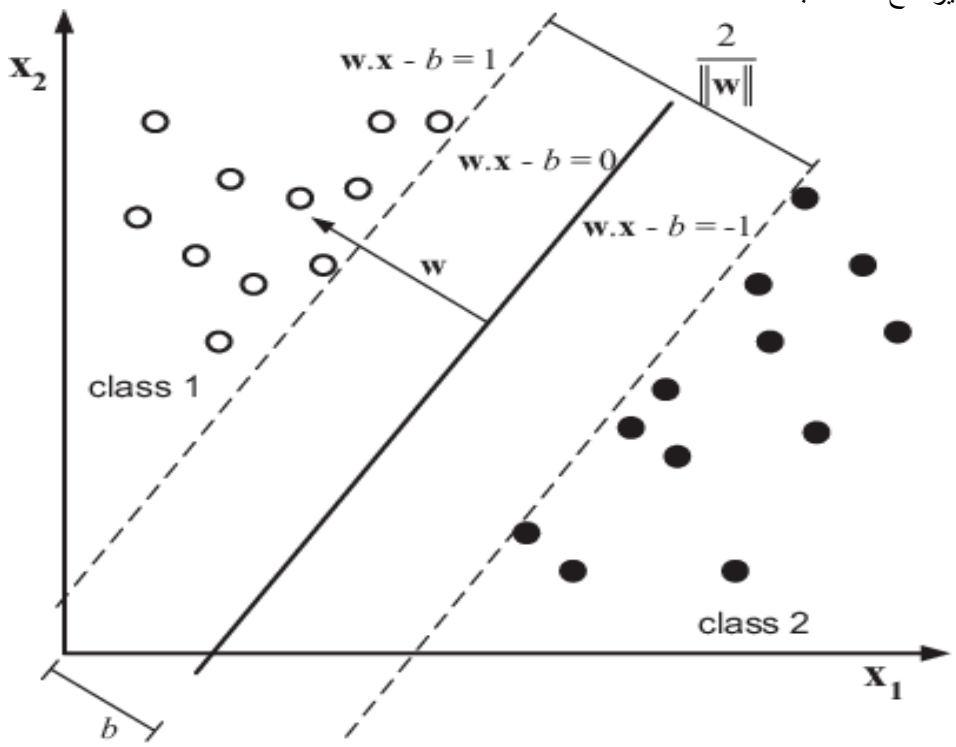
ثم يأخذ المعدل لكل معادلة نحصل على الاتي

$$b = \frac{1}{N} \sum_{i=1}^{N} (y - \sum_i^n \alpha_i \, y_i \, x_i \, . \, x) \tag{17}$$

وبذلك تكون معادلة التصنيف كالاتي:

$$y = (w * x_i + b) \tag{18}$$

فيما يأتي مخطط يوضح آلة المتجه المساند



شكل (4): مخطط يوضح آلة المتجه الداعم

Source: Papadonikolakis, M., & Bouganis, C. S. (2012). Novel cascade FPGA accelerator for support vector machines classification. IEEE transactions on neural networks and learning systems, 23(7), 1040-1052.

3. الجانب التطبيقي (Practical Part)

تم في الجانب التطبيقي استخدام آلة المتجه الداعم svm وكذلك استخدام الرسم لتوضيح كفاءة طريقة آلة المتجه الداعم في تصنيف اتجاهات الأسهم المالية فيما اذا كانت صاعدة او هابطة وتم استخدام بيانات الاوراق المالية للأعوام 2019- 2020 للمصرف العراقي الاسلامي للاستثمار والتنمية حيث استخدمت المتغيرات المتعلقة بالأسهم وذلك للقيام بعملية التصنيف وتكون هذه المتغيرات كالاتي:

1. x_1: ويمثل متغير قيمة السهم
2. x_2 : ويمثل متغير قيمة السهم الابتدائية
3. x_3: ويمثل متغير قيمة السهم الاعلى
4. x_4: ويمثل متغير قيمة السهم الاقل
5. x_5 : ويمثل متغير قيمة السهم الحقيقية
6. y : ويمثل متغير قيمة التغير في السهم

ان تطبيق آلة المتجه الداعم وذلك من خلال استخدام بيانات سوق العراق للأوراق المالية المصرف العراقي الاسلامي للاستثمار والتنمية وهذه البيانات تتكون من خمسة متغيرات وهي قيمة السهم و قيمة السهم الابتدائي وقيمة

السهم الاعلى وقيمة السهم الاقل وقيمة السهم الحقيقي و 105 قيود اذ تم تطبيق آلة المتجه الداعم على بيانات التدريب ثم اختبار الطريقة على بيانات الاختبار وذلك لاختبار دقة التصنيف لطريقة آلة المتجه الداعم وتم تطبيق طريقة آلة المتجه الداعم عن طريق كتابة برنامج باستخدام Matlab 2018b وكذلك تم حساب معايير دقة الطريقة بالاعتماد على مجموعة مؤشرات وهي كالاتي:

1. مؤشر تشخيص العينة غير المعالجة (TP - True Positive)
2. مؤشر تشخيص الصحيح للعينة المعالجة (TN- True Negative)
3. مؤشر عينة غير معالجة تم تشخيصها بشكل صحيح (FP- False Positive)
4. مؤشر تشخيص العينة تم معالجتها بصورة غير صحيحة (FN- False Negative)

وبالاعتماد على هذه المؤشرات نجد المعايير التالية:

1. الدقة: وهو المعيار الذي يمثل النسبة بين النتائج التي تم تشخيصها بشكل صحيح الى النتائج الكلية

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (19)$$

2. النوعية: وهو المعيار الذي يمثل النسبة بين العينة التي تم تشخيصها ومعالجتها بشكل صحيح الى مجموع العينات التي تم تشخيصها بشكل صحيح و معالجتها مع العينات التي تم تشخيصها بشكل غير صحيح

$$SPE = \frac{TN}{TN+FP} * 100\% \quad (20)$$

3. الحساسية: وهو المعيار الذي يمثل النسبة بين العينات غير المعالجة والمشخصة بشكل صحيح الى جميع العينات غير المعالجة والمشخصة بشكل صحيح وبشكل غير صحيح

$$SEN = \frac{TP}{TP+FP} * 100\% \quad (21)$$

4. نسبة الخطاء الموجب: وهو المعيار الذي يمثل النسبة بين العينات غير المعالجة والمشخصة بشكل غير صحيح الى جميع العينات المشخصة بشكل صحيح

$$FPR = \frac{FP}{TN+FP} * 100\% \quad (22)$$

وهنا يتم تطبيق آلة الموجه الداعم وتم الحصول على النتائج وكالاتي:

جدول (1): نتائج Svm باستخدام دالة linear kernel

n	نوع kernel	C	Acc
1	linear	0.001	69.524
2	linear	0.01	69.524
3	linear	0.1	67.619
4	linear	1	69.524
5	linear	10	71.429
6	linear	100	71.429
7	linear	1000	70.476

الجدول 1 يمثل نتائج التصنيف باستخدام Svm مع دالة النوى الخطية حيث تم استخدام عدة قيم للمعلمة c. ويظهر الجدول ان أعلى دقة (71.429%) تحققت عند القيمتين 10 و 100 للمعلمة c.

جدول (2): نتائج Svm باستخدام دالة rbf kernel

n	نوع kernel	C	G	Acc
1	rbf	0.001	0.001	86.667
2	rbf	0.001	0.01	86.667
3	rbf	0.001	0.1	86.667
4	rbf	0.001	1	86.667
5	rbf	0.001	10	86.667
6	rbf	0.001	100	86.667
7	rbf	0.001	1000	86.667
8	rbf	0.01	0.001	86.667
9	rbf	0.01	0.01	86.667
10	rbf	0.01	0.1	86.667

11	rbf	0.01	1	86.667
12	rbf	0.01	10	86.667
13	rbf	0.01	100	86.667
14	rbf	0.01	1000	86.667
15	rbf	0.1	0.001	86.667
16	rbf	0.1	0.01	86.667
17	rbf	0.1	0.1	86.667
18	rbf	0.1	1	86.667
19	rbf	0.1	10	86.667
20	rbf	0.1	100	86.667
21	rbf	0.1	1000	86.667
22	rbf	1	0.001	91.429
23	rbf	1	0.01	91.429
24	rbf	1	0.1	91.429
25	rbf	1	1	91.429
26	rbf	1	10	91.429
27	rbf	1	100	91.429
28	rbf	1	1000	91.429
29	rbf	10	0.001	91.429
30	rbf	10	0.01	91.429
31	rbf	10	0.1	91.429
32	rbf	10	1	91.429
33	rbf	10	10	91.429
34	rbf	10	100	91.429
35	rbf	10	1000	91.429
36	rbf	100	0.001	91.429
37	rbf	100	0.01	91.429
38	rbf	100	0.1	91.429
39	rbf	100	1	91.429
40	rbf	100	10	91.429
41	rbf	100	100	91.429
42	rbf	100	1000	91.429
43	rbf	1000	0.001	91.429
44	rbf	1000	0.01	91.429
45	rbf	1000	0.1	91.429
46	rbf	1000	1	91.429
47	rbf	1000	10	91.429
53	rbf	1000	100	91.429
59	rbf	1000	1000	91.429

جدول (2) يمثل نتائج التصنيف باستخدام Svm مع دالة النوى rbf حيث تم استخدام عدة قيم للمعلمة G و C ولوحظ انه نحصل على اعلى دقة 91.429%) عندما تكون قيمة c من 0.1 الى 1000.

جدول (3): نتائج Svm باستخدام دالة polynomial kernel

n	نوع kernel	C	G	d	R	Acc
1	polynomial	0.001	0.001	1	1	72.381
2	polynomial	0.001	0.01	1	1	71.429
3	polynomial	0.001	0.1	1	1	71.429
4	polynomial	0.001	1	1	1	69.524
5	polynomial	0.001	10	1	1	71.429
6	polynomial	0.001	100	1	1	69.524
7	polynomial	0.001	1000	1	1	70.476
8	polynomial	0.01	0.001	1	1	72.381
9	polynomial	0.01	0.01	1	1	71.429
10	polynomial	0.01	0.01	1	1	69.524

11	polynomial	0.01	0.01	1	1	69.524
12	polynomial	0.01	10	1	1	69.524
13	polynomial	0.01	100	1	1	69.524
14	polynomial	0.01	1000	1	1	70.476
15	polynomial	0.1	0.001	1	1	74.286
16	polynomial	0.1	0.01	1	1	70.476
17	polynomial	0.1	0.1	1	1	66.667
18	polynomial	0.1	1	1	1	68.571
19	polynomial	0.1	10	1	1	69.524
20	polynomial	0.1	100	1	1	66.667
21	polynomial	0.1	1000	1	1	69.524
22	polynomial	1	0.001	1	1	71.429
23	polynomial	1	0.01	1	1	70.476
24	polynomial	1	0.1	1	1	66.667
25	polynomial	1	0.01	1	1	69.524
26	polynomial	1	10	1	1	68.571
27	polynomial	1	100	1	1	66.667
28	polynomial	1	100	1	1	69.524
29	polynomial	10	0.001	1	1	72.381
30	polynomial	10	0.01	1	1	70.476
31	polynomial	10	0.1	1	1	71.429
32	polynomial	10	1	1	1	71.429
33	polynomial	10	10	1	1	70.476
34	polynomial	10	100	1	1	68.571
35	polynomial	10	1000	1	1	69.524
36	polynomial	100	0.001	1	1	71.429
37	polynomial	100	0.01	1	1	66.667
38	polynomial	100	0.01	1	1	68.571
39	polynomial	100	0.1	1	1	65.714
40	polynomial	100	0.1	1	1	70.476
41	polynomial	100	10	1	1	68.571
42	polynomial	100	100	1	1	69.524
43	polynomial	100	1000	1	1	70.476
44	polynomial	1000	0.001	1	1	72.381
45	polynomial	1000	0.01	1	1	71.429
46	polynomial	1000	0.1	1	1	70.476
47	polynomial	1000	1	1	1	71.429
48	polynomial	1000	10	1	1	68.571
49	polynomial	1000	100	1	1	69.524
50	polynomial	1000	1000	1	1	69.524
51	polynomial	0.001	0.001	2	1	67.619
52	polynomial	0.001	0.01	2	1	73.333
53	polynomial	0.001	0.1	2	1	71.429
54	polynomial	0.001	10	2	1	47.619
55	polynomial	0.001	100	2	1	47.619
56	polynomial	0.001	1000	2	1	47.619
57	polynomial	0.01	0.001	2	1	44.762
58	polynomial	0.01	0.01	2	1	71.429
59	polynomial	0.01	0.1	2	1	71.429
60	polynomial	0.01	10	2	1	47.619
61	polynomial	0.01	100	2	1	47.619
62	polynomial	0.01	1000	2	1	47.619

63	polynomial	0.01	0.1	2	1	70.476
64	polynomial	0.1	0.01	2	1	72.381
65	polynomial	0.1	0.1	2	1	71.429
66	polynomial	0.1	10	2	1	47.619
67	polynomial	0.1	100	2	1	47.619
68	polynomial	0.1	1000	2	1	47.619
69	polynomial	1	0.001	2	1	66.667
70	polynomial	1	0.01	2	1	70.476
71	polynomial	1	0.1	2	1	71.429
72	polynomial	1	10	2	1	47.619
73	polynomial	1	100	2	1	47.619
74	polynomial	1	1000	2	1	47.619
75	polynomial	1	0.001	2	1	66.667
76	polynomial	10	0.01	2	1	70.476
77	polynomial	10	10	2	1	47.619
78	polynomial	10	100	2	1	47.619
79	polynomial	10	1000	2	1	47.619
80	polynomial	10	0.1	2	1	71.429
81	polynomial	100	0.001	2	1	68.571
82	polynomial	100	0.01	2	1	74.286
83	polynomial	100	0.1	2	1	71.429
84	polynomial	100	10	2	1	47.619
85	polynomial	100	100	2	1	47.619
86	polynomial	100	1000	2	1	47.619
87	polynomial	1000	0.001	2	1	71.429
88	polynomial	1000	0.01	2	1	70.476
89	polynomial	1000	0.1	2	1	71.429
90	polynomial	1000	10	2	1	47.619
91	polynomial	1000	100	2	1	47.619
92	polynomial	1000	1000	2	1	47.619
93	polynomial	0.001	0.001	3	1	47.619
94	polynomial	0.001	0.01	3	1	47.619
95	polynomial	0.001	0.1	3	1	47.619
96	polynomial	0.001	1	3	1	47.619
97	polynomial	0.001	10	3	1	47.619
98	polynomial	0.001	100	3	1	47.619
99	polynomial	0.001	1000	3	1	47.619
100	polynomial	0.01	0.001	3	1	47.619
101	polynomial	0.01	0.01	3	1	47.619
102	polynomial	0.01	0.01	3	1	47.619
103	polynomial	0.01	0.01	3	1	47.619
104	polynomial	0.01	10	3	1	47.619
105	polynomial	0.01	100	3	1	47.619
106	polynomial	0.01	1000	3	1	47.619
107	polynomial	0.1	0.001	3	1	47.619
108	polynomial	0.1	0.01	3	1	47.619
109	polynomial	0.1	0.1	3	1	47.619
110	polynomial	0.1	1	3	1	47.619
111	polynomial	0.1	10	3	1	47.619
112	polynomial	0.1	100	3	1	47.619
113	polynomial	0.1	1000	3	1	47.619
114	polynomial	1	0.001	3	1	47.619
115	polynomial	1	0.01	3	1	47.619
116	polynomial	1	0.1	3	1	47.619

117	polynomial	1	0.01	3	1	47.619
118	polynomial	1	10	3	1	47.619
119	polynomial	1	100	3	1	47.619
120	polynomial	1	100	3	1	47.619
121	polynomial	10	0.001	3	1	47.619
122	polynomial	10	0.01	3	1	47.619
123	polynomial	10	0.1	3	1	47.619
124	polynomial	10	1	3	1	47.619
125	polynomial	10	10	3	1	47.619
126	polynomial	10	100	3	1	47.619
127	polynomial	10	1000	3	1	47.619
128	polynomial	100	0.001	3	1	47.619
129	polynomial	100	0.01	3	1	47.619
130	polynomial	100	0.1	3	1	47.619
131	polynomial	100	0.1	3	1	47.619
132	polynomial	100	10	3	1	47.619
133	polynomial	100	100	3	1	47.619
134	polynomial	100	1000	3	1	47.619
135	polynomial	1000	0.001	3	1	47.619
136	polynomial	1000	0.01	3	1	47.619
137	polynomial	1000	0.1	3	1	47.619
138	polynomial	1000	1	3	1	47.619
139	polynomial	1000	10	3	1	47.619
140	polynomial	1000	100	3	1	47.619
141	polynomial	1000	1000	3	1	47.619
142	polynomial	0.001	0.001	4	1	47.619
143	polynomial	0.001	0.01	4	1	47.619
144	polynomial	0.001	0.1	4	1	47.619
145	polynomial	0.001	1	4	1	47.619
146	polynomial	0.001	10	4	1	47.619
147	polynomial	0.001	100	4	1	47.619
148	polynomial	0.001	1000	4	1	47.619
149	polynomial	0.01	0.001	4	1	47.619
150	polynomial	0.01	0.01	4	1	47.619
151	polynomial	0.01	0.01	4	1	47.619
152	polynomial	0.01	0.01	4	1	47.619
153	polynomial	0.01	10	4	1	47.619
154	polynomial	0.01	100	4	1	47.619
155	polynomial	0.01	1000	4	1	47.619
156	polynomial	0.1	0.001	4	1	47.619
157	polynomial	0.1	0.01	4	1	47.619
158	polynomial	0.1	0.1	4	1	47.619
159	polynomial	0.1	1	4	1	47.619
160	polynomial	0.1	10	4	1	47.619
161	polynomial	0.1	100	4	1	47.619
162	polynomial	0.1	1000	4	1	47.619
163	polynomial	1	0.001	4	1	47.619
164	polynomial	1	0.01	4	1	47.619
165	polynomial	1	0.1	4	1	47.619
166	polynomial	1	0.01	4	1	47.619
167	polynomial	1	10	4	1	47.619
168	polynomial	1	100	4	1	47.619
169	polynomial	1	100	4	1	47.619
170	polynomial	10	0.001	4	1	47.619

171	polynomial	10	0.01	4	1	47.619
172	polynomial	10	0.1	4	1	47.619
173	polynomial	10	1	4	1	47.619
174	polynomial	10	10	4	1	47.619
175	polynomial	10	100	4	1	47.619
176	polynomial	10	1000	4	1	47.619
177	polynomial	100	0.001	4	1	47.619
178	polynomial	100	0.01	4	1	47.619
179	polynomial	100	0.1	4	1	47.619
180	polynomial	100	0.1	4	1	47.619
181	polynomial	100	10	4	1	47.619
182	polynomial	100	100	4	1	47.619
183	polynomial	100	1000	4	1	47.619
184	polynomial	1000	0.001	4	1	47.619
185	polynomial	1000	0.01	4	1	47.619
186	polynomial	1000	0.1	4	1	47.619
187	polynomial	1000	1	4	1	47.619
188	polynomial	1000	10	4	1	47.619
189	polynomial	1000	100	4	1	47.619
190	polynomial	1000	1000	4	1	47.619
191	polynomial	0.001	0.001	5	1	47.619
192	polynomial	0.001	0.01	5	1	47.619
193	polynomial	0.001	0.1	5	1	47.619
194	polynomial	0.001	1	5	1	47.619
195	polynomial	0.001	10	5	1	47.619
196	polynomial	0.001	100	5	1	47.619
197	polynomial	0.001	1000	5	1	47.619
198	polynomial	0.01	0.001	5	1	47.619
199	polynomial	0.01	0.01	5	1	47.619
200	polynomial	0.01	0.01	5	1	47.619
201	polynomial	0.01	0.01	5	1	47.619
202	polynomial	0.01	10	5	1	47.619
203	polynomial	0.01	100	5	1	47.619
204	polynomial	0.01	1000	5	1	47.619
205	polynomial	0.1	0.001	5	1	47.619
206	polynomial	0.1	0.01	5	1	47.619
207	polynomial	0.1	0.1	5	1	47.619
208	polynomial	0.1	1	5	1	47.619
209	polynomial	0.1	10	5	1	47.619
210	polynomial	0.1	100	5	1	47.619
211	polynomial	0.1	1000	5	1	47.619
212	polynomial	1	0.001	5	1	47.619
213	polynomial	1	0.01	5	1	47.619
214	polynomial	1	0.1	5	1	47.619
215	polynomial	1	0.01	5	1	47.619
216	polynomial	1	10	5	1	47.619
217	polynomial	1	100	5	1	47.619
218	polynomial	1	100	5	1	47.619
219	polynomial	10	0.001	5	1	47.619
220	polynomial	10	0.01	5	1	47.619
221	polynomial	10	0.1	5	1	47.619
222	polynomial	10	1	5	1	47.619
223	polynomial	10	10	5	1	47.619
224	polynomial	10	100	5	1	47.619

225	polynomial	10	1000	5	1	47.619
226	polynomial	100	0.001	5	1	47.619
227	polynomial	100	0.01	5	1	47.619
228	polynomial	100	0.1	5	1	47.619
229	polynomial	100	0.1	5	1	47.619
230	polynomial	100	10	5	1	47.619
231	polynomial	100	100	5	1	47.619
232	polynomial	100	1000	5	1	47.619
233	polynomial	1000	0.001	5	1	47.619
234	polynomial	1000	0.01	5	1	47.619
235	polynomial	1000	0.1	5	1	47.619
236	polynomial	1000	1	5	1	47.619
237	polynomial	1000	10	5	1	47.619
238	polynomial	1000	100	5	1	47.619
239	polynomial	1000	1000	5	1	47.619

جدول (3) يمثل نتائج التصنيف باستخدام Svm مع دالة النوى polynomial حيث تم استخدام عدة قيم للمعلمات r و d و G وC علما ان اعلى دقة (74.286%) قد تحققت عند قيمة C (0.1) و (100) وقيمة G (0.001) و (0.01) وقيمة d (1) و (2).

جدول (4): يمثل اعلى قيم نتائج الدقة في Svm باستخدام دوال kernel

N	Kernel	Acc
1	Linear	71.429
2	Rbf	91.429
3	Polynomial	74.286

من خلال الجداول 1 و2و3 و 4 نلاحظ ان هنالك زيادة كبيرة في الدقة عند استخدام دالة النوى rbf والتي تصل الى (91.429%). وعند التطبيق دالة النوى rbf تم الحصول على نتائج المؤشرات وكالآتي:

جدول (5): مؤشرات Svm باستخدام دالة النوى rbf

TN	TP	FN	FP
41	55	0	9

من هذه المؤشرات يتم حساب اداء آلة المتجه الداعم باستخدام دالة النوى rbf باستخدام المعايير وهي كالاتي

جدول(6): يمثل معايير Svm باستخدام دالة النوى rbf

ACC	SPE	SEN	FPR
91.429	82	85.9375	18

من خلال المعايير المذكورة في جدول (6) نلاحظ أنه تم تصنيف البيانات الى اسهم صاعدة واسهم هابطة وذلك يدل على ان طريقة svm تعمل بشكل جيد جدا.

اظهرت النتائج svm باستخدام دالة النوى rbf kernel ما يلي:

1. ان قيمة المعلمة c لها تأثير في svm عند استخدام دالة النوى rbf kernel
   a. عندما تكون قيمة c من 0.1 الى 1000 نحصل على أعلى دقة
   b. عندما تكون قيمة c من 0.01 الى 0.001 نحصل على أعلى دقة
2. ان قيمة المعلمة g ليس لها تأثير يذكر في svm عند استخدام دالة النوى rbf kernel.

4. الاستنتاجات والتوصيات Conclusions and Recommendation

4.1. الاستنتاجات

1. اظهرت النتائج ان دقة آلة المتجه الداعم كانت جيدة جدا والنتائج كانت فعالة في تصنيف فئات الأسهم الصاعدة والأسهم الهابطة
2. تم استخدام دالة اللب rbf وكانت نتائج الدقة 71.429 وتم استخدام دالة اللب Linear وكانت نتائج الدقة % 91.429. وتم استخدام دالة اللب Polynomial وكانت نتائج الدقة 74.286 وذلك يدل على ان دالة اللب rbf قد حققت اعلى نتائج للدقة.
3. تم الحصول على اعلى قيمة للدقة في دالة اللب rbf وذلك عندما تكون قيمة المعلمة c من 0.1 الى 1000.

4.2. التوصيات

نوصي باستخدام خوارزميات الذكاء الاصطناعي لطرق المقارنة مثل الخوارزمية الجينية، شجرة القرار وغيرها.

5. المصادر

[1] Ma, Y., & Guo, G., (2014). (Eds.). Support vector machines applications (Vol. 649). New York, NY, USA: Springer.

[2] Xydas, E. S., Marmaras, C. E., Cipcigan, L. M., Hassan, A. S., & Jenkins, N. (2013, September). "Forecasting electric vehicle charging demand using support vector machines". 48th International Universities Power Engineering Conference (UPEC) (pp. 1-6). IEEE.

[3] Pham, H. (Ed.). (2006), Springer handbook of engineering statistics. Springer Science & Business Media.

[4] Deng, N., Tian, Y., & Zhang, C., (2012), Support vector machines: optimization based theory, algorithms, and extensions. CRC press.

[5] Steinwart, I., & Christmann, A., (2008), Support vector machines. Springer Science & Business Media.

[6] Wang, L. (Ed.). (2005), Support vector machines: theory and applications (Vol. 177). Springer Science & Business Media.

[7] Stanevski, N., & Tsvetkov, D. (2005), Using support vector machine as a binary classifier. In International Conference on Computer Systems and Technologies–CompSys Tech.

[8] Kecman, V., (2001), Learning and soft computing: support vector machines, neural networks, and fuzzy logic models. MIT press.

[9] Suykens, J., (2003), A (short) introduction to Support Vector Machines and kernel based learning. Katholieke Universitiy Leuven ESANN. -SCD-SISTA, Belgium.

[10] Hastie, T., Tibshirani, R., & Friedman, J., (2009), The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

[11] Maronna, R. Alan julian izenman. (2008), Modern multivariate statistical techniques: regression, classification and manifold learning. Statistical Papers, 52(3), 733.

[12] Ren, R., Wu, D. D., & Liu, T., (2018) Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Systems Journal, 13(1), 760-770.

Journal of AL-Rafidain University College for Sciences (2022); Issue 51; 104-117

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Classification of Financial Stock Data Using the Vector Technology in Statistical Learning

Mohamed H. Ibrahim	Assist. Prof. Dr. Asmaa G. Jaber
mohammed.ibrahim1201@coadec.uobaghdad.edu.iq	drasmaa.ghalib@coadec.uobaghdad.edu.iq
Department of Statistics - College of Administration and Economics - University of Baghdad, Baghdad - Iraq.	

Article Information

Article History:
Received: June, 21, 2021
Accepted: September, 13, 2021
Available Online: June, 25, 2022

Keywords:
Support vector machine, classification, financial stock

Correspondence:
Mohamed H. Ibrahim
mohammed.ibrahim1201@coadec.uobaghdad.edu.iq

Abstract

The financial markets are considered to be of continuous, dynamic and accelerating movement, which leads to the need for techniques methods and means for analysis and decision-making, which push investors and analysts in the financial markets to use various analysis methods and forecasting approaches for the direction of the financial market movement to make decisions for different investments in order to determine the direction of the stock movement, thus the Support Vector Machine method (SVM) is used to classify financial stock data to determine the direction of the stock, whether it is in an upward or downward trend. The aim of the research is to classify the financial stock data using five variables, where data from The Iraqi Islamic Bank for Investment and Development is used to clarify the method of the Support Vector Machine. The results showed accuracy of this technique and the performance of the Support Vector Machine is good.

DOI: https://doi.org/10.55562/jrucs.v51i1.526