پژوهشنامه اقتصاد انرژی ایران
سال هفتم، شماره 25، زمستان 1396، صفحات 60-41

پیش‌بینی قیمت روزانه نفت خام برنت با ترکیب روش‌های آنالیز مؤلفه‌های اصلی و رگرسیون بردار پشتیبان

الهام حاجی‌کرم[1]
رویا دارابی[2]

تاریخ پذیرش:1397/06/20 تاریخ دریافت:1396/07/24

چکیده

پیش‌بینی روند قیمت نفت خام و نوسانات آن همواره یکی از چالش‌های پیش روی معامله‌گران در بازارهای نفتی بوده است. این مقاله به پیش‌بینی قیمت روزانه نفت خام برنت با یک مدل ترکیبی پیشنهادی می‌پردازد. نمونه آماری قیمت روزانه نفت خام برنت دریای شمال از ژوئیه سال 2008 تا ژوئیه سال 2016 می‌باشد که از میان کل قیمت‌های روزانه نفت در تمام بازارهای نفتی انتخاب شده است. در این پژوهش، برای پیش‌بینی مدلی از ترکیب روش‌های آماری و هوش مصنوعی (PCA-SVR) ارائه می‌شود. با توجه به اثبات برتری دقت پیش‌بینی مدل رگرسیون بردار پشتیبان (SVR)[3] نسبت به سایر روش‌های پیش‌بینی در مطالعات گذشته هدف اصلی در این پژوهش، بهبود پیش‌بینی رگرسیون بردار پشتیبان با استفاده از پیش‌پردازش اولیه داده‌ها به وسیله آنالیز مولفه-های اصلی (PCA)[4] است. جهت انجام پژوهش پس از انجام آزمون مانایی، با استفاده از آنالیز مولفه‌های اصلی متغیرهای ورودی را به مولفه‌های اصلی که کل پراکندگی داده‌ها را پوشش می‌دهد تبدیل نموده و به عنوان ورودی برای مدل پیش‌بینی در نظر گرفتیم. سپس با استفاده از مدل رگرسیون بردار پشتیبان و شبیه‌سازی آن در نرم‌افزار متلب اقدام به پیش‌بینی قیمت روزانه نفت خام برنت نمودیم. به منظور مقایسه عملکرد مدل‌های SVR و PCA-SVR از آزمون مقایسات زوجی استفاده نمودیم. نتیجه پژوهش بیانگر این موضوع بود که پیش‌پردازش اولیه به وسیله آنالیز مولفه‌های اصلی بر روی داده‌ها باعث کاهش خطای مدل پیشنهادی گردیده است.

طبقه‌بندی JEL: G10,G19,G17,C02

کلیدواژه‌ها: آنالیز مولفه‌های اصلی، رگرسیون بردار پشتیبان، مدل ترکیبی، نفت خام

1.دانشجوی کارشناسی ارشد مدیریت بازرگانی- مالی دانشگاه آزاد اسلامی واحدعلوم وتحقیقات (نویسنده مسئول)
elhamhajikaram@srbiau.ac.ir

2.دانشیار گروه حسابداری دانشگاه آزاد اسلامی واحد تهران جنوب
royadarabi110@yahoo.com

3. Support Vector Regression
4. Principal Components Analyses

Archive of SID
www.SID.ir

1. مقدمه

ایران یکی ازکشورهای بزرگ دارای ذخایر غنی نفتی درجهان است و درآمدهای حاصل از نفت و محصولات نفتی نقش بسزائی در سطح معیشتی مردم و همچنین تصمیم‌گیری‌های مسئولان کشور دارد. مسئله پیش‌بینی قیمت نفت از دیرباز مورد توجه اقتصاددانان قرارداشته و بدین منظور ازمدل‌های خطی و غیرخطی زیادی استفاده شده است. اگر پیش‌بینی قیمت نفت به درستی اطلاعات مربوط به روند آتی این متغیر را منعکس کند، آنگاه می‌توان از آن به عنوان یک متغیر پیشرو برای پیش‌بینی نوسان فعالیت‌های اقتصادی استفاده‌کرد. بنابراین باید پیش بینی‌ای را مورد توجه قرار داد که با دقت بیشتری صورت گیرد و نسبت به نتایج واقعی مشاهده شده خطای کمتری داشته باشد.

اهمیت روزافزون پیش‌بینی متغیرهای کلان اقتصادی از یک سو و کمبود مدل‌های ساختاری مناسب برای پیش‌بینی از سوی دیگر باعث ایجاد مدل‌های سری‌های زمانی شده است. هدف اصلی مدلسازی سری‌های زمانی، دادن نظم خاص به مشاهدات وابسته به زمان است تا بر اساس آن بتوان آینده را پیش‌بینی کرد؛ به عبارت دیگر، مهمترین هدف تجزیه و تحلیل سری‌های زمانی یافتن مدل تغییرات و پیش‌بینی آینده آن است (فرج‌زاده، 1394).

پیش‌بینی سری‌های زمانی از مهم‌ترین روش‌های پیش‌بینی است که برمشکلات روش-های ساختاری غلبه می‌کند و درآن تلاش می‌شود متغیرها را تنها با استفاده از اطلاعات موجود در مدل، مقادیر گذشته‌شان و مقادیرفعلی پیش‌بینی و مدل‌سازی‌کرد. این عمل را می‌توان با مدل‌های ساختاری که ماهیتاً چندمتغیره هستند و تلاش می‌کنند تغییرات یک متغیر را بر مبنای تغییرات مقادیر جاری و گذشته سایر متغیرها (متغیرهای توضیحی) توضیح دهند، مقایسه نمود (بروکز[1]، 1389).

محمدی (1394) در مقاله‌ای تحت عنوان " پیش‌بینی قیمت هفتگی نفت خام از طریق مدل ترکیبی ماشین بردار پشتیبان و خودرگرسیو میانگین متحرک انباشته" به منظور

1 .Brooks

Archive of SID

www.SID.ir

پیش‌بینی قیمت هفتگی نفت خام برنت به عنوان یک نفت شاخص با توجه به دشوار بودن شناسایی دقیق الگوهای خطی و غیرخطی در سری‌های زمانی اقتصادی و مالی از ترکیب مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان استفاده نمود. و نهایتا، نتایج بیانگر این موضوع بودند که مدل ترکیبی خطای کمتری در پیش‌بینی قیمت نفت خام نسبت به کاربرد مجزای مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان دارد.

در این پژوهش نیز با توجه به اثبات برتری دقت پیش‌بینی مدل رگرسیون بردار پشتیبان نسبت به سایر روش‌های پیش‌بینی، از آنجایی که متغیرهای زیادی می‌توان شناسایی کرد که بر روی قیمت نفت اثر گذار باشد و همچنین به دلیل این که سری زمانی قیمت نفت از یک الگوی خطی پیروی نمی‌کند، برای کاهش خطای پیش‌بینی، ترکیبی از روش‌های آماری و هوش مصنوعی استفاده شده است بدین معنا که جهت بهبود پیش‌بینی رگرسیون بردار پشتیبان از پیش‌پردازش داده‌ها به وسیله آنالیز مولفه‌های اصلی استفاده شده است. بنابراین فرضیه پژوهش به صورت زیر بیان می‌گردد:

پیش‌بینی مدل ترکیبی PCA-SVR نسبت به پیش‌بینی مدل SVR خطای کمتری دارد. ابتدا به مبانی نظری و سپس به پیشینه پژوهش در داخل و خارج پرداخته و در ادامه روش‌شناسی پژوهش ارائه شده است. بخش پایانی مقاله نیز به آزمون فرضیه، نتیجه‌گیری و پیشنهادها اختصاص دارد.

2. مبانی نظری و مروری بر پیشینه پژوهش

2-1. تجزیه و تحلیل مولفه‌های اصلی

تحلیل مولفه اصلی به تبیین ساختار واریانس-کوواریانس به کمک چند ترکیب خطی از متغیرهای اصلی سر و کار دارد. اهداف کلی آن عبارتند از 1- کاهش حجم داده‌ها 2- تعبیر و تفسیر آن‌ها.

Archive of SID

www.SID.ir

اگر چه برای مطالعه تغییرپذیری کل سیستم، p مولفه نیاز است، ولی اغلب بیشتر این تغییرپذیری را می‌توان با تعداد کمتر مثلا k مولفه اصلی بیان نمود. در این صورت میزان اطلاعی که در k مولفه وجود دارد (تقریبا) میزان اطلاع در p متغیر اولیه است. بنابراین k مولفه اصلی را می‌توان به جای p متغیر اولیه به کار برد و مجموعه داده‌های اولیه که شامل n اندازه روی p متغیر است به مجموعه‌ای از داده‌های شامل n اندازه درمورد k مولفه اصلی کاهش داد.

آنالیز مولفه‌های اصلی وسیله‌ای برای رسیدن به هدف است تا این‌که خودش هدف باشد، زیرا اغلب به عنوان مراحل میانی در وضعیت‌های بزرگ‌تر به کار می‌آید. برای مثال مولفه‌های اصلی می‌توانند ورودی‌های یک رگرسیون چندگانه یا تحلیل خوشه‌ای باشند (جانسون و ویچرن[1]، 1393).

مولفه‌های اصلی از نظر جبری ترکیبات خطی ویژه p متغیر تصادفی $X_1, X_2, ..., X_p$ هستند. این ترکیبات خطی از نظر هندسی انتخاب یک دستگاه مختصات جدید را نشان می‌دهند که از دوران دستگاه اولیه با $X_1, X_2, ..., X_p$ به عنوان محورهای مختصات به دست می‌آیند. محورهای جدید جهت‌ها را با بیشترین تغییرپذیری نشان می‌دهند و بیان ساده‌تر و ممسک‌تری از ساختمان کوواریانس‌ها را فراهم می‌کنند.

چنان‌که ملاحظه خواهیم نمود مولفه‌های اصلی تنها به ماتریس کوواریانس Σ (یا ماتریس همبستگی ρ) $X_1, X_2, ..., X_p$ مربوط می‌شوند. برای گسترش آن‌ها، گسترش نرمال چند متغیری لازم نیست. از سوی دیگر مولفه‌های اصلی جامعه‌های نرمال چند متغیری، تعابیر مفیدی بر حسب بیضوی‌های چگالی ثابت دارند. علاوه براین، در جامعه نرمال چند متغیری، می‌توان به استنباط‌هایی از مولفه‌های نمونه دست یافت.

فرض کنید بردار تصادفی $X' = [X_1, X_2, ..., X_p]$ دارای ماتریس کوواریانس Σ با مقادیر ویژه $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_p \geq 0$ است.

ترکیبات خطی زیر را در نظر می‌گیریم:

1. Johnson and Wichern (2015)

Archive of SID

www.SID.ir

$$Y_1 = e_1'X = e_{11}X_1 + e_{21}X_2 + ... + e_{p1}X_p$$

$$Y_2 = e_2'X = e_{12}X_1 + e_{22}X_2 + ... + e_{p2}X_p \quad (1)$$

$$Y_p = e_p'X = e_{1p}X_1 + e_{2p}X_2 + ... + e_{pp}X_p$$

در این صورت داریم:

$$Var(Y_i) = e_i' \sum e_i \qquad i = 1,2,...,p \quad (2)$$

$$Cov(Y_i, Y_k) = e_i' \sum e_k \qquad i,k = 1,2,...,p \quad (3)$$

مولفه‌های اصلی آن، ترکیبات خطی ناهمبسته $Y_1, Y_2, ..., Y_P$ هستند که واریانس‌های آن‌ها تا حد ممکن بزرگ می‌باشند.

نخستین مولفه اصلی یک ترکیب خطی با واریانس ماکزیمم است. یعنی $Var(Y_i) = e_1' \sum e_1$ را ماکزیمم می‌کند. واریانس دومین مولفه اصلی، کمتر از مولفه نخست می‌باشد. واریانس مولفه ها کاهش می‌یابد تا این‌که مولفه p ام کمترین واریانس را دارا می‌باشد (جانسون و ویچرن ،1393).

از تجزیه و تحلیل مولفه‌های اصلی نتیجه زیر حاصل می‌شود:

اگر $Y_1 = e_1'X, Y_2 = e_2'X, ..., Y_p = e_p'X$ ، مولفه‌های اصلی به‌دست آمده ازماتریس کوواریانس Σ باشد آن‌گاه

$$\rho_{Y_i, X_k} = \frac{e_{ki}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i,k = 1,2,...,p \quad (4)$$

ضرایب همبستگی بین مؤلفه‌های Y_i و متغیرهای X_k است. در این جا

$(\lambda_1, e_1), (\lambda_2, e_2), ..., (\lambda_p, e_p)$ زوج‌های مقدار ویژه-بردار ویژه $\sum$ هستند (جانسون و ویچرن، 1393).

Archive of SID

www.SID.ir

2-2. ماشین بردار پشتیبان

روش ماشین بردار پشتیبان یکی از روش‌های یادگیری ماشینی است که بر مبنای تئوری یادگیری آماری وپنیک[1] در دهه 90 میلادی توسط وی و همکارانش ارائه گردید. در SVM از اصول کمینه سازی ریسک ساختاری(SRM)[2] استفاده شده است، در حالی که سایر روش‌ها از اصول کمینه‌سازی ریسک تجربی(ERM)[3] بهره می‌برند (لیپو[4]، 2005).

از ماشین بردار پشتیبان به طور کلی در مسائل طبقه‌بندی دو یا چند کلاسه و رگرسیون استفاده می‌شود. مانند بسیاری از روش‌های یادگیری ماشینی، در ماشین بردار پشتیبان نیز فرآیند ساخت مدل شامل دو مرحله آموزش و آزمایش می‌باشد. در انتهای فاز آموزش قابلیت تعمیم‌یابی مدل آموزش داده شده با استفاده از داده‌های آزمایش ارزیابی می‌شود (لیپو، 2005).

به طور خلاصه سازوکار اصلی SVM در حل مساله رگرسیون به صورت زیر بیان می‌شود:

1- ماشین بردار پشتیبان، تابع رگرسیون را با به کارگیری یک دسته تابع خطی تخمین می‌زند.

2- ماشین بردارپشتیبان، عملیات رگرسیون را با تابعی که انحراف از مقدار واقعی در آن به میزان کمتر از ε مجاز است انجام می‌دهد.(تابع ضرر[5])

3- ماشین بردارپشتیبان، با کمینه‌کردن ریسک ساختاری، بهترین جواب را ارائه می‌دهد (سانچز[6]، 2003).

در ماشین بردار پشتیبان برای حل مساله رگرسیون یک تابع خطی به شکل $f(x)=<w.x>+b$ ، که بر روی یک مجموعه شامل k نمونه مانند

1 . Vapnik
2. Structural Risk Minimization
3 . Emprical Risk Minimization
4 . Lipo (2005)
5 . Loss Function
6 . Sanches

Archive of SID

www.SID.ir

$\{(x_1,y_1),...,(x_k,y_k)\} \in R^n, y \in R$ است، سعی در تخمین مقادیر خروجی بر مبنای مقادیر ورودی دارد. در رابطه فوق x بردار مقادیر ورودی و $(w,b) \in R^n \times R$ مولفه‌های کنترل کننده تابع f هستند. $< w.x >$ نشانگر ضرب داخلی می‌باشد. برای حل مساله رگرسیون، تابع ضرر vapnik که در آن حداقل خطایی به میزان ε قابل صرفنظر کردن است، مورد استفاده قرار می‌گیرد. این تابع ضرر را می‌توان به صورت ذیل نمایش داد.

$$L_\varepsilon\left(X,Y,f\left(x\right)\right) = \begin{cases} |y - f(x)| - \varepsilon & |y - f(x)| \le \varepsilon \\ 0 & \text{در غیر اینصورت} \end{cases} \quad (5)$$

$L_\varepsilon(y)$ معرف تابع ضرر و ε خطای مجاز در تابع ضرر می‌باشد. مولفه‌های کنترل کننده تابع رگرسیون بهینه با حل مساله بهینه‌سازی زیر به‌دست می‌آیند.

$$minimize\ \phi\left(w,\theta^*,\theta\right) = \frac{\|w\|^2}{2} + c\left(\sum \theta_j^* + \sum \theta\right)$$

subject to:

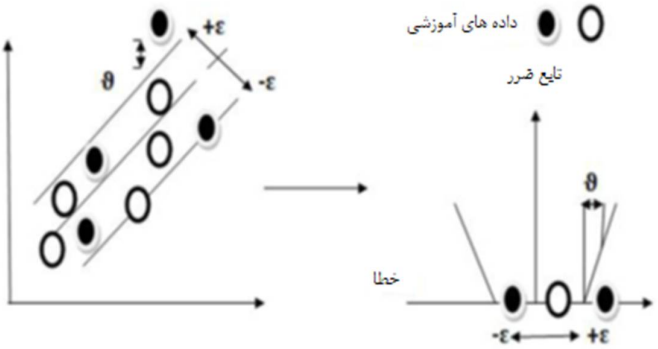
$$\begin{aligned} &\left(\left(w.x_j\right)+b\right) - y_j \le \varepsilon + \theta_j^* \\ &y_j - \left(\left(w.x_j\right)+b\right) \le \varepsilon + \theta_j \\ &\theta_j, \theta_j^* \ge 0 \end{aligned} \quad (6)$$

در رابطه فوق θ و θ^* متغیرهای slack هستند. این متغیرها به همراه تابع ضرر در شکل (1) نشان داده شده‌اند (گان[1]، 1998).

1 . Gunn (1998)

Archive of SID

www.SID.ir



شکل(1): تابع ضرر وپنیک و متغیرهای slack

منبع: گان (1998)

برای حل مساله بهینه‌سازی فوق، به کمک تئوری لاگرانژ، تابع لاگرانژ به صورت زیر نوشته می‌شود.

$$L(a^*,a) = -\varepsilon\sum_{i=1}^{k}(a_i^* + a_i) + \sum_{i=1}^{k} y_i(a_i^* - a_i) - \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(a_i^*,a_i)(a_j^*,a_j)(x_i,x_j) \quad (7)$$

با بیشینه شدن تابع فوق تحت قیدهای زیر، مقادیر ضرایب a و a^* به‌دست می‌آیند. این ضرایب، ضرایب لاگرانژ نامیده می‌شوند.

$$\begin{cases} \sum a_i^* = \sum a_i \\ 0 \le a_i^* \le C \\ 0 \le a_i \le C \end{cases} \qquad \text{For } i = 1,2,3,\ldots,k \quad (8)$$

Archive of SID

www.SID.ir

مساله بهینه سازی فوق می‌تواند به کمک روش های برنامه ریزی درجه دو (QP) حل شود، در نتیجه رسیدن به اکسترمم کلی نیز قطعی خواهد بود وخطر به دام افتادن در اکسترمم محلی وجود ندارد. داده هایی‌که ضرایب لاگرانژ متناظر با آنها غیرصفر باشد، به عنوان بردار پشتیبان شناخته می‌شوند. از نظر هندسی این داده‌ها دارای خطای پیش‌بینی بزرگتر از $\mp\varepsilon$ هستند. بنابراین بردارهای پشتیبان درون باند $\mp\varepsilon$ قرار نمی‌گیرند. بنابراین مقدارε تعداد بردارهای پشتیبان را کنترل می‌کند (گان، 1998).

به کمک ضرایب لاگرانژ و بردارهای پشتیبان، مولفه‌های کنترل کننده پاسخ بهینه به صورت زیر محاسبه می‌شود.

$$w_0 = \sum\left(a_i^* - a_i\right)x_i \tag{9}$$

$$b_0 = -\left(\frac{1}{2}\right).w_0.\left[x_r + x_s\right] \tag{10}$$

$$f\left(x\right) = \sum\left(a_i^* - a_i\right)\left(x_i - x\right) + b_0 \tag{11}$$

در رابطه فوق x_S و x_r دو بردار پشتیبان هستند.

برای ساخت مدل ماشین بردار پشتیبان، مولفه‌های C و ε توسط کاربر تعریف می-شوند. مولفه C یک مولفه تنظیمی است و می‌تواند مقادیر صفر تا بی نهایت را بپذیرد. نقش این مولفه ایجاد تعادل میان کمینه کردن ریسک تجربی وبیشینه کردن قابلیت تعمیم یابی است. مولفه ε نیز میتواند مقادیر صفر تا بی نهایت را بپذیرد. مقدار این مولفه در وضعیت بردارهای پشتیبان ودر نتیجه کارایی مدل بسیار موثر است (گان، 1998).

مساله رگرسیون خطی در SVM به آسانی قابل گسترش به رگرسیون غیرخطی است. بدین منظور از توابع کرنل استفاده می‌شود. تاکنون کرنل‌های گوناگونی شناخته شده‌اند. از جمله کرنل‌های چند جمله‌ای و پایه شعاعی(RBF) بدین ترتیب در حالت

Archive of SID

www.SID.ir

رگرسیون غیرخطی در SVM مولفه‌های کنترل کننده تابع بهینه با روابط زیر محاسبه می-شوند (سانچز، 2003).

$$W.X = \sum\left(a_i^* - a_i\right)K\left(x_i, x\right) \quad (12)$$

$$b_0 = -\frac{1}{2}\sum\left(a_i^* - a_i\right)\left[K\left(x_r, x_i\right) + K\left(x_s, x_i\right)\right] \quad (13)$$

در این روابط $K(x_i, x)$ نشانگر تابع کرنل می‌باشد.

2-3. پیشینه پژوهش

همانطور که در مقدمه نیز اشاره شد محمدی (1394) در مقاله‌ای تحت عنوان " پیش‌بینی قیمت هفتگی نفت خام از طریق مدل ترکیبی ماشین بردار پشتیبان و خودرگرسیو میانگین متحرک انباشته" به پیش‌بینی قیمت هفتگی نفت خام برنت به عنوان یک نفت شاخص پرداخته است، وی با توجه به دشوار بودن شناسایی دقیق الگوهای خطی و غیرخطی در سری‌های زمانی اقتصادی و مالی از ترکیب مدل‌های خودرگرسیو میانگین متحرک انباشته، با این پیش فرض که سری زمانی دارای الگوی خطی می‌باشد و ماشین بردار پشتیبان که توانایی بالایی در مدل‌سازی الگوهای غیرخطی دارد به‌منظور افزایش دقت پیش‌بینی استفاده نموده است. با توجه به آزمون مقایسه زوجی معیارهای ارزیابی عملکرد ریشه میانگین مجذور خطا و میانگین قدرمطلق درصد خطا که حاصل مقادیر پیش‌بینی شده و مقادیر واقعی هر یک از مدل‌ها هستند، نتایج بیانگر این موضوع بودند که در بیشتر موارد مدل ترکیبی خطای کمتری در پیش‌بینی قیمت نفت خام نسبت به کاربرد مجزای مدل‌های خودرگرسیو میانگین متحرک انباشته و ماشین بردار پشتیبان دارد.

فغانی ماکرانی (1395) در مقاله‌ای تحت عنوان "پیش‌بینی مدیریت سود مبتنی بر مدل جونز تعدیل شده با استفاده از مدل شبکه عصبی مصنوعی و الگوریتم ژنتیک" از دو مدل شبکه عصبی مصنوعی و مدل ترکیبی الگوریتم ژنتیک – شبکه عصبی به عنوان الگوی

Archive of SID

www.SID.ir

موفق جهت پیش‌بینی مدیریت سود استفاده کرده است و هدف پیش‌بینی مدیریت سود از طریق اقلام تعهدی اختیاری مبتنی بر جونز تعدیل شده در بورس اوراق بهادار تهران بوده است. نمونه مورد استفاده در این پژوهش شامل 570 سال-شرکت بین سال‌های 1387 تا 1392 می‌باشد. یافته‌های پژوهش نشان داد که شبکه عصبی مصنوعی از توانایی بالایی در پیش‌بینی مدیریت سود، نسبت به مدل خطی جونز تعدیل شده برخوردار است. همچنین یافته‌ها حاکی از آن است که الگوریتم ژنتیک به عنوان مدل بهینه‌ساز می‌تواند در افزایش توان پیش‌بینی شبکه عصبی مصنوعی و بهینه کردن وزن‌های آن برای پیش‌بینی مدیریت سود مبتنی بر مدل جونز تعدیل شده تأثیر بسزایی داشته باشد.

زی و همکاران (2006)[1] در مقاله‌ای تحت عنوان " یک روش جدید برای پیش‌بینی قیمت نفت خام بر اساس ماشین‌های بردار پشتیبان" روش جدیدی برای پیش‌بینی قیمت نفت خام پیشنهاد می‌کند. در این مقاله به ارزیابی توانایی پیش‌بینی روش SVM پرداخته شده و عملکرد آن را با روش‌های ARIMA و BPNN مقایسه می‌کند. ابتدا تئوری پایه‌ای را معرفی می‌کند. بزرگترین مزیت SVM این واقعیت است که یک راه حل جهانی است که می‌توان از آن استفاده کرد و در پیش‌بینی سری‌های زمانی مناسب است. نتایج پژوهش نشان می‌دهد که SVM در پیش‌بینی قیمت نفت خام بهتر از دو روش دیگر عمل می‌کند و یک کاندید نسبتا خوب برای پیش‌بینی قیمت نفت خام است.

لو (2013)[2] در مقاله‌ای تحت عنوان "پیش‌بینی شاخص با مدل هیبریدی آنالیز مولفه‌های مستقل و رگرسیون بردارپشتیبان با بهینه‌سازی حرکت تجمعی ذرات" به پیش‌بینی شاخص بورس‌های چین، تایوان و هند پرداخت. وی در این مقاله بیشترین، کم‌ترین و مقدار پایانی شاخص و مقدار باز شدن شاخص در روز جاری را به عنوان متغییر جهت پیش‌بینی در نظر گرفت. در پایان به این نتیجه رسید که مدل ترکیبی شاخص را با خطای کمتری نسبت به مدل SVR پیش‌بینی می‌کند.

1. Xie, et al. (2006)
2 .Lu (2013)

Archive of SID

www.SID.ir

لی و همکاران (2014)[1] در مقاله‌ای تحت عنوان " پیش‌بینی نفت خام با مدل‌های نوفه‌زدایی شده چندمقیاسه" اقدام به پیس‌بینی قیمت نفت خام نمودند. آن‌ها از مدل ARIMA جهت پیش‌بینی قیمت نفت استفاده نموده و چون این مدل یک مدل خطی می-باشد برای پیش‌بینی بخش غیرخطی سری زمانی از مدل ماشین بردار پشتیبان استفاده نمودند. در پژوهش جهت جلوگیری از تاثیر نوسان‌های قیمت نفت از نوفه‌زدایی موجک استفاده نمودند که نتایج حاکی از بهبود قدرت پیش‌بینی مدل بوده است.

کوازی منسا (2016)[2] در مقاله خود تحت عنوان " مدل سازی باکس-جنکینز و پیش بینی قیمت نفت خام برنت " پویایی ماهانه قیمت نفت برنت در دو دهه گذشته با استفاده از تکنیک جعبه جنکینز ARIMA را بررسی و نشان داد که چنین مدلی قادر به بررسی نوسانات ذاتی در قیمت نفت خام برای پیش‌بینی دقیق آن نیست. او داده‌ها را به دو قسمت تقسیم نموده است. از داده‌های هفده سال اول برای مدل‌سازی و داده‌های سه سال بعد برای بررسی اعتبار دقت مدل پیش‌بینی استفاده نموده است. عملکرد مدل‌های مختلف مقایسه شدند و نهایتا نشان می‌دهد که در مقایسه با مدل قبلی یک مدل موجز (ARIMA) بهترین مدل پیش‌بینی برای قیمت نفت است.

همانطور که ملاحظه می‌شود مطالعات انجام شده روش‌های مختلفی برای پیش‌بینی انواع شاخص‌ها و یا قیمت هفتگی نفت خام ارائه نموده‌اند ولی در هیچ کدام از روش‌ها با استفاده از پیش پردازش اولیه داده‌ها، پیش‌بینی روزانه قیمت نفت خام انجام نشده است. در این پژوهش از روش تجزیه و تحلیل مولفه‌های اساسی برای پالایش اولیه داده‌ها استفاده شده سپس به کمک مدل رگرسیون بردار پشتیبان که نوع خاصی از ماشین‌های بردار پشتیبان می‌باشد، به پیش‌بینی قیمت روزانه نفت خام پرداخته می‌شود.

1 . Li, et al.(2014)
2 .Kwasi Mensah (2016)

Archive of SID

www.SID.ir

3. روش شناسی پژوهش

در پژوهش حاضر تلاش بر این است که مدلی از ترکیب روش‌های آماری و هوش مصنوعی جهت پیش‌بینی قیمت روزانه نفت خام برنت ارائه شود.

در این پژوهش جهت پیش‌بینی قیمت روزانه نفت خام برنت، مطابق با پژوهش لو (2013)، از 4 متغیر که شامل مقدار آغازین، بیشترین مقدار، کمترین مقدار و مقدار پایانی قیمت روزانه نفت می‌باشد، استفاده شده است. به دلیل امکان وجود همبستگی و هم‌خطی میان داده‌ها که باعث ایجاد تورش در مقادیر پیش‌بینی می‌گردد، با استفاده از روش آنالیز مولفه‌های اصلی چهار متغیر ورودی به چهار مولفه اصلی که کل پراکندگی داده‌ها را پوشش می‌دهد تبدیل شده است. چهار مولفه اصلی استخراج شده از داده‌های ورودی به طور کامل مستقل از یکدیگر می‌باشند که به عنوان ورودی مدل جهت پیش‌بینی مورد استفاده قرار می‌گیرد.

همانطور که می‌دانیم برای برازش مدل باید سری زمانی مانا مورد استفاده قرار گیرد. بنابراین از آن‌جایی که در این پژوهش نامانایی سری زمانی قیمت نفت براساس آزمون دیکی فولر رد نشد، از سری زمانی بازده قیمت نفت استفاده می‌نماییم و با توجه به محاسبات انجام شده در نرم افزار (Eviews) مبنی بر مانایی سری زمانی بازده قیمت نفت از آن برای ادامه پژوهش بهره می‌بریم. سپس از روش تجزیه و تحلیل مولفه‌های اساسی(PCA)[1] برای پالایش اولیه داده‌ها استفاده شده آنگاه با استفاده از رگرسیون بردار پشتیبان(SVR)[2] که نوع خاصی از ماشین‌های بردار پشتیبان(SVM)[3] می‌باشد، به پیش‌بینی قیمت نفت اقدام می‌شود.

با توجه به توضیحات ارائه شده درباره مفهوم ریاضی مدل SVR، این مدل را در نرم‌افزار MATLAB شبیه‌سازی می‌نماییم. سپس مولفه‌های اصلی خروجی از مدل PCA به عنوان ورودی برای این مدل در نظر گرفته و مدل را با توجه به داده‌های آموزش برازش

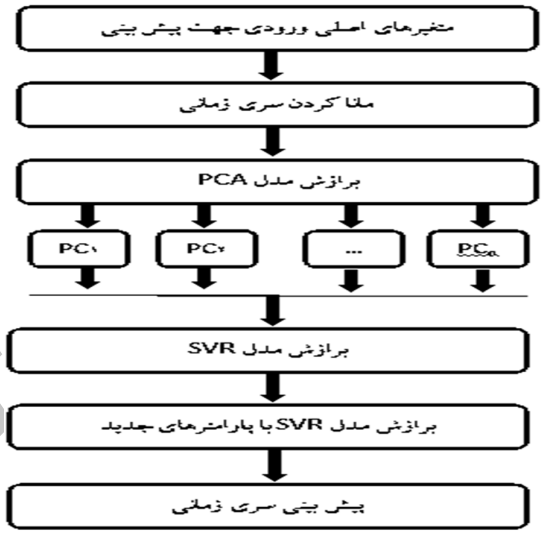
1. Principal Components Analyses
2 . Support Vector Regression
3 . Support Vector Machine

Archive of SID

www.SID.ir

می‌کنیم. در مدل سازی SVR مولفه‌های c و ε و σ توسط کاربر تعریف می‌شوند. مولفه c یک مولفه تنظیمی است و می‌تواند مقادیر صفر تا بی‌نهایت را بپذیرد. نقش این مولفه ایجاد تعادل میان کمینه کردن ریسک تجربی و بیشینه کردن قابلیت تعمیم‌یابی است. مولفه ε نیز می‌تواند مقادیر صفر تا بی‌نهایت را بپذیرد، مقدار این مولفه در وضعیت بردارهای پشتیبان و در نتیجه کارایی مدل بسیار موثر است. σ نیز عرض کرنل با پایه شعاعی می‌باشد و مقدار آن توسط کاربر تعیین می‌شود. سپس با توجه به ورودی‌های تست، مقدار قیمت نفت برای 20 روز آینده را پیش‌بینی کرده و با مقادیر واقعی مقایسه می‌نماییم و با توجه به معیارهای ارزیابی MAPE و RMSE دقت پیش‌بینی مدل را اندازه‌گیری می‌نماییم.

الگوریتم کلی مدل پیش‌بینی در شکل (2) نشان داده شده است.



شکل(2): الگوریتم کلی مدل پیش‌بینی

منبع: مفروضات پژوهش

Archive of SID

www.SID.ir

جامعه آماری پژوهش حاضر کل قیمت‌های روزانه نفت در تمام بازارهای نفتی می‌باشد و نمونه آماری قیمت روزانه نفت خام برنت دریای شمال از ژوئیه سال 2008 تا ژوئیه سال 2016 می‌باشد.

برای پیش‌بینی قیمت روزانه نفت خام برنت از مقدار سری در روزهای قبل استفاده شده است. چرا که فرض بر این است که اطلاعات در مورد مقدار آتی قیمت نفت در گذشته سری زمانی نهفته است. بنابراین مقدار قیمت نفت در روزهای قبل متغیر مستقل، و مقدار آتی قیمت نفت متغیر وابسته خواهد بود. داده‌های مربوط به قیمت روزانه نفت خام برنت از سایت هات فارکس[1] استخراج شده است. بنابراین قیمت نفت در روزهای گذشته به عنوان متغیرهای اصلی این پژوهش می‌باشد. برای پیش‌بینی قیمت نفت از اندیکاتور تصادفی[2] K% و D%، شاخص قدرت نسبی[3] 9، 14 و 21 روزه، میانگین متحرک همگرا/واگرا[4]، نرخ تغییر قیمت[5] 12 و 25 روزه، شاخص کانال کالا[6] 14، 20 و 34 روزه و نوسان نمای ویلیامز[7] R% با استفاده از مدل‌های مورد اشاره و مدل ترکیبی پیشنهادی انجام می‌گیرد.

4. یافته‌های پژوهش

در این قسمت یافته‌های پژوهش و تجزیه و تحلیل آن‌ها آورده شده است. به منظور بررسی بهبود در قدرت پیش‌بینی مدل، داده‌های قیمت روزانه نفت خام برنت از سال 2008 تا سال 2016، با استفاده از تکنیک پنجره غلتان[8] به 5 دوره زمانی برابر تقسیم شده و سپس برای هر دوره پیش‌بینی با مدل های SVR (استفاده از داده‌های معمولی جهت پیش‌بینی با رگرسیون بردار پشتیبان) و PCA-SVR (استفاده از مولفه‌های اصلی خروجی از روش آنالیز مولفه‌های اصلی، به عنوان ورودی مدل رگرسیون بردار پشتیبان جهت پیش‌بینی)

1 . www.hotforex.com
2 . Stochastic
3 . Relative Strength Index
4 . Moving Average Convergence Divergence
5 . Rate Of Change
6 . Commodity Channel Index
7 . Williams
8 . Rolling Window

Archive of SID

www.SID.ir

برای 20 روز آینده انجام گرفته و سپس با مقادیر واقعی مقایسه شده و معیارهای میانگین مربع خطا و درصد قدر مطلق خطا استخراج گردیده است.

1-4. محاسبه معیار ارزیابی عملکرد ریشه میانگین مربع خطا

با برازش هر یک از مدل‌های SVR و PCA-SVR ، معیار ارزیابی عملکرد ریشه میانگین مربع خطا در جدول (1) نشان داده شده است. همان‌طور که ملاحظه می‌شود مدل PCA-SVR نسبت به مدل SVR در تمامی دوره‌ها، عملکرد بهتری داشته است و از لحاظ میانگین 5 دوره نیز خطای کمتری را نمایش می‌دهد. برای آزمون فرضیه پژوهش مبنی بر عملکرد بهتر مدل PCA-SVR نسبت به مدل SVR معنادار بودن تفاوت میانگین این دو مدل را بررسی شده است. در صورت معنی‌دار بودن تفاوت، نشان می‌دهد که عملکرد مدل PCA-SVR به صورت معنی‌داری از مدل SVR بهتر است.

جدول (1) : ریشه میانگین مربع خطا

دوره	قیمت نفت	
	SVR	PCA-SVR
1	0/00203	0/00051
2	0/00184	0/00048
3	0/00168	0/00044
4	0/00243	0/00068
5	0/00373	0/00096
میانگین	0/00234	0/00061

منبع: یافته‌های پژوهش

Archive of SID

www.SID.ir

2-4. محاسبه معیار ارزیابی عملکرد میانگین درصد قدرمطلق خطا

با برازش هر یک از مدل‌های SVR و PCA-SVR ، معیار ارزیابی عملکرد میانگین درصد قدرمطلق خطا در جدول (2) نشان داده شده است. همان‌طور که ملاحظه می‌شود مدل PCA-SVR نسبت به مدل SVR در تعداد روزهای بیشتری، عملکرد بهتری داشته است و از لحاظ میانگین پیش‌بینی 20 روزه نیز خطای کمتری را نمایش می‌دهد. برای آزمون فرضیه پژوهش مبنی بر عملکرد بهتر مدل PCA-SVR نسبت به مدل SVR معنادار بودن تفاوت میانگین این دو مدل بررسی شده است. معنی‌دار بودن تفاوت، نشان می‌دهد که عملکرد مدل PCA-SVR به صورت معنی‌داری از مدل SVR بهتر است.

جدول (2) : درصد قدرمطلق خطا

دوره	قیمت نفت	
	SVR	PCA-SVR
1	10/02352	2/61192
2	12/03232	3/21932
3	15/02480	4/00381
4	11/09754	2/97672
5	13/08045	3/40106
میانگین	12/25173	3/24257

منبع: یافته‌های پژوهش

5. نتیجه‌گیری و بحث

قیمت نفت از مهمترین شاخص‌های تاثیرگذار بر عوامل اقتصادی در ایران است و به عنوان یک متغیر برونزای قدرتمند می‌تواند متغیرهای کلان اقتصادی را تحت تاثیر قرار دهد. گروه‌های مختلف شرکت کنندگان در بازار، سرمایه‌گذاران و کارشناسان اقتصادی همچنین سیاست‌گذاران و برنامه‌ریزان اقتصادی به ویژه در کشورهای صادرکننده نفت

Archive of SID

www.SID.ir

نظیر ایران به بررسی روش‌هایی جهت پیش‌بینی بهتر قیمت نفت نیاز دارند و خواستار راهکارهایی هستند که بتوانند با دقت بیشتری این متغیر را پیش‌بینی نمایند. بنابراین در این پژوهش با استفاده از داده‌های روزانه قیمت نفت خام برنت طی دوره زمانی ژوئیه 2008 تا ژوئیه 2016 و استفاده از مدل ترکیبی آنالیز مولفه‌های اصلی و رگرسیون بردار پشتیبان به این امر پرداخته‌ایم. ما فرض خود را بر این گذاشتیم که پالایش اولیه داده‌ها منجر به کاهش خطای پیش‌بینی می‌گردد. در ابتدا مدل SVR را جهت پیش‌بینی قیمت روزانه نفت خام برنت در نرم‌افزار متلب شبیه سازی نموده و سپس با استفاده از PCA متغیرهای ورودی را به مولفه‌های اصلی تجزیه نموده و به عنوان ورودی برای مدل SVR انتخاب می‌نماییم. پس از برازش مدل‌ها، مقدار قیمت نفت را برای یک روز آینده تا 20 روز پیش‌بینی نموده، سپس دقت مدل‌ها در پیش‌بینی قیمت نفت با معیارهای ارزیابی عملکرد MAPE و RMSE اندازه گیری شده است. به منظور مقایسه عملکرد مدل‌های SVR و PCA-SVR از آزمون مقایسات زوجی استفاده نموده که در نهایت نتایج حاکی از بالاتر بودن دقت پیش‌بینی مدل PCA-SVR نسبت به مدل SVR بود و فرضیه تحقیق مبنی بر کمتر بودن خطای پیش‌بینی مدل ترکیبی PCA-SVR نسبت به مدل SVR مورد تایید قرار گرفت.

بنابراین نتیجه می‌گیریم که انتخاب صحیح ورودی‌ها و کاهش بعد داده‌ها می‌تواند باعث بهبود عملکرد ماشین‌های بردار پشتیبان در پیش‌بینی سری زمانی قیمت روزانه نفت خام برنت گردد.

در پژوهش حاضر به مدل‌سازی جهت پیش‌بینی قیمت روزانه نفت خام برنت با ترکیب روش‌های آنالیز مؤلفه‌های اصلی و رگرسیون بردار پشتیبان پرداختیم ولی برای بررسی کارایی و مقایسه نتایج می‌توان این سری زمانی قیمت‌ها را به دیگر انواع شبکه‌های عصبی از جمله شبکه عصبی پیش‌خور داد تا دقت آن را نسبت به رگرسیون بردار پشتیبان مورد سنجش قرار داد و دقت پیش‌بینی ترکیب شبکه عصبی و PCA را نسبت به کاربرد مجزای هر یک از شبکه‌های عصبی مورد بررسی قرار داد.

Archive of SID

www.SID.ir

6.منابع

الف) فارسی

بروکز،کریس (1389). *اقتصاد سنجی مالی و تجزیه و تحلیل داده‌ها در علوم انسانی*، ترجمه احمد بدری و عبدالمجید عبدالباقی، چاپ اول. تهران: موسسه عملی فرهنگی نص.

جانسون،ریچارد آرنولد. و ویچرن،دین دبلیو. (1393). *تحلیل آماری چند متغیری کاربردی*. ترجمه حسینعلی نیرومند، چاپ هفتم. مشهد: دانشگاه فردوسی مشهد.

فرج‌زاده، منوچهر (1394). *تکنیک‌های اقلیم شناسی*. تهران: انتشارات سمت.

فغانی ماکرانی، خسرو، صالح‌نژاد،سیدحسن و امین،وحید (1395). پیش‌بینی مدیریت سود مبتنی بر مدل جونز تعدیل شده با استفاده از مدل شبکه عصبی مصنوعی و الگوریتم ژنتیک، *فصلنامه مهندسی مالی ومدیریت اوراق بهادار*، شماره 28، صص 136-117.

محمدی، شاپور، راعی، رضا و کرمی، حسین.(1394). پیش‌بینی قیمت هفتگی نفت خام از طریق مدل ترکیبی ماشین بردار پشتیبان و خودرگرسیو میانگین متحرک انباشته. *فصلنامه دانش سرمایه‌گذاری*، سال چهارم، شماره14، صص 159-145.

ب) انگلیسی

Gunn, S. (1998). “Support Vector Machines for Classification and Regression”. Technical Report. Dept. of Electronics and computer Science. University of Southampton, Southampton, UK, pp.1-52.

Kwasi Mensah, E. (2016). “Box-Jenkins modelling and forecasting of Brent crude oil price”, MPRA Paper, No. 67748, pp. 1-10.

Lipo, W. (2005). “Support Vector Machines: Theory and Applications”, Springer-Verlag Berlin Heidelberg, Berlin, Germany, pp.335-363.

Li, X.; He, K.; Lai, K. and Zou,Y. (2014). “Forecasting Crude Oil Price With Multiscale Denoising Ensemble Model”, *Mathematic Problems in Engineering*, No. 4, pp. 1-9.

Lu, C.-J. (2013). “Hybridizing nonlinear independent component analysis and support vector regression with particle swarm optimization for stock index forecasting”, *Neural Comput & Applic*, No. 23, pp. 2417-2427.

Sanches, D. (2003). “Advanced Support Vector Machines and Kernel methods”, *Neurocomputing*, Vol.55, Issue.1-2, pp.5-20.

Archive of SID

www.SID.ir

Xie, W.,Yu, L., Xu, S. and Wang, S. (2006), "A New Method for Crude oil Price Forecasting Based on Support Vector Machines", Chinese Academy of Sciences, Beijing 100080, China, pp. 444-4

Archive of SID

www.SID.ir