

پیش‌بینی قیمت مس با استفاده از ابزارهای رگرسیون بردار پشتیبان، بیز ساده، حافظه کوتاه‌مدت و K-نزدیکترین همسایگی

امیرعلی فریدی

دانشجوی کارشناسی، دانشکده مهندسی کامپیوتر، دانشگاه غیرانتفاعی شمال آمل

چکیده

این مقاله یک مطالعه مقایسه‌ای از چهار مدل یادگیری ماشین، یعنی حافظه کوتاه مدت (LSTM)، K-نزدیکترین همسایگی (KNN)، رگرسیون بردار پشتیبان (SVR) و بیز ساده (naïve bayes) برای پیش‌بینی قیمت مس ارائه می‌دهد. هدف این مطالعه ارزیابی عملکرد این مدل‌ها در پیش‌بینی قیمت‌های آتی مس بر اساس داده‌های تاریخی است. مجموعه داده مورد استفاده در این مطالعه شامل داده‌های تاریخی قیمت مس و ویژگی‌های مربوطه است که ممکن است بر نوسانات قیمت تأثیر بگذارد. هر مدل بر روی زیرمجموعه‌ای از داده‌ها آموزش داده می‌شود و در یک مجموعه آزمون جداگانه ارزیابی می‌شود تا دقت پیش‌بینی آنها ارزیابی شود. معیارهای عملکرد مختلفی مانند میانگین خطای مطلق (MAE)، خطای میانگین مربعات ریشه (RMSE) و R-squared برای مقایسه قابلیت‌های پیش‌بینی مدل‌ها استفاده می‌شوند. نتایج نشان می‌دهد که مدل K-نزدیکترین همسایگی از نظر دقت پیش‌بینی برای پیش‌بینی قیمت مس از مدل‌های دیگر بهتر عمل می‌کند. KNN عملکرد برتر را در ثبت الگوها و روندهای اساسی در داده‌ها نشان می‌دهد که منجر به پیش‌بینی دقیق‌تر قیمت می‌شود. توانایی مدل برای شناسایی الگوهای مشابه در داده‌های تاریخی و استفاده از آنها برای پیش‌بینی قیمت‌های آینده به موفقیت آن در این کار کمک می‌کند. به طور کلی، این مطالعه اهمیت انتخاب یک مدل یادگیری ماشین مناسب برای پیش‌بینی وظایف را برجسته می‌کند و اثربخشی KNN را در پیش‌بینی قیمت مس نشان می‌دهد. این یافته‌ها بینش‌های ارزشمندی را برای محققان و دست‌اندرکاران بازارهای مالی که برای اهداف تصمیم‌گیری بر پیش‌بینی قیمت دقیق تکیه می‌کنند، ارائه می‌کند.

واژگان کلیدی: پیش‌بینی قیمت مس، رگرسیون بردار پشتیبان، بیز ساده، حافظه کوتاه‌مدت و K-نزدیکترین همسایگی.

مقدمه

مس که به دلیل شهرتش به عنوان یک فشارسنگ برای اقتصاد جهانی، اغلب به عنوان "دکتر مس" شناخته می‌شود، نقشی محوری در صنایع و بخش‌های مختلف در سراسر جهان ایفا می‌کند. مس به عنوان یک فلز کلیدی صنعتی، به دلیل رسانایی، چکش‌خواری و مقاومت در برابر خوردگی، تقاضای بالایی دارد و آن را به یک جزء ضروری در تولید سیم‌کشی برق، لوله‌کشی، الکترونیک و مصالح ساختمانی تبدیل می‌کند. با توجه به کاربرد گسترده و اهمیت اقتصادی آن، پیش‌بینی دقیق قیمت مس برای بسیاری از سهامداران از جمله سرمایه‌گذاران، تجار، سیاست‌گذاران و کارشناسان صنعت بسیار مهم است. در سال‌های اخیر، استفاده از تکنیک‌های یادگیری ماشین در پیش‌بینی مالی مورد توجه قرار گرفته است و راه‌های جدیدی را برای پیش‌بینی قیمت کالاها با دقت و کارایی بیشتر ارائه می‌دهد. در میان مدل‌های مختلف یادگیری ماشینی موجود، حافظه کوتاه‌مدت بلند مدت (LSTM)، K-نزدیکترین همسایگی (KNN)، رگرسیون بردار پشتیبانی (SVR) و بیز ساده به عنوان گزینه‌های محبوب برای وظایف پیش‌بینی سری‌های زمانی ظاهر شده‌اند. این مدل‌ها از داده‌های تاریخی و ویژگی‌های مرتبط برای پیش‌بینی تغییرات قیمت در آینده استفاده می‌کنند و بینش‌های ارزشمندی را برای تصمیم‌گیری در بازارهای مالی ارائه می‌کنند. پیش‌بینی قیمت مس به دلیل تأثیر متقابل پیچیده عواملی که بر بازارهای کالایی تأثیر می‌گذارند، از جمله پویایی عرضه-تقاضا، شاخص‌های اقتصادی، رویدادهای ژئوپلیتیکی و پیشرفت‌های فناوری، چالشی منحصربه‌فرد است.

مدل‌های اقتصادسنجی سنتی اغلب برای به دست آوردن روابط غیرخطی و الگوهای ذاتی در داده‌های قیمت کالا تلاش می‌کنند، که باعث می‌شود روش‌های یادگیری ماشین برای پیش‌بینی‌های دقیق‌تر و قوی‌تر کاوش شود. در این زمینه، مطالعه ما بر مقایسه عملکرد چهار مدل یادگیری ماشین رگرسیون بردار پشتیبانی، بیز ساده، حافظه کوتاه‌مدت و K-نزدیکترین همسایگی در پیش‌بینی قیمت مس بر اساس داده‌های تاریخی متمرکز است. هدف شناسایی موثرترین مدل برای پیش‌بینی قیمت‌های آینده مس و ارزیابی دقت پیش‌بینی آن نسبت به مدل‌های دیگر است. با تجزیه و تحلیل نقاط قوت و محدودیت‌های هر مدل در ثبت الگوهای اساسی در داده‌ها، هدف ما ارائه بینش‌های ارزشمندی در مورد کارایی تکنیک‌های یادگیری ماشین برای پیش‌بینی قیمت کالاها است.

در میان مدل‌های ارزیابی شده K-نزدیکترین همسایگی (KNN) به دلیل سادگی، انعطاف‌پذیری و توانایی آن در به تصویر کشیدن الگوهای محلی در داده‌ها، به عنوان یک نامزد امیدوارکننده برای پیش‌بینی قیمت‌های مس برجسته است. KNN بر اساس اصل شباهت عمل می‌کند، جایی که K-نزدیکترین همسایگی را در داده‌های آموزشی شناسایی می‌کند تا برای نقاط داده جدید پیش‌بینی کند. این مکانیسم به KNN اجازه می‌دهد تا روابط و روندهای اساسی در داده‌ها را به‌طور مؤثر ثبت کند، و آن را برای کارهای پیش‌بینی سری‌های زمانی مانند پیش‌بینی قیمت کالاها مناسب می‌سازد. ماهیت ناپارامتری KNN باعث می‌شود که آن را در برابر نویزهای دورافتاده و داده‌ها مقاوم کند، و آن را قادر می‌سازد به خوبی با شرایط متغیر بازار و نوسانات قیمت سازگار شود. این سازگاری به ویژه در زمینه پیش‌بینی قیمت‌های کالا، که در آن عوامل خارجی می‌توانند منجر به تغییرات ناگهانی و غیرقابل پیش‌بینی در پویایی بازار شوند، ارزشمند است. KNN با استفاده از توانایی خود در شناسایی الگوهای مشابه در داده‌های تاریخی و برون‌یابی آنها برای پیش‌بینی قیمت‌های آتی، عملکرد برتر را در گرفتن روابط پیچیده ذاتی در بازارهای کالا نشان می‌دهد.

هدف تجزیه و تحلیل مقایسه‌ای ما از مدل‌های یادگیری ماشین برای پیش‌بینی قیمت مس، روشن کردن نقاط قوت و ضعف هر مدل در پیش‌بینی دقیق قیمت‌های کالا است. با ارزیابی عملکرد حافظه کوتاه‌مدت بلند مدت (LSTM)، K-نزدیکترین همسایگی (KNN)، رگرسیون بردار پشتیبانی (SVR) و بیز ساده بر روی داده‌های تاریخی قیمت مس، ما به دنبال شناسایی بهترین مدل برای پیش‌بینی قیمت مس و ارائه بینش‌های ارزشمند برای سهامداران در بازارهای مالی هستیم.

در نتیجه، تحلیل مقایسه‌ای مدل‌های یادگیری ماشین برای پیش‌بینی قیمت‌های مس، اثربخشی K-نزدیکترین همسایگی (KNN) را به عنوان بهترین عملکرد در میان حافظه کوتاه‌مدت بلند مدت (LSTM)، رگرسیون بردار پشتیبانی (SVR) و بیز ساده برجسته می‌کند. توانایی KNN برای ثبت الگوهای اساسی در داده‌ها، سادگی، استحکام در برابر نویز و سازگاری با شرایط متغیر بازار، آن را به مدلی امیدوارکننده برای پیش‌بینی قیمت کالاها

تبدیل کرده است. این مطالعه بر اهمیت انتخاب مدل مناسب برای وظایف پیش‌بینی تأکید می‌کند و بر ارزش KNN در پیش‌بینی دقیق قیمت مس و سایر کالاها تأکید می‌کند.

پیشینه تحقیق

هوش مصنوعی توسط فلاسفه و ریاضی‌دانانی نظیر جرج بول که اقدام به ارائه قوانین و نظریه‌هایی در مورد منطق نمودند، مطرح شده بود. با اختراع رایانه‌های الکترونیکی در سال ۱۹۴۳، هوش مصنوعی، دانشمندان آن زمان را به چالشی بزرگ فراخواند. در این شرایط، چنین به نظر می‌رسید که این فناوری قادر به شبیه‌سازی رفتارهای هوشمندانه خواهد بود. با وجود مخالفت گروهی از متفکران با هوش مصنوعی که با تردید به کارآمدی آن می‌نگریستند تنها پس از چهار دهه، شاهد تولد ماشین‌های شطرنج باز و دیگر سامانه‌های هوشمند در صنایع گوناگون شدیم.

حوزه پژوهش در زمینه هوش مصنوعی در یک کارگاه آموزشی در کالج دارتموت در سال ۱۹۵۶ متولد شد. شرکت‌کنندگان آلن نیول (دانشگاه کارنگی ملون)، هربرت سیمون (دانشگاه کارنگی ملون)، جان مک‌کارتی (مؤسسه فناوری ماساچوست)، ماروین منسکی (مؤسسه فناوری ماساچوست) و آرتور ساموئل (آی بی ام) از بنیان‌گذاران و رهبران پژوهش در زمینه هوش مصنوعی شدند. آن‌ها به همراه دانشجویانشان برنامه‌هایی نوشتند که مطبوعات آن را «شگفت‌آور» توصیف می‌کردند، رایانه‌ها استراتژی‌های برد بازی چکرز را فرا می‌گرفتند، سوالاتی در جبر حل می‌کردند، قضیه‌های منطقی اثبات می‌کردند و انگلیسی صحبت می‌کردند. در اواسط دهه ۱۹۶۰ میلادی وزارت دفاع آمریکا سرمایه‌گذاری‌های سنگینی در حوزه پژوهش در زمینه هوش مصنوعی انجام می‌داد، در آن دهه آزمایشگاه‌های فراوانی در سراسر جهان تأسیس شد. بنیانگذاران هوش مصنوعی در مورد آینده خوشبین بودند: هربرت سیمون پیش‌بینی کرد «ماشین‌ها ظرف بیست سال قادر به انجام هر کاری هستند که یک انسان می‌تواند انجام دهد». ماروین منسکی، نوشت: در طی یک نسل ... مسئله هوش مصنوعی اساساً حل خواهد شد.

نام هوش مصنوعی در سال ۱۹۶۵ میلادی به عنوان یک دانش جدید ابداع گردید. البته فعالیت در این زمینه از سال ۱۹۶۰ میلادی شروع شد. بیشتر کارهای پژوهشی اولیه در هوش مصنوعی بر روی انجام ماشین‌های بازی‌ها و نیز اثبات قضیه‌های ریاضی با کمک رایانه‌ها بود. در آغاز چنین به نظر می‌آمد که رایانه‌ها قادر خواهند بود چنین فعالیت‌هایی را تنها با بهره گرفتن از تعداد بسیار زیادی کشف و جستجو برای مسیرهای حل مسئله و سپس انتخاب بهترین روش برای حل آن‌ها به انجام رسانند. اصطلاح هوش مصنوعی برای اولین بار توسط جان مک‌کارتی (که از آن به عنوان پدر علم و دانش تولید ماشین‌های هوشمند یاد می‌شود) استفاده شد. وی مخترع یکی از زبان‌های برنامه‌نویسی هوش مصنوعی به نام لیسپ (Lisp) است. با این عنوان می‌توان به هویت رفتارهای هوشمندانه یک ابزار مصنوعی پی برد. (ساخته دست بشر، غیرطبیعی، مصنوعی) حال آنکه هوش مصنوعی به عنوان یک اصطلاح عمومی پذیرفته شده که شامل محاسبات هوشمندانه و ترکیبی (مركب از مواد مصنوعی) است.

روش تحقیق

الگوریتم K-نزدیکترین همسایه

الگوریتم K-نزدیکترین همسایه یا KNN یکی از ساده‌ترین الگوریتم‌های یادگیری ماشین با ناظر است که برای حل مسائل طبقه‌بندی و رگرسیون استفاده می‌شود. KNN همچنین به عنوان یک مدل مبتنی بر نمونه یا یک یادگیرنده‌ی تنبل شناخته می‌شود؛ زیرا یک مدل داخلی ایجاد نمی‌کند و از داده‌های آموزش عملکرد متمایز را یاد نمی‌گیرد؛ فقط نمونه‌های آموزشی را حفظ می‌کند که به عنوان «دانش» برای مرحله‌ی پیش‌بینی استفاده می‌شود.

این الگوریتم برای مسائل طبقه‌بندی K -نزدیکترین همسایه را پیدا و با اکثریت آرا نزدیکترین همسایگی کلاس را پیش‌بینی می‌کند. برای مسائل رگرسیون K -نزدیکترین همسایه را پیدا و با محاسبه‌ی میانگین مقدار نزدیک‌ترین همسایه‌ها، مقدار مدنظر را پیش‌بینی می‌کند. مراحل الگوریتم K -نزدیکترین همسایه:

- داده‌ها بارگذاری می‌شود.
- مقدار K را تعیین می‌شود که همان تعداد نزدیک‌ترین همسایه‌ها هستند.
- برای هر نمونه داده، فاصله‌ی میان نمونه داده‌ی جدید با نمونه داده‌های موجود محاسبه می‌شود.
- فاصله و شاخص هر نمونه به یک فهرست وارد می‌شود.
- کل لیست براساس فاصله‌ی نمونه داده‌ها، از کمترین به بیشترین فاصله، مرتب می‌شود.
- K تا از اولین نمونه‌های فهرست مرتب‌شده، به‌عنوان K -نزدیک‌ترین همسایه انتخاب می‌شود.
- برچسب این K نمونه بررسی می‌شود.
- اگر مسئله رگرسیون باشد، میانگین برچسب‌های این K نمونه داده برچسب نمونه داده جدید خواهد بود.
- در صورتی که مسئله طبقه‌بندی باشد، نمونه‌ی جدید هم همان برچسب K همسایه را خواهد داشت.

پیدا کردن مقدار بهینه K یا Optimal K Value

انتخاب مقدار مناسب K تنظیم پارامتر نامیده می‌شود و برای نتایج بهتر ضروری است. برای انتخاب مقدار K ریشه‌ی مربع یا همان جذر یا رادیکال تعداد کل نقاط داده‌ی موجود در مجموعه‌ی داده محاسبه می‌شود. این موضوع را در نظر داشته باشیم که در نهایت مقدار فرد K همیشه برای جلوگیری از سردرگمی میان دو کلاس انتخاب می‌شود.

رگرسیون بردار پشتیبان (SVR)

SVR مخفف عبارت Support Vector Regression است که یک تکنیک یادگیری ماشینی برای پیش‌بینی مقادیر پیوسته مانند قیمت سهام، دما یا قیمت خانه است. SVR مبتنی بر ایده یافتن تابعی است که تا حد امکان با نقاط داده مطابقت داشته باشد و در عین حال خطا یا انحراف از مقادیر واقعی را نیز به حداقل برساند. SVR از مفهومی به نام هسته برای تبدیل داده‌ها به فضایی با ابعاد بالاتر استفاده می‌کند، جایی که یافتن یک تابع مناسب آسان‌تر است. برخی از هسته‌های رایج تابع پایه خطی، چند جمله‌ای و شعاعی (RBF) هستند.

SVR نسبت به سایر روش‌های رگرسیون مزایایی دارد، مانند توانایی مدیریت روابط غیرخطی، قوی بودن نسبت به نقاط پرت و داشتن پارامترهای کمتر برای تنظیم. با این حال، SVR همچنین دارای معایبی است، مانند گران بودن محاسباتی، نیاز به مقیاس بندی داده‌ها، و حساس بودن به انتخاب هسته و پارامترهای آن.

SVR در حوزه‌های مختلفی مانند مالی، مهندسی، پزشکی و علوم اجتماعی اعمال شده است. یک مثال پیش‌بینی قیمت سهام با استفاده از SVR است که شامل استفاده از داده‌های تاریخی قیمت سهام و سایر ویژگی‌های مرتبط مانند حجم، شاخص‌ها یا احساسات برای آموزش یک مدل SVR است که می‌تواند قیمت‌های آینده را پیش‌بینی کند. این می‌تواند به سرمایه‌گذاران یا معامله‌گران کمک کند تا تصمیمات بهتری بگیرند و سود خود را بهینه کنند. با این حال، پیش‌بینی قیمت سهام نیز یک کار چالش‌برانگیز است، زیرا بازار تحت تأثیر عوامل بسیاری مانند اخبار، رویدادها، روندها یا رفتار انسانی است که به سختی می‌توان آن‌ها را به دست آورد یا کمیت آن را تعیین کرد. بنابراین، مدل‌های SVR ممکن است همیشه دقیق یا قابل اعتماد نباشند و باید با احتیاط و اعتبار استفاده شوند.

رگرسیون بردار پشتیبان (SVR) Support Vector Regression یک الگوریتم یادگیری ماشینی قدرتمند است که به طور گسترده برای وظایف رگرسیون استفاده می‌شود. این یک توسعه از ماشین‌های بردار پشتیبانی (SVM) است که عمدتاً برای مشکلات طبقه‌بندی استفاده می‌شود. از سوی دیگر، SVR به طور خاص برای رسیدگی به مشکلات رگرسیون طراحی شده است، جایی که هدف پیش‌بینی یک متغیر خروجی پیوسته است. مقدمه

ای بر پشتیبانی از رگرسیون برداری به عبارت ساده، رگرسیون یک تکنیک آماری است که برای مدل سازی رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل استفاده می شود.

بیز ساده (Naïve Bayes)

بیز ساده (Naive Bayes) ساده ترین الگوریتم یادگیری ماشین است که می توانیم روی داده های خود اعمال کنیم. همان طور که از نامش پیداست، این الگوریتم فرض می کند همه ی متغیرهای مجموعه ی داده ساده (Naïve) هستند، یعنی با یکدیگر ارتباط ندارند. Naive Bayes یک الگوریتم طبقه بندی محبوب در یادگیری ماشین است. یک الگوریتم احتمالی است که مبتنی بر قضیه بیز است، که احتمال وقوع یک رویداد را بر اساس دانش قبلی از شرایطی که ممکن است با رویداد مرتبط باشد محاسبه می کند. در Naive Bayes، الگوریتم فرض می کند که ویژگی های داده های ورودی مستقل از یکدیگر هستند، که این موضوع محاسبات احتمال را ساده می کند و الگوریتم را از نظر محاسباتی کارآمد می کند. علی رغم این فرض ساده کننده، مشخص شده است که Naive Bayes در بسیاری از کاربردهای دنیای واقعی، به ویژه در پردازش زبان طبیعی (NLP) و وظایف طبقه بندی متن، بسیار مؤثر است. الگوریتم Naive Bayes با یادگیری احتمال شرطی هر ویژگی داده شده در هر کلاس در داده های آموزشی کار می کند. هنگامی که داده های جدید و دیده نشده به الگوریتم ارائه می شود، الگوریتم احتمال هر کلاس را بر اساس ویژگی های مشاهده شده محاسبه می کند و کلاسی را که بیشترین احتمال را دارد به عنوان کلاس پیش بینی شده برای آن داده انتخاب می کند. با وجود سادگی و فرضیات گفته شده، مشخص شده است که Naive Bayes در طیف گسترده ای از وظایف طبقه بندی به خوبی عمل می کند، به خصوص زمانی که میزان داده های آموزشی در مقایسه با تعداد ویژگی ها کم باشد.

حافظه کوتاه مدت (LSTM)

حافظه کوتاه مدت (Long short-term memory (LSTM)، یک شبکه عصبی مصنوعی است که در حوزه های موضوعی هوش مصنوعی و یادگیری عمیق به کار گرفته می شود. بر خلاف شبکه عصبی پیشخور، LSTM دارای اتصالات بازخوردی است. LSTM مانند شبکه عصبی بازگشتی، نه تنها به پردازش داده های مفرد مثل تصاویر، می پردازد، بلکه آن می تواند داده های متوالی مثل سخنرانی یا فیلم را هم پردازش کند. برای مثال LSTM برای کارهایی مثل شناسایی دست خط، شناسایی سخنرانی، ترجمه ماشینی، کنترل ربات، بازی های ویدیویی و مراقبت های بهداشتی کاربرد دارد. LSTM در قرن بیستم به عنوان بهترین نوع شبکه عصبی شناخته شده است. LSTM توسط Yoshua Bengio و دیگران در سال ۱۹۹۴ در مقاله ای با عنوان Learning long term Dependencies with gradient descent is difficult مطرح شد. LSTM به زبان ساده یعنی مدلی که برای دنباله های طولانی به کار می رود. LSTM لایه پیشرفته تری نسبت به simple RNN است. برخی اوقات برای داشتن خروجی مناسب باید به ورودی های اخیر نگاه کرد. LSTM می تواند هر دنباله ای با طول طولانی را پردازش کند. برای مثال در شکل زیر قابلیت LSTM نشان داده می شود. مثلاً اطلاعاتی در قسمت اول وجود دارد که در قسمت آخر استفاده خواهد شد. $T=1$ بازه زمانی اول است و اطلاعات پردازش شده قرار است در بازه زمانی ۵۰۰ استفاده شود. پس LSTM بازه زمانی اول را یادگیری می کند و آن را تا بازه زمانی ۵۰۰ که برایش در نظر گرفته شده ذخیره می کند تا در صورت آن نیاز آن اطلاعات استفاده شود.

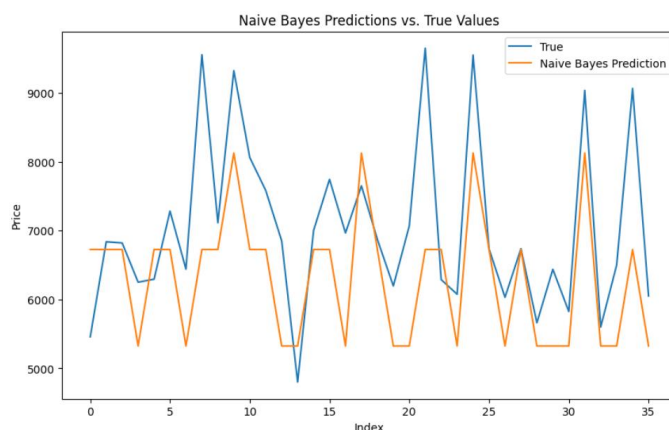
پیاده سازی مدل ها

برای پیش بینی قیمت مس در سال های آینده می توان از الگوریتم های یادگیری ماشین استفاده کرد. مجموعه داده مورد استفاده ما شامل داده های قیمت دلاری مس از سال ۲۰۰۹ تا سال ۲۰۲۳ میلادی است. برای پیش بینی دقیق تر نیاز به ویژگی ها و همچنین تعداد رکورد بیشتر داده توصیه می شود؛ با این حال در این بخش سعی شده است، الگوریتم های مناسب جهت این پیش بینی را معرفی کرده و یک پیش بینی تقریبی از قیمت مس تا سال ۲۰۲۹

میلادی ارائه شود. همچنین عملکرد هریک از مدل‌ها به ازای داده‌های آزمون و آموزش با توجه به چهار شاخص $MAPE$, $RMSE$, MAE و R^2 سنجیده می‌شود.

مدل بیز ساده Naïve Bayes

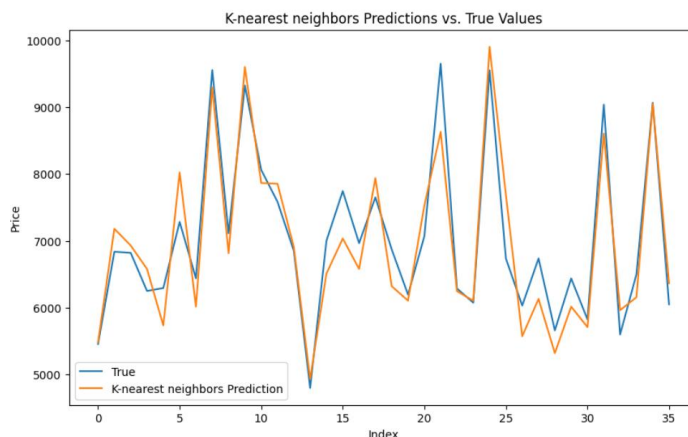
با توجه به این مدل و رابطه‌ی بدست آمده از این مدل پیش‌بینی، برای داده‌های سال ۲۰۰۹ تا ۲۰۲۳ نمودار آن به شکل زیر بدست می‌آید. این مدل هر چند همبستگی کاملی با داده‌های واقعی ندارد اما به نسبت مدل‌های $LSTM$ و SVR نمودار هم‌بسته‌تر و ضریب خطای (R^2) بهتری دارد که داده‌های آنالیز حساسیت بدست آمده برای تحلیل خطا و نمودار این مدل در زیر قابل مشاهده است (خطاها در جدول ۱ موجود است).



نمودار ۱: پیش‌بینی با بیز ساده

مدل K-نزدیکترین همسایگی

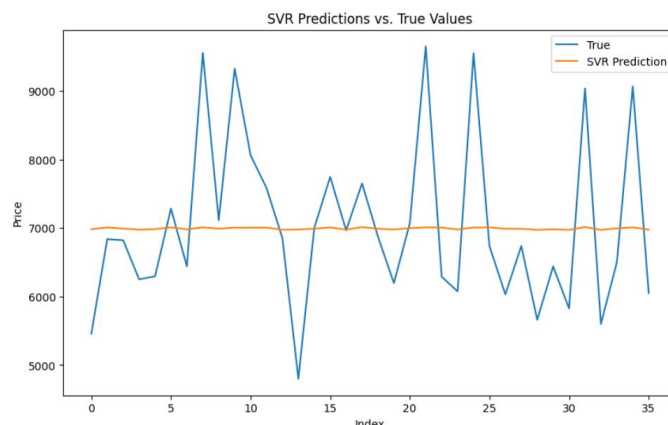
با توجه به این مدل و رابطه‌ی بدست آمده از این مدل پیش‌بینی، برای داده‌های سال ۲۰۰۹ تا ۲۰۲۳ نمودار آن به شکل زیر بدست می‌آید. این مدل نسبت به سایر مدل‌ها همبستگی کاملی با داده‌های واقعی دارد و به نسبت مدل‌های $LSTM$ و SVR حتی بیز ساده نمودار هم‌بسته‌تر و ضریب خطای (R^2) بهتری دارد که داده‌های آنالیز حساسیت بدست آمده برای تحلیل خطا و نمودار این مدل در زیر قابل مشاهده است (خطاها در جدول ۱).



نمودار ۲: پیش‌بینی با K-نزدیکترین همسایگی

رگرسیون بردار پشتیبان یا SVR

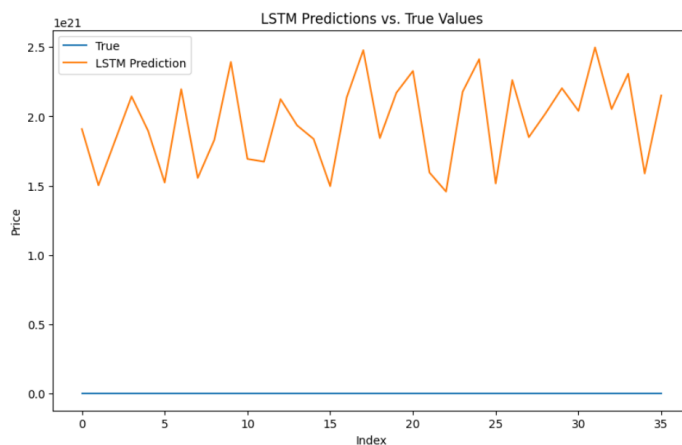
با توجه به این مدل و رابطه‌ی بدست آمده از این مدل پیش بینی، برای داده‌های سال ۲۰۰۹ تا ۲۰۲۳ نمودار آن به شکل زیر بدست می‌آید. این مدل نسبت به سایر مدل‌ها همبستگی با داده‌های واقعی ندارد اما به نسبت مدل LSTM نمودار هم‌بسته‌تر و ضریب خطای (R^2) بهتری دارد که داده‌های آنالیز حساسیت بدست آمده برای تحلیل خطا (جدول ۱) و نمودار این مدل در زیر قابل مشاهده است.



نمودار ۳: پیش‌بینی با مدل رگرسیون بردار پشتیبان

حافظه کوتاه‌مدت (LSTM)

با توجه به این مدل و رابطه‌ی بدست آمده از این مدل پیش بینی، برای داده‌های سال ۲۰۰۹ تا ۲۰۲۳ نمودار آن به شکل زیر بدست می‌آید. این مدل نسبت به سایر مدل‌های دیگر همبستگی با داده‌های واقعی ندارد و اصلاً نمودار هم‌بسته و ضریب خطای (R^2) خوبی ندارد که داده‌های آنالیز حساسیت بدست آمده برای تحلیل خطا (جدول ۱) و نمودار این مدل در زیر قابل مشاهده است.

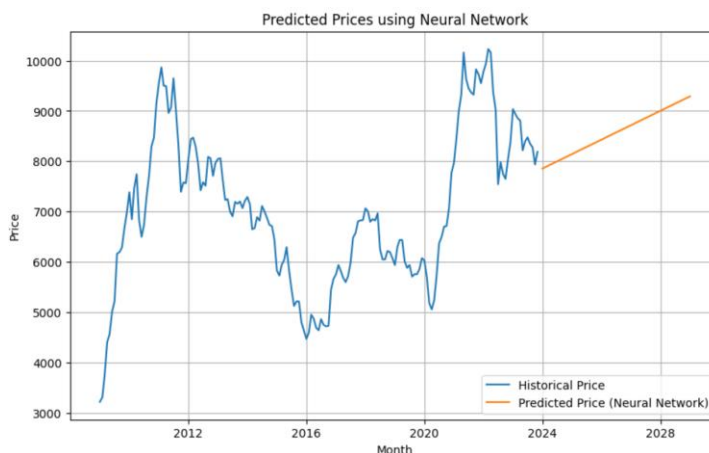


نمودار ۴: پیش‌بینی با مدل حافظه کوتاه‌مدت

در تصاویر بالا مشاهده می‌شود که بهترین نتایج را الگوریتم‌های KNN و Naïve Bayes داشته‌اند. بیشترین خطا مربوط به الگوریتم‌های SVR و LSTM است. برای الگوریتم LSTM نیاز است تا ویژگی‌های بیشتری در مجموعه داده‌ها قرار بگیرد تا دقت این روش شبکه عصبی بالاتر رود. همچنین نمودار پیش‌بینی دقیق‌تر روش KNN به شکل زیر است.

همانطور که در جدول زیر مشاهده می‌شود روش KNN کمترین خطا را داشته است. البته این روش در پیش‌بینی قیمت‌ها تا سال ۲۰۲۹ میلادی چندان خوب عمل نکرده و مقادیر پیش‌بینی آن حول قیمت ۸۰۰۰ تا ۸۵۰۰ است. بنابراین سعی کردیم تا از روش دقیق‌تری استفاده

کنیم. با توجه به کمبود داده و ویژگی، یک مدل شبکه عصبی با دو لایه مخفی را روی داده‌های خود آموزش دادیم که نتایج شبیه‌سازی آن تا سال ۲۰۲۹ در شکل زیر آمده است.



نمودار ۵: پیش‌بینی قیمت مس تا سال ۲۰۲۹

مدل	R^2	RMSE	MAE	MAPE
بیز ساده	۰.۱۱۳۳	۱۱۶۱.۴۲۵۳	۹۱۶.۵۱۱۱	۱۲.۴۷۵۷
نزدیکترین همسایگی	۰.۸۷۸۸	۴۲۹.۳۴۸۰	۳۵۵.۷۰۶۳	۵.۰۴۳۷
رگرسیون بردار پشتیبان	۰.۰۱۵۶	۱۲۲۳.۷۰۸۲	۹۳۲.۱۰۳۲	۱۳.۰۵۹۴
الگوریتم شبکه عصبی	-۲.۵۹۴۳	۱.۹۸۶۶	۱.۹۶۲۷	۲.۸۷۵۶

جدول ۱: خطاهای مدل‌ها

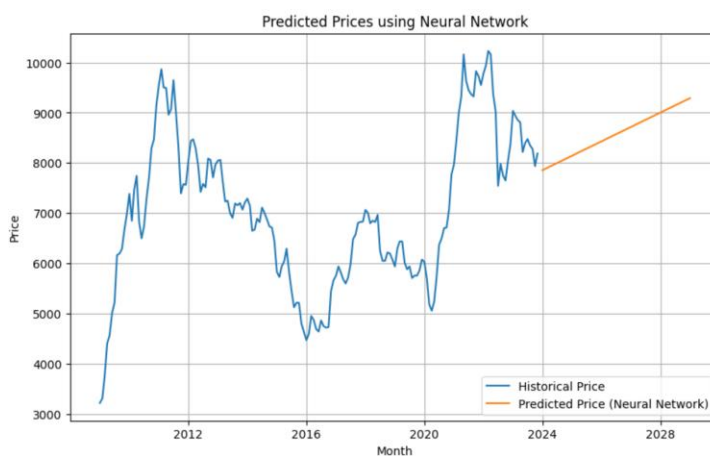
نتیجه‌گیری

پس به طور کلی می‌توان گفت، همانطور که در فصول قبل به این موضوع پرداخته شد، هدف این پژوهش پیش‌بینی قیمت مس با سال ۲۰۲۹ با مدل‌های هوش مصنوعی بیز ساده، رگرسیون بردار پشتیبان، نزدیکترین همسایگی، الگوریتم شبکه عصبی LSTM و قیمت مس از ۲۰۰۹ تا ۲۰۲۳ بوده‌است. صحت-سنجی این مدل‌ها با توجه به چهار شاخص MAPE، MAE، RMSE و R^2 سنجیده شده‌است. که این شاخص‌ها به طور کلی در جدول زیر موجود است:

مدل	R^2	RMSE	MAE	MAPE
بیز ساده	۰.۱۱۳۳	۱۱۶۱.۴۲۵۳	۹۱۶.۵۱۱۱	۱۲.۴۷۵۷
نزدیکترین همسایگی	۰.۸۷۸۸	۴۲۹.۳۴۸۰	۳۵۵.۷۰۶۳	۵.۰۴۳۷
رگرسیون بردار پشتیبان	۰.۰۱۵۶	۱۲۲۳.۷۰۸۲	۹۳۲.۱۰۳۲	۱۳.۰۵۹۴
الگوریتم شبکه عصبی	-۲.۵۹۴۳	۱.۹۸۶۶	۱.۹۶۲۷	۲.۸۷۵۶

جدول ۱: خطاهای مدل‌ها

همانطور که از داده‌های جدول بالا، نمودارها و تحلیل‌های فصل قبل مشخص است بهترین مدل در بین این مدل‌ها برای پیش‌بینی روش KNN کمترین خطا را داشته است. البته این روش در پیش‌بینی قیمت‌ها تا سال ۲۰۲۹ میلادی چندان خوب عمل نکرده و مقادیر پیش‌بینی آن حول قیمت ۸۰۰۰ تا ۸۵۰۰ است. بنابراین سعی شده‌است تا از روش دقیق‌تری استفاده شود. با توجه به کمبود داده و ویژگی، یک مدل شبکه عصبی با دو لایه مخفی روی داده‌های آموزش داده شده است که نتایج شبیه‌سازی آن تا سال ۲۰۲۹ در شکل زیر آمده است:



نمودار ۵: پیش‌بینی قیمت مس تا سال ۲۰۲۹

منابع

- Lahmirmi, S. (2018). Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression. In *Applied Mathematics and Computation* (Vol. 320, pp. 444–451). Elsevier BV. <https://doi.org/10.1016/j.amc.2017.09.049>
- Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. In *Mathematical and Computer Modelling* (Vol. 58, Issues 3–4, pp. 458–465). Elsevier BV. <https://doi.org/10.1016/j.mcm.2011.11.021>
- Ahmed, A. M., Rizaner, A., & Ulusoy, A. H. (2016). Using data Mining to Predict Instructor Performance. In *Procedia Computer Science* (Vol. 102, pp. 137–142). Elsevier BV. <https://doi.org/10.1016/j.procs.2016.09.380>
- Koc, L., Mazzuchi, T. A., & Sarkani, S. (2012). A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. In *Expert Systems with Applications* (Vol. 39, Issue 18, pp. 13492–13500). Elsevier BV. <https://doi.org/10.1016/j.eswa.2012.07.009>
- Zhang, Y., Cao, G., Wang, B., & Li, X. (2019). A novel ensemble method for k-nearest neighbor. In *Pattern Recognition* (Vol. 85, pp. 13–25). Elsevier BV. <https://doi.org/10.1016/j.patcog.2018.08.003>
- Park, Y., Park, S., Jung, W., & Lee, S. (2015). Reversed CF: A fast collaborative filtering algorithm using a k-nearest neighbor graph. In *Expert Systems with Applications* (Vol. 42, Issue 8, pp. 4022–4028). Elsevier BV. <https://doi.org/10.1016/j.eswa.2015.01.001>
- Rao, G., Huang, W., Feng, Z., & Cong, Q. (2018). LSTM with sentence representations for document-level sentiment classification. In *Neurocomputing* (Vol. 308, pp. 49–57). Elsevier BV. <https://doi.org/10.1016/j.neucom.2018.04.045>
- عباس زاده، ملیحه. (۱۳۹۸). تخمین عیار کانسار فسفات اسفوردی با روش رگرسیون بردار پشتیبان. مهندسی منابع معدنی، ۴(۴)، ۱۶–۱. SID. <https://sid.ir/paper/265413/fa>
- عباسی حسن آبادی، نسترن، فیروزی جهان تیغ، فرزاد، و طبرسی، پیام. (۱۳۹۸). تشخیص بیماری سل ریوی با استفاده از هوش مصنوعی (الگوریتم بیز ساده). پیاورد سلامت، ۱۳(۶)، ۴۱۹–۴۲۸. SID. <https://sid.ir/paper/367988/fa>
- فردمال، جواد، امیدی، طاهره، پورالعجل، جلال، و روشنایی، قدرت اله. (۱۳۹۴). مقایسه عملکرد مدل کاکس و روش K-نزدیکترین همسایگی در تخمین بقای بیماران پیوند کلیه. مجله پزشکی بالینی ابن سینا (مجله علمی دانشگاه علوم پزشکی و خدمات بهداشتی درمانی همدان)، ۲۲(۴) (مسلول ۷۸)، ۳۰۸–۳۰۰. SID. <https://sid.ir/paper/17816/fa>
- معلم، محمود، و پویان، علی اکبر. (۱۳۹۸). کشف ناهنجاری با استفاده از کد کننده خودکار مبتنی بر بلوک های LSTM. مدل سازی در مهندسی، ۱۷(۵۶)، ۱۹۱–۲۱۱. SID. <https://sid.ir/paper/358500/fa>

Copper price prediction using support vector regression, naive Bayes, short-term memory, and K-nearest neighbor tools

Amir Ali Faridi

Bachelor's student, Faculty of Computer Engineering, Shomal Amol Non-Profit University

1-1- Abstract

This paper presents a comparative study of four machine learning models, namely Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN), Support Vector Regression (SVR), and Naive Bayes, for the forecasting of copper prices. The study aims to evaluate the performance of these models in predicting future copper prices based on historical data. The dataset used in this study consists of historical copper price data and relevant features that may impact price fluctuations. Each model is trained on a subset of the data and evaluated on a separate test set to assess their predictive accuracy. Various performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared are used to compare the models' forecasting capabilities. The results indicate that the K-Nearest Neighbors (KNN) model outperforms the other models' predictive accuracy for forecasting copper prices. KNN demonstrates superior performance in capturing the underlying patterns and trends in the data, leading to more accurate price predictions. The model's ability to identify similar patterns in the historical data and use them to predict future prices contributes to its success in this task. Overall, this study highlights the importance of choosing an appropriate machine learning model for forecasting tasks and demonstrates the effectiveness of KNN in predicting copper prices. The findings provide valuable insights for researchers and practitioners in the financial markets who rely on accurate price forecasts for decision-making purposes.

Keywords: Copper price prediction, support vector regression, simple Bayes, short-term memory, and K-nearest neighbor.