



پرديس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

پیش‌بینی روند جفت‌ارزها در بازار فارکس (Forex) با استفاده از روش‌های یادگیری ماشین

نگارنده: مهرداد ریاحی

استاد راهنما: دکتر باقر باباعلی

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

مرداد ماه ۱۴۰۰

چکیده

پیش‌بینی داده‌های سری زمانی^۱ همواره موضوع جذابی برای محققین در حوزه یادگیری ماشین بوده است. بازارهای سهام و ارزهای مختلف همواره به عنوان عرصه‌هایی برای به چالش کشیدن قدرت پیش‌بینی و کنجکاوی انسان‌ها مطرح بوده‌اند. بزرگترین این بازارها Forex^۲ نام دارد که در آن جفت‌ارزهای مختلف مورد معامله قرار می‌گیرند، که حجم بسیار بالایی معاملات، آن را به بزرگترین بازار خرید و فروش جفت‌ارزها تبدیل کرده است. نوسانات قیمتی جفت‌ارزها را می‌توان به صورت نموداری برحسب زمان نمایش داد که همین امر باعث می‌شود که بتوان به این موضوع فکر کرد که آیا می‌توان نموداری یافت که بر این نوسانات منطبق باشد؟ در این پژوهش سعی شده با استفاده از ابزار Prophet صرفاً صعودی یا نزولی بودن روند^۳ جفت‌ارز EURUSD^۴ بررسی شود. دادگان استفاده شده در این پژوهش شامل اطلاعات حدوداً ۴۵۰۰ روز از معاملات این جفت‌ارز است که با استفاده از آن‌ها به حدود دقت ۷۵ درصد در پیش‌بینی روند روزانه دست یافته‌ایم.

^۱ Time Series Data

^۲ Foreign Exchange Market

^۳ Up or Down Trend

^۴ Euro vs US Dollar

فهرست مطالب

۱	مفاهیم مقدماتی	۱
۱	۱.۱ قیمت	۱
۲	۱.۱.۱ شمع	۲
۲	۲.۱.۱ انواع شمع	۲
۳	۳.۱.۱ جزئیات شمع	۳
۴	۲.۱ زمان	۴
۴	۱.۲.۱ بازه زمانی	۴
۴	۳.۱ تحلیل در بازارهای مالی	۴
۵	۱.۳.۱ تحلیل فنی	۵
۵	۲.۳.۱ تحلیل بنیادی	۵
۶	۲ آزمایش‌ها و نتایج	۶
۶	۱.۲ جعبه ابزار Prophet	۶
۶	۱.۱.۲ معادلات	۶
۷	۱.۱.۱.۲ $g(t)$	۷
۸	۲.۱.۱.۲ $s(t)$	۸
۸	۳.۱.۱.۲ $h(t)$	۸
۹	۲.۲ پیش‌بینی روند روزانه	۹
۹	۱.۲.۲ صورت مسئله	۹
۹	۲.۲.۲ Regressors	۹
۱۰	۳.۲.۲ داده	۱۰
۱۰	۴.۲.۲ ضریب اطمینان	۱۰

۱۱	پیاده‌سازی	۵.۲.۲
۱۱	تحلیل و بررسی نتایج	۶.۲.۲
۱۳	جداول به دست آمده برای α و <i>LookBack</i> های مختلف	۷.۲.۲
۱۷		جمع بندی و پیشنهادات برای ادامه کار	۳
۱۸		منابع	

فصل ۱

مفاهیم مقدماتی

بازارهای مالی گسترده‌ای امروزه در حوزه های مختلفی مانند بازار مالی خرید و فروش اوراق بهادار، بازار خرید و فروش ارزهای دیجیتال و... وجود دارند. از جمله این بازارها، بازار مالی خرید و فروش جفت‌ارزها معروف به Forex (FX) است. در این بازار معامله‌گران با خرید یک ارز در مقابل ارزی دیگر و بالارفتن ارزش ارزی که در اختیار دارند، سود می‌کنند. از جمله جفت‌ارزهایی که در این بازار معامله می‌شود، جفت‌ارز EURUSD می‌باشد که بیشترین حجم معاملاتی در بازار Forex را به خود اختصاص می‌دهد و بسیاری از تحلیل‌گران این بازار معتقدند این جفت‌ارز به دلیل اثرپذیری شدید از اتفاقات سیاسی، اقتصادی، اجتماعی و... غیر قابل پیش‌بینی است. در ادامه و پیش از بیان روش استفاده شده در این پژوهش برای پیش‌بینی روند این جفت‌ارز، در ابتدا به بیان مفاهیم قیمت و زمان، دو عنصر اساسی در بررسی روند نوسانات بازارهای مالی می‌پردازیم، سپس مواردی را در مورد انواع روش‌های تحلیل در بازارهای مالی بررسی خواهیم کرد.

۱.۱ قیمت

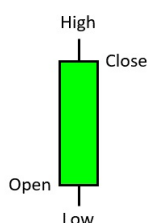
در بازارهای مالی، معاملات صورت گرفته بر روی کالای قابل معامله باعث نوسان قیمتی آن کالا می‌شود. برای بیان این نوسان از بازه‌های زمانی استفاده می‌شود، به طور مثال گفته می‌شود از روز ۱ تیر الی ۵ تیر قیمت x تومان تغییر داشته است. این تغییرات قیمتی در بازه‌های زمانی مشخص معمولاً توسط مفهومی به نام شمع^۱ بیان می‌شود. اسم شمع از روی نحوه نمایش این تغییرات قیمتی بر روی نمودارها برداشته شده است.

^۱Candle

۱.۱.۱ شمع

تعریف ۱.۱ (شمع). شمع از کنار هم قرار گرفتن اولین و آخرین قیمت و بالاترین و پایین‌ترین قیمتی که کالا در آن قیمت‌ها در یک بازه زمانی معامله شده است، تشکیل شده است.

همانطور که انتظار می‌رود می‌توان برای هر بازه زمانی یک شمع قیمتی به وجود آورد مثل ۵ سال، ۵ ساعت، ۵ ماه و... در اصطلاح به اولین قیمتی که کالا در آن معامله می‌شود، قیمت باز شدن^۲ و به آخرین قیمت، قیمت بسته شدن^۳ می‌گویند. همچنین بالاترین قیمت و پایین‌ترین قیمت به ترتیب بالا^۴ و پایین^۵ خوانده می‌شوند. در شکل ۱.۱ نمونه ای از یک شمع ساده را مشاهده می‌کنید.

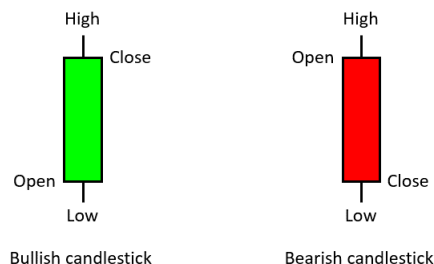


شکل ۱.۱: یک شمع ساده

۲.۱.۱ انواع شمع

شمع‌ها به طور کلی به دو دسته شمع صعودی و نزولی تقسیم می‌شوند: صعودی یا نزولی بودن شمع بر اساس تغییرات قیمت بسته شدن نسبت به قیمت باز شدن شمع تعیین می‌شود. اگر قیمت بسته شدن شمعی بیشتر از قیمت باز شدن آن باشد ($Close > Open$) آنگاه می‌گویند که شمع صعودی است و اگر قیمت بسته شدن شمعی از قیمت باز شدن آن کمتر باشد ($Close < Open$) آنگاه شمع نزولی خوانده می‌شود. شکل ۲.۱ نشان دهنده دو شمع صعودی و نزولی است.

Open^۲
Close^۳
High^۴
Low^۵



شکل ۲.۱: شمع صعودی و نزولی

۳.۱.۱ جزئیات شمع

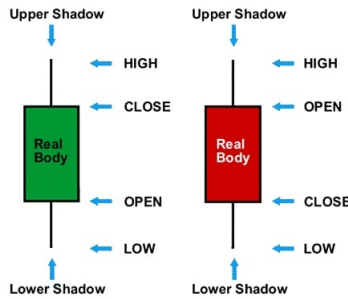
تعریف ۲.۱ (سایه^۶). به فاصله میان بالاترین قیمت و قیمت بسته شدن و همچنین فاصله بین باز شدن و پایین ترین قیمت در شمع صعودی، همچنین به فاصله بین قیمت باز شدن و بالاترین قیمت و همچنین به فاصله بین قیمت بسته شدن و پایین ترین قیمت در شمع نزولی، سایه گفته می شود.

تعریف ۳.۱ (بدنه^۷). به فاصله بین قیمت بسته شدن و قیمت باز شدن، بدنه شمع گفته می شود.

اصولاً به سایه‌ای که بالای قسمت بدنه در شمع ایجاد می شود، سایه بالایی^۸ و به سایه‌ای که در پایین قسمت بدنه ایجاد می شود، سایه پایینی^۹ گفته می شود. تعریف سایه بالایی و پایینی در دو شمع صعودی و نزولی همانطور که انتظار می رود با یکدیگر متفاوت است به طوری که سایه بالایی در شمع صعودی، فاصله بین قیمت بسته شدن با بالاترین قیمت است در حالی که در شمع نزولی، سایه بالایی فاصله بین قیمت باز شدن و بالاترین قیمت است.

تذکر ۴.۱. دقت شود که ممکن است سایه‌های بالایی و پایینی در شمع وجود نداشته باشند به طور مثال اگر قیمت بسته شدن در شمع صعودی همان بالاترین قیمت باشد، آنگاه این شمع، سایه‌ی بالایی نخواهد داشت.

Shadow^۶
 Body^۷
 Upper Shadow^۸
 Lower Shadow^۹



شکل ۳.۱: سایه بالایی و پایینی و بدنه شمع

۲.۱ زمان

برای بیان کردن و نشان دادن تغییرات قیمت به زمان نیاز است، به طور مثال گفته می‌شود یک شمع نشان دهنده تغییرات قیمت در طول دو ساعت گذشته است.

۱.۲.۱ بازه زمانی

برای نشان دادن نوسانات قیمتی، بازه‌های زمانی مرسوم وجود دارد: سالیانه - ماهیانه - هفتگی - روزانه - ۴ ساعته - ۱ ساعته و... منظور از این بازه‌های زمانی این است که به طور مثال اگر بازه زمانی ما روزانه باشد آنگاه هر شمع نشان دهنده تغییرات قیمتی در یک روز است و اینگونه به دست می‌آید که به طور مثال به ازای هر روز ۲۴ شمع ساعتی داریم، آنگاه قیمت باز شدن شمع روزانه برابر است با قیمت باز شدن اولین شمع ساعتی و قیمت بسته شدن آن برابر است با قیمت بسته شدن آخرین شمع ساعتی. بالاترین و پایین‌ترین قیمت نیز بالاترین قیمت و پایین‌ترین قیمت در این ۲۴ شمع در نظر گرفته می‌شود.

۳.۱ تحلیل در بازارهای مالی

تجزیه و تحلیل مالی به طور کلی فرایند ارزیابی مبادلات مربوط به امور مالی برای تعیین عملکرد آنها است. تجزیه و تحلیل مالی معمولاً برای تحلیل اینکه آیا یک موجودیت مالی دارای ثبات، نقدینگی و سودآوری کافی برای تضمین سرمایه‌گذاری است یا خیر، مورد استفاده قرار می‌گیرد. به طور کلی تجزیه و تحلیل مالی به دو نوع مختلف تقسیم می‌شود:

- تحلیل فنی^{۱۰}
- تحلیل بنیادی^{۱۱}

در ادامه به توضیح مختصری از هر کدام از موارد فوق خواهیم پرداخت.

۱.۳.۱ تحلیل فنی

در تحلیل فنی یا تکنیکال سعی می‌شود که با جستجوی روندها و الگوها در نمودار قیمت، روند حرکتی قیمت درک شود. به همین دلیل است که به آن تحلیل نموداری نیز گفته می‌شود. در واقع در تحلیل تکنیکال فرض می‌شود که قیمت یک موجودیت، منعکس کننده تمام اطلاعات موجود، راجع به آن موجودیت می‌باشد.

در تحلیل تکنیکال از شاخص‌های فنی^{۱۲} برای تفسیر حرکات قیمت بر روی نمودارها کمک گرفته می‌شود. [۲] این شاخص‌ها از اطلاعات معاملات صورت گرفته بر روی سهام، کالا، جفت‌ارز و به طور کلی هر گونه موجودیت مالی به دست می‌آیند، به طور مثال یکی از این شاخص‌ها میانگین متحرک^{۱۳} است که میانگین قیمت بسته شدن را در k روز گذشته برای هر روز به دست می‌دهد.

۲.۳.۱ تحلیل بنیادی

تحلیل بنیادی، در واقع به دنبال یافتن ارزش ذاتی موجودیت‌ها با استفاده از بررسی عوامل اقتصادی مرتبط است. در این نوع تحلیل، تحلیل‌گران هر عاملی را که بتواند بر ارزش ذاتی یک سهم، جفت‌ارز و... تأثیر بگذارد، از عوامل و متغیرهای کلان اقتصادی مانند شرایط یک صنعت، میزان تورم و... تا عوامل خرد مانند تأثیر مدیریت یک شرکت بر سودآوری آن، مطالعه می‌کنند.

در تحلیل بنیادی به طور عمده از گزارش‌های اقتصادی، اخبار سیاسی و اقتصادی و شاخص‌های کلان اقتصادی برای تفسیر معاملات صورت گرفته و نوسانات قیمت استفاده می‌شود. [۲]

Technical Analysis^{۱۰}
 Fundamental Analysis^{۱۱}
 Technical Indicators^{۱۲}
 Moving Average^{۱۳}

فصل ۲

آزمایش‌ها و نتایج

پیش از توضیحات در مورد نحوه پیش‌بینی روند جفت‌ارز EURUSD به توضیح مختصری در مورد ابزار Prophet که ابزار مورد استفاده در این پژوهش بوده است، پرداخته می‌شود.

۱.۲ جعبه ابزار Prophet

Prophet یک ابزار برای پیش‌بینی داده‌های سری زمانی است که تلاش می‌کند روندهای غیرخطی^۱ داده را با توجه به تغییرات روزانه، هفتگی و فصلی و سالیانه به دست بیاورد.

۱.۱.۲ معادلات

ابزار Prophet سعی می‌کند که خواسته مسئله $y(t)$ را بر اساس عوامل دیگری به دست بیاورد که فرمول کلی که این ابزار برای پیش‌بینی خواسته مسئله استفاده می‌کند، بدین شرح است:

$$y(t) = g(t) + s(t) + h(t) + \beta X(t) + \epsilon(t)$$

- $g(t)$: بخش مربوط به روند کلی
- $s(t)$: بخش مربوط به تغییرات دوره‌ای و تناوبی (مانند سالیانه ، ماهیانه ، هفتگی)
- $h(t)$: بخش مربوط به تأثیرات روزهای تعطیل

^۱Non-Linear Trends

• $\beta X(t)$: بخش مربوط به تأثیر regressor ها

• $\epsilon(t)$: بخش مربوط به outlier ها

در ادامه برای هر کدام از این بخش ها به تناسب آنچه برای این مسئله نیاز است، توضیحاتی داده می شود.

۱.۱.۱.۲ $g(t)$

در Prophet به طور کلی دو نوع روند برای مدل ها پیشنهاد می شود:

• Logistic : اگر با توجه به داده ها به گونه ای انتظار نوعی اشباع شدن^۲ را در نمودار خواسته مسئله داشته باشیم، استفاده از این نوع روند مناسب تر است.

• Linear : اگر نمودار خواسته مسئله به صورت قطعه قطعه خطی باشد، استفاده از روند Linear بهتر است.

برای روشن تر شدن تفاوت بین این دو نوع روند توجه کنید که به طور مثال برای مدل کردن روند افزایش جمعیت یا افزایش کاربران یک شبکه اجتماعی از روند Logistic استفاده می کنیم. دقت شود که در این نوع روندها همواره مقدار خواسته مسئله به سمت مقداری به نام ظرفیت اشباع^۳ نزدیک می شود. در مورد مثالی که برای کاربران یک شبکه اجتماعی ذکر شد، ظرفیت اشباع برای کاربران این شبکه در یک منطقه می تواند همه افرادی باشد که در این منطقه به اینترنت دسترسی دارند. برای مسائلی که در آن ها رشد اشباع مشاهده نمی شود، استفاده از روند Linear بهتر است. [۱] طبق شکل ۱.۲ مشاهده می شود که استفاده از روند Linear برای پیش بینی روند جفت ارز EURUSD مناسب تر است.



شکل ۱.۲: بخشی از نمودار قیمت جفت ارز EURUSD

Saturation^۲
Carrying Capacity^۳

$$s(t) \quad ۲.۱.۱.۲$$

اکثر داده‌های سری زمانی که رفتار انسان‌ها بر روی آن‌ها تأثیرگذار است معمولاً تحت تأثیر این رفتارها دارای نتایج تناوبی هستند. به طور مثال ۵ روز کاری در هر هفته می‌تواند اثری بر روی داده‌ها داشته باشد که هر هفته تکرار می‌شود. بخش مربوط به تغییرات تناوبی سعی می‌کند، انعطافی به مدل بدهد تا مدل بتواند بر تغییرات روزانه، هفتگی و سالیانه منطبق شود.

Prophet برای دستیابی به مدلی که با اثرات تناوبی انطباق داشته باشد، از سری فوریه استفاده می‌کند. بنابراین، تخمین تقریبی اثرات تناوبی دلخواه با یک سری استاندارد فوریه مرتبط است:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

که در فرمول بالا P نشان دهنده دوره تناوب است. برای مثال برای دوره‌ی هفتگی داریم: $P = 7$. در نهایت از فرمول بالا متوجه می‌شویم که Prophet برای پیدا کردن این الگوهای تناوبی باید $2N$ متغیر را تخمین بزند: $\alpha = [a_1, b_1, \dots, a_N, b_N]$. همچنین مقدار N توسط خود ابزار Prophet تعیین می‌شود. [۱]

$$h(t) \quad ۳.۱.۱.۲$$

تأثیرات تعطیلات در روند داده‌های سری زمانی ممکن است تکرار شوند و به همین دلیل است که این بخش اهمیت پیدا می‌کند. واضح است که این بخش برای بررسی مسئله مورد بحث در این پژوهش نیز اهمیت دارد، چرا که برای مثال در زمان تعطیلات بازارهای مالی آمریکا، نسبت یورو به دلار تحت تأثیر این تعطیلات قرار می‌گیرند.

ابزار Prophet برای مدل کردن تأثیرات تعطیلات به گونه‌ای شبیه به بخش دوره‌ای عمل می‌کند. بدین صورت که اگر فرض شود برای هر روز تعطیل i ، D_i مجموعه‌ای از تاریخ‌های گذشته و آینده این روز تعطیل باشد. آنگاه برای هر روز مثل t یک تابع مشخصه در نظر گرفته می‌شود که یکسان بودن این روز با روز تعطیل i را نشان می‌دهد ($t \in D_i$). سپس برای هر تعطیلی متناسب با تغییراتی که در پیش‌بینی ایجاد می‌کند، پارامتری مانند κ_i در نظر گرفته می‌شود. در نهایت می‌توان فرمول کلی را که برای به دست آوردن $h(t)$ در نظر گرفته می‌شود، اینگونه نمایش داد:

$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_n)]$$

$$\kappa = [\kappa_1, \dots, \kappa_n]$$

$$h(t) = Z(t)\kappa$$

که k توسط ابزار Prophet تخمین زده می‌شود. [۱]

یادداشت ۱.۲. در مقاله ای که در مورد ابزار Prophet نوشته شده است، [۱] در مورد بخش‌های $X(t)$ و $\epsilon(t)$ توضیحات بیشتری داده نشده است.

۲.۲ پیش‌بینی روند روزانه

در این بخش به توضیح روش استفاده شده در این پژوهش برای پیش‌بینی روند جفت‌ارز EURUSD در بازه زمانی روزانه پرداخته می‌شود. در این پژوهش سعی شده که با تکیه بر تحلیل تکنیکال و استفاده از ابزار Prophet و استخراج چندین ویژگی^۴، روند این جفت‌ارز پیش‌بینی شود. پس از شناسایی این ویژگی‌ها همانطور که در فرمول ۱.۱.۲ آمده است، آن‌ها را تحت عنوان Regressor به مدل می‌دهیم تا مدل به کمک آن‌ها بهتر بتواند پیش‌بینی‌ها را انجام دهد. تنوع و تعداد ویژگی‌هایی که ممکن است وجود داشته باشند، به طرز گسترده‌ای زیاد است، به گونه‌ای که نمی‌توان همه آن‌ها را مورد آزمایش قرار داد. همچنین کمبود داده نیز به طور کلی یکی از معضلات این مسئله است زیرا اگر حتی هر روز سال هم در بازار FX معامله‌ای ثبت شود، باز هم به ازای هر سال 365 داده وجود خواهد داشت که به نسبت کم است (در ضمن می‌دانیم آخر هفته‌ها این بازار تعطیل است).

۱.۲.۲ صورت مسئله

مسئله اینگونه است که می‌خواهیم صعودی یا نزولی بودن روند جفت‌ارز مورد نظر را در یک روز پیش‌بینی کنیم. ورودی، اطلاعات معاملاتی مربوط به k روز قبل از این روز است و انتظار داریم مدل، یکی از سه وضعیت صعودی، نزولی و یا عدم اطمینان از پیش‌بینی انجام شده، را در خروجی اعلام کند.

۲.۲.۲ Regressors

- برای بخش Regressor ها همانطور که اشاره شد، تنوع و تعداد Regressor ها میتواند خیلی زیاد باشد. ویژگی‌هایی که در این پژوهش از آن‌ها استفاده شده بدین شرح هستند:
- صعودی یا نزولی بودن k روز اخیر: به این صورت که به ازای روزهایی که صعودی هستند ($Close > Open$) این مقدار برابر 1 در نظر گرفته شده و به ازای روزهایی که نزولی هستند ($Close < Open$) این مقدار برابر 0 در نظر گرفته شده است.
 - سایه‌های بالایی k روز اخیر: یعنی مقدار سایه‌های بالایی در شمع‌های k روز گذشته به ازای هر روز به مدل داده

^۴Feature

می شود.

- سایه‌های پایینی k روز اخیر: به طور مشابه با سایه بالایی، سایه پایینی نیز محاسبه می‌شود و به مدل داده می‌شود.
 - بدنه‌های شمع‌های k روز اخیر: به این صورت که بدنه شمع در k روز گذشته به مدل داده می‌شود.
- با توجه به موارد فوق، واضح است که اگر برای پیش‌بینی هر روز از داده‌های k روز قبل آن استفاده شود، آنگاه به ازای هر روز $4k$ ویژگی به مدل داده می‌شود.

۳.۲.۲ داده

مجموعه دادگان استفاده شده در این پژوهش، شامل اطلاعات حدود ۴۵۰۰ روز از معاملات جفت‌ارز EURUSD است که با توجه توضیحات داده شده در بخش قبل به ۴ ستون (*Open, High, Low, Close*) از این اطلاعات نیاز داریم. با استفاده از این ستون‌ها و تشخیص صعودی یا نزولی بودن شمع‌ها، می‌توان سایه‌های بالایی و پایینی را نیز برای هر شمع به دست آورد.

برای بررسی تأثیرگذاری میزان داده‌ها بر روند کار، تعداد داده‌های موجود در مجموعه‌ی آموزش^۵ تغییر می‌کند. اما تعداد داده‌های مجموعه تست^۶ در کل این آزمایش‌ها ثابت و برابر ۵۰۰ باقی می‌ماند. در واقع روند کار اینگونه است که اگر تعداد کل داده‌ها n باشد و m ، تعداد داده‌ها در مجموعه داده‌های تست باشد، m روز آخر از مجموعه دادگان جدا شده و بقیه به عنوان داده آموزشی استفاده می‌شوند. به همین دلیل است که با وجود تفاوت در تعداد داده‌های آموزشی اما داده‌های تست یکسان می‌مانند (چون در هر حالت m روز آخر مجموعه دادگان مربوط به داده‌های تست هستند).

۴.۲.۲ ضریب اطمینان

همانطور که توضیح داده شد، ابزار Prophet سعی می‌کند تا بتواند تابعی بر روی نمودار مقدار خواسته شده (صعودی یا نزولی بودن روز که مقداری در مجموعه $\{0, 1\}$ است). بر حسب زمان پیدا کند، به طوری که مقدار این تابع به مقدار خواسته شده در هر زمان نزدیک باشد. اگر فرض کنیم که محور پارامتر خواسته شده y نامیده شود. همانطور که انتظار می‌رود تابعی که Prophet پیدا می‌کند، تابعی است که به طور کلی حول خط $y = 0.5$ نوسان می‌کند. حال در هر نقطه که مقدار این تابع به 1 نزدیک‌تر باشد، بدین معنی است که Prophet با اطمینان بیشتری صعودی بودن آن روز را پیش‌بینی می‌کند و برعکس هر چقدر که این تابع نسبت به خط $y = 0.5$ به مقدار 0 نزدیک‌تر باشد، یعنی Prophet اطمینان بیشتری از نزولی بودن آن روز دارد. حال با استفاده از این ویژگی می‌توان پارامتری به نام

^۵ Train Set

^۶ Test Set

ضریب اطمینان تعریف کرد که آن را با نماد α نشان می‌دهیم. با توجه به این ضریب اطمینان می‌توان پیش‌بینی‌های مدل را به سه دسته تقسیم‌بندی کرد:

• $\hat{y} > 0.5 + \alpha$: در این صورت می‌توان گفت پیش‌بینی مدل این است که این روز، روند صعودی خواهد داشت.

• $|\hat{y} - 0.5| \leq \alpha$: در این صورت می‌توان گفت که پیش‌بینی مدل در محدوده مطمئن قرار ندارد و پیش‌بینی آن معتبر نیست.

• $\hat{y} < 0.5 - \alpha$: در این صورت می‌توان گفت که پیش‌بینی مدل این است که این روز، روند نزولی خواهد داشت.

یادداشت ۲.۲. در این پژوهش مسئله مورد بحث پیش‌بینی روند روزانه است ولی اصولاً برای هر بازه زمانی که داده‌های آن موجود باشند، می‌توان این پیش‌بینی‌ها را انجام داد. [بخش ۱.۲.۱] هر چند که ابزار Prophet در پیش‌بینی روندها در بازه‌های زمانی کوچک‌تر از یک روز مثل ۴ ساعت و... مشکلاتی دارد.

۵.۲.۲ پیاده‌سازی

برای پیاده‌سازی روش استفاده شده در این پژوهش از زبان پایتون استفاده شده است. روند کلی کار به اینگونه است:

ابتدا داده‌های ورودی وارد مرحله پیش‌پردازش^۷ می‌شوند که در این مرحله با توجه به اینکه از اطلاعات چه تعدادی از روزهای اخیر می‌خواهیم در پیش‌بینی استفاده کنیم، این اطلاعات استخراج می‌شوند. سپس مدل با توجه به Regressor های توضیح داده شده و همچنین مشخص کردن روزهای تعطیل [بخش ۳.۱.۱.۲] (در پیش‌بینی جفت‌ارز EURUSD تعطیلات کشور آمریکا اهمیت دارد.) ساخته می‌شود. پس از آموزش مدل با مجموعه دادگان آموزشی، دادگان تست برای پیش‌بینی به مدل داده می‌شوند. نتایج به دست آمده در این مرحله با توجه به ضریب اطمینان تعیین شده، نتایج نهایی را به دست می‌دهند.

۶.۲.۲ تحلیل و بررسی نتایج

پیش از تحلیل و بررسی نتایج برای سادگی در توضیحات مواردی را تعریف می‌کنیم.

تعریف ۳.۲ (تصمیم معتبر (پیش‌بینی معتبر)). اگر ضریب اطمینان α باشد، گوییم مدل برای روز t تصمیم معتبر گرفته است، هرگاه پیش‌بینی مدل برای آن روز در بازه اطمینان [بخش ۴.۲.۲] قرار بگیرد $(|\hat{y}(t) - 0.5| > \alpha)$.

^۷Preprocessing

تعریف ۴.۲ (*LookBack*). همانطور که در بخش ۲.۲.۲ اشاره شد، برای پیش‌بینی روند هر روز به اطلاعات k روز قبل از آن نیاز است، در اینجا مقدار k را *LookBack* می‌نامیم.

حال به تحلیل نتایج به دست آمده به ازای α و *LookBack* های مختلف می‌پردازیم.

تذکر ۵.۲. در بخش ۷.۲.۲ برای برخی از مقادیر α و *LookBack* جداولی آورده شده است. در این جداول منظور از *N.O.D* تعداد روزهایی هست که مدل توانسته برای آن‌ها با در نظر گرفتن ضریب اطمینان، پیش‌بینی معتبر انجام دهد. همچنین در مواردی درصد دقت ۵- گزارش شده است، این بدان معنی است که مدل نتوانسته برای هیچ کدام از داده‌های این مجموعه با توجه به ضریب اطمینان، تصمیم معتبری بگیرد.

به طور کلی در این مسئله یکی از مشکلات همانطور که اشاره شد، کم بودن تعداد داده است. همانطور که در بعضی موارد و برای برخی مقادیر α ، تعداد روزهایی که مدل توانسته برای آن‌ها تصمیم معتبر بگیرد، کمتر از ۱۰ تا است. به همین دلیل تفسیر کردن درصد دقت مدل در این موارد اعتبار زیادی نخواهد داشت.

همانطور که از جداول در بخش ۷.۲.۲ مشخص است، به طور کلی وقتی که مقدار α افزایش می‌یابد، دقت مدل افزایش پیدا می‌کند. از طرفی هر چقدر ضریب اطمینان (α) افزایش پیدا کند، تعداد روزهایی که مدل می‌تواند در مورد آن‌ها پیش‌بینی معتبر انجام دهد، کاهش می‌یابد. به طور کلی می‌توان گفت که مقادیر $\alpha > 0.2$ در عمل کاربردی ندارند، زیرا تعداد روزهایی که مدل می‌تواند برای آن‌ها پیش‌بینی معتبر انجام دهد، بسیار کم خواهد بود.

در مورد مقدار متغیر *LookBack* نیز برای مقادیر $LookBack < 3$ و $LookBack > 6$ ، نتایج به دست آمده نسبت به نتایج به دست آمده برای این متغیر در محدوده $\{3, \dots, 6\}$ بهبود چشم‌گیری ندارند و حتی گاهی دقت آن‌ها کمتر است. از طرفی هر چقدر مقدار متغیر *LookBack* افزایش می‌یابد، بار محاسباتی مرحله پیش‌پردازش دادگان افزایش می‌یابد. به همین دلیل استفاده از مقادیر بزرگ برای متغیر *LookBack* توصیه نمی‌شود. به طور کلی می‌توان گفت که وقتی متغیر *LookBack* افزایش می‌یابد، تعداد روزهایی که مدل می‌تواند برای آن‌ها تصمیم معتبر بگیرد، افزایش می‌یابد. این افزایش مستقل از دقت مدل است، یعنی احتمالاً اگر با مقدار α و مجموعه آموزشی یکسان دو مدل، یکی با مقدار $LookBack = 9$ و دیگری با مقدار $LookBack = 3$ ساخته شوند، مدل اول در مورد تعداد روزهای بیشتری پیش‌بینی معتبر انجام می‌دهد اما دقت این پیش‌بینی‌ها از مدل دوم پایین‌تر خواهند بود.

در نهایت می‌توان گفت که با توجه به این مقدار داده آموزشی، بهترین مقدار برای متغیر *LookBack*، 4 یا 5 است (با توجه به تعداد روزهایی که نیاز به پیش‌بینی معتبر داریم، بین این دو مقدار باید تصمیم‌گیری شود). برای انتخاب مقدار α نیز مبادله و معاوضه بین دقت و تعداد پیش‌بینی‌های معتبر باید مدنظر قرار بگیرد.

۷.۲.۲ جداول به دست آمده برای α و $LookBack$ های مختلف

جدول ۱.۲: جدول مقادیر دقت و تعداد تصمیم‌ها برای $LookBack = 3$

Look Back = 3							
	Train Size	Train Acc %	Test Acc %	Train N.O.D	Test N.O.D	Train Acc Avg %	Test Acc Avg %
$\alpha = 0.10$	2500	64.65	44.33	447	97	64.48	48.65
	3000	64.9	47.31	396	93		
	3500	63.24	52.94	272	34		
	4000	65.16	50	353	38		
$\alpha = 0.15$	2500	71.43	38.89	105	18	71.39	47.45
	3000	69.89	42.86	105	18		
	3500	74.07	44.44	93	14		
	4000	70.18	63.64	57	11		
$\alpha = 0.20$	2500	77.19	55.56	57	9	78.16	67.16
	3000	75.47	71.43	53	7		
	3500	76.19	66.67	42	6		
	4000	83.61	75	37	4		
$\alpha = 0.25$	2500	78.57	60	42	5	79.47	71.66
	3000	80.95	60	42	5		
	3500	75.76	100	33	3		
	4000	82.61	66.67	23	3		
$\alpha = 0.30$	2500	84.21	50	19	2	82.06	75
	3000	75	50	28	2		
	3500	85.71	100	14	1		
	4000	83.33	100	12	1		

جدول ۲.۲: جدول مقادیر دقت و تعداد تصمیم‌ها برای $LookBack = 4$

Look Back = 4							
	Train Size	Train Acc %	Test Acc %	Train N.O.D	Test N.O.D	Train Acc Avg %	Test Acc Avg %
$\alpha = 0.10$	2500	65.87	44.27	460	110	65.16	44.23
	3000	64.39	44.38	410	69		
	3500	65.31	44.74	320	38		
	4000	65.09	44.44	424	36		
$\alpha = 0.15$	2500	72.12	36.36	104	22	71.27	56.1
	3000	71.28	55.56	94	9		
	3500	74.19	62.5	62	8		
	4000	67.5	70	80	10		
$\alpha = 0.20$	2500	72.73	25	33	4	74.84	57.29
	3000	75.47	62.5	53	8		
	3500	76.19	75	42	4		
	4000	75	66.67	44	6		
$\alpha = 0.25$	2500	90	0	10	1	86.29	56.66
	3000	82.5	60	40	5		
	3500	86.96	100	23	1		
	4000	85.71	66.67	28	3		
$\alpha = 0.30$	2500	100	0	1	1	87.02	62.5
	3000	73.08	50	26	2		
	3500	91.67	100	12	1		
	4000	83.33	100	18	1		

جدول ۳.۲: جدول مقادیر دقت و تعداد تصمیم‌ها برای $LookBack = 5$

Look Back = 5							
	Train Size	Train Acc %	Test Acc %	Train N.O.D	Test N.O.D	Train Acc Avg %	Test Acc Avg %
$\alpha = 0.10$	2500	66.73	49.17	520	120	65.57	51.08
	3000	66.67	50.54	519	93		
	3500	66.39	50.94	482	53		
	4000	62.5	53.66	472	41		
$\alpha = 0.15$	2500	73.15	50	158	22	72.03	56.79
	3000	73.1	60	145	15		
	3500	73	53.85	100	13		
	4000	68.87	63.34	106	11		
$\alpha = 0.20$	2500	77.05	45.45	61	11	74.48	64.49
	3000	75	62.5	56	8		
	3500	72.97	75	37	4		
	4000	72.92	75	48	4		
$\alpha = 0.25$	2500	70.27	60	37	5	81.01	47.09
	3000	75	66.67	44	6		
	3500	89.47	-5	19	0		
	4000	89.29	66.67	28	3		
$\alpha = 0.30$	2500	80	33.33	20	3	81.02	40.41
	3000	74.07	33.33	27	3		
	3500	83.33	-5	6	0		
	4000	86.67	100	15	1		

جدول ۴.۲: جدول مقادیر دقت و تعداد تصمیم‌ها برای $LookBack = 6$

Look Back = 6							
	Train Size	Train Acc %	Test Acc %	Train N.O.D	Test N.O.D	Train Acc Avg %	Test Acc Avg %
$\alpha = 0.10$	2500	68.12	47.86	549	117	65.71	53.12
	3000	64.32	47.31	555	93		
	3500	65.61	58.82	535	68		
	4000	64.82	58.49	577	53		
$\alpha = 0.15$	2500	71.43	52.94	168	34	69.36	58.85
	3000	69.62	52.17	158	23		
	3500	69.74	57.14	152	14		
	4000	66.67	57.14	132	14		
$\alpha = 0.20$	2500	73.44	37.5	64	8	72.81	61.51
	3000	71.7	57.14	53	7		
	3500	75.41	71.43	61	7		
	4000	70.69	80	58	5		
$\alpha = 0.25$	2500	69.44	60	36	5	71.69	64.17
	3000	75	50	32	4		
	3500	75.68	80	37	5		
	4000	66.67	66.67	34	3		
$\alpha = 0.30$	2500	85	33.33	20	3	79.38	57.08
	3000	64.29	-5	14	0		
	3500	80	100	25	1		
	4000	88.24	100	17	1		

فصل ۳

جمع بندی و پیشنهادات برای ادامه کار

آنچه در این پژوهش انجام گرفت، گوشه‌ای از هزاران پژوهشی است که می‌توان در حوزه پیش‌بینی بازارهای مالی بالاخص پیش‌بینی جفت‌ارزها در بازار Forex انجام داد. علاوه بر تغییر متد و استفاده از انواع روش‌های پیش‌بینی، می‌توان ویژگی‌های متنوعی با تکیه بر دیدگاه‌های بنیادی و تکنیکال استخراج نموده و با استفاده از این ویژگی‌ها به پیش‌بینی روند حرکتی قیمت این جفت‌ارزها پرداخت.

در ادامه پیشنهاداتی برای ادامه کار در این حوزه مطرح می‌شوند که احتمالاً به بهبود دقت پیش‌بینی کمک نموده و می‌توانند مورد آزمایش و بررسی قرار بگیرند:

- استفاده از شاخص‌های کلان موثر بر روند جفت‌ارزها بالاخص EURUSD با تکیه بر تحلیل بنیادی، مانند شاخص S&P 500^۱.
- استفاده از سایر شاخص‌های تحلیل تکنیکال مانند RSI^۲، MACD^۳ و انواع میانگین‌های متحرک.
- استفاده از انواع شبکه‌های عصبی بازگشتی (RNN)^۴ در پیش‌بینی روند حرکتی قیمت. [۳]

موارد فوق را به صورت ترکیبی با یکدیگر نیز می‌توان استفاده کرد. به طور مثال استفاده از متغیرهای کلان در کنار استفاده از شاخص‌های تکنیکال می‌تواند بسیار موثر باشد. [۴]

Standard and Poor's^۱

Relative Strength Index^۲

Moving Average Convergence/Divergence^۳

Recurrent Neural Network^۴

- [١] Sean J Taylor, Benjamin Letham, *Forecasting at scale*, Facebook, Menlo Park, California, United States, 2017
- [٢] Alex Nekritin, Walter Peters, *Naked Forex: High-Probability Techniques for Trading Without Indicators*, Wiley, 2012
- [٣] Francesca Lazzeri, *Machine Learning for Time Series Forecasting with Python*, Wiley, 2020
- [٤] Deniz Can Yildirim, Ismail Hakkı Toroslu, Ugo Fiore, *Forecasting directional movement of Forex data using LSTM with technical and macroeconomic indicators*, 2021, <https://doi.org/10.1186/s40854-020-00220-2>

Abstract

Time series forecasting has always been an attractive topic for machine learning researchers. Various stock markets and currencies have always been considered as arenas to challenge the power of prediction and curiosity of men. The largest of these markets is called Forex, in which different currency pairs are traded that the high volume of transactions has made it to the largest market for trading currency pairs. The price fluctuations of the currency pairs can be displayed as a chart according to the time, which makes it possible to think whether we can find a chart that matches those volatilities? In this research, we tried to investigate upside or downside trend of the EURUSD currency pair using the Prophet toolbox. The data used in this research include information about 4,500 days of transactions of this pair, which by using them we could achieve about 75% in predicting trend in daily time frame.



College of Science
School of Mathematics, Statistics, and Computer Science

Predicting Currency Pair Trends In Forex Market Using Machine Learning Methods

Mehrdad Riyahi

Supervisor: Bagher BabaAli

A thesis submitted in partial fulfillment of the requirements for
the degree of B.Sc. in Computer Science

August 2021