



Project report

STOCK PREDICTION

Phase 1

BY:

Sharif Indicator Group

Negin Hashemi

Ali Abdollahi

Alireza Taherian

Soroush vafaietabar

گزارش فاز اول پروژه

یادگیری مفاهیم پیشرفته یادگیری ماشین

استاد

دکتر زارع زاده

اعضای گروه

نگین هاشمی

علی عبدالهی

علیرضا طاهریان

سروش وفایی تبار

دانشگاه صنعتی شریف

بهار ۱۴۰۲

فهرست

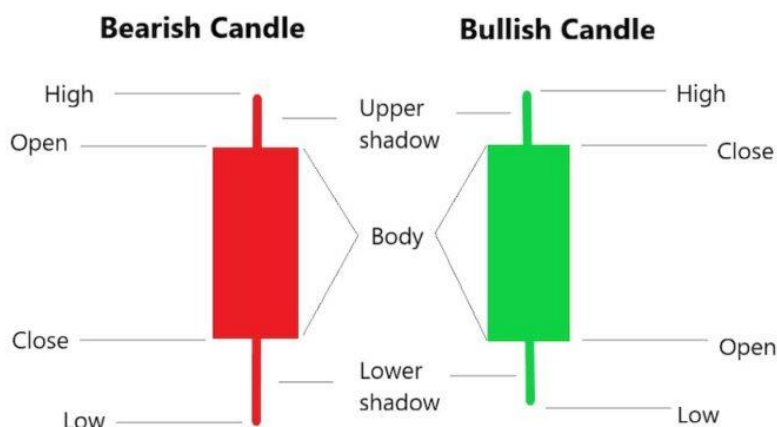
۱ مقدمه
۳ جمع‌آوری داده
۵ آماده‌سازی داده
۷ اندیکاتور RSI
۷ Stochastic Oscillator
۸ اندیکاتور Willams %R
۹ اندیکاتور Accumulation/Distribution
۹ اندیکاتور CMF
۱۰ اندیکاتور OBV
۱۰ اندیکاتور MACD

فهرست اشکال

- شکل ۱- بخش های مختلف کندل قیمتی ۱
- شکل ۲- نمودار کندل های قیمتی سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۴
- شکل ۳- نمودار کندل های قیمتی سهم بوعلی در تایم فریم ۳۰ دقیقه ای پس از حذف مجمع ۶
- شکل ۴- نمودار اندیکاتور RSI سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۷
- شکل ۵- نمودار اندیکاتور Stochastic Oscillator سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۸
- شکل ۶- نمودار اندیکاتور Williams%R سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۸
- شکل ۷- نمودار اندیکاتور CMF سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۹
- شکل ۸- نمودار اندیکاتور MACD سهم بوعلی در تایم فریم ۳۰ دقیقه ای ۱۰
- شکل ۹- ماتریس correlation ویژگی ها ۱۱

مقدمه

بازار سهام بی‌ثبات، پیچیده و پویا است و بدون شک پیش‌بینی آن یکی از پرچالش‌ترین کارها در پیش‌بینی سری‌های زمانی^۱ است. از جمله اطلاعاتی که برای سهام در پیش‌بینی مورد استفاده قرار می‌گیرد، اطلاعات مربوط به کندل‌های قیمتی^۲ است. هر کندل نشان‌دهنده‌ی قیمت در دوره‌های زمانی یکسان است. در واقع هر کندل شامل بیشترین (high) و کمترین قیمت (low)، قیمت باز شدن (open) و بسته شدن (close) یک سهم در یک بازه‌ی زمانی مشخص است. که این کندل‌ها روی یک بازه‌ی زمانی (تایم فریم)^۳ ایجاد شده و مقادیر کندل‌ها در آن تایم‌فریم محاسبه می‌شود. به طور مثال در یک تایم‌فریم ۱ دقیقه‌ای، هر کندل شامل یک قیمت ابتدایی (open)، انتهایی (close)، بالاترین (high) و پایین‌ترین (low) در طول ۱ دقیقه‌ای است. بخش‌های مختلف کندل قیمتی در شکل ۱ آمده است.



شکل ۱- بخش‌های مختلف کندل قیمتی

^۱Time series

^۲Candlestick

^۳Time frame

نوسانات قیمتی در تایم فریم های کوچک، زیاد و در تایم فریم های بزرگ، کم است. به همین منظور از مقادیر اندیکاتورهای روندنما^۴ جهت نوسان گیری استفاده می شوند. اندیکاتورهای پرکاربردترین ابزارهای تحلیل تکنیکال هستند که با تجزیه و تحلیل اطلاعات کندل ها در یک پنجره با سایز مناسب، در پیدا کردن جهت روند قیمت، شتاب حرکتی و نوسانات قیمت کمک کنند که توضیحات اندیکاتورهای مختلفی از قبیل RSI^۵ و OBV^۶ ... در بخش آماده سازی داده آمده است.

^۴Indicator

^۵Relative Strength Index

^۶On-Balance Volume

جمع‌آوری داده

در بازار بورس ایران هیچ API وجود ندارد که بتوان از آن تمام داده‌های قیمتی سهام را گرفت و مطابق آن‌ها کندل‌های زمانی در تایم‌فریم‌های مختلف ساخت. اما یک کتابخانه پایتون وجود دارد به نام `finpy_tse` که داده‌های مختلفی از بازار بورس ایران را از وبسایت `TseTmc.ir` (شرکت مدیریت فناوری بورس تهران) استخراج کرده و در اختیار قرار می‌دهد که یکی از آن‌ها ریز معاملات یک نماد است که می‌توان با استفاده از آن کندل‌های زمانی در تایم‌فریم‌های مختلف زمانی را ساخت و همچنین حجم معاملات در هر کندل زمانی را استخراج کرد.

در جمع‌آوری اولیه داده‌ها ریز معاملات ۲۲ سهام بورس ایران از چهار گروه صنعتی را استخراج کرده و سپس با استفاده از آن‌ها کندل‌های قیمتی و حجمی در تایم‌فریم‌های مختلف زمانی از قبیل ۵ دقیقه‌ای، ۳۰ دقیقه‌ای، ۱ ساعته و یک روزه را ساختیم.

اسامی سهام استخراج شده به صورت زیر می‌باشد:

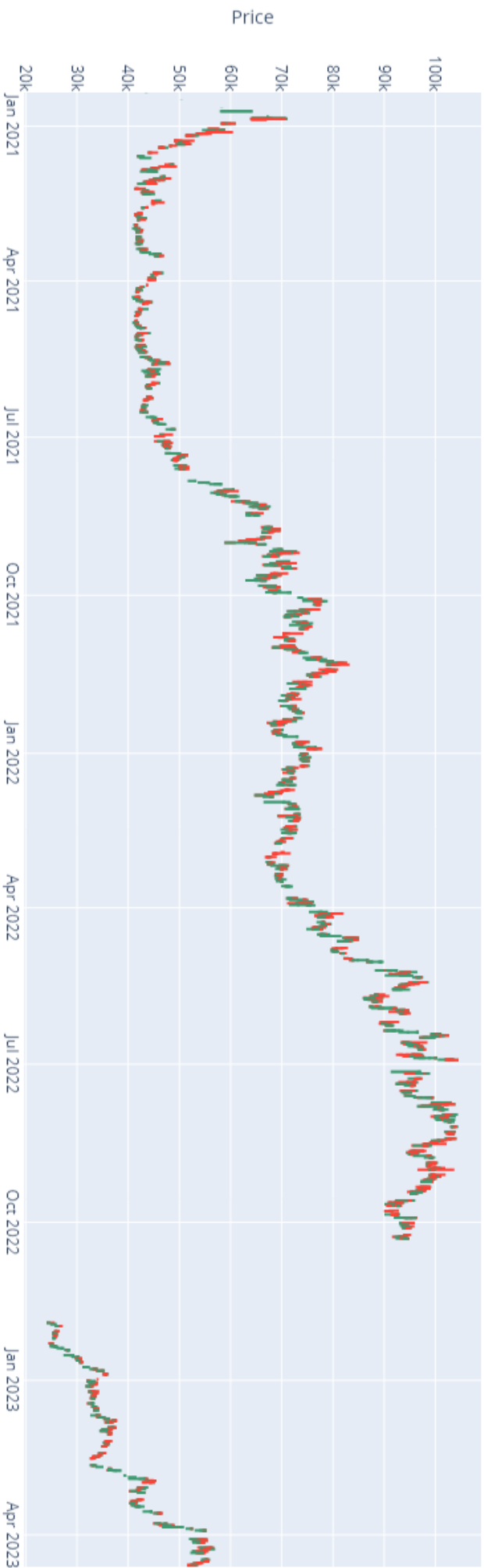
بوعلی، دارا یکم، دی، فارس، نوری، پارسان، شاوان، شاهر، شبندر، شگویا، شلرد، شرانل، تاپیکو، وبملت، وبصادر، وخارزم، وپارس، وساپا، وسپهر، وتجارت، وتجارت، وتوصا، پترول، پالایش

کدهای مربوط به این بخش در فایل به آدرس زیر قابل مشاهده است.

<https://github.com/NeginHashemi/StockPrediction/tree/main/crawl>

نمونه نمودار کندل‌های قیمتی برای سهم بوعلی در تایم‌فریم ۳۰ دقیقه‌ای در شکل ۲ آمده است.

Candlestick Chart



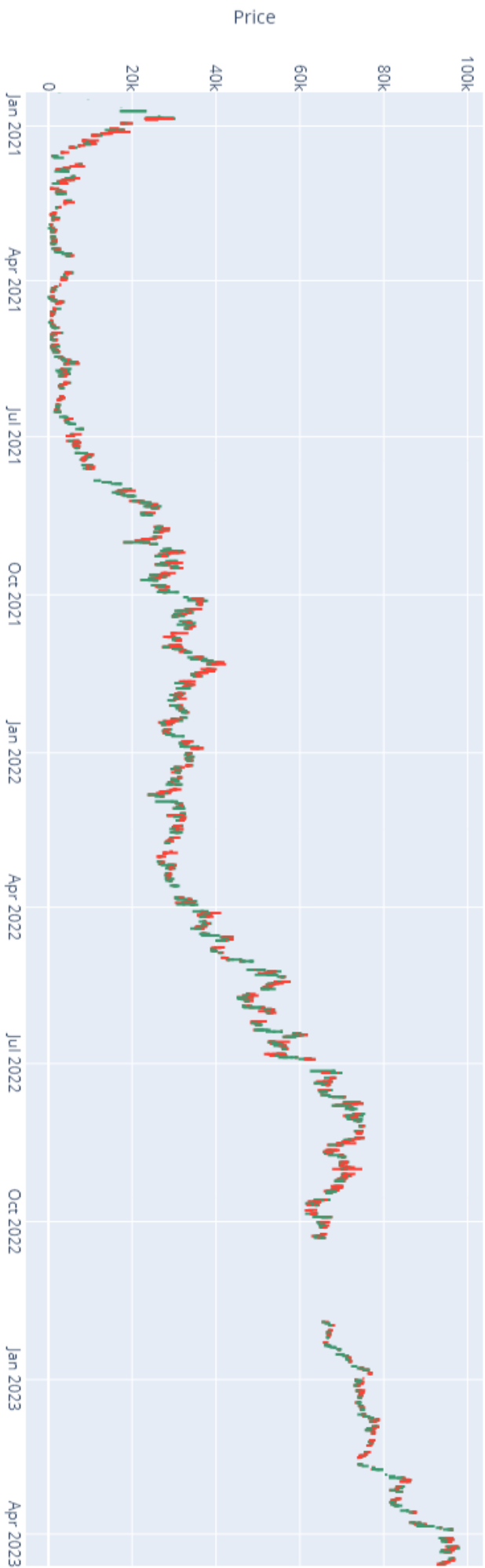
شکل ۲- نمودار کندل‌های قیمتی سهم بوعلی در تایم‌فریم ۳۰ دقیقه‌ای

آماده‌سازی داده

از جمله پیش‌پردازش‌های لازم برای دادگان استخراج شده می‌توان به افت قیمت پس از مجمع اشاره کرد که در آن پس از رفتن یک سهم به مجمع و بسته بودن آن برای مدتی، قیمت سهم افت کرده و به منظور حفظ ارزش سهام، تعداد سهم افراد متناسب با آن افزایش می‌یابد. در این پروژه به منظور رفع تفاوت قیمت سهم پس از مجمع، فاصله‌ی قیمتی ایجاد شده با مبنا قرار دادن قیمت فعلی، پوشش داده شده‌است تا روند سهم بصورت پیوسته و بدون فاصله‌ی قیمتی حفظ شود. نمونه نمودار کندل‌های قیمتی سهم شکل ۲ در شکل ۳ آمده‌است.

از طرفی بدلیل اینکه قیمت سهم‌های مختلف دارای دامنه‌ی متفاوتی بود، به جای قرار دادن قیمت برای ویژگی‌های open، close، high و low، میزان درصد رشد یا نزول هر یک نسبت به کندل قبلی محاسبه شده‌است.

Candlestick Chart

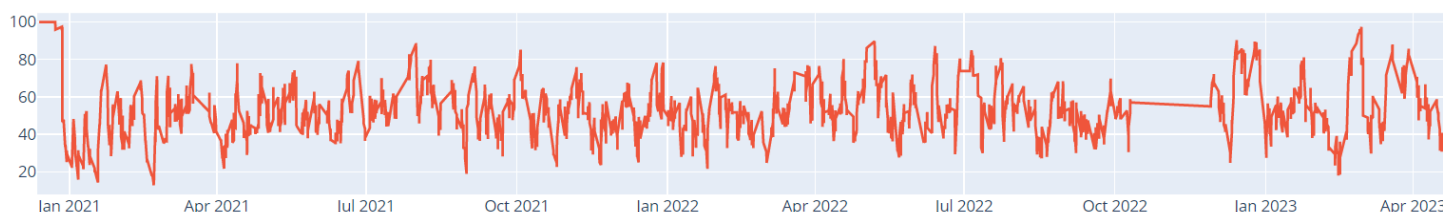


شکل ۳- نمودار کندل‌های قیمتی سهم بوعلی در تایم‌فریم ۳۰ دقیقه‌ای پس از حذف مجموع

از مقادیر اندیکاتورهای مختلفی نظیر RSI، CMF و ... به منظور ویژگی‌های مناسب جهت تسک موردنظر استفاده شده‌است که توضیحات هر کدام در ذیل آمده است.

• اندیکاتور RSI

ازجمله پرکاربردترین اندیکاتورهای بازگشتی در تحلیل تکنیکال، اندیکاتور RSI یا شاخص قدرت نسبی می‌باشد. اندیکاتور RSI را می‌توان از بهترین اندیکاتورهای تشخیص روند معرفی کرد. این اندیکاتور نشان دهنده قدرت بازار و قدرت فروشندگان و خریداران نسبت به یکدیگر می‌باشد. به همین دلیل به آن شاخص قدرت نسبی گفته می‌شود. اندیکاتور RSI جزء اسیلاتورها محسوب می‌شود، زیرا نوسان آن بین ۰ و ۱۰۰ بوده و در زیر چارت قرار دارد. به دلیل این که این اندیکاتور، شامل میانگین قیمت‌ها در بازه‌های زمانی مختلف است، یک اسیلاتور قیمتی محسوب می‌شود. نمونه نمودار اندیکاتور RSI در شکل ۴ آمده‌است.

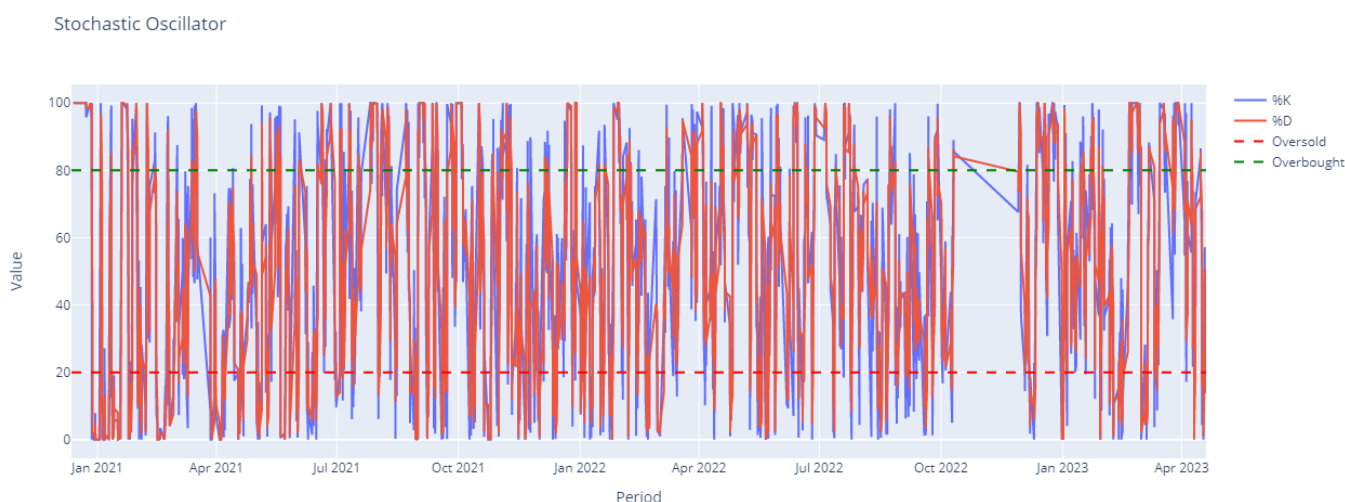


شکل ۴- نمودار اندیکاتور RSI سهم بوعلی در تایم فریم ۳۰ دقیقه‌ای

• Stochastic Oscillator

این اندیکاتور یک شاخص برای momentum است که قیمت بسته شدن یک سهم را با طیفی از قیمت‌های آن در طول یک بازه زمانی معین مقایسه می‌کند. حساسیت این نوسانگر به حرکات بازار را می‌توان با تنظیم آن با بازه زمانی یا با گرفتن یک میانگین متحرک از نتیجه کاهش داد. از این نوسانگر با استفاده از مقادیر محدود صفر تا صد برای صدور سیگنال‌های معاملاتی فروش بیش از حد (oversold) و خرید بیش از حد (overbought) استفاده می‌شود.

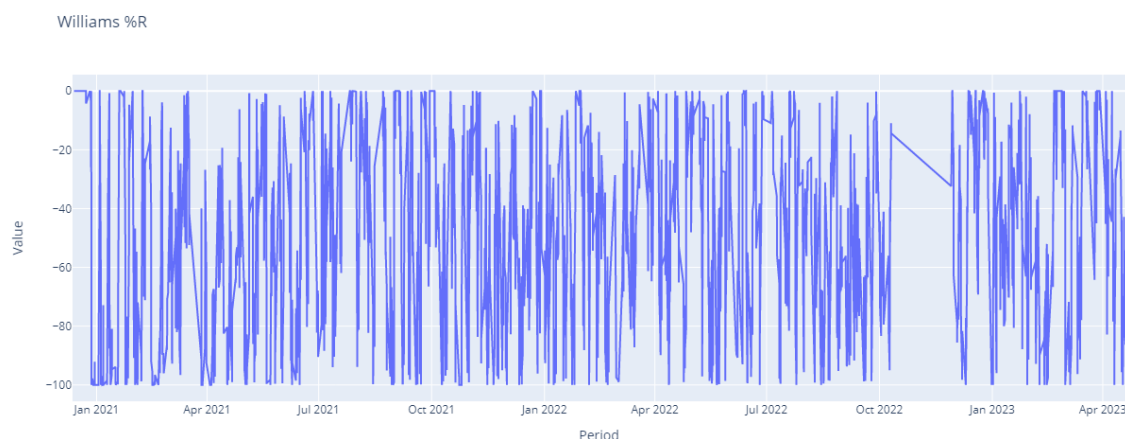
نمونه نمودار اندیکاتور Stochastic Oscillator در شکل ۵ آمده است.



شکل ۵- نمودار اندیکاتور Stochastic Oscillator سهم بوعلی در تایم فریم ۳۰ دقیقه‌ای

• اندیکاتور Williams %R

اندیکاتور Williams % R که همچنین به عنوان محدوده درصدی ویلیامز شناخته می‌شود، نوعی اندیکاتور است که بین ۰ تا -۱۰۰ حرکت می‌کند و میزان خرید بیش از حد و فروش بیش از حد را اندازه‌گیری می‌کند. اندیکاتور ویلیامز می‌تواند برای یافتن نقاط ورود و خروج در بازار استفاده شود. این اندیکاتور قیمت بسته شدن سهم یا ارز مورد نظر را با محدوده بالا و پایین در یک دوره خاص، به طور معمول ۱۴ روز یا دوره مقایسه می‌کند. نمونه نمودار اندیکاتور Williams %R در شکل ۶ آمده است.



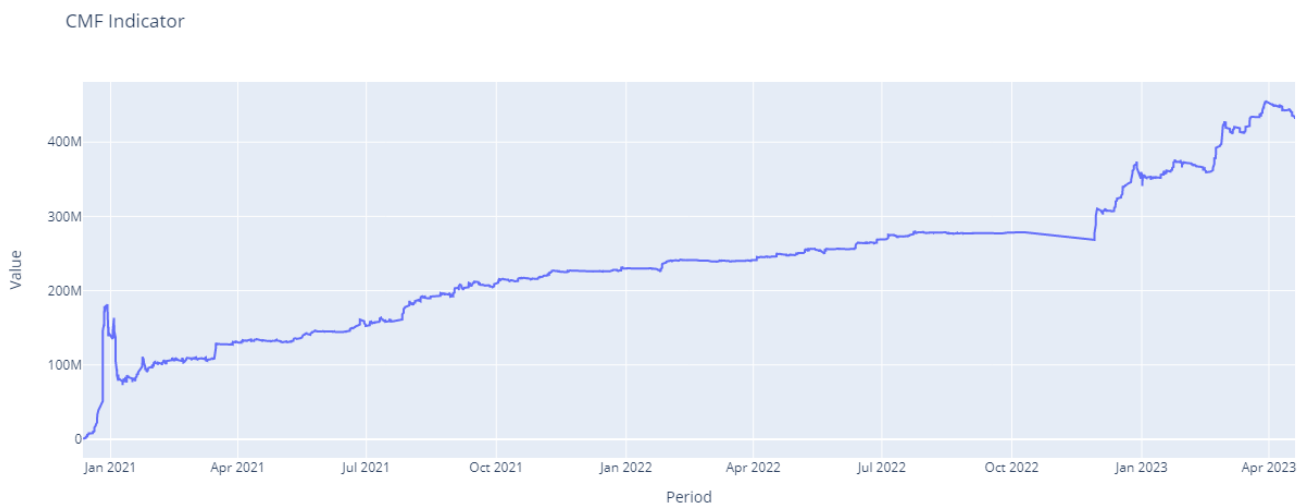
شکل ۶- نمودار اندیکاتور Williams %R سهم بوعلی در تایم فریم ۳۰ دقیقه‌ای

- اندیکاتور Accumulation/Distribution

اندیکاتور Accumulation/Distribution (تراکم/توزیع) یک اندیکاتور تحلیل تکنیکال بر مبنای حجم است که با مقایسه قیمت های بسته شدن همراه با صعودها و نزولها و سنجیدن ارتباط بوسیله حجمهای معامله، تراکمهای ورود و خروج پول برای یک سرمایه را بازتاب می دهد.

- اندیکاتور CMF

اندیکاتور CMF که یکی از اندیکاتورهای مهم تحلیل تکنیکال است، با نام اختصاری Chaikin Money Flow شناخته می شود و به معنی، اندازه گیری میزان حجم جریان پول در یک دوره ی زمانی مشخص است. نمونه نمودار اندیکاتور CMF در شکل ۷ آمده است.



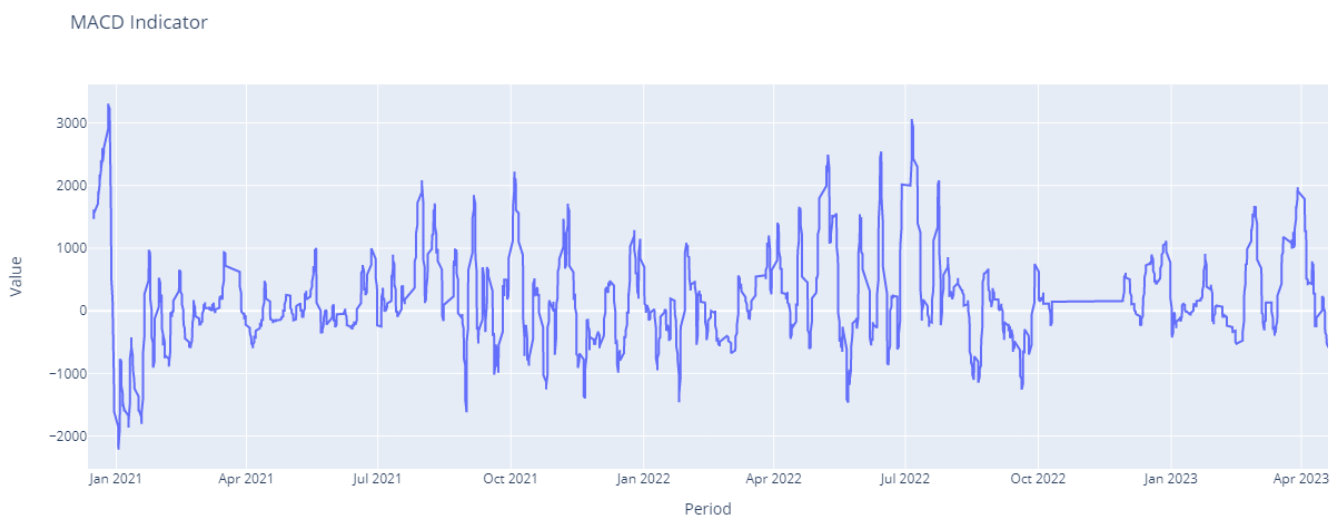
شکل ۷- نمودار اندیکاتور CMF سهم بوعلی در تایم فریم ۳۰ دقیقه ای

• اندیکاتور OBV

اندیکاتور حجم تعادلی (OBV) که مخفف کلمه انگلیسی On-Balance Volume می‌باشد، یک اندیکاتور تکنیکال از نوع momentum است که جریان حجم در سهام مختلف را می‌سنجد و از آن برای پیش‌بینی تغییرات قیمتی بعدی استفاده می‌کند. کار آن اینست که تغییرات حجم در سهام را بر اساس حرکت قیمتی سهم بسنجد تا بتواند برای حرکت بعدی، پیش‌بینی کند.

• اندیکاتور MACD

اندیکاتور MACD مختصر شده عبارت Moving Average Convergence Divergence است. این عبارت به معنی همگرایی (Convergence) و واگرایی (Divergence) میانگین متحرک است و میانگین متحرک نقش اصلی را در ساختار این اندیکاتور ایفا می‌کند. این اندیکاتور با استفاده از ۳ میانگین متحرک ۹، ۱۲ و ۲۶ روزه کمک می‌کند تا قدرت، جهت و شتاب یک روند صعودی یا نزولی را تشخیص داده شود. نمونه نمودار اندیکاتور MACD در شکل ۸ آمده‌است.



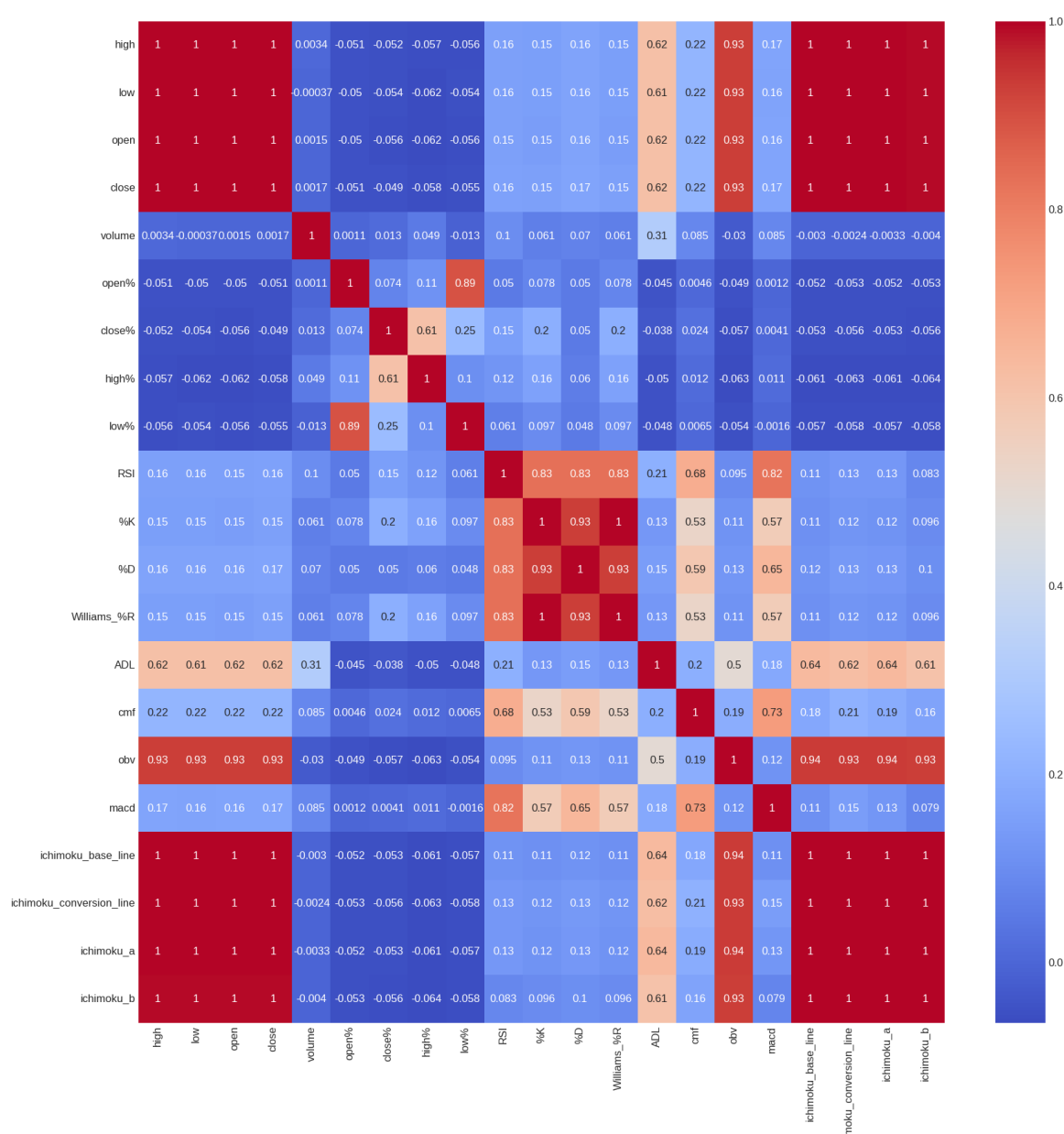
شکل ۸- نمودار اندیکاتور MACD سهم بوعلی در تایم‌فریم ۳۰ دقیقه‌ای

از طرفی در محاسبه‌ی هریک از این اندیکاتورها با توجه به طول پنجره موردنظر مقادیر nan برای کندل‌های ابتدایی ایجاد می‌شد که برای آماده‌سازی داده برای مدل، این کندل‌ها حذف شدند.

درنهایت به منظور نشان دادن میزان وابستگی ویژگی‌های استخراجی، ماتریس correlation ویژگی‌ها در شکل ۹ آمده‌است.

کدهای مربوط به این بخش در فایل به آدرس زیر قابل مشاهده است.

<https://github.com/NeginHashemi/StockPrediction/blob/main/preprocessing.ipynb>



شکل ۹- ماتریس correlation ویژگی‌ها