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

دكتر زارع زاده

اعضای گروه

نگین هاشمی

على عبدالهي

عليرضا طاهريان

سروش وفایی تبار

دانشگاه صنعتی شریف بهار ۱۴۰۲

جمع آوری 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 یک کتابخانه پایتون وجود دارد به نام TseTmc.ir که دادههای استخراج ایران را از وبسایت) استخراج شرکت مدیریت فناوری بورس تهران) استخراج کرده و در اختیار قرار می دهد که یکی از آنها ریز معاملات یک نماد است که می توان با استفاده از آن کندلهای زمانی در تایمفریمهای مختلف زمانی را ساخت. و همچنین حجم معاملات در هر کندل زمانی را استخراج کرد.

در جمع آوری اولیه داده ها ریز معاملات ۲۲ سهام بورس ایران از چهار گروه صنعتی را استخراج کرده و سپس با استفاده از آنها کندلهای قیمتی و حجمی در تایم فریمهای مختلف زمانی را ساختیم.

اسامی سهام استخراج شده به صورت زیر میباشد:

بوعلی، دارا یکم، دی، فارس، نوری، پارسان، شاوان، شبهرن، 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

آمادهسازی داده

از جمله پیشپردازشهای لازم برای دادگان استخراج شده می توان 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 می یابد. در این پروژه به منظور رفع تفاوت قیمت سهم پس از مجمع، فاصله ی قیمتی ایجاد شده با مبنا قرار دادن قیمت فعلی، پوشش داده شده است تا روند سهم بصورت پیوسته و بدون فاصله ی قیمتی حفظ شود.

از طرفی بدلیل اینکه قیمت سهم های مختلف دارای دامنه ی متفاوتی بود، به جای قرار دادن قیمت برای ویژگیهای high ،close ،open و high میزلن درصد رشد یا نزول هر یک نسبت به کندل قبلی محاسبه شدهاست.

از مقادیر اندیکاتورهای مختلفی نظیر CMF ،RSI 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 IRSI 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 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.

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 (overbought) استفاده میشود.

• اندیکاتور Willams %R

اندیکاتور R % Williams که همچنین به عنوان محدوده درصدی ویلیامز شناخته می شود، نوعی اندیکاتور است که بین ۰ تا ۱۰۰- حرکت می کند و میزان خرید بیش از حد و فروش بیش از حد را اندازه گیری می کند. اندیکاتور ویلیامز می تواند 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 می کند.

• 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 میدهد.

• اندیکاتور CMF

اندیکاتور CMF که یکی از اندیکاتورهای مهم تحلیل تکنیکال است، با نام اختصاری Chaikin Money Flow شناخته میشود و به معنی، اندازه گیری میزان حجم جریان پول در یک دورهی زمانی مشخص است.

• اندیکاتور OBV

اندیکاتور حجم تعادلی (OBV) که مخفف کلمه انگلیسی On-Balance اندیکاتور حجم تعادلی (OBV) که جریان Wolume است که جریان حجم در سهام مختلف را می سنجد و از آن برای پیشبینی تغییرات قیمتی بعدی استفاده می کند. کار آن اینست که تغییرات حجم در سهام را بر اساس حرکت قیمتی سهم بسنجد تا بتواند برای حرکت بعدی، پیشبینی کند.

• اندیکاتور MACD

Moving Average Convergence مختصر شده عبارت MACD اندیکاتور (Convergence) و واگرایی Divergence است. این عبارت به معنی همگرایی (Divergence) میانگین متحرک است و میانگین متحرک نقش اصلی را در ساختار این اندیکاتور ایفا می کند. این اندیکاتور با استفاده از Υ میانگین متحرک Υ و Υ روزه کمک می کند تا قدرت، جهت و شتاب یک روند صعودی یا نزولی را تشخیص داده شود.

از طرفی در محاسبه ی هریک از این اندیکاتورها با توجه به طول پنجره موردنظر مقادیر nan برای کندلهای ابتدایی ایجاد می شد که برای آماده سازی داده برای مدل، این کندلها حذف شدند.

درنهایت به منظور نشان دادن میزان وابستگی ویژگیهای استخراجی، ماتریس correlation

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

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

