Shahid Beheshti University
Journal of Financial Management Perspective
No.36, Winter 2022
PP: 57-89

P-ISSN: 2645-4637
E-ISSN: 2645-4645

Investigating the impact of sentiments on stock returns: evidence from reactions to social media content

Reza Majidi Zavieh*, Ehsan Hajizadeh **

Research Paper

Abstract

Due to the increasing growth of social networks in recent years, investors in various markets, in addition to reviewing and analyzing classic market information, also pay attention to news and information published on social networks. By examining and evaluating the relationship between news and information published on social networks and changes in stock prices, it is possible to understand the impact of the information on stock prices and predict the future trend. In this article, using the methods of sentiment analysis and text mining, the impact of public thoughts and feelings caused by news on the Internet and cyberspace on stock prices is examined. The information used in this research includes content published on the social network Twitter about stocksand real stock price data of the top 5 companies on the US stock exchange. Using the presented method, general feelings about a text are estimated and a general score is considered for it. Then, using back testing methods and adopting different trading strategies, the impact of these emotions on the share price trend will be examined and the obtained results will be compared with and without the effect of emotion analysis. According to the results of this study, the effectiveness of strategies based on sentiments analysis is significantly higher than technical analysis-based methods.

Keywords: Stock Returns; Sentiment analysis; text mining; Trading Strategy; backtesting.

Received: 2021. December.14, Accepted: 2022.March.26.

*BSc in industrial engineering, Amirkabir University of Technology, Tehran, Iran.
**Assistant Prof, Department of Management Systems, Amirkabir University of Technology, Tehran, Iran (Corresponding Author). E-Mail: ehsanhajizadeh@aut.ac.ir

دانشگاه شهیدبهشتی
نشریه چشم‌انداز مدیریت مالی
شماره۳۶ ـ زمستان ۱۴۰۰
ص‌ص ۸۹-۵۷

شاپای چاپی: ۴۶۳۷- ۲۶۴۵
شاپای الکترونیکی: ۴۶۴۵- ۲۶۴۵

بررسی تاثیر احساسات بر بازدهی سهام: شواهدی از واکنش به مطالب منتشره در فضای مجازی

رضا مجیدی زاویه*، احسان حاجی زاده**

مقاله پژوهشی

چکیده

با توجه به رشد روزافزون شبکه‌های اجتماعی در سال‌های اخیر، سرمایه‌گذاران در بازارهای مختلف علاوه بر بررسی و تحلیل اطلاعات کلاسیک بازار، به اخبار و اطلاعات منتشره در شبکه‌های اجتماعی نیز توجه می‌کنند. با بررسی و ارزیابی میزان ارتباط اخبار و اطلاعات منتشره در شبکه‌های اجتماعی و تغییرات قیمت سهام می‌توان به میزان تاثیرگذاری این اطلاعات بر قیمت سهام پی برد و از طریق آن روند آینده را پیش بینی نمود.

در این مقاله با استفاده از روش تحلیل احساسات و متن‌کاوی، به بررسی میزان اثرگذاری افکار و احساسات عمومی ناشی از اخبار در اینترنت و فضای مجازی بر قیمت سهام پرداخته می‌شود. اطلاعات استفاده‌شده در این پروژه شامل مطالب منتشرشده در شبکه اجتماعی توییتر پیرامون سهام و همچنین داده‌های واقعی قیمتی سهام ۵ شرکت برتر بورس آمریکا است. با استفاده از روش ارائه‌شده، احساسات کلی در مورد یک متن برآورد شده و امتیازی کلی برای آن در نظر گرفته می‌شود. سپس با استفاده از روش‌های بک‌تستینگ و با اتخاذ استراتژی‌های مختلف معاملاتی، تاثیر این احساسات در روند قیمتی سهم بررسی خواهد شد و بازدهی به‌دست آمده، با و بدون اثردهی آنالیز احساسات، مقایسه خواهد شد. طبق نتایج این پژوهش، بازدهی استراتژی‌های مبتنی بر آنالیز احساسات به طور قابل توجهی بیشتر از روش‌های تحلیل تکنیکال هستند.

کلیدواژه‌ها: بازدهی سهام؛ تحلیل احساسات؛ متن کاوی؛ استراتژی معاملاتی؛ بک تست.

تاریخ دریافت مقاله: ۱۴۰۰/۰۹/۲۳، تاریخ پذیرش مقاله: ۱۴۰۱/۰۱/۰۶.

* کارشناسی مهندسی صنایع، دانشگاه صنعتی امیرکبیر، تهران، ایران.
** استادیار، گروه مدیریت سیستم و بهره‌وری، دانشگاه صنعتی امیرکبیر، تهران، ایران(نویسنده مسئول).
E-Mail: ehsanhajizadeh@aut.ac.ir

۱. مقدمه

بازار سهام محلی برای مبادله و انتقال مالکیت سهام شرکت‌های پذیرفته‌شده در این بازار بوده و می‌تواند نقش بسیار زیادی در توسعه اقتصادی کشورها ایفا نماید. با توجه به اینکه سرمایه‌گذاران به دنبال بیشینه‌سازی سود خود در این بازار هستند، همواره به دنبال روش‌های نوین و دقیق برای پیش‌بینی روند بازار هستند. با رشد اینترنت و استفاده گسترده از شبکه‌های اجتماعی، مطالب منتشره در این فضا بر تصمیم سرمایه‌گذاران تاثیر می‌گذارد[۲۳]. لذا با تحلیل و بررسی مطالب منتشرشده در شبکه‌های اجتماعی می‌توان با تحلیل احساسات[1] به برآوردی از تغییرات روند قیمت یک سهم دست یافت.

در بازارهای مالی، واکنش به اطلاعات به معنی تغییر ناگهانی، اما معمولا کوتاه‌مدت قیمت‌های سهام بعد از انتشار اخبار و اطلاعاتی پیرامون آن سهام یا شرکت، به‌طور مستقیم یا غیرمستقیم، است. یک واکنش ممکن است تنها مدتی پس از ایجاد آن باقی بماند و سپس توسط حرکت‌های اصلاحی خنثی شود.

در این مقاله سعی بر آن است رابطه میزان احساسات کلی منتشرشده در شبکه‌های اجتماعی و تاثیر آن بر تغییرات قیمت سهام، با تکیه بر روش‌های تحلیل احساسات و به کارگیری استراتژی‌های مختلف معاملاتی مورد بررسی قرار گیرد.

شناخت رابطه میان احساسات کلی پیرامون شرکت‌ها و سهام آن‌ها و تغییرات قیمتی آن‌ها می‌تواند برای دو دسته افراد مناسب باشد. نخست، تحلیل‌گران و معامله‌گران سهام که به دنبال استفاده از روش‌های مختلف تحلیلی برای بیشینه‌سازی سود خود هستند. دسته دوم مدیران ارشد و تصمیم‌گیران شرکت‌هایی هستند که می‌خواهند با درنظرگرفتن میزان اثرگذاری اخبار منتشره پیرامون شرکت، بتوانند سیستم‌های بهینه اطلاع‌رسانی و تبلیغاتی را ایجاد و ساماندهی کنند.

در واقع می‌توان سوالاتی که این مقاله به دنبال پاسخ آن است را به شرح زیر بیان کرد:

آیا اخبار و مطالب مرتبط با یک شرکت در فضای مجازی تاثیر قابل توجهی بر تعیین روند آتی سهام دارد یا قیمت‌ها تابعی از رفتارهای تصادفی بوده و روند تصادفی را دنبال می کنند؟ آیا با به‌کارگیری روش‌های تحلیل احساسات، می‌توان بازدهی آینده سهام را پیش‌بینی کرد؟ آیا ترکیب روش آنالیز احساسات و استراتژی‌های تحلیل تکنیکال می‌تواند کمکی به بهترشدن نتیجه پیش‌بینی روند آتی سهام بکند؟ در این پژوهش، ابتدا به مبانی نظری و پیشینه پژوهش و روش‌شناسی پژوهش پرداخته می‌شود. سپس، در بخش روش شناسی پژوهش، روش‌های ارزیابی احساسات توییت‌ها و امتیازدهی به آن‌ها بیان خواهد شد. در بخش تحلیل داده‌ها و یافته‌ها، روش‌های پاک‌سازی و آماده‌سازی داده‌های مورد نیاز تحقیق مورد بررسی قرار خواهند گرفت. در بخش بحث و نتیجه گیری، به عنوان نتایج تحقیق ارتباط میان حجم توییت‌ها و حجم معاملات

[1] Sentiment Analysis

سهام، ارتباط میان تعداد توییت‌های مثبت و قیمت سهام و ارتباط میان تعداد توییت‌های منفی و قیمت سهام ارزیابی خواهند شد. در بخش آخر نیز، با استفاده از روش‌های بک‌تستینگ، به بررسی بازدهی سهام با استفاده از روش‌های تحلیل احساسات پرداخته خواهد شد و مقایسه‌ای از بازدهی بدون و با به‌کارگیری این روش‌ها ارائه خواهد شد.

۲. مبانی نظری و پیشینه پژوهش

مسئله پیش‌بینی روند قیمتی سهام همواره موضوعی بحث‌برانگیز و قابل توجه برای تحلیل‌گران، سرمایه‌گذاران، معامله‌گران و مدیران شرکت‌ها بوده است. طبق نظریه گشت تصادفی، تغییرات قیمت سهام و بازدهی‌های متاثر از این تغییرات، همواره تصادفی و غیر قابل پیش‌بینی هستند. این نظریه از ابتدایی‌ترین نظریات در مورد بازار سهام می‌باشد و کارایی بازار را در تغییر تصادفی قیمت‌ها می‌داند. امروزه و طبق نظریه‌های اقتصاد رفتاری، قیمت سهام ممکن است تحت تاثیر اخبار و اطلاعات سایت‌ها و شبکه‌های اجتماعی دستخوش تغییراتی شود.

در سال‌های اخیر، با توجه به عدم کارایی بازارها و قابل پیش‌بینی بودن آن‌ها، روش‌های مختلفی برای پیش‌بینی این بازارها ارائه شده است. در این میان عمده روش‌های ارائه شده، متمرکز بر استفاده از روش‌های هوش مصنوعی بوده است. در ادامه برخی پژوهش این حوزه مورد بررسی قرار می گیرد.

یافتیان و رستگار (۱۳۹۹) یک سیستم معاملاتی خودکار با استفاده از پردازش تصویر به وسیله‌ی شبکه عصبی پیچشی ارائه کردند. در مدل ارائه شده، در ابتدا پس از دریافت داده‌های مورد نیاز برای سهام منتخب، ۲۸ اندیکاتور تحلیل تکنیکال انتخاب و مقادیر هرکدام به صورت جداگانه برای هر سهم محاسبه شد. سپس سری‌های زمانی این اندیکاتورها به تصاویر ۲ بعدی تبدیل شده و در نتیجه برای هر داده روی سری زمانی قیمت سهم، یک تصویر دو بعدی با ابعاد ۲۸×۲۸ ساخته شد. نتایج آن‌ها نشان می‌دهد که در ۸۰٪ موارد، این روش بازدهی بیشتری نسبت به استراتژی مرسوم خرید و نگهداری کسب کرده است [۲۷]. محبی و همکاران، نیز الگوریتمی را برای انتخاب ویژگی‌های مناسب به منظور پیش‌بینی شاخص بورس اوراق بهادار تهران ارائه کردند. آن‌ها، برای پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران بر اساس متغیرهای مالی و اقتصادی، ابتدا اقدام به اولویت‌بندی ویژگی‌ها با الگوریتم MID[1] نموده، سپس از ۴ مدل مختلف شبکه عصبی (MLP[2], SVR[3], RBF[4], DNN[5]) که از مهم‌ترین و بدیع‌ترین مدل‌های پیش‌بینی است استفاده کردند. نتایج بدست‌آمده نشان می‌دهد که با روش

[1] Mid-Point Algorithm
[2] Multi-Layer Perceptron
[3] Support-Vector Regression
[4] Radial Basis Function
[5] Deep Neural Network

پیشنهادی، می‌توان با ۷ ویژگی انتخابی به دقت بالایی در پیش‌بینی روزانه شاخص کل بورس اوراق بهادار تهران دست یافت [۲۰].

رویکرد دیگری که در سال‌های اخیر مورد توجه فعالان بازار به منظور پیش‌بینی روند آتی سهام قرار گرفته است، بهره‌گیری از تحلیل احساسات در بررسی اخبار و اطلاعات منتشرشده در فضای مجازی است. لذا بدین منظور و در مورد موضوع بررسی رابطه میان داده‌های اینترنت و بازارهای مالی، سه نوع داده مختلف مورد بررسی قرار گرفته‌اند: اخبار سایت‌ها، کوئری‌های موتورهای جستجو و شبکه‌های اجتماعی. در مورد اخبار، مطالعات مختلفی از قبیل بررسی ارتباط اخبار با تغییرات قیمت [۸]، واکنش قیمت سهام به اخبار [۷، ۲۶]، رابطه بین اشاره به نام یک شرکت در اخبار مالی و قیمت آن [۱]، خوش‌بینی رسانه‌ها [۲۴] و ارتباط بین احساسات اخبار و پیش‌بینی‌پذیری بازده سهام [۲۵]، نقش اخبار در استراتژی‌های معاملاتی [۱۷]، نقش اخبار اقتصاد کلان در بازدهی سهام[۲] و واکنش‌های پرتکرار بازار به اخبار [۱۲] صورت گرفته‌اند.

جینگ[1] و همکاران، مدلی ترکیبی مبتنی بر روش‌های یادگیری ماشین و به صورت مشخص شبکه‌های عصبی پیچشی و روش‌های مبتنی بر تحلیل احساسات در بازار بورس شانگهای چین ارائه کردند. نتایج این پژوهش نشان می دهد نتایج مدل ترکیبی (درنظر گرفتن همزمان روش‌های مبتنی بر تحلیل احساسات و یادگیری ماشین) بهتر از هر یک از روش‌ها به تنهایی عمل می کند [۱۵].

در پژوهش دی آلوریا کاروسیا و همکاران(۲۰۲۱)[2] به صورت مشخص روی طراحی استراتژی‌های معاملاتی مبتنی بر تحلیل احساسات متمرکز شده است. استراتژی‌های ارائه شده روی بازارهای بورس برزیل و پرتغال بررسی شده است. نتایج این تحقیق نشان می دهد مدل ارائه شده در کوتاه مدت و بلند مدت از استراتژی‌های متداول مانند حرکت تصادفی و خرید و نگهداری بهتر عمل می کند[۱۰].

در مورد کوئری‌های[3] موتورهای جستجو نیز پژوهش‌ها زیادی وجود دارد. از جمله این مقالات می‌توان به ارتباط بین تعداد جستجوی‌های روزانه یک سهم خاص و حجم معاملات آن اشاره کرد [۵، ۶]. تحلیل مشابهی نیز در مورد شاخص ۳۰۰۰ سهم راسل انجام شده که نشان می‌دهد افزایش در تعداد جستجوها منجر به افزایش قیمت‌ها در دو هفته آتی می‌شود [۹]. استفاده از ترندهای موتور جستجوی گوگل برای ارزیابی ریسک سهام از دیگر مقالات در این زمینه می‌باشد[۱۴، ۱۵].

در بحث شبکه‌های اجتماعی، توییتر به عنوان یکی از محبوب‌ترین پلتفرم‌های میکروبلاگینک برای پیش‌بینی مالی مورد استفاده قرار گرفته است [۱۱، ۲۱]. تعدادی از پژوهش‌ها، رابطه بین

---

[1] Jing
[2] De Oliveira Carosia et al
[3] Queries

حجم توییت‌ها و بازارهای مالی را مورد بررسی قرار داده‌اند، به عنوان مثال مائو[1] وهمکاران(۲۰۱۲)در پژوهش خود به این سوال پاسخ داده‌اند که آیا تعداد توییت‌های روزانه می‌تواند اندیکاتورهای سهام شاخص S&P 500 را پیش‌بینی نماید؟ دسته دوم مقالات محتوای توییت‌ها را مورد بررسی قرار داده است. به عنوان نمونه مقالات [۳، ۴ و ۱۸] رابطه بین احساسات حاکم بر توییت‌ها و میانگین صنعتی داوجونز (DJIA) را بررسی کرده اند[۱۹]. سوزا[2] و همکاران(۲۰۱۵) نشان داده‌اند که احساسات توییتری ۵ شرکت خرده‌فروش از نظر آماری ارتباط قابل توجهی با بازدهی و نوسان‌پذیری سهام داشته است[۲۳].

در نتیجه، در بیشتر پژوهش‌ها رابطه‌ی معناداری بین احساسات مثبت شبکه‌های اجتماعی و افزایش اقبال سرمایه‌گذاران و به تبع آن افزایش مقدار تقاضا و قیمت سهام وجود دارد. اگرچه عوامل موثر بر تغییرات قیمت سهام بسیار زیاد می‌باشند و تنها با تکیه بر احساسات نمی‌توان استراتژی‌های معاملاتی بدون ریسک و یا با ریسک کم ابداع کرد، اما این احساسات می‌تواند دید مناسبی به سرمایه‌گذار بدهد تا با ترکیب آن با سایر استراتژی‌ها بتواند معاملاتی با بازده مناسب داشته باشد. البته بیشتر این پژوهش‌ها تاکید داشته‌اند که در بازارهای بسیار کارا، معمولا قیمت‌ها از گشت تصادفی پیروی می‌کنند و نمی‌توان به بازدهی‌های غیرمعمول دست یافت. به همین دلیل، معمولا این استراتژی‌ها در کوتاه‌مدت موثرتر هستند. با توجه به نتایج گوناگونی که در پژوهش‌های مختلف در این باره گرفته شده است، در این پژوهش به ارزیابی این مسئله پرداخته خواهد شد که آیا فضای مجازی تاثیر قابل توجهی بر تعیین روند آتی سهام دارد یا قیمت‌ها تابعی از رفتارهای تصادفی بوده و روند تصادفی را دنبال می کنند؟ و اینکه آیا امکان پیش‌بینی روند آتی سهام با تکیه بر تحلیل احساسات منتشره در فضای وجود دارد یا خیر؟ آیا استراتژی‌های مبتنی بر تحلیل احساسات عملکرد بهتری نسبت به استراتژی‌های مبتنی بر تحلیل تکنیکال دارند؟ در واقع به طور مشخص نوآوری این مقاله را می توان به این صورت بیان نمود که ضمن پیاده سازی استراتژی‌های معاملاتی مبتنی بر تحلیل احساسات و اخبار فضای مجازی و مقایسه سودآوری آن با استراتژی‌های مبتنی بر تحلیل تکنیکال، روش جدیدی جهت تفسیر اخبار فضای مجازی و تاثیر آن بر روند قیمتی سهام چند شرکت معتبر ارائه شده است.

۳. روش‌شناسی پژوهش

تجزیه و تحلیل احساسات که به نظرکاوی[3] یا هوش مصنوعی احساسات[4] نیز معروف است، به معنی استفاده از تکنیک‌های پردازش زبان‌های طبیعی[5] (NLP) ، متن‌کاوی[6] و زبان‌شناسی

[1] Mao et al
[2] Souza et al
[3] Opinion mining
[4] Emotion AI
[5] Natural Language Processing
[6] Text mining

محاسباتی[1] برای شناسایی، استخراج، کمّی‌سازی و مطالعه حالات عاطفی و اطلاعات ذهنی به شکلی سیستماتیک می‌باشد.

امروزه آنالیز احساسات کاربردهای فراوانی به خصوص در زمینه ارتباط با مشتریان و پیش‌بینی حرکت‌های آتی بازارها دارد. به کمک این ابزار و با استفاده از تکنیک‌های NLP می‌توان احساسات کلی حاکم بر متون را استخراج کرد و در متعارف‌ترین تعریف، متون را به سه دسته مثبت، منفی و خنثی از نظر احساسات تقسیم کرد.

تجزیه و تحلیل احساسات، فرایند تشخیص جملات مثبت یا منفی در متون است. این تکنیک غالبا توسط کسب و کارها برای تشخیص احساسات در داده‌های اجتماعی، اندازه‌گیری میزان شهرت برند و فهم نیازهای مشتریان استفاده می‌شود. به دلیل اینکه امروزه مشتریان، افکار و احساسات خود را بیش از پیش ابراز می‌کنند، تجزیه و تحلیل احساسات در حال تبدیل به ابزاری ضروری و اساسی برای ارزیابی و فهم احساسات می‌باشد.

در این مقاله، برای ارزیابی احساسات توییت‌ها، از کتابخانه vaderSentiment که یک کتابخانه شخص ثالث[2] بر پایه زبان برنامه‌نویسی پایتون می‌باشد، استفاده شده است[۱۳]. VADER[3] یک ابزار تجزیه و تحلیل احساسات از نوع واژه‌نامه‌ای[4] و برپایه الگوریتم‌ها می‌باشد که به طور خاص برای ارزیابی احساسات بیان شده در رسانه‌های اجتماعی ساخته شده است.

این کتابخانه، با دریافت یک متن، نسبت به تجزیه و تحلیل آن از نظر احساسات کلی حاکم بر آن اقدام می‌کند و خروجی آن به صورت یک دیکشنری پایتون[5] که دارای چهار مشخصه به صورت زیر است، می‌باشد:

{pos': 0.746, 'compound': 0.8316, 'neu': 0.254, 'neg': 0.0'}

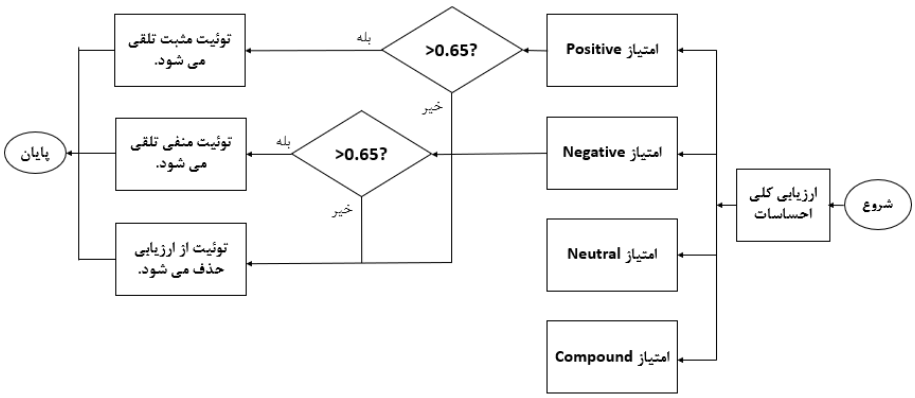
اعداد روبروی هر مشخصه، مربوط به یک نمونه تجزیه و تحلیل هستند و در هر بار تست، تغییر می‌کنند. در ادامه، به توضیح هر کدام از این مشخصه‌ها پرداخته می‌شود؛ که شامل موارد زیر هستند:

- pos: امتیاز مثبت برآورد شده برای متن، امتیازی است بین صفر و یک که امتیاز یک به معنی کاملا مثبت و صفر به معنی عدم برخورداری از هیچ مشخصه مثبتی می‌باشد.
- neg: امتیاز منفی برآورد شده برای متن، امتیازی است بین صفر و یک که امتیاز یک به معنی کاملا منفی و صفر به معنی عدم برخورداری از هیچ مشخصه منفی‌ای می‌باشد.

[1] Computational linguistics
[2] Third party library
[3] Valence Aware Dictionary and Sentiment Reasoner
[4] Lexicon
[5] Python dictionary

- neu: امتیاز خنثی برآورد شده برای متن، امتیازی است بین صفر و یک که امتیاز یک به معنی کاملا خنثی (نه منفی و نه مثبت) و صفر به معنی عدم برخورداری از هیچ مشخصه خنثی‌ای می‌باشد.
- compound: امتیاز کل، حاصل‌جمع امتیازهای هر لغت در متن، که توسط قوانین کتابخانه تصحیح شده و سپس به صورت عددی بین منفی ۱ (کاملا منفی) و مثبت ۱ (کاملا مثبت) نرمال شده است.

طبق اعلام سایت vader، امتیاز ترکیبی[1] امتیازی است که بیشتر از همه مورد استفاده تحلیل‌گران و سایر کاربران قرار می‌گیرد و به صورت کلی، برآوردی مناسب از احساسات متون ارائه می‌کند. شایان ذکر است که جمع امتیازهای pos، neg و neu برابر یک می‌باشد. شکل (۱) روند کلی امتیازدهی به هر توئیت را نشان می دهد.



شکل ۱. روند کلی امتیازدهی به توییت‌ها.

تغییرات اعمالی بر روی کتابخانه

با توجه به اینکه vader یک کتابخانه عمومی است و این مقاله بر روی بازار سهام تمرکز دارد، لازم است که تغییراتی را در کتابخانه اعمال کرده و تا دقت کار بالا رود. برای اعمال این تغییرات نیاز به مجموعه لغات مثبت و منفی استفاده شده در شبکه‌های اجتماعی و مربوط به بازار سهام داریم. این مجموعه لغات را توسط دستورات زبان پایتون به کتابخانه اضافه شده است تا دقت ارزیابی متون مربوط به بازار سهام افزایش یابد.

مجموعه لغات مثبت اضافه شده به کتابخانه شامل موارد زیر است:

'buy bull long support awesome undervalued underpriced cheap upward rising trend moon rocket hold breakout call beat support buying holding high-profit growth'

و مجموعه لغات منفی اضافه شده نیز شامل موارد زیر است:

[1]Compound

'sell bear bubble bearish short overvalued overbought overpriced expensive downward falling sold sell low put miss resistance squeeze cover seller overrated'

که لازم است این لغات به کتابخانه vader اضافه گردد.

ارزیابی احساسات توییت‌ها

پس از اعمال تغییراتی در کتابخانه vader، نوبت به ارزیابی احساسات توییت‌های واقعی کاربران، در مورد سهام انتخابی بورس آمریکا می‌رسد. همانطور که پیش‌تر گفته شد، داده‌های مورد استفاده در این مقاله، داده‌های واقعی توییت‌های کاربران از سال ۲۰۱۵ تا ۲۰۲۰ در مورد پنج سهم برتر بورس آمریکا، یعنی آمازون[1] ، اپل[2]، گوگل[3] ، مایکروسافت[4] و تسلا[5] می‌باشد. همانطور که بیان شد، این داده‌ها با توجه به معیارهایی پاک‌سازی شدند و سپس آماده استفاده در کتابخانه vader هستند.

داده‌های توییتر شامل ستون‌هایی از جمله نام نماد و متن توییت هستند. نام نماد با شناسه انگلیسی ticker symbol مشخص شده است که برای هر شرکت به شرح زیر است:

آمازون: AMZN، اپل: AAPL، گوگل: GOOGL، مایکروسافت: MSFT، تسلا: TSLA

در ادامه باید برای هر شرکت، توییت‌های مربوطه تفکیک شوند و سپس وارد تابع کتابخانه vader شده تا امتیاز احساسات توییت، استخراج شود. برای تفکیک، از دستور loc در زبان برنامه‌نویسی پایتون استفاده خواهد شد. طبق نتایج تفکیک، تعداد توییت‌های هر شرکت در بازه زمانی ۲۰۱۵ تا ۲۰۲۰ به شرح جدول (۱) است.

جدول۱. تعداد توییت‌های هر شرکت

نام شرکت	نماد	تعداد
آمازون	AMZN	۷۱۸۷۱۵
اپل	AAPL	۱۴۲۵۰۱۳
گوگل	GOOGL	۳۲۷۵۶۹
مایکروسافت	MSFT	۳۷۵۷۱۱
تسلا	TSLA	۱۰۹۶۸۶۸

[1] Amazon
[2] Apple
[3] Google
[4] Microsoft
[5] Tesla

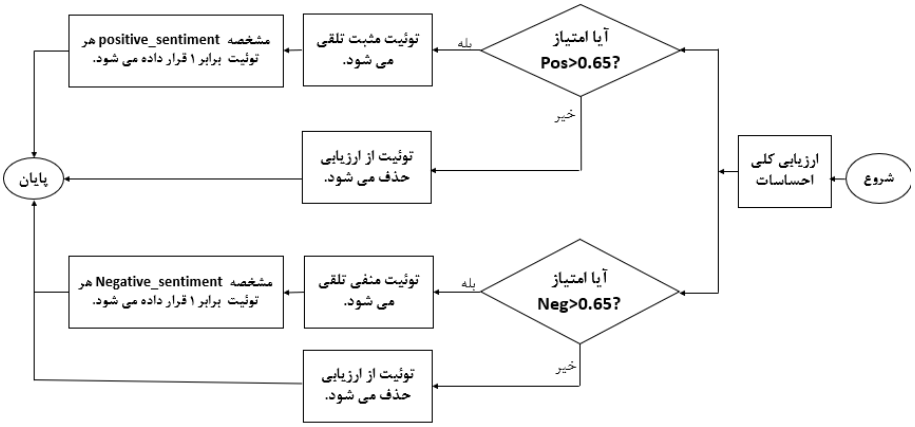
امتیازدهی به توییت‌ها

پس از تفکیک، با کمک کتابخانه vader، به توییت‌ها امتیاز می‌دهیم. شیوه کار انتخابی در این مقاله به این شکل است که دو ستون positive_sentiment و negative_sentiment به توییت‌ها اضافه شده است. اگر توییت دارای احساسات مثبت باشد، دارای positive_sentiment برابر عدد یک و negative_sentiment برابر عدد صفر خواهد بود. برای اینکار، دو تابع positivityCheck و negativityCheck نوشته شده است. تابع positivityCheck، متن توییت‌ها را مورد ارزیابی قرار می‌دهد و با شرایط زیر مقدارها را پر می‌کند:

اگر مقدار امتیاز pos (که یک خروجی کتابخانه vader است) بیشتر از ۰.۶۵ باشد، توییت را دارای احساسات مثبت فرض کرده و مقدار ۱ را برای ستون positive_sentiment و مقدار ۰ را برای negative sentiment برمی‌گرداند.

اگر مقدار امتیاز neg (که یک خروجی کتابخانه vader است) بیشتر از ۰.۶۵ باشد، توییت را دارای احساسات منفی فرض کرده و مقدار ۱ را برای ستون negative_sentiment و مقدار ۰ را برای positive sentiment برمی‌گرداند.

شکل (۲) روش برچسب زنی به هر توئیت را نشان می دهد.



شکل۲. روش برچسب‌زنی توییت‌ها.

تعداد توییت‌های مربوط به هر شرکت، پس از اجرای کد ارزیابی احساسات، مطابق جدول زیر می‌باشد:

جدول ۲. تعداد توییت‌های هر شرکت پس از ارزیابی احساسات.

نام شرکت	نماد	تعداد
آمازون	AMZN	۳۲۸۲
اپل	AAPL	۵۲۸۱
گوگل	GOOGL	۱۳۳۶
مایکروسافت	MSFT	۱۰۲۹
تسلا	TSLA	۵۳۱۰

استخراج تعداد توییت‌های مثبت و منفی و ادغام جدول تعداد توییت‌ها

حال برای استفاده از این داده‌ها در بخش بعد، تعداد توییت‌های منفی و مثبت هر شرکت در تاریخ‌های مختلف استخراج شده است. در پایان، لازم است که دو دیتافریم به دست آمده از اجرای کدهای بالا ادغام شود تا دیتافریمی شامل سه ستون تاریخ، تعداد توییت‌های منفی و تعداد توییت‌های مثبت ایجاد شود. برای ادغام، با توجه به اینکه بعضی روزها دارای توییت‌های مثبت و منفی نمی‌باشند، لازم است برای روزهای خالی و تعداد مربوطه عدد صفر قرار داده شود.

٤. تحلیل داده‌ها و یافته‌ها

با توجه به اینکه روش کلی مورد استفاده در این پژوهش نیازمند برنامه‌نویسی است، برای پیش‌برد اهداف مقاله نیاز به یک زبان برنامه‌نویسی جدید، پویا و دارای جامعه استفاده‌کنندگان پیشرفته می‌باشد. همچنین برای به کارگیری این زبان وجود یک محیط برنامه‌نویسی مناسب نیز لازم است. در این مقاله از زبان پایتون (Python) و محیط برنامه‌نویسی نوت‌بوک جوپیتر (Jupyter notebook) استفاده شده است. همچنین در طول مقاله، از کتابخانه‌هایی نظیر pandas، numpy، vaderSentiment، nltk، datetime، matplotlib، backtesting در مقاله استفاده شده که در ادامه و به فراخور موضوع، در مورد آن‌ها توضیح داده خواهد شد.

داده‌های توییتر

در این مقاله، از داده‌های توییت‌های مربوط به ۵ شرکت بزرگ و مشهور آمریکایی شامل آمازون(Amazon)، اپل(Apple)، گوگل(Google)، مایکروسافت(Microsoft) و تسلا(Tesla) استفاده شده است. این توییت‌ها مربوط به سال‌های ۲۰۱۵ تا ۲۰۲۰ می‌باشند که شامل ۳ میلیون توییت منحصر به فرد راجع به شرکت‌های مذکور با اطلاعاتی نظیر شناسه، نویسنده، تاریخ، تعداد لایک، تعداد کامنت، تعداد ریتوییت و در نهایت متن توییت است. منبع این داده‌ها سایت Kaggle بوده که یکی از بزرگترین پایگاه‌های داده برای انواع داده‌های مورد

استفاده در پژوهش‌ها است. برای جمع‌آوری این داده‌ها از روش‌های استخراج وب[1] و کتابخانه مشهور سلنیوم[2] استفاده شده است. شکل(۳) ساختار بندی کلی توییت‌های جمع آوری شده را نشان می‌دهد.

```
In [4]: tweets.head()
```

Out[4]:

	tweet_id	ticker_symbol	writer	post_date	body	comment_num	retweet_num	like_num
0	550803612197457920	AAPL	SentiQuant	1420156789	#TOPTICKERTWEETS $AAPL $MRS $BABA EBAY $AMZN...	0	0	1
1	550803612197457920	AMZN	SentiQuant	1420156789	#TOPTICKERTWEETS $AAPL $MRS $BABA EBAY $AMZN...	0	0	1
2	550803610825928706	AAPL	SentiQuant	1420156788	#SENTISHIFTUP $KFB $GOOGL $GS $GOLD T $AAPL...	0	0	1
3	550803610825928706	GOOGL	SentiQuant	1420156788	#SENTISHIFTUP $KFB $GOOGL $GS $GOLD T $AAPL...	0	0	1
4	550803610825928706	AMZN	SentiQuant	1420156788	#SENTISHIFTUP $KFB $GOOGL $GS $GOLD T $AAPL...	0	0	1

شکل ۳. نمونه ساختار کلی توییت‌ها.

داده‌های قیمت سهام

برای ارزیابی تاثیر آنالیز احساسات بر بازدهی سهام، نیاز به داده‌های قیمتی واقعی سهام پنج شرکت مورد نظر است، این داده‌ها اکنون در فضای وب و در بسیاری از منابع قابل یافت هست و در این مقاله از داده‌های روزانه تاریخی سایت یاهوفایننس[3] استفاده شده است. این داده‌ها حاوی اطلاعات حجم سهام در بازه تاریخی مدنظر (۲۰۱۵ تا ۲۰۲۰) هستند که شامل تاریخ معامله، قیمت آغازین[4]، بالاترین قیمت[5]، پایین‌ترین قیمت[6]، قیمت پایانی[7] (که در دنیای مالی به این چهار قیمت به اختصار OHLC گفته می‌شود) و حجم معاملات می‌باشند. شکل (۴) ساختار کلی داده‌های تاریخی قیمت سهام را نشان می دهد.

```
In [136]: prices = pd.read_csv('./CompanyValues.csv')
          prices
```

Out[136]:

	ticker_symbol	day_date	close_value	volume	open_value	high_value	low_value
0	AAPL	2020-05-29	317.94	38399530	319.25	321.15	316.4700
1	AAPL	2020-05-28	318.25	33449100	316.77	323.44	315.6300
2	AAPL	2020-05-27	318.11	28236270	316.14	318.71	313.0900
3	AAPL	2020-05-26	316.73	31380450	323.50	324.24	316.5000
4	AAPL	2020-05-22	318.89	29450750	315.77	319.23	315.3500
...	...	...	...	...	...	...	...
17523	TSLA	2019-12-21	405.59	14785210	410.29	413.00	400.1850
17524	TSLA	2019-12-22	405.59	14785210	410.29	413.00	400.1850
17525	TSLA	2019-12-25	425.25	8054720	418.36	425.47	412.6875
17526	TSLA	2019-12-28	430.38	9956827	435.00	435.31	426.1100
17527	TSLA	2019-12-29	430.38	9956827	435.00	435.31	426.1100

17528 rows × 7 columns

شکل۴. نمونه ساختار کلی داده‌های تاریخی قیمت سهام

[1] Web Scraping
[2] Selenium
[3] Yahoo Finance
[4] Open
[5] High
[6] Low
[7] High

پاک‌سازی داده‌ها[1]

با توجه به اینکه داده‌های مورد استفاده در مقاله در ابتدا داده‌های خام می‌باشند نیاز به پاک‌سازی و پیش‌پردازش داده‌ها وجود دارد. پاک‌سازی و پیش‌پردازش داده‌ها باعث دقیق‌تر شدن نتایج خواهد شد چرا که داده‌های خام ممکن باعث ایجاد اشکالاتی به خصوص در زمینه امتیازدهی به توییت‌ها و استفاده از داده‌های قیمتی در بک‌تستینگ شود. به همین دلیل، بر روی دو دسته از داده‌ها اقدام به پاک‌سازی شده است: نخست، داده‌های توییتر و گروه دوم، داده‌های قیمتی ۵ شرکت مد نظر.

داده‌های توییتر دارای ساختار اولیه شامل متن توییت، شناسه، نویسنده، تاریخ، تعداد لایک، تعداد کامنت، تعداد ریتوییت می‌باشند. برای افزایش دقت تحلیل، نیاز داریم که متن توییت‌ها را پاک‌سازی کنیم. متن توییت‌های استفاده شده در این مقاله، به زبان انگلیسی می‌باشند. به کمک RegEx و ماژول re در زبان پایتون، موارد زیر از توییت‌ها حذف شده است تا ساختاری منظم‌تر و مرتب‌تر داشته باشند:

فضاهای خالی اضافی (White Spaces):

فواصل اضافی که احتمالا به خاطر خطای کاربر یا برای زیبایی به متن اضافه شده‌اند.

لینک‌های وب:

آدرس سایت‌های مختلف که با اهداف متفاوتی از سوی کاربران در توییت استفاده می‌شوند.

نام کاربری اشخاص (Usernames):

نام کاربری اشخاص در توییتر که احتمالا به دلیل بازنشر[2] یا موارد دیگر به متن توییت اضافه شده‌اند.

ایست‌واژه‌ها (Stop words):

کلماتی که علی‌رغم تکرار بسیار، معنای خاصی ندارند؛ مثل as، the، a، is، are و ...

نکته دیگر اینکه، در ساختار توییت‌ها تاریخ به شکل post_date و در قالب ثانیه‌های سپری شده از تاریخ ۱۹۷۰/۰۱/۰۱ بیان شده است که لازم است تبدیل به شکل متعارف تاریخ با فرمت روبرو شود: Y-M-D که Y بیانگر سال، M بیانگر ماه و D بیانگر روز می‌باشند. برای این تبدیل، از تابع to_datetime موجود در کتابخانه پانداسِ پایتون استفاده شده است.

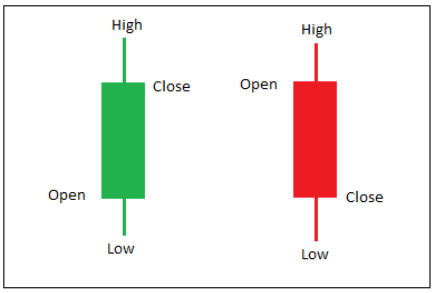
توییت‌ها با در نظرگرفتن موارد ذکر شده و به کمک توابعی در زبان پایتون پاک‌سازی گردید تا آماده استفاده در کتابخانه‌های مورد نظر برای امتیازدهی شوند.

داده‌های قیمتی مورد استفاده از سایت یاهو فایننس استخراج شده‌اند و دارای ساختار اولیه شامل نام نماد (ticker_symbol)، تاریخ (day_date)، قیمت پایانی (close_value)،

[1] Data cleaning
[2] Retweet

قیمت اولیه (open_value)، بالاترین قیمت (high_value)، پایین‌ترین قیمت (low_value) و حجم معاملات (volume) می‌باشند. با توجه به این‌که کتابخانه مورد استفاده برای بک‌تستینگ، کتابخانه backtesting.py می باشد و این کتاب‌خانه نیاز به ساختار خاصی از داده‌های ورودی دارد، باید ساختار داده‌های خام اولیه استخراجی از یاهوفایننس را تغییر داده تا ساختار دلخواه کتابخانه ایجاد شود.

شایان ذکر است که به طور کلی در بازارهای مالی و به خصوص بازار سهام، روند قیمتی سهام و سایر ابزارهای مالی در طول زمان، به کمک ساختارهایی به نام کَندِل[1] در نمودارها نمایش داده می‌شوند. کندل‌ها معمولا دارای اجزایی با نام اختصاری OHLC هستند که به ترتیب شامل قیمت اولیه[2] ، بالاترین قیمت[3] ، پایین‌ترین قیمت[4] و قیمت پایانی[5] می‌باشند که به صورت شکل (۵) در نمودارها نمایش داده می‌شوند.



شکل۵. ساختار یک کندل، سمت راست کندل منفی و سمت چپ کندل مثبت

معمولاً وقتی قیمت پایانی بالاتر از قیمت اولیه باشد کندل به رنگ سبز (شکل ۳ سمت چپ) و اگر قیمت پایانی پایین‌تر از قیمت اولیه باشد کندل به رنگ قرمز (شکل ۳ سمت راست) نمایش داده می‌شود. ساختار ورودی برای کتابخانه backtesting.py باید یک دیتافریم پانداس[6] شامل چهار ستون اجباری به ترتیب با نام‌های Open، High، Low و Close (به ترتیب قیمت اولیه، بالاترین قیمت، پایین‌ترین قیمت و قیمت پایانی) و همچنین یک ستون اختیاری با نام Volume (مربوط به حجم معاملات) می باشد که به کمک کتابخانه پانداس در پایتون، داده‌ها را در چارچوب ساختار مشخص شده در شکل(۶) تبدیل شده است.

---

[1] Candlestick
[2] Open price
[3] High price
[4] Low price
[5] Close price
[6] Pandas data frame

Out[1]:

day_date	Open	High	Low	Close	Volume
2010-06-01	37.0986	37.9914	36.9943	37.2614	218455576
2010-06-02	37.7914	37.8286	37.1907	37.7071	171594061
2010-06-03	37.8828	37.9357	37.2014	37.5886	162341809
2010-06-04	36.8871	37.4143	36.3757	36.5664	189045767
2010-06-07	36.8986	37.0214	35.7928	35.8486	221253336
...	...	...	...	...	...
2020-05-22	315.7700	319.2300	315.3500	318.8900	20450750
2020-05-26	323.5000	324.2400	316.5000	316.7300	31380450
2020-05-27	316.1400	318.7100	313.0900	318.1100	28236270
2020-05-28	316.7700	323.4400	315.6300	318.2500	33449100
2020-05-29	319.2500	321.1500	316.4700	317.9400	38399530

3085 rows × 5 columns

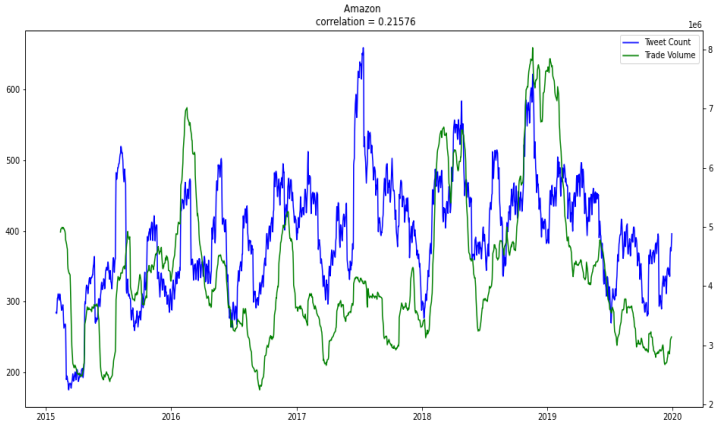
شکل ۶. ساختار نهایی داده‌های قیمتی برای استفاده در backtesting.py

پس از ارزیابی احساسات توییت‌ها و دسته‌بندی آن‌ها که در بخش قبلی بررسی شد، نوبت به بررسی ارتباط میان احساسات توییت‌ها و روند قیمتی سهام ۵ شرکت انتخابی می‌رسد. در بخش بعدی، ابتدا به بررسی ارتباط تعداد توییت‌ها و حجم معاملات پرداخته شده است. سپس ارتباط تعداد توییت‌های مثبت و منفی و روند قیمتی سهم بررسی و در نهایت، به کمک کتابخانه backtesting.py، میزان بازدهی با استفاده از استراتژی آنالیز احساسات محاسبه شده است.

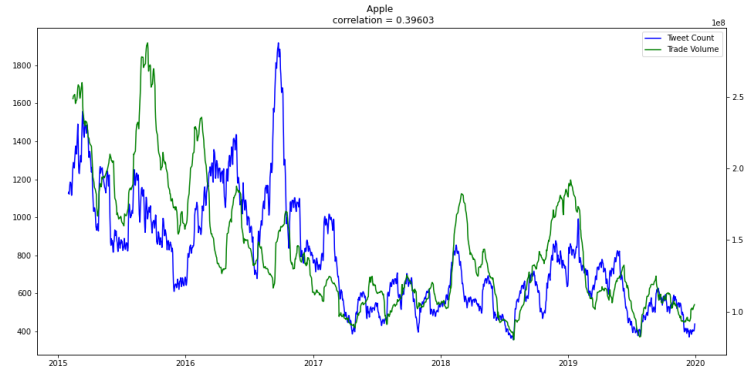
ارتباط حجم معاملات و تعداد توییت‌ها

در این بخش، رابطه میان تعداد توییت‌های مربوط به هر شرکت در روزهای مختلف، با حجم معاملات آن در بورس آمریکا بررسی شده است. حال به کمک داده‌های قیمتی که از کتابخانه یاهوفایننس استخراج شده و با استفاده از ستون حجم معامله (volume) رابطه میان حجم توییت‌ها و معاملات برای پنج شرکت مورد نظر ارزیابی شده است. برای این‌کار، نموداری شامل تعداد توییت و حجم معاملات در بازه زمانی مد نظر، ترسیم شده است. برای رسم نمودار، از کتابخانه matplotlib که یک کتابخانه بسیار مشهور برای رسم انواع نمودار در زبان برنامه نویسی پایتون است، استفاده می شود.

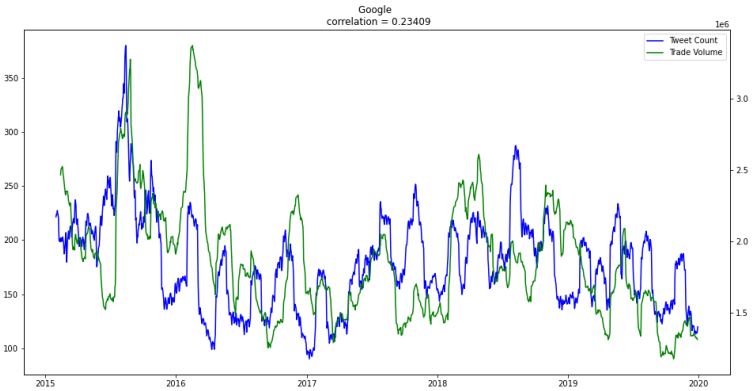
در ادامه، نمودارهای(۱) تا(۵) ارتباط بین تعداد توییت و حجم معاملات را برای ۵ شرکت منتخب بورس آمریکا در بازه سال‌های ۲۰۱۵ تا ۲۰۲۰ نشان می دهد. لازم به ذکر است که در هر نمودار، محور افقی مربوط به زمان، محور عمودی سمت چپ مربوط به تعداد توییت‌ها و محور عمودی سمت راست مربوط به حجم معاملات سهم می‌باشد. تعداد توییت‌ها با رنگ آبی و حجم معاملات با رنگ سبز نشان داده شده‌اند.



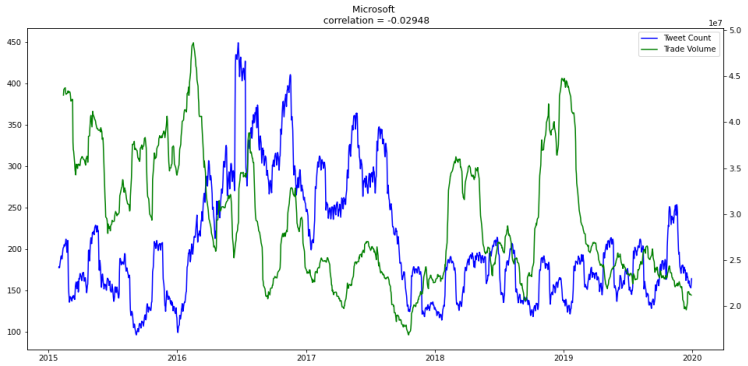
نمودار ۱. روند تعداد توییت–حجم معاملات شرکت آمازون.



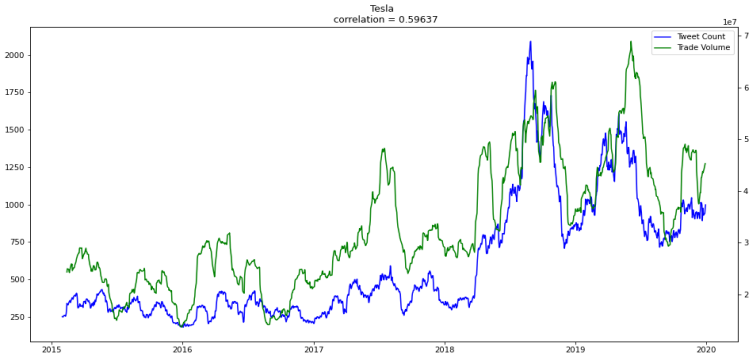
نمودار ۲. روند تعداد توییت–حجم معاملات شرکت اپل.



نمودار ۳. روند تعداد توییت–حجم معاملات شرکت گوگل.



نمودار ۴. روند تعداد توییت–حجم معاملات شرکت مایکروسافت.

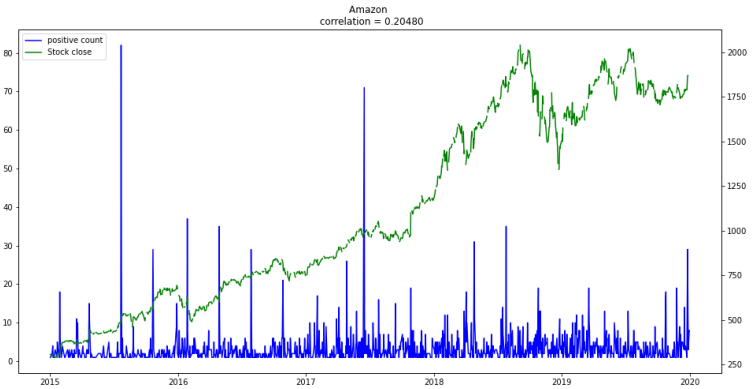


نمودار ۵. روند تعداد توییت–حجم معاملات شرکت تسلا.

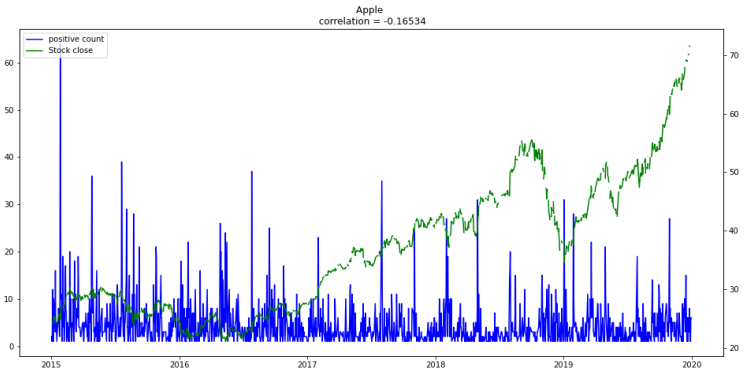
ارتباط احساسات مثبت و قیمت سهام

در این بخش و بخش بعدی، ارتباط بین تعداد توییت‌های مثبت و منفی و قیمت پایانی سهام ۵ شرکت مد نظر بررسی می شود. نحوه استخراج تعداد توییت‌های مثبت و منفی را در بخش‌های قبلی توضیح داده شده است. از داده‌های قیمتی، قیمت پایانی مورد استفاده قرار گرفته است.

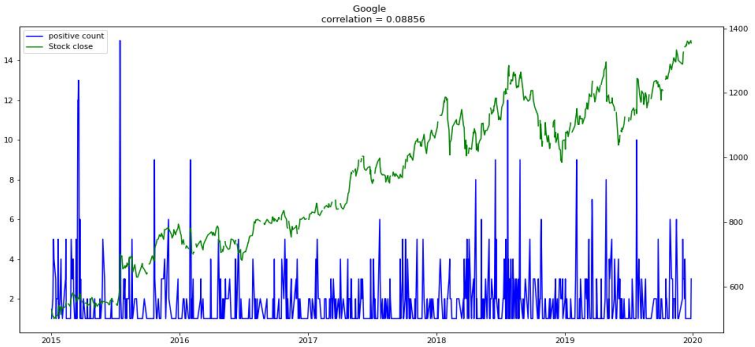
در ادامه، نمودارهای (۶) الی (۱۰) ارتباط بین تعداد توییت‌های مثبت و قیمت پایانی سهام را برای ۵ شرکت منتخب بورس آمریکا را نشان می دهد. لازم به ذکر است که در هر نمودار، محور افقی مربوط به زمان، محور عمودی سمت چپ مربوط به تعداد توییت‌های مثبت و محور عمودی سمت راست مربوط به قیمت پایانی سهام می‌باشد. تعداد توییت‌ها با رنگ آبی و قیمت با رنگ سبز نشان داده شده‌اند.



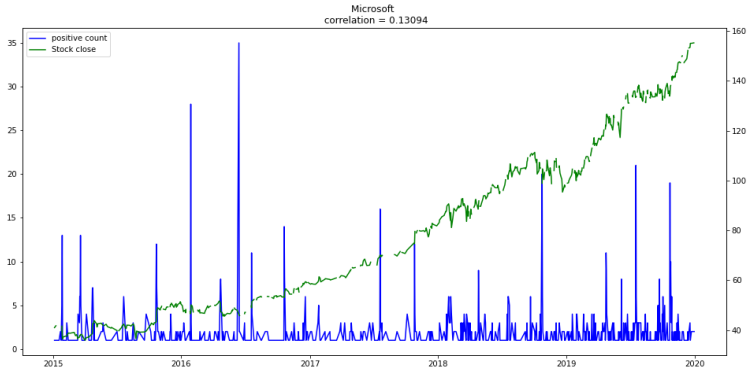
نمودار ۶. روند تعداد توییت مثبت–قیمت پایانی شرکت آمازون.



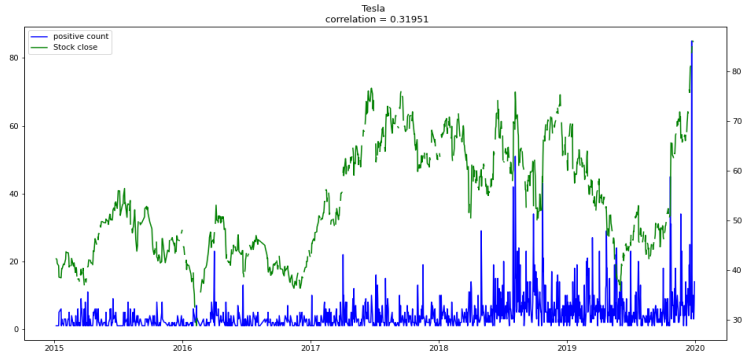
نمودار ۷. روند تعداد توییت مثبت–قیمت پایانی شرکت اپل.



نمودار ۸. روند تعداد توییت مثبت–قیمت پایانی شرکت گوگل.



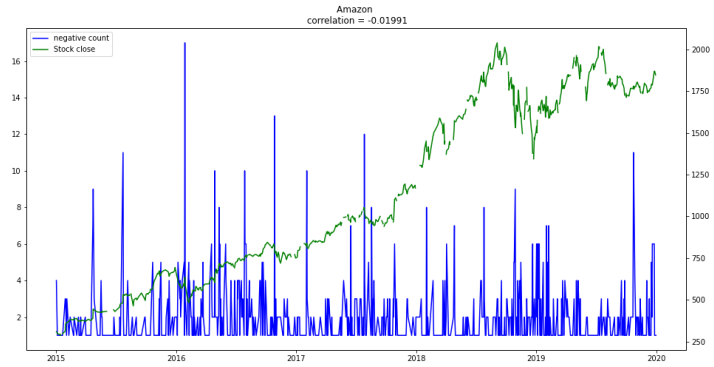
نمودار ۹. روند تعداد توییت مثبت-قیمت پایانی شرکت مایکروسافت.



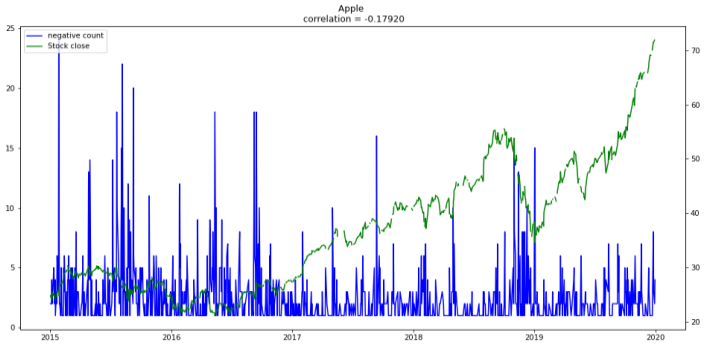
نمودار ۱۰. روند تعداد توییت مثبت-قیمت پایانی شرکت تسلا.

ارتباط احساسات منفی و قیمت سهام

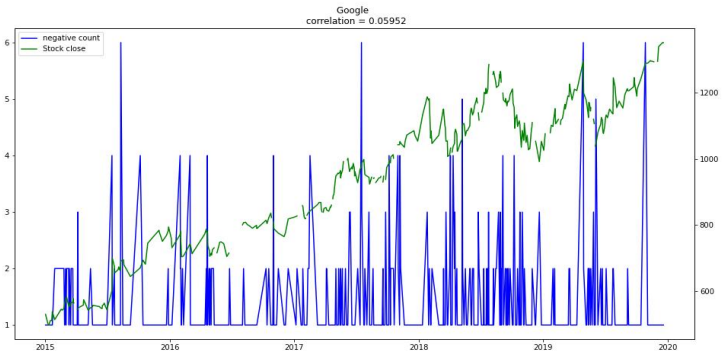
در این قسمت، نمودارهای (۱۱) الی (۱۵) ارتباط تعداد توییت‌های منفی و قیمت پایانی سهام را برای ۵ شرکت منتخب بورس آمریکا نشان می دهد. لازم به ذکر است که در هر نمودار، محور افقی مربوط به زمان، محور عمودی سمت چپ مربوط به تعداد توییت‌های مثبت و محور عمودی سمت راست مربوط به قیمت پایانی سهام می‌باشد. تعداد توییت‌ها با رنگ آبی و قیمت پایانی با رنگ سبز نشان داده شده‌اند.



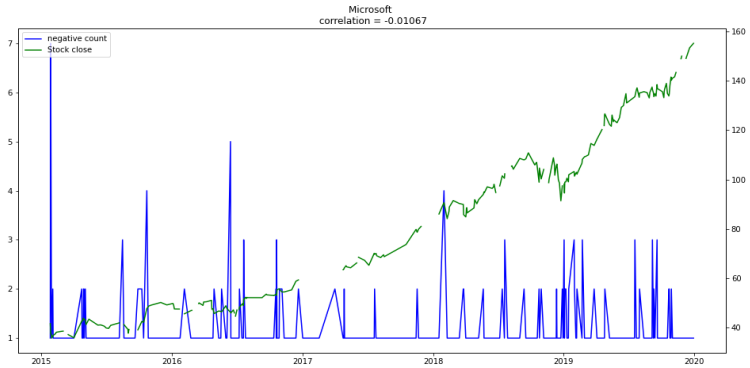
شکل ۱۱. روند تعداد توییت منفی–قیمت پایانی شرکت آمازون.



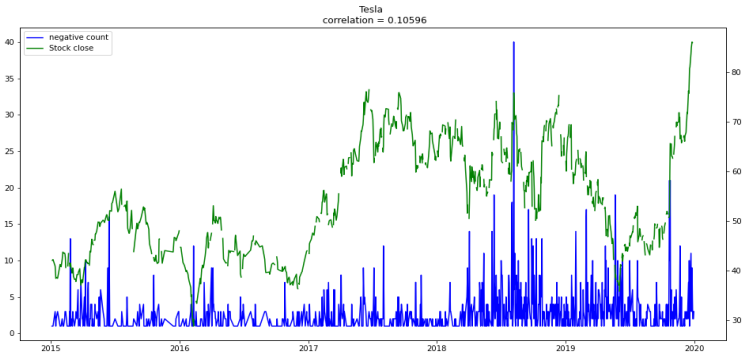
نمودار ۱۲. روند تعداد توییت منفی–قیمت پایانی شرکت اپل.



نمودار ۱۳. روند تعداد توییت منفی–قیمت پایانی شرکت گوگل.



نمودار ۱۴. روند تعداد توییت منفی-قیمت پایانی شرکت مایکروسافت.



نمودار ۱۵. روند تعداد توییت منفی-قیمت پایانی شرکت تسلا.

بک تستینگ

بک‌تستینگ[1] روشی کلی برای مشاهده عملکرد یک استراتژی یا مدل با توجه به نمودارهای گذشته می‌باشد. این روش، ضریب اطمینان یک استراتژی معاملاتی را با توجه به عملکرد آن با داده‌های تاریخی ارزیابی می‌کند. اگر بک‌تستینگ نتیجه خوبی داشته باشد، معامله‌گران و تحلیل‌گران بازار می‌توانند با اطمینان از این استراتژی‌ها برای آینده نیز استفاده کنند.

کتابخانه مورد استفاده برای بک‌تستینگ در این مقاله، backtesting.py نام دارد که یک کتابخانه پرکاربرد پایتون می‌باشد. برای استفاده از این کتابخانه، باید داده‌های ورودی دارای ساختار خاصی باشند. این ساختار شامل یک دیتافریم با ستون‌های تاریخ، قیمت آغازین، بیشترین قیمت، کمترین قیمت، قیمت پایانی و حجم معاملات باشد. با توجه به اینکه سعی برآن است داده‌های احساسات نیز دخیل داده شود، با استفاده از تابع merge پانداس در پایتون، دو ستون دیگر شامل تعداد توییت‌های مثبت و تعداد توییت‌های منفی را نیز به ساختار داده‌ها اضافه شده

[1] Backtesting

است. در نهایت، شکل(۷) ساختار مورد نظر برای استفاده در کتابخانه backtesting به صورت زیر خواهد بود:

Date	Open	High	Low	Close	Volume	positiveCount	negativeCount
2014-12-31	44.618000	45.136002	44.450001	44.481998	11487500	0.0	0.0
2015-01-02	44.574001	44.650002	42.652000	43.862000	23822000	0.0	0.0
2015-01-05	42.910000	43.299999	41.431999	42.018002	26842500	0.0	1.0
2015-01-06	42.012001	42.840000	40.841999	42.256001	31309500	1.0	0.0
2015-01-07	42.669998	42.956001	41.956001	42.189999	14842000	1.0	1.0
...	...	...	...	...	...	...	...
2019-12-23	82.356003	84.402000	82.000000	83.844002	66598000	85.0	11.0
2019-12-24	83.671997	85.094002	82.538002	85.050003	40273500	16.0	3.0
2019-12-26	85.582001	86.695999	85.269997	86.188004	53169500	8.0	9.0
2019-12-27	87.000000	87.061996	85.222000	86.075996	49728500	10.0	5.0
2019-12-30	85.758003	85.800003	81.851997	82.940002	62932000	7.0	2.0

1258 rows × 7 columns

شکل ۷. ساختار نهایی داده‌های مورد استفاده کتابخانه backtesting.

۱- استراتژی

برای ارزیابی بک تست استراتژی، دو عدد n_1 و n_2 را در نظر می‌گیریم. وقتی تعداد توییت-های مثبت در تاریخی معین، بیشتر از n_1 باشد دستور خرید صادر می‌شود و هرگاه تعداد توییت-های منفی بیشتر از n_2 باشد دستور فروش صادر می‌شود.

۲- نتایج بک تستینگ

بسته به استفاده از اعداد مختلف برای n_1 و n_2، بازدهی‌های متفاوتی برای معاملات سهام موردنظر ایجاد می‌شود. به عنوان نمونه، با اعداد $n_1=$ ۶ و $n_2=$ ۵ برای سهام آمازون و طبق داده‌های احساساتی که قبلا تخمین زده شده است، شکل (۸) نتایج این نمونه را نشان می دهد. پارامترهای محاسبه شده برای این استراتژی معاملاتی، شامل مواردی مثل بازدهی، بازدهی استراتژی خرید و نگهداری و ... می‌باشد که در زیر هرکدام تعریف شده‌اند:

Start: تاریخ شروع بازه زمانی

End: تاریخ پایان بازه زمانی

Duration: طول مدت بازه زمانی

Exposure time: درصدی از کل بازه زمانی که موقعیت باز خرید یا فروش داشته‌ایم.

Equity final: مقدار نهایی سرمایه

Equity peak: بیشترین مقدار سرمایه

Return: درصد بازدهی

Buy and hold return: درصد بازدهی استراتژی خرید و نگهداری
Volatility: میزان نوسان
Sharpe ratio: نسبت شارپ (پاداش به تغییرپذیری)
Drawdown: زیان
Trades: تعداد معاملات
Win rate: درصد پیروزی (معامله با سود)
که در این استراتژی، برای شرکت آمازون، بازدهی ۱۲۹ درصد بوده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	95.786963
Equity Final [$]	22925.375757
Equity Peak [$]	37173.027608
Return [%]	129.253758
Buy & Hold Return [%]	495.099075
Return (Ann.) [%]	18.080412
Volatility (Ann.) [%]	33.355299
Sharpe Ratio	0.542055
Sortino Ratio	0.951525
Calmar Ratio	0.443312
Max. Drawdown [%]	-40.78485
Avg. Drawdown [%]	-4.017662
Max. Drawdown Duration	489 days 00:00:00
Avg. Drawdown Duration	33 days 00:00:00
# Trades	141
Win Rate [%]	56.028369
Best Trade [%]	30.313972
Worst Trade [%]	-17.192865

شکل ۸. نتایج عددی استراتژی نمونه برای شرکت آمازون.

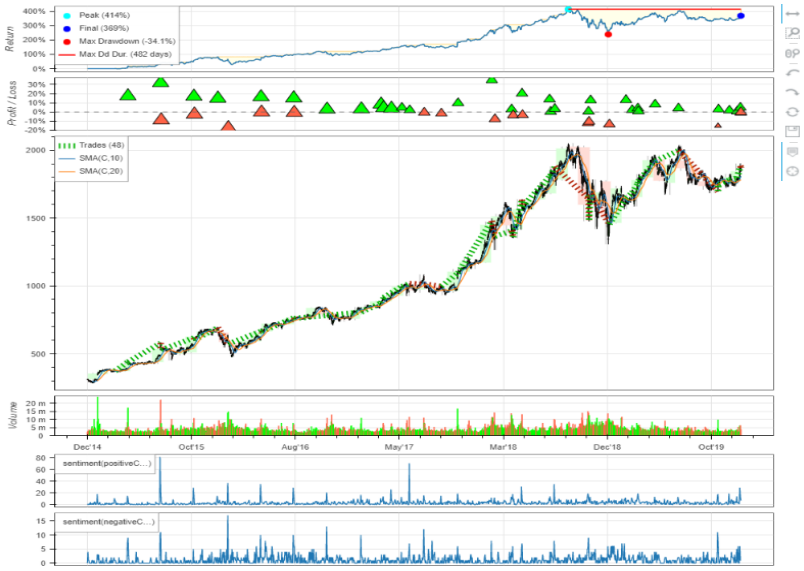
در نمودار(۱۶) این اطلاعات و همچنین نقاط معامله را به صورت گرافیکی و بر روی نمودار قیمت گزارش شده است. در مورد نمودار(۱۶)، لازم به توضیح است که در بالاترین قسمت، بازدهی به مرور زمان دیده می‌شود. در قسمت بعدی، نقاط خرید و فروش همراه با سود و ضرر متناسب با آن و قسمت سوم نیز نمودار کلی روند قیمت سهم و نقاط سود و زیان را نمایش می‌دهد. سه بخش بعدی نیز به ترتیب حجم معاملات، تعداد توییت‌های با احساسات مثبت و تعداد توییت‌های با احساسات منفی را طی دوره زمانی مربوطه، نشان می‌دهند.

نمودار ۱۶. نمایش گرافیکی استراتژی نمونه برای شرکت آمازون.

در ادامه با استفاده از تکنیک‌های بهینه‌سازی کتابخانه backtesting پایتون، بهترین استراتژی موجود برای هر سهم برآورد شده است؛ این کار در واقع معادل است با محاسبه اعداد n_1 و n_2 طوری که بازدهی بیشینه شود. نتایج این محاسبات در ادامه آمده است. برای شرکت آمازون، بهترین اعداد $n_1=$ ۱۰ و $n_2=$ ۱۰ به دست آمدند. نتایج عددی این استراتژی و نمایش گرافیکی مربوطه به ترتیب در شکل(۹) و نمودار(۱۷) نشان داده شده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	95.786963
Equity Final [$]	46856.260889
Equity Peak [$]	51429.953414
Return [%]	368.562609
Buy & Hold Return [%]	495.099075
Return (Ann.) [%]	36.259505
Volatility (Ann.) [%]	38.53678
Sharpe Ratio	0.940906
Sortino Ratio	2.026967
Calmar Ratio	1.062541
Max. Drawdown [%]	-34.125269
Avg. Drawdown [%]	-3.622064
Max. Drawdown Duration	482 days 00:00:00
Avg. Drawdown Duration	25 days 00:00:00
# Trades	48
Win Rate [%]	66.666667
Best Trade [%]	34.785444
Worst Trade [%]	-17.413381

شکل ۹. نتایج استراتژی بهینه برای شرکت آمازون.

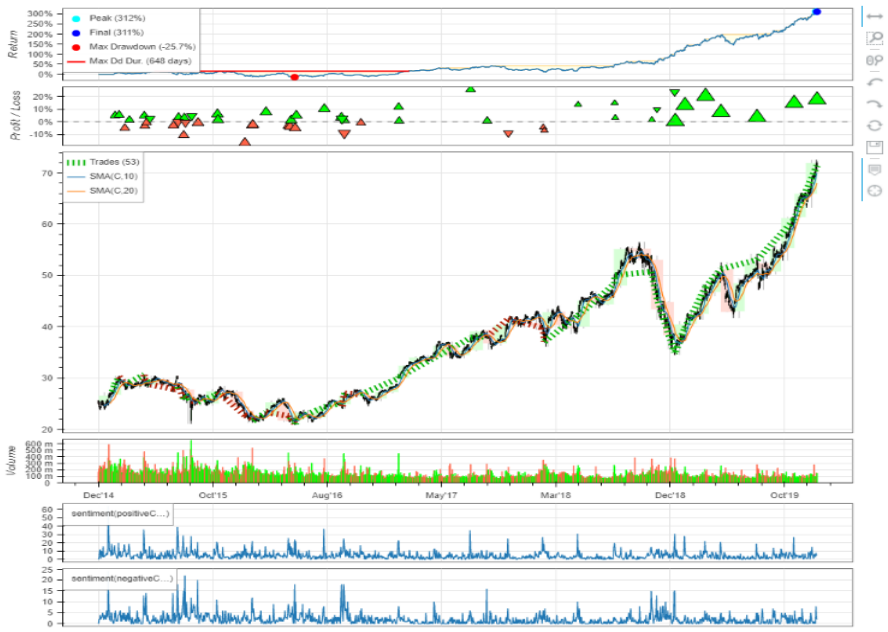


نمودار ۱۷. نمایش گرافیکی استراتژی بهینه برای شرکت آمازون.

برای شرکت اپل، بهترین اعداد $n_1=$ ۱۰ و $n_2=$ ۱۵ به دست آمدند. نتایج عددی این استراتژی و نمایش گرافیکی مربوطه به ترتیب در شکل(۱۰) و نمودار(۱۸) نشان داده شده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	98.09221
Equity Final [$]	41088.830961
Equity Peak [$]	41152.669455
Return [%]	310.88831
Buy & Hold Return [%]	187.441037
Return (Ann.) [%]	32.721078
Volatility (Ann.) [%]	33.108866
Sharpe Ratio	0.988287
Sortino Ratio	1.99558
Calmar Ratio	1.272276
Max. Drawdown [%]	-25.718533
Avg. Drawdown [%]	-3.198152
Max. Drawdown Duration	649 days 00:00:00
Avg. Drawdown Duration	29 days 00:00:00
# Trades	53
Win Rate [%]	64.150943
Best Trade [%]	25.387063
Worst Trade [%]	-16.593601

شکل ۱۰. نتایج استراتژی بهینه برای شرکت اپل.

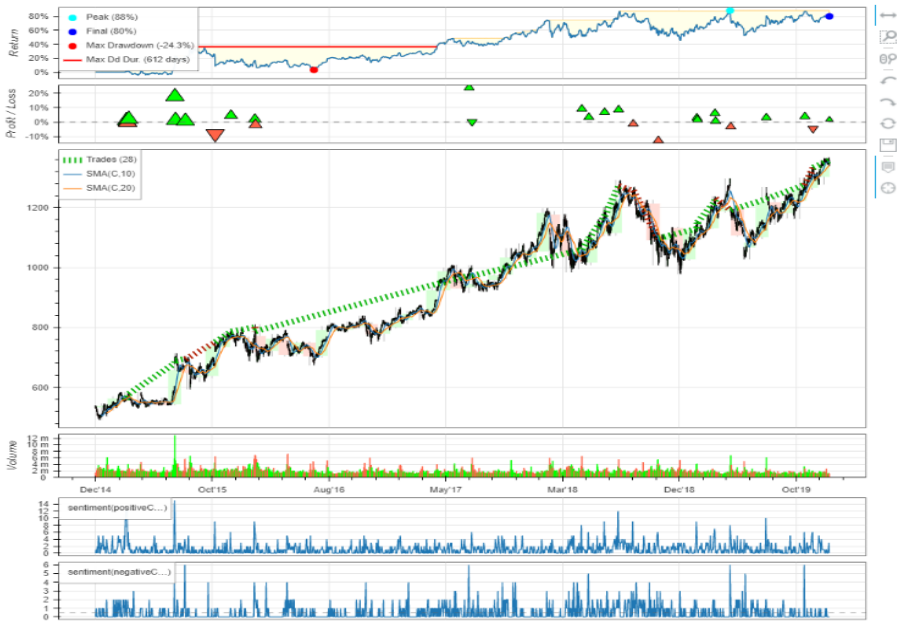


نمودار ۱۸. نمایش گرافیکی استراتژی بهینه برای شرکت اپل.

برای شرکت گوگل، بهترین اعداد $n_1=$ ۵ و $n_2=$ ۵ به دست آمدند. نتایج عددی این استراتژی و نمایش گرافیکی مربوطه به ترتیب در شکل(۹) و نمودار(۱۹) نشان داده شده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	95.866455
Equity Final [$]	17994.44489
Equity Peak [$]	18816.937762
Return [%]	79.944449
Buy & Hold Return [%]	152.461092
Return (Ann.) [%]	12.488679
Volatility (Ann.) [%]	25.838003
Sharpe Ratio	0.483345
Sortino Ratio	0.791464
Calmar Ratio	0.513909
Max. Drawdown [%]	-24.301322
Avg. Drawdown [%]	-4.806665
Max. Drawdown Duration	612 days 00:00:00
Avg. Drawdown Duration	66 days 00:00:00
# Trades	28
Win Rate [%]	75.0
Best Trade [%]	23.669375
Worst Trade [%]	-12.776131

شکل ۹. نتایج استراتژی بهینه برای شرکت گوگل.

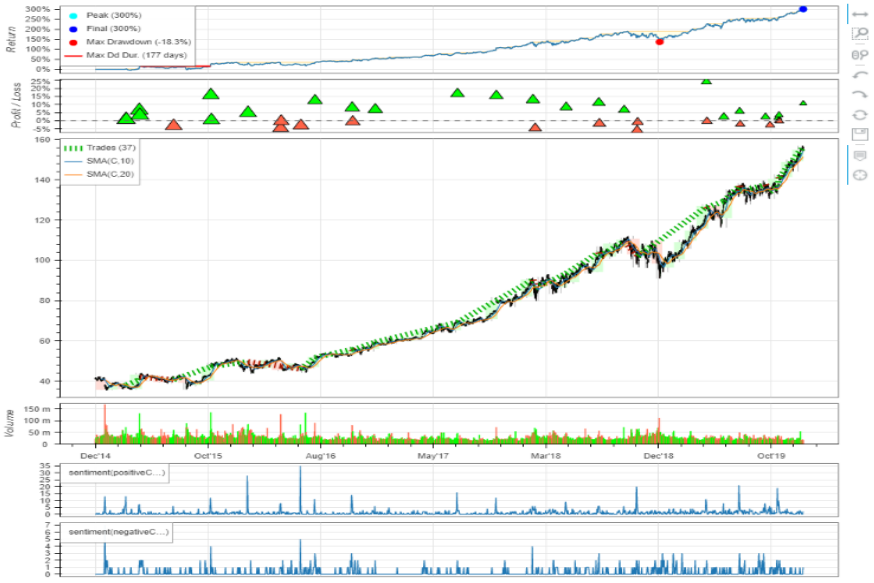


نمودار ۱۹. نمایش گرافیکی استراتژی بهینه برای شرکت گوگل.

برای شرکت مایکروسافت، بهترین اعداد n۱= ۵ و n۲= ۵ به دست آمدند. نتایج عددی این استراتژی و نمایش گرافیکی مربوطه به ترتیب در شکل(۱۰) و نمودار(۲۰) نشان داده شده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	95.786963
Equity Final [$]	39978.893127
Equity Peak [$]	39978.893127
Return [%]	299.788931
Buy & Hold Return [%]	277.905626
Return (Ann.) [%]	31.99501
Volatility (Ann.) [%]	29.919164
Sharpe Ratio	1.069382
Sortino Ratio	2.204615
Calmar Ratio	1.744251
Max. Drawdown [%]	-18.343124
Avg. Drawdown [%]	-2.680061
Max. Drawdown Duration	178 days 00:00:00
Avg. Drawdown Duration	20 days 00:00:00
# Trades	37
Win Rate [%]	62.162162
Best Trade [%]	23.993431
Worst Trade [%]	-5.827166

شکل ۱۰. نتایج استراتژی بهینه برای شرکت مایکروسافت.

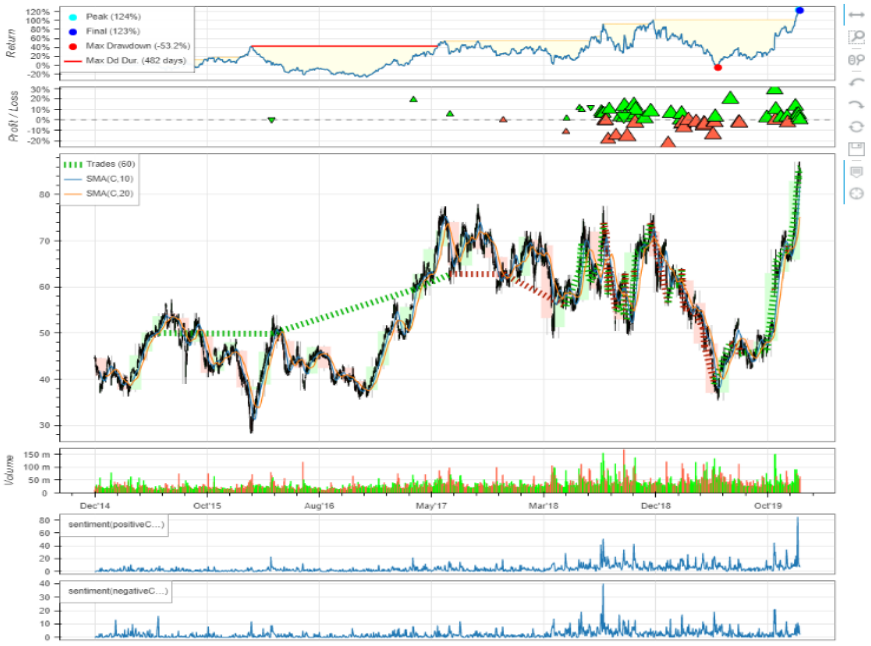


نمودار ۲۰. نمایش گرافیکی استراتژی بهینه برای شرکت مایکروسافت.

برای شرکت تسلا، بهترین اعداد n۱= ۱۵ و n۲= ۱۲ به دست آمدند. نتایج عددی این استراتژی و نمایش گرافیکی مربوطه به ترتیب در شکل(۱۱) و نمودار(۲۱) نشان داده شده است.

Start	2014-12-31 00:00:00
End	2019-12-30 00:00:00
Duration	1825 days 00:00:00
Exposure Time [%]	90.937997
Equity Final [$]	22298.986752
Equity Peak [$]	22410.786831
Return [%]	122.989868
Buy & Hold Return [%]	86.457455
Return (Ann.) [%]	17.426947
Volatility (Ann.) [%]	52.372593
Sharpe Ratio	0.332749
Sortino Ratio	0.621011
Calmar Ratio	0.327669
Max. Drawdown [%]	-53.184664
Avg. Drawdown [%]	-11.115169
Max. Drawdown Duration	483 days 00:00:00
Avg. Drawdown Duration	65 days 00:00:00
# Trades	60
Win Rate [%]	58.333333
Best Trade [%]	28.695625
Worst Trade [%]	-23.374733

شکل ۱۱. نتایج استراتژی بهینه برای شرکت تسلا.



نمودار۲۲. نمایش گرافیکی استراتژی بهینه برای شرکت تسلا.

۳- نتایج استراتژی SMA CROSS تحلیل تکنیکال

در این بخش و برای مقایسه نتایج به‌دست‌آمده از به‌کارگیری استراتژی‌های آنالیز احساسات، از یک استراتژی معروف در تحلیل تکنیکال به نام SMA CROSS[1] استفاده شده است. این استراتژی نمونه‌ای از استفاده از میانگین متحرک برای پیش‌بینی روند آتی سهام می‌باشد. طبق این استراتژی، هرگاه برای سهمی، SMA کوتاه‌مدتی‌تر، SMA بلندمدتی‌تر را به سمت بالا قطع کند دستور خرید صادر می‌شود و هرگاه SMA بلندمدتی‌تر، SMA کوتاه‌مدتی‌تر را به سمت بالا قطع کند دستور فروش صادر می‌شود. در اینجا از میانگین متحرک‌های ۱۰ و ۲۰ روزه استفاده شده است که متداول‌ترین اعداد برای این استراتژی هستند. این استراتژی با کمک دستور crossover در کتابخانه backtesting در پایتون اجرا شده است. نتایج بازدهی این استراتژی برای ۵ شرکت مورد نظر به شرح جدول(۳) می‌باشد:

[1] The Simple Moving-Average Cross

جدول ۳. مقایسه بازدهی استراتژی SMA CROSS و تحلیل احساسات برای ۵ شرکت.

نام شرکت	نماد	بازدهی SMA CROSS (درصد)	بازدهی تحلیل احساسات (درصد)
آمازون	AMZN	۴۷	۳۶۸
اپل	AAPL	۷۱	۳۱۰
گوگل	GOOGL	-۵۸	۷۹
مایکروسافت	MSFT	-۲۹	۲۹۹
تسلا	TSLA	-۴۴	۱۲۲

همانطور که نتایج ارائه شده در جدول(۳) نشان می‌دهد، بازدهی استراتژی مبتنی بر تحلیل احساسات در مقایسه با بازدهی استراتژی SMA CROSS در سهام آمازون، اپل، گوگل، مایکروسافت و تسلا به ترتیب ۳۲۱٪، ۲۳۹٪، ۱۳۷٪، ۳۲۸٪ و ۱۶۶٪ افزایش یافته است.

۵. بحث و نتیجه‌گیری

در این پژوهش، مبحث تحلیل احساسات و میزان تاثیر آن در روند حرکتی سهام مورد بررسی قرار گرفت. لذا با استفاده از داده‌های توییت‌های شبکه اجتماعی توییتر و داده‌های قیمتی سهام ۵ شرکت برتر آمریکا شامل آمازون، اپل، گوگل، مایکروسافت و تسلا به بررسی ارتباط میان حجم معاملات و تعداد توییت‌ها، احساسات مثبت و قیمت سهام، احساسات منفی و قیمت سهام پرداخته شد. همچنین نتایج بازدهی سهام را با اتخاذ استراتژی‌های معاملاتی مختلف با تکیه بر احساسات به‌دست‌آمده از توییت‌ها ارزیابی کردیم. با توجه به تحقیق انجام شده، به نظر می‌رسد حجم معاملات سهام ارتباط خوبی با تعداد توییت‌های مربوط به آن سهم دارد. این ارتباط برای شرکت تسلا با ضریب همبستگی ۰.۵۹، بیشترین مقدار را داشت و سایر شرکت‌ها نیز ضریب همبستگی نسبتا خوبی را ثبت کردند. البته به جز شرکت آمازون که با ضریب همبستگی ۰.۰۲- ، ظاهرا ارتباط خاصی بین حجم معاملات و تعداد توییت‌های مربوط خود ندارد. در مورد بعضی شرکت‌ها مانند تسلا و آمازون، ارتباط خوبی بین احساسات مثبت در شبکه اجتماعی توییتر و قیمت سهام مشاهده شد که به این معنی است که توییت‌های مثبت منتشر شده در این شبکه اجتماعی مردم را تشویق کرده است تا اقدام به خرید سهام این شرکت‌ها نمایند. در بخش ارتباط توییت‌های منفی و قیمت سهام ارتباط خاصی بین این دو مورد مشاهده نشد، اگرچه باز هم تسلا بیشترین ضریب همبستگی را در این بخش دارا بود. به نظر می‌رسد دلیل این امر آن است که بازار سهام آمریکا و به تبع سهام این ۵ شرکت بزرگ، در این بازه ۵ ساله همواره صعودی بوده‌اند و بازدهی‌های خوبی را ثبت کرده‌اند. بنابراین، سایر عوامل موثر بر افزایش قیمت سهام این شرکت‌ها بر احساسات منفی در شبکه‌های اجتماعی غلبه داشته است. در نهایت و در بخش بک-تستینگ، بازدهی به‌دست‌آمده برای شرکت‌های اپل، مایکروسافت و تسلا بیشتر از بازدهی

استراتژی خرید و نگهداری بود اما شرکت‌های آمازون و گوگل بازدهی کمتر از استراتژی خرید و نگهداری داشتند. با توجه به این نکته که بازار آمریکا در این بازه زمانی همواره روند صعودی داشته است، به نظر می‌رسد این نتایج مورد قبول هستند. ضمنا نتایج بازده این استراتژی‌ها در مقایسه با یک استراتژی تحلیل تکنیکال (SMA CROSS) به طور قابل توجهی افزایش یافته است. به نظر می‌رسد با تکیه بر اطلاعات احساسات سهام و بررسی مطالب منتشره در شبکه‌های اجتماعی در مورد شرکت‌های مختلف می‌توان به بازدهی معقولی دست یافت. اما عوامل موثر بر قیمت سهام بسیار زیاد می‌باشند و حتی در بازارهای بسیار کارا، قیمت‌ها از گشت تصادفی پیروی می‌کنند. برای پیشنهاد آتی نیز می توان استفاده از داده‌های سایر شبکه‌های اجتماعی مانند تلگرام، فیسبوک و ... در آنالیز احساسات، استفاده از سایر تکنیک‌های آنالیز احساسات مانند روش‌های یادگیری ماشین، استفاده از سایر استراتژی‌های بک‌تستینگ، متفاوت از آنچه در این مقاله انجام شد و ترکیب استراتژی‌های آنالیز احساسات و تحلیل تکنیکال برای ارزیابی بازدهی سهام را مدنظر قرار داد.

منابع

1. Alanyali, M., Moat, H. S., & Preis, T. (2013). Quantifying the relationship between financial news and the stock market. *Scientific reports*, *3*(1), 1-6.

2. Birz, G., & Lott Jr, J. R. (2011). The effect of macroeconomic news on stock returns: New evidence from newspaper coverage. *Journal of Banking & Finance*, *35*(11), 2791-2800.

3. Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Proceedings of the international AAAI conference on web and social media,

4. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8.

5. Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. *PloS one*, *7*(7), e40014.

6. Bordino, I., Kourtellis, N., Laptev, N., & Billawala, Y. (2014). Stock trade volume prediction with yahoo finance user browsing behavior. 2014 IEEE 30th International Conference on Data Engineering,

7. Chan, W. S. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics*, *70*(2), 223-260.

8. Cutler, D. M., Poterba, J. M., & Summers, L. H. (1988). What moves stock prices? In: National Bureau of Economic Research Cambridge, Mass., USA.

9. Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of* Finance, *66*(5), 1461-1499.

10. De Oliveira Carosia, A. E., Coelho, G. P., & da Silva, A. E. A. (2021). Investment strategies applied to the Brazilian stock market: a methodology based on sentiment analysis with deep learning. *Expert Systems with Applications*, 184, 115470.

11. Graham, M., Hale, S. A., & Gaffney, D. (2014). Where in the world are you? Geolocation and language identification in Twitter. *The Professional Geographer*, *66*(4), 568-578.

12. Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high-frequency news-implied market reactions. *Journal of Empirical Finance*, *18*(2), 321-340.

13. https://github.com/cjhutto/vaderSentiment.

14. Huang, M. Y., Rojas, R. R., & Convery, P. D. (2020). Forecasting stock market movements using google trend searches. *Empirical Economics*, *59*(6), 2821-2839.

15. Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, 178, 115019.

16. Kristoufek, L. (2013). Can Google Trends search queries contribute to risk diversification? *Scientific reports*, *3*(1), 1-5.

17. Lillo, F., Micciche, S., Tumminello, M., Piilo, J., & Mantegna, R. N. (2015). How news affects the trading behavior of different categories of investors in a financial market. *Quantitative Finance*, *15*(2), 213-229.

18. Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, Twitter and search engine data. *arXiv preprint arXiv:1112.1051*.

19. Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. Proceedings of the First ACM international workshop on hot topics on interdisciplinary social networks research,

20. Mohebbi, S., Fadaienejad, M. E., & Hamidizadeh, M. r. (2021). The Proposed Algorithm to Select Appropriate Features for Predicting Tehran Stock Exchange Index. *-Journal of Financial Management Perspective*, -. https://doi.org/10.52547/jfmp.2021.220761.1015

21. Nguyen, V. D., Varghese, B., & Barker, A. (2013). The royal birth of 2013: Analysing and visualizing public sentiment in the UK using Twitter. 2013 IEEE International Conference on Big Data,

22. Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1933), 5707-5719.

23. Souza, T. T. P., Kolchyna, O., Treleaven, P. C., & Aste, T. (2015). Twitter sentiment analysis applied to finance: A case study in the retail industry. *arXiv preprint arXiv:1507.00784*.

24. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of* Finance, *62*(3), 1139-1168.

25. Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of* Finance, *63*(3), 1437-1467.

26. Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics*, *82*(1), 103-133.

27. Yaftian, A. H., & Rastegar, M. A. (2020). Designing an Automated Trading System Using Convolutional Neural Network. *-Journal of Financial Management Perspective*, *10*(31), 153-184. https://doi.org/10.52547/jfmp.10.31.153