



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

گزارش پروژه

پایش رویدادهای فضای فضای مجازی

نگارش

بردیا اردکانیان

محمدرضا حسنزاده

موژان میرجلیلی

فرهان کیهان

استاد راهنما

دکتر میثم نظریانی

استاد مشاور

دکتر میثم نظریانی

خرداد ۱۴۰۱



سپاس‌گزاری

در اینجا بر خود لازم می‌دانیم که از راهنمای‌های استاد گرانقدر جناب آقای دکتر نظریای کمال تشکر را داشته باشیم و همچنین از تمام کسانی که امکان کرده‌اند، اظهار قدردانی می‌کنیم.

چکیده

در سال‌های اخیر بورس اوراق بهادار مورد توجه افراد زیادی قرار گرفته است. علت این موضوع می‌تواند از شرایط اقتصادی تا مناسب بودن شرایط سرمایه‌گذاری در بازار اوراق بهادار باشد. به همین سبب افرادی که تجربه کمتری در زمینه بورس داشتند به دنبال افراد/افرادی گشتند که چراغ راهشان در این بازار جدید باشد.

افراد بسیاری در شبکه‌های مجازی با نام تحلیلگر بورسی پیشنهادهای مبنی بر خرید یا فروش سهام‌ها می‌دهند. گاهی افرادی پیدا می‌شوند که با سوء استفاده از «اطلاعات نهانی» و «سیگنال دهی» نسبت به سهام داران برتری پیدا می‌کنند. این سوء استفاده باعث کاهش اعتماد عمومی نسبت به بازار اوراق بهادار می‌شود.

در این پروژه به بررسی چند طرح برای بررسی و کشف تقلب در بازار اوراق بهادار می‌پردازیم. طرح مطرح شده در این پروژه روش‌های اخیر و مرسوم دنیای دیجیتال می‌باشد که در این پروژه علاوه بر الگوریتم‌های این روش‌ها برتری هر کدام را مورد بررسی قرار می‌دهیم که باعث جلب اعتماد عمومی نسب به بازار اوراق بهادار خواهد شد.

واژه‌های کلیدی:

بازار اوراق بهادار، سیگنال دهی، کشف تقلب، اطلاعات نهانی، پایش

چکیده.....	أ
فصل اول مقدمه.....	۱
۲-۱ اهداف پروژه.....	۲
۲-۳-۱ اهداف آرمانی.....	۲
۲-۳-۲ هدف کلی.....	۲
۲-۳-۳ اهداف ویژه.....	۳
۲-۳-۴ اهداف کاربردی.....	۳
۲-۲ راه حل ایده آل و نیازمندی های مرتبط با پایش رویدادهای فضای مجازی به منظور نظارت بر بازار سرمایه	۳
۲-۳-۱ فضاهای مجازی هدف.....	۳
۲-۳-۲ نیازمندی های مرتبط با پایش رویدادهای فضای مجازی.....	۴
۲-۳-۳ تحلیل به منظور نظارت بر بازار سرمایه.....	۷
فصل دوم مشخص کردن محدوده و ساده سازی پروژه.....	۹
تفاوت پروژه ایده آل و پروژه هدف.....	۱۰
۲-۱ مشخص کردن Scope و Not Scope.....	۱۰
۲-۳-۴ محدوده پروژه.....	۱۰
۲-۳-۵ خارج محدوده پروژه.....	۱۱
۲-۲ معرفی نیازمندی های شناسایی شده در قالب Product Backlog.....	۱۱
۲-۳-۶ Product Backlog مربوط به تیم توسعه محصول.....	۱۲
۲-۳-۷ تیم تحلیل گر.....	۱۳
۲-۳ برنامه ریزی اسپرینت و جلسات اسکرام.....	۱۴
۲-۳-۸ مکان جلسات.....	۱۵
۲-۳-۹ مراحل اتمام هر Product Backlog.....	۱۵
۲-۳-۱۰ اسپرینت اول پیاده سازی پایگاه داده، باتنت و استخراج نمادها.....	۱۶
۲-۳-۱۱ اسپرینت دوم پیاده سازی تحلیل گر و تحلیل متخصصین.....	۱۹
۲-۳-۱۲ اسپرینت سوم: تحلیل و خروجی نهایی.....	۲۱
فصل سوم چالش های تیم توسعه محصول.....	۲۴
ایده اولیه.....	۲۵
۲-۴ چالش های مربوط به پایگاه داده.....	۲۵

۲۵.....	۲-۵- چالش‌های مربوط به صف کار.....
۲۶.....	۲-۶- چالش‌های مربوط به استفاده از پایتون به صورت خام.....
۲۶.....	۲-۷- چالش‌های مربوط به ربات تلگرام.....
۲۷.....	۲-۳-۱۳- مهیا سازی و آماده سازی های اولیه.....
۲۷.....	۲-۳-۱۴- پیاده سازی ربات تلگرامی با استفاده از Telethon.....
۲۹.....	۲-۳-۱۵- چالش‌های پیاده‌سازی ربات تلگرامی.....
۲۹.....	۲-۸- چالش‌های پیاده‌سازی پایگاه‌داده.....
۳۰.....	۲-۹- چالش‌های مربوط به تحلیلگر متن.....
۳۰.....	۲-۳-۱۶- پیاده‌سازی تشخیص سیگنال.....
۳۲.....	۲-۳-۱۷- بهبود و افزایش کارایی الگوریتم.....
۳۳.....	۲-۳-۱۸- چالش‌های پیاده‌سازی تحلیلگر پیام.....
۳۴.....	فصل چهارم چالش‌های تیم تحلیل چالش‌های مربوط به نرم‌افزار Power BI.....
۳۵.....	۲-۱۰- چالش‌های سخت‌افزاری مربوط به تیم تحلیل.....
۳۶.....	۲-۱۱- چالش‌های مربوط به ایجاد حساب.....
۳۶.....	۲-۱۲- چالش‌های مربوط به فرمت ورودی Power BI.....
۳۷.....	۲-۱۳- آموزش و یادگیری Power BI.....
۳۷.....	۲-۳-۱۹- یادگیری با داده آموزشی.....
۳۸.....	۲-۳-۲۰- چالش‌های مربوط به دسته‌بندی مناسب داده‌های خروجی.....
۳۹.....	۲-۳-۲۱- چالش مربوط به اکسل.....
۳۹.....	۲-۳-۲۲- چالش‌های مربوط به رابط کاربری و رنگ‌بندی نمودارها.....
۳۹.....	۲-۳-۲۳- چالش‌های مربوط به مدل‌سازی و ارتباط بین نمودارها.....
۴۰.....	۲-۳-۲۴- چالش‌های مربوط به استفاده از Color Saturation در Tree-map.....
۴۰.....	۲-۱۴- چالش‌های مربوط به تولید کلیدواژه‌ها.....
۴۱.....	۲-۱۵- چالش‌های مربوط به ارزش کلیدواژه‌ها.....
۴۱.....	۲-۱۶- چالش بدست آوردن فرمول تحلیل گر.....
۴۱.....	۲-۱۷- چالش مقایسه نتایج با سایت کدال.....
۴۴.....	فصل پنجم خروجی نرم‌افزار و تیم تحلیل.....
۴۵.....	خروجی نهایی و نمودارهای خروجی.....
۴۵.....	۲-۱۸- نمودارهای گروهی.....
۴۶.....	۲-۳-۲۵- نمودار اول: بررسی مختصر.....
۵۲.....	۲-۳-۲۶- نمودار دوم: بررسی دقیق تر.....
۵۷.....	۲-۱۹- نمودارهای تکی.....
۵۷.....	۲-۳-۲۷- نمودار تکی مربوط به کانال GoldenSignalChannel.....

۵۹	Aryaz_1	نمودار تکی مربوط به کانال	۲-۳-۲۸
۵۹	Bartarinsahm	نمودار تکی مربوط به کانال	۲-۳-۲۹
۶۰	BourseSignalTrade	نمودار تکی مربوط به کانال	۲-۳-۳۰
۶۰	Bozorgane_tse	نمودار تکی مربوط به کانال	۲-۳-۳۱
۶۱	Hamian_bourse	نمودار تکی مربوط به کانال	۲-۳-۳۲
۶۱	Matinmentalofficial	نمودار تکی مربوط به کانال	۲-۳-۳۳
۶۲	Meta_bourse	نمودار تکی مربوط به کانال	۲-۳-۳۴
۶۲	Navasangir100	نمودار تکی مربوط به کانال	۲-۳-۳۵
۶۳	TopSignalExchange	نمودار تکی مربوط به کانال	۲-۳-۳۶
۶۳	Burs_signal	نمودار تکی مربوط به کانال	۲-۳-۳۷

فصل ششم جمع‌بندی و نتیجه‌گیری و پیشنهادات جمع‌بندی و نتیجه‌گیری ۶۵

منابع و مراجع ۶۷

فصل اول

مقدمه

مقدمه

به منظور پایش رویدادهای فضای مجازی و کشف تقلب در سیگنال دهی می‌بایست نیازمندی‌ها را بررسی کنیم و یک پروژه آرمانی طراحی کنیم. در مراحل بعدی با بررسی‌های بیشتر این پروژه آرمانی به پروژه دست یافتنی نزدیک تر خواهد شد تا مرحله‌ای که پیاده سازی آن امکان پذیر باشد.

۱-۲- اهداف پروژه

پروژه متشکل از چهار هدف آرمانی، کلی، ویژه و کاربردی می‌باشد.

۱-۳-۲- اهداف آرمانی

اهداف آرمانی نشان می‌دهد که این پروژه - با بررسی توسعه با زمان هدف اصلی خود - به رسیدن به کدام هدف یا اهداف ایده‌آل کمک می‌کند. در پروژه پایش رویدادهای فضای مجازی هدف ایده‌آل رسیدن به الگوریتم و روش‌هایی که در کمترین زمان و بهترین کیفیت متقلبین را شناسایی و رفتار آنها را تحلیل کند.

۲-۳-۲- هدف کلی

هدف کلی پروژه آن چیزی است که با انجام این پروژه می‌توان به آن دست یافت. در پروژه فعلی هدف کلی یافتن راهی به اندازه کافی مناسب به منظور پایش رویدادهای فضای مجازی می‌باشد.

۲-۳-۳ اهداف ویژه

اهداف ویژه این پروژه شناخت رویدادهای فضای مجازی و تعیین روشی کاربردی برای پایش اطلاعات و تحلیل آنها می‌باشد.

۲-۳-۴ اهداف کاربردی

اهداف کاربردی مشخص می‌کنند که پژوهش در چه زمینه‌هایی و برای چه مصرف‌کنندگانی کاربرد دارد. پروژه مورد بررسی سرمایه‌گذاران را مورد هدف قرار می‌دهد و در زمینه کشف تقلب فعالیت دارد.

۲-۲ راه‌حل ایده‌آل و نیازمندی‌های مرتبط با پایش رویدادهای فضای مجازی به

منظور نظارت بر بازار سرمایه

اگر تصور کنیم که زمانی نامحدود و منابع نامحدود در دست داشته باشیم چگونه به پیاده‌سازی این پروژه می‌پردازیم. در مرحله اول می‌بایست فضاهای مجازی هدف را مورد شناسایی قرار دهیم. در مرحله بعد راه‌حلی ایده‌آل را پیشنهاد می‌دهیم تا تمامی این فضاهای مجازی را پوشش دهد.

۲-۳-۱ فضاهای مجازی هدف

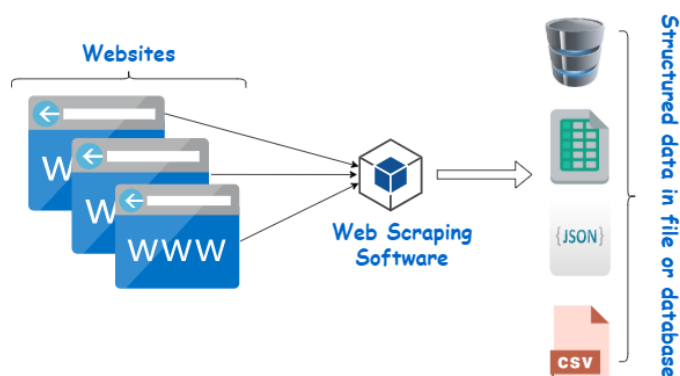
با یک جست‌وجو ساده در گوگل می‌توان سایت‌ها، کانال‌های خبری، گروه‌ها و صفحه‌های زیادی را پیدا کرد که سیگنال دهی انجام می‌دهند. یک پروژه ایده‌آل تمامی این فضاهای مجازی را پوشش می‌دهد.

۲-۳-۲- نیازمندی‌های مرتبط با پایش رویدادهای فضای مجازی

به منظور پایش فضای مجازی می‌توان از برخی مفاهیم برای استخراج و جمع‌آوری داده به نام Web Scraping و Web Crawling استفاده کرد.

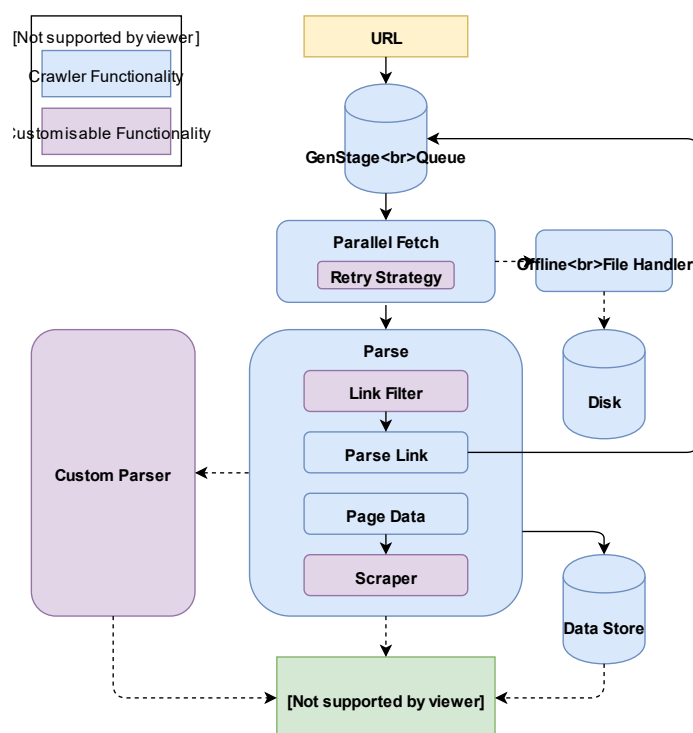
خزنده وب نوعی ربات است که معمولاً توسط موتورهای جستجو مانند گوگل و بینگ اداره می‌شود. هدف آن‌ها ایندکس کردن محتوای وبسایت‌ها در سراسر اینترنت است تا آن وبسایت‌ها بتوانند در نتایج موتورهای جستجو ظاهر شوند. همچنین خزنده وب به دنبال پیدا کردن دامنه‌ها و لینک‌ها است.

Web Scraping خراش دادن داده‌هایی است که برای استخراج داده‌ها از وبسایت‌ها استفاده می‌شود. نرم‌افزار خراش‌دهنده وب ممکن است مستقیماً با استفاده از پروتکل انتقال ابرمتن HTTP یا یک مرورگر وب به وب جهانی دسترسی پیدا کند. درحالی‌که Scraper وب را می‌توان به صورت دستی توسط یک کاربر نرم‌افزار انجام داد، این اصطلاح معمولاً به فرآیندهای خودکاری اشاره دارد که با استفاده از یک ربات یا خزنده وب اجرا می‌شوند. این نوعی کپی است که در آن داده‌های خاصی از وب برای بازیابی یا تجزیه و تحلیل آن‌ها در یک پایگاه داده مرکزی جمع‌آوری و کپی می‌شوند. به صورت اجمالی می‌توان گفت Web Scraping به معنای استخراج داده‌ها از سایت‌های مختلف است. Web scraping شامل دریافت و استخراج یک صفحه وب می‌شود.



عکس ۱-۲

Web Crawling یکی از مهم‌ترین اجزای تشکیل‌دهنده Web Scraping به‌منظور دریافت کردن یک صفحه برای فرآیندهای بعدی است. بعد از مرحله دریافت fetch یک صفحه نوبت استخراج مطالب آن می‌رسد. در این مرحله محتوای موجود در یک صفحه وب تجزیه، جستجو و دوباره پیکربندی می‌شود و سپس داده‌های آن در یک پایگاه داده ذخیره می‌شود. Web Scraper ها معمولاً داده‌هایی را از صفحات وب استخراج می‌کنند در جای دیگر از آن‌ها را به‌عنوان هدفی دیگر مورد استفاده قرار دهند. برای مثال در اینجا ما می‌خواهیم داده‌هایی را به‌منظور پایش اخبار، اطلاعات و سیگنال‌هایی که در بازار سرمایه داده می‌شود استفاده کنیم.



عکس ۲-۲

با توجه به تعاریف بالا، با استفاده از Web Scraping و Web Crawling می‌توان فضای مجازی را پایش کرد. حال اگر که بخواهیم پایش دقیقی از فضای مجازی داشته باشیم تا بتوانیم بر روی بازار سرمایه نظارت و پیش‌بینی داشته باشیم، باید در دو بخش مشغول به جست‌وجو و استخراج داده‌ها شویم.

یکی از داده‌های معتبر به‌منظور استخراج، سایت‌هایی مانند TSETMC است که داده‌های آن را به‌منظور تحلیل می‌توان ذخیره کرد. به‌عنوان Web Scraper می‌تواند در ۳۰ روز متوالی پارامترهایی نظیر خرید، معامله، بازه روز و غیره را پایش کند و اطلاعات آن را در یک پایگاه داده ذخیره کند و سپس در مراحل بعدی از آن ۳۰ روز تحلیلی به دست آید.

بخش دیگر آن مربوط به پایش سیگنال‌ها است. بدین منظور باید Web Scraper داده‌هایی که به نظر می‌رسد تأثیری بر بازار سرمایه دارد را نیز استخراج کند. به‌عنوان مثال باید بتواند در برخی از سایت‌های سیگنال دهی جست‌وجو و سیگنال‌های مرتبط را استخراج کند. همچنین اخبار موجود نیز می‌تواند بر بازار سرمایه نقش داشته باشد به‌عنوان مثال خبر واردات ۷۰۰۰۰ دستگاه خودروی خارجی یا خبرهایی از این‌دست، می‌تواند در بازار سرمایه نقش داشته باشد که از این اخبار به‌منظور تحلیل بازار سرمایه استفاده می‌شود و پیش‌بینی‌هایی در این مورد نیز انجام می‌گیرد. بدین منظور نیاز است تا یک لغت‌نامه تهیه شود تا Web Scraper به آن‌ها حساس باشد تا در صورتی که مشاهده آن اخبار یا لغت‌ها بتواند آن خبر را ذخیره کند.

پس در نتیجه نیازمندی‌های این بخش به‌منظور پایش فضای مجازی این خواهد بود که در ابتدا یک پایگاه داده باید تهیه شود. سپس با استفاده از زبان‌های برنامه‌نویسی مختلف، یک Web Scraper و یک Web Crawler طراحی شود که به‌صورت ترکیب با یکدیگر کار می‌کنند و با توجه به سلیت‌های داده‌شده و همچنین لغت‌نامه داده‌شده، به ترتیب شروع به پایش سایت‌های مرتبط باقیمت‌گذاری بازار سرمایه و همچنین شروع به پایش برخی از سایت‌های خبری و سیگنال دهی می‌کند، داده‌ها را استخراج می‌کند و سپس موارد استخراج‌شده از این سایت‌ها را در داخل پایگاه داده ذخیره می‌کند تا بخش تحلیل، بتواند از این داده‌ها به‌منظور بررسی و تحلیل بازار سرمایه استفاده کند.

۳-۳-۲ تحلیل به منظور نظارت بر بازار سرمایه

به صورت کلی تمامی اخبار، اطلاعات و پیام‌های پایش شده به دو شکل پیام تصویری و متنی تقسیم می‌شوند.

۱-۱-۱-۱ تحلیل پیام تصویری

به منظور تحلیل پیام تصویری می‌بایست از ابزارهایی مانند پردازش تصویر با کمک هوش مصنوعی استفاده کرد. بدین منظور که هر اطلاعات پایش شده تصویری که توسط بخش پایش اطلاعات، در پایگاه داده موجود است را پردازش و تحلیل کنیم. به عنوان مثال پستی در شبکه مجازی اینستاگرام پایش شده است که در آن سیگنال فروش سهم خاص داده شده است. این اطلاعات می‌تواند مفید باشد و به کمک آن‌ها کشف تقلب صورت بگیرد.

۱-۱-۱-۲ تحلیل پیام متنی

به منظور تحلیل پیام متنی می‌بایست از ابزارهایی مانند پردازش زبان طبیعی یا NLP استفاده کنیم. به طور خاص تر از تکنیک NLP Labeling استفاده می‌کنیم. با استفاده از NLP، معامله‌گران و سرمایه‌گذاران می‌توانند تشخیص دهند که آیا مفهوم کلی بیان شده در گزارش سالانه مثبت است، منفی است، یا هیچ کدام و غیره. نتیجه این تحلیل مفهوم کلی می‌تواند برای تصمیم‌گیری معاملات مورد استفاده قرار گیرد. NLP همچنین برای تجزیه و تحلیل رونوشت‌های تماس‌های کسب درآمد برای شناسایی سیگنال‌ها استفاده می‌شود. در ادامه با بررسی سیگنال دهی انجام شده و مقایسه نتایج با وضعیت بازار می‌توانیم تحلیل خوبی را انجام دهیم و افرادی که مستعد به تقلب هستند را شناسایی کنیم.

با استفاده از این تکنیک سایت‌های موجود در ایندکس گردآوری‌شده را می‌توانیم تحلیل و نتیجه‌گیری کنیم. مرحله بعد کشف معنا و مفهوم با decode کردن syntax است که برای این کار به NLP نیاز داریم. به‌عنوان یک تعمیم گسترده، می‌توان گفت که NLP اساساً یعنی گرفتن یک سند غیر شفاف، متشکل از مجموعه منظمی از نمادها که به یک دستور زبان کاملاً تعریف‌شده به نحو مناسب پایبند هستند، و درنهایت استنتاج کردن معنا و مفهوم مرتبط با آن نمادها. اولین گام در اکثر خطوط لوله‌های (Pipeline) پردازش زبان طبیعی در تشخیص جملات، ممکن است فقط شمردن نقطه‌ها، علامت‌های سؤال و علامت تعجب در جمله باشد. این واضح‌ترین ابتکار (heuristic) برای شروع است، اما کاملاً خام است و پتانسیل ایجاد خطای بسیار بالایی را دارد.

توانایی انجام و اجرای یک تشخیص نسبتاً خوب از جملات به‌عنوان بخشی از رویکرد NLP برای استخراج داده‌های بدون ساختار، می‌تواند برخی از قابلیت‌های متن‌کاوی بسیار قدرتمند، مانند تلاش‌های خام اما بسیار منطقی برای خلاصه‌سازی اسناد را برای ما امکان‌پذیر سازد؛ و احتمالات و رویکردهای متعددی وجود دارد. ما می‌بایست شباهت در متن را بیابیم، اطلاعات را استخراج کنیم، رابطه میان معنا و مفهوم متن را بیابیم تا درنهایت خلاصه‌ای منسجم و روان شامل تمام نکات اصلی متن را به دست آوریم.

فصل دوم

مشخص کردن محدوده و ساده سازی پروژه

تفاوت پروژه ایده آل و پروژه هدف

نیازمندی‌های مشخص شده و اطلاعات جمع‌آوری شده مربوط به یک پروژه ایده آل می‌باشد و باوجود زمان محدود و منابع محدود فعلی امکان پذیر نمی‌باشند. در نتیجه در جلسات مطرح شده بین تیم توسعه محصول و سهام‌داران پروژه به ساده‌سازی و مشخص کردن محدوده پروژه پرداختیم.

۱-۲- مشخص کردن Scope و Not Scope

یکی از مهم‌ترین قسمت‌های هر پروژه مشخص کردن محدوده پیاده‌سازی و موارد خارج محدوده هست. در صورتی که این مرحله به خوبی صورت نگیرد پروژه در انتها با شکست روبه‌رو خواهد شد. طی جلسات پی‌درپی تیم توسعه محصول و سهام‌داران پروژه (استاد درس) محدوده پروژه دچار تغییراتی شد. همچنین برخی از این تغییرات در حین پیاده‌سازی رخ داده. به ترتیب به معرفی محدوده و خارج محدوده می‌پردازیم.

۴-۳-۲- محدوده پروژه

محدوده پروژه شامل موارد ذیل می‌باشد:

- ✓ جمع‌آوری اطلاعات و پایش رویدادها مربوط به شبکه اجتماعی تلگرام می‌باشد.
- ✓ تنها محتوای متنی مورد بررسی قرار می‌گیرد.
- ✓ پیام‌های یک الی هفت روز گذشته نیاز به تحلیل دارند.
- ✓ پیام‌های کانال‌های تلگرامی مورد بررسی قرار گیرند.
- ✓ پیام‌های کانال‌های پربازدید مورد بررسی قرار گیرند.
- ✓ تحلیل پیام‌ها با استفاده از نمادها صورت گیرد.
- ✓ تحلیل پیام‌ها با استفاده از تعداد بازدید و ارسال صورت گیرد.
- ✓ خروجی تحلیل گر مشخص‌کننده سیگنال دهی و یا عدم سیگنال دهی باشد.

- ✓ خروجی تحلیل گر مشخص کننده سیگنال خرید یا فروش باشد.
- ✓ از زبان برنامه نویسی پایتون در پیاده سازی استفاده شود.
- ✓ از فریمورک جانگو در پیاده سازی استفاده شود.
- ✓ نتایج تحلیل گر در قالب نمودار نمایش داده شود.

۵-۳-۲- خارج محدوده پروژه

محدوده پروژه شامل موارد ذیل نمی باشد.

- ✓ جمع آوری اطلاعات از شبکه های اجتماعی اینستاگرام، توییتر، واتس آپ، رد دیت و غیره نمی باشد.
- ✓ محتوای تصویری مورد بررسی قرار نمی گیرد.
- ✓ پیام های قدیمی تر مورد بررسی قرار نمی گیرد.
- ✓ برای تحلیل از هوش مصنوعی کمک گرفته نمی شود.
- ✓ در تحلیل اطلاعات فرستنده پیام استخراج نمی شود.
- ✓ از زبان های برنامه نویسی جاوا، سی، سی پلاس پلاس، سی شارپ، گولنگ و غیره استفاده نمی شود.
- ✓ خروجی تحلیل گر کشف تقلب انجام نمی دهد.

۲-۲- معرفی نیازمندی های شناسایی شده در قالب Product Backlog

کاربرد متدولوژی های چابک در پروژه هایی که ایده اصلی مشخص ولی جزئیات نامشخص هستند بسیار پرکاربرد است. ما نیز به همین دلیل از متدولوژی های چابک استفاده می کنیم.

برای این منظور ابتدا خواسته ها را مشخص می کنیم. خواسته های واقعی تر و قابل پیاده سازی در شرایط فعلی پروژه را انتخاب می کنیم و آن ها را خواسته می نامیم. در نهایت خواسته هایی که در قالب بازه زمانی

دو الی چهار هفته قابل پیاده سازی هستند را انتخاب می کنیم و در قالب نیازها مطرح می کنیم. در نهایت نیازها را به Product Backlog اضافه می کنیم و پیاده سازی می کنیم.

بعد از جلسات مکرر با اعضا تیم توسعه و سهام داران Product Backlog پیاده سازی شد که به توضیح آن می پردازیم. ابزار مورد استفاده ما نرم افزار تحت وب ماکروسافت Azure هست. در مرحله اول Epic ها را مشخص می کنیم. هر Epic به تعداد زیادی Feature شکسته می شوند. Feature ها نیز به Product Backlog شکسته می شوند که افزونه های اصلی پروژه می باشند. باید در تعریف Product Backlog حواسمان باشد که تمامی موارد در دوره دو الی چهار هفته ای قابل پیاده سازی باشند. در غیر این صورت به موارد کوچک تر تقسیم شوند.

۶-۳-۲- Product Backlog مربوط به تیم توسعه محصول

تیم توسعه محصول در پیاده سازی نرم افزار پروژه نقش دارد. این پروژه به صورت یک نرم افزار تحت وب پیاده سازی شده است و از فریمورک جنگو استفاده می کند. استفاده از این فریمورک بسیار مفید است چراکه از کتابخانه های بسیار گسترده و مفید زبان پایتون استفاده می کند.

PBI های مربوط به تیم توسعه محصول (عکس ۲-۳) در سه دسته باتنت، پایگاه داده و تحلیل گر تقسیم می شود. ماژول باتنت با کمک یک کلید و اتصال به تلگرام پیام های کانال تلگرامی از پیش مشخص شده را که در ۲۴ ساعت گذشته پیامی را به اشتراک گذاشته باشند می خواند و در پایگاه داده قرار می دهد.

ماژول پایگاه داده مربوط به تمامی کاربردهای پایگاه داده و پیاده سازی توابع پایگاه داده می باشد. در این پروژه از پایگاه داده Sqlite3 استفاده شده است.

ماژول تحلیل گر با کمک نمادهایی که در اختیار دارد اطلاعات را از پایگاه داده می‌خولند و پیام‌ها را تحلیل می‌کند. دوباره تحلیل‌های صورت گرفته را به پایگاه داده برمی‌گرداند تا تیم تحلیل گر از آن‌ها در ساختن نمودارها استفاده کنند.

Order	Work Item Type	Title	State	Effort	Busin...	Value Area	Tags
1	Epic	Botnet	New			Business	
	Feature	Web Crawler	New			Business	
	Product Backlog	Web Crawler API	New			Business	
	Product Backlog	Telegram bot	New			Business	
	Feature	Web Scraper	New			Business	
	Product Backlog	Web Scraper API	New			Business	
2	Epic	Database	New			Business	
	Feature	Database Initialization	New			Business	
	Product Backlog	SQL Database	New			Business	
	Product Backlog	Database Connectivity API	New			Business	
	Product Backlog	Database Session Manager	New			Business	
	Feature	Database Functionality	New			Business	
	Product Backlog	Database ORM	New			Business	
	Product Backlog	Database Entities CRUD	New			Business	
3	Epic	Data Analysis	New			Business	
	Feature	Data Analysis	New			Business	
	Product Backlog	Keywords List	New			Business	
	Product Backlog	Keywords Values	New			Business	
	Product Backlog	Analysis Formula	New			Business	
	Product Backlog	Symbols	New			Business	
	Feature	Data Reader	New			Business	
	Product Backlog	Database Read Operation	New			Business	
	Product Backlog	Database Modification Operation	New			Business	

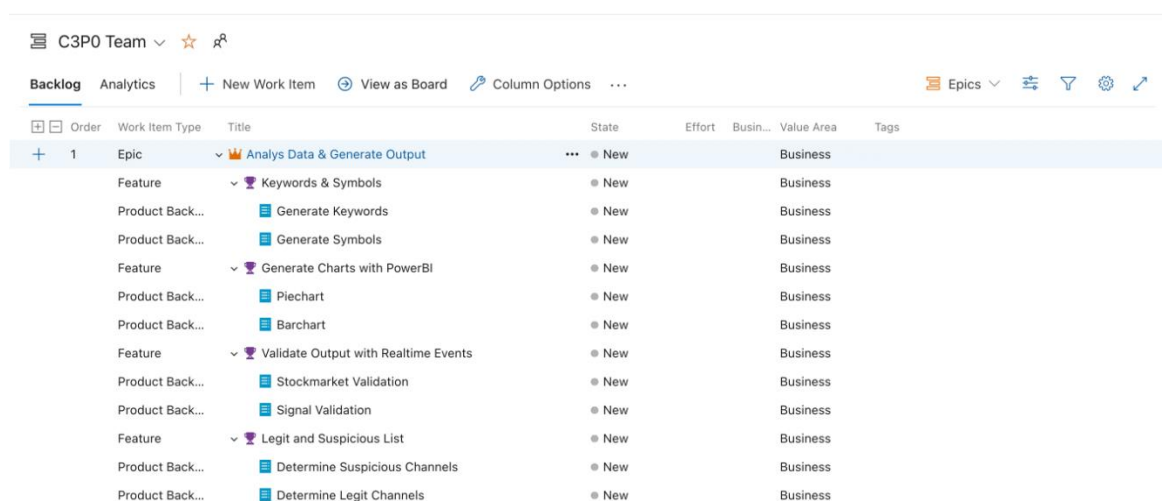
عکس ۲-۳

۷-۳-۲ تیم تحلیل گر

تیم تحلیل گر با استفاده از خروجی تحلیل شده تیم توسعه محصول نمودارهایی را تشکیلی داده و آن‌ها را تحلیل می‌کنند. در صورتی که در تحلیل صورت گرفته رفتار مشکوکی مشاهده شود و یا تقلبی کشف شود این تقلب گزارش داده خواهد شد.

تیم تحلیل گر با دریافت اطلاعات تحلیل شده توسط تیم توسعه و تطابق آن با رویدادهای بورس اوراق بهادار کشف تقلب را انجام می دهند. این عملیات به این صورت است اگر کانال تلگرامی وجود داشته باشد که بازدید بالایی دارد و در اکثر موارد سیگنال دهی درستی انجام می دهد ممکن است اطلاعات نهانی - هرگونه اطلاعات افشاء نشده برای عموم که به طور مستقیم و یا غیرمستقیم به اوراق بهادار، معاملات یا ناشر آن مربوط می شود و در صورت انتشار بر قیمت و یا تصمیم سرمایه گذاران برای معامله اوراق بهادار مربوط تأثیر می گذارد - داشته باشد. این صرفاً حدس است و نیاز به بررسی های عمیق تر دارد.

PBI های تیم تحلیل (عکس ۲-۴) تنها در آنالیز کردن و تولید خروجی خلاصه می شود. این تحلیل با کمک متخصصین صورت می گیرد و ربات ها.



Order	Work Item Type	Title	State	Effort	Busin...	Value Area	Tags
1	Epic	Analys Data & Generate Output	New			Business	
	Feature	Keywords & Symbols	New			Business	
	Product Backlog Item	Generate Keywords	New			Business	
	Product Backlog Item	Generate Symbols	New			Business	
	Feature	Generate Charts with PowerBI	New			Business	
	Product Backlog Item	Piechart	New			Business	
	Product Backlog Item	Barchart	New			Business	
	Feature	Validate Output with Realtime Events	New			Business	
	Product Backlog Item	Stockmarket Validation	New			Business	
	Product Backlog Item	Signal Validation	New			Business	
	Feature	Legit and Suspicious List	New			Business	
	Product Backlog Item	Determine Suspicious Channels	New			Business	
	Product Backlog Item	Determine Legit Channels	New			Business	

عکس ۲-۴

۲-۳- برنامه ریزی اسپرینت و جلسات اسکرام

مهم ترین رکن یک تیم موفق توانای برنامه ریزی اهدافش در مدت زمانی مشخص است. برای پیاده سازی پروژه سه اسپرینت خواهیم داشت که هر اسپرینت متشکل از ۱۲ روز کاری می باشد. با احتساب روزهای تعطیل بین اسپرینت ها هر اسپرینت ۲ هفته طول می کشد.

۸-۳-۲- مکان جلسات

طبق توافق اعضا تیم تمامی جلسات در سایت دانشکده مهندسی کامپیوتر دانشگاه امیرکبیر صورت گرفته‌اند. مکان جلسات همیشه ثابت بوده و گوشه‌ای از سایت می‌باشد. زمان اجرای جلسات نیز در تمامی مراحل یکسان بوده و به هیچ وجه تغییر نکرده‌اند.

جلسات Daily به مدت ۱۵ دقیقه روزهای زوج رأس ساعت ۹ صبح اجرا شدند. جلسات Review به مدت دو ساعت در انتهای هر اسپرینت ساعت ۹ الی ۱۱ برگزار شدند. جلسات Planning بلافاصله بعد از جلسه Review به مدت یک ساعت و ۲۰ دقیقه برگزار شدند. با اتمام جلسات فوق جلسه Retrospective را خواهیم داشت که مدت زمان ۳۰ دقیقه طول می‌کشند. مکان تمامی جلسات سایت دانشکده می‌باشد.

۹-۳-۲- مراحل اتمام هر Product Backlog

هر PBI که در اسپرینت آورده می‌شود توسط تیم توسعه محصول به تعدادی Task شکسته می‌شود. هر Task مراحل ذیل را طی می‌کند:

- ✓ To Do
- ✓ In Progress
- ✓ Technical Review
- ✓ Test
- ✓ Done

هر وظیفه در ابتدا در وضعیت To Do قرار دارد و بعد از اینکه توسعه‌دهنده شروع به کار کردن می‌کند به وضعیت In Progress وارد می‌شود. بعد از اینکه توسعه‌دهنده وظیفه را به اتمام برساند وضعیت آن را به Technical Review تغییر می‌دهد تا مدیر پروژه وظیفه را بررسی کند و در صورتی که با

Definition Of Done تطابق نداشته باشد به توسعه‌دهنده اطلاع دهد و وضعیتش را به To Do تغییر دهد.

در صورتی که تأیید مدیر پروژه را داشته باشد به وضعیت Test وارد می‌شود تا توسط تیم تحلیل آزموده شود. با اتمام آزمایش‌ها تیم تحلیل به وضعیت اتمام یا Done وارد می‌شود و بر افزونه‌ها پروژه اضافه می‌شود. در نهایت توسعه‌دهنده سراغ وظیفه دیگری می‌رود ولی آخر.

۱۰-۳-۲ - اسپرینت اول پیاده‌سازی پایگاه داده، بات‌نت و استخراج نمادها.

در اسپرینت اول بیشتر کارها مربوط به تیم توسعه محصول می‌باشد. در این اسپرینت PBI‌های انتخاب و پیاده‌سازی شدند. همچنین تیم تحلیل به دلیل اینکه تیم توسعه برای استخراج و تحلیل اولیه پیام‌ها نیازمند نمادها، کلیدواژه‌ها و ارزش کلیدواژه‌ها بودند به جمع‌آوری این موارد پرداختند.

مواردی که در این اسپرینت به آن‌ها پرداخته شد و وظایفی که به آن‌ها شکسته شدند در عکس ۱-۳ قابل مشاهده است.

The screenshot shows a Jira backlog for the 'C3P0 Team'. The main table lists tasks with columns for Order, Title, State, Assigned To, and Rem... (Remaining). The tasks are grouped by sprint. The 'Planning' sidebar on the right shows a list of sprints (Sprint 1 to Sprint 5) with their planned effort and current status. Sprint 1 is the current sprint.

Order	Title	State	Assigned To	Rem...
1	Telegram bot	New	kamyar.hass...	
2	SQL Database	New	bardia ardak...	
3	Database ORM	New	bardia ardak...	
4	Database Entities CRUD	In Progress	bardia ardak...	
5	Database Connectivity API	New	bardia ardak...	
6	Web Crawler API	New	kamyar.hass...	
7	Web Scraper API	New	kamyar.hass...	
8	Generate Keywords	New	moujanmirjali...	
9	Generate Symbols	New	moujanmirjali...	

The 'Planning' sidebar shows the following sprints:

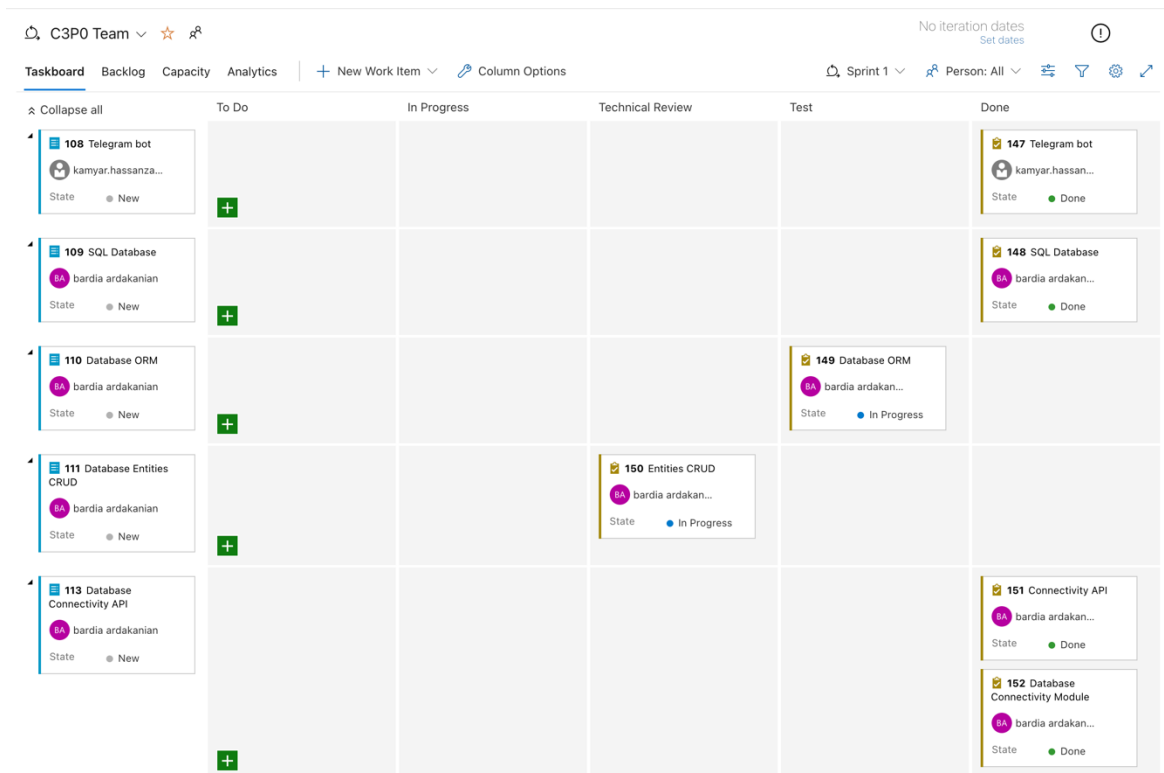
- Sprint 1**: Planned Effort: - (Current)
- Sprint 2**: Planned Effort: -
- Sprint 3**: Planned Effort: -
- Sprint 4**: No work scheduled yet
- Sprint 5**: No work scheduled yet

عکس ۳-۱

در بین وظایف، طبق بررسی‌های صورت گرفته وظیفه Web Scraper API خارج از محدوده پروژه بود و حذف گردید. (عکس ۳-۲) همچنین دو وظیفه در وضعیت تست و بررسی تکنیکال مانده بودند که به اسپرینت بعدی فرستاده شدند. بقیه وظایف در این اسپرینت به پایان رسیدند.

در جلسه Review به بررسی عملکرد تیم توسعه پرداختیم. در این بررسی به نتیجه رسیدیم که بیشتر از ظرفیت تیم وظیفه اختصاص داده‌ایم و این باعث کاهش بازدهی افراد شده است. همچنین تخمین زمان اختصاص یافته به وظایف غیرواقعی و از مقدار واقعی کمتر بوده‌اند. در اسپرینت بعدی این موارد اصلاح خواهند شد.

نتیجه جلسه Planning اسپرینت دوم در بخش بعدی اعلام شده است. در جلسه Retrospective عوامل موفقیت شناخته شدند و در اسپرینت بعدی به کار گرفته شدند. همچنین راه‌حل‌های جدیدی پیشنهاد داده شدند.



عکس ۲-۳

۳-۱-۱-۱ عوامل پیروزی در اسپرینت اول: جلسات روزانه

جلسات روزانه بیشتر از آنچه انتظار داشتیم نیاز به جلسات دیگر را کاهش داد. در نتیجه این بررسی تصمیم گرفته شد تا جلسات روزانه بجای روزهای زوج به هرروز تغییر پیدا کند.

۴-۱-۱-۱ عوامل پیروزی در اسپرینت اول: کار کردن در محیط اشتراکی

در اسپرینت اول یکی از عوامل موفقیت کار کردن در کنار یکدیگر بود که باعث بالا رفتن بازدهی افراد تیم شد. تلاش کردیم تا در اسپرینت‌های بعدی نیز از این کیفیت استفاده کنیم. به همین جهت ساعاتی در روز (معمولا ۱۵ الی ۱۸) در سایت دانشکده و یا سالن مطالعه دانشکده دورهم جمع شدیم و به پیاده‌سازی پرداختیم.

۱۱-۳-۲ - اسپرینت دوم پیاده سازی تحلیل گر و تحلیل متخصصین

در اسپرینت دوم با بهره برداری از تجربیات اسپرینت اول، برنامه ای منطقی تر برای اسپرینت دوم در نظر گرفتیم. همچنین وارد فاز دوم پیاده سازی یعنی تحلیل شدیم. در این اسپرینت ابتدا تابعی پیاده سازی کردیم که در ادامه با کارکرد آن به طور دقیق آشنا خواهیم شد. این تابع با کمک ارزش نمادها امتیازی به هر پیام داده و سیگنال بودن/نبودن و سیگنال خرید/فروش بودن را تشخیص می دهد.

برای بدست آوردن فرمول ذکر شده نرمال سازی هایی صورت گرفت که به بهترین فرمول ممکن دست یابیم. مواردی که در این اسپرینت به آن ها پرداخته شد و وظایفی که به آن ها شکسته شدند در عکس ۳-۳ قابل مشاهده است.

The screenshot displays a Jira backlog for the C3P0 Team. The main table lists tasks with columns: Order, Title, State, Assigned To, and Rem... The tasks are grouped under the 'Database ORM' category. The right sidebar shows the 'Planning' section with a list of sprints (Sprint 1 to Sprint 6) and their planned effort.

Order	Title	State	Assigned To	Rem...
1	Database ORM	New	bardia ardak...	
2	Database Entities CRUD	New	bardia ardak...	
3	Entities CRUD	New	bardia ardak...	
4	Keywords List	New	kamyar.hass...	
5	Read Keywords from DB	Done	kamyar.hass...	
6	Keywords Values	New	kamyar.hass...	
7	Assign Values to Keywords from DB	Done	kamyar.hass...	
8	Analysis Formula	New	kamyar.hass...	
9	Make Analysis Formula	Done	kamyar.hass...	
10	Analysis Formula	Done	kamyar.hass...	
11	Normalize Formula	Done	kamyar.hass...	
12	Database Read Operation	New	bardia ardak...	
13	Database Read	Done	bardia ardak...	
14	Database Modification Operation	New	bardia ardak...	
15	Database Update	Done	bardia ardak...	
16	Symbols	New	kamyar.hass...	
17	Extract Symbols from Messages	Done	kamyar.hass...	
18	Update Database	Done	kamyar.hass...	
19	Determine Suspicious Channels	New	bardia ardak...	
20	Detect Suspicious Activities	In Progress	bardia ardak...	
21	Detect Suspicious Channels	In Progress	bardia ardak...	
22	Determine Legit Channels	New	kamyar.hass...	
23	Detect Legit Activities	In Progress	kamyar.hass...	
24	Detect Legit Channels	In Progress	kamyar.hass...	

The right sidebar shows the 'Planning' section with a list of sprints (Sprint 1 to Sprint 6) and their planned effort.

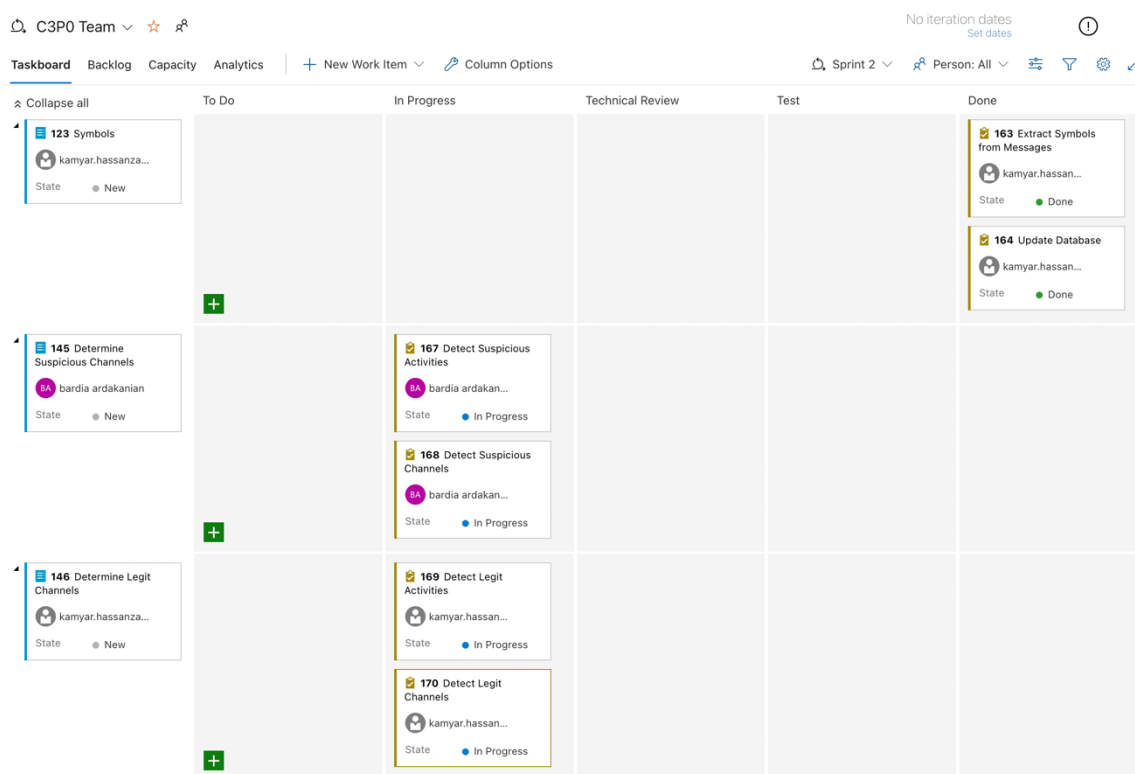
Sprint	Planned Effort
Sprint 1	Planned Effort: -
Sprint 2	Planned Effort: -
Sprint 3	Planned Effort: -
Sprint 4	Planned Effort: -
Sprint 5	Planned Effort: -
Sprint 6	Planned Effort: -

عکس ۳-۳

با توجه به پیچیدگی تشخیص مشکوک و غیر مشکوک بودن کانال‌ها این وظیفه تکمیل نشد (عکس ۳-۴) و به انتها نرسید. بنابراین به اسپرینت بعد انتقال داده شد تا در اسپرینت بعدی مورد بررسی قرار گیرد.

در جلسه Review به بررسی عملکرد تیم توسعه و تحلیل پرداختیم. در این بررسی به نتیجه رسیدیم که اختصاص وظایف با کیفیت بهتری انجام شده و همین روند باید ادامه یابد.

نتیجه جلسه Planning اسپرینت سوم در بخش بعدی اعلام شده است. در جلسه Retrospective به بررسی عوامل موفقیت پرداختیم و به نتیجه رسیدیم که عملکرد خوبی داشتیم و عوامل موفقیت شناخته شده در اسپرینت اول را باید ادامه دهیم. همچنین به پیشنهاد یکی از افراد تیم توسعه برای توسعه نرم افزار روزهایی را به باغ کتاب رفتیم تا در کنار هم و سکوت بیشتر به فعالیت بپردازیم که نتیجه بسیار مطلوبی داشت.



عکس ۳-۴

۱۲-۳-۲ اسپرینت سوم: تحلیل و خروجی نهایی

طبق جدول زمان بندی اسپرینت سوم اسپرینت آخر پروژه می باشد که در آن می بایست نرم افزاری کاربردی داشته باشیم و بتوانیم داده هایی مفید را از آن استخراج کنیم. همچنین در این اسپرینت می بایست داده های استخراج شده و نمودارهای ایجاد شده را تحلیل کنیم و با اخبار بورس اوراق بهادار مقایسه کنیم. در صورتی که رفتاری مشکوک مشاهده کنیم گزارش می دهیم. به عنوان مثال اگر کانالی به صورت عمومی سیگنال دهی درست انجام دهد و بعضی از این سیگنال دهی ها برخلاف شواهد فعلی بازار بورس اوراق بهادار باشد به احتمال زیاد اطلاعات نهانی دارد و از این اطلاعات سوء استفاده می کند. این تخلفات باعث از بین رفتن اعتماد میان سهام داران می شود و باید با افرادی از این دست برخورد شود.

برای محقق شدن این امر چندین وظیفه تعریف و میان اعضا تیم تقسیم شد. این وظایف در عکس

۵-۳ قابل مشاهده هستند.

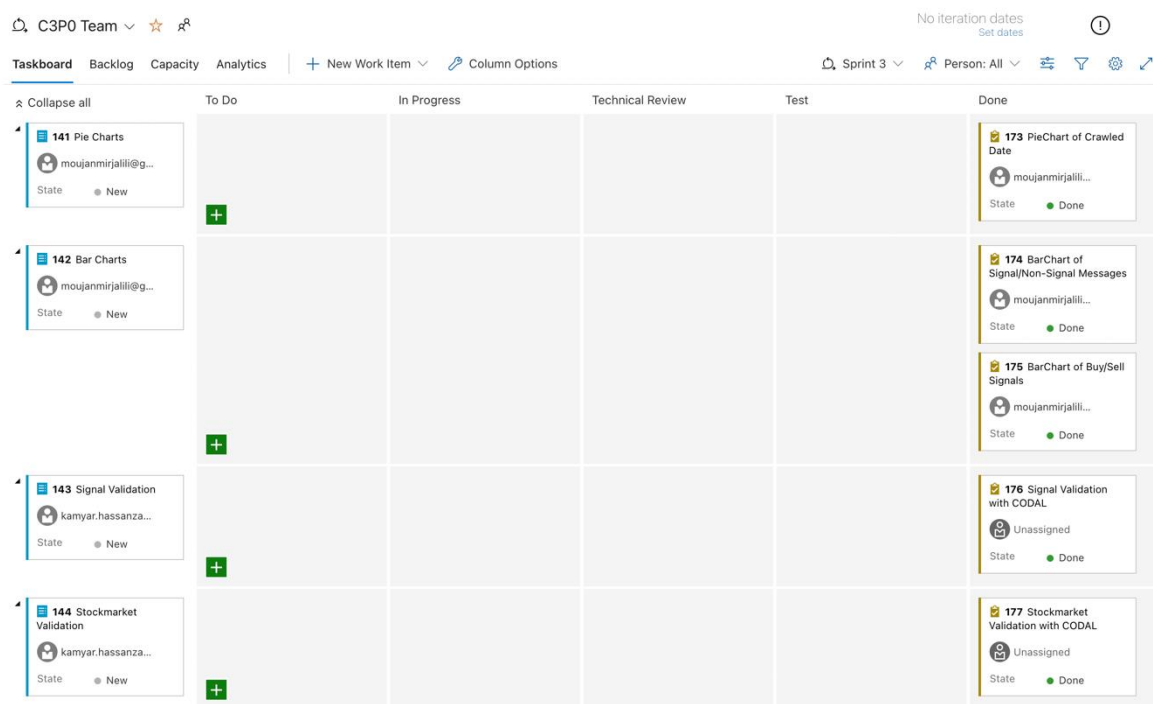
The screenshot displays a Jira backlog for the 'C3P0 Team'. The main table lists tasks under the 'Pie Charts' category. The tasks are as follows:

Order	Title	State	Assigned To
1	PieChart of Crawled Date	Done	moujanmirjali...
2	Bar Charts	New	moujanmirjali...
	BarChart of Signal/Non-Signal Messages	Done	moujanmirjali...
	BarChart of Buy/Sell Signals	Done	moujanmirjali...
3	Signal Validation	New	kamyar.hass...
	Signal Validation with CODAL	Done	
4	Stockmarket Validation	New	kamyar.hass...
	Stockmarket Validation with CODAL	Done	
5	Determine Suspicious Channels	New	bardia.ardak...
	Detect Suspicious Activities	Done	bardia.ardak...
	Detect Suspicious Channels	Done	bardia.ardak...
6	Determine Legit Channels	New	kamyar.hass...
	Detect Legit Activities	Done	kamyar.hass...
	Detect Legit Channels	Done	kamyar.hass...

On the right side, the 'Planning' panel shows a list of sprints:

- Sprint 1**: Planned Effort: -, Current
- Sprint 2**: Planned Effort: -
- Sprint 3**: Planned Effort: -
- Sprint 4**: No work scheduled yet

عکس ۳-۵



عکس ۳-۶

با نتایج حاصل شده توسط نمودارها و طبق این اطلاعات با سایت CODAL توانستیم کانالی را به عنوان کانال مشکوک نشانه گذاری کنیم. البته این بررسی باید عمیق تر باشد و نمی توان نتیجه قطعی بدست آورد.

در جلسه Review به بررسی عملکرد تیم توسعه و تحلیل پرداختیم. در این بررسی به نتیجه رسیدیم که عملکرد ضعیف تری داشتیم و بازدهی تیم توسعه و تحلیل نزدیک به بیست درصد کاهش داشته است. البته مواردی همچون بازه امتحانات در این موضوع تأثیر بسیاری داشته است.

در جلسه Retrospective به بررسی عوامل موفقیت پرداختیم و به نتیجه رسیدیم که برای بهبود بازدهی می بایست ساعاتی مشخص و روتین را در روز به پیاده سازی پروژه اختصاص دهیم. ساعات شناور بسیار تأثیر نامطلوبی داشته اند.

فصل سوم

چالش‌های تیم توسعه محصول

ایده اولیه

در ابتدا پیاده‌سازی تیم توسعه قصد داشت تا با کمک خزنده وب اطلاعات را از سایت‌های مختلف من جمله کانال‌های تلگرامی جمع‌آوری کند و در پایگاه داده MongoDB ذخیره کند. در محله‌های بعدی این اطلاعات را از پایگاه داده برداشته، گزینش کند و منتخبین را در پایگاه داده Sqlite3 ذخیره کند. این کاوش اطلاعات و استخراج از پایگاه داده اول به صورت موازی و با کمک یک صف کاری (Job Queue) پیاده‌سازی شود.

همچنین در ابتدا خبری از فریمورک جنگو نبود و به پیاده‌سازی به صورت خالص با پایتون بسنده کرده بودیم. در ادامه کار با چالش‌هایی روبه‌رو شدیم که نیازمند تغییر در ساختار پیاده‌سازی پروژه بود.

۴-۲- چالش‌های مربوط به پایگاه داده

به علت دوباره کاری شدن و استفاده نکردن بهینه از فضا و منابع بجای استفاده از دو پایگاه داده تنها از یک پایگاه داده استفاده کنیم. در این حالت مشکلاتی من جمله موازی‌سازی عملیات استخراج پیام و تحلیل نخواهیم داشت چراکه کماکان می‌توانند به صورت موازی از هم فعالیت کنند.

۵-۲- چالش‌های مربوط به صف کار

پیاده‌سازی این بخش خارج از محدوده پروژه بود و وقت و انرژی زیادی از تیم توسعه محصول می‌گرفت. در نتیجه پروژه به صورت Real-Time پیاده‌سازی شد و صف کار از آن حذف گردید.

۶-۲- چالش‌های مربوط به استفاده از پایتون به صورت خام

یکی از بزرگ‌ترین بخش‌های پروژه مربوط به پایگاه داده و ارتباط بین جداول مختلف بود. به منظور قابل‌گسترش بودن پروژه در آینده از فریمورک جنگو استفاده شد چون هم از ORM بسیار قدرتمندی برخوردار هست و هم دارای Generic Relation بین جداول می‌باشد.

با توجه به اینکه قصد داریم پروژه در آینده گسترش یابد از مفهوم رابطه کلی استفاده می‌کنیم به این صورت که با اضافه شدن انواع مختلف محتوا متنی (پیام کانال، اطلاعات صفحه وب و غیره) و یا محتوا تصویری بتوانیم به راحتی ارتباط جدیدی بین جداول داشته باشیم تا نیازی به پیاده‌سازی مفاهیم جدید و پیچیده‌ای نباشد. این امر باعث گسترش‌پذیری و آسودگی برنامه پیاده‌سازی شده می‌شود. علاوه بر موارد ذکر شده فریمورک جنگو توابع و ماژول‌های بسیار زیادی دارد که کار پیاده‌سازی را بسیار آسان‌تر می‌کند و باعث می‌شود تیم توسعه انرژی خود را برای مباحث پیچیده‌تر بگذارند.

۷-۲- چالش‌های مربوط به ربات تلگرام

در این بخش ابتدا به منظور پیاده‌سازی ربات بررسی‌های لازم انجام شد. نتایج به این صورت بود که با استفاده از زبان پایتون رباتی به منظور جمع‌آوری داده‌ها نوشته شود. کتابخانه‌ی python-telegram-bot به منظور پیاده‌سازی این ربات استفاده شد. به هنگام پیاده‌سازی طبق بررسی‌های مجدد انجام شده، بیان شده است که ربات‌های نوشته شده در تلگرام نمی‌توانند عضو کانال‌های عمومی تلگرام شوند. در نتیجه به منظور برطرف کردن این مشکل به جای استفاده از ربات، از یک واسطی به نام Telethon که در زبان پایتون پیاده‌سازی شده است استفاده شد.

کتابخانه‌ی Telethon مستقیماً به اکانت تلگرامی کاربران متصل می‌شود و سپس با استفاده از جست‌وجو و پیدا کردن کانال می‌توان بر روی بخش‌های مختلف پیام‌ها پیمایش کند و داده‌ها را استخراج

کند. این کتابخانه توابع بسیار زیادی را در اختیار توسعه‌دهندگان قرار می‌دهد که پیاده‌سازی را ساده می‌کند.

۲-۳-۱۳- مهیاسازی و آماده‌سازی‌های اولیه

در ابتدا نیاز است که اکانتی برای ربات درست شود. سپس کتابخانه‌ی Telethon با استفاده از رمز مربوط به نرم‌افزار تحت وب به اکانت مربوطه متصل می‌شود و اعتبار سنجی انجام می‌دهد. این اعتبار سنجی به‌منظور امنیت بیشتر دومارحله‌ای بوده. بعد از اعتبار سنجی می‌توانید مانند هر کاربری کانال‌های تلگرامی را پیدا کرده و از محتوای آنان استفاده کنید.

البته با توجه به تحریم‌ها و فیلترینگ هوشمند در کشور نیاز به یک فیلترشکن داریم تا بدون مشکل متصل شویم. در صورتی که فیلترشکن موجود نباشد بعد از چند ثانیه تمامی بسته‌های ارسال شده Timeout می‌شوند و ارتباط بسته می‌شود.

۲-۳-۱۴- پیاده‌سازی ربات تلگرامی با استفاده از Telethon

در ابتدا با استفاده از app_id و hash مربوطه که از سایت تلگرام دریافت شده است به اکانت ربات با استفاده از دستور connect() متصل می‌شویم. توجه کنید بیش از یک اتصال می‌تواند برقرار باشد و برای هر اتصال نشست جداگانه ایجاد می‌شود.

در مرحله بعدی لیست کانال‌های انتخاب‌شده را از پایگاه داده می‌خواند و با پیمایش بر روی این لیست کانالی را که اطلاعات از آن جمع‌آوری می‌شود، قابل دسترسی است. بعد از دسترسی به کانال‌های تلگرامی شاخص‌های منتخب پیام‌ها استخراج شده و در پایگاه داده وارد می‌شوند. این استخراج تنها پیام‌های محدودی را برمی‌دارد و این محدودیت توسط توسعه‌دهنده مشخص می‌شود. شاخص‌های انتخاب‌شده برای پیام‌ها به صورت زیر می‌باشند:

- ✓ ID: شماره‌ی اختصاص‌یافته‌ی یکتا به پیغام در کانال
 - ✓ Text: متن پیغام نوشته‌شده
 - ✓ Pub_date: تاریخ دقیق پیغام منتشرشده
 - ✓ Forwards: تعداد forward های این پیغام در این متغیر ذخیره می‌شود.
 - ✓ Name: نام کانالی که پیام از آن دریافت شده است.
 - ✓ From_id: شماره آی دی کاربری که این پیام را ارسال کرده است.
 - ✓ Forward_from: شماره آی دی کاربری که این پیام از آن ارسال شده است.
 - ✓ Edit_date: در صورت ویرایش، تاریخ ویرایش پیام ذخیره می‌شود.
 - ✓ Edit_hide: در صورتی که ادمین کانال تغییری را اعمال کرده باشد و با استفاده از گزینه‌های تلگرام، پیغام edited را از پست حذف کرده باشد، مقدار آن True خواهد شد.
 - ✓ Is_reply: آیا پیغام مربوطه reply به پیغام دیگری می‌باشد یا خیر.
 - ✓ Num_replies: تعداد پیام‌هایی که به این پیام reply شده‌اند.
 - ✓ Reply_to: شماره آی دی پیامی که این پیام به آن reply کرده است.
 - ✓ Views: بیانگر تعداد دفعاتی که در لحظه دریافت پیام توسط بات، دیده‌شده است.
- برنامه به‌صورت کلی از ساعت ۸ صبح روزهای شنبه تا پنج‌شنبه به‌غیر از ایام تعطیل تا ساعت ۱ اجرا می‌شود. طبق بررسی‌ها و پایش های کانال‌های استفاده‌شده به نظر می‌رسد کانال‌های بورسی در بین این ساعات بیشترین فعالیت خود را دارند.
- روند اجرای ربات بدین‌صورت می‌باشد که در ابتدا تابعی برای استفاده سرور از آن نوشته‌شده است. این تابع سه ورودی می‌گیرد که به ترتیب ابتدا نام کانال، تعداد پیام‌ها و تاریخ پیام‌ها را دریافت می‌کند. سپس به دنبال پیام‌هایی از یک کانال می‌رود که ابتدا تاریخ انتشارشان از تاریخی ورودی داده‌شده است

بزرگ‌تر باشد و سپس با استفاده از یک حلقه for به تعداد پیام‌های ورودی داده‌شده عملیات دریافت و ذخیره‌سازی پیام‌ها را انجام می‌دهد. بعد از دریافت پیام‌ها آن‌ها را در پایگاه داده ذخیره می‌کنیم.

۲-۳-۱۵ - چالش‌های پیاده‌سازی ربات تلگرامی

در ابتدا به منظور استفاده از ربات و دسترسی داشتن به تلگرام به وسیله پایتون، نیاز است تا در خود سایت تلگرام کاربری که می‌خواهد به تلگرام متصل شود، فرآیند تصدیق را طی کند. بعد از طی کردن این مراحل، سایت تلگرام کدی را تحت عنوان api_id و api_hash نمایش می‌دهد که کتابخانه‌ی Telethon از این دو رمز برای ورود به تلگرام استفاده می‌کند. فرآیند طی شده نیاز به فیلترشکن دارد. همچنین در هنگام اجرای کد نیز فیلترشکن باید متصل باشد. (دقت شود استفاده از پروکسی در اجرای کد بررسی شد، اما در هنگام اجرای اولیه کد خطا داده شد)

چالش بعدی در ذخیره‌سازی و مقایسه‌ی زمان‌ها و فیلد Date در پیام‌های خود تلگرام می‌باشد. به دلیل متفاوت بودن time zone مقایسه به راحتی انجام نمی‌شود. ابتدا باید time zone پیام‌ها را با تاریخ مطرح‌شده یکسان‌سازی کرد و سپس عملیات مقایسه و ذخیره‌سازی را انجام داد.

چالش مهم بعدی هنگام ذخیره‌سازی پیام‌ها مشاهده شد. مشکل بدین صورت بود که در ابتدا پایگاه داده با استفاده از SQL به صورت کدهای پایتونی خام، پیاده‌سازی شده بود که به هنگام ذخیره‌سازی داده‌هایی که در آن‌ها حروف ' و ' وجود دارد، خطابه وجود می‌آید. (این خطا ضمنی بوده و مشکل امنیتی تزریق کد را به همراه خواهد داشت) به همین جهت پیاده‌سازی دیتابیس با استفاده از جنگو ابتدا بررسی و سپس اجرا شد.

۲-۸ - چالش‌های پیاده‌سازی پایگاه داده

در بخش پیاده‌سازی پایگاه داده با توجه به اینکه متن‌هایی که از کانال‌های موردنظر به دست می‌آید

همگی خام هستند و نیاز به پردازش یا عملیات Text Filtering دارند، به نظر می‌رسد که با استفاده از MongoDB در ابتدا بتوان داده‌ها را ذخیره کرد و سپس با استفاده از الگوریتم‌های احتمالی مانند استفاده از دیکشنری، بتوان داده‌ها را از پیام‌ها خام استخراج کرد. ولی با توجه به توضیحات داده‌شده در بخش‌های نخستین این فرایند بررسی و رد شد. همچنین به منظور ایجاد ارتباط و ذخیره سازی داده‌ها، یک API نوشته می‌شود تا ربات تلگرامی استفاده شده بتواند این داده‌ها را نیز به سمت پایگاه داده انتقال دهد.

۲-۹- چالش‌های مربوط به تحلیلگر متن

بعد از ذخیره‌سازی داده‌ها در داخل پایگاه داده نیاز است تا هر کدام از پیام‌های ذخیره‌شده تحلیل شوند. تحلیل پیام‌ها بدین صورت می‌باشد که تعیین می‌کند آیا پیام مورد نظر سیگنال می‌باشد یا خیر و در صورتی که سیگنال بود بررسی می‌کند آیا سیگنال فروش است یا خرید؟

۲-۳-۱۶- پیاده‌سازی تشخیص سیگنال

به منظور تشخیص سیگنال بودن یا نبود الگوریتمی مبتنی بر شمارش کلمات پیاده‌سازی شد. در ابتدا یک فایلی از تمامی کلید واژه‌های ممکن برای تشخیص سیگنال بودن یک پیام یا نبود آن تهیه شد. به منظور تهیه‌ی این لیست، پیام‌های کانال‌های تلگرامی به صورت دستی بررسی شدند تا کلمات مشترک آن‌ها بدست آید. کلماتی مانند خرید مهم، سیگنال خرید، سیگنال فروش به عنوان مثالی از این کلمات کلیدی می‌باشند.

کلمات کلیدی پیداشده به صورت دستی نیز دسته‌بندی شدند. بیش از ۶۵۰ کلمات کلیدی پیداشده و دسته‌بندی شدند. پنج دسته مربوط به هر کلیدواژه به صورت زیر می‌باشند.

✓ غیرقطعی

✓ خرید

✓ فروش

✓ گروه سهام

✓ اخبار

برخی از کلمات کلیدی مربوط به خرید و فروش می‌باشند و برخی از کلمات نیز خرید یا فروش بودن سیگنال را تعیین نمی‌کنند اما می‌توانند سیگنال باشند. در نتیجه یک دسته‌ای مربوط به غیرقطعی در نظر گرفته شد. همچنین دسته‌ای تحت عنوان گروه اخبار در نظر گرفته شد چراکه برخی از کانال‌های تلگرامی به برخی از اخبار در مورد سیاست و اجرای توافق‌نامه‌ها و سیاست بین‌الملل مانند بگرام، می‌پرداختند که تحلیل آن‌ها به احتمال زیاد سیگنالی برای بورس نخواهند بود.

در پایان کار به منظور افزایش دقت و کارایی الگوریتم به هر کدام از کلمات کلیدی ارزشی بین صفر تا صد نسبت داده شد. ارزش نسبت داده شده به هر کدام از کلمات به این دلیل نسبت داده شد که بتوان بین کلمات یک دسته نیز مقایسه انجام داد، چراکه برخی از کلمات در دقت و اطمینان حاصل کردن نسبت به سیگنال بودن یا نبودن تفاوت داشتند. به عنوان مثال کلمه‌ای مانند سیگنال خرید در مقابل نگهداری دارایی ارزش خیلی بالاتری دارد و در تعیین نوع سیگنال یا سیگنال بودن یا نبودن کمک می‌کند.

ارزش‌گذاری نیز به صورت کاملاً دستی و با بررسی‌های کانال‌های تلگرامی صورت گرفته است. لازم به ذکر است این ارزش‌گذاری‌ها بررسی شده‌اند و در داده‌های مختلف تست شده‌اند تا بیشترین دقت ممکن را داشته باشند. ادامه الگوریتم بدین صورت می‌باشد که متن یک پیام را دریافت می‌کند و در ابتدا یک لیستی از ویژگی‌ها داخل یک دیکشنری درست می‌کند. ویژگی‌ها به صورت زیر هستند.

✓ Is_signal: آیا پیام سیگنال می‌باشد یا خیر.

✓ Is_buy: آیا سیگنال خرید است یا خیر.

✓ Is_sell: آیا سیگنال فروش است یا خیر.

✓ Category: لیستی از نوع صنعتی که این پیام به آن اشاره می‌کند. (مانند دارویی، خودرو)

بعد از این مرحله حلقه‌ی for بر روی لیست کلیدواژه‌های سیگنال بودن پیمایش می‌کند و تعداد کل کلماتی که به عنوان کلیدواژه بوده‌اند و در داخل متن رؤیت شده‌اند را محاسبه می‌کند. همچنین با استفاده از یک حلقه‌ی for دیگر، کلماتی که در دسته‌بندی اخبار بوده‌اند را نیز می‌شمارد.

برای تشخیص سیگنال بودن دو عدد بدست آمده در مرحله‌ی قبل را از یک دیگر کم می‌کند و بر روی تعداد کلمات کل متن تقسیم می‌کند تا عملیات نرمال‌سازی و مقایسه‌ی آن بهتر صورت پذیرد. اگر عدد بدست آمده از حاصل عملیات بزرگ‌تر از ۰.۴۴۳ باشد، پیام به عنوان سیگنال در نظر گرفته می‌شود. عدد ۰.۴۴۳ عددی تصادفی نیست و با بررسی‌های گوناگون به عنوان بهترین مقدار انتخاب شده است. این مقدار می‌تواند پویا باشد ولی پیاده‌سازی آن از محدوده پروژه خارج است. تشخیص سیگنال خرید یا فروش بودن نیز به همین صورت پیاده‌سازی انجام شده است.

$$\frac{\sum Signal - \sum News}{WordCount} \geq 0.443$$

علت این کاهش مقادیر این است که گاهی پیام‌ها سیگنال نمی‌باشند و یا پیام‌های طولانی می‌باشند که حاوی چندین اطلاعات است و پایش را دشوار می‌کند. با کاهش مقادیر سیگنال از اخبار احتمال خطا را کاهش می‌دهیم.

۲-۳-۱۷ بهبود و افزایش کارایی الگوریتم

به منظور بهبود الگوریتم، یک عملیات پاک‌سازی نیز بر روی داده‌های بدست آمده انجام شد. عملیات

پاک‌سازی بدین‌صورت است که اگر الگوریتم تشخیص داد که پیام سیگنال نیست اما یکی از دو ویژگی `is_buy` یا `is_sell` مقدارشان `True` باشد بدین معناست که الگوریتم درست کار نکرده است. به همین جهت در صورت `False` بودن ویژگی `is_signal` حتماً باید دو ویژگی دیگر مقدار `False` را به خود بگیرند. همچنین دو ویژگی `is_buy` و `is_sell` نمی‌توانند هم‌زمان برای یک پیام `True` باشند. به همین منظور بررسی می‌شود در صورت رخ دادن چنین موردی الگوریتم نمی‌تواند به‌صورت دقیق نوع سیگنال را تعیین کند. در نتیجه مقدار هر دو ویژگی را `False` قرار می‌دهیم و تنها مقدار `is_signal` را `True` قرار می‌دهیم.

۲-۳-۱۸ - چالش‌های پیاده‌سازی تحلیلگر پیام

در ابتدا استفاده از الگوریتم‌های هوش مصنوعی نظیر الگوریتم‌های `NLP` و تحلیل عواطف بررسی شدند اما با توجه به زمان محدود پروژه و همچنین عدم دانش کافی در این حوزه و همچنین تعیین شدن محدوده‌ی پروژه، الگوریتم تحلیلگر به‌صورت دستی و بدون استفاده از کتابخانه‌های هوش مصنوعی پیاده‌سازی شد.

به هنگام پیاده‌سازی الگوریتم فرآیند نرماله کردن از اهمیت بالایی برخوردار است چرا که در بررسی یک متن تعداد کل کلمات می‌تواند تأثیرگذار باشد. به همین دلیل زمانی برای بررسی متن‌های هر پیغام قرار داده شد تا بتوان به تحلیل درست و دقیقی از متن‌های پیام‌ها دست پیدا کرد. بدون استفاده از نرماله سازی پیام‌ها بر روی تعداد کلمات پیام امکان دارد تا در تشخیص سیگنال بودن یک پیام یا نبودن آن دچار اشتباه شویم. با استفاده از نرماله سازی، عملکرد الگوریتم به‌صورت بهینه‌تری کار خواهد کرد.

فصل چهارم چالش‌های تیم تحلیل

چالش‌های مربوط به نرم‌افزار Power BI

Power BI یک محصول نرم‌افزاری تجسم داده تعاملی است که توسط مایکروسافت با تمرکز اصلی بر هوش تجاری توسعه یافته است. این بخشی از پلتفرم مایکروسافت پاور است. Power BI مجموعه‌ای از خدمات نرم‌افزاری، برنامه‌ها و رابط‌ها است که باهم کار می‌کنند تا منابع نامرتب داده را به بینش‌های منسجم، غوطه‌ور کننده بصری و تعاملی تبدیل کنند. داده‌ها ممکن است با خواندن مستقیم از پایگاه داده، صفحه وب یا فایل‌های ساختاریافته مانند صفحات گسترده، CSV، XML و JSON وارد شوند.

برای نمایش خروجی بدست آمده و تحلیل داده‌ها از نرم‌افزار Power BI استفاده می‌کنیم. دلیل این انتخاب با بررسی‌های صورت گرفته این است که نرم‌افزاری بسیار گسترده است و به آسانی می‌توان خروجی تیم توسعه نرم‌افزار را در قالب فایل اکسل، SQL گرفته و تحلیل‌های بسیاری را انجام دهیم. یکی از نقاط قوت این نرم‌افزار قابلیت اجرا کردن دستورات SQL برای فیلتر کردن روی داده‌ها می‌باشد.

۱۰-۲- چالش‌های سخت‌افزاری مربوط به تیم تحلیل

یکی از چالش‌های که با آن مواجه شدیم نبودن نرم‌افزار Power BI برای کامپیوترهای اپل بود. اپل در سال ۲۰۱۹ نسل جدید پردازنده‌های M1 خود را رونمایی کرد. این پردازنده‌ها از خانواده پردازنده‌های ARM می‌باشند که معماری متفاوتی نسبت به پردازنده‌های ۳۲ و ۶۴ بیتی دارند. از آن جهت که عده کمی (کمتر از ۱۰ درصد کاربران) از این کامپیوترها استفاده می‌کنند خیلی از نرم‌افزارها از معماری ARM پشتیبانی نمی‌کنند. به همین دلیل مجبور شدیم تنها از دو سیستم دیگر استفاده کنیم که به علت کندتر بودن و قدیمی‌تر بودن چالش‌هایی را به همراه داشته‌اند.

چالش بعدی که با آن مواجه شدیم خراب شدن یکی از سیستم‌های اعضا تیم تحلیل است. این امر به شدت سرعت ما را پایین آورد ولی بعد از بررسی‌های صورت گرفته لپتاپی از نزدیکان قرض کردیم و از نسخه وب نرم‌افزار استفاده کردیم. این چالش باعث شد مقداری از برنامه پیاده‌سازی و توسعه عقب بیفتیم ولی با مقداری کوشش به برنامه رسیدیم و ادامه دادیم.

۱۱-۲- چالش‌های مربوط به ایجاد حساب

برای استفاده از این نرم‌افزار می‌بایست حساب کاربری خود را داشته باشیم که با توجه به تحریم‌های موجود کاری دشوار می‌باشد. در صورتی که حساب کاربری نداشته باشیم امکان استفاده از ابزارهای نرم‌افزار وجود ندارد.

برای ایجاد حساب کاربری ابتدا از ایمیل دانشگاه استفاده کردیم تا از امکانات نرم‌افزار بهره ببریم. در مرحله بعدی با وارد شدن به حساب توانستیم از امکانات استفاده کنیم و نمودارهای دلخواه را رسم کنیم.

۱۲-۲- چالش‌های مربوط به فرمت ورودی Power BI

با توجه به اینکه از پایگاه داده SQL در پیاده‌سازی استفاده کردیم می‌توانستیم خروجی پایگاه داده را به سادگی فیلتر کنیم و در هر فرمتی خروجی دهیم. همچنین می‌توانستیم فایل database.sqlite3 موجود در پیاده‌سازی را در Power BI به صورت مستقیم وارد کنیم و با سینتکس SQL فیلترهای لازم را اعمال کنیم.

در تولید نمودارها هر دو روش استفاده شده به صورتی که برای تعدادی از نمودارها از فایل اکسل و برای تعدادی دیگر از SQL استفاده کردیم. در استفاده از SQL به سادگی می‌توانستیم فیلترهایی را اعمال و تقریباً اتومات نمودارها را تشکیل دهیم. بنابراین برای پروژه‌های بعدی تنها از SQL استفاده خواهیم کرد.

۱۳-۲- آموزش و یادگیری Power BI

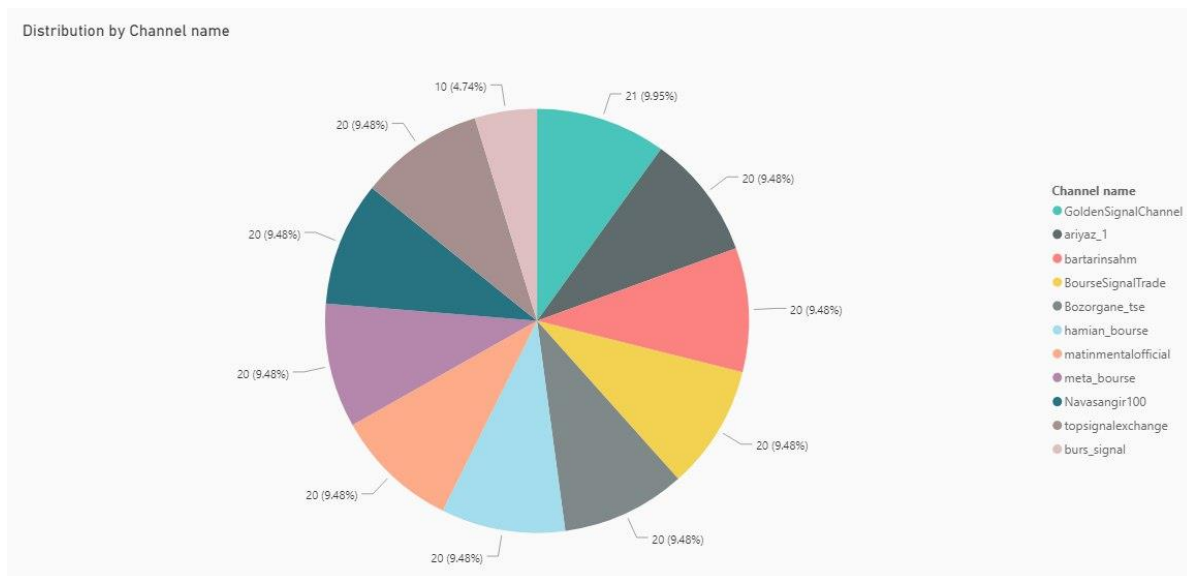
به منظور آموزش و یادگیری این نرم‌افزار از فضاهای مجازی مانند یوتیوب استفاده کردیم. همچنین مستندات ماکروسافت بسیار کامل و دقیق بودند و کمک بسیاری به یادگیری کردند. همچنین فرمی برای پرسش و پاسخ‌های فراوان وجود داشت تا در صورتی که به مشکلی شناخته شده برخوردیم به سادگی مشکل را برطرف کرده و بر آن غلبه کنیم.

۱۹-۳-۲- یادگیری با داده آموزشی

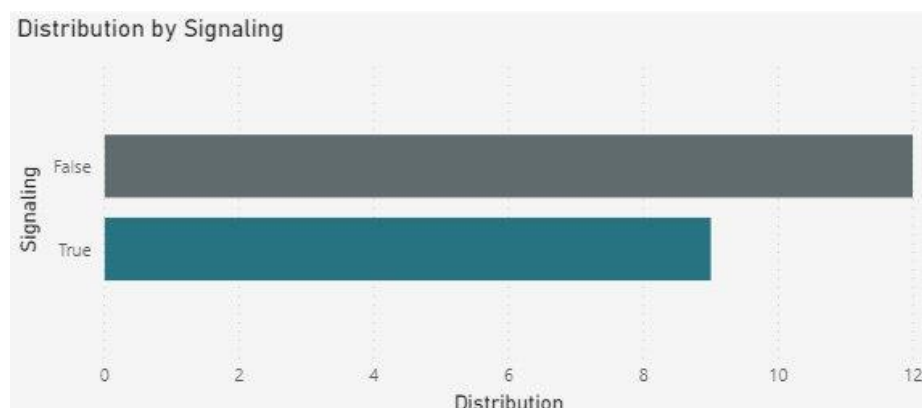
تیم تحلیل به موازات تیم توسعه به یادگیری نرم‌افزار پرداختند. این یادگیری و ارتقا دانش زمانی حدود ۱۲ ساعت در هفته طول می‌کشید و این روند به اندازه یک اسپرینت و نصف طول کشید. در انتها تیم تحلیل به تسلط نسبی و خوبی رسیدند و آمادگی این را داشتند تا با داده تصادفی ولی در فرمت خروجی نرم‌افزار توسعه یافته کار کنند.

تیم توسعه فایل اکسلی که حاوی تمامی ردیف‌های جدول خروجی پایگاه داده هست را در اختیار تیم تحلیل قرار دادند. تیم تحلیل نیست با کمک این داده آموزشی نمودارهای خام و اولیه‌ای تولید کردند. در نهایت در جلساتی تصمیم گرفتیم از چه داده‌هایی نمودار تولید کنیم و نمودارهای تولیدی به چه شکلی باشند.

به عنوان مثال از نمودار دایره‌ای برای نمایش تعداد پیام استخراج شده از کلنال‌های مختلف را با یکدیگر مقایسه می‌کنیم (عکس ۴-۱)، و یا با استفاده از نمودار افقی تعداد پیام‌هایی که سیگنال هستند یا نیستند را در کانال مربوطه با یکدیگر مقایسه کنیم (عکس ۴-۲).



عکس ۴-۱



عکس ۴-۲

۲۰-۳-۲ چالش‌های مربوط به دسته‌بندی مناسب داده‌های خروجی

یکی از چالش‌ها دسته‌بندی مناسب ستون‌های اکسل و SQL می‌باشد. در نرم‌افزار Power BI باید مشخص کنیم کدام ستون‌ها را برای رسم نمودار انتخاب کند. به‌عنوان مثال ستونی که اطلاعات زمانی به فرمت سال:ماه:روز ساعت:دقیقه:ثانیه ذخیره می‌کند نباید به‌صورت خام استفاده شود چراکه ممکن است اختلاف زمانی ردیف‌ها در حد ساعت و یا روز باشد و نمایش با سال بسیار نمودار نادقیق و اشتباهی تولید

می‌کند. برای تولید هر نمودار ستون‌های متفاوتی انتخاب شدند و نتایج مقایسه شدند تا بهترین نمودار انتخاب شود.

۲۱-۳-۲ - چالش مربوط به اکسل

در اکسل با مشکلی مواجه شدیم. ستون‌هایی که خالی بودند با نام `column_name` و داده `null` مقداردهی شده بودند و این تولید نمودار را با مشکل مواجه کرد به‌طوری‌که مجبور به پاک کردن اطلاعات اکسل و گرفتن خروجی دوباره شدیم. احتمالاً دلیل این امر قدیمی بودن نسخه اکسل بر روی کامپیوتر اعضا بود. باین‌حال این مشکل در زمانی کوتاه برطرف شد.

۲۲-۳-۲ - چالش‌های مربوط به رابط کاربری و رنگ‌بندی نمودارها

یکی از چالش‌ها انتخاب نمودار مناسب می‌باشد. در صورتی‌که نمودار اشتباه برای نمایش داده‌هایمان انتخاب کنیم می‌توانیم پیامی نادرست ارسال کنیم. به‌عنوان مثال بهتر از پراکندگی را به کمک نمودار Pie-Chart نمایش دهیم و نمودارهای زمانی را به‌صورت پیوسته نمایش دهیم.

چالش دیگری که با آن روبه‌رو شدیم رنگ‌بندی نمودارها می‌باشد. با توجه به اینکه نمودارهای گوناگون مفاهیم مختلفی را در بردارند نباید رنگ‌بندی یکسانی نیز داشته باشند در غیر این صورت اطلاعات غلطی را انتشار می‌دهند. سعی کردیم تا برای نمودارها از رنگ‌های گوناگون استفاده کنیم تا بهترین کیفیت را داشته باشند.

۲۳-۳-۲ - چالش‌های مربوط به مدل‌سازی و ارتباط بین نمودارها

یکی از قابلیت‌های بسیار جذاب نرم‌افزار Power BI توانای برقراری ارتباط بین نمودارها می‌باشد. به این صورت که اگر در نقاط خاصی از نمودار مادر کلیک شود یا تغییری ایجاد شود تمامی نمودارهای فرزند

نیز دچار تغییراتی می‌شوند. در برخی نمودارها از این قابلیت جذاب استفاده کردیم تا اطلاعات را به شکل صحیح و مفید انتقال دهیم.

۲۴-۳-۲ - چالش‌های مربوط به استفاده از Color Saturation در Tree-map

در تولید برخی نمودارها که قصد داشتیم از Color Saturation استفاده کنیم، به این صورت که نمودار با پیشروی به سبز و با پسروی به قرمز متمایل شود و شهود بهتری از پیشروی نمودار در اختیار کاربران قرار می‌دهد.

چالشی که با آن روبه‌رو شدیم به این دلیل بود که بخش Data Color در نسخه‌های جدید Power BI حذف شده است و به شکلی ساختار قابل استفاده است. اما این ساختار جدید قابلیت‌های نسخه قبلی را ندارد و باعث شد از این نمودار گذر کنیم و به نمودارهای دیگر بپردازیم.

۱۴-۲ - چالش‌های مربوط به تولید کلیدواژه‌ها

یکی از رکن‌های اصلی نرم‌افزار تحلیل‌گر استفاده از کلیدواژه‌ها و ارزش‌های آن‌ها بود. در اصل تمام تحلیل و تشخیص سیگنال و عدم سیگنال از همین طریق صورت می‌گرفت. در نتیجه تیم تحلیل در طول اسپرینت اول بیست کانال بورس و سیگنال که کاربران زیادی داشتند را انتخاب کردند و به دنبال کلیدواژه‌ها گشتند. این کلیدواژه‌ها به صورت یک نداشت یک به چند بود. به عنوان مثال کلیدواژه‌های خرید، ارزش، دارای و نگهداری سهم معنی سیگنال خرید را می‌دهند.

این کلیدواژه‌ها توسط تیم تحلیل استخراج و در دسترس تیم توسعه قرار گرفت. بعد تیم توسعه این اطلاعات را در پایگاه داده ثبت کردند تا بخش تحلیل‌گر نرم‌افزار از آن‌ها استفاده کند.

۱۵-۲- چالش‌های مربوط به ارزش کلیدواژه‌ها

هر کلیدواژه ارزشی دارد. این ارزش باید از بقیه کلیدواژه‌ها متفاوت باشد تا تحلیل به درستی انجام گیرد. به عنوان مثال ارزش کلیدواژه «خرید» به مراتب بالاتر از «شرایط مثبت» می‌باشد. تیم تحلیل در طول اسپرینت اول و دوم کلیدواژه‌های استخراج شده را ارزش گذاری کردند و این ارزش‌ها را در اختیار تیم توسعه قرار دادند. بعد تیم توسعه این اطلاعات را در پایگاه داده ثبت کردند تا بخش تحلیل گر نرم افزار از آن‌ها استفاده کند.

۱۶-۲- چالش بدست آوردن فرمول تحلیل گر

با داشتن کلیدواژه‌ها و ارزش‌ها می‌توان ارزش خام هر فرایند را بدست آورد. ولی در مواردی مشاهده کردیم پیامی طولانی وجود دارد که از شرایط فعلی بازار و یا سیگنال دهی اشتباه شکایت به عمل می‌آورند و یا پیامی طولانی با چندین سیگنال نامشخص وجود دارد. در نتیجه پیام‌های یک ماه را از کانالی دریافت و ارزش گذاری کردیم، این ارزش‌ها را مقایسه کردیم و Threshold نرمال شده‌ای را بدست آوردیم به صورتی که ارزش کلیدواژه‌ها بر ارزش غیر کلیدواژه‌ها باید از مقداری بیشتر باشد در غیر این صورت پیام معتبر نمی‌باشد. می‌توان در تکامل‌های بعدی پروژه روشی برای پیاده سازی فرمولی پویا بدست آورد ولی در شرایط فعلی از محدوده پروژه خارج بوده است.

۱۷-۲- چالش مقایسه نتایج با سایت کدال

به منظور کشف تقلب قصد داریم تا از نتایج تحلیل شده توسط نرم افزار نمودارهایی را تولید کنیم و با استفاده از نمودارها و مقایسه نتایج با اطلاعات سایت کدال کانال‌های مشکوک را شناسایی کنیم و گزارش دهیم. برای این منظور نمودارهایی را با کمک اطلاعات ذیل ایجاد می‌کنیم.

✓ تنوع پیام‌ها از کانال‌های مختلف.

✓ پیک بازدید از هر کانال در ساعات گوناگون روز.

✓ تعداد پیام‌هایی که سیگنال هستند/نیستند.

✓ تعداد پیام‌هایی که سیگنال خرید/فروش هستند.

✓ تعداد دفعات فوروارد شدن پیام کانال‌ها

✓ ادیت شدن یا نشدن پیام‌های کانال

✓ و ...

در مرحله آخر این نمودارها را با درصد باارزش/کم‌ارزش شدن سهم سیگنال دهی شده مقایسه می‌کنیم. در صورتی که باوجود اطلاعات ضدونقیض، شرایط خوب و بد بورس اوراق بهادار و تعداد دفعات سیگنال دهی درست می‌توان تا حد خوبی کشف تقلب انجام دهیم.

چالشی که در این مرحله با آن برخوردیم نیاز به بررسی دستی و بدون استفاده از کامپیوتر بود که کار را بسیار دشوار کرد و در صورتی که قصد گسترش پروژه در سطح تمامی شبکه‌های اجتماعی داشته باشیم تیم تحلیل بسیار بزرگ متشکل از متخصصین بسیار زیاد نیاز داریم که هزینه و وقت بسیار زیادی هدر می‌دهد.

ولی در سطح محدوده پروژه همین میزان کافی است و تنها قصد Proof-of-Concept - اثبات مفهوم (POC یا PoC) که به‌عنوان اثبات اصل نیز شناخته می‌شود، تحقق یک روش یا ایده معین به‌منظور اثبات امکان‌پذیری آن، یا اثبات اصولی باهدف تأیید اینکه یک مفهوم یا نظریه دارای پتانسیل عملی است. اثبات مفهوم معمولاً کوچک است و ممکن است کامل باشد یا نباشد - را داشتیم که مشخص کنیم آیا این امر شدنی است یا راه دیگری را باید پیش گرفت.

فصل پنجم

خروجی نرم افزار و تیم تحلیل

خروجی نهایی و نمودارهای خروجی

به منظور تحلیل خروجی و بررسی های انجام شده از نمودارهای گوناگون استفاده می کنیم. تحلیل ها به دو صورت گروهی و تکی صورت می گیرد. تحلیل های گروهی مربوط به تمامی کانال ها می باشند مانند تعداد بازدید، تعداد بازنشر و غیره در صورتی که نمودارهای تکی داده های هر کانال را نسبت به خودش بررسی می کند مانند تعداد سیگنال خرید/فروش و یا اوج بازدید کانال ها در ساعاتی خاص. در این پروژه کانال های ذیل مورد بررسی قرار می گیرند. توجه کنید داده های استخراج شده مربوط به هفته های انتهای خرداد ماه سال ۱۴۰۱ می باشد.

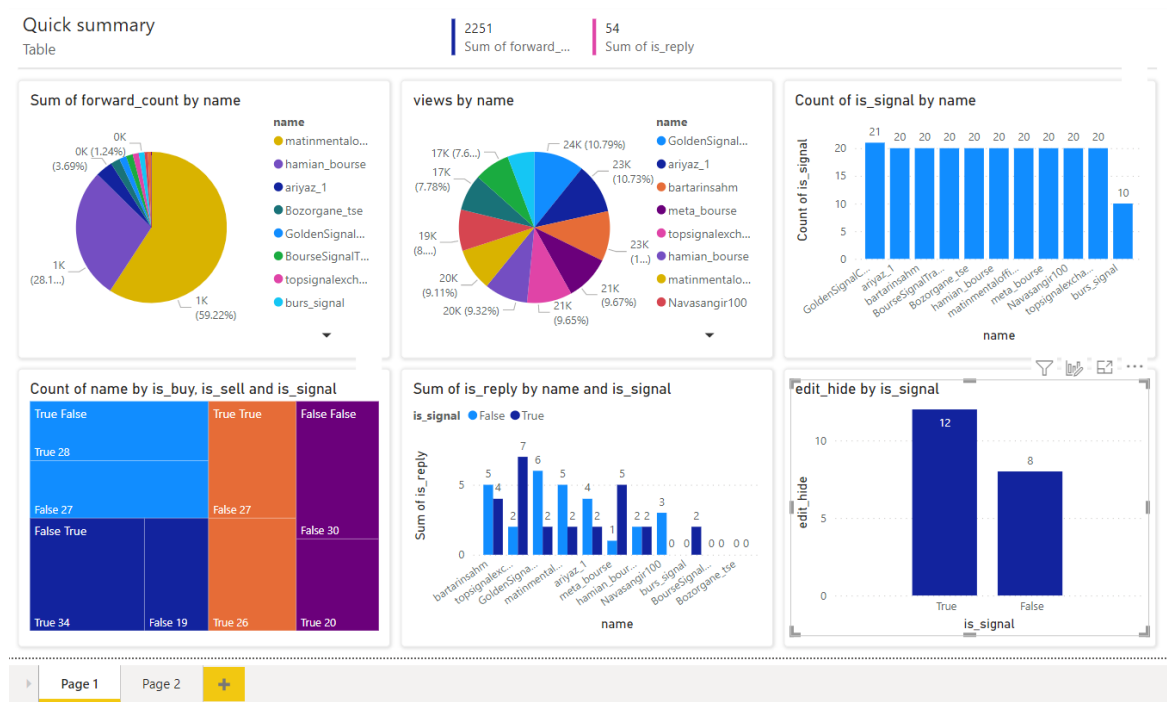
- ✓ https://t.me/ariyaz_1
- ✓ <https://t.me/bartarinsahm>
- ✓ <https://t.me/topsignalexchange>
- ✓ https://t.me/Bozorgane_tse
- ✓ <https://t.me/BourseSignalTrade>
- ✓ https://t.me/hamian_bourse
- ✓ <https://t.me/GoldenSignalChannel>
- ✓ <https://t.me/Navasangir100>
- ✓ <https://t.me/matinmentalofficial>
- ✓ https://t.me/burs_signal
- ✓ https://t.me/meta_bourse

۱۸-۲- نمودارهای گروهی

دو نمودار گروهی تولید شده که به نمایش و توضیح هریک می پردازیم

۲۵-۳-۲- نمودار اول: بررسی مختصر

نمودار ۱-۵ نمودار بررسی مختصر کانال های انتخاب شده می باشد. این نمودار از شش نمودار کوچکتر تقسیم شده که تمامی اجزاء آن به یکدیگر متصل هستند و با تغییر یکی مابقی نیز تغییر می کنند. این تغییر در مقادیر و شکل نمودار نمی باشد و صرفاً در نمایش اطلاعات می باشد.

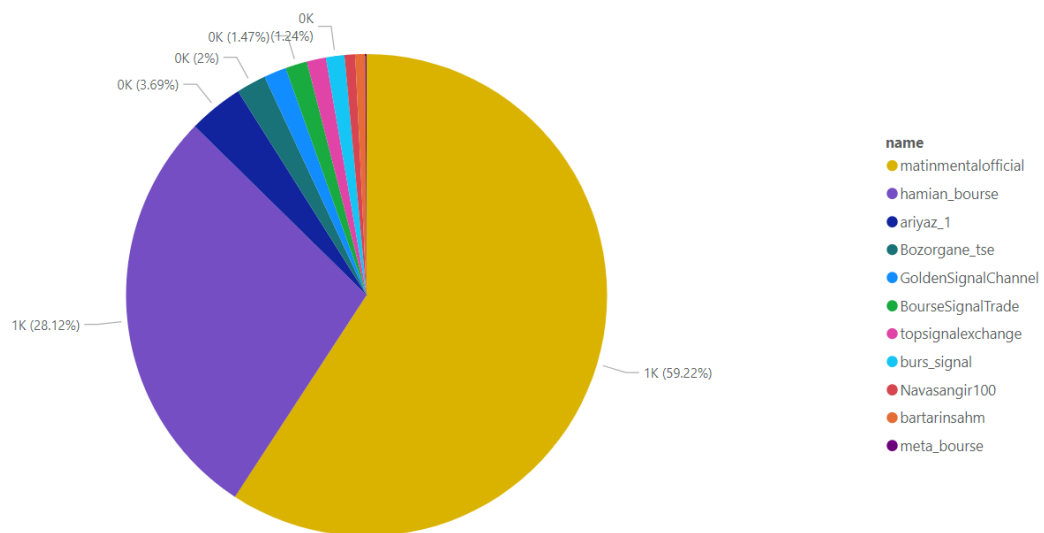


نمودار ۱-۵

یکی از عواملی که کانال متقلب را مورد توجه سازمان بورس اوراق بهادر قرار می دهد تعداد بازدید و باز نشر پیام هایش است. به این منظور کانال های بررسی شده را از این منظور با یکدیگر مقایسه می کنیم. دو نمودار اول از چپ بالا بیانگر این مفاهیم هستند.

۵-۱-۱-۱ تعداد بازنشر پیام‌های بارگذاری شده

همان‌طور که گفته شد تعداد بازنشر یکی از فاکتورهای بسیار مهم در کشف تقلب می‌باشد. در صورتی که دو کانال متقلب داشته باشیم اولویت کشف تقلب با کانالی است که بازنشر بیشتری دارد و بر روی بازار بورس اوراق بهادار تأثیر بیشتری می‌گذارد. نمودار ۵-۲ بیانگر این مفهوم می‌باشد.

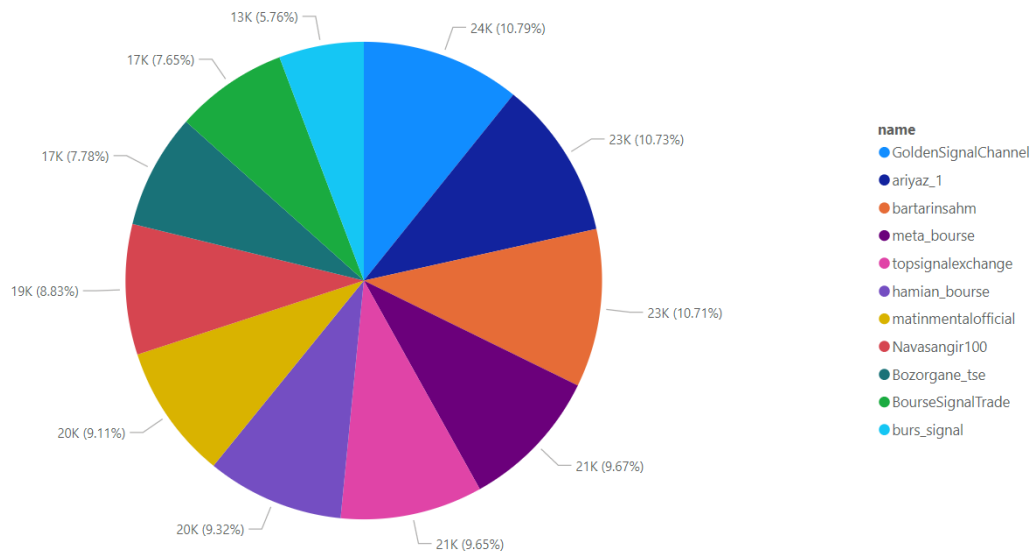


نمودار ۵-۲

۶-۱-۱-۱ تعداد بازدید پیام‌های بارگذاری شده

تعداد بازنشر تأثیر مستقیم بر تعداد بازدید پیام‌ها دارد و هرچه تعداد بازنشر بالاتر رود تعداد بازدید پیام‌ها نیز بالاتر می‌رود و بر میزان خطر افزوده می‌شود. نمودار ۵-۳ بیانگر تعداد بازدید کل پیام‌های استخراج شده از کانال‌های مربوطه می‌باشد. با مقایسه‌ای ساده به درستی گزاره «هرچه تعداد بازنشر بالاتر باشد تعداد بازدید نیز بالاتر می‌باشد» می‌رسیم.

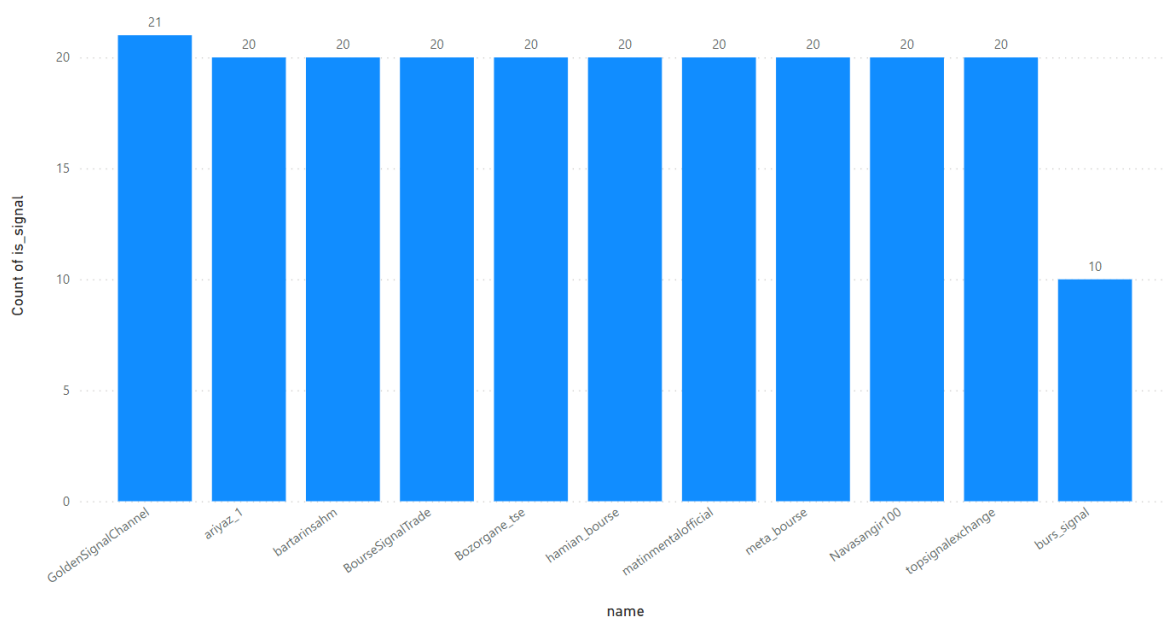
همچنین هرچه تعداد بازدید پیام های کانالی بیشتر باشد سیگنال های خرید و فروش کانال مربوطه به مراتب بیشتر مورد توجه قرار می گیرند. پس معیار بعدی که مورد مطالعه قرار می گیرد تعداد سیگنال های داده شده توسط هر کانال تلگرامی می باشد.



نمودار ۳-۵

۷-۱-۱-۱ سیگنال دهی انجام شده توسط کانال ها

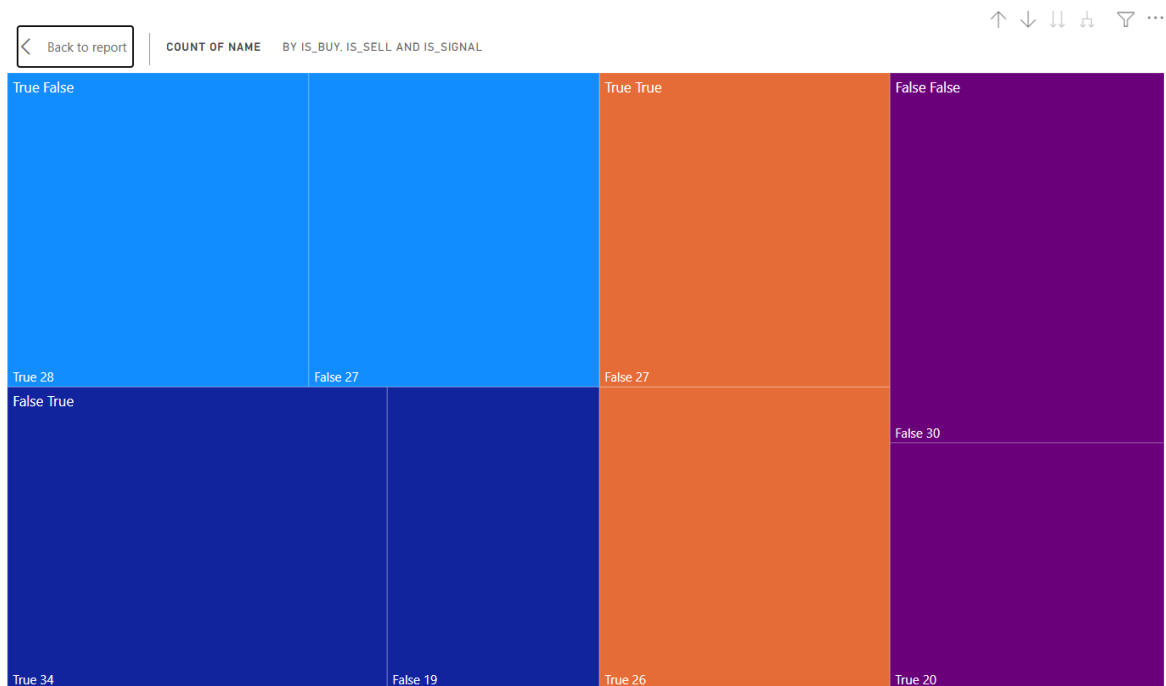
این موضوع که هر کانال چه تعداد سیگنال دهی داشته اند در کنار تعداد بازدید و باز نشر عمر بسیار مهمی در کشف تقلب می باشد. نمودار ۴-۵ نموداری ستونی می باشد که این مقایسه را برایمان آسان می سازد.



نمودار ۴-۵

۸-۱-۱-۱- تعداد سیگنال‌ها، سیگنال‌های خرید و سیگنال‌های فروش

در نمودار ۵-۵ می‌توانید گستردگی سیگنال‌ها، سیگنال‌های خرید و فروش را به صورت جمعی مشاهده کنیم. این نمودار به ما کمک می‌کند تا بررسی کنیم کدام کانال‌ها بیشتر در چه سیگنال دهی‌هایی دخیل هستند.



نمودار ۵-۵

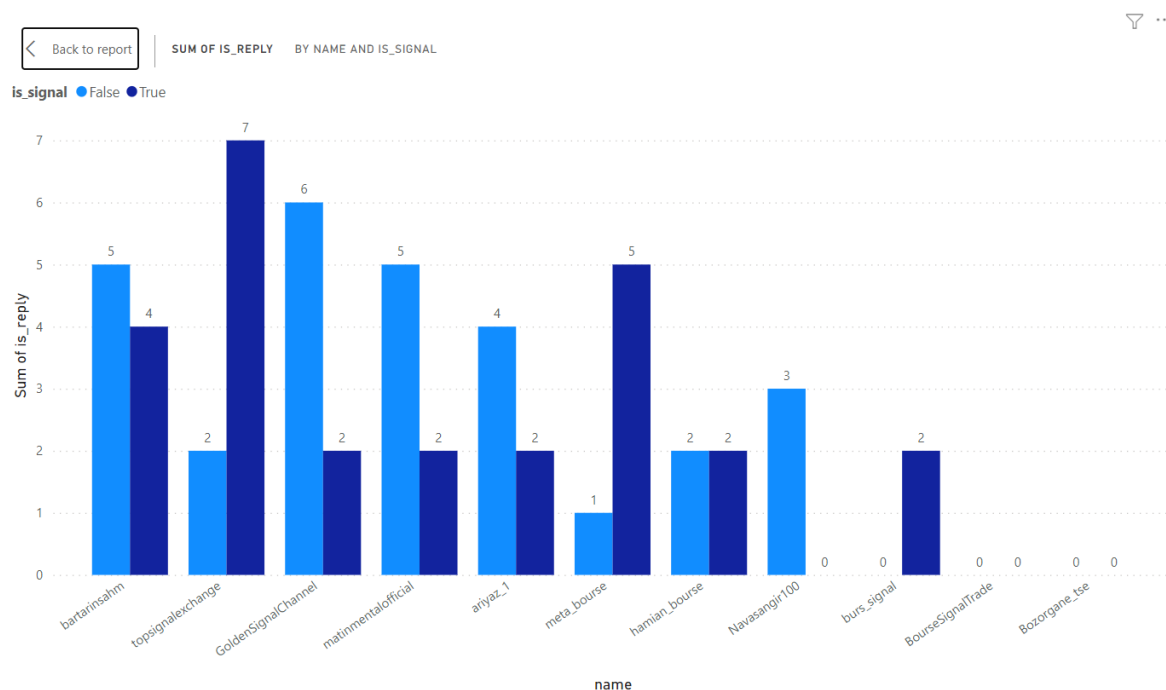
۹-۱-۱-۱- تعداد پاسخ های سیگنال دهی شده

گاهی در کانال های تلگرامی عکس، نموداری یا پیامی را ارسال می کنند و در پاسخ به محتوای مربوطه سیگنال دهی انجام می دهند. در نتیجه بررسی پیام هایی که پاسخ به محتوایی هستند و سیگنال دهی انجام می دهند اطلاعات مفیدی را در اختیار ما قرار می دهد. نمودار ۵-۶ این امر را مورد بررسی قرار داده است.

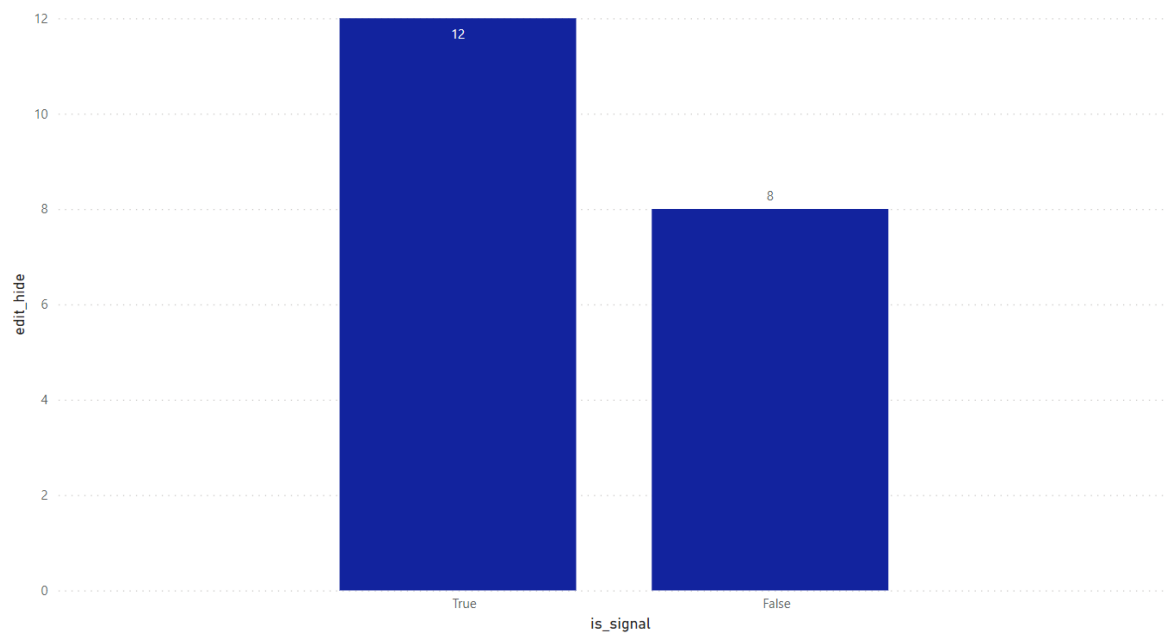
۱۰-۱-۱-۱- سیگنال دهی هایی که ادیت پیام پنهان شده است

ممکن است کانالی سیگنال دهی غلط یا درستی انجام دهد و بعد از گذر زمان پیام مربوطه را تغییر دهد. این در صورتی مشکل ساز است که ادمین کانال در تنظیمات تغییراتی را ایجاد کرده باشد تا کاربران متوجه تغییر داده شدن پیام نشوند. به عنوان مثال ممکن است سیگنال دهی غلطی انجام دهد و

بعد از مشخص شدن نتیجه واقعی این سیگنال دهی را تغییر دهد و به کاربران بگوید که سیگنال دهی او صحیح بوده است. نمودار ۵-۷ این موضوع را مورد بررسی قرار می دهد.



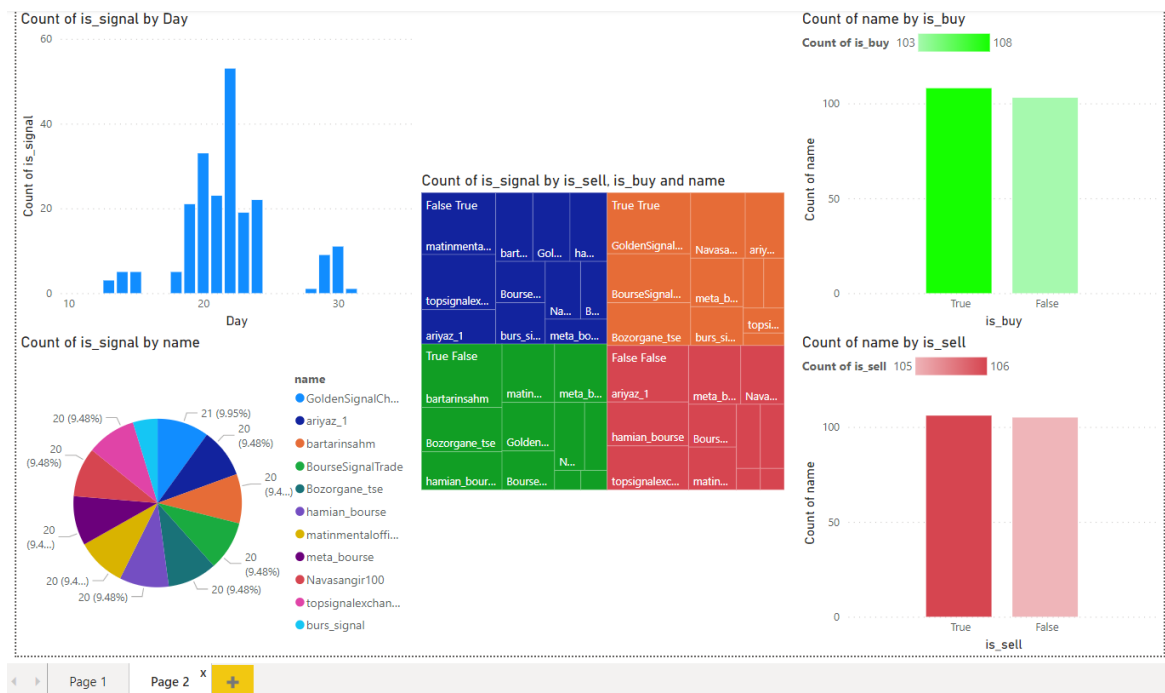
نمودار ۵-۶



نمودار ۵-۷

۲۶-۳-۲- نمودار دوم: بررسی دقیق تر

نمودار ۵-۸ مواردی را همچون تعداد سیگنال، تعداد سیگنال خرید، تعداد سیگنال فروش، سیگنال های فرستاده شده در روزهای خاص و گستردگی سیگنال های تحلیل شده را به نمایش می گذارد.



نمودار ۵-۸

۱-۱-۱-۱- گستردگی سیگنال دهی در روزهای مختلف

گاهی ممکن است با یک سیگنال دهی تمامی کانال ها آن را به اشتراک بگذارند یا به صورت بازه ای سیگنال دهی زیاد باشد و در آینده کاهش یابد. این امر می تواند به علت نشر شدن اطلاعاتی نهانی باشد و باید رصد شود. نمودار ۵-۹ این مقایسه را انجام داده است. همان طور که مشاهده می کنید در بازه ۲۰ الی

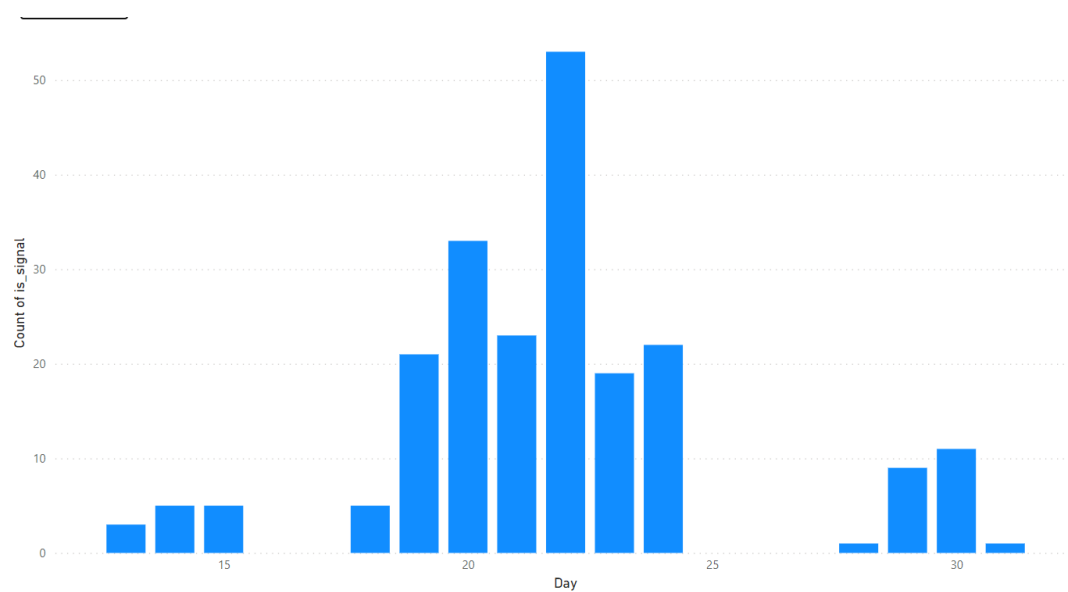
۱۲۵م خردادماه سیگنال دهی بیشتری صورت گرفته. البته تعطیلات و یا بد بودن شرایط بازار در این موضوع تأثیرگذار می باشد.

۱-۱-۱-۱ سیگنال دهی انجام شده توسط کانال ها

این موضوع که هر کانال چه تعداد سیگنال دهی داشته اند در نمودار ۵-۱۰ که نمودار Pie-Chart می باشد قابل مقایسه است.

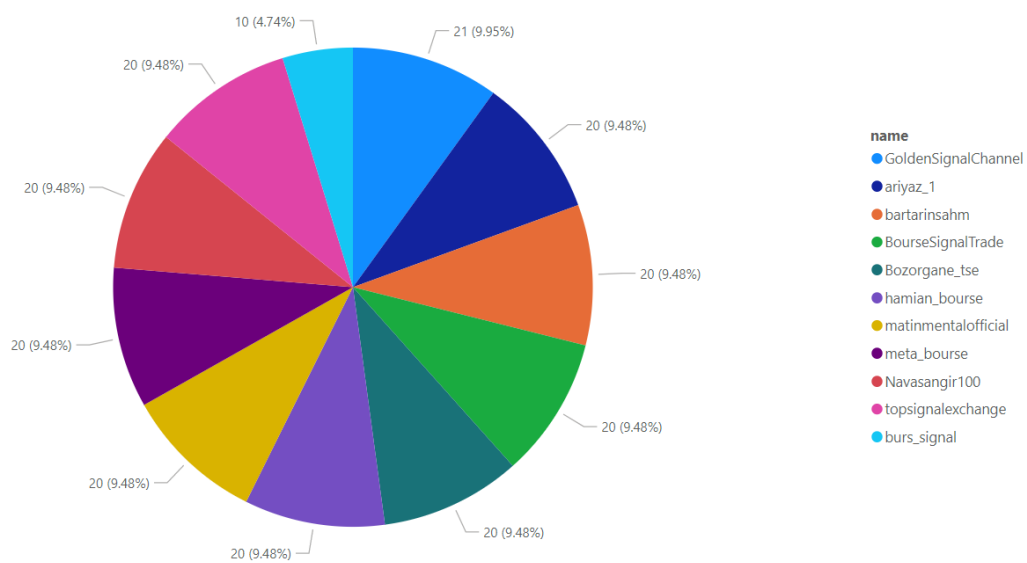
۱-۱-۱-۲ گستردگی سیگنال خرید و سیگنال فروش

همان طور که در بخش های قبل توضیح دادیم ممکن است بعضی پیام ها سخت باشند و تحلیلی دشوار داشته باشند. و یا پیامی بلند باشند که چند سیگنال را در بردارند. گاهی ممکن است کانال متقلبی سیگنالی صحیح با اطلاعات نهانی را در دل سیگنال های دیگر جای دهد تا به چشم نیاید و تصادفی تصور شود. بنابراین مهم است تا گستردگی سیگنال خرید و فروش در بین کانال ها مورد بررسی قرار گیرد تا مواردی از این دست مشاهده شوند. مقایسه گستردگی در نمودار ۵-۱۱ صورت گرفته است.

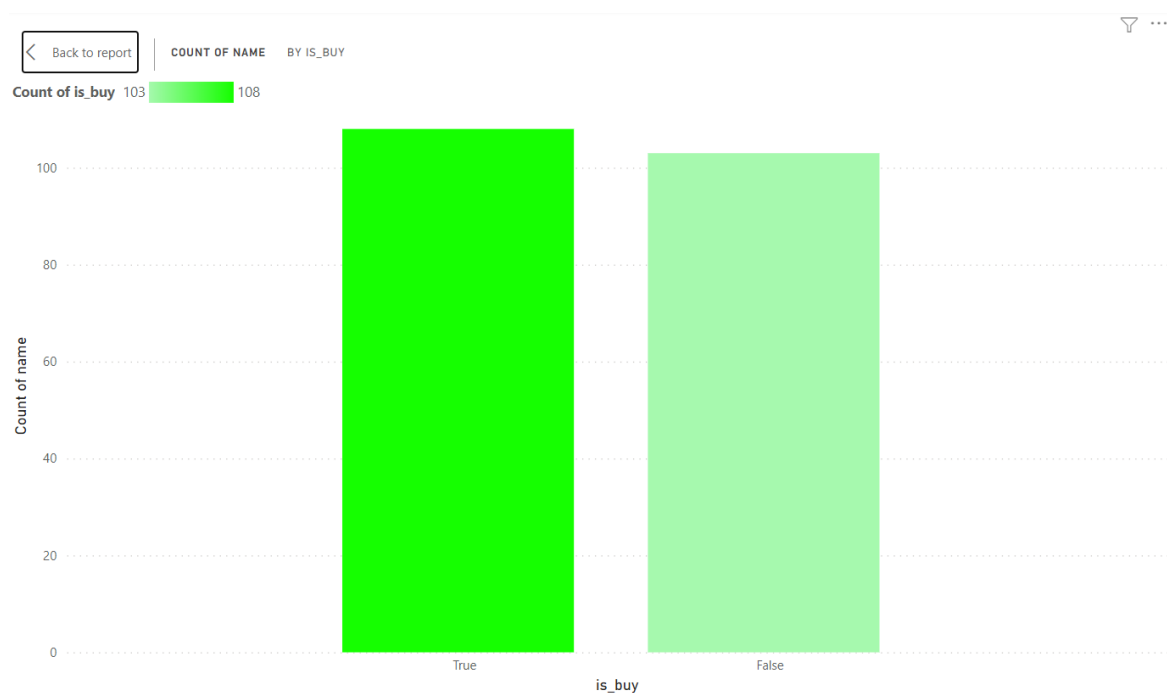


نمودار ۵-۹

COUNT OF IS_SIGNAL BY NAME



نمودار ۵-۱۰



نمودار ۵-۱۲



نمودار ۵-۱۳

۱۹-۲- نمودارهای تکی

نمودارهای تکی مواردی را همچون تعداد پیام استخراج شده از هر کانال، تعداد سیگنال/غیر سیگنال هر کانال، تعداد سیگنال خریدوفروش و تعداد بازدید رد ساعات مختلف روز را در بردارد. از این نمودارها می توان تحلیل هایی همچون ساعت اوج فعالیت کانال ها و گستردگی سیگنال های هر کانال را با یکدیگر مقایسه کرد و اطلاعات مفیدی را درباره کانال بدست آورد. نمودار ۵-۱۴ گستردگی پیام های استخراج شده را به نمایش می گذارد. این نمودار در تمامی نمودارهای تکی تکرار خواهد شد.

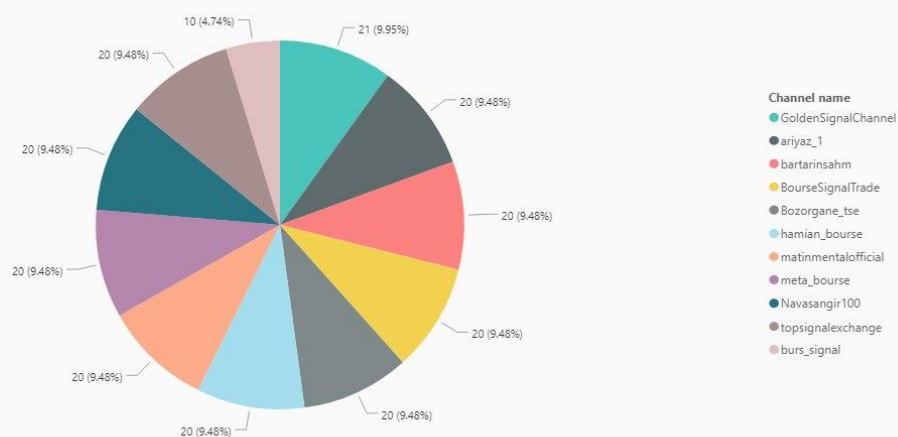
27-3-2- نمودار تکی مربوط به کانال GoldenSignalChannel

این نمودار (نمودار ۵-۱۵) شامل ۳ نمودار کوچک تر می باشد. نمودار بالا چپ همان نمودار ۵-۱۴ می باشد که بخش مربوط به کانال مربوطه را نشان می دهد. نمودار بالا راست تعداد سیگنال ها را با تعداد غیر سیگنال ها مقایسه می کند همچنین یک نمودار افقی وجود دارد که از بین پیام های سیگنال تعداد سیگنال های خرید را با سیگنال های فروش مقایسه می کند. نمودار سوم نیز مشخص می کند پیام های ارسال شده در ساعات مختلف روز چقدر بازدید داشته اند. این اطلاعات مفیدی همچون ساعت اوج کانال را به نمایش می گذارد. همچنین نشان می دهد ادمین کانال در چه ساعاتی فعالیت بیشتری دارد.

توضیح تمام نمودارهای تکی به همین شکل می باشد و تنها کانال مربوطه متفاوت است. در نتیجه از زیاده گویی پرهیز می کنیم و فقط شرح نمودار اول می پردازیم.

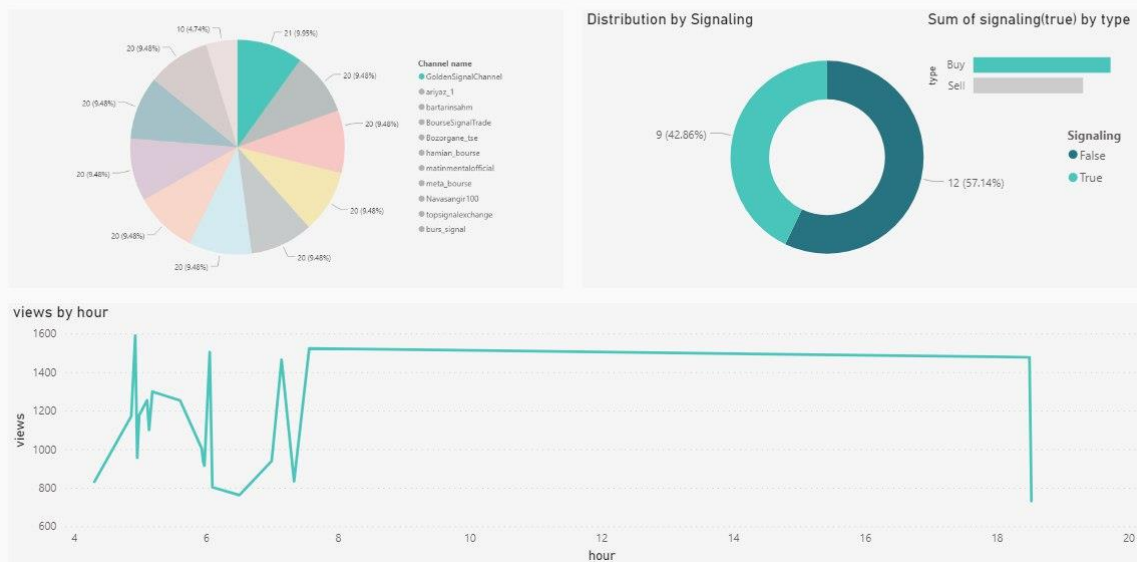
Message Distribution

Distribution by Channel name



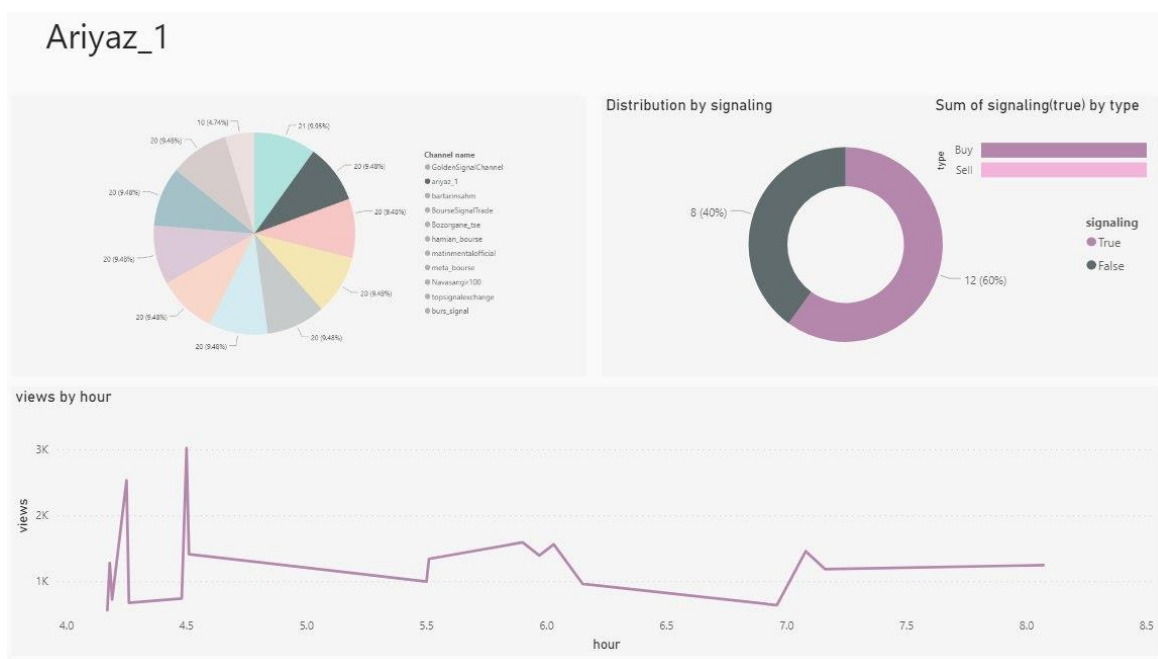
نمودار ۵-۱۴

GoldenSignalChannel



نمودار ۵-۱۵

۲۸-۳-۲ - نمودار تکی مربوط به کانال Aryaz_1



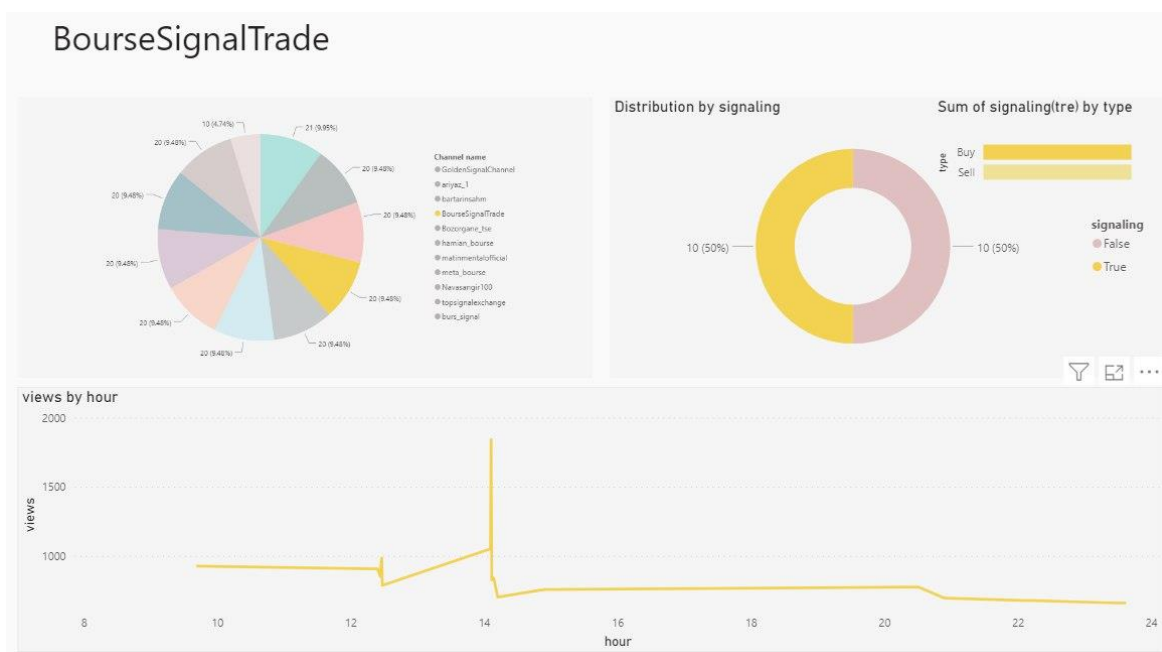
نمودار ۵-۱۶

۲۹-۳-۲ - نمودار تکی مربوط به کانال Bartarinsahm



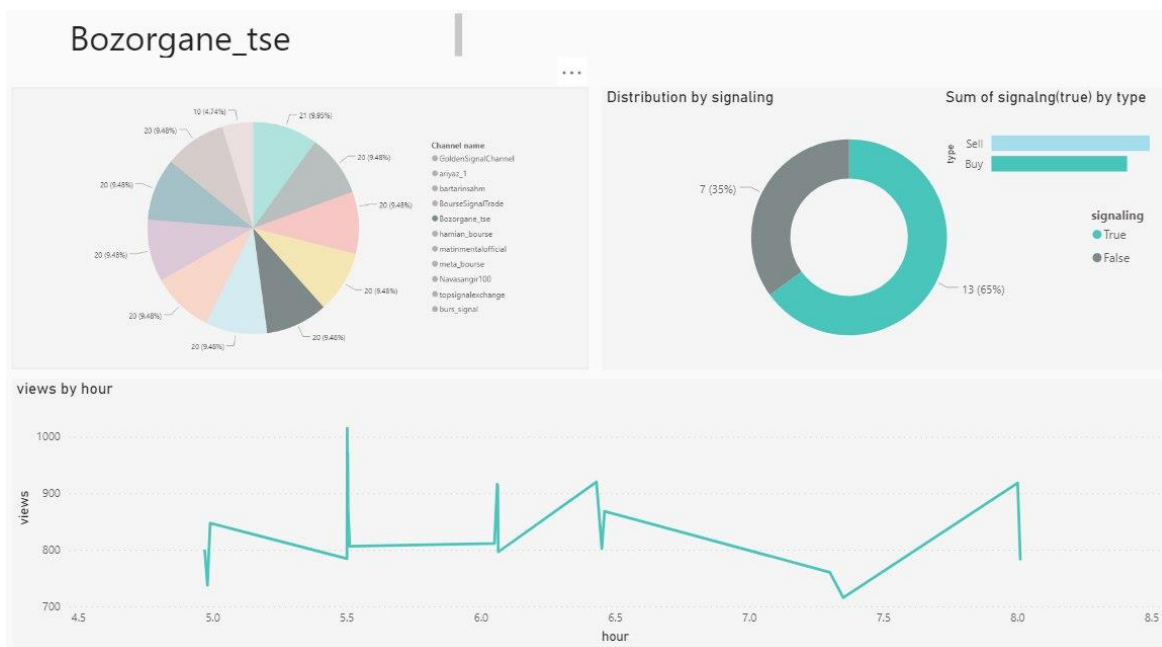
نمودار ۵-۱۷

۳۰-۳-۲ نمودار تکی مربوط به کانال BourseSignalTrade



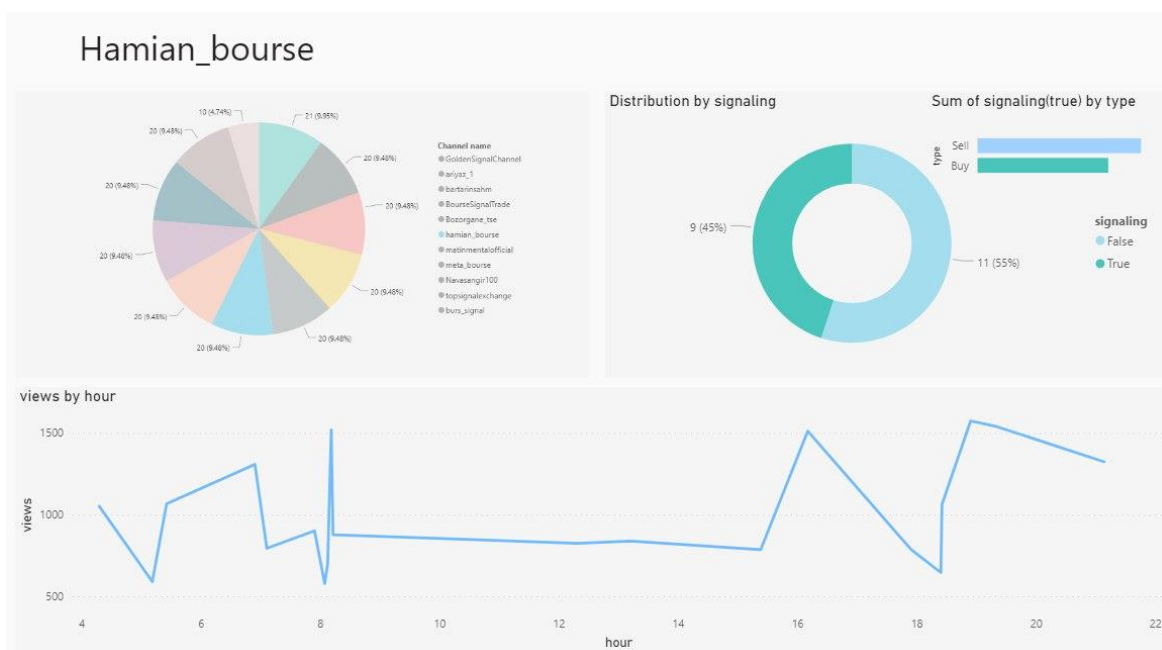
نمودار ۵-۱۸

۳۱-۳-۲ نمودار تکی مربوط به کانال Bozorgane_tse



نمودار ۵-۱۹

۲-۳-۳۲ - نمودار تکی مربوط به کانال Hamian_bourse



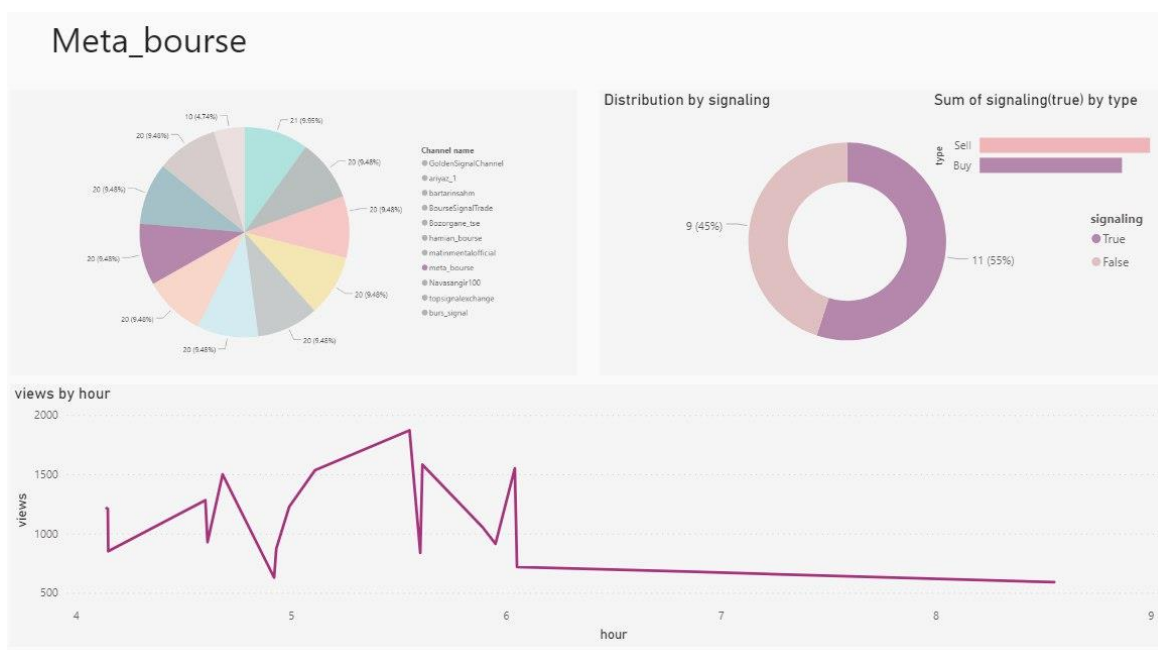
نمودار ۵-۲۰

۲-۳-۳۳ - نمودار تکی مربوط به کانال Matinmentalofficial



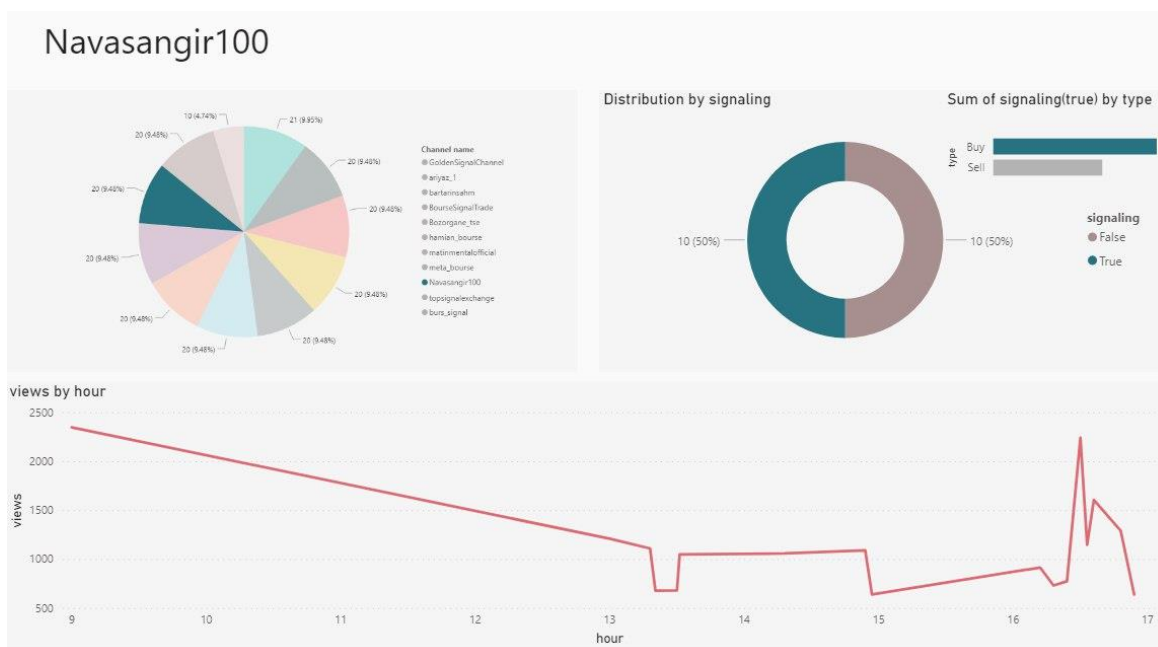
نمودار ۵-۲۱

۳۴-۲-۳- نمودار تکی مربوط به کانال Meta_bourse



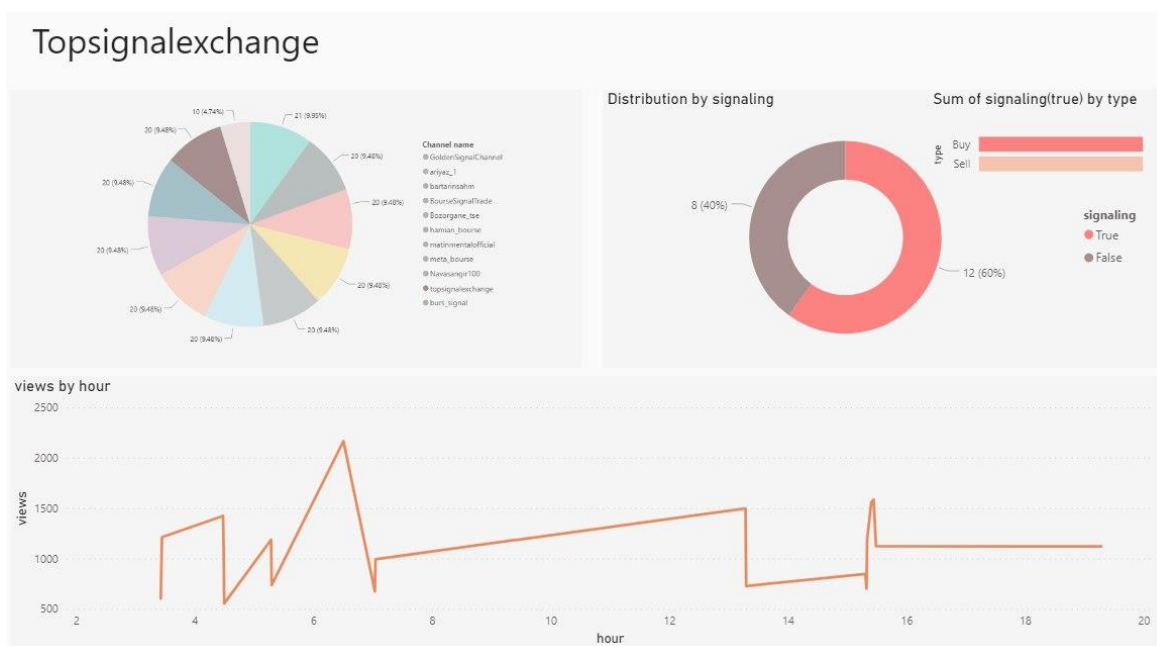
نمودار ۵-۲۲

۳۵-۲-۳- نمودار تکی مربوط به کانال Navasangir100



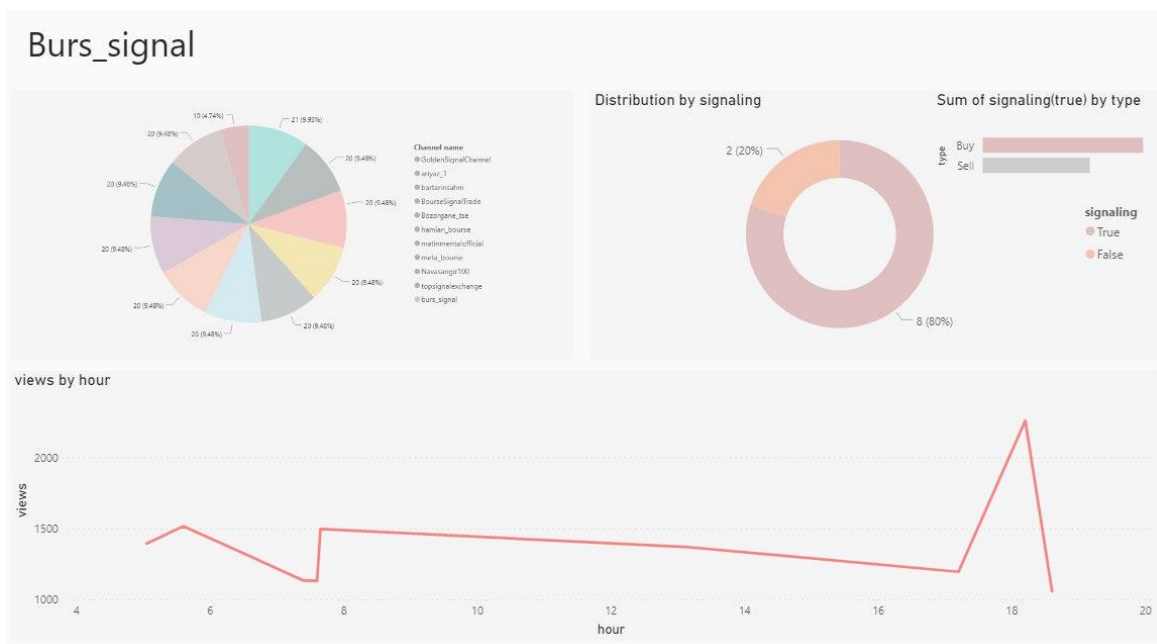
نمودار ۵-۲۳

۲-۳-۳۶ - نمودار تکی مربوط به کانال TopSignalExchange



نمودار ۲۴-۵

۲-۳-۳۷ - نمودار تکی مربوط به کانال Burs_signal



نمودار ۲۵-۵

فصل ششم

جمع‌بندی و نتیجه‌گیری و پیشنهادات

جمع‌بندی و نتیجه‌گیری

در این پروژه اثبات کردیم که امکان پایش رویدادهای فضای مجازی امکان‌پذیر می‌باشد. همچنین چندین روش من‌جمله تحلیل گروهی و تکی را نیز معرفی کردیم. این پروژه به علت اینکه در آینده دور می‌تولند بیشتر تخلفات را در بورس اوراق بهادار تشخیص دهد اهمیت بسزایی دارد. همچنین با انجام پژوهش‌ها و پروژه‌های آینده می‌توان این پروژه را گسترش داده و برای تمامی شبکه‌های اجتماعی اعمال کنیم.

پیشنهادهای

نیازمند به پژوهش‌های بیشتری در این زمینه هستیم چراکه اعتماد عمومی به بورس اوراق بهادار کلید موفقیت این بازار می‌باشد و پژوهش‌هایی از این دست به محقق شدن اعتماد عمومی کمک بسزایی می‌کند.

منابع و مراجع

- [1]
ایران بورس آنلاین، "سیگنال‌های لحظه‌ای فارکس، اونس طلا و نفت خام"
<https://iranbourseonline.info/forex-signals>
- [2]
"Home." <https://www.stlouisfed.org/>
- [3]
<https://khanesarmaye.com/>، "خانه سرمایه"، آموزش سرمایه گذاری در بورس
- [4]
آموزش فارسی، "یا سبد سهام چیست؟ (Portefeuille: پرتفوی (به فرانسوی)، colorlib
<http://amoozeshfarsi.com/portfolio/>
- [5]
"PDF + تابلو خوانی و بازار خوانی بورس چیست - آموزش کامل و حرفه ای تابلو خوانی بورس"
<https://chartiran.com/2364/tablokhani/>، می ۲۰۲۰، چارت ایران
- [6]
/ بورس ۲۴، آخرین اخبار و تحلیل های بورس تهران، "تحلیل ها سیگنال ۲۴"، Tarnama.ir
[Bourse24. https://www.bourse24.ir/articles](https://www.bourse24.ir/articles)
- [7]
چارت ایران، "چارت ایران - سایت آموزش بورس از مقدماتی تا پیشرفته رایگان"
<https://chartiran.com/>
- [8]
اکتبر ۲۰۲۱، همشهری آنلاین، "چگونه در بورس سود کنیم؟"
<https://www.hamshahrionline.ir/news/633188/چگونه-در-بورس-سود-کنیم>

[9]

“سامانه سایت کدال چیست و چه کاربردی برای سهامداران دارد؟”

<https://www.sdbroker.ir/article/172/%D8%B3%D8%A7%DB%8C%D8%AA-%DA%A9%D8%AF%D8%A7%D9%84>

[10]

سیگنال “چیست؟ آموزش تصویری سایت کدال (codal.ir) سایت کدال”

<https://isignal.ir/tutorial/معرفی-سایت-کدال/>

[11]

ژوئن 1، آموزش فارسی “چیست و چه سهامی در بورس معامله می شود؟ (Share) سهام”، colorlib

۲۰۲۰. <http://amoozeshfarsi.com/what-is-share/>

[12]

سامانه کپی ترید سهم تو “سهم تو | بهترین های بازار ارز دیجیتال و بورس در یک نگاه”

<https://sahmeto.com>

[13]

جولای 14، فرادرس - مجله “(سیگنال بورس چیست ؟ — به زبان ساده (+ فیلم آموزش رایگان)

۲۰۲۰. <https://blog.faradars.org/سیگنال-بورس-چیست/>

[14]

آوریل ۲۰۲۰ 13، آموزش مفید “کارگزاران بورس چه وظایفی برعهده دارند؟”

https://learning.emofid.com/what_is_brokers_job/

[15]

کدال چیست - آموزش کامل سامانه سایت کدال ۳۶۰ در بورس - آموزش کدال خوانی با فیلم و

<https://chartiran.com/2817/codal/> ژوئن ۲۰۲۰ 22، چارت / ایران، PDF

[16]

سیگنال “یادگیری صفر تا صد آموزش بورس | معرفی نمادها و ارائه تحلیل | سایت سیگنال”

<https://isignal.ir/سهام/>

Abstract

In recent years, the stock market has attracted a lot of attention. The reason for this can be from economic conditions to suitable investment conditions in the stock market. That's why people with less experience in the stock market were looking for people who would be their mentor in this new market

Many people in social networks with the name of stock market analyst make offers to buy or sell stocks. Sometimes there are people who take advantage of shareholders by abusing "confidential information" and "signaling". This abuse undermines public confidence in the stock market

In this project, we review several plans to investigate and detect fraud in the stock market. The design proposed in this project is the recent and conventional methods of the digital world. In this project, in addition to the algorithms of these methods, we examine ways to gain people's trust in the stock market

Key Words: Stock market, Signaling, Fraud detection, Confidential information, Monitoring



Amirkabir University of Technology
(Tehran Polytechnic)
Department of Computer Engineering

Project Report

**Monitoring cyber events to detect fraud in stock
market**

By
Bardia Ardakanian
Mohammadreza Hassanzadeh
Moujan Mirjalili
Farhan Keihan

Supervisor
Dr. Meysam Nazariani

Advisor
Dr. Meysam Nazariani

June & 2022