**فصل اول مقدمه**

**۱-۱ اهمیت و ضرورت تحقیق**

در سال ۲۰۰۹ با انتشار مستند یک شبکه مالی نظیر به نظیر1 تحت عنوان "بیت‌کوین: یک سیستم پول الکترونیک نظیر به نظیر2" توسط شخص گمنامی به نام ساتوشی ناکاموتو3 فصل جدیدی در سیستم‌های مالی و اقتصادی جهان آغاز شد. پس از روی کار آمدن شبکه رمزارز دیجیتال بیت‌کوین4 و با توجه به ساختار تبادل مالی این شبکه علاقه مندی و فراگیری، استفاده از تکنولوژی زنجیره‌های بلوکی5 تحت این ساختار به شدت افزایش یافت و از سوی دیگر این امر باعث به وجود آمدن نسل جدیدی از سیستم‌های نقل و انتقال مالی و همچنین ارزهای مورد تبادل بین مردم شد.

مهم‌ترین و منحصر به فرد ترین ویژگی‌های اصلی شبکه بیت‌کوین را میتوان این گونه بیان کرد:

۱- امنیت بالا در نقل و انتقال تبادل مالی بین دو شخص

۲- حذف شخص ثالث در کنترل شبکه مالی و نقل و انتقالات بین اشخاص

۳- شفافیت در انتقالات

پس از ظهور بیت‌کوین و افزایش علاقه‌مندی‌ها به تکنولوژی شبکه‌های زنجیره بلوکی رمزارزهای دیجیتال بسیار زیادی به وجود آمد که تا فبریه ۲۰۲۱ درحدود ۴.۵۰۱ رمزارز درحال فعالیت می‌باشند.

(تصویر از سایت statista.comافزوده شود   
<https://www.statista.com/statistics/863917/number-crypto-coins-tokens/>)

در اوایل سال ۲۰۱۵ میلادی نیز نسل جدیدی از رمزارزهای دیجیتال تحت عنوان رمزارزهای با ثبات6 روانه بازار رمزارزهای دیجیتال شدند که سعی در تثبیت نرخ تبدیل خود با یک ارز واحد دیگر همچون دلار را داشتند. یکی از نمونه‌های موفق در این حوزه را نیز می‌توان تتر18 معرفی کرد که تا به این لحظه توانسته است نرخ تبدیل خود به یک دلار آمریکا را تا حدودی ثابت نگهدارد و تغییرات بسیار شدیدی را درون خود مشاهده نکرده است. درواقع میتوان گفت یکی از چالش‌های اصلی تا به آن لحظه در بازار رمزارزهای موجود عدم ثبات نرخ تبدیل آن رمزارزها با دارایی دیگری همچون دلار آمریکا بود، به گونه‌ای که نرخ تبدیل هریک از ارزهای دیجیتال با افزایش میزان درخواست آن‌ها به صورت سعودی افزایش می‌یافت ویا با کاهش درخواست‌‌ها نرخ تبدیل آن‌ها کاهش چشم‌گیری داشتند. از طرفی دیگر برای همه گیر شدن استفاده از ارزهای دیجیتال در زندگی روزمره مردم نیاز بود تا بتوان نرخ تبدیل این دارایی‌ها به دارایی‌های دیگر ثابت بماند تا میزان ریسک استفاده از آن‌ها در زندگی روزمره کاهش یابد به همین منظور محققان حوزه کامپیوتر و اقتصاد با ارائه روش‌های مختلف سعی در ایجاد شبکه‌های زنجیره بلوکی‌ای داشتند تا بتوانند نرخ تبدیل این ارزهارا با استفاده از الگوریتم‌های پیاده‌سازی شده در شبکه‌های زنجیره بلوکی کنترل کنند.

در [1] به این موضوع اشاره شده است که دلایل اصلی که تحلیل و برسی داده‌های شبکه‌های زنجیره بلوکی از اهمیت بالایی برخوردار است را در سه موضوع اشاره کرده است:

1. تشخیص تقلب و پول‌شویی در شبکه‌های مالی زنجیره‌بلوکی
2. افزایش کارایی و تحقیق و برسی کردن میزان کارایی یک شبکه زنجیره بلوکی
3. تحلیل و برسی میزان دسترس پذیری به اشخاص و میزان پنهان بود هویت آن‌ها

یکی از چالش‌های اصلی‌ای که محققین این حوزه با آن رو به رو بودند اولا عدم دسترسی به داده‌های تبادل مالی رمزارزهای متفاوت درکنار یک دیگر بود و همچنین اطمینان از صحت و درستی آن داده‌ها می‌باشد. از طرفی دیگر موضوع مهمی که حتی در دنیای اقتصاد نیز یک چالش بزرگ به حساب می‌آید این است که بتوانند با استفاده از داده‌های واقعی، یک بازار مالی را شبیه‌سازی کنند، تا شناخت بهتری نسبت به آن بازار و قوانین حاکم بر آن پیدا کنند. این امر را می‌توان یکی از مهم ترین موضوعات و چالش‌های پیش‌روی بازارهای مالی سنتی و بازارهای مالی ارزهای دیجیتال دانست.

هدف از انجام این طرح تحقیق ارائه مجموعه داده جدید از تبادلات مالی رمزارزهای دیجیتال درکنار یکدیگر و شناخت بهتر کاربران شبکه تبادل دارایی توزیع شده به نام استلار با استفاده از روش‌های یادگیری ماشین می‌باشد، تا درک بهتری نسبت به نحوه فعالیت کاربران این حوزه و نیز دست‌بندی کاربران از نظر نوع رفتار آن‌ها داشته باشیم.

**۲-۱ بیان مسئله**

همانگونه که پیش‌ از این نیز اشاره شد شناخت بازارهای اقتصادی و بخصوص رمزارزها با استفاده از داده‌های تبادلات آن‌ها که قابل اطمینان باشد و همچنین بتوان داده‌های تبادل رمزارزهای متفاوت را در کنار یک دیگر داشته باشیم یک چالش بسیار بزرگ در حوزه تحقیق بازار رمزارزهای دیجیتال می‌باشد.

در این طرح تحقیق ما در ابتدا داده‌های شبکه استلار12 را جمع‌آوری کرده و به معرفی و توصیف آن می‌پردازیم، پس از آن نیز سعی میکنی با استفاده از روش‌های خوشه‌بندی داده‌ها کاربران فعل در این شبکه‌ را به دسته‌های مختلفی تقسیم کرده و شناخت بهتری نسبت به کاربران این سیستم ارائه دهیم.

**۳-۱ چالش‌های تحقیق**

در مقاله [1] به این موضوع اشاره شده است که یکی از چالش‌های اصلی در شبیه‌سازی و یادگیری رفتار کاربران شبکه‌های اقتصادی پیچیدگی این سیستم‌ها هستش که امکان شناخت هرچه بهتر قوانین حاکم بر این نوع از سیستم‌ها را با دشواری رو به رو می‌کند. در ابتدای این طرح تحقیق یکی از چالش‌های اصلی فرآیند کار موضوع جمع‌آوری داده‌ها از شبکه‌ استلار بود.

شبکه‌ استلار برای برقراری ارتباط بین نرم‌افزارهایی که توسعه دهندگان مختلف ایجاد می‌کنند تا بتوانند با شبکه استلار تعامل داشته باشند، این تعامل را از طریق سرورهایی تحت عنوان سرورهای چشم‌انداز14 (horizon) برقرار می‌کند. مشکل اصلی در دریافت اطلاعات از طریق این سرورها محدودیت تعداد درخواست در هرثانیه می‌باشد، به گونه‌ای که این سرورها در هر ثانیه تنها تعداد ۱۵ درخواست را پاسخ ‌‌می‌دهند برای رفع این چالش ما با استفاده از سیستم‌های کامپیوتری متعدد سعی در دریافت و جمع‌آوری داده‌ها کردیم تا در زمان کوتاه‌تر بتوانیم حجم داده بیشتری را دریافت کنیم.

پس از رفع این مشکل و دریافت داده‌های ۱ سال تراکنش‌های انجام شده در شبکه استلار چالش بعدی در پردازش داده‌ها بود. در این مرحله به دلیل حجم بالای داده‌های موجود ما در ابتدا تنها داده‌های سه ماه آخر سال ۲۰۱۹ را مورد تحلیل و برسی قرار دادیم و برای کاهش هزینه پردازش داده‌های سه ماه آخر نیز مراحل پردازش‌ را به قسمت‌های مختلفی تقسیم کرده و با استفاده از روش برنامه نویسی تقسیم و غلبه و چندپردازنده‌ای سعی در پردازش داده‌های خود کردیم.

**3-1 جنبه‌های نوآوری**

* **جدید بودن مجموعه داده:**

مجموعه داده‌ جمع‌آوری شده از شبکه تبادل مالی استلار یک مجموعه داده بسیار جدید است از تبادلات مالی لحظه‌ای کاربران این شبکه درخصوص تبادل دارایی‌های مختلف همچون ارزهای دیجیتال و ارزهای سنتی رایج با یکدیگر که تا به این لحظه در تحقیقات دیگر از این مجموعه داده استفاده نشده است و گام بزرگی را در رفتار شناسی کاربران این شبکه ارائه می‌کند چرا که مزایا و ویژگی‌های این مجموعه داده‌ را فصل‌های ۳ و ۴ به تفصیل مورد تحلیل و برسی قرار می‌دهیم.

* **استفاده از روش‌های مختلف خوشه‌بندی داده‌ها در تحلیل رفتار کاربران**:

در تمامی طرح‌های تحقیق پیشین توجه خاصی به پتانسیل و توانایی روش‌های مختلف خوشه‌بندی موجود در شناخت رفتار کاربران بازارهای رمزارزهای دیجیتال نشده است این موضوع در حالی است که در حوزه بازارهای سنتی تحقیقات فراوانی بر روی تحلیل بازارهای اقتصادی و رفتار کاربران از روش‌های خوشه‌بندی استفاده شده است، در این طرح تحقیق نیز برای اولین بار ما با بهره‌گیری از این الگوریتم‌ها سعی در شناخت هرچه بهتره رفتار کاربران در بازارهای ارزهای دیجیتال داریم.

* **ارائه راهکار برای پردازش ‌داده‌های بزرگ با استفاده از منابع پردازشی کم:**

در تمامی مراحل این طرح تحقیق ما با چالش اصلی کمبود منابع پردازشی رو به رو بودیم و از آنجایی که تعداد تراکنش‌های یک باز سه ماهه چیزی در حدود ۳ میلیون تراکنش بوده که ما با استفاده از روش‌های برنامه نویسی تقسیم و غلبه و همچنین تکنیک چندپردازنده‌گی سعی در تقسیم حجم پردازش در قسمت‌های مختلف و کاهش زمان پردازش کردیم.

**۴-۱ ساختار پایان نامه**

در این رساله در فصل دوم ما به توصیف و توضیح مفاهیم پایه‌ای شبکه‌های زنجیره‌بلوکی، توصیف آن‌ها در شبکه مالی استلار که مبتنی بر زنجیره‌بلوکی است و در انتها به توضیح موضوعات پایه‌ای که در این پایان نامه مطرح شده‌اند می‌پردازیم. در فصل سوم مروری بر کارهای مرتبط پیشین در چند‌سال اخیر می‌شود، در فصل چهارم در ابتدا به توصیف مجموعه‌داده جمع‌آوری شده از شبکه‌مالی استلار پرداخته و پس از آن به روش‌های پیاده‌سازی برای تحلیل این مجموعه داده‌ می‌پردازیم. در فصل‌های پنج و ششم نیز به ترتیب به ارزیابی کار خود و مقایسه ‌آن با کارهای پیشین پرداخته و پس از آن نتیجه‌گیری بدست آمده از این طرح تحقیق و کارهای پیش رو را بیان می‌کنیم.

**فصل دوم مفاهیم پایه**

**۱-۲ مقدمه**

در این بخش مفاهیم پایه استفاده شده در این رساله را توصیف می‌کنیم تا درک بهتری نسبت به مسائله مطرح شده در ادامه داشته باشیم. در ابتدای این بخش مفاهیم پایه‌ای درخصوص شبکه‌های زنجیره بلوکی همچون،‌ الگوریتم‌ اجماع،‌ دفترحساب کل، تراکنش‌ها می پردازیم و در ادامه این موضوعات را در شبکه‌ زنجیره بلوکی استلار بیان می‌کنیم، همچنین در ادامه عملیات‌ها در شبکه استلار، عملیات پیشنهاد خرید ویا فروش یک دارایی و به جزییات یک پیشنهاد خرید ویا فروش یک دارایی می‌پردازیم.

**۲-۲ تعاریف و اصلاحات**

در این بخش در ابتدا ما مفاهیم شبکه‌های زنجیره بلوکی را به صورت عام منظوره توصیف می‌کنیم و پس از آن، این مفاهیم را در شبکه تبادل‌مالی استلار توضیح می‌دهیم.

**۱-۲-۲ مفاهیم شبکه‌های زنجیره بلوکی به طور عام**

* دفترحساب کل توزیع شده

در شبکه‌های زنجیره بلوکی برای ذخیرسازی و ثبت رویدادهای شبکه از مفهمومی تحت عنوان دفترحساب‌کل استفاده می‌کنند که تحت عنوان یک پایگاه‌داده‌ای1 در ساختار شبکه‌های زنجیره‌بلوکی عمل می‌کند و بین کاربران ویا به عبارتی دیگر گره‌های2 داخل شبکه بلوکی توزیع شده است و با یکدیگر همسان می‌باشند.

* زنجیره بلوکی

در یک شبکه نظیر به نظیر مبتنی بر زنجیره بلوکی همانگونه که پیش از این توضیح داده شده رویدادها و داده‌های آن‌ها در داخل یک دفتر‌حساب کل ذخیره‌سازی می‌شوند. اجزا‌ء این دفترحساب کل را بلوک‌هایی از داده‌ها ویا تراکنش‌ها3 تشکیل‌ می‌دهند. این مجموعه بلوک‌ها که در داخل دفترحساب‌‌های کل ذخیره‌سازی شده‌اند به وسیله یک فیلد تحت عنوان هَش به یکدیگر متصل می‌شوند که درواقع مقدار این فیلد با درهم‌سازی سربرگ4 بلوک‌های قبلی به یکدیگر متصل می‌شوند و به اتصال این بلوک‌های ذخیره شده در داخل دفتر‌حساب‌های کل (بسته به ساختار شبکه زنجیره بلوکی در یک یا چند دفترحساب کل) تشکل یک زنجیره را می‌دهند که به آن‌ها زنجیره بلوکی می‌گویند.

* الگوریتم‌ اجماع

پیش از این نیز به این موضوع اشاره کردیم که در شبکه‌های زنجیره بلوکی گره‌ها به صورت نظیر به نظیر به یکدیگر متصل هستند و با استفاده از دفترحساب‌های کل اطلاعات ذخیره‌سازی می‌شود و این دفترحساب‌های کل در بین گره‌های شبکه توزیع شده است. در این شبکه‌ها کاربران برای ثبت یک داده‌جدید بایستی بر سر ذخیره‌سازی آن داده با یکدیگر رای گیری کنند، این عملیات رای‌گیری درون شبکه‌های زنجیره بلوکی توسط الگوریتم‌های اجماع صورت می‌گیرد. الگوریتم‌های اجماع ساختارهای متفاوتی را باهم دارند و بر اساس ساختار هرکدام از آن‌ها این موضوع مشخص می‌شود که در انجام الگوریتم‌ اجماع درون شبکه چه تعداد کاربر رای مثبت دهند آن داده در دفترحساب‌ کل ثبت و ذخیره می‌شود.

**۲-۲-۲ مفاهیم شبکه‌های زنجیره بلوکی در شبکه استلار**

مفاهیم بیان شده در این بخش تمامی‌ آن‌ها برگرفته از مستندات و مطالب بیان شده توسط شبکه تبادل مالی استلار می‌باشد و هرگونه بروز رسانی در ساختار ویا مستندات این شبکه می‌تواند تعاریف بیان شده در این بخش را تحت تاثیر قرار دهد.

* **دفترحساب کل توزیع شده**

دفترحساب‌کل توزیع شده در شبکه استلار موقعیت کل شبکه را در زمان های مختلف ذخیره می‌کند و بین تمامی گره‌هایی که شبکه استلار را ایجاد می‌کند توزیع می‌شود. در این دفترحساب‌کل لیست تمامی کاربران، موجودی آن‌ها و تمامی سفارشات1 درون شبکه تبادل توزیع شده ذخیره می‌شود. در هر دور اجرا الگوریتم اجماع در شبکه‌ استلار، شبکه تصمیم می‌گیرد که کدام یک ازمجموعه تراکنش‌ها بایستی بر روی آخرین دفترحساب‌کل بسته شده1 بایستی اعمال شود و زمانی که آن مجموعه تراکنش تایید می‌شود یک دفترحساب‌کل بسته شده جدید ایجاد می‌شود. دفترحساب‌های کل با استفاده از الگوریتم‌های رمزنگاری2 به دفترحساب‌کل واحد دیگری متصل می‌شود و آن‌ها با یکدیگر تشکیل یک زنجیره را می‌دهند تا به دفترحساب‌کل پیدایش3 متصل شوند که دفترحساب کل‌ پیدایش اولین دفترحساب‌کل ایجاد شده در داخل شبکه می‌باشد. لازم به ذکر است که هر یک از دفترحساب‌های کل یک شماره دارند که در شبکه استلار این شمارنده به صورت بازگشتی4 عمل می‌کند و اولین آن شماره یک می‌باشد و به همین صورت افزوده می‌شود[2].

* **الگوریتم اجماع**

نام این الگوریتم قرارداد فدرال بیزانس4 می‌باشد که توسط توسعه‌دهندگان شبکه زنجیره بلوکی استلار ایجاد شده است. در این الگوریتم نیز برای حل مشکل بروز رسانی وضعیت‌های تکرار شده7 ، مانند یک دفترحساب تراکنش ویا درخت گواهی‌ها8 همانند الگوریتم قرارداد غیر فدرال بیزانس5 عمل می‌کند. برای این موضوع آن‌ها هر بروز رسانی را در یک لبه9 تعریف می‌کنند تا متوجه شوند که کدام وابستگی‌های بروزرسانی میانی می‌تواند استنباط شود[3].

* **تراکنش‌ها**

تراکنش‌ها در داخل شبکه استلار مجموعه‌ای از دستورات هستند که وضعیت دفترحساب‌های کل را تغییر می‌دهند. در واقع تراکنش‌ها مجموعه‌ای از عملیات‌ها (هرتراکنش می‌تواند بین یک تا ۱۰۰ عملیات را شامل شود) هستند که امضا شده و به داخل شبکه ارسال می‌شوند و بارای جلوگیری از ایجاد نقص درون تراکنش‌ها از الگوریتم‌ اجماع شبکه استلار9 استفاده می‌شود. در واقع تراکنش‌ها را می‌توانند شامل عملیات‌هایی همچون ویرایش حساب‌کاربری، ارسال عملیات‌‌های پرداخت و وارد کردن سفارشات در داخل شبکه تبادل مالی توزیع شده شوند[2].

* **عملیات‌ها**

در واقع تنها عناصری که در داخل شبکه توزیع شده استلار می‌توانند وضعیت دفترحساب‌های کل را تغییر دهند عملیات‌ها هستند که در داخل یک تراکنش قرار می‌گیرند (پیش از این نیز اشاره شد که یک تراکنش‌ می‌تواند یک ویا چندین عملیات را درخود جای دهد). در داخل شبکه‌ استلار ۱۳ نوع تراکنش مختلف وجود دارد که می‌توان لیست آن‌ها را در [4] مشاهده کرد که در این طرح تحقیق تنها عملیات‌های پیشنهادات خرید/فروش دارایی برای ما اهمیت دارد و داده‌های جمع‌آوری شده در این طرح تحقیق درواقع، همین عملیات‌های خرید ویا فروش در داخل شبکه تبادل مالی توزیع شده استلار می‌باشند.

* **عملیات خرید و فروش**

در واقع ایجاد، بروز رسانی ویا حذف یک پیشنهاد خرید ویا فروش یک دارایی درون شبکه تبادل مالی توزیع شده استلار توسط این عملیات‌ها انجام می‌شوند[5]. برای ایجاد یک پیشنهاد خرید ویا فروش یک دارایی در این شبکه تنها لازم است فیلدی تحت عنوان ایدی پیشنهاد11 را برابر با مقدار صفر قرار دهیم. یکی از نکات اصلی در این عملیات‌ها در شبکه تبادل مالی توزیع شده استلار این است که در این شبکه به تمامی پیشنهادات خرید ویا فروش دارایی، تحت عنوان یک پیشنهاد خرید دیده می‌شود. این موضوع از آن‌جا شکل می‌گیرد که اگر شخص الف بخواهد یک بیت‌کوین خود را در مقابل ۱۰۰۰ اتریوم به فروش برسد پیشنهادی که آن شخص در شبکه ایجاد می‌کند را تحت عنوان پیشنهاد خرید اتریوم ثبت می‌شود و همچنین اگر همان شخص مجددا پیشنهادی را بخواهد ثبت کند که در آن ۴۰۰ لایت‌کوین12 خود را در مقابل یک اتریوم تعویض کند آن پیشنهاد نیز مجدد تحت عنوان پیشنهاد خرید اتریوم مشاهده می‌شود.

این موضوع درواقع در شبکه استلار به این شکل بیان می‌شود که ما می‌توانیم بر اساس برداشت خود تمامی پیشنهادات ثبت شده در داخل شبکه تبادل مالی توزیع شده یا به شکل پیشنهادات خرید مشاهده کنیم ویا به شکل پیشنهادات فروش به آن‌ها نگاه کنیم چرا که تنها یک نوع از آن‌ها در شبکه ثبت شده و از بهم پیوستن دو پیشنهاد متفاوت توسط افراد متفاوت عملیات تبادل دارایی‌ها بین دو شخص متفاوت انجام می‌شود.

**۳-۲-۲ داده‌های سری زمانی**

در دنیای امروزه داده‌های سری‌زمانی از جمله داده‌های بسیار پر اهمیت هستند که رشد روز افزونی دارند، ساختار و ماهیت این نوع از داده‌ها چالش‌های متفاوتی را در جمع‌آوری و تحلیل آن‌ها به وجود می‌آورد، در واقع داده‌های سری‌زمانی به در کنار هم قرار گرفتن مجموعه‌ای از توالی‌های پشت سر هم در زمان گفته می‌شود که درصورتی که شما بخواهید این داده‌ها را با استفاده از نمودارهای رایج به نمایش بگذارید حتما یکی از بعد‌های شما را در این نمودار بعد زمان تشکیل می‌دهد. برخی از مثال‌های داده‌های سری‌زمانی را می‌توان توالی‌های ایجاد شده توسط سنسورهای یک ماشین، تبادلات مالی انجام شده توسط کاربران بازارهای مالی، لاگ1ایجاد شده توسط سیستم‌های امنیتی و … را مثال زد که هریک از این مثال‌ها خود به تنهایی از اهمیت بالای داده‌های سری زمانی را نشان می‌دهند.

**۴-۲-۲ پنجره‌های زمانی**

در سیستم‌های سری‌زمانی محققین برای آنکه بتوانند بازه زمانی مورد تحلیل خود را کوچک‌تر کنند و تحلیل‌های خود را بر روی سیستم مورد نظر خود انجام دهند از تکنیک‌های بازه زمانی استفاده می‌کنند. در [13] برای پیش‌بینی قیمت بیت‌کوین بازه‌های زمانی ۱۰ دقیقه و ۱۰ ثانیه را انتخاب کرده‌اند و به این نتیجه رسیده‌اند که در بازه‌های زمانی بیشتر میزان درستی تحلیل انجام شده و دانش بدست‌آمده از داده‌های بازارهای رمزارز بهتر می‌باشد و اگر بازه زمانی کوچکی انتخاب شود این موضوع باعث می‌شود تا میزان درستی تحلیل‌های انجام شده از صحت پایین‌تری برخوردار شود.

**۲-۲-۵ الگوریتم‌های خوشه بندی**

خوشه‌بندی یک تکنیک یادگیری‌ماشین است که گروهی از داده‌ها را در یک خوشه در کنار یکدیگر قرار می‌دهد به طوری که هرگروه قوانین خودشان را دارا می‌باشند[16].از دیدگاهی دیگر می‌توان اینگونه بیان کرد که اعضای یک خوشه‌4 که به وسیله‌ الگوریتم‌های خوشه‌بندی در کنار یکدیگر قرار گرفته‌اند از نظر میزان شباهت بایکدیگر بیشترین شباهت را دارا می‌باشند یا به عبارتی دیگر کمترین فاصله را با یکدیگر دارند.

* **سلسله مراتبی**

الگوریتم‌ سلسله مراتبی جزو پر طرفدار ترین الگوریتم در بین محققینی است که می‌خواهند داده‌های بدون برچسب را با روش‌های یادگیری ماشین ویا داده‌کاوی مورد تحلیل و برسی قرار دهند. به طور کلی الگوریتم‌های سلسله مراتبی به دو دسته کلی تقسیم می‌شوند: ۱- خوشه‌بندی تجمعی4 ۲- خوشه‌بندی تقسیمی5 . تفاوت اصلی این دو روش را می‌توان از این جنبه برسی کرد که در روش خوشه‌بندی تجمعی درخت تقسیم گره‌های یک مجموعه داده از سمت پایین به بالا مورد پیمایش قرار می‌گیرد و این موضوع درحالی است که در روش تقسیمی درخت تقسیم گره‌ها از بالا به پایین مورد پیمایش قرار می‌گیرد، به عبارتی دیگر می‌توان گفت در روش پایین اگر تعداد رکورد‌های یک مجموعه داده را N در نظر بگیریم در ابتدای شروع کار الگوریتم ما N خوشه مجزا داریم و در هر دور از اجرا الگوریتم سعی می‌شود تا خوشه‌ها را با یکدیگر ادغام کنیم تا در نهایت به ۱ خوشه واحد برسیم. این موضوع در روش‌های بالا به پایین به این شکل است که در ابتدای شروع الگوریتم ما یک خوشه با N عنصر عضو داریم و در هر دور از اجرا الگوریتم سعی می‌شود تا تعداد خوشه‌ها را افزایش داده و عناصر با میزان شباهت بیشتر را در خوشه‌های مجزا قرار بدهیم و این موضوع می‌تواند تا جایی ادامه پیدا کند که تعداد خوشه‌های موجود در داده‌ها برابر با تعداد رکورد‌های مجموعه داده شود.

* **Kmeans**

**6-2-2 الگوریتم‌های محاسبه فاصله دو بردار**

به طور کلی برای محاسبه فاصله بین دو بردار الگوریتم‌های متفاوتی وجود دارد از جمله روش اقلیدسی، منهتن، روش مینکوسکی و …. . از جمله معروف ترین روشی که در الگوریتم‌های خوشه‌بندی برای محاسبه فاصله بین دو بردار از آن‌ها استفاده می‌شود روش اقلیدسی است که درصورتی که فاصله دو بردار از یکدیگر امکان محاسبه فاصله آن دو بردار به این روش وجود دارد. اما محققین در اواخر ده ۹۰ میلادی در حوزه تشخیص گفتار3 توانستند با ترکیب روش برنامه نویسی پویا و انطباق دو بردار زمانی بر روی یکدیگر روش جدیدی را کشف کنند که در آن محاسبه دو بردار زمانی با طول‌های متفاوت انجام شود[17]. پس از آن در [18] نیز از روش ارائه شده در حوزه تشخیص گفتار برای محاسبه فاصله بین دو بردار زمانی در بازارهای اقتصادی و دیگر مجموعه داده‌هایی که ذات داده‌ای آن‌ها به صورت سری زمانی است بهره گرفته و برای محاسبه فاصله بین دو بردار فاصله زمانی از این روش استفاده کرده‌اند.

* **انطباق زمانی پویا**

هماهنگونه که پیش از این اشاره شد اولین بار از روش انطباق زمانی پویا4 در حوزه تشخیص گفتار استفاده شد و پس از آن از این الگوریتم بر روی داده‌های سری‌زمانی در حوزه‌های مالی نیز استفاده شد. فرآیند کلی این الگوریتم به این شکل است که سعی دارد دو بردار زمانی را به گونه‌ای بر یکدیگر منطبق کند به گونه‌ای که برخی فاصله‌ها کمینه شوند. به طور کلی می‌توان گفت فاصله هر دو بردار زمانی که توانایی تبدیل شدن به یک دنباله خطی را دارا می‌باشند را می‌توان، با استفاده از این روش محاسبه کرد. به عنوان مثال دو بردار S و T را ما داریم که فرآیند انطباق زمانی پویا در واقع جستجوی پیدا کردن یک الگوی انطباق را به جستجوی دنباله سری زمانی S برای نمونه‌های از نمونه T گفته می‌شود[18].

S = s1 + s2 + … + sn [۱۸]

T = t1 + t2 + t3  + … + tm [۱۸]

در واقع دنباله‌های S و T می‌توانند در قالب یک سطح5 ویا توری6 n-by-m بشوند، به طوری که هر نقطه از توری در قالب (i, j) برابر است با انطباق نقطه si  در بردار S با نقطه tj در بردار T می‌باشد[18]. یک فاصله انطباقی W دو بردار S و T را به یکدیگر انطباق می‌دهد به گونه‌ای که فاصله بین نقاط انطباقی با یکدیگر کمینه شود.

W = w1 + w2 + w3  + … + wk [۱۸]

برای فرموله کردن عملیات انطباق بردار زمانی می‌بایستی یک معیار محاسبه فاصله داشته‌ باشیم، که پیش از این معیار‌های فاصله متفاوتی توضیح داده شده‌اند، برای فرموله کردن این معیار‌های ما از نماد δ استفاده می‌کنیم. و به یکی از شکل‌های زیر آن معیار را تعریف می‌کنیم.

δ(i,j) = | si – tj | [۱۸]

δ(i,j) = (si – tj)2 [۱۸]

پس از تعریف معیار اندازه‌گیری فاصله بین دو نقطه می‌توانیم الگوریتم انطباق پویا را به شکل زیر بنویسیم.

DTW(S, T) = minw[Σpk=1 δ(wk)] [۱۸]

**۷-۲-۲ باکت**

در این طرح تحقیق برای از آن‌جایی که در بازارهای تبادل مالی سنتی در هر تبادل دارایی که در آن‌ها انجام می‌شود یک سوی هر تبادل یک دارایی ثبات می‌باشد، به عنوان مثال در بازار سرمایه ایران یک طرف تمامی تبادلات ریال است، و از آن‌ جایی که در بازار مورد تحلیل ما دارایی‌هایی متفاوتی با یکدیگر تبادل می‌شود، به عنوان مثال ممکن است شخصی دارایی بیت‌کوین خود را در ازای اتریوم تبادل کند و به طور هم زمان شخص دیگری دارایی لایت‌کوین خود را در ازای دلار آمریکا تبادل کند، ما برای ثبات کردن این موضوع داده‌های خود را به باکت‌هایی تقسیم کردیم که در هر باکت همواره یک دارایی در یک سوی تبادل قرار بگیرد. به عنوان مثال در باکت بیت‌کوین در هر طرف تبادل همواره یک سوی آن بیت‌کوین قرار می‌گیرد تا بتوان مقایسه‌های خود را با بازارهای سنتی و روش‌های پیشین پیش ببریم.

**فصل سوم**

**پیشینه پژوهش**

**۱-3 مقدمه**

در این فصل به پیشینه پژوهش‌های انجام شده در حوزه تحلیل بازارهای مالی سنتی و بازر رمزارزهای دیجیتال می‌پردازیم. همانگونه که ذکر شد پژوهش‌هایی که در این تحقیق مورد تحلیل و برسی قرار میگیرند در ابتدای امر به دو دسته تقسیم می‌شوند، ۱- پژوهش‌هایی که در حوزه تحلیل بازار رمزارزهای دیجیتال انجام شده‌اند. ۲- پژوهش‌هایی که در حوزه تحلیل بازارهای سنتی هستند.

بر این اساس مقالات پژوهش‌ها از سال ۲۰۱۷ میلادی تا سال ۲۰۲۱ میلادی مورد برسی قرار گرفته‌اند و هرکدام را به تفکیک مورد برسی قرار می‌دهیم.

**۲-3 استراتژی جستجوی مقالات**

تاریخچه پژوهش مربوط به سال‌های ۱۹۹۴ تا ۲۰۲۱ می‌باشد، اصلی‌ترین این مقالات در حوزه تحلیل رفتار بازارهای سنتی قرار می‌گیرد و در واقع پژوهش‌های اولیه در حوزه تحلیل رفتار بازارهای رمزارزهای دیجیتال از سال ۲۰۱۳ شروع می‌شوند[23]. پس از آن از سال ۲۰۱۶ به بعد از ابزارهای یادگیری‌ماشین در خصوص تحلیل رفتاری بازارهای ارزهای دیجیتال استفاده شده است. به همین منظور فرآیند جستجوی کارهای پیشین در دو حوزه مختلف انجام شده است، درخصوص جستجوی مقالات حوزه بازارهای تبادلات مالی سنتی از کلیدواژه‌های جدول-۱ استفاده شده است.

|  |
| --- |
| کیلد واژه |
| User Behavior in financial markets |
| Stock market behaviors |
| Traders clustering in stock markets |

جدول-۱ لیست کلید‌ واژه‌های استفاده شده درخصوص پیشینه پژوهش رفتار شناسی در بازارهای سنتی

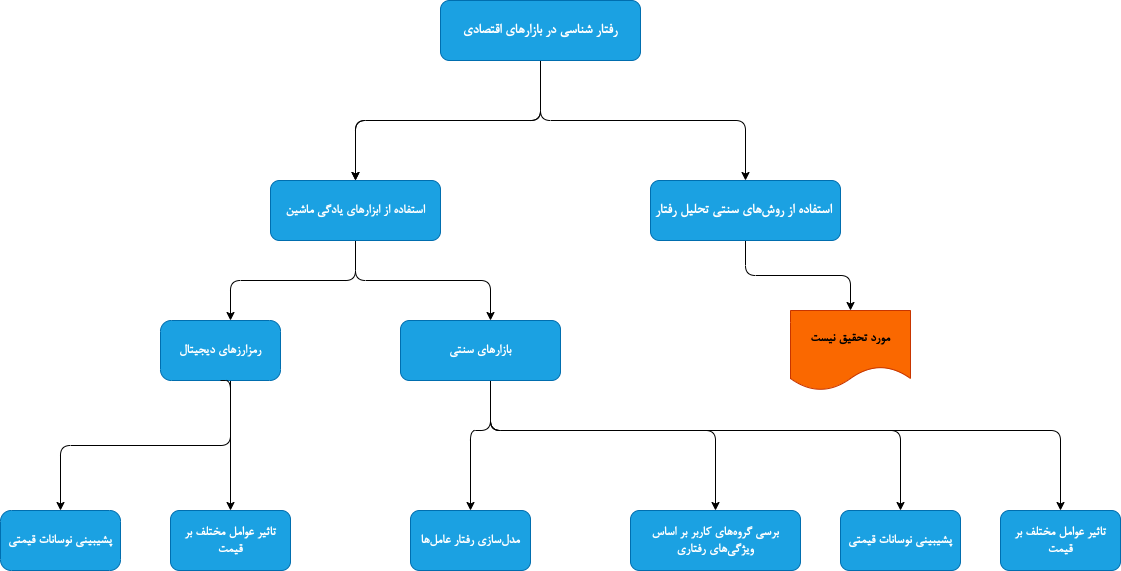
از میان مقالات جستجو شده با استفاده از ۳ کلید واژه مطرح شده در جدول-۱ تنها مقالات سال‌های ۲۰۱۷-۲۰۲۱ که با استفاده از روش‌های یادگیری ماشین عملیات رفتار شناسی را انجام می‌دهند استخراج شده و مورد تحلیل و برسی قرار گرفته‌اند.

کلید واژه‌های استفاده شده درخصوص پیشینه پژوهش رفتار شناسی در بازار ارزهای دیجیتال نیز در جدول-۲ ارائه شده است.

|  |
| --- |
| کیلد واژه |
| User Behavior in cryptocurrency markets |
| Stock market behaviors |
| Clustering traders in cryptocurrency markets |
| Behavior pattern clustering in cryptocurrency markets |

جدول-۲ کلید واژه‌ّای استفاده شده در خصوص پیشینه پژوهش‌ رفتارشناسی در بازار رمزارزهای دیجیتال

از میان مقالات جستجو شده با استفاده از ۴ کلید واژه مطرح شد در جدول-۱ نیز تنها مقالات سال‌های ۲۰۱۷-۲۰۲۱ اسختراج شده و مورد تحلیل و برسی قرار گرفته‌اند. به طور کلی موضوع تحلیل رفتاری بازارهای اقتصادی از جنبه‌های متفاوتی مورد پژوهش قرار می‌گیرد و هرکدام از این جنبه‌ها در جهت اهداف متفاوتی استفاده می‌شوند. در شکل-۱ فرآیند کلی تحلیل و ارزیابی بازارهای اقتصادی به تفصیل ارائه می‌شود.

شکل-۱ ساختار کلی تحقیقات رفتارشناسی در بازارهای اقتصادی

همانگونه که در شکل-۱ مشاهده می‌شود با استفاده از روش‌های یادگیری تحقیقات متفاوتی تا به این لحظه بر روی بازارهای سنتی انجام شده است و این موضوع درحالی است که تا به این لحظه برخی تحقیقات تا به این لحظه بر روی بازار رمزارزهای دیجیتال انجام نشده است و موضوعات بسیار پر اهمیت و تاثیر گذاری همچنان در این حوزه باقی مانده است، حوزه اصلی تحقیق انجام شده در این پژوهش مربوط به برسی گروه‌های کاربر بر اساس ویژگی‌های رفتاری آن‌ها در بازار است که تا به این لحظه در حوزه رمزارزهای دیجیتال انجام نشده است می‌باشد و از آن‌جایی که این موضوع تا به الان مورد تحلیل و برسی قرار نگرفته است مطالعات انجام شده درخصوص تحقیقات پیشین در این حوزه بخصوص موجود نمی‌باشد و در بخش بعد تنها به مرور کارهای پیشین رفتارشناسی در بازارهای دیجیتال می‌پردازیم که ارتباط چندانی با حوزه اصلی فعالیت این پژوهش ندارن اما از جمله فعالیت‌های تحقیقاتی در حوزه رفتارشناسی می‌باشند.

**۱-۲-۳ تحلیل بازار رمزارزهای دیجیتال**

در [5] با استفاده از داده‌های تبادلات انجام شده در شبکه‌ مالی بیت‌کوین و روش‌های خوشه‌بندی داده‌های شبکه بیت‌کوین که به صورت عمومی در دسترس اعضای شبکه بیت‌کوین وجود دارد تلاش شده است رفتارهای کاربران این شبکه را مورد تحلیل و برسی قرار دهد. در این تحقیق به این موضوع اشاره شده است که یکی از موفق‌ترین شبکه‌های رمزارزهای دیجیتال تا به این لحظه (۳۱ تیرماه ۱۳۹۹) بیت‌کوین بوده است و به این موضوع اشاره کرده است که یکی از اصلی‌ترین ویژگی‌های این شبکه نام مستعار1 کاربران در شبکه می‌باشد.

در واقع این ویژگی به کاربران این موضوع را ارائه می‌دهد که تنها با داشتند آدرس یک حساب کاربری در شبکه‌ بیت‌کوین نتوان به راحتی هویت صاحب آن حساب را بدست آورد. در همین جهت در این تحقیق سعی شده است با استفاده از آدرس‌های کاربران شبکه بیت‌کوین و روش‌های خوشه‌بندی داده‌ها کاربران موجود در شبکه بیت‌کوین را بر اساس ویژگی‌های رفتاری آن‌ها دسته‌بندی کند به گونه‌ای که آدرس‌های حسابی که در یک خوشه قرار می‌گیرند بر اساس متعلق به یک کاربر باشند.

در این تحقیق یک روش خوشه‌بندی اکتشافی2 (Heuristic) ارائه شده است که بر اساس دو ویژگی مهم در تراکنش‌ها جستجو کرده و آن‌ها را در خوشه‌های مختلف قرار داده‌است ولی این موضوع اشاره شده است که روش ارائه شده در این تحقیق نسبت به داده‌هایی که خوشه‌های بسیار زیادی را ایجاد می‌کنند درصد خطای این الگوریتم می‌تواند بالا برود.

در واقع منظور از خطا در این الگوریتم این است که تراکنش‌هایی را در یک خوشه قرار دهد که متعلق به یک کاربر در شبکه‌ بیت‌کوین نباشد.

در این تحقیق با استفاده از داده‌های بری منابع همچون توییتر3 و وب‌سایت‌های دیگر دسته‌بندی هایی را ایجاد کرده‌اند به گونه‌ای که در هر یک از دسته‌بندی های در نظر گرفته شده تگ‌هایی4 را در نظر گرفته‌اند که به داده‌های دسته‌بندی شده در یک خوشه نسبت ‌داده‌ می‌شوند و با استفاده از این داده‌ها ارزیابی‌های خود را مورد تحلیل و برسی قرار داده‌اند.

یکی از ویژگی‌های اصلی که در این تحقیق مورد برسی قرار گرفته‌ است و در فرآیند پژوهش ما تاثیر گذار است بیان این موضوع می‌باشد که با استفاده‌ از داده‌های تبادل دارایی در شبکه‌ بیت‌کوین توانسته‌اند رفتار کاربران را مورد تحلیل و برسی قرار دهند،‌ به گونه‌ای که کاربرانی که تراکنش‌هایی که ویژگی‌های نزدیکی با یکدیگر دارند را در یک خوشه‌ قرار داده و آن‌ها را تحلیل کرده‌اند.

اما یکی از چالش‌هایی اولیه وارده بر این تحقیق این موضوع است که در ابتدای امر تنها این تحقیق بر روی یک شبکه مورد برسی انجام گرفته است که در این تحقیق برای ما بسیار پر اهمیت و مهم بوده است.

در [8] با استفاده از داده‌های تراکنشی شبکه بیت‌کوین تلاش شده است دید بهتری نسبت به میزان الگوی رفتاری تراکنش‌ها در این شبکه بدست بیاورد و از سوی دیگر تلاش شده است که با استفاده از الگوریتم‌های خوشه‌بندی توانسته‌ است الگوی رفتاری که در تراکنش‌های شبکه بیت‌کوین وجود دارد را کشف کند. یکی از موضوعات بسیار مهمی که در این مقاله به آن اشاره شده است استفاده از الگوریتم‌های خوشه‌بندی مقایسه دودویی9 به جای استفاده از الگوریتم‌های مقایسه مرکزی10 می‌باشد. در واقع در الگوریتم‌های مقایسه دودویی تعداد خوشه‌های ما در ابتدای کار مشخص نمی‌شود و توسط خود الگوریتم تعداد خوشه‌ها تشخیص داده می‌شود و یکی از مزایای دیگری که این دسته از الگوریتم‌ها به نسبت به الگوریتم‌های خوشه‌بندی مرکزی همانند Kmeans دارند در این است که اولا تعداد ورودی داده‌ها در این الگوریتم‌ها اهمیتی ندارد و ثانیه سرعت و سربار پردازشی کمتری نیز دارند. در همین راستا در این مقاله محققین از الگوریتم Domain set برای شناخت رفتاری استفاده کرده‌اند و آن را با نتایج الگوریتم‌ Kmeans مقایسه کرده‌اند.

بر اساس یافته‌های این مقاله، مشخص شده است که میزان معیار Silhouette که بیان کننده میزان تفکیک خوشه‌ها از یک دیگر می‌باشد، در روش‌های مقایسه دودویی (Domain set) در مقابل روش‌های متمرکز‌(Kmeans) همواره میزان کمتری داشته است و این موضوع درحالی بیان شده است که الگوریتم‌ها داده‌ها را به خوشه‌هایی با تعداد ۳ تا۷ خوشه تقسیم کرده‌اند. اما در مقابل میزان معیار Rand Index که میزان درستی خوشه‌بندی داده‌ها توسط الگوریتم‌های خوشه‌بندی را نمایش‌ می‌دهد، روش‌ مقایسه دودویی (Domain set) همواره از روش مقایسه متمرکز (Kmeans) بیشتر بوده است.

یکی از چالش‌ها و موضوعات بحث برانگیز این مقاله حجم کم داده‌های انتخابی است که ۲،۰۴۸ نمونه را برای ارزیابی دو روش انتخاب کرده‌اند. همچنین در ادامه این موضوع را نیز می‌توان بیان کرد که درصورتی که تراکنش‌ها از شبکه‌های متفاوت گردهم بیایند آیا همچنان الگوریتم‌های مقایسه دودویی با روش‌های مقایسه متمرکز امکان رقابت کردن را دارند یا خیر؟

در [20] سعی شده است با استفاده از داده‌های شبکه بیت‌کوین و شبکه‌های عصبی هوش‌مصنوعی عملیات خرید و فروش بیت‌کوین را انجام دهند، در این مقاله محققان داده‌های روزانه از خرید و فروش بیت‌کوین را از یک بازار تبادل مالی در پنجره‌های زمانی ۱۵ جمع‌آوری کرده و داده‌های تکنیکال مورد نیاز برای خرید و فروش بیت‌کوین را از آن‌ها استخراج کرده‌اند تا بتوانند با استفاده از ارائه داده‌های تکنیکال بدست آمده از این مجموعه داده تحت عنوان داده‌های ورود به مدل خود نوسان بازه زمانی روز بعد بیت‌کوین را پیش‌بینی کنند.با اعمال مدل شبکه‌های عصبی بر روی مجموعه داده بدست آمده، نتایج آن‌ها نشان داده است که استفاده از شاخص‌های تکنیکال در فرآیند برچسب‌‌زنی بر روی داده‌ها با استفاده از شبکه‌های عصبی هوش مصنوعی می‌تواند، از انطباق بیش از حد4 (over-fitting) مدل بر روی داده‌ها در طی فرآیند برچسب زنی بر روی داده‌ها غیرثبات سری‌ زمانی بازارهای اقتصادی جلوگیری کند که این موضوع باعث بحبود نتایج تبادلات با روش ارائه شده در مقابل روش‌های تکنیکال سنتی هماهنند روش خرید-نگهداشتن5 دارایی می‌شود.

در [21] با استفاده از داده‌های تبادلات مالی رمرزارزهای دیجیتال در شبکه تبادل مالی بایننس5 و بینت‌فینیکس6 و استخراج داده‌های قیمت‌شروع، بالاترین قیمت، کمترین قیمت و قیمت پایان روز رمزارزهای دیجیتال مختلف با فرکناس بازه زمانی ۵ دقیقه سعی در پیداه سازی مدل پیش‌بینی کننده برای نوسانات بازارهای رمزارزها دارند، برای همین منظور برای ایجاد مدل یادگیرنده خود از جنگل‌تصادفی (Random Forest) استفاده کرده‌اند. در این مقاله محققان برای بحث مهندسی ویژگی‌ها از فاکتور alpha101 استفاده کرده‌اند، بر این اساس نتیجه گرفته‌اند استفاده از برخی از ویژگی‌ها در بازه‌های زمانی بیشتر باعث افزایش بهبود صحت7 مدل پیش‌بینی کنندشان شده است. پس از اعمال مدل جنگل تصادفی بر روی داده‌های خود نتیجه پیش‌بینی خود را با مدل‌های یادگیری‌عمیق8 با استفاده از ماتریس‌های F1 precision و recall مقایسه کرده‌اند و به این نتیجه رسیده‌اند که مدل جنگل‌تصادفی نتیجه بهتری نسبت به روش‌های یادگیری عمیق در پیش‌بینی نوسانات تغییر قیمت رمزارزها دارد. در انتها نیز محققان تلاش کرده‌اند با استفاده از مدل ارائه شده استراتژی تبادل بیت‌کوین برای کاربران با فرکانس بالا ارائه دهند.

در [22] سعی شده است با استفاده از الگوریتم SVM استراتژی خرید و فروش رمزارزهایی را ارائه بدهد تا بتواند به وسیله آن رمزارزها با میزان سود دهی بالاتر ویا پایین‌تر که توسط مدل ارائه شده پیش‌بینی شده اند را بدست بیاورد. نتایج این تحقیق نشان داد که مدل ارائه شده با استفاده از روش SVM چهار برابر بهتر از روش‌های استراتژی سنتی هماهنند S&P و B&H می‌باشد. در استراتژی S&P تبادل‌گر یک دارایی را در یک نقطه خریداری می‌کند و پس از آن برای سعود و نزول آن دارایی (سهام) یک کران بالا و کران پایین درنظر می‌گیرد، به این صورت که روبات تبادل‌گر با استفاده از این کران‌ها برای وضعیت آن سهام تصمیم می‌گیرد. به عنوان مثلا اگر دارایی مورد نظر از کران پایین مشخص شده توسط کاربر پایین‌تر بیاید روبات تبادل‌گر به صورت اتوماتیک آن سهام را به فروش می‌رساند. استراتژی B&H نیز این‌گونه عمل می‌کند که در آن تبادل‌گر اطلاعات بنیادینی از یک سهام را در اختیار دارد، به عنوان مثال از نقشه راه آن شرکت ویا اهداف آن شرکت و … اطلاعات دارد و بر این اساس سهام آن شرکت را در زمانی که قیمت پایینی دارد خریداری می‌کند و پس از گذشت چند ماه آن را می‌فروشد به ایم عمل، عمل خرید و نگهداری می‌گویند که در بازه نگهداری سهام سود بسیار زیادی بر اساس رشد آن سهام به تبادل کننده داده می‌شود. اما در این تحقیق، محققین این موضوع را بیان کرده‌اند که از آن جایی که روش SVM پارامتر‌های بسیار زیادی دارد و همچنین بسیار به over-fit کردن داده‌ها تمایل دارد، از همین رو مدل بهینه‌ای نمی‌باشد.

**۲-۲-۳ تحلیل بازار سنتی**

در [6] با استفاده از داده‌های تبادلات دارایی در بازار قراردادهای آتی چین (CFFEX۳) و روش‌های خوشه‌بندی سعی در شناخت رفتار کاربران این بازار کرده‌اند. در این تحقیق در ابتدا با استفاده از داده‌های ساختگی و بهره‌گیری از دو الگوریتم SLF-PARAFAC و

۴SPM این دو الگوریتم را مورد تحلیل قرار داده‌اند و پس از آن داده‌های یک ماه سال ۲۰۱۲ این بازار را جمع‌آوری کرده و از روی این داده‌ها ویژگی‌های زیر را بدست آورده و با استفاده از الگوریتم‌های ذکر شده این داده‌ها را مورد تحلیل و برسی قرار داده‌اند.

ویژگی‌های ایجاد شده از روی داده‌های جمع‌آوری شده:

* Number of trades
* Trading volume for each trader
* Change in inventory
* Cumulative net inventory
* Inter-trade duration

ویژگی‌های بیان شده در این قسمت به دلیل آن در این تحقیق نیز بر اساس این پژوهش ویژگی‌ها ایجاد شده‌اند در فصل ۳ به تفکیک توضیح داده‌ می‌شوند و مورد برسی قرار می‌گیرند.

در این پژوهش هر کدام از این ویژگی‌ها را در بازه های ۱۵ دقیقه ای برای هر یک از کاربران ایجاد شده‌اند و در الگوریتم‌های خوشه‌بندی ذکر شده مورد استفاده قرار گرفته‌اند.

در این مقاله تعداد خوشه‌هایی که برای تحلیل رفتار کاربران انتخاب شده است ۵ خوشه می‌باشد که از آنجایی که تحقیقات ارائه شده در پژوهش ما نیز بر پایه این مقاله می‌باشد هر یک از خوشه‌ها به تفکیک در فصل ۴ مورد تحلیل و برسی قرار می‌گیرند. موضوع پر اهمیت در این مقاله این است که در این مقاله در نهایت پس از خوشه‌بندی کاربران فعال در بازار مذکور با استفاده از دو ویژگی Avrage volume share وAverage inventory/trade ratio رفتار کاربران را مورد تحلیل و برسی قرار داده‌اند.

در واقع با استفاده از این ویژگی‌ها محققان نشان‌ داده‌اند که کاربرانی که در هر یک از خوشه‌ها قرار داده‌اند چه میزان از حجم بازار را به خود اختصاص داده‌اند و همچنین میزان ریسک پذیری کاربران هر خوشه به چه صورت می‌باشد.

یکی از چالش‌های اصلی در این تحقیق این موضوع می‌باشد که اولا باز زمانی انتخاب شده برای تحلیل رفتار کاربران یک بازار تبادل مالی به یک ماه محدود می‌شود و همچنین با توجه به ذات بازارهای سنتی امکان تحلیل کاربران چندین بازار با یکدیگر وجود ندارد که در واقع در پژوهش ما سعی شده است با تحلیل دارایی‌های متفاوت در باز زمانی بلندتری تحلیل بهتری از رفتار کاربران بازارهای مختلف و همچنین تحلیل بهتری بر کلیت رفتار آن‌ها نسبت با دارایی‌های مختلف را ارائه بدهیم.

در [7] به این موضوع اشاره شده است که تحلیل و پیش‌بینی5 بازارهای مالی که حجم زیادی تبادلات6 دارند، یک چالش بسیار بزرگ برای تحلیل گران بازارهای تبادل مالی و اقتصاد دانان است. به همین منظور در این پژوهش، دانشمندان مجموعه‌داده‌ای را از بازار Nasdaq Nordic stock جمع آوری کرده‌اند که شامل حدودا ۴ میلیون، داده‌های نرمال سازی شده، سری زمانی، تبادلات مالی در این بازار می‌باشد و پس از آن سعی شده است با استفاده از روش‌های با ناظر7 یادگیری ماشین به تخمین قیمت یک دارایی در این بازار بپردازن.

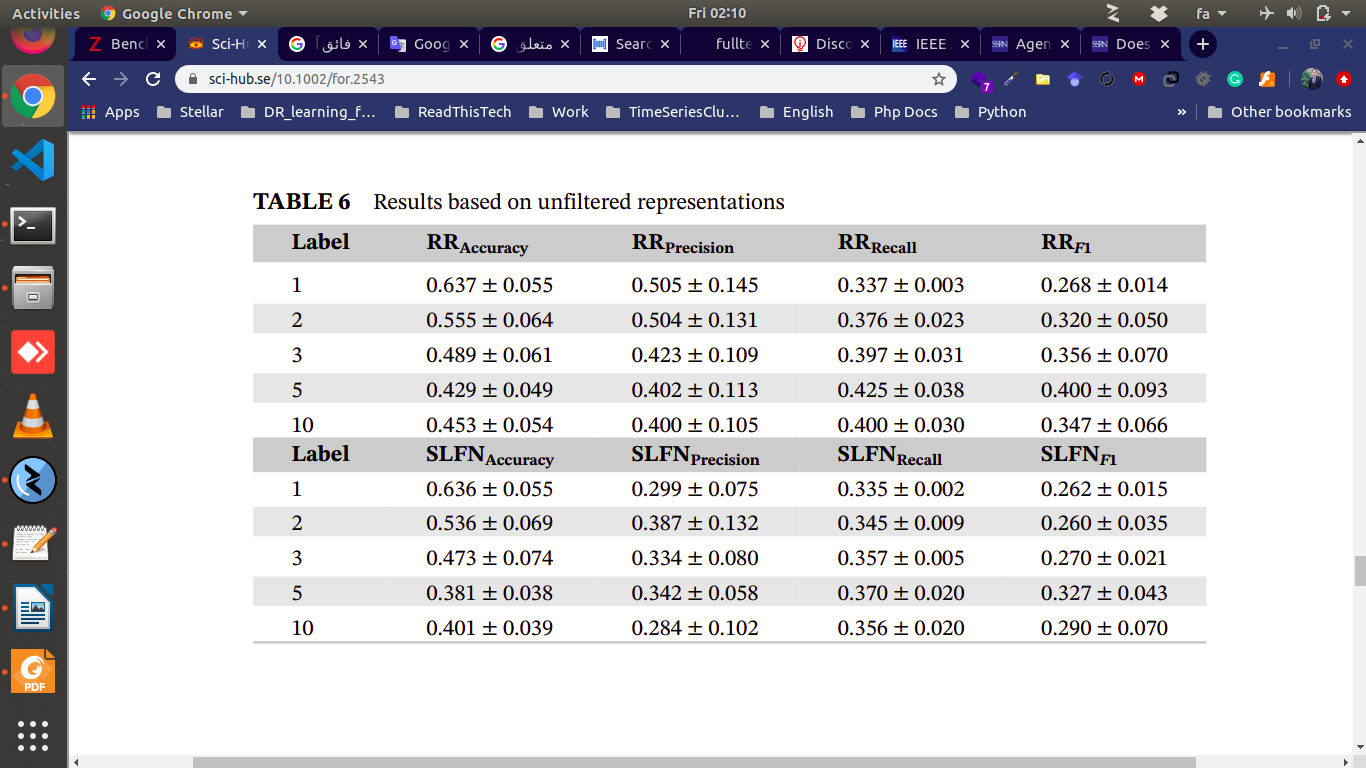
در این پژوهش نویسندگان داده‌های خود را بر اساس دو بازه زمانی اکتبر 2007 و ژانویه 2008 جمع آوری کرده‌اند و تحقیق خود را پیش برده‌اند. از آنجایی که یکی از چالش‌های اساسی در تحلیل داده‌های سری زمان با فرکانس بالا تعدد تراکنش‌ها در فاصله‌های زمانی میلی ثانیه‌ای وجود دارد،‌ محقیقین به این موضوع اشاره کرده‌اند که برای فائق آمدن بر این موضوع دو روش وجود دارد. یکی از این روش‌ها استفاده از بازه‌های زمانی با طول ثابت است به گونه‌ای که از ابتدای مجموعه داده‌ای تا انتها، آن را به بازه‌های زمانی کوچیک‌تری (در حدود چند دقیقا،‌ به عنوان مثال ۲۴ ویا ۱۵ دقیقه) تقسیم کنند و ویژگی‌های مورد نظر خود را در هر یک از این بازه‌ها به ازای هر یک از کاربران فعال در آن بازار محاسبه شود.

یکی دیگر از روش‌هایی که برای حل این مشکل مطرح است استفاده از بازه‌های زمانی با طول‌های متفاوت است، به گونه‌ای که در این روش بازهای زمانی بر اساس رویداد‌ها ایجاد می‌شوند و محاسبه ویژگی‌های مورد نظر برای هر یک از کاربران در این بازه‌ها انجام می‌گیرد. شایان ذکر است که با توجه به توضیحات مطرح شده در مقاله مذکور تنها در صورتی می‌توان داده‌ها مورد نظر خود را با استفاده از روش تقسیم بازه‌های زمان از طریق رویدادها به بازه‌های زمانی کوچیک‌تر تقسیم کرد که مجموعه‌ داده مورد استفاده رویداد‌های ایجاد شده در یک بازار را شامل شده باشد.

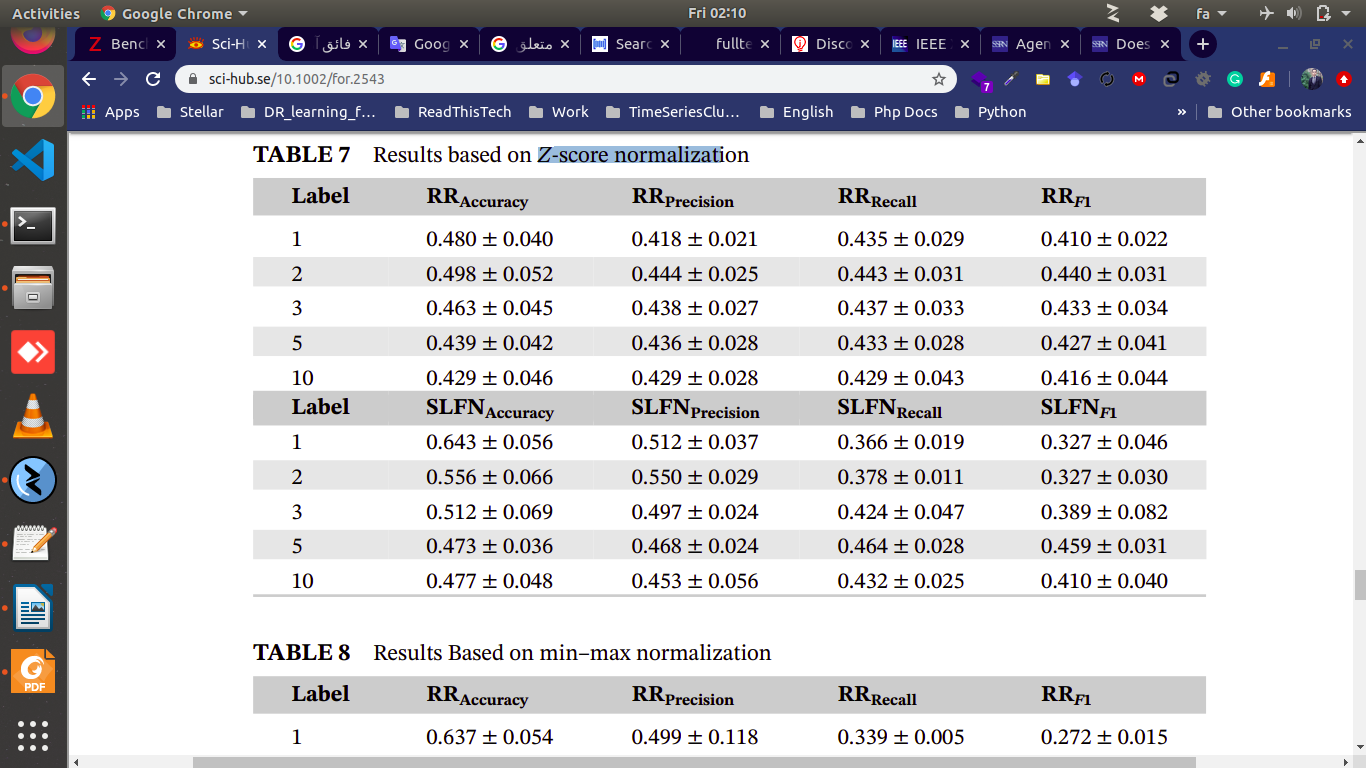
در این تحقیق نویسندگان این مقاله برای پیش‌ بینی حرکت‌های متوسط قیمت یک دارایی از دو مدل رگرسیون استفاده کرده‌اند که اولین مدل،‌ مدل Ridge regression بوده و دومین مدل استفاده در این تحقیق نیز SLF network-based nonlinear regression می باشد.

در این مقاله محققین نتیجه‌های خود را به چهار صورت متفاوت ارائه کرده‌اند، در ابتدا بدون اعمال فیلتر کردن دادها و پس از آن نتیجه‌های خود را بر اساس نرمال‌سازی داده‌ها با استفاده از روش‌های نرمال‌سازی Z-score ،min-max وdecimal precision ارائه کرده‌اند، که هدف اصلی آن‌ها از ارائه این چهار نتیجه متفاوت، این است که بر اساس قوانین مشخص شده توسط خود بازار Nasdaq Nordic stock داده‌هایی که به صورت عمومی در اختیار سایر محققین قرار می‌گیرد بایستی در ابتدا فیلتر شده و پس از آن نرمال‌سازی شود تا هیچ‌گونه امکان تشخیص هویت کاربران این بازار برای سایرین وجود نداشته باشد.

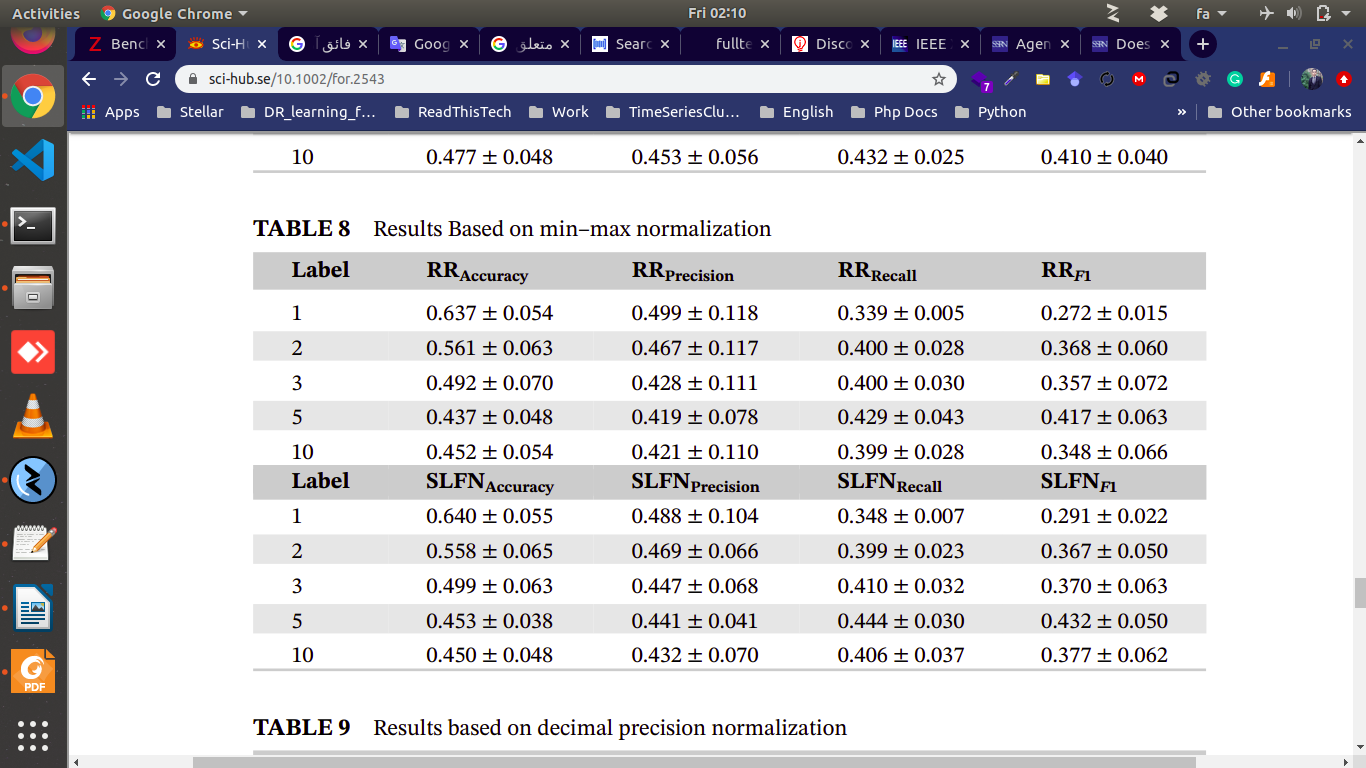
نتایج بدست آمده در این طرح تحقیق برای بیان میزان درستی مدل‌های اعمال شده بر روی این مجموعه داده بر اساس هر یک از برچسب‌های داده‌ای که در مجموعه داده‌های آن‌ها وجود دارد نیز به تفکیک روش‌های مطرح شده در بالا برای ارائه نتیجه‌ها به شرح جدول‌های ۱-4 می باشند.



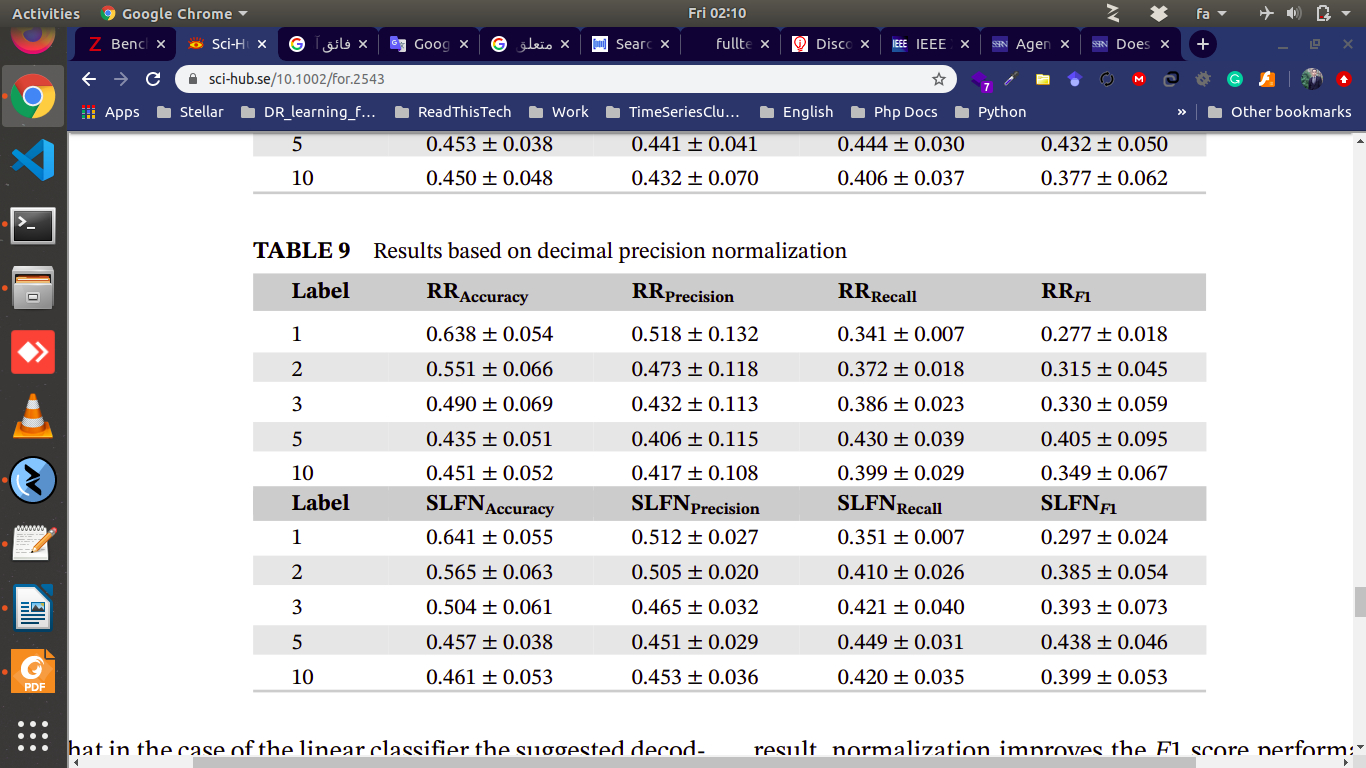
جدول-۱ نتایج [7] بدون فیلتر کردن داده‌ها.



جدول-۲ نتایج [7] با اعمال نرمال‌سازی Z-score .



جدول-۳ نتایج [7] با اعمال نرمال‌سازی min-max .



جدول-۴ نتایج [۷] با اعمال نرمال‌سازی decimal precision .

در [10] به این موضوع اشاره شده است که،‌ تحلیل و پیش‌بینی بازارهای تبادلات سهام یکی از چالش‌های پر اهمیت برای محققان این حوزه می‌باشد و اهمیت این موضوع از آن‌جا نشات می‌گیرد که در واقع بازارهای تبادلات مالی در یک جامعه نقش حیاتی را در رشد و تکامل بازار اقتصادی آن جامعه را ایفا می‌کنند. همچنین به این موضوع اشاره شده است که درخصوص تحلیل و خرید ویا فروش یک سهام در یک بازار تبادل مالی نیازمند شناخت و دانش‌کافی نسبت به آن سهام و همچنین بازار برای هر یک از تبادل‌کنندگان می‌باشد. این موضوع در این پژوهش به این شکل بیان شده است که تبادل (خرید/فروش) سهام در یک بازار سرمایه بدون داشتن دانش کافی همانند این است که گلف را با چشمانی بازی کنید.

در جهت تحلیل و پیش‌بینی نوسانات بازارهای تبادل مالی تحلیل‌ها و پژوه‌های فراوانی انجام شده است که با استفاده از ابزارهایی همچون هوش‌مصنوعی،‌ ابزارهای تحلیل تکنیکال،‌ استفاده از منطق فازی1 و بسیاری روش‌های دیگری تا به امروز انجام شده است. اما یک نکته بسیار پراهمیت در آن‌ها وجود داشته، که تحلیل و پیش‌بینی نوسانات قیمت را تنها در یک روز انجام می‌دادند اما در این روش با استفاده از روش خو‌شبندی ارائه شده توانسته‌اند نوسانات و تغییر قیمت برای روز بعد چندین سهام را با هم انجام دهند.

در این مقاله با استفاده از داده‌های ۴ماه اول و دوم سال ۲۰۱۸ از بازار تبادل‌ مالی، ملی هند که در هر یک از ماه‌‌ها تنها بازه ۱۰ روز را به عنوان بازه زمانی مورد تحلیل خود قرار داده‌اند و بدست آوردن ویژگی‌هایOHLC که شامل داده‌های قیمت‌ شروع یک سهام در ابتدای بازه زمانی2، حداکثر قیمت آن سهام در طول بازه زمانی3، حداقل قیمت‌ آن در طول بازه زمانی4 و همچنین قیمت بسته شده انتهای بازه زمانی5 و الگوریتم خوشه‌بندی BRICH برای تحلیل و گروه‌بندی داده‌های خود استفاده کرده‌اند و پس از آن با گرفتن متوسط نرخ تغییر سهام‌ها در هر گروه و برسی ‌آن با متوسط کلی نرخ نوسانات آن سهام تصمیم گرفته‌اند که آیا هر یک از سهام‌ها در روز بعد سعودی هستند ویا نزولی.

لازم به ذکر است که در این مقاله پیش از اعمال داده‌های خود به الگوریتم خوشه‌بندی،‌ آن‌ها را با ورش نرمال‌سازی Z score نرمال کرده و سپس داده‌های نرمال شده را به الگوریتم خوشه‌بندی خود داده‌اند.

در انتها در این پژوهش میزان درستی مدل پیشنهادی خود در تحلیل و پیش‌بینی نوسان یک سهام در روز بعد، روش‌ خود را با روش‌های معمول خوشه‌بندی و روش تحلیل تکنیکال با استفاده از نمودارهای شعمی ارزیابی کرده‌اند. بر اساس مشاهدات انجام شده در این مقاله،‌ روش‌ پیشنهادی این مقاله به طول کلی میزان درستی ۷۲.۹۹٪ را بدست‌آورد و این درحالی است که دو روش‌ دیگر به ترتیب میزان درستی ۵۰.۸۰٪ و ۵۰.۹۹٪ را بدست آورده است.

در این پژوهش، نقش مهم یادگیری ماشین و تاثیر آن در پیش‌بینی نوسانات مالی در یک بازار تبادل سرمایه به خوبه‌ نشان‌ داده‌شده است و همچنین برای اولین بار چندین سهام مختلف را در کنار یکدیگر مورد تحلیل و برسی قرار داده‌اند. اما یکی از چالش‌های اصلی این مقاله این موضوع است که بازه زمانی انتخابی برای یک پنجره‌زمانی را یک روز کامل که معادل با ۲۴ ساعت می‌باشد، در نظر گرفته شده است و این موضوع مطرح می‌شود که درصورتی که بازه‌های پنجره‌ زمانی داده‌ها اگر تغییر کند چه تاثیری در میزان پیش‌بنی نوسان سهام‌ها می‌گذارد.

در [11] این موضوع بیان شده است که تاثیر تبادلات مالی انجام شده توسط تبادل‌کنندگان با نرخ تبادل بالا1 بر روی نرخ تغییر قیمت در بازار چگونه می‌باشد و همچنین معکوس این موضوع را مورد تحلیل و برسی قرار داده‌اند. در همین جهت محققان داده‌های بازار تبادل مالی شیکاگو مرکٍنتیل2 را برای تحقیق خود انتخاب کرده و داده‌های قرارداده‌های عاطی E-Mini S&P 500 را در بازه‌های زمانی آگوست ۲۰۱۰ تا سپتامبر ۲۰۱۰ و همچنین آگوست – سپتامبر ۲۰۱۱ را جمع‌آوری کرده و مورد تحلیل قرار داده‌اند.

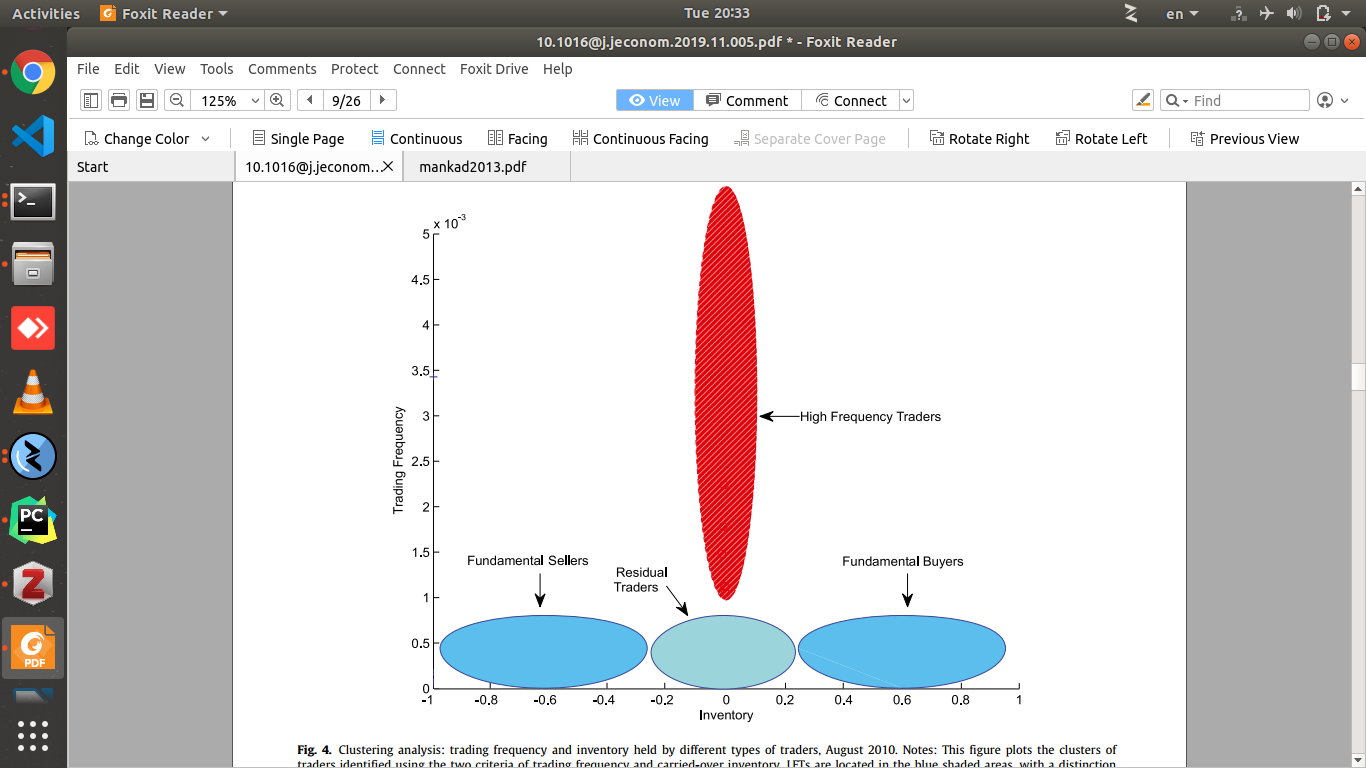
ویژگی‌های استفاده شده در این مقاله درخصوص ایجاد تفاوت در بین تبادل کنندگان با نرخ بالا و تبادل‌کنندگان با نرخ‌پایین2 شامل ویژگی‌های زیر می‌باشند:

* تعداد تراکنش‌های هرکاربر و حجم خرید ویا فروش هر کاربر در روز
* متوسط نرخ تبادل هرکاربر در روز در قالب میلی‌ثانیه
* درصد دارایی موجود کاربر در پایان روز

با توجه به توضیحات بیان شده در این مقاله برای ایجاد تفکیک در بین کاربران با نرخ تبادل بالا و کاربران با نرخ تبادل پایین از ویژگی‌های فوق استفاده شده است و این موضوع بیان شده است که کاربران با نرخ تبادل بالا همواره تعداد تبادلات بسیار بالایی در روز دارند و از سویی دیگر حجم تبادلات آن‌ها در بسیار بالا است و همین امر باعث می‌شود که نرخ تبادل این کاربران به نسبت کاربران با نرخ تبادل پایین کمتر باشد، از سوی دیگر این موضوع نیز بیان شده‌است که کاربران با نرخ تبادل بالا همواره تلاش ‌می‌کنند تا در انتهای روز درصد دارایی‌های موجود خود را بسیار اندک نگهدارند.

اولین قدم در این تحقیق ایجاد تفکیک در بین کاربران با نرخ تبادل بالا با کاربران با نرخ تبادل پایین است، در همین راستا برای خوشه‌بندی رفتار کاربران از الگوریتم‌های خوشه‌بندی استفاده کرده‌اند که الگوریتم‌های استفاده شده در این مقاله شامل الگوریتم سلسله مراتبی در خوشه‌بندی (Agglomerative Clustering) و الگوریتم جدا‌کننده3 (Divisive clustering) می‌باشد.

شکل-۱ در تکمیل بحث شناخت کاربران با نرخ تبادل بالا و کاربران با نرخ تبادل پایین می‌باشد که در آن بر اساس ویژگی‌ متوسط نرخ تبادل در میلی‌ثانیه در محور عمودی و حجم دارایی‌های موجود در انتهای روز در محور افقی قرار گرفته است و این تفکیک پذیری را به صورت شفاف تر بیان کرده است.



شکل-۱ نمایش قرار گیری گروه‌های کاربران مختلف بر اساس متوسط تبادل دارایی و حجم دارایی موجود در پایان روز می‌باشد[11].

در این تحقیق همچنین به این موضوع اشاره شده است که کاربران با نرخ تبادل دارایی بالا متوسط تعداد تبادل و همچنین حجم تبادل بیشتری نسبت به کاربران با نرخ تبادل پایین دارند و از سوی دیگر این کاربران انحراف از معیار کمتری را در این دو ویژگی دارا می‌باشند.

در قدم بعد پس از گروه‌بندی کاربران با استفاده از الگوریتم‌های با ناظر سعی در شناخت گروه‌های تراکنش‌ها را داشتن به گونه‌ای که مشخص کنند که هر تراکنش متعلق به کدام یک از گروه‌های فوق می باشد. در جدول-۱ انواع گروه‌های تراکنشی در این طرح تحقیق آورده شده است.

|  |  |
| --- | --- |
| HL | تراکنش بین یک کاربر با نرخ تبادل بالا و یک کاربر با نرخ تبادل پایین می‌باشد |
| HH | تراکنش بین دو کاربر با نرخ تبادل بالا می‌باشد |
| LH | تراکنش بین یک کاربر با نرخ تبادل پایین و یک کاربر با نرخ تبادل بالا می‌باشد |
| LL | تراکنش بین دو کاربر با نرخ تبادل بالا می‌باشد. |

جدول-5 برچسب‌های استفاده شده در [10] برای برچسب‌گذاری بر روی تراکنش‌ها

پس از برچسب‌گذاری بر روی تراکنش‌ها در آخرین قدم سعی بر شناخت تاثیرات تبادلات کاربران با نرخ تبادل بالا بر روی تغییرات پول در بازار را داشتند. بر اساس یافته‌های بدست‌امده در این مقاله مشخص شد که نوسانات و پرش‌های قیمتی در بازارهای تبادل مالی ناشی از فعالیت کاربران با نرخ تبادل بالا نمی‌باشد، این موضوع درحالی است که با فعالیت کاربران با نرخ تبادل بالا پدیده نوسان ویا پرش‌ قیمتی اتفاق نمی‌افتد بلکه این دو رویداد در هنگام غیبت این کاربران رخ می‌دهند، به عبارتی دیگر فعالیت کاربران با نرخ تبادل بالا باعث تثبیت نرخ تبدیل دارایی‌ها در بازار نیز می‌شود. از سویی دیگر بروز اختلالات در بازار خود باعث ایجاد نیروی محرکه در تشویق کاربران با نرخ بالا برای انجام تبادلات بیشتر در بازار می‌باشد.

در پایان نیز به این موضوع اشاره شده است که کاربران با نرخ تبادل بالا همواره شانس بیشتری را در پیروزی یک تبادل در بازار را نسبت به کاربران با نرخ تبادل پایین دارند، این موضوع درحالی است که این کاربران در مقابل یکدیگر شانس زیادی را برای پیروزی در تبادلات مالی را دارا نمی‌باشند و بر طبق یک قانون نانوشته در بازارهای تبادل مالی، کاربران با نرخ تبادل بالا همواره تلاش می‌کنند تا با کاربران با نرخ تبادل بالا، تبادلات دارایی انجام ندهند چرا که شانس پیروزی برای سود بیشتر آن‌ها در این صورت بسیار کم می‌باشد.

در [12] سعی شده است میزان تاثیر گزاری تبادل گران اختلال گر1 بر بازده غیرمتعارف2 را دریابند. به طور کلی منظور از تبادل گران اختلال گر در بازارهای سرمایه به تبادل کنندگانی گفته می‌شود که به صورت هیجانی در بازار دست به خرید ویا فروش می‌زنند و هیچگونه اطلاعات تکنیکال ویا بنیادی نسبت به وضعیت سهام‌ها در بازار را ندارند و بر اساس موج‌های خبری و نوسانات به صورت احساسی تصمیم می‌گیرند. در این مقاله با استفاده از داده‌های ایندکس SSE50 در بازار شانگ‌های که بیان کننده سهام ۵۰ شرکت برتر در این بازار می‌باشد که بازه مورد مطالعه این سهام ها از تاریخ یکم، ژانویه ۲۰۱۲ تا ۳۰ام دسامبر ۲۰۱۸ می‌باشد. درخصوص شناخت میزان تاثیر گزاری این تبادل کنندگان بر سودهای غیر متعارف و بازده مازاد2 فرآیند کار خود را به شکل زیر پی‌گرفته اند.

در ابتدای امر با استفاده از الگوریتم‌های خوشه‌بندی سهام‌های بازار که در تاریخ ذکر شده جمع‌آوری شده است را دسته‌بندی کرده‌اند و برای این منظور از الگوریتم خوشه‌بندی سلسله مراتبی بهره گرفته‌اند.

پس از آن با استفاده از الگوریتم PCA۳ داده‌‌های خرید و فروش جمع آوری شده از این بازار را کاهش ابعاد کرده‌اند و سپس با استفاده از روش‌های آماری و آنالیز علیت4 سعی در شناخت و درک بهتر میزان تاثیر گزاری تبادل کنندگان اختلال گر بر روی بازده غیر متعارف و بازده مازاد دارند.

براساس گفته‌های این مقاله قیمت سهام بازتاب کننده چهار دیدگاه اطلاعاتی می‌باشد که این چهار دیدگاه شامل، بازار، صنعت، ویژگی‌های شرکت‌های متفاوت و احساس سرمایه‌گذاران است. بر اساس این اطلاعات، بازده غیرمتعارف می‌ةواند تحت بازتابی از احساسات یک سرماه گذار دیده شود و با توجه به مدل 4CAMP ارتباط قطعی بین عملکرد بازار و بازده هر یک از سهام‌های بازار وجود دارد که اگر نتایج این مدل را با مدل APT ۵ادغام کنیم، این موضوع می‌تواند دریافت شود که اطلاعات بازار و اطلاعات صنعت می‌توانند بر روی قیمت هریک از سهام‌های بازار به طور هم‌زمان تاثیر بگذارند.

در ابتدای امر برای برسی صحت و درستی برچسب‌های داده‌های خود با استفاده از الگوریتم خوشه‌بندی سلسله‌مراتبی سعی در دسته‌بندی داده‌ها کرده‌اند و نتیجه‌ای از پیش انتظار آن را داشتند بدست آورده‌اند. پس از آن با استفاده از الگوریتم PCA سعی در کاهش‌ابعاد داشتند که در واقع می‌توان گفت با بهره‌گیری از این الگوریتم اطلاعات بنیادی شرکت‌ها را بدست آورده‌اند و در نهایت نیز با استفاده از رگرسیون پانل‌6 میان تبادل‌گران اختلال‌گر و سوده‌بازده نامتعارف را مورد تحلیل و برسی قرارداده‌اند. پس از اعمال تست Granger causality

درخصوص تحلیل این ارتباط دریافتند که ضریب همبستگی مابین تبادل‌گران اختلال‌گر و سود‌بازده نامتعارف برابر با 0.8254 است که این ظریب همبستگی نشانه از تاثیر قابل ملاحظه تبادل‌گران اختلال‌گر بر روی بازده نامتعارف می‌باشد.

**۳-۳ مروری بر مجموعه داده‌های استفاده شده در مقالات**

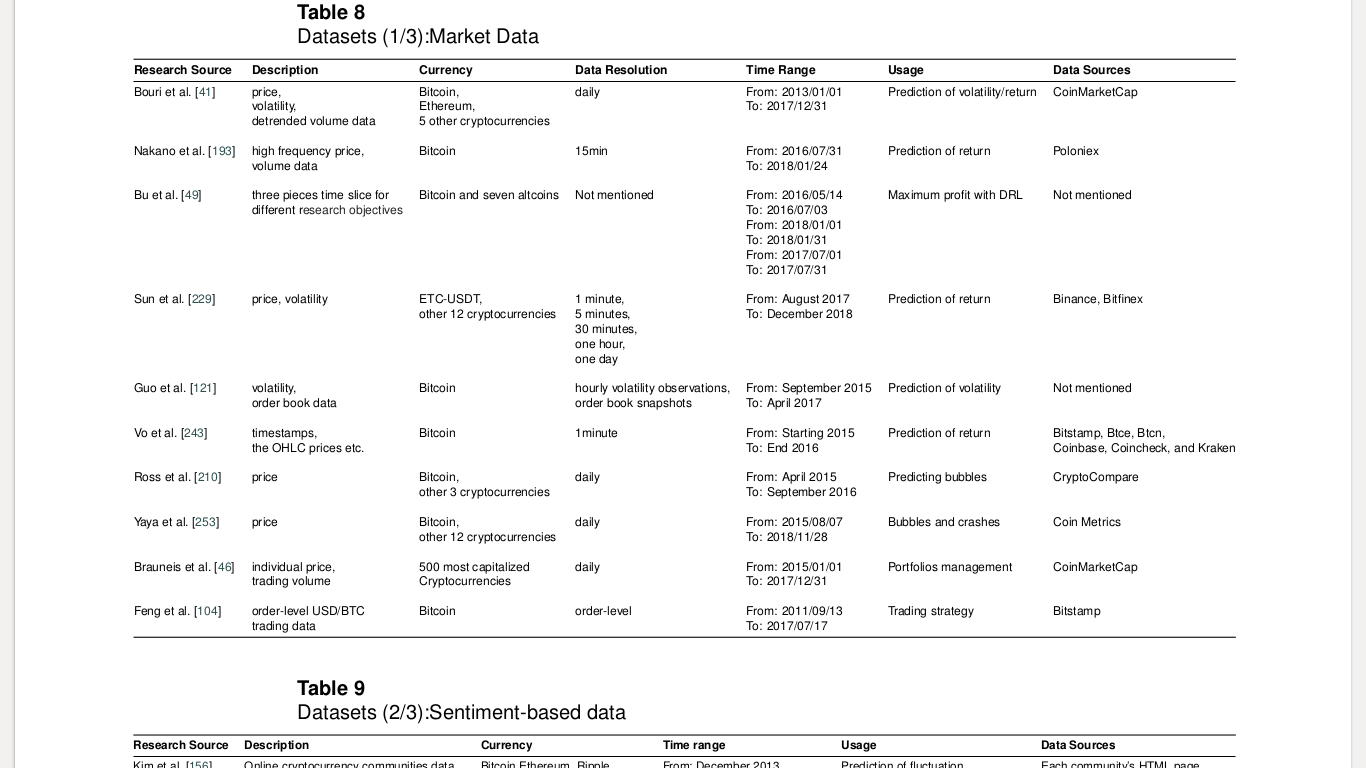
در این بخش ما مجموعه داده‌هایی که در مقالات تحلیل رفتار بازارهای رمزارزهای دیجیتال استفاده شده‌اند را مورد تحلیل و برسی قرار می‌دهیم، دلیل آنکه در این بخش مجموعه داده‌های استفاده شده در پژوهش‌های بازارهای سنتی مورد برسی قرار نمی‌گیرد این است که در ابتدا حوزه اصلی پژوهش ما در خصوص رفتارهای بازارهای رمزارزهای دیجیتال می‌باشد و داده‌های استفاده شده در این حوزه به طور کامل متفاوت از بازارهای سنتی هستند.

در [23] که مرور کلی بر روی مقالات حوزه رمزارزهای دیجیتال از سال ۲۰۱۳ تا ۲۰۱۹ می‌باشد لیستی از تمامی مجموعه داده‌های استفاده شده در مقالات آورده شده است را ارائه کرده است، بر اساس این تحقیق مجموعه داده‌های استفاده شده در بازارهای رمزارزهای دیجیتال به سه دسته متفاوت تقسیم شده‌اند، که شامل :

۱- داده‌های صرافی‌های تبادل رمزارزها.

۲- داده‌های شبکه‌های اجتماعی که به داده‌های احساسی4 معروف می‌باشند.

۳- داده‌های شبکه‌های رمزارزهای دیجیتال

در شکل-۲ لیست مجموعه‌‌داده‌های مورد ۱ [23].

حوزه اصلی تحقیق و پژوهش ما در این رساله بر اساس داده‌های صرافی‌های تبادل رمزارزها می‌باشد، اصلی ترین و مهم‌ترین تفاوت استفاده در داده‌های مورد استفاده در این پژوهش را می‌توان در این مورد دانست که علاوه بر آنکه داده‌های جمع‌آوری شده در این پژوهش از یک صرافی برخط می‌باشد و این صرافی توانایی تبادل رمزارزها و ارزهای سنتی را در یک زمان فراهم می‌کند، همانگونه‌ که پیش از این هم اشاره شد، ساختار پیاده‌سازی این بازار نیز مبتنی بر تکنولوژی زنجیره بلوکی می‌باشد. در واقع یکی دیگر از دلایلی که در این پژوهش ما سعی در جمع‌آوری و استخراج داده از بازار تبادل مالی استلار داشته‌ایم این است که این شبکه خود بر پایه زنجیره‌بلوکی است و فرکانس تبادلات در این بازار بر حسب ثانیه می‌باشد.

**مراجع**

1. S. S. Chawathe, “Clustering Blockchain Data,” in *Clustering Methods for Big Data Analytics: Techniques, Toolboxes and Applications*, O. Nasraoui and C.-E. Ben N’Cir, Eds. Cham: Springer International Publishing, 2019, pp. 43–72.
2. Stellar Development Foundation, “Ledger, Stellar Documentation.” Stellar Development Foundation, Jun. 29, 2021, Accessed: Jul. 09, 2021. [Online]. Available: https://developers.stellar.org/docs/glossary/ledger/.
3. Federation Stellar Protocol
4. <https://developers.stellar.org/docs/start/list-of-operations/>
5. D. Ermilov, M. Panov, and Y. Yanovich, “Automatic Bitcoin Address Clustering,” in *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2017, pp. 461–466, doi: 10.1109/ICMLA.2017.0-118.
6. G. Shi, L. Ren, Z. Miao, J. Gao, Y. Che, and J. Lu, “Discovering the Trading Pattern of Financial Market Participants: Comparison of Two Co-Clustering Methods,” *IEEE Access*, vol. 6, pp. 14431–14438, Feb. 2018, doi: 10.1109/ACCESS.2018.2801263.
7. A. Ntakaris, M. Magris, J. Kanniainen, M. Gabbouj, and A. Iosifidis, “Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods,” *Journal of Forecasting*, vol. 37, no. 8, pp. 852–866, 2018, doi: 10.1002/for.2543.
8. M. K. Awan and A. Cortesi, “Blockchain Transaction Analysis Using Dominant Sets,” in *Computer Information Systems and Industrial Management*, Cham, 2017, pp. 229–239, doi: 10.1007/978-3-319-59105-6\_20.
9. J. Y. Song, W. Chang, and J. W. Song, “Cluster analysis on the structure of the cryptocurrency market via Bitcoin–Ethereum filtering,” *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121339, Aug. 2019, doi: 10.1016/j.physa.2019.121339.
10. Y. Patil and J. Manish, “Cluster Driven Candlestick Method for Stock Market Prediction,” in *2020 International Conference on System, Computation, Automation and Networking (ICSCAN)*, Jul. 2020, pp. 1–5, doi: 10.1109/ICSCAN49426.2020.9262356.
11. A.-S. Yacine and B. Celso, “High frequency traders and the price process,” Dec. 2019, doi: <https://doi.org/10.1016/j.jeconom.2019.11.005>.
12. C. Lan Chang and F. Ming, “Noise Trading and Abnormal Return in Stock Market,” *Revista Argentina de Clínica Psicológica*, vol. 29, no. 4, 2020, doi: [10.24205/03276716.2020.831](https://doi.org/10.24205/03276716.2020.831).
13. P. Grimberg, T. Lauinger, and D. McCoy, “Empirical Analysis of Indirect Internal Conversions in Cryptocurrency Exchanges,” Feb. 27, 2020. Accessed: Aug. 25, 2021. [Online].
14. H. Masanori, H. MATSUSHIMA, K. IZUMI, and H. SAKAJI, “STBM: Stochastic Trading Behavior Model for Financial Markets Based on Long Short-Term Memory,” presented at the The Japanese Society for Artificial Intelligence, Jun. 2020. doi: <https://doi.org/10.11517/pjsai.JSAI2020.0_1K4ES204>.
15. I. Madan, S. Saluja, and Z. Aojia, “Automated Bitcoin Trading via Machine Learning Algorithms.” 2015. Accessed: Aug. 28, 2021. [Online]. Available: http://cs229. stanford. edu/proj2014/Isaac% 20Madan 20
16. M. Jianliang, S. Haikun, and B. Ling, “The Application on Intrusion Detection Based on K-means Cluster Algorithm,” in *2009 International Forum on Information Technology and Applications*, May 2009, vol. 1, pp. 150–152. doi: [10.1109/IFITA.2009.34](https://doi.org/10.1109/IFITA.2009.34).
17. L. Rabiner and S. Levinson, “Isolated and Connected Word Recognition - Theory and Selected Applications,” *IEEE Transactions on Communications*, vol. 29, no. 5, pp. 621–659, May 1981, doi: [10.1109/TCOM.1981.1095031](https://doi.org/10.1109/TCOM.1981.1095031).
18. D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, Jul. 1994, vol. 10, pp. 359–370.
19. D. J. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, Jul. 1994, pp. 359–370.
20. M. Nakano, A. Takahashi, and S. Takahashi, “Bitcoin technical trading with artificial neural network,” *Physica A: Statistical Mechanics and its Applications*, vol. 510, pp. 587–609, Nov. 2018, doi: [10.1016/j.physa.2018.07.017](https://doi.org/10.1016/j.physa.2018.07.017).
21. J. Sun, Y. Zhou, and J. Lin, “Using machine learning for cryptocurrency trading,” in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, Taipei, Taiwan, May 2019, pp. 647–652. doi: [10.1109/ICPHYS.2019.8780358](https://doi.org/10.1109/ICPHYS.2019.8780358).
22. Ś. Robert and Z. Maryna, “Robustness of Support Vector Machines in Algorithmic Trading on Cryptocurrency Market,” *Central European Economic Journal*, vol. 5, no. 52, pp. 186–205, 2018, doi: [10.1515/ceej-2018-0022](https://doi.org/10.1515/ceej-2018-0022).
23. F. Fang *et al.*, “Cryptocurrency Trading: A Comprehensive Survey,” *arXiv:2003.11352 [q-fin]*, Jan. 2021, Accessed: Aug. 23, 2021. [Online]. Available: <http://arxiv.org/abs/2003.11352>