Archive of SID

دسترسی در سایتِ http://jnrm.srbiau.ac.ir
سال ششم، شماره بیست و چهارم، خرداد و تیر ۱۳۹۹
شماره شاپا: ۲۵۸۸-۵۸۸X

JNRM
پژوهش‌های نوین در ریاضی

دانشگاه آزاد اسلامی، واحد علوم و تحقیقات

طراحی یک مدل هوشمند جهت تعیین سیگنال‌های معاملات سهام با رویکرد داده‌کاوی

پانتهآ ملکی مقدم[۱*]، اکبر عالم تبریز[۲]، سید اسماعیل نجفی[۳]

[۱] دانشجوی دکتری، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد تهران مرکز، تهران، ایران
[۲] استاد، گروه مدیریت صنعتی، دانشکده مدیریت، دانشگاه شهید بهشتی، تهران، ایران
[۳] استادیار، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

تاریخ ارسال مقاله: ۱۳۹۷/۱۱/۱۰ تاریخ پذیرش مقاله:۱۳۹۸/۷/۱۵

چکیده

یکی از مهمترین مسائل در بازارهای مالی مدرن یافتن راه‌های کارآمد برای تلخیص و تجسم کردن اطلاعات بازار بورس می‌باشد. با حجم انبوه از داده‌هایی که در بازار بورس تهران در هر لحظه ایجاد می‌گردد برای بررسی روابط میان داده‌ها و دست یافتن به اطلاعات نهفته آنها که تاثیر قابل ملاحظه‌ای در تصمیمات سرمایه‌گذاران دارد به مدل‌هایی دست یافتیم. با استفاده از کلان داده‌های ارزشمند تولید شده توسط بازار سهام با استفاده از روش خوشه‌بندی افرازی و به کمک الگوریتم k-means به تعیین نقاط سیگنال معاملات سهام پرداخته شده است. در این پژوهش از داده‌های صنایع خودرو و فرآورده‌های نفتی طی سال ۱۳۸۷ تا ۱۳۹۶ که با کمک بیست شاخص تکنیکی مدل‌سازی انجام پذیرفت. نتایج این پژوهش نشان داد که مدل مورد استفاده در شناسایی و پیش بینی سیگنال‌های فروش صادره در نقاط حداکثری دارای عملکرد قابل توجهی بوده و با دقت قابل قبولی قابل پیش‌بینی می‌باشند. در واقع این سیگنال‌ها دارای خطای کمتری بوده و بهتر پیش‌بینی گردیده است.

واژه‌های کلیدی: تحلیل بازار سهام، پیش‌بینی مالی، داده‌کاوی، تحلیل خوشه‌ای، الگوریتم K-MEANS.

* عهده‌دار مکاتبات: Email: pan.maleki.eng@iauctb.ac.ir

www.SID.ir

Archive of SID

۱- مقدمه

روش‌های مختلفی برای تحلیل بازار سهام در جهت کمک به تصمیم‌گیری سرمایه‌گذاران ارایه شده است که سرمایه‌گذاران و تحلیگران مالی بتوانند بیشترین بازده را در کنار کمترین ریسک ممکن داشته باشند. دو نوع تحلیل در این حوزه وجود دارد: تحلیل بنیادی و تحلیل تکنیکال. در هر دو روش سعی بر پیش‌بینی جهت قیمت‌ها دارند. در واقع آنها از دو راه متفاوت به یک مسئله واحد می‌پردازند. تحلیلگر بنیادی با تمرکز بر روی داده‌ها و صورت‌های مالی شرکت‌ها که در طی زمان منتشر شده با استفاده از روش‌های ارزش‌گذاری مختلف همچون گوردن، تنزیل جریانات نقدی و قیمت‌گذاری نسبی و ... به پیش بینی آینده شرکت و برآورد ارزش ذاتی سهام جهت تصمیم‌گیری برای خرید یا فروش آن می‌پردازد. اما تحلیلگر تکنیکال تمرکز بر روی تغییرات قیمت دارد و فقط به روند قیمتی گذشته دارایی موردنظر توجه می‌شود این مهم به این دلیل می‌باشد که پایه مباحث فلسفی و منطقی تکنیکال به سه مبحث برمی‌گردد اینکه همه چیز در قیمت لحاظ شده، قیمت‌ها براساس روند حرکت می‌کنند و تاریخ تکرار می‌شود.

در این میان با توجه به تفاوت‌های این دو نوع تحلیل به تعریف ابتدایی بازار کارا برمیخوریم که بر این فرض استوار است که تمام سرمایه‌گذاران به کلیه اطلاعات گذشته، موجود و آینده دسترسی دارند. طبق این فرضیه تغییر قیمت سهام ناشی از اطلاعات موجود بوده و این بازار موجب تعیین عادلانه قیمت و تخصیص بهینه سرمایه می‌شود. دربازار کارای کامل، قیمت‌های اوراق بهادار متاثراز اطلاعات موجود در بازار می‌باشد و سرمایه گذاران نمی‌توانند بخاطر داشتن اطلاعات، سود غیر عادی بدست آورند بنابراین در این بازار قیمت اوراق بهادار معادل ارزش ذاتی آن است. اگر برخی از اطلاعات به طور کامل در قیمت سهام منعکس نشود، درآن بازار از کارایی کامل برخوردار نخواهد بود. در واقع بازار از لحاظ کارایی دارای سه شکل می‌باشد. شکل قوی بازار کارا که در این بازار قیمت سهام کاملاً تحت تاثیر اطلاعات، اعم از اطلاعات عمومی و غیر عمومی می‌باشد. در این بازار بازده غیرعادی بر اساس اطلاعات گذشته، حال و آینده امکان‌پذیر نبوده و سرمایه‌گذاران نمی‌توانند بیش از نرخ بازده عادی در یک برهه از زمان منتفع شوند. در بازار با کارایی نیمه قوی، اطلاعات شناخته شده و در دسترس به سرعت بر قیمت‌ها اثر می‌گذارند. اما در بازار با کارایی ضعیف، اطلاعات موجود مربوط به دوره‌های قبل بوده و افراد براساس اطلاعات گذشته قادر به کسب بازده غیرعادی نمی‌باشند. [2] ,[1]

عمده مطالعاتی که در ایران در رابطه با کارایی انجام شده، حاکی از این است که بورس اوراق بهادار تهران از کارایی ضعیف برخوردار می‌باشد، این امر سبب شده است تا قیمت سهام تفاوت زیادی با ارزش ذاتی آن داشته باشد .سرمایه گذاری در چنین بازاری مستلزم تجزیه و تحلیل اوراق بهادار و همچنین زمانبندی خرید و فروش آنها می‌باشد .به طورکلی، روش‌ها و فنون مختلفی در رابطه با سرمایه گذاری در بازار سرمایه وجود دارد که آگاهی از روشهایی که برای فضای سرمایه‌گذاری امروزی مؤثر باشد، می‌تواند به سرمایه‌گذاران کمک نماید تا بیشترین بازدهی را حاصل نمایند.

با توجه به درجات مختلف کارایی که در بازار وجود دارد، محققان با در نظر گرفتن فرضیه‌های بازار کارا و به کمک داده‌های موجود به کشف الگوهای پنهان و پیش بینی روندهای آتی و رفتارها در بازار سهام می‌پردازند.[3] بنابراین در این مقاله نیز با داده‌کاوی به شناسایی سیگنال‌های خرید و فروش سهام به کمک نماگرهای تکنیکال مورد بررسی قرار گرفته است.

۲- مبانی نظری و مروری بر پیشینه تحقیق

کشف دانش و داده‌کاوی، مقوله‌ای میان رشته‌ای و رو به رشد است که با علوم مختلفی مانند پایگاه داده‌ها، آمار، ماشین و مقوله‌هایی در این زمره در ارتباط است. موضوع اصلی این مقوله داده، اطلاعات و دانش است. کشف دانش و داده کاوی را به اختصار KDD[1] می‌نامند. کشف دانش از پایگاه داده‌ها، فرآیند تشخیص الگوها[2] و مدل‌های موجود در داده‌ها است. این الگو و مدل‌ها، ویژگی‌های مانند معتبر بودن، بدیع[3] بودن، بالقوه مفید، سادگی، افزونگی و قابل فهم بودن را داراست.[4]

داده‌کاوی مرحله‌ای از کشف دانش است که با کمک الگوریتم‌های مختص داده‌کاوی با در نظر گرفتن کارایی

1. Knowledge discovery and data mining
2. patterns
3. novel

www.SID.ir

Archive of SID

قابل قبول، الگوها یا مدل‌ها را در داده‌ها پیدا می‌کند. به عبارتی داده‌کاوی یک گام اساسی در فرآیند کشف دانش است که الگوریتم استخراج و اثبات فرضیه را اعمال می‌کند. [5]

داده‌کاوی توسط یک فرد با یک هدف خاص و همراه با یک مجموعه‌ی داده (data set) صورت می‌گیرد. در واقع این فرد با بکارگیری الگوریتم‌های یادگیری ماشینی قصد دارد الگوی موردنیاز خود را از یک مجموعه داده دریافت کند. در نتیجه با استفاده از یادگیری ماشینی به انبوهی از متغیرهای اقتصادی و مالی در راستای پیش بینی آتی در این حوزه‌ها می‌توان دست یافت. در مقالات دیگر به کمک داده‌کاوی با استفاده از داده‌های بازار سهام به ارتباطات شبکه‌ای دست یافتند که توسط آن ساختار نقشه بازار بر اساس قیمت بازگشایی سهام در یک دوره مشخص است، حاصل می‌گردد. در مقاله‌ای که توسط دیویس و همکاران در سال ۲۰۱۲ ارائه شده است به طور گسترده به بررسی قوانین داده‌های عظیم می پردازند. [6]، [7]

در زمینه مدیریت پرتفولیو نیز مقالات بسیاری نیز از داده کاوی کمک گرفته‌اند که با طبقه‌بندی و خوشه‌بندی شرکت‌های مختلف در بازار سهام به روشی جدید در این زمینه دست یافتند. همچنین بهینه سازی پرتفو با استفاده از تحلیل خوشه ای و پیش بینی بازار سهام با روش ترکیبی رگرسیون چندگانه، شبکه عصبی و خوشه‌بندی دو مرحله‌ای فازی می‌توان نام برد.[9]، [8]

هم چنین از الگوریتم ژنتیک جهت بهینه‌سازی نماگرهای تکنیکال در بازار بورس نیویورک بر روی سهام‌های شاخص اس اند پی ۵۰۰ در سالهای ۱۹۲۸ تا ۱۹۹۵ استفاده شده است که بازدهی بدست امده از این روش از بازدهی استراتژی خرید و نگهداری بیشتر بوده است. [10]

در پژوهشی دیگر با استفاده از تلفیق منطق فازی با نماگر میانگین متحرک ۵۰ روزه روش تکنیکال، سیگنال‌های خروجی برای جفت ارزهای دلار به یورو از این روش بررسی شده‌اند. در این روش استفاده از منطق فازی منجر به کاهش سیگنال‌های خرید و فروش نادرست می‌شود و در نهایت بازده بیشتر و کاهش هزینه معاملاتی در پی دارد. [11] تحقیق حاضر به دنبال ارائه یک الگوریتم هوشمند و کارا برای تحلیل داده‌های بازار بورس و ارزیابی مدل ارئه شده با قابلیت استفاده در صنایع مختلفی بورسی می‌باشد. در این راستا پیاده‌سازی الگوریتم بر روی کلان داده‌های بازار بورس اوراق بهادار تهران جهت پیش‌بینی و تحلیل بازارها در دو گروه منتخب ارائه می‌گردد.

۳- روش‌شناسی و مراحل پژوهش

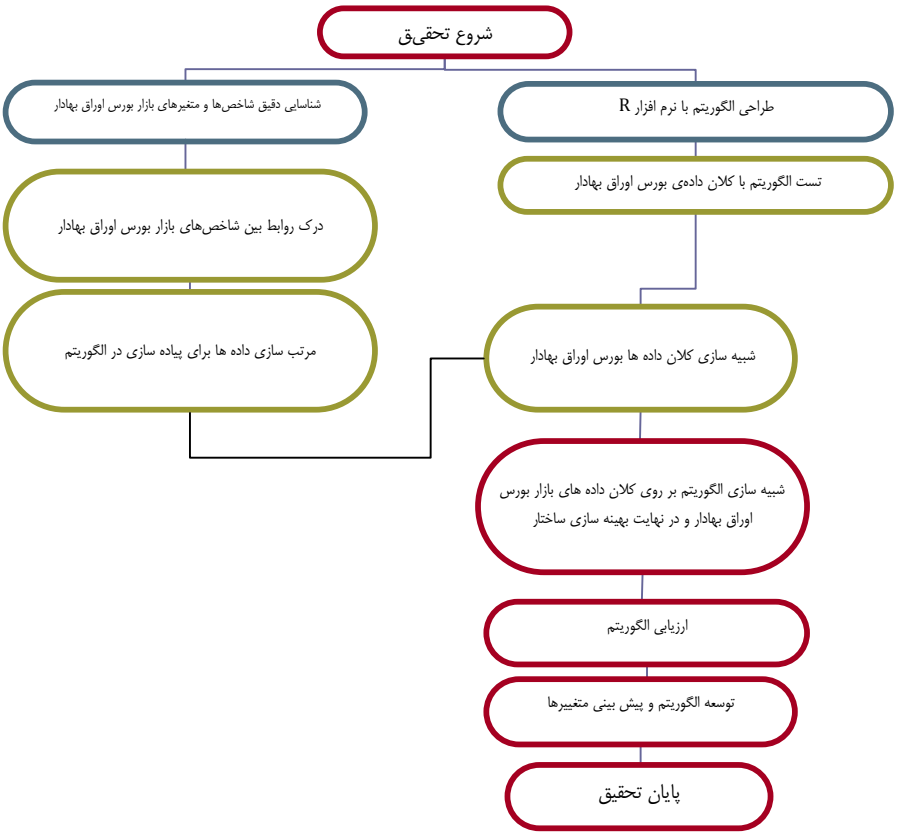
این پژوهش، از نظر هدف، کاربردی و از نظر نوع روش توصیفی است. از لحاظ جهت نیز پس‌رویدادی است. جامعه آماری این تحقیق شرکت‌های پذیرفته شده در بورس اوراق بهادار تهران می باشد و از انجایی که داده کاوی بر روی نماگرهای[1] تحلیل تکنیکال صورت می‌گیرد شرکت‌های مورد بررسی به لحاظ داده تاریخی از مقدار کافی برخوردار می‌باشند. مدل‌سازی به کمک ۲۰ نماگر تکنیکال انجام گرفته است و از ۴۰ شرکت در دو گروه طی یک دوره ده ساله شرکتهای فعال در بازار بورس تهران از سال ۱۳۸۷ تا سال ۱۳۹۶ جهت مدلسازی استفاده شده است.

برای بدست آوردن نماگرها نیازمند استفاده از بانک اطلاعات بورس اوراق بهادار تهران با استفاده همزمان از سایت شرکت مدیریت فناوری بورس تهران، نرم افزار TSE Client، نرم‌افزارهای ره‌آورد نوین و تدبیرپرداز جهت استخراج اطلاعات قیمتی روزانه شرکتها استفاده گردیده است. سپس با استفاده از آمی بروکر نرم افزار مرتبط به تحلیل‌های تکنیکال بازار سرمایه با کمک برنامه‌نویسی و کدیابی مقادیر عددی هر یک از نماگرهای موردنیاز استخراج شده‌اند. در نهایت طراحی مدل هوشمند با برنامه نویسی به کمک نرم افزار R به عنوان بهترین نرم افزار شناخته شده در زمینه داده‌کاوی مورد استفاده قرار گرفته است. مراحل تحقیق حاضر در شکل ۱ نمایش داده شده است.

1. Indicator

www.SID.ir

Archive of SID



شکل ۱ – مراحل تحقیق

٤- ادبیات تحقیق

۴-۱- روشهای افرازی

فرض کنید یک پایگاه داده با n شیء داریم. یک روش افرازی، K افراز از این داده های اشیاء درست می کند به طوریکه هر افراز یک خوشه را نشان می‌دهد و $K<n$. پس داده های اشیاء در k گروه خوشه بندی شده و دارای شرط زیر می‌باشند:

– هر گروه حداقل یک شیء دارد.

– هر شیء تنها به یک گروه تعلق دارد.

در روش افرازی برای k معلوم، یک افراز ابتدایی ایجاد می‌شود. سپس یک روش جابجایی تکراری را به کار برده که تلاش به بهبود افراز بندی دارد. به این صورت که اشیاء را از یک گروه به دیگر گروه ها می برد. یک معیار عمومی برای یک افرازبندی خوب این است که اشیاء در یک خوشه به هم نزدیک یا به یکدیگر وابسته باشند و در مقابل اشیاء در خوشه‌های مختلف، از یکدیگر دور یا تا حد امکان متفاوت باشند. برای دستیابی به خوشه بندی بهینه در روش افرازی، به شمارش کامل همه افرازهای ممکن نیاز خواهد بود یعنی تمام حالات ممکن باید بررسی شوند که این روش برای پایگاه داده های بزرگ ناممکن است. [12]

لذا الگوریتمهای هیوریستیک زیر برای بررسی این گونه موارد استفاده می شوند: الگوریتم k-means که هر خوشه با میانگین اشیاء آن خوشه یا مرکز خوشه، نمایش داده می‌شود. الگوریتم k-medoids که هر خوشه با یکی از اشیاء که در نزدیکی مرکز خوشه جای گرفته است، نمایش داده می‌شود.

۴-۲- الگوریتم k-means

الگوریتم k-means یکی از معروف‌ترین روش‌های بخش‌بندی[1] محسوب می‌شود. این الگوریتم مبتنی بر

1. Partitioning

www.SID.ir

Archive of SID

مرکز[1] نیز نامیده می‌شود. در این الگوریتم هر خوشه به صورت میانگین ارزش یا مقدار اعضای آن خوشه نمایش داده می‌شود. روش کار الگوریتم k-means به شرح زیر است.

دریافت پارامتر ورودی k

بخش‌بندی مجموعه n عضوی به k خوشه به صورتی که در داخل هر خوشه شباهت اعضا زیاد و تفاوت با اعضای خوشه‌های دیگر زیاد باشد.

همانطور که در تعریف الگوریتم k-means اشاره شد، معیار شباهت در خوشه بر اساس میانگین ارزش یا مقدار آن خوشه است. هم‌چنین در مرحله اول انتخاب تعداد خوشه‌ها به صورت تصادفی انجام می‌شود.

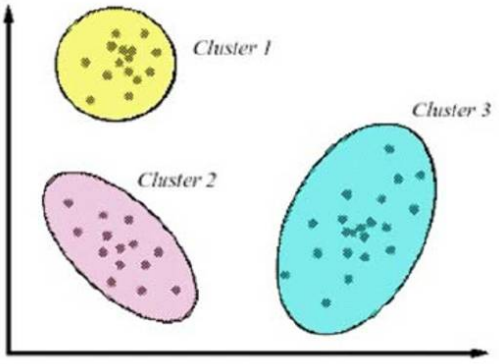
تکرار مرحله قبل تا قرارگیری اعضای مشابه در یک خوشه به طور معمول معیار عضویت در خوشه‌ها براساس معیار مربع خطا به کار گرفته خواهد شد. نحوه محاسبه معیار عضویت براساس معیار مربع به صورت زیر است:

$$E = \sum_{i=1}^{K} \sum_{p \in C_i} |p - m_i|^2$$

به عبارتی E مجموع مربع خطای تمام اعضای مجموعه، P نقطه ای از فضا به نمایندگی اعضا، m_i میانگین خوشه‌ی c_i. مثالی از خوشه بندی در الگوریتم k-means در شکل ۲ نمایش داده شده است.[13]

۵-۱- معرفی نماگرها و متغییرهای تحقیق

نماگرهای تکنیکی مورد استفاده در این تحقیق در جدول ۱ نمایش داده شده است و در ادامه فرمول‌های هر یک از نماگرها در جدول ۲ شرح داده شده است.



شکل ۲- مثالی از خوشه بندی در الگوریتم k-means

جدول ۱ : متغییرهای مورد استفاده در تحقیق

ردیف	نام متغیر	کلمه اختصاری در مدل
۱	قیمت آغازین	O
۲	بالاترین قیمت	H
۳	پایینترین قیمت	L
۴	قیمت پایانی	C
۵	حجم معاملات	Volume
۶	شاخص قدرت نسبی	RSI
۷	شاخص جریان پول	MFI
۸	شاخص کانال کالا	CCI

1. Centroid-based Technique

www.SID.ir

Archive of SID

۹	میانگین متحرک	MA
۱۰	میانگین متحرک نمایی	EMA
۱۱	میانگین متحرک موزون	WMA
۱۲	میانگین متحرک هال	HMA
۱۳	واگرایی و همگرایی متحرک	MACD
۱۴	متوسط شاخص جهت حرکت	ADX
۱۵	دامنه متوسط واقعی	ATR
۱۶	شاخص حرکت نسبی	RMI
۱۷	شاخص گام تصادفی	RWI
۱۸	استوکاستیک (D%)	D%
۱۹	استوکاستیک (K%)	K%
۲۰	تغییرات میانگین متحرک نمایی	TRIX

جدول ۲ : فرمول‌های متغیرهای تکنیکال

نماگرهای تکنیکال	فرمول	مقدار متغیر
RSI	$RS(n) = 100 - \frac{100}{(1+RS(n))}$ $RS(n) = \sum_{i=0}^{n-1} UP_{t-i} \;/\; Down_{t-i}$ Where UP_t ($Down_t$) is upward (downward) price change	n=14
MFI	$MFI = 100 - \frac{100}{(100+MR)}$ $TP = \frac{H+L+C}{3},\ MF = TP \times V$ $MR = \frac{Positive\ MF}{Negative\ MF}$	
CCI	$CCI(n) = \frac{M-\overline{M}(n)}{d(n)\times 0.015}$ $M = \frac{H+L+C}{3}$ $d(n) = \frac{1}{n}\sum_{i=0}^{n-1}\left\lvert M_{t-1} - \overline{M}_t\ (n)\right\rvert$	n=9
MA	$MA = (\mathrm{Cn} + \mathrm{C_{n-1}} + \cdots . \mathrm{C})/\mathrm{n}$	n=14
EMA	$\mathrm{EMA} = \sum_{i=0}^{n-1} \mathrm{wi} \times \mathrm{V_{t-1}}$	
WMA	$MA = \mathrm{nCm} + (\mathrm{n}-1)\mathrm{Cm} - 1 + \cdots . + 2\mathrm{C}(\mathrm{m}-\mathrm{n}+2) + \mathrm{C}(\mathrm{m}-\mathrm{n}+1)/\mathrm{n} + (\mathrm{n}-1) + \cdots + 2 + 1$	n=14
HMA	$HMA[2 \times \mathrm{WMA(n/2)} - \mathrm{WMA(n)}], \mathrm{sqrt(n)}$	n=14
MACD	$MACD = \mathrm{EMA(C, m)} - \mathrm{EMA(C, n)}$	m=5, n=26
ADX	$ADX = \mathrm{MA}[((+\mathrm{DI}) - (-\mathrm{DI}))/((+\mathrm{DI}) + (-\mathrm{DI}))] \times 100$	
ATR	$ATR = \frac{1}{n}\sum_{i=1}^{n} TR_i$	n=14
RMI	$RMI(n) = \frac{\sum_{i=0}^{n-1} UP_{t-i}}{\sum_{i=0}^{n-1} UP_{t-i} + \sum_{i=0}^{n-1} Down_{t-i}}$	n=14
RWI	$RWI = (\mathrm{H1} - \mathrm{n} - \mathrm{L})/\mathrm{ATRn} * \sqrt{\mathrm{n}})$	n=14
%D	$\%D(\mathrm{n}) = \sum_{i=0}^{n-1} \%\, K_{t-1}/n$	n=14
%K	$\%K(n) = \frac{C - L_{n,min}}{H_{n,max} - L_{n,min}} \times 100$	n=14
TRIX	$TRIX(n) = \frac{EMA^3(C,n) - EMA^3_{t-1}(C,n)}{EMA^3_{t-1}(C,n)}$	n=12

www.SID.ir

Archive of SID

٤-٤- ارزیابی اولیه مدل

شاخص‌های مختلفی برای طبقه‌بندی مورد استفاده قرار می‌گیرند و در شرایط مختلف، رفتارهای متفاوتی از خود نشان می‌دهند. این شاخص‌ها عبارتند از:

صحت مدل[1]: صحت یک روش دسته‌بندی، بستگی به تعداد پیش‌بینی‌های درستی است که آن مدل انجام داده است.

سرعت: زمان لازم جهت ساخت و استفاده از مدل در دسته‌بندی است.

پایداری[2]: این شاخص توانایی برخورد مدل در مواجهه با داده‌های غیر معمول و یا مقادیر مفقوده را نشان می‌دهد.

تفسیر پذیری[3]: این شاخص نشانگر قابل فهم بودن مدل است. به عبارتی شفافیت نحوه دسته بندی و نوع دسته ها قابل درک باشند.

جمع و جور بودن مدل[4]: این شاخص به اندازه درخت و یا تعداد قواعد ایجاد شده توسط مدل مرتبط است. [14] میزان صحت یک روش طبقه‌بندی بر روی مجموعه داده‌های آموزشی، درصد مشاهده‌هایی از مجموعه آموزشی است که به درستی توسط روش مورد استفاده، طبقه‌بندی شده‌اند. ماتریس اغتشاش[5](پریشانی)، ابزار مفیدی برای ارزیابی اولیه تحلیل چگونگی عملکرد روش طبقه بندی در تشخیص داده ها یا مشاهده‌ی دسته‌های مختلف است. برای اینکه یک روش طبقه بندی، صحت بالایی داشته باشد. حالت ایده آل آنست که بیشتر داده‌های مرتبط به مشاهده‌ها بر روی قطر اصلی ماتریس قرار گرفته باشند و بقیه مقادیر مارتیس صفر یا نزدیک به صفر باشند. ماتریس ممکن است سطر یا ستون اضافی داشته باشد که نشان‌دهنده مجموع عناصر یا درصد شناخت می باشد. در جدول ۳ شمای کلی ماتریس اغتشاش نمایش داده شده است.

مفاهیم استفاده شده در ماتریس عبارت است از: مثبت واقعی(TP[6]) و منفی واقعی(TN[7]) که دو دستبند صحیح می‌باشند. حالت مثبت اشتباه (FP[8]) زمانی رخ می‌دهد نمونه‌ی به طور واقعی منفی است، مثبت تشخیص داده شود و منفی اشتباه (FN[9]) برای نمونه هایی که به طور واقعی مثبت هستند ولی منفی تشخیص داده می‌شوند، رخ می‌دهد. نحوه ارزیابی با استفاده از فرمول‌های زیر انجام شده است:

محاسبه صحت مدل:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

درصورتی که فراوانی کلاس ها به صورت متناسب نباشند، محاسبه صحت کل کفایت نمی‌کند. بنابراین محاسبه پارامترهای نرخ منفی واقعی (TNR)[10]، دقت[11]، صحت حاشیه‌ای[12] و اندازه‌ی F[13] نیز لازم است انجام شود.[16],[15]

محاسبه نرخ منفی واقعی:

$$True\ Negative\ Rate = TN/(TN + FP)$$

محاسبه دقت:

$$Precision = TP/(TP + FP)$$

محاسبه صحت حاشیه‌ای:

$$Recall = TP/(TP + FN)$$

محاسبه اندازه‌ی F:

$$F - Measure = (\frac{Precision * Recall}{Precision + Recall}$$

۶-۴- ارزیابی و مقایسه نهایی مدل

جهت ارزیابی نهایی مدل از معیارهای سه گانه در جدول ۴ استفاده شده است.

1. Accuracy
2. Robustness
3. Interpretability
4. Compactness
5. Confusion Matrix

6. True positive
7. True negative
8. False positive
9. False negative
10. True Negative Rate
11. Precision
12. Recall
13. F-measure

www.SID.ir

Archive of SID

جدول ۳: ماتریس اغتشاش

کلاس پیش بینی شده			
		C1	C2
کلاس واقعی	C1	مثبت واقعی	منفی اشتباه
	C2	مثبت اشتباه	منفی واقعی

جدول ۴ : فرمول معیارهای مورد استفاده جهت ارزیابی مدل

معیار	فرمول
Accuracy	$Accuracy = (TP + TN)/(TP + TN + FP + FN)$
Classification error	$classificatio_error = 1 - accuracy(ACC)$
Corelation	$corr(x,y) = \frac{cov(x,y)}{\sigma_x \sigma_y}$

۷- مدل‌سازی و تحلیل یافته‌های پژوهش

در این قسمت نتایج حاصل از مدل‌سازی با الگوریتم‌های تحقیق جهت پیش‌بینی نقاط سیگنال به تفکیک شرکتها در دو صنعت در جدول زیر به تفصیل آورده شده است.

الگوریتم‌ها برای دو گروه خودرو و ساخت قطعات و گروه فرآورده‌های نفتی، کک و سوخت هسته‌ای مورد بررسی قرار گرفته است.

جدول ۵: نتایج و یافته‌های تحقیق

گروه بندی	نام شرکت	نام نماد	خوشه‌بندی ۱		خوشه‌بندی ۲	
افزایش: صفر/کاهش: یک			صفر	یک	صفر	یک
خودرو و ساخت قطعات	ایران خودرو	خودرو	٪۸۲،۳۹	٪۱۷،۶۱	٪۱۹,۴۴	٪۸۰,۵۶
	سایپا	خساپا	٪۷۹،۴۴	٪۲۰،۵۶	٪۲۳,۸۲	٪۷۶,۱۸
	پارس خودرو	خپارس	٪۶۷،۱۲	٪۳۲،۸۸	٪۳۴,۴۵	٪۶۵,۵۵
	زامیاد	خزامیا	٪۵۳،۰۷	٪۴۶،۹۳	٪۴۸,۳۵	٪۵۱,۶۵
	ایران خوردو دیزل	خاور	٪۵۴،۷۲	٪۴۵،۲۸	٪۴۷،۷۸	٪۵۲،۲۲
	گسترش سرمایه‌گذاری ایران خودرو	خگستر	٪۷۹،۵۹	٪۲۰،۴۱	٪۲۲،۷۸	٪۷۷،۲۲
	کمک فنرایندامین	خکمک	٪۵۹،۱۱	٪۴۰،۸۹	٪۴۲،۶۷	٪۵۷،۳۳
	سرمایه‌گذاری رنا	ورنا	٪۵۷،۰۶	٪۴۲،۹۴	٪۴۴،۷۸	٪۵۵،۲۲
	گروه بهمن	خبهمن	٪۵۸،۲۶	٪۴۱،۷۴	٪۴۳،۵۳	٪۵۶،۴۷
	سایپا آذین	خاذین	٪۵۱،۲۲	٪۴۸،۷۸	٪۵۰،۱۲	٪۴۹،۸۸
	نیرو محرکه	خمحرکه	٪۶۶،۷۲	٪۳۲،۲۸	٪۳۴،۱۳	٪۶۵،۸۷
	قطعات اتومبیل ایران	ختوقا	٪۵۷،۸۴	٪۴۲،۱۶	٪۴۳،۵۳	٪۵۶،۴۷
	رینگ سازی مشهد	خرینگ	٪۷۰،۱۱	٪۲۹،۸۹	٪۳۱،۸۹	٪۶۸،۱۱
	ایرکا پارت صنعت	خکار	٪۶۰،۸۵	٪۳۹،۱۵	٪۳۹،۵۰	٪۶۰،۵
	آهنگری تراکتورسازی ایران	خاهن	٪۷۸،۵۶	٪۲۱،۴۴	٪۲۷،۸۵	٪۷۲،۱۵

www.SID.ir

Archive of SID

	چرخشگر	خچرخش	٪۵۳،۷۲	٪۴۶،۲۸	٪۵،۵۶	٪۹۴،۴۴
	فنرسازی زر	خزر	٪۵۵،۲۶	٪۴۴،۷۴	٪۴۷،۳۵	٪۵۲،۶۵
	محورسازان ایران خودرو	خوساز	٪۵۵،۸۷	٪۴۴،۱۳	٪۴۵،۰۲	٪۵۴،۹۸
	صنایع ریخته‌گری ایران	خریخت	٪۵۸،۲۲	٪۴۱،۷۸	٪۴۳،۵۹	٪۵۶،۴۱
	موتورسازان تراکتورسازی ایران	خموتور	٪۶۳،۳۴	٪۳۶،۶۶	٪۳۷،۵۰	٪۶۲،۵
	مهندسی صنعتی روان فن‌آور	خفناور	٪۵۱،۶۵	٪۴۸،۳۵	٪۵۱،۷۵	٪۴۸،۵۲
	رادیاتور ایران	ختور	٪۷۰،۲۹	٪۲۹،۷۱	٪۳۱،۵۲	٪۶۸،۱۸
	لنت ترمز ایران	خلنت	٪۶۳،۴۵	٪۳۶،۵۵	٪۳۷،۴۵	٪۶۲،۵۵
	تولید محور خودرو	خمحور	٪۶۳،۶۴	٪۳۶،۳۶	٪۳۹،۱۳	٪۶۰،۸۷
	ریخته‌گری تراکتورسازی ایران	ختراک	٪۵۶،۱۳	٪۴۳،۸۷	٪۴۴،۱۱	٪۵۵،۸۹
	الکترونیک خودرو شرق	خشرق	٪۶۹،۴۵	٪۳۰،۵۵	٪۳۴،۲۰	٪۶۵،۸
	مهندسی نصیر ماشین	خنصیر	٪۵۳،۳۵	٪۴۶،۶۵	٪۴۷،۱۶	٪۵۲،۸۴
	بهمن دیزل	خدیزل	٪۶۲،۴۸	٪۳۷،۵۴	٪۴۰،۲۰	٪۵۹،۸
فرآورده های نفتی، کک و سوخت هسته‌ای	پالایش نفت اصفهان	شپنا	٪۸۰،۳۵	٪۱۹،۶۵	٪۲۱،۷۷	٪۷۸،۲۲
	پالایش نفت تهران	شتران	٪۶۷،۴۸	٪۳۲،۵۲	٪۳۴،۰۱	٪۶۵،۹۹
	پالایش نفت تبریز	شبریز	٪۷۵،۴۴	٪۲۴،۵۶	٪۲۵،۸۸	٪۷۴،۱۲
	پالایش نفت بندرعباس	شبندر	٪۸۰،۲	٪۱۹،۸	٪۲۰،۷۸	٪۷۹،۲۲
	سرمایه‌گذاری صنعت نفت	ونفت	٪۵۶،۳۹	٪۴۳،۶۱	٪۴۵،۱۹	٪۵۴،۸۱
	نفت سپاهان	شسپا	٪۵۴،۸۱	٪۴۵،۱۹	٪۴۶،۸۶	٪۵۳،۱۴
	پالایش نفت لاوان	شاوان	٪۷۷،۶۵	٪۲۲،۳۵	٪۲۳،۴۶	٪۷۶،۵۴
	نفت ایرانول	شرانل	٪۶۷،۹۸	٪۳۲،۰۳	٪۳۵،۸۷	٪۶۴،۱۳
	پالایش نفت شیراز	شراز	٪۶۰،۱۹	٪۳۹،۸۱	٪۴۰،۸۲	٪۵۹،۱۸
	نفت بهران	شبهرن	٪۵۳،۹۸	٪۴۶،۰۲	٪۴۸،۵۵	٪۵۱،۴۵
	نفت پارس	شنفت	٪۶۸،۳۵	٪۳۱،۶۵	٪۳۲،۴۴	٪۶۷،۵۶
	نفت پاسارگاد	شپاس	٪۵۲،۹۸	٪۴۷،۰۲	٪۴۸،۴۵	٪۵۱،۵۵

همانطور که در جدول شماره ۵ مشاهده می‌شود، خوشه بندی در دو گروه انجام گرفته است که گروه اول از نقاط کف نمودارهای هر شرکت استخراج شده و سپس در مدل جهت خوشه بندی استفاده شده است و در گروه دوم نیز نتایج حاصل از نقاط سقف نمودارها مورد تجزیه وتحلیل قرا گرفته است. در این تحقیق فقط نقاط سقف و کف مورد ارزیابی قرار گرفته است و سایر نقاط که نه در سقف و نه در کف باشند در مدل استفاده نشده است. چهارده شرکت در گروه خودرو و ساخت قطعات و هفت شرکت درگروه فرآورده‌های نفتی، کک و سوخت هسته‌ای در پیش بینی نقاط سقف از دقت بالاتری برخوردار بوده اند و سیگنال فروش صادر کرده‌اند. این بدین معنی است که در مدل پیشنهادی در گروه خودرو در ۵۰ درصد و گروه فرآورده های نفتی در ۵۸ درصد

www.SID.ir

Archive of SID

موارد پیش‌بینی نقاط سیگنال فروش یا نقاط در سقف بهتر انجام داده است. در واقع در گروه فرآورده‌های نفتی تنها در پنج مورد نقاط کف بهتر از سایر نقاط پیش بینی شده است.

جدول ۶: محاسبه دقت نتایج مدل

گروه‌بندی	نام شرکت	correlation	Classification error(%)	Accuracy(%)
خودرو و ساخت قطعات	ایران خودرو	۰،۴۲	۱۶،۵	۸۳،۵
	سایپا	۰،۵۱	۲۱،۳۵	۷۸،۶۵
	پارس خودرو	۰،۶۷	۲۴،۷۷	۷۵،۲۳
	زامیاد	۰،۵۶	۲۳،۲	۷۶،۸
	ایران خوردو دیزل	۰،۴۶	۲۸،۱	۷۱،۹
	گسترش سرمایه‌گذاری ایران خودرو	۰،۶۴	۲۰،۴۵	۷۹،۵۵
	کمک فنرایندامین	۰،۲۱	۳۱،۵	۶۸،۵
	سرمایه‌گذاری رنا	۰،۴۰	۳۴،۳۴	۶۵،۶۶
	گروه بهمن	۰،۵۳	۳۲،۵	۶۷،۵
	سایپا آذین	۰،۵۴	۴۹،۵	۵۸،۵
	نیرو محرکه	۰،۷۳	۳۸،۴	۶۱،۶
	قطعات اتومبیل ایران	۰،۷	۴۱،۳۱	۵۸،۶۹
	رینگ سازی مشهد	۰،۳۶	۳۸،۲	۶۱،۸
	ایرکا پارت صنعت	۰،۴۱	۳۷،۶	۶۲،۴
	آهنگری تراکتورسازی ایران	۰،۷۱	۴۲،۷	۶۷،۳
	چرخشگر	۰،۴۵	۴۵،۴	۶۴،۶
	فنرسازی زر	۰،۶۳	۴۷،۳	۷۲،۷
	محورسازان ایران خودرو	۰،۴۲	۴۸،۲	۶۱،۸
	صنایع ریخته گری ایران	۰،۴۴	۳۸،۲	۷۱،۸
	موتورسازان تراکتورسازی ایران	۰،۲۲	۳۷،۸	۷۲،۲

www.SID.ir

Archive of SID

	مهندسی صنعتی روان فن‌آور	۰،۴۰	۴۰،۷	۵۹،۳
	رادیاتور ایران	۰،۴۲	۴۳،۴	۵۶،۶
	لنت ترمز ایران	۰،۲۱	۴۱،۸	۵۸،۲
	تولید محور خودرو	۰،۳۶	۳۶،۷	۷۳،۳
	ریخته گری تراکتورسازی ایران	۰،۲۸	۳۵،۵	۷۴،۵
	الکترونیک خودرو شرق	۰،۴۸	۳۴،۸	۶۵،۲
	مهندسی نصیر ماشین	۰،۴۴	۴۴،۸	۷۵،۲
	بهمن دیزل	۰،۶۷	۳۸،۲	۷۱،۸
فرآورده های نفتی، کک و سوخت هسته‌ای	پالایش نفت اصفهان	۰،۷۵	۱۴،۸	۸۵،۲
	پالایش نفت تهران	۰،۵۴	۲۱،۱	۷۸،۹
	پالایش نفت تبریز	۰،۴۸	۱۷،۴	۸۲،۶
	پالایش نفت بندرعباس	۰،۶۹	۱۱،۱	۸۸،۹
	سرمایه‌گذاری صنعت نفت	۰،۵۱	۲۱،۴	۷۸،۶
	نفت سپاهان	۰،۴۸	۲۴،۴	۷۵،۶
	پالایش نفت لاوان	۰،۷۱	۲۱،۷	۷۸،۳
	نفت ایرانول	۰،۶۵	۲۷،۲	۷۲،۸
	پالایش نفت شیراز	۰،۴۵	۴۱،۴	۷۸،۶
	نفت بهران	۰،۴۸	۴۴،۴	۶۵،۶
	نفت پارس	۰،۳۸	۴۵،۳	۶۴،۷
	نفت پاسارگاد	۰،۴۱	۴۹،۴	۷۰،۶

نتایج حاصل از ارزیابی اولیه مدل پیشنهادی در جدول ۶ نشان می‌دهد که دقت مدل در تمامی شرکت‌ها در هر دو گروه خودرو و ساخت قطعات و گروه فرآورده‌های نفتی، کک و سوخت هسته‌ای به‌طور میانگین بیش از ۷۰ درصد می‌باشد حداکثر دقت مدل در گروه فرآورده‌های نفتی شامل اوراق بهادر شرکت پالایش نفت بندر عباس می‌شود که با دقتی در حدود ۸۸،۹ درصد می‌باشد. در این تحقیق سیگنال‌های فروش صادره در نقاط حداکثری

www.SID.ir

Archive of SID

دارای عملکرد قابل توجهی بوده است. در واقع این سیگنال‌ها دارای خطای کمتری بوده و بهتر پیش‌بینی گردیده است. نتایج تحقیق نشان می‌دهد که دقت مدل در کل شرکت‌های مورد بررسی به جز ۵ شرکت از ۴۰ شرکت تماما بالای ۷۰ درصد می‌باشد. درواقع خطای مدل کمتر از ۳۵ درصد می‌باشد و مدل پیشنهادی دارای دقت حداقل ۷۰ درصدی در دو گروه مورد بررسی می‌باشد. نتایج اولیه جهت تشخیص عملکرد مدل جهت پیش‌بینی نقاط سیگنال به تفکیک هر شرکت در دو گروه خودرو و فرآورده های نفتی درج گردیده است.

با توجه به اینکه از شاخص‌های یکسانی در هر دو گروه خودرو و ساخت قطعات و گروه فرآورده‌های نفتی، کک و سوخت هسته‌ای استفاده شده است اما دقت در هر یک از شرکتهای مورد بررسی دارای تفاوت معناداری است. به نوعی می‌توان نتیجه گرفت که در نقاطی که سیگنال خرید می‌توان صادر کرد تا حدودی شاهد نوعی رفتار غیر قابل پیش‌بینی و منطبق بر ساختار آشوب‌گونه یا نوعی بی‌نظمی می‌باشیم به طوری که دقت کمتری در مقایسه با نقاط حداکثری می‌باشد.

www.SID.ir

Archive of SID

فهرست منابع

[1] Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. International Journal of forecasting, 20(1), 15-27.

[2] Fama, E. (1991). Efficient capital markets. Journal of Finance, XLVI, 1575–1617.

[3] Skabar, A., & Cloete, I. (2002). Neural networks, financial trading and the efficient markets hypothesis. Australian Computer Science Communications, 24(1), 241-249.

[4] Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. Expert Systems with Applications, 29, 927–940.

[5] Kamber, J. H. A. M. (2006). Data Mining Concepts and Techniques.
[6] K. Davis, D. Patterson, Ethics of Big Data: Balancing Risk and Innovation, O'Reilly Media, 2012.

[7] Kumar, M., & Thenmozhi, M. (2006). Forecasting stock index movement: A comparison of support vector machines and random forest.

[8] Hsu, C. M. (2011). A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. Expert Systems with Applications, 38(11), 14026-14036.

[9] Vrahatis, M. N., Boutsinas, B., Alevizos, P., & Pavlides, G. (2002). The new k-windows algorithm for improving the k-means clustering algorithm. Journal of Complexity, 18(1), 375–391.

[10] Allen, F., Karjalainen, R., (1999), using genetic algorithms to find technical trading rules, Journal of Financial Economics, 51, 245-271.

[11] Kuo, R. J., Chen, C. H., & Hwang, Y. C. (2001). An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy sets and systems*, *118*(1), 21-4

[12] Kuo, R. J., Liao, J. L., & Tu, C. (2005). Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce. Decision Support Systems, 40(2), 355–374.

[13] Padmanabhan, B., & Tuzhilin, A. (2002). Knowledge refinement based on the discovery of unexpected patterns in data mining. Decision Support Systems, 33, 309–321.

[14] S.R. Nanda, B. Mahanty, M.K. Tiwari.(2010).Clustering Indian stock market data for portfolio management.Expert Systems with Applications ,37 ,8793–8798.

[15] Haugen, R. (1997). Modern investment theory. Upper Saddle River, NJ: Prentice-Hall.

[16] Ballings, M., Van den Poel, D., Hespeels, N., & Gryp, R. (2015). Evaluating multiple classifiers for stock price direction prediction. Expert Systems with Applications, 42(20), 7046-705.

www.SID.ir

Archive of SID

www.SID.ir