چشم‌انداز مدیریت مالی
شماره ۲۱ ـ بهار ۱۳۹۷
ص‌ص ۲۷ـ۹

پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه سهام با استفاده از مدل‌های نزدیک‌ترین همسایگی و ماشین بردار پشتیبان

زهرا اسماعیلی*، ابراهیم عباسی، میرفیض فلاح *****

چکیده

اولین انتشار عمومی سهام توسط یک شرکت، عرضه عمومی اولیه نامیده می‌شود. در سال‌های اخیر تحقیقات زیادی عملکرد کوتاه‌مدت IPO را موردبرسی قرار داده است. هدف این پژوهش، آزمون مدل‌های طبقه‌بندی مختلف برای یافتن مدلی که از کارایی و دقت بالایی در پیش‌بینی عملکرد کوتاه‌مدت IPO برخوردار است. این پژوهش ۶۰ مورد IPO عرضه‌شده در بورس اوراق بهادار تهران را طی دوره ۱۳۹۴-۱۳۸۴ موردبررسی قرار می‌دهد. در چارچوب ارائه‌شده، میانگین بازده مازاد سه روز اول کاری سهم عرضه‌شده دارای مقداری مثبت و برابر با ۱/۳ درصد است، هرچند این مقدار به اندازه بازارهای توسعه‌یافته بالا نیست. روش اعتبارسنجی متقابل۱۰-تایی برای ارزیابی و کنترل عملکرد مدل‌های طبقه‌بندی نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، درخت تصمیم‌گیری کارت و بیز ساده مورداستفاده قرار گرفت؛ در نتیجه در بین مدل‌ها، مدل نزدیک‌ترین همسایگی و ماشین بردار پشتیبان از دقت و کارایی بالایی در پیش‌بینی عملکرد کوتاه‌مدت IPO برخوردار است.

کلیدواژه‌ها: اعتبارسنجی متقابل۱۰ تایی؛ عرضه عمومی اولیه؛ عملکرد کوتاه‌مدت IPO؛ مدل‌های طبقه‌بندی.

تاریخ دریافت مقاله: ۱۳۹۵/۰۸/۱۸، تاریخ پذیرش مقاله: ۱۳۹۶/۱۰/۰۶

* کارشناس ارشد، دانشگاه الزهراء (نویسنده مسئول).

E-mail: Zahra.esmaeili۶۹@gmail.com

** دانشیار، دانشگاه الزهراء.

*** دانشیار، دانشگاه آزاد اسلامی.

مقدمه

اولین فروش سهام به عموم، عرضه عمومی اولیه (IPO[۱]) نامیده می‌شود. قیمت‌گذاری اوراق شرکت‌هایی که برای اولین بار عرضه می‌شود مشکل‌ترین و چالش‌برانگیزترین گام عرضه اوراق در بازار اولیه است؛ چراکه قیمت‌گذاری واقعی اوراق عرضه‌شده با درنظرگرفتن منافع سرمایه‌گذاران و شرکت منتشرکننده اوراق، نیازمند دخالت‌دادن متغیرهای زیادی است و بدین منظور باید فرایند دقیق و کاملی اجرا شود [۹]. قیمت‌گذاری کمتر از ارزش واقعی باعث جذابیت سهم برای خریداران می‌شود. پژوهش‌های انجام‌شده در کشورهای مختلف حاکی از آن است که سهم تازه عرضه‌شده در کوتاه‌مدت عملکرد بهتری نسبت به شاخص بورس در همان دوره زمانی دارد که این موضوع توسط محققان ایرانی نیز موردتایید قرار گرفته است.

بحث بازده‌های کوتاه‌مدت برای عرضه عمومی اولیه سال‌ها است در ادبیات تحقیقاتی و تجربی موردتوجه است. تحقیقات نشان می‌دهد ارزان قیمت‌گذاری باعث بازده‌های فوق‌العاده در کوتاه‌مدت برای مدت سه روز معاملاتی بعد از عرضه اتفاق می‌افتد و اما عملکرد بلندمدت آنها بازده‌های غیرعادی ندارد [۲].

ازآنجاکه در بیشتر پژوهش‌ها در زمینه بازده‌های کوتاه‌مدت عرضه‌های اولیه سهام، تنها به بررسی رویکرد آمار کلاسیک از جمله رگرسیون پرداخته‌اند؛ بنابراین این پرسش که آیا می‌توان مدل طبقه‌بندی[۲] آموزش داد که بتوان عملکرد کوتاه‌مدت سهام تازه عرضه‌شده را با دقت بالایی پیش‌بینی کرد و اینکه کدام مدل دقت بیش‌تری در پیش‌بینی دارد، یکی از موضوع‌های حائز اهمیت در این زمینه است [۵].

از سال‌ها قبل، تکنیک‌های بسیاری برای پیش‌بینی عملکرد کوتاه‌مدت سهام عرضه عمومی اولیه توسعه یافته‌اند. در ابتدا روش رگرسیون کلاسیک برای پیش‌بینی روند سهام استفاده می‌شد؛ اما ازآنجاکه داده‌های سهام می‌تواند به‌عنوان داده‌های سری زمانی غیرثابت طبقه‌بندی شوند، از روش‌های یادگیری ماشین غیرخطی نیز استفاده شده است که رایج‌ترین آنها شبکه عصبی است؛ با این حال شبکه عصبی کاستی‌هایی دارد مانند نیاز به پارامترهای کنترلی زیاد، دشواری رسیدن به نتیجه پایدار و غیره. الگوریتم نزدیک‌ترین همسایگی و ماشین بردار پشتیبانی الگوریتم‌های یادگیری ماشین هستند که به‌طور گسترده برای پیش‌بینی سهام استفاده می‌شود. هر الگوریتم برای یادگیری، الگوهای منحصربه‌فردی دارد [۱۰].

با توجه به روش‌های تحلیل، از رایج‌ترین روش‌های یادگیری ماشینی استفاده می‌شود. انتخاب روش‌های ما براساس نتایج به‌دست‌آمده از مطالعات تفضیلی قبلی [۸] و نیز براساس آزمایش‌های محققان این پژوهش بوده است. بعد از اعتبارسنجی این مطالعات، پی برده شد که

---

۱. Initia Public Offer
۲. Classification Model

درختان تصمیم‌گیری و ماشین‌های بردار پشتیبان و بیز ساده و الگوریتم نزدیک‌ترین همسایگی عملکرد بهتری از روش‌های آماری، شبکه‌های عصبی و رگرسیون لجستیک داشتند. چون این روش‌ها مزیت‌های بهتری از حیث سهولت درک، شفاف‌بودن، ظاهر جذاب و استفاده آسان دارند، در این پژوهش از محبوب‌ترین روش درخت تصمیم‌گیری CART در تحلیل تفضیلی استفاده شده است.

این مطالعه بر روی مقایسه عملکرد پیش‌بینی الگوریتم نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، درخت تصمیم‌گیری C&R و الگوریتم بیز ساده برای کار پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه است. پانزده پارامتر سابقه ۲۱ روزه از شاخص کل بازار، نرخ تورم، سن شرکت، مجموع دارایی‌ها، حقوق صاحبان سهام، فروش سالانه، سود عملیاتی، سود خالص، جریان‌های نقدی عملیاتی، بازده دارایی (ROA)، بازده حقوق صاحبان سهام (ROI)، گردش دارایی، نسبت بدهی، نسبت قیمت به سود (P/E) و قیمت انتشار سهام عرضه اولیه به‌عنوان متغیر ورودی به این مدل‌ها استفاده می‌شود.

هدف پژوهش پیش‌رو نیز دادن بهترین مدل برای پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه است تا از این طریق بتوان به تصمیم‌گیری درست سرمایه‌گذاران کمک کرد.

۲. مبانی نظری و پیشینه پژوهش

اصطلاح عرضه عمومی اولیه که رواج آن به بازارهای ۱۹۹۰ باز می‌گردد، عبارت است از این که یک شرکت برای اولین بار در طول دوران فعالیت خود اقدام به عرضه سهام به عموم می‌نماید [۳]. عرضه عمومی اولیه سهام توسط شرکت‌هایی صورت می‌گیرد که درمورد این شرکت‌ها بازار شناخت کمی از مدیریت و مالکیت آن دارد. بدین ترتیب بین ناشران در زمان عرضه عمومی و مالکان بالقوه بیرونی، عدم تقارن اطلاعاتی وجود دارد. برای سرمایه‌گذاران، قضاوت درمورد مناسب‌بودن اقلام تعهدی شرکت‌هایی که برای اولین بار سهام خود را در بازار عرضه می‌کنند، به منظور ارزیابی عملکرد آتی امری دشوار است [۱۸].

شواهد تجربی فراوانی بر عملکرد کوتاه‌مدت عرضه عمومی اولیه تاکید دارد که سرمایه‌گذارانی که سهام عرضه‌های اولیه را در روز اول عرضه خریداری کرده و پس از مدتی کوتاهی آن را به فروش می‌رسانند، به بازدهی غیرعادی قابل‌توجهی دست می‌یابند؛ در نتیجه عرضه عموم اولیه عملکرد کوتاه‌مدت مثبتی دارد. در این زمینه، تحقیقات وسیعی نیز در سرتاسر دنیا انجام شده است. برای مثال لوگران و ریتر (۲۰۰۲)، در مطالعات خود درباره عملکرد کوتاه‌مدت IPO بیان نمودند که بازده کوتاه‌مدت در آمریکا در سال ۱۹۸۰ در حدود هفت درصد و در سال ۱۹۹۰ این بازده در حدود پانزده درصد است. کمرمان و والچی (۲۰۰۵) میانگین بازده

کوتاه‌مدت در بازار بورس کشور سوئیس را در حدود ۳۴/۵ درصد بیان می‌دارد. ظریف‌فرد و مهرجو (۱۳۸۳) نشان می‌دهد که بازده کوتاه‌مدت سهام شرکت‌های تازه پذیرفته‌شده در بورس اوراق بهادار تهران بیشتر از بازده پرتفوی بازار بوده است.

متغیرهای تحقیق. در پیش‌بینی عملکرد سهام تازه عرضه‌شده، دقت پیش‌بینی تحت تاثیر تعداد ویژگی‌ها و متغیرهای ورودی مدل قرار می‌گیرد. استفاده از بردارهای ویژگی با ابعاد بزرگ هزینه محاسباتی زیاد و ریسک بیش برازشی مدل را به همراه دارد [۱۰].

بازده کوتاه‌مدت بازار. بازده کل بازار در یک دوره زمانی ۳ روزه، از اولین روز پس از عرضه سهم تازه عرضه‌شده تا سه روز کاری بعد از آن را نشان می‌دهد [۱].

بازده کوتاه‌مدت سهام جدید. بازده سهام جدید در سه روز اول کاری پس از عرضه نشان می‌دهد [۲].

بازده مازاد سهم عرضه‌شده نسبت به بازار (میزان ارزان یا گران قیمت‌گذاری). تفاضل بازده کوتامدت ۳ روزه سهام جدید و بازده کوتاه‌مدت بازار برای دورهٔ کوتاه‌مدت مشابه [۲].

تمایلات بازار. بازده ۲۱ روزه شاخص کل بازار به‌عنوان نماینده‌ای برای تمایلات بازار پیش از عرضه اولیه درنظر گرفته خواهد شد [۵].

سن شرکت. تاریخ اولین سال معاملاتی منهای تاریخ سال ثبت شرکت، سن شرکت را نشان می‌دهد [۵].

قیمت انتشار سهام عرضه اولیه. قیمت پیشنهادی خریدار است که اولین معامله بر آن صورت می‌گیرد و از سایت بورس تهران به‌دست می‌آید [۱۱].

نسبت P/E. نسبت قیمت به درآمد سهم بیانگر آن است که قیمت سهام شرکت، چند برابر میزان سود نقدی که شرکت به هر سهم خود تخصیص می‌دهد که از تقسیم قیمت انتشار سهام عرضه اولیه به آخرین سود به‌دست می‌آید [۱۱].

نرخ تورم. نشان‌دهنده افزایش یا کاهش یا ثبات در قیمت مصرف‌کننده کالاها و خدمات دانست و از سایت بانک مرکزی به‌دست می‌آید [۸].

سایر متغیرها نیز سن شرکت، نسبت بدهی، مجموع دارایی‌ها، حقوق صاحبان سهام، فروش سالانه، سود عملیاتی، سود خالص، جریان‌های نقدی عملیاتی، بازده دارایی (ROA)، بازده حقوق صاحبان سهام (ROI) و گردش دارایی است که از صورت‌های مالی آخرین سال قبل از عرضه عمومی قابل دستیابی هستند.

پیشینه تجربی پژوهش. روند مطالعات نشان می‌دهد هرچند مدل‌های آماری توانسته‌اند پیش‌بینی‌های خوبی را درمورد عملکرد سهام تازه عرضه‌شده ارائه دهند، در عین حال مفروضات محدودکننده برخی از این مدل‌ها بر اثربخشی این مدل‌ها موثر بوده است؛ بنابراین به تدریج روش‌های دیگری برای مقابله با این محدودیت‌ها و بهبود عملکرد پیش‌بینی معرفی شد. در پژوهش‌های داخلی و خارجی مطالعات زیادی انجام شده که برتری روش‌های ماشین برداری را بر روش‌های پیشین، از جمله روش شبکه عصبی و نیز روش‌های سنتی نشان می‌دهد؛ ازاین‌رو تنها به ذکر چند نمونه از مطالعاتی پرداخته می‌شود که اخیرا در داخل و خارج از کشور انجام گرفته و بیشترین ارتباط را با موضوع پژوهش دارند.

خداپرستی، زمانیان، سنگنیان (۱۳۹۲)، در تحقیقی به‌بررسی بازده کوتاه‌مدت و بلندمدت سهام عرضه‌شده در عرضه‌های اولیه در بورس اوراق بهادار پرداخته است. یافته‌های تحقیق نشان می‌دهد متوسط بازده کوتاه‌مدت شرکت‌های واگذارشده دولتی در این دوره ۵/۱۹٪ (تعدیل‌شده با بازده بازار) و برای شرکت‌های خصوصی فقط ۸ درصد بوده است؛ همچنین براساس یافته‌های این تحقیق، پیش‌بینی نظریه‌های مبتنی بر عدم تقارن اطلاعاتی را در بازار سرمایه ایران موردتایید قرار می‌دهد [۱۱].

فلاح‌پور و همکاران (۱۳۹۲)، یک مدل ترکیبی برای پیش‌بینی روند حرکتی قیمت سهام با استفاده از مدل ماشین بردار پشتیبان بر پایه الگوریتم ژنتیک ارائه کرده‌اند. نتایج نشان می‌دهد، این مدل ترکیبی در پیش‌بینی روند حرکتی قیمت سهام بسیار بهتر عمل کرده و در مقایسه با روش ماشین بردار پشتیبان ساده، از دقت بالاتری برخوردار است [۱۰].

بولو و فلاح برندق (۱۳۹۲)، در تحقیقی به‌بررسی بازده غیرعادی عرضه عمومی اولیه پرداخته است که نتایج حکایت از این دارد که سهام تازه عرضه‌شده در بازده چهار هفته بعد از تاریخ عرضه، به‌طور متوسط ۲۶/۰۷ درصد بازده اولیه عادی و ۱۹/۶۱ درصد بازده اولیه غیرعادی ایجاد کرده‌اند [۶].

عباسی، بالاورد (۱۳۸۷) در تحقیقی به آزمون پدیده ارزان قیمت‌گذاری عرضه عمومی اولیه سهام و عملکرد بلندمدت آن در بورس تهران موردبررسی قرار داده است. نتایج نشان داد که پدیده ارزان قیمت‌گذاری عرضه عمومی اولیه در بورس تهران وجود دارد [۲].

باستی و همکاران (۲۰۱۵)، در پژوهشی عرضه‌های عمومی اولیه شرکت‌های ترکی را بررسی کردند. در چارچوب ارائه‌شده، عواملی که عملکرد کوتاه‌مدت را تحت‌تاثیر قرار می‌دهند تجزیه و تحلیل شده‌اند. در این پژوهش نشان داده شد که عرضه عمومی اولیه شرکت‌ها ارزان قیمت‌گذاری شده‌اند. تمایلات بازار، میزان فروش سالانه، نرخ گردش کل دارایی‌ها، روش‌های فروش سهام، روش‌های پذیره‌نویسی IPO، قیمت پیشنهادی، نرخ بدهی، تعداد سهام فروش رفته عوامل بسیار مهمی هستند که عملکرد کوتاه‌مدت عرضه‌های عمومی اولیه شرکت‌های ترکی را تحت‌تاثیر قرار می‌دهند [۵].

سرن و ریچاردسون (۲۰۱۶)، در پژوهشی با رویکرد جدیدی از داده‌کاوی و اعمال نفوذ اطلاعات مربوط به سهام به‌منظور پیش‌بینی عملکرد سهام جدید عرضه عمومی اولیه معرفی کردند. با استفاده از توابع کاهشی یکنواختی، رابطه بین نوسانات قیمت سی روزه هر همسایه سهام تازه عرضه‌شده توسط تغییر انتشار سی روزه هر یک از همسایه‌ها به نسبت فاصله آن بررسی شد. ترکیب حساسیت مکانی روابط جالبی بین عرضه اولیه اوراق بهادار و نزدیکی همسایگی کشف کرد؛ همچنین رابطه قوی بین فاصله سهام همسایه و عملکرد سی روزه مشاهده شد [۷].

مولی و همکاران (۲۰۱۴)، در پژوهشی به‌بررسی پیش‌بینی سهام برنده عرضه عمومی اولیه پرداخته‌اند. این پژوهش روش جدیدی برای کمک به سرمایه‌گذاران عرضه عمومی اولیه اوراق بهادار برای توزیع بازده رو به بالا پیشنهاد می‌کند. این پژوهش مدل رگرسیون لجستیک را برای پیش‌بینی این‌که آیا یک شرکت ۲،۱ یا ۳ سال بعد از فهرست‌شدن، تنها با تکیه به اطلاعات در دسترس عموم هنوز عملکرد بهتری دارند. سرمایه‌گذارانی که از مدل این پژوهش استفاده می‌کنند همواره منجر به نرخ تعدیل پیش‌بینی بالاتر می‌شود، براساس یک طبقه‌بندی ساده همه عرضه اولیه اوراق بهادار به‌عنوان «برندگان» می‌باشند [۱۷].

تانگ سنگ و کیان چونگ (۲۰۱۶)، به‌پژوهشی برای پیش‌بینی عملکرد عرضه اولیه اوراق بهادار پرداخته است. اگر چه بسیاری از نتایج تحقیقات نشان می‌دهد که شبکه‌های عصبی می‌تواند تقریبا تمام مشکلات را موثرتر از روش مدل‌های آماری و سنتی حل کند؛ اما نتایج تحقیقات نشان می‌دهد که روش‌های آماری در نمونه داده‌های خاص بهتر از شبکه‌های عصبی عمل می‌کند. ازآنجاکه عرضه اولیه اوراق بهادار که اطلاعات قبل از خود، مقادیر پیش‌بینی خواهد

بود در سری زمانی مرتبط و متغیرهای از عوامل خاص شرکت است. نتایج تجربی براساس عرضه اولیه اوراق بهادار از بورس سنگاپور به ارزیابی عملکرد پیش‌بینی شده است [۱۹].

با توجه به عدم پیش‌بینی‌پذیری اندازه حرکت یا تغییر قیمت دارایی‌ها در اکثر بازارها، محققان به سراغ روش‌هایی رفتند که صرفا جهت حرکت قیمت دارایی را پیش‌بینی نماید. این روش‌ها شامل روش‌های سنتی از قبیل لاجیت، پروبیت، گشت تصادفی و روش‌های جدیدتر که رایج‌ترین آنها شبکه عصبی می‌باشد؛ با این حال شبکه عصبی کاستی‌هایی دارد مانند نیاز به پارامترهای کنترلی زیاد، دشواری رسیدن به نتیجه پایدار و غیره. الگوریتم نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، بیز ساده و درخت تصمیم‌گیری است. پژوهش‌های انجام‌شده در زمینه پیش‌بینی عملکرد سهام نشان می‌دهد که مدل‌های غیرخطی در مقایسه با مدل‌های پیش‌بینی سنتی از دقت بالاتری و عملکرد بهتری برخوردار است [۱۰].

مدل نزدیک‌ترین همسایگی[۱]. الگوریتم نزدیک‌ترین همسایگی یک روش غیرپارامتری مورداستفاده برای طبقه‌بندی و رگرسیون است. در دو مورد، شامل K نزدیک‌ترین نمونه‌های آموزشی در فضای ویژگی می‌باشد. خروجی KNN به اینکه برای طبقه‌بندی استفاده می‌شود و یا رگرسیون بستگی دارد [۱۸].

خروجی مدل طبقه‌بندی نزدیک‌ترین همسایگی عضویت در کلاس است. یک شی با رای اکثریت همسایگی خود طبقه‌بندی می‌شود، با این هدف که هر شی به رایج‌ترین کلاس در میان K نزدیک‌ترین همسایگی خود اختصاص داده می‌شود (K یک عدد صحیح مثبت، به‌طور معمول کوچک است). اگر K=۱ باشد، پس از آن شی به سادگی به کلاس نزدیک‌ترین همسایه واحد خود اختصاص داده می‌شود. در رگرسیون KNN خروجی یک مقدار ارزش (ضریب) برای شی موردنظر می‌باشد که مقدار آن برابر با میانگین ارزش‌های K نزدیک‌ترین همسایه تعیین‌شده برای آن است.

KNN یک نوع یادگیری براساس مثال و یا یادگیری تنبل است که در آن تابع فقط به‌صورت محلی تقریب و محاسبه می‌شود. الگوریتم KNN یکی از ساده‌ترین الگوریتم یادگیری ماشینی است. در هر دو مورد طبقه‌بندی و رگرسیون، می‌توان به کمک همسایگی وزن‌دهی را انجام داد؛ به‌طوری‌که همسایه نزدیک‌تر اهمیت بیشتری نسبت به همسایه دورتر دارد؛ برای مثال، یک طرح وزن معمول عبارت است از دادن وزن $1/d$ به هر یک از همسایگان که در آن d فاصله به همسایه است.

---

۱. Nearest Neighbor Model

برای به‌دست‌آوردن نزدیک‌ترین همسایگان یک نمونه، معمولا از فاصله اقلیدسی طبق رابطه زیر استفاده می‌شود [۳]:

$$d_{eucl}(x,t) = \sqrt{\sum d^{i}_{eucl}(x,t)}$$ رابطه (۱)

اگر مقادیر خصوصیات عددی و پیوسته باشد، d^{i}_{eucl} از رابطه زیر به‌دست می‌آید:

$$d^{i}_{eucl}(x,t) = (a_i(x) - a_i(t))^2$$ رابطه (۲)

و اگر خصوصیات مقادیر اسمی داشته باشند، رابطه بالا به‌صورت زیر تغییر پیدا خواهد کرد:

$$d^{i}_{eucl}(x,t) = \begin{cases} 1 & if\ a_i(x) \neq a_i(t) \\ 0 & if\ a_i(x) = a_i(t) \end{cases}$$ رابطه (۳)

مدل طبقه‌بندی ماشین بردار پشتیبان[۱]. مدل ماشین بردار پشتیبان یکی از مدل‌هایی است که جهت طبقه‌بندی و تفکیک گروه‌ها به‌کار می‌رود. با مشخص‌شدن داده‌های مورد پژوهش، مدل ماشین بردار پشتیبان داده‌ها را به گروه‌های متمایزی تقسیم می‌کند. این مدل‌ها دارای خواص کلی زیر هستند [۱۰]:

ـ طبقه‌بندی داده‌ها با حداکثر قابلیت تعمیم

ـ رسیدن به نقطه بهینه تفکیک داده‌ها

ـ تعیین خودکار ساختار بهینه برای طبقه‌بندی کننده

ـ امکان مدل‌کردن داده‌های غیرخطی با استفاده از تجزیه‌وتحلیل اجزاء اصلی

ماشین بردار پشتیبان الگوریتمی است که نوع خاصی از مدل‌های خطی را می‌یابدکه موجب حداکثرشدن تفکیک بین طبقات می‌شود. بردارهای پشتیبان در واقع نزدیک‌ترین نقاط به حاشیه ابر صفحه هستند. تنها از این بردارها (نقاط) برای مشخص‌کردن مرز بین طبقات استفاده می‌شود.

اگر داده‌ها به‌صورت خطی مجزا از هم باشند، SVM به ماشین‌های خطی برای تولید یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیک‌ترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید. اگر نقاط آموزشی را به‌صورت $[x_i, y_i]$ و بردار ورودی

۱. Support Vector Machines

$x_i \in R^n$ وارزش طبقه را به‌صورت$\{-۱,۱\} \in y_i$ تعریف کنیم و $i = ۱, \dots, l$ آنگاه در حالتی که داده‌ها به‌صورت خطی قابل تفکیک هستند قواعد تصمیم‌گیری که تعریف می‌شود و توسط یک صفحه بهینه که طبقات تصمیم‌گیری باینری را تفکیک می‌کند به‌صورت معادله زیر است:

رابطه (۴) $Y=\text{sign}(\sum_{i=۱}^{N} y_i \alpha_i (X.X) + b)$

که در آن Y خروجی معادله، y_i ارزش طبقه نمونه آموزشی X_i . نشان‌دهنده ضرب داخلی است. بردار نشان‌دهنده یک داده ورودی و بردارهای پشتیبان هستند. در رابطه ۴ پارامترهای b و α_i تعیین‌کننده ابرصفحه هستند.

اگر داده‌ها به‌صورت خطی قابل تفکیک نباشند، این معادله به شکل زیر تغییر می‌یابد:

رابطه (۵) $Y=\text{sign}(\sum_{i=۱}^{N} y_i \alpha_i K(X, X_i) + b)$

تابع $K(X, X_i)$ تابعی کرنلی است که برای ایجاد ماشین‌هایی با انواعی از سطوح تصمیم‌گیری غیرخطی در فضای داده‌ها، ضرب‌های داخلی تولید می‌کند. برای مثال، سه نوع تابع کرنل که در مدل SVM به‌کار می‌رود، عبارت‌اند از:

ـ ماشین چندجمله‌ای با تابع کرنل

رابطه (۶) $K(X, X_i) = (X.X_i + ۱)^d$

که در آن d درجه کرنل چندجمله‌ای است.

ـ ماشین تابع پایه شعاعی با تابع پایه کرنل

رابطه (۷) $K(X, X_i) = \exp\left(-\frac{۱}{\delta^۲}(X.X_i)^۲\right)$

که در آن δ پهنای باند کرنل تابع پایه شعاعی است.

ـ ماشین NN دو لایه با تابع کرنل

$$K(X, X_i) = S[(X.X_i)] = {}^{1}/_{[1 + \exp\{v(X.X_i) - c\}]}$$ رابطه (۸)

که در آن c و v پارامترهای تابع زیگموئیدی $S[(X.X_i)]$ هستند؛ به‌نحوی‌که نامعادله $c \geq v$ برقرار باشد.

فرآیند یادگیری برای ایجاد توابع تصمیم‌گیری دارای ساختاری دو لایه است. ماشین بردار پشتیبان از تئوری بهینه‌سازی برای طبقه‌بندی استفاده می‌کند و براساس تئوری یادگیری آماری، خطای طبقه‌بندی را به حداقل می‌رساند.

روش اعتبارسنجی متقابل[۱]. اعتبارسنجی متقابل یک روش محبوب برای برآورد صحت غیراریب عملکرد مدل پیش‌بینی است. این روش گاهی موسوم به برآورد چرخش بوده و هدف از این روش ارزیابی این است که چگونه نتایج روش تحلیل پیش‌بینی را می‌توان به مجموعه داده‌های مستقل تعمیم داد. اعتبارسنجی متقابل ۱۰ تایی، به فراوانی در تحقیقات داده‌کاوی استفاده می‌شوند؛ زیرا مطالعات تجربی نشان داده‌اند که ۱۰ تعداد تاهای بهینه می‌باشند که یک تعادل ظریف و خوب را بین اریبی نمونه‌برداری (یعنی تنوع‌بخشی به زیر نمونه‌های آموزشی و آزمایشی) و زمان‌بندی (زمان کامل‌شدن مدل‌سازی و فعالیت‌های آزمایشی) ایجاد می‌کند. ضرورتا، در اعتبارسنجی متقابل ۱۰ تایی، مجموعه داده‌ها به‌طور تصادفی به ۱۰ زیرمجموعه انحصاری متقابل با اندازه تقریبا برابر تقسیم می‌شود. مدل‌ها در ابتدا آموزش دیده و توسعه می‌یابند و سپس آزمایش می‌شوند و فرایند ۱۰ بار تکرار می‌شود. هر بار، مدل بر روی ۹ بار (به‌صورت داده‌های آموزشی ترکیبی که شامل ۹۰ درصد مجموعه داده کل است) آموزش دیده و بر روی ۱ بار (۱۰ درصد مجموعه داده کل) دیگر آزمایش می‌شود. برآورد اعتبارسنجی متقابل از صحت کل مدل با میانگین‌گیری از ۱۰ شاخص صحت متعلق به هر بار که در معادله زیر ذکر شده، محاسبه می‌شود.

$$CVA = \frac{1}{k}\sum_{i=1}^{k} A_i$$ رابطه (۹)

۱. K-fold Cross Volidation

که CVA بیانگر صحت اعتبارسنجی متقابل، k تعداد بارها (k =۱۰ در اینجا درنظرگرفته شد) و A شاخص صحت است [۱۳].

اندازه‌گیری دقت و کارایی مدل‌های پیش‌بینی. یک ماتریس بهم‌ریختگی که موسوم به ماتریس دقت است، برای تعیین عملکرد مدل‌های مورداستفاده در پیش‌بینی برآیندهای دودویی استفاده می‌شود. این ماتریس دارای اطلاعات ارزشمندی درخصوص طبقه‌بندی‌های واقعی و پیش‌بینی‌شده توسط مدل پیش‌بینی است [۱۳].

به گفته کارونا و همکاران (۲۰۰۸)، ارزیابی عملکرد روش‌های یادگیری بسیار مهم است. ازاین‌رو در این پژوهش از شاخص‌های عملکرد شناخته‌شده نظیر صحت کلی و شاخص F- استفاده شدند.

TN: بیانگر تعداد رکوردهایی است که دسته واقعی آنها منفی بوده و الگوریتم طبقه‌بندی نیز دسته آنها را به درستی منفی تشخیص داده است (منفی واقعی).

TP: بیانگر تعداد رکوردهایی است که دسته واقعی آنها مثبت بوده و الگوریتم طبقه‌بندی نیز دسته آنها را به‌درستی مثبت تشخیص داده است (مثبت واقعی).

FP: بیانگر تعداد رکوردهایی است که دسته واقعی آنها منفی بوده و الگوریتم طبقه‌بندی نیز دسته آنها را به اشتباه مثبت تشخیص داده است (مثبت کاذب).

FN: بیانگر تعداد رکوردهایی است که دسته واقعی آنها مثبت بوده و الگوریتم طبقه‌بندی دسته آنها را به اشتباه منفی تشخیص داده است (منفی کاذب).

صحت کلی AC به‌صورت درصدی از اطلاعاتی که به‌طور صحیح توسط مدل پیش‌بینی می‌شوند تعریف می‌شود. این شاخص همچنین به‌صورت نسبت موارد پیش‌بینی‌شده صحیح به تعداد کل موارد پیش‌بینی تعریف می‌شود.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$ رابطه (۱۰)

پارامترهای موردنیاز برای محاسبه شاخص F- در معادلات زیر تعریف شده است:

$$Precision_{positive} = \frac{TP}{TP+FP}$$ رابطه (۱۱)

$$Precision_{negative} = \frac{TN}{TN+FN}$$ رابطه (۱۲)

$$Recall_{positive} = \frac{TP}{TP+FN}$$ رابطه (۱۳)

$$Recall_{negative} = \frac{TN}{TN+FP}$$ رابطه (۱۴)

شاخص_F، به‌صورت میانگین همساز شاخص‌های عملکرد ریکال ودقت می‌باشند که با استفاده از معادله زیر به‌دست می‌آید:

$$F - measure = \frac{۲\times Precision\times Recall}{Precision+Recall}$$ رابطه (۱۵)

فرضیه‌های پژوهش

۱. مدل ماشین بردار پشتیبان در پیش‌بینی عملکرد کوتامدت IPO از کارایی و دقت بالایی[۱] برخوردار است.

۲. مدل نزدیک‌ترین همسایگی در پیش‌بینی عملکرد کوتامدت IPO از کارایی و دقت بالایی برخوردار است.

۳. مدل درخت تصمیم‌گیری CART در پیش‌بینی عملکرد کوتامدت IPO از کارایی و دقت بالایی برخوردار است.

۴. بیز ساده در پیش‌بینی عملکرد کوتامدت IPO از کارایی و دقت بالایی برخوردار است.

۳. روش‌شناسی پژوهش

در این پژوهش از الگوریتم‌های نزدیک‌ترین همسایگی، ماشین بردار پشتیبان، درخت تصمیم‌گیری C&R و بیز ساده به‌عنوان مدل‌های پیش‌بینی استفاده شده است. جامعه آماری این پژوهش عبارت است از کلیه شرکت‌های تازه پذیرفته‌شده در بورس اوراق بهادار تهران در فاصله فروردین ۱۳۸۴ تا اسفندماه ۱۳۹۴ که سهام آنها در این فاصله در بورس برای اولین بار مورد معامله قرار گرفته‌اند که تعداد آنها برابر با ۸۰ شرکت است. از میان این ۸۰ شرکت، نمونه‌ای گرفته شد که تعداد ۶۰ شرکت وارد نمونه پژوهش شدند. تعداد ۲۰ شرکت به این دلیل که اطلاعات موردنیاز از آنها بر نرم‌افزارهای اطلاعاتی بورس موجود نبود و یا با اشتباه همراه بود، وارد هیچ‌یک از نمونه‌ها نشد.

در پژوهش حاضر برای تجزیه‌وتحلیل داده‌ها و انجام آزمون‌های آماری از نرم‌افزارهای EXCEL و MATLAB استفاده شده است. عملکرد کوتاه‌مدت سهام تازه عرضه‌شده را با استفاده از تغییرات قیمت محاسبه‌شده و با عملکرد کوتاه‌مدت بازار در همان دوره زمانی موردمقایسه قرار خواهد گرفت که این دوره این عملکرد کوتاه‌مدت سه روز اول کاری

۱. درصورتی یک مدل از دقت و کارایی بالایی برخوردار است که دقت کل و شاخص–اف آن بالای هفتاد درصد باشد.

می‌باشد که این دوره زمانی براساس پژوهش‌های پیشین و تجربه محققان این پـژوهش انتخـاب شده است.

برای شروع بازدهی سه روزه از سهام جدید به وسیله فرمول زیر محاسبه می‌شود:

$$R_{i,t}= (\frac{P_{i.t}}{P_{i.Ipo}}) -۱$$

$R_{i,t}$: نشان‌دهنده بازدهی سه روزه سهام (i)

$P_{i.t}$: قیمت پایانی سهام (i) در سومین روز معاملاتی

$P_{i.Ipo}$: قیمت انتشار سهام عرضه اولیه (i)

براساس این، بازدهی شاخص بازار بورس تهـران طـی دوره مشـابه بـه‌صـورت فرمـول زیـر محاسبه می‌شود:

$$R_{m,t} = \left(\frac{P_{m,t}}{P_{m,t-۳}}\right) - ۱$$

$R_{m,t}$: بازدهی سه روزه شاخص بازار

$P_{m,t}$: قیمت پایانی از شاخص بازار در روز t

$P_{m,t-۳}$: قیمت پایانی از شاخص بازار در روز t-۳

پس از آن، بازدهی شاخص بازار از بازدهی سهام شرکت عرضه اولیه در جهت یافتن بـازدهی مازاد شرکت عرضه اولیه در سه روز اول معاملاتی (سه روز کاری) کم خواهد شد:

$$E(R_{i,t}) = R_{i,t} - R_{m,t}$$

در این پژوهش با استفاده از پانزده متغیر پیش‌بینی و بازده مازاد سهم عرضه‌شده نسبت به بازار به‌عنوان خروجی به مقایسه و اعتبارسنجی مدل‌های طبقه‌بندی مختلف پرداخته شد. با توجه به روش‌های مختلف طبقه‌بندی، از رایج‌ترین روش‌های یادگیری ماشینی استفاده شده است.

متغیرهای مورد استفاده در این تجزیه‌وتحلیل عبارت‌اند از:

سابقه ۲۱ روزه از شاخص کل بازار، نرخ تورم، سن شرکت، مجموع دارایی‌ها، حقوق صاحبان سهام، فروش سالانه، سود عملیاتی، سود خالص، جریان‌های نقدی عملیاتی، بازده دارایی (ROA)، بازده حقوق صاحبان سهام (ROI)، گردش دارایی، نسبت بدهی، نسبت قیمت به سود (P/E)، قیمت انتشار سهام عرضه اولیه.

٤. تحلیل یافته‌ها

ابتدا به‌بررسی آمار توصیفی متغیر وابسته پرداخته شد و جدول ۱ نتیجه این بررسی است:

جدول ۱. آمار توصیفی متغیر وابسته به‌کاررفته در آزمون فرضیه‌ها

متغیر وابسته	حداقل	حداکثر	میانگین	انحراف معیار
بازده مازاد سهام تازه عرضه‌شده	-۰/۳۸۶	۰/۱۲۱	۰/۰۱۳	۰/۰۸۴

برای این‌که درستی دقت پیش‌بینی چهار مدل مقایسه شود، داده‌های تمامی شرکت‌های تازه پذیرفته‌شده در بورس اوراق بهادار به‌عنوان ورودی، به مدل‌ها داده شد و نتایج پیش‌بینی برای تمام مدل‌های به‌کاررفته در پژوهش محاسبه شده است. برای ارزیابی همه‌جانبه توانمندی مدل‌ها، باید مقدار صحت کل AC و شاخص-F محاسبه شود که این کار برای هریک از روش‌ها به‌طور جداگانه انجام گرفت و نتایج این محاسبات در جدول‌های جداگانه آمده است. صحت کلی AC به‌صورت درصدی از اطلاعاتی که به‌طور صحیح توسط مدل پیش‌بینی می‌شوند، تعریف می‌شود. این شاخص همچنین به‌صورت نسبت موارد پیش‌بینی‌شده صحیح به تعداد کل موارد پیش‌بینی تعریف می‌شود. شاخص-F، به‌صورت میانگین همساز شاخص‌های عملکرد ریکال و دقت می‌باشند. ویژگی موسوم به نرخ منفی حقیقی است که به‌صورت نسبت تعداد منفی حقیقی به جمع منفی حقیقی و مثبت کاذب درنظر گرفته می‌شود. این نسبت‌ها خروجی ماتریس دقت است که در این قسمت به‌صورت خلاصه ارائه خواهد شد. این ماتریس برای تعیین عملکرد مدل‌های مورداستفاده در پیش‌بینی برآیندهای دودویی استفاده می‌شود که دارای اطلاعات ارزشمندی درخصوص طبقه‌بندی‌های واقعی و پیش‌بینی‌شده توسط مدل پیش‌بینی است.

جدول ۲. نتایج عددی ماتریس دقت برای مدل‌های طبقه‌بندی

مدل طبقه‌بندی	ماتریس دقت		پیش‌بینی	
			منفی	مثبت
ماشین بردار پشتیبان	واقعیت	منفی	۴۵	۲۴۰
		مثبت	۰	۶۱۵
نزدیک‌ترین همسایگی	واقعیت	منفی	۱۶۵	۱۲۰
		مثبت	۱۰۵	۵۱۰
درخت تصمیم‌گیری R&C	واقعیت	منفی	۱۳۵	۱۵۰
		مثبت	۱۹۵	۴۲۰
بیز ساده	واقعیت	منفی	۲۴۰	۴۵
		مثبت	۴۰۵	۲۱۰

جدول ۳. نتایج درصدی ماتریس دقت برای مدل‌های طبقه‌بندی

مدل طبقه‌بندی	ماتریس دقت		پیش‌بینی	
			منفی	مثبت
ماشین بردار پشتیبان	واقعیت	منفی	۰/۱۶	۰/۸۴
		مثبت	۰	۱
نزدیک‌ترین همسایگی	واقعیت	منفی	۰/۵۸	۰/۴۲
		مثبت	۰/۱۷	۰/۸۳
درخت تصمیم‌گیری R&C	واقعیت	منفی	۰/۴۷	۰/۵۳
		مثبت	۰/۳۲	۰/۶۸
بیز ساده	واقعیت	منفی	۰/۸۴	۰/۱۶
		مثبت	۰/۶۶	۰/۳۴

جدول ۳ از نتایج ماتریس دقت برای هر چهار مدل به‌دست آمده است. داده‌های قیدشده در خانه‌های جدول ۲، مجموع نتایج ماتریس دقت برای هر شصت شرکت نمونه است؛ در جدول ۳ نیز نسبت درصدی آنها ارائه شده است. همان‌طور که ملاحظه می‌شود در هر چهار مدل، مجموع تعداد رکوردهایی که مثبت یا منفی بوده و به درستی پیش‌بینی شده است، در مدل نزدیک‌ترین همسایگی بیشتر از سایر مدل‌ها است که برتری این مدل نسبت به سه مدل دیگر را در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه نشان می‌دهد. براساس این نتایج، مقایسه‌ای برای ارزیابی دقت کل و کارایی پیش‌بینی بین مدل‌های مورداستفاده در این پژوهش انجام شده است (جدول ۴).

جدول ۴. نتایج دقت پیش‌بینی برای مدل‌های طبقه‌بندی

مدل طبقه‌بندی	دقت کل	شاخص-F
SVM	۷۱/۷	۸۰
K-NN	۷۳/۳	۸۱
C&R	۶۱/۶	۷۰/۹
بیز ساده	۵۰	۴۸/۱

جدول ۴ نتایج کلی به‌دست‌آمده از شصت شرکت را برای مدل‌های ماشین بردار پشتیبان، نزدیک‌ترین همسایگی، درخت تصمیم‌گیری CART و بیز ساده نشان می‌دهد. براساس تحقیقات پیشین که در این زمینه صورت گرفته است، مدلی دارای دقت لازم و قابل‌قبول در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه برخوردار است که دارای دقت کل بالای ۷۰ درصد و شاخص-F بالای ۸۰ درصد باشد. همان‌طورکه از نتایج جدول مشخص است، مدل

نزدیک‌ترین همسایگی در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه از کارایی لازم برخوردار است؛ همچنین دقت کل این مدل از سایر مدل‌ها بیش‌تر می‌باشد که نشان از برتری این مدل در مقابل سایر مدل‌ها است. در این پژوهش، مدل ماشین بردار پشتیبان دومین مدل از نظر دقت و صحت در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه است، این در حالی است که دو مدل درخت تصمیم‌گیری CART و بیز ساده از دقت کافی برخوردار نیستند.

۵. بحث و نتیجه‌گیری

در این پژوهش بازده کوتاه‌مدت سهام جدید با استفاده از تغییرات قیمت سهام شرکت‌های تازه پذیرفته‌شده در بورس طی دوره سه روز اول معاملاتی پس از عرضه عمومی اولیه آنها در این بازار محاسبه شده است و سپس با مقایسه آن با بازده کوتاه‌مدت بازار طی دوره مشابه این پژوهش موردآزمون قرار گرفته است. بر طبق جدول ۱ میانگین بازده مازاد کوتاه‌مدت سه روزه سهام شصت شرکت تازه عرضه‌شده به بازار در طی دوره سه روزه پژوهش ۱/۳ درصد بوده است و این نشان‌دهنده عملکرد بهتر سهام تازه عرضه‌شده به بازار است.

مقایسه دقت کل پیش‌بینی برای چهار روش KNN، SVM، C&R Tree و بیز ساده نشان داد مدل نزدیک‌ترین همسایگی قوی‌تر از سایر مدل‌های موجود در این پژوهش عمل کرده است و پیش‌بینی را با دقت بالاتری انجام داده است. نتایج مقایسه مدل‌های به‌کاررفته در پژوهش نشان داد که بعد از مدل نزدیک‌ترین همسایگی، مدل ماشین بردار پشتیبان با تفاوت دو درصد بالاترین دقت را در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه دارد. نتایج حاکی از قابلیت پیش‌بینی مدل نزدیک‌ترین همسایگی و ماشین بردار پشتیبان برای عملکرد کوتاه‌مدت عرضه عمومی اولیه بود. این در حالی است که دو مدل درخت تصمیم‌گیری CART و بیز ساده از حداقل دقت و کارایی برخوردار نیستند؛ به‌عبارتی دیگر این دو مدل قابلیت پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه را ندارند.

پژوهش‌های مختلفی که در این زمینه انجام شده است، نشان می‌دهد که مدل‌های KNN و SVM مدل‌های مناسبی برای پیش‌بینی عملکرد کوتاه‌مدت IPO می‌باشد که دراین تحقیق نیز تایید شد؛ اما با توجه به پژوهش باستی و همکاران (۲۰۱۵) که توانایی مدل درخت تصمیم‌گیری C&R در پیش‌بینی عملکرد کوتاه‌مدت عرضه عمومی اولیه انجام داده است، دقت کلی مدل ۸۰/۵ درصد است که با مقدار دقت به‌دست‌آمده در این پژوهش تفاوت قابل‌توجهی دارد؛ به‌گونه‌ای‌که در آن پژوهش مدل درخت تصمیم‌گیری C&R به‌عنوان یکی از بهترین مدل‌ها برای پیش‌بینی توصیه می‌شود؛ درحالی‌که در این پژوهش این مدل حداقل دقت لازم را دارد. تفاوت دقت این مدل پیش‌بینی در پژوهش حاضر و پژوهش ذکرشده، نخست به این دلیل است

که این دو پژوهش در دو بورس جدا که یکی بورس اوراق بهادار تهران و دیگری بورس اوراق بهادار استانبول صورت گرفته است، دوم متغیرهای دو تحقیق با یکدیگر متفاوت است و در پژوهش باستی از متغیرهایی استفاده شده است که در درخت تصمیم‌گیری بیش‌تر از سایر متغیرها قابلیت طبقه‌بندی دارد، مانند قیمت پیشنهادی، درصد سهام عرضه‌شده و درصد سهام نگهداری‌شده که در بورس ایران قابل دستیابی نیست.

در زمینه پیش‌بینی عملکرد سهام تازه عرضه‌شده، در سطح جهان به‌خصوص کشورهای توسعه‌یافته مطالعات متعدد و گسترده‌ای صورت گرفته است؛ بنابراین پیشنهاد می‌شود در بورس اوراق بهادار تهران نیز در مورد پیش‌بینی عملکرد سهام جدید بررسی‌ها و تحقیقات دیگری انجام گیرد.

با توجه به این‌که روش‌های نزدیک‌ترین همسایگی و ماشین بردار پشتیبان دقت بالایی در پیش‌بینی عملکرد کوتاه‌مدت IPO دارند، توصیه می‌شود که از این الگوریتم‌ها در جهت تصمیم‌گیری بهتر استفاده شود.

پیشنهاد می‌شود برای دست‌یافتن به نتیجه بهتر، از روش‌های انتخاب ویژگی استفاده شود و تعداد ویژگی‌ها کاهش یافته و دقت مدل‌ها افزایش یابد.

همچنین در پژوهش‌های آتی می‌توان سایر ویژگی‌های پیش‌بینی تاثیرگزار بر عملکرد کوتاه‌مدت عرضه عمومی اولیه همانند: روش‌های عرضه سهام،روش‌های پذیره‌نویسی، درصد سهام عرضه‌شده، نوع صنعت و غیره بررسی شود. قدرت پیش‌بینی سایر مدل‌های طبقه‌بندی همچون شبکه عصبی، درخت تصمیم‌گیری C۵ و غیره در عملکرد کوتاه‌مدت IPO بررسی شود و در نهایت تحلیل حساسیت ویژگی‌های پیش‌بینی و درصد تاثیر آنها بر ارزان قیمت‌گذاری با استفاده از دو مدل نزدیک‌ترین همسایگی و ماشین بردار پشتیبان اندازه‌گیری شود.

منابع

۱. Abbasi, E and Teimourpour, B. (۲۰۱۷). The application of conditional value at risk in portfolio optimization based on structural Breakpoint approach In Tehran Securities Exchange. *Financial Management Perspective Financial Management Perspective, ۲.*

۲. Abbasi, E and Balaverdi, M. (۲۰۰۸). Empirical review of low pricing and long-term returns of the initial public offering of shares. (Master's thesis). Islamic Azad University, Iran.

۳. Alimi, Amir & Kordestani, GHolamreza. (۲۰۰۹). The use of Residual Income Valuation Model as a Benchmark for Portfolio Selection. *Financial Management perspective, ۱.*

٤. Arefin, F and Eftekhari, M. (۲۰۱۳). New K-nearest neighbor method of fuzzy and rough to semi-supervised classification. Conference on Computer Engineering and Sustainable Development.

٥. Basti, E. (۲۰۱٥). Analyaing initial public offerings' short-term performance using decision tree and svm. *Decision support systems, ۱۳:* ۱٥-۲۷.

٦. Bolo, Gh and Falah barandegh.M. (۲۰۱۳). The relationship between conservatism and short-term abnormal returns of primary public supply stocks, with an emphasis on the role of information asymmetry models. *Quarterly Journal of Empirical Accounting Studies. ۳۹:* ٥۷-۸۲.

۷. Ceren, R and Richardson, W. (۲۰۱٦). Predicting IPO Performance from Nearest Neighbors Using TF-IDF Weighted Word Count Vectors. Department of Computer Science, Athens, GA ۳۰٦۰۲.

۸. Delen, D and Kuzey, H. (۲۰۱۳). A comparative analysis of machine learning systems for measuring the impact of knowledge manegmentparactices. *Decision Support System, ٥٤(۲):* ۱۱٥۰-۱۱٦۰.

۹. Dmory, D. (۲۰۰۲). Identification of factors affecting the long-run returns of newly accepted shares in Tehran Stock Exchange and analyzing the interactions of these factors. (Ph.D. Financial Management). University of Tehran, Iran.

۱۰. Falah por, S. (۲۰۱۲). Investigating price manipulation in Tehran Stock Exchange using a backup car model. *Journal of Financial Research, ۱٤:* ٦۹-۸٤.

۱۱. Khodaparasti, S. (۲۰۱۳). Factors Affecting the Short-Term and Long-Term Returns of Shares Delivered in Early Issues in the Tehran Stock Exchange. *Quarterly Journal of Empirical Accounting Studies. ۱۱:* ۱۷۹-۲۰۰.

۱۲. Kohavi, R. (۱۹۹٥). A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceedings of the ۱٤th International Conference, l:* ۱۱۳۷-۱۱٤٥.

۱۳. Kohavi, R and Provost, F. (۱۹۹۸). Editorial for the Special issue on applications of machins learning and the knowledge discovery prosecc. *machine learning, ۳۰(۲):* ۲۷۱-۲۷٤.

۱٤. Ljungqvist, A (۲۰۰٥). IPO Underpricing. Tuck School Of Business at Dartmouth, handbook of corporat finance: *Empirical Corporate Finance, ۱(۱۲).*

۱٥. Malaei, Masoud & SHeikhi, Mohammad Javad & KHodamoradi, Saeed. (۲۰۰۹). Optimization of Markowitz Risk Management Models, Value at Risk and Value at Risk Parametric Using local and Global Algorithms in Tehran Stock Exchange. *Financial Management Perspective, ۱:* ٦۷-۹۷.

۱٦. Maleki, R. (۲۰۰۸). IPO public offering. Tehran Stock Exchange.

۱۷. Moli, M and Migliorati,K. (۲۰۱٤). Forecasting winner IPOs. *Investment Analysts Journal, ۷۹.*
۱۸. Poyan far, A and Falah por, S. (۲۰۱۵). Using the Hybrid Method Feature Selection and the Nearest Neighborhood Algorithm for predicting the daily turnover of ۵۰ more active companies in Tehran Stock Exchange Using the Hybrid Method Feature Selection and the Nearest Neighborhood Algorithm for predicting daily movements of ۵۰ more active companies of Tehran Stock Exchange. Journal of Financial Engineering and Management of Securities, No. ۲۵. *Journal of Financial Engineering and Management of Securities, ۲۵.*
۱۹. Zarif fard, A and Mehr joo, H. (۲۰۰٤). Investigate Stock Price Performance at the First Stock Exchanges in Tehran Stock Exchange. *Quarterly Journal of Empirical Financial Accounting Studies, ۸.*
۲۰. Tong, S and Kian, CH. (۲۰۱۲). Predicting IPOs Performance Using Generalized Growing and Pruning Algorithm for Radial Basis Function (GGAP-RBF) Network. *۱۲(۱).*