



استفاده از انتخاب ویژگی جهت بهبود صحت پیش‌بینی نتایج مسابقات بسکتبال

مارال حقیقت^۱، حمید رستگاری^۲، نسیم نورافزا^۳

^۱دانشجوی کارشناسی ارشد دانشکده کامپیوتر، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد نجف‌آباد، اصفهان؛ maral.haghighat.۹۳۶۶@gmail.com

^۲استادیار و عضو هیئت‌علمی دانشکده کامپیوتر، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد نجف‌آباد، اصفهان؛ hamid_rastegari@yahoo.com

^۳عضو هیئت‌علمی دانشکده کامپیوتر، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد نجف‌آباد، اصفهان؛ n_noorafza@yahoo.com

چکیده

امروزه پیش‌بینی نتایج مسابقات ورزشی موضوعی جالب است که توجهات زیادی را به خود جلب کرده است. از این رو از داده‌کاوی به عنوان تکنیکی برای تجزیه و تحلیل داده‌های جمع‌آوری شده در ورزش استفاده شده است. یکی از این ورزش‌ها بسکتبال است که به علت تحرکش، تعداد زیادی رویداد در طول یک بازی اتفاق می‌افتد. این رویدادها باعث ایجاد مجموعه داده‌هایی با ابعاد بالا می‌شوند و از آنجا که تعدادی از این ویژگی‌ها ممکن است در پیش‌بینی نتیجه یک مسابقه بی‌تأثیر باشند و یا تأثیر منفی داشته باشند، ما در این مقاله روشی مبتنی بر الگوریتم ژنتیک برای انتخاب ویژگی‌های مؤثر در پیش‌بینی نتیجه مسابقات بسکتبال ارائه کردیم که بر طبق آزمایشات انجام شده، صحت پیش‌بینی نتایج را افزایش داده است.

کلمات کلیدی

پیش‌بینی، انتخاب ویژگی، الگوریتم ژنتیک، صحت پیش‌بینی

۱- مقدمه

دنیای ورزش با اطلاعات آماری گسترده‌ای برای هر بازیکن، تیم، بازی و فصل شناخته شده است. از این رو سازمان‌های ورزشی از داده‌های فرم تجزیه و تحلیل آماری، کشف الگو، پیش‌بینی نتیجه، ارزیابی عملکرد، ارزیابی حریف آینده، انتخاب بازیکن و غیره استفاده می‌کنند [۶]. امروزه پیش‌بینی نتایج مسابقات در میان هواداران سرتاسر دنیا بسیار محبوب شده است که این امر در ورزش‌هایی مثل فوتبال و بسکتبال مشهود است. به علت تحرکات زیادی که در بسکتبال وجود دارد، در طول یک بازی تعداد زیادی رویداد اتفاق می‌افتد که بر روی بازی تأثیر می‌گذارند. این رویدادها عبارت‌اند از: شوت‌ها، ریباندها، از دست دادن توپ، پاس منجر به گل و غیره [۱۱]. در کنار این آمارها، آمارهای دیگری در رابطه با وضعیت تیم‌ها در طول فصل وجود دارد که با قرار دادن تمام این ویژگی‌های آماری در کنار هم مجموعه داده‌ای با ابعاد بالا ایجاد می‌شود که ممکن است عمل پیش‌بینی را با مشکل مواجه کند و هر دفعه زمان زیادی از سیستم پیش‌بینی صرف بررسی تمام ویژگی‌ها شود. الگوریتم‌های جمعیت‌گرا همچون الگوریتم ژنتیک، به اکتشاف و استخراج فضای جستجو می‌پردازند و فضای جستجو را با عامل‌های خود بررسی می‌کنند. این روش‌ها سعی می‌کنند با استفاده از دانش موجود و تکرارهای قبلی به جواب‌های بهتر برسند. این روش برای ایجاد راه‌حل‌های مناسب در مسائل بهینه‌سازی و جستجو به کار برده می‌شوند.

در این مقاله ما سعی داریم روشی مبتنی بر الگوریتم ژنتیک برای انتخاب ویژگی ارائه دهیم تا صحت دسته‌بندی را توسط پیدا کردن ویژگی‌های مهم‌تر افزایش دهیم. این مقاله بصورت زیر سازمان‌دهی شده است: در بخش ۲ نحو انتخاب ویژگی بین داده‌های موجود و کارهای مرتبط در این زمینه ارائه شده است. بخش ۳ مجموعه داده مورد استفاده را معرفی کرده و در بخش ۴ روش پیشنهادی را شرح داده شده است. در بخش ۵ نتایج بدست آمده بیان شده و در نهایت در بخش ۶ نتیجه‌گیری نهایی صورت گرفته است.

۲- انتخاب ویژگی

ابعاد بالای ویژگی‌ها، چالش جدیدی را در برابر فرایند یادگیری در الگوریتم‌های یادگیری ماشین قرار داده است. با وجود ویژگی‌های نامرتبط، مدل یادگیری جامعیت کمتری خواهد داشت و به سمت یادگیری بیشتر از حد سوق پیدا می‌کند. انتخاب ویژگی یک روش مؤثر برای شناسایی ویژگی‌های مرتبط جهت کاهش ابعاد ویژگی‌ها است [۴،۷]. مطالعات زیادی نشان دادند که ویژگی‌ها حذف شدند بدون اینکه کارایی کاهش پیدا کند. الگوریتم‌های ژنتیک بسته به استراتژی‌های طراحی به ۳ دسته تقسیم می‌شوند: فیلتر، بسته‌بندی و تعبیه شده.

روش‌های مبتنی بر فیلتر بدون نیاز به یک دسته‌بندی، ویژگی‌ها را انتخاب می‌کنند و برای ارزیابی یک زیرمجموعه از ویژگی‌ها از معیارهای مستقل بهره می‌گیرند. آن‌ها زیرمجموعه‌ای بهینه از ویژگی‌ها را بر مبنای خصوصیات آماری داده مانند انحراف معیار، وابستگی و سازگاری، انتخاب می‌کنند. بنابراین سرعت اجرای بالایی دارند، مثل روش CFS. روش‌های مبتنی بر بسته‌بندی این از یک الگوریتم یادگیری از پیش تعریف شده استفاده می‌کنند و از معیار کارایی الگوریتم یادگیری برای ارزیابی زیرمجموعه بهینه از ویژگی‌ها استفاده می‌کنند. آن‌ها ویژگی‌های مناسب‌تری را نسبت به الگوریتم‌های فیلتر انتخاب می‌کنند اما کندتر هستند، مثل الگوریتم ژنتیک. در روش‌های تعبیه شده الگوریتم دسته‌بندی در درون خود، انتخاب ویژگی را (بطور صریح یا ضمنی) انجام می‌دهد. این الگوریتم‌ها سعی در تمرکز بر روی ویژگی‌های مرتبط دارند و از ویژگی‌های غیرمرتبط در حین یادگیری صرف‌نظر می‌کنند، مثل درخت تصمیم [۱۲،۱۰].

Bor-Wen و Shu-Ting [۱۳] از دو روش forward selection و backward selection برای انتخاب ویژگی در تصاویر ماموگرافی استفاده کردند و با حذف ویژگی‌های نامرتبط باعث بهبود در پیش‌بینی سرطان سینه شدند. Nithya و همکارانش [۹] از روش Maximum Difference برای انتخاب ویژگی استفاده کردند و صحت دسته‌بندی شبکه عصبی را به میزان زیادی افزایش دادند. Tsymbal و همکارانش [۱۶] از دو روش جدید انتخاب ویژگی مبتنی بر استراتژی sequential-search-based استفاده کردند و صحت دسته‌بندی بیزین ساده را برای مسئله تشخیص دردهای شکمی حاد افزایش دادند. Abeel و همکارانش [۱] از یک انتخاب ویژگی ensemble استفاده کردند و به این نتیجه رسیدند که روش آن‌ها تأثیر خوبی بر روی دسته‌بندی توسط ماشین بردار پشتیبان می‌گذارد. و همچنین Lu و همکارانش [۸] روشی را بر مبنای الگوریتم ژنتیک برای انتخاب ویژگی‌های مؤثر در بهبود تفسیر تصاویر در حوزه پردازش تصویر ارائه کردند.

۳- مجموعه داده

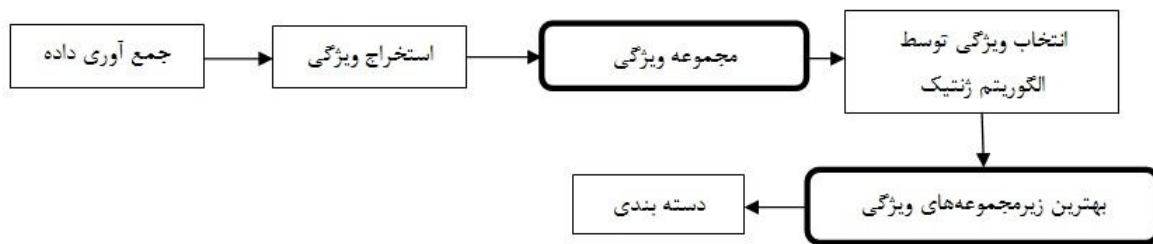
وب سایت Basketball-reference.com منبع اصلی داده ما برای مدل آموزش و تست است. Basketball-reference.com وب سایتی است که

برای جمع‌آوری داده بویژه برای بسکتبال استفاده می‌شود و همچنین داده‌های ورزش‌های دیگر مثل بیس‌بال و فوتبال را هم جمع‌آوری می‌کند. این سایت یکی از معروف‌ترین سایت‌ها برای کاوش داده‌های ورزشی است و در آزمایشات زیادی مورد استفاده قرار گرفته است. پس از جمع‌آوری داده‌های مورد نیاز از این سایت و انجام استخراج ویژگی بر روی آن‌ها، در نهایت مجموعه داده نهایی بدست آمده شامل ۳۶ ویژگی شد که ۳۵ ویژگی از آن ویژگی‌های آماری دو تیم و یک ویژگی آن نتیجه بازی (برد و باخت) است. این مجموعه داده شامل ۱۵۵۸ نمونه داده است که تمام این داده‌ها مربوط به لیگ ۲۰۱۲-۲۰۱۳ بسکتبال آمریکا (NBA) می‌باشند.

۴- روش پیشنهادی

پس از جمع‌آوری داده‌ها از اینترنت و استخراج ویژگی‌های مورد نیاز از روی این داده‌ها، از روش انتخاب ویژگی جهت کاهش ابعاد مجموعه داده استفاده می‌کنیم. روند کار الگوریتم انتخاب ویژگی تحقیق ما بصورت زیر است:

الگوریتم ژنتیک به جستجوی فضای ویژگی‌ها می‌پردازد و زیرمجموعه بهینه‌ای از ویژگی‌هایی که کارایی دسته‌بندی کننده را نسبت به جمعیت قبل افزایش می‌دهد، انتخاب می‌کند. پس از اتمام عمل انتخاب ویژگی توسط الگوریتم ژنتیک، ویژگی‌های انتخاب شده توسط الگوریتم‌های دسته‌بندی کننده مورد استفاده قرار می‌گیرند. جزئیات بیشتر درباره الگوریتم ژنتیک در قسمت بعد عنوان می‌شود. شمای کلی سیستم در شکل (۱) آورده شده است.



شکل (۱): شمای کلی سیستم

۴-۱- الگوریتم ژنتیک

الگوریتم ژنتیک الهامی از علم ژنتیک و نظریه تکامل داروین است و بر اساس بقای برترین‌ها یا انتخاب طبیعی استوار است. یک کاربرد متداول الگوریتم ژنتیک، استفاده از آن به عنوان تابع بهینه‌کننده است. الگوریتم ژنتیک ابزار سودمندی در بازشناسی الگو، انتخاب ویژگی، پردازش تصویر، دسته‌بندی و پیش‌بینی است. در الگوریتم‌های ژنتیک، نحوه تکامل ژنتیکی موجودات زنده شبیه‌سازی می‌شود [۱۵]. الگوریتم ژنتیک یکی از قویترین روش‌های برگرفته از طبیعت است که به جستجوی فضای مسأله به صورت تصادفی هدایت شده پرداخته که این جستجو در قالب تلاش جهت ایجاد جواب‌هایی بهتر، در هر نسل نسبت به جواب‌های نسل قبل صورت می‌گیرد [۱۴].

ویژگی اساسی الگوریتم ژنتیک ساده بودن آن می‌باشد. در الگوریتم ژنتیک ابتدا پاسخ مسئله در قالب یک ساختار کروموزومی تعریف می‌شود. با معرفی تابع تناسب، کیفیت پاسخ‌های ارائه شده در هر کروموزوم به صورت عدد مشخص می‌شود. سپس تعداد معینی کروموزوم به صورت تصادفی (یا شبه تصادفی) تولید می‌کند که این کروموزوم‌ها به عنوان جمعیت اولیه شناخته می‌شوند. در این مرحله تعدادی پاسخ برای مسئله وجود دارد که عمدتاً از کیفیت پایینی برخوردار هستند. میزان کیفیت هر کروموزوم از جمعیت بر اساس تابع تناسب مشخص می‌شود. حال با استفاده از روشی مناسب (روشی که در آن احتمال انتخاب کروموزوم با مقدار تناسب بهتر بیشتر از کروموزوم دارای مقدار تناسب ضعیف‌تر باشد) دو کروموزوم جهت تولیدمثل انتخاب می‌شوند. سپس با استفاده از این دو کروموزوم، کروموزوم جدیدی ایجاد می‌شود. سپس با یک احتمال مشخص تعدادی از ژن‌های بعضی از کروموزوم‌ها را تغییر می‌یابد. انجام مراحل بالا باعث ایجاد جمعیت جدید از کروموزوم‌ها می‌گردد. در صورت همگرایی کروموزوم‌ها به پاسخ مطلوب عملیات تولید نسل متوقف می‌گردد در غیر این صورت ایجاد هر نسل از نسل قبلی تا رسیدن به جواب مطلوب یا برقراری شرط پایان الگوریتم ادامه می‌یابد [۵].

مسئله انتخاب ویژگی را می‌توان جز مسائل NP کامل در نظر گرفت [۲]. الگوریتم ژنتیک زمانی که با مسائل NP کامل سروکار دارد بسیار خوب عمل می‌کند. الگوریتم ژنتیک را می‌توان برای مسئله انتخاب ویژگی بکار برد زیرا آن در فضاهای جستجو با ابعاد بالا بسیار مؤثر عمل می‌کند.

وظیفه الگوریتم ژنتیک این است که زیرمجموعه بهینه m را از فضای جستجو با ابعاد d بیابد که $m < d$ است. برای اینکه بتوانیم از الگوریتم ژنتیک برای انتخاب ویژگی استفاده کنیم، باید ساختار مناسب برای آن تعریف شود. اجزای الگوریتم ژنتیک عبارتند از: نحوه نمایش کروموزوم، تابع تناسب، عملگر انتخاب، عملگرهای تقاطع و جهش. جزئیات هر یک از این اجزا در ادامه آورده شده است.

۴-۱-۱- نحوه نمایش کروموزوم

وقتی که الگوریتم ژنتیک برای انتخاب ویژگی استفاده می‌شود، کروموزوم‌ها به دو روش قابل نمایش هستند. روش اول نمایش بصورت بردار باینری از n ژن‌ها است که هر ژن نشان دهنده یک ویژگی است. زمانی که ژن i م برابر با یک باشد نشان دهنده این است که ویژگی i م انتخاب شده و اگر برابر با صفر باشد یعنی ویژگی i م انتخاب نشده است. در روش دوم نمایش به هر ویژگی وزنی اختصاص داده می‌شود و ویژگی با وزن بیشتر نشان دهنده اهمیت بالای ویژگی است. در این تحقیق نحوه نمایش باینری جهت نمایش کروموزوم‌ها انتخاب شده است و طول هر کروموزوم برابر با ۳۵ ویژگی (تعداد کل ویژگی‌های موجود در مجموعه داده به غیر از ویژگی نتیجه) است. در شکل (۲) این نحوه نمایش، نشان داده شده است.

ویژگی ۱	ویژگی ۲	ویژگی ۳	ویژگی ۴	...	ویژگی n
۱	۰	۰	۱	...	۱

۱: ویژگی انتخاب شده، ۰: ویژگی انتخاب نشده، n: تعداد کل ویژگی‌ها

شکل (۲): نحوه نمایش باینری کروموزوم برای انتخاب ویژگی توسط الگوریتم ژنتیک

۴-۱-۲- تابع تناسب

برای این که بتوانیم هر جواب (کروموزوم) را ارزیابی کنیم، یک تابع تناسب باید استفاده شود. یک تابع تناسب نوع خاصی از تابع هدف است که میزان کیفیت یک جواب را نشان می‌دهد. به علت اینکه هدف ما افزایش صحت پیش‌بینی است، ما تابع تناسب را مطابق فرمول صحت انتخاب می‌کنیم:

$$F(i) = \frac{TP+TN}{TP+TN+FP+FN}, \quad i = 1, 2, 3, \dots, n \quad (1)$$

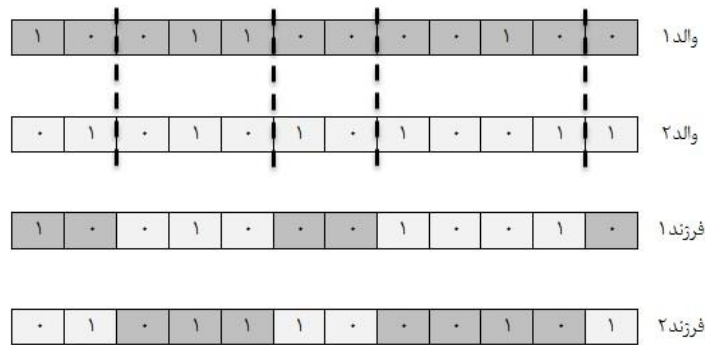
فرمول بالا به ازای هر کروموزوم در جمعیت محاسبه می‌شود که TP مثبت درست، TN منفی درست، FP مثبت غلط و FN منفی غلط می‌باشند.

۴-۱-۳- عملگر انتخاب

در فرایند انتخاب، افراد جمعیت بر طبق تابع تناسب انتخاب می‌شوند و یک جمعیت جدید را شکل می‌دهند. افراد جمعیت با تابع تناسب بالاتر با احتمال بیشتری برای تشکیل جمعیت جدید انتخاب می‌شوند. در این تحقیق ما از عملگر انتخاب چرخ رولت استفاده کرده‌ایم.

۴-۱-۴- عملگر تقاطع

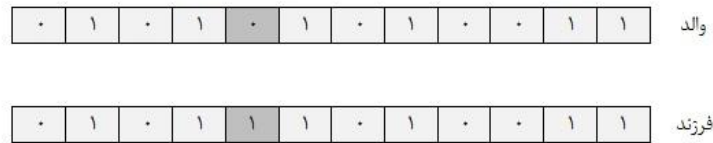
عملگر تقاطع برای ایجاد تغییر در کروموزوم‌ها از یک نسل به نسلی دیگری استفاده می‌شود. این عملگر زیرمجموعه‌ای از ژن‌های دو کروموزوم را با هم تعویض کرده و دو فرزند جدید ایجاد می‌کند. فرایند کلی عملگر تقاطع با یک احتمال انجام می‌شود که این احتمال به هر جفت از والدین تعلق می‌گیرد تا فرزندان جدیدی را تولید کنند. در این تحقیق ما از عملگر تقاطع چهار نقطه‌ای برای ایجاد گوناگونی بهتر در میان کروموزوم‌ها استفاده کرده‌ایم. روند کار به این صورت است که ابتدا ۴ نقطه تصادفی در طول کروموزوم ایجاد می‌شود و سپس والدین در این نقاط به ۵ قسمت تقسیم می‌شوند و هر فرزند با انتخاب تکه‌های فرد از یک والد و تکه‌های زوج از والد دیگر ساخته می‌شود. شکل (۳) نحوه عملکرد این روش را نشان می‌دهد.



شکل (۳): تقاطع چهار نقطه‌ای

۴-۱-۵- عملگر جهش

عملگر جهش برای ایجاد تنوع در یک نسل از جمعیت کروموزوم‌ها بکار برده می‌شود. این عملگر به صورت تصادفی برخی از ژن‌های کروموزوم را با یک احتمال ثابت در هر نسل از جمعیت تغییر می‌دهد. هدف این عملگر ایجاد تنوع بیشتر در جستجو است و به الگوریتم ژنتیک کمک می‌کند از نقاط بهینه محلی فرار کند. در این تحقیق با توجه به احتمال جهش، هر یک از ژن‌های کروموزوم از یک به صفر یا از صفر به یک تغییر می‌کنند. شکل (۴) چگونگی عملکرد این عملگر را نشان می‌دهد.



شکل (۴): عملگر جهش

پارامترهای تنظیم شده برای الگوریتم ژنتیک ما در جدول (۱) ارائه شده است.

جدول (۱): پارامترهای الگوریتم ژنتیک

تعداد افراد	تعداد نسل	احتمال تقاطع	احتمال جهش
۱۰۰	۱۰۰	۰.۵	۰.۰۲

۵- نتایج تجربی

برای ارزیابی معیار صحت پیش‌بینی روش خود، آزمایشاتی بر روی مجموعه داده بسکتبال انجام دادیم. در جدول (۲) چهار زیرمجموعه از ویژگی‌های انتخاب شده بعد از ۲۰ بار اجرای الگوریتم پیشنهادی نشان داده شده است.

جدول (۲): چهار زیرمجموعه ویژگی بر اساس الگوریتم ژنتیک

نام	تعداد ویژگی‌ها	لیست ویژگی‌ها (برحسب شماره ویژگی)
۱#	۵	۵، ۴، ۳، ۲، ۱
۲#	۷	۷، ۶، ۵، ۴، ۳، ۲، ۱
۳#	۷	۲۷، ۱۴، ۵، ۴، ۳، ۲، ۱
۴#	۹	۲۸، ۲۷، ۱۵، ۱۴، ۵، ۴، ۳، ۲، ۱

با توجه به این نکته که در بسکتبال مجموعه داده عمومی برای مقایسه کارمان با کار دیگر محققان وجود نداریم، برای سنجش کار خود الگوریتم‌های دسته‌بندی استفاده شده توسط Cao [۱۵] را بر روی کل مجموعه داده و زیرمجموعه ویژگی‌های انتخاب شده اجرا کرده‌ایم. میزان صحت دسته‌بندی بصورت ۱۰-fold cross validation بدست آمده است. جدول (۳) نتایج بدست آمده از این آزمایشات را نشان می‌دهد.

جدول (۳): محاسبه صحت دسته‌بندی (%) برای کل ویژگی‌ها و زیرمجموعه‌های ویژگی‌های انتخاب شده

ویژگی‌ها	ماشین بردار پشتیبان	شبکه عصبی	بیز ساده
تمام ویژگی‌ها	۵۶.۲	۶۶	۶۶.۲
۱#	۵۷.۱	۶۷.۶	۶۸.۵
۲#	۵۶.۶	۶۷.۸	۶۸.۵
۳#	۵۸.۳	۶۷.۲	۶۸
۴#	۵۹.۱	۶۷.۳	۶۷.۳

با بررسی جدول (۳) به این نتیجه می‌رسیم که توسط انتخاب ویژگی با الگوریتم ژنتیک توانسته‌ایم تا صحت دسته‌بندی ماشین بردار پشتیبان، شبکه عصبی و بیز ساده را به ترتیب در حدود ۰.۳٪، ۰.۲٪ و ۰.۵٪ را افزایش دهیم.

۶- نتیجه‌گیری

امروزه پیش‌بینی نتایج بازی‌ها به یکی از موضوعات مورد علاقه سازمان‌های ورزشی تبدیل شده‌است. تا به امروز سیستم‌های محدودی در زمینه پیش‌بینی نتایج بازی‌های بسکتبال توسط محققان با استفاده از روش‌های داده‌کاوی ایجاد شده‌اند که کمک زیادی به مدیران، مربیان و متخصصان در جهت تصمیم‌گیری‌های آتی‌شان کرده‌اند و مانع تحلیل اشتباه و گمراه‌کننده از آمارهای موجود می‌شوند. اما با این حال، صحت و درستی پیش‌بینی در سیستم‌های پیش‌بینی نتایج مسابقات بسکتبال موجود کافی نبوده و پایین است و هنوز راه زیادی تا کسب اطمینان سازمان‌های ورزشی نسبت به نتایج این سیستم‌ها در پیش است.

در این مقاله برای افزایش صحت پیش‌بینی نتایج مسابقات بسکتبال NBA روشی مبتنی بر الگوریتم ژنتیک ارائه شد تا عمل انتخاب ویژگی را بر روی ویژگی‌های مجموعه داده انجام دهد و با پیدا کردن ویژگی‌های مهم‌تر در پیش‌بینی به الگوریتم‌های موجود در این حوزه کمک کند. بر طبق آزمایشات، از میان ۳۵ ویژگی، دسته ویژگی‌های ۵ تایی، ۷ تایی و ۹ تایی انتخاب شدند که در نهایت تمام این مجموعه‌ها به بهتر شدن صحت پیش‌بینی منجر شدند.

۷- مراجع

- [۱] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. saeye. Robust biomark identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics*, ۲۶(۳):۳۹۲-۳۹۸, Feb ۲۰۱۰.
- [۲] E. Amaldi and V. Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, ۲۰۹(۱-۲):۲۳۷-۲۶۰, December ۱۹۹۸.
- [۳] C. Cao. *Sports data mining technology used in basketball outcome prediction*. Master dissertation, Dublin Institute of technology, ۲۰۱۲.
- [۴] I. Guyon and A. Elisseeff. An introduction to variable a feature selection. *Journal of Machine Learning Research*, ۳:۱۱۵۷-۱۱۸۲, ۲۰۰۳.
- [۵] R. L. Haupt, and S. E. Haupt. *Parctical Genetic Algorithm (۳rd ed)*. Wiley, USA, ۲۰۰۴.
- [۶] M. Haghighat, H. Rastegari, and N. Nourafza. A Review of Data Mining Techniques for Result Prediction in Sports. *ACSIJ Advances in Computer Science: an International Journal*, ۲(۵):۷-۱۲, November ۲۰۱۳.
- [۷] H. Liu and L. Yu. Toward Integrating Feature Selection Algorithms for Classification and Clustering. The ۱۷th *IEEE Transaction on Knowledge and Data Engineering*, pages ۴۹۱-۵۰۲, ۲۰۰۵.
- [۸] J. Lu, T. Zhao, and Y. Zhang. Feature selection based-on genetic algorithm for image annotation. *Expert system with applications*, ۳۱(۸):۸۸۷-۸۹۱, ۲۰۰۸.
- [۹] R. Nithya and B. Santh. Mammogram Classification using Maximum Difference Feature Selection Method. *Journal of Theoretical and Applied Information Technology*, ۳۳(۲):۱۹۷-۲۰۴, November ۲۰۱۱.
- [۱۰] D. Meladenic. Feature selection for dimensionality reduction. *Lecture Notes in Computer Science*, ۳۹۴۰:۸۴-۱۰۲, ۲۰۰۶.
- [۱۱] D. Miljkovic, L. Gajic, A. Kovacevic, and Z. Konjovic. The use of data mining for basketball matches outcomes prediction. *IEEE 14th International Symposium on intelligent and informatics*, pages ۳۰۹-۳۱۲, Subotica, Serbia, September ۲۰۱۰.
- [۱۲] L. C. Molina, L. Belanche, and A. Nebot. Feature selection algorithms: a survey and experimental evaluation. *IEEE*



International Conference on Data Mining, pages:۳۰۶-۳۱۳, ۲۰۰۲.

[۱۳] L. Shu-Ting and Ch. Bor-Wen. Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods. *J Med Syst*, DOI ۱۰.۱۰۰۷/s۱۰۹۱۶-۰۱۰-۹۵۱۸-۸, ۲۰۱۰.

[۱۴] R. Sikaro and S. Piramuthu. Framework for efficient feature selection in genetic algorithm based data mining. *European Journal of Operational Research*, ۱۸۰(۲):۷۲۳-۷۳۷, July ۲۰۰۷.

[۱۵] L. Y. Tseng and S. Yang. Genetic algorithms for clustering, feature selection and classification. *IEEE Int. Conference on Neural Networks*, pages ۱۶۱۲-۱۶۱۶, ۱۹۹۷.

[۱۶] A. Tsymbal, P. Cunningham, M. Pechenizkiy, and S. Puuronen. Search Strategies for Ensemble Feature Selection in Medical Diagnostics. *Proceedings of the ۱۶th IEEE Symposium on Computer-Based Medical Systems*, pages ۱۲۴-۱۲۹, June ۲۰۰۳.