
پیش‌بینی موفقیت کشورهای شرکت‌کننده در بازی‌های آسیایی بر اساس متغیرهای کلان اقتصادی، سیاسی، اجتماعی و فرهنگی

دکتر مهرعلی همتی‌نژاد^۱، دکتر محمد رحیم رضائیان^۲، دکتر محمد حسن قلیزاده^۳، شهرام

شفیعی^۴، امین قاضی زاهدی^۵

ص ص: ۸۶-۵۷

تاریخ دریافت: ۱۳۹۱/۴/۸

تاریخ تصویب: ۱۳۹۱/۱۰/۲۰

چکیده

پیروزی‌های بین‌المللی بویژه پیروزی در بازی‌های آسیایی به شکل چشمگیری برای بسیاری از کشورها حائز اهمیت شده است. اگرچه بسیاری از کشورها مبالغ هنگفتی را برای رقابت با کشورهای دیگر در ورزش هزینه می‌کنند، اما مدارک و دلایل کافی وجود ندارد که نشان بدهد متغیرهای ورزشی چگونه می‌توانند بر پیروزی کشورها در سطوح بین‌المللی اثرگذار باشند؟ در این مقاله ما تلاش کردیم تا پیروزی کشورها را در بازی‌های آسیایی از راه متغیرهای کلان اقتصادی، سیاسی، فرهنگی و اجتماعی پیش‌بینی کنیم. اطلاعات همه متغیرهای جمعیت شهری، هزینه آموزش و پرورش، ساختار سنی، تولید واقعی ناخالص داخلی، سرانه تولید ناخالص داخلی، بیکاری، جمعیت، نرخ تورم، تعادل حساب جاری، امید به زندگی و تراز بازرگانی برای همه کشورهای شرکت‌کننده در بازی‌های آسیایی از سال ۱۹۷۰ تا ۲۰۰۶ به منظور طراحی مدل مورد استفاده قرار گرفت و مدل برای سال ۲۰۱۰ آزمایش شد. پیش‌بینی رتبه کشورها بر اساس تعداد مدال‌های طلا کسب شده آنها بود. در این پژوهش نرم‌افزار WEKA (که یک نرم‌افزار ماشین یادگیری است) مورد استفاده قرار گرفت. نتایج پژوهش نشان داد که ژاپن در همه دوره‌های برگزاری مسابقه‌های خود دارای ثبات بیشتری بود. ضریب همبستگی بین رتبه‌های پیش‌بینی شده و واقعی بر اساس مدل درختی ۷۵٫۵٪ دیده شد. بیکاری، جمعیت و ساختار سنی، بالاترین سهم (به ترتیب ۲۴٫۶٪ و ۱۸٫۵٪ و ۱۶٫۱٪) را در پیش‌بینی پیروزی کشورهای شرکت‌کننده در بازی‌های آسیایی ۲۰۱۰ به خود اختصاص می‌دهند. سهم متغیرهای جمعیت شهری، تولید واقعی ناخالص داخلی، امید به زندگی، هزینه آموزش و پرورش، سرانه تولید ناخالص داخلی، تراز بازرگانی، نرخ تورم و تعادل حساب جاری به ترتیب برابر با ۹٫۷۱٪، ۸٫۳٪، ۷٫۱۱٪، ۶٫۷۷٪، ۴٫۹۶٪، ۳٫۵۸٪، ۰٫۲۷٪ و ۰٫۱٪ است. با استفاده از مدل طراحی شده،

۳،۲،۱- دانشیار دانشگاه گیلان

۴- استادیار دانشگاه گیلان، shafieeshahram@gmail.com

۵- کارشناس ارشد دانشگاه امیر کبیر تهران

برای هر کشور این امکان فراهم است تا در راستای توسعه ورزش و کسب کرسی‌های بین‌المللی با توجه به امکانات و منابع بالقوه خود و مقایسه با کشورهای دیگر، سیاست‌های مناسبی را اتخاذ کند؛ به گونه‌ای که مدیران و برنامه‌ریزان اهداف بلندمدت، میان‌مدت و کوتاه‌مدت خود را بر اساس فاکتورهای اقتصادی، سیاسی، اجتماعی و فرهنگی تعیین کنند.

واژگان کلیدی: پیش‌بینی، بازی‌های آسیایی، متغیرهای کلان و مدل درخت

مقدمه

در سالیان اخیر شاهد حرکتی مستمر از پژوهش‌های کاملاً نظری نسبت به پژوهش‌های کاربردی بویژه در زمینه پردازش اطلاعات بوده ایم آن هم برای مسائلی که یا برای آنها راه حلی وجود ندارد و یا به سادگی قابل حل نیستند. با توجه به این امر، میل به توسعه تئوریک سیستم‌های دینامیکی هوشمند مدل آزاد^۱ که مبتنی بر داده‌های تجربی هستند، رو به افزایش است. در این میان، هوش محاسباتی^۲ یا محاسبات نرم^۳ به معنای هوش، دانش، الگوریتم یا نگاهت از دل محاسبات عددی بر اساس ارائه به روز داده‌های عددی نقش کلیدی را ایفا می‌کند که مولفه‌های اساسی آن شبکه‌های عصبی مصنوعی^۴ (محاسبات نرونی)، منطق فازی^۵ (محاسبات تقریبی) و الگوریتم ژنتیک^۶ (محاسبات ژنتیکی) هستند و هر یک مغز انسان را الگو قرار داده‌اند. شبکه‌های عصبی، ارتباط‌های سیناپسی و ساختار نرونی، منطق فازی استنتاج‌های تقریبی، محاسبه‌های ژنتیکی و محاسبه‌های موتاسیونی مغز را مدل می‌کنند(۲).

یکی از امور مورد علاقه هوش محاسباتی، پیش‌بینی^۷ اموری است که در آینده رخ می‌دهد. هر چند که این پدیده در طول تاریخ همواره مورد علاقه انسان بوده است، اما هوش محاسباتی از منظری نوین و متفاوت به آن می‌نگرد. پیش‌بینی روند تغییرات در همه حوزه‌ها یکی از دغدغه‌های اصلی مدیران و برنامه‌ریزان است؛ اما در مسیر پیش‌بینی همواره مشکلاتی وجود داشته که انجام دادن پیش‌بینی‌های دقیق و قابل اعتماد را تقریباً غیرممکن ساخته است. وجود پارامترهای زیاد و در بسیاری موارد پنهان، پیش‌بینی را به مسأله‌ای پیچیده تبدیل کرده که الگوریتم‌های غول‌پیکر ریاضی را نیز از ارائه راهکارهای مناسب برای ساخت یک الگو پیش‌بینی کارآمد عاجز کرده است. پیش‌بینی برای نخستین بار در برنامه‌ریزی مورد استفاده قرار گرفت. مشاهده تکرار در وقوع

1- Model Free

2- Computational Intelligence (CI)

3- Soft Computing

4- Artificial Neural Network (ANN)

5- Fuzzy Logic

6- Genetic algorithm

7- Prediction

رویدادها این پیش فرض را به وجود آورد که می توان نتایج و تبعات آن را به همه رویدادهایی از آن جنس، تعمیم داد. در دو دهه اخیر با پیدایش هوش مصنوعی و ترکیب آن با علم پرسابقه آمار، در کنار الگوریتم های پیشرفته و ابتکاری همچون الگوریتم ژنتیک، روش های متاهوریستیک، شبکه های عصبی مصنوعی و... تحول گسترده ای در این گستره ایجاد شده است (۳).

پرسشی که در اینجا مطرح می شود، این است که آیا پیش بینی هایی که با استفاده از مدل های آماری صورت می گیرد، دقیق تر و قابل اعتمادتر از پیش بینی هایی است که افراد خبره بر اساس قضاوت های ذهنی خود صورت انجام می دهند؟ افراد خبره با قضاوت های ذهنی ممکن است در شرایط تا حدودی نامعلوم^۱ و بیشتر نامطمئن^۲ به علت استفاده از معیار و ملاک های کیفی در مقایسه با مدل های آماری پیش بینی بهتری را انجام دهند (۳۳). برای مقایسه این دو روش در زمینه های گوناگون همچون، پزشکی^۳، موفقیت پیروزی های دانشگاهی^۴، تصمیم گیری در تجارت^۵، پیش بینی وضعیت آب و هوا^۶، پیش بینی های اقتصاد کلان^۷، نرخ تورم، نتایج انتخاب های سیاسی^۸ و... بررسی های بی شماری صورت پذیرفته است که در بیشتر آنها مدل های آماری پیش بینی بهتری کردند (۳۶، ۳۵، ۲۱، ۲۰، ۱۳، ۱۰)، در ورزش نیز این پژوهش ها، پیاپی انجام شده است (۱۸، ۱۷، ۴۸).

در دنیای پیچیده و پیشرفته ورزش کنونی تصمیم گیری درست، علمی و به موقع نقش بسیار مهم و تعیین کننده ای در شکست یا موفقیت دارد. در این میان، تعداد معیارها، پیچیدگی داده ها و پویایی محیط از جمله عواملی هستند که در دهه اخیر مساله تصمیم گیری در ورزش را با چالشی جدی روبرو کرده است. امروزه برای پیش بینی نتایج رویدادهای ورزشی، روش های علمی

- 1- Less Routine Situations
- 2- More Uncertain Situations
- 3- Medicine
- 4- College Success
- 5- Business Decision-Making
- 6- Weather Forecasting
- 7- Macroeconomics Prediction
- 8- political elections

نوینی برگزیده شده که با استفاده از این روش‌ها می‌توان با شناخت عوامل اثرگذار بر نتایج به دست آمده، نتایج رویدادها و رده‌بندی تیم‌های ورزشی را نیز پیش‌بینی کرد (۲۶). بر اساس مقاله‌های بسیاری مشخص شده که پیش‌بینی‌های ورزشی مبتنی بر داده و اطلاعات، با آنچه که به صورت تصادفی انجام می‌پذیرد (مانند پیش‌بینی‌های شانسی^۱) کاملاً متفاوت است (۱۵).
پیش‌بینی‌های ورزشی معمولاً از سه راه برحسب نوع پیش‌بینی انجام می‌پذیرد:
الف) پیش‌بینی‌هایی که نتیجه برگزاری یک مسابقه ویژه ما بین دو تیم را حدس می‌زند (۲۵)،
(۸، ۱۱).

ب) پیش‌بینی‌هایی که امتیازهای دو رقیب را گمانه زنی می‌کند (۳۲).

ج) پیش‌بینی‌هایی که برنده و فاتح چندین مسابقه را پیش‌بینی می‌کند همانند، تورنمنت‌ها^۲ (۴ و ۱۲) لیگ‌ها^۳ (۳۰) یا مسابقه‌های اسب‌دوانی^۴ (۶).

پیروزی‌های بین‌المللی بویژه پیروزی در بازی‌های آسیایی به شکل چشمگیری برای بسیاری از کشورهای آسیایی حائز اهمیت شده است. هم سیاستمداران و برنامه‌ریزان و هم اصحاب رسانه و مطبوعات مدال‌های کسب شده در مسابقه‌ها را به عنوان شاخصی از موفقیت بین‌المللی قلمداد می‌کنند علی‌رغم اعتراض و اظهارنظر جدی و صریح کمیته برگزاری بازی‌های آسیایی مبنی بر این است که جدول مدال‌ها نمی‌تواند نشانه شایستگی^۵ یک کشور بر کشور دیگر باشد. در نتیجه، موفقیت در ورزش‌هایی که در کانون توجه جهان^۶ هستند (ورزش نخبگان) به عنوان یک منبع بارز برای رسیدن به اهداف غیرورزشی رو به افزایش است. به خاطر بسیاری از مطالعات، دانش نحوه توسعه ورزش‌هایی که در کانون توجه جهانیان قرار دارند، افزایش یافت و به کمک این مطالعات، سیستم‌های این ورزش‌ها و فاکتورهای شکل

1- lottery

2- tournaments

3- leagues

4- horse races

5- Merit

6- Elite Sports

دهنده آنها هر چه بهتر شناسایی شدند (۲۸ و ۱۹، ۵) که این باعث افزایش تمایل و مداخله مستقیم دولت‌ها برای توسعه این ورزش‌ها از طریق سرمایه‌گذاری مالی قابل توجه شده است (۵). در اینجا پرسشی پیش می‌آید که چرا برخی از کشورها در مسابقه‌ها و رویدادهای بین‌المللی ورزشی نسبت به کشورهای دیگر موفق‌ترند؟ این موضوع به صورت آشکاری با کار سیاستمداران و برنامه‌ریزان است. کسانی که تمایل دارند پیروزی و جایگاه خود را در جدول بازی‌ها بهبود بخشند (۷). هرچند شمار کشورهای کمی که به منظور رقابت با سایر کشورها، هزینه‌های زیادی را صرف ورزش می‌کنند رو به افزایش است، اما مدارک و شواهد تعیین‌کننده عوامل اثرگذار در پیروزی ورزشی سطوح بین‌المللی بسیار اندک است (۲۲). چنانچه عوامل اساسی به گونه‌ای مناسب شناسایی شوند، مدیران و برنامه‌ریزان با صرف زمان و انرژی کمتر می‌توانند دستیابی به پیروزی ورزشی یک کشور را هدفمند سازند تا بتوان با هزینه بهینه، بهترین عملکرد را از کاروان ورزشی پیش بینی کرد.

در این پژوهش پیروزی کلی کشورها در بازی‌های آسیایی مورد تجزیه و تحلیل قرار خواهد گرفت نه پیروزی ورزشکاران به صورت فردی یا گروهی. برای رسیدن به این هدف، عوامل سطح کلان و متوسط (عوامل اقتصادی، سیاسی، اجتماعی و فرهنگی همانند؛ جمعیت^۱، مساحت^۲، نرخ تورم^۳، تولید ناخالص داخلی^۴، هزینه‌های آموزش و پرورش^۵ و...) که همپوشانی بالایی با هم دارند، مورد بررسی قرار خواهند گرفت و عوامل در سطح خرد حذف شده‌اند؛ علی‌رغم پذیرفتن این فرض که این سه سطح همواره با هم در تعامل هستند. البته در این میان، عواملی دیگر همچون؛ بخش خصوصی، فرهنگ ورزشی و سنت‌های ورزشی خاص هر کشور، رسانه‌های جمعی به عنوان ترویجگر علاقه به ورزش، شنوندگان و تماشاگران و... که در توسعه ورزش بسیار اثر گذارند، به علت اثرگذاری غیر مستقیم آنان بر توسعه ورزش در این پژوهش

1- Population

2- Area

3- Inflation Rate

4- Gross Domestic Production (GDP)

5- Education Expenditures

مورد بحث و بررسی قرار نگرفته‌اند.

به طور خلاصه نتایج پژوهش‌های مهمی که در زمینه پیش‌بینی‌های ورزشی صورت پذیرفته‌اند، در زیر آورده شده‌اند:

دروکو و همکاران (۲۰۰۲)^۱ با استفاده از نرم‌افزارهای کامپیوتری، عملکرد ورزشکاران نخبه پنج ماده دو و میدانی در ۱۰ سال آینده را پیش‌بینی کرده‌اند و پیش‌بینی میزان بهبود رکوردها تا سال ۲۰۱۰ میلادی از نظر آنها بین ۰/۲ درصد تا ۱۰/۳ درصد بوده است. آنها همچنین خاطر نشان کرده‌اند که ارزش‌های به دست آمده از این پیش‌بینی رایانه‌ای می‌تواند تحت عواملی همچون، تجهیزات ورزشی بهتر، تغذیه و تمرین بهتر و بویژه اثر مواد نیروزا در آینده ورزش قرار گیرد (۱۶). در این ارتباط لوکاس و لواگلیا (۲۰۰۵)^۲ نیز تأثیر شیوه‌های مدیریتی، مربیان، تسهیلات تیم‌ها و شانس را در پیش‌بینی رده‌بندی تیم‌ها مؤثر دانسته‌اند (۲۶).

بوایر و همکاران (۱۹۹۹)^۳ به ارزیابی این فرضیه پرداختند که آیا رتبه‌بندی‌های ورزشی، پیش‌بینی‌کننده‌های خوبی هستند؟ این پژوهش بر روی تیم‌های بسکتبال دانشگاهی آمریکا و تیم‌های حرفه‌ای تنیس انجام شد. از رگرسیون آماری با تفاوت‌های موجود در رنکینگ‌ها به عنوان عامل پیش‌بینی نتایج مسابقه‌های تیمی و انفرادی استفاده شد. یافته‌های پژوهش نشان داد که رتبه‌بندی‌ها به خودی خود عامل مؤثری در پیش‌بینی نتایج مسابقه‌ها هستند و درستی تخمین‌ها را بالا می‌برند (۹).

چیپه‌نی و برودر (۲۰۰۷)^۴ نتایج مسابقه‌های تنیس ویمبلدون سال ۲۰۰۵ را از طریق نام بازیکنان پیش‌بینی کردند. در این پژوهش بازیکنان آماتور تنیس (۷۹ نفر) و افراد عادی (۱۰۵ نفر) با بیان نام بازیکنانی که می‌شناختند، نتایج مسابقه‌ها را پیش‌بینی کردند. پیش‌بینی‌های انجام شده بر اساس شناسایی رتبه‌ها (که در میان همه شرکت‌کنندگان توزیع و انجام شد) نشان داد که پیش‌بینی بالغ بر ۷۰ درصد همه نتایج به درستی انجام پذیرفته است (۳۱).

1- Dereveco P. *et al.*

2- Lovalgia, M. J., & Lucas, J. W.

3- Boulier, B. L., & Stekler, H.

4- Scheibehenne, B., Broder, A.

آیر و شاردا (۲۰۰۹)^۱ شبکه‌های عصبی را برای پیش بینی عملکرد آینده بازیکنان (به صورت موردی کریکت) بر اساس عملکرد گذشته آنان از سال ۱۹۸۵ تا فصل ۲۰۰۷-۲۰۰۶ مورد استفاده قرار دادند. نتایج پژوهش نشان داد که شبکه‌های عصبی می‌توانند در واقع، پشتیبانی تصمیم‌گیری ارزشمندی را در فرایند انتخاب بازیکنان فراهم سازد (۳۱).

بویلر و همکاران (۲۰۰۳)^۲ طی پژوهشی نتایج مسابقه‌های لیگ ملی فوتبال^۲ را مورد پیش بینی قرار دادند. این مقاله نتایج نهایی را به عنوان ابزارهای پیش بینی در فصل‌های ۲۰۰۰-۱۹۹۴ در نظر می‌گیرد. این ارزشیابی، مقایسه‌ای از پیش بینی‌های حاصل از رگرسیون را بر اساس نتایج منتشر شده در نیویورک تایمز^۳ با مدل‌های محلی، شرط‌بندی‌ها و آرای ویراستاران ورزشی نیویورک تایمز ارائه می‌دهد. نتایج نشان داد که شرط‌بندان پس از پیش بینی‌های انجام شده بر اساس رگرسیون، بهترین پیش بینی کنندگان مسابقه‌های ورزشی بودند. همچنین نتایج نشان داد که آرای ویراستاران ورزشی نیویورک تایمز از کل پیش بینی‌ها (حتی پیش بینی‌های محلی) ارزش و جایگاه پایین تری دارد (۸).

مین و همکاران (۲۰۰۸)^۴ چهارچوبی مرکب برای پیش بینی نتایج مسابقه‌های ورزشی (به صورت موردی فوتبال) ارائه کردند و نشان دادند که پیش بینی‌های سیستم خبره نتایج فوتبال (که در این تحقیق مورد استفاده قرار گرفت) پیش بینی‌های منطقی و یکنواختی هستند (۲۷).

کاندن و همکاران (۱۹۹۹)^۵ به پیش بینی پیروزی کشورها در بازی‌های المپیک سال ۱۹۹۶ از طریق شبکه‌های عصبی پرداختند. در این پژوهش، اطلاعات ۲۷۱ رشته ورزشی از ۱۹۵ کشور بر روی ۱۷ متغیر مستقل جمع‌آوری شد و در بررسی خود به این نتیجه رسیدند که برای پیش بینی موفقیت کشورها در بازی‌های المپیک الگوی شبکه‌های عصبی از الگوی رگرسیون ابزار

1- Iyer, S. R., & Sharda, R.

2- National Football League game (NFL)

3- New York Times

4- Min, B., et al.

5- Condon, E. et al.

مناسب تری هستند (۱۴).

اندرسون و همکاران (۲۰۰۵)^۱ در پیش‌بینی نتایج جام جهانی ۲۰۰۲ متوجه شدند که افراد آشنا به فوتبال نسبت به افراد ناآشنا با اعتماد به نفس و دقت بیشتری نتایج را پیش‌بینی می‌کنند (۴).

سونگ و همکاران^۲ (۲۰۰۷) پژوهشی با عنوان مقایسه دقت پیش‌بینی به کمک مدل و داوری در بازی‌های فوتبال آمریکا انجام دادند. آنها با بهره‌گرفتن از ۳۱ الگوی آماری و به کمک ۷۰ متخصص (که در پیشگویی برندگان ۴۹۶ بازی NFL در سال ۲۰۰۰ تا ۲۰۰۱ موفق بودند) نتایج بازی‌ها را پیش‌بینی کردند. نتایج نشان داد که تفاوت در دقت پیش‌بینی سیستم‌های آماری و متخصصان برای پیشگویی برندگان بازی‌ها معنی‌دار نبود. تغییرات در میزان پیروزی پیش‌بینی در متخصصان نسبت به سیستم‌های آماری بالاتر بود؛ با این حال، داشتن اطلاعات بسیار، پیش‌بینی را بهبود نمی‌بخشد. نه تنها متخصصان بلکه، سیستم‌های آماری نیز نتوانستند به شکل مفیدی عمل پیشگویی را درست انجام دهند (۳۳).

محمدی (۱۳۸۹) در پژوهشی با عنوان "ارائه مدل ریاضی برای رتبه‌بندی کشورهای شرکت‌کننده در بازی‌های آسیایی ۲۰۰۶ قطر" به این نتیجه رسید که رویکرد جایگزینی رتبه‌بندی منصفانه‌تر از رویکرد رایج است (۱).

روش پژوهش

مراحل پژوهش شامل سه مرحله است:

الف) از طریق مطالعه پیشینه پژوهش، متغیرهای حیطه‌های اقتصادی، سیاسی، اجتماعی و فرهنگی شناسایی شدند. سپس متغیرهای هر حیطه در اختیار خبرگان قرار گرفت تا از حیث

1- Andersson, P. et al.

2- Song, ch., et al.

اهمیت، رتبه بندی شوند. پس از آن، هر متغیر با توجه به رتبه کسب شده بر اساس روش دلفی امتیازی را به خود اختصاص داد که سرانجام ۱۱ متغیر نخست به عنوان متغیرهای اثرگذار جهت پیش بینی رتبه کشورهای شرکت کننده در بازی‌های آسیایی برگزیده شدند.

ب) در گام دوم، اطلاعات متغیرهای انتخاب شده برای کشورهای شرکت کننده در بازی‌های آسیایی از سال ۱۹۷۴ تا ۲۰۱۰ جمع آوری شد. در ضمن اطلاعات کشورهای ازبکستان، قزاقستان، تاجیکستان و قرقیزستان از سال ۱۹۹۴ تا ۲۰۱۰ آورده شده و کشورهای افغانستان، کره شمالی و عراق را به علت در دسترس نبودن اطلاعات از طرح پژوهش خارج کردیم.

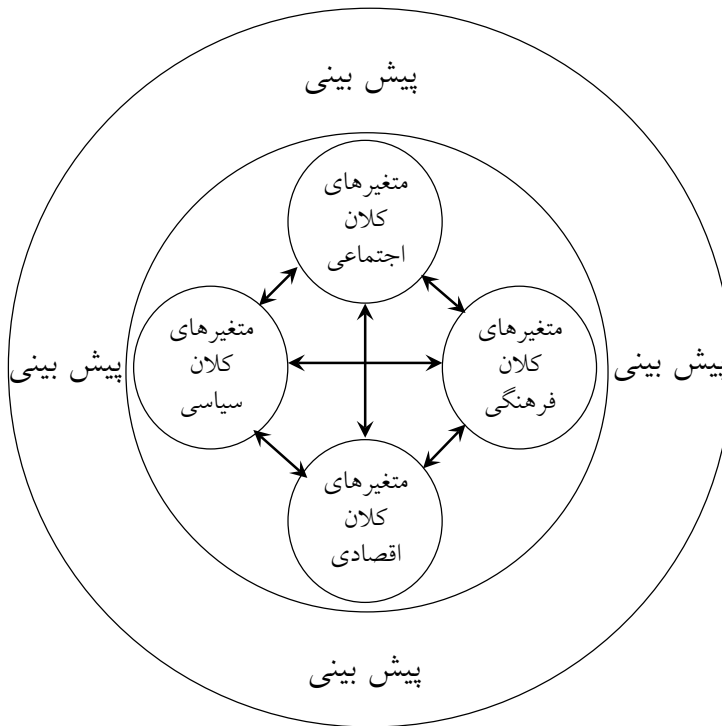
ج) رتبه بندی کشورهای شرکت کننده در بازی‌های آسیایی به ۲ شیوه صورت گرفته است::

۱- رتبه بندی بر اساس تعداد مدال‌های طلای کسب شده هر کشور

۲- رتبه بندی بر اساس مجموع مدال‌های کسب شده هر کشور

در طرح پژوهش، پیش بینی کشورها بر اساس تعداد مدال‌های طلای کسب شده هر کشور صورت پذیرفته و تعداد کشورها در دوره‌های مختلف برگزاری بازی‌های آسیایی متغیر بوده است.

در این پژوهش متغیر مستقل شامل؛ متغیرهای کلان اقتصادی، سیاسی، اجتماعی و فرهنگی است و متغیر وابسته، پیروزی کشورهای شرکت کننده در بازی‌های آسیایی بر اساس تعداد مدال‌های طلای اخذ شده، خواهد بود.



نمودار ۱ مدل مفهومی پیش‌بینی پیروزی کشورهای شرکت‌کننده در بازی‌های آسیایی

در این پژوهش از مدل درختی استفاده شده است. مدل درختی، تکنیک و روشی است که با ساختاری از داده‌ها و یا المان‌های تکه ای خطی از مجموعه ای خاص مرتبط است. مدل درختی، مجموعه دستورهای درختی مرسوم دارد که در فرایند ساخت دستورهای به جای استفاده از عناوین مجزا، از توابعی خطی در برگها استفاده می‌کند نخستین اجرا و الگوی مدل درختی را کوئین‌لن^۱ (۱۹۹۲) با عنوان M5 تعریف کرد. وانگ و ویتن^۲ این الگو را با عنوان جدیدی به نام M5 در سال ۱۹۹۷ بازسازی و ارتقا دادند (۳۴). همانند دستورهای درختی مرسوم، M5^۱ نیز درخت را از راه تقسیم کردن داده‌ها بر اساس مقادیر خصوصیات پیش بینی شده شکل

1- Quinlan, J. R.

2- Wang, Y., Witten, I. H.

می‌دهد (۲۳).

M5 یک سیستم جدید از مدل‌هایی است که مقادیر را پیش بینی می‌کند، و بر اساس مدل درختی ساخته می‌شود. درخت‌های ساخته شده توسط M5 می‌توانند مدل‌های خطی چند متغیره داشته باشند که این مدل‌های خطی مشابه توابع تکه ای خطی (گسسته) هستند. M5 راندمان بالایی دارد و می‌تواند پروژه‌هایی با تعداد زیادی داده را نیز انجام دهد (بالای ۱۰۰ متغیر)، این ویژگی باعث می‌شود که M5 و CART^۱ از روش MARS^۲ جدا شوند؛ چرا که روش MARS قابلیت اجرای آن به ازای تعداد محدودی متغیر است. مزیت M5 بر CART این است که مدل درختی به طور کلی کوچکتر از رگرسیون درختی است و ثابت شده که در بررسی نتایج و پیش بینی نیز دقت به مراتب بیشتری دارد.

ساخت الگوی درختی

چنانچه مجموعه ای با T داده داشته باشیم، به گونه ای که مشخصه هر مورد بر اساس مقدار ویژگی‌های ثابت آن مورد به صورت گسسته یا عددی تعیین شود و هدف ساختن مدلی جهت مرتبط ساختن مقدار مورد نظر هر داده به ویژگی‌های دیگرش باشد. ارزش مدل، بر اساس دقت پیش بینی داده‌ها با مقادیر نهان محک زده می‌شود.

مدل‌های درختی بر روش‌های حل و تقسیم استوار است. مجموعه T یا با ۱ برگ مربوط می‌شود و یا با تعدادی آزمون که به منظور تقسیم کردن مجموعه T به زیرمجموعه‌ها (که متناظر با نتایج آزمون‌ها هستند) انتخاب می‌شوند و فرایند مشابهی را به صورت بازگشتی روی زیرمجموعه‌ها به کار می‌برد. این فرایند تقسیم اغلب به ایجاد ساختاری با جزئیات زیاد می‌انجامد که حتماً باید هرس شود، که این هرس کردن به طور مثال با جایگزین کردن ۱ گره به جای ۱ زیر درخت صورت می‌پذیرد.

1- Classification and Regression Trees

2- Multivariate Adaptive Regression splines

نخستین گام برای ایجاد یک مدل درختی، محاسبه مقدار انحراف استاندارد موردنظر داده‌ها در T است. اگر T شامل تعداد خیلی کمی داده باشد و یا مقادیر این داده‌ها خیلی کم باهم متفاوت باشند (به عبارتی نزدیک به هم باشند) ناگزیریم که T را به نتایج آزمون تقسیم کنیم و تقسیم بندی تنها شامل نتایج آزمون باشد. هرآزمون بالقوه از راه مشخص کردن همه زیرمجموعه‌های مرتبط با نتیجه آن قابل ارزیابی است. اگر T_i معرف زیرمجموعه‌ای از حالت‌هایی باشد که دارای آیین نتیجه آزمون بالقوه است و اگر انحراف استاندارد $Sd(T_i)$ مقادیر مورد نظر از همه حالت‌ها را به عنوان یک پارامتر خطا در نظر بگیریم میزان خطا عبارت است از:

$$\Delta error = sd(T) - \sum_i \frac{|T_i|}{|T|} \times sd(T_i)$$

پس از بررسی تمام حالات ممکن آزمون، $M5$ حالتی را انتخاب کنیم که بیشترین کاهش خطای قابل قبول را دارد. این ویژگی برجسته $M5$ پس از اینکه درخت اولیه رشد یابد، وارد عمل می‌شود.

تخمین خطا: $M5$ اغلب دقت مدل را به کمک حالت‌هایی که داده‌ها نهان هستند، ارزیابی می‌کند. باقیمانده^۲ یک مدل در واقع از تفاضل کامل مقدار واقعی و مقدار پیش‌بینی شده از جانب مدل به دست می‌آید. برای ارزیابی خطای یک مدل که نشأت گرفته از مجموعه‌ای از داده‌ها است، $M5$ ابتدا متوسط باقیمانده مدل براساس داده‌ها را تعیین می‌کند، این عموماً خطایی را که بر پایه داده‌های نهان است، ناچیز می‌پندارد، بنابراین $M5$ مقدار خطا را در عبارت $(n+v)/(n-v)$ ضرب می‌کند که در آن n تعداد داده‌ها و v تعداد پارامترهای موجود در مدل است. با توجه به تعداد زیاد پارامترهایی که نشأت گرفته از تعداد کم داده‌ها هستند، نتیجه به افزایش خطای تخمین زده شده مدل می‌انجامد.

1 - Error Estimates

2- Residual

مدل‌های خطی^۱: در ساختار مدل درختی برای ایجاد برگ‌ها از توابع خطی استفاده می‌شود. یک مدل خطی چند متغیره با داده‌هایی در هر گره، از مدل درختی ساخته شده که از فنون و روش‌های رگرسیون استاندارد استفاده کرده است.

ساده سازی مدل‌های خطی: پس از اینکه مدل خطی به روش بالا به دست آمد، سپس می‌توان مدل را با حذف پارامترهایی به منظور به حداقل رساندن خطا، ساده سازی کرد. اگرچه حذف پارامترها عموماً خود باعث می‌شود که متوسط باقیمانده افزایش یابد، این مسأله همچنین باعث می‌شود که مقدار فاکتور ضرب کننده که در بالا آمد نیز کاهش یابد و در نتیجه موجب کاهش خطای تخمین زده شده می‌شود. M5 همچنین سعی در حذف کردن متغیرهایی دارد که تاثیر آنها بر مدل کمتر است. در بعضی مواقع M5 تمام متغیرها را حذف می‌کند و تنها ۱ ثابت را باقی می‌گذارد.

هرس کردن^۲: هر گره از مدل درختی از پایین بررسی می‌شود. M5 به عنوان مدل نهایی این گره، از میان مدل خطی ساده شده بالا و مدل زیر درخت، یکی را بر اساس اینکه کدام یک خطای تخمینی کمتری دارد انتخاب می‌کند. در واقع عملیات هرس کردن مادامی که خطای تخمینی قابل قبول حاصل شود، ادامه دارد. این پروسه از ساخت درخت منجر به وقوع گسستگی بسیار زیاد مدل‌های خطی مجاور در برگ‌های درخت هرس شده می‌شود. برای رفع این مشکل از فرایندی موسوم به "هموارسازی" استفاده می‌کنند که می‌توان این تفاضل را جبران کند. هموارسازی^۳: پریگیون^۴ مشاهده کرد که می‌توان دقت پیشگویی را در مدل‌های درختی با بهره گیری از "پروسه هموارسازی" ارتقا داد. هنگامی که مدل درختی مقدار یک داده را پیش بینی می‌کند، مقداری که مدل ارائه می‌دهد، در برگ مناسبی به منظور منعکس کردن

1- Liner Model

2- Pruning

3- Smoothing

4- Pregibon

مقادیر پیش‌بینی شده در گره‌هایی که در امتداد مسیری از ریشه تا آن برگ وجود دارند، تنظیم می‌شود. فرم هموارسازی که M5 استفاده می‌کند با فرمی که پرگیبون آن را ارتقا داد، فرق دارد؛ اما انگیزه هر دو یکسان است. در واقع، در فرایند هموارسازی، پیش‌بینی مقداری که از مدل برگ استفاده شده، تخمین زده می‌شود. سپس مقدار حاصله در امتداد پایین درخت تا ریشه مورد بررسی قرار می‌گیرد و هموارسازی در هر گره از طریق ترکیب با مقدار پیش‌بینی شده مدل خطی آن گره قابل اجراست.

• مقدار پیش‌بینی شده در برگ همان مقداری است که توسط مدل آن برگ محاسبه می‌شود.

• اگر شاخه S_i از زیر درخت S و n_i تعداد موارد آموزش در S_i و $PV(S_i)$ مقدار پیش‌بینی شده در S_i و $M(S)$ مقدار داده شده مدل در S و k ثابت هموارسازی باشد، مقدار پیش‌بینی شده

ای که به S برمی‌گردد برابر است با:

$$PV(S) = \frac{n_i \times PV(S_i) + k \times M(S)}{n_i + k}$$

بیشترین تاثیر هموارسازی مربوط به مواردی است که مدل مقادیر بسیار متفاوتی را پیش‌بینی می‌کند و یا در شرایطی که مدل از تعداد کمی متغیر ساخته شده است.

هرگاه با حجم گسترده‌ای از داده مواجه هستیم، مدل درختی مانند رگرسیون درختی کارایی خوبی دارد. از مزیت‌های دیگر مدل درختی، دقت بالا در پیشگویی و توانایی این مدل در استفاده از خطی‌سازی محلی داده هاست. از دیگر تفاوت‌های دیگر رگرسیون درختی و مدل درختی این است که رگرسیون در پروسه پیش‌بینی هرگز توانایی ارائه مقادیر پیش‌بینی در خارج از دامنه مجموعه دیتاهایش را ندارند، درحالی که مدل درختی این قابلیت را دارد (۲۹).

نتایج پژوهش

با توجه به تجزیه و تحلیل داده‌ها از طریق مدل درختی پس از هرس نهایی، ۱۰ قانون جهت پیش بینی رتبه کشورهای شرکت کننده در بازی‌های آسیایی ۲۰۱۰ مبتنی بر متغیرهای کلان طراحی شد.

به طور مثال، قانون نخست این گونه تعریف می‌شود:

$$\begin{aligned} \text{قانون نخست: مدال طلا} &= \text{جمعیت شهری} (-0,1092) + \text{هزینه آموزش و پرورش} (-0,541) \\ &+ \text{ساختار سنی} (-0,297) + \text{تولید واقعی ناخالص داخلی} (-0,016) + \text{بیکاری} (0,1796) + \\ &\text{جمعیت} (-0,3529) + \text{نرخ تورم} (-0,0056) + \text{امید به زندگی} (0,4474) + \text{تراز بازرگانی} \\ &+ (0,0033) + 15,081 \end{aligned}$$

در جدول ۱ تمامی قوانین به تفکیک آورده شده است.

جدول ۱. قوانین ۱۰ گانه طراحی شده جهت پیش‌بینی رتبه کشورهای شرکت‌کننده در بازی‌های آسیایی ۲۰۱۰*

| | قانون ۱ | قانون ۲ | قانون ۳ | قانون ۴ | قانون ۵ | قانون ۶ | قانون ۷ | قانون ۸ | قانون ۹ | قانون ۱۰ |
|--|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| جمعیت شهری ^۱ | -۰.۱۰۹۲ | -۰.۰۸۹۳ | -۰.۱۱۲۷ | -۰.۱۱۴۱ | -۰.۰۵۴ | -۰.۰۵۴ | -۰.۰۵۴ | ۰.۲۸۸ | ۰.۰۷۶۵ | -۰.۰۸۸۳ |
| هزینه آموزش و پرورش | -۰.۵۴۱ | -۰.۴۸۱۴ | -۰.۴۸۱۴ | -۰.۴۸۱۴ | -۲.۲۶۲۵ | -۱.۶۹۰۸ | -۱.۶۹۰۸ | -۰.۴۸۱۴ | -۰.۹۸۹۶ | ۰.۰۶۱۷ |
| ساختار سنی ^۲ | -۰.۲۹۷ | ۰.۵۵۱۷ | ۰.۵۵۱۷ | ۰.۵۵۱۷ | ۰.۵۶۶۳ | ۰.۵۶۶۳ | ۰.۵۶۶۳ | ۰.۳۴۴ | -۰.۰۳۰۶ | -۰.۰۳۰۶ |
| تولید واقعی ناخالص داخلی ^۳ | -۰.۰۱۶ | -۰.۵۲۵۶ | -۰.۵۳۲۲ | -۰.۵۳۲۲ | ۰.۲۶۷۳ | ۰.۲۶۷۳ | ۰.۲۶۷۳ | ۰.۰۱۷۱ | ۰.۰۱۷۱ | ۰.۰۱۷۱ |
| سرمایه تولید ناخالص داخلی ^۴ | . | -۰.۰۰۳۶ | -۰.۰۰۳۶ | -۰.۰۰۳۶ | -۰.۰۰۵۵ | -۰.۰۰۵۵ | -۰.۰۰۵۵ | -۰.۰۰۳۱ | -۰.۰۰۳۱ | -۰.۰۰۳۱ |
| بیکاری ^۵ | ۰.۱۷۹۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ | ۰.۰۳۴۶ |
| جمعیت | -۰.۵۳۲۹ | -۰.۰۰۹۱ | -۰.۰۰۹۱ | -۰.۰۰۹۱ | -۰.۰۰۹۱ | -۰.۰۰۹۱ | -۰.۰۰۹۱ | -۰.۰۱۴۸ | -۰.۰۰۵۷ | -۰.۰۰۴۵ |
| نرخ تورم | -۰.۰۰۵۶ | ۰.۰۷۵۶ | ۰.۰۲۲۵ | ۰.۰۲۲۵ | ۰.۲۲۵ | ۰.۲۲۵ | ۰.۲۲۵ | -۰.۰۰۰۸ | -۰.۰۰۰۸ | -۰.۰۰۰۸ |
| تعادل حساب جاری ^۶ | --- | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۵۴۴ | ۰.۹۵۵ | --- |
| امید به زندگی ^۷ | ۰.۴۴۷۴ | ۰.۱۷۶۳ | ۰.۱۷۶۳ | ۰.۱۷۶۳ | ۰.۴۹۲۱ | ۰.۴۹۲۱ | ۰.۴۷۵۷ | ۰.۲۳۶۹ | ۰.۳۰۸ | ۰.۹۲۸ |
| تراز بازرگانی ^۸ | ۰.۰۰۳۳ | -۰.۴۱۳ | -۰.۴۱۳ | -۰.۴۱۳ | -۰.۶۹۱ | -۰.۶۹۱ | -۰.۶۹۳ | -۰.۰۴۲۵ | -۰.۰۰۳۶ | -۰.۰۰۳۶ |
| | ۱۵.۰۸۱ | -۱۸.۳۰۹ | -۱۶.۶۷۶ | -۱۶.۷۰۱ | -۳۲.۲۰۶ | -۲۵.۸۲۹ | -۳۴.۴۲۵ | -۱۵.۹۸۴ | -۰.۵۳۸۲ | ۶.۸۵۹۱ |

^۱ - urban population
^۲ - Age Structure
^۳ - GDP Real Growth Rate
^۴ - GDP Per Capita
^۵ - Unemployment Rate
^۶ - current account balance
^۷ - life expectancy at birth
^۸ - Merchandise Trade

در جدول ۲ اطلاعات همه متغیرهای تمامی کلیه کشورهای مدال آور در بازی‌های آسیایی ۲۰۱۰ آورده شده است. در ۲ ستون نهایی، رتبه‌های واقعی و پیش بینی شده کشورها گزارش شده است. مدل، از ۲۸ کشوری که موفق به کسب مدال طلا در بازی‌های آسیایی شدند، رتبه پیش بینی شده ۱۸ (۶۴,۲۸٪) کشور را حداکثر با ۳ اختلاف، ۶ (۲۱,۴۳٪) کشور را حداکثر بین ۴ تا ۷ اختلاف و ۴ (۱۴,۲۹٪) کشور را با بیش از ۷ اختلاف به نسبت رتبه واقعی آنها پیش بینی کرد.

* جمعیت شهری، تورم و بیکاری (%)، هزینه آموزش و پرورش (% of GNI)، جمعیت (میلیون نفر)، ساختار سنی (درصد ۱۵ تا ۶۴ سال)، تولید واقعی ناخالص داخلی (درصد رشد سالانه)، سرانه تولید ناخالص داخلی (بر حسب دلار)، امید به زندگی (سال)، تعادل حساب جاری و تراز بازرگانی (% of GDP)

جدول ۲. اطلاعات ۱۱ متغیر کلان کلیه کشورهای مدال آور در بازی‌های آسیایی ۲۰۱۰ به همراه رتبه واقعی و پیش‌بینی شده آنها

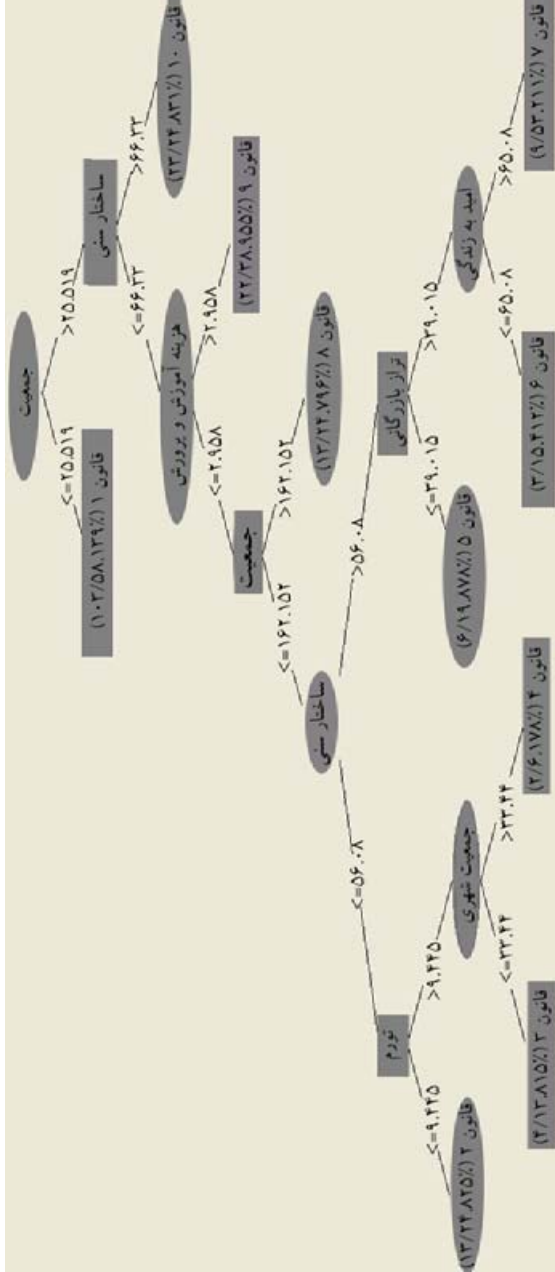
| رتبه پیش‌بینی شده | رتبه واقعی | تراز پارکتلی | امید به زندگی | کسر پاری مثال | روز پریم | جمعیت | بهره‌وری | ناتخالص داخلی سرانه تولید | ناتخالص داخلی تولید واقعی | شاخص سنی | هزینه آموزش و پرورش | جمعیت شهری |
|-------------------|------------|--------------|---------------|---------------|----------|---------|----------|---------------------------|---------------------------|----------|---------------------|------------|
| ۲ | ۱ | ۵۹.۲ | ۷۴.۶۸ | ۶.۱۳۹ | ۳.۱۲ | ۱۳۴۱.۴۱ | ۴.۱ | ۲۲۰۲ | ۸۵۰۴ | ۷۲.۱ | ۱.۹ | ۴۴ |
| ۳ | ۲ | ۹۲.۲۷ | ۷۹.۰۵ | ۱.۶۰۷ | ۲.۹ | ۴۸.۹۱ | ۳.۲ | ۲۰۲۳۹ | -۰.۹۸۷ | ۷۲.۳ | ۴.۲ | ۸۱.۷ |
| ۱ | ۳ | ۳۱.۴۵ | ۸۲.۳۵ | ۲.۸۴ | -۱.۴۰۷ | ۱۲۷.۳۷۱ | ۵.۱۴۵ | ۳۷۵۵۵ | -۵.۲۳۹ | ۶۴.۳ | ۳.۷ | ۶۶.۶ |
| ۵ | ۴ | ۶۴.۷۳ | ۷۰.۰۶ | ۲.۳ | ۸.۵ | ۷۵.۳۵ | ۱۲.۵۷ | ۳۴۱۱ | ۱.۴۴۴ | ۷۲.۹ | ۴.۸ | ۶۹ |
| ۱۲ | ۵ | ۸۱.۷۴ | ۶۸.۸۱ | -۰.۷۱۵ | ۷.۳۰۳ | ۱۵.۵۸۴ | ۷.۸ | ۶۳۴۶ | ۱.۴۶۵ | ۷۰.۲ | ۴.۴۱ | ۵۸.۲ |
| ۶ | ۶ | ۴۰.۶ | ۶۶.۸ | -۲.۱۷۲ | ۱۳.۱۶۲ | ۱۲۱۵.۹۴ | ۱۰.۷ | ۸۷۱ | ۵.۳۵۵ | ۶۴.۳ | ۳.۲ | ۳۹.۸ |
| ۸ | ۷ | ۵۵.۹۲ | ۷۲.۵۱ | ۵.۰۵۵ | ۹.۱۵۱ | ۲۸.۲۴۶ | -۰.۲ | ۷۶۴ | ۶.۹۷۸ | ۶۷ | ۹.۴ | ۳۶.۹ |
| ۷ | ۸ | ۱۲۰.۸۶ | ۳۶.۶ | ۲.۴۹۶ | ۳.۲۴۵ | ۶۷.۶۵۳ | ۱.۳۹ | ۳۵۷۷ | -۲.۴۲۶ | ۴۹ | ۷.۰ | ۳۲.۷ |
| ۱۱ | ۹ | ۱۶۰.۷۱ | ۷۴.۱۲ | ۱۵.۳۹۹ | ۲ | ۲۸.۲۳۳ | ۳.۵ | ۶۲.۴۷ | -۳.۶۲۱ | ۶۳.۶ | ۴.۵ | ۷۱.۲ |
| ۲۲ | ۱۰ | ۳۵۳.۳۹ | ۸۲.۳۴ | ۱۲.۰۵۲ | ۲ | ۷.۱۲۲ | ۴.۳۸۷ | ۳۹۲۷۳ | -۳.۶۳۳ | ۷۴.۶ | ۳.۲ | ۱۰۰ |
| ۴ | ۱۱ | ۹۴.۰۳ | ۷۳.۱۲ | ۹.۱ | ۵.۲ | ۲۶.۱۰۶ | ۱۰.۲۷۶ | ۱۵۸۸۶ | -۰.۸ | ۵۹.۵ | ۵.۷ | ۸۲.۳ |
| ۲۵ | ۱۲ | ۱۴۳.۲۴ | ۷۵.۹۱ | ۵.۴۸۶ | ۲.۳۹۳ | ۱.۰۶ | ۱۵ | ۱۴۹۰۸ | ۳.۰۴ | ۷۰.۱ | ۲.۹ | ۸۳.۹ |
| ۱۰ | ۱۳ | ۵۱.۹۸ | ۷۰.۷۹ | ۱.۴۱۴ | ۴.۷۲۴ | ۱۳۴.۵۵۷ | ۷.۵ | ۱۷۵۳ | ۳.۹۹ | ۶۶ | ۳.۵ | ۵۲.۶ |
| ۱۸ | ۱۴ | ۳۶۱.۶۲ | ۸۰.۳۴ | ۲۱.۹۶۶ | ۲.۰۹۷ | ۴.۸۲۲ | ۲.۰۷۸ | ۳۳۱۷۴ | -۲.۳۲۸ | ۷۶.۷ | ۳.۲ | ۱۰۰ |
| ۱۳ | ۱۵ | ۹۰.۱۲ | ۷۵.۹۴ | ۲۵.۱۱۱ | ۱.۰۳۳ | ۱.۳۲۵ | -۰.۵ | ۳۴۴۴۹ | ۱۱.۳۶۷ | ۶۸.۸ | ۳.۲ | ۹۵.۷ |
| ۱۹ | ۱۶ | ۷۹.۹۲ | ۷۷.۹۷ | ۳۱.۶۲ | ۴.۶۶۳ | ۳.۶۰۶ | ۱.۶۳۹ | ۱۸۷۵۶ | -۱.۵۱ | ۷۰.۷ | ۳.۸ | ۹۸.۴ |
| ۱۴ | ۱۷ | ۳۸.۱۱ | ۶۶.۵۳ | -۳.۸۳۴ | ۱۱.۵ | ۱۶۹.۳۸ | ۶.۱۹۵ | ۸۶۶ | ۱.۹۶۶ | ۵۹.۱ | ۲.۹ | ۳۶.۶ |
| ۱۵ | ۱۸ | ۶۴.۸۲ | ۷۱.۸۳ | ۳.۴۹۲ | ۴.۹۵۳ | ۹۴.۰۱۳ | ۷.۲ | ۱۴۹۹ | -۰.۹۹۴ | ۶۰.۶ | ۲.۶ | ۶۵.۷ |
| ۲۰ | ۱۹ | ۱۱۷.۰۷ | ۶۶.۵۷ | -۶.۶۴۱ | ۷.۳۴۱ | ۲.۷۳۴ | ۳ | ۱۳۲۲ | -۰.۵ | ۶۷.۹ | ۵.۱ | ۵۷.۳ |
| ۲۷ | ۲۰ | ۱۱۶.۲ | ۷۲.۷۱ | -۸.۹۱۱ | ۵.۲۷۸ | ۶.۱۲۶ | ۱۲ | ۲۷۰۷ | ۳ | ۵۹.۴ | ۲ | ۷۸.۵ |
| ۱۷ | ۲۳ | ۴۹.۳۱ | ۶۶.۱۵ | ۲۰.۸۸ | ۷.۳۸۵ | ۱۶۷.۶۷۱ | ۵.۱ | ۴۲۶ | ۵.۴۱۹ | ۶۱.۴ | ۲.۴ | ۲۷.۶ |
| ۲۴ | ۲۵ | ۱۱۲.۶۶ | ۶۷.۲۷ | -۱۲.۴۷ | ۸.۴۲۸ | ۵.۴۳۱ | ۵.۵۷ | ۵۸۹ | ۱.۴۶۵ | ۶۴.۵ | ۶.۶ | ۳۶.۴ |
| ۹ | ۲۱ | ۱۵۸.۱۱ | ۷۴.۲۷ | -۶.۹۰۷ | ۱۲ | ۸۸.۲۵۷ | ۵ | ۷۷۸ | ۴.۶۰۶ | ۶۸.۳ | ۵.۲ | ۲۸.۲ |
| ۲۱ | ۲۲ | ۵۹.۰۹ | ۷۴.۲۳ | ۱.۱۳ | ۵.۰۲۷ | ۲۱.۷۶۲ | ۸.۵ | ۱۸۹۳ | ۳.۰۱۹ | ۵۹.۹ | ۴.۹ | ۵۴.۶ |
| ۲۶ | ۲۴ | ۹۱.۰۸ | ۶۶.۷۵ | -۷.۲۷ | ۷.۰۳۹ | ۶.۵۳۶ | ۲.۲ | ۴۶۵ | ۲ | ۶۲.۱ | ۳.۵ | ۲۶.۵ |
| ۲۸ | ۲۶ | ۴۴.۵۶ | ۶۴.۹۷ | -۱۰.۱۳ | ۶.۸۶۵ | ۶.۴۹۷ | ۲.۵ | ۶۵۶ | ۴.۵۸۴ | ۵۶.۲ | ۲.۳ | ۳۲ |
| ۲۳ | ۲۷ | ۷۲.۴۷ | ۷۲.۰۵ | -۱۲.۷۹ | ۵.۰۰۳ | ۳.۹۰۸ | ۹.۲ | ۶۱۶۳ | ۷ | ۶۷.۱ | ۲ | ۸۷.۱ |
| ۱۶ | ۲۸ | ۳۷.۰۲ | ۶۶.۳۹ | -۲.۷۷ | ۱۱.۷۶۲ | ۲۸.۲۸۵ | ۴.۶ | ۳۵۵ | ۳.۹۹۵ | ۵۹.۲ | ۳.۸ | ۱۷.۷ |

در جدول ۳ اطلاعات مربوط به همبستگی بین رتبه‌های واقعی و پیش بینی شده کشورها و میزان خطاها را پیش و پس از هرس آورده است. میزان همبستگی پس از هرس به مقدار ۷۵,۵٪ رسیده و میزان خطا به نسبت قابل توجهی در مقایسه با مرحله پیش از هرس، کاهش یافته است.

جدول ۳. اطلاعات مربوط به همبستگی و میزان خطاها بین رتبه‌های واقعی و پیش بینی شده کشورها

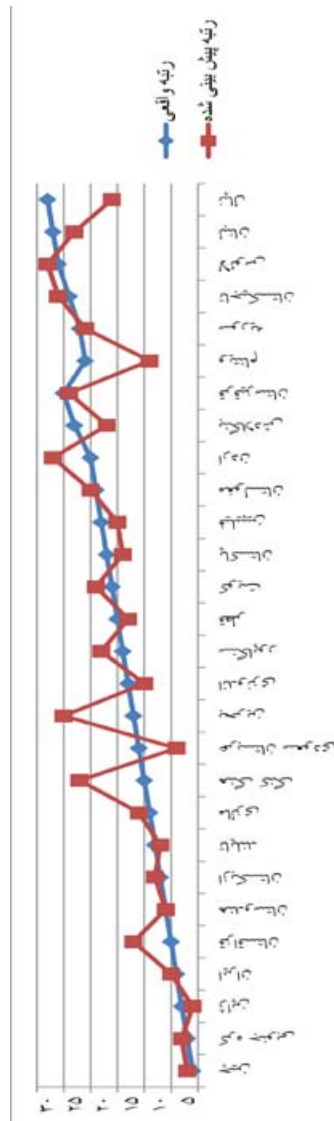
| | تعداد قانون | ضریب همبستگی | خطای مطلق میانگین | تعداد کشورها |
|------------|-------------|--------------|-------------------|--------------|
| M5 | ۸۲ | ۰.۷۳۴۵ | ۴.۹۰۱۷ | ۲۸ |
| M5- Pruned | ۱۰ | ۰.۷۵۵ | ۳.۵۳ | ۲۸ |

در شکل ۱ مدل درختی نشأت گرفته از متغیرهای کلان اقتصادی، سیاسی، اجتماعی و فرهنگی که منجر به موفقیت کشورهای آسیایی مدال آور در بازی‌های آسیایی می‌شود آورده شده است.



شکل ۱. مدل درختی پیش‌بینی کننده موفقیت کشورهای آسیایی بر اساس متغیرهای کلان اقتصادی، سیاسی، اجتماعی و فرهنگی*

در شکل ۲ رتبه‌های واقعی و پیش‌بینی شده کشورهای گزارش شده است. به طور مثال بین رتبه پیش‌بینی شده کشور ایران برای بازی‌های آسیایی ۲۰۱۰ با رتبه واقعی آن فقط ۱ اختلاف وجود دارد. این بیانگر آن است که کشور ایران در زمره کشورهایی قرار می‌گیرد که به طور کلی از مدل ارائه شده پیروی می‌کند و می‌توان از پیش‌بینی این مدل برای سال ۲۰۱۴ استفاده کرد.



شکل ۲. رتبه واقعی و پیش‌بینی شده کشورهای شرکت کننده در بازی‌های آسیایی ۲۰۱۰

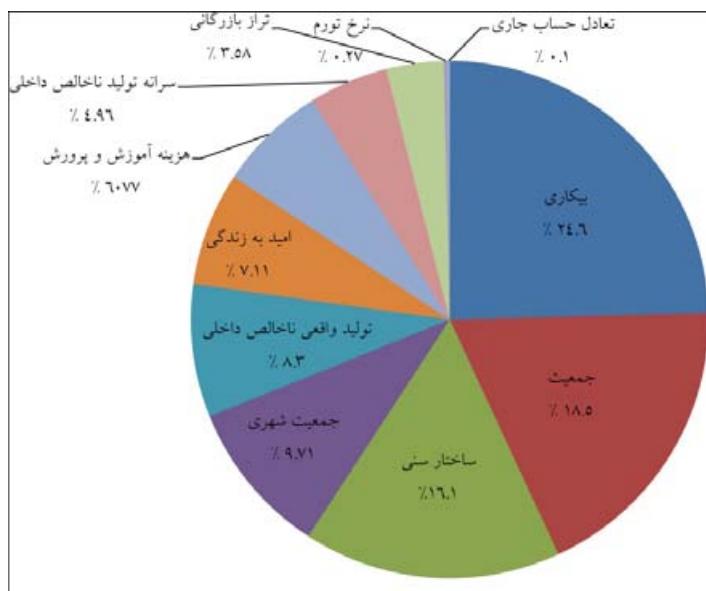
تعیین ضرائب متغیرهای اثرگذار در پیش‌بینی رتبه کشورهای

شرکت‌کننده در بازی‌های آسیایی ۲۰۱۰

بر اساس یافته‌های پژوهش پیش‌بینی تابعی است از متغیرهای ۱۱ گانه با ضرائبی که در جدول (۴-۱۰) و شکل (۴-۵) ارائه شده است.

جدول (۴-۱۰) ضرائب متغیرهای اثرگذار در پیش‌بینی رتبه کشورهای شرکت‌کننده در بازی‌های آسیایی ۲۰۱۰

| میزان اثرگذاری | نام متغیر پیش‌بین | میزان اثرگذاری | نام متغیر پیش‌بین |
|----------------|---------------------------|----------------|--------------------------|
| ۶.۷۷٪ | هزینه آموزش و پرورش | ۲۴.۶٪ | بیکاری |
| ۴.۹۶٪ | سرنانه تولید ناخالص داخلی | ۱۸.۵٪ | جمعیت |
| ۳.۵۸٪ | تراز بازرگانی | ۱۶.۱٪ | ساختار سنی |
| ۰.۲۷٪ | نرخ تورم | ۹.۷۱٪ | جمعیت شهری |
| ۰.۱٪ | تعالد حساب جاری | ۸.۳٪ | تولید واقعی ناخالص داخلی |
| | | ۷.۱۱٪ | امید به زندگی |



شکل (۴-۵) ضرائب متغیرهای اثرگذار در پیش‌بینی رتبه کشورهای شرکت‌کننده در بازی‌های آسیایی ۲۰۱۰

بحث و بررسی

روش‌های مبتنی بر درخت برای پیش‌بینی کردن، نسبت به سایر روش‌ها (مانند شبکه‌های عصبی) مناسب‌تر هستند؛ زیرا امکان تولید و بازیابی قوانین را دارند؛ از این رو می‌توانند علاوه بر کارایی مناسبشان در تخمین‌ها، دیدگاه مناسبی را نیز ایجاد کنند. در این کاربرد روش M5 گزینه مناسبی را برای تخمین داده‌های عددی و تولید رگرسیون‌های خطی مناسب با شرایط متغیرهای معرف کشورها ایجاد می‌کند. با هرس درخت از تولید قوانین اضافی جلوگیری می‌شود. علاوه بر آن، نتایج این آزمایش نشان می‌دهد که خطا هم کاهش می‌یابد. کشورهایی که درصد خطای تخمین آنها بالاست، یعنی دارای ساختاری هستند که به نوعی از فرم عمومی مدل پیروی نمی‌کنند. بر اساس مدل ساخته شده مبتنی بر متغیرها می‌توانیم برنامه ریزی‌هایی را به منظور تغییر، اصلاح و بهبود متغیرهای کلان یک کشور ایجاد کنیم. همچنین از روی این ساختار می‌توان ویژگی‌های کشورهای مناسب را بررسی کرد تا بتوان با برنامه ریزی برای تغییرات در این پارامترها موجب کسب نتایج مناسب برای کشورهای با وضعیت مشابه شد؛ زیرا کشورهای دارای ساختارهای کلان مشابه، می‌توانند الگوهای مناسب تری برای تغییر ارائه کنند. از طریق ساخت قانون‌ها، کشورهای با ساختار مشابه در کنار هم طبقه بندی می‌شوند و از آن طریق، امکان مقایسه و سرانجام برنامه ریزی به منظور کسب کرسی‌های بهتر فراهم خواهد شد.

مقدار ضریب همبستگی بین رتبه‌های پیش‌بینی شده و واقعی ۷۵٫۵٪ است و در شرایطی که ما با متغیرهای اقتصادی، سیاسی، اجتماعی و فرهنگی مواجه بودیم و با علم به وجود عامل انسانی در این متغیرها و غیر قابل پیش‌بینی بودن "انسان" این مقدار، مناسبی بوده است. بر این اساس می‌توان تغییرات جزئی در مدل را بررسی کرد. به طور مثال، رتبه پیش‌بینی شده کشور ژاپن بهتر از رتبه واقعی آن در سال ۲۰۱۰ است و این نشانگر ثبات کشور ژاپن در کل دوره‌های برگزاری شده بازی‌های آسیایی می‌باشد زیرا کشور چین اگرچه به نظر می‌رسد باید حائز رتبه نخست در دوره‌های بعد برگزاری بازی‌های آسیایی باشد، اما هرگز در کلیه دوره‌ها با

ثبات تر از کشور ژاپن نبوده است. و این ثبات کشور ژاپن در همه دوره‌های بازی‌های آسیایی با متغیرهای آن همخوانی دارد. البته از دست دادن جایگاه نخست کشور چین اگرچه دور از ذهن است، اما یافته‌های پژوهش با توجه به مدل درختی اهمیت و اثرگذاری ثبات یک کشور در گذر زمان را نشان می‌دهد و می‌توان نتیجه گرفت که تنها کشوری که توانایی گرفتن جایگاه کشور چین را در سال‌های آینده بر حسب متغیرهای کلان مورد استفاده شده دارد، کشور ژاپن است. با توجه به تعداد و کلان بودن متغیرهای مورد استفاده و اینکه امکان توسعه آنها در کوتاه مدت میسر نیست، سیاستمداران و برنامه‌ریزان کلان کشور می‌توانند برای بهبود این متغیرها راهکارهای کوتاه مدت، میان مدت و بلند مدت را طراحی کنند. بنابراین مهمترین و اثرگذارترین متغیری که در آغاز بر روی آن باید متمرکز شد، « درصد بیکاری » افراد جامعه است. بیکاری سال ۲۰۱۰ در کشورهای چین، کره جنوبی و ژاپن به ترتیب ۴،۱٪، ۳،۳٪ و ۵،۱٪ و در کشور ما ۱۲،۵۷٪ است. میانگین درصد بیکاری نیز از سال ۱۹۷۴ تا سال ۲۰۱۰ در کشورهای چین، کره جنوبی و ژاپن و ایران به ترتیب ۳،۰۴٪، ۳،۷۶٪، ۳،۱۷٪ و ۱۱،۳۹٪ بوده است. بنابراین وضعیت کشور ایران در مقابل ۳ کشور برتر آسیا بر اساس اثرگذارترین متغیر پیش‌بینی‌کننده بازی‌های آسیایی در بازه زمانی ۳۶ ساله که پژوهش انجام گرفت، بسیار ضعیف است. بنابراین با توجه به اطلاعات موجود امکان دستیابی به جایگاه بالاتر نه تنها برای کشور ایران فراهم نیست، بلکه بیم از دست دادن رتبه چهارمی هم می‌رود.

بنابراین ما تلاش کردیم تا الگویی را ایجاد کنیم تا از این طریق:

برای مدیران و برنامه‌ریزان این امکان فراهم شود تا برای توسعه ورزش و کسب کرسی‌های بین‌المللی با توجه به امکانات و منابع بالقوه کشور و مقایسه با کشورهای دیگر سیاست‌های مناسب اتخاذ شود؛ به گونه‌ای که اهداف بلندمدت، میان‌مدت و کوتاه‌مدت خود را بر اساس فاکتورهای اقتصادی، سیاسی، اجتماعی و فرهنگی تعیین کنند.

به ورزشکاران می‌تواند این فرصت را بدهد تا از طریق مقایسه خود با ورزشکاران کشورهای

دیگر جایگاه خود را شناسایی کرده و برنامه‌های تمرینی لازم را به منظور کسب رکوردهای بهتر بر اساس الگوهای استاندارد طراحی کنند.

شرایطی فراهم می‌کند تا انتظارات و توقعات تماشاگران، صاحبان رسانه و مطبوعات، منتقدان، کارشناسان یا تجزیه و تحلیلگران ورزشی و... واقع‌بینانه و عقلانی باشد و با توجه به پتانسیل موجود و همچنین آگاهی از توان کشورهای رقیب از توقع‌های احساسی دست بردارند تا از فشار و تنش بیش از حد و ویرانگر بر کاروان ورزشی اعزامی به مسابقه‌ها بکاهند.

سرانجام برای توسعه ورزش راهکارهای زیر پیشنهاد می‌شود:

تمرکز بر روی متغیرهای اثرگذار در توسعه ورزش

درک روشن از نقش سازمان‌های مختلف فعال و شبکه ارتباطی موثر نگهدارنده سیستم

تسهیل مدیریت از طریق ورزش عمومی و مرزهای سیاسی

ایجاد سیستم موثر برای کنترل و شناسایی میزان پیشرفت ورزشکاران نخبه و مستعد

نظارت بر خدمات ورزشی برای ایجاد فرهنگی عالی و برتر در تمام اعضای تیم (ورزشکاران،

مربی‌ان، مدیران و متخصصان) به منظور تعامل با دیگر اعضا به روش رسمی و غیر رسمی)

برنامه‌های رقابتی ساختار یافته با نمایش مداوم در سطح بین‌المللی

تجهیزات ویژه و پیشرفته

تمرکز منابع بر روی تعداد نسبتاً کمی از ورزش‌هایی که شانس بیشتری برای موفقیت در

سطح جهان دارند، البته این به معنای نادیده گرفتن ورزش‌های دیگر نیست.

برنامه‌ریزی جامع برای هریک از نیازهای ورزشی

شناخت میزان هزینه‌ها به همراه سرمایه‌گذاری مناسب برای زیر ساخت‌ها و مردم

حمایت از سبک زندگی سالم و فعال و آماده کردن برای زندگی پس از ورزش

توجه به ورزش مدارس و دانشگاه‌ها

توجه به ورزش همگانی

منابع

۱. محمدی علی، ۱۳۸۹، ارائه مدل ریاضی برای رتبه بندی کشورهای شرکت کننده در بازی‌های آسیایی ۲۰۰۶ قطر، فصلنامه المپیک، سال هجدهم، شماره ۳ (پیاپی ۵۱)، ۷-۱۹.
۲. منهاج، محمدباقر، مبانی شبکه‌های عصبی (هوش محاسباتی)، انتشارات دانشگاه صنعتی امیر کبیر (پلی تکنیک تهران)، ۱۳۸۸.
۳. ناظمی امیر، قدیری روح الله، آینده نگری از مفهوم تا اجرا، مرکز صنایع نوین، فراندیش، ۱۳۸۵.
4. Anderson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting* , 21, 565–576.
5. Bergsgard, N. A., Houlihan, B., Mangset, P., Nodland, S. I., & Rommetveldt, H. (2007). *Sport policy. A comparative analysis of stability and change*. London: elsevier.
6. Bolton, R., & Chapman, R. (1986). Searching for positive returns at the track: a multinomial logit model for handicapping horse races. *Management Science* , 32, 1040–1060.
7. Bosscher, V. D., Knop, P. D., Bottenburg, M. V., & Shibli, S. (2006). A Conceptual Framework for Analysing Sports Policy Factors Leading to International Sporting Success. *European Sport Management Quarterly* , 6, 185- 215.
8. Boulrier, B. L., & Stekler, H. (2003). Predicting the outcomes of National Football League games. 19, 257–270.
9. Boulrier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors?: an evaluation. *International Journal of Forecasting* , 15, 83–91.
10. Bunn, d., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science* , 37, 501–518.

11. Caudill, S. (2003). Predicting discrete outcomes with the maximum score estimator: the case of the ncaa men's basketball tournament. *International Journal of Forecasting* , 19, 313–317.
12. Clarke, S., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research* , 7, 585–594.
13. Collopy, F., Lawrence, M. J., & Wright, G. (1996). The role and validity of judgment in forecasting. *International Journal of Forecasting* , 12, 1-8.
14. Condon, E. M., Golden, B. L., & Wasil, E. A. (1999). Predicting the success of nations at the Summer Olympics using neural networks. 26, 1243-1265.
15. Corral, J. d., & Rodriguez, J. P. (2010). Are differences in ranks good predictors for Grand Slam? *International Journal of Forecasting* , 26, 551–563.
16. Derevenco, P., Albu, M., & Duma, E. (2002). Forecasting of top athletic performance. *Rom J Physiol* , 39, 57- 62.
17. Forrest, D., & Simmons, R. (2000). Forecasting sports results: the behaviour and performance of football tipsters. *International Journal of Forecasting* , 16, 317–331.
18. Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: the case of english football. *International Journal of Forecasting* , 21, 551–564.
19. Green, M., & Houlihan, B. (2005). *Elite sport development. Policy learning and political priorities.* London and new York: Routledge.
20. Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law* , 2, 293–323.

21. Grove, W., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment* , 12, 19–30.
22. Heinila, K. (1982). The totalisation process in international sport. Toward a theory of the totalization of competition in top-level sport. *Sportwissenschaft* , 3, 235- 253.
23. Holmes, G., Hall, M., & Frank, E. (۱۹۹۹). Generating Rule Sets from Model Trees, AI '99 Proceedings of the 12th Australian Joint Conference on Artificial Intelligence: Advanced Topics in Artificial Intelligence, **Springer-Verlag London**.
24. Iyer, S. R., & Sharda, R. (2009). Prediction of athletes performance using neural networks: An application in cricket team selection. *Expert Systems with Applications* , 36, 5510–5522.
25. Klaassen, F., & Magnus, J. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research* , 148, 257–267.
26. Lovalgia, M. J., & Lucus, J. W. (2005). High visibility athletic programs in the Football. *The Sport Journal* , 8 (2), 1- 5.
27. Min, B., Kim, J., Choe, C., Eom, H., & McKay, R. (. (2008). A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems* , 21, 551–562.
28. Oakley, B., & Green, m. (2001). The Production of Olympic Champions: International Perspectives on Elite Sport Development System. *European Journal for Sport Management* , 8, 83–105.
29. Quinlan, J. R. (1992). Learning with continuous classes. In Proc. of the Fifth

- Australian Joint Conference on Artificial Intelligence, World Scientific, Singapore, 343-348.
30. Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *The statistician* , 49 (3), 399-418.
31. Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting* , 23, 415-426.
32. Smith, T., & Schwertman, N. (1999). Can the ncaa basketball tournament seeding be used to predict margin of victory? *The american statistician* , 53 (2), 94-98.
33. Song, C., Boulier, B. L., & Stekle, H. O. (2007). The comparative accuracy of judgmental and model forecasts of American football games. *International Journal of Forecasting* , 23, 405- 413.
34. Wang, Y., Witten, I. H. (1997). Induction of model trees for predicting continuous classes. In *Proc. of the poster papers of the European Conference on Machine Learning*, pages 128-137, Prague, Czech Republic
35. Webby, R., & O'Connor, M. (1996). Judgmental and statistical time series forecasting: A review of the literature. *International Journal of Forecasting* , 12, 91-118.
36. Wright, G., Lawrence, M. J., & Collopy, F. (1996). Editorial: The role and validity of judgment in forecasting. *International Journal of Forecasting* , 12, 1-8.