

به نام خداوند بخشنده مهربان



عنوان پژوهش و گزارش:

تحلیل استنتاج های آماری داده های مربوط به مسابقات فوتبال بین المللی

نگارش:

نگار درویشی

فروردین ۱۴۰۰

چکیده

فوتبال یکی از ورزشهای پرطرفدار در سراسر جهان است که احتمالاً بیش از هر ورزش دیگری از آن تجلیل و بازدید می شود. اثبات این گزارش هنگامی است که ایالات متحده آمریکا ۵ - ۲ ژاپن را در جام جهانی فوتبال زنان ۲۰۱۵ شکست داد و با میانگین رکورد ۲۵,۴ میلیون دلار طرفدار تماشاگران فقط در ایالات متحده ، تیم خود را برنده کرد. تجزیه و تحلیل ورزشی در حال افزایش است و به طور گسترده ای برای تجارت ، شرط بندی و بهبود تکنیک ها و بهره وری هر باشگاه ، هر تیم اهمیت پیدا می کند. این پروژه تجزیه و تحلیل های مشابه (توصیفی و پیش بینی شده) را با برای مثال توجه به لیگ فوتبال اسپانیا La liga برجسته می کند. تیم های اسپانیایی در چند وقت اخیر بر دنیای فوتبال تسلط یافته اند و به همین دلیل , دلیل زیادی برای یادگیری در مورد آنها و مدل سازی ایجاد کرده اند.

پیش بینی نتایج مسابقات فوتبال ، چالش بسیار جالبی برای اکثر طرفداران فوتبال و صاحب نظران به وجود می آورد. با این حال ، درست مانند پیش بینی بازار سهام ، مسئله این است که این یک مسئله غیر قطعی است. در این مسئله نمی توانید انتظار ۱۰۰٪ دقت داشته باشید زیرا این امر نشان می دهد که می توانید آینده را پیش بینی کنید! در طول توسعه پروژه ، چندین روش را امتحان کردیم و متعاقباً بهترین مدلی را انتخاب کردیم که نتایج بهینه را ارائه می داد.

فهرست مطالب

مقدمه.....	۶
ابزارهای به کار برده شده.....	۹
درک داده ها.....	۱۰
آماده سازی داده ها.....	۱۱
مدل سازی.....	۱۱
برآورد عملکرد برای طبقه بندی برای مدل های طبقه بندی.....	۱۴
اعتبار سنجی متقابل K-Fold.....	۱۴
تقسیم های تکراری آموزش / آزمون.....	۱۵
مرحله ارزیابی.....	۲۱
مدل اول.....	۲۱
درخت تصمیم.....	۲۲
جنگل تصادفی.....	۲۳

مدل تقویت گرادیان.....۲۵

بیز.....۲۶

مدل ارزیابی دوم.....۳۰

درخت تصمیم.....۳۰

جنگل تصادفی.....۳۱

بیز.....۳۲

افزایش شیب (Gradient Boosting).....۳۴

نتایج نهایی (به طور خلاصه نشان داده شده).....۳۸

رسم مقایسه بین کلاس پیش بینی شده ما با کلاس واقعی.....۳۹

مقایسه احتمالات.....۴۱

نتیجه.....۴۳

مراجع.....۴۴

مقدمه

فوتبال Laliga Div-I هر ساله از تابستان برگزار می شود و یک سال تقویمی ادامه دارد. این بازی مسابقات آدرنالین بالا بین برخی از تیم های بسیار معتبر در جهان است (که از این واقعیت ناشی می شود که این لیگ بهترین بازیکنان جهان ، بهترین تیم ها را تشکیل می دهد و معمولاً بیشتر اوقات به عنوان قهرمان لیگ قهرمانان انتخاب می شوند - بازی برای تیم هایی از لیگ های مختلف.) در مجموع تیم های ۲۰ دلاری برای مسابقات قهرمانی لیگ داخلی در سال مبارزه می کنند ، هر تیم ۳۸ بازی انجام می دهد ، در خانه با تیم مشابه ملاقات می کند و در مسابقات خارج از خانه نیز حضور دارد. پیروزی های رو در رو معمولاً در انتها برای تعیین برنده نهایی بسیار مهم تلقی می شوند. به عنوان مثال - اگر تیمی که پس از ۳۸ مسابقه ، با تیم دیگری که امتیازات یکسانی را کسب می کند مساوی شود ، برنده با توجه به گلزنی و خصوصیات کلی با عملکرد سربازی این دو تیم تصمیم می گیرد. علاوه بر این ، میلیون ها نفر شرط بندی های خود را روی بازی قرار می دهند. برای کسب درآمد ، کتابهای ورزشی شانس بازیها را طوری تنظیم می کنند که احتمال نتیجه بازی را منعکس کنند. وقتی توزیع متعادل بین شرط بندی ها برای هر دو تیم وجود دارد ، کتابهای ورزشی بیشترین سود را کسب می کنند ،

بنابراین ممکن است شانس را برای تشویق بیشتر شرط بندی ها به انتخاب یک طرف خاص تغییر دهد. شرط بندی ها به دنبال پیدا کردن ناکارآمدی در شانس هستند ، یعنی شانس هایی که به طور دقیق احتمال نتیجه یک بازی را منعکس نمی کنند.

بنابراین ، در این ساخت مدل ، هدف ما تولید احتمال از دست دادن یا تساوی بردن ، و مقایسه آن با شانس شرط بندی (احتمالات) تعیین شده است و نحوه عملکرد ما را خواهید دید.

برای شرط بندی ورزشی ، یک شرط گذار می تواند با مقایسه شانس کتابهای ورزشی با برداشت شخصی خود از احتمال واقعی بودن ، بفهمد که آیا لبه دارد.

این شانس ها نیز با احتمال پیروزی مطابقت دارند ، با استفاده از فرمول های

$$\text{احتمال برنده برای شانس مثبت} = ۱۰۰ / (\text{به علاوه شانس}) + ۱۰۰$$

$$\text{احتمال برنده برای منهای شانس} = - (\text{منهای شانس}) / - (\text{منهای شانس}) + ۱۰۰$$

تبدیل فرد از مبدل ها می تواند به ما امکان ایجاد احتمالات پیروزی مربوطه را بدهد. توجه داشته باشید که این درصدها تا ۱۰۰٪ جمع نمی شوند. در پایان ، علاقه مند به داشتن یک شرط بندی متعادل برای هر دو طرف هستند ، بنابراین آنها شانس را بر این اساس تنظیم می کنند تا انگیزه بیشتری فراهم کنند. همچنین ، از دیدگاه شرط بندی کنندگان ، آنها شانس را کمی بدتر از شانس واقعی قرار می دهند.

ابزارهای به کار برده شده:

همانطور که می دانیم پایتون در استخراج داده و دستکاری بسیار مهارت دارد. در اینجا ما از آن برای گرفتن داده های ساخت یافته و بدون ساختار و تراز کردن آنها با یک قالب خاص که کار با آن آسان است ، استفاده می کنیم. سپس این داده های ساخت یافته به اتصال پایگاه داده NoSQL با MongoDB برای مدیریت داده ها و به روزرسانی آنها در صورت لزوم ارسال می شود. این داده ها به Elasticsearch منتقل می شوند که نمایه سازی انحراف را برای کمک به جستجوی سریعتر داده ها فراهم می کند. این داده ها سپس با R دنبال می شوند ، با اتصال پایگاه داده از انتهای دیگر برای ایجاد جریان خط لوله مجازی. داده های وارد شده به R برای انجام تجزیه و تحلیل و پیش بینی داده های توضیحی استفاده می شود ، از تجزیه و تحلیل داده های توصیفی تا تجزیه و تحلیل نسخه ای با تمرکز بر به دست آوردن آمار خلاصه از چندین متغیر (متغیرهای اصلی) و استفاده از مدل های کارآمد و مناسب پس از آزمایش، برای دقت علاوه بر این ، از جدول ، برای کاوش داده ها در مرحله اولیه و کسب اطلاعات در مورد ساختار داده ها استفاده شده است.

درک داده ها

داده ها حاوی اطلاعاتی مانند گل‌های تیم میزبان ، تیم خارج از خانه ، خطاهای مرتکب شده ، کرنرهای گرفته شده ، ضربه های شلیک شده و هدف و همچنین نتیجه تمام وقت است که می خواهیم پیش بینی کنیم. توجه داشته باشید که مجموعه داده ها همچنین حاوی شانس شرط بندی از کارشناسان شرط بندی است که بعداً برای مقایسه استفاده خواهیم کرد. داده ها گسترده و جامع هستند و برای پیش بینی مناسب هستند. این شامل جزئیات مربوط به هر تیم و معیارهای مختلف مرتبط با آن است. همچنین برای لیگ های مهم فوتبال اروپا آمار بازی ها مانند شوت به سمت دروازه ، کرنر ، خطا ، آفساید ، رزرو ، کارت قرمز و داوران درج شده است.

آماده سازی داده ها

داده ها تصفیه و زیر مجموعه شدند تا فقط شامل متغیرهای مهم باشند.

معیارهای دیگری که بعداً اضافه می شوند عبارتند از: آب و هوا ، مسافرت به کشورهای مختلف ، وزن به چهار تیم اول هر فصل تا آنها را از تیم های ضعیف تر متمایز کند زیرا این قطعاً باید بر نتایج نهایی تأثیر بگذارد.

مدل سازی

خلاصه ای کوتاه از تکنیک هایی که ما امتحان کردیم و استدلال ما برای گنجاندن آنها:

درخت تصمیم - درختان تصمیم برای تفسیر پذیری و داده های پیچیده مناسب هستند. علاوه بر این ، ما می خواستیم عملکرد یک درخت تصمیم واحد را با رویکرد (جنگل تصادفی) مقایسه کنیم.

جنگل تصادفی - همانطور که می دانیم این یک کیسه درخت است که روی چندین درخت تصمیم با هم ساخته شده است. آنها از شاخص های خوبی برای طبقه بندی

استفاده می کنند و همچنین بیشترین استفاده را دارند.

Naive Bayes - برای خاتمه یافتن عادلانه ارزیابی مورد استفاده قرار گرفت و ما نتایج خوبی از nb به ویژه در مدل دوم خود دیدیم ، جایی که از نتیجه نیمه وقت نمی دانستیم. نتیجه در مجموعه تمرینات خوب بود اما تغییر سریعی در مجموعه اعتبار سنجی / آزمایشی ما نشان داد.

ماشین بردار پشتیبان - می خواستیم از هر دو روش خطی و هسته ای که در svm ارائه شده است استفاده کنیم. جالب توجه است که انتخاب های خطی و غیرخطی نتایج خوبی را به همراه داشته است. علاوه بر انجام طبقه بندی خطی ، SVM ها می توانند با استفاده از آنچه ترفند هسته نامیده می شود ، طبقه بندی غیرخطی را به طور موثر انجام دهند و به طور ضمنی ورودی های آنها را در فضاها ی ویژگی های بعد بالا نگاشت کنند.

تقویت گرادیان - این یک تکنیک کاربردی است که به شدت در رگرسیون و همچنین طبقه بندی به کار گرفته می شود و با ساخت مدل بر روی چندین مدل پیش بینی کننده ضعیف ، به طور معمول درخت تصمیم و `esemblin` ، نتایج قدمتی را ایجاد می کند تا در نهایت با بهینه سازی عملکرد زیان ، مدل بهتری را تشکیل دهد.

رگرسیون خطی تعمیم یافته - مدل های خطی تعمیم یافته (GLM) مدل های رگرسیون را برای نتایج زیر توزیع نمایی تخمین می زنند. علاوه بر توزیع گوسی (به عنوان مثال طبیعی) ، این موارد شامل توزیع پواسون ، دوجمله ای ، گاما ، مولتی نوم و توئیدی است. هر یک هدف متفاوتی دارند و بسته به توزیع و انتخاب عملکرد پیوند ، می توانند برای پیش بینی یا طبقه بندی استفاده شوند.

برآورد عملکرد برای طبقه بندی برای مدل های طبقه بندی:

می توان از دقت کلی استفاده کرد ، اما این ممکن است وقتی کلاس ها متعادل نیستند مشکل ساز شود.

$$k = \frac{O-E}{1-E} \text{ : میزان خطای مورد انتظار}$$

جایی که O دقت مشاهده شده و E دقت مورد انتظار تحت توافق شانس است.

اعتبار سنجی متقابل K-Fold

در اینجا ، ما داده ها را به طور تصادفی به بلوک های مشخص K با اندازه تقریباً

یکسان تقسیم می کنیم. ۱- اولین بخش داده ها را کنار می گذاریم و یک مدل را

متناسب می کنیم. ۲- این مدل برای پیش بینی بلوک نگهدارنده استفاده می شود -۳ ما

این روند را ادامه می دهیم تا زمانی که همه بلوک های نگه داشته شده K را پیش بینی

کنیم.

عملکرد نهایی بر اساس پیش بینی های متوقف شده K است که معمولاً ۵ یا ۱۰ در نظر گرفته می شود و یک اعتبار سنجی متقابل را کنار می گذارد که هر نمونه به عنوان یک بلوک CV تکرار شده K-fold نسخه های مختلفی ایجاد می کند و نتایج را جمع می کند (این روش را ترجیح می دهیم).

تقسیم های تکراری آموزش / آزمون (با اعتبار سنجی متمایز از گروه)

نسبت تصادفی داده ها (مثلاً ۸۰٪) برای آموزش مدل استفاده می شود در حالی که بقیه برای پیش بینی استفاده می شود. این فرآیند بارها تکرار می شود و از عملکرد متوسط استفاده می شود. این انشعابات را می توان با استفاده از نمونه گیری طبقه ای نیز ایجاد کرد. با تکرارهای زیاد (۲۰ تا ۱۰۰)، این روش از CV , k برابر کمتر است، اما احتمالاً جانبدارانه و مغرضانه است. ما از هر دو روش برای اعتبار سنجی استفاده خواهیم کرد.

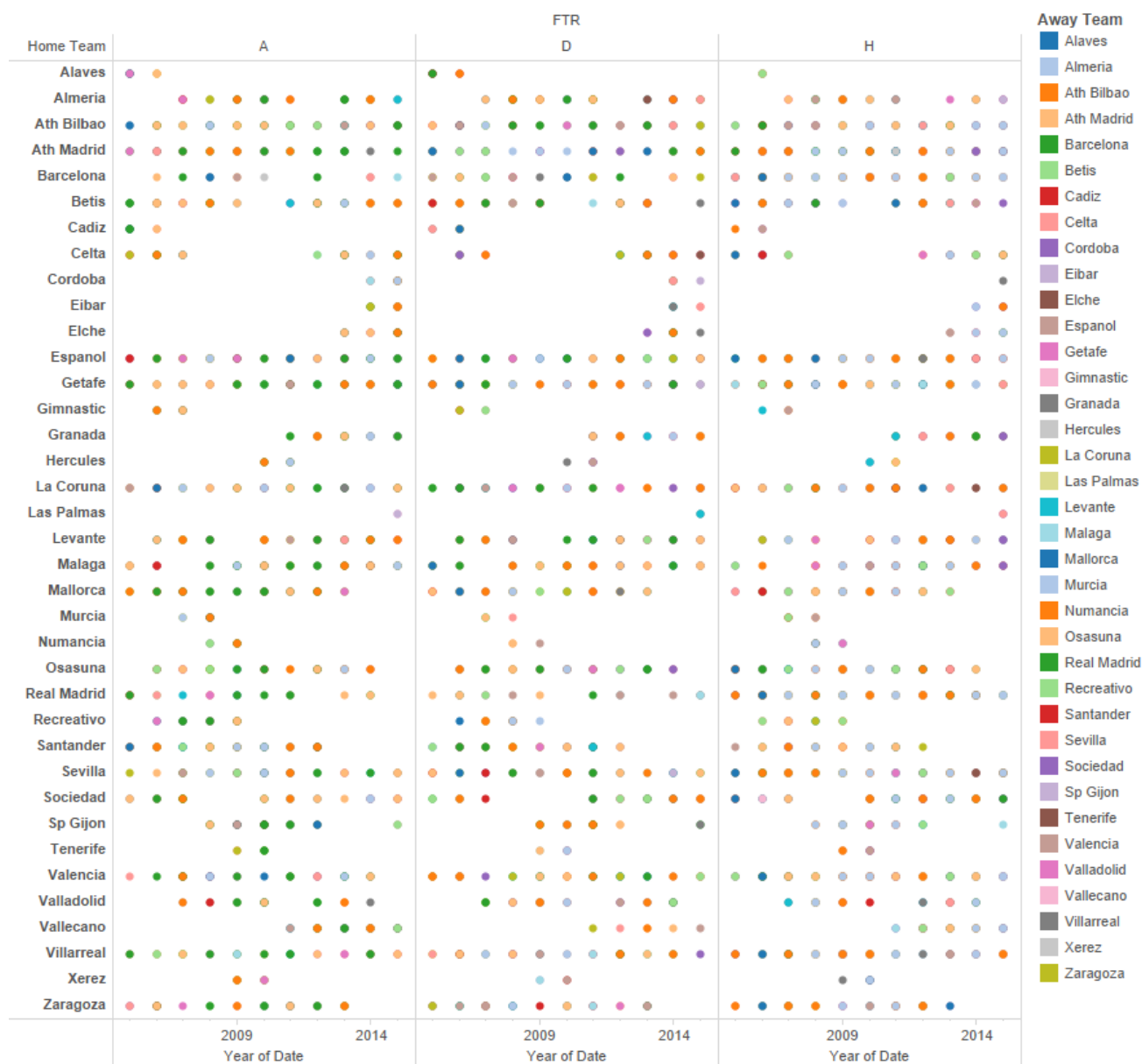
همچنین از تکنیک های تجسم داده با استفاده از Tableau برای بررسی اهمیت

برخی متغیرها و همچنین تعیین تأثیر متغیرهای خاص بر متغیر هدف علاوه بر انجام

آن در R استفاده کردیم. در زیر چند نمونه تجسم در مورد چگونگی تأثیر معیارهای

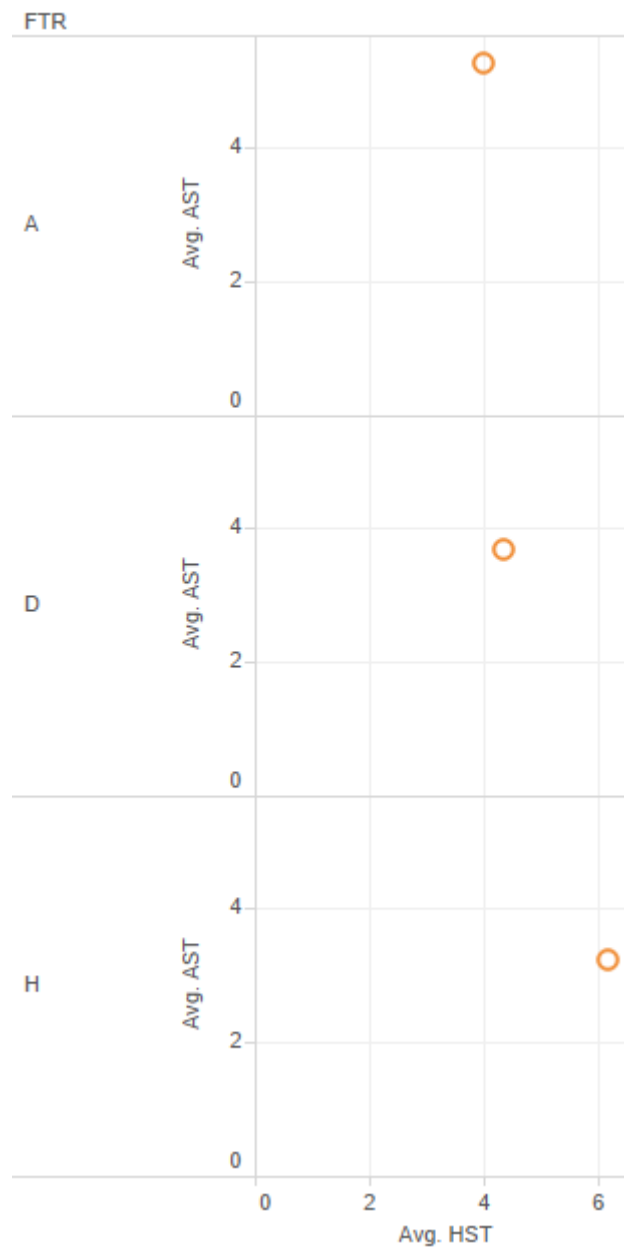
مشخصی بر بازی وجود دارد و در تعیین نتایج مطابقت مهم است.

Sheet 2

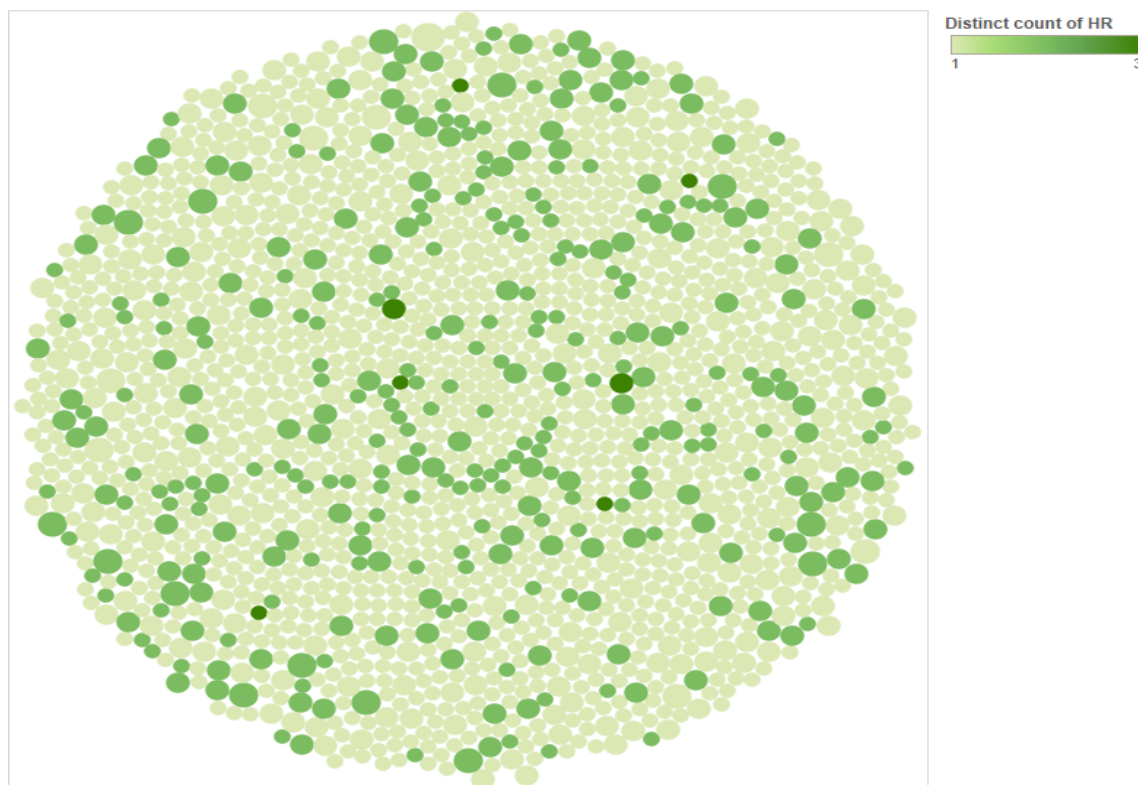


ما ابتدا با بازی های برد ، تساوی یا باخت تیم های میزبان در مقابل تیم های خارج از خانه برای هر سال (در کل مجموعه داده) ، نتایج رو در رو شروع می کنیم.

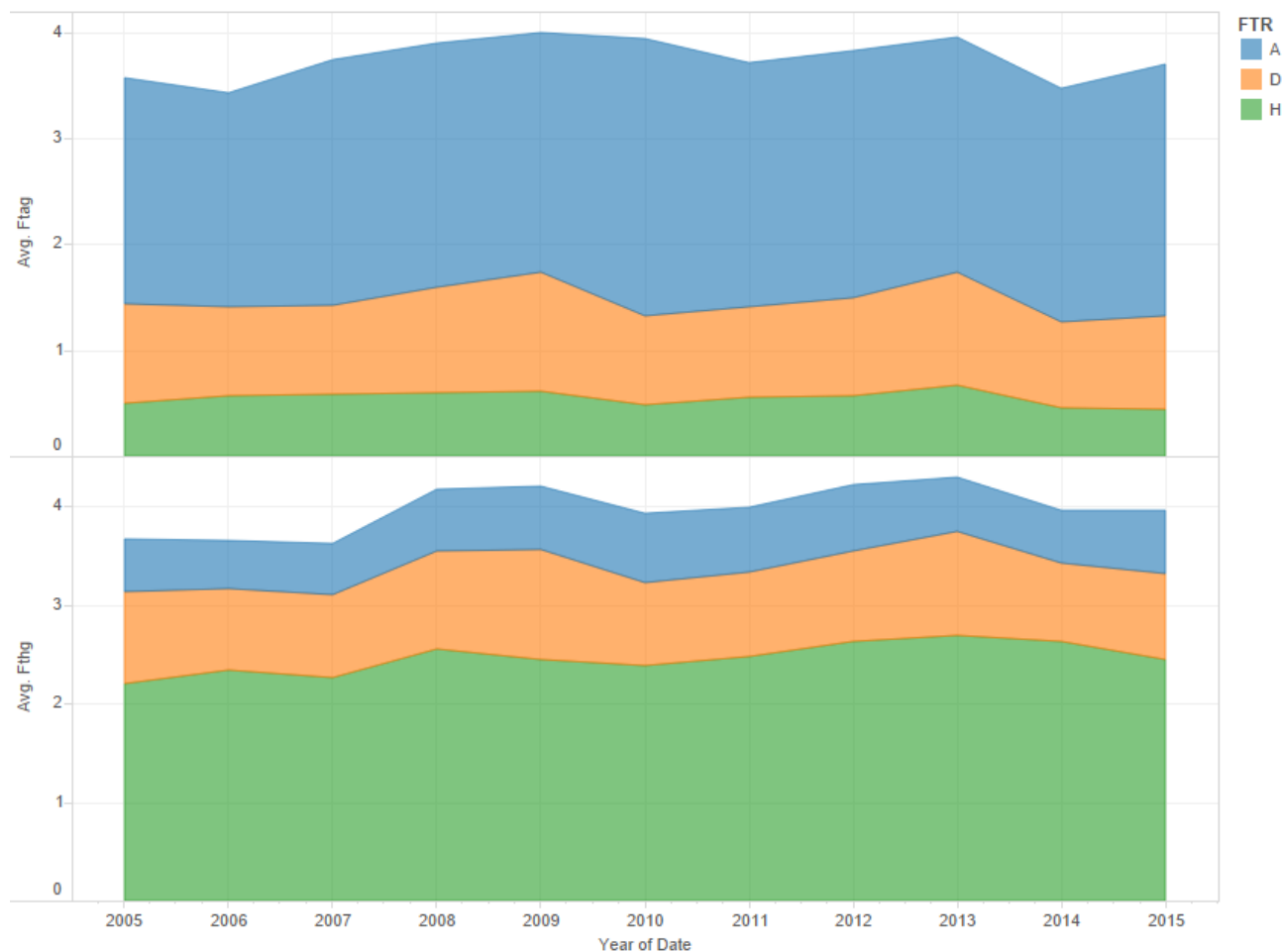
Sheet 1



وقتی تیم میهمان برنده می شود ، میانگین تعداد شوت های وارد شده به سمت دروازه (هدف) در پاسخ به تلاش تیم میزبان در حدود ۳,۹ برابر ۵,۲ بوده است. برای تیم میزبان بردن میانگین شوتها به سمت دروازه ۶,۱۶ در پاسخ به ۳,۲ توسط تیم میهمان بوده است. در حالی که بازی به سود تساوی خاتمه یافت ، ۳,۶ تلاش برای هدف توسط تیم میهمان به ۴,۳ توسط تیم میزبان انجام شد. این نشان می دهد تیم میزبان برای داشتن نتیجه مطلوب ، باید بیشتر از تیم خارج از خانه شوتزنی کند.



این تعداد کارت قرمز مشخصی را که هر تیم کسب کرده نشان می دهد. اگر کارت های قرمز بیشتر برای تیم میزبان باشد ، معمولاً ضرر می کند ، اما یک بینش هیجان انگیز به آن پیوسته است ، این زمانی است که تیمی که در خانه بازی می کند قوی است و کارت قرمز بیشتری نسبت به تیم خارج از خانه (تیم ضعیف تر) دارد ، آنها هنوز هم در نهایت برنده می شود ، که نشان دهنده قدرت تیم در تلاشهای کلی برای گلزنی خطاها و کارت های قرمز است. همچنین ، یک تئوری جالب دیگر که به آن پیوسته است این است که می توان کارتهای قرمز را به پایان بازی داد ، بنابراین ممکن است در تعیین نتیجه مسابقه نقش مهمی نداشته باشد (البته فقط در بعضی موارد)



اگر میانگین گل‌های تمام وقت ، برای تیم میهمان بیش از ۲ گل باشد (به طور متوسط) ، داده های تاریخی نشان می دهد که تیم میهمان تیم در پایان به پیروزی می رسد ، در صورتی که تیم میزبان ۱ یا کمتر امتیاز کسب کند.

اگر گل‌های تمام وقت 2 avergae یا بیشتر، تیم میزبان باشد و برعکس در مورد ۰,۵۰ برای تیم خارج از خانه تیم میزبان برنده می شود.

شانس تساوی ، همانطور که فکر می کردیم ، زمانی زیاد است که هر دو تیم تعداد گل های یکسانی را به ثمر برسانند. در بیشتر مواقع ، برای هر دو تیم به طور متوسط ۱,۵۰۸ است.

مرحله ارزیابی

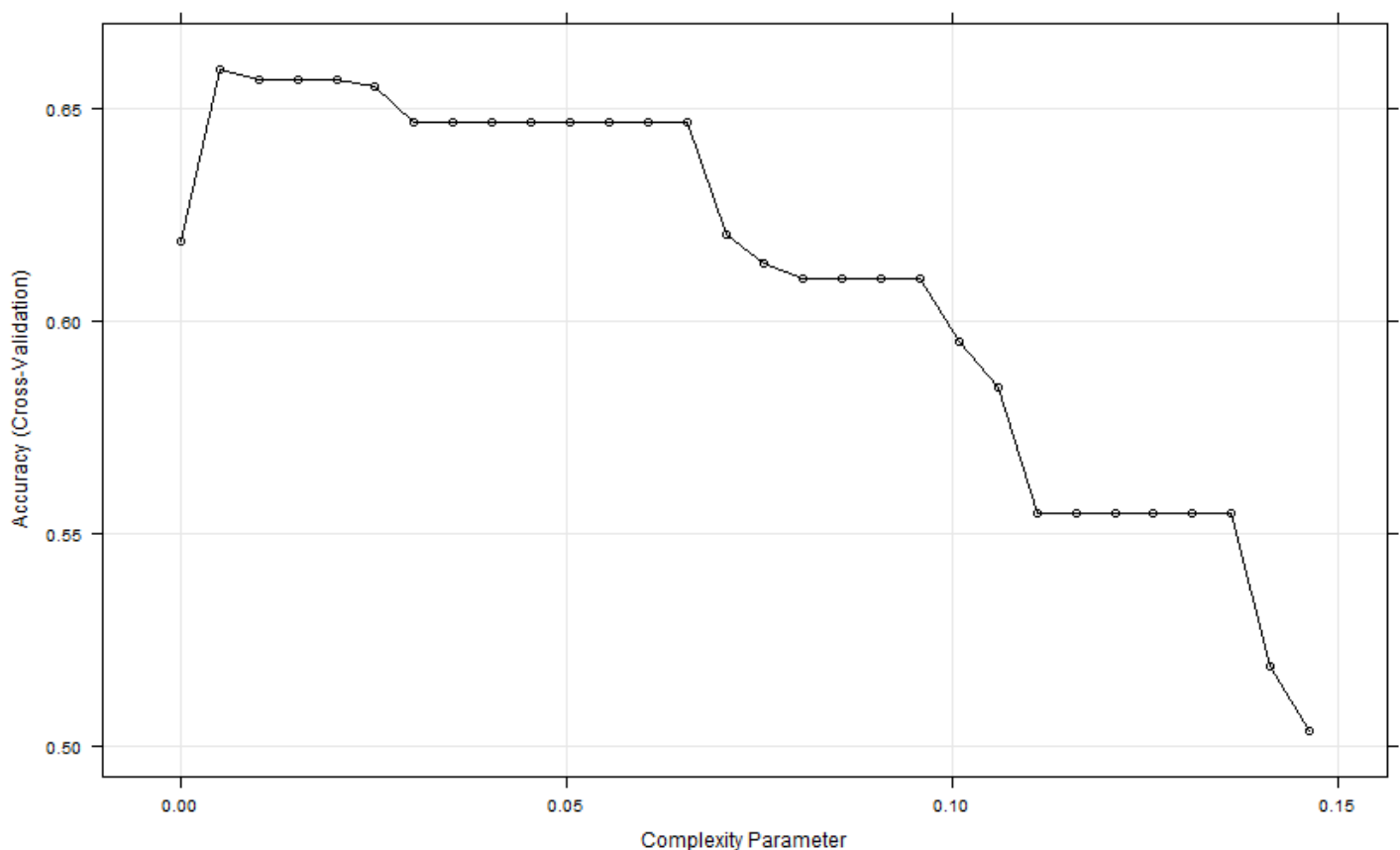
مدل اول

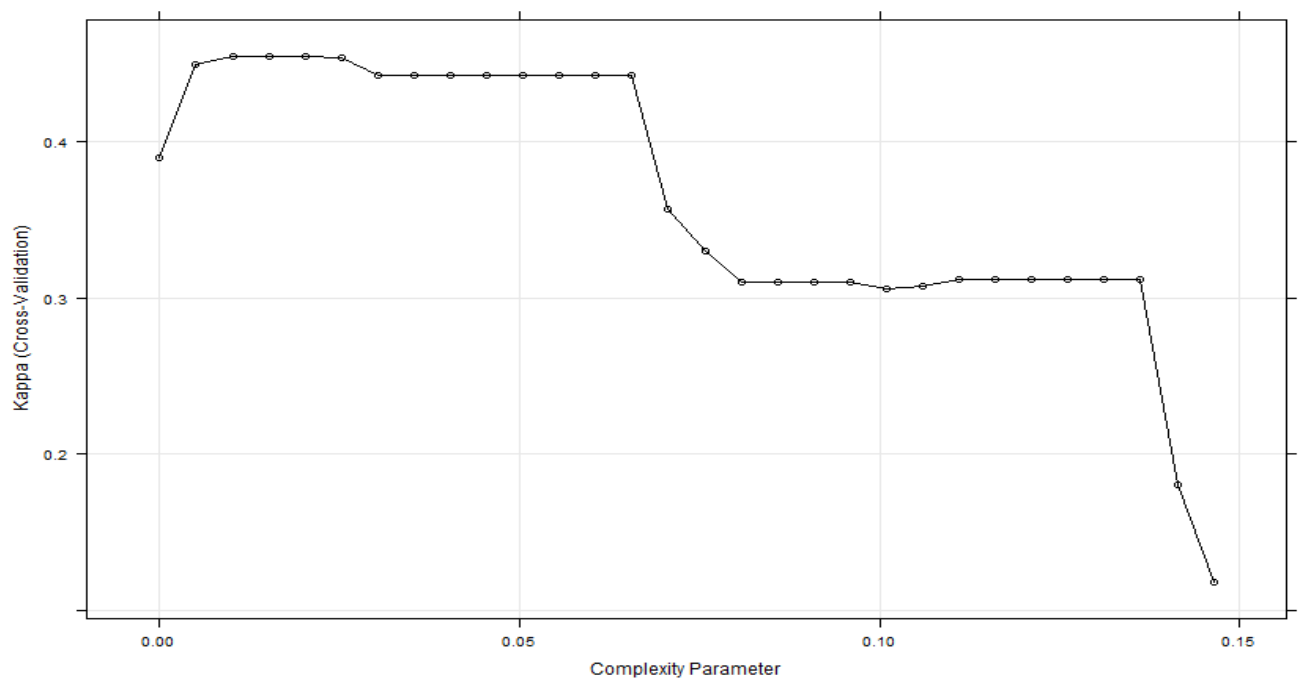
در مدل اول ، مجموعه پیش بینی یا متریک ما شامل اهداف نیمه وقت و نتایج نیمه وقت برای هر دو تیم است ، این اولین مدل است ، وقتی می خواهیم برای یک بازی پیش بینی کنیم در حالی که می دانیم نتایج نیمه وقت و عملکرد هر دو تیم ها در این مرحله ارزیابی همه مدل های فوق الذکر انجام شد و در نهایت بهترین مدل بر اساس عملکرد آموزش (اعتبار سنجی) با استفاده از اعتبار سنجی / تنظیم که به عنوان یاداش برای تأیید درک ما از مدل انجام شد ، انتخاب شد. اعتبارسنجی متقابل / تنظیم / نمونه برداری نتایج متفاوتی را ایجاد کرد که در انتخاب مدل مناسب به ما کمک کرد.

ما از دقت و مقادیر کاپا به عنوان شاخص اصلی عملکرد استفاده کردیم (برای مقادیر ROC چند طبقه معمولاً وجود ندارد و از این رو اینها ترجیح داده می شوند). آزمون t و تفاوت بین عملکرد هر مدل نیز انجام شد. اینها به شرح زیر است:

درخت تصمیم

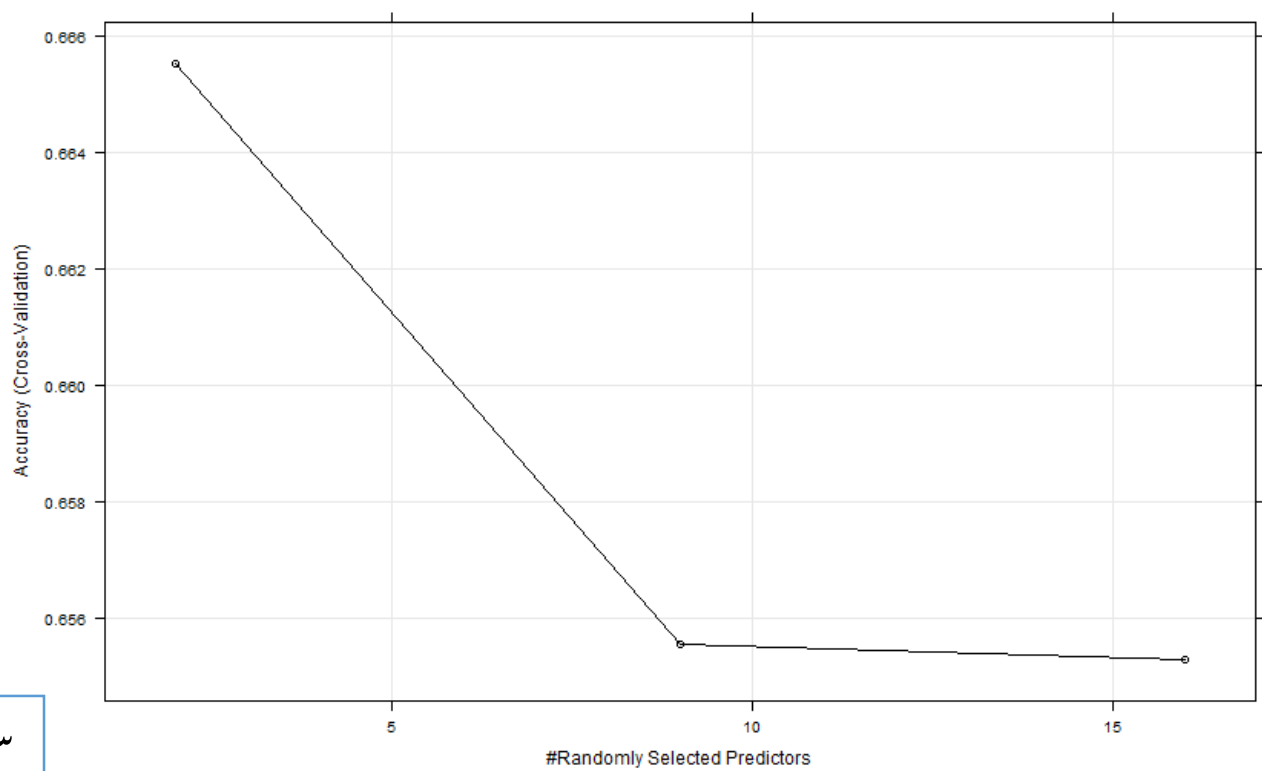
– دقت

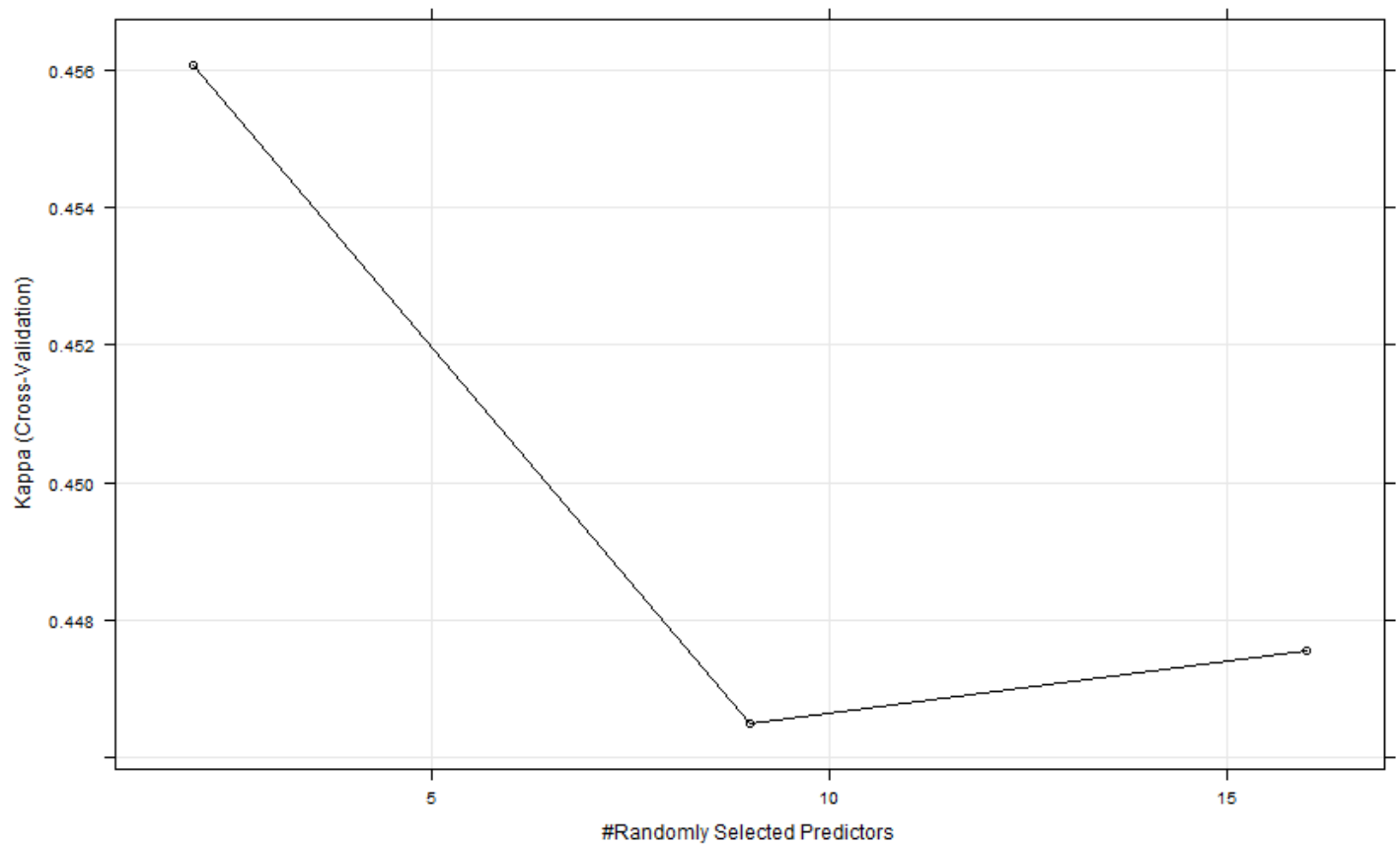




جنگل تصادفی

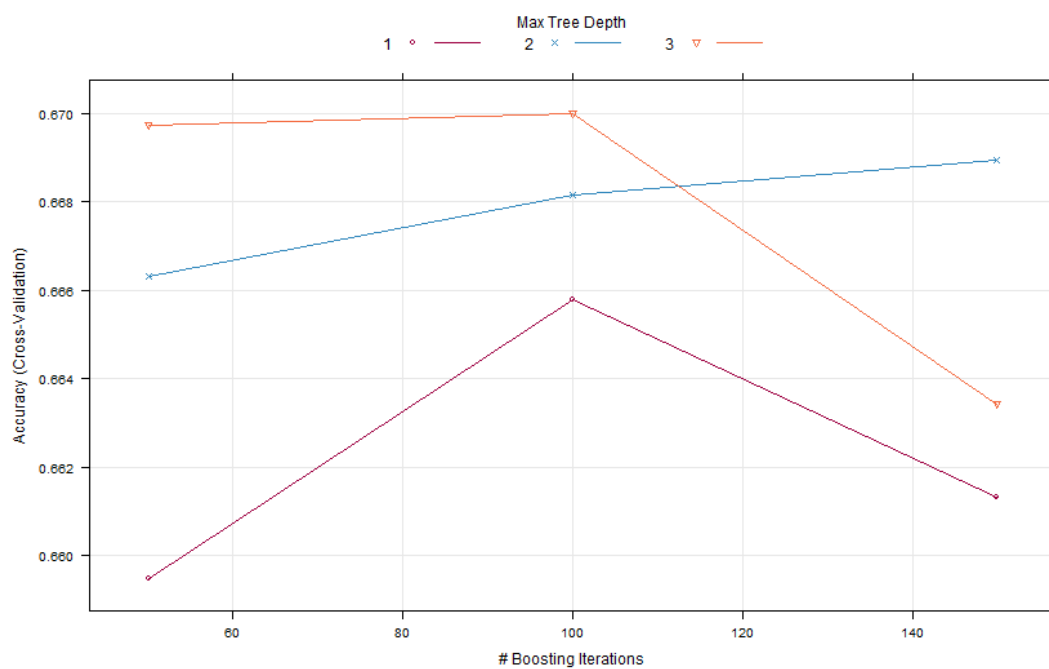
– دقت



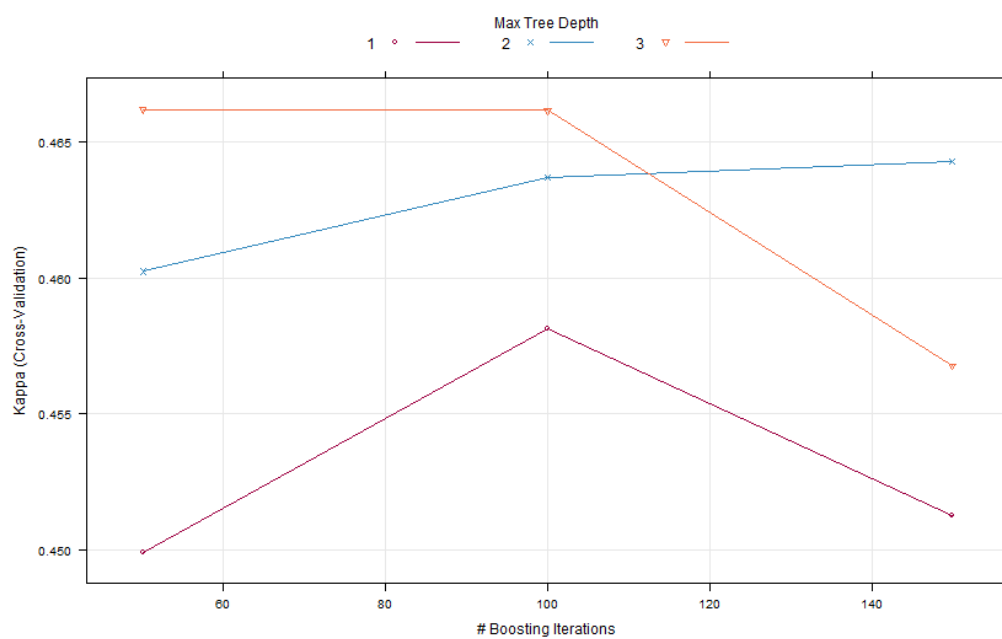


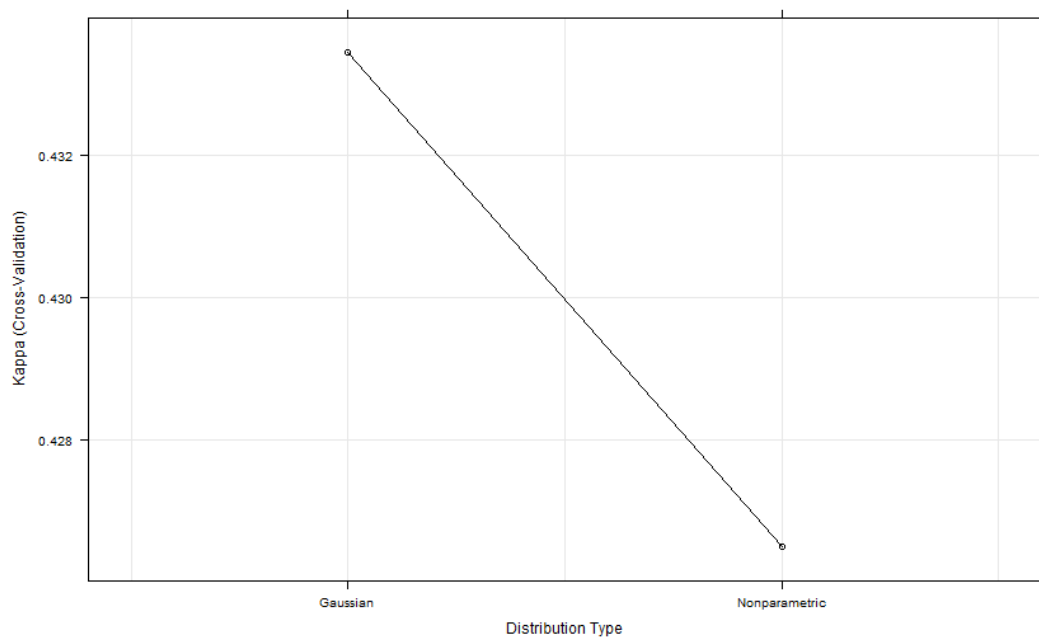
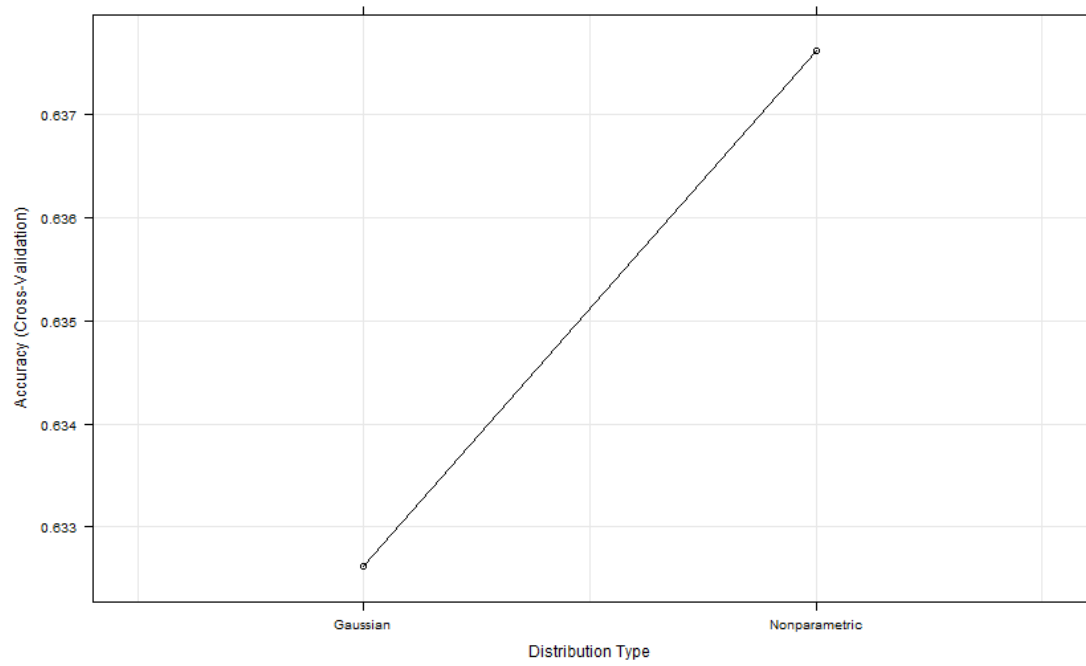
مدل تقویت گرادیان

دقت

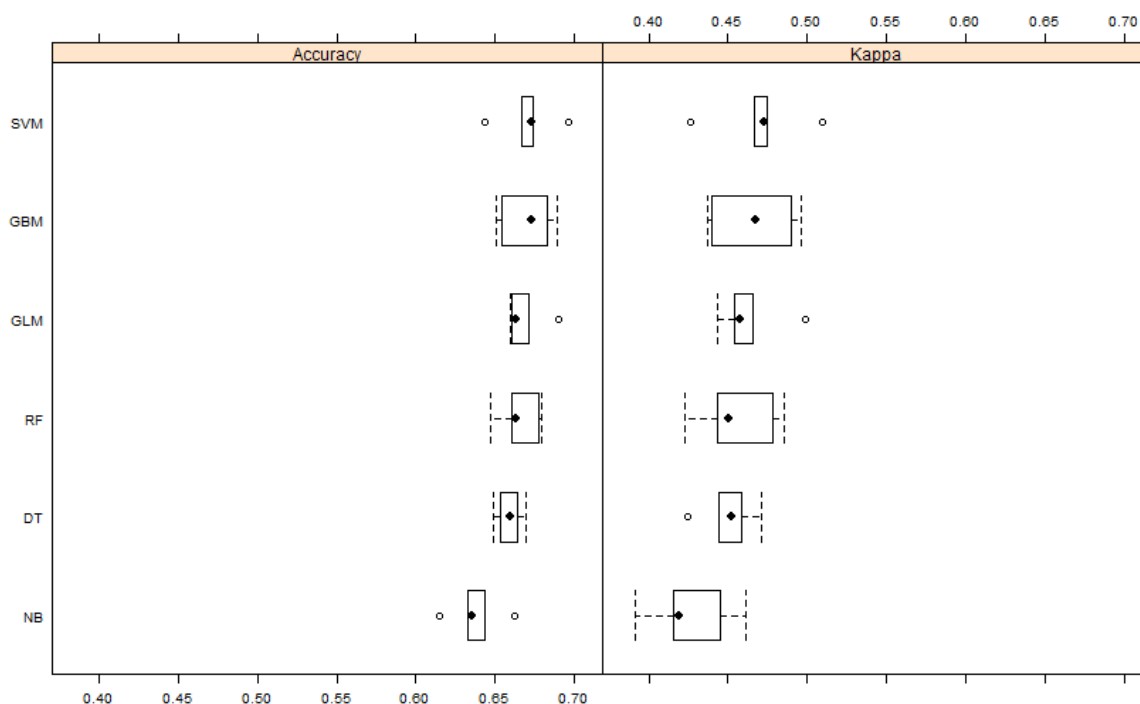


کاپا

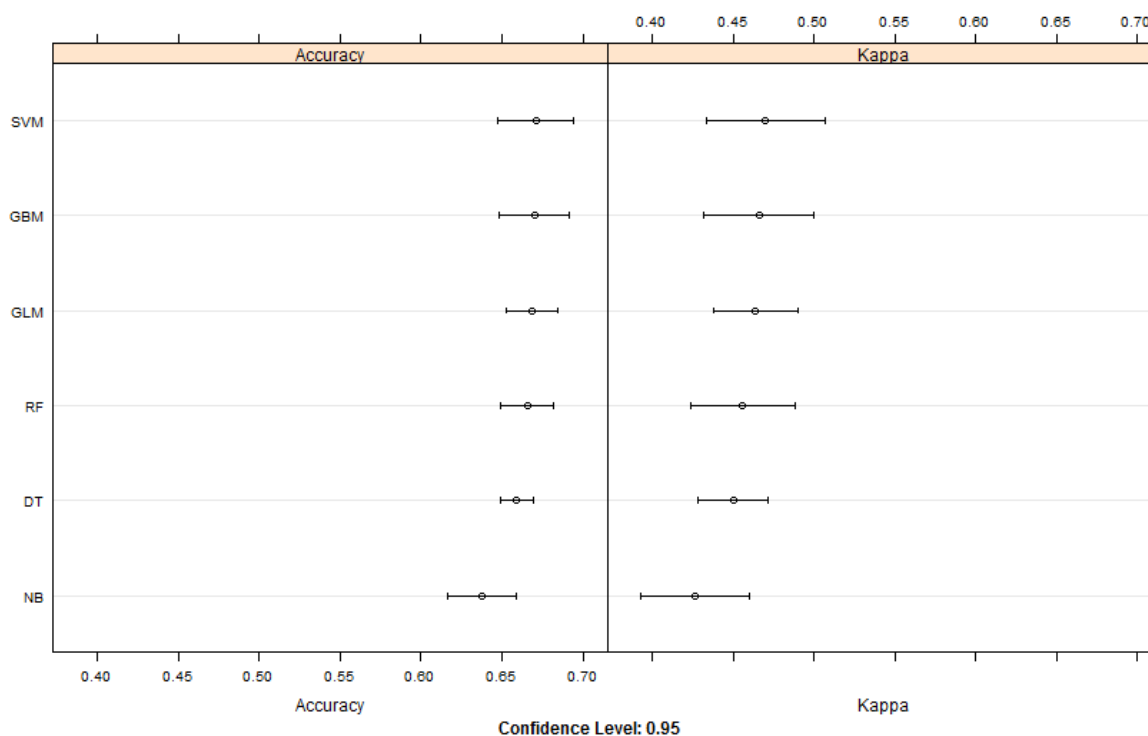




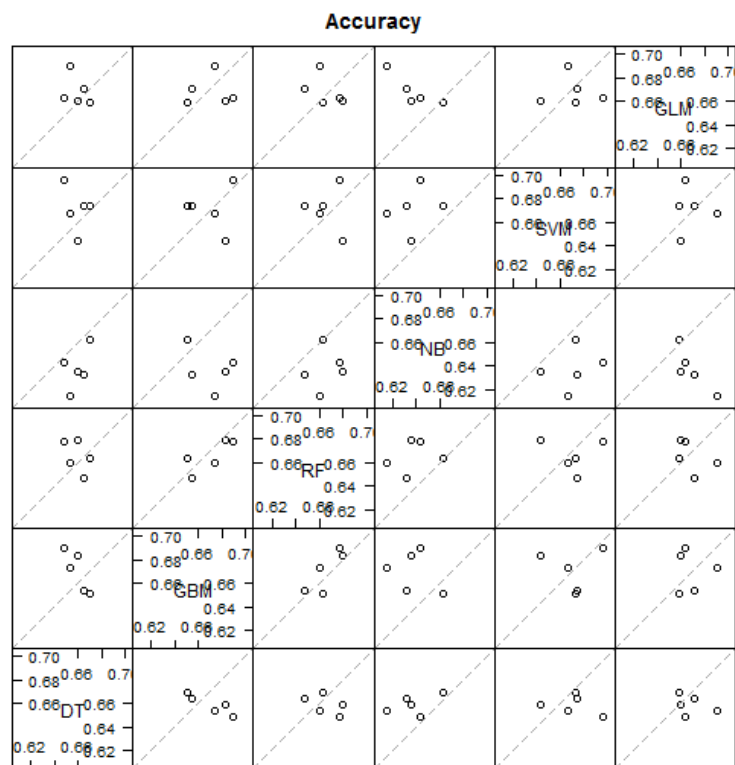
نمودار BW برای همه مدل ها که با روش نمونه گیری مجدد ترکیب شده اند.



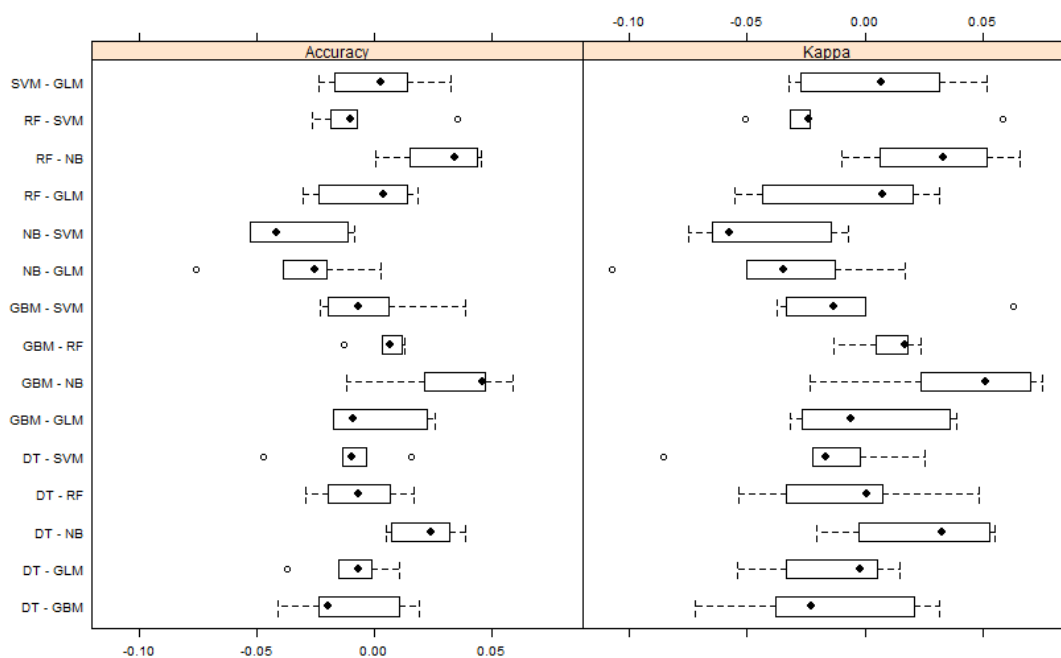
نمودار Dotplot

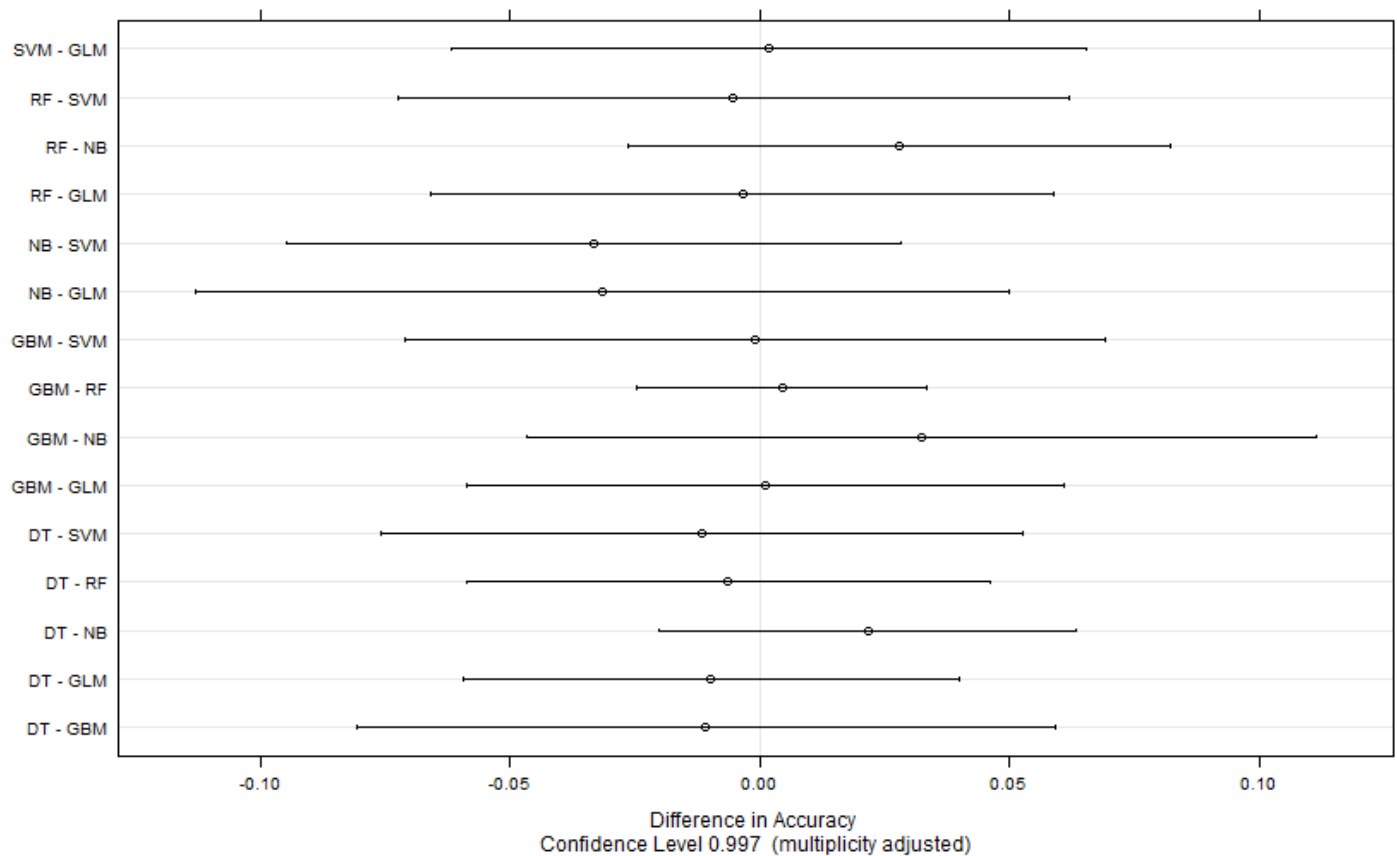


نمودار نقطه ای



مقایسه مستقیم





SVM-Linear به عنوان مدل نهایی ما برای پیاده سازی بر اساس عملکرد مجموعه داده های آزمایشی ذکر شده در بالا انتخاب شد ، SVM تولید می کند دقت متوسط ۶۶٫۸۷٪ و ارزش $cappa$ 46.84٪ .

در مجموعه تست ، دقت ۶۶٫۳۶٪ و کاپا ۴۵٫۸۵٪ مورد توجه قرار گرفت.

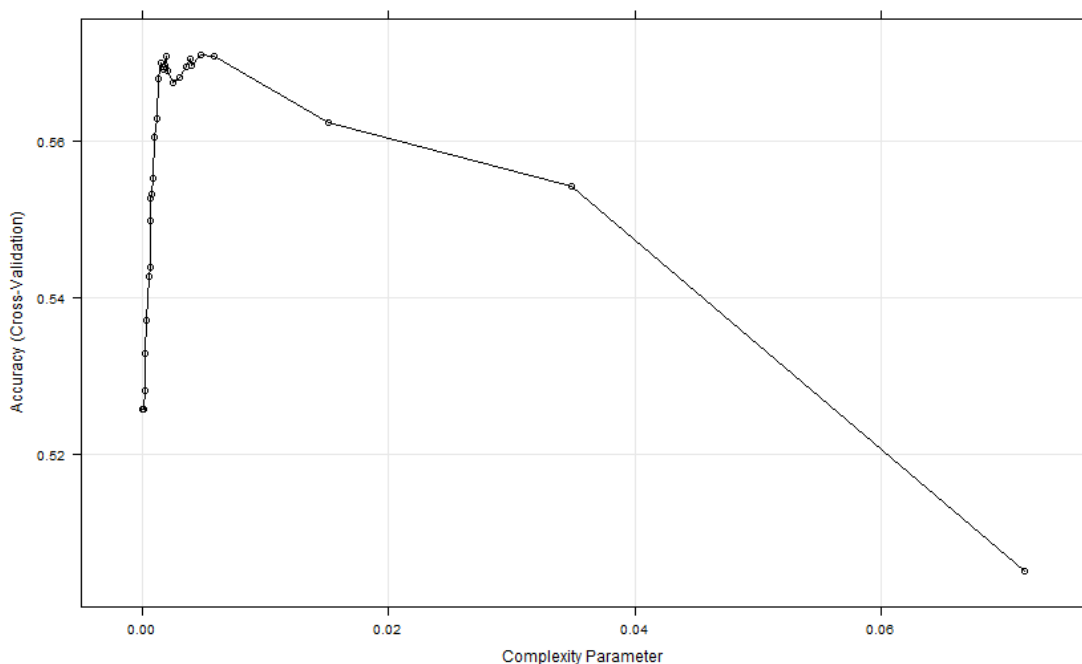
مدل ارزیابی دوم

در مدل دوم ، ما پیش بینی های خود را اصلاح می کنیم و نتایج و ویژگی های نیمه وقت را حذف می کنیم تا بدون دانستن عملکرد نیمه وقت هر تیم ، بتوانیم در ابتدای بازی پیش بینی کنیم.

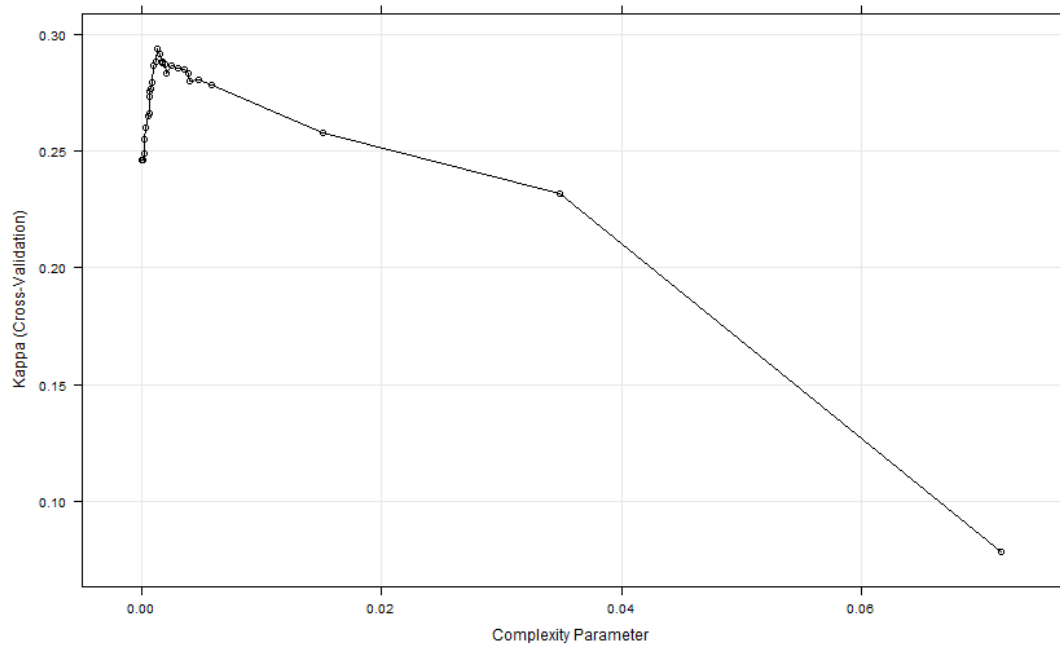
از این رو هر دو مدل ، پیش بینی های با الزامات مختلف در هنگام شرط بندی یا استفاده از سایر ارزش های تجاری را امکان پذیر می کند.

درخت تصمیم

—دقت

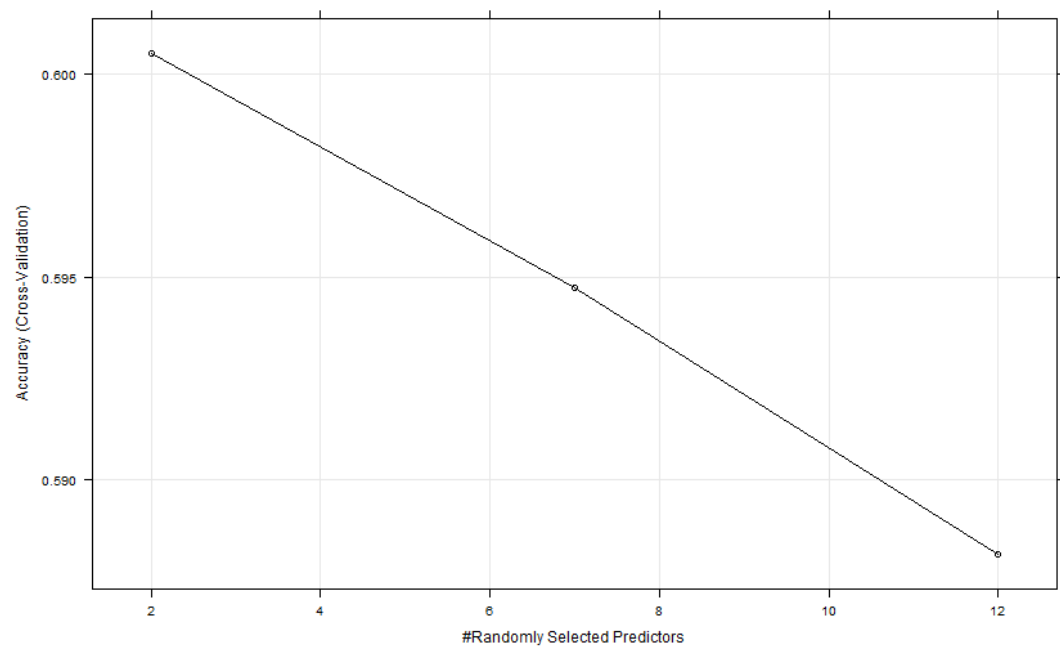


–کاپا

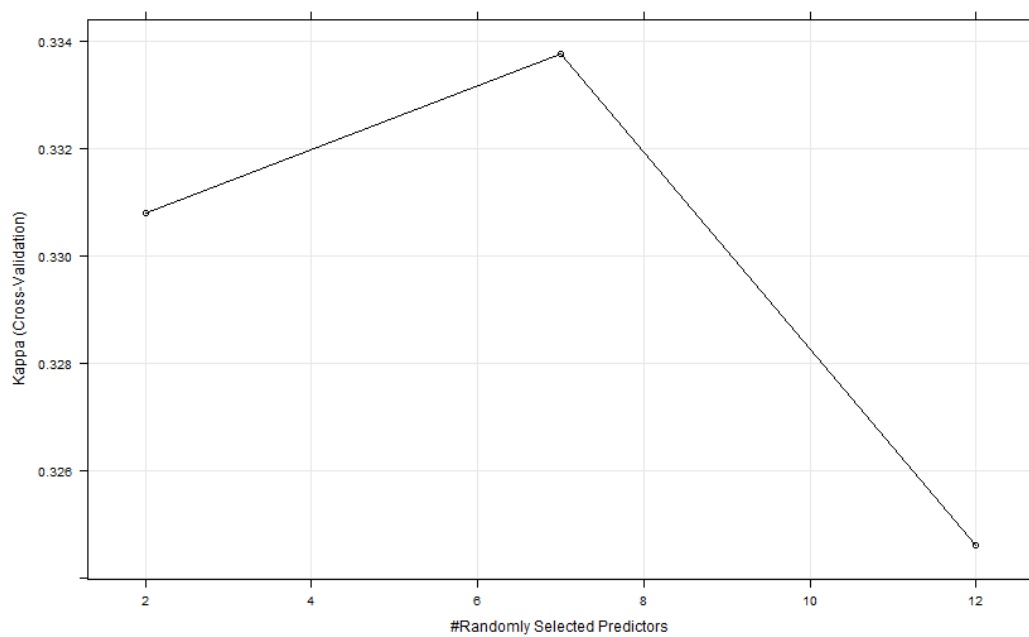


جنگل تصادفی

–دقت

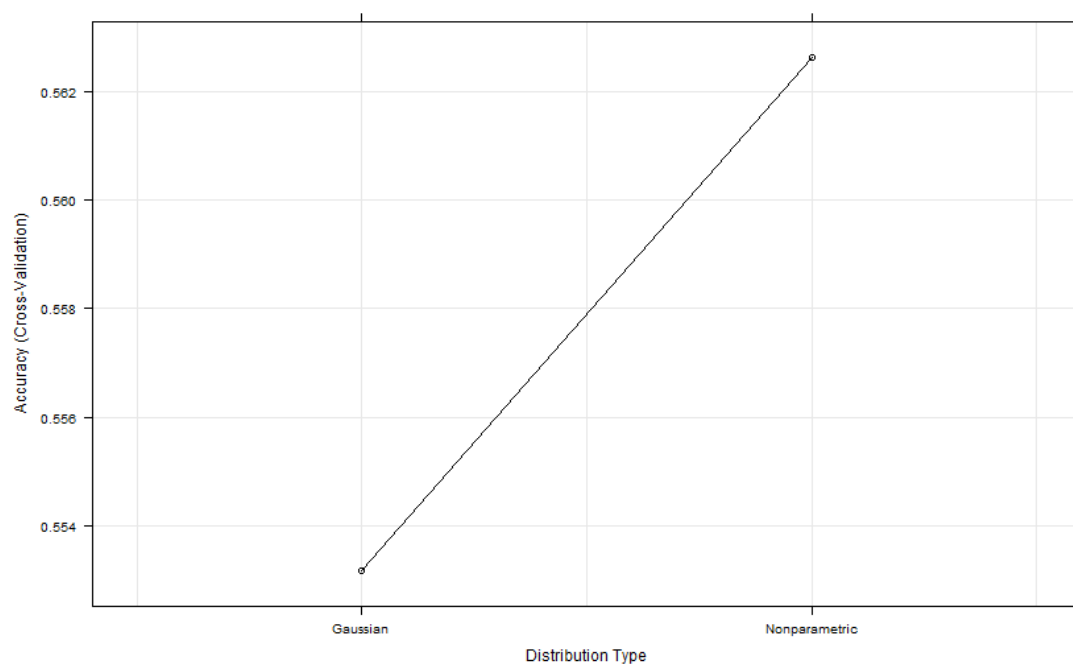


–کاپا

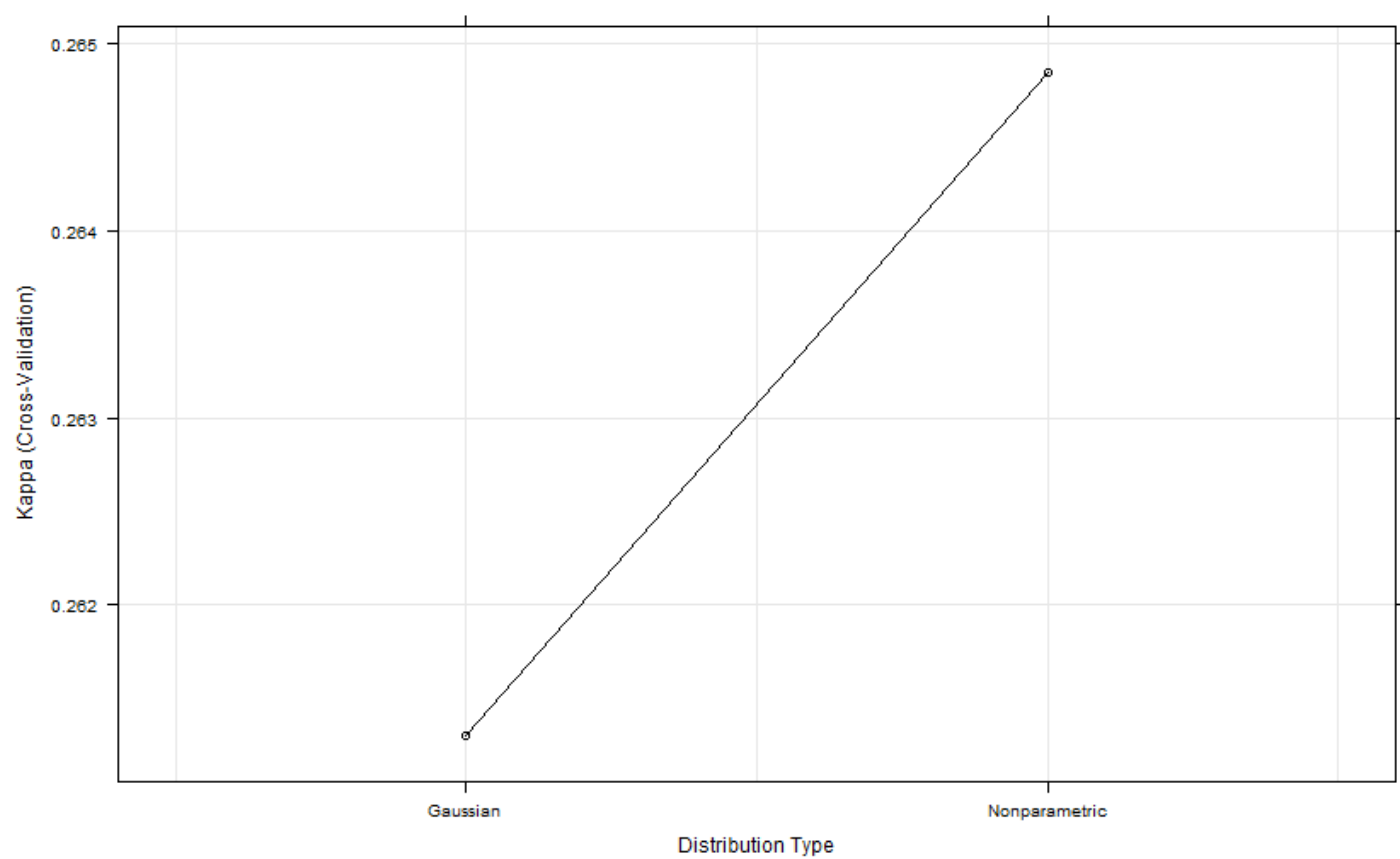


بیز

–دقت

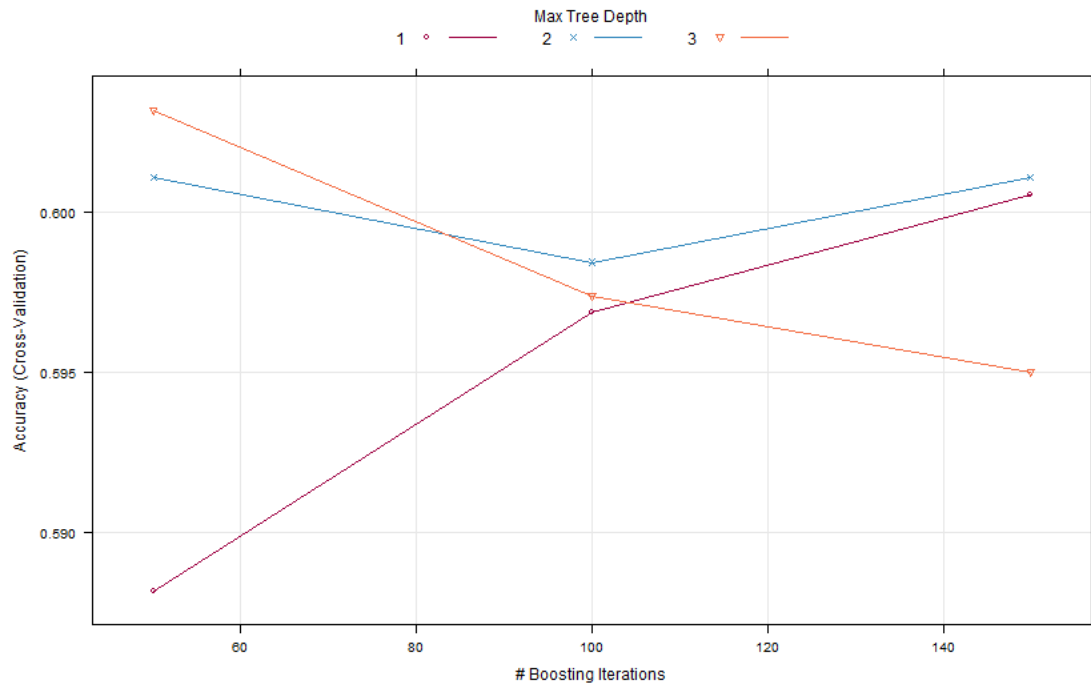


– کاپا

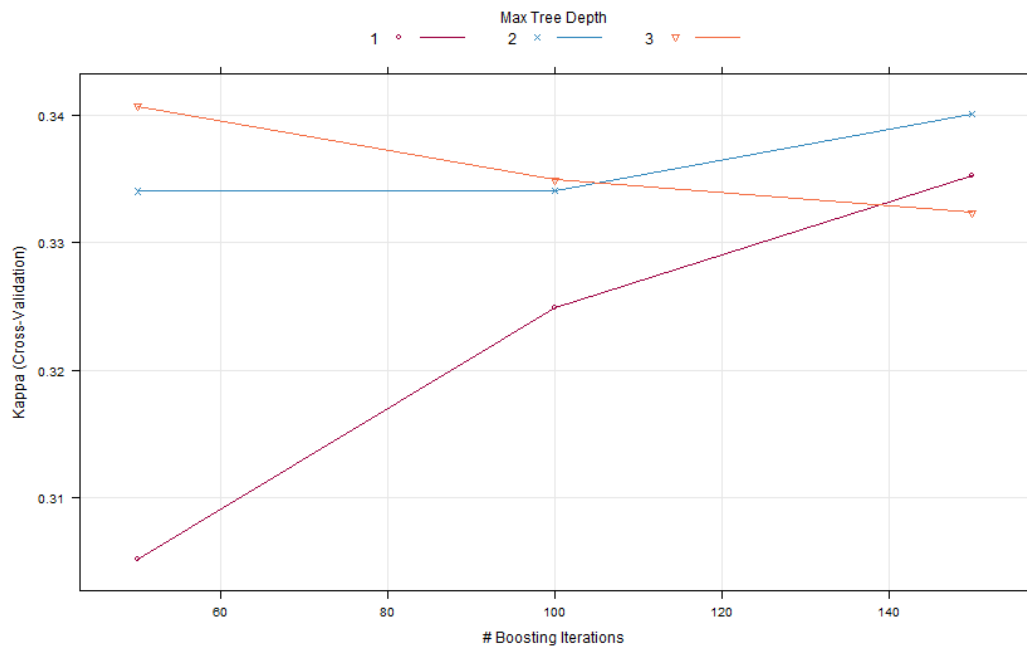


افزایش شیب (Gradient Boosting)

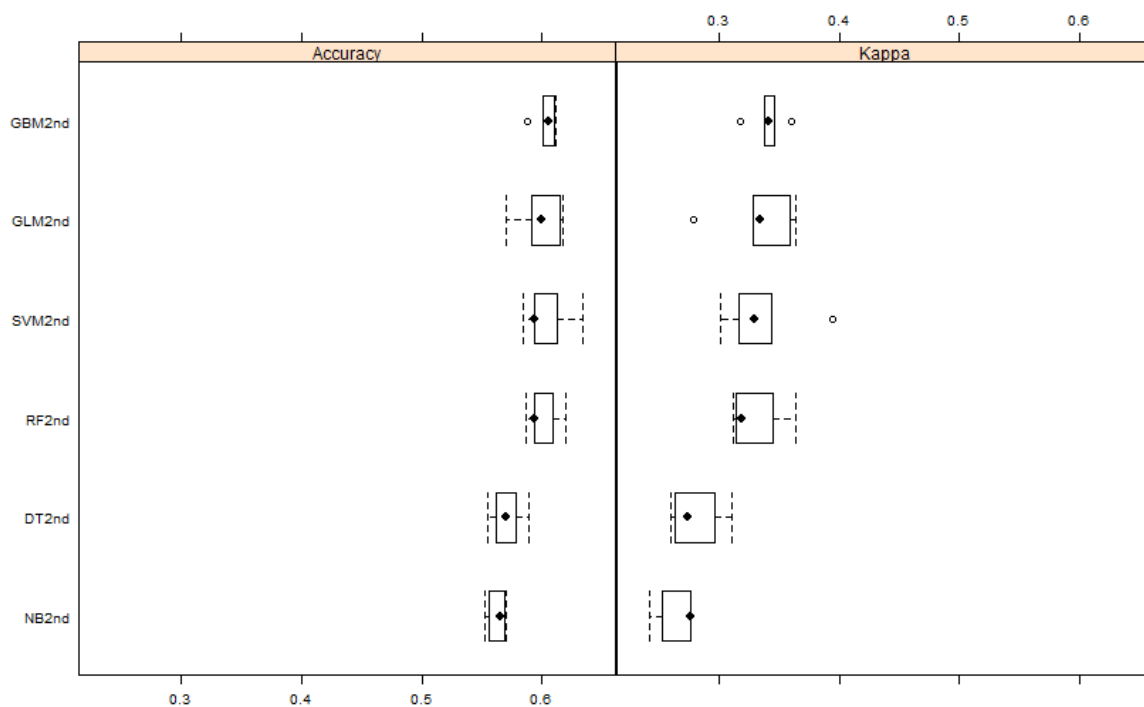
دقت



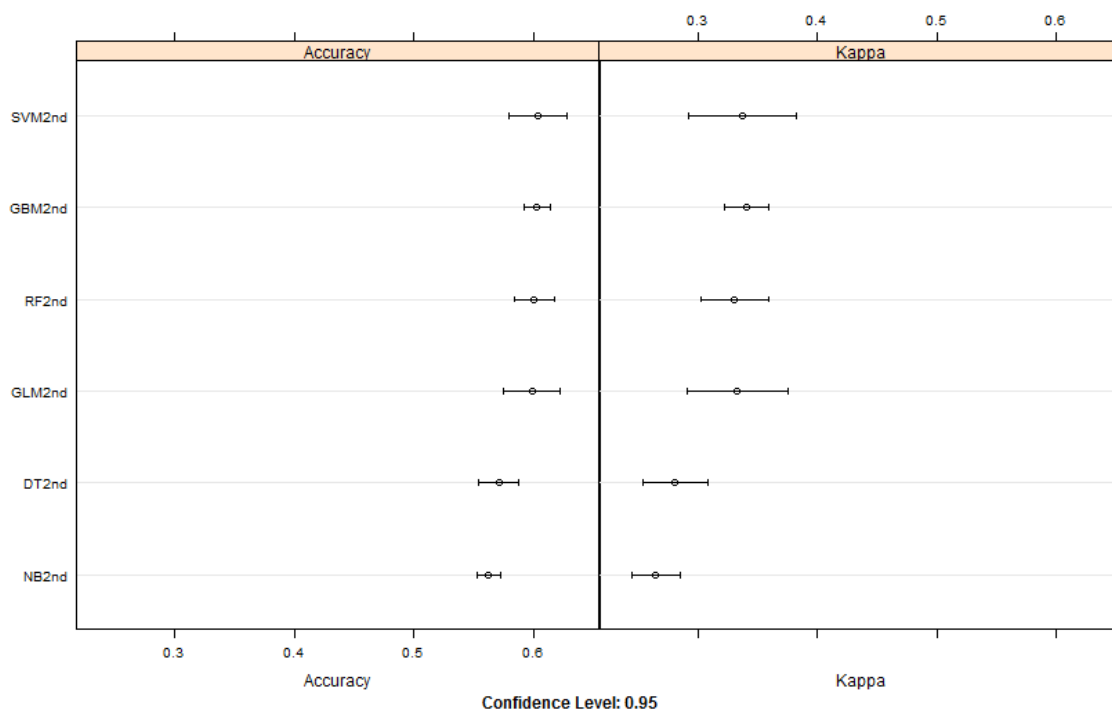
کاپا



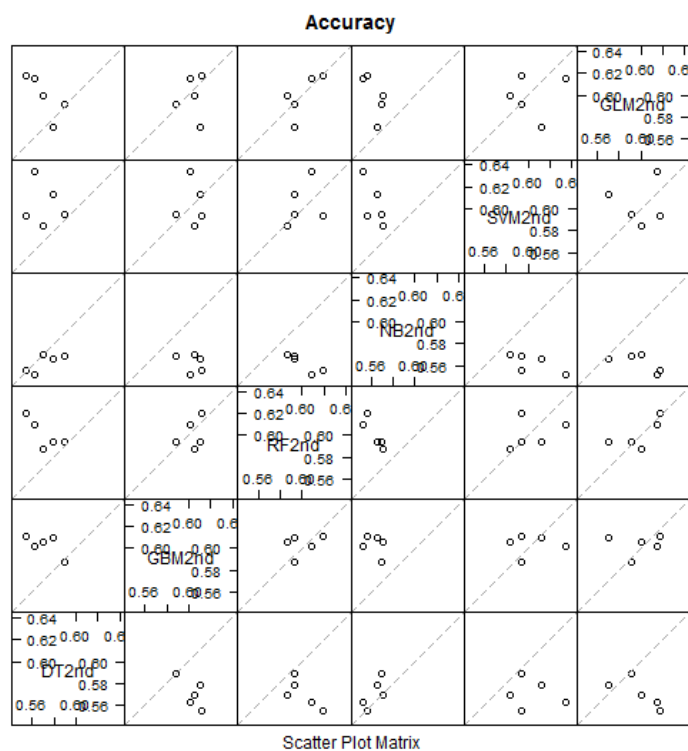
نمودار BW برای همه مدل ها که با روش نمونه گیری مجدد ترکیب شده اند.



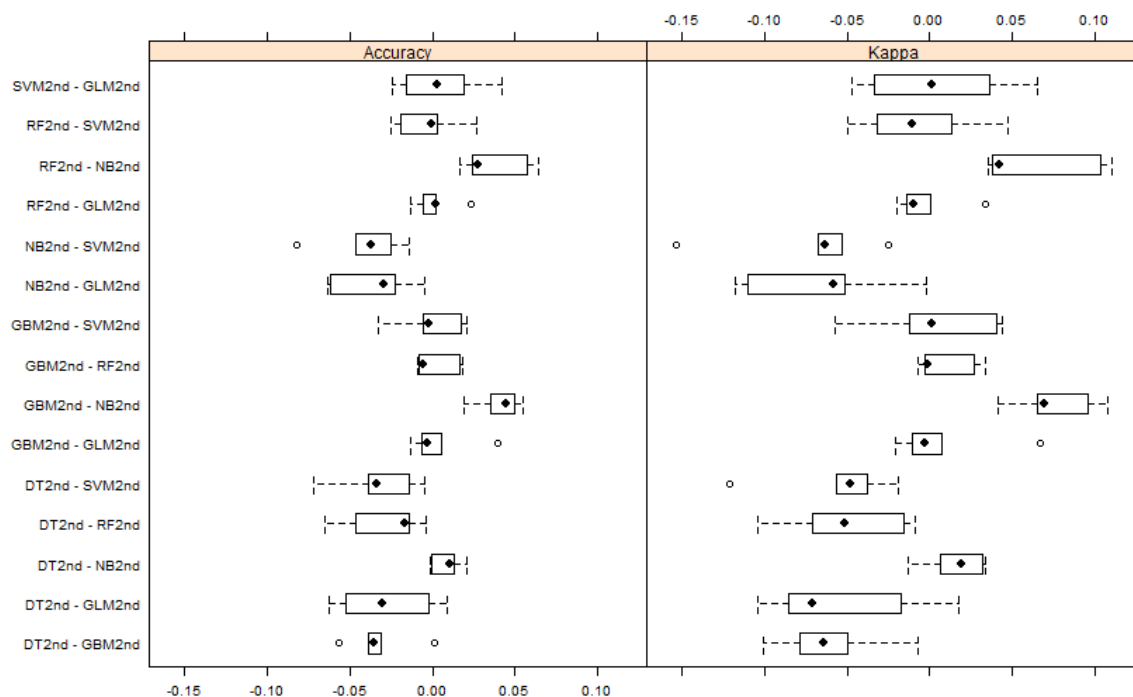
نمودار Dotplot

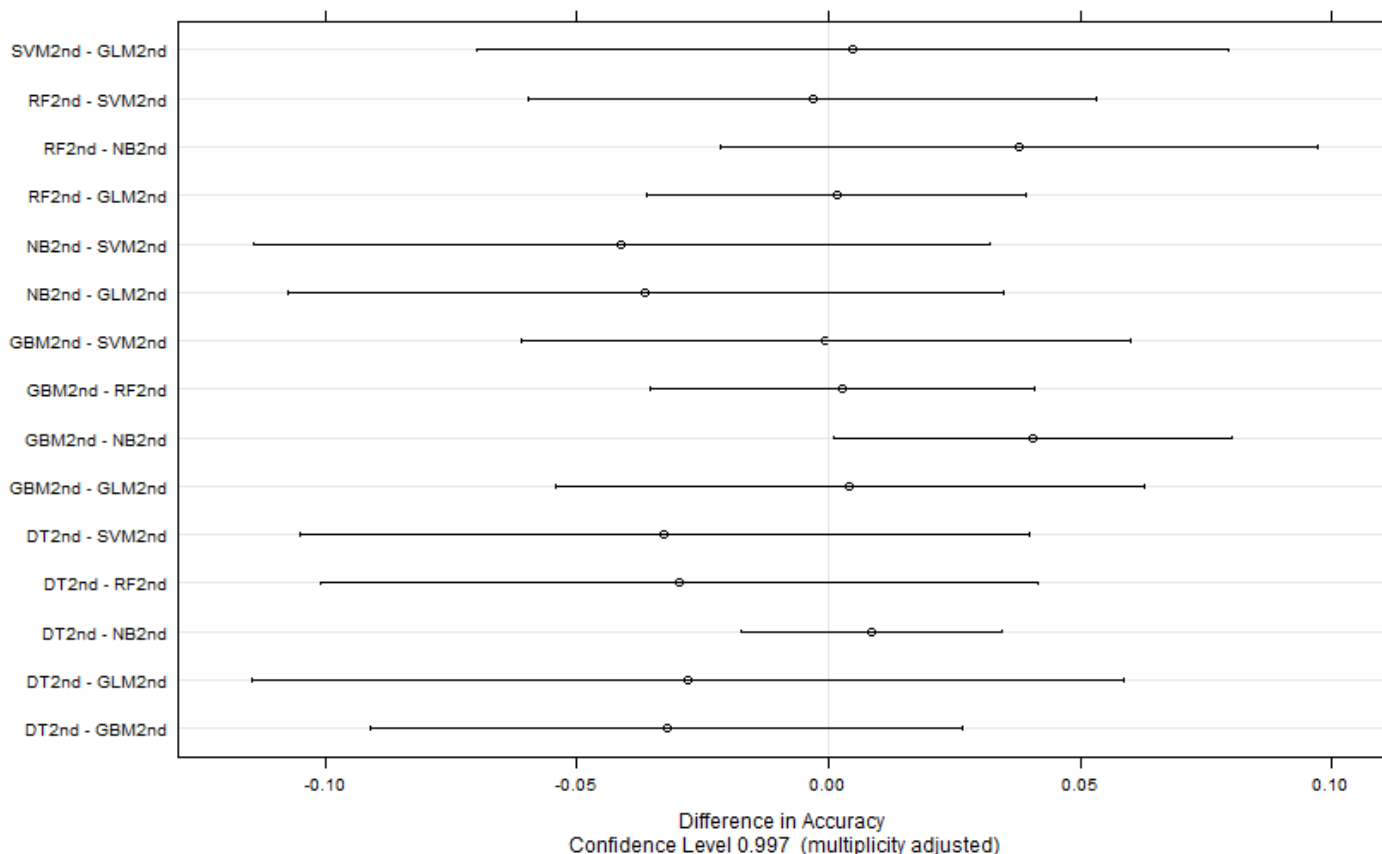


نمودار نقطه ای



مقایسه مستقیم





بر اساس نتایج مدل و مقایسه بین عملکرد مدل ها در بین یکدیگر ، مدل-SVM
 Linear به طور قابل توجهی بهتر از مدل باقیمانده عمل کرده و بنابراین برای
 آزمایش مدل نهایی مورد استفاده قرار گرفت.

SVM برای مدل دوم دقت متوسط ۶۰٫۳۷٪ و ارزش cappa 33.72٪ تولید کرد.

مجموعه تست دقت ۵۸٫۱۸٪ و کاپا ۲۹٫۹۱٪ تولید کرد.

نتایج نهایی (به طور خلاصه نشان داده شده)

توجه: جالب اینکه **SVM-Linear** برای هر دو مدل انتخاب شده است ، حتی با پیش بینی های اصلاح شده / جایگزین شده.

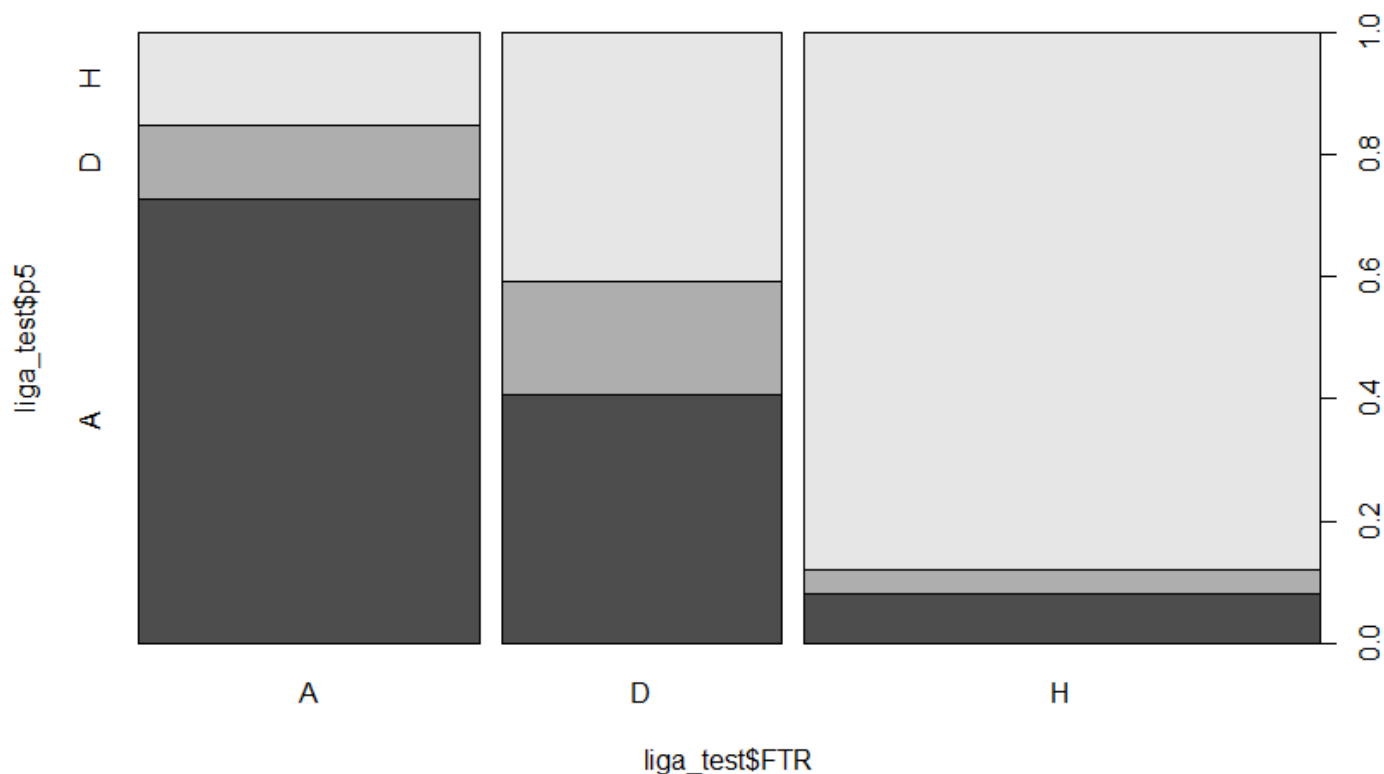
FTR p5 p5new PredictedFTRProbWhenHTRisknown.A PredictedFTRProbWhenHTRisknown.D PredictedFTRProbWhenHTRisknown.H						
1	D	D	A	0.3175107	0.38527311	0.297216146
2	A	D	A	0.3539532	0.40278740	0.243259361
3	D	D	D	0.2983670	0.49595873	0.205674279
4	D	A	H	0.6413772	0.22022013	0.138402709
5	H	H	H	0.1149321	0.29441837	0.590649528
6	A	A	A	0.9388621	0.05450384	0.006634032
PredictedFTRProbWhenHTRisUnknown.A PredictedFTRProbWhenHTRisUnknown.D PredictedFTRProbWhenHTRisUnknown.H						
1				0.3672793	0.2605115	0.37220919
2				0.3918856	0.2878935	0.32022088
3				0.3872056	0.4117744	0.20102002
4				0.2738267	0.3096866	0.41648670
5				0.3115845	0.3400149	0.34840060
6				0.8488061	0.1140005	0.03719338
B365H B365A B365D						
1	0.2777778	0.4878049	0.2857143			
2	0.1538462	0.6666667	0.2309469			
3	0.5405405	0.2222222	0.2857143			
4	0.2105263	0.5714286	0.2666667			
5	0.4347826	0.3125000	0.3030303			
6	0.4761905	0.2857143	0.2857143			

رسم مقایسه بین کلاس پیش بینی شده ما با کلاس واقعی (این به

نوعی تجسم غیر مستقیم ماتریس سردرگمی است.)

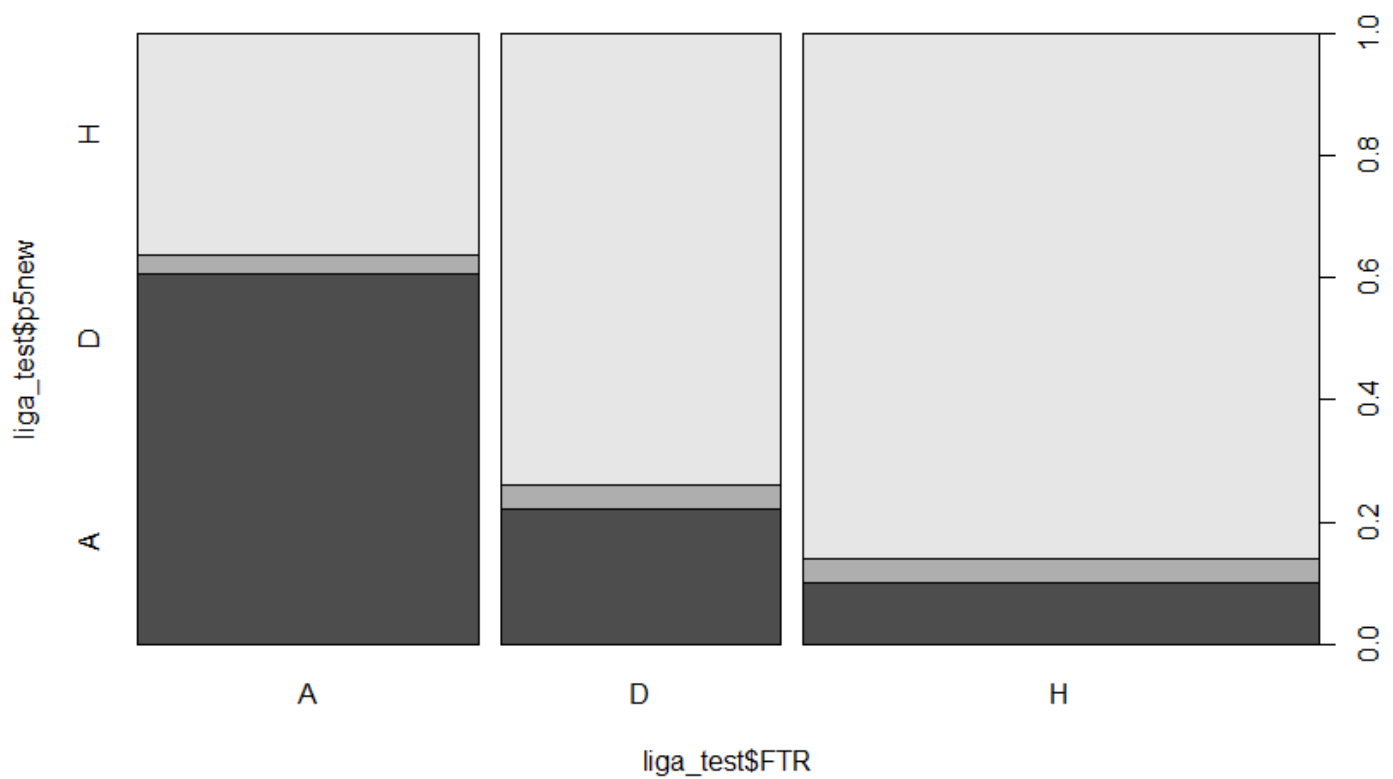
ابتدا HTR شناخته می شود. (اولین مدل ما)

- نتایج قرعه کشی باعث سردرگمی بیشتر در اینجا و حداقل برد تیم میزبان می شود.



اکنون چه زمانی HTR ناشناخته است. (مدل دوم ما)

-نتایج دور از خانه باعث سردرگمی بیشتر و حداقل برد تیم میزبان می شود.

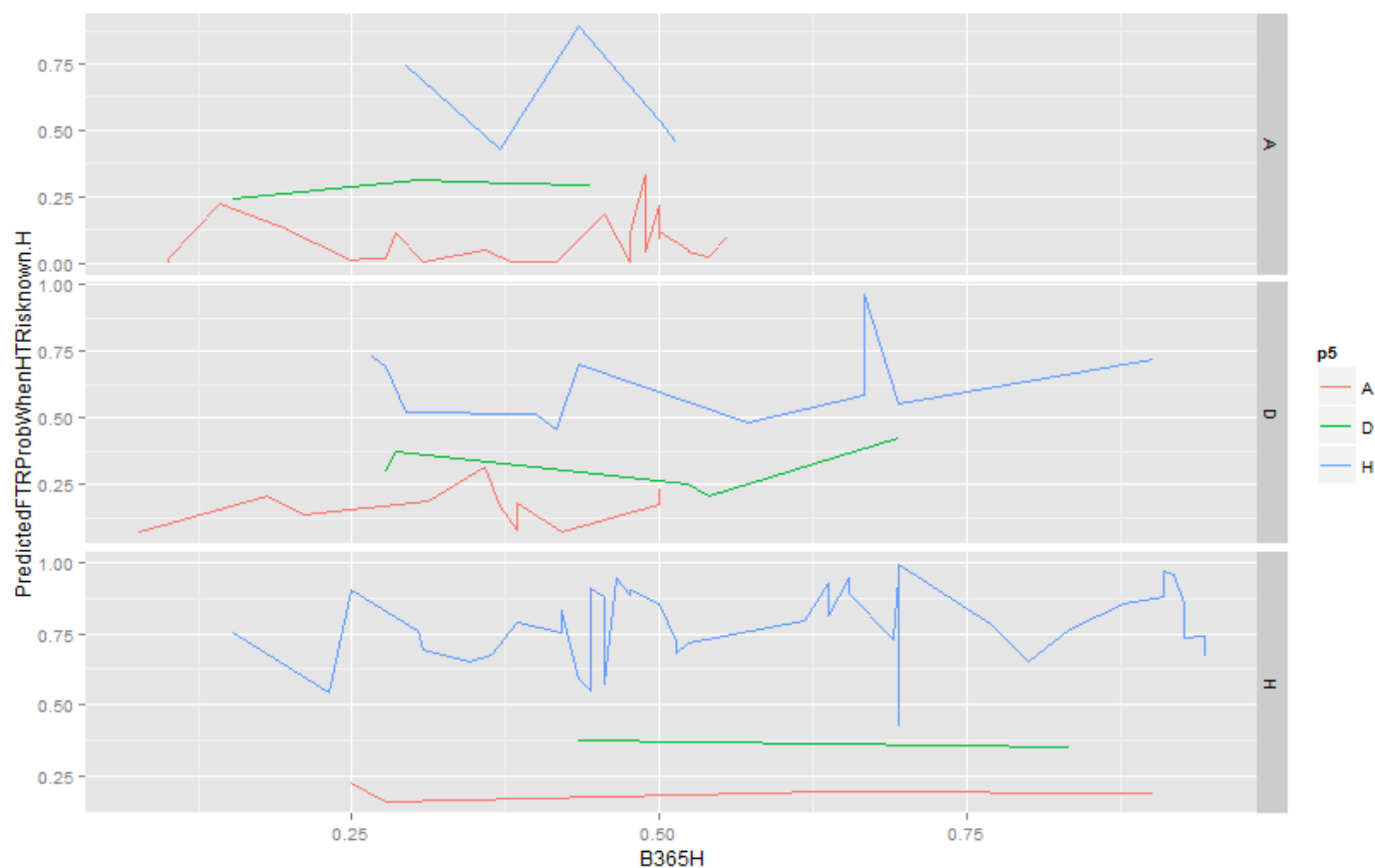


مقایسه احتمالات (پیش بینی شده در مقابل احتمالات خبرگان شرط

(بندی)

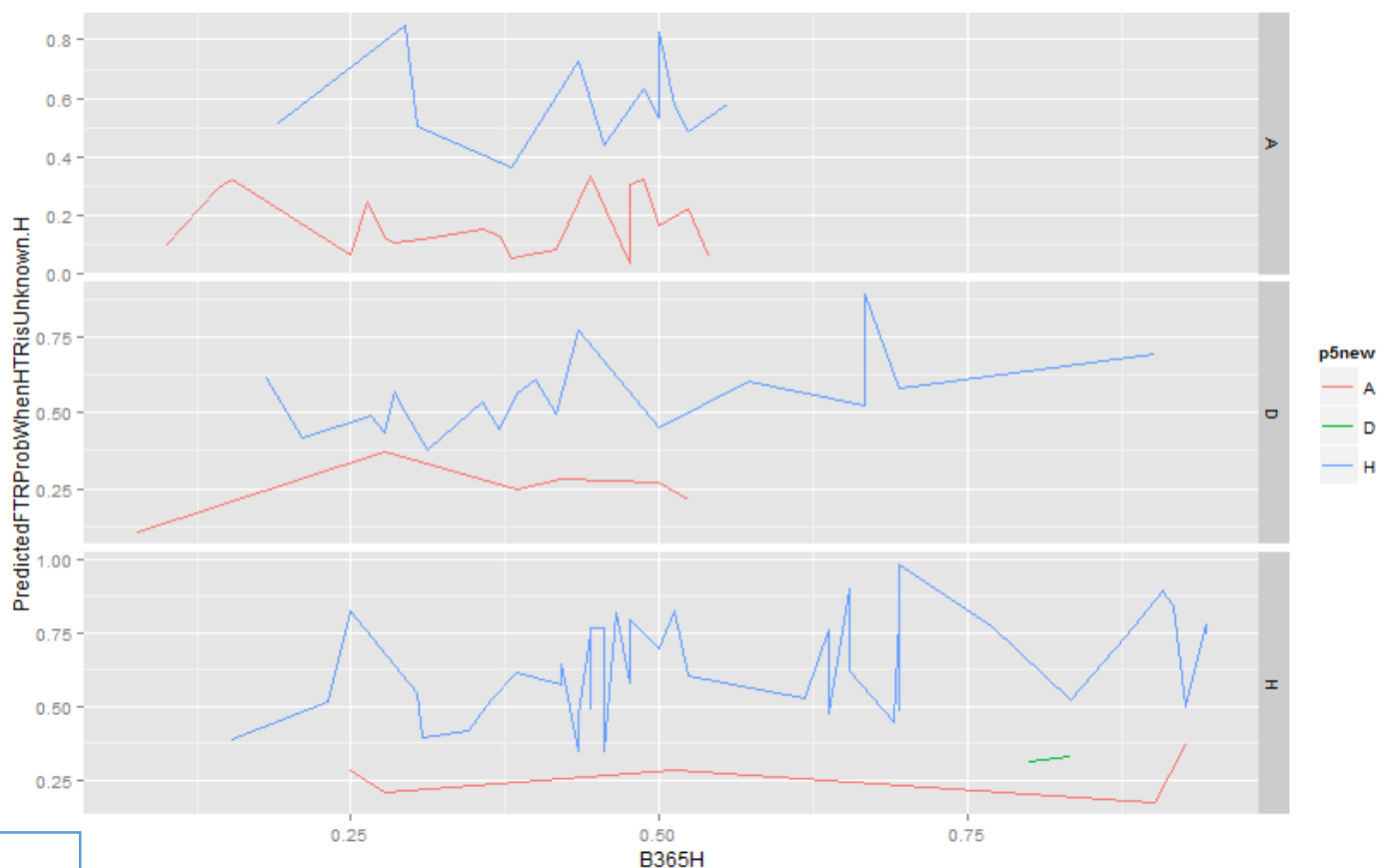
ما فقط در مورد دو مثال بحث خواهیم کرد ، یک مورد برای هر مدل:

وقتی HTR شناخته می شود. (مدل اول)



ما به شبکه "میزبان" نگاه خواهیم کرد (در پایان تیم میزبان در واقع برنده می شود) ، جایی که وقتی کارشناسان شرط بندی احتمال برد خانگی را پیش بینی می کنند ، نشان می دهد که ما برخی از نتایج را برای تیم میزبان به سود تیم خارج از خانه باختیم (همانطور که رنگ نشان می دهد) و بسیار کم به عنوان "رسم" نیز ، اما علی رغم آن ، با مقایسه اعداد احتمال ، این نشان می دهد که احتمال های پیش بینی ما بالاتر و از این رو قابل اطمینان تر هستند.

هنگامی که HTR ناشناخته است (مدل دوم)



ما به شبکه "میزبان" نگاه خواهیم کرد (در پایان تیم میزبان در واقع برنده می شود) ، جایی که وقتی کارشناسان شرط بندی احتمال برد خانگی را پیش بینی می کنند ، نشان می دهد که ما برخی از نتایج را برای تیم میزبان به سود تیم خارج از خانه باختیم (همانطور که رنگ نشان می دهد) و بسیار کم به عنوان "رسم" نیز ، اما علی رغم آن ، با مقایسه اعداد احتمال ، این نشان می دهد که احتمالات پیش بینی ما بالاتر و از این رو قابل اطمینان تر هست.

نتیجه

این ابزار باید به عنوان یکی از بسیاری از ابزارها مورد استفاده قرار گیرد ، که به طور کلی احتمالاً می تواند تصویری دقیق و قابل اعتماد در مورد بازی های فوتبال از آنچه این سیستم به تنهایی ارائه می دهد ، ارائه دهد.

این روش برای پیش بینی بازی های فوتبال می توان از آن برای پیش بینی بازی ها برای مسابقات آینده به دلیل کارکرد خوب آن استفاده کرد.

مراجعے

– آمار ریاضی و کاربردهای آن _ {جان فروند}

– داده کاوی و مفاهیم _ {ژیائوس هان}

– دانشنامه ویکی پدیا

– سایت rulesofsport.com

