



هشتمین کنفرانس بین‌المللی مهندسی
صنایع

دانشگاه صنعتی امیرکبیر، تهران

1390 و 27 بهمن 26

کاربرد داده کاوی پیش‌بینی کننده در پزشکی: مقایسه الگوهای منتخب در شناسایی
زودهنگام بیماری منژیت باکتریایی در ایران

عطیه سادات میرخانی^۱, امید پورنیک^۲, شهرام توفیقی^۳, جمال شهرابی^۴

^۱ کارشناس ارشد مهندسی پزشکی - مدیریت فن آوری اطلاعات پزشکی، دانشگاه صنعتی امیرکبیر، mirkhani_th@aut.ac.ir

^۲ گروه انفورماتیک پزشکی، دانشکده پزشکی، دانشگاه علوم پزشکی مشهد، pournik@gmail.com

^۳ عضو هیئت علمی مرکز تحقیقات سلامت، دانشگاه علوم پزشکی بقیه الله، shr_tofiqhi@yahoo.com

^۴ عضو هیئت علمی گروه مهندسی صنایع، دانشگاه صنعتی امیرکبیر، jamalshahrabi@gmail.com

چکیده

کاربرد داده کاوی پیش‌بینی کننده در پزشکی، در واقع استفاده از مدل‌های یادگیرنده جهت پیش‌بینی وضعیت سلامت بیماران است که با استفاده از ابزارها و روش‌های تحلیلی تفسیر خودکار داده‌های بیماران از تصمیم گیری‌های مناسب در مراقبت سلامت پشتیبانی می‌کند. هدف داده کاوی پیش‌بینی کننده در پزشکی ایجاد مدل‌هایی است که می‌توانند از اطلاعات ویژه بیماران، نتیجه جالبی را پیش‌بینی کرده که در خصوص شناسایی و گزارش دهی dataacademy.ir سریع بیماری‌های عفونی و مسری نظیر منژیت باکتریایی، پیش‌گیری از گسترش اپیدمی‌های خطرناک و مرگ و میر حاصل از آن بسیار حائز اهمیت می‌باشد. داده‌های مورد استفاده جهت ایجاد مدل‌های پیش‌بینی کننده در شناسایی زودهنگام بیماری منژیت باکتریایی مربوط به بانک اطلاعاتی بیماران منژیتی سال‌های 1388 و 1389 مرکز مدیریت بیماری‌های وزارت بهداشت درمان و آموزش پزشکی می‌باشد که با استفاده از تکنیک‌های داده کاوی Naïve Bayes SOTA دو مدل پیش‌بینی کننده برتر ساخته شد. داده‌ها با دقت پیش فرض 70/30 درصد به دو دسته Train جهت ساخت مدل و Test جهت ارزیابی مدل با هریک از تکنیک‌های ذکر شده تقسیم شده و نتیجه حاصل از مقایسه ساخت و ارزیابی مدل‌ها بررسی شد که با دقت و حساسیت بالای ۹۰٪ وجود میکروارگانیسم‌های مولد بیماری منژیت را شناسایی کردند.

کلمات کلیدی

داده کاوی پیش‌بینی کننده، مدل‌های یادگیرنده، منژیت باکتریایی، مراقبت سلامت

Application of Predictive Data Mining in Medicine: Comparison Selected Patterns in Early Detection of Bacterial Meningitis in Iran

Atieh Sadat Mirkhani, Omid Pournik, Shahram Tofiqhi and Jamal Shahrabi

^۱ عطیه سادات میرخانی، تهران، میدان سپاه، کوچه احمد نیا، پلاک ۱، واحد ۱، ۰۲۱۷۷۵۰۰۱۶۵، ۰۹۱۲۴۴۵۸۶۹۵

ABSTRACT

Usage of Predictive data mining in medicine, indeed, is to use learning models for predicting of patients' health situation. It uses tools and analytical methods to interpret patients' data automatically and supports proper decision making in healthcare. The objective of predictive data mining in medicine is to construct the models, which can predict interesting outcome from specific patients' information; detects and reports rapidly outbreak infectious diseases like Bacterial Meningitis; prevents dangerous epidemic extension and mortality follow that. This study is performed on meningitis patients' data set of Iran health ministry' CDC in 2009-2010. Separated data set into train and test sets with accuracy default 70/30%, then the models was built on the train set and its quality estimate on the test set with data mining techniques like SOTA and Naïve Bayes. Comparison the finding results showed accuracy and sensitivity upper than 90%. the built models with SOTA and Naïve Bayes detect the people who have one of the bacterial meningitis with sensitivity upper 90% .

KEYWORDS

Predictive Data Mining, Learning Models, Bacterial Meningitis, HealthCare.

برای شناسایی زودهنگام و پاسخ بلاذرنگ به بیماری‌ها یک کلید در سلامت عمومی می‌باشد^[2]. هر ساله حدود 2000 مورد جدید مبتلا به منژیت از طریق مراکز بهداشتی درمانی و بیمارستان‌های سطح کشور گزارش می‌شوند. پیشگیری صحیح و به هنگام در نظام مراقبت این بیماری به گونه‌ای که پاسخگوی نیازهای پیشگیری و کنترل آن در کشور باشد، پیش از پیش احساس می‌شود بتایران به کارگیری روش‌های داده‌کاوی می‌تواند در تحلیل داده‌های نظام پایش و نظارت به منظور تشخیص زودهنگام بیماری‌ها کمک کرده و ما را به سوی برنامه ریزی، اجرا و ارزیابی برنامه‌های بلند مدت جهت پیشگیری و کنترل بیماری مانند توزیع دارو، برنامه‌های واکسیناسیون و اختصاص منابع درمانی دیگر سوق دهد^[4].

پیشینه طرح موضوع داده‌کاوی به دهه 1980 و به صورت جدی، به دهه 1990 برمی‌گردد. به مرور زمان، استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه داده‌ها به عنوان داده‌کاوی مورد توجه قرار گرفت^[6]. از اما فیا در 1996 داده‌کاوی را این گونه تعریف می‌کند: «داده‌کاوی به تحلیل مجموعه‌ای از داده‌ها یا اطلاعات به منظور شناسایی الگوهای مفید بالقوه و جدید می‌پردازد»^[7]. یکی از مهم‌ترین کاربردهای داده‌کاوی در پژوهشی، به کارگیری مدل‌های داده‌کاوی پیش‌بینی کننده^۳ است زیرا این مدل‌ها با ترکیب دو یا چند آیتم از داده‌های بیمار به عنوان ابزارهایی جهت پیش‌بینی نتایج بالینی و کمک به تصمیم‌گیری به کار می‌روند^[8]. بانک اطلاعاتی مورد استفاده در این پژوهش که تکنیک‌های داده کاوی بر روی آن اجرا شد، شامل یک جدول با 30 متغیر و 7574 سطر از داده‌های گردآوری شده سال‌های 1388 و 1389 بیماران منژیت از کل کشور می‌باشد. مالکیت این بانک مربوط به وزارت بهداشت درمان و آموزش پزشکی، معاونت بهداشت، مرکز مدیریت بیماری‌های واگیر بوده و کلیه داده‌های مربوطه از طریق فرم

1- مقدمه

توانایی پیش‌بینی وضعیت سلامت بی‌شک یکی از چالش‌های بزرگ در حوزه تحقیقات سلامت می‌باشد. پیش‌بینی‌ها ممکن است از طبقه بندی ساده بیماران بر اساس فاکتورهای نظریه‌سن، جنس، وضعیت زندگی و ... تا پیش‌بینی تاثیر یک دارو یا درمان بر زندگی بیمار را در برگیرد. مدل‌های حاصل از داده‌کاوی پیش‌بینی کننده می‌توانند از درمان‌گران در انجام وظایف تشخیصی، درمانی و پایش و نظارتی حمایت کنند^[1]. پایش و نظارت در نظام سلامت به معنی گردآوری به هنگام داده‌های بیماری و تحلیل و گزارش دهی سریع از داده‌های گردآوری شده به منظور شناسایی شیوع بیماری‌های عفونی مسری نظری منژیت باکتریایی که موضوع این پژوهش است و پیش‌گیری از گسترش اپیدمی‌های خطرناک آنها می‌باشد^[2]. بیماری منژیت^۱ یا شامه‌آماس عفونت و التهاب پرده منژ و مایع مفرزی - نخاعی می‌باشد که دور مغز و نخاع را احاطه کرده‌اند^[3]. این بیماری از جمله بیماری‌هایی است که از میزان مرگ و میر و عوارض نسبتاً بالایی برخوردار است. از آنجایی که زمان ابتلاء به بیماری اغلب در دوران کودکی است، نه تنها عوارض بیماری و معلوایت‌های ناشی از آن افراد مبتلا را سالیان متمادی درگیر می‌نماید، بلکه عواقب اقتصادی و اجتماعی نامطلوب آن جامعه را نیز متأثر می‌سازد^[4]. منژیت باکتریایی بیماری شایعی نیست ولی ترس آور است زیرا می‌تواند در مدت بیست و چهار ساعت انسان را بکشد. افرادی که از منژیت جان سالم به در می‌برند، به میزان زیادی دچار آسیب‌های برگشت ناپذیر عصبی می‌شوند^[5]. تعداد موارد ابتلاء به منژیت در دنیا سالانه ۱/۲ میلیون نفر و مرگ ناشی از آن ۱۳۵/۰۰۰ نفر برآورد می‌شود^[4]. در دنیای امروز که همواره خطر تهدید بیماری‌های عفونی ناشی از بحران‌های طبیعی یا حملات بیوتوریسم وجود دارد، آمادگی

27.51	134	ثبت	گسسته	رنگ آمیزی گرم
	1923	منفی		
99.98	1687	بلی	گسسته	تشخیص محتمل
	5787	خبر		
99.98	276	بلی	گسسته	تشخیص قطعی
	71998	خبر		
55.27	87	هموفیلوس انفلوانزا نوع b	گسسته	
	55	ناپسربا مننژیتیدیس		
	93	استرپیتوکوک پنومونی		
	894	ویروسی		
	2256	سایر موارد		
	747	نامشخص		
				تشخیص نهایی
100	1819	بلی	گسسته	آنتی بیوتیک قبل
	5655	خبر		
46,76	3496	12 حالت	گسسته	آنتی بیوتیک بعد
95.11	6780	درمان	گسسته	نتیجه درمان
	121	فوت		
100	9	بلی	گسسته	دربافت و اکسن Hib
	7465	خبر		
100	30	بلی	گسسته	دربافت و اکسن مننگوکوکی
	7444	خبر		

برای مفید بودن اهداف داده کاوی، پایگاه‌های داده به پیش dataacademy.ir پردازشی به صورت پاکسازی و انتقال داده ها نیاز دارند. از آنجا که داده کاوی اغلب بر روی داده‌هایی انجام می‌شود که برای سال ها به آنها نگاهی نشده است، بنابراین داده‌ها شامل مقادیری هستند که چندان مربوط نیستند یا به سادگی گم می‌شوند. بنابراین پس از شناسایی و ادراک داده‌ها به آماده‌سازی داده‌ها می‌پردازیم. با توجه به هدف پژوهش که شناسایی زودهنگام بیماری مننژیت باکتریایی است، متغیر هدفمان متغیر اسمی جدیدی به نام تشخیص نهایی گروه بندی شده بر حسب موارد باکتریایی (شامل موارد هموفیلوس انفلوآنزا، نایسربا مننژیتیدیس و استرپیتوکوک پنومونی) و غیر باکتریایی (شامل موارد ویروسی، نامشخص و سایر موارد) است. مقیاس و نوع متغیر هدف، نوع متاداده کاوی که در اینجا طبقه بندی^۴ است را مشخص می‌کند.

به دلیل وجود انواع مقادیر سن ابتدا بر اساس دستورالعمل سازمان جهانی بهداشت، سن‌ها گروه بندی شده [۹] و در هر یک از گروه‌های سنی مقدار متوسط هر گروه در نظر گرفته شده و در فیلد جدید به نام سن - روز قرار داده شد. همچنین به دلیل وجود متغیر سن به روز (برای نوزادان ۰-۲۹ روز)، کلیه مقادیر متغیر سن به سال نیز به معادل آنها به روز تبدیل شده و در فیلد سن - روز قرار داده می‌شود. در خصوص فیلد جنسیت در مواردی که جنسیت ذکر نشده است، بر اساس نام و نام خانوادگی افراد جنسیت مشخص می‌شود. در

الکترونیکی به مرکز مدیریت بیماری‌ها ارسال و در پایگاه داده این مرکز ذخیره می‌گردد.

با توجه به موارد گزارش‌دهی سالانه این بیماری در کشور و وجود بستر فن آوری اطلاعات در گردآوری داده‌های این بیماری، ایجاد مدل‌های پیش‌بینی کننده با استفاده از روش‌های داده‌کاوی می‌تواند به شناسایی زودهنگام این بیماری و پیشگیری از اpidemi خطرناک و کشنده آن کمک کرده و نظام سلامت کشور را جهت برنامه‌ریزی‌های بهداشتی در خصوص این بیماری یاری نماید.

2- متن

بانک اطلاعاتی مورد استفاده دارای 30 متغیر می‌باشد که تعدادی از آنها مانند نام و نام خانوادگی به دلیل رعایت حقوق محترمانگی بیماران در نظر گرفته نشد. سایر متغیرها به دو دسته متغیرها پیوسته و گسسته تقسیم می‌شوند و فراوانی و درصد قابل استفاده از آنها در این پژوهش به تفکیک طبقات هر کدام از متغیرها با استفاده از نرم افزار Clementine 12 محاسبه شده که به قرار جدول (۱) است.

جدول (1): متغیرها به تفکیک نوع، طبقه شناخته شده هر طبقه، فراوانی هر طبقه و درصد قابل استفاده

نام متغیر	نوع متغیر	طبقه شناخته شده	طبقه فراوانی هر طبقه	درصد قابل استفاده
سن (سال)	پیوسته	-	-	99.98
روز	پیوسته	-	-	19.53
جنس	گسسته	مرد	4363	99.81
		زن	3098	
شغل	گسسته	36 حالت		88.84
LP	گسسته	بلی	6486	100
		خبر	988	
ظاهر نمونه	گسسته	کدر	790	53,13
		شفاف	2757	
		خونی	94	
		نامشخص	331	
تعداد گلوبول سفید	پیوسته	-	-	45,15
درصد لنفوسیت	پیوسته	-	-	45,12
PMN	پیوسته	-	-	45,83
مقدار پروتئین	پیوسته	-	-	53,15
مقدار قند	پیوسته	-	-	53,75
CSF	گسسته	ثبت	235	46,52
		منفی	3243	
کشت خون	گسسته	ثبت	80	34,1
		منفی	2442	
لاتکس	گسسته	ثبت	6	27.14
		منفی	2023	

PMN فقط از مواردی که انجام LP بر روی آن ها صورت می گیرد حاصل می شود. مواردی که انجام LP روی آن ها صورت نگرفته است، ایجاد داده های گمشده می کنند و از آنجا که این تکنیک به موارد داده های گمشده بسیار حساس بوده و نتیجه را به شکل قابل توجهی تحت تاثیر قرار می دهد، در نتیجه فقط مواردی را که انجام LP بر روی آن ها صورت گرفته است را در نظر می گیرد.

پس از انجام فیلترینگ مجدد، رکوردهایی را که شامل متغیرهای مورد نظر می باشند، وارد مدل شده و داده ها با دقت پیش فرض 70/30 درصد به دو قسمت Train چهت ساخت مدل و Test ارزیابی مدل تقسیم می شوند. نتیجه حاصل از ساخت و ارزیابی با یکدیگر مقایسه شده و نتیجه با پیش بینی حساسیت (sensitivity) 94.7% و دقت 88.8% را به ما نشان می دهد.

جدول (2): جدول مقایسه نتیجه حاصل از ساخت و ارزیابی مدل شناسایی زودهنگام بیماری منژیت با استفاده از تکنیک SOTA

	TP	FP	TN	FN	Sensitivity	accuracy
F	720	59	68	40	0.947	
T	68	40	720	59	0.535	
overall						0.88.8

Naïve Bayes 2-2

قبل از ورود داده ها به مدل همانند مدل قبلی با استفاده از تکنیک Feature Selection فیلترینگ بر روی آنها صورت می گیرد و با توجه به متغیر هدف (تشخیص نهایی گروه بندی شده بر حسب موارد باکتریایی و غیر باکتریایی)، متغیرهایی که بیشترین ارتباط را با متغیر هدف داشته انتخاب می شوند. همچنین متغیرهای که بیشترین همبستگی با متغیر هدف داشتند را به دلیل ایجاد نتیجه مشابه حذف می کنیم. این متغیرها شامل کشت خون، CSF، رنگ آمیزی گرم، آزمایش لاتکس، تشخیص احتمالی و تشخیص قطعی می باشند. متغیرهایی که وارد مدل می شوند شامل سن، جنس، انجام آزمایش LP، ظاهر مایع مغزی نخاعی، تعداد گلbulus سفید خون، درصد لنفوسیت، درصد PMN، مقدار پروتئین و مقدار سفید خون می باشد.

با توجه به نامتعادل بودن مجموعه داده ها - وجود اقلیت تعداد موارد بیمار (کلاس اقلیت) نسبت به موارد غیر بیمار (کلاس اکثربیت) - از تکنیک SMOTE جهت قدرتمند شدن کلاس اقلیت استفاده شده است. با توجه به متغیر هدف، موارد داده های گمشده را حذف کرده و سپس رکوردهایی را که شامل متغیرهای مورد نظر می باشند، وارد مدل می شوند. در اینجا به دلیل اینکه تکنیک Naïve Bayes جزو تکنیک های تبلیغی¹¹ بوده و زمان زیادی جهت ساخت مدل می برد، به خاطر کاهش زمان، حداقل مقدار جهت رسیدن به بیشترین دقت در نظر گرفته می شود [12]. در نتیجه فقط 20 درصد از داده ها جهت Train وارد مدل می شوند. نتیجه حاصل از

مورد متغیرهای کشت مایع مغزی نخاعی یا CSF، کشت خون، آزمایش آگلوتیناسیون لاتکس و آزمایش رنگ آمیزی گرم، در مواردی که طبقه انجام نشده ذکر شده است با طبقه منفی جایگزین می شود زیرا طبقه انجام نشده با مفهوم منفی در انجام تست های آزمایشگاهی برابر است. موارد داده های گمشده⁵ در کلیه سطرهای مربوط به متغیر هدف (تشخیص نهایی بر حسب باکتریایی و غیر باکتریایی) حذف شد. در مورد متغیرهای انجام LP ، استعمال آنتی بیوتیک قبل، تشخیص محتمل، تشخیص قطعی، واکسن مننگوکوک و واکسن Hib هیچ گونه داده گمشده وجود ندارد. مقادیر داده پرت خفیف⁶ با استفاده از روش Coerce نرم افزار Clementine 12.0 با بالاترین یا پائین ترین مقدار درون محدوده 3 انحراف معیار و مقادیر داده پرت شدید⁷ با بالاترین یا پائین ترین مقدار درون محدوده 5 انحراف معیار جایگزین شدند. با توجه به هدف پژوهش و با درنظر گرفتن متغیر هدف، متغیرهای مرتب جهت ورود به مدل و انتخاب مدل برتر بررسی می شوند. روش انتخاب متغیرها با استفاده از Feature Selection صورت می گیرد. این روش بر اساس رتبه بندی⁸ متغیرها و میزان همبستگی⁹ با متغیر هدف عمل می کند [10].

برای رسیدن به هدف پژوهش مدل های مختلف را آزمایش کرده و در نهایت به دو مدل برتر که بیشترین حد تفکیک را نشان دادند رسیدیم. این دو مدل با استفاده از تکنیک های SOTA و Naïve Bayes جهت شناسایی زودهنگام بیماری می باشند.

(Self Organizing Tree SOTA 1-2 Algorithm)

قبل از ورود داده ها به مدل با استفاده از تکنیک Feature Selection فیلترینگ بر روی آنها صورت می گیرد و با توجه به متغیر هدف (تشخیص نهایی گروه بندی شده بر حسب باکتریایی و غیر باکتریایی)، متغیرهایی که بیشترین ارتباط را با متغیر هدف داشته انتخاب می شوند. همچنین متغیرهای مشابه حذف می کنیم. این متغیرها شامل کشت مایع مغزی نخاعی، رنگ آمیزی گرم، آزمایش لاتکس، تشخیص احتمالی و تشخیص قطعی می باشند. متغیرهایی که وارد مدل می شوند شامل سن، جنس، انجام آزمایش LP، ظاهر مایع مغزی نخاعی، تعداد گلbulus سفید خون، درصد لنفوسیت، درصد PMN، مقدار پروتئین و مقدار قند می باشد.

با توجه به نامتعادل بودن مجموعه داده ها - وجود اقلیت تعداد موارد بیمار (کلاس اقلیت) نسبت به موارد غیر بیمار (کلاس اکثربیت) از تکنیک¹⁰ SMOTE جهت قدرتمند شدن کلاس اقلیت استفاده شده است [11]. سپس فیلترینگ مجدد بر مبنای مثبت بودن انجام آزمایش LP صورت می گیرد زیرا متغیرهای آزمایشگاهی مانند تعداد سلول، قند، پروتئین، درصد لنفوسیت و

- focus on selected methods and applications. WIREs Data Mining and Knowledge Discovery, 2011. 1: p. 430-416
- [2] Chen, H., D. Zeng, and P. Yan, INFECTIOUS DISEASE INFORMATICS Syndromic Surveillance for Public Health and BioDefense. Vol. 1. 2009, New York: Springer.
- [3] Mandell, G.L., J.E. Bennett, and R. Dolin, Principles and Practice of Infectious Diseases, ed. six. Vol. 1. 2005, Philadelphia, Pennsylvania CHURCHILL LIVINGSTONE. [4] استقلمنی, ع. و ف. عسکری, دستورالعمل مراقبت منژیت. 1384, تهران: مرکز مدیریت بیماری های وزارت بهداشت درمان و آموزش پزشکی.
- [5] R.Murray, P., K. S.Rosenthal, and M. L.Pfaller, Medical Microbiology. 2005, Philadelphia: Mosby. [6] زرزری, ق.م, داده کاری یک ایزار آنلاین مدیریتی, ایرانیکا.
- [7] Fayyad, U. and G. piatetsky, Knowledge Discovery and Data Mining: Towards a Unifying Framework. KDD processing, .1996
- [8] Riccardo Bellazzi , B.Z., Predictive data mining in clinical medicine: Current issues and guidelines. international journal of medical informatics, 2006: p. 17
- [9] WorldHealthOrganization. Bacterial meningitis (including Haemophilus influenzae type b(Hib),Neisseria meningitis, and streptococcus pneumoniae). .2001
- [10] Bontempi, G. and B. Haibe-Kains, Feature selection methods for mining bioinformatics data. Machine Learning Group Departement d'Informatique ULB, Université Libre de Bruxelles Boulevard de Triomphe.
- [11] Chawla, N.V. and K.W. Bowyer, SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 2002. 16: p. 321–357
- [12] Lowd, D. and P. Domingos, Naive Bayes Models for Probability Estimation. Department of Computer Science and Engineering, University of Washington, Seattle, USA.

زیرنویس‌ها

- ¹
- Meningitis
-
- ²
- Osama Fayyad
-
- ³
- Predictive Data Mining Models
-
- ⁴
- Classification
-
- ⁵
- Missing values
-
- ⁶
- outliers
-
- ⁷
- extreme
-
- ⁸
- Ranking
-
- ⁹
- Correlation
-
- ¹⁰
- Synthetic Minority Over-sampling Technique
-
- ¹¹
- Laziness

و Test با هم مقایسه شده و نتیجه با پیش‌بینی حساسیت و دقیقی (%) ۹۹/۷ و دقیقی (%) ۹۳/۱ را به ما نشان می‌دهد.

جدول (3) مقایسه نتیجه حاصل از ساخت و ارزیابی مدل شناسایی زودهنگام بیماری منژیت با استفاده از تکنیک Naïve Bayes

	TP	FP	TN	FN	Sensitivity	accuracy
F	15123	1087	44	42	0.997	
T	44	42	15123	1087	0.039	
overall						0.931

3- نتیجه و جمع بندی

پژوهش حاضر با هدف بررسی کاربرد مدل‌های پیش‌بینی کننده در پزشکی به منظور شناسایی زودهنگام بیماری منژیت باکتریایی حاد در کشور ایران می‌باشد که با استفاده از عناصر داده‌ای متناسب با اهداف پایش و نظارتی بیماری منژیت مربوط به سال‌های ۱۳۸۸ و ۱۳۸۹ که از طریق فرم‌های الکترونیکی از سراسر کشور گردآوری شده و در پایگاه داده مرکز مدیریت بیماری‌های وزارت بهداشت درمان و آموزش پزشکی ذخیره می‌شود، انجام شد.

در رسیدن به این هدف، مدل‌سازی بر اساس متغیر هدف تشخیص نهایی گروه‌بندی شده بر حسب موارد باکتریایی و غیرباکتریایی انجام شد. هدف از ساخت دو مدل پیش‌بینی کننده با تکنیک‌های Naïve Bayes و SOTA به منظور شناسایی زودهنگام بیماری منژیت، ایجاد مدل‌هایی حساس یا Sensitive می‌باشد تا بتواند افرادی را که هر کدام از سه میکروارگانیسم False باکتریایی را دارند شناسایی کند. بنابراین تا حد امکان Negative آن باید پایین باشد تا اگر اعلام کرد که هر یک از این سه میکروارگانیسم هموفیلوس آنفلوانزا نوع b نایسیریامنژیتیدیس و استرپتوکوک پنومونی را ندارد اشتباه تکرده باشد. بنابراین همانطور که در جداول 2 و 3 مشاهده می‌شود، جداول Score کمترین مقادیر و False Negative و حساسیت بالای ۹۰٪ را نشان می‌دهند. یعنی در مدل ساخته شده با تکنیک SOTA با حساسیت ۹۴٪ و در مدل ساخته شده با تکنیک Naïve Bayes با حساسیت ۹۹.۷٪ می‌توان اعلام کرد که یکی از این سه میکروارگانیسم که مولد اپیدمی‌های خطرناک در جامعه می‌باشند شناسایی شده‌اند. کاربرد این مدل‌ها در سیستم‌های پایش و نظارتی در مدیریت سلامت می‌باشد زیرا به محض شناسایی هر کدام از این سه میکروارگانیسم باکتریایی پیشگیری دارویی درخصوص فرد بیمار باید آغاز گردد و از خطرگسترش اپیدمی در جامعه و مرگ و میر بیماران در پی آن جلوگیری شود.

4- مراجع

- [1] Bellazzi, R., F. Ferrazzi, and L. Sacchi, Predictive data mining in clinical medicine: a focus on selectedmedicine: a