تشخیص زودهنگام مننژیت باکتریایی با استفاده از Medical Calculator

عطیه‌سادات میرخانی[1]
دانشگاه صنعتی امیرکبیر
Mirkhani_th@aut.ac.ir

دکتر امید پورنیک
دانشگاه علوم پزشکی مشهد
Pourniko۸۸۱@mums.ac.ir

فرزاد کاوه
مرکز مدیریت بیماری های وزارت بهداشت درمان و آموزش پزشکی
fkaveh۴۴@yahoo.com

دکتر محمود نبوی
معاون مرکز مدیریت بیماری های واگیر وزارت بهداشت درمان و آموزش پزشکی
mahmoodnabavi۵۳@yahoo.com

دکتر شهرام توفیقی
دانشگاه علوم پزشکی بقیه الله (عج)
Shr_tofighi@yahoo.com

نتایج دومین بررسی سالانه جامعه سیستم های مدیریت و اطلاعات مراقبت سلامت که در سال ۲۰۱۲ منتشر شد، افزایش بی سابقه استفاده پزشکان و متخصصان مراقبت سلامت از تکنولوژی موبایل را نشان داد. در این میان کاربرد نرم افزارهای Medical Calculator که از طریق تجهیزات ارتباطی موبایل مانند تلفن همراه، تبلت، PDA ... در دریافت خدمات و اطلاعات سلامت به پزشکان و درمانگران کمک می کند، افزایش چشم گیری داشته است. این مقاله به کاربرد الگوریتم Naïve Bayes در تشخیص زودهنگام مننژیت باکتریایی که می تواند در قالب یک Medical calculator قابل استفاده بر روی تلفن همراه یا تبلت یا PDA به کار رود، می پردازد. این الگوریتم بیماری را با sensitivity ۹۹/۷٪ و accuracy ۹۳/۱٪ پیش بینی کرده و از این طریق به تشخیص زودهنگام بیماری پیش از انجام کشت های زمان بر و هزینه بر تشخیصی می پردازد. تشخیص زودهنگام این بیماری از طریق این الگوریتم علاوه بر ایجاد صرفه جویی اقتصادی در نظام مراقبت و سلامت از مرگ و میر و عوارض ناشی از تاخیر در تشخیص بیماری، مقاومت آنتی بیوتیکی و در نهایت پیشگیری از اپیدمی شدن آن کمک می کند.

واژگان کلیدی:

مننژیت باکتریایی، Medical Calculator، Naïve Bayes، تشخیص زودهنگام، نظام مراقبت

[1] عطیه سادات میرخانی – mirkhani_th@aut.ac.ir

[Downloaded from journals.tums.ac.ir on 2024-10-19]

۱- مقدمه

توانایی پیش بینی وضعیت سلامت بی شک یکی از چالش‌های بزرگ در حوزه تحقیقات سلامت می‌باشد. پیش‌بینی‌ها ممکن است از طبقه بندی ساده بیماران بر اساس فاکتورهایی نظیر سن، جنس، وضعیت زندگی و ... تا پیش بینی تاثیر یک دارو یا درمان بر زندگی بیمار را در برگیرد. مدل‌های حاصل از داده‌کاوی پیش‌بینی کننده می‌توانند در ترکیب با تکنولوژی‌های نوین ICT از درمان‌گران در انجام وظایف تشخیصی، درمانی و پایش و نظارتی حمایت کنند[۱]. پایش و نظارت در نظام سلامت به معنی گردآوری به هنگام داده‌های بیماری و تحلیل و گزارش‌دهی سریع از داده‌های گردآوری شده به منظور شناسایی شیوع بیماری‌های عفونی مسری نظیر منتژیت باکتریایی و پیش‌گیری از گسترش اپیدمی‌های خطرناک آنها می باشد[۲].بیماری منتژیت[۱] یا شامه‌آماس عفونت و التهاب پرده منتز و مایع مغزی – نخاعی می‌باشد که دور مغز و نخاع را احاطه کرده اند[۳]. این بیماری از جمله بیماری‌هایی است که از میزان مرگ و میر و عوارض نسبتا بالایی برخوردار است. از آنجایی که زمان ابتلاء به بیماری اغلب در دوران کودکی است، نه تنها عوارض بیماری و معلولیت‌های ناشی از آن افراد مبتلا را سالیان متمادی درگیر می‌نماید، بلکه عواقب اقتصادی و اجتماعی نامطلوب آن جامعه را نیز متاثر می‌سازد[۴]. منتژیت باکتریایی بیماری شایعی نیست ولی ترس‌آور است زیرا می‌تواند در مدت بیست و چهار ساعت انسان را بکشد. افرادی که از منتژیت جان سالم به در می‌برند، به میزان زیادی دچار آسیب‌های برگشت ناپذیر عصبی می‌شوند[۵].تعداد موارد ابتلاء به منتژیت در دنیا سالانه ۱/۲ میلیون نفر و مرگ ناشی از آن ۱۳۵/۰۰۰ نفر برآورد می‌شود[۴].در دنیای امروز که همواره خطر تهدید بیماری‌های عفونی ناشی از بحران‌های طبیعی وجود دارد، آمادگی برای شناسایی زود هنگام و پاسخ بلادرنگ به بیماری‌ها یک کلید در سلامت عمومی می‌باشد[۲]. هر ساله حدود ۲۰۰۰ مورد جدید مبتلا به منتژیت 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، بیش از پیش احساس می‌شود بنابراین به کارگیری روش‌های داده‌کاوی و ایجاد مدل های پیش بینی کننده 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 داروی مناسب و اختصاص منابع درمانی دیگر سوق دهد[۴].

پیشینه طرح موضوع داده‌کاوی به دهه ۱۹۸۰ و به صورت جدی، به دهه ۱۹۹۰ برمی‌گردد. به مرور زمان، استخراج و کشف سریع و دقیق اطلاعات با ارزش و پنهان از پایگاه داده‌ها به‌عنوان داده‌کاوی مورد توجه قرار گرفت[۶]. ازاما فیاد[۲] در ۱۹۹۶ داده‌کاوی را این گونه تعریف می‌کند: « داده‌کاوی به تحلیل مجموعه‌ای از داده‌ها یا اطلاعات به منظور شناسایی الگوهای مفید بالقوه و جدید می‌پردازد»[۷]. یکی از مهم‌ترین کاربردهای داده‌کاوی در پزشکی، 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[۸].

از آنجا که این بیماری عفونی یکی از اولویت‌های WHO[۴] و وزارت بهداشت درمان و آموزش پزشکی است، همچنین با توجه به موارد گزارش‌دهی سالانه این بیماری در کشور و وجود بستر فن آوری اطلاعات در گردآوری داده‌های این بیماری، این پژوهش به منظور ایجاد مدل‌های پیش‌بینی کننده با استفاده از روش‌های داده‌کاوی در شناسایی زودهنگام این بیماری و پیشگیری از اپیدمی خطرناک و کشنده آن انجام گرفت و انتظار می رود استفاده از این مدل‌ها نظام سلامت کشور را جهت برنامه‌ریزی‌های بهداشتی در خصوص این بیماری یاری نماید.

۲- روش کار

بانک اطلاعاتی مورد استفاده در این پژوهش که تکنیک‌های داده کاوی بر روی آن اجرا شد، شامل یک جدول با ۳۰ متغیر و ۷۴۷۵ سطر از داده‌های گردآوری شده سال‌های ۱۳۸۸ و ۱۳۸۹ بیماران مشکوک به منتژیت باکتریال 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 الکترونیکی از مراکز بهداشت در سراسر کشور به مرکز مدیریت بیماری‌ها ارسال و در پایگاه داده این مرکز ذخیره می‌گردد. متغیرهای این بانک به جز نام و نام خانوادگی به دلیل رعایت اصل محرمانگی، و تاریخ بروز علائم و تاریخ بستری شدن بیمار به دلیل عدم کاربرد در پژوهش حذف شدند مابقی در جدول شماره ۱ آمده است :

---

[۱] Meningitis

[۲] Osama Fayyad

[۳] Predictive Data Mining Models

[۴] World Health Organization

[Downloaded from journals.tums.ac.ir on 2024-10-19]

جدول ۱متغیرهای بانک اطلاعاتی نظام مراقبت بیماری مننژیت باکتریال

نام متغیر	نام متغیر
سن (سال)	مقدار پروتئین مایع مغزی نخاعی
روز	مقدار قند مایع مغزی نخاعی
جنس	کشت CSF (مایع مغزی نخاعی)
شغل	کشت خون
استان	لاتکس
دانشگاه	رنگ آمیزی گرم
شهر	تشخیص محتمل
منطقه	تشخیص قطعی
انجام LP	تشخیص نهایی
ظاهر نمونه	آنتی بیوتیک قبل
تعداد گلبول سفید	آنتی بیوتیک بعد
درصد لنفورسیت گلبول سفید	نتیجه درمان
درصد PMN[1] گلبول سفید	دریافت واکسن Hib

برای مفید بودن اهداف داده کاوی، پایگاه‌های داده به پیش 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، نایسریا مننژیتدیس و استپتوکوک پنمونی) و غیر باکتریایی (شامل موارد ویروسی، نامشخص و سایر موارد) است. مقیاس و نوع متغیر هدف، نوع متد داده‌کاوی که در این‌جا طبقه بندی[2] 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 نام‌خانوادگی افراد جنسیت مشخص می‌شود. در 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 تست‌های آزمایشگاهی برابر است. موارد داده‌های گمشده[3] 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 خفیف[4] با استفاده از روش Coerce نرم افزار Clementine ۱۲,۰ با بالاترین یا پائین ترین مقدار درون محدوده ۳ انحراف معیار و مقادیر داده پرت شدید[5] با بالاترین یا پایین ترین مقدار درون محدوده ۵ انحراف معیار جایگزین شدند. با توجه به هدف پژوهش و با درنظر گرفتن متغیر هدف، متغیرهای مرتبط جهت ورود به مدل و انتخاب مدل برتر بررسی می‌شوند. روش انتخاب متغیرها با استفاده از روش Feature Selection صورت می‌گیرد. این روش بر اساس رتبه بندی[6] متغیرها و میزان هم‌بستگی[7] با متغیر هدف عمل می‌کند[۱۰]. متغیرهای منتخب برای ورود به مدل در جدول شماره ۲ آورده شده است:

جدول ۲متغیرهای منتخب با استفاده از روش feature selection برای ورود به مدل

نام متغیر	نام متغیر
سن بر حسب روز	مقدار قند
تعداد گلبول سفید	جنسیت بیمار
درصد لنفوسیت	انجام LP
درصد PMN	ظاهر نمونه
	مقدار پروتئین

برای رسیدن به هدف پژوهش مدل‌های مختف را آزمایش کرده و در نهایت به مدل برتری که بیشترین حد تفکیک را نشان دادند رسیدیم. این مدل با استفاده از تکنیک Naïve Bayes جهت شناسایی زودهنگام بیماری می‌باشند.

قبل از ورود داده‌ها به مدل با استفاده از تکنیک Feature Selection فیلترینگ بر روی آنها صورت می‌گیرد و با توجه به متغیر هدف (تشخیص نهایی گروه‌بندی شده بر حسب موارد باکتریایی و غیر باکتریایی)، متغیرهایی که بیشترین ارتباط را با متغیر هدف داشته انتخاب می شوند. همچنین متغیرهایی که بیشترین هم‌بستگی با متغیر هدف داشتند را به دلیل ایجاد نتیجه مشابه حذف می‌کنیم. این متغیرها شامل کشت CSF، کشت خون، رنگ آمیزی گرم، آزمایش لاتکس، تشخیص احتمالی و تشخیص

[1] polymorphonuclear
[2] Classification
[3] Missing values
[4] outliers
[5] extreme
[6] Ranking
[7] Correlation

[Downloaded from journals.tums.ac.ir on 2024-10-19]

قطعی می‌باشند. متغیرهایی که وارد مدل می شوند شامل سن، جنس، انجام آزمایش LP، ظاهر مایع مغزی نخاعی، تعداد گلبول سفید خون، درصد لنفوسیت، درصد PMN، مقدار پروتئین و مقدار قند می‌باشد.

با توجه به نامتعادل بودن مجموعه داده ها - وجود اقلیت تعداد موارد بیمار (کلاس اقلیت) نسبت به موارد غیر بیمار(کلاس اکثریت)- از تکنیک SMOTE جهت قدرتمند شدن کلاس اقلیت استفاده شده است. با توجه به متغیر هدف، موارد داده‌های گمشده را حذف کرده و سپس رکوردهایی را که شامل متغیرهای مورد نظر می‌باشند، وارد مدل می‌شوند. در اینجا به دلیل اینکه تکنیک Naïve Bayes جزو تکنیک های تنبل[1] بوده و زمان زیادی جهت ساخت مدل می‌برد، به خاطر کاهش زمان، حداقل مقدار جهت رسیدن به بیشترین دقت در نظر گرفته می‌شود[۱۲].در نتیجه فقط ۲۰ درصد از داده ها جهت Train وارد مدل می‌شوند. نتیجه حاصل از Train و Test با هم مقایسه شده و نتیجه با پیش‌بینی حساسیت (sensitivity) ۹۹/۷ ٪ و دقت ۹۳/۱٪را به ما نشان می‌دهد.

Naïve Bayes Result	Sensitivity	accuracy
	۰,۹۹۷	۰,۹۳۱

جدول ۳ جدول مقایسه نتیجه حاصل از ساخت و ارزیابی مدل شناسایی زودهنگام بیماری مننژیت با استفاده از تکنیک Naïve Bayes

۳- نتیجه و جمع بندی

پژوهش حاضر با هدف بررسی کاربرد مدل‌های پیش‌بینی کننده در پزشکی به منظور شناسایی زودهنگام بیماری مننژیت باکتریایی حاد در کشور ایران می باشد که با استفاده از عناصر داده‌ای متناسب با اهداف پایش و نظارتی بیماری مننژیت مربوط به سال های ۱۳۸۸ و ۱۳۸۹ که از طریق فرم های 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 مدل پیش‌بینی کننده با تکنیک Naïve Bayes ، ایجاد مدل‌هایی حساس یا Sensitive می‌باشد تا بتواند افرادی را که هر کدام از سه میکروارگانیسم باکتریایی را دارند شناسایی کند. بنابراین تا حد امکان False Negative آن باید پایین باشد تا اگر اعلام کرد که هر یک از این سه میکروارگانیسم هموفیلوس آنفلوانزا نوع b، نایسریامننژیتیدیس و استرپتوکوک پنمونی را ندارد اشتباه نکرده باشد. بنابراین همانطور که در جدول شماره ۳ مشاهده می‌شود، این مدل با حساسیت ۹۹,۷٪ می‌تواند اعلام کند که یکی از این سه میکروارگانیسم که مولد اپیدمی‌های خطرناک در جامعه می‌باشند شناسایی شده‌اند. این مدل در قالب یک Medical calculator بر روی گوشی‌های تلفن همراه یا دیگر تجهیزات الکترونیکی کمک بزرگی به پزشک در امر تشخیص می کند از آنجا که انجام تست‌های آزمایشگاهی مانند کشت CSF، لاتکس ویا انجام CT Scan زمان بر و هزینه بر بوده[۱۳] و از طرفی امکان ایجاد آزمایشگاه تشخیصی و نیروی انسانی مرتبط با آن در همه روستاها یا خانه‌های بهداشت وجود ندارد و هم چنین چه بسا تاخیر در امر تشخیص سبب ایجاد عوارض مانند معلولیت‌ها و گاها مرگ شود، پس استفاده از این مدل‌ها علاوه بر تشخیص زودهنگام بیماری، به غربالگری بیماران مننژیت باکتریایی از غیر باکتریایی، ایجاد برنامه درمانی و پیشگیری از مقاومت آنتی بیوتیکی کمک بزرگی خواهد کرد.

[1] Laziness

۴- مراجع:

۱. Bellazzi, R., F. Ferrazzi, and L. Sacchi, *Predictive data mining in clinical medicine: a focus on selectedmedicine: a focus on selected methods and applications.* WIREs Data Mining and Knowledge Discovery, ۲۰۱۱. ۱: p. -٤١٦ .٤٣٠

۲. Chen, H., D. Zeng, and P. Yan, *INFECTIOUS DISEASE INFORMATICS Syndromic Surveillance for Public Health and BioDefense*. Vol. ۱. ۲۰۰۹, New York: Springer.

۳. Mandell, G.L., J.E. Bennett, and R. Dolin, *Principles and Practice of Infectious Diseases*, ed. six. Vol. ۱. ۲۰۰۵, Philadelphia, Pennsylvania CHURCHILL LIVINGSTONE.

٤. استقامتی, ع. و ف. عسکری, دستورالعمل مراقبت مننژیت. ۱۳۸۴, تهران: مرکز مدیریت بیماری های وزارت بهداشت درمان و آموزش پزشکی.

۵. R.Murray, P., K. S.Rosenthal, and M. L.Pfaller, *Medical Microbiology*. ۲۰۰۵, Philadelphia: Mosby.

٦. زرزری, ق.م., داده کاوی یک ابزار آنالیز مدیریتی, ایرانیکا.

۷. Fayyad, U. and G. piatetsky, *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. KDD processing, ۱۹۹٦.

۸. Riccardo Bellazzi , B.Z., *Predictive data mining in clinical medicine: Current issues and guidelines.* international journal of medical informatics, ۲۰۰٦: p. .۱۷

۹. WorldHealthOrganization. *Bacterial meningitis (including Haemophilus influanzae type b(Hib),Neisseria meningitis, and streptoccus pneumoniae)*. .۲۰۰۱

۵

۱۰. Bontempi, G. and B. Haibe-Kains, *Feature selection methods for mining bioinformatics*

[Downloaded from journals.tums.ac.ir on 2024-10-19]

(including Haemophilus influanzae type b(Hib),Neisseria meningitis, and streptoccus pneumoniae). 2001; Available from:

data. Machine Learning Group Departement d'Informatique ULB, Université Libre de Bruxelles Boulevard de Triomphe.

.۱۱ Chawla, N.V. and K.W. Bowyer, *SMOTE: Synthetic Minority Over-sampling Technique.* Journal of Artificial Intelligence Research, ۲۰۰۲. ۱٦: p. ۳۲۱–.۳۵۷

.۱۲ Lowd, D. and P. Domingos, *Naive Bayes Models for Probability Estimation.* Department of Computer Science and Engineering, University of Washington, Seattle, USA.

.۱۳WorldHealthOrganization. Bacterial meningitis

[Downloaded from journals.tums.ac.ir on 2024-10-19]

[Downloaded from journals.tums.ac.ir on 2024-10-19]