# داده‌کاوی بیماران افسرده در جهت بهبود و بررسی ارتباط آن با موسیقی

**مهتاب جمالی،* زهره فرجی،** محمد ربیعی*****

**تاریخ دریافت:** ۱۴۰۱/۸/۲ **تاریخ پذیرش:** ۱۴۰۱/۸/۲۸ **نوع مقاله:** پژوهشی

**چکیده**

جمع‌آوری داده‌های بیماری‌ها، به جهت شناسایی و درمان آنها از اهمیت زیادی برخوردار است. به منظور کشف الگوهای پنهان در داده‌ها می‌توان از روش‌های داده‌کاوی استفاده کرد. این نتایج کمک می‌کند تا بتوانند راه حل های جدیدی را برای درمان یا پیشگیری بیماری ها پیداکنند. افسردگی همراه با اختلال در اندیشه و بدن ایجاد می‌شود. هدف این پژوهش، ارائه مدلی در جهت تشخیص میزان افسردگی و بررسی ارتباط آن با موسیقی در جهت ارائه راه حل‌های مفید به منظور بهبود این بیماران است. در این تحقیق، ۴۷۰ نفر از بیماران مبتلا به افسردگی ۲ شهر تهران و کرج مورد بررسی قرار گرفتند تا ارتباط بین پارامترهای سبک زندگی و موسیقی مورد علاقه آنها با بیماری افسردگی کشف شود. بیماری‌های فیزیکی مختلفی می‌تواند منجر به افسردگی شود. از این رو، تعدادی از این بیماری‌ها در این پژوهش مورد بررسی قرار گرفتند. از الگوریتم‌های درخت تصمیم‌گیر و ماشین‌بردار پشتیبان و برنامه رپیدماینر استفاده شد. نتایج نشان داد که موسیقی و ورزش نقش مهمی در میزان افسردگی افراد دارد و گوش کردن به موسیقی‌های غمگین و راک و متال می‌تواند به افسرده شدن فرد کمک کند.

**واژگان کلیدی**: افسردگی، داده‌کاوی، موسیقی، درخت تصمیم گیر، ماشین بردار پشتیبان

---

*. گروه کارشناسی ارشد فناوری اطلاعات، دانشگاه غیاث الدین جمشید کاشانی، آبیک ، قزوین ، ایران
Mahtab.jamaly@gmail.com

**. گروه کارشناسی ارشد فناوری اطلاعات، دانشگاه غیاث الدین جمشید کاشانی، آبیک ، قزوین ، ایران
Zohre.faraji.212@gmail.com

***. استادیار گروه کامپیوتر، دانشگاه ایوان کی، ایوان کی، سمنان، ایران Rabiei.eyc@gmail.com

**مقدمه**

امروزه در دانش پزشکی جمع‌آوری داده‌های فراوان در مورد بیماری‌های مختلف دارای اهمیت فراوانی است. صنعت سلامت به طور مستمر در حال تولید میزان زیادی داده است و افرادی که با این نوع داده ها مواجه هستند، دریافتند که بین جمع‌آوری تا تفسیر داده شکاف وسیعی وجود دارد و داده‌کاوی از جمله شیوه‌هایی است که می‌تواند این صنعت را از تحلیل عمیق این داده‌ها بهره‌مند سازد و به توسعه تحقیقات پزشکی و تصمیم گیری‌های علمی در زمینه تشخیص و درمان کمک کند (ملک‌پور، اسماعیل‌پور، ۱۳۹۶). اکتشاف الگوهای پنهان یکی از اساسی‌ترین کاربردهای داده‌کاوی است. این الگوها می‌توانند سوی پزشکان برای تشخیص، پیش‌بینی و درمان بیماران در سازمان‌های بهداشتی استفاده شوند. افسردگی یک بیماری جدی پزشکی است که همراه با اختلال در خلق و خوی، اندیشه و بدن ایجاد می‌شود و باعث می‌شود فرد احساس ناراحتی و بی فایده بودن کند و به طور مداوم فاقد توانایی زندگی طبیعی باشد (دایمی، بانیتان،۲۰۱۴). امروزه افسردگی یکی از شایع‌ترین بیماری‌های روانی در جهان محسوب می‌شود. به گزارش سازمان بهداشت جهانی ۳۵۰ میلیون نفر که ۵ درصد جمعیت کل جهان را تشکیل می‌دهند از بیماری افسردگی رنج می‌برند. نرخ ابتلا به افسردگی نسبت به یک دهه گذشته نزدیک به ۲۰ درصد افزایش یافته است. بنا بر آمار و با توجه به آخرین نظرات کارشناسان، مسئولان و وزیر بهداشت در سال ۳۹۶، آمار افسردگی و بیماری‌های روحی در ایران را به طور میانگین باید مابین ۲۰ تا ۲۵ درصد دانست. اگر افسردگی درمان نشود، می‌تواند باعث افزایش احتمال وابستگی به مشروبات الکلی و مواد مخدر شود (هانگ،۲۰۱۷). همچنین احتمال خودکشی در فرد بیمار بیشتر خواهد شد و می‌تواند منجر به بیماری‌های فیزیکی در بدن فرد بیمار شود. از این رو، بررسی بیماری‌های جسمانی افراد افسرده کمک شایانی به تحقیقات در این زمینه خواهد کرد. افسردگی می‌تواند علل بسیاری داشته باشد که از جمله آن می‌توان به عوامل ژنتیکی، عوامل محیطی، بدبینی، اضطراب و ... اشاره کرد (دایمی،بانیتان،۲۰۱۴) . بر اساس راهنمای تشخیص اختلالات روانی[1] افراد افسرده نشانه‌هایی دارند؛ از جمله، کاهش قابل ملاحظه علاقه و احساس لذت نسبت به بیشتر فعالیت‌ها، کاهش و یا افزایش وزن بدون پرهیز و رژیم غذایی، بیخوابی یا پرخوابی، خستگی یا فقدان انرژی، احساس بی ارزشی و یا گناه. تاکنون برای درمان افسردگی، روش‌های مختلفی کشف و به کار گرفته شده است؛ مانند: دارو درمانی، نور درمانی، درمان الکتروشوک، موسیقی درمانی و ... که موسیقی درمانی یکی از جدیدترین روش‌های مؤثر در درمان افسردگی است که تأثیر مثبتی بر روی بیماران گذاشته و باعث بهبود آنان شده است. در این تحقیق، ۴۷۰ نفر از افسردگان مورد بررسی قرار گرفتند تا ابتدا مشخص شود چه

1. DSM-IV

عواملی بیشترین تأثیر را روی افسردگی دارد و ارتباط آن با موسیقی چیست. همچنین بیماری‌های فیزیکی که می‌تواند به مبتلا شدن فرد به افسردگی نقش داشته باشد، مشخص شوند. بنابراین، با استفاده از الگوریتم‌های درخت تصمیم‌گیر و ماشین‌بردار پشتیبان و روش‌های داده‌کاوی و همچنین برنامه رپیرماینر به بررسی داده‌ها پرداختیم و بهترین نتایج استخراج و دقت مدل‌ها بررسی شد.

## جمع‌آوری داده‌ها و پیش‌پردازش آنها

در طی این پژوهش، تعداد ۸۵۰ نفر از شهرهای تهران و کرج مورد بررسی قرار گرفتند. از این تعداد، ۴۷۰ نفر از آنها از طریق تست افسردگی بک و مراکز درمانی روان‌پزشکی افسرده تشخیص داده شدند که پرسشنامه‌ای برای جمع‌آوری اطلاعات آنان به صورت آنلاین تهیه و در اختیارشان قرار گرفت تا به آن پاسخ دهند. این پرسشنامه در قالب ۱۷ سؤال تهیه شد. طی مدت ۱ ماه جواب پرسشنامه‌ها آماده و جمع‌آوری شدند و در فایل اکسل درج شدند.

داده‌های ما دارای ۱۷ ویژگی است. مقادیر ویژگی‌های به صورت جدول ۱ آمده است:

**جدول۱: ویژگی‌های مورد بررسی در پرسشنامه و دیتاست**

| ردیف | نام ویژگی | نوع داده | محدوده |
|---|---|---|---|
| ۱ | جنسیت | اسمی | (زن) ، (مرد) |
| ۲ | سن | عددی | (۱۷-۵۰) |
| ۳ | تک فرزند بودن | اسمی | (خیر) ، (بله) |
| ۴ | فرزند چندم خانواده | عددی | (۱-۵) |
| ۵ | مدرک تحصیلی | اسمی | بی سواد، سیکل، دیپلم، لیسانس، ارشد، دکترا |
| ۶ | شغل | اسمی | بیکار، خانه دار، کارگر، کارمند، معلم، آزاد، دکتر |
| ۷ | تاهل | اسمی | (مجرد) ، (متاهل) |
| ۸ | تعداد فرزندان | عددی | (۰-۳) |
| ۹ | میزان گوش دادن به موسیقی | اسمی | (کم) ، (متوسط) ، (زیاد) |
| ۱۰ | موسیقی مورد علاقه | اسمی | غمگین، رپ، سنتی ، پاپ، راک و متال |
| ۱۱ | زمان گوش کردن به موسیقی | اسمی | صبح، ظهر، شب |
| ۱۲ | میزان مسافرت کردن | اسمی | کم، متوسط، زیاد |
| ۱۳ | میزان خواب شبانه | اسمی | کم، متوسط، زیاد |
| ۱۴ | اشتها | اسمی | کم شده، زیاد شده ، تغییر نکرده |
| ۱۵ | بیماری های فیزیکی | اسمی | گوارشی، هورمونی، تب رماتیسم، کیست، آرتروز، دیسک، قلبی، سرطان، کبد چرب، فشار چشم، کلیوی، میگرن، |
| ۱۶ | میزان ورزش کردن | اسمی | خیلی کم، متوسط ، حرفه ای |
| ۱۷ | میزان افسردگی | اسمی | کم، زیاد، متوسط |

**۱. پیش‌پردازش داده‌ها**

در بیماری افسردگی، از آنجایی که بسیاری از اطلاعات بیماران از طریق پرسشنامه جمع‌آوری شده است، داده‌های از دست رفته زیادی ممکن است در دیتاست ما وجود داشته باشد که نتیجه کار را تحت تأثیر قرار دهد. به این منظور، قبل از انجام هرکاری باید داده‌های خود را آماده کنیم. آماده‌سازی داده‌ها شامل: تمیز کردن داده‌ها، گسسته‌سازی داده‌ها و تبدیل آنها به فرمت مناسب با الگوریتم مورد نظر است که این امر سبب خروجی با کیفیت‌تر خواهد شد (خاریا، ۲۰۱۲ ). برای تمیز کردن داده‌ها اگر رکوردی مقادیر از دست رفته دارد، باید آن مقادیر حذف و یا جایگزین شوند و در صورتی که تعداد داده‌های از دست رفته در یک ویژگی بیش از ۵۰ درصد رکوردها را در بر گیرد، آن ویژگی باید حذف شود. ویژگی تاریخ پرکردن پرسشنامه که در ابتدا در نظر گرفته شده بود، بعدها به دلیل کم اهمیت بودن در نتیجه پایانی حذف شد. ویژگی میزان درآمد نیز به دلیل اینکه تعداد داده‌های از دست رفته ما در این ویژگی بسیار زیاد بود و این ویژگی را از لیست دیتاست خود حذف کردیم. همچنین رکوردهای ویژگی‌های میزان مسافرت کردن، شغل و موسیقی مورد علاقه که دارای مقادیر از دست رفته بودند را در برنامه رپیدماینر حذف یا جایگزین کردیم. از اپراتور فیلتر استفاده کردیم و فیلتر را برای نمایش رکوردهایی که هیچ مقدار از دست رفته‌ای ندارند قرار می‌دهیم تا اطمینان حاصل شود که دیگر رکوردی با مقادیر از دست رفته در داده‌های ما وجود نخواهد داشت (تانجا،۲۰۱۳)

سه ویژگی شغل، موسیقی مورد علاقه و مسافرت دارای مقادیر از دست رفته هستند و الگوریتم‌هایی مثل ماشین بردار پشتیبان توانایی کار کردن با ویژگی‌هایی با مقادیر از دست رفته را ندارند. در مرحله اول بعد از مشخص شدن وجود مقادیر از دست رفته آنها را حذف و یا جایگزین می‌کنیم. در صورت داشتن داده‌های نا صحیح که بیشتر به دلیل اشتباهات سهوی به وجود می‌آیند، داده‌های خود را نرمال‌سازی می‌کنیم تا نتیجه پایانی صحیح‌تری داشته باشیم. برای هر ویژگی یک فاصله اقلیدسی تعریف می‌شود و داده‌ها نسبت به آن فاصله سنجیده میشوند و اگر در آن بازه باشند داده نرمال محسوب می‌شوند و اگر در آن باز نباشند داده نادرست محسوب می‌شوند (هان، کمبر،۲۰۰۶)

**۲. مدل ها و الگوریتم‌های داده‌کاوی مورد استفاده در این پژوهش**

مدل‌های داده‌کاوی مورد استفاده در این تحقیق عبارت‌اند از: الگوریتم درخت تصمیم‌گیر، الگوریتم ماشین‌بردار پشتیبان، مدل ترکیبی الگوریتم ماشین‌بردار پشتیبان.

**الف) درخت تصمیم‌گیر**

درخت تصمیم‌گیر یک ابزار کمکی برای تصمیم‌گیری است و از یک گراف یا مدل درختی که تصمیمات و عواقب محتمل آنها را نمایش می‌دهد، تشکیل شده است. الگوریتم درخت تصمیم قادر است، علاوه بر متغیرهای کمی، متغیرهای کیفی را نیز پیش‌بینی کند. درخت تصمیم درختی است که در آن نمونه‌ها را به نحوی دسته‌بندی می‌کند که از ریشه به سمت پایین رشد می‌کنند و در نهایت، به گره‌های برگ می‌رسد(چن، زو،۲۰۱۷). هر گره غیر برگ با یک ویژگی مشخص می‌شود. قوانین تولیدشده و به کارگرفته شده قابل استخراج و قابل فهم هستند و همچنین توانایی کار با داده‌های پیوسته و گسسته را دارد. آماده‌سازی داده‌ها برای یک درخت تصمیم، ساده یا غیرضروری است در حالی که روش‌های دیگر اغلب نیاز به نرمال‌سازی داده یا حذف مقادیر خالی یا ایجاد متغیرهای پوچ دارند. در این پژوهش، انواع محتلف الگوریتم‌های خانواده درخت بررسی می‌شوند (میلویک،۲۰۱۲).

**ب) ماشین‌بردار پشتیبان**

ماشین‌بردار پشتیبان به عنوان یکی از بهترین تکنیک‌های دسته‌بندی و پیش‌بینی و تشخیص داده‌های خارج از محدوده شناخته می شود و برخلاف الگوریتم‌های خوشه‌بندی در دسته یادگیری با نظارت محسوب می‌شود. این الگوریتم با افزایش ابعاد مسئله و با استفاده از نگاشت کرنل، یک چارچوب کاری یکپارچه را برای اکثر مدل‌ها فراهم می‌کنند. استفاده از این روش، برای مجموعه داده‌هایی با ابعاد ویژگی بالا و حجم داده‌های پایین، کارایی بالایی از خود نشان می‌دهد و حاشیه جداسازی برای دسته‌های مختلف کاملاً واضح است. ماشین‌بردار پشتیبان مرزی است که به بهترین شکل دسته‌های داده‌ها را از یکدیگر جدا می‌کند. با فرض اینکه دسته‌ها به صورت خطی جداپذیر باشند، ابر صفحه‌هایی با حداکثر حاشیه را به دست می‌آورد که دسته‌ها را جدا کنند (هالاریس، ۲۰۱۱).

**معیارهای ارزیابی**

در همه مدل‌ها پارامترهای زیر محاسبه می‌شود:

**۱. ماتریس در هم ریختگی**

ماتریس در هم ریختگی[1] به ماتریسی گفته می‌شود که در آن عملکرد الگوریتم‌های مربوطه را نشان می‌دهد. این ماتریس یک ماتریس مربعی N در N است. در این ماتریس، تعدادی نمونه وجود دارد

1. Confusion Matrix

که قرار است این نمونه‌ها را در دسته نمونه‌های مثبت و یا منفی قرار گیرند و میزان درست و نادرست بودن دسته‌بندی را برای ما مشخص کند (کدی، امیت،۲۰۱۵) این ماتریس ۲ قطر دارد و قطر اصلی ماتریس نشان‌دهنده تعداد داده‌هایی است که به درستی دسته‌بندی شده‌اند و قطر دیگر نشان دهنده داده‌هایی است که به اشتباه دسته بندی شده‌اند. زیاد بودن مقادیر قطر اصلی و کم بودن مقادیر قطر فرعی نشان‌دهنده این است که دسته‌بندی داده‌ها به درستی انجام شده است. عناصر این ماتریس به شرح زیر است:

TP تعداد نمونه‌های مثبتی که به درستی مثبت طبقه بندی شده‌اند.

FP تعداد نمونه‌های منفی که به اشتباه مثبت طبقه‌بندی شده‌اند.

TN تعداد نمونه‌های منفی که به درستی منفی طبقه‌بندی شده‌اند.

FN تعداد نمونه‌های مثبتی که به اشتباه منفی طبقه بندی شده‌اند.

### ۲. نمودارهای ROC و AUC

ROC یک ابزار مدل‌سازی قوی است که در تصمیم‌گیری ها و در زمانی که ارزش‌های آستانه‌ای مد نظر است، استفاده می‌شود. این منحن یک نمودار پراگندگی از حساسیت[1] برای یک سیستم طبقه بندی کننده دودویی است. از طریق این نمودار می‌توان ۳ معیار اساسی در بررسی صحت یک مدل را بررسی کرد. این ۳ معیار شامل دقت، ویژگی و حساسیت است که به شرح زیر محاسبه می‌شوند:

الف) دقت[2]: معیاری برای دسته‌بندی درست است. یا به عبارتی مشخص می‌کند که چه تعداد از نمونه‌های مثبت در کلاس مثبت و چه تعداد از نمونه‌های منفی در کلاس منفی دسته‌بندی شده‌اند.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{All Data}}$$

ب) ویژگی[3]: این معیار درصد نمونه های منفی که به درستی طبقه بندی شده‌اند را نشان می‌دهد.

$$\text{Specificity} = \frac{\text{TN}}{\text{FP+TN}}$$

ج) حساسیت[4] : این معیار درصد نمونه‌های مثبت که به درستی مثبت تشخیص داده شده‌اند را مشخص می‌کند.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}}$$



1. sensitivity
2. Accuracy
3. Specificity
4. Sensitivity

سطح زیر نمودار ROC یک معیار استاندارد از میزان تشخیص مربوط به روش شناسایی بوده که به آن AUC می‌گویند. این مقدار کارایی یک مدل را نشان می‌دهد و هر چه این مقدار بیشتر باشد، بهتر است.

**۳. ضریب کاپای کوهن**

ضریب کاپای کوهن یک روش آماری است که در طبقه‌بندی استفاده می‌شود. این ضریب به عنوان یک اندازه‌گیری قوی برای محاسبه درصد پیش‌بینی درست استفاده می‌شود. ضریب کاپا اندازه‌ای عددی بین ۱- تا ۱+ است که هر چه به ۱+ نزدیک‌تر باشد، بیانگر وجود توافق متناسب و مستقیم است. اندازه‌های نزدیک به ۱- نشان‌دهنده وجود توافق وارون و عکس و اندازه‌های نزدیک به صفر عدم توافق را نشان می‌دهد (هان،کمبر،۲۰۰۶).

$$\text{Kappa} = \frac{(\text{Po}-\text{Pe})}{(1-\text{Pe})}$$

**۴. ماتریس همبستگی**

این ماتریس مشخص می‌کند که در یک پژوهش آیا دو ویژگی با هم ارتباط دارند یا خیر. این میزان عددی بین ۱- و ۱+ است. مقدار مثبت نشان‌دهنده رابطه مثبت بین دو ویژگی X وY است. مقدار منفی نشان‌دهنده ارتباط منفی و صفر نشان‌دهنده عدم ارتباط است ( متیو،۱۹۷۹). در این پژوهش، این ماتریس بر روی بیماری‌های ذکرشده نظیر: گوارشی، دیسک کمر، سرطان، هورمونی و ... اعمال می‌شود (اختر، ۲۰۱۲).

## پیاده‌سازی الگوریتم درخت تصمیم

بعد از آماه‌سازی داده‌ها ابتدا داده‌ها را برچسب دار کنیم و تعیین کنیم که هدف ما پیش‌بینی نتیجه کدام ویژگی است. برچسب را بر روی میزان افسردگی قرار می‌دهیم.

**۱. قوانین استخراج شده از درخت تصمیم**

با استفاده از درخت تصمیم، تعدادی قانون برای تشخیص میزان افسردگی به دست آمد. بخشی از ساختار درخت تصمیم در شکل ۱ آورده شده است. مشاهده می شود دو ویژگی ورزش و موسیقی مورد علاقه در بالای درخت قرار دارند. بالا بودن ویژگی در درخت نشان‌دهنده اهمیت و تأثیرگذار بودن آن ویژگی است.


**شکل۱: بخشی از نتیجه درخت تصیم در رپیدماینر**

حال با استفاده از اپراتورهای این الگوریتم دقت مدل درخت تصمیم خود را اندازه‌گیری می‌کنیم:

accuracy: 90.91%

| | true Medium | true Much | true low | class precision |
|---|---|---|---|---|
| pred. Medium | 105 | 7 | 13 | 84.00% |
| pred. Much | 8 | 180 | 7 | 92.31% |
| pred. low | 5 | 2 | 135 | 95.07% |
| class recall | 88.98% | 95.24% | 87.10% | |

**شکل ۲: دقت الگوریتم درخت تصمیم‌گیر بر روی دیتاست این پژوهش**

با استفاده از الگوریتم درخت تصمیم‌گیر تعداد ۲۶۳ قانون برای این مدل تولید شد که از این میان قانون‌هایی که بیشترین تکرار و اطمینان را دارند، بررسی می‌کنیم. بخشی از این قوانین پر تکرار تولیدشده به شرح زیر است:

**جدول۲: قوانین تولیدشده بیماران افسرده با الگوریتم درخت تصمیم‌گیری قبل از کاهش ویژگی‌ها**

| قوانین تولید شده | تکرار | اطمینان |
|---|---|---|
| ۱) ورزش کم و مدرک تحصیلی لیسانس و موسیقی غمگین ← میزان افسردگی زیاد | (۱۴ / ۱۱۰ / ۱۱) | ۱ |
| ۲) خواب شبانه متوسط و متعهل و موسیقی مورد علاقه موسیقی پاپ ← میزان افسردگی کم | (۵۱ / ۰ / ۴) | ۱ |
| ۳) ورزش گاهی اوقات و سن ≥ ۳۳,۵۰۰ ← میزان افسردگی متوسط | (۳۲ / ۴ / ۶۰) | ۱ |
| ۴) مسافرت کم و موسیقی مورد علاقه غمگین ← میزان افسردگی زیاد | (۶ / ۴۶ / ۶) | ۱ |
| ۵) اشتها تغییر نکرده است و ورزش حرفه ای ← میزان افسردگی کم | (۱۸ / ۰ / ۱) | ۱ |
| ۶) موسیقی مورد علاقه غمگین و جنسیت زن ← میزان افسردگی متوسط | (۶ / ۹ / ۱۹) | ۰,۹۶ |
| ۷) زمان گوش کردن به موسیقی شب و موسیقی مورد علاقه راک و متال ← میزان افسردگی زیاد | (۰ / ۱۰ / ۲) | ۰,۹۳ |
| ۸) ورزش متوسط و موسیقی مورد علاقه سنتی ← میزان افسردگی کم | (۶۱ / ۰ / ۳) | ۰,۸۹ |
| ۸) ورزش متوسط و موسیقی مورد علاقه سنتی ← میزان افسردگی کم | (۶۱ / ۰ / ۳) | ۰,۸۹ |

دو ویژگی موسیقی مورد علاقه و ورزش در نتایج زیاد به چشم می‌خورد. بعد از به دست آمدن نتایج، سه ویژگی تعداد فرزندان، تک فرزند بودن و فرزند چندم بودن را از ویژگی‌های خود حذف کردیم و بعد از حذف دقت مدل ما به ۹۱,۱۳ ٪ رسید و افزایش یافت. پس این ۳ ویژگی کم اهمیت بوده و آنها را از لیست ویژگی‌های خود حذف می‌کنیم.

دوباره قانون های ایجاد شده را در جدول ۳ بررسی می‌کنیم.

**جدول۳: قوانین تولید شده بیماران افسرده با الگوریتم درخت تصمیم‌گیری بعد از کاهش ویژگی‌ها**

| قوانین تولیدشده | تکرار | اطمینان |
|---|---|---|
| ۱) ورزش کم و موسیقی مورد علاقه غمگین ← میزان افسردگی زیاد | (۱۰ / ۱۴۹ / ۲۷) | ۱ |
| ۲) خواب شبانه متوسط و اشتها تغییری نکرده است ← میزان افسردگی کم | (۴۱ / ۰ / ۲) | ۱ |
| ۳) موسیقی مورد علاقه پاپ و ورزش حرفه ای ← میزان افسردگی کم | (۲۱ / ۰ / ۰) | ۱ |
| ۴) مجرد باشد و میزان ورزش گاهی اوقات ← میزان افسردگی متوسط | (۴۷/۱۱/۱۶) | ۱ |

| قوانین تولیدشده | تکرار | اطمینان |
|---|---|---|
| ۵) میزان خواب شبانه متوسط و میزان موسیقی متوسط ← میزان افسردگی کم | (۲/۰/۱۹) | ۱ |
| ۶) اگر اشتها تغییری نکند و جنسیت زن باشد ← میزان افسردگی کم | (۵/۰/۱۷) | ۱ |
| ۷) موسیقی مورد علاقه غمگین باشد و سن ≥۳۴,۵۰۰ ← میزان افسردگی متوسط | (۹/۱/۲) | ۰,۹۷ |
| ۸) موسیقی مورد علاقه پاپ و میزان خواب شبانه کم ← میزان افسردگی کم | (۱/۱/۷) | ۰,۹۲ |

در این مرحله هم ویژگی‌های ورزش و موسیقی مورد علاقه جزو ویژگی های پرتکرار هستند. در حالت کلی می‌توان گفت که :

**جدول۴: علائم افراد افسرده با توجه به درخت تصمیم**

| علائم | میزان افسردگی |
|---|---|
| کسی که کم ورزش می‌کند.<br>موسیقی غمگین گوش می‌دهد.<br>کم به مسافرت می‌رود .<br>موسیقی را شب هنگام گوش می‌دهد.<br>سن > ۲۸,۵۰ باشد.<br>میزان موسیقی زیاد باشد .<br>اشتها کاهش یافته کرده باشد.<br>میزان خواب شبانه زیاد باشد . | زیاد |
| کسی که گاهی اوقات ورزش می‌کند.<br>موسیقی غمگین گوش می دهد .<br>جنسیت زن باشد.<br>سن <=۳۳,۵۰<br>میزان موسیقی متوسط باشد. | متوسط |
| کسی که موسیقی پاپ گوش می‌دهد .<br>میزان خواب شبانه متوسط دارد.<br>متاهل است.<br>اشتهای آنها تغییری نکرده است.<br>گاهی اوقات ورزش می‌کند.<br>جنسیت زن باشد.<br>سن <=۳۱,۵۰ | کم |

**جدول۵: ماتریس در هم ریختگی در مدل درخت تصمیم‌گیری**

| کلاس | افسردگی متوسط | افسردگی زیاد | افسردگی کم | مجموع |
|---|---|---|---|---|
| افسردگی متوسط | ۱۰۷ | ۹ | ۱۵ | ۱۳۱ |
| افسردگی زیاد | ۵ | ۱۷۹ | ۵ | ۱۸۹ |
| افسردگی کم | ۶ | ۱ | ۱۳۵ | ۱۴۲ |
| مجموع | ۱۱۸ | ۱۸۹ | ۱۵۵ | ۴۶۲ |

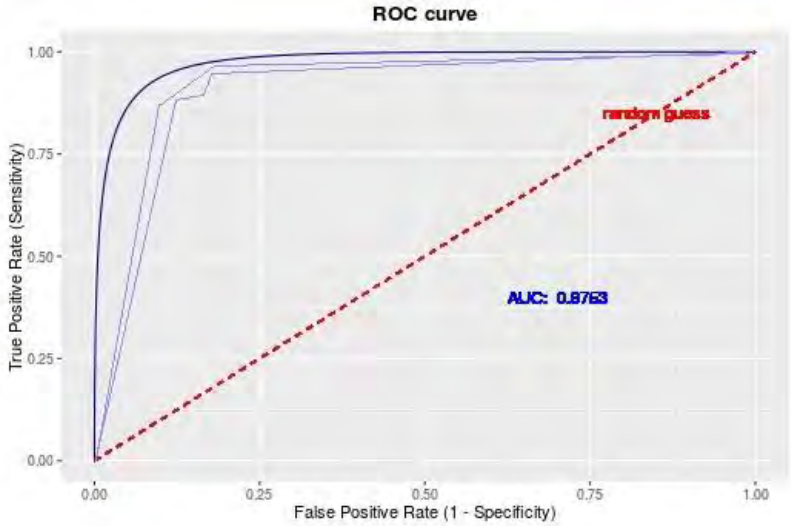

**شکل۳: مقادیر قابل ارزیابی ماتریس در هم ریختگی درخت تصمیم در ۳ کلاس میزان افسردگی**

طبق معیارهای اندازه‌گیری، معیار TP که تعداد افراد افسرده‌ای است که به درستی دسته‌بندی شده‌اند، نشان می دهد که بر اساس الگوریتم درخت تصمیم برای کلاس افسردگی زیاد ۰/۹۳۲ است که به این معناست که ۰/۰۶۸ از نمونه‌هایی که متعلق به کلاس افسردگی زیاد است، به اشتباه در کلاس‌های دیگر طبقه‌بندی شده‌اند. مقدار TP برای کلاس‌های افسردگی متوسط و افسردگی کم به ترتیب ۰/۸۹۹ و ۰/۸۴۹ است.

**جدول۶. ارزیابی مدل درخت تصمیم‌گیری با توجه به ماتریس در هم ریختگی**

| کلاس | حساسیت | ویژگی | دقت |
|---|---|---|---|
| افسردگی متوسط | ٪۹۰,۶ | ٪۹۶,۶ | ٪۸۱,۶ |
| افسردگی زیاد | ٪۹۴,۷ | ٪۹۶,۳ | ٪۹۴,۷ |
| افسردگی کم | ٪۸۷,۱ | ٪۹۳,۷ | ٪۹۵ |

دقت در این مدل ٪ ۹۱,۱۳ است. حساسیت = ٪۹۰,۸ ، ویژگی مدل = ٪۹۵,۵ و ضریب کاپا = ٪۷۴,۲ است.




**شکل۴: نمودار ROC الگوریتم درخت تصمیم‌گیری**

## ۲. اعمال انواع مختلف درخت و مقایسه دقت آنها

خانواده الگوریتم درخت انواع مختلفی دارد. برروی این دیتاست ما انواع آنها را تست کرده تا دریابیم که کدام نوع جواب بهتر و دقیق‌تری با توجه به دیتاست می‌دهد که به شرح زیر است:

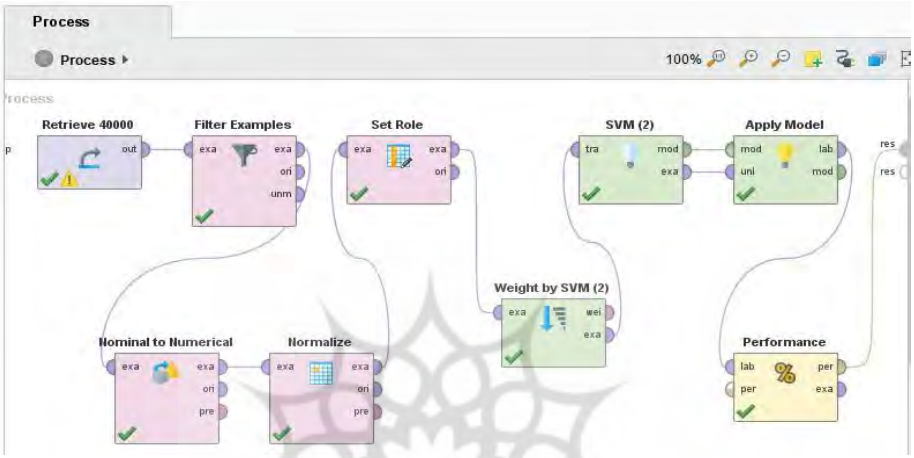
**جدول۷: مقایسه میزان دقت مدل‌های الگوریتم‌های درخت**

| نوع مدل درخت | میزان دقت مدل |
|---|---|
| Decision Tree | ٪۹۱,۱۳ |
| Random Forest | ٪۸۹,۸۳ |
| Random Tree | ٪۶۲,۹۹ |
| Gradient Boosted Trees | ٪۹۳,۲۹ |
| CHAID | ٪۸۷,۶۶ |
| Decision Stump | ٪۶۲,۹۹ |

با توجه به مقادیر به دست آمده مدل‌های درخت تصمیم گیری با ٪۹۱,۱۳ دقت و ارتقای گرادیان با ٪۹۳,۲۹ دقت بیشترین میزان دقت را در برخورد با دیتاست دارند. پس از دسته الگوریتم های درخت این دو مدل در پیش‌بینی بیماری افسردگی بیشترین کاربرد را دارند.

## پیاده‌سازی الگوریتم ماشین‌بردار پشتیبان[1]

در ماشین‌بردار پشتیبان نرمال‌سازی داده‌ها از اهمیت زیادی برخوردار است به همین منظور، باید داده‌ها در محدوده‌های (۰و۱) و یا (۱- و ۱) نرمال‌سازی کنیم. این الگوریتم با داده‌های چندمقداره یا چندمقداره سازگار نیست. از این رو، باید ابتدا داده‌های چندمقداره را به مقادیر عددی تبدیل کنیم که در این میان ویژگی میزان افسردگی به دلیل اینکه اپراتور libSVM استفاده شده است، داده‌های چندمقداره را برای داده برچسب زده شده قبول می‌کند، از این قاعده مستثناست و به صورت چندمقداره باقی می‌ماند.




**شکل۵: پیاده‌سازی الگوریتم ماشین‌بردار پشتیبان در رپیدماینر**

اکنون میزان دقت مدل را قبل از کم و زیاد کردن ویژگی‌ها به دست می‌آوریم که به شرح زیر است:

accuracy: 72.08%

| | true Medium | true Much | true low | class precision |
|---|---|---|---|---|
| pred. Medium | 8 | 0 | 0 | 100.00% |
| pred. Much | 55 | 185 | 15 | 72.55% |
| pred. low | 55 | 4 | 140 | 70.35% |
| class recall | 6.78% | 97.88% | 90.32% | |

**شکل ۶: میزان دقت مدل ماشین‌بردار پشتیبان قبل از تغییر داده‌ها**

1 SVM

**۱. انواع مختلف ماشین‌بردار پشتیبان و پیاده‌سازی آنها**

این الگوریتم انواع مختلفی دارد که شامل موارد زیر است:

۱) c-SVC ( برای مسائل دسته‌بندی استفاده می‌شود )؛

۲) nu-SVC ( برای مسائل دسته‌بندی استفاده می شود )؛

۳) Epsilon-SVR ( برای مسائل رگرسیونی استفاده می شود )؛

۴) nu-SVR ( برای مسائل رگرسیونی استفاده می‌شود )؛

که ما از دو مدل c-SVC و nu-SVC به دلیل دسته‌بندی بودن مدل استفاده می‌کنیم و تفاوت آنها در نوع پارامترهایی که دارند، مشخص می‌شود. در c-SVC پارامتر c مد نظر است و در nu-SVC پارامتر nu. رنج پارامتر c از صفر تا بینهایت است ولی رنج پارامتر nu بین ۰ و ۱ است. یک ویژگی خوب در مورد پارامتر nu این است که به ضریب بردارهای پشتیبان و خطاهای آموزشی مربوط می‌شود. برنامه رپیدماینر به صورت پیش‌فرض بر روی نوع c-SVC قرار دارد. یک بار هم باید دقت مدل خود را با استفاده از nu-SVC به دست آوریم و جواب‌ها را با هم مقایسه کنیم.

دقت مدل ما بعد از اجرای nu-SVC از ۷۲,۰۸٪ به ۸۳,۳۳٪ افزایش یافت . پس می‌توان گفت که این مدل ، مدل بهتری برای نتیجه کار ماست.

**۲. انواع کرنل‌ها در ماشین‌بردار پشتیبان**

ماشین بردار پشتیبان برای اینکه داده های غیر خطی را از هم تفکیک کند، باید از کرنل‌های مختلف استفاده کند. برای این کار دیگر در فضای دو بعدی کار نمی‌کند بلکه داده‌ها به فضایی با ابعاد بیشتر نگاشت داده می‌شوند تا بتوان آنها را در این فضای جدید به صورت خطی تفکیک کرد. انواع مختلف کرنل‌ها شامل: RBF ، Linear ، Polynomial، Sigmoid و Precomputed هستند. RBF محبوب‌ترین و پرکاربردترین کرنل ماشین‌بردار پشتیبان است و کرنل چمد جمله‌ای در پردازش زبان طبیعی محبوبیت دارد. درجه متداول آن ۲ بوده و درجه‌های بیشتر از این مقدار باعث بیش پردازش در مسائل پردازش زبان طبیعی می‌شود. زمانی که شخص مطمئن نیست برای نتیجه بهتر از کدام یک از کرنل‌ها استفاده کند، می‌تواند از تکنیک‌های انتخاب خودکار مثل: اعتبارسنجی متقابل استفاده کند و یا از ترکیبی از کلاسیفایرها که با کرنل‌های مختلف به دست می‌آید. RBF از منحنی‌های نرمال در اطراف نقاط داده استفاده می‌کند و اینها را طوری با هم جمع می‌کند که مرز تصمیم را بتوان به نوعی تعریف کرد. با استفاده از کرنلی مثل RBF فضای ویژگی از طریق یک تبدیل غیر خطی به دست می‌آید.

انواع مختلف کرنل و دقت آنها در مدل ما به شرح زیر است:

**جدول۸: اندازه دقت مدل با استفاده از کرنل‌های مختلف**

| نوع کرنل | دقت مدل با استفاده از این کرنل |
|---|---|
| RBF | ٪۸۳,۳۳ |
| Linear | ٪۸۳,۱۲ |
| Polynomial | ٪۶۰,۳۹ |
| Sigmoid | ٪۸۳,۱۲ |
| Precomputed | ٪۳۳,۵۵ |

به صورت پیش فرش رپیدماینر بر روی کرنل RBF قرار دارد و به طور کلی، در اکثر موارد این کرنل جواب منطقی‌تر و بهتری نسبت به سایر کرنل‌ها می‌دهد . در مدل‌ها هم کرنل RBF بیشترین دقت را داشت. پس در نتیجه، ما از مدل nu-SVC و کرنل RBF استفاده می‌کنیم تا دقت مدل را بالا ببریم و جواب بهتری به دست آوریم.

در مرحله بعد ۳ ویژگی تعداد فرزندان، تک فرزند بودن و فرزند چندم بودن را حذف کردیم. مشاهده شد که دقت ما همچنان ٪۸۳,۳۳ است پس این ۳ ویژگی در این مدل ما تأثیر بسیار کمی در میزان افسردگی افراد داشتند. این مدل برای ما ۲۲۱ عدد قانون تولید کرد که تعدادی از قوانین تولیدشده پرتکرار در جدول ۹ آمده است:

**جدول ۹: قوانین تولیدشده بیماران افسرده با الگوریتم ماشین‌بردار پشتیبان**

| قوانین تولیدشده | تکرار | اطمینان |
|---|---|---|
| ۱) موسیقی مورد علاقه غمگین و ورزش کم و سن > ۰,۵۰۰ ← میزان افسردگی زیاد | (۵ / ۷۵ / ۰) | ۱ |
| ۲) اگر میزان ورزش کردن زیاد باشد شغل آزاد باشد ← میزان افسردگی کم | (۱ / ۰ / ۳۲) | ۱ |
| ۳) مسافرت کم و ورزش کم و بیماری فیزیکی نداشته باشد و سن≥۰,۵۶۱ و شغل کارمند باشد ← میزان افسردگی متوسط | (۱۵ / ۰ / ۰) | ۱ |
| ۴) مسافرت کم باشد و موسیقی مورد علاقه راک و متال و مدرک تحصیلی لیسانس ← میزان افسردگی زیاد | (۰ / ۱۲ / ۰) | ۱ |
| ۵) میزان موسیقی زیاد باشد و شغل کارمند و متعل باید و سن ≥۰,۴۵ باشد ← میزان افسردگی کم | (۰ / ۰ / ۸) | ۱ |
| ۶) زمان گوش کردن به موسیقی شب باشد و جنسیت زن باشد و موسیقی مورد علاقه غمگین باشد ← میزان افسردگی زیاد | (۰ / ۱۴ / ۱) | ۰,۹۸ |
| ۷) ورزش کم باشد و جنسیت زن باشد و خواب شبانه زیاد باشد و خانه دار باشد ← میزان افسردگی متوسط | (۱۰ / ۰ / ۱) | ۰,۹۱ |
| ۸) جنسیت زن باشد و موسیقی مورد علاقه غمگین باشد و خواب شبانه زیاد باشد ← میزان افسردگی زیاد | (۲ / ۲۱ / ۰) | ۰,۸۷ |

طبق قوانین تولیدشده می‌توان ویژگی‌های بیماران افسرده را با توجه به الگوریتم ماشین بردار پشتیبان به شرح زیر دسته بندی کرد:

**جدول۱۰ : علائم افراد افسرده با توجه به الگوریتم ماشین‌بردار پشتیبان**

| علائم | میزان افسردگی |
|---|---|
| موسیقی مورد علاقه غمگین و راک و متال<br>زمان گوش کردن به موسیقی شب هنگام<br>ورزش کم<br>مسافرت کم<br>مدرک تحصیلی لیسانس و سیکل<br>جنسیت زن<br>خواب شبانه متوسط<br>نداشتن بیماری فیزیکی<br>کاهش اشتها<br>سن > ۰,۵۰ | زیاد |
| مسافرت کم<br>ورزش کم<br>نداشتن بیماری فیزیکی<br>شغل کارمند<br>جنسیت زن<br>خواب شبانه زیاد<br>میزان گوش کردن به موسیقی متوسط | متوسط |
| ورزش زیاد و کم<br>میزان گوش کردن به موسیقی زیاد<br>کارمند باشد<br>متاهل باشد<br>سن < ۰,۴۵<br>نداشتن بیماری فیزیکی<br>اشتها تغییر نکرده | کم |

با توجه به نتایج به دست آمده می‌توان گفت که ورزش و موسیقی دو فاکتور تأثیرگذار در میزان افسردگی هستند. اکثر بیماران مورد بررسی زن هستند و شغل تأثیر چندانی بر روی میزان افسردگی آنان نداشت. همچنین ۳ فاکتور تک فرزند بودن، فرزند چندم بودن و تعداد فرزندان به طور کامل بی تأثیر بودند و داشتن مدرک تحصیلی بالا دلیلی بر افسرده نبودن فرد نبود. در

افسردگی‌های شدید ورزش کم، موسیقی غمگین، خواب شبانه کم، مسافرت کم و جنسیت زن عناصر اصلی هستند. در افسردگی‌های متوسط ورزش گاهی اوقات، موسیقی غمگین، خواب شبانه زیاد، مسافرت متوسط و جنسیت مرد و در افسردگی‌های کم، ورزش گاهی اوقات، موسیقی راک و متال، خواب شبانه متوسط، مسافرت متوسط و جنسیت زن فاکتورهای اساسی هستند. در اکثر موارد، جنسیت زن بر مرد اولویت دارد؛ پس می‌توان گفت که میزان افسردگی در این مدل نیز در زن‌ها بیشتر است.

**جدول۱۱ : ماتریس در هم ریختگی در مدل ماشین‌بردار پشتیبان**

| کلاس | افسردگی متوسط | افسردگی زیاد | افسردگی کم | مجموع |
|---|---|---|---|---|
| افسردگی متوسط | ۷۸ | ۱۱ | ۱۵ | ۱۰۴ |
| افسردگی زیاد | ۲۶ | ۱۷۷ | ۱۰ | ۲۱۳ |
| افسردگی کم | ۱۴ | ۱ | ۱۳۰ | ۱۴۵ |
| مجموع | ۱۱۸ | ۱۸۹ | ۱۵۵ | ۴۶۲ |

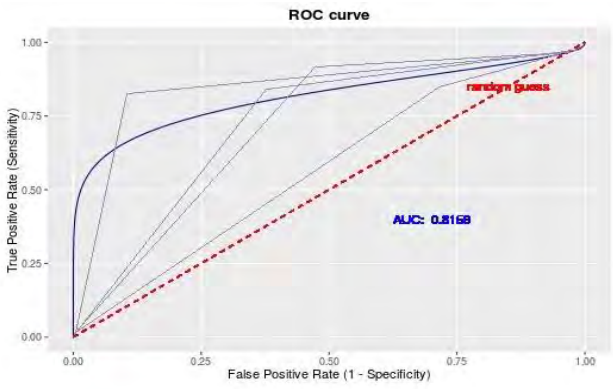

**شکل۷: مقادیر قابل ارزیابی ماتریس در هم ریختگی ماشین بردار پشتیبان در ۳ کلاس میزان افسردگی**

معیار TP در مدل الگوریتم ماشین‌بردار پشتیبان برای کلاس افسردگی زیاد، ۰/۹۲۱ است که به این معناست که ۰/۰۷۹ از نمونه‌هایی که متعلق به کلاس افسردگی زیاد است، به اشتباه در کلاس‌های دیگر طبقه‌بندی شده‌اند. مقدار TP برای کلاس‌های افسردگی متوسط و افسردگی کم به ترتیب ۰/۶۵۵ و ۰/۸۱۷ است.

**جدول۱۲: ارزیابی مدل ماشین‌بردار پشتیبان با توجه به ماتریس در هم ریختگی**

| کلاس | حساسیت | ویژگی | دقت |
|---|---|---|---|
| افسردگی متوسط | ٪۶۶,۱۰ | ٪۸۸,۸ | ٪۷۵,۰۰ |
| افسردگی زیاد | ٪۹۳,۶۵ | ٪۹۵,۱ | ٪۸۳,۱۰ |
| افسردگی کم | ٪۸۳,۸۷ | ٪۹۲,۱ | ٪۸۹,۶۶ |

دقت در این مدل ۸۳,۳۳ ٪ است. حساسیت = ۸۱,۸ ٪ ، ویژگی مدل= ۹۲ ٪ و ضریب کاپا= ۵۸,۶ ٪ است.




**شکل۸: نمودار ROC مدل ماشین‌بردار پشتیبان**

## مدل ترکیبی درخت تصمیم‌گیری و ماشین‌بردار پشتیبان

در پی یافتن مدلی جدید که بیشترین دقت و حساسیت را بر روی داده‌های بیماران افسرده داشته باشد، مدل ترکیبی ۲ الگوریتم درخت تصمیم‌گیری و ماشین‌بردار پشتیبان ارائه شد. این مدل در برنامه رپیدماینر دقت ۹۴,۸۱٪ را نشان می دهد که نسبت به دو روش قبلی از دقت بالاتری برخوردار است.

**جدول۱۳ : ماتریس در هم ریختگی در مدل ترکیبی در تصمیم‌گیری و ماشین‌بردار پشتیبان**

| کلاس | افسردگی متوسط | افسردگی زیاد | افسردگی کم | مجموع |
|---|---|---|---|---|
| افسردگی متوسط | ۱۱۵ | ۷ | ۱۰ | ۱۳۲ |
| افسردگی زیاد | ۲ | ۱۸۲ | ۴ | ۱۸۸ |
| افسردگی کم | ۱ | ۰ | ۱۴۱ | ۱۴۲ |
| مجموع | ۱۱۸ | ۱۸۹ | ۱۵۵ | ۴۶۲ |

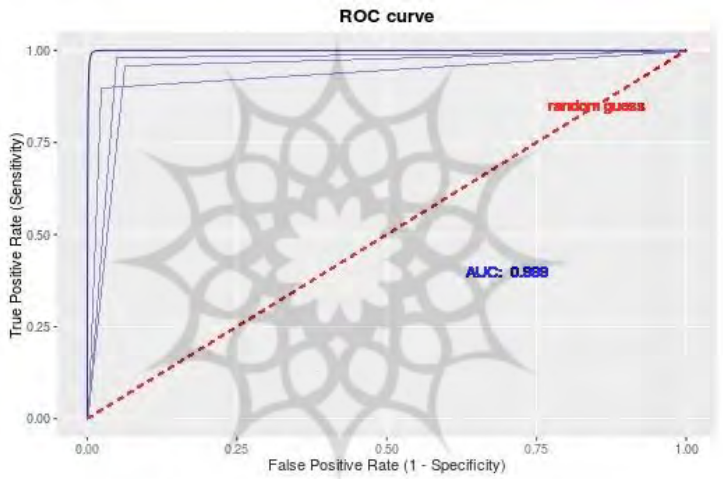

**شکل۹: مقادیر قابل ارزیابی ماتریس در هم ریختگی مدل ترکیبی درخت تصمیم و ماشین‌بردار پشتیبان**

معیار TP در این روش هیبرید برای ۳ کلاس افسردگی زیاد، افسردگی متوسط و افسردگی کم به ترتیب مقادیر ۰/۹۴۷ – ۰/۹۶۶ و ۰/۸۸۶ است که با توجه به ۲ مدل قبلی تعداد مقادیری که به درستی دسته‌بندی شده‌اند بیشتر است.

**جدول۱۴: ارزیابی مدل هیبرید ترکیبی درخت تصمیم‌گیری و ماشین‌بردار پشتیبان**

| کلاس | حساسیت | ویژگی | دقت |
|---|---|---|---|
| افسردگی متوسط | ٪۹۷,۴۶ | ٪۹۹ | ٪۸۷,۱۲ |
| افسردگی زیاد | ٪۹۶,۳۰ | ٪۹۷,۴ | ٪۹۶,۸۱ |
| افسردگی کم | ٪۹۰,۹۷ | ٪۹۵,۶ | ٪۹۹,۳۰ |

دقت این مدل ۹۴,۸۱ ٪ است. حساسیت= ۹۴,۸ ٪ ، ویژگی مدل= ۹۴,۴ ٪ و ضریب کاپا= ۸۱,۳ است.




**شکل۱۰: نمودار ROC مدل ترکیبی درخت تصمیم‌گیری و ماشین‌بردار پشتیبان**

### بررسی سه مدل ارائه‌شده در این پژوهش

سه مدل درخت تصمیم‌گیری، ماشین‌بردار پشتیبان و مدل ترکیبی درخت تصمیم و ماشین بردار پشتیبان بر روی داده‌های بیماران افسرده پیاده‌سازی شدند. این سه مدل از لحاظ حساسیت، ویژگی، دقت، ضریب کاپا و کارایی با هم مقایسه شدند. مدل درخت تصمیم‌گیری با ٪۹۱,۱۳ دقت از مدل ماشین بردار پشتیبان با ٪۸۳,۳۳ دقت جواب بهتری به ما ارائه داد و همچنین درخت

تصمیم گیری قوانین جامع تر و مفیدتری در رابطه با بیماران افسرده در اختیار ما گذاشت. به منظور بهبود نتایج به دست آمده مدل ترکیبی درخت تصمیم و ماشین بردار پشتیبان پیشنهاد شد. این مدل بهترین مدل در پژوهش انجام شده است. مدل هیبرید با ۹۴,۸۱٪ دقت، بیشترین دقت را داراست. همچنین در ۴ فاکتور: حساسیت، دقت، ضریب کاپا و کارایی نتایج بهتری را به ما ارائه داد. می‌توان نتیجه گرفت که ترکیب دو الگوریتم درخت تصمیم‌گیری و ماشین‌بردار پشتیبان باعث افزایش دقت، حساسیت و کارایی مدل می‌شود. معیار AUC که دقت یک طبقه بندی را نشان می‌دهد، برای مدل درخت تصمیم‌گیری ۸۷,۶٪ ، برای مدل ماشین‌بردار پشتیبان ۸۱,۶٪ و برای مدل ترکیبی درخت تصمیم و ماشین‌بردار پشتیبان ۸۸,۸٪ است. بالاتر بودن این مقدار در مدل ترکیبی نشان می‌دهد که کارایی این مدل بیشتر از دو مدل دیگر است. تعداد نمونه‌هایی که در کلاس‌های اشتباه دسته‌بندی شده‌اند، در این مدل کمتر از دو مدل دیگر است و برای کلاس‌های افسردگی زیاد، متوسط و کم به ترتیب تعداد ۰/۰۵۳ ، ۰/۰۳۴ و ۰/۱۱۴ تعداد نمونه به اشتباه دسته‌بندی شدند. با توجه به جدول ۱۵ می‌توان موارد ارزیابی ۳ مدل را مشاهده کرد.

**جدول۱۵: ارزیابی مدل‌های پیش‌بینی ایجادشده در این پژوهش**

| نام مدل | حساسیت | ویژگی | دقت | ضریب کاپا | کارایی |
|---|---|---|---|---|---|
| درخت تصمیم‌گیری | ٪۹۰,۸ | ٪۹۵,۵ | ٪۹۱,۱۳ | ٪۷۴,۲ | ٪۸۷,۶ |
| ماشین‌بردار پشتیبان | ٪۸۱,۱ | ٪۹۲ | ٪۸۳,۳۳ | ٪۵۸,۶ | ٪۸۱,۶ |
| مدل ترکیبی درخت تصمیم و ماشین‌بردار پشتیبان | ٪۹۴,۸ | ٪۹۴,۴ | ٪۹۴,۸۱ | ٪۸۱,۳ | ٪۸۸,۸ |

## بررسی ارتباط بیماری‌های فیزیکی با بیماری افسردگی

برای بررسی میزان تأثیرذاری بیماری‌های فیزیکی بر ریسک مبتلا شدن به افسردگی از ضریب همبستگی استفاده شد و اثرگذاری و ارتباط بین بیماری‌هایی نظیر: بیماری‌های قلبی، گوارشی، هورمونی، میگرن و سایر بیماری‌های فیزیکی مورد پژوهش با افسردگی به دست آمد. نتایج به دست آمده در جدول ۱۶ قابل مشاهده است. نتایج نشان می‌دهد که در افسردگی حاد بیماری‌های کلیوی با ضریب همبستگی ۰/۰۹۸ و بیماری قلبی با ۰/۰۴۸ و در افسردگی خفیف بیماری‌های هورمونی با ضریب همبستگی ۰/۰۵۹ ، بیماری گوارشی با ۰/۰۴۶ و میگرنی با ۰/۰۱۱ بیشترین تأثیر و ارتباط را با افسردگی دارند.

**افسردگی بیماری مرتبط با بیماری‌های همبستگی جدول۱۶: ماتریس**

| Attribut... | . = no | . = gova... | . = kolie | . = hor... | . = ghalbi | . = migr... | . = disk | . = sara... | .. = Med... | .. = Much | .. = low |
|---|---|---|---|---|---|---|---|---|---|---|---|
| . = no | 1 | -0.333 | -0.190 | -0.173 | -0.452 | -0.603 | -0.205 | -0.077 | 0.074 | -0.030 | -0.036 |
| . = govar... | -0.333 | 1 | -0.024 | -0.021 | -0.056 | -0.075 | -0.025 | -0.010 | -0.038 | -0.011 | 0.046 |
| . = kolie | -0.190 | -0.024 | 1 | -0.012 | -0.032 | -0.043 | -0.015 | -0.005 | -0.066 | 0.098 | -0.042 |
| . = horm... | -0.173 | -0.021 | -0.012 | 1 | -0.029 | -0.039 | -0.013 | -0.005 | -0.012 | -0.046 | 0.059 |
| . = ghalbi | -0.452 | -0.056 | -0.032 | -0.029 | 1 | -0.101 | -0.035 | -0.013 | -0.058 | 0.048 | 0.003 |
| . = migren | -0.603 | -0.075 | -0.043 | -0.039 | -0.101 | 1 | -0.046 | -0.017 | -0.021 | 0.008 | 0.011 |
| . = disk | -0.205 | -0.025 | -0.015 | -0.013 | -0.035 | -0.046 | 1 | -0.006 | 0.011 | -0.033 | 0.024 |
| . = sarat... | -0.077 | -0.010 | -0.005 | -0.005 | -0.013 | -0.017 | -0.006 | 1 | 0.082 | -0.040 | -0.034 |
| .. = Medi... | 0.074 | -0.038 | -0.066 | -0.012 | -0.058 | -0.021 | 0.011 | 0.082 | 1 | -0.483 | -0.409 |
| .. = Much | -0.030 | -0.011 | 0.098 | -0.046 | 0.048 | 0.008 | -0.033 | -0.040 | -0.483 | 1 | -0.602 |
| .. = low | -0.036 | 0.046 | -0.042 | 0.059 | 0.003 | 0.011 | 0.024 | -0.034 | -0.409 | -0.602 | 1 |

## مقایسه مدل پیشنهادی با مدل‌های ایجادشده در سال‌های اخیر

در شهر همدان ( سهیلا ملک‌پور و منصور اسماعیل‌پور،۱۳۹۶) به داده‌کاوی ۳۹۱ بیمار مبتلا به افسردگی پرداختند. در پژوهش آنان ویژگی‌هایی نظیر: سن، جنسیت ، مسافرت ، عقاید مذهبی، ورزش و ... مورد بررسی قرار گرفت. آنان با استفاده از الگوریتم‌های درخت تصمیم، الگوریتم شبکه‌های عصبی و رافست به داده‌کاوی داده‌ها پرداختند. دقت مدل شبکه‌های عصبی در پژوهش آنان ۸۸,۴۹٪ ، در مدل رافست ۸۲,۵٪ و در مدل درخت تصمیم ۹۵,۴٪ بود که این مدل به عنوان بهترین مدل ارائه شد (ملک‌پور، اسماعیل پور، ۱۳۹۶).

در آمریکا (شادی بانیتان، کوین دایمی، ۲۰۱۴) به بررسی عوامل خلقی مؤثر بر روی افسردگی پرداختند. در این پژوهش،آنان ویژگی‌های خلقی نظیر: پرخاشگری، اضطراب، میزان گریه کردن، غم و اندوه، بد خلقی، مصرف مواد و ... را مورد بررسی قرار دادند. آنان از الگوریتم C4.5 استفاده کردند.آنان ۶۰۰ نفر را مورد بررسی قرار دادند. مدلی که ارائه دادند، ۹۲,۵٪دقت دارد.

در تحقیقی دیگر(فادا ازاب، مهدی محمدی، ۲۰۱۵) مطالعه‌ای را روی ۵۳ نفر بزرگسال انجام دادند که به این سؤوال پاسخ دهند که آیا از روی نوارهای مغضی افراد میتوان افسرده بودن یا نبودن آنها را تشخیص داد؟ آنها داده های مورد نیاز را از نوارهای مغضی این افراد به دست آورند و با استفاده از الگوریتم های آنالیز تشخیص خطی[1] و ژنتیک به داده‌کاوی آنان پرداختند. مدل ترکیبی

1. LDA

آنان با ۸۰ درصد صحت، ۷۰ درصد حساسیت و ۷۶ درصد ویژگی ارائه شد و نتیجه به دست آمده حاکی از آن بود که نوارهای مغزی می‌تواند معیار خوبی برای تشخیص افراد سالم از افسرده باشد.

**جدول۱۷: مقایسه مدل پیشنهادی با مدل‌های ایجادشده در سال‌های اخیر**

| سال | تعداد ویژگی | تعداد رکورد | برنامه | مدل پیشنهادی | نتیجه |
|---|---|---|---|---|---|
| ۱۳۹۶ | ۸ | ۳۹۱ | اس.پی.اس.اس وکا | الگوریتم درخت تصمیم | دقت : ٪۹۵,۴ حساسیت: ٪۹۱,۶ |
| ۲۰۱۴ | ۲۱ | ۶۰۰ | وکا | C4.5 | دقت: ٪۹۲,۵ |
| ۲۰۱۵ | ۹ | ۵۳ | وکا | ترکیب LDA و ژنتیک | صحت: ٪۸۰ حساسیت: ٪۷۱ ویژگی: ٪۷۶ |
| مدل پیشنهادی در این پایان نامه ۱۳۹۷ | ۱۷ | ۴۷۰ | رپیدماینر اس.پی.اس.اس | ترکیب درخت تصمیم و ماشین‌بردار پشتیبان | دقت: ٪۹۴,۸۱ حساسیت: ٪۹۴,۸ ویژگی: ٪۹۴,۴ کارایی: ٪۸۸,۸ ضریب کاپا: ٪۸۱,۳ |

## نتیجه‌گیری

در این پژوهش، هدف اصلی بررسی پارامترهای سبک زندگی، موسیقی مورد علاقه و ارتباط آن با بیماری افسردگی بوده است تا بتوان ارتباط بین آنها را بررسی و راه‌حلی در جهت درمان ارائه داد. داده‌ها از طریق مراکز روان درمانی و همچنین تست افسردگی بک جمع‌آوری شدند. ۸۵۰ نفر مورد بررسی قرار گرفتند که از این میان، ۴۷۰ نفر افسردگی داشتند که اطلاعات دیتاست اصلی بر روی این ۴۷۰ نفر اعمال شد. دو الگوریتم درخت تصمیم‌گیری و ماشین‌بردار پشتیبان برای کشف الگوهای پنهان روی داده های بیماران افسرده اعمال شدند. نتایج به دست آمده نشان داد که دو فاکتور موسیقی مورد علاقه‌ای که بیمار گوش می‌دهد و میزان ورزشی که می‌کند، بر روی میزان افسردگی بیماران تأثیر زیادی دارد. ویژگی‌های تعداد فرزندان، تک فرزند بودن و فرزند چندم بودن تأثیر کمی بر روی میزان افسردگی افراد دارد و همچنین بیماری‌های فیزیکی نظیر: بیماری‌های قلبی، کبدی و کلیوی در افراد با افسردگی زیاد و بیماری‌های: هورمونی، گوارشی و میگرنی در افراد با افسردگی کم بیشتر دیده می‌شود. نتایج نشان داد که افسردگی حاد در خانم‌ها بیشتر از آقایان است. نتایج پژوهش

حاکی از این است که افراد با افسردگی زیاد تمایل به گوش دادن به سبک‌های غمگین و راک و متال دارند و همچنین اکثراً در شب‌ها موسیقی گوش می‌دهند و از طرفی، گوش دادن به موسیقی پاپ و سنتی در افراد با افسردگی کم و کسانی که افسرده نیستند بیشتر به چشم می‌خورد. گوش کردن به موسیقی‌های غمگین و راک و متال به خصوص شب‌ها می‌تواند میزان افسردگی در افراد را افزایش دهد. از این رو، این سبک از موسیقی به افراد افسرده پیشنهاد نمی‌شود و بهتر است که موسیقی‌های شاد و سنتی ایرانی گوش دهند. موسیقی‌های آرامش‌بخش و شاد در بهبود میزان افسردگی بسیار مؤثر است و همچنین از مبتلا شدن فرد به افسردگی تا حدی جلوگیری می‌کند. تشویق جامعه به ورزش کردن و همچنین گوش کردن به موسیقی‌های آرامش‌بخش و شاد در طول روز می‌تواند راه حلی برای داشتن جامعه‌ای سالم‌تر، شادتر و به دور از افسردگی باشد.

## منابع


Akthar F, Hahne C.RapidMiner 5 Operator Reference.2012.

Alizadehsani R, Habibi J, Hosseini MJ, Mashayekhi H, Boghrati R, Ghandeharioun A, Bahadorian B, Sani ZA. A data mining approach for diagnosis of coronary artery disease. 2013. Comput Methods Programs Biomed *111:52–61. [PubMed: 23537611]*.

Caddy C, Amit B.H, McCloud T.L et al. Ketamine and other glutamate receptor modulators for depression in adults. 2015. *Cochrane Database of Systematic Reviews*. no. 9. article CD011612.

ChaitraliDangare S, SulabaApte S. Improved study of disease prediction using data mining classification techmiques. 2010. Int.J.Comp.Appl.

Chen L, Zhou S, Bryant J. Temporal changes in mood repair through music consumption: Effects of mood, mood salience, and individual differences. 2007. Media Psychology.

Corporation T C .*Introduction to Data Mining and Knowledge Disocvery*.2005

Daimi K, Banitaan S. Using Data Mining to Predict Possible Future Depression Cases. 2014. Vol.3. No.4.

Dipnall Joanna F, Pasco Julie A and Berk M. Fusing Data Mining, Machine Learning and Traditional Statistics to Detect Biomarkers Associated with Depression. 2016.

Esfandiari N, Mansouri S. The effect of listening to light and heavy Music on reducing the symptoms of depression among female students. 2014. *Arts Psychother. 41. 211–213.*

FaiChan M, EstherMok. Effects of music on depression and sleep quality in elderly people: A randomised controlled trial. 2010. Complementary Therapiesin Medicine.

Halaris A. A primary focus on the diagnosis and treatment of major depressive disorder in adults. 2011.

Han J, Kamber M, Pei J. Data mining: concepts and techniques: Morgan kaufmann. 2006.

Hossein zadeh S. Data mining and its application to health at the first specialized conference on electricity and computers. 1363. No.4. In persian

Huang Yu-Jhen, Hsien-Yuan Lane, and Chieh-Hsin Lin. New Treatment Strategies of Depression: Based on Mechanisms Related to Neuroplasticity. 2017. Article ID 4605971. 11 pages.

Kharya S. Using data mining techniques for diagnosis and prognosis of cancer disease.2012. Int. J. Comput. Sci. Inf. Technol. *vol. 2. no. 2. pp*. 55–66.

Leubner D, Hinterberger T. Reviewing the Effectiveness of Music Interventions in Treating Depression. 2017.

Malekpour S, Esaeilpour M. Investigating Factors Affecting Depression in People Using Data Mining Methods.1396. In Persian

Mathew RJ, Largen J, Claghorn JL. Biological Symptoms of Depression. 1979. Psychosomatic Medicine. vol/issue: 41(6). pp. 439-443.

Milovic B, Milovic M. Prediction and Decision Making in Health Care using Data Mining. *2012. International Journal of Public Health Science (IJPHS)*. vol/issue: 1(2). pp. 69-78.

Milovic B, Milovic M. Prediction and Decision Making in Health Care using Data Mining. *2012. International Journal of Public Health Science (IJPHS)*. vol/issue: 1(2). pp. 69-78.

Mohammadi M, Al-Azab F, Raahemi B. Data mining EEG signals in depression for their diagnostic value. 2015

Perez S, Perez V. Affects of music therapy on depression compared with psychotherapy. 2010. *The Arts in Psychotherapy 37*. 387–390.

Taneja A. Heart Disease Prediction System Using Data Mining Techniques. 2013.

Yoon S, Bakken S. Using a Data Mining Approach to Discover Behavior Correlates of Chronic Disease: A Case Study of Depression.2014. *Stud Health Technol Inform*.