

مهندسی برق و کامپیوتر

چشم انداز علمی
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

بررسی و مقایسه اثربخشی الگوریتم های داده کاوی جهت پیش بینی بیماری پارکینسون

Comparison of the effectiveness of data mining algorithms to predict Parkinson's disease

محسن غلامی^{۱*}، ناصر نجفی^۲

۱- دانشجوی کارشناسی ارشد رشته نرم افزار دانشگاه آزاد اسلامی واحد بوشهر، Mohsen.gholami18@yahoo.com

۲- کارشناس مهندس صنایع، Naser.najafi@aol.com

خلاصه

بیماری پارکینسون پس از بیماری آلزایمر، شایع ترین بیماری در آسیب رسانی به مغز می باشد. این بیماری باعث می گردد ترشح دوپامین در مغز با مشکل روبه رو شده و از آن جایی که وظیفه دوپامین تسهیل و تنظیم حرکات بدن است، بدن با مشکلات حرکتی و کندی مواجه می شود. تحقیقات علمی نشان می دهد، حدود چهار درصد از بیماران با گذشت زمان، دچار اختلال حافظه نیز می شوند. جالب این که ده درصد این بیماران دارای سنین کمتر از چهل سال هستند. در این پژوهش سعی گردیده تا عوامل موثر بر تشخیص با استفاده از فرکانس صوتی بیماران شناسایی و با مقایسه انواع الگوریتم های داده کاوی، پارامترهای مختلف بررسی و شناسایی گردند. داده ها از مخزن یادگیری UCI دریافت شده و شامل بیست و سه متغیر می باشد.

کلمات کلیدی: طبقه بندی، کاپای کوهن، بیز ساده، درخت تصمیم، شبکه های عصبی، نزدیک ترین همسایه

۱. مقدمه

گسترش حوزه داده کاوی به دلیل ادغام تکنیک ها و حوزه هایی مانند یادگیری ماشین، آمار، تشخیص الگو، هوش مصنوعی و پایگاه داده جهت کار با حجم بالای داده ها است. از این رو است که تعداد زیادی از الگوریتم های داده کاوی موجود برگرفته از این علوم است. [1]

مسائل پیچیده را می توان با شکستن به مسایل ساده تر، بررسی، مدل سازی و حل نمود. ما نیز در این مقاله، سعی در ساده سازی و کاهش متغیرهای غیر ضروری جهت پیش بینی وجود بیماری در افراد مورد مطالعه را داریم.

آدرس ایمیل و نام نویسنده مسئول: محسن غلامی

Mohsen.gholami18@yahoo.com

مهندسی برق و کامپیوتر

پیشروان علمی
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

با روند رو به رشد تولید داده و اطلاعات، دسترسی به مخازن داده‌های مورد نیاز در رسته‌های مختلف تسهیل شده است و برای بهره برداری از این حجم از داده و کشف روابط بین آنها، دانش داده‌کاوی توسعه فراوان یافته است. داده‌کاوی از ترکیب تکنیک‌های آماری و یادگیری ماشین، پایگاه داده، محاسبات و تطبیق الگو جهت به کارگیری در زمینه‌های مختلف استفاده می‌کند. [2]

بیماری پارکینسون یک بیماری کمتر شناخته شده است، همین امر باعث شده متاسفانه بسیاری از افراد بر این باور باشند که علائم ناشی از آن با افزایش سن به وجود می‌آید اما باید به این نکته توجه داشت که تقریباً ده درصد مبتلایان به این بیماری افراد زیر چهل سال هستند. این باور نادرست باعث شده تا شناسایی و غربالگری در مراحل اولیه و درست زمانی که بیماری علایم پنهان تری از خود نشان می‌دهد، کمتر مورد توجه قرار گیرد.

اختلالات زمانی هنگامی رخ می‌دهد که نواحی خاصی از مغز توانایی خود در تولید دوپامین (یکی از ناقلین عصبی در مغز) را از دست می‌دهند. این بیماری یک بیماری عصبی بوده و لرزش، سفتی عضلات، کندگی و افسردگی از نشانه‌های شایع این بیماری است. [3]

این پژوهش به بررسی و مقایسه الگوریتم‌های برتر داده‌کاوی در خصوص شناسایی این بیماری می‌پردازد و سعی دارد تا با بررسی پارامترهای مختلف، الگوریتم‌های برتر را پیشنهاد دهد. مدل سازی از طریق داده‌های مخزن یادگیر UCI که شامل صدای ضبط شده از افراد است، صورت گرفته و این مقاله نیز بر شناسایی کاراترین الگوریتم جهت شناسایی بیماران از طریق فرکانس صوتی متمرکز شده است.

dataacademy.ir

۲. عوامل شایع بیماری

- لرزش، به خصوص در حالت عدم حرکت اندام در دست و بازو و پاها و فک در حدود هفتاد تا نود درصد بیماران [3]
- سفتی عضلانی و کندگی حرکت در کل بدن
- راه رفتن نامتناسب به حالتی که پاها به زمین کشیده شده و فاصله پاها بیشتر از حالت طبیعی است.
- قامت خمیده
- از بین رفتن حالت چهره
- تغییرات صدا؛ صدا ضعیف و بم می‌شود.
- اختلال بلع، آبریزش دهان
- توانایی ذهنی تا مراحل پیشرفته، بدون تغییر می‌ماند و در مراحل پیشرفته به آهستگی کاهش می‌یابد.
- افسردگی، عصبی بودن

در این مقاله سعی داریم در ابتدا پارامترهایی همچون Accuracy و Kappa را مورد بررسی قرار داده و پس از آن به مقایسه الگوریتم‌های مختلف بپردازیم.

مهندسی برق و کامپیوتر

پیشروان علمی
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

۳. مفاهیم اولیه

۳-۱. فرمول ها و محاسبات

Accuracy: دقت تست درصد درستی طبقه بندی مجموعه را مشخص می کند.

(1)

$$Accuracy = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}} * 100$$

فرمول محاسبه Accuracy

Recall (Sensitivity): حساسیت، نسبت تعداد نمونه های مثبتی است که الگوریتم نیز آن ها را درست تخمین زده به

کل نمونه های مثبت واقعی.

(2)

$$Sensitivity = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} * 100\%$$

فرمول محاسبه Recall

Precision: نسبت میزان نمونه هایی که مقدارشان مثبت بوده و سیستم نیز به درستی پیش بینی کرده به روی کل نمونه هایی که سیستم به عنوان مثبت تخمین زده است. به عبارتی منظور از کل تخمین مثبت در اینجا مجموع نمونه های مثبتی است که سیستم نیز برچسب مثبت به آن داده با نمونه هایی است که سیستم به اشتباه برچسب مثبت به آن داده است.

(3)

$$Precision = \frac{\text{number of true positives}}{\text{number of true positives} + \text{false positives}} * 100\%$$

فرمول محاسبه Precision

Specificity: نسبت میزان نمونه هایی است که مقدارشان منفی بوده و سیستم به درستی مقدار منفی را برای آن پیش بینی کرده است به روی مجموع نرخ منفی صحیح پیش بینی شده و نمونه هایی که به اشتباه توسط سیستم مقدار مثبت داده شده است.

در واقع می توان Specificity را نسبت نمونه هایی که برچسب منفی داشته و سیستم نیز به درستی برچسب منفی به آن ها داده است به کل نمونه هایی که برچسب هایی که مقدار منفی دارند.

(4)

$$Specificity = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} * 100\%$$

فرمول محاسبه Specificity

Confusion Matrix: این ماتریس رابطه مابین نتایج واقعی و پیش‌بینی را نشان می‌دهد.

۲-۳. انواع الگوریتم‌ها

الگوریتم‌های طبقه‌بندی:

الگوریتم‌های آموزشی که به صورتی گسترده استفاده می‌شوند، حداقل دارای دو مرحله آموزشی و تست هستند در مرحله آموزشی مدلی ایجاد گردیده و عملکرد آن در مرحله دوم توسط مجموعه داده تست سنجیده می‌شود. [4] نحوه کار در الگوریتم‌های طبقه‌بندی به این صورت است که در مرحله اول از مجموعه داده آموزشی با استفاده از الگوریتم‌های طبقه‌بندی مدلی ایجاد می‌گردد که با استفاده از این مدل، پیش‌بینی وضعیت سایر نمونه‌ها امکان پذیر می‌شود. آنچه مسلم است این است که در این مرحله ما جامعه آماری را در اختیار داریم که دارای چندین متغیر مستقل و یک متغیر پیوسته است که این متغیر، صفت خاصه هدف نیز نامیده می‌شود.

در مرحله تست، هدف ما تخمین کارایی سیستم از جنبه‌های مختلف است، در نتیجه ما با مجموعه داده‌ای که مجموعه آزمایشی نامیده می‌شود و دارای تمامی متغیرهای مستقل مشابه مجموعه آموزشی است، سروکار داریم. این مجموعه داده فاقد متغیر وابسته (صفت خاصه هدف) است چرا که پیش‌بینی مقدار این متغیر با استفاده از مدل ایجاد گردیده در مرحله اول صورت خواهد گرفت.

dataacademy.ir

خوشه‌بندی

خوشه‌بندی تکنیکی است که ایجاد خوشه‌هایی با توجه به میزان شباهت بین داده‌ها را ممکن می‌سازد. در واقع در هر خوشه، اعضای آن با هم، شباهت دارند و از اعضای سایر خوشه‌ها قابل تشخیص هستند. در این مدل، نتایج از قبل مشخص نبوده و با استفاده از معیارهای آماری مثل فاصله بین داده‌ها خوشه‌بندی صورت می‌گیرد. [5]

خوشه‌بندی با طبقه‌بندی که در ادامه به آن پرداخته خواهد شد، تفاوت‌هایی دارد مثلاً در طبقه‌بندی اشیا با توجه به طبقات از پیش تعریف شده در طبقات مناسب قرار می‌گیرند [6] اما در خوشه‌بندی، از قبل هیچگونه طبقه‌بندی یا مدل تعریف شده‌ای وجود ندارد که به آن روش بدون نظارت نیز گفته می‌شود.

K-Means

این الگوریتم در سال ۱۹۶۷ توسط James Macqueen مورد استفاده قرار گرفت. از خصوصیات بارز این الگوریتم سادگی آن است که باعث شده بیشتر از سایر الگوریتم‌های خوشه‌بندی مورد استفاده قرار بگیرد.

روند کار به صورت انتخاب k نقطه به عنوان نماینده k خوشه به صورت اولیه و تخصیص نمونه‌های باقیمانده به یکی از خوشه‌ها و محاسبه مجدد مرکز خوشه و در نهایت تکرار این فرآیند تا زمانی که مرکز خوشه تغییر نکند یا تغییرات آن به کمتر از یک درصد برسد، می‌باشد. [7]

مهندسی برق و کامپیوتر

پیشینه علمی
اولین نشر ایرانی

1st International Conference New Perspective in
Electrical & Computer Engineering

طبقه بندی

طبقه بندی روشی است در داده کاوی که وظیفه دسته بندی با توجه به متغیر هدف را بر عهده دارد. هدف از طبقه بندی، پیش بینی متغیر هدف برای همه داده ها است. طبقه بندی را یادگیری با ناظر نیز گویند. در این روش ابتدا با استفاده از مجموعه داده های آموزشی که شامل تعدادی متغیر و دارای یک متغیر هدف است مدلی ایجاد گردیده که قادر به پیش بینی صفت خاصه هدف در مجموعه داده آزمایشی و رکوردهای آینده که فاقد صفت هدف است می باشد. [8]

الگوریتم بیز ساده

در هنگام کار با دیتاست های بزرگ، این الگوریتم با دقت بالایی می تواند کارایی مناسبی به دست دهد. روش کار آن طبقه بندی با استفاده از قضیه بیز می باشد. این الگوریتم نوعی یادگیری با ناظر می باشد. [9]

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

قضیه بیز

$P(h)$ = احتمال اولیه ای که فرضیه h قبل از مشاهده مثال آموزشی D داشته است.

$P(D)$ = احتمال اولیه ای که داده آموزشی D مشاهده خواهد شد.

$P(D|h)$ = احتمال مشاهده داده آموزشی D به فرض آنکه فرضیه h صادق باشد.

dataacademy.ir

درخت تصمیم

روش طبقه بندی داده ها به صورت نظارت شده است که هدف آن استفاده از قوانین تصمیم گیری جهت رسیدن به متغیر هدف است. ساختار آن مانند درخت بوده از ریشه شروع شده و به گره برگ می رسد. این روش ساده و به راحتی قابل درک و تفسیر بوده و امکان هرس کردن تصمیم هایی که قابل تعمیم نیستند را دارد. [10] درخت تصمیم از معیارهای متفاوتی بسته به روش انتخابی جهت انتخاب گره ها استفاده می کند که یکی از آن ها استفاده از بهره اطلاعاتی می باشد. از سایر معیارها می توان $Gain$ ، $Index$ ، $Ratio$ ، $Likelihood$ و DKM را نام برد.

معیار Information Gain

از معروف ترین معیارهای ساخت درخت و انتخاب گره ها که خود بر اساس معیار آنتروپی بنا شده $Information$ $Gain$ می باشد. (6)

$$Information\ Gain(A) = Entropy(D) - Entropy_A(D)$$

$$Entropy(D) = - \sum_{i=1}^c P_i \times \log_2(P_i)$$

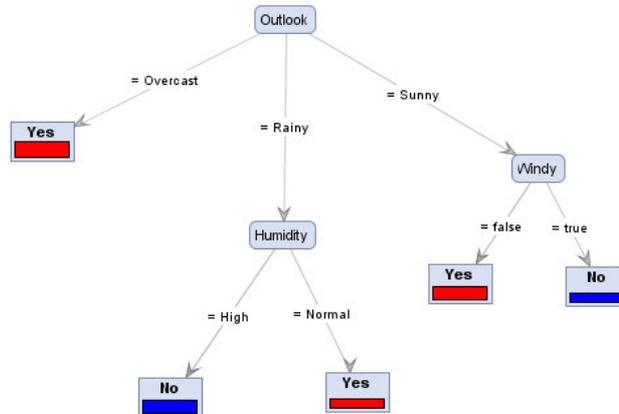
$$Entropy_A(D) = \sum_{j=1}^v \frac{|D_j|}{D} \times Entropy(D_j)$$

محاسبه بهره اطلاعاتی و آنتروپی

مهندسی برق و کامپیوتر

پیشروان علمی
اولین نشر ایرانی

1st International Conference New Perspective in
Electrical & Computer Engineering



شکل ۱- ساختار درخت تصمیم

ضریب همبستگی پیرسون

این تکنیک آماری میزان همبستگی خطی بین دو متغیر را نشان می‌دهد که مقدار آن بین -۱ و ۱ است. مقدار یک، نشان دهنده همبستگی کامل و ضریب صفر عدم همبستگی و ضریب -۱ نشان دهنده نسبت منفی کامل بین دو متغیر است. [11]

(7)

$$r_{X,Y} = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

همبستگی بین دو متغیر تصادفی X و Y

(8)

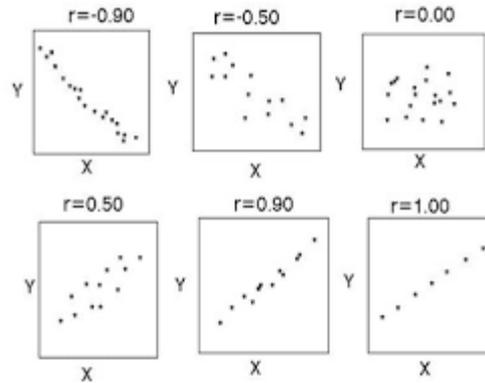
$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

ضریب همبستگی نمونه آماری پیرسون

مهندسی برق و کامپیوتر

پیشروان علمی
اوکنین آری

1st International Conference New Perspective in
Electrical & Computer Engineering



شکل 2- پراکندگی قرارگیری داده‌ها بر روی محور مختصات و مقدار همبستگی

۴. داده‌ها

داده‌های این بررسی، از مخزن یادگیری سایت UCI دریافت شده و شامل ۲۳ متغیر جهت تجزیه و تحلیل بیماران پارکینسون است. این مجموعه داده شامل طیف وسیعی از اندازه‌گیری صدای پزشکی و ارزیابی ۳۱ نفر که ۲۳ نفر از آن مبتلا به بیماری پارکینسون می‌باشند. هدف ما پیش بینی و جداسازی بیماران پارکینسون از افراد سالم است و به افراد سالم در متغیر Status مقدار صفر و به افراد مبتلا به پارکینسون مقدار یک تعلق خواهد گرفت. تقریباً از هر بیمار شش بار صدا ضبط شده است.

جدول ۱- متغیرهای موجود در مخزن داده

RD	نام متغیر	توضیحات
1	MDVP: Fo(Hz)	میانگین فرکانس صوتی
2	MDVP: Fhi(Hz)	حداکثر فرکانس صوتی
3	MDVP: Flo(Hz)	حداقل فرکانس صوتی
4	MDVP: Jitter(%)	تغییرات فرکانس صوتی به عنوان درصد
5	MDVP: Jitter(Abs)	تغییرات فرکانس صوتی در میکرو ثانیه
6	MDVP: RAP	تغییرات فرکانس صوتی (دامنه نسبی)
7	MDVP: PPQ	تغییرات فرکانس صوتی
8	Jitter:DDP	تغییرات فرکانس صوتی
9	MDVP: shimmer	معیار مربوط به تغییرات دامنه
10	MDVP: shimmer(Db)	
11	Shimmer: APQ3	
12	Shimmer: APQ5	
13	MDVP: APQ	
14	Shimmer: DDA	
15	NHR	معیار وجود نویز
16	HNR	معیار وجود نویز

مهندسی برق و کامپیوتر

پیشروان دانش
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

17	RPDE	
18	DFA	تجزیه و تحلیل نوسانات Detrended
19	Spread1	اندازه گیری غیر خطی فرکانس
20	Spread2	
21	D2	اندازه گیری همبستگی
22	PPE	اندازه گیری غیر خطی فرکانس
23	Status	متغیر هدف مقدار ۰ برای افراد سالم و مقدار ۱ افراد مبتلا به پارکینسون

۵. آموزش و تست

در این بخش، سعی شده تا برخی از الگوریتم‌های مختلف داده‌کاوی از جهات مختلف و با معیارهای متفاوت ارزیابی گردند. بدیهی است تشریح محاسبات همه این الگوریتم‌ها در این مقاله امکان پذیر نمی باشد.

۵-۱. الگوریتم K-NN

با تست مدل، نتایج بدست آمده به صورت ماتریس Confusion Matrix پیاده شده که محاسبه پارامترهای مختلف آن به صورت ذیل انجام شده است. این ماتریس مقادیر و نتایج مختلف را نشان می دهد.

جدول ۲- نتایج حاصل از مدل سازی الگوریتم K-NN dataacademy.ir

	true 1	true 0
pred. 1	144	21
pred. 0	3	27

(9)

$$Accuracy = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}} * 100$$

$$Accuracy = \frac{144+27}{144+27+21+3} * 100 = 87.69$$

محاسبه Accuracy

با انجام محاسبه، مقدار کاپای کوهن مقدار آن ۰.۶۲۱ بدست آمد. این معیار هرچه به یک نزدیک تر باشد، عملکرد الگوریتم قابل قبول تر است.

جهت محاسبه Recall به این صورت عمل می گردد.

(10)

مهندسی برق و کامپیوتر

پیشروان علمی
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

$$\text{Sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} = 97.96\%$$

$$\frac{144}{144 + 2} * 100 = 97.96$$

$$\frac{27}{27 + 21} * 100 = 56.25$$

محاسبه Sensitivity

۵-۲. الگوریتم درخت تصمیم

از جمله خصوصیات این الگوریتم استفاده از معیارهای متفاوت در انتخاب گره مناسب است. در این مقاله کارایی انتخاب گره بر اساس Gini index و Information Gain مقایسه گردیده است. مقایسه هر دو درخت با اینکه معیار Gini Index دقت بیشتری نسبت به معیار Information Gain نشان می دهد اما این روش خطا و تلورانس احتمالی بیشتری دارد. در نهایت با مقایسه معیار Kappa متوجه می شویم معیار Gini Index تا حدودی عملکرد بهتری از خود نشان می دهد.

جدول ۲- ماتریس Confusion Matrix جهت روش تصمیم با روش انتخاب گره Gini Index

	true 1	true 0
pred. 1	133	9
pred. 0	14	39

(11)

$$\text{Accuracy} = \frac{133 + 39}{133 + 9 + 14 + 39} * 100 = 88.21$$

محاسبه Accuracy

(12)

$$\frac{133}{133 + 14} * 100 = 90.48$$

$$\frac{39}{39 + 9} * 100 = 81.25$$

محاسبه Recall

جدول 3- Confusion Matrix مدل درخت تصمیم با انتخاب گره به روش Information Gain

	true 1	true 0
pred. 1	129	12
pred. 0	18	36

مهندسی برق و کامپیوتر

چشم اندازهای نو
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

(13)

$$\text{Accuracy} = \frac{129+36}{129+12+18+36} * 100 = 84.62$$

محاسبه Accuracy

$$\frac{129}{129+18} * 100 = 87.76$$

$$\frac{36}{36+12} * 100 = 75.00$$

۳-۵. الگوریتم شبکه های عصبی

جدول 4- ماتریس Confusion و نتایج بدست آمده به شرح زیر می باشد

	true 1	true 0
pred. 1	140	5
pred. 0	7	43

نتایج حاصله در اجرای الگوریتم شبکه های عصبی نشان می دهد از ۱۹۵ مورد الگوریتم فقط ۱۲ مورد را درست تشخیص نداده است.

dataacademy.ir

(14)

$$\text{Accuracy} = \frac{140+43}{140+5+7+43} * 100 = 93.85$$

محاسبه Accuracy

$$\frac{140}{140+7} * 100 = 95.24$$

$$\frac{43}{43+5} * 100 = 89.58$$

۶

. مدل سازی و مقایسه

مهندسی برق و کامپیوتر

پیشروان علمی
اولین کنفرانس بین المللی

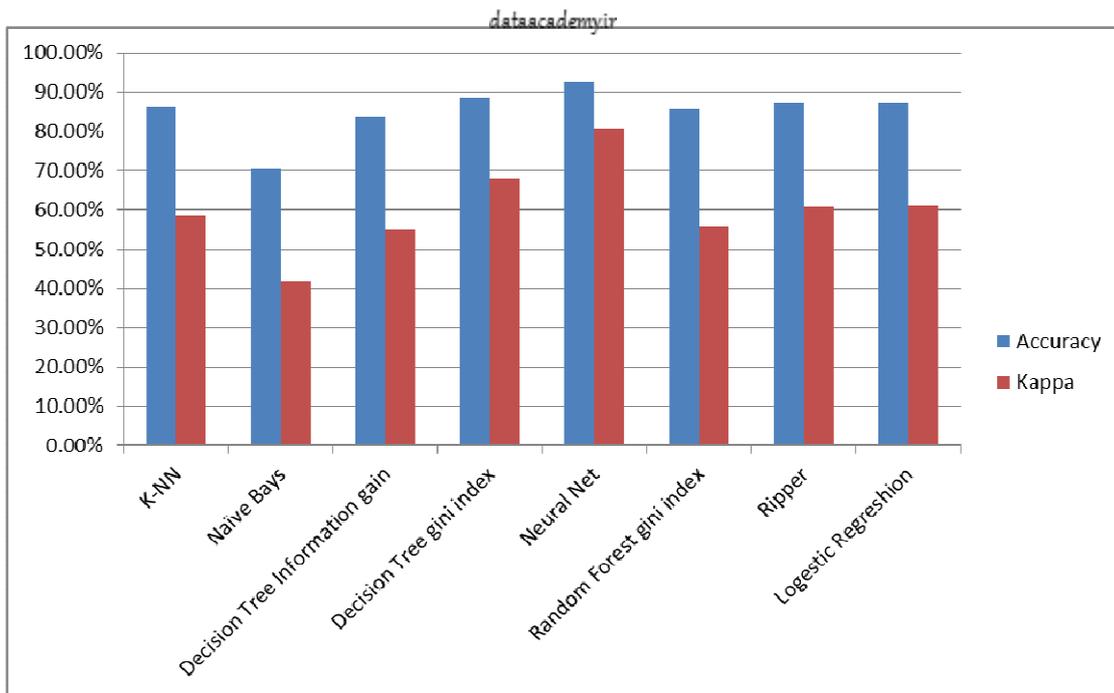
1st International Conference New Perspective in
Electrical & Computer Engineering

معیار Kappa ذکر شده است.

جدول 5- جدول نتایج به دست آمده و مقایسه الگوریتمها

RD	Algorithm	Accuracy	Recall	Precision	Kappa
1	K-NN	87.69% +/- 3.77%	77.31% +/- 5.66%	88.57% +/- 5.50%	0.621
2	Naive Bays	70.26% +/- 4.47%	77.36% +/- 2.28%	70.67% +/- 1.39%	0.408
3	Decision Tree Information Gain	84.62% +/- 3.63%	81.17% +/- 2.77%	81.51% +/- 6.75%	0.605
۴	Decision Tree Gini Index	88.21% +/- 5.28%	85.90% +/- 6.62%	84.96% +/- 7.63%	0.697
۵	Neural Net	93.85% +/- 4.47%	92.51% +/- 5.35%	91.96% +/- 6.49%	0.839
۶	Random Forest Gini index	87.69% +/- 4.70%	77.97% +/- 6.74%	87.60% +/- 7.17%	0.627
۷	Ripper	87.18% +/- 5.38%	78.04% +/- 10.53%	86.60% +/- 8.97%	0.607
۸	Logistic Regression	87.18% +/- 4.59%	77.66% +/- 7.52%	86.65% +/- 7.53%	0.610

با مقایسه الگوریتم های مختلف با توجه به معیارهای متفاوت الگوریتم، شبکه های عصبی کارایی بهتری از خود نشان می دهد. معیار دیگری که باید مورد توجه قرار گیرد معیار Kappa است که هر چه به یک نزدیک تر باشد، الگوریتم کارایی بهتری را از خود نشان داده است.



شکل 3- مقایسه دو پارامتر Accuracy و Kappa در الگوریتم های مختلف

۷. بحث و نتیجه گیری

مهندسی برق و کامپیوتر

چشم اندازهای نو
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

با شناسایی زودهنگام بیماری می توان به بهبود زودتری در روند رشد این بیماری که با کندی نمودن روند پیشرفت آن منتهی می شود، دست یافت. یکی از راه های شناسایی افراد مبتلا، مقایسه فرکانس صوتی در شرایط لازم جهت تشخیص به موقع بیماری را فراهم نمود. مقایسه فرکانس صوتی در شرایط کنترل شده است که این مقاله به بررسی جنبه های مختلف کارایی این الگوریتم ها در پردازش این داده ها و تشخیص بیماری پرداخته و نتایج حاصله نیز مقایسه و بررسی شده است. با مقایسه الگوریتم های برتر داده کاوی جهت پیش بینی بیماران مبتلا به پارکینسون، این نتیجه حاصل شده که الگوریتم شبکه های عصبی کارایی بهتری از خود نشان داده و با دقت ۹۲.۸۲٪، میانگین تلورانس خطای ۴.۱۸٪ در هر بار آزمایش و نرخ کاپای کوهن ۰.۸۰۸ کارایی بهتری از خود نشان داده است. از سوی دیگر با توجه به ماهیت این مدل، زمان مورد نیاز جهت پردازش بیشتر از سایر مدل ها بوده و سرعت محاسبه نتایج، پایین تر از سایرین است. امید است با تهیه دیتاست کشوری و استفاده از روش ذکر شده در این مقاله، بتوان با تشخیص به موقع این بیماری، اقدام های عملی مناسبی را در کشور شاهد باشیم.

۸. مراجع

1. Wu, x. and Kumar, v. and et al. (2008), "Top 10 algorithms in data mining," Springer-verlag, 1-37.
2. Ashish Patel, J. (2015), "Classification Algorithms and Comparison in Data Mining," International Journal of Innovations & Advancement in Computer Science, vol 4, 206-210.
3. Ganesh, H. and Annamary, G. (2014), "Comparative study of Data Mining Approaches for Parkinson's Diseases," International Journal of Advanced Research in Computer Engineering & Technology, Vol 3 Issue 9, 3062-3068.
4. Venkata Ramana, B. and Surendra Prasad Babu, M. and et al. (2011), "A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis," International Journal of Database Management Systems, Vol 3, No 2, 101-114.
5. Kaur Mann, A. and et al. (2013), "Survey Paper on Clustering Techniques," International Journal of Science, Engineering and Technology Research, Vol 2, Issue 4, 803-806.
6. Krishnaiah, V. and Narsimha, G. and Subhash Chandra, N (2016), "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review," International Journal of Computer Applications, Vol 136, No 2, 43-51.
7. Guru Rao, C.V. and Sreenivasa Rao, M (2016), "Cluster Analysis of Medical Research Data using R," Global Journal of Computer Science and Technology: C Software & Data Engineering, Vol 16, 17-22.
8. Purusothaman, G. and Krishnakumari, P (2015), "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease," Indian Journal of Science and Technology, Vol 8(12), 2-5.
9. Mir Abedini, S.J. and Gholami, M (2016), "A Method For Accuracy Of Heart disease Diagnosis (MAWB)," Computer Engineering And IT Management, 3th, Tehran, 1-8.

مهندسی برق و کامپیوتر

چشم اندازهای نو
اولین کنفرانس بین المللی

1st International Conference New Perspective in
Electrical & Computer Engineering

10. Venkatalakshmi, B. and Sivarajah, S. (2016), "Cancer Diagnosis Using Predictive Data mining," International Journal of Innovative Research in Science, Engineering and Technology, Vol 3, Issue 3, 17–22.
11. MirAbedini, S.J. and Gholami, M. and Najafi, N. (2016), "A method for simultaneous use of pearson linear coordination and data-minig algorithms to improve the diagnosis of cancerous tomor type forecasting," International Conference on Computer Science, Electrical and Electronics Engineering, 5th, 1097–1105.

dataacademy.ir