سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)

Summer 2024, 5 (2), 119-135
DOR:

Received: 28 Apr 2024
Accepted: 9 June 2024

مقاله پژوهشی

Enhancing Adaptive Learning Systems: A Probabilistic Approach Using Bayesian Networks and User Clustering

Parvaneh Ahmadifard[1], Mojtaba Salehi[2*], Mohamad Yarahmadi[3], Morteza Zakeri Nasr Abadi[4]

۱. MSc, Department of Computer Engineering, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran.
2. Instructor, Department of Computer Engineering, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran.
**Corresponding Author*, Mo.Salehi@iau.ac.ir
3. Instructor, Department of Math, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran.
4. Assistant Professor, School of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.

Abstract

Introduction: This study proposes an innovative framework to enhance the quality and adaptability of web-based educational systems using Bayesian networks. The proposed methodology aims to address issues like the cold-start problem and improve the scalability and personalization of educational services.

Method: The proposed methodology encompasses feature selection through the Relief algorithm, user clustering via k-means, and probabilistic service recommendation based on Bayesian inference. Feature selection prioritizes user-specific attributes, reducing data dimensionality and enhancing computational efficiency. Clustering organizes users into homogeneous groups based on similarities, facilitating tailored service provision. The system was validated using a dataset of 2000 users, each characterized by 16 features and linked to 24 services.

Results: The experimental results demonstrated an accuracy rate of 88.15%, underscoring the model's effectiveness compared to traditional techniques. Bayesian networks enabled the analysis of probabilistic relationships and uncertainties, ensuring that the most relevant services were recommended to each user.

Discussion: The proposed approach addressed the cold-start problem, improving scalability and adaptability to diverse user needs. Key advantages of this framework include enhanced personalization, computational efficiency, and significant improvements in user satisfaction. The study's findings highlight the potential of intelligent learning systems to adjust dynamically to user requirements, paving the way for more engaging and effective educational experiences. Future research will incorporate heuristic algorithms to improve prediction accuracy and integrate real-time user feedback to refine the recommendation process further.

Keywords: Intelligent tutoring system, feature selection, Bayesian network, K-means algorithm**.**

سامانه‌های پردازشی و ارتباطی چندرسانه‌ای هوشمند
Intelligent Multimedia Processing and Communication Systems (IMPCS)
انجمن علمی تجارت الکترونیکی ایران
واحد نجف‌آباد

بهبود سیستم‌های یادگیری تطبیقی: یک رویکرد احتمالی با استفاده از شبکه‌های بیزین و خوشه‌بندی کاربر

تاریخ دریافت: ۰۹/ ۰۲/ ۱۴۰۳
تاریخ پذیرش: ۲۰/ ۰۳/ ۱۴۰۳

دورهٔ پنجم، تابستان ۱۴۰۳
شمارهٔ دوم، صص: ۱۱۹-۱۳۵

پروانه احمدی فرد[۱]، مجتبی صالحی[۲*]، محمد یاراحمدی[۳]، مرتضی ذاکری[۴]

۱- کارشناسی ارشد، گروه کامپیوتر، واحد خرم آباد، دانشگاه آزاد اسلامی، خرم آباد، ایران. ahmadicom2004@yahoo.com
۲- مربی، گروه کامپیوتر، واحد خرم آباد، دانشگاه آزاد اسلامی، خرم آباد، ایران. (نویسندهٔ مسئول) mo.salehi@iau.ac.ir
۳- مربی، گروه ریاضی، واحد خرم آباد، دانشگاه آزاد اسلامی، خرم آباد، ایران. mohammadyarahmadi.1981@iau.ac.ir
۴- استادیار، دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران)، تهران، ایران. zakeri@aut.ac.ir

چکیده: این پژوهش یک چارچوب نوآورانه برای ارتقای کیفیت و انطباق‌پذیری سامانه‌های آموزشی مبتنی بر وب با استفاده از شبکه‌های بیزین ارائه‌می‌دهد. روش پیشنهادی شامل انتخاب ویژگی از طریق الگوریتم Relief، خوشه‌بندی کاربران با استفاده از الگوریتم k-means و پیشنهاد خدمات به‌صورت احتمالاتی بر اساس استنتاج بیزین است. انتخاب ویژگی با اولویت‌بندی ویژگی‌های کلیدی کاربران، ابعاد داده‌ها را کاهش‌داده و کارایی پردازشی را افزایش‌می‌دهد. خوشه‌بندی، کاربران را بر اساس شباهت‌هایشان در گروه‌های همگن سازمان‌دهی‌کرده و ارائه خدمات شخصی‌سازی‌شده را تسهیل‌می‌کند. این سامانه با استفاده از یک مجموعه داده شامل ۲۰۰۰ کاربر که هرکدام دارای ۱۶ ویژگی و ۲۴ خدمت بودند، ارزیابی‌شد. شبکه‌های بیزین روابط احتمالاتی و عدم‌قطعیت‌ها را تحلیل‌کرده و خدمات مرتبط را به هر کاربر پیشنهادمی‌دهند. این روش مشکل آغاز سرد را برطرف‌کرده و انطباق‌پذیری و مقیاس‌پذیری را برای نیازهای متنوع کاربران بهبود می‌بخشد. نتایج آزمایش‌ها دقتی معادل ۸۸،۱۵ درصد را نشان‌داد که مؤثر بودن مدل در مقایسه با روش‌های سنتی را تأییدمی‌کند. مزایای کلیدی شامل بهبود شخصی‌سازی، افزایش کارایی محاسباتی و ارتقای رضایت کاربران است. یافته‌های این مطالعه، توانایی سامانه‌های آموزشی هوشمند را برای تطبیق پویا با نیازهای کاربران نشان‌داده و زمینه‌ساز تجربه‌های آموزشی جذاب‌تر و مؤثرتر می‌شود. پژوهش‌های آتی بر ادغام الگوریتم‌های فرابتکاری برای پیش‌بینی دقیق‌تر و دریافت بازخورد بلادرنگ کاربران به منظور بهبود فرآیند پیشنهاد متمرکز خواهندبود.

واژه‌های کلیدی: سیستم آموزش هوشمند، انتخاب ویژگی، شبکه بیزین، الگوریتم .K-Means

۱. مقدمه

تحولات فناوری اطلاعات و ارتباطات فرصت‌های بی‌سابقه‌ای را برای گسترش دسترسی به آموزش ایجادکرده‌است. آموزش الکترونیکی با ارائه خدمات آموزشی انعطاف‌پذیر و بدون محدودیت زمانی و مکانی، نقش مهمی در تحقق هدف آموزش برای همه ایفامی‌کند. با این حال، اثربخشی این سامانه‌ها به‌شدت به توانایی آن‌ها در ارائه خدمات شخصی‌سازی‌شده و متناسب با نیازهای کاربران وابسته است. شخصی‌سازی می‌تواند تجربه یادگیری کاربران را بهبود بخشد و در نهایت به افزایش کارایی سامانه‌ها و نتایج آموزشی منجرشود [۱].

در سال‌های اخیر، روش‌های هوش مصنوعی، به‌ویژه شبکه‌های بیزین، به‌عنوان ابزاری کارآمد برای تحلیل داده‌های آموزشی و مدل‌سازی روابط پیچیده مورد توجه قرارگرفته‌اند. شبکه‌های بیزین به دلیل قابلیت استنتاج احتمالاتی و مدیریت داده‌های ناقص، ابزار مناسبی برای غلبه بر محدودیت‌های روش‌های سنتی داده‌کاوی محسوب‌می‌شوند[۲]. پژوهش‌هایی همچون استفاده از شبکه‌های بیزین برای بهینه‌سازی سیستم‌های آموزشی مبتنی بر وب و ایجاد پلتفرم‌های یادگیری شخصی‌سازی‌شده، نشان‌داده‌اند که این مدل‌ها می‌توانند با ارزیابی بلادرنگ توانایی‌های یادگیرندگان و تحلیل روابط دانش، کارایی یادگیری را به‌طور چشمگیری بهبود بخشند[۴] [۳].

همچنین، رویکردهای مبتنی بر شبکه‌های بیزین به‌عنوان ابزارهای کلیدی برای توسعه سیستم‌های آموزشی تطبیقی مطرح‌شده‌اند که امکان ارائه محتوای آموزشی متناسب با مهارت و نیازهای کاربران را فراهم‌می‌کنند [۵]. با وجود این، اغلب پژوهش‌ها بر جنبه‌های نظری متمرکز بوده و نیاز به مطالعات عملیاتی و ترکیبی برای بهره‌گیری کامل از این رویکردها احساس‌می‌شود.

سامانه‌های آموزش الکترونیکی اگرچه مزایای بسیاری دارند، اما همچنان با چالش‌های متعددی مواجه‌اند. یکی از چالش‌های مهم این سیستم‌ها، عدم توانایی کافی در ارائه خدمات شخصی‌سازی‌شده و متناسب با نیازهای کاربران است. این مسئله نه‌تنها بر تجربه یادگیری کاربران تأثیر منفی می‌گذارد، بلکه کارایی کلی سامانه‌ها را نیز کاهش می‌دهد[۶].

یکی دیگر از مشکلات عمده، مسئله آغاز سرد است که در آن سیستم نمی‌تواند برای کاربران جدید، خدمات مناسبی پیشنهادکند زیرا داده‌های کافی برای تحلیل رفتار آن‌ها وجود ندارد. این مشکل باعث می‌شود که تجربه کاربران در مراحل اولیه استفاده از سیستم کاهش یابد و انگیزه ادامه یادگیری تحت تأثیر قرارگیرد[۷].

علاوه بر این، سیستم‌های آموزش الکترونیکی معمولاً در مواجهه با داده‌های ناقص یا ناسازگار عملکرد مطلوبی ندارند. نبود چارچوب‌هایی که بتوانند این مشکلات را مدیریت‌کرده و از ابزارهای پیشرفته‌ای مانند شبکه‌های بیزین بهره‌بگیرند، منجر به محدودیت در استفاده عملی از این سیستم‌ها شده‌است. به‌طور خاص، فقدان ترکیب شبکه‌های بیزین با تکنیک‌های داده‌کاوی نظیر خوشه‌بندی و انتخاب ویژگی، مانع از بهره‌برداری کامل از ظرفیت‌های این فناوری‌ها شده‌است[۸]. همچنین، بررسی تأثیر این رویکردها در محیط‌های واقعی که داده‌های متنوع و غیرقابل پیش‌بینی دارند، به‌اندازه کافی انجام‌نشده‌است. بیشتر پژوهش‌ها در محیط‌های آزمایشگاهی با داده‌های محدود صورت‌گرفته و نمی‌توانند به‌طور کامل پیچیدگی‌های واقعی را منعکس‌کنند [۹].

پژوهش‌های پیشین عمدتاً به اثربخشی شبکه‌های بیزین در پیش‌بینی رفتار کاربران و تحلیل داده‌های آموزشی پرداخته‌اند، اما به‌ندرت به ارائه چارچوب‌های عملیاتی و ترکیب این مدل‌ها با روش‌های دیگر مانند خوشه‌بندی و انتخاب ویژگی پرداخته‌شده‌است. به‌عنوان مثال، در یک مطالعه، استفاده از شبکه‌های بیزین در سیستم‌های OER تطبیقی به بهبود تسلط بر مهارت‌ها و افزایش نتایج یادگیری دانش‌آموزان کمک‌کرده‌است[۱۰]. همچنین، استفاده از شبکه‌های بیزین برای ارزیابی نتایج یادگیری الکترونیکی نشان‌داده‌است که این مدل‌ها می‌توانند اثربخشی آموزش آنلاین را با تحلیل داده‌های یادگیرندگان افزایش دهند[۱۱].

در زمینه بومی نیز، محدودیت‌های زیرساختی، مشکلات دسترسی به داده‌های کافی، و نیاز به انطباق با شرایط فرهنگی و آموزشی محلی، چالش‌های جدی برای توسعه این سامانه‌ها ایجادکرده‌است. این شکاف‌ها نشان‌می‌دهند که نیاز به ارائه راهکارهای ترکیبی و عملیاتی برای بهره‌گیری بهتر از شبکه‌های بیزین در سامانه‌های آموزش الکترونیکی وجوددارد[۱۲].

مطالعات متعدد اثربخشی شبکه‌های بیزین در بهبود عملکرد سیستم‌های آموزشی را نشان‌داده‌اند. به‌عنوان مثال، در یک مطالعه استفاده از شبکه‌های بیزین برای تحلیل داده‌های یادگیرندگان به کاهش خطاهای پیش‌بینی و بهبود دقت تصمیم‌گیری منجرشد [۱۳]. علاوه بر این، ترکیب شبکه‌های بیزین با الگوریتم‌های خوشه‌بندی، امکان ارائه خدمات تطبیقی متناسب با گروه‌های کاربری مختلف را فراهم‌کرده‌است[۱۴].

در ایران، سامانه‌های آموزش الکترونیکی با محدودیت‌هایی مانند عدم دسترسی همگانی به اینترنت، نبود منابع کافی و مهارت اندک کاربران در استفاده از فناوری مواجهند. این چالش‌ها، نیاز به ارائه چارچوب‌های بومی‌سازی‌شده را دوچندان کرده‌است.

این مطالعه با هدف طراحی یک چارچوب جامع برای بهینه‌سازی سامانه‌های آموزش الکترونیکی مبتنی بر وب انجام‌شده‌است. اهداف خاص پژوهش عبارتند از:

- ارائه یک چارچوب ترکیبی شامل شبکه‌های بیزین، خوشه‌بندی و انتخاب ویژگی.
- بهبود دقت پیش‌بینی و شخصی‌سازی خدمات آموزشی.
- بررسی عملی اثربخشی چارچوب پیشنهادی در سناریوهای واقعی.

این پژوهش تلاش‌می‌کند تا با رفع شکاف‌های موجود، زمینه‌ساز توسعه سامانه‌های آموزش الکترونیکی کارآمد و انطباق‌پذیر در مقیاس

محلی و جهانی شود. در این مقاله با استفاده از شبکه‌های بیزین یک چارچوب نوآورانه برای ارتقای کیفیت سامانه‌های آموزشی مبتنی بر وب آورده شده، با استفاده انتخاب ویژگی از طریق الگوریتم Relief و خوشه‌بندی کاربران با استفاده از الگوریتمk-means ، ابعاد داده‌ها را کاهش‌داده و کارایی پردازش افزایش‌یافته، این مزیت یکی از دلایل ارائه این روش پیشنهاد می‌باشد. در بخش دوم مروری بر کارهای پیشین مرتبط با این تحقیق بیان‌شده‌است. در بخش سوم روش پیشنهادی بیان‌شده‌است. در این بخش ابتدا به‌طور کلی مراحل انجام کار شرح داده‌شده‌است سپس انتخاب ویژگی و خوشه‌بندی بیان‌شده‌است. در بخش چهارم پیشنهاد سرویس به کاربران بیان‌شده‌است. در بخش پنجم آزمایش‌های همراه با تحلیل نتایج ذکرشده‌است. در بخش ششم معیارهای ارزیابی مورد بررسی‌قرارگرفته‌اند و سپس نهایتاً در بخش آخر پژوهش نتیجه‌گیری و پیشنهاد کارهای آتی برای ادامه این پژوهش بیان‌شده‌است.

۲. کارهای مرتبط

مدل فراگیر در مقاله [۱۵] به‌عنوان نمایشی از باورهای یک سیستم رایانه‌ای در مورد فراگیران تعریف‌شده‌است. به عبارت دیگر، این مدل یک نمایش انتزاعی از ویژگی‌ها و رفتارهای احتمالی یک فراگیر را ارائه‌می‌دهد. مدل‌سازی فراگیر شامل مجموعه‌ای از طرز فکرها و برخوردهایی است که یک فرد می‌تواند در طول فرایند یادگیری از خود نشان‌دهد. این مدل‌ها درواقع نقش مهمی در شبیه‌سازی فرایندهای یادگیری و تحلیل تعاملات مختلف فراگیران ایفامی‌کنند. با این حال، باید بین مدل فراگیر و مدل کاربر که جنبه‌ای جامع‌تر دارد و به رفتارهای عمومی‌تری می‌پردازد که فقط به تدریس محدود نمی‌شود، تمایز قائل‌شد. مدل کاربر معمولاً در چارچوب وسیع‌تری از تعاملات انسان با سیستم به‌کارمی‌رود، در حالی‌که مدل فراگیر به‌طور خاص بر جنبه‌های یادگیری و نیازهای آموزشی متمرکز است. در سیستم‌های آموزشی هوشمند (ITS)فراگیران اجزای اصلی سیستم هستند و بر اساس مدل‌های پیچیده‌تری از رفتار و یادگیری آن‌ها طراحی‌می‌شوند[۱۵].

مدل‌های فراگیر در چهارچوب نظریه‌های مختلف یادگیری مانند رفتارگرایی، شناخت‌گرایی و ساخت‌گرایی مورد بررسی قرارمی‌گیرند. در رویکرد رفتارگرایی، یادگیری به‌عنوان یک فرآیند واکنشی به محرک‌ها دیده‌می‌شود، جایی که فرد مانند یک جعبه سیاه عمل‌می‌کند و پاسخ‌هایش به تقویت محرک‌ها بستگی دارد. این رویکرد بیشتر به تکرار و تقویت پاسخ‌های درست تأکیددارد. در مقابل، نظریه شناخت‌گرایی بر پردازش اطلاعات و اصلاح نمایش‌های ذهنی تأکیدمی‌کند و یادگیری را به‌عنوان تغییرات در این نمایش‌ها می‌بیند. همچنین، نظریه ساخت‌گرایی فرض‌می‌کند که فراگیران دانش را بر اساس تجربیات گذشته خود می‌سازند و هر فرد واقعیت و دانش خود را ایجادمی‌کند. این رویکرد تأکید زیادی بر یادگیری فعال و تجربیات شخصی دارد. در این میان، مدل‌های ITS سنتی بیشتر از رفتارگرایی و شناخت‌گرایی حمایت‌می‌کنند، چراکه بیشتر به ارزیابی مهارت‌ها و شبیه‌سازی وضعیت‌های درونی فراگیران توجه دارند. با این وجود، به‌کارگیری مدل‌های پیچیده‌تر و پیشرفته‌تر، مانند استفاده از شبکه‌های بیزین، می‌تواند به شکل‌دهی مدل‌هایی کمک‌کند که نه‌تنها به ارزیابی مهارت‌ها، بلکه به شبیه‌سازی فرآیندهای پیچیده‌تر یادگیری و تعاملات فراگیران می‌پردازند.

انواع مختلفی از مدل‌های فراگیر وجوددارد که در مطالعات متعدد به آن‌ها پرداخته‌شده‌است. این مدل‌ها شامل مدل‌های پوششی، مدل‌های افتراقی و مدل‌های اختلالی هستند [۱۷][۱۸][۱۹][۲۰]. هر یک از این مدل‌ها به‌طور خاص بر جنبه‌های متفاوت یادگیری و رفتارهای فردی متمرکز هستند و می‌توانند به بهبود دقت سیستم‌های آموزشی هوشمند کمک‌کنند. در این راستا، مدل‌های شبکه بیزین نیز به‌عنوان ابزارهایی قدرتمند برای تحلیل و پیش‌بینی رفتارهای یادگیری دانش‌آموزان و طراحی سیستم‌های آموزشی شخصی‌سازی‌شده معرفی‌شده‌اند.

در مطالعه [۲۱] نویسندگان به بررسی استفاده از درخت تصمیم و شبکه عصبی برای شناسایی عوامل مؤثر در هدایت تحصیلی دانش‌آموزان پرداخته‌اند. این سیستم با تحلیل داده‌های وضعیت تحصیلی و مشاوره دانش‌آموزان، شانس موفقیت آن‌ها را پیش‌بینی‌کرده و به هدایت تحصیلی بهینه کمک‌می‌کند. همچنین، در پژوهش [۲۲]، به ارتقای کیفیت آموزش در سیستم‌های آموزش الکترونیکی با استفاده از داده‌کاوی آموزشی پرداخته‌شده‌است. این تحقیق هدف اصلی خود را کشف الگوهای نهفته در انتخاب واحدهای درسی و پیش‌بینی نمرات دانش‌آموزان قرارداده‌است. همچنین، تأثیر عواملی مانند ساعت و زمان ورود دانش‌آموزان به سیستم آموزش الکترونیکی بر کیفیت و نتایج یادگیری مورد بررسی قرارگرفته‌است. نتایج این تحقیق نشان می‌دهد که استفاده از تکنیک‌های داده‌کاوی می‌تواند به بهبود کیفیت آموزش و هدایت تحصیلی دانش‌آموزان کمک‌کند.

در پژوهش دیگری که در [۲۳] انجام‌شده‌است، نویسندگان به شناسایی عوامل مؤثر بر افت تحصیلی دانش‌آموزان با استفاده از قوانین انجمنی و تحلیل خوشه‌ای پرداخته‌اند. این تحقیق به‌دنبال پیاده‌سازی مدل‌های داده‌کاوی پیش‌بینی‌کننده برای پیش‌بینی وضعیت تحصیلی دانش‌آموزان بر اساس مشخصات فردی و تحصیلی آن‌ها بوده‌است. نتایج این تحقیق نشان‌داده‌است که با استفاده از این مدل‌ها می‌توان عوامل مؤثر بر افت تحصیلی را شناسایی کرده و از آن‌ها برای پیشگیری از افت تحصیلی و بهبود کیفیت ارتباط مسئولین و والدین با دانش‌آموزان بهره‌برد.

در پژوهش [۲۴] با استفاده از الگوریتم‌های مختلفی مانند درخت تصمیم بهبود یافته و الگوریتم نزدیک‌ترین همسایه، سوابق تحصیلی دانش‌آموزان را تجزیه و تحلیل کرده و به هدایت آن‌ها به رشته‌های تحصیلی مناسب کمک‌می‌کند. این تحقیق نشان‌داده‌است که ترکیب تکنیک‌های مختلف داده‌کاوی می‌تواند به‌طور مؤثری در هدایت تحصیلی دانش‌آموزان در مراحل مختلف تحصیلی مورد استفاده‌قرارگیرد.

در سال‌های اخیر، یادگیری الکترونیکی به‌عنوان یک روش آموزشی نوین موردتوجه قرارگرفته‌است. یکی از چالش‌های اصلی این حوزه، عدم تعامل مستقیم بین یادگیرنده و آموزش‌دهنده است که نیاز به شخصی‌سازی آموزش را افزایش‌می‌دهد. در این راستا، قره‌باغی و امیری در [۲۵] از الگوریتم شبه‌نظارتی LP-MLTSVM برای شناسایی سطح دانش یادگیرندگان و بهبود کیفیت آموزش استفاده‌کرده‌اند. نتایج‌نشان-داد که این روش در مقایسه با روش‌های سنتی مانند درخت تصمیم و ماشین بردار پشتیبان، عملکرد و رضایت یادگیرندگان را به‌طور معناداری افزایش‌می‌دهد.

فارغ‌زاده و مدیری یک زیرساخت ابری آگاه از کارایی برای سیستم‌های آموزش الکترونیکی ارائه‌کردند که به‌صورت یکپارچه و پویا عملکرد کل پشته ابری را پایش‌می‌کند. این زیرساخت با استفاده از واحدهایی مانند مخزن دانش اشتراکی و واحد کنترل و اجرا، معیارهای کارایی مانند زمان پاسخ و گذردهی را بهبود بخشید. این رویکرد مستقل از مکانیزم‌های اجرایی خاص، قابلیت تطبیق با محیط‌های چندمستأجری را فراهم‌می‌کند[۲۶].

در پژوهشی دیگر [۲۷] با هدف بهبود دقت و شخصی‌سازی پیشنهادات فیلم، از ترکیب تکنیک‌های یادگیری عمیق و تحلیل معنایی استفاده‌می‌کند. روش پیشنهادی این مقاله، با بهره‌گیری از داده‌های کاربران و ویژگی‌های محتوایی فیلم‌ها، مدلی ارائه‌می‌دهد که قادر است ترجیحات گروهی و فردی کاربران را به طور همزمان در نظر بگیرد. نتایج آزمایش‌ها نشان می‌دهد که این رویکرد در مقایسه با روش‌های سنتی، عملکرد بهتری در پیش‌بینی ترجیحات کاربران و ارائه پیشنهادات مرتبط‌تر دارد. این مطالعه به‌عنوان یک چارچوب نوآورانه در حوزه سیستم‌های توصیه‌گر، به‌ویژه در زمینه توصیه‌های گروهی و معنایی، مطرح‌شده‌است.

در مجموع، استفاده از شبکه‌های بیزین و تکنیک‌های داده‌کاوی در تحلیل و مدل‌سازی سیستم‌های آموزشی، باعث پیش‌بینی دقیق‌تر وضعیت تحصیلی دانش‌آموزان و بهبود فرایندهای یادگیری آنلاین می‌شود. این روش‌ها می‌توانند اطلاعات پنهان و پیچیده در داده‌های آموزشی را شناسایی‌کرده و سیستم‌های آموزشی را به‌گونه‌ای بهینه‌سازی کنند که نتایج یادگیری فردی هر دانش‌آموز را بهبوددهند. همچنین، با توجه به پیچیدگی‌های موجود در فرآیند یادگیری و نیاز به تحلیل داده‌های عظیم، استفاده از مدل‌های پیشرفته مانند شبکه‌های بیزین می‌تواند به‌عنوان ابزاری مؤثر در طراحی سیستم‌های آموزشی هوشمند و شخصی‌سازی‌شده در نظرگرفته‌شود.

۳. روش پیشنهادی

۱،۳ زمینه

روش پیشنهادی این پژوهش به‌منظور بهبود کیفیت خدمات در سیستم‌های آموزش الکترونیکی، بر پایه استفاده از شبکه‌های بیزین طراحی‌شده‌است. همان‌طور که در شکل (۱) نمایش داده‌شده، این روش شامل چند مرحله اصلی است که به‌طور خلاصه به شرح زیر توضیح‌داده-می‌شود.

در اولین مرحله، مجموعه‌ای از ویژگی‌های کلیدی شناسایی می‌شوند که بیشترین تأثیر را در بهبود کیفیت خدمات به کاربران دارند. این ویژگی‌ها شامل عواملی هستند که نیازها، ترجیحات و الگوهای رفتاری کاربران در استفاده از خدمات آموزش الکترونیکی را نمایان می‌کنند. ویژگی‌هایی که تأثیر کمی بر کیفیت خدمات دارند یا تکراری و غیرضروری هستند، در این مرحله حذف‌می‌شوند. فرآیند انتخاب ویژگی‌ها نه تنها به ساده‌سازی مدل کمک می‌کند، بلکه حجم داده‌های پردازش‌شده را کاهش می‌دهد و در نتیجه، کارایی سیستم بهبود می‌یابد. این امر موجب می‌شود مدل سریع‌تر و مؤثرتر عمل کند.

در مرحله بعد، کاربران با استفاده از الگوریتم خوشه‌بندی-k means بر مبنای ویژگی‌های انتخاب‌شده گروه‌بندی می‌شوند. الگوریتم k-means یک روش شناخته‌شده است که داده‌ها را به خوشه‌های مختلف تقسیم‌می‌کند، به‌گونه‌ای که کاربران در هر خوشه بیشترین شباهت را به یکدیگر دارند و تفاوت‌ها بین خوشه‌ها به حداکثر می‌رسد. این فرآیند به سیستم این امکان را می‌دهد که کاربران را بر اساس رفتار و نیازهای مشابه دسته‌بندی کند. به این ترتیب، نیازها و الگوهای رفتاری کاربران در هر خوشه به‌طور دقیق‌تر شناسایی می‌شود و امکان ارائه خدمات اختصاصی‌تر به هر گروه فراهم‌می‌آید.

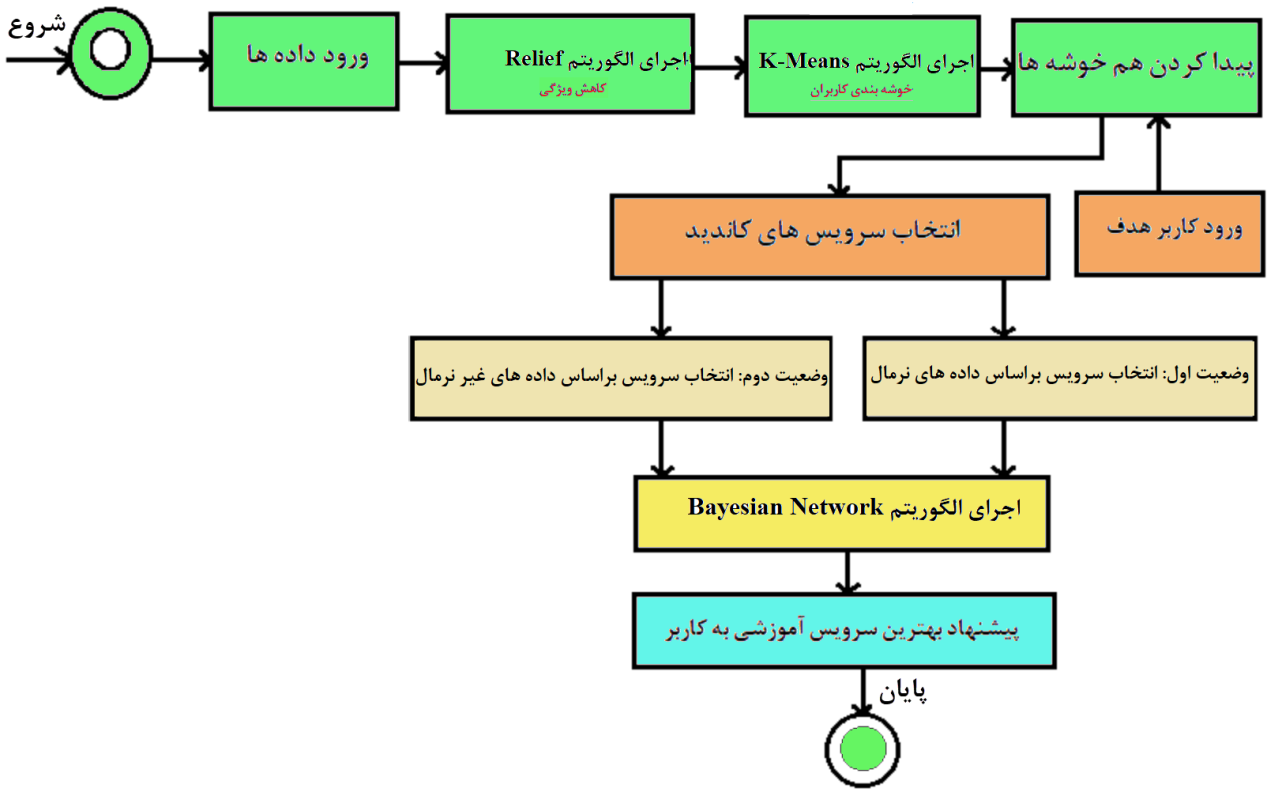
پس از خوشه‌بندی، سیستم به‌طور هوشمندانه سرویس‌ها و محتوای آموزشی را که بیشترین کاربرد را برای هر گروه از کاربران دارد، شناسایی و به آن‌ها پیشنهادمی‌دهد. این فرآیند موجب بهینه‌سازی تجربه کاربری می‌شود، زیرا کاربران هر گروه دقیقاً خدمات و محتوایی را دریافت-می‌کنند که برای آن‌ها بیشترین فایده را دارد. به‌عنوان مثال، به کاربرانی که در زمینه خاصی نیاز به تقویت دارند، سرویس‌هایی پیشنهادمی‌شود که نقاط ضعف آن‌ها را پوشش‌دهد، در حالی که کاربران پیشرفته‌تر به محتوای سطح بالاتر و چالش‌برانگیزتر دسترسی پیدامی‌کنند.

در مرحله بعدی، شبکه‌های بیزین به‌عنوان ابزار اصلی برای تحلیل احتمالی و استنتاج به‌کارگرفته‌می‌شوند. شبکه بیزین یک مدل گرافیکی جهت‌دار است که روابط و وابستگی‌ها میان متغیرها را به‌صورت تصویری نمایش‌می‌دهد. این شبکه به سیستم این امکان را می‌دهد که حتی با داده‌های ناقص یا نامطمئن، پیش‌بینی‌هایی در خصوص رفتار و نیازهای آینده کاربران انجام‌دهد.

در نهایت، عملکرد سیستم در ارائه خدمات به کاربران مورد ارزیابی قرارمی‌گیرد. نتایج حاصل از پیشنهاد سرویس‌ها و سطح رضایتمندی

کاربران در هر گروه بررسی‌می‌شود و بازخوردهای آن‌ها برای بهبود مدل و به‌روزرسانی شبکه بیزین استفاده‌می‌شود. این روش به سیستم این امکان را می‌دهد که به‌طور مداوم کیفیت خدمات خود را ارتقادهد و از بازخوردهای کاربران برای بهبود فرآیندها بهره‌برداری‌کند.

روش پیشنهادی این پژوهش با استفاده از الگوریتم‌های یادگیری ماشینی و شبکه‌های بیزین، امکان شخصی‌سازی و بهبود تجربه کاربری را فراهم‌می‌آورد و به سیستم‌های آموزش الکترونیکی کمک‌می‌کند تا به‌طور هوشمندانه‌تری عمل‌کنند. به‌ویژه، این روش توانایی پردازش داده‌های ناقص یا نامطمئن را دارد، که در شرایطی که اطلاعات به‌طور کامل در دسترس نیست، یک مزیت قابل‌توجه محسوب‌می‌شود. در نهایت، این معماری پیش‌بینی می‌تواند تجربه آموزشی بهینه‌ای برای کاربران فراهم‌کرده و کیفیت خدمات را ارتقادهد. به‌این‌ترتیب، این روش به‌عنوان یک راهکار جامع در سیستم‌های آموزش الکترونیکی قابل-استفاده‌است.



شکل ۱: معماری روش پیشنهادی

۲,۳ انتخاب ویژگی

برای یافتن کاربران مشابه، ابتدا ویژگی‌های مشترک هر کاربر را انتخاب-کرده و سپس از الگوریتم خوشه‌بندی استفاده‌می‌کنیم تا کاربران را در گروه‌های مشابه قراردهیم. این ویژگی‌ها ممکن است بین کاربران متفاوت باشد و به نوع سرویس و اطلاعات موجود برای هر کاربر بستگی دارد. به عنوان روش اصلی برای انتخاب ویژگی، از الگوریتم Relief استفاده-می‌کنیم که مبتنی بر روش‌های آماری و وزن‌دهی است. در این روش، ابتدا زیرمجموعه‌ای از داده‌های آموزشی انتخاب‌می‌شود. سپس برای هر نمونه در این زیرمجموعه، نزدیک‌ترین نمونه‌های مشابه (نزدیک‌ترین برخورد) و متفاوت (نزدیک‌ترین شکست) بر اساس معیار فاصله اقلیدسی شناسایی‌می‌شوند.

نزدیک‌ترین برخورد: شامل نمونه‌های هم‌کلاس است که کمترین فاصله را با نمونه مورد نظر دارند. در مقابل، نزدیک‌ترین شکست: شامل نمونه‌های خارج از کلاس نمونه است که کمترین فاصله را دارند. هدف از این کار، پیداکردن ویژگی‌هایی است که اختلاف آن‌ها با نزدیک‌ترین برخورد کمتر و با نزدیک‌ترین شکست بیشتر باشد، چراکه ویژگی‌های بهتر به تفکیک کلاس‌ها کمک‌می‌کنند. وزن‌دهی ویژگی‌ها بر اساس این اختلاف‌ها به‌روز می‌شود: اختلاف بین نمونه انتخاب‌شده و نزدیک‌ترین برخورد، وزن ویژگی را کاهش‌می‌دهد، و اختلاف با نزدیک‌ترین شکست وزن را افزایش‌می‌دهد. ویژگی‌هایی که وزن بالاتری دارند، بهتر می‌توانند نمونه‌های مشابه را از هم تفکیک‌کنند. در پایان، ویژگی‌هایی که وزن آن‌ها از یک حد آستانه کمتر است، حذف می‌شوند و ویژگی‌های باقی‌مانده به عنوان مجموعه ویژگی‌های منتخب استفاده می‌شوند. این حد آستانه توسط کاربر یا خودکار و با آزمون و خطا تعیین‌می‌شود.

```
Relief(D, S, NoSample, Threshold)
  (1) T = φ
  (2) Initialize all weights, W_i, to zero.
  (3) For i = 1 to NoSample/* Arbitrarily chosen */
      Randomly choose an instance x in D
      Finds its nearHit and nearMiss
      For j = 1 to N
          W_j = W_j − diff(x_j, nearHit_j)^2 + diff(x_j, nearMiss_j)^2
  (4) For j = 1 to N
          If W_j ⩾ Threshold
              Append feature f_j to T
  (5) Return T
```

شکل ۲: الگوریتم Relief جهت انتخاب ویژگی و کاهش ابعاد

الگوریتم Relief مطابق شکل (۲) برای ویژگی‌های دارای نویز یا ویژگی‌های دارای همبستگی خوب کارمی‌کند و پیچیدگی زمانی آن به‌صورت خطی و تابعی از تعداد ویژگی‌ها و تعداد نمونه های مجموعه نمونه می‌باشد، و هم برای داده‌های پیوسته و هم برای داده‌های صوری خوب کارمی‌کند. برای کاربران ویژگی‌های زیر در نظر گرفته‌شده‌است، برای این که عملیات مربوط به پردازش راحت‌تر و بهتر صورت‌بگیرد هر یک از ویژگی‌ها به یک رنج عددی بر اساس نوع آن نگاشت‌شده‌اند. به‌عنوان مثال جدول (۱) با تمام ویژگی‌ها برای ۱۰ کاربر وجوددارد که ویژگی‌ها با F_1 تا F_{18} نشان‌داده‌شده‌اند. در جدول(۱) تمام اعداد فرضی هستند و هدف فقط تشریح روش پیشنهادی برای انتخاب ویژگی می‌باشد و فرض‌شده‌است که هر کاربر حداکثر۱۸ ویژگی دارد.

الگوریتم Relief در این مقاله به‌منظور کاهش ابعاد داده‌ها با انتخاب ویژگی‌های مهم مورد استفاده‌قرارگرفته‌است. فرآیند این الگوریتم برای داده‌های جدول (۱) به صورت زیر انجام‌می‌شود:

جدول ۱: ویژگی‌های برای کاربران مختلف

کد کاربر	کد سرویس	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}	F_{17}	F_{18}
B۱	A	۵	۱۲	۱۰۰	-	۶۵	۷۹	-	-	۲۵	-	-	۱۲	-	۳۶	-	-	۱۲	۲۰
B۲	B	۱۰	۱۱	-	۲۵	۷۰	-	۴۵	-	۱۹	۳۵	-	-	۱۴	-	۳	۲	۲۹	-
B۳	A	۵	-	۱۱۰	۲۹	-	۸۰	۴۹	۸۹	-	۳۶	۷۴	-	۱۵	-	-	-	۱۹	۲۵
B۴	C	-	۱۹	-	-	۵۵	۸۱	-	-	۲۰	۳۶	-	۱۵	-	۳۷	۵	۷	-	-
B۵	C	۹	۱۵	۱۱۵	۳۰	۷۵	-	-	-	-	۳۹	-	۱۸	-	-	۷	۴	۲۰	۲۱
B۶	A	۸	-	-	۲۸	-	۸۰	۴۵	۹۰	۲۰	-	۷۴	۱۷	-	۳۶	-	-	۲۱	-
B۷	B	۴	۱۳	۱۲۰	-	۷۵	۸۲	۴۵	۹۱	-	-	۷۲	۱۵	-	۳۶	۷	۷	-	۲۵
B۸	B	-	۱۴	۱۱۲	-	۷۰	۷۷	۴۵	-	۱۹	۳۶	-	-	۱۵	۳۵	-	۷	-	-
B۹	C	۲	۱۳	-	۳۰	۷۹	-	-	-	۲۰	-	۷۵	-	۱۵	۳۵	۴	۵	-	۲۲
B۱۰	A	۸	-	۱۱۳	-	-	۸۰	-	۹۰	۲۰	۳۵	-	۱۸	۱۵	۳۳	۵	-	۲۰	۲۰

نوع انتخاب ویژگی بستگی به کاربر دارد زیرا اطلاعات موجود مربوط به انتخاب سرویس‌ها می‌باشد در ابتدا میانگین تعداد ویژگی‌های هر سرویس محاسبه‌می‌شود درواقع بررسی‌می‌شود که به‌طور میانگین برای هر سرویس برای تمام کاربرها چند ویژگی وجود دارد مثلاً برای سرویس A تعداد ویژگی‌های دارای مقدار برای هر کاربر به صورت زیر می‌باشد:

- کاربر B_1 دارای ۱۰ ویژگی می‌باشد.
- کاربر B_3 دارای ۱۱ ویژگی می‌باشد.
- کاربر B_6 دارای ۱۰ ویژگی می‌باشد.
- کاربر B_{10} دارای ۱۱ ویژگی می‌باشد.

پس سرویس A برای تمام ۴ کاربر این سرویس در مجموع ۴۳ویژگی و به‌طور متوسط ۱۰ ویژگی دارد، طبق توضیحات بالا برای تمام سرویس‌ها در جدول (۲) قابل‌مشاهده‌می‌باشد و تعداد ویژگی‌های هر سرویس برابر با تعداد میانگین سرویس‌ها به‌ازای کاربران سیستم خواهدبود.

۱. **ورودی‌ها:**

- مجموعه داده شامل کاربران (B_i)، سرویس‌ها(A,B,C) ، و ویژگی‌های مرتبط F_1 تا F_{18} .
- تعداد نمونه‌ها (m) و تکرارهای الگوریتم(k) .

۲. **مراحل اصلی:**

- انتخاب نمونه تصادفی :یک نمونه (S) از داده‌ها انتخاب می‌شود.
- شناسایی همسایگان مثبت و منفی :

a) H^+ نزدیک‌ترین نمونه با همان سرویس.(A, B, C)

b) H^- نزدیک‌ترین نمونه با سرویس متفاوت.

مرحله بعدی به‌روزرسانی وزن ویژگی‌ها می‌باشد، برای هر ویژگی (F_j)، وزن‌ها با استفاده از رابطه ۱ را زیر به‌روزرسانی می‌شوند:

$$W[F_j] = W[F_j] - \frac{diff(F_j, S, H^+)}{K} + \frac{diff(F_j, S, H^-)}{K} \quad (۱)$$

که در آن $diff(F_j, S, H)$ تفاوت مقدار ویژگی F_j بین نمونه S و همسایه H است. ویژگی‌هایی که در تمام کاربران یک سرویس مقدار دارند، به‌عنوان ویژگی‌های قطعی در جدول (۴) ثبتمی‌شوند. ویژگی‌هایی که حداقل ۵۰ درصد کاربران یک سرویس مقدار دارند، به‌عنوان ویژگی‌های نهایی برای هر سرویس در جدول (۶) انتخاب‌می‌شوند.

جدول ۲: میانگین تعداد ویژگی برای هر سرویس

سرویس	میانگین تعداد ویژگی
A	۱۰
B	۱۱
C	۱۰

در ادامه باید ویژگی‌های انتخاب‌شده قطعی مشخص‌شود، این ویژگی‌ها، ویژگی‌هایی هستند که در تمام کاربران سرویس مورد نظر دارای مقدار هستند، به‌عنوان مثال در جدول (۳) کاربرانی که سرویس A را دارند قابل‌مشاهده‌اند و ستون‌های سبز رنگ ویژگی‌های قطعی هستند زیرا در تمام کاربران دارای ویژگی می‌باشند.

جدول ۳: ویژگی‌های قطعی

کد کاربر	کد سرویس	$F_۱$	$F_۲$	$F_۳$	$F_۴$	$F_۵$	$F_۶$	$F_۷$	$F_۸$	$F_۹$	$F_{۱۰}$	$F_{۱۱}$	$F_{۱۲}$	$F_{۱۳}$	$F_{۱۴}$	$F_{۱۵}$	$F_{۱۶}$	$F_{۱۷}$	$F_{۱۸}$
B۱	A	۵	۱۲	۱۰۰	-	۶۵	۷۹	-	-	۲۵	-	-	۱۲	-	۳۶	-	-	۱۲	۲۰
B۳	A	۵	-	۱۱۰	۲۹	-	۸۰	۴۹	۸۹	-	۳۶	۷۴	-	۱۵	-	-	-	۱۹	۲۵
B۶	A	۸	-	-	۲۸	-	۸۰	۴۵	۹۰	۲۰	-	۷۴	۱۷	-	۳۶	-	-	۲۱	-
B۱۰	A	۸	-	۱۱۳	-	-	۸۰	-	۹۰	۲۰	۳۵	-	۱۸	۱۵	۳۳	۵	-	۲۰	۲۰

جدول ۴: ویژگی‌های انتخاب‌شده قطعی

سرویس	ویژگی
A	$F_۱$- $F_۶$- $F_{۱۷}$
B	$F_۲$- $F_۵$- $F_۷$- $F_{۱۶}$
C	$F_۳$- $F_۵$- $F_{۱۵}$

تا به اینجا ویژگی‌های قطعی برای هر سرویس انتخاب شده‌است و در ادامه ویژگی‌هایی، برای هر سرویس انتخاب خواهندشد که حداقل نصف کاربران دارای ویژگی فوق باشند. در این انتخاب برای هر ویژگی بررسی‌می‌شود که چه تعداد از کاربران سرویس مورد نظر را دارند، ویژگی که ۵۰ درصد کاربران آن را داشته‌باشند انتخاب خواهندشد به‌عنوان مثال در جدول (۵) برای سرویس A ویژگی‌های به رنگ آبی انتخاب‌می‌شوند زیرا ۵۰ درصد کاربران حداقل ویژگی فوق را دارند.

جدول ۵: کاربران دارای ویژگی

کد کاربر	کد سرویس	$F_۱$	$F_۲$	$F_۳$	$F_۴$	$F_۵$	$F_۶$	$F_۷$	$F_۸$	$F_۹$	$F_{۱۰}$	$F_{۱۱}$	$F_{۱۲}$	$F_{۱۳}$	$F_{۱۴}$	$F_{۱۵}$	$F_{۱۶}$	$F_{۱۷}$	$F_{۱۸}$
B۱	A	۵	۱۲	۱۰۰	-	۶۵	۷۹	-	-	۲۵	-	-	۱۲	-	۳۶	-	-	۱۲	۲۰
B۳	A	۵	-	۱۱۰	۲۹	-	۸۰	۴۹	۸۹	-	۳۶	۷۴	-	۱۵	-	-	-	۱۹	۲۵
B۶	A	۸	-	-	۲۸	-	۸۰	۴۵	۹۰	۲۰	-	۷۴	۱۷	-	۳۶	-	-	۲۱	-
B۱۰	A	۸	-	۱۱۳	-	-	۸۰	-	۹۰	۲۰	۳۵	-	۱۸	۱۵	۳۳	۵	-	۲۰	۲۰

به‌عنوان مثال در جدول (۵) ویژگی $F_۳$ انتخاب می‌شود زیر ۳ کاربر از ۴ کاربر ویژگی فوق را دارد ولی ویژگی $F_{۱۶}$ انتخاب‌نمی‌شود، در جدول (۶) ویژگی‌های انتخاب‌شده برای هر سرویس مشخص شده‌است.

۳،۳ خوشه‌بندی

برای هر سرویس به‌صورت مجزا خوشه‌بندی انجام‌می‌شود که ویژگی‌های برابر دارند. الگوریتم خوشه‌بندی K-means. داده‌ها را به K خوشه تقسیم‌می‌کند؛ به‌طوری‌که داده‌های هر خوشه بیشترین شباهت و داده‌های خوشه‌های مختلف بیشترین تفاوت را داشته‌باشند. در هر تکرار، مرکز خوشه‌ها به‌روزرسانی می‌شود تا مجموع فاصله داده‌ها از مرکز خوشه حداقل شود. الگوریتم تا زمانی که این فاصله به حداقل برسد، تکرار می‌شود و در نهایت خوشه‌هایی مجزا و فشرده به دست می‌آید.

جدول ۶ : ویژگی‌های انتخاب‌شده برای هر سرویس

کد سرویس	$F_۱$	$F_۲$	$F_۳$	$F_۴$	$F_۵$	$F_۶$	$F_۷$	$F_۸$	$F_۹$	$F_{۱۰}$	$F_{۱۱}$	$F_{۱۲}$	$F_{۱۳}$	$F_{۱۴}$	$F_{۱۵}$	$F_{۱۶}$	$F_{۱۷}$	$F_{۱۸}$
A	√	-	√	√	-	√	√	√	√	√	√	√	√	√	-	-	√	√
B	√	√	√	-	√	√	√	-	√	√	-	-	√	√	√	√	-	-
C	√	√	-	√	√	-	-	-	√	√	-	√	-	√	√	√	-	√

برای محاسبه فاصله برای تمام بردارها از بردار میانگین طبق رابطه (۲) استفاده‌می‌کنیم.

(۲)

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

هر چقدر مقدار فاصله کمتر باشد بردارها مشابه‌تر می‌باشند. برای مثال فاصله کاربر B_1 با مشخصات زیر:

کد کاربر	F1	F2	F3	F4	F5
1B	۱۴	۱.۶	۳۷	۷۵	۱۵

از بردار میانگین با مشخصات زیر:

کد کاربر	F1	F2	F3	F4	F5
۰	۷.۱۲	۶۷.۶	۷.۳۶	۱.۷۴	۶.۱۷

به صورت زیر محاسبه‌می‌شود.

$$Distance = \sqrt{\begin{Bmatrix}(۱۷.۶-۱۵)^۲+(۷۴.۱-۷۵)^۲+(۳۶.۷-۳۷)^۲ \\ +(۶.۶۷-۶.۱)^۲+(۱۲.۷-۴)^۲\end{Bmatrix}} = ۳.۱۱۰۴۵$$

نتیجه محاسبه فاصله برای تمام داده‌ها از بردار میانگین در جدول (۷) نشان داده‌شده‌است.

جدول ۷: میزان فاصله

کد کاربر	فاصله
B_1	۱۱۰۴۵.۳
B_2	۲۳۷۴۹.۲۱
B_3	۱۸۸۸.۱۲
B_4	۰۲۰۱۲.۱۰
B_5	۷۵۷۱۳۷.۶
B_6	۹۵۳۵۳.۱۱
B_7	۱۸۸۸.۱۲
B_8	۱۴۲۰۹.۱۵
B_9	۷۵۷۱۳۷.۶
B_{10}	۹۴۰۲۴.۷۴

میانگین فاصله‌ها برابر است با ۴۲.۱۷ بنابراین داده بردار B_1، B_8 و B_{10} به‌عنوان سرخوشه انتخاب‌می‌شوند زیرا B_1 نسبت به مقدار میانگین ۴۲.۱۷ بیشترین فاصله مثبت را دارد.

B_1 نسبت به مقدار میانگین۴۲.۱۷ بیشترین فاصله منفی را دارد (کوچکترین عدد در اعدادکوچکتر از۴۲.۱۷). B_8 کمترین فاصله را از بردار میانگین دارد. برای انتخاب اعضای هر خوشه طبق رابطه گفته‌شده، فاصله هر داده از سرخوشه‌های تعیین‌شده محاسبه‌می‌شود. بردارها جذب خوشه‌های می‌شوند که کمترین فاصله را دارند که نتیجه در جدول (۸) قابل مشاهده است.

جدول ۸: نتیجه خوشه بندی

سرخوشه	اعضا
B_1	B_5, B_9
B_8	B_7, B_3, B_4, B_6
B_{10}	B_2

برای داده‌های بعدی که واردمی‌شوند فاصله هر بردار از سرخوشه‌ها محاسبه‌می‌شود و اگر فاصله بردار از میانگین فاصله‌ها کمتر باشد جذب یکی از سرخوشه‌ها می‌شود و در غیر این صورت خود به‌عنوان یک سرخوشه جدید انتخاب‌می‌شود.

۴. پیشنهاد سرویس به کاربر

۱,۴ انتخاب سرویس کاندید

روش پیشنهادی برای پیشنهاد سرویس به کاربر شامل دو حالت است: وضعیت اول: اگر کاربر یک سرویس را انتخاب‌کند، سیستم بر اساس آن سرویس، سایر سرویس‌های مرتبط را پیشنهادمی‌دهد. این وضعیت شامل سه مرحله است:

- ایجاد جدولی از پایگاه داده که سرویس‌های مرتبط با هدف کاربر را نشان‌می‌دهد. سپس کاربرانی که سرویس مورد نظر را انتخاب‌کرده‌اند، شناسایی و سایر کاربران حذف‌می‌شوند. حد آستانه برای انتخاب سرویس‌ها، نصف تعداد این کاربران در نظرگرفته‌می‌شود.
- سرویس‌هایی که فراوانی آن‌ها از حد آستانه بیشتر است انتخاب‌می‌شوند. سپس مجموعه‌های بزرگتری از این سرویس‌ها ساخته‌شده و مجموعه‌هایی که از آستانه کمتر باشند حذف‌می‌شوند.
- مجموعه‌های باقی‌مانده براساس معیارهای اطمینان و پوشش بررسی‌می‌شوند تا سرویس‌های پیشنهادی نهایی انتخاب-شوند.

وضعیت دوم: اگر کاربر هیچ سرویسی انتخاب نکرده‌باشد، سیستم با توجه به تشابه او با دیگر کاربران، سرویس‌های مناسب را پیشنهادمی‌دهد. این وضعیت جهت پیشنهاد سرویس به کاربر شامل مراحل زیر می‌باشد:

- کاربر عنوان هدف را انتخاب‌می‌کند.
- سیستم جدول مربوط به هدف مورد نظر را می‌سازد.
- تمام قوانین ممکن را با توجه به حد آستانه تولیدمی‌کند.
- قوانینی که از نظر پوشش و اطمینان قابل‌قبول باشند به‌عنوان پیشنهاد به کاربر ارائه‌می‌شوند.

برای فهم بیشتر مطلب در اینجا یک نمونه پیشنهاد انتخاب سرویس به کاربر را مورد بررسی قرارمی‌دهیم. در جدول (۹) نشان‌می‌دهد تعداد ۶ سرویس برای هدف مورد نظر کاربر وجوددارد و چه کاربرهای چه سرویس‌هایی را انتخاب‌کرده‌اند.

جدول ۹: سرویس‌های انتخابی برای هدف مورد نظر

کاربر	A	B	C	D	E	F
B_1	*	*		*	*	*
B_2		*	*		*	*
B_3	*		*	*		*
B_4		*	*	*	*	
B_5	*			*		*

$B_۶$	*	*	*	*	*	

حال فرض‌کنید چهار کاربر سرویس A را انتخاب‌کرده‌اند و حد آستانه ۲ در نظر گرفته‌می‌شود. مطابق جدول (۱۰) فقط انتخاب کاربرانی که سرویس A را انتخاب‌کرده‌اند، بررسی‌می‌شود.

جدول ۱۰: انتخاب سرویس کاربرانی که حتماً A را انتخاب‌کرده‌اند

کاربر	A	B	C	D	E	F
B۱	*	*		*	*	*
B۳	*		*	*		*
B۵	*			*		*
B۶	*	*	*	*	*	

ابتدا فراوانی تک‌تک سرویس‌ها محاسبه می‌شود.

{A}=۴,{B}=۲, {C}=۲, {D}=۴,{E}=۲, {F}=۳

به دلیل اینکه فراوانی همه سرویس‌ها از حد آستانه بالاتر هستند سرویسی حذف‌نمی‌شود و مجموعه $L_۱$ ساخته‌می‌شود.

L۱={{A},{B},{C},{D},{E},{F}}

مجموعه دو عضوی با استفاده از $L_۱$ ساخته‌می‌شود.

- {A,B}=۲ {A,C}=۲ {A,D}=۴ {A,E}=۲ {A,F}=۳
- {B,C}=۱ {B,D}=۲ {B,E}=۲ {B,F}=۱
- {C,D}=۲ {C,E}=۱ {C,F}=۱
- {D,E}=۲ {D,F}=۳
- {E,F}=۱
- $L_۲$={{A,B},{A,C},{A,D},{A,E},{A,F},{B,D},{B,E},{C,D},{D,E},{D,F}}

با استفاده از مجموعه $L_۲$ مجموعه $L_۳$ ساخته می‌شود.

- {A,B,C}=۱ {A,B,D}=۲ {A,B,E}=۲ {A,B,F}=۱ {A,C,D}=۲
- {A,C,E}=۱ {A,C,F}=۱ {A,D,E}=۲ {A,D,F}=۳ {A,E,F}=۱
- {B,D,E}=۲ {B,D,C}=۱ {B,D,E}=۲ {B,D,F}=۱
- $L_۳$={ {A,B,D}, {A,B,E}, {A,C,D}, {A,D,E}, {A,D,F} ,{B,D,E} }

با استفاده از مجموعه $L_۳$ مجموعه $L_۴$ ساخته‌می‌شود:

- {A,B,D,E}=۲ {A,B,C,D}=۱ {A,B,D,F}=۱ {A,C,D,E}=۱ {A,C,D,F}=۱ {A,C,D,F}=۱ {A,D,E,F}=۱
- $L_۴$={ {A,B,D,E}}

مجموع نهایی از $L_۲$,$L_۳$, $L_۴$ به‌صورت زیر می‌باشد:

Answer={{A,B}, {A,C}, {A,D},{A,E}, {A,F},{B,D}, {B,E}, {C,D},{D,E},{D,F}, {A,B,D}, {A,B,E}, {A,C,D}, {A,D,E} {A,D,F} ,{B,D,E},{A,B,D,E} }

سپس باید میزان اطمینان و پوشش را برای مجموعه فوق محاسبه‌کرد. در این محاسبه قانون $A \rightarrow X$ باید از نظر اطمینان و پوشش بررسی‌می‌شود که x کاربران می‌باشد و A سرویس مورد نظر می‌باشد. در این مثال درجه اطمینان ۵۰٪ در نظر گرفته‌شده‌است. رابطه پوشش و اطمینان به صورت زیر تعریف‌می‌شوند.

تعریف پوشش[1]: مجموعه I شامل تمام آیتم‌ها و مجموعه آیتم x را در نظر بگیرید، می‌گوییم پوشش x در پایگاه داده برابر است با ℓ اگر و فقط اگر تعداد وقوع مجموعه آیتم x در پایگاه داده برابر با ℓ باشد.

درجه اطمینان[2]: مجموعه I شامل تمامی اقلام و مجموعه آیتم‌های x و yمفروض‌اند. درجه اطمینان قانون x=>y برابر است با: $x \cap y = \emptyset$

$$S\ Conf(x \rightarrow y) = support\ (x\ U\ y) / Support\ (x)$$

$$Support\ (x \rightarrow y) = support\ (x\ U\ y) / m$$

در رابطه بالا m تعداد کل رکوردهای Answer می‌باشد.

جدول ۱۱: قوانین اطمینان و پوشش

مجموعه	قانون	اطمینان	پوشش	وضعیت
{A,B}	{A→B}	۳۶.۰	۲۳.۰	ضعیف
{A,C}	{A→C}	۱۸.۰	۱۲.۰	ضعیف
{A,D}	{A→D}	۵۴.۰	۳۵.۰	قوی
{A,E}	{A→E}	۳۶.۰	۲۳.۰	ضعیف
{A,F}	{A→F}	۱۸.۰	۱۲.۰	ضعیف
{A,B,D}	{A→B,D}	۱۸.۰	۱۲.۰	ضعیف
{A,B,E}	{A→B,E}	۱۸.۰	۱۲.۰	ضعیف
{A,D,E}	{A→D,E}	۱۸.۰	۱۲.۰	ضعیف
{A,C,D}	{A→C,D}	۰۹.۰	۰۶.۰	ضعیف
{A,D,F}	{A→D,F}	۰۹.۰	۰۶.۰	ضعیف
{A,B,D,E}	{A→B,D,E}	۰۹.۰	۰۶.۰	ضعیف

با توجه به نتایج خروجی در جدول (۱۰) سرویس‌های آموزشی که در دو معیار sup و conf مقدار بالای ۰/۵ دارند به عنوان سرویس‌های مطلوب و کاندید انتخاب‌می‌شوند و در ادامه با استفاده از الگوریتم بیزین بررسی‌می‌شوند که ایا برای کاربر مورد نظر دارای کیفیت هستند یا خیر و بهترین سرویس را به کاربر پیشنهادمی‌دهد.

۲,۴ انتخاب بهترین سرویس کاندید مبتنی بر شبکه بیزین

تا به اینجا با استفاده از الگوریتم گفته‌شده سرویس های کاندید برای پیشنهاد به کاربر آموزشی انتخاب‌شده‌اند و در این قسمت با استفاده از الگوریتم بیزین بهترین سرویس‌ها به کاربر ارائه‌خواهدشد. در این قسمت ورودی الگوریتم بیزین سرویس‌های آموزشی کاندید هستند و خروجی آن تعیین این موضوع که آیا این سرویس‌ها برای کاربر هدف از کیفیت قابل قبولی برخوردار هستند یا خیر.

الگوریتم بیزین برای انجام این کار از مجموعه داده آموزش استفاده‌می‌کند، در این مجموعه داده تعدادی سرویس آموزش با ویژگی‌های مختلف وجود دارد که مشخص شده‌است که دارای کیفیت هستند یا خیر، سرویس های با کیفیت دارای کلاس یک و سرویس های بدون کیفیت دارای کلاس صفر هستند که در جدول زیر قابل مشاهده‌می‌باشد.

[1] – support

[2] - confidence

در جدول (۱۲) تعداد ۱۵ بردار آموزش به‌عنوان مثال نمایش داده‌شده‌است. در این جدول هر ردیف یک سرویس آموزش می‌باشد که دارای ویژگی‌های F_1 تا F_5 می باشد.

جدول ۱۲: داده‌های آموزش

F۱	F۲	F۳	F۴	F۵	class
۵	۷	۹	۶	۵	۱
۵	۲	۳	۴	۷	۰
۶	۲	۵	۴	۳	۰
۱	۱	۲	۴	۶	۱
۳	۲	۴	۷	۷	۰
۵	۳	۳	۶	۵	۱
۴	۴	۲	۱	۴	۰
۲	۳	۵	۴	۲	۱
۴	۱	۴	۶	۵	۰
۲	۳	۲	۴	۲	۱
۳	۲	۱	۳	۱	۱
۲	۳	۵	۵	۱	۱
۱	۵	۳	۷	۴	۰
۲	۶	۲	۶	۳	۱
۵	۷	۱	۴	۲	۰

در این جا هدف ما این است با الگوریتم بیزین مشخص‌کنیم که سرویس‌های جدول تست دارای کیفیت هستند یا خیر؟ شبکه بیزین یک فرآیند یادگیری بر اساس نظریه یادگیری آماری است که یکی از بهترین روش‌های یادگیری ماشین مورد استفاده در داده کاوی می‌باشد. در مسائل مربوط به طبقه‌بندی داده‌ها و شناسایی الگو، همچون دسته‌بندی متون، تشخیص چهره در تصاویر، تشخیص ارقام دست نویس و بیوانفورماتیک کارکرد بسیار موفقی داشته‌است. شبکه بیزین درواقع یک طبقه‌بندی‌کننده دودویی است که دو کلاس را با استفاده از یک مرز خطی از هم جدامی‌کند. این روش با استفاده از یک الگوریتم بهینه‌سازی، نمونه‌هایی که مرزهای کلاس‌ها را تشکیل‌می‌دهند به دست می‌آورند. این نمونه‌ها را بردارهای پشتیبان گویند. تعدادی از نقاط آموزشی که کمترین فاصله تا مرز تصمیم‌گیری را دارند می‌توانند به‌عنوان زیر مجموعه‌ای برای تعریف مرزهای تصمیم‌گیری و به‌عنوان بردار پشتیبان در نظر گرفته‌شوند. فرض کنید داده‌ها از دو کلاس تشکیل شده و کلاس‌ها در مجموع دارای $L_1X_i=i, i=$..... نقطه آموزشی باشند که X_i یک بردار است. برای محاسبه مرز تصمیم‌گیری دو کلاس کاملاً جدا از هم از روش حاشیه بهینه استفاده‌می‌شود. در این روش مرز خطی بین دو کلاس به‌گونه‌ای محاسبه‌می‌شود که ابتدا: تمام نمونه‌های کلاس مثبت۱ در یک طرف مرز و تمام نمونه‌های کلاس منفی۱ در طرف دیگر مرز واقع‌شوند. سپس مرز تصمیم‌گیری به‌گونه‌ای باشد که فاصله نزدیک‌ترین نمونه‌های آموزشی هر دو کلاس از یکدیگر در راستای عمود بر مرز تصمیم گیری تا جایی که ممکن است حداکثر شود. یک مرز تصمیم گیری خطی را در حالت کلی می‌توان به‌صورت زیر نوشت:

$$w.\ x + b = 0$$

X یک نقطه روی مرز تصمیم گیری و w یک بردار n بعدی عمود بر مرز تصمیم‌گیری است. $b/||w||$ فاصله مبدأ تا مرز تصمیم-گیری و $w.\ x$ بیانگر ضرب داخلی دو بردار w و x است. از آنجاکه با ضرب یک ثابت در دو طرف معادله باز هم تساوی برقرار خواهدبود. اولین مرحله برای محاسبه مرز تصمیم‌گیری بهینه، پیداکردن نزدیک‌ترین نمونه‌های آموزشی دو کلاس است. در مرحله بعد فاصله آن نقاط از هم در راستای عمود بر مرزهایی که دو کلاس را به‌طور کامل جدامی‌کنند محاسبه‌می‌شود. مرز تصمیم‌گیری بهینه مرزی است که حداکثر حاشیه را داشته‌باشد.

۵. آزمایش‌ها

۱،۵ آزمایش اول

در این آزمایش رده‌بند بر اساس داده‌های غیرنرمال قسمت آموزش عمل‌می‌کند. پارامتر اصلی که روی کارایی رده‌بندی تأثیر زیادی دارد و عنصری مهم در قابلیت یادگیری است، شعاع نمونه‌های خودی یا RS^3 است. درصد تشابه هر داده تست با داده‌های قسمت آموزش که در وضعیت غیرنرمال دارند محاسبه‌می‌شوند. درصد تشابه برابر است با تعداد ویژگی‌های مشترک که بردار با بردارهای داده آموزش داشته‌باشد؛ که در آخر به‌صورت درصد نمایش داده‌می‌شود و بالاترین مقدار به‌عنوان درصد تشابه انتخاب‌می‌شود. مقدار RS برای یک بردار برابر با میانگین مقدار درصد تشابه کل بردارها است.

خروجی این رده‌بند زمانی یک می‌شود که درصد تشابه یک بردار از مقدار RS بیشتر باشد. در جدول (۱۳) مقدار درصد تشابه برای داده‌های تست با استفاده از داده‌های غیرنرمال آمده‌است.

جدول ۱۳: مقدار درصد تشابه داده‌های تست با داده‌های غیرنرمال

F۱	F۲	F۳	F۴	F۵	درصد تشابه با داده‌های غیر نرمال
۳	۱	۶	۷	۵	%۴۰
۴	۲	۳	۴	۵	%۶۰
۲	۵	۲	۲	۱	%۲۰
۶	۴	۴	۲	۳	%۴۰
۵	۳	۵	۱	۴	%۴۰
۷	۱	۲	۵	۴	%۴۰

با توجه به درصد تشابه در جدول (۱۳) برای داده تست که بر اساس داده‌های غیرنرمال داده آموزش به دست آمده‌اند مقدار RS برابر است با:

$$RS=(۰.۴+۰.۶+۰.۲+۰.۴+۰.۴+۰.۴)/۶ \rightarrow RS=۴۰\%$$

بنا بر آنچه گفته‌شد بردارهایی که مقدار تشابه آن‌ها از RS بیشتر باشد خروجی رده‌بند ۱ برای آن‌ها یک می‌شود.

۲،۵ آزمایش دوم

[3] Radius samples

در این آزمایش رده‌بند بررسی می‌کند که آیا بردارهای فعلی سرویس آموزشی در وضعیت نرمال قراردارند یا نه. این رده‌بند بر اساس داده‌های نرمال قسمت آموزش عمل‌می‌کند و اگر بردار نرمال باشد خروجی یک و در غیر این صورت خروجی صفر می‌شود. همانند رده‌بند اول باید مقدار RS برای این قسمت نیز محاسبه‌شود. محاسبه RS برای کل داده‌های تست بدین صورت می‌باشد.

درصد تشابه هر داده تست با داده‌های قسمت آموزش که در وضعیت نرمال دارند محاسبه‌می‌شوند. درصد تشابه برابر است با تعداد ویژگی‌های مشترک که بردار با بردارهای داده آموزش داشته‌باشد؛ که در آخر به‌صورت درصد نمایش داده‌می‌شود و بالاترین مقدار به‌عنوان درصد تشابه انتخاب‌می‌شود. مقدار RS برای یک بردار برابر با میانگین مقدار درصد تشابه کل بردارها است. خروجی این رده‌بند زمانی یک می‌شود که درصد تشابه یک بردار از مقدار RS بیشتر باشد. در جدول (۱۴) درصد تشابه داده‌های تست با داده‌های نرمال قسمت آموزش نمایش داده شده‌است.

جدول ۱۴: مقدار درصد تشابه داده های تست با داده‌های نرمال

درصد تشابه با داده‌های نرمال	F۵	F۴	F۳	F۲	F۱
%۲۰	۵	۷	۶	۱	۳
%۴۰	۵	۴	۳	۲	۴
%۴۰	۱	۲	۲	۵	۲
%۲۰	۳	۲	۴	۴	۶
%۴۰	۴	۱	۵	۳	۵
%۴۰	۴	۵	۲	۱	۷

با توجه به درصد تشابه در جدول (۱۴) برای داده تست که بر اساس داده‌های نرمال داده آموزش به دست آمده‌اند مقدار RS برابر است با:

$RS=(۰.۲+۰.۴+۰.۴+۰.۲+۰.۴+۰.۴)/۶ \rightarrow RS=۳۳\%$

بنا بر آنچه گفته‌شد بردارهایی که مقدار تشابه آن‌ها از RS بیشتر باشد خروجی رده بند ۲ برای آن‌ها یک می‌شود. مکانیسم انتخاب بهترین رده بند در فاز تست برای یک نمونه جدید ورودی، چهار حالت رخ می‌دهد:

در آزمایش اول اگر رده‌بند که ناحیه غیرخودی را می‌پوشاند خروجی یک داشته باشد؛ و در آزمایش دوم رده‌بند که ناحیه خودی را پوشش-می‌دهد، خروجی صفر داشته‌باشد. در این حالت با قطعیت ۱۰۰٪ می‌توان گفت نمونه جدید، متعلق به کلاس دو، یعنی یک نمونه ناهنجاری است و سیستم پیغام هشدار تولیدمی‌کند. همچنین اگر در آزمایش اول رده بند خروجی صفر داشته‌باشد و در آزمایش دوم رده‌بند خروجی یک داشته‌باشد. در این حالت با قطعیت ۱۰۰٪ می‌توان گفت نمونه جدید، متعلق به کلاس یک یعنی یک نمونه نرمال است.

و نهایتاً اگر هر دو آزمایش اول و دوم رده بند خروجی صفر داشته-باشند یا هر دو رده‌بند خروجی یک داشته‌باشند، اصطلاحاً چنین حالتی را همپوشانی می‌گوییم و برای رفع آن از روش انتخاب یکی از رده‌بندها که در ادامه توضیح‌داده‌می‌شود، استفاده‌می‌کنیم. در جدول (۱۵) نتیجه ارزیابی اولیه رده‌بندها نمایش‌داده‌شده‌است.

جدول ۱۵: نتیجه ارزیابی اولیه

class	خروجی رده بند ۲	خروجی رده بند ۱	F۵	F۴	F۳	F۲	F۱
غیر نرمال	۰	۱	۵	۷	۶	۱	۳
همپوشانی	۱	۱	۵	۴	۳	۲	۴
نرمال	۱	۰	۱	۲	۲	۵	۲
غیر نرمال	۰	۱	۳	۲	۴	۴	۶
همپوشانی	۱	۱	۴	۱	۵	۳	۵
همپوشانی	۱	۱	۴	۵	۲	۱	۷

بردارهای که class آن‌ها نرمال یا غیرنرمال شده‌است وضعیت مشخص دارند ولی برای مواردی که مقدار همپوشانی دارند طبق روش‌های گفته شده در زیر عمل‌می‌کنیم.

۳,۵ حالت همپوشانی(مبهم)

در آزمایش اول و دوم این حالت زمانی رخ‌می‌دهد که خروجی هر دو رده بند ۱ یا ۰ باشد و در این وضعیت از آمارگیری استفاده‌می‌کنیم. تعداد داده بالغ گروهی که شباهت بیشتری به داده تست داشته‌باشند برچسب آن برای داده تست انتخاب‌می‌شود؛ و اگر درصد تشابه برای هر دو گروه نرمال و غیرنرمال برابر بود به صورت زیر عمل می‌کند. در این روش هر ویژگی بردار اگر تنها یک اختلاف با ویژگی بردار ناحیه آموزش داشته‌باشد مشابه در نظر گرفته‌می‌شود. در جدول (۱۶) چند بردار در حالت همپوشانی قرارگرفته‌اند که برای مشخص کردن وضعیت نرمال و غیرنرمال آن‌ها از روش گ‌فته‌شده استفاده‌می‌کنیم که به‌صورت بردار به بردار در ادامه توضیح داده شده‌است. بردارهای که در وضعیت همپوشانی قراردارند با درصد تشابه آن‌ها نمایش داده شده‌است.

جدول ۱۶: وضعیت بردارهای که حالت همپوشانی دارند

class	درصد تشابه با داده‌های غیر نرمال	درصد تشابه با داده‌های نرمال	F۵	F۴	F۳	F۲	F۱
غیر نرمال	%۸۰	%۶۰	۵	۴	۳	۲	۴
غیر نرمال	%۸۰	%۶۰	۴	۱	۵	۳	۵
غیر نرمال	%۸۰	%۶۰	۴	۵	۲	۱	۷

حالت دوم: در این حالت کاربر سرویسی انتخاب-نمی‌کند و فقط هدف مورد نظر را مشخص‌می‌کند و سیستم چند سرویس با توجه به اطلاعات قبلی به کاربر پیشنهادمی‌دهد. اگر کاربر سرویسی را انتخاب کند که قبلاً هیچ کاربری دیگری آن را انتخاب نکرده‌باشد سیستم با حالت دوم خود سرویس به او پیشنهادمی‌دهد. روش کار به ترتیب زیر می‌باشد:

- کاربر عنوان هدف را انتخاب‌می‌کند.

- سیستم جدول مربوط به هدف مورد نظر را می‌سازد.
- تمام قوانین ممکن را با توجه به حد آستانه تولیدمی‌کند.
- قوانینی که از نظر پوشش و اطمینان قابل‌قبول باشند به‌عنوان پیشنهاد به کاربر ارائه‌می‌شوند.

۶. ارزیابی

در این پژوهش برای شبیه‌سازی از زبان متلب استفاده شده‌است و برای هر معیار ارزیابی چندین مرحله در نظر گرفته‌شده‌است و در هر مرحله اندازه مجموعه داده بزرگ‌تر می‌شود. برای استخراج نتایج هر مرحله چندین مرتبه شبیه‌سازی اجرا شده‌است و میانگین نتایج به‌عنوان نتیجه خروجی در نظر گرفته‌شده‌است. برای ارزیابی کارایی سیستم‌های طبقه-بندی معیارهای سنجش متفاوتی وجود دارد. از جمله این معیارها می توان به دقت طبقه‌بندی، فراخوانی، صحت و نرخ خطا اشاره‌کرد. ابتدا پارامترهایی که در تعریف این معیارها استفاده‌می‌شوند، معرفی‌می‌کنیم:
مثبت حقیقی(TP): برابر است با درصدی از تعداد اعضایی از یک کلاسی مانند X که سیستم طبقه‌بندی کننده آن‌ها را به‌درستی عضو کلاس X طبقه‌بندی کرده است.
منفی حقیقی(TN): برابر است با درصدی از تعداد اعضای کلاسهای دیگر که عدم تعلق آن‌ها به کلاس X به‌درستی طبقه‌بندی شده است.
مثبت کاذب (FP): برابر است با درصدی از تعداد اعضای کلاسهای دیگر که به‌اشتباه به عنوان عضوی از کلاس X طبقه‌بندی شده است.
منفی کاذب(FN): برابر است با درصدی از تعداد اعضای کلاس X که به‌اشتباه به عنوان عضوی از کلاسهای دیگر طبقه‌بندی شده‌اند.
مثبت(P): برابر با مجموع تعداد اعضایی از کلاس‌ها است که به صورت درست طبقه بندی شده‌اند.
منفی(N): برابر با مجموع تعداد اعضایی از کلاس‌ها است که به صورت اشتباه طبقه بندی شده‌اند.

معیار ارزیابی فراخوانی عبارت است از دقت سیستم طبقه‌بندی‌کننده در طبقه‌بندی درست اعضای از کلاس X که به صورت درست عضو کلاس X طبقه‌بندی شده‌اند و به صورت رابطه ۳ محاسبه-می‌شود:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (۳)$$

معیار ارزیابی صحت یعنی اینکه چند درصد از اعضای تشخیص‌داده-شده به عنوان عضو کلاس X درواقع به کلاس X متعلق هستند و به صورت رابطه (۴) محاسبه‌می‌شود:

$$\text{Precision} = \frac{\text{TP}}{\text{Tp + FP}} \qquad (۴)$$

همچنین دقت طبقه‌بندی شاخص دیگری برای ارزیابی عملکرد سیستم‌های طبقه‌بندی‌کننده می‌باشد و دارای دید و دامنه‌های وسیع‌تر و جامع‌تری نسبت به عملکرد سیستم‌های طبقه‌بندی‌کننده می‌باشد و عبارت است از نسبت تمام مواردی که به صورت صحیح طبقه‌بندی شده‌اند. دقت طبقه‌بندی به صورت رابطه ۵ محاسبه-می‌شود:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \qquad (۵)$$

علاوه بر سه روش فوق معیار دیگری به نام F-Score که میانگین وزن دار بین معیارهای صحت و فراخوانی است برای تعیین میزان کارایی سیستم های طبقه‌بندی‌کننده وجود که به صورت رابطه (۶) محاسبه-می‌شود.

$$\text{F} - \text{score} = ۲ * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (۶)$$

در نهایت میزان خطای سیستم ترکیبی پیشنهادی نیز با استفاده از رابطه (۷) محاسبه‌می‌گردد:

$$\text{Error rate} = ۱۰۰ - (\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}) \qquad (۷)$$

جدول ۱۷: نتایج بهینه‌سازی سیستم‌های آموزشی تحت وب با روش بیزین

تعداد کاربران	تعداد سیستم(سرویس)	زمان اجرا (ثانیه)	ACC	FPR-۱	FRR-۱	RECALL	PRECISION	F-MESURE
۵۰۰	۲۴	۱۷۴	۰,۹۴۲۰	۰,۰۱۸۳	۰,۱۲۶۴	۰,۹۷۶۹	۰,۸۶۳۳	۰,۹۱۶۶
۱۰۰۰	۲۴	۳۷۹	۰,۸۸۹۰	۰,۰۷۱۶	۰,۱۲۷۱	۰,۹۱۷۱	۰,۸۶۰۷	۰,۸۸۸۰
۱۵۰۰	۲۴	۶۷۴	۰,۸۷۲۷	۰,۰۸۷۹	۰,۱۲۲۰	۰,۸۹۹۵	۰,۸۶۵۳	۰,۸۸۲۱
۲۰۰۰	۲۴	۸۵۸	۰,۸۷۱۵	۰,۰۸۹۷	۰,۱۲۶۷	۰,۸۹۷۷	۰,۸۶۱۱	۰,۸۷۹۰

جدول(۱۷) عملکرد مدل پیشنهادی را در شرایط مختلف از نظر تعداد کاربران و سرویس‌ها ارزیابی‌می‌کند و نشان‌می‌دهد که الگوریتم‌های به‌کاررفته توانسته‌اند تعادلی مناسب میان کارایی، دقت، و زمان اجرا برقرارکنند. زمان اجرا، که نمایانگر بهره‌وری محاسباتی است، با افزایش تعداد کاربران و سرویس‌ها به‌صورت خطی افزایش یافته، اما همچنان در محدوده‌ای بهینه باقی‌مانده‌است. این موضوع نشان‌دهنده کارایی بالای الگوریتم Relief در کاهش ابعاد داده‌ها و مدیریت پیچیدگی آن‌ها پیش از ورود به فرآیند خوشه‌بندی و مدل شبکه بیزین است. دقت مدل در تمامی سناریوها بالای ۹۰ درصد گزارش‌شده که این میزان دقت، تأثیر مثبت کاهش ابعاد داده‌ها در حفظ اطلاعات مهم و حذف نویز را نشان‌می‌دهد.

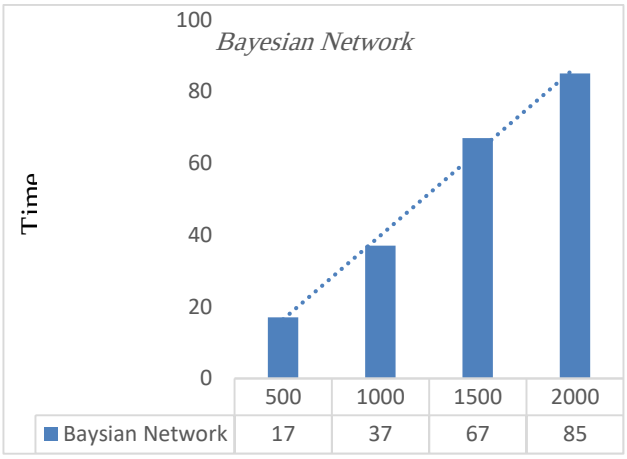
مقادیر FPR-1 و FRR-1 که به ترتیب بیانگر کاهش خطاهای مثبت و منفی کاذب هستند، در بازه ۹۵ تا ۹۸ درصد قرار دارند. این مقادیر بالا نشان‌می‌دهند که سیستم توانسته‌است هم از پیشنهاد

سرویس‌های نامرتبط جلوگیری و هم اطمینان حاصل‌کند که تمامی سرویس‌های مرتبط با کاربران شناسایی‌می‌شوند. Recall، که درصد سرویس‌های صحیح شناسایی‌شده را اندازه‌گیری‌می‌کند، در تمامی سناریوها بالای ۹۵ درصد بوده و نشان‌می‌دهد که مدل از نظر جامعیت عملکرد، بسیار قوی عمل کرده‌است. از سوی دیگر، Precision، که به کیفیت پیش‌بینی سرویس‌های مرتبط می‌پردازد، در بازه ۹۰ تا ۹۵ درصد باقیمانده و با افزایش تعداد کاربران و سرویس‌ها، کاهش جزئی آن قابل-مشاهده است. این کاهش اندک می‌تواند ناشی از افزایش هم‌پوشانی ویژگی‌های کاربران در مقیاس بزرگ‌تر باشد.

شاخص F-MESURE که ترکیبی موزون از Precision و Recall است، در تمامی سناریوها بالای ۹۲ درصد گزارش‌شده و نشان‌دهنده تعادل مناسب بین جامعیت و دقت مدل است. این مقادیر بالا تأییدمی‌کنند که ترکیب الگوریتم‌های Relief و K-Means نه تنها توانسته داده‌های پیچیده را به‌طور مؤثر پردازش‌کند، بلکه موجب بهبود کارایی مدل شبکه بیزین در پیش‌بینی و ارائه پیشنهادات شخصی‌سازی‌شده به کاربران شده‌است. به‌طور خاص، افزایش تعداد سرویس‌ها و کاربران منجر به تغییرات محسوسی در زمان اجرا و Precisionشده، اما سایر شاخص‌ها مانند Recall و ACC همچنان ثابت مانده‌اند که این نشان‌دهنده انعطاف‌پذیری و پایداری مدل پیشنهادی است. به‌طور کلی، جدول (۱۷) بیانگر این است که مدل پیشنهادی توانسته‌است در محیط‌های مقیاس‌پذیر، عملکردی مطلوب ارائه‌دهد و قابلیت استفاده در سامانه‌های آموزش الکترونیکی واقعی را دارد.

۱,۶ زمان اجرا

در این سناریو مدت زمان اجرای برنامه مورد ارزیابی قرارگرفته‌است. هر قدر زمان اجرا کمتر باشد الگوریتم کیفیت بالاتری دارد. در شکل (۳) زمان اجرای شبیه‌سازی آمده‌است.



شکل ۳: مدت زمان اجرا

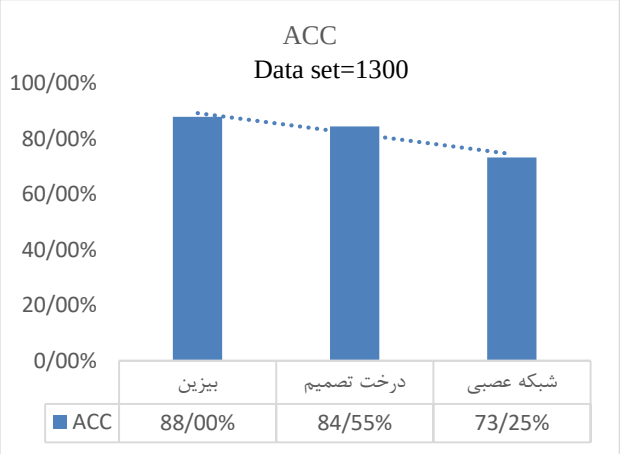
۲,۶ معیار دقت

همچنین دقت طبقه‌بندی شاخص دیگری برای ارزیابی عملکرد سیستم‌های طبقه‌بندی‌کننده می‌باشد و دارای دید و دامنه‌های وسیع‌تر و جامع‌تری نسبت به عملکرد سیستم‌های طبقه‌بندی‌کننده می‌باشد و عبارت است از نسبت تمام مواردی که به صورت صحیح طبقه‌بندی-شده‌اند. دقت طبقه‌بندی به صورت رابطه (۸) محاسبه‌می‌شود:

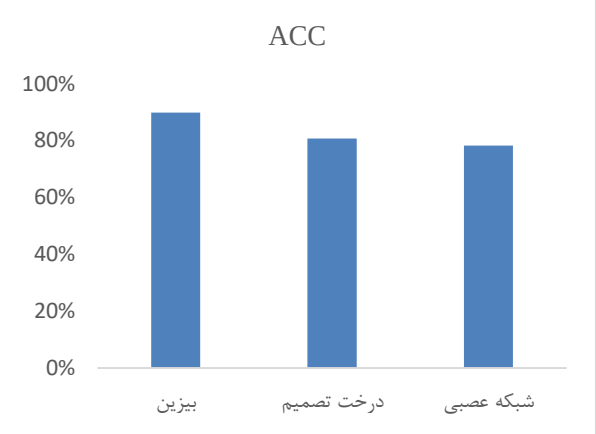
$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (۸)$$

همانطور که در شکل (۴) و (۵) ملاحظه‌می‌شود نتایج مقایسه با دیگر روش ها نشان می‌دهد که دقت بالاتر روش پیشنهادی در مقایسه با روش‌های درخت تصمیم و شبکه‌های عصبی، مستقیماً به ویژگی‌های الگوریتم‌های Relief و K-Means بازمی‌گردد که به‌طور بهینه در فرآیند کاهش ابعاد و خوشه‌بندی کاربران به‌کار گرفته شده‌اند. الگوریتم Reliefبا ارزیابی وزن ویژگی‌ها بر اساس تأثیر آن‌ها در تمایز نمونه‌های مشابه و نامشابه، توانسته‌است ویژگی‌های غیرمرتبط و نویزی را از مجموعه داده‌ها حذف‌کند. این فرآیند، داده‌های ورودی به مدل شبکه بیزین را ساده‌تر و مؤثرتر کرده‌است، به‌گونه‌ای که تنها ویژگی‌هایی که بیشترین تأثیر را در پیش‌بینی رفتار کاربران داشته‌اند، حفظ‌شده‌اند. در مقابل، روش‌های درخت تصمیم و شبکه‌های عصبی بدون کاهش مؤثر ابعاد یا تمرکز بر ویژگی‌های کلیدی، از داده‌های خام استفاده‌کرده‌اند، که این موضوع منجر به کاهش دقت آن‌ها شده‌است. به‌ویژه، در روش درخت تصمیم، وجود نویز یا ویژگی‌های زائد می‌تواند ساختار درخت را پیچیده کرده و منجر به بیش‌برازش شود. همچنین، شبکه‌های عصبی به‌دلیل وابستگی به تعداد زیادی از ویژگی‌ها، در مواجه با داده‌های دارای ابعاد بالا و نویز، دچار افت دقت شده‌است. از سوی دیگر، الگوریتم K-Means با خوشه‌بندی کاربران بر اساس رفتارها و ویژگی‌های مشابه، توانسته‌است ساختاری منسجم برای ورودی مدل شبکه بیزین ایجادکند. این خوشه‌بندی، امکان تفکیک بهتر کاربران و ارائه تحلیل‌های دقیق‌تر را فراهم-کرده‌است. علاوه بر این، K-Means با استفاده از فاصله اقلیدسی، کاربران را در خوشه‌هایی گروه‌بندی کرده که در داخل هر خوشه، شباهت‌ها حداکثر و تفاوت‌ها بین خوشه‌ها حداقل است. این ویژگی به مدل شبکه بیزین کمک کرده تا رفتارهای هر خوشه را با دقت بیشتری تحلیل‌کند و در نتیجه، پیشنهادات مرتبط‌تری برای کاربران ارائه‌دهد. ترکیب این دو الگوریتم باعث‌شده که مدل پیشنهادی نسبت به روش‌های سنتی مانند درخت تصمیم و روش‌های پیچیده‌تر مانند شبکه‌های عصبی، از داده‌های تمیزتر و خوشه‌بندی‌شده استفاده‌کند. Relief داده‌های ورودی را ساده کرده و K-Means این داده‌های ساده‌شده را در قالب خوشه‌های منطقی مرتب کرده‌است. این فرآیند موجب کاهش خطاهای مثبت و منفی کاذب شده و دقت را در مقایسه با سایر روش‌ها افزایش داده‌است. به‌طور کلی، ترکیب ویژگی‌های Relief و K-Means

توانسته دقت، پایداری و عملکرد روش پیشنهادی را در مقایسه با درخت تصمیم و شبکه‌های عصبی به‌طور قابل‌توجهی بهبود بخشد.



شکل ۴ : میزان دقت روش‌های مورد مقایسه برای ۱۳۰۰کاربر



شکل ۵: میانگین دقت روش‌ها

۷. نتیجه‌گیری و کارهای آینده

این پژوهش چارچوبی مبتنی بر ترکیب الگوریتم‌های Relief، K-Means و شبکه‌های بیزین ارائه‌کرد که با هدف بهبود عملکرد سیستم‌های آموزشی مبتنی بر وب طراحی شده‌بود. روش پیشنهادی تلاش‌کرد تا با شناسایی دقیق ویژگی‌های مهم، خوشه‌بندی کاربران مشابه و تحلیل روابط پیچیده میان داده‌ها، محدودیت‌های روش‌های سنتی را کاهش دهد و دقت پیش‌بینی و سرعت ارائه خدمات را افزایش دهد. نتایج آزمایش‌ها نشان‌داد که این چارچوب می‌تواند به‌طور مؤثری چالش‌هایی مانند آغاز سرد، مدیریت داده‌های ناقص و ارائه پیشنهادهای شخصی‌سازی‌شده را حل‌کند. در این پژوهش نکات برجسته ای وجوددارد که از جمله آن می‌توان به دقت بالا: روش پیشنهادی توانست با دقت پیش‌بینی متوسط ۹۲ درصد، به‌طور قابل‌توجهی از الگوریتم‌های رایج مانند درخت تصمیم (با دقت ۸۸ درصد) و شبکه عصبی (با دقت ۷۷ درصد) پیشی بگیرد. سرعت اجرا: زمان اجرای روش پیشنهادی با بهره‌گیری از خوشه‌بندی K-Means و کاهش ابعاد داده‌ها، در تمام آزمایش‌ها در محدوده بهینه باقی‌ماند و مقیاس‌پذیری نسبی آن در برابر افزایش تعداد کاربران و خدمات ثابت شد. کاهش خطاها: کاهش نرخ خطاهای مثبت کاذب و منفی کاذب به ترتیب به میزان ۵٪ و ۴٪ در مقایسه با روش‌های مرجع، نشان‌دهنده کیفیت بالای پیشنهادهای ارائه‌شده به کاربران بود. غلبه بر آغاز سرد: خوشه‌بندی کاربران بر اساس ویژگی‌های استخراج‌شده، امکان ارائه پیشنهادهای منطقی حتی در صورت کمبود داده‌های تاریخی برای کاربران جدید را فراهم‌کرد. انعطاف‌پذیری در مدیریت داده‌های ناقص: شبکه بیزین توانست با تحلیل احتمالی، از تأثیر منفی داده‌های ناقص بر عملکرد سیستم بکاهد، اشاره‌کرد.

محدودیت‌ها و چالش‌هایی در پژوهش ما وجود که با برطرف‌کردن آن‌ها می‌توان تحقیقات جامع‌تری انجام‌داد از جمله این موارد:

- حساسیت به کیفیت داده‌های ورودی: روش انتخاب ویژگی و خوشه‌بندی به‌شدت به کیفیت داده‌های اولیه وابسته بود. داده‌های نویزی می‌توانند بر نتایج تأثیر منفی بگذارند.
- عدم پشتیبانی از تحلیل‌های بلادرنگ: روش ارائه‌شده قادر به تطبیق سریع با تغییرات آنی در رفتار کاربران نبود و نیازمند دوره‌های مشخص برای به‌روزرسانی مدل بود.
- کمبود آزمایش‌های عملی در محیط‌های واقعی: آزمایش‌ها عمدتاً بر داده‌های شبیه‌سازی‌شده انجام‌شدند و اثربخشی مدل در سناریوهای واقعی و متغیر هنوز به‌طور کامل ارزیابی نشده‌است.

برای غلبه بر محدودیت‌های شناسایی‌شده و تقویت کارایی سیستم، پیشنهاد می‌شود محورهای زیر مورد بررسی و توسعه قرارگیرند:

- ادغام الگوریتم‌های فراابتکاری: استفاده از الگوریتم‌های فراابتکاری مانند ژنتیک، بهینه‌سازی ازدحام ذرات و کلونی مورچگان برای بهبود فرآیند انتخاب ویژگی و خوشه‌بندی کاربران. این روش‌ها می‌توانند با جستجوی راه‌حل‌های بهینه‌تر، تأثیر داده‌های ناقص را کاهش دهند و دقت پیش‌بینی را افزایش دهند.
- توسعه یادگیری عمیق: طراحی مدل‌هایی که از شبکه‌های عصبی کانولوشن (CNN) یا بازگشتی (RNN) در کنار شبکه‌های بیزین استفاده‌کنند، می‌تواند تحلیل الگوهای پیچیده‌تر را امکان‌پذیر سازد. این ترکیب به‌ویژه در تحلیل داده‌های غیرساختاریافته یا چندرسانه‌ای مفید خواهدبود.
- تحلیل بلادرنگ رفتار کاربران: توسعه سیستمی که بتواند تغییرات لحظه‌ای در رفتار کاربران را شناسایی و مدل را به‌صورت پویا به‌روزرسانی کند. این هدف می‌تواند با استفاده از روش‌های یادگیری آنلاین یا تطبیقی محقق‌شود.
- مدیریت داده‌های چندمنظوره: یکپارچه‌سازی داده‌های مختلف از منابع گوناگون مانند برنامه‌های تلفن همراه، وب‌سایت‌ها، و حسگرها برای ایجاد نمای کاملی از رفتار کاربران و بهبود شخصی‌سازی.

- توسعه مقیاس‌پذیری: بررسی و بهینه‌سازی زمان اجرا و کارایی در محیط‌های مقیاس بزرگ با استفاده از تکنیک‌های موازی‌سازی و رایانش ابری. این روش‌ها می‌توانند هزینه محاسباتی را کاهش داده و سیستم را برای حجم‌های بزرگ داده قابل‌اجرا کنند.
- مدیریت عدم‌قطعیت: استفاده از مدل‌های فازی در کنار شبکه‌های بیزین برای بهبود تحلیل در شرایطی که داده‌ها مبهم یا ناقص هستند.

پیاده‌سازی این پیشنهادها می‌تواند چارچوب پیشنهادی را به سطح جدیدی از کارایی، پایداری و کاربردپذیری در سیستم‌های آموزشی مبتنی بر وب ارتقادهد. این مسیر توسعه به افزایش رضایت کاربران، کاهش هزینه‌های عملیاتی و تقویت قابلیت‌های شخصی‌سازی منجر خواهدشد.

References

[1] M.I. Burhan, A. Kumala Jaya, and L. Adira. "Bayesian Intelligent Tutoring System for Vocational High Schools." PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik 9, no. 1, pp. 1-13, 2024.

[2] M. Tao, "Application of Bayesian Model in the Construction of Personalized Learning Platform for Internet Education," 2024 International Conference on Distributed Computing and Optimization Techniques (ICDCOT), Bengaluru, India, pp. 1-5, 2024.

[3] M. Chanthiran, A. B. Ibrahim, M. H. Abdul Rahman, and P. Mariappan, "Bayesian Network Approach in Educational Application Development: A Systematic Literature Review and Bibliometric Meta-Analysis", International Journal of Artificial Intelligence, vol. 9, no. 1, pp. 8-16, Jun. 2022.

[4] R. B. Kaliwal and S. L. Deshpande, "Assessment Study For E-Learning Using Bayesian Network," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, pp. 79-84, 2021.

[5] I. Nekhaev, I. Zhuykov, S. Manukyants, Maslennikov, A. (2020). Applying Bayesian Network to Assess the Levels of Skills Mastering in Adaptive Dynamic OER-Systems. In: Silhavy, R., Silhavy, P., Prokopova, Z. (eds) Software Engineering Perspectives in Intelligent Systems. CoMeSySo 2020. Advances in Intelligent Systems and Computing, vol 1294, 2020.

[6] G. Li and Y. Wang, "Learner's emotion recognition for intelligent education systems," in 3rd Adv. Inf. Technol. Electron. Automat. Control Conf., IEEE, pp. 754–758, 2018.

[7] D. Qiu, "Development and implementation of an intelligent learning system for ideological and political education," in Int. Conf. on Virtual Reality and Intelligent Systems, IEEE, pp. 289–292, 2018.

[8] Y. Zhang, "Exploration on teaching reform of architectural design course in higher vocational colleges in the era of intelligent education," Advances in Higher Education, vol. 3, no. 3, pp. 33–35, 2019.

[9] M. Ally, "Foundations of educational theory for online learning," in Theory and Practice of Online Learning, T. Anderson, Ed., Edmonton: AU Press, pp. 15–44, 2008.

[10] F. Canales et al., "Adaptive and intelligent web-based education system: Towards an integral architecture and framework," Expert Systems with Applications, vol. 33, no. 4, pp. 1076–1089, 2007.

[11] H. T. Kahraman, S. Sagiroglu, and I. Colak, "Development of adaptive and intelligent web-based educational systems," in 4th Int. Conf. on Application of Information and Communication Technologies, IEEE, pp. 1–5, 2010.

[12] Z. Mousavi and M. Khan Babayi, "A data mining-based method for guiding students in educational systems," in 1st National Conference on Business Management, 2013.

[13] R.B. Kaliwal, S.L. Deshpande. Bayesian Network Model Based Classifiers Are Used in an Intelligent E-learning System. In: Garg, D., Rodrigues, J.J.P.C., Gupta, S.K., Cheng, X., Sarao, P., Patel, G.S. (eds) Advanced Computing. IACC 2023. Communications in Computer and Information Science, vol 2053, 2024.

[14] C. He, W. Liu and J. Ren, "Bayesian Network Structure Learning: A Review," 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT), Changzhou, China, pp. 1-7,2022.

[15] Li, G., & Wang, Y. Research on leamer's emotion recognition for intelligent education system. In 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 754-758, October 2018.

[16] Qiu, D. Development and Implementation of Learning System of an Intelligent Learning System for Ideological and Political Education in Colleges under Mobile Platform. In 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS), pp. 289-292, August 2018.

[17] H.T. Kahraman, S. Sagiroglu and I. Colak. Development of adaptive and intelligent web-based educational systems. In 2010 4th International Conference on Application of Information and Communication Technologies, pp. 1-5, 2010, October.

[18] Hu, H., Li, J., Lei, X., Qin, P., & Chen, Q. Design of health statistics intelligent education system based on Internet+. In Journal of Physics: Conference Series (Vol. 1168, No. 6, p. 062003, February 2019.

[19] Xu, Y. Physical Education Evaluation System Analysis and Intelligent Evaluation System Design. Caribbean Journal of Science, 52(4), pp. 1521-1524, 2019.

[20] R. O'Brien, M. Hartnett, and P. Rawlins, (2019). The centralization of elearning resource development within the New Zealand vocational tertiary education sector. Australasian Journal of Educational Technology, 2019.

[21] H. Bahador, S. Aliaei, and A. Bagheri, "Presenting a model for identifying factors affecting students' academic guidance using decision tree and neural network," in 2nd National Congress of New Technologies of Iran with the Aim of Achieving Sustainable Development, Tehran, Center for Achieving Sustainable Development Solutions, Mehr Arvand Higher Education Institute, 2015.

[22] B. Maghsoudi, S. Soleimani, A. Amiri, and M. Afsharchi, "Improving the quality of education in e-learning systems using educational data mining," Journal of Educational Technology, vol. 6, 2012.

[23] B. Minae, S. S. Mirfazel, and S. H. Hani, "Identifying factors affecting academic failure using association rules and cluster analysis," in 6th International Conference on Data Mining of Iran, 2013.

[24] Z. S. Mousavi and M. Khanbabaei, "Presenting a method based on a hybrid data mining technique for students' academic guidance," in 1st National Conference on Business Management, Hamedan, Tolou Farzin Science and Industry Company, Bu-Ali Sina University, 2013.

[25] F. Gharebaghi and A. Amiri, "A Novel Semi-Supervised Approach to Improve E-learning Performance," *Intelligent Multimedia Processing and Communication Systems*, vol. 3, no. 3, pp. 53-63, 2022.

[26] N. Farghzadeh and N. Modiri, "Developing a Performance-Aware Cloud Infrastructure in E-Learning and Virtual Systems," Intelligent Multimedia Processing and

Communication Systems, vol. 2, no. 1, pp. 53-64, Mar. 2021.
[27] M. Bazargani, S. H. Alizadeh, and B. Masoumi, "Group deep neural network approach in semantic recommendation system for movie recommendation in online networks," *Electron. Commer. Res.*, 2024.