Islamic Azad University , Shiraz Branch
نشریه تحلیل مدارها، داده ها و سامانه ها
Journal of Circuits, Data and Systems Analysis
sanad.iau.ir/journal/jcdsa

بهبود سیستم‌های توصیه‌گر وب با استفاده از مهندسی ویژگی برای پیش‌بینی لینک‌های بعدی کاربران

وحید صفاری[۱]، کرم الله باقری فرد[۲*]، حمید پروین[۳]، صمد نجاتیان[۴]، وحیده رضایی[۵]

۱- گروه مهندسی کامپیوتر، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
Vahid.Saffari@iau.ac.ir
۲- گروه مهندسی کامپیوتر، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
ka.bagherifard@iau.ac.ir
۳- گروه مهندسی کامپیوتر، واحد نورآباد ممسنی، دانشگاه آزاد اسلامی، نورآباد، ایران
parvinhamid@gmail.com
۴- گروه مهندسی برق، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
Sa.Nejatian@iau.ac.ir
۵- گروه ریاضی، واحد یاسوج، دانشگاه آزاد اسلامی، یاسوج، ایران
vahidehrezaie80@gmail.com

چکیده: در دوران رشد چشم‌گیر محتوای آنلاین، درک رفتار کاربر و ارائه پیشنهادات دقیق محتوا، چالش اساسی است. با توجه به محدود بودن ویژگی‌های موجود در لاگ سرور، روش‌های قبلی صرفاً با توجه به این ویژگی‌ها اقدام به طراحی سیستم توصیه‌گر نموده‌اند؛ ولی لازم است با استفاده از ویژگی‌ها و روش‌های جدید، دقت مدل افزایش یابد. این مقاله رویکرد جامعی را برای افزایش دقت پیشنهاد صفحات وب ارائه می‌دهد؛ به صورتی که با استخراج ویژگی "مدت ارتباط کاربر با صفحه"، خوشه‌بندی کاربران انجام گرفته و در نهایت برای هر کدام از کاربران یک پروفایل با استفاده از دنباله کاوی ایجاد می‌گردد. این ویژگی جدید توانسته به طور قابل‌توجهی در بهبود پروفایل‌های کاربر موثر باشد و پیش‌بینی بازدید بعدی کاربر را تسهیل کند. ارزیابی این مدل، نشان می‌دهد که اضافه کردن این ویژگی باعث افزایش قابل ملاحظه‌ای در دقت پیش‌بینی می‌شود. همچنین، تحلیلی از خوشه‌بندی با استفاده از الگوریتم‌های k-means و k-medoids نشان می‌دهد که روش دوم تنوع بیشتری در دسته‌بندی نمونه‌ها داشته و برتری آن را در این زمینه نشان می‌دهد. در نهایت، این مقاله به توسعه یک سیستم پیشنهاد وب جهت پیش‌بینی مقصد وب بعدی کاربر، منجر شد.

واژه های کلیدی: وب کاوی، مهندسی ویژگی، مدل‌سازی رفتار کاربر، پیش‌بینی صفحه بعدی

Improving Web Recommender Systems Using Feature Engineering to Predict Users' Next Links

Vahid Saffari[1], KaramAllah Bagherifard[2*], Hamid Parvin[3], Samad Nejatian[4], Vahideh Rezaie[5]

[1] Department of Computer Engineering, Yasuj Branch, Islamic Azad University, Yasuj, Iran
Vahid.Saffari@iau.ac.ir
[2] Department of Computer Engineering, Yasuj Branch, Islamic Azad University, Yasuj, Iran
ka.bagherifard@iau.ac.ir
[3] Department of Computer Engineering, Nurabad Branch, Islamic Azad University, Nurabad, Iran
parvinhamid@gmail.com
[4] Department of Electrical Engineering, Yasuj Branch, Islamic Azad University, Yasuj, Iran
Sa.Nejatian@iau.ac.ir
[5] Department of Mathematics, Yasuj Branch, Islamic Azad University, Yasuj, Iran
vahidehrezaie80@gmail.com

* نویسنده مسئول

Abstract:
In the era of exponential growth in online content, understanding user behavior and providing accurate content recommendations pose fundamental challenges. Due to the limited features available in server logs, previous methods solely designed recommendation systems based on these features. However, there is a need to enhance model accuracy through new features and approaches. This article introduces a comprehensive approach to improving the precision of web page recommendations. By extracting the feature "user's duration of interaction with a page," user clustering is performed, resulting in the creation of profiles for each user using Mining Sequence. This new feature significantly contributes to enhancing user profiles, facilitating the prediction of a user's next web page visit. Evaluation of this model demonstrates a noticeable increase in prediction accuracy with the addition of this feature. Furthermore, clustering analysis using k-means and k-medoids algorithms indicates that the latter provides greater diversity in sample categorization, showcasing the superiority of using k-medoids in this domain. Ultimately, this article contributes to the development of a web recommendation system capable of predicting the next web destination for users.
Keywords: Web mining, Future Engineering, User Behavior Modeling, Next Web Page Prediction

نوع مقاله:
DOI: 00.00000/0000
تاریخ ارسال مقاله: ۱۴۰۲/۱۰/۱۳ تاریخ پذیرش مقاله: ۱۴۰۲/۱۲/۰۳ تاریخ چاپ مقاله: ۱۴۰۲/۱۲/۲۷

۱- مقدمه

وب‌کاوی، فرایند خودکاراستخراج اطلاعات از اینترنت با استفاده از تکنیک‌های داده‌کاوی است؛ که به سه دسته اصلی تقسیم می‌شود: وب‌کاوی ساختاروب (WSM[1])، وب‌کاوی محتواوب (WCM[2]) وب‌کاوی استفاده از وب (WUM[3]). وب‌کاوی محتوای وب، استخراج نکات ارزشمند از انواع مختلف محتوای صفحات وب شامل متن، تصاویر، ویدیوها و صداها است. وب‌کاوی ساختار وب بر روی کشف اطلاعات مفید از پیوندهای صفحات وب تمرکز دارد و همچنین تجزیه و تحلیل پیوندهای خروجی و ورودی را بررسی می‌کند که اغلب برای رتبه‌بندی صفحات وب استفاده می‌شود. وب‌کاوی استفاده از وب فعالیت و ورودی‌های جستجو را برای شناسایی الگوها و رفتارهای مشترک کاربران مورد بررسی قرار می‌دهد و کاربردهای آن شامل توسعه پروفایل کاربر است [۱].

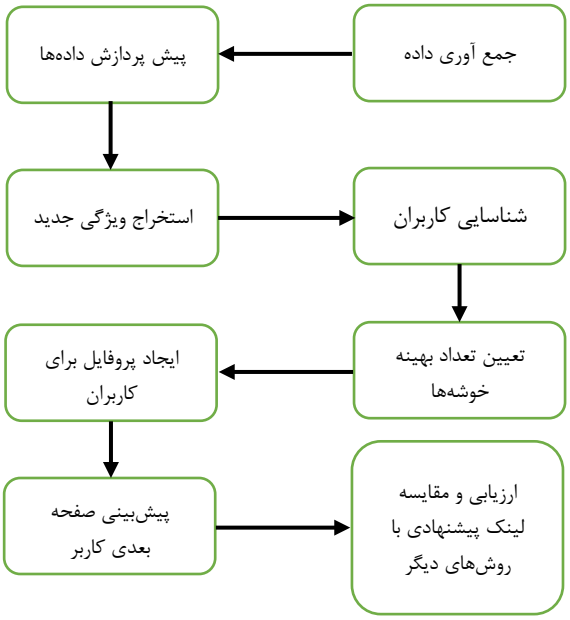
در سیستم‌های توصیه‌گر وب در دوران دیجیتالی، ارائه توصیه‌های شخصی‌سازی شده از اهمیت بسزایی برخوردارند. آن‌ها به کاربران کمک می‌کنند تا در حضور گسترده‌ی اطلاعات آنلاین، گام‌های بعدی خود را بردارند. پیش‌بینی لینک یا توصیه بعدی در توصیه‌های وب یک چالش مهم است [۲]. در میان اصولی که در این تحقیق به آن توجه شده، نقش حیاتی مهندسی ویژگی در بهینه‌سازی سیستم‌های پیشنهادی وب قرار دارد. مهندسی ویژگی به عنوان یک ابزار اساسی، عملکرد را بهبود بخشیده و امکان ساخت نمایش‌های دقیق و مفیدتری از رفتار و ترجیحات کاربران را فراهم می‌کند. اهمیت بیشتر مهندسی ویژگی در این است که می‌تواند الگوهای پیچیده در تعاملات کاربران را آشکار سازد. از طریق مهندسی ویژگی موثر، هدف ما گشایش یافتن فهم عمیق‌تری از ترجیحات هر فردی است که این امکان را فراهم می‌کند که سیستم‌های پیشنهادی وب، پیشنهاداتی ارائه دهند که نه تنها با علایق کاربران همخوانی داشته باشند، بلکه با پیشروی و تحولات مستمر با مسیر پیمایش‌های آنلاین آن‌ها نیز هماهنگی داشته باشند [۳]. از آنجایی که اکثرا کارهای مشابه، روش‌های پیشنهادی خود را با استفاده از ویژگی‌های موجود در فایل لاگ‌سرور ارائه داده‌اند این تحقیق در نظر دارد با استخراج ویژگی جدید، نشان دهد که در مقایسه با روش‌های موجود، دقت مدل سیستم توصیه‌گر بهبود پیدا می کند. این تحقیق اهداف متعددی در حوزه سیستم‌های توصیه وب دارد. در ابتدا، به‌طور دقیق اثربخشی تکنیک‌های مهندسی ویژگی در سیستم‌های توصیه وب بررسی می‌گردد. سپس از طریق تحلیل دقیق، پتانسیل این تکنیک‌ها برای بهبود کیفیت توصیه و بهبود انتظارات کاربر بررسی می‌شود. جنبه دوم تحقیق بر تأثیر مهندسی ویژگی بر پیش‌بینی لینک بعدی و دقت آن تمرکز دارد. در این مقاله به بررسی اهداف کلیدی تحقیق که شامل استخراج ویژگی، روش‌های نوآورانه و پتانسیل تحولی آن‌ها می‌شود، پرداخته شده است.

روش‌شناسی تحقیق، با فرآیندهای جمع‌آوری داده‌های دقیق و گوناگون شروع می‌شود. سپس تکنیک‌های مهندسی ویژگی توصیف می‌شود که نقش ویژگی‌های مهندسی در بهبود دقت پیش‌بینی لینک بعدی را برجسته می‌کند. انتخاب مدل و معیارهای ارزیابی مورد بحث قرار می‌گیرد تا رویکرد تحقیق را روشن کند. فلوچارت کلی تحقیق در شکل (۱) نشان داده شده است. ساختار این مقاله شامل بررسی ادبیات و کارهای پیشین، توضیحات دقیق از روش پیشنهادی، آزمایشات و نتایج، بحث درباره یافته‌ها و جهت‌های تحقیقات آینده است.

۲- پیشینه تحقیق

با توجه به تمرکز روزافزون جامعه توسعه‌دهندگان بر بهبود تجربه کاربری (UX[4]) به دلیل دیجیتالی شدن، این تحقیق یک سیستم توصیه‌گر وب نوآورانه را معرفی می‌کند که از داده‌های دنباله‌ی مرور وب کاربر استفاده می‌کند. با استفاده از خوشه‌بندی Fuzzy C-Means (FCM)، سیستم خوشه‌های بهتر N-tuples و کاربران مشابه

[1] Web structure Mining
[2] Web Content Mining
[3] Web Usage Mining
[4] User Experience



شکل (۱): فلوچارت کلی تحقیق

برای کاربر هدف را شناسایی می‌کند. همچنین وزن هر صفحه وب را ارزیابی کرده و مسائل موجود در سیستم‌های توصیه را مورد بررسی قرار داده تا توصیه‌هایی برای پیش‌بینی بازدید صفحه وب بعدی کاربر ارائه دهد. از طریق آزمایش‌ها با استفاده از داده‌های واقعی مجموعه‌داده MSNBC که شامل ۵۰۰۰ ورودی کاربر می‌باشد، دقت مدل پیشنهادی به‌طور تقریبی سه برابر برخی سیستم‌های موجود با نرخ دقت حدود ۳۳٪ است [۴]. تحقیق دیگری بر روی چالش توسعه یک سیستم توصیه بسیار دقیق برای صفحات وب جدید براساس ترجیحات کاربران تمرکز دارد. از یک الگوریتم ژنتیک خودکار خوشه‌بندی (ACGA) و یک مدل مارکوف سومین درجه اصلاح‌شده استفاده می‌کند. داده‌های رفتار کاربر، مانند فرکانس وقوع و زمان بازدید از صفحات، از فایل‌های log سرور وب استخراج می‌شود. ACGA از دو مرحله خوشه‌بندی و یک الگوریتم ژنتیک برای ایجاد خوشه‌ها تشکیل شده است. مدل مارکوف سومین درجه اصلاح‌شده به طور موثر صفحه بعدی را در سیستم توصیه پیش‌بینی می‌کند. آزمایشات با مجموعه داده CTI واقعی نشان می‌دهد که سیستم توصیه پیشنهادی بهتر از دیگر روش‌ها عمل می‌کند و دقت بالاتری دارد [۵]. در سال‌های اخیر، تحقیقات گسترده‌ای بر روی بهبود سیستم‌های وب با رویکردهای توصیه‌گر کارآمد و قابل اعتماد تمرکز داشته است. در حالی که انواع مختلفی از روش‌ها برای بهبود توصیه‌های کاربر بررسی شده‌اند، چندین چالش هنوز هم در سیستم‌های فعلی توصیه‌گر وجود دارند. مطالعه بعدی یک چارچوب جامع ارائه می‌دهد که به چالش‌های اساسی مانند قابلیت مقیاس‌پذیری، پراکندگی داده، مشکلات شروع سرد و مسائل نقطه مبنای توصیه پرداخته است. این چارچوب اساساً برای سیستم‌های توصیه فیلم طراحی شده و در این حوزه اعتبارسنجی شده است، اما به راحتی می‌تواند به حوزه‌های دیگری نیز تطبیق یابد. این چارچوب، مدل‌های پیش‌بینی مختلف را مدیریت می‌کند و از داده‌های ساختارمند مشتق‌شده از الگوریتم‌های داده‌کاوی استفاده می‌کند تا توصیه‌های وابسته به زمینه مبتنی بر ویژگی‌های محصول و کاربر در یک انتزاع حوزه خاص ارائه دهد [۷].

یک سیستم توصیه‌گر به عنوان یک عنصر بی‌نیاز برای وب‌سایت‌ها و برنامه‌ها حضور دارد و به کاربران یک رابط قابل اعتماد ارائه می‌دهد. این سیستم کاربران را در جستجو و بهره‌برداری بهینه از منابع وب کمک می‌کند. عوامل متعددی در جمع‌آوری داده‌های مرتبط با علایق کاربر و بهره‌گیری از آن برای ساخت یک سیستم توصیه‌گر ماهر مؤثر هستند. وب معنایی نقش حیاتی در ترویج پروتکل‌های مبتنی بر انتولوژی بازی می‌کند که ماهیت و کارایی سیستم توصیه‌گر مورد نظر را تعریف می‌کنند. علاوه بر این، اصول مختلفی در توسعه این سیستم‌ها به کار می‌رود. سیستم‌های توصیه‌گر، تأثیر بسیار مهمی در ارتقای تجربه کاربری وب‌سایت‌ها و برنامه‌ها دارند و اطمینان می‌دهند که محتواها و منابع مناسبی به کاربران ارائه می‌شود[۴]. سیستم‌های توصیه‌گر نقش حیاتی در دوران دیجیتال ایفا می‌کنند و با ارائه محتواها و توصیه‌های محصول سفارشی، به کاربران کمک می‌کنند تا در سایه گسترده‌ی خدمات آنلاین، گام‌های خود را بردارند. همانطور که حجم داده‌های اینترنت به افزایش ادامه می‌دهد، دقت و تنوع این توصیه‌ها در اهمیت زیادی افزایش می‌یابد. یک چالش مهم در مواجهه با این سیستم‌ها، "مشکل شروع سرد[1]" است که به توصیه‌هایی برای کاربران جدید یا معرفی محتوای تازه به کاربرانی که با آن‌ها قبلاً تعاملی نداشته‌اند، اشاره دارد. برای رفع این چالش‌ها، الگوریتم‌های پیشرفته مانند KNN2 وSVD استفاده می‌شوند که بر اساس مفاهیم تجزیه ماتریس عمل می‌کنند. این سیستم‌ها در تولید توصیه‌های دقیق و بهبود تجربه کاربری از طریق کاهش بارهای تصمیم‌گیری عالی عمل می‌کنند. موفقیت آن‌ها بستگی به توانایی درک ترجیحات کاربران، حفظ پایداری در میان تغییرات پروفایل‌های کاربری و افزایش تنوع توصیه‌ها دارد. مقالات تحقیقاتی در این حوزه، بینش‌ها و ارزیابی‌های عملی از انواع مدل‌های توصیه‌گر را ارائه می‌دهند [۸-۱۲].

سیستم‌های توصیه‌گر که از تکنیک‌های یادگیری ماشین استفاده می‌کنند، در حوزه سلامت به‌طور قابل توجهی جلب توجه کرده‌اند. با گسترش اطلاعات درمان آنلاین، بیماران به دنبال یافتن پزشکان مناسب آنلاین بیشتر می‌گردند. برای پاسخ به این نیاز، یک سیستم توصیه دهنده دکتر نوعی از سیستم توصیه‌گر ترکیبی ارائه داده است که از روش‌های متنوع توصیه‌گری مانند مبتنی بر محتوا، همکاری و فیلترینگ مشارکتی استفاده می‌کند. این سیستم با تحلیل ترجیحات بیماران برای انتخاب پزشک و استفاده از الگوریتمی تطبیقی برای ایجاد یک تابع رتبه‌بندی پزشکان، اولویت‌بندی شخصی‌سازی را در نظر می‌گیرد. این سیستم رتبه‌بندی معیارهای بیمار را کمّی کرده و

[1] Cold Start

[2] k-nearest neighbors

توصیه‌های منطقی و موثر برای انتخاب پزشکان ارائه می‌دهد که توسط ارزیابی‌های جامع تأیید شده است [۹]. پژوهشی دیگر، یک سیستم توصیه صفحات وب نوآورانه را معرفی می‌کند که هدف آن بهبود تجربه مرور برای کاربران ناشناس است. برخلاف تکنیک‌های سنتی وب کاوی، این سیستم از یک رویکرد چند مرحله‌ای برای پیش‌پردازش داده، کشف ارتباط بین صفحات وب، خوشه‌بندی و طبقه‌بندی داده استفاده می‌کند که اندازه‌گیری‌های منحصر به فردی مانند ماتریس فواصل، ماتریس فراوانی و ماتریس ارتباطات برای مدل‌سازی ارتباطات صفحات وب استفاده می‌شوند. گراف مجازی بر اساس ماتریس ارتباطی ایجاد می‌شود و سپس به خوشه‌هایی مبتنی بر الگوریتم جستجوی عمقی بهبود یافته، به عنوان نماینده الگوهای مسیریابی تقسیم می‌شود. کاربر فعال با استفاده از الگوریتم LCS به یکی از این خوشه‌ها دسته‌بندی می‌شود و مقدار آستانه‌ای به کار گرفته می‌شود تا تعداد بهینه‌ای از صفحات وب توصیه شود. اندازه‌گیری‌های نوآورانه این روش، استفاده از مقدار آستانه در هنگام تشکیل و توصیه خوشه، منجر به بهبود معیارهای عملکرد می‌شود. این بهبودها شامل دقت بیشینه ۶۱٪، پوشش میانگین ۴۹/۲٪، امتیاز F1 میانگین ۲۸/۸۷٪، تطابق بازدید میانگین ۵۷/۸٪ در تشکیل خوشه و حداقل ۱۵٪ نقاط پرت در فرآیند تشخیص پرت داده می‌شود [۱۰].

در حال حاضر، برای توصیه‌های مبتنی بر محتوا، تحلیل معنایی صفحات وب به چالش اساسی تبدیل شده است. در مطالعه دیگری، یک رویکرد استخراج معنایی محتوای وب برای سیستم‌های توصیه در فروشگاه‌های آنلاین معرفی شده است. این روش بر اساس دو مرحله اصلی استوار است: مرحله اول شامل پیش‌پردازش معنایی داده‌های متنی با استفاده از ترکیبی از انتزاعات واژگان توسعه یافته و موجود است. مرحله دوم از الگوریتم Naive Bayes برای ارائه توصیه‌ها استفاده می‌کند. خروجی سیستم با استفاده از دقت، بازخوانی و امتیاز F ارزیابی می‌شود. نتایج نشان دادند که پیش‌پردازش معنایی، دقت سیستم توصیه را نسبت به روش KNN، ۵/۲ ٪ افزایش داده است. علاوه بر این، سیستم توانایی ارائه یک بستر برای توصیه‌های مبتنی بر محتوا در فروشگاه‌های آنلاین دارد. این سیستم به روش‌های توصیه موجود با این ویژگی برتری دارد که می‌تواند بازخورد کاربران را در مورد محتوای متنی مرتبط با محصول تحلیل کرده و توصیه‌های لازم را ارائه دهد [۵]. تحقیق بعدی به چالش ارائه اطلاعات بهداشتی کاربرپسند از طریق توسعه یک سیستم توصیه صفحات وب می‌پردازد. سیستم معروف به PASO-WPR از بهینه‌سازی اجتماع ذرات و دو عامل کلیدی یعنی عامل ذره کاربر و عامل ذره معنایی استفاده می‌کند. این سیستم با بهره‌گیری از اطلاعات معنایی و تکنیک‌های وب‌کاوی، صفحات وب را بر اساس شباهت‌های معنایی آن‌ها به‌جای مکان‌های یکسان به منابع متحد کلاس‌بندی می‌کند. این خوشه‌ها برای ارائه توصیه‌های صفحات وب منطبق با نیازهای هر کاربر به کار می‌روند. سیستم PASO-WPR نسبت به روش‌های دیگر در شناسایی صفحات وب که کاربران قصد بازدید از آن‌ها را دارند، عملکرد بهتری داشته و دقت و پوشش را نسبت به سیستم‌های توصیه مبتنی بر محتوا بهبود می‌بخشد [۱۱-۱۵].

در تحقیق دیگری، از یک سامانه توصیه‌گر مبتنی بر وب برای هدایت مشاوران در راهنمایی دانش‌آموزان برای انتخاب مسیر شغلی و حرفه‌ای استفاده شده است. با توجه به عدم اطمینان دانش‌آموزان در مورد انتخاب مسیر تحصیلی و افزایش نرخ انصراف از تحصیل، این سامانه با بهره‌گیری از الگوریتم شبکه عصبی عمیق، موفق به پیش‌بینی مسیر تحصیلی دانش‌آموزان با دقت ۸۳/۱۱ ٪ شده است. این مطالعه با استفاده از نمرات و پروفایل‌های اجتماعی به عنوان عوامل تعیین کننده، شامل ۱۵۰۰ دانش‌آموز از پایه اول تا سوم برنامه K-12 را مورد بررسی قرار داده است. یافته‌های تحقیق نشان می‌دهد که الگوریتم شبکه عصبی عمیق به‌عنوان ابزاری موثر در پیش‌بینی مسیر تحصیلی می‌تواند نقش مهمی در بهبود عملکرد مشاوران راهنمایی ایفا کند. به این ترتیب، سامانه توصیه‌گر مسیر حرفه‌ای مبتنی بر وب به عنوان یک ابزار تصمیم‌گیری جهت هدایت دانش‌آموزان به مسیرهای تحصیلی اثربخش و یکپارچه می‌باشد [۱۶]. در مطالعه بعدی، اهمیت بررسی نقش توصیه‌گر صفحات وب در چارچوب‌های وب هوش مصنوعی و استخراج اطلاعات معنادار از وب به وسیله توصیه به صفحات وب به عنوان یک مسئله کلیدی مورد توجه قرار گرفته است. نتایج مدل‌های معمولی مطابقت کلیدی و آماری نشان داده که اغلب به‌طور ناسازگار با پرس و جوهای کاربر عمل می‌کنند. این تحقیق به یادگیری نیازهای هر کاربر بر اساس داده‌های اولویتی که در اختیار دارد، توجه کرده و از چارچوب گراف با استفاده از الگوریتم وزن‌دار گستون برای تمثیل مناسب داده‌های کاربر مورد استفاده نموده است. با استفاده از مدل توصیه جدید THHO_DFC با بهینه‌سازی تیلور هرد و ترکیب یادگیری عمیق فازی، بهبود چشمگیری در دقت، بازخوانی، اندازه‌گیری F و دقت مدل K-Nearest Neighbor مبتنی بر اصلاح لاپلاس در توصیه به صفحات وب مناسب به دست آمده است. این پژوهش نشان می‌دهد که مدل پیشنهادی عملکرد بهتری در زمینه کارایی و توصیه به صفحات وب ارائه می‌دهد [۱۷]. در مقاله دیگر، نگاهی جامع به چالش‌های یافتن اطلاعات در دوران داده‌های بزرگ ارائه شده و نیاز به سیستم‌های فیلترینگ اطلاعات برای کمک به کاربران در پیدا کردن اطلاعات مورد نیاز آنها بررسی شده است. این مقاله با تأکید بر سیستم‌های توصیه‌گر به برنامه‌های عملیاتی واقعی، تکنیک‌ها و توسعه‌های مختلف در زمینه‌های متنوع از جمله تجارت الکترونیک، گردشگری الکترونیک، منابع الکترونیکی، دولت الکترونیک، یادگیری الکترونیکی و کتابخانه الکترونیکی را مورد بررسی قرار می‌دهد. تحلیل این مقاله امکان ارائه نگاه دقیقتری به توسعه‌های کنونی و شناسایی چالش‌ها و مشکلات موجود در سیستم‌های توصیه‌گر را فراهم می‌آورد و نتایج به عنوان راهنمایی و پیشنهادات برای افراد عملی و محققان در زمینه سیستم‌های توصیه‌گر ارائه شده است [۱۸].

۳- روش پیشنهادی

۳-۱- جمع‌آوری داده

سایت Ehadish.com یک بازار آنلاین معروف است که محصولات و خدمات متنوعی را ارائه می‌دهد. این وب‌سایت با رابط کاربری دوستانه و دسته‌بندی‌ها و فیلترهای جستجوی واضح، امکان مرور آسان را برای کاربران فراهم می‌کند و محصولاتی از جمله تجهیزات الکترونیک، دستگاه‌های موبایل و... را ارائه می‌دهد [۶]. این سایت به دلیل حضور قوی در رسانه‌های اجتماعی و تعهد به رضایت مشتریان شناخته شده است. این مطالعه از داده‌های سال ۲۰۲۲ تا ۲۰۲۳ که شامل فایل‌های لاگ سرور و اطلاعات مربوط به محتوای وب‌سایت می‌شود، استفاده کرده است. جدول (۱)، یک مرور بر تعداد رکوردهای مرتبط با فایل‌های لاگ قبل و بعد از پیش‌پردازش و تقسیم جلسه ارائه می‌دهد. این جدول نشان می‌دهد که پایگاه‌داده نهایی شامل ۲۱۹۱۶ جلسه با میانگین طول جلسه ۶/۵ لینک در هر جلسه است.

۳-۲- پیش پردازش

وب کاوی فرآیندی است که اطلاعات ارزشمندی را از داده‌های وب استخراج می‌کند. پیش‌پردازش داده مرحله‌ای حیاتی در این فرآیند است که شامل پاکسازی و آماده‌سازی داده‌های خام می‌شود. فایل‌های لاگ سرور وب، معمولاً به فرمت CLF[1]، شامل فیلدهایی همچون اطلاعات مشتری، زمان، جزئیات درخواست و غیره هستند. در وب‌کاوی، پیش‌پردازش داده‌ها لاگ‌های خام را به منظور نمایش رفتارهای کاربران به ساختاری جدولی تبدیل می‌کند. این پردازش بر روی تعاملات اصلی مانند کلیک‌ها بر صفحات وب تمرکز دارد و درخواست‌های ثانویه و محتواهای غیرمرتبط همچون تصاویر و ویدیوها را استثنا می‌کند. فرمت‌های URL استاندارد می‌شوند، پارامترهای تکراری حذف می‌شوند و درخواست‌های خطا، فیلتر می‌شوند. شناسایی کاربر و تعیین جلسه با استفاده از روش‌هایی همچون آدرس‌های IP و زمان رویت صفحه انجام می‌شود. این فرایندها رفتارهای کاربر را به جلساتی تجمیع می‌کنند تا برای تحلیل‌های بیشتر، تشخیص الگو و توصیه‌های شخصی‌سازی شده قابل انجام شود. پیش‌پردازش داده از اهمیت بسزایی برخوردار است زیرا داده‌های وب را برای تحلیل معنادار و برداشت‌های قابل اجرا آماده می‌کند [۱۳-۱۴].

۳-۳- استخراج ویژگی جدید

تجزیه و تحلیل و استخراج داده‌ها از گام‌های حیاتی در وب کاوی داده و تحلیل آن است. در این زمینه، یک ویژگی نوآورانه معرفی می‌شود که میزان زمان صرف‌شده بر روی هر صفحه وب داخل مجموعه داده را اندازه‌گیری می‌کند. این ویژگی اهمیت چشمگیری در درک رفتار کاربر در دامنه وب و در ارائه توصیه‌های شخصی‌سازی شده با استفاده از داده‌های زمانی دارد. فرایند ایجاد این ویژگی شامل مراحل زیر است:

مرحله ۱- ساخت ماتریس IP-URL: فرآیند ایجاد ویژگی را با ساخت یک ماتریس آغاز می گردد که در آن ردیف‌ها نشان‌دهنده آدرس‌های IP منحصربه‌فرد و ستون‌ها نمایانگر URLهای متمایز در مجموعه داده هستند. این ماتریس به عنوان پایه‌ای برای محاسبه زمان صرف‌شده توسط کاربران بر روی هر صفحه وب از طریق آدرس‌های IP خاص عمل می‌کند که محاسبه زمان بازدید یک آدرس IP از یک URL خاص به شکل زیر است:

$$Visit\ Time_{(IP-URL)} = \sum \text{Time Spent on URL by IP} \quad (۱)$$

که $Visit\ Time_{(IP-URL)}$ نمایانگر زمان بازدید یک IP خاص از یک URL مشخص است.

مرحله ۲- محاسبه‌ی میانگین زمان بازدید برای هر URL: بعد از تعیین زمان‌های بازدید برای هر IP بر روی هر URL، گام بعدی شامل محاسبه‌ی میانگین زمان بازدید برای هر URL است. این کار با استفاده از فرمول زیر انجام می‌شود:

$$\text{Average Visit Time}_{\text{URL}} = \frac{\sum Visit\ Time_{(IP-URL)}}{Number\ of\ visit\ to\ URL} \quad (۲)$$

که $\text{Average Visit Time}_{\text{URL}}$ میانگین زمان بازدید است.

مرحله ۳- اعمال وزن‌دهی بر اساس مقادیر مشخص شده: بعد از به دست آوردن میانگین مدت زمان بازدید برای هر URL، این مقادیر با استفاده از عوامل وزن‌دار مشخص شده، تجمیع می‌شوند. این تجمیع به شکل زیر قابل نمایش است:

$$\begin{aligned}\text{Score}_{\text{URL}} &= W_1 \times \text{Average Visit Time}_{\text{URL}} \\ &+ W_2 \times \text{Average Visit Time}_{\text{URL}} \\ &+ W_3 \times \text{Average Visit Time}_{\text{URL}}\end{aligned} \quad (۳)$$

در اینجا، امتیاز ترکیبی برای یک URL خاص به عنوان $\mathbf{Score_{URL}}$ نمایش داده شده است که $\mathbf{W_1}$، $\mathbf{W_2}$ و $\mathbf{W_3}$ مقادیر وزن‌دهی مشخص شده هستند که به ترتیب برای هر کدام از وزنها مقادیر ۱، ۳ و ۵ را در نظر گرفته ایم. با پیروی از این مراحل و استفاده از فرمول‌های ریاضی برای ایجاد ویژگی‌ها، با موفقیت یک ویژگی جدید برای اندازه‌گیری زمان صرف شده در هر صفحه وب ایجاد می گردد. این ویژگی پتانسیل بسیاری را برای بهبود تجزیه و تحلیل داده‌ها دارد و امکان فراهم کردن توصیه‌های شخصی بر اساس زمان را فراهم می‌کند. در نتیجه، این رویکرد ساختاری در مهندسی ویژگی در داده‌کاوی وب، یک دارایی ارزشمند در درک رفتار کاربر است.

جدول (۱) : رکوردهای فایل لاگ

	تعداد
کل رکوردها	4,477,843
بعد از پیش پردازش	715,014
تعداد جلسات کاربر	21,916

[1] Common Log Format

جدول (۲) : محاسبه تعداد خوشه

تعداد خوشه ها (K)	معیارهای ارزیابی		
	Precision	*Recall*	*F-Measure*
20	0.702	0.402	0.511
80	0.721	0.419	0.53
120	0.732	0.431	0.543
200	0.728	0.430	0.541
300	0.730	0.425	0.537

۳-۴- شناسایی کاربران

شناسایی کاربر یک عنصر اساسی در سیستم‌های کاوش وب است که امکان تفکیک کاربران فردی بر اساس آدرس‌های IP آن‌ها را فراهم می‌کند. برای افزایش دقت این فرایند، یک عامل محلی مورد استفاده قرار می‌گیرد تا به چالش‌های احتمالی ناشی از چند کاربر با آدرس‌های IP یکسان در شبکه‌های محلی پاسخ دهد. با اجرای شناسایی کاربر، امکان دسته‌بندی دقیق‌تر کاربران از طریق تقسیم‌بندی جلسات به وجود می‌آید که معمولاً با تعریف جلسات به عنوان دنباله‌های درخواست‌های کاربر در یک بازه زمانی ۳۰ دقیقه‌ای انجام می‌شود. این رویکرد، تجزیه و تحلیل دقیق‌تری از رفتارها و ترجیحات کاربران را فراهم می‌کند، که منجر به تولید توصیه‌های شخصی، بهبود طراحی وب‌سایت و ارتقای تجربه کلی کاربر می‌شود. این رویکرد چندوجهی در شناسایی کاربر و تقسیم‌بندی جلسات به عنوان یک روش اساسی در کاوش وب، پایه‌ای برای تحلیل عمیق رفتار کاربر در محیط وب فراهم می‌کند.

۳-۵- کشف الگوی رفتاری کاربران

در این بخش، به بررسی مراحل مختلف کشف الگوهای رفتاری کاربران در سیستم‌های کاوش وب پرداخته می‌شود. این بخش شامل چهار زیربخش اساسی است: تقسیم داده، مقدار بهینه خوشه، خوشه‌بندی و ساخت پروفایل. در این بخش، به طور مختصر بر اهمیت و نقش هر یک از این زیربخش‌ها در فرایند کشف الگوهای رفتاری کاربران تأکید می‌شود. لازم به ذکر است که هر زیربخش با همکاری و هماهنگی با یکدیگر، منجر به کسب بینش‌های دقیق و تنزیل شده‌تری درباره الگوهای رفتاری کاربران می‌شود.

زیر بخش اول - تقسیم داده: در این بخش مجموعه دادگان به سه زیرمجموعه تقسیم شده است: مجموعه آموزش، مجموعه اعتبارسنجی و مجموعه آزمون با تقسیمی به نسبت ۲۰-۲۰-۶۰ درصد. هدف این تقسیم‌بندی سه‌گانه به این صورت است: اولاً، از داده‌های اعتبارسنجی برای تعیین مقدار بهینه برای پارامتر k که نمایانگر تعداد مطلوب خوشه‌ها برای تحلیل خوشه‌بندی می‌باشد، استفاده می‌گردد. ثانیاً، داده‌های آموزش برای آموزش مدل خوشه‌بندی مورد استفاده قرار می‌گیرد. در آخر، داده‌های آزمون برای ارزیابی عملکرد مدل آموزش دیده استفاده می‌شود. این رویکرد منطقی تقسیم داده اجازه می‌دهد تا به طور موثر مدل خوشه‌بندی را بهینه کرده و عملکرد آن را بر روی داده‌های ناشناخته ارزیابی کنیم که منجر به کشف الگوهای رفتاری قوی و دقیق می‌شود. این روش دقیق تقسیم داده اطمینان می‌دهد که هر زیرمجموعه درون فرآیند خوشه‌بندی، یک هدف مشخص را مشخص می‌کند و در نتیجه کارایی مدل را بهینه می‌کند و قابلیت اطمینان در شناسایی الگوهای رفتاری را افزایش می‌دهد.

زیر بخش دوم - مقدار بهینه خوشه: در زمینه تحلیل خوشه‌بندی، تعیین تعداد بهینه از خوشه‌ها یک گام حیاتی است. هدف این فرایند، یافتن تعداد مناسبی از خوشه‌ها است و اغلب از داده‌های اعتبارسنجی بهره می‌برد. در این مقاله، از مجموعه داده اعتبارسنجی برای تخمین تعداد بهینه از خوشه‌ها استفاده کرده و با مشخص کردن یک دامنه از ۲۰ تا ۳۰۰ برای هر یک از این سناریوهای خوشه‌بندی، معیارهای عملکرد از جمله دقت، فراخوانی و اندازه F همان طور که در جدول (۲) نشان داده شده، محاسبه انجام پذیرفته است. از طریق این ارزیابی دقیق، تعداد مطلوب خوشه ها به مقدار عدد ۱۲۰ شناسایی گردید. این روش دقیق، اطمینان حاصل می‌کند که فرآیند خوشه‌بندی از اطلاعات کافی برخوردار بوده و در زمینه داده‌ها کارایی دارد.

زیر بخش سوم - خوشه بندی: در تلاش برای یافتن بهترین روش خوشه‌بندی، دو تکنیک معروف k-means و k-medoids با استفاده از معیار شباهت Levenshtein مقایسه گردیده است. همانطور که پیش‌تر بیان گردید در شروع خوشه‌بندی، هدف شناسایی مقدار مطلوب برای تعداد خوشه بود و مشخص گردید که استفاده از ۱۲۰ خوشه بهترین عملکرد را دارد. این رویکرد در گروه‌بندی دنباله‌های مشابه URLها عالی بوده و سازماندهی روشنی از داده‌ها را ارائه داده است. پس از ارزیابی دقیق، مشخص گردید که روش k-medoids نسبت به k-means از نظر تنوع و کارایی خوشه‌بندی بهتر عمل کرده است. به همین دلیل، تصمیم بر آن شد که از تکنیک k-medoids به عنوان پایه برای تحلیل‌های بیشتر اشتفاده شود.

زیر بخش چهارم - ساخت پروفایل: در مرحله "پروفایل‌سازی"، هدف ایجاد پروفایل برای هر خوشه و در نهایت ایجاد پروفایل کاربری با شناسایی دنباله‌های در هر خوشه با استفاده از الگوریتم FP-growth می‌باشد که این دنباله‌های شناسایی شده به عنوان پروفایل‌های خوشه عمل می‌کنند. برای در نظر گرفتن تنوع در دنباله لینک‌ها در هر خوشه، به دنبال سناریویی هستیم که یک خوشه ممکن است بیش از یک پروفایل را به خود اختصاص دهد که این امر باعث دقت بیشتر در پیش‌بینی لینک بعدی در مرحله توصیه لینک می‌شود. در این مرحله، برای سه خوشه امکان تعیین پروفایل‌ها براساس دنباله‌های لینک موجود وجود نداشت. پس از یافتن پروفایل‌ها، وزنی برای هر پروفایل بر اساس ویژگی زمان بازدید استخراج شده محاسبه می‌شود. این وزن‌ها هنگام انتخاب URLها و ارائه توصیه‌ها برای لینک‌های بعدی استفاده خواهند شد. این رویکرد پروفایل‌سازی، با ترکیب کاوش دنباله با اطلاعات زمانی، دقت توصیه لینک‌ها در چارچوب کاوش استفاده از وب را افزایش می‌دهد.

۳-۶- پیش‌بینی صفحه بعدی کاربر

در این مرحله، از داده‌های موجود در مجموعه داده آزمون برای ارزیابی عملکرد مدل خود در پیش‌بینی صفحه وب بعدی که کاربر احتمالاً به آن سر خواهد زد، استفاده شده است. این رویکرد شامل استفاده از الگوریتم همسایگی‌های نزدیکترین است که ورودی‌های آن شامل URLهای ورودی و پروفایل‌های خوشه‌هایی است که در مرحله "پروفایل‌سازی" تولید شده‌اند. همانطور که پیش‌تر توضیح داده شد، این پروفایل‌ها با ویژگی‌های وزن‌دار مدت زمان بازدید برای هر صفحه وب غنی شده‌اند که این امکان را می‌دهد تا الگوهای مرور صفحات توسط کاربر را به صورت جامع‌تری بررسی شود. در این مرحله، از الگوریتم همسایگی‌های نزدیکترین برای انتخاب مناسب‌ترین پروفایل برای URLهای ورودی داده شده استفاده گردیده که این فرایند انتخاب عمدتاً بر اساس ویژگی وزنی مدت زمان بازدید انجام می‌شود. پروفایل خوشه با بیشترین وزن شباهت به عنوان پروفایلی که ارتباطش با جلسه کاربری فعلی را نشان می‌دهد، انتخاب می‌شود. پس از انتخاب، مدل خود را ارزیابی نموده که شامل معیارهای دقت، فراخوانی و اندازه F می‌باشد. هدف این فرایند، اطمینان از این است که مدل پیش‌بینی با الگوهای مرور صفحات توسط کاربر هماهنگ باشد و این باعث تسهیل تجربه وب به صورت شخصی‌سازی شده می‌شود. با بهینه‌سازی تعاملات کاربر، وب‌سایت‌ها می‌توانند درگیری و رضایت کاربران را افزایش دهند.

۴- آزمایشات و نتایج

در بخش اول، ابتدا به ارزیابی عملکرد مدل خوشه‌بندی پرداخته شده است. معیارهای ارزیابی از جمله دقت، فراخوانی و لندازه F، به عنوان نشانگرهای حیاتی برای اندازه‌گیری کیفیت خوشه‌های تولید شده توسط مدل عمل می‌کنند. همانطور که ذکر گردید از داده‌های اعتبارسنجی برای تعیین تعداد مناسب خوشه‌ها (K) برای موثر بودن آموزش مدل استفاده شده است. برای به اشتراک گذاشتن نتایج به طور جامع، یک خلاصه در جدول (۲) نشان داده شده که عملکرد مدل برای انواع مختلف خوشه‌بندی را نشان می‌دهد. این جدول شامل ستون‌های تعداد خوشه (K)، دقت، فراخوانی و اندازه F است که نتایج برای تعداد مختلف خوشه‌ها از ۲۰ تا ۳۰۰ ارائه شده تا پیکربندی بهینه را شناسایی کند. بر اساس نتایج آزمایشی که در جدول (۲) ارائه شده است، مشخص است که اندازه خوشه ۱۲۰ نسبت به پیکربندی‌های دیگر عملکرد بهتری داشته و بیشترین اندازه F را به دست آورده است که نشانگر بهترین نتایج خوشه‌بندی می‌باشد.

در خوشه‌بندی، از الگوریتم های k-means و k-medoids برای تقسیم داده‌ها به خوشه‌ها استفاده شد. سپس، عملکرد این دو روش را مقایسه نموده و مشاهده شد که روش k-medoids به طور مداوم از k-means عملکرد بهتری داشته است. برای نشان دادن این موضوع، نتایج در جدول (۳) نشان داده شده که ستون‌های آن به ترتیب نشان دهنده روش استفاده شده، دقت، فراخوانی و اندازه F هستند.

در مراحل قبلی، از داده‌های اعتبارسنجی برای تعیین تعداد بهینه خوشه‌ها و شناسایی الگوریتم خوشه‌بندی استفاده شد و سپس، از داده‌های آموزش برای آموزش و ارزیابی مدل استفاده گردید. در ادامه برای ارزیابی مدل پیش‌بینی برای صفحه وب بعدی، دو سناریو متمایز مورد بررسی قرار داده شده است. ابتدا، ویژگی نوآورانه‌ای به نام مدت زمان صرف شده در هر صفحه به پروفایل کاربری اضافه گردید که این امر ایجاد پروفایل‌های دقیق‌تر را ممکن ساخت و این پروفایل‌ها برای پیش‌بینی صفحه وب بعدی برای هر کاربر استفاده شدند. در دومین سناریو، پیش‌بینی‌ها بدون اضافه کردن این ویژگی جدید انجام گرفت. نتایج نشان داد که با اضافه کردن این ویژگی جدید، دقت پیش‌بینی به طور قابل توجهی افزایش می یابد که دقت مدل با استفاده از معیارهای ارزیابی استاندارد مانند دقت، فراخوانی و اندازه F به شکل جدول (۴) برای موارد 'پیش‌بینی با ویژگی جدید' و 'پیش‌بینی بدون ویژگی جدید' نمایش داده شده است.

۵- نتیجه گیری و کارهای آتی

این مقاله یک بررسی گسترده از مدل‌سازی رفتار کاربران و پیش‌بینی در حوزه استخراج داده وب ارائه می‌دهد. در این مقاله یک رویکرد نوآورانه معرفی شده است که مدت زمان جلسات کاربران را در نظر می‌گیرد و این امر باعث بهبود دقت در پیش‌بینی انتخاب‌های صفحات وب بعدی کاربران می‌شود. در ادامه با استفاده از تکنیک‌های خوشه‌بندی مانند k-means و k-medoids، کاربران به طور موثر در خوشه‌ها گروه‌بندی گردیده که منجر به بهبود دقت و بازخوانی مدل پیش‌بینی شده شده است. به‌طور خاص، در این مقاله اندازه خوشه ۱۲۰ را به عنوان مقدار مطلوب شناسایی کرده که به طور پیوسته از سایر گزینه‌ها بهتر عمل کرده است. آزمایشات و یافته‌ها به طور روشن افزایشات قابل توجهی در دقت پیش‌بینی را نشان می‌دهد که به دلیل اضافه کردن مدت زمان جلسه به عنوان ویژگی جدید در مدل به‌دست آمده است. با استفاده از معیارهای استاندارد مانند دقت، فراخوانی و اندازه F، عملکرد مدل را با دقت ارزیابی کرده و نتایج ارائه شده نشان می‌دهد که این ویژگی جدید چگونگی موثر بودن مدل پیش‌بینی را افزایش داده است. علاوه بر این، تحلیل مقایسه‌ای از خوشه‌بندی k-means و k-medoids نشان می‌دهد که عملکرد برتر k-medoids، کمک به پروفایل‌دهی دقیق‌تر به کاربر می‌کند. برای کارهای آتی، این تحقیق پیشنهاد می گردد که راه‌های جدیدتری برای مدل‌سازی و پیش‌بینی رفتار کاربران را بررسی گردد که بررسی ویژگی‌ها و الگوریتم‌های اضافی، می‌تواند دقت مدل را بیشتر افزایش دهد. علاوه بر آن، این مطالعه می‌تواند به دامنه‌های گسترده‌تری مانند تجارت الکترونیک، پیشنهاد محتوا و تبلیغات شخصی‌سازی شده گسترش یابد تا به تقاضای رو به رشد خدمات محور برای کاربران پاسخ دهد.

جدول (۳) : مقایسه K-MEANS و K-MEDOIDS

روش	معیار ارزیابی		
	Precision	***Recall***	***F-Measure***
k-means	0.732	0.431	0.543
k-medoids	0.812	0.691	0.747

جدول (۴) : مقایسه پیش‌بینی صفحه بعدی با ویژگی جدید

روش	معیار ارزیابی		
	Precision	***Recall***	***F-Measure***
پیش‌بینی با ویژگی جدید	0.89	0.92	0.90
پیش‌بینی بدون ویژگی جدید	0.78	0.86	0.82

مراجع

[1] L. WangBin, "Web Mining Research," in *5th International Conference on Computational Intelligence and Multimedia Applications*, 2003.

[2] F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh,, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal,* vol. 16, no. Issue 3, pp. 261-273, 2015.

[3] Zhang, Q., Lu, J. & Jin, Y, " Artificial intelligence in recommender systems," *Complex Intell,* vol. 7, no. 1, p. 439–457, 2021.

[4] V. P. a. M. P. F. Mansur, "A review on recommender systems," in *International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, India, 2017.

[5] I. T. Afolabi, "Semantic Web mining for Content-Based Online Shopping Recommender Systems," . *International Journal of Intelligent Information Technologies (IJIIT),* vol. 15, no. 4, pp. 41-56, 2019.

[6] "www.Ehadish.com," [Online].

[7] H. Hasija, "Recommender system with web usage mining based on fuzzy c means and neural networks," in *International Conference on Next Generation Computing Technologies*, Dehradun, India, 2015.

[8] Katarya, R., Verma, O.P, "An effective web page recommender system with fuzzy," *Multimed Tools Appl,* vol. 76, no. 20, p. 21481–21496, 2017.

[9] F. Darbandi Monfared, "A novel web page recommender using data automatic clustering," *SN Applied Sciences,* vol. 1, p. 1719, 2019.

[10] María N. Moreno, Saddys Segrera, Vivian F. López, María Dolores Muñoz, "Web mining based framework for solving usual problems," *Neurocomputing,* vol. 176, pp. 72-80, 2016.

[11] T. Bhattacharya, A. Jaiswal and V. Nagpal, "Web usage mining and text mining in the environment of web personalization for ontology development of recommender systems," in *5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, 2016.

[12] A. G. M. E. Badr Hssina, "Recommendation system using the k-nearest neighbors and singular value decomposition algorithms," *International Journal of Electrical and Computer Engineering (IJECE),* Vols. 11,No 6, pp. 5541-5548, 2021.

[13] Muhammad Waqar, Nadeem Majeed, Hassan Dawood, Ali Daud & Naif Radi, "An adaptive doctor-recommender system," *Behaviour & Information Technology,* vol. 38, no. 9, pp. 959-973, 2019.

[14] R. &. P. J. Wagh, "A Novel Web Page Recommender System for Anonymous Users Based on Clustering," *Asian Journal For Convergence In Technology (AJCT),* vol. 5, no. 1, 2019.

[15] Manikandan, R., Saravanan, V, "A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective," *Multimed Tools Appl,* vol. 79, p. 3807–3829, 2020.

[16] J. R. D. Atienza, R. M. (2022). A Deep Neural Network in a Web-based Career Track Recommender System for Lower Secondary Education. 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), , (pp. pp. 1-6). Ravet, India.

[17] Jayalakshmi, N. (January 2023). Taylor Horse Herd Optimized Deep Fuzzy clustering and Laplace based K-nearest neighbor for web page recommendation. Advances in Engineering Software, Volume 175.

[18] Imran Hossain, M. A. (2023). A Survey o f Recommender System Techniques and the E-commerce Domain. Computer Science , Information Retrieval.

COPYRIGHTS

©2024 by the authors. Published by the Islamic Azad University Shiraz Branch. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0)

https://creativecommons.org/licenses/by/4.0