

الأساليب الخوارزمية لتطوير نظم التوصية ومجالات تطبيقها في

فهارس المكتبات

The algorithmic methods for developing recommendation systems and their application areas in library catalogs*1 طلال ناظم الزهيري، ضحى عبد الأمير شبل²¹ جامعة المستنصرية- العراق² جامعة المستنصرية- العراق

تاريخ النشر: 2024/12/20

تاريخ القبول: 2024/06/16

تاريخ الاستلام: 2024/03/16

Abstract

Library catalogs remain one of the most critical services offered by libraries to their users. With advancements in technology and the shift from card-based systems to automated catalogs resembling internet search engines, integrating concepts and applications of recommendation systems has become essential. This study explores the philosophy and operational mechanisms of recommendation systems, focusing on various types of algorithms. Its aim is to apply these concepts to build and enhance online library catalogs that meet current and future user expectations. By analyzing three types of algorithms—collaborative, content-based, and hybrid filtering—the study found that each has its strengths and weaknesses. However, the hybrid approach proved to be the most effective, leveraging the strengths of both collaborative and content-based methods to create a more robust recommendation system. Ultimately, the choice of approach depends on the specific characteristics and requirements of the library catalog and user interactions.

Keywords: LibraryCatalogs; RecommendationSystems;
Recommendation System Algorithms; Collaborative
Recommendation; Content-Based Recommendation; Hybrid
Recommendation.

المستخلص

لا تزال فهراس المكتبات تعد من أهم الخدمات التي تقدمها المكتبات للمستخدمين، ومع التطورات التقنية وتحول الفهراس من البطاقية إلى الآلية التي أصبحت تحاكي محركات البحث على الإنترنت. كان لا بد من الاستفادة من مفاهيم وتطبيقات التوصيات ومحاولة استثمارها في بناء وتطوير فهراس المكتبات المتاحة عبر الإنترنت. لذا؛ تهدف هذه الدراسة إلى التعرف على فلسفة وآلية عمل نظم التوصيات وفقاً لمختلف أنواع الخوارزميات المعتمدة في إطار آلية عمل نظم التوصيات ومحاولة توظيفها في تطوير فهراس المكتبات بما يتناسب مع تطلعات المستخدمين الآنية والمستقبلية. ومن خلال تحليل الأنواع الثلاث من الخوارزميات توصلنا إلى أن لكل منها نقاط قوة وضعف. لكن من الناحية العملية، وجدنا أن النهج الهجين يعد اختياراً مناسباً كونه يحقق الاستفادة من نقاط القوة في كل من التصنيفية التعاونية والمستندة إلى المحتوى، مما يوفر نظام توصيات أكثر قوة. ومع ذلك، يعتمد الاختيار في النهاية على الخصائص والمتطلبات المحددة لفهرس المكتبة وتفاعلات المستخدم.

الكلمات المفتاحية: فهراس المكتبات؛ نظم التوصية؛ خوارزميات نظم التوصية؛ التوصية التعاونية؛ التوصية على أساس المحتوى؛ التوصية الهجينة.

المقدمة

لعمدود طويلة من الزمن أسهمت الفهراس البطاقية في المكتبات بدور فعال في تمكين

المستفيدين من التعرف على مقتنيات المكتبات من الكتب والوصول لمصادرها وفقا لمطالباتهم الشخصية. ومع التطور التكنولوجي في مجال النظم البرمجية، وتحول الفهارس البطاقية إلى فهارس آلية، لاتزال الغاية والأهداف هي ذاتها بفارق السرعة والشمولية وتعدد نقاط الإتاحة التي حققتها الفهارس الآلية مقارنة بالبطاقية التي كانت تقتصر على مفاتيح استرجاع مرتبطة بالمؤلف والعنوان والموضوع فقط. واليوم ومع التطور الكبير في مجال النظم ولغات البرمجة، فضلا عن التنامي الكبير في الموارد الرقمية المتاحة عبر الإنترنت، أصبحت رغبات المستفيدين أوسع مما كانت عليه سابقا إذ لم تعد عملية الوصول للمصادر التي يبحثون عنها هي المشكلة كما كانت. بل تجددت رغباتهم في الوصول إلى مصادر ذات أهمية بالنسبة لهم ولم يكونوا يعرفون عنها مسبقا. وأمام هذه التطلعات، تهدف هذه الدراسة إلى التعرف على فلسفة وآلية عمل نظم التوصيات ومجالات استثمارها في تطوير فهارس المكتبات بما يتناسب مع تطلعات المستفيدين الآنية والمستقبلية.

الأهمية

التعرف على الخوارزميات الخاصة بتطوير نظم التوصية يمكن أن يسهم في تمكين العاملين في المكتبات أنفسهم؛ أو من خلال التعاون مع محترفي البرمجة من اختيار الخوارزمية الأنسب لتحقيق أهدافهم في بناء فهارس قادرة على تقديم توصيات نافعة للمستفيدين تتضمن الإشارة إلى مصادر معلومات تقترب موضوعيا من خياراتهم البحثية الأولية. وهذا الأمر سوف يشكل أهمية كبيرة للمكتبات والمستفيدين، خاصة على مستوى الشمول والسرعة ودقة المعلومات. وفي نظرة مستقبلية قريبة يمكن أن تنتقل المكتبات في فهارسها إلى جيل جديد ينسجم مع التطورات التقنية الكبيرة في مجال الإتاحة الرقمية لمصادر المعلومات.

الدراسات السابقة

يستعرض كل من (Konstan & Riedl, 2012) الموضوعات المتعلقة بنظم التوصية بالتفصيل مع التأكيد على بعض التحديات التي واجهها المطورون في هذا المجال وطرق تجاوزها، مثل مشكلة الأبعاد المستقلة في التصنيف الأساسي لبيانات تفضيلات المستخدمين للمنتجات. أشار البحث أيضا إلى الطرق التي يمكن استخدامها للتغلب على هذه التحديات بتطوير خوارزميات جديدة تستند إلى التحليل العاملي المفرد وتطبيقها على بيانات المستخدمين.

كما يتحدث عن تحسين جودة التوصية من خلال تحليل تفاعل المستخدمين مع التوصيات، بحيث يعمل النظام على تحسين جودة التوصية باستمرار من خلال فهم متطلبات المستخدمين بشكل أفضل والاستجابة لها. ويتطرق أيضا إلى التحديات التي يتعرض لها نظام التوصية عندما يتعلق الأمر بمنتجات جديدة ومستخدمين جدد، وطرق تجاوزها من خلال دمج توصيات الأحداث المعرفية والمحتوى لمساعدة النظام على فهم المنتجات والمستخدمين الجدد بشكل أفضل.

يتوقف الباحثان عند موضوع مهم يتعلق بتقييم جودة نظام التوصية، حيث يشير إلى الحاجة إلى استخدام مقياس جديد يأخذ في الاعتبار الخصوصيات والجوانب الاجتماعية والتنظيمية، بالإضافة إلى الدقة التنبؤية للنظام. من خلال بعض النظائر التي يمكن استخدامها لتقييم جودة نظام التوصية، مثل مقاييس الاعتمادية، والكفاءة، والرضى ومقاييس القابلية للتطبيق.

وقدم كل من (Rajpurkar, Bhatt, Malhotra, & Rajpurkar, 2015) تطبيق مبسط لنظام توصيات الكتب الذي يتكون من خصائص تصفية المحتوى، والتصفية التعاونية، واستخراج قواعد الارتباط بتوفير توصيات بالكتب للمستخدمين بناء على ما

يهتمون به. وقد تم استخدام نموذج تعاوني لتحسين نتائج التوصيات بشكل دقيق. يبحث النظام في الكتب التي أعجب بها المستخدمون الآخرون الذين يشاركونه نفس الاهتمامات ثم يوصي باستخدامها. وقد أظهر النظام نتائج فعالة وفعالية في التوصيات وليس لديه مشكلات في الأداء بسبب استخدامه نموذج تعاوني. ويتوفر النظام بمرونة في تغيير أو استخدام تقنيات جديدة.

وبدرجة اقرب لموضوع البحث قد كل من (Ali & Ullah, 2016) طريقة لتطوير نظام توصية للكتب من خلال استخدام تقنيات تصفية المعلومات بهدف تقليل مشاكل تكس المعلومات والضغط الذهني عن طريق توفير نتائج ذات صلة لمساعدة المستخدمين في تلبية احتياجاتهم من المعلومات. اذ اشار الباحثان الى ان هناك محاولات عديدة لتطوير مختلف أنظمة التوصية في مجالات متنوعة، ولكن هناك نقصا في الأنظمة التي تم إنشاؤها بشكل خاص للكتب ويمكن أن توصي بالكتب بناء على محتواها. ولهذا الاسباب، قام الباحثان بتصميم نظام مزدوج لتوصية الكتب. يستخدم النظام المقترح محتوى الكتب، ونهج التصفية بين العناصر، واستخراج قواعد الارتباط لتوصية بكتب أكثر دقة وذات صلة تلي احتياجات قارئ الكتب. ولقد اكد الباحثان ومن خلال تجاربهم، أن دمج تقنيات التصفية القائمة على المحتوى وتلك القائمة على العلاقة بين العناصر مع خوارزميات استخراج قواعد الارتباط لتشكيل تقنية مزدوجة يعطي توصيات ممتازة. اذ تم التغلب أيضا على مشاكل بداية التشغيل وتحليل المحتوى المحدود وقلة تقييم المشكلات باستخدام هذا النهج.

وفي سياق متصل قدم كل من (Patil & Patil, 2019) طريقة إنشاء نظام توصية بالكتب عبر الإنترنت باستخدام التصفية التعاونية واستخراج قواعد الارتباط لمساعدة المستخدمين في العثور على المنتجات بناءً على تفضيلاتهم وتقييماتهم. من منطلق إن أنظمة التوصية تعمل على تحسين رضا العملاء والخدمة من خلال تقديم توصيات مخصصة. يتم استخدام هذه الأنظمة عبر منصات مختلفة مثل Netflix و Instagram

لمساعدة المستخدمين في العثور على المنتجات أو المحتوى الذي يهتمون به. يستخدم نظام توصية الكتب التوصية التعاونية واستخراج قواعد الارتباط للتوصية بالكتب، ومعالجة تحديات تناثر البيانات ومشكلة بداية التشغيل.

نظم التوصيات

مفهوم التوصيات على مستوى الممارسة يكاد يكون متجذرا في العلاقات التجارية، اذ غالبا ما يلجأ الباعة والتجار الى توجيه انظار الزبون الى بضاعة او منتج لم يكن ينوى شراءه اصلا او كان يفكر في شراء منتج اخر، وهنا تكمن الحرفية بين بائع واخر في القدرة على دراسة سلوك الزبون والتعرف على رغبته حتى تكون التوصية ذات جدوى وتحقق الغاية منها، في ذات السياق تعد هذه الممارسة واحدة من اهم الممارسات التسويقية في البيئتين الواقعية والافتراضية. على سبيل المثال في مراكز التسوق تجد ان هناك آلية في عرض البضائع والمنتجاج بطريقة تعتمد على العلاقة الاستهلاكية فيما بينها. اذ غالبا ما تجد (السكر قرب الشاي) وهكذا بالنسبة لبقية المنتجات التي تكون مترابطة على مستوى الاستهلاك او تكمل بعضها بعض. في المقابل تكون هذه الصورة اوضع في مواقع التسوق الالكترونية حيث يعمل نظام التوصيات [Recommends System] بطريقة الية بحيث تختلف التوصية باختلاف الزبون وتوجهاته.

وفي هذا الصدد يشير (Shah, Gaudani, & Balani, 2016) إلى أن نظام التوصية هو أحد المجالات الكبيرة جدا في مجال التنقيب عن البيانات. والذي يمكن تقسيمه إلى نهجين رئيسيين: التوصية الشخصية وغير الشخصية. كل نهج لديه إحصائيات وتقنيات تعلم آلي مختلفة. في التوصية غير الشخصية، يقدم النظام نفس التوصيات لجميع المستخدمين من خلال تقييم بيانات النظام بأكملها في وقت واحد بدلا من بيانات كل

مستخدم على حدة. يستخدم هذا النهج في موقع اليوتيوب على سبيل المثال، حيث يتم تقديم أشهر الفيديوهات بطريقة الجمع بين ذوق جميع المستخدمين عبر الوسائل احصائية كالوسط الحسابي والمتوسط وغيرها. في النهج الآخر، يظهر نظام التوصية الشخصي، الذي يأخذ في اعتباره تفضيلات أو اهتمامات كل مستخدم على حدة، وبالتالي يقدم توصيات أكثر فعالية للعناصر المعنية للمستخدم أو مجموعة من المستخدمين . وفي سياق متصل يذكر كل من (Sharma & Mann, 2013) إن نظام التوصية يعد نظاما ذكيا يقدم اقتراحات حول العناصر للمستخدمين وفقا لاهتماماتهم المحتملة. تستخدم هذه الأنظمة في تطبيقات عملية مثل مواقع التجارة الإلكترونية ومواقع بيع الكتب وعرض الأفلام والموسيقى، والتي تعمل على التوصية بفحص تفضيلات المستخدم وتقديم اقتراحات تناسب تلك التفضيلات، مما يعزز تجربة المستخدم ويتعامل مع حمل المعلومات الزائد.

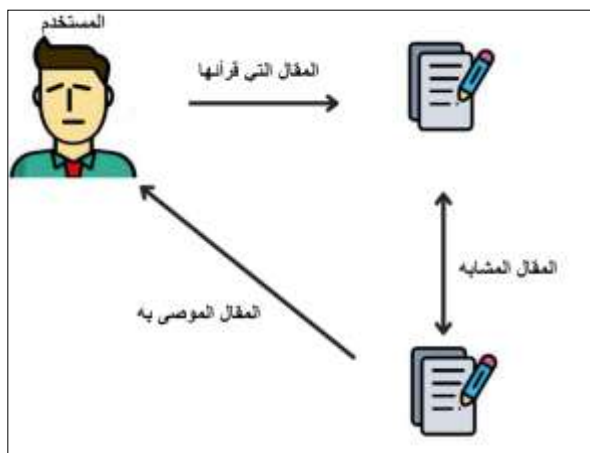
تقنيات نظم التوصية

لتحليل وفهم التطبيقات الرئيسية لنظام التوصية في المجالات المختلفة، نحاول هنا ان نستعرض تقنيات او طرق التصنيفية التي تعمل بها نظم التوصية بشكل عام والتي تقسم الى ثلاثة تقنيات اساسية وكالاتي :

اولا: التصنيفية القائمة على المحتوى

يشير كل من (Thorat, Goudar, & Barve, 2015) إلى أن تقنية التصنيفية القائمة على المحتوى [Content base filtering] ، تعمل من خلال تقديم التوصيات للمستخدمين اعتمادا على سلوك وبيانات مستخدم واحد فقط إذ يؤدي وصف العنصر وملف تعريف

المستخدم دورا مهما في نطاق عمل تقنية التوصية القائمة على المحتوى. ويضيف (Lu, Wu, & Mao, 2015) إلى أن المفهوم الأساسي لأنظمة التوصية القائمة على المحتوى يعتمد على مبدئين. الأول هو تحليل وصف العناصر التي يفضلها مستخدم واحد معين لتحديد التفضيلات التي يمكن استخدامها لوصف هذه العناصر. ويتم تخزين هذه التفضيلات في ملف تعريف المستخدم. ثانيا، هو مقارنة سمة كل عنصر مع ملف تعريف المستخدم بحيث تتم التوصية فقط بالعناصر ذات الصلة التي لها درجة عالية من التشابه مع ملف تعريف المستخدم لهذا الشخص المعين. وعليه يمكن القول ان مايميز نظام التوصية القائم على المحتوى عن أنظمة التوصية الأخرى هو أنه لا يحتاج حقًا إلى بيانات الأشخاص الآخرين. يتم تقديم جميع التوصيات بناءً على بيانات وتفضيلات المستخدم فقط. وكما ممثل في الشكل (1):



الشكل (1) مخطط التوصية على اساس المحتوى

في ضوء ما تقدم يمكن القول ان نظام التوصية القائم على المحتوى سوف يركز على وصف خصائص أو ميزات العناصر ذاتها لتحديد مدى تشابهها على اساس محتواها. بصرف النظر عن بيانات تفاعل المستخدم، اذ يمكن تقديم التوصيات استنادًا إلى التشابه بين العناصر من خلال تحليل المحتوى نفسه. على سبيل المثال، يمكن لنظام التوصية القائم على المحتوى أن يوصي بمقالات أخرى إذا كان لديه مواضيع أو كُتَّاب أو قصص مشابهة. ولغرض انشاء نظام توصيات قائم على المحتوى يجب اتخاذ الخطوات التالية :

1. جمع مجموعة بيانات تحتوي على تفاصيل حول العناصر وخصائصها. يتضمن تجهيز وتنظيف وتنسيق البيانات، مثل تحويل بيانات النص إلى تمثيلات رياضية وإزالة كلمات التوقف والرموز الخاصة.
2. استخدام مقياس تشابه مثل معامل ارتباط بيرسون، لتحديد مدى تشابه محتوى العناصر فيما بينها. وتكون مصفوفة التشابه النهائية هي نتيجة هذه العملية، وتظهر النقاط المتشابهة زوجيًا بين كل عنصرين في المجموعة.
3. إنشاء اقتراحات عن طريق تحديد العناصر الأكثر تشابهًا مع منتج محدد والتوصية بها للمستخدم. يمكن اختيار أفضل عدد من عناصر المتشابهة بعد فرز درجات التشابه.

بالنتيجة يمكن القول إن تطبيق هذا النهج في فهارس المكتبات الآلية سوف يحقق منفعة كبيرة للمستخدم حيث يمكنه من الوصول إلى كتب أو مقالات تقترب موضوعيا من الكتب الذي تم استرجاعه لها. وفي الوقت الذي ينظر إلى ايجابيات هذه الطريقة خاصة موضوع الاستقلالية ودرجة تطابق التوصية مع اهتمامات المستخدم نفسه. إلا أنها تعاني من عيوب أيضا على سبيل المثال إذا كان المحتوى لا يحتوي على معلومات كافية لتمييز العناصر بدقة، فلن تكون التوصية دقيقة في النهاية. كما يمكن القول إنها توفر درجة

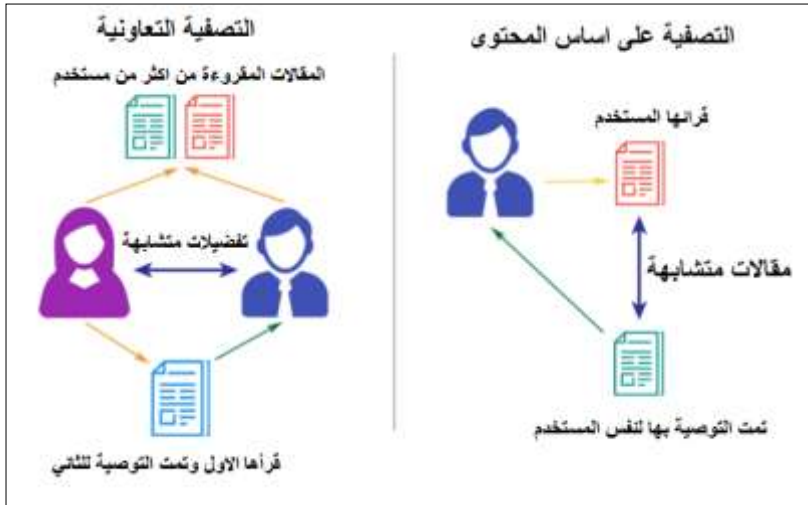
محدودة من الحداثة، حيث يجب أن تتوافق مع ميزات الملف الشخصي والعناصر الموصى بها بالتالي قد لا تشير عملية التصفية الكاملة المستندة إلى المحتوى إلى أي شيء جديد. ومن العيوب الأخرى أنها قد لا تقدم توصية بشكل صحيح إذا لم تكن هناك معلومات كافية لإنشاء ملف تعريف قوي للمستخدم.

ثانيا: التصفية التعاونية

يشير كل من (Schafer, Frankowski, & Herlocker, 2007) الى ان التصفية التعاونية [Collaborative filtering] هي عملية تصفية العناصر أو تقييمها من خلال آراء الأشخاص الآخرين. تجمع تقنية CF آراء المجتمعات الكبيرة المترابطة على الويب، مما يدعم تصفية كميات كبيرة من البيانات.

ويضيف كل من (Ekstrand, Riedl, & Konstan, 2011) إلى أن مجال المعلومات لنظام التصفية التعاوني يتكون من المستخدمين الذين أعربوا عن تفضيلاتهم لمختلف العناصر. هذا التفضيل يطلق على التأثير الذي يعبر عنه المستخدم لعنصر ما ويتم ذلك بشكل متكرر حتى يتم تمثيلها على شكل ثلاثية (المستخدم، العنصر، التقييم) ويمكن أن تأخذ هذه التقييمات أشكالاً عديدة، اعتماداً على النظام المعنى. بمعنى آخر أن نظام التوصية التعاونية سوف يعتمد على السلوك المباشر للمستخدم. أي أن هذا النظام يقوم ببناء نموذج للمستخدم بناء على الاختيارات والأنشطة والتفضيلات السابقة. ثم يستخدم هذه المعرفة للتنبؤ بما يفضلها المستخدم بناء على تشابهه مع ملفات تعريف المستخدمين الأخرى. ويمكن تصور الفرق بين نظام التصفية التعاونية مقارنة مع التصفية على أساس المحتوى بأن الأول يكون قائماً على اهتمامات وتفضيلات عدد من

المستخدمين في الوقت الذي يعتمد فيه نظام المحتوى على تفضيلات المستخدم نفسه وكما مبين في الشكل (2):



الشكل (2) الفرق بين التصفية التعاونية و التصفية على أساس المحتوى (1).

وفقا لهذا التصور يمكن أن يكون تطبيق التصفية التعاونية في فهارس المكتبات أقل نفعاً من التصفية على أساس المحتوى خاصة وأن تفضيلات المستخدمين الآخرين قد تتدخل فيها عوامل كثيرة تبعا للهدف من القراءة والتفضيل لهذا المستخدم عن ذاك. فضلا عن مخاطر ومخاوف تتعلق بالخصوصية. فبينما نتحرك نحو الحوسبة في كل مكان ويشارك الأفراد المزيد من المعلومات حول الأماكن والأشياء التي يمكنهم القيام بها ورؤيتها وشراؤها، تصبح مخاطر الخصوصية شديدة. في الوقت نفسه قد يجد البعض أن وضع خوارزمية

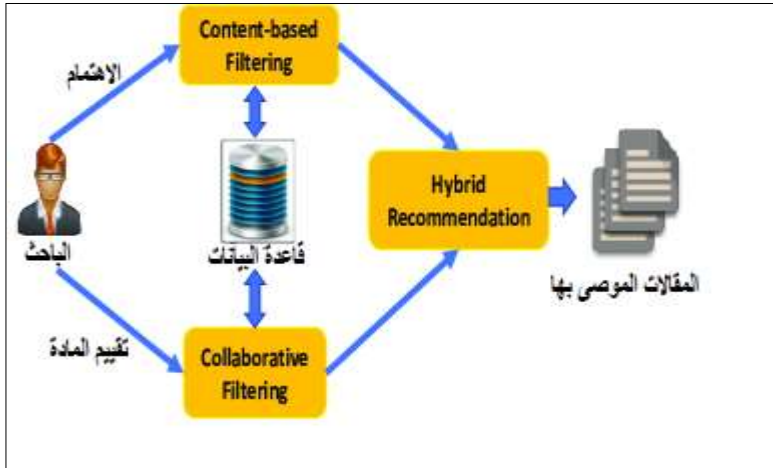
¹ . الشكل منقول بتصرف من <https://pritamaich.github.io/Books-Recommendation-System>

محكمة على أساس جمع ملفات تعريف المستخدمين على أساس التقارب في التخصص والاهتمام قد يحقق فائدة مهمة للمستخدمين. بالتالي يجد البعض أن التصفية التعاونية قد تكون أفضل للتطبيق في مواقع عرض الأفلام منها في مجال الكتب. وهو ما أثبتاه (Koren & Bell, 2021) في تحليل تطبيق هذا النوع من التصفية في موقع Netflix .

ثالثاً: التصفية الهجينة

نظام التوصية الهجين [Hybrid Recommender System] . تشترك أنظمة التوصية التعاونية وتلك القائمة على أساس المحتوى بنقاط قوة ونقاط ضعف في الوقت نفسه، لكنها تتكامل في نقاط القوى في حالة دمجها ليتم تجاوز نقاط الضعف في كل منها. لذلك تم تطوير نظام توصية هجين يعمل على التصفية التعاونية القائمة على التشابه بين تفضيلات المستخدمين المطبقة في التوصية التعاونية، وتفضيلات المستخدم نفسه التي يدعمها نظام التوصية على أساس المحتوى. إذ يتم الجمع بين عوامل التصفية القائمة على المحتوى والتعاونية، ويقوم كل منها بإنشاء ملف بقائمة التوصيات دون الجمع بينهما لتقديم التنبؤ المشترك للمستخدمين (Li & Kim, 2003).

وفي سياق متصل يشير (Verma, 2021) إلى أن الجمع بين التصفية التعاونية والمستندة إلى المحتوى معاً سوف يساعد في التغلب على أوجه القصور عند استخدامها بشكل منفصل، كما يمكن أن يكون أكثر فعالية في بعض الحالات. حيث يمكن تنفيذ أساليب نظام التوصية الهجين بطرق مختلفة مثل استخدام المحتوى والأساليب القائمة على التعاون لإنشاء تنبؤات بشكل منفصل ثم دمج التنبؤ، أو يمكننا فقط إضافة إمكانات الأساليب القائمة على التعاون إلى النهج القائم على المحتوى وبالعكس. وكما في الشكل (3)



الشكل (3) يمثل الية عمل نظام التوصية الهجين².

بشكل عام تؤدي أساليب التصفية دورا محوريا في فعالية أنظمة التوصية . سواء كان ذلك من خلال التصفية التعاونية أو التصفية القائمة على المحتوى أو الأساليب الهجينة. فإن الهدف هو تعزيز تجربة المستخدم من خلال تقديم اقتراحات مخصصة وذات صلة. تتمتع كل طريقة بنقاط قوتها وتحدياتها، وتعد معالجة مشكلات مثل مشكلة البداية الباردة أمراً بالغ الأهمية لضمان توصيات دقيقة، خاصة للمستخدمين أو العناصر الجديدة.

². مقتبس بتصرف من: <https://analyticsindiamag.com/a-guide-to-building-hybrid-recommendation-systems-for-beginners>

*1 المؤلف المرسل: طلال ناظم الزهيري ، الإيميل: drazzuhaire@gmail.com

مراحل تطوير فهرس المكتبات

لا شك أن فهرس المكتبات تعد من أفضل الخيارات لتطبيق نظم التوصية خاصة ونحن أمام كم هائل من الموارد الرقمية التي أصبح من الصعوبة في مكان التعرف عليها وحصرها أو المفاضلة فيما بينها. بالتالي فإن تطوير فهرس تعمل بنظام التوصية سوف يساعد المستفيدين في الحصول على المعلومات بأسرع وأدق الطرق. فضلا عن أهميتها في توجيه أنظار المستخدمين نحو العلاقات الموضوعية بين الكتب فضلا عن تمكين العاملين في المكتبات من اختيار أفضل الكتب لأغراض الاقتناء بناء على تقييمات المستخدمين. وسوف نحاول هنا أن نستعرض طريقة بناء الخوارزميات بالطرق الثلاث التي تطرقنا إليها :

أولا: خوارزمية التصفية على اساس المحتوى

تتضمن التصفية على أساس المحتوى لفهرس المكتبة التوصية بالعناصر بناء على ميزاتها وسماتها. في هذه الحالة، يوصى بالكتب للمستخدمين بناء على الخصائص المتعلقة بالمحتوى مثل العنوان والموضوعات والمؤلفين والملخصات. وفيما يلي خوارزمية بسيطة للتصفية على اساس المحتوى تصلح للتطبيق في فهرس المكتبة.

• مدخلات

- معرف بيانات المستخدم user_id
- بيانات الكتاب مع السمات: book_id، العنوان، الموضوع، المؤلفين، الملخص، إلخ.
- مخرجات:

- الكتب الموصى بها للمستخدم على وفق التصفية على أساس المحتوى.

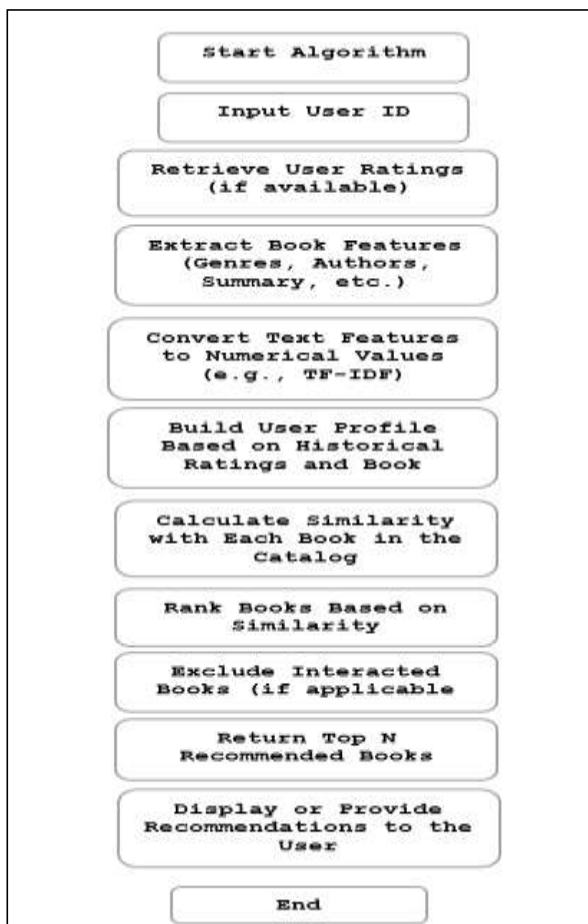
1. أدخل تفضيلات المستخدم أو اهتماماته (على سبيل المثال، user_id).

2. استرجاع تقييمات الكتب التاريخية للمستخدم، إذا كانت متوفرة.
3. استخراج الميزات من بيانات الكتاب.
4. تحويل ميزات النص (الموضوعات، المؤلفون، الملخص) إلى تمثيلات رقمية (على سبيل المثال، باستخدام TF-IDF³).
5. إنشاء ملف تعريف مستخدم بناءً على التقييمات السابقة للمستخدم وميزات الكتب التي تفاعل معها.
6. حساب التشابه بين ملف تعريف المستخدم وكل كتاب في الفهرس:
7. استخدم مقياس التشابه مثل تشابه جيب التمام.
8. ترتيب الكتب بناءً على درجات التشابه مع ملف تعريف المستخدم.
9. استبعاد الكتب التي سبق للمستخدم التفاعل معها (إن أمكن).
10. إرجاع أهم الكتب الموصى بها إلى المستخدم.
11. عرض أو توفير الكتب الموصى بها للمستخدم.

النهاية...

ويمكن تمثيل تلك الخوارزمية في المخطط الآتي :

³. مصطلح تردد الوثيقة العكسي] هو مقياس إحصائي يقوم بتقييم مدى صلة الكلمة بمستند في مجموعة من المستندات. ويتم ذلك عن طريق ضرب مقياسين: عدد المرات التي تظهر فيها الكلمة في المستند، والتكرار العكسي للكلمة عبر مجموعة من المستندات.



الشكل (4) مخطط خوارزمية نظام التوصية فهرس المكتبة على اساس المحتوى

تتضمن خوارزميات التوصية القائمة على المحتوى إنشاء ملفات تعريف المستخدمين

وتمثيلات العناصر بناءً على ميزات المحتوى الخاصة بها. هنا سوف نعبر عن خوارزمية التوصية الأساسية القائمة على المحتوى باستخدام التدوين الرياضي وكالاتي:

1. مصفوفة عنصر المستخدم:
- (U) مصفوفة عنصر المستخدم حيث تمثل (u_{ij}) تفاعل المستخدم (i) مع العنصر (j)
2. مصفوفة ميزات العنصر:
- (i) مصفوفة ميزات العنصر حيث يمثل (i_{jk}) قيمة الميزة (k) للعنصر (j).
3. مصفوفة ملف تعريف المستخدم:
- (P) مصفوفة ملف تعريف المستخدم حيث يمثل (P_{ik}) تفضيل المستخدم (i) للميزة (k).
4. التنبؤ:
- يتم حساب التفاعل المتوقع (u_{ij}) كمنتج نقطي لملف تعريف المستخدم وميزات العنصر.
5. التحسين:

$$\hat{u}_{ij} = \sum_{k=1}^K P_{ik} \cdot I_{jk}$$

تم تحسين مصفوفة ملف تعريف

المستخدم (P) بناءً على التفاعلات التاريخية لتقليل الفرق بين التفاعلات المتوقعة والفعالية. تتضمن خوارزمية التوصية القائمة على المحتوى تمثيل العناصر والمستخدمين من حيث الميزات. ويتم التنبؤ عن طريق حساب التشابه بين ملف تعريف المستخدم وميزات العنصر. يتم إجراء التحسين للتعرف على مصفوفة ملف تعريف المستخدم بناءً على التفاعلات التاريخية.

ثانيا: خوارزمية التوصية التعاونية

تتضمن التوصية التعاونية تقديم توصيات بناء على التقييمات المشتركة لمجموعة من المستفيدين وعليه فان تطوير فهرس المكتبة الذي يعمل بنظام التوصية التعاونية يتطلب وضع خوارزمية متكاملة وكالاتي:

مدخل :

1. بيانات تفاعل عنصر المستخدم (user_id ، book_id، التقييمات)

2. عدد أقرب الجيران (K)

المخرجات

نموذج التوصية التعاونية

1. أدخل بيانات تفاعل عنصر المستخدم، والتي تتكون من معرفات المستخدمين،

ومعرفات الكتب، والتقييمات.

2. قم بإنشاء مصفوفة عنصر المستخدم حيث تمثل الصفوف المستخدمين،

وتمثل الأعمدة الكتب، وتمثل الإدخالات التقييمات. التعامل مع القيم المفقودة

بشكل مناسب.

3. تنفيذ التوصية التعاونية :

أ. لكل مستخدم، احسب التشابه مع جميع المستخدمين الآخرين باستخدام

مقياس التشابه (على سبيل المثال، تشابه جيب التمام).

ب. حدد أفضل اقرب الجيران لكل مستخدم بناءً على درجات التشابه

4. توقع تقييمات الكتب التي لم يتفاعل معها المستخدم :

أ. لكل كتاب لم يتفاعل معه المستخدم، قم بحساب المتوسط المرجح للتقييمات من أقرب الجيران الذين تفاعلوا مع هذا الكتاب. استخدم درجات التشابه كأوزان.

5. توليد التوصيات :

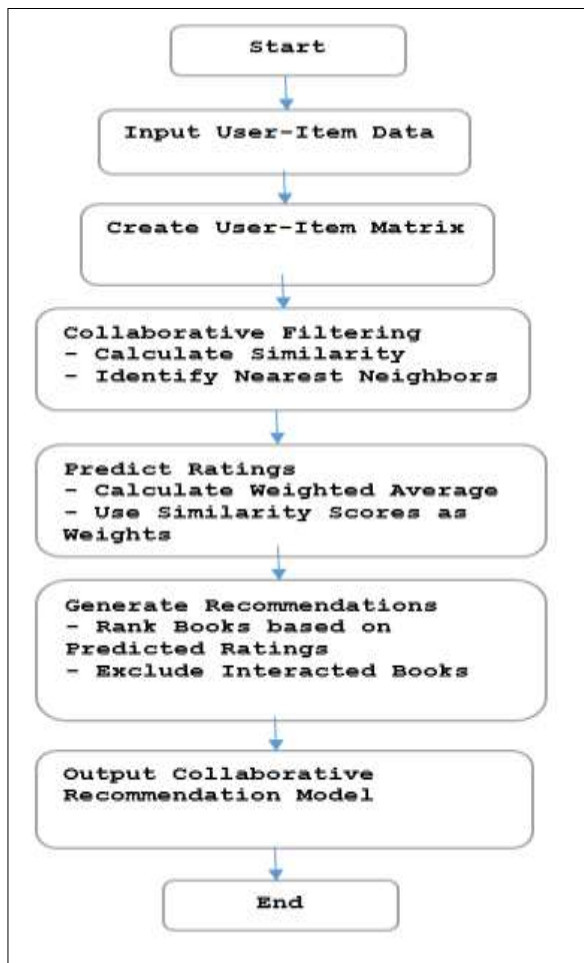
أ. رتب الكتب بناءً على التقييمات المتوقعة للمستخدم.

ب. استبعاد الكتب التي تفاعل معها المستخدم بالفعل.

6. إخراج نموذج التوصية التعاونية، بما في ذلك درجات التشابه والكتب الموصى بها لكل مستخدم.

النهاية...

ولتمثيل هذه الخوارزمية كمخطط سوف تكون بالشكل الآتي:



الشكل (5) مخطط خوارزمية نظام التصفية التعاونية لفهرس المكتبة

غالبًا ما يتم التعبير عن خوارزميات التصفية التعاونية رياضياً في سياق تحليل

$$R \text{ where } r_{ij}$$

المصفوفات. نفترض مصفوفة عنصر المستخدم التقييم الذي يمنحه المستخدم (i) للعنصر (j).

الهدف هو تقريب هذه المصفوفة عن طريق تحليلها إلى مصفوفتين ذات رتبة أقل (U) و (V) بحيث يكون $R \approx UV^T$.

وفيما يلي تمثيل رياضي مبسط لخوارزمية التصفية التعاونية باستخدام تحليل المصفوفة:

1. مصفوفة عنصر المستخدم:

- (R): مصفوفة تقييمات عنصر المستخدم، حيث (r_{ij}) هو التقييم الذي قدمه المستخدم (i) للعنصر (j).

2. تحليل المصفوفة:

- (U): مصفوفة المستخدم بالحجم $(m \times k)$ ، حيث (u_{ik}) تمثل تقارب المستخدم (i) مع العامل الكامن (k).

- (V): مصفوفة العنصر بالحجم $(n \times k)$ ، حيث (v_{jk}) تمثل أهمية العامل الكامن (k) للعنصر (j).

3. التنبؤ:

- يتم حساب التصنيف المتوقع $(\hat{r}_{ij})$ بواسطة المنتج النقطي للمستخدم المقابل ومتجهات العنصر:

$$\hat{r}_{ij} = \sum_{k=1}^K u_{ik} \cdot v_{jk}$$

4. التحسين:

تم تحسين المصفوفتين (U) و (V) لتقليل الفرق بين (R) و (UV^T) باستخدام تقنيات مثل نزول التدرج العشوائي (SGD) أو المربعات الصغرى بالتناوب (ALS).

باختصار، تهدف خوارزمية التصفية التعاونية إلى تحليل مصفوفة عنصر المستخدم إلى مصفوفتين أقل رتبة، والتقاط العوامل الكامنة التي تمثل تفضيلات

المستخدم وخصائص العنصر. ثم يتم حساب التصنيفات المتوقعة باستخدام المصفوفات المستفاد. يتم إجراء التحسين لتقليل الفرق بين التقييمات المتوقعة والفعالية.

ثالثا: خوارزمية التصفية الهجينة

يجمع نظام التوصية الهجين بين أساليب التوصيات المتعددة، مثل التصفية التعاونية، أو التصفية المستندة إلى المحتوى، أو تقنيات أخرى، لتعزيز دقة التوصيات وتغطيتها. ومن خلال الاستفادة من نقاط القوة في الأساليب المختلفة، لذا فإن تطوير نظام لفهرس كتب المكتبة اعتمادا على هذا النهج غالبا ما يهدف إلى التغلب على القيود الفردية وتقديم اقتراحات أكثر تخصيصا وتنوعا للمستخدمين. وفيما يأتي خوارزمية الفهرس وفقا لهذه التوصية:

مدخل - :

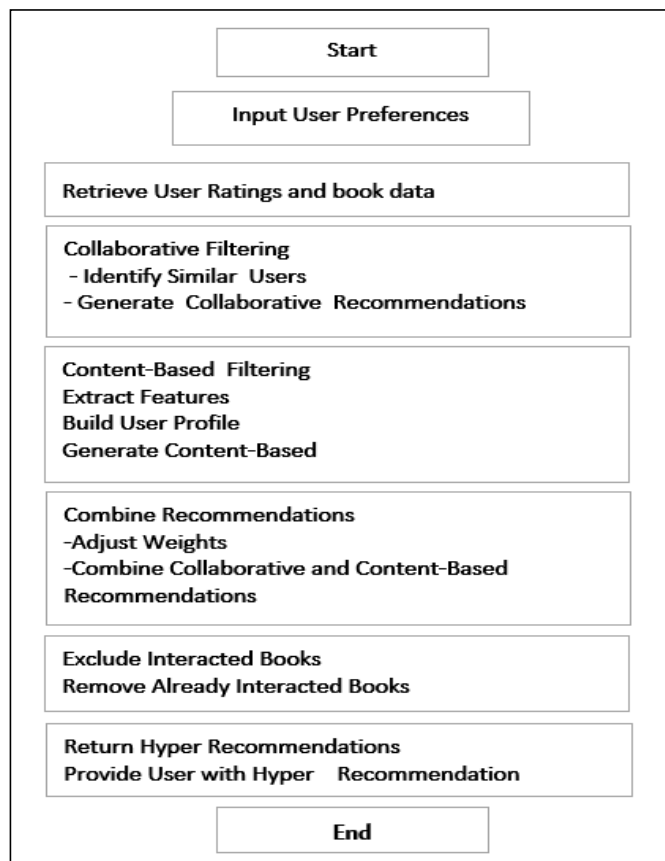
1. معرف المستخدم.
2. بيانات تفاعل عنصر المستخدم (بيانات المستخدم، بيانات الكتب، التقييمات).
3. بيانات الكتاب مع السمات (بيانات الكتب، العنوان، الموضوع، المؤلفين، الملخص، وما إلى ذلك)

مخرجات

التوصية الهجينة للمستخدم

1. أدخل تفضيلات المستخدم أو اهتماماته (على سبيل المثال، user_id)
2. إسترجاع تقييمات الكتب السابقة للمستخدم، إذا كانت متوفرة.
3. التصفية التعاونية:

- أ. استخدم التصفية التعاونية لتحديد المستخدمين ذوي الأذواق المماثلة.
 - ب. قم بإنشاء توصيات تعاونية بناءً على تفضيلات المستخدمين المماثلين
 4. التصفية على أساس المحتوى :
 - أ. استخراج الميزات من بيانات الكتاب (مثل الأنواع والمؤلفين والملخص).
 - ب. أنشئ ملفًا شخصيًا للمستخدم بناءً على التقييمات السابقة للمستخدم وميزات الكتب التي تفاعل معها.
 - ت. استخدم التصفية المستندة إلى المحتوى للتوصية بالكتب المشابهة لتلك الموجودة في الملف الشخصي للمستخدم
 5. الجمع بين التوصيات :
 - أ. اجمع بين التوصيات التعاونية والقائمة على المحتوى باستخدام نهج مرجح أو أسلوب دمج آخر
 - ب. اضبط الأوزان بناءً على أداء أو خصائص كل نهج توصية
 6. استبعاد الكتب المتفاعلة بالفعل :
 - أ. قم بإزالة الكتب التي تفاعل معها المستخدم بالفعل من قائمة التوصيات المجمعة
 7. قم بإرجاع التوصية الهجينة
 - أ. تزويد المستخدم بقائمة التوصيات النهائية .
- ويمكن تمثيل تلك الخوارزمية في المخطط الآتي :



الشكل (6) مخطط خوارزمية نظام التوصية الهجينة لفهرس المكتبة

تجمع خوارزمية التوصية الهجينة بين التنبؤات من التوصية التعاونية والتوصية القائمة على المحتوى باستخدام المتوسطات المرجحة. يمكن تعديل الأوزان لإعطاء الأفضلية لأسلوب واحد على الآخر بناءً على متطلبات التطبيق أو تفضيلات المستخدم. ويمكن التعبير عنها رياضياً بالآتي :

1. مصفوفة عنصر المستخدم :

(R) مصفوفة عنصر المستخدم حيث (r_{ij}) هو التصنيف الذي قدمه المستخدم (i) للعنصر (j).

2. التصفية التعاونية:

التقييم المتوقع للمستخدم (i) على العنصر (j) باستخدام التصفية التعاونية. $\hat{r}_{ij}^{CF}$

3. التصفية على اساس المحتوى:

التقييم المتوقع للمستخدم (i) على العنصر (j) باستخدام التصفية المستندة الى المحتوى. $\hat{r}_{ij}^{CBF}$

4. التنبؤ الهجين:

التنبؤ المشترك باستخدام المتوسط المرجح أو تقنية الدمج الأخرى. $\hat{r}_{ij}^{Hybrid}$

$$\hat{r}_{ij}^{Hybrid} = w_{CF} \cdot \hat{r}_{ij}^{CF} + w_{CBF} \cdot \hat{r}_{ij}^{CBF}$$

يتم ضبط الأوزان w_{CF} و w_{CBF} بناءً على الأداء أو الخصائص المطلوبة.

المناقشة

يعتمد الاختيار بين التصفية التعاونية أو التصفية القائمة على المحتوى أو النهج الهجين على عوامل مختلفة، ولكل منها نقاط القوة والضعف الخاصة به. ويمكن الاختيار ما بين الأساليب الثلاثة في سياق بناء نظام فهرس لكتب المكتبة يعمل بنظام التوصية وفقاً لطبيعة وأهداف فهرس المكتبة وحجم وتنوع الكتب فضلاً عن حجم وتنوع المستخدمين منه. وفيما يلي عرض لجوانب القوة والضعف لكل من مناهج التصفية.

فالتصفية التعاونية تتميز بقدرتها على جمع تفضيلات المستخدم بشكل فعال دون الاعتماد على ميزات العنصر الواضحة. كما يمكن ان تقدم توصيات متوافقة بناء على سلوك المستخدم. في الجانب السلبي تعاني التصفية التعاونية من مشكلة البداية الباردة للعناصر الجديدة أو المستخدمين ذوي سجل التفاعل المحدود. وقد تواجه مشكلات تشتت البيانات التي تكون تفاعلاتها قليلة. مع هذا تعد هذه التصفية فعالة عندما تكون هناك بيانات كافية لتفاعل المستخدم ولا تشكل مشكلة البدايات الباردة مشكلة صعبة.

اما التصفية على اساس المحتوى فتتميز بقدرتها على تجاوز مشكلة البداية الباردة من خلال الاعتماد على ميزات العنصر. فضلا عن امكانية تقديم توصيات مخصصة بناء على تفضيلات المستخدم الصريحة. لكنها تعاني من ضعف في جانب التوصيات الملائمة كونها تعتمد بشكل كبير على جودة وأهمية ميزات العنصر فقط. مع هذا تعد ملائمة لبناء فهرس المكتبات عندما تكون هناك ميزات عناصر غنية (مثل الموضوعات والمؤلفين والملخصات).

في ما يخص نهج التصفية الهجينة التي تجمع بين نقاط القوة في كل من الأساليب التعاونية والقائمة على المحتوى. الامر الذي يساهم في قدرتها على معالجة القيود المفروضة على الحالات الفردية، بالتالي يمكن ان تقدم توصيات أكثر قوة ودقة. الا ان التعقيد في إدارة وتحسين نماذج التوصيات الهجينة التي تتطلب ضبطا دقيقا للأوزان للتنبؤات المجمعة. يمكن ان تعد جوانب سلبية ترافق تطبيقها. لكنها غالبا ما تكون فعالة عندما تكون هناك رغبة في تحسين دقة التوصية وتغطية ومرونة.

الاستنتاجات

في ضوء عوامل القوة والضعف المرافقة للتوصيات الثلاث يمكن أن نستنتج أن اختيار أي من التوصيات لبناء فهارس المكتبات سوف يعتمد على:

1. مدى توفر البيانات: تتطلب التصفية التعاونية تفاعلات بين المستخدم والعنصر وتعتمد التصفية المستندة إلى المحتوى على ميزات العنصر ويحتاج النهج الهجين إلى كليهما. النظر في مدى توفر وجودة هذه البيانات يعد عاملا مهما في اتخاذ القرار المناسب.
2. تفضيلات المستخدم: تعتبر التصفية التعاونية مفيدة لالتقاط تفضيلات المستخدم بناء على السلوكيات بينما تعد التصفية المستندة إلى المحتوى أفضل للتفضيلات الصريحة.
3. البداية الباردة: يعد نظام التوصية المستند إلى المحتوى فعالا في معالجة مشكلات البداية الباردة بينما قد تواجه التصفية التعاونية صعوبات في مثل هذه السيناريوهات.
4. تنوع التوصيات: غالبا ما توفر الأساليب الهجينة توازنا جيدا بين الملاءمة والتوصيات المخصصة.

من الناحية العملية غالبا ما يفضل النهج الهجين لأنه يمكنه الاستفادة من نقاط القوة في كل من التصفية التعاونية والمستندة إلى المحتوى، مما يوفر نظام توصيات أكثر قوة. ومع ذلك، يعتمد الاختيار في النهاية على الخصائص والمتطلبات المحددة لفهرس المكتبة وتفاعلات المستخدم.

References

1. Verma , Y. (2021, 3 5). *A Guide to Building Hybrid Recommendation Systems for Beginners*. Retrieved from analytics indiamag: <https://analyticsindiamag.com/a-guide-to-building-hybrid-recommendation-systems-for-beginners/>
2. Ali, Z. K., & Ullah, I. (2016). A hybrid book recommender system based on table of contents (toc) and association rule mining. In *Proceedings of the 10th International Conference on Informatics and Systems*, (pp. 68-74).
3. Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*,, 81-173. Retrieved 3 2, 2024, from file:///C:/Users/Administrator/Downloads/9781601984432-summary-1.pdf
4. Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms. *User Model User-Adap Inter*, 23-101. doi:DOI 10.1007/s11257-011-9112-x
5. Koren, Y., & Bell, R. (2021). Advances in Collaborative Filtering. In *Recommender systems handbook*. Retrieved 3 2, 2024, from <https://datajobs.com/data-science-repo/Collaborative-Filtering-%5BKoren-and-Bell%5D.pdf>
6. Li, Q., & Kim, B. M. (2003). Clustering Approach for Hybrid Recommender System. *Proceedings IEEE/WIC International Conference on Web Intelligence* (pp. 33-38). IEEE. Retrieved from <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.453.5105&rep=rep1&type=pdf>
7. Lu, J., Wu, D., & Mao, M. (2015). Recommender system application developments: a survey. *Decision support systems*, 12-32.
8. Patil, A. E., & Patil, S. S. (2019). Online book recommendation system using association rule mining and collaborative filtering. *International Journal of Computer Science and Mobile Computing*, 8(4), 83-87.

9. Rajpurkar, S., Bhatt, D., Malhotra, P., & Rajpurkar, P. (2015). Book recommendation system. *International Journal for Innovative Research in Science & Technology*, 314-316.
10. Schafer, J. B., Frankowski, D., & Herlocker, J. (2007). Collaborative Filtering Recommender Systems. *The Adaptive Web*, 291 – 324. Retrieved 3 2, 2024, from https://www.researchgate.net/profile/Shilad-Sen/publication/200121027_Collaborative_Filtering_Recommender_Systems/links/547691850cf2778985b08077/Collaborative-Filtering-Recommender-Systems.pdf
11. Shah, L., Gaudani, H., & Balani, P. (2016). Survey on Recommendation System. *International Journal of Computer Applications*, 0975 – 8887. Retrieved 3 1, 2024, from https://www.researchgate.net/profile/Hetal-Gaudani/publication/298801724_Survey_on_Recommendation_System/links/5732fe1f08ae9f741b25ba01/Survey-on-Recommendation-System.pdf
12. Sharma, M., & Mann, S. (2013). A Survey of Recommender Systems: Approaches and Limitations. *International Journal of Innovations in Engineering and Technology*, 2(2), 8-14. Retrieved 3 1, 2024, from <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=fa41dc4b60eccedf1c41e2ae488044827dd79384>
13. Thorat, P. B., Goudar, R. M., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 0975 – 8887.