APJ 7(2):51-61, Summer 2026

Arman Process Journal (APJ)

Arman Process Journal (APJ)

Homepage: https://www.armanprocessjournal.ir

# An Intelligent Link Recommendation Approach in Virtual Social Networks

**M. Baghaei** [1*]

[1] Department of Computer Engineering, Mara.C., Islamic Azad University, Marand, Iran

**RESEARCH PAPER**

Received: 2026-3-27
Accepted: 2026-8-14

**KEYWORDS:**
Virtual Social Networks,
Link Prediction,
Similarity Measure,
Machine Learning

Corresponding author:
mo.baghaei@iau.ac.ir

## ABSTRACT

The expansion of online social networks has attracted a vast number of users. Link prediction is a crucial aspect of social network analysis; its objective is to identify potential connections within a network that are not currently present or observed. In social networks, there is a potential for links to exist between entities—links that may emerge in the future or were simply missing due to data loss. While link prediction is not a form of social network analysis in itself, it is a vital component of the analytical process. Furthermore, recommending interesting friends—in a way that pleasantly surprises new users—has become a major challenge for social networks. This challenge is particularly significant due to both the massive volume of data and the lack of historical activity data for new users upon joining. This paper aims to design a hybrid recommendation system—leveraging novel knowledge and ensemble learning—to facilitate friend recommendations and enhance the accuracy and precision of these suggestions within social networks.

Copyright © Author(s).

نشریه تخصصی آرمان پردازش، دوره ۷، شماره ۲، تابستان ۱۴۰۵

فصلنامه تخصصی آرمان پردازش
(APJ)
Homepage: www.armanprocessjournal.ir

فصلنامه تخصصی فناوری اطلاعات و ارتباطات
شماره مجوز: ۸۷۰۹۰

# رویکرد هوشمند پیش‌بینی و پیشنهاد پیوند در شبکه های اجتماعی مجازی

**محمدعلی بقائی*[۱]**

[۱] گروه مهندسی کامپیوتر، واحد مرند، دانشگاه آزاد اسلامی، مرند، ایران

## چکیده

**مقاله پژوهشی**

با گسترش فضای اجتماعی شبکه‌های اجتماعی بر خط کاربران زیادی را به‌سوی خود جلب کرده است. یکی از مهم‌ترین جنبه‌های تحلیل شبکه‌های اجتماعی پیش بینی پیوند است. هدف از پیش بینی پیوند، یافتن پیوندهای است که در شبکه مورد مطالعه در دسترس نیست و هنوز مشاهده نشده است در یک شبکه اجتماعی، امکان وجود پیوند بین بسیاری از موجودیت‌ها است که در حال حاضر این پیوندها در دسترس نیستند و ممکن است این پیوندها در آینده به وجود بیاید و یا اینکه بر حسب اتفاقاتی اطلاعات از دست رفته باشد پیش بینی پیوند به خودی خود یک تحلیل شبکه اجتماعی نیست اما از مهم‌ترین کارهایی که در جهت تحلیل شبکه نیاز است تا انجام شود. از سوی دیگر معرفی دوستان جذاب، بطوریکه برای کاربر جدیدالورود تعجب برانگیز باشد امروزه تبدیل به یک چالش اساسی برای شبکه‌های اجتماعی شده است. این چالش هم بخاطر وجود حجم عظیم اطلاعات و هم بخاطر عدم وجود عملکرد قبلی کاربر جدیدالورود در ابتدای عضویت بسیار حائز اهمیت است. این مقاله درصدد طراحی یک سیستم پیشنهاد دهنده هوشمند ترکیبی مبتنی بر دانش جدید و سیستم یادگیری جمعی جهت معرفی دوست در شبکه‌های اجتماعی و بهبود دقت و صحت پیشنهادات دوست در شبکه اجتماعی با استفاده از سیستم یادگیری تجمعی می‌باشد.

**واژگان کلیدی:**
شبکه‌های اجتماعی مجازی،
پیش بینی لینک،
معیار شباهت،
یادگیری خودکار

## ۱- مقدمه

معرفی دوستان جذاب، بطوریکه برای کاربر جدیدالورود تعجب برانگیز باشد امروزه تبدیل به یک چالش اساسی برای شبکه‌های اجتماعی شده است. این چالش هم بخاطر وجود حجم عظیم اطلاعات و هم بخاطر عدم وجود عملکرد قبلی کاربر جدیدالورود در ابتدای عضویت بسیار حائز اهمیت است. راه حل کلیدی چالش ذکر شده بهره گیری از سیستم‌های پیشنهاد دهنده[1] می باشد [۱]. یک سیستم پیشنهاد دهنده با جمع آوری داده‌های مورد نیاز خود سعی می‌کند پیشنهادات مناسبی را برای کاربران تولید نماید. هدف سیستم پیشنهاد دهنده ارائه توصیه‌های معنی‌دار به مجموعه‌ای از کاربران برای استفاده از کالا یا خدماتی است که ممکن است مورد علاقه آنها باشد. سیستم‌های پیشنهاد دهنده در برخی از صنایع به شدت ضروری و حیاتی هستند، زیرا در صورت کارآیی می‌توانند درآمد زیادی ایجاد نموده و یا راهی برای پیشی گرفتن از رقبا باشند. امروزه شرکت‌های خرده‌فروشی و فروشگاه‌های آنلاین مختلف در سراسر دنیا، سعی دارند با استفاده از سیستم پیشنهاد دهنده و با بهره‌گیری از قدرت بی‌نظیر داده، تجربه‌ی مطلوبی برای مشتریان خلق کنند و از این راه، فروش خود را افزایش دهند [۲]. یک سیستم پیشنهاد دهنده با گردآوری اطلاعات پیرامون کاربر و اقلام و دریافت بازخورد صریح و ضمنی از کاربر یاد می‌گیرد که چگونه آیتمی را پیش‌بینی نماید و بر این اساس، عموما پیشنهاداتی را به کاربران سیستم ارائه می‌نماید. جریان های کاری و فازهای اصلی اجرائی در سیستم های پیشنهاد دهنده عموما شامل مراحل: گردآوری اطلاعات، انواع بازخورد، یادگیری و پیش‌بینی می باشد [۳].

امروزه سیستم های توصیه گر به شکل گسترده ای برای ارائه توصیه هایی به کاربران استفاده می شوند. اولویت استفاده از این سیستم ها به دلیل افزایش روز افزون حجم اطلاعات آنلاین و جلوگیری از سرریز اطلاعات بالا رفته است و این سیستم ها را به ابزار مفیدی برای غلبه به این چالش ها تبدیل کرده است. اخیرا سیستم های توصیه گر امروزه در حوزه های عملیاتی زیادی نفوذ کرده اند و روزانه در حال گسترش هستند. با وجود این پیشرفت ها سیستم های توصیه گر دچار چالش های پیچیده تری نیز شده اند. همچنین عملکرد سیستم های توصیه‌گر برای مدیران در حوزه های مختلف بسیار حائز اهمیت است و نیاز به دیدگاهها و ابزارهای مناسبی جهت ارزیابی و سنجش عملکردها و رفع چالش های موجود وجود دارد [۴].

در سیستم های توصیه گر یک چالش اساسی از دیرباز، مشکل شروع سرد می باشد. این مشکل زمانی رخ می دهد که یک کاربر جدید وارد سیستم می شود و چون اطلاعات کافی از این کاربر در سیستم موجود نمی باشد، سیستم قادر به ارائه ی پیشنهادات مناسب نخواهد بود. این به این معنا است که در این حالت نمی‌توان چیزی را به کاربران جدید توصیه کرد، چرا که سیستم هیچ گونه اطلاعی از نظرات یا علایق آنها ندارد و همچنین یک آیتم جدید را هم نمی‌توان به کاربران پیشنهاد داد زیرا که هیچ کاربری در مورد آن نظر نداده است. لذا، کاربرانی که اخیرا در شبکه‌های اجتماعی ثبت نام نموده اند و فعالیت قابل توجه‌ای در آن حوزه ندارند، اصطلاحاً با وضعیت شروع سرد مواجه خواهند بود [۵]. این مشکل عموما زمانی به وجود می‌آید که یک کاربر یا قلم‌داده جدید وارد سیستم می‌شود و عدم وجود اطلاعات قبلی از این درون داده باعث می‌شود سیستم نتواند به طور عادی پیشنهادها را تولید کند. به دلیل اهمیت موضوع، مرتفع نمودن این مشکل در شبکه‌های اجتماعی بسیار مورد توجه قرار گرفته و روشهای مختلفی نیز به منظور بهبود آن ارائه شده است. از این رو، ارائه یک راحل مطلوب که بتواند بدون در دست داشتن اطلاعات مربوط به تاریخچه فعالیت و عملکرد کاربران در سیستم شبکه‌های اجتماعی، دوستانی را به آنها پیشنهاد دهد، بسیار مهم و ضروری به نظر می‌رسد [۶]. بنابراین با توجه به نیاز جامعه مجازی امروزی به وجود یک سیستم پیشنهاد دهنده که بدون در دست داشتن اطلاعات عملکردی خاصی از کاربران جدیدالورود، بتواند دوستانی را پیشنهاد دهد یک امر لازم و ضروری است که این رویکرد در حوزه تمرکز پژوهش حاضر قرار دارد. در این مقاله قصد داریم ابتدا یک نمای کلی از وضعیت فعلی انواع سیستم های توصیه گر، چالش ها، محدودیت ها و کاربردهای آنها را بررسی نموده و سپس درجهت رفع برخی از چالش های مهم موجود، یک سیستم پیشنهاد دهنده ترکیبی مبتنی بر دانش جدید[2] و یادگیری تجمعی[3] جهت معرفی دوست در شبکه‌های اجتماعی طراحی و پیشنهاد نمائیم. در مقاله حاضر، یک سیستم پیشنهاد دهنده جدید را مبتنی بر اطلاعات جمعیت شناختی کاربران جدیدالورود و اطلاعات مشارکتی سایر کاربران موجود در شبکه اجتماعی، ارائه خواهیم نمود که علیرغم داشتن دقت کافی در ارائه پیشنهاد دوست به کاربران جدید، مشکلات شروع سرد و همچنین پراکندگی داده‌ها در این شبکه‌ها را مرتفع نماید و نرخ خطا را به حداقل رساند. لذا، ایده‌های اصلی و اهداف پژوهش حاضر طراحی یک پیشنهاد دهنده ترکیبی مبتنی بر دانش جدید و سیستم یادگیری جمعی جهت معرفی دوست در شبکه‌های اجتماعی و بهبود دقت و صحت پیشنهادات دوست در شبکه اجتماعی با استفاده از سیستم یادگیری تجمعی می‌باشد.

در ادامه مقاله ابتدا برخی از مهمترین پژوهش‌های انجام شده در زمینه‌های مرتبط به موضوع مورد بررسی قرار می‌گیرند. سپس تشریح کامل دیدگاه پیشنهادی با ارائه جزئیات طراحی و همچنین الگوریتم های پیشنهادی توصیف می‌گردد و نهایتا مراحل شبیه سازی دیدگاه پیشنهادی با نرم افزارهای مربوطه و مقایسه با سایر روشها مورد بحث

1 Recommender System

2 New Knowledge

3 Ensemble Learning

قرار می گیرد و براساس یافته‌های پژوهش، نتیجه گیری نهایی و پیشنهادات آتی ارائه می گردند.

## ۲- پیشینه تحقیق

با توجه به اهمیت توصیه پیوند و دوست در شبکه های اجتماعی، امروزه این موضوع مورد توجه پژوهشگران زیادی قرار گرفته است، بطوریکه هر کدام بر اساس روشها و تکنیک‌های مختلفی سعی کرده اند توصیه هائی را به افراد در شبکه‌های آنلاین ارائه نمایند.

نگوین-تی و همکارانش در مقاله [۱۶]، یک روش پیش بینی لینک مبتنی بر الگوریتم یادگیری ماشین آدابوست و ماشین بردار پشتیبان در شبکه‌های اجتماعی ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های Epinions، Slashdot و Wiki انجام دادند. همچنین معیارهای شباهت جاکارد و کسینوسی را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیارهای دقت و صحت استفاده نمودند. یائو و همکارانش در سال ۲۰۱۶ در مقاله [۱۷]، یک روش ترکیبی مبتنی بر پیش بینی لینک با همسایه مشترک در شبکه اجتماعی پویا ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های DBLP۱ انجام دادند. همچنین معیارهای شباهت جاکارد را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیار TPR و زمان اجرا استفاده نمودند. ژانگ و همکارانش در سال ۲۰۱۷ در مقاله [۱۸]، یک روش پیش‌بینی پیوند دینامیکی افزایشی کارآمد در شبکه اجتماعی ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های Facebook و Youtube انجام دادند. همچنین معیارهای شباهت کسینوسی را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیار زمان اجرا، دقت و AUC استفاده نمودند. لی و همکارانش در مقاله [۱۹]، یک روش پیش‌بینی لینک مبتنی بر کاربرد در شبکه اجتماعی ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست Facebook۲ انجام دادند. همچنین معیارهای شباهت جاکارد، RA۳ را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیار AUC، صحت، TPR استفاده نمودند.

چن و همکارانش در سال ۲۰۱۸ و در مقاله [۲۰]، یک روش تشخیص دهنده ارتباطات با استفاده از پیش‌بینی لینک در شبکه‌های اجتماعی چند ارتباطی ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های YouTube, disease-gene، network, climate network, و DBLP انجام دادند. همچنین معیارهای شباهت همسایه مشترک۴، جاکارد، Adamic/Adar، معیار PA۵ را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیار AUC، صحت، فراخوانی و F-Measure استفاده نمودند. زو و همکارانش در مقاله [۲۱]، یک الگوریتم طبقه بندی جدید مبتنی بر پیش بینی لینک در شبکه اجتماعی ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های سایت UCI انجام دادند. همچنین معیارهای شباهت نزدیکترین همسایه را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیار دقت، F-Measure و ARR۶ استفاده نمودند. مالک و همکارانش در سال ۲۰۱۹ در مقاله [۲۲]، یک روش مبتنی بر پیش‌بینی لینک در شبکه‌های اجتماعی بر اساس اطلاعات ساختاری و اجتماعی شبکه ارائه نمودند. آنها در تحقیق خود از دیتاست‌های BlogCatalog، Flickr و Facebook استفاده نمودند. همچنین معیارهای شباهت همسایه مشترک، جاکارد، Adamic/Adar، معیار PA را مورد استفاده قرار دادند. در نهایت نیز جهت ارزیابی روش مطرح شده از متریک‌های AUC و صحت استفاده نمودند. شارما و همکارانش در مقاله [۲۳]، یک مدلسازی مبتنی بر یادگیری چندسطحی برای پیش‌بینی لینک و ترجیح مصرف کنندگان در شبکه‌های اجتماعی آنلاین ارائه نمودند. آنها آزمایشات خود را بر روی دیتاست‌های Amazon و +Google انجام دادند. همچنین معیارهای شباهت همسایه مشترک و جاکارد را مورد استفاده قرار دادند. آنها در تحقیق خود جهت ارزیابی عملکرد روش مطرح شده از معیارهای AUC، TPR۷، F-Measure، r-Measure ، p-Measure استفاده نمودند.

یوان و همکارانش در مقاله [۲۴]، یک روش پیش‌بینی لینک مبتنی بر هسته گراف و الگوریتم SVM۸ برای شبکه‌های اجتماعی ارائه نمودند. آنها در تحقیق خود، شبیه سازی روش مطرح شده را بر روی دیتاست Epinions آزمایش نموده و همچنین از معیارهای ارزیابی دقت و F1-Score استفاده کردند. همچنین از معیارهای شباهت همسایه مشترک، جاکارد و Adamic/Adar را مورد استفاده قرار دادند. یو و همکارانش در سال ۲۰۱۹ در مقاله [۲۵]، یک روش پیش بینی لینک برای رفع شروع سرد کاربران در شبکه‌های اجتماعی چند رابطه‌ای مبتنی بر تحلیل وابستگی شبکه ارائه نمودند. کوی و همکاران در سال ۲۰۲۰ و در مقاله [۲۶]، سیستم‌های پیشنهاد دهنده مبتنی بر فیلتر مشارکتی در اینترنت اشیاء را بررسی نمودند. در این مقاله یک مدل جدید بر اساس ضریب همبستگی زمان و الگوریتم k-means و جستجوی فاخته، به نام TCCF پیشنهاد نمودند. علاوه بر این ، یک مدل پیشنهادی موثر و شخصی مبتنی بر الگوی ترجیحی (PTCCF) برای بهبود کیفیت TCCF طراحی شده است.

[1] http://dblp.uni-trier.de/
[2] http://snap.stanford.edu/data/egonets-Facebook.html
[3] Resource Allocation index
[4] Common Neighbors
[5] Preferential Attachment
[6] Adjusted Ratio of Ratios
[7] True positive rate
[8] Support Vector Machine

نصار و همکاران در مقاله [۲۷]، یک مدل جدید فیلتر مشارکتی چندمعیاره در سیستم‌های پیشنهاددهنده را بررسی نمودند. در این مقاله یک مدل فیلتر چندگانه چند معیار جدید مبتنی بر یادگیری عمیق پیشنهاد گردید. ماشال و همکاران در مقاله [۲۸]، به تجزیه و تحلیل سیستم‌های پیشنهاددهنده چندمعیاره در اینترنت اشیاء پرداختند. در این مقاله یک سیستم پیشنهاددهنده با استفاده از یک رویکرد تصمیم گیری چند معیاره ترکیبی مبتنی بر فرآیند سلسله مراتبی تحلیلی و روش‌های ساده وزن افزودنی پیشنهاد گردید. برکانی و همکاران در سال ۲۰۲۱ در مقاله [۲۹]، سیستم‌های پیشنهاددهنده رایج در شبکه‌های اجتماعی را مورد بررسی قرار دادند. روش پیشنهادی این مقاله مبتنی بر الگوی نمایش معنایی و اجتماعی از پروفایل کاربران است.

در مقایسه رویکرد پژوهش حاضر با تحقیقات بررسی شده لازم به ذکر است، بطور کلی هدف اصلی سیستم پیشنهاد شده در این مقاله، پیشنهاد دوست در شبکه اجتماعی به کاربران جدیدالورود که با مشکل شروع سرد مواجه هستند، با استفاده از اطلاعات دموگرافیک کاربران است. درواقع اغلب راه حل‌های ذکر شده مساله را بصورت پراکنده و موردی با تاکید بر کاربرد و یا استراتژی خاصی بررسی نموده اند. در سیستم پیشنهادی در این مقاله سعی شده است که حجم عظیمی از پردازش‌ها بر روی داده‌ها اعم از پیش پردازش بر روی داده‌ها، آماده سازی داده، انتخاب ویژگی‌های بهینه و خوشه بندی اولیه کاربران در پس زمینه انجام شود تا ارائه پیشنهادات دوست با صحت و با دقت بیشتر صورت پذیرد. پس از شبیه سازی روش پیشنهادی و محاسبه میزان خطا و دقت مشاهده گردید، میزان دقت نهایی پیشنهاد دوست به کاربر جدیدالورود در این سیستم برابر با ۹۰٪ می‌باشد. میانگین صحت پیشنهاد دوست در روش پیشنهادی برابر با ۰،۰۱۹۷ است که نسبت به سایر روشهایی که اخیرا بر روی دیتاست مشابه استفاده شده، بهبود قابل توجهی داشته است. همچنین سیستم پیشنهادی علاوه بر مرتفع نمودن چالش شروع سرد کاربران جدیدالورود، دقت ارائه دوستان جذاب به کاربران را افزایش داده و نرخ خطا را به حداقل می رساند. در بخش بحث در نتایج و مقایسه در مقاله حاضر، راهکار پیشنهادی این پژوهش را با مهم ترین تحقیقات مرتبط از طریق نمودار مقایسه نموده ایم. در بخش بعدی کلیات و ابعاد دیدگاه پیشنهادی را تشریح می نمائیم.

## ۳- سیستم پیشنهادی

هدف اصلی روش مطرح شده در این پژوهش، پیشنهاد دوست در شبکه اجتماعی به کاربران جدیدالورود که با مشکل شروع سرد مواجه هستند، با استفاده از اطلاعات دموگرافیک کاربران است. کاربران جدید الورود آن دسته از کاربرانی هستند که عملکرد و فعالیتی در سیستم ندارند. در این سیستم سعی بر این شده است که حجم عظیمی از پردازش‌ها بر روی داده‌ها اعم از پیش پردازش بر روی داده‌ها، آماده سازی داده، انتخاب ویژگی‌های بهینه و خوشه بندی اولیه کاربران در پس زمینه انجام شود تا بتوان ارائه پیشنهادات دوست با سرعت قابل قبول و با دقت مطلوبی صورت پذیرد. یکی از مهمترین مزایای اینکه عملیات ذکر شده در پس زمینه سیستم پیشنهادی انجام می‌شود این است که پیشنهادات دوست به کاربران جدیدالورود با سرعت بیشتری صورت پذیرد. به عنوان مثال فرآیندهای ذکر شده در دوره‌های یک ماهه یا چند ماهه تنها یکبار لازم است اجرا شوند. از طرفی از بخش انتخاب دسته برای کاربر جدیدالورود در سیستم پیشنهادی به ازای هر کاربر جدید لازم الاجرا است. روش پیشنهادی دارای فازهای اجرایی و جریان های کاری مرتبط به شرح زیر می‌باشد:

**فاز ۱: ورود داده‌ها:** ابتدا داده‌های مربوط به کاربران پوکک همراه با ۵۹ مشخصه و ویژگی موجود به سیستم پیشنهادی وارد می‌شوند.

**فاز ۲: پیش پردازش بر روی داده‌ها:** پس از ورود داده ها عملیات پیش پردازش داده بر روی این اطلاعات صورت می‌پذیرد. در این مرحله استراتژی استفاده شده بدین صورت است که، کاربرانی را که دارای اطلاعات ناقص هستند استخراج می‌شود. با توجه به این معمولا کاربران ویژگی‌هایی را تکمیل نمی‌کنند، در این پژوهش ویژگی‌های ناقص هر کاربر را توسط سیستم پیشنهادی با توجه به اطلاعات تکمیل شده توسط کاربران مشابه دیگر تکمیل می‌شود. فرض کنید U1, U2 دارای اطلاعات کاملی بود و U3 اطلاعاتی اعم از علاقه مندی به ورزش و موزیک ناقص است. از میان کاربران موجود یک شباهت گیری بر اساس ویژگی‌ها صورت می‌گیرد. بعنوان مثال کاربر ۱و۲ که میزان مشابهت بیشتری با کاربر ۱ دارند، ویژگی‌های علاقه مندی به موزیک و ورزش این دو کاربر ۱ است. بنابراین ویژگی‌های ذکر شده مربوط به کاربر ۳ را ۱ قرار می‌دهیم. با این استراتژی کاربرانی که دارای اطلاعات ناقصی هستند، تکمیل شده و در فرآیند پردازش مشکل زا نخواهند بود. پس از اعمال پیش پردازش بر روی داده‌ها، آن دسته از کاربرانی که دارای اطلاعاتی ناقص و بیش از نصف ویژگی باشد کلا از داده‌ها حذف می‌شوند.

**فاز ۳: آماده سازی داده‌ها:** جهت اجرای داده‌ها در نرم افزارهای برنامه نویسی و داده کاوی می‌بایست به فرمت قابل قبولی تبدیل شوند. در این فاز کلیه داده‌ها در قالب فایل متنی و اکسل مکانیزه می‌شوند.

**فاز ۴: تفکیک کاربران با ویژگی مشترک:** در این مرحله از میان کل ویژگی‌ها، تنها ویژگی‌های انتخاب شده، برگزیده می‌شود و مابقی ویژگی‌ها حذف می‌شوند.

**فاز ۵: اعمال خوشه بندی اولیه با الگوریتم K-Means :** داده‌های مرحله قبل به الگوریتم خوشه بندی X-Means جهت اعمال خوشه بندی اولیه داده‌ها وارد می‌شوند. این الگوریتم نسخه توسعه یافته الگوریتم K-Means است. در الگوریتم K-Means لازم

است تعداد خوشه‌ها از قبل تعیین گردد. با توجه به اینکه تعداد خوشه‌های بهینه از قبل مشخص نیست، بسیاری از پژوهشگران یا با مشکل مواجه خواهند شد یا مجبور هستند تعداد خوشه‌های متنوعی را به الگوریتم داده و در تکرارهای مختلفی اجرا نمایند. این فرآیند هم وقت گیر بوده و هم با خطا مواجه است. با توجه به وجود چنین محدودیتی در الگوریتم K-Means، از الگوریتم توسعه یافته X-Means استفاده می‌گردد. در این الگوریتم لازم به وارد نمودن K نبوده و خود الگوریتم در تعداد تکرارهای بهینه، تعداد خوشه یا K بهینه را انتخاب می‌کند [۳۰]. بنابراین در این مرحله کلیه کاربران موجود در سیستم بر اساس اطلاعات دموگرافیکشان خوشه بندی می‌شوند. این خوشه بندی اولیه به عنوان یک روش دسته بندی بر روی داده‌های بدون نظارت، اعمال شده تا بتوان در الگوریتم‌های یادگیری جمعی جهت آموزش مدلها مورد استفاده قرار گیرد.

**فاز ۶: تفکیک داده‌های آموزشی و آزمایشی:** پس از خوشه بندی داده‌ها لازم است تعدادی از کاربران را به عنوان کاربر جدیدالورود و تعدادی از کاربران را جهت آموزش مدلهای الگوریتم‌های سیستم یادگیری جمعی تفکیک شوند. بدین منظور ۸۰٪ از داده‌ها به عنوان داده‌های آموزشی و ۲۰ درصد از داده‌ها را به عنوان داده‌های آزمایش تفکیک خواهد شد. لازم به ذکر است که تا این مرحله در پس زمینه سیستم انجام شده و کاربر هیچ اطلاعی از آن ندارد.

**فاز ۷: انتخاب دسته بهینه با کمک رویکرد یادگیری جمعی:** پس از اینکه داده‌های خوشه بندی شده تفکیک شدند، لازم است کاربری که جدیدالورود است به سیستم وارد شود. جهت ارائه پیشنهادات دوست به کاربر جدیدالورد، ابتدا کلیه الگوریتم‌های درخت تصمیم اعم از الگوریتم C4.5, CHAID,CART بر اساس داده‌های آموزشی، مدلهای خود را تولید نموده و درختهایی را نیز ایجاد می‌نمایند. جهت تعیین یک دسته یا اصطلاحا یک خوشه به کاربر جدیدالورود لازم است تا کاربر جدید به سیستم یادگیری جمعی اعمال شده و به درخت‌های تولید شده نیز وارد شود. این فرآیند بسیار مناسب تر از این حالتی است که کاربر جدیدالورد از طریق تکنیک خوشه بندی، دسته مورد نظرش تعیین شود. بنابراین پس از تعیین دسته مربوط به کاربر جدیدالورود توسط هر الگوریتم، پاسخ مورد نظر به یک هسته وارد می‌شود. همانند استراتژی استفاده شده در سیستم بوستینگ، در سیستم یادگیری جمعی دسته‌هایی را که دارای بیشترین تعداد هستند به عنوان دسته اصلی برای کاربر هدف انتخاب می‌گردد.

**فاز ۸: تشکیل ماتریس مجاورتی:** در این مرحله پس از اینکه دسته ای برای کاربر جدید انتخاب شد، کلیه کاربرانی که در آن دسته هستند با کاربر جدید در یک ماتریس مجاورتی قرار می‌گیرند. سطر ماتریس را شماره کاربران و ستون‌ها را ویژگی‌های آنها تشکیل داده است.

**فاز ۹: محاسبه شباهت و پیشنهاد دوست:** این مرحله که آخرین مرحله در روش پیشنهادی است، بر اساس معیارهای شباهت ویژگی‌ها، شباهت بین تک تک کاربران را محاسبه نموده و در نهایت آن دسته از کاربرانی که دارای مشابهت بیشتری هستند به کاربر جدید پیشنهاد می‌گردد. در این مرحله نیز از ترکیب معیارهای شباهت گیری استفاده می‌گردد. بنابراین با استراتژی فوق می‌توان پیشنهادات مطلوب تری را ارائه نمود.

## ۴- شبیه سازی و ارزیابی

در این پژوهش از دیتاست محبوب شبکه اجتماعی اسلواکی استفاده شده است. این دیتاست که با نام pokec[1] معروف است که متشکل از دو فایل کلی می‌باشد [۳۱]. فایل اول، پروفایل مربوط به ۱۶۳۲۸۰۳ کاربر را نمایش میدهد. فایل دوم نیز دوستان و افرادی را که هر کدام از کاربران با آنها در ارتباط است را نیز نشان می‌دهد. در این فایل بیش از ۱٫۶ میلیون رکورد وجود دارد. این فایل شامل مجموعه ۵۹ ویژگی درخصوص اطلاعات شخصی، زمینه های کاری، علائق و خصوصیات افراد و مقادیر مرتبط می باشد. درواقع تعداد ویژگی‌های دموگرافیک افراد در حدود ۵۹ ویژگی است. به طور کلی این فایل شامل دو ویژگی بوده که نود یا کاربر مبدا و مقصد را مشخص می‌کند.

در فاز پیش پردازش و پاکسازی داده‌ها، با توجه با این که منبع داده استفاده شده دارای مقادیر تهی و بلا استفاده است و همچنین شامل برخی ویژگی‌های اضافی نیز میباشد، نمونه‌هایی را که شامل داده‌های پرت بوده و همچنین ویژگی‌هایی را که بلا استفاده و اضافه هستند حذف می‌کنیم. عملیات پیش پردازش بروی داده‌ها توسط نرم افزار داده کاوی رپیدماینر و نرم افزار اکسل انجام می‌شود. روال کار بدین صورت است که ابتدا ویژگی‌هایی را که شامل مقادیر پرت است را از طریق نرم افزار داده کاوی رپیدماینر شناسایی کرده و سپس از طریق نرم افزار اکسل یا همان رپید ماینر اقدام به حذف آن ویژگی می‌نمائیم. در آخر نیز با حذف ویژگی‌های بلا استفاده، نمونه رکوردهایی که شامل داده‌های پرت هستند از نمونه‌ها حذف کرده تا در نتایج تاثیر منفی نداشته باشند. نکته قابل تاملی که در پیش پردازش داده‌ها وجود دارد این است که در فیچرهای رشته ای که مستقیم توسط کاربران مقدار دهی شده است مورد قبول نبوده و در پیش پردازش حذف می‌شوند. در مرحله پیش پردازش برخی از فیلدها که دارای مقادیر مختلفی از رشته‌ها هستند تحلیل شده و با مقادیر ۰ و ۱ جایگزین می‌شود. داده های پرت و نمونه‌هایی که فیچرهایشان دارای مقادیر null هستند و برخی نیز دارای مقدار واقعی هستند، به ازای مقادیر null، با مقادیر پیش فرض در نظر گرفته می‌شود. در

[1] http://snap.stanford.edu/data/soc-pokec.html

بخش بعد بهینه ترین تعداد خوشه را با استفاده نرم افزار داده کاوی weka تعیین نمونه، تعداد k=n که n تعداد خوشه بهینه است به الگوریتم k-means اعمال شده و در نهایت تعداد 2,000 کاربر خوشه بندی می‌گردد.

جهت خوشه بندی کاربران میبایست ابتدا داده‌های مربوطه را به فرم پذیرفته شده نرم افزار داده کاوی وکا تبدیل نمود. ابتدا داده‌های مربوط به اطلاعات دموگرافیک کاربران را تبدیل به فرمت قابل قبول وکا می‌کنیم. سپس با استفاده از نرم افزار داده کاوی وکا، داده‌ها را به برنامه ایمپورت نموده و ادامه مراحل انجام می‌گردد. پس از وارد نمودن دیتاست کاربران به نرم افزار، میبایست جهت انسجام داده‌ها یک فیلتر Normalize بروی دیتاست اعمال شود.

به کلیه مراحل انجام شده بالا، پیش پردازش بروی داده‌ها گفته میشود. بعد از انجام مراحل فوق از منوی Cluster عملیات خوشه بندی را انجام می‌دهیم. لازم به ذکر است که جهت خوشه بندی داده‌ها و انتخاب ورودی‌های مسئله راه‌های مختلفی وجود دارد که در پژوهش فعلی انتخاب نمونه‌ها به صورت 70% داده‌های آموزشی و ٪۳۰ داده‌های آزمایشی انتخاب شده است. در این مرحله کافیست از Clusterer الگوریتم k-means را انتخاب نموده و با کلیک بروی نام آن به مراتب kهای مختلف را مورد ارزیابی قرار دهیم. در نهایت به ازای کلیه kهای تنظیم شده برای الگوریتم k-means میزان خطا را محاسبه و نسبت به سایر مقادیر، ارزیابی می‌کنیم. لذا جهت انجام خوشه بندی و استخراج خوشه‌های تهیه شده، این عملیات را توسط نرم افزار داده کاوی rapidminer انجام داده تا بتوانیم در مراحل دیگر از خروجی مربوطه استفاده کنیم. با توجه به انجام آزمون و خطاهای تست شده مشاهده گردید که تعداد k=100 تعداد خوشه‌های بهینه است و به همین جهت k=100 به عنوان تعداد خوشه‌های پیش فرض برای الگوریتم خوشه بندی در نظر گرفته می‌شود.

یکی از مهمترین مراحلی که در این پژوهش ضروریست اعمال خوشه بندی کاربران بر اساس اطلاعات دموگرافیک است. مهمترین علت انتخاب الگوریتم k-means سرعت خوشه بندی و اجرای بروی حجم بالای داده‌ها است. لذا جهت محاسبه فاصله بین نقاط داده تا مرکز خوشه از تابع فاصله اقلیدوسی بدست آمده است. با توجه به اینکه تعداد k یا نقاط مرکز داده در الگوریتم خوشه بندی k-means می‌بایست از طرف محقق وارد شود، نیاز مند محاسبه میزان خطای خوشه بندی با توجه به تعداد k است. بنابراین جهت اعمال خوشه بندی کاربران با استفاده از نرم افزار رپیدماینر مراحل زیر لازمم است انجام شود:

- ابتدا کنترل readexcel را به پروسس اضافه نمود
- داده‌ها را از طریق این کنترل به برنامه وارد نمود
- با استفاده از کنترل Normalize داده‌ها را بین اعداد ۰ و ۱ نرمال نمود
- کنترل k-means را اضافه کرده و تعداد k=200 یا k=100 تعیین کرده و خروجی این کنترل به یک کنترل writeExcel متصل شده تا خروجی در قالب فایل اکسل ذخیره سازی گردد.

پس از این مرحله، با استفاده از الگوریتم درخت تصمیم CHAID، خروجی بخش قبل را به داده‌های آزمایشی و داده‌های آموزشی تقسیم بندی می‌کنیم. به منظور آموزش درخت تصمیم لازم است که مجموعه داده‌هایی را به عنوان داده‌های آموزشی استفاده نمود. همچنین بخش بسیار کمی از داده‌ها را به عنوان داده‌های تست در نظر گرفته و جهت ارزیابی مدل تهیه شده استفاده گردد. بنابراین حدود ٪۹۸ از داده‌ها را به عنوان داده‌های آموزشی و چندین مورد از داده‌ها در حدود ٪۲ را به عنوان داده‌های آزمایشی به منظور ارزیابی الگوریتم تفکیک می‌کنیم. ٪۹۸ از خروجی بخش خوشه بندی به عنوان داده‌های آموزشی به درخت تصمیم وارد می‌گردد. آموزش‌ها و قوانین بر اساس معیار information_gain به درخت اعمال می‌شود. در انتها داده‌های آموزشی که در حدود ۵ نمونه است به درخت تصمیم داده شده و دسته مربوطه شناسایی می‌گردد.

پس از اجرای الگوریتم درخت تصمیم بر روی داده‌های فوق مشاهده گردید که تنها یکی از دسته‌ها جزو دسته‌های اصلی نبوده و نزدیک به دسته مشابه است و لذا دقت روش انتخاب در حدود ۹۰ ٪ می‌باشد. در نهایت میزان خطای این درخت برابر با ۰,۳۰۷۴ است. پس از اینکه خوشه یا دسته مورد نظر برای کاربر جدید انتخاب شد، می‌بایست همسایه‌های مشابه هر کاربر $n_j \in N$ استخراج شود. الگوریتم مربوطه جهت یافتن نزدیکترین همسایه‌ها در ادامه نشان داده شده است. چس از اینکه کاربران همسایه در قالب آرایه NG مقدار دهی شدند می بایست شباهت بین کاربران جدید و کاربران مشابه در لیست NG را محاسبه نمود. در نهایت نیز پیش بینی نهایی بر اساس امتیازات کاربران همسایه است. نتایج معیارهای شباهت گیری بر اساس ویژگی‌های دموگرافیک کاربران بوده که با مجموعه D نشان داده شده است. معادله نهایی جهت محاسبه میزان شباهت بین کاربر جدید با کاربران همسایه بصورت زیر می باشد:

$$sim(n,u) = \frac{\sum_{j=1}^{I} SF_{j*} w_j}{\sum_{j=1}^{I} w_j} \qquad (۱)$$

که در آن SFj مقدار مشابهت jمین ویژگی است و wj وزن مربوط به ویژگی مورد نظر است. جهت خوشه بندی کاربران و محاسبه میزان شباهت آنها از ویژگی‌های سن، جنسیت و شغل و غیره استفاده شده است که به هر کدام وزن‌هایی با توجه به میزان اهمیتشان داده شده است. به عنوان مثال به ویژگی سن وزن 0,5، به ویژگی جنسیت وزن 0,25 و به ویژگی شغل نیز وزن 0.25 اطلاق گردیده است. مجموع وزن ها باید ۱ باشد. بنابراین: $D = \{d1 = age, d2 = gender, d3 = Occupation\}$ و مجموعه وزن‌ها عبارتند از:

$W = \{w1 = 0.5\ , w2 = 0.25\ , w3 = 0.25\} = 1$. در این پژوهش از یک معیار ترکیبی جهت محاسبه میزان شباهت بین کاربران استفاده شده است. برای هر ویژگی dj ما یک تابع SF(at1,at2) تعریف کرده ایم که مقادیری بین [0,1] دارند. این تابع میزان مشابهت دو ویژگی مربوط به یک جفت کاربر را محاسبه می‌کند. با توجه به ماهیت ویژگی‌هایی که جهت محاسبه شباهت‌ها در نظر گرفته ایم، دو گروه کلی شامل ویژگی های عددی و ویژگی های رشته ای وجود دارد. برای ویژگی‌های عددی مثل سن یک معیار شباهت به صورت زیر در نظر گرفته ایم:

$$w_{age} = \begin{cases} (1 - \frac{|Diff|}{Diff_{max}})^{\beta} & if\ |Diff| \le Diff_{max} \\ 0 & if\ |Diff| > Diff_{max} \end{cases} \quad (۲)$$

در رابطه بالا Diff اختلاف سن بین کاربران را نشان میدهد و Diffmax یک حداکثر مقدار اختلاف است که توسط محقق تعریف می‌شود. β نیز پارامتری است که در صورتیکه محقق بخواهد مقدار وزن wage را نسبت به مقدار Diff افزایش دهد کافیست مقدار بتا را کمتر از ۱ قرار دهد. از آنجائی که برخی از ویژگی‌های دیتاست شبکه اجتماعی اسلواکی در قالب مقادیر رشته ای بیان شده اند، لذا جهت محاسبه میزان شباهت این ویژگی‌ها از رابطه زیر استفاده می‌شود:

$$w_{str_attr} = \frac{|n_j.StrAttribute \cap u_j.StrAttribute|}{MAX(LEN(n_j.StrAttribute), LEN(u_j.StrAttribute))} \quad (۳)$$

Nj بیانگر کاربر جدید و uj بیانگر کاربر موجود در خوشه است. همانطور که از فرمول بالا مشخص است، ویژگی رشته ای دو کاربر دریافت شده و در قالب آرایه رشته ای قرار داده می‌شود. در صورت فرمول تعداد اشتراکات بین ویژگی مربوطه محاسبه می‌شود. مخرج کسر نیز ماکسیمم تعداد اقلام آرایه می‌باشد. فرض کنید کاربر جدید الورود با سه زیان دنیا در تعامل است که عبارتند از $n_j.StrAttribute = \{EN, SP, US\}$ و همچنین یکی از کاربران موجود در خوشه مربوطه با زبان $U_j.StrAttribute = \{EN, SP\}$ در تعامل می‌باشد. لذا مخرج کسر برابر با ۲ و صورت کسر برابر با ۳ است. در نهایت میزان شباهت بین این دو کاربر برابر با ۰,۶۶ می‌باشد. زبان گویش، سرگرمی ها و فعالیت های ورزشی کاربران از رابطه فوق استفاده می‌کنند. همچنین مقادیر بولین نیز از معیار شباهت زیر استفاده می‌گردد:

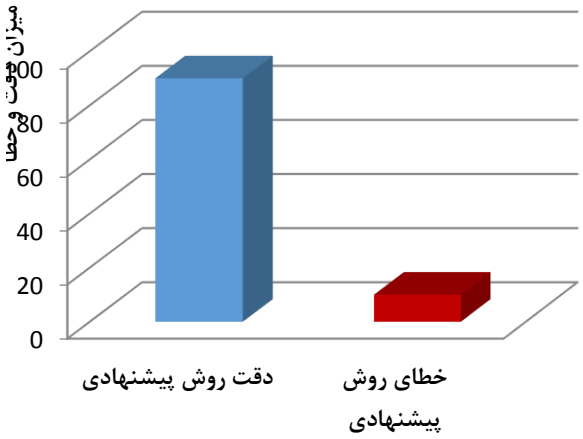
$$W_{gender} = \begin{cases} 1 & \text{if att1 == att2} \\ 0 & \text{if at1 <> att2} \end{cases} \quad (۴)$$

با توجه به اینکه مقادیر ویژگی‌های ۱ و ۲ با هم یکسان هستند یا خیر، مقدار ۱ یا ۰ برگردانده می‌شود. بدلیل اینکه برخی از ویژگی‌های کاربران به صورت بولین در دیتاست تعریف شده است، می‌توان از رابطه بالا به عنوان یک فرمول واحد برای محاسبه میزان شباهت بین این دو ویژگی‌ها استفاده نمود. ویژگی هایی که از فرمول بالا استفاده می‌کنند ویژگی علاقه مندی به فیلم، ویژگی علاقه مندی به موزیک، ویژگی ارتباط کاربر با الکل، ویژگی ارتباط به کودکان، ویژگی نحوه استفاده از سیگار و ویژگی عمومی‌بودن حساب کاربری کاربران می باشد. در الگوریتم ترکیب، کلیه معیارهای شباهت ترکیب شده و در نهایت یک عدد بین ۰ و ۱ محاسبه می‌گردد که این عدد بیانگر میزان شباهت بین دو کاربر با روش پیشنهادی است. پس از اینکه کاربر جدید به خوشه مربوطه از طریق درخت تصمیم C4.5 اضافه شد، کلیه معیارهای شباهت ترکیبی بروی کاربر جدید با تمامی کاربران موجود در خوشه مورد نظر اعمال می‌گردد. پس از محاسبه میزان شباهت کاربر جدید با سایر کاربران، با توجه به حد آستانه تعدادی از دوستانی که دارای بالاترین میزان شباهت هستند به کاربر جدید پیشنهاد می‌گردد. پس از آن، کاربر جدید به گراف موجود اضافه شده و دوستان آن نیز مشخص می‌گردد، سپس الگوریتم friendLink که از محبوب ترین الگوریتم‌های پیش بینی لینک است بر روی گراف جدید اعمال شده و دوستان دیگر را با دوستان ارائه شده ترکیب نموده و به عنوان پیشنهاد به کاربر جدید ارائه می‌گردد [۳۲]. رابطه مربوط به الگوریتم پیش بینی لینک بصورت زیر می باشد:

$$Sim(n_j, u_j) = \sum_{i=2}^{l} \frac{1}{i-1} \cdot \frac{|path^{i}_{n_j,u_j}|}{\prod_{y=2}^{i}(n-y)} \quad (۵)$$

همانطور که در فرمول بالا مشخص است، ni کاربر جدید که به تازگی در گراف شبکه اجتماعی مستقر شده است و uj تک به تک کاربرانی است که دارای بیشترین شباهت با کاربر جدید هستند. L طول مسیر را مشخص می‌کند. n تعداد کل گره‌های گراف را مشخص می‌کند. $\sum_{i=2}^{l} \frac{1}{i-1}$ یک فاکتور وزن دهی است که برای مسیرهای بیشتر از L=2 تاثیر گذاری بیشتری دارد. فرض کنید حداکثر طول مسیر مد نظر برابر با ۲ است(L=2)، در این شرایط $\sum_{i=2}^{l} \frac{1}{i-1} = 1$ بوده و تاثیری در شباهت نهایی ندارد. در حالتی که L=3 باشد این وزن به 0.5 تغییر یافته و در مشابهت تاثیر مطلوبی دارد. $|path^{i}_{n_j,u_j}|$ تعداد مسیرهای بین کاربر جدید و کاربری که با دوستان کاربر جدید در ارتباط است با توجه به تعداد طولشان نشان داده شده است. به عنوان مثال، با فرض اینکه کاربر جدید u1 است و دوستان پیشنهاد شده به این کاربر u2,u3,u5,u6,u8 می‌باشد، حداکثر طول مسیر به کاربر u4 که یکی از دوستان کاربر پیشنهاد شده به کاربر جدید(u2,u3) است برابر با ۳ می‌باشد. $\prod_{y=2}^{i}(n-y)$ تعداد مسیرهای ممکن از کاربر جدید تا کاربر هدف است. فرایند پیشنهاد پیوند چندین پارامتر را به عنوان ورودی دریافت می‌کند که عبارتند از: ۱) N بیانگر تعداد کاربران موجود در شبکه است که در

واقع شناسه‌های کاربران مورد بررسی را نشان می‌دهد. ۲) newUser بیانگر کاربر جدیدالورود است با شروع سرد مواجه می‌باشد.۳) NewNodesنیز لیست کاربرانی است که بیشترین شباهت با کاربر جدیدالورد در خوشه انتخابی در مراحل قبل را دارد. روال کار الگوریتم پایه friendLink بدین صورت است که ابتدا گراف کل شبکه را دریافت نموده و پس از محاسبه همسایه‌های کاربر جدید الورود، بهترین مسیرهای موجود در جهت رسیدن به سایر دوستان کاربران همسایه با توجه به طول مسیر را محاسبه نموده و در نهایت با توجه به فرمول ارائه شده در قسمت بالا، میزان شباهت را محاسبه می‌نماید. یکی از تفاوت‌های رویکرد پیشنهادی در این قسمت به الگوریتم پایه friendLink این است که بر روی کل گراف شبکه پردازش را انجام نداده و تنها کاربرانی را مورد پویش قرار می‌دهد که با استفاده از معیارهای شباهت گیری ترکیبی از خوشه انتخاب شده کاربر جدیدالورود در مرحله قبل استخراج شده اند. روال کلی الگوریتم فوق بدین صورت است که یک گراف ارتباطی بین کاربر جدید با کاربرانی که بیشترین مشابهت را با توجه به یک حد آستانه دارا می‌باشند تشکیل می‌دهد. لازم به ذکر است که در این گراف کلیه دوستان کاربران استخراج شده توسط روش شباهت گیری ترکیبی در گراف شرکت داشته و الگوریتم پیشنهادی بر روی این دوستان اجرا می‌گردد. بنابراین لازم نیست مجدد کل گراف شبکه پیمایش شود. پس از اینکه میزان شباهت کاربر جدید با سایر دوستان کاربران مشابه بر اساس زوال مسیر و حد آستانه محاسبه شد؛ نتایج این قسمت با مرحله قبل ترکیب شده و کاربران نهایی با بالاترین امتیاز به کاربر جدیدالورود به عنوان دوست پیشنهاد می‌گردد. پس از شبیه سازی روش پیشنهادی و محاسبه میزان خطا و دقت روش مطرح شده در این پژوهش، مشاهده گردید میزان دقت نهایی پیشنهاد دوست به کاربر جدیدالورود برابر با ۹۰٪ می‌باشد. لازم به ذکر است که پارامتر دقت این بدین صورت محاسبه شده است که یک کاربر موجود در سیستم را که از قبل تعداد دوستانش مشخص است را انتخاب نموده و پس از اجرای روش پیشنهادی، نتایج بدست آمده با دوستان واقعی مقایسه گردید است. پس از مقایسه نتایج مشاهده گردید که ۹۰ درصد از دوستان دقیقا مشابه بوده و ۱۰ درصد از دوستان پیشنهاد شده نیز جزو دوستان قبل کاربر جدید الورود نمی‌باشند. لذا با توجه به این دیدگاه میانگین صحت روش پیشنهادی برای کلیه کاربران جدیدالورود در بخش بعد مورد بررسی قرار گرفته است. در نمودار زیر میزان دقت و خطای روش پیشنهادی نشان داده شده است.




شکل ۴- محاسبه میزان دقت و خطای روش پیشنهادی

## ۵- نتیجه‌گیری

با توجه به نیاز شبکه‌های اجتماعی جدیدالورود در این پژوهش سیستم پیشنهاد دهنده ترکیبی براساس الگوهای دانش جدید و یادگیری تجمعی جهت پیشنهاد دوست در شبکه های اجتماعی طراحی گردید است. پس از شبیه سازی روش پیشنهادی و محاسبه میزان خطا و دقت روش مطرح شده در این پژوهش، میزان دقت نهایی پیشنهاد دوست به کاربر جدیدالورود برابر با ۹۰٪ می‌باشد. میانگین صحت پیشنهاد دوست در روش پیشنهادی برابر با ۰،۰۱۹۷ است که نسبت به سایر روشهایی که اخیرا بر روی دیتاست استفاده شده در این پژوهش انجام شده است بهبود قابل توجهی داشته است. بهره گیری از سایر روش ها و الگوریتم های توصیه هوشمند و همچنین استفاده از سایر معیارهای موثر شباهت سنجی می تواند از راهکارهای آتی این پژوهش باشد.

## تعارض منافع

هیچ‌گونه تعارض منافع توسط نویسندگان بیان نشده است.

## منابع

[1] L. Quijano-Sánchez, I. Cantador, M. E. Cortés-Cediel, and O. Gil, “Recommender systems for smart cities,” Information Systems, vol. 92, Art. no. 101545, 2020.

[2] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” Decision Support Systems, vol. 74, pp. 12–32, 2015.

[3] B. Alhijawi and Y. Kilani, "The recommender system: A survey," International Journal of Advanced Intelligence Paradigms, vol. 15, no. 3, pp. 229–251, 2020.

[4] H. Yuan and A. A. Hernandez, "User cold start problem in recommendation systems: A systematic review," IEEE Access, vol. 11, pp. 136958–136977, 2023.

[5] R. Sethi and M. Mehrotra, "Cold start in recommender systems—A survey from domain perspective," in Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, Singapore: Springer, 2021, pp. 223–232.

[6] M. Rahman, I. A. Shama, S. Rahman, and R. A. H. M. A. T. U. L. L. A. H. Nabil, "Hybrid recommendation system to solve cold start problem," Journal of Theoretical and Applied Information Technology, vol. 100, no. 11, pp. 3562–3578, 2022.

[7] D. Jannach, A. Manzoor, W. Cai, and L. Chen, "A survey on conversational recommender systems," ACM Computing Surveys, vol. 54, no. 5, pp. 1–36, 2021.

[8] A.-T. Nguyen-Thi, P. Q. Nguyen, T. D. Ngo, and T.-A. Nguyen-Hoang, "Transfer AdaBoost SVM for link prediction in newly signed social networks using explicit and PNR features," Procedia Computer Science, vol. 60, pp. 332–341, 2015.

[9] L. Yao, L. Wang, L. Pan, and K. Yao, "Link prediction based on common-neighbors for dynamic social network," Procedia Computer Science, vol. 83, pp. 82–89, 2016.

[10] Z. Zhang, J. Wen, L. Sun, Q. Deng, S. Su, and P. Yao, "Efficient incremental dynamic link prediction algorithms in social network," Knowledge-Based Systems, vol. 132, pp. 226–235, 2017.

[11] Y. Li, P. Luo, Z.-p. Fan, K. Chen, and J. Liu, "A utility-based link prediction method in social networks," European Journal of Operational Research, vol. 260, no. 2, pp. 693–705, 2017.

[12] L. Chen, M. Gao, B. Li, W. Liu, and B. Chen, "Detect potential relations by link prediction in multi-relational social networks," Decision Support Systems, vol. 115, pp. 78–91, 2018.

[13] X. Zhu, X. Yang, C. Ying, and G. Wang, "A new classification algorithm recommendation method based on link prediction," Knowledge-Based Systems, vol. 159, pp. 171–185, 2018.

[14] S. Mallek, I. Boukhris, Z. Elouedi, and E. Lefèvre, "Evidential link prediction in social networks based on structural and social information," Journal of Computational Science, vol. 30, pp. 98–107, 2019.

[15] P. K. Sharma, S. Rathore, and J. H. Park, "Multilevel learning based modeling for link prediction and users' consumption preference in online social networks," Future Generation Computer Systems, vol. 93, pp. 952–961, 2019.

[16] W. Yuan, K. He, D. Guan, L. Zhou, and C. Li, "Graph kernel based link prediction for signed social networks," Information Fusion, vol. 46, pp. 1–10, 2019.

[17] S.-Y. Wu, Q. Zhang, C.-Y. Xue, and X.-Y. Liao, "Cold-start link prediction in multi-relational networks based on network dependence analysis," Physica A: Statistical Mechanics and its Applications, vol. 515, pp. 558–565, 2019.

[18] Z. Cui, X. Xu, X. Fei, X. Cai, Y. Cao, W. Zhang, and J. Chen, "Personalized recommendation system based on collaborative filtering for IoT scenarios," IEEE Transactions on Services Computing, vol. 13, no. 4, pp. 685–695, 2020.

[19] N. Nassar, A. Jafar, and Y. Rahhal, "A novel deep multi-criteria collaborative filtering model for recommendation system," Knowledge-Based Systems, vol. 187, Art. no. 104811, 2020.

[20] I. Mashal, O. Alsaryrah, T.-Y. Chung, and F.-C. Yuan, "A multi-criteria analysis for an Internet of Things application recommendation system," Technology in Society, vol. 60, Art. no. 101216, 2020.

[21] L. Berkani, S. Belkacem, M. Ouafi, and A. Guessoum, "Recommendation of users in social networks: A semantic and social based classification approach," Expert Systems, vol. 38, no. 2, Art. no. e12634, 2021.

[22] H. Zarzour, F. Maazouzi, M. Al-Zinati, A. Nusayr, M. Alsmirat, M. Al-Ayyoub, and Y. Jararweh, "Using K-means clustering ensemble to improve the performance in recommender systems," in Proc. 2022 Int. Conf. Intelligent Data Science Technologies and Applications (IDSTA), pp. 176–180, 2022.

[23] J. Yuan, W. Wu, Y. Li, and D. Du, "Active friending in online social networks," in Proc. Fourth IEEE/ACM Int. Conf. Big Data Computing, Applications and Technologies, pp. 139–148, 2017.

[24] W. Yuan, K. He, D. Guan, L. Zhou, and C. Li, "Graph kernel based link prediction for signed social networks," Information Fusion, vol. 46, pp. 1–10, 2019.

[25] S.-Y. Wu, Q. Zhang, C.-Y. Xue, and X.-Y. Liao, "Cold-start link prediction in multi-relational networks based on network dependence analysis," Physica A: Statistical Mechanics and its Applications, vol. 515, pp. 558–565, 2019.

[26] Z. Cui, X. Xu, X. Fei, X. Cai, Y. Cao, W. Zhang, and J. Chen, "Personalized recommendation system based on collaborative filtering for IoT scenarios," IEEE Transactions on Services Computing, vol. 13, no. 4, pp. 685–695, 2020.

[27] N. Nassar, A. Jafar, and Y. Rahhal, “A novel deep multi-criteria collaborative filtering model for recommendation system,” Knowledge-Based Systems, vol. 187, Art. no. 104811, 2020.

[28] I. Mashal, O. Alsaryrah, T.-Y. Chung, and F.-C. Yuan, “A multi-criteria analysis for an Internet of Things application recommendation system,” Technology in Society, vol. 60, Art. no. 101216, 2020.

[29] L. Berkani, S. Belkacem, M. Ouafi, and A. Guessoum, “Recommendation of users in social networks: A semantic and social based classification approach,” Expert Systems, vol. 38, no. 2, Art. no. e12634, 2021.

[30] H. Zarzour, F. Maazouzi, M. Al-Zinati, A. Nusayr, M. Alsmirat, M. Al-Ayyoub, and Y. Jararweh, “Using K-means clustering ensemble to improve the performance in recommender systems,” in Proc. 2022 Int. Conf. Intelligent Data Science Technologies and Applications (IDSTA), pp. 176–180, 2022.