شناسایی گره‌های قدرت در شبکه‌های اجتماعی به کمک داده‌کاوی

الهام مظاهری حسین آبادی[۱]، علیرضا طالب پور [۲*]، ،علی رضاییان[۳]

۱. کارشناس ارشد مدیریت فناوری اطلاعات، دانشکده مدیریت و حسابداری، دانشگاه شهید بهشتی ، تهران،ایران
۲. استادیار دانشکده علوم و مهندسی کامپیوتر دانشگاه شهید بهشتی، تهران،ایران
۳. استاد گروه مدیریت، دانشکده مدیریت و حسابداری دانشگاه شهید بهشتی ، تهران،ایران

دریافت: ۹۳/۱۰/۲۳ پذیرش: ۹٤/۰٤/۰۵

چکیده

با ظهور شبکه‌های اجتماعی، روابط انسان‌ها دراینترنت شکل تازه‌ای به خود گرفته است. امروزه شمار کثیری از کاربران با اهداف مختلف در شبکه‌های اجتماعی عضو شده و به فعالیت‌های گوناگون می‌پردازند. از سوی دیگراین کاربران بر یکدیگر تأثیر می‌گذارند که نقش و تأثیرگذاری کاربران مختلف یکسان نیست. برخی ازاین کاربران به دلایل مختلفی از جمله موقعیت شغلی، میزان تحصیلات یا نوشته‌های جذاب، تأثیر بیشتری بر کاربران دیگر دارند.این کاربران که ما از آن‌ها به عنوان گره‌های قدرت نام می‌بریم می‌توانند نظر و دیدگاه کاربران زیادی را تغییر داده و به سمت و سوی خاصی سوق دهند. هدف‌این پژوهش شناسایی گره‌های قدرت در شبکه اجتماعی تبیان از طریق کشف الگوهای پنهان در خصیصه‌های کاربران است. تحقیق حاضر بر اساس روش‌شناسی استاندارد CRISP-DM انجام شده است. بدین منظور پس از شناسایی گره‌های قدرت بر اساس مجموع امتیازات به دست‌آمده از سه شاخص مرکزیت بینیت، رتبه صفحه و فعالیت‌های کاربران ،با استفاده از الگوریتم‌های درخت تصمیم به طبقه‌بندی‌این کاربران اقدام شده است. نتایج به دست آمده نشان داد که مهم‌ترین ویژگی‌ها در تعیین میزان قدرت یک کاربر در شبکه اجتماعی تبیان تحصیلات، سن و جنسیت می‌باشند.

کلیدواژه‌ها: داده‌کاوی، دسته‌بندی ،شبکه اجتماعی، گره‌های قدرت.

---

* نویسندهٔ مسؤول مقاله: E-mail: talebpour@sbu.ac.ir

۱- مقدمه

ما در عصر شبکه‌ها زندگی می‌کنیم. عصری که در آن شکل‌گیری شبکه‌های گوناگون اجتماعی آنلاین، شیوه‌های ارتباطی و اطلاع‌رسانی نوینی به عرصه گسترده ارتباطات اجتماعی معرفی کرده است. سایت‌های شبکه‌های اجتماعی از محبوب‌ترین مقاصد سال‌های اخیر شده‌اند و تنها در چند دهه اخیر، میلیاردها نفر با استفاده خلاقانه از رسانه‌های اجتماعی و با اتصال به شبکه‌ای جهان‌گستر، زندگی‌شان را تغییر داده‌اند [۱، ص۱۱]. ما از رسانه‌های اجتماعی استفاده می‌کنیم تا دوستان و خانواده‌هایمان را به خود نزدیک‌تر کنیم، به همسایگان و هم‌دانشگاهیانمان دسترسی یابیم، به بازار محصولات و خدمات نیرو بخشیم [۲] و حتی بر محدودیت‌های مکانی و زمانی فائق‌اییم و گاه، مانند آنچه در تحول‌های اخیر خاورمیانه دیده‌ایم، جریان‌های قدرتمند (موافق و مخالف) اجتماعی به راه بیندازیم و از جامعه‌هایمان تصاویری تازه، با ابعادی نوین به جهان عرضه کنیم [۱، ص۱۱].

از اولین سؤال‌هایی که یک نفر هنگام مشاهده یک شبکه اجتماعی می‌پرسد،این است که چه کسی دراین شبکه مهم‌تر است یا چه کسی قدرت دارد؟ بنابراین مهم است که بدانیم درون یک شبکه اجتماعی چه می‌گذرد. نقاط خاصی از شبکه توانایی بیشتری برای اثرگذاری، نشر افکار و اطلاعات دارند و نقاطی هم دور افتاده و راکد هستند [۱، ص۱۷]. بنابراین دریافت محدودیت‌ها و امکانات بالقوه شبکه‌های اجتماعی، چه برای محققان شبکه و چه برای موسسان آن، امری لازم و حیاتی است.

دراین راستا محقق تلاش کرده است که با توجه به اهمیت‌این موضوع پژوهشی را در مورد گره‌های قدرت و کاربران تاثیرگذار در شبکه اجتماعی انجام دهد.

۱-۱- بیان مسئله

از آغاز قرن اخیر، رسانه‌های اجتماعی مجازی که گونه‌ای از وب ۲/۰ محسوب می‌شوند، در رتبه‌بندی پایگاه الکسا[۱]، همواره جزء جذاب‌ترین و پرمخاطب‌ترین گونه‌های «رسانه‌های مبتنی براینترنت[۲]» بوده‌اند [۳].

استقبال فزاینده و پرشتاب کاربران ازاین رسانه‌ها، حتی در میان کاربران‌ایرانی با وجود محدودیت‌های موجود در فضای مجازی، غیر قابل انکار است [٤] . شبکه‌های اجتماعی مجازی

نیز به عنوان زیرمجموعه‌ای از رسانه‌های اجتماعی حجم بسیار بالایی از کاربران را به خود جذب کرده‌اند به طوری که امروزه فعالیت در شبکه‌های اجتماعی به جزء لاینفکی از زندگی بخش عظیمی از کاربران‌اینترنت تبدیل شده است. کاربران دراین سایت‌ها به ارتباط با دیگر کاربران، تولید محتوا و به اشتراک‌گذاری آن می‌پردازند.این امر باعث شده است که اکثر کاربران در تولید محتوا نقش داشته باشند. محتوای کاربران بسته به شخصیت و علایق کاربر در حوزه‌های متفاوتی قرار می‌گیرند [۵].

از طرف دیگر کاربران در شبکه‌های اجتماعی بر یکدیگر تاثیرمی‌گذارند واین تأثیر معمولاً از طریق مشاهده پست‌های یکدیگر رخ می‌دهد. می‌توان تأثیر کاربر را به‌این صورت تعریف کرد: تشویق کاربران دیگر به انجام یک عمل و یا هدایت آن‌ها به سمت و سویی خاص؛ امااین تأثیر برای همه کاربران به یک میزان نبوده و بعضی از کاربران تأثیر بسیار زیادتری نسبت به عموم کاربران دارند. به عبارتی کاربرانی دراین شبکه‌ها وجود دارند که به دلایل مختلفی از جمله موقعیت شغلی، نوع شخصیت، سن، تحصیلات و یا نوشته‌های جذاب، نفوذ و تأثیر بیشتری بر کاربران دیگر می‌گذارند.این کاربران که ما از آن‌ها به عنوان گره‌های قدرت نام می‌بریم می‌توانند کارکردهای زیادی داشته باشند از جمله استخراج دیدگاه عمومی در رابطه با یک مفهوم ، تشخیص سمت و سوی بازار، تبلیغات مؤثر از طریق‌این افراد معتبر، پیش بینی نتایج انتخابات و تأثیر بر آن و همچنین تشخیص افراد کلیدی گروه‌های مخرب.

هدف از انجام‌این پژوهش شناسایی گره‌های قدرت در شبکه اجتماعی تبیان از طریق کشف روابط و الگوهای پنهان در ویژگی‌های کاربران است. پژوهش پیش رو تلاش می‌کند تا به پرسش‌های زیر پاسخ دهد:

گره‌های قدرت در شبکه اجتماعی چگونه با کشف روابط و الگوهای پنهان بین ویژگی‌های کاربران شناسایی می‌گردند؟

کدام‌یک از ویژگی‌های یک کاربر تأثیر بیشتری بر میزان قدرت و تأثیرگذاری او دارد؟

آیا امکان ارائه مدلی برای طبقه‌بندی کاربران شبکه به لحاظ قدرت و تأثیرگذاری با استفاده از تکنیک‌ها و الگوریتم‌های داده‌کاوی وجود دارد؟

آیا استفاده از تکنیک درخت تصمیم c5. 0 طبقه‌بندی صحیح‌تری را نسبت به سایر طبقه‌بندی‌ها حاصل خواهد کرد؟

۲- پیشینه پژوهش

۲-۱- پیشینه نظری

شبکه‌های اجتماعی بسیار قدیمی هستند؛ از زمانی که کسی به کسی کمک کرده است، شبکه‌های اجتماعی، حتی اگر آشکار نبودند، وجود داشته‌اند [۱، ص۱۸]

شبکه‌های اجتماعی: بوید و الیسون شبکه اجتماعی رااین‌گونه تعریف می‌کنند: شبکه‌های اجتماعی، سرویس‌های مبتنی بر وب هستند که به کاربران اجازه می‌دهند:۱. یک نمایه [۳]شخصی یا نیمه شخصی از مشخصات خود بسازند. ۲. فهرستی از کاربران که با آن‌ها در ارتباطاندایجاد کنند. ۳. فهرست پیوند[۴]های خود و کسانی که توسط دیگران در سیستم‌ایجاد شده است را مشاهده و پیمایش کنند [٦] .

قدرت: در کلی‌ترین معنا، قدرت به معنای توان‌ایجاد یا سهم داشتن درایجاد نتایجی است برای پدیدآوردن تفاوتی در جهان. می‌توانیم بگوییم که در زندگی اجتماعی قدرت به معنای توان انجام چنین کاری از طریق روابط اجتماعی است، یعنی توان‌ایجاد یا داشتن سهمی درایجاد نتایج از طریق تأثیر گذاردن بر دیگری یا دیگران [۷] . در تحقیق حاضر قدرت به معنی توانایی تأثیرگذاری بر رفتار دیگران است به گونه‌ای که موجب اتفاق افتادن آن چیزی که ما می‌خواهیم شود.

گره قدرت: گره قدرت در شبکه اجتماعی کاربری است تأثیرگذار که قابلیت نفوذ در رفتار دیگران برای دستیابی به نتایج دلخواه را دارد.

داده کاوی و کشف دانش :تاریخچه کشف دانش از پایگاه‌های اطلاعاتی قدمت چندانی ندارد و امروزه به داده‌کاوی مشهور است. اصطلاح کشف دانش برای نخستین بار در دهه ۱۹۹۰ مطرح شد و توجه پژوهشگران را به سمت الگوریتم‌های داده‌کاوی معطوف کرد. هدف داده‌کاوی، کشف دانش جدید، معتبر و قابل پیگیری با استفاده از ابزارهای هوش مصنوعی و آماری در حجم بالایی از داده‌ها است [۸] . داده‌کاوی استخراج یا اقتباس دانش از مجموعه داده است و به فرایندی گفته می‌شود که دانش را از داده‌ها استخراج می‌کند واین دانش در قالب الگوها و مدل‌ها بیان می‌شود [۹] . در شکل ۱ مراحل کشف دانش از پایگاه داده‌ها نشان داده شده است [۱۰] ،این مراحل به شرح زیر هستند:

۱. انتخاب داده‌ها: داده‌های مربوط به تجزیه و تحلیل و تصمیم‌گیری از داده‌های دیگر جدا

می‌شوند.

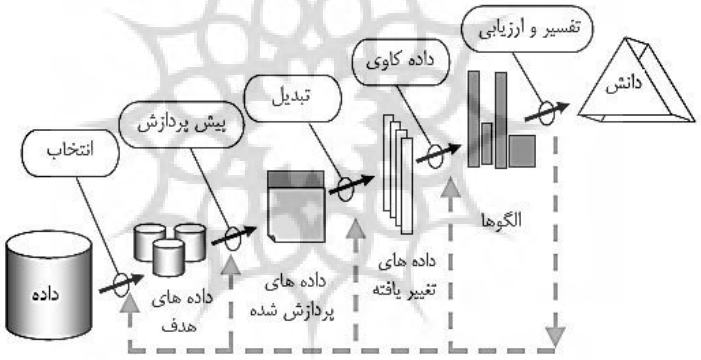
۲. پیش‌پردازش اطلاعات: پردازش، تمیز کردن و یکپارچه سازی داده‌ها انجام می‌گیرد.

۳. تبدیل داده‌ها: داده‌های انتخاب شده به شکل مناسبی برای روش داده‌کاوی تبدیل می‌شوند.

٤. داده‌کاوی: دراین مرحله از روش‌های هوشمندانه‌ای برای استخراج الگوهای بالقوه مفید استفاده می‌شود و در مورداین روش‌ها تصمیم‌گیری می‌شود.

۵. تفسیر و ارزیابی: دراین مرحله، الگوهای جالب توجه نشان‌دهندهٔ دانش، بر اساس اقدامات انجام شده شناخته می‌شوند و دانش کشف شده در اختیار کاربر قرار می‌گیرد.

دراین مرحله استفاده از روش‌های تجسم‌سازی برای کمک به کاربران ضروری است.



شکل۱ مراحل کشف دانش از پایگاه داده‌ها [۱۱]

دو هدف اصلی داده‌کاوی، پیش‌بینی و توصیف است [۱۲]:

۱. داده‌کاوی پیش‌گویی کننده، مدلی را از سیستم ارائه می‌دهد که‌این مدل را مجموعه‌ای از داده‌های مشخص، پیش‌بینی می‌کنند. هدف کلی آن‌ایجاد الگویی برای طبقه‌بندی، پیش‌بینی و تخمین داده‌ها است.

۲. داده‌کاوی توصیفی، اطلاعات جدید و غیربدیهی را بر اساس مجموعه‌ای از داده‌های موجود ارائه می‌دهد و هدف کلی آن درک و شناخت سیستم‌های تجزیه و تحلیل شده با استفاده

از الگوها و روابط موجود است.

شایان ذکر است که داده‌کاوی از ترکیب چندین رشته نشات می‌گیرد. آمار، یادگیری ماشین، روش‌های بهینه‌سازی، روش‌های تشخیص و شناخت الگو، بانک اطلاعاتی، تجسم سازی، شبکه‌های عصبی، مدل‌های ریاضی، بازیابی اطلاعات، الگوریتم ژنتیک و هوش مصنوعی فنونی هستند که داده‌کاوی از آن‌ها بهره می‌برد [۱۳].

۲-۲-پیشینه تجربی

«استفاده از شاخص‌های مرکزیت شبکه‌های اجتماعی برای شناسایی بازیگران پرنفوذ در شبکه‌های اجتماعی مجازی» موضوع پژوهشی است که توسط پارسا و کجباف در سال ۱۳۹۱ انجام شده است. آن‌ها با استفاده از شاخص‌های مرکزیت بینیت، مرکزیت نزدیکی و مرکزیت بردار ویژه به دنبال پاسخ به‌این سؤال بودند که چه کسانی در شبکه اجتماعی مهم‌تر هستند. آن‌ها روش پیشنهادی خود را بر روی دو مجموعه داده استاندارد باشگاه فوتبال و باشگاه کاراته شبیه سازی و اعمال نمودند و نتیجه گرفتند که‌این پارامترها در تعیین نقش یک گره بسیار مؤثر هستند [١٤].

ریگی و همکاران در مقاله‌ای با عنوان«روشی جهت شناسایی بازیگران اصلی در شبکه‌های اجتماعی» در سال ۱۳۹۲ الگوریتمی جهت شناسایی گره‌های اصلی در شبکه‌های اجتماعی ارائه نمودند. بدین منظور از ترکیب معیار مرکزیت نزدیکی و مرکزیت بینیت استفاده کردند. الگوریتم پیشنهادی آن‌ها بر روی داده‌های باشگاه کاراته زاخاری شبیه سازی شده است [۱۵].

بادامچی و همکاران در سال ۱۳۹۰ با ارائه مقاله‌ای با عنوان رتبه‌بندی کاربران بر اساس میزان تأثیرگذاری در شبکه اجتماعی توییتر به بررسی کاربران توییتر ، تشخیص کاربران تأثیرگذار و رتبه‌بندی آن‌ها پرداختند. دراین مقاله میزان تأثیرگذاری کاربران بر اساس چهار پارامتر تعداد دنبال‌کنندگان یک کاربر، نرخ تعداد دنبال شوندگان بر دنبال کنندگان، تعداد توییت‌های باز ارسال شده و تعداد توییت‌های اشاره شده به کاربر محاسبه شده است.این روش بر روی مجموعه داده‌ای مشتمل بر ۱۰۰ توییت اخیر ۱۰۰۰ کاربر مشهور سایت توییتر مورد بررسی قرار گرفته است. بررسی‌ها نشان دادند که تعداد دنبال کنندگان یک کاربر، تنها پارامتر مؤثر در میزان تأثیرگذاری کاربر نمی‌باشد. همچنین مشخص شد تأثیر اشاره شدن به نام کاربر و همچنین میزان بازارسال

شدن توییت‌های او نقش بیشتری در تأثیرگذاری کاربر دارند [۵].

در تحقیقی در سال ۲۰۰۹، کیم و هان روشی را برای یافتن افراد بانفوذ در شبکه‌های اجتماعی مطرح نمودند. دراین روش افراد بر اساس دو فاکتور ارزیابی می‌شوند:۱. تعداد همسایه‌های هر کاربر در شبکه اجتماعی (مرکزیت درجه) و ۲. پیشینه فعالیت هر کاربر شامل تعداد گروه‌هایی که کاربر در آن عضویت دارد، تعداد صفحات لینک شده در صفحه اطلاعات و. .. آن‌ها با استفاده از نظر خبرگان برای هر معیار وزنی تعیین کردند و در نهایت با ترکیب‌این دو فاکتور توسط‌این روش، توانستند بانفوذترین افراد در شبکه‌های اجتماعی برای انجام بازاریابی ویروسی را شناسایی نمایند [١٦].

چا و همکاران در سال ۲۰۱۰ در پژوهشی به اندازه گیری نفوذ کاربران در توییتر پرداختند. آن‌ها برای‌این منظور از سه سنجه زیر استفاده کردند:

۱. درجه ورودی که تعداد دنبال کنندگان یک کاربر است و نشان‌دهنده محبوبیت یک کاربر است.

۲. تعداد توییت‌های باز ارسال شده که ارزش محتویات و مطالب توییت‌های یک فرد را نشان می‌دهد و در واقع نشان دهنده توانایی کاربر برای تولید محتوای با ارزش است.

۳. تعداد توییت‌های اشاره شده به کاربر که ارزش نام کاربر را تعیین کرده و نشان دهنده توانایی کاربر در جذب دیگران برای محاوره و گفتگو است.

نتایج آن‌ها نشان داد کاربران محبوبی که تعداد دنبال‌کنندگان زیادی دارند لزوماً از لحاظ باز ارسال توییت و اشاره‌ها تأثیرگذار نیستند. به علاوه تأثیر خود به خود یا به طور تصادفی به دست نمی‌آید بلکه از طریق تلاش‌های برنامه‌ریزی‌شده از قبیل محدود کردن توییت‌ها به یک موضوع واحد به دست می‌آید [۱۷].

در سال ۲۰۱۲ایریناکی و همکاران در پژوهشی با عنوان «شناسایی افراد بانفوذ در شبکه‌های اجتماعی» به منظور شناسایی کاربران بانفوذ سنجه‌ای با نام profilerank را معرفی کردند که از ویژگی‌های محبوبیت و فعالیت کاربر برای رتبه‌بندی افراد استفاده می‌کرد. آن‌ها یک سری پارامتر برای هر یک ازاین ویژگی‌ها تعریف کردند و در نهایت از مجموع وزنی برخی ازاین پارامترها برای شناسایی کاربران برتر استفاده کردند. آن‌ها نتایج حاصل ازاین روش را با دیگر سنجه‌های تحلیل شبکه اجتماعی از جمله مرکزیت بینیت و رتبه صفحه مقایسه نمودند و در نهایت‌این سه سنجه را با هم ادغام کردند [۱۸] .

در سال ۲۰۱۳ سینگ و همکاران در مقاله‌ای به معرفی روش‌های مختلف برای تعیین کاربران بانفوذ پرداختند. آن‌ها به مرور دو رویکرد کلی شناسایی کاربران بانفوذ پرداختند که اولی مبتنی بر ساختار شبکه اجتماعی و تئوری گراف است و دومین رویکرد مبتنی بر لینک است. آن‌ها رویکرد جدیدی را پیشنهاد کردند که مبتنی بر فعالیت‌هایی است که کاربران در شبکه‌های اجتماعی انجام می‌دهند مثل انتشار مطالب، نظردهی، لایک[۵]، دنبال کردن، باز ارسال توییت‌ها و. .. . [۱۹].

در سال ۲۰۱٤ پژوهشی با عنوان«افراد بانفوذ در شبکه‌های اجتماعی: نمونه‌ای از فیس‌بوک در گرجستان» توسط بابوتسیدز و همکاران انجام شد. آن‌ها یک مجموعه اطلاعات در مورد ویژگی‌های دموگرافیک و رفتار کاربران گرجستان در فیس‌بوک را تجزیه و تحلیل کردند. آن‌ها از ابزارهای تجزیه و تحلیل شبکه اجتماعی به منظور توصیف قدرت و موقعیت افراد در شبکه استفاده کردند. سنجه‌های مرکزیت درجه، مرکزیت بینیت و مرکزیت نزدیکی توسط آن‌ها محاسبه شد. نتایج نشان داد زنان بیش از مردان افراد را تحت تأثیر قرار می‌دهند و نتیجه دیگرآنکه افراد بین ۱۷ تا ۲۵ سال کمتر از سایرین تأثیرگذار هستند که نشان می‌دهد کاربران جوان فیس‌بوک از قدرتشان در شبکه محلی کمتر استفاده می‌کنند [۲۰].

رابادی و همکاران در سال ۲۰۱٤ در پژوهشی با عنوان«بررسی تکنیک‌های شناسایی کاربران بانفوذ در شبکه‌های اجتماعی» به معرفی فنون و الگوریتم‌های مختلف برای شناسایی کاربران فعال پرداختند.این فنون مبتنی بر ساختار شبکه اجتماعی، رتبه‌بندی لینک‌ها، مدل‌های انتشار، انجمن کاوی، محتوا کاوی و همچنین تکنیک‌های شناسایی کاربران بانفوذ در بازاریابی وبلاگی است [۲۱].

با توجه به پیشینه مطرح شده،این پژوهش به ارائه یک چارچوب پیشنهادی نوین برای طبقه‌بندی کاربران شبکه اجتماعی در دو دسته گره قدرت و کاربران عادی به کمک مدل فرایندی CRISP-DM پرداخته و درخت‌های تصمیم‌گیری منتج به انتخاب الگوریتم C5 مقایسه شده‌اند.این امر به کشف الگوهای پنهان موجود در داده‌های مربوط به ویژگی‌های کاربران، به صورت مجموعه‌ای از قوانین کمک می‌کند.

۳- روش‌شناسی پژوهش

پژوهش پیش‌رو از نظر هدف، کاربردی و از منظر نحوه گردآوری داده‌ها از نوع تحقیقات

توصیفی-همبستگی محسوب می‌شود و از منظر قطعیت داده‌ها در زمره تحقیقات اکتشافی به شمار می‌رود. دراین پژوهش برای شناسایی شاخص‌های تاثیرگذار بر میزان قدرت یک کاربر (فیلتر نمودن خصیصه‌ها)، از مطالعه ادبیات تحقیق و مصاحبه آزاد با خبرگان شبکه استفاده شده است. همچنین به منظور تعیین وزن مربوط به هر شاخص از پرسشنامه خبره و فرایند تحلیل سلسله مراتبی (AHP) استفاده شده است. نوع داده‌های‌این پژوهش به صورت عددی(پیوسته) و اسمی(گسسته) است.

دراین پژوهش از مدل استاندارد CRISP-DM استفاده شده است که یک روش استاندارد داده‌کاوی بوده و در اواخر سال ۱۹۹٦، سه شرکت بزرگ دایملر کرایسلر(بنز)، اس. پی. اس. اس.[٦] و ان. سی. آر.[٧] آن راایجاد کرده‌اند[۲۲].

این روش، یک مدل فرایندی برای داده‌کاوی ارائه می‌دهد که مروری بر چرخه عمر هر پروژه داده‌کاوی است، چرخه عمر یک پروژه داده‌کاوی، شامل شش مرحله است: درک مسئلهٔ کسب و کار، درک داده‌ها، آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی نتایج و به‌کارگیری مدل که در شکل ۲ نشان داده شده است [۲۳، صص ۲۲-۳].

شکل۲ چرخه عمر یک پروژه داده‌کاوی [۲۲]

با توجه به‌اینکه برای انجام فرایند پـژوهش، اسـتاندارد جهـانی CRISP-DM مـورد اسـتفاده قرار گرفته است؛ در زیـر سـاختار اجرایـی پـژوهش بـر اسـاس مراحـل‌این اسـتاندارد تشـریح می‌شود:

۱. درک مسئلهٔ کسب و کار[۸]:این مرحله ابتدا بـر درک اهـداف و ملزومـات پـروژه از دیـدگاه کسب و کار متمرکز می‌شود که اصلی‌ترین هدف درایـن تحقیـق، شناسـایی گـره‌هـای قـدرت و شاخص‌های تعیین‌کننده آن در شبکه اجتماعی و در نهایت کشـف الگوهـای پنهـان موجـود بـین خصیصه‌ها و ویژگی‌های کاربران از طریق داده‌کاوی است به نحوی که از طریق‌این ویژگی‌ها و قواعد بتوان به شناسایی و پیش‌بینی وضعیت سایر کـاربران شـبکه پرداخـت. هـدف داده‌کـاوی دراین پژوهش «طبقه‌بندی» و نوع آن «پیش‌بینی کننده» است؛ به‌این نحو کـه ابتـدا بـا تعیـین یـک امتیاز برای هر کاربر و تعریف یک مقدار آستانه، کاربران به کلاس‌های با برچسـب گـره قـدرت (کاربران تأثیرگذار) و کاربر عادی (کاربران غیر تأثیرگذار) تقسیم شده‌اند، سپس دانـش مربـوط به پیش‌بینی وضعیت یک کاربر جدید بر اساس الگوریتم‌های داده‌کاوی استخراج شده است.

۲. درک داده‌ها[۹]:این مرحله با جمع‌آوری اولیه داده‌ها شروع می‌شود و به توصیف داده‌هـا و تعیین کیفیت آن‌ها می‌پردازد. به همین منظور و با توجـه بـه هـدف‌ایـن پـژوهش کـه شناسـایی گره‌های قدرت می‌باشد باید به یکسری از اطلاعات مربوط به کاربران دسترسی داشت تا بتـوان به کمک آن‌ها و با کمک روش داده‌کاوی وضعیت قدرت کاربران را تعیین نمـود. پـس از مطالعـه ادبیات تحقیق، یک‌سری از ویژگی‌ها استخراج گردید. دراین مرحله برای تعیـین فیلـدهای نهـایی، پس از چندین جلسه مصاحبه با کارشناسان شـبکه و توصـیف آنـان از داده‌هـا، فیلـدهای قابـل استفاده مشخص گردید. پس از جمع‌آوری و انتخاب داده‌ها برای شروع عملیـات پـیش‌پـردازش، ابتدا کیفیت داده‌های موجود مورد سنجش قرار گرفت.این ارزیابی به‌این دلیل انجام گرفت کـه در داده‌ها، خصیصه‌های اصلی تاثیرگذار در قدرت کاربران وجود داشته باشد و داده‌های گـم‌شـده یا ناقص در میان داده‌ها، میزان قابل قبولی داشته باشند.

۳. آماده‌سازی داده‌ها[۱۰]: مرحله آماده‌سازی داده‌ها، شامل کلیه فعالیت‌هایی اسـت کـه بـرای ساختن مجموعه داده‌های نهایی (داده‌هایی که برای مدل‌سازی آماده شده‌انـد) از داده‌هـای خـام اولیه به کار می‌رود. هر چه کیفیت‌این آمـاده‌سـازی بهتـر باشـد، مـدل‌سـازی نیـز بهتـر خواهـد بود[۲۳، صص۲۲-۳]. وظایف آماده‌سازی داده‌ها در چند دوره انجام می‌گیرد و هـیچ ترتیـب از

پیش تعریف شده‌ای ندارد.این وظایف شامل انتخاب جداول، رکوردها و خصیصه‌ها و همچنین انتقال و پاک‌سازی داده برای مدل‌سازی است. دراین مرحله، داده‌های مرتبط با کاربران شبکه از پایگاه‌داده‌هایی که در مرحله قبل شناسایی شدند، استخراج و ثبت شدند. سپس داده‌ها در یک پایگاه داده جامع و یکپارچه(البته پالایش نشده) قرار گرفتند. سپس داده‌ها پالایش شده و ساختار موردنظر برای مدل‌سازی روی آن‌ها اعمال شد. برای پاک‌سازی و پیش‌پردازش داده‌ها، دو عملیات کاهش داده‌ها و اعمال تغییرات در شکل داده‌ها، روی پایگاه داده صورت گرفت.

٤. مدل‌سازی[11]: دراین مرحله، انواع روش‌های مدل‌سازی انتخاب شده و به کار گرفته می‌شوند. در کل، برای یک نوع مسئله داده‌کاوی چندین روش وجود دارد. برخی از روش‌ها نیازمند فرمت ویژه‌ای از داده‌ها هستند. بنابراین اغلب لازم است که به مرحله آماده‌سازی داده بازگشت. برای اعمال روش‌ها در پژوهش حاضر، از نرم افزار Clementine12 استفاده شده است. برای مدل‌سازی داده‌ها، به طور جداگانه روی هر الگوریتم، طبقه‌بندی انجام گرفت و الگوریتمی که بالاترین صحت را داشت، مبنای مدل‌سازی و استخراج دانش از آن قرار گرفت. هدف ازاین کار، استخراج دانش با توجه به داده‌های آموزشی از الگوریتم انتخابی است.

٥. ارزیابی نتایج[12] : دراین مرحله از پروژه، مدلی که از دیدگاه تحلیل داده، کیفیت بالایی دارد، ساخته شده است. پیش از اقدام برای به‌کارگیری مدل، باید مدل به طور کلی ارزیابی شده و گام‌های اجرایی برای تطابق با اهداف کسب و کار مرور شوند. دراینجا با توجه به مقایسه‌های صورت گرفته، روشی که از بقیه پیش‌بینی دقیق‌تری را انجام می‌دهد، برای استفاده بهره‌وران پیشنهاد می‌شود.

٦. به‌کارگیری مدل[13]:ایجاد مدل، به معنای پایان پروژه نیست؛ حتی اگر هدف مدل ارتقای دانش از داده‌ها باشد، باز هم دانش حاصل شده به سازماندهی نیاز دارد و باید به شکلی ارائه شود که بهره‌وران بتوانند از آن استفاده کنند. بسته به ملزومات ، فاز به‌کارگیری می‌تواند به سادگی‌ایجاد یک گزارش، یا به پیچیدگی اجرای یک فرایند قابل تکرار داده‌کاوی باشد.

۴- یافته‌های پژوهش

مراحل اجرایی پژوهش به سه قسمت توصیف داده‌ها، تحلیل داده‌ها و ارزیابی مدل، تفکیک

شده‌اند که در ادامه جزییات آن‌ها تشریح شده است.

۴-۱- توصیف داده‌ها

جامعه آماری پژوهش متشکل از داده‌های مربوط به ۳۰۷۱۲ مورد از کـاربران شـبکه اجتمـاعی تبیان است که در بازه یکساله در سال ۱۳۹۲ در شبکه عضویت داشـته و بـه فعالیـت مشـغول بوده‌اند.

با توجه به‌اینکه داده‌های موجود در جامعه آماری از لحاظ کیفیت مورد بررسی قرار گرفتـه و تا حد امکان پالایش شده‌اند، بنابراین مدل‌های پـژوهش بـرای نمونـه نهـایی، فقـط روی ۵۲۹۲ مورد از داده‌های کاربران اجرا شد که‌این داده‌ها در یک پایگاه داده در نرم افـزار اکسـل ذخیـره شده‌اند. رکوردهای حذف شده مربوط به کاربران با اطلاعات ناقص و داده‌های پرت بوده است. هفت خصیصه مربوط به کاربران، متغیرهای مستقل پژوهش و طبقه کاربران (وضـعیت قـدرت کاربر)، متغیر وابسته آن هستند که در جدول ۱ نشان داده شده است.

جدول ۱ عناوین فیلدها

نام فیلد	سن	جنسیت	وضعیت تاهل	استان	شغل	تحصیلات	سن عضویت در شبکه	وضعیت قدرت کاربر
برچسب متغیر در پایگاه داده	age	Sex	Marital Status	State	Job	Education	Ozviat (month)	Power_Node

نوع داده‌های‌این پژوهش از هر دو نوع پیوسته و گسسته است. گفتنی است، برخی داده‌هـای مربوط به متغیرهای عددی، به داده‌های اسمی تبدیل شـده‌انـد و هـر کـدام از متغیرهـای اسـمی دراین داده‌ها، انواع مختلفی دارند که در جدول ۲ مشاهده می‌شود.

جدول۲ انواع متغیرهای اسمی

متغیرهای اسمی	نوع مربوطه
سطح تحصیلات	حوزوی(اجتهاد، خارج، سطح و طلبه)،زیردیپلم(ابتدایی، راهنمایی، سیکل، دبیرستان و کم سواد)،دیپلم، فوق دیپلم، دانشجو، لیسانس، فوق لیسانس، دکتری و بالاتر(دانشجوی دکترا، دکترا و فوق دکترا)
جنسیت	مرد، زن
استان	تهران، جمعیت زیاد(خراسان رضوی، اصفهان، فارس، خوزستان، آذربایجان غربی، آذربایجان شرقی، مازندران، کرمان، سیستان و بلوچستان، گیلان و البرز)، جمعیت متوسط(کرمانشاه، گلستان، همدان، لرستان، هرمزگان، کردستان، مرکزی، اردبیل، قزوین، قم، یزد، بوشهر، زنجان)، جمعیت کم(چهار محال و بختیاری، خراسان شمالی، کهگیلویه و بویر احمد، سمنان، خراسان جنوبی وایلام)
شغل	کارمند-کارشناس(کارمند، کارشناس)آزاد(آزاد، صنعتگر)، پزشک، دانشجو، طلبه، کارگر،خانه‌دار، محصل، مدرس،معلم، مدیر
وضعیت تاهل	مجرد، متاهل
سن	a(زیر ۱۸ سال)، b(۲٦-۱۸)،c (۳۵-۲۷)،d (٤٤-۳٦)، e(٤۵ و بالاتر)

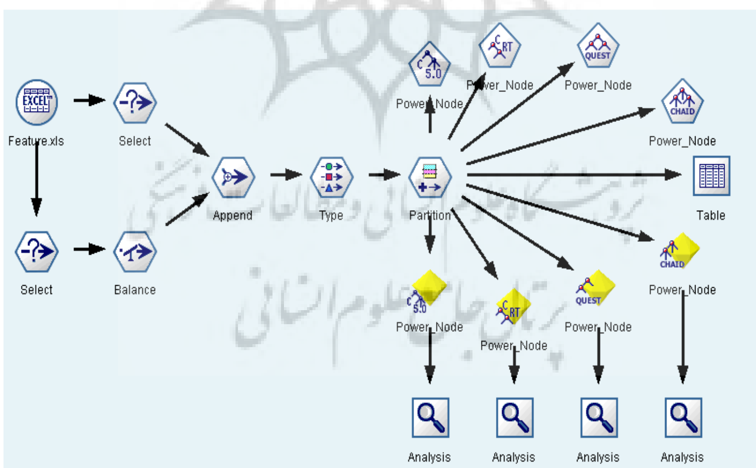
فیلد وضعیت قدرت کاربر مربوط به طبقه کاربران است که بر اساس میزان فعالیت و جایگاهشان در ساختار شبکه به دو وضعیت yes(گره قدرت) و No(کاربر عادی) تفسیر می‌شود. هفت شاخص مربوط به فعالیت کاربران (نظرات دریافتی، لایک دریافتی، پست‌ها[۱٤]، دنبال کنندگان، نظرات کاربر برای دیگران، لایک کاربر برای دیگران و تعداد دوستان) و دو شاخص تحلیل شبکه (مرکزیت بینیت و رتبه صفحه) معیارهای تعیین کننده برای محاسبه امتیاز قدرت کاربر هستند. برای وزن‌دهی به شاخص‌های مربوط به فعالیت کاربران از پرسشنامه خبره و تحلیل سلسله مراتبی و هفت نفر خبره کمک گرفته شد و هر هفت نفر به شاخص‌های مطرح شده وزنی دادند که بعد از تجمیع نظر آنان، هر یک از فعالیت‌های کاربران، وزن مربوطه را گرفتند. هر کاربر بر اساس فعالیت‌هایش و با توجه به وزن مربوطه و امتیاز مرکزیت بینیت و رتبه صفحه امتیازی دریافت کرده است. مقدار صدک ۹۵ معیار قرار گرفتن کاربران در هر یک از طبقات می‌باشد. کاربرانی که امتیاز آنها از مقدار صدک ۹۵ بیشتر باشد در طبقه گره‌های قدرت، و مقادیر کمتر در طبقه کاربران عادی قرار گرفتند.

۲-۴- تحلیل داده‌ها

دراین بخش داده‌ها را مدل‌سازی کرده و تحلیل‌های لازم در خصوص اعتبارسنجی مدل و کارایی آن ارائه می‌شود.

پس از حذف فیلدهای غیرمفید، پایگاه داده نهایی برای طبقه‌بندی کاربران به منظور ورود به نرم افزار، دارای ۵۲۹۲رکورد و هفت فیلد از خصیصه‌های مربوط به کاربران و فیلد وضعیت قدرت کاربر بود. به علاوه با توجه به‌اینکه داده‌های اولیه به شدت نامتوازن بوده است ناگزیر از تکنیک Over_Sampling برای مقادیر صدک ۹۵ استفاده شده است. به منظورایجاد تعادل بین دو طبقه، از یک گره در محیط نرم افزار Clementine به نام Balance استفاده شده است تا به شکل مجازی تعداد رکوردهای با وضعیت گره قدرت در مجموعه آموزشی افزایش یافته در نتیجه تأثیر آن‌ها در ساخت مدل نهایی زیادتر شود. برای‌این منظور ضریب ٤ به رکوردهای دارای وضعیت گره قدرت اختصاص یافت.

برای فراخوانی و اجرای مدل، باید فیلد وضعیت قدرت کاربر را خروجی و بقیه فیلدها را به عنوان ورودی معرفی کرد. الگوریتم‌های اصلی اعمال شده دراین پژوهش، درخت‌های تصمیم بوده و استریم حاصل از مدلسازی در نرم افزار Clementine12 به صورت شکل ۳ است.



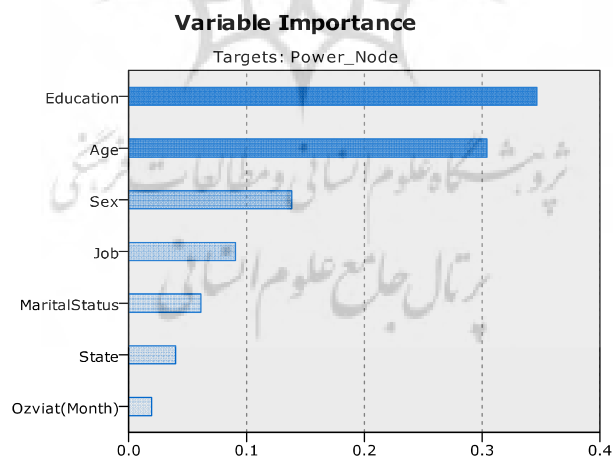
شکل۳ مدل‌ایجاد شده برای طبقه‌بندی کاربران در نرم افزار Spss Clementine

الگوریتم اصلی اعمال شده دراین پژوهش با توجه به مقادیر صحت و دقت، درخت تصمیم C5 بوده است. میزان دقت‌این الگوریتم به لحاظ طبقه‌بندی صحیح کاربران به صورت جدول ۳ است. همان‌گونه که دراین جدول مشخص است، برای داده‌های آموزشی از کل ٤٢٤٠ رکورد، تعداد ۳۹۸۵ رکورد به طور صحیح طبقه‌بندی شدند و تعداد ۲۵۵ رکورد طبقه‌بندی نادرستی داشتند. برای داده‌های آزمایشی، از کل ۱۸۵۰رکورد، تعداد ۱۷۱۷ رکورد طبقه‌بندی درست و تعداد ۱۳۳ رکورد نادرست طبقه‌بندی شدند.

جدول۳ دقت الگوریتم درخت تصمیم C5. 0

	داده‌های آموزشی		داده‌های آزمایشی	
داده‌های طبقه‌بندی‌شده به طور صحیح	۳۹۸۵	۹۳/۹۹ درصد	۱۷۱۷	۹۲/۸۱ درصد
داده‌های طبقه‌بندی‌شده به طور غلط	۲۵۵	٦/۰۱ درصد	۱۳۳	۷/۱۹ درصد
کل نمونه‌ها	٤٢٤٠		۱۸۵۰	

الگوریتم C5 اهمیت خصیصه‌های کاربران را به صورت شکل ٤ اولویت‌بندی می‌کند.



شکل۴ اهمیت شاخص‌ها

۴-۳- ارزیابی مدل

از آنجا که باید اعتبار روش ارائه شده در هر پژوهشی مورد سنجش قرار گیرد، برای سنجش اعتبار و صحت مدل، از تقسیم نمونه به دو مجموعهٔ دادهٔ آموزشی و آزمایشی استفاده شده است. میزان اعتبار با نتایج داده‌های جدید آزمون می‌شود و داده‌های آزمایشی به منزلهٔ ناظر به الگوریتم وارد شده و نتایج میزان صحت آن را ارزیابی می‌کند. در این پژوهش، ۷۰ درصد از داده‌ها مجموعهٔ داده‌های آموزشی و ۳۰ درصد نیز، مجموعه داده‌های آزمایشی انتخاب شده‌اند.

معیار اعتبار و صحت مدل، بسته به صحت طبقه بندی یا تفکیک داده‌های آزمایشی است. دراین پژوهش از اعتبارسنجی متقابل با ده تکرار استفاده شده است؛ به این معنا که مجموعهٔ داده‌ها به ده قسمت تقسیم می‌شود و هر بار ۷۰ درصد از مجموعهٔ داده‌های آموزشی و ۳۰ درصد از مجموعهٔ داده‌های آزمایشی انتخاب شده و دقت طبقه بندی سنجیده می‌شود، پس از اینکه فرایند ده بار تکرار شد، دقت نهایی مدل ارائه می‌شود.

برای سنجش کارایی مدل ارائه شده درخت تصمیم c5، صحت نتایج‌این مدل با الگوریتم‌های طبقه‌بندی دیگری چون شبکه عصبی، درخت تصمیم CHAID و C&R Tree وQUEST مورد مقایسه قرار گرفت که نتایج آن در جدول ٤ نمایش داده شده است.

جدول۴ مقایسه تطبیقی صحت الگوریتم‌ها

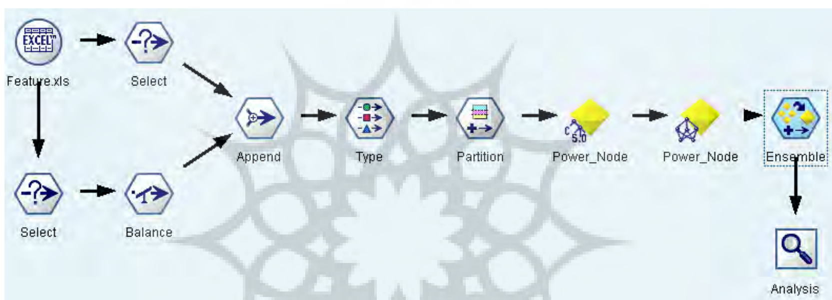
الگوریتم	C5. 0	Neural net	C&R Tree	QUEST	CHAID
صحت	۹۲/۸۱۱	۹۳/۰۸۱	۹۱/۵۱٤	۸٦/٥٤۱	۹۰/۸۱۱

همانطور که در جدول ٤ مشاهده می‌شود در بین درخت‌ها، درخت تصمیم‌گیری c5 دارای دقت بالاتری نسبت به بقیه درختان است. عدم انتخاب شبکه عصبی علیرغم داشتن مقدار صحت کل بالاتر، به دلیل عدم ارائه قانون توسط‌این الگوریتم بوده است. با توجه به‌اینکه در تحقیق حاضر به دنبال شناسایی عوامل مؤثر در میزان قدرت یک کاربر و کشف روابط پنهان به صورت قوانین اگر–آنگاه هستیم، الگوریتم درخت تصمیم c5. 0 گزینه مناسب‌تری تشخیص داده شده است.

با توجه به‌اینکه تفاوت معناداری میان مقادیر صحت الگوریتم C5. 0 و شبکه عصبی وجود

نداشته و تا حد بسیار زیادی مطابقت دارند مدل از لحاظ پایداری نتایج قابل تأیید می‌باشد.

الگوریتم استخراج شده قابلیت طبقه‌بندی رکوردهای جدید کاربران را داشته و می‌توان به میزان صحت آن در پیش‌بینی وضعیت قدرت کاربران اطمینان کرد. برای پیش‌بینی وضعیت کاربران جدید از ترکیب مدل c5 و شبکه عصبی به کمک گره انسمبل استفاده شده است که استریم مربوط به آن در شکل ٥ نشان داده شده است. لازم به ذکر است صحت حاصل از ترکیب درخت C5. 0 و شبکه‌های عصبی مصنوعی ۹۳/۱۹ درصد می‌باشد که نسبت به درخت c5. 0 عملکری بهتر را فراهم می‌آورد.



شکل٥ ترکیب الگوریتم درخت تصمیم C5. 0 با شبکه‌های عصبی

٥-نتیجه‌گیری و پیشنهادها

داده‌کاوی ابزار مهمی برای استفاده سودمند از داده‌ها به شمار می‌رود و یکی از مهم‌ترین فناوری‌ها برای بهره‌برداری موثر و دقیق از داده‌های حجیم است. شناسایی کاربران تاثیرگذار در شبکه اجتماعی یکی از جنبه‌های تحقیقاتی مورد توجه در سالیان اخیر است و دراین راستا پژوهش پیش رو از الگوریتم‌های طبقه‌بندی و ساخت درختان تصمیم‌گیری استفاده کرده است.

طبق بهترین دانش ما تاکنون رویکرد داده‌کاوی دراین حوزه مورد استفاده قرار نگرفته است. روش پیشنهادی با تعیین و اولویت‌بندی یک سری از شاخص‌ها، مدلی را برای شناسایی افراد تأثیرگذار ارائه می‌کند.

از آنجا که درخت تصمیم c5 بهینگی و دقت بالاتری نسبت به کلیه الگوریتم‌های طبقه‌بندی

داده‌کاوی مورد استفاده در پژوهش دارد، بنابراین دانش استخراج شده ازاین درخت، مورد اعتمادترین دانش حاصل از داده‌های مورد بررسی به شمار رفته و می‌تواند مبنای استخراج قوانین داده‌کاوی باشد. مهم‌ترین شاخص‌ها دراین الگوریتم، تحصیلات، سن و جنسیت است و شاخص‌های شغل، وضعیت تاهل، استان و مدت زمان عضویت به ترتیب از اهمیت کمتری برخوردارند.

در بین رکوردهای موجود در پایگاه داده قوانین زیادی به دست می‌آید که تعدادی از آن‌ها که اطلاعات بهتر و کاربردی‌تری را ارائه می‌دهند را در جدول زیر آورده‌ایم. قوانین به دست آمده نتیجه به‌کارگیری الگوریتم‌های هوشمند بوده و تفسیراین قوانین بر اساس اصول و مبانی نظری در حوزه تخصص خبرگان شبکه اجتماعی و اساتید مربوطه می‌باشد. همچنین‌این قوانین از پایگاه داده در دسترس ما استخراج شده و مسلماً قابل تعمیم به سایر موارد نمی‌باشد.

جدول ۵ قوانین مربوط به گره‌های قدرت (Power_Node=Yes)

ردیف	قانون	تعداد نمونه‌ها	میزان اطمینان
۱	اگر سن= "b "و جنسیت="مرد" و تحصیلات="حوزوی" و وضعیت تأهل="متأهل" آنگاه کاربر گره قدرت است.	٤٦	۰/۸۰٦
۲	اگر سن= "b" و جنسیت="مرد" و تحصیلات="حوزوی" و وضعیت تأهل="مجرد" و استان="تهران" آنگاه کاربر گره قدرت است.	٢٤	۰/۸۱٤
۳	اگر سن= "c"و جنسیت="مرد" و تحصیلات="لیسانس" و شغل= "دانشجو" و وضعیت تأهل="مجرد" و استان="تهران و جمعیت زیاد" آنگاه کاربر گره قدرت است.	۱۵۷	۰/۸۱۵
٤	اگر سن="c" و جنسیت="مرد" و تحصیلات="فوق‌لیسانس" و شغل="دانشجو و کارمند-کارشناس" آنگاه کاربر گره قدرت است.	۱۷۰	۰/۸۷٦

با توجه به جمع‌بندی نظرات خبرگان از قواعد نهایی منتج شده دیدگاه‌هایی به شرح زیر حاصل گشت:

-گره‌های قدرت اکثراً در رده سنی ۲٦-۱۸ و ۳۵-۲۷ قرار دارند.

-سطح تحصیلات گره‌های قدرت حوزوی، لیسانس و فوق‌لیسانس می‌باشد.

– در شبکه اجتماعی تبیان مردان تأثیرگذاری بسیار بیشتری دارند.

– شغل گره‌های قدرت در اکثر موارد کارمند، کارشناس، دانشجو و یا طلبه است.

با در نظر گرفتن محدوده پژوهش، موارد زیر برای مطالعات بعدی پیشنهاد می‌شود:

۱. دراین پژوهش با توجه به امکانات خاص شبکه تبیان، محدودیت زمانی برای انجام پژوهش و در نهایت نظر خبرگان، ۷ مورد از فعالیت‌های کاربران به عنوان شاخص‌های مناسب در تعیین امتیاز قدرت کاربر معرفی گردید. حال می‌توان با اضافه کردن شاخص یا شاخص‌های دیگر اطلاعات و نتایج جالبی پیرامون وضعیت گره‌های قدرت به دست آورد.

۲. پیشنهاد می‌شود پژوهش حاضر و روش‌های به کارگرفته شده در آن در سایر شبکه‌های اجتماعی پیاده‌سازی شده و نتایج آن‌ها با هم مقایسه شود.

۳. پیشنهاد می‌شود در پژوهش‌های آتی از ترکیب روش خوشه‌بندی و دسته‌بندی برای شناسایی و دسته‌بندی کاربران شبکه و گره‌های قدرت استفاده گردد.

٤. پیشنهاد می‌شود در پژوهش‌های‌اینده به ویژگی‌های محتوا از قبیل محتوای کاربر، موضوع آن، جنس محتوا از لحاظ متن، تصویر و. .. و شرایط زمانی ارائه محتوا به عنوان عوامل مهم در تأثیرگذاری بر مخاطب توجه شود.

۶- پی‌نوشت‌ها

1. www. alexa. com
2. Internet Based Media
3. Profile
4. connection
5. like
6. SPSS
7. NCR
8. Business Understanding
9. Data Understanding
10. Data Preparation
11. Modeling
12. Evaluation
13. Deployment
14. Posts

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

۷- منابع

[1] Mir Mohammad Sadeghi, M., *An Analysis of Social Networks by Node XL*, Kiyan Rayaneh, Tehran, 2012.

[2] Hansen, D., Shneiderman, B., & Smith, M. A., Analyzing the social media networks with NodeXL: Insights from a connected world, Morgan Kaufmann, 2010, p. 4.

[3] Khaniki, H. & Basirian Jahromi, H., "Effectiveness and power in virtual social networks " A study on the functions of Facebook in real world", *Quarterly Journal of Social Sciences*, 61, 2013, pp. 45-81.

[4] Kosari, M., *The Cultural World of Iranian Users in Friendship Networks*, Institute of Culture, Arts and Communication, Tehran, 2007, p. 20.

[5] Badamchi, A., Fani, M. R. & Shirigheidari, S., "The rating of users according to their effectiveness in Twitter, as a social network", *Paper Presented at the first National Conference on Researchers of Computer and InformationTechnology*, University of Tabriz, 2011.

[6] Boyd, D. M., & Ellison, N. B., "Social Network Sites: Definition, History, and Scholarship", *Journal of Computer-Mediated Communication, 13*(1), 2007, pp. 210-230.

[7] Outhwaite, W. & Batamur, T., *The Culture of Social Sciences in the Twentieth Century*, Transletad by Chavoshiyan, H. Ney Publisher, Tehran, 2013.

[8] Marban, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C., "Toward data mining engineering: A software engineering approach", *Information Systems, 34*(1), 2009, pp. 87-107.

[9] Taghvafard, M. T. & Nadali, A., "Classification of the bank applicants looking for a beneficial credit, by analysis and Fuzzy logic", *Journal of Industrial Management Studies*, 52(9), 2012, pp. 52-91.

[10] Deshpande, M. S., & Thakare, D. V., "Data mining system and applications: A review", *International Journal of Distributed and Parallel Systems (IJDPS),*

1(1), 2010, pp. 32-44.

[11] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P., "From data mining to knowledge discovery in databases", *AI Magazine, 17*(3), 1996, p. 37.

[12] Kiss, F., Credit scoring processes from a knowledge management perspective, *Social and Management Sciences, 11*(1), 2003, pp. 95-110.

[13] Mahmoudi, K., Rostami, H., Sayebani, M. & Moradi, A., "An overview of the analytical science and its functions in offshore industry", *Paper Presented at the Fifth National Conference on Offshore Industries*, 21-22 May. Sharif Industry University, 2013.

[14] Parsa, S. & Kajbaf, M., "The use of centrality parameters of social networks for recognition of affective users in virtual social networks", *Paper Presented at the Second International Conference on Software Engineering Branch*, 31 October-1 November, Lahijan, Lahijan Islamic Azad University, 2012.

[15] Reegi, R., Jalali, M. & Moatar, M. H., "A method for recognition of main users in social networks", *Paper Presented at the First National Conference on Computer Engineering and Information Technology*, 13 March. Shoshtar. Young Researchers Club, Unit Shoshtar, 2013.

[16] Kim, E. S., & Han, S. S., "An analytical way to find influencers on social networks and validate their effects in disseminating social games", Paper presented at International Conference on Advances the Social Network Analysis and Mining, 2009. ASONAM'09, 2009.

[17] Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K., Measuring User Influence in Twitter: The Million Follower Fallacy. ICWSM, 10, 2010, pp.10-17.

[18] Eirinaki, M., Monga, S. P. S., & Sundaram, S., "Identification of influential social networkers", International Journal of Web Based Communities, 8(2), 2012, 136-158.

[19] Singh, S., Mishra, N., & Sharma, S., "*Survey of various techniques for determining influential users in social networks*", Paper presented at International

Conference on the Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN), 2013.

[20] Babutsidze, Z., Lomitashvili, T., & Turmanidze, K., "Influential individuals on social networks: An example of facebook In Georgia", *European Scientific Journal, 9*(10), 2014.

[21] Rabade, R., Mishra, N., & Sharma, S., Survey of influential user identification techniques in online social networks *Recent Advances in Intelligent Informatics,* Springer, 2014, pp. 359-370.

[22] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R., CRISP-DM 1. 0 Step-by-step Data Mining Guide, 2000.

[23] Azar, A., Ahmadi, P. & Cebt, M., "An analytical approach to devising a placement model for recruiting employees (the employment of applicants participating in the entrance examination of a trading bank in Iran)", *Journal of Information Technology Management*, 2(4), 2010, pp. 3-22.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی