

## تشخیص سطح علمی دانشجویان با بررسی کلیک روزانه در سامانه‌های آموزش الکترونیکی

حدیث عزیزی\*

دانشجوی ارشد مهندسی نرم‌افزار، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج

پست الکترونیکی: hadis.azizi@uok.ac.ir

صادق سلیمانی

استادیار، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج

پست الکترونیکی: s.sulaimany@uok.ac.ir

پرهم مرادی

دانشیار، دانشکده مهندسی کامپیوتر، دانشگاه کردستان، سنندج

پست الکترونیکی: p.moradi@uok.ac.ir

### چکیده

داده‌های پژوهش از سامانه آموزشی متن باز OULAD که حاوی اطلاعات تعامل دانشجویان با سامانه است، انتخاب شده است. نتایج بیانگر این مفهوم است که بالاتر بودن مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه متمایز کننده دانشجو ممتاز از ضعیف و تایید کننده عدم شباهت رفتاری دانشجویان براساس میانگین کلیک‌هایشان در سامانه است. میانگین دقت این معیارها با الگوریتم‌های درخت تصمیم، جنگل تصادفی و KNN و ... بعد از اجرا در ده لایه صورت گرفته که الگوریتم درخت تصمیم با ۷۸ درصد بهترین میانگین دقت تفکیک کنندگی را داشته است.

**واژه‌های کلیدی:** داده‌های جریان کلیک، داده‌های سری زمانی آموزش الکترونیکی، گراف پدیداری طبیعی، تحلیل گراف پدیداری، شبکه‌های پیچیده، تحلیل شبکه

داده‌های جریان کلیک سامانه‌های آموزش الکترونیکی اهمیت زیادی در رابطه با تشخیص سطح علمی دانشجویان آن سامانه‌ها دارد. بیشتر مطالعات قبلی صورت گرفته بر داده‌های جریان کلیک، مبتنی بر الگوریتم‌های یادگیری ماشین بوده است. با توجه به اهمیت الگوریتم‌های گراف پدیداری و کارایی مطلوب آن‌ها در تحلیل داده‌های سری زمانی و تولید آورده‌های جدید محاسباتی، کارایی مطلوب در پیش‌بینی و تشخیص قابلیت‌های متمایز دو سری زمانی مختلف، با نداشتن آن به حیطه داده‌های آموزشی نیز می‌توان از این پتانسیل‌ها بهره برد. هدف این پژوهش تحلیل میانگین کلیک روزانه دانشجویان ممتاز و ضعیف و تبدیل به گراف پدیداری طبیعی و بررسی با معیارهای تحلیل شبکه پیچیده مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه برای تشخیص تفاوت رفتاری آن‌ها است.

\* نویسنده مسئول

جریان کلیک، مستعد استفاده از روش‌هایی است که می‌تواند آن را به فضای تحلیل گراف نگاشت کند و سبب ایجاد ارزش افزوده و انجام تحلیل‌های جدید شود. این مهم با کمک گراف پدیداری میسر است و تاکنون مورد بررسی قرار نگرفته است. پس از تبدیل داده‌های سری زمانی جریان کلیک به گراف پدیداری، می‌توان از تحلیل گراف پدیداری به منظورهای مختلف بهره برد. مزیت استفاده از روش‌های تحلیلی مرتبط با الگوریتم‌های گراف پدیداری این است که: تفکیک و طبقه‌بندی نمونه‌ها بدون وابستگی به مفروضات از قبل انجام می‌شود. همچنین تحلیل گراف پدیداری به کمک معیارهای تحلیل شبکه پیچیده نظیر توزیع درجه، میانگین درجه و... سبب انتخاب ویژگی‌هایی مرتبط شده که بیشترین تاثیر را در تفکیک دو دسته یا دو سری زمانی دارند و این کار را با دقت مطلوب و نمایش موثر انجام می‌دهد [۴]. همچنین الگوریتم‌های گراف پدیداری تحلیل بصری موثری ارائه می‌دهد و این تحلیل‌ها نسبت به سایر روش‌های یادگیری ماشین قابل تفسیرتر هستند، زیرا نمایشی بصری از داده‌ها و روابط بین موجودیت‌ها ارائه می‌دهند. همچنین این الگوریتم‌ها در کشف الگوها و روابط پنهان در داده‌های جریان کلیک که ممکن است برای الگوریتم‌های یادگیری ماشین آشکار نباشند، بهتر عمل می‌کنند.

روش‌های تحلیل گراف پدیداری تاکنون در حیطه‌های مختلف علمی مانند پزشکی، اقتصاد، معماری، مدیریت آب و غیره مورد بهره‌برداری قرار گرفته است [۵].

در این مقاله هدف آن است که ایده تحلیل گراف پدیداری در داده‌های جریان کلیک آموزشی به صورت روزانه مورد استفاده قرار گیرد. بدین منظور ابتدا با در اختیار داشتن اطلاعات داده‌های آموزش الکترونیکی و جریان کلیک مرتبط با کاربران، نسبت به پالایش و استخراج جریان کلیک مطلوب اقدام خواهد شد. سپس دو سری زمانی مربوط به الگوی کلیک دانشجویان ممتاز و ضعیف ایجاد می‌شود و پس از تبدیل به گراف پدیداری طبیعی، از دیدگاه تحلیل

تحلیل داده‌های جریان کلیک جمع‌آوری شده از سامانه‌های آموزشی برای تصمیم‌گیران و مدیران حیطه یادگیری الکترونیکی، اهمیت برجسته‌ای داشته است. آموزش مدرن مجازی به‌طور فزاینده‌ای بر تحلیل داده‌های این سامانه‌ها برای کمک به تدریس و نمره‌دهی بهتر، مدیریت محتوای یادگیری، ارائه تابلوهای بحث و گفتگو مفیدتر، تسهیل کار گروهی، یا به‌طور کلی جایگزینی محیط کلاسی سنتی متکی است. داده‌های جریان کلیک سامانه‌های آموزش الکترونیکی تاکنون توسط الگوریتم‌ها و روش‌های متنوعی در حیطه داده‌کاوی آموزشی مانند الگوریتم‌های یادگیری ماشین مورد مطالعه قرار گرفته است. خروجی تحلیل این بررسی‌ها، باعث نتایج مفیدی جهت سازمان‌دهی مجدد و بگانه‌های آموزشی و ارائه اصلاحات و پیشنهادهای برای بهبود فرایند آموزش و کشف الگوهای پنهان در این داده‌ها بوده است [۱]. روش‌های مبتنی بر الگوریتم‌های یادگیری ماشین کار تفکیک نمونه‌ها را با تکیه بر مفروضات از قبل انجام داده‌اند و عملکرد این مدل‌ها به اندازه نمونه‌های آموزشی وابسته بوده است. بنابراین زمانی که دانشجویان تعاملی با سامانه نداشته یا در فعالیت‌های یادگیری شرکت نکردند، عملکرد این مدل‌ها برای پیش‌بینی نرخ ترک تحصیل کاهش یافته است، زیرا این نتیجه برای پیش‌بینی دانش‌آموزانی که برای کنجکاوی در دوره‌های برخط ثبت نام کرده بودند، رایج بوده است [۲]. همچنین در مطالعات اخیر به بررسی روش‌هایی پرداخته شده است که در آن‌ها به صورت خودکار ویژگی‌های زمانی را از داده‌های خام استخراج کرده‌اند و سپس، با استفاده از روش‌های یادگیری عمیق همانند توابع هم‌آمیخت<sup>۱</sup>، برای استخراج ویژگی‌ها اقدام شده است. علیرغم سودمندی این روش‌ها، نمایش‌های تولید شده اغلب بسیار پراکنده بوده و برای مقایسه رفتار دانشجویان نسبت به همدیگر مطلوب نبوده است [۳]. از طرفی ماهیت سری زمانی داده‌های

1- Clickstream Data  
2- Convolution

گراف پدیداری مورد بررسی قرار می‌گیرد. سپس نتایج تفکیک‌پذیری حاصل شده از تحلیل گراف پدیداری توسط معیارهای مختلف ارزیابی و صحت‌سنجی می‌شود.

ساختار مقاله به صورت کلی به این شرح است: ابتدا در فصل ۲، مروری بر تحلیل داده‌های سری زمانی آموزش الکترونیکی کرده و به زمینه‌های کاربردی مطرح شده در زمینه تحلیل گراف پدیداری پرداخته می‌شود. سپس در بخش ۳ به شرح مجموعه داده و معرفی روش پژوهش اشاره می‌شود و مفهوم گراف پدیداری را شرح داده و معیارهای تحلیل شبکه‌های پیچیده و شبه‌کدهای مراحل اجرا و ارزیابی الگوریتم معرفی می‌شود. سپس به تحلیل پیچیدگی زمانی هر الگوریتم به صورت دقیق پرداخته می‌شود. در بخش ۴ نتایج تحلیل گراف پدیداری ارائه می‌شود. نتایج موجود برای دانشجویان ممتاز و دانشجویان ضعیف با پارامترهای تحلیل شبکه پیچیده و الگوریتم‌های دسته‌بندی گوناگون مقایسه و تحلیل می‌شوند. درگام آخر در بخش ۵ به نتیجه‌گیری و کارهای آینده اشاره خواهد شد.

## ۲. پیشینه پژوهش

### ۲-۱ مروری بر تحلیل داده‌های سری زمانی آموزش الکترونیکی

تاکنون پژوهش‌های متنوعی در زمینه تحلیل داده‌های آموزشی مانند مسئله پیش‌بینی عملکرد دانش‌آموزان با استفاده از یک مدل حافظه کوتاه‌مدت عمیق<sup>۳</sup> با استفاده از مجموعه داده‌های تجزیه و تحلیل یادگیری دانشگاه صورت گرفته است [۶]. در پژوهشی دیگر در سال ۲۰۲۱ یک مدل پیش‌بینی پیشنهاد شده که با الگوریتم‌های مختلف یادگیری ماشین و یادگیری عمیق آموزش داده می‌شود تا رفتار یادگیری دانش‌آموزان را با توجه به متغیرهای مطالعه آن‌ها مشخص کند، تا مشکلاتی را که دانش‌آموزان در معرض خطر با آن‌ها مواجه هستند، تجزیه و تحلیل کند، و متعاقباً، مربیان با مداخله به موقع، مشکل را تسهیل

کنند [۷]. در مطالعه بعدی با تحلیل داده‌های جریان کلیک برای سنجش، درک و حمایت از یادگیری خودتنظیمی<sup>۴</sup> در دوره‌های برخط استفاده می‌شود، این پژوهش از داده‌های جریان کلیک جمع‌آوری شده از یک سیستم مدیریت یادگیری، برای اندازه‌گیری دو جنبه یادگیری خودتنظیمی استفاده می‌کند: مدیریت زمان و تنظیم تلاش. که درک فرآیند یادگیری خودتنظیمی و شناسایی دانش‌آموزانی که ممکن است با شرکت در کلاس‌های برخط به خوبی به آن‌ها خدمت نکنند را راهنمایی می‌کند [۸]. در مطالعه‌ای در سال ۲۰۲۰ به تحلیل بصری داده‌های جریان کلیک ویدیویی پرداخته شد و پیش‌بینی عملکرد فراگیران با استفاده از مدل‌های یادگیری عمیق در دوره‌های MOOC<sup>۵</sup> بررسی شد [۹]. در پژوهشی دیگر در سال ۲۰۲۲ با تحلیل داده‌های جریان کلیک توسط الگوریتم‌های یادگیری ماشین، به بررسی الگوهای رفتاری دانشجویان مرتبط با موفقیت و شکست آن‌ها در وظایف تعاملی پرداخته شده است [۱۰].

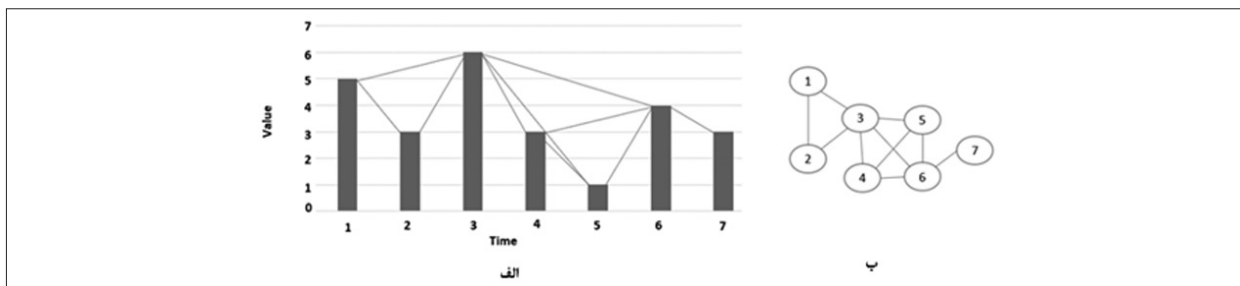
با توجه به بررسی مطالعات پیشین بیشتر مطالعات صورت گرفته در زمینه آموزش الکترونیکی در موضوعی مانند پیش‌بینی عملکرد دانشجویان از روش‌ها و مدل‌های مبتنی بر الگوریتم‌های یادگیری ماشین استفاده کرده‌اند که بعضاً این روش‌ها قادر بوده‌اند با دقت بالایی عملکرد دانشجویان را پیش‌بینی کنند، اما چالش‌هایی که با آن مواجه بوده‌اند این بوده که این روش‌ها الگوها و روابط درون داده‌های سری زمانی را با تکیه بر مدل‌ها یا مفروضات از قبل شناسایی می‌کنند و عموماً تفسیر آن‌ها دشوارتر است. از طرفی دیگر به ویژه در مجموعه داده‌های بزرگ، الگوریتم‌های یادگیری ماشین ممکن است زمان پردازش بالاتری نیاز داشته باشند تا نتایج و تحلیل کارآمدی حاصل کنند.

لازم به ذکر است در این پژوهش تفکیک دانشجویان ممتاز و ضعیف براساس میانگین کلیک روزانه و با استفاده از تحلیل گراف پدیداری ساده انجام می‌شود.

4- Self-Regulated Learning

5- Massive Open Online Course (MOOC)

3- Long Short-Term Memory



شکل ۱: الف) نمودار میله‌ای سری زمانی (نقاط زمانی بر محور افقی، گره‌های گراف هستند و خطوط متصل کننده مقادیر، یال‌های گراف هستند که با شرط صدق کردن در رابطه ۱ ترسیم شده‌اند) ب)

از همدیگر با وجود میله میانی است. به زبان ریاضی شرط ایجاد اتصال (یال) بین دو سرمیله  $(t_i, y_i)$  و  $(t_j, y_j)$  با وجود میله میانی  $(t_n, y_n)$  صدق کردن در معادله (۱) است [۱۱].

$$y_n < y_j + (y_i - y_j) \frac{t_j - t_n}{t_j - t_i} \quad (1)$$

### ب) گراف حاصل از نگاشت سری زمانی

گراف رسم شده در شکل (۱) گراف پدیداری طبیعی بدون جهت، بدون وزن و همبند است. انواع دیگری از گراف پدیداری برحسب روش ساخت آن وجود دارد که می‌توان گراف پدیداری افقی [۱۲] و یا گراف پدیداری نفوذپذیر محدود را نام برد [۱۳]. در این پژوهش از گراف پدیداری طبیعی استفاده خواهد شد.

تحلیل گراف پدیداری به‌عنوان ابزاری مفید برای تحلیل سری‌های زمانی در زمینه‌های متنوع پژوهشی به کار می‌رود و نتایج مفیدی در حیطه‌های مختلفی مانند پزشکی، زیست‌شناسی، اقتصادی، زمین‌شناسی، مدیریت آب و پردازش تصویر و غیره تولید می‌کند. برخی از خروجی‌های تحلیل این نوع بررسی بازده بازار با راهکارهای راهبردی از طریق تحلیل ارزش‌های دیجیتالی مثل تحلیل سوابق قیمت بیت‌کوین جهت درک تعمیق این بازار پرداخته شده است [۱۴]. استفاده از تحلیل گراف پدیداری با پارامتر PSVG<sup>۶</sup>، برای داده‌های سری زمانی ضربان قلب برای افراد مبتلا به CHF<sup>۷</sup> و افراد سالم، نشان می‌دهد که گراف پدیداری

تمایز و برتری روش‌های تحلیل مبتنی بر الگوریتم‌های گراف پدیداری این است که:

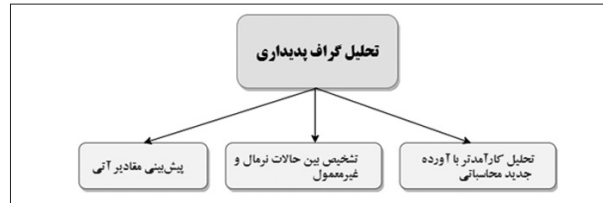
- این الگوریتم‌ها در کشف الگوها و روابط پنهان در داده‌های جریان کلیک که ممکن است برای الگوریتم‌های یادگیری ماشین آشکار نباشند، بهتر هستند زیرا به مفروضاتی از قبیل وابستگی ندارند. تجزیه و تحلیل مبتنی بر الگوریتم‌های گراف پدیداری یک رویکرد ناپارامتریک است، به این معنی که به هیچ فرضی در مورد توزیع اساسی داده‌ها نیاز ندارد. لذا باعث می‌شود که نسبت به برخی از الگوریتم‌های یادگیری ماشین که نیاز به توزیع‌های خاصی دارند، انعطاف‌پذیرتر باشد.

- روش‌های تجزیه و تحلیل مبتنی بر الگوریتم‌های گراف پدیداری عموماً به‌صورت بدون ناظر هستند و لذا برخلاف الگوریتم‌های یادگیری ماشین نیاز به یادگیری ندارند در نتیجه زمان کمتری برای انجام آن‌ها طول می‌کشد و زودتر به نتیجه می‌رسند.

### ۲-۲- مروری بر گراف پدیداری و تحلیل آن

الگوریتم‌های گراف پدیداری ساده نخستین بار توسط لاکاسا و همکارانش در سال ۲۰۰۸ معرفی شد. گراف طبیعی یا ساده با جایگذاری نقاط متوالی سری زمانی بر روی محور زمان که هر نقطه نیز دارای مقدار است که به‌صورت میله‌های عمودی نشان داده می‌شود، مدل می‌شود. چنانچه اگر سر این میله‌های عمودی از هم قابل رویت باشد (میله‌ای بین خط دید مستقیم آن‌ها حایل نباشد) یک یال بین آن دو نقطه گراف رسم می‌شود. به عبارت دیگر شرط رسم یال، قابل مشاهده بودن سر میله‌ها

6- The Power of Scale-freeness in Visibility Graph (PSVG)  
7- Congestive heart failure (CHF)



شکل ۲: دلایل استفاده از تحلیل گراف پدیداری

افراد سالم به‌صورت قابل توجهی PSVG بالاتری نسبت به گراف پدیداری افراد سالم دارا هستند و می‌تواند به‌عنوان یک ابزار بالینی جدید و مفید برای طبقه‌بندی خطر حمله قلبی عمل کند [۱۵].

به‌صورت کلی با توجه به شکل (۲) می‌توان از تحلیل گراف پدیداری به سه منظور استفاده کرد:

منظور از تحلیل کارآمدتر آن است که با توجه به نگاه متفاوت گراف پدیداری به داده‌های سری زمانی و برخی مفاهیم مانند مرکزیت، کشف اجتماع، میانگین کوتاه‌ترین طول مسیر، توزیع درجه، جهان کوچک و ... ممکن است کمک کند که موارد جدیدی در تحلیل به چشم بیاید و یا نکات اضافی علاوه بر روش‌های کنونی در تحلیل ارائه دهد و شناخت بهتری از سری‌های زمانی ارائه دهد. به‌عنوان نمونه داده‌های سری زمانی قیمت بیت‌کوین را می‌توان با روش متفاوت مدل نمود که سبب ایجاد نماهای جدیدی و شناخت بیشتری از آن‌ها می‌شود [۱۴]. اما در تشخیص با دو دسته از سری‌های زمانی مواجه هستیم: سری زمانی مرتبط با حالت معمول (نرمال) و سری زمانی حالت غیر معمول. مثلاً در بیماری‌ها با سری زمانی مرتبط با علائم حیاتی فرد سالم و سری زمانی مرتبط با علائم حیاتی فرد بیمار مواجه هستیم، که تحلیل گراف پدیداری به ما کمک می‌کند که بدون نیاز به روش‌های یادگیری ماشین و روش‌های پیچیده‌تر تمایز و تشخیص بهتری بین این دو دسته قائل شویم [۴]. در نهایت پیش‌بینی مرتبط با تحلیل گراف پدیداری به ما کمک می‌کند به شیوه‌های جدیدی نسبت به یافتن مقادیر جدید در سری‌های زمانی مانند سری زمانی آب و هوا، سری زمانی مرتبط با گسترش بیماری و ... دست پیدا کنیم و راهکارهای مفیدتری برای

مقادیر پیش‌بینی شده در گام‌های بعدی ارائه کنیم [۱۶][۱۷]. همان‌طور که پیش‌تر دلایل و نحوه استفاده از تحلیل گراف پدیداری در سایر حوزه‌ها بیان گردید، اکنون این پرسش مطرح می‌شود که چرا باید از تحلیل گراف پدیداری در حوزه آموزش الکترونیکی استفاده کرد؟ چه ارزش افزوده و مزیتی در این انتخاب است؟ در پاسخ می‌توان اشاره کرد که تحلیل گراف پدیداری به‌عنوان ابزاری قدرتمند برای تحلیل سری زمانی جریان کلیک دانشجویان به کار بسته شده و قادر به تولید آورده محاسباتی جدید در حوزه آموزشی با استفاده از معیارهای مرکزیت و توزیع درجه بوده است. از طرفی دیگر این تنها مزیت استفاده از این تحلیل نبوده، زیرا تحلیل گراف پدیداری در مدت زمان کمتر و بدون نیاز حتمی به استفاده از الگوریتم‌ها و روش‌های پیچیده و مدل‌های یادگیری ماشین که نیاز به یادگیری دارند، قادر است براساس الگوی کلیک روزانه در سامانه آموزش برخط بین دانشجویان ممتاز و ضعیف تفکیک مطلوب ایجاد کند.

### ۳- روش تحقیق

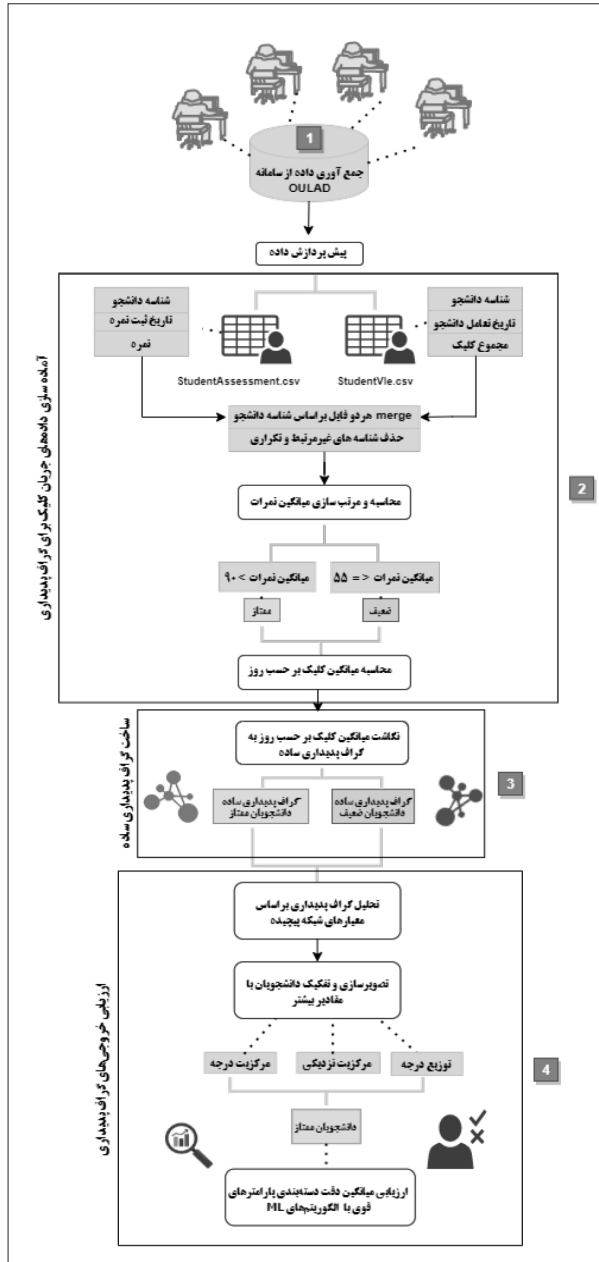
با توجه به بررسی مقالات مرتبط با روش‌های تحلیل داده‌های سری زمانی آموزش الکترونیکی و اهمیت بالای تحلیل داده‌های جریان کلیک، مشخص است براساس بررسی‌های ما، تاکنون دید شبکه‌ای به این نوع داده‌ها صورت نگرفته است. نوآوری این پژوهش عبارت است از نگاشت داده‌های جریان کلیک آموزش الکترونیکی به گراف پدیداری برای نخستین بار و تحلیل الگوی میانگین کلیک دانشجویان ممتاز و ضعیف براساس پارامترهای متمایز کننده تحلیل شبکه‌های پیچیده است. در ادامه به جزئیات داده‌ها و روش پژوهش پرداخته می‌شود.

#### ۳-۱ روش پژوهش و شرح مجموعه داده

جزئیات گام به گام روند انجام پژوهش در شکل (۳) به‌صورت دقیق‌تر در ۴ گام کلی توصیف شده است. در گام ۱ ابتدا داده‌های سری زمانی از سامانه آموزش

الکترونیکی OULAD جمع‌آوری شده و داده‌های مرتبط با پژوهش را ذخیره کرده و در گام ۲ روی آن‌ها پیش‌پردازش انجام شده است. به این منظور ابتدا تمام سطرها را جدا کرده و میانگین نمرات دانشجویان را محاسبه کرده سپس به فایل دیگر studentVle، مراجعه کرده که حاوی تعاملات دانشجویان با سامانه برخط است، و الگوی کلیک دانشجویانی که معدل‌شان را محاسبه کردیم، پیدا کرده و آن دسته از دانشجویانی که الگوی کلیک‌شان موجود نبود، حذف شده‌اند. همچنین شناسه‌های تکراری و غیر مرتبط نیز حذف شدند.

سپس دو سری زمانی از دانشجویان براساس معدل محاسبه شده تشکیل دادیم و دانشجویانی که میانگین نمراتشان بالاتر از ۹۰ بود را به‌عنوان سری دانشجویان ممتاز و دانشجویانی که میانگین نمراتشان کوچک‌تر یا مساوی ۵۵ بود را به‌عنوان سری دانشجویان ضعیف در نظر گرفتیم و میانگین کلیک دانشجویان مرحله قبل را بر حسب روز محاسبه کردیم. در گام ۳ گراف پدیداری طبیعی میانگین کلیک بر حسب روز را برای هر دو سری دانشجویان برتر و ضعیف رسم کردیم. در گام ۴ گراف پدیداری حاصل شده برای هر دوسری شبکه پیچیده عنوان شده مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه تحلیل کردیم. نتایج تحلیل دو گراف پدیداری طبیعی حاصل شده براساس پارامترهای متمایز کننده در این تحلیل مانند مرکزیت درجه، مرکزیت نزدیکی، توزیع درجه که مشخص کننده دانشجویان ممتاز است، را با الگوریتم‌های دسته‌بندی یادگیری ماشین مانند تحلیل جداساز خطی<sup>۸</sup>، درخت تصمیم<sup>۹</sup>، جنگل تصادفی<sup>۱۰</sup> و KNN<sup>۱۱</sup>، بیز ساده، ماشین بردار پشتیبان، پرسپترون چندلایه و رگرسیون منطقی پس از اجرا با 10 fold validation cross ارزیابی کردیم که در نهایت دقت دسته‌بندی و جداسازی



شکل ۳: فرایند گام به گام تمایز دانشجویان ممتاز از ضعیف با کمک نگاهت الگوی جریان کلیک به گراف پدیداری و استفاده از تحلیل گراف پدیداری

دانشجو ممتاز از ضعیف را با دقت مطلوبی نتیجه گرفتیم. در این پژوهش از مجموعه داده متن باز OULAD<sup>۱۲</sup> که شامل داده‌هایی درباره دوره‌ها، دانشجویان و تعامل آن‌ها با محیط یادگیری مجازی برای هفت دوره انتخابی (موسوم به ماژول) است، استفاده کردیم. از فایل studentAssessment که شامل نتایج ارزیابی دانشجویان است استفاده کردیم و ستون‌های مرتبط آن مانند شناسه

12- [https://analyse.kmi.open.ac.uk/open\\_dataset](https://analyse.kmi.open.ac.uk/open_dataset)

8- Linear Discriminant Analysis (LDA)

9- Decision Tree

10- Random Forest

11- K Nearest Neighbor (KNN)

جدول ۱: شرح مجموعه داده انتخاب شده

فایل انتخابی از مجموعه داده	شرح فایل	ستون انتخابی
OULAD		
studentAssessment	حاوی نتایج ارزیابی دانشجویان	id_student, date_submitted, score
studentVle	حاوی اطلاعات تعامل دانشجویان با سامانه	id_student, date, sum_click

کلیک‌شان بر حسب تعداد روزهایی که از شروع ارائه مازول اندازه‌گیری می‌شود را ساختیم.

در واقع سری زمانی از میانگین کلیک این دو سری دانشجویان از ۲۴ روز مانده به شروع کلاس‌های برخط، تا آخرین روز کلیک‌شان در سامانه برخط تشکیل داده شده و معیارهای تحلیل شبکه را برای دو گراف پدیداری ساخته شده تحلیل و بررسی کردیم.

### ۳-۲ شبهه کد مراحل اجرای الگوریتم

در این بخش شبهه‌کد مراحل اجرا و ارزیابی الگوریتم در شکل‌های (۴) و (۵) بررسی شده است. قابل توجه است که فرض بر این بوده که مراحل ۱ و ۲ اجرای الگوریتم که در شکل (۳) آمده است و شامل جمع‌آوری و پیش‌پردازش داده‌ها است، انجام شده و اکنون میانگین کلیک بر حسب روز به‌عنوان ورودی مناسب برای گراف پدیداری آماده شده است. به این منظور ابتدا در شکل (۴) شیوه ساخت دوگراف پدیداری طبیعی بیان شده است. سپس، در شکل (۵) معیارهای قوی تفکیک‌کننده معرفی شده توسط تحلیل گراف پدیداری توسط الگوریتم‌های دسته‌بندی یادگیری ماشین مورد دقت سنجی قرار گرفته است.

همان‌طور که در شکل (۴) مشخص است، ابتدا در ورودی الگوریتم NVG دو آرایه دو بعدی از میانگین کلیک دانشجویان ممتاز و ضعیف بر حسب تعداد روزهایی که از شروع ارائه مازول گذشته است دریافت شده است. سپس به ازای تعداد روزهایی که از شروع ارائه مازول برای هر دو گروه دانشجویان گذشته است، شرط ساخت گراف پدیداری طبیعی اعمال می‌شود. شیوه ساخت دو گراف پدیداری طبیعی دانشجویان ممتاز و ضعیف از خطوط ۱ تا ۹ بیان شده است، با این تعریف که به ازای هر دو نقطه زمانی مشخص  $(t_i, \gamma_i)$  و  $(t_j, \gamma_j)$  با وجود نقطه میانی  $(t_n, \gamma_n)$  که هر کدام از این نقاط زمانی دارای مقدار نیز هستند، در صورتی یالی بین این دو نقطه زمانی رسم می‌شود که ارتفاع نقطه میانی  $(t_n, \gamma_n)$  بین این دو نقطه زمانی در دستور خط ۴ صدق کند. در غیر این صورت یالی نیز رسم نمی‌شود.

دانشجو، تعداد روزهایی که از شروع ارائه مازول می‌گذرد و نمره‌ای که تا آن روز مشخص برای آن دانشجو خاص ثبت شده است را استخراج کردیم. سپس از فایل studentVle که حاوی اطلاعاتی در مورد تعاملات هر دانشجو با گزینه‌های موجود در محیط یادگیری مجازی است، استفاده کردیم و ستون‌های مرتبط آن مانند شناسه دانشجو، تاریخ تعامل دانشجو با وبگاه آموزشی و مجموع کلیک‌های هر دانشجو که تا آن روز مشخص ثبت شده است را در نظر گرفتیم. شروع دوره برخط از روز صفر است، همچنین قابل توجه است که تاریخ تعامل و کلیک دانشجو با وبگاه آموزشی قبل از شروع دوره آموزشی نیز ثبت شده است. یعنی تعداد روزهایی که به شروع دوره برخط زمان باقی مانده است اما دانشجو وارد سامانه شده و کلیک کرده است که با تاریخ منفی در مجموعه داده ذخیره شده است.

بر اساس جدول (۱)، ابتدا از فایل studentAssessment تمام سطرها را جدا کرده و نمرات دانشجویان و زمان فعال در سامانه، را در نظر گرفته و میانگین نمرات دانشجویانی که الگوی کلیک‌شان در فایل studentVle موجود بود را محاسبه کردیم. با استفاده از میانگین نمرات هر دانشجو، دانشجویان را دسته‌بندی کرده و از بزرگ به کوچک مرتب کردیم. سپس دانشجویان ممتاز را با فرض میانگین نمرات بالاتر از ۹۰ و دانشجویان ضعیف را با فرض میانگین نمرات پایین‌تر از ۵۵ به‌عنوان دانشجویان ضعیف در نظر گرفتیم، سپس برای هر دو سری دانشجویان ممتاز و ضعیف گراف پدیداری طبیعی‌شان را به ازای میانگین

**#Algorithm 2: The pseudo-code of the proposed S\_NVG\_ACC method**

**Input:**

# Strong differentiating parameters for classifying excellent students from weak ones  
D [X, Y]: X (arrays of inputs with m features), Y(array of class labels)  
Y = array(C) // Class label

**Output:**

Average accuracy of strong discriminating parameters for each Classifier

**Begin:**

```
1: Avg_accuracy_list ← []
2: Classifier = ['DT', 'NB', 'RF', 'SVM', 'KNN', 'MLP', 'LDA', 'LR']
3: Split D [X, Y] into train and test

#Each time select one of the Classifier for classifying inputs
4: While (Classifier is not empty):
    5: Select one Classifier each time
    6: Fit the Classifier with train and test data
    7: Evaluate the fitted Classifier with 10-fold cross validation
    8: Calculate Avg_accuracy
    9: Append (Avg_accuracy_list, Avg_accuracy)
10: end While
11: Return Avg_accuracy_list
```

End

شکل ۵: شبه کد ارزیابی میانگین دقت پارامترهای قوی تفکیک کننده با الگوریتم های دسته بندی

است که خروجی های تحلیل گراف پدیداری با دقت مطلوبی قادر بودند کار دسته بندی دانشجویان ممتاز از ضعیف را به درستی انجام دهند.

### ۳-۳- تحلیل پیچیدگی زمانی

مراحل اجرای ساخت دو گراف پدیداری دانشجویان ممتاز و ضعیف و ارزیابی میانگین دقت پارامترهای قوی تفکیک کننده معرفی شده توسط تحلیل گراف پدیداری در شکل های (۴) و (۵) آورده شده است. الگوریتم موجود در شکل (۴) دارای پیچیدگی زمانی از مرتبه  $O(n^2)$  بوده است، زیرا این الگوریتم دو حلقه تو در تو برای پیمایش تعداد روزهای شروع شده از ارائه ماژول برای ایجاد دو گراف داشته است.

پیچیدگی زمانی الگوریتم S\_NVG\_ACC موجود در شکل (۵) که میانگین دقت پارامترهای قوی تفکیک کننده را مورد بررسی قرار داده است، به خاطر وابستگی به اجرای

**#Algorithm 1: The pseudocode of the proposed NVG method**

**Input:**

excellent\_students [m][n] # Mean\_click per day of excellent\_students  
weak\_students [m1][n1] # Mean\_click per day of weak\_students  
m = Mean\_click\_excellent  
n = day\_excellent  
m1 = Mean\_click\_weak  
n1 = day\_weak

**Output:**

G (V,E)\_weak : Natural visibility graph extracted from mean click per day of weak stu  
G (V,E)\_excellent : Natural visibility graph extracted from mean click per day of excellent stu

**Begin:**

```
# Creating the visibility graphs of excellent and weak timeseries
1: For (i=1 to n) #until last day of click
2:   For (j=1 to n1)
3:     For each two time nodes (ti,yi) and (tj,yj) with (ta,ya) between them
4:       If (ya < yi+yj)*(tj-ta) / (tj-ti) then draw an Edge between (ti,tj)
5:       else break
6:     End If
7:   End For
8: End For
9: End For
```

End

شکل ۴: شبه کد ساخت گراف های پدیداری دانشجویان ممتاز و ضعیف

در ادامه به ارزیابی میانگین دقت پارامترهای قوی تفکیک کننده در شکل (۵) با استفاده از الگوریتم S\_NVG\_ACC پرداخته شده است. همان طور که مشخص است، ابتدا آرایه هایی از مقادیر پارامترهای قوی تفکیک کننده همانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه با تعداد ویژگی مشخص به ازای برچسب کلاس های ضعیف و قوی به ورودی الگوریتم های دسته بندی مانند درخت تصمیم، بیز ساده، جنگل تصادفی، ماشین بردار پشتیبان، K نزدیک ترین همسایه، پرسپترون چند لایه، تحلیل جداساز خطی و رگرسیون منطقی داده شده است. سپس، به ازای تمامی الگوریتم های دسته بندی و تمامی پارامترهای قوی تفکیک کننده، میانگین دقت دسته بندی هر الگوریتم با مقادیر ورودی تعریف شده با استفاده از روش اعتبارسنجی متقاطع در ۱۰ fold بررسی شده است و در نهایت نیز میانگین دقت گزارش شده است. اهمیت ارزیابی میانگین دقت پارامترهای قوی تفکیک کننده، این مفهوم



الگوریتم‌های دسته‌بندی بسیار متفاوت‌تر از الگوریتم قبلی بوده است. زیرا پیچیدگی زمانی کل الگوریتم  $S\_NVG\_ACC$  وابسته به پیچیدگی زمانی الگوریتم‌های دسته‌بندی در حالت میانگین و بدترین بوده است. از طرفی هر کدام از این الگوریتم‌های دسته‌بندی در بدنه یک حلقه `while` اجرا شده‌اند، بنابراین پیچیدگی زمانی این الگوریتم‌ها چه در حالت میانگین و چه در حالت بدترین، در یک حلقه از مرتبه  $O(n)$  ضرب شده است. قابل ذکر است که پیچیدگی زمانی الگوریتم  $S\_NVG\_ACC$  با احتساب اجرا در بدنه `while` در وضعیت میانگین حالت، حداقل از مرتبه  $O(n^2)$  بوده است. همچنین پیچیدگی زمانی همین الگوریتم در وضعیت بدترین حالت حداقل از مرتبه  $O(n^3)$  بوده است.

### ۳-۴- معیارهای تحلیل شبکه پیچیده

بعد از ساخت گراف‌های پدیداری براساس میانگین کلیک‌های دو دسته دانشجوی ممتاز و ضعیف بر حسب تعداد روزهایی که از شروع ارائه مازول می‌گذرد، مقایسه توسط معیارهای مرکزیت درجه، مرکزیت نزدیکی، توزیع درجه انجام خواهد شد و خروجی این تحلیل‌ها با جزییات بیشتر بررسی و مصورسازی می‌شود. انتخاب این معیارها به ترتیب براساس کارکرد مشابه در حیطه‌های متفاوت غیرآموزشی [۱۸]، [۱۹]، [۲۰] بوده است.

#### ۳-۴-۱ مرکزیت درجه

در یک گراف مرکزیت درجه<sup>۱۴</sup> با تعداد کل پیوندهای مستقیم با سایر گره‌ها اندازه‌گیری می‌شود. طوری که در رابطه (۲) داریم:

$$C'd(Ni) = \frac{\sum_{j=1}^n x_{ij}}{(n-1)(n-2)} \quad (i \neq j) \quad (2)$$

$C'd$  به معنی تعداد پیوندهایی است که مستقیماً به گره  $N$  متصل می‌شوند و  $n$  به معنای تعداد کل گره‌ها در شبکه کانونی است [۲۱].

#### ۳-۴-۲ مرکزیت نزدیکی

مرکزیت نزدیکی<sup>۱۵</sup>، به معنای اندازه‌گیری فاصله‌های مجموع یک گره نسبت به گره‌های دیگر است، اگر طول کوتاه‌ترین مسیرهای گره  $N$  با سایر گره‌های شبکه کوچک باشد، گره  $N$  مرکزیت نزدیکی بالایی دارد. این نشان دهنده راحتی و سهولت اتصالات بین گره متمرکز و سایر گره‌ها است. همان‌طور که در رابطه (۳) مشخص است:  $\sum_{j=a}^n d(Ni, Nj)$  به معنای تعداد کل گام‌ها از گره  $N$  به سایر گره‌های شبکه است [۲۱].

$$C'c(Ni) = \frac{n-1}{\sum_{j=a}^n d(Ni, Nj)} \quad (i \neq j) \quad (3)$$

#### ۳-۴-۳ توزیع درجه

در شبکه‌های پیچیده توزیع درجه<sup>۱۶</sup> معیار مهمی برای تحلیل انواع شبکه است. احتمال توزیع درجه گره با درجه  $K$  به صورت  $PK$  نشان داده می‌شود و به صورت رابطه (۴) تعریف می‌شود:

$$P(K) = \frac{N_k}{n} \quad (4)$$

که در آن  $N_k$  تعداد گره‌های با درجه  $k$  است و  $n$  تعداد کل گره‌ها است [۲۲].

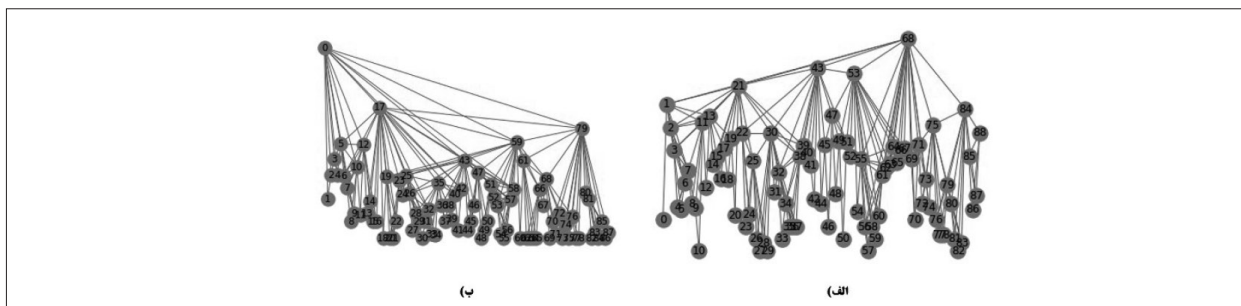
### ۴- نتایج و بحث

در این قسمت تلاش خواهد شد تا یافته‌های پژوهش از دیدگاه‌های مختلف عرضه و تحلیل شود. تصویر کلی گراف با تعداد گره کمتر بر حسب شکل (۶) است و آمارهای اولیه این گراف‌ها طبق جدول (۲) می‌باشد:

در این قسمت گراف پدیداری دانشجویان ممتاز و ضعیف را براساس میانگین کلیک‌شان بر حسب روز در سامانه آموزش برخط OULAD ساختیم و معیارهای تحلیل شبکه که در جدول (۲) قسمت الف و ب نشان داده شده است، را تحلیل و بررسی کردیم. تحلیل گراف پدیداری طبیعی میانگین کلیک دانشجویان ممتاز نسبت به

14- Closeness Centrality  
15- Degree Distribution

13- Degree Centrality



شکل ۶: گراف پدیداری طبیعی حاصل از نگاهت سری زمانی میانگین کلیک دانشجویان بر حسب روز (الف) دانشجویان ممتاز (ب) دانشجویان ضعیف

جدول ۲: (ب) ادامه شاخصه‌های اصلی شبکه

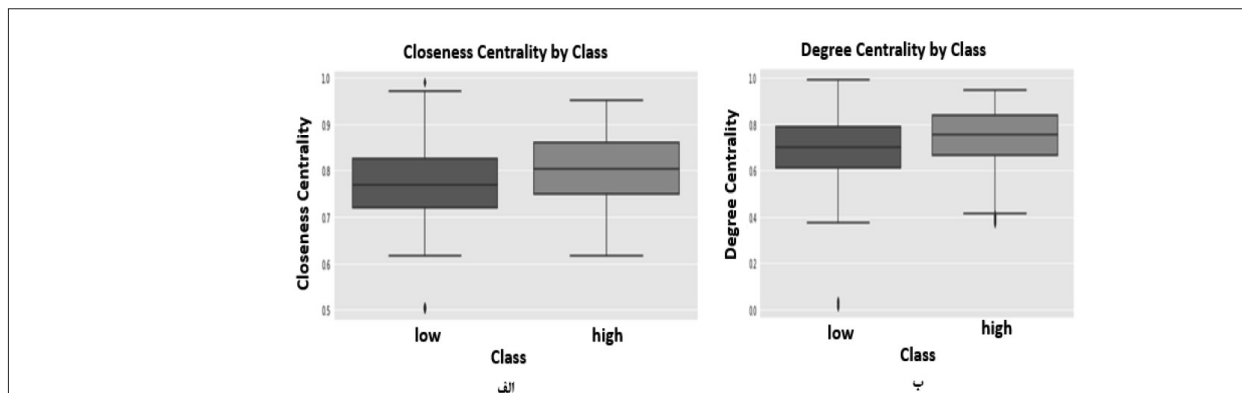
Student Group	Density	Shannon Entropy
High_level	0.73	6.76
Low_level	0.66	6.60
Student Group	Average Clustering Coefficient	Average Shortest Path Length
High_level	0.8132	1.26
Low_level	0.8130	1.33

جدول ۲: (الف) شاخصه‌های اصلی شبکه ایجاد شده براساس گراف پدیداری میانگین کلیک دانشجویان ضعیف و ممتاز

Student Group	Node Count	Edge Count	Mean Degree
High_level	289	30386	210.28
Low_level	292	28119	192.59
Student Group	Diameter	Max Degree	Transitivity
High_level	2	273	0.82
Low_level	3	288	0.81

باشد که دانشجویان ضعیف زمانشان را برای کلیک بر روی قسمت‌هایی از سامانه سپری می‌کنند که در ارزیابی نهایی نمرات آن‌ها خیلی تاثیرگذار نیست. به عنوان مثال این دسته دانشجویان کلیک‌هایشان را در بخش‌های مرتبط با تغییر عکس پروفایل، ویرایش نام کاربری و سایر اطلاعات شخصی مانند بررسی پیام‌های داخل تالار گفتگو خود و غیره انجام می‌دهند. همچنین از آنجایی که هر دو گروه دانشجویان میانگین نمرات آن‌ها متفاوت است، ضریب شباهت جاکارد میانگین کلیک دانشجویان ضعیف و ممتاز، عدد نزدیک به صفر یعنی  $0/019$  است، یعنی این‌که این دانشجویان علاوه بر این که نمرات مشابهی نداشتند، رفتارهای مشابهی از نظر میانگین کلیک در سامانه نیز نداشته‌اند و گرنه در صورت شباهت بالای این دو گروه بایستی ضریب شباهت جاکارد عددی نزدیک به ۱ می‌شد. همان‌طور در شکل (۷) قسمت الف مشاهده می‌شود، مرکزیت نزدیکی دانشجویان با high که در سمت راست مشخص است، بیانگر آن است که گراف دانشجویان ممتاز دارای مقادیر بالاتری نسبت به دانشجویان ضعیف است. یعنی این که دانشجویان سطح بالا و پایین‌الگویی

دانشجویان ضعیف، مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه بالاتری را نشان می‌دهد. که در شکل‌های (۷) و (۸) این اختلاف نشان داده شده است. به نظر می‌رسد دانشجویان ضعیف در تعامل با سامانه آموزش الکترونیکی کم‌دقت‌تر و کم‌حوصله‌تر هستند و با تمرکز کمتری کلاس را دنبال می‌کنند و تمایلی به بررسی تمرینات بارگذاری شده در طول کلاس برخط ندارند و بیشتر تمرکزشان بر ثبت حضور در سامانه و گاه رها کردن کلاس برخط بدون توجه به مقررات است. همان‌طور که در جدول (۲) قسمت الف ستون چهارم مشخص است، گراف پدیداری طبیعی دانشجویان ممتاز دارای میانگین درجه بالاتری می‌باشد، چندین دلیل می‌تواند در شکل‌گیری این نتیجه موثر باشد. با احتمال زیاد الگوی رفتاری کلیک دانشجویان ممتاز سریع است یعنی دانشجویان سطح بالا با مشاهده تمرینات درسی و پیدا کردن مسیر محتوای آموزشی مانند مشاهده ویدیوها و بارگذاری پاسخ تکالیف، فکر می‌کنند و سپس سریع کلیک می‌کنند و می‌توان الگوی رفتاری آن‌ها را تحلیلی و سرعت کلیک آن‌ها را سریع به شمار آورد. دلیل دیگر می‌تواند آن



شکل ۷: الف) مرکزیت نزدیکی گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف  
 ب) مرکزیت درجه گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف

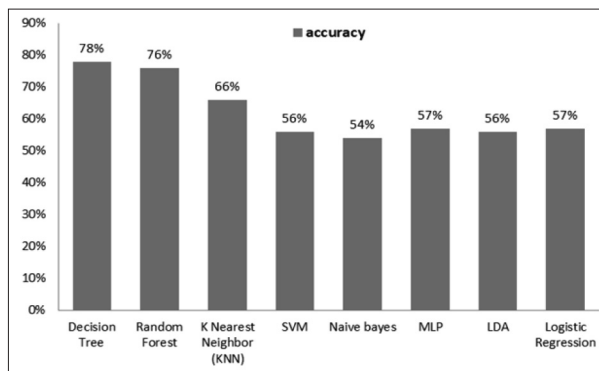
اکثر حالات از میانگین دانشجویان ضعیف بیشتر کلیک کرده‌اند شاید به خاطر این‌که دانشجویان ممتاز سازگاری بیشتری با محتوای درسی سخت بارگذاری شده دارند و زودتر تکالیف و روند دستورالعمل‌های موجود در سامانه آموزشی مانند تماشای ویدیوها، بازکردن مسیر فایل‌ها و محتوای آموزشی و نرم‌افزارهای کمک آموزشی را درک کرده‌اند. در حالی که دانشجویان ضعیف‌تر تمایل بیشتری به رد کردن صفحات کلاس برخط داشته‌اند. همچنین بیشتر به سمت عدم طی کردن گام به گام مراحل کار و نصب مثلاً یک نرم‌افزار آموزشی جدید و چالش برانگیز بوده‌اند. از طرفی عدم بررسی گزینه‌های جدید و فایل‌های کمک درسی و دکمه‌های مربوط با این مسئله خود سبب تعداد کمتر کلیک توسط دانشجویان ضعیف بوده است.

در نمودار (۸) همان‌طور که مشخص است، نحوه عملکرد میانگین کلیک هر دو گروه دانشجویان برای درجه‌های بین ۱۴۰ تا ۱۸۰ با هم دیگر همپوشانی داشته است، طوری که نقاط کلیک این دو نمودار همدیگر را کاملاً پوشانده‌اند. یعنی این‌طور برآورد می‌شود هر دو گروه دانشجویان با انگیزه یکسانی شروع به شرکت در کلاس‌های برخط کرده‌اند. از طرفی افزایش درجات گره‌ها برای مقادیر بیشتر از ۲۴۰ تا ۲۸۰ برای هر دو گروه تأیید کننده این قضیه است که روند تغییرات توزیع درجه در طی زمان برای هر دو سری دانشجویان تقریباً نزدیک بهم شده است اما دانشجویان ممتاز با کمی اختلاف بیشتر در میانگین کلیک روبه رو

رفتاری یکسانی از نظر تعامل و کلیک با سامانه نداشتند که نشان دهنده این است پویایی مراجعات دانشجویان ممتاز به سامانه برخط در رابطه با حل تمرین و حضور در سامانه برخط و پیگیری تکالیفی که در حین کلاس برخط معلم به آن‌ها برای فهم بیشتر مسئله می‌سپارد، نسبت به دانشجویان سطح پایین، بیشتر بوده است هر چند ممکن است که هر دو گروه دانشجویان با انگیزه‌های یکسانی شروع به فعالیت در سامانه می‌کنند و تمایل دارند که فعالیت در سامانه را به اتمام برسانند، اما دانشجویان سطح پایین چون کم‌دقت‌تر هستند و در روند حل تکالیف همراه با آزمون و خطا عمل می‌کنند، ممکن است در فرایند پاسخ دهی به تکالیف سامانه را ترک کرده باشند.

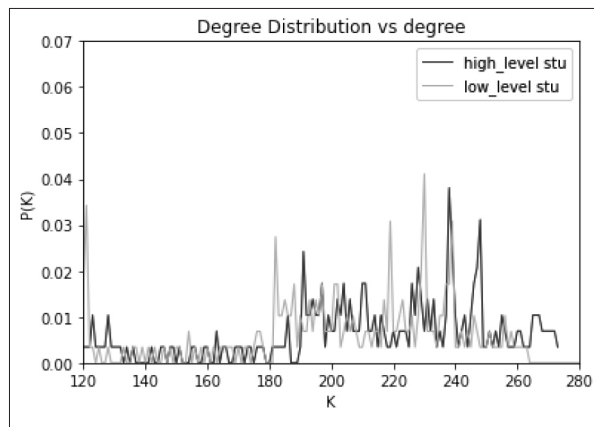
همان‌طور در شکل (۷) ب مشاهده می‌شود، مرکزیت درجه دانشجویان با برچسب high دارای مقادیر بالاتری نسبت به دانشجویان ضعیف است. یعنی این‌که دانشجویان سطح بالا در حین دوره برخط تمایل بیشتری به تماشای مجدد فایل‌های ضبط شده و محتوای درسی داشته‌اند همچنین آن‌ها تمایل بیشتری به متوقف کردن فایل‌های ضبط شده برای یادداشت‌برداری نکات درسی مهم و کلیک بیشتر داشته‌اند.

با بررسی شکل (۸) می‌توان دریافت که به‌صورت کلی روند توزیع درجه برحسب درجه گراف پدیداری میانگین کلیک هر دو سری دانشجو در روزهای مختلف، تقریباً مشابه است به‌صورت کلی میانگین دانشجویان ممتاز در



شکل ۹: میانگین دقت الگوریتم‌های دسته‌بندی استفاده شده پس از اجرا با 10 fold cross validation

طبیعی نگاشت کردیم و در نهایت با استفاده از معیارهای تحلیل شبکه پیچیده مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه تحلیل کردیم. سپس خروجی‌های گراف پدیداری دانشجویان ممتاز و ضعیف را از نظر میزان دقت تفکیک‌پذیری با الگوریتم‌های یادگیری ماشین بررسی کردیم. خروجی نتایج بیانگر این است که بالاتر بودن مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه می‌تواند معیار مهمی در دسته‌بندی دانشجویان ممتاز و ضعیف از نظر میانگین کلیک در سامانه باشد و دقت این ادعا با الگوریتم‌های گوناگون تحلیل شده است که میانگین دقت مطلوبی را نشان داده است. برای توسعه پژوهش، می‌توان داده‌های بایگانی را با گراف پدیداری وزن‌دار جهت بررسی شباهت‌های رفتاری دانشجویان در چندین مجموعه داده بررسی و مقایسه کرد. همچنین می‌توان برای بررسی ارتباط بین نرخ ثبت نام و میزان ترک تحصیل دانشجویان پیش‌بینی انجام داد. از طرفی دیگر می‌توان با تحلیل سایر ویژگی‌ها مانند جنسیت، سن، دوره‌های آموزشی گذرانده شده، نوع فعالیت‌های آموزشی از قبیل نوع آزمون، تعداد صفحات درسی، کارگاه‌ها و فایل‌های بارگیری که دانشجویان در آن‌ها شرکت کرده است به پیاده‌سازی مسیرهای یادگیری تطبیقی متناسب با سطح علمی، علایق و ترجیحات فردی هر دانشجو به صورت مجزا پرداخت. همچنین می‌توان از تجزیه و تحلیل شبکه‌های اجتماعی استفاده کرد و به ایجاد جامعه‌ای از دانشجویان و تعامل با آن‌ها به روشی مشارکتی‌تر کمک کرد. که این می‌تواند سبب



شکل ۸: توزیع درجه بر حسب درجه گراف پدیداری طبیعی میانگین کلیک دانشجو ممتاز و ضعیف

هستند و این کلیک‌ها نشانگر آن است که با پیشرفت دوره تحصیلی هر دو گروه دانشجویان سعی می‌کنند که در سامانه فعالیت کنند و پیگیر تکالیف باشند اما به دلیل عدم دقت و تمرکز دانشجویان سطح پایین در حل صحیح تکالیف، نتیجه نهایی نمرات آن‌ها متفاوت بوده است. برای ارزیابی این‌که معیارهای متمایز کننده توسط تحلیل گراف پدیداری مانند مرکزیت درجه، مرکزیت نزدیکی و توزیع درجه، با چه دقتی کار دسته‌بندی دانشجویان ممتاز از ضعیف را انجام می‌دهند از الگوریتم‌های دسته‌بندی موجود در شکل (۹) استفاده شده است. همان‌طور که مشخص است، میانگین دقت تفکیک دانشجو ممتاز از ضعیف، برای معیارهای معرفی شده توسط تحلیل گراف پدیداری با الگوریتم‌های مختلف یادگیری ماشین دقت مطلوبی را نشان داده است. در این بین الگوریتم درخت تصمیم چون مدل‌سازی ساده و قابل فهمی دارد قادر بوده با دقت ۷۸ درصد بهترین میانگین دقت را در میان سایر الگوریتم‌های دسته‌بندی داشته باشد.

## ۵ نتیجه‌گیری و کارهای آینده

در این مقاله ابتدا داده‌های جریان کلیک سامانه‌های آموزشی دانشجویان را جمع‌آوری کرده و پیش‌پردازش کردیم و به شکلی که برای تحلیل توسط گراف پدیداری مناسب باشد تبدیل کرده، سپس آن را به گراف پدیداری

ibility graph for establishing complex network from time series," *Acta Phys. Sin.*, vol. 61, no. 3, 2012.

[14] K. Liu, T. Weng, C. Gu, and H. Yang, "Visibility graph analysis of Bitcoin price series," *Phys. A Stat. Mech. its Appl.*, vol. 538, p. 122952, 2020, doi: 10.1016/j.physa.2019.122952.

[15] A. Bhaduri, S. Bhaduri, and D. Ghosh, "Visibility graph analysis of heart rate time series and bio-marker of congestive heart failure," *Physica A*, vol. 482, pp. 786–795, 2017, doi: 10.1016/j.physa.2017.04.091.

[16] G. R. Ghimire, N. Jadidoleslam, W. F. Krajewski, and A. A. Tsonis, "Insights On Streamflow Predictability Across Scales Using Horizontal Visibility Graph Based Networks," *Front. Water*, vol. 2, p. 17, 2020.

[17] D. Tsiotas and L. Magafas, "The effect of anti-COVID-19 policies on the evolution of the disease: A complex network analysis of the successful case of Greece," *Physics (College. Park. Md.)*, vol. 2, no. 2, pp. 325–339, 2020.

[18] N. Makaram and R. Swaminathan, "Analysis of dynamic contractions from biceps brachii muscle using surface electromyography signals and multiscale visibility graph features," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 2653–2656.

[19] T. Plocoște, R. Carmona-Cabezas, F. J. Jiménez-Hornero, E. G. de Ravé, and R. Calif, "Multifractal characterisation of particulate matter (PM10) time series in the Caribbean basin using visibility graphs," *Atmos. Pollut. Res.*, vol. 12, no. 1, pp. 100–110, 2021.

[20] J. Gómez-Gómez, R. Carmona-Cabezas, E. Sánchez-López, E. Gutiérrez de Ravé, and F. J. Jiménez-Hornero, "Analysis of Air Mean Temperature Anomalies by Using Horizontal Visibility Graphs," *Entropy*, vol. 23, no. 2, p. 207, 2021.

[21] J. Zhang and Y. Luo, "Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network," vol. 132, no. Msam, pp. 300–303, 2017, doi: 10.2991/msam-17.2017.68.

[22] F. Liu, N. Wang, and D. Wei, "Analysis of chinese stock market by using the method of visibility graph," *Open Cybern. Syst. J.*, vol. 11, pp. 36–43, 2017, doi: 10.2174/1874110X01711010036.

شناسایی رهبران نوظهور در گروه دانشجویانی که علائق مشابهی دارند و می‌توانند از فعالیت‌های یادگیری مشترک بهره‌مند شوند، کاربردی باشد.

## مراجع

[1] K. Dobashi, "Real-Time Web Scraping for Analyzing Moodle Course Material Clickstream," in 2021 19th International Conference on Information Technology Based Higher Education and Training (ITHET), 2021, pp. 1–10.

[2] C. Jin, "Dropout prediction model in MOOC based on clickstream data and student sample weight," *Soft Comput.*, vol. 25, no. 14, pp. 8971–8988, 2021, doi: 10.1007/s00500-021-05795-1.

[3] A. Alhothali, M. Albsisi, H. Assalahi, and T. Aldosemani, "Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review," *Sustain.*, vol. 14, no. 10, pp. 1–23, 2022, doi: 10.3390/su14106199.

[4] G. Zhu, Y. Li, and P. Wen, "Analysing epileptic EEGs with a visibility graph algorithm," in 2012 5th international conference on biomedical engineering and informatics, 2012, pp. 432–436.

[5] عزیزی، حدیث و سلیمانی، صادق، ۱۴۰۰، "رویکرد شبکه‌ای به داده‌های سری زمانی: مروری بر تحلیل گراف پدیداری"، اولین کنفرانس ملی سیستم‌های پیچیده با محوریت علم شبکه، تهران،

<https://civilica.com/doc/1444424>

[6] N. R. Aljohani, A. Fayoumi, and S.-U. Hassan, "Predicting at-risk students using clickstream data in the virtual learning environment," *Sustainability*, vol. 11, no. 24, p. 7238, 2019.

[7] M. Adnan et al., "Predicting at-risk students at different percentages of course length for early intervention using machine learning models," *Ieee Access*, vol. 9, pp. 7519–7539, 2021.

[8] Q. Li, R. Baker, and M. Warschauer, "Using clickstream data to measure, understand, and support self-regulated learning in online courses," *Internet High. Educ.*, vol. 45, p. 100727, 2020.

[9] A. A. Mubarak, H. Cao, W. Zhang, and W. Zhang, "Visual analytics of video-clickstream data and prediction of learners' performance using deep learning models in MOOCs' courses," *Comput. Appl. Eng. Educ.*, vol. 29, no. 4, pp. 710–732, 2021.

[10] E. Ulitzsch, V. Ulitzsch, Q. He, and O. Lütke, "A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks," *Behav. Res. Methods*, pp. 1–21, 2022.

[11] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: The visibility graph," *Proc. Natl. Acad. Sci.*, vol. 105, no. 13, pp. 4972–4975, 2008.

[12] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, "Horizontal visibility graphs: Exact results for random time series," *Phys. Rev. E*, vol. 80, no. 4, p. 46103, 2009.

[13] Z. T.-T. J. Ning-De and G. Bin, "Limited penetrable vis-