

## پیش بینی عوامل مؤثر در موفقیت تحصیلی دانشجویان دانشگاه پیام نور با کمک تکنیک‌های داده کاوی

سارا خواجه‌وند<sup>۱</sup>، عبدالله چاله‌چاله<sup>۲</sup>، بهمن سعیدی پور<sup>۳</sup>

<sup>۱</sup> کارشناس ارشد مهندسی کامپیوتر، دانشگاه پیام نور، ایران

<sup>۲</sup> استادیار مهندسی کامپیوتر، دانشگاه رازی، رازی، ایران

<sup>۳</sup> دانشیار مدیریت آموزشی، دانشگاه پیام نور، ایران

---

### چکیده

هدف از این مقاله پیش بینی عوامل مؤثر در موفقیت تحصیلی دانشجویان دانشگاه پیام نور با کمک تکنیک‌های داده کاوی می‌باشد. جامعه آماری شامل کلیه دانش‌آموختگان تعدادی از مراکز پیام نور استان کرمانشاه یعنی مراکز دانشگاهی کرمانشاه، هرسین، سرپل ذهاب، اسلام آباد غرب، پاوه و کنگاور است. این پژوهش روی داده‌های استخراج شده از سیستم جامع گلستان انجام و پس از گردآوری داده‌ها، از روش simple k-means به منظور خوشه‌بندی و تعریف دانشجویان ضعیف، متوسط و موفق استفاده شده است. سپس چند دسته‌بندی مختلف از جمله؛ درخت تصمیم‌گیری بر آن اعمال گشته است. یافته‌های پژوهش نشان می‌دهد که بهترین پاسخ‌ها، با بیش از ۹۳.۲٪ صحت است که با استفاده از درخت تصمیم‌گیری به دست آمده است.

**واژه‌های کلیدی:** وضعیت تحصیلی دانشجویان، داده کاوی، خوشه‌بندی، دسته‌بندی، درخت تصمیم

---

## ۱- مقدمه

یکی از وظایف دانشگاه‌ها تربیت نیروهای انسانی کارآمد است. دانشجویان هر دانشگاه سرمایه‌های گران‌قدر هر کشور به شمار می‌روند و توجه به تربیت آنها از نظر علمی، پژوهشی و فرهنگی بسیاری حائز اهمیت است و برنامه ریزی صحیح در آموزش و پرورش آنها، یکی از وظایف اصلی دانشگاه‌ها می‌باشد. دانشگاه‌ها سالانه دانشجویان جدیدی را پذیرش کرده، دانشجویانی را فارغ التحصیل می‌کند که توجه به کیفیت آموزشی آنها از اهمیت ویژه‌ای برخوردار است (نجیمی و همکاران، ۱۳۹۰). دانشگاه‌ها به عنوان مراکز تولید کننده‌ی علم و فرهنگ و تربیت متخصصان مورد نیاز کشور، خاستگاه بسیاری از نوآوری‌ها و خلاقیت‌ها برای حل مسائل روز جامعه هستند. یکی از مراکز آموزشی فعال در نظام آموزش عالی، دانشگاه پیام نور است که در مهرماه سال ۱۳۶۷ تأسیس شد. با توجه به آنکه جامعه‌ی آماری بزرگی در دانشگاه پیام نور در ایران وجود دارد، اطلاعات فارغ‌التحصیلان آن می‌تواند یکی از منابع مفید برای استخراج عوامل تأثیرگذار بر وضعیت تحصیلی دانشجویان باشد. افت تحصیلی، یکی از مسائل گریبان گیر سیستم آموزشی در همه‌ی مقاطع تحصیلی است و موجب هدر رفتن نیروی انسانی و هزینه‌های صرف شده می‌شود (بختیار و همکاران، ۱۳۸۱). از سوی دیگر، دانش آموختگانی که با افت تحصیلی دانشگاه را به اتمام می‌رسانند، توانمندی علمی مورد نیاز را به نحو مطلوب کسب نمی‌کنند. از داده‌های دریافتی می‌توان فهمید ۴۲٪ از دانشجویان دانشگاه پیام نور کرمانشاه در دوران تحصیل خود دچار مشروطی شده‌اند. این مشروطی می‌تواند دانشجویان را از نظر روحی دچار مشکل کند. مطالعه‌ی در آمریکا نشان می‌دهد که ۱۲.۶٪ از دانشجویان پزشکی یک دانشگاه دچار افت تحصیلی بوده‌اند که اغلب به دلیل نارسایی علمی آنها بوده است (لازین و نیومان، ۱۹۹۱). در مطالعه دیگری در آمریکا، شایعترین علت خودکشی دانشجویان، نارسایی یا اختلال تحصیلی یا شغلی گزارش شده است (میلمن و همکاران، ۱۹۹۴). با نگاهی به روند تحولات جاری نظام آموزش عالی، می‌توان دریافت که آموزش عالی باید ضمن توجه به بحران افزایش کمی و تنگناهای مالی، به حفظ و بهبود ارتقای کیفیت نیز بپردازد. رشد و توسعه‌ی هر جامعه متأثر از نیروی آموزش دیده‌ی آن جامعه بوده، شناسایی فاکتورهای مؤثر در پیشرفت تحصیلی دانشجویان و بذل توجه به آنها، گامی به سوی توسعه پایدار است. این مطالعه با هدف تعیین عوامل تأثیرگذار بر وضعیت تحصیلی دانشجویان پیام نور مراکز کرمانشاه، اسلام آباد غرب، پاره، جوانرود، هرسین، کنگاور و سرپل ذهاب انجام شده است. با توجه به رشد روز افزون پایگاه‌های داده‌ی مورد استفاده در دانشگاهها، فروشگاهها، بانک‌های املاک، مخابرات و موارد دیگر کسب دانش و الگوها توسط انسان، کار بسیار دشواری شده است. از طرفی چون اهمیت کسب دانش از اطلاعات موجود و تراکنش‌های انجام شده در پیشرفت یک سازمان را نمی‌توان نادیده گرفت. امروزه استفاده از تکنیک‌های کاوش داده از متن، صدا، تصویر و غیره رونق بسیار زیادی یافته و در مقالات متعدد، روش‌ها و تکنیک‌های بسیاری ارائه شده است. در سال ۱۹۶۰ آماردانان اصطلاح "data fishing" یا "Data Dredging" به معنای "صید داده" را جهت کشف هر گونه ارتباط در حجم بسیار بزرگی از داده‌ها بدون در نظر گرفتن هیچگونه پیش فرضی به کار بردند. بعد از سی سال و با انباشته شدن داده‌ها در پایگاه‌های داده، اصطلاح "Data Mining" در حدود سال ۱۹۹۰ رواج بیشتری یافت. اصطلاح Data Mining همان طور که از ترجمه آن بر می‌آید، به معنی داده کاوی بوده و مفهوم آن استخراج اطلاعات نهان و یا الگوها و روابط مشخص در حجم زیادی از داده‌ها در یک یا چند بانک اطلاعاتی بزرگ است

## ۲- بیان مسأله

دانشگاه یکی از سازمان‌هایی که پایگاه داده‌ی بسیار بزرگ و البته بسیار ارزشمندی را دارا می‌باشد. با توجه به رشد گرایش به تحصیل در کشور ایران، این پایگاه‌های داده به صورت روز افزون در حال گسترش هستند. بر اساس آمار وزارت علوم در سال

<sup>1</sup>Lazin and Neumann

<sup>2</sup>Meilman et al

<sup>۳</sup> صید داده

<sup>۴</sup> داده کاوی

۱۳۹۳، تعداد کل دانشجویان نسبت به سال ۱۳۵۷، ۲۶.۷ برابر شده است؛ به این ترتیب که تعداد دانشجویان کاردانی ۱۷.۷ برابر، تعداد دانشجویان کارشناسی ۲۹.۵ برابر، تعداد دانشجویان کارشناسی ارشد ۹۱.۵ برابر، تعداد دانشجویان دکتری حرفه ای ۶.۵ برابر و تعداد دانشجویان دکتری تخصصی ۵۸.۵ برابر گردیده است. با استناد به این آمار، می‌توان نتیجه گرفت که تعداد فارغ التحصیلان رشته‌های دانشگاهی نیز به همین ترتیب اضافه گردیده است. یکی از مراکز علمی که در سال‌های اخیر جای خود را میان دانشگاه‌های ایران به خوبی باز کرده، دانشگاه پیام نور است که این دانشگاه بر اساس رتبه‌بندی جدید موسسه QS، بیست و دومین دانشگاه برتر ایران است. بر اساس رتبه‌بندی سایت *webometrics*، وب سایت این دانشگاه نوزدهمین وب سایت برتر دانشگاه‌های ایران و دویست و سی و دومین دانشگاه برتر آسیاست؛ همچنین بر اساس رتبه‌بندی سایت *AICU*، وب‌سایت دانشگاه پیام‌نور نهمین وب‌سایت برتر میان دانشگاه‌های ایران است. دانشگاه پیام‌نور بزرگ‌ترین شبکه آموزش عالی دولتی ایران و در بین دانشگاه‌های باز، در آسیا دوم و در جهان ششم مرتبه را به خود اختصاص داده است. بنابراین انتظار یک پایگاه داده‌ی بسیار بزرگ از دانش‌آموختگان این دانشگاه امری دور از انتظار نیست. حال آنکه این پایگاه داده‌ی گسترده می‌تواند شامل اطلاعات ارزشمندی باشد. داده کاوی آموزشی یک تکنیک جدید ایجاد شده در حوزه ی آموزش است که می‌تواند روی داده‌های مرتبط با این حوزه فعالیت کند (کومار و سارابث، ۲۰۱۲)<sup>۱</sup> استفاده از داده کاوی آموزشی روشی مناسب برای درک بهتر دانشجویان و مجموعه ای است که در آن آموزش می‌بینید. داده کاوی را در سیستم های آموزشی می‌توان به سه گروه تقسیم کرد: ۱- داده کاوی اطلاعات دانشجویان ۲- داده کاوی اطلاعات آموزش دهندگان ۳- داده کاوی اطلاعات مسئولان و مدیران آموزشی (زواره، ۱۳۹۲) به دلیل افزایش اهمیت آموزش در ایران و گرایش به افزایش کیفیت تحصیل در سال‌های اخیر، استفاده از داده‌های آموزشی دانشگاه پیام نور، می‌تواند بسیار ارزشمند باشد. به منظور افزایش رشد هرچه بیشتر کشور، نیاز به تربیت دانشجویانی زنده احساس می‌گردد. هرچند که این نیاز برای تمامی مقاطع تحصیلی احساس می‌گردد؛ اما تمرکز این پژوهش بر روی فارغ التحصیلان مقطع کارشناسی است. با استفاده از کشف الگوها و استخراج دانش از پایگاه داده‌ی مجموعه‌ی دانشجویان فارغ‌التحصیل از دانشگاه‌های پیام نور مراکز کرمانشاه، اسلام آباد غرب، پاوه، جوانرود، هرسین، کنگاور و سرپل ذهاب، می‌توان اطلاعاتی ارزشمندی را در مورد کیفیت تحصیلی دانشجویان و عوامل مؤثر بر کیفیت تحصیلی آنها به دست آورد.

### ۳- اهمیت و ضرورت پژوهش

از آنجائیکه موفقیت تحصیلی دانشجویان یکی از مهمترین امور در این حوزه است در این پژوهش سعی شده است که از روابط پنهان موجود بین داده‌ها برای پیش بینی فاکتورهای موفقیت دانشجویان استفاده شود. مسأله موفقیت و شکست تحصیلی سابقه‌ای طولانی دارد و تاریخ آن مقارن با آغاز خواندن و نوشتن توسط بشر است؛ اما به طور رسمی پس از اجباری شدن آموزش در اواخر قرن ۱۹ و همگام با تحولات صنعتی و نیاز به نیروهای تربیت شده، تمرکز دولتمردان به مسأله پیشرفت و شکست تحصیلی معطوف شد. متأسفانه همیشه مسأله افت تحصیلی برای اولیاء و معلمان و مربیان مبهم و پیچیده بوده است و کمتر کسی توانسته است به ریشه چنین نتایج نامطلوبی که برای برخی از دانش‌آموزان و دانشجویان به وجود می‌آید، پی ببرد و حقایق پنهان و بیشماری را روشن سازد و پاسخگوی این سوال باشد که چه عواملی باعث بروز و ظهور این مشکل اجتماعی می‌شوند (بیابانگرد، ۱۳۸۴) به طور کلی محرومیت‌های محیطی و کمبود محرک‌های رشد ذهنی و اجتماعی از عوامل مهم درافت تحصیلی است. « درصد قابل توجهی از افراد مشغول به تحصیلی که به طور جدی دچار افت تحصیلی شده‌اند افرادی هستند که متعلق به خانواده‌هایی از تبار اقتصادی واجتماعی فوق العاده پایینی هستند که با فقر اقتصادی و فرهنگی مواجه واز انگیزه تحصیلی بسیار پائین برخوردارند» (افروز، ۱۳۷۴) «وضعیت اقتصادی تعیین کننده‌ی محل سکونت است و اگر افراد به دلیل فقر مالی نتوانند محل مناسبی داشته باشند برای انجام تکالیف و مطالعه‌ی دروس با مشکل روبرو خواهند بود. محل

<sup>۱</sup>Kumar and Saurabh

سکونت تعیین کننده‌ی نوع مدرسه‌ای است که کودک در آن درس می‌خواند. وقتی محل فرد در یک محل فقیر باشد از نظر امکانات آموزشی و معلمان با کمبودهایی روبرو می‌شود که مستقیماً بر پیشرفت تحصیلی او اثر می‌گذارد» (بیابانگرد، ۱۳۸۴). اندازه‌گیریهای مربوط به چگونگی ارتباط پیشرفت تحصیلی با وضعیت اقتصادی - اجتماعی موقعیت خانواده در تمامی کشورها و در تمامی سطوح سنی و تحصیلی و در مورد تمامی موارد درسی حکایت از وجود یک رابطه مثبت دارد. در واقع عامل اقتصادی یکی از عواملی است که تأثیر آن بر روی افراد مشغول به تحصیل به وضوح دیده می‌شود. عواملی چون سن افراد تحصیل کننده، مشکلات خانوادگی شاغل بودن یا شاغل نبودن و بسیاری از موارد دیگر نیز ممکن است که بر نتایج تأثیرگذار باشد. همچنین جنسیت و دانشکده از جمله عوامل مؤثر بر پیشرفت تحصیلی می‌باشند. پذیرش دانشجویان تکمیل ظرفیت، استفاده از سهمیه - های بومی و ارائه امکانات رفاهی می‌تواند بر این پیشرفت مؤثر باشد. در این مطالعه سعی بر این است که سایر الگوهای دیگر نیز شناسایی شود. الگوهایی که شاید به وضوح قابل درک و یافتن نباشند. در ضمن، بیشتر تلاش‌هایی که روی شناسایی عوامل مؤثر بر پیشرفت تحصیلی انجام گرفته است بر دانش آموزان بوده و این عوامل می‌تواند برای دانشجویان متفاوت باشد. با توجه به آنکه سن دانشجویان به مراتب نسبت به دانش آموزان بیشتر بوده و در رده‌ی سنی جوانان قرار می‌گیرند، عوامل و مشکلات آنها با دانش آموزانی که در رده‌های کودک و نوجوان قرار دارد می‌تواند بسیار متفاوت باشد. یکی از دلایل مهم بررسی این موضوع، اثرات مخرب ناشی از تداوم مشکلات تحصیلی بر نیروی کار کشور است. در صورتی که روندهای مخرب اثر گذار بر روی وضعیت تحصیلی دانشجویان ادامه داشته باشد، نیروی کاری که به مرور زمان وارد بازار کار می‌شوند، افرادی ناکارآمد خواهند بود و منجر به خسارات اقتصادی متعددی به اقتصاد کشور خواهند شد. روی دیگر سکه نیز زمانی است که نیروی کار متخصص برای اکتساب شغل‌های موجود وجود نداشته باشد. در صورتی که الگوها و اثرات مخرب بر روی وضعیت تحصیلی دانشجویان ادامه داشته باشد، سطح علمی دانشجویان فارغ‌التحصیل از کیفیتی مناسب برای برآزش توسط کارفرمایان برخوردار نخواهد بود.

#### ۴- پیشینه تحقیق

موسوی و خان‌بابایی (۱۳۹۲) مطالعه‌ای در مورد تکنیک‌های ترکیبی داده‌کاوی جهت هدایت تحصیلی دانش‌آموزان ارائه کردند. آنها برای ساخت مدل‌های مورد نظر از تکنیک‌های مختلفی نظیر الگوریتم درخت تصمیم CART، الگوریتم K نزدیکترین همسایه و بهبود یافته الگوریتم نزدیکترین همسایه با استفاده از الگوریتم ژنتیک بر روی اطلاعات ۹۶۹ دانش آموز و با استفاده از متدولوژی کریسپ در نرم افزارهای متلب و کلمنتاین استفاده کردند. این پژوهشگران نتیجه گرفتند که در الگوریتم بهبود یافته K نزدیکترین همسایه بالاترین دقت را دارا می‌باشد (موسوی و خان‌بابایی، ۱۳۹۲). علی زواره و همکارانش (۱۳۹۲) به بررسی اطلاعات آموزشی دانش‌آموختگان رشته ی دندانپزشکی بین سالهای ۱۳۷۵ تا ۱۳۹۰ پرداختند و ایشان مدلی جهت تشخیص دانشجویان ضعیف و قوی ارائه دادند و عوامل مؤثر بر نمرات و رفتار آموزشی آنان را بیان کردند. در این تحقیق، داده کاوی داده‌ها با استفاده از یک چند مرحله‌ایبه نام CHAID انجام گردید و برای پیش‌بینی از مدل‌های Two-Step و K-means برای خوشه بندی استفاده گردید. نتایج این مطالعه نشان داد که فیلدهای جنسیت، سن، سهمیه، معدل کل و تعداد ترم مشروطی در خوشه بندی به عنوان فیلدهای با اهمیت به دست آمدند [۳]. تقوی و همکارانش (۱۳۹۲) با استفاده از تکنیک های داده کاوی به پیش بینی معدل ترمی و بررسی موفقیت یا عدم موفقیت دانشجویان رشته عمران دانشگاه شاهرود در فارغ-التحصیلی پرداختند و سعی کردند با توجه به پارامترهایی همچون تعداد دروس (عمومی، پایه، اصلی) و معدل دروس (عمومی، پایه، اصلی...) دانشجویان و با استفاده از تکنیک درخت تصمیم بر روی داده‌های مورد نظر، عوامل و روابط موجود و تأثیرگذار در بهبود عملکرد و موفقیت دانشجویان را در طول تحصیل شناسایی کنند که با توجه به مدل بدست آمده، ترم هشت به عنوان بهترین حالت و بیشترین وابستگی برای پیش بینی معدل شناسایی شد (تقوی و همکاران، ۱۳۹۲). محمد زاده (۱۳۹۱) این پژوهش به منظور ارزشیابی کیفیت برنامه‌های آموزشی با توجه به استفاده از فناوری اطلاعات در چهار دانشکده ی علوم انسانی دانشگاه تهران، علمی کاربردی، پیام نور و تربیت مدرس در شهر تهران انجام شد. با توجه به تحلیل-های حاصل از داده‌های این پژوهش مشخص شد که به آموزش‌های نرم‌افزاری در دانشکده‌های علوم انسانی در واحدهای برنامه

ریزی کمتر توجه شده، در حالی که استقبال دانشجویان برای استفاده از این برنامه‌های آموزشی بسیار بالا می‌باشد؛ اما رضایت آنها از برنامه‌های آموزشی با توجه به هزینه‌های بالایی که برای تحصیلات دانشگاهی متحمل می‌شوند، بسیار پایین است. سنایی نسب و همکاران (۱۳۹۱) به بررسی عوامل مؤثر بر پیشرفت تحصیلی دانشجویان با نمونه برداری ۱۸۰ دانشجویان از رشته‌ها و مقاطع مختلف دانشگاه علوم پزشکی تهران پرداختند. آنها دریافتند که متغیرهای جنسیت، گروه سنی، مقطع تحصیلی، رشته تحصیلی، علاقه به رشته تحصیلی و سبک یادگیری، تفاوت معنی داری را در شاخص پیشرفت تحصیلی شرکت کننده‌ها نشان دادند (سنایی نسب، ۱۳۸۹) توسرکانی و نصیری (۱۳۹۰) با استفاده از تکنیکهای داده کاوی وضعیت تحصیلی آتی دانش‌آموزان شامل دوره‌ی فارغ‌التحصیلی از تحصیلات متوسطه و وضعیت فارغ‌التحصیلی از مدارس روزانه یا بزرگسال پیشبینی شده‌است. برای ساخت مدل‌های مورد نظر از تکنیکهای مختلفی نظیر درخت تصمیم SVM و MLP و شبکه‌های عصبی استفاده گردیده‌است. اعتبار سنجی انجام شده روی مدلها اثبات می‌کند که نتایج به دست آمده دقیق و قابل اعتماد بوده اند. استفاده از نتایج حاصل می‌تواند کمک زیادی برای مدیران دبیرستانها جهت بهینه کردن فرآیندهای آموزشی باشد (مرادی توپسرکانی و نصیری، ۱۳۹۱). محمد تازی و همکارانش (۱۳۸۹) الگوریتم داده کاوی را روی بانک‌های اطلاعاتی دانشگاه پیام نور استان قم به کار برده و از طریق دسته بندی انتخاب واحد دانشجویان با کمک الگوریتم APRIORI، به استخراج الگوهای پنهان پرداخته و از مدل‌های ایجاد شده به عنوان یک ابزار پشتیبان تصمیم‌گیری در سیستم‌های آموزشی مشابه منظور ارتقاء سطح علمی دانشگاه‌ها استفاده کردند. آنها در این مطالعه با استفاده از نرم‌افزار SPSS و الگوریتم Apriori مدلی برای وابستگی بین دروس را ایجاد نمودند. این مطالعه می‌تواند برای ارائه‌ی دروس توسط گروه‌های دانشگاهی بسیار مفید باشد (تازی، ۱۳۸۹). صادقی و همکارانش (۱۳۸۹) تحقیقی روی ۹۹ دانشجوی جدیدالورود دانشکده توانبخشی که ۳۴ نفر آنها مرد و ۶۵ نفر زن بودند و از بین آنها ۲۷ نفر رشته کاردرمانی، ۲۵ نفر رشته فیزیوتراپی، ۲۲ نفر رشته بینایی سنجی و ۲۵ نفر رشته شنوایی سنجی بودند انجام دادند و متوجه شدند که هوش هیجانی با جنسیت، سطح سواد مادران و درآمد خانواده رابطه‌ی مستقیم دارد (میرزا خانی، ۱۳۸۹). رودباری و همکاران (۱۳۸۹) به مطالعه‌ی عوامل مؤثر بر پیشرفت تحصیلی دانشجویان دانشگاه علوم پزشکی تهران (پردیس همت) در سال تحصیلی ۸۸-۸۹ پرداختند. اطلاعات آنها با استفاده از جمع‌آوری ۶۷۰ نمونه از پایگاه سیستم سما (سیستم آموزش دانشگاه) شامل اطلاعات دموگرافیک و آموزشی (معدل کل، تعداد دروس مردودی و ترم‌های مشروطی) جمع‌آوری گردید. آنها دریافتند که جنسیت، مقطع تحصیلی و دانشکده از جمله عوامل مؤثر بر پیشرفت تحصیلی می‌باشند. پذیرش دانشجویان تکمیلی و دختر، استفاده از سهمیه‌های بومی و ارائه‌ی امکانات رفاهی می‌تواند بر این پیشرفت مؤثر باشد. پذیرش تعداد زیاد دانشجو در مقاطع پزشکی عمومی و تحصیلات تکمیلی، نیز خود از جمله دلایل عدم موفقیت این دانشجویان می‌باشد. سوچیتا بوکار و همکارانش<sup>۱</sup> (۲۰۱۳) به ارزیابی کارایی دانش‌آموزان با استفاده از الگوریتم استخراج رابطه پرداختند. آنها قوانین مهم را تولید و از آن برای اندازه‌گیری ارتباط میان ویژگی‌های مختلف برای بهبود عملکرد تحصیلی دانش‌آموز کمک گرفتند. نتایج نشان داد که سطح عملکرد دانش‌آموزان را می‌توانتوسط شناسایی دانش‌آموزانی که در امتحان، حضور و غیاب و فارغ‌التحصیلی، ضعیف هستند بهبود بخشید. سورجیت کومارو سائوراب پال (۲۰۱۲) مدل‌های داده کاوی را برای پیشگویی کارایی دانش‌آموزان توسعه دادند و از طریق آنها دانش‌آموزان ضعیف و قوی را تشخیص داده، نتیجه گرفتند که عواملی همچون: متوسط آموزش، محل زندگی دانش‌آموز، صلاحیت مادر دانش‌آموز، عادات و درآمد سالانه‌ی خانواده در وضع درسی و نمره‌ی دانش‌آموز تأثیر دارد [۱۹]. زلاتکو<sup>۲</sup> (۲۰۱۰) نشان داد که متغیرهایی مثل (سن، جنسیت، تحصیلات، وضعیت کاری، ناتوانایی‌ها و محیط مطالعه) ممکن است روی پایداری ترک تحصیل دانش‌آموزان در پلی تکنیک باز نیوزلند تأثیرگذار باشند. امانوئل اوگار<sup>۳</sup> (۲۰۰۷) متدولوژی استخراج شاخص‌های پیش‌بینی عملکرد را برای گسترش ارزیابی ساده‌ی عملکرد دانش‌آموز و نظارت سیستم در طول آموزش و یادگیری محیط با تمرکز عمده‌ای روی بازبینی عملکرد

<sup>1</sup>Borkar and Rajeswari<sup>2</sup>Zlatko<sup>3</sup>Ogor

دانش آموزان و نمره ی آزمون برای پیش بینی وضعیت نهایی خود پس از فارغ التحصیلی، با کمک تکنیک‌های مختلف داده کاوی ارائه داد و نتیجهی پروژه یاو نشان داد که قابلیت های DMT ابزار نظارت مؤثر برای عملکرد تحصیلی دانش آموزان رافراهم می کند سوپر بای و مسکنس<sup>۱</sup> (۲۰۰۶) به بررسی فاکتورهای تاثیرگذار بر موفقیت دانشجویان سال اول دانشگاهی با استفاده از روش های کاوش داده پرداختند. آنها دانشجویان سال اولی را به سه گروه بر اساس معدل تقسیم بندی کردند و با استفاده از شبکه های عصبی، جنگل های تصادفی و درخت های تصمیم گیری به ایجاد مدل هایی برای آنها اقدام نمودند.

### ۵- فرضیه های پژوهش

در زیر فرضیه های عینی ارائه شده است که ناشی از پردازش داده ها نبوده و رسیدن به این نتایج با استفاده از داده ها محتمل به نظر می رسد.

۱. بین موفقیت تحصیلی دانشجویان بر حسب وضعیت اقتصادی تفاوت معنا داری وجود دارد.
۲. بین موفقیت تحصیلی دانشجویان بر حسب رشته تحصیلی تفاوت معنا داری وجود دارد.
۳. بین موفقیت تحصیلی دانشجویان بر حسب جنسیت تفاوت معنا داری وجود دارد.
۴. بین موفقیت تحصیلی دانشجویان بر حسب نظام آموزشی تفاوت معنا داری وجود دارد.
۵. موفقیت تحصیلی سهمیه های پذیرش تفاوت معنا داری دارند.

### ۶- روش پژوهش

این پژوهش از نوع پژوهش های بنیادی- کاربردی است؛ جامعه آماری شامل دانشجویان دانشگاه پیام نور کل کشور می باشد و نمونه ای که برای این جمعیت بزرگ در نظر گرفته شده است، مجموعه دانشجویان دانشگاه پیام نور در واحدهای کرمانشاه، اسلام آباد غرب، هرسین، جوانرود، پاوه، سرپل ذهاب و کنگاور می باشد که مجموعه ای بیش از ۶۰۰۰۰ رکورد مشتمل بر اطلاعات دانشجویان فارغ التحصیل، در حال تحصیل، انصراف داده و اخراج شده است که منبع غنی برای تحقیق و پژوهش خواهد بود. در این پژوهش به منظور جمع آوری اطلاعات نیازی به استفاده از مشاهده، مصاحبه و پرسشنامه نبوده و داده ها به طور مستقیم از سیستم جامع دانشگاه پیام نور به دست آمده است. مزیت این روش موجود بودن و ارزان بودن دسترسی به اطلاعات است. ولی مشکل این سیستم این می باشد که تمامی اطلاعات قابل دسترسی نبوده و می توان گفت که اطلاعات بهتری می توانست به منظور بررسی وضعیت تحصیلی دانشجویان در بر داشته باشد. **روایی این پژوهش به شرح زیر است:** برای روایی ابزار از روایی محتوایی و صوری استفاده شده است. در این نوع روایی برای تعیین روایی محتوایی یک آزمون از قضاوت متخصصان در این باره که داده ها تا چه میزان معرف محتوا و هدف های برنامه یا حوزه محتوایی هستند، استفاده می شود. در این پژوهش ابزار تهیه شده در اختیار استاد راهنما، مشاور و پنج نفر از صاحب نظران قرار داده شد که پس از بررسی و مطالعه آن اشکالات برطرف و ابزار نهایی تدوین گردید. **و پایایی این پژوهش به شرح زیر است:** پس از تهیه ابزار گردآوری داده ها و تایید روایی آن به منظور تعیین میزان پایایی آن از آلفای کرونباخ استفاده شد. با توجه به تعداد بسیار بالای رکوردها، عموماً پایایی نشان داده شده پس از خوشه بندی بر اساس خصوصیات، بیشتر از ۰.۸۵ بود که بیانگر پایا بودن ابزار گردآوری داده ها می باشد. همچنین به منظور تست پایایی، مجموعه ی داده به دو قسمت تقسیم گردید که در آن یک قسمت بر اساس قسمت دیگر پیش بینی شد. در نرم افزار Weka، مجموعه ی اول که اصولاً باید حدود ۷۰ درصد کل مجموعه داده را داشته باشد، مجموعه ی Train<sup>۲</sup> و مجموعه ای که شامل داده های کمتر است، مجموعه ی Test<sup>۳</sup> نامیده می شود. مجموعه

<sup>۱</sup>Superby and Meskens

<sup>۲</sup>مجموعه ی آموزش

<sup>۳</sup>مجموعه ی آزمایش

**Train** به منظور آموزش و ایجاد مدل به کار گرفته شده و مجموعه‌ی **Test** به منظور آزمایش میزان پایایی آن استفاده می‌گردد. عمل جدا سازی مجموعه‌ی **Test** از **Train** را می‌توان به صورت دستی انجام داد؛ اما در نرم‌افزار وکا قابلیت جداسازی آنها با استفاده از تعیین میزان مجموعه‌ی **Train** با استفاده از مشخص کردن آن بر اساس درصد، وجود دارد. این مطالعه با استفاده از داده‌های پراکنده (به دلیلی آنکه داده‌هایی که پشت سر هم وجود دارند مربوط به یک مرکز دانشگاهی پیام نور هستند) انجام شد تا بتوان درصد روایی و پایایی را با دقت هرچه بیشتر تخمین زد. در اکثر موارد و در نتایج گزارش شده، مقادیر پیش بینی شده بین ۸۰ تا ۹۷ درصد قابل اطمینان بودند که اصولاً درصد‌های زیر ۶۰ درصد نادیده گرفته شده و به عنوان "عدم وجود رابطه‌ی قابل توجه گزارش" گردیده‌اند.

## ۷- تجزیه و تحلیل داده‌ها

برای داده کاوی نرم افزارهای مختلفی عرضه شده است که هر یک از آنها با توجه به نوع داده‌هایی که مورد کاوش قرار می‌دهند الگوریتم خاصی دارند. بنابراین شاید قیاس آنها بجا نباشد. مقایسه دقیق و علمی این ابزارها باید از جنبه‌های متفاوت و متعددی مانند تنوع انواع و فرمت داده‌های ورودی، حجم ممکن برای پردازش داده‌ها، الگوریتم‌های پیاده سازی شده، روشهای ارزیابی نتایج، روشهای مصور سازی، روشهای پیش پردازش داده‌ها، واسط‌های کاربر پسند، پلت‌فرمهای سازگار برای اجرا، قیمت و در دسترس بودن نرم افزار صورت گیرد. از این میان، نرم افزار **Weka** با داشتن امکانات بسیار گسترده، امکان مقایسه خروجی روشهای مختلف با هم، راهنمای خوب، واسط گرافیکی کارا، سازگاری با سایر برنامه‌های ویندوزی، و از همه مهمتر وجود کتابی بسیار جامع و مرتبط با آن معرفی می‌شود. برنامه وکا مجموعه‌ای از الگوریتم‌های روز است. همچنین این امکان وجود دارد که الگوریتم‌های جدیدتر به آن اضافه گردد و یا حتی الگوریتم‌هایی توسط کاربران نوشته و بکار رود. این سیستم به زبان جاوا نوشته شده و تقریباً روی هر پلت فرمی اجرا می‌شود. این نرم افزار، یک واسط همگون برای بسیاری از الگوریتم‌های یادگیری متفاوت، فراهم کرده است که از طریق آن روشهای پیش پردازش، پس از پردازش و ارزیابی نتایج طرح های یادگیری روی همه مجموعه‌های داده موجود، قابل اعمال است. این محیط، شامل روشهایی برای همه مسائل استاندارد داده کاوی مانند رگرسیون، رده بندی، خوشه بندی، کاوش قواعد انجمنی و انتخاب ویژگی می‌باشد. با در نظر گرفتن اینکه، داده‌ها بخش مکمل کار هستند، بسیاری از ابزارهای پیش پردازش داده‌ها و مصورسازی آنها فراهم گشته است. همه الگوریتم ها، ورودیهای خود را به صورت یک جدول رابطه‌های به فرمت **ARFF** دریافت می‌کنند. این فرمت داده‌ها، میتواند از یک فایل خوانده شده یا به وسیله یک درخواست از پایگاه داده‌های تولید گردد. یکی از راههای به کارگیر **Weka**، اعمال یک روش یادگیری به یک مجموعه داده و تحلیل خروجی آن برای شناخت چیزهای بیشتری راجع به آن اطلاعات می‌باشد. راه دیگر استفاده از مدل یادگیری شده برای تولید پیشبینی‌هایی در مورد نمونه‌های جدید است. سومین راه، اعمال یادگیرنده‌های مختلف و مقایسه کارایی آنها به منظور انتخاب یکی از آنها برای تخمین می‌باشد. روشهای یادگیری **Classifier** نامیده می‌شوند و در واسط تعاملی **Weka**، می‌توان هر یک از آنها را از منو انتخاب نمود. بسیاری از **classifier**ها پارامترهای قابل تنظیم دارند که می‌توان از طریق صفحه ویژگی‌ها یا **object editor** به آنها دسترسی داشت. پس از آماده سازی و پیش پردازش داده‌ها، داده‌های آماده سازی شده را در برنامه وکا وارد کرده و خوشه بندی روی آن اعمال می‌کنیم. همان طور که پیش از این نیز گفته شده بود، ما از تکنیک جداسازی مجموعه‌ی آموزشی و آزمایشی در بخش دسته بندی به منظور یافتن میزان اطمینان و پیش بینی استفاده خواهیم کرد، به همین دلیل در اینجا دلیلی برای جدا سازی آنها نداریم و گزینه‌ی انتخابی ما **use training set** می‌باشد. بنابراین در اینجا خوشه بندی بر روی تمامی داده‌های موجود انجام خواهد گرفت تا بتوان خوشه‌هایی را پیدا کرد که در آن وضعیت دانشجویان از یکدیگر متفاوت باشد. باید گفت، **Simple K Means** تنها بر روی فیلدهایی کار می‌کند که مقادیر آنها کمی باشد و نه کیفی. بنابراین در اینجا، روند خوشه بندی با در نظر گرفتن ۴ فیلد زیر وضعیت تحصیلی دانشجویان را تعیین می‌کند: معدل، تعداد مشروطی، تعداد مشروطی متوالی و مدت زمان تحصیل. تعداد خوشه‌ها را ۳ و از آنجایی که نقطه‌ی شروع به کار برای **Simple K Means** در فضا به صورت تصادفی انتخاب می‌شود و با

انتخاب مقدار هسته می توانیم مقدار تصادفی آن را تغییر دهیم، این مقدار را آنقدر تغییر دهیم که فواصل دسته‌ها به مقدار قابل قبول و کاملاً متمایزی برسد. پس از انتخاب هسته مناسب سه خوشه دانشجویان ضعیف، متوسط و قوی ایجاد می شود که تحلیل داده‌های این خوشه‌ها نشان می‌دهد که:

دانشجویان موفق به نسبت در دانشگاه مرکز کرمانشاه، دانشجویان متوسط در اسلام آبادغرب و دانشجویان ضعیف در پناه بیشتر هستند.

دانشجویان ضعیف زن به نسبت از دانشجویان ضعیف مرد بیشتر است استفاده از سهمیه بومی بودن به پذیرش در دانشگاه کمک کرده ولی باعث می شود دانشجویان ضعیف تری وارد دانشگاه شوند افراد بالای ۳۰ سال به نسبت موفق تر هستند و دانشجویان زیر ۲۲ سال به نسبت بیشتر در دسته دانشجویان ضعیف قرار میگیرند.

دانشجویان رشته علوم تربیتی و روانشناسی نسبت به رشته‌های دیگر کمترین دانشجویان ضعیف و بیشترین دانشجویان موفق را دارد.

دانشجویانی که محل تحصیل و زندگی آنها در یکجا باشد به طور نسبی دانشجویان ضعیف تری هستند. پس از بدست آمدن خوشه‌ها، برای اطمینان هرچه بیشتر و اندازه گیری میزان اطمینان به خوشه‌های به دست آمده و همچنین به منظور پیش‌بینی، از دسته بندی استفاده می‌کنیم. دسته بندی به عمل طبقه بندی دسته‌ها بر اساس خوشه‌های از پیش تعیین شده، گفته می‌شود. روش‌های دسته بندی متفاوتی وجود دارد که به واسطه‌ی آن بتوان پیش‌بینی را انجام داد. در این پژوهش ما از دسته‌بندی‌های ماشین‌های بردار پشتیبان، درخت تصمیم، شبکه‌ی عصبی و شبکه‌های بیز استفاده می‌نماییم که تمامی این موارد با یکدیگر مقایسه شد و بهترین حالت درخت تصمیم‌گیری بود که با مقدار ۹۳.۲۸۱۵٪ پیش‌بینی درست، بهترین کارایی را داشت. در عین حال، زمان ایجاد درخت تصمیم‌گیری بسیار پایین بود. از آنجایی که هدف نهایی ما تحلیل درخت تصمیم‌گیری و پیش‌بینی بر اساس آن است، بهترین و مناسب‌ترین گزینه درخت تصمیم‌گیری J48 بود. بنابراین در مراحل بعدی، از مدل ایجاد شده توسط درخت تصمیم‌گیری برای پیش‌بینی عوامل موثر استفاده خواهیم کرد.

جدول ۱-مقایسه‌ی روش‌های دسته بندی مختلف

| دسته بندی           | درصد نمونه‌هایی که به درستی دسته بندی شده اند | خطای دسته بندی | زمان تولید مدل |
|---------------------|---|----------------|----------------|
| درخت تصمیم‌گیری     | ۹۳.۲۸۱۵ درصد                                  | ۰.۱۷۰۷         | ۰.۰۳ ثانیه     |
| شبکه عصبی           | ۹۳.۲۲۹۶ درصد                                  | ۰.۲۰۲۲         | ۱۹۸.۳۴ ثانیه   |
| شبکه بیز            | ۹۲.۲۶۹۸ درصد                                  | ۰.۲۵۷          | ۰.۱ ثانیه      |
| ماشین بردار پشتیبان | ۹۳.۰۴۸ درصد                                   | ۰.۲۱۵۳         | ۲۸.۱۶ ثانیه    |

## ۸- بحث و نتیجه گیری

داده کاوی فرایند اکتشاف اطلاعات و الگوهای نهفته از درون حجم بسیار زیاد داده‌هایی است که در قالب پایگاه‌های داده ای، انبار داده ای، یا هر نوع انبار اطلاعاتی ذخیره می شود. داده کاوی اطلاعات موجود در انبار داده‌ها را استخراج و داده‌ها را به دانش حیاتی تبدیل می نماید. با توجه به سناریوهایی مذکور می‌توان نتیجه گرفت که دو نظام سنتی و تجمیع در مرحله اول بیشترین تأثیر را برای جداسازی داده‌ها داشته و پس از آن به ترتیب سهمیه (مناطق - ایثارگری - آزاد)، رشته تحصیلی، محل تحصیل دانشجو، سن دانشجو هنگام ورود به دانشگاه، سهمیه بومی بودن، جنسیت، یکسان بودن محل زندگی و تحصیل دانشجو و دوره تحصیلی (رسمی - فراگیر) عوامل موثر و مهم در موفقیت تحصیلی دانشجو می‌باشند. این پژوهش دارای ۵ فرضیه می‌باشد که با توجه به نتایج حاصل از خوشه بندی و دسته بندی تمامی این فرضیات تایید می شود. با توجه به اینکه برای این مطالعه تنها از یک روش؛ یعنی روش درخت تصمیم‌گیر استفاده گردیده است، پیشنهاد می‌گردد به منظور دستیابی



به پاسخ‌های بهتر، در تحقیقات آتی از روش‌های ترکیبی استفاده شود. همچنین پیشنهاد می‌گردد که از پایگاه داده‌های کامل-تر که به طور کامل شامل فیلدهای وارد شده‌ی وضعیت تاهل و سطح درآمد افراد یا محل زندگی آنها می‌باشد، استفاده گردد تا میزان تأثیر این موارد نیز به وضوح مشخص شود. همچنین می‌توان از پایگاه داده‌ای دانشگاه‌های دیگر همانند دانشگاه آزاد اسلامی و دانشگاه سراسری برای دستیابی به فیلدهای ارزشمند دیگر نیز استفاده نمود.

#### منابع

۱. افروز، غلامعلی (۱۳۷۴) جامعه فرهنگ و تدوین شخصیت کودکان و نوجوانان، پیوند-نشریه ماهانه آموزشی تربیتی انجمن اولیاء و مربیان جمهوری اسلامی ایران
۲. بختیار نصرآبادی، حسنعلی، بهجنی اردکانی، بتول (۱۳۸۱)، علل آموزشی افت تحصیلی و راهکارهای جلوگیری از آن، فولاد، شماره ۹۶
۳. بیابانگرد، اسماعیل (۱۳۸۴) روش‌های تحقیق در روانشناسی و علوم تربیتی، تهران، انتشارات دوران، چاپ اول
۴. تاری، محمد (۱۳۸۹). اثر بخشی داده کاوی در مدیریت آموزش عالی و مطالعه موردی آن در دانشگاه پیام نور استان قم، سومین همایش ملی مهندسی کامپیوتر و فناوری اطلاعات
۵. تقوی، مطهره السادات؛ جلالی، محمد مهدی و طلوع بیدختی، ندا (۱۳۹۲). بررسی عوامل مؤثر در بهبود عملکرد دانشجویان رشته عمران در طول ترم تحصیلی پیشبینی معدل ترمی با بکارگیری تکنیکهای داده‌کاوی، موسسه آموزش عالی خاوران مشهد،
۶. رودباری، مسعود؛ احمدی، آزاده و عبادی فردآذر، فرید (۱۳۸۹). تعیین عوامل مؤثر بر پیشرفت تحصیلی دانشجویان علوم پزشکی تهران (پردیس همت) در سال تحصیلی ۸۸-۸۹، نشریه طب و تزکیه نفس
۷. زواره، علی (۱۳۹۲) کاربرد داده کاوی بروی داده‌های آموزش عالی دانشجویان دندانپزشکی شهر رشت با استفاده از تکنیک‌های طبقه بندی و خوشه بندی، اولین کنفرانس ملی رویکردهای نوین در مهندسی کامپیوتر و بازیابی اطلاعات ایران.
۸. سنایی نسب، هرمز، رشیدی جهان، حجت و صفاری، محسن (۱۳۸۹). عوامل مؤثر بر پیشرفت تحصیلی دانشجویان، دو ماهنامه راهبردهای آموزش در علوم پزشکی، سال پنجم، شماره ۴، ص ۲۴۳-۲۴۹
۹. محمد زاده، سپیده و محمد زاده، هانیه (۱۳۹۱). ارزیابی کیفیت برنامه‌های آموزشی در دانشگاه‌ها تسط داده کاوی، اولین همایش‌های ملی کاربردهای سیستم‌های هوشمند در علوم صنایع، دانشگاه آزاد اسلامی واحد قوچان،
۱۰. مرادی توپسرکانی، م. نصیری، پیشبینی دوره‌ی فارغ‌التحصیلی دانش‌آموزان با استفاده از تکنیکهای داده‌کاوی، اولین کنفرانس ملی دانش‌پژوهان کامپیوتر و فناوری اطلاعات، ۱۳۹۱
۱۱. موسوی، زهرا السادات و خان‌بابایی، محمد (۱۳۹۲) ارائه روشی مبتنی بر تکنیک ترکیبی داده کاوی جهت هدایت تحصیلی دانش‌آموزان، اولین همایش ملی مدیریت کسب و کار.
۱۲. میرزا خانی، نوید؛ صادقی، اسماعیل؛ یاسایی، راما و طباطبایی، سید مهدی (۱۳۸۹) بررسی هوش هیجانی دانشجویان کارشناسی جدید ورود دانشگاه توانبخشی دانشگاه علوم پزشکی شهید بهشتی
۱۳. نجیمی، آرش؛ مکفتاق، سید داوود و شریفی راد، غلامرضا (۱۳۹۰). عوامل مؤثر بر افت تحصیلی، از دیدگاه دانشجویان دانشگاه علوم پزشکی، مجله تحقیقات نظام سلامت، سال هفتم، شماره ۶، ص ۷۹۰-۷۹۷
14. Borkar, S and Rajeswari, K., (2013) "Predicting Students Academic performance Using Education Data Mining", IJCSMC, 2 (7), pg.273 – 27,
15. Kumar, Y.S and P.Saurabh,(2012) "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification "World of Computer Science and Information Technology Journal (WCSIT), 2 (2), PP,51-56,

16. Lazin, R and Neumann, L (1991)“Student characteristics as predictors of drop-out from medical school : admission to Beer-sheva over a decade”, *MedEduc*; 25(5): 390-404
17. Meilman, P.M, Pattis, J.A and Kraus –Zeilmann, D., (1994)“Suicide attempts and threats on one college campus: policy and practice”, *J Am coll Health*; 42(4), PP.147-54,
18. Ogor, E.N (2007)“ Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques”, *Proc. Electron., Robot. Automotive Mech. Conf.* , pp.354 - 359 ,
19. Superby, J.F and Meskens , J.P , (2006). Determination of factors influencing the achievement of the first-year university students using data mining methods, *Proc. Int. Conf. Intell. Tutoring Syst. Workshop Educ. Data Mining* , pp.1 -8.
20. Zlatko N J.K(2010),“Early Prediction of Student Success: Mining Students Enrolment Data” *Proceedings of Informing Science & IT Education Conference ( InSITE )* ,