



Investigating Novel Applications of Data Mining in Industry and Business

A.R Salehzadeh ^{*,1}

¹ Department of Computer Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran

ABSTRACT

Received: 7 June 2022

Accepted: 25 September 2022

KEYWORDS:

Data Mining

Knowledge Management

Distributed Processing

Database

Information Technology

In today's technology-oriented world, we are always faced with the processing of a huge amount of data as a major challenge in various fields of application. Therefore, we need knowledge discovery tools for such massive processing. Data mining is used as an advanced capability in data analysis and knowledge discovery. Data mining is the process of discovering patterns and regular and hidden trends in large and distributed data, using a wide set of algorithms based on mathematical and statistical sciences. In general, data mining can be considered the result of the natural evolution of information technology, which originated from the evolution of the database industry. The main reason for the emergence of data mining science has been the availability of a large amount of data and the strong need to extract useful knowledge and information from these data. Due to the importance of this issue, in this article we reviewed the important aspects of the data mining problem including goals, applications, models and related tools.

¹ Corresponding author

 a.r.salehzadeh@gmail.com



NUMBER OF REFERENCES

15



NUMBER OF FIGURES

0



NUMBER OF TABLES

0

بررسی کاربردهای نوین داده کاوی در صنعت و تجارت

امیررضا صالح زاده^{۱*}

^۱ گروه مهندسی کامپیوتر، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

چکیده

در دنیای فناوری محور امروزی عموماً در حوزه های کاربردی مختلف همواره با پردازش حجم عظیمی از داده ها، به عنوان یک چالش اساسی روبرو هستیم. لذا، برای چنین پردازش های حجیمی به ابزارهای کشف دانش نیاز داریم. داده کاوی به عنوان یک توانایی پیشرفته در تحلیل داده و کشف دانش مورد استفاده قرار می گیرد. داده کاوی عبارت است از فرآیند اکتشاف الگو و روندهای منظم و پنهان در داده های بزرگ و توزیع شده، با استفاده از مجموعه وسیعی از الگوریتم های مبتنی بر علوم ریاضی و آمار. به طور کلی داده کاوی را می توان نتیجه سیر تکاملی طبیعی فناوری اطلاعات دانست که این سیر نشأت گرفته از تکامل صنعت پایگاه داده است. اصلی ترین علت بوجود آمدن علم داده کاوی در دسترس بودن حجم وسیعی از داده ها و نیاز شدید به استخراج دانش و اطلاعات مفید از این داده ها بوده است. به دلیل اهمیت این موضوع، در این مقاله به بررسی اجمالی ابعاد مهم مساله داده کاوی شامل اهداف، کاربردهای نوین داده کاوی در حوزه صنعت و تجارت، مدل ها و ابزارهای مرتبط پرداختیم.

واژگان کلیدی:

داده کاوی
مدیریت دانش
پردازش توزیعی
پایگاه داده
فناوری اطلاعات


تعداد مراجع
۱۵


تعداد شکل ها
♦


تعداد جداول
♦

مقدمه

در سال‌های اخیر، با رشد روزافزون قدرت محاسباتی رایانه‌ها و امکان دستیابی به نتایج حاصل از محاسبات پیچیده در مدت زمان کوتاه، سبب شده تا الگوریتم‌های پیشرفته ریاضی مورد توجه قرار گیرند. این الگوریتم‌ها با در نظر گرفتن ابعاد مختلف داده، به پالایش و تحلیل آن پرداخته و الگوهای پیچیده و غیرقابل شناسایی توسط روش‌های قدیمی را استخراج و ارائه می‌کنند. رایانه‌ها کمک کرده‌اند تا فرآیند استخراج، پالایش، پیش پردازش و مدل‌سازی داده‌ها و همچنین اعتبارسنجی یافته‌ها با دقت بیشتر و سرعتی بی‌نظیر انجام شود. داده‌کاوی عبارت است از فرآیند اکتشاف الگو و روندهای منظم و پنهان در داده‌های بزرگ و توزیع شده، با استفاده از مجموعه وسیعی از الگوریتم‌های مبتنی بر علوم ریاضی و آمار. این الگوریتم‌ها معمولاً بر روی مقادیر عددی اعمال می‌شوند و برای داده‌های متنی، از الگوریتم‌های متن‌کاوی استفاده می‌شود [۱].

داده‌کاوی از علومی مانند هوش مصنوعی، یادگیری ماشینی، آمار، پژوهش عملیاتی و مدیریت پایگاه‌های داده برای ساخت مدل‌ها و پاسخ به سوالات دانش محور بهره می‌برد. داده‌ها امروز نقش زیادی در موفقیت یک کسب‌وکار و پیش‌برد اهداف یک سازمان دارد. سازمان‌ها به متخصصان علوم داده نیاز دارند تا با استفاده از تحلیل و پردازش داده افق‌های روش‌تری را پیش‌روی آنها قرار دهند. بنابراین نقش داده‌کاوی و تحلیل داده در آینده خدمات سازمان‌ها بسیار حیاتی است [۲]. این در حالی است که هم‌اکنون بسیاری از مدیران شرکت‌ها و کسب‌وکارها درک چندانی از این مفاهیم ندارند. در این مقاله قصد داریم تا به مفهوم داده‌کاوی، مزایا و پیش‌نیازهای آن بیشتر اشاره کنیم و کاربردهای مهم این علم در حوزه صنعت و تجارت را بررسی نمائیم. همانطور که اشاره شد، داده‌کاوی به معنای استخراج دانش از حجم عظیمی از داده‌ها و همچنین کشف الگوهای پنهان در میان انبوهی از داده‌ها می‌باشد. این رویکرد امکان استخراج روابط معنایی و عمقی را از انبارهای داده فراهم می‌نماید. وقتی در انباره‌ای حجم عظیمی از داده وجود داشته باشد، داده‌کاوی مفهوم ملموس‌تری می‌یابد. درحقیقت بیشتر الگوریتم‌های داده‌کاوی نیاز به حجم بالایی از داده جهت ساخت و آموزش مدل دارند. بمنظور ایجاد ساختارهای پردازشی روی داده‌های حجیم، هرکدام از روشهای داده‌کاوی یک یا تعداد بیشتری از انواع مدل‌سازی داده را بنا به نیازمندی‌های عملیاتی و اجرایی ذخیره و بازیابی داده‌ها انجام می‌دهند. انواع مدل‌سازی عبارتند از: مدل‌سازی پیوستگی، دسته‌بندی، خوشه‌بندی، پیش‌بینی، رگرسیون، تشخیص توالی و مصورسازی و ... [۳-۴]. برای هرکدام از مدل‌های نام برده شده، روشهای یادگیری ماشین متعددی وجود دارد که

انتخاب آنها بستگی به ویژگیهای داده و نیاز کسب و کار دارد. از کاربردهای اولیه و مهم داده‌کاوی در حوزه صنعت و تجارت، استفاده از آن در مدیریت ارتباط با مشتری است که در آن داده‌کاوی به عنوان یک ابزار بنیادی برای آشکارسازی خصوصیات جمعیت شناختی مشتریان محسوب می‌شود و با بکارگیری فرایندهای پردازشی مناسب می‌توان پرس و جوها و گزارش‌گیری‌های آماری سلسله‌مراتبی از پایگاه داده‌های توزیعی را ایجاد نمود. داده‌کاوی در هر زمینه‌ای که به آن نیاز است، می‌تواند کاربرد داشته باشد. امروزه دیتا ماینینگ در مواردی مانند موارد زیر درحوزه صنعت و تجارت قابل بکارگیری می‌باشد [۵-۷]:

- تحقیقات بازار خرید مشتریان: این مبحث که به نوعی کاربرد داده‌کاوی در مدیریت است در پی شناسایی کالاهای مرتبط با سبد خرید مشتری است تا امکان خرید آن‌ها را افزایش دهد.
- مدیریت پرس و جو و گزارش‌گیری از پایگاه‌های داده: با بکارگیری روش‌های مناسب و ابزارهای بنیادین این حوزه می‌توان مدیریت موثر جستجو، بازیابی و گزارش اطلاعات موردنیاز از پایگاه‌های داده توزیعی و تجمیعی حجیم را ایجاد نمود.
- آموزش: فعالیت این زمینه در جهت بهبود کیفیت سیستم آموزشی و هدایت صحیح کاربران سیستم‌های تجاری و مالی در حوزه صنعت و تجارت می‌باشد.
- ساخت و عمران: تلاش این حوزه در جهت تسهیل راه‌سازی و الگوهای بهینه‌عمرانی در حوزه صنعت است.
- مدیریت ارتباطات مشتریان: اهداف اصلی این حوزه بهبود روابط مشتریان با شرکت‌ها و افزایش بهره‌وری در حوزه‌های اقتصادی و تجاری است.
- مدیریت و جلوگیری از حملات الکترونیکی: به منظور شناسایی الگوریتم‌های حمله به سیستم‌های صنعتی و تجاری مورد استفاده قرار می‌گیرد.
- تحقیقات جرم‌شناسی: از داده‌کاوی می‌توان برای بررسی ارتباطات بین کارکردهای مجرمانه و خلاف قوانین بخصوص در حوزه‌های تجاری استفاده کرد.

لازم به ذکر است، اساساً برای بکارگیری استفاده از داده کاوی در صنایع مختلف فرایند و مرحله‌ای تعریف شده است که در ادامه به آنها اشاره می‌کنیم [۸-۹]:

- ثبت و مستندسازی داده‌ها

در اولین قدم نیاز است تا اطلاعاتی که در زمینه مورد نظر در یک دوره زمانی ثبت شده است را در اختیار داشته باشیم، که در این مرحله لازم است داده‌ها ابتدا جمع‌آوری و ثبت شده و سپس جمع‌آوری و مستندسازی گردند.

- پیش پردازش و تبدیل به داده‌های قابل پردازش

برای این که بتوان از داده‌های جمع‌آوری شده از پایگاه‌های داده‌های مختلف به خوبی استفاده کرد نیاز است که داده‌ها پیش‌پردازش شوند و تنها داده‌های مورد نیاز برای پردازش نهایی مورد استفاده قرار گیرند. این مرحله بیشترین زمان را در فرآیند داده کاوی به خود اختصاص می‌دهد.

- مشخص کردن ویژگی‌های داده کاوی

در این گام لازم است اطلاعاتی که باید فرآیند داده کاوی بر اساس آنها انجام گیرد را مد نظر قرار داد و ویژگی‌ها و چارچوب داده کاوی را براساس نوع کاربرد مشخص نمود.

- اجرای عملیات داده کاوی و یافتن الگو

در این مرحله فرایند داده کاوی و بررسی‌های لازم براساس ساختار داده‌ها انجام می‌گیرد. در میان داده‌هایی که جمع‌آوری شده است باید به دنبال یک الگوی مشخص بود که در اکثر مناطق مورد نظر رعایت شده باشد. کشف الگو عمدتاً براساس الگوریتم‌های احتمالی تشخیص الگو صورت می‌پذیرد.

- تفسیر نتایج داده‌ها و استخراج نتیجه تحلیل

بعد از این که الگوهای رفتاری را به خوبی بررسی کرده و الگوهای پیشنهادی را ارائه کردیم، نوبت به تفسیر نتایج می‌رسد. در این بخش باید نتایج به دست آمده به خوبی تجزیه و تحلیل شوند و شکل نهایی تحقیق برای اجرایی شدن آماده شود.

نکته قابل تامل در این حوزه این است که، اساساً داده کاوی پیشرفت قابل ملاحظه‌ای را در نوع ابزارهای تحلیل موجود نشان می‌دهد اما محدودیت‌هایی نیز دارد. یکی از این محدودیت‌ها این است که با وجود اینکه داده‌سازی به آشکارسازی الگوها و روابط کمک می‌کند اما اطلاعاتی را درباره ارزش یا میزان اهمیت آن‌ها به دست نمی‌دهد. دومین محدودیت آن این است که با وجود توانایی شناسایی روابط بین رفتارها یا متغیرها لزوماً قادر به کشف روابط علت و معلولی نیست. لذا، موفقیت داده کاوی عموماً در گرو بهره‌گیری از کارشناسان فنی و تحلیل گرانی است که از توانایی کافی برای طبقه‌بندی تحلیل‌ها و تغییر و مدیریت آن داده‌ها

برخوردار هستند [۱۰]. در بخش بعدی مقاله حاضر اهداف اصلی، روش‌های و مدل‌های مهم داده کاوی در حوزه صنعت و تجارت را بررسی می‌نمائیم.

اهداف، روش‌ها و مدل‌ها

داده کاوی در حوزه صنعت و تجارت با اهداف، روش‌ها و مدل‌های مختلفی بکار گرفته می‌شود. در این راستا اولین گام شناخت اهداف داده کاوی در زمینه موردنظر است. اساساً اهداف مهم داده کاوی شامل پیش‌بینی و توصیف الگوهای عملکردی یا ترکیبی از آنهاست. هدف پیش‌بینی تمرکز بر روی دقت در توانایی پیش‌بینی بوده و "توصیف" بر درک فرآیند تولید داده‌ها تمرکز دارد. در پیش‌بینی تا زمانی که مدل قدرت پیش‌بینی دارد، کاربر توجهی به این ندارد که مدل انعکاس دهنده واقعیت است؛ مثلاً مدلهایی که شاخصهای مالی در حوزه‌های تجاری را به شکل غیر خطی ترکیب می‌کند تا نرخ تبادل ارز را پیش‌بینی نمایند. از سویی دیگر مدل توصیفی به مثابه واقعیت تفسیر میشود [۱۱]. برای مثال مدلهایی که متغیرهای اقتصادی و جمعیتی را به پیشرفت‌های آموزشی مرتبط می‌سازد، به عنوان مبنایی برای توصیه‌های سیاست اجتماعی توصیف میشوند. در عمل، اغلب کاربردهای اکتشاف دانش به درج‌های از هر دو مدل سازی توصیفی و پیش‌بینی نیاز دارند. به هر ترتیب، آنچه در این حوزه اهمیت دارد این است که اهداف داده کاوی با استفاده از روشهای داده کاوی، محقق میشوند. اصطلاح روشهای داده کاوی در واقع بیانگر جمع‌کثیری از الگوریتمها و فنون است که از علوم مانند آمار، یادگیری ماشین، پایگاه داده و تجسم سازی استنتاج شده اند. برخی از این روشها عمدتاً در مورد داده‌های مالی و درحوزه‌های تجاری به کار رفته‌اند. روشهای داده کاوی مشهور، شامل شبکه‌های عصبی، الگوریتمهای ژنتیک درختان تصمیم، نظریه مجموعه اولیه، استدلال مبتنی بر مورد، استدلال موردگرا و برنامه‌نویسی ریاضی هستند [۱۲-۱۳]. بدیهی است که این فهرست از روشها، فهرستی کامل نبوده و ترتیبی برای اولویت بندی در کاربرد این روشها، پیشنهاد نمی‌شود.

درحوزه صنعت و تجارت اساساً، هرکدام از عناصر و اجزاء داده کاوی می‌توانند با یک یا چند روش داده کاوی مدلسازی شوند. عمده‌ترین مدل‌هایی که در این حوزه‌ها بکار گرفته میشوند در ادامه به صورت خلاصه مرور شده‌اند [۱۴]:

- پیوستگی

عموماً قوانین تلازمی به فرایند کشف وابستگی‌ها میان حجم عظیمی از داده‌ها و نمایش آنها به فرم عبارتهای "اگر-آنگاه" اطلاق میشود. هدف این قوانین پیدا کردن بیشترین روابط علت و معلولی

بین داده هاست. این قوانین، از داده و روابط منطقی حاکم بر آنها استنباط میشوند که طبعاً احتمالی هستند. هدف پیوستگی، برقراری ارتباط بین عناصری است که با یکدیگر در یک تراکنش گرد هم آمده اند. در حوزه تجارت برنامه های فروش جانبی و فروش بیشتر، مثالهای معمول از به کارگیری پیوستگی میباشد. برای نمونه، با استفاده از پیوستگی، سازمانهای کاتالوگ محور، کاتالوگهای فروش محصولاتی خاص را به مشتریانی ارائه میدهند که احتمال خرید آن توسط شخص با توجه به محصولات خریداری شده، از یک مقدار آستانه بیشتر باشد.

- دسته بندی

هدف دسته بندی ایجاد مدلی برای پیشبینی رفتار مشتریان بر اساس طبقه بندی رکوردهای داده به دسته های از پیش تعیین شده است. دسته بندی شامل بررسی داده جدید و انتصاب آن به یکی از دسته های شناخته شده قبلی روشهای رایج در این زمینه شامل شبکه عصبی، درخت تصمیم و قواعد اگر-آنگاه هستند. دسته بندی شامل الگوریتم های با نظارت جهت ایجاد رده های مرجع می باشد.

- خوشه بندی

روشهای خوشه بندی داده ها را به بخش های همگن تقسیم می کنند، به گونه ای که داده های موجود در یک خوشه دارای حداکثر شباهت با یکدیگر بوده و فاصله بین خوشه ها نیز حداکثر باشد. تفاوت خوشه بندی با دسته بندی در این است که خوشه ها دارای عنوان از پیش تعیین شده نمی باشند.

- پیش بینی

روشهای پیش بینی، مقادیر آینده را بر اساس الگوهای کشف شده تاکنون تخمین میزنند. پیش بینی نیاز مشتریان یک مثال معمول از یک مدل پیش بینی به شمار میرود. روشهای رایج پیش بینی شامل شبکه عصبی و سری های زمانی میباشد. داده کاوی معمولاً با آنچه در طول زمان اتفاق می افتد، گره خورده است. تحلیل سری زمانی نیاز به انتخاب فریم زمانی مناسب برای داده دارد و علاوه بر آن، مشخص کردن نقطه آغاز بررسی نیز حائز اهمیت است. مرحله بعد، رسم سری زمانی است چراکه نمودار سری زمانی اطلاعات مفید مختلفی را فراهم میکند، هرچند ایده های در رابطه با اینکه آیا تغییرات در طول زمانی مورد انتظار بوده یا نبوده است نمی دهد.

- رگرسیون

رگرسیون نوعی روش آماری جهت نگاشت هر داده به یک مقدار واقعی است تا بتوان مقدار یک ویژگی را بر اساس مقدار سایر ویژگیها تخمین زد. کاربردهای رگرسیون شامل برازش منحنی، پیش بینی، مدلسازی روابط علی و آزمایش فرضیات علمی در رابطه با روابط بین متغیرهاست. تشخیص توالی، شناسایی پیوستگیها یا الگوها در طول

زمان است. هدف این رویکرد، مدلسازی فرآیند ایجاد کننده دنباله و یا استخراج و گزارش انحراف و روند در طول زمان میباشد.

- مصورسازی

هدف روشهای مصورسازی، نمایش مناسب داده ها به گونه ای است که الگوهای پیچیده درون آن استخراج شود. این روشها در ترکیب با سایر مدل ها جهت شفاف سازی الگوی استخراج شده یا روابط به کار میرود. بررسی داده ها با استفاده از ابزارهای مصورسازی به ایجاد درک اولیه از داده، جهت انتخاب متغیرها و انتخاب مکانیزم و روش مناسب داده کاوی کمک میکند.

بمنظور تکمیل بحث و ارائه رویکرد کاربردی در انتهای این مقاله مروری بر ابزارهای اصلی داده کاوی خواهیم داشت. لازم به ذکر است، استخراج دستی داده ها از صدها سال پیش وجود داشته است. این در حالی است که بخصوص در سال های اخیر ماشینی شدن فرآیند داده کاوی او بکارگیری ابزارهای مرتبط در جهت تسریع فرایندهای داده کاوی شایع شده است. ابزارهای داده کاوی نرم افزارهایی هستند که به کاربران اجازه استخراج اطلاعات از داده ها را بصورت مکانیزه میدهند. این ابزارها توانایی گردآوری داده ها و کاربرد آنها به منظور تصمیم گیری در خصوص مصرف کننده خاص یا گروهی از مصرف کنندگان را، به سازمانها و افراد می دهند. هدف نهایی این ابزارها ظاهر ساختن الگوهای پنهان میباشد. به هر حال ابزارهای داده کاوی میتواند شامل نرم افزارهای رایج زیر باشد:

- 1- Data to knowledge
- 2- Sas25
- 3- Clementine
- 4- Intelligent-miner
- 5- Insightful miner
- 6- Weka
- 7- Microsoft excel
- 8- ACL29

کارکرد و نقاط قوت و ضعف هریک از نرم افزارهای نام برده شده براساس حوزه کاربرد، حجم اطلاعات، میزان پراکندگی و نوع الگوهای داده کاوی مورد استفاده تعیین می گردد. لازم به ذکر است، در حالیکه محصولات داده کاوی ابزارهای قدرتمندی می باشند، اما در نوع کاربردی کافی نیستند. برای کسب موفقیت، داده کاوی نیازمند تحلیل گران حرفه ای و متخصصان ماهر می باشد که بتوانند ترکیب خروجی به وجود آمده را تحلیل و تفسیر نمایند. در نتیجه محدودیت های داده کاوی مربوط به داده اولیه یا افراد است تا اینکه مربوط به تکنولوژی باشد. همچنین درنهایت باید تاکید نمود، اگرچه داده کاوی به الگوهای مشخص و روابط آنها کمک می کند، اما برای کاربر اهمیت و ارزش این الگوها را بیان نمی کند. تصمیماتی از این قبیل بر عهده خود کاربر است.

نتیجه گیری

به طور کلی داده کاوی را می توان نتیجه سیر تکاملی طبیعی فناوری اطلاعات دانست که این سیر نشأت گرفته از تکامل صنعت پایگاه داده است. اصلی ترین علت بوجود آمدن علم داده کاوی در دسترس بودن حجم و سیعی از داده ها و نیاز شدید به استخراج دانش و اطلاعات مفید از این داده ها بود. این دانش و اطلاعات کاربرد وسیعی در مدیریت کسب و کار و کنترل تولید، تحقیقات علمی، تحلیل بازار و طراحی های مهندسی دارد. این سیر شامل فرآیندهای جمع آوری داده ها، ایجاد پایگاه داده، مدیریت داده و تحلیل و درک داده ها می باشد. تکامل فناوری پایگاه داده و کاربرد فراوان آن در حوزه های مختلف موجب جمع آوری حجم انبوهی از داده ها شده است. این حجم از داده ها نیازمند ابزارهای قدرتمندی برای تحلیل آنها می باشد، زیرا در دنیای کنونی از نظر داشتن داده غنی هستیم اما کمبود اطلاعات داریم. به دلیل اهمیت موضوع، در این مقاله به بررسی اجمالی ابعاد مهم مساله داده کاوی شامل اهداف، کاربردهای نوین داده کاوی در حوزه صنعت و تجارت، مدل ها و ابزارهای مرتبط پرداختیم.

منابع

- introduction to data mining (Vol. 4). John Wiley & Sons.
- [6] Mughal, M. J. H. (2018). Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, 9(6).
 - [7] Kumar, T. S. (2020). Data mining based marketing decision support system using hybrid machine learning algorithm. *Journal of Artificial Intelligence*, 2(03), 185-193.
 - [8] Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1355.
 - [9] Ageed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Rashid, Z. N., Salih, A. A., & Abdullah, W. M. (2021). A survey of data mining implementation in smart city applications. *Qubahan Academic Journal*, 1(2), 91-99.
 - [10] Yan, H., Yang, N., Peng, Y., & Ren, Y. (2020). Data mining in the construction industry: Present status, opportunities, and future trends. *Automation in Construction*, 119, 103331.
 - [11] Jalota, C., & Agrawal, R. (2019, February). Analysis of educational data mining using classification. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 243-247). IEEE.
 - [12] Shin, D., & Shim, J. (2021). A systematic review on data mining for mathematics and science education. *International Journal of Science and Mathematics Education*, 19(4), 639-659.
 - [13] Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining
 - [1] Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., ... & Lyu, J. (2020). Brief introduction of medical database and data mining technology in big data era. *Journal of Evidence-Based Medicine*, 13(1), 57-69.
 - [2] Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, 166, 114060.
 - [3] Roiger, R. J. (2017). Data mining: a tutorial-based primer. Chapman and Hall/CRC.
 - [4] Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118.
 - [5] Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an*

techniques applied in educational environments: Literature review. *Digital Education Review*, (33), 235-266.

- [14] Hernández-Nieves, E., Parra-Domínguez, J., Chamoso, P., Rodríguez-González, S., & Corchado, J. M. (2021). A Data Mining and Analysis Platform for Investment Recommendations. *Electronics*, 10(7), 859.
- [15] Islam, M. S., Hasan, M. M., Wang, X., Germack, H. D., & Noor-E-Alam, M. (2018, May). A systematic review on healthcare analytics: application and theoretical perspective of data mining. In *Healthcare* (Vol. 6, No. 2, p. 54). MDPI.