

مقاله علمی-پژوهشی

## مقایسه‌ی الگوریتم‌های داده‌کاوی در پیش‌بینی نتیجه‌ی مناقصات بین‌المللی برگزار شده برای انتخاب مشاوران صنعت آب (به‌منظور استفاده در سیستم‌های پشتیبان تصمیم‌گیری)

هادی شکبیا زاهد<sup>۱</sup>، مجید پرچمی جلال<sup>۲\*</sup>، محمدمین قادری<sup>۳</sup>

تاریخ دریافت: ۱۳۹۹/۲/۲۸ تاریخ پذیرش: ۱۳۹۹/۴/۳۱

### چکیده

در کشورهای درحال توسعه با ارزش پولی پایین، شرکت‌های مشاوره برای کسب سود بیشتر به دنبال شرکت کردن در پروژه‌های بین‌المللی می‌باشند. اما شرکت در مناقصات بین‌المللی نیازمند صرف منابع زیادی (زمان، هزینه و منابع انسانی) برای خرید اسناد مناقصه، ارزیابی شرایط پروژه و آماده کردن پیشنهاد مناقصه است. از این رو پیش‌بینی نتیجه مناقصات بین‌المللی می‌تواند باعث جلوگیری از مصرف منابع برای شرکت در مناقصات نامناسب شود. هدف از این پژوهش شناسایی معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزار شده برای انتخاب مشاوران صنعت آب و مقایسه الگوریتم‌های داده‌کاوی در پیش‌بینی نتیجه این مناقصات است. معیارهای شناسایی شده شامل نوع امور تخصصی، نحوه ارسال پیشنهاد مناقصه، چگونگی آشنایی با مناقصه، نوع مناقصه، تأمین‌کننده مالی، داشتن همکار در کشور مبدأ و نوع پروژه می‌باشند و الگوریتم‌های مقایسه شده به ترتیب دقت شامل ماشین بردار پشتیبانی، ID3، Chaid، درخت تصمیم، بیز ساده و نزدیک‌ترین همسایه هستند. از میان الگوریتم‌های مذکور، الگوریتم ماشین بردار پشتیبانی با دقت ۸۹،۳۱ درصد دارای بیشترین دقت می‌باشد؛ بنابراین پیشنهاد می‌شود که از این الگوریتم به‌عنوان پردازشگر در سیستم‌های پشتیبان تصمیم مناقصات استفاده شود تا موجب بهبود تصمیم به شرکت/عدم شرکت مشاوران در مناقصات بین‌المللی صنعت آب شود.

**واژه‌های کلیدی:** الگوریتم‌های دسته‌بندی، تصمیم‌سازی، داده‌کاوی، مشاوران، مناقصات صنعت آب

### مقدمه

صنعت آب تصمیم می‌گیرد تا مشاوره را از طریق مناقصه انتخاب کند، یکی از مهم‌ترین تصمیم‌ها برای شرکت‌های مشاوره، تصمیم به شرکت یا عدم شرکت در این مناقصات می‌باشد. بهره‌وری این تصمیم مهم را می‌توان با بهره‌گیری از سیستم‌های پشتیبان تصمیم<sup>۴</sup>، ارتقا داد.

فرآیند شرکت کردن مشاوران در مناقصات بین‌المللی صنعت آب شامل چهار سناریوی محتمل می‌باشد که در شکل ۱ نمایش داده شده است. هدف از این پژوهش، شناسایی معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزار شده برای انتخاب مشاوران صنعت آب و مقایسه الگوریتم‌های داده‌کاوی در پیش‌بینی نتیجه این مناقصات است. نتایج این پژوهش در توسعه سیستم پشتیبان تصمیم مناقصات برای شرکت‌های مشاوره در صنعت آب کاربرد دارد و سیستم پشتیبان تصمیم توسعه داده شده بر اساس نتایج این پژوهش می‌تواند موجب بیشینه‌سازی احتمال رخداد سناریوی اول (مطلوب‌ترین سناریو) شود. این سناریو در شکل ۱ با خط‌چین نمایش داده شده است.

مطابق پژوهش جاسیزیک و همکاران، یکی از استراتژی‌های رایج شرکت‌های مشاور فعال در صنعت آب در کشورهای درحال توسعه با ارزش پولی پایین، شرکت کردن در مناقصات بین‌المللی می‌باشد، زیرا درآمد چشمگیری به همراه دارد و این امکان را فراهم می‌کند تا فعالیت در بازارهای جدید را تجربه کنند. شرکت نکردن در مناقصات بین‌المللی می‌تواند موجب از دست دادن یک فرصت برای کسب درآمد چشمگیری باشد و معمولاً به‌عنوان یک تهدید شناخته می‌شود (Juszczyk et al., 2014). با این وجود، شرکت کردن در مناقصات بین‌المللی نامناسب می‌تواند موجب هدر دادن منابع و حتی شکست سازمان شود. بنابراین هنگامی که یک کارفرما در

۱- دانشجوی دکتری، گروه مدیریت پروژه و ساخت، دانشگاه تهران، تهران، ایران

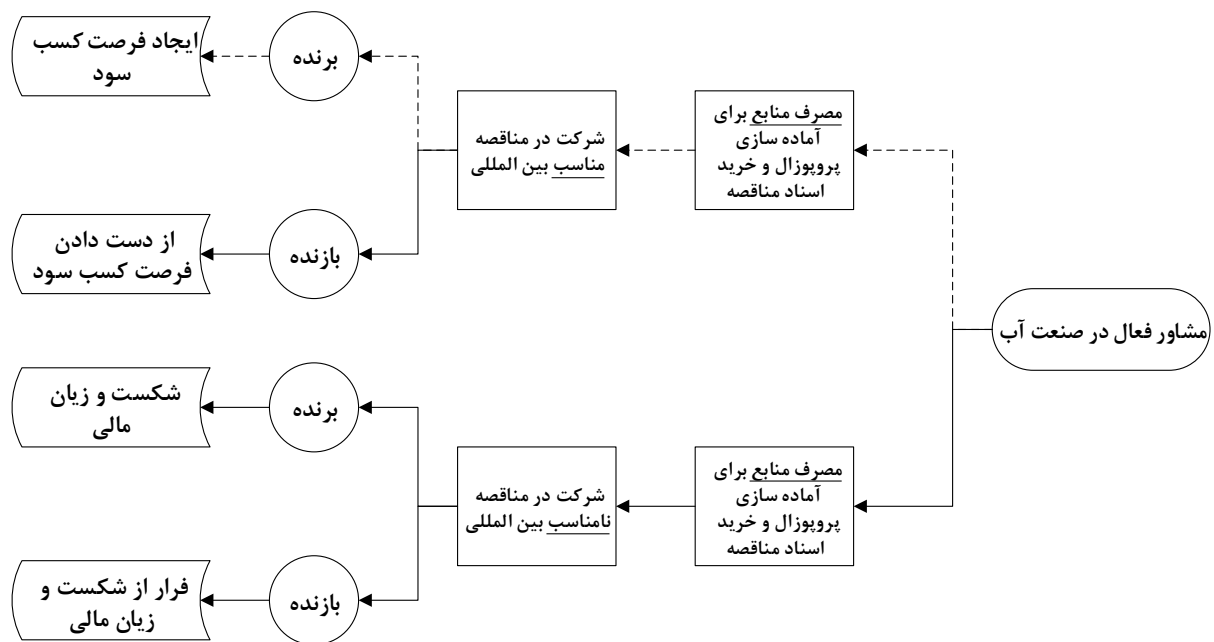
۲- دانشیار، گروه مدیریت پروژه و ساخت، دانشگاه تهران، تهران، ایران

۳- دانشجوی کارشناسی ارشد، گروه مدیریت پروژه و ساخت، دانشگاه تهران،

تهران، ایران

(Email: parchamijalal@ut.ac.ir

\*) نویسنده مسئول:



شکل ۱- سناریوهای محتمل هنگام شرکت کردن مشاوران در مناقصات بین‌المللی صنعت آب

را با بررسی پژوهش‌های قبلی شناسایی کردند. آن‌ها با ضریب همبستگی<sup>۳</sup> به‌عنوان یک شاخص آماری نشان دادند که تنها هشت معیار از ۲۱ معیار شناسایی شده ارتباط خطی قابل توجهی با تصمیم به شرکت در مناقصه دارند و با استفاده از این هشت معیار، یک مدل رگرسیون لجستیک با دقت ۹۶ درصد برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات، توسعه دادند (Lowe and Parvar, 2004). اوو و همکاران دو مدل مبتنی بر رگرسیون برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات ارائه کردند. مدل اول با استفاده از معیارهای مرتبط با مشخصات پروژه (مانند نوع و ابعاد پروژه) و محیط تجاری (مانند تعداد رقبا و شرایط تجاری) و مدل دوم با استفاده از معیارهای مرتبط با مشخصات کارفرما (مانند تجربه و ابعاد سازمان کارفرما) توسعه داده شدند، تا تأثیر هر یک از این معیارها در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات مشخص شود (Oo et al., 2007). تن و همکاران استراتژی‌های رقابتی در مناقصات صنعت ساخت هنگ کنگ را بررسی کردند. آن‌ها سیزده استراتژی، فراوانی استفاده هر یک از آن‌ها و تأثیر این استراتژی‌ها در پیروزی پیمانکاران را تشریح کرده و پژوهش آن‌ها افق‌های جدیدی درباره‌ی استراتژی‌های رقابتی پیمانکاران در مناقصات را آشکار کرد (Tan et al., 2010). چنگ و همکاران معیارهای کلیدی مؤثر در تصمیم به شرکت/عدم شرکت در

وانوس و همکاران یک مدل پارامتریک برای تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات ارائه کرده‌اند. آن‌ها ۳۸ معیار مؤثر در تصمیم به شرکت/عدم شرکت در مناقصه را از طریق مصاحبه و پرسشنامه شناسایی کردند، ۲۰ معیار با اهمیت کمتر را حذف کردند و ۱۸ معیار باقیمانده را به دو گروه معیارهای مثبت و معیارهای منفی تقسیم‌بندی کردند. با استفاده از این معیارها، دیدگاه خبرگان در مورد تصمیم به شرکت/عدم شرکت در مناقصات، به یک مدل ساختاریافته تبدیل می‌شود (Wanous et al., 2000). وانوس و همکاران یک مدل مبتنی بر شبکه‌های عصبی مصنوعی<sup>۱</sup> برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات توسعه داده‌اند. مدل آن‌ها هشت گره ورودی، دولایه مخفی و یک گره خروجی دارد. آن‌ها معیارهای کلیدی مؤثر در تصمیم به شرکت/عدم شرکت در مناقصات را با استفاده از پرسشنامه شناسایی کرده و به‌عنوان گره‌های ورودی تعریف کردند. داده‌های ۱۵۷ مناقصه برگزار شده در کشور سوریه برای یادگیری مدل و ۲۰ پروژه جدید برای ارزیابی مدل مورد استفاده قرار گرفتند. دقت مدل آن‌ها تقریباً ۹۰ درصد بوده است (Wanous et al., 2003). لاو و پروار با بهره‌گیری از رگرسیون لجستیک<sup>۲</sup> یک مدل برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات توسعه دادند. آن‌ها ۲۱ معیار مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات

1- Artificial Neural Network  
2- Logistic Regression

3- Correlation

کارهای فعال، تجربه در پروژه‌های مشابه، ابعاد پروژه، شهرت کارفرما، وضعیت و ثبات مالی کارفرما، وجود مناقصات مشابه، سرعت و توانایی کارفرما در پرداختی‌ها و جزئیات اسناد مناقصه (Jarkas et al., 2014). پولات و همکاران با مرور پژوهش‌های قبلی و استفاده از پرسشنامه، ۵۲ معیار مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت را شناسایی کردند و سپس یک سیستم پشتیبان تصمیم مبتنی بر شبکه عصبی استنتاجی توسعه دادند (Polat et al., 2014). ژانگ و همکاران شش مدل جدید برای پیش‌بینی ریسک‌های موجود در مناقصات صنعت ساخت ارائه کردند. مدل‌های آن‌ها از تئوری تصمیم‌سازی چندمعیاره برای مدیریت سود مورد انتظار پیمانکاران استفاده کرده‌اند (Zhang et al., 2014). الحمیدی یک مدل تاپسیس فازی برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت توسعه داد. این مدل با رتبه‌بندی مناقصات، از تصمیم‌گیرنده‌ها حمایت می‌کند. این مدل دو محدودیت دارد، محدودیت اول این است که حداکثر از ۱۰ تصمیم‌گیرنده پشتیبانی می‌کند و محدودیت دوم این است که حداکثر پنج پروژه را می‌تواند در نظر گرفته و رتبه‌بندی کند (AI-Humaidi, 2016). لزنیاک و رادزیجوسکا یک مدل تصمیم‌گیری چند معیاره ارائه کردند که پیمانکاران را قادر می‌سازد تا بهترین تصمیم را در شرکت/عدم شرکت در مناقصات اتخاذ کنند. مدل آن‌ها تلفیقی از روش مجموع ساده وزنی، تحلیل سلسله مراتبی و سازمان‌دهی رتبه‌بندی اولویت می‌باشد (Leśniak and Radziejowska, 2017). بیروک و همکاران مدلی برای پشتیبانی تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات ارائه کردند. مدل پیشنهادی آن‌ها شامل سه جز می‌باشد: یک مدل تصمیم‌گیری چند معیاره برای ارزیابی مطلوبیت مناقصه و دو مدل برنامه‌ریزی خطی، یکی برای پیش‌بینی هزینه کل پروژه و دیگری برای توزیع هزینه کل پروژه در بین اقلام لایحه مقادیر برای به حداکثر رساندن جریان نقدی پیمانکار (Biruk et al., 2017). چیسالا ۷۵ معیار مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت را شناسایی کرده، آن‌ها را به دو گروه معیارهای خارجی و معیارهای داخلی دسته‌بندی کرده است، با استفاده از مدل مجموع وزنی ساده، وزن این معیارها را تعیین کرده و درنهایت یک مدل پشتیبان تصمیم با دقت ۸۶ درصد توسعه داده است (Chisala, 2017). کالان و اوزبک با بهره‌گیری از تجزیه و تحلیل سلسله مراتبی، یک مدل پشتیبان تصمیم مناقصه برای پیمانکاران صنعت ساخت توسعه داده و شیوهی کارکرد این مدل را با بررسی دو نمونه موردی تشریح کردند (Kalan and Ozbek, 2020). پرچمی جلال و همکاران معیارهای مؤثر در تصمیم به شرکت/عدم شرکت مشاوران

مناقصات را شناسایی کرده و با استفاده از روابط ترجیحی فازی<sup>۱</sup>، وزن هر یک از معیارها را تعیین کردند. اگر تصمیم به شرکت در مناقصه باشد، مدلی را برای تعیین میانگین قیمت پیشنهادهای ارائه‌شده با تلفیق مدل چشم‌انداز چندمعیاره<sup>۲</sup> و نظریه چشم‌انداز تجمعی<sup>۳</sup> ارائه کردند (Cheng et al., 2011). کستانتینو و همکاران مدلی را برای پیش‌بینی روند قیمت‌های پیشنهادی در مناقصات عمومی قبل از بازگشایی پاکت پیشنهادها ارائه کردند. شرکت‌ها با استفاده از این مدل می‌توانند توزیع احتمال قیمت‌های پیشنهادی را تعیین کنند و سود مورد انتظار خود را با در نظر گرفتن سایر پیشنهادها و احتمال برنده شدن، به حداکثر برسانند (Costantino et al., 2011). مینلی و شانشان معیارهای مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات را شناسایی کردند و با بهره‌گیری از این معیارها به عنوان گره ورودی، یک شبکه عصبی مصنوعی پیش‌خور<sup>۴</sup> برای پشتیبانی از تصمیم به شرکت در مناقصات توسعه دادند (Minli and Shanshan, 2012). چوو و همکاران با تلفیق تحلیل سلسله مراتبی فازی<sup>۵</sup> و رگرسیون، یک سیستم پشتیبان تصمیم مناقصات توسعه دادند. در مدل آن‌ها از تحلیل سلسله مراتبی فازی برای تعیین وزن معیارهای مؤثر در هزینه پروژه و از رگرسیون برای تعیین قیمت پیشنهادی در مناقصه استفاده شده است (Chou et al., 2013). الماشالا یک چارچوب عملی برای تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت توسعه داد. این چارچوب دارای دو جز می‌باشد؛ جز اول در تعیین معیارهای کلیدی ارزیابی مناقصات کاربرد دارد و این معیارها را از طریق مصاحبه با خبرگان استخراج کرده است. جز دوم با بهره‌گیری از تجزیه و تحلیل پوششی داده‌ها، توصیه به شرکت/عدم شرکت در مناقصات می‌کند (El-Mashaleh, 2013). لزنیاک و پلبانکیویز مروری بر مدل‌های پشتیبان تصمیم به شرکت/عدم شرکت در مناقصات صنعت ساخت انجام دادند و سپس یک مدل پشتیبان تصمیم مبتنی بر منطق فازی توسعه دادند. مطابق یافته‌های پژوهش آن‌ها، سه معیار مهم در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت عبارت‌اند از نوع پروژه، تجربه پروژه‌های مشابه و شرایط قراردادی پروژه (Leśniak and Plebankiewicz, 2013). جرکاس و همکاران با بهره‌گیری از پرسشنامه‌های ساختاریافته، معیارهای مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات صنعت ساخت کشور قطر را شناسایی کردند. معیارهای شناسایی‌شده در پژوهش آن‌ها عبارت‌اند از تجربه ارتباط قبلی با کارفرما، میزان نیاز به شرکت در مناقصه جدید، حجم

- 1- Fuzzy Preference Relations
- 2- Multi-Criteria Prospect
- 3- Cumulative Prospect Theory
- 4- Back-Propagation
- 5- Fuzzy Analytic Hierarchy Process (AHP)

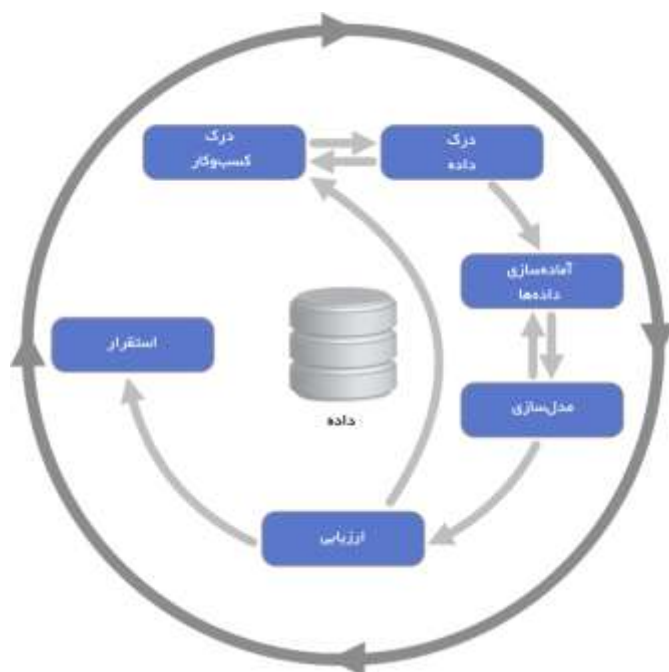
برای انتخاب مشاوران صنعت آب پرداخته است و همچنین در این پژوهش برای اولین بار، الگوریتم‌های داده‌کاوی برای پیش‌بینی نتیجه مناقصات مورد استفاده قرار گرفته‌اند.

### مواد و روش‌ها

در این پژوهش از متدولوژی CRISP-DM استفاده شده است. CRISP-DM یک متدولوژی صنعتی داده‌کاوی است که در این پژوهش با ایجاد تغییراتی مورد استفاده قرار گرفته است. این متدولوژی شش مرحله مختلف داده‌کاوی در پروژه‌های صنعتی را تشریح می‌کند. در این پژوهش از پنج مرحله استفاده شده، زیرا مرحله آخر (پیاده‌سازی و استقرار) کاربردی در فعالیتهای پژوهشی ندارد. مراحل متدولوژی CRISP-DM در شکل ۲ نمایش داده شده است.

در مناقصات بین‌المللی صنعت آب را با بهره‌گیری از دو ابزار مختلف شناسایی کردند. ابزار اول شامل تحلیل آماری پرسشنامه‌های ساختاریافته و ابزار دوم شامل متن‌کاوی می‌باشد. یافته‌های این پژوهش عبارت‌اند از پنج معیار اصلی و ۶۸ زیرمعیار. معیارهای اصلی به ترتیب اهمیت عبارت‌اند از وضعیت مالی، ملاحظات سازمان کارفرما، ویژگی‌های نیروی کار، ویژگی‌های پروژه، وضعیت رقبا و مفاد قرارداد (Parchamijalal et al., 2020).

همان‌طور که شرح داده شد، مقالات زیادی به بررسی معیارهای مؤثر در تصمیم به شرکت/عدم شرکت پیمانکاران در مناقصات پرداخته‌اند اما تعداد بسیار کمی، الگوریتم‌های پیش‌بینی را مقایسه کرده‌اند و همچنین اکثر پژوهش‌های قبلی بر روی مناقصات صنعت ساخت متمرکز بوده و جامعه هدف آن‌ها پیمانکاران بوده است. جنبه نوآورانه این پژوهش این است که به مناقصات بین‌المللی برگزار شده



شکل ۲- مراحل متدولوژی CRISP-DM (Chapman et al., 1999)

پژوهش مدارک ۱۱۸۴ پیشنهاد مناقصه یکی از شرکت‌های مشاوره فعال در صنعت آب در ایران مورد بررسی قرار گرفتند و اطلاعات این پیشنهادها در یک پایگاه داده با هشت ستون و ۱۱۷۹ ردیف جمع‌آوری شدند.

در مرحله سوم (آماده‌سازی داده‌ها) عملیات پیش‌پردازش بر روی داده‌ها انجام شده است که شامل مدیریت داده‌های تکراری<sup>۱</sup>، مقادیر گم‌شده<sup>۲</sup> و داده‌های پرت<sup>۱</sup> می‌باشد. بخش عمده‌ای از زمان

در مرحله اول (درک کسب‌وکار) باید شناخت کاملی از نیازمندی‌های کارفرما ایجاد شود. در این پژوهش می‌توان شرکت‌های مشاوره فعال در صنعت آب که به دنبال شرکت در مناقصات بین‌المللی هستند را به‌عنوان کارفرما و مخاطب پژوهش در نظر گرفت. در این مرحله معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزار شده برای انتخاب مشاوران صنعت آب با استفاده از روش دلفی شناسایی شدند.

مرحله دوم (درک داده‌ها) با جمع‌آوری داده‌ها شروع می‌شود و با آمار توصیفی ادامه پیدا می‌کند تا کیفیت داده‌ها ارزیابی شود. در این

1- Duplicate Values  
2- Missing Values

نزدیک‌ترین همسایگان خود شناسایی و سپس حذف می‌کند. مرحله چهارم (مدل‌سازی) شامل انتخاب مدل و الگوریتم مناسب و توسعه مدل می‌باشد. در این پژوهش از الگوریتم‌های درخت تصمیم<sup>۲</sup>، ID3، ماشین بردار پشتیبانی<sup>۳</sup>، Chaid<sup>۴</sup>، بیز ساده<sup>۴</sup> و نزدیک‌ترین همسایه<sup>۵</sup> برای مدل‌سازی استفاده شده است. شرح مختصری از هر الگوریتم و نحوه عملکرد آن‌ها در ذیل آورده شده است:

درخت تصمیم نقشه‌ای از نتایج احتمالی یکسری از انتخاب‌ها یا گزینه‌های مرتبط به هم است به طوری که به یک فرد یا سازمان اجازه می‌دهد تا اقدامات محتمل را از لحاظ هزینه‌ها، احتمالات و مزایا بسنجد. یک درخت تصمیم‌گیری به طور معمول با یک گره اولیه شروع می‌شود که پس از آن پیامدهای احتمالی به صورت شاخه‌هایی از آن منشعب شده و هر کدام از آن پیامدها به گره دیگری منجر شده که آن‌ها هم به نوبه‌ی خود شاخه‌هایی از احتمالات دیگر را ایجاد می‌کنند که این ساختار شاخه‌شاخه سرانجام به نموداری شبیه به یک درخت مبدل می‌شود. از درخت تصمیم می‌توان به منظور ایجاد مدل‌های پیش‌گویانه استفاده کرد که در حوزه‌ی یادگیری ماشینی، استخراج داده و آمار کاربردی هستند. این نوع درخت‌ها همچنین تحت عنوان Classification Tree نیز شناخته می‌شوند به طوری که هر شاخه دربرگیرنده مجموعه‌ای از ویژگی‌ها یا قوانین طبقه‌بندی داده است و مرتبط با یک دسته‌ی خاص می‌باشد که در انتهای هر شاخه یافت می‌شود. این دست قوانین که تحت عنوان Decision Rules شناخته می‌شوند قابل بیان به صورت جملات شرطی می‌باشند (مثلاً اگر شرایط ۱ و ۲ و ۳ محقق شوند، با قطعیت می‌توان گفت که X نتیجه‌ای همچون Y بر خواهد گرداند) هر مقدار داده‌ی اضافی به مدل کمک می‌کند تا دقیق‌تر پیش‌بینی کند که مسئله‌ی موردنظر به کدام مجموعه از مقادیر متعلق می‌باشد و این در حالی است که از این اطلاعات بعداً می‌توان به عنوان ورودی در یک مدل تصمیم‌گیری بزرگ‌تر استفاده کرد.

الگوریتم ID3 و الگوریتم Chaid از الگوریتم‌های زیرمجموعه‌ی درخت تصمیم می‌باشند و تفاوت آن‌ها در شیوه‌ی توسعه‌ی درخت می‌باشد. ماشین بردار پشتیبانی یک الگوریتم یادگیری ماشینی به صورت نظارتی است که می‌تواند در مسائل دسته‌بندی و رگرسیون مورد استفاده قرار بگیرد. نحوه عملکرد ماشین بردار پشتیبانی در قسمت بحث و نتایج تشریح شده است. الگوریتم بیز ساده یکی از ساده‌ترین الگوریتم‌های دسته‌بندی می‌باشد. اساس و مبنای الگوریتم دسته‌بندی بیز، قضیه بیز است و در بیشتر مواقع، زمانی که تعداد

پژوهش بر روی این مرحله و آماده‌سازی پایگاه داده نهایی صرف شده است. داده‌های تکراری شامل ردیف‌هایی در پایگاه داده می‌باشند که مقادیر تمامی ستون‌های آن‌ها باهم برابر می‌باشند. معمولاً داده‌های تکراری در اثر بروز اشتباه هنگام واردکردن مقادیر در پایگاه داده، ایجاد شده و باعث کاهش یا افزایش کاذب دقت مدل نهایی می‌شوند، بنابراین در مرحله آماده‌سازی داده‌ها باید داده‌های تکراری شناسایی و حذف شوند. در نرم‌افزار Rapidminer از عملگر "Remove Duplicates" برای شناسایی و حذف داده‌های تکراری استفاده شده است. این عملگر ابتدا تمامی داده‌ها را دوبه‌دو مقایسه کرده و پس از شناسایی داده‌های تکراری، آن‌ها را حذف می‌کند. مقادیر گمشده هنگامی ایجاد می‌شوند که یک یا چند ردیف (مشاهده) دارای مقادیر ثبت‌نشده یا ناموجود در ستون‌ها باشد. در این حالت گفته می‌شود که آن ردیف دارای مقادیر گمشده است. دو رویکرد در مقابله با مقادیر گمشده وجود دارد؛ حذف و جایگزینی. در رویکرد حذف که از رویکرد دیگر ساده‌تر می‌باشد، ردیف‌هایی که مقادیر گمشده دارند، به صورت کامل از پایگاه داده حذف می‌شوند. این رویکرد بسیار سریع و ساده می‌باشد اما کیفیت داده‌های موجود در پایگاه داده را کاهش می‌دهد. در رویکرد جایگزینی، مقدار گمشده با میانگین، میانه، مُد، کمینه و یا بیشینه‌ی سایر مقادیر موجود، جایگزین می‌شود تا تعداد ردیف‌ها کاهش نیابد. مزیت این رویکرد این است که کیفیت داده‌های موجود را کاهش نمی‌دهد؛ بنابراین در این پژوهش تصمیم گرفته شد که از رویکرد جایگزینی استفاده شود. در نرم‌افزار Rapidminer از عملگر "Replace Missing Values" برای شناسایی مقادیر گمشده و جایگزینی آن‌ها استفاده شده است. این عملگر ابتدا مقادیر گمشده را شناسایی کرده و مقادیر گمشده با مُد سایر مقادیر موجود جایگزین می‌شود. دلیل استفاده از مُد به جای میانگین این است که مُد نسبت به مقادیر پرت، حساس نیست. داده‌ی پرت به داده‌ای اطلاق می‌شود که با دیگر داده‌های هم‌گروه فاصله چشمگیری داشته باشد، به عبارت دیگر داده‌ی پرت به داده‌ای گفته می‌شود که نسبت به یک نقطه مرکزی (مثل میانگین) فاصله زیادی برحسب یک شاخص پراکندگی (مثال انحراف معیار) داشته باشد. یکی از دلایل ایجاد داده‌ی پرت، بروز اشتباه هنگام واردکردن مقادیر در پایگاه داده می‌باشد. داده‌های پرت نیز همانند داده‌های تکراری می‌توانند باعث کاهش یا افزایش کاذب دقت مدل نهایی شوند و در مرحله آماده‌سازی داده‌ها، باید شناسایی و حذف شوند. در نرم‌افزار Rapidminer از عملگر "Detect Outlier (Distances)" برای شناسایی و حذف داده‌های پرت استفاده شده است. این عملگر، داده‌های پرت را از طریق اندازه‌گیری فاصله‌ی یک داده از

2- Decision Tree

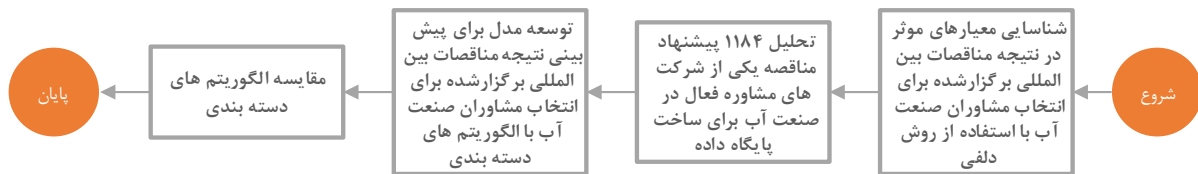
3- Support Vector Machine (SVM)

4- Naïve Bayes

5- K-Nearest Neighbor (KNN)

1- Outliers

نزدیک دسته‌بندی می‌کند. نمونه‌های یادگیری به‌عنوان بردارهایی در فضای ویژگی چندبعدی مطرح می‌شوند. فضا به ناحیه‌هایی با نمونه‌های یادگیری پارتیشن‌بندی می‌شود. یک نقطه در فضا به کلاسی تعلق می‌یابد که بیشترین نقاط یادگیری متعلق به آن کلاس در داخل نزدیک‌ترین نمونه‌ی یادگیری به  $k$  در آن باشد. در مرحله پنجم (ارزیابی) باید شایستگی مدل توسعه داده‌شده در مرحله چهارم مورد ارزیابی قرار بگیرد. روند کلی پژوهش در شکل ۳ نمایش داده‌شده است.



شکل ۳- روند کلی پژوهش

توافق ۸۰ درصد و بالاتر و تغییر امتیاز بیش از ۲۰ درصد برای دور بعدی دلفی باقی نماند، معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزار شده برای مشاوران صنعت آب با دو دور دلفی، نهایی شدند. به‌عبارت‌دیگر در طی دو دور دلفی، اتفاق نظر بر روی معیارهای مؤثر ایجاد شد. این معیارهای شناسایی شده، در پایگاه داده به‌عنوان صفت خاصه<sup>۱</sup> (ستون) در نظر گرفته شده‌اند. این معیارها در جدول ۱ نمایش داده شده‌اند.

تأمین‌کننده‌های مالی:

ADB = بانک توسعه آسیا، AFD = آژانس توسعه فرانسه، AFDB = بانک توسعه آفریقا، AIIB = بانک سرمایه‌گذاری زیرساختی آسیا، EBRD = بانک توسعه و بازسازی اروپا، EIB = بانک سرمایه‌گذاری اروپا، EU = اتحادیه اروپا، IBRD = بانک بین‌المللی بازسازی و توسعه، IDA = سازمان توسعه بین‌المللی، IFAD = صندوق توسعه کشاورزی، ISDB = بانک توسعه اسلامی، JICA = آژانس همکاری بین‌المللی ژاپن، KFW = بانک توسعه آلمان، UNDP = برنامه‌های توسعه سازمان ملل، UNIDO = سازمان توسعه صنعتی سازمان ملل، WB = بانک جهانی.

نوع پروژه‌ها: (با توجه به اینکه این دسته‌بندی بر اساس عرف بین‌المللی انجام شده، از ترجمه آن‌ها صرف‌نظر شده است)

IWSES= Integrated water, soil and environmental studies, WQMLS= Water Quality Monitoring and Limnology Studies, WUAPP= Water User Association and Public Participation, IWRM= Integrated Water Resources management, GMM= Groundwater Modeling and Management, IFIES= Institutional,

متغیرها کم ولی مشاهدات زیاد هستند الگوریتم بیز ساده برای تشخیص دسته‌ها مناسب است. الگوریتم نزدیک‌ترین همسایه در مسائل دسته‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. در این الگوریتم یک نمونه بارای اکثریت از همسایه‌های دسته‌بندی می‌شود و این نمونه در عمومی‌ترین کلاس مابین  $k$  همسایه نزدیک تعیین می‌شود.  $K$  یک مقدار مثبت صحیح و عموماً کوچک است. اگر  $k=1$  باشد نمونه به‌سادگی در کلاس همسایگان نزدیکش تعیین می‌گردد. فرد بودن مقدار  $k$  مفید می‌باشد چون با این کار جلوی آراء برابر گرفته می‌شود. این الگوریتم یک نمونه اعتبارسنجی را بر اساس  $k$  همسایه

## نتایج و بحث

با همکاری دفتر بازاریابی بین‌المللی یکی از شرکت‌های مشاوره فعال در صنعت آب در ایران، داده‌های این پژوهش در یک پایگاه داده با هشت ستون و ۱۱۷۹ ردیف جمع‌آوری شد. به‌عبارت‌دیگر، ۱۱۸۴ پیشنهاد مناقصه بین‌المللی مورد بررسی قرار گرفتند و پایگاه داده توسعه داده شد. روش دلفی به‌عنوان یک روش ساختاریافته برای ایجاد اتفاق نظر میان خبرگان، به‌منظور شناسایی معیارهای مؤثر در نتیجه مناقصات بین‌المللی صنعت آب استفاده شده است. در این پژوهش اعضای پانل دلفی ۱۵ نفر از خبرگان فعال در دفتر بازاریابی بین‌المللی بوده و از آن‌ها خواسته شد تا اهمیت معیارهای مؤثر را تعیین کنند. در دور اول از اعضای پانل دلفی خواسته شد تا بر اساس مقیاس هفت‌قسمتی لیکرت، به معیارهای مؤثر امتیاز دهند و نظرات اصلاحی و پیشنهادهای خود را نیز در سؤال باز ارائه دهند.

تحلیل آماری پاسخ‌ها با استفاده از محاسبه شاخص‌های مرکزی و شاخص‌های پراکندگی در نرم‌افزار SPSS 25 انجام شد و محدوده میان چارکی، میانگین و انحراف معیار امتیاز معیارهای مؤثر محاسبه گردید.

معیارهای با صدک ۸۰ و بالاتر مورد پذیرش قرار گرفتند، معیارهای با صدک کمتر از ۲۰ کنار گذاشته شدند و معیارهای با صدک بین ۲۰ تا ۸۰ برای دور دوم دلفی باقی ماندند (Landeta, 2006). در دور دوم دلفی نیز از اعضای پانل خواسته شد تا همانند دور اول میزان موافقت خود را با وجود هر یک از معیارها اعلام نمایند. در این دور، معیارهایی با سطح توافق ۸۰ درصد و بالاتر و تغییر امتیاز کمتر از ۲۰ درصد مورد پذیرش قرار گرفتند و چون معیاری با سطح

Aquaculture, and Animal Husbandry, SSDS= Soil survey and Drainage Studies, AIAE= Agricultural investigation and agro product economics, GI= Geotechnical Investigation, WSTL= Water and Sewage and transmission lines, WS= Water Supply, DN= Distribution Network, SN= Sewerage Network, WTP= Water Treatment Plant, RUSWN= Rehabilitation and Upgrading of Sewage and Water Network and Facilities, SSMP= Sewerage and Stormwater Master Plan, RET= Reservoir / Elevated Tank, PS= Pumping Station.

Financing, Investment, and Economic Studies, CDM= Clean Development Mechanism, HPP= Hydroelectric Power Plant, CD= Concrete Dam, ED= Embankment Dam, UCT= Underground Cavern and Tunnels, DS= Dam Safety, RHR= Road, Highway and Railway, GSB= Geology, Seismicity, Borrow area, TD= Tender documents, PS= Power supply, ID= Irrigation and Drainage, MSID= Main and Secondary Irrigation and Drainage, FCDB= Flood Control and Dam Break, DDI= Diversion Dams and Intakes, IIRID= Integration, Rehabilitation and Renovation of Irrigation and Drainage, WM= Watershed Management, AAH=

جدول ۱- معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزارشده برای مشاوران صنعت آب

شماره	معیار (صفت خاصه)	نوع معیار	مقادیر
1	نوع امور تخصصی	Polynomial	۱ = مطالعات یکپارچه منابع آب، خاک و محیط‌زیست، ۲ = سد و نیروگاه و تأسیسات زیربنایی، ۳ = آبیاری، زهکشی و کشاورزی، ۴ = آب و فاضلاب و خطوط انتقال سیالات، ۵ = ژئوماتیک
2	نحوه ارسال پیشنهاد مناقصه	Polynomial	۱ = وب‌سایت، ۲ = ایمیل، ۳ = تحویل فیزیکی
3	چگونگی آشنایی با مناقصه	Polynomial	۱ = ارتباط قبلی با کارفرما، ۲ = اینترنت، ۳ = سایر، ۴ = شریک خارجی، ۵ = کنسرسیوم
4	نوع مناقصه	Polynomial	۱ = مناقصه عمومی (EOI) و ۲ = مناقصه محدود (RFP)
5	تأمین‌کننده مالی	Polynomial	1= ADB, 2= AFD, 3= AFDB, 4= AIIB, 5= Arab Fund, 6= EBRD, 7= EIB, 8= EU, 9= IBRD, 10= IDA, 11= IFAD, 12= ISDB, 13= JICA, 14= KFW, 15= UNDP, 16= UNICEF, 17= UNIDO, 18= WB, 19= employer, 20= Other
6	داشتن همکار در کشور مبدأ	Binomial	۱ = بله، ۲ = خیر
7	نتیجه نهایی مناقصه	Binomial	۱ = برنده، ۲ = بازنده
8	نوع پروژه	Polynomial	200= IWSES, 205= WQMLS, 206= WUAPP, 207= IWRM, 208= GMM, 210= IFIES, 211= CDM, 300= DAM, 301= HPP, 302= CD, 303= ED, 304= UCT, 306= DS, 307= RHR, 310= GSB, 312= TD, 313= PS, 400= ID, 401= MSID, 404= FCDB, 405= DDI, 407= IIRID, 408= WM, 410= AAH, 412= SSDS, 413= AIAE, 414= GI, 500= WSTL, 501= WS, 502= DN, 503= SN, 505= WTP, 507= RUSWN, 508= SSMP, 509= RET, 510= PS

Rapid Miner بر اساس چهار مشخصه انجام می‌شود. به عبارت دیگر برای تعیین معیارهای مؤثر در پیش‌بینی نتیجه مناقصات بین‌المللی، مشخصه‌های زیر در مورد هریک از معیارها بررسی می‌شوند:

۱- Correlation: بیانگر همبستگی معیارها با معیار هدف (نتیجه نهایی مناقصه) می‌باشد

۲- ID-ness: بیانگر میزان تمایز در مقادیر هریک از معیارها می‌باشد

۳- Stability: بیانگر میزان تشابه در مقادیر هریک از معیارها می‌باشد

۴- Missing: بیانگر میزان مقادیر گم‌شده در هریک از معیارها می‌باشد

در این پژوهش پس از انجام عملیات انتخاب معیار، این نتیجه حاصل شد که تمامی هفت معیار شناسایی شده (جدول ۱) در پیش‌بینی نتیجه مناقصات مؤثر هستند.

در مرحله آماده‌سازی داده‌ها، مقادیر تکراری در نرم‌افزار Rapid Miner با عملگر Remove Duplicates حذف شدند و مقادیر

پایگاه داده با ۱۱۷۹ ردیف و هشت ستون ایجاد گردید و "نتیجه نهایی مناقصه" به عنوان متغیر هدف در نظر گرفته شد. با توجه به اینکه "نتیجه نهایی مناقصه" فقط دارای دو مقدار "بازنده" و "برنده" و از نوع دودویی<sup>۱</sup> می‌باشد، مسئله پیش‌بینی نتیجه مناقصه از نوع دسته‌بندی است و در این پژوهش نیز از الگوریتم‌های دسته‌بندی<sup>۲</sup> در نرم‌افزار Rapid Miner برای مدل‌سازی استفاده شده است.

در مرحله درک داده‌ها از تحلیل‌های آمار توصیفی استفاده شد و مشخص گردید که حدوداً ۵۴ درصد از پیشنهادها مناقصه برنده شده و ۴۶ درصد از پیشنهادها بازنده شده‌اند. برابری تقریبی تعداد پیشنهادها برنده شده و تعداد پیشنهادها بازنده نشان‌دهنده توزیع مناسب داده‌ها در پایگاه داده می‌باشد. در گام بعدی باید معیارهایی که می‌توانند در پیش‌بینی نتیجه مناقصات مؤثر باشند شناسایی شوند که این عملیات، انتخاب معیار<sup>۳</sup> نامیده می‌شود. انتخاب معیار در نرم‌افزار

- 1- Binomial
- 2- Classification
- 3- Attribute Selection

دادن به سوالات می‌سجند. منابع مختلفی به تعیین نسبت داده‌های یادگیری و داده‌های اعتبارسنجی پرداخته‌اند. جردن و میچل در کتاب "یادگیری ماشین، چشم‌انداز و روندها"، نسبت ۳۰-۷۰ (۷۰ درصد برای داده‌های یادگیری و ۳۰ درصد برای داده‌های اعتبارسنجی) را پیشنهاد کرده‌اند (Jordan and Mitchell, 2015). در حال حاضر این نسبت، بیشترین کاربرد را در یادگیری ماشین دارد.

ماتریس آشفتگی یک جدول است که از آن برای تشریح عملکرد مدل‌های پیش‌گویانه ساخته شده با الگوریتم‌های دسته‌بندی استفاده می‌شود. ماتریس آشفتگی، نتایج حاصل از طبقه‌بندی را بر اساس اطلاعات واقعی موجود، نمایش می‌دهد. حال بر اساس این مقادیر می‌توان معیارهای مختلف ارزیابی دسته‌بندی و اندازه‌گیری دقت را تعریف کرد. ماتریس آشفتگی هر یک از مدل‌های توسعه داده شده، در شکل‌های زیر نمایش داده شده است.

ماتریس آشفتگی مدل توسعه داده شده با الگوریتم درخت تصمیم، در جدول ۲ نمایش داده شده است.

با توجه به جدول ۲، مدل توسعه داده شده با درخت تصمیم، از میان ۵۳۱ پیشنهاد بازنده شده در واقعیت، ۴۵۱ پیشنهاد را به‌عنوان بازنده پیش‌بینی کرده است و از میان ۶۴۸ پیشنهاد برنده شده در واقعیت، ۵۵۸ پیشنهاد را به‌عنوان برنده پیش‌بینی کرده است. در این پژوهش از دقت صحیح<sup>۴</sup>، دقت تکرارپذیری<sup>۵</sup> و نرخ یادآوری<sup>۶</sup> برای ارزیابی عملکرد مدل‌ها و الگوریتم‌ها استفاده شده است. دقت صحیح عبارت است از نسبت تعداد مقادیری که به‌طور صحیح پیش‌بینی شده‌اند به کل تعداد مقادیر. به عبارت دیگر معیار دقت صحیح، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک الگوریتم دسته‌بندی است و عبارت است از میزان تشخیص صحیح الگوریتم در مجموع دودسته. این پارامتر در واقع نشان‌گر میزان الگوهای است که درست تشخیص داده شده‌اند. دقت تکرارپذیری عبارت است از نسبت تعداد مقادیر مثبتی (برنده) که به‌طور صحیح پیش‌بینی شده‌اند به کل تعداد مقادیر مثبت (برنده). مقدار بالا برای معیار دقت تکرارپذیری، بیانگر تعداد کم داده‌هایی است که به اشتباه، در کلاس خاص دسته‌بندی شده‌اند. شایان توجه است که معیار دقت تکرارپذیری، فقط برای مواردی ارزیابی می‌شود که در آن‌ها، مدل دسته‌بندی تعلق یک نمونه به یک کلاس خاص را پیش‌بینی کرده باشد. نرخ یادآوری عبارت است از نسبت تعداد مقادیر مثبتی که به‌طور صحیح پیش‌بینی شده‌اند به کل تعداد مقادیر. مقدار بالا برای معیار نرخ یادآوری، بیانگر تعداد کم داده‌هایی است که به اشتباه، در آن کلاس خاص دسته‌بندی نشده‌اند. استفاده از این معیار، به‌تنهایی، برای ارزیابی عملکرد سیستم درست نیست و باید در کنار معیار دقت

گمشده با استفاده از عملگر Replace Missing جایگذاری شدند، این عملگر با توجه با سایر داده‌ها، مقادیر گمشده را جایگذاری می‌کند. خروجی این مرحله پایگاه داده نهایی می‌باشد و این پایگاه داده در مدل‌سازی مورد استفاده قرار می‌گیرد.

با توجه به ماهیت مسئله، الگوریتم‌های دسته‌بندی زیر برای مدل‌سازی استفاده شده‌اند:

- درخت تصمیم
- ID3
- Chaid
- بیز ساده
- ماشین بردار پشتیبانی
- نزدیک‌ترین همسایه

در این پژوهش، مدل‌سازی بر اساس مراحل زیر انجام شده است:

- ۱- بارگذاری پایگاه داده در نرم‌افزار Rapid Miner
  - ۲- تعیین تناسب داده‌های یادگیری<sup>۱</sup> و داده‌های اعتبارسنجی<sup>۲</sup>. در این مسئله از ۷۰ درصد داده‌ها برای یادگیری و توسعه مدل و از ۳۰ درصد داده‌ها برای اعتبارسنجی مدل استفاده شده است.
  - ۳- شروع عملیات یادگیری و توسعه مدل با هر یک از شش الگوریتم مذکور
  - ۴- ارزیابی و اعتبارسنجی مدل‌های تولید شده با هر یک از الگوریتم‌ها با استفاده از ماتریس آشفتگی<sup>۳</sup>.
- برای درک بهتر داده‌های یادگیری و داده‌های اعتبارسنجی، مثال زیر ارائه شده است:

فرض کنید یک معلم ۱۰۰ نمونه سؤال همراه با جواب در اختیار دانشجو قرار داده است. دانشجو باید با خواندن این ۱۰۰ سؤال خود را برای امتحان آماده کند. به این ۱۰۰ سؤال به‌نوعی داده‌های یادگیری گفته می‌شود زیرا دانشجو از این داده‌ها برای آموزش خود و آمادگی برای امتحان اصلی، استفاده می‌کند. البته از آنجاکه فرض بر این است که دانشجو فقط همین ۱۰۰ سؤال را دارد و هیچ منبع دیگری در اختیار ندارد، نمی‌تواند خود را قبل از امتحان ارزیابی کند. پس منطقی است که به‌صورت تصادفی، از میان این ۱۰۰ سؤال، مثلاً ۷۰ سؤال را جدا کرده، آن‌ها را بخواند و یاد بگیرد. ۳۰ سؤال باقی‌مانده، داده‌های اعتبارسنجی برای ارزیابی هستند که دانشجو باید توسط آن‌ها، خود را قبل از آزمون واقعی بیازماید. در واقع با خواندن ۷۰ سؤال و دیدن جواب‌های آن‌ها، یادگیری انجام می‌شود و سپس ۳۰ سؤال باقی‌مانده را برای ارزیابی خود استفاده می‌کند. ۳۰ سؤال را خوانده و برای خود جواب می‌دهد، سپس جواب‌های داده شده را با جواب‌هایی واقعی همان ۳۰ سؤال، مقایسه می‌کند و دقت و صحت خود را در پاسخ

4 Accuracy  
5 Precision  
6 Recall

1- Train  
2- Test  
3- Confusion Matrix



مقدار پیش‌بینی شده نیز همین را نشان می‌دهد. مثبت‌های کاذب (False Positive): مقدار واقعی "بازنده" بوده است اما مقدار پیش‌بینی شده "برنده" است. به‌طور مثال مقدار واقعی نشان می‌دهد که مشاور در مناقصه بازنده شده است اما مقدار پیش‌بینی شده نشان می‌دهد که مشاور در مناقصه برنده شده است. منفی‌های کاذب (False Negative): مقدار واقعی "برنده" بوده است اما مقدار پیش‌بینی شده "بازنده" است. به‌طور مثال مقدار واقعی نشان می‌دهد که مشاور در مناقصه برنده شده است اما مقدار پیش‌بینی شده نشان می‌دهد که مشاور در مناقصه بازنده شده است. ماتریس آشفتگی مدل توسعه داده شده با الگوریتم ID3، در جدول ۳ نمایش داده شده است. ماتریس آشفتگی مدل توسعه داده شده با الگوریتم Chaid، در جدول ۴ نمایش داده شده است. ماتریس آشفتگی مدل توسعه داده شده با الگوریتم نزدیک‌ترین همسایه، در جدول ۵ نمایش داده شده است.

تکرارپذیر مورد استفاده قرار بگیرد؛ زیرا به راحتی می‌شود مدل‌های دسته‌بندی متنی طراحی کرد که نرخ یادآوری بالایی داشته باشند و این لزوماً به معنای دقت تکرارپذیر بالا نیست.

- دقت صحیح:  $TP+TN/TP+FP+FN+TN$
- دقت تکرارپذیر:  $TP/TP+FP$
- نرخ یادآوری:  $TP/TP+FN$

مثبت‌های واقعی (True Positive): مقادیر مثبتی هستند که به‌طور صحیح پیش‌بینی شده‌اند. به این معنی که مقدار واقعی "برنده" بوده است و مقدار پیش‌بینی شده نیز "برنده" است. به‌طور مثال مقدار واقعی نشان می‌دهد که مشاور در مناقصه برنده شده است و مقدار پیش‌بینی شده نیز همین را نشان می‌دهد.

منفی‌های واقعی (True Negative): مقادیر منفی هستند که به‌طور صحیح پیش‌بینی شده‌اند. به این معنی که مقدار واقعی "بازنده" بوده است و مقدار پیش‌بینی شده نیز "بازنده" است. به‌طور مثال مقدار واقعی نشان می‌دهد که مشاور در مناقصه بازنده شده است و

جدول ۲- ماتریس آشفتگی مدل توسعه داده شده با درخت تصمیم

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
حقیقی	برنده	558	90	86.11%
	بازنده	80	451	84.93%
Class Recall		87.46%	83.36%	Accuracy=85.58%

جدول ۳- ماتریس آشفتگی مدل توسعه داده شده با الگوریتم ID3

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
حقیقی	برنده	558	88	86.38%
	بازنده	80	453	84.99%
Class Recall		87.46%	83.73%	Accuracy=85.75%

جدول ۴- ماتریس آشفتگی مدل توسعه داده شده با الگوریتم Chaid

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
حقیقی	برنده	566	92	86.02%
	بازنده	72	449	86.18%
Class Recall		88.71%	82.99%	Accuracy=86.08%

جدول ۵- ماتریس آشفتگی مدل توسعه داده شده با الگوریتم نزدیک‌ترین همسایه

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
حقیقی	برنده	489	129	79.13%
	بازنده	149	412	73.44%
Class Recall		76.65%	76.16%	Accuracy=76.42%

جدول ۶- ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم بیز ساده

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
خطای	برنده	547	86	86.41%
	بازنده	91	455	83.33%
	Class Recall	85.74%	84.10%	Accuracy=84.98%

ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم بیز ساده، در جدول ۶ نمایش داده‌شده است. ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی، در جدول ۷ نمایش داده‌شده است. دقت صحیح مدل‌های توسعه داده‌شده با هر یک از الگوریتم‌ها در جدول ۱۰ نمایش داده‌شده است.

جدول ۷- ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
خطای	برنده	599	87	87.32%
	بازنده	39	454	92.09%
	Class Recall	93.89%	83.92%	Accuracy=89.31%

جدول ۸- ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی تکاملی

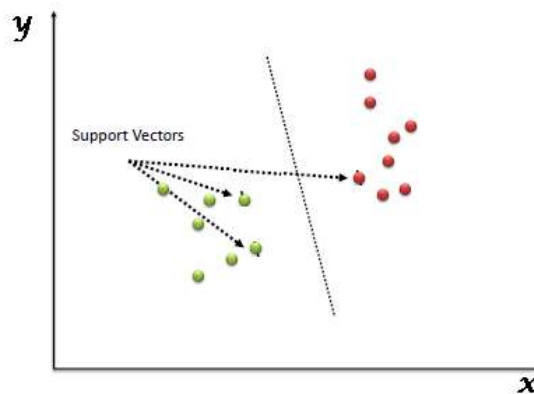
پیش‌بینی شده				
	برنده	بازنده	Class Precision	
خطای	برنده	536	94	85.08%
	بازنده	102	447	81.42%
	Class Recall	84.01%	82.62%	Accuracy=83.38%

جدول ۹- ماتریس آشفتگی مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی PSO

پیش‌بینی شده				
	برنده	بازنده	Class Precision	
خطای	برنده	427	61	87.50%
	بازنده	211	480	69.46%
	Class Recall	66.93%	88.72%	Accuracy=76.92%

جدول ۲- دقت صحیح و خطای دسته‌بندی مدل‌های توسعه داده‌شده با هر یک از الگوریتم‌ها

الگوریتم	دقت صحیح	خطای دسته‌بندی
درخت تصمیم	85.58%	14.42%
ID3	85.75%	14.25%
Chaid	86.08%	13.92%
نزدیک‌ترین همسایه	76.42%	23.58%
بیز ساده	84.98%	15.02%
ماشین بردار پشتیبانی	89.31%	10.69%
ماشین بردار پشتیبانی تکاملی	83.38%	16.62%
ماشین بردار پشتیبانی PSO	76.92%	23.08%



شکل ۴- ترسیم داده‌ها در فضای n بعدی در الگوریتم ماشین بردار پشتیبانی

مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی دارای مشخصات زیر می‌باشد (جدول ۱۱).

جدول ۱۱- مشخصات مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی

Parameters	Values
SVM Type	C-SVC
Kernel Type	rbf
Gamma	0.099
C	10
Cache Size	80
Epsilon	0.001

بین‌المللی هستند.

### نتیجه‌گیری

شرکت/عدم شرکت در مناقصات بین‌المللی صنعت آب، یکی از تصمیم‌های مهم شرکت مشاوره است که تحت تأثیر عوامل مختلفی می‌باشد. این تصمیم نقش مهمی را در موفقیت یا شکست سازمان ایجاد می‌کند. شرکت‌های مشاوره فعال در صنعت آب، مخاطب و کاربر اصلی این پژوهش هستند و هدف این مطالعه عبارت است از شناسایی معیارهای مؤثر در نتیجه مناقصات بین‌المللی برگزارشده برای انتخاب مشاوران صنعت آب از طریق روش دلفی و همچنین مقایسه الگوریتم‌های داده‌کاوی در پیش‌بینی نتیجه این مناقصات.

معیارهای مؤثر در نتیجه این مناقصات عبارت‌اند از نوع امور تخصصی، نحوه ارسال پیشنهاد مناقصه، چگونگی آشنایی با مناقصه، نوع مناقصه، تأمین‌کننده مالی، داشتن همکار در کشور مبدأ و نوع پروژه. الگوریتم‌های داده‌کاوی مقایسه شده عبارت‌اند از درخت تصمیم، Chaid، ID3، نزدیک‌ترین همسایه، بیز ساده و ماشین بردار پشتیبانی. در این پژوهش ۱۱۸۴ پیشنهاد مناقصه بین‌المللی به‌عنوان داده مورد استفاده قرار گرفته است.

الگوریتم‌هایی که بیشترین دقت را در پیش‌بینی نتیجه مناقصات

همان‌طور که در جدول ۱۰ نمایش داده‌شده است، مدل توسعه داده‌شده با الگوریتم ماشین بردار پشتیبانی دارای بیشترین دقت صحیح در پیش‌بینی نتیجه مناقصات بین‌المللی برگزارشده برای انتخاب مشاوران صنعت آب را دارد. ماشین بردار پشتیبانی یک الگوریتم یادگیری ماشین به‌صورت نظارتی<sup>۱</sup> است که می‌تواند در مسائل دسته‌بندی و رگرسیون مورد استفاده قرار بگیرد، با این وجود کاربرد عمده آن در مسائل دسته‌بندی می‌باشد. در الگوریتم ماشین بردار پشتیبانی، هر نمونه داده به‌عنوان یک نقطه در فضای n بعدی روی نمودار پراکنده‌گی داده‌ها ترسیم می‌شود (n عبارت است از تعداد معیارها) و مقدار هر معیار مربوط به داده‌ها، یکی از مؤلفه‌های مختصات نقطه روی نمودار را مشخص می‌کند. سپس با ترسیم یک خط راست، داده‌های مختلف و متمایز از یکدیگر را دسته‌بندی می‌کند (مطابق شکل ۴).

از آنجایی که الگوریتم ماشین بردار پشتیبانی نسبت به سایر الگوریتم‌های دسته‌بندی عملکرد بهتری در پیش‌بینی نتیجه مناقصات بین‌المللی داشت، پیشنهاد می‌شود که این الگوریتم به‌عنوان پردازشگر در سیستم‌های پشتیبان تصمیم مناقصات استفاده شود تا موجب بهبود تصمیم مشاورانی شود که در صنعت آب به دنبال شرکت در مناقصات

of Science, 349(6245): 255–260.

Juszczyk, M. et al. 2014. 'Errors in the preparation of design documentation in public procurement in Poland', *Procedia Engineering*. Elsevier, 85: 283–292.

Kalan, D. and Ozbek, M.E. 2020. 'Development of a Construction Project Bidding Decision-Making Tool', *Practice Periodical on Structural Design and Construction*. American Society of Civil Engineers, 25(1): 4019032.

Landeta, J. 2006. 'Current validity of the Delphi method in social sciences', *Technological forecasting and social change*. Elsevier, 73(5): 467–482.

Leśniak, A. and Plebankiewicz, E. 2013. 'Modeling the decision-making process concerning participation in construction bidding', *Journal of Management in Engineering*. American Society of Civil Engineers, 31(2): 4014032.

Leśniak, A. and Radziejowska, A. 2017. 'Supporting bidding decision using multi-criteria analysis methods', *Procedia engineering*. Elsevier, 208: 76–81.

Lowe, D.J. and Parvar, J. 2004. 'A logistic regression approach to modelling the contractor's decision to bid', *Construction Management and Economics*. Taylor & Francis, 22(6): 643–653.

Minli, Z. and Shanshan, Q. 2012. 'Research on the application of artificial neural networks in tender offer for construction projects', *Physics Procedia*. Elsevier, 24: 1781–1788.

Oo, B., Drew, D.S. and Lo, H.P. 2007. 'Applying a random coefficients logistic model to contractors' decision to bid', *Construction Management and Economics*. Taylor & Francis, 25(4): 387–398.

Parchamijalal, M., Shakibazahed, H. and Ghaaderi, M.A. 2020. 'Comparison of questionnaire and text mining in identifying the criteria that affect the bid/no-bid decision of Consultant firms in international bidding for the water industry', *Iranian Journal of Irrigation & Drainage*, 14(2).

Polat, G., Bingol, B.N. and Uysalol, E. 2014. 'Modeling Bid/No bid decision using adaptive neuro fuzzy inference system (ANFIS): a case study', in *Construction Research Congress*, pp. 1083–1092.

Tan, Y., Shen, L. and Langston, C. 2010. 'Contractors' competition strategies in bidding: Hong Kong study', *Journal of construction engineering and management*. American Society of Civil Engineers, 136(10): 1069–1077.

Wanous, M., Boussabaine, A. and Lewis, J. 2000. 'To bid or not to bid: a parametric solution', *Construction Management & Economics*. Taylor & Francis, 18(4): 457–466.

بین‌المللی برگزار شده برای انتخاب مشاوران صنعت آب، دارند، به ترتیب عبارت‌اند از: ۱- ماشین بردار پشتیبانی، ۲- Chaid، ۳- ID3، ۴- درخت تصمیم، ۵- بیز ساده و ۶- نزدیک‌ترین همسایه؛ بنابراین پیشنهاد می‌شود که در پژوهش‌های بعدی از الگوریتم ماشین بردار پشتیبانی به‌عنوان پردازشگر در سیستم‌های پشتیبان تصمیم مناقصات استفاده شود. این سیستم از تصمیم مشاوران برای شرکت/عدم شرکت در مناقصات بین‌المللی صنعت آب پشتیبانی می‌کند.

## منابع

Al-Humaidi, H.M. 2016. 'Construction projects bid or not bid approach using the fuzzy technique for order preference by similarity FTOPSIS method', *Journal of Construction Engineering and Management*. American Society of Civil Engineers, 142(12): 4016068.

Biruk, S., Jaśkowski, P. and Czarnigowska, A. 2017. 'Modeling contractor's bidding decisions', *Procedia Engineering*. Elsevier, 182: 91–98.

Chapman, P. et al. 1999. 'The CRISP-DM user guide', in 4<sup>th</sup> CRISP-DM SIG Workshop in Brussels in March.

Cheng, M.-Y. et al. 2011. 'Bidding decision making for construction company using a multi-criteria prospect model', *Journal of Civil Engineering and Management*. Taylor & Francis, 17(3): 424–436.

Chisala, M.L. 2017. 'Quantitative bid or no-bid decision-support model for contractors', *Journal of Construction Engineering and Management*. American Society of Civil Engineers, 143(12): 4017088.

Chou, J.-S., Pham, A.-D. and Wang, H. 2013. 'Bidding strategy to support decision-making by integrating fuzzy AHP and regression-based simulation', *Automation in Construction*. Elsevier, 35: 517–527.

Costantino, N. et al. 2011. 'A model for predicting the bid distribution in public tenders', in *Proceedings International Public Procurement Conference IPPC4*, pp. 1–15.

El-Mashaleh, M.S. 2013. 'Empirical framework for making the bid/no-bid decision', *Journal of Management in Engineering*. American Society of Civil Engineers, 29(3): 200–205.

Jarkas, A.M., Mubarak, S.A. and Kadri, C.Y. 2014. 'Critical factors determining bid/no bid decisions of contractors in Qatar', *Journal of Management in Engineering*. American Society of Civil Engineers, 30(4): 5014007.

Jordan, M.I. and Mitchell, T.M. 2015. 'Machine learning: Trends, perspectives, and prospects', *Science*. American Association for the Advancement

Bidding Risk Prediction Model Based on Multiple Attribute Decision-Making Theories and Its Application', in CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems, pp. 3280-3291.

Wanous, M., Boussabaine, H.A. and Lewis, J. 2003. 'A neural network bid/no bid model: the case for contractors in Syria', Construction Management and Economics. Taylor & Francis, 21(7): 737-744.

Zhang, F., Wang, W. and Xu, S. 2014. 'Six Kinds of

## Comparison of Data Mining Algorithms in Predicting the Results of International Tenders Held to Select Water Industry Consultants (For Use in Decision Support Systems)

H. Shakiba Zahed<sup>1</sup>, M. Parchamijalal<sup>2\*</sup>, M.A. Ghaaderi<sup>3</sup>

Received: May.17, 2020

Accepted: Jul.21, 2020

### Abstract

In developing countries where the value of money is low, consultant firms are keen to participate in international tenders. Participating in international tenders requires a lot of resources (time, cost, etc.) to evaluate the project condition and prepare a suitable proposal. Predicting the outcome of these tenders is important because it can prevent the use of resources to participate in inappropriate tenders. The aim of this paper is to identify factors that affect the outcome of water industry international tenders holding for selection of consultants and compare the classification algorithms in predicting the outcome of this tenders. effective factors include Technical classification, Documents delivery method, Tender Identify Method, Type of Tender, Financer, Having partner in origin country, Final Status and Project Type and Compared algorithms include Decision Tree, ID3, Chaid, K-Nearest Neighbor (KNN), Naïve Bayes and Support Vector Machine (SVM). The most accurate algorithms are 1-SVM, 2-Chaid, 3-ID3, 4-Decision Tree, 5-Naïve Bayes and 6-KNN. So it is suggested to use the SVM algorithm as the processor in decision support systems to improve the bid/no-bid decision for consultant firms seeking to participate in the water industry international tenders.

**Keywords:** Classification algorithms, Data mining, Decision making, Water industry tenders

---

1- Ph.D. Candidate in Project and Construction Management, University of Tehran, Tehran, Iran

2- Associate Professor in Project and Construction Management, University of Tehran, Tehran, Iran

3- Masters Student in Project and Construction Management, University of Tehran, Tehran, Iran

(\* - Corresponding Aurther Email: parchamijalal@ut.ac.ir)