

ارائه مدل ترکیبی هوشمند تعیین روش مشارکت عمومی - خصوصی صنعت آب و فاضلاب ایران بر مبنای الگوریتم‌های جمعی درختی

چکیده:

دسترسی به آب سالم و دفع بهداشتی فاضلاب از ارکان توسعه هر کشور است و ضروری است که این طرح‌ها در کوتاه‌ترین زمان تکمیل گردد. با توجه به ظهور انواع روش‌های مشارکتی، انتخاب روش مناسب، یکی از مسائل مهم بوده و این صنعت نیازمند مدلی جهت تصمیم‌گیری در خصوص نحوه و روش سرمایه‌گذاری در منطقه و یا طرح خاص است. با توجه به وجود پایگاه داده از پروژه‌های مشارکتی در بخش آب و فاضلاب، میتوان با استفاده از اطلاعات به‌دست‌آمده و الگوریتم‌های کشف الگو و تصمیم‌گیری داده‌کاوی، مدل سرمایه‌گذاری و روش مناسب اجرای پروژه را پیش‌بینی نمود. این تحقیق با استفاده از داده‌های ۱۷۶ پروژه و بهره‌گیری از فرآیند کریسپ در داده‌کاوی انجام شده است. پس از تشریح و درک داده، مراحل پاکسازی و حذف داده پرت اجرا و در مرحله دسته‌بندی با تکنیک‌های درختی و یادگیری ماشین، طبقه‌بندی موفقیت و شکست پروژه‌ها و تحلیل‌های لازم انجام گرفته و شاخص‌های مشارکت عمومی - خصوصی به ترتیب اولویت استخراج گردید. بر مبنای یافته‌های مدل‌سازی، روش ترکیبی COF^۱ جهت حذف داده پرت و شاخص آنتروپی برای انتخاب ویژگی و روش استکینگ با دقت ۸۶,۲۷٪، مدل پیشنهادی جهت پیش‌بینی موفقیت پروژه‌ها هستند. با توجه به مدل پیشنهادی، میتوان علاوه بر معرفی قالب قراردادی مناسب اجرای هر گروه از پروژه‌های بخش آب و فاضلاب، میزان موفقیت هر طرح را در هر یک از قالب‌های قراردادی پیش‌بینی و تأثیر بهبود هر یک از شاخص‌ها را بررسی نمود.

واژه‌های کلیدی: داده‌کاوی، پیش‌بینی، حذف داده پرت، روش‌های ترکیبی، شاخص سود اطلاعاتی، گرادیان بوستینگ، جنگل تصادفی

مقدمه:

در دنیای امروز و با توجه به پیشرفت‌های روزافزون اطلاعات محور، لزوم استفاده از فناوری اطلاعات و ارتباطات در برنامه‌ریزی و تصمیم‌گیری بر کسی پوشیده نیست (Mutula and van Brakel, 2006). صنعت آب و فاضلاب نیز از این قاعده مستثنا نبوده و همواره سعی داشته تا با به‌کارگیری مدل‌ها و ابزارهای جدید، کارایی را در حوزه‌های کاری مختلف خود بهبود بخشد.

ازجمله مسائل موثر در ایجاد اخلاص در تکمیل طرح‌های صنعت آب و فاضلاب، کمبود منابع مالی می‌باشد. بر مبنای مستندات بانک جهانی^۲، سازمان بهداشت جهانی^۳ و یونسف^۴، به‌منظور بهره‌مندی مردم از سیستم دفع بهداشتی فاضلاب در کشورهای آسیایی در حدود ۳۰۰ میلیارد دلار اعتبار نیاز است (Hall et al., 2005) و با توجه به وابستگی این طرح‌ها به منابع دولتی در اکثر این کشورها، تحقق این امر با مشکلات فراوانی روبه‌رو است. درواقع، حجم بالای منابع مالی موردنیاز جهت نیل به اهداف برنامه‌ریزی‌شده توسعه‌ای و محدودیت منابع مالی دولتی در سالیان اخیر، رویکرد استفاده از روش‌های نوین تأمین منابع مالی و سرمایه‌گذاری بخش غیردولتی جهت تأمین منابع مالی موردنیاز طرح‌های بخش آب و فاضلاب را نه تنها یک راه حل، بلکه تبدیل به یک ضرورت نموده تا بتوان از توان اقتصادی، مدیریتی و اجرایی بخش خصوصی کشور استفاده نمود.

در کشور ایران، شرکت‌های آب و فاضلاب بر مبنای رسالت اصلی خود، مسئول تأمین آب سالم و دفع بهداشتی فاضلاب می‌باشند که متأسفانه در این مسیر و به‌منظور تأمین منابع مالی، همواره با مشکلاتی مواجه بوده‌اند. از یک سو هزینه بری بالای طرح‌های آب و فاضلاب که بر مبنای مطالعات انجام شده در این صنعت به میزان ۲۰۶ هزار میلیارد ریال در بخش شهری و ۲۵ هزار میلیارد ریال در بخش روستایی می‌باشد (Company, 2009) و از سوی دیگر محدودیت‌های منابع مالی، اجرایی و تخصصی، سبب گردیده که اجرای پروژه‌ها توسط بخش دولتی معمولاً با زمان بری بالایی تحقق یافته و هزینه‌های اضافی بیشتری نیز به پروژه تحمیل شود. بر این اساس، ورود بخش خصوصی به طرح‌های صنعت آب و فاضلاب در قالب سرمایه‌گذار، مجری و بهره‌بردار اجتناب‌ناپذیر می‌باشد. این امر در سایر کشورها نیز به‌گونه‌ای مناسب صورت پذیرفته است و از دهه ۱۹۸۰، کشورهای توسعه‌یافته به‌طور فزاینده‌ای برای احداث و توسعه تأسیسات زیربنایی به راهکارهای مشارکتی روی آورده‌اند که این امر در ابتدا به دلیل تأمین بودجه و تمایل برای بازدهی بیشتر طرح‌ها بوده است.

در این راستا و با توجه به وجود انواع روش‌های مشارکت بخش خصوصی، انتخاب روش مناسب که با مقتضیات پروژه، محل اجرای آن و سایر عوامل مرتبط با اجرای طرح سازگاری داشته باشد، به یکی از مسائل مهم در این صنعت تبدیل شده و همواره نیازمند مدلی بوده تا به راحتی به مدیران ارشد و سرمایه‌گذاران در خصوص انتخاب روش مناسب برای هر طرح و یا منطقه خاص کمک نماید. به‌منظور تحقق این امر و ارائه مدل مشارکت عمومی - خصوصی در صنعت آب و فاضلاب، ضروری است تا پس از شناسایی نیازهای این صنعت، انواع روش‌های مشارکت بخش خصوصی در اجرای طرح‌ها شناسایی شوند. در قدم بعدی ضروری است تا عواملی که در جذب سرمایه‌گذاران تأثیرگذار هستند موردبررسی قرار گرفته و ارتباط میان آن‌ها مشخص شود. در نهایت و با استفاده از اطلاعات به‌دست آمده و الگوریتم‌های داده‌کاوی، در حوزه انتخاب ویژگی و طبقه‌بندی با روش‌های جنگل تصادفی، گرادیان بوستینگ و استکینگ می‌توان مدل سرمایه‌گذاری بخش خصوصی در این

صنعت را ترسیم نموده و در خصوص هر یک از طرح‌ها و مناطق، روش مناسب اجرای پروژه را پیش‌بینی کرد. از این رو در ادامه در دو بخش اهداف مشارکت بخش خصوصی و روشهای داده‌کاوی، مروری بر کارهای انجام شده خواهیم داشت. سپس تعاریف و الگوریتم‌های داده‌کاوی کار شده در بخش دسته‌بندی پروژه‌ها را تشریح کرده و براساس روش پژوهش، مدل پیشنهادی در بخش پیش پردازش و پاکسازی، بررسی شاخص‌های تأثیرگذار با سود اطلاعاتی و مرحله دسته‌بندی ارایه خواهد شد و در نهایت نتایج اجرا، گزارش و تحلیل می‌شود.

پیشینه تحقیق

مشارکت عمومی- خصوصی (PPP) رویکردی است که برای افزایش ارزش اقتصادی خروجی زیرساخت‌ها مورد استفاده قرار می‌گیرد و شامل طیف وسیعی از زیرساخت‌های بخش عمومی می‌شود (Cui et al., 2018). این مشارکت یک قرارداد بلندمدت بین یک سازمان عمومی و نماینده خصوصی است که به‌موجب آن منابع و ریسک‌های پروژه بین طرفین قرارداد تقسیم می‌گردد و هدف از این قراردادها توسعه و یا نوسازی تسهیلات عمومی است (McQuaid and Scherrer, 2010).

در سال‌های اخیر به دلیل رشد روزافزون جمعیت و توسعه اقتصادی کشورهای مختلف، نیاز شدیدی به ساختارهای زیربنایی در بسیاری از کشورها احساس می‌شود (Hodge et al., 2010) و این در حالی است که بودجه‌های دولتی برای توسعه زیرساخت‌های مورد نیاز یک کشور به‌صورت متعارف محدود و غیر بهینه است و به همین دلیل دولت‌ها تلاش کرده اند با به‌کارگیری بخش خصوصی و فعال کردن این بخش در پروژه‌های زیربنایی، راه‌حلی نوین برای این کاستی‌ها بیابند. در این تحقیق، ضروری است که ابتدا اهداف مشارکت بخش خصوصی در صنعت آب و فاضلاب و در نتیجه شاخص‌های مؤثر در ورود بخش خصوصی به طرح‌های آب و فاضلاب و نیز اقدامات صورت گرفته در زمینه استفاده از داده‌کاوی در مدل سازی و پیش‌بینی مورد مطالعه قرار گیرد. با توجه به این امر، می‌توان پیشینه تحقیق را در ۲ بخش مطرح نمود:

- شناسایی اهداف و در نتیجه استخراج شاخص‌های مشارکت بخش خصوصی در صنعت آب و فاضلاب
- شناسایی و بررسی مطالعات انجام شده در زمینه استفاده از داده‌کاوی در پیش‌بینی

شناسایی اهداف و شاخص‌های مشارکت عمومی- خصوصی در صنعت آب و فاضلاب ایران

بر اساس مطالعات انجام شده در صنعت آب و فاضلاب در سایر کشورها و تطابق آن با مطالعات انجام شده در شرکت مهندسی آب و فاضلاب کشور، مشارکت بخش خصوصی در طرح‌های آب و فاضلاب بر مبنای ۷ هدف کلی دانش و تجربیات تکنیکی، اثربخشی عملیاتی، پاسخگویی به مشتریان، کارایی اقتصادی، سرمایه‌گذاری، دانش و تجربیات مدیریتی و مصون ماندن از مداخلات سیاسی صورت می‌پذیرد (Böhl, 2007, Company, 2009).

به‌منظور استفاده از این اهداف و شناسایی شاخص‌های مشارکت عمومی- خصوصی، می‌توان در هر یک از اهداف تعدادی از شاخص‌های کلیدی عملکرد (KPI)^۵ را دسته‌بندی نمود (Alegre and Association, 2006). با توجه به نتیجه مطالعات انجام شده در سایر کشورها و در قالب هر یک از اهداف ۷ گانه مشارکت بخش خصوصی، می‌توان کلیه KPIهای مؤثر و موجود را شناسایی کرد.

در سال‌های اولیه و با توجه به نوظهور بودن روش‌های مشارکت عمومی - خصوص، مهم‌ترین شاخص‌های این‌گونه طرح‌ها را تکمیل طرح در زمان مقرر و ایجاد حداکثر نتایج عنوان می‌کردند (Yu et al., 2007). لیکن در سال‌های بعد و با توجه به تجربیات به وجود آمده در اجرای طرح‌های مختلف به‌ویژه در حوزه ایجاد زیرساخت‌های شهری، انتخاب روش بهینه به‌منظور موفقیت اجرای پروژه‌ها، مورد توجه محققان قرار گرفته و شاخص‌های متعددی به‌منظور بررسی اجرای طرح‌ها به روش‌های مختلف مورد تحلیل و بررسی قرار گرفت.

با توجه به رشد بالای استفاده از روش‌های مشارکت عمومی - خصوصی از سال ۲۰۰۰ به بعد، در این پژوهش، مبنای مطالعات در شناسایی شاخص‌های مشارکت عمومی - خصوصی، سال ۲۰۰۰ در نظر گرفته شده و شاخص‌های مربوطه استخراج گردیده است.

ازجمله مطالعاتی که در زمینه شناسایی شاخص‌های مشارکت انجام شده است می‌توان به مطالعات تیونگ و همکاران در سال ۲۰۰۱ اشاره کرد که شاخص‌های حیاتی موفقیت (CSF) در پروژه‌های BOT را مورد بررسی قرار داده و ۶ شاخص اصلی نوآوری و کارآفرینی، انتخاب پروژه مناسب، تشکیل ترکیبی قوی از سهامداران، راهکارهای تکنیکی مناسب، وجود پروپوزال مالی قوی و شرایط ویژه مناقصه، را به‌عنوان عوامل حیاتی موفقیت عنوان نموده‌اند (Qiao et al., 2001).

در سال ۲۰۰۳ Akintoye و همکاران، مهم‌ترین شاخص‌های مشارکت بهینه بخش عمومی - خصوصی در اجرای پروژه‌ها را بهبود کیفیت، افزایش کارایی و اثربخشی، استفاده بهینه از منابع مالی و بهره‌مندی از استانداردهای بهتر کارایی، عنوان نمودند (Akintoye et al., 2003). در تحقیقی دیگر، ژنگ و همکارانشان (Zhang, 2005) پس از بررسی و تجمیع کلیه مطالعاتی که تا این سال بر روی شناسایی شاخص‌های مشارکت عمومی - خصوصی انجام شده بود، CSF های اصلی مشارکت عمومی - خصوصی در پروژه‌های زیرساختی را وجود محیط مساعد برای سرمایه‌گذاری، قابلیت حیات اقتصادی، تسهیم ریسک مناسب و وجود قراردادهای مطمئن، وجود مشارکت سرمایه‌گذاران بانفوذ و دارای قدرت تکنیکی بالا و ارائه بسته‌های مالی مناسب، ارائه کرده‌اند. لی و همکاران نیز (Li et al., 2005) با بررسی پروژه‌های مشارکتی انجام شده در سال ۲۰۰۵، CSF های مهم در اجرای طرح‌های مشارکت عمومی - خصوصی را وجود ترکیب مناسب و قوی در بخش خصوصی، ارزیابی واقع‌بینانه هزینه - فایده، وجود بازارهای مالی در دسترس، تقسیم و تسهیم مناسب ریسک، امکان‌سنجی تکنیکی پروژه، حمایت سیاسی، وجود فرآیند رقابتی در تدارکات، وجود شفافیت در فرآیند تدارکات، وجود اهداف و فواید چندبعدی، تعهد و مسئولیت‌پذیری بخش عمومی و خصوصی، حکمرانی مناسب، مشارکت دولت در تأمین ضمانت‌های لازم، وجود اقتصاد کلان باثبات، وجود چارچوب مناسب قانونی، شرایط سیاسی و اقتصادی، وجود ساختارهای خصوصی مناسب، تقسیم حکمرانی بین بخش عمومی و خصوصی، پشتیبانی اجتماعی، انتقال تکنولوژی، عنوان نمودند. بوهل و همکارانش (Böhl, 2007) با انجام تحقیقی در زمینه مشارکت عمومی - خصوصی در صنعت آب و فاضلاب، اقدام به شناسایی این شاخص‌ها در کشور آلمان نمودند. این گروه با شناسایی اهداف مشارکت عمومی - خصوصی در این صنعت، ۳۶ شاخص را به‌عنوان شاخص‌های کلیدی کارایی و موفقیت پروژه‌های PPP اعلام نموده و از طریق بررسی پروژه‌های انجام شده در برخی از کشورها به وزن دهی شاخص‌ها پرداختند. ساش و همکاران (Sachs et al., 2007) نیز طی تحقیقی به بررسی فرصت‌ها و تأثیرات ریسک‌های سیاسی بر روی پروژه‌های مشارکت عمومی - خصوصی حوزه زیرساخت‌ها در چین و ۱۴ کشور دیگر آسیا پرداختند. بسیاری از این پروژه‌ها در حوزه آب و فاضلاب و به‌ویژه ساخت تصفیه‌خانه‌های آب و فاضلاب با مشارکت عمومی - خصوصی بوده است و در طی آن تأثیر ۶ فاکتور و ریسک سیاسی عدم سهولت تسعیر ارز و محدودیت‌های تبادل آن، سلب مالکیت‌ها و انحصارهای دولتی، عدم ایفای تعهدات، تهدیدها و مشکلات

سیاسی، ریسک‌های حقوقی، قانونی و ناشی از بوروکراسی و ریسک‌های غیر متأثر از دولت، بر روی این گروه از طرح‌ها مورد بررسی قرار گرفت. این تحقیق در ادامه، وضعیت هر یک از کشورها را در ۶ حوزه مورد اشاره بررسی کرده و راهکارهای مناسب در هر یک از کشورها را مطرح می‌نماید. در همان سال، سالم‌ن و همکارانش (Salman et al., 2007) بررسی‌هایی را بر روی حوزه پروژه‌های مشارکت عمومی-خصوصی و به‌ویژه پروژه‌های BOT که به‌منظور تکمیل زیرساخت‌های اساسی انجام می‌شود، انجام داده و مهم‌ترین شاخص‌های مؤثر در تصمیم‌گیری و موفقیت را در حوزه‌های مختلف استخراج و دسته‌بندی نموده‌اند. این گروه ۲۱ شاخص را در ۳ دسته عوامل محیطی و حقوقی، عوامل فنی و تکنیکی و عوامل مالی و تجاری مطرح کردند. یان و همکاران (Yuan et al., 2009) انتخاب مدل مشارکت عمومی - خصوص را از دید کلیه ذینفعان پروژه مورد بررسی قرار دادند. ذینفعان مورد بررسی در این تحقیق بخش‌های اصلی جامعه شامل بخش خصوصی، بخش عمومی و دولتی و جامعه دانشگاهی را شامل می‌گردید. این تحقیق ۱۵ شاخص را به‌عنوان اصلی‌ترین شاخص‌های عملکردی در این پروژه‌ها مطرح نمود و اهمیت این شاخص‌ها را در هر یک از گروه‌های ذینفع به تفکیک مشخص کرده است. وی در ادامه با استفاده از روش‌های فازی، شاخص‌های مورد اشاره را وزن دهی و اهمیت آن‌ها را در هر یک از گروه‌های ذینفع مشخص نموده است. همین گروه در تحقیقی دیگر (Yuan et al., 2011) با بررسی پروژه‌هایی که در حوزه مشارکت عمومی-خصوصی انجام شده است، شاخص‌های کلیدی موفقیت را در پنج حوزه اصلی تقسیم کرده‌اند و با استفاده از پرسش‌نامه و ارائه آن به ذینفعان درگیر در پروژه‌های مشارکت عمومی - خصوصی، در مجموع ۴۸ شاخص را به‌عنوان عوامل کلیدی کارایی در پروژه‌های مشارکتی شناسایی کرده و پس از شناسایی، به‌منظور تعیین اهمیت هر یک از شاخص‌ها از پرسش‌نامه‌های مبتنی بر طیف لیکرت^۷ استفاده کرد.

گلابچی و نورزایی (Golabchi, 2015) تحقیقی را با عنوان انتخاب روش مناسب مشارکت خصوصی-دولتی به‌منظور تأمین مالی پروژه‌های آزادراه ایران انجام داده و پس از ارائه دو پرسش‌نامه به خبرگان حوزه مورد پژوهش و تحلیل نتایج با استفاده از روش آماره تی، ۱۹ معیار مهم در انتخاب روش مناسب مشارکت عمومی خصوصی را عنوان نمودند. بانک جهانی نیز در گزارش خود با عنوان ایجاد چارچوب برای مشارکت عمومی-خصوصی در سال ۲۰۱۵، اصلی‌ترین محدودیت‌ها در پیشبرد طرح‌های مشارکتی را محدودیت‌های قانونی عنوان کرده و عوامل مهم را در قالب ۱۵ آیتم عنوان کرده است (Delmon, 2015). مرزوک و همکاران نیز (Marzouk and Fayez, 2018) از طریق استفاده از پرسش‌نامه در حوزه مشارکت عمومی-خصوصی، ۴۶ شاخص را در ۶ حوزه عنوان نمودند.

اگرچه در سال‌های اخیر تحقیقات بسیار زیادی بر روی قراردادهای PPP انجام شده است، اما بررسی‌های انجام شده بر روی این مطالعات نشان داد، شناسایی و ارزیابی معیارهای مؤثر به منظور انتخاب نوع قرارداد مورد استفاده جهت تأمین مالی پروژه‌های صنعت آب و فاضلاب در دنیا و به‌ویژه در ایران، کمتر مورد توجه قرار گرفته است.

مطالعات انجام شده در زمینه استفاده از داده‌کاوی در پیش‌بینی

داده‌کاوی در حوزه‌های زیادی از جمله بازاریابی، تأمین منابع مالی، بانکداری، تولید، سلامت، مدیریت روابط مشتریان، تشخیص خطا، پیش‌بینی و یادگیری سازمانی استفاده شده است که در ادامه به اختصار به برخی از این مطالعات اشاره می‌شود.

بیکزاده و همکاران (Beikzadeh et al., 2008) از داده‌کاوی در استخراج قواعد و طبقه‌بندی داده‌ها به‌منظور پیش‌بینی و کمک به بهبود فرآیند تصمیم‌گیری در دستگاه‌های آموزشی و مؤسسات آموزش عالی پرداختند. ال‌ریدیه و

همکاران (Al-Radaideh and Al Nagi, 2012) از داده‌کاوی و تکنیک‌های طبقه‌بندی، جهت بهبود عملکرد دستگاه‌های آموزشی در مؤسسات آموزش عالی استفاده نمودند. در این تحقیق در ابتدا عوامل اصلی اثرگذار بر کارایی و عملکرد دانشجویان شناسایی و اطلاعات آن‌ها جمع‌آوری گردید و در نهایت با استفاده از درخت تصمیم و داده‌های مربوط به هر دانشجو، پیش‌بینی عملکرد و نمرات دانشجو انجام گرفت. این مطلب توسط کیریمی و همکاران (Kirimi and Moturi, 2016) نیز مورد ارزیابی قرار گرفت که در این پژوهش بسیاری از داده‌های کیفی نیز به مطالعه افزوده شد.

لاوراک و همکاران (Lavrač et al., 2007) از الگوریتم‌های طبقه‌بندی در خصوص پیش‌بینی مراجعه و بررسی لزوم ارجاع بیماران به متخصصین استفاده نمودند. در این تحقیق، دسته‌بندی بیماران در مراکز بهداشت از خدمات اولیه سلامت تا رجوع به متخصصین، با استفاده از ترکیبی از داده‌کاوی و رویکردهای پشتیبان تصمیم مورد بررسی قرار گرفته و در نهایت یک واسط کاربری مناسب جهت کمک به مدیریت دانش و پشتیبانی تصمیم ارائه گردیده است.

چنگ و همکاران (Cheng et al., 2009) از الگوریتم‌های طبقه‌بندی، در طراحی سیستم مدیریت دانش مالی^۸ (FKMS) استفاده نموده و از ترکیب الگوریتم‌های خوشه‌بندی و تکنیک‌های طبقه‌بندی داده‌کاوی جهت طبقه‌بندی داده‌های گروهی و منسجم استفاده نمودند. چن و همکارانش (Chen and Du, 2009) نیز در همان سال از تکنیک‌های داده‌کاوی و شبکه‌های عصبی به منظور پیش‌بینی بحران‌های مالی بهره بردند. در این پژوهش، مدلی طراحی گردید که با دریافت اطلاعات مربوط به شرکت‌ها، وضعیت مالی آن‌ها مورد بررسی قرار گرفته و احتمال بروز بحران مالی را پیش‌بینی شود.

لی و همکاران (Li et al., 2010) نمونه‌ای از الگوریتم طبقه‌بندی را در صنایع کوچک و متوسط پیاده‌سازی نمودند. در این روش، ترکیبی از داده‌کاوی و مدیریت دانش به منظور طراحی یک سیستم پشتیبان تصمیم مورد استفاده قرار گرفته است. در این تحقیق که به منظور پیش‌بینی و بهینه‌سازی زمان رسیدن غذا در زنجیره تأمین یک شرکت تأمین غذا صورت گرفت، در ابتدا عوامل مؤثر بر روی زمان رسیدن غذا شناسایی گردید و سپس با استفاده از درخت تصمیم و شبکه‌های عصبی، زمان رسیدن غذا پیش‌بینی گردید.

ریبیرو و همکاران (Ribeiro et al., 2013) از تکنیک‌های داده‌کاوی به منظور بررسی و پیش‌بینی کارایی در تصفیه‌خانه‌های فاضلاب استفاده کردند. در این پژوهش از تکنیک‌های^۹ SVM بر روی داده‌های موجود استفاده شده و کارایی تصفیه‌خانه پیش‌بینی شد.

سان و همکاران (Sun et al., 2014) میزان شکستگی و اتفاقات صورت گرفته در شبکه‌های آب را توسط خوشه‌بندی فضایی مورد تجزیه و تحلیل قرار داده و سپس با خوشه‌بندی مبتنی بر چگالی، محدوده‌هایی از شبکه تأمین آب که در معرض اتفاقات هستند را شناسایی کردند.

یانگ و همکاران (Qiu et al., 2018) در پژوهشی، یک سیستم مبتنی بر داده‌کاوی را طراحی و پیاده‌سازی نمودند که فرآیند طراحی و ساخت تصفیه‌خانه‌های فاضلاب را در کشور چین بهینه نماید.

جنابی و همکاران (Jenabi and Mirroshandel, 2014) از این تکنیک جهت بهبود مدیریت امور مشتریان استفاده کردند.

علاوه بر آنچه در بالا آورده شده است، در تحقیقات بسیاری نیز (Azhar et al., 2013, OGWUELEKA and OGWUELEKA, 2010, Ali and Qamar, 2013, Wen et al., 2013) پیش‌بینی کیفیت آب با استفاده از الگوریتم‌های داده‌کاوی انجام شده است.

مبانی نظری

در این پژوهش از تکنیک های داده کاوی جهت تحقق ۲ هدف اصلی استفاده می گردد. در ابتدا لازم است که بتوان موفقیت و یا عدم موفقیت یک پروژه را در قالب قراردادی مشخص پیش بینی نمود و در مرحله بعد بتوان عدد مربوط به میزان موفقیت را به عنوان خروجی ارائه کرد. در این تحقیق الگوریتم های مختلفی مورد استفاده قرار گرفته اند که شامل گرادیان بوسستینگ، شبکه عصبی، رگرسیون لاجستیک، یادگیری عمیق و بگینگ و استکینگ است.

➤ شبکه عصبی

شبکه های عصبی مصنوعی الگویی برای پردازش اطلاعات می باشند که با تقلید از شبکه های عصبی بیولوژیکی مثل مغز انسان ساخته شده اند. عنصر کلیدی این الگو ساختار جدید سیستم پردازش اطلاعات آن است و از تعداد زیادی عناصر (نورون) با ارتباطات قوی داخلی که هماهنگ باهم برای حل مسائل مخصوص کار می کنند، تشکیل شده است. شبکه های عصبی مصنوعی با پردازش روی داده های تجربی، دانش یا قانون نهفته در ورای داده ها را به ساختار شبکه منتقل می کند که به این عمل یادگیری می گویند. هر شبکه عصبی شامل یک لایه ورودی است که هر گره در این لایه معادل یکی از متغیرهای پیش بینی است. گره های موجود در لایه میانی به تعدادی گره در لایه نهان، وصل می شوند. هر گره ورودی به همه گره های لایه نهان متصل می شود. برای شبیه سازی یک روش طبقه بندی بر روی شبکه های عصبی، میبایست در لایه ی ورودی به تعداد صفات خاصه نورون قرار دهیم. در لایه خروجی نیز باید به تعداد کلاس ها نورون داشته باشیم. شبکه های عصبی برای دسته بندی و پیشگویی در داده کاوی کاربردهای فراوان دارند. شبکه های عصبی می توانند هم برای دسته بندی (برای پیش بینی برچسب کلاس یک چندتایی داده شده) و هم برای پیشگویی (برای پیش بینی مقادیر پیوسته خروجی) استفاده شود (HAN, PEI & KAMBER, 2011).

➤ درخت تصمیم:

درختان تصمیم، درختانی هستند که نمونه ها را توسط دسته ی آن ها بر اساس مقادیر ویژگی هایشان دسته بندی می کنند و در درخت تصمیم دسته بندی با معیار چند دسته ای انجام می شود. درخت تصمیم شبیه ساختار درختی است که نمونه ها را با مرتب سازی آن ها بر اساس مقادیر ویژگی دسته بندی می کند. هر گره در این درخت نشان دهنده ی ویژگی ها در نمونه طبقه بندی شده است. همه شاخه ها نشان دهنده ی خروجی تست هستند و هر برگ برچسب کلاس را در خود نگه می دارد. نمونه ها بر اساس مقادیر ویژگی ها دسته بندی می شود. روش های زیادی برای پیدا کردن ویژگی که بهترین تقسیم داده های آموزشی را بدهد وجود دارد.

رایج ترین روش ساخت درخت تصمیم روش حریصانه بالا به پایین است. درخت تصمیم قوانینی برای دسته بندی دیتا ست تولید می کند. سه الگوریتم رایج C4.5، ID3 و CART وجود دارد (HAN, PEI & KAMBER, 2011).

از جمله مفاهیم مطرح در این حوزه، آنتروپی میباشد که میزان خلوص (بی نظمی یا عدم خالص بودن) مجموعه ای از مثال ها را مشخص می کند و هر چه این مقدار بیشتر باشد، یعنی داده ها اطلاعات کمتری را در خود دارند. اگر مجموعه S

شامل مثال‌های مثبت و منفی از یک مفهوم هدف باشد، آنتروپی S نسبت به این دسته‌بندی بولی به صورت رابطه ۱ تعریف می‌شود:

$$\log_2 P^- \times P^- - \log_2 P^+ \times \text{Entropy}(S) = -P^+ \quad (1)$$

➤ آدابوست^{۱۰}:

مخفف بوستینگ تطبیقی بوده و یک الگوریتم یادگیری ماشین است که توسط یاو فروند و رابرت شاپیر ابداع شد (Freund and Schapire 1995). این روش از روش‌های جمعی یا ترکیبی است که یک مجموعه از دسته‌بندی‌های ضعیف را گرفته و خروجی آن‌ها را با یکدیگر ترکیب می‌کند تا دسته‌بند نهایی را به گونه‌ای بسازد، که کارایی آن از کارایی تک تک دسته‌بندی‌های استفاده شده در الگوریتم بیشتر باشد. درواقع آدابوست یک متا الگوریتم است که به منظور ارتقاء عملکرد، و رفع مشکل رده‌های نامتوازن همراه دیگر الگوریتم‌های یادگیری استفاده می‌شود. در این الگوریتم، طبقه بند هر مرحله جدید به نفع نمونه‌های غلط طبقه‌بندی شده در مراحل قبل تنظیم می‌گردد. آدابوست نسبت به داده‌های نویزی و پرت حساس است. ولی نسبت به مشکل بیش برآزش از بیشتر الگوریتم‌های یادگیری برتری دارد. طبقه بند پایه که در اینجا استفاده می‌شود فقط کافی است از طبقه بند تصادفی (۵۰٪) بهتر باشد و به این ترتیب بهبود عملکرد الگوریتم با تکرارهای بیشتر بهبود می‌یابد. حتی طبقه بندی‌های با خطای بالاتر از تصادفی با گرفتن ضریب منفی عملکرد کلی را بهبود می‌بخشند.

➤ پشته سازی یا استکینگ^{۱۱}:

در این روش مجموعه‌ای از طبقه‌بندی کننده‌ها با استفاده از نمونه بوت استرپ داده‌های آموزشی، آموزش داده می‌شوند. ایده اصلی استکینگ این است که ببینیم آیا داده‌های آموزشی به خوبی آموزش دیده است یا نه. مثلاً اگر یک طبقه‌بندی کننده‌ی خاص در منطقه خاصی از ویژگی‌ها اشتباه آموزش داده شده باشد و به همین دلیل نمونه‌های آن منطقه را به اشتباه دسته‌بندی کرده باشد، ردیف دوم طبقه‌بندی کننده‌ها قادر به یادگیری این رفتار است و می‌تواند با استفاده از این رفتاری که یاد گرفته و رفتارهای یاد گرفته شده از دیگر طبقه‌بندی کننده‌ها، چنین آموزش نادرستی را اصلاح کند. انتخاب نوع ارزیابی کراس برای آموزش طبقه‌بندی کننده‌های ردیف اول استفاده می‌شود. کل داده‌های آموزشی به T بلوک تقسیم می‌شوند و هر طبقه‌بندی کننده‌ی ردیف اول، اولین بار بر روی مجموعه‌ای متفاوت از $T-1$ بلوک از داده‌های آموزش داده می‌شود. سپس هر طبقه‌بندی کننده بر روی T امین بلوک که در طول آموزش با آن مواجه نشده است، ارزیابی می‌شود. خروجی این طبقه‌بندی کننده‌ها همراه با برچسب واقعی درست برای آن بلوک‌ها، مجموعه داده آموزشی برای ردیف دوم طبقه‌بندی کننده‌ها را تشکیل می‌دهند (Rokach, 2010).

➤ جنگل تصادفی^{۱۲}:

جنگل تصادفی درخت تصمیم‌های زیادی را تولید می‌کند. برای طبقه‌بندی یک شیء جدید، بردار ورودی را در انتهای هر یک از درختان جنگل تصادفی قرار می‌دهد. هر درخت یک طبقه‌بندی را ارائه می‌کند و اصطلاحاً گفته می‌شود این درخت به آن کلاس «رأی» می‌دهد. جنگل، طبقه‌بندی که بیشترین رأی را داشته باشد (بین همه درخت‌های جنگل) انتخاب می‌کند (کرمیچ و همکاران، ۲۰۱۵). هر درخت به صورت زیر تشکیل می‌شود:

۱. اگر N تعداد حالت‌ها در مجموعه داده‌های train (مجموعه‌ی کار) باشد، N حالت را به‌صورت تصادفی با جایگذاری از داده‌های اصلی، نمونه‌گیری می‌کنیم. این نمونه مجموعه‌ی کار برای این درخت است.

۲. اگر M متغیر داشته باشیم و m را کوچکتر از M در نظر بگیریم به‌طوری‌که در هر گره، m متغیر به‌صورت تصادفی از M انتخاب می‌شوند و بهترین جداسازی روی این m متغیر برای جداسازی گره استفاده می‌شود. مقدار m در طول ساخت جنگل ثابت در نظر گرفته می‌شود.

۳. هر درخت به‌اندازه‌ی ممکن بزرگ می‌شود. هیچ هرسی وجود ندارد.

➤ گرادیان بوستینگ^{۱۳}:

روش گرادیان بوستینگ^۱ مانند جنگل تصادفی از درخت‌های تصمیم «ضعیف» استفاده می‌کند. تفاوت بزرگ این دو روش آن است که در روش گرادیان بوستینگ، درخت‌ها یکی پس از دیگری آموزش داده می‌شوند. هر درخت زیرمجموعه، در درجه اول با داده‌هایی که به اشتباه توسط درخت پیشین پیش‌بینی شده‌اند آموزش داده می‌شوند. این امر موجب می‌شود مدل کمتر بر مسائلی که پیش‌بینی در آن‌ها آسان است تمرکز داشته و بیشتر روی موارد پیچیده متمرکز شود.

روش تحقیق

در این تحقیق به‌منظور شناسایی روش‌های مشارکت بخش خصوصی در طرح‌های آب و فاضلاب و نیز عوامل مرتبط با آن‌ها، از مطالعات کتابخانه‌ای و روش‌های میدانی استفاده می‌شود. در ادامه به‌منظور درک تأثیرات متقابل هر یک از این عوامل بر روی یکدیگر و درنهایت ارائه مدل مفهومی مشارکت بخش خصوصی در طرح‌های آب و فاضلاب، از الگوریتم‌های داده‌کاوی استفاده می‌گردد. به منظور انجام داده‌کاوی، از متدولوژی CRISP-DM استفاده شده و پیاده‌سازی با نرم افزار Rapid miner ورژن ۹ انجام شده است. در ادامه گام‌های روش تحقیق کریسپ در این پژوهش به اختصار تشریح می‌شود:

➤ شناخت بازار(درک فضای کسب‌وکار): بر مبنای اطلاعات موجود در وب‌سایت رسمی تجهیز منابع مالی و

توسعه مشارکت بخش غیردولتی صنعت آب و فاضلاب^{۱۴}، این بخش به‌عنوان صنعت پیشرو در حوزه ارائه روش‌های نوین جذب مشارکت بخش خصوصی همواره در سطح کشور مطرح بوده است و توانسته با ابداع قالب‌های مختلف قراردادهای مشارکت عمومی- خصوصی و انعقاد قراردادهایی در هر قالب، تعداد بسیاری از پروژه‌های زیرساختی این حوزه را در کمترین زمان و با بالاترین کیفیت به بهره‌برداری رساند. با توجه به گذشت بیش از ۱۷ سال از انعقاد قراردادهای مشارکتی در این صنعت، تجربه و آمار و اطلاعات مناسبی در این حوزه وجود دارد که می‌تواند کمک شایانی به پیشبرد تحقیق نماید. در حال حاضر، دغدغه فعلی تصمیم‌گیران در این صنعت، انتخاب روش مناسب مشارکت عمومی- خصوصی در پیشبرد طرح‌ها است که با توجه به حجم بالای سرمایه‌گذاری موردنیاز و اهمیت بسیار زیاد اجرای این طرح‌ها در زندگی افراد جامعه، این موضوع به یکی از دغدغه‌های اصلی مبدل شده است. از سوی دیگر، انجام داده‌کاوی نیازمند دسترسی به داده‌های گذشته است

که با توجه به وجود ۱۷۶ پروژه مشارکت عمومی- خصوصی در این صنعت، می‌توان به بررسی روند گذشته پرداخت.

➤ **درک داده‌ها:** پس از جمع آوری داده از منابع مختلف جهت اجرای پروژه‌های داده‌کاوی، فرمت داده، تعداد متغیرها، نوع متغیرها (پیوسته و گسسته) و مفهوم متغیرها، موردبررسی قرار گرفت. هم‌چنین با استفاده از روش‌های آماری مشخصات اصلی داده‌ها استخراج شد. خروجی این مرحله معمولاً به‌صورت گزارشی است که در آن اشتباهات داده‌ها، مقادیر ازدست‌رفته و وجود رکوردهای تکراری مشخص می‌گردد. منظور از اشتباه‌های داده‌ها، مقادیر بی‌معنی و داده دورافتاده و پرت است که در نتایج تأثیر می‌گذارد.

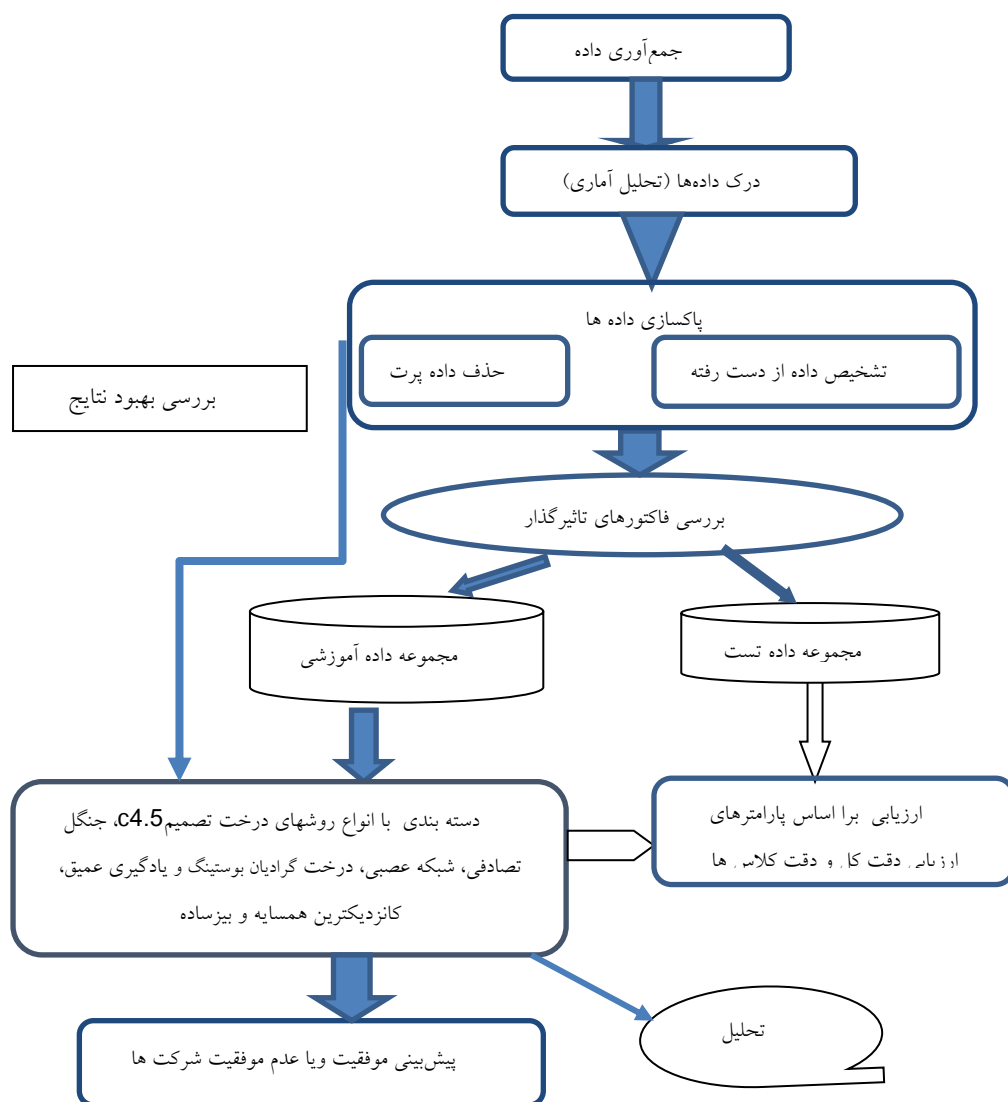
➤ **آماده سازی داده‌ها:** با استفاده از گزارش خروجی بررسی کیفیت داده‌ها در مرحله قبل، در صورت وجود داده‌های از دست رفته، داده‌های نویز و پرت، سطرهای تکراری باید حذف و داده پاکسازی شده جهت مدل‌سازی آماده شود.

➤ **مدل‌سازی:** در این مرحله ابتدا چندین تکنیک دسته‌بندی یا خوشه‌بندی براساس نوع داده‌ها انتخاب خواهد شد. با توجه به مرحله پیش‌پردازش، ممکن است در این مرحله براساس پارامترهای ارزیابی، نیاز به بازگشت به عقب داشته باشیم و تکنیک‌های مختلف دسته‌بندی و پارامترها را بررسی کنیم.

➤ **ارزیابی:** در این مرحله می‌توان با توجه به نوع داده‌ها، شاخص‌های ارزیابی مدل را انتخاب نمود؛ و بهترین روش و دسته‌بند در هر مرحله بر اساس این ویژگی انتخاب می‌شود.

مدل پیشنهادی

با توجه به روش اجرای پژوهش، مدل پیشنهادی براساس تکنیک های داده کاوی، به شرح نمودار ۱ است.



شکل (۱): نمودار مدل کلی تحقیق

- **جمع آوری داده‌ها:** به‌منظور استخراج لیست پروژه‌ها، تمامی استان‌های کل کشور مورد مطالعه قرار گرفته و پروژه‌های مشارکتی که در این استان‌ها با روش‌های مختلف محقق گردیده است استخراج شده‌اند. جداول ارسالی به شرکت‌های آب و فاضلاب کل کشور، شامل لیست پروژه‌ها و شاخص‌هایی بوده که می‌بایست برای آن‌ها ورود اطلاعات انجام می‌شده است. با توجه به اهمیت این پروژه در بخش آب و فاضلاب، اطلاعات تکمیل شده هر استان، پس از امضاء مدیرعامل محترم آن استان به شرکت مهندسی آب و فاضلاب کشور ارسال شده است و به‌منظور بررسی و صحت سنجی اولیه اطلاعات تکمیل شده، مجدداً اطلاعات با بانک داده مجامع شرکت‌ها که در شرکت مهندسی آب و فاضلاب کشور قرار دارد، تطبیق داده شده است.

- **شناخت داده‌ها:** در این مجموعه داده‌ها، ۱۷۶ پروژه با ۲۷ شاخص در ۴ قالب قراردادی BOO، BOT، بیع متقابل و فاینانس جاری اجرا می‌شود. درصد موفقیت پروژه‌ها به‌صورت دقیق مشخص است و برای ورود به الگوریتم اعداد در ۳ گروه Success (موفقیت بالای ۷۵ درصد)، semi-success (موفقیت بین ۵۰ تا ۷۵ درصد) و Failure (موفقیت زیر ۵۰ درصد) طبقه‌بندی شده‌اند.

شاخص‌های استخراج شده شامل قالب قراردادی، سال اجرای طرح، قابلیت فروش محصول در استان، تصفیه فاضلاب (درصد)، آب به حساب نیامده (درصد)، نسبت حوادث شبکه به طول کل شبکه (درصد)، نسبت حوادث انشعابات به کل انشعابات (درصد)، شکایات مربوط به کیفیت خدمات‌رسانی، کیفیت آب مصرفی (درصد)، جمعیت تحت پوشش خدمات آب (درصد)، جمعیت تحت میزان پوشش خدمات فاضلاب (درصد)، قطعی آب نسبت به تعداد انشعابات (درصد)، کمبود فشار آب (درصد)، تعرفه‌های متعارف (فروش به قیمت تمام‌شده در نقطه سربه‌سر)، سود (زیان) عملیاتی (درصد)، دوره وصول مطالبات، نسبت کاری (نسبت پوشش هزینه‌ها)، نسبت سرمایه‌گذاری طی سال به فروش (درصد)، نسبت گردش دارایی‌های ثابت (درصد)، سطح بدهکاری (نسبت پوشش بدهی)، سطح اندازه‌گیری (درصد)، هزینه تمام‌شده آب در نقطه سربه‌سر (ریال)، هزینه تمام‌شده فاضلاب دفع شده در نقطه سربه‌سر (ریال)، بهره‌وری نیروی انسانی (ناخالص)، اثربخشی نیروی انسانی (ناخالص)، سرانه هزینه نیروی انسانی (میلیون ریال)، درجه حرفه‌ای شدن شرکت، آموزش و غیبت میباشند که به غیر از قالب قراردادی، بقیه متغیرها به صورت عددی می‌باشند. نام پروژه به عنوان ID و میزان موفقیت پروژه نیز متغیر هدف تعریف شده است.

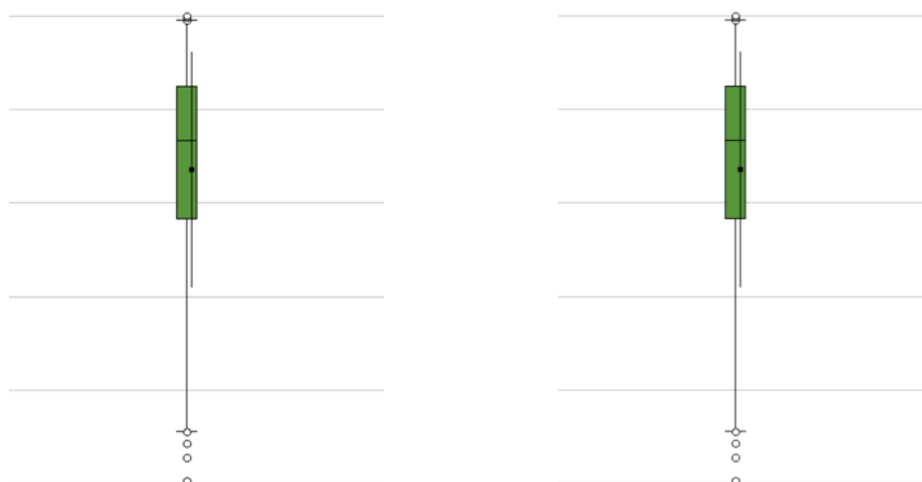
- **تحلیل آماری:** اطلاعات آماری مجموعه داده برای تعدادی از شاخص‌ها به عنوان نمونه در جدول شماره ۱ آورده شده است. ستون اول نوع داده، ستون دوم تعداد داده‌های از دست‌رفته هر متغیر، ستون سوم و چهارم ماکسیمم و مینیمم متغیرهای عددی و کمترین و بیشترین مقدار متغیرهای گسسته و ستون آخر مقدار میانگین و انحراف معیار متغیرهای عددی و مقدار مد متغیرهای گسسته را نشان می‌دهد. نتایج حاصله نشان می‌دهد که ۹۰ درصد شاخص‌ها از نوع عددی هستند و هیچ یک از متغیرها مقادیر از دست رفته ندارند. علاوه بر این، با توجه به این که در برخی از شاخص‌ها انحراف معیار بالایی مشاهده می‌گردد، پیش‌بینی می‌شود که داده‌های پرت وجود دارد.

جدول (۱): اطلاعات آماری مجموعه داده

اطلاعات آماری			داده از دست رفته	نوع	نام شاخص
Least Failure(14)	Max Success(135)	Values: Success& Semi	0	Polynomial	میزان موفقیت پروژه

					Success & Failure
سال اجرای طرح	Integer	0	Min 1385	Max 1395	Average 1391
قابلیت فروش محصول در استان	Integer	0	Min 1	Max 9	Average 4.1
تصفیه فاضلاب (درصد)	Integer	0	Min 0	Max 79.6	Average 33.48
آب به حساب نیامده (درصد)	Integer	0	Min 17.5	Max 41	Average 24.6
نسبت حوادث شبکه به طول کل شبکه (درصد)	Integer	0	Min 0.9	Max 29	Average 2.89
نسبت حوادث انشعابات به کل انشعابات (درصد)	Integer	0	Min 2	Max 27	Average 12.37

- **مصورسازی داده‌ها :** در این مرحله، نمودار quartile مربوط به هر یک از شاخص‌ها در نرم افزار رسم شد که شکل های ۱ و ۲ نمونه‌هایی از این نمودارها را نشان میدهند. در این نوع از نمودارها که برای تشخیص داده‌های دورافتاده استفاده می‌شود، حباب‌ها نقاط پرت را مشخص می‌کنند که با توجه به این نکته، در برخی از شاخص‌ها مقادیر دور افتاده زیادی داریم.



شکل (۲): نمودار quartile شاخص سود و زیان عملیاتی

شکل (۳): نمودار quartile نسبت کاری

- **پاکسازی داده‌ها (تشخیص داده پرت):** با توجه به این که بر اساس واریانس‌ها و نمودارهای مشخص شده در مبحث داده‌های ورودی، برخی از داده‌ها پرت می‌باشند، حذف این داده‌ها می‌تواند به بهبود مدل کمک کند. داده پرت مشاهده‌ای است که در فاصله دورتری از سایر داده‌ها قرار می‌گیرد و با مقدار مورد انتظاری که داریم متفاوت است. بر روی مجموعه داده دو روش LOF^{15} و COF برای تشخیص نقاط پرت و سپس حذف آن اعمال شد. این روش‌ها چگالی هر نقطه را محاسبه کرده و رتبه داده پرت را گزارش می‌دهند.

الگوریتم فاکتور داده پرت محلی یا LOF اولین بار در مقاله ای در سال ۲۰۰۰، معرفی شد. در این روش پرکاربرد، در مرحله اول برای هر نقطه K تا نزدیک‌ترین همسایه یک نقطه محاسبه می‌شود. سپس برای محاسبه فاصله قابل دسترسی یک

نقطه نسبت به سایر نقاط، بیشترین مقدار فاصله دو نقطه و K تا نزدیک‌ترین همسایه‌هایش در نظر گرفته می‌شود. در مرحله بعد جهت محاسبه چگالی محلی هر نقطه، معکوس میانگین فاصله قابل دسترسی نسبت به سایر نقاط محاسبه می‌شود و در آخر، رتبه داده پرت میانگین نرخ چگالی محلی هر نمونه داده نسبت به همسایگانش معرفی می‌شود. این الگوریتم نرخ داده پرت را به‌طور تقریبی بالاتر از ۱ و ۱,۱ معرفی می‌کند. یعنی داده‌های نرمال را کمتر از ۱,۱ و ۱ تعریف می‌کنیم. از این‌رو در نرم‌افزار بین انتخاب $LOF\ outlier\ score \leq 1.1$ و $LOF\ outlier\ score \leq 1.15$ نتایج را شبیه‌سازی کردیم و بهترین رنج، مقدار ۱,۱۵ در نظر گرفته شد و داده‌هایی با وزن کمتر از ۱,۱۵ به‌عنوان داده نرمال و مابقی حذف شدند. پس از اعمال این الگوریتم، از مجموع ۱۷۶ پروژ، ۱۲۱ پروژ باقی ماندند.

در الگوریتم COF که برای بهبود روش LOF پیشنهاد شد، مقدار نمونه‌های پرت را از مقدار ۱ تا ۱,۱۵ شبیه‌سازی کرده و بهترین حد آستانه نقاط پرت را برای آن در نظر گرفتیم که تا ۱,۱۵ محاسبه شد. پس از اعمال این الگوریتم، از مجموع ۱۷۶ پروژ، ۱۰۹ پروژ باقی ماندند.

نتایج اجرای روش‌های مذکور با مقدار $k=10$ (مقدار نزدیک‌ترین همسایه) در جدول ۲ گزارش شده است که رنج رتبه داده پرت، نمونه‌های دورافتاده که میبایست حذف شوند را اثبات می‌کند. بر مبنای نتایج به دست آمده، روش COF دامنه بیشتری برای داده پرت محاسبه و نمونه‌های دورافتاده را بهتر شناسایی کرده است.

در مرحله دسته‌بندی، تفاوت دقت هر دسته‌بند بدون حذف داده پرت و با حذف داده پرت مشخص خواهد شد. و نتایج دو روش LOF و COF مقایسه خواهد شد.

جدول (۲): رنج نمونه‌های پرت در هر روش

روش	تنظیمات	متغیر «داده پرت»: عددی حقیقی
LOF	K کوچک: ۱۰ K بزرگ: ۲۰ فاصله: اقلیدسی نوع اندازه‌گیری: اندازه‌گیری ترکیبی	رنج داده پرت: 0.983-1.192 میانگین: 1.113 انحراف معیار: 0.0056
COF	K کوچک: ۱۰ فاصله: اقلیدسی نوع اندازه‌گیری: اندازه‌گیری ترکیبی	رنج داده پرت: 0.469-2.765 میانگین: 1.110 انحراف معیار: 0.432

- **انتخاب ویژگی:** در این بخش روش انتخاب ویژگی شاخص آنتروپی (سود اطلاعاتی) را جهت ارزیابی و انتخاب متغیرهای تأثیرگذار انتخاب کردیم. این روش وزن تأثیری ویژگی‌ها براساس سود اطلاعاتی هر شاخص، در طبقه‌بندی شکست و یا موفقیت پروژ با مقداری بین صفر و یک در نرم افزار رپیدماینر، گزارش داد که براساس آن نتایج زیر حاصل شد:
- با توجه به وزن ویژگی‌ها، ۹۸ درصد ویژگی‌ها وزن بالایی دارند و در طبقه‌بندی و پیش‌بینی شکست و یا موفقیت پروژه‌ها، تأثیر زیادی دارند.

- سطح شکایات مربوط به کیفیت خدمات‌رسانی، به‌طور مستقیم با بالاترین وزن ۱ روی پیش‌بینی شکست تأثیر دارد.
- نوع قالب قراردادی، نسبت پوشش هزینه‌ها و تعرفه‌های متفاوت تأثیرگذارترین متغیرها هستند.
- بحث تصفیه فاضلاب، درجه حرفه‌ای شدن و سطح بدهکاری با وزن پایین، تأثیر بسیار کمی در پیش‌بینی شکست و یا موفقیت پروژه‌ها دارد.

مدل‌سازی (طبقه‌بندی)

- در این پژوهش، در گام دسته‌بندی نتایج دسته بندها یکبار بدون انتخاب ویژگی و حذف داده پرت، یکبار با روش LOF و انتخاب ویژگی سود اطلاعاتی و یکبار با روش COF و انتخاب ویژگی پیاده‌سازی و نتایج مقایسه و تحلیل شده است.
- جهت اعتبارسنجی داده‌ها و تقسیم داده‌ها به دو قسمت آموزشی و آزمایشی از روش هوشمندانه اعتبارسنجی ضربدری به جای اعتبارسنجی ساده استفاده شده است. پس از مراحل پیش پردازش مجموعه داده‌ها، از رده بندهای پایه مانند درخت تصمیم، بیزساده و شبکه عصبی و در ادامه از روش‌های ترکیبی مانند یادگیری عمیق، آدابوست درخت تصمیم، جنگل تصادفی و گرادیان بوستینگ و یک روش که از ترکیب شبکه عصبی و درخت تصمیم با نام استکینگ در نرم افزار استفاده کردیم. پارامترهای در نظر گرفته جهت درخت تصمیم شاخص سود اطلاعاتی، حداکثر عمق ۱۵، حداقل سائز برگ ۳ و حداقل تعداد رکورد جداسازی ۴ در نظر گرفته شد. شبکه عصبی پرسپترون با ضریب مومنتوم ۰.۸ و نرخ یادگیری ۰.۳ تعیین شد. گرادیان بوستینگ با ۲۰ درخت، جنگل تصادفی با ۲۰ مدل درخت در نظر گرفته شد.
- از مجموع ۲۹ شاخص پس از شبیه سازی، ۲۷ شاخص انتخاب و شاخصهای تصفیه فاضلاب و درجه حرفه ای شدن شرکت که وزن بسیار پایینی داشتند، حذف شدند. زیرا حذف این دو شاخص دقت دسته‌بندها را افزایش داد. نتایج براساس پارامترهای ارزیابی مقایسه شد.

پارامترهای ارزیابی

- قبل از آنکه به بررسی انواع مهم معیارهای دسته‌بندی بپردازیم، لازم است که مفهوم ماتریس درهم‌ریختگی را که در این بخش مورد استفاده قرار گرفته، تشریح شود، این ماتریس چگونگی عملکرد الگوریتم دسته‌بندی را با توجه به مجموعه داده ورودی به تفکیک انواع رده‌های مسئله دسته‌بندی نشان می‌دهد. با توجه به جدول ۳، مفاهیم مثبت درست^{۱۶}، مثبت غلط^{۱۷}، منفی درست^{۱۸} و منفی غلط^{۱۹} به شرح ذیل است.

جدول (۳): ماتریس درهم‌ریختگی

		نمونه‌های تخمینی	
		دسته منفی (-)	دسته مثبت (+)
نمونه‌های واقعی	دسته منفی (-)	TN	FP
	دسته مثبت (+)	FN	TP

- ✓ مثبت درست: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و دسته‌بند نیز رده آن‌ها را به‌درستی مثبت تشخیص داده است.
 - ✓ مثبت غلط: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و دسته‌بند رده آن‌ها را به نادرستی مثبت تشخیص داده است.
 - ✓ منفی درست: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها منفی بوده و دسته‌بند نیز دسته آن‌ها را به‌درستی منفی تشخیص داده است.
 - ✓ منفی غلط: این مقدار بیانگر تعداد رکوردهایی است که دسته واقعی آن‌ها مثبت بوده و دسته‌بند رده آن‌ها را به نادرستی منفی تشخیص داده است.
- اکنون به تشریح انواع مهم معیارهای ارزیابی الگوریتم‌های دسته‌بندی می‌پردازیم. اولین معیار، دقت یا نرخ دسته‌بندی است و به این مفهوم است که دسته‌بند، چند درصد رکوردهای آزمایشی را به‌درستی تشخیص داده است. این دقت براساس مفاهیم ماتریس با توجه به رابطه ۲ قابل محاسبه می‌شود:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (۲)$$

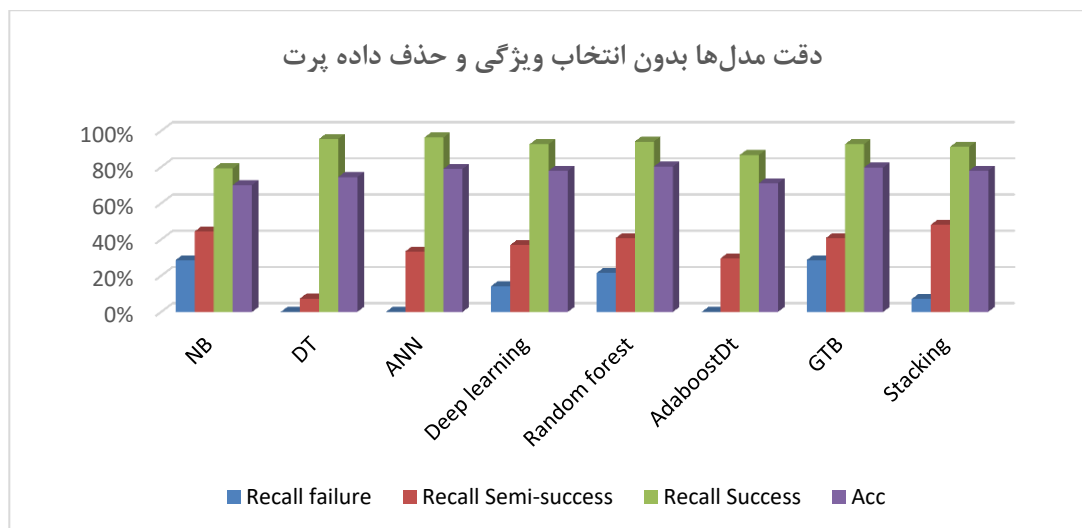
معیار دیگر، Recall یا دقت کلاس‌های مثبت و یا منفی است. در این معیار صورت کسر تعداد تشخیص‌های درست کلاس مثبت (منفی) نسبت به کل نمونه‌هایی که واقعاً مثبت (منفی) بوده‌اند.

$$\text{Recall}^+ = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (۳)$$

$$\text{Recall}^- = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (۴)$$

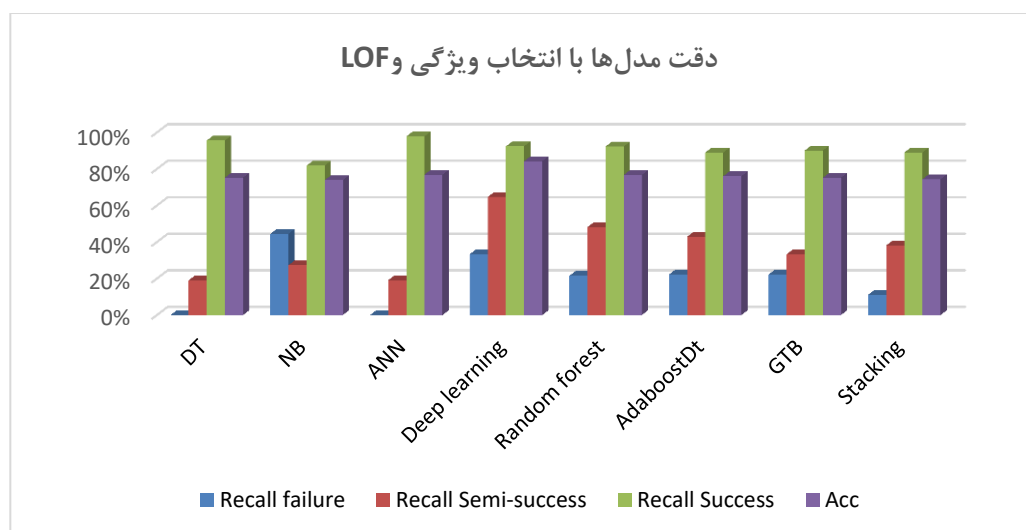
نتایج و بحث

نتایج تکنیک‌های دسته‌بندی را در سه مرحله برای هر دسته‌بند اجرا و نتایج دقت کل، و دقت هر کلاس نوع پروژه، به شرح شکل‌های ۴، ۵ و ۶ است.



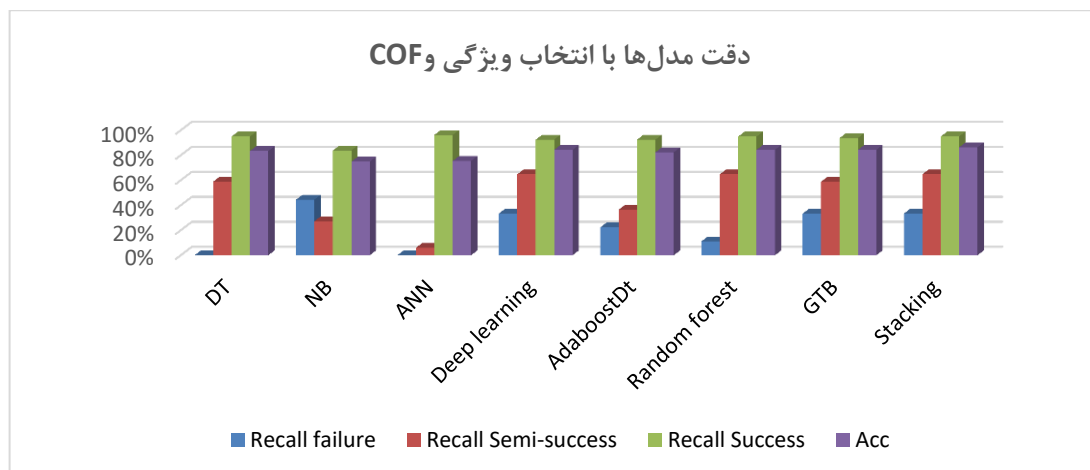
شکل (۴): نتایج دقت مدل‌ها بدون انتخاب ویژگی و حذف داده پرت

- ✓ با توجه به شکل ۴، دقت کلاس شکست در تمام دسته بندها به دلیل نمونه پروژدهای کم شکست (عدم توازن در کلاس) بسیار کمتر از کلاس‌های دیگر است. اما به طور کل در هر سه دقت در همه دسته بندها نتایج نسبتاً متوسطی داریم.
- ✓ روش‌های جنگل تصادفی و نایو بیز دقت بهتری نسبت به سایر دسته بندها دارند.



شکل (۵): نتایج دقت مدل‌ها با انتخاب ویژگی و حذف داده پرت LOF

- ✓ با توجه به شکل ۵، دقت کل مدل‌ها با حذف نمونه‌های پرت، افزایش داشته است. در روش‌های درختی تا میانگین ۷ درصد با حذف نمونه‌های دورافتاده افزایش دقت داشتیم.
- ✓ الگوریتم‌های پایه مانند شبکه عصبی و درخت تصمیم در کلاس failure نتایج ضعیفی دارند. اما روش‌های ترکیبی آدابوست، جنگل تصادفی، گرادیان و استکینگ به دلیل تقسیم داده و متوازن شدن کلاس‌ها دقت بهتری گزارش دادند.
- ✓ در مجموع بالاترین دقت را با جنگل تصادفی در هر سه کلاس دقت گزارش شد.



شکل (۶): نتایج دقت مدل‌ها با انتخاب ویژگی و حذف داده پرت COF

✓ بالاترین دقت کل برای الگوریتم استکینگ با روش حذف داده پرت COF، با دقت ۸۶٫۲۷٪ گزارش شده است. این مدل همچنین در مقایسه با سایر مدل‌ها در کلاسهای موفقیت، شکست و نیمه موفق نیز دقت بالایی گزارش داد در صورتیکه مدل‌های دیگر مانند شبکه عصبی و درخت تصمیم در کلاسهای نیمه موفق و شکست به دلیل عدم توازن نمونه‌ها، دقت بسیار پایین و صفر درصد گزارش شد.

✓ روش حذف داده پرت COF نتایج بسیار بهتری نسبت به LOF گزارش داد و در اکثر مدل‌ها دقت بالاتری داشت که نشان می‌دهد این روش، داده‌های پرت را بهتر حذف کرده است.

✓ مدلی موفق است که در هر چهار دقت نتایج خوبی داشته باشد. از آن جهت که داده‌های ورودی در این پژوهش نامتوازن است (۹۰ درصد پروژه‌ها موفق هستند) مدل بر روی نمونه‌های کلاس اکثریت دقت بالایی دارند از این رو دقت در کلاس شکست پایین است، اما در مقایسه مدل‌ها و با توجه به نمودارها، مدل‌های استکینگ، یادگیری عمیق و گرادیان بوستینگ در مقایسه با سایر مدل‌ها جزو روش‌های برتر این پژوهش است. دقت در کلاس شکست بسیار بهتر از سایر روش‌ها است.

✓ دقت درخت تصمیم پایه با روش‌های ترکیبی آدابوست، جنگل تصادفی و استکینگ افزایش زیادی داشته است، این روش‌ها که از تعدادی درخت استفاده می‌کنند، براساس رأی اکثریت مدل‌ها، کلاس را پیش‌بینی کرده و به‌طورکلی دقت درخت را به‌خصوص روی داده‌های نامتوازن افزایش می‌دهند زیرا بخشی از مجموعه داده را به هر درخت اعمال شده و نسبت کلاس‌ها در حالت ترکیبی متوازن‌تر از حالت یک روش درختی است. هم‌چنین در روش‌های درختی مانند استکینگ، دقت در فاز دوم پیش‌بینی تقویت شده و افزایش می‌یابد.

تحلیل شاخص‌های تأثیرگذار

یکی از مواردی که از ابتدای شروع پژوهش همواره مورد تاکید و درخواست صنعت آب و فاضلاب بوده است، رتبه‌بندی شاخص‌ها بوده تا از طریق بهبود شاخص‌های با اولویت بالاتر بتوانند میزان موفقیت پروژه‌های مشارکتی را به طور موثرتری بهبود بخشند. رتبه‌بندی شاخص‌ها از نظر میزان تأثیرگذاری در موفقیت پروژه‌های مشارکت عمومی- خصوصی، بر مبنای روش شاخص سود اطلاعاتی (آنتروپی) انجام شده که ۱۵ شاخص اول مؤثر در مشارکت عمومی- خصوصی به ترتیب اولویت درجدول ۴ آورده شده است.

جدول (۴): ۱۵ شاخص کلیدی موفقیت پروژه‌های مشارکت عمومی- خصوصی در صنعت آب و فاضلاب

اولویت	نام شاخص
۱	شکایات مربوط به کیفیت خدمات‌رسانی
۲	قالب قراردادی
۳	تعرفه‌های متفاوت (فروش به قیمت)
۴	نسبت کاری (نسبت پوشش هزینه‌ها)
۵	هزینه تمام شده فاضلاب
۶	دوره وصول مطالبات
۷	نسبت گردش در دارایی‌های ثابت
۸	هزینه تمام شده آب در نقطه سر به سر
۹	جمعیت تحت پوشش خدمات آب
۱۰	غیبت
۱۱	سود و زیان عملیاتی (به درصد)
۱۲	کمبود فشار آب (درصد)
۱۳	نسبت سرمایه‌گذاری طی سال به فروش
۱۴	سال اجرای طرح
۱۵	کیفیت آب مصرفی (درصد)

همانگونه که مشاهده می‌گردد، در صنعت آب و فاضلاب ایران، موفقیت پروژه‌های مشارکت عمومی-خصوصی بیشترین تأثیر را از شاخص "شکایات مربوط به کیفیت خدمات‌رسانی" می‌پذیرد. با توجه به این که بیش از ۴۰ درصد از پروژه‌های موجود به صورت فاینانس جاری و با مشارکت مردم در تامین مالی طرح انجام شده است، این مطلب گویای تأثیر رضایت مردم از کیفیت خدمات بر روی موفقیت و استمرار پروژه می‌باشد. بر مبنای اطلاعات حاصل از جدول شماره ۴، در بین ۱۵ شاخص کلیدی موفقیت پروژه‌های مشارکت عمومی- خصوصی در صنعت آب و فاضلاب، ۸ شاخص آن به طور مستقیم با مسائل مربوط به تامین مالی طرح‌ها در ارتباط می‌باشند. این امر بیانگر این است که در صنعت آب و فاضلاب ایران، برخلاف

سایر کشورهای توسعه یافته که حرکت به سمت اجرای مشارکتی طرح‌ها به دلیل افزایش بهره‌وری، کیفی‌سازی طرح‌ها و استفاده از دانش و تجربیات بخش خصوصی است، همچنان دغدغه اصلی تامین مالی طرح‌ها می‌باشد.

بر مبنای نتایج حاصل شده، قالب قراردادی به عنوان دومین شاخص مهم و تأثیرگذار در موفقیت پروژه مشارکتی مطرح است و در نتیجه، شرکت‌های آب و فاضلاب به راحتی می‌توانند قبل از شروع به عملیات اجرایی، نتایج حاصل از اجرای پروژه در هر یک از قالب‌ها را مورد ارزیابی و بررسی قرار داده و مناسب‌ترین قالب قراردادی برای اجرای پروژه را انتخاب نمایند. تعرفه‌های متعارف سومین شاخص تأثیرگذار در موفقیت یک پروژه مشارکتی است. این امر نشان می‌دهد که در صورت حذف تعرفه‌های تکلیفی و از طریق ایجاد فضای رقابتی بر مبنای هزینه‌های واقعی تولید آب و پساب، می‌توان بسیاری از پروژه‌های بخش آب و فاضلاب را به صورت خودگردان و با مشارکت بخش خصوصی تکمیل نموده و این صنعت تا حد زیادی از وابستگی به منابع دولتی دور شود.

تحلیل قوانین

با توجه به مدل‌های درختی گرادیان بوسستینگ، جنگل تصادفی و استکینگ که در این پژوهش بر روی داده‌ها دقت بالایی دارند (نمونه‌ای از خروجی در شکل ۶ آمده است)، در این بخش به بررسی مهم‌ترین قوانین قابل تولید و الگوهای پنهان موجود در داده‌ها پرداخته شده است. البته بیشتر قوانین گزارش شده، آن دسته از قوانینی هستند که احتمال وقوع بیشتر و خطای کمتری داشته و در نتیجه مفیدتر هستند.



شکل (۷): مدل

درخت روش استکینگ

قوانین استخراج شده به صورت زیر می‌باشند:

۱. "اگر تعرفه‌های متعارف بیشتر از ۱۹,۳ درصد باشد، و شکایت کمتر از ۳۸,۵ پروژه با احتمال زیاد موفق می‌شود".
این امر نشان دهنده اهمیت بالای این دو شاخص بوده که در صورتی که شرکت‌های آب و فاضلاب بتوانند در بهبود

آنها قدم بردارند، در کوتاهترین زمان میتوانند به موفقیت پروژه‌ها نزدیک شوند. این قانون شامل دو فاکتور اساسی میباشد که یکی از آنها تعرفه‌های متعارف برای خدمات صنعت میباشد و با توجه به تکلیفی بودن این تعرفه‌ها، نیازمند سیاست‌گذاری درخصوص بهبود آنها میباشیم. بخش دیگر افزایش رضایت مشترکین است. در واقع در صورتی که بخش آب و فاضلاب بتواند از طریق بررسی دقیق سرمایه‌گذاران در زمان مناقصات و مزایدات، رضایت مشترکین را جلب نماید، احتمال موفقیت پروژه مشارکت عمومی - خصوصی تا حد بالایی افزایش می‌یابد.

۲. "اگر سود و زیان عملیاتی بیشتر از ۱۴۵,۹ و قالب قراردادی بیع متقابل باشد، احتمالاً پروژه موفق داریم". استخراج این قانون بیانگر نقش شاخص سود و زیان عملیاتی در موفقیت یا عدم موفقیت پروژه‌های بیع متقابل است. با توجه به استقرار پروژه‌های بیع متقابل در فلات مرکزی ایران، شرکتهای آب و فاضلاب مستقر در این استانها میبایست تلاش بیشتری در زمینه بهبود این شاخص داشته باشند.

۳. "اگر غیبت کمتر از ۷۹,۵ روز باشد، احتمالاً شکست داریم". شاخص غیبت در واقع شاخص شماره ۱۰ در رتبه بندی به دست آمده بوده و تنها شاخصی است که با عوامل انسانی موجود در بخش آب و فاضلاب در ارتباط است. بر مبنای بررسی‌های به عمل آمده، یکی از دلایل استقبال بخش خصوصی از مشارکت در طرح‌های صنعت آب و فاضلاب، حضور و پاسخگویی نیروهای سازمانی این صنعت میباشد. با توجه به ساختار سازمانی ایجاد شده در بخش آب و فاضلاب و اختصاص دفتر جداگانه در زیرمجموعه مدیرعامل جهت تجهیز منابع مالی و توسعه مشارکت بخش غیردولتی، درخواست‌های سرمایه‌گذاران به خوبی در این صنعت مورد توجه قرار گرفته و در نتیجه می‌تواند پروژه‌های با موفقیت بالاتری را به همراه داشته باشد.

۴. "اگر دوره مطالبات کمتر از ۹۰,۵ روز باشد، پروژه احتمالاً نیمه موفق است". از جمله عوامل بسیار موثر در بقا و تداوم یک پروژه مشارکت عمومی - خصوصی، تامین جریان نقدینگی سرمایه‌گذار است. با توجه به وجود سرمایه در گردش و توان مالی سرمایه‌گذاران موجود در صنعت آب و فاضلاب، ضروری است این بخش، برنامه‌ریزی‌های خود را به گونه‌ای انجام دهد که مطالبات معوق بخش خصوصی، بیشتر از ۳ ماه نباشد.

۵. "اگر دوره وصول مطالبات به‌روز بیشتر از ۵۳,۵ باشد، آنگاه با احتمال زیاد شرکت موفق داریم". این قانون در کنار قانون شماره ۴، ضرورت پرداخت معوقات بخش خصوصی در بازه ۲ تا ۳ ماه را نشان میدهد.

نتیجه‌گیری و تحقیقات آینده

در این پژوهش، برای دسته‌بندی پروژه‌های شکست‌خورده و موفق شده و سپس ساختن مدل پیش‌بینی، ۱۷۶ سطر اطلاعات پروژه‌ها از کلیه شرکت‌های آب و فاضلاب سراسر کشور، جمع‌آوری شد. ۱۷۶ پروژه با ۲۷ شاخص (ستون قابلیت فروش محصول در استان تا ستون آموزش) در ۴ قالب قراردادی BOO, BOT, بیع متقابل و فاینانس جاری گزارش شد. جهت درک داده‌ها، خصوصیات آماری و نمودار هیستوگرام و پراکندگی برخی متغیرها رسم و تحلیل شد. به دلیل انحراف معیار بالای برخی شاخص‌ها، در بخش پیش‌پردازش از تکنیک‌های تشخیص داده پرت LOF و COF استفاده کردیم. سپس در بخش دسته‌بندی الگوریتم‌های درخت تصمیم، یادگیری عمیق، شبکه عصبی چندلایه، بیزساده، آدابوست درخت تصمیم، جنگل تصادفی و گرادیان بوستینگ، ابتدا بدون انتخاب ویژگی و حذف داده پرت، سپس با حذف داده پرت و انتخاب متغیرهای تأثیرگذار مدل‌سازی، در نرم افزار رپید ماینر پیاده‌سازی گردید.

با توجه به نتایج به دست آمده، شاخص‌های مشارکت عمومی - خصوصی به ترتیب اولویت استخراج گردید. بخش آب و فاضلاب می‌تواند با بهبود شاخص‌های با اولویت بالاتر، احتمال موفقیت پروژه‌ها را در قالب‌های قراردادی مختلف افزایش دهد. بر مبنای یافته‌های مدل‌سازی، الگوریتم‌های آدابوست، استکینگ و جنگل تصادفی با دقت بالاتر از ۸۶ درصد، روش‌های برتر در پیاده‌سازی مدل بوده‌اند. روش حذف داده پرت COF به دلیل پاکسازی و تشخیص نمونه‌های پرت بیشتر، بهبود بسیار خوبی در دقت مدل‌های دسته‌بندی داشت و لزوم اجرای مرحله پاکسازی و پیش‌پردازش را در داده‌های سازمانی در پروژه‌های داده‌کاوی اثبات کرد. با توجه به این امر، تحلیل قوانین بر مبنای خروجی‌های درخت تصمیم و تحلیل نتایج بر اساس خروجی الگوریتم استکینگ انجام شد. با توجه به مدل پیشنهادی، به راحتی می‌توان با ورود اطلاعات مربوط به هر طرح جدید، میزان موفقیت آن را در هر یک از قالب‌های قراردادی پیش‌بینی نموده و در نتیجه بهترین روش پیاده‌سازی را برای پروژه مورد نظر انتخاب کرد.

بر مبنای نتایج حاصل از پژوهش و به طور کلی، بهترین روش اجرای مشارکتی براساس نوع پروژه‌های موجود در صنعت آب و فاضلاب به صورت جدول شماره ۵ می‌باشد.

جدول (۵): روش بهینه اجرای مشارکتی براساس نوع پروژه‌ها

قراردادهای بخش فاضلاب	احداث تصفيه‌خانه‌ها و ايستگاه‌های پمپاژ	BOT/ بيع متقابل
	احداث شبکه‌های فاضلاب	بيع متقابل
قراردادهای بخش آب	احداث تصفيه‌خانه‌ها و خطوط انتقال	BOT
	احداث تأسيسات آب‌شیرین‌کن و احیای کیفی	BOO

همان‌گونه که مشاهده می‌شود، قالب قراردادی بيع متقابل در پروژه‌های فاضلاب می‌تواند بسیار موفق باشد، لیکن با توجه به تحلیل انجام شده درخصوص استان‌های مختلف و با توجه به این که کالای پساب هنوز ارزش واقعی خود را در تمامی نقاط کشور به دست نیاورده است، در حال حاضر بخش آب و فاضلاب این قالب قراردادی را در مناطق مرکزی که صنعت و مشتریان بالقوه وجود داشته و مشکلات کمبود آب نیز بالاتر از سایر مناطق است، مورد استفاده قرار میدهد. در سایر نقاط، روش بهینه مشارکت جهت اجرای طرح‌های فاضلاب روش BOT است که به دلیل محدودیت منابع مالی دولت چندان مورد استقبال صنعت آب و فاضلاب نبوده و تمایل این بخش اجرای پروژه‌های تصفیه‌خانه به صورت BOT خودگردان است. در بخش احداث شبکه‌های فاضلاب، با توجه به مستندات قانونی جدید و امکان افزایش میزان حق انشعاب دریافتی از مردم (با موافقت شورای شهر)، می‌توان این پروژه‌ها از طریق درآمدهای حاصل از این حوزه و در قالب قراردادی فاینانس جاری انجام داد.

1. AKINTOYE, A., HARDCASTLE, C., BECK, M., CHINYIO, E. & ASENOVA, D. 2003. Achieving best value in private finance initiative project procurement. *Construction management and economics*, 21, 461-470.
2. AL-RADAIDEH, Q. A. & AL NAGI, E. 2012. Using data mining techniques to build a classification model for predicting employees performance. *International Journal of Advanced Computer Science and Applications*, 3.
3. ALEGRE, H. & ASSOCIATION, I. W. 2006. *Performance indicators for water supply services*, IWA publishing.
4. ALI, M. & QAMAR, A. M. Data analysis, quality indexing and prediction of water quality for the management of rawal watershed in Pakistan. Eighth International Conference on Digital Information Management (ICDIM 2013), 2013. IEEE, 108-113.
5. AZHAR, S. A. S., JOHAR, H., BAKI, S. R. M. S. & TAHIR, N. M. Optimization of water quality monitoring based on fuzzy algorithms. 2013 IEEE Conference on Systems, Process & Control (ICSPC), 2013. IEEE, 283-288.
6. BEIKZADEH, M. R., PHON-AMNUAISUK, S. & DELAVARI, N. 2008. Data mining application in higher learning institutions. *Informatics in Education-An International Journal*, 31-54.
7. BÖHL, C. G. P. 2007. *Development of a knowledge based decision support system for private sector participation in water and sanitation utilities*, Oldenbourg Industrieverlag.
8. CHEN, W.-S. & DU, Y.-K. 2009. Using neural networks and data mining techniques for the financial distress prediction model. *Expert Systems with Applications*, 36, 4075-4086.
9. CHENG, H., LU, Y.-C. & SHEU, C. 2009. An ontology-based business intelligence application in a financial knowledge management system. *Expert Systems with Applications*, 36, 3614-3622.
10. COMPANY, N. W. A. W. 2009. National Wastewater Financing and investment strategy. Ministry of Energy.
11. CUI, C., LIU, Y., HOPE, A. & WANG, J. 2018. Review of studies on the public-private partnerships (PPP) for infrastructure projects. *International Journal of Project Management*, 36, 773-794.
12. DELMON, J. 2015. *Creating a Framework for Public-Private Partnership Programs: A Practical Guide for Decision-makers*, World Bank.
13. FREUND, Y. & SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. European conference on computational learning theory, 1995. Springer, 23-37.
14. GOLABCHI, M. N. 2015. Selecting the best PPP method in rail projects by using AHP methods. *Journal of Transportation Engineering*, 6.
15. HALL, D., LOBINA, E. & MOTTE, R. D. L. 2005. Public resistance to privatisation in water and energy. *Development in practice*, 15, 286-301.
16. HODGE, G. A., GREVE, C. & BOARDMAN, A. 2010. *International handbook on public-private partnerships*, Edward Elgar Publishing.
17. JENABI, G. & MIRROSHANDEL, S. A. 2014. Using data mining techniques for improving customer relationship management. *European Online Journal of Natural and Social Sciences: Proceedings*, 2, pp. 3143-3149.
18. KIRIMI, J. M. & MOTURI, C. A. 2016. Application of Data Mining Classification in Employee Performance Prediction. *International Journal of Computer Applications*, 146, 28-35.
19. LAVRAČ, N., BOHANEČ, M., PUR, A., CESTNIK, B., DEBELJAK, M. & KOBLEK, A. 2007. Data mining and visualization for decision support and modeling of public health-care resources. *Journal of biomedical informatics*, 40, 438-447.
20. LI, B., AKINTOYE, A., EDWARDS, P. J. & HARDCASTLE, C. 2005. Critical success factors for PPP/PFI projects in the UK construction industry. *Construction management and economics*, 23, 459-471.
21. LI, X., ZHU, Z. & PAN, X. 2010. Knowledge cultivating for intelligent decision making in small & middle businesses. *Procedia Computer Science*, 1, 2479-2488.
22. MARZOUK, M. & FAYEZ, E. 2018. Public Private Partnership Projects: Concessionaire Performance Measurement. *Journal of Al Azhar University Engineering Sector*, 13, 466-480.
23. MCQUAID, R. W. & SCHERRER, W. 2010. Changing reasons for public-private partnerships (PPPs). *Public Money & Management*, 30, 27-34.
24. MUTULA, S. M. & VAN BRAKEL, P. 2006. E-readiness of SMEs in the ICT sector in Botswana with respect to information access. *The electronic library*, 24, 402-417.
25. OGWUELEKA, T. C. & OGWUELEKA, F. N. 2010. Data mining application in predicting *Cryptosporidium* spp. oocysts and *Giardia* spp. cysts concentrations in rivers. *Journal of Engineering Science and Technology*, 5, 342-349.
26. QIAO, L., WANG, S. Q., TIONG, R. L. & CHAN, T.-S. 2001. Framework for critical success factors of BOT projects in China. *The Journal of Structured Finance*, 7, 53-61.
27. QIU, Y., LI, J., HUANG, X. & SHI, H. 2018. A Feasible Data-Driven Mining System to Optimize Wastewater Treatment Process Design and Operation. *Water*, 10, 1342.
28. RIBEIRO, D., SANFINS, A. & BELO, O. Wastewater treatment plant performance prediction with support vector machines. Industrial Conference on Data Mining, 2013. Springer, 99-111.
29. ROKACH, L. 2010. Ensemble-based classifiers. *Artificial Intelligence Review*, 33, 1-39.

30. SACHS, T., TIONG, R. & QING WANG, S. 2007. Analysis of political risks and opportunities in public private partnerships (PPP) in China and selected Asian countries: Survey results. *Chinese Management Studies*, 1, 126-148.
31. SALMAN, A. F., SKIBNIEWSKI, M. J. & BASHA, I. 2007. BOT viability model for large-scale infrastructure projects. *Journal of Construction Engineering and Management*, 133, 50-63.
32. SUN, J., WANG, R., WANG, X., YANG, H. & PING, J. 2014. Spatial cluster analysis of bursting pipes in water supply networks. *Procedia Engineering*, 70, 1610-1618.
33. WEN, Y. Y., HUANG, W. M., WU, J., CHEN, Y. & SONG, J. Q. Water consumption analysis system based on data mining. *Applied Mechanics and Materials*, 2013. Trans Tech Publ, 1093-1097.
34. YU, I., KIM, K., JUNG, Y. & CHIN, S. 2007. Comparable performance measurement system for construction companies. *Journal of Management in Engineering*, 23, 131-139.
35. YUAN, J., SKIBNIEWSKI, M. J., LI, Q. & ZHENG, L. 2009. Performance objectives selection model in public-private partnership projects based on the perspective of stakeholders. *Journal of Management in Engineering*, 26, 89-104.
36. YUAN, J., WANG, C., SKIBNIEWSKI, M. J. & LI, Q. 2011. Developing key performance indicators for public-private partnership projects: questionnaire survey and analysis. *Journal of Management in Engineering*, 28, 252-264.
37. ZHANG, X. 2005. Critical success factors for public-private partnerships in infrastructure development. *Journal of construction engineering and management*, 131, 3-14.

An Intelligent Hybrid Model for Determining Public-Private Partnership Based on Collective Tree Algorithms in Iranian Water and Wastewater Industry

Malihe Eskandari^{24*}, Mohammad Taghi Taghavifard²⁵, Iman Raeisi Vanani²⁶, Seyed Soroush Ghazi Noori²⁷

Abstract

One of the pillars of any country's development is the access to safe water and sanitation and so, these projects should be implemented in the shortest possible time. In this regard, considering the existence and emergence of various methods of private sector participation, choosing the right approach has become one of the most important issues in this industry. The decision makers always need a model for selection of the best public-private partnership method in specific region or project. Using the database of partnership projects, existing information, and pattern and data mining algorithms in the water and wastewater sector, we have designed a public-private partnership (PPP) model to predict and propose a proper way to execute such projects. In this research, CRISP data mining method was applied to the data from 176 projects. After describing and understanding the data, the purging and deletion steps were performed. In the process of classification with tree techniques and machine learning, the classification and success of the projects were applied to the data and necessary analysis were performed. Based on the results, the indicators of public-private participation were extracted and prioritized. Based on the research findings, the combined COF[\] method for deleting outliers, Entropy indices for feature selection, and Stacking methods are applied to predict project success or failure with an accuracy of 86.27%. According to the proposed model, one can easily predict the success rate of each of the contractual templates in addition to introducing the appropriate contractual template for any water and wastewater project by entering information on each new project and examine the impact of the improvement of each of the indicators.

Key words: Data Mining, Forecasting, Outsourcing, Combined Methods, Information Gain Index, Gradient Boosting, Random Forest

[\] Connectivity-based Outlier Factor

Connectivity-based Outlier Factor	.۱
World Bank	.۲
World Health Organization (WHO)	.۳
UNICEF	.۴
Key Performance Indicators	.۵
Economic Viability	.۶
Likert-style rating questions	.۷
Financial Knowledge Management System	.۸
Support Vector Machines	.۹
Adaboost	.۱۰
Stacking	.۱۱
Random Forest	.۱۲
Gradient Boosting	.۱۳
http://invest.nww.ir	.۱۴
Local Output Factor	.۱۵
True Positive-TP	.۱۶
False Positive _FP	.۱۷
True Negative-TN	.۱۸
False Negative _FN	.۱۹