

واکاوی و کشف الگوی تصمیم‌گیری در بودجه عمومی جمهوری اسلامی ایران با رویکرد داده‌کاوی

سیدمحمدصادق داودی^۱

غلامرضا گودرزی^۲

عباس طلوعی اشلقی^۳

چکیده:

امروزه یکی از مهم‌ترین گام‌های مؤثر برای داشتن یک موفقیت بزرگ در سازمان‌ها بحث تصمیم‌گیری به‌موقع و مبتنی بر داده است. تحلیل یکی از اساسی‌ترین بخش‌های مهم اقتصاد یعنی بودجه می‌تواند مدیران را در تصمیم‌گیری هرچه کارآمدتر و مؤثرتر کمک نماید. با توجه به حجم عظیم دستگاه‌های دولتی و حجم داده‌های مرتبط با آن‌ها یکی از اصلی‌ترین روش‌هایی که می‌تواند به ما در تحلیل این دسته از داده‌ها کمک نماید داده‌کاوی است. داده‌کاوی با گردآوری مجموعه‌ای از علوم مختلف و باهدف کشف دانش نهفته در دل داده‌ها به دنبال آن است تا الگوها و نظم‌های پنهان موجود در دل داده‌ها را شناسایی کند تا بتواند به مدیران در تصمیم‌گیری دقیق و مبتنی بر داده کمک نماید. پژوهش حاضر به دنبال کشف و تحلیل الگوی کارآمد در بودجه دستگاه‌های دولتی در سال‌های ۱۳۹۶-۱۳۸۱ بر اساس تکنیک‌های داده‌کاوی از جمله خوشه‌بندی است. یافته‌های پژوهش نشان می‌دهد که اولاً بررسی حالات مختلف داده در نوع خوشه‌بندی آن‌ها اثر چندانی ندارد و ثانیاً آن‌که دستگاه‌های بزرگ و مهم کشور دارای یک الگوی خاص در دریافت بودجه می‌باشند.

کلمات کلیدی: بودجه، بودجه عمومی، داده‌کاوی، تصمیم‌گیری، خوشه‌بندی

۱. کارشناسی ارشد، دانشکده معارف اسلامی و مدیریت، دانشگاه امام صادق علیه‌السلام، تهران، ایران. نویسنده مسئول. پست الکترونیک: davoudi@isu.ac.ir

۲. استاد، گروه مدیریت صنعتی، دانشکده معارف اسلامی و مدیریت، دانشگاه امام صادق علیه‌السلام، تهران، ایران

۳. استاد، گروه مدیریت صنعتی، دانشکده مدیریت و اقتصاد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران، ایران

مقدمه و بیان مساله

بودجه شاه‌رگ حیاتی دولت است و در واقع یک آئینه تمام‌نما از همه‌ی برنامه‌ها و فعالیت‌های دولت بوده و نقش بسیار مهم و حیاتی در توسعه اقتصاد ملی ایفا می‌نماید. با توجه به سهم بالای دولت در اقتصاد، اهمیت و نقش بودجه در هدایت و جهت‌دهی به بخش‌های مختلف اقتصاد آشکار می‌شود به طوری که به جرئت می‌توان گفت هر نوع پیشرفت و توسعه در بخش‌های گوناگون اقتصاد جامعه، مدیون برنامه‌ریزی عملیاتی دولت در قالب بودجه بوده و برعکس، هر نوع عقب‌ماندگی و عدم توسعه در بخش‌های مختلف اقتصادی را می‌توان به نواقص و کمبودهای نظام بودجه‌بندی نسبت داد.

نهادهای بودجه‌ریزی در طی تاریخ، در روند تدریجی حرکت کشورها به سوی حکمرانی شایسته و پاسخ‌گو، نقش قابل توجهی ایفا کردند. بودجه‌ریزی ابزار راهبردی برای انضباط اقتصادی و مالی دولت‌هاست دولت‌هاست (آذر و امیرخانی، ۱۳۹۱، ص ۱). به طوری که از این ابزار به عنوان مهم‌ترین و مؤثرترین عامل برنامه‌ریزی و کنترل یاد شده است (آذر و خدیور، ۱۳۹۶، ص ۳۳). در واقع بودجه مهم‌ترین ابزاری است که از طریق آن مأموریت‌های دولت صورت کمی به خود گرفته و طبق آن اولویت‌های زمان‌بندی و به‌طور کلی تعبیرهای عملیاتی مأموریت‌های دولت تعیین می‌شود (آذر و خدیور، ۱۳۹۶، ص ۳۳). بنابراین بودجه‌ریزی عمومی، نوعی ابزار حکمرانی جدید است. این ابزار می‌تواند کارکنان بخش عمومی را پاسخگو نگه دارد.

بودجه کل کشور به دودسته کلی تقسیم‌بندی می‌شود: (۱) بودجه عمومی و (۲) بودجه شرکت‌های دولتی، بانک‌ها و مؤسسات انتفاعی وابسته به دولت. بودجه عمومی شامل اجزای زیر است: الف) پیش‌بینی دریافت‌ها و منابع تأمین اعتبار که به‌طور مستقیم و غیرمستقیم در سال مالی قانون بودجه به‌وسیله دستگاه‌ها از طریق حساب‌های خزانه‌داری کل اخذ می‌گردد. ب) پیش‌بینی پرداخت‌هایی که از محل درآمد عمومی و یا اختصاصی برای اعتبارات جاری و عمرانی و اختصاصی دستگاه‌های اجرائی می‌تواند در سال مالی مربوط انجام دهد. (مقیمی، ۱۳۹۶، ص ۶۵). برآورد بودجه در دو مرحله اول صورت می‌گیرد و چون این برآوردها با تأخیر زمانی در مرحله سوم اجرا می‌شود؛ لذا مطابق با مقادیر در هنگام اجرا نیست و بودجه تخصیصی با عملکرد بودجه در مقایسه با بودجه پیشنهادی دولت و بودجه مصوب دچار تغییرات می‌شود (اربابیان و همکاران، ۱۳۹۱، ص ۳۳). با توجه به این تغییرات و حجم عظیم داده‌هایی که در بودجه عمومی وجود دارد، یکی از مهم‌ترین و اساسی‌ترین کارهایی که می‌بایست صورت گیرد تحلیل این داده‌ها و اطلاعات است. در این پژوهش به بررسی و تحلیل داده‌های بودجه عمومی به تفکیک دستگاه و برنامه پرداخته می‌شود. با توجه به این که تصمیم‌گیری داده‌محور^۱ به‌عنوان یک رکن اساسی در راستای بالابردن کارایی و اثربخشی تصمیمات است، تصمیمات بر خواسته از دل داده‌ها می‌تواند تصویر دقیقی از عملکرد فعالیت‌های سازمان را ارائه نماید. اهمیت این تصمیم‌گیری تاجایی است که

برخی از دانشمندان علم مدیریت نظیر هربرت سایمون^۱ اعتقاد دارند تصمیم‌گیری مترادف با کل فرآیند مدیریت است. یکی از علومی که به تحلیل داده‌های عظیم^۲ می‌پردازد، علم داده‌کاوی است. داده‌کاوی با نشئت گرفته از علوم مختلف از جمله علم آمار، هوش مصنوعی، یادگیری ماشین، شناسایی الگو و پایگاه داده است که می‌تواند حجم عظیمی از داده‌ها را مورد پردازش قرار دهد و کلیه نظم‌هایی را که در عمق داده وجود دارند، به صورت دانشی با ارزش کشف کرده و جهت استفاده و بهبود تصمیم‌گیری عرضه نماید.

پرسش‌های اصلی

- پرسش‌های اصلی این پژوهش عبارتند از:
- الگوهای داده در بودجه عمومی چگونه است؟
 - کشف دانش در بودجه عمومی چگونه به تصمیم‌گیری‌های مدیران کمک خواهد کرد؟

۱. ادبیات موضوع

۱.۱. تعریف لغوی بودجه

ریشه‌ی واژه‌ی «باجت»^۳ برگرفته از کلمه‌ی فرانسوی «بوژت»^۴ به معنی کیف کوچک است. در انگلستان این واژه برای توصیف «کیف چرمی»^۵ که وزیر دارایی هنگام عزیمت به مجلس با خود حمل می‌کرد، مورد استفاده قرار می‌گرفت که در این کیف چرمی، اسناد مربوط به صورت وضعیت دخل و خرج دولت قرار داشت [داشت (مالگوی و ونگبو، ۲۰۱۲)].^۶ زمانی که سند موجود درون کیف را حاوی فهرست دخل و خرج دولت بود بیرون می‌آوردند تا در مجلس قرائت کنند، نمایندگان می‌گفتند دولت بودجه‌اش را باز کرد (عباسی، ۱۳۸۳، ص ۳۳). بعدها، واژه‌ی بودجه به اسناد برنامه‌های دولت اطلاق گردید که در قالب ارقام پولی، تنظیم و در داخل کیفی قرار داده می‌شد و برای تصویب به مجلس تقدیم می‌گردید (مالگوی و ونگبو، ۲۰۱۲).

۲.۱. تعریف اصطلاحی بودجه

تعاریف مختلفی از حیث اصطلاح بودجه بیان شده است که می‌توان آن‌ها را در دو تعریف خلاصه کرد:

اول: بودجه دخل و خرج دولت است. دوم: بودجه یک سند مالی است، مربوط به درآمدها و هزینه‌های دولت (طالب نیا و طالبی، ۱۳۹۷).

1. Herbert Simon
2. Big Data
3. budget
4. bougette
5. Leather bag
6. Malgwi & Unegbu, 2012

۳,۱. تقسیم‌بندی بودجه کل کشور

طبق ماده ۱ قانون محاسبات عمومی، بودجه دولتی ایران از سه قسمت تشکیل می‌شود که عبارتند از: بوجه عمومی، بودجه شرکت‌های دولتی و بانک‌ها، بودجه سایر مؤسسات (مقیمی، ۱۳۹۶، ص ۲۸).

۱,۳,۱. بودجه عمومی

بودجه عمومی دولت که شامل اجزاء زیر است:
 الف) پیش‌بینی دریافت‌ها و منابع تأمین اعتبار که به‌طور مستقیم و یا غیرمستقیم در سال مالی قانون بودجه به‌وسیله دستگاه‌ها از طریق حساب‌های خزانه‌داری کل اخذ می‌گردد.
 ب) پیش‌پرداخت‌های که از محل درآمد عمومی و یا اختصاصی برای اعتبارات جاری و عمرانی و اختصاصی دستگاه‌های اجرائی می‌تواند در سال مالی مربوط انجام دهد.

۲,۳,۱. بودجه شرکت‌های دولتی و بانک‌ها

بودجه شرکت‌های دولتی و بانک‌ها شامل پیش‌بینی درآمدها و سایر منابع تأمین اعتبار است. بودجه شرکت‌های دولتی، بانک‌ها و مؤسسات انتفاعی وابسته به دولت در سند بودجه کل کشور، در کنار بودجه عمومی دولت منعکس می‌شود (مقیمی، ۱۳۹۶، ص ۲۹).

۴,۱. چرخه بودجه‌ریزی دولتی

نحوه‌ی بودجه‌ریزی در کشورهای مختلف، متنوع است و به چارچوب حقوقی و قانون کشورها و همچنین ساختار سازمانی بستگی دارد. فرآیند بودجه‌ریزی تعاملات در میان مشارکت‌کنندگان متعدد، از شهروندان گرفته تا بوجه معمولاً برای یک سال مالی تنظیم می‌شود، ولی فرآیند بودجه‌ریزی در یک دوره طولانی‌تر از یک سال امتداد پیدا می‌کند و چرخه‌ای است که دارای مراحل است که بعضاً با یکدیگر هم‌پوشانی نیز دارند.
 مراحل کلیدی سیستم بودجه‌ریزی دولتی عبارتند از: «تنظیم و تقدیم پیشنهاد»^۱، «مباحثه و تصویب»^۲، «اجرا»^۳ و «گزارشگری، حسابرسی و ارزیابی»^۴ (مقیمی، ۱۳۹۶، ص ۱۴).

۵,۱. تصمیم‌گیری

یکی از مسائل مهم مورد بحث مدیریت موضوع تصمیم‌گیری است. در اداره امور یک سازمان اعم از یک وزارتخانه یا یک شرکت تعاونی یا یک مؤسسه تولیدی تصمیم‌گیری از فعالیت‌های اساسی در همه سطح‌های مختلف سلسله‌مراتب سازمانی به شمار می‌آید. تصمیم‌گیری نه‌تنها

1. formulation and submission
2. debate and adoption
3. execution
4. reporting, auditing and evaluating

در یک سازمان یا یک شرکت بلکه در زندگی روزانه هرکس دیده می‌شود. تصمیم‌گیری جز لاینفک وظائف مدیریت هست. در حقیقت بنا به ضرورت حرفه‌ای، مدیران تصمیم‌گیر هستند ولی بسیاری از آنان از تحلیل اینکه چگونه به تصمیم می‌رسند ناتوانند.

یکی از دانشمندان رشته مدیریت به نام پیتر دراکر^۱ می‌گوید که تأکید مدیریت آینده بر درک فرایند تصمیم‌گیری است. با توجه به این واقعیت که مدیریت را می‌توان مترادف با تصمیم‌گیری دانست و عمل تصمیم‌گیری رکن اصلی و اساسی مدیریت به شمار می‌آید اهمیت موضوع و ضرورت شناخت فرایند اتخاذ تصمیم به خوبی نمایان می‌گردد. به دلیل همین اهمیت موضوع است که پاره‌ای مدیران تصمیم‌گیری را کار اصلی خود می‌دانند (معتدوزیری، ۱۳۷۲، ص ۳۲). اطلاعات پایه و اساس تصمیم‌گیری است و این موضوع یعنی فرایند جریان اطلاعات به‌عنوان جنبه‌ای از سازماندهی و مدیریت در سال‌های اخیر مورد توجه بسیار قرار گرفته است. به گفته یکی از دانشمندان رشته مدیریت برای اتخاذ تصمیم خوب نود درصد (۹۰٪) اطلاعات و ده درصد (۱۰٪) الهام و بینش لازم است. ستاینر^۲ یکی دیگر از علمای مدیریت می‌گوید: «جریان اطلاعات همانند جریان خون برای حیات و سلامت هر واحد لازم است» (معتدوزیری، ۱۳۷۲، ص ۳۲).

از آنجاکه تصمیم‌گیری از حساس‌ترین وظایف مدیران سازمان‌ها در عرصه‌های مختلف عملیات به شمار می‌رود، شناخت فرایند، ساختار، سازوکارها و ویژگی‌های تصمیم‌ها برای هدایت و برقراری ارتباط در سازمان اهمیت فراوان دارد. از نظر علم مدیریت وجود دوگانگی بین هدف‌ها و عملکردها و ناتوانی مدیران در دستیابی به هدف‌های طراحی‌شده، تنها یک علت کلیدی دارد و آن ضعف نظام تصمیم‌گیری است (حمیدی زاده، ۱۳۸۰، ص ۴۴).

۱،۵،۱. گام‌های تصمیم

پنج گام در تحلیل تصمیم‌گیری عمومی وجود دارد که به ترتیب عبارت‌اند از: شناسایی و تعریف مسئله، جستجو برای یافتن راه‌حل‌های احتمالی، بررسی نمودن عواقب ناشی از هر راه‌حل، انتخاب یکی از مدل‌های تصمیم‌گیری، بکار بردن یکی از مدل‌های تصمیم‌گیری و اتخاذ تصمیم (عزیزی، ۱۳۹۶، ص ۲۹). برخی دیگر از دانشمندان مرحله ششمی را نیز به این گام‌ها اضافه نموده‌اند که همانا پی‌گیری باز خورهای ناشی از نتایج تصمیم‌هاست (دویس، ۱۹۵۸)^۳

۱،۶. داده‌کاوی و آشنایی با فرایند کشف دانش

در عصر اطلاعات، حجم بسیار زیادی از داده‌ها در صنایع مختلف موجود است. در دسترس بودن این داده عظیم هیچ فایده‌ای ندارد مگر اینکه به اطلاعات ارزشمندی تبدیل شود. داده‌کاوی مجموعه‌ای از تکنیک‌های برای کشف خودکار کارآمد، معتبر، جدید، مفید و قابل فهم در پایگاه

1. Peter Drucker
2. Steiner
3. Davis, 1958

داده‌های بزرگ است. این الگوها باید قابل اجرا باشند تا آن‌ها بتوانند در تصمیم‌گیری شرکت به کار روند. داده‌کاوی فرایندی از کشف خودکار الگوهای ناشناخته در حجم زیاد داده‌ها است. این حجم بزرگ از داده‌ها معمولاً داده‌های تاریخی یک سازمان شناخته شده به عنوان انبار داده^۱ هست (بهاتیا، ۲۰۱۹، ص ۱۴). داده‌کاوی بخش مهمی از کشف دانش در پایگاه‌های داده^۲ (KDD) است، که فرآیند کلی تبدیل داده‌های خام به اطلاعات مفید است (تن و همکاران، ۲۰۰۷، ص ۳).^۳ داده‌کاوی در حقیقت کشف ساختارهای جالب توجه، غیرمنتظره و بارز از روی مجموعه گستره‌ای از داده‌ها است و فعالیتی است که اساساً با آمار و تحلیل موشکافانه داده منطبق است (دیویدهند، ۱۳۸۱، ص ۸۷).

۲. پیشینه تحقیق

آذر و همکاران (۱۳۹۰) در پژوهشی تحت عنوان "مدل‌سازی ریاضی استوار، رویکردی نوین در بودجه‌ریزی عمومی ایران" با به‌کارگیری روش برنامه‌ریزی خطی مدلی ریاضی با رویکرد بهینه‌سازی استوار به بررسی بودجه‌ریزی در بخش عمومی ایران پرداخته‌اند. نتایج این تحقیق حاکی از آن است که مدل ریاضی طراحی شده از قابلیت تحلیل سناریوهای مختلف در طی سال بودجه‌ریزی برخوردار است و امکان عکس‌العمل مناسب نسبت به تغییرات محیطی را برای کاربران به وجود خواهد آورد.

همچنین در همین راستا آذر و نجفی (۱۳۹۰) به بررسی "مدل ریاضی بودجه‌ریزی در بخش عمومی: رویکرد بهینه‌سازی استوار" پرداخته‌اند. در این مقاله، رویکرد بهینه‌سازی استوار انتخاب شده است. مدل بودجه‌ریزی در بخش عمومی با مدل‌های استوار سویستر^۴، بن-تال و نمیروفسکی^۵، و برتسیماس و سیم^۶ توسعه داده شده است تا بر اساس میزان محافظه‌کاری تصمیم‌گیران - در زمینه وضعیت حال و آینده کشور - بتوان از تک‌تک مدل‌ها استفاده کرد.

لی و همکاران^۷ (۲۰۱۳) در کتاب خود تحت عنوان «دستگاه‌های بوجه‌ریزی دولتی» به بررسی ابعاد مختلف این مسئله پرداخته است. از جمله چرخه بودجه، آماده‌سازی بودجه، تصویب بودجه و مدیریت سرمایه.

کفورد^۸ (۲۰۱۰) در پژوهشی با موضوع «گزینه‌های بودجه عمومی و تمایلات خصوص برای پرداخت» بیان می‌کند که افراد مایل به پرداخت بیش‌ترین سهم برای توسعه خدمات آموزشی و پس‌از آن مراقبت‌های بهداشتی هستند.

1. data warehouse
2. knowledge discovery from data(KDD)
3. Tan, Steinbach, & Kumar, 2007
4. Soyester
5. Ben-Tal And Nemirovski
6. Bertsimas And Sim
7. Lee, Johnson, & Joyce, 2013
8. Koford, 2010

اسکات و همکاران^۱ (۲۰۱۳) به تجزیه و تحلیل مشتری و کسب‌وکار: استخراج داده‌های کاربردی برای تصمیم‌گیری تجاری با استفاده از زبان برنامه‌نویسی R پرداخته است. زارتزکی^۲ (۲۰۱۱) در مقاله خود با عنوان «بررسی هوشمندی کسب‌وکار: داده‌کاوی و بهینه‌سازی برای تصمیم‌گیری» به اهمیت داده‌کاوی در تصمیم‌گیری پرداخته است.

تقوی فرد و همکاران (۱۳۸۶) بیان می‌کنند که ابهامات محیطی ناشی از مدل‌های نوین کسب‌وکار سبب تشدید پیچیدگی در تصمیم‌گیری، به‌ویژه در حوزه کسب‌وکار شده است. به‌طوری‌که بسیاری از متغیرهای تأثیرگذار ناشناخته بوده و روابط میان آن‌ها نیز غیرخطی و پیچیده است. در این پژوهش یک مطالعه موردی بر روی پایگاه داده مشتریان یکی از بانک‌ها با استفاده از داده‌کاوی ارائه گردیده است.

دستگیر و شفیعی سردشت (۱۳۹۰) در پژوهشی تحت عنوان «فناوری داده‌کاوی؛ رویکردی نوین در حوزه مالی» به بررسی اجمالی این فناوری جدید و تحقیقات مرتبط با آن، به‌منظور نمایان ساختن کارکردهای داده‌کاوی در امور مالی و حسابداری و فرصت‌های تحقیقاتی آن پرداخته است.

کیقبادی و خدابی (۱۳۹۲) پژوهشی باهدف مدل‌سازی اعتبارسنجی مشتریان در بانک به روش‌های شبکه عصبی، درخت تصمیم و ماشین بردار پشتیبان انجام داده‌اند. در این تحقیق پس از بررسی پرونده‌های اعتباری هریک از مشتریان، ۳۱ متغیر توضیح‌دهنده مورد ارزیابی قرار گرفته و نتایج ضمن دلالت بر تأیید نظریه‌های اقتصادی و مالی نشان می‌دهد که تکنیک‌های داده‌کاوی جهت اعتبارسنجی مشتریان از کارایی بالایی برخوردار هست و همچنین عملکرد پیش‌بینی الگوی شبکه عصبی به‌مراتب بهتر از سایر الگوها است.

هاشمی و همکاران (۱۳۹۱) پژوهشی باهدف مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای ارزیابی عملکرد مالی واحدهای تجاری با استفاده از مدل‌های داده‌کاوی انجام داده‌اند. یافته‌های پژوهش نشان می‌دهد نسبت‌های نقدی در پیش‌بینی عملکرد مالی شرکت‌های قوی و نسبت‌های تعهدی در پیش‌بینی عملکرد مالی شرکت‌های ضعیف از محتوای اطلاعاتی بالاتری در هر دو مدل مورد استفاده برخوردارند.

مسلم‌زاده گتایی (۱۳۹۴) در پژوهش خود به بررسی روش‌های داده‌کاوی برای کشف تقلب در حسابداری صورت‌های مالی پرداخته است.

قوسی و همکاران (۱۳۹۷) در پژوهش خود تحت عنوان "طراحی سامانه تصمیم‌یار با استفاده از ابزار داده‌کاوی (مطالعه موردی معاونت فرهنگی دانشگاه علم و صنعت)"، ابتدا ادبیات مربوط به سیستم پشتیبان تصمیم بررسی کرده است؛ سپس داده‌کاوی به‌عنوان ابزاری برای استخراج دانش و اطلاعات کاربردی از داده‌های خام سازمانی معرفی شده است. این دانش به‌دست‌آمده حاوی مفاهیم و اطلاعاتی است که ممکن است پیش‌ازاین در سازمان مغفول مانده باشد لذا کمک

1. Scott, 2013

2. Zaretski, 2011

شایانی به روند تصمیم‌گیری مدیران خواهد داشت. در نهایت یافته‌های این مطالعه در دانشگاه علم و صنعت به کار گرفته شده است تا راهنما و پشتیبان تصمیمات مدیران دانشگاهی باشد.

آنتی^۱ (۲۰۱۰) در کتاب خود تحت عنوان «داده‌کاوی در بخش‌های دولتی و خصوصی: کاربردهای سازمانی و دولتی» به موضوعاتی از قبیل مطالعات داده‌کاوی در مدیریت و دولت، داده‌کاوی به‌عنوان حریم خصوصی، امنیت و حفظ اطلاعات و دانش، داده‌کاوی در موقعیت‌های سازمانی برای آماده‌سازی و پیش‌بینی داده‌کاوی به‌عنوان برنامه‌ها و رویکردهای مربوط به صحنه‌سازمانی پرداخته است.

تاکنون پژوهشی در ایران در ارتباط با بررسی داده‌های بودجه عمومی دستگاه‌های دولتی از طریق الگوریتم‌های داده‌کاوی انجام نشده است و پژوهش حاضر از این منظر دارای نوآوری است.

۳. روش‌شناسی پژوهش

پژوهش حاضر یک پژوهش اکتشافی از نوع کاربردی است. داده‌های پژوهش حاضر مربوط به بودجه دستگاه‌های دولتی از سال ۱۳۸۱ تا ۱۳۹۶ در سه متغیر اصلی یعنی اعتبار (مقدار بودجه نهایی مصوب شده توسط مجلس برای هر دستگاه)، تخصیص (سازمان برنامه‌بودجه به چه میزان به هر دستگاه تخصیص داده است) و دریافتی (میزان دریافتی هر دستگاه از خزانه دولت) است. دو سناریو اصلی برای بررسی داده‌ها در نظر گرفته شد. سناریو اول بررسی داده‌های مربوط به دستگاه‌ها به تفکیک سال و سناریو دوم بررسی داده‌های مربوط به دستگاه‌ها بر اساس مجموع سال‌های ۱۳۸۱ تا ۱۳۹۶. در هر سناریو ۵ حالت مختلف بررسی داده‌ها صورت گرفته است که عبارتند از: بررسی کلیه دستگاه‌ها، دستگاه‌های دارای حداقل ده سال سابقه بودجه، دستگاه‌های دارای دقیقاً ۱۶ سال سابقه بودجه، دستگاه‌های دارای نسبت ۹۰ درصد تخصیص نسبت به اعتبار و هم‌چنین دستگاه‌های موجود در بازه سه انحراف معیار از میانگین می‌شود.

داده‌ها بر اساس رویکرد CRISP-DM^۲ که از رویکردهای جامع و پراستفاده در داده‌کاوی هست، و با استفاده از الگوریتم‌های خوشه‌بندی k-means، two-step و kohonen مورد تحلیل قرار گرفتند. در ادامه نتایج حاصل از الگوریتم‌های خوشه‌بندی، شاخص‌های تعیین تعداد خوشه بهینه، شاخص‌های مربوط به اعتبار خوشه‌بندی توضیح داده شده است.

۱،۳. روش‌شناسی CRISP-DM

روش‌شناسی‌های مختلفی از جمله SEMMA^۴، KDD^۳، CRISP-DM و VC-DM^۵ برای انجام تحلیل داده‌ها در داده‌کاوی وجود دارد که با توجه به جامعیت و میزان استفاده و

1. Antti, 2010

2. Data mining (CRISP-DM) Cross-Industry Standard Process for

3. knowledge discovery from data (KDD)

4. Sample, Explore, Modify, Model, and Assess (SEMMA)

5. Virtuous Cycle of Data Mining (VC-DM)

کاربرد از روش‌شناسی CRISP-DM استفاده شده است (روگالویکز و سیکا، ۲۰۱۶، ص ۹۹). این روش‌شناسی شامل ۶ گام اصلی است که عبارتند از:

در ابتدا کلیه داده‌های مربوط به دستگاه‌های دولتی از سال ۸۱ تا ۹۶ جمع گردید. برای ایجاد یک تصویر اولیه از داده‌ها شاخص‌های مرکزی و پراکندگی داده‌ها مثل: میانگین، میانه، مد، انحراف معیار، دامنه تغییرات و هم‌چنین مقادیر کشیدگی و چولگی متغیرها محاسبه گردید. سپس در مرحله آماده‌سازی داده‌ها، مقادیر گم‌شده یا مفقود شناسایی و اقدامات لازم از جمله حذف یا جایگزین کردن آن‌ها صورت گرفت. به هر دستگاه در هر سال یک کد منحصر به فرد اختصاص داده شد. بعد از نهایی شدن مرحله پاک‌سازی داده‌ها، سه متغیر دیگر به متغیرهای سه‌گانه اصلی ما اضافه شد: اول متغیر "اختلاف" که اختلاف بین میزان اعتبار و تخصیص را نشان می‌دهد و دوم متغیر "نسبت" که نسبت بین تخصیص و اعتبار را بیان می‌کند و سوم متغیر "سابقه" که نشان می‌دهد هر دستگاه دارای چند سال سابقه اخذ بودجه در سال‌های ۸۱ تا ۹۶ است. دوباره شاخص‌های مرکزی و پراکندگی برای این ۶ متغیر محاسبه شد که نشان از عدم توزیع نرمال داده‌ها بود. هم‌چنین نتیجه آزمون کولموگروف-اسمیرنوف هم حاکی از توزیع غیر نرمال داده‌ها بود که در جدول زیر قابل مشاهده است.

جدول شماره ۱: نتیجه آزمون کولموگروف-اسمیرنوف

One-Sample Kolmogorov-Smirnov Test						
	SETEBAR	STAKHSIS	SDARYAFTI	Sekhtelaf	AVnesbat	
N	29326	29326	29326	29326	29326	
Normal Parameters ^{a,b}	Mean	5.36E+11	4.99E+11	4.88E+11	36221665841	.9309775173
	Std. Deviation	5.075E+12	4.892E+12	4.835E+12	6.456E+11	.1759251063
Most Extreme Differences	Absolute	.458	.459	.460	.478	.347
	Positive	.414	.416	.418	.440	.347
	Negative	-.458	-.459	-.460	-.478	-.319
Kolmogorov-Smirnov Z	78.427	78.661	78.738	81.793	59.492	
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	

a. Test distribution is Normal.

b. Calculated from data.

با توجه به متفاوت بودن مقیاس متغیرها، از روش نرمال‌سازی z-score برای نرمال‌سازی متغیرها استفاده شده است. حال داده‌ها آماده ورود به مرحله مدل‌سازی است. در مرحله مدل‌سازی از سه الگوریتم خوشه‌بندی two-step، k-means، و kohonen استفاده شده است که هر سه آن‌ها در گره auto cluster در نرم افزار spss modeler هستند.

الگوریتم k-means داده‌ها را به خوشه‌های مجزا خوشه‌بندی می‌کند. این روش تعداد خوشه‌های ثابتی را در نظر گرفته، به‌طور تکراری رکوردها را به خوشه‌ها تخصیص می‌دهد و مراکز خوشه‌ها را تا هنگامی که مدل بهبود نیابد تنظیم می‌کند. در عوض تلاش برای پیش‌بینی یک خروجی، k-means از یک فرآیند به نام یادگیری بدون نظارت برای کشف الگوها در مجموعه‌ای

از فیلدهای ورودی استفاده می‌کند (علیزاده و ملک‌محمدی، ۱۳۹۳، ص ۱۵۲). الگوریتم TwoStep از یک روش خوشه‌بندی دومرحله‌ای استفاده می‌کند. مرحله اول با یک بار گذر از داده‌ها، آن‌ها را در مجموعه قابل قبولی از زیر خوشه‌ها فشرده می‌کند. قدم دوم از یک روش خوشه‌بندی سلسله‌مراتبی، به منظور ادغام تکاملی این زیر خوشه‌ها به خوشه‌های بزرگ‌تر، بهره می‌برد. روش TwoStep مزیت تخمین خودکار تعداد بهینه خوشه‌ها را داراست. او می‌تواند فیلدهای با نوع مختلف و مجموعه داده‌های بزرگ را به خوبی مدیریت کند (علیزاده و ملک‌محمدی، ۱۳۹۳، ص ۱۵۳).

الگوریتم Kohonen از نوعی شبکه عصبی به منظور خوشه‌بندی مجموعه داده‌ها به خوشه‌های مجزا استفاده می‌کند. هنگامی که شبکه به طور کامل آموزش دید، رکوردهای شبیه هم در نقشه خروجی مجاور هم قرار می‌گیرند، درحالی‌که رکوردهایی که متفاوت‌اند دور از هم واقع می‌شوند. شما می‌توانید تعداد مشاهده‌های کسب‌شده توسط هر واحد را در قطعه مدل ببینید تا واحدهای قوی را شناسایی کنید. این مسئله ممکن است شهودی از تعداد مناسب خوشه‌ها به شما بدهد (علیزاده و ملک‌محمدی، ۱۳۹۳، ص ۱۵۲).

همچنین برای بررسی تعداد خوشه‌های بهینه از شاخص‌های AIC و BIC در نرم‌افزار SPSS استفاده شد که نتایج آن در مدل‌های K -means و towstep به ترتیب ۵ و ۲ خوشه به دست آمد. یکی از روش‌های ارزیابی خوشه‌بندی، معیار نیم‌رخ^۱ است. این معیار هم به پیوستگی^۲ درون خوشه‌ها و هم به میزان تفکیک‌پذیری آن‌ها بستگی دارد. مقدار نیم‌رخ برای هر نقطه، میزان تعلق آن را به خوشه‌اش در مقایسه با خوشه مجاور اندازه می‌گیرد. در جدول زیر مقدار شاخص نیم‌رخ برای سناریوهای مختلف بررسی داده در الگوریتم‌های خوشه‌بندی محاسبه شده است.

جدول شماره ۲: مقدار شاخص نیم‌رخ در سناریوهای مختلف

مقدار شاخص نیم‌رخ در حالات مختلف بررسی دستگاه‌ها						
سناریوها	مدل‌ها	کل دستگاه‌ها	بالای ده سال سابقه	دقیقاً ۱۶ سال سابقه	با نسبت بالای ۹۰ درصد تخصیص	سه انحراف معیار از میانگین
-auto cluster به تفکیک سال	kmean	۰,۹۸۴	۰,۹۸۷	۰,۹۶۱	۰,۹۸۵	۰,۹۳۳
	two step	۰,۹۸۲	۰,۹۸	۰,۹۵۴	۰,۹۸۱	۰,۹۱۷
	kohonen	۰,۵۵۱	۰,۵۳۹	۰,۶۳۲	۰,۶۰۷	۰,۵۱۶
-auto cluster با مجموع سال‌ها	kmean	۰,۹۹۳	۰,۹۷۲	۰,۹۳۷	۰,۹۹۳	۰,۹۳۳
	two step	۰,۹۸۴	۰,۹۵۷	۰,۹۰۱	۰,۹۸۹	۰,۸۹۴
	kohonen	۰,۵۸۷	۰,۵۱۳	۰,۵۹۲	۰,۶۴۱	۰,۵۶۳

1. Bayesian Information Criterion
2. Akaikes Information Criterion
3. Silhouette
4. Cohesion

بررسی خوشه‌بندی‌ها نشان می‌دهد که حالات مختلف بررسی داده تأثیر به‌سزایی در تغییر خوشه‌بندی دستگاه‌ها ایجاد نمی‌کند به‌طوری‌که در مدل TWOSTEP، ۹۶ درصد دستگاه‌ها دقیقاً در همان خوشه‌ای قرار گرفتند که در سایر حالات نیز بررسی شده‌اند. هم‌چنین در مدل خوشه‌بندی k-means در سناریو اول اگر خوشه‌های اول تا پنجم را به دودسته خوشه اول و سایر خوشه‌ها تقسیم‌بندی کنیم، ۹۵ درصد دستگاه‌ها در حالات مختلف بررسی داده در یک خوشه مشابه قرار گرفته‌اند.

با توجه به زیاد بودن تعداد اعضای خوشه اول در مدل twostep و k-means، این دو خوشه کنار گذاشته شدند و تحلیل‌ها بر مبنای سایر خوشه‌ها انجام شد. با توجه به بالاترین میزان شاخص نیم‌رخ که مربوط به بررسی داده‌ها در الگوریتم k-means در سناریو دوم (مجموع سال‌ها) است، به همین منظور دستگاه‌های مربوط به خوشه‌های دوم تا پنجم در مدل k-means در سناریو دوم کنار یکدیگر قرار داده شد و دستگاه‌های تکراری حذف شد که در نهایت به ۴۷ دستگاه رسید. این ۴۷ دستگاه مجدداً بر اساس مدل‌های k-means، twostep و kohonen مورد ارزیابی و خوشه‌بندی قرار گرفت که نتایج زیر حاصل شد:

تصویر شماره ۱: اجرای مدل خوشه‌بندی برای ۴۷ دستگاه باقی‌مانده



جدول شماره ۳: خلاصه نتایج گره auto cluster برای ۴۷ دستگاه باقی‌مانده

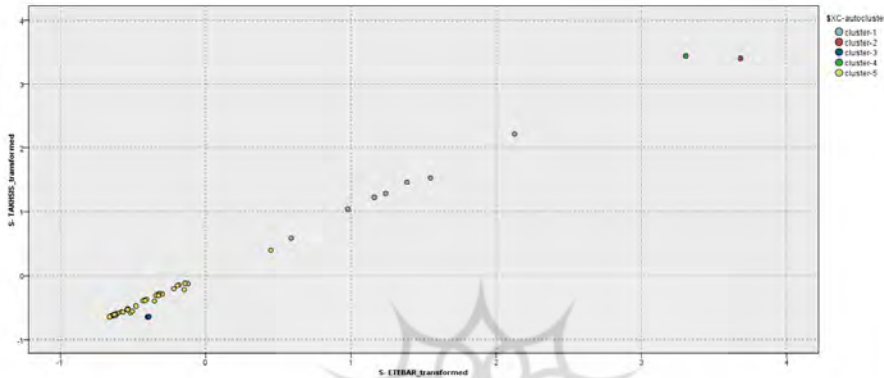
Graph	Model	Silhouette	Number of Clusters	Smallest Cluster (N)	Smallest Cluster (%)	Largest Cluster (N)	Largest Cluster (%)	Smallest/Largest
	K-means 1	0.821	5	1	2	36	76	0.028
	TwoStep 1	0.795	2	9	19	38	80	0.237
	Kohonen 1	0.329	11	1	2	17	36	0.059

بر اساس شاخص نیم‌رخ خوشه‌بندی k-means بهتر از twostep عمل کرده است. تعداد اعضای هر خوشه در مدل k-means عبارتند از:

جدول شماره ۴: تعداد اعضای هر خوشه در مدل k-means

cluster-۵	cluster-۴	cluster-۳	cluster-۲	cluster-۱
۳۶	۱	۲	۱	۷

نمودار شماره ۱: خوشه بندی دستگاه ها و نحوه پراکندگی آنها



جدول شماره ۵: نتایج مربوط به خوشه بندی و مشخص شدن خوشه هر دستگاه

sabegheh	Av-nesbat	autocluster	code
۱۳	۰٫۹۵	cluster-۱	A۱۶۱۱
۵	۰٫۹۵		A۳۲۵۳
۷	۱٫۰۰		A۳۶۵۵
۱۶	۰٫۹۹		A۳۷۰۴
۱۶	۰٫۹۹		A۳۷۱۴
۷	۱٫۰۰		A۴۰۱۲
۱۶	۰٫۹۸		A۵۰۰۴
۱۴	۱٫۰۰	cluster-۴	A۳۲۶۲
۱۵	۰٫۹۲	cluster-۲	A۴۵۱۰
۴	۰٫۴۷	cluster-۳	A۵۰۶۴
۶	۰٫۳۹	cluster-۵	A۵۰۸۰
۱۲	۰٫۹۱		A۳۳۷۵
۱۵	۰٫۸۱		A۱۵۶۷
۱۶	۰٫۸۸		A۲۰۳۷
۱۳	۰٫۹۹		A۳۰۹۹

sabegheh	Av-nesbat	autocluster	code
۱۳	۱,۰۰	cluster-۵	A۳۱۰۱
۱۲	۱,۰۰		A۳۱۰۷
۱۳	۰,۹۳		A۳۱۰۹
۱۱	۱,۰۰		A۳۱۱۳
۱۳	۱,۰۰		A۳۱۱۷
۱۳	۱,۰۰		A۳۱۲۳
۱۵	۱,۰۰		A۳۱۵۹
۱۲	۱,۰۰		A۳۱۶۰
۱۶	۰,۹۱		A۳۲۶۱
۱۶	۰,۹۲		A۳۴۴۶
۱۰	۰,۶۴		A۴۵۱۶
۱۶	۰,۹۳		A۵۰۲۹
۱۶	۰,۹۶		A۵۰۵۳
۱۶	۰,۵۸		A۵۱۱۳
۱۶	۰,۷۷		A۵۱۷۸
۱۶	۰,۹۵		A۵۲۳۴
۱۰	۰,۸۴		A۵۲۴۷
۱۶	۰,۷۵		A۱۹۳۶
۱۶	۰,۸۷		A۳۰۵۱
۱۶	۰,۸۶		A۳۲۵۱
۱۶	۰,۸۱		A۳۳۱۹
۱۶	۰,۶۹		A۳۳۲۱
۱۶	۰,۷۲		A۳۳۸۳
۱۶	۰,۶۸		A۳۴۲۰
۱۶	۰,۹۶		A۳۶۷۱
۱۶	۰,۸۸		A۴۶۳۵
۱۶	۰,۶۰		A۴۹۶۱
۱۶	۰,۸۲		A۵۰۳۹
۱۶	۰,۷۷		A۵۰۶۱
۱۶	۰,۶۸		A۵۲۹۱
۱۶	۰,۸۵	A۵۴۲۶	
۱۶	۰,۷۸	A۵۴۵۰	

۴. بحث و نتیجه‌گیری

تحلیل داده‌های بودجه به‌عنوان یک بخش مهم اقتصاد امری بسیار مهم است. تاکنون داده‌های مربوط بودجه دستگاه‌های دولتی به‌صورت دقیق و بر اساس الگوریتم‌های داده‌کاوی مورد بررسی قرار نگرفته بود. در این پژوهش بر اساس یکی از مهم‌ترین مدل‌های داده‌کاوی یعنی خوشه‌بندی به تحلیل داده‌های دستگاه‌های دولتی از سال ۱۳۸۱ تا ۱۳۹۶ پرداخته شد. متغیرهای پژوهش که شامل: میزان اعتبار، تخصیص، دریافتی، اختلاف اعتبار و تخصیص و نسب تخصیص به اعتبار هستند بر اساس الگوریتم‌های مختلف خوشه‌بندی و در حالات مختلف مورد بررسی قرار گرفتند. نتایج حاصل از خوشه‌بندی نشان می‌دهد که اولاً در نظر گرفتن حالات مختلف برای تحلیل داده‌ها، تأثیری به‌سزایی در ارائه خوشه‌های متفاوت‌تر و متنوع‌تر ندارد. ثانیاً آن‌که دستگاه‌های دارای نسبت تخصیص به اعتبار بالا، دارای یک الگوی مشابه دریافت بودجه در سال‌های اخیر هستند و از طرف مجلس، سازمان برنامه بودجه و وزارت اقتصاد و دارایی نگاه ویژه‌ای به این دستگاه‌ها می‌شود.

پیشنهاد می‌شود پژوهشگران علاقه‌مند به تحلیل داده به‌ویژه در حوزه بودجه، به بررسی ارتباط سایر متغیرهای کلیدی مثل تعداد کارکنان، استان، نوع سازمان با میزان پرداخت بودجه بپردازند. همچنین بررسی روند پرداخت بودجه دستگاه‌های دولتی برای پیدا کردن بهترین روش پیش‌بینی میزان تخصیص و اعتبار دستگاه‌ها کمک شایانی به تمامی سازمان‌های مرتبط و درگیر در امر بودجه برای برنامه‌ریزی و تصمیم‌گیری‌های مؤثر و دقیق مبتنی بر روند گذشته نماید تا بتوانند با یک دید صحیح و جامع به پیش‌بینی و تخصیص منابع بودجه‌ای به دستگاه‌های دولتی بر اساس اولویت‌ها بپردازند.

۵. فهرست منابع

الف) منابع فارسی

۱. آذر، عادل؛ نجفی، ابراهیم؛ و نجفی، سجاد. (۱۳۹۰). مدلسازی ریاضی استوار، رویکردی نوین در بودجه‌ریزی عمومی ایران. پژوهش‌های مدیریت در ایران، ۱۵(۲)، ۱-۲۰.
۲. آذر، عادل؛ و امیرخانی، طیبه. (۱۳۹۱). کتاب بودجه‌ریزی عمومی: نهادهای بودجه‌ریزی و بودجه محلی. تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت).
۳. آذر، عادل؛ و خدیور، آمنه. (۱۳۹۶). بودجه‌ریزی بر مبنای عملکرد: پارادایم مدل سازی. مرکز پژوهش‌های مجلس شورای اسلامی.
۴. آذر، عادل؛ و نجفی، سجاد. (۱۳۹۰). مدل ریاضی بودجه‌ریزی در بخش عمومی: رویکرد بهینه سازی استوار. چشم انداز مدیریت دولتی، ۸۳-۹۸، (۸)، ۰.
۵. اربابیان، شیرین؛ قاسمی، محمد رضا؛ و وفائی، وجیهه. (۱۳۹۱). بررسی خطای پیش‌بینی بودجه ایران طی سال‌های ۸۷ - ۱۳۴۴. تحقیقات توسعه اقتصادی، ۷۰(۷)، ۵۱-۶۸.
۶. تقوی فرد، محمد تقی؛ منصوری، طاهار؛ ناصرزاده، سید محمد رضا؛ و فراست، علی رضا. (۱۳۸۶). داده‌کاوی و کاربرد آن در تصمیم‌گیری‌ها. دانش مدیریت، ۲۰(۷۹)، ۳-۱۴.
۷. حمیدی زاده، محمد رضا. (۱۳۸۰). ساختار و فرایند تصمیم‌گیری سازمانی (۱). توسعه مدیریت، ۹(۲۴-۲۵)، ۴۴-۴۶.

۸. دستگیر، محسن؛ و شفیعی سردشت، مرتضی. (۱۳۹۰). فناوری داده‌کاوی؛ رویکردی نوین در حوزه مالی. دانش‌های حسابرسی، ۱۱(۴۳)، ۶-۲۷.
۹. رازینی، روح‌الله؛ و عزیزی، مهدی. (۱۳۹۴). طراحی الگوی تصمیم‌گیری با رویکرد اسلامی. مدیریت اسلامی، ۲۳(۱۲)، ۷۳-۱۰۰.
۱۰. طالب‌نیا، قدرت‌الله؛ و طالبی، مریم. (۱۳۹۷). نظام بودجه‌ریزی در ایران و ساختار آن. رویکردهای پژوهشی نو در علوم مدیریت، ۱(۶)، ۴۱-۶۸.
۱۱. عباسی، ابراهیم. (۱۳۸۳). اصلاحات ساختاری در نظام بودجه‌ریزی کشور. مطالعات مدیریت، بهبود و تحول، ۱۵(۴۲-۴۱)، ۷۹-۱۲۴.
۱۲. علیزاده، سمیه؛ و ملک‌محمدی، سمیرا. (۱۳۹۳). داده‌کاوی و کشف دانش گام به گام با نرم‌افزار. تهران: دانشگاه صنعتی خواجه نصیرالدین طوسی.
۱۳. قوسی، روزبه؛ چیذری، عماد؛ و وحدانی، هانی. (۱۳۹۷). طراحی سامانه تصمیم‌یار با استفاده از ابزار داده‌کاوی (مطالعه موردی معاونت فرهنگی دانشگاه علم و صنعت). مدیریت فردا، ۱۷(۵۴)، ۱۶۳-۱۷۶.
۱۴. کقیبادی، امیررضا؛ و خدای، وحید. (۱۳۹۲). داده‌کاوی صورت‌های مالی جهت اعطای تسهیلات مالی. پژوهش‌های حسابداری مالی و حسابرسی، ۵(۱۷)، ۱۷۹-۲۱۲.
۱۵. مسلم‌زاده گنابلی، آزاده. (۱۳۹۴). روش‌های داده‌کاوی برای کشف تقلب در حسابرسی صورت‌های مالی. حسابدار رسمی، ۱۲(۲۹)، ۲۲-۳۲.
۱۶. معتمدوزیری، فریدون. (۱۳۷۲). فرایند تصمیم‌گیری. تعاون و کشاورزی، ۱۹-۱۲۰(۱)، ۳۲-۳۷.
۱۷. معتمدوزیری، فریدون. (۱۳۷۲). فرایند تصمیم‌گیری. تعاون و کشاورزی، ۱۹-۱۲۰(۱)، ۳۲-۳۷.
۱۸. مقیمی، سیدمحمد. (۱۳۹۶). بودجه‌ریزی دولتی: نظریه‌ها، اصول و کاربردها از تنظیم تا کنترل. تهران: انتشارات راه دان.
۱۹. هاشمی، سید عباس؛ حسینی، سید محسن؛ و برعندان، سجاد. (۱۳۹۱). مقایسه محتوای فزاینده اطلاعاتی نسبت‌های نقدی و تعهدی برای ارزیابی عملکرد مالی شرکت‌ها با رویکرد داده‌کاوی. پژوهش‌های حسابداری مالی، ۴(۱۲)، ۶۳-۸۲.

ب) منابع انگلیسی

1. Antti, Syvajarvi. (2010). Data Mining in Public and Private Sectors: Organizational and Government Applications: Organizational and Government Applications. IGI Global.
2. Bhatia, Parteek. (2019). Data mining and data warehousing: principles and practical techniques. Cambridge, United Kingdom ; New York, NY: Cambridge University Press.
3. Davis, Ralph C. (1958). A Philosophy of Management*. Academy of Management Journal. <https://doi.org/10.5465/254624>
4. Koford, Brandon C. (2010). Public Budget Choices and Private Willingness to Pay (SSRN Scholarly Paper No. ID 1620651). Rochester, NY: Social Science Research Network.
5. Lee, Robert D.; Johnson, Ronald Wayne; & Joyce, Philip G. (2013). Public Budgeting Systems. Jones & Bartlett Publishers.
6. Malgwi, Abdullahi A.; & Unegbu, Angus O. (2012). Budget in Nigerian Public Sector: Need for Balanced Scorecard Perspective. International Journal of Finance and Accounting, 1(2), 1-6.
7. Rogalewicz, Michał; & Sika, Robert. (2016). Methodologies of Knowledge Discovery from Data and Data Mining Methods in Mechanical Engineering. Management and Production Engineering Review, 7(4), 97-108. <https://doi.org/10.1515/mper-2016-0040>
8. Scott, David. (2013). [Review of Review of Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R, Criticism D. S. Putler & R. E. Kridler]. International Statistical Review / Revue Internationale de Statistique, 81(2), 328-328.
9. Tan, Pang-Ning; Steinbach, Michael; & Kumar, Vipin. (2007). Introduction to data mining.

- Zaretzki, Russell L. (2011). [Review of Review of Business Intelligence: Data Mining and Optimization for Decision Making, Criticism C. Vercellis]. Journal of the American Statistical Association, 106(493), 376-377.

