فصلنامه علمی- پژوهشی
فناوری اطلاعات و ارتباطات ایران

سال هشتم، شماره‌های ۲۷ و ۲۸، بهار و تابستان ۱۳۹۵
صص: ۶۱ - ۸۰

تلفیق مدل تحلیل پوششی داده‌ها و درخت تصمیم به منظور ارزیابی واحدهای مبتنی بر تکنولوژی اطلاعات

[*]امیر امینی [**] علیرضا علی‌نژاد [***]سمیه شفقی‌زاده
*دانشجوی کارشناسی‌ارشد، مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه صنعتی ارومیه، ارومیه، ایران
**دانشیار دانشگاه آزاد اسلامی، واحد قزوین، دانشکده مهندسی صنایع و مکانیک، قزوین، ایران
***دانش‌آموخته کارشناسی‌ارشد مهندسی صنایع، دانشکده مهندسی صنایع، دانشگاه آزاد اسلامی، واحد علوم تحقیقات قزوین، قزوین؛ ایران
تاریخ دریافت: ۹۳/۰۸/۳۰ تاریخ پذیرش: ۹۵/۰۸/۲۳

چکیده

هر سازمان به منظور آگاهی از میزان عملکرد و مطلوبیت فعالیت واحدهای خود به یک نظام ارزشیابی جهت سنجش این مطلوبیت نیاز دارد و این موضوع برای مؤسسات مالی از جمله شرکت‌های مبتنی بر تکنولوژی اطلاعات اهمیت بیشتری دارد. تحلیل پوششی داده‌ها[۱] یک روش غیرپارامتری برای اندازه‌گیری کارایی و بهره‌وری از واحدهای تصمیم‌گیری[۲](DMUs)هاست. از طرف دیگر تکنیک داده‌کاوی[۳] به (DMUs)ها اجازه کاوش و کشف اطلاعات معنی‌داری، که قبلاً در پایگاه داده‌های بزرگ پنهان بوده‌اند را می‌دهد. این مقاله یک چارچوب کلی ترکیب تحلیل پوششی داده‌ها بادرخت رگرسیون [۴]را برای ارزیابی کارایی و بهره‌وری ازDMUها پیشنهاد می‌کند. نتیجه مدل ترکیبی مجموعه‌ای از قوانین است که می‌تواند توسط سیاست‌گذاران برای کشف دلایلDMUهای کارآمد و ناکارآمد مورد استفاده قرار گیرد. به عنوان مطالعه موردی با استفاده از روش پیشنهادی برای بررسی عوامل مرتبط با بهره‌وری نمونه‌ای شامل ۱۸ شعبه از بیمه ایران در شهر تهران انتخاب گردید و پس از مدل‌سازی بر مبنای مدل پیشرفته LVM[۵] ورودی محور با دسترسی پذیری ضعیف درتحلیل پوششی داده‌ها با خروجی نامطلوب[۶]محاسبه گردید و با تکنیک درخت تصمیم، به استخراج قوانین برای کشف دلایل افزایش بهره‌وری و پسرفت بهره‌وری می‌پردازد.

واژه‌های کلیدی: تحلیل پوششی داده‌ها، داده‌کاوی، طبقه‌بندی و رگرسیون، درخت تصمیم، خروجی نامطلوب

[۱] Data Envelopment Analysis
[۲] Decision Making Units
[۳] Data Mining
[۴] C&R
[۵] Latent Variable Model
[۶] Undesirable Output

نویسنده عهده‌دار مکاتبات : امیر امینی mr۶۲.amini@gmail.com

۱- مقدمه

اندازه‌گیری راندمان یا کارایی، همیشه یکی از مباحث مهم مدیریت بوده است. آنچه پیکره دفاعی یک واحد اقتصادی را استحکام می‌دهد و بر دیگر رقبا ارتقاء می‌بخشد، نگاه تیزبینانه‌ای است که برنامه‌ریزان آن واحد با ایجاد بستری مناسب سعی می‌کنند که ضمن شناسایی نقاط ضعف و قوت موجود، درصدد رفع نقایص و ارتقاء قوت در کل مجموعه برآیند و این میسر نمی‌شود، مگر با استفاده از روش‌هایی که در نهایت صحت و به دور از هرگونه اغواگری، موجب شناسایی واحدهای کارا از ناکارا شود.

امروزه صنایع با یک محیط به شدت رقابتی مواجه‌اند. صنعت بیمه نیز از این محیط به شدت رقابتی بی بهره نبوده و در سال‌های اخیر شاهد رقابت فزاینده‌ای بین شرکت‌های بیمه بوده‌ایم، در مواجهه با یک محیط رقابتی، تدوین استراتژی‌های رقابتی، تقویت فعالیت‌های شرکت وارتقای کیفیت خدمات برای بقاء ضروری است، در جهان امروز، بیمه از مهمترین بخش‌های اقتصادی در جامعه است که علاوه بر تسهیل روابط بین الملل در پیشگیری از تحقق ریسک و نیز تأمین آرامش خاطر برای فعالیت‌های مختلف نقش شگرفی را ایفا می‌نماید. شعب مختلف یک شرکت بیمه به‌رغم استفاده از منابع مادی و انسانی مشابه، عملکردهای متفاوتی دارند. یکی از دلایل، تفاوت در عملکرد شعب و نمایندگی‌های بیمه در نحوه به‌کارگیری منابع انسانی و مدیریت می‌توان دانست [۴].

این تحقیق به دنبال ارزیابی کارایی شعب بیمه ایران است تا واحدهای تصمیم‌گیرنده بتوانند به کمک نتایج تحقیق یا بسط آن اهداف زیر را پیگیری نمایند:

- ایجاد فضای رقابت سالم بین شعب با توجه به نتایج تحقیق
- استفاده از تجربیات واحدهای کارا و انتقال تجربه و دانش آن‌ها به سایر شعب
- اهداف کاربردی مستتر در این ارزیابی، استفاده از نتایج جهت سیاست‌گذاری شرکت و استفاده ازآن به عنوان محک مدیران شعب است.

اما این سؤال مطرح می‌شود که شرکت‌های بیمه‌ای موجود در کشور تا چه اندازه از منابع در دسترس خود به صورت کارا استفاده می‌کنند و نحوه عملکرد شرکت‌های بیمه‌ای در استفاده بهینه از منابع در اختیار آن‌ها از دیگر عوامل مؤثر برتوسعه بازارهای بیمه و افزایش کارایی آنها محسوب می‌گردد.

امروزه شرکت‌ها سعی می‌کنند عملکردشان را اندازه‌گیری کرده و آن را با رقبایشان مقایسه نمایند تا بتوانند جهت حصول به سطح عملکردی که بتواند آنها را در بازار حفظ نماید، اقدام مقتضی را انجام دهند [۲۹].

چارنز و همکاران (۱۹۷۸) به معرفی یک روش غیرپارامتری اندازه‌گیری کارایی تکنیکی با یک مجموعه از واحدهای تصمیم‌گیرنده قابل مقیاس با استفاده از روش تحلیل پوششی داده‌ها پرداختند.

اندازه‌گیری کارایی واحدهای تصمیم‌گیرنده (DMUs)ها مانند بیمارستان‌ها، بانک‌ها، شرکت‌های بیمه و... که با چندین ورودی، چندین خروجی تولید می‌کنند کار بسیار پیچیده‌ای است. تحلیل پوششی داده‌ها از مفهوم ارزیابی کارایی یک گزینه تصمیم‌گیری بر مبنای عملکردش در ایجاد خروجی‌ها بواسطه مصرف ورودی بسط یافته است [۳۵]. کارایی هر واحد تصمیم‌گیری در مقایسه با واحد دیگر به عنوان نسبت مجموع وزنی خروجی های آن عضو به مجموع وزنی ورودی ها تعریف می‌شود [۲۷].

در طول دهه گذشته DEA به عنوان یک ابزار مدیریتی قدرتمند برای ارزیابی عملکرد به طور گسترده، توجه قابل ملاحظه‌ای را به دست آورده است و از طرف دیگر درخت تصمیم یک تکنیک داده کاوی برای طبقه‌بندی نمونه‌ها یا عملکرد ثابت تکه‌ای در رگرسیون غیرخطی است، این الگوریتم توسط برایمن[۷] و همکاران (۱۹۸۴) ابداع شده است.

برای ارزیابی کارایی شرکت‌های بیمه از مدل تحلیل پوششی داده‌ها با خروجی نامطلوب استفاده شده به این دلیل که فرآیند تولید صنعت بیمه، در حالت کلی شامل دو فعالیت است؛ یکی فعالیت بازاریابی که مربوط می‌شود به بازاریابی، بیمه‌کردن و پرداختن ادعای خسارت‌ها که به عنوان خروجی

[۷] Breiman

که تولید کمتر آنها به صرفه‌تر است وآن را عوامل نامطلوب می‌نامیم و دیگری فعالیت سرمایه‌گذاری که عبارت است از منافع حاصل از سرمایه‌گذاری منابع مالی مؤسسه بیمه.

این تحقیق بر آن است تا با استفاده از مدل LVM[۸]، تغییرات کارایی و بهره‌وری شعب بیمه ایران در طی سال (۱۳۸۹-۱۳۸۷) در شهر تهران را باهم مقایسه نموده، شاخصی برای تعیین کارایی و ناکارایی یک واحد با استفاده از تغییرات مرز حاصل از ورودی واحدها و تغییرات کارایی آن‌ها صورت دهد و در نهایت با ارائه نتایج، کارایی و بهره‌وری شعب را مورد بحث و بررسی قرار داده و جایگاه هریک از شعب را در این مدل تعیین کند.

نهاده‌هایی که در این تحقیق مورد استفاده قرار گرفته‌اند، تعداد نمایندگی‌ها، هزینه‌های اداری و بیمه‌گری هستند و ستانده‌های مورد استفاده، درآمد حاصل از حق بیمه و ارزش وام‌های پرداختی به عنوان خروجی مطلوب و ارزش خسارت پرداختی به عنوان خروجی نامطلوب در نظرگرفته شده است، که شناسایی شاخص‌ها و معیارهای عملکردی از طریق مصاحبه و نظرسنجی با کارشناسان و مدیران ارشد بیمه ایران و مطالعه اسنادی صورت پذیرفت. این داده‌ها با مراجعه مستقیم امور مالی و شعب بیمه ایران در شهر تهران جمع‌آوری شده است.

تحقیق حاضر از نظر هدف، از نوع تحقیقات کاربردی محسوب می‌شود. همچنین از نظر نحوه گردآوری داده‌ها این تحقیق توصیفی و از نوع علّی- مقایسه‌ای می‌باشد. جامعه آماری این تحقیق کلیه شعب بیمه ایران، درشهر تهران می‌باشند.

۲. بیان مسأله

مفهوم کارایی در اقتصاد، تخصیص مطلوب منابع است. اما از نظر اهداف کاربردی، تعاریف گوناگونی بیان شده است. به طور کلی کارایی، معرف نسبت ستانده‌ها به نهاده‌ها در مقایسه با یک استاندارد مشخص است [۷].

در عصر ما کارایی و بهره‌وری از بالاترین اهداف مدیران و ارزشمندترین مقصد همه سازمان‌ها می‌باشد.

تلاش برای افزایش بهره‌وری جدی‌ترین مبارزه‌ای است که مدیریت در آستانه قرن بیست‌ویکم با آن روبروست. در نگاهی ساده به بهره‌وری، نسبت بازده به منابع مصرف شده است. لذا اگر با هزینه یکسان محصول بیشتری تولید شود یا مقدار معینی محصول با هزینه کمتری تولید شود، منافعی برای کل جامعه حاصل می‌شود که می‌تواند برای تأمین زندگی بهتر برای افراد جامعه مؤثر باشد [۱].

اندازه‌گیری کارایی و روش‌های محاسباتی آن از اواخر قرن بیستم توجه بیشتر اقتصاددانان را برانگیخته است[۳۷].

روش‌های اندازه‌گیری اغلب برآورد مرز کارای تولید و به دست آوردن منحنی مرزی یا تابع تولید مرزی و کلاً به دو روش پارامتریک و ناپارامتریک بوده است. روش تحلیل پوششی داده‌ها[۹] که در سال ۱۹۷۸ چارنز و همکارانش[۱۰] ارائه کردند، نخست برای ارزیابی اقتصادی و فنی واحدهای تولیدی معرفی گردید. امروزه این روش برای ارزیابی عملکرد واحدهای خدماتی، اعم از دولتی و غیردولتی و واحدهای تولیدی [۲۵] و [۳۴] و [۱۹] و [۶] کاربرد گسترده‌ای دارد.

کائو و هوآنگ [۲۴] برای ارزیابی کارایی شرکت‌های بیمه غیرعمر تایوان از مدل دو مرحله‌ای و در واقع DEA دو مرحله‌ای استفاده کرده‌اند. مدل ارتباطی DEA ارتباطی متداول است تا بتواند رابطه فیزیکی بین فرآیند کل و زیرتوسعه یافته مدل DEA دو مرحله‌ای، محدودیت‌های DEA و فرآیندهای جزء را توصیف کند. در مدل ارتباطی هر دو زیر فرآیند، به محدودیت فرآیند کل اضافه می‌شوند.

در صنعت بیمه نیز تحقیقات متعددی از این روش جهت ارزیابی عملکرد نمایندگی‌های بیمه بهره گرفته‌اند [۵] و [۲۸] و [۳۹].

در همین سال تحقیقات متعددی به طرق گوناگون در زمینه ارزیابی کارایی شرکت‌های بیمه با استفاده از روش تحلیل پوششی داده‌ها در کشورهای آلمان [۲۸]، چین [۱۴]، تایوان [۱۸]، ایالات متحده [۳۰] و بسیاری از کشورهای دیگر انجام شده است.

[۸] Latent Variable Model

[۹] Data Envelopment Analysis

[۱۰] Charnens et al, ۱۹۷۸

چن و همکارانش[۱۱]، کارایی شرکت‌های بیمه تایوان را با استفاده از روش دومرحله‌ای تحلیل پوششی داده‌ها بررسی کردند [۱۴].

گونزالس و همکاران[۱۲]در سال ۲۰۰۹ به تحلیل DEA از خطر، هزینه و درآمد در بیمه پرداختند، با استفاده از تحلیل پوششی داده‌ها به مطالعه در زمینه تجزیه و تحلیل نتایج حاصل از DEA در زمینه چند متغیره، تحلیل آماری با استفاده از تحلیل عاملی تفسیر شده و تأثیرعوامل خطر در بهره‌وری مشخص شده است.

زای[۱۳]، کارایی شرکت‌های بیمه دولتی و خصوصی ایالات متحده را با استفاده از روش تحلیل پوششی داده‌ها مقایسه کرد و دریافت به‌رغم کارایی بالای شرکت‌های بیمه خصوصی، علت تمایل ایشان به دولتی شدن به دلیل دسترسی به منابع و سرمایه بیشتر است [۳۸].

باروز و همکارانش[۱۴] نیز در همین سال با استفاده از روش دو مرحله‌ای تحلیل پوششی داده‌ها کارایی شرکت‌های بیمه یونان را در دوره ۹ ساله بررسی کردند [۹].

حمید اسفاندرانی [۴] در تحقیق خودکه جامعه آماری آن شامل ۲۴ شعبه و۱۰۰ نمایندگی واقع در تهران شرکت بیمه ایران می‌باشد. در این تحقیق انتخاب نهاده‌ها و ستانده‌ها با توجه به تحقیقات مشابه بوده و مصاحبه و نظرسنجی از کارشناسان و خبرگان بیمه انجام گرفته است. در نهایت با توجه به تجزیه و تحلیل‌های انجام گرفته از مدل های متعدد DEA، مدلBCC نهاده‌گرا به منظور اندازه‌گیری کارآیی شبکه فروش بیمه عمر مدلی مناسب تشخیص داده شده است.

شاهرودی و همکاران[۱۵](۲۰۱۱) به تجزیه و تحلیل کارایی بیمه‌های خصوصی در ایران با روش تحلیل پوششی داده‌ها، به محاسبه کارایی بیمه‌های خصوصی بین سال‌های۲۰۰۷-۲۰۰۶ با روش DEA دومرحله‌ای پرداختند. سئول و همکاران (۲۰۰۷) از تلفیق تحلیل پوششی داده‌ها و درخت تصمیم به عنوان یک روش الگوبرداری در فرایند خدمت استفاده کرده‌اند [۳۲]. بابازاده و همکاران (۲۰۱۵) رویکرد یکپارچه دوگانه‌ای از تحلیل پوششی داده‌ها و تکنیک‌های برنامه‌ریزی ریاضی را به منظور طراحی راهبردی شبکه زنجیره تأمین زیست – دیزل در ایران ارائه داده اند [۸].

۳. مواد و روش

در این مقاله، دو روش تلفیقی تحلیل پوششی داده‌ها و داده کاوی که برای تخمین بهره‌وری واحدهای تصمیم گیرنده (DMUs) مورد استفاده قرار می‌گیرد، در نظر گرفته شده است. در روش تحلیل پوششی داده‌ها، تخمین کارایی و صرفه‌های مقیاس، با استفاده از سیستم برنامه‌ریزی ریاضی صورت می‌گیرد. در این روش محدودیتی در انتخاب تعداد نهاده‌ها و ستانده‌ها نداریم، البته به شرط آنکه تعداد واحدهای تصمیم‌گیرنده به اندازه کافی زیاد باشند، هیچ گونه نیازی به انتخاب نوع تابع تولید نیست و تکنیک داده کاوی اجازه کشف و بررسی معنادار اطلاعاتی که قبلاً از یک پایگاه داده بزرگ پنهان بوده‌اند را می‌دهد، C&R همان درخت تصمیم در داده‌کاوی است که توسط برایمن و همکاران در سال۱۹۸۴ توسعه یافته است و بعدها توسط رایپلی[۱۶] بهبود یافته است.

۳-۱- روش تحلیل پوششی داده‌ها (DEA)

به طور کلی شالوده روش ناپارامتری برای اندازه‌گیری کارایی با انتشار مقاله‌ای از فارل[۱۷](۱۹۵۷) بنیان نهاده شد. وی با یک رهیافت صرفاً ریاضی، روش جدیدی را برای اندازه‌گیری کارایی مدل در مقابل روش‌های پارامتری معرفی نمود [۲۳]. تحلیل پوششی داده‌ها روشی ناپارامتریک بر مبنای برنامه‌ریزی خطی است و نخستین بار توسط چارنز و همکاران ارائه شد [۱۳].

[۱۱] Chen et al, ۲۰۰۹
[۱۲] MM Segovia-Gonzalez, Contreras and C Mar-Molinero
[۱۳] Xie, ۲۰۱۰
[۱۴] Barros et al, ۲۰۱۰
[۱۵] Shahroudi, ۲۰۱۱
[۱۶] Ripley
[۱۷] Farrel ۱۹۵۷

امروزه روش تحلیل پوششی داده‌ها به عنوان یک روش مناسب برای ارزیابی عملکرد واحدها شناخته شده است که در آن بر پایه اطلاعات موجود، مرز کارا به صورت تجربی برآورد می شود. از آنجایی که در دستیابی به تابع مرزی، همه داده ها پوشش داده می‌شوند، آن را تحلیل پوششی داده ها نام نهاده‌اند.

کارایی تکنیکی در روند تولید صنعت بیمه، ورودی‌ها را به خروجی تبدیل می‌کند (هزینه‌های اداری و بیمه‌گری و تعداد نمایندگی‌ها به عنوان ورودی و ارزش وام‌های پرداختی و درآمد حاصل از حق بیمه به عنوان خروجی مطلوب و ارزش خسارت پرداختی به عنوان خروجی نامطلوب). ارتباط بین ورودی‌ها و خروجی‌ها می‌تواند به صورت یک تابع تولید که بیشترین خروجی ممکن برای یک سطح از ورودی‌های ارائه شده را نشان می‌دهد بیان شود، کارایی تکنیکی یک اندازه‌گیری است که نشان می‌دهد، چگونه میزان حداکثر خروجی از ورودی‌های دسترس به دست می آید.

۳-۲- کارایی تکنیکی با چندین ورودی و چندین خروجی

فرض کنید، n شعبه بیمه وجود دارد (j=۱،۲،..,n) با m ورودی(x_{ij} ,i=۱،۲،…,m) و s خروجی تولید می‌کند (y_{rj} , r=۱،۲،…,s). DEA کارایی تکنیکی شعبه j_0 را در مقایسه با گروه همتا از n شعبه را به صورت زیر اندازه‌گیری می‌کند:

- کارایی تکنیکی: یک مدل برنامه ریزی خطی برای بازده به مقیاس متغیر(VRS)

$$\text{Max } h_{j_0} = \sum_{r=1}^{s} u_r y_{rj_0} + u_{j_0} \qquad (1)$$

st:

$$\sum_{i=1}^{m} v_i x_{ij_0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + u_{j_0} \le 0 \quad j=1, 2, \ldots, n$$

$$u_r , v_i \ge \varepsilon \quad i= 1, \ldots, m, r=1, \ldots, s$$

u_{j_0} free

در این مدل، علامت u_0 بازده به مقیاس را تعریف می‌کند: $u_0 > 0$ به معنای بازده به مقیاس افزایشی است؛ $u_0 = 0$ به معنای بازده به مقیاس ثابت و $u_0 < 0$ نشان دهنده بازده به مقیاس کاهشی است.

مدل (۱) یک مدل وزنی است و پوششی این مدل درمدل (۲) نشان داده شده است.

کارایی تکنیکی: یک مدل برنامه ریزی خطی برای بازده به مقیاس متغیر: مدل پوششی، ورودی‌محور

$$\min h_{j_0} = \varepsilon (s^+_r + s_i^-) \qquad (2)$$

St:

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s^-_I = h_{j_0} x_{ij_0} \quad \forall \ i = 1,\ldots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s^+_r = y_{rj_0} \quad \forall \ r= 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \ge 0 \qquad \forall j=1, \ldots, n$$

در این مدل، محدودیت اول نشان‌دهنده این است که همه ترکیب‌های ورودی مشاهده شده برای jامین واحد تصمیم‌گیرنده بر روی مرز یا زیر مرز تولید قرار دارد و دومین محدودیت نشان می‌دهد که سطح‌های خروجی از مشاهدات ناکارایی با سطح‌های واحد مرجع قابل مقایسه است، که از ترکیب محدب خروجی مشاهده تشکیل شده است.

محدودیت سوم معرف بازده به مقدار متغیر است و محدودیت آخر تضمین می‌کند که همه مقادیر تولید اوزان تحدب بزرگتر مساوی یک است، بنابراین واحد فرضی مرجع درون مجموعه امکان است، DMU_{j0} کاراست اگر و تنها اگر $h_0 = 1$ و تمام متغیرهای کمکی برابر صفر باشند ($\forall s_r^+, s_i^- = 0$) [۲۵].

معرفی مدل

۳-۳- مدل پیشنهادی روش ناپارامتری خروجیهای نامطلوب با اصل دسترس‌پذیری ضعیف

- دسترس پذیری ضعیف ورودی‌ها و متغیر پنهان

برتولت و پان برای زمانی که تکنولوژی شامل خروجی‌های نامطلوب باشد، روشی را معرفی کردند که براساس اصول زیر ساخته می‌شود [۱۷].

فرض کنید در یک فعالیت تولیدی، P ورودی در تولید، Q خروجی مطلوب و R خروجی نامطلوب به کار برده شود. $X \in \mathrm{R}_+^P$ بردار ورودی مصرف شده و $Y \in R_+^Q$ ، $U \in R_+^R$ به ترتیب نشان‌دهنده‌ی بردارهای خروجی‌های مطلوب و خروجی‌های نامطلوب باشند، همچنین $u \in \mathrm{R}_+^P$ ، $v \in R_+^Q$ و

$w \in R_+^R$ به ترتیب نشان‌دهنده قیمت ورودی‌ها، قیمت خروجی‌های مطلوب و قیمت خروجی‌های نامطلوب باشند و مجموعه مشاهدات با S نشان داده شود.

تمام اصول موضوعه، همراه یک فاکتور انقباضی برای خروجی‌های نامطلوب برای تشکیل مدل LV ضروری است. هایلو و ویمن ادعا کردند که اگر قید ورودی‌ها یا خروجی‌های نامطلوب را با فرم $X = \sum_{j=1}^{J} Z_j X_j$ تساوی جایگزین شود مدل DEA حاصل، دسترسی‌پذیری ضعیف را نشان می‌دهد .

VRS LV Min α:
{∀ DMU ‖ j=۱, ۲, … J: K, L}
St. $\sum_{j=۱}^{J} Z_j X^t_{pj} = \alpha X^t_{p\cdot}$, p= ۱, ۲, …, P (۳)
$\sum_{j=۱}^{J} Z_j y_{qj}{}^t \geq y_{q\cdot}{}^t$, q= ۱, ۲, …, Q
$\sum_{j=۱}^{J} Z_j u_{rj}{}^t \leq \lambda\, u^t_{r\cdot}$, r= ۱, ۲,…, R
$\sum_{j=۱}^{J} Z_j$=۱
Z
$Z_j \geq$ ۰

Latent Variable $۰ \leq \lambda = \frac{\sum_{j=۱}^{J} Z_j U^t_{rj}}{U} \leq ۱$
{∀ DMU ‖r= ۱, ۲, …, P}

در این مقاله یک مدل جدید برای ارزیابی کارایی، خروجی‌ها به عنوان ورودی پیشنهاد شده است، با در نظر گرفتن یک فاکتور انقباضی برای خروجی‌های نامطلوب، در اینجا با استفاده از شاخص مالم کوییست به بررسی پراکندگی بین دوره ها نیز می‌پردازیم.

پس از به دست آورد مقادیر کارایی واحدهای تصمیم گیرنده، بهره‌وری را برای بازه زمانی سال‌های۱۳۸۸-۱۳۸۷و ۱۳۸۹-۱۳۸۸ با توجه به فرمول‌های زیر به دست می‌آوریم.

شاخص بهره‌وری مالم کوییست برای ارزیابی تغییرات تکنولوژی واثر تغییرات روی ورودی‌ها و خـروجی‌ها اسـتفاده می‌شود [۲۸].

مالمکوئیست به صورت بیشترین فاکتور به وسیله ورودی‌هایی که روی یک دوره که می‌تواند کاهش یابد، تعریف می‌شود و تولید کردن همان خروجی در یک دوره دوم را می‌تواند محاسبه کند.

فرض کنید تکنولوژی تولید در دوره K، به طوری‌که نمایش اصلی ضریب کاهش به این صورت باشد.

$\lambda_j{}^k (X_j{}^L, Y_j{}^L, U_j{}^L)$ (۴)

به طوری که مقادیر هدف در دوره L هستند، به بیان مختصرتر با ساده‌سازی نماد اصلی را به $\lambda_j^k(U_j^L)$ نشان می‌دهیم، که این نشان می‌دهد اندازه‌گیری عملکرد برای خروجی نامطلوب در دوره L در حالی‌که با تکنولوژی مرجع دوره K محاسبه شده است.

به طور مشابه، $\alpha_j^k(X_j^L)$ اندازه‌گیری عملکرد برای ورودی‌ها در دوره L چنانچه تکنولوژی مرجع در دوره K محاسبه شده است.

در حالت کلی شاخص بهره وری مالم کوییست(MPI) به دو مؤلفه تجزیه می‌شود:

تغییرات کارایی تکنیکی[۱۸]

و تغییر مرز تولید کارایی[۱۹]

که در واقع نماد اصلی را با مؤلفه‌های MPI در معادله(۳) ترکیب می‌کنیم. هم هدف کاهش α و هم متغیر پنهان λ، ضرایب کاهشی برای این روش تجزیه هستند. وقتی از یک LVM متراکم ضرایب کاهش همبستگی نسبی ناشی می‌شود، که به نوبه خود تا به حداقل رساندن واریانس در این مدل کمک می‌کند.

[۱۸] Technical Efficiency Change
[۱۹] Efficient Production Frontier shift

به طور کلی معرفی تکنولوژی متغیر پنهان تحلیل پوششی داده‌ها یک گام نخست به سوی نه تنها تحلیل خروجی‌های نامطلوب بلکه برای گنجاندن پیامدهای خارجی بر روی شرکت‌ها و جامعه است.

این مدل با استفاده از مدل کاهش متغیر پنهان متراکم، تئوری تولید کاهش همزمان خروجی‌های نامطلوب و ورودی‌ها را از طریق رابطه علی با ورودی‌ها نشان می‌دهد.

نتایج بهره‌وریDMUها از رابطه(۵) به‌دست می‌آید، چنانچه MPI>۱ (progress) نشان‌دهنده روند پیشرفت در طی دوره و اگر MPI=۱ (No change) نشان‌دهنده عدم تغییر و اگر MPI<۱ (Regress) نشان‌دهنده پسرفت در طی دوره است.

۳-۴- ترکیب DEA با تکنیک داده‌کاوی

محققان DEA در گذشته عمدتاً بر روی ارزیابی و کنترل عملکرد متمرکز شده‌اند و تنها تلاش اندکی از ترکیب DEA با تکنیک داده‌کاوی گزارش شده است.

برای مثال، سون و مون[۲۰](۲۰۰۴) امکان استفاده از درخت تصمیم‌گیری را با DEA برای پروژه‌های R&D مورد بررسی قرار دادند، وقتی که یک شرکت سعی بر انتقال یا گسترش تکنولوژی جدید دارد.

مطالعات بسیار اندکی برای DEA برای اهداف پیشگویانه به کار رفته است از جمله، پیش‌بینی شکست بانک (بار و همکاران، ۱۹۹۴) و پیش بینی شکست اتحادیه اعتباری (پایل و پاردی، ۱۹۹۷). اما هیچ مطالعه‌ای پیرامون DEA و درخت C&R برای کارایی و بهره وری شعب بیمه گزارش نشده است.

۳-۵- طبقه بندی و درخت رگرسیون(C&R)

تکنیک داده کاوی به DMUها اجازه کشف وکاوش معنادار اطلاعاتی که قبلاً در پایگاه داده های بزرگ پنهان بوده‌اند را می‌دهد.

C&Rدرخت تصمیم‌گیری است که به طور معمول در داده کاوی مورد استفاده قرار می‌گیرد، به وسیله برایمن و همکاران در سال ۱۹۸۴ توسعه داده شده است و سپس توسط رایپلی در سال ۱۹۹۶بهبود یافت. ساختار درختی نشان‌دهنده مسأله تصمیم داده شده است به طوری که هریک از گره‌های غیر برگ با یکی از متغیرهای تصمیم در ارتباط است، هر شاخه از گره غیر برگ با زیرمجموعه‌ای از مقادیر متغیر تصمیم‌گیری مربوطه در ارتباط است، و هر گره برگ با یک مقدار از متغیر هدف(متغیر وابسته) وابسته شده است.

برای هر برگ درخت مرتبط با مقدار میانگین متغیر هدف است، بنابراین یک درخت، یک روش جایگزین برای مدل‌های خطی پیوسته برای مشکلات رگرسیون و مدل‌های خطی لجستیک برای مسائل طبقه بندی است.

به طور کلی، درختان C&R برخی مزایا بیش از مدل های رگرسیون دارند. اولاً، مدل ایجاد شده توسط یک درخت قابل فهم‌تر و نسبتاً ساده برای تفسیر غیرآماری است [۲۲].

ثانیاً، ذاتاً غیر پارامتری است یعنی بدون نیاز به هیچ فرضی با توجه به توزیع اساسی از مقادیر متغیرهای مستقل ساخته شده است. بنابراین، درختان C&R می‌توانند داده‌های عددی که چولگی زیاد و یا چند وجهی هستند و همچنین پیش‌بینی های رده‌ای با هر دو ساختار ترتیبی و غیرترتیبی را اداره کند.

ثالثاً، درختان C&R روش‌های پیچیده‌ای برای مقابله با متغیرهای از دست رفته در مقایسه با رگرسیون دارد.

در رگرسیون، داده‌هایی که دارای هر گونه مقدار از دست رفته باشند به طور خودکار حذف می کنیم.

بنابراین، درختان C&R می‌توانند حتی هنگامی که متغیرهای مستقل مهم برای برخی ازDMUها ناشناخته است ایجاد شوند.

در نهایت درختان C&R، نسبتاً یک روش یادگیری ماشین خودکار است. درختان C&R، کارایی محاسباتی ارائه می‌کنند تا جایی‌که آنها زمان کمتری را درمحاسبات و ذخیره‌سازی کمتری نیاز داشته باشند.

در تولید یک درخت C&R، مجموعه داده‌ها حداقل به دو بخش تقسیم می‌شوند: مجموعه داده‌های آموزشی و مجموعه داده‌های آزمایشی (عموماً به عنوان مجموعه داده‌های تست ارجاع داده می‌شود) [۲۲]. سپس تحت دو فاز فرایند اصلی: مرحله رشد و مرحله هرس قرار می‌گیرد [۲۶].

[۲۰] Sohn & Moon

در مرحله رشد C&R، یک درخت از مجموعه داده‌های آموزشی ساخته می‌شود. در این مرحله، هر گره برگ با یک کلاس در ارتباط است با تقسیم‌بندی بیشتری از برگ مفروض در تعدادی از موارد به یک یا دو گره بعدی که در زیر بعضی آستانه مشخص شده، منجر می‌شود.

دراین مرحله، درخت C&R برخلاف صحت یا مجموعه داده‌های آموزشی برطبق تولید یک زیر درخت با کمترین نرخ خطای مجموعه داده‌ها آزمایشی ارزیابی شده است. برای الگوریتم دقیق درختان C&R، هند و همکاران(۲۰۰۱) را می‌بینیم.

در این مقاله، از تحلیل درخت C&R برای کشف و بررسی عوامل درونی و بیرونی همچون تعداد خسارت‌های پرداختی، تعداد و تحصیلات کارکنان، درجه شعب، قدمت شعب، تعداد وام‌ها، تعداد بیمه نامه‌ها و...که برروی بهره‌وری شعب بیمه مؤثرند، استفاده می‌شود.

مقدار هدف برای درخت رتبه‌کارایی و بهره‌وری به دست آمده از DEA است. بنابراین، DMUها به دو دسته گروه‌های پیشرفت بهره‌وری و پسرفت بهره‌وری تقسیم می‌شود و بنابراین هدف درخت ما یک متغیر گسسته (طبقه‌ای) است.

۳-۶- ترکیب DEA با درخت C&R

درختC&R که در این مقاله پیشنهاد شده است، شامل چهارمؤلفه اصلی است.

مؤلفه اول، متغیر خروجی یا متغیر وابسته است. به طور کلی این متغیر مشخصه‌ای است که امیدواریم بتوانیم بر اساس متغیرهای مستقل یا پیش‌بینی کننده‌ها، پیش‌بینی کنیم. در این مطالعه متغیر خروجی همان نمرات به دست آمده از بهره‌وری است که به سه دسته پیشرفت بهره‌وری (۱<هدف)، پسرفت (۱>هدف)، بدون تغییر (۱=هدف) دسته‌بندی می‌شود.

مؤلفه دوم از درخت C&R، متغیرهای مستقل یا پیش‌بینی کننده است، تعدادی متغیرهای مستقل ممکن بسته به هدف از تجزیه و تحلیل وجود دارد.

در این مطالعه موردی، متغیرهای مستقل عوامل درونی و بیرونی هستند که در جدول۱ ذکر شده است.

سومین مؤلفه از یک درخت C&R مجموعه داده آموزش است. این یک پایگاه داده شامل مقادیر متغیرهایی برای هم متغیر خروجی و هم متغیرهای پیش‌بینی‌کننده (مستقل)، از یک گروه از DMUهایی که می‌خواهیم نتایج خروجی را پیش‌بینی کنیم.

مؤلفه چهارم از درخت C&R، آزمون یا مجموعه داده‌های بیشتر که متشکل از DMUهایی است که ما خواهان پیشگویی دقیق آنها هستیم. این مجموعه داده‌های آزمون ممکن است در عمل وجود نداشته باشد در حالی که به طور معمول براین باوریم که یک مجموعه داده تست مورد نیاز است تا یک طبقه یا قوانین تصمیم را بدان تنفیذ کنیم، یک پایگاه داده آزمایش جداگانه همیشه نیاز به تعیین عملکرد یک قوانین تصمیم ندارد.

شکل۱ مراحل ارزیابی کارایی و بهره‌وری شعب بیمه با استفاده از DEA/C&R را نشان می‌دهد. همانطور که در شکل مشاهده می‌شود، ابتدا از DEA برای اندازه‌گیری کارایی و بهره‌وری هر شعبه با ۳ ورودی (هزینه‌های اداری، هزینه‌های بیمه‌گری، تعداد نمایندگی‌ها) و ۳ خروجی (درآمد حاصل از حق بیمه، ارزش وام‌های پرداختی، ارزش خسارت پرداختی) استفاده شده است، به عنوان نتایج این مرحله را به سه گروه شعب کارآمد، ناکارآمد و بدون تغییر تقسیم می‌کنیم.

بنابراین در مرحله دوم تحلیل درخت C&R از عوامل محیطی وابسته به سیستم بیمه همچون قدمت شعب، درجه شعب، تعداد بیمه‌نامه‌های صادره و.... به عنوان عوامل ورودی به تحلیل درخت C&R و نمرات بهره‌وری بدست آمده در مرحله نخست به عنوان خروجی درخت C&R (برای لیست کامل متغیرها به جدول ۱مراجعه کنید).

واضح است که، این یک چارچوب کلی می‌تواند برای هرگونه تجزیه و تحلیل در بخش بیمه و بانک و یا بخش‌های دیگر کاربرد داشته باشد.

در مرحله اول این روش، نمرات بهره‌وری DMUها به عنوان متغیر هدف (مستقل) وارد درخت رگرسیون می‌شوند. سپس عوامل محیطی در مرحله دوم به عنوان متغیرهای پیش بینی(وابسته)از یک طرف و نتایج بهره وری در مرحله اول از طرف دیگر وارد درخت رگرسیون می‌شوند و در نهایت

مجموعه قوانینی از درخت استخراج می‌شود.

شکل۱: DEA/C&R متدلوژی برای ارزیابی شعب بیمه ایران

جدول۱: عوامل ورودی در درخت تصمیم(۱۳۸۹-۱۳۸۷)

[۱]درجه شعبه:۱، مجتمع ؛ ۲، ممتاز ؛ ۳، درجه۱

[۲]موقعیت جغرافیایی شعبه:۱،شمال؛ ۲، جنوب؛ ۳، شرق؛ ۴، غرب؛ ۵، مرکز

[۳]تحصیلات کارکنان:۱، زیردیپلم؛ ۲، دیپلم؛ ۳، فوق دیپلم؛ ۴، لیسانس؛ ۵، فوق لیسانس؛ ۶، دکترا

[۴]جنسیت کارکنان:۱، زن؛ ۲، مرد

۴. یافته‌های پژوهش

۴-۱- روش بوت استرپینگ DEA/C&R برای ارزیابی شعب بیمه

یکی از مشکلات استفاده از DEA/C&R این است که در بسیاری از مطالعات DEA، داده‌های کافی برای تولید درخت تصمیم در دسترس وجود ندارد. بنابراین، روش Bootstrapping زیر برای افزایش تعداد DMUها قبل از ایجاد درخت تصمیم C&R پیشنهاد شده است [۱۵].

این روش از سه مرحله تشکیل شده است، مرحله اول: مقادیر کارایی و بهره‌وری را برای هر شعبه محاسبه نموده و بر طبق این نمرات شعب را به سه گروه پیشرفت بهره‌وری (Progress) (MPI>۱⟵Target>۱)، پسرفت بهره‌وری (Regress) (MPI<۱ ⟵ Target<۱) و بدون تغییر (No change)(MPI=۱ Target=۱) تقسیم می‌شوند.

یک درخت C&R دقیق، به یک پایگاه داده بزرگ نیاز دارد؛ بنابراین چون نمونه ما تنها شامل ۱۸ شعبه بیمه است، پس با افزایش پایگاه داده اصلی با۱۰۰بار استفاده از تکنیک نمونه‌برداری مجدد بوت استرپینگ که توسط افرون و تیبشیرانی توصیف شد است، استفاده می‌شود [۱۶].

بنابراین در مرحله دوم، به طور تصادفی ۱۸واحد (با جایگزینی) انتخاب می‌کنیم و این نمونه‌ها را ۱۰۰ بار تا به دست آوردن ۱۸۰۰واحد تکرار می‌کنیم. بعد از نمونه‌برداری مجدد[۲۱]۱۰۰بار پایگاه داده اصلی، پایگاه داده به دو گروه آموزشی و آزمایشی به نسبت ۷ به ۳ تقسیم می‌شود.

درمرحله سوم با استفاده از نمرات بهره‌وری طبقه‌بندی شده (۱> ،۱=،۱<) به عنوان متغیرهای هدف درخت C&R و دیگر متغیرهای غیرقابل کنترل مانند درجه شعب، موقعیت جغرافیایی، تعداد کارکنان و... به عنوان ورودی درخت تصمیم قوانین منطقی استخراج می‌شود.

جدول۲: متغیرهای ورودی/خروجی در مدلDEA

[۲۱] Re-sampling

در طی بررسی روند بهره‌وری در سال‌های ۱۳۸۸-۱۳۸۷، ۵ واحد روند پیشرفت و ۱۳ واحد روند پسرفت داشته‌اند و در طی سال‌های ۱۳۸۹-۱۳۸۸، ۸ واحد بهره‌ور و ۱۰ واحد عدم بهره‌وری را نشان می‌دهند.

به طور میانگین، ۳۶/۲٪ واحدها دارای پیشرفت بهره‌وری و۶۶/۲٪ واحدها پسرفت بهره‌وری را نشان می‌دهند. البته به دلیل پراکندگی بالای مقادیر داده‌های ورودی درخت تمام مقادیر قبل از ورود به درخت نرمال شده و سپس وارد درخت شده است تا سبب کاهش دقت پیش‌بینی نشود.

- مرحله دوم: بوت استرپینگ (Bootstrapping)

همان‌طور که در بخش قبل به طور کامل توضیح داده شد، با انتخاب تصادفی ۱۸ واحد با جایگذاری و تکرار این نمونه‌ها ۱۰۰بار یک پایگاه داده با ۱۸۰۰ واحد به دست آمد. که این اطمینان را برای رسیدن به دقت بالاتری در پیش‌بینی درخت C&R به ما خواهد داد.

- مرحله سوم: تحلیل C&R

بر طبق ارزیابی DEA شعب بیمه به سه گروه، پیشرفت بهره‌وری (MI>۱)، پسرفت بهره‌وری (MI<۱) و بدون تغییر بهره‌وری (MI=۱) تقسیم می‌شود. این گروه‌ها، به عنوان متغیر هدف در درخت C&R مورد استفاده قرار می‌گیرند نمرات کارایی و بهره‌وری شعب با شاخص مالمکوییست در جدول۳ نشان داده شده است.

جدول۳: نمرات بهره وری به دست آمده از مدلLVM با شاخص مالمکوییست

واحدها	از سال	تا سال	TC	EC	MI
DMU۱	۱۳۸۷	۱۳۸۸	۰,۰۱۲۹	۱۱,۷۲۳۶	۰,۱۵۱۲
DMU۱	۱۳۸۸	۱۳۸۹	۲,۴۰۷۷	۰,۱۱۱۱	۰,۲۶۷۵
DMU۲	۱۳۸۷	۱۳۸۸	۰,۰۰۶۳	۴۰,۴۵۳۱	۰,۲۵۴۹
DMU۲	۱۳۸۸	۱۳۸۹	۰,۹۰۶۷	۰,۴۰۹۷	۰,۳۷۱۵
DMU۳	۱۳۸۷	۱۳۸۸	۰,۰۱۲۲	۱۳۲,۲۴۶۹	۱,۶۱۳۴
DMU۳	۱۳۸۸	۱۳۸۹	۰,۹۰۷۵	۰,۲۹۹۷	۰,۲۷۲
DMU٤	۱۳۸۷	۱۳۸۸	۰,۰۱۲۳	۳,۰۹۴۶	۰,۰۳۸۱
DMU٤	۱۳۸۸	۱۳۸۹	۱,۰۵۱۱	۰,۴۴۸۳	۰,۴۷۱۲
DMU٥	۱۳۸۷	۱۳۸۸	۰,۰۱۰۳	۲۱,۲۲۴۳	۰,۲۱۸۶
DMU٥	۱۳۸۸	۱۳۸۹	۰,۹۴۶۱	۰,۲۵۶۵	۰,۲۴۲۷
DMU٦	۱۳۸۷	۱۳۸۸	۰,۰۰۶۳	۱۱,۹۶۰۶	۰,۰۷۵۴
DMU٦	۱۳۸۸	۱۳۸۹	۲,۵۴۶	۰,۶۳۷	۱,۶۲۲۷
DMU۷	۱۳۸۷	۱۳۸۸	۰,۰۰۸۵	۴,۵۷۹۲	۰,۰۳۸۹
DMU۷	۱۳۸۸	۱۳۸۹	۱,۰۸۳۹	۰,۲۳۲۸	۰,۲۵۲۳
DMU۸	۱۳۸۷	۱۳۸۸	۰,۰۰۵	۱۰,۰۳۵۴	۰,۰۵۰۲
DMU۸	۱۳۸۸	۱۳۸۹	۲,۲۵۵	۰,۷۹۶۴	۱,۷۹۵۳
DMU۹	۱۳۸۷	۱۳۸۸	۰,۰۰۵	۳,۹۵۲۵	۰,۰۱۹۸
DMU۹	۱۳۸۸	۱۳۸۹	۰,۹۵۳۳	۰,۵۳۴۷	۰,۵۰۹۷
DMU۱۰	۱۳۸۷	۱۳۸۸	۰,۰۰۸	۱۱,۷۱۷۷	۰,۰۹۳۷
DMU۱۰	۱۳۸۸	۱۳۸۹	۱,۶۰۹۲	۰,۲۶۴۶	۰,۴۲۵۸
DMU۱۱	۱۳۸۷	۱۳۸۸	۰,۰۰۷۴	۰,۴۹	۰,۰۰۳۶
DMU۱۱	۱۳۸۸	۱۳۸۹	۱,۰۳۶۲	۰,۲۵۷۳	۰,۲۶۶۶
DMU۱۲	۱۳۸۷	۱۳۸۸	۰,۰۱۱۸	۱,۱۷۷۴	۰,۰۱۳۹
DMU۱۲	۱۳۸۸	۱۳۸۹	۱,۲۴۲۱	۱,۳۸۶۷	۱,۷۲۲۵
DMU۱۳	۱۳۸۷	۱۳۸۸	۰,۰۰۵۵	۰,۲۵۹۹	۰,۰۰۱۴
DMU۱۳	۱۳۸۸	۱۳۸۹	۱,۱۹۵۵	۱,۲۰۵۹	۱,۴۴۱۷
DMU۱٤	۱۳۸۷	۱۳۸۸	۰,۰۱۱۴	۱۴,۶۹۷۷	۰,۱۶۷۶
DMU۱٤	۱۳۸۸	۱۳۸۹	۱,۲۴۴۵	۱,۲۸۸	۱,۶۰۲۳
DMU۱٥	۱۳۸۷	۱۳۸۸	۰,۰۱۰۷	۱۱,۰۴۸	۱,۱۸۲
DMU۱٥	۱۳۸۸	۱۳۸۹	۱,۳۷۹	۰,۳۳۱۷	۰,۴۵۷۴
DMU۱٦	۱۳۸۷	۱۳۸۸	۰,۰۱۰۳	۹۷۵,۷۳۹۶۸	۱۰,۵۰۱۲
DMU۱٦	۱۳۸۸	۱۳۸۹	۳,۸۹۵۲	۲,۰۶۳۵	۸,۰۳۷۷
DMU۱۷	۱۳۸۷	۱۳۸۸	۰,۰۰۸۲	۳۲۴,۸۱۸۱۸	۲,۶۶۳۵۱
DMU۱۷	۱۳۸۸	۱۳۸۹	۳,۰۸۳۵	۲,۰۶۳۵	۶,۳۶۲۸
DMU۱۸	۱۳۸۷	۱۳۸۸	۰,۰۱	۲۰۳,۱۴۷	۲,۰۳۱۴۷
DMU۱۸	۱۳۸۸	۱۳۸۹	۳,۰۲۹۹	۰,۳۹۳۹	۱,۱۹۳۵

۵. بحث و نتیجه‌گیری

با توجه به انتخاب درخت C&R، با شاخصی به نام (Impurity) بر حسب کمترین مقدار ناخالصی در هر گره، برای تمام خصیصه‌ها[۲۲] میزان ناخالصی را قبل و بعد از شاخه‌زدن محاسبه نموده و خصیصه‌ای که ناخالصی را بیشتر کاهش دهد را انتخاب می‌کند. به این ترتیب چندین درخت تصمیم‌گیری رگرسیون، برای هر دوره رسم می‌شود.

۵-۱- تحلیل نتایج درخت رگرسیون طی دوره (۱۳۸۷-۱۳۸۸)

ابتدا درخت پیش‌بینی را برای سال (۱۳۸۸-۱۳۸۷) با متغیرهای، درجه شعبه، قدمت، موقعیت مکانی، تعداد کارمندان فوق لیسانس و دکترا، تعداد وام‌های این دوره، تعداد بیمه‌نامه‌های صادر شده در این دوره، تعداد خسارت‌های پرداختی در این دوره، به عنوان ورودی و طبقه‌بندی بهره‌وری به عنوان خروجی درخت در نظر گرفته می‌شود. توجه داشته باشید که یکی از ویژگی‌های ذاتی درخت، حذف برخی از ویژگی‌ها برحسب اهمیت یا حداقل همبستگی است، بنابراین درخت تنها با متغیرهای تعداد بیمه‌نامه‌های صادره، قدمت و موقعیت مکانی به عنوان متغیرهای محیطی وارد می‌کنیم.

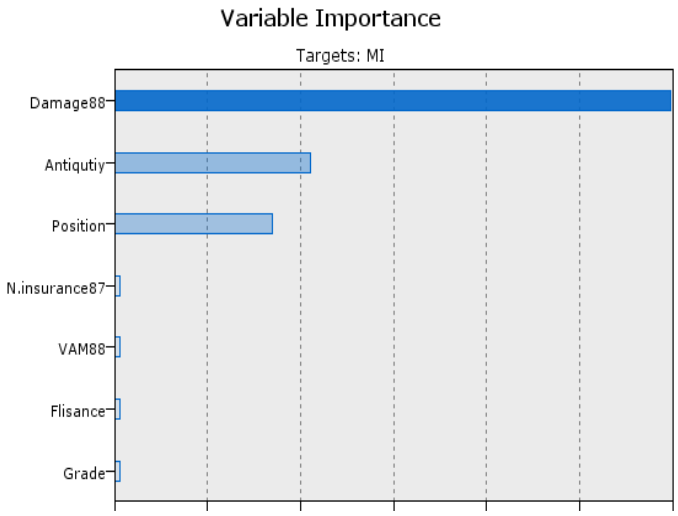
شکل۲، اهمیت متغیرهای محیطی را نشان می‌دهد. در این شکل مشاهده می‌کنیم، که تعداد خسارت‌های پرداختی دارای بیشترین اهمیت در تعیین طبقه‌بندی را دارد (۵۹٪) و قدمت با (۲۱٪) و موقعیت جغرافیایی (۰,۱۷٪) در جایگاه دوم و سوم و بقیه متغیرها با(۰,۵٪) در مقام‌های بعدی اهمیت قرار دارند، اما چون متغیرهای بعدی در جایگاه یکسان اهمیت قرار دارند تا رسیدن به نتیجه مطلوب و دقت بیشتر درخت را با متغیرهای هم‌رده دوباره می‌سازیم.

جدول۶-۳ دقت پیش‌بینی درخت تولید شده را نشان می‌دهد. به واسطه ۱۸۰۰مورد:

۱۲۹۰مورد در طی سال۱۸۷-۱۳۸۸، شاخص بهره‌وری زیر یک و۵۱۰ مورد بالای یک را پیش‌بینی می‌کند و از ۱۲۴۶ مورد داده برای آزمایش (Training) داده‌ها صورت گرفته است، ۱۱۵۶مورد به صورت صحیح پیش‌بینی شده است و با دقت ۹۲,۷۸٪ این پیش‌بینی همراه است و۹۰ مورد به اشتباه پیش بینی شده است.

سطح دقت کلی درخت پیش‌بینی C&R،۹۸,۰۲٪ است، که سطح اطمینان بالایی را نشان می‌دهد.

در شکل۳درخت تولید شده C&R، با ۸ گره نمایش داده شده است.



شکل۲: اهمیت متغیرهای محیطی درخت (۱۳۸۷-۱۳۸۸)

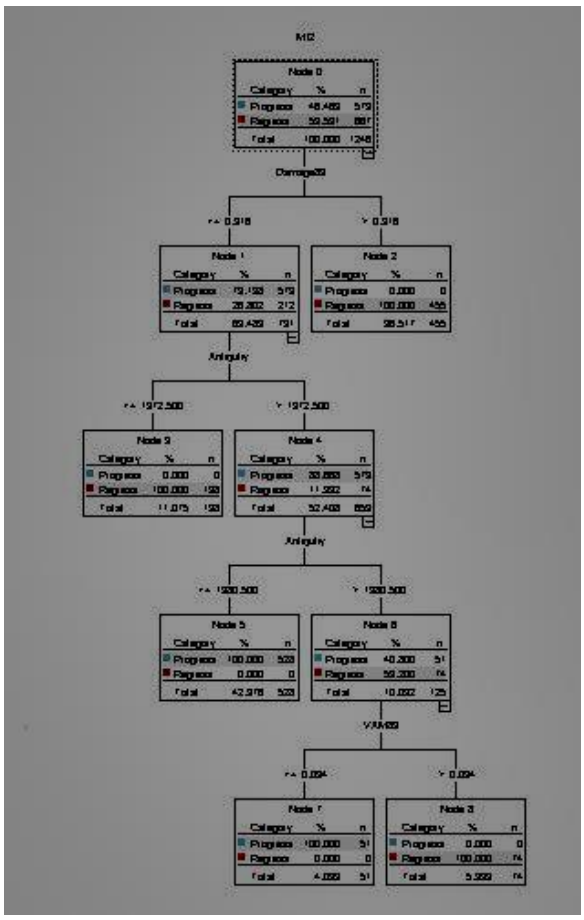
جدول۴: دقت پیش‌بینی درخت تولید شده (شکل۶-۲)

Results for output field MI

Comparing $R-MI with MI

'Partition'	1_Training		2_Testing	
Correct	1,226	98.39%	541	97.65%
Wrong	20	1.61%	13	2.35%
Total	1,246		554	

[۲۲] Attributes



شکل۳: درخت تولید شده با متغیرهای محیطی (۱۳۸۸-۱۳۸۷)

بر طبق این درخت قوانین زیر استخراج می‌شود:

استخراج قوانین برای شعب با پیشرفت بهره‌وری

شعب با روند پیشرفت بهره‌وری (در مجموع ۳۲۵ مورد از۱۲۴۶)، اگر قوانین به صورت زیر دنبال شود:

- قانون۱: اگر تعداد خسارت پرداختی کوچکتر مساوی ۰,۰۲۰باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است (Progress)؛ (۱۹۲ مورد).
- قانون۲: اگر تعدادخسارت پرداختی هر شعبه بزرگتر از ۰,۰۲۰و قدمت تأسیس مربوط به کمتر مساوی از سال۱۳۸۰ باشد و موقعیت مکانی در غرب شهر تهران باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است. (Progress)؛ (۵۰مورد).
- قانون۳: اگر تعداد خسارت پرداختی شعب بیشتر از ۰,۰۲۰ و قدمت تأسیس مربوط به بیشتر از سال ۱۳۸۰ باشد، آنگاه شعبه دارای پیشرفت بهره‌وری است (Progress) ؛ (۷۲ مورد).

۵-۲- استخراج قوانین برای شعب با پسرفت بهره‌وری

شعب با روند پسرفت بهره‌وری (در مجموع ۹۱۴ مورد از۱۲۴۶)، اگر قوانین به صورت زیر دنبال شود:

- قانون۴: اگر تعداد خسارت پرداختی بزرگتر مساوی ۰,۰۲۰ و قدمت شعبه کمتر مساوی سال۱۳۸۰ و موقعیت مکانی شعبه (مرکز، شرق، شمال وجنوب) باشد، آنگاه شعبه دارای روند پسرفت بهره‌وری است (Regress) ؛ (۹۱۲ مورد).

۵-۳- تحلیل نتایج درخت رگرسیون طی دوره (۱۳۸۸-۱۳۸۹)

جدول۵ دقت پیش‌بینی درخت تولید شده را نشان می‌دهد. به واسطه ۱۸۰۰مورد:

۹۹۷ مورد درطی سال۱۳۸۹-۱۳۸۸، شاخص بهره‌وری زیر یک و۸۰۳ مورد بالای یک را پیش‌بینی می‌کند و از ۱۲۴۶ مورد داده برای آزمایش (Training) داده‌ها صورت گرفته است، کل ۱۲۴۶ مورد به صورت صحیح پیش‌بینی شده است و با دقت۱۰۰٪ این پیش‌بینی همراه است.

سطح دقت کلی درخت پیش‌بینی C&R، ۱۰۰٪ است، که سطح اطمینان بالایی را نشان می‌دهد.

در شکل۵ درخت تولید شده C&R، با ۵۷۹ مورد پیشرفت بهره‌وری و ۶۶۷ مورد کاهش بهره‌وری، با ۸ گره نمایش داده شده است.

با توجه به نمودار تشکیل شده در شکل۴ که اهمیت متغیرهای ورودی را در درخت رگرسیون نشان می‌دهد، تعداد خسارات پرداختی در سال۱۳۸۹با (۶۷٪) در جایگاه اول و قدمت شعبه با (۳۰٪) اهمیت در جایگاه دوم و بقیه متغیرها (تعداد کارمندان با تحصیلات دکترا و فوق لیسانس، درجه شعبه و تعداد بیمه‌نامه‌های صادره در سال۸۹) با (۶٪) در جایگاه بعدی با اهمیت یکسان قرار دارند.

شکل۴: اهمیت متغیرهای محیطی در درخت ایجاد شده

جدول۵: دقت پیش‌بینی درخت شکل ۵

شکل۵: درخت ایجادشده با متغیرهای اهمیت جدول۴

۵-۴- استخراج قوانین برای شعب با پیشرفت بهره‌وری (۵۷۹ مورد)

- قانون۱: اگر تعداد خسارات در سال۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس مربوط به سال‌های بعد ۱۳۷۲ و قبل ۱۳۸۰ باشد، آنگاه شعبه دارای روند پیشرفت بهره‌وری است (Progress)؛ (۵۲۸ مورد).
- قانون۲: اگر تعداد خسارت در سال ۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس مربوط به سال‌های بعد ۱۳۸۰ باشد و تعداد وام‌های پرداختی کمتر مساوی۰,۰۳۴ باشد، شعبه دارای پیشرفت بهره‌وری است (Progress) (۵۱ مورد).

۵-۵- استخراج قوانین برای شعب با پسرفت بهره‌وری(۶۶۷ مورد)

- قانون۳: اگر تعداد خسارت پرداختی سال۱۳۸۹ کمتر مساوی ۰,۳۱۶ باشد و قدمت تأسیس شعبه مربوط به سال‌های قبل از ۱۳۷۲ باشد، شعبه دارای روند کاهشی یا پسرفت بهره‌وری است (Regress)؛ (۱۳۸مورد).
- قانون۴: اگر تعداد خسارت پرداختی سال۱۳۸۹ کمتر مساوی ۰,۳۱۶ و قدمت تأسیس شعبه مربوط به سال‌های بعد از۱۳۷۲ باشد وتعداد وام‌های پرداختی بیشتر از ۰,۰۳۴باشد، شعبه دارای روند کاهشی یا پسرفت بهره‌وری است (Regress)؛ (۷۴ مورد).
- قانون۵: اگر تعداد خسارت پرداختی بیش از ۰,۳۱۶ پس شعبه دارای روند پسرفت بهره‌وری است (Regress)؛ (۴۵۵ مورد).

۵-۶- ارزیابی نهایی

به طورخلاصه این مقاله، به معرفی ترکیب DEA و رویکرد درخت C&R می‌پردازد. مطالعه موردی ما در این مطالعه شعب بیمه ایران در شهر تهران می باشد.

به طور کلی، تحلیل پوششی داده‌ها نمرات کارایی و با استفاده از شاخص مالمکوییست، نمرات بهره‌وری را می‌توان محاسبه نمود، اما این روش نمی‌تواند عوامل مربوط به ناکارایی و عدم بهره‌وری را توضیح دهد، خصوصاً اگر این عوامل در شکل متغیرهای غیرعددی مانند درجه شعب باشند.

استفاده از درختC&R می‌تواند درک بهتری از نتایج به دست آمده از تحلیل پوششی داده‌ها را نشان دهد و با بررسی عوامل مربوط به کارایی و بهره‌وری و عدم بهره‌وری که یک مسئله مهم برای سیاست گزاران است، در نظر گرفته شود.

همچنین ما روش پیشنهادی را در صنعت بیمه مورد مطالعه قرار داده، اما روش پیشنهادی در این مطالعه به طور بالقوه دارای کاربردهای بسیار گسترده‌تری است.

روش پیشنهادی DEA/C&R یک چارچوب برای هر مطالعه دیگری است که می تواند برای ارزیابی کارایی و بهره وری واحدهای تصمیم گیرنده(DMUs) مورد استفاده قرار گیرد.

نتایج مدل ترکیبی، یک مجموعه قوانینی است که می‌تواند توسط سیاست‌گزاران برای کشف دلایل کارایی و ناکارایی واحدهای تصمیم‌گیرنده مورد استفاده قرار گیرد. اگرچه، برخی از محدودیت‌ها، زمانی‌که DEA با C&R ترکیب می‌شود، وجود دارد.

ایجاد یک درخت خوب و قابل اطمینان C&R، معمولاً به یک پایگاه داده بزرگ با مشاهدات زیادی نیازمند است.

در اغلب گزارشات DEA در ادبیات موضوع، تعداد واحدهای تصمیم‌گیرنده (DMUs) به اندازه کافی بزرگ نیست که بتواند یک درختC&R را تغذیه کند.

در این مطالعه، روش بوت استرپینگ (Bootstrapping) را برای رفع این مشکل پیشنهاد کردیم؛ با این حال به نظر می‌رسد که این موضوع نیاز به تحقیقات بیشتری برای برنامه‌های کاربردی دارد.

۵-۷- نتیجه گیری

تحلیل پوششی داده‌ها (DEA)، یک ابزار مدیریتی برای ارزیابی کارایی و بهره‌وری است. این مقاله به معرفی یک چارچوب ترکیب DEA با طبقه‌بندی و تحلیل رگرسیون می‌پردازد.

در حالی که، DEA نتایج قابل قبول و ارزشمندی ارایه می‌دهد، از طرفی تحلیل C&R، مبتنی برتحلیل یافته‌های حقایقی اضافی که در مطالعات قبلی نامشخص بودند را برای ما آشکار می‌کند. برخلاف مطالعات قبل در حوزه تحلیل پوششی داده‌ها وصنعت بیمه که به شناسایی تأثیر یکسانی

از عوامل بر روی کارایی می‌پردازد، درختC&R، مبتنی بر تحلیل پیشنهادهایی است که سطح[۲۳] تأثیر عوامل وابسته به کارایی و بهره‌وری را بر روی شعب نشان می‌دهد.

با بررسی اهمیت متغیرها با نرم‌افزار کلمنتاین[۲۴] و تأثیر قدرت آنها در وابستگی متغیرها با کمترین میزان ناخالصی تا رسیدن به گره هدف، این روند می‌تواند به تفکیک دقیق عوامل محیطی با نتایج کارایی و بهره‌وری شعب که از مرحله قبل و با روش DEA به دست آمده، تحلیلی قابل تعمق و با کمترین میزان خطا را در پی داشته باشد.

با بررسی مطالعات گذشته بر روی صنعت بیمه، تاکنون تنها به ارزیابی شعب بر اساس شاخص‌های مؤثر و تنها به تحلیل کارایی یا ناکارایی و یا بهره‌وری و عدم بهره‌وری در طی دوره پرداخته شده است و به کشف عوامل محیطی مرتبط با پیشرفت بهره‌وری(Progress) و یا پسرفت بهره‌وری (Regress) پرداخته نشده است.

به عنوان مثال، با این که تعداد خسارت و تعداد وام‌های پرداختی و قدمت تأسیس شعبه به عنوان شاخص مهم بیمه در روند کارایی و ناکارایی محسوب نمی‌شود، اما همین طور که از قوانین استخراج شده استنباط می‌کنیم، به عنوان عوامل وابسته در این روند با سطح اهمیت متفاوت دخیل هستند.

از طرف دیگر، توانستیم به استخراج قوانین برای هر واحد تصمیم‌گیرنده (DMU) با استفاده از متغیرهای عددی[۲۵]و رده‌ای[۲۶] که دارای درصدی از اهمیت بودند، بپردازیم و برای شناسایی بهره‌وری و عدم بهره‌وری در انتخاب شعب استفاده کنیم.

برخلاف مطالعات قبلی کاربردهای DEA، که تمرکز فقط در زمینه عددی برای محاسبه کارایی و بهره‌وری تصمیم‌گیری می‌شد، در این مطالعه با استفاده از درخت C&R، به توسعه عوامل مرتبط با بهره‌وری و عدم بهره‌وری می‌پردازد و هرقانونی که می‌تواند برای یک واحد بهره‌ور به دست آید با استفاده از متغیرهای عددی و - متغیرهای قیاسی را استخراج کند. بدیهی است که این قوانین برای سیاست‌گزاران بیمه بسیار مفید و روند تصمیم‌گیری آنان را تحت‌الشعاع قرار می‌دهد.

۵-۸- پیشنهادات آتی

تعدادی موضوعات فراتر از محدوده تحقیق ما وجود دارد که از لحاظ اهمیت کاربردی برای اشخاصی که با درخت C&R به تحلیل و ارزیابی موضوعات می‌پردازند، بسیار حائز اهمیت است. این عوامل عبارتند از فاکتورهای مستقل برای بخش‌های بیمه و استفاده از قوانین مختلف و اندازه‌گیری دقیق همچنین بهبود روش بوت استرپینگ، این موارد می‌تواند زمینه را برای توسعه مطالعات آینده گسترده‌تر کند.

در تحقیقات آتی می‌توان، پایگاه داده‌ای با رکوردهای بالا انتخاب نمود که مجبور به استفاده از بوت‌استرپینگ برای تولید پایگاه داده معتبر نبوده وبه راحتی درختی با دقت بالا داشته باشیم، البته استفاده از شبیه‌سازی داده‌ها در این تحقیق جزء محدودیت‌های تحقیق در زمینه دسترسی به داده‌های مورد نظر بوده است.

همچنین می‌توان از درخت تصمیم فازی به جای درخت تصمیم در حالت دقیق[۲۷] استفاده نمود که برای داده‌های کیفی صنعت بیمه، نتایج مفیدی را ارائه می‌دهد و یا خروجی‌های درخت را نتایج کارایی و ناکارایی مدل DEA قرار داد و یا بسته به نوع داده‌ها و اهمیت متغیرهای ورودی از درختان دیگری چون $C5$،$C4.5$ ،$CART$و.. استفاده نمود.

منابع

۱. آذر عادل؛ مؤمنی منصور. ۱۳۸۳. اندازه‌گیری بهره‌وری در شرکت‌های تولیدی به وسیله مدل‌های تحلیل پوششی داده‌ها، دو ماهنامه علمی پژوهشی دانشور رفتار، دانشگاه شاهد، سال یازدهم، شماره ۸.

۲. امامی میبدی علی. ۱۳۸۴. اصول و اندازه‌گیری کارایی و بهره‌وری (علمی- کاربردی)، مؤسسه مطالعات پژوهش‌های بازرگانی، چاپ دوم، تهران.

۳. اسفاندرانی حمید. ۱۳۹۰. طراحی شبکه فروش بیمه‌های عمر (مورد مطالعاتی شرکت سهامی بیمه ایران به روش

[۲۳] Level
[۲۴] clementine
[۲۵] Numerical
[۲۶] Categorical
[۲۷] Crisp

تحلیل پوششی داده‌ها(DEA)، پایان‌نامه کارشناسی‌ارشد، دانشگاه تربیت مدرس.

۴. دعایی حبیب ا..؛ نیکخواه فرخانی زهرا. ۱۳۸۸. ارزیابی عملکرد عملیاتی و منابع انسانی نمایندگی‌های بیمه کارافرین در استان خوزستان با نگرش چندگانه به روش تحلیل پوششی داده‌ها، فصلنامه صنعت بیمه، سال بیست و چهارم، شماره۳و۴، پاییز و زمستان، شماره مسلسل۹۶-۹۵.

۵. سلطان‌پناه هیرش و همکاران. ۱۳۸۶. ارزیابی کارایی شعب بیمه البرز با استفاده از تحلیل پوششی داده‌ها، فصلنامه صنعت بیمه، سال بیست و دوم، شماره۴، زمستان، شماره مسلسل۸۸، ۱۵۱-۱۷۷.

۶. مشیری سعید؛ رضوان مهدی. ۱۳۸۵. اثر به‌کارگیری فناوری ارتباطات و اطلاعات در کارایی صنعت خدمات هوایی ایران، فصلنامه پژوهش های اقتصادی ایران، سال هشتم، ش ۲۶.

۷. مهرگان محمدرضا. ۱۳۸۳. مدل‌های کمی در ارزیابی عملکرد سازمان‌ها، انتشارات دانشگاه تهران.

۸. Babazadeh, R. Razmi, J. Rabbani, M. Pishvaee, M. S. ۲۰۱۵. An integrated data envelopment analysis-mathematical programming approach to strategic biodiesel supply chain network design problem, Journal of Cleaner Production (Article in press).

۹. Barros, C. P., Nektarios, M. & Assaf, A. ۲۰۱۰. Efficiency in the Greek insurance industry, European Journal of Operational Research DEA technology. Omega 41,PP. ۳۱۵-۳۲۵.

۱۰. Bretholt A, Pan J. 2013. Evolving the latent variable model as a an environmental ۹. Charnes, A, Cooper, WW, Rhodes, E. ۱۹۷۸. Measuring the efficiency of decision making units, European journal of operational research, ۲, ٤۲۹-٤٤٤.

۱۱. Barr, R., L.M. Siford and T.F. Simes. ۱۹۹٤. Forecasting bank failure: a non-parametric approach, Recherches Economiques de Louvain.

۱۲. Breiman, L., J. Friedman, R. Olshenand C. Stone. ۱۹۸٤. Classification and Regression Trees, Pacific Grove, CA: Wadsworth-Monterey.

۱۳. Charnes A., W. W. Cooper , Rhodes E. ۱۹۷۸. Measuring the Efficiency, European Journal of Operations Research, No ۲.

۱٤. Chen, Y., Cook, W.D., Li, N., Zhu, J. ۲۰۰۹. Additive efficiency decomposition in two stage DEA. European Journal of Operational Research.

۱۵. Emrouznejad A., Anouze A. ۲۰۱۰. Data envelopment analysis with classification and regression tree-a case of banking efficiency, Expert Systems.

۱٦. Efron, B. and Tibishirani R. ۱۹۹۳. An Introduction to the Bootstrap, New York: Chapman and Hall.

۱۷. Fakhari A., Eftekhari Moghadam A.M. ۲۰۱۳. Combination of classification and regression in decision tree for multi-labeling image annotation and retrieval, Applied Soft Computing.

۱۸.Fan Ch.K, & Cheng, Sh.W. ۲۰۰۹. Using Analytic Hierarchy Process Method and Technique for Order Preference by Similarity to Ideal Solution to Evaluate Curriculum in Department of Risk Management and Insurance, J. Soc. Sci., ۱۹(۱).

۱۹. Farzipoor R. ۲۰۰۷. Suppliers selection in the presence of both cardinal and ordinal data, European Journal of Operational Research.

۲۰. Emroznejad A., DEA Home page ,http://www.deazone.com/tutorial.

۲۱. Han J., Kamber M. 2006. Data Mining Concept and Techniques. 2 nd Edition. San Francisco, Elsvier.

۲۲. Hand D.J., Manilla H., Smyth P. ۲۰۰۱. Principles of Data, Cambridge, MA: MIT Press.

۲۳. Hosseini Bamakan, S. M. Gholami, P. ۲۰۱٤. A Novel Feature Selection Method Based on an Integrated Data Envelopment Analysis and Entropy Model, ۲nd International Conference on Information Technology and Quantitative Management, ITQM ۲۰۱٤, Procedia Computer Science ۳۱, ٦۳۲ – ٦۳۸.

۲٤. Hwang S., Kao T.L. ۲۰۰۷. Measuring Managerial Efficiency in Non-Life Insurance Companies: An Application of Two-Stage Data Envelopment Analysis, International Journal of a Management, vol.1, No. 3.

۲۵. Jahanshahloo GR. , Alirezaee MR. ۱۹۹۲. Measuring the efficiency of academic units at the Teacher Training University, Procedings of 26 th Annual Iranian math conference.

۲۶. Kim, H. and G.J. Koehler.۱۹۹۵. Theory and practice of decision tree induction, Omega.
۲۷. Lee, S. ۲۰۱۰. Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B²C controls, Decision Support Systems, ٤۹, ٤۸٦–٤۹۷.
۲۸. Luhnen, M. ۲۰۰۹. Efficiency and Competition in Insurance Markets, Dissertation no. ۳٦۷٥.
۲۸. Mahlberg,B & Url, Th. ۲۰۱۰. Single Market effects on productivity in the German insurance industry, Journal of Banking & Finance.
۲۸. Malmquist, S. ۱۹٥۳. Index numbers and indifference surfaces. Trabajos de Estatis- tica.
۲۹. Nagurur, N. N. Rajbhanari, B. ۲۰۰۱. Data envelopment analysis for the performance evaluation of air conditioning and refrigeration companies in Thailand, Business Performance Management.
۳۰. Park, J. Lee, D.S. Christakis, N, and Barabasi,A.L.. ۲۰۰۹. The impact of cellular networks on disease comorbidity, Molecular Systems Biology.
۳۱. Pille, P. and Paradi J. ۱۹۹۷. Facets at the frontier and efficiency measurement in DEA, Paper presented at the Fifth European Workshop on Efficiency and Productivity Analysis.
۳۲. Seol, H. Choi, J. Park, G. Park, Y. ۲۰۰۷. A framework for benchmarking service process using data Envelopment analysis and decision tree, Expert Systems with Applications, ۳۲, ٤۳۲–٤٤۰.
۳۳. Shahroudi K., Taleghani M., Mohammadi G. ۲۰۰۱. Efficiency Decomposition in Data Envelopment Analysis: An application to Insurance companies in Iran.
۳٤.Sueyoshi T., Goto P. ۲۰۰۹. DEA-discriminate analysis Methodological comparison among eight discriminant analysis approaches, European Journal of Operational Research.
۳٥. Tavana, M., Keramatpour, M., Santos-Arteaga, F.J., Ghorbaniane, E. ۲۰۱٥. A Fuzzy Hybrid Project Portfolio Selection Method Using Data Envelopment Analysis, TOPSIS and Integer Programming, Expert Systems with Applications (Article in press).
۳٦. Torgo L. ۱۹۹۷. Functional models for regression tree leaves, Proceedings of the 14th International Conference on Machine Learning.
۳۷. Wang, C. H. Chuang, J. J. ۲۰۱٥. Integrating decision tree With back propagation network to conduct business diagnosis and performance simulation For solar companies, Decision Support Systems (Article in press).
۳۸. Xie X. ۲۰۱۰. Are publicly held firms less efficient? Evidence from the US property-liability insurance industry, Journal of Banking & Finance.
۳۹. Yao, Sh., Han,zh.,& Feng, G. ۲۰۰۷. On technical efficiency of China's insurance industry after WTO accession, China Economic Review.