طبقه بندی داده‌های ماشین با استفاده از تحلیل پوششی داده ها

بهرنگ برزگر

کارشناسی ارشد مهندسی کامپیوتر، گرایش نرم افزار

چکیده

تکنیک تحلیل پو ششی داده ها یک روش برنامه ریزی ریا ضی برای ارزیابی کارایی واحدهای تصمیم گیرنده ا ست که چندین ورودی و چندین خروجی دارند. اندازه گیری کارایی به دلیل اهمیت آن در ارزیابی عملکرد یک شرکت یا سازمان، همواره مورد توجه محققان قرار دا شته ا ست. در واقع می توان گفت تکنیک تحلیل پو ششی داده ها مبتنی بر یک‌سری بهینه سازی با استفاده از برنامه ریزی خطی می باشد که به آن روش ناپارامتریک نیز گفته می شود. طبقه بندی چیست، یک روش ساختن دسته‌های اشیاء یا خوشه‌ها است، طوری که اشیاء در یک خوشه بسیار مشابه هم و در خوشه‌های مختلف کاملاً از هم متمایز ه ستند. در این مقاله داده‌ای بازه‌ای را با ا ستفاده از تحلیل پو ششی داده ها طبقه بندی می کنیم و با ا ستفاده از تحلیل پو ششی داده‌ها مدل BCC برای محا سبه کارایی ن سبی داده های بازه ای ا ستفاده می شود. که در این مدل به دلیل بازه ای بودن داده‌ها، کران بالا و پایینی را برای مقدار بهینه تابع هدف را بایستی به د ست آوریم. با ا ستفاده از روش بیان شده، یک مثال عددی بازه ای را کران بالا و پایین آن را محا سبه کرده و تابع هدف میانگین را به عنوان مقدار تابع هدف بهینه بد ست می آید. و از این مقادیر برای طبقه بندی داده ها در دو کلاس پذیرش و رد و حالت هایی که بین این دو م شترک ه ستند استفاده می شود.

کلمات کلیدی: تحلیل پوششی داده‌ها، واحد های تصمیم گیرنده ، طبقه بندی، داده های بازه ای

۱- مقدمه:

یکی از شاخه های مهم علم تحقیق در عملیات تحلیل پو ش شی داده ها (DEA) ا ست که یک روش غیرپارامتری برای ارزیابی کارایی و یا محا سبه بهره وری تعداد متناهی از واحدهای ت صمیم گیرنده متجانس در حالت چند ورودی و چندخروج ـی است. (آذر، ۱۳۷۹). به عب ـارت دیگر تحلیل پوششی دادهها، یک روش برنامه ریزی ریاضی، برای ارزیابی کارآیی واحدهای تصمیم گیرنده[1](DMUs) ای است که چندین ورودی و چندین خروجی دارند. تحلیل پوششی دادهها[2] (DEA)، شامل تکنیکها و روشهایی برای ارزیابی کارایی و یا سنجش بهره وری واحدهای ت صمیم گیرنده (DMU) است. DEA در واقع تعمیم کار فارل[3] در ابداع اولین روش غیر پارامتری ا ست. فارل با ا ستفاده از ورودیها و خروجیهای واحدهای ت صمیم گیرنده و ا صول حاکم بر آنها، مجموعهای با عنوان مجموعه امکان تولید، ارائه و ق سمتی از مرز آن را به عنوان تابع تولید معرفی نمود. این مرز را مرز کارا نیز می نامند و واحدهای ت صمیم گیرندهای که روی این مرز قرار میگیرند، کارا ارزیابی می شوند. از آنجایی که DEA، تکنیک ارزیابی کارایی ن سبی واحدهای ت صمیم گیرنده ا ست، حداقل یکی از واحدها روی مرز و بقیه واحدها در زیر آن قرار دارند. (آذر، ۱۳۷۹). نام تحلیل پو ششی دادهها از ویژگی پو ششی بودن، منشأ گرفته ا ست. DEA همچنین فر صتهای زیادی را برای همکاری میان تحلیلگر و ت صمیم گیرنده ایجاد میکند. این همکاریها میتواند در را ستای انتخاب ورودی و خروجی واحدهای تحت ارزیابی و چگونگی عملکرد و الگویابی ن سبت به مرز کارا با شد. در واقعیت ساختار بسیاری از واحدهای تصمیم گیرنده به صورت یک جعبه که تنها ورودی و خروجی دا شته با شند نیست بلکه در حقیقت واحدهای تصمیم گیرنده شامل زیر واحدهای تصمیم گیرنده ای هستند که با هم در ارتباط می باشند، بنابراین ساختار واقعی واحدهای تصمیم گیرنده بسیار پیچیده و مانند شبکه های متفاوت است. برای این منظور محققیق تحلیل پوششی داده های شبکه ای (Network DEA) را معرفی کردند.(لطفی و همکاران، ۱۳۹۰)

در این روش نیازی به تعیین شکل صریح تابع تولید نی ست و از برنامه ریزی خطی برای ساختن یک سطح قطعه-قطعه خطی (یا مرز) برای پوشاندن (نام تحلیل پوششی داده ها از این ویژگی منشأ گرفته است) تمام داده ها استفاده می شود و سپس کارایی هر یک از واحدهای تصمیم گیرنده نسبت به این مرز محاسبه می شود. مرز به دست آمده همان مرز کارآیی است که نقاط واقع بر آن نقاط کارآ هستند. سایر واحدها که در داخل سطح پو ششی قرار می گیرند ناکارآ هستند و با یکی از تصاویر خود بر روی سطح پو ششی مقایسه می شود که نحوهٔ تصویر شدن واحدهای ناکارآ بر روی مرز در مدلهای مختلف و بسته به ماهیت مدل متفاوت است. (مهرگان، ۱۳۹۱).

این روش بر پایه کار اقتصاددانی به نام فارل ا ست که پایه گذار روش های غیرپارامتری در ارزیابی کارایی و محا سبه بهره وری واحدهای تصمیم گیرنده ا ست. او در سال ۱۹۵۷، اولین روش غیر پارامتری جهت تعیین کارایی را در حالت دو ورودی

---

[1] Decision Making Unit
[2] Data Envelopment Analysis
[3] Farrell

و یک خروجی ارائه نمود و روش پو سته محدب قطعه-قطعه خطی برای تقریب مرز را ارائه کرد. برای تعیین اندازه کارایی واحدهای تصمیم گیرنده، فارل پیشنهاد کرد که ابتدا بایستی یک مرز کارای مفروض را مشخص کرد و سپس فا صله از مرز کارا را به عنوان یک اندازه ناکارایی تعبیر نمود. وی به جای برآورد تابع تولید، مرز کارای قطعه-قطعه خطی را با اعمال فرض های زیر بااستفاده از تبدیل یک به یک بدست آورد:

۱. شیب پاره‌خط ها، منفی یا صفراست.

۲. هیچ واحدی بین مرز و مبدا قرار نمی‌گیرد. به عبارت دیگر تمام نقاط م شاهده شده در سمتی از مرز قرار می‌گیرند که مبدا در آن واقع نباشد.

۳. نقاطی که روی مرز قرار می‌گیرند نقاط کارا و بقیه ناکارا هستند و میزان ناکارایی آنها نسبت به مرز محاسبه می شود.

روش فارل با اینکه مشکل مربوط به انتخاب تابع تولید را رفع کرد ولی هنوز مشکل تعداد ورودی و خروجی را داشت. در سال ۱۹۷۸ چارنز، کوپر و رودز با ا ستفاده از برنامه ریزی ریا ضی روش غیرپارامتری فارل را، در مقاله ای تحت عنوان "اندازه گیری کارایی واحدهای ت صمیم گیرنده"، برای سی ستمی با ورودی ها و خروجی های چندگانه تعمیم دادند و عنوان "تحلیل پو ششی داده ها" (Data Envelopment Analysis) را به آن دادند. البته این تکنیک قبل از آن در سال ۱۹۷۶ در رساله دکتری رودز به راهنمایی کوپر تحت عنوان "ارزیابی پیشرفت تحصیلی دانش آموزان مدارس ملی آمریکا" مورد استفاده قرار گرفته بود. مدلی که تو سط چارنز، کوپر و رودز معرفی شد مدل معروف CCR ا ست که با فرض بازده به مقیاس ثابت ارائه شد و یکی از اساسی ترین مدل های تحلیل پوششی داده ها است. از آن زمان به بعد مطالعات زیادی در زمینه کاربرد و توسیع این روش صورت گرفت و مدل های جدید و مقالات زیادی در این زمینه ارائه شد. به عنوان نمونه، در سال ۱۹۸۴، بنکر، چارنز و کوپر مدل CCR را برای بازده به مقیاس متغیر بسط دادند و مدل معروف BCC را ارائه نمودند.

۲- بیان مساله

تحلیل پو ششی داده ها، برنامه ریزی ریا ضی ا ست که برای ارزیابی عملکرد واحدهای ت صمیم گیرنده متجانس که دارای ورودی ها و خروجی های چندگانه می با شند به کار می رود. از روش تحلیل پو ش شی داده‌ها می‌توان به عنوان یکی از پرکاربردترین روش‌های ارزیابی عملکرد یاد نمود. اندازه گیری کارایی به دلیل اهمیت آن در ارزیابی عملکرد یک شرکت یا سازمان، همواره مورد توجه محققان قرار دا شته ا ست. (حمزه پور و همکاران، ۱۳۹۱). در واقع می توان گفت تکنیک تحلیل پوششی داده ها مبتنی بر بهینه سازی با استفاده از برنامه ریزی خطی می باشد که به آن روش غیر پارامتریک نیز گفته می شود. طبقه بندی را می توان به صورت د سته بندی یک ماهیت به چند ماهیت دیگر، به ق سمی که ماهیت های تق سیمی از ماهیت ا صلی یک یا چند ویژگی را به ارث برند، تعریف نمود. طبقه بندی ها ابزاری ه ستند که جهت شنا سایی دقیق و کامل از یک ماهیت، مانند طبقه بندی موجودات زنده، طبقه بندی ا شیا، طبقه بندی م شاغل، طبقه بندی اطلاعات، طبقه بندی ا سناد، طبقه بندی ح سابها و طبقه بندی کالا و خدمات مورد ا ستفاده قرار می گیرند. در ادبیات طبقه بندی داده ها، ممکن ا ست کلمات

مختلفی برای بیان یک چیز به کار رود. برای مثال، برای یک بانک اطلاعات فر ضی که شامل بـ سیاری گزار شات ا ست، ا صطلاحات نقطه داده، مورد نمونه، مشاهده، شیء، جزء، عنصر و چندتایی، همگی برای بیان یک نقطه تنها به کار می رود. (لطفی و همکاران، ۱۹۰)

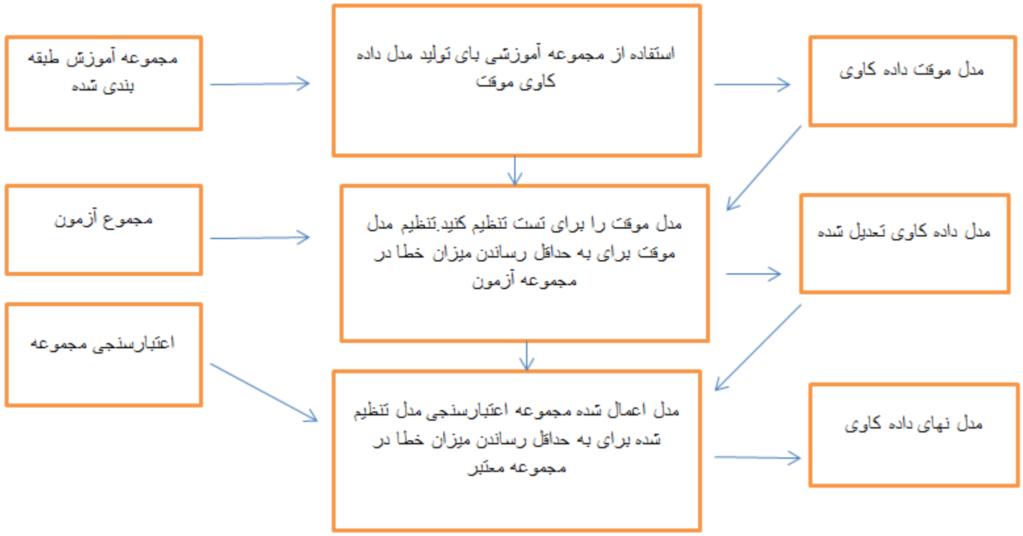
در این مقاله رابطه ی معادل میان ما شین طبقه بندی داده DCM[4] و مدل تحلیل پو شـ شی داده ها (DEA) را برر سی نموده و یک ماشین طبقه بندی بر پایه ی تحلیل پوششی داده را به وجود می آورد . داده توسط مجموعه ای از مقادیر مشخص می گردد. طبقه بندی بیان می کند که آیا یک داده ی مشخص، متعلق به یک گروه مشخص بر پایه ی مجموعه ای از مقادیر، ویژگی ها و یا مشخصه ی از پیش تعیین شده، است و یا خیر. ما به چنین داده ای به عنوان واحد تصمیم گیرنده (DMU) نگاه می کنیم. همچنین مقادیر ویژگی های داده شده را به عنوان ورودی در نظر می گیریم. ما شین طبقه بندی تحلیل پو ششی داده های پیـ شنهادی شامل یک حوزه ی پذیرش[5] و یک تابع طبقه بندی[6] ا ست . حوزه ی پذیرش تو سط یک سیـ ستم شفاف و آ شکار نامعادلات خطی ارائه می شود. که این باعث سهولت محاسبات فرایند طبقه بندی می گردد. این روش می تواند برای طبقه بندی حجم وسیعی از داده استفاده شود.

۳- طبقه‌بندی

طبقه‌بندی را می‌توان دسته‌بندی یک ماهیت به چند ماهیت دیگر به صورتی که ماهیت‌های تقسیمی از ماهیت اصلی یک یا چند ویژگی را به ارث ببرند تعریف نمود. طبقه‌بندی داده‌ها که تحلیل خو شه‌ای، تحلیل چندبخـ شی یا تحلیل طبقه‌بندی بی ناظر نیز نامیده می شود، یک روش ساختن د سته‌های ا شیاء یا خو شه‌ها ا ست، طوری که ا شیاء در یک خوشه بسیار مشابه هم و در خوشه‌های مختلف کاملاً از هم متمایز هستند. خوشه‌بندی داده‌ها، اغلب با طبقه‌بندی، که در آن ا شیاء به د سته‌های از پیش تعریف شده اختـ صاص می‌یابند، ا شتباه می شود. در خو شه‌بندی داده‌ها، د سته‌ها نیز باید تعریف‌شده باشند. طبقه‌بندی یک روش کشف داده قوی برای تشکیل گروه داده‌ها و نمایان ساختن اطلاعات ساختاری اجزاء یک مجموعه از داده های مفروض ا ست. امروزه طبقه‌بندی در نواحی مختلف کاربرد های گوناگونی دارد. دسته‌بندی درواقع ارزشیابی ویژگی‌های مجموعه‌ای از داده‌ها و سپس اختصاص دادن آن‌ها به مجموعه‌ای از گروه‌های از پیش تعریف‌شده است. این متداول‌ترین قابلیت داده‌کاوی هست. داده‌کاوی را می‌توان با استفاده از داده‌های تاریخی برای تولید یک مدل یا نمایی از یک گروه بر اساس ویژگی‌های داده‌ها به کاربرد. سپس می‌توان از این مدل تعریف‌شده برای طبقه‌بندی مجموعه داده‌های جدید ا ستفاده کرد. همچنین می‌توان با تعیین نمایی که با آن سازگار ا ست برای پیش‌بینی‌های آتی ازآن بهره گرفت.در دنیای امروز بحثclassification اطلاعات اهمیت بـسیاری دارد،اینکه بتوان

[4] Data Classification Machine
[5] Acceptance Domain
[6] Classification Function

مدلی مناسب برای تحلیل داده‌هایی خاص به دست آورد و بتوان با بررسی اولیه ویژگی‌های یک عنصر خاص ، الگوی رفتاری آن عنصر را پیش‌بینی کرد.



شکل ۱ : طبقه بندی

در مسائل دسته‌بندی هدف شناسایی ویژگی هایی است که گروهی را که هر مورد به آن تعلق دارد را نشان دهند. از این الگو می‌توان هم برای فهم داده‌های موجود و هم پیش‌بینی نحوه رفتار مواد جدید ا ستفاده کرد. در داده کاوی مبحث طبقه بندی اطلاعات به برر سی اینگونه مدل ها و متدها می پردازد .در د سته‌بندی اطلاعات هدف بد ست آوردن مدلی برای الگوی رفتاری و ویژگی های مجموعه ایی از داده ها ا ست تا با کمک آن بتوان بدون دان ستن رفتار یک موجودیت، با توجه به ویژگی های آن و با استفاده از مدل بدست آورده شده، رفتار آن را تشخیص داد و آن موجودیت را در گروه خا صی طبقه بندی کرد . داده‌کاوی مدلهای د سته‌بندی را با برر سی داده‌های د سته‌بندی شده قبلی ایجاد می‌کند و یک الگوی پیش‌بینی کننده را ب صورت ا ستقرایی می‌یابند. این موارد موجود ممکن ا ست از یک پایگاه داده تاریخی آمده باشند.

داده کاوی مدلهای د سته‌بندی را به و سیله امتحان کردن داده طبقه بندی شده(موارد) و نهایتا یافتن یک الگوی پیش گو ایجاد می کند. این موارد موجود می تواند از یک پایگاه داده تاریخی نا شی شود مانند اطلاعات افرادی که تحت معالجه دارویی خاصی هستند و یا به سمت یک خدمت با مسافت دور جذب شده اند.یا اینکه از تجربه هایی که طی آن یک نمونه از تمام پایگاه داده در جهان واقعی تست شده باشد و نتایج آن برای ایجاد یک گروه بندی استفاده شده باشند منتج شود.

از جمله تکنیک های داده کاوی که برای طبقه بندی به کار می آیند می توان از الگورریتم طبقه بندیKNN و تئوری بیز نام برد.

K-nearest neighbor یکی از الگوریتم های طبقه بندی می باشد. مبنای الگوریتم KNN پیدا کردن تعداد معینی از نزدیکترین عناصر موجود در جامعه آماری به عنصر جدید واردشده در آن جامعه است که بر اساس آن بتوان نزدیکترین داده (داده ها) موجود به عنصر جدید را از لحاظ ویژگی های مختلف پیدا کرد تا عنصر جدید را در همان طبقه ای قرار داد که عناصر نزدیک به آن قرار دارند. Knn یکی از روش های غیر پارامتریک برای بدست آوردن تابع توزیع از روی داده های توزیع شده می باشد. همچنین این روش یکی از متداول ترین روش ها برای دسته بندی داده ها می باشد.

تئوری بیز نیز یکی از مهم‌ترین روش های آماری بوده که از آن می توان برای طبقه بندی داده ها در داده کاوی ا ستفاده کرد. در این روش هر کدام از کلاس های مختلف به شکل یک فر ضیه دارای احتمال در نظر گرفته می شوند. در رویکرد ذکر شده هر رکورد آموزشی جدید احتمال درست بودن فرضیه های پیشین را افزایش و یا کاهش داده و در نهایت فرضیاتی که دارای بالاترین احتمال شوند به عنوان یک کلاس در نظر گرفته شده و برچسبی بر آنها زده می شود. در نگرش بیزی، با فرض اینکه مدل های احتمالی دارای پارامترهای ناشناخته‌ای هستند که برای هر یک از مقادیر ناشناخته، توزیع احتمال اولیه‌ای در نظر گرفته می گیریم که بازگو کننده باور ما به محتمل بودن هر یک ازاین مقادیر بدون دیدن دادها ست. با جمع آوری داده و مشاهده آن، مقدار توزیع احتمال ثانویه را محاسبه و با استفاده از آن به یک نتیجه گیری در مورد عدم قطعیت می‌رسیم. همچنین با میانگین گیری روی مقادیر احتمال ثانویه پیش بینی انجام می‌دهیم به گونه ای که خطای ثانویه مورد انتظار کاهش یابد . همچنین، اگر در فضای فرضیه H بدنبال بهترین فرضیه‌ای باشیم که درمورد داده‌های آموزشی D صدق کند، یک راه تعیین بهترین فرضیه، انتخاب محتمل ترین فرضیه بر مبنای داده‌های آموزشی D است. می‌توان انتظار داشت تئوری بیز چنین راه حلی را ارائه می‌دهد. این تئوری امکان محاسبه احتمال ثانویه را بر مبنای احتمالات اولیه می‌دهد:

(۱) $P(h|D) = \frac{P(D|H)P(H)}{P(D)}$

همانطور که مشاهده می شود با افزایش $P(D)$ مقدار $P(h|D)$ کاهش می‌یابد. زیرا هر چه احتمال مشاهده D مستقل از h بیشتر باشد به این معنا خواهد بود که D شواهد کمتری در حمایت از hدر بر دارد. در حالی که هرچه احتمال مشاهده D (همراه با پذیرش فرضیه h) بیشتر باشد داده های D شواهد بیشتری در حمایت از h در بر دارد.

۴- اهداف تحقیق

٤-١- هدف اصلی

بررسی رابطه میان طبقه بندی داده های ماشین و تحلیل پوششی داده (DEA)

٤-٢- اهداف فرعی

ارایه روش جدید برای طبقه بندی داده ها

ایجاد یک ماشین طبقه بندی بر پایه ی تحلیل پوششی داده

۵- فرضیات تحقیق

۵-۱- فرضیه اصلی

تحلیل پو ش شی داده ها یک روش برنامه ریزی ریا ضی برای ارزیابی کارایی واحدهای ت صمیم گیرنده متجانس با چند ورودی و چند خروجی است.

۵-۲- فرضیه های فرعی

تحلیل پو ش شی داده‌ها برای طبقه بندی داده ها مورد ا ستفاده قرار می گیرد.همه ی واحدهای ت صمیم گیرنده با م صرف ورودی ها، خروجی ها را تولید می کند.

٦- روش تحقیق

این تحقیق از جمله تحقیقات کمی و از حیث شیوه آزمایشی است. این تحقیق از منظرهای مختلف بشرح زیر است و از نظر تعداد محققان: این تحقیق از نوع تحقیقات فردی است زیرا جهت انجام پایان نامه دوره کارشناسی ارشد انجام شده است. همچنین از نظر رویکرد، تحقیق مذکور جز تحقیقات کمی و از حیث شیوه آزمایشی محسوب می‌شود.

٦-١- مدل CCR

این مدل اولین مدل تحلیل پوششی داده‌ها است که توسط چارنز، کوپر و رودز در سال ۱۹۷۸ ارائه شد (چارنز و همکاران، ۱۹۷۸). نام این مدل از نام پدیدآورندگان مدل گرفته شده ا ست و مجموعه امکان تولید این مدل را T_c ا ست. برای بیان این مدل مجموعه‌ی T_c و DMU_0 که روی مرز کارایی قرار ندارد را در نظر بگیرید. به روش‌های مختلف می‌توان آن را به سوی مرز T_c سوق داد که مهم‌ترین این روش‌ها عبارتند از:

الف) کاهش ورودی‌ها

ب) افزایش خروجی‌ها

در واقع در ق سمت (الف) برای ارزیابی DMU_0 سعی بر حداکثر کاهش شعاع ورودی‌های آن با ن سبت θ داریم، تا جائیکه از PPS خارج ن شویم. به عبارت دیگر دنبال یک DMU (حقیقی یا مجازی) ه ستیم، که با ورودی‌های کمتر همان

مقدار خروجی‌ها را تولید کند. در صورت وجود چنین DMUای، DMU_0 ناکار ا ست. برای ارزیابی DMU_0 م سأله بهینه‌سازی زیر را بایستی حل کنیم:

$$\begin{aligned} &\text{Min}\ \theta \\ &s.t.(\theta x_0, y_0) \in T_c. \end{aligned} \qquad (۲)$$

شرط عضویت در T_c با توجه به تعریف T_c آن است که $\theta x_0 \geq \sum_{j=1}^{n} x_j\lambda_j$ و $y_0 \leq \sum_{j=1}^{n} y_j\lambda_j$.

پس مسألهٔ (۲) به‌صورت زیر خواهد بود:

$$\begin{aligned} &\text{Min}\ \theta \\ &s.t. \sum_{j=1}^{n} \lambda_j x_j \leq \theta x_0 \\ &\sum_{j=1}^{n} \lambda_j y_j \geq y_0 \\ &\lambda_j \geq 0, j = 1, \dots, n, \theta \ \text{آزاد}. \end{aligned} \qquad (۳)$$

مسألهٔ (۳) یک مسألهٔ برنامه‌ریزی خطی است که در آن مجهولات عبارتند از $\lambda_1, \lambda_2, \dots, \lambda_n$ و θ. این مسأله معروف به‌صورت مدل پوششی CCR با اندازه‌گیری ورودی محور است.

اگر در جواب بهین (۳)، $\theta^* < 1$ باشد، آن‌گاه DMU_0 روی مرز قرار ندارد. درنتیجه ناکاراست. زیرا $(\theta^* x_0, y_0) \in T_c$ و هم‌چنین $\theta^* x_0 < x_0$ و این یعنی $(\theta^* x_0, y_0)$ که عضوی از T_c است، بر (x_0, y_0) غالب می‌باشد.

همانطور که اشاره شد در قسمت (الف) سعی بر حداکثر کاهش متناسب ورودی‌ها ضمن حفظ خروجی‌ها در همان سطح قبلی شان ا ست تا جائیکه از PPS خارج ن شویم. در حالت (ب) اگر ورودی‌ها را در همان سطح قبلی شان نگه داریم و خروجی‌ها را با نسبت φ افزایش دهیم، تا جائیکه در PPS باقی بمانیم، مسألهٔ زیر را داریم:

$$\begin{aligned} &Max\ \phi \\ &s.t.(x_0, \theta y_0) \in T_c \end{aligned} \qquad (۴)$$

با توجه به ساختار مجموعهٔ T_c مسألهٔ (۴) به‌صورت زیر خواهد بود:

$$\begin{aligned} &Max\ \phi \\ &s.t. \sum_{j=1}^{n} \lambda_j x_j \leq x_0 \\ &\sum_{j=1}^{n} \lambda_j y_j \geq \phi y_0 \\ &\lambda_j \geq 0, j = 1, \dots, n, \phi \ \text{آزاد}. \end{aligned} \qquad (۵)$$

م سألهٔ (۵) نیز یک م سألهٔ برنامه‌ریزی خطی ا ست. اگر ϕ^* جواب بهین آن با شد و $\phi^* > 1$ در این صورت نقطه‌ای مانند $(x_0, \phi^* y_0)$ در T_c وجود دارد که $\varphi y_0 > y_0$ و این یعنی $(x_0, \phi^* y_0)$ بر (x_0, y_0) غالب ا ست. پس DMU_0 روی مرز کارایی قرار ندارد و این یعنی DMU_0 ناکاراست. دوگان مدل (۵) به‌صورت زیر است:

$$\begin{aligned} &Max\ u^T y_0 \\ &s.t. v^T y_j - v^T x_j \leq 0, j = 1, \dots, n \\ &v^T x_0 = 1 \end{aligned} \qquad (۶)$$

$u \geq 0 , v \geq 0$

مدل (۶) صورت مضربی مدل CCR با اندازه‌گیری ورودی محور است. در این مدل مجهولات $u = (u_1, \dots, u_s)$ و $\upsilon = (v_1, \dots, v_m)$ است. تعداد قیود مدل (۳) برابر $m + s$ و تعداد قیود مدل (۶) برابر $n + 1$ است.

برای تعریف دقیق کارایی ابتدا با افزودن متغیرهای کمکی s^{-19} و s^{+} به قیود مسأله‌ی (۳) آن را به فرم استاندارد زیر درمی‌آوریم:

$$
\begin{aligned}
& Min\ \theta \\
& s.t.\ \theta x_{i_0} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_i^- = 0 , i = 1, \dots, m \\
& \textstyle\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{r_0}, r = 1, .., s \qquad (۷) \\
& \lambda_j \geq 0, j = 1, \dots, n \\
& s_i^- \geq 0, i = 1, \dots, m \\
& s_r^+ \geq 0\ r = 1, \dots, s.
\end{aligned}
$$

حل مسأله‌ی (۷) که همان مسأله‌ی (۳) است، اطلاعات بسیار سودمندی در اختیار ما قرار می‌دهد که عبارتند از (تون، ۲۰۰):

اگر مقدار بهینه θ یعنی θ^* برابر با یک باشد و در یک جواب بهینه بعضی از متغیرهای کمکی مخالف صفر باشد آن‌گاه DMU_0 توسط نقاطی در مجموعه امکان تولید مغلوب خواهد شد.

تعریف ۱- ۴.۱. اگر در همه جواب‌های بهینه مدل (۷) $\theta^* = 1$ و $s^{-*} \neq 0$ یا $s^{+*} \neq 0$، آن‌گاه DMU_0 را کارای ضعیف گوییم (تون، ۲۰۰۰).

تعریف ۱- ۴.۲. اگر در مسأله‌ی (۷) $\theta^* < 1$، آن‌گاه DMU_0 را به‌طور اکید ناکارا گوییم(تون، ۲۰۰۰).

لذا برای کارا بودن $\theta^* = 1$ لازم است، ولی کافی نیست و از حل مدل (۱.۶) نیز نمی‌توان کارای قوی را بدست آورد. یعنی چنانچه در ارزیابی DMU_0 داشته باشیم $\theta^* = 1$ و تمام متغیرهای کمکی صفر باشند، نمی‌توان به‌طور قاطع اظهارنظر کرد DMU_0 کارای قوی است، زیرا ممکن است جواب بهینه‌ای وجود داشته باشد که متغیرهای کمکی آن دارای مؤلفه‌ی مثبت باشند. لذا برای رفع این مشکل از روش دوفازی استفاده می‌کنیم (تون، ۲۰۰۰).

فاز اول: ابتدا مدل (۷) را حل می‌کنیم. فرض کنیم جواب بهینه (θ^*, λ^*) به‌دست آید. اگر $\theta^* < 1$ باشد، DMU_0 ناکاراست. اما اگر $\theta^* = 1$ باشد و بعضی از متغیرهای کمکی مخالف صفر باشند، DMU_0 کارای ضعیف می‌باشد. اما اگر $\theta^* = 1$ و همه‌ی متغیرهای کمکی برابر صفر باشند در این‌صورت مسأله‌ی فاز دوم زیر را حل می‌کنیم.

$$
\begin{aligned}
& Max \sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- \\
& s.t. \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta^* x_{i_0}, i = 1, \dots, m \\
& \textstyle\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{r_0}, r = 1, \dots, s \qquad (٨) \\
& \lambda_j \geq 0, j = 1, \dots, n \\
& s_i^- \geq 0 \;\; i = 1, \dots, m \\
& s_r^+ \geq 0 \;\; r = 1, \dots, s.
\end{aligned}
$$

اگر در جواب بهینه، مقدار تابع هدف (٨) برابر صفر باشد آن‌گاه DMU_0 به مفهوم پاراتو کارا است. در غیر این صورت DMU_0 کارای ضعیف است.

٦-٢- مدل BCC

مجموعه امکان تولید این مدل T_v است که قبلاً به آن اشاره کردیم. فرم پوششی مدل BCC با ماهیت ورودی:

در این مدل هدف، حداکثر کاهش ورودی‌های DMU_0 با نسبت θ و ثابت نگه داشتن مقدار خروجی‌ها است، به‌طوری که $(\theta x_0, y_0)$ به T_v متعلق باشد. یعنی حل مسأله‌ی زیر:

$$
\begin{aligned}
& Min\ \theta \\
& s.t. (\theta x_0, y_0) \in T_v \qquad (٩)
\end{aligned}
$$

که با توجه به ساختار T_v مسأله‌ی (٩) به‌صورت زیر خواهد بود:

$$
\begin{aligned}
& Min\ \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\
& s.t. \sum_{j=1}^{n} x_{ij} \lambda_j + s_i^- = \theta x_{i_0}, i = 1, \dots, m \\
& \textstyle\sum_{j=1}^{n} y_{rj} \lambda_j - s_r^+ = y_{r_0}, r = 1, \dots, s \qquad (١٠) \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \lambda_j \geq 0, j = 1, \dots, n \\
& s_i^- \geq 0 \; i = 1, \dots, m \\
& s_r^+ \geq 0 \;\; r = 1, \dots, s.
\end{aligned}
$$

که در آن ε یک عدد غیر ارشمیدسی است و به‌صورت عدد مثبت کوچک‌تر از هر عدد حقیقی مثبت تعریف می‌شود. این بدان معنی است که ε یک عدد حقیقی نیست. این مدل توسط بانکر، چارنز و کوپر در سال ۱۹۸۴ ارائه شد و با حروف BCC شناخته می‌شود که از حرف اول نام پدیدآورندگان آن گرفته شده است (بانکر و همکاران، ۱۹۸۴).

٦-٣- فرم مضربی مدل BCC با ماهیت ورودی:

همانند CCR، با گرفتن دوگان از فرم پوششی مدل BCC، فرم مضربی آن را به‌دست می‌آوریم که به صورت زیر است:

$$
\begin{aligned}
& Max \sum_{r=1}^{s} u_r y_{r_0} - u_0 \\
& s.t. \sum_{i=1}^{m} x_{i_0} v_i = 1 \\
& \textstyle\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} x_{ij} - u_0 \le 0, j = 1, \dots, n \qquad (۱۱) \\
& u_r \ge \varepsilon, \qquad r = 1, \dots, s \\
& v_i \ge \varepsilon, \qquad i = 1, \dots, m \\
& u_0 \text{ آزاد.}
\end{aligned}
$$

تعاریفی از کارایی که برای مدل CCR ارائه شده است برای مدل BCC نیز برقرار هستند.

۷- طبقه بندی برای داده های بازه‌ای

در مبحث طبقه بندی با استفاده از DEA برای دو حالت پذیرش و رد از دو مدل استفاده می شد. در آن فصل باید این مدل هارا حل می کردیم تا بتوانیم مرزهای بین مجموعه واحدهای تصمیم گیرنده ای که پذیرش کردیم یا رد کردیم را رسم کنیم. وتمایز شان را متوجه شویم . آن دو مدل که برای داده های غیربازه ای بودند دوباره در این فصل آورده شدند که به صوررت (۱۰-۴) و(۱۱-۴) هستند.

اگر برای برخی از n، ... ، $t = 1$، راه حل بازده‌های فرمول (۱۲) یک ارزش$\mathcal{E}^t = 1$ باشد سپس حالتی در نظر گرفته می‌شود که در مرز کلاس پذیرش DEA[7] باشد. یعنی مدل زیر برای کلاس پذیرش استفاده می شود.

$$
\begin{aligned}
& \text{Min } \mathcal{E}^t \text{ ،} t = 1 \text{ ، } \dots \text{ ،} n \\
& \text{S.t} \\
& \textstyle\sum_{i=1}^{n} \lambda_i x_{ij} - \mathcal{E}^t x_{tj} \le 0 \quad j = 1 \text{ ، } \dots \text{ ،} m \ \ t = 1 \text{ ، } \dots \text{ ،} n \\
& \textstyle\sum_{i=1}^{n} \lambda_i = 1 \qquad\qquad (۱۲) \\
& \lambda_i \ge 0 \quad i = 1 \text{ ، } \dots \text{ ،} n
\end{aligned}
$$

اگر برای برخی ار ، راه حل بازده‌های فرمول (۱۳) یک ارزش $\mathcal{E}^t \ge 1$ باشد سپس حالتی در نظر گرفته می‌شود که در مرز کلاس رد DEA باشد. یعنی مدل زیر برای کلاس موارد ردی استفاده می شود.

$$
\begin{aligned}
& \text{Max } \mathcal{E}^t \text{ ،} t = 1 \text{ ، } \dots \text{ ،} n \\
& \text{S.t} \\
& \textstyle\sum_{i=1}^{n} \lambda_i x_{ij} - \mathcal{E}^t x_{tj} \ge 0 \quad j = 1 \text{ ، } \dots \text{ ،} m \ \ t = 1 \text{ ، } \dots \text{ ،} n \\
& \textstyle\sum_{i=1}^{n} \lambda_i = 1 \qquad\qquad (۱۳) \\
& \lambda_i \ge 0 \quad i = 1 \text{ ، } \dots \text{ ،} n
\end{aligned}
$$

[7] Accept class

دو مدل بالا برای طبقه بندی داده های غیربازه ای استفاده می شد که آن ها را در دو کلاس پذیرش یا رد قرار می داد. در مبحث زیر مدل های بالا برای داده های بازه ای پیاده سازی شده است.

در زیر مدل مطرح شده برای کلاس پذیرش در مبحث طبقه بندی برای داده های بازه ای به صورت (۱۴) تعریف می شود. که این مدل با فرض وجود کران های بالا و پایین برای ورودی ها و خروجی ها به صورت زیر خواهد بود.

$$
\begin{aligned}
&\text{Min } \mathcal{E}^t \qquad t = 1 ، \dots ، n\\
&\text{s.t}\\
&\textstyle\sum_{i=1}^{n} \lambda_i x_{ij} - \mathcal{E}^t x_{tj} \le 0 \quad j = 1 ، \dots ، m \quad t = 1 ، \dots ، n\\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1\\
&\lambda_i \ge 0 \quad i = 1 ، \dots ، n\\
&x_{ij} \in \left[x_{ij}^{L} ، x_{ij}^{U}\right] \quad i = 1 ، \dots ، m \quad j = 1 ، \dots ، s
\end{aligned} \tag{۱۴}
$$

به دلیل بازه ای بودن داده‌ها، کران بالا و پایینی را برای مقدار بهینه تابع هدف مدل فوق بایستی به دست آوریم. بسته به اینکه واحد تحت ارزیابی در چه موقعیتی و سایر واحدها در چه موقعیتی قرار داشته باشند دو حالت مختلف را برای این مدل در نظر می گیریم که در این دوحالت مقادیر Θ^U و Θ^L به عنوان کران هایی برای تابع هدف مدل (۱۴) به دست می آید. و دوحالت زیر را متناظر با این مقادیر در نظر می‌گیریم:

حالت اول: واحد تحت ارزیابی دربدترین حالت خود($x_{ik} = x^U{}_{ik}$ و $y_{rk} = y^L{}_{rk}$) قرار داشته. وبقیه واحدها در بهترین موقعیت خود ($x_{ik} = x^L{}_{ik}$ و $y_{rk} = y^U{}_{rk}$) واقع شده باشند. لذا از مدل(۱۵) استفاده می شود

$$
\begin{aligned}
&\theta^L = \text{Min } \mathcal{E}^t \qquad t = 1 ، \dots ، n\\
&\text{s.t}\\
&\textstyle\sum_{i=1}^{n} \lambda_i x_{ij}^{L} - \mathcal{E}^t x_{tj}^{U} \le 0\\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1\\
&\lambda_i \ge 0
\end{aligned} \tag{۱۵}
$$

حالت دوم: جهت یافتن کران بالای تابع هدف θ^U در مدل (۱۶)بایستی واحد تحت ارزیابی در بهترین موقعیت خود ($x_{ik} = x^L{}_{ik}$ و $y_{rk} = y^U{}_{rk}$) و بقیه واحدها در بدترین موقعیت خود ($x_{ik} = x^U{}_{ik}$ و $y_{rk} = y^L{}_{rk}$) واقع باشند. مدل (۱۶) θ^U رابدست می آورد.

(۱۶)

$\theta^{U} = \text{Min}\, \mathcal{E}^{t} \qquad t = 1 ، \ldots ،n$

s.t

$\sum_{i=1}^{n} \lambda_i x^{U}{}_{ij} - \mathcal{E}^{t} x^{L}{}_{tj} \leq 0$

$\sum_{i=1}^{n} \lambda_i = 1$

$\lambda_i \geq 0$

در ادامه ثابت می شود اگر ورودی و خروجی ها دارای کران بالا و پایین با شند. آنگاه می توان برای مقدار بهینه تابع هدف مدل (۱۴) کران بالا θ^{U} و پایین θ^{L} را بدست آورد.

قضیه(۲-۴): اگر θ^{*} جواب بهینه مدل (۱۴) با شد. همچنین θ^{U} و θ^{L} جواب های بهینه مدل (۱۶) و(۱۵) با شد. آنگاه $\theta^{L} \leq \theta^{*} \leq \theta^{U}$.

برهان:جهت اثبات $\theta^{L} \leq \theta^{*} \leq \theta^{U}$ کافی ا ست ثابت کنیم که جواب بهینه مدل (۱۶) متناظر با θ^{U} یک جواب شدنی برای مدل(۱۴) متناظر با θ^{*} می باشد. همچنین جواب بهینه مدل (۱۴)یک جواب شدنی برای (۱۵)است.

اثبات $\theta^{*} \leq \theta^{U}$: فرض کنید λ_i^{U*} جواب بهینه مدل (۱۷) با شد. ثابت میکنیم که یک جواب شدنی برای مدل خواهد (۱۶) بود. بدین منظور درنظر بگیرید:

$$\sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x_{ij} \leq \sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x^{U}{}_{ij} \qquad (۱۷)$$

s.t

$$\sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x_{ij} \leq \sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x^{U}{}_{ij} \leq \mathcal{E}^{t} x^{L}{}_{tj} \leq \mathcal{E}^{t} x_{\ \ tj}$$

باتوجه به رابطه (۱۷) داریم:

$$\sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x_{ij} - \mathcal{E}^{t} x_{tj} \leq 0 \qquad (۱۸)$$

که نشان می دهدجواب نی برای مدل(۱۶) می باشد درنتیجه $\theta^{*} \leq \theta^{U}$

اثبات $\theta^{L} \leq \theta^{*}$: فرض کنید λ_i^{*} جواب بهینه مدل (۱۵) با شد. ثابت میکنیم که یک جواب شدنی برای مدل (۱۴)خواهد بود. بدین منظور درنظر بگیرید:

$$\sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x_{ij} \leq \sum_{i=1}^{n} \lambda^{U*}{}_{i}\, x^{U}{}_{ij} \qquad (۱۹)$$

$$\mathcal{E}^{t} x^{U}{}_{tj} \leq \mathcal{E}^{t} x_{tj} \sum_{i=1}^{n} \lambda^{*}{}_{i}\, x^{L}{}_{ij} \leq \sum_{i=1}^{n} \lambda^{*}{}_{i}\, x_{ij} \leq$$

باتوجه به رابطه‌ی (۱۹) داریم:

$$\sum_{i=1}^{n} \lambda^{*}{}_{i}\, x^{L}{}_{ij} - \mathcal{E}^{t} x_{tj} \leq 0 \qquad (۲۰)$$

که نشان می دهد λ_i^* یک جواب شدنی برای مدل(۱۵) می باشد و درنتیجه $\theta^L \leq \theta^*$ لذا حکم برقرار بوده و $\theta^L \leq \theta^* \leq \theta^U$.

مدل (۱۳) که برای کلاس رد داده های غیر بازه ای استفاده می شد برای کلاس رد داده های بازه ای یعنی با فرض وجود کران های بالا و پایین برای ورودی ها و خروجی ها به صورت مدل (۲۱) است.

$$
\begin{aligned}
&\text{Max } \mathcal{E}^t ، \quad t = 1 ، \dots ، n \\
&\text{S.t} \\
&\textstyle\sum_{i=1}^{n} \lambda_i x_{ij} - \mathcal{E}^t x_{tj} \geq 0 \quad j = 1 ، \dots ، m \quad t = 1 ، \dots ، n \\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1 \\
&\lambda_i \geq 0 \quad i = 1 ، \dots ، n \\
&x_{ij} \in \left[x_{ij}^L ، x_{ij}^U\right] \quad i = 1 ، \dots ، m \quad j = 1 ، \dots ، s
\end{aligned}
\tag{۲۱}
$$

بسته به موقعیت واحد تحت ارزیابی نسبت به سایر واحدها، همچنین به دلیل بازه ای بودن داده ها کران بالا و کران پایینی را باید برای مقدار بهینه تابع هدف مدل (۲۱) بدست آوریم. مقدار بهینه تابع هدف مدل(۲۱)متناظر با کران های بالا و پایین را به ترتیب θ^U و θ^L نامیده می شود و دو حالت زیر متناظر با این مقادیر درنظر گرفته شده است.

حالت اول: واحد تحت ارزیابی دربدترین حالت خود($x_{ik} = x^U{}_{ik}$ و $y_{rk} = y^L{}_{rk}$) قرار داشته. وبقیه واحدها در بهترین موقعیت خود ($x_{ik} = x^L{}_{ik}$ و $y_{rk} = y^U{}_{rk}$) واقع شده باشند.لذا از مدل(۲۲) استفاده می کنیم.

$$
\begin{aligned}
&\theta^L = \text{Max } \mathcal{E}^t ، \quad t = 1 ، \dots ، n \\
&\text{S. t} \\
&\textstyle\sum_{i=1}^{n} \lambda_i x_{ij}^L - \mathcal{E}^t x_{tj}^U \geq 0 \quad j = 1 ، \dots ، m \quad t = 1 ، \dots ، n \\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1
\end{aligned}
\tag{۲۲}
$$

$\lambda_i \geq 0 \quad i = 1 ، \dots ، n$

حالت دوم: جهت یافتن کران بالای تابع هدف θ^U در مدل (۲۱)بایستی واحد تحت ارزیابی در بهترین موقعیت خود ($x_{ik} = x^L{}_{ik}$ و $y_{rk} = y^U{}_{rk}$) و بقیه واحدها در بدترین موقعیت خود ($x_{ik} = x^U{}_{ik}$ و $y_{rk} = y^L{}_{rk}$) واقع باشند. مدل (۲۳) θ^U را بدست می آورد.

$$
\begin{aligned}
&\theta^U = \text{Max } \mathcal{E}^t ، \quad t = 1 ، \dots ، n \\
&\text{S.t} \\
&\textstyle\sum_{i=1}^{n} \lambda_i x_{ij}^U - \mathcal{E}^t x_{tj}^L \geq 0 \quad j = 1 ، \dots ، m \quad t = 1 ، \dots ، n \\
&\textstyle\sum_{i=1}^{n} \lambda_i = 1 \\
&\lambda_i \geq 0 \quad i = 1 ، \dots ، n
\end{aligned}
\tag{۲۳}
$$

در ادامه ثابت می شود اگر ورودی و خروجی ها دارای کران بالا و پایین باشند. آنگاه می توان برای مقدار بهینه تابع هدف مدل (۲۱) کران بالا θ^U و پایین θ^L را بدست آورد.

قضیه(۴-3): اگر θ^* جواب بهینه مدل (۲۱) باشد. همچنین θ^U و θ^L جواب های بهینه مدل (۲۳) و(۲۲) باشد. آنگاه $\theta^L \leq \theta^* \leq \theta^U$.

۸- نتیجه گیری

از جمله تعریفی درباره تحلیل پوششی داده ها و واحدهای تصمیم گیرنده که در این مبحث آورده شده است. و تاریخچه این موضوع علمی و مراحل پیشرفت تحلیل پوششی داده ها از زمان ابداع تا اکنون است. و همینطور مزایا و معایب روش های تحلیل پوششی داده ای است. روش های گردآوری اطلاعات و اینکه چه نوع داده هایی باید برای تحلیل با استفاده از DEA مورد استفاده قرار می گیرند. و روش تجزیه و تحلیل داده ها برا ساس مدل های تحلیل پوششی داده ها که با استفاده از نرم افزار GAMS است هم گنجانده شده اند. در ادامه این مبحث مجموعه امکان تولید که مجموعه فعالیت‌های شدنی است که هر کدام از مدل‌های DEA به یک مجموعه امکان تولید یکتا وابسته هستند آورده شده است. و همینطور اصولی که تحت آنها این مجموعه امکان ایجاد می شوند. بعد از مجموعه امکان مدل CCR به همراه تمام حالت های این مدل که شامل مدل های پوششی و مضربی در ماهیت های ورودی محور و خروجی محور و اینکه در چه شرایطی از این مدل ها دارای جواب بهینه هستیم به این صورت است که در مدل پوششی ورودی محور با در صورتی تابع هدف در محدوده $1 \leq \theta^* <$ ۰ باشد دارای جواب بهینه هستیم. و شرط لازم کارایی تحت مدل فوق این است که $\theta^* = 1$. برای حل مشکل ناکارایی واحدهایی که تابع هدف آنها مقداری کمتراز ۱ دارد راه حلی که ارائه شده است این است که از مدل پوششی CCR با ماهیت خروجی محور استفاده کنیم. بعد از مدل پوششی CCR مدل مضربی آن در هردو حالت ورودی محور و خروجی محور ارائه شده است. مدل BCC دومین مدل تحلیل پوششی داده ها است که در دوحالت پوششی و مضربی با دو ماهیت ورودی محور و خروجی محور وجود دارد.

با استفاده از مدل های کلاسیک تحلیل پوششی داده ها قابل رتبه بندی نیستند .بدیهی است که رتبه بندی واحد های کارا به جهت تعیین کارا ترین واحد ها ، اهمیت زیادی دارد و جهت انجام این طبقه بندی از روش اندر سون- پتر سون استفاده می شود. همچنین به تعریف طبقه بندی و داده کاوی و روش های داده کاوای و شناسایی الگو پرداخته شده است. روش های داده کاوی شامل طبقه بندی و رگرسیون و رگرسیون منطقی است. و همینطور مبحث خوشه بندی و تفاوت آن با دسته بندی بیان شده است.بعد از بیان مسئله طبقه بندی الگوریتم های دسته بندی درخت تصمیم گیری که مدل خود را بر اساس یک درخت پیاده سازی می کند . در این الگوریتم با توجه به مجموعه آموزش یک درخت بر اساس ویژگی های مختلف آن در ست می شود که با استفاده از این درخت باید بتوان یک عضو جدید را در دسته خاصی طبقه بندی کرد. و الگوریتم KNN که مبنای آن پیدا کردن تعداد معینی از نزدیکترین عناصر موجود در جامعه آماری به عنصر جدید واردشده در آن جامعه است که بر اساس آن بتوان نزدیکترین داده (داده ها) موجود به عنصر جدید را از لحاظ ویژگی های مختلف پیدا کرد تا عنصر جدید را در همان طبقه ای قرار داد که عناصر نزدیک به آن قرار دارنددر ادامه مبحث تحلیل پوششی داده ها پرداخته شده است. که شامل مدل CCR و BCC هستند. در مدل CCR هدف اندازه گیری و مقایسه کارایی نسبی واحدهای سازمانی با چندین ورودی و خروجی است. و همین طور در ادامه به طبقه بندی داده ها که تقسیم بندی داده‌ها و قرار دادن آنها در بخشهای متفاوت اما با توجه به یک الگوی از پیش تعریف شده می‌باشد. روشی که برای طبقه بندی ارائه

شده است به این صورت است که از ابتدا تصمیم گیرنده یک آستانه قطعی را برای داده ها در نظر می گیرد که طبق آن داده ها را در دوکلاس رد و پذیرش طبقه بندی می کند. سپس داده های کلاس پذیرش و رد را هرکدام تو سط مدل مربوط به کلاس خود شان اجرا می کنند. تا داده هایی که روی مرز پذیرش یا رد و داده هایی که بالای مرز پذیرش و رد و داده هایی که زیر مرز پذیرش و رد هستند را د سته بندی کنند سپس با استفاده از یک روش دو مرحله ای داده هایی که به عنوان حالت م شکل طبقه بندی یا به عبارتی م شترک در بین دو کلاس ه ستند را شنا سایی می کند. طبقه بندی تعاملی کاربران را قادر می‌سازد تا مشمول دو موضوع اصلی طبقه بندی شوند که در ابتدا به کاربر اجازه می‌دهد انتخاب و بیان نماید که کدام گرانول طبقه بندی شود. کاربر می‌تواند اولویت یا علاقه خود را برای اجرای طبقه بندی با انتخاب آزاد گرانول جهت طبقه بندی ارائه دهند. ثانیاً، سیستم به کاربر اجازه می هد تا انتخاب و بیان نماید که چگونه طبقه بندی انجام شود که این کار را با فراهم آوری ارزیابی‌های متعدد آماری انجام می‌دهد. در ادامه مبحث تحلیل پوششی داده ها برای داده های بازه ای بررسی شدند و از مدل تحلیل پوششی داده ای BCC برای محاسبه کارایی نسبی داده های بازه ای استفاده می شود. که در این مدل به دلیل بازه ای بودن داده‌ها، کران بالا و پایینی را برای مقدار بهینه تابع هدف را بای ستی به د ست آوریم. برای این حالت دو مقدار Θ^U که در آن دادهای واحد تصمیم گیرنده در بهترین حالت و بقیه واحدها در بدترین حالت هستند و Θ^L که واحد تصمیم گیرنده دربدترین حالت و بقیه واحدها در بهترین حالت هستند. را برای کران های بالا و پایین به د ست می آوریم. در ادامه یک مثال عددی با داده های بازه ای برای محاسبه کارایی که براساس مدل های استفاده شده که برای طبقه بندی آورده شده است پس از محاسبه کران بالا و پایین تابع هدف برای این داده ها میانگینشان را به عنوان مقدار تابع هدف بهینه بدست می آید. و از این مقادیر برای طبقه بندی داده ها در دو کلاس پذیرش و رد و حالت هایی که بین این دو م شترک ه ستند ا ستفاده می شود که برای اولین بار طبقه بندی با استفاده از داده های بازه ای بیان شده اند.

منابع:

آذر، عادل؛ پاییز و زم ستان ۱۳۷۹، تحلیل پو ش شی داده‌ها و فرایند تحلیل سل سله مراتبی، مجله مدیریت دانشگاه علامه طباطبایی، ص ۱۴۳-۱۲۹.

حسین‌زاده لطفی، فرهاد.، بوبهرز، زهرا.، رستمی، محمد.، و احدزاده نمین، محمد. (۱۳۹۰). معرفی یک روش DEA-TOPSIS برای داده‌های بازه‌ای. سومین همایش ملی تحلیل پوششی داده‌ها، فیروزکوه.

حمزه پور، مهدی؛ محمدی، روح اله؛ تابستان ۱۳۹۱، برر سی کارایی شعب سازمان بیمه تأمین اجتماعی در استان تهران با استفاده از روش تحلیل پوششی داده‌ها (DEA)، مجله الکترونیک پارس مدیر، شماره ٤، ص ۹٤-۱۱۷.

مهرگان، محمدرضا. تحلیل پوششی داده‌ها (مدل‌های کمی در ارزیابی عملکرد سازمانها)، تهران: نشر کتاب دانشگاهی، ۱۳۹۱.

Banker, R. D., Charnes, A., Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science., Vol. 30, No. 9, pp. 1078–1092.
Charnes, A., Cooper, W. W., Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research., Vol. 2, pp. 429 – 444.

Tone, K. (2000). Data Envelopment Analysis. United States of America: Kluwer Academic Publishers.