

## ارائه مدلی با رویکرد داده کاوی جهت اعتبار سنجی دفترچه های خدمات درمانی (مطالعه موردی سازمان تامین اجتماعی شعبه کوار)

جواد ایرانبان\*<sup>۱</sup>، مریم اسماعیل نژاد<sup>۲</sup>، محمدرضا معتدل<sup>۳</sup>

دکترای مدیریت صنعتی، عضو هیئت علمی مدعو واحد الکترونیکی، دانشگاه آزاد اسلامی، واحد الکترونیکی، گروه مدیریت فناوری اطلاعات، تهران، ایران، [airanban@yahoo.com](mailto:airanban@yahoo.com)

کارشناسی ارشد مدیریت فناوری اطلاعات (IT)، دانشگاه آزاد اسلامی، واحد الکترونیکی، گروه مدیریت فناوری اطلاعات، تهران، ایران، [esmailnezhad@ymail.com](mailto:esmailnezhad@ymail.com)

دکترای مدیریت صنعتی، عضو هیئت علمی مدعو واحد الکترونیکی، دانشگاه آزاد اسلامی، واحد الکترونیکی، گروه مدیریت فناوری اطلاعات، تهران، ایران، [Dr.motadel@yahoo.com](mailto:Dr.motadel@yahoo.com)

### خلاصه

با توجه به اینکه یکی از خدمات تامین اجتماعی کمک به درمان بیمه شدگان میباشد و این امر در حال حاضر بواسطه دفترچه های درمانی کاغذی و به یاری پزشکان و مراکز درمانی تحت پوشش، یا افراد و شرکتهای دارای قرارداد با تامین اجتماعی حاصل میشود، لذا در این تحقیق سعی شده است که مدلی جهت ارائه هر چه دقیقتر تاریخ اعتبار دفترچه درمان ارائه گردد. در حال حاضر تاریخ مذکور توسط سیستم از طریق روندی درختی و ایستا تعیین میگردد. مدلی که در این تحقیق ارائه شده، پویا و از طریق داده کاوی می باشد. این روش از دقت بیشتری برخوردار است. در این تحقیق از نرم افزار کلمنتاین ۱۲ و الگوریتم تحلیل عاملی<sup>۴</sup> استفاده شده است و فرمولی جهت اعتبار سنجی دفترچه های تامین اجتماعی و مدلی جهت تامین اعتبار بیمه شده از طریق الگوریتم داده کاوی به دست آمده است. همچنین با توجه به اینکه داده های شعب تامین اجتماعی به طور مستقل در سرور های هر شعبه موجود است و استفاده از نتایج داده کاوی در روند جاری امور مکانیزه شعب مستلزم ارائه راهکاری مناسب است در این پژوهش روش اجرایی حل این مشکل به دست آمده است. این روش جهت سازمانهای مشابه نیز قابلیت استفاده دارد.

کلمات کلیدی: تامین اجتماعی، داده کاوی، الگوریتم تحلیل عاملی، دفترچه درمانی

### ۱. مقدمه

یکی از ابزارهای ارائه مزایا به ارباب رجوع، دفترچه های تامین اجتماعی میباشد که اعتبار استفاده از آن منوط به تاریخ اعتبار آن میباشد که به صورتهای گوناگون در آن ثبت میگردد. در دفترچه های درمانی در حال حاضر تاریخ اعتبار، بسته به پارامترهای سابقه و کارگاهی که بیمه شده در آن شاغل است و سن افراد تحت تکفل (تبعی) تعیین و در تمامی برگه های آن چاپ میشود هدف این تحقیق بدست آوردن پارامترهای دخیل در تاریخ اعتبار درمان و تعیین فرمول مربوطه و

<sup>۱</sup> Javad Iranban

<sup>۲</sup> Maryam Esmailnezhad

<sup>۳</sup> Mohammadreza Motadel

<sup>۴</sup> PCA

\*نویسنده مسئول مقاله: جواد ایرانبان [airanban@yahoo.com](mailto:airanban@yahoo.com)

مسئول مکاتبات: مریم اسماعیل نژاد [esmailnezhad@ymail.com](mailto:esmailnezhad@ymail.com)

تشریح کلیه مزایا و معایب انواع واسطه های شناسایی بیمه شده (دفترچه کاغذی/کارت هوشمند/سایر) میباشد.

برای پیاده سازی و اجرای پروژه های داده کاوی نرم افزار های مختلفی وجود دارد. کلمتاین یکی از قوی ترین این نرم افزارها می باشد [۱]. در این تحقیق مساله این است که با استفاده از داده های واقعی سازمان، قصد کشف روابطی پنهان بین داده ها را داریم. تحقیق، مربوط به بخش اعتبار دفترچه میباشد و به دنبال کشف مدلی جهت اعتبار سنجی دفترچه های تامین اجتماعی با استفاده از داده کاوی، بمنظور دقت بیشتر در تعیین تاریخ اعتبار میباشیم. سازمان تامین اجتماعی از دو گروه متخصص تشکیل شده که عبارت است از کارمندان بخش کامپیوتر و دیگری کارمندان واحد. گروه اول کارهای تخصصی چون کد نویسی، اعمال شبکه، محافظت از داده های سازمان وغیره را انجام میدهند و گروه دوم از کارمندانی تشکیل شده است که با سیستم بیمه ای جهت راه اندازی کار ارباب رجوع کار میکنند. در این تحقیق از هر دو گروه استفاده شده بدین صورت که جهت تحلیل و کارهای تخصصی گروه اول و جهت اطلاعات سازمانی از گروه دوم بهره گرفته شده است. سعی بر این است که با استفاده از فنون داده کاوی، قواعد و روابط میان تاریخ اعتبار دفترچه و عوامل موثر بر آن شناسایی شود و اطلاعات مربوطه از طریق الگوریتمهای داده کاوی بررسی و نتایج بصورت تحلیلی ارائه شود. ابتدا با مطالعه و بررسی پایگاههای داده، جداول اطلاعاتی موثر در اعتبار دفترچه شناسایی شد و سپس با استفاده از کدهای SQL فیلدهای مورد نظر از هر بانک در کنار هم قرار داده شد و جدولی سفارشی جهت داده کاوی آن ساخته شد. در گام بعدی نیز با توجه به مطالعات انجام شده پیشین درباره داده کاوی و کاربرد آن بر شرکتهای بیمه ای و پایگاه داده حجیم، مدل مفهومی مناسب برای اجرای داده کاوی طراحی شد و سعی بر الگوبرداری مناسب از مطالعات گذشته و رفع مشکلات و نواقص مدلهای پیشین داریم. از طرفی از زمانی که رایانه در ذخیره سازی داده های سازمان بکار رفت، حجم داده در پایگاه داده ها چند برابر شد بطوریکه هم اکنون برخی پایگاه داده ها شامل چند صد میلیون رکورد میباشند. در این شرایط فن داده کاوی، ما را قادر می سازد تا از داده های خود بهره برداری نموده و از آن به عنوان ابزاری برای پشتیبانی فرآیند تصمیم گیری استفاده کنند، بنابراین انبار داده تشکیل شد. (شهرابی، ۱۳۹۰) براساس فنون داده کاوی، داده های پیش پردازش و پاکسازی شدند، یعنی متغیرهایی که برای هدف مورد نظر مناسب نبودند، حذف و اطلاعات تکراری یکی شدند. پس از اصلاح متغیرها، رکوردهای مورد پالایش، حذف و اصلاح قرار گرفتند. پس از آماده سازی، داده ها به نرم افزار کلمتاین داده شد و با استفاده از مدل تحلیل عاملی نتایج حاصل شد و سپس مورد تحلیل قرار گرفت.

## ۲. نحوه جمع آوری داده

داده هادراین تحقیق در دو مرحله گردآوری شده اند:

پرسش نامه

پرسشنامه مجموعه ای است از سولات کتبی و غالباً مبتنی بر گزینه های مشخص که پاسخ دهنده جوابهای خود را بر آن درج می کند. به هنگامی که پژوهشگر دقیقاً می داند در پی چیست و چگونه باید متغیرهای خود را اندازه گیری کند. پرسشنامه ابزاری کارآمد برای گردآوری اطلاعات به شمار می رود. پرسشنامه را می توان به طور حضوری یا از طریق پست میان پاسخ دهندگان توزیع کرد. همچنین می توان با اجرای آن به طور الکترونیکی داده های لازم را گردآوری کرد.

به اعتقاد ساندرز و همکاران (۲۰۰۰) بسیاری از افراد از پرسشنامه برای گردآوری داده ها استفاده میکنند بدون اینکه سایر شیوه های گردآوری داده ها را در نظر بگیرند. به اعتقاد این صاحب نظران باید همه شیوه های گردآوری داده را ارزیابی کرد و متناسب با سوالها و اهداف پژوهش بهترین شیوه را انتخاب کرد. از پرسشنامه میتوان برای پژوهشهای توصیفی یا تبیینی (پیمایش و همبستگی) استفاده کرد. پژوهشهای توصیفی (در استراتژی پیمایش) نظیر پژوهشهایی که از پرسشنامه نگرش سنجی یا افکار سنجی و پرسشنامه های اعمال سازمانی استفاده میکنند، پژوهشگر را به شناسایی و توصیف تغییر پذیری در پدیده های متفاوت قادر میسازد. در مقابل پژوهشهای تبیینی و تحلیلی (همبستگی و تجربی) پژوهشگر را قادر میسازد رابط بین متغیرها خصوصاً رابطه علت - معلولی را بررسی کند. (گیل و جانسون، ۱۹۹۷) در این تحقیق از پرسشنامه و مقیاس فاصله ای استفاده شده است.

پرسشنامه این تحقیق از نوع حضوری است چون این پژوهش محدود به حوزه ی سازمانی است و مزیت این نوع این است که میتوان همه پاسخها را در یک

دوره زمانی کوتاه گردآوری کرد [۲]. سولات این پرسشنامه بسته میباشند. در سولات پرسشنامه سعی شده که پاسخ دهنده را به مطابق جدول ۱ هدایت کنیم:

### جدول ۱. نمودار هدایت پاسخ دهنده



شیراز - دی ماه ۹۲

دسته الف	دسته ب	دسته ج	دسته د
روش کنونی	روش آرمانی و مزایا	توانمندیها و نیازمندیها	پیاده سازی و جایگزینی

(دسته الف) سوالات ۱ و ۲ و ۳ و ۴

(دسته ب) سوالات ۵ و ۷-۱

(دسته ج) سوال ۶

با توجه به نظر سنجی انجام شده کارمندان روش فعلی را موثر و کارا نمیدانند ضمناً پتانسیل و توانایی سازمان را در حرکت به انجام روش جدید بسیار خوب برآورد کرده اند و آمادگی پذیرش روش جدید وجود دارد. از طرفی امکاناتی که داریم نشاندهنده این است که امکان انجام سریع و کم هزینه می باشد. سیستمهای سازمانی

اطلاعات شعب با استفاده از کد های sql از پایگاه داده استخراج شده است. چون داده های مورد نیاز در فایل پایگاه داده ای اوراکل شعبه موجود بودند ابتدا با استفاده از دستورات پرس و جوی ساختار یافته<sup>۱</sup> یک سری فایل با فرمت جدا شده با کاما (CSV)<sup>۲</sup> که قابل استفاده در نرم افزار میکروسافت اکسل باشد تهیه شدند. از نرم افزار اکسل برای تبدیل فرمت و محتوای داده ها استفاده گردید. مصادیق این تبدیلات در جدول ۲ آمده است.

جدول ۲- نمونه ای از تبدیلات داده ای اکسل، به کار گرفته شده در تحقیق

ردیف	نوع تبدیل	دلیل کاربرد	عملیات به کار گرفته شده
۱	نگاشت انواع فاصله دار به مقادیر ترتیبی	مثلاً اعداد ۱، ۳، ۷ را به اعداد ترتیبی ۱، ۲ و ۳ نگاشت میدهیم زیرا در غیر اینصورت برخورد نرم افزار داده کاوی با این فواصل متفاوت خواهد بود.	تهیه جدول نگاشت و استفاده از تابع VLookup
۲	نگاشت نوع عددی به تاریخ	مثلاً عدد ۸۹۰۶۰۵ که به تاریخ ۸۹/۰۶/۰۵ تبدیل شده است.	fdate(a)
۳	تفریق دو تاریخ	برخی فیلدها با تفریق دو فیلد تاریخ بدست آمده است.	Diff(a,b)

همچنین برخی تبدیلات نیز به طور مستقیم در نرم افزار داده کاوی کلمنتاین انجام شده است که در جدول ۳ نشان داده شده است.

جدول ۳- انواع داده ای مورد استفاده و تعاریف آنها به اختصار

ردیف	نام متغیر (قلم داده ای)	مفهوم	مثال نمونه
۱	تاریخ درخواست	تاریخی تقاضای اعتبار یا صدور دفترچه بیمه شده	۱۳۹۲/۰۴/۰۷
۲	تاریخ تولد	روز، ماه، سال تولد بیمه شده.	۱۳۵۲/۰۶/۰۵
۳	تاریخ اعتبار	تاریخی که دفترچه بیمه شده تا آن زمان معتبر است و امکان استفاده دارد.	۱۳۹۲/۰۱/۰۵

فیلد های جدیدی با انجام عملیات بر فیلد های موجود بدست آمده است که عبارتند از:

(۱) تاریخ درخواست بیمه شده - تاریخ تولد فرد = سن کنونی فرد بیمه شده

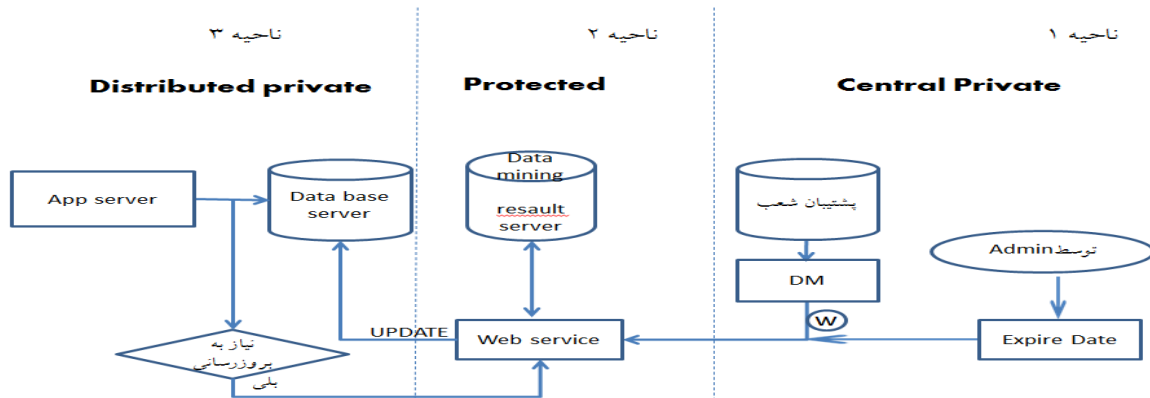
(۲) تاریخ درخواست - تاریخ اعتبار = روزهای اعتبار دفترچه بیمه شده

<sup>۱</sup> SQL = Structured Query Language

<sup>۲</sup> CSV = Comma separated values

### ۳. اعتبار مدل و قابلیت اطمینان مدل

مدل به کار رفته در این تحقیق بر اساس استاندارد CRISP-DM میباشد و اعتبار و قابلیت این مدل به دلیل استاندارد بودن تایید شده است. شکل ۱ مراحل انجام کلی روند سازمان و جایگاه داده کاوی در عملیات کلی را نشان میدهد.



شکل ۱ - مدلی کلی جهت اعتبار سنجی دفترچه

مزایای مجزای بودن سرور نتیجه از شعب عبارت است از:

جدا سازی پروسه داده کاوی از کل عملیات، بنابراین تغییرات در یک مرکز جدا از شعب، انجام میشود. (یکسان سازی شیوه کل کشور)  
 انقضاء<sup>۱</sup> به صورت دستی انجام میشود بنابراین با توجه به توان پردازشی مرکز میتوان این تاریخ را کم یا زیاد در نظر گرفت. چون سرور نتیجه<sup>۲</sup> از سرور شعب جدا شده، توان پاسخگویی در هر لحظه را دارد بنابراین میتوان امکانات پایین تری برای این سرور در نظر گرفت.  
 دسترسی به سرور اصلی از طریق اینترنت/اینترنت انجام میگردد که باعث دسترسی راحت تر میشود.  
 اطلاعات داده کاوی، شامل اطلاعات سازمان است زیرا بخش اطلاعاتی در سرور عملیاتی آست و مجزا میباشد، بنابراین خطراتی ندارد.  
 سه ناحیه مجزا در نمودار در نظر گرفته شده است:

سرور پایگاه داده<sup>۴</sup>: اطلاعاتی چون شناسه، تاریخ انقضاء و وزن متغیر را شامل میشود:

شناسه: شناسه<sup>۵</sup> به معنای نسخه ماتریس وزن، که در سراسر کشور یکسان است، میباشد. این فیلد به عنوان یک فیلد راجعی در محاسبات بعدی استفاده میشود. از لحاظ جداول پایگاه داده این فیلد، کلید اصلی میباشد. اطلاعات شناسه در مسائل بازرسی و پیگیریها (به عنوان مثال کشف تقلب) به کار میآید.

تاریخ انقضاء: اعتبار استفاده از فرمول را تعیین میکند. یعنی تا این تاریخ میتوانند از این فرمول استفاده کنند و پس از آن ارزشی ندارد. به منظور توان

<sup>۱</sup> Expire

<sup>۲</sup> Result

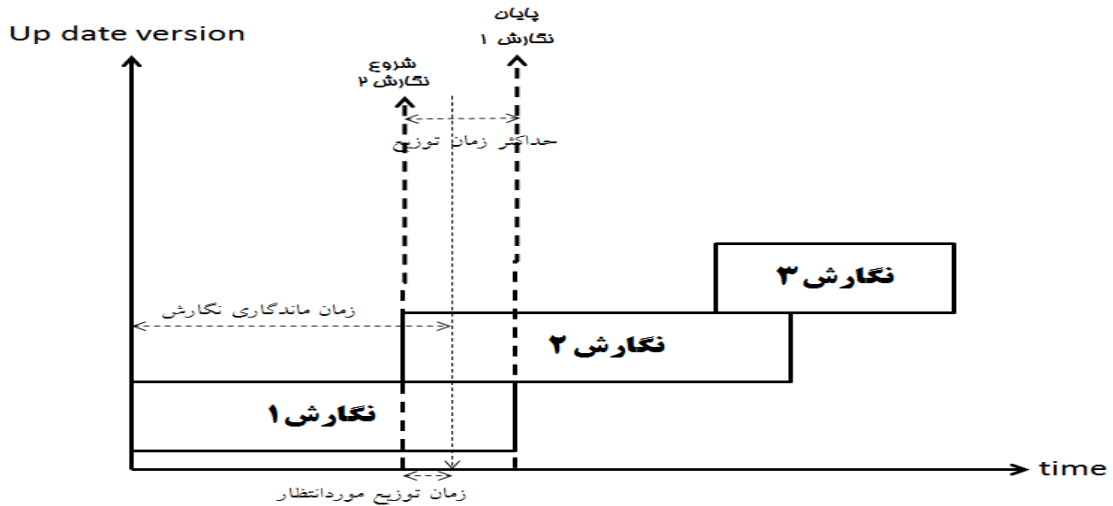
<sup>۳</sup> Application

<sup>۴</sup> Data base server

<sup>۵</sup> identification

پاسخگویی عملیاتی سیستم، مرکز داده کاوی بایستی قبل از انقضاء آخرین ماتریس وزنی نسبت به روز رسانی آن مقدار اقدام، و در نتیجه ماتریس وزنی مورد نظر را از اعتبار ساقط کرده و آن مقدار جدید را با اعتبار جدید جایگزین نماید.

شکل ۲ جهت فهم بهتر مفهوم تاریخ انقضاء میباشد:



شکل ۲. نمودار رابطه زمان و نگارش

$$T_{dn} = t_{d \text{ last } n} - t_{d \text{ first } n} = \text{زمان انتظار واقعی} \quad (۳)$$

$$dn+1 = \text{start } n+1 - \text{expire } n = \text{حد اکثر زمان توزیع} \quad (۴)$$

$$t_{d \text{ last } n} = \text{زمان درج فرمول در آخرین شعبه} \quad (۵)$$

$$t_{d \text{ first } n} = \text{زمان درج فرمول در اولین شعبه} \quad (۶)$$

با استفاده از به روز رسانی خودکار و نگهداری خطوط بصورت برخط (online) تلاش به سمت کاستن از میزان  $dn$  ها میباشد.

فرض کنیم فرمول شماره ۱ (نگارش ۱) در تاریخ ۹۲/۰۶/۲۵ منقضی میشود. همکاران مرکز داده کاوی یک نسخه جدید (نگارش ۲) مورخ ۹۲/۰۶/۱۲ برای

توزیع آماده کرده اند، از اولین شعبه ای که این فرمول در آن درج و عملیاتی میشود، تا آخرین شعبه حداکثر ۱۳ روز فرصت وجود دارد. هدف این است که این زمان را تا آنجا که ممکن است کم کنیم تا قوانین در کل کشور بصورت یکسان اجرا شود. بنابراین پیشنهاد میشود از تکنیکهای به روز رسانی خودکار<sup>۱</sup>، استفاده از خطوط داده ای و شبکه موجود بهره گیری شود.

وزن متغیر:  $k$  متغیر داریم که از دید داده کاوی این متغیرها در محاسبه فرمول نهایی موثر بوده اند، بنابراین برای هر  $X_i$  یک  $W_i$  وجود دارد بطوریکه:

$$W = [w_1, w_2, \dots, w_k] \quad (۷)$$

$$F = \sum_{i=1}^n W_i X_i \quad (۸)$$

#### ۴. آماده سازی فایل

<sup>۱</sup> replication



جهت خارج کردن فیلدهای مورد نظر از بانک اطلاعاتی از کد نویسی SQL استفاده شده است. کد مربوطه به صورت زیر میباشد:

```

select t1.فیلد اول,t1.فیلد دوم,t1.فیلد سوم,t1.فیلد سوم,t1.فیلد چهارم,t1.فیلد پنجم,t1.فیلد ششم,t2.فیلد اول,t2.فیلد دوم,t2.فیلد سوم,t2.فیلد چهارم,t2.فیلد پنجم,t2.فیلد ششم,t2.فیلد اول,t2.فیلد دوم,t2.فیلد سوم,t2.فیلد چهارم,t2.فیلد پنجم,t2.فیلد ششم,t3.فیلد اول,t3.فیلد دوم,t3.فیلد سوم,t3.فیلد چهارم,t3.فیلد پنجم,t3.فیلد ششم,t3.فیلد اول,t3.فیلد دوم,t3.فیلد سوم,t3.فیلد چهارم,t3.فیلد پنجم,t3.فیلد ششم,t3.فیلد اول,t3.فیلد دوم,t3.فیلد سوم,t3.فیلد چهارم,t3.فیلد پنجم,t3.فیلد ششم,t4.فیلد اول,t4.فیلد دوم,t4.فیلد سوم,t4.فیلد چهارم,t4.فیلد پنجم,t4.فیلد ششم,t4.فیلد اول,t4.فیلد دوم,t4.فیلد سوم,t4.فیلد چهارم,t4.فیلد پنجم,t4.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم,t5.فیلد اول,t5.فیلد دوم,t5.فیلد سوم,t5.فیلد چهارم,t5.فیلد پنجم,t5.فیلد ششم
from t1 جدول اول,t2 جدول دوم,t3 جدول سوم,t4 جدول چهارم,t5 جدول پنجم
where t1.فیلد اول=t2.فیلد اول
and t2.فیلد اول=t3.فیلد اول
and t3.فیلد اول=t4.فیلد اول
and t4.فیلد اول=t5.فیلد اول

```

لازم به ذکر است که به علت مسائل امنیتی مجاز به بازگویی دقیق نام جداول و فیلدها نبوده ایم، بنابراین از اسامی مجازی چون فیلد اول و یا جدول اول

استفاده کرده ایم.

فایل با پسوند CSV استخراج شده و سپس با استفاده از قابلیت دریافت داده خارجی<sup>۱</sup> در اکسل ۲۰۱۰ به فایلی با پسوند XLS تبدیل شده است. قابل ذکر است

که جهت ورود فایل به نرم افزار کلمنتاین حتما فایلهای باید با پسوند Xls باشد نه Xlsx.

جهت آماده سازی فایل بدست آمده تغییراتی در آن ایجاد شد به عنوان مثال فیلد تاریخ تولد به این صورت از بانک اطلاعاتی خارج شد ۱۳۴۸۰۲۲۸ به این

معنا که بیمه شده متولد بیست و هشتم اردی بهشت ماه هزار و سیصد و چهل و هشت میباشد که این برای نرم افزار کلمنتاین مفهوم نیست لذا با استفاده از توابع اکسل

که به صورت Add ins موجود میباشد از جمله Fdate, J\_DIFF, تبدیل انجام شد. بدینصورت که تابع Fdate تاریخ را بدین شکل نمایش می دهد

(۱۳۴۸/۰۲/۲۸). این تبدیل برای تمام فیلدهایی که شامل تاریخ میشدند انجام شد و سپس با استفاده از تفاضل تاریخ تولد فرد و تاریخی که جهت اعتبار دفترچه درمان

مراجعه کرده است، سن شخص بدست آمد. عمل تفریق دو تاریخ با استفاده از تابع J\_DIFF انجام گرفت. در نهایت فایلی با ۴۳ ستون جهت تحلیل با کلمنتاین

آماده شد.

## ۵. بدست آوردن فرمولی دینامیک با استفاده از داده کاوی

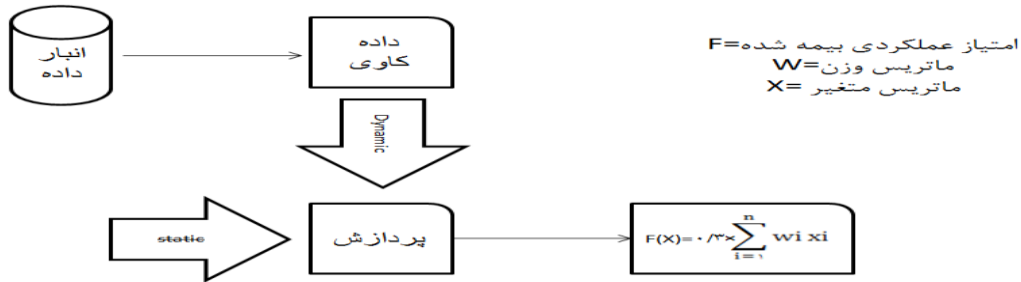
مدل اکنون اعتبار سنجی دفترچه ها در برنامه سازمان شامل مجموعه ای از اگرها میباشد، که مجموعه ای از شرایط را چک کرده و در نهایت نتیجه

میگیرد، در واقع مدل کنونی استفاده از یک درخت بصورت استاتیک میباشد.

در شکل ۳ روند بدست آوردن فرمولی کلی جهت تاریخ اعتبار، با استفاده از داده کاوی و باروندی پویا نشان داده شده است:

<sup>۱</sup> Get external data

<sup>۲</sup> if



شکل ۳. روند بدست آوردن فرمول کلی

برای بدست آوردن فرمول دینامیک از ابزار PCA<sup>۱</sup> (تحلیل عاملی) در کلمنتاین استفاده شده است. خروجی استفاده از این الگوریتم PCA به صورت زیر می باشد:

Initial گویای تمامی اشتراکهای قبل از استخراج است، بنابراین تمامی آنها برابر یک هستند. همانطور که در جدول ۴ مشاهده میشود بیش از نیمی از میزان اشتراکها بالاتر از ۵۰ درصد است و بیانگر توانایی عامل های تعیین شده در تبیین واریانس متغیر های مورد مطالعه است. با وجود این، در بین مقادیر اشتراک، تفاوت هایی نیز مشاهده میشود. برای مثال مقدار اشتراک مربوط به متغیر "روز کارکرد آخرین ماه" ۰/۹۷۴ و متغیر "دوره ارسال لیست" ۰/۹۷۷ می باشد.

جدول ۴- میزان اشتراک اولیه و بعد از استخراج عاملها برای متغیر های وارد شده در تحلیل عاملی

	Initial	Extraction
نوع سابقه	۱	.۶۵۱
روز کارکرد آخرین ماه	۱	.۹۷۴
کارکرد ماهانه	۱	.۹۷۶
نوع شغل	۱	.۰۳۶
نوع کارگاه	۱	.۹۳۶
دوره ارسال لیست	۱	.۹۷۷
نحوه شناسایی	۱	.۹۲۳
نرخ حق بیمه	۱	.۹۷۷
وضعیت	۱	.۷۹۳
شخصیت	۱	.۷۰۶
دوره فعالیت	۱	.۹۱۰
سن بیمه شده	۱	.۷۳۸
سابقه به ماه	۱	.۷۴۷

## ۶. تعریف متغیرها

موثرترین فیلهای بدست آمده از الگوریتم PCA در جدول ۵ آمده است:

جدول ۵- تعریف موثرترین فیلهای بدست آمده از طریق الگوریتم PCA

نام متغیر	تعریف متغیر	واحد اندازه گیری متغیر
دوره ارسال لیست	از آنجا که کارفرما موظف است همه ماهه لیست ماه قبل کارگران را تنظیم و به همراه مبلغ حق بیمه به سازمان پرداخت نماید. لذا مدت زمانی که مهلت قانونی برای ارسال لیست و پرداخت حق بیمه در قانون برای کارفرما در نظر گرفته شده	ماه

<sup>۱</sup> Principal component analysis

	است دوره ارسال لیست ماهانه میباید که طی سال ۱۲ بار تکرار میشود.	
ریال	از آنجا که عمده ترین محل درآمد سازمان وصول حق بیمه میباشد، لذا از طرف سازمان درصدی از حقوق و مزایای مشمول و قابل پرداخت و متعلق به کارگران بعنوان حق بیمه وصول میشود.	نرخ حق بیمه
ماه	تعداد روزهایی که بیمه شده در ماه به کار پرداخته و حق بیمه توسط کارفرما پرداخت شده است.	کارکرد ماهانه
روز	این متغیر در واقع آخرین ارتباط بیمه شده با سازمان رامیرساند که آیا ارتباط بیمه شده تا تاریخ مراجعه اش، با سازمان برقرار بوده است یا نه؟ هرگونه قطع ارتباط باعث عدم اعتبار دفترچه درمان میشود.	روز کارکرد آخرین ماه
حق بیمه	تعیین نحوه ی محاسبات حق بیمه ی کارگاههای تحت پوشش، بکار میرود.	نوع کارگاه
روز	به مدت زمانی که کارگاه فعالیت داشته و دارای کارگر میباشد، دوره فعالیت اطلاق میشود.	دوره فعالیت
ارسال لیست-بازرسی	نحوه شناسایی کارگاه یا از طریق ارسال لیست از طرف کارفرما یا بر اساس گزارش بازرسی صورت می پذیرد.	نحوه شناسایی
فعال-نیمه فعال-غیر فعال	فعال، نیمه فعال و یا غیر فعال بودن کارگاه بر مبنای بدهی و یا عدم بدهی تعریف میگردد. کارگاه فعال یعنی کارگاهی که کارگر دارد (در صورت پرداخت و یا عدم پرداخت حق بیمه) کارگاه نیمه فعال یعنی فاقد کارگر باشد یا بدهی داشته باشد، ولی غیر فعال باشد. کارگاه غیر فعال کارگاهی که بسته و فاقد بدهی باشد.	وضعیت

در پایان دیدیم که با توجه به جدول ۴ وزن هر متغیر داده شد، بنابراین با توجه به فرمول و داده های موجود برای هر بیمه شده، میتوان امتیاز عملکردی را برای تک تک بیمه شدگان در هر شعبه بدست آورد. تناسب اعتبار واقعی بر امتیاز عملکردی برای ۱۰ بیمه شده که به طور اتفاقی از بانک اطلاعاتی انتخاب شده اند حساب شده است و تقریباً ۰/۳ بدست آمده است. بنا براین فرمول محاسبه اعتبار به صورت زیر می باشد:

$$F(X) = 0.3 \times \text{امتیاز عملکردی بیمه شده} \quad (9)$$

$$\text{امتیاز عملکردی بیمه شده} = \sum_{i=1}^n W_i X_i \quad (10)$$

## ۷. نتیجه گیری

تحلیل داده ها با استفاده از الگوریتم تحلیل عاملی بررسی شد. با استفاده از الگوریتم PCA\FACTOR مراحل بدست آوردن امتیاز عملکردی بیمه شده به صورت دینامیک نشان داده میشود. استفاده از روند دینامیک باعث میشود که اعتبار دقیقتری برای بیمه شدگان به دست آید. این الگوریتم مواردی چون فیلدهای مهم را تعیین کرده و میزان اهمیت هر یک را در قالب وزن بیان میکند و مقدار هر متغیر را در بانک اطلاعاتی، مربوط به هر بیمه شده رامیتوان به دست آورد. بنا براین با توجه به فرمول عنوان شده در فصل ۴ با به دست آوردن W و X میتوان میزان امتیاز عملکردی بیمه شده را بدست آورد. همانطور که دیدیم فرمول به دست آمده عبارت است از:

$$F = \sum_{i=1}^n W_i X_i \quad (11)$$

از آنجا که نسبت اعتبار قبلی موجود در بانک اطلاعاتی بر امتیاز عملکردی فرد تقریباً ۰/۳ بدست آمده است لذا فرمول را به این صورت می توان نوشت:

$$F(X) = 0.3 \times \sum_{i=1}^n W_i X_i \quad (12)$$

## ۱۱. قدردانی

از سازمان تامین اجتماعی بابت در اختیار قرار دادن اطلاعات تشکر و قدر دانی میشود.

## مراجع





شیراز - دی ماه ۹۲

۱. مینایی، بهروز؛ نصیری، مهدی؛ حسنی، دانیال و شناسا، ابراهیم (۱۳۹۰). آموزش داده کاوی با کلمنتاین (چاپ اول). انتشارات گروه پژوهشی ساحر.
۲. دانایی فرد، حسن؛ الوانی، سید مهدی و آذر، عادل (۱۳۸۹). روش شناسی پژوهش کمی در مدیریت: رویکردی جامع. تهران: انتشارات صفار- اشراقی.
۳. Oracle, *Oracle Data Mining Concepts, 11g Release 1 (11.1)*: Oracle, ۲۰۰۸.
۴. D. Delen, G. Walker, and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods," *Artificial intelligence in medicine*, vol. ۳۴, pp. ۱۱۳-۱۲۷, ۲۰۰۵.
۵. سکاران، اوما. روشهای تحقیق در مدیریت. ترجمه: محمد صائبی و محمود شیرازی (۱۳۸۹) تهران: انتشارات موسسه عالی آموزش و پژوهش مدیریت و برنامه ریزی.
۶. باقرزاده، فاطمه. (۱۳۸۹) داده کاوی و کشف دانش در *sql server ۲۰۰۸*. (چاپ اول). تهران: نشر اندیشه سرا.