

# استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم تصمیم‌یار (موردکاوی شرکت اتوبوسرانی تهران)

اردلان بادامی، دانش آموخته کارشناسی ارشد، دانشکده مهندسی صنایع، دانشگاه آزاد اسلامی واحد تهران جنوب، تهران، ایران

روزبه قوسی (مسئول مکاتبات)، استادیار، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران

محمد سعیدی مهرآباد، استاد، دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران، تهران، ایران

E-mail: ghousei@iust.ac.ir

دریافت: ۱۳۹۶/۰۶/۱۲ پذیرش: ۱۳۹۶/۱۲/۰۸

## چکیده

در این مطالعه داده‌های تخلفات شرکت اتوبوسرانی تهران با استفاده از تکنیک‌های داده‌کاوی مورد بررسی و به کشف دانش پنهان میان داده‌ها پرداخته شده است. در ابتدا داده‌های تخلفات شرکت‌های پیمانکار اتوبوسرانی تهران در حواصل سال‌های ۱۳۹۱ تا ۱۳۹۴ جمع‌آوری و جهت ورود به نرم‌افزار آماده‌سازی گردید. در مرحله بعد با بهره‌مندی از الگوریتم‌های قوانین انجمنی و درخت تصمیم جهت استخراج دانش استفاده شده است. تعیین شرکت‌های با میزان عملکرد پایین و تخلفات با فراوانی زیاد و ارتباط رخداد آن‌ها در فصول مختلف از جمله نتایج قابل توجهی است که بعنوان خروجی الگوریتم‌های مورد استفاده محسوب می‌گردد. مدیران عالی سازمان‌ها به منظور تصمیم‌گیری در شرایط عدم قطعیت نیاز دارند که از سیستم‌های تصمیم‌یار بهره‌مند شوند. نتایج خروجی تحقیق این مکان را فراهم می‌نماید که به منظور دستیابی یکی از اهداف عالی سازمان که رضایت شهروندی است، مدیران سازمان را در تامین دانش برای اتخاذ تصمیم یاری نمایند. ریشه‌یابی علل رخداد تخلفات و تعیین فصول و ماه‌های پر تکرار از منظر تخلفات، می‌تواند باعث تدوین دستورالعمل‌ها و رویه‌های بهبود در شرکت اتوبوسرانی گردد که ضمن کنترل شرایط موجود، شرکت‌های پیمانکار را ملزم به اجرای دقیق قوانین و کاهش رخداد تخلفات می‌نماید. به هر حال این یک واقعیت است که بمنظور دستیابی به شرایط باثبات و پایدار برای تمامی واحدهای تولیدی، صنعتی و خدماتی همانند مطالعه موردی مورد نظر، توجه به رضایت مشتری لازم و ضروری است.

واژه‌های کلیدی: داده‌کاوی، درخت تصمیم، سیستم پشتیبان تصمیم، قوانین انجمنی

## ۱. مقدمه

امروزه سازمان‌های تولیدی و شرکت‌های خدماتی بمنظور ارتقاء بهره‌وری، لازم است کیفیت ارائه محصولات و خدمات ارائه شده به مشتریان را افزایش دهند. توجه به این مهم باعث می‌گردد ضمن افزایش درآمد و عایدی خود در بازار رقابتی بین شرکت‌ها از رقبای خود سبقت بگیرند. به منظور افزایش کیفیت محصول و خدمت می‌بایست عواملی تاثیرگذار را کشف نمود. ضمن ارزیابی تمامی عوامل، موارد نامطلوب را حذف و در راستای بهبود شرایط، راهکارهای عملی ارائه نمود. در این خصوص از مفهوم داده‌کاوی بعنوان یکی از روش‌های محاسبات نرم و تجزیه و تحلیل داده می‌توان اشاره کرد.

داده‌کاوی فرآیند کشف الگوها و پیش‌بینی روندهای ناشناخته تکرار شونده از مقادیر زیادی داده به صورت خودکار است. داده‌کاوی مجموعه‌ای است از توابع چند جمله‌ای تجزیه و تحلیل ارتباطات، طبقه بندی و پیش‌بینی، تجزیه و تحلیل خوشه‌ها و داده‌های پرت است [Sun and Li, 2008]. در این تحقیق که مطابق با مدل کریسپ نوشته شده است، ابتدا داده‌های شرکت اتوبوسرانی جمع آوری شده، سپس مجموعه داده‌ها پاکسازی و تحت فرآیند پیش پردازش قرار گرفت. با بهره‌مندی از الگوریتم جی آر آی از روش‌های قوانین تلازمی همچنین مدل سی فایو از الگوریتم‌های دسته بندی، داده‌های تحقیق تحت مدلسازی و واکاوی قرار گرفته و نتایج بدست آمده، در تامین دانش مورد نیاز برای ایجاد سیستمی تصمیم یار مورد استفاده قرار گرفت.

## ۲. مروری بر ادبیات

در فاصله زمانی بین سالهای ۲۰۰۵ تا ۲۰۱۷ مقالات در حوزه داده‌کاوی مورد بررسی قرار گرفته است. این مقالات از منظر موضوع و محتوا در دو دسته قابل تقسیم‌بندی هستند: ۱- حوزه های مهندسی سلامت و حوادث ناشی از کار ۲- حوزه مهندسی حمل و نقل. در دسته‌بندی اول تاکنون تحقیقات زیادی در حوزه‌های مختلف انجام شده است. دلن و همکاران تحقیقی در زمینه پیش‌بینی سرطان سینه انجام داده‌اند. در این تحقیق از دو الگوریتم داده‌کاوی پرکاربرد شبکه‌های عصبی و درخت تصمیم گیری همراه با روش آماری رگرسیون لجستیک برای توسعه مدل‌های پیش‌بینی استفاده گردیده است. در این

مطالعه داده‌های بیش از ۲۰۰/۰۰۰ نفر بیماران مبتلا به سرطان سینه مطالعه و روش درخت تصمیم در این مطالعه دارای دقت بالاتری گردید [Delen, Walker and Kadam, 2005]. تسو و یو (۲۰۰۷) به مقایسه سه تکنیک مدل سازی شامل تجزیه و تحلیل رگرسیون، درخت تصمیم و شبکه‌های عصبی بمنظور پیش‌بینی مصرف انرژی برق پرداختند. در راستای پیش‌بینی مصرف انرژی برق، درخت تصمیم و مدل شبکه عصبی به نظر جایگزین مناسبی برای مدل رگرسیون در درک الگوهای مصرف انرژی و پیش‌بینی سطح مصرف انرژی میباشند [Tso and Yau, 2007]. یانگ و همکاران (۲۰۰۹) بخش اورژانس یک بیمارستان مورد مطالعه قرار دادند. در این پژوهش با استفاده از روش‌های داده‌کاوی همانند خوشه بندی و درخت تصمیم به تجزیه و تحلیل مدل پیش‌بینی تقاضای بیماران بخش اورژانس در شرایط واقعی درمان پرداخته‌اند [Yang, 2009].

قوسی و همکاران (۲۰۱۲) به مطالعه یک روش مناسب و سریع برای پیش‌بینی مصرف داروهای مختلف پرداختند. بر همین اساس، الگوریتم پیش‌بینی همانند، مدل‌های رگرسیون، شبکه‌های عصبی و درخت‌های تصمیم‌گیری مقایسه شده است [Ghousi, Mehrani and Momeni, 2012]. سدنو و همکاران (۲۰۱۳) از داده‌کاوی برای توانبخشی بیماران مبتلا به آسیب دیدگی مغزی استفاده کردند، آسیب مغزی یکی از علل اصلی مرگ و میر و ناتوانی در جهان است که هزینه‌های مراقبتی بالایی دارد. با استفاده از داده‌کاوی سه مدل پیش‌بینی با درخت تصمیم‌گیری همانند شبکه عصبی ارائه گردید. بمنظور آزمایش الگوریتم‌ها از پایگاه داده بیماران یک بیمارستان استفاده شد و مدل پیش‌بینی درخت تصمیم‌گیری با ۹۰/۳۸٪ دقت تعیین گردید. [Cedeno et al. 2013].

منگ و همکاران (۲۰۱۳) به مقایسه سه مدل استخراج اطلاعات با داده‌کاوی برای پیش‌بینی دیابت یا به تاخیر انداختن عوامل ریسک آن پرداختند. هدف از این مطالعه مقایسه عملکرد رگرسیون لجستیک، شبکه‌های عصبی و مدل درخت تصمیم‌گیری برای پیش‌بینی دیابت یا به تاخیر انداختن آن با استفاده از عوامل تاثیرگذار است. افراد شرکت کننده در این پژوهش از دو جامعه آماری در گوانجو چین، یعنی ۷۳۵ نفر دارای دیابت و ۷۵۲ نفر سالم انتخاب شدند. دقت سه مدل

شبکه عصبی مدل‌سازی شد [Izad Shenan et al. 2014]. قوسی (۲۰۱۵) به مطالعه مجموعه داده های زیادی در مورد حوادث در یک واحد صنعتی با استفاده از ابزارهای داده‌کاوی پرداخته است. در مرحله اول، اطلاعات پیش پردازش و سپس با انجام روش‌های خوشه بندی، داده‌ها را به گروه‌های مجزا تقسیم‌بندی و تعداد خوشه های مشخص شده است. در مرحله بعد، با استفاده از الگوریتم آپریوری بعنوان یکی از روش‌های قوانین انجمنی روابط میان حوادث استخراج شده و مورد تجزیه و تحلیل قرار گرفت. بر اساس این تحقیق، سهم قابل توجهی از حوادث به کارگران جوان دارای تحصیلات دیپلم با سن بین ۲۰ تا ۲۵ سال و سابقه کاری کمتر از ۲ سال و همچنین کارگران با تجربه کاری بین ۳ تا ۸ سال و میزان تحصیلات کمتر از دیپلم اختصاص داشته است [Ghousi, 2015]. اسماعیلی و همکاران (۲۰۱۵) با در نظر گرفتن این موضوع که آموزش کارکنان فرآیندی بسیار وقت‌گیر و هزینه‌بر می باشد. مدل پیش‌بینی بر اساس داده‌های تاریخی سازمان و همچنین اطلاعات و ویژگی‌های شخصی کارکنان از جمله سن، مهارت‌های فنی و تجربه کاری آنان به منظور جلوگیری از رویگردانی و ترک شغل کارکنان و کاهش کارایی سازمان ارائه نمودند [Esmiaeeli, Ghousi and Esmiaeeli, 2015].

در پژوهش دیگری میستی کوقلو و همکاران (۲۰۱۵) حوادث سقوط از ارتفاع در صنعت ساخت و ساز را با استفاده از درخت تصمیم مورد تجزیه و تحلیل قرار دادند. در این تحقیق با استفاده از الگوریتم‌های سی‌فایو و کای‌دو حادثه سقوط از ارتفاع تحت مورد واکاوی قرار گرفته و با استفاده از واکاوی قواعد استخراج شده، روابط پنهان میان ویژگی‌های مختلف ارائه گردید [Mistikoglu et al. 2015].

صاحبی و همکاران (۱۳۹۴) با هدف پیش‌بینی و ارزیابی عوامل موثر بر شدت تصادفات عابران پیاده در راه‌های برون شهری عوامل مربوط به تصادفات منجر به جراحت و فوت در استان تهران طی سالهای ۱۳۸۸ الی ۱۳۹۰ با استفاده از مدل لوجیت دوتایی مورد تحلیل قرار دادند. نتایج تحقیق میتواند مدیران را در شناخت بهتر ماهیت تصادفات کمک بسزایی نماید [Sahebi et al. 2015]. بشیر و همکاران (۲۰۱۶) سیستم پشتیبان تصمیمی در حوزه سلامت MDSS<sup>1</sup> را ارائه کردند. در این تحقیق داده های متنوعی در خصوص بیماران‌های

مورد استفاده در این پژوهش از نظر دقت، حساسیت و ویژگی مورد ارزیابی قرار گرفت و مدل درخت تصمیم، بالاترین دقت را به خود اختصاص داد [Meng et al. 2013]. چنگ و همکاران (۲۰۱۳) از تکنیک‌های داده‌کاوی بمنظور تجزیه و تحلیل علل حوادث عمده شغلی در صنعت پتروشیمی استفاده کردند. در این مطالعه ۳۴۹ مورد از حوادث مهم در صنعت پتروشیمی بین سال‌های ۲۰۰۰ و ۲۰۱۰ در کشور تایوان بمنظور تجزیه و تحلیل جمع‌آوری شده‌اند. در این تحقیق از طبقه بندی و درخت تصمیم در بررسی توابع توزیع و قوانین حاکم بر عوامل و علل حوادث استفاده شده است. این مطالعه نشان داد که تجهیزاتی همانند خطوط لوله و دریچه‌های کنترل، طرح ریزی دستگاه‌هایی با ایمنی و کیفیت بالای نگهداری و برنامه‌های نوسازی می‌تواند به طور موثر از حوادثی همانند آتش سوزی، انفجار و مسمومیت ناشی از نشت مواد جلوگیری کند. در ضمن، اجرای اقدامات مدیریت ایمنی، مانند آموزش‌های ایمنی به کارگران، اجرای استاندارد بازرسی و ارزیابی ریسک پرسنل، عامل مهمی در پیشگیری از حوادث میباشد [Cheng et al. 2013]. شرکت‌های تولیدی می‌توانند درآمد و سود قابل توجهی را از طریق ارتقاء کیفیت خدمات پس از فروش به دست آورند.

ابراهیم‌زاده و همکاران (۲۰۱۴) به بررسی یک چارچوب متشکل از تکنولوژی داده‌کاوی و برنامه نویسی پویا برای تحلیل روابط مختلف در میان خرابی‌های بخش خودرو پرداختند. مدل ارائه شده در ابتدا به جمع‌آوری داده‌های ثبت مناسب می‌پردازد و پس از پردازش، در دو جریان مجزا روابط مهم و قوانین انجمنی را جستجو می‌کند. از الگوریتم کائین جهت تعیین ارتباط بین خرابی‌های بخش‌های متفاوت استفاده شد [Ebrahimzadeh, Mahdavi and Ghousi, 2014].

در تحقیق دیگری ایزد شناس و همکاران (۲۰۱۴) به شناسایی بیماران با هزینه های درمانی بالا با استفاده از تکنیک‌های داده‌کاوی همانند شبکه‌های عصبی و درخت‌های تصمیم پرداختند. در این تحقیق مجموعه ای به تعداد ۹۸۱۷۵ پرونده پزشکی با مخارج بالا مورد بررسی و مطالعه قرار گرفت. این در حالی است که پس از عملیات پیش‌پردازش، طبقه بندی و نرمال‌سازی داده‌ها، مجموعه داده‌های پاکسازی شده به کمتر از ۳۱۷۰۴ مورد پرونده کاهش یافته و توسط درخت تصمیم و

## استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم.....

سطوح تحصیلات و همچنین سن با میزان مرگ و میر مورد تایید قرار گرفت [Sami et al. 2013]. توکلی و همکاران (۲۰۱۴) با توجه به اینکه مسافران ترک سوار موتورسیکلت شامل بخش قابل توجهی از قربانیان تصادفات رانندگی هستند به بررسی عوامل موثر در شدت تصادفات پرداختند. برای تجزیه و تحلیل شدت آسیب مسافر ترک سوار در ایران، داده‌های بیش از یک دوره چهار ساله حداقل سالهای ۲۰۰۹ تا ۲۰۱۲ با استفاده از روش طبقه بندی و درخت رگرسیون بررسی شد. نتایج نشان داد که نوع منطقه، کاربری زمین، و بخش آسیب دیده بدن مصدومین حادثه (همانند سر و گردن) موثرترین عوامل بر مرگ و میر موتور سیکلت است [Tavakoli, Rabieyan and Besharati, 2014] جیانگین و همکاران (۲۰۱۵) به تجزیه و تحلیل فاصله زمانی در تاکسی‌های شهر پکن پرداختند و از داده‌های GPS<sup>2</sup> تاکسی‌ها استفاده کردند. داده‌ها شامل عملیات روزانه، محدوده عملیاتی، محل اقامت رانندگان، استراحت دوره ای و شاخص‌های دیگر بودند. اطلاعات بدست آمده از این بررسی می‌تواند اساس تصمیم‌گیری در سطح مدیریت عملیاتی حوزه تاکسیرانی پکن باشد [Jianqin et al. 2015].

وانگ و همکاران (۲۰۱۵) به بررسی داده‌های رانندگی در جاده‌های چین پرداختند. هدف آن‌ها از این تحقیق در ابتدا خوشه‌بندی سطوح مختلف رانندگی در ارتباط با خطر ناشی از تصادفات و یافتن عواملی که تا حد زیادی بر سطح خطر رانندگی تاثیر می‌گذارد. از طبقه‌بندی و درخت رگرسیون برای تجزیه و تحلیل داده‌ها استفاده شد. نتایج نشان می‌دهد که سرعت بالا در هنگام ترمز، عوامل تحریک آور و نوع تصادف احتمالی باعث بیشترین اثر در خطرات رانندگی است [Wang et al. 2015].

وانگ و همکاران (۲۰۱۶) به بررسی ۳۶ پارامتر مرتبط با ایمنی با استفاده از داده‌های تقاطع‌ها در شهر آلاباما پرداختند. چهار تابع عملکرد ایمنی مختلف از جمله رگرسیون پواسون، رگرسیون دو جمله‌ای منفی، مدل خطی عمومی و درختان رگرسیون مورد بررسی و مقایسه قرار گرفتند. مدل درخت رگرسیون عملکرد بهتری نسبت به سایر مدل‌ها داشته است [Wang et al. 2016]. وانگ و همکاران (۲۰۱۶) به بررسی ویژگی‌های تلفات تصادفات و عوامل موثر بر آن در

قلبی، سرطان پستان، دیابت، بیماری کبد، پارکینسون و هپاتیت با استفاده از درخت تصمیم، شبکه عصبی و ماشین بردار پشتیبان مورد تحلیل قرار گرفت [Bashir et al. 2016].

در حوزه مهندسی حمل و نقل بعنوان دومین دسته بندی تحقیقات بسیار محدودی انجام شده است. میرآبادی و شریفیان (۲۰۱۰) با تجزیه و تحلیل اطلاعات تاریخی در خطوط راه آهن ایران و اعمال قوانین انجمنی از تکنیک‌های داده‌کاوی به منظور کشف ارتباطات ناشناخته و الگوهای میان داده‌ها پرداختند. در این پژوهش اطلاعات مربوط به ۶۵۰۰ تصادف از میان پایگاه داده تصادفات بین سال‌های ۱۹۹۶ تا ۲۰۰۵ انتخاب و مورد بررسی قرار گرفت [Mirabadi and Sharifian, 2010].

گوربوز و همکاران (۲۰۱۱) با توجه به اهمیت موضوع ایمنی در حمل و نقل هوایی و از طرفی رشد چشمگیر مسافرت‌های هوایی، تاخیر و لغو پرواز و همینطور بروز حوادث و تصادفات هوایی و تولید و تجمع حجم زیادی داده و دانش در صنعت حمل و نقل هوایی در قالب گزارش‌های خلبان، تعمیر و نگهداری، حادثه و یا تاخیر به پردازش و تجزیه و تحلیل گزارش‌های یک شرکت هواپیمایی متعلق به کشور ترکیه پرداخته‌اند [Gurbuz, Ozbakir and Yapici, 2011].

پاک گوهر و همکاران (۲۰۱۱) در پژوهشی به بررسی حوادث راهنمایی و رانندگی در کشور ایران پرداختند. در این مطالعه داده‌های ترافیکی توسط تکنیک‌های داده‌کاوی مانند رگرسیون لجستیک، طبقه بندی و روش رگرسیون مورد بررسی قرار گرفت [Pakgozar et al. 2011]. اونا و همکاران (۲۰۱۲) با توجه به اهمیت ارزیابی کیفیت خدمات حمل و نقل و بررسی رضایت مشتریان، از روش درخت رگرسیون که هیچ نیازی به رابطه اساسی از پیش تعیین شده بین متغیرهای مستقل و وابسته ندارد، استفاده کردند و به شناسایی عوامل کلیدی موثر بر کیفیت خدمات حمل و نقل اتوبوسرانی با توجه به داده‌های جمع آوری شده رضایت مشتریان سیستم حمل و نقل شهری گرانا‌دا پرداختند [Ona, Ona and Calvo, 2012].

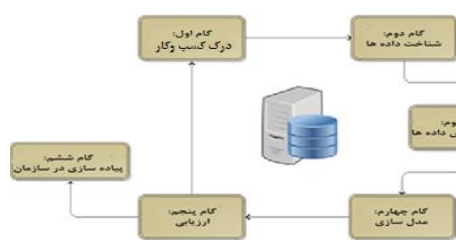
سامی و همکاران (۲۰۱۳) به تجزیه و تحلیل داده‌های تصادفات جاده‌ای در استان فارس به منظور کشف الگوهایی که به پیشگیری یا کاهش تصادفات کمک می‌کند پرداختند.

داده‌ها در این تحقیق با نرم افزار کلمنتاین مورد واکاوی قرار گرفت و از جمله نتایج بدست آمده، معنی داری روابط بین

بر مبنای دستورالعمل واگذاری اتوبوس و بهره برداری از خطوط اتوبوسرانی شهری و حومه شهری به بخش خصوصی، سیاست شرکت اتوبوسرانی به سمت خصوصی سازی میل داشته است و این شرکت با پیمانکاران مختلف برای سرویس دهی در خطوط مختلف قرارداد منعقد کرده است، ولی این رویکرد با مشکلاتی هم رو به رو بوده است. برخی از این شرکتها استانداردهای لازم را در زمینه های مختلف اعم از زمانبندی خدمات دهی، استفاده از پرسنل مجرب، کنترل کیفیت خدمات خود را رعایت نمی کنند. در این تحقیق اطلاعات تمامی شرکت های خصوصی که در تهران فعالیت می کنند مورد واکاوی قرار داده شده است. با توجه به بررسی مقالات مختلف داده کاوی مشخص گردید اکثر تحقیقات انجام شده در زمینه اعمال تکنیک های داده کاوی در حوزه های مهندسی پزشکی، داده های آموزشی و خدمات پس از فروش واحدهای خودروسازی بوده است و تحقیقات کمی در زمینه حمل و نقل عمومی انجام شده است، ضمناً تاکنون هیچ تحقیقی در زمینه ی تخلفات شرکت های حمل و نقل عمومی ایرانی انجام نشده است و همچنین در هیچ یک از تحقیقات مرتبط گذشته در ایران از داده های جمع آوری شده با سیستم GPS استفاده نشده است. ضمناً در هیچ یک از تحقیقات حوزه تحلیل عملکرد انسانی نیز از تکنیک های داده کاوی بعنوان مدلهای پیش بینی استفاده نشده است.

### ۳. متدولوژی تحقیق

در این تحقیق متدولوژی کریسپ- دی ام ۲ بعنوان متدولوژی تحقیق مورد توجه قرار گرفت. این متدولوژی روشی توانمند و منعطف جهت ارتقای داده کاوی در حل مسائل سازمانی است.



شکل ۱. چرخه داده کاوی با مدل کریسپ

تجزیه و تحلیل ایمنی ترافیک پرداختند. آنان به وسیله قوانین انجمنی این تجزیه و تحلیل را انجام دادند. از جمله نتایج بدست آمده نوشیدن هنگام رانندگی، سرعت بیش از ۴۰ مایل در ساعت و استفاده نکردن از دستگاه های کنترل ترافیکی می باشد [Weng et al. 2016]. با توجه به این موضوع که ارائه خدمات حمل و نقل در مناطق شهری ضروری بوده و سطح درآمد برای رانندگان مهم است کین و همکاران (۲۰۱۷) به بررسی عوامل موثر بر درآمد رانندگان پرداختند. محققین تعداد ۱۶۷ میلیون سطر داده را از ۸۰۰۰ تاکسی در شهر شانگهای چین بررسی نموده و در پایان یک مدل چند سطحی از عوامل مهم و تاثیرگذار ارائه نمودند [Qin et al. 2017].

پراتی و همکاران (۲۰۱۷) به بررسی عوامل پیش بینی شدت تصادفات دوچرخه در ایتالیا پرداختند، آنها برای این پژوهش از دو تکنیک داده کاوی درخت تصمیم و شبکه بیز استفاده کردند. محققین تعداد ۴۹۶۲۱ تصادف که حداقل یک دوچرخه سوار مجروح شده است را بررسی کردند که مواردی مانند سقوط، نوع جاده و نوع وسیله نقلیه طرف مقابل از مهمترین عوامل تاثیرگذار در شدت تصادفات در این تحقیق محسوب گردیدند [Prati, Pietrantonio and Fraboni, 2017]. با بررسی ادبیات تحقیق و مقالات و مطالعات انجام شده مشخص گردیده است که در هیچیک از پژوهش های انجام شده، داده های GPS برای ارزیابی پیمانکاران در حوزه حمل و نقل شهری استفاده نگردیده است و همچنین در ایران پژوهشی در زمینه تخلفات رانندگی رانندگان شرکت اتوبوسرانی صورت نگرفته است.

با توسعه روز افزون شهرها، امروزه حمل و نقل عمومی به عنوان یکی از ضروری ترین انواع خدمات شهری، راهگشای بسیاری از مشکلات زندگی در کلان شهرهایی همانند شهر تهران، درخصوص ترافیک و آلودگی هوا میباشد.

شهر تهران با توجه به اینکه یکی از پرجمعیت ترین شهرهای جهان است، کیفیت حمل و نقل عمومی از اهمیت بسزایی برخوردار است و باید توجه ویژه ای به این حوزه، به خصوص ارزیابی عملکرد واحدهای زیرمجموعه آن همانند شرکت اتوبوسرانی تهران مد نظر قرار داد. در سالیان اخیر با توجه ضرورت ارتقای سطح کمی و کیفی ارائه خدمات به شهروندان،

## استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم.....

کاوی تجزیه و تحلیل داده های تولید شده در شرکت اتوبوسرانی جهت پاسخگویی به این سوال است که عملکرد پیمانکاران و رانندگان زیر مجموعه این شرکتهای پیمانکاری ارزیابی و رتبه بندی گردند. این فرآیند بایستی قادر باشد که پیمانکارانی را که عملکرد مناسبی ندارند را شناسایی نموده و نتایج حاصل از این واکاوی، پیمانکاران با عملکرد و کارایی بالا را انتخاب نموده و شرکت اتوبوسرانی را در راستای هدف غایی یعنی ارتقاء رضایتمندی مشتریان سوق دهد.

طبق آخرین اطلاعات اعلام شده در فروردین ۱۳۹۴ این شرکت دارای ۴۵ خط در بخش عمومی و ۱۹۵ خط در بخش خصوصی و ۱۰ خط تندرو است که این خطوط 8231 ایستگاه را پوشش می‌دهند، این شرکت دارای ۸۸۰۹ نفر پرسنل اعم از راننده، عوامل فنی، عوامل بهره برداری و سایر(ستادی، خدماتی، پشتیبانی) می باشد. هجده شرکت در بخش خصوصی با اتوبوسرانی تهران در حال همکاری هستند و وظیفه جابه‌جایی مسافران در کلان شهر تهران را بر عهده دارند.

### ۲-۳ شناخت داده‌ها

مجموعه داده‌های مورد استفاده در تحقیق، میزان تخلفات بین سال های ۱۳۹۱ تا ۱۳۹۴ اتوبوس های بخش خصوصی شرکت اتوبوسرانی تهران شامل ۶۵۰۰ سطر است. ویژگیهای داده های ورودی شامل نام شرکت متخلف، کد تخلف انجام شده، زمان وقوع تخلف، تعداد اتوبوس هر شرکت در هر سال، همچنین شرکت اتوبوسرانی برای تخلفاتی که توسط راننده و یا شرکت انجام می شود به صورت جداگانه امتیاز منفی در نظر گرفته است، که هر کدام را به اختصار شرح داده می‌شود:

جدول ۱. معرفی داده ها

ردیف	داده های ثبت شده	شرح
۱	نام شرکت	نام شرکت متخلف را نشان می‌دهد
۲	زمان وقوع	زمان دقیق تخلف را معین می‌کند
۳	درجه تخلف	میزان اهمیت تخلف را نشان می‌دهد
۴	امتیاز منفی شرکت	میزان امتیاز منفی شرکت را بر حسب نوع تخلف را نشان می‌دهد
۵	امتیاز منفی راننده	میزان امتیاز منفی راننده را بر حسب نوع تخلف را نشان می‌دهد
۶	کد تخلف	تخلفات به صورت کد سه رقمی ثبت شده‌اند
۷	تعداد اتوبوسهای هر شرکت	تعداد اتوبوس‌های هر شرکت پیمانکار را نشان می‌دهد

با توجه به این روش، یک پروژه داده‌کاوی چرخه عمری متشکل از شش گام دارد [Weng et al. 2016]. مدل فرآیند کریسپ که در شکل شماره یک مشاهده می‌شود. نمایش کلی از چرخه پروژه‌های داده‌کاوی است که در برگزیده مراحل پروژه و ارتباط بین آن‌ها است. این متدولوژی شامل شش گام می باشد که در ادامه به معرفی و بررسی آن‌ها پرداخته خواهد شد.

### ۳-۱ درک کسب و کار

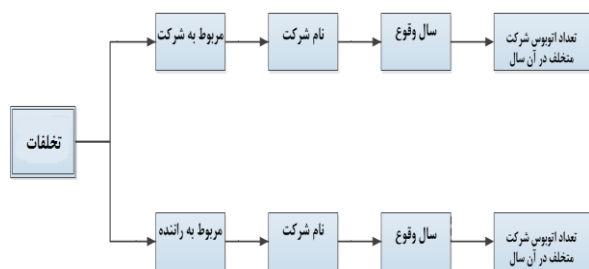
اولین مرحله از چرخه کریسپ، درک کسب و کار است. در این مرحله کارفرما مشکل سازمان و کسب و کار خود را شناسایی نموده و به تیم داده‌کاوی گزارش می‌دهد. این یک واقعیت است که مدیران سازمانها تا زمانی که نتوانند وضعیت واحدهای زیرنظر خود را با استفاده از شاخص های مناسب اندازه گیری نمایند، قطعاً مدیریت این مجموعه ها کار بسیار دشواری خواهد بود. در این راستا با توجه به نارضایتی‌های اعلام شده مردم در سامانه اتوبوسرانی تهران، مدیران شرکت را بر آن داشت تا به دنبال پاسخگویی به این سوال باشند که چگونه میتوانند عملکرد پیمانکاران خود را ارزیابی نموده و داده های تولید شده از فعالیتهای پیمانکاران شرکت اتوبوسرانی و رانندگان زیرمجموعه این شرکتهای پیمانکاری را مورد استفاده قرار داده و به تبع این ارزیابی قرارداد پیمانکارانی که خودشان و یا رانندگان آنها مرتکب تخلف شده اند را تمدید نمایند.

به همین منظور در این تحقیق استفاده از تکنیکهای متنوع داده کاوی مورد توجه قرار گرفت. هدف استفاده از رویکرد داده

چه فصلی رخ داده است، این فیلد با توجه به تاریخ دقیق وقوع تخلف بدست آمده است.

**ماه وقوع تخلف:** این ویژگی نشان می‌دهد که تخلف در چه ماهی رخ داده است، این ویژگی نشان دهنده تاریخ دقیق وقوع تخلف است. استفاده از فیلدها و ویژگیهای موجود در داده‌های تحقیق و یا ساخت ویژگی‌های جدید که بطور مثال می‌توان به سال، فصل و ماه وقوع تخلف رانندگان وابسته به هر شرکت پیمانکار نیز اشاره نمود، این امکان را فراهم می‌نماید که وقوع تخلفات و امتیاز منفی رانندگان و شرکت پیمانکار را از منظر زمانی نیز اشاره نمود و شاید بتوان برای زمان‌هایی که تواتر این تخلفات افزایش می‌یابد، راهکارهای اصلاحی مناسبی پیشنهاد نمود.

امتیاز منفی شرکت، راننده و تعداد اتوبوس: با توجه به این که تعداد اتوبوس متعلق به شرکت‌ها دارای مقادیر متنوعی است، فرآیند نرمال‌سازی بوسیله نرم‌افزار متلب<sup>۵</sup> انجام شد. همان‌طور که در شکل دو نشان داده شده است ابتدا الگوریتم مشخص می‌کند تخلف را اتوبوسران و یا شرکت مرتکب شده است و در ادامه نام شرکت، سال وقوع و تعداد اتوبوس آن شرکت را مشخص می‌کند.



شکل ۲. روند نرمال‌سازی در نرم‌افزار متلب

با توجه به امتیاز منفی تخلف و تعداد اتوبوس متعلق به هر یک از شرکت‌های پیمانکار متخلف، عدد بدست آمده که نشان دهنده کیفیت عملکرد شرکت پیمانکار به پنج سطح (A,B,C,D,E) دسته‌بندی می‌گردد. لازم به توضیح است که بیشتر شرکت‌های مورد بررسی مطابق شکل پنج در رده‌ی A قرار گرفته‌اند.

**۱-۲-۳ نام شرکت:** نام شرکت متخلف تحت قرارداد اتوبوسرانی تهران را نشان می‌دهد، این یک فیلد اسمی است که شامل ۱۸ شرکت می‌باشند.

**۲-۲-۳ زمان وقوع:** زمان دقیق وقوع تخلف بر حسب سال، ماه و روز را نشان می‌دهد. این فیلد به صورت عددی است.

**۳-۲-۳ درجه تخلف:** نشان دهنده میزان اهمیت تخلف انجام شده است که درجه یک (اهمیت بالا)، درجه دو (اهمیت متوسط) و درجه سه (اهمیت پایین‌تر) است. این فیلد نیز همانند زمان وقوع از نوع عددی است.

**۴-۲-۳ امتیاز منفی شرکت:** در این فیلد شرکت اتوبوسرانی بر حسب نوع تخلف شرکت پیمانکار، امتیاز منفی برای شرکت در نظر گرفته و این فیلد بصورت عددی است.

**۵-۲-۳ امتیاز منفی راننده:** شرکت اتوبوسرانی برای تخلفاتی که راننده مرتکب شده است بر حسب نوع تخلف امتیاز منفی منظور مینماید. این فیلد از نوع عددی است.

**۶-۲-۳ کد تخلف:** تخلفات در شرکت اتوبوسرانی تهران به صورت کد در قالب یک عدد سه رقمی ثبت می‌گردد.

**۷-۲-۳ تعداد اتوبوس‌های هر شرکت:** تعداد اتوبوس شرکت‌ها در هر سال نشان داده شده است که این فیلد عددی است.

### ۳-۳ پیش پردازش داده‌ها

کلیه عملیاتی که باعث آماده‌سازی مجموعه داده‌ها جهت ورود به نرم‌افزار و شروع فرآیند مدل‌سازی را شامل می‌شود را پیش‌پردازش داده‌ها گویند. پنج مرحله در آماده‌سازی داده‌ها از جمله انتخاب داده‌ها، تمیز کردن داده‌ها، ساخت و ساز داده‌ها، یکپارچه‌سازی داده‌ها و قالب بندی داده‌ها اهمیت بسزایی دارد.

برای ورود داده‌ها به نرم‌افزار کلمنتاین<sup>۶</sup> باید اقداماتی بر روی داده‌ها انجام شود. به این منظور با بررسی و ارزیابی داده‌های موجود، سال و فصل و ماه وقوع تخلف بدست آمد. همچنین با توجه به تعداد اتوبوس‌های هر شرکت و نیز امتیاز منفی شرکت‌ها و رانندگان کیفیت عملکردی آنان در پنج سطح بخش بندی شد. سال وقوع تخلف: این ویژگی به سال وقوع تخلف اشاره می‌نماید و با بررسی زمان وقوع تخلف‌ها حاصل می‌شود. فصل وقوع: این ویژگی نشان دهنده این است که تخلف در

## استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم.....

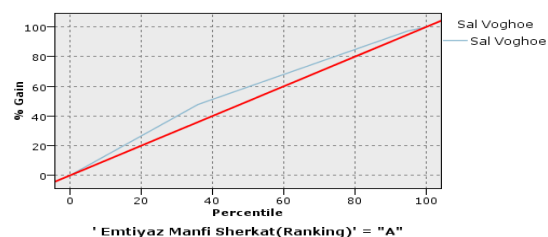
### ۳-۴-۱ الگوریتم جی آر ای

این الگوریتم بعنوان یکی از الگوریتم‌های روش قوانین انجمنی محسوب می‌گردد. هریک از قواعد استخراج شده توسط الگوریتم‌های قوانین انجمنی با سه شاخص پشتیبان، اطمینان و لیفت ارزشیابی می‌گردند. شاخص‌های پشتیبان و اطمینان در خصوص مفید بودن و اطمینان از قاعده کشف شده از میان داده‌های موجود در قالب جذابیت و میزان اهمیت آن قاعده اظهار نظر مینماید. در نظر گرفتن همزمان شاخص‌های پشتیبان و اطمینان باعث ایجاد شاخص لیفت میشود که بر اساس ادبیات تحقیق قواعدی که میزان لیفت بالای یک دارند میتوانند بعنوان قواعد قابل قبول برای ارزیابی و بررسی مساله تحقیق مورد توجه قرار گیرند. این در حالی است که حداقل مقادیر شاخص‌های پشتیبان و اطمینان نیز در این تحقیق بترتیب ۱۰ درصد و ۷۰ درصد منظور گردیده است [Ghousi, 2015].

این الگوریتم می‌تواند، قواعدی را که در مجموعه داده‌ها به دفعات رخ داده است، پیدا نموده و به صورت قواعدی جداگانه در اختیار تصمیم‌گیرندگان و مدیران قرار دهد. شرکت‌های الف و ب، با توجه به درخت تصمیم نیز عملکرد نامناسبی داشته‌اند در قواعد بدست آمده از الگوریتم جی آر ای نیز به آنها اشاره شده است. بعلاوه این شرکتها تخلفات مختلفی را در فصولی خاص مرتکب شده اند که در ادامه به چند نمونه از قواعد بدست آمده اشاره می‌گردد

### جدول ۲. نمونه ای از قواعد انجمنی استخراج شده

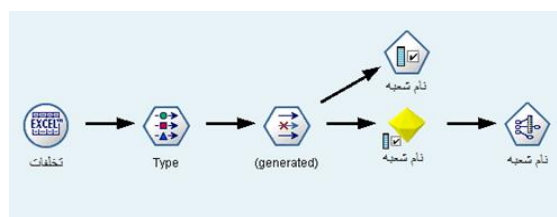
ردیف	تالی/پیرو	مقدم	پشتیبان	اطمینان	لیفت
۱	نام شعبه: الف	فصل وقوع: بهار کد تخلف: ۳۴۵ امتیاز منفی شرکت (E)	۰/۳۶	٪۱۰۰	۳۵/۶۸
۲	نام شعبه: الف	فصل وقوع: تابستان کد تخلف: ۳۴۵ امتیاز منفی شرکت (E)	۰/۳۶	٪۱۰۰	۳۵/۶۸
۳	نام شعبه: الف	فصل وقوع: تابستان کد تخلف: ۳۲۷ امتیاز منفی شرکت (E)	۰/۳۹	٪۱۰۰	۳۵/۶۸



شکل ۳. رتبه بندی فراوانی شرکت‌ها بر حسب سال

### ۳-۴-۲ مدل سازی

جهت ساخت مدلی از داده‌های شرکت اتوبوسرانی از تکنیک‌های مختلف داده‌کاوی نظیر جی آر ای و سی فایو استفاده گردید. الگوریتم جی آر ای از دسته قوانین تلازمی وسی فایو مربوط به درخت تصمیم می‌باشند. در این تحقیق از هر دو روش قوانین تلازمی و دسته بندی استفاده شده است و در بین ۱۸ شرکت پیمانکار دو شرکت (الف) و (ب) در هر دو روش جزء شرکت‌های با تخلف بالا تعیین شدند.



شکل ۴. نمایی از نرم افزار کلمنتاین



اردلان بادامی، روزبه قوسی، محمد سعیدی مهرآباد

۴	نام شعبه: ب	فصل وقوع: پاییز کد تخلف: ۲۲۲ امتیاز منفی شرکت (E)	۰/۰۲	٪۱۰۰	۴۱/۵۱
۵	نام شعبه: ب	فصل وقوع: تابستان کد تخلف: ۱۴۱ امتیاز منفی شرکت (E)	۰/۰۲	٪۱۰۰	۴۱/۵۱
۶	نام شعبه: ب	فصل وقوع: زمستان کد تخلف: ۳۴۵ امتیاز منفی شرکت (D)	۰/۰۲	٪۱۰۰	۴۱/۵۱

معیار لیفت ۴۱/۵۱٪ باشد، شرکت متخلف، شرکت ب محسوب می‌گردد.

**قاعده شماره ۵:** این قاعده نشان می‌دهد که اگر در فصل تابستان، تخلف ۱۴۱ (عدم رعایت برنامه فعالیت در خطوط و یا زمانبندی سرفاصله حرکت ناوگان برابر مقررات اتوبوسرانی و شرکت در روزهای تعطیل و ایام کشیک) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه E باشد، با معیار های پشتیبان، اطمینان و لیفت بترتیب ۰/۰۲٪ و ۱۰۰٪ و ۴۱/۵۱٪ شرکت متخلف، شرکت ب است.

**قاعده شماره ۶:** این قاعده نشان می‌دهد که اگر در فصل زمستان، تخلف ۳۴۵ (عدم رعایت برنامه فعالیت در خطوط، ساعت کار روزانه و یا زمانبندی سرفاصله حرکت ناوگان برابر مقررات اتوبوسرانی و شرکت در روزهای غیر تعطیل) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه D باشد، با معیار پشتیبان ۰/۰۲٪ و معیار اطمینان ۱۰۰٪ و معیار لیفت ۳۵/۶۸٪ باشد، شرکت متخلف، شرکت ب می‌باشد. همانطوریکه با مرور قواعد مکشوفه توسط الگوریتم جی آر آی در جدول شماره ۲ ملاحظه میگردد، از میان سایر شرکتهای پیمانکار، شرکتهای الف و ب جزء پیمانکاران پرتخلف می‌باشند و این در حالیست که در بررسی درخت تصمیم نیز به این موضوع اشاره شده است. این دانش کشف شده میتواند در فرآیند ارزیابی عملکرد پیمانکاران شرکت اتوبوسرانی در کنار سایر موارد، جهت تمديد قرارداد پیمانکاران با عملکرد مطلوب و عدم تمديد قرارداد پیمانکاران فاقد عملکرد مطلوب مورد توجه قرار گیرد.

**قاعده شماره ۱:** این قاعده نشان می‌دهد در صورتیکه در فصل بهار، تخلف ۳۴۵ (عدم رعایت برنامه فعالیت در خطوط، ساعت کار روزانه و یا زمانبندی سرفاصله حرکت ناوگان برابر مقررات اتوبوسرانی و شرکت در روزهای غیر تعطیل) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه E باشد، با معیار پشتیبان ۰/۳۶٪ و اطمینان ۱۰۰٪ و لیفت ۳۵/۶۸٪ باشد، شرکت متخلف، شرکت الف می‌باشد.

**قاعده شماره ۲:** این قاعده نشان دهنده این است که اگر در فصل تابستان، تخلف ۳۴۵ (عدم رعایت برنامه فعالیت در خطوط، ساعت کار روزانه و یا زمانبندی سرفاصله حرکت ناوگان برابر مقررات اتوبوسرانی و شرکت در روزهای غیر تعطیل) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه E باشد، با معیار پشتیبان ۰/۳۶٪ و اطمینان ۱۰۰٪ و لیفت ۳۵/۶۸٪ باشد، شرکت متخلف، شرکت الف می‌باشد.

**قاعده شماره ۳:** این قاعده نشان می‌دهد که اگر در فصل تابستان، تخلف ۳۲۷ (عدم توجه به اخطار کتبی و تذکرات بازرسان و مسئولین ذیربط) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه E باشد، با معیار پشتیبان ۰/۳۹٪ و معیار اطمینان ۱۰۰٪ و معیار لیفت ۳۵/۶۸٪ باشد، شرکت متخلف، شرکت الف محسوب می‌گردد.

**قاعده شماره ۴:** این قاعده نشان می‌دهد که اگر در فصل پاییز، تخلف ۲۲۲ (حمل مسافر در محلی غیر از مسیر تعیین شده توسط اتوبوسرانی) اتفاق بیافتد و امتیاز منفی شرکت دارای رتبه E باشد، با معیار پشتیبان ۰/۰۲٪ و معیار اطمینان ۱۰۰٪ و

## استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم.....

### ۳-۴-۲ درخت تصمیم

در سطح دوم این شاخه درجه تخلف قرار دارد، در تخلفاتی که با درجه دو و سه رخ داده اند ۱۰۰٪ رده امتیاز منفی شرکت E بوده است.

۲- شرکت ب: همان طور که در شکل ۵ نمایان است ۲/۳۹ درصد از تخلفات مربوط به این شرکت می باشد و ۵۰٪ از امتیاز منفی شرکت دارای رتبه E و بیش از ۳۴٪ از آن در رده D است.

در سطح دوم این شاخه درجه تخلف قرار دارد، که درجات تخلف یک و دو در رده E و درجه تخلف سه در رده D قرار دارد. همان‌طور که در شکل ۵ ملاحظه می‌گردد، شرکت‌های الف و ب جزء شرکت‌های با میزان تخلف بالا و امتیاز منفی E و D محسوب شده و این قاعده کشف شده در درخت تصمیم میتواند در روند ارزیابی پیمانکاران مورد استفاده قرارگیرد.

یکی از مزیت‌های درخت تصمیم بعنوان یکی از روشهای دسته بندی این است که میزان اهمیت عوامل تاثیرگذار را نشان می‌دهد و این عوامل را در تشکیل درخت تصمیم اولویت بندی کرده و نیز می‌تواند ابزاری برای پیش بینی آینده باشد. در این مرحله از تحقیق تلاش شده است امتیاز منفی شرکت رتبه بندی شده (هدف) براساس نام شرکت (متغیر ورودی)، فصل وقوع (متغیر ورودی)، ماه وقوع تخلف (متغیر ورودی)، کد تخلف (متغیر ورودی) پیش‌بینی شود.

از بین ۱۸ شرکت خصوصی، زیرمجموعه شرکت اتوبوسرانی شهر تهران، دو شرکت الف و ب در سطح امتیاز منفی (E) از نظر تخلفات شرکت و رانندگان شان قرار گرفتند.

### ۳-۵-۱ ارزیابی

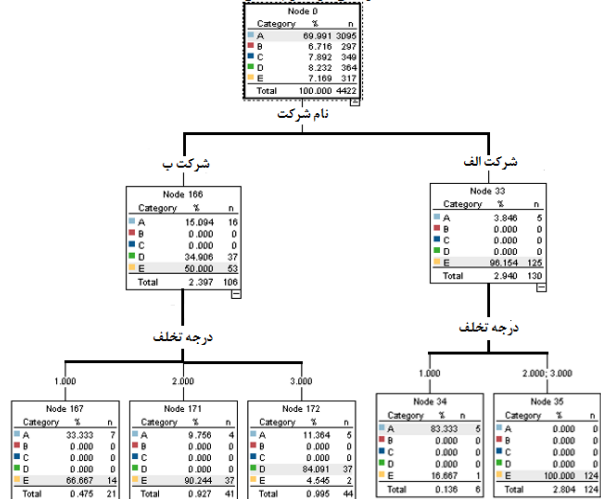
در این مرحله از مدل کریسپ دی ام، بمنظور ارزیابی قواعد کشف شده توسط الگوریتم‌های جی آر آی بعنوان یکی از روشهای قوانین انجمنی لازم است از شاخص های پشتیبان، اطمینان و لیفت استفاده نماییم.

اما در مورد ارزیابی درخت ایجاد شده توسط مدل سی فایو بعنوان یکی از مدل‌های روش دسته بندی لازم است داده های مساله به دو دسته آموزشی و آزمایشی تقسیم بندی شده و ضمن تعیین میزان دقت پیش بینی مدل سی فایو، از دیگر مدل‌های موجود در حوزه ی دسته بندی میتوان دقت مدل بکار رفته در تحقیق را تعیین نمود.

هدف از این مرحله، پاسخ به این سوال است که آیا مدل انتخابی، ما را به هدفی که در مرحله اول تعیین کرده ایم می‌رساند یا خیر؟ در حقیقت صحت نتایج و اهداف داده کاوی با دیدگاه و نقطه نظر سازمانی ارزیابی می‌گردد تا مشخص شود نتایج بدست آمده برای ادامه کار مناسب است و یا باید در مراحل پیشین تجدید نظر گردد [Chapman et al. 2000].

در این بخش میزان دقت الگوریتم سی فایو بعنوان مدل تحقیق با مدل شبکه عصبی مقایسه گردید، که نتیجه اعتبارسنجی و صحت مدل تحقیق با استفاده از الگوریتم C5 معادل ۸۳/۶۳

امتیاز منفی شرکت (رتبه بندی شده)



شکل ۵. نمونه‌ای از درخت تصمیم تحقیق

مطابق شکل شماره ۵ بمنظور تحلیل درخت تصمیم در ابتدا از سطح صفر امتیاز منفی شرکت (رتبه بندی) شروع می‌کنیم. همان طور که در شکل هفت ملاحظه می‌گردد، در بیش از ۶۹٪ از تخلفات امتیاز منفی شرکت ها در رده A بوده است. در سطح یک نام شرکت‌ها نشان داده شده است، در این تحقیق دو شاخه مهم از این سطح بررسی شده است که دو شرکت الف) و ب) رتبه امتیاز منفی E داشته اند.

۱- شرکت الف: همان طور که در شکل چهار نشان داده شده است این شرکت ۲/۹۴ درصد از تخلفات را به خود اختصاص داده است و بیش از ۹۶٪ از تخلفات آن در رده E بوده است.

درصد و شبکه عصبی ۷۹/۹۵ درصد را نشان داد، که این میزان صحت نتایج پیش بینی حاصل از مدل را تایید می نماید.

### ۳-۶ پیاده سازی و توسعه

در این مرحله از نتایج و قواعد بدست آمده از الگوریتم قوانین انجمنی و درخت تصمیم که در قسمت مدل سازی بدست آمده است استفاده می کنیم. مدیران سازمان می توانند از این اطلاعات بدست آمده برای چشم انداز آینده و نیز طرح قوانین جدید استفاده کنند. همچنین این نتایج می تواند در تصمیم گیری جهت تمدید قرارداد شرکت های خصوصی پیمانکار تحت نظارت شرکت اتوبوسرانی راه گشا بوده و باعث حذف شرکت های با میزان کارآیی پایین و تمدید قرارداد شرکت های با میزان کارآیی بالا می شود. همانطوریکه ملاحظه می گردد بعنوان نمونه با بررسی تمامی اطلاعات و نتایج بدست آمده از الگوریتم های قوانین انجمنی و دسته بندی بصورت قواعد کشف شده و درخت تصمیم در خصوص پیمانکاران شرکت اتوبوسرانی، قطع همکاری شرکت های الف و ب با شرکت اتوبوسرانی ضروری به نظر می رسد.

به هر حال در آخرین مرحله مدل کریسپ که جاری سازی و توسعه است، هدف از انجام این مدل، کشف قواعد و رویدادهای بحرانی و نامطلوب است که در بصورت دانش در اختیار تحلیل گران و تصمیم گیرندگان و به تعبیری مدیران عالی سازمانها قرار میگیرد تا با استفاده از این قواعد نامطلوب با تواتر بالا قابلیت اطمینان تصمیم های خود را به حداکثر ممکن ارتقاء بخشیده و ضمن راهبری سازمان با ثبات و پایداری بالاتر، رضایت مشتریان مجموعه خود را افزایش دهند.

### ۴. نتیجه گیری

در تمامی سازمانها بصورت مستمر میزان قابل توجهی داده تولید و بر روی هم انباشته میشود. در صورتیکه بتوانیم با استفاده از روشهای مختلف داده کاوی داده ها را مورد واکاوی و تجزیه و تحلیل قرار دهیم، دانش استخراج شده از درون داده ها میتواند در تصمیم گیری های درون سازمان مورد استفاده قرار گیرد، و عملا در تامین دانش برای اتخاذ تصمیم مدیران عالی سازمان در سطوح مختلف سازمانی راه گشا باشد. زیر مجموعه های ناکارآمد در یک سازمان میتوانند باعث کاهش

سطح کیفیت ارائه خدمات شوند. مدیران سازمان های مختلف با ارزیابی کارکنان و سازمان خود می توانند کیفیت خدمات سازمان خود را افزایش دهند و این باعث رضایت بیشتر مشتریان و در نتیجه افزایش بهره وری و عایدی سازمان می گردد.

در این پژوهش برای اولین بار از داده های GPS برای ارزیابی عملکرد پیمانکاران وابسته به شرکت اتوبوسرانی تهران استفاده شده است. با توجه به فراوانی بروز تخلفات در برخی فصل های سال می توان در آن فصول، نظارت بیشتری بر روی رانندگان شرکت اعمال نمود. همچنین با توجه به تخلفاتی که بیش از سایر تخلفات رخ داده است یا تخلفات مهم و حادثه خیز می توانند جرائم خاصی را در نظر گرفت. بعلاوه با برگزاری کلاس های آموزشی ویژه اتوبوسرانان و عوامل کنترل شرکت های مذکور از بروز و تعدد این تخلفات کاست و درجهت ارتقا کیفیت سرویس دهی به شهروندان محترم از آن بهره برد.

### ۵. پی نوشت ها

- 1- Medical Decision Support Systems
- 2- Global Positioning System
- 3- Crisp DM
- 4- Clementine SPSS-12.0
- 5- Matlab

### ۶. مراجع

- صاحبی، سینا، میریها، بابک، ماهپور، علیرضا و نوروز علیایی، محمد حسین (۱۳۹۴) "ارائه مدل پیش بینی شدت تصادفات عابران پیاده در راه های برون شهری"، فصلنامه مهندسی حمل و نقل، سال ششم، شماره ۴، ص. ۵۸۱-۵۹۲

- Bashir, S., Qamar, U., Hassan Khana, F. and Naseem, L. (2016) "HMV: A medical decision support framework using multi-layer classifiers for disease prediction", Journal of Computational Science, Vol. 13, pp.10-25

- Cedeno, A., Chausa, P., Garcia, A., Caceres, C. and Tormos, J. (2013) "Data mining applied to the cognitive rehabilitation of patients with

- Izad Shenas, A., Raahemi, B., Tekieh, M. and Kuziemy, C. (2014) "Identifying high-cost patients using data mining techniques and a small set of non-trivial attributes", *Computers in Biology and Medicine* Vol. 53, pp. 9–18
- Jianqin, Z., Peiyuan, Q., Yingchao, D., Mingyi, D. and Feng, L. (2015) "A space-time visualization analysis method for taxi operation in Beijing", *Journal of Visual Languages and Computing Part A*, Vol. 31, pp.1–8
- Meng, X., Huang, Y., Roa, D., Zhang, Q. and Liu, Q. (2013) "Comparison of three data mining models for predicting diabetes or prediabetes by risk factors", *Kaohsiung Journal of Medical Sciences*, Vol. 29, No.2, pp. 93-99
- Mirabadi, A. and Sharifian, S. (2010) "Application of association rules in Iranian Railways (RAI) accident data analysis", *Safety Science*, Vol. 48, No.10, pp. 1427-1435
- Mistikoglu, G., Halil Gerek, I., Erdis, E., Usmen, M., Cakan, H. and Kazan, E. (2015) "Decision tree analysis of construction fall accidents involving roofers", *Expert Systems with Applications*, Vol. 42, No. 4, pp. 2256–2263
- Ona, J., Ona, R. and Calvo, F. (2012) "A classification tree approach to identify key factors of transit service quality", *Expert Systems with Applications*, Vol. 39, No. 12, pp. 11164 –11171
- Pakgozar, A., Sigari Tabrizi, R., Khalili, M. and Esmaeili, A. (2011) "The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach", *Procedia Computer Science* Vol. 3, pp. 764–769
- Prati, G., Pietrantoni, L. and Fraboni, F. (2017) "Using data mining techniques to predict the severity of bicycle crashes", *Accident Analysis and Prevention*, Vol. 101, pp. 44–54
- Qin, G., Li, T., Yu, B., Wang, Y., Huang, Z. and Sun, J. (2017) "Mining factors affecting taxi drivers' incomes using GPS trajectories", *Journal of Transportation Research*, Part C Vol. 79, pp. 103–118
- Sami, A., Najafi, A., Yamini, N., Moafian, G. Aghabeigi, M. and Lankarani, K. (2013) "Educational level and age as contributing acquired brain injury", *Expert Systems with Applications*, Vol. 40, No. 4, pp. 1054-1060
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T. and Reinartz, T. (2000) "CRISP-DM 1.0: Step-By-Step Data Mining Guide", pp. 26-27
- Ching, C., Yao, H. and Wu, T. (2013) "Applying data mining techniques to analyze the causes of major occupational accidents in the petrochemical industry", *Journal of Loss Prevention in the Process Industries*, Vol. 8, No. 6, pp.1269-1278
- Delen, D., Walker, G. and Kadam, A. (2005) "Predicting breast cancer survivability: a comparison of three data mining methods", *Artificial Intelligence in Medicine* Vol. 34, No. 2, pp. 113-127
- Ebrahimzadeh Pilerood, A., Mahdavi Mazdeh, M. and Ghousi, R. (2014) "Integration of data mining and stochastic dynamic programming to present a research framework for after-sales service", *Applied Mathematics in Engineering, Management and Technology*, Vol. 2, No. 1, pp. 155-165
- Esmaieeli Sikaroudi, A., Ghousi, R. and Esmaieeli Sikaroudi, A. (2015) "A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing)", *Journal of Industrial and Systems Engineering*, Vol. 8, No. 4, pp.106-121
- Ghousi, R. (2015) "Applying a decision support system for accident analysis by using data mining approach: A case study on one of the Iranian manufactures", *Journal of Industrial and Systems Engineering*, Vol. 8, No. 3, pp. 59-76
- Ghousi, R., Mehrani, S. and Momeni, M. (2012) "Application of data mining techniques in drug consumption forecasting to help pharmaceutical industry production planning", *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management Istanbul, Turkey, July 3 – 6*, pp. 1521-1530
- Gurbuz, F., Ozbakir, L. and Yapici, H. (2011) "Data mining and preprocessing application on component reports of an airline company in Turkey", *Expert Systems with Applications*, Vol. 38, No. 6, pp. 6618-6626

- Wang, J., Zheng, Y., Li, X., Yu, C, Kodaka, K. and Li, K. (2015) "Driving risk assessment using near-crash database through data mining of tree-based model", *Accident Analysis and Prevention*, Vol. 84, pp. 54-64
- Wang, K., Simandl, J., Porter, M., Graettinger, A. and Smith, R. (2016) "How the choice of safety performance function affects the identification of important crash prediction variables", *Accident Analysis and Prevention* vol. 88, pp. 1–8
- Weng, J., Zhu, J., Yan, X. and Liu, Z. (2016) "Investigation of work zone crash casualty patterns using association rules", *Accident Analysis and Prevention*, Vol. 92, pp. 43–52
- Yang, C., Lin, T., Chen, M. and Shi, H. (2009) "improving scheduling of emergency physicians using data mining analysis", *Expert Systems with Applications* Vol. 36, No.2, pp. 3378-3387
- factors to road traffic accidents, *Chinese Journal of Traumatology*", Vol. 16, No. 5, pp. 281–285
- Shearar, C. (2000) "The CRISP-DM model: The new blueprint for data mining", *Journal of Data Ware Housing*, Vol. 5, pp.13-22
- Sun, J. and Li, H. (2008) "Data mining method for listed companies' financial distress prediction", *Knowledge-Based Systems*, Vol. 21, No.1, pp. 1–5.
- Tavakoli Kashani, A., Rabieyan, R. and Besharati, M. (2014) "A data mining approach to investigate the factors influencing the crash severity of motorcycle pillion passengers", *Journal of Safety Research*, Vol. 51, pp. 93–98
- Tso, G. and Yau, K. (2007) "Predicting electricity energy consumption: A comparison of regression analysis", *Decision Tree And Neural Networks* Vol. 32, No. 9, pp. 1761–1768

## استفاده از تکنیک‌های داده‌کاوی در تامین دانش برای یک سیستم.....

اردلان بادامی، درجه کارشناسی خود در رشته مهندسی صنایع را در سال ۱۳۹۲ از دانشگاه آزاد اسلامی واحد فیروزکوه و درجه کارشناسی ارشد خود در رشته مهندسی صنایع - صنایع را در سال ۱۳۹۴ از دانشگاه آزاد اسلامی واحد تهران جنوب اخذ نمود. زمینه های پژوهشی مورد علاقه ایشان تجزیه و تحلیل داده های مهندسی در حوزه های کاربردی مهندسی صنایع است.



روزبه قوسی، درجه دکتری خود در رشته مهندسی صنایع را در سال ۱۳۹۱ از دانشگاه علم و صنعت ایران اخذ نمود. در حال حاضر عضو هیات علمی دانشکده مهندسی صنایع، دانشگاه علم و صنعت ایران است. زمینه های پژوهشی مورد علاقه ایشان مدیریت ریسک، تجزیه و تحلیل داده ها، داده کاوی و کشف دانش، مدلسازی فازی و کاربرد آن در حوزه مهندسی صنایع و عارضه یابی واحدهای صنعتی است.



محمد سعیدی مهرآباد، درجه دکتری خود در رشته مهندسی صنایع را در سال ۱۳۷۱ از دانشگاه ویرجینیای غربی کشور آمریکا اخذ نمود. وی از سال ۱۳۷۱ تاکنون عضو هیات علمی دانشکده مهندسی صنایع دانشگاه علم و صنعت ایران است. زمینه های پژوهشی مورد علاقه نامبرده مدل سازی سیستم های تولید سلولی، تحلیل و بررسی اقتصادی هزینه های تولید و قابلیت اطمینان محصول و مدل سازی ریاضی فرایندهای ساخت و بهینه سازی آنهاست

