تحلیل داده‌های خرابی تجهیزات علائم در راه‌آهن ایران با استفاده از تکنیک‌های داده کاوی

مسعود یقینی*، دانشیار، دانشکدهٔ مهندسی راه‌آهن، دانشگاه علم و صنعت ایران، تهران، ایران
فائزه غفرانی، دانشجوی کارشناسی ارشد، دانشکدهٔ مهندسی راه‌آهن، دانشگاه علم و صنعت ایران، تهران، ایران
سمیه ملا، دانشجوی کارشناسی ارشد، دانشکدهٔ مهندسی راه‌آهن، دانشگاه علم و صنعت ایران، تهران، ایران
مهدی عامره، دانشجوی کارشناسی ارشد، دانشکدهٔ مهندسی راه‌آهن، دانشگاه علم و صنعت ایران، تهران، ایران
بهاره جوانبخت، دانشجوی کارشناسی ارشد، دانشکدهٔ مهندسی راه‌آهن، دانشگاه علم و صنعت ایران، تهران، ایران
پست الکترونیکی نویسنده مسئول: yaghini@iust.ac.ir

دریافت: ۱۳۹۳/۰۵/۲۰– پذیرش: ۱۳۹۳/۱۱/۲۰

Archive of SID

چکیده

حمل ایمن بار و مسافر برای مدیران حمل‌ونقل ریلی، از اهمیت بسیار بالایی برخوردار است. به‌کارگیری روش‌های ایمن حمل‌ونقل، نیازمند شناخت کامل از عوامل به وجود آورنده شرایط غیر ایمن است که این امر با یادگیری از تجربیات گذشته محقق می‌گردد. سیستم علائم و ارتباطات راه‌آهن یکی از بخش‌های بسیار مهمی است که کنترل خرابی قطعات و تجهیزات علائمی و ارتباطی و در نتیجه فراهم نمودن شرایط ایمن حرکت قطارها بخشی از وظایف آن است. در این مقاله به تحلیل داده‌های مربوط به خرابی تجهیزات علائم الکتریکی و ارتباطی در راه‌آهن جمهوری اسلامی ایران پرداخته شده است. برای این منظور، تعداد ۳۰۲۳۵ اطلاعات ثبت شده مربوط به خرابی‌های تجهیزات علائم الکتریکی و ارتباطی از پایگاه داده راه‌آهن جمهوری اسلامی ایران مربوط به سال‌های ۱۳۹۰ و ۱۳۹۱انتخاب شده است. برای تحلیل این اطلاعات از تکنیک قوانین انجمنی Apriori و تکنیک درخت تصمیم C5.0 استفاده شده است. مراحل اجرای پروژه داده کاوی بر اساس متدلوژی CRISP-DM انجام شده است. نتیجه این تحقیق می‌تواند در برنامه‌ریزی اقدامات لازم برای پیشگیری از وقوع خرابی‌ها و کاهش آن‌ها و در نتیجه تأمین ایمنی تردد قطارها استفاده شود.

واژه‌های کلیدی: راه‌آهن ایران، خرابی تجهیزات علائم الکتریکی و ارتباطی، تکنیک‌های داده کاوی، متدولوژی *CRISP-DM*

۱- مقدمه

بشر امروزی با حجم عظیمی از انواع داده‌های حاصل از آزمایش‌ها، نمونه‌ها، الگوها، تصاویر و مشاهدات مواجه است که از طریق حسگرها، دوربین‌ها، میکروفون‌ها، نرم افزارها و یا سایر تجهیزات ساخته دست بشر جمع‌آوری شده‌اند. حجم این داده‌ها بسیار زیاد بوده و پایگاه‌های داده‌ای بزرگی را به وجود آورده‌اند. برای استخراج اطلاعات از این پایگاه‌های داده‌ای از تکنیک‌های داده کاوی استفاده می شود.

داده کاوی یک علم میان رشته‌ای[1] است که تکنیک‌های مختلفی

www.SID.ir

را نظیر یادگیری ماشینی، شناسایی الگو، آمار و پایگاه داده را در کنار هم قرار می‌دهد تا امکان استخراج اطلاعات از پایگاه‌های داده‌ای را ممکن سازد. داده کاوی امکان تحلیل انبوهی از داده‌ها را جهت دستیابی به روابطی که قبلاً مشاهده نشده بود و نیز شناسایی روابط و الگوهای پنهان موجود در داده‌ها را فراهم می‌نماید (Han et al., 2012). یکی از زمینه‌هایی که شاهد سرعت رشد اطلاعات در آن هستیم، سیستم‌های حمل و نقلی و به طور ویژه راه آهن‌های جهان می‌باشد که یکی از پایه‌های اصلی توسعه پایدار و متوازن در جوامع بشری محسوب می‌شوند. یکی از مهم‌ترین نقاط ضعف در شبکه‌های حمل و نقلی به ویژه راه آهن‌ها که عامل افزایش هزینه‌ها و زایل نمودن سرمایه های گزافی می گردد، فقدان ایمنی مناسب می‌باشد. حفظ و ارتقای سطح ایمنی یکی از الزامات اساسی در توسعه صنعت حمل و نقل ریلی محسوب می‌شود. توسعه سیستم‌های ایمنی نیازمند شناخت کامل از نارسایی‌ها و مشکلات موجود در سطح شبکه ریلی می‌باشد. در این میان راه آهن جمهوری اسلامی ایران نیز از این قاعده مستثنی نیست. با توجه به تولید حجم انبوهی از اطلاعات در راه آهن جمهوری اسلامی ایران به خصوص در سیستم‌های علائم و ارتباطات، نیاز به ابزاری مناسب برای تحلیل این داده‌ها وجود دارد تا دستگاه‌های مدرن با قابلیت و ضریب اطمینان بیشتری در زمینه علائم و ارتباطات به کار گرفته شوند و میزان خرابی‌های ناشی از دستگاه‌ها کاهش و در نتیجه ایمنی تردد قطارها تأمین گردد.

در این مطالعه به دنبال شناسایی الگوهای معنادار در داده‌های مربوط به خرابی‌های تجهیزات علائمی و ارتباطی هستیم تا بتوان اقدامات لازم به منظور پیشگیری از وقوع آنها یا کاهش دفعات آنها را اتخاذ نموده و ایمنی تردد قطارها فراهم شود.

بر اساس بررسی‌های انجام شده توسط نویسندگان این مقاله در زمینه تحلیل داده‌های مربوط به خرابی سیستم‌های علائم و ارتباطات راه آهن در ادبیات موضوع مطالعه انجام شده‌ای وجود ندارد و تنها مطالعات اندکی نظیر (Rubaai, 2003)، (Hou and Gupta, 2004) و (Nefti and Oussalah, 2004) در زمینه تحلیل داده‌های خرابی خطوط ریلی انجام شده است.

در این مطالعه، با به کارگیری متدولوژی استاندارد CRISP-DM به عنوان ابزار مورد استفاده برای داده کاوی، به تحلیل تعداد ۳۰, ۲۳۵ داده ثبت شده مربوط به خرابی‌های تجهیزات علائم الکتریکی و ارتباطی در پایگاه داده راه آهن جمهوری اسلامی ایران در سال‌های ۱۳۹۰ و ۱۳۹۱ پرداخته شده است. برای مدلسازی داده‌ها، دو تکنیک استخراج قوانین انجمنی و پیش بینی مورد استفاده قرار گرفته‌اند. به منظور استخراج قوانین انجمنی از الگوریتم *Apriori* و برای پیش بینی مشخصه‌های خرابی تجهیزات از الگوریتم درخت تصمیم *C5.0* استفاده شده است.

سایر بخش‌های مقاله به صورتی که در ادامه بیان می‌شود، سازماندهی شده‌اند. بخش دوم، شامل معرفی متدولوژی و ابزار مورد استفاده برای دستیابی به اهداف مطالعه می‌باشد. در بخش سوم به معرفی راه آهن ایران و به طور ویژه اداره کل علائم و ارتباطات به عنوان سازمان مورد بررسی پرداخته می‌شود. بخش چهارم به شناخت داده‌های مربوط به سیستم علائم و ارتباطات و آماده سازی آنها می‌پردازد. در بخش پنجم مدلسازی و نتایج حاصل از آن ارایه می‌شود. در نهایت بخش ششم شامل نتیجه‌گیری و دستاوردهای مطالعه می‌باشد.

۲- متدولوژی و ابزار مورد استفاده

متدولوژی مورد استفاده در این مطالعه، مدل مرجع *CRISP-DM* می‌باشد. این متدولوژی در بردارنده کلیه مراحل یک پروژه داده کاوی شامل فازها، فعالیت‌های هر فاز و روابط میان آنها می‌باشد. یک پروژه داده کاوی شامل شش مرحله اصلی می‌باشد که در شکل ۱ نشان داده شده‌اند. توالی فازها منعطف است و لازم است به طور مکرر میان فازها رفت و برگشت وجود داشته باشد. خروجی‌های هر فاز تعیین کننده این است که در قدم بعد چه فاز و حتی چه قدم به خصوصی از فاز بایستی صورت بگیرد. جهت فلش‌ها روی شکل ۱، متداول‌ترین ترتیب و توالی میان فازها را نشان می‌دهد (Chapman (et al.), 2000).

یک روش مناسب برای به کارگیری داده کاوی، استفاده از یک برنامه نرم‌افزاری مناسب است که در بردارنده امکانات لازم جهت کاوش داده به طرق متفاوت باشد. با توجه به امکانات تصویری، گستره الگوریتمی و پیروی از مدل مرجع *CRISP*، نرم افزار

Archive of SID
www.SID.ir

Clementine 12.0 که محصولی از شرکت Integral Sol., Ltd. می‌باشد، به عنوان یک ابزار مناسب برای پیاده‌سازی متدولوژی *CRISP* در این مقاله مورد استفاده قرار گرفته است.

هر یک از مراحل اصلی متدولوژی *CRISP* برای داده‌های خرابی‌های سیستم علائم و ارتباطات در راه آهن ایران پیاده سازی و نتایج آنها در بخش‌های بعدی مقاله ارایه شده است.

شکل ۱. فازهای مدل مرجع CRISP- DM (Chapman et al., 2000)

۳- شناخت سازمان: اداره کل ارتباطات و علائم الکتریکی در راه آهن ایران

راه آهن جمهوری اسلامی ایران، راه آهن ملی دولتی می باشد. در سال ۱۳۹۱، ایران با مساحت ۱٦٤٨۱۹٥ کیلومتر مربع و حدود ۷۰ میلیون جمعیت دارای حدود ۱۰۲۲۳ کیلومتر خط ریلی اصلی در ۱۸ ناحیه متفاوت بوده است (سالنامه راه آهن، ۱۳۹۱).

در راه آهن ایران، اداره کل ارتباطات و علائم الکتریکی در تلاش است تا در راستای پاسخگویی به نیازها در زمینه تعمیر، نگهداری، بازسازی، نوسازی و توسعه شبکه و واحدهای فرسوده و جایگزینی آنها با سیستم‌های جدید ارتباطی و علائم الکتریکی، ارزش شرکت راه آهن را نزد ذینفعان (مشتریان، کارکنان، جامعه و مدیران ارشد راه آهن) افزایش دهد. اهداف کاری این اداره بر مبنای توسعه و تعمیر و نگهداری سیستم‌ها بوده و در این راستا در دو بخش علائم الکتریکی و ارتباطات فعالیت می‌نماید.

در بخش علائم الکتریکی، سیستم‌های علائم الکتریکی که شامل سیستم اینترلاکینگ، سیستم بلاک، سیستم محور شمار، سیستم تراک، چراغ علائم (سیگنال)، ماشین سوزن الکتریکی و مگنت می باشند.

Archive of SID

www.SID.ir

می‌باشند، به سه طریقی که در ادامه بیان می‌شود، کنترل می‌گردند:

- کنترل از راه دور (RC) – امکان کنترل دو یا چند ایستگاه را از یک محل با سیستم‌های کنترل از راه دور فراهم می‌آورد. در این سیستم یک مرکز کنترل، توانایی کنترل ترافیک از راه دور برای حداکثر تا چهار ایستگاه را داراست.
- کنترل مرکزی ترافیک (C.T.C)[۴] این سیستم در هر بخش مرکب از یک مرکز کنترل و تعدادی ایستگاه تابعه است و مرکز کنترل می تواند علائم، سوزنها و راه بندهای هر یک از ایستگاه‌های تابعه خود را کنترل نماید. در این سیستم یک مرکز کنترل، توانایی کنترل ترافیک ٤٠ ایستگاه و بیشتر را دارا می باشد.

در بخش ارتباطی، شبکه مخابراتی راه آهن که در سطح کل شبکه ریلی کشور گسترش و توسعه پیدا کرده است، ارتباطات شبکه را تامین نموده و روزانه هزاران پیام و دستور فنی و اداری را از طریق شبکه کابلی و رادیویی در سراسر شبکه راه‌آهن مبادله می‌نماید. تجهیزات ارتباطی شامل خطوط انتقال، کاریر و ماکس، پارتی لاین تصویری، رادیویی و شبکه سراسری داده که مسئولیت ایجاد شبکه‌های بیسیم و مراکز تلفن می‌باشد و در پنج بخش اساسی مورد استفاده قرار می‌گیرند که عبارتند از: شبکه سراسری مراکز تلفن خودکار، بستر خطوط ارتباطی شامل شبکه سراسری فیبر نوری، شبکه سراسری سیستم انتقال (کاریر و ماکس)، شبکه سراسری ارتباطات یکپارچه داده محلی، منطقه‌ای و سراسری راه آهن را عهده دار می‌باشد. یکی از وظایف اداره کل علائم و ارتباطات راه آهن جمهوری اسلامی ایران، گرد آوری، نگهداری و تحلیل داده‌های مربوط به خرابی تجهیزات و سیستم‌های علائمی و ارتباطی است. تحلیل‌های صورت گرفته بر روی این داده‌ها شامل تحلیل‌های آماری توصیفی ساده و مقایسه‌ای با دوره‌های قبل می باشد. برای این منظور آمار و اطلاعات دقیق خرابی‌های بوجود آمده در هر یک از سیستم‌ها و تجهیزات علائمی یا ارتباطی به تفکیک نواحی و ایستگاه‌ها، در طول ٢٤ ساعت شبانه روز توسط ارتباط تلفنی از ایستگاه ها به مراکز نواحی اعلام شده و از طریق این مراکز، به عنوان یک منبع اطلاعاتی درج و به صورت برنامه آمار خرابی توسط اداره کل

- کنترل محلی (LC) – امکان کنترل کلیه تجهیزات مستقر در ایستگاه را از یک محل فراهم می‌نماید.

ارتباطات و علائم الکتریکی منتشر می‌گردد. با استفاده از این سیستم می‌توان با دریافت پیوسته اطلاعات مرتبط با خرابی‌ها، بر کارکرد صحیح شبکه علائم و ارتباطات نظارت داشت. روابط و الگوهای موجود در این داده‌ها را می‌توان با استفاده از تکنیک‌های داده کاوی استخراج نموده و به منظور جلوگیری از وقوع و یا کاهش آنها، تحلیل و بررسی نمود.

٤- شناخت و آماده سازی داده‌ها

٤-١- شناخت داده‌ها

اطلاعات مورد نیاز برای پیش برد مطالعه پیش رو از سوابق، مستندات، اطلاعات و داده‌های موجود در پایگاه داده اداره کل علائم الکتریکی و ارتباطات راه آهن جمهوری اسلامی ایران تهیه شده است. پایگاه داده‌ای مورد استفاده در تحقیق پیش رو، دارای ٨ مشخصه و ٣٠٢٣٥ رکورد بوده و اطلاعات مرتبط با خرابی تجهیزات علائمی و ارتباطی را در بازه زمانی دو سال برای ١٥ ناحیه، در سال‌های ١٣٩٠ و ١٣٩١ را در بر می‌گیرد. اطلاعات قابل استخراج از این پایگاه داده شامل نام ناحیه، نام ایستگاه، نوع تجهیز یا قطعه خرابی، کد خرابی، تاریخ خرابی تجهیز یا قطعه، ساعت شروع خرابی، مدت زمان خرابی و علت خرابی می‌باشد. نواحی مورد بررسی برای مطالعه پیش رو شامل شانزده ناحیه اراک، آذربایجان، اصفهان، زاگرس، یزد، تهران، شمالشرق، شمالغرب، شمال، شرق، قم، لرستان، خراسان، کرمان، جنوب و هرمزگان می‌باشد. دو ناحیه جنوب شرق و فارس به دلیل دارا نبودن سیستم‌های علائم و ارتباطات در لیست نواحی پایگاه داده موجود نمی‌باشند. هریک از نواحی دارای تعدادی ایستگاه می‌باشد که آمار خرابی هر ایستگاه به تفکیک موجود می‌باشد. تجهیزات یا قطعات موجود در پایگاه داده شامل ٧٢ نوع می‌باشند که از نمونه‌های آن می‌توان به ماشین سوزن الکتریکی، بلاک، مدار تراک و اینترلاکینگ اشاره نمود. تعداد ٧٠ کد خرابی در پایگاه داده موجود می‌باشد که برای هر خرابی یک کد در نظر گرفته شده و ثبت شده است. تاریخ خرابی تجهیز یا ایستگاه به

Archive of SID
www.SID.ir

صورت یک مشخصه دسته ای با چهار دسته بهار، تابستان، پاییز و زمستان می‌باشد. مشخصه دسته ای ساعت شروع خرابی شامل چهار دسته نیمه شب (از ساعت ۰۰:۰۰ تا ساعت ۰۵:۵۹)، صبح (از ساعت ۰٦:۰۰ تا ساعت ۱۱:۵۹)، بعداز ظهر (از ساعت ۱۲:۰۰ تا ساعت ۱۷:۵۹) و شب (از ساعت ۱۸:۰۰ تا ساعت ۲۳:۵۹) می‌باشد. مدت زمان خرابی نیز یک مشخصه دسته‌ای با دسته‌های (٤-۰ ساعت)، (۸-٤ ساعت)، (۱۲-۸ ساعت)، (۱٦-۱۲ ساعت)، (۲۰-۱٦ ساعت) و (۲٤-۲۰) می‌باشد. خرابی تجهیزات در پایگاه داده موجود شامل ۱۲۱ علت می‌باشد که از جمله آنها می توان به تجهیزات مکانیکی، تجهیزات الکترونیکی، ریل، افتادن شیء فلزی از قطار و محوطه روی دستگاه‌ها اشاره نمود. نمونه‌ای از اطلاعات موجود در پایگاه داده‌ای مورد استفاده، در جدول ۱ ارایه شده است.

جدول ۱. نمونه‌ای از پایگاه داده به کار رفته برای داده کاوی داده‌های خرابی سیستم‌های علائم و ارتباطات

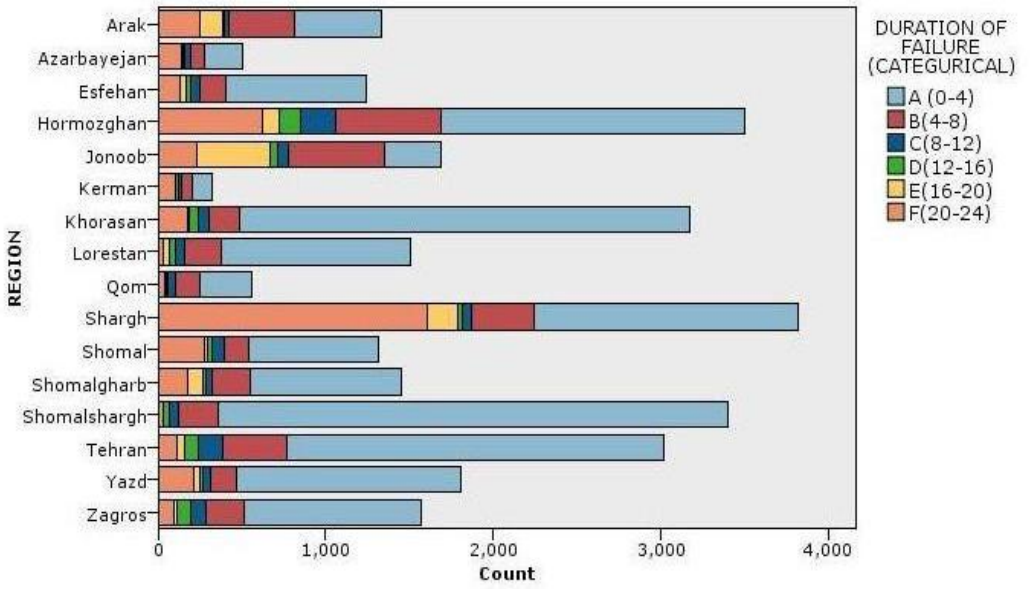
ناحیه	ایستگاه	نوع تجهیز	کد خرابی	علت خرابی	مدت خرابی (ساعت)	زمان شروع خرابی	تاریخ خرابی
اصفهان	ساسان	ماشین سوزن الکتریکی	۱۱	تجهیزات مکانیکی	٤-۰	شب	بهار
لرستان	کشور	بلاک	۵۸	علت نامشخص	۱۲-۸	نیمه شب	تابستان
خراسان	کاشمر	بلاک	۱۲	تجهیزات الکترونیکی	۸-٤	صبح	پاییز
تهران	تهران	مدار تراک	۱۱	تجهیزات مکانیکی	۸-٤	صبح	پاییز
اراک	نورآباد	مدار تراک	۲۲	ریل	٤-۰	بعد از ظهر	زمستان

با بررسی داده‌های موجود در پایگاه داده مورد نظر، مطابق شکل ۲ تعداد تکرار خرابی بر حسب ناحیه و مدت زمان خرابی نشان داده شده است. همانگونه که در شکل ۲ مشاهده می شود، در اکثر نواحی مدت زمان خرابی اکثراً بین صفر تا چهار ساعت به طول انجامیده است. همچنین مشاهده می‌شود که در ناحیه شمال شرق مدت زمان خرابی‌ها حداکثر تا شانزده ساعت بوده است. در سایر نواحی این میزان تا ۲٤ ساعت نیز رسیده است. انجام عملیات نگهداری بهتر و وجود پرسنل کارآزموده تر و نیز نوع و زمان نصب تجهیزات موجود در ناحیه می تواند از دلایل این موضوع باشد. تعداد تکرار خرابی سیستم‌های علائم و تجهیزات بر حسب ساعت شروع و مدت زمان خرابی در نمودار ارایه شده در شکل ۳ نمایش داده شده است.

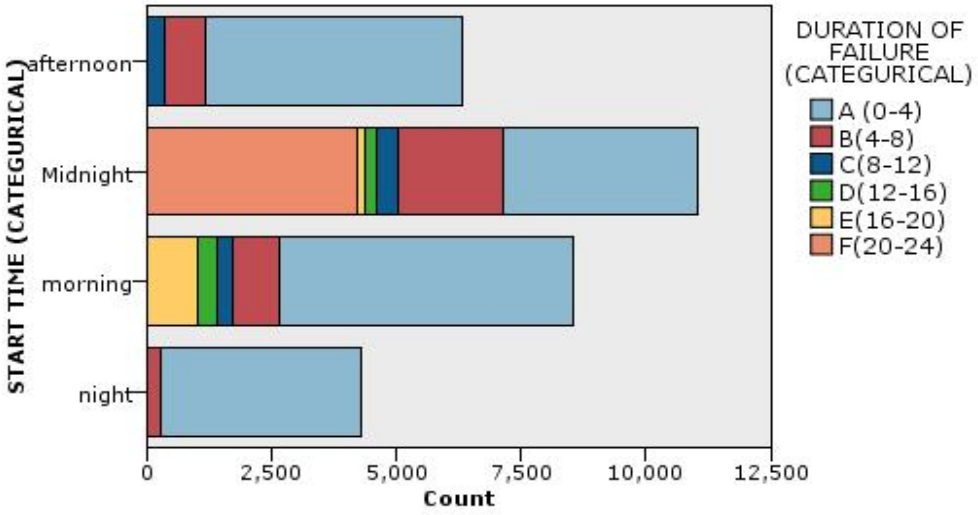
همانگونه که در این شکل مشاهده می‌شود اکثر خرابی‌های رخ داده در نیمه شب آغاز شده‌اند و اکثر آنها بیش از ۲٤ ساعت به طول انجامیده است. به علاوه صرفاً خرابی‌هایی که از نیمه شب آغاز شده‌اند مدت زمان خرابی آنها بیش از ۲٤ ساعت بوده است و سایر خرابی‌ها در مدت زمان کمتر از یک شبانه روز رفع شده‌اند. همچنین این نمودار نشان می‌دهد که خرابی‌های به وجود آمده در شب، نهایتاً هشت ساعت به طول انجامیده است. نمودار ارایه شده در شکل ٤، فصل بروز خرابی را بر حسب تعداد تکرار خرابی‌ها نمایش می‌دهد. طبق این نمودار، فصل بروز خرابی تأثیر چندانی در بروز خرابی نداشته و تعداد تکرار خرابی‌ها در کلیه فصول تقریباً نزدیک به یکدیگر می‌باشد.

Archive of SID

www.SID.ir

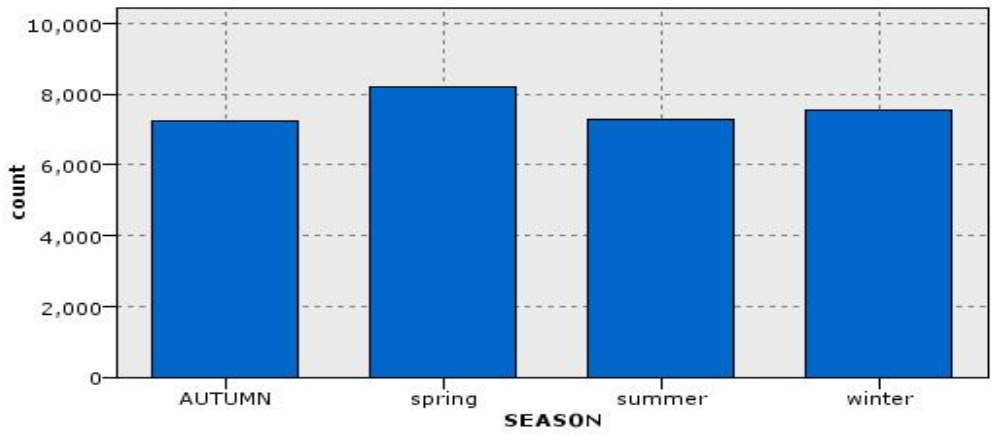


شکل۲. تعداد خرابی‌های تجهیزات علائم و ارتباطات بر حسب ناحیه و مدت زمان خرابی



شکل ۳. تعداد خرابی‌های تجهیزات علائم و ارتباطات بر حسب ساعت شروع و مدت زمان خرابی

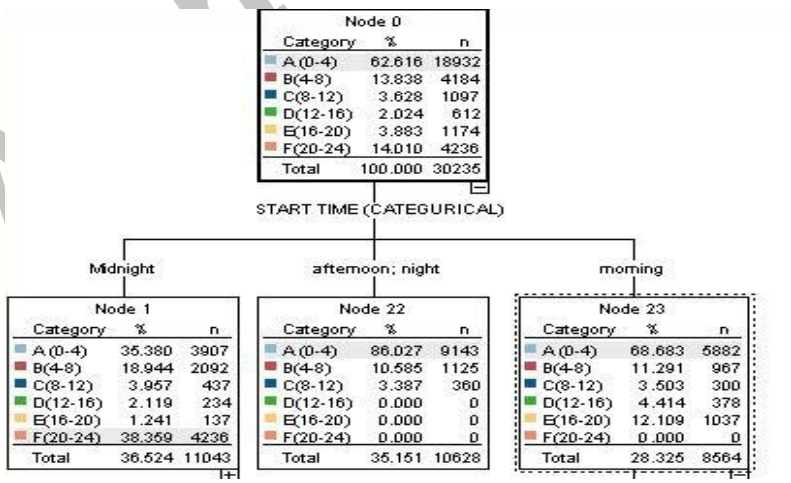
www.SID.ir



شکل٤. نمودار فصل بروز خرابی بر حسب تعداد تکرار

جدول ۲. قوانین استخراج شده با استفاده از الگوریتم *Apriori*

شماره رابطه	اگر	آنگاه	درصد پشتیبانی	درصد اطمینان
۱	مدت زمان خرابی=۲٤-۲۰ ساعت	زمان شروع خرابی= نیمه شب	۱۳/۹۹	۱۰۰/۰۰
۲	زمان شروع= بعداز ظهر	مدت زمان خرابی= ٤-۰ ساعت	۲۰/۹۵	۸۱/۰۵
۳	زمان شروع= شب	مدت زمان خرابی= ٤-۰ ساعت	۱٤/۲۰	۹۳/۳۸
٤	ناحیه= شمالشرق	مدت زمان خرابی= ٤-۰ ساعت	۱۱/۲٦	۸۹/۳٦
٥	ناحیه خراسان	مدت زمان خرابی= ٤-۰ ساعت	۱۰/٤۹	۸٤/٦۱
٦	تجهیز= محور شمار	مدت زمان خرابی= ٤-۰ ساعت	۱۳/٤٦	۸٦/۰۹
۷	تجهیز= مدار تراک	مدت زمان خرابی= ٤-۰ ساعت	۱٥/۳۱	۸۱/۹٤



شکل ٥. بخشی از نمودار درختی مربوط به رابطه بین مدت زمان خرابی و سایر مشخصه‌ها

www.SID.ir

جدول ۳. دقت مدل پیش بینی درخت تصمیم برای مدت زمان خرابی تجهیزات

وضعیت پیش‌بینی	فراوانی	درصد
درست	۲٤۰٦۸	۷۹/٦
نادرست	٦۱٦۷	۲۰/٤
مجموع	۳۰۲۳۵	۱۰۰

شکل ٦. خروجی درخت تصمیم برای تعیین رابطه بین مدت زمان خرابی و سایر مشخصه‌ها

www.SID.ir

٤-٢- آماده‌سازی‌داده‌ها: مناسب‌سازی داده‌ها برای کاوش

در دنیای واقعی، وارد کردن داده‌های خطا و مقادیر غیر معمول امری نسبتاً طبیعی می‌باشد. همچنین امکان این امر وجود دارد که فرمت و شکل ذخیره داده‌ها، با اهداف تحلیل داده‌ها، ناسازگار باشند. لذا آماده سازی داده‌ها پیش از شروع تحلیل داده‌ها امری ضروری است. به بیان دیگر، در دنیای واقعی داده‌هایی که برای تحلیل با تکنیک‌های داده کاوی مورد استفاده قرار می‌گیرند، دارای نقص (نظیر ناقص بودن برخی از مقادیر مشخصه‌ها و عدم وجود برخی از مشخصه‌های مورد نظر)، خطا (نظیر وجود خطا در مقادیر داده‌ها، وجود داده‌های پرت که از میزان مورد انتظار تخطی کرده‌اند) و ناسازگاری (موجود اختلاف در کدگذاری‌های مربوط به مشخصه‌ها در بخش‌های مختلف سازمان) می‌باشند (Han et al., 2012). در این مطالعه، با توجه به تعداد اندک داده‌های از دست رفته در پایگاه داده‌ای موجود، این داده‌ها به صورت دستی اصلاح شده‌اند. به علت تعداد بسیار اندک داده‌های دارای خطا، این نوع داده‌ها از پایگاه داده حذف گردیده‌اند.

٥- مدلسازی و نتایج به‌دست‌آمده

٥-١- تعیین روابط میان مشخصه‌های خرابی تجهیزات با استفاده از قوانین انجمنی

قوانین انجمنی روابط و وابستگی‌های متقابل بین مجموعه بزرگی از اقلام داده‌ای را نشان می‌دهد. وقتی حجم داده‌ها بسیار زیاد باشد، پیدا کردن چنین وابستگی‌ها و قوانینی می‌تواند بسیار مفید بوده و نقش موثری در تحلیل داده ایفا نماید. استخراج این قوانین، به شکل "اگر...آنگاه" با در نظر گرفتن مقادیر مشخصی از درصد پشتیبانی و درصد اطمینان که وابسته به قانون مورد نظر می‌باشد، صورت می‌گیرد (Larose, 2005). استخراج قوانین انجمنی از پایگاه داده مربوط به خرابی تجهیزات علایم الکتریکی و ارتباطی می‌تواند به تحلیل علل خرابی‌ها و ایجاد تمهیدات لازم جهت جلوگیری از وقوع آنها یاری رساند. استخراج قوانین از داده‌های مربوط به خرابی‌های تجهیزات علائم و ارتباطات در این مطالعه، با استفاده از مدل *Apriori* و با در نظر گرفتن حداقل درصد پشتیبانی برابر ١٠ و حداقل درصد اطمینان برابر ٨٠ صورت گرفته است. نتیجه حاصل، شامل ٧ قانون انجمنی استخراج شده از پایگاه داده می‌باشد که در جدول ٢ ارایه شده است. هر یک از قوانین عنوان شده می‌تواند در طراحی عملیات تعمیر و نگهداری تجهیزات و ارتقای آنها موثر باشد. بدیهی است با در نظر گرفتن درصد پشتیبانی و اطمینان بالاتر، تعداد قوانین استخراج شده کاهش می‌یابد. قانون اول از جدول شماره ١ نشان دهنده این است که خرابی‌هایی که مدت زمان آنها بیش از بیست ساعت می‌باشد، از نیمه شب آغاز شده است. این قانون دارای ١٣/٩٩ درصد پشتیبانی و ١٠٠ درصد اطمینان می‌باشد. به این معنا که از کل رکوردهای پایگاه داده ١٣/٩٩٪ مدت زمان خرابی بین ٢٠ تا ٢٤ و از بین این داده‌ها، ١٠٠ درصد آنها زمان شروع خرابی نیمه شب داشته‌اند. دلایل این مسئله را می‌توان در عدم وجود نیروی متخصص در دپوها در نیمه شب و یا عدم حضور آن‌ها در پست کاری، عدم دسترسی به قطعات یدکی و انبارها در نیمه شب یا تأخیر در اعلام خرابی در نظر گرفت. با توجه به دلایل احتمالی اشاره شده، آماده باش و تخصیص پرسنل بیشتر به دپوهای تعمیر و نگهداری، تدارک حداقل قطعات یدکی مورد نیاز و امکان دسترسی سریع به آن در نیمه شب و حضور و توجه کافی پرسنل نگهداری شیفت نیمه شب می‌تواند به کاهش ساعت خرابی‌هایی که از نیمه شب آغاز می‌شود، کمک نماید. قوانین دوم و سوم این جدول حاکی از این است که عملیات نگهداری و تعمیر در بعد از ظهر و شب نسبت به عملیات نگهداری در نیمه شب به طور نسبی موفق تر بوده است و مدت زمان خرابی برای خرابی‌هایی که در این بازه زمانی رخ می‌دهد، صفر تا چهار ساعت به طول انجامیده است. قوانین چهارم و پنجم به دلیل پشتیبانی بالای ١٠ درصد نشان می‌دهند که نواحی شمال شرق و خراسان، دارای تعداد خرابی زیادی در سیستم‌های علائم و ارتباطی بوده‌اند. همچنین به دلیل درصد اطمینان بالای ٨٠ برای دوره خرابی‌های صفر تا چهار ساعت، می‌توان نتیجه گرفت که عملکرد خوبی در تعمیرات خرابی‌ها داشته‌اند. از دلایل آن می‌توان به وجود پرسنل متخصص جهت نگهداری، وجود الگوی مناسب جهت انجام عملیات نگهداری و عمر تجهیزات منصوبه اشاره کرد. لذا عملیات تعمیر و نگهداری انجام شده در این نواحی می‌تواند الگوی سایر نواحی قرار گیرد. قوانین ششم و

Archive of SID
www.SID.ir

هفتم بیان کننده این است که اولاً تعداد زیادی از خرابی‌ها مربوط به سیستم محور شمار و سیستم تراک بوده‌اند. دوماً عملیات رفع خرابی این سیستم‌ها در کمتر از چهار ساعت انجام شده است. وجود نیروی متخصص و آموزش دیده در خصوص نگهداری و تعمیر سیستم محور شمار و سیستم تراک و نیز فراهم بودن قطعه یدکی مورد نیاز این سیستم‌ها را می‌توان از دلایل آن برشمرد.

۵-۲- پیش بینی مشخصه‌های خرابی تجهیزات با استفاده از درخت تصمیم

با استفاده از پیش‌بینی دسته‌ای و عددی، مقدار یک مشخصه دسته‌ای و یا عددی از طریق مدلسازی پیش‌بینی می‌شود. برای ارزیابی مدل پیش بینی ساخته شده، بخشی از داده‌های گذشته در مدل قرار گرفته و با نتیجه به دست آمده در مدل سنجیده می‌شود. مدل‌های پیش‌بینی دسته‌ای می‌توانند خروجی‌های متفاوتی به شکل درخت تصمیم، قوانین و یا شبکه‌های عصبی داشته باشند. مدل‌های پیش‌بینی عددی می‌توانند خروجی‌های متفاوتی به شکل مدل‌های رگرسیونی، قوانین و یا درخت رگرسیون داشته باشند. در ادامه این بخش با استفاده از پایگاه داده موجود در خصوص خرابی تجهیزات علایم الکتریکی و ارتباطی و با استفاده از تکنیک درخت تصمیم، مدت زمان خرابی تجهیزات مورد پیش‌بینی قرار می‌گیرند. درخت تصمیم در ساختن مدل‌های پیش بینی کاربرد زیادی دارد. این درخت مانند یک نمودار است که گره‌های داخلی آن مقدار یک مشخصه خاص را ارزیابی کرده و بر اساس مقادیر موجود، شاخه‌ها و برگ‌های مختلف تولید می‌نماید. برای ساختن درخت تصمیم از الگوریتم‌های متفاوتی استفاده می‌شود. در این مطالعه، از الگوریتم C5.0 جهت پیش بینی مدت زمان خرابی تجهیزات و ساعت شروع خرابی تجهیزات استفاده شده است. اساس کار الگوریتم C5.0 شکستن یک نمونه، به زیر بخش‌های کوچک‌تر بر اساس یک مشخصه می‌باشد به گونه‌ای که بیشترین اطلاعات ممکن را در اختیار ما قرار دهد. هر زیر بخش از نمونه که خود توسط بخش اول تعریف می‌شود، بر مبنای یک مشخصه دیگر به زیربخش‌های کوچک‌تر شکسته می‌شود و این فرآیند تا زمانی که امکان شکستن زیربخش‌ها فراهم نباشد، ادامه پیدا می‌کند. در نهایت پایین ترین سطح درخت مورد بررسی قرار می‌گیرد و آنهایی که به ارزش مدل کمکی نمی‌کنند حذف شده و یا هرس می‌شوند. خروجی مدل C5.0 به شکل نمودار درختی و نیز به شکل مجموعه ای از قوانین می‌باشد. بخش کوچکی از نمودار درختی مربوط به خروجی درخت تصمیم برای پیش‌بینی مدت زمان خرابی تجهیزات از نرم افزار *Clementine* در شکل ۵ ارایه شده است. همانگونه که در شکل ۵ مشاهده می‌شود، گره اولیه در درخت تصمیم بر مبنای مشخصه مربوط به زمان شروع خرابی به سه زیر بخش شکسته شده است. سه زیرگره ایجاد شده نشان می‌دهد که در ۳۸٪ خرابی‌هایی که نیمه شب، در ۸۸٪ خرابی‌هایی که بعد از ظهر و شب و در ٦٨٪ خرابی‌هایی که صبح بوقوع پیوسته است، مدت زمان خرابی به ترتیب بین (۲٤-۲۰)، (٤-۰) و (٤-۰) ساعت بوده است. با توجه به گستردگی شکل ایجاد شده توسط درخت تصمیم، از نمایش سایر زیربخش‌ها در شکل ۵ صرف نظر شده است است. بخشی از قوانین خراج شده از درخت تصمیم مربوط به رابطه بین مدت زمان خرابی و دیگر مشخصه‌ها در شکل ٦ نشان داده شده و در ادامه تشریح می‌شود.

- ✓در صورتی که نوع تجهیز خراب شده فیبر نوری باشد و خرابی به ترتیب در فصل‌های پاییز، بهار، تابستان و زمستان رخ داده باشد مدت زمان خرابی به ترتیب بین (١٦-١٢)، (٨-٤)، (١٢-٨) و (٤-۰) ساعت خواهد بود.
- ✓چنانچه نوع تجهیز مربوط به *NEC12*، *SIEMENS*، پارتی لاین تصویری، روتر ٣٦٠٠ و یا روتر ٣٧٠٠ باشد، مدت زمان خرابی بین ۲۰ تا ٢٤ ساعت طول می‌کشد. همانگونه که در قوانین استخراج شده از درخت تصمیم مشاهده می‌کنیم، مجموعه قوانین بر اساس رابطه میان مدت زمان خرابی با نوع تجهیزات، ساعت شروع خرابی، فصل و ناحیه‌ای که در آن خرابی به وقوع پیوسته است، مشخص شده‌اند. مدل ارایه شده دارای دقت ۷۹٪ می‌باشد که میزان قابل قبولی را دارا است و لذا می‌تواند در زمینه خرید قطعات یدکی مورد نیاز، خرید تجهیزات جدید جهت جایگزینی به جای تجهیزات قدیمی و تدارک انبارها یاری رساند.

Archive of SID
www.SID.ir

۷- نتیجه‌گیری

این مطالعه با هدف تحلیل خرابی‌های تجهیزات علائمی نواحی راه آهن جمهوری اسلامی ایران برای افزایش ایمنی تردد قطارها تعریف شده است. با بکارگیری متدولوژی *CRISP*، تعداد ۳۰۲۳۵ اطلاعات ثبت شده مربوط به خرابی‌های تجهیزات علائم الکتریکی و ارتباطی از پایگاه داده راه آهن جمهوری اسلامی ایران مربوط به سال‌های ۱۳۹۰ و ۱۳۹۱ برای کاوش مورد استفاده قرار گرفت. قوانین انجمنی و درخت تصمیم به عنوان تکنیک‌های مدلسازی پیاده سازی و اجرا شدند. به منظور استخراج قوانین انجمنی الگوریتم *Apriori* و برای درخت تصمیم الگوریتم *C5.0* به کار گرفته شد. نتایج حاصل از مدلسازی حاکی از آن است که در شیفت نیمه شب، با توجه به پرسنل ناکافی و عدم دسترسی مناسب به انبارها و تجهیزات، در اکثر نواحی مدت زمان خرابی طولانی بوده و در حرکت قطارها اختلال ایجاد می‌شود. همچنین نتایج نشان دهنده عملکرد مناسب دو ناحیه خراسان و شمال شرق در خصوص تعمیر و نگهداری تجهیزات است و مدت زمان خرابی تجهیزات در این نواحی نسبت به سایر نواحی کم می‌باشد. در نتیجه الگوی تعمیر و نگهداری این دو ناحیه می‌تواند در سایر نواحی مورد توجه قرار گیرد. از سوی دیگر، ناحیه شرق، به ویژه در شیفت نیمه شب دارای عملکرد نسبتاً ضعیفی نسبت به دیگر نواحی می‌باشد و در اکثر موارد مدت زمان خرابی تجهیزات شیفت نیمه شب در این ناحیه بیش از بیست ساعت به طول انجامیده است. لذا با تخصیص نیروی متخصص بیشتر در این نواحی و نیز با فراهم کردن قطعات مورد نیاز لازم است که در خصوص بهبود عملکرد و کاهش زمان خرابی تجهیزات این نواحی اقدام شود.

۸- منابع

- راه آهن جمهوری اسلامی ایران، دفتر آمار و فناوری اطلاعات، سالنامه آماری حمل‌ونقل ریلی کشور، ۱۳۹۱.

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.

- Han, J., Kamber, M., & Pei, J. (2012). Data mining: concepts and techniques. Morgan kaufmann.

- Hou, Z. G., & Gupta, M. M. (2004, July). A rail damage detection and measurement system using neural networks. In Computational Intelligence for Measurement Systems and Applications, 20.

- Rubaai, A. (2003). A neural-net-based device for monitoring Amtrak railroad track system. Industry Applications, IEEE Transactions on, 39(2), 374-381.

- Nefti, S., & Oussalah, M. (2004, October). A neural network approach for railway safety prediction. In Systems, Man and Cybernetics, 2004 IEEE International Conference on Vol. 4, pp. 3915-3920. IEEE.

Archive of SID

www.SID.ir