پیش بینی تاخیرات قطارهای مسافری با استفاده از یادگیری ماشین مطالعه موردی راه اهن جمهوری اسلامی ایران

مقاله علمی – پژوهشی

سید حسام الدین ذگردی*، استاد، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران
فرحان ثاقب ملکی، دانش آموخته کارشناسی ارشد، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران
الهام آخوندزاده، استادیار، دانشکده مهندسی صنایع و سیستم‌ها، دانشگاه تربیت مدرس، تهران، ایران
*پست الکترونیکی نویسنده مسئول: zegordi@modares.ac.ir

دریافت: ۱۴۰۳/۰۲/۰۸– پذیرش: ۱۴۰۳/۰۸/۲۰
صفحه ۲۵۲-۲۴۱

چکیده

هدف از این مقاله، پیش‌بینی تاخیرات قطارهای مسافری در نواحی ریلی راه آهن جمهوری اسلامی ایران با استفاده از داده‌های تاریخی حرکت قطارهای مسافری و داده‌های آب و هوایی با روش‌های یادگیری ماشین است. پیش‌بینی تاخیرات در نواحی ریلی با شرایط آب و هوایی در آن ناحیه در فصل زمستان می تواند برای تصمیم گیری و اقدامات پیشگیرانه موثر واقع شود. داده‌های بکار گرفته شده در این مطالعه شامل داده های تاخیرات قطارهای مسافری از سال ۱۳۹۶ تا ۱۴۰۰ و داده‌های آب و هوایی ایستگاه‌های سینوپتیک از سال ۱۳۹۶ تا ۱۴۰۰ است. که در بردارنده ۴۶۵۹۶ رکورد است. متغیرهای مستقل شامل سال، ماه، روز ماه، روز هفته، محور حرکت، نوع قطار، ناحیه ریلی، حداکثر سرعت باد، حداقل دید افقی، دمای کمینه، دمای بیشینه، تعداد گزارش‌های یخ‌زدگی در سطح زمین و بارش ۲۴ ساعته باران و برف هستند. روش پیشنهادی برای حل مسئله موجود مبتنی بر CRISP-DM یک متدولوژی برتر در زمینه پیاده‌سازی تکنیک‌های یادگیری ماشین و داده کاوی در حوزه تحقیقاتی و اجرایی است. مدلسازی پیش‌بینی به صورت طبقه‌بندی انجام شده است. جهت پیش‌بینی طبقه‌بندی متغیر وابسته تاخیر به دو طبقه به موقع و با تاخیر تقسیم شده‌اند. از روش‌های یادگیری نظارت شده از نوع طبقه‌بندی برای پیش بینی تثیر عوامل آب و هوایی در به وجود آمدن تاخیر در نواحی ریلی استفاده شده‌است. برای ارزیابی نتایج پیش‌بینی از اعتبارسنجی متقابل برای بررسی صحت مدل استفاده شده‌است. نتایج نشان می‌دهد که تاثیر عوامل آب و هوایی در فصل زمستان در طی دوره ۵ ساله در بررسی‌های سال به سال در به وجود آمدن تاخیر در نواحی ریلی تاثیر مثبت، منفی یا خنثی داشته‌است. در انتهای مقاله برای انطباق صنعت ریلی با تهدیدات اقلیمی در آینده اقدامات پیشگیرانه‌ای ارائه شده‌است.

واژه‌های کلیدی: آب و هوا، پیش‌بینی، تاخیرات، راه آهن، قطارهای مسافری، یادگیری ماشین

۱- مقدمه

قابلیت اطمینان و وقت شناسی از معیارهای کلیدی ارزیابی خدمات راه آهن برای مسافران و اپراتورها است. تأخیرها در ابعاد مکانی و زمانی به طور قابل توجهی بر قابلیت اطمینان و زمانبندی عملکرد قطار تأثیر میگذارد. بهینه سازی استفاده از ظرفیت و طراحی جدول زمان‌بندی مستلزم پیش‌بینی قابلیت اطمینان و زمان دقیق عملیات قطار است که با تأخیر و انتشار تأخیر قطار تعیین می‌شود. برای پیش‌بینی زمان دقیق عملیات قطار، توزیع تاخیرهای ورود و خروج باید با در نظر گرفتن ساختار پیچیده شبکه راه آهن و انواع مختلف تاخیر ناشی از درگیری مسیرها و سفرهای متصل تا حد ممکن واقع بینانه تخمین زده می‌شود [Li et al., 2021]. تحلیل و پیش‌بینی تأخیر قطارهای مسافری موضوع مهمی در حوزه حمل و نقل ریلی است. بسیاری از روش‌های پیش‌بینی که در حال حاضر توسط اپراتورهای راه آهن مورد استفاده قرار می‌گیرد، هنوز بر

اساس قوانین ساده ایستا یا بر اساس دانش متخصصان استوار است و نتایج آنها اغلب مبهم است. پیش‌بینی‌های نادرست روند برنامه‌ریزی زمان واقعی برای اپراتورهای راه آهن را پیچیده می‌کند و بر رضایت مشتری تأثیر منفی می‌گذارد. با توجه به این که امروزه حجم زیادی از داده‌های تأخیر وجود دارند، تحلیل های آماری نمی توانند موثر باشند روش‌های یادگیری ماشین به طور گسترده‌ای در پیشبینی تاخیر قطار مورد استفاده قرار می‌گیرد [2020,Hauck & Kliewer]. علیرغم این واقعیت که وقت‌شناسی در مقایسه با سایر حمل و نقل‌ها از مزایای سفرهای ریلی است، اغلب در قطارها تأخیر رخ می‌دهد. با توجه به این که برنامه زمانبندی سیر و حرکت قطارها براساس برنامه‌ریزی ارائه شده از سوی راه آهن جمهوری اسلامی ایران انجام می‌شود. تأخیر قطارها از مبدأ تقریباً در حد صفر است، اما بیشتر این تأخیرها در زمان سیر و در طول مسیر اتفاق می‌افتد که یک خطه بودن اغلب خطوط راه آهن، محدودیت سرعت در برخی از مناطق ریلی، طول عمر ناوگان، تلاقی قطارهای باری و مسافری، طول عمر لوکوموتیوها، شرایط زیر ساختی، و شرایط جوی از جمله مسائلی است که موجب تاخیر در سیر قطارهای مسافری و افزایش نارضایتی مسافران شده است. به دلیل تأثیرات غیرقابل کنترل خارجی و استفاده زیاد از زیرساخت‌ها، نمی‌توان به طور کامل از تاخیر قطارها جلوگیری کرد. به ویژه در شبکه‌های بزرگ راه آهن، که تأخیرها به صورت روزانه رخ می‌دهد و هنوز هم یک مشکل عمده برای شرکت‌های راه آهن است. مطالعه ادبیات نشان می‌دهد که عامل شرایط آب و هوایی، نقش مهمی در سیستم راه آهن ایفا می‌کند و می توانند باعث ایجاد تاخیر در توسعه قطارهای آینده شود. در این مطالعه برای اولین بار در شبکه راه آهن جمهوری اسلامی ایران داده‌های تاخیر قطارهای مسافری در نواحی ریلی را با عوامل آب و هوایی مختص آن نواحی برای بررسی تاثیر شرایط آب و هوایی ترکیب کرده است. ساختار این مقاله به این شکل صورت گرفته است که ابتدا در بخش دوم، پیشینه تحقیق، در بخش سوم، مطابق روش CRISP-DM مسئله موجود حل خواهد شد در نهایت در بخش چهارم به جمع بندی، نتیجه‌گیری و پیشنهاد برای تحقیقات آتی پرداخته می‌شود.

۲-پیشینه تحقیق

در ادامه به بررسی تعدادی از مقالات از سال ۱۹۹٤ تا ۲۰۲۰ در رابطه با پیش‌بینی تاخیر منتشر شده‌است، پرداخته می‌شود. [Carey and Kwieciński, 1994] وابستگی های قطار را به‌طور تحلیلی مدل‌سازی می‌کنند و تأثیر محدودیت‌های پیشروی را ارزیابی می‌کنند. [Hallowell and Harker, 1996] تأخیر قطارهای مسافری را در یک خط نیمه دو ریل، از جمله عدم قطعیت زمان حرکت، پیش‌بینی کردند. [Yuan, 2007] یک مدل پیچیدگی تصادفی تحلیلی را برای تخمین انتشار تاخیر در ایستگاه‌ها با در نظر گرفتن اتصالات انتقال برنامه‌ریزی‌شده و درگیری‌های قطار پیشنهاد می‌کنند. [Berger et al., 2011] از سیاست‌های انتظار، نمایه‌های زمان رانندگی و پتانسیل‌های جبران به صراحت برای ایجاد یک مدل نمودار تصادفی برای پیش‌بینی تاخیر قطار در شبکه‌های بزرگ استفاده می‌کند. [Büker and Seybold, 2012] یک مدل پیچیدگی توزیع احتمال را برای شبکه‌های راه‌آهن بزرگ با فرض زمان‌های فرآیند توزیع نمایی پیشنهاد می‌کنند. [Keyhani et al., 2012, Lemnian et al., 2014] رویکردهای نمودار تصادفی را برای پیش‌بینی قابلیت اطمینان مورد انتظار یک اتصال قطار برنامه‌ریزی شده ارائه می‌کند. [Zilko et al., 2016] روش Copula BN را برای پیش‌بینی طول یک اختلال راه‌آهن پیشنهاد کرد. [Corman and Kecman, 2018] اولین کسانی هستند که BN را برای پیش‌بینی مستقیم تاخیرهای قطار، با فرض تغییرات معمولی تاخیر از گره به گره، اعمال کردند. آنها تجزیه و تحلیل می‌کنند که چگونه اطلاعات جدید در مورد آخرین تاخیرهای قطار مشاهده شده، احتمال تاخیر پیش بینی شده را در سراسر شبکه راه آهن تغییر می‌دهد. [Shi et al., 2019] یک DT تقویت کننده گرادیان را برای به تاخیر انداختن پیش بینی اعمال می‌کند. [Zhang et al., 2020] از یک مدل PN برای پیش بینی تاخیر قطار استفاده می‌کند. آنها یک منطق فازی را برای مدل‌سازی انتشار تاخیر قطار تحت اطلاعات ناقص در مورد وضعیت‌های تاخیر قطار پیاده‌سازی می‌کنند. [Hauck and Kliewer, 2020] از یک مدل LR برای پیش‌بینی تاخیر قطار در شبکه راه‌آهن آلمان استفاده می‌کنند و نشان می‌دهند که مدل LR آنها نتایج مشابهی را با اجرای NN در همان سناریوی مورد آزمایش ارائه می‌دهد. [Huang et al., 2020b] انواع مختلفی از مدل‌های

NN را برای پیش‌بینی تـأخیر قطار با گنجاندن داده‌های ورودی مبتنی بر وضـــعیت، مکان و مســـیر اعمال می‌کند. ما می‌توانیم مشـــاهده کنیم که NN بیشـــتر برای پیش‌بینی تاخیر قطارهای مسـافری با تمرکز اخیر بر استفاده بلادرنگ استفاده شده است. [Huang et al., 2020c] از ترکیبی از SVM و فیلتر کـالمن برای پیش بینی زمـان حرکت قطار در هنگام اختلالات راه آهن اسـتفاده می‌کند. آنها نشان می‌دهند که کاهش خطای پیش بینی با رویکرد آنها امکان پذیر اسـت، به ویژه برای قطارهایی که در حال حاضـر با تاخیر در حال حرکت هستند. در نهایت، [Gao et al., 2020] از یک مدل RF دو مرحله‌ای برای افزایش دقت پیش بینی زمـان بـازیـابی قطـار اســتفـاده می‌کند. با توجه به بررسی‌های صورت گرفته در پیشینه تحقیق در پایگاه‌های علمی در دسترس به نظر می‌رسد تحقیقی با در نظر گرفتن ویژگی‌های اب و هوایی در شـرایط شـبکه ریلی ایران انجام نشـده اسـت. مطالعه حاضـر مهم‌ترین عوامل موثر بر تاخیر قطارها در نواحی ریلی شـــبکـه راه آهن ایران از داده‌های آب‌وهوایی و داده‌های تاخیر قطارهای مســـافری را ترکیب می‌کند. روش‌های مختلف یادگیری ماشین به طور گسترده در پیش‌بینی تاخیر قطار استفاده شـــده اســـت. با این حال، مدل با عملکرد بهتر از یک مطالعه به مطالعه دیگر بسـته به مورد داده اسـتفاده شده متفاوت است. این تحقیق، اندازه مجموعه داده‌ها را با افزودن ۵ سـال و ادغام سـایر عوامل خارجی مانند شـرایط آب و هوایی که می‌تواند بر تأخیر قطار در سیستم‌های راه آهن تأثیر گذارد را بررسیمی‌کند.

۳- متدلوژی تحقیق

رویکرد پژوهشی برای حل مسئله موجود متدلوژی CRISP-DM اسـت. این روش یک متدولوژی برتر در زمینه پیاده‌سازی تکنیـک‌های یادگیری و داده کاوی در حوزه‌های تحقیقاتی و اجرایی است. یک استاندارد واقعی و یک مدل فرآیندی مستقل از صـنعت برای به کارگیری پروژه‌های داده کاوی اسـت. این تحقیق به دنبال پیش بینی تاخیرات قطارهای مسافری در نواحی ریلی با اسـتفاده از روش‌های یادگیری ماشـین در شرایط شبکه ریلی جمهوری اســـلامی ایران بـا اســـتفاده از داده‌های تاخیر قطارهای مســـافری و داده‌های آب و هوایی در پاســـخ به این سوال که آیا عوامل آب و هوایی در پیش‌بینی تاخیرات قطارهای مســـافری در نواحی ریلی ایران در فصـــل زمســـتان تاثیر گذار هسـت را بررسـی می‌کند. این متدلوژی در برگیرنده شش فاز اصـلی شـامل: ۱) شـناخت کسـب و کار، ۲) شـناخت داده‌ها، ۳) آماده‌ســـازی داده‌ها، ٤) مدلســـازی، ٥) ارزیابی و ٦) پیاده سازی است. مراحل بعدی این تحقیق براساس فازهای شماره ۲ تا ۵ این متدلوژی محسوب می‌شود.

۳-۱- شناخت کسب و کار

شـبکه ریلی ایران از نظر مدیریت، تعمیر و نگهداری خطوط و وسـائط نقلیه ریلی، بهره برداری و غیره به ۲۱ ناحیه تقسـیم شده‌است. حدود این نواحی، به نوع عوارض جغرافیایی بستگی دارد و دارای مسـافتهای متفاوتی بوده و با تقسـیمات کشـوری مطابقت ندارد. در شبکه ریلی چند ناحیه پشت سرهم یک مسیر کـه در اصـــطلاح راه آهن محور می‌گویند تشـــکیل می‌دهند، بنابراین، شـبکه راه آهن به ۷ محور تقسـیم شـده‌اسـت که این محورهـا عبـارتند از: آذربایجان، آذربایجان خراســـان، جنوب، جنوب خراسـان، جنوب شرق، جنوب شرق خراسان، و شمال. قطارهای مســـافری بر اســـاس نوع موتور به دو نوع تقســـیم می‌شـوند: قطارهای خودکشـش که شـامل قطارهای ترن‌ست و توربوترن هسـتند و سـایر قطارها که از نوع عادی می‌باشـند. در این تحقیق، منظور از تأخیر قطار در ناحیه که به اختصـــار تأخیر گفته می شـــود اختلاف زمان واقعی ورود قطار به ناحیه مقصد با زمان از پیش تعیین شده در برنامه زمانبندی قطاراست. علت‌هایی که می‌توانند باعث تأخیر قطار شوند شامل: تأخیر در مبدأ، تلاقی با قطارهای باری و مســـافری، نرســـیدن قطارهای باری و مســـافری، خرابی دیزل قطار یا قطارهای دیگر، خرابی ســـالن، خرابی خط، خرابی علائم، ســـانحه و نماز، عوامل آب وهوایی و خطاهای انسـانی هسـتند. همه این موارد در صورتی باعث تأخیر می‌شـوند که در برنامه زمانبندی قطار در نظرگرفته نشـده باشند. برای مثال: تلاقی با قطارهای باری و مسافری که در برنامه زمانبندی قطار در نظرگرفته شـــده‌اند، منجر به تأخیر قطـار نمی‌شـــونـد. تلاقی با قطارهای دیگر در مورد قطارهایی اتفاق می افتد که در مسیر مخالف در حرکت هستند و به مدت زمـانی می‌گویند که یک قطار منتظر می‌ماند تا یک بلاک خالی شود و آن قطار بتواند در آن بلاک سیر کند. درحالیکه، نرسیدن قطـارهای دیگر در مورد قطارهایی اتفاق می‌افتد که در مســـیر مشـترک در حرکت هسـتند و به مدت زمانی می‌گویند که یک قطار منتظر می‌ماند تا قطارهای دیگر به آن رســـیده، و بر طبق زمـانبنـدی یـا اولویـت از آن ســـبقت بگیرند [Yaghini et

[al.,2013]. وظیفه اصلی سازمان هواشناسی کشور در راستای تامین اهداف ملی و بین المللی، حفاظت از جان و مال مردم در مقابل مخاطرات مرتبط با آب، هوا و اقلیم می‌باشد. تحقیق و شناخت کامل اتمسفر و کلیه پدیده‌های جوی وجمع آوری، تنظیم و تجزیه و تحلیل اطلاعات و آمارهای هواشناسی و ارائه و کشف نتایج حاصله بمنظور کمک به برنامه ریزی، پیشرفت و افزایش بازدهی موثر در فعالیت‌های کشاورزی، دامپروری، آب، آبیاری، مخازن آبهای زیر زمینی، آبهای سطحی، شهرسازی، سدسازی، تامین انرژی، تاسیسات شهری، آبرسانی، سلامتی و بهبود ارتباطات هوایی، در یایی، زمینی و مطالعات محیط زیست و سایر فعالیتهای دریایی، آلودگی هوا، تعدیل وضع هوا و غیره از مهمترین کاربرد های هواشناسی می‌باشد.

۳-۲- شناخت داده‌ها

پایگاه داده مورد استفاده در این تحقیق برای داده‌های تاریخی تاخیر قطارهای مسافری، بخشی از داده‌های تأخیر شرکت قطارهای مسافری رجا است که شامل تأخیر قطارهای مسافری در مسیرهایی می‌باشد که عمده تأخیرات شبکه راه آهن است. در این تحقیق محدوده مورد مطالعه شامل قطارهای محلی، قطارهای بین المللی و گردشگری نیست و بازه زمانی از سال‌های ۱۳۹۶ تا ۱۴۰۰ که در آن برای هر سفر ریلی با یک قطار مسافری در تاریخ مشخص، یک رکورد ثبت شده است. این پایگاه داده دارای ۱۲۴۹۱۵ رکورد و دارای ویژگی‌های مربوط به علت‌های تاخیر و تاخیرها در نواحی ریلی و اطلاعات حرکت قطار است. جدول ۱ شامل بعضی از ویژگی‌های این پایگاه داده است.

جدول ۱. ویژگی‌های پایگاه داده تاخیر قطارهای مسافری

عنوان ویژگی	توضیح ویژگی
تاخیر	مدت زمان تاخیر در ناحیه
تاخیر در مبدا	مدت تاخیر در زمان حرکت از مبدا سفر
نوع قطار	براساس نوع موتور تعیین شده است.
نام مالک	نام شرکتی که مالک قطار است.
تاریخ حرکت	تاریخ شمسی که سفر در آن آغاز شده است.
محور	محوری که سفر در آن صورت گرفته است.
علت تاخیر	علتی که باعث تاخیر شده است.
ناحیه ریلی	ناحیه‌ای که سفر در آن صورت گرفته است.
شماره قطار	شماره منحصر به فرد که به قطار اختصاص یافته است.

داده‌های وضعیت آب و هوایی از سازمان هواشناسی کشور از طریق فرم تعاملی آنلاین در دسترس جمع‌آوری شد. داده‌های جمع‌آوری‌شده شامل تاریخ، زمان و شناسایی مختصات مکان و اندازه‌گیری‌های روزانه‌ای بود که میانگین سرعت باد، بالاترین سرعت باد، میانگین دما، حداقل دما، بارش اندازه‌گیری‌شده، حداقل فاصله دید افقی مقیاس‌شده و. مه، باران، برف، طوفان یا تشکل‌های یخی تشکیل شده‌است. داده‌ها از ایستگاه هواشناسی سینوپتیک در ایران سرچشمه می‌گیرند. با این حال، همه ایستگاه‌های هواشناسی همه داده‌های انتخاب شده را ارائه نکرده‌است، بنابراین برخی از آنها نادیده گرفته شده یا نزدیک‌ترین آنها به ایستگاه‌های راه هواشناسی انتخاب شده‌است. جدول ۲ شامل بعضی از ویژگی‌های این پایگاه داده است.

۳-۳- آماده سازی داده‌ها

در این فاز فعالیت‌هایی جهت آماده سازی داده برای مدل‌سازی انجام می‌پذیرد. به اعتقاد بسیاری از متخصصان، این فاز مهم‌ترین و زمان برترین فاز در در فرآیند یادگیری ماشین محسوب می‌شود و گاهی بین ۶۵ تا ۹۰ درصد زمان و هزینه پروژه‌های یادگیری ماشین صرف این فاز می‌شود که شامل تمام فعالیت‌هایی می‌گردد روی داده‌های خام صورت می‌پذیرد تا ورودی‌های مناسبی را برای ابزارهای یادگیری ماشین مهیا سازد در این فاز سه مرحله آماده سازی صورت می‌پذیرد. ۱) آماده

سازی پایگاه داده تاخیر قطارها، ۲) آماده سازی پایگاد داده آب و هوایی و ۳) آماده‌سازی پایگاه داده نهایی

۳-۳-۱-آماده سازی پایگاه داده تاخیر قطارها

در این مرحله همه داده‌های سالیانه در یک پایگاه داده ادغام شدند. رکوردهای نامعتبر شامل قطارهای بین المللی، گردشگری و بدون اعزام از پایگاه داده حذف شده است. رکوردهایی با مقادیر غیر منتظره با پرسش و پاسخ از متخصصان راه آهن پر شده است و بعضی از رکوردها حذف شده است. به منظور کار کردن راحت با ویژگی تاریخ و بررسی دقیق‌تر اجزای تاریخ حرکت، ویژگی تاریخ حرکت در هر دو پایگاه داده آب و هوایی و حرکت قطارها بصورت سال، ماه و روز یکسان سازی شدند و ویژگی‌های روز هفته، روزماه، ماه و سال از تاریخ حرکت استخراج شدند و به پایگاه داده اضافه شدند و ویژگی تاریخ حرکت از پایگاه داده حذف شد. ویژگی نوع قطار که در واقع نوع موتور قطار را مشخص می‌کند و نشان می‌دهد که قطار عادی یا خودکشش است. به صورت یک متغیر طبقه‌بندی به پایگاه داده افزوده شد و در نهایت رکوردهای تکراری از پایگاه داده حذف شد. پس از استخراج و افزودن ویژگی‌های جدید و حذف ویژگی‌های زائد پایگاه داده‌ها تاخیر در جدول۳ خلاصه شده است.

جدول ۲. ویژگی‌های پایگاه داده هواشناسی

عنوان ویژگی	توضیح ویژگی
تاریخ	تاریخی که رکورد ثبت شده‌است.
نام ایستگاه	اسم منحصر به فرد ایستگاه هواشناسی
VVMin	حداقل دید افقی
FF_Max	حداکثر سرعت باد
TMax	دمای بیشینه
Tmin	دمای کمینه
NGF	تعداد گزارشات یخزدگی در سطح زمین
SS24	میزان بارش برف ۲٤ ساعته
RRR24	**میزان بارش باران ۲٤ ساعته**

جدول ٤. ویژگی های پایگاه داده نهایی تاخیر قطار

عنوان ویژگی	توضیح ویژگی
تاخیر	مدت زمان تاخیر در ناحیه
نوع قطار	براساس نوع موتور تعیین شده است.
سال	سال شمسی که سفر در آن انجام شده‌است.
ماه	ماهی که سفر در آن انجام شده‌است.
روز ماه	روز ماه که سفر در آن انجام شده‌است.
روز هفته	روز هفته که سفر در آن هفته انجام شده‌است.
محور	محوری که سفر در آن صورت گرفته است.
ناحیه ریلی	ناحیه‌ای که سفر در آن صورت گرفته است.
شماره قطار	شماره منحصر به فرد که به قطار اختاص یافته است.

۳-۳-۲- آماده سازی پایگاد داده آب و هوایی

در میان ویژگی‌های انتخاب شده، برای تعداد اندکی از رکوردها برخی از آنها توسط ایستگاه‌های هواشناسی ارائه نشده است. بنابراین، ابتدا با درونیابی خطی بر روی ویژگی‌های گمشده با استفاده از ارزش ویژگی‌های رکوردهای دیگر در همان ایستگاه هواشناسی در روزهای قبل و بعد پر شد. با این حال، پس از این مرحله، هنوز برخی از ویژگی‌های برخی رکوردها داده نشده‌است، زیرا از آنجایی که برخی از ایستگاه‌های هواشناسی به هیچ وجه چند مشخصات را اندازه‌گیری نمی‌کنند، درونیابی خطی غیرممکن بود. پس از آن، در مرحله دوم، ویژگی‌های گمشده با رویکرد نزدیک‌ترین فاصله اقلیدسی به ایستگاه هواشناسی برآورد شد. جدول ٤ پایگاه داده نهایی هواشناسی را خلاصه شده‌است.

جدول ٤. ویژگی‌های پایگاه داده نهایی هواشناسی

عنوان ویژگی	توضیح ویژگی
تاریخ	تاریخی که رکورد ثبت شده‌است.
نام ایستگاه	اسم منحصر به فرد ایستگاه هواشناسی
VVMin	حداقل دید افقی
FF_Max	حداکثر سرعت باد
TMax	دمای بیشینه
TMin	دمای کمینه
NGF	تعداد گزارشات یخزدگی در سطح زمین
SS24	میزان بارش برف ٢٤ ساعته
RRR24	میزان بارش باران ٢٤ ساعته

۳-۳-۳- آماده سازی پایگاد داده نهایی

در این مرحله دو پایگاه داده آماده سازی شده را تبدیل به یک پایگاه داده که هر رکورد از این پایگاه داده شامل تاخیر در نواحی ریلی با ویژگی‌های آب و و هوایی و ویژگی‌های تاخیر قطارها است. در این پایگاه داده نهایی فعالیت‌های آماده‌سازی داده‌ها نظیر مقیاس‌بندی ویژگی‌ها، متوازن کردن ویژگی تاخیر در نواحی ریلی و گسسته سازی ویژگی تاخیر در نواحی ریلی به دو طبقه به موقع و با تاخیر انجام شده‌است. در این تحقیق به منظور بررسی تاثیر عوامل آب وهوایی در فصل زمستان داده‌های مربوط به این فصل در طی دوره ۵ ساله از پایگاه داده نهایی استخراج شد.

۳-٤- مدلسازی به صورت طبقه‌ای بر روی پایگاه داده نهایی

۳-٤-۱- داده‌های آموزشی و آزمایشی و اعتبارسنجی

داده‌ها به دو دسته شامل مجموعه داده‌های آموزشی و داده‌های آزمایشی به ترتیب ۸۰ ٪ و ۲٪ تقسیم شده‌اند و ارزیابی مدل‌های طبقه‌بندی بر اساس اعتبارسنجی متقابل (k-fold Cross-Validation) سنجیده شده‌است. از داده‌های آزمایشی برای مقایسه میزان واقعی آن‌ها با میزان تخمین آنها استفاده می‌شود تا دقت مدل دسته‌بندی تعیین شود. ابتدا مدل بر اساس داده‌های آموزشی آموزش می‌بیند و سپس پس از اعتبارسنجی متقابل مدل آموزش دیده به دسته بندی تأخیر در نواحی ریلی با عوامل آب و هوایی برای داده‌های آزمایشی پرداخته می‌شود.

۳-٤-۲- مدل‌سازی تاخیر در نواحی ریلی با و بدون عوامل آب و هوایی با الگوریتم‌های یادگیری نظارت شده از نوع طبقه‌بندی

در این مرحله از داده‌های پیش پردازش‌شده به طبقه‌بندی تاخیر در نواحی ریلی با و بدون عوامل آب و هوایی برای هر سال از دوره ۵ ساله در فصل زمستان با استفاده از الگوریتم‌های شبکه بردار پشتیبانی، جنگل تصادفی و رگرسیون لجستیک پرداخته شده است. ضمناً از کتابخانه Scikit Learn در محیط نرم افزار Python 3.10.7 برای پیاده‌سازی مدل‌های پیشنهادی استفاده شده است.

۳-٤-۳- ورودی‌های مدل

ویژگی‌های ورودی مورد استفاده در مدل‌های پیش‌بینی یکبار ویژگی‌های آب و هوایی همراه ویژگی تاخیر قطارها و یکبار ویژگی تاخیر قطارها، به دلیل بررسی تاثیر تاثیر عوامل آب و هوایی در به وجودآمدن تاخیر در نواحی ریلی در فصل زمستان است.

۳-٤-٤- خروجی‌های مدل

در مساله کلاس بندی تاخیر، خروجی مدل دقیقاً مشخص نیست. هر یک از بر چسب‌های بدست آمده بعد از گسسته‌سازی ویژگی تاخیر در نواحی ریلی یک واحد خروجی را ایجاد می‌کنند. در نهایت مدل پیشنهادی می‌تواند یکی از این برچسب‌ها را پیش‌بینی کند.

۳-٤-٤- ارزیابی نتایج مدل

برای اینکه توانایی مدل آموزش دیده با روش‌های پیشنهادی در دسته‌بندی تاخیر برای نواحی ریلی برای داده‌های آتی مشخص گردد، مدل بر اساس داده های آموزشی که ۸۰٪ از داده ها را شامل می‌شود آموزش می‌بینید، سپس از مدل آموزش دیده برای دسته‌بندی تأخیر در نواحی ریلی با استفاده از ۲۰ ٪ داده‌های آزمایشی استفاده شده‌است. نتایج بعد از پیاده‌سازی الگوریتم‌ها در دو جدول جداگانه به صورت زیر است.

به منظور ارزیابی مدل، گام های فرایند اعتبارسنجی مدل به شرح زیر انجام شده‌اند.

- اعتبارسنجی پایگاه داده

همان طور که در بخش‌های قبلی بیان شده است، از روش‌های استاندارد برای تبدیلات و پیش پردازش روی داده‌ها استفاده شده است؛ در نتیجه، داده مورد استفاده در این تحقیق معتبر است.

- اعتبارسنجی مدل نرم افزاری

به جهت استفاده از روش‌های موجود در نرم افزار پایتون و کتابخانه Scikit Learn استاندارد در این پژوهش، اعتبار مدل نرم افزاری اثبات شده است.

- اعتبارسنجی اجرایی

به منظور اعتبارسنجی اجرایی از صحت کافی مدل، از مدل پیش‌بینی آموزش دیده برای داده‌های آزمایشی نیز استفاده شده‌است. در مقایسه سال به سال نتایج براساس بالاترین صحت پیش‌بینی در همان سال و مقایسه آن با شرایط بدون عوامل آب و هوایی نشان می‌دهد تاثیر عوامل آب و هوایی می‌تواند تاثیر مثبت، منفی و خنثی داشته باشد. پیش‌بینی‌ها به شدت به مجموعه داده‌های در نظر گرفته شده و همچنین منطقه و ساختار سیستم راه آهن در نظر گرفته شده در مطالعه‌های موردی متکی است.به طور خلاصه، مدل‌ها می توانند به طور اساسی تاخیر قطار را پیش بینی کنند، اما همچنان از کاستی های زیر رنج می برند: (۱) دقت این الگوریتم ها محدود است و (۲) نیاز مبرمی به مدل‌های پیش‌بینی بهبود یافته برای پیش‌بینی تاخیرات شبکه ریلی با در نظر گرفتن عوامل خارجی وجود دارد که داده‌های عملیات قطار در دنیای واقعی را پیش‌بینی کند.

جدول ۵. صحت پیش‌بینی الگوریتم‌ها با عوامل آب و هوایی

سال	مدل با داده‌های آب و هوایی	صحت داده‌های آموزشی	صحت داده‌های اعتبارسنجی	صحت داده‌های آزمایشی
۱	جنگل تصادفی	٪ ۸۰,۵۶	٪ ۵۸,۵۰	٪ ۶۲
۱	شبکه بردار پشتیانی	٪ ۶۵,۹۸	٪ ۵۸,۰۵	٪ ۶۷
۱	رگرسیون لجستیک	٪ ۶۷,۷۳	٪ ۶۲,۹۲	٪ ۶۵
۲	جنگل تصادفی	٪ ۷۱,۱۱	٪ ۶۵,۹۶	٪ ۶۸
۲	شبکه بردار پشتیانی	٪ ۷۰,۳۸	٪ ۶۶,۶۷	٪ ۶۷
۲	رگرسیون لجستیک	٪ ۶۶,۰۵	٪ ۶۵,۰۶	٪ ۶۷
۳	جنگل تصادفی	٪ ۷۱,۸۵	٪ ۶۰,۶۳	٪ ۶۱
۳	شبکه بردار پشتیانی	٪ ۶۹,۳۹	٪ ۶۱,۶۲	٪ ۶۱
۳	رگرسیون لجستیک	٪ ۶۲,۱۵	٪ ۶۰,۲۳	٪ ۶۱
٤	جنگل تصادفی	٪ ۷۱,٦٤	٪ ۶۰,۳۳	٪ ۶۱
٤	شبکه بردار پشتیانی	٪ ٦٨,٥٨	٪ ۶۰,۳۳	٪ ۶۰
٤	رگرسیون لجستیک	٪ ٦٨,٤٨	٪ ۵۹,۹۰	٪ ۶۲
۵	جنگل تصادفی	٪ ۸۷,۸۷	٪ ۶۳,۱۱	٪ ۶۸
۵	شبکه بردار پشتیانی	٪ ۷٤,۰٤	٪ ۶۸	٪ ۶۸
۵	رگرسیون لجستیک	٪ ۶۹,۲۵	۶۷,۳۲	٪ ۶۸

جدول ٦. صحت پیش‌بینی الگوریتم‌ها بدون عوامل آب و هوایی

سال	مدل بدون داده‌های آب و هوایی	صحت داده‌های آموزشی	صحت داده‌های اعتبارسنجی	صحت داده‌های آزمایشی
۱	جنگل تصادفی	٪ ۸۰,۵۸	٪ ٥٦,٤٤	٪ ۶۲
۱	شبکه بردار پشتیانی	٪ ۷۳,٤٥	٪ ۶۱,۰۷	٪ ۶۵
۱	رگرسیون لجستیک	٪ ۶۷,۵۰	٪ ۶۶,۳۳	٪ ۶۶
۲	جنگل تصادفی	٪ ۶۹,۸۰	٪ ۶۶,۳۶	٪ ۶۷
۲	شبکه بردار پشتیانی	٪ ۶۹,۵٦	٪ ۶۷,۲۳	٪ ۶۸
۲	رگرسیون لجستیک	٪ ٦٥,٨٤	٪ ۶۵,۱۱	٪ ۶۷
۳	جنگل تصادفی	٪ ۶۹,۹٦	٪ ۶۱,۱۳	٪ ۶۱
۳	شبکه بردار پشتیانی	٪ ۶۹,۳۵	٪ ۶۱,۵۷	٪ ۶۱
۳	رگرسیون لجستیک	٪ ۶۱,۹۸	٪ ۵۵,۱٤	٪ ۶۱
٤	جنگل تصادفی	٪ ۸۲,۲۲	٪ ۶۰,۲۳	٪ ۶۲
٤	شبکه بردار پشتیانی	٪ ٦٨,٧٣	٪ ۶۰,۲۳	٪ ۶۱
٤	رگرسیون لجستیک	٪ ۶۱,۲۱	٪ ۵۹,۹۰	٪ ۶۲
۵	جنگل تصادفی	٪ ۸۳,۸۵	٪ ۶۲,۸۰	٪ ۶۸
۵	شبکه بردار پشتیانی	٪ ۷۳,۲٤	٪ ٦٨,٤٨	٪ ۶۹
۵	رگرسیون لجستیک	٪ ٦٨,٥٩	٪ ٦٦,٥٣	٪ ۶۹

۵-نتیجه‌گیری

در این پژوهش، به پیش‌بینی تاخیرات قطارهای مسافری در نواحی ریلی شبکه راه آهن جمهوری اسلامی ایران با استفاده از یادگیری ماشین پرداخته شد. برای حل مسئله موجود مراحل متدلوژی CRISP-DM برای پایگاه داده تاخیر قطارهای مسافری و پایگاه داده آب و هوایی به کار گرفته شد. داده‌های مورد استفاده در این تحقیق از سال ۱۳۹۶ تا ۱۴۰۰ برای تاخیر قطارهای مسافری و آب و هوایی بود. پیش از انجام مدلسازی، روی داده‌های تأخیر قطار آماده‌سازی و پیش پردازش انجام شد. و متغیر هدف برای دسته‌بندی تاخیر در نواحی شبکه ریلی به دو کلاس به موقع و با تاخیر تقسیم شد. مدلسازی پیش‌بینی تاخیر قطارها در نواحی ریلی بر روی پایگاه داده نهایی با استفاده از روش‌های طبقه‌بندی برای پیش‌بینی کلاس تاخیر و همچنین بررسی تاثیر عوامل آب و هوایی در نواحی ریلی در فصل زمستان در طی دوره ۵ ساله انجام شد. نتایج نشان داد که در مقایسه سال به سال براساس بالاترین صحت پیش‌بینی در همان سال و مقایسه آن با شرایط بدون عوامل آب و هوایی تاثیر عوامل آب و هوایی می‌تواند تاثیر مثبت، منفی و خنثی داشته باشد. پیش‌بینی‌ها به شدت به مجموعه داده‌های در نظر گرفته شده و همچنین منطقه و ساختار سیستم راه آهن در نظر گرفته شده در مطالعه‌های موردی متکی است. به طور خلاصه، مدل‌ها می توانند به طور اساسی تاخیر قطار را پیش بینی کنند، اما همچنان از کاستی های زیر رنج می برند: (۱) دقت این الگوریتم‌ها محدود است و (۲) نیاز مبرمی به مدل‌های پیش‌بینی بهبود یافته برای پیش‌بینی تاخیرات شبکه ریلی با در نظر گرفتن عوامل خارجی وجود دارد که داده‌های عملیات قطار در دنیای واقعی را پیش‌بینی کند. بسیاری از جهان هنوز به شدت به راه آهن برای جابجایی افراد و محصولات متکی هستند. اما زیرساخت‌های راه‌آهن از سیم‌های هوایی گرفته تا ریل‌ها در معرض خطر زیادی از تغییرات آب و هوا و آب و هوای شدید مرتبط با آن هستند. بدون تلاش برای انطباق با تهدیدات اقلیمی آینده، صنعت راه آهن با زیرساخت‌های تخریب‌کننده، خطرات ایمنی و خسارت‌های ناشی از تاخیرات مواجه خواهد شد. اما تاکنون، راه‌آهن برای آن استراتژی چندانی تعبیه نکرده است. آب و هوای شدید ناشی از آب و هوا، از گرم تا سرد و مرطوب تا خشک، چالش‌های جدی برای زیرساخت‌های راه آهن ایجاد می‌کند. امواج گرما می‌تواند باعث کمانش و انبساط مسیرها شود و باعث خروج قطار از ریل شود، در حالی که یخ زدگی می‌تواند به خطوط برق هوایی آسیب برساند. سیل، خاکریزهایی را که خطوط ریل را پوشش می‌دهند، از بین می‌برد و می‌تواند باعث رانش زمین شود که قطارها را مسدود می‌کند، در حالی که خشکسالی باعث فرونشست و خشکی خاک می‌شود، بنابراین مسیرها ناهم تراز هستند. که این عوامل مستعد به وجود آوردن تاخیر در قطارها هستند. پس اقدامات پیشگیرانه‌ای برای مقابله این عوامل در فصل زمستان ارائه می‌شود، که به شرح زیر است.

۱. پیش‌بینی تخصصی آب و هوا مختص شبکه ریلی: ارائه یک پیش‌بینی تخصصی در مورد شرایط خاص آب و هوایی برای شبکه ریلی که می‌تواند روی مسیرها و احتمال تشکیل یخ در ریل سوم تأثیر بگذارد، مفید واقع شود.

۲. تجهیز کل شبکه ریلی به گرمکن های سوزن‌ها و بازرسی‌های به موقع آن‌ها: گرمکن‌های سوزن از یخ زدگی نقاط جلوگیری می‌کنند. هنگامی که دمای ریل به زیر یک سطح معین می‌رسد، به‌طور خودکار فعال می‌شوند. در شرایط سخت آب و هوایی، با بررسی‌های به موقع می‌توان از یخ زدگی آنها جلوگیری کرد.

۳. حصارهای برفی: در مکان‌های کلیدی خاصی که مستعد رانش هستند، حصارهای برفی برای جلوگیری از سرازیر شدن برف روی مسیرها نصب شود.

٤. توسعه جداول زمانی ویژه زمستانی با اپراتورهای قطار: تهیه جدول‌های زمانی برای شرایط اضطراری که از قبل با اپراتورهای قطار توافق شده باشد می‌تواند فعال شود و در صورت احتمال اختلال به مسافران اطلاع داده شود.

۵. استفاده از سیالات ضد یخ و نوارهای گرمایش در ریل¬های هادی برای جلوگیری از تجمع یخ

٦. اطمینان از در دسترس بودن قطعات یدکی برای هر چیزی در راه آهن که به برف و یخ حساس است.

۷. تشکیل ستاد مدیریت بحران شامل نیروهای مکانیکی، مهندسی، عملیات و خدمه اضافی برای اطمینان از عملکرد ایمن و کارآمد راه آهن در شرایط آب و هوایی شدید

۸. پیش‌بینی‌های لازم برای روشن نگه داشتن لکوموتیو و چک کردن خشک کن آنها برای جلوگیری از یخ زدن سیستم ترمز واگن‌ها

این مطالعه فقط به روش یادگیری ماشین نظارت شده از نوع دسته بندی محدود شد. روش‌ها بیشتری باید بررسی شوند.

این تحقیق یک بار در یک بازه زمانی محدود انجام شده است. بدیهی است هرچقدر بازه زمانی گسترده‌تری مورد بررسی قرار گیرد امکان پیش‌بینی دقیق‌تر تأخیر فراهم می‌شود.

دقت مدل را می‌توان از طریق روش‌های فراابتکاری مانند الگوریتم ژنتیک، روش‌های ترکیبی یا یادگیری گروهی بهبود بخشید.

پایگاه داده تأخیر قطارهای مسافری بدلیل نبود اطلاعات جزئی درباره تأخیری است که اتفاق می‌افتد. با مشکل مواجه است با در دسترس بودن از اطلاعات تکمیلی درباره تأخیر مانند: اطلاعات زیر ساختی، خطای عوامل انسانی، اطلاعات مسافران، اطلاعات عملیاتی، اطلاعات دیزل و واگن های قطار، زمان ورود و خروج دقیق از ایستگاه‌های راه آهن، امکان تحلیل و بررسی دقیق تر تأخیر را فراهم می‌کند.

٦- مراجع

-Berger, A., Gebhardt, A., Müller-Hannemann, M., & Ostrowski, M. (2011). Stochastic delay prediction in large train networks. In 11th Workshop on Algorithmic Approaches for Transportation Modelling, *Optimization and Systems*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

-Büker, T., & Seybold, B. (2012). Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning & Management,* 2(1-2), 34-50.

-Carey, M., & Kwieciński, A. (1994). Stochastic approximation to the effects of headways on knock-on delays of trains. *Transportation Research Part B: Methodological*, *28*(4), 251-267.

-Corman, F., & Kecman, P. (2018). Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies,* 95, 599-615.

-Gao, B., Ou, D., Dong, D., & Wu, Y. (2020). A data-driven two-stage prediction model for train primary-delay recovery time. *International Journal of Software Engineering and Knowledge Engineering*, 30(07), 921-940.

-Hallowell, S. F., & Harker, P. T. (1996). Predicting on-time line-haul performance in scheduled railroad operations. *Transportation Science,* 30(4), 364-378.

-Hauck, F., & Kliewer, N. (2020). Data analytics in railway operations: Using machine learning to predict train delays. *In Operations Research Proceedings 2019*. Springer, Cham, 741-747.

-Hauck, Florian, & Kliewer, Natalia. (2020). Data analytics in railway operations: Using machine learning to predict train delays. *In Operations Research Proceedings 2019,* Springer.741-747.

-Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., Peng, Q., & Xu, X. (2020). Modeling train operation as sequences: A study of delay prediction with operation and weather data. *Transportation Research Part E: Logistics and Transportation Review*, 141, 102022.

-Huang, P., Wen, C., Fu, L., Peng, Q., & Li, Z. (2020). A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. *Safety Science,* 122, 104510.

-Lemnian, M., Rückert, R., Rechner, S., Blendinger, C., & Müller-Hannemann, M. (2014). Timing of train disposition: Towards early passenger rerouting in case of delays. In 14th Workshop on Algorithmic Approaches for Transportation Modelling, *Optimization, and Systems*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

-Li, ZhongCan, Wen, Chao, Hu, Rui, Xu, Chuanlin, Huang, Ping, & Jiang, Xi. (2021). Near-term train delay prediction in the Dutch railways network. *International Journal of Rail Transportation,* 9(6), 520-539.

-Marković, Nikola, Milinković, Sanjin, Tikhonov, Konstantin S, & Schonfeld, Paul. (2015). Analyzing passenger train arrival delays with support vector regression. *Transportation Research Part C: Emerging Technologies*, 56, 251-262.

-Shi, R., Wang, J., Xu, X., Wang, M., & Li, J. (2020). Arrival train delays prediction based on gradient boosting regression tress. In Proceedings of the 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2019: Rail Transportation Information Processing and Operational Management Technologies. *Springer Singapore,* 307-315.

-Yaghini, M., Khoshraftar, M. M., & Seyedabadi, M. (2013). Railway passenger train delay prediction via neural network model. *Journal of Advanced Transportation,* 47(3), 355-368.

-Yuan, J. (2007). Dealing with stochastic dependence in the modeling of train delays and delay propagation. *In International Conference on Transportation Engineering,* 3908-3914.

-Zhang, L., Feng, X., Ding, C., & Liu, Y. (2020). Mitigating errors of predicted delays of a train at neighbouring stops. *IET Intelligent Transport Systems,* 14(8), 873-879.

-Zilko, A. A., Kurowicka, D., & Goverde, R. M. (2016). Modeling railway disruption lengths with Copula Bayesian Networks. *Transportation Research Part C: Emerging Technologies*, 68, 350-368.

Passenger Trains Delay Prediction via Machine Learning (Case Study: Railways of the Islamic Republic of Iran)

Seyed Hessameddin Zegordi, Professor, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran.

Farhan Sagheb Maleki, M.Sc., Grad., Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran.

Elham Akhondzadeh, Assistant Professor, Faculty of Industrial & Systems Engineering, Tarbiat Modares University, Tehran, Iran.

E-mail: zegordi@modares.ac.ir

Received: June 2024- Accepted: September 2024

ABSTRACT

The purpose of this article is to predict the delays of passenger trains in the railway areas of the Islamic Republic of Iran using historical data on passenger trains and weather data with machine learning methods. Forecasting delays in railway areas with weather conditions in that area in the winter season can be effective for decision-making and preventive measures. The data used in this study includes passenger train delay data from 2017 to 2022 and weather data from synoptic stations from 1396 to 1400, which contains 46596 records. Independent variables include year, month, day of the month, day of the week, axis of movement, type of train, railway area, maximum wind speed, minimum horizontal visibility, minimum temperature, maximum temperature, number of frost reports on the ground surface, and precipitation. It rains and snows 24 hours a day. The proposed method for solving the existing problem based on CRISP-DM is a superior methodology in the field of implementing machine learning and data mining techniques in research and executive fields. Predictive modeling has been carried out in the form of classification. In order to predict the classification of the delay dependent variable, they are divided into two classes: on time and with delay. Supervised learning methods of the classification type have been used to predict the influence of weather factors on the occurrence of delays in railway areas. Cross-validation was used to evaluate the prediction results and check the validity of the model. The results show that the influence of weather factors in the winter season during the 5 years years in the year-by-year surveys had a positive, negative, or neutral effect on the occurrence of delays in railway areas. At the end of the article, preventive measures are presented to adapt the railway industry to climate threats in the future.

Keywords: Weather, Predicting, Delays, Passenger Trains, Machine Learning