Journal of Intelligent Strategic Management
Vol(4), 209-224
Bumara.ir

Predicting Flight Delays at Imam Khomeini International Airport: An XGBoost-Based Approach

Katayoun Kahrizi — Information Technology Management Department, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Abbas Asadi* — Marketing Management Department, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran.

Seyed Ahmad Shayannia — Industrial Management Department, Firouzkouh Branch, Islamic Azad University, Firouzkouh, Iran.

Abstract

Flight delays represent a significant challenge in the aviation industry, leading to passenger dissatisfaction, financial losses for airlines, and disruptions in airport operations. While numerous studies have investigated the factors influencing delays, further research is needed, particularly in utilizing advanced machine learning algorithms and examining the impact of various factors on delays at specific airports. In this study, we leverage the XGBoost tree-based learning algorithm and the R programming language to investigate the influence of weather conditions and flight-related factors on flight delays at Imam Khomeini International Airport (IKIA). The objective is to develop an accurate flight delay prediction model and identify the key factors contributing to delays. By analyzing historical flight data and meteorological information, the developed prediction model effectively identifies key factors influencing delays, achieving a 99.63% accuracy in delay detection. The results indicate that factors such as wind speed, precipitation, visibility, aircraft type, departure time, and flight route significantly impact the occurrence and duration of delays. These findings can facilitate informed decision-making in airport management, flight scheduling, and delay mitigation, ultimately leading to improved operational efficiency and enhanced passenger satisfaction.

Keywords: Flight Delay, XGBoost Algorithm, Weather Conditions, Imam Khomeini International Airport, Prediction Model

How to Cite: Kahrizi, K. , Asadi, A. and Shayannia, S. A. (2025). Predicting Flight Delays at Imam Khomeini International Airport: An XGBoost-Based Approach.Journal of Intelligent Strategic Management, 4(1), 209-224.

doi: bumara . 3.2.11235564.35886779

Intelligent Strategic Management (JISM) in Development and Evolution is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.

© Authors

* Corresponding Author: Abas.asadi@iau.ac.ir

Original Research
Received: 2024/12/25
Review: 2025/03/05
Accepted: 2025/04/19
eISSN: 2891-4128

نشریه مدیریت استراتژیک هوشمند
دوره ۴ شماره۱، بهار، ۱۴۰۴، ۲۲۴-۲۰۹
Bumara.ir

پیش‌بینی تأخیر در پروازها در فرودگاه بین‌المللی امام خمینی: یک رویکرد مبتنی بر XGBoost

کتایون کهریزی گروه مدیریت فناوری اطلاعات، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

عباس اسدی* گروه مدیریت بازاریابی، واحد ورامین-پیشوا، دانشگاه آزاد اسلامی، ورامین، ایران.

سید احمد شایان نیا گروه مدیریت صنعتی، واحد فیروزکوه، دانشگاه آزاد اسلامی، فیروزکوه، ایران.

مقاله پژوهشی
تاریخ دریافت:۱۴۰۳/۱۰/۰۵
بازنگری:۱۴۰۳/۱۲/۱۵
تاریخ پذیرش: ۱۴۰۴/۰۱/۳۰
eISSN:2891-4128

چکیده

تاخیر در پروازها یکی از چالش‌های اساسی در صنعت هوانوردی است که منجر به نارضایتی مسافران، خسارات مالی برای شرکت‌های هواپیمایی و اختلال در عملیات فرودگاهی می‌شود. اگرچه مطالعات بسیاری به بررسی عوامل موثر بر تاخیر پرداخته‌اند، اما همچنان نیاز به تحقیقات بیشتری در این زمینه، به ویژه در خصوص استفاده از الگوریتم‌های پیشرفته یادگیری ماشین و بررسی تأثیر عوامل مختلف بر تاخیرها در فرودگاه‌های خاص وجود دارد. در این پژوهش، با استفاده از الگوریتم یادگیری درختی XGBoost و زبان برنامه‌نویسی R، تأثیر شرایط آب و هوایی و عوامل مرتبط با پرواز بر تاخیر پروازها در فرودگاه امام خمینی مورد بررسی قرار گرفته است. هدف این پژوهش، توسعه یک مدل پیش‌بینی دقیق برای تاخیر پروازها و شناسایی عوامل کلیدی مؤثر بر آن‌ها است. با تحلیل داده‌های تاریخی پروازها و اطلاعات هواشناسی، مدل پیش‌بینی توسعه داده شده قادر به شناسایی عوامل کلیدی مؤثر بر تاخیرها بوده و به دقت ۹۹,۶۳ درصد در تشخیص تاخیر دست یافته است. نتایج این پژوهش نشان می‌دهد که عواملی همچون سرعت باد، بارش، دید، نوع هواپیما، زمان حرکت و مسیر پرواز، تأثیر قابل توجهی بر وقوع و مدت زمان تاخیرها دارند. این یافته‌ها می‌توانند به تصمیم‌گیری آگاهانه‌تر در مدیریت فرودگاه، برنامه‌ریزی پروازها و کاهش تأخیرها کمک کنند و در نهایت منجر به بهبود کارایی عملیاتی و افزایش رضایت مسافران شوند.

کلیدواژه‌ها: تاخیر پرواز، الگوریتم XGBoost، شرایط آب و هوایی، فرودگاه امام خمینی، مدل پیش بینی

استناد به این مقاله: کهریزی، کتایون و اسدی، عباس و شایان نیا، سید احمد . (۱۴۰۴). پیش‌بینی تأخیر در پروازها در فرودگاه بین‌المللی امام خمینی: یک رویکرد مبتنی بر .XGBoost. مدیریت استراتژیک هوشمند, ۴(۱), ۲۲۴-۲۰۹.

مدیریت استراتژیک هوشمند (JISM) در توسعه و تکامل تحت مجوز بین‌المللی کریتیو کامنز با شرایط انتساب-غیرتجاری ۴,۰ منتشر می‌شود.

©نویسندگان

* نویسنده مسئول: Abas.asadi@iau.ac.ir

مقدمه:

صنعت حمل‌ونقل هوایی یکی از بخش‌های حیاتی و پیچیده در سراسر جهان است که هر ساله میلیاردها نفر را جابجا می‌کند. با توجه به اهمیت زمانی در مسافرت‌های هوایی، تاخیر در پروازها یکی از چالش‌های عمده‌ای است که می‌تواند تجربه مسافران را تحت تأثیر قرار دهد و خسارات مالی قابل توجهی را برای شرکت‌های هواپیمایی به دنبال داشته باشد.

تاخیر در پروازها می‌تواند ناشی از عوامل متعددی از جمله شرایط جوی، مشکلات فنی، ازدحام ترافیک هوایی، و حتی عوامل انسانی باشد. در این میان، استفاده از روش‌های پیش‌بینی دقیق تاخیر پرواز می‌تواند به بهبود عملکرد سیستم‌های حمل‌ونقل هوایی کمک کند و از تأخیرات جلوگیری نماید. این تاخیرها نه تنها برای مسافران ناخوشایند هستند، بلکه باعث افزایش هزینه‌ها برای شرکت‌های هواپیمایی و اختلال در برنامه‌ریزی فرودگاه‌ها می‌شوند (IATA,۲۰۲۳). فرودگاه بین‌المللی امام خمینی به عنوان یکی از مهم‌ترین فرودگاه‌های ایران، از این قاعده مستثنی نیست و با توجه به حجم بالای ترافیک هوایی، با مشکل تاخیر در پروازها مواجه است (شرکت فرودگاه‌ها و ناوبری هوایی ایران, ۲۰۲۲). تاخیر در پروازها تبعات اقتصادی، اجتماعی و زیست‌محیطی زیادی دارد. از نظر اقتصادی، این تاخیرها باعث افزایش هزینه‌های سوخت، نیروی انسانی و تعمیر و نگهداری هواپیماها می‌شوند (Cook et al,۲۰۱۱). همچنین، مسافران به دلیل از دست دادن زمان و برنامه‌های خود، متحمل خسارات مالی و غیرمالی می‌شوند (IATA, ۲۰۲۳). از نظر اجتماعی، تاخیرها باعث نارضایتی مسافران، استرس و اضطراب آن‌ها می‌شود (EUROCONTROL , ۲۰۱۸).

از نظر زیست‌محیطی نیز، تاخیرها به دلیل افزایش زمان پرواز و مصرف سوخت بیشتر، باعث افزایش آلودگی هوا و انتشار گازهای گلخانه‌ای می‌شوند (ICAO , ۲۰۱۹). با وجود اهمیت پیش‌بینی تاخیر در پروازها، مطالعات اندکی در زمینه پیش‌بینی تاخیر در فرودگاه امام خمینی انجام شده است. اکثر مطالعات موجود در این زمینه به بررسی تاخیر در پروازها در فرودگاه‌های دیگر کشورها پرداخته‌اند و نتایج آن‌ها ممکن است به طور کامل قابل تعمیم به فرودگاه امام خمینی نباشد. علاوه بر این، اکثر مطالعات قبلی از روش‌های آماری سنتی برای پیش‌بینی تاخیر استفاده کرده‌اند که ممکن است در مواجهه با داده‌های پیچیده و حجیم امروزی کارایی لازم را نداشته باشند (Ahmed et al, ۲۰۲۱). در این پژوهش، با استفاده از الگوریتم‌های یادگیری ماشین پیشرفته، به پیش‌بینی تاخیر در پروازهای خروجی از فرودگاه امام خمینی پرداخته می‌شود. الگوریتم‌های یادگیری ماشین، به ویژه الگوریتم‌های یادگیری درختی مانند XGBoost، به دلیل قابلیت بالا در مدل‌سازی روابط پیچیده و غیرخطی بین متغیرها، برای پیش‌بینی تاخیر در پروازها مناسب هستند (Guestrin & Chen, ۲۰۱۶). استفاده از این الگوریتم‌ها در کنار داده‌های دقیق و جامع مربوط به پروازهای فرودگاه امام خمینی، می‌تواند به بهبود دقت پیش‌بینی تاخیر در پروازها و کاهش اثرات منفی آن کمک کند.

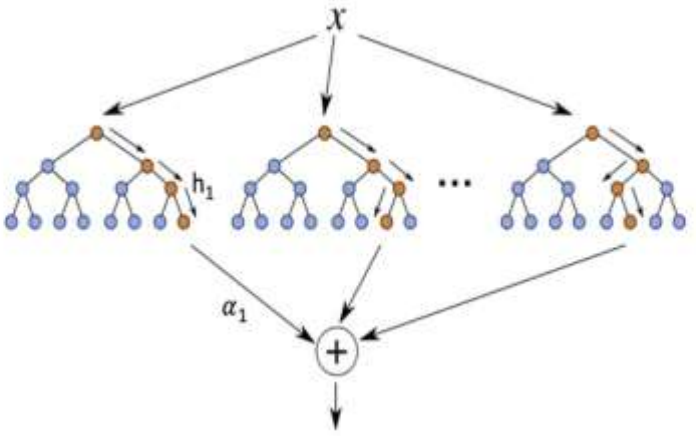
هدف اصلی این پژوهش، توسعه یک مدل پیش‌بینی دقیق برای تاخیر پروازها و شناسایی عوامل کلیدی مؤثر بر آن‌ها است. در بخش بعدی، به مرور ادبیات و پیشینه تحقیقات مرتبط پرداخته می‌شود. سپس، روش پژوهش و مدل پیشنهادی تشریح خواهد شد. در ادامه، نتایج حاصل از پیاده‌سازی مدل و ارزیابی آن ارائه می‌شود. در نهایت، نتیجه‌گیری و پیشنهادهایی برای تحقیقات آینده ارائه خواهد شد.

پیشینه نظری تحقیق:

در دهه‌های اخیر، توسعه الگوریتم‌های یادگیری ماشین و تکنیک‌های تحلیل داده‌های بزرگ (Big Data) امکان پیش‌بینی دقیق‌تر تاخیرات پروازی را فراهم کرده است. این الگوریتم‌ها با تحلیل داده‌های گذشته و بررسی روابط پیچیده بین عوامل مختلف، قادر به ارائه پیش‌بینی‌های قابل اعتمادی هستند که می‌توانند به بهبود کارایی و بهره‌وری در سیستم‌های حمل‌ونقل هوایی منجر شوند.

یکی از الگوریتم‌های قدرتمند و پرکاربرد در حوزه یادگیری ماشین برای پیش‌بینی تاخیرات پروازی، الگوریتم XGBoost. یک الگوریتم تقویتی (Boosting) است که با استفاده از تکنیک گرادیان تقویتی، مدل‌های یادگیری ماشین ضعیف‌تر را به یک مدل قوی‌تر تبدیل می‌کند. این الگوریتم به دلیل کارایی بالا، سرعت پردازش سریع، و دقت بالای پیش‌بینی، به طور گسترده‌ای در مسائل مختلف از جمله پیش‌بینی تاخیرات پروازی مورد استفاده قرار می‌گیرد.

الگوریتم XGBoost، که به معنای تقویت گرادیان شدید است، یک نسخه پیشرفته و تکامل‌یافته از الگوریتم‌های تقویت گرادیان محسوب می‌شود و با آنها در بسیاری از جنبه‌ها همخوانی دارد. تفاوت عمده XGBoost در به‌کارگیری فرآیند پردازش موازی در سطح گره‌ها است که به افزایش سرعت و کارایی آن منجر می‌شود. همچنین، XGBoost با استفاده از تکنیک‌های منظم‌سازی پیشرفته، خطر بیش‌برازش (overfitting) را کاهش داده و کیفیت مدل‌سازی را بهبود می‌بخشد (Chen & Guestrin, 2016). یک ویژگی برجسته دیگر این است که XGBoost قادر به مدیریت داده‌های ناقص است و می‌تواند در طول فرآیند آموزش، مکان مناسب برای داده‌های گمشده را تعیین کند. همانطور که در شکل ۱ نشان داده شده است، درختان متنوعی ایجاد می‌شوند که با تفاوت‌هایی همراه هستند و مدل به سمت پیش‌بینی دقیق‌ترین نتیجه پیش می‌رود.



شکل ۱. طرز کار الگوریتم XGBoost

با توجه به توانایی XGBoost در مدل‌سازی روابط پیچیده و مدیریت داده‌های ناقص، این الگوریتم به عنوان یک گزینه مناسب برای پیش‌بینی تأخیر پروازها مطرح می‌شود. در بخش پیشینه تحقیق، به بررسی کاربرد این الگوریتم و سایر روش‌های یادگیری ماشین در مطالعات پیشین می‌پردازیم.

پیشینه تحقیق:

مطالعه‌ی حاضر (Jing, & Li ۲۰۲۱) مدل‌هایی را برای پیش‌بینی تأخیر پرواز ارائه می‌دهد که اثرات محلی و شبکه‌ای را در فرودگاه‌ها بررسی می‌کند. با تحلیل جداگانه‌ی این اثرات، مشخص شد تأخیرهای بلندمدت عمدتاً ناشی از شبکه و تأخیرهای کوتاه‌مدت مرتبط با عوامل محلی هستند. در این راستا، عوامل جدیدی جایگزین عوامل سنتی شدند و عملکرد پیش‌بینی مدل‌ها در طبقه‌بندی و رگرسیون بهبود یافت. با استفاده از الگوریتم جنگل تصادفی بر روی داده‌های پروازهای داخلی آمریکا، دقت مدل طبقه‌بندی به ۹۶,۴۸٪ و برای مدل رگرسیون، ۹۳,۹۲٪ خطاها در بازه ۱۵ دقیقه بودند.

(Unterberger& Borsky , ۲۰۱۹) بیان می‌کند که شوک‌های آب و هوایی اغلب عملکرد روان سیستم‌های حمل و نقل را مختل می‌کند. آنها از مجموعه داده‌ای متشکل از ۲,۱۴ میلیون پرواز از ده فرودگاه بزرگ ایالات متحده بین ژانویه ۲۰۱۲ تا سپتامبر ۲۰۱۷ استفاده می‌کنند و اثرات رویدادهای شروع ناگهانی، یعنی بارش و باد و رویدادهای شروع آهسته، یعنی دما، را بر تأخیر پرواز تخمین می‌زنند. برای رویدادهای شروع ناگهانی، آنها یک چارچوب تفاوت در تفاوت را اعمال می‌کنند که امکان استنتاج در سطح ساعتی را فراهم می‌کند. اثرات رویدادهای شروع آهسته بر تأخیر خروج بر اساس برآوردگر پریس وینستین با خطاهای استاندارد تصحیح شده توسط پانل برآورد شده است. برآوردهای آنها افزایش قابل توجهی در تأخیر خروج تا ۲۳ دقیقه بسته به نوع آب و هوا و شدت اختلال نشان می‌دهد. با توجه به هزینه‌های اجتماعی تأخیر برنامه، این نتایج از اهمیت اقتصادی بالایی برخوردار است.

(Li & Jing, ۲۰۲۲) مدلی برای پیش‌بینی تأخیر پرواز با استفاده از شبکه حافظه کوتاه‌مدت پیش‌بینی تأخیر با مکانیسم توجه (LSTM-AM) برای یک فرودگاه منفرد ارائه کردند. هدف این مدل، پیش‌بینی دقیق تأخیر پرواز و تجزیه و تحلیل علل اولیه آن بود. نوآوری اصلی این مدل، استفاده از مکانیسم توجه در شبکه LSTM-AM برای تمرکز بر نقاط زمانی بحرانی و بهبود دقت پیش‌بینی است. آنها از داده‌های عملیاتی واقعی فرودگاه بین المللی پکن برای آموزش و ارزیابی مدل استفاده کردند. نتایج نشان داد که LSTM-AM دقت پیش‌بینی بهتری نسبت به الگوریتم‌های پایه دارد و میانگین خطای مطلق آن در مجموعه داده آزمایشی حدود ۸,۱۵ دقیقه است. با این حال، این مطالعه محدود به فرودگاه بین المللی پکن بود و ممکن است نتایج آن به سایر فرودگاه‌ها قابل تعمیم نباشد.

de Oliveira و همکاران (۲۰۲۱) یک چارچوب پیش‌بینی جدید به نام ST-Random Forest را برای پیش‌بینی تأخیر پرواز از منظر زمانی و مکانی پیشنهاد کردند. هدف آنها توسعه یک مدل پیش‌بینی بود که همبستگی زمانی و مکانی را در نظر بگیرد. نوآوری اصلی این مطالعه، استفاده از تئوری شبکه پیچیده برای استخراج ویژگی‌های فضایی و LSTM برای استخراج ویژگی‌های زمانی بود. آنها از داده‌های پروازهای داخلی چین بین ژوئن و اوت ۲۰۱۶ استفاده کردند و با اعمال تئوری شبکه پیچیده و LSTM، ویژگی‌های مرتبط را استخراج کرده و سپس از جنگل تصادفی برای پیش‌بینی تأخیر پرواز استفاده کردند. نتایج نشان داد که مدل ST-Random Forest به دقت ۹۲,۳۹٪ در پیش‌بینی تأخیر پرواز دست یافته است. با این حال، این مطالعه محدود به پروازهای داخلی چین بود و ممکن است نتایج آن به سایر مناطق قابل تعمیم نباشد.

(Li et al., ۲۰۲۳) بیان می‌کند که عملیات ترافیک هوایی به طور قابل توجهی تحت تأثیر شرایط آب و هوایی قرار دارد. این عوامل خارجی ممکن است محدودیت‌های عملیاتی را تحمیل کنند و عدم تعادل تقاضا-ظرفیت را ایجاد کنند که منجر به کاهش عملکرد به موقع، هزینه‌های اضافی خطوط هوایی و ناراحتی برای مسافران شود. مدیریت کارآمد چنین اختلالاتی مستلزم درک علل اصلی تأخیر پرواز به منظور افزایش قابلیت پیش‌بینی آنها است. در این مطالعه، تأثیرات شرایط آب و هوایی سطح فرودگاه را بر احتمال تأخیر پرواز برای سیستم حمل‌ونقل هوایی داخلی برزیل بررسی می‌کنند. آنها از برنامه تاریخی پرواز، عملکرد به‌موقع و داده‌های آب‌وهوا استفاده می‌کنند و یک مدل لاجیت را تخمین می‌زنند تا چگونگی تأثیر متغیرهای مختلف هواشناسی در فرودگاه مقصد را بر احتمال تأخیر رسیدن به مقصد بررسی کنند. شواهد تجربی برای تأثیرات شرایط دید، بارش و وزش باد بر احتمال تأخیر رسیدن برای مجموعه فرودگاه‌های برزیلی تحلیل شده به دست می‌آورند.

(Kaewunruen et al., ۲۰۲۱) پیش‌بینی تأخیر پرواز، به عنوان یک چالش اساسی در صنعت هوانوردی، پیامدهای قابل توجهی برای مسافران، شرکت‌های هواپیمایی و فرودگاه‌ها به همراه دارد. هدف تحقیقات پیشین، توسعه مدل‌های پیش‌بینی تأخیر پرواز با دقت بالا بوده است. نوآوری‌های صورت گرفته در این زمینه، شامل بهره‌گیری از الگوریتم‌های یادگیری ماشین و رویکردهای مبتنی بر شبکه به منظور بهبود دقت پیش‌بینی است. در

این راستا، از روش‌های متنوعی از جمله مدل‌های آماری، یادگیری ماشین و رویکردهای مبتنی بر شبکه استفاده شده است. داده‌های مورد استفاده در این پژوهش‌ها، شامل اطلاعات پرواز، شرایط آب و هوایی، ترافیک هوایی و عملیات فرودگاهی می‌شود. نتایج تحقیقات پیشین نشان می‌دهد که مدل‌های پیش‌بینی توسعه‌یافته، قادر به پیش‌بینی تأخیر پرواز با دقت قابل قبولی هستند. با این حال، محدودیت‌هایی همچنان چالش در پیش‌بینی دقیق تأخیر در شرایط آب و هوایی نامساعد و بازه‌های زمانی طولانی‌تر وجود دارد.

مطالعه‌ی وو و همکاران (۲۰۲۲) به بررسی تأثیر شرایط آب‌وهوایی شدید بر عملکرد فرودگاه‌ها و پروازها پرداخته است. هدف این پژوهش، شناسایی عوامل جوی مؤثر بر وقت‌شناسی پرواز و تعیین مقیاس پیامدهای ناشی از آن‌ها بود. با استفاده از یادگیری ماشین و داده‌های آب‌وهوایی و عملیاتی فرودگاه بیرمنگام، مدل‌هایی برای پیش‌بینی وقت‌شناسی پرواز توسعه داده شد. نتایج نشان داد که مدل جنگل تصادفی با R^2 برابر ۰٫۸۰ و RMSE کمتر از ۱۵٪، بهترین عملکرد را دارد. این یافته‌ها به مقامات فرودگاهی و صنعت بیمه در پیش‌بینی و مدیریت تأثیرات شرایط آب‌وهوایی کمک می‌کند.

مطالعه کیم و پارک (۲۰۲۴) با توجه به افزایش تقاضای سفرهای هوایی و تأثیر تغییرات اقلیمی بر تأخیر پروازها، به پیش‌بینی تأخیر خروج پروازها در بازه‌های زمانی طولانی پرداخته است. آنها از روش‌های داده‌محور و الگوریتم‌های یادگیری ماشین و یادگیری عمیق بر روی داده‌های سه فرودگاه ICN، JFK، MDW استفاده کردند. نتایج نشان داد که مدل‌های آنها در پیش‌بینی‌های ۲ ساعته و ۴۸ ساعته به دقت قابل قبولی دست یافته‌اند، به ویژه برای فرودگاه JFK. این پژوهش اهمیت پیش‌بینی تأخیر در بازه‌های زمانی بلندمدت را برجسته می‌کند و نشان می‌دهد که روش‌های یادگیری ماشین می‌توانند در این زمینه مفید باشند.

جدول ۱. خلاصه پیشینه تحقیق

ردیف	عنوان	سال	روش	نتیجه
۱	پیش‌بینی تأخیر پرواز با در نظر گرفتن اثرات محلی و اثرات شبکه	۲۰۲۱	رویکرد شبکه پیچیده، الگوریتم جنگل تصادفی	دقت بالای مدل در پیش‌بینی تأخیرهای بلندمدت و کوتاه‌مدت
۲	اثرات شوک‌های آب و هوایی بر تأخیر پرواز	۲۰۱۹	چارچوب تفاوت در تفاوت، برآوردگر پریس وینستین	افزایش قابل توجه در تأخیر خروج تا ۲۳ دقیقه بسته به نوع آب و هوا و شدت اختلال
۳	پیش‌بینی تأخیر پرواز با استفاده از شبکه حافظه کوتاه‌مدت پیش‌بینی تأخیر با مکانیسم توجه (LSTM-AM)	۲۰۲۲	شبکه LSTM-AM	دقت پیش‌بینی بهتر LSTM-AM نسبت به الگوریتم‌های پایه
۴	پیش‌بینی تأخیر پرواز از منظر زمانی و مکانی با استفاده از ST-Random Forest	۲۰۲۱	تئوری شبکه پیچیده، LSTM، جنگل تصادفی	دقت بالای مدل در پیش‌بینی تأخیر پرواز ۳۹.۹۲٪
۵	تأثیر شرایط آب و هوایی سطح فرودگاه بر احتمال تأخیر پرواز	۲۰۲۳	مدل لاجیت	شواهد تجربی برای تأثیرات شرایط دید، بارش و وزش باد بر احتمال تأخیر رسیدن
۶	پیش‌بینی تأخیر پرواز با در نظر گرفتن اطلاعات اولویت آب و هوا و ویژگی‌های غیر آب و هوایی	۲۰۲۱	جنگل تصادفی، الگوریتم خوشه‌بندی، روش نمونه‌گیری احتمالی	مدل پیشنهادی می‌تواند دقت پیش‌بینی تأخیر پرواز را به میزان قابل توجهی افزایش دهد
۷	پیش‌بینی وقت‌شناسی پرواز با استفاده از یادگیری ماشین در ارتباط با شرایط آب و هوایی	۲۰۲۲	یادگیری ماشین، جستجوی شبکه‌ای، اعتبارسنجی متقابل	مدل جنگل تصادفی می‌تواند روند نرخ‌های وقت‌شناسی پرواز را به خوبی درک کند
۸	پیش‌بینی تأخیرهای خروج پرواز با استفاده از روش‌های داده محور	۲۰۲۴	درخت تصمیم، جنگل تصادفی، ماشین بردار پشتیبان، K-نزدیکترین همسایه، رگرسیون لجستیک، افزایش گرادیان شدید، و حافظه کوتاه مدت طولانی	نرخ‌های دقت ۰٫۷۴۹ برای فرودگاه ICN، ۰٫۸۵۲ برای فرودگاه JFK و ۰٫۷۸۵ برای فرودگاه MDW در پیش‌بینی‌های ۲ ساعته

روش تحقیق:

پژوهش حاضر با هدف بررسی تأثیر شرایط آب و هوایی و عوامل پروازی بر تأخیرهای فرودگاه بین‌المللی امام خمینی (ره)، با بهره‌گیری از الگوریتم یادگیری ماشین XGBoost در محیط R انجام پذیرفته است مجموعه داده های عمومی پروازهای فرودگاه مهر آباد از مرکز داده ازمان هواپیمایی کشوری وسازمانهای مربوطه به دست می آید. این داده ها شامل ۸ دسته داده از جمله زمان حرکت، تاخیر خروج، زمان رسیدن، تاخیر ورود، حامل و شماره دم است. مجموعه داده های عمومی شرایط هواشناسی از خدمات ملی هواشناسی و آب و هوای زیرزمینی، شامل ۲۴ نوع داده مانند دما، نقطه شبنم، بارش، عمق برف و رطوبت به دست آمده است. این الگوریتم با استفاده از داده‌های مختلفی از جمله شرایط جوی، زمانبندی پروازها، و سایر عوامل مرتبط، به تحلیل و پیش‌بینی دقیق تاخیرات پروازی می‌پردازد.

معیارهای ارزیابی :

در این تحقیق، به منظور ارزیابی و مقایسه کارآیی مدل‌های مختلف یادگیری ما شین در پیش‌بینی و تشخیص تاخیر پروازها، از معیارهای معتبر به شرح موارد زیر استفاده شده است:

ماتریس درهم ریختگی (Confusion Matrix)

در دسته‌بندی با یادگیری ماشین، هدف دقت بالا در تشخیص دسته‌هاست. اما گاهی تشخیص درست یک دسته خاص اهمیت بیشتری دارد، مثلاً در تشخیص بیماری‌های خطرناک. در این موارد، تشخیص همه بیماران، حتی با وجود تشخیص اشتباه برخی افراد سالم، حیاتی است. ماتریس درهم‌ریختگی به ارزیابی این دقت کمک می‌کند و نمونه‌ها را به دو دسته مثبت (بیمار) و منفی (سالم) تقسیم می‌کند.

- مثبت صحیح[1] (TP): نمونه عضو دسته مثبت باشد و عضو همین کلاس تشخیص داده شود.
- منفی اشتباه[2] (FN): نمونه عضو کلاس مثبت باشد و عضو کلاس منفی تشخیص داده شود.
- منفی صحیح[3] (TN): نمونه عضو کلاس منفی باشد و عضو همین کلاس تشخیص داده شود.
- مثبت اشتباه[4] (FP): و در نهایت، نمونه عضو کلاس منفی باشد و عضو کلاس مثبت تشخیص داده شود.

پس از اجرای الگوریتم د سته‌بندی، با توجه به تو ضیحات و تعاریف ذکر شده، می‌توان عملکرد یک طبقه‌بند را به کمک جدولی مطابق با جدول زیر بررسی کرد.

[1] True Positive
[2] False Negative
[3] True Negative
[4] False Positive

جدول **۲.** ماتریس درهم ریختگی

		برچسب شناخته شده	
		مثبت	منفی
برچسب پیش‌بینی شده	مثبت	TP	FP
	منفی	FN	TN

این جدول را اصطلاحا ماتریس درهم ریختگی می گویند. جدول یا ماتریس درهم ریختگی، نتایج حاصل از طبقه‌بندی را بر اساس اطلاعات واقعی موجود، نمایش می‌دهد.

صحت

پارامتر صحت، متداول‌ترین، اساسی‌ترین و ساده‌ترین معیار اندازه‌گیری کیفیت یک دسته‌بند است و عبارت است از میزان تشخیص صحیح دسته‌بند در مجموع دو دسته. این پارامتر در واقع نشان‌گر میزان الگوهایی است که درست تشخیص داده شده‌اند و بر اساس ماتریس ارائه شده در بالا، به شکل رابطه زیر فرموله و تعریف می‌شود.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

لذا در زمینه تشخیص های پزشکی دو معیار حساسیت[5] و خصوصیت[6] نیز تعریف می شود که اولی معیاری از توانایی مدل طبقه‌بندی کننده در تشخیص درست بیماری و دومی معیاری از توانایی مدل طبقه‌بندی کننده در تشخیص درست سالم بودن می‌باشد.

حساسیت

پارامتر دیگری نیز علاوه بر معیار صحت وجود دارد که می‌توان به سادگی از ماتریس درهم ریختگی استخراج کرد. یکی از متداول‌ترین آن‌ها، معیار حساسیت (Sensitivity) است که آن را نرخ پاسخ‌های مثبت درست[7] نیز می گویند. حساسیت به معنی نسبتی از موارد مثبت است که آزمایش آن‌ها را به درستی به عنوان نمونه مثبت تشخیص داده است. این پارامتر به صورت رابطه زیر محاسبه می‌شود.

$$Sensitivity\ (TPR) = \frac{TP}{TP + FN}$$

[5] Sensitivity
[6] Specifiacity
[7] True Positive Rate (TPR)

دقت

پارامتر دقت[8] نشان می‌دهد که چه نسبتی از مثبت‌ها واقعا مثبت هستند یعنی چه نسبتی از داده‌هایی که دسته‌بند، مثبت تشخیص داده واقعا مثبت هستند. این پارامتر به صورت زیر تعریف می‌شود.

$$Precision = \frac{TP}{TP + FP}$$

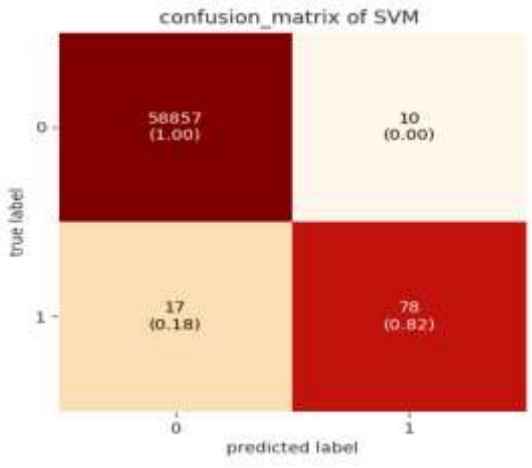
معیار F-Score

علاوه بر دو پارامتر حساسیت و خصوصیت، پارامترهای دیگری هم از ماتریس درهم‌ریختگی استخراج می شوند که هر یک بیان کننده مفهومی هستند و کاربردهای متفاوتی دارند. پارامتر دیگری به نام «معیار اف» (F-Score) وجود دارد که برای ارزیابی عملکرد دسته‌بندها بسیار مورد استفاده قرار می‌گیرد. این پارامتر از ترکیب دو پارامتر حساسیت (Recall) و پارامتر دقت (Precision)، که در واقع میانگین هارمونیک این دو پارامتر است، به صورت رابطه زیر تعریف می‌شود.

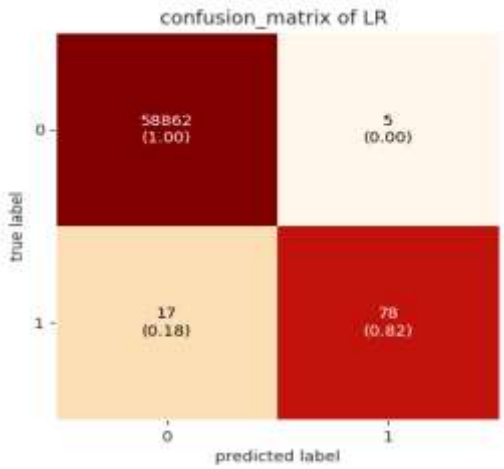
$$F - Score = 2 \times \frac{(\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})}$$

نتایج

ماتریس درهم ریختگی برای روش پیشنهادی و تکنیک های یادگیری ماشین برای مجموعه داده فرودگاه امام خمینی طبق شکل ۴-۲ است. در روش پیشنهادی، از ۵۸۸۶۷ رکورد مربوط به کلاس (بدون تاخیر)۱، تعداد ۵۸۸۶۵ رکورد درست و تعداد ۲ رکورد اشتباه تشخیص داده است. همچنین، از ۹۵ رکورد مربوط به کلاس ۰ (با تاخیر)، تعداد ۷۷ رکورد درست و تعداد ۱۸ رکورد اشتباه تشخیص داده است.

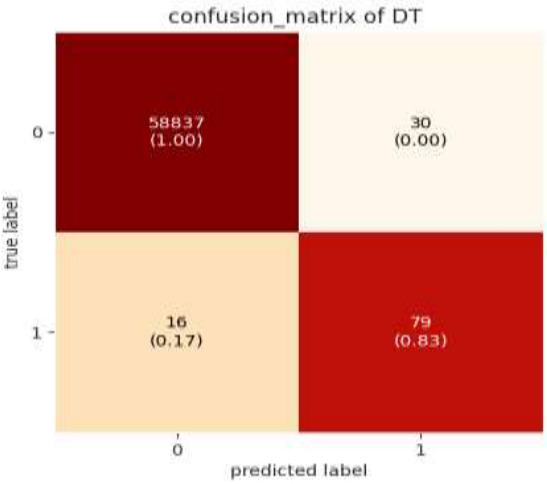


ب) ماشین بردار پشتیبان

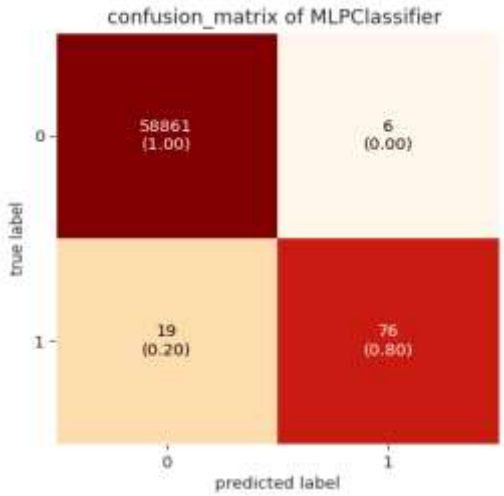


الف) رگرسیون لجستیک

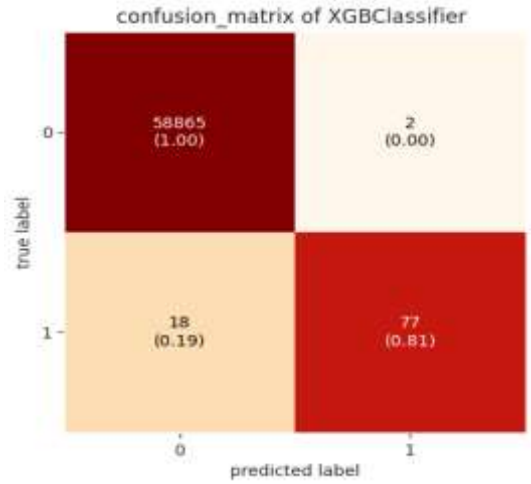
[8] Precision



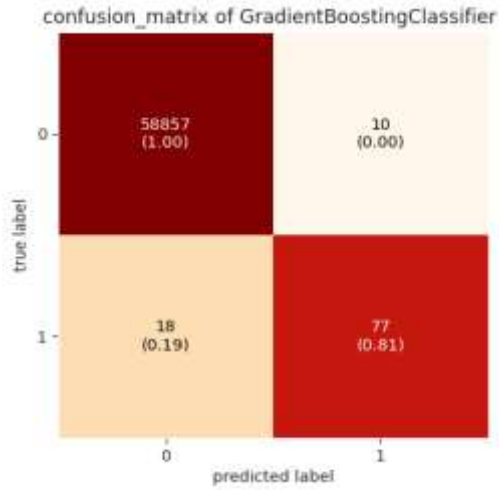
د)درخت تصمیم



چ) شبکه عصبی M LP



س)XGBoost (روش پیشنهادی)



ژ)GradientBoost

شکل ۳. ماتریس درهم ریختگی برای روش پیشنهادی و تکنیک های یادگیری ماشین برای داده فرودگاه امام خمینی

در ادامه، به منظور مقایسه عملکرد الگوریتم XGBoost با سایر مدل‌های یادگیری ماشین، نتایج ارزیابی آن‌ها بر اساس معیارهای مختلف در جدول زیر ارائه شده است.

جدول ۳. پارامتر ارزیابی روش پیشنهادی در مقایسه تکنیک های یادگیری ماشین روی مجموعه داده فرودگاه امام خمینی

	CatBoost	GradientBoost	XGBoost (روش پیشنهادی)	جنگل تصادفی	درخت تصمیم	پرستون چند لایه	ماشین بردار پشتیبان	رگرسیون لجستیک
دقت	۰.۹۹۹۶۹	۰.۹۹۹۵۳	۰.۹۹۹۷۱	۰.۹۹۹۶۶	۰.۹۹۹۲۲	۰.۹۹۹۵۸	۰.۹۹۹۵۴	۰.۹۹۹۶۳
حساسیت	۰.۹۸	۰.۹۴	۰.۹۹	۰.۹۹	۰.۸۶	۰.۹۶	۰.۹۴	۰.۹۷
دقت	۰.۹۸	۰.۹۴	۰.۹۹	۰.۹۷	۰.۸۶	۰.۹۶	۰.۹۴	۰.۹۷
بازخوانی	۰.۹۲	۰,۹۱	۰.۹۳	۰.۹۱	۰.۹۲	۰.۹۰	۱.۰	۰.۹۱
معیار F Score۱	۰.۹۵	۰.۹۲	۰.۹۵	۰.۹۴	۰.۸۹	۰.۹۳	۰.۹۳	۰.۹۴

بحث و نتیجه گیری

پژوهش حاضر با بهره‌گیری از الگوریتم یادگیری ماشین XGBoost، به بررسی تأثیر شرایط آب و هوایی و عوامل پروازی بر تأخیرهای پرواز در فرودگاه بین‌المللی امام خمینی (ره) پرداخت. نتایج حاصل از این پژوهش، نشانگر توانایی قابل‌توجه این الگوریتم در پیش‌بینی دقیق احتمال و مدت تأخیرهای پروازی است. این یافته‌ها با مطالعات پیشین که به بررسی عوامل مؤثر بر تأخیرهای پروازی پرداخته‌اند، همسو بوده و بر اهمیت شرایط آب و هوایی و عوامل پروازی در این زمینه تأکید می‌نماید. با این حال، پژوهش حاضر با به‌کارگیری الگوریتم XGBoost و تحلیل جامع داده‌های پروازی و هواشناسی، توانسته است مدلی با دقت پیش‌بینی بالا (۹۹,۶۳٪) ارائه دهد که نسبت به مدل‌های پیشین، از عملکرد بهتری برخوردار است.

یکی از نوآوری‌های این پژوهش، بررسی هم‌زمان تأثیر شرایط آب و هوایی و عوامل پروازی بر تأخیرهای پروازی است. در مطالعات پیشین، اغلب این عوامل به‌صورت جداگانه مورد بررسی قرار می‌گرفتند. اما در این پژوهش، با تلفیق این دو دسته از عوامل، مدلی جامع‌تر و واقع‌گرایانه‌تر برای پیش‌بینی تأخیرها ارائه شده است.

یافته‌های این پژوهش می‌تواند برای ذینفعان صنعت هوانوردی، از جمله مدیران فرودگاه‌ها، خطوط هوایی و مسافران، مفید واقع گردد. با استفاده از مدل پیش‌بینی توسعه‌یافته در این پژوهش، می‌توان تصمیمات آگاهانه‌تری در زمینه برنامه‌ریزی پروازها، تخصیص منابع و مدیریت عملیات فرودگاهی اتخاذ نمود. این امر می‌تواند به بهبود کارایی عملیاتی، کاهش تأخیرها و افزایش رضایت مسافران منجر شود.

در نهایت، این پژوهش با ارائه یک مدل پیش‌بینی دقیق و جامع، گامی مؤثر در جهت پیشبرد تصمیم‌گیری مبتنی بر داده در صنعت هوانوردی برداشته است. با این حال، برای تعمیم‌پذیری بیشتر نتایج و افزایش کارایی مدل، پیشنهاد می‌شود در پژوهش‌های آتی، عوامل دیگری نظیر بی‌ثباتی سیاسی، ازدحام فضای هوایی و نوسانات اقتصادی نیز مورد بررسی قرار گیرند. همچنین، به‌منظور تسهیل تفسیرپذیری مدل برای ذینفعان غیرمتخصص، می‌توان از روش‌های ساده‌سازی مدل و ارائه توضیحات شفاف‌تر بهره برد.

مراجع:

صنیعی آباده, م., محمودی, س., & طاهرپرور, م. (۱۳۹۶). داده‌کاوی کاربردی. نیاز دانش.

اسماعیلی, م. (۱۳۹۸). آموزش گام به گام داده کاوی با RapidMiner. آتی نگر.

غضنفری, م., علیزاده, س., & تیمورپور, ب. (۱۳۹۸). داده کاوی و کشف دانش. دانشگاه علم و صنعت ایران.

Berry, M. J., & Linoff, G. S. (2009). Data mining techniques. John Wiley & Sons.

Borsky, S., & Unterberger, C. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events. Economics of transportation, 18, 10-26.

de Oliveira, M., Eufrásio, A. B. R., Guterres, M. X., Murça, M. C. R., & de Arantes Gomes, R. (2021). Analysis of airport weather impact on on-time performance of arrival flights for the Brazilian domestic air transportation system. Journal of Air Transport Management, 91, 101974.

Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.

https://en.wikipedia.org/wiki/Decision_tree.

Kaewunruen, S., Sresakoolchai, J., & Xiang, Y. (2021). Identification of weather influences on flight punctuality using machine learning approach. Climate, 9(8), 127.

Kim, S., & Park, E. (2024). Prediction of flight departure delays caused by weather conditions adopting data-driven approaches. Journal of Big Data, 11(1), 11.

Kleinberg, E. M. (1990). Stochastic discrimination. Annals of Mathematics and Artificial intelligence, 1, 207-239.

Kotsiantis, S. (2011). Combining bagging, boosting, rotation forest and random subspace methods. Artificial Intelligence Review, 35, 223-240.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.

Li, K., & Jing, Y. (2021). Airport individual flight delay prediction model considering local effects and network effects. Journal of Air Transport Management, 95, 102088.

Li, Q., & Jing, R. (2021). Generation and prediction of flight delays in air transport. IET Intelligent Transport Systems, 15(6), 740-753.

Li, Q., & Jing, R. (2022). Flight delay prediction from spatial and temporal perspective. Expert Systems with Applications, 205, 117662.

Li, Q., Jing, R., & Dong, Z. S. (2023). Flight delay prediction with priority information of weather and non-weather features. IEEE Transactions on Intelligent Transportation Systems.

Mokhtarimousavi, S., & Mehrabi, A. (2023). Flight delay causality: Machine learning technique in conjunction with random parameter statistical analysis. International Journal of Transportation Science and Technology, 12(1), 230-244.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6), 275-285.

Qi, Y. (2012). Random forest for bioinformatics. Ensemble machine learning: Methods and applications, 307-323.

Song, Y.-Y., & Ying, L. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

Wu, Y., Mei, G., & Shao, K. (2022). Revealing influence of meteorological conditions and flight factors on delays using XGBoost. Journal of Computational Mathematics and Data Science, 3, 100030.

Yan, A. (2020). A Survey on Tree-Based Machine Learning Methods: The Path to XGBoost and Generalized Random Forests.