



مقایسه تطبیقی رگرسیون‌های خطی چندگانه و جنگل تصادفی در تخمین متوسط دمای سطح زمین: مطالعه موردی شهر تبریز

محمدعلی کوشش وطن*، اکبر اصغری زمانی، شهرپور روستایی

- ۱- محمدعلی کوشش وطن: دانشجوی دکتری جغرافیا و برنامه ریزی شهری. گروه، جغرافیا و برنامه ریزی شهری، دانشکده برنامه ریزی و علوم محیطی، دانشگاه تبریز، شهر تبریز، کشور ایران. ایمیل m.alikousheshvatan@gmail.com
- ۲- اکبر اصغری زمانی: هیات علمی گروه جغرافیا و برنامه ریزی شهری، دانشکده برنامه ریزی و علوم محیطی، دانشگاه تبریز، شهر تبریز، کشور ایران. ایمیل azamani621@gmail.com
- ۳- شهرپور روستایی: هیات علمی گروه جغرافیا و برنامه ریزی شهری، دانشکده برنامه ریزی و علوم محیطی، دانشگاه تبریز، شهر تبریز، کشور ایران. ایمیل sroustaie@gmail.com

چکیده

دمای سطح زمین به عنوان یکی از پارامترهای مهم و پایه‌ای در مباحث اقلیمی، نشان‌دهنده رابطه بین اتمسفر و زمین می‌باشد. با در نظر داشتن مشکلات زیست‌محیطی شهرها از جمله شدت یافتن جزایر حرارتی شهری، تخمین دمای سطح زمین با دقت مطلوب و همچنین استخراج عوامل مؤثر بر آن نقش قابل توجهی را در مدیریت حرارتی محیط‌های شهری و همچنین اتخاذ استراتژی‌های انطباقی با جزایر حرارتی دارد. پژوهش حاضر در این راستا به مقایسه دو روش رگرسیون خطی چندگانه و جنگل تصادفی پرداخته است. از تصاویر مادیس روزانه در بازه شب (۱۰:۳۰) جهت استخراج دمای سطح زمین فصل تابستان شهر تبریز بهره گرفته شد. تصاویر مذکور در سامانه گوگل ارث انجین پردازش شده و برای بازه ۲۰۱۸ الی ۲۰۲۲ میانگین‌گیری گردید. بر اساس نتایج پژوهش، جنگل تصادفی با ضریب تعیین $0/924$ ($0/009$) = RMS عملکرد بسیار بهتری را نسبت به رگرسیون خطی چندگانه از خود نشان داد. جهت استخراج اهمیت شاخص‌ها نیز از جنگل تصادفی بهره گرفته شد. بر اساس نتایج اهمیت شاخص‌ها، شاخص‌های نور شب ($51/06$ درصد)، ضریب دید به آسمان ($48/01$ درصد) و مساحت ناحیه رو به باد ($45/27$) به ترتیب مهم‌ترین شاخص‌های اثرگذار بر متوسط دمای سطح زمین شبانه فصل تابستان شهر تبریز هستند. نتایج پژوهش حاضر علاوه بر آشکار نمودن قوت رگرسیون جنگل تصادفی در تخمین دمای سطح زمین، اهمیت شاخص‌های متعدد بر آن را نیز آشکار می‌کند. در این راستا، یافته‌های این مطالعه برای مدیریت حرارتی محیط شهری تبریز و اتخاذ استراتژی‌های سازگاری با جزایر حرارتی آن کاربردی خواهد بود.

کلمات کلیدی: دمای سطح زمین، رگرسیون جنگل تصادفی، رگرسیون خطی چندگانه، یادگیری ماشینی، تبریز.

۱- مقدمه

دمای سطح زمین یک پارامتر مهم محیطی است که به طور قابل توجهی بر مدل سازی اقلیمی، پایش کشاورزی و برنامه ریزی شهری تأثیرگذار است. از کاربردهای قابل توجه داده های دمای سطح زمین در مطالعه اثر جزیره گرمایی شهری است؛ جایی که مناطق شهری به دلیل فعالیت های انسانی و تغییر سطوح زمین دمای بالاتری نسبت به مناطق روستایی اطراف خود دارند (Voogt & Oke, 2003). این پدیده پیامدهای جدی برای پایداری شهری، بهداشت عمومی و مصرف انرژی دارد (Li et al., 2013; Sobrino et al., 2016). دماهای بالا در مناطق شهری می تواند استرس گرمایی را در میان ساکنان تشدید کرده، تقاضای انرژی برای خنک سازی را افزایش داده و کیفیت هوا را در سطح قابل توجهی کاهش می یابد (Li et al., 2013; Zhan et al., 2020)؛ بنابراین، برآورد دقیق دمای سطح زمین برای مدیریت حرارتی مؤثر شهری و استراتژی های کاهش اثرات جزایر گرمای شهری بسیار مهم است.

روش های سنتی برای برآورد دمای سطح زمین به شکل عمده به داده های سنجش از دور و الگوریتم های تجربی متکی است که اغلب به دلیل تداخل جوی، نویز حسگر و ماهیت ناهمگن سطح زمین با محدودیت هایی مواجه می گردد (Li et al., 2013; Sobrino et al., 2016). در سال های اخیر، تکنیک های یادگیری ماشینی به عنوان ابزارهای قدرتمند برای افزایش دقت و پایداری تخمین دمای سطح زمین استفاده شده اند. در میان این تکنیک ها، الگوریتم جنگل تصادفی به دلیل توانایی آن در مدیریت مجموعه داده های بزرگ، مقادیر ناموجود و مدل سازی روابط پیچیده و غیرخطی بین ویژگی های ورودی مورد توجه قرار گرفته است (Breiman, 2011). این الگوریتم یک روش یادگیری تجمعی مبتنی بر درخت های تصمیم می باشد که پیش بینی های چندین درخت را ترکیب کرده و عملکرد پیش بینی را بهبود می دهد (Rodriguez-Galiano et al., 2012).

مطالعات انجام شده بر عملکرد مطلوب الگوریتم جنگل تصادفی در کاربردهای مختلف سنجش از دور، از جمله طبقه بندی پوشش زمین، برآورد رطوبت خاک و نقشه برداری پوشش گیاهی اشاره کرده اند (Belgiu & Drăgut, 2016; Chan & Paelinckx, 2008). در زمینه برآورد دمای سطح زمین نیز این روش عملکرد بهتری نسبت به روش های سنتی مانند رگرسیون خطی چندگانه را از خود نشان داده است. به عنوان نمونه،

¹ Random Forest (RF)

فان^۱ و همکاران (۲۰۱۸) از جنگل تصادفی جهت برآورد دمای سطح زمین حاصل از لندست ۸ پرداخته‌اند. نتایج ایشان نشان می‌دهد که رگرسیون جنگل تصادفی عملکرد بهتری را نیست به رگرسیون خطی چندگانه دارد. منگ^۲ و همکاران (۲۰۲۰) نیز دریافتند که رگرسیون جنگل تصادفی نیست به رگرسیون خطی چندگانه برآورد دقیق‌تری را دمای سطح زمین حاصل از ماهواره مادیس ارائه می‌دهد. علاوه بر این، حقوردی و همکاران (۲۰۲۱) در مقایسه این دو روش در برآورد دمای سطح زمین با استفاده از داده‌های مادیس دریافتند که رگرسیون جنگل تصادفی به طور قابل توجهی دارای دقت و پایداری بالایی است. این مطالعه توانایی جنگل تصادفی در کاهش بیش‌برازش و همچنین اثربخشی آن در مدیریت داده‌های با حجم بالا که از چالش‌های رایج در برآورد دمای سطح زمین هستند را برجسته کرده است. این مطالعات همچنین نیاز به بررسی بیشتر ویژگی‌های ورودی بهینه، تنظیمات پارامترها و عملکرد الگوریتم در مناطق مختلف را اشاره کرده‌اند. با این حال، هنوز یک شکاف در ارزیابی سیستماتیک قابلیت‌ها و محدودیت‌های رگرسیون جنگل تصادفی در برآورد دمای سطح زمین به ویژه در مناطق شهری تحت تأثیر اثر جزایر حرارتی وجود دارد. علاوه بر شاخص‌های معمول مانند شاخص‌های طیفی و داده‌های هواشناسی، متغیرهای مورفولوژی شهر همچون ارتفاع، تراکم و حجم ساختمان به عنوان تعیین‌کننده‌های مهم دمای سطح زمین در مطالعات متعدد استفاده شده‌اند. ویژگی‌های مورفولوژیکی بر خواص حرارتی سطوح شهری و ریزاقلمیم تأثیر آن تأثیر قابل توجهی دارند و در نتیجه بر توزیع دمای سطح زمین تأثیر می‌گذارند. ادغام این متغیرها در مدل‌های پیش‌بینی کننده، می‌تواند وضوح فضایی و دقت برآورد دما را افزایش دهد. خروجی این عمل نیز در شکل‌دهی بینش‌های ضروری برای برنامه‌ریزی شهری و استراتژی‌های کاهش جزایر حرارتی بسیار مفید خواهد بود (Zhan et al., 2020; Zhang & Yu, 2023).

باتوجه به موارد مطرح شده، هدف پژوهش حاضر مقایسه دو روش رگرسیونی خطی چندگانه و جنگل تصادفی در پیش‌بینی دمای سطح زمین شبانه فصل تابستان و همچنین استخراج اهمیت مؤلفه‌های مختلف در راستای تخمین آن است. مؤلفه‌های مذکور شامل شاخص‌های مساحت نفوذناپذیری سطح، دید به آسمان، شاخص مساحت رو به باد، شاخص تفاضلی پوشش گیاهی، آلبدو و نور شب می‌باشد.

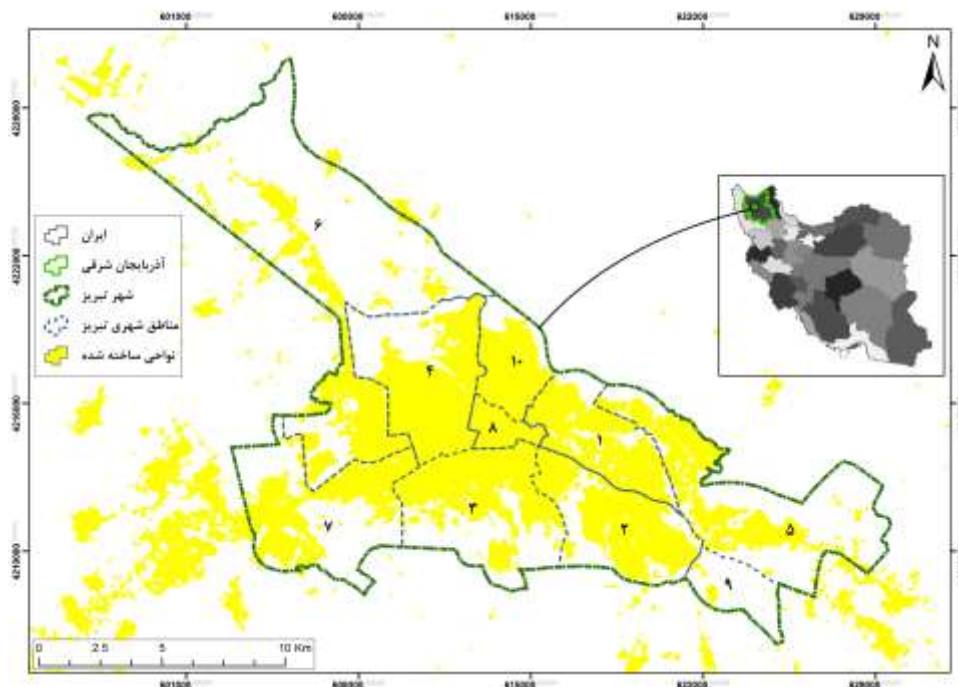
۲- مواد و روش تحقیق

۲-۱- منطقه مورد مطالعه

¹ Phan

² Meng

تبریز (شرقی ۱۷' ۴۶°، شمالی ۰۵' ۳۸°) یکی از شهرهای مهم شمال غرب کشور در استان آذربایجان شرقی است. بر اساس آخرین سرشماری این شهر دارای جمعیتی حدود دو میلیون نفر است. متوسط ارتفاع شهر تبریز ۱۳۵۰ متر بالاتر از آب‌های آزاد بوده که توسط کوه‌های عینالی و سه‌پند از شمال و جنوب محدود شده است (کوشش وطن و اصغری زمانی، ۱۴۰۰). این شهر به لحاظ طبقه‌بندی اقلیمی کوپن در دسته اقلیم سرد نیمه خشک واقع شده است؛ بنابراین، تغییرات فصلی قابل توجهی را دارا می‌باشد. فصل تابستان آن اغلب گرم و خشک و زمستان آن سرد و همراه با بارش برف است. متوسط دمای هوای شهر تبریز بر اساس دوره آماری ۱۹۷۰ الی ۲۰۲۲ برابر با ۱۲/۷ درجه سانتی‌گراد است. شکل (۱) نقشه موقعیت محدوده مورد مطالعه را نشان می‌دهد.



شکل (۱): موقعیت جغرافیایی شهر تبریز
Fig. (1): The geographical location of Tabriz City

۲-۲- داده‌های مورد استفاده

جهت استخراج دمای سطح زمین شهر تبریز از پروداکت شبانه MODIS Terra به صورت روزانه استفاده شد. تصاویر موردنظر برای فصل تابستان در بازه ۲۰۱۸ الی ۲۰۲۲ در سامانه گوگل ارث انجین مورد پردازش قرار

گرفته و دانلود گردید. کیفیت تصاویر نیز با استفاده از باند QA کنترل شد. جهت آماده‌سازی شاخص‌های آلبدو و شاخص تفاضلی پوشش گیاهی از تصاویر لندست ۸ به صورت تصحیح شده بهره گرفته شد. شاخص‌های موردنظر برای بازه ۲۰۱۸ الی ۲۰۲۲ در فصل تابستان در سامانه گوگل ارث انجین پردازش شده و دانلود گردید. شاخص مساحت سطح نفوذناپذیر^۱ نیز با استفاده از تصویر سنتینل ۲ در سامانه مذکور تهیه شد. همچنین، شاخص‌های دید به آسمان^۲ و مساحت ناحیه رو به باد^۳ از مدل رقمی ارتفاعی سطح^۴ تبریته تهیه شد. شاخص نور شب نیز از تصاویر VIIRS stray light corrected night-time برای بازه مورد پژوهش به صورت میانگین فصل تابستان از سامانه ارث انجین تهیه شد. جدول (۱) شاخص‌های پژوهش را نشان می‌دهد. این شاخص‌ها از جمله شاخص‌های مهم در راستای تخمین دمای سطح زمین و همچنین جزایر حرارتی هستند که در پژوهش‌های متعدد مورد استفاده قرار گرفته است.

جدول شماره (۱): متغیرهای پژوهش جهت تخمین متوسط دمای سطح زمین

| منبع | شاخص |
|---|------------------------|
| Pan et al., 2023; Zhang et al., 2021 | مساحت سطح نفوذناپذیر |
| Chen et al., 2021; Bernard et al., 2018; Hodul et al., 2016 | ضریب دید به آسمان |
| Gong et al., 2023; Keerthi Naidu and Chundeli, 2023; Li et al., 2022 | مساحت ناحیه رو به باد |
| Gadekar et al., 2023; Guha and Ghovil 2020; Parastatidis et al., 2017 | شاخص تفاضلی پوشش گیاهی |
| Mansourmoghaddam et al., 2024; Andrés-Anaya, 2021 | آلبدو |
| Li et al., 2022; Mpakairi & Muvengwi, 2019 | نور شب |

۲-۲- روش اجرای پژوهش

در پژوهش حاضر، جهت مدل‌سازی دمای سطح زمین شبانه، متغیرهای متعددی استفاده شد. این شاخص‌ها عبارت‌اند از: مساحت سطح نفوذناپذیر، ضریب دید به آسمان، شاخص مساحت ناحیه رو به باد، شاخص تفاضلی پوشش گیاهی، آلبدو و نور شب. شاخص مساحت سطح نفوذناپذیر با استفاده از تصویر سنتینل ۲ تهیه گردید. جهت استخراج آن از الگوریتم طبقه‌بندی جنگل تصادفی استفاده شد. اراضی ساخته شده به‌عنوان سطوح نفوذناپذیر در نظر گرفته شد. تخمین شاخص ضریب دید به آسمان نیز با استفاده از مدل رقمی ارتفاعی سطح

¹ Normalized difference vegetation index (NDVI)
² Impervious surface area (ISA)
³ Sky view factor (SVF)
⁴ Frontal area index (FAI)
⁵ Digital Surface Model (DSM)

شهر تبریز با استفاده از نرم‌افزار SAGA تهیه گردید. در ادامه، شاخص مساحت ناحیه رو به باد با استفاده از DSM برآورد گردید. رابطه (۱) فرمول محاسبه این شاخص را نشان می‌دهد. این شاخص λ_f از نسبت مجموع مساحت وجوه ساختمان رو به جهت باد (A_{facets}) با مساحت حجمی ساختمان (A_{plane}) حاصل می‌گردد (Wong et al., 2010).

$$\lambda_f = A_{facets}/A_{plane} \quad (1)$$

شاخص NDVI نیز با استفاده از رابطه (۲) برآورد گردید. بدین جهت از باندهای ۴ (RED) و ۵ (NIR) ماهواره لندست ۸ استفاده شد (Carlson & Riziley, 1997).

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (2)$$

جهت تخمین آلودگی در بازه مورد مطالعه نیز از رابطه (۳) استفاده شد که توسط Liang (2000) ارائه شده است. در این رابطه ρ نشانگر باندهای لندست می‌باشد.

$$a_{TOA} = \frac{0.356\rho_1 + 0.13\rho_3 + 0.373\rho_4 + 0.085\rho_5 + 0.072\rho_7 - 0.0018}{0.356 + 0.13 + 0.373 + 0.085 + 0.072} \quad (3)$$

تصاویر نور شب نیز در سامانه ارت انجین برای بازه مورد مطالعه تهیه گردید. این شاخص به‌عنوان مولفه‌ای از فعالیت‌های انسانی در نواحی شهری توسط پژوهش‌های متعددی استفاده شده است (Mpakairi & Muvengqi, 2019). در نهایت تمام متغیرهای تهیه شده، در قالب شبکه ۳۰۰*۳۰۰ میانگین‌گیری شد که نشانگر مقیاس واحد همسایگی می‌باشد. لازم به ذکر است همه تصاویر تولید شده در یک سیستم تصویر (UTM) قرار داده شده و به لحاظ تفکیک مکانی نیز به دقت مکانی ۱۰۰ متر ریسپل شدند. بدین جهت از الگوریتم Bilinear Interpolation شد. در نهایت جهت سنجش میزان هم‌خطی چندگانه شاخص‌ها از عامل تورم واریانس استفاده شد. در پژوهش حاضر، آستانه این عامل کمتر از ضریب ۱۰ در نظر گرفته شد. در این

¹ Variance Inflation Factor (VIF)

راستا، هم خطی چندگانه برای هیچ یک از متغیرها مشاهده نشد ($VIF < 10$). واحد همه شاخص‌های مورد استفاده در پژوهش نیز با استفاده از لگاریتم طبیعی یکسان‌سازی گردید.

۳-۲-۱- رگرسیون خطی چندگانه

رگرسیون خطی چندگانه، یکی از روش‌های آماری است که به مدل‌سازی رابطه بین متغیرهای مستقل و وابسته بر مبنای معادله خطی در داده‌های مشاهده‌شده می‌پردازد (Abrougui et al., 2019). این رگرسیون مدل گسترش‌یافته رگرسیون خطی ساده بوده و تأثیر دو یا چند متغیر بر یک متغیر وابسته را نشان می‌دهد (Abbot, 2020; George & Mallery, 2017). رابطه (۴) معادله عمومی رگرسیون خطی چندگانه به شرح زیر است:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon_i \quad (4)$$

در این معادله Y_i معرف میزان پیش‌بینی شده برای متغیر وابسته، X معرف متغیرهای پیش‌بین، β_0 معرف عرض از مبدأ، β معرف شیب خطوط و ε_i معرف خطای مدل است (Richardson, 2015; Pal & Bharati, 2019).

۳-۲-۲- رگرسیون جنگل تصادفی

رگرسیون جنگل تصادفی یکی از قوی‌ترین الگوریتم‌های یادگیری ماشینی است که توسط بریمن در سال ۲۰۰۱ ارائه شده است (کوشش وطن و همکاران، ۱۴۰۰). این روش از نوع نظارت‌شده است که برای مسائل رگرسیون و طبقه‌بندی بر اساس توسعه درخت تصمیم ارائه شده است (Makhija et al., 2021; Sharma & Rani, 2021). این روش با استفاده از یک درخت تصمیم ضعیف، یک یادگیرنده قوی‌تر تشکیل می‌دهد. همچنین، رگرسیون جنگل تصادفی، توانایی مناسبی را در راستایی کنترل داده‌های از دست‌رفته و نامتوازن از خود نشان می‌دهد. علاوه بر این، آن‌ها در مقایسه با سایر الگوریتم‌های یادگیری ماشینی، کمتر مستعد بیش‌برازش هستند (Wlodarczak, 2020).

¹ Breiman

² Overfitting

در پژوهش حاضر از ضریب تعیین^۱ (۵)، خطای جذر میانگین مربعات^۲ (۶) برای ارزیابی عملکرد دو مدل رگرسیون خطی چندگانه و رگرسیون جنگل تصادفی بهره گرفته شد. در فرمول‌های زیر N بیانگر تعداد مشاهدات، A_i ، \hat{P}_i و \bar{A} به ترتیب بیانگر اعداد مشاهده شده، اعداد پیش‌بینی شده و میانگین اعداد مشاهده شده است (کوشش وطن و همکاران، ۱۴۰۰).

$$R^2 \left(1 - \frac{\sum_{i=1}^N (A_i - \hat{P}_i)^2}{\sum_{i=1}^N (A_i - \bar{A})^2} \right) \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - \hat{P}_i)^2} \quad (6)$$

جهت انجام مدل‌سازی و همچنین استخراج عوامل مؤثر در دمای سطح زمین شبانه فصل تابستان، داده‌های پژوهش به دودسته آزمایش (۳۰ درصد) و آموزش (۷۰ درصد) تقسیم شد. جهت اجرای رگرسیون جنگل تصادفی نیاز است تا دو پارامتر $mtry$ و $ntree$ با مقدار مطلوب بهینه گردد. $ntree$ بیانگر فراوانی درخت در جنگل تصادفی بوده و $mtry$ نشانگر تعداد متغیرهای موجود برای تقسیم در هر گره درخت است. در این الگوریتم، هرچه تعداد درختان تا حدی بیشتر باشد، مدل عملکرد بهتری خواهد داشت (Breiman, 2002; Liaw & Wiener, 2001). در پژوهش حاضر، تعداد درختان در عدد ۱۵۰۰ بهینه شده و تعداد متغیرهای موجود برای تقسیم در هر گره نیز ۲ عدد در نظر گرفته شد. هر دو روش رگرسیونی در نرم‌افزار RStudio اجرا شد. همچنین، جهت اجرای رگرسیون جنگل تصادفی نیز از بسته `randomForest` بهره گرفته شد.

۳- نتایج و بحث

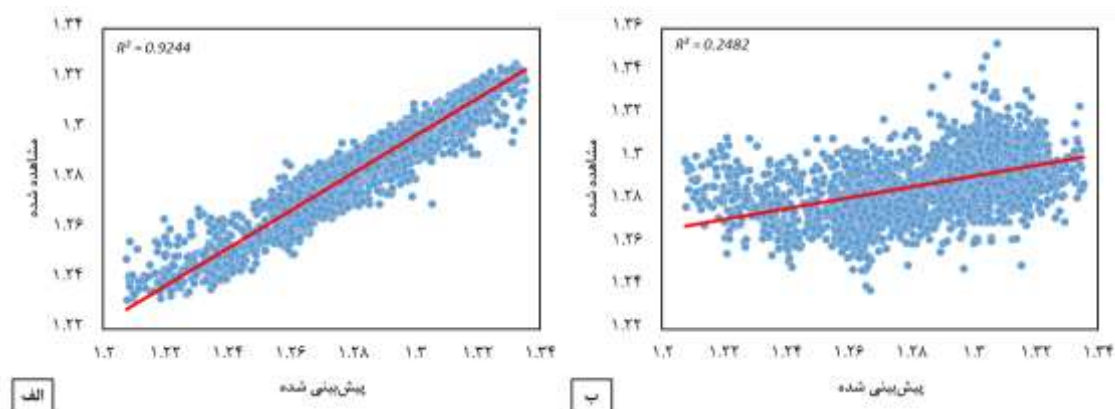
۳-۱- مقایسه دقت الگوریتم جنگل تصادفی و رگرسیون خطی چندگانه

شکل (۲) نتایج حاصل از آموزش رگرسیون‌های مورد استفاده در پژوهش را نشان می‌دهد. در این راستا، ضریب تعیین رگرسیون جنگل تصادفی برابر با ۰/۹۲۴ و ضریب تعیین رگرسیون خطی چندگانه برابر با ۰/۲۴۸ است. همچنین خطای رگرسیون جنگل تصادفی برابر با $RMSE = ۰/۰۰۹$ و خطای رگرسیون خطی چندگانه برابر با

¹ Coefficient Of Determination

² Root Mean Squared Error (RMSE)

۰/۰۲۴ $RMSE =$ است. بدین ترتیب، رگرسیون جنگل تصادفی با خطای کمتر و ضریب تعیین بالاتر نسبت به رگرسیون خطی چندگانه عملکرد بهتری در مدلسازی دمای سطح زمین شبانه فصل تابستان داشته است. مپاکایری و موونگوی (۲۰۱۹) در پژوهش خود از رگرسیون جنگل تصادفی برای مدلسازی دمای سطح زمین استفاده کرده‌اند که نتایج ایشان ضریب تعیین ۰/۶۵ را نشان می‌دهد. مطالعه حاضر مدلسازی بهتری را با استفاده از این روش رگرسیونی انجام داده است. تنگ و همکاران (۲۰۲۲) نیز در مطالعه خود با استفاده از رگرسیون جنگل تصادفی تخمین مطلوبی ($R^2 \geq 0/9$) را از دمای سطح زمین انجام داده‌اند که هم راستا با یافته پژوهش حاضر است.



شکل (۲): نمودار ضریب تعیین (الف) الگوریتم جنگل تصادفی و (ب) رگرسیون خطی چندگانه

Fig. (2): Determination coefficient of (a) the random forest regression and (b) the multiple linear regression

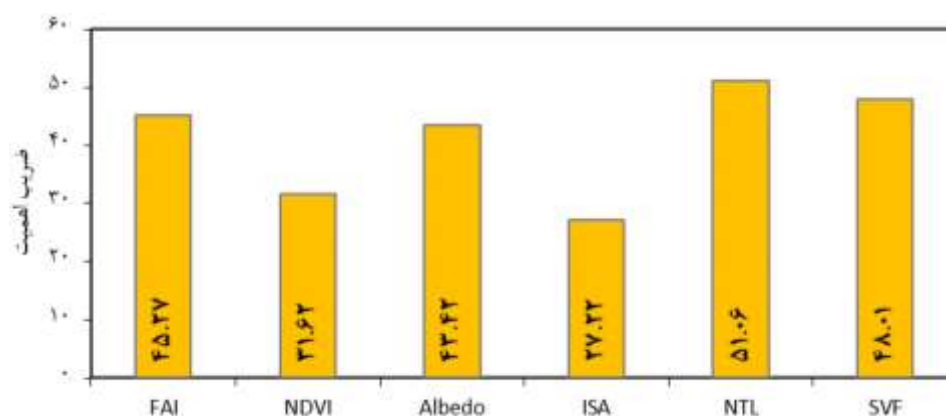
۲-۳- اهمیت شاخص‌ها

به جهتی که رگرسیون جنگل تصادفی عملکرد بهتری داشت، نتایج مربوط به اهمیت شاخص این روش گزارش می‌شود. شکل (۳) اهمیت شاخص‌های مورد استفاده در پژوهش در راستای دمای سطح زمین را نشان می‌دهد. در بین شاخص‌های استفاده شده، شاخص NTL با ۵۱/۰۶ درصد دارای بیشترین اهمیت در راستای دمای سطح زمین شبانه فصل تابستان است. در زمینه نتایج هم راستا، نتیجه مطالعه مپاکایری و موونگوی (۲۰۱۹) نشان می‌دهد که شاخص نور شب مهم‌ترین عامل موثر بر دمای سطح زمین است. همچنین نتایج مطالعه تیموری و همکاران (۱۴۰۲) نیز نتایج مشابهی را در این زمینه نشان می‌دهد. شاخص SVF نیز دومین (۴۸/۰۱)

¹ Tang

² Mpakairi & Muvengwi

درصد) شاخص مهم در راستای دمای سطح زمین شبانه فصل تابستان می‌باشد. نتیجه مشابه در این زمینه در پژوهش چن^۱ و همکاران (۲۰۲۱) مشاهده شده است. ایشان همبستگی قوی را بین SVF و LST روز و شب با استفاده از تصاویر لندست و استر به دست آورده‌اند. شاخص FAI نیز با ۴۵/۲۷ درصد سومین شاخص موثر در دمای سطح زمین شبانه فصل تابستان است. هم راستا با نتیجه مطالعه حاضر، در پژوهش دیگری نیز تاثیر مثبتی ($R^2 = 0/59$) بین FAI و LST مشاهده شده است (Li et al., 2022). علاوه بر این، نتایج لی^۲ و همکاران (۲۰۲۱) خلاف یافته مطالعه حاضر در زمینه شاخص FAI در وهان چین را نشان می‌دهد. نتایج ایشان با استفاده از رگرسیون جنگل تصادفی نشان می‌دهد که شاخص FAI دارای کمترین تاثیر بر LST در تمام فصول می‌باشد. شاخص آلبدو نیز با ۴۳/۴۲ درصد چهارمین مولفه موثر در دمای سطح زمین است که تاثیر آن بر دما در پژوهش‌های متعددی مشاهده شده است (Andrés-Anaya et al., 2021). شاخص‌های NDVI (۳۱/۶۲) و ISA (۲۷/۲۵) نیز به ترتیب کمترین اهمیت را در مدلسازی دمای سطح زمین شبانه فصل تابستان شهر تبریز دارند. همچنین، تاثیر ضعیف NDVI در LST روز نیز در پژوهش دیگری در شهر تبریز مشاهده شده است (آزادی مبارکی و احمدی، ۱۴۰۰). در زمینه شاخص ISA نیز تاثیر قابل توجهی (۰/۶۹) الی ($R^2 = 0/85$) بین این مولفه با LST شبانه در پژوهش‌های دیگر مشاهده شده است که مغایر با نتیجه پژوهش حاضر می‌باشد (Mallick et al., 2013; Mathew et al., 2022).



شکل (۳): نمودار اهمیت شاخص‌های موثر بر متوسط دمای سطح زمین

Fig. (2): Variable importance of the mean LST drivers

¹ Chen

² Li

۴- نتیجه گیری

شناسایی عوامل مؤثر بر دمای سطح زمین در شهرها به جهت مدیریت حرارتی مؤثر نواحی شهری و همچنین اتخاذ استراتژی‌های انطباقی با جزایر حرارتی امری ضروری است. در این راستا، هدف پژوهش حاضر مطالعه تطبیقی رگرسیون‌های خطی چندگانه و جنگل تصادفی در تخمین دمای سطح زمین شبانه فصل تابستان شهر تبریز و همچنین رتبه‌بندی عوامل مؤثر بر آن بود.

بر اساس نتایج عملکرد دو روش رگرسیونی، روش رگرسیونی جنگل تصادفی خطای کمتری را نسبت به رگرسیونی خطی چندگانه از خود نشان داد؛ بنابراین، از این روش جهت استخراج عوامل مؤثر استفاده شد. بدین اساس، مهم‌ترین شاخص در دمای سطح زمین شبانه تابستان شهر تبریز، شاخص نور شب است. این شاخص دارای همبستگی قابل توجهی با فعالیت‌های انسانی (فعالیت‌های صنعتی و غیره) است که در طول شب رخ می‌دهد. دومین شاخص نیز شاخص ضریب دید به آسمان می‌باشد. مورفولوژی شهر تبریز به‌ویژه نواحی مرکزی آن دارای بافت فشرده و همچنین بلندمرتبه است که نقش مهم این عامل را در تغییرات دمایی محیط شهری توجیه می‌نماید. سومین عامل مهم نیز شاخص مساحت رو به باد بود. این عامل از موارد مهم در مطالعه جزایر حرارتی است که محصورشدگی محیط شهری به جهت عدم نفوذپذیری باد را نشان می‌دهد. هر چه محیط شهری فشرده باشد، ضریب نفوذ باد نیز کاهش یافته و روند تخلیه حرارتی با سرعت کمتری انجام می‌شود. چهارمین عامل مؤثر بر دمای سطح زمین نیز آلبدو بود که از مؤلفه‌های اساسی کنترل و کاهش جزایر حرارتی است. شهر تبریز نیز دارای وسعت بالایی از سطوح تیره و غیرقابل نفوذ است که باعث ذخیره گرمایی در طول روز شده و در شب آن را در محیط شهری آزاد می‌نماید. شاخص‌های پوشش گیاهی و مساحت سطح نفوذناپذیر نیز به ترتیب کمترین اهمیت را در دمای سطح زمین شبانه فصل تابستان داشته‌اند. کمبود پوشش گیاهی نیز از دیگر مشکلات شهر تبریز است که از آن رنج می‌برد. کمبود این مقوله و پراکنش نامناسب آن و همچنین کمبود نسبت مساحتی پهنه‌های سبز نیز منجر به کاهش اهمیت آن در تخمین دمای سطح زمین شده است. پوشش گیاهی از جمله مهم‌ترین استراتژی‌های کاهش جزایر حرارتی است که در وهله اول مورد توجه برنامه‌ریزان می‌باشد. البته این استراتژی باید در قالب توجه به منابع آبی و همچنین نوع اقلیم به مرحله اجرا در آید.

۵-منابع

1. Abbot, M. L. (2017). *Using Statistics in the Social and Health Sciences with SPSS and Excel*. John Wiley & Sons. DOI: 10.1002/9781119121077
2. Abrougui, K. & Gabsi, K. & Mercatoris, B. & Khemis, C. & Amami, R. & Chehaibi, Sa. (2019). Prediction of organic potato yield using tillage systems and soil properties by artificial neural network and multiple linear regressions. *Soil and Tillage Research*. Vol. 190. pp 202-208. DOI: 10.1016/j.still.2019.01.011.
3. Andrés-Anaya, P., Sánchez-Aparicio, M., del Pozo, S., & Lagüela, S. (2021). Correlation of Land Surface Temperature with IR Albedo for the Analysis of Urban Heat Island. *Engineering Proceedings*, 8(1). <https://doi.org/10.3390/engproc2021008009>
4. Azadi Mubarak, M., & Ahmadi, M. (2021). Investigation of urban heat islands of Tabriz metropolis using multi-time data of LANDSAT8 satellite based on hot spot analysis method. *Regional Planning*, 11(43), 47–63. <https://doi.org/10.30495/jzpm.2021.3992> (In Persian).
5. Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. <https://doi.org/10.1016/j.isprsjprs.2016.01.011>
6. Bernard, J., Bocher, E., Petit, G., Palominos, S. (2018). "Sky View Factor Calculation in Urban Context: Computational Performance and Accuracy Analysis of Two Open and Free GIS Tools" *Climate* 6(3). <https://doi.org/10.3390/cli6030060>
7. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
8. Breiman, L. (2002). *Manual on setting up, using, and understanding random forests v3.1*. https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf
9. Carlson, T. N., & Rizley, D. A. (1997). On the Relation between NDVI, Fractional Vegetation Cover, and Leaf Area Index. *Remote Sensing of Environment*, 62(3), 241–252. [https://doi.org/10.1016/S0034-4257\(97\)00104-1](https://doi.org/10.1016/S0034-4257(97)00104-1)
10. Chan, J. C.-W., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne

- hyperspectral imagery. *Remote Sensing of Environment*, 112(6), 2999-3011. <https://doi.org/10.1016/j.rse.2008.02.011>
11. Chen, Q., Cheng, Q., Chen, Y., Li, K., Wang, D., & Cao, S. (2021). The Influence of Sky View Factor on Daytime and Nighttime Urban Land Surface Temperature in Different Spatial-Temporal Scales: A Case Study of Beijing. *Remote Sensing*, 13(20). <https://doi.org/10.3390/rs13204117>
 12. Gadekar, K., Pande, C.B., Rajesh, J., Gorantiwar, S.D., Atre, A.A. (2023). Estimation of Land Surface Temperature and Urban Heat Island by Using Google Earth Engine and Remote Sensing Data. In: Pande, C.B., Moharir, K.N., Singh, S.K., Pham, Q.B., Elbeltagi, A. (eds) *Climate Change Impacts on Natural Resources, Ecosystems and Agricultural Systems*. Springer Climate. Springer, Cham. https://doi.org/10.1007/978-3-031-19059-9_14
 13. George, D. & Mallery, P. (2020). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. 16th Edition. Routledge: Taylor & Francis Group. <https://www.routledge.com/IBM-SPSS-Statistics-26-Step-by-Step-A-Simple-Guide-and-Reference/George-Mallery/p/book/9780367174354>
 14. Gong, D., Dai, X., Zhou, L. (2023). Satellite-Based Optimization and Planning of Urban Ventilation Corridors for a Healthy Microclimate Environment. *Sustainability*, 15(21):15653. <https://doi.org/10.3390/su152115653>
 15. Guha, S., Govil, H. (2020). Land surface temperature and normalized difference vegetation index relationship: a seasonal study on a tropical city. *SN Appl. Sci.* 2, 1661. <https://doi.org/10.1007/s42452-020-03458-8>
 16. Haghverdi, A., Noori, R., & Tisseyre, B. (2021). Estimating Land Surface Temperature using MODIS data: A Random Forest approach. *Journal of Environmental Management*, 298, 113491. <https://doi.org/10.1016/j.jenvman.2021.113491>
 17. Hodul, M., Knudby, A., Chak, H. (2016). "Estimation of Continuous Urban Sky View Factor from Landsat Data Using Shadow Detection" *Remote Sensing* 8(7). <https://doi.org/10.3390/rs8070568>
 18. Keerthi Naidu, B., Chundeli, F.A. (2023). Assessing LULC changes and LST through NDVI and NDBI spatial indicators: a case of Bengaluru, India. *GeoJournal* 88, 4335–4350. <https://doi.org/10.1007/s10708-023-10862-1>
 19. Koushesh Vatan, M. A., & Asghari Zamani, A. (2021). Study of land surface temperature concerning land-use in Tabriz city using the Landsat 8 data. *Journal of Economic Geography Researches*, 2(3), 49–58. https://jurs.znu.ac.ir/article_244529.html (In Persian).
 20. Koushesh Vatan, M. A., Asghari Zamani, A., Nemati, M., & Pourmohammad, F. (2021). A Comparison of Multiple Linear Regression and Random Forest Regression to Evaluate the Price of Residential Units (Case Study: North Valiasr, Tabriz). *Geography and Territorial Spatial Arrangement*, 11(40), 57–82. <https://doi.org/10.22111/gaj.2021.6486> (In Persian).
 21. Li, H., Li, Y., Wang, T., Wang, Z., Gao, M., & Shen, H. (2021). Quantifying 3D building form effects on urban land surface temperature and modeling seasonal correlation patterns.

22. Building and Environment, 204, 108132.
<https://doi.org/https://doi.org/10.1016/j.buildenv.2021.108132>
23. Li, L., Zhao, Z., Wang, H., Shen, L., Liu, N., & He, B.-J. (2022). Variabilities of Land Surface Temperature and Frontal Area Index Based on Local Climate Zone. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2166–2174.
<https://doi.org/10.1109/JSTARS.2022.3153958>
24. Li, Z. L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., ... & Zhang, X. (2013). A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors*, 13(5), 5977-6010. <https://doi.org/10.3390/s130505977>
25. Li, Z., Wu, F., Ma, H., Xu, Z., Wang, S. (2022). Spatiotemporal Evolution and Relationship between Night Time Light and Land Surface Temperature: A Case Study of Beijing, China. *Land*. 211(4):548. <https://doi.org/10.3390/land11040548>
26. Liang, S. (2000). Narrowband to broadband conversions of land surface albedo I Algorithms. *Remote Sensing of Environment*, 76, 213–238. www.elsevier.com/locate/rse
27. Liaw, Andy & Wiener, Matthew. (2001). Classification and Regression by RandomForest. *RNews*. 23.
https://www.researchgate.net/publication/228451484_Classification_and_Regression_by_RandomForest
28. Makhija R., Ali S., Jaya Krishna R. (2021) Detecting Influencers in Social Networks Through Machine Learning Techniques. In: Hassanien A., Bhatnagar R., Darwish A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, vol 1141. Springer, Singapore. https://doi.org/10.1007/978-981-15-3383-9_23.
29. Mallick, J., Rahman, A., & Singh, C. K. (2013). Modeling urban heat islands in heterogeneous land surface and its correlation with impervious surface area by using night-time ASTER satellite data in highly urbanizing city, Delhi-India. *Advances in Space Research*, 52(4), 639–655. <https://doi.org/10.1016/j.asr.2013.04.025>
30. Mansourmoghaddam, M., Rousta, I., Ghafarian Malamiri, H., Sadeghnejad, M., Krzyszczyk, J., Ferreira, CSS. (2024). Modeling and Estimating the Land Surface Temperature (LST) Using Remote Sensing and Machine Learning (Case Study: Yazd, Iran). *Remote Sensing*, 16(3):454. <https://doi.org/10.3390/rs16030454>



31. Mathew, A., P. S., & Khandelwal, S. (2022). Investigating the contrast diurnal relationship of land surface temperatures with various surface parameters represent vegetation, soil, water, and urbanization over Ahmedabad city in India. *Energy Nexus*, 5. <https://doi.org/10.1016/j.nexus.2022.100044>
32. Meng, Q., Li, L., & Dong, Q. (2020). Land surface temperature estimation from MODIS data using the Random Forests regression algorithm. *Remote Sensing*, 12(3), 507. <https://doi.org/10.3390/rs12030507>
33. Mpakairi, K. S., & Muvengwi, J. (2019). Night-time lights and their influence on summer night land surface temperature in two urban cities of Zimbabwe: A geospatial perspective. *Urban Climate*, 29. <https://doi.org/10.1016/j.uclim.2019.100468>
34. Pal, M. & Bharati, P. (2019). *Applications of Regression Techniques*. Springer Singapore. <https://www.springer.com/gp/book/9789811393136>
35. Pan, T., Li, B., Ning, L. (2023). "Impervious Surface Area Patterns and Their Response to Land Surface Temperature Mechanism in Urban–Rural Regions of Qingdao, China" *Remote Sensing* 15(17). <https://doi.org/10.3390/rs15174265>
36. Parastatidis D, Mitraka Z, Chrysoulakis N, Abrams M. (2017). Online Global Land Surface Temperature Estimation from Landsat. *Remote Sensing*, 9(12):1208. <https://doi.org/10.3390/rs9121208>
37. Phan, T. L., Kappas, M., Degener, J., & Nguyen, L. D. (2018). Application of Random Forest for spatial distribution of tropical forest aboveground biomass estimation using multi-spectral remote sensing data. *International Journal of Digital Earth*, 11(7), 640-660. <https://doi.org/10.1080/17538947.2017.1365965>
38. Richardson, R. (2015). *Business Applications of Multiple Regression, Second Edition - Quantitative approaches to decision making collection*. Business Expert Press. <https://www.businessexpertpress.com/books/business-applications-multiple-regression-second-edition/>
39. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a Random Forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93-104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
40. Sharma R., Rani S. (2021) A Novel Approach for Smart-Healthcare Recommender System. In: Hassanien A., Bhatnagar R., Darwish A. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2020. Advances in Intelligent Systems and Computing*, vol 1141. Springer, Singapore. https://doi.org/10.1007/978-981-15-3383-9_46.
41. Sobrino, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2016). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90(4), 434-440. <https://doi.org/10.1016/j.rse.2004.02.003>
42. Song, C., Woodcock, C. E., Seto, K. C., Lenney, M. P., & Macomber, S. A. (2014). Classification and change detection using Landsat TM data: When and how to correct

- atmospheric effects? *Remote Sensing of Environment*, 75(2), 230-244. <https://doi.org/10.1016/j.rse.2000.11.011>
43. Tang, J., Lan, X., Lian, Y., Zhao, F., & Li, T. (2022). Estimation of Urban-Rural Land Surface Temperature Difference at Different Elevations in the Qinling-Daba Mountains Using MODIS and the Random Forest Model. *International journal of environmental research and public health*, 19(18), 11442. <https://doi.org/10.3390/ijerph191811442>
 44. Teymouri, I., Asghari Zamani, A., Torabi, S. (2023). Investigating the impact of urban morphological factors on thermal islands in cities, a case study: Urmia city (2021-2015 period). *Remote Sensing and GIS Applications in Environmental Sciences*, 2(5), 111-97. doi: 10.22034/rsgi.2023.16594 (In Persian)
 45. Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86(3), 370-384. [https://doi.org/10.1016/S0034-4257\(03\)00079-8](https://doi.org/10.1016/S0034-4257(03)00079-8)
 46. Wlodarczak, Peter (2020). *Machine Learning and its applications*, First edition. Boca Raton, FL: CRC Press/Taylor & Francis Group. DOI: 10.1201/9780429448782
 47. Wong, M. S., Nichol, J. E., To, P. H., & Wang, J. (2010). A simple method for designation of urban ventilation corridors and its application to urban heat island analysis. *Building and Environment*, 45(8), 1880-1889. <https://doi.org/10.1016/j.buildenv.2010.02.019>
 48. Zhang, Q., Wu, Z., Singh, V., Liu, C. (2021). "Impacts of Spatial Configuration of Land Surface Features on Land Surface Temperature across Urban Agglomerations, China" *Remote Sensing* 13(19). <https://doi.org/10.3390/rs13194008>

A Comparative Study of Multiple Linear Regression and Random Forest in the Estimation of Land Surface Temperature: A Case Study of Tabriz City

Abstract

Land surface temperature, as one of the important and fundamental parameters in climatology, indicates the relationship between the atmosphere and the Earth. Considering the environmental issues of cities, including the intensification of urban heat islands, accurately estimating LST and identifying its influencing factors play a significant role in urban thermal management and adopting adaptive strategies for heat islands. In this regard, this study compares two regression methods: multiple linear regression and random forest in order to estimate the LST. Daily nighttime MODIS images were used to extract the LST of Tabriz city during the summer. These images were processed in the Google Earth Engine platform and averaged for the period from 2018 to 2022. According to the results, the random forest showed significantly better performance with a coefficient of determination of 0.924 (RMS = 0.009) compared to multiple linear regression. The random forest was also used to determine the importance of the indices. Based on the index importance results, night lights (51/06%), sky view factor (48/01%) and frontal area index (45/27%) were the most important factor affecting the nighttime summer LST in Tabriz city, respectively. The findings of this study, in addition to revealing the strength of the random forest regression in estimating LST, also highlight the importance of various indices in the LST. In this context, the study's results will be practical for managing the thermal environment of Tabriz city and adopting mitigation strategies for its heat islands.

Keywords: Land Surface Temperature, Random Forest Regression, Linear Regression, Machine learning, Tabriz.