

پیش‌بینی عملکرد گندم با داده‌های بارش و شاخص نرمال شده تفاوت پوشش گیاهی توسط الگوریتم‌های یادگیری ماشین

سید آرش خسروانی شریعتی^۱، علی عباسی^۲

۱- دانشجوی کارشناسی ارشد مهندسی عمران، دانشگاه فردوسی

۲- استادیار گروه مهندسی عمران، دانشگاه فردوسی مشهد (نویسنده مسئول)

aabbasi@um.ac.ir

چکیده

گندم یک محصول استراتژیک در دنیا است و تغییرات میزان تولید آن می‌تواند قیمت این محصول و مواد غذایی وابسته به آن را تحت تاثیر قرار دهد. بنابراین گندم نقش تعیین‌کننده‌ای در امنیت غذایی هر کشور دارد. در این پژوهش عملکرد گندم زمستانه از سال ۲۰۰۸ تا ۲۰۲۱ در ۵۳ شهرستان ایالت کنزاس در آمریکا مورد بررسی قرار گرفته است. برای پیش‌بینی عملکرد گندم از شاخص بارش استاندارد شده (SPI) در دو مقیاس یک و شش ماهه و شاخص نرمال شده تفاوت پوشش گیاهی (NDVI) به عنوان متغیرهای ورودی استفاده شده است. دو الگوریتم یادگیری ماشین غیرخطی به نام جنگل تصادفی و ماشین بردار پشتیبان برای انجام این پیش‌بینی به کار گرفته شدند. بررسی همبستگی بین متغیرهای ورودی با عملکرد گندم نشان داد که شاخص NDVI در ماه می و شاخص SPI با مقیاس شش ماهه در ماه آوریل بیش‌ترین همبستگی را با عملکرد گندم دارند. عملکرد دو الگوریتم یادگیری ماشین تقریباً مشابه یکدیگر بوده و مقدار ضریب تعیین در هر دو مدل پیش‌بینی در حدود ۰.۴ بوده است.

کلمات کلیدی: عملکرد گندم، یادگیری ماشین، بارش، شاخص نرمال شده تفاوت پوشش گیاهی، سنجش از دور

۱- مقدمه

عملکرد محصولات کشاورزی نقش مهمی در برنامه‌ریزی و امنیت غذایی کشور دارد. این پدیده تابع عواملی مانند بارش، دما و نوع خاک است. شناخت تاثیر این عوامل بر میزان تولید محصولات، می‌تواند منجر به برنامه‌ریزی بهتر، جهت دهی به سیاست‌های کلان کشور و کاهش اثرات پدیده‌هایی مانند خشکسالی شود. علاوه بر این، درآمد کشاورزان نیز تحت تاثیر میزان تولید محصولات و عوامل اقلیمی قرار می‌گیرد.

طبق گزارش سازمان فائو^۱، تعداد افرادی که با ناامنی غذایی حاد مواجه هستند و نیاز به کمک‌های غذایی و حمایت معیشتی فوری دارند با سرعت نگران‌کننده‌ای در حال افزایش است. طبق این گزارش در سال ۲۰۲۱، ۱۹۳ میلیون نفر در ۵۱ کشور دنیا دچار عدم امنیت غذایی بودند [1].

با توجه به کاهش دسترسی به منابع آب در طی سال‌های اخیر و افزایش جمعیت، نیاز به برآورد میزان تولید محصولات کشاورزی بیش از پیش مورد توجه قرار گرفته است. علاوه بر این، دانستن میزان برداشت محصولات کشاورزی می‌تواند به سیاست‌گذاران در زمینه‌ی میزان واردات، صادرات محصولات و قیمت‌گذاری آن‌ها کمک فراوانی کند. بنابراین پایش و پیش‌بینی وضعیت محصولات کشاورزی یکی از چالش‌های مهم بوده و جزئی ضروری برای امنیت غذایی و اقتصادی هر کشور است.

به منظور پیش‌بینی عملکرد محصولات می‌توان از روش‌های نمونه‌گیری، مدل‌های رشد گیاه و مدل‌سازی اطلاعات (مدل‌سازی آماری و یادگیری ماشین) استفاده نمود [2]. با توجه به اینکه دو روش اول زمان‌بر هستند و نیاز به داده‌های زیادی مانند نوع خاک و عوامل مدیریتی زمین‌های کشاورزی دارند در این پژوهش از مدل‌های یادگیری ماشین استفاده شده است.

به کمک علم سنجش از دور که بر مبنای استفاده از تصاویر ماهواره‌ای می‌باشد، امکان سنجش و پایش سلامت محصولات کشاورزی با استفاده از تصاویر ماهواره‌ای فراهم شده است و می‌توان سبزی‌نگی و میزان رشد آن‌ها را با سال‌های دیگر مقایسه نمود. در سال‌های اخیر، محققان در تلاش بوده‌اند

¹ FAO (Food and Agriculture Organization of the United Nations)

که رابطه‌ای بین عواملی مانند سبزیگی محصول و بارش با عملکرد محصولات کشاورزی گزارش شده پیدا کنند، اما اثر این عوامل و شدت آن‌ها هنوز جای بحث دارد و پیدا کردن ترکیب موثری از این پارامترها همواره چالش برانگیز بوده است.

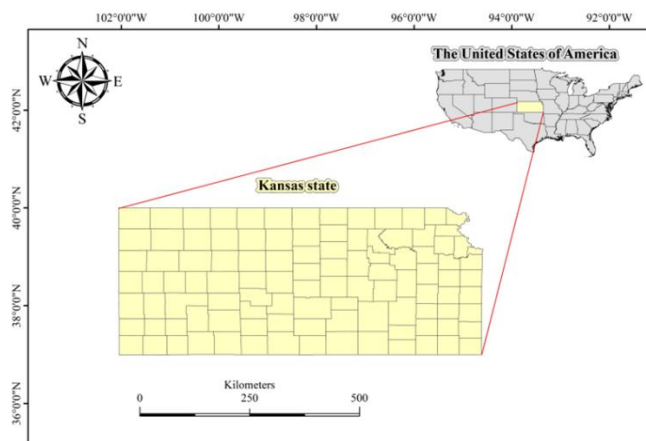
در تحقیقی آراسته و همکاران (۱۳۹۱) رابطه‌ی عوامل هواشناسی و چهار شاخص خشکسالی را با عملکرد گندم و جوی دیم در استان خراسان در بازه‌ی سال‌های ۱۳۷۳ تا ۱۳۹۱ سنجیدند. آن‌ها روابط آماری بین عملکرد محصولات و شاخص‌های خشکسالی را با همبستگی و رگرسیون تحلیل کردند. بر اساس همبستگی بین عملکرد گندم و جو و پارامترهای هواشناسی، نقش بارندگی و رطوبت نسبی از سایر عوامل بیش‌تر بود. آن‌ها بیان کردند که روابط رگرسیونی متغیرهای اقلیمی با عملکرد گندم و جو در اغلب موارد معنادار بود، اما این انطباق در مناطق با بارش کم ناچیز است [3].

زارع ایبانه (۱۳۹۲) با انتخاب ده پارامتر هواشناسی و هفت شاخص خشکسالی به عنوان ورودی مدل رگرسیونی، عملکرد چهار محصول دیم شامل گندم، جو، هندوانه و نخود در شهرهای بیرجند و مشهد را پیش‌بینی کرد. نتایج بیانگر این بود که از بین متغیرهای هواشناسی، دو متغیر تعداد روزهای بارانی در شهر بیرجند و دمای پیش‌بینی هوا در مشهد، بیش‌ترین همبستگی را با عملکرد محصولات زراعی داشتند. از بین شاخص‌های خشکسالی، شاخص نگوین نیز همبستگی بالایی با عملکرد محصولات داشت [4].

زو و همکاران (۲۰۲۱) از شاخص‌های هواشناسی-کشاورزی و سنجش از راه دور برای تخمین عملکرد ذرت در دو استان جیلین و لیائونینگ چین استفاده کردند. دو روش انتخاب یک مرحله‌ای و دو مرحله‌ای پارامترها، برای برگزیدن متغیرهای ورودی بهینه با توجه به تأثیرگذاری‌شان در پیش‌بینی استفاده شده است. از الگوریتم جنگل تصادفی نیز برای ایجاد مدل‌های پیش‌بینی عملکرد محصول استفاده کردند و صحت مدل‌ها را در سه سالی که این دو منطقه دچار خشکسالی بوده‌اند (۱۹۹۷، ۲۰۰۰ و ۲۰۰۱) مورد سنجش قرار دادند. نتایج نشان داد که مقادیر ضریب تعیین (R^2) مدل‌های برآورد عملکرد ذرت در دو استان، همگی بالای ۰.۷ بوده است. البته مدلی که در آن پارامترها به صورت دو مرحله‌ای انتخاب شدند، دقت بالاتر و پارامترهای ورودی کمتری نیاز داشت. از بین بازده پارامتر ورودی، پارامتر مجموع دماهای بالای ۱۰ درجه، تولید اولیه ناخالص در ماه آگوست، آنومالی شاخص نرمال شده تفاوت پوشش گیاهی ($NDVI^1$) در ماه آگوست و شاخص بارش استاندارد شده (SPI^2) با مقیاس دو ماهه در ماه جولای مهم‌ترین پارامترها بودند [2]. هدف از این پژوهش پیش‌بینی عملکرد گندم زمستانه در ایالت کنزاس آمریکا با استفاده از الگوریتم‌های یادگیری ماشین می‌باشد. برای این منظور از شاخص $NDVI$ و SPI به عنوان پارامترهای ورودی استفاده شده است و رابطه‌ی آن‌ها با عملکرد گندم در ماه‌های مختلف رشد آن بررسی شده است.

۲- منطقه‌ی مورد مطالعه و ویژگی‌های گندم زمستانه

در این پژوهش منطقه‌ی مورد مطالعه انتخابی برای پیش‌بینی عملکرد گندم، ایالت کنزاس آمریکا می‌باشد. دلیل انتخاب این منطقه این بوده است که این ایالت بیش‌ترین تولید گندم زمستانه را در بین تمام ایالت‌های آمریکا دارد و طبق آمار سازمان خدمات ملی آمار کشاورزی آمریکا در سال ۲۰۲۱ در این ایالت در حدود سه میلیون هکتار گندم زمستانی کاشت شده است. موقعیت این ایالت در شکل شماره‌ی ۱ آورده شده است:



شکل شماره‌ی ۱- ایالت کنزاس

1 Normalized Difference Vegetation Index

2 Standardized Precipitation Index

از بین ۱۰۵ شهرستان در این ایالت، ابتدا شهرستان‌ها بر اساس میزان تولید گندم از سال ۲۰۰۸ تا ۲۰۲۱ مرتب شدند و شهرستان‌هایی که میزان تولید آن‌ها بالای صدک سی بود، انتخاب شدند و سپس از بین آن‌ها شهرستان‌هایی که بیش‌تر از ۳ مقدار نامعلوم در عملکرد گندم داشتند، حذف شدند. با توجه به این روش، در انتها ۵۳ شهرستان انتخاب شدند و پیش‌بینی عملکرد گندم بر روی این ۵۳ شهرستان انجام شد. بازه زمانی بررسی عملکرد گندم در این پژوهش از سال ۲۰۰۸ تا ۲۰۲۱ در نظر گرفته شده است.

گندم زمستانه در این ایالت معمولاً در پاییز کاشته می‌شود. جوانه زنی و سبز شدن محصول معمولاً در اواخر پاییز قبل از زمستان اتفاق می‌افتد. گندم زمستانه برای گل‌دهی در فصل بهار پیش رو، نیاز به فرآیند بهاره‌سازی دارد. بهاره‌سازی عبارت است از نیاز به قرار گرفتن در دمای بین ۵ تا ۱۰ درجه سانتی‌گراد برای یک دوره ۶ تا ۸ هفته‌ای و محصول باید حداقل چهار هفته قبل از بهاره‌شدن از نظر بیولوژیکی فعال باشد. در مرحله بهاره‌سازی، رشد گندم متوقف می‌شود و هنگامی که خاک در بهار گرم می‌شود، گندم زمستانه رشد رویشی خود را از سر می‌گیرد و در اوایل تا اواسط تابستان به بلوغ می‌رسد [5]. زمان برداشت آن هم از اواسط ماه ژوئن تا اواسط جولای است. بنابراین در این پژوهش از ماه اکتبر تا جولای سال بعد آن به عنوان بازه‌ی زمانی رشد گندم در هر سال در نظر گرفته شده است.

۳- مواد و روش‌ها

۳-۱ داده‌های مورد استفاده

داده‌های مورد استفاده در پژوهش حاضر عبارتند از: ۱- عملکرد گندم زمستانه، ۲- لایه‌ی اراضی زیر کشت محصولات کشاورزی، ۳- شاخص NDVI از تصاویر سنجنده مودیس^۱، ۴- شاخص SPI در مقیاس زمانی یک و شش ماهه.

۳-۱-۱ عملکرد گندم زمستانه

عملکرد محصولات کشاورزی هر ساله توسط سازمان خدمات ملی آمار کشاورزی آمریکا منتشر می‌شود. این گزارش در سطح ملی، ایالتی، بخش‌های کشاورزی و شهرستان می‌باشد. در این پژوهش از داده‌های سطح شهرستان استفاده شده است. واحد اندازه‌گیری عملکرد گندم در این گزارش‌ها بر اساس بوشل بر جریب است. هر بوشل گندم تقریباً معادل با ۲۷.۲ کیلوگرم و هر جریب معادل با ۴۰۴۶.۹ متر مربع است. در این تحقیق نتایج به کیلوگرم بر هکتار تبدیل شده است.

۳-۱-۲ لایه‌ی اراضی زیر کشت محصولات کشاورزی

سازمان خدمات ملی آمار کشاورزی آمریکا از سال ۱۹۹۷ تحت عنوان پروژه‌ی لایه‌ی داده‌ی اراضی کشاورزی شروع به تهیه نقشه‌ی کاشت سالانه محصولات کشاورزی کرده است. تا قبل از سال ۲۰۰۸ این نقشه‌ها محدود به تعداد اندکی ایالت بودند، اما بعد از این سال محدوده‌ی داده‌ها کامل‌تر شد و سراسر آمریکا را پوشش می‌دهند.

۳-۱-۳ شاخص NDVI

شاخص NDVI به صورت زیر محاسبه می‌شود:

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (1)$$

که در آن NIR مادون قرمز نزدیک و RED نور قرمز می‌باشد. پوشش گیاهی سالم دارای کلروفیل بیش‌تری بوده و باعث می‌شود که میزان انعکاس مادون قرمز نزدیک بیش‌تری نسبت به گیاه ناسالم داشته باشد. علاوه بر این، گیاه سالم نور قرمز را بیش‌تر جذب می‌کند؛ در نتیجه میزان NDVI بیش‌تر می‌شود. این شاخص از لحاظ تئوری بین ۱- تا ۱ نوسان دارد. در این پژوهش برای محاسبه‌ی این شاخص از تصاویر سنجنده‌ی مودیس با نام اختصاصی "MCD43A4" استفاده شده است. این تصاویر از ترکیب ۱۶ روزه‌ی تصاویر ماهواره‌ی Terra و Aqua به صورت روزانه تولید می‌شوند.

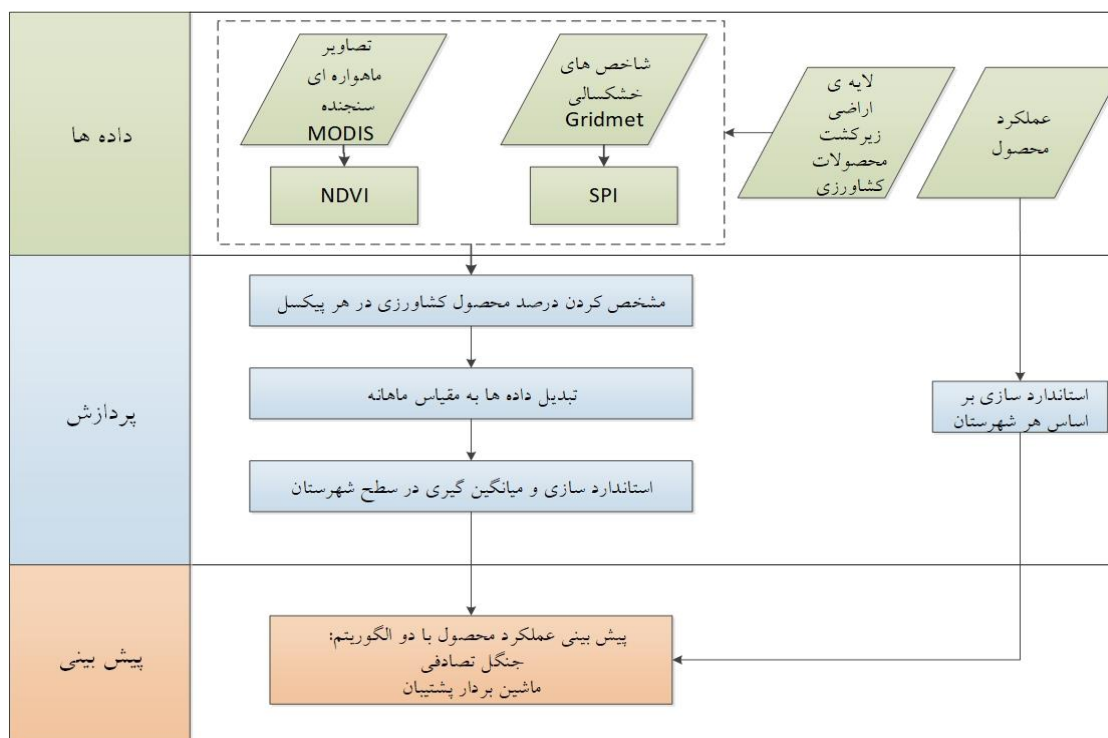
¹ MODIS (Moderate Resolution Imaging Spectroradiometer)

۳-۱-۴ شاخص SPI

شاخص SPI توسط مک کی در سال ۱۹۹۳ بر اساس بارش معرفی شد. این شاخص جهت پایش خشکسالی در مقیاس‌های زمانی مختلف (۱ تا ۳۶ ماهه) طراحی شده است. برای محاسبه‌ی این شاخص داده‌های خام بارش معمولاً به یک توزیع گاما یا پیرسون تیب ۳ برازش داده می‌شوند و سپس به توزیع نرمال تبدیل می‌شوند، به طوری که میانگین SPI در دوره‌ی مورد نظر صفر باشد. این شاخص مقادیری بین ۲- تا ۲ می‌تواند داشته باشد که مقادیر بیشتر از صفر نمایانگر مقدار بارش بیشتر از میانگین در مکان مورد نظر است. برای دریافت شاخص SPI از داده‌های gridMET در این پژوهش استفاده شده است.

۳-۲ روش انجام پژوهش

در ابتدا تصاویر سنجنده‌ی مودیس دریافت شدند و شاخص NDVI بر اساس باند قرمز و مادون قرمز نزدیک در بازه‌ی زمانی مورد مطالعه محاسبه شد. شاخص SPI با مقیاس یک و شش ماهه از gridMET دریافت شد. سپس بر اساس لایه‌ی اراضی کشت محصولات کشاورزی، میزان درصد گندم در هر پیکسل برای شاخص‌های NDVI و SPI محاسبه شد. با توجه به اینکه داده‌های ورودی تفکیک زمانی مختلفی دارند، برای پیش‌بینی عملکرد گندم، باید یک مقیاس زمانی یکسان برای داده‌ها در نظر گرفته شود. بنابراین برای ساده‌تر شدن محاسبات، تمامی داده‌ها به مقیاس ماهانه تبدیل شدند. با توجه به اینکه شرایط رشد گندم زمستانه در هر شهرستان متفاوت بوده است، تمامی پارامترهای دریافتی نسبت به هر پیکسل استانداردسازی شدند؛ به عبارت دیگر تغییرات عوامل ذکر شده سنجیده شدند. سپس در هر شهرستان میانگین پیکسل‌های هر پارامتر با توجه به درصد گندم در هر پیکسل برای آن شهرستان محاسبه شده است. در نهایت این پارامترها به عنوان داده‌ی ورودی برای پیش‌بینی عملکرد گندم زمستانه توسط دو الگوریتم یادگیری ماشین جنگل تصادفی و ماشین بردار پشتیبان استفاده شدند. مراحل مختلف انجام پژوهش در شکل شماره‌ی ۲ نشان داده شده است.



شکل شماره‌ی ۲- روند کلی انجام پژوهش

۳-۳ الگوریتم‌های یادگیری ماشین

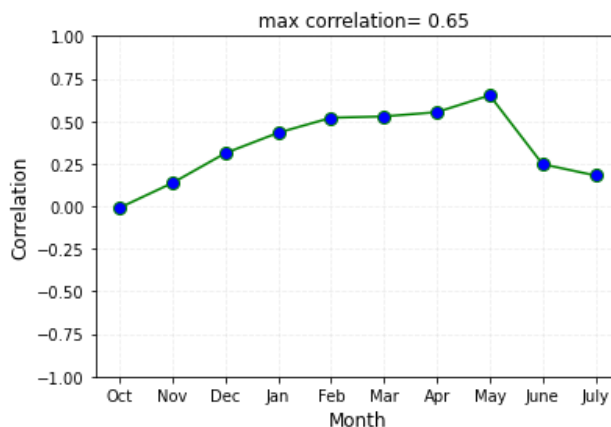
در این پژوهش به منظور پیش‌بینی عملکرد گندم از دو الگوریتم جنگل تصادفی و ماشین بردار پشتیبان استفاده شده است.

الگوریتم جنگل تصادفی توسط لئو برایمن و ادل کانلر در سال ۲۰۰۱ معرفی شد و برای طبقه‌بندی و رگرسیون مورد استفاده قرار می‌گیرد. الگوریتم رگرسیونی جنگل تصادفی مجموعه‌ای از درخت‌های تصمیم است که هر درخت پیش‌بینی خود را انجام می‌دهد و پیش‌بینی نهایی با میانگین‌گیری پیش‌بینی همه‌ی درخت‌های تصمیم در این الگوریتم به دست می‌آید. ماشین بردار پشتیبان در ابتدا برای حل مسائل دسته‌بندی طراحی شده بود. اما در سال ۱۹۹۶، واپنیک و همکارانش نسخه‌ای از این الگوریتم را توسعه دادند که عمل رگرسیون را انجام می‌دهد که به نام رگرسیون ماشین بردار پشتیبان معروف است.

۴- نتایج و بحث

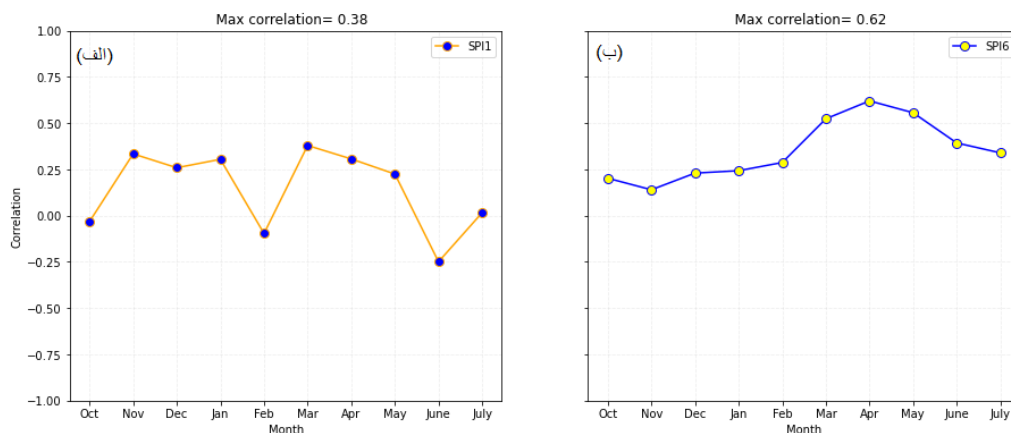
۱-۴ بررسی همبستگی پارامترها با عملکرد گندم

در شکل شماره‌ی ۳ رابطه‌ی همبستگی بین شاخص NDVI و عملکرد گندم استاندارد شده در بین تمامی شهرستان‌های بررسی شده نشان داده می‌شود. در زمان شروع فصل رشد گندم در ماه اکتبر، این همبستگی تقریباً نزدیک به صفر بوده است که با جوانه زدن محصول و رشد آن این همبستگی افزایش پیدا می‌کند. این همبستگی در ماه می به مقدار حداکثر خود (۰.۶۵) می‌رسد و دلیل آن این است که این ماه اوج سبزیگی گندم است.



شکل شماره‌ی ۳- همبستگی شاخص استاندارد شده NDVI با عملکرد گندم استاندارد شده از سال ۲۰۰۸ تا ۲۰۲۱

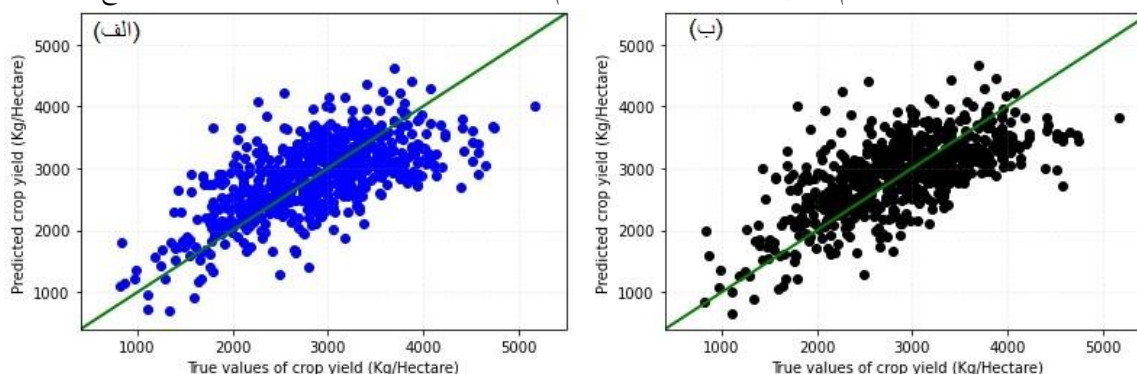
همبستگی شاخص SPI در دو مقیاس زمانی یک و شش ماهه در شکل شماره‌ی ۴ مقایسه شده است. در مقیاس یک ماهه، مشاهده می‌شود که در ماه‌های ابتدایی رشد همبستگی معناداری با عملکرد محصول وجود دارد، پس از آن در ماه فوریه این همبستگی کم شده است و در نهایت در مرحله‌ی ظهور سنبله و گلدهی این همبستگی افزایش یافته است. در مقیاس شش ماهه همبستگی از حدود صفر در ابتدای فصل رشد شروع می‌شود و در ماه آوریل به حداکثر مقدار خود می‌رسد اما پس از ماه آوریل همبستگی آن با عملکرد گندم زمستانه کاهش می‌یابد.



شکل شماری ۴- بررسی همبستگی شاخص‌های SPI با عملکرد گندم استاندارد شده: الف) در مقیاس زمانی یک ماهه، ب) در مقیاس زمانی شش ماهه

۲-۴ نتایج مدل پیش‌بینی گندم

در این پژوهش از دو الگوریتم جنگل تصادفی و ماشین بردار پشتیبان جهت پیش‌بینی عملکرد گندم استفاده شده است. برای انجام این پیش‌بینی از دو پارامتر ورودی شاخص NDVI در ماه می و شاخص SPI با مقیاس شش ماهه در ماه آوریل استفاده شده است. دلیل انتخاب این دو پارامتر این است که بیش‌ترین همبستگی را با عملکرد گندم دارند. در شکل شماری ۵ مقادیر پیش‌بینی شده عملکرد گندم در برابر مقدار واقعی آن نمایش داده شده است. واحد در نظر گرفته شده برای عملکرد گندم در این دو نمودار بر حسب کیلوگرم بر هکتار بوده است که هر هکتار معادل با ده هزار متر مربع است.



شکل شماری ۵- مقایسه عملکرد پیش‌بینی شده با عملکرد واقعی گندم: الف) با الگوریتم جنگل تصادفی، ب) با الگوریتم ماشین بردار پشتیبان

در جدول شماری ۱ نتایج این دو مدل با معیارهای آماری شامل ضریب تعیین (R^2) و خطای میانگین مطلق (MAE) سنجیده شده است. همان‌طور که مشاهده می‌شود دقت این دو الگوریتم یادگیری ماشین تقریباً نزدیک به هم بوده است و مقدار ضریب تعیین هر دوی آن‌ها در حدود ۰.۴ بوده است. دلیل اینکه این ضریب تعیین بالا نمی‌باشد این است که پارامترهایی مانند دما، نوع خاک و عوامل مدیریتی زمین‌های کشاورزی در این دو مدل یادگیری ماشین در نظر نگرفته شدند و در صورت اضافه کردن این داده‌ها به عنوان پارامترهای ورودی احتمالاً دقت مدل افزایش پیدا کند.

جدول شماری ۱- نتایج آماری پیش‌بینی عملکرد گندم

MAE (kg/hectare)	R^2	الگوریتم یادگیری ماشین
۴۴۴.۷۵	۰.۴۰	جنگل تصادفی
۴۳۲.۴	۰.۴۲	ماشین بردار پشتیبان



۵- نتیجه‌گیری

دو شاخص NDVI و SPI رابطه‌ی معناداری با عملکرد گندم زمستانه دارند و با توجه به نتایج حاصل از پیش‌بینی عملکرد گندم، می‌توان نتیجه گرفت که با ترکیب این دو پارامتر می‌توان به نتایج قابل قبولی در پیش‌بینی تغییرات عملکرد گندم دست یافت. دو الگوریتم یادگیری ماشین استفاده شده در این تحقیق دقت پیش‌بینی تقریباً مشابه‌ای داشتند.

۶- مراجع

1. FAO, "Global Report on Food Crises: acute food insecurity hits new highs" <https://www.fao.org/newsroom/detail/global-report-on-food-crises-acute-food-insecurity-hits-new-highs/en> (Accessed Date December 30, 2022).
2. X. Zhu, R. Guo, T. Liu, and K. Xu, "Crop yield prediction based on agrometeorological indexes and remote sensing data," *Remote Sens.*, vol. 13, no. 10, 2021, doi: 10.3390/rs13102016.
۳. آراسته، م. و کابلی، س. و یزدانی، م. "بررسی تأثیرات خشکسالی‌های هواشناسی بر عملکرد دو محصول گندم و جو دیم (مطالعه موردی: استان خراسان رضوی)،" *هواشناسی کشاورزی*، ۱۳۹۶، ۵(۱)، ۱۵-۲۵.
۴. زارع ایبانه، ح. "بررسی نقش عوامل اقلیمی و خشکسالی بر تغییرپذیری عملکرد چهار محصول دیم در مشهد و بیرجند،" *دانش آب و خاک (دانش کشاورزی)*، ۱۳۹۲، ۲۳(۱)، ۳۹-۵۶.
5. I. Becker-Reshef, E. Vermote, M. Lindeman, and C. Justice, "A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data," *Remote Sens. Environ.*, vol. 114, no. 6, pp. 1312–1323, 2010, doi: 10.1016/j.rse.2010.01.010.