2024, Vol.11, No.2, Pages 39 to 56
Journal of Water and Sustainable Development

سال یازدهم، شماره ۲، ۱۴۰۳، صفحات ۳۹ تا ۵۶
نشریه آب و توسعه پایدار

Article Type: Technical paper

نوع مقاله: فنی و ترویجی

A Review of the Application of Data Science and Machine Learning in Agricultural Water Management

R. Delbaz[1], H. Ebrahimian[2*]

1, 2- PhD Candidate and Associate Professor, Department of Irrigation and Reclamation Engineering, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran.

*(Corresponding Author Email: ebrahimian@ut.ac.ir)

Received: 08-02-2024 Revised: 27-04-2024
Accepted: 11-05-2024 Available Online: 29-08-2024

مروری بر کاربرد علم داده و یادگیری ماشین در مدیریت آب کشاورزی

رضا دلباز[1]، حامد ابراهیمیان[2*]

۱ و ۲- به ترتیب دانشجوی دکتری و دانشیار، گروه مهندسی آبیاری و آبادانی، دانشکدگان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

*(رایانامه نویسنده‌ی مسئول، E-Mail: ebrahimian@ut.ac.ir)

تاریخ دریافت: ۱۴۰۲/۱۱/۱۹ تاریخ بازنگری: ۱۴۰۳/۰۲/۰۸
تاریخ پذیرش: ۱۴۰۳/۰۲/۲۲ تاریخ انتشار: ۱۴۰۳/۰۶/۰۸

Abstract

New technologies and innovations can improve water management in agriculture. Data science and machine learning are emerging technologies. Data science is a growing field in the world of technology that helps analyze, extract information, and understand patterns and relationships in big data. It plays a pivotal role in a wide range of industries, including agriculture and environmental science. One field in which data science has a significant impact is water science and engineering. The aim of this research is to provide a comprehensive definition of data science and review existing studies in this field. According to the results, 10% of the studies conducted in the field of machine learning in agriculture are related to water management. Furthermore, among all studies conducted in this field from 2018 to 2020, Iran accounted for 5.62% of the total. This field of research has primarily focused on determining crop evapotranspiration, predicting yield, and assessing water quality. However, given the novelty of this technology, there are still gaps in studies in this field, which is expected to be attracted by researchers in the future. On the other hand, like other emerging technologies, there are challenges in the implementation and execution of data science that require collaborative efforts among policymakers, researchers, and farmers to address. To resolve these challenges, it is necessary for these stakeholders to propose solutions that can optimally leverage the benefits of data science while simultaneously addressing the existing challenges and problems.

Keywords: Irrigation, Data Mining, Remote Sensing, Water Management, Artificial Intelligence.

چکیده

مدیریت آب در کشاورزی از طریق فناوری‌ها و نوآوری‌های جدید قابل ارتقا است. از جمله فناوری‌های رو به رشد، علم داده و یادگیری ماشین است. علم داده یک حوزه رو به رشد در دنیای فناوری اطلاعات است که به تحلیل، استخراج اطلاعات، و فهم الگوها و روابط در داده‌های بزرگ کمک می‌کند. این حوزه در صنایع مختلف و بخصوص در زمینه‌های کشاورزی و محیط‌زیست نقش مهمی ایفا می‌کند. یکی از زمینه‌هایی که علم داده در آن تأثیر بسزایی دارد، زمینه علوم و مهندسی آب است. هدف این پژوهش ارائه یک تعریف جامع از علم داده و بررسی مطالعات انجام شده در این زمینه است. بر اساس نتایج به‌دست آمده، ۱۰ درصد مطالعات انجام شده این حوزه در کشاورزی، به موضوع مدیریت آب اختصاص یافته‌اند. همچنین در بین کل مطالعات انجام شده در این حوزه بین سال‌های ۲۰۱۸ تا ۲۰۲۰، ایران سهم ۵/۶۲ درصدی بین این مطالعات داشته است. تمرکز پژوهشگران نیز، بیشتر در موضوعات تعیین تبخیر-تعرق گیاه، پیش‌بینی عملکرد و تعیین کیفیت آب بوده است. با این حال با توجه به نوظهور بودن این فناوری، هنوز خلاءهای مطالعاتی در این حوزه وجود دارد که انتظار می‌رود در آینده نظر محققین به آن‌ها جلب شود. از طرف دیگر، همانند سایر فناوری‌های جدید، مشکلاتی در زمینه اجرا و پیاده‌سازی آن وجود دارد که برای حل این مشکلات، لازم است تا سیاستمداران، محققین، و کشاورزان با همکاری یکدیگر، راهکارهایی را ارائه کنند که بتوانند از مزایای علم داده به شکل بهینه استفاده کنند و در عین حال به چالش‌ها و مشکلات موجود پاسخ دهند.

واژه‌های کلیدی: آبیاری، داده‌کاوی، سنجش از دور، مدیریت آب، هوش مصنوعی.

How to cite this article: Delbaz, R., & Ebrahimian, H. (2024). A Review of the Application of Data Science and Machine Learning in Agricultural Water Management. Journal of Water and Sustainable Development, 11(2), 39-56. doi: http://dx.doi.org/10.22067/jwsd.v11i2.2402-1310

مقدمه

آب راهبردی‌ترین کالای روی کره زمین است و اگر از آن محافظت نشود، کمیاب می‌شود. به طور متوسط، ۷۰ درصد آب برداشتی در کشاورزی مصرف می‌شود (Wang و همکاران، ۲۰۰۲). به گفته فائو (FAO[1])، کشاورزی باید با چالش‌های پیچیده‌ای که به دلیل افزایش جمعیت، رشد اقتصادی، تغییر ترجیحات مصرف‌کنندگان و رقابت برای آب به وجود آمده، سازگار شود (FAO، ۲۰۲۰). چالش‌هایی مانند افزایش جمعیت و تقاضا برای غذا، تخریب خاک‌ها و اکوسیستم‌های مرتبط با آب، آلودگی، و تغییرات آب و هوایی عدم قطعیت‌هایی را در مورد در دسترس بودن منابع آب به همراه دارد (Rosegrant و همکاران، ۲۰۰۹؛ Clark و Tilman، ۲۰۱۷؛ Sundmaeker و همکاران، ۲۰۲۲). این چالش‌ها شامل عدم قطعیت بالا در مورد اثرات تغییر آب و هوا و همچنین افزایش هزینه‌های استخراج آب است (Rosegrant و همکاران، ۲۰۰۹). همچنین گزارش برنامه توسعه سازمان ملل در سال ۲۰۲۱ درباره "استفاده از فناوری دیجیتال برای کشاورزی پایدار" بیان می‌کند که تولید جهانی غذا باید ۹۸ درصد افزایش یابد تا نیاز جمعیت رو به رشد انسانی (۹/۹ میلیارد نفر تا سال ۲۰۵۰) را تأمین کند (Burra و همکاران، ۲۰۲۱). بنابراین، چگونگی مدیریت و کاهش مصرف آب بدون به خطرانداختن بهره‌وری، حیاتی است. اگر از فناوری‌ها و نوآوری‌های جدید استفاده شود، می‌توان به شیوه‌های مناسب کشاورزی دست یافت (Sundmaeker و همکاران، ۲۰۲۲). از جمله فناوری‌های رو به رشد در مطالعات مختلف، علم داده[2] و یادگیری ماشین[3] است.

اصطلاح علم داده در اواخر دهه ۱۹۹۰ در بحث‌های مربوط به نیاز آماردانان به پیوستن به دانشمندان رایانه برای ایجاد دقت ریاضی در تجزیه و تحلیل محاسباتی مجموعه داده‌های بزرگ مطرح شد. در سال ۲۰۰۱ ویلیام اس. کلیولند[4] برنامه عملیاتی را برای تأسیس دانشگاهی در زمینه علم داده منتشر کرد. این برنامه بر لزوم درک علم داده به عنوان یک مبحث چند رشته‌ای تأکید داشت. علم داده نشان‌دهنده بهینه‌سازی فرآیندها و منابع است (Kelleher و Tierney، ۲۰۱۸). علم داده بینش‌های داده را تولید می‌کند؛ نتیجه‌گیری‌ها یا پیش‌بینی‌های عملی و مبتنی بر داده‌ها که می‌توان از آن‌ها برای درک و بهبود حوزه‌های مختلف علمی استفاده کرد. گفته می‌شود استفاده از بینش علم داده مانند توانایی دیدن در تاریکی است (Pierson، ۲۰۲۱). به طور کلی هدف علم داده بهبود تصمیم‌گیری بر مبنای بینش‌هایی است که از مجموعه داده‌های بزرگ استخراج شده‌اند.

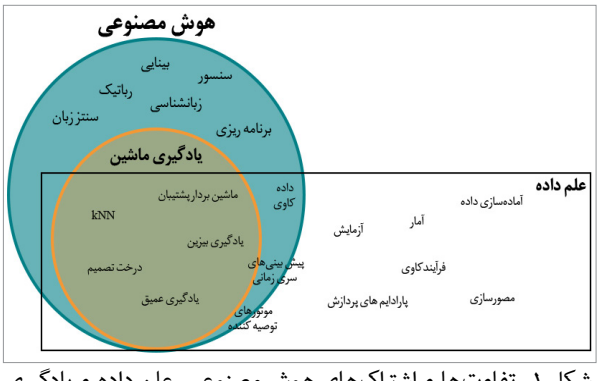
بر این اساس، علوم داده حوزه نسبتاً جدیدی است که بسیاری از صنایع از جمله کشاورزی را متحول می‌کند. کشاورزی دقیق که اکنون کشاورزی دیجیتال نیز نامیده می‌شود، به عنوان زمینه‌های علمی جدیدی مطرح شده است که از رویکردهای مبتنی بر داده‌ها برای افزایش بهره‌وری کشاورزی و در عین حال به حداقل رساندن اثرات محیط‌زیستی آن استفاده می‌کنند. داده‌های تولید شده در کشاورزی مدرن که توسط انواع حسگرهای مختلف فراهم و تولید می‌شوند سبب درک بهتر محیط عملیاتی (تعامل پویا با محصول، خاک و شرایط آب و هوایی) می‌شود و در نهایت، منجر به تصمیم‌گیری دقیق‌تر و سریع‌تر می‌شود. با این وجود مدیریت آب در کشاورزی به تلاش‌های قابل توجهی نیاز دارد و این موضوع نقش مهمی در تعادل هیدرولوژیکی، اقلیمی و کشاورزی ایفا می‌کند (Liakos و همکاران، ۲۰۱۸). به طور کلی هدف از این پژوهش، تعریف دقیق علم داده و سایر اصطلاحات مشتق شده از آن و یا مشابه آن مانند هوش مصنوعی، یادگیری ماشین و یادگیری عمیق در مدیریت آب کشاورزی است. معمولاً این تعاریف به اشتباه به جای همدیگر استفاده می‌شوند. در این پژوهش همچنین ابزارهای کلی مورد استفاده در علم داده معرفی و مورد بحث قرار گرفته و در انتها مطالعات انجام شده نیز در این حوزه بررسی خواهد شد.

مروری بر تعاریف هوش مصنوعی[5]، علم داده، یادگیری ماشین و یادگیری عمیق[6]

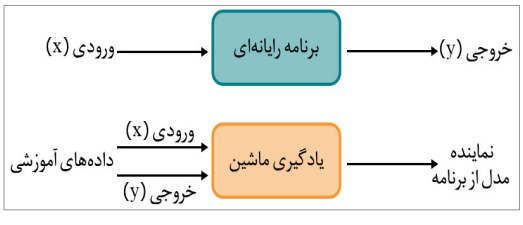
امروزه در کنار واژه علم داده، معمولاً واژه‌های هوش مصنوعی، یادگیری ماشین و یادگیری عمیق نیز به کار برده می‌شود. در شکل (۱) وجوه تمایز و اشتراک هوش مصنوعی، علم داده، یادگیری ماشین و یادگیری عمیق نشان داده شده است. هوش مصنوعی در واقع اصطلاحی است برای توصیف عواملی که به طور مستقل عمل می‌کنند. در برخی موارد عوامل هوش مصنوعی روبات‌ها و در برخی دیگر برنامه‌های کاربردی نرم‌افزاری هستند (Pierson، ۲۰۲۱). به طور کلی در تعریف هوش مصنوعی می‌توان گفت هر عملی که به ماشین قابلیت تقلید از رفتار انسانی به خصوص عملکردهای شناختی را بدهد، در حیطه هوش مصنوعی قرار خواهد گرفت. به عنوان مثال تشخیص چهره، زبان‌شناسی، رانندگی خودکار و سایر مواردی که در شکل (۱) مشاهده می‌شود در زمره هوش مصنوعی قرار می‌گیرند (Kotu و Deshpande، ۲۰۱۸). یادگیری ماشین که زیرمجموعه هوش مصنوعی است، در واقع ابزارهایی هستند که از آن‌ها برای انتقال تجربیات انسانی به ماشین استفاده می‌شوند. این تجربیات در واقع همان داده‌ها هستند. داده‌هایی که برای یادگیری ماشین استفاده می‌شوند اصطلاحاً «داده‌های آموزشی[7]» گفته می‌شوند.

یادگیری ماشین مدل برنامه‌نویسی سنتی را وارونه می‌کند (شکل ۲). در یک برنامه رایانه‌ای، مجموعه‌ای از قوانین و روابط به رایانه داده شده و سپس رایانه داده‌های ورودی را با استفاده از روابط از پیش تعیین شده به خروجی تبدیل می‌کند. اما الگوریتم‌های

یادگیری ماشین که یادگیرندگان نیز نامیده می‌شوند، هم ورودی و هم خروجی شناخته شده (داده‌های آموزشی) را گرفته تا مدلی را برای برنامه‌ای که ورودی را به خروجی تبدیل می کند، پیدا کنند (Kotu و Deshpande، ۲۰۱۸).



شکل ۱- تفاوت‌ها و اشتراک‌های هوش مصنوعی، علم داده و یادگیری ماشین (Kotu و Deshpande، ۲۰۱۸)



شکل ۲- تفاوت برنامه‌های سنتی رایانه‌ای در مقابل یادگیری ماشین (Kotu و Deshpande، ۲۰۱۸)

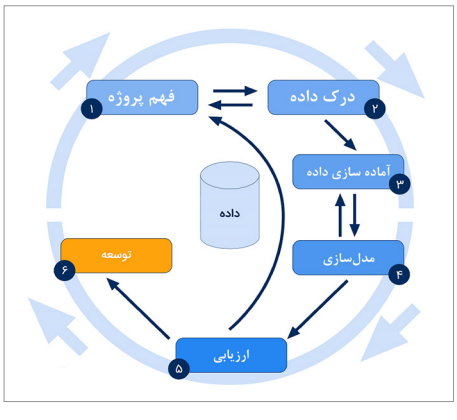
در ده سال گذشته پیشرفت‌های عمده‌ای در یادگیری ماشین، به‌ویژه یادگیری عمیق، ایجاد شد که در نحوه پردازش رایانه‌ها به منظور پردازش داده‌های زبان و تصویر، تغییرات اساسی را ایجاد کرد. اصطلاح یادگیری عمیق خانواده‌ای از مدل‌های شبکه عصبی را توصیف می‌کند که از لایه‌های متعددی از واحدهای موجود در شبکه تشکیل شدند. شبکه‌های عصبی از دهه ۱۹۴۰ وجود داشتند؛ اما این شبکه‌ها با مجموعه داده‌های بزرگ و پیچیده بهترین عملکرد را داشتند و از منابع محاسباتی زیادی برای آموزش استفاده می‌کنند. بنابراین، ظهور یادگیری عمیق با رشد کلان داده[۸] و قدرت محاسبات مرتبط است. با این تفاسیر می‌توان گفت، یادگیری عمیق، در زیرمجموعه یادگیری ماشین قرار گرفته و به عنوان یکی از ابزارهای خارق‌العاده آن شناخته می‌شود (Kelleher و Tierney، ۲۰۱۸). حال آن‌که علم داده، براساس آن‌چه که پیش‌تر تعریف شد به مجموعه‌ی روش‌ها و فعالیت‌هایی گفته می‌شود که بر روی داده‌ها انجام می‌شود. در علوم داده، تحلیل اولیه داده‌ها، پیش‌پردازش داده‌ها و تفسیر و نمایش داده‌ها وجود دارد. واکشی داده‌ها و ذخیره‌سازی آن‌ها نیز جزو فعالیت‌های علوم داده قرار می‌گیرد. در تعریف دیگر برای علم داده آمده است که علم داده کاربرد تجاری یادگیری ماشین، هوش مصنوعی و سایر زمینه‌های کمی مانند آمار، مصورسازی و ریاضیات است. در حقیقت علم داده یک زمینه بین رشته‌ای است که ارزش را از داده‌ها استخراج می‌کند. علم داده امروزه، به شدت به یادگیری ماشین متکی است (Kotu و Deshpande، ۲۰۱۸).

روش‌شناسی علم داده

• تهیه و آماده‌سازی داده

علم داده اساساً به داده وابسته است. یک مجموعه داده از داده‌های مربوط به مجموعه‌ای از موجودیت‌ها[۹] شکل می‌گیرد که هر موجودیت بر اساس مجموعه‌ای از ویژگی‌ها[۱۰] توصیف می‌شود. یک مجموعه داده در ابتدایی‌ترین شکل در یک ماتریس n×m به نام رکورد تحلیل‌ها[۱۱] (مجموعه داده‌ها[۱]) سازمان یافته است که در آن n تعداد موجودیت‌ها (سطرها) و m تعداد ویژگی‌ها (ستون‌ها) هستند. ساخت رکورد تحلیلی پیش نیاز انجام علم داده است. در واقع بیشترین زمان و تلاش در پروژه‌های علم داده صرف ایجاد، متناسب کردن و به‌روزرسانی رکورد تحلیلی می‌شود (Kelleher و Tierney، ۲۰۱۸). در پروژه‌های علم داده در این زمینه که چه ویژگی‌هایی را می‌توان انتخاب کرد و چه ویژگی‌هایی بر اساس دانش متخصص انتخاب می‌شوند، محدودیت‌هایی وجود دارد. ویژگی‌ها انواع مختلفی داشته و برای هر نوع ویژگی تجزیه و تحلیل مختلفی وجود دارد. بنابراین، آگاهی و شناخت انواع مختلف ویژگی، مهارت اصلی دانشمند داده است. ویژگی‌های عددی[۱۳]، اسمی[۱۴] و ترتیبی[۱۵] انواع استاندارد ویژگی هستند (Kelleher و Tierney، ۲۰۱۸). ویژگی‌های عددی مقادیر قابل اندازه‌گیری را توصیف می‌کنند. به عنوان مثال، عملکرد گیاه، مقدار آب آبیاری، مقدار کود و موارد این چنینی که با مقادیر صحیح یا واقعی نشان داده می‌شوند. ویژگی‌های اسمی (که به دسته‌ای نیز معروف است) مقادیر را از مجموعه متناهی می‌گیرد. این مقادیر برای دسته‌ها، کلاس‌ها یا حالت‌های اسامی هستند. به عنوان مثال نوع سامانه آبیاری، نوع کود و نام منطقه مزرعه در این دسته از ویژگی‌ها قرار می‌گیرند. ویژگی‌های ترتیبی مشابه ویژگی‌های اسمی هستند با این تفاوت که اعمال ترتیب روی دسته‌های ترتیبی اعمال می‌شود. به عنوان مثال نظر کارشناس در مورد عملکرد یک مجموعه زراعی می‌تواند به صورت خوب، بد و متوسط باشد. داده‌ها نیز به شکل دیگری دسته‌بندی می‌شوند. داده‌ها می‌توانند ساختاریافته، نیمه ساختاریافته و بدون ساختار باشند. داده‌های ساختاریافته، که تنها ۵ درصد از کل داده‌های موجود را تشکیل می‌دهد، به داده‌های جدولی موجود در صفحات گسترده یا پایگاه داده‌های رابطه‌ای اشاره دارد. متن، تصاویر، صدا و ویدئو نمونه‌هایی از داده‌های بدون ساختار هستند که گاهی فاقد سازماندهی ساختاری مورد نیاز ماشین‌ها برای تجزیه و تحلیل هستند. داده‌های نیمه ساختاریافته

زنجیره‌ای بین داده‌های کاملاً ساختاریافته و بدون ساختار را در بر می‌گیرد. زبان نشانه‌گذاری توسعه‌پذیر (XML[16])، یک زبان متنی برای تبادل داده‌ها در وب، یک نمونه معمولی از داده‌های نیمه ساختاریافته است. اسناد XML حاوی تگ‌های داده تعریف شده توسط کاربر هستند که آنها را برای ماشین قابل خواندن می‌کند. این تفاوت‌ها باعث شده است تا انتخاب بهترین و بهینه‌ترین روش‌های بررسی داده‌ها بسته به نوع پروژه دشوار باشد. به همین منظور متخصصین حوزه علم داده برای رسیدن به نتیجه نهایی، شیوه‌نامه‌های مختلفی را ارائه کرده‌اند که معروف‌ترین آن فرآیند کریسپ[17] است (Chapman، ۱۹۹۹). فرآیند کریسپ در ابتدا توسط کنسرسیومی از سازمان‌ها متشکل از فروشندگان برجسته علم داده، شرکت‌های مشاور و محققان در کارگاهی با حمایت مالی کمیسیونی اروپایی تحت برنامه ESPRIT ارائه شد. از آن زمان تاکنون اقدامات متعددی برای به‌روزرسانی این فرآیند انجام شد؛ با این حال هنوز هم غالباً از نسخه اصلی استفاده می‌شود. این فرآیند شامل شش مرحله است که تحت عنوان چرخه عمر کریسپ شناخته می‌شود. این چرخه عمر شامل فهم پروژه، درک و آماده‌سازی داده‌ها، مدل‌سازی، ارزیابی و توسعه است (شکل ۳).

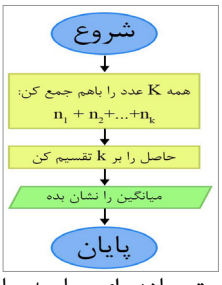


شکل ۳- چرخه عمر کریسپ (Chapman، ۱۹۹۹)

در این چرخه، داده‌ها در مرکز فعالیت‌های علم داده قرار دارند. پیکان‌های بین مراحل مختلف جهت معمول این فرآیند را نشان می‌دهند. این فرآیند نیمه ساختاریافته است یعنی دانشمند داده بسته به نتیجه یک مرحله خاص می‌تواند به یکی از مراحل قبلی برگشته، مرحله فعلی را تکرار یا به مرحله بعدی حرکت کند. در فهم پروژه، دانشمند داده بایستی پروژه‌ای که می‌خواهد بر روی آن کار کند را به خوبی بشناسد. در این مرحله، بایستی زوایای مختلف آن پروژه، محدودیت‌ها، شرایط موجود و اهداف آن را بررسی نماید. در مرحله درک داده‌ها، متخصص علم داده، به سراغ داده‌های موجود رفته و آن را به لحاظ کمیت و کیفیت برای شروع پروژه بررسی می‌کند. در این مرحله، عملیاتی مانند خلاصه‌سازی داده‌ها و ساخت گزارش‌های اولیه از داده‌ها می‌تواند بسیار کمک‌کننده باشد. با فهم داده‌ها و درک ابعاد و ویژگی‌های مختلف آن، می‌توان ایده‌های مختلف را مطرح کرد و ساختار اصلی پروژه را تعیین نمود. در مرحله آماده‌سازی داده‌ها، هدف ایجاد مجموعه داده‌ای است که برای تجزیه و تحلیل داده‌ها استفاده می‌شود. معمولاً ایجاد مجموعه داده شامل ادغام منابع داده از تعدادی پایگاه داده است. بعد از این که مجموعه داده‌ها ایجاد شد، کیفیت داده‌ها باید بررسی شود. مشکلاتی که معمولاً در کیفیت داده‌ها وجود دارند شامل مقادیر پرت و مفقود شده است. از آن‌جایی که خطاهای موجود در داده‌ها بر عملکرد الگوریتم‌ها تأثیرگذار هستند، بررسی کیفیت داده‌ها بسیار مهم است. معمولاً دانشمند داده در مرحله مدل‌سازی از تعدادی الگوریتم مختلف یادگیری ماشین برای آموزش تعدادی از مدل‌های مختلف در مجموعه داده استفاده می‌کند. هر مدل با نوع خاصی از الگوریتم‌های یادگیری ماشین آموزش دیده و هر الگوریتم در جستجوی انواع مختلفی از الگوها در داده‌ها است. معمولاً دانشمند داده در این مرحله نمی‌داند که بهترین الگوها برای جستجوی داده‌ها کدامند، بنابراین منطقی است که الگوریتم‌های مختلفی را در این زمینه امتحان نماید تا متوجه شود کدام الگوریتم هنگام اجرا روی داده‌ها بیشترین بازده را دارد. در ادامه الگوریتم‌های مهم در عرصه علم داده و یادگیری ماشین معرفی خواهند شد. انجام یک بررسی کلی به منظور تضمین کیفیت اعمال پروژه در مرحله ارزیابی ضروری است. در این مرحله سوالاتی همچون: آیا چیزی جا مانده است؟ آیا می‌توان کار بهتری انجام داد؟ بررسی می‌شود و بر اساس ارزیابی کلی مدل‌ها، تصمیم اصلی که گرفته می‌شود این است که آیا هیچ یک از این مدل‌ها مناسب پروژه خواهند بود یا اینکه نیاز است به منظور ایجاد مدل‌های مناسب، فرآیند کریسپ مجدد تکرار شود. در نهایت، مرحله توسعه شامل بررسی نحوه توسعه مدل‌های انتخاب شده برای رسیدن به یک نرم‌افزار است. این نرم‌افزار به کاربران معمولی نیز این امکان را می‌دهد تا با استفاده از آن، بتوانند مشکلی را که به دنبال آن هستند، بررسی و رفع نمایند (Kelleher و Tierney، ۲۰۱۸).

• معرفی الگوریتم‌های پرکاربرد یادگیری ماشین در حوزه علوم و مهندسی آب

در یادگیری ماشین، الگوریتم‌هایی ایجاد می‌شود که رایانه به منظور شناسایی و استخراج الگوها از داده‌ها، آن‌ها را درک می‌کند. الگوریتم یک روش منطقی گام به گام برای حل یک مسئله است. در واقع، چگونگی حل یک مشکل خاص با استفاده از داده‌ها مشخص می‌شود. بسیاری از الگوریتم‌های یادگیری بازگشتی هستند؛ یعنی در آن مجموعه‌ای از مراحل چندین بار تکرار می‌شوند تا زمانی که یک شرط محدودکننده برآورده شود. هر الگوریتم دارای ورودی، خروجی و دستورالعمل‌های قطعی انجام‌پذیر است. به عنوان مثال، در شکل (۴) یک الگوریتم ساده برای محاسبه میانگین K عدد نمایش داده شده است.

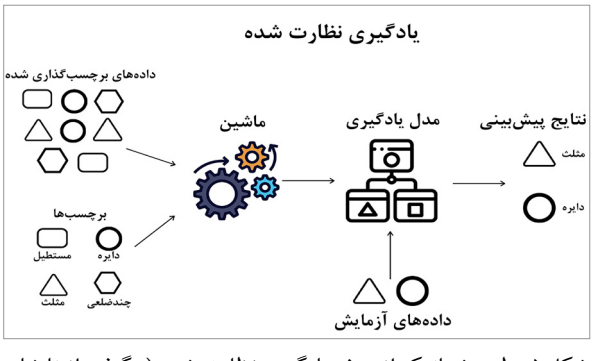


شکل ۴- نمونه یک الگوریتم ساده برای محاسبه میانگین (نیکخواه،۱۳۹۷)

در ادامه الگوریتم‌های پرکاربرد در علوم و مهندسی آب به صورت مختصر معرفی می‌شوند. در یک دسته‌بندی کلی، الگوریتم‌های یادگیری ماشین در سه دسته یادگیری نظارت شده[۱۸]، یادگیری بدون نظارت[۱۹] و یادگیری نیمه نظارت شده[۲۰] تقسیم‌بندی می‌شوند (کاویانی، ۱۴۰۱).

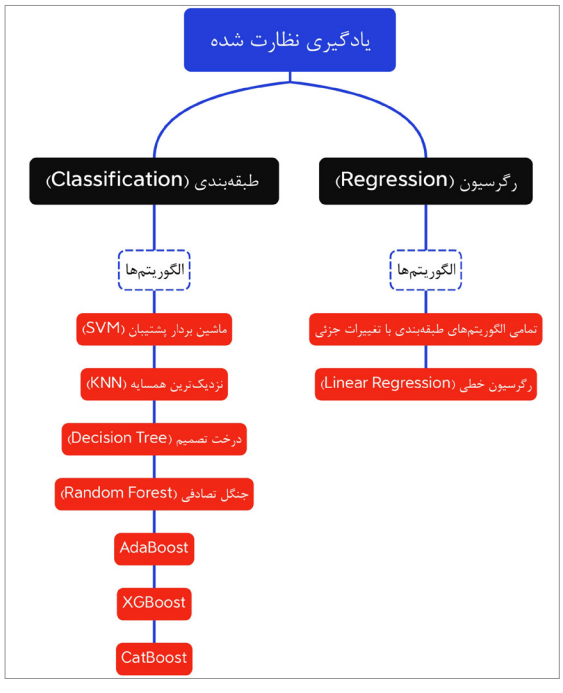
۱- یادگیری نظارت شده:

روش‌ها یا الگوریتم‌های نظارت شده، دسته‌ای از الگوریتم‌های یادگیری ماشین هستند که از نمونه داده شناخته شده (که مجموعه داده‌های آموزشی نیز نامیده می‌شود) و خروجی مرتبط (که برچسب یا ستون هدف نیز معروفند) با هر نمونه داده، در فرآیند یادگیری مدل استفاده می‌کنند. هدف اصلی یادگیری، نگاشت یا کشف ارتباط بین نمونه داده‌های ورودی (x) با خروجی‌های متناظر آن‌ها (y) بر مبنای نمونه داده‌های آموزشی است. سپس، الگوریتم آموزش داده شده برای پیش‌بینی خروجی (y') از هر نمونه داده جدید (x')، که قبلاً در فرآیند آموزش مدل دیده نشده و ناشناس است، به کار می‌رود (شکل ۵).



شکل ۵- طرح شماتیک از روش یادگیری نظارت شده (برگرفته از تارنمای ((Raj)lenjoyalgorithms.com، ۲۰۲۱))

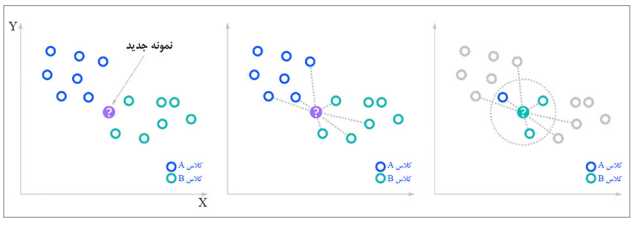
روش‌های یادگیری نظارت شده در دو دسته کلی طبقه‌بندی[۲۱] و رگرسیون[۲۲] قرار می‌گیرند (نیکخواه، ۱۳۹۷). برای هر دو روش ذکر شده، الگوریتم‌های مختلفی معرفی شده‌اند که از این بین می‌توان به الگوریتم‌های رگرسیون خطی، ماشین÷÷ بردار پشتیبان، درخت تصمیم و جنگل‌های تصادفی اشاره کرد (شکل ۶). با توجه به اهمیت این الگوریتم‌ها در کشاورزی (به علت وجود داده‌های برچسب‌گذاری شده در این حوزه) در ادامه به معرفی اجمالی برخی از الگوریتم‌های پرکاربرد پرداخته می‌شود.



شکل۶- الگوریتم‌های مختلف در روش یادگیری نظارت شده (کاویانی،۱۴۰۱)

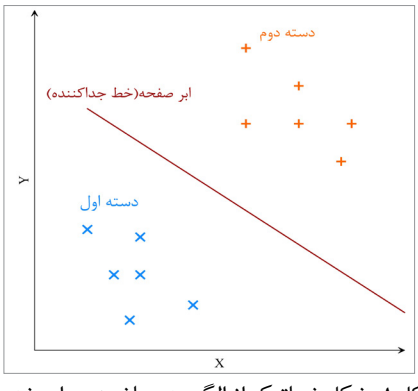
- **الگوریتم رگرسیون خطی:** وقتی مجموعه داده‌ای از ویژگی‌های عددی تشکیل شده باشد، غالباً از مدل‌های پیش‌بینی مبتنی بر رگرسیون استفاده می‌شود. اولین مرحله در تحلیل رگرسیون، فرضیه‌پردازی ساختار رابطه بین ویژگی‌های ورودی و هدف است. سپس مدل ریاضی پارامتری[۲۳] از رابطه فرضی تعریف می‌شود. به این مدل پارامتری شده تابع رگرسیون می‌گویند. می‌توان تابع رگرسیون را به عنوان ماشینی که ورودی‌ها را به مقدار خروجی تبدیل می‌کند و پارامترها را به عنوان تنظیماتی که رفتار این ماشین را کنترل می‌کند در نظر گرفت (Kelleher و Tierney، ۲۰۱۸). ممکن است با استفاده از تحلیل رگرسیون بتوان انواع بسیار متفاوتی از روابط بین ویژگی‌ها را فرضیه‌پردازی و مدل‌سازی کرد. در برخی حوزه‌ها، دلایل علمی فراوانی برای اثبات نوع خاصی از رابطه بین متغیرها وجود دارد؛ اما در غیاب این دلایل علمی، بهتر است که با فرض ساده‌ترین شکل رابطه یعنی "رابطه خطی" شروع کرده و سپس در صورت لزوم روابط پیچیده‌تر مدل‌سازی شود. دلیل دیگر برای شروع با روابط خطی این است که تفسیر نسبتاً راحتی دارند (Kotu و Deshpande، ۲۰۱۸). هنگامی که رابطه‌ای خطی فرض شود به تحلیل رگرسیون، رگرسیون خطی گفته می‌شود. ساده‌ترین کاربرد رگرسیون خطی، مدل‌سازی رابطه بین دو ویژگی ورودی (X) و ویژگی هدف (Y) است. اما همیشه مسائل صرفاً دارای یک متغیر وابسته (Y) و یک متغیر مستقل (X) نخواهند بود؛ بلکه ممکن است چندین متغیر مستقل به صورت همزمان بر روی متغیر وابسته اثر بگذارند. در این حالت باید رگرسیون خطی چندگانه[۲۴] استفاده شود (Kotu و Deshpande، ۲۰۱۸).

- **الگوریتم نزدیک‌ترین همسایه (KNN[۲۵]):** الگوریتم طبقه‌بندی نزدیک‌ترین همسایه، برای پیدا کردن برچسب داده‌ی جدید، فاصله نمونه جدید را با تمامی نمونه‌های قبلی از مجموعه داده‌ی آموزشی مقایسه کرده، نزدیک‌ترین نمونه‌ها را انتخاب و برچسبِ آن‌ها را بر روی نمونه‌های جدید می‌گذارد. این کار می‌تواند با استفاده از یک معیار فاصله (مانند فاصله اقلیدسی[۲۶]، منهتن[۲۷]، مینکووسکی[۲۸] و هَمینگ[۲۹]) انجام شود (کاویانی، ۱۴۰۱). این الگوریتم با توجه به پارامتری تحت عنوان K، تعداد K عدد از نزدیک‌ترین نمونه‌ها را از مجموعه آموزشی (مجموعه‌ای که قبلاً توسط متخصص برچسب‌گذاری شده است) با توجه به فاصله اقلیدسی (رابطه ۴) آن‌ها از نمونه جدید در نظر گرفته و این نمونه بر اساس آن‌ها برچسب خواهد خورد (شکل ۷). دقت و صحت بالا، از نقاط قابل توجه در این الگوریتم است. در مورد نقاط ضعف آن نیز می‌توان به محاسبات زیاد و سرعت پایین اشاره کرد.



شکل ۷- نمایی از الگوریتم نزدیک‌ترین همسایه‌ها (برگرفته از وبسایت ibm.com)

- **ماشین بردار پشتیبان (SVM[۳۰]):** یکی از مشکلات الگوریتم نزدیک‌ترین همسایه، قرار گرفتن آن در دسته الگوریتم‌های تنبل[۳۱] است (Raschka، ۲۰۱۸). در واقع این الگوریتم، داده‌های مجموعه‌ی آموزشی را (بعد از پیش پردازش) بدون تغییر ذخیره و نمونه جدید را با تمامی نمونه‌های موجود در مجموعه‌ی داده، مقایسه می‌کند. این کار باعث حجم بالای محاسبات و به تبع آن کاهش سرعت الگوریتم می‌شود. اما الگوریتم ماشین بردار پشتیبان با استفاده از روش ایجاد خط یا ابرصفحه[۳۲] این مشکل را حل می‌کند (شکل ۸). این الگوریتم فضا را به بخش‌های منفک تقسیم‌بندی کرده تا سرعت پیش‌بینی افزایش یابد (کاویانی، ۱۴۰۱). در شکل (۸) مجموع داده‌ها دو ویژگی (متغیر) با نام‌های X و Y دارند. دو نوع برچسب، کلاس و یا دسته نیز وجود دارد که با رنگ‌های متفاوت مشخص شده‌اند؛ الگوریتم ماشین بردار پشتیبان با استفاده از یک خط متمایزکننده، دو نوع دسته را از یکدیگر جدا کرده است، به گونه‌ای که هر نمونه پایین این خط جداکننده برچسب دسته اول و هر نمونه بالای این خط، برچسب دسته دوم را خواهد گرفت. اگر تعداد ویژگی‌ها بیشتر از دو مورد باشد، آنگاه دیگر نمی‌توان با یک خط آن‌ها را از یکدیگر تفکیک کرد بلکه باید از صفحه استفاده شود. اگر تعداد ویژگی‌ها (متغیرها) بیشتر از سه ویژگی باشد، بایستی از اَبَرصفحه برای جداسازی نمونه‌ها استفاده شود. ابرصفحه در واقع همان خط یا صفحه در فضاهایی با n بُعد است (کاویانی، ۱۴۰۱). الگوریتم SVM سعی می‌کند خط، صفحه یا ابرصفحه را به گونه‌ای رسم کند که بیشترین فاصله را با مرز نمونه‌ها داشته باشد. به این کار در اصطلاح بیشنه‌سازی حاشیه[۳۳] در الگوریتم می‌گویند.

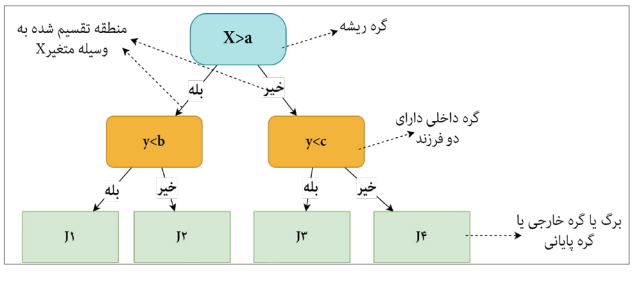


شکل ۸- شکل شماتیک از الگوریتم ماشین بردار پشتیبان

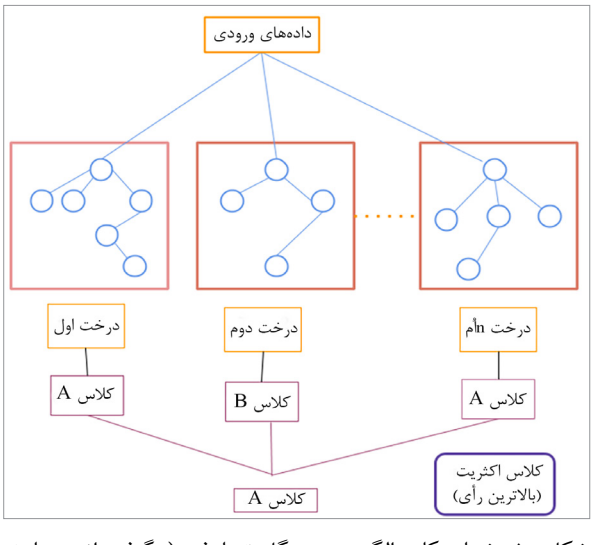
- **الگوریتم درخت تصمیم:** الگوریتم درخت تصمیم یکی از الگوریتم‌های ساده و پر کاربرد در داده کاوی و یادگیری ماشین است. پیش درآمد اکثر الگوریتم‌های مدرن برای یادگیری درخت تصمیم، الگوریتم ID3[۳۴] است (Quinlan، ۱۹۸۶). ID3 یک درخت تصمیم را به صورت بازگشتی و عمقی ایجاد می‌کند و هر بار یک گره را اضافه می‌کند. درخت تصمیم با گره ریشه شروع می‌شود. سپس با انتخاب یک ویژگی برای آزمایش در گره ریشه، فرآیند رشد درخت از ریشه شروع می‌شود (Kelleher و Tierney، ۲۰۱۸). این الگوریتم در واقع از نوعی if/else برای ایجاد یک ساختار درختی استفاده می‌کند (کاویانی، ۱۴۰۱). در واقع یک درخت تصمیم ساختاری درختی است که از گره‌های داخلی و گره‌های خارجی تشکیل شده است. این گره‌ها به وسیله شاخه‌ها به یکدیگر متصل می‌شوند و مجموعه ورودی را به مناطق اختصاصی و متناقض تقسیم می‌کند. به هر یک از این مناطق یک برچسب، یک مقدار یا یک عمل اختصاص داده می‌شود. گره‌های داخلی که به عنوان واحد تصمیم‌گیری شناخته می‌شوند، یک تابع تصمیم را ارزیابی می‌کنند تا تصمیم بگیرند کدام گره فرزند بعدی را باید بیاورند. گره‌های انتهایی که با برچسب‌ها (مثلاً رویدادهای آبیاری یا غیرآبیاری) مشخص می‌شوند، به عنوان برگ‌ها یا گره‌های پایانی (گره‌های خارجی) شناخته می‌شوند و فرزندی ندارند (Perea و همکاران، ۲۰۱۹). در شکل (۹) نمونه یک ساختار کلاسیک از درخت تصمیم آورده شده است. در این شکل J1، J2، J3 و J4 در حقیقت همان برچسب‌های نهایی برای طبقه‌بندی نمونه‌ها هستند که به عنوان برگ نیز شناخته می‌شوند.

- **الگوریتم جنگل‌های تصادفی:** جنگل مجموعه‌ای از درخت‌ها است. الگوریتم جنگل تصادفی نیز از مجموعه‌ای از درخت‌های تصمیم برای ساخت مدل روی مجموعه‌ی داده‌ها و پیش‌بینی

نمونه‌های جدید استفاده می‌کند. شمای کلی الگوریتم جنگل تصادفی در شکل (۱۰) آورده شده است. الگوریتم جنگل تصادفی ابتدا در فاز یادگیری، چندین درخت از روی مجموعه‌ی داده‌های آموزشی ایجاد می‌کند. این الگوریتم به هر کدام از درخت‌ها یک زیرمجموعه از داده‌های ورودی در هنگام ساخت تزریق می‌کند. در واقع، هر کدام از درخت‌ها با مشاهده‌ی زیرمجموعه‌ای از ویژگی‌ها (متغیرها) در مجموعه آموزشی، یادگیری را انجام می‌دهند (به این صورت معمولاً دو درخت در جنگل تصادفی شبیه به هم نمی‌شوند) و الگوریتم جنگل تصادفی این درخت‌های تصمیم را به عنوان مدل ذخیره می‌کند. هنگامی که نمونه جدید برای طبقه‌بندی از راه می‌رسد، جنگل تصادفی این نمونه را به هر کدام از درخت‌ها داده و درخت‌ها هر کدام یک برچسب (براساس یادگیری که انجام داده‌اند) برمی‌گردانند. سپس الگوریتم، برچسب یا کلاس اکثریت را به عنوان برچسب نهایی برای آن نمونه‌ی جدید انتخاب می‌کند. الگوریتم درخت تصادفی نسبت به الگوریتم درخت تصمیم، معمولاً دقت و صحت بالاتری در مسائل مختلف تولید می‌کند. اگر چه سرعت این الگوریتم به واسطه‌ی درخت‌های متعدد پایین است و یادگیری و ساخت مدل معمولاً فضای زیادی نسبت به درخت‌های تصمیم اشغال می‌کند (کاویانی، ۱۴۰۱).



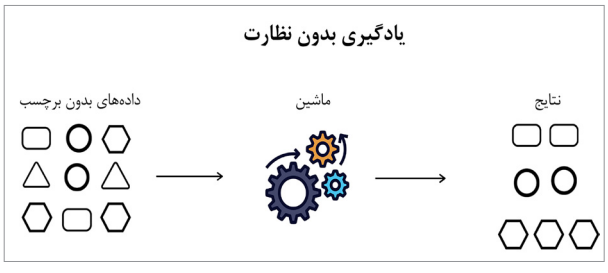
شکل ۹- ساختار کلاسیک یک درخت (Perea و همکاران، ۲۰۱۹)



شکل ۱۰- شمای کلی الگوریتم جنگل تصادفی (برگرفته از وبسایت analyticssteps.com)

۲- یادگیری بدون نظارت

در یادگیری بدون نظارت مجموعه داده‌های ورودی بدون قالب‌های برچسب‌دار هستند. این روش‌ها بدون نظارت نامیده می‌شوند؛ زیرا مدل یا الگوریتم تلاش می‌کند ساختارها، الگوها و روابط ذاتی را از داده‌های مفروض بدون هیچ گونه کمک و نظارتی یاد بگیرد. یادگیری بدون نظارت به جای پیش‌بینی برخی نتایج بر اساس داده‌های آموزشی، تلاش می‌کند تا بینش یا اطلاعات معنادار از داده‌ها استخراج کند (شکل ۱۱).

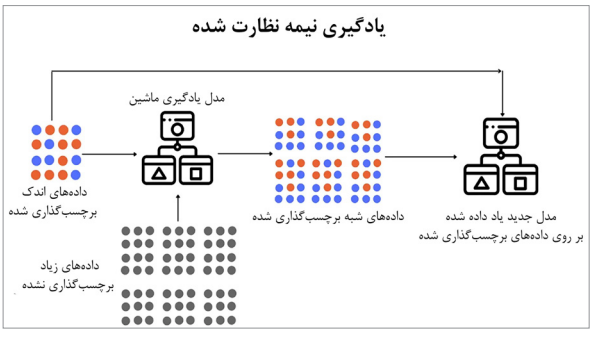


شکل ۱۱- طرح شماتیک از روش یادگیری بدون نظارت (برگرفته از تارنمای enjoyalgorithms.com (Raj)، ۲۰۲۱))

الگوریتم‌های خوشه‌بندی[۳۵]، کاهش بعد[۳۶] و کشف ناهنجاری[۳۷] از روش‌های یادگیری بدون نظارت هستند (نیکخواه، ۱۳۹۷). الگوریتم‌های خوشه‌بندی شامل K میانگین[۳۸]، خوشه‌بندی مبتنی بر غلظت[۳۹] و الگوریتم تغییر میانگین[۴۰] می‌شوند (کاویانی، ۱۴۰۱).

۳- یادگیری نیمه نظارتی

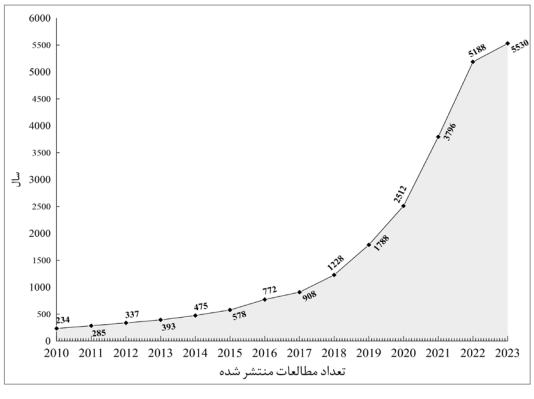
یادگیری نیمه‌نظارتی نوعی یادگیری ماشین است که در آن الگوریتم براساس ترکیبی از داده‌های برچسب‌دار و بدون برچسب آموزش داده می‌شود. به‌طور معمول، این ترکیب مقدار بسیار کمی از داده‌های برچسب‌دار و مقدار بسیار زیادی از داده‌های بدون برچسب را شامل خواهد بود (نیکخواه، ۱۳۹۷؛ شکل ۱۲). روش یادگیری نیمه نظارتی به عنوان یکی از مدل‌های یادگیری ماشین، راهکاری مؤثر برای رفع چالش‌های مربوط به داده‌های برچسب‌دار می‌باشد. در یادگیری نیمه نظارتی به طور همزمان از داده‌های برچسب خورده و از داده‌های برچسب نخورده استفاده می‌شود تا بتوان دقت یادگیری را بهبود بخشید. این نوع از مدل یادگیری به عنوان روشی که می‌تواند از داده‌های بدون برچسب بیشترین استفاده را بکند، از منظر کاربرد عملی ارزش فوق‌العاده‌ای دارد. در این حوزه روش‌های متعددی در قالب روش‌های مولد[۴۱]، روش‌های مبتنی بر نمودار[۴۲] و روش‌های اکتشاف محور[۴۳] در دسترس هستند (نیکخواه، ۱۳۹۷).



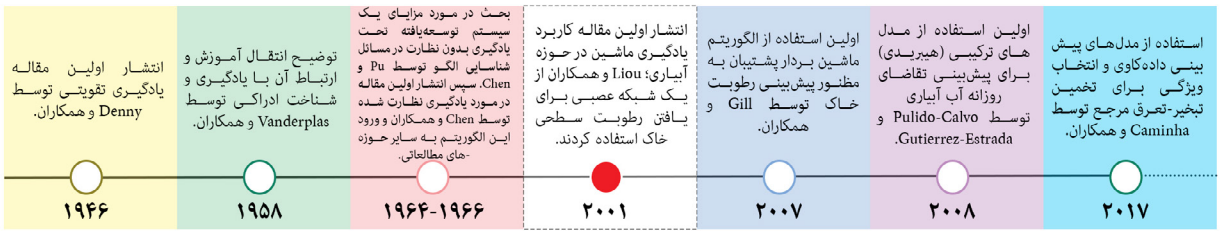
شکل ۱۲- طرح شماتیک از روش یادگیری نیمه نظارت شده (برگرفته از تارنمای enjoyalgorithms.com (Raj، ۲۰۲۱))

کاربردهای علم داده در علوم و مهندسی آب

نتایج جستجو در پایگاه داده‌های مطالعاتی مختلف حاکی از رشد چشمگیر انتشار مطالعات در این حوزه است که نشان‌دهنده اهمیت موضوع در جامعه جهانی است. به عنوان مثال شکل (۱۳)، نتیجه جستجو با سه کلید واژه "Water"، "Machine learning" و "Agriculture" در پایگاه ساینس‌دایرکت[۴۴] را نشان می‌دهد. همان‌گونه که مشاهده می‌شود تعداد مطالعات مرتبط با این کلیدواژه از ۲۳۴ عدد در سال ۲۰۱۰ به تعداد ۵۵۳۰ عدد در سال ۲۰۲۳ رسیده است. در پژوهش دیگری که Gao و همکاران (۲۰۲۳) با دو کلیدواژه "machine learning" و "irrigation" بر روی پایگاه داده وب‌آف‌ساینس[۴۵] انجام دادند، نشان داد که مطالعات در این زمینه به طور میانگین ۴۴/۶۷ درصد در سال افزایش می‌یابد. نتایج این پژوهش رشد سریع این مطالعات از سال ۲۰۱۹ به بعد را نشان داد که این موضوع در پایگاه داده ساینس‌دایرکت[۴۶] نیز مشاهده می‌شود (شکل ۱۳). در شکل (۱۴) سیر تکاملی الگوریتم‌های یادگیری ماشین در زمینه آبیاری آورده شده است. در این روند مشاهده می‌شود که تحقیقات یادگیری ماشین در زمینه آبیاری از پیش‌بینی رطوبت خاک سرچشمه می‌گیرد. با افزایش تحقیقات در این زمینه، مدل‌ها به تدریج از یک مدل منفرد به یک مدل ترکیبی در جهت و محتوای تحقیق عمیق‌تر شدند.



شکل ۱۳- مطالعات منتشر شده در زمینه استفاده از یادگیری ماشین در حوزه کشاورزی و آب طی ۱۳ سال گذشته (منبع: Sciencedirect.com)



شکل ۱۴- تاریخچه یادگیری ماشین در حوزه علوم آبیاری (Gao و همکاران، ۲۰۲۳)

نتیجه یک بررسی انجام شده بین کل مطالعات با محوریت یادگیری ماشین در کشاورزی بین سال‌های ۲۰۱۸ تا ۲۰۲۰ نشان داد که به طور کلی یادگیری ماشین در چهار حوزه اصلی شامل مدیریت گیاه، دام، آب و خاک مورد استفاده قرار می‌گیرد که بخش آب ۱۰ درصد این مطالعات را شامل می‌شوند (Benos و همکاران، ۲۰۲۱).

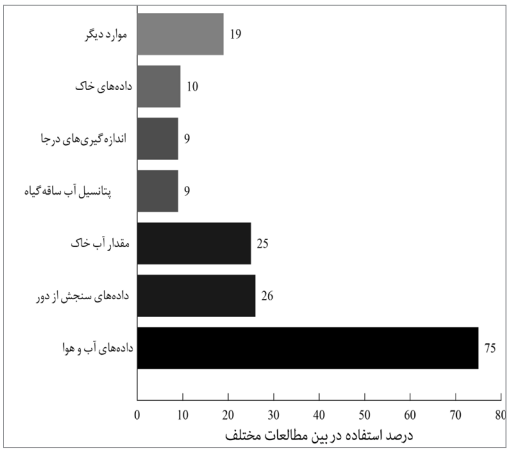
بالتبع در ایران نیز در سال‌های اخیر نظر پژوهشگران حوزه کشاورزی به این موضوعات معطوف شده است. بر اساس نتایج Benos و همکاران (۲۰۲۱)، ایران سهم ۵/۶۲ درصدی از مطالعات با محوریت استفاده از یادگیری ماشین در کشاورزی داشته است. در این بین کشور چین، با سهم ۲۴/۹ درصدی در جایگاه نخست و ایالت متحده آمریکا با سهم ۲۰/۷ درصدی در جایگاه دوم قرار گرفته است. نتایج فراتحلیل[۴۷] Ahmadi و همکاران (۲۰۲۲) نشان داد که ایران با سهم ۲۴ درصدی در بین مطالعات انجام شده در زمینه استفاده از یادگیری ماشین در پایش و پیش‌بینی سطح آب‌های زیرزمینی، در بین سایر کشورها پیشگام بوده است. بر اساس نتایج این پژوهش کشورهای با وابستگی بیشتر به آب زیرزمینی به عنوان یک منبع آب شیرین، اکثر مطالعات در مورد کاربرد یادگیری ماشین در مدل‌سازی آب زیرزمینی را تولید کردند.

در ادامه این نوشتار، موضوعاتی که بیشتر مورد توجه پژوهشگران در حوزه کاربرد علم داده در علوم و مهندسی آب بوده‌اند مورد بحث قرار گرفته است.

• مدیریت آب در کشاورزی

یکی از مهم‌ترین کاربردهای علم داده در کشاورزی، مدیریت آب است. با استفاده از علم داده و یادگیری ماشین می‌توان آبیاری را با نرخ متغیر به جای استفاده از نرخ یکنواخت برای کل مزرعه

انجام داد. به این معنا که هر نقطه از زمین کشاورزی بر اساس داده‌های مختلف نرخ متفاوتی از آب دریافت کند. اثربخشی و امکان‌سنجی روش آبیاری با نرخ متغیر به عوامل زراعی از جمله توپوگرافی، خواص خاک و تأثیر آنها بر آبِ خاک به منظور صرفه‌جویی در مصرف آب و بهینه‌سازی عملکرد بستگی دارد. پایش دقیق وضعیت آبِ خاک، شرایط رشد محصول و الگوهای زمانی و مکانی در ترکیب با پایش و پیش‌بینی شرایط آب و هوایی، می‌تواند به برنامه‌ریزی آبیاری و مدیریت کارآمد آب کمک کند (Benos و همکاران، ۲۰۲۱). در موضوع مدیریت آب، داده‌های آب و هوا رایج‌ترین داده‌های ورودی برای تخمین تبخیر-تعرق (ET) هستند. در بسیاری از موارد، برآورد دقیق تبخیر-تعرق یکی از محوری‌ترین عناصر چرخه هیدرولوژیکی برای مدیریت بهینه منابع آب است. داده‌های حسگرهای از راه دور و اندازه‌گیری مقدار آبِ خاک نیز به طور گسترده در این دسته، استفاده می‌شود. در شکل (۱۵) معمول‌ترین داده‌هایی که به عنوان ورودی الگوریتم‌های یادگیری ماشین در بحث مدیریت آب در بین مطالعات مختلف استفاده شده‌اند آورده شده است (Benos و همکاران، ۲۰۲۱).



شکل۱۵- درصد شکل‌های مختلف داده‌های ورودی به الگوریتم‌های یادگیری ماشین با هدف مدیریت آب در کشاورزی (Benos و همکاران، ۲۰۲۱)

• تعیین تبخیر-تعرق با استفاده از یادگیری ماشین

پیش‌بینی تبخیر-تعرق می تواند نقش مهمی در برنامه‌ریزی آبیاری و مدیریت منابع آب داشته باشد. تخمین دقیق تبخیر-تعرق یک فرآیند پیچیده است و برای مدیریت منابع در تولید محصول و همچنین برای طراحی و مدیریت عملیات سامانه‌های آبیاری اهمیت بالایی دارد. در اکثر مطالعات انجام شده در این زمینه، نتیجه یک معادله پایه‌ای تبخیر-تعرق مانند پنمن-مانتیث به عنوان ستون هدف در نظر گرفته شده و مقادیر ورودی الگوریتم، شامل شاخص‌های تأثیرگذار بر این معادله هستند. این شاخص‌ها به طور معمول شامل حداقل، حداکثر و میانگین دمای روزانه، رطوبت نسبی، مدت زمان تابش خورشید و سرعت باد است. تاکنون مطالعات متعددی دقت بالای الگوریتم‌های یادگیری ماشین را در برآورد تبخیر-تعرق گزارش کرده‌اند (Mehdizadeh و همکاران، ۲۰۱۷؛ Fan و همکاران، ۲۰۱۸؛ Yaseen و همکاران، ۲۰۲۰؛ Xing و همکاران، ۲۰۲۲). ویژگی‌های ورودی معمول به الگوریتم‌های یادگیری ماشین به منظور محاسبه تبخیر-تعرق شامل حداکثر، حداقل و میانگین دما، رطوبت نسبی، تابش خورشید و سرعت باد است. استفاده از یادگیری عمیق و شبکه‌های عصبی مصنوعی در این زمینه نسبت به سایر الگوریتم‌ها محبوبیت بیشتری دارد.

• تعیین پارامترهای کیفیت آب با یادگیری ماشین

استفاده از روش‌های معمولی برای پیش‌بینی کیفیت آب آبیاری، عموماً برای کشاورزان به ویژه در کشورهای در حال توسعه، پرهزینه و دشوار است (Jayaraman و همکاران، ۲۰۲۲). از طرف دیگر شکل‌های مختلف منابع آب شامل فاضلاب، آب زیرزمینی، آب سطحی، آب دریا و آب شیرین باعث شده است انواع مختلف آب، ویژگی‌های متفاوتی داشته باشند که خود باعث پیچیدگی بیشتر بررسی کیفیت آب خواهد شد (Maxwell، ۲۰۱۵). از این رو استفاده از مدل‌های مختلف یادگیری ماشین می‌تواند به اقتصاد و همچنین بهداشت کشاورزان کمک به‌سزایی کرده و فرآیند بررسی کیفیت آب را آسان‌تر و سریع‌تر کند. اخیراً مطالعاتی با استفاده از الگوریتم‌های مختلف یادگیری ماشین نیز به تعیین کیفیت آب و سایر جنبه‌های آن پرداختند (Haghiabi و همکاران، ۲۰۱۸؛ Pu و همکاران، ۲۰۱۹؛ Ly و همکاران، ۲۰۲۱؛ Zhi و همکاران، ۲۰۲۱). در این دست مطالعات معمولاً مجموعه‌ای از داده‌های از قبل تهیه شده شامل عوامل مختلف کیفیت آب به مدل داده شده و ماشین بر اساس آن‌ها آموزش داده می‌شود. در ستون برچسب [۴۸](هدف) به صورت متغیر، غلظت عناصر مختلف و یا شاخص‌های دیگر کیفیت آب قرار می‌گیرد. سپس هر کدام از این شاخص‌ها براساس سایر شاخص‌ها تعیین می‌شود.

• تشخیص نشتی و تلفات آب در کانال‌ها

از جمله مسائل دیگری که در حوزه استفاده از علوم داده در مدیریت آب بررسی می‌شود، تشخیص تلفات آب در سازه‌های انتقال آب است. وجود درز و ترک‌ها و حفره‌ها در کانال‌های انتقال آب باعث از دست رفتن آب در حین انتقال خواهد شد. این موضوع بر راندمان انتقال آب اثر منفی خواهد گذاشت. از طرف دیگر برای برنامه‌ریزی تعمیر و نگهداری زیرساخت‌های انتقال آب به بازرسی میدانی و متناوب نیاز است. این موضوع می‌تواند فرآیندی زمان‌بر و طاقت‌فرسا باشد. چرا که در عمل، کانال‌ها می‌توانند ده‌ها یا حتی صدها کیلومتر طول داشته باشند. این موضوع ممکن است باعث شود که بازرسی‌های میدانی کانال هفته‌ها طول بکشد. مشکل دیگر این است که بیشتر ترک‌ها در زیر آب ایجاد می‌شوند. بازرسان بدون

خشک‌کردن کانال‌ها قادر نخواهند بود شرایط زیر آب را مشاهده کنند. حتی بازرسان باتجربه نیز به زمان زیادی برای تکمیل یک ارزیابی قابل اعتماد از وضعیت کانال نیاز دارند. امروزه مطالعاتی با ترکیب تصاویر ماهواره‌ای و یادگیری ماشین توانسته‌اند نشتی در کانال‌ها را با دقت بسیار بالایی شناسایی کنند. Zanganeh و همکاران (۲۰۱۶) از روش‌های پردازش تصویر ماهواره‌ای لندست ۸ برای تعیین محل نشتی استفاده کردند. آن‌ها نشان دادند که روش‌های سنجش از دور با استفاده از تصاویر ماهواره‌ای با وضوح متوسط می‌توانند به تشخیص زودهنگام نشت‌ها دست یابند. ایشان از الگوریتم توسعه‌یافته از NDVI برای تجزیه و تحلیل توزیع پوشش گیاهی در طول کانال‌ها استفاده کردند. سپس از روش طبقه‌بندی K-means روی نقشه NDVI برای شناسایی نشتی‌ها استفاده شد. در مطالعه‌ای دیگر Chen و همکاران (۲۰۲۰) با استفاده از تصاویر ماهواره‌ای و یادگیری عمیق، نشتی آب در کانال‌ها را شناسایی کردند. آن‌ها با استفاده از تصاویر ماهواره‌ای و الگوریتم (Physics-guided Neural Networks (PGNN در یادگیری ماشین، توانستند با دقت ۸۶ درصد نشتی آب در کانال‌ها را شناسایی کنند.

• مدیریت و اجرای سامانه‌های آبیاری

ترکیب دو علم داده و اینترنت اشیاء[۴۹] می‌تواند از جمله راهکارهای مفید در زمینه افزایش بهره‌وری آب باشد. اینترنت اشیاء ترکیبی از چندین فناوری است که بوسیله آن حسگرها، عملگرها و سایر نقش¬های یک پروژه به صورت هوشمند به یکدیگر متصل می-شوند. در این روش، به طور معمول مجموعه‌ای از داده‌ها از حسگرهای مختلف مانند حسگر رطوبت خاک، دمای خاک، دما و رطوبت محیط و همچنین داده‌های ماهواره‌ای دریافت شده و در یک پایگاه داده ذخیره می‌شود. پس از ذخیره‌سازی، با استفاده از الگوریتم‌های یادگیری ماشین، زمان مناسب آبیاری تعیین شده و عملیات آبیاری نیز به صورت خودکار انجام خواهد شد. Singh و همکاران (۲۰۲۱) در پژوهشی یک سامانه آبیاری هوشمند با استفاده از روش‌های یادگیری ماشین پیاده‌سازی کردند. ایشان برای کنترل پمپ از یک حسگر رطوبت برای بررسی میزان رطوبت خاک و نیاز محصولات مختلف استفاده کردند. این حسگر مقادیر آنالوگ را به میکروکنترلر مورد استفاده در سامانه ارسال کرده و میکروکنترلر پس از مقایسه این مقادیر با مقدار آستانه، پمپ را روشن یا خاموش خواهد کرد. با استفاده از این سامانه، کشاورزان نیازی به آبیاری دستی در شرایط آب و هوایی نامناسب و شب‌ها نخواهند داشت. همچنین تلفات آب کاهش و به تبع آن بهره‌وری آب افزایش خواهد یافت.

• تشخیص نوع سامانه‌های آبیاری با استفاده از تصاویر ماهواره‌ای و یادگیری عمیق

دانستن نوع سامانه‌های آبیاری در هر منطقه به صورت یکپارچه می‌تواند در سطح کلان و فراحوضه‌ای باعث بهبود بهره‌وری آب با اتخاذ سیاست‌های درخور آن شود. از این رو داشتن اطلاعات و آمار بروز و دقیق از این سامانه‌ها اهمیت ویژه‌ای دارد. در این عرصه نیز با کمک ترکیب تصاویر ماهواره‌ای و یادگیری ماشین می‌توان اطلاعاتی از نوع سامانه‌های آبیاری موجود در هر منطقه به دست آورد. به عنوان مثال Saraiva و همکاران (۲۰۲۰) در پژوهشی به نقشه‌برداری خودکار سامانه‌های آبیاری دوار مرکزی[۵۰] با استفاده از تصاویر ماهواره‌ای و یادگیری عمیق پرداختند. نتایج آنان نشان داد مدل پیاده شده در این تحقیق دقت ۹۹ درصد و بازیابی ۸۸ درصد برای تشخیص و نقشه‌برداری سامانه‌های آبیاری دوار مرکزی (سنترپیوت) در منطقه مورد مطالعه (ایالت‌های گوییاس[۵۱] و مینا ژِرایس[۵۲] در برزیل) را دارا است.

• مدیریت و پایش گیاه و پیش‌بینی عملکرد

استفاده از روش‌های مختلف علم داده و یادگیری ماشین برای مدیریت محصولات شامل حیطه‌های متنوعی مانند پیش‌بینی عملکرد، تشخیص بیماری، تشخیص علف‌های هرز، تشخیص محصول و کیفیت محصول است. این روش‌ها به افزایش بهره‌وری و در نتیجه، درآمد مالی کمک می‌کند. همچنین علم داده می‌تواند برای ایجاد سامانه‌های نظارتی محصول استفادهشود. با استفاده از حسگرها، پهپادها و ماهواره‌ها داده‌های مختلف جمع‌آوری شده که می‌تواند برای شناسایی زودهنگام مشکلات و انجام اقدامات اصلاحی تجزیه و تحلیل شود. این به بهبود عملکرد و جلوگیری از تلفات ناشی از آفات یا بیماری‌ها، کم آبی و کمبود مواد مغذی کمک می‌کند (Virnodkar و همکاران، ۲۰۲۰). با تجزیه و تحلیل داده‌های تاریخی در مورد الگوهای آب و هوایی و عملکرد محصولات، محققین می‌توانند مدل‌هایی را توسعه دهند که بتوانند پیش‌بینی کنند وضعیت یک محصول خاص در آینده تحت شرایط مختلف چگونه خواهد بود (Kouadio و همکاران، ۲۰۱۸؛ Pourmohammadali و همکاران، ۲۰۱۹؛ Khosla و همکاران، ۲۰۲۰؛ Alibabaei و همکاران، ۲۰۲۱؛ Cedric و همکاران، ۲۰۲۲). این اطلاعات می‌تواند به کشاورزان کمک کند تا در مورد کاشت نوع محصول و زمان برداشت آن به درستی تصمیم‌گیری کنند.

• تعیین الگوی کشت

برآورد مساحت محصول به روش سنتی زمان‌بر، پرهزینه و با احتمال خطا و سوگیری انسان است. همچنین برآورد مساحت محصول در روستاهای دور افتاده، زمین‌های صعب‌العبور و مزارع شیب‌دار بسیار دشوار است. به منظور غلبه بر این مشکلات، برای برآورد سطح محصول می‌توان از ترکیب فناوری سنجش از دور و یادگیری ماشین استفاده کرد. ماهواره‌های سنجش از دور، تصاویر زمانی، سینوپتیکی، چند طیفی و با وضوح چندگانه از پوشش زمین را فراهم

می‌کند و توانایی طبقه‌بندی محصولات مختلف را ارائه می‌دهد. تاکنون مطالعات متعددی در این زمینه انجام شده است که حاکی از دقت خوب این روش‌ها در تعیین الگوی کشت محصولات است (شریفی و همکاران، ۱۴۰۰، باقری، ۱۴۰۱؛ Pan و Zhao، ۲۰۱۸؛ Kranjčić و همکاران، ۲۰۱۹؛ Blickensdörfer و همکاران، ۲۰۲۲؛ Ghassemi و همکاران، ۲۰۲۲). برای این منظور در فناوری سنجش از دور، بازتاب طیفی بررسی شده و گزارشی از سلامت پوشش گیاهی به عنوان مثال، ساختار گیاه، رطوبت و محتوای کلروفیل بررسی می‌شود. سنسورهای نوری از راه دور در باندهای طیفی مرئی، NIR[۵۳] و SWIR[۵۴] کار می‌کنند. به کمک باندهای مختلف تصاویر ماهواره‌ای می‌توان شاخص‌های مختلفی را محاسبه نمود. برخی از پرکاربردترین شاخص‌ها با عنوان‌های NDVI[۵۵]، GNDVI[۵۶] و LAI[۵۷] شناخته می‌شوند. شاخص‌های محاسبه در هر سلول از تصویر که عوارض آن متفاوت باشد، مقدار متفاوتی می‌دهد. همین امر مبنای طبقه‌بندی تصاویر می‌باشد. نتایج الگوریتم در پژوهش‌های مختلف از باندهای تصاویر ماهواره‌ای و یا شاخص اختلاف گیاهی نرمال شده (NDVI) به عنوان ورودی و محدوده‌های داده‌برداری شده با کلاس یا طبقه معین به‌دست می‌آیند (Hudait و Patel، ۲۰۲۲؛ Gumma و همکاران، ۲۰۲۲؛ Ranjan و Parida، ۲۰۱۹).

• سایر کاربردهای علم داده در حوزه مدیریت آب کشاورزی

استفاده از یادگیری ماشین در حوزه مدیریت آب تنها معطوف به موارد ذکر شده نبوده و در بخش‌های دیگری نیز استفاده می‌شود. به عنوان مثال می‌توان به پایش و پیش‌بینی خشکسالی، تغییرات اقلیمی، سطح آب‌های زیرزمینی و... اشاره کرد.

خشکسالی یکی از پیچیده‌ترین بلایای طبیعی است که شناسایی، پیش بینی و کاهش آن اغلب دشوار است (Aadhar و Mishra، ۲۰۱۷). علاوه بر این، شاخص‌های مختلفی به منظور پایش خشکسالی ارائه شده است که به علت وابستگی آن‌ها به اندازه‌گیری‌های زمینی مانند اندازه‌گیری دما و بارندگی، برای نشان دادن توزیع مکانی دقیق در مقیاس بزرگ مناسب نیستند. امروزه روش‌های سنجش از دور با پوشش یک منطقه جغرافیایی بزرگ که برای پایش خشکسالی مناسب‌تر از روش مبتنی بر زمین است، کاربرد گسترده‌تری پیدا کرده‌اند (Prodhan و همکاران، ۲۰۲۱). اطلاعات دقیق و کامل خشکسالی با سیستم پایش مناسب در آشکار ساختن ماهیت پیچیده خشکسالی و عوامل مرتبط با آن بسیار مهم است. بسیاری از شاخص‌های خشکسالی برای توصیف خشکسالی کشاورزی، با تکیه بر داده‌های آب و هوایی یا عوامل مختلف خشکسالی سنجش از راه دور ایجاد شده‌اند (Feng و همکاران، ۲۰۱۹). چندین شاخص خشکسالی سنجش از دور (بیش از ۱۶۰ شاخص) از جمله شاخص تغییرات پوشش گیاهی نرمال شده (NDVI)، شاخص وضعیت پوشش گیاهی (VCI[۵۸])، شاخص وضعیت رطوبتی خاک (SMCI[۵۹]) و ... وجود دارند. می‌توان گفت در بحث خشکسالی، چالش اصلی این است که این شاخص‌ها قادر به انعکاس صحیح اطلاعات نیستند؛ زیرا شرایط خشکسالی به عوامل متعددی مرتبط است. در نتیجه، شاخص‌های خشکسالی منفرد اطلاعات پیچیده خشکسالی را نشان نمی‌دهند و لازم است که ترکیبی از شاخص‌ها مورد بررسی قرار گیرند. به همین علت مدل‌های مختلف یادگیری ماشین مانند رگرسیون خطی، سری‌های زمانی، یادگیری عمیق و حتی جنگل‌های تصادفی به منظور پیدا کردن روابط غیرخطی بین علل مختلف خشکسالی استفاده می‌شود (Chen و همکاران، ۲۰۱۲؛ Khadr، ۲۰۱۶؛ Li و همکاران، ۲۰۱۶؛ Shen و همکاران، ۲۰۱۹؛ Prodhan و همکاران، ۲۰۲۱).

با در نظر گرفتن تخلیه سریع بسیاری از سفره‌های زیرزمینی با تغذیه ناچیز، مدیریت مؤثرتر آب به منظور صرفه‌جویی بهتر آب از نظر دستیابی به تولید محصول پایدار مورد نیاز است (El Bilali و همکاران، ۲۰۲۲). در حوزه آب‌های زیرزمینی به طور معمول از علم داده به منظور پایش کمی و کیفی آب‌های زیرزمینی استفاده می‌شود. به طور کلی مدل‌های فیزیکی و عددی ابزار اصلی در مدل‌سازی و پیش‌بینی سطح آب زیرزمینی بوده‌اند. با این حال، از آنجا که این روش‌ها نیز بر روی ورودی‌های مختلف تکیه دارند و سازوکارهای زیربنایی معمولاً برای درک آن‌ها بسیار پیچیده هستند، اخیراً از روش‌های مبتنی بر داده استفاده می‌شود (Azari و همکاران، ۲۰۲۱؛ Osman، ۲۰۲۲). Ahmadi و همکاران (۲۰۲۲) در یک مطالعه فراتحلیل و مرور سیستماتیک بر روی ۱۹۷ مطالعه انجام شده در سراسر دنیا از سال‌های ۲۰۱۰ تا ۲۰۲۰، نتیجه گرفتند که روش‌های یادگیری ماشین می‌توانند با دقت بسیار بالایی سطوح آب زیرزمینی را نظارت و پیش‌بینی کنند. براساس نتایج این پژوهش الگوریتم‌های شبکه عصبی مصنوعی (ANN[۶۰]) با سهم ۵۳ درصدی در این مطالعات بیشترین کاربرد را داشته‌اند. داده‌های هواشناسی و هیدرولوژیکی عمده داده‌های هستند که در این حوزه به عنوان متغیرهای ورودی به الگوریتم‌های یادگیری ماشین استفاده می‌شوند (Osman، ۲۰۲۲).

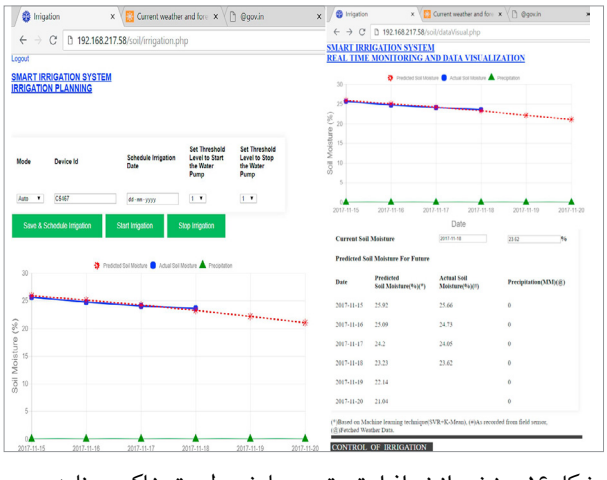
علم داده از حوزه مزرعه نیز فراتر رفته و در بحث مدیریت گلخانه‌ها نقش بسزایی ایفا می‌کند. این فناوری علاوه بر تسهیل و بهبود روش‌های آبیاری گلخانه‌ها می‌تواند در زمینه پایش و کنترل اقلیم گلخانه‌ها که تحت عنوان خرداقلیم[۶۱] نیز دسته‌بندی می‌شوند نقش بسزایی ایفا کند. مطالعه Chen و همکاران (۲۰۲۲) نشان داد که سامانه‌های کنترل هوشمند گلخانه‌ها[۶۲] می‌تواند در مقایسه با روش‌های سنتی به میزان ۶۶/۸ درصد به صرفه‌جویی در مصرف آب و انرژی دست یابد.

مدل‌های یادگیری ماشین در پیش‌بینی تولید گازهای گلخانه‌ای تولید شده در مزارع نیز به خوبی عمل کرده‌اند. Hamrani و همکاران (۲۰۲۰) در مطالعه‌ای با استفاده از شبکه‌های عصبی عمیق توانستند با دقت بالایی نیتروژن اکسید و کربن‌دی‌اکسید حاصل از فعالیت‌های

کشاورزی را پیش‌بینی کنند. این موضوع با توجه به نقش پر رنگ گازهای گلخانه‌ای در تغییرات اقلیمی بسیار حائز اهمیت است. همان‌طور که پیش‌تر اشاره شد هدف این پژوهش بررسی مهم‌ترین کاربردهای علم داده در حوزه مدیریت آب کشاورزی است و واضح است که این فناوری به شکل‌های دیگری نیز می‌تواند استفاده شود.

سامانه‌های توسعه یافته

مرحله انتهایی چرخه کریسپ فاز توسعه بود که به طور معمول به توسعه یک نرم افزار کاربردی ختم می‌شود. در این بخش به برخی از مطالعاتی پرداخته می‌شود که به طور خاص نتیجه کار آن‌ها به توسعه یک سامانه ختم شده است. برای بهره‌برداری مؤثر و بهینه از آب در مزرعه، توسعه سامانه‌های آبیاری هوشمند مبتنی بر پیش‌بینی الگوی رطوبتی خاک مزرعه و اطلاعات بارش روزهای آتی ضروری است. برای پیش‌بینی رطوبت خاک می‌توان از الگوریتم‌های مربوط به یادگیری ماشین بهره گرفت (Goap و همکاران، ۲۰۱۸). تمامی اطلاعات ذکر شده را می‌توان با توسعه یک پورتال وب به صورت آنی مشاهده کرد. این پورتال باید شامل خدماتی برای جمع‌آوری داده‌های حسگر میدانی، سرویس جمع‌آوری اطلاعات هواشناسی در دسترس موجود در اینترنت، سرویس برای کنترل پمپ آب، الگوریتم‌های پیش‌بینی رطوبت خاک و ناظر بر مزرعه در هر لحظه باشد. Goap و همکاران (۲۰۱۸) با استفاده از این اطلاعات به توسعه یک نرم‌افزار تحت وب پرداختند. ایشان در این نرم‌افزار با استفاده از الگوریتم‌های ترکیبی (ترکیب الگوریتم‌های SVR و K-means) یادگیری ماشین توانستند با دقت خوبی (بیش از ۹۶ درصد) رطوبت خاک را پیش‌بینی کنند. با این وجود این سامانه تحت وب به صورت محلی[۶۳] در دسترس بوده و جنبه استفاده عمومی تا زمان نگارش این پژوهش پیدا نکرده است. نمایی از این نرم‌افزار در شکل (۱۶) نشان داده شده است.



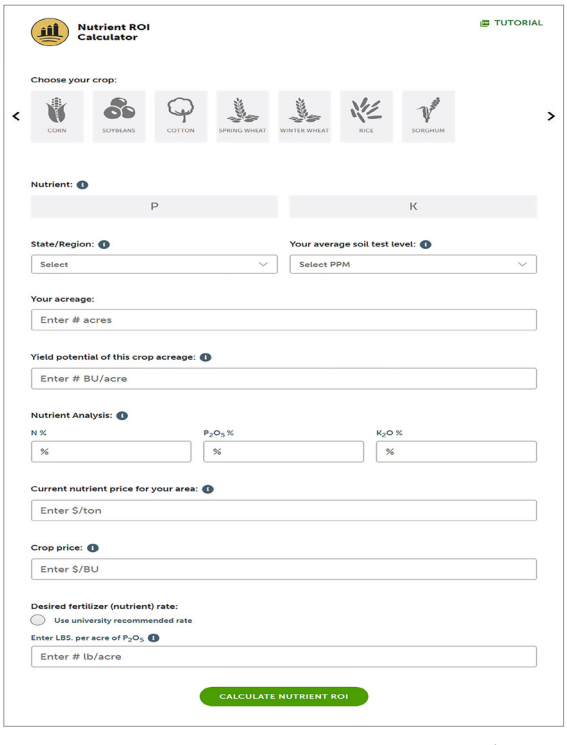
شکل ۱۶- بخشی از نرم‌افزار تحت وب پایش رطوبت خاک و برنامه‌ریزی زمان آبیاری (Goap و همکاران، ۲۰۱۸)

Gong و همکاران (۲۰۲۲) با ترکیب داده‌های مختلف گذشته (شامل داده‌های قدیمی هواشناسی و آبیاری)، حال (داده‌های حسگرها) و آینده (پیش‌بینی هواشناسی) به طراحی یک سامانه هوشمند آبیاری پرداختند. نتایج ایشان نشان داد که این سامانه می‌تواند باعث بهبود چشم‌گیر راندمان ذخیره آب شود. در این مطالعه نیز یک نرم‌افزار تلفن همراه ارائه شده است که کاربر می‌تواند به وسیله آن آبیاری را به صورت دستی و یا هوشمند انجام دهد (شکل ۱۷).

شکل ۱۷- نرم افزار تلفن همراه آبیاری هوشمند (Gong و همکاران، ۲۰۲۲)

به طورکلی در این زمینه مطالعات متعددی انجام شده است که برخی از آن‌ها برای سایر محیط‌های کشت مانند گلخانه‌ها نیز ارائه شده‌اند (Nie و همکاران، ۲۰۱۸؛ Chang و همکاران، ۲۰۱۹؛ Moghadas و همکاران، ۲۰۱۹؛ Muñoz و همکاران، ۲۰۲۰). برخی از مطالعات مباحث اقتصادی در زمینه کشاورزی را نیز در نظر گرفته‌اند. به عنوان مثال Nutrient ROI calculator[۶۴] یک سامانه تعیین بازده سرمایه‌گذاری در کشاورزی است که هدف آن کمک به کشاورزان در بهینه‌سازی عملکرد و درآمد است (Lindemann، ۲۰۱۹؛ Sinha و Dhanalakshmi، ۲۰۲۲). در واقع کشاورزان می‌توانند با استفاده از پیش‌بینی‌های این سامانه، راهبردهای دقیق‌تری برای کاربرد کود ایجاد کنند و در نتیجه سود حاصل از نهاده‌های مزرعه را بهینه کنند و بنابراین می‌توانند سودآوری خود را افزایش دهند (شکل ۱۸). پروژه دیگری تحت عنوان OpenIot به منظور بررسی شرایط مختلف برای کشت گندم توسعه یافته است. این بستر به کشاورزان کمک می‌کند تا زمان برداشت، زمان آبیاری و مواد مغذی مورد نیاز برای رشد گیاهان را پیش‌بینی کنند (Bhagat و همکاران، ۲۰۱۹).

برخی مطالعات نیز مانند پروژه SWAMP[۶۵] در حال توسعه یک مفهوم یا به عبارت دیگر معماری مناسب سامانه آبیاری هوشمند با دقت بالا برای کشاورزی هستند (Kamienski و همکاران، ۲۰۱۹). با این حال بر اساس نتایج جستجوهای انجام شده تا زمان نگارش این پژوهش، هنوز موضوع استفاده از یادگیری ماشین در زمینه مدیریت آب کشاورزی به گونه‌ای فراگیر نشده است که در مقیاس اجرایی و در دستان کشاورز مورد بهره قرار گیرد. در واقع اکثر مطالعات در فاز مطالعه باقی مانده و در سطح کلان استفاده نشده‌اند. همین امر نشان‌دهنده ظرفیت‌های بالقوه موجود فراوان در این حوزه است.



شکل ۱۸- نمایی از نرم افزار Nutrient ROI calculator

نتیجه‌گیری کلی

به طور کلی، هدف علم داده بهینه‌سازی عملکرد یک کار، از طریق بهره‌برداری از مثال‌ها یا تجربه‌های گذشته است. به طور خاص، یادگیری ماشین می‌تواند روابط کارآمدی بین داده‌های ورودی ایجاد و از آن‌ها به بینش و الگوی جدیدی برسد. الگوریتم‌های یادگیری ماشین قادر هستند روابط غیرخطی بین داده‌ها را کشف و مدل‌سازی کنند. نتایج مطالعات مختلف نشان می‌دهد که علم داده و یادگیری ماشین قدرت کافی در زمینه برآورد و پیش‌بینی شاخص‌های مختلف در حوزه کشاورزی را دارا هستند. بر این اساس استفاده از علم داده در بخش کشاورزی، علی‌الخصوص در حوزه آب کشاورزی بسیار کمک‌کننده خواهد بود. طبق جستجوهای انجام شده در پژوهش حاضر، تمرکز پژوهشگران بیشتر در موضوعات تعیین تبخیر-تعرق گیاه، پیش‌بینی عملکرد و تعیین کیفیت آب بوده است. در ایران نیز در سال‌های اخیر توجه محققین حوزه مهندسی آب به کاربرد علم داده جلب شده است. در بین کل مطالعات انجام شده این حوزه در جهان بین سال‌های ۲۰۱۸ تا ۲۰۲۰، ایران سهم ۵/۶۲ درصدی بین این مطالعات داشته است. در کل مطالعات انجام شده در این حوزه استفاده از داده‌های آب و هوا کاربرد بیشتری دارد که این موضوع می‌تواند به طور ضمنی به اهمیت برآورد تبخیر-تعرق در مدیریت آب کشاورزی اشاره داشته باشد.

رهیافت ترویجی

استفاده از علم داده در کشاورزی مزایای زیادی دارد، اما بالتّبع هر فناوری نوظهوری می‌تواند با مشکلاتی روبرو باشد. برخی از مشکلات موجود در استفاده از علم داده در حوزه مدیریت آب می‌تواند شامل موارد زیر باشد:

۱- مشکلات مربوط به داده‌ها: در برخی مناطق کشاورزی، داده‌های دقیق و کافی ممکن است موجود نباشد. این کمبود داده می‌تواند به کارآمدی مدل‌های پیش‌بینی و تحلیل‌های داده‌محور آسیب برساند. از سوی دیگر حجم بالای داده‌های جمع‌آوری شده از حسگرها و دستگاه‌های مختلف ممکن است باعث پیچیدگی در مدیریت، ذخیره‌سازی، و پردازش این داده‌ها شود.

۲- پیچیدگی فرآیندهای مختلف حوزه مدیریت آب و تعدد پارامترها: عوامل متعددی در فرآیندهای کشاورزی درگیر هستند که ممکن است باعث افزایش پیچیدگی مدل‌ها و تحلیل‌ها شود. این پیچیدگی ممکن است باعث کاهش تفسیرپذیری مدل‌ها شود. از طرف دیگر تدوین مدل‌های دقیق و قابل اعتماد برای پیش‌بینی‌ها و تحلیل‌ها ممکن است دشوار باشد، به خصوص زمانی که داده‌های ناقص وجود داشته باشد.

۳- هزینه‌های پیاده‌سازی: پیاده‌سازی فناوری‌های علم داده گاهاً می‌تواند نیازمند سرمایه‌گذاری هزینه‌بر باشد. این هزینه‌ها می‌توانند شامل هزینه تهیه داده، ابزار و رایانه باشند. در بعضی از مناطق، زیرساخت‌های فنی برای ارتباطات اینترنتی و پردازش داده‌ها به اندازه کافی توسعه یافته نیستند که می‌توانند استفاده از علم داده را با چالش روبرو کند.

۴- تخصص: نباید فراموش کرد که ساخت و پیاده‌سازی مدل‌های یادگیری ماشین و سامانه‌های مبتنی بر اینترنت اشیاء می‌تواند به دانش و منابع تخصصی نیاز داشته باشد که ممکن است به راحتی در بخش کشاورزی در دسترس نباشد. این امر می‌تواند استفاده از این فناوری‌ها را برای کشاورزان و سایر ذی‌نفعان دشوار کند.

درک دقیق از این محدودیت‌ها و توجه به آن‌ها در طراحی و پیاده‌سازی مفاهیم علم داده در حوزه علوم و مهندسی آب، می‌تواند کمک کند تا به چالش‌ها پاسخ داده شود و بهبودهای مؤثری در این

زمینه ایجاد شود. برای حل این مشکلات، لازم است تا سیاستمداران، محققین، و کشاورزان با همکاری یکدیگر، راهکارهایی را ارائه کنند که بتوانند از مزایای علم داده بهینه استفاده کنند و در عین حال به چالش‌ها و مشکلات موجود پاسخ دهند. به طور کلی، توسعه نرم‌افزارهایی که به‌صراحت پیرامون راه‌حل‌های مبتنی بر یادگیری ماشین و اینترنت اشیا طراحی شده‌اند، نیازمند سرمایه‌گذاری قابل توجهی است که ممکن است برای کشاورزان خرده‌مالک که به منابع بیشتری برای حمایت مالی از این طرح‌ها نیاز دارند، بسیار پرهزینه باشد. خوشبختانه، راه‌های مختلفی وجود دارد که می‌توان این موانع را از طریق سیستم‌های ابری با ارائه سکوهای دسترسی باز که در آن کاربران می‌توانند مجموعه داده‌های خود را بارگذاری کنند، بینش‌های خود را به اشتراک بگذارند، و در پروژه‌های تحقیقاتی همکاری کنند، برطرف کرد.

با این حال با توجه به نوظهور بودن این فناوری، هنوز خلاءهای مطالعاتی در این حوزه وجود دارد که انتظار می‌رود در آینده نظر محققین به آن‌ها جلب شود. به عنوان مثال می‌توان به کاربرد کمتر علم داده در بررسی تأثیرات زیست‌محیطی فعالیت‌های کشاورزی، بهینه‌تر کردن الگوریتم‌های فعلی، توسعه الگوریتم‌های جدید مناسب‌تر و همچنین تبدیل الگوریتم‌ها به سامانه‌های کاربردی در حوزه کشاورزی اشاره کرد. می‌توان از الگوریتم‌های یادگیری ماشین برای پیش‌بینی عملکرد محصول، قیمت‌ها و سایر روندهای کشاورزی آتی استفاده کرد. پیش‌بینی‌های کشاورزی می‌تواند برای تصمیم‌گیری در مورد راهبردهای کاشت، برداشت و بازاریابی استفاده شود. از طرف دیگر می‌تواند به کشاورزان و کسب‌وکارهای کشاورزی در پیش‌بینی الگوهای آب‌وهوا و برنامه‌ریزی برای بلایای طبیعی نیز کمک کند. این باعث می‌شود که آنها به ویژه برای پیش‌بینی‌های بلندمدت نیز مناسب باشند. کاربردهای بالقوه زیادی برای پیش‌بینی کشاورزی وجود دارد؛ از جمله کمک به کشاورزان برای تصمیم‌گیری در مورد بهترین زمان کاشت یا برداشت محصولات خود، حمایت از مشاغل کشاورزی در تصمیم‌گیری در مورد قیمت‌گذاری و در دسترس بودن محصول و اطلاع‌رسانی سیاست‌های دولت که از بخش کشاورزی حمایت می‌کنند. این موضوع می‌تواند در امر پیاده‌سازی الگوی کشت مناسب در مقیاس کشور نیز بسیار حائز اهمیت باشد.

پی‌نوشت‌ها

1-Food and Agriculture Organization
2-Data Science
3-Machine learning
4-Cleveland
5-Artificial Intelligence (AI)
6-Deep learning
7-Training data
8-Big Data
9-Entities
10-Attributes
11-Analytics record
12-Data set
13-Numeric
14-Nominal
15-Ordinal
16-Extensible Markup Language
17-CRISP-DM
18-Supervised Learning
19-Un supervised Learning
20-Semi-supervised Learning
21-Classification
22-Regression
23-Parameterized mathematical model
24-Multiple linear regression
25-K-Nearest Neighbors
26-Euclidean
27-Manhattan
28-Minkowski
29-Hamming
30-Support Vector Machine
31-Lazy learning algorithm
32-Hyper plane
33-Margin maximization
34-Iterative Dichotomiser 3
35-Clustering
36-Dimensionality Reduction
37-Anomaly Detection
38-K-means
39-DBSCAN
40-Mean-shift
41-Generative Methods
42-Graph Based Methods
43-Heuristic Based Methods
44-Science Direct
45-https://mjl.clarivate.com/; Web of Science
46-sciencedirect.com
47-Meta-Analysis
48-Label
49-Internet of things
50-Center pivot
51-Goiás
52-Minas Gerais
53-Near infrared
54-shortwave infrared
55-Normalized Difference Vegetation Index
56-Green Normalized Difference Vegetation Index
57-Leaf area index
58-Vegetation condition index
59-Soil Moisture Condition Index
60-Artificial neural network
61-Microclimate
62-Smart microclimate-control system
63-Local
64-return on investment
65-Smart water management platform

منابع

باقری، حمید. (۱۴۰۱). استفاده از الگوریتم فراابتکاری کرم شب تاب در بهبود دقت طبقه‌بندی تصویر ماهواره ای، مطالعه موردی: شهر رفسنجان. سنجش از دور و سامانه اطلاعات جغرافیایی در منابع طبیعی، انتشار آنلاین ۶ تیرماه ۱۴۰۱. doi: 10.30495/girs.2022.692233

شریفی، عبید، اصغری بیرامی، بهنام، و مختارزاده، مهدی. (۱۴۰۰).

طبقه‌بندی تصاویر ابرطیفی با استفاده از ادغام ویژگی‌های طیفی و مکانی در شبکه‌های عصبی پیچشی. نشریه مهندسی فناوری اطلاعات مکانی، ۹(۲)، ۲۷-۱. doi: 10.52547/jgit.9.2.1

کاویانی، مسعود. (۱۴۰۱). یادگیری ماشین و یادگیری عمیق با زبان‌های پایتون و R. انتشارات دیباگران. چاپ اول. ۲۴۶ صفحه. تهران، ایران.

نیکخواه، امیره. (۱۳۹۷). واژه نامه علم داده. انتشارات لیلاژ. چاپ اول. ۱۴۲ صفحه. تهران، ایران.

Aadhar, S., & Mishra, V. (2017). High-resolution near real-time drought monitoring in South Asia. Scientific Data, 4(1), 1-14. doi: 10.1038/sdata.2017.145

Ahmadi, A., Olyaei, M., Heydari, Z., Emami, M., Zeynolabedin, A., Ghomlaghi, A., ... & Sadegh, M. (2022). Groundwater level modeling with machine learning: a systematic review and meta-analysis. Water, 14(6), 949. doi: 10.3390/w14060949

Alibabaei, K., Gaspar, P. D., & Lima, T. M. (2021). Crop yield estimation using deep learning based on climate big data and irrigation scheduling. Energies, 14(11), 3004. doi: 10.3390/en14113004

Azari, A., Zeynoddin, M., Ebtehaj, I., Sattar, A. M., Gharabaghi, B., & Bonakdari, H. (2021). Integrated preprocessing techniques with linear stochastic approaches in groundwater level forecasting. Acta Geophysica, 69(4), 1395-1411. doi: 10.1007/s11600-021-00617-2

Benos, L., Tagarakis, A. C., Dolias, G., Berruto, R., Kateris, D., & Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. Sensors, 21(11), 3758. doi: 10.3390/s21113758

Bhagat, M., Kumar, D., & Kumar, D. (2019). Role of Internet of Things (IoT) in smart farming: A brief survey. 2019 Devices for Integrated Circuit (DevIC), 141-145. doi: 10.1109/DEVIC.2019.8783800

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., & Hostert, P. (2022). Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. Remote sensing of environment, 269, 112831. doi: 10.1016/j.rse.2021.112831

Burra, D. D., Hildebrand, J., Giles, J., Nguyen, T., Hasiner, E., Schroeder, K., ... & Kropff, W. (2021). Digital Agriculture Profile: Viet Nam. Food and Agriculture Organisation of the United Nations. Rome, Italy.

Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M., & Kimpolo, C. L. M. (2022). Crops yield prediction based on machine learning models: Case of West African countries. Smart Agricultural Technology, 2, 100049. doi: 10.1016/j.atech.2022.100049

Chang, Y. C., Huang, T. W., & Huang, N. F. (2019, September). A machine learning based smart irrigation system with LoRa P2P networks. In 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS) (pp. 1-4). IEEE. https://doi.org/10.23919/APNOMS.2019.8893034

Chapman, P., (1999). Julian Clinton (SPSS), Randy Kerber (NCR), Thomas Khabaza (SPSS), Thomas Reinartz (DaimlerChrysler), Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler),". CRISP-DM 1.0. Step-by-step data mining guide. Available online: https://www.coursehero.com/file/14884931/CRISP-DM-Process-Model-User-Guide

Chen, J., Li, M., & Wang, W. (2012). Statistical uncertainty estimation using random forests and its application to drought forecast. Mathematical Problems in Engineering, 2012(1), 915053, 1-12. doi: 10.1155/2012/915053

Chen, T. H., Lee, M. H., Hsia, I. W., Hsu, C. H., Yao, M. H., & Chang, F. J. (2022). Develop a Smart Microclimate Control System for Greenhouses through System Dynamics and Machine Learning Techniques. Water, 14(23), 3941. doi: 10.3390/w14233941

Chen, Z., Zhu, Z., Jiang, H., & Sun, S. (2020). Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods. Journal of Hydrology, 591, 125286. doi: 10.1016/j.jhydrol.2020.125286

Clark, M., & Tilman, D. (2017). Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. Environmental Research Letters, 12(6), 064016. doi: 10.1088/1748-9326/aa6cd5

Dwivedi, R. (2020). How to use the Random Forest classifier in Machine learning?. [Online]. Available: https://www.analyticssteps.com/blogs/how-use-random-forest-classifier-machine-learning. Updated: 1/18/2021.

Sentinel-2 satellite data with focus on field-level information. Geocarto International, 37(7), 1833-1849. doi: 10.1080/10106049.2020.1805029

Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. Water Quality Research Journal, 53(1), 3-13. doi: 10.2166/wqrj.2018.025

Hamrani, A., Akbarzadeh, A., & Madramootoo, C. A. (2020). Machine learning for predicting greenhouse gas emissions from agricultural soils. Science of The Total Environment, 741, 140338. doi: 10.1016/j.scitotenv.2020.140338

Hudait, M., & Patel, P. P. (2022). Crop-type mapping and acreage estimation in smallholding plots using Sentinel-2 images and machine learning algorithms: Some comparisons. The Egyptian Journal of Remote Sensing and Space Science, 25(1), 147-156. doi: 10.1016/j.ejrs.2022.01.004

IBM. (n.d.). K-Nearest Neighbors (KNN). IBM. Retrieved September 26, 2024, from https://www.ibm.com/topics/knn?mhsrc=ibmsearch_a&mhq=KNN

Jayaraman, P., Nagarajan, K. K., & Partheeban, P. (2022). A Review on Artificial intelligence Algorithms and Machine Learning to Predict the Quality of Groundwater for Irrigation Purposes. In 2022 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) (Vol. 1, pp. 1-8). IEEE. doi: 10.1109/ICDSAAI55433.2022.10028857

Kamienski, C., Soininen, J. P., Taumberger, M., Dantas, R., Toscano, A., Salmon Cinotti, T., ... & Torre Neto, A. (2019). Smart water management platform: IoT-based precision irrigation for agriculture. Sensors, 19(2), 276. doi: 10.3390/s19020276

Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press. Massachusetts, USA. doi: 10.7551/mitpress/11140.001.0001

Khadr, M. (2016). Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia). Ain Shams Engineering Journal, 7(1), 47-56. doi: 10.1016/j.asej.2015.11.005

Khosla, E., Dharavath, R., & Priya, R. (2020). Crop yield prediction using aggregated rainfall-based modular

El Bilali, A., Taleb, A. and Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agricultural Water Management, 245, p.106625. doi: 10.1016/j.agwat.2020.106625

FAO. (2020). The State of Food and Agriculture 2020. Overcoming water challenges in agriculture. Rome.

Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., ... & Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and forest meteorology, 263, 225-241. doi: 10.1016/j.agrformet.2018.08.019

Feng, P., Wang, B., Li Liu, D., & Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173, 303-316. doi: 10.1016/j.agsy.2019.03.015

Gao, H., Zhangzhong, L., Zheng, W., & Chen, G. (2023). How can agricultural water production be promoted? a review on machine learning for irrigation. Journal of Cleaner Production, 414, 137687. doi: 10.1016/j.jclepro.2023.137687

Ghassemi, B., Dujakovic, A., Żółtak, M., Immitzer, M., Atzberger, C., & Vuolo, F. (2022). Designing a European-wide crop type mapping approach based on machine learning algorithms using LUCAS field survey and Sentinel-2 data. Remote sensing, 14(3), 541. doi: 10.3390/rs14030541

Goap, A., Sharma, D., Shukla, A. K., & Krishna, C. R. (2018). An IoT based smart irrigation management system using Machine learning and open source technologies. Computers and electronics in agriculture, 155, 41-49. doi: 10.1016/j.compag.2018.09.040

Gong, L., Yan, J., Chen, Y., An, J., He, L., Zheng, L., & Zou, Z. (2022). An IoT-based intelligent irrigation system with data fusion and a self-powered wide-area network. Journal of Industrial Information Integration, 29, 100367. doi: 10.1016/j.jii.2022.100367

Gumma, M. K., Tummala, K., Dixit, S., Collivignarelli, F., Holecz, F., Kolli, R. N., & Whitbread, A. M. (2022). Crop type identification and spatial mapping using

Moghadas, D., Jadoon, K. Z., & McCabe, M. F. (2019). Spatiotemporal monitoring of soil moisture from EMI data using DCT-based Bayesian inference and neural network. Journal of Applied Geophysics, 169, 226-238. doi: 10.1016/j.jappgeo.2019.07.004

Muñoz, M., Guzmán, J. L., Sánchez-Molina, J. A., Rodríguez, F., Torres, M., & Berenguel, M. (2020). A new IoT-based platform for greenhouse crop production. IEEE Internet of Things Journal, 9(9), 6325-6334. doi: 10.1109/JIOT.2020.2996081

Nie, H., Yang, L., Li, X., Ren, L., Xu, J., & Feng, Y. (2018). Spatial prediction of soil moisture content in winter wheat based on machine learning model. In 2018 26th International Conference on Geoinformatics (pp. 1-6). IEEE. doi: 10.1109/GEOINFORMATICS.2018.8557119

Osman, A. I. A., Ahmed, A. N., Huang, Y. F., Kumar, P., Birima, A. H., Sherif, M., ... & El-Shafie, A. (2022). Past, present and perspective methodology for groundwater modeling-based machine learning approaches. Archives of Computational Methods in Engineering, 29(6), 3843-3859. doi: 10.1007/s11831-022-09715-w

Pan, X., & Zhao, J. (2018). High-resolution remote sensing image classification method based on convolutional neural network and restricted conditional random field. Remote Sensing, 10(6), 920. doi: 10.3390/rs10060920

Perea, R. G., Poyato, E. C., Montesinos, P., & Díaz, J. R. (2019). Prediction of irrigation event occurrence at farm level using optimal decision trees. Computers and electronics in agriculture, 157, 173-180. doi: 10.1016/j.compag.2018.12.043

Pierson, L. (2021). Data science for dummies. John Wiley & Sons, NewYork, USA.

Pourmohammadali, B., Hosseinifard, S. J., Salehi, M. H., Shirani, H., & Boroujeni, I. E. (2019). Effects of soil properties, water quality and management practices on pistachio yield in Rafsanjan region, southeast of Iran. Agricultural water management, 213, 894-902. doi: 10.1016/j.agwat.2018.12.005

Prodhan, F. A., Zhang, J., Yao, F., Shi, L., Pangali Sharma, T. P., Zhang, D., ... & Mohana, H. P. (2021). Deep learning for monitoring agricultural drought in South artificial neural networks and support vector regression. Environment. Development and Sustainability, 22, 5687-5708. doi: 10.1007/s10668-019-00445-x

Kotu, V., & Deshpande, B. (2018). Data science: concepts and practice. Morgan Kaufmann. Massachusetts, USA.

Kouadio, L., Deo, R. C., Byrareddy, V., Adamowski, J. F., & Mushtaq, S. (2018). Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties. Computers and electronics in agriculture, 155, 324-338. doi: j.compag.2018.10.014

Kranjčić, N., Medak, D., Župan, R., & Rezo, M. (2019). Machine learning methods for classification of the green infrastructure in city areas. ISPRS International Journal of Geo-Information, 8(10), 463. doi: 10.3390/ijgi8100463

Li, J., Zhou, S., & Hu, R. (2016). Hydrological drought class transition using SPI and SRI time series by log-linear regression. Water resources management, 30, 669-684. doi: 10.1007/s11269-015-1184-7

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674. doi: 10.3390/s18082674

Lindemann, M. D. (2019). 169 awardee talk-nutrition from a risk management perspective. Journal of Animal Science, 97(Supplement_3), 174-175. doi: 10.1093/jas/skz258.358

Ly, Q. V., Nguyen, X. C., Lê, N. C., Truong, T. D., Hoang, T. H. T., Park, T. J., ... & Hur, J. (2021). Application of Machine Learning for eutrophication analysis and algal bloom prediction in an urban river: A 10-year study of the Han River, South Korea. Science of The Total Environment, 797, 149040. doi: 10.1016/j.scitotenv.2021.149040

Maxwell, S. (2015). One water: the need for more holistic thinking, analysis, and policymaking in water. Journal-American Water Works Association, 107(3), 21-24. doi: 10.5942/jawwa.2015.107.0048

Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2017). Using MARS, SVM, GEP and empirical equations for estimation of monthly mean reference evapotranspiration. Computers and electronics in agriculture, 139, 103-114. doi: 10.1016/j.compag.2017.05.002

299-318). CRC Press, Florida, USA.

Sinha, B. B., & Dhanalakshmi, R. (2022). Recent advancements and challenges of Internet of Things in smart agriculture: A survey. Future Generation Computer Systems, 126, 169-184. doi: 10.1016/j.future.2021.08.006

Sundmaeker, H., Verdouw, C., Wolfert, S., & Freire, L. P. (2022). Internet of food and farm 2020. In Digitising the Industry Internet of Things Connecting the Physical, Digital and VirtualWorlds (pp. 129-151). River Publishers.

Virnodkar, S. S., Pachghare, V. K., Patil, V. C., & Jha, S. K. (2020). Remote sensing and machine learning for crop water stress determination in various crops: a critical review. Precision Agriculture, 21(5), 1121-1155. doi: 10.1007/s11119-020-09711-9

Wang, H., Liu, C., & Zhang, L. (2002). Water-saving agriculture in China: an overview. Advances in Agronomy, 75, 135-171. doi: 10.1016/S0065-2113(02)75004-9

Xing, L., Cui, N., Guo, L., Du, T., Gong, D., Zhan, C., ... & Wu, Z. (2022). Estimating daily reference evapotranspiration using a novel hybrid deep learning model. Journal of Hydrology, 614, 128567. doi: 10.1016/j.jhydrol.2022.128567

Yaseen, Z. M., Al-Juboori, A. M., Beyaztas, U., Al-Ansari, N., Chau, K. W., Qi, C., ... & Shahid, S. (2020). Prediction of evaporation in arid and semi-arid regions: A comparative study using different machine learning models. Engineering applications of computational fluid mechanics, 14(1), 70-89. doi: 10.1080/19942060.2019.1680576

Zanganeh, R., Mojaradi, B., & Jabbari, E. (2016). Leak Detection from the Buried Water Transmission Pipeline Using Landsat 8 Satellite Images (Case Study of the Kosar Water Transmission Pipeline). In Proceedings of the International Conference on Civil Engineering, Tehran, Iran (pp. 8-10).

Zhi, W., Feng, D., Tsai, W. P., Sterle, G., Harpold, A., Shen, C., & Li, L. (2021). From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale?. Environmental science & technology, 55(4), 2357-2368. doi: 10.1021/acs.est.0c06783

Asia using remote sensing data. Remote Sensing, 13(9), 1715. doi: 10.3390/rs13091715

Pu, F., Ding, C., Chao, Z., Yu, Y., & Xu, X. (2019). Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks. Remote Sensing, 11(14), 1674. doi: 10.3390/rs11141674

Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1, 81-106.

Raj. R. (2021). Supervised, Unsupervised, And Semi-Supervised Learning With Real-Life Usecase. Enjoy Algorithm. [Online]. Available: https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning. Updated: 07/08/2021.

Ranjan, A. K., & Parida, B. R. (2019). Paddy acreage mapping and yield prediction using sentinel-based optical and SAR data in Sahibganj district, Jharkhand (India). Spatial Information Research, 27(4), 399-410. doi: 10.1007/s41324-019-00246-4

Raschka, S. (2018). STAT 479: Machine Learning Lecture Notes. Department of Statistics, University of Wisconsin–Madison. https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/07_ensembles_notes.pdf

Rosegrant, M. W., Ringler, C., & Zhu, T. (2009). Water for agriculture: maintaining food security under growing scarcity. Annual review of Environment and resources, 34, 205-222. doi: 10.1146/annurev.environ.030308.090351

Saraiva, M., Protas, É., Salgado, M., & Souza Jr, C. (2020). Automatic mapping of center pivot irrigation systems from satellite images using deep learning. Remote Sensing, 12(3), 558. doi: 10.3390/rs12030558

ScienceDirect. (n.d.). Search results for "water, machine learning, agriculture". Retrieved December 2023, from https://www.sciencedirect.com

Shen, R., Huang, A., Li, B., & Guo, J. (2019). Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. International Journal of Applied Earth Observation and Geoinformation, 79, 48-57. doi: 10.1016/j.jag.2019.03.006

Singh, R., Deshwal, A., & Kumar, K. (2021). Implementation of smart irrigation system using intelligent systems and machine learning approaches. In Data Science and Innovations for Intelligent Systems (pp.