1. **مقدمه و اهمیت مسئله**کیفیت آب به‌عنوان یکی از منابع حیاتی برای بقای انسان، حیات موجودات و توسعه صنایع شناخته می‌شود. در سال‌های اخیر، آلودگی ناشی از فعالیت‌های انسانی و تغییرات محیطی باعث کاهش کیفیت منابع آبی شده است. در همین راستا، ارزیابی و پیش‌بینی شاخص کیفیت آب (WQI) و طبقه‌بندی کیفیت آب (WQC) اهمیت ویژه‌ای یافته‌اند. مقاله حاضر با بهره‌گیری از روش‌های یادگیری ماشین و به‌کارگیری تکنیک grid search به منظور بهینه‌سازی پارامترهای مدل‌های مختلف، به دنبال ارائه یک رویکرد نوین و کارا جهت پیش‌بینی وضعیت کیفیت آب است. در این پژوهش، مدل‌های طبقه‌بندی (RF، XGBoost، GB، AdaBoost) و مدل‌های رگرسیون (KNN، DT، SVR، MLP) مورد بررسی قرار گرفته‌اند تا از میان آن‌ها بهترین عملکرد انتخاب شود.
2. **مرور ادبیات و چالش‌های پیش‌بینی کیفیت آب**پیش‌بینی کیفیت آب همواره یکی از چالش‌های بزرگ در مطالعات محیط زیست بوده است. در مطالعات پیشین، از روش‌های سنتی مانند تحلیل‌های شیمیایی و آزمایشگاهی استفاده می‌شد که همزمان با هزینه بالا و زمان‌بر بودن مواجه بودند. پژوهش‌های اخیر تلاش کرده‌اند با بهره‌گیری از هوش مصنوعی و یادگیری ماشین، روند پایش و ارزیابی کیفیت آب را بهبود بخشند. مدل‌هایی نظیر ANN، SVR، LSTM و سایر الگوریتم‌های پیشرفته به‌منظور مدل‌سازی روابط غیرخطی بین ویژگی‌های مختلف آب و شاخص کیفیت به کار گرفته شده‌اند. از جمله چالش‌های اساسی می‌توان به نبود داده‌های کامل، پیچیدگی روابط میان متغیرها و نیاز به تنظیم دقیق پارامترهای مدل اشاره کرد. مقاله حاضر در راستای رفع این چالش‌ها، از تکنیک grid search جهت یافتن بهترین ترکیب پارامترها برای مدل‌های انتخاب شده بهره برده است.
3. **دیتاست و پیش‌پردازش داده‌ها**برای ارزیابی عملکرد مدل‌های پیشنهادی، از دیتاست واقعی جمع‌آوری‌شده از دریاچه‌ها و رودخانه‌های هند (از سال ۲۰۰۵ تا ۲۰۱۴) استفاده شده است. این دیتاست شامل ۱۹۹۱ نمونه و ۷ ویژگی کلیدی نظیر اکسیژن محلول، pH، هدایت الکتریکی، نیاز بیولوژیکی به اکسیژن (BOD)، نیترات، فکال کلرفورم و کلرفورم کلی است.  
   در گام اولیه، پیش‌پردازش داده‌ها شامل تکمیل مقادیر گمشده با استفاده از میانگین‌گیری و اعمال نرمال‌سازی جهت کاهش اثر تفاوت مقیاسی بین ویژگی‌ها انجام شد. همچنین، به منظور کاهش اثر نویز و آماده‌سازی داده‌ها برای مدل‌سازی، ماتریس همبستگی بین ویژگی‌ها تحلیل شده و از نمودارهای حرارتی برای شناسایی الگوهای پنهان استفاده گردید.
4. **مدل‌های طبقه‌بندی: معرفی و کاربرد**برای پیش‌بینی طبقه‌بندی کیفیت آب (WQC)، چهار مدل طبقه‌بندی به‌کار گرفته شده‌اند:  
   • جنگل تصادفی (RF): با ترکیب چندین درخت تصمیم، این مدل توانایی کاهش خطاهای تعمیم و اجتناب از هم‌بستگی میان ویژگی‌ها را داراست.  
   • XGBoost: روشی مبتنی بر افزایش گرادیان با استفاده از مشتقات درجه اول و دوم، که دقت بالایی در طبقه‌بندی داده‌های پیچیده دارد.  
   • گرادیان بوستینگ (GB): با ترکیب مدل‌های ضعیف و تمرکز بر نمونه‌های اشتباه طبقه‌بندی‌شده، عملکرد بهتری در پیش‌بینی ارائه می‌دهد.  
   • AdaBoost: این مدل با تنظیم مجدد وزن نمونه‌ها در هر مرحله، تلاش می‌کند اشتباهات مدل‌های اولیه را کاهش دهد.  
   با بهره‌گیری از تکنیک grid search، هر یک از این مدل‌ها از نظر تنظیم پارامترهایی نظیر تعداد برآوردگرها، عمق درخت‌ها و نرخ یادگیری بهینه شده و بهترین عملکرد برای پیش‌بینی WQC به دست آمده است.
5. **مدل‌های رگرسیون: تحلیل و توضیحات**در بخش پیش‌بینی شاخص کیفیت آب (WQI)، از چهار مدل رگرسیون استفاده شده است:  
   • K‑Nearest Neighbors (KNN): مدلی غیرپارامتری که از نزدیک‌ترین همسایگان جهت پیش‌بینی استفاده می‌کند.  
   • درخت تصمیم (DT): این مدل با ساختار درختی تصمیم‌گیری، قادر به مدل‌سازی روابط غیرخطی میان ویژگی‌ها است.  
   • پشتیبانی برداری (SVR): الگوریتمی مبتنی بر تئوری پشتیبانی برداری که برای مسائل رگرسیون با استفاده از توابع کرنل مناسب است.  
   • شبکه چندلایه پرسپترون (MLP): شبکه عصبی پیشرو که با استفاده از چندین لایه پنهان می‌تواند روابط پیچیده و غیرخطی بین متغیرها را یاد بگیرد.  
   در این بخش نیز همانند مدل‌های طبقه‌بندی، از grid search جهت تنظیم دقیق پارامترهایی مانند تعداد همسایگان، عمق درخت، پارامترهای کرنل و تنظیمات شبکه عصبی استفاده گردید تا عملکرد مدل‌ها به حداکثر برسد.
6. **بهینه‌سازی پارامترها با Grid Search**یکی از نقاط قوت این مقاله، استفاده از روش grid search جهت بهینه‌سازی پارامترها است.  
   در این روش، تمامی ترکیب‌های ممکن از مقادیر تعریف‌شده برای هر پارامتر در یک شبکه جستجو مورد بررسی قرار می‌گیرند. به عنوان مثال، برای مدل‌های طبقه‌بندی پارامترهایی مانند تعداد درخت‌ها (N\_estimators)، عمق درخت (Max\_depth) و معیار تقسیم (Criterion) مورد آزمایش قرار گرفتند. در بخش رگرسیون نیز پارامترهایی همچون تعداد همسایگان در KNN، پارامترهای تنظیم SVR (C، Epsilon و Kernel) و تنظیمات مربوط به لایه‌های مخفی در MLP بهینه شدند. استفاده از اعتبارسنجی متقابل (CV=۵) در هر مرحله، امکان انتخاب ترکیب بهینه پارامترها را فراهم ساخت. این رویکرد باعث شد تا مدل‌ها از نظر دقت پیش‌بینی و کاهش خطا به نتایج بسیار مطلوب دست یابند.
7. **نتایج مدل‌های طبقه‌بندی و معیارهای ارزیابی**پس از بهینه‌سازی پارامترها، عملکرد مدل‌های طبقه‌بندی با استفاده از معیارهایی همچون دقت، فراخوانی (Recall)، دقت (Precision)، میانگین F1 و ضریب همبستگی متمتوس (MCC) مورد ارزیابی قرار گرفت. نتایج نشان داد که:  
   • مدل گرادیان بوستینگ (GB) با دقت ۹۹.۵۰ درصد، عملکرد برتری نسبت به سایر مدل‌ها ارائه می‌دهد.  
   • نمودارهای اهمیت ویژگی‌ها نیز نشان می‌دهد که برخی ویژگی‌ها (به‌ویژه BOD و اکسیژن محلول) تأثیر قابل توجهی در تعیین طبقه‌بندی کیفیت آب دارند.  
   این نتایج علاوه بر نشان دادن قابلیت بالای مدل GB، اهمیت تنظیم دقیق پارامترها در بهبود عملکرد مدل‌های طبقه‌بندی را نیز برجسته می‌کند.
8. **نتایج مدل‌های رگرسیون و تحلیل معیارهای پیش‌بینی**برای پیش‌بینی شاخص کیفیت آب (WQI) نیز مدل‌های رگرسیون مورد آزمایش قرار گرفتند. معیارهای مورد استفاده شامل میانگین خطای مطلق (MAE)، میانگین خطای مربعی (MSE)، خطای میانه مطلق (MedAE) و ضریب تعیین (R2) بودند. نتایج به دست آمده نشان داد که:  
   • مدل شبکه چندلایه پرسپترون (MLP) با ضریب تعیین برابر با ۹۹.۸ درصد و مقدار بسیار پایین MSE (در حدود 2.8×10⁻⁵) بهترین عملکرد را در پیش‌بینی WQI ارائه می‌دهد.  
   • مدل‌های دیگر مانند KNN، DT و SVR نیز عملکرد مناسبی داشتند، اما نتایج آن‌ها نسبت به مدل MLP کمی پایین‌تر ارزیابی شد.  
   این دستاوردها نشان می‌دهد که شبکه‌های عصبی با توانایی یادگیری روابط غیرخطی پیچیده، برای مسائل رگرسیونی نظیر پیش‌بینی کیفیت آب بسیار مناسب هستند.
9. **بحث و تحلیل نتایج و مقایسه با مطالعات پیشین**در بخش بحث، پژوهشگران به تحلیل دقیق نتایج حاصل از مدل‌های مختلف پرداختند. از منظر مقایسه با مطالعات پیشین، نتایج به دست آمده در این تحقیق از دقت بسیار بالاتری برخوردار بودند؛ به عنوان مثال، مدل MLP با MSE بسیار پایین نسبت به مدل‌های NARNET و ANFIS نشان‌دهنده پیشرفت قابل توجه در ارزیابی کیفیت آب است. همچنین، نمودارهای «Actual vs Predicted» برای هر یک از مدل‌های رگرسیون به خوبی همخوانی میان مقادیر پیش‌بینی‌شده و واقعی را نشان داد که بیانگر قابلیت بالای مدل‌های بهینه‌شده است. از سوی دیگر، نتایج مدل‌های طبقه‌بندی نشان داد که استفاده از ensemble learning (مانند GB) موجب بهبود دقت پیش‌بینی شده و نمونه‌های اشتباه در طبقه‌بندی به حداقل رسیده‌اند. به‌طور کلی، یافته‌های این مقاله اهمیت بهینه‌سازی پارامترها در مدل‌های یادگیری ماشین و کاربرد این روش‌ها در مدیریت منابع آبی را به‌خوبی نشان می‌دهد.
10. **نتیجه‌گیری، توصیه‌ها و چشم‌انداز پژوهشی**نتیجه‌گیری نهایی مقاله نشان می‌دهد که استفاده از مدل‌های یادگیری ماشین همراه با بهینه‌سازی پارامترها (با استفاده از grid search) می‌تواند روشی مؤثر و کارآمد برای پیش‌بینی و ارزیابی کیفیت آب باشد.  
    از نکات برجسته این تحقیق می‌توان به موارد زیر اشاره کرد:  
    • مدل گرادیان بوستینگ (GB) در حوزه طبقه‌بندی کیفیت آب (WQC) با دقت ۹۹.۵۰ درصد عملکرد ممتاز داشته است.  
    • مدل شبکه چندلایه پرسپترون (MLP) در حوزه رگرسیون جهت پیش‌بینی شاخص کیفیت آب (WQI) با ضریب تعیین ۹۹.۸ درصد و خطاهای بسیار پایین، بهترین عملکرد را ارائه داده است.  
    • بهینه‌سازی دقیق پارامترها از طریق grid search به عنوان عاملی کلیدی در افزایش دقت مدل‌های یادگیری ماشین عمل کرده و امکان استفاده از این مدل‌ها در سامانه‌های پایش و مدیریت بحران آبی را فراهم ساخته است.

توصیه‌های پژوهشی برای تحقیقات آتی شامل موارد زیر است:  
۱. استفاده از مدل‌های پیشرفته‌تر عمیق مانند شبکه‌های بازگشتی (LSTM) برای تحلیل سری‌های زمانی کیفیت آب در مواجهه با تغییرات اقلیمی.  
۲. افزودن ویژگی‌های جدید مرتبط با شرایط محیطی، هواشناسی و آلودگی‌های صنعتی به منظور بهبود عملکرد مدل‌ها.  
۳. به کارگیری روش‌های انتخاب ویژگی (Feature Selection) جهت کاهش ابعاد داده و افزایش سرعت و دقت مدل‌های پیش‌بینی.  
۴. توسعه سیستم‌های ترکیبی (Hybrid Models) که با تلفیق چندین الگوریتم یادگیری ماشین، نقاط قوت هر کدام را به بهره‌گیری گرفته و عملکرد کلی سامانه‌های پیش‌بینی کیفیت آب را ارتقا دهند.

در پایان، مقاله حاضر با ارائه یک چارچوب جامع برای پیش‌بینی و طبقه‌بندی کیفیت آب، نشان‌دهنده اهمیت همگام‌سازی دانش داده و محیط زیست است. این پژوهش می‌تواند به عنوان یک الگوی موفق در طراحی سیستم‌های پایش هوشمند و مدیریت بهینه منابع آب در مقیاس‌های مختلف مورد استفاده قرار گیرد و زمینه پژوهش‌های آینده را در حوزه‌های مرتبط گسترش دهد.

**جمع‌بندی**مقاله حاضر نشان می‌دهد که استفاده از مدل‌های یادگیری ماشین، به‌ویژه در ترکیب با بهینه‌سازی پارامترها با روش grid search، می‌تواند به‌طور قابل توجهی در پیش‌بینی شاخص کیفیت آب (WQI) و طبقه‌بندی کیفیت آب (WQC) موفق عمل کند. نتایج به دست آمده نشانگر قابلیت بالای مدل‌های گرادیان بوستینگ و شبکه‌های پرسپترونی در پردازش داده‌های پیچیده و ارائه نتایج دقیق است. از منظر کاربردی، این رویکرد می‌تواند در طراحی سامانه‌های پایش و مدیریت منابع آب، سیستم‌های هشداردهنده و برنامه‌های کنترل آلودگی مورد استفاده قرار گیرد. همچنین، یافته‌های این تحقیق زمینه‌ساز پژوهش‌های آتی در زمینه‌های مرتبط با کیفیت آب و تغییرات اقلیمی بوده و امکان استفاده از الگوریتم‌های پیشرفته‌تر را فراهم می‌سازد.

در نهایت، تأکید می‌شود که پیشرفت‌های حاصل در این مقاله، نشان‌دهنده اهمیت همگام‌سازی علوم داده با مسائل زیست‌محیطی بوده و می‌تواند راهگشای توسعه راهکارهای نوین در مدیریت پایدار منابع آبی در سراسر جهان گردد.