

رویکرد درخت تصمیم در پیش‌بینی جریان رودخانه: مطالعه موردی رودخانه کشکان

سعید زارع اندلانی^۱، محمد ذاکر مشفق^۲

۱- دانشجوی کارشناسی مهندسی عمران-آب، دانشگاه صنعتی جندی شاپور دزفول

۲- استادیار گروه مهندسی عمران، دانشگاه صنعتی جندی شاپور دزفول

moshfegh@jsu.ac.ir

خلاصه

بحران کمبود آب در دهه‌های اخیر سبب توجه جدی‌تر به مدیریت منابع آب شده است. یکی از مهم‌ترین مباحث مطرح در مدیریت منابع آب، پیش‌بینی سیلاب است که یک فرآیند پیچیده و غیر خطی می‌باشد. از یک دیدگاه می‌توان روش‌های پیش‌بینی سیل را به دو دسته مدل‌های مفهومی و روش‌های داده محور تقسیم کرد. یکی از جدیدترین روش‌های داده‌محور، درخت تصمیم‌گیری می‌باشد. در این روش، داده‌های آموزش، با حرکت از ریشه‌ی درخت آموزش یافته با داده‌های آموزشی، به سمت برگ هدایت می‌شوند که هدایت این داده‌ها بر اساس قوانین کشف شده توسط درخت در مجموعه‌ی داده‌های آموزشی انجام می‌پذیرد. این پژوهش در ابتدا به تحلیل هیدروگراف سیل به کمک درخت تصمیم‌گیری و حساسیت درخت نسبت به متغیرهای مستقل می‌پردازد و سپس عملکرد درخت برای پیش‌بینی دبی روزانه هنگام آموزش آن با در نظر گرفتن همه متغیرهای مستقل و همچنین عملکرد آن در صورت حذف متغیرهای مستقل کم اهمیت شناسایی شده در مرحله‌ی تحلیل حساسیت، مورد ارزیابی قرار می‌گیرد. نتایج حاکی از اهمیت مقدار دبی روز قبل به عنوان تأثیرگذارترین متغیر و بارش ایستگاه‌های چمن‌انجیر و کرم‌بست می‌باشد. شش دسته حاصل تفکیک داده‌ها به کمک این سه متغیر به گونه‌ای است که قوانین حاکم بر داده‌های هر دسته با مفاهیم هیدرولوژیکی و هیدرولیکی همخوانی خوبی داشته و منطقی به نظر می‌رسند. نتایج مطلوب حذف متغیرهای کم‌اهمیت را می‌توان در انطباق بهتر هیدروگراف سیل پیش‌بینی شده و هیدروگراف سیل مشاهداتی و همچنین بهتر شدن شاخص‌های نکویی برازش به خوبی مشاهده کرد. مقادیر تعدادی از این شاخص‌ها مانند جذر متوسط مربع خطا ($RMSE$) و ضریب همبستگی (R^2) برای درخت ساخته شده با داده‌های فاقد مقادیر مربوط به متغیرهای کم‌اهمیت بهبود می‌یابند.

واژگان کلیدی: پیش‌بینی سیلاب، درخت تصمیم‌گیری، الگوریتم $CART$ ، تحلیل حساسیت، رودخانه کشکان.

۱- مقدمه

همواره پیش‌بینی سیلاب یکی از مهم‌ترین اهداف مهندسی آب به منظور برنامه‌ریزی منابع آب و مدیریت سیلاب بوده است. نتایج مربوط به تحلیل‌های آماری و داده‌های ثبت شده از این پدیده طبیعی نشان دهنده آن است که وقوع آن از قوانین خاصی پیروی می‌کند؛ هرچند ممکن است این قوانین به قدری پیچیده، غیر خطی و تابع عوامل متعدد باشند که کشف رابطه بین آن‌ها بسیار دشوار به نظر برسد. روش‌های مختلفی برای کشف این روابط ارائه شده‌اند که در یک نوع دسته‌بندی، شامل مدل‌های مفهومی و مدل‌های مبتنی بر داده‌کاوی^۱ می‌باشند. مدل‌های مفهومی به دلیل نیاز به کالیبراسیون پارامترها دارای مشکلات و پیچیدگی‌های خاص خود هستند و از این نظر این امر سبب محبوبیت مدل‌های مبتنی بر داده‌کاوی می‌گردد. یکی از جدیدترین مدل‌های داده‌کاوی، درخت تصمیم‌گیری^۲ می‌باشد (Mitchell 1997). این روش، مشاهدات و اندازه‌گیری‌های انجام شده در مورد مقادیر پارامترهای تأثیرگذار بر مقدار یک پدیده مورد نظر را به شکل قوانینی در آورده و از آن‌ها به منظور دسته‌بندی و پیش‌بینی استفاده می‌کند. به کمک درخت تصمیم‌گیری و با استفاده از داده‌های ثبت شده (داده‌های آموزشی) می‌توان یک ماشین را چنان آموزش داد که با معلوم بودن عوامل تأثیرگذار، مقدار تابع مورد نظر را پیش‌بینی کرد. انجام این کار توسط درخت تصمیم‌گیری به این صورت است که با پرسش یک سری سوال و تعیین مقادیر ممکن برای جواب این سوال، داده‌های ورودی از ریشه را در مسیرهای موجود حرکت داده و در نهایت به یک برگ می‌رساند. مقدار متغیر هدف در آن برگ، همان مقدار پیش‌بینی شده توسط درخت است. داده‌های قرار گرفته در برگ همگی ویژگی یکسانی دارند و آن ویژگی برگ است. این ویژگی می‌تواند یک عدد، بازه‌ای از اعداد یا یک عبارت باشد.