

مقایسه عملکرد روش‌های جنگل تصادفی و XGBoost نسبت به روش درخت تصمیم

فرزانه زارعیان

دانشجوی کارشناسی ارشد، دانشکده عمران و محیط زیست، دانشگاه صنعتی امیرکبیر

محمدسجاد زارعیان

استادیار گروه مهندسی عمران، دانشگاه شهاب دانش

چکیده

الگوریتم درخت تصمیم یکی از الگوریتم‌های مهم در یادگیری ماشین برای اهداف طبقه‌بندی و رگرسیون می‌باشد. این الگوریتم در عین سادگی دارای ضعف‌هایی می‌باشد که در کاربردهای عملی استفاده مستقیم از آن خیلی مناسب نمی‌باشد. بنابراین در سال‌های اخیر با استفاده از شیوه‌های مختلف به بهبود عملکرد این روش پرداخته شده است. در این مطالعه با استفاده از ۳ روش درخت تصمیم، جنگل تصادفی و XGBoost به پیش‌بینی خرید خودرو با استفاده از مجموعه داده ارزیابی خودرو که از مخزن داده خودروهای اهدایی مارکوبوهانک در دانشگاه کالیفرنیا بدست آمده است، پرداخته می‌شود. برای حل مسئله موردنظر، با بهینه کردن یک هایپرپارامتر مشترک بین تمامی روش‌ها و آموزش مدل‌ها بر اساس آن، عملکرد روش‌ها با یکدیگر مقایسه می‌گردد. با توجه به نتایج بدست آمده، الگوریتم جنگل تصادفی و XGBoost با وجود این که زمان آموزش بیشتری نسبت به الگوریتم درخت تصمیم دارند، دقت‌های بسیار خوب و پیش‌بینی‌های قابل اعتمادتری را از حل مسئله ارائه می‌دهند.

واژگان کلیدی: پیش‌بینی خرید خودرو، درخت تصمیم، جنگل تصادفی، XGBoost