سیویلیکا - ناشر تخصصی مقالات کنفرانس ها و ژورنال ها
گواهی ثبت مقاله در سیویلیکا
CIVILICA.com

عنوان مقاله:

بهینه سازی پارامترهای روش های یادگیری ماشین بر ارزیابی ژنومی صفات گسسته دودویی با در نظر گرفتن ساختار جمعیت و توزیع های متفاوت فنوتیپ در جمعیت مرجع

محل انتشار:

فصلنامه پژوهشهای علوم دامی ایران, دوره 12, شماره 1 (سال: 1399)

تعداد صفحات اصل مقاله: 12

نویسنده:

یوسف نادری - *دانشگاه آزاد اسلامی واحد آستارا*

خلاصه مقاله:

تنظیم اولیه و بهینه سازی پارامترهای ورودی روش های یادگیری ماشین گامی اساسی جهت دستیابی به حداکثر صحت پیش بینی ژنومی می باشد. در این تحقیق، جمعیت های ژنومی برای سطوح مختلف وراثت پذیری (۰/۵ و ۰/۲)، عدم تعادل پیوستگی (پایین و بالا) و تعداد متفاوت جایگاه صفات کمی (۲۰۰ و ۶۰۰) بر روی ۲۹ کروموزوم شبیه سازی شد. جهت ایجاد نسبت های مختلف فنوتیپ آستانه ای دودویی، فنوتیپ افراد جمعیت مرجع وابسته به اینکه باقی مانده آنها کمتر از ē-۱SDe (رویکرد اول) یا ۵۰ درصد افراد جمعیت (رویکرد دوم) باشد کد یک (فنوتیپ نامطلوب) و سایر حیوانات کد صفر (فنوتیپ مطلوب) اختصاص داده شد. برای بهینه سازی پارامترهای ورودی مدل، سطوح مختلف تعداد SNP نمونه گیری شده (۱۰۰، ۱۰۰۰ و mtry=۲۰۰۰)، تعداد بوت استراپ (۵۰۰، ۱۰۰۰ و ntree=۲۰۰۰) و حداقل اندازه گره پایانی (۱ و node size=۵) برای جنگل تصادفی و سطوح مختلف تعداد درخت (۱۰۰، ۱۰۰۰ و ntree=۲۰۰۰)، عمق درخت (۱، ۵ و tc=۱۰) و نرخ یادگیری (۰/۱ و lc=۰/۵) برای Boosting در نظر گرفته شد. کمترین میزان خطای خارج از کیسه برای mtry برابر با ۲۰۰۰، ntree برابر با ۱۰۰۰ و node size برابر با ۱ و کمترین خطای اعتبارسنجی در روش Boosting برای tc، ntree و lr به ترتیب ۱۰۰۰، ۱۰ و ۰/۵ مشاهده شد. صحت پیش بینی ژنومی روش های جنگل تصادفی و Boosting با کاهش فنوتیپ نامطلوب (رویکرد اول) افزایش یافت. بطور کلی در تمام سناریوها روش Boosting عملکرد بهتری نسبت به روش جنگل تصادفی داشت که دلیل این امر را می توان لحاظ کردن اثرات متقابل بین نشانگرها، خود ترمیمی و قدرت بالای این روش در کاهش خطای مدل دانست.

کلمات کلیدی:

اعتبارسنجی, صفات آستانه ای, عدم تعادل پیوستگی, وراثت پذیری, یادگیری ماشین

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/1417405