

مقایسه روشهای نزدیکترین همسایه مجاور و نیویز برای رده‌بندی داده‌ها

آرش رضایی^{۶۳}، رضا انتظاری‌ملکی^{۶۴}، بهروز مینایی‌بیدگلی^{۶۵}

چکیده

در این مقاله دقت تکنیک‌های k -NN و NB برای رده‌بندی داده‌ها با توجه به معیار "سطح زیر منحنی"ⁱ مقایسه شده است. تاثیر عواملی مانند اندازه مجموعه داده، تعداد صفات پیوسته و گسسته‌ی مستقل و عدم وجود داده نامعلومⁱⁱ در شرایط استقلال مجموعه داده از مسئله خاص، مورد بررسی قرار گرفته است.

نتایج بدست آمده بیانگر آن است که k -NN در اکثر موارد بهتر از NB عمل می‌نماید. در شرایطی که ترکیبی از صفات پیوسته و گسسته داشته باشیم و نسبت تعداد صفات پیوسته به گسسته بیشتر باشد NB خیلی ضعیف‌تر از k -NN عمل می‌کند. اما زمانی که این نسبت کمتر می‌شود، دقت NB بهبود می‌یابد ولی همواره کمتر از k -NN است.

کلمات کلیدی

روش نیویز، نزدیکترین همسایه مجاور، مجموعه داده‌ها، ناحیه زیر منحنی، تولید داده‌های تصادفی.

⁶³ دانشجوی کارشناسی ارشد مهندسی نرم‌افزار دانشگاه علم و صنعت ایران arash_rezaei@comp.iust.ac.ir

⁶⁴ دانشجوی کارشناسی ارشد مهندسی نرم‌افزار دانشگاه علم و صنعت ایران - r_entezari@comp.iust.ac.ir

⁶⁵ استادیار و عضو هیئت علمی دانشگاه علم و صنعت ایران - b_minai@iust.ac.ir