

پروژه درس داده کاوی ترم دوم سال تحصیلی ۱۴۰۰ - ۱۳۹۹

سجاد ابراهیمی ۹۷۱۲۷۶۲۴۶۵

فاز چهارم پروژه: طبقه بندی

برای یک مجموعه داده که مربوط به دیتاست دیوار و دیجی کالا می باشد. موارد ذیل را برای این دیتاست ها انجام دهید.

در این فاز باید یکی از پروژه های زیر انتخاب و در صورت تمایل پیاده سازی شوند.

- تعیین پرت بودن یا نبودن یک قیمت پیشنهاد داده شده
- تعیین قیمت برای یک محصول
- تعیین قیمت براساس توضیحات درج شده برای یک محصول

بخش انتخاب شده برای انجام تعیین قیمت برای یک محصول است.

در این بخش داده ها را با استفاده از فقط ستون ها cat1, cat2, cat3 به عنوان ورودی classifier به آن داده و label ما قیمت ها تعیین می شوند. سپس قیمت هایی که برابر ۱- هستند (لیبل قیمت توافقی در دیوار) را حذف کرده و سپس با تعیین مرز های پرت بودن قیمت یک داده، داده های دارای قیمت صحیح را نگهداری میکنیم. حال با استفاده از الگوریتم طبقه بندی KNN که شیوه عملکرد آن به این صورت که نزدیکترین n رکورد مشابه به رکورد ورودی را پیدا کرده و با میانگین گیری از label آنها، label داده ورودی را حدس می زند، اقدام به پیش بینی قیمت داده های تست می نماییم. در اینجا ما n را برابر ۵ تعیین کردیم یعنی هر قیمتی که حدس زده ایم میانگینی از قیمت ۵ محصول مشابه آن است.

برای اندازه گیری دقت از دو روش دقت دقیق و دقت حدودی استفاده کردیم بدین صورت که در دقت دقیق میزان قیمت هایی که دقیقاً صحیح اندازه گیری شده اند را نمایش و دقت حدودی قیمت هایی که تا $\alpha\%$ با قیمت اصلی اختلاف داشته اند را نیز به عنوان قیمت صحیح در نظر میگیرد که با تعیین اختلاف 30% ای برای α شاهد دقت حدودی 16% هستیم.