فاز اول

مقدمه

ج) میزان دقت طبقه بندها به صورت زیر است :

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| J48 | NBS | OneR | ZeroR | Validation/Classifier |
| 100% | 92.8571% | 71.4286% | 64.2857% | Use Training Set |
| 40% | 60% | 40% | 60% | Percentage Split (66%) |
| 0% | 33.3333% | 33.3333% | 33.3333% | Percentage Split (80%) |
| 0% | 0% | 0% | 0% | Percentage Split (88%) |

Accuracy Of Different Method in Weka

* روش ZeroR طبق ویژگی کلاس هدف دسته بندی را انجام می دهد و تمامی دسته بندی هایش را طبق بیشترین کلاسی که در رکوردهای کلاس هدف دیده می شود انتخاب می کند. از آنجا که ما 9 رکورد با کلاس هدف مثبت و 5 رکورد با کلاس هدف منفی داریم، این روش تمامی رکوردها را در کلاس مثبت قرار می دهد. نتایج این روش به عنوان پایه ای برای ارزیابی عملکرد دیگر روش ها استفاده می شود به این معنی که اگر روش های دیگر نتوانند دقتی بیشتر از این روش داشته باشند، روشی کارا برای ارزیابی نخواهند بود. پس به نوعی این روش کمترین دقت قابل پیش بینی را بر روی داده های آموزشی خواهد داشت.
* روش OneR با استفاده از یکی از ویژگی ها و قوانینی که برای آن وضع می کند تابع هدف را پیش بینی می کند و دیگر ویژگی ها را در آموزش پیش بینی خود دخیل نمی کند.
* روش NBS(Naïve Bayes Simple) به دلیل عدم توازن کلاس هدف رکوردها قادر به پیش بینی کاملا درست داده های آموزشی نبوده، اما عمومی سازی بهتری انجام داده است و بر روی داده های تست بیشترین درصد دقت را در مقایسه با مدل های دیگر داده است.
* روش J48 یا درخت تصمیم به طور کامل توانسته است که داده های آموزشی را پیش بینی کند زیرا در هر برگ از این درخت فقط رکوردهایی از داده های آموزشی که با کلاس هدف یکسان هستند، قرار دارند. از جهت دیگر قادر به عمومی سازی خوبی در مقایسه با طبقه بندهای دیگر نشده است. با بیشتر کردن رکوردها می توان دقت این طبقه بند را بهتر کرد.
* طبق نتایج بدست آمده روش NBS قدرت عمومی سازی بالاتری نسبت به دیگر روش ها را داراست، اما روش درخت تصمیم قادر است داده های آموزشی خود را به طور کامل حفظ کند.

د) از آنجا که ما نیازمند آنیم که روش ما برای رکوردهای جدید پیش بینی خوبی را انجام دهد، تست باید روی داده هایی که روی آن ها آموزشی انجام نگرفته نیز انجام شود، تا قادر به اندازه گیری میزان عمومی سازی روش خود باشیم. از نتایج این داده ها به این نتایج می رسیم :

* درخت تصمیم داده های آموزشی را به خوبی حفظ می کند اما قادر به عمومی سازی خوبی نسبت به روش های دیگر نبوده است.
* NBS فادر به حفظ کامل تمامی داده های آموزشی نبودی ولی به خوبی عمومی سازی کرده است.
* OneR نتوانسته است به کمک فقط یک ویژگی داده های آموزشی و داده های تست را به خوبی پیش بینی کند و دارای مشکل UnderFitting است.
* ZeroR نیز دارای مشکل underfitting است.

نکته : هرچند روش NBS روی داده های تست دفت بالاتری را نسبت به روش درخت تصمیم داشته است اما در هر صورت هر دوی آن ها دارای مشکل Overfitting هستند، مخصوصا درخت تصمیم.