

قسمت اول پروژه پایانی :

ذبیح از دانشجویهای دانشگاه علوم پایه زنجان هستند که یه پسر کوچیک به اسم عطا داره. ذبیح تصمیم داره تا با استفاده از یک دیتاست که در اختیار داره یک مدل بسازه تا عطا کوچولو بدونه در نبود ذبیح چه موقع می تونه بره بیرون و بازی کنه و چه موقع باید داخل خونه بمونه. چون ذبیح از داده کاوی سر در نیاره از شما به عنوان یه متخصص داده کاوی درخواست می کنه تا تو این راه بهش کمک کنین.

توضیحات دیتاست:

- دیتاست به دو فایل جدا برای یادگیری و تست به نام های `dataSetTrain.csv` و `dataSetTest.csv` تقسیم شده است.
- این دیتاست از پنج قسمت (ستون) مختلف تشکیل شده که شامل وضعیت هوا، دما، رطوبت، وضعیت ورزش باد، و اجازه بیرون رفتن یا نرفتن می شود.
- دیتاست یادگیری نیاز به پیش پردازش دارد.

کارهایی که ذبیح از شما انتظار داره تا با برنامه weka براش انجام بدین:

- ۱- پیش پردازش دیتاست یادگیری را انجام دهید و تصمیمات گرفته شده را برای هر ستون به صورت مستقل توضیح دهید. اجازه استفاده از فیلتر `replaceMissingValue` را ندارید ولی از بقیه فیلترها می توانید استفاده کنید.
- ۲- بر روی دیتاست یادگیری، الگوریتم `Naïve Bayse` را یکبار با انجام تست بر روی خود دیتاست یادگیری و یک بار با انجام تست بر روی دیتاست تست پیاده سازی کنین. نتایج بدست آمده را تشریح کنید و در صورت وجود تفاوت در نتایج این دو پیاده سازی، دلیل وجود تفاوت را از نظر خودتان بیان کنید.
- ۳- بر روی دیتاست یادگیری، الگوریتم درخت تصمیم را با انجام تست بر روی دیتاست تست ارائه شده پیاده سازی کنید و از روی درخت تصمیم، مدل بدست آمده را تشریح کنید. همچنین تصویر درخت را به صورت بصری چاپ و در داکيومنت قرار دهید.
- ۴- مدل های بدست آمده از سوال ۲ و ۳ را باهم مقایسه کنید و برداشت خودتان را به صورت کامل توضیح دهید.
- ۵- با استفاده از مدل بدست آمده از کدام الگوریتم عطا بهتر می تواند تصمیم بگیرد. دلیل خود را شرح دهید.

نکته ها:

قرار دادن تصویر مرحله به مرحله کارهای انجام شده داخل برنامه و کا نمره دارد.

پیدا کردن الگوریتم های خواسته شده از داخل الگوریتم های موجود در برنامه یکی از چالش های این پروژه می باشد.