

به نام خداوند

گزارش فاز سوم تمرین

در این فاز طول هر نمونه در غزلیات یک بیت و در اشعار نو دو عبارت در نظر گرفته شده است. فیچرهای انتخاب شده برای کلاسیفیکیشن شامل کلمه‌ی آخر هر نمونه (که در اشعار کلاسیک قافیه بیت است و در اشعار نو قاعده‌ی خاصی ندارد)، وجود کلماتی مانند عشق، جان، من و تو (کلماتی که بین دو دیتاست اختلاف زیادی داشتند)، نقش دو کلمه‌ی ابتدایی و انتهایی (جابه‌جایی نقش‌ها در ادبیات شعر بسیار متداول است؛ مخصوصاً شعر کهن) و همچنین تعداد توکن‌های هر نمونه هستند (وجه تمایز بارز دیگر بین اشعار نو و کلاسیک اختلاف طول بین عبارات شعر نوست). با این فیچرها دو الگوریتم بیشترین آنتروپی و نایبویز زده شده‌اند که دقت‌ها به شرح زیر می‌باشد:

در بیشترین آنتروپی:

غزلیات :

- Recall= 0.95
- Presicion = 0.74
- F1 = 0.83

اشعار نو:

- Recall= 0.67
- Presicion = 0.94
- F1 = 0.78

در نایویز:

غزلیات :

- Recall= 0.88 •
- Presicion = 0.78 •
- F1 = 0.82 •

اشعار نو:

- Recall= 0.75 •
- Presicion = 0.86 •
- F1 = 0.80 •