

به نام خدا

محمدرضا عرب زاده 810195429

Prj3

برای یادگیری برنامه این را در نظر گرفتیم که در ابتدا تعداد رخ دادن هر کلمه خاص برای هر کدام از دو شاعر را در یک دیکشنری قرار می دهیم و سپس آن مقدار را به تعداد کلمات هر کدام از شاعر ها تقسیم می کنیم که در واقع احتمال رخ دادن کلمه به شرط احتمال رخ دادن شاعر ذخیره می شود. سپس برای شناسایی یک بیت این احتمالا هارا برای هر کدام از کلمات و برای هر کدام از شاعر های سعدی و حافظ جداگانه محاسبه و ضرب می شود و در نهایت احتمال بدست آمده ضرب در تعداد کلمات حافظ برای احتمال حافظ و سعدی برای احتمال سعدی می شود.

احتمال شاعر به شرط بیت: Posterior

احتمال کلمه برای حافظ: Prior

احتمال کلمه به شرط شاعر: Likelihood

تعداد کل کلمات: Evidence

سوال اضافی یک: در این صورت بررسی نمی شود که چند درصد از جواب های موجود شناسایی شده اند برای مثال در این مسئله اگر فقط به precision توجه شود آنگاه برنامه آنهایی را که حافظ تشخیص داده است درست تشخیص داده ولی مشکل آنجاست که بسیاری از حافظ هارا اشتباها سعدی تشخیص می دهد.

لاپلاس:

مشکل آن است که در صورتی که یک کلمه برای فقط یکی از شاعر ها دیده شده باشد احتمال آن برای شاعر دیگر برابر صفر است و برای همین هنگامی که در یک بیت شعر احتمالات کلمات در هم ضرب می شود، احتمال شاعری که آن کلمه در دیتاست ما نیست صفر می شود و برنامه ممکن است اشتباها شاعر دیگر را تشخیص دهد.

برای حل این مشکل می توان گفت اگر تعداد تکرار یک کلمه برای یک شاعر صفر بود، آن را یک در نظر گرفت تا این مشکل حل شود.

تغییرات صورت گرفته در خروجی بعد از احتمال لاپلاس به این صورت است:

Before:

Accuracy= .58975 , recall= .5134, precision= .7242

After:

Accuracy= .7775 , recall= .7382, precision= .7172