



بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

روش‌های یادگیری ماشین در پردازش زبان طبیعی - نیم سال اول ۱۴۰۱-۱۴۰۲

توسط: هادی ویسی

تمرین شماره ۲

بارم کلیه نمرات از ۲۰ است

| ردیف | شماره دانشجویی | سوال ۱<br>15% | سوال ۲<br>50% | سوال ۳<br>35% | تاخیر<br>ساعت | نمره کل<br>از 20 |
|------|----------------|---------------|---------------|---------------|---------------|------------------|
| 1    | 830400035      | 20.0          | 20.0          | 20.0          | 2.0           | 19.60            |
| 2    | 830400016      | 20.0          | 20.0          | 18.3          |               | 19.39            |
| 3    | 830400047      | 20.0          | 19.0          | 20.0          | 1.0           | 19.31            |
| 4    | 830400060      | 0.0           | 0.0           | 0.0           |               | 0.00             |
| 5    | 830400036      | 0.0           | 0.0           | 0.0           |               | 0.00             |
| 6    | 830499024      | 18.0          | 15.0          | 16.0          | 28.0          | 11.38            |
| 7    | 830400023      | 20.0          | 20.0          | 20.0          |               | 20.00            |
| 8    | 830400043      | 0.0           | 0.0           | 0.0           |               | 0.00             |
| 9    | 830400028      | 20.0          | 20.0          | 20.0          |               | 20.00            |
|      | کمینه          | 0.0           | 0.0           | 0.0           | 1.0           | 0.0              |
|      | بیشینه         | 20.0          | 20.0          | 20.0          | 28.0          | 20.0             |
|      | میانگین        | 13.1          | 12.7          | 12.7          | 10.3          | 12.2             |

سوال سوم: در پیاده سازی روش بیز ساده عباراتی که باید خارج از حلقه قرار می‌گرفته اند داخل حلقه هستند و به همین علت، احتمالات نهایی اسپم و غیراسپم بودن اشتباه محاسبه شده و در نتیجه تشخیص نهایی غلط است.

سوال دوم: به علت اشتباهی در پیاده‌سازی برداری کردن جملات، ماتریس نهایی ضریب همبستگی با جواب مورد انتظار مطابقت ندارد.

سوال اول: مقدارهای گزارش شده شباهت کسینوسی برای داکيومنت‌های مختلف درست نیست. و از آنجایی که روش محاسبه بردارها و مقادیر ویژه درست است، احتمالاً در محاسبات ریاضی خطا رخ داده است. / سوال دوم: مقادیر گزارش شده برای معیار جاکارد غیر قابل قبول است (حاوی مقادیر بزرگ‌تر از ۱). همچنین مقادیر گزارش شده ضریب همبستگی حاوی مقادیر منفی است در صورتی که با پیاده‌سازی درست، برای تمامی حالت‌ها مقدار مثبت به دست می‌آید. / سوال سوم: هیچ پیش‌پردازشی روی داکيومنت‌ها انجام نشده و جواب نهایی گزارش نشده. پس از ران کد ارائه شده نیز جواب نهایی مبنی بر غیراسپم بودن کوئری است که غلط است.