

گزارش تمرین چهارم یادگیری ماشین

نام : حامد خشه چی

شماره دانشجویی : 9723494

جدول زیر را پر کنید.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Class6 | Class5 | Class4 | Class3 | Class2 | Class1 | Best word of |
| Guns | Rutgers | Encryption | Autos | Forsale | Graphics | 1 |
| Politics | Athos | Key | Rec | Misc | Ompass | 2 |
| Gun | Christian | Crypt | Car | Computer | op | 3 |
| Ifas | Od | Sci | Cars | Sale | Polygon | 4 |
| Gnv | Esus | Security | Dealer | Shipping | Raphics | 5 |
| Firearms | Religion | Org | Engine | Music | Animation | 6 |
| Ufl | Soc | Privacy | Vw | Sunysb | Mage | 7 |
| Talk | Hedrick | Eff | Cactus | Clone | Cols | 8 |
| Weapons | Aramis | Escrow | Cdac | Condition | Rows | 9 |
| Firearm | Igor | Technology | Ole | Oit | Au | 10 |
| Utk | Pproved | Lipper | Aturn | Unc | Acsc | 11 |
| Weapon | Hristians | Cryptography | W | Pc | Int | 12 |
| Ulowell | Hrist | Government | Tires | Marketplace | Row | 13 |
| Iftccu | Hristianity | Keys | Ntegra | Gibbs | Image | 14 |
| Utkvm | Truth | Answers | Fuel | Sking | Points | 15 |
| Laws | Faith | Enforcement | Kocrsv | Ba | Eality | 16 |
| Police | Hristian | Hip | Brandeis | Ale | Raf | 17 |
| Crime | Ible | Dministation | Pockets | Pchang | Irtual | 18 |
| Defense | Dziuxsolim | Secure | Boyle | 96 | Color | 19 |
| Arms | church | data | radar | sbcs | ys | 20 |

['graphics', 'ompass', 'op', 'polygon', 'raphics', 'animation', 'mage', 'cols', 'rows', 'au', 'acsc', 'int', 'row', 'image', 'points', 'eality', 'raf', 'irtual', 'color', 'ys']

['forsale', 'misc', 'computers', 'sale', 'shipping', 'music', 'sunysb', 'clone', 'condition', 'oit', 'unc', 'pc', 'marketplace', 'gibbs', 'sking', 'ba', 'ale', 'pchang', '96', 'sbcs']

['autos', 'rec', 'car', 'cars', 'dealer', 'engine', 'vw', 'cactus', 'cdac', 'ole', 'aturn', 'W', 'tires', 'ntegra', 'fuel', 'kocrsv', 'brandeis', 'pockets', 'boyle', 'radar']

['encryption', 'key', 'crypt', 'sci', 'security', 'org', 'privacy', 'eff', 'escrow', 'technology', 'lipper', 'cryptography', 'government', 'keys', 'answers', 'enforcement', 'hip', 'dministration', 'secure', 'data']

['rutgers', 'athos', 'christian', 'od', 'esus', 'religion', 'soc', 'hedrick', 'aramis', 'igor', 'pproved', 'hristians', 'hrist', 'hristianity', 'truth', 'faith', 'hristian', 'ible', 'dziuxsolim', 'church']

['guns', 'politics', 'gun', 'ifas', 'gnv', 'firearms', 'ufl', 'talk', 'weapons', 'firearm', 'utk', 'weapon', 'ulowell', 'iftccu', 'utkvm', 'laws', 'police', 'crime', 'defense', 'arms']

خروجی های ما بوده اند برای هر کلاس

که توسط نوشتن رگولار اکسپرشن روبرو به دست آمده اند

(?<!\d)[<>(),|\*-\_.?$!:\'`@"]|[<>()$,.!-\_?|\*`\':@"](?!\d)|\s+

که یعنی هر چی به جز وایت اسپیس ها و علامات خاص بود جدا کن و به ما بده که کلمات اند که فایل وکا نیز بر اساس همین ها به دست آمده و دقت نیو ما و وکا هردو یک بوده است

1.0

1.0

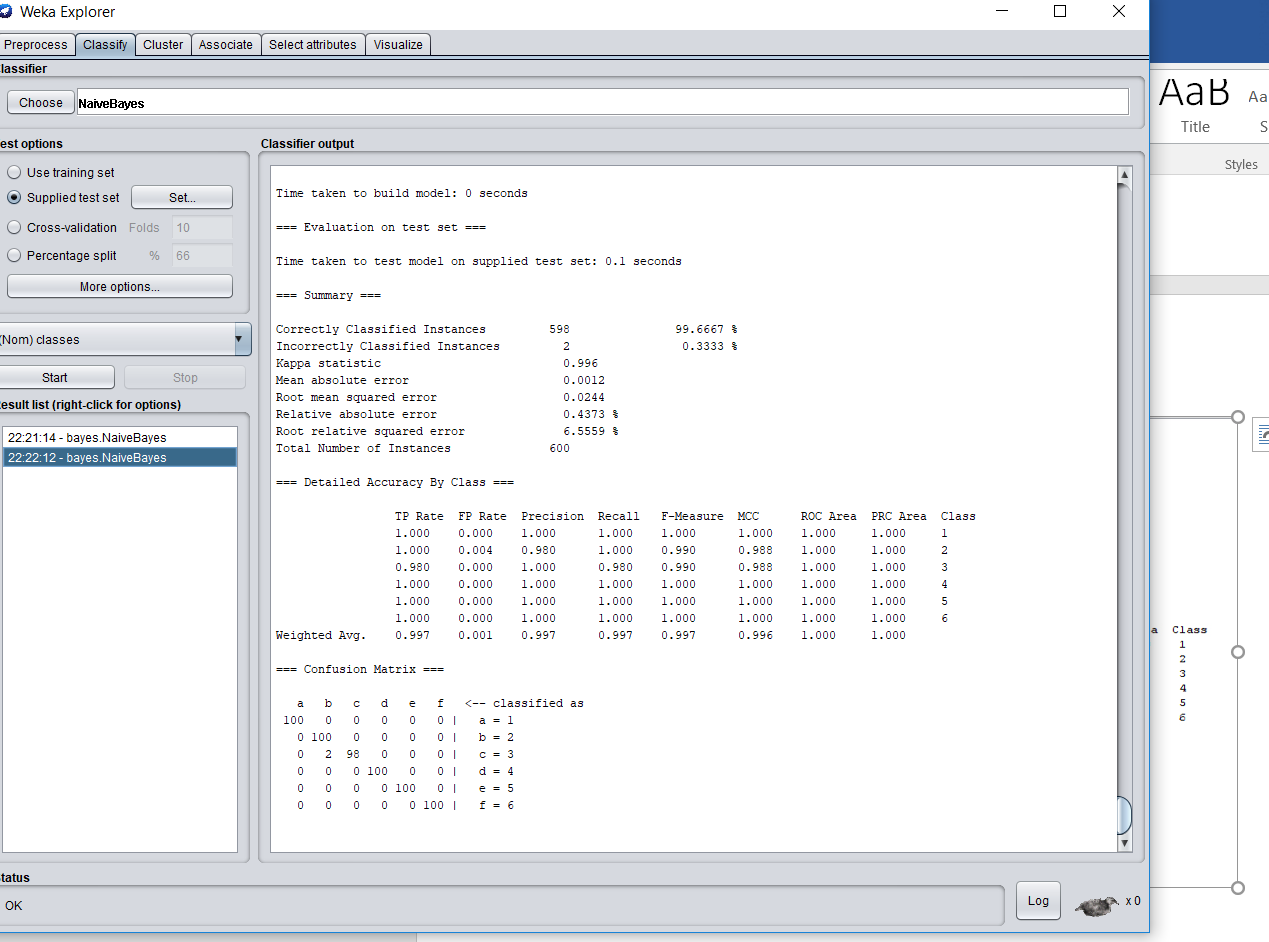
1.0

1.0

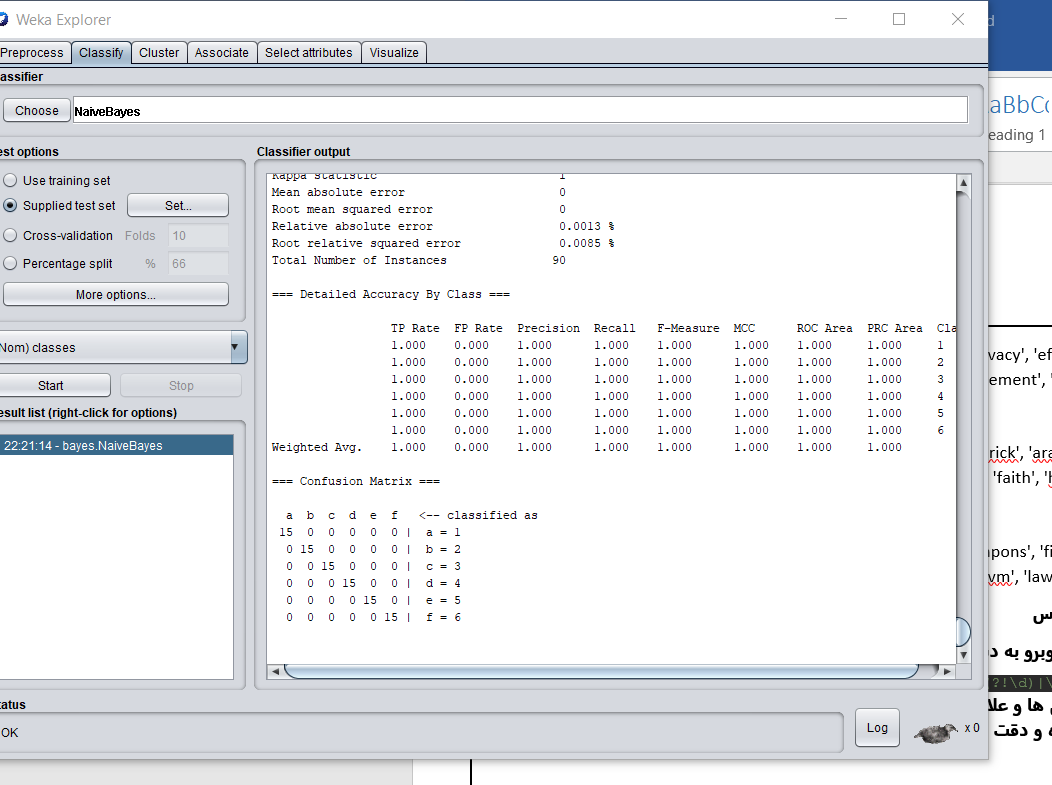
1.0

1.0

خروجی classifier naïve bayes برای داده های آموزشی (خروجی نرم افزار weka )



خروجی classifier naïve bayes برای داده های تست (خروجی نرم افزار weka )



و برای نوشتن کد از فایل کمکی استفاده کردم که یعنی من هم مقدار موثر را طبق همان فرمول حساب می کنم. که بیشترین تکرار در خود که در باقی نیست باشد!