

به نام خدا

تمرین اول درس پردازش زبان طبیعی دکتر ممتازی

دانشجو: محمد قربانی-97131099

بخش اول: 200 عدد از بهترین ویژگی‌های بدست آمده به همراه امتیاز مربوط به هر یک از ویژگی‌ها

- 1 (ورزش, '0.6952145174887998')
- 2 (ورزشی, '0.6116539756422967')
- 3 (اقتصاد, '0.5884575410801618')
- 4 (سیاسی, '0.5551796797227384')
- 5 (یاسی, '0.5533237607723467')
- 6 (ورز, '0.5184873311882234')
- 7 (اجتماعی, '0.5107256525551107')
- 8 (جتماع, '0.4978206923133697')
- 9 (اجتماع, '0.4978206923133697')
- 10 (صاد, '0.45748536722254896')
- 11 (تیم, '0.45235936889162054')
- 12 (جت, '0.4257065728472127')
- 13 (یاس, '0.41579508012283206')
- 14 (قت, '0.3410064713527685')
- 15 (بازی, '0.31199273280795525')
- 16 (سیا, '0.301315146315257')
- 17 (فوتبال, '0.2835028784871567')
- 18 (اقتصادی, '0.27109093113500693')
- 19 (وتب, '0.26876924158735616')

- (صادی, '0.2659805821972949) 20
- (فوت, '0.25732443567005303) 21
- (ادب, '0.24948856169706657) 22
- (بازیگ, '0.2237384045512243) 23
- (بازیکن, '0.2237384045512243) 24
- (قهرمان, '0.2195650706297516) 25
- (ازیگ, '0.21323339796046614) 26
- (هنر, '0.2104195325501197) 27
- (قهر, '0.20741183802368224) 28
- (اسلام, '0.20635954569673243) 29
- (رز, '0.20521145022849674) 30
- (یکن, '0.20187027412349567) 31
- (اق, '0.19966858888198313) 32
- (قهرمانی, '0.19480737508162882) 33
- (سلام, '0.19205359705007385) 34
- (اج, '0.19179431593452678) 35
- (باشگاه, '0.18928725931830748) 36
- (تولید, '0.1845028738968928) 37
- (تم, '0.1798992778114199) 38
- (ولید, '0.17795818661211626) 39
- (اسلامی, '0.17764796723637444) 40
- (مربی, '0.1773729138806157) 41
- (سلامی, '0.1765585467243107) 42
- (لامی, '0.1728576521032732) 43
- (شی, '0.16787921753324753) 44
- (اشگ, '0.16757934570710686) 45
- (تولی, '0.16619798983083234) 46

- 47 (هرم, '0.15806693721696918)
- 48 (لام, '0.1568969112686518)
- 49 (فدراسیو, '0.1555718870858973)
- 50 (فدراسیون, '0.1555718870858973)
- 51 (رقابت, '0.1526701511713986)
- 52 (رمانی, '0.15068864351520173)
- 53 (بازیکنان, '0.15016815209160672)
- 54 (مجلس, '0.1433324504047735)
- 55 (شورای, '0.13982995354124483)
- 56 (شورا, '0.13828710326435667)
- 57 (بازار, '0.137000762018594)
- 58 (ساب, '0.13564971052181596)
- 59 (مسابقات, '0.1348427960425842)
- 60 (ورای, '0.134831237096106)
- 61 (دلار, '0.13448476810239862)
- 62 (سابقا, '0.13353727442929486)
- 63 (نر, '0.13260463030737712)
- 64 (مسابقه, '0.1312882711815715)
- 65 (بازیها, '0.12899061903939124)
- 66 (ابقا, '0.12736771210233178)
- 67 (لید, '0.12277219951456875)
- 68 (باشگاهها, '0.12264897668823815)
- 69 (بال, '0.11806287867091658)
- 70 (آسیا, '0.11794287890618937)
- 71 (سرمایه, '0.11772966400091311)
- 72 (جمهور, '0.11711943382430778)
- 73 (صنعت, '0.11697992164691051)

- (فو, '0.11601983550004746) 74
- (باشگاههای, '0.11545480845356426) 75
- (سرمای, '0.11485026129553777) 76
- (صنع, '0.1148467860550153) 77
- (جمهوری, '0.11464472762774536) 78
- (قیمت, '0.11451373457382763) 79
- (صنایع, '0.11415249482323153) 80
- (حریف, '0.11276107192339513) 81
- (هوری, '0.11251197183681372) 82
- (دادگاه, '0.11249098495556487) 83
- (نعت, '0.1112345321024133) 84
- (بازیهایی, '0.11118048331973451) 85
- (راسی, '0.10996437884796828) 86
- (سرما, '0.10991457092466828) 87
- (نفت, '0.10855331348405617) 88
- (سل, '0.10611774068697999) 89
- (پیک, '0.10377235414836727) 90
- (صادرات, '0.10184105196462179) 91
- (هور, '0.09851897364467277) 92
- (مایه, '0.09847725898527626) 93
- (انقلاب, '0.09433894846255342) 94
- (قلاب, '0.09433122228103041) 95
- (گروه, '0.0941223452570017) 96
- (برتر, '0.09403003653408115) 97
- (دولت, '0.09346817743785207) 98
- (نظام, '0.09288518590287342) 99
- (بقا, '0.0922998875619232) 100

- 101 (گرو', '0.09223148894286615')
- 102 (انتخابات', '0.0911139177784794')
- 103 (ربی', '0.09077520130658279')
- 104 (برت', '0.09056572664735096')
- 105 (زنا', '0.08911934351863326')
- 106 (صنعتی', '0.08906850003025957')
- 107 (جل', '0.08902079380844485')
- 108 (آیت', '0.08738008663410346')
- 109 (توسعه', '0.08733215009014916')
- 110 (سعه', '0.08671633588439676')
- 111 (قانو', '0.08566107490956032')
- 112 (قانون', '0.08566107490956032')
- 113 (ریال', '0.08562362312013994')
- 114 (قه', '0.08523553999726996')
- 115 (ولت', '0.08505548569796417')
- 116 (فزایش', '0.08438159967899006')
- 117 (افزایش', '0.08438159967899006')
- 118 (بق', '0.08404664769915149')
- 119 (معظم', '0.08392042608062411')
- 120 (سینما', '0.08370252718282067')
- 121 (زایش', '0.08362842270300419')
- 122 (وسع', '0.08266818793918729')
- 123 (زندان', '0.08251046970554698')
- 124 (لس', '0.0821811196232698')
- 125 (قان', '0.08183258076751576')
- 126 (سابقه', '0.08017660469216903')
- 127 (خاتمی', '0.07999249051943558')

- 128 (کنان, '0.07995436059992134)
- 129 (ورزشگاه, '0.07958977995026872)
- 130 (نم, '0.07957347253690994)
- 131 (افزا, '0.07952641066171218)
- 132 (مانی, '0.07934393972757992)
- 133 (هنری, '0.07918643677895743)
- 134 (قاب, '0.07907817400277839)
- 135 (جامع, '0.07898750281181899)
- 136 (پیکا, '0.07847122319639022)
- 137 (وزیر, '0.07828524695275219)
- 138 (افزای, '0.07817858130504751)
- 139 (پرسپو, '0.07788515447184352)
- 140 (پرسپولیس, '0.07788515447184352)
- 141 (فیلم, '0.07773941381643024)
- 142 (مربیان, '0.07772044632678021)
- 143 (نما, '0.07746505371955426)
- 144 (شگاه, '0.0773715682020577)
- 145 (تولیدی, '0.07700042335000035)
- 146 (سی, '0.07692590945325506)
- 147 (نون, '0.07663316155727129)
- 148 (شگاههای, '0.07626192937644571)
- 149 (ایرنا, '0.07621186714570083)
- 150 (ولیس, '0.07620611897210017)
- 151 (تاکید, '0.07619160097858169)
- 152 (اکید, '0.07618188953612481)
- 153 (روزنامه, '0.07611103161921928)
- 154 (جامعه, '0.07601370235660321)

- 155 (عات, '0.07584949988159995)
- 156 (پولیس, '0.07565711944842457)
- 157 (تاکي, '0.0755783127279872)
- 158 (رئيس, '0.07547446708914274)
- 159 (ديدا, '0.07529048142942729)
- 160 (ارش, '0.0751837018913255)
- 161 (رهن, '0.07505785638855311)
- 162 (فضاي, '0.07481822407749439)
- 163 (قيم, '0.07414845347015042)
- 164 (لار, '0.0740630398409734)
- 165 (سها, '0.07397577638155672)
- 166 (تيمي, '0.07391636113215183)
- 167 (نماي, '0.07383697341790318)
- 168 (رمان, '0.07382679569094108)
- 169 (کيد, '0.07335168604442144)
- 170 (زن, '0.07301257663468874)
- 171 (ديدار, '0.07297497338123549)
- 172 (دلار, '0.07154264505931751)
- 173 (وت, '0.07140289941662115)
- 174 (صلاح, '0.07139159269372497)
- 175 (فرهنگ, '0.07133663244972244)
- 176 (طب, '0.07133075188381621)
- 177 (پترو, '0.07132666089527406)
- 178 (پتروشيمي, '0.07106379550890463)
- 179 (اتمي, '0.0710374109473868)
- 180 (ليگ, '0.07084929790483008)
- 181 (رهبر, '0.07081936512619125)

- 182 (مطب', '0.07075541602015312')
- 183 (مطبو', '0.07053061959529838')
- 184 (مطبوع', '0.07053061959529838')
- 185 (روزن', '0.07051007060480075')
- 186 (ورا', '0.0704993618690235')
- 187 (قضا', '0.07023335734651193')
- 188 (شکست', '0.069922336343053')
- 189 (ابا', '0.06992057067439372')
- 190 (ایش', '0.06972160589245924')
- 191 (صادر', '0.06936627229079573')
- 192 (مطبوعا', '0.06892270864548378')
- 193 (مطبوعات', '0.06892270864548378')
- 194 (درصد', '0.06826272575697256')
- 195 (حک', '0.06826233984245222')
- 196 (المپ', '0.06823192129566236')
- 197 (امن', '0.06819024568158305')
- 198 (مصاف', '0.06817071325695201')
- 199 (خانواده', '0.0681197607306876')
- 200 (گزارش', '0.06809681038396187')

بخش دوم:

unigram بدون feature selection:

Confusion Matrix:

شمارش tp و fn در زیر آمده است:

{(0.1, 'ادب و هنر', 'tp'): 17, (0.3, 'ادب و هنر', 'tp'): 17, (0.5, 'ادب و هنر', 'tp'): 17, (0.1, 'ورزش', 'fn'): 147, (0.3, 'ورزش', 'fn'): 147, (0.5, 'ورزش', 'fn'): 147, (0.1, 'سیاسی', 'tp'): 46, (0.3, 'سیاسی', 'tp'): 46, (0.5, 'سیاسی', 'tp'): 46, (0.1, 'اجتماعی', 'fn'): 134, (0.3, 'اجتماعی', 'fn'): 134, (0.5, 'اجتماعی', 'fn'): 134, (0.1, 'سیاسی', 'fn'): 154, (0.3, 'سیاسی', 'fn'): 154, (0.5, 'سیاسی', 'fn'): 154, (0.1, 'اقتصاد', 'tp'): 134, (0.3, 'اقتصاد', 'tp'): 134, (0.5, 'اقتصاد', 'tp'): 134, (0.1, 'ورزش', 'tp'): 81, (0.3, 'ورزش', 'tp'): 81, (0.5, 'ورزش', 'tp'): 81, (0.1, 'اقتصاد', 'fn'): 77, (0.3, 'اقتصاد', 'fn'): 77, (0.5, 'اقتصاد', 'fn'): 77, (0.1, 'ادب و هنر', 'fn'): 41, (0.3, 'ادب و هنر', 'fn'): 41, (0.5, 'ادب و هنر', 'fn'): 41, (0.1, 'اجتماعی', 'tp'): 29, (0.3, 'اجتماعی', 'tp'): 29, (0.5, 'اجتماعی', 'tp'): 29)}

اطلاعات کامل تر از این که در مواردی که سیستم به نادرستی تشخیص داده است. (چه دسته ایی را تشخیص داده است که به دین صورت می باشد: (counter =(label, real, delta)

{(0.1, 'ورزش', 'اقتصاد'): 119, (0.3, 'ورزش', 'اقتصاد'): 116, (0.5, 'ورزش', 'اقتصاد'): 114, (0.1, 'اجتماعی', 'سیاسی'): 8, (0.3, 'اجتماعی', 'سیاسی'): 8, (0.5, 'اجتماعی', 'سیاسی'): 86, (0.1, 'سیاسی', 'اقتصاد'): 8, (0.3, 'سیاسی', 'اقتصاد'): 86, (0.5, 'سیاسی', 'اقتصاد'): 86, (0.1, 'سیاسی', 'ادب و هنر'): 21, (0.3, 'سیاسی', 'ادب و هنر'): 21, (0.5, 'سیاسی', 'ادب و هنر'): 48, (0.1, 'ادب و هنر', 'اقتصاد'): 86, (0.3, 'ادب و هنر', 'اقتصاد'): 86, (0.5, 'ادب و هنر', 'اقتصاد'): 48, (0.1, 'ادب و هنر', 'اجتماعی'): 34, (0.3, 'ادب و هنر', 'اجتماعی'): 34, (0.5, 'ادب و هنر', 'اجتماعی'): 33, (0.1, 'اجتماعی', 'اقتصاد'): 33, (0.3, 'اجتماعی', 'اقتصاد'): 33, (0.5, 'اجتماعی', 'اقتصاد'): 33, (0.1, 'اجتماعی', 'ورزش'): 34, (0.3, 'اجتماعی', 'ورزش'): 34, (0.5, 'اجتماعی', 'ورزش'): 35, (0.1, 'ورزش', 'ادب و هنر'): 7, (0.3, 'ورزش', 'ادب و هنر'): 7, (0.5, 'ورزش', 'ادب و هنر'): 14, (0.1, 'ورزش', 'سیاسی'): 46, (0.3, 'ورزش', 'سیاسی'): 46, (0.5, 'ورزش', 'سیاسی'): 54, (0.1, 'سیاسی', 'اقتصاد'): 52, (0.3, 'سیاسی', 'اقتصاد'): 52, (0.5, 'سیاسی', 'اقتصاد'): 54, (0.1, 'سیاسی', 'ورزش'): 54, (0.3, 'سیاسی', 'ورزش'): 54, (0.5, 'سیاسی', 'ورزش'): 54, (0.1, 'اقتصاد', 'اجتماعی'): 80, (0.3, 'اقتصاد', 'اجتماعی'): 80, (0.5, 'اقتصاد', 'اجتماعی'): 80, (0.1, 'ادب و هنر', 'سیاسی'): 2, (0.3, 'ادب و هنر', 'سیاسی'): 2, (0.5, 'ادب و هنر', 'سیاسی'): 13, (0.1, 'ورزش', 'اجتماعی'): 10, (0.3, 'ورزش', 'اجتماعی'): 9, (0.5, 'ورزش', 'اجتماعی'): 18, (0.1, 'اجتماعی', 'اقتصاد'): 3, (0.3, 'اجتماعی', 'اقتصاد'): 3, (0.5, 'اجتماعی', 'اقتصاد'): 3, (0.1, 'سیاسی', 'ادب و هنر'): 15, (0.3, 'سیاسی', 'ادب و هنر'): 15, (0.5, 'سیاسی', 'ادب و هنر'): 13, (0.1, 'اجتماعی', 'ادب و هنر'): 12, (0.3, 'اجتماعی', 'ادب و هنر'): 11, (0.5, 'اجتماعی', 'ادب و هنر'): 3, (0.1, 'ورزش', 'اقتصاد'): 4, (0.3, 'ورزش', 'اقتصاد'): 4, (0.5, 'ورزش', 'اقتصاد'): 8, (0.1, 'ادب و هنر', 'اجتماعی'): 2, (0.3, 'ادب و هنر', 'اجتماعی'): 2, (0.5, 'ادب و هنر', 'اجتماعی'): 2, (0.1, 'سیاسی', 'اقتصاد'): 1, (0.3, 'سیاسی', 'اقتصاد'): 1, (0.5, 'سیاسی', 'اقتصاد'): 1}

Precision:

{(0.1, 'اقتصاد'): 0.2958057395143488, (0.1, 'سیاسی'): 0.6301369863013698, (0.1, 'ادب و هنر'): 0.2328767123287671, (0.1, 'اجتماعی'): 0.6304347826086957, (0.1, 'ورزش'): 0.3767441860465116, (0.3, 'اقتصاد'): 0.2537313432835821, (0.3, 'سیاسی'): 0.6388888888888888, (0.3, 'ادب و هنر'): 0.2977777777777777, (0.3, 'ورزش'): 0.3665158371040724, (0.5, 'اقتصاد'): 0.29977628635346754, (0.5, 'سیاسی'): 0.29977628635346754, (0.5, 'ادب و هنر'): 0.29977628635346754, (0.5, 'ورزش'): 0.29977628635346754}

0.5974025974025974, (0.5, 'ادب و هنر'): 0.26153846153846155, (0.5, 'اجتماعی'): 0.5918367346938775, (0.5, 'ورزش'): 0.36486486486486486

Recall:

{ (0.1, 'اقتصاد'): 0.3383838383838384, (0.1, 'سیاسی'): 0.2541436464088398, (0.1, 'ادب و هنر'): 0.17525773195876287, (0.1, 'اجتماعی'): 0.19205298013245034, (0.1, 'ورزش'): 0.28825622775800713, (0.3, 'اقتصاد'): 0.34096692111959287, (0.3, 'سیاسی'): 0.25555555555555554, (0.3, 'ادب و هنر'): 0.18681318681318682, (0.3, 'اجتماعی'): 0.1870967741935484, (0.3, 'ورزش'): 0.28222996515679444, (0.5, 'اقتصاد'): 0.3435897435897436, (0.5, 'سیاسی'): 0.24864864864864866, (0.5, 'ادب و هنر'): 0.19101123595505617, (0.5, 'اجتماعی'): 0.18831168831168832, (0.5, 'ورزش'): 0.28125 }

F-measure:

{ (0.1, 'اقتصاد'): 0.31566548881036516, (0.1, 'سیاسی'): 0.36220472440944884, (0.1, 'ادب و هنر'): 0.19999999999999998, (0.1, 'اجتماعی'): 0.29441624365482233, (0.1, 'ورزش'): 0.3266129032258065, (0.3, 'اقتصاد'): 0.31791221826809013, (0.3, 'سیاسی'): 0.36507936507936506, (0.3, 'ادب و هنر'): 0.21518987341772153, (0.3, 'اجتماعی'): 0.28292682926829266, (0.3, 'ورزش'): 0.3188976377952756, (0.5, 'اقتصاد'): 0.3201911589008363, (0.5, 'سیاسی'): 0.35114503816793896, (0.5, 'ادب و هنر'): 0.2207792207792208, (0.5, 'اجتماعی'): 0.28571428571428575, (0.5, 'ورزش'): 0.3176470588235294 }

Macro Average:

{(0.1, 'Macro_precision'): 0.43319968135993864, (0.1, 'Macro_recalls'): 0.24961888492837972, (0.1, 'Macro_f_measure'): 0.29977987202008854, (0.3, 'Macro_precision'): 0.4273827694108642, (0.3, 'Macro_recalls'): 0.25053248056773564, (0.3, 'Macro_f_measure'): 0.300001184765749, (0.5, 'Macro_precision'): 0.4230837889706538, (0.5, 'Macro_recalls'): 0.2505622633010273, (0.5, 'Macro_f_measure'): 0.29909535247716224}

Micro Average:

{(0.1, 'Micro_precision'): 0.4541951181356884, (0.1, 'Micro_recalls'): 0.26676712014509435, (0.1, 'micro_f_measure'): 0.31756273566522875, (0.3, 'Micro_precision'): 0.44584990425472687, (0.3, 'Micro_recalls'): 0.265971514595125, (0.3, 'micro_f_measure'): 0.3155838352520028, (0.5, 'Micro_precision'): 0.439024562342585, (0.5, 'Micro_recalls'): 0.2652623517527489, (0.5, 'micro_f_measure'): 0.3134761569186655}

Confusion Matrix:

{(0.1, 'ادب و هنر', 'tp'): 17, (0.3, 'ادب و هنر', 'tp'): 17, (0.5, 'ادب و هنر', 'tp'): 17, (0.1, 'ورزش', 'tp'): 119, (0.3, 'ورزش', 'tp'): 119, (0.5, 'ورزش', 'tp'): 118, (0.1, 'سیاسی', 'fn'): 116, (0.3, 'سیاسی', 'fn'): 116, (0.5, 'سیاسی', 'fn'): 116, (0.1, 'اجتماعی', 'fn'): 132, (0.3, 'اجتماعی', 'fn'): 132, (0.5, 'اجتماعی', 'fn'): 132, (0.1, 'اقتصاد', 'fn'): 118, (0.3, 'اقتصاد', 'fn'): 118, (0.5, 'اقتصاد', 'fn'): 118, (0.1, 'ادب و هنر', 'fn'): 41, (0.3, 'ادب و هنر', 'fn'): 41, (0.5, 'ادب و هنر', 'fn'): 41, (0.1, 'ورزش', 'fn'): 109, (0.3, 'ورزش', 'fn'): 109, (0.5, 'ورزش', 'fn'): 110, (0.1, 'سیاسی', 'tp'): 84, (0.3, 'سیاسی', 'tp'): 84, (0.5, 'سیاسی', 'tp'): 84, (0.1, 'اجتماعی', 'tp'): 93, (0.3, 'اجتماعی', 'tp'): 93, (0.5, 'اجتماعی', 'tp'): 93, (0.1, 'اقتصاد', 'tp'): 93, (0.3, 'اقتصاد', 'tp'): 93, (0.5, 'اقتصاد', 'tp'): 93, (0.1, 'ادب و هنر', 'tp'): 31, (0.3, 'ادب و هنر', 'tp'): 31, (0.5, 'ادب و هنر', 'tp'): 31}

اطلاعات کامل تر از این که در مواردی که سیستم به نادرستی تشخیص داده است. (چه دسته ایی را تشخیص داده است که به دین صورت می باشد: (counter =(label, real, delta)

{(0.1, 'سیاسی', 'ورزش'): 86, (0.3, 'سیاسی', 'ورزش'): 86, (0.5, 'سیاسی', 'ورزش'): 86, (0.1, 'اجتماعی', 'سیاسی'): 30, (0.3, 'اجتماعی', 'سیاسی'): 30, (0.5, 'اجتماعی', 'سیاسی'): 30, (0.1, 'اقتصاد', 'سیاسی'): 12, (0.3, 'اقتصاد', 'سیاسی'): 12, (0.5, 'اقتصاد', 'سیاسی'): 12, (0.1, 'ادب و هنر', 'سیاسی'): 21, (0.3, 'ادب و هنر', 'سیاسی'): 21, (0.5, 'ادب و هنر', 'سیاسی'): 21, (0.1, 'ورزش', 'اجتماعی'): 67, (0.3, 'ورزش', 'اجتماعی'): 67, (0.5, 'ورزش', 'اجتماعی'): 69, (0.1, 'ورزش', 'اقتصاد'): 104, (0.3, 'ورزش', 'اقتصاد'): 104, (0.5, 'ورزش', 'اقتصاد'): 105, (0.1, 'ورزش', 'ادب و هنر'): 13, (0.3, 'ورزش', 'ادب و هنر'): 13, (0.5, 'ورزش', 'ادب و هنر'): 13, (0.1, 'اجتماعی', 'ادب و هنر'): 18, (0.3, 'اجتماعی', 'ادب و هنر'): 18, (0.5, 'اجتماعی', 'ادب و هنر'): 16, (0.1, 'اقتصاد', 'ادب و هنر'): 16, (0.3, 'اقتصاد', 'ادب و هنر'): 16, (0.5, 'اقتصاد', 'ادب و هنر'): 16, (0.1, 'ادب و هنر', 'اقتصاد'): 3, (0.3, 'ادب و هنر', 'اقتصاد'): 3, (0.5, 'ادب و هنر', 'اقتصاد'): 3, (0.1, 'ادب و هنر', 'اجتماعی'): 3, (0.3, 'ادب و هنر', 'اجتماعی'): 3, (0.5, 'ادب و هنر', 'اجتماعی'): 3, (0.1, 'ادب و هنر', 'سیاسی'): 5, (0.3, 'ادب و هنر', 'سیاسی'): 5, (0.5, 'ادب و هنر', 'سیاسی'): 5, (0.1, 'ورزش', 'سیاسی'): 1, (0.3, 'ورزش', 'سیاسی'): 1, (0.5, 'ورزش', 'سیاسی'): 1, (0.1, 'اجتماعی', 'سیاسی'): 3, (0.3, 'اجتماعی', 'سیاسی'): 3, (0.5, 'اجتماعی', 'سیاسی'): 3, (0.1, 'اقتصاد', 'سیاسی'): 6, (0.3, 'اقتصاد', 'سیاسی'): 6, (0.5, 'اقتصاد', 'سیاسی'): 6, (0.1, 'ادب و هنر', 'سیاسی'): 1, (0.3, 'ادب و هنر', 'سیاسی'): 1, (0.5, 'ادب و هنر', 'سیاسی'): 1, (0.1, 'ورزش', 'اجتماعی'): 2, (0.3, 'ورزش', 'اجتماعی'): 2, (0.5, 'ورزش', 'اجتماعی'): 2, (0.1, 'ورزش', 'اقتصاد'): 1, (0.3, 'ورزش', 'اقتصاد'): 1, (0.5, 'ورزش', 'اقتصاد'): 1, (0.1, 'ورزش', 'ادب و هنر'): 1, (0.3, 'ورزش', 'ادب و هنر'): 1, (0.5, 'ورزش', 'ادب و هنر'): 1}

Precision:

{(0.1, 'اقتصاد'): 0.38271604938271603, (0.1, 'سیاسی'): 0.6363636363636364, (0.1, 'ادب و هنر'): 0.32075471698113206, (0.1, 'اجتماعی'): 0.8857142857142857, (0.1, 'ورزش'): 0.2236842105263158, (0.3, 'اقتصاد'): 0.384297520661157, (0.3, 'سیاسی'): 0.631578947368421, (0.3, 'ادب و هنر'): 0.3181818181818182, (0.3, 'ورزش'): 0.8857142857142857, (0.3, 'اجتماعی'): 0.2236842105263158, (0.3, 'اقتصاد'): 0.3163538873994638, (0.5, 'اقتصاد'): 0.38271604938271603, (0.5, 'سیاسی'): 0.631578947368421, (0.5, 'ادب و هنر'): 0.3181818181818182, (0.5, 'ورزش'): 0.8857142857142857, (0.5, 'اجتماعی'): 0.2236842105263158}

Recall:

{ (0.1, 'اقتصاد'): 0.34701492537313433, (0.1, 'سیاسی'): 0.5121951219512195, (0.1, 'ادب و هنر'): 0.1650485436893204, (0.1, 'اجتماعی'): 0.22794117647058823, (0.1, 'ورزش'): 0.3296398891966759, (0.3, 'اقتصاد'): 0.34831460674157305, (0.3, 'سیاسی'): 0.509090909090909, (0.3, 'ادب و هنر'): 0.17, (0.3, 'اجتماعی'): 0.22794117647058823, (0.3, 'ورزش'): 0.34701492537313433, (0.5, 'اقتصاد'): 0.3269230769230769, (0.5, 'سیاسی'): 0.509090909090909, (0.5, 'ادب و هنر'): 0.17, (0.5, 'اجتماعی'): 0.22794117647058823, (0.5, 'ورزش'): 0.3232876712328767 }

F-measure:

{ (0.1, 'اقتصاد'): 0.36399217221135033, (0.1, 'سیاسی'): 0.5675675675675675, (0.1, 'ادب و هنر'): 0.18681318681318684, (0.1, 'اجتماعی'): 0.3625730994152046, (0.1, 'ورزش'): 0.32513661202185795, (0.3, 'اقتصاد'): 0.36542239685658157, (0.3, 'سیاسی'): 0.5637583892617449, (0.3, 'ادب و هنر'): 0.19318181818181818, (0.3, 'اجتماعی'): 0.3625730994152046, (0.3, 'ورزش'): 0.32249322493224936, (0.5, 'اقتصاد'): 0.36399217221135033, (0.5, 'سیاسی'): 0.5637583892617449, (0.5, 'ادب و هنر'): 0.19318181818181818, (0.5, 'اجتماعی'): 0.3625730994152046, (0.5, 'ورزش'): 0.3197831978319783 }

Macro Average:

{(0.1, 'Macro_precision'): 0.48814771237189836, (0.1, 'Macro_recalls'): 0.3163679313361877, (0.1, 'Macro_f_measure'): 0.36121652760583345, (0.3, 'Macro_precision'): 0.4886913564903995, (0.3, 'Macro_recalls'): 0.31645395384522945, (0.3, 'Macro_f_measure'): 0.3614857857295197, (0.5, 'Macro_precision'): 0.4880094760782404, (0.5, 'Macro_recalls'): 0.31546693643350165, (0.5, 'Macro_f_measure'): 0.36065773538041923 }

Micro Average:

{(0.1, 'Micro_precision'): 0.5093143376672499, (0.1, 'Micro_recalls'): 0.34598173916221375, (0.1, 'micro_f_measure'): 0.3888158016638782, (0.3, 'Micro_precision'): 0.5084803879774393, (0.3, 'Micro_recalls'): 0.345192368771885, (0.3, 'micro_f_measure'): 0.38800955759635386, (0.5, 'Micro_precision'): 0.5076077616307789, (0.5, 'Micro_recalls'): 0.34390968822990103, (0.5, 'micro_f_measure'): 0.3869401813416962 }

Confusion Matrix:

{(0.1, 'ادب و هنر', 'fn'): 46, (0.1, 'ورزش', 'fn'): 157, (0.1, 'سیاسی', 'fn'): 147, (0.1, 'اجتماعی', 'fn'): 122, (0.1, 'سیاسی', 'tp'): 53, (0.1, 'اقتصاد', 'fn'): 125, (0.1, 'اجتماعی', 'tp'): 41, (0.1, 'اقتصاد', 'tp'): 86, (0.1, 'ورزش', 'tp'): 71, (0.3, 'ادب و هنر', 'fn'): 12, (0.3, 'ادب و هنر', 'fn'): 46, (0.3, 'ورزش', 'fn'): 154, (0.3, 'سیاسی', 'fn'): 148, (0.3, 'اجتماعی', 'fn'): 127, (0.3, 'اقتصاد', 'fn'): 133, (0.3, 'اجتماعی', 'tp'): 36, (0.3, 'اقتصاد', 'tp'): 78, (0.3, 'ورزش', 'tp'): 74, (0.5, 'ادب و هنر', 'fn'): 12, (0.5, 'سیاسی', 'tp'): 52, (0.5, 'ادب و هنر', 'fn'): 47, (0.5, 'ورزش', 'tp'): 68, (0.5, 'سیاسی', 'fn'): 160, (0.5, 'اجتماعی', 'fn'): 123, (0.5, 'اقتصاد', 'fn'): 136, (0.5, 'ورزش', 'fn'): 160, (0.5, 'اجتماعی', 'tp'): 40, (0.5, 'اقتصاد', 'tp'): 75, (0.5, 'سیاسی', 'tp'): 40, (0.5, 'ادب و هنر', 'tp'): 11}

اطلاعات کامل تر از این که در مواردی که سیستم به نادرستی تشخیص داده است. (چه دسته ایی را تشخیص داده است که به دین صورت می باشد: (counter =(label, real, delta)

{(0.1, 'ادب و هنر', 'سیاسی'): 8, (0.1, 'ورزش', 'اجتماعی'): 44, (0.1, 'سیاسی', 'ادب و هنر'): 28, (0.1, 'سیاسی', 'اجتماعی'): 30, (0.1, 'اجتماعی', 'ورزش'): 35, (0.1, 'اقتصاد', 'ورزش'): 50, (0.1, 'سیاسی', 'ورزش'): 46, (0.1, 'اقتصاد', 'اجتماعی'): 34, (0.1, 'ورزش', 'اقتصاد'): 53, (0.1, 'ادب و هنر', 'اقتصاد'): 23, (0.1, 'ادب و هنر', 'اجتماعی'): 10, (0.1, 'اقتصاد', 'ادب و هنر'): 19, (0.1, 'ادب و هنر', 'ورزش'): 5, (0.1, 'ورزش', 'سیاسی'): 39, (0.1, 'سیاسی', 'اقتصاد'): 43, (0.1, 'اجتماعی', 'اقتصاد'): 40, (0.1, 'اجتماعی', 'سیاسی'): 26, (0.1, 'اقتصاد', 'سیاسی'): 22, (0.1, 'ورزش', 'ادب و هنر'): 21, (0.1, 'اجتماعی', 'ادب و هنر'): 21, (0.3, 'ادب و هنر', 'سیاسی'): 10, (0.3, 'ورزش', 'سیاسی'): 42, (0.3, 'سیاسی', 'ادب و هنر'): 46, (0.3, 'ادب و هنر', 'سیاسی'): 31, (0.3, 'سیاسی', 'اجتماعی'): 40, (0.3, 'اجتماعی', 'ورزش'): 29, (0.3, 'اقتصاد', 'ورزش'): 46, (0.3, 'ورزش', 'اجتماعی'): 50, (0.3, 'سیاسی', 'ورزش'): 42, (0.3, 'اقتصاد', 'اجتماعی'): 45, (0.3, 'ورزش', 'ادب و هنر'): 19, (0.3, 'ورزش', 'اقتصاد'): 43, (0.3, 'ادب و هنر', 'اقتصاد'): 17, (0.3, 'ادب و هنر', 'اجتماعی'): 14, (0.3, 'اقتصاد', 'ادب و هنر'): 19, (0.3, 'اقتصاد', 'سیاسی'): 23, (0.3, 'سیاسی', 'اقتصاد'): 35, (0.3, 'اجتماعی', 'سیاسی'): 33, (0.3, 'اجتماعی', 'اقتصاد'): 37, (0.3, 'ادب و هنر', 'ورزش'): 5, (0.3, 'اجتماعی', 'ادب و هنر'): 28, (0.5, 'ادب و هنر', 'سیاسی'): 11, (0.5, 'سیاسی', 'ادب و هنر'): 29, (0.5, 'سیاسی', 'اجتماعی'): 48, (0.5, 'اجتماعی', 'ورزش'): 26, (0.5, 'اقتصاد', 'ورزش'): 45, (0.5, 'ورزش', 'اجتماعی'): 49, (0.5, 'سیاسی', 'ورزش'): 44, (0.5, 'اقتصاد', 'اجتماعی'): 50, (0.5, 'ورزش', 'اقتصاد'): 40, (0.5, 'ادب و هنر', 'اقتصاد'): 17, (0.5, 'ادب و هنر', 'اجتماعی'): 15, (0.5, 'ورزش', 'ادب و هنر'): 25, (0.5, 'ورزش', 'سیاسی'): 46, (0.5, 'سیاسی', 'اقتصاد'): 39, (0.5, 'اجتماعی', 'اقتصاد'): 38, (0.5, 'اجتماعی', 'سیاسی'): 31, (0.5, 'ادب و هنر', 'ورزش'): 4, (0.5, 'اقتصاد', 'ادب و هنر'): 16, (0.5, 'اجتماعی', 'ادب و هنر'): 28, (0.5, 'اقتصاد', 'سیاسی'): 25}

Precision:

{(0.1, 'اقتصاد'): 0.3510204081632653, (0.1, 'سیاسی'): 0.3581081081081081, (0.1, 'ادب و هنر'): 0.1188118811881188, (0.1, 'اجتماعی'): 0.2578616352201258, (0.1, 'ورزش'): 0.34299516908212563, (0.3, 'اقتصاد'): 0.37142857142857144, (0.3, 'سیاسی'): 0.325, (0.3, 'ادب و هنر'): 0.11009174311926606, (0.3, 'اجتماعی'): 0.1945945945945946, (0.3, 'ورزش'): 0.37755102040816324, (0.5, 'اقتصاد'): 0.3588516746411483, (0.5, 'سیاسی'): 0.19801980198019803, (0.5, 'ادب و هنر'): 0.26143790849673204, (0.5, 'ورزش'): 0.36363636363636365}

Recall:

{ (0.1, 'اقتصاد'): 0.3028169014084507, (0.1, 'سیاسی'): 0.2190082644628099, (0.1, 'ادب و هنر'): 0.24232081911262798, (0.3, 'اجتماعی'): 0.08888888888888889, (0.1, 'ورزش'): 0.17083333333333334, (0.3, 'اقتصاد'): 0.2943396226415094, (0.3, 'سیاسی'): 0.203125, (0.3, 'ادب و هنر'): 0.08391608391608392, (0.5, 'اجتماعی'): 0.13043478260869565, (0.3, 'ورزش'): 0.26811594202898553, (0.5, 'اقتصاد'): 0.2777777777777778, (0.5, 'سیاسی'): 0.14652014652014653, (0.5, 'ادب و هنر'): 0.07586206896551724, (0.5, 'اجتماعی'): 0.24372759856630824, (0.5, 'ورزش'): 0.14035087719298245 }

F-measure:

{ (0.1, 'اقتصاد'): 0.3251417769376181, (0.1, 'سیاسی'): 0.2717948717948718, (0.1, 'ادب و هنر'): 0.284, (0.3, 'اجتماعی'): 0.10169491525423728, (0.1, 'ورزش'): 0.2055137844611529, (0.3, 'اقتصاد'): 0.328421052631579, (0.3, 'سیاسی'): 0.25000000000000006, (0.3, 'ادب و هنر'): 0.09523809523809525, (0.5, 'اجتماعی'): 0.15618221258134488, (0.3, 'ورزش'): 0.31355932203389836, (0.5, 'اقتصاد'): 0.31315240083507306, (0.5, 'سیاسی'): 0.18779342723004697, (0.5, 'ادب و هنر'): 0.08661417322834647, (0.5, 'اجتماعی'): 0.2918454935622318, (0.5, 'ورزش'): 0.16427104722792607 }

Macro Average:

{(0.1, 'Macro_precision'): 0.28575944035234874, (0.1, 'Macro_recalls'): 0.20477364144122218, (0.1, 'Macro_f_measure'): 0.23762906968957598, (0.3, 'Macro_precision'): 0.2757331859101191, (0.3, 'Macro_recalls'): 0.1959862862390549, (0.3, 'Macro_f_measure'): 0.22868013649698354, (0.5, 'Macro_precision'): 0.2565726359894205, (0.5, 'Macro_recalls'): 0.17684769380454646, (0.5, 'Macro_f_measure'): 0.20873530841672486 }

Micro Average:

{(0.1, 'Micro_precision'): 0.3172236766797519, (0.1, 'Micro_recalls'): 0.2278448310887362, (0.1, 'micro_f_measure'): 0.26408481540084355, (0.3, 'Micro_precision'): 0.31111314098177456, (0.3, 'Micro_recalls'): 0.22091755533990398, (0.3, 'micro_f_measure'): 0.2578719508180939, (0.5, 'Micro_precision'): 0.28958699382519326, (0.5, 'Micro_recalls'): 0.19856072775664516, (0.5, 'micro_f_measure'): 0.23485397360441246 }

بررسی تاثیر مقادیر مختلف دلتا:

در مدل unigram: با بررسی‌هایی که انجام شد متوجه شدیم که افزایش مقدار دلتا در اکثر مواقع منجر به افزایش precision و recall شده است. هر چند که این افزایش ناچیز است (حدوداً کوچکتر از 0.01).

در مدل bigram: در این مدل با افزایش دلتا متوجه کاهش مقادیر precision و recall شدیم.