1 Text Classification and Feature Selection

1.3 χ^2 -Feature selection

$$\chi^2(\text{mathematics,composition}) = \frac{370 * (30 * 340 - 120 * 60)^2}{(30 + 120) * (30 + 60) * (120 + 340) * (60 + 340)} = 13.405797101$$

$$\chi^2(\text{mathematics,gravity}) = \frac{370 * (3 * 367 - 87 * 87)^2}{(3 + 87) * (3 + 87) * (367 + 87) * (376 + 87)} = 9.27139367$$

$$\chi^2(\text{mathematics,differential}) = \frac{370 * (50 * 320 - 40 * 50)^2}{(50 + 40) * (50 + 50) * (320 + 40) * (320 + 50)} = 60.49382716$$

$$\chi^2(\text{mathematics,theory}) = \frac{370 * (7 * 363 - 83 * 23)^2}{(7 + 83) * (7 + 23) * (363 + 83) * (363 + 23)} = 0,317943502$$

$$\chi^2(\text{chemistry,composition}) = \frac{370(43 * 337 - 107 * 7)^2}{(43 + 107) * (43 + 7) * (337 + 107) * (337 + 7)} = 60.995660207$$

$$\chi^2(\text{chemistry,gravity}) = \frac{370 * (-50 * 90)^2}{(370 + 50) * (370 + 90)} = 387810.55900621$$

$$\chi^2(\text{chemestry,differential}) = \frac{370 * (2 * 368 - 48 * 98)^2}{(2 + 48) * (2 + 98) * (368 + 48) * (368 + 98)} = 6.010295147$$

$$\chi^2(\text{chemistry,theory}) = \frac{370 * (5 * 365 - 45 * 25)^2}{(5 + 45) * (5 + 25) * (365 + 45) * (365 + 25)} = 0.755889097$$

$$\chi^2(\text{astronomy,composition}) = \frac{370 * (47 * 323 - 103 * 73)^2}{(47 + 103) * (47 + 73) * (323 + 103) * (323 + 73)} = 7.15333772$$

$$\chi^2(\text{astronomy,gravity}) = \frac{370 * (53 * 317 - 77 * 37)^2}{(53 + 77) * (53 + 37) * (317 + 77) * (317 + 37)} = 44,135628321$$

$$\chi^2(\text{astronomy,differential}) = \frac{370 * (19 * 351 - 111 * 81)}{(19 + 111) * (19 + 81) * (351 + 111) * (351 + 81)} = 0.768877373$$

$$\chi^2(\text{astronomy,theory}) = \frac{370 * (11 * 359 - 19 * 119)^2}{(11 + 19) * (11 + 119) * (359 + 19) * (359 + 119)} = 714.906766087$$

$$\chi^2(\text{physics,composition}) = \frac{370 * (30 * 340 - 120 * 70)}{(30 + 120) * (30 + 70) * (340 + 120) * (340 + 70)} = 0.412301166$$

$$\chi^2(\text{physics,gravity}) = \frac{370 * (34 * 336 - 56 * 56)^2}{(34 + 56) * (34 + 56) * (336 + 56) * (336 + 56)} = 20.419450743$$

$$\chi^2(\text{physics,differential}) = \frac{370 * (29 * 341 - 71 * 71)^2}{(29 + 71) * (29 + 71) * (341 + 71) * (341 + 71)} = 5.123096239$$

$$\chi^2(\text{physics,theory}) = \frac{370 * (7 * 363 - 23 * 93)^2}{(7 + 23) * (7 + 93) * (363 + 23) * (363 + 93)} = 0.113234933$$

$$\chi^2 avg(\text{composition}) = \frac{90}{370} * 13.41 + \frac{50}{370} * 61 + \frac{130}{370} * 7.15 + \frac{100}{370} * 0.41 = 15,221891892$$

This means compostion would be a good feature for chemisry

$$\chi^2 avg(\text{gravity}) = \frac{90}{370} * 9.27 + \frac{50}{370} * 387810 + \frac{130}{370} * 44.1 + \frac{100}{370} * 20.42 = 2311152,353918919$$

This means gravity would be a good feature for chemistry

$$\chi^2 avg(\text{differential}) = \frac{90}{370} * 60.5 + \frac{50}{370} * 6.01 + \frac{130}{370} * 0.77 + \frac{100}{370} * 5.12 = 17,182702703$$

This means differential would be a good feature for mathematics

$$\chi^2 avg(\text{theory}) = \frac{90}{370} * 0.32 + \frac{50}{370} * 0.76 + \frac{130}{370} * 714.9 + \frac{100}{370} * 0.11 = 251,391351351$$

This means theory would be a good feature for physics
Since we have no feature above the threshhold for astronomy the next best feature would be gravity. This is caused by the realy high value caused by 0 cases of gravity in the class chemistry.