**Data Evaluation:**

Data evaluation involves better understanding and visualisation of data for finding out patterns. By performing descriptive analysis on data help us to build right predictive model.

For Data Evaluation we used Weka and Excel

1) Out of total 1054 intense in Class 326 is No and 728 is Yes.

2) Attribute A1 has 460 intense for score 0 where 264 is belongs to no and 196 belongs to yes class and score 1 has 594 intense where 62 belongs to no and 532 belongs to yes class.

3) Attribute A2 has 581 intense for score 0 where 289 are belong to no and 292 belong to yes class and score 1 has 473 intense where 34 belong to no and 439 belong to yes class.

4) Attribute A3 has 631 intense for score 0 where 293 are belong to no and 338 belong to yes class and score 1 has 423 intense where 33 belong to no and 390 belong to yes class.

5) Attribute A4 has 514 intense for score 0 where 282 are belong to no and 232 belong to yes class and score 1 has 540 intense where 44 belong to no and 496 belong to yes class.

6) Attribute A5 has 601 intense for score 0 where 292 are belong to no and 209 belong to yes class and score 1 has 553 intense where 34 belong to no and 519 belong to yes class.

7) Attribute A6 has 446 intense for score 0 where 275 are belong to no and 171 belong to yes class and score 1 has 608 intense where 52 belong to no and 556 belong to yes class.

8) Attribute A7 has 369 intense for score 0 where 245 are belong to no and 124 belong to yes class and score 1 has 685 intense where 81 belong to no and 604 belong to yes class.

9) Attribute A8 has 570 intense for score 0 where 280 are belong to no and 290 belong to yes class and score 1 has 484 intense where 46 belong to no and 438 belong to yes class.

10) Attribute A9 has 538 intense for score 0 where 307 are belong to no and 231 belong to yes class and score 1 has 516 intense where 19 belong to no and 497 belong to yes class.

11) Attribute A10 has 436 intense for score 0 where 178 are belong to no and 258 belong to yes class and score 1 has 618 intense where 148 belong to no and 470 belong to yes class.

12) Attribute sex has 319 intense for Female where 125 are belong to no and 194 belong to yes class and male has 735 intense where 201 belong to no and 534 belong to yes class.

13) Attribute Jaundice has 288 intense for Yes where 73 are belong to no and 215 belong to yes class and No has 766 intense where 253 belong to no and 513 belong to yes class.

14) Attribute Family with ASD has 884 intense for No, where 271 are belong to no and 613 belong to yes class and Yes has 170 intense where 55 belong to no and 116 belong to yes class.

15) ) Attribute Who completed the test has 1018 intense for family member, where 316 are belong to no and 702 belong to yes class ,For Health Care Professional total 33 intense where no class has 10 and yes class has 23 intense and For Others total has 3 intense all belong to yes class.

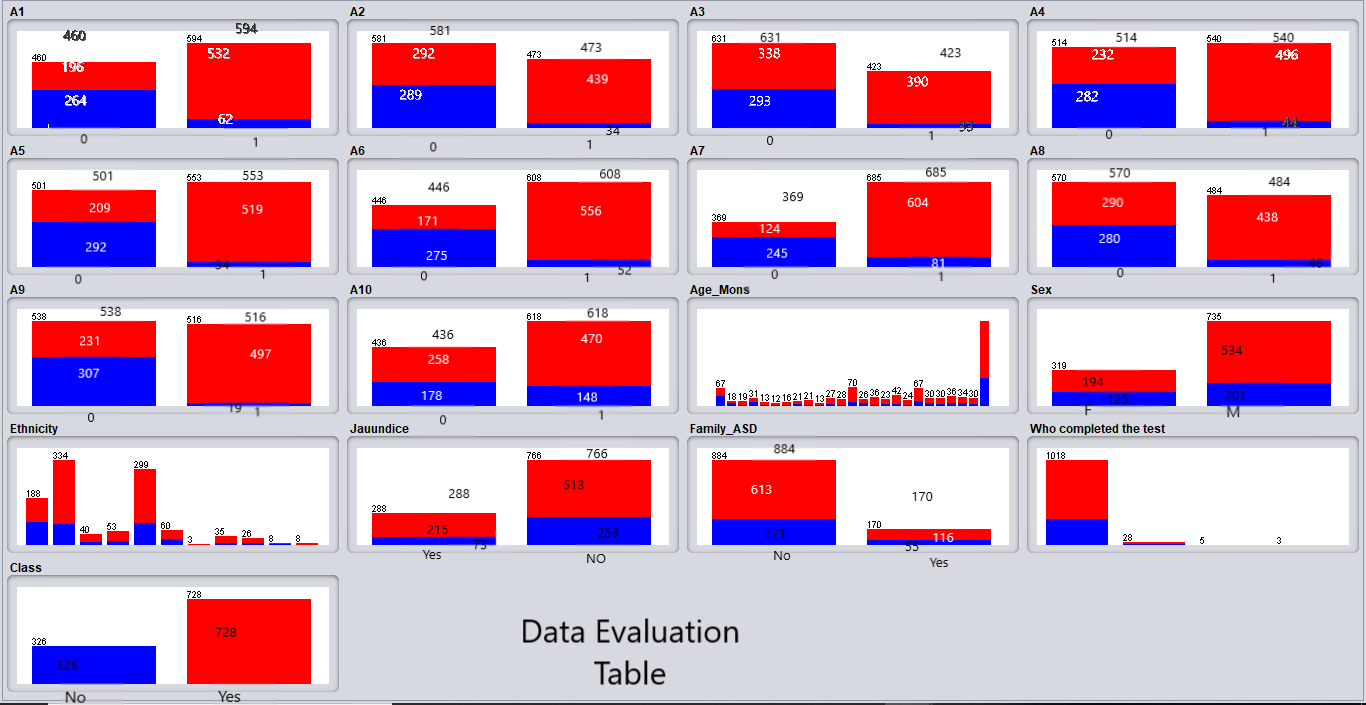
16) Attribute Ethnicity out of total 1054 observations break down is given under in table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Class | |  |
|  | No | Yes | total |
| asian | 87 | 212 | 299 |
| black | 14 | 39 | 53 |
| Hispanic | 10 | 30 | 40 |
| Latino | 6 | 20 | 26 |
| middle eastern | 92 | 96 | 188 |
| mixed | 3 | 5 | 8 |
| Native Indian | 0 | 3 | 3 |
| Others | 6 | 29 | 35 |
| Pacifica | 1 | 7 | 8 |
| south asian | 23 | 37 | 60 |
| White European | 84 | 250 | 334 |
| Total | 326 | 728 | 1054 |

17) Attribute Age\_Mons Total 1054 observation break down is below

|  |  |  |  |
| --- | --- | --- | --- |
|  | Class | |  |
| Age in Months | No | Yes | total |
| 12 | 37 | 30 | 67 |
| 13 | 10 | 8 | 18 |
| 14 | 3 | 16 | 19 |
| 15 | 16 | 15 | 31 |
| 16 | 2 | 11 | 13 |
| 17 | 5 | 7 | 12 |
| 18 | 4 | 12 | 16 |
| 19 | 8 | 13 | 21 |
| 20 | 4 | 17 | 21 |
| 21 | 5 | 8 | 13 |
| 22 | 6 | 21 | 27 |
| 23 | 5 | 23 | 28 |
| 24 | 15 | 55 | 70 |
| 25 | 11 | 15 | 26 |
| 26 | 5 | 31 | 36 |
| 27 | 6 | 17 | 23 |
| 28 | 9 | 33 | 42 |
| 29 | 4 | 20 | 24 |
| 30 | 19 | 48 | 67 |
| 31 | 10 | 20 | 30 |
| 32 | 8 | 22 | 30 |
| 33 | 11 | 25 | 36 |
| 34 | 10 | 24 | 34 |
| 35 | 6 | 24 | 30 |
| 36 | 107 | 213 | 320 |
| Total | 326 | 728 | 1054 |

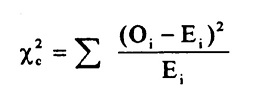
By Data Evaluation table it can be more easily visualize



**Feature Evaluation**:

Finding of correlation between dependent variable and independent variable we choose chi sq. test. Chi-square test is one of the most useful testing methods when variables are nominal or categorical. We want to test whether the incidence of a specific feature and the incidence of specific class are independent or not. If two feature is dependent than we can use the incidence of the feature to predict the incidence of the class.

Formula for chi-square test

The (Hopkins., 2008) chi-square 1

Where

c= the degrees of freedom(Df)

O= observed value

E = expected value

Df=(total row-1)(total col-1)[[1]](#endnote-1)

R Studio is used for conducting Chi-Square test.

Chi Square Test Table for Class Vs. Every variable

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | X-squared Value | Degree Of Freedom(df) | p-value |
| A1 | 265.34 | 1 | < 2.2e-16 |
| A2 | 224.39 | 1 | < 2.2e-16 |
| A3 | 175.12 | 1 | < 2.2e-16 |
| A4 | 266.83 | 1 | < 2.2e-16 |
| A5 | 332 | 1 | < 2.2e-16 |
| A6 | 339.26 | 1 | < 2.2e-16 |
| A7 | 331.75 | 1 | < 2.2e-16 |
| A8 | 190.46 | 1 | < 2.2e-16 |
| A9 | 348.82 | 1 | < 2.2e-16 |
| A10 | 33.301 | 1 | = 7.894e-09 |
| Age\_Mons | 56.625 | 24 | = 0.0001883 |
| Sex | 14.044 | 1 | = 0.0001786 |
| Ethnicity | 43.571 | 10 | = 3.93e-06 |
| Jauundice | 5.427 | 1 | = 0.01983 |
| Family\_ASD | 0.12094 | 1 | **= 0.728** |
| Who.completed.the.test | 1.3554 | 2 | **= 0.5078** |

**Result Analysis of Chi-square test:**

Setting up p-value >= 0.05 for accepting Null hypothesis means there is no association between variables

P-value <0.05 means rejecting Null hypothesis and accepting alternating hypothesis means there is association or relationship between variables.

So for ASD toddler data set by p-value we can say that variable family with ASD and who completed the test has no association or correlation with Class. By chi-sq. test we can say that these two variables are not making any significance for predicting class. Thus for building predictive model these two variables can be dropped.

1. (Hopkins., 2008) [↑](#endnote-ref-1)