Future work

- Modification of proposed algorithm to find the patterns which are **favourable** and to **discover the rare rules**. Rare rules are prone to have **low statistical occurence**. **Weakly Associated Pattern Tree** may help in discovering useful patterns that are rare.
- Find interesting patterns utilizing **different constraints** on **selection of items at each level**. These constraints would be **utility based**, **similarity distance based** (similarity between candidate item and the parent item), **mutual entropy**, **fuzzy relations**, and so on.
- The proposed approach can be studied under the big data analytics. The **current implementation** requires the data to reside in the **memory** (This limits the amount of data to be processed). This problem can be addressed by **utilizing disk memory** (computational time goes up !). What we can do? **Algorithms can be modified for multi-threaded computation.**

Connect dataset
67557 obs. of 43 variables

Pumsb dataset
49046 obs. of 74 variables

Diabetes dataset
101766 obs. of 50 variables

Assocation Rule Mining in R

```
> head(titanic.raw)
  Class  Sex   Age Survived
1   3rd Male Child       No
2   3rd Male Child       No
3   3rd Male Child       No
4   3rd Male Child       No
5   3rd Male Child       No
6   3rd Male Child       No

> library(arules)
> rules <- apriori(titanic.raw, control = list(verbose = F))
Apriori
   lhs                                rhs           support   confidence lift
1  {}                              => {Age=Adult}   0.9504771 0.9504771  1.0000000
2  {Class=2nd}                     => {Age=Adult}   0.1185825 0.9157895  0.9635051
3  {Class=1st}                     => {Age=Adult}   0.1449341 0.9815385  1.0326798
4  {Sex=Female}                    => {Age=Adult}   0.1930940 0.9042553  0.9513700
5  {Class=3rd}                     => {Age=Adult}   0.2848705 0.8881020  0.9343750
6  {Survived=Yes}                  => {Age=Adult}   0.2971377 0.9198312  0.9677574
7  {Class=Crew}                    => {Sex=Male}    0.3916402 0.9740113  1.2384742
8  {Class=Crew}                    => {Age=Adult}   0.4020900 1.0000000  1.0521033
9  {Survived=No}                   => {Sex=Male}    0.6197183 0.9154362  1.1639949
10 {Survived=No}                   => {Age=Adult}   0.6533394 0.9651007  1.0153856
11 {Sex=Male}                      => {Age=Adult}   0.7573830 0.9630272  1.0132040
12 {Sex=Female,Survived=Yes}       => {Age=Adult}   0.1435711 0.9186047  0.9664669
13 {Class=3rd,Sex=Male}            => {Survived=No} 0.1917310 0.8274510  1.2222950
14 {Class=3rd,Survived=No}         => {Age=Adult}   0.2162653 0.9015152  0.9484870
15 {Class=3rd,Sex=Male}            => {Age=Adult}   0.2099046 0.9058824  0.9530818
16 {Sex=Male,Survived=Yes}         => {Age=Adult}   0.1535666 0.9209809  0.9689670
17 {Class=Crew,Survived=No}        => {Sex=Male}    0.3044071 0.9955423  1.2658514
18 {Class=Crew,Survived=No}        => {Age=Adult}   0.3057701 1.0000000  1.0521033
19 {Class=Crew,Sex=Male}           => {Age=Adult}   0.3916402 1.0000000  1.0521033
20 {Class=Crew,Age=Adult}          => {Sex=Male}    0.3916402 0.9740113  1.2384742
21 {Sex=Male,Survived=No}          => {Age=Adult}   0.6038164 0.9743402  1.0251065
22 {Age=Adult,Survived=No}         => {Sex=Male}    0.6038164 0.9242003  1.1751385
23 {Class=3rd,Sex=Male,Survived=No} => {Age=Adult}  0.1758292 0.9170616  0.9648435
```

```
24 {Class=3rd,Age=Adult,Survived=No}  => {Sex=Male}    0.1758292 0.8130252  1.0337773
25 {Class=3rd,Sex=Male,Age=Adult}     => {Survived=No} 0.1758292 0.8376623  1.2373791
26 {Class=Crew,Sex=Male,Survived=No}  => {Age=Adult}   0.3044071 1.0000000  1.0521033
27 {Class=Crew,Age=Adult,Survived=No} => {Sex=Male}    0.3044071 0.9955423  1.2658514
```

```
rules <- apriori(titanic.raw, parameter = list(minlen = 2, supp = 0.005, conf = 0.8),
+ appearance = list(rhs = c("Survived=No", "Survived=Yes"), default = "lhs"),
+ control = list(verbose = F))
```

```
> inspect(rules)
   lhs                                rhs            support     confidence lift
1  {Class=2nd,Age=Child}           => {Survived=Yes} 0.010904134 1.0000000  3.095640
2  {Class=2nd,Sex=Female}          => {Survived=Yes} 0.042253521 0.8773585  2.715986
3  {Class=2nd,Sex=Male}            => {Survived=No}  0.069968196 0.8603352  1.270871
4  {Class=1st,Sex=Female}          => {Survived=Yes} 0.064061790 0.9724138  3.010243
5  {Class=Crew,Sex=Female}         => {Survived=Yes} 0.009086779 0.8695652  2.691861
6  {Class=3rd,Sex=Male}            => {Survived=No}  0.191731031 0.8274510  1.222295
7  {Class=2nd,Sex=Female,Age=Child}  => {Survived=Yes} 0.005906406 1.0000000  3.095640
8  {Class=2nd,Sex=Female,Age=Adult}  => {Survived=Yes} 0.036347115 0.8602151  2.662916
9  {Class=2nd,Sex=Male,Age=Adult}    => {Survived=No}  0.069968196 0.9166667  1.354083
10 {Class=1st,Sex=Female,Age=Adult}  => {Survived=Yes} 0.063607451 0.9722222  3.009650
11 {Class=Crew,Sex=Female,Age=Adult} => {Survived=Yes} 0.009086779 0.8695652  2.691861
12 {Class=3rd,Sex=Male,Age=Adult}    => {Survived=No}  0.175829169 0.8376623  1.237379
```

```
> subset.matrix <- is.subset(rules, rules)
> subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
> redundant <- colSums(subset.matrix, na.rm=T) >= 1
> rules.pruned <- rules[!redundant]
> which(redundant)
7    8    9   10   11   12
> inspect(rules.pruned)
  lhs                       rhs            support     confidence lift
1 {Class=2nd,Age=Child}   => {Survived=Yes} 0.010904134 1.0000000  3.095640
2 {Class=2nd,Sex=Female}  => {Survived=Yes} 0.042253521 0.8773585  2.715986
3 {Class=2nd,Sex=Male}    => {Survived=No}  0.069968196 0.8603352  1.270871
4 {Class=1st,Sex=Female}  => {Survived=Yes} 0.064061790 0.9724138  3.010243
5 {Class=Crew,Sex=Female} => {Survived=Yes} 0.009086779 0.8695652  2.691861
6 {Class=3rd,Sex=Male}    => {Survived=No}  0.191731031 0.8274510  1.222295
```