

Estimation of Frequently Effectuated Diseases using Improved FP-Tree

Yamala Sandhya¹, Dr. Halavath Balaji²

¹ Research scholar in Computer Science Engineering, Sreenidhi Institute of Science and Technology, Hyderabad (T.G), India

² Computer Science Engineering, Sreenidhi Institute of Science and Technology, Hyderabad (T.G), India.

Abstract : The population of India is 1/5th of the total world population. The country India is the second largest country in the population. The each and every state of India has different climatic situations. Based upon the climatic changes and many factors the diseases in India are growing day by day. The health of people in India is the major issue in now-a-days. There are so many inequalities between the states about the health. Health care system in India is mainly controlled or administrated by the respective state governments. To improve the health of the population in each and every state we require collecting the data from the state government and trends. The dataset consists of 30 states and union territory states except the Delhi. The experiment result shows the diseases which are mostly effect the people across the India. In this research we use the one of the data mining technique to produce the mostly effectuated diseases.

IndexTerms - Data Mining, Improved FP-Tree, Support Count, Pruning.

I. INTRODUCTION

India has 30 states and 7 union territories. All of them have population larger than the other countries except China. India has the large amount of population. The death of people in India is increased day by day. There are many reasons for the cause of death. All of these factors affect the health outcomes of India. In this research paper we take the some of the reasons of death all over the India. By using all of these data we can find the main reasons or diseases for the death of people. So, we can take the appropriate methods to control the spread of the diseases. The data mining technique of Improved FP-Tree is finding the much more suitability in circumstances that are required to interrogate for the independent state of relations. The Improved FP-tree is used to produce the frequently effectuated diseases all over the India. This technique can be used to interrogate the mostly effectuated diseases in India.

II. RELATED WORK

The Improved FP-Tree produces the recurring diseases in the given data. By using Improved FP-Tree we can discover the associations and correlations among the data set. The data mining has many techniques to extract the data from the large amount of data. The data mining has the some techniques to find the repeated data of the given data sets. There are two algorithms to finding the frequently produced data one is Apriori and another one is FP-Tree. The Apriori algorithm is earliest methods which find the frequently produced data by generating the candidates. The Improved FP-Tree finds the frequently produced data without using the candidate generation. The Improved FP-Tree uses the divide-and-conquer method to produce the frequently used data. At first the Improved FP-Tree scans the database only once and then compress the dataset into a tree called as the Improve FP-Tree which consist the information about the association of each item set. The algorithm divides the large amount of database into the number of databases based upon the condition; in which every item of the database has the association with at least one frequent item or also known as the "pattern fragment", and then it mines the each and every such database separately.

III. PROPOSED RESEARCH WORK

In this research paper, an algorithm of Improvised FP-growth is used which is an advanced form of Frequent Pattern-growth algorithm. Improved FP-growth algorithm the structure of the node is modified. It has four attributes of item name, count, node link and flag. Here, node link plays an act of connect nodes with the similar item. It helps to search for particular node quickly. Thus the proposed algorithm quickly traverses the tree in relation than the Frequent Pattern-growth algorithm. It reduces the runtime and memory of IFP-growth algorithm. IFP-growth algorithm shortens the database into IFP-tree. It also maintains the data between item sets. IFP-growth algorithm uses the divide and conquers method, which uses the two stages: they are build and mine. Structure of proposed IFP-growth algorithm is as follows:

An Improved FP-tree consist the three attributes, they are:

- i. Root node.
- ii. Child node.
- iii. Header Table.

Attributes of each and every node are:

- i. Item Name :The names of the item are stored in this field.
- ii. Count : Number of times the item occurs is updated in this field.
- iii. Node Link : Node link that points the node that contains the same item name.
- iv. Flag :The branch information is noted in this field. Every entry of the item in the header table consists of two Characteristics: Item-name and Head of node-link.

IV. EXAMPLE VALIDATION

All the diseases details of every state in India are displayed in Table 1. The table consist all the states in India and each and every state has many number of diseases. The below table consist thirty states of India and another place that is UTs other then Delhi. So, the table consist thirty one entries.

| | |
|----------------------|---|
| Chhattisgarh | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Stroke, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Diabetes, Other neonatal. |
| Bihar | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Low back & neck pain, Diabetes, Other neonatal. |
| Jharkhand | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Madhya Pradesh | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Odisha | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Stroke, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Rajasthan | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Uttar Pradesh | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Other neonatal. |
| Uttarakhand | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Road injuries, Self-harm, Diabetes, Other neonatal. |
| Meghalaya | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Self-harm, Other neonatal. |
| Assam | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Stroke, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Diabetes, Other neonatal. |
| Arunachal Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Self-harm, Other neonatal. |
| Mizoram | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Other neonatal. |
| Nagaland | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Tuberculosis, Other neonatal. |
| Tripura | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Stroke, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Sikkim | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Manipur | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Diabetes, Other neonatal. |
| Gujarat | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Other neonatal. |
| Haryana | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Road injuries, Self-harm, Diabetes, Other neonatal. |
| Delhi | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Diabetes, Other neonatal. |
| Telangana | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Diabetes, Other neonatal. |
| Andhra Pradesh | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Diabetes, Other neonatal. |
| Jammu and Kashmir | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Road injuries, Other neonatal. |
| Karnataka | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Diabetes, Other neonatal. |
| West Bengal | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Stroke, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Maharashtra | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Diabetes, Other neonatal. |
| UTs other than Delhi | Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Diabetes, Other neonatal. |
| Himachal Pradesh | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Other neonatal. |
| Punjab | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Road injuries, Diabetes, Other neonatal. |
| Goa | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Lower respiratory infections, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Low back & neck pain, Diabetes, Other neonatal. |
| Tamil Nadu | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis, Sense organ diseases, Road injuries, Self-harm, Diabetes. |
| Kerala | Ischaemic heart disease, Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Preterm birth complications, Tuberculosis, Sense organ diseases, Self-harm, Low back & neck pain, Diabetes. |

Table 1: List of diseases of India according to state wise
The database is scanned once and counts the frequently occurred diseases in each and every state. The count is known as the support count. The Table 2 shows the support count of every disease.

| Items | Support count |
|---------------------------------------|---------------|
| Ischaemic heart disease | 21 |
| Chronic obstructive pulmonary disease | 31 |
| Diarrhoeal diseases | 31 |
| Lower respiratory infections | 10 |
| Stroke | 5 |
| Iron deficiency anemia | 31 |
| Preterm birth conditions | 13 |
| Tuberculosis | 31 |
| Sense organ diseases | 28 |
| Road injuries | 5 |
| Self-harm | 24 |
| Low back&neck pain | 8 |
| Diabetes | 13 |
| Other neonatal | 29 |

Table 2: Support count of diseases

| Items | Support count |
|---------------------------------------|---------------|
| Chronic obstructive pulmonary disease | 31 |
| Diarrhoeal diseases | 31 |
| Iron deficiency anemia | 31 |
| Tuberculosis | 31 |
| Other neonatal | 29 |
| Sense organ diseases | 28 |
| Self-harm | 24 |
| Ischaemic heart disease | 21 |
| Diabetes | 16 |
| Preterm birth conditions | 13 |
| Lower respiratory infections | 10 |
| Low back&neck pain | 8 |
| Stroke | 5 |
| Road injuries | 5 |

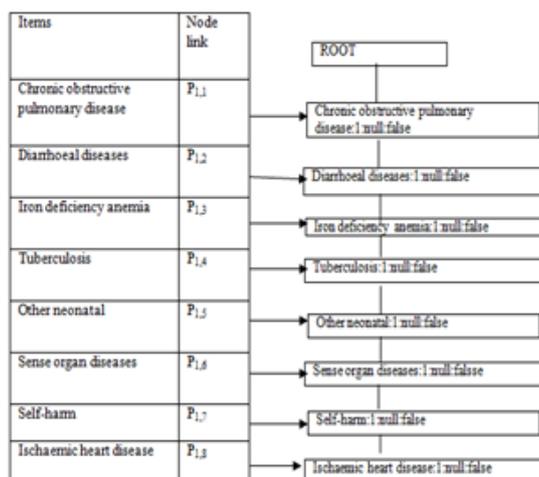
Table 3: Rearrangement of diseases support count in descending order

| | |
|----------------------|---|
| Chhattisgarh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Bihar | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Ischaemic heart disease. |
| Jharkhand | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Madhya Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Odisha | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm. |
| Rajasthan | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Uttar Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Uttarakhand | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Meghalaya | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Self-harm. |
| Assam | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm. |
| Arunachal Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Self-harm. |
| Mizoram | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease. |
| Nagaland | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal. |
| Tripura | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Sikkim | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm. |
| Manipur | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm. |
| Gujarat | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Haryana | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Delhi | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease. |
| Telangana | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Andhra Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Jammu and Kashmir | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Ischaemic heart disease. |
| Karnataka | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| West Bengal | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Maharashtra | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| UTs other than Delhi | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm. |
| Himachal Pradesh | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Punjab | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Ischaemic heart disease. |
| Goa | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Ischaemic heart disease. |
| Tamil Nadu | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Sense organ disease, Self-harm, Ischaemic heart disease. |
| Kerala | Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Sense organ disease, Self-harm, Ischaemic heart disease. |

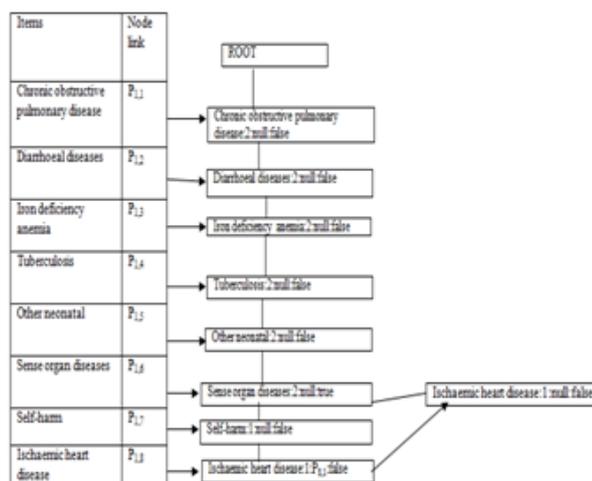
Table 4: List of diseases of every state after pruning

Now we have to arrange the diseases in the descending order with respect to the support count. Table 3 shows the details of diseases after rearrangement of support count in descending order. The database has the highest support count 31 and the lowest support count 5. In this example we take the minimum support count is 17. So, the items which have the support count less than 17 are pruned. In the above table the diseases Diabetes, Preterm birth conditions, Lower respiratory infections, Low back & neck pain, Stroke and Road injuries are pruned because these diseases are have the support count less than 17. After pruning we have to rearrange the diseases of each and every state. Table 4 shows the rearrangement of states after pruning. After finishing of the pruning process we have to construct the Tree for the given database. The below steps shows the construction of IFP-Tree for the given example.

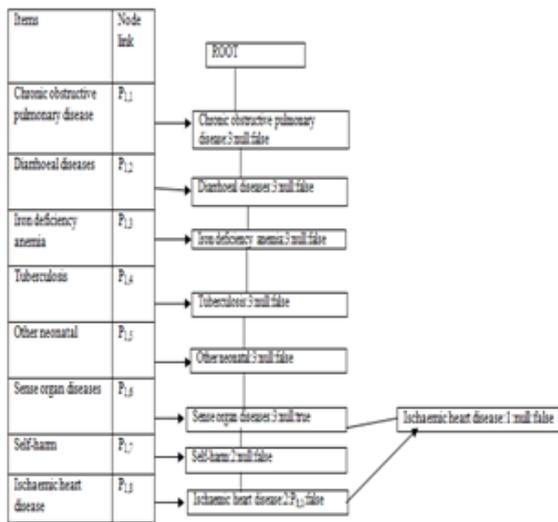
Step 1: insertion of state Chhattisgarh.



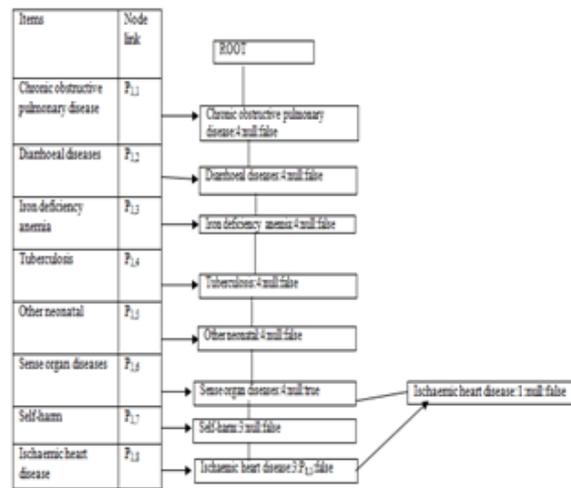
Step 2: insertion of state Bihar.



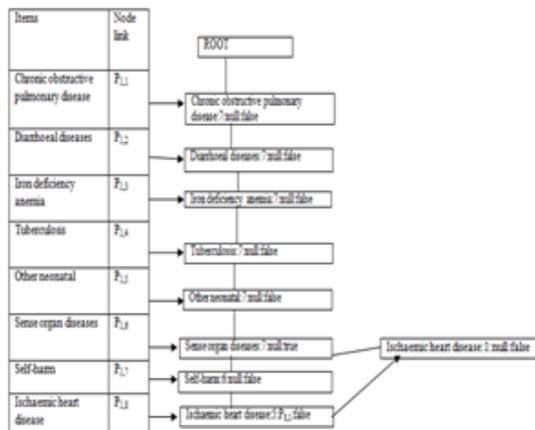
Step 3 : insertion of state Jharkhand.



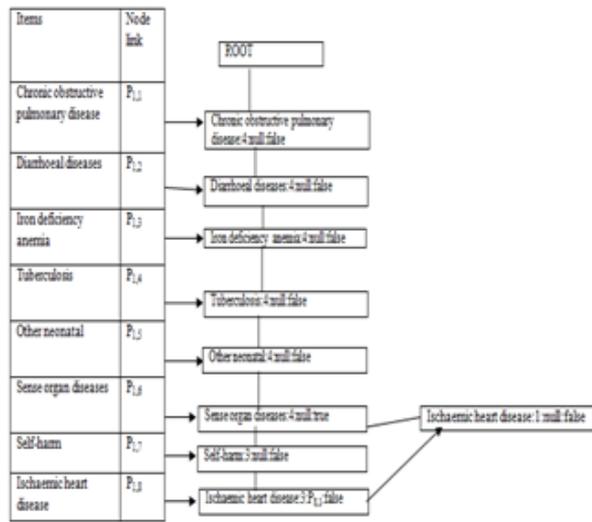
Step 5 : insertion of state Odisha.



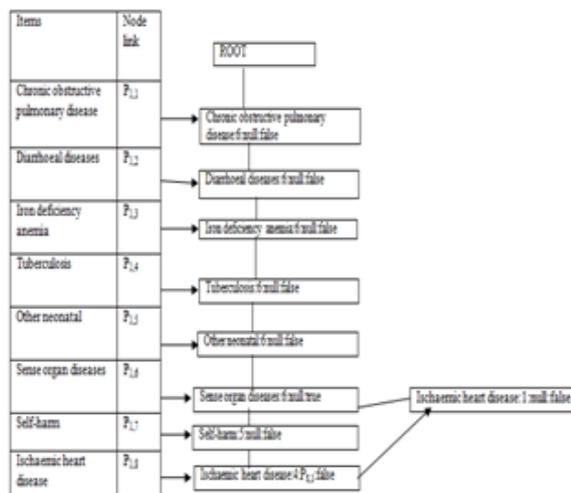
Step 7 : insertion of state Uttarpradesh.



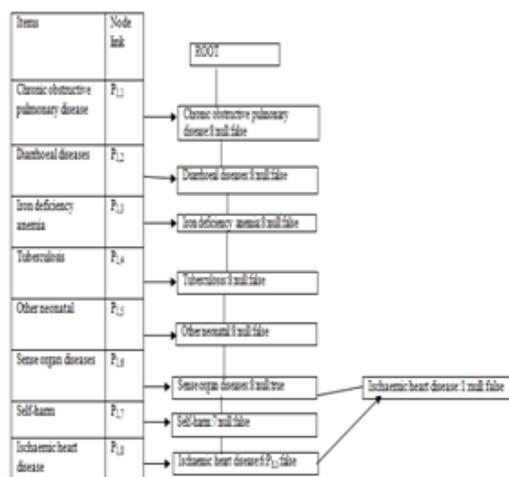
Step 4 : insertion of state Madhyapradesh.



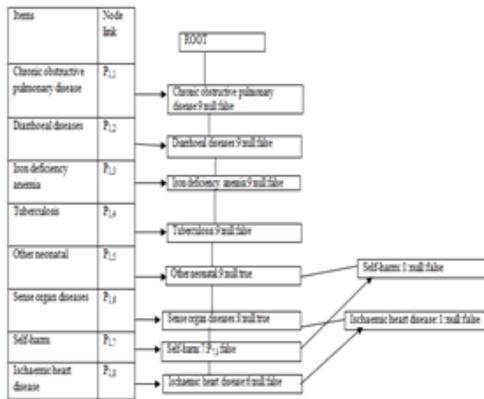
Step 6 : insertion of state Rajasthan.



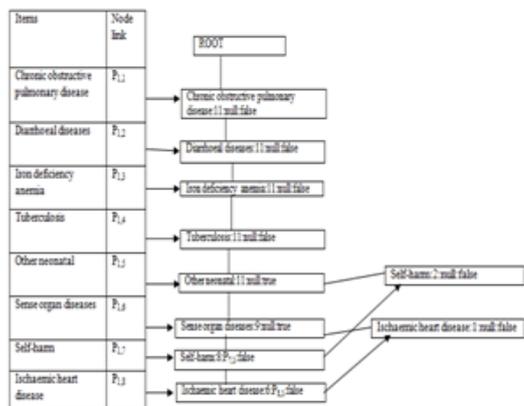
Step 8 : insertion of state Uttarakhand.



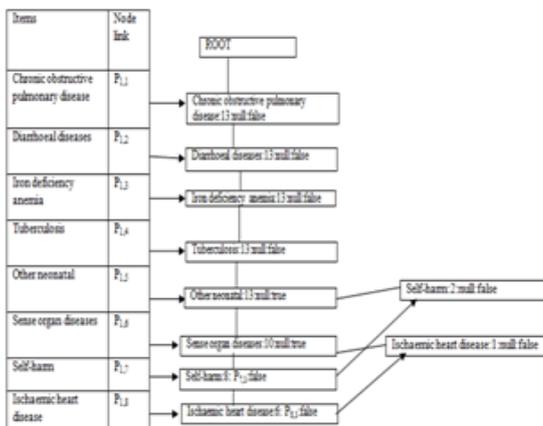
Step 9: insertion of state Meghalaya.



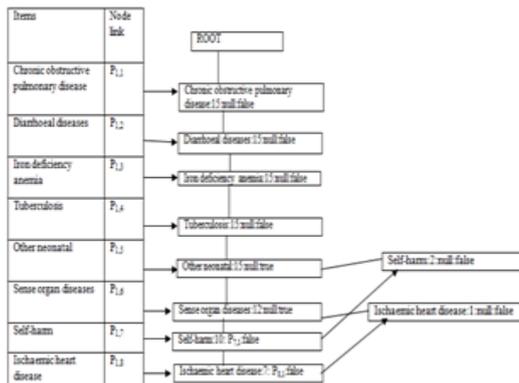
Step 11: insertion of state Arunachal Pradesh.



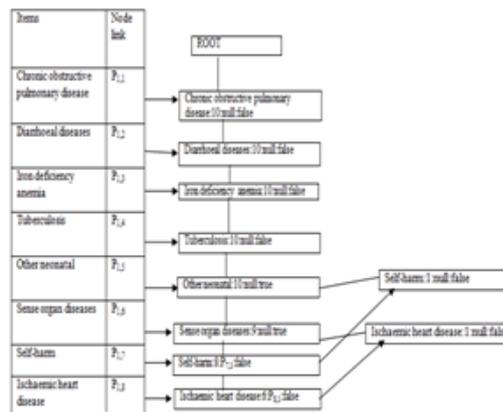
Step 13: insertion of state Nagaland.



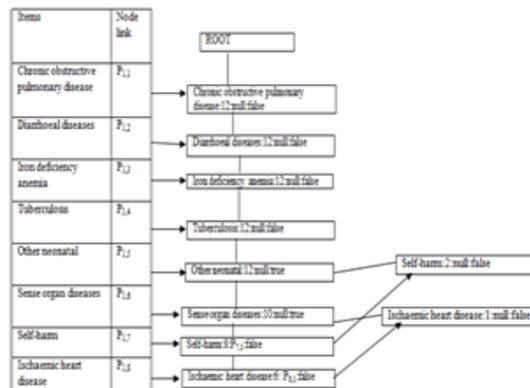
Step 15: insertion of state Sikkim.



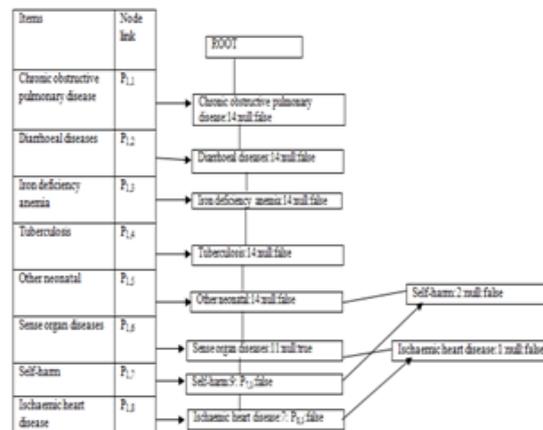
Step 10: insertion of state Assam.



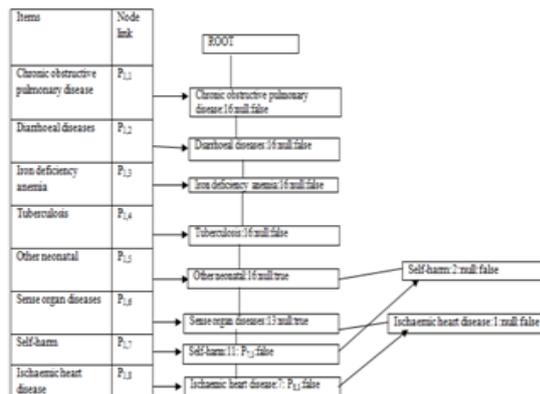
Step 12: insertion of state Mizoram.



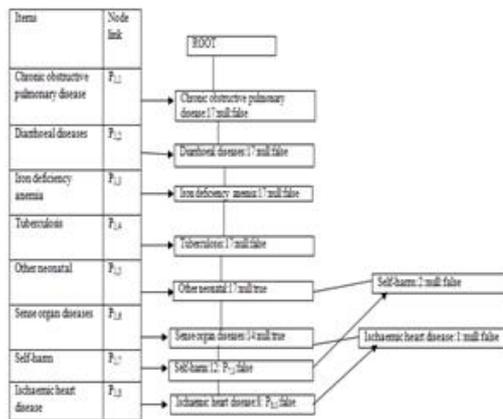
Step 14: insertion of state Tripura.



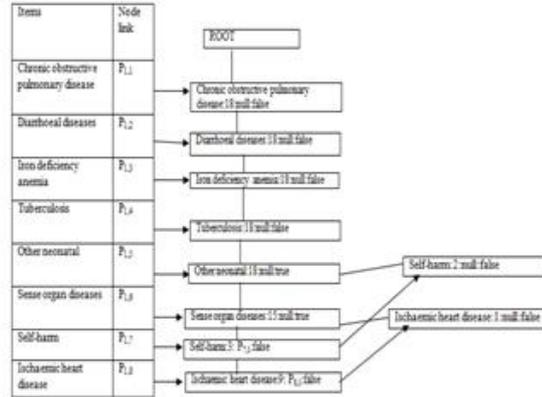
Step 16: insertion of state Manipur.



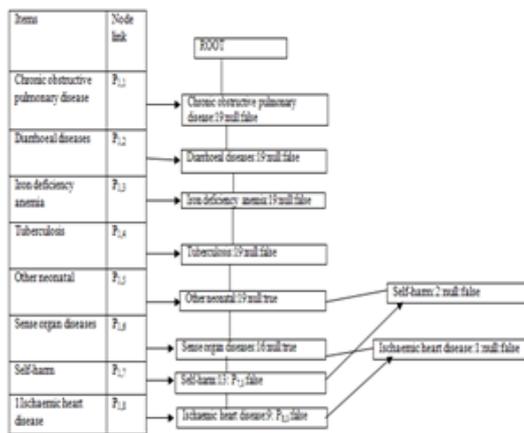
Step 17: insertion of state Gujarat.



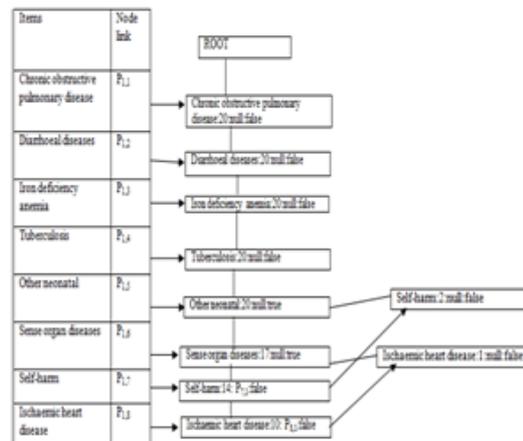
Step 18: insertion of state Haryana.



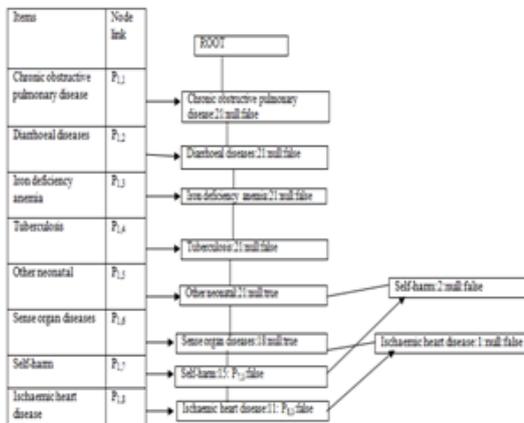
Step 19: insertion of state Delhi.



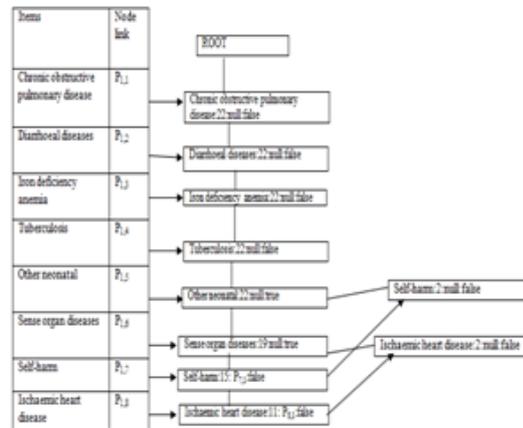
Step 20: insertion of state Telangana.



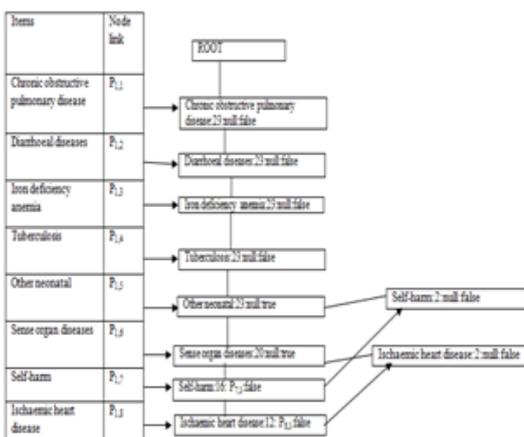
Step 21: insertion of state Andhra Pradesh.



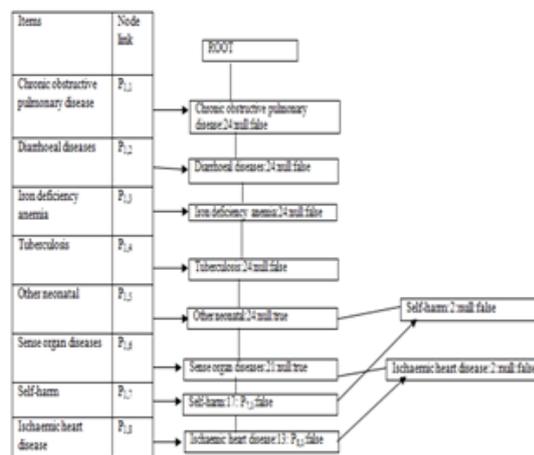
Step 22: insertion of state Jammu & Kashmir.



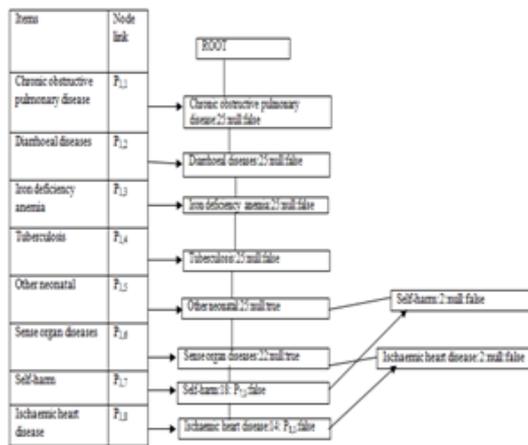
Step 23: insertion of state Karnataka.



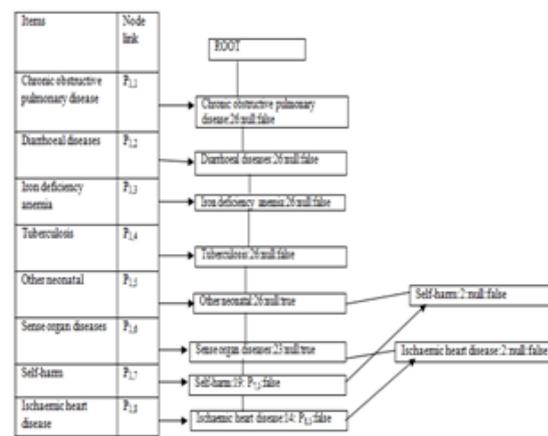
Step 24: insertion of state West Bengal.



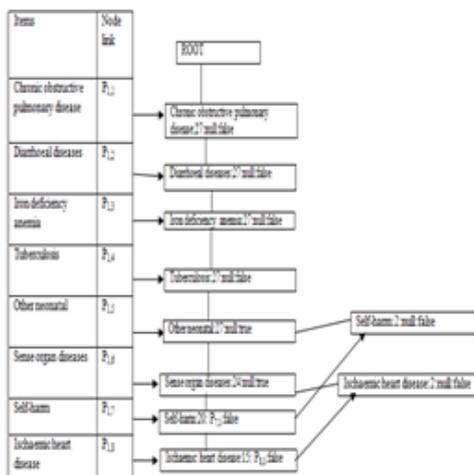
Step 25: insertion of state Maharashtra.



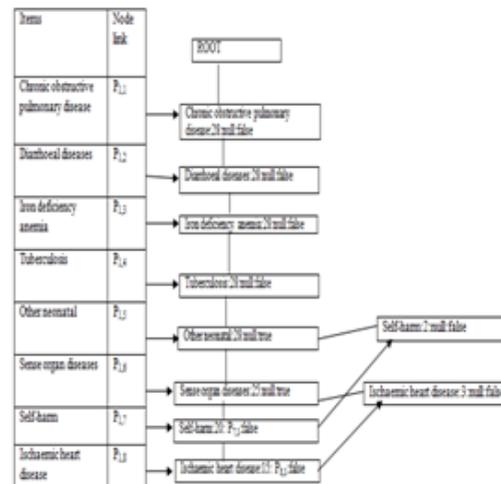
Step 26: insertion of state Uts other than Delhi.



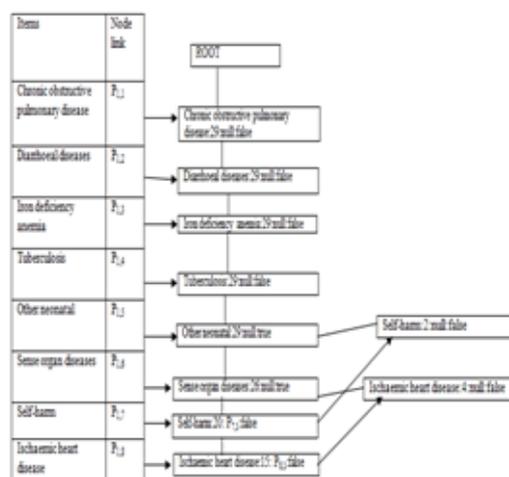
Step 27: insertion of state Himachal Pradesh.



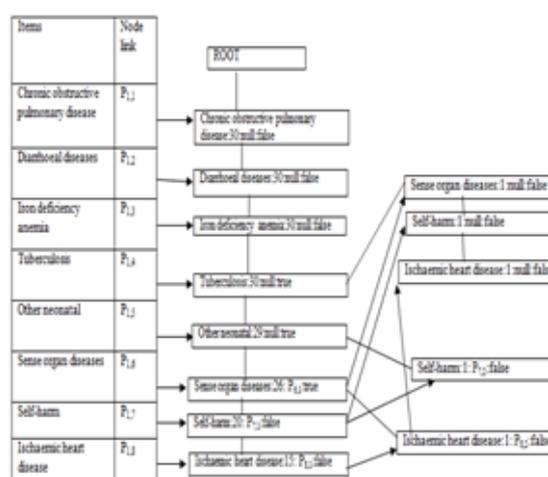
Step 28: insertion of state Punjab.



Step 29: insertion of state Goa.



Step 30: insertion of state Tamilnadu.



Step 31: insertion of state Kerala.

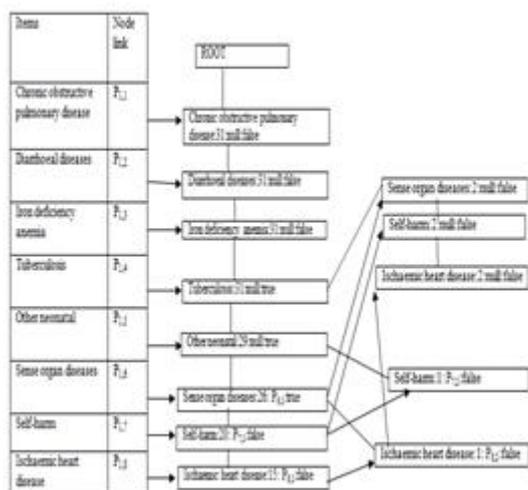
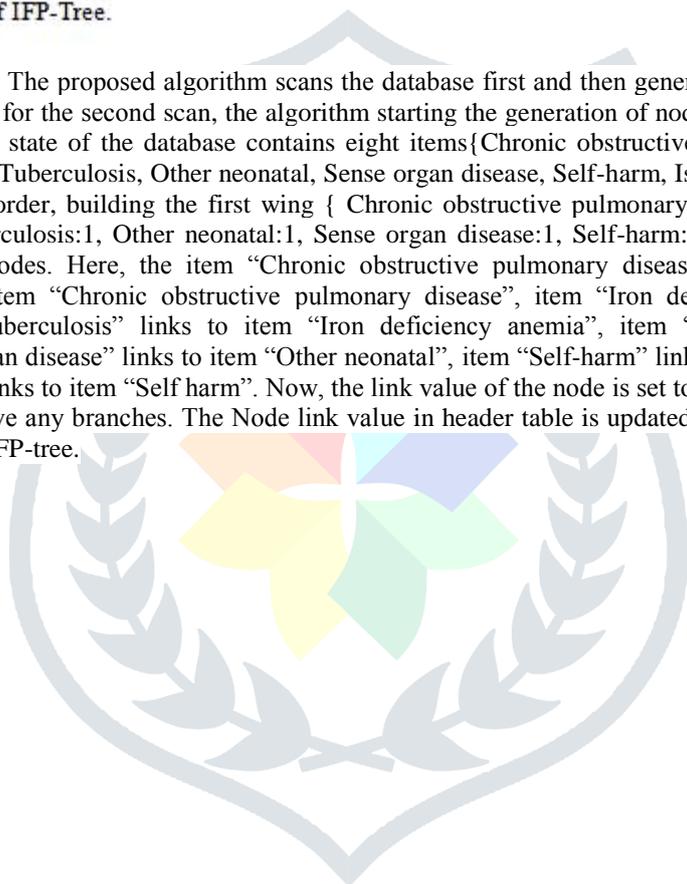


Fig 1: Construction of IFP-Tree.

Improved FP-tree is constructed. The proposed algorithm scans the database first and then generates a node name as “root” with the initial tag value “null”. Then for the second scan, the algorithm starting the generation of nodes for each item in the database. For the given example, the first state of the database contains eight items{Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm, Ischaemic heart disease} based on the arrangement of descending order, building the first wing { Chronic obstructive pulmonary disease:1, Diarrhoeal disease:1, Iron deficiency anemia:1, Tuberculosis:1, Other neonatal:1, Sense organ disease:1, Self-harm:1, Ischaemic heart disease:1} in Improved FP-tree with eight nodes. Here, the item “Chronic obstructive pulmonary disease” is the child of “root”, item “Diarrhoeal disease” links to item “Chronic obstructive pulmonary disease”, item “Iron deficiency anemia” links to item “Diarrhoeal disease”, item “Tuberculosis” links to item “Iron deficiency anemia”, item “Other neonatal” links to item “Tuberculosis”, item “Sense organ disease” links to item “Other neonatal”, item “Self-harm” links to item “Sense organ disease”, item “Ischaemic heart disease” links to item “Self harm”. Now, the link value of the node is set to “null” is updated, the flag value is “false” because it does not have any branches. The Node link value in header table is updated for each and every item when a state is inserted in the Improved FP-tree.



The below Table 5 shows the generated frequent patterns for the given example.

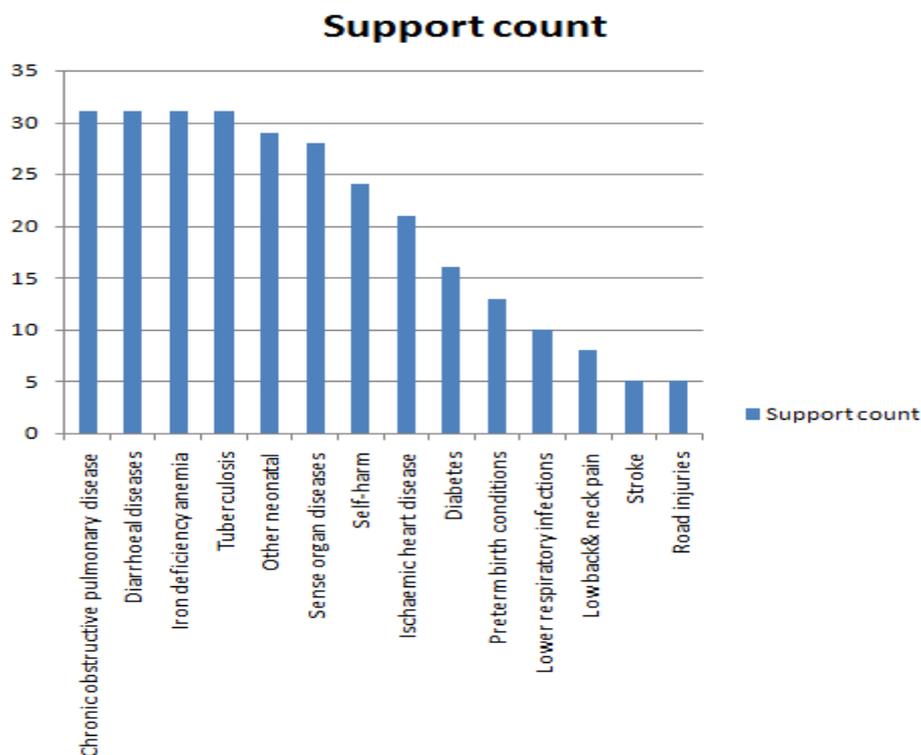
| Item(suffix) | Condition pattern base(prefix) | Conditional FP-Tree | Frequent Patterns Generated |
|---------------------------------------|---|--|---|
| Ischaemic heart disease | {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm.15}{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease.4}{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Sense organ disease, Self-harm.2} | {Chronic obstructive pulmonary disease:21, Diarrhoeal disease:21, Iron deficiency anemia:21, Tuberculosis:21, Other neonatal:19, Sense organ disease:21, Self-harm:21} | {Chronic obstructive pulmonary disease, Ischaemic heart disease:21}{ Diarrhoeal disease, Ischaemic heart disease :21}{ Iron deficiency anemia ,Ischaemic heart disease:21 }{ Tuberculosis ,Ischaemic heart disease:21 }{ Other neonatal ,Ischaemic heart disease:19 }{ Sense organ disease ,Ischaemic heart disease:21 }{ Self-harm ,Ischaemic heart disease }{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease, Self-harm.15} |
| Self-harm | {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease:20}{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal:2}{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Sense organ disease:2} | { Chronic obstructive pulmonary disease:24, Diarrhoeal disease:24, Iron deficiency anemia:24, Tuberculosis:24, Other neonatal:22, Sense organ disease:22} | {Chronic obstructive pulmonary disease, Self-harm:24}{ Diarrhoeal disease, Self-harm:24}{ Iron deficiency anemia , Self-harm:24 }{ Tuberculosis, Self-harm:24 }{ Other neonatal , Self-harm:22 }{ Sense organ disease , Self-harm:22 }{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Sense organ disease :20} |
| Sense organ disease | {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal:26}{Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis:2} | { Chronic obstructive pulmonary disease:28, Diarrhoeal disease:28, Iron deficiency anemia:28, Tuberculosis:28, Other neonatal:26} | {Chronic obstructive pulmonary disease, Sense organ disease:28}{ Diarrhoeal disease, Sense organ disease:28}{ Iron deficiency anemia , Sense organ disease:28 }{ Tuberculosis, Sense organ disease:28 }{ Other neonatal , Sense organ disease:26 }{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal:26} |
| Other neonatal | {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis:29} | { Chronic obstructive pulmonary disease:29, Diarrhoeal disease:29, Iron deficiency anemia:29, Tuberculosis:29} | {Chronic obstructive pulmonary disease, Other neonatal:29}{ Diarrhoeal disease, Other neonatal:29}{ Iron deficiency anemia , Other neonatal:29 }{ Tuberculosis, Other neonatal:29 }{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis:29} |
| Tuberculosis | {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia:31} | { Chronic obstructive pulmonary disease:31, Diarrhoeal disease:31, Iron deficiency anemia:31} | {Chronic obstructive pulmonary disease, Tuberculosis:31}{ Diarrhoeal disease, Tuberculosis:31}{ Iron deficiency anemia, Tuberculosis:31 }{ Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia:31} |
| Iron deficiency anemia | {Chronic obstructive pulmonary disease, Diarrhoeal disease:31} | { Chronic obstructive pulmonary disease:31, Diarrhoeal disease:31} | {Chronic obstructive pulmonary disease, Iron deficiency anemia:31}{ Diarrhoeal disease, Iron deficiency anemia:31}{ Chronic obstructive pulmonary disease, Diarrhoeal disease:31} |
| Diarrhoeal disease | {Chronic obstructive pulmonary disease:31} | { Chronic obstructive pulmonary disease:31} | {Chronic obstructive pulmonary disease, Diarrhoeal disease:31}{ Chronic obstructive pulmonary disease:31} |
| Chronic obstructive pulmonary disease | | | |

Table 5: Generated Frequent Patterns.

The second entry contains seven items {Chronic obstructive pulmonary disease:1, Diarrhoeal disease:1, Iron deficiency anemia:1, Tuberculosis:1, Other neonatal:1, Sense organ disease:1, Ischaemic heart disease:1} where seven nodes are generated. Here, the branch shares the prefix “Chronic obstructive pulmonary disease”, “Diarrhoeal disease”, “Iron deficiency anemia”, “Tuberculosis”, “Other neonatal”, “Sense organ disease”. For the item “Ischaemic heart disease“ new node is generated from the “Sense organ disease”, then the flag value of “Sense organ disease” is updated as “true” because it consist a branch and the link value of “Ischaemic heart disease” is updated to P_{8,1}. The count value is also updated. Similarly all the states are inserted. The ninth state of the database contains {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Other neonatal, Self-harm}. In this situation the items shares the previous link upto the “Other neonatal”. For the item “Self harm” it shares the link from “Other neonatal”. The flag value of “Other neonatal” is updated to “true” and the link value of “Self harm” is updated to P_{7,1}. For the entry of thirty state {Chronic obstructive pulmonary disease, Diarrhoeal disease, Iron deficiency anemia, Tuberculosis, Sense organ disease, Self harm, Ischaemic heart disease}.this entry shares the link upto the “Tuberculosis”, the item “sense organ disease” shares the link with “Tuberculosis”, the item “Self harm” shares the link with “Sense organ disease”, the item “Ischaemic heart disease” shares the link with “Self harm”. The flag value of “Tuberculosis” is updated to “true” and the node value of “Sense organ disease” is updated to P_{6,1}.

V. EXPERIMENTAL RESULTS ANALYSIS

In this particular experiment we can observe that the frequently produced diseases in India. From the graph we can conclude that Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis are the four diseases which are frequently occurred in all the states of India. So, the Health Organization of India or Government of India can take the appropriate measures to control spread of these diseases and can provide sufficient medicines and treatment to the people of India.



For the above graph we have taken the diseases list of all states on X-axis and the support count of diseases on the Y-axis. From the above graph we can clearly observe that the Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis are the four diseases have the support count 31 because these diseases are presented in all the states of India.

VI. CONCLUSION

In this research paper we find out the frequently occurred diseases in all the states of India by using the Improved FP-Tree. By using the Improved FP-Tree algorithm we can easily find the frequently occurred diseases in India. The Improved FP-tree used in this research paper uses the divide and conquer technique. This algorithm reduces the usage of the space while constructing the Tree. By observing the results we can conclude that the Improved FP-Tree find the frequently occurred diseases in India. In this research paper we find out the occurrence of Chronic obstructive pulmonary disease, Diarrhoeal diseases, Iron deficiency anemia, Tuberculosis are more in the all the 31 states of India. With the help of these results the central government of India can take the necessary actions to reduce the spread of these diseases all over the country and provide the sufficient hospitals, medicines to the effected people to survive from these diseases.

REFERENCES

- [1] J. Schafer, J. Konstan, and J. Riedl, "Electronic Commerce Recommender Applications," *Journal of Data Mining and Knowledge Discovery*, vol. 5, pp. 115-152, 2001.
- [2] Zhaopeng Pan, Peiyu Liu, Jing Yi. "An Improved FP-Tree Algorithm for Mining Maximal Frequent Patterns", 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2018.
- [3] W. Lin, "Association Rule Mining for Collaborative Recommender Systems," in *Computer Science*.
- [4] Ning Hui, Wang Suhong, Cui Ligang, Guo Xiaoyan, Xu Li. Based on the improved FP-tree maximum frequent pattern mining algorithm [J/OL]. *Application technology*, 2016, 43(02):37-43.
- [5] Li Yibai, Tang Hui, Zhang Chun, He Yuming. Frequency pattern mining algorithm based on improved FP-tree [J]. *Journal of Computer Applications*, 2011, 31(01):101-103.
- [6] Liu Nai-li. Design and implementation of association rule mining algorithm based on FP-tree [D]. Shandong University, 2005.
- [7] Feng Zhixin. Based on FP-tree maximum frequent pattern mining algorithm [D]. Guangxi University, 2003.
- [8] QI Xue-zhong, HUO Liang. Multifrequent pattern mining algorithm for improved FP-tree in association rules [J]. *Computer Engineering and Design*, 2010, 31 (21): 4635-4638.
- [9] Ali Anaissi, Madhu Goyal. "chapter 12 Blending Association Rules for Knowledge Discovery in Big Data", IGI Global, 2019.
- [10] S. Langenfeld, "CRM and the Customer Driven Demand Chain," vol. 2004, 2004.
- [11] R. Agarwal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Datasets," in *Proceedings of the ACM SIGMOD Conference on Management of Data*. Washington, D.C.: ACM, 1993, pp. 207-216.
- [12] B. Goethals, "Survey on Frequent Pattern Mining," vol. 2004, 2003.
- [13] Yong Qiu, Yongjie Lan, Qing-Song Xie "An Improved Algorithm Of Mining From FP-Tree" *Information & Electronic Engineering School, Shandong Institute of Business and Technology*, YanTai 264005, China.
- [14] Han, J. and Kamber, M. *Data mining: Concepts and Techniques*, Academic Press, 2001.
- [15] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *VLDB'1994*, pp. 487-499.
- [16] G. Grahn, L. Lakshmanan, and X. Wang. Efficient mining of constrained correlated sets. In *ICDE2000*.
- [17] B. Lent, A. Swami, and J. Widom. Clustering association rules. In *ICDE1997*, pp. 220-231.
- [18] A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large database. In *VLDB'1995*, pp. 432-443.

- [19] R.Srikant, Q. Vu, and R.Agrawal. Mining association rules with item constraints. In KDD' 1997, pp.594-605.
- [20] Nasreen, Shamila, Muhammad Awais Azam, Khurram Shehzad, Usman Naeem, and Mustansar Ali Ghazanfar. "Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams: A Survey", Procedia Computer Science, 2014.

