We Claim:

1. A method to construct a decision tree from a dataset(s), the method comprising:

building a prefix tree to store the dataset using a predictor attribute(s) value(s) of a set of records of the dataset;

    counting a frequencies of the predictor attributes’ values from the prefix tree and storing the frequencies in a frequency table;

    determining a best split condition based on the frequencies of the predictor attributes’ values stored in the frequency table;

    splitting one or more portion of the prefix tree to generate at least one new prefix tree(s) based on the best split condition; and

    constructing a decision tree by creating a decision tree node and at least one of a decision subtree(s) at the decision tree node; wherein the construction of at least one of the decision subtrees occurs in a recursive process by using the new prefix tree.

2. The method as recited in claim 1, wherein the set of records of the dataset stored in a relational database format is converted into a set of sequences in a sequential database format before building the prefix tree.

3. The method of claim 1, wherein the set of records of the dataset converted to sequential format considering at least one of a primary predictor attribute value(s) or the predictor attribute value(s) or the predictor attribute(s).
4. The method of claim 2, wherein the set of records of the dataset in the sequential database format comprises at least one of the primary predictor attributes values or a class attributes values or both.

5. The method of claim 1, wherein the dataset used to build the prefix tree includes at least one of a continuous dataset or a categorical dataset or combinations thereof.

6. The method of claim 1, wherein building the prefix tree starts with a prefix tree root node, wherein the prefix tree root node of the prefix tree comprises a null value.

7. The method of claim 1, wherein the built prefix tree includes at least one or more of the prefix subtrees.

8. The method of claim 7, wherein the prefix subtrees of the prefix tree are created in an order of occurrence of the predictor attributes’ values.

9. The method of claim 1, wherein the building of the prefix tree is performed in a sequential mode by using the set of sequences in the sequential database.

10. The method of claim 1, wherein the prefix tree is built by creating a path for each set of sequences and incrementing the class label frequencies at every prefix tree node in the path.
11. The method of claim 1, wherein counting the frequencies of the predictor attributes values from the prefix tree is performed by a first traversing process of the prefix tree.

12. The method as recited in claim 1, wherein the best split condition is the set of records' a best split predictor attribute derived out of the frequency table.

13. The method of claim 1, wherein the creation of the decision tree node includes storing the best split condition in the decision tree node as a decision tree condition.

14. The method of claim 1, wherein the new prefix tree is built by second traversing the prefix tree and by splitting one or more portion of the prefix tree based on the best predictor attribute value.

15. The method as recited in claim 1, wherein at least one of the new prefix trees stores the set of records of the dataset in relation with the best split predictor attribute value.

16. The method as recited in claim 1, wherein at least one of the new prefix tree stores the set of records of the dataset which is not in relation with the best split predictor attribute value.
17. The method of claim 1, wherein the creation of the decision subtrees occurring in a recursive process ends when the corresponding prefix subtree satisfies at least one or more set of predefined conditions.

18. The method of claim 17, wherein the set of predefined conditions includes at least one of the frequencies of the sequences' class attribute at the prefix tree node in the path belonging to one class value or all of the predictor attributes has been selected as the best split condition or the frequency of the sequences' class attribute is null or combinations thereof.

19. The method of splitting a prefix tree(s) into at least one or more new prefix tree(s), the method comprising:

   traversing a path of the prefix tree in a sequential order to locate a prefix tree node comprising a predictor attributes' value(s) matching a best split condition;

   adding a new path from a new prefix tree root node of a new prefix tree; wherein the new path added from the new prefix tree root node is the path selected from the prefix tree while traversing the prefix tree till the best split condition is located; and

   removing a prefix subtree(s) of the prefix tree in entirety from the prefix tree node of the prefix tree where the best split condition is located and appending the prefix subtrees to the end of the path of the new prefix tree.

20. The method of claim 19, further comprising traversing the path at each prefix tree node of the prefix tree for locating the best split condition while splitting the prefix tree.
21. The method of claim 19, further comprising retaining the path of the prefix tree till the prefix tree node where the best split condition is not located in the prefix tree while traversing the prefix tree for splitting the prefix tree.

22. The method of claim 21, wherein retaining the path of the prefix tree is performed after traversing the path at each prefix tree node of the prefix tree where the best split condition in not located while splitting the prefix tree.

23. The method of claim 19, further comprising incrementing a class label frequencies of a new prefix tree node in the first path of the new prefix tree with the class label frequencies of the prefix tree node where the best split condition is located in the prefix tree.

24. The method of claim 19, further comprises decrementing the class label frequencies of the prefix tree node in the path of the prefix tree with the class label frequencies of the prefix tree node where the best split condition is located in the prefix tree.

25. The method of claim 19, wherein the prefix tree node storing the predictor attribute value and matching the best split condition is deleted after splitting the prefix tree into new prefix tree.

26. A computer program product comprising a computer usable medium having a computer readable program code embodied therein to construct a decision tree from a dataset(s), the method comprising:
program code adapted for building a prefix tree to store the dataset using a predictor attribute(s) value(s) of a set of record(s) of the dataset;

program code adapted for counting a frequencies of the predictor attributes’ values from the prefix tree and storing the frequencies in a frequency table;

program code adapted for determining a best split condition based on the frequencies of the predictor attributes’ values stored in the frequency table;

program code adapted for splitting one or more portion of the prefix tree to generate at least one of a new prefix tree(s) based on the best split condition; and

program code adapted for creating a decision tree node and constructing at least one of a decision subtree(s) at the decision tree node; wherein the construction of at least one of the decision subtrees occurs in a recursive process by using the new prefix trees.

27. The product of claim 26, further comprising program code adapted for counting the frequencies of the predictor attributes values from the prefix tree is performed by a first traversing process of the prefix tree.

28. The product of claim 26, further comprising program code adapted to create the decision tree node includes storing the best split condition in the decision tree node as a decision tree condition.

29. The product of claim 26, further comprising program code adapted to end the recursive process of creation of the decision subtrees on approaching an event of a set of predefined conditions.

Dated this 20th day of August, 2007

[Signature]

MANISHA SINGH NAIR
Agent for the Applicant [IN/PA-740]
LEX ORBIS
Intellectual Property Practice
709 / 710, Tolstoy House,
15-17, Tolstoy Marg,
New Delhi-110 001