CLAIMS

1. A computer-implemented method comprising:

receiving, using a processor, a plurality of data inputs as a training set from a process;

selecting, using the processor, a data subset I_0 from the plurality of data inputs by solving linear programming to obtain a solution resulting in a subset of data inputs, wherein at least some of the subset of data inputs in data subset I_0 with label y are correctly classified and were assigned into a same leaf node, wherein all of the subset of data inputs in the data subset I_0 with a same label y inside a convex hull of I_o conv($\{x_i \mid i \in I_o\}$) are dropped from the training set, where x_i are removed training samples, and wherein correctly classifying the subset of data inputs in I_0 also correctly classifies remaining data inputs inside the convex hull;

building, using the processor, an optimal decision tree ("ODT") based on the data subset I_0 in which all of the subset of data inputs in the data subset I_0 with the same label y inside the convex hull of I_o conv($\{x_i \mid i \in I_o\}$) have been dropped from the training set;

alerting, using the processor, a user when a prediction of a process failure results from the optimal decision tree is greater than a threshold value, in response to checking the prediction against the threshold value; and

in response to alerting the user when the prediction of the optimal decision tree is greater than the threshold value, implementing, by the processor, a control action based on analysis of the optimal decision tree, the control action adjusting a value of a feature vector input that results in a change to a prediction output of the prediction.

2. The computer-implemented method of claim 1, wherein selecting, using the processor, a data subset from the plurality of data inputs comprises performing, using the processor, a hyperplane data-selection to select the data subset.

3. The computer-implemented method of claim 1 further comprising performing, using the processor, data selection within each leaf node of the ODT to select the data subset.

4. The computer-implemented method of claim 1 further comprising using, using the processor, a set of cutting planes to strengthen the ODT.

5. The computer-implemented method of claim 1, wherein selecting, using the processor, a data subset from the plurality of data inputs comprises selecting the data subset such that points inside a convex hull of the data subset are maximized.

6. The computer-implemented method of claim 1, wherein the ODT comprises a multivariable decision tree.

7. A system comprising:

a memory;

a processor communicatively coupled to the memory, the processor operable to execute instructions stored in the memory, the instructions causing the processor to:

receive a plurality of data inputs as a training set from a process;

select a data subset I_0 from the plurality of data inputs by solving linear programming to obtain a solution resulting in a subset of data inputs, wherein at least some of the subset of data inputs in data subset I_0 with label y are correctly classified and were assigned into a same leaf node, wherein all of the subset of data inputs in the data subset I_0 with a same label y inside a convex hull of I_o conv($\{x_i \mid i \in I_o\}$) are dropped from the training set, where x_i are removed training samples, and wherein correctly classifying the subset of data inputs in I_0 also correctly classifies remaining data inputs inside the convex hull;

build an optimal decision tree ("ODT") based on the data subset I_0 in which all of the subset of data inputs in the data subset I_0 with the same label y inside the convex hull of I_o conv($\{x_i \mid i \in I_o\}$) have been dropped from the training set;

alert a user when a prediction of a process failure results from the optimal decision tree is greater than a threshold value, in response to checking the prediction against the threshold value; and

in response to alerting the user when the prediction of the optimal decision tree is greater than the threshold value, implement a control action based on analysis of the optimal decision tree, the control action adjusting a value of a feature vector input that results in a change to a prediction output of the prediction.

8. The system of claim 7, wherein selecting a data subset from the plurality of data inputs comprises performing a hyperplane data-selection to select the data subset.

9. The system of claim 7, wherein the instructions further cause the processor to perform data selection within each leaf node of the ODT to select the data subset.

10. The system of claim 7, wherein the instructions further cause the processor to use a set of cutting planes to strengthen the ODT.

11. The system of claim 7, wherein selecting a data subset from the plurality of data inputs comprises selecting the data subset such that points inside a convex hull of the data subset are maximized.

12. The system of claim 7, wherein the ODT comprises a multivariable decision tree.

13. A computer program product for creating and optimizing a decision tree, the computer program product comprising a computer readable storage medium having program instructions embodied therewith, the program instructions executable by a computer, to cause the computer to perform a method comprising:

receiving a plurality of data inputs as a training set from a process;

selecting a data subset I_0 from the plurality of data inputs by solving linear programming to obtain a solution resulting in a subset of data inputs, wherein at least some of the subset of data inputs in data subset I_0 with label y are correctly classified and were assigned into a same leaf node, wherein all of the subset of data inputs in the data subset I_0 with a same label y inside a convex hull of I_o conv($\{x_i \mid i \in I_o\}$) are dropped from the training set, where x_i are removed training samples, and wherein correctly classifying the subset of data inputs in I_0 also correctly classifies remaining data inputs inside the convex hull;

building and optimizing an optimal decision tree ("ODT") based on the data subset I_0 in which all of the subset of data inputs in the data subset I_0 with the same label y inside the convex hull of I_o conv($\{x_i \mid i \in I_o\}$) have been dropped from the training set; and

alerting a user when a prediction of a process failure results from the optimal decision tree is greater than a threshold value, in response to checking the prediction against the threshold value; and

in response to alerting the user when the prediction of the optimal decision tree is greater than the threshold value, implementing a control action based on analysis of the optimal decision tree, the control action adjusting a value of a feature vector input that results in a change to a prediction output of the prediction.

14. The computer program product of claim 13, wherein selecting a data subset from the plurality of data inputs comprises performing a hyperplane data-selection to select the data subset.

15. The computer program product of claim 13, wherein the instructions further cause the computer to perform data selection within each leaf node of the ODT to select the data subset.

16. The computer program product of claim 13, wherein the instructions further cause the computer to use a set of cutting planes to strengthen the ODT.

17. A computer-implemented method of claim 1, further comprising training an artificial intelligence model to anticipate events by using predictions from the ODT.