CLAIMS

1. A method, comprising:

receiving from a client device an input dataset and a selection of a feature of the input dataset for which a prediction is to be made, wherein the input dataset is split into training data and validation data;

generating a plurality of machine learning models to make a prediction for the selected feature based on the training data until a stopping condition has been reached, wherein generating the plurality of machine learning models includes:

determining a plurality of initial machine learning models based on a plurality of original features included in the input dataset;

filtering the plurality of initial machine learning models based at least in part on corresponding validation scores associated with the plurality of initial machine learning models, wherein filtering the plurality of initial machine learning models at least includes pairing a first initial machine learning model with a second initial machine learning model, pairing a third initial machine learning model with a fourth initial machine learning model, and selecting a subset of the initial machine learning models as one or more surviving machine learning models, wherein the first initial machine learning model or the second initial machine learning model is selected based on which corresponding validation score is higher and wherein the third initial machine learning model or the fourth initial machine learning model is selected based on which corresponding validation score is higher;

generating one or more evolved machine learning models, wherein at least one of the evolved machine learning models is based at least in part on one or more new features, wherein the one or more new features are based at least in part on a transformation of at least one feature of the one or more surviving machine learning models;

comparing corresponding validation scores associated with the one or more evolved machine learning models and the one or more surviving machine learning models including comparing a corresponding validation score associated with a first evolved machine learning model with a corresponding validation score associated with a first surviving machine learning model; and

selecting based on a comparison of the corresponding validation scores, at least one of the one or more evolved machine learning models or the one or more surviving machine learning models as one or more new surviving machine learning models, wherein the first evolved machine

learning model or the first surviving machine learning model is retained based on which has a higher corresponding validation score; and

providing to the client device a prediction for the selected feature using a retained machine learning model.

2. The method of claim 1, wherein the received input dataset includes corresponding feature values associated with the plurality of original features.

3. The method of claim 2, wherein the corresponding feature values have corresponding data types.

4. The method of claim 3, wherein the transformation is selected based at least in part on the corresponding data types associated with the plurality of original features.

5. The method of claim 2, wherein the plurality of initial machine learning models are based on a subset of the original features.

6. The method of claim 1, wherein generating the one or more evolved machine learning models includes determining one or more important features of the plurality of original features.

7. The method of claim 6, wherein the one or more important features are determined based on a number of times a feature is used in a decision tree.

8. The method of claim 6, wherein determining one or more important features of the plurality of original features includes:

selecting one of the original features;

randomizing feature values associated with the selected feature; and

determining an overall validation score of an initial machine learning model based at least in part on the selected feature having the randomized feature values.

9. The method of claim 8, wherein the selected feature is determined to be an important feature in response to determining that the overall validation score of the initial machine learning model decreases more than a threshold statistical amount.

10. The method of claim 2, wherein filtering the plurality of initial machine learning models includes:

randomly selecting a predetermined number of entries of the input data;

determining a validation score associated with each entry of the randomly selected entries; and

determining an overall validation score based at least in part on the validation score associated with each entry of the randomly selected entries.

11. The method of claim 1, wherein one of the evolved machine learning models is selected in the event an overall validation score of the evolved machine learning model is greater than an overall validation score of at least one of the surviving machine learning models.

12. The method of claim 1, wherein an evolved machine learning model is selected in the event an overall accuracy of the evolved machine learning model is greater than a statistical measure of at least one of the surviving machine learning models.

13. The method of claim 1, further comprising:

determining that the stopping condition has not occurred; and

filtering the one or more surviving machine learning models and the one or more evolved machine learning models based at least in part on corresponding validation scores associated with the one or more surviving machine learning models and the one or more evolved machine learning models.

14. The method of claim 1, further comprising:

determining that the stopping condition has occurred;

selecting a machine learning model with the highest validation score; and

outputting a prediction label corresponding to the selected feature of the plurality of original features.

15. The method of claim 1, wherein one of the one or more new features is based in part on a first feature from a first surviving machine learning model and a second feature from a second surviving machine learning model.

16. The method of claim 1, further comprising:

filtering the one or more new selected surviving machine learning models based at least in part on corresponding validation scores associated with the one or more new selected surviving machine learning models, wherein filtering the one or more new selected surviving machine learning models includes selecting a subset of the one or more new selected surviving machine learning models as the one or more new surviving machine learning models;

generating one or more new evolved machine learning models, wherein at least one of the new evolved machine learning models is based at least in part on a second set of one or more new features, wherein the second set of one or more new features are based at least in part on a second transformation of at least one of many features of the one or more new surviving machine learning models;

comparing corresponding validation scores associated with the one or more new evolved machine learning models and the one or more new surviving machine learning models; and

selecting based on the comparison of the corresponding validation scores, at least one of the one or more new evolved machine learning models or the one or more new surviving machine learning models as a second set of one or more new surviving machine learning models.

17. A system, comprising:

a processor configured to:

receive from a client device an input dataset and a selection of a feature of the input dataset for which a prediction is to be made, wherein the input dataset is split into training data and validation data;

generate a plurality of machine learning models to make a prediction for the selected feature based on the training data until a stopping condition has been reached, wherein to generate the plurality of machine learning models, the processor is configured to:

determine a plurality of initial machine learning models based on a plurality of original features included in the input dataset;

filter the plurality of initial machine learning models based at least in part on corresponding validation scores associated with the plurality of initial machine learning models, wherein filtering the plurality of initial machine learning models at least includes pairing a first initial machine learning model with a second initial machine learning model, pairing a third initial machine learning model with a fourth initial machine learning model, and selecting a subset of the initial machine learning models as one or more surviving machine learning models, wherein the first initial machine learning model or the second initial machine learning model is selected based on which corresponding validation score is higher and wherein the third initial machine learning model or the fourth initial machine learning model is selected based on which corresponding validation score is higher;

generate one or more evolved machine learning models, wherein at least one of the evolved machine learning models is based at least in part on one or more new features, wherein the one or more new features are based at least in part on a transformation of at least one feature of the one or more surviving machine learning models;

compare corresponding validation scores associated with the one or more evolved machine learning models and the surviving machine learning models including comparing a corresponding validation score associated with a first evolved machine learning model with a corresponding validation score associated with a first surviving machine learning model; and

selecting based on the comparison of the corresponding validation scores, at least one of the one or more evolved machine learning models or the one or more surviving

machine learning models as one or more new selected surviving machine learning models, wherein the first evolved machine learning model or the first surviving machine learning model is retained based on which has a higher corresponding validation score; and

provide to the client device a prediction for the selected feature using a retained machine learning model; and

a memory coupled to the processor and configured to provide the processor with instructions.

18. The system of claim 17, wherein to generate the one or more evolved machine learning models, the processor is further configured to determine one or more important features of the plurality of original features.

19. A computer program product, the computer program product being embodied in a non-transitory computer readable storage medium and comprising computer instructions for:

receiving from a client device an input dataset and a selection of a feature of the input dataset for which a prediction is to be made, wherein the input dataset is split into training data and validation data;

generating a plurality of machine learning models to make a prediction for the selected feature based on the training data until a stopping condition has been reached, wherein generating the plurality of machine learning models includes:

determining a plurality of initial machine learning models based on a plurality of original features included in the input dataset;

filtering the plurality of initial machine learning models based at least in part on corresponding validation scores associated with the plurality of initial machine learning models, wherein filtering the plurality of initial machine learning models at least includes pairing a first initial machine learning model with a second initial machine learning model, pairing a third initial machine learning model with a fourth initial machine learning model, and selecting a subset of the initial machine learning models as one or more surviving machine learning models, wherein the first initial machine learning model or the second initial machine learning model is selected based on which corresponding validation score is higher and wherein the third initial machine learning model or the fourth initial machine learning model is selected based on which corresponding validation score is higher;

generating one or more evolved machine learning models, wherein at least one of the evolved machine learning models is based at least in part on one or more new features, wherein the one or more new features are based at least in part on a transformation of at least one feature of the one or more surviving machine learning models;

comparing corresponding validation scores associated with the one or more evolved machine learning models and the one or more surviving machine learning models including comparing a corresponding validation score associated with a first evolved machine learning model with a corresponding validation score associated with a first surviving machine learning model; and

selecting based on the comparison of the corresponding validation scores, at least one of the one or more evolved machine learning models or the one or more surviving machine learning models as one or more new surviving machine learning models, wherein the first evolved machine learning model or the first surviving machine learning model is retained based on which has a higher corresponding validation score; and

providing to the client device a prediction for the selected feature using a retained machine learning model.