

CLAIMS

What is claimed is:

1. A method comprising: with a machine learning platform:
 - accessing user data;
 - accessing purpose information identifying a purpose for a model;
 - identifying canonical features by using the purpose information;
 - detecting one or more of the canonical features from the accessed user data;
 - selecting a model type in accordance with the purpose information;
 - selecting a target in accordance with the purpose information;
 - select model parameters in accordance with the purpose information;
 - generating a model having the selected model type by using the accessed user data, wherein the model uses the detected canonical features as inputs, predicts values for the selected target, and includes the selected model parameters;
 - generating business analysis information for the generated model, in accordance with the purpose information; and
 - providing the business analysis information to at least one system external to the machine learning platform.
2. The method of Claim 1, wherein the machine learning platform automatically identifies canonical features, detects canonical features from the accessed user data, selects the model type, selects the target, selects the model parameters, generates the model, generates the business analysis information, and provides the business analysis information in response to accessing the user data and the purpose information from a user system via a user interface system included in the machine learning platform.
3. The method of Claim 1, wherein the identified purpose is one of: automotive loan origination, consumer loan origination, business loan origination, loan repayment prediction, new loan solicitation, curable loan identification, applicant identification, and business loan repayment.
4. The method of Claim 1,

wherein accessing purpose information identifying a purpose for a model comprises: accessing model purpose data that is stored in association with the identified model purpose,

wherein the accessed model purpose data defines the canonical features to be used as model inputs, the model type, the target, and the model parameters, and

wherein identifying canonical features comprises: identifying canonical features defined by the accessed model purpose data.

5. The method of Claim 4, wherein detecting one or more of the canonical features from the accessed user data comprises: extracting canonical features from the accessed user data by applying at least one predetermined transformation rule.

6. The method of Claim 5, wherein the accessed model purpose data defines the at least one predetermined transformation rule used to extract the canonical features from the accessed user data.

7. The method of Claim 4,

wherein the accessed model purpose data defines a business analysis process, and

wherein generating business analysis information for the generated model comprises: performing the business analysis process defined by the accessed model purpose data.

8. The method of Claim 4,

further comprising: with the machine learning platform: evaluating the generated model,

wherein the accessed model purpose data defines at least one business metric, and

wherein evaluating the generated model comprises:

computing a business metric value for each business metric defined by the model purpose data for the generated model,

computing a business metric value for each business metric defined by the model purpose data for an original model,

comparing the business metric values for the original model with the corresponding business metric values for the generated model, and generating evaluation information that includes results of the comparison between the business metric values for the original model and the business metric values for the generated model.

9. The method of Claim 8, wherein evaluating the generated model further comprises: performing fair lending disparate impact analysis, and generating evaluation information includes results of the fair lending disparate impact analysis.
10. The method of Claim 9, wherein evaluating the generated model further comprises: performing model accuracy analysis, and generating evaluation information includes results of the model accuracy analysis.
11. The method of Claim 10,
further comprising: with the machine learning platform: generating explanation information for model output generated by the model.
12. The method of Claim 11, wherein the model is a credit model that generates a credit score for a credit application, and wherein the explanation information generated for the model output includes FCRA Adverse Action Reason Codes.
13. The method of Claim 11,
further comprising: with the machine learning platform:
monitoring the model to detect at least one of feature drift, unexpected inputs, unexpected outputs, population instability, and unexpected economic performance; and
providing an alert to at least one system in response to detecting at least one of feature drift, unexpected inputs, unexpected outputs, population instability, and unexpected economic performance.
14. The method of Claim 1,
further comprising: with the machine learning platform:
automatically generating documentation for the model, wherein the documentation includes:
documentation information identifying the accessed user data,
documentation information identifying the identified canonical features,

documentation information identifying the detected canonical features,
documentation information identifying the selected model type,
documentation information identifying the selected target,
documentation information identifying the selected model parameters,
information describing generating of the model, and
the business analysis information; and

providing the generated documentation to a system external to the machine learning platform.

15. The method of Claim 4, wherein the user data and the purpose information are received one or more of an external loan origination system and an external loan management system.

16. The method of Claim 15, wherein the model purpose data is received from an external computing system of a domain expert.

17. The method of Claim 1, wherein the generated model includes at least a gradient boosted tree forest (GBM) coupled to base signals, and a smoothed approximate empirical cumulative distribution function (ECDF) coupled to output of the GMB, wherein output values of the GBM are transformed by using the ECDF and presented as a credit score.

18. The method of Claim 1, wherein the generated model includes submodels including at least a GMB, a neural network, and an Extremely Random Forest (ETF), wherein outputs of the submodels are ensembled together using one of a stacking function and a combining function, and wherein an ensembled output is presented as a credit score.

19. The method of Claim 1, wherein the generated model includes submodels including at least a neural network (NN), a GBM, and an ETF, wherein outputs of the submodels are ensembled by a linear ensembling module, wherein an output of the linear ensembling module is processed by a differentiable function, and wherein an output of the differentiable function is presented as a credit score.

20. The method of Claim 1, wherein the generated model includes at least a neural network (NN), a GBM, and a neural network ensembling module, wherein an output of the neural network ensembling module is processed by a differentiable function.