

**(12) STANDARD PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

(11) Application No. **AU 2016304571 B2**

(54) Title  
**Model integration tool**

(51) International Patent Classification(s)  
**G06F 9/44** (2006.01)

(21) Application No: **2016304571** (22) Date of Filing: **2016.08.05**

(87) WIPO No: **WO17/024242**

(30) Priority Data

(31) Number	(32) Date	(33) Country
<b>62/201,258</b>	<b>2015.08.05</b>	<b>US</b>

(43) Publication Date: **2017.02.09**

(44) Accepted Journal Date: **2021.09.16**

(71) Applicant(s)  
**Equifax Inc.**

(72) Inventor(s)  
**Indurthivenkata, Rajesh;Thakur, Nandita;Venkataramani, Lalithadevi;Somaka, Aparna;Zhang, Xingjun;Turner, Matthew;Koshyari, Bhawana;Nagarajan, Vijay;Reid, James**

(74) Agent / Attorney  
**FB Rice Pty Ltd, Level 23 44 Market Street, Sydney, NSW, 2000, AU**

(56) Related Art  
**US 20040250255 A1**  
**FLINT A et al;"Scorecard Element in PMML 4.1 Provides Rich, Accurate Exchange of Predictive Models for Improved Business Decisions" 21 August 2011**



- (51) International Patent Classification:  
G06F 9/44 (2006.01)
- (21) International Application Number:  
PCT/US2016/045822
- (22) International Filing Date:  
5 August 2016 (05.08.2016)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
62/201,258 5 August 2015 (05.08.2015) US
- (71) Applicant: EQUIFAX INC. [US/US]; 1550 Peachtree Street, N.W., Atlanta, Georgia 30309 (US).
- (72) Inventors: INDURTHIVENKATA, Rajesh; 4290 Ridge-fair Drive, Cumming, Georgia 30040 (US). THAKUR, Nandita; 1345 Hartford Terrace, Alpharetta, Georgia 30004 (US). VENKATARAMANI, Lalithadevi; 565 Midhurst Place, Johns Creek, Georgia 30024 (US). SOMAKA, Aparna; 1520 Riva Ridge, Swanee, Georgia 30024 (US). ZHANG, Xingjun; 5630 Upper Creek Court, Suwanee, Georgia 30024 (US). TURNER, Matthew; 1840 Bitternut Lane, Cumming, Georgia 30041 (US).

KOSHYARI, Bhawana; 5415 Fairmont Glen, Alpharetta, Georgia 30004 (US). NAGARAJAN, Vijay; 3105 Stonecrest Drive, Cumming, Georgia 30041 (US). REID, James; 2410 Langstrath Lane, Cumming, Georgia 30041 (US).

(74) Agents: GARDNER, Jason D. et al.; Kilpatrick Townsend & Stockton LLP, Suite 1400, 4208 Six Forks Road, Raleigh, North Carolina 27609 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,

[Continued on next page]

(54) Title: MODEL INTEGRATION TOOL

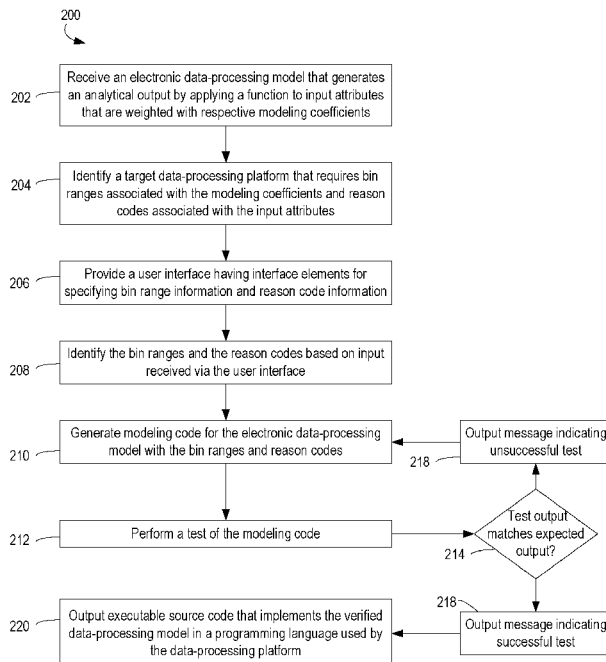


FIG. 2

(57) Abstract: Certain aspects involve building and debugging models for generating source code executed on data-processing platforms. A target data-processing platform is identified that requires bin ranges for modeling coefficients and reason codes for input attributes. A processor outputs source code, which is generated from a modeling code, in a programming language used by the target data-processing platform.

WO 2017/024242 A1

TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG). **Published:** — *with international search report (Art. 21(3))*

## MODEL INTEGRATION TOOL

### Cross-Reference to Related Applications

[0001] This disclosure claims priority to U.S. Provisional Application No. 62/201,258, entitled “MODEL INTEGRATION TOOL,” filed August 5, 2015, the entirety of which is hereby incorporated by reference herein.

### Technical Field

[0002] This disclosure generally relates to software executed on a computer or digital data-processing system for creating and managing computer program code, and more particularly relates to a model integration tool for building and debugging predictive or analytical models that are used to generate source code for execution on data-processing platforms.

### Background

[0003] Software development tools can be used for generating, testing, and deploying program code that implements the functionality of analytical or predictive models. For example, an analytical model can be used for identifying relationships between sets of predictor variables (e.g., datasets representing actions performed by entities or transactions between entities) and one or more output variables (e.g., the probability of an entity performing some task or meeting some criteria).

[0004] A software development tool can be used to develop a model using data extracted from a data archive or other data source. Data mining and statistical methods can be used to develop model specifications. The software development tool can generate program code that is executed by data-processing platforms models in production environments. A production environment can include one or more computing systems for testing and implementing various operations that are represented by a model (e.g., risk prediction). When the program code is executed in the production environment, the efficacy of a model

can be evaluated, and the program code for implementing the model can be updated based on the evaluation.

[0005] Relying on the evaluation of the deployed programming code can present disadvantages. For example, the deployment and execution of program code for implementing a model can utilize computing resources (e.g., network bandwidth, processing cycles, memory, etc.) that may be otherwise deployed in a data-processing environment. If an executing of program code for implementing a model results in errors, a debugging process must distinguish errors specific to the software platform (e.g., coding errors for C++, Java, etc.) from errors in the modeling logic (e.g., improper training or calibration of the model). Thus, in some cases, relying on platform-specific evaluations of an analytical model may inefficiently utilize available computing resources.

[0006] Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

[0007] Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present disclosure as it existed before the priority date of each of the appended claims.

#### Summary

[0008] This disclosure relates to building and auditing predictive or analytical models that are used to generate source code for execution on data-processing platforms. There is provided a model integration system comprising: a processing device communicatively coupled via a data network to a client computing system; and a computer-readable medium

storing program code that provides a model integration tool and that is executable by the processing device; wherein the processing device is configured for executing the program code to perform operations comprising: receiving an electronic data-processing model that generates an analytical output by applying a function to one or more input attributes that are weighted with one or more respective modeling coefficients; identifying a target data-processing platform that requires one or more bin ranges associated with the one or more modeling coefficients and one or more reason codes associated with the one or more input attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output; providing a user interface having interface elements for: specifying bin range information identifying the one or more bin ranges and reason code information identifying the one or more reason codes; and uploading, to the computer-readable medium or an additional computer-readable medium accessible to the model integration tool, an audit file having an expected analytical output and an expected reason code; identifying the one or more bin ranges and the one or more reason codes based on user input received via the user interface and indicating the bin range information and the reason code information; generating Predictive Model Markup Language (PMML) code that implements the electronic data-processing model with the one or more bin ranges and the one or more reason codes; and outputting an executable source code in a programming language used by the target data-processing platform, wherein the executable source code is generated from the PMML code and implements the electronic data-processing model with the one or more bin ranges and the one or more reason codes, wherein the processing device is further configured for verifying, from a test of the PMML code, that a test analytical output matches the expected analytical output and that a test reason code matches the expected reason code, wherein verifying that

the test analytical output matches the expected analytical output and the test reason code matches the expected reason code comprises: accessing the audit file from the computer-readable medium or the additional computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code, and wherein prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code, the processing device performs further operations comprising: performing an additional test of the PMML code; determining that the additional test results in a failed test, the failed test comprising an additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code; outputting an indication that the additional test results in a failed test; in response to outputting the indication, receiving user input, via the user interface, specifying one or more updated characteristics of the electronic data-processing model; and generating, subsequent to the failed test, updated PMML code including the updated characteristics of the electronic data-processing model, wherein the updated PMML code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

[0009] There is also provided a method comprising: receiving, by a model integration tool executed by a processing device, an electronic data-processing model that generates an analytical output by applying a function to one or more input attributes that are weighted with one or more respective modeling coefficients, wherein the received electronic data-processing model lacks bin range information associated with the one or more modeling coefficients and reason code information associated with the one or more input attributes; providing, from the model integration tool to a client computing system, a user interface having interface elements for: specifying the bin range information and the reason code

information; and uploading, to a computer-readable medium accessible to the model integration tool, an audit file having the expected analytical output and the expected reason code; identifying, by the model integration tool and based on user input received via the interface elements, one or more bin ranges for the one or more modeling coefficients and one or more reason codes for the one or more input attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output; generating, by the model integration tool, modeling code that specifies the electronic data-processing model with the one or more bin ranges and the one or more reason codes; verifying, with the model integration tool, that a test of the modeling code results in a test analytical output matching an expected analytical output and a test reason code matching an expected reason code, wherein the verification comprises: accessing the audit file from the computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code; generating, by the model integration tool and subsequent to the verification, an executable source code that implements the electronic data-processing model on a data-processing platform that is accessible by the client computing system, wherein the executable source code is generated from the modeling code; and prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code: performing, with the model integration tool, an additional test of the modeling code; determining, by the model integration tool, that the additional test results in an additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code; outputting, by the model integration tool, an indication that the additional test results in a failed test; in response to outputting the indication, receiving user

input, via the interface elements, specifying one or more updated characteristics of the electronic data-processing model; and generating, with the model integration tool and subsequent to the additional test resulting in a failed test, updated modeling code including the updated characteristics of the electronic data-processing model, wherein the updated modeling code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

[0010] There is also provided a computer-readable medium having program code of a model integration tool that is executable by a processing device to cause a computing device to perform operations, the operations comprising: receiving an electronic data-processing model that generates an analytical output by applying a function to input attributes that are weighted with respective modeling coefficients; providing, to a client computing system, a user interface having interface elements for: specifying bin range information associated with the one or more modeling coefficients and reason code information associated with the one or more input attributes; and identifying, based on input received via the interface elements, bin ranges for the one or more modeling coefficients and reason codes for the one or more input attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output; generating Predictive Model Markup Language (PMML) code that specifies the electronic data-processing model with the one or more bin ranges and the one or more reason codes; verifying that a test of the PMML code results in a test analytical output matching an expected analytical output and a test reason code matching an expected reason code; and generating, subsequent to the verification, an executable source code that implements the electronic data-processing model on a data-processing platform that is accessible by the client computing system, wherein the executable source code is generated

from the PMML code, wherein verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code comprises: accessing the audit file from the computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code, and wherein prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code, the processing device performs further operations comprising: performing an additional test of the PMML code; determining that the additional test results in a failed test, the failed test comprising an additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code; outputting an indication that the additional test results in a failed test; in response to outputting the indication, receiving user input, via the user interface, specifying one or more updated characteristics of the electronic data-processing model; and generating, subsequent to the failed test, updated PMML code including the updated characteristics of the electronic data-processing model, wherein the updated PMML code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

[0011] In one example, a processor receives an electronic data-processing model, which generates an analytical output by applying a function to one or more input attributes that are weighted with respective modeling coefficients. The processor identifies a target data-processing platform that requires bin ranges associated with the modeling coefficients and reason codes associated with the input attributes. Each bin range identifies a range of input attribute values for which a modeling coefficient is valid. Each reason code identifies a reason code output indicating an impact of a weighted input attribute on the analytical output. The processor provides a user interface for specifying bin range information and reason code

information. The processor identifies the bin ranges and the reason codes based on user input received via the user interface. The processor generates Predictive Model Markup Language (“PMML”) code or other suitable modeling code that implements the electronic data-processing model with the bin ranges and the reason codes. The processor outputs source code, which is generated from the PMML code and implements the electronic data-processing model, in a programming language used by the target data-processing platform.

[0012] This illustrative example is mentioned not to limit or define the invention, but to aid understanding thereof. Other aspects, advantages, and features of embodiments of the invention will become apparent after review of the entire description and figures, including the following sections: Brief Description of the Figures, Detailed Description, and Claims.

#### Brief Description of the Figures

[0013] Some embodiments of the present invention are hereinafter described, by way of example only, with reference to the accompanying drawings, wherein:

[0014] FIG. 1 depicts an example of a system in which a model integration tool can be used for building and testing predictive or analytical models prior to generating source code for implementing the models on data-processing platforms, according to some aspects of the present disclosure.

[0015] FIG. 2 depicts an example of a method for building and testing predictive or analytical models prior to generating source code for implementing the models on data-processing platforms, according to some aspects of the present disclosure.

[0016] FIG. 3 depicts an example of an interface for mapping model attributes to production attributes of a data processing platform using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0017] FIG. 4 depicts an example of another interface for mapping model attributes to production attributes of a data processing platform using the system depicted in FIG. 1,

according to some aspects of the present disclosure.

[0018] FIG. 5 depicts an example of an interface for adding bin range information and reason code information to a model using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0019] FIG. 6 depicts an example of an interface for mapping model attributes involving decision trees to production attributes of a data processing platform using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0020] FIG. 7 depicts an example of an interface for configuring a model that involves decision trees, according to some aspects of the present disclosure.

[0021] FIG. 8 depicts an example of an audit file used for testing a model prior to source code generation using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0022] FIG. 9 depicts an example of an interface showing the results of an audit used for testing a model prior to source code generation using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0023] FIG. 10 depicts an example of an executing trace of an audit used for testing a model prior to source code generation using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0024] FIG. 11 depicts an example generating source code for a tested model using the system depicted in FIG. 1, according to some aspects of the present disclosure.

[0025] FIG. 12 depicts an example of a computing system for implementing certain features, according to some aspects of the present disclosure.

Detailed Description

[0026] This disclosure describes a model integration tool. The model integration tool can be executed by a suitable computing system to facilitate model the building and debugging

predictive or analytical models that implemented via source code deployed to data-processing platforms. The model integration tool can be used to perform one or more of building an analytical or predictive model, training the model, and testing the model. These operations can be performed using a model in an industry-standard modeling language or other suitable modeling language, prior to generating source code (e.g., C++, Java, etc.) that implements the model. An example of an industry-standard modeling language is the Predictive Model Markup Language (“PMML”) format. Building and testing the model in PMML or another modeling language can allow a model to be developed and tested without extensively debugging and revising the source code that is used to implement the modeling logic on data-processing platforms.

[0027] The model integration tool can facilitate customizing a model for a specific data-processing platform. For example, the model integration tool can provide features such as mapping production attributes to model attributes for a model, adding bin range and reason code data to the model, and testing the model using the mapping and the added bin range and reason code information. A model that is created using the model integration tool can be tested using a model execution engine without generating source code specific to certain data-processing platforms (e.g., C++ platforms, Java platforms, etc.).

[0028] This testing process, which uses PMML or other modeling code, allows errors in the modeling logic to be identified and corrected. In some aspects, these errors can be identified and corrected prior to generating source code for a target data-processing platform. When the model is deployed to a data-processing platform, computing resources are used more efficiently when evaluating the efficacy of the deployed model. For example, errors in the deployed model may be limited to platform-specific issues (e.g., issues specific to C++, Java, etc.) rather than errors in the modeling logic. Thus, a debugging process can focus on these platform-specific errors. In this manner, a computing system that utilizes the model

integration tool (e.g., a combination of computing devices used to generate a model and data-processing platforms to which the model are deployed) can reduce the inefficient use of computing resources, can increase the efficiency of debugging processes related to the deployed model, or both.

[0029] The features discussed herein are not limited to any particular hardware architecture or configuration. A computing device can include any suitable arrangement of components that provide a result conditioned on one or more inputs. Suitable computing devices include multipurpose, microprocessor-based computer systems accessing stored software that programs or configures the computing system from a general-purpose computing apparatus to a specialized computing apparatus implementing one or more aspects of the present subject matter. Any suitable programming, scripting, or other type of language or combinations of languages may be used to implement the teachings contained herein in software to be used in programming or configuring a computing device.

[0030] Referring now to the drawings, FIG. 1 is a block diagram illustrating an example of a model integration system 100 that can be used to generate and audit modeling code for implementing electronic data-processing models 110. Examples of a data-processing model 110 include a predictive model, an analytical model, or some other type of model.

[0031] The model integration system 100 can include a computing device or group of computing devices configured to execute the model integration tool 102. The model integration system 100 can be communicatively coupled, via a data network 119, to one or more client computing systems 124 that can be accessed by one or more users.

[0032] A client computing system 124 can include or can access a data-processing system 126. The data-processing system 126 can include one or more processing devices that implement a data-processing platform 128 using production data 130. A data-processing platform 128 can include one or more software components that use source code in a

particular programming language, such as C++ software, Java software, proprietary software, or any other programming language or architecture.

[0033] The model integration tool 102, which is executed by the model integration system 100, can perform one or more functions for generating and auditing modeling code. For example, the model integration tool 102 can include program instructions executable by one or more processing devices of the model integration system 100 to perform one or more operations. These operations can include creating, testing, and deploying one or more data-processing models 110. A data-processing models 110 can created and tested in an industry-standard modeling language or other suitable modeling language. One example of a modeling language is a set of functions and attributes specified in Extensible Markup Language (“XML”) or another suitable language for encoding functions and inputs to functions. Another example of an industry-standard modeling language is the Predictive Model Markup Language (“PMML”) format.

[0034] A data-processing model 110, such as an analytical model or predictive model, can be used for identifying relationships between sets of predictor variables and one or more output variables. Examples of predictor variables include attributes of an individual or other entity that can be obtained from credit files, financial records, or other data about the activities or characteristics of the entity. Attributes can include, for example, a credit data-related element, a single data element from a credit file, or an aggregation, calculation, or derivation of credit data to form a new data element. Examples of an output variable include a probability of an individual or other entity performing some task (e.g., defaulting on a financial obligation, responding to a sales offer, etc.), a probability of an individual or other entity meeting some criteria (e.g., being approved for a loan), etc.

[0035] The model integration tool 102 can facilitate creating and testing data-processing models by generating and providing one or more user interfaces 106. The user interfaces 106

can include interface elements for eliciting data required by a data-processing model 110 for deployment to a particular data-processing platform 128. Examples of these interfaces are described in detail herein.

[0036] The model integration tool 102 can have one or more features for using the data-processing model 110. Examples of these features include model creation, model maintenance, model auditing, model deployment, model execution, model importing, model downloading, and model specification generating. Model creation can involve creating logistics models, linear regression models, or other suitable analytical or predictive models. Functions for creating a model include one or more of configuring model metadata, setting up reason codes, creation of scorecards with segmentation, creation of scorecards with calibration, etc. Model maintenance can include updating and versioning of models. Model auditing can include testing a model by uploading an audit file containing attribute values as well as expected scores and reason codes, as described in detail herein. Model deployment can involve generating PMML, executable source code, or both, and providing the PMML or source code to a data-processing system 126 (either directly or via a client computing system 124). Model execution can involve providing the model to a model execution engine 103 (e.g., a predictive analytics decision engine such as the Adaptive Decision and Predictive Analytics (“ADAPA”) platform). Importing and downloading models can involve importing a model from a model development system 120 or a client computing system 124 and making the updated model available for download by a client computing system 124, a data-processing system 126, or both. Generating a model specification can include generating a human-readable file (e.g., a PDF document) that lists the model specifications for a data-processing model 110 that has been modified and audited by the model integration tool 102.

[0037] The model integration system 100 can include, or can otherwise access, a storage medium 108. The storage medium 108 can include one or more computer-readable media

included in or accessible to the model integration system 100. The storage medium 108 can store one or more of data-processing models 110, modeling coefficients 112, production attribute data 114, reason code data 116, and training data 118.

[0038] Modeling coefficients 112 can include regression coefficients or other coefficients that are used to weight various attributes in a data-processing model 110. In some aspects, modeling coefficients 112 can be generated using a model development platform 122 and provided to the model integration tool 102 with a data-processing model 110. In additional or alternative aspects, modeling coefficients 112 can be generated or modified at a client computing system 124 and transmitted to the model integration tool 102 via a data network 119 using one or more of the user interfaces 106.

[0039] Production attribute data 114 can include data describing one or more production attributes used by the data-processing platform 128. In some aspects, the model integration tool 102 can communicate with the data-processing platform 128 via the data network 119 to obtain the production attribute data 114. For example, the model integration tool 102 can transmit a request to the data-processing platform 128 for the production attribute data 114 and receive the production attribute data 114 in response. In additional or alternative aspects, production attributed data can be generated, modified, or otherwise obtained using a client computing system 124 and transmitted from the client computing system 124 to the model integration system 100 via a data network 119 using one or more of the user interfaces 106.

[0040] The reason code data 116 can include one or more reason codes. Reason codes can include data that identifies an impact of a particular attribute in a data-processing model 110 on an outcome generated by the data-processing model 110. For example, a reason code indicates the rationale for one or more types of information in a credit report (e.g., the aspects of an entity that resulted in a given credit score) that is generated using a data-processing model 110.

[0041] The training data 118 can include data used by the model integration tool 102 to train a data-processing model 110 that has been generated, updated, or modified using the model integration tool 102. In one example, training data 118 may include training inputs (e.g., credit files) and training outputs (e.g., credit scores, associated reason codes, etc.). One or more suitable training algorithms can be used to train the data-processing model 110 to identify relationships between the training inputs and the training outputs of the training data 118, as described in detail herein.

[0042] In some aspects, the model integration tool 102 can be accessed by one or more client computing systems 124 via an online service 104. In some aspects, the online service 104 can be a separate service that is executed by the model integration system 100 and that is accessible by the model integration tool 102. In other aspects, the online service 104 can be integrated with the model integration tool 102.

[0043] In some aspects, the model integration system 100 can communicate, via a data network 119, with a model development system 120 that is separate from the model integration system 100. The model development system 120 can include any computing device or group of computing devices that execute a suitable model development platform 122. The model integration system 100 can receive modeling code for one or more data-processing models from the model development system 120. The model integration tool 102 can consume this modeling code, which is generated and trained using the model development platform 122. Examples of suitable model development platforms 122 include Statistical Analysis System (“SAS”), Statistical Package for the Social Sciences (“SPSS”), R-programming, etc.

[0044] For example, a data-processing model 110 can be developed with a model development platform 122 such as SAS or SPSS. The data-processing model 110 can be modified using the model integration tool 102 and exported in a PMML format.

[0045] A data-processing model 110 that is created, modified, or otherwise used by the model integration tool 102 can be deployed to a model execution engine 103. The model execution engine 103 can perform some or all of the functions specified by the data-processing model 110. In some aspects, the model execution engine 103 can perform functions specified by the data-processing model 110 without the need to recode the data-processing model 110 for specific operating environments. An example of a specific operating environment is a data-processing system 126.

[0046] In some aspects, a user of the client computing system 124 can be authenticated by one or more of the model integration tool 102 and the online service 104. This authentication can verify that a user of the client computing system 124 is permitted to access the functionality of the model integration tool 102. In some aspects, one or more of the model integration tool 102 and the online service 104 can access subscriber accounts that are stored in, for example, a database or other suitable data source. The subscriber accounts can include data regarding respective users that are subscribers to an online service 104.

[0047] The model integration system 100 can electronically communicate with one or more client computing systems 124 via one or more data networks 119. For example, the online service 104 can receive requests from one or more users via the data network 119 to establish one or more subscriber accounts. The online service 104 can establish the subscriber accounts by storing authentication information (e.g., user names, passwords, etc.) that can be used to authenticate one or more users. Subsequent to establishing the subscriber accounts, the online service 104 that includes the model integration tool 102 can authenticate one or more users, thereby allowing the users to access the model integration tool 102.

[0048] In some aspects, the online service 104 can include or communicate with one or more of a web server module, a web services module, and an enterprise services module. For example, a web server module can be executed by a suitable processing device to provide one

or more web pages or other interfaces to the client computing system 124. The web services module can generate this content by executing one or more analytical algorithms using information retrieved from the storage medium 108. The enterprise services module can be executed to retrieve the information from the storage medium 108.

[0049] In some aspects, the model integration tool 102 provides functionality for managing the names of the production attributes that can be used in one or more models. Attributes can include factors used as variables in determining the score of an entity. The value of the attribute can be a natural number, a floating point number, a Boolean value, a character, or a series of characters. An attribute may be the result of the execution of another rule or a combination of other attributes. The model integration tool 102 can be used to create attribute catalogs that will be used in creating the model specification. An Attribute Catalog can include a list of attributes created for a specific organization, user, or other entity.

[0050] One example of a data-processing model 110 that can be tested and updated with the model integration tool 102 is a scorecard model. A scorecard model is an analytical function that can be used to calculate one or more metrics on one or more subject entities that assists in making a decision with respect to the subject entity. A complex scorecard (also known as “nested” or “segmented” scorecards) can include multiple intricate scorecards within one model that uses segmentation logic to identify the dependency, flow, and order in which the scoring from each of the scorecards will be included in the overall score.

[0051] In some aspects, the model integration tool 102 can facilitate, execute, or otherwise enable one or more operations for creating a scorecard. For example, FIG. 2 is a flow chart illustrating an example of a process 200 for generating and auditing modeling code for an electronic data-processing model. For illustrative purposes, the process 200 is described with reference to the implementation depicted in FIG. 1 and various other examples described herein. But other implementations are possible.

[0052] The process 200 can include receiving an electronic data-processing model that generates an analytical output by applying a function to input attributes that are weighted with respective modeling coefficients, as depicted at block 202. For example, one or more processing devices of model integration system 100 can execute the model integration tool 102 to receive a data-processing model 110. In some aspects, the model integration tool 102 can receive the data-processing model 110 from a client computing system 124. In additional or alternative aspects, the model integration system 100 can receive the data-processing model 110 from a model development system 120.

[0053] An input attribute can be an attribute that is identified as one of the inputs to the function included in the data-processing model. For example, an attributes can be a factor used as a variable in determining an analytical output (e.g., credit-related elements used as variables in determining credit scores). The value of the attribute can be a natural number, a floating point number, a Boolean value, a character, or a series of characters. In some aspects, an attribute may be the result of the execution of a rule or a combination of other attributes.

[0054] An analytical output can be a number (e.g., a score), a string (e.g., an identification of a reason code), or any other suitable output data that is generated by executing the data-processing model. In some aspects, analytical outputs indicate predictions of behavior by individuals or other entities. For example, an analytical output can have a value (e.g., a credit score) indicating a risk prediction. In additional or alternative aspects, analytical outputs indicate relationships among input attributes. For example, an analytical output can have a value (e.g., a credit score) indicating a degree to which certain input attributes impact other input attributes.

[0055] In some aspects, the received data-processing model 110 may lack certain data or functionality that is used with the data-processing model 110 when implemented at a data-

processing system 126. In one example, the received data-processing model 110 may lack one or more of bin range information associated with modeling coefficients or reason code information associated with the input attributes. For instance, a model development platform 122, which may be used to generate an initial data-processing model 110, may lack features for specifying bin range information, reason code information, or both. The model integration tool 102 can be used to update the data-processing model 110 to include one or more of bin range information, reason code information, or both, as described herein. Bin range information can include any data that specifies or otherwise indicates a bin range for one or more modeling coefficients. Reason code information can include any data that specifies or otherwise indicates one or more reason codes for a data-processing model 110.

[0056] A bin range can include a range of attributes values for which a given modeling coefficient is applicable to an attribute, or a set of ranges of attribute values for which different modeling coefficients are applicable to the attribute. In a simplified example, a data-processing model can include a function  $y = B_1X_1 + B_2X_2$ , where an output  $y$  is modeled as a function of attributes  $X_1$  and  $X_2$  that are weighted by the modeling coefficients  $B_1$  and  $B_2$ . The model may accurately model the behavior of a dataset for certain values of the attributes  $X_1$  and  $X_2$ . The bin ranges for this simplified example are the ranges of values of the attributes  $X_1$  and  $X_2$  that will result in accurate outputs from a model using the modeling coefficients  $B_1$  and  $B_2$ .

[0057] The process 200 can also include identifying a target data-processing platform that requires bin ranges associated with the modeling coefficients and reason codes associated with the input attributes, as depicted at block 204. For example, one or more processing devices of model integration system 100 can execute the model integration tool 102 to identify one or more data-processing platforms 126 to which a received data-processing model 110 will be deployed.

[0058] In some aspects, the model integration tool 102 identifies the data-processing system 126 based on user input that is received via one or more of the interfaces 106 and that specifies one or more features of the relevant data-processing platform 128. For example, the model integration tool 102 can store, in the storage medium 108 or another suitable computer-readable medium, data identifying one or more data-processing platforms 128 to which a data-processing model 110 may be deployed. The model integration tool 102 can use the stored data to generate an interface that lists one or more of the available data-processing platforms 128. The model integration tool 102 can provide the interface 106 to the client computing system 124. The model integration tool 102 can identify the target data-processing system 126 based on user input that is received from the client computing system 124 via the interface and that specifies one or more features of the relevant data-processing system 126.

[0059] The process 200 can also include providing a user interface having interface elements for specifying bin range information and reason code information, as depicted at block 206. For example, one or more processing devices of the model integration system 100 can execute the model integration tool 102 to provide one or more interfaces for specifying bin range information and reason code information. In some aspects, the model integration tool 102 can also provide one or more interfaces for mapping production attributes, as used by a data-processing platform 128, to model attributes included in a data-processing model 110. Examples of interfaces for mapping production attributes to model attributes and for specifying bin range information and reason code information are described herein with respect to FIGs. 3-5.

[0060] The model integration tool 102 can provide access to the user interface by a client computing system 124. In some aspects, providing access to the user interface involves transmitting data for the user interface from the model integration system 100 to the client

computing system 124. For example, the model integration tool 102 can configure the online service 104 to provide the user interface in a web page. The online service 104 can receive inputs from the client computing system 124 via the user interface. In additional or alternative aspects, providing access to the user interface involves transmitting data for the user interface from the model integration system 100 to a web server or other third-party system that is accessible by the client. For example, the model integration system 100 can transmit one or more instructions to a web server that cause the web server to present the user interface to the client computing system 124. The web server can receive inputs from the client computing system 124 via the user interface and transmit data indicating the user inputs to the model integration system 100.

[0061] The process 200 can also include identifying the bin ranges and the reason codes based on input received via the user interface, as depicted at block 208. For example, the model integration tool 102 can receive input from a client computing system 124 that was elicited via one or more of the interfaces provided at block 206. The model integration tool 102 can identify, from this received input, the bin ranges and reason code information to be used in a data-processing model 110 that is to be deployed to a given data-processing platform 128.

[0062] The process 200 can also include generating modeling code for the electronic data-processing model with the bin ranges and reason codes, as depicted at block 210. For example, one or more processing devices of the model integration system 100 can execute the model integration tool 102 to generate modeling code, such as PMML code. The modeling code can specify various aspects of a model, such as field names, function names, etc. The PMML or other suitable modeling code can be executed by a model execution engine 103, such as the ADAPA platform, without creating platform-specific source code (e.g., C++, Java, etc.) for implementing the model.

[0063] The process 200 can also include performing a test of the modeling code, as depicted at block 212. For example, one or more processing devices of the model integration system 100 can execute the model execution engine 103 in response to one or more commands received via the model integration tool 102. Examples of the testing process are described herein with respect to FIGs. 8-10.

[0064] Performing the test can involve, for example, providing an interface for uploading an audit file to a computer-readable medium that is accessible by the model integration tool 102. An audit file can include one or more expected analytical outputs, one or more expected reason codes, or some combination thereof. The audit file can also include test values for attributes used by the data-processing model 110 to be tested.

[0065] The model execution engine 103 can execute the PMML or other modeling code generated at block 210, using data from the audit file as input data, and thereby generate various test outputs. The test outputs can include test analytical outputs, such as, for example, scores generated by a scorecard data-processing model 110. The model integration tool 102 can compare the test analytical outputs to corresponding expected analytical outputs from the audit file. The test outputs can also include test reason codes. The model integration tool 102 can compare the test reason codes to corresponding expected reason codes from the audit file.

[0066] The process 200 can also include determining whether test outputs (e.g., one or more test analytical outputs, one or more test reason codes) match expected outputs (e.g., one or more expected analytical outputs, one or more expected reason codes), as depicted at block 214. If test analytical outputs match the expected outputs, the model integration tool 102 can output one or more messages indicating that the data-processing model 110 has been successfully tested or audited, as depicted at block 216. In some aspects, a match can be an exact match (e.g., a test score exactly matching an expected score). In additional or

alternative aspects, a test value can be considered a match to an expected value if the test value falls within a threshold range of expected values (e.g., within a certain range of an expected value identified in an audit file).

[0067] If the test analytical outputs match the expected outputs, the model integration tool 102 can output one or more messages indicating that the data-processing model 110 has failed the test or audit, as depicted at block 218. In one example, a test analytical output (e.g., a test score) may be sufficiently different from an expected analytical output that the data-processing model 110 does not provide the desired modeling results. In another example, a test reason code may be different from an expected reason code.

[0068] If the data-processing model 110 has failed the test or audit, the process can return to block 210. The model integration tool 102 can be used to update one or more characteristics of the data-processing model 110 (e.g., bin range information, coefficient information, processing logic, etc.). The model integration tool 102 can generate updated PMML code or other modeling code that includes the updates to the data-processing model 110 that were performed subsequent to the unsuccessful test or audit.

[0069] The model integration tool 102 can return to block 212 to perform additional tests or audits of the updated PMML code or other modeling code. If the additional tests or audits are successful, the process 200 can include outputting executable source code that implements the verified data-processing model in a programming language used by the data-processing platform, as depicted at block 220. For example, one or more processing devices of the model integration system 100 can generate executable source code (e.g., C++, Java, etc.) that is used by the data-processing platform 128 identified at block 204. The source code generated at block 220 can implement the data-processing model 110 with the bin range information and reason code information described at block 206, 208. The model integration tool 102 can execute a code generation process in which the successfully audited data-

processing model 110 is parsed, corresponding source code components are identified, and one or more source code files are generated that implement the data-processing model 110. An example of the code generation process is described herein with respect to FIG. 11.

[0070] The model development described in FIG. 2 is provided for illustrative purposes. Other implementations, however, are possible. In some aspects, one or more of the blocks depicted in FIG. 2 may be omitted. For example, the model integration tool 102 may generate and output executable source code after generating the PMML or other modeling code, without performing a test of the PMML code or other modeling code.

[0071] As discussed with respect to FIG. 2, the model integration tool 102 can be used for mapping the model attributes from a data-processing model 110 to production attributes in the production data 130. For example, a model attribute can include the attribute name that is used in PMML code or other modeling code that is outputted from a model development platform 122. The name of the model attribute may differ from a name of a corresponding attribute in production data 130 that is used by a data-processing platform 128 for which the data-processing model 110 is being developed. The model integration tool 102 can be used to create a mapping between the model attribute and the production attribute, thereby allowing the production data 130 or corresponding training data 118 to be used with the data-processing model 110 under development.

[0072] For example, the model integration tool 102 can generate and provide an interface that includes one or more interface elements for creating a mapping between an attribute from a trained model and an attribute from a production environment that includes the data-processing platform 128 and the production data 130. For example, the model integration tool 102 can communicate with the client computing system 124, the data-processing system 128, or another suitable computing device to identify one or more attributes from the production data 130. The model integration tool 102 can generate an interface that displays

one or more of the identified attributes from the production data 130. When generating the interface, the model integration tool 102 can also include one or more interface elements that display one or more of the attributes from the training model. For example, the model integration tool 102 can access an imported file with the attributes from the training model, retrieve the attributes from the imported file, and include the retrieved attributes in the interface.

[0073] The model integration tool 102 can receive input from a user identifying a mapping between an attribute from the training model and a corresponding attribute from the production data 130. For example, the input can include a selection of an attribute from the training model as displayed in the interface, and a selection of an attribute from the production data 130 as displayed in the interface, and a selection of a command to create a mapping. In response to receiving the input, the model integration tool 102 can create a mapping between an attribute from the training model and a corresponding attribute from the production data 130. For example, the model integration tool 102 can create an entry in a table or other data structure stored in the storage medium 108 (or another suitable computer-readable medium) that identifies the association between the attribute from the training model and the corresponding attribute from the production data 130.

[0074] In the example depicted in FIG. 3, the user can access different production attributes from the production attribute data 114 via the interface 300. For instance, the menu element 302 can allow the user to indicate a production attribute from a drop-down menu. The interface 300 can also depict the names of model attributes from the data-processing model 110, one of which can be selected via the radio buttons 304. To complete the mapping of model attributes to production attributes, the user can click the appropriate one of the radio buttons 304. In response to receiving user input indicating a selection of a production attribute and a modeling attribute via the interface 300, the model integration tool 102 can

create a mapping between the production attribute and the modeling attribute in the storage medium 108.

[0075] FIG. 4 depicts an example of another interface 400 for mapping production attributes to model attributes. In the example depicted in FIG. 4, the user can access different production attributes from the production attribute data 114 via the interface 400. For instance, the menu elements 402 can allow the user to indicate different production attributes from drop-down menus. The interface 400 can also depict the names of model attributes from the data-processing model 110 (e.g., the model attributes listed under the “SAS Name” column). To complete the mapping of model attributes to production attributes, the user can select production attributes to be mapped to the displayed model attributes. In response to receiving user input indicating a selection of various production attributes via the interface 400, the model integration tool 102 can create a mapping between the production attribute and the modeling attribute in the storage medium 108.

[0076] Another operation for creating a scorecard involves obtaining regression coefficients for a model. The model integration tool 102 can be used to identify, input, or otherwise provide regression coefficients from a training model. The model integration tool 102 can provide an interface that includes one or more interface elements for identifying regression coefficients. For example, the interface 400 can also include fields 404 that allow the user to enter regression coefficients or other modeling coefficients to be applied to the attributes used in a data-processing model 110. (Other implementations can use other interface elements for entering regression coefficients.)

[0077] In some aspects, an interface presented by the model integration tool 102 can include one or more interface elements, such as fields, for identifying the location of a file in which the regression coefficients are stored. The interface can also include a command element that allows a user to send an “import” command to the model integration tool 102.

The model integration tool 102 can receive input from a user identifying the regression coefficients. For example, the model integration tool 102 can receive data that is entered into the field element and that identifies the location of the file in which modeling coefficients 112 (e.g., regression coefficients) are stored. The interface can also receive an import command that is triggered by the command element being clicked or otherwise selected in the interface.

[0078] In response to receiving the input, the model integration tool 102 can import the file that includes the modeling coefficients 112. In some aspects, the model integration tool 102 can retrieve modeling coefficients 112 from a file system that is stored on the storage medium 108. In additional or alternative aspects, the model integration tool 102 can send a request for the file to the data source at which the modeling coefficients 112 are stored, and can receive a responsive transmission from the data source that includes the modeling coefficients 112.

[0079] In one example, the user can add a scorecard for a model by importing a regression table, which can include a set of regression coefficients, or other modeling coefficients 112. The regression table can be a delimited file that generated using, for example, a model development platform 122 that has been used to develop, at least partially, a scorecard model. A user can choose and upload the delimited file with regression coefficients. The imported coefficients can be used to populate the values under the “Points” column depicted in the interface 400 of FIG. 4.

[0080] As discussed above with respect to FIG. 2, the model integration tool 102 can be used to add bin range information to a data-processing model 110. In some aspects, bin ranges are a one-to-many relationship with respect to the attributes utilized in the scorecard or other data-processing model 110.

[0081] The model integration tool 102 can provide an interface that includes one or more interface elements for obtaining bin ranges. For example, FIG. 5 depicts an example of an interface 500 that can be used by the model integration tool 102 to obtain bin range information. The interface 500 can include a table 501 that is used to identify bin ranges for different regression coefficients. The model integration tool 102 can receive input from a user identifying bin ranges. For example, a user can modify the table 501 to specify bin ranges for different regression coefficients. In response to receiving the input, the model integration tool 102 can store the bin ranges in a data file that is associated with a given data-processing model 110.

[0082] In the example depicted in FIG. 5, the interface 500 displays mapped attributes and the corresponding points for a scorecard. Examples of the fields that can be included in this interface are fields 504 (labeled “Attribute ID and Description”), 406 (labeled “Bin Range”), 408 (labeled “Points”), 410 (labeled “Bin Type (Original)”), 412 (labeled “Transformation”), 414 (labeled “Reason Code”), and 416 (labeled “Return 5th RC?”).

[0083] The field 504 can display production attributes from the production attribute data 114 that have been mapped to model attributes from a data-processing model 110 (e.g., an uploaded scorecard model file). In the example depicted in FIG. 5, field 504 displays an attribute identifier for a production attribute. In some aspects, multiple model attributes from the data-processing model 110 can be mapped to the same attribute identifier for a production attribute.

[0084] The field 506 can allow a user to input a bin range. The bin range can be a range of attribute values for which a corresponding modeling coefficient (e.g., a regression coefficient from an uploaded scorecard model) is applicable. An attribute can have multiple bin ranges associated with it. In the example depicted in FIG. 5, each bin range is displayed in a separate row under the attribute. The field 508 can be used to display or input a

regression coefficient associated with a particular bin range for a particular attribute.

[0085] The field 510 can allow a user to specify whether the attribute values for a given modeling coefficient can be used as-is or whether the attribute values must be transformed in some way (e.g., through an equation to normalize the attribute values). For example, if the “original” attribute values can be used, a user can select “Yes” for the bin type. If the “original” attribute values must be transformed for use in the model, a user can select “No” for the bin type.

[0086] If a bin must undergo transformation, the transformation expression can be entered in to the field 512. For example, a user can enter a mathematical expression using, as inputs, the attribute value and the modeling coefficient (e.g., “points”) for the corresponding bin range.

[0087] The field 514 can allow a user to select a reason code for the model. For example, the field 514 can be a drop-down menu with available reason codes for the model. A reason code can be selected that corresponds to the value in the bin range.

[0088] As discussed above with respect to FIG. 2, the model integration tool 102 can be used to add reason code information to a data-processing model 110. Reason codes can, for example, indicate the rationale for one or more types of information in a credit report (e.g., the aspects of an entity that resulted in a given credit score). The model integration tool 102 can provide an interface that includes one or more interface elements for identifying reason codes. For example, a user interface provided by the model integration tool 102 can include a field element for identifying the location of a file in which the reason codes are stored. The interface can also include a command element that allows a user to send an “import” command to the model integration tool 102.

[0089] The model integration tool 102 can receive input from a user identifying the reason codes. For example, the model integration tool 102 can receive data that is entered

into the field element and that identifies the location of the file in which the reason codes are stored. The interface can also receive an import command that is triggered by the command element being clicked or otherwise selected in the interface. In response to receiving the input, the model integration tool 102 can import the file that includes the reason codes. In one example, the model integration tool 102 can retrieve the file from a file system that is stored on the storage medium 108 or another computer-readable medium accessible to the model integration system 100. In another example, the model integration tool 102 can send a request for the file to the data source at which the file is stored and can receive a responsive transmission from the data source that includes the file.

[0090] In some aspects, a data-processing model 110 that includes scorecard functionality can include multiple scorecard functions. Each of the scorecard functions can be referred to as “Segments” in PMML model that is generated from the data-processing model 110. For example, when a scorecard model is executed, only one scorecard (segment) may be used in calculating the model score. This segment is called the “winning segment.” The model integration tool 102 can present an interface to a user for defining segmentation logic, which is used to determine which segment is to be selected as the winning segment when a data-processing platform 128 executes the data-processing model 110. The segmentation logic can include an expression that defines the conditions on which a given segment (e.g., a particular scorecard function) will be selected, at runtime, to generate an analytical or predictive output (e.g., by calculating a score using the model). If a data-processing model 110 has more than one scorecard, then each scorecard can be addressed via the segmentation logic.

[0091] Segmentation logic can be defined using one or more suitable interfaces. For example, the model integration tool 102 can present an interface for selecting, entering, or otherwise inputting a segmentation expression. The segmentation expression can include

mathematical conditions applied to one or more attributes that are used in a data-processing model 110. In some aspects, the segmentation expression for a data-processing model 110 can also refer to attributes that are not included in the data-processing model 110. For example, the segmentation logic can refer to an attribute in the production attribute data 114 or any other suitable attribute that is defined in an attribute catalog associated with a data-processing model 110 that is under development.

[0092] In additional or alternative aspects, the model integration tool 102 can facilitate, execute, or otherwise enable one or more operations for developing a data-processing model 110 that incorporates multiple decision trees or other analytical functions. In one example, a data-processing model 110 may use a prediction of card usage to develop a spending estimate for one or more entities. The data-processing model 110 can use different decision trees to develop different predictions of card usage. For instance, a first decision tree may be developed using a “thick” credit file (e.g., a credit file having a number of tradelines that exceeds a threshold number). A second decision tree may be developed using a “thin” credit file (e.g., a credit file having a number of tradelines that is less than the threshold number). Modeling logic in a data-processing model 110 can obtain a prediction of card usage from one of the decision trees and use the prediction to develop a spending estimate. During execution of the data-processing model 110, a particular decision tree can be selected based on whether a credit file to be used is a thin file or a thick file.

[0093] One operation for developing a model that uses multiple decision trees involves obtaining a decision tree or multiple decision trees that are to be added to or otherwise included in a model. The model integration tool 102 can be used to identify, input, or otherwise provide decision trees that are developed using a training model. The model integration tool 102 can provide an interface that includes one or more interface elements for identifying decision trees. For example, the interface can include a field element for

identifying the location of a PMML file or other modeling code in which one or more decision trees are stored. The interface can also include a command element that allows a user to send an “import” command to the model integration tool 102.

[0094] In some aspects, the model integration tool 102 can receive user input identifying the decision trees. For example, the model integration tool 102 can receive data that is entered into an appropriate user interface and that identifies the location of the file in which the decision trees are stored. The interface can also receive an import command that is triggered by the command element being clicked or otherwise selected in the interface. In response to receiving the user input obtained via the interface, the model integration tool 102 can import the file that includes the decision trees. For example, the model integration tool 102 can send a request for the file to the data source at which the file is stored and can receive a responsive transmission from the data source that includes the file.

[0095] Another operation for developing a model that uses multiple decision trees involves mapping attributes used in one or more decision trees to attributes in a production environment, such as the data-processing system 126. The model integration tool 102 can provide an interface, such as the interface 600 depicted in FIG. 6, that includes one or more interface elements for creating a mapping between an attribute from a model and an attribute from a production environment. The model integration tool 102 can access the production attribute data 114 to identify one or more attributes from the production environment. The model integration tool 102 can generate the interface 600, which includes one or more interface elements 502 for displaying or selecting one or more of the identified attributes from the production attribute data 114. The model integration tool 102 can also include one or more interface elements 604 that display one or more of the attributes used in the decision tree. For example, the model integration tool 102 can access an imported file with the

attributes used in the decision tree, retrieve the attributes from the imported file, and include the retrieved attributes in the interface.

[0096] The model integration tool 102 can receive input from a user identifying a mapping between an attribute used in the decision tree and a corresponding attribute from the production environment. For example, the input can include a selection of an attribute used in the decision tree and a selection, via one or more of the interface elements 602, of a production attribute from the production environment. The input can also include a selection of a command to create a mapping. In response to receiving the input, the model integration tool 102 can create a mapping between an attribute used in the decision tree and a corresponding attribute from the production environment. For example, the model integration tool 102 can create an entry in a table or other data structure stored in a computer-readable medium that identifies the association between the attribute used in the decision tree and the corresponding attribute from the production environment.

[0097] The interface 700 depicted in FIG. 7 can be used to import a decision tree file or otherwise configure a data-processing model 110 for using a decision tree function. The “Predicted Output” field can allow a user to select the name of the output that the decision tree is expected to return. The “Output Feature Type” field can allow a user to select the type of output that the decision tree is expected to return (e.g., “Probability” and “Predicted”). The “Output Data Type” field can allow a user to select whether the decision tree output is Numeric or String type. If “Probability” is selected as the “Output Feature Type,” the “Output Data Type” may be limited to “Numeric.” If “Predicted” is selected as the “Output Feature Type,” the “Output Data Type” may be “Numeric” or “String.”

[0098] Another operation for developing a model that uses multiple decision trees involves providing segmentation logic for selecting a given decision tree. The model integration tool 102 can provide an interface that includes one or more interface elements for

these criteria. A segmentation expression entered into this interface can include mathematical conditions applied to one or more attributes that are used in the model. For example, a user can input data indicating that a number of trade lines above a threshold should trigger the use of a first decision tree and that a number of trade lines below a threshold should trigger the use of a second decision tree. In response to receiving the input, the model integration tool 102 can store the criteria in a data file that is associated with the model to be developed. Similarly to the selection of different modeling segments, as described above, the segmentation expression for a decision tree can also contain attributes that are not included in the decision tree.

[0099] In additional or alternative aspects, the model integration tool 102 can provide one or more suitable interfaces that elicit input identifying how output from the decision tree is outputted. For example, if a data-processing model 110 has multiple decision trees, the final output can be a combination of the output of those trees or it can be the output of the last decision tree executed.

[0100] The model integration tool 102 can use the data obtained via the various interfaces described herein to generate and build modeling code for implementing the functionality of a data-processing model 110. In some aspects, the model integration tool 102 can use the data obtained via the various interfaces described above to build a specification for a model that uses multiple scorecards, multiple decision trees, or some combination thereof. The model integration tool 102 can generate modeling code, such as PMML code, for implementing one or more functions of a data-processing model 110. For instance, the model integration tool 102 can generate PMML code or other suitable modeling code for implementing the data-processing model 110 with multiple scorecards, multiple decision trees, or both. The PMML code or other suitable modeling code can be generated based on the specification. In some aspects, generating the PMML code or other modeling code can involve updating PMML

code or other modeling code that has been received from a model development system 120 or a client computing system 124. In additional or alternative aspects, generating the PMML code or other modeling code can involve generating a new set of PMML code or other modeling code to implement a model that has been developed using the model integration system 100.

[0101] In some aspects, the PMML code or other modeling code can also be used to train the model with the aspects identified in a model specification that is generated via the model integration tool 102. For example, the model integration tool 102 can access training data 118 from the storage medium 108 or other computer-readable medium. The training data 118 can include training inputs (e.g., credit files) and training outputs (e.g., credit scores, associated reason codes, etc.). In some aspects, machine-learning algorithms (e.g., decision trees, neural nets, random forest, etc.) can be generated using model development platforms 122, such as SPSS and SAS eMiner. These machine-learning algorithms can be used to train or otherwise modify a data-processing model 110 that is managed using the model integration tool 102.

[0102] Training the model can include modifying PMML code or other modeling code such that a data-processing model 110 provides desired analytical or predictive outputs based on certain attribute values. For example, a data-processing model 110 can be trained to assign reason codes to certain attributes based on the modeling coefficients 112, the reason code data 116, and the bin ranges that have been obtained using the interfaces described herein.

[0103] As described herein with respect to FIG. 2, the model integration tool 102 can also provide model-auditing features. For instance, the PMML code or other modeling code that implements a data-processing model 110 can be used to audit the data-processing model 110. In one example, a data file (e.g., a delimited file such as a comma-separated values (“CSV”))

file) can be imported into the model integration tool 102. The delimited data file can include sample input data for auditing the model and sample expected outputs. Examples of sample input data include a set of attribute values calculated from credit data, financial data, or other data indicating various characteristics or activities of one or more entities. Examples of sample expected outputs include credit scores, risk scores, reason codes, etc.

[0104] The model integration tool 102 can use the model execution engine 103 to test or otherwise execute the data-processing model 110 using the sample input data. The model integration tool 102 can determine whether the sample expected outputs match the outputs that have been obtained by executing the model using the sample input data. The extent to which the sample expected outputs match the obtained outputs can indicate the success of the auditing process.

[0105] FIG. 8 depicts an example of a delimited data file 800 that can be used for auditing a data-processing model 110. The field 802 can include unique numeric identifiers for different records in a test data set contained in the delimited data file 800. The field 804 can identify an expected predictive or analytical output (e.g., a score) based on a given set of input attribute values included in the delimited data file 800. The model integration tool 102 can compare an expected predictive or analytical output with a calculated predictive or analytical output generated by the model integration tool 102 performing a test of the data-processing model 110 (e.g., auditing the data-processing model 110).

[0106] The fields 806 can identify different expected reason codes that are to be outputted when implementing the data-processing model 110. In this example, five reason codes are returned for a scorecard model, so the delimited data file 800 includes a field for each of the five reason codes. The expected reason codes are compared with the calculated reason codes generated by the model integration tool 102 performing a test of the data-processing model 110 (e.g., auditing the data-processing model 110).

[0107] The fields 808 can identify a set of attribute values to be used in a test of the data-processing model 110. In the example depicted in FIG. 8, this set of attribute values can also include attributes that are used by segmentation logic used within a data-processing model 110 (e.g., segmentation logic for selecting a specific decision tree or other analytical function).

[0108] The results of an audit can be depicted in a suitable interface. FIG. 9 depicts an example of a results interface 900 generated by a test of a data-processing model 110 by the model integration tool 102. In this example, the results interface 900 identifies a name for the audit, a file name for the audit file (e.g., the delimited data file 800) that was used to perform a test of a data-processing model 110, a model baseline for which the audit was run. The results interface 900 can also identify a number of records that were successfully processed from the audit file, a percentage of records in the audit file for which both the returned score and the returned reason codes matched the expected score and reason codes from the uploaded file, a percentage of records in the file for which the returned score matched the expected score from the uploaded file, and a percentage of records in the file for which the returned reason codes matched the expected reason codes from the uploaded audit file. The results interface 900 can also identify the date of the audit and the execution status (e.g., “Complete” or “Failed”) for the audit.

[0109] In some aspects, the model integration tool 102 can generate an execution trace for an audited data-processing model 110. The execution trace can include details of the calculated values of attributes, scores, etc. at each stage of the model execution. FIG. 10 depicts an example of an execution trace 1000.

[0110] As described herein with respect to FIG. 2, the model integration tool 102 can deploy a data-processing model 110 into a production environment that includes the data-processing system 126. For example, the model integration tool 102 can receive, via an

interface, input from a user indicating that the user wishes to deploy the model (e.g., a selection of a “Deploy” command in a menu). The model integration tool 102 can identify any platform-specific aspects of the production environment. For example, the production environment may include one or more computing systems that perform data processing using C++ software, Java software, proprietary software, or any other programming language or architecture that is different from a modeling language that is used to generate and audit the model.

[0111] In some aspects, the model integration tool 102 can identify platform-specific aspects based on one or more user inputs received via an interface that specify the platform-specific aspects of the production environment (e.g., receiving input specifying that the audited model should be deployed to a C++ production environment). In additional or alternative aspects, the model integration tool 102 can identify the production environment based on one or more user inputs received via an interface (e.g., selection of “Production Environment 1” from a list of production environments) and can identify the platform-specific aspects based on stored data about the selected environment (e.g., one or more data files indicating that the Production Environment 1 uses C++).

[0112] The model integration tool 102 can include a code generator that creates an object to collect model information required for the code generation. The code generator can employ a chain of code generation commands responsible of generating the C++ project. For example, FIG. 11 depicts an example of a process flow for deploying a model to a platform-specific environment.

[0113] In the example depicted in FIG. 11, a controller bean 1102 of the model integration tool 102 can call a `transformToCPP()` method on a transformation service in response to a transform command receive via a user interface 1100 of the model integration tool 102. The method can create a `PredictiveModel` object containing model data. The

method can provide the PredictiveModel object and a code generation chain name to a code generator module that is included in or accessible to the model integration tool 102. The code generator 1104 of the model integration tool 102 can output C++ codes via a ScoreModel Code Generation Chain 1106.

[0114] An example of a model outputted for a C++ platform can include the following files:

- Make file: makefile
- Model entry: SCPD{model id}.cpp
- Test stub: Test{mode id}.cpp
- Model header: CMDL{model id}.h
- Model implementation: CMDL{model id}.cpp
- Scorecard implementation (one per scorecard): CMDL{model id}Scorecard{scorecard id}.cpp

[0115] One example of generating C++ code from a PMML-based model is implementing a reject code function in C++. A reject code can be assigned when a score cannot be derived. Normally, one reject code per scorecard only is assigned. To implement the reject code function, the code generator looks for reject codes contained in logic of a “Reject Code” output field in the PMML file. If the code generator finds a supported reject code, the code generator can generate the corresponding C++ codes in the reject () method in the model implementation class.

[0116] Another example of generating C++ code from a PMML-based model is implementing a 5th reason code function in C++. A 5th reason code can be a 'Y' or blank flag that is returned for a model and can indicate whether inquiries affected a score. To implement the 5th reason code function, the code generator looks for the 5th reason code logic from a “Reason Code 5” output field in the PMML file. A method (e.g.

“calculate5thReasonCode”) can be generated in a model implementation class to calculate the 5th reason code. The method can take the winning segment/scorecard identifier as argument. Depending on the actual 5th reason code logic, some class level attributes may also be calculated and used along with the winning segment/scorecard identifier to determine the inquiry score impact flag value (“Y” or “ ”), and then set the flag to the score object. At the execution time, the method is called with the score() method after a winning scorecard is selected and calculated, but before a finalCalculations() method.

[0117] Another example of generating C++ code from a PMML-based model is implementing attribute definition and calculations in C++. Each attribute can be coded as a C++ class. Some attributes can have dependencies on others. The code generator can load the attribute dependency information from data files and automatically generate codes to calculate dependent attributes as well as the primary attributes. Attributes used in the segmentation and 5th reason code logic can be calculated at a score model level. Attributes used in the scorecards can be calculated within scorecard implementation codes. For example, the model integration tool 102 can read credit records from a suitable dataset, and supply the parsed credit record information to aggregate all the attributes required for a score model.

[0118] Any suitable computing system or group of computing systems can be used to perform the operations described herein. For example, FIG. 12 is a block diagram depicting an example of a model integration system 100. The example of the model integration system 100 can include various devices for communicating with one or more of the model development system 120, the client computing system 124, and the data-processing system 126. The example of the model integration system 100 can also include various devices for storing the datasets used for developing and auditing data-processing model 110 (e.g., the modeling coefficients, production attribute data, reason code data, and training code data), as

described above with respect to FIG. 1. The example of the model integration system 100 can also include various devices for performing one or more operations described above with respect to FIGs. 1-11.

[0119] The model integration system 100 can include a processor 1202 that is communicatively coupled to a memory 1204. The processor 1202 executes computer-executable program code stored in the memory 1204, accesses information stored in the memory 1204, or both. Examples of a processor 1202 include a microprocessor, an application-specific integrated circuit, a field-programmable gate array, or any other suitable processing device. The processor 1202 can include any number of processing devices, including one. The processor 1202 can include or communicate with a memory 1204. The memory 1204 stores program code that, when executed by the processor 1202, causes the processor to perform the operations described in this disclosure.

[0120] The memory 1204 can include any suitable computer-readable medium (e.g., the storage medium 108 depicted in FIG. 1). The computer-readable medium can include any electronic, optical, magnetic, or other storage device capable of providing a processor with computer-readable program code or other program code. Non-limiting examples of a computer-readable medium include a magnetic disk, memory chip, optical storage, flash memory, storage class memory, a CD-ROM, DVD, ROM, RAM, an ASIC, magnetic tape or other magnetic storage, or any other medium from which a computer processor can read and execute program code. The program code may include processor-specific program code generated by a compiler or an interpreter from code written in any suitable computer-programming language. Examples of suitable programming language include C, C++, C#, Visual Basic, Java, Python, Perl, JavaScript, ActionScript, etc.

[0121] The model integration system 100 may also include a number of external or internal devices such as input or output devices. For example, the model integration system

100 is shown with an input/output interface 1208 that can receive input from input devices or provide output to output devices. A bus 1206 can also be included in the model integration system 100. The bus 1206 can communicatively couple one or more components of the model integration system 100.

[0122] The model integration system 100 can execute program code that includes the model integration tool 102, the model execution engine 103, and the online service 104. Although FIGs. 1 and 12 depict the model integration tool 102, the model execution engine 103, and the online service 104 as separate software modules executed by the same system, other implementations are possible. In some aspects, a model integration tool 102 can be an application that includes software modules that implement one or more of the model execution engine 103 and the online service 104. In other aspects, a model integration tool 102 can be an application that communicates with one or more computing devices over a data network 119 that execute one or more of the model execution engine 103 and the online service 104.

[0123] The program code for the model integration tool 102, the model execution engine 103, and the online service 104 may be resident in any suitable computer-readable medium and may be executed on any suitable processing device. For example, as depicted in FIG. 12, the program code for the model integration tool 102, the model execution engine 103, and the online service 104 can reside in the memory 1204 at the model integration system 100. Executing one or more of the model integration tool 102, the model execution engine 103, and the online service 104 can configure the processor 1202 to perform the operations described herein.

[0124] In some aspects, the model integration system 100 can include one or more output devices. One example of an output device is the network interface device 1210 depicted in FIG. 12. A network interface device 1210 can include any device or group of devices

suitable for establishing a wired or wireless data connection to one or more data networks 119. Non-limiting examples of the network interface device 1210 include an Ethernet network adapter, a modem, etc. Another example of an output device is the presentation device 1212 depicted in FIG. 12. A presentation device 1212 can include any device or group of devices suitable for providing visual, auditory, or other suitable sensory output indicating recommendations generated by the model integration tool 102. Non-limiting examples of the presentation device 1212 include a touchscreen, a monitor, a speaker, a separate mobile computing device, etc.

[0125] General Considerations

[0126] Numerous specific details are set forth herein to provide a thorough understanding of the claimed subject matter. However, those skilled in the art will understand that the claimed subject matter may be practiced without these specific details. In other instances, methods, apparatuses, or systems that would be known by one of ordinary skill have not been described in detail so as not to obscure claimed subject matter.

[0127] Unless specifically stated otherwise, it is appreciated that throughout this specification that terms such as “processing,” “computing,” “calculating,” “determining,” and “identifying” or the like refer to actions or processes of a computing device, such as one or more computers or a similar electronic computing device or devices, that manipulate or transform data represented as physical electronic or magnetic quantities within memories, registers, or other information storage devices, transmission devices, or display devices of the computing platform.

[0128] The system or systems discussed herein are not limited to any particular hardware architecture or configuration. A computing device can include any suitable arrangement of components that provides a result conditioned on one or more inputs. Suitable computing devices include multipurpose microprocessor-based computer systems accessing stored

software that programs or configures the computing system from a general purpose computing apparatus to a specialized computing apparatus implementing one or more aspects of the present subject matter. Any suitable programming, scripting, or other type of language or combinations of languages may be used to implement the teachings contained herein in software to be used in programming or configuring a computing device.

[0129] Aspects of the methods disclosed herein may be performed in the operation of such computing devices. The order of the blocks presented in the examples above can be varied—for example, blocks can be re-ordered, combined, or broken into sub-blocks. Certain blocks or processes can be performed in parallel.

[0130] The use of “adapted to” or “configured to” herein is meant as open and inclusive language that does not foreclose devices adapted to or configured to perform additional tasks or steps. Additionally, the use of “based on” is meant to be open and inclusive, in that a process, step, calculation, or other action “based on” one or more recited conditions or values may, in practice, be based on additional conditions or values beyond those recited. Headings, lists, and numbering included herein are for ease of explanation only and are not meant to be limiting.

[0131] While the present subject matter has been described in detail with respect to specific aspects thereof, it will be appreciated that those skilled in the art, upon attaining an understanding of the foregoing, may readily produce alterations to, variations of, and equivalents to such aspects. Any aspects or examples may be combined with any other aspects or examples. Accordingly, it should be understood that the present disclosure has been presented for purposes of example rather than limitation, and does not preclude inclusion of such modifications, variations, or additions to the present subject matter as would be readily apparent to one of ordinary skill in the art.

Claims

1. A model integration system comprising:
  - a processing device communicatively coupled via a data network to a client computing system; and
  - computer-readable medium storing program code that provides a model integration tool and that is executable by the processing device;
  - wherein the processing device is configured for executing the program code to perform operations comprising:
    - receiving an electronic data-processing model that generates an analytical output by applying a function to one or more input attributes that are weighted with one or more respective modeling coefficients;
    - identifying a target data-processing platform that requires one or more bin ranges associated with the one or more modeling coefficients and one or more reason codes associated with the one or more input attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output;
    - providing a user interface having interface elements for:
      - specifying bin range information identifying the one or more bin ranges and reason code information identifying the one or more reason codes; and
      - uploading, to the computer-readable medium or an additional computer-readable medium accessible to the model integration tool, an

audit file having an expected analytical output and an expected reason code;

identifying the one or more bin ranges and the one or more reason codes based on user input received via the user interface and indicating the bin range information and the reason code information;

generating Predictive Model Markup Language (PMML) code that implements the electronic data-processing model with the one or more bin ranges and the one or more reason codes; and

outputting an executable source code in a programming language used by the target data-processing platform, wherein the executable source code is generated from the PMML code and implements the electronic data-processing model with the one or more bin ranges and the one or more reason codes,

wherein the processing device is further configured for verifying, from a test of the PMML code, that a test analytical output matches the expected analytical output and that a test reason code matches the expected reason code, wherein verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code comprises: accessing the audit file from the computer-readable medium or the additional computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code, and

wherein prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code, the processing device performs further operations comprising:

performing an additional test of the PMML code;

determining that the additional test results in a failed test, the failed test

comprising an additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code;

outputting an indication that the additional test results in a failed test;

in response to outputting the indication, receiving user input, via the user interface, specifying one or more updated characteristics of the electronic data-processing model; and

generating, subsequent to the failed test, updated PMML code including the updated characteristics of the electronic data-processing model, wherein the updated PMML code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

2. The model integration system of claim 1, wherein the user interface further comprises one or more additional interface elements for:

identifying a mapping of the one or more input attributes to one or more production attributes from the target data-processing platform, and

identifying one or more transformations of one or more respective values of the one or more production attributes into one or more respective values that are compliant with the electronic data-processing model,

wherein the PMML code implements the mapping and the one or more transformations.

3. The model integration system of any of the preceding claims, wherein the electronic data-processing model comprises a regression model and wherein the one or more modeling

coefficients comprise regression coefficients.

4. The model integration system of any of the preceding claims, wherein the user interface further comprises additional interface elements for identifying a first analytical function to be added to the electronic data-processing model, identifying a second analytical function to be added to the electronic data-processing model, and indicating a specific attribute from the one or more input attributes whose value controls a selection of the first analytical function or the second analytical function;

wherein the PMML code implements the first analytical function, the second analytical function, and segmentation logic that selects the first analytical function based on the input attribute having a first value and selects the second analytical function based on the input attribute having a second value.

5. The model integration system of claim 4, wherein each of the first analytical function and the second analytical function comprises a respective decision tree function.

6. A method comprising:

receiving, by a model integration tool executed by a processing device, an electronic data-processing model that generates an analytical output by applying a function to one or more input attributes that are weighted with one or more respective modeling coefficients, wherein the received electronic data-processing model lacks bin range information associated with the one or more modeling coefficients and reason code information associated with the one or more input attributes;

providing, from the model integration tool to a client computing system, a user interface having interface elements for:

specifying the bin range information and the reason code information; and  
uploading, to a computer-readable medium accessible to the model integration tool, an audit file having the expected analytical output and the expected reason code;  
identifying, by the model integration tool and based on user input received via the interface elements, one or more bin ranges for the one or more modeling coefficients and one or more reason codes for the one or more input attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output;

generating, by the model integration tool, modeling code that specifies the electronic data-processing model with the one or more bin ranges and the one or more reason codes;

verifying, with the model integration tool, that a test of the modeling code results in a test analytical output matching an expected analytical output and a test reason code matching an expected reason code,

wherein the verification comprises: accessing the audit file from the computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code;

generating, by the model integration tool and subsequent to the verification, an executable source code that implements the electronic data-processing model on a data-processing platform that is accessible by the client computing system, wherein the executable source code is generated from the modeling code; and

prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code:

performing, with the model integration tool, an additional test of the modeling code;

determining, by the model integration tool, that the additional test results in an

additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code;

outputting, by the model integration tool, an indication that the additional test results in a failed test;

in response to outputting the indication, receiving user input, via the interface elements, specifying one or more updated characteristics of the electronic data-processing model; and

generating, with the model integration tool and subsequent to the additional test resulting in a failed test, updated modeling code including the updated characteristics of the electronic data-processing model,

wherein the updated modeling code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

7. The method of claim 6, wherein the modeling code comprises a specification of the electronic data-processing model in a Predictive Model Markup Language.

8. The method of claim 7, wherein the user interface further comprises one or more additional interface elements for identifying a mapping of the one or more input attributes to one or more production attributes from the data-processing platform and for identifying one or more transformations of one or more respective values of the one or more production attributes into one or more respective values that are compliant with the electronic data-processing model,

wherein the modeling code implements the mapping and the one or more transformations.

9. The method of any of claims 7 to 8, wherein the electronic data-processing model comprises a regression model and wherein the one or more modeling coefficients comprise one or more regression coefficients.

10. The method of any of claims 6 to 9, wherein the user interface further comprises one or more additional interface elements for identifying a first analytical function to be added to the electronic data-processing model, identifying a second analytical function to be added to the electronic data-processing model, and indicating a specific attribute from the one or more attributes whose value controls a selection of the first analytical function or the second analytical function;

wherein the modeling code implements the first analytical function, the second analytical function, and segmentation logic that selects the first analytical function based on the specific attribute having a first value and selects the second analytical function based on the specific attribute having a second value.

11. The method of claim 10, wherein each of the first analytical function and the second analytical function comprises a respective decision tree function.

12. The method of any of claims 7 to 11, wherein the modeling code comprises a specification of the model in a predictive model interchange format and the executable source code comprises source code to be compiled by the data-processing platform,

wherein the user interface further comprises additional interface elements for:

(i) identifying a mapping of the one or more input attributes to one or more production attributes from the data-processing platform,

(ii) identifying one or more transformations of one or more values of the one

or more production attributes into one or more values compliant with the electronic data-processing model,

(i) uploading, to a computer-readable medium accessible to the model integration tool, first PMML code implementing a first decision tree and second PMML code implementing a second decision tree, and

(ii) indicating an input attribute whose value controls a selection of the first decision tree or the second decision tree,

wherein the modeling code implements the mapping, the one or more transformations, the first decision tree, the second decision tree, and segmentation logic that selects the first decision tree based on the input attribute having a first value and selects the second decision tree based on the input attribute having a second value.

13. A computer-readable medium having program code of a model integration tool that is executable by a processing device to cause a computing device to perform operations, the operations comprising:

receiving an electronic data-processing model that generates an analytical output by applying a function to input attributes that are weighted with respective modeling coefficients;

providing, to a client computing system, a user interface having interface elements for:

specifying bin range information associated with the one or more modeling coefficients and reason code information associated with the one or more input attributes; and

identifying, based on input received via the interface elements, bin ranges for the one or more modeling coefficients and reason codes for the one or more input

attributes, wherein each bin range identifies a range of input attribute values for which a respective modeling coefficient is valid, wherein each reason code identifies a respective reason code output indicating a respective impact of a respective weighted input attribute on the analytical output;

generating Predictive Model Markup Language (PMML) code that specifies the electronic data-processing model with the one or more bin ranges and the one or more reason codes;

verifying that a test of the PMML code results in a test analytical output matching an expected analytical output and a test reason code matching an expected reason code; and

generating, subsequent to the verification, an executable source code that implements the electronic data-processing model on a data-processing platform that is accessible by the client computing system, wherein the executable source code is generated from the PMML code,

wherein verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code comprises: accessing the audit file from the computer-readable medium; and comparing the expected analytical output and the expected reason code stored in the audit file with the test analytical output and the test reason code, and

wherein prior to verifying that the test analytical output matches the expected analytical output and the test reason code matches the expected reason code, the processing device performs further operations comprising:

performing an additional test of the PMML code;

determining that the additional test results in a failed test, the failed test comprising an additional test analytical output that fails to match the expected analytical output or an additional test reason code that fails to match the expected reason code;

outputting an indication that the additional test results in a failed test;

in response to outputting the indication, receiving user input, via the user interface, specifying one or more updated characteristics of the electronic data-processing model; and

generating, subsequent to the failed test, updated PMML code including the updated characteristics of the electronic data-processing model,

wherein the updated PMML code results in the test analytical output and the test reason code matching the expected analytical output and the expected reason code.

14. The computer-readable medium of claim 13, wherein the user interface further comprises additional interface elements for identifying a mapping of the one or more input attributes to production attributes from the data-processing platform and for identifying one or more transformations of one or more respective values of the one or more production attributes into one or more values that are compliant with the electronic data-processing model,

wherein the PMML code implements the mapping and the one or more transformations.

15. The computer-readable medium of any of claims 13 to 14, wherein the user interface further comprises one or more additional interface elements for identifying a first analytical function to be added to the electronic data-processing model, identifying a second analytical function to be added to the electronic data-processing model, and indicating a specific attribute from the one or more attributes whose value controls a selection of the first analytical function or the second analytical function;

wherein the PMML code implements the first analytical function, the second analytical function, and segmentation logic that selects the first analytical function based on

the specific attribute having a first value and selects the second analytical function based on the specific attribute having a second value.

16. A computer program that, when executed by a computer, causes the computer to perform the method of any of claims 6 to 12.

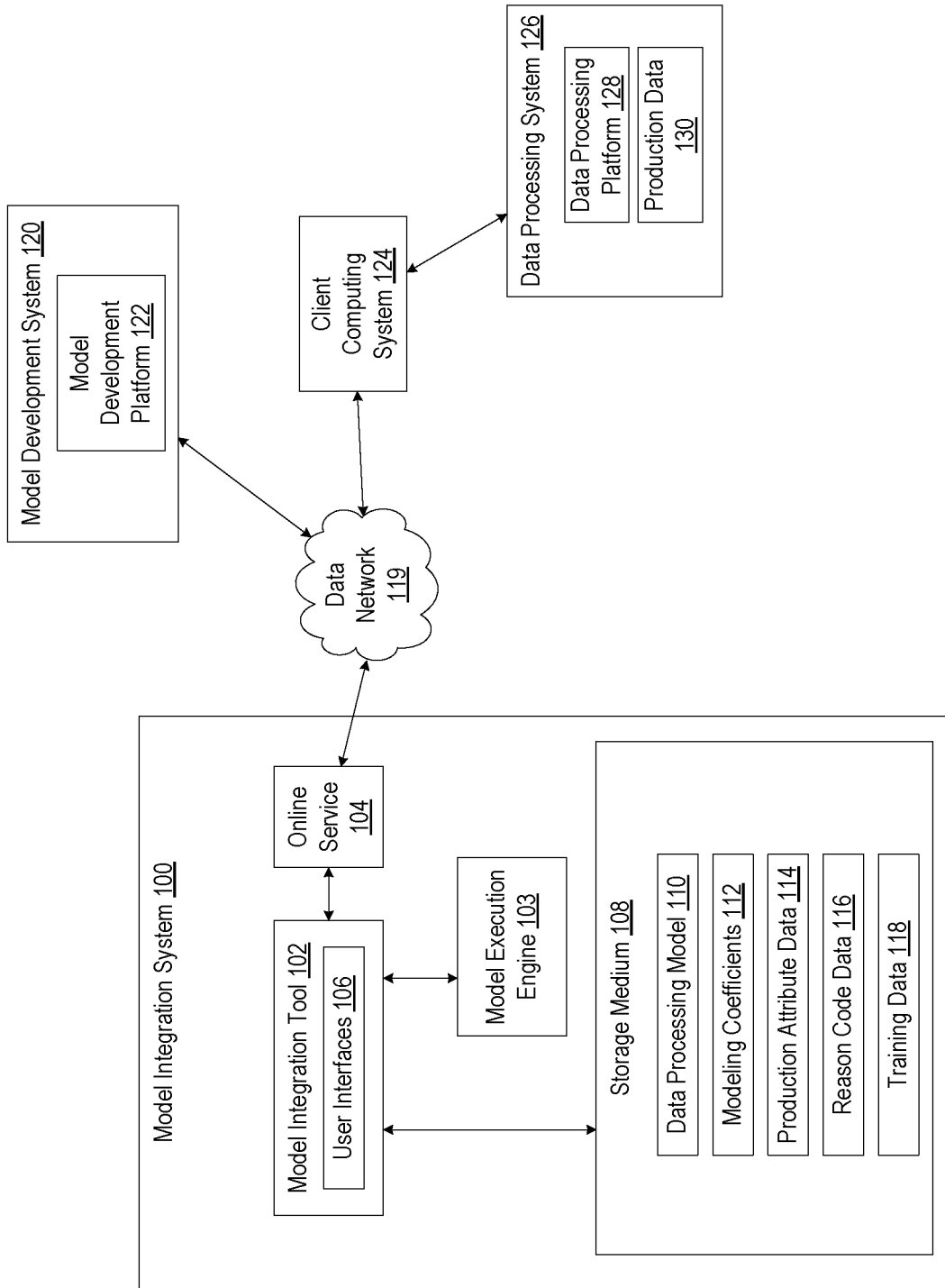


FIG. 1

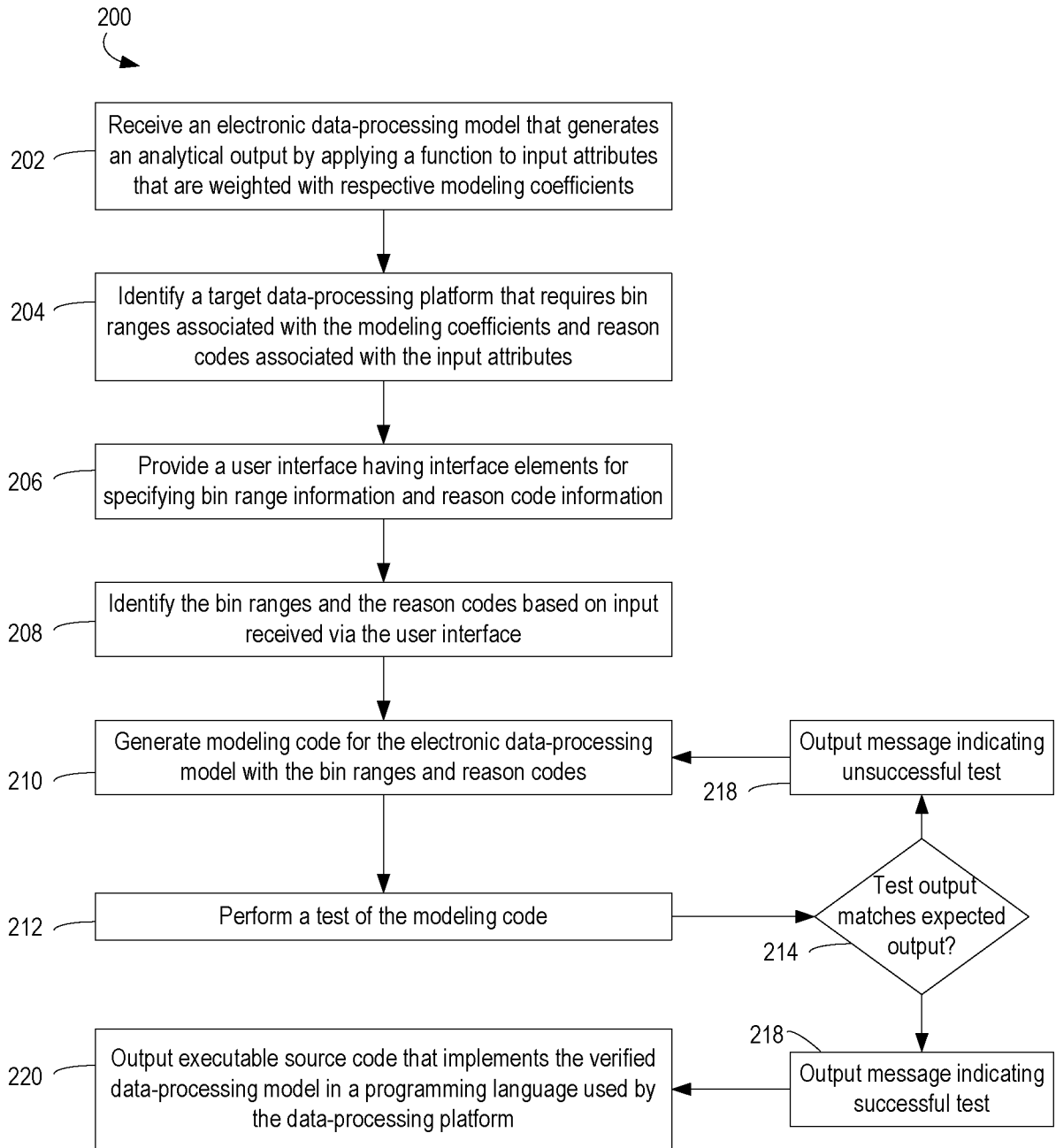


FIG. 2

Select Attribute

IFS Attributes

10 (1 of 36)

Attribute Id	Description
<input type="radio"/> 998	Inquiry Fails eID Standard Quality Checks
<input type="radio"/> 999	System is unavailable
<input type="radio"/> AA0001	Bad SSN in client inquiry
<input type="radio"/> AA0002	Bad Addr in client inquiry
<input type="radio"/> AA0003	Bad Ph in client inquiry
<input type="radio"/> AA0004	Bad Name in client inquiry
<input type="radio"/> AA0005	Bad DOB in client inquiry
<input type="radio"/> AA0006	Stated age of consumer (in years)
<input type="radio"/> AF1001	Address fraud list match on CID
<input type="radio"/> AF1002	Address fraud list match on SSN

302

304

Save

Cancel

300

FIG. 3

400

SAS ATTRIBUTE MAPPING

1 (1 of 1) 10

Attribute Id	Description	SAS Name	Points
		BPFLAG1	1.981160309
		BPFLAG2	1.981160309
		CLU1	-0.013453711

402

404

FIG. 4

500

Attribute Id and Description	Max Points	Bin Range	Points	Bin Type (Original)	Transformation	Reason Code
1001	1.981160309		1.981160309	Yes		R01
			1.981160309	No	$(1/("1001"))^{1.981160309}$ Editor	

502

504

506

508

510

512

[+] Add a New Bin Range

FIG. 5

600

Attribute Id	Decision Tree Attribute Name	Description
Use_of_car	Use_of_car	
A_833IQ12	A_833IQ12	
A_839OLDES	A_839OLDES	
A_1264IQ3M	A_1264IQ3M	
A_1279miq3	A_1279miq3	
A_1477SCHE	A_1477SCHE	
A_1773SUM_	A_1773SUM_	
flag_segment	flag_segment	

602 604

FIG. 6

700

Name:\*  
Decision Tree 01

Predicted Output:\*  
Predicted Spend

Output Feature Type:\*  
Probability

Output Data Type:\*  
Numeric

File Upload:\*  
Choose Upload Cancel

FIG. 7

802		804		806		808																
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Record#	Expected	Expected	Expected	Expected	Expected	301	3796	3807	3812	3813	3903	3002	3909	3905	3100	91	3113	3159	3177	3181	3188	3224
2	363 C91				98	99	99	99	9999	9999	9	1	8	3	0	0	264	34148	1	0	9999996	0
3	308 C91	203	304	109	2	7	9998	7	9998	9998	0	1	8	3	0	0	153	9999998	2	9999996	9999996	0
4	333 C91	203	109		99	8	9998	8	9998	9998	0	1	8	3	0	0	114	18171	0	9999996	9999998	0
5	338 C91	203	109		99	1	9998	1	9998	9998	0	1	8	3	0	0	267	45950	1	0	9999996	0
6	347 C91	304			1	99	9999	99	9999	9999	9	1	8	3	0	0	105	9999998	1	0	9999997	0
7	337 C91	203	109		99	3	9998	3	9998	9998	0	1	8	3	0	0	71	56166	1	9999996	9999996	4
8	338 C91	203	109		99	1	9998	1	9998	9998	0	1	8	3	0	0	306	9999998	5	1445	9999998	0
9	363 C91				98	99	9999	99	9999	9999	9	1	8	3	0	0	264	9999996	3	0	9999996	0
10	338 C91	203	109		99	2	9998	2	9998	9998	0	1	8	3	0	0	9998	9999998	2	9999997	9999998	98
11	325 C92	304	299	36	6	99	9998	99	9998	9998	1	1	8	3	1	0	264	34148	1	0	9999996	0

FIG. 8

Overview Reason Code Score cards Audit: Deploy

Results: 2 Filter:  << 1 >> 10 (1 of 1)

Audit Name	File Name	Baseline	Records Processed	Overall Matched Rate	Matched Score Rate	Matched Reason Code Rate	Run Date	Run Status
Audit2	DemoModeAudit.csv	Work in Progress	10	0	0	20	03/27/2015 03:22:04 PM	Complete
Audit1	DemoModeAudit.csv	Work in Progress	10	0	0	20	03/27/2015 03:18:58 PM	Complete

FIG. 9

```

Model ID: Model_1
Baseline: Baseline1
Organization: Test_org
Audit Name: Audit3
File Name: DemoModelAudit.csv
Record Number: 17

[3010] := null
[3022] := null
[3100] := 0 (INTEGER)

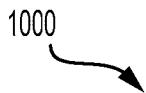
[MiningSchema]
[Mining Field]: Input field [3010] has no value. Returning missing value
replacement [null].
[3010] := null
[Mining Field]: Input field [3022] has no value. Returning missing value
replacement [null].
[3022] := null
[3100] := 0 (INTEGER)

[ModelEnsemble]
The predicate of segment model [1] evaluates [False]
All [OutputField] elements associated with this segment and any encapsulating
segments will be considered [null]

The predicate of segment model [2] evaluates [False]
All [OutputField] elements associated with this segment and any encapsulating
segments will be considered [null]

[Output]
[WinningSegmentId] :=null
The [predictedValue] is [null]
[predicted score] := null
[S1 Reason Code 1] := null
[S1 Reason Code 2] := null
[S1 Reason Code 3] := null
[S1 Reason Code 4] := null
[S2 Reason Code 1] := null
[S2 Reason Code 2] := null
[S2 Reason Code 3] := null
[S2 Reason Code 4] := null
[model score] := null
[Reason Code 1] := null
[Reason Code 2] := null
[Reason Code 3] := null
[Reason Code 4] := null
[uncapped final score] := null
[final score] := null
    
```

FIG. 10



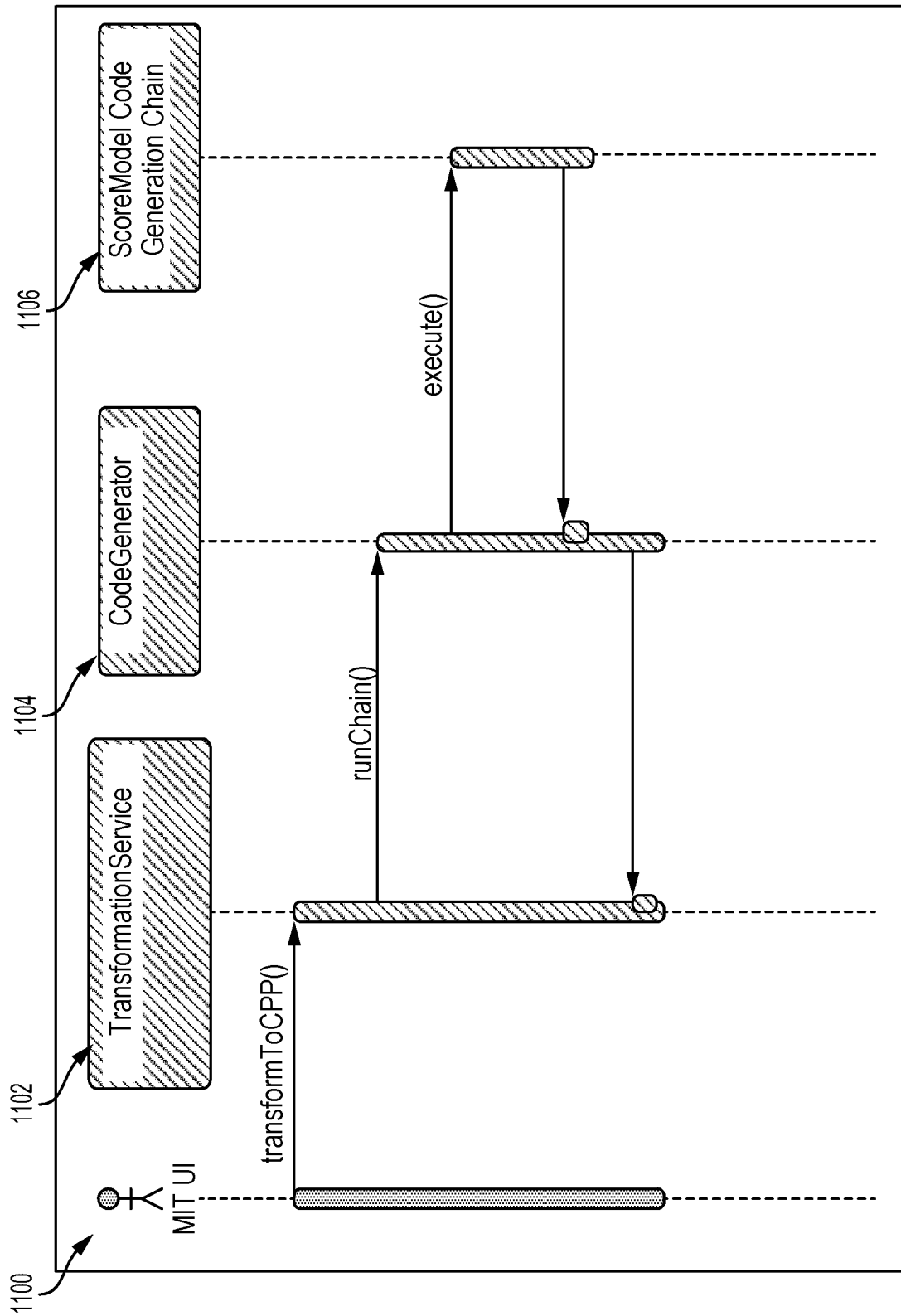


FIG. 11

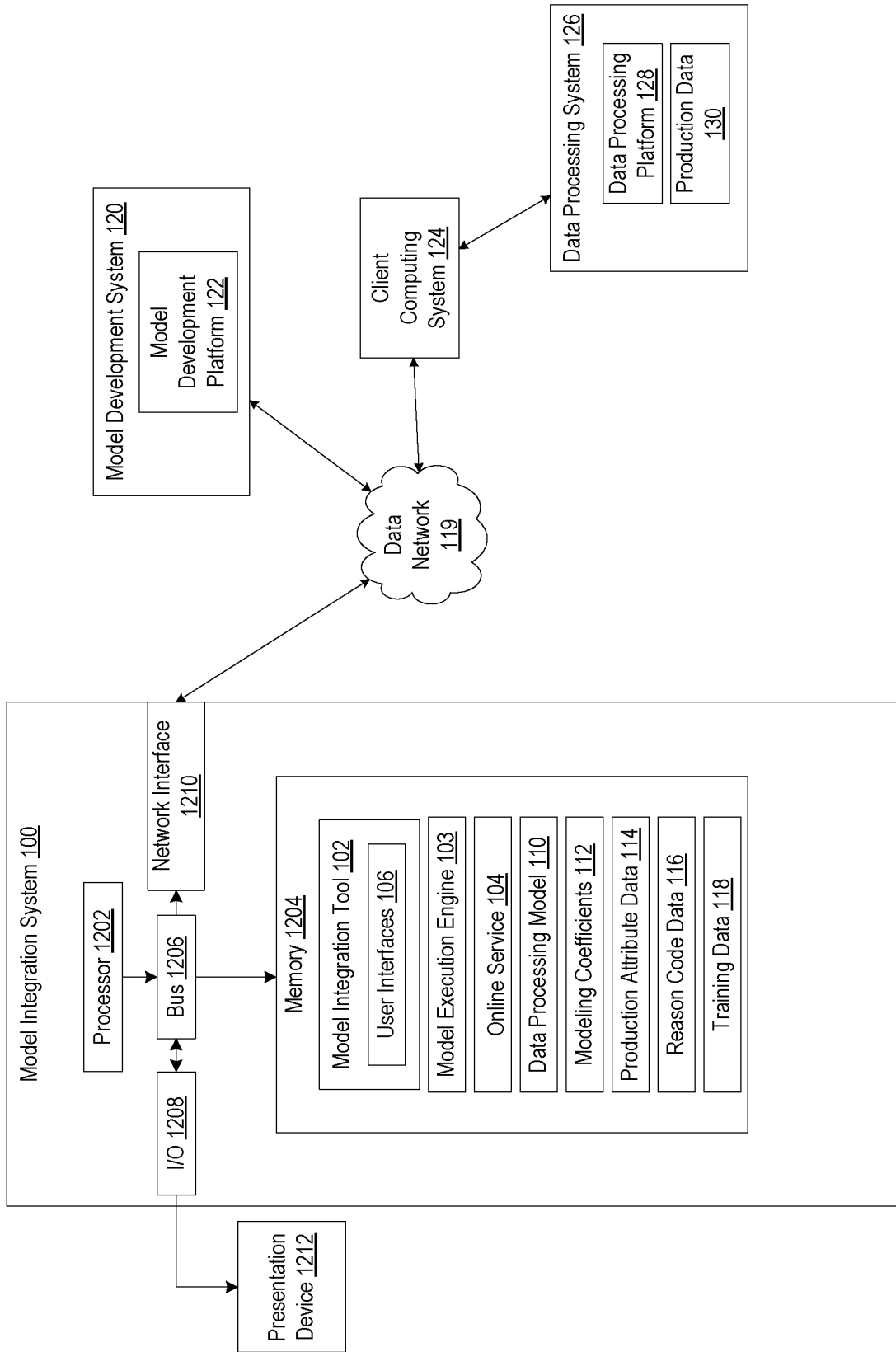


FIG. 12