

METHOD, SYSTEM, AND COMPUTER PROGRAM PRODUCT FOR UNSUPERVISED FEATURE ENGINEERING IN MACHINE LEARNING MODELS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/472,359, filed June 12, 2023, titled “Method, System, and Computer Program Product for Unsupervised Feature Engineering in Machine Learning Models”, the disclosure of which is hereby incorporated by reference in its entirety.

BACKGROUND

1. Technical Field

[0002] This disclosure relates generally to machine learning models and, in non-limiting embodiments or aspects, to methods, systems, and computer program products for performing automatic feature engineering and selection in an unsupervised setting for use in machine learning models.

2. Technical Considerations

[0003] Machine learning models (*e.g.*, change detection models) evaluate features (*e.g.*, one or more data fields including and/or representing raw data and/or values derived from the raw data) to generate predictions (*e.g.*, classifications). Certain machine learning techniques may require initial knowledge about the underlying data to select relevant features that will produce effective predictions. Moreover, certain machine learning techniques may require significant human input for model generation, such as to label data fields, identify which fields are likely to be important for feature engineering, and/or the like. Relying on human input slows generation of a new machine learning model (*e.g.*, increases time spent waiting for human analysis and/or human input) and introduces biases and assumptions about the data that may not be true. Without intelligent, unbiased feature engineering, machine learning models may suffer from excessive false positives (*e.g.*, detecting a change in data from expected behavior where there is none) and/or excessive false negatives (*e.g.*, failing to detect a change in data from expected behavior), which directly correlate with computational inefficiencies in machine learning models and inaccuracies in the overall modeling system.

[0004] Furthermore, it may be difficult to ascertain which features, from all possible features, may be relevant for training machine learning models. Some possible features may be redundant with other features or may not have a meaningful effect on the output of a machine learning model that is to be trained. Overinclusion of features may increase computer resources required to train and execute a trained machine learning model, without improving the overall performance of the machine learning model.

[0005] There is a need in the art for a technical solution to automatically, and in an at least partially unsupervised setting, generate and identify features for inclusion in machine learning models to improve the accuracy of such machine learning models while also reducing the time spent generating and/or deploying such models. There is a further need in the art for a technical solution to automatically, and in an unsupervised setting, identify features for exclusion from machine learning models that might decrease the accuracy of such machine learning models.

SUMMARY

[0006] Accordingly, provided are improved methods, systems, and computer program products for unsupervised feature engineering in machine learning models.

[0007] According to non-limiting embodiments or aspects, provided is a computer-implemented method for unsupervised feature engineering in machine learning models. The method includes receiving, with at least one processor, a dataset including a plurality of records, each record of the plurality of records including a plurality of fields. The method also includes determining, with at least one processor, a data type of a field of the plurality of fields. The method further includes determining, with at least one processor, a distribution of values of the field based on the plurality of records. The method further includes generating, with at least one processor, configuration data based on the data type and the distribution of values of the field. The method further includes generating, with at least one processor, a plurality of time-series features based on the configuration data and the plurality of records. The method further includes determining, with at least one processor, a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof. The method further includes determining, with at least one processor, a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of

the plurality of time-series features. The method further includes determining, with at least one processor, a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0008] In some non-limiting embodiments or aspects, the method may include generating, with at least one processor, at least one higher-level time-series feature based on at least one time-series feature of the plurality of time-series features. The method may also include updating, with at least one processor, the subset of the plurality of time-series features to include the at least one higher-level time-series feature.

[0009] In some non-limiting embodiments or aspects, determining the data type of the field of the plurality of fields may include determining, with at least one processor, the data type of the field of the plurality of fields, wherein the data type includes a categorical data type or a continuous data type. The method may also include processing, with at least one processor, a plurality of transactions between at least one payment device and at least one merchant to produce transaction data. The method may further include storing, with at least one processor, the transaction data in at least one database in a time-series format, wherein the transaction data includes the dataset.

[0010] In some non-limiting embodiments or aspects, determining the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features may include determining that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold, determining that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold, including the first feature in the subset of the plurality of time-series features, and excluding the second feature from the subset of the plurality of time-series features.

[0011] In some non-limiting embodiments or aspects, the at least one threshold may include a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. Determining the subset of the plurality of time-series features may further include determining that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0012] In some non-limiting embodiments or aspects, the method may further include communicating, with at least one processor, the subset of the plurality of time-series features to a computing device of a user. The method may further include receiving, with at least one processor, a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user. The method may further include updating, with at least one processor, the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0013] In some non-limiting embodiments or aspects, generating the plurality of time-series features may further include converting, with at least one processor, the dataset to a time-series format.

[0014] In some non-limiting embodiments or aspects, the method may further include training the machine learning model based on the subset of the plurality of fields.

[0015] In some non-limiting embodiments or aspects, the method may further include receiving, with at least one processor, a second dataset including a second plurality of records. The method may further include generating, with at least one processor, a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records. The method may further include performing, with at least one processor, at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action includes at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0016] In some non-limiting embodiments or aspects, the method may further include determining, with at least one processor, a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model. Determining the relevance score may include determining a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. The relevance score may include a combined

relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features. Determining the subset of the plurality of time-series features based on the relevance score further may include determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0017] In some non-limiting embodiments or aspects, determining the first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model may include determining the first relevance score for each time-series feature of the plurality of time-series features based on the at least one supervised machine learning model, wherein the at least one supervised machine learning model includes at least one gradient boosting model.

[0018] In some non-limiting embodiments or aspects, determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features may include comparing the first relevance score of each time-series feature of the plurality of time-series features to a first threshold, determining that the first relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the first threshold, comparing the second relevance score for each time-series feature of the plurality of time-series features to a second threshold, and determining that the second relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the second threshold.

[0019] In some non-limiting embodiments or aspects, the method may further include generating, with at least one processor, at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of time-series features.

[0020] According to non-limiting embodiments or aspects, provided is a system for unsupervised feature engineering in machine learning models. The system includes at least one processor configured to receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields. The at least one processor is also configured to determine a data type of a field of the plurality of fields. The at least one processor is further configured to determine a distribution of values of the field based on the plurality of records. The at least one processor is further

configured to generate configuration data based on the data type and the distribution of values of the field. The at least one processor is further configured to generate a plurality of time-series features based on the configuration data and the plurality of records. The at least one processor is further configured to determine a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof. The at least one processor is further configured to determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features. The at least one processor is further configured to determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0021] In some non-limiting embodiments or aspects, while determining the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features, the at least one processor may be configured to determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold, determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold, include the first feature in the subset of the plurality of time-series features, and exclude the second feature from the subset of the plurality of time-series features.

[0022] In some non-limiting embodiments or aspects, the at least one processor may be further configured to determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model. While determining the relevance score, the at least one processor may be further configured to determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. The relevance score may include a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features. While determining the subset of the plurality of time-series features based on the relevance score, the at least one processor may be further configured to determine the subset of the plurality of time-series features based on the

first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0023] In some non-limiting embodiments or aspects, the at least one threshold may include a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. While determining the subset of the plurality of time-series features, the at least one processor may be further configured to determine that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0024] In some non-limiting embodiments or aspects, the at least one processor may be further configured to communicate the subset of the plurality of time-series features to a computing device of a user. The at least one processor may be further configured to receive a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user. The at least one processor may be further configured to update the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0025] In some non-limiting embodiments or aspects, the at least one processor may be further configured to train the machine learning model based on the subset of the plurality of fields. The at least one processor may be further configured to receive a second dataset including a second plurality of records. The at least one processor may be further configured to generate a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records. The at least one processor may be further configured to perform at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action includes at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0026] According to non-limiting embodiments or aspects, provided is a computer program product for unsupervised feature engineering in machine learning models. The computer program product includes at least one non-transitory computer-readable medium including program instructions that, when executed by at least one

processor, cause the at least one processor to receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields. The program instructions also cause the at least one processor to determine a data type of a field of the plurality of fields. The program instructions further cause the at least one processor to determine a distribution of values of the field based on the plurality of records. The program instructions further cause the at least one processor to generate configuration data based on the data type and the distribution of values of the field. The program instructions further cause the at least one processor to generate a plurality of time-series features based on the configuration data and the plurality of records. The program instructions further cause the at least one processor to determine a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof. The program instructions further cause the at least one processor to determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features. The program instructions further cause the at least one processor to determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0027] In some non-limiting embodiments or aspects, the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features may cause the at least one processor to determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold, determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold, include the first feature in the subset of the plurality of time-series features, and exclude the second feature from the subset of the plurality of time-series features.

[0028] In some non-limiting embodiments or aspects, the program instructions may further cause the at least one processor to determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model. The program instructions that cause the at least one processor to determine the relevance score may cause the at least one processor to determine a second relevance score for each time-series feature of the plurality of

time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. The relevance score may include a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features. The program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score, cause the at least one processor to determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0029] In some non-limiting embodiments or aspects, the at least one threshold may include a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. The program instructions that cause the at least one processor to determine the subset of the plurality of time-series features may cause the at least one processor to determine that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0030] In some non-limiting embodiments or aspects, the program instructions may further cause the at least one processor to communicate the subset of the plurality of time-series features to a computing device of a user. The program instructions may further cause the at least one processor to receive a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user. The program instructions may further cause the at least one processor to update the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0031] In some non-limiting embodiments or aspects, the program instructions may further cause the at least one processor to train the machine learning model based on the subset of the plurality of fields. The program instructions may further cause the at least one processor to receive a second dataset including a second plurality of records. The program instructions may further cause the at least one processor to generate a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records. The program

instructions may further cause the at least one processor to perform at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action including at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0032] According to some non-limiting embodiments or aspects, provided is a computer-implemented method for combined supervised and unsupervised feature engineering in machine learning models. The method includes receiving, with at least one processor, a dataset including a plurality of records, each record of the plurality of records including a plurality of fields associated with a plurality of features. The method also includes determining, with at least one processor, a first relevance score for each feature of the plurality of features based on at least one first machine learning model. The method further includes determining, with at least one processor, a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric. The method further includes determining, with at least one processor, a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features. The method further includes determining, with at least one processor, a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0033] In some non-limiting embodiments or aspects, determining the first relevance score for each feature of the plurality of features based on the at least one first machine learning model may include determining the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model includes at least one supervised machine learning model including at least one gradient boosting model.

[0034] In some non-limiting embodiments or aspects, determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features may include comparing the first relevance score of each feature of the plurality of features to a first threshold, determining that the first relevance score of each feature of the subset of the plurality of features satisfies the first threshold, comparing the second relevance score for each

feature of the plurality of features to a second threshold, and determining that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0035] In some non-limiting embodiments or aspects, the method may include generating, with at least one processor, at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0036] In some non-limiting embodiments or aspects, generating at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features further may include repeating, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features: determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features and based on the first threshold and the second threshold, and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjusting at least one of the first threshold and the second threshold.

[0037] In some non-limiting embodiments or aspects, the method may include receiving, with at least one processor, at least one of the first threshold and the second threshold from a computing device of a user.

[0038] In some non-limiting embodiments or aspects, the method may include communicating, with at least one processor, the subset of the plurality of features to a computing device of a user. The method may also include receiving, with at least one processor, a selection of a second subset of features from the subset of the plurality of features from the computing device of the user. The method may further include updating, with at least one processor, the subset of the plurality of features to (i) remove the second subset of features from the subset of the plurality of features, or (ii) exclude all other features from the subset of the plurality of features that are not in the second subset of features.

[0039] In some non-limiting embodiments or aspects, the method may include training, with at least one processor, the at least one second machine learning model based on the subset of the plurality of features in a first time period. The method may also include receiving, with at least one processor, transaction data associated with a plurality of transactions completed in a second time period subsequent to the first time

period. The method may further include generating, with at least one processor, a prediction using the at least one second machine learning model and based on the transaction data. The method may further include performing, with at least one processor, at least one remediative action based on the prediction, wherein the at least one remediative action includes at least one of transmitting a message to a computing device, disabling a user account, assigning a label associated with the prediction to at least one transaction of the plurality of transactions, or any combination thereof.

[0040] In some non-limiting embodiments or aspects, determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features may include determining that the first relevance score of a first feature of the plurality of features satisfies at least one threshold, determining that the first relevance score of a second feature of the plurality of features does not satisfy the at least one threshold, including the first feature in the subset of the plurality of features, and excluding the second feature from the subset of the plurality of features.

[0041] In some non-limiting embodiments or aspects, the at least one threshold may include a plurality of thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. Determining the subset of the plurality of features may further include determining that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of thresholds.

[0042] According to non-limiting embodiments or aspects, provided is a system for combined supervised and unsupervised feature engineering in machine learning models. The system includes at least one processor configured to receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields associated with a plurality of features. The at least one processor is also configured to determine a first relevance score for each feature of the plurality of features based on at least one first machine learning model. The at least one processor is further configured to determine a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric. The at least one processor is further configured to determine a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the

subset of the plurality of features. The at least one processor is further configured to determine a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0043] In some non-limiting embodiments or aspects, while determining the first relevance score for each feature of the plurality of features based on at least one first machine learning model, the at least one processor may be configured to determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model includes at least one supervised machine learning model including at least one gradient boosting model.

[0044] In some non-limiting embodiments or aspects, while determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, the at least one processor may be configured to compare the first relevance score of each feature of the plurality of features to a first threshold, determine that the first relevance score for each feature of the subset of the plurality of features satisfies the first threshold, compare the second relevance score for each feature of the plurality of features to a second threshold, and determine that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0045] In some non-limiting embodiments or aspects, the at least one processor may be further configured to generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0046] In some non-limiting embodiments or aspects, while generating at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features, the at least one processor may be further configured to repeat, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features, determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features and based on the first threshold and the second threshold, and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjust at least one of the first threshold and the second threshold.

[0047] According to some non-limiting embodiments or aspects, provided is a computer program product for combined supervised and unsupervised feature engineering in machine learning models. The computer program product includes at least one non-transitory computer-readable medium including program instructions. The program instructions, when executed by at least one processor, cause the at least one processor to receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields associated with a plurality of features. The program instructions also cause the at least one processor to determine a first relevance score for each feature of the plurality of features based on at least one first machine learning model. The program instructions further cause the at least one processor to determine a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric. The program instructions further cause the at least one processor to determine a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features. The program instructions further cause the at least one processor to determine a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0048] In some non-limiting embodiments or aspects, the program instructions that cause the at least one processor to determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model may cause the at least one processor to determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model includes at least one supervised machine learning model including at least one gradient boosting model.

[0049] In some non-limiting embodiments or aspects, the program instructions that cause the at least one processor to determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features may cause the at least one processor to compare the first relevance score of each feature of the plurality of features to a first threshold, determine that the first relevance score for each feature of the subset of the plurality of features satisfies the first threshold, compare the second relevance score for each feature of the plurality of features to a second threshold, and determine that the second

relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0050] In some non-limiting embodiments or aspects, the program instructions may further cause the at least one processor to generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0051] In some non-limiting embodiments or aspects, the program instructions that cause the at least one processor to generate at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features may cause the at least one processor to repeat, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features, determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features and based on the first threshold and the second threshold, and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjust at least one of the first threshold and the second threshold.

[0052] Further non-limiting embodiments or aspects are set forth in the following numbered clauses:

[0053] Clause 1: A computer-implemented method including: receiving, with at least one processor, a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields; determining, with at least one processor, a data type of a field of the plurality of fields; determining, with at least one processor, a distribution of values of the field based on the plurality of records; generating, with at least one processor, configuration data based on the data type and the distribution of values of the field; generating, with at least one processor, a plurality of time-series features based on the configuration data and the plurality of records; determining, with at least one processor, a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof; determining, with at least one processor, a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features; and determining, with at least one

processor, a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0054] Clause 2: The computer-implemented method of clause 1, further comprising generating, with at least one processor, at least one higher-level time-series feature based on at least one time-series feature of the plurality of time-series features; and updating, with at least one processor, the subset of the plurality of time-series features to include the at least one higher-level time-series feature.

[0055] Clause 3: The computer-implemented method of clause 1 or clause 2, wherein determining the data type of the field of the plurality of fields comprises: determining, with at least one processor, the data type of the field of the plurality of fields, wherein the data type comprises a categorical data type or a continuous data type.

[0056] Clause 4: The computer-implemented method of any of clauses 1-3, further comprising: processing, with at least one processor, a plurality of transactions between at least one payment device and at least one merchant to produce transaction data; and storing, with at least one processor, the transaction data in at least one database in a time-series format, wherein the transaction data comprises the dataset.

[0057] Clause 5: The computer-implemented method of any of clauses 1-4, wherein determining the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features comprises: determining that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold; determining that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold; including the first feature in the subset of the plurality of time-series features; and excluding the second feature from the subset of the plurality of time-series features.

[0058] Clause 6: The computer-implemented method of any of clauses 1-5, wherein the at least one threshold comprises a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, and wherein determining the subset of the plurality of time-series features further comprises: determining that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0059] Clause 7: The computer-implemented method of any of clauses 1-6, further comprising: communicating, with at least one processor, the subset of the plurality of time-series features to a computing device of a user; receiving, with at least one processor, a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user; and updating, with at least one processor, the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0060] Clause 8: The computer-implemented method of any of clauses 1-7, wherein generating the plurality of time-series features further comprises converting, with at least one processor, the dataset to a time-series format.

[0061] Clause 9: The computer-implemented method of any of clauses 1-8, further comprising: training the machine learning model based on the subset of the plurality of fields.

[0062] Clause 10: The computer-implemented method of any of clauses 1-9, further comprising: receiving, with at least one processor, a second dataset comprising a second plurality of records; generating, with at least one processor, a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records; and performing, with at least one processor, at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action comprises at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0063] Clause 11: The computer-implemented method of any of clauses 1-10, the method further comprising: determining, with at least one processor, a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model; wherein determining the relevance score further comprises: determining a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score of each time-series feature of the

plurality of time-series features; and wherein determining the subset of the plurality of time-series features based on the relevance score further comprises: determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0064] Clause 12: The computer-implemented method of any of clauses 1-11, wherein determining the first relevance score for each time-series feature of the plurality of time-series features based on the at least one supervised machine learning model comprises: determining the first relevance score for each time-series feature of the plurality of time-series features based on the at least one supervised machine learning model, wherein the at least one supervised machine learning model comprises at least one gradient boosting model.

[0065] Clause 13: The computer-implemented method of any of clauses 1-12, wherein determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features comprises: comparing the first relevance score of each time-series feature of the plurality of time-series features to a first threshold; determining that the first relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the first threshold; comparing the second relevance score for each time-series feature of the plurality of time-series features to a second threshold; and determining that the second relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the second threshold.

[0066] Clause 14: The computer-implemented method of any of clauses 1-13, further comprising generating, with at least one processor, at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of time-series features.

[0067] Clause 15: A system comprising at least one processor configured to: receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields; determine a data type of a field of the plurality of fields; determine a distribution of values of the field based on the plurality of records; generate configuration data based on the data type and the distribution of values of the field; generate a plurality of time-series features based on the configuration data and the plurality of records; determine a relevance score for each time-series feature

of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof; determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features; and determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0068] Clause 16: The system of clause 15, wherein, while determining the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features, the at least one processor is further configured to: determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold; determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold; include the first feature in the subset of the plurality of time-series features; and exclude the second feature from the subset of the plurality of time-series features.

[0069] Clause 17: The system of clause 15 or clause 16, wherein the at least one processor is further configured to determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model; wherein, while determining the relevance score, the at least one processor is further configured to: determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features; and wherein, while determining the subset of the plurality of time-series features based on the relevance score, the at least one processor is further configured to: determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0070] Clause 18: The system of any of clauses 15-17, wherein the at least one threshold comprises a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, and wherein, while determining the subset of the plurality of time-series features, the at least one processor is further configured to: determine that at

least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0071] Clause 19: The system of any of clauses 15-18, wherein the at least one processor is further configured to: communicate the subset of the plurality of time-series features to a computing device of a user; receive a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user; and update the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0072] Clause 20: The system of any of clauses 15-19, wherein the at least one processor is further configured to: train the machine learning model based on the subset of the plurality of fields; receive a second dataset comprising a second plurality of records; generate a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records; and perform at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action comprises at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0073] Clause 21: A computer program product comprising at least one non-transitory computer-readable medium comprising program instructions that, when executed by at least one processor, cause the at least one processor to: receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields; determine a data type of a field of the plurality of fields; determine a distribution of values of the field based on the plurality of records; generate configuration data based on the data type and the distribution of values of the field; generate a plurality of time-series features based on the configuration data and the plurality of records; determine a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof; determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features;

and determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

[0074] Clause 22: The computer program product of clause 21, wherein the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features cause the at least one processor to: determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold; determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold; include the first feature in the subset of the plurality of time-series features; and exclude the second feature from the subset of the plurality of time-series features.

[0075] Clause 23: The computer program product of clause 21 or clause 22, wherein the at least one threshold comprises a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, and wherein the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features cause the at least one processor to: determine that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0076] Clause 24: The computer program product of any of clauses 21-23, wherein the program instructions further cause the at least one processor to: communicate the subset of the plurality of time-series features to a computing device of a user; receive a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user; and update the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0077] Clause 25: The computer program product of any of clauses 21-24, wherein the program instructions further cause the at least one processor to: train the machine learning model based on the subset of the plurality of fields; receive a second dataset comprising a second plurality of records; generate a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records; and perform at least one responsive action based on the

prediction that the change in expected behavior has occurred, wherein the at least one responsive action comprises at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

[0078] Clause 26: The computer program product of any of clauses 21-25, wherein the program instructions further cause the at least one processor to determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model; wherein the program instructions that cause the at least one processor to determine the relevance score cause the at least one processor to: determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features; and wherein the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score cause the at least one processor to: determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

[0079] Clause 27: A computer-implemented method comprising: receiving, with at least one processor, a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields associated with a plurality of features; determining, with at least one processor, a first relevance score for each feature of the plurality of features based on at least one first machine learning model; determining, with at least one processor, a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric; determining, with at least one processor, a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features; and determining, with at least one processor, a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0080] Clause 28: The computer-implemented method of clause 27, wherein determining the first relevance score for each feature of the plurality of features based on the at least one first machine learning model comprises: determining the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model comprises at least one supervised machine learning model comprising at least one gradient boosting model.

[0081] Clause 29: The computer-implemented method of clause 27 or clause 28, wherein determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features comprises: comparing the first relevance score of each feature of the plurality of features to a first threshold; determining that the first relevance score of each feature of the subset of the plurality of features satisfies the first threshold; comparing the second relevance score for each feature of the plurality of features to a second threshold; and determining that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0082] Clause 30: The computer-implemented method of any of clauses 27-29, further comprising generating, with at least one processor, at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0083] Clause 31: The computer-implemented method of any of clauses 27-30, wherein generating at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features further comprises repeating, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features: determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features and based on the first threshold and the second threshold; and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjusting at least one of the first threshold and the second threshold.

[0084] Clause 32: The computer-implemented method of any of clauses 27-31, further comprising receiving, with at least one processor, at least one of the first threshold and the second threshold from a computing device of a user.

[0085] Clause 33: The computer-implemented method of any of clauses 27-32, further comprising: communicating, with at least one processor, the subset of the plurality of features to a computing device of a user; receiving, with at least one processor, a selection of a second subset of features from the subset of the plurality of features from the computing device of the user; and updating, with at least one processor, the subset of the plurality of features to (i) remove the second subset of features from the subset of the plurality of features, or (ii) exclude all other features from the subset of the plurality of features that are not in the second subset of features.

[0086] Clause 34: The computer-implemented method of any of clauses 27-33, further comprising training, with at least one processor, the at least one second machine learning model based on the subset of the plurality of features in a first time period; receiving, with at least one processor, transaction data associated with a plurality of transactions completed in a second time period subsequent to the first time period; generating, with at least one processor, a prediction using the at least one second machine learning model and based on the transaction data; and performing, with at least one processor, at least one remediative action based on the prediction, wherein the at least one remediative action comprises at least one of transmitting a message to a computing device, disabling a user account, assigning a label associated with the prediction to at least one transaction of the plurality of transactions, or any combination thereof.

[0087] Clause 35: The computer-implemented method of any of clauses 27-34, wherein determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features comprises: determining that the first relevance score of a first feature of the plurality of features satisfies at least one threshold; determining that the first relevance score of a second feature of the plurality of features does not satisfy the at least one threshold; including the first feature in the subset of the plurality of features; and excluding the second feature from the subset of the plurality of features.

[0088] Clause 36: The computer-implemented method of any of clauses 27-35, wherein the at least one threshold comprises a plurality of thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, and wherein determining the subset of the plurality of features further comprises: determining that at least two of the variance metric, the

correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of thresholds.

[0089] Clause 37: A system comprising at least one processor configured to: receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields associated with a plurality of features; determine a first relevance score for each feature of the plurality of features based on at least one first machine learning model; determine a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric; determine a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features; and determine a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0090] Clause 38: The system of clause 37, wherein, while determining the first relevance score for each feature of the plurality of features based on at least one first machine learning model, the at least one processor is configured to: determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model comprises at least one supervised machine learning model comprising at least one gradient boosting model.

[0091] Clause 39: The system of clause 37 or clause 38, wherein, while determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, the at least one processor is configured to: compare the first relevance score of each feature of the plurality of features to a first threshold; determine that the first relevance score for each feature of the subset of the plurality of features satisfies the first threshold; compare the second relevance score for each feature of the plurality of features to a second threshold; and determine that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0092] Clause 40: The system of any of clauses 37-39, wherein the at least one processor is further configured to generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0093] Clause 41: The system of any of clauses 37-40, wherein, while generating at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features, the at least one processor is further configured to repeat, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features: determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, and based on the first threshold and the second threshold; and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjust at least one of the first threshold and the second threshold.

[0094] Clause 42: A computer program product comprising at least one non-transitory computer-readable medium comprising program instructions that, when executed by at least one processor, cause the at least one processor to: receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields associated with a plurality of features; determine a first relevance score for each feature of the plurality of features based on at least one first machine learning model; determine a second relevance score for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric; determine a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features; and determine a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features.

[0095] Clause 43: The computer program product of clause 42, wherein the program instructions that cause the at least one processor to determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model cause the at least one processor to: determine the first relevance score for each feature of the plurality of features based on the at least one first machine learning model, wherein the at least one first machine learning model comprises at least one supervised machine learning model comprising at least one gradient boosting model.

[0096] Clause 44: The computer program product of clause 42 or clause 43, wherein the program instructions that cause the at least one processor to determine

the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features cause the at least one processor to: compare the first relevance score of each feature of the plurality of features to a first threshold; determine that the first relevance score for each feature of the subset of the plurality of features satisfies the first threshold; compare the second relevance score for each feature of the plurality of features to a second threshold; and determine that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0097] Clause 45: The computer program product of any of clauses 42-44, wherein the program instructions further cause the at least one processor to generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features.

[0098] Clause 46: The computer program product of any of clauses 42-45, wherein the program instructions that cause the at least one processor to generate at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features cause the at least one processor to repeat, until a number of features in the subset of the plurality of features satisfies the predetermined target number of features: determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, and based on the first threshold and the second threshold; and in response to the number of features in the subset of the plurality of features not satisfying the predetermined target number of features, adjust at least one of the first threshold and the second threshold.

[0099] These and other features and characteristics of the present disclosure, as well as the methods of operation and functions of the related elements of structures and the combination of parts and economies of manufacture, will become more apparent upon consideration of the following description and the appended claims with reference to the accompanying drawings, all of which form a part of this specification, wherein like reference numerals designate corresponding parts in the various figures. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended as a definition of the limits of the disclosed subject matter.

BRIEF DESCRIPTION OF THE DRAWINGS

[0100] Additional advantages and details are explained in greater detail below with reference to the non-limiting, exemplary embodiments that are illustrated in the accompanying schematic figures, in which:

[0101] FIG. 1 is a schematic diagram of a system for supervised and/or unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0102] FIG. 2 is a schematic diagram of example components of one or more devices of FIG. 1, according to some non-limiting embodiments or aspects;

[0103] FIG. 3 is a flow diagram of a method for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0104] FIG. 4A is a schematic diagram of a process for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0105] FIG. 4B a schematic diagram continuing the process of FIG. 4A, for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0106] FIG. 5 is a flow diagram of a method for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0107] FIG. 6 is a flow diagram of a method for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects;

[0108] FIG. 7 is a flow diagram of a method for combined unsupervised and supervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects; and

[0109] FIG. 8 is a flow diagram of a method for combined unsupervised and supervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects.

DETAILED DESCRIPTION

[0110] For purposes of the description hereinafter, the terms “end,” “upper,” “lower,” “right,” “left,” “vertical,” “horizontal,” “top,” “bottom,” “lateral,” “longitudinal,” and derivatives thereof shall relate to the embodiments as they are oriented in the drawing figures. However, it is to be understood that the present disclosure may assume various alternative variations and step sequences, except where expressly specified to the contrary. It is also to be understood that the specific devices and processes

illustrated in the attached drawings, and described in the following specification, are simply exemplary and non-limiting embodiments or aspects of the disclosed subject matter. Hence, specific dimensions and other physical characteristics related to the embodiments or aspects disclosed herein are not to be considered as limiting.

[0111] Some non-limiting embodiments or aspects are described herein in connection with thresholds. As used herein, satisfying a threshold may refer to a value being greater than the threshold, more than the threshold, higher than the threshold, greater than or equal to the threshold, less than the threshold, fewer than the threshold, lower than the threshold, less than or equal to the threshold, equal to the threshold, etc.

[0112] No aspect, component, element, structure, act, step, function, instruction, and/or the like used herein should be construed as critical or essential unless explicitly described as such. Also, as used herein, the articles “a” and “an” are intended to include one or more items and may be used interchangeably with “one or more” and “at least one.” Furthermore, as used herein, the term “set” is intended to include one or more items (*e.g.*, related items, unrelated items, a combination of related and unrelated items, and/or the like) and may be used interchangeably with “one or more” or “at least one.” Where only one item is intended, the term “one” or similar language is used. Also, as used herein, the terms “has,” “have,” “having,” or the like are intended to be open-ended terms. Further, the phrase “based on” is intended to mean “based at least partially on” unless explicitly stated otherwise. In addition, reference to an action being “based on” a condition may refer to the action being “in response to” the condition. For example, the phrases “based on” and “in response to” may, in some non-limiting embodiments or aspects, refer to a condition for automatically triggering an action (*e.g.*, a specific operation of an electronic device, such as a computing device, a processor, and/or the like).

[0113] As used herein, the term “communication” may refer to the reception, receipt, transmission, transfer, provision, and/or the like of data (*e.g.*, information, signals, messages, instructions, commands, and/or the like). For one unit (*e.g.*, a device, a system, a component of a device or system, combinations thereof, and/or the like) to be in communication with another unit means that the one unit is able to directly or indirectly receive information from and/or transmit information to the other unit. This may refer to a direct or indirect connection (*e.g.*, a direct communication connection, an indirect communication connection, and/or the like) that is wired and/or

wireless in nature. Additionally, two units may be in communication with each other even though the information transmitted may be modified, processed, relayed, and/or routed between the first and second unit. For example, a first unit may be in communication with a second unit even though the first unit passively receives information and does not actively transmit information to the second unit. As another example, a first unit may be in communication with a second unit if at least one intermediary unit processes information received from the first unit and communicates the processed information to the second unit. In some non-limiting embodiments or aspects, a message may refer to a network packet (*e.g.*, a data packet and/or the like) that includes data. It will be appreciated that numerous other arrangements are possible.

[0114] As used herein, the term “computing device” may refer to one or more electronic devices configured to process data. A computing device may, in some examples, include the necessary components to receive, process, and output data, such as a processor, a display, a memory, an input device, a network interface, and/or the like. A computing device may be a mobile device. As an example, a mobile device may include a cellular phone (*e.g.*, a smartphone or standard cellular phone), a portable computer, a wearable device (*e.g.*, watches, glasses, lenses, clothing, and/or the like), a personal digital assistant (PDA), and/or other like devices. A computing device may also be a desktop computer or other form of non-mobile computer.

[0115] As used herein, the term “server” may refer to or include one or more computing devices that are operated by or facilitate communication and processing for multiple parties in a network environment, such as the Internet, although it will be appreciated that communication may be facilitated over one or more public or private network environments and that various other arrangements are possible. Further, multiple computing devices (*e.g.*, servers, point-of-sale (POS) devices, mobile devices, etc.) directly or indirectly communicating in the network environment may constitute a “system.”

[0116] As used herein, the term “system” may refer to one or more computing devices or combinations of computing devices (*e.g.*, processors, servers, client devices, software applications, components of such, and/or the like). Reference to “a device,” “a server,” “a processor,” and/or the like, as used herein, may refer to a previously-recited device, server, or processor that is recited as performing a previous step or function, a different device, server, or processor, and/or a combination of

devices, servers, and/or processors. For example, as used in the specification and the claims, a first device, a first server, or a first processor that is recited as performing a first step or a first function may refer to the same or different device, server, or processor recited as performing a second step or a second function.

[0117] As used herein, the term “acquirer institution” may refer to an entity licensed and/or approved by a transaction service provider to originate transactions (*e.g.*, payment transactions) using a payment device associated with the transaction service provider. The transactions the acquirer institution may originate may include payment transactions (*e.g.*, purchases, original credit transactions (OCTs), account funding transactions (AFTs), and/or the like). In some non-limiting embodiments or aspects, an acquirer institution may be a financial institution, such as a bank. As used herein, the term “acquirer system” may refer to one or more computing devices operated by or on behalf of an acquirer institution, such as a server computer executing one or more software applications.

[0118] As used herein, the term “account identifier” may include one or more primary account numbers (PANs), tokens, or other identifiers associated with a customer account. The term “token” may refer to an identifier that is used as a substitute or replacement identifier for an original account identifier, such as a PAN. Account identifiers may be alphanumeric or any combination of characters and/or symbols. Tokens may be associated with a PAN or other original account identifier in one or more data structures (*e.g.*, one or more databases, and/or the like) such that they may be used to conduct a transaction without directly using the original account identifier. In some examples, an original account identifier, such as a PAN, may be associated with a plurality of tokens for different individuals or purposes.

[0119] As used herein, the terms “client” and “client device” may refer to one or more client-side devices or systems (*e.g.*, remote from a transaction service provider) used to initiate or facilitate a transaction (*e.g.*, a payment transaction). As an example, a “client device” may refer to one or more POS devices used by a merchant, one or more acquirer host computers used by an acquirer, one or more mobile devices used by a user, one or more computing devices used by a payment device provider system, and/or the like. In some non-limiting embodiments or aspects, a client device may be an electronic device configured to communicate with one or more networks and initiate or facilitate transactions. For example, a client device may include one or more computers, portable computers, laptop computers, tablet computers, mobile devices,

cellular phones, wearable devices (*e.g.*, watches, glasses, lenses, clothing, and/or the like), PDAs, and/or the like. Moreover, a “client” may also refer to an entity (*e.g.*, a merchant, an acquirer, and/or the like) that owns, utilizes, and/or operates a client device for initiating transactions (*e.g.*, for initiating transactions with a transaction service provider).

[0120] As used herein, the terms “electronic wallet” and “electronic wallet application” refer to one or more electronic devices and/or software applications configured to initiate and/or conduct payment transactions. For example, an electronic wallet may include a mobile device executing an electronic wallet application, and may further include server-side software and/or databases for maintaining and providing transaction data to the mobile device. An “electronic wallet provider” may include an entity that provides and/or maintains an electronic wallet for a customer, such as Google Pay®, Android Pay®, Apple Pay®, Samsung Pay®, and/or other like electronic payment systems. In some non-limiting examples, an issuer bank may be an electronic wallet provider.

[0121] As used herein, the term “issuer institution” may refer to one or more entities, such as a bank, that provide accounts to customers for conducting transactions (*e.g.*, payment transactions), such as initiating credit and/or debit payments. For example, an issuer institution may provide an account identifier, such as a PAN, to a customer that uniquely identifies one or more accounts associated with that customer. The account identifier may be embodied on a portable financial device, such as a physical financial instrument, *e.g.*, a payment card, and/or may be electronic and used for electronic payments. The term “issuer system” refers to one or more computer devices operated by or on behalf of an issuer institution, such as a server computer executing one or more software applications. For example, an issuer system may include one or more authorization servers for authorizing a transaction.

[0122] As used herein, the term “merchant” may refer to an individual or entity that provides goods and/or services, or access to goods and/or services, to customers based on a transaction, such as a payment transaction. The term “merchant” or “merchant system” may also refer to one or more computer systems operated by or on behalf of a merchant, such as a server computer executing one or more software applications.

[0123] As used herein, a “point-of-sale (POS) device” may refer to one or more devices, which may be used by a merchant to conduct a transaction (*e.g.*, a payment

transaction) and/or process a transaction. For example, a POS device may include one or more client devices. Additionally or alternatively, a POS device may include peripheral devices, card readers, scanning devices (*e.g.*, code scanners), Bluetooth® communication receivers, near-field communication (NFC) receivers, radio frequency identification (RFID) receivers, and/or other contactless transceivers or receivers, contact-based receivers, payment terminals, and/or the like. As used herein, a “point-of-sale (POS) system” may refer to one or more client devices and/or peripheral devices used by a merchant to conduct a transaction. For example, a POS system may include one or more POS devices and/or other like devices that may be used to conduct a payment transaction. In some non-limiting embodiments or aspects, a POS system (*e.g.*, a merchant POS system) may include one or more server computers configured to process online payment transactions through webpages, mobile applications, and/or the like.

[0124] As used herein, the term “payment device” may refer to an electronic payment device, a portable financial device, a payment card (*e.g.*, a credit or debit card), a gift card, a smartcard, smart media, a payroll card, a healthcare card, a wristband, a machine-readable medium containing account information, a keychain device or fob, an RFID transponder, a retailer discount or loyalty card, a cellular phone, an electronic wallet mobile application, a PDA, a pager, a security card, a computing device, an access card, a wireless terminal, a transponder, and/or the like. In some non-limiting embodiments or aspects, the payment device may include volatile or non-volatile memory to store information (*e.g.*, an account identifier, a name of the account holder, and/or the like).

[0125] As used herein, the term “transaction service provider” may refer to an entity that receives transaction authorization requests from merchants or other entities and provides guarantees of payment, in some cases through an agreement between the transaction service provider and an issuer institution. For example, a transaction service provider may include a payment network such as Visa® or any other entity that processes transactions. The term “transaction processing system” may refer to one or more computer systems operated by or on behalf of a transaction service provider, such as a transaction processing server executing one or more software applications. A transaction processing server may include one or more processors and, in some non-limiting embodiments or aspects, may be operated by or on behalf of a transaction service provider.

[0126] The systems, methods, and computer program products described herein provide numerous technical advantages in systems for machine learning models, including models that detect changes in expected data behavior. First, by generating configuration data based on the data types and distribution of values of fields in a dataset, the disclosed systems may adapt to the nature of the fields being used for change detection and may automatically determine how to aggregate fields for feature engineering. Such techniques reduce computation time, time to generate and deploy the models, reduce data preprocessing requirements (*e.g.*, pre-labeling), and are more accurate, given the adaption to the nature of the field itself (*e.g.*, a field's data type and distribution of values). Second, by generating a plurality of time-series features based on the generated configuration data, the disclosed systems may automatically identify and prepare features for evaluation and possible inclusion in a machine learning model. Such techniques reduce time requirements from data input to training completion, reduce data preprocessing requirements, and are more accurate, given that the feature table is adapted to the evaluated field. Third, by determining one or more scores for each feature and comparing the scores to thresholds (which may determine inclusion or exclusion in the underlying model), disclosed systems may readily determine which features are sufficiently meaningful for inclusion in the underlying machine learning model that will be generated. Such techniques improve the accuracy of the machine learning model, since the use of one or more statistical metrics to evaluate the generated features will provide information of the effectiveness of the feature being able to predict changes in expected data behavior.

[0127] The systems, methods, and computer program products described herein further provide numerous technical advantages in systems for training and executing machine learning models. Sets of features that are candidates for use in training machine learning models may be reduced in size to include only features that are likely to affect the output of the machine learning models to be trained. By using both unsupervised and supervised scoring models, features may be evaluated for relevance and inclusion in training machine learning models. Identifying a subset of features from a plurality of possible features, as described herein, allows the model training systems to use fewer computer resources and take less time to produce trained machine learning models that are more reliable in their output, because the subset of features includes the most relevant features.

[0128] In some non-limiting embodiments or aspects, described techniques and systems do not require (or require relatively less, *e.g.*, compared to other techniques) pre-labeling of fields or features by manual user input and may be completed without human intervention. Not only does removing human labeling reduce overall time required to produce a trained machine learning model, but it removes human bias from the feature engineering process. Moreover, the disclosed techniques provide a generalizable, scalable approach that can be used for a multitude of change detection tasks.

[0129] In some non-limiting embodiments or aspects, duplicated or non-informative features may influence performance of machine learning models by producing false positives (*e.g.*, generating anomaly alerts) where changes to expected behavior do not exist, or by producing false negatives (*e.g.*, failing to generate an anomaly alert) where changes to expected behavior do exist. For example, individual records may appear non-anomalous without context. Consider the illustrative Table 1 (below), which shows an individual record where each field is given a change score between 0 and 10, where 0 indicates no likelihood of change, and 10 indicates a high likelihood of change. Most of the fields have low change scores (*e.g.*, merchant name, weather, temperature, mode, and utility fee are all have low values for change score), but one field has a high change score (*e.g.*, the change score of transaction count is high, but its anomalous effect is mitigated by the other fields).

Table 1

Trans. Count	Merchant Name	Weather	Temperature	Mode	Utility Fee
10	0	1	0	2	0

[0130] In the above example, the average change score value for the fields of the record are 2.2, which may not stand out in a change detection model that may be looking for an change score satisfying a higher threshold (*e.g.*, an aggregate change score of 5 or higher). This is because non-informative values like merchant name, weather, temperature, mode, and utility fee may be adversely affecting the accuracy of the model. To compare, consider where the change detection model is configured to use features that have been determined (*e.g.*, by modeling system 102) to be informative. Consider the second illustrative Table 2 (below), in which feature

engineering has produced a change detection model that considers only relevant features. The exemplary record below may be produced by the techniques described herein, to identify relevant features (*e.g.*, transaction count, transaction amount, declined transaction count, declined transaction amount, etc.) for change detection.

Table 2

Trans. Count	Trans. Amount 1	Trans. Amount 2	Trans. Amount 3	Trans. Amount 4	Declined Trans. Count	Declined Trans. Amount
0	10	10	10	10	0	10

[0131] In the above example, the average change score for the fields of the record is 7.1, which may satisfy (*e.g.*, meet, exceed, fall below, etc.) a threshold value (*e.g.*, a predetermined threshold, a dynamic threshold, etc.) of the change detection model and may appropriately identify the transaction records that have the above features as being anomalous. By conducting unsupervised feature engineering, machine learning models have improved accuracy in detecting changes to expected behavior, which may further improve response processes that rely on accurate change detection.

[0132] Referring now to FIG. 1, FIG. 1 is a diagram of an example *system 100* in which devices, systems, and/or methods, described herein, may be implemented. As shown in FIG. 1, *system 100* may include modeling system 102, memory 104, computing device 106, and/or communication network 108. Modeling system 102, memory 104, and computing device 106 may interconnect (*e.g.*, establish a connection to communicate) via wired connections, wireless connections, or a combination of wired and wireless connections, which may be separate from or include communication network 108 (or a portion thereof).

[0133] Modeling system 102 may include one or more computing devices configured to communicate with memory 104 and/or computing device 106 (*e.g.*, over communication network 108 and/or the like). Modeling system 102 may be configured to receive data for feature engineering, perform unsupervised feature engineering, use features produced by unsupervised feature engineering to train a machine learning model, execute the trained machine learning model, and perform one or more responsive actions based on a prediction generated by the trained machine learning model. Modeling system 102 may include or be in communication with memory 104.

Modeling system 102 may be associated with, or included in a same system as, a natural language processing system, a fraud detection system, an advertising system, and/or a transaction processing system. In general, modeling system 102 may carry out unsupervised feature engineering for use in generating machine learning models according to the following processes: (i) a first process for profiling raw data to determine underlying data distributions; (ii) a second process for automatically creating features using the data distributions learned in the first process; and (iii) a third process for selecting a representative set of features for inclusion in training a machine learning model, and eliminating other features by excluding those features from inclusion in training the machine learning model, to improve machine learning model performance on future data records.

[0134] Memory 104 may include one or more computing devices configured to communicate with modeling system 102 and/or computing device 106 at least partly over communication network 108. Memory 104 may be configured to store raw data (*e.g.*, transaction data) in one or more non-transitory computer readable storage media. The raw data stored by memory 104 may be used by modeling system 102 to perform unsupervised feature engineering. Memory 104 may further store one or more features determined by modeling system 102 to be relevant to machine learning model training. Memory 104 may further store parameters of one or more machine learning models that have been trained. Memory 104 may communicate with and/or be included in modeling system 102.

[0135] Computing device 106 may include one or more processors that are configured to communicate with modeling system 102 and/or memory 104 at least partly over communication network 108. Computing device 106 may be associated with a user and may include at least one user interface for transmitting data to and receiving data from modeling system 102 and/or memory 104. For example, computing device 106 may show, on a display of computing device 106, one or more outputs of an unsupervised feature engineering process and/or a trained machine learning model executed by modeling system 102. By way of further example, one or more inputs for the unsupervised feature engineering process and/or the trained machine learning model may be determined or received by modeling system 102 via a user interface of computing device 106. Computing device 106 may have an input component for a user to provide input (*e.g.*, text, a selection, an approval, etc.) to the unsupervised feature engineering process and/or the trained machine learning model.

[0136] Communication network 108 may include one or more wired and/or wireless networks over which the systems and devices of *system 100* may communicate. For example, communication network 108 may include a cellular network (*e.g.*, a long-term evolution (LTE®) network, a third generation (3G) network, a fourth generation (4G) network, a fifth generation (5G) network, a code division multiple access (CDMA) network, etc.), a public land mobile network (PLMN), a local area network (LAN), a wide area network (WAN), a metropolitan area network (MAN), a telephone network (*e.g.*, the public switched telephone network (PSTN)), a private network, an ad hoc network, an intranet, the Internet, a fiber optic-based network, a cloud computing network, and/or the like, and/or a combination of these or other types of networks.

[0137] In some non-limiting embodiments or aspects, modeling system 102 may be configured to perform one or more steps of a process for unsupervised feature engineering for use in machine learning models. For example, modeling system 102 may receive (*e.g.*, from memory 104, from transaction processing system, etc.) a dataset (*e.g.*, one or more tables of data, which may be stored in time-series format or converted to time-series format by modeling system 102) including a plurality of records (*e.g.*, rows of data in a table), wherein each record of the plurality of records includes a plurality of fields (*e.g.*, cells of data in a table). Modeling system 102 may determine a data type (*e.g.*, bitwise; continuous, such as numeric, integer, float, real; chronological, such as date, time, timestamp, year; informational, such as character, text, variable-character; categorical, such as labels, classes, groups, etc.; Unicode; string; XML; and/or the like) of a field of the plurality of fields. In some non-limiting embodiments or aspects, modeling system 102 may determine that the data type includes a categorical data type or a continuous data type. Modeling system 102 may further determine a distribution of values of the field based on the plurality of records. The distribution of values may include a set of all values appearing in a record of the plurality of records for the column of a given field. By way of further example, modeling system 102 may identify a distribution of bitwise values for a bitwise data type of the field, based on the actual bitwise values for the column of the field in the dataset. In such an example, suppose the field of a record is associated with “Response Code”, which may have eight different bitwise values across all records in the dataset; modeling system 102 may determine a distribution of values as follows: 000, 001, 010, 011, 100, 101, 110, and 111.

[0138] In some non-limiting embodiments or aspects, modeling system 102 may generate configuration data (*e.g.*, a JSON data file including structured arguments defining parameters for performing an aggregation process) based on the data type and the distribution of values of the field. For example, the configuration data may specify a column name for the field, a data type for the field, one or more aggregation filters (*e.g.*, whether to include or exclude certain data), one or more of values from the distribution of values, an aggregation rule (*e.g.*, calculate a count of values), one or more partition rules (*e.g.*, one or more parameters for grouping the data of the field), and an aggregated field name for use in a plurality of time-series features to be generated from the configuration data. Modeling system 102 may input the configuration data to a feature creation module of modeling system 102 (*e.g.*, a subprocess for generating a plurality of time-series features).

[0139] In some non-limiting embodiments or aspects, modeling system 102 may generate a plurality of time-series features (*e.g.*, a feature one or more fields, organized by time-interval, that may be aggregated, such as “count of transactions where ECI Code equals 05”) based on the configuration data and the plurality of records. When generating the plurality of time-series features, modeling system 102 may convert the dataset to a time-series format. Alternatively, the dataset may be received by modeling system 102 in a time-series format. The configuration data may define the form of aggregation of the plurality of records, the field, the field’s values, the field’s distribution of values, and the format of the plurality of time-series features, as described above. See FIG. 4B for further non-limiting embodiments or aspects of a method for generating a plurality of time-series features 409.

[0140] In some non-limiting embodiments or aspects, modeling system 102 may determine a relevance score (*e.g.*, a score indicating the likelihood of a feature being correlated with a modeled event, such as a change in data behavior from an expected behavior) for each time-series feature of the plurality of time-series features based on at least one metric. The relevance score may be quantitative (*e.g.*, a normalized value between 0 and 1, where 0 is not relevant and 1 is relevant) and/or qualitative (*e.g.*, an indication of how many evaluated metrics support a finding that the feature is relevant, a categorical assessment of the prediction, etc.). Metrics used to determine the relevance score may include, but are not limited to, a variance metric (*e.g.*, a measurement expressing the expectation of the squared deviation of a random variable from its mean), a correlation metric (*e.g.*, a measurement expressing the

extent that two variables are linearly related), an entropy metric (*e.g.*, a measurement of randomness in a system), a statistical importance metric (*e.g.*, simple or complex importance, which may include a measurement indicating a likelihood that a relationship between two variables is caused by something other than chance), or any combination thereof. Modeling system 102 may generate a score for each metric (*e.g.*, a subscore for the relevance score) and combine the scores of two or more metrics to determine the relevance score, in a form of ensemble analysis. For example, for a value-based relevance score, a subscore for the variance metric may be 0.5, and the subscore for the correlation metric may be 1.0, for which an ensemble relevance score may be 0.75. By way of further example, for a binary-based (*e.g.*, flag-based) relevance score, subscores for the variance metric, correlation metric, and entropy metric may all indicate “relevant”, while the subscore for the statistical importance metric may indicate “not relevant”, in which case the relevance score may be set to “relevant” if a threshold number of subscores indicate “relevant” (*e.g.*, two or more, three or more, majority, etc.).

[0141] In some non-limiting embodiments or aspects, modeling system 102 may determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features. While determining the subset of the plurality of time-series features, modeling system 102 may determine that the relevance score of a first feature of the plurality of features satisfies (*e.g.*, meets, exceeds, or falls below) at least one threshold (*e.g.*, a predetermined or dynamic threshold value for relevance, such as a value that is indicative of a feature being more likely relevant than not, over >0.5, etc.). The at least one threshold may include a plurality of different thresholds, each corresponding to a metric (*e.g.*, a variance metric, a correlation metric, an entropy metric, a statistical metric) that the relevance score is based on. In response to modeling system 102 determining that the relevance score of the first feature satisfies (*e.g.*, meets, exceeds, falls below, etc.) at least one threshold, modeling system 102 may include the first feature in the subset of the plurality of time-series features. Further, while determining the subset of the plurality of time-series features, modeling system 102 may determine that the relevance score of a second feature of the plurality of features does not satisfy (*e.g.*, does not meet, exceed, fall below, etc.) the at least one threshold. In response to modeling system 102 determining that the relevance score of a second feature of the plurality of features does not satisfy the at least one threshold, modeling system

102 may exclude the second feature from the subset of the plurality of time-series features. When determining the subset of the plurality of time-series features, modeling system 102 may compare the subscore for each metric of the relevance score to a corresponding threshold of the at least one threshold (*e.g.*, and include features that satisfy a minimum number of thresholds). After determining the subset of the plurality of time-series features, modeling system 102 may determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features (*e.g.*, by decoding the subset of the plurality of time-series features, such as shown in process 503 of FIG. 5).

[0142] In some non-limiting embodiments or aspects, modeling system 102 may determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model (*e.g.*, a light gradient boosting machine learning model, a regularized distributed gradient boosting machine learning model, etc.). In some non-limiting embodiments or aspects, the first relevance score may be a combined score (*e.g.*, a mean, a median, a mode, a weighted average, and/or the like). Additionally or alternatively, one or more gradient boosting models may be used to generate a categorical score (*e.g.*, 1 or 0) indicating whether a feature is relevant for inclusion (*e.g.*, a score of 1) or not (*e.g.*, a score of 0), and a mathematic mode of all gradient boosting model scores may be used as the first relevance score (*e.g.*, if most scores are 1, then the first relevance score of the supervised models is 1). In some non-limiting embodiments or aspects, modeling system 102 may determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric. For example, a relevance subscore (*e.g.*, a categorical score of 1, indicating the feature is relevant, or a score of 0, indicating the feature is not relevant) may be produced for each metric for the feature. The second relevance score may then be computed based on the subscore for each metric (*e.g.*, a mode or majority of scores being 1 would indicate that a feature is relevant). In some non-limiting embodiments or aspects, the relevance score may include a combined relevance score based on the first relevance score and the second relevance score. In some non-limiting embodiments or aspects, modeling system 102 may determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time series feature of the subset of the plurality of time series features. For example, modeling

system 102 may include a feature in the subset of the plurality of time-series features if the first relevance score indicates the feature is relevant (*e.g.*, a score of 1), and the second relevance score also indicates the features is relevant (*e.g.*, a score of 1).

[0143] In some non-limiting embodiments or aspects, the at least one supervised machine learning model may include one or more gradient boosting models (*e.g.*, Light Gradient Boosting Machine (GBM), XGBoost, etc.). When modeling system 102 determines the subset of the plurality of time-series features based on the first relevance score and the second relevance score, modeling system 102 may execute a threshold comparison process. The threshold comparison process may include comparing, by modeling system 102, the first relevance score of each time-series feature of the plurality of time-series features to a first threshold (*e.g.*, a categorical threshold such as 1 or 0, a numerical threshold, such as 0.00 to 1.00, etc.). The threshold comparison process may further include determining, by modeling system 102, that the first relevance score for each time-series feature of the subset of the plurality of time-series features satisfies (*e.g.*, meets or exceeds) the first threshold. The threshold comparison process may further include comparing, by modeling system 102, the second relevance score for each time-series feature of the plurality of time-series features to a second threshold (*e.g.*, a categorical threshold such as 1 or 0, a numerical threshold, such as 0.00 to 1.00, etc.). The threshold comparison process may further include determining, by modeling system 102, that the second relevance score for each time-series feature of the subset of the plurality of time-series features satisfies (*e.g.*, meets or exceeds) the second threshold. Modeling system 102 may generate and/or determine the first threshold and/or the second threshold based on a predetermined target number of features (*e.g.*, 80 to 100 features) in the subset of the plurality of time-series features. In some non-limiting embodiments or aspects, the predetermined target number of features may include a range of a number of features (*e.g.*, between 80 and 100 features). In this manner, the thresholds may be set such that the topmost number of features that satisfy the thresholds are equal to or in the range of the target number of features.

[0144] In some non-limiting embodiments or aspects, modeling system 102 may further train the machine learning model based on the subset of the plurality of fields, and implement the machine learning model in a live environment (*e.g.*, such as a change detection model in a change detection and response system). While training the machine learning model, modeling system 102 may input data from a training

dataset based on the subset of the plurality of fields to the machine learning model (*e.g.*, inputting training field data associated with the subset of the plurality of fields into a forward-propagating input layer). Modeling system 102 may further determine a loss based on a loss function and an output from at least one output layer of the machine learning model. Modeling system 102 may feed the output and the loss into the next forward-propagating input layer of machine learning model until the loss converges. Additionally or alternatively, modeling system 102 may back-propagate the loss (*e.g.*, use gradients based on the loss) to update the parameters of the machine learning model. After training (*e.g.*, convergence of the loss), the parameters of the machine learning model may be used for the machine learning model in a live environment (*e.g.*, a production environment, such as in a change detection and response system).

[0145] In some non-limiting embodiments or aspects, modeling system 102 may be included in or associated with a transaction processing system. In such a case, modeling system 102 and/or transaction processing system may process a plurality of transactions between at least one payment device and at least one merchant over a plurality of time intervals (*e.g.*, days, weeks, months) to produce transaction data (*e.g.*, records of transactions including data such as, but not limited to, transaction amount, transaction time, transaction description, merchant identifier, payment device identifier, merchant category code, and/or the like). Modeling system 102 may store the transaction data in at least one database (*e.g.*, memory 104) in a time-series format (*e.g.*, wherein records of transactions are labeled and/or ordered according to time interval), and the transaction data may include the dataset for training the machine learning model.

[0146] In some non-limiting embodiments or aspects, and before determining the subset of the plurality of fields for training a machine learning model, modeling system 102 may communicate the subset of the plurality of time-series features to computing device 106 of a user. The subset of the plurality of time-series features may be included in a message that is configured to cause a display of computing device 106 of the user to display the subset of the plurality of time-series features in a user interface. The user of computing device 106 may use one or more input components of computing device 106 to provide a selection of a second subset of features from the subset of the plurality of time-series features. Modeling system 102 may receive the selection of the second subset of features from the subset of the plurality of time-series

features from computing device 106 of the user. Modeling system 102, in response to receiving the selection, may update the subset of the plurality of time-series features to either remove the second subset of features from the subset of the plurality of time-series features, or exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features. Modeling system 102 may, additionally or alternatively, remove some features of the second subset of features from the subset of the plurality of time-series features based on the selection, and include some other features of the second subset of features in the subset of the plurality of time-series features based on the selection. After refining the subset of the plurality of time-series features based on the selection, modeling system 102 may determine the subset of the plurality of fields for training a machine learning model.

[0147] In some non-limiting embodiments or aspects, modeling system 102 may generate at least one higher-level time-series feature (*e.g.*, a feature based on the aggregation of two or more fields) based on at least one time-series feature of the plurality of time-series features. Modeling system 102 may further update the subset of the plurality of time-series features to include the at least one higher-level time-series feature.

[0148] In some non-limiting embodiments or aspects, and after determining the subset of the plurality of fields for training a machine learning model, modeling system 102 may train the machine learning model based on the subset of the plurality of fields. Modeling system 102 may then receive a second dataset including a second plurality of records. Modeling system 102 may generate a prediction (*e.g.*, a prediction that a change in expected behavior has occurred) using the trained machine learning model (*e.g.*, a trained change detection model), and based on at least one feature of the subset of the plurality of time-series features and at least a portion of data (*e.g.*, input to the change detection model) from the second dataset. Modeling system 102 may then perform, in response to and based on the prediction, at least one responsive action (*e.g.*, one or more steps taken in response to a condition). The at least one responsive action may include, but is not limited to, transmitting a message to computing device 106 (*e.g.*, transmitting an alert, transmitting information about the prediction, transmitting a targeted advertisement, etc.), disabling a user account (*e.g.*, disabling a transaction account associated with a payment device that is exhibiting anomalous behavior associated with the prediction), assigning a label associated with the change to at least one record of the second dataset (*e.g.*, to modify the record or

a group of records as stored in memory 104 to identify the record or group of records as exhibiting a change in data behavior, as indicated by the prediction), and/or the like. Modeling system 102 may further decode the at least one feature to determine at least one field associated with the at least one feature (*e.g.*, if the prediction indicates that merchant category code “06” is associated with the prediction that a change in behavior has occurred, modeling system 102 may decode the feature to indicate that the associated field is “merchant category code”). Modeling system 102 may then transmit the message to computing device 106 including the at least one field (*e.g.*, to display the at least one field on a display of computing device 106 in a user interface to inform a user that the at least one field is associated with the prediction that a change in behavior has occurred).

[0149] In some non-limiting embodiments or aspects, modeling system 102 may perform unsupervised feature engineering, as described herein, to produce automatic recommendations of features based on parameters set by a user. A user may monitor and give feedback to the process via a pipeline (*e.g.*, a step-by-step process, which may be displayed on subsequent windows) in a user interface of computing device 106. For example, the user may first identify a data source (*e.g.*, a location in memory 104 where data is stored, such as on a Hadoop cluster) of the dataset for modeling system 102 to analyze (*e.g.*, in a “Data Source” window of the user interface of the pipeline). Next, the user may select one or more columns of the data source for data profiling (*e.g.*, in a “Columns Selection” window of the user interface of the pipeline). The user may view and confirm the dataset name (*e.g.*, in a text field), indicate whether or not the data is aggregated (*e.g.*, with a checkbox or radio button), and identify one or more columns of data for modeling system 102 to analyze (*e.g.*, in a dropdown). To prevent excessive computation time, the user may be limited in the number of columns that may be selected if the columns do not already contain aggregated data (*e.g.*, 20 columns). The user may select the columns to use from a dropdown that automatically populates based on the data source identified by the user in the first step of the pipeline. The user may then proceed to the next step of the pipeline by selecting a confirmation input (*e.g.*, clicking a “Start Data Profiling” button). Modeling system 102 may then perform an initial data profiling process (*see* process 403 shown in FIG. 4A).

[0150] Further in the process pipeline, modeling system 102 may then generate and display, for the user, information about the data profiling process (*e.g.*, in a

“Columns Summary” window of the pipeline in the user interface). The user may be presented with sample data (*e.g.*, showing a subset of the dataset identified by the user, including a number of record rows and field columns, which may be paginated if the sample data takes up substantial space). The user may also be presented with a column-level profiling summary, which indicates how modeling system 102 proposes to profile the dataset. The column-level profiling summary may include a table showing a column for field name, a column for whether the field is produced from a partition of data, a column for the type of the field (*e.g.*, having a data type including categorical, continuous, etc.), a column for a recommendation of whether to include the field in the analysis (*e.g.*, “Yes” or “No”), a count column for the number of distinct values for the field, and a grouping column that indicates how categorical fields are being grouped (*e.g.*, to group similar categories, such as response code fields that all indicate approval). The grouping column may include a selection option in each row (*e.g.*, an embedded link) to allow the user to view, in greater detail (*e.g.*, in a pop-up window or other screen) how categorical fields are being grouped. The column-level profiling summary may further include a control (*e.g.*, an edit button in an edit column, each button including a pencil icon) for the user to edit each row of the data profiling summary, to allow the user to modify how the eventual profiled data table (*see* profiled data table 404 of FIG. 4A) will be produced. Any edits the user makes to the column-level profiling summary may be transmitted to modeling system 102 to modify how the raw data will be profiled.

[0151] Further in the user pipeline, modeling system 102 may generate and display information and controls for the user to modify parameters for the configuration data (*e.g.*, in the “Columns Summary” window of the user interface of the pipeline). For example, the user interface of computing device 106 may display, below the column-level profiling summary, a list of parameters of configuration data (*e.g.*, under a heading titled “Data Selection & Aggregation Info”). Such parameters may include, but are not limited to: aggregation keys (*e.g.*, displayed in a dropdown for a user to select), aggregation time granularity (*e.g.*, by days; the granularity options may be presented in a dropdown for a user to select), time column (*e.g.*, the column of the data to use as the time definition, which may be presented in a dropdown for the user to select), time column format (*e.g.*, YYYY-MM-DD; the options for time format may be presented in a dropdown for a user to select), training start time (*e.g.*, a time value to set as the start point for selecting training data from the dataset for a machine

learning model, which may be set by user in a time selector field), partition date column (*e.g.*, the column of the data to use as a partition by date, which may be presented in a dropdown for the user to select), and partition date format (*e.g.*, YYYY-MM-DD; the options for date format may be presented in a dropdown for the user to select). All parameters may be pre-populated by modeling system 102. After the user views and/or modifies any of the parameters, the user may prompt modeling system 102 to proceed to generate the plurality of time-series features (*e.g.*, in a time-series feature aggregation table) and determine a subset of the plurality of time-series features. The user may prompt modeling system 102 to proceed, for example, through input in the user interface of computing device 106 (*e.g.*, a button labeled “Trigger Data Aggregation”).

[0152] Further in the process pipeline, modeling system 102 may generate and display information for the user (*e.g.*, in the user interface of computing device 106) concerning a plurality of time-series features that may be selected from for training a machine learning model (*e.g.*, in a window titled “Feature List”). The list of features may include a list of primary continuous features, a list of primary categorical features, and/or a list of possible features that may be derived from the primary features (*e.g.*, by calculating a ratio between a first feature and a second feature). The user may be permitted to approve or disapprove of any features to include in the modeling system’s 102 analysis. Once the user has reviewed, modified, and/or approved the list of features to be analyzed by modeling system 102, the user may instruct modeling system 102 to proceed to analyze those features (*e.g.*, by clicking a button named “Derive Features”).

[0153] Further in the process pipeline, modeling system 102 may generate and display information for the user (*e.g.*, in the user interface of computing device 106) concerning the features that are recommended for training a machine learning model (*e.g.*, in a window titled “Feature Selection”). For example, modeling system 102 may display the list of features that were generated by modeling system 102. The list of features generated by modeling system 102 may specifically indicate which features are recommended for inclusion in training a machine learning model (*e.g.*, a subset of the plurality of time-series features). The user may be presented the option to select which features to keep or exclude from training (*e.g.*, a checkbox next to each generated feature in a column titled “Include in Modeling”, which can be checked “Yes” or “No”). Features that modeling system 102 recommends to keep (*e.g.*, based on

score analysis, such as shown in FIG. 6) may automatically be selected for inclusion in training. Once the user has reviewed, modified, and/or approved the list of features for use by modeling system 102 to train the machine learning model, the user may instruct modeling system 102 to proceed to train the machine learning model (*e.g.*, by clicking a button named “Start Model Training”).

[0154] Further in the user pipeline, modeling system 102 may generate and display information for the user (in the user interface of computing device 106) concerning the execution of the machine learning model that was trained on the selected features (*e.g.*, in a window titled “Alerts”). For example, modeling system 102 may run the trained machine learning model and detect likely changes in data behavior (*e.g.*, anomalies) and display such changes to the user in the user interface of computing device 106. The user may also be presented with options of setting up parameters for alerts for trained models being executed on ongoing processed data. With further reference to the foregoing pipeline and at any time in the process, the user may restart the pipeline process (*e.g.*, by selecting a button named “Reset Pipeline”) or may prompt the modeling system 102 to proceed automatically to deploy the model (*e.g.*, by selecting a button named “Deploy Model”). In the latter instance, modeling system 102 may automatically proceed and set any parameters not already reviewed and approved by the user.

[0155] In some non-limiting embodiments or aspects, modeling system 102 may use a combination of supervised and unsupervised feature importance processes to determine a subset of relevant features from a plurality of candidate features, for the purposes of model training. For example, modeling system 102 may receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields that are associated with a plurality of features (*e.g.*, each feature associated with one or more fields of data records) that are candidates for training one or more machine learning models. Modeling system 102 may determine a first relevance score (*e.g.*, a first score associated with an impact that a feature has on an output of a machine learning model) for each feature of the plurality of features based on at least one first machine learning model (*e.g.*, a supervised machine learning model, such as a gradient boosting model). Modeling system 102 may determine a second relevance score (*e.g.*, a second score associated with an impact that a feature has on an output of a machine learning model) for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy

metric, and a statistical metric. Modeling system 102 may determine a subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features. Modeling system 102 may further determine a subset of the plurality of fields for training at least one second machine learning model based on the subset of the plurality of features. Modeling system 102 may further train at least one second machine learning model (*e.g.*, a transaction fraud detection model, a targeted advertisement model, a network security model, and/or the like) based on the subset of the plurality of fields.

[0156] In some non-limiting embodiments or aspects, while determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, modeling system 102 may compare the first relevance score (*e.g.*, a value that is 0 or 1, a value between 0.00 and 1.00, etc.) for each feature of the plurality of features to a first threshold (*e.g.*, a value that is 0 or 1, a value between 0.00 and 1.00, etc.). Modeling system 102 may further compare the second relevance score (*e.g.*, a value that is 0 or 1, a value between 0.00 and 1.00, etc.) for each feature of the plurality of features to a second threshold (*e.g.*, a value that is 0 or 1, a value between 0.00 and 1.00, etc.). Modeling system 102 may further determine that the first relevance score and second relevance score of each feature of the subset of the plurality of features satisfy the first threshold and second threshold, respectively.

[0157] In some non-limiting embodiments or aspects, modeling system 102 may generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features. For example, modeling system 102 may set the first threshold and/or the second threshold at values such that only x features will satisfy the first threshold and the second threshold, where x is the target number of features. The target number of features may be a number that is determined to produce reliable results when training a machine learning model, but to prevent computational bloat by avoiding inclusion of too many features (*e.g.*, between 80 and 100 features). In some non-limiting embodiments or aspects, x may be a numerical range of a target number of features.

[0158] In some non-limiting embodiments or aspects, modeling system 102 may determine the first threshold and/or the second threshold based at least partly on prior iterations of a feature engineering and model training process (*e.g.*, in a feedback loop), such that information about the number of features that satisfied the thresholds

and the values of the thresholds is fed forward to the next iteration of feature engineering and model training. Additionally or alternatively, modeling system 102 may iteratively adjust the thresholds up and/or down to yield a subset of the plurality of features that satisfies the thresholds at or within a range of the target number of features. For example, modeling system 102 may, while generating the first threshold and the second threshold, repeatedly perform an iterative adjustment process until a number of features in the subset of the plurality of features satisfies the predetermined target number of features. The iterative adjustment process may include determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, and based on the first threshold and the second threshold. In response to the number of features in the subset of the plurality of features not satisfying (*e.g.*, meeting, exceeding, falling below, etc.) the predetermined target number of features, modeling system 102 may adjust (*e.g.*, increase, decrease, etc.) the first threshold and/or the second threshold.

[0159] In some non-limiting embodiments, modeling system 102 may receive the first threshold and/or the second threshold from a computing device 106 of a user, such as a data engineer that has expertise in the field. The user may, via a user interface on computing device 106, input the first threshold and/or the second threshold, which may be transmitted to modeling system 102 for use in the above-described processes. Modeling system 102 may communicate back and forth with computing device 106 (*e.g.*, via communication network 108) throughout the feature engineering process, such as to display results, receive user input, allow for process monitoring, etc. For example, modeling system 102 may communicate the subset of the plurality of features to computing device 106 of the user, such as for presentation to the user on a display of computing device 106. The user may then make a selection, in a user interface of computing device 106, of features within the subset of the plurality of features (*e.g.*, a second subset including none, one, some, or all of the features within the subset). Computing device 106 may transmit the selection to modeling system 102. Modeling system 102 may receive the selection of the second subset of features from the subset of the plurality of features from computing device 106. Modeling system 102 may then update the subset of the plurality of features to (i) remove the second subset of features from the subset of the plurality of features, or (ii) exclude all other features from the subset of the plurality of features that are not in the second subset of features.

[0160] In some non-limiting embodiments or aspects, modeling system 102 may train the at least one second machine learning model based on the subset of the plurality of features (*e.g.*, using fields corresponding to the features from records of historic data) in a first time period (*e.g.*, a day, a week, a month, etc.). Modeling system 102 may further receive transaction data (*e.g.*, transaction authorization request records, transaction clearance/settlement records, etc.) associated with a plurality of transactions completed in a second time period (*e.g.*, a day, a week, a month, etc.) subsequent the first time period. Modeling system 102 may generate a prediction (*e.g.*, a classification based on a model output) that a change in expected behavior (*e.g.*, transaction activity) has occurred, using the at least one second machine learning model and based on the transaction data (*e.g.*, by inputting at least a portion of the transaction data to the at least one second machine learning model after it is trained, including on historic transaction data). In response to generating the prediction that the change in expected behavior has occurred, modeling system 102 may perform at least one responsive action (*e.g.*, transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one transaction of the plurality of transaction, etc.) based on the prediction that the change in expected behavior has occurred.

[0161] In some non-limiting embodiments or aspects, modeling system 102 may use the first relevance score and second relevance score in combination to include features in the subset of the plurality of features, and/or to exclude features from the subset of the plurality of features. For example, modeling system 102 may determine that the first relevance score (*e.g.*, a value of 1) of a first feature of the plurality of features satisfies at least one threshold (*e.g.*, a value of 1) and, in response to the determination, include the first feature in the subset of the plurality of features. Modeling system 102 may further determine that the second relevance score (*e.g.*, a value of 1) of the first feature also satisfies the at least one threshold (*e.g.*, a value of 1) and, in response to the determination, include the first feature in the subset of the plurality of features. By way of further example, modeling system 102 may determine that either or both of the first relevance score and the second relevance score of a second feature of the plurality of features does not satisfy the at least one threshold, and in response to the determination, exclude the second feature from the subset of the plurality of features. The at least one threshold may include, for the first relevance score, a plurality of different thresholds corresponding to two or more of the variance

metric, the correlation metric, the entropy metric, and the statistical importance metric. As such, modeling system 102 may determine (*e.g.*, in an ensemble analysis) that two or more values of the variance metric, the correlation metric, the entropy metric, and the statistical metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0162] The number and arrangement of systems and devices shown in FIG. 1 are provided as an example. There may be additional systems and/or devices, fewer systems and/or devices, different systems and/or devices, or differently arranged systems and/or devices than those shown in FIG. 1. Furthermore, two or more systems or devices shown in FIG. 1 may be implemented within a single system or device, or a single system or device shown in FIG. 1 may be implemented as multiple, distributed systems or devices. Additionally or alternatively, a set of systems (*e.g.*, one or more systems) or a set of devices (*e.g.*, one or more devices) of system 100 may perform one or more functions described as being performed by another set of systems or another set of devices of system 100.

[0163] Referring now to FIG. 2, shown is a diagram of example components of a device 200, according to non-limiting embodiments. Device 200 may correspond to one or more devices of modeling system 102, memory 104, computing device 106, and/or communication network 108, as an example. In some non-limiting embodiments, such systems or devices may include at least one device 200 and/or at least one component of device 200. The number and arrangement of components shown are provided as an example. In some non-limiting embodiments, device 200 may include additional components, fewer components, different components, or differently arranged components than those shown. Additionally, or alternatively, a set of components (*e.g.*, one or more components) of device 200 may perform one or more functions described as being performed by another set of components of device 200.

[0164] As shown in FIG. 2, device 200 may include bus 202, processor 204, memory 206, storage component 208, input component 210, output component 212, and communication interface 214. Bus 202 may include a component that permits communication among the components of device 200. In some non-limiting embodiments or aspects, processor 204 may be implemented in hardware, firmware, or a combination of hardware and software. For example, processor 204 may include a processor (*e.g.*, a central processing unit (CPU), a graphics processing unit (GPU),

an accelerated processing unit (APU), etc.), a microprocessor, a digital signal processor (DSP), and/or any processing component (*e.g.*, a field-programmable gate array (FPGA), an application-specific integrated circuit (ASIC), etc.) that can be programmed to perform a function. Memory 206 may include random access memory (RAM), read only memory (ROM), and/or another type of dynamic or static storage device (*e.g.*, flash memory, magnetic memory, optical memory, etc.) that stores information and/or instructions for use by processor 204.

[0165] With continued reference to FIG. 2, storage component 208 may store information and/or software related to the operation and use of device 200. For example, storage component 208 may include a hard disk (*e.g.*, a magnetic disk, an optical disk, a magneto-optic disk, a solid state disk, etc.) and/or another type of computer-readable medium. Input component 210 may include a component that permits device 200 to receive information, such as via user input (*e.g.*, a touch screen display, a keyboard, a keypad, a mouse, a button, a switch, a microphone, etc.). Additionally, or alternatively, input component 210 may include a sensor for sensing information (*e.g.*, a global positioning system (GPS) component, an accelerometer, a gyroscope, an actuator, etc.). Output component 212 may include a component that provides output information from device 200 (*e.g.*, a display, a speaker, one or more light-emitting diodes (LEDs), etc.). Communication interface 214 may include a transceiver-like component (*e.g.*, a transceiver, a separate receiver and transmitter, etc.) that enables device 200 to communicate with other devices, such as via a wired connection, a wireless connection, or a combination of wired and wireless connections. Communication interface 214 may permit device 200 to receive information from another device and/or provide information to another device. For example, communication interface 214 may include an Ethernet interface, an optical interface, a coaxial interface, an infrared interface, a radio frequency (RF) interface, a universal serial bus (USB) interface, a Wi-Fi® interface, a cellular network interface, and/or the like.

[0166] Device 200 may perform one or more processes described herein. Device 200 may perform these processes based on processor 204 executing software instructions stored by a computer-readable medium, such as memory 206 and/or storage component 208. A computer-readable medium may include any non-transitory memory device. A memory device includes memory space located inside of a single physical storage device or memory space spread across multiple physical

storage devices. Software instructions may be read into memory 206 and/or storage component 208 from another computer-readable medium or from another device via communication interface 214. When executed, software instructions stored in memory 206 and/or storage component 208 may cause processor 204 to perform one or more processes described herein. Additionally, or alternatively, hardwired circuitry may be used in place of or in combination with software instructions to perform one or more processes described herein. Thus, embodiments described herein are not limited to any specific combination of hardware circuitry and software. The term “configured to,” as used herein, may refer to an arrangement of software, device(s), and/or hardware for performing and/or enabling one or more functions (e.g., actions, processes, steps of a process, and/or the like). For example, “a processor configured to” may refer to a processor that executes software instructions (e.g., program code) that cause the processor to perform one or more functions.

[0167] Referring now to FIG. 3, shown is a flow diagram for a method for unsupervised feature engineering for use in machine learning models, according to some non-limiting embodiments or aspects. The steps shown in FIG. 3 are for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, a step may be automatically performed in response to performance and/or completion of a prior step. In some non-limiting embodiments or aspects, one or more of the steps of process 300 may be performed (e.g., completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps of process 300 may be performed (e.g., completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102.

[0168] As shown in FIG. 3, at step 302, process 300 may include receiving a dataset. For example, modeling system 102 may receive a dataset including a plurality of records, each record of the plurality of records including a plurality of fields. In some non-limiting embodiments or aspects, modeling system 102 may process a plurality of transactions between at least one payment device and at least one merchant over a plurality of time intervals to produce transaction data, and modeling system 102 may store the transaction data (e.g., in time-series format) in at least one

database (e.g., memory 104). The transaction data may include the dataset received in step 302.

[0169] As shown in FIG. 3, at step 304, process 300 may include determining a data type of a field. For example, modeling system 102 may determine a data type of a field of the plurality of fields of the dataset. In some non-limiting embodiments or aspects, modeling system 102 may repeat step 304 for each field of the plurality of fields. In some non-limiting embodiments or aspects, modeling system 102 may, while determining the data type of the field of the plurality of fields, determine that the data type includes at least one of a categorical data type or a continuous data type.

[0170] As shown in FIG. 3, at step 306, process 300 may include determining a distribution of values of the field. For example, modeling system 102 may determine a distribution of values of the field based on the plurality of records (e.g., by evaluating all values in the plurality of records corresponding to the given field). In some non-limiting embodiments or aspects, modeling system 102 may repeat step 306 for each field of the plurality of fields.

[0171] As shown in FIG. 3, at step 308, process 300 may include generating configuration data based on the data type and the distribution of values of the field. For example, modeling system 102 may generate configuration data based on the data type and the distribution of values of the field. In some non-limiting embodiments or aspects, the configuration data may include a JSON file including structured arguments that define parameters for generating a plurality of time-series features (e.g., in a time-series aggregated feature table).

[0172] As shown in FIG. 3, at step 310, process 300 may include generating a plurality of time-series features based on the configuration data. For example, modeling system 102 may generate the plurality of time-series features based on the configuration data and the plurality of records. In some non-limiting embodiments or aspects, generating the plurality of time-series features may include, in part, converting the dataset to a time-series format.

[0173] As shown in FIG. 3, at step 312, process 300 may include determining a relevance subscore for each time-series feature of the plurality of time-series features. For example, modeling system 102 may determine a relevance subscore for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric (e.g., simple and/or complex importance), or any combination thereof. In some non-

limiting embodiments or aspects, modeling system 102 may determine a first relevance subscore for each feature based on the variance metric of the feature, determine a second relevance subscore for each feature based on the correlation metric of the feature with other features, determine a third relevance subscore for each feature based on the entropy metric of the feature, and determine a fourth relevance subscore of each feature based on the statistical importance metric of the feature. The relevance score may be based on the first relevance subscore, the second relevance subscore, the third relevance subscore, and/or the fourth relevance subscore (*e.g.*, as a combined score, such as a normalized mean; as a count of individual scores that indicate the feature is likely relevant; and/or the like).

[0174] As shown in FIG. 3, at step 314, process 300 may include determining a subset of the plurality of time-series features based on the relevance score. For example, modeling system 102 may determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features. While determining the subset of the plurality of time-series features, modeling system 102 may determine that the relevance score of a first feature of the plurality of features satisfies at least one threshold, and include the first feature in the subset of the plurality of time-series features. While determining the subset of the plurality of time-series features, modeling system 102 may further determine that the relevance score of a second feature of the plurality of features does not satisfy the at least one threshold, and exclude the second feature from the subset of the plurality of time-series features. In some non-limiting embodiments or aspects, the at least one threshold may include a plurality of different thresholds corresponding to each metric that the relevance score is based on (*e.g.*, a composite relevance score determined from a plurality of relevance subscores of individual metrics), and while determining the subset of the plurality of time-series features, modeling system 102 may compare each metric that the relevance score is based on to a corresponding threshold of the at least one threshold (*e.g.*, comparing a relevance subscore of a variance metric to a variance threshold for a feature, comparing a relevance subscore of a correlation metric to a correlation threshold for a feature, comparing a relevance subscore of an entropy metric to an entropy threshold for a feature, comparing a relevance subscore of a statistical importance metric to a statistical importance threshold for a feature, etc.). The composite relevance score may be determined based on how many relevance subscores satisfy a corresponding threshold (*e.g.*, two

or more satisfied thresholds may determine the composite relevance score as indicating the feature is relevant), and/or may be computed from the subscores (*e.g.*, as a mean).

[0175] In some non-limiting embodiments or aspects, modeling system 102 may generate at least one higher-level time-series feature (*e.g.*, a feature based on the aggregation of two or more fields) based on at least one time-series feature of the plurality of time-series features. Modeling system 102 may further update the subset of the plurality of time-series features to include the at least one higher-level time-series feature.

[0176] In some non-limiting embodiments or aspects, at step 314, process 300 may further include modifying the subset of the plurality of time-series features based on a selection of a user. For example, modeling system 102 may communicate the subset of the plurality of time-series features to a computing device (*e.g.*, computing device 106) of a user. Modeling system 102 may receive a selection of a second subset of features from the subset of the plurality of time-series features from the computing device 106 of the user. Modeling system 102 may further update the subset of the plurality of time-series features to remove the second subset of features from the subset of the plurality of time-series features, or exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

[0177] As shown in FIG. 3, at step 316, process 300 may further include determining a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features. For example, modeling system 102 may determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features. In some non-limiting embodiments or aspects, modeling system 102 may decode one or more features to determine one or more fields (*e.g.*, deconstruct and/or identify fields from aggregated time-series features).

[0178] In some non-limiting embodiments or aspects, at step 316, process 300 may include training the machine learning model and using the trained machine learning model to make a prediction and perform at least one responsive action based on the prediction. For example, modeling system 102 may train the machine learning model based on the subset of the plurality of fields. Modeling system 102 may receive a second dataset including a second plurality of records. Modeling system 102 may

generate a prediction that a change in expected behavior has occurred (*e.g.*, detected transaction fraud, detected change in user spending pattern, detected change in user network resource access pattern, detected change in user network access time, etc.) using the machine learning model, and based on the second plurality of records. Modeling system 102 may perform at least one responsive action based on the prediction that the change in expected behavior has occurred. The at least one responsive action may include, but is not limited to, transmitting a message to a computing device (*e.g.*, computing device 106), disabling a user account, assigning a label associated with the change to at least one record of the second dataset, and/or the like.

[0179] Referring now to FIG. 4A, depicted is a schematic diagram of a process for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects. The process shown in FIG. 4A is for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 4A may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 4A may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102. It will also be appreciated that the tables and arrangements of data shown in FIG. 4A are for illustration and are not to be interpreted as limiting.

[0180] As shown in FIG. 4A, the process may include receiving raw data 402 from a dataset including a plurality of records (*e.g.*, rows), wherein each record includes a plurality of fields (*e.g.*, cells corresponding to columns in the row). Each field may represent a different aspect of the data record. As illustrated, in the example of transaction records of raw data 402, fields may include, but are not limited to, entity type (field “Entity”), merchant category code (field “Merch. Code”), response code (field “Resp. Code”), e-commerce indicator code (field “ECI Code”), and/or the like. As illustrated, a first record of raw data 402 indicates an Entity of “A”, a Merch. Code of 6011, a Resp. Code of 00, and an ECI Code of 05. A second record of raw data 402 indicates an Entity of A, a Merch. Code of 5041, a Resp. Code of 00, and an ECI Code of 06. A third record of raw data 402 indicates an Entity of “A”, a Merch. Code of 5024,

a Resp. Code of 10, and an ECI Code of 05. It will be appreciated that the raw data may include a plurality of additional records, as indicated by the “-” symbol in illustration of the raw data 402.

[0181] As shown in FIG. 4A, the process may include a data profiling process 403 to profile raw data 402 (*e.g.*, organize aspects of raw data 402 into profiles, such as rows of profiled data table 404). The profiled data of raw data 402 may form profiled data table 404. Each row of profiled data table 404 may be a profile of the data of raw data 402. For example, profiled data table 404 may include a column for the field name (column “Field”), a column for the number of distinct values for that field name (column “N”), a column for the distribution of values within that field name (column “Values”), and a column of the field values at relevant percentiles to demonstrate a statistical distribution of the field (column “Percentiles”). As shown, a first profile of profiled data table 404 indicates that for field Resp. Code, there are 8 distinct field values having a distribution of values that includes at least values 000, 100, 011, among other values. A second profile of profiled data table 404 indicates that for field ECI Code, there are 3 distinct field values having a distribution of values that includes values 05, 06, and 07. A third profile of data table 404 indicates that for field Merch. Code, there are 45 distinct field values having a distribution of values that includes at least values 6011, 5041, among other values. As part of the data profiling process 403, modeling system 102 may determine the statistical distribution of a field and profile the fields of raw data 402 accordingly. For example, modeling system 102 may determine that a field is categorical in nature and may group similar categorical values into a group field (*e.g.*, where multiple categories for response code indicate approval, the multiple fields may be combined into a single categorical value for approval). By way of another example, modeling system 102 may determine that a field is continuous in nature and may determine a range of values (*e.g.*, a minimum and maximum) for the continuous range of the field. It will be appreciated that modeling system 102 may also combine a continuous field (*e.g.*, transaction amount) with a categorical field (*e.g.*, ECI Code) to produce a composite field.

[0182] As shown in FIG. 4A, the process may include a configuration data generation process 405. For example, modeling system 102 may convert the tabular data of profiled data table 404 into a structured configuration argument (*e.g.*, a JSON data file), to produce configuration data 406. Each profile (*e.g.*, row) of the profiled data table 404 may be converted, in the configuration generation process 405, into a

respective set of configuration data 406. For example, if the profiled field is e-commerce indicator (ECI Code), which has a categorical data type with a distribution of three values including 05, 06, and 07, then configuration data 406 may be specified as the following exemplary JSON argument (*e.g.*, specified in a JSON file): {"columnName": ["ECI Code"], "columnType": "categorical", "columnFilters": ["in","in","in"], "columnValues": [{"05"}, {"06"}, {"07"}], "calculation": "count", "calculationColumn": "", "partitionColumn": ["Entity"], "newColumnPrefix": "count_trans_eci_"}.

In some non-limiting embodiments or aspects, modeling system 102 may generate proposed configuration data 406 and transmit the proposed configuration data 406 to computing device 106 of a user, for approval, rejection, and/or modification by the user. The user may use computing device 106 to optionally modify configuration data 406 and transmit the optionally modified configuration data 406 back to modeling system 102 for use in subsequent steps.

[0183] Referring now to FIG. 4B, depicted is a schematic diagram continuing the process of FIG. 4A, for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects. The process shown in FIG. 4B is for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 4B may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 4B may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102. It will also be appreciated that the tables and arrangements of data shown in FIG. 4B are for illustration and are not to be interpreted as limiting.

[0184] As shown in FIG. 4B, after configuration data 406 is generated, the process may include a feature creation process 408. For example, modeling system 102 may use configuration data 406 to generate a plurality of time-series features 409. Configuration data 406 may act as an input for feature creation process 408. Based on the type of aggregation specified in configuration data 406, modeling system 102 may aggregate the data of the plurality of records of the dataset to generate plurality of time-series features 409. As illustrated, the exemplary plurality of time-series features 409 indicate a first column for the feature of a count of transactions when ECI

Code is 05 (column "Count_trans_eci_05"), a second column for the feature of a count of transactions when ECI Code is 06 (column "Count_trans_eci_06"), and a third column for the feature of a count of transactions when ECI Code is 07 (column "Count_trans_eci_07"). Plurality of time-series features 409 may include additional columns for additional features. Further non-limiting embodiments or aspects of the feature creation process 408 are illustrated and described in connection with FIG. 5.

[0185] As shown in FIG. 4B, after plurality of time-series features 409 is generated, the process may include a feature selection process 410. For example, modeling system 102 may determine a relevance score for each feature of plurality of time-series features 409 based on one or more metrics of the feature. This provides the benefit of modeling system 102 determining, in an unsupervised fashion, a subset of the plurality of fields for training a machine learning model. Further non-limiting embodiments or aspects of the feature selection process 410 are illustrated and described in connection with FIG. 6. In some non-limiting embodiments or aspects, the subset of the plurality of time-series features may be presented as a proposed list of features to the user by modeling system 102 in computing device 106, to allow the user to modify the proposed list of features and have a final say on how the machine learning model may be trained. Even in such scenarios, modeling system 102 alleviates the analytical burden and removes user bias through a predominantly unsupervised feature engineering process. As shown, the subset of the plurality of time-series features may be presented in feature table 411, which includes a column for an evaluated feature (column "Feature") and a column for a recommendation by modeling system 102 (column "KEEP", where inclusion of the feature in the subset of the plurality of time-series features is indicated by "Yes" or "No"). As illustrated, the exemplary feature of transaction count when ECI Code is 05 ("Count_trans_eci_05") is included in the subset of the plurality of time-series features ("Yes"), the exemplary feature of transaction count when ECI Code is 06 is not included in the subset of the plurality of time-series features ("No"), and the exemplary feature of transaction count when Resp. Code is 00 is included in the subset of the plurality of time-series features ("Yes").

[0186] Referring now to FIG. 5, depicted is a flow diagram of a method for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects. The process shown in FIG. 5 is for example purposes only. It will be appreciated that additional, fewer, different, and/or a different

order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 5 may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 5 may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102. FIG. 5 further illustrates a first process 501 for feature creation (shown as process 408 in FIG. 4B), and further illustrates a second process 503 for feature decoding, to decode features into related fields that may be used to train a machine learning model.

[0187] As shown in FIG. 5, the first process 501 may begin with receiving raw data 502 (*e.g.*, raw transaction data), which may or may not be aggregated as stored (*e.g.*, in memory 104). The raw data 502 is then processed according to the parameters of configuration data. Given that raw data 502 may not be in time-series format, and given that the plurality of time-series features are, by nature, arranged according to time interval, the first process 501 may include processing raw data 502 in a sequence generator 504 (*e.g.*, a process executed by modeling system 102 that generates time-series data, such as sequential data). Sequence generator 504 may generate features as time-series variables, where data is aggregated for each time interval of a plurality of time intervals in the records of raw data 502. For example, where raw data 502 is transaction data, sequence generator 504 may aggregate total transaction amount from individual transaction amount fields, aggregate total approved transaction amount from individual transaction amount fields where response codes indicate approval, aggregate total declined transaction amount from individual transaction amount fields where response codes indicate disapproval, and/or the like. In some non-limiting embodiments or aspects, sequence generator 504 may be skipped for one or more fields of raw data 502 if those fields are already aggregated and/or time-series. The first process 501, therefore, proceeds to identify (through optional sequence generator 504) aggregated time-series features 506.

[0188] As shown in FIG. 5, after the identification of aggregated time-series features 506, first process 501 may include determining calculated features 510 from aggregated time-series features 506. Based on the parameters of the configuration data, modeling system 102 may produce calculated features 510 (*e.g.*, ratios, conditional relationships, etc.) from aggregated time-series features 506. Aggregated

time-series features 506 may be input to a feature generator 508 (*e.g.*, a process of modeling system 102 that generates features) to produce calculated features 510. If the user modified the configuration data to specify additional calculations, feature generator 508 may include the additional calculations as specified by the user. If the aggregated time-series features 506 are already in a calculated format (*e.g.*, configured as ratios), then processing by feature generator 508 may be skipped. The final output of calculated features 510 may be used to produce the features for inclusion in a plurality of time-series features.

[0189] As shown in FIG. 5, sequence generator 504 and feature generator 508 may be included in a feature mapper 514 (*e.g.*, a process of modeling system 102 for mapping fields to features). Feature mapper 514 may generate and store relationships between generated features and the underlying fields of raw data 502. For example, if sequence generator 504 converts non-time-series data of raw data 502 to aggregated time-series features 506, the relationship between the non-time-series data and aggregated time-series features 506 may be mapped (*e.g.*, stored in a relationship table) by feature mapper 514. By way of further example, if feature generator 508 generates calculated features 510 from aggregated time-series features 506, the relationship between the aggregated time-series features 506 and the calculated features 510 may be mapped by feature mapper 514.

[0190] As shown in FIG. 5, the process may include a second process 503 for feature decoding, to decode features into related fields that may be used to train machine learning models. As noted above, feature mapper 514 may map relationships between raw data 502, aggregated time-series features 506, and/or calculated features 510. As such, to determine fields 516 that are associated with a feature in the subset of the plurality of time-series features, modeling system 102 may perform the second process 503 to identify fields 516 that are related to a given feature. For example, modeling system 102 may select (*e.g.*, automatically, via user prompting, etc.) one or more features in the subset of the plurality of time-series features and determine the underlying fields 516 of raw data 502. In second process 503, a given feature may be input to a decoder 512 (*e.g.*, a process of modeling system 102 to decode a feature into one or more fields 516). Decoder 512 may convert the time-series aggregated feature back to raw data based on the feature mapping of feature mapper 514. In some non-limiting embodiments or aspects, decoder 512 may make use of machine learning (*e.g.*, natural language processing) to determine the

level of aggregation to decode, the selected features to decode, the potential feature relationships, and/or the feature mapping, based on input by user.

[0191] Referring now to FIG. 6, depicted is a flow diagram of a method for unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects. The process shown in FIG. 6 is for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 6 may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 6 may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102. FIG. 6 further illustrates non-limiting embodiments or aspects of a process for feature selection (illustrated as feature selection process 410 in FIG. 4B).

[0192] As shown in FIG. 6, modeling system 102 may determine one or more relevance scores 604, 606, 608, 610, 612 (*e.g.*, subscores) for one or more features 602, so that modeling system 102 may determine a subset of the plurality of time-series features. Accordingly, modeling system 102 may input features 602 into an analysis process 603 (*e.g.*, an ensemble analysis to generate multiple statistical scores) to generate one or more relevance scores 604, 606, 608, 610, 612. Each relevance score 604, 606, 608, 610, 612 may be a quantitative value (*e.g.*, between 0 and 1, 0 and 10, and/or the like) or a qualitative value (*e.g.*, a category indicating “relevant” or not, etc.). For example, modeling system 102 may generate a variance relevance score 604 based on a variance metric for each feature of features 602, which may be a relevance score indicating an expectation of the squared deviation of a random variable from its mean. A feature with high internal variance (*e.g.*, above 0.5) may indicate a high likelihood that the feature is relevant and, therefore, variance relevance score 604 may also be high. Modeling system 102 may also generate an entropy relevance score 606 based on an entropy metric for each feature of features 602, which may be a relevance score indicating randomness of the feature. Entropy relevance score 606 may be a measure of information (*e.g.*, distance-based entropy), where high entropy (*e.g.*, above 0.1) indicates high randomness and low information of a given feature, which may indicate a high likelihood that a feature is relevant (which

may result in a high entropy relevance score 606). Modeling system 102 may also generate a correlation relevance score 608 based on a correlation metric for each feature of features 602, which may be a measure of how correlated the feature is with other features. Low correlation of the feature with other features may indicate a high likelihood that the feature is relevant and, therefore, correlation relevance score 608 may also be high. It will be appreciated that the correlation relevance score 608 may be inverted (low correlation = low score) for salience of correlation relevance score 608, but in such instances, the threshold comparisons may also be inverted. Modeling system 102 may also generate importance relevance scores 610, 612 based on simple and complex statistical importance metrics, respectively, which may indicate the ability for a record to be recreated based on the information contained in a feature. High importance may result in high importance relevance scores 610, 612, indicative of a higher likelihood that the feature is relevant. Importance relevance scores 610, 612 may include simple importance relevance score 610 and complex importance relevance score 612. In some non-limiting embodiments or aspects, simple importance relevance score 610 may make use of principal component analysis, and complex importance relevance score 612 may make use of autoencoder.

[0193] As shown in FIG. 6, modeling system 102 may compare one or more of the relevance scores 604, 606, 608, 610, 612 to a corresponding threshold (*e.g.*, predetermined threshold, dynamic threshold, etc.) for each score. At minimum, modeling system 102 may include a feature in a subset of the plurality of time-series features if a relevance score for the feature satisfies at least one threshold. In some non-limiting embodiments or aspects, modeling system 102 may include a feature in a subset of the plurality of time-series features if two or more scores for the feature satisfy two or more thresholds. Additionally or alternatively, modeling system 102 may ensemble (*e.g.*, combine, such as through an average or a weighted average) the sub-scores into a final score and compare the final score to a single threshold.

[0194] As shown in FIG. 6, modeling system 102 may first rank features 602 based on one or more relevance scores 604, 606, 608, 610, 612 in a ranking process 614. For example, each feature of features 602 may be ranked according to how many (*e.g.*, a count) of relevance scores 604, 606, 608, 610, 612 for the feature that satisfied a corresponding threshold (*e.g.*, where a feature having four relevance scores that satisfied a corresponding threshold would be ranked higher than a feature having only one relevance score that satisfied a corresponding threshold). Additionally or

alternatively, features 602 may be quantitatively ranked based on individual relevance scores or composite final relevance scores (*e.g.*, wherein a feature having a higher quantitative relevance score for one or more scores may be ranked higher than a feature with a lower quantitative relevance score). Modeling system 102 may then select, in a selection process 616, the features that were highest ranked from the ranking process 614. For example, modeling system 102 may select top N features that are ranked the highest and include the top N features in the subset of the plurality of time-series features.

[0195] Referring now to FIG. 7, depicted is a flow diagram of a method for combined supervised and unsupervised feature engineering in machine learning models, according to some non-limiting embodiments or aspects. The process shown in FIG. 7 is for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 7 may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps shown in FIG. 7 may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102.

[0196] As shown in FIG. 7, modeling system 102 may receive a feature set 702 including a plurality of features (*e.g.*, hundreds of features, thousands of features, etc.) that are possible candidates for training one or more machine learning models. Modeling system 102 may then use unsupervised feature importance 704 analysis and supervised feature importance 706 analysis to reduce the feature set 702 to a subset of the plurality of features (*e.g.*, tens of features, hundreds of features, etc.) that is smaller than the initial feature set 702.

[0197] In some non-limiting embodiments or aspects, unsupervised feature importance 704 may include using one or more unsupervised metric scores, in comparison to one or more thresholds, to determine whether a given feature should be included in the subset of the plurality of features. The unsupervised metric scores may include, but are not limited to, a variance score 724, an entropy score 726, a correlation score 728, a simple importance score 730, a complex importance score 732, and/or the like. See analysis process 603 (shown in FIG. 6) for further description of determining whether a feature is relevant (*e.g.*, should be included in the subset)

based on the unsupervised metric scores. Each of variance score 724, entropy score 726, correlation score 728, simple importance score 730, and complex importance score 732 may be the same or similar as variance score 604, entropy score 606, correlation score 608, simple importance score 610, and complex importance score 612, respectively as shown in FIG. 6. From the unsupervised metric scores, modeling system 102 may determine an unsupervised relevance score 712. The unsupervised relevance score 712 may include, but is not limited to, a statistical mean (*e.g.*, a sum of each unsupervised metric score divided by the number of unsupervised metric scores), a statistical median (*e.g.*, the value in the middle of a set of unsupervised metric scores), a statistical mode (*e.g.*, a value that appears most often in the set of unsupervised metric scores), and/or the like. For unsupervised metric scores that are binary categorical scores (*e.g.*, where 0 is not relevant and 1 is relevant), the statistical mode may be used to identify whether a majority of the unsupervised metric scores indicate that a feature should be included in the subset of the plurality of features.

[0198] In some non-limiting embodiments or aspects, supervised feature importance 706 may include using one or more supervised machine learning model scores, in comparison to one or more thresholds, to determine whether a given feature should be included in the subset of the plurality of features. For example, the one or more supervised machine learning model scores may include a first supervised machine learning model 708 and a second supervised machine learning model 710. In some non-limiting embodiments or aspects, first supervised machine learning model 708 may include a first gradient boosting model score 708 (*e.g.*, produced by Light GBM), and second supervised machine learning model 710 may include a second gradient boosting model score 710 (*e.g.*, produced by XGBoost). If the values of the supervised machine learning model scores are categorical (*e.g.*, 0 is associated with a non-relevant feature, 1 is associated with a relevant feature), then the thresholds may likewise be categorical (*e.g.*, where the threshold for including a feature is a value of 1). If the values of the supervised machine learning model scores are numerical (*e.g.*, 0.00 is associated with a least likely relevant feature, 1.00 is associated with a most likely relevant feature, and values may fall on an inclusive range therebetween), then the thresholds may likewise be numerical (*e.g.*, a value of 0.50, a value of 0.75, a value of 0.90, a value of 0.95, etc.). For a plurality of supervised machine learning models, modeling system 102 may determine a supervised relevance score 714, which may include, but is not limited to, a statistical mean, a statistical median, a

statistical mode, and/or the like. For supervised model scores that are binary categorical scores (e.g., where 0 is not relevant and 1 is relevant), the statistical mode may be used to identify whether a majority of the supervised model scores indicate that a feature should be included in the subset of the plurality of features.

[0199] In some non-limiting embodiments, unsupervised relevance score 712 and supervised relevance score 714 may be evaluated in comparison to corresponding thresholds. Unsupervised relevance score 712 may or may not satisfy its corresponding threshold, and such a comparison is independent of whether supervised relevance score 714 satisfies its corresponding threshold. The evaluation of relevance scores 712, 714 may produce indications of whether the feature is relevant (e.g., if relevance score 712, 714 satisfies a corresponding threshold, relevance score 712, 714 indicates that the feature should be included). The evaluation may be disjunctive, such that if either unsupervised relevance score 712 or supervised relevance score 714 satisfies a corresponding threshold and indicates that a feature should be included (e.g., is relevant), then modeling system 102 will include the feature in the subset. The evaluation may also be conjunctive, such that both unsupervised relevance score 712 and supervised relevance score 714 must both satisfy a corresponding threshold to indicate that a feature should be included (e.g., is relevant), and for modeling system 102 to include the feature in the subset. In some non-limiting embodiments or aspects, a conjunctive evaluation of unsupervised relevance score 712 and supervised relevance score 714 may itself be disjunctively evaluated with a user input indicating whether or not the feature is relevant. For example, user input from computing device 106 indicating that a feature is relevant may alone be enough for modeling system 102 to include the feature in the subset. Likewise, user input from computing device 106 indicating that a feature is not relevant may alone be enough for modeling system 102 to exclude the feature from the subset, even to override the indications of unsupervised relevance score 712 and supervised relevance score 714. In this manner, the logic operator for determining whether to include a feature may be represented by the following:

Formula 1

$$(\text{If } \text{URS} \geq X \text{ and } \text{SRS} \geq Y) \text{ OR } (\text{UCI} == \text{TRUE})$$

where URS represents unsupervised relevance score 712, SRS represents supervised relevance score 714, X represents a threshold associated with

unsupervised relevance score 712, Y represents a threshold associated with supervised relevance score 714, and UCI represents user computer input.

[0200] It will be appreciated that the above formula may be rewritten if either or both of unsupervised relevance score 712 and supervised relevance score 714 are binary categorical values. For example, in the scenario where unsupervised relevance score 712 is binary, X is the value “TRUE”, and supervised relevance score 714 is numerical, Formula 1 may be rewritten as:

Formula 2

$$(If\ URS == TRUE\ and\ SRS \geq Y)\ OR\ (UCI == TRUE)$$

[0201] In some non-limiting embodiments or aspects, the relationship of the user input with the evaluation of unsupervised relevance score 712 and supervised relevance score 714 may be made conjunctive, where the user input must confirm the evaluations of unsupervised relevance score 712 and supervised relevance score 714 rather than override them disjunctively. In such a scenario, Formula 2 may be rewritten as:

Formula 3

$$(If\ URS == TRUE\ and\ SRS \geq Y)\ AND\ (UCI == TRUE)$$

[0202] Depicted below is an exemplary table showing the evaluations of a plurality of features for inclusion in a final subset of features, according to non-limiting embodiments or aspects of the present disclosure. The first column (“FEAT”) shows feature identifier, the second column (“SMLS1”) shows a first supervised machine learning score 708, the third column (“GBS2”) shows a second supervised machine learning score 710, the fourth column (“SRS”) shows supervised relevance score 714, the fifth column (“URS”) shows a binary unsupervised relevance score 712 (*e.g.*, where “TRUE” indicates a feature is relevant and should be included), the sixth column (“UCI”) shows user input (*e.g.*, where “TRUE” indicates a feature is relevant and should be included), and the seventh column (“FINAL”) shows the final determination by modeling system 102 of whether the feature should be included in the subset of the feature set 702 that will be used to train one or more machine learning models. The below table shows a scenario where user input is treated disjunctively, such that user input may, on its own, be used as a basis to include a feature in the final subset.

Table 3

FEAT	SMLS1	SMLS2	SRS	URS	UCI	FINAL
A	0.1081	0.1333	0.1207	0	TRUE	TRUE
B	0.0901	0.3000	0.1950	1	TRUE	TRUE
C	0.3153	0.0333	0.1743	1	-	TRUE
D	0.0811	0.0667	0.0739	0	TRUE	TRUE
E	0.0180	0.0333	0.0257	0	TRUE	TRUE
F	0.1081	0.1667	0.1374	0	-	FALSE
G	0.0900	0.1333	0.1117	0	-	FALSE

As shown in the above illustrative Table 3, supervised relevance score 714 is determined by a statistical mean of first GB score 708 and second GB score 710, the threshold for supervised relevance score 714 is 0.1500, unsupervised relevance score 712 is determined by a statistical mode of an ensemble of unsupervised metric scores, and the threshold for unsupervised relevance score 712 is 1. In some non-limiting embodiments or aspects, modeling system 102 may transmit the data for the first five columns to computing device 106, for display in a user interface to user. User may view the first five columns and provide their input in the sixth column. The user's input may then be communicated back from computing device 106 to modeling system 102.

[0203] Referring now to FIG. 8, FIG. 8 is a flow diagram of a non-limiting embodiment or aspect of a process 800 for combined supervised and unsupervised feature engineering for use in machine learning models, according to some non-limiting embodiments or aspects. The steps shown in FIG. 8 are for example purposes only. It will be appreciated that additional, fewer, different, and/or a different order of steps may be used in non-limiting embodiments or aspects. In some non-limiting embodiments or aspects, one or more of the steps of process 800 may be performed (*e.g.*, completely, partially, and/or the like) by modeling system 102. In some non-limiting embodiments or aspects, one or more of the steps of process 800 may be performed (*e.g.*, completely, partially, and/or the like) by another system, another device, another group of systems, or another group of devices, separate from or including modeling system 102.

[0204] As shown in FIG. 8, at step 802, process 800 may include receiving a dataset comprising a plurality of records. For example, modeling system 102 may

receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields associated with a plurality of features (*e.g.*, feature set 702) that are candidate features for training one or more machine learning models. Each feature may be associated with one or more fields of data records of a dataset (*e.g.*, a dataset of transaction data).

[0205] As shown in FIG. 8, at step 804, process 800 may include determining a first relevance score for each feature of the plurality of features based on a first machine learning model. For example, modeling system 102 may determine a first relevance score (*e.g.*, supervised relevance score 714) based on at least one first machine learning model (*e.g.*, a supervised machine learning model, such as a gradient boosting model). The first relevance score may be associated with a relevance of a feature for training at least one second machine learning model, which may indicate whether a feature should be included in a final subset of the plurality of features, from which a final subset of the plurality of fields may be determined for training the at least one second machine learning model.

[0206] As shown in FIG. 8, at step 806, process 800 may include determining a second relevance score for each feature of the plurality of features based on at least one metric. For example, modeling system 102 may determine a second relevance score (*e.g.*, unsupervised relevance score 712) for each feature of the plurality of features based on at least one of a variance metric, a correlation metric, an entropy metric, and a statistical importance metric.

[0207] As shown in FIG. 8, at step 808, process 800 may include determining a subset of the plurality of features based on the first relevance score and the second relevance score for each feature. For example, modeling system 102 may determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features.

[0208] In some non-limiting embodiments or aspects, modeling system 102 may, while determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, compare the first relevance score of each feature of the plurality of features to a first threshold, determine that the first relevance score of each feature of the subset of the plurality of features satisfies the first threshold, compare the second relevance score for each feature of the plurality of features to a second threshold, and

determine that the second relevance score for each feature of the subset of the plurality of features satisfies the second threshold.

[0209] In some non-limiting embodiments or aspects, modeling system 102 may, while determining the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, determine that the first relevance score of a first feature of the plurality of features satisfies at least one threshold and determine that the first relevance score of a second feature of the plurality of features does not satisfy the at least one threshold. In response, modeling system 102 may include the first feature in the subset of the plurality of features and exclude the second feature from the subset of the plurality of features. The at least one threshold may include a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical metric. Modeling system 102 may determine the subset of the plurality of features by determining that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

[0210] In some non-limiting embodiments or aspects, modeling system 102 may generate at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of features. For example, modeling system 102 may, while generating at least one of the first threshold and the second threshold based on the predetermined target number of features in the subset of the plurality of features, repeat an adjustment process until a number of features in the subset of the plurality of features satisfies the predetermined target number of features. Modeling system 102 may, in the adjustment process, determine the subset of the plurality of features based on the first relevance score and the second relevance score for each feature of the subset of the plurality of features, and based on the first threshold and the second threshold. Modeling system 102 may further, in the adjustment process, and in response to the number of features in the subset of the plurality of features not satisfying (*e.g.*, meeting, exceeding, falling below, etc.) the predetermined target number of features, adjust (*e.g.*, increase, decrease, etc.) at least one of the first threshold and the second threshold (*e.g.*, and loop again). Modeling system 102 may further, in the adjustment process, and in response to the number of features in the subset of the plurality of features satisfying the

predetermined target number of features, halt the adjustment process and keep the final subset of the plurality of features.

[0211] In some non-limiting embodiments or aspects, modeling system 102 may receive the first threshold and/or the second threshold from computing device 106 of a user.

[0212] In some non-limiting embodiments or aspects, modeling system 102 may communicate the subset of the plurality of features to computing device 106 of a user and receive a selection, from computing device 106, of a second subset of features from the subset of the plurality of features. Modeling system 102 may, in response to receiving the selection, update the subset of the plurality of features to (i) remove the second subset of features from the subset of the plurality of features, or (ii) exclude all other features from the subset of the plurality of features that are not in the second subset of features. Additionally or alternatively, modeling system 102 may receive a selection of a third subset of features from the plurality of features (*e.g.*, feature set 702) from computing device 106, which may or may not include features from the initial subset as determined by modeling system 102. In response to receiving the selection of the third subset of features from computing device 106, modeling system 102 may update the initial subset of features to include any additional features from the third subset that were not yet included in the initial subset of features.

[0213] As shown in FIG. 8, at step 810, process 800 may include determining a subset of the plurality of fields for training at least one second machine learning model. For example, modeling system 102 may determine a subset of the plurality of fields for training at least one machine learning model based on the subset of the plurality of features. In some non-limiting embodiments or aspects, modeling system 102 may further train the at least one second machine learning model (*e.g.*, a fraud detection model, a targeted advertising model, a network security model, etc.) based on the subset of the plurality of fields.

[0214] In some non-limiting embodiments or aspects, modeling system 102 may train the at least one second machine learning model in a first time period. Modeling system 102 may further receive transaction data associated with a plurality of transactions completed in a second time period subsequent the first time period. Modeling system 102 may generate a prediction (*e.g.*, a prediction that a change in expected behavior (*e.g.*, a change in transaction pattern, parameters, activity, etc.) has occurred) using the at least one second machine learning model that was trained

in the first time period, and based on the transaction data. Modeling system 102 may perform at least one remediative action (*e.g.*, transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one transaction of the plurality of transactions, such as to a transaction record in a transaction database, etc.), based on the prediction.

[0215] Although embodiments have been described in detail for the purpose of illustration, it is to be understood that such detail is solely for that purpose and that the disclosure is not limited to the disclosed embodiments or aspects, but, on the contrary, is intended to cover modifications and equivalent arrangements that are within the spirit and scope of the appended claims. For example, it is to be understood that the present disclosure contemplates that, to the extent possible, one or more features of any embodiment or aspect can be combined with one or more features of any other embodiment or aspect.

WHAT IS CLAIMED IS:

1. A computer-implemented method comprising:
 - receiving, with at least one processor, a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields;
 - determining, with at least one processor, a data type of a field of the plurality of fields;
 - determining, with at least one processor, a distribution of values of the field based on the plurality of records;
 - generating, with at least one processor, configuration data based on the data type and the distribution of values of the field;
 - generating, with at least one processor, a plurality of time-series features based on the configuration data and the plurality of records;
 - determining, with at least one processor, a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof;
 - determining, with at least one processor, a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features; and
 - determining, with at least one processor, a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.
2. The computer-implemented method of claim 1, further comprising:
 - generating, with at least one processor, at least one higher-level time-series feature based on at least one time-series feature of the plurality of time-series features; and
 - updating, with at least one processor, the subset of the plurality of time-series features to include the at least one higher-level time-series feature.
3. The computer-implemented method of claim 2, wherein determining the data type of the field of the plurality of fields comprises:

determining, with at least one processor, the data type of the field of the plurality of fields, wherein the data type comprises a categorical data type or a continuous data type.

4. The computer-implemented method of claim 3, further comprising:

processing, with at least one processor, a plurality of transactions between at least one payment device and at least one merchant to produce transaction data; and

storing, with at least one processor, the transaction data in at least one database in a time-series format, wherein the transaction data comprises the dataset.

5. The computer-implemented method of claim 1, wherein determining the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features comprises:

determining that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold;

determining that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold;

including the first feature in the subset of the plurality of time-series features; and

excluding the second feature from the subset of the plurality of time-series features.

6. The computer-implemented method of claim 5, wherein the at least one threshold comprises a plurality of different thresholds corresponding to each of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, and wherein determining the subset of the plurality of time-series features further comprises:

determining that at least two of the variance metric, the correlation metric, the entropy metric, and the statistical importance metric for the first feature satisfy at least two thresholds of the plurality of different thresholds.

7. The computer-implemented method of claim 1, further comprising:

communicating, with at least one processor, the subset of the plurality of time-series features to a computing device of a user;

receiving, with at least one processor, a selection of a second subset of features from the subset of the plurality of time-series features from the computing device of the user; and

updating, with at least one processor, the subset of the plurality of time-series features to (i) remove the second subset of features from the subset of the plurality of time-series features, or (ii) exclude all other features from the subset of the plurality of time-series features that are not in the second subset of features.

8. The computer-implemented method of claim 1, wherein generating the plurality of time-series features further comprises converting, with at least one processor, the dataset to a time-series format.

9. The computer-implemented method of claim 1, further comprising training the machine learning model based on the subset of the plurality of fields.

10. The computer-implemented method of claim 9, further comprising:

receiving, with at least one processor, a second dataset comprising a second plurality of records;

generating, with at least one processor, a prediction that a change in expected behavior has occurred using the machine learning model, and based on the second plurality of records; and

performing, with at least one processor, at least one responsive action based on the prediction that the change in expected behavior has occurred, wherein the at least one responsive action comprises at least one of the following: transmitting a message to a computing device, disabling a user account, assigning a label associated with the change to at least one record of the second dataset, or any combination thereof.

11. The computer-implemented method of claim 1, the method further comprising:

determining, with at least one processor, a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model; and

wherein determining the relevance score further comprises:

determining a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features; and

wherein determining the subset of the plurality of time-series features based on the relevance score further comprises:

determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

12. The computer-implemented method of claim 11, wherein determining the first relevance score for each time-series feature of the plurality of time-series features based on the at least one supervised machine learning model comprises:

determining the first relevance score for each time-series feature of the plurality of time-series features based on the at least one supervised machine learning model, wherein the at least one supervised machine learning model comprises at least one gradient boosting model.

13. The computer-implemented method of claim 11, wherein determining the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features comprises:

comparing the first relevance score of each time-series feature of the plurality of time-series features to a first threshold;

determining that the first relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the first threshold;

comparing the second relevance score for each time-series feature of the plurality of time-series features to a second threshold; and

determining that the second relevance score for each time-series feature of the subset of the plurality of time-series features satisfies the second threshold.

14. The computer-implemented method of claim 13, further comprising generating, with at least one processor, at least one of the first threshold and the second threshold based on a predetermined target number of features in the subset of the plurality of time-series features.

15. A system comprising:

at least one processor configured to:

receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields;

determine a data type of a field of the plurality of fields;

determine a distribution of values of the field based on the plurality of records;

generate configuration data based on the data type and the distribution of values of the field;

generate a plurality of time-series features based on the configuration data and the plurality of records;

determine a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof;

determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features; and

determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

16. The system of claim 15, wherein, while determining the subset of the plurality of time-series features based on the relevance score for each time-series

feature of the subset of the plurality of time-series features, the at least one processor is further configured to:

- determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold;

- determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold;

- include the first feature in the subset of the plurality of time-series features; and

- exclude the second feature from the subset of the plurality of time-series features.

17. The system of claim 15, wherein the at least one processor is further configured to determine a first relevance score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model;

- wherein, while determining the relevance score, the at least one processor is further configured to:

- determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features; and

- wherein, while determining the subset of the plurality of time-series features based on the relevance score, the at least one processor is further configured to:

- determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.

18. A computer program product comprising at least one non-transitory computer-readable medium comprising program instructions that, when executed by at least one processor, cause the at least one processor to:

receive a dataset comprising a plurality of records, each record of the plurality of records comprising a plurality of fields;
determine a data type of a field of the plurality of fields;
determine a distribution of values of the field based on the plurality of records;
generate configuration data based on the data type and the distribution of values of the field;
generate a plurality of time-series features based on the configuration data and the plurality of records;
determine a relevance score for each time-series feature of the plurality of time-series features based on at least one of a variance metric, a correlation metric, an entropy metric, a statistical importance metric, or any combination thereof;
determine a subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features; and
determine a subset of the plurality of fields for training a machine learning model based on the subset of the plurality of time-series features.

19. The computer program product of claim 18, wherein the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score for each time-series feature of the subset of the plurality of time-series features cause the at least one processor to:

determine that the relevance score of a first feature of the plurality of time-series features satisfies at least one threshold;
determine that the relevance score of a second feature of the plurality of time-series features does not satisfy the at least one threshold;
include the first feature in the subset of the plurality of time-series features; and
exclude the second feature from the subset of the plurality of time-series features.

20. The computer program product of claim 18, wherein the program instructions further cause the at least one processor to determine a first relevance

score for each time-series feature of the plurality of time-series features based on at least one supervised machine learning model;

wherein the program instructions that cause the at least one processor to determine the relevance score cause the at least one processor to:

determine a second relevance score for each time-series feature of the plurality of time-series features based on the variance metric, the correlation metric, the entropy metric, and the statistical importance metric, wherein the relevance score comprises a combined relevance score based on the first relevance score and the second relevance score for each time-series feature of the plurality of time-series features; and

wherein the program instructions that cause the at least one processor to determine the subset of the plurality of time-series features based on the relevance score cause the at least one processor to:

determine the subset of the plurality of time-series features based on the first relevance score and the second relevance score for each time-series feature of the subset of the plurality of time-series features.