CLAIMS

What is claimed is:

1. A fraud prevention system comprising:

a client server; and

a fraud prevention server including an electronic processor and a memory, the memory including an online application origination (OAO) service,

wherein, when executing the OAO service, the electronic processor is configured to

receive a training dataset of fraudulent applications,

generate, with machine learning, an online application origination (OAO) model of the OAO service based on the training dataset, the OAO model differentiating between a behavior of a normal user and a fraudulent behavior of a nefarious actor that occurs only during a submission of an online application on a device,

perform feature encoding on one or more categorical variables in a first matrix to generate one or more feature encoded categorical variables, the first matrix including the one or more categorical variables and numerical features, and the one or more feature encoded categorical variables are feature encoded with one or more numerical values of the numerical features,

generate a second matrix including the one or more feature encoded categorical variables, and

generate a feature encoded OAO model by training the OAO model with the second matrix,

receive information regarding the submission of the online application on the device,

determine a fraud score of the online application based on the information that is received and the feature encoded OAO model, the feature encoded OAO model being more precise than the OAO model relative to differentiating between the behaviour of the normal user and the fraudulent behaviour of the nefarious actor that occurs only during the submission of the online application on a device, and the feature encoded OAO model requiring less resources of the client server relating to the OAO model because the feature encoded OAO model is more precise than the OAO model, and

control the client server to approve, hold, or deny the online application based on the fraud score that is determined.

2. The fraud prevention system of claim 1, wherein, to perform the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables, the electronic processor is further configured to

determine an average feature value of one of the numerical features in the training dataset for one of the one or more categorical variables,

determine a number of times the one of the one or more categorical variables was recorded in the training dataset,

determine a population average feature value of the one of the numerical features across a specific time period in the training dataset, and

adjust the average feature value based on the number of times the one of the one or more categorical variables was recorded in the training dataset and the population average feature value.

3. The fraud prevention system of claim 2, wherein, to perform the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables, the electronic processor is further configured to

determine a smoothing value based on a first integer parameter and a second integer parameter,

wherein the first integer parameter defines a minimum number of samples required to determine the average feature value, and

wherein the second integer parameter balances an effect of the average feature value and the population average feature value.

4. The fraud prevention system of claim 3, wherein, to perform the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables, the electronic processor is further configured to

determine an encoded categorical variable value that is equal to the population average feature value multiplied by one minus the smoothing value and an addition of the average feature value multiplied by the smoothing value.

5. The fraud prevention system of claim 4, wherein, to perform the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables, the electronic processor is further configured to add random noise to the encoded categorical variable value.

6. The fraud prevention system of claim 5, wherein, to generate the second matrix including the one or more feature encoded categorical variables, the electronic processor is further configured to generate the second matrix by determining the encoded categorical variable value for each categorical variable.

7. The fraud prevention system of claim 5, wherein the first matrix is a one-dimensional array, and wherein the second matrix is a two-dimensional array.

8. The fraud prevention system of claim 1, wherein the one or more categorical variables include a country categorical variable, and wherein the numerical features include a words-per-minute numerical feature and a time-on-page numerical feature.

9. A method for operating a fraud prevention system, the method comprising:

receiving, with an electronic processor, a training dataset of fraudulent applications;

generating, with the electronic processor and machine learning, an online application origination (OAO) model of the OAO service based on the training dataset, the OAO model differentiating between a behavior of a normal user and a fraudulent behavior of a nefarious actor that occurs only during a submission of an online application on a device;

performing, with the electronic processor, feature encoding on one or more categorical variables in a first matrix to generate one or more feature encoded categorical variables, the first matrix including the one or more categorical variables and numerical features, and the one or more feature encoded categorical variables are feature encoded with one or more numerical values of the numerical features;

generating, with the electronic processor, a second matrix including the one or more feature encoded categorical variables; and

generating, with the electronic processor, a feature encoded OAO model by training the OAO model with the second matrix;

receiving, with the electronic processor, information regarding the submission of the online application on the device;

determining, with the electronic processor, a fraud score of the online application based on the information that is received and the feature encoded OAO model, the feature encoded OAO model being more precise than the OAO model relative to differentiating between the behaviour of the normal user and the fraudulent behaviour of the nefarious actor that occurs only during the submission of the online application on a device, and the feature encoded OAO model requiring less resources of the client server relating to the OAO model because the feature encoded OAO model is more precise than the OAO model; and

controlling, with the electronic processor, a server to approve, hold, or deny the online application based on the fraud score that is determined.

10. The method of claim 9, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining an average feature value of one of the numerical features in the training dataset for one of the one or more categorical variables,

determining a number of times the one of the one or more categorical variables was recorded in the training dataset,

determining a population average feature value of the one of the numerical features across a specific time period in the training dataset, and

adjusting the average feature value based on the number of times the one of the one or more categorical variables was recorded in the training dataset and the population average feature value.

11. The method of claim 10, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining a smoothing value based on a first integer parameter and a second integer parameter,

wherein the first integer parameter defines a minimum number of samples required to determine the average feature value, and

wherein the second integer parameter balances an effect of the average feature value and the population average feature value.

12. The method of claim 11, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining an encoded categorical variable value that is equal to the population average feature value multiplied by one minus the smoothing value and an addition of the average feature value multiplied by the smoothing value.

13. The method of claim 12, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes adding random noise to the encoded categorical variable value.

14. The method of claim 13, wherein generating the second matrix including the one or more feature encoded categorical variables further includes generating the second matrix by determining the encoded categorical variable value for each categorical variable.

15. The method of claim 14, wherein the first matrix is a one-dimensional array, and wherein the second matrix is a two-dimensional array.

16. The method of claim 9, wherein the one or more categorical variables include a country categorical variable, and wherein the numerical features include a words-per-minute numerical feature and a time-on-page numerical feature.

17. A non-transitory computer-readable medium comprising instructions that, when executed by an electronic processor, causes the electronic processor to perform a set of operations comprising:

receiving a training dataset of fraudulent applications;

generating, with machine learning, an online application origination (OAO) model of the OAO service based on the training dataset, the OAO model differentiating between a behavior of a normal user and a fraudulent behavior of a nefarious actor that occurs only during a submission of an online application on a device;

performing feature encoding on one or more categorical variables in a first matrix to generate one or more feature encoded categorical variables, the first matrix including the one or more categorical variables and numerical features, and the one or more feature encoded

categorical variables are feature encoded with one or more numerical values of the numerical features;

generating a second matrix including the one or more feature encoded categorical variables;

generating a feature encoded OAO model by training the OAO model with the second matrix;

receiving information regarding the submission of the online application on the device;

determining a fraud score of the online application based on the information that is received and the feature encoded OAO model, the feature encoded OAO model being more precise than the OAO model relative to differentiating between the behaviour of the normal user and the fraudulent behaviour of the nefarious actor that occurs only during the submission of the online application on a device, and the feature encoded OAO model requiring less resources of the client server relating to the OAO model because the feature encoded OAO model is more precise than the OAO model; and

controlling a server to approve, hold, or deny the online application based on the fraud score that is determined.

18. The non-transitory computer-readable medium of claim 17, performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining an average feature value of one of the numerical features in the training dataset for one of the one or more categorical variables,

determining a number of times the one of the one or more categorical variables was recorded in the training dataset,

determining a population average feature value of the one of the numerical features across a specific time period in the training dataset, and

adjusting the average feature value based on the number of times the one of the one or more categorical variables was recorded in the training dataset and the population average feature value.

19. The non-transitory computer-readable medium of claim 18, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining a smoothing value based on a first integer parameter and a second integer parameter,

wherein the first integer parameter defines a minimum number of samples required to determine the average feature value, and

wherein the second integer parameter balances an effect of the average feature value and the population average feature value.

20. The non-transitory computer-readable medium of claim 19, wherein performing the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables further includes

determining an encoded categorical variable value that is equal to the population average feature value multiplied by one minus the smoothing value and an addition of the average feature value multiplied by the smoothing value.

21. The fraud prevention system of claim 1, wherein, to perform the feature encoding on the one or more categorical variables in the first matrix to generate the one or more feature encoded categorical variables, the electronic processor is further configured to translate a one-dimensional array of strings of the first matrix into a two-dimensional array of real numbers of the second matrix.

22. The fraud prevention system of claim 21, wherein the feature encoded OAO model is more precise than the OAO model by up to 15% after translating the one-dimensional array of strings of the first matrix into the two-dimensional array of real nymbers of the second matrix.