Fortanix®

Examples

FORTANIX CONFIDENTIAL AI: CLASSIFICATION LEARNING AND INFERENCE USING k-NEAREST NEIGHBORs (KNN)

VERSION 1.1

Fortanix Inc. | 800 West El Camino Real | Suit 180 | Mountain View, CA 94040| United States of America | +1 (650) 943-2484
| www.fortanix.cominfo@fortanix.com | www.fortanix.com

TABLE OF CONTENTS

1.0 INTRODUCTION

This document describes how to build and test a model using the K-Nearest Neighbor (KNN) algorithm in Fortanix Confidential AI.

2.0 PREREQUISITES

- A user signed in to a Fortanix Confidential AI account.

For instructions to sign up and log in, refer to our User's Guide: Sign up for Confidential AI.

3.0 BUILDING THE MODEL

3.1 DATA INGESTION

1. On the **DATA INGESTION** page, click **CREATE DATASET** and select **CSV Dataset.**
2. **Dataset name** – Enter a name for your dataset.
3. Select the **Upload a file** option if you want to upload your data directly to the Fortanix Confidential AI platform.
   a. In the **File Upload** section, upload the CSV file. After the file is uploaded, the headers (column names). For example: UserID, Gender, Age, and so on. The number of rows is also detected and displayed.
4. To track what the data is used for, add Labels in the form of "Key:Value" pairs.
5. Click **CREATE DATASET** to save the data.
6. You will now see the saved dataset in the dataset table.

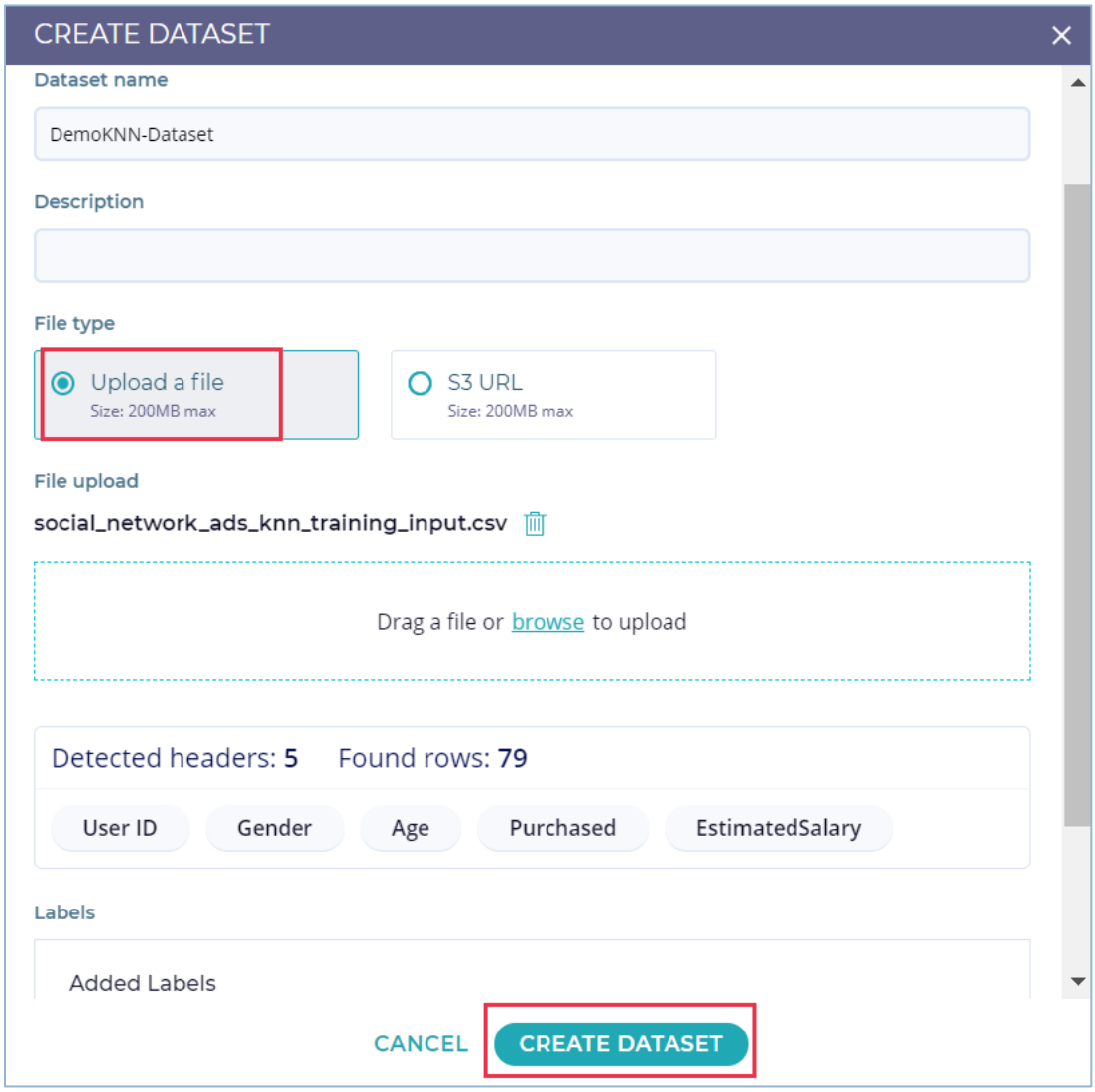

FIGURE 1: CSV DETAILS

3.2 DATA PREPARATION

In this phase, you can choose the column names as a set of features and a target for the tabular dataset. You can choose multiple combinations of features and targets (collectively called **Variables**).

1. In the Data Preparation page, click **ADD VARIABLES** to select the features and target.

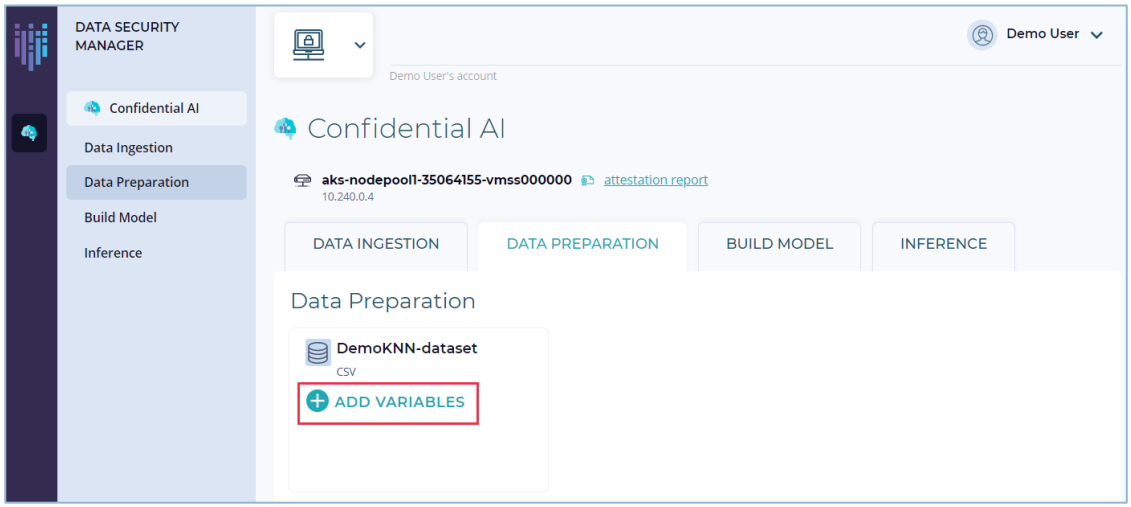

FIGURE 2: ADD VARIABLES

2. Select one or more features from the **SET A FEATURE** column and select one target from the **SET A TAREGT** column.
3. Click **ADD** to add the variables.

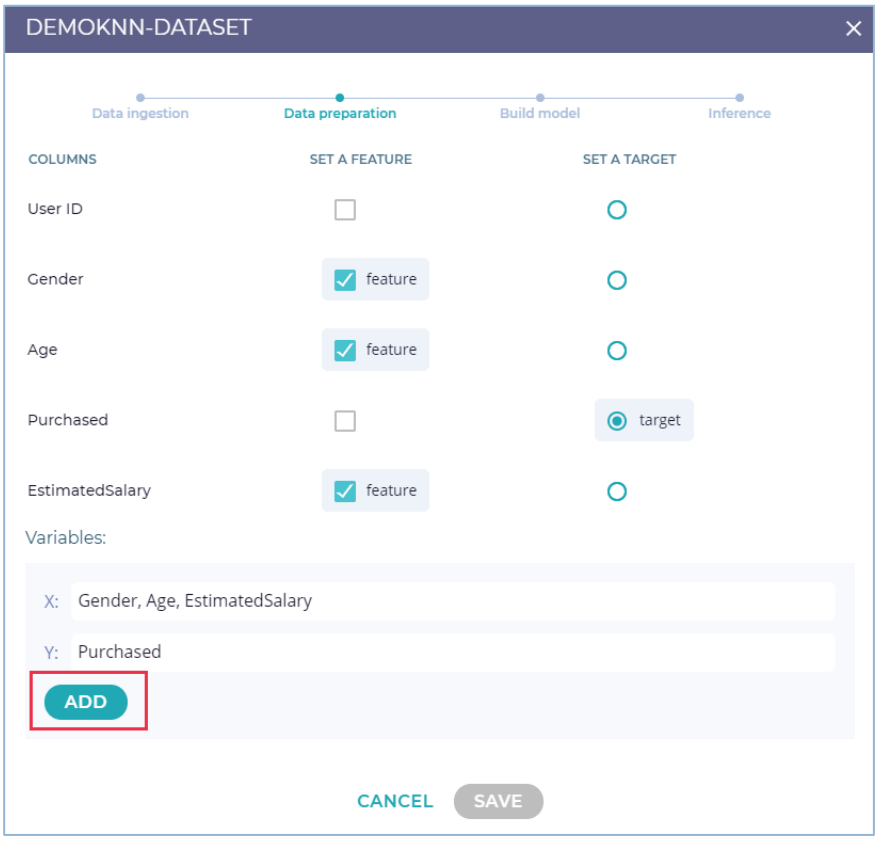

FIGURE 3: SELECT THE FEATURES AND TARGET

4. Click **SAVE** to save the variables and proceed to the next phase, that is, build a model.

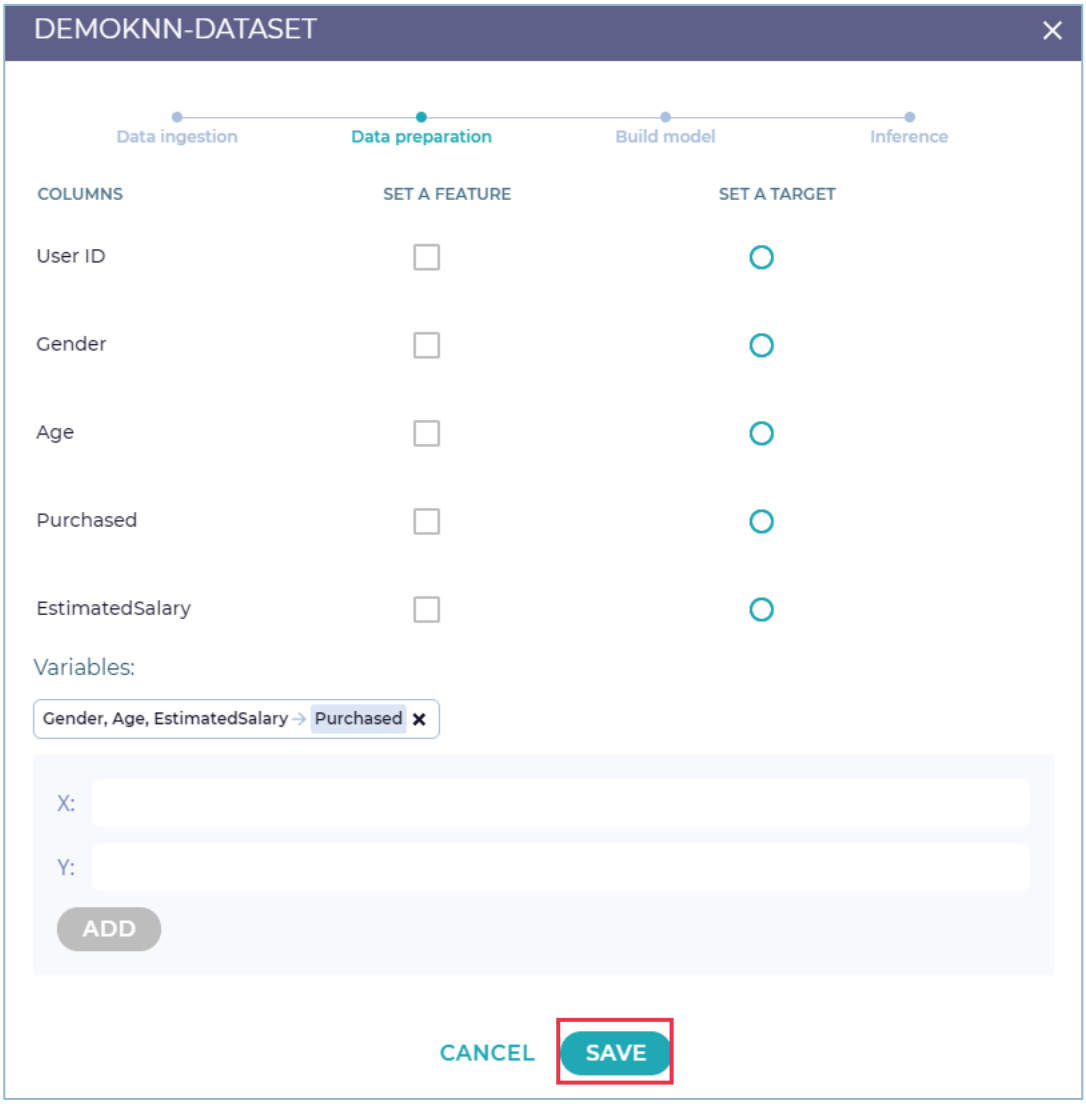

FIGURE 4: VARIABLES ADDED

3.3 BUILD MODEL

In the build a model stage, you can select the Logistic Regression algorithm to run on the dataset defined in the previous phases, to analyze and build AI models.

In the "Build a Model" form:

1. Select the **BUILD A MODEL** tab and click **BUILD MODEL** to build a training model for the dataset created in the previous phase.
2. Enter the **Training flow name,** that is, the name of the model.

3. In the **Training Dataset** field, select the training dataset on which you want to run the KNN algorithm and build a trained model.
4. In the **Algorithm** field, select the **k-Nearest Neighbors** algorithm.
5. **Select ML variables** that you created in the Data Preparation phase.
6. Enter the **Neighbor count** which is the number of nearest neighbors (k) to use in the kNN algorithm.
7. In the **Model name** field, enter a name of the output dataset. This is the output model that will be used in the data inference phase.
8. Click **BUILD MODEL** to run the selected algorithm on the training data and build the model for inference

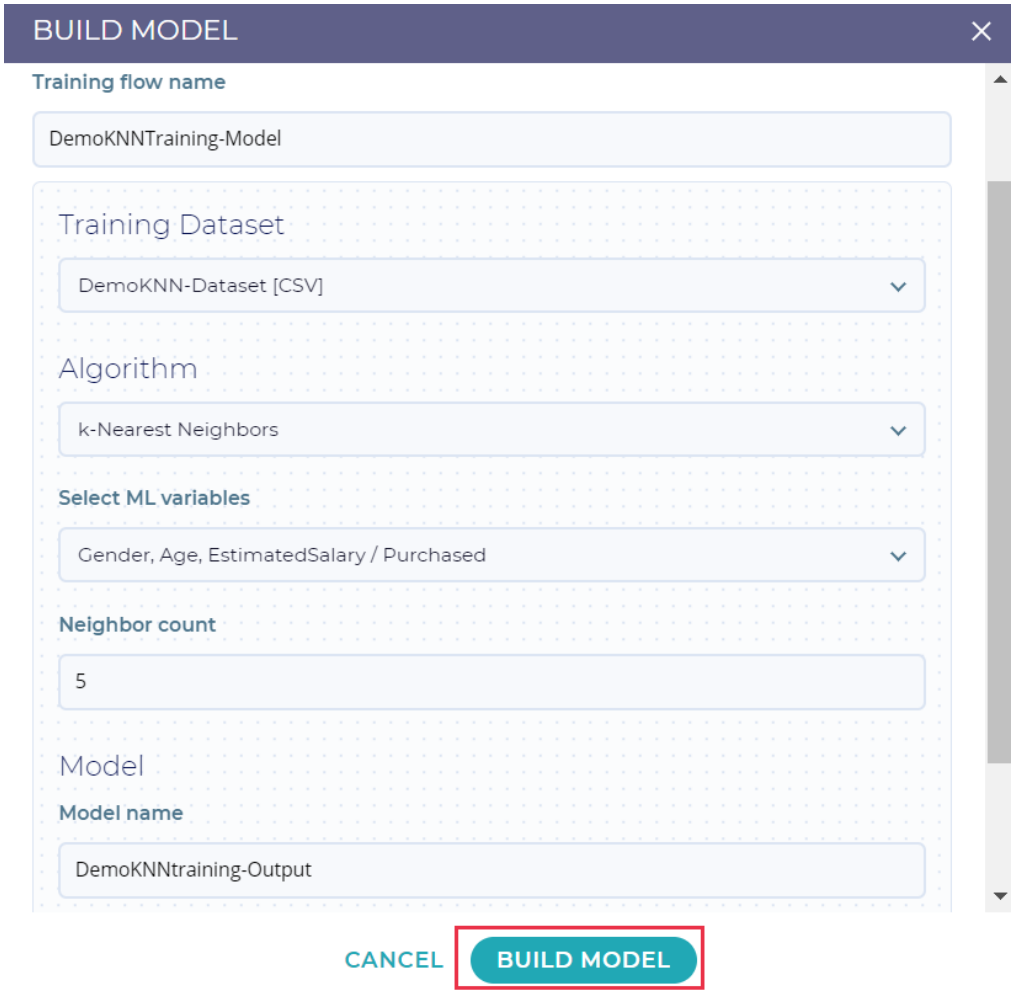

FIGURE 5: BUILD A MODEL

9. After the training model is built, you will see the model created under the Training flows. To run the training model, click the **RUN** button below the model.

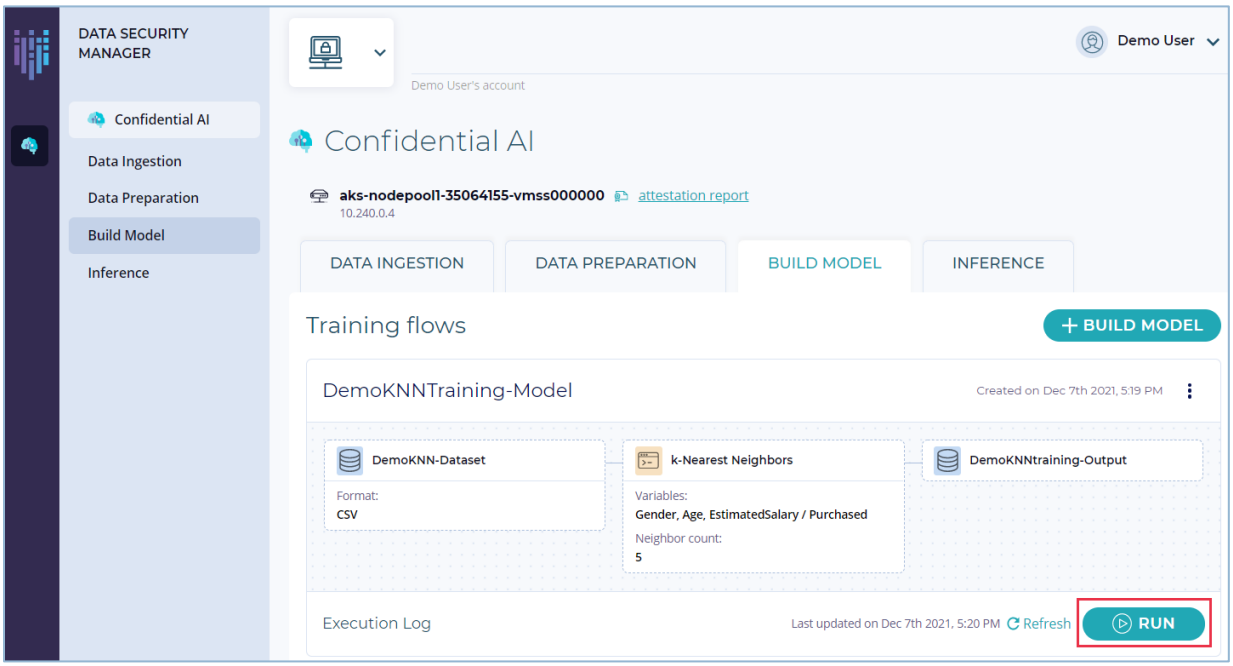

FIGURE 6: RUN TRAINING MODEL

10. If the model executed successfully, you would see the status of the execution under the **Execution Log**. Click the **Execution log** link to view the log details.

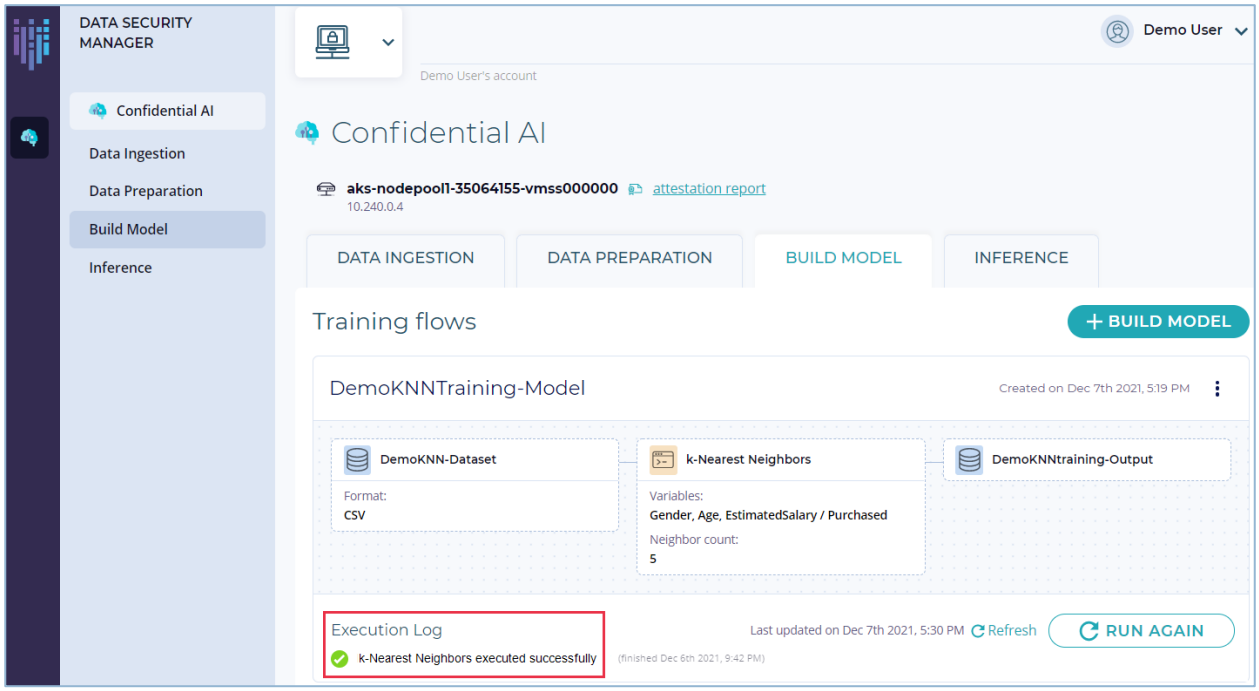

FIGURE 7: MODEL TRAINING SUCCESS

11. Click the download report icon to download the execution log report.

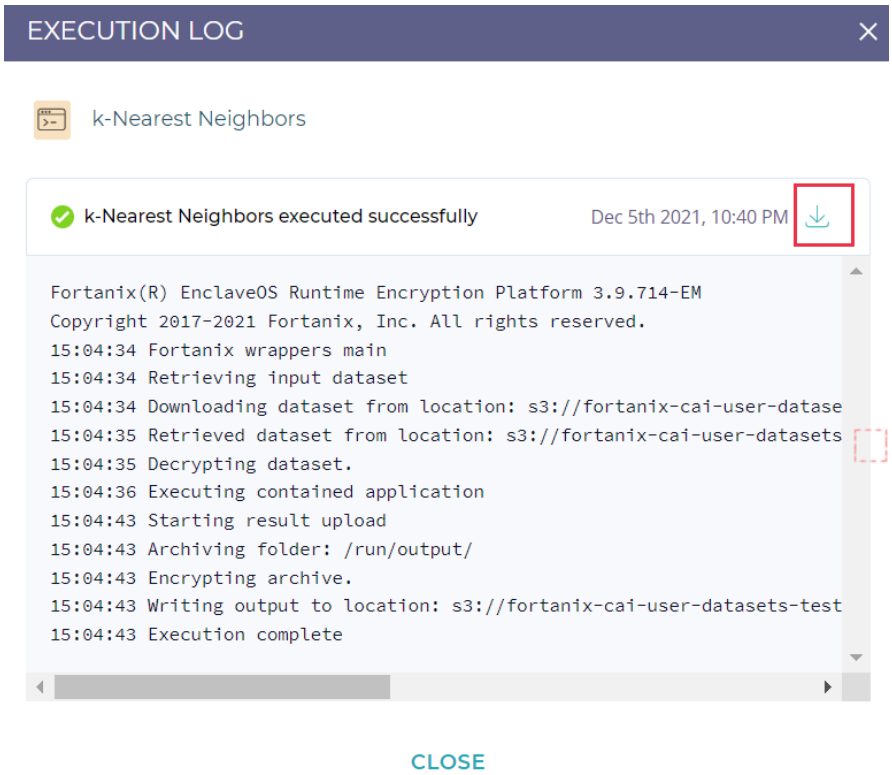

FIGURE 8: DOWNLOAD EXECUTION REPORT

12. After the execution completed successfully, the model is now trained and ready for inference where it will be passed through a machine learning model for output data prediction.

3.4 DATA INFERENCE

In this stage, the data is passed through a machine learning model to identify and predict the output from the data.

1. In the **INFERENCE** tab, click **BUILD INFERENCE** to predict the data output.
2. In the Build Inference form, enter the **Inference flow name**, that is, the name of the inference model.
3. In the **Select model** section, select **TRAINED,** and select the trained model that was built in the "build a model" stage from the drop-down.
4. In the **Select input dataset** field, select the input dataset you created in the first stage that you want to pass through a machine learning model.
5. In the **Select inference application** section, select the prediction algorithm.

6. Select the ML variables.
7. In the **Output Configuration** field, enter a name of the output dataset that will contain the predicted output.
8. The **Encrypt Dataset** option is selected by default to generate an encryption key and add an extra layer of protection to the output data. Copy or download the key to decrypt the output data for viewing.
9. Click **CREATE INFERENCE FLOW** to pass the data through a machine learning model and predict the output.

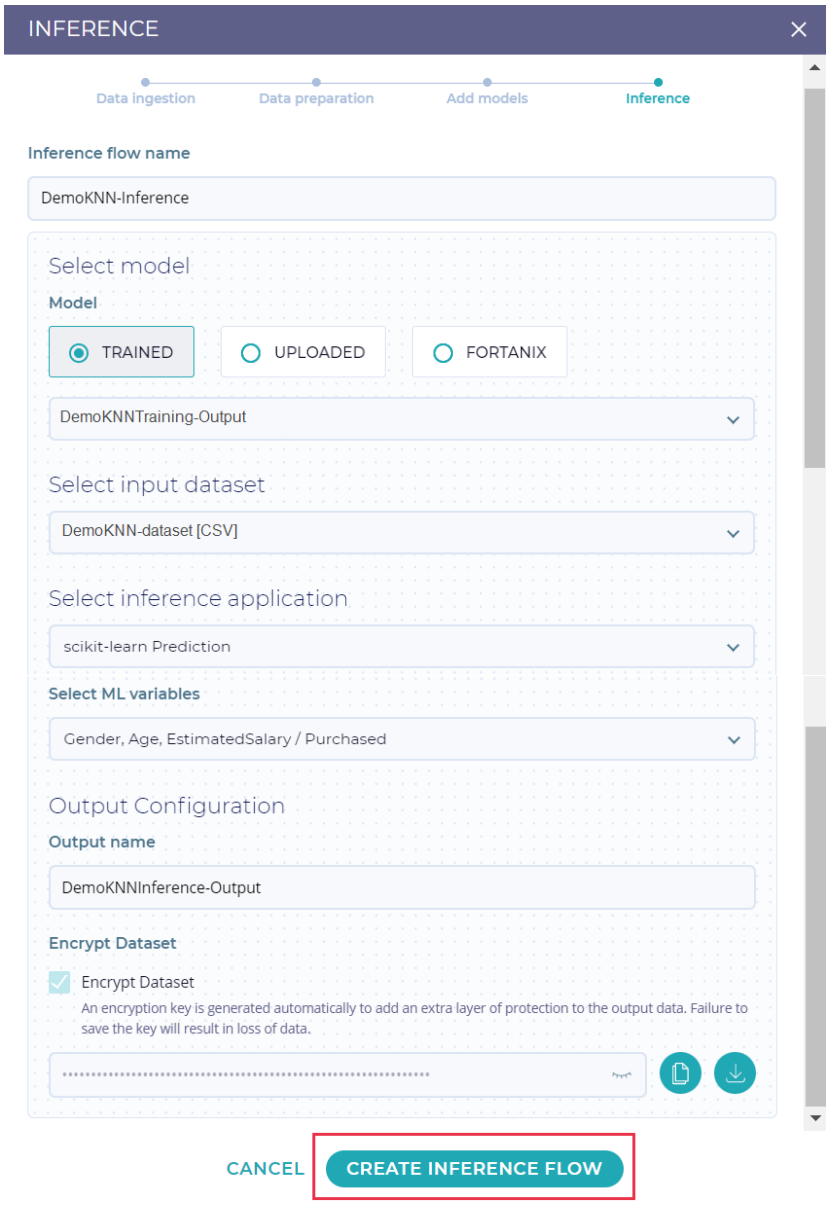

FIGURE 9: BUILD INFERENCE

10. The inference is successfully created. Click **RUN** below the inference workflow to run the model and predict the output.

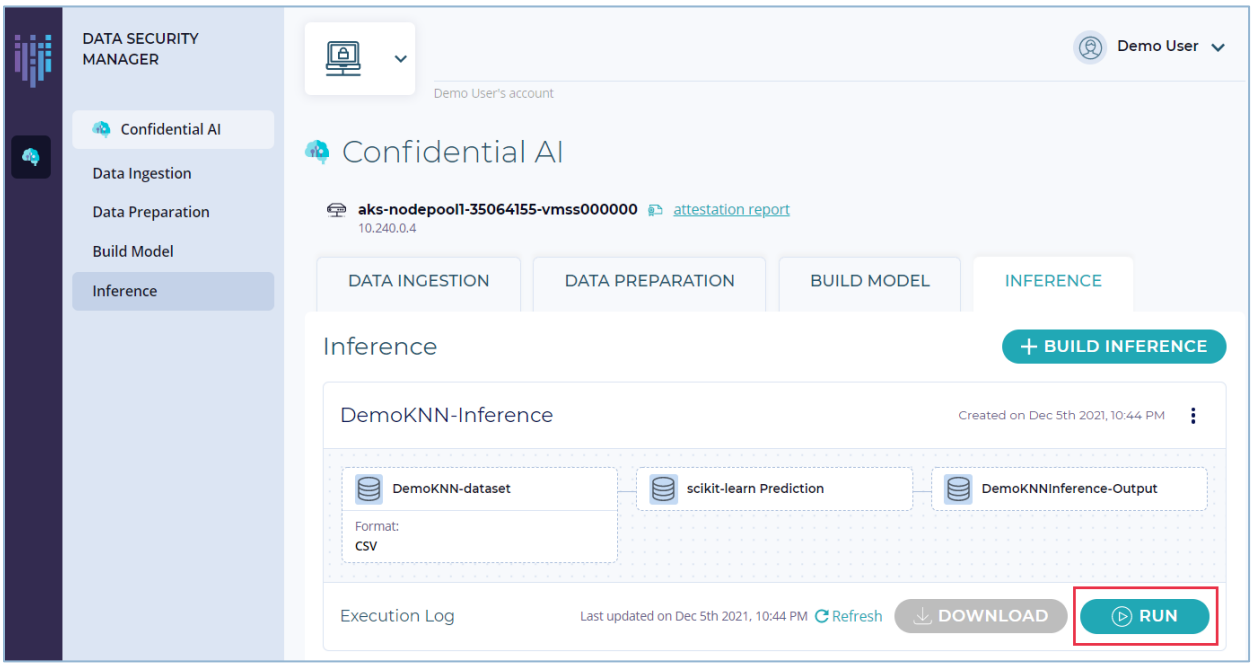

FIGURE 10: RUN INFERENCE

11. If the model executed successfully, you would see the status of the execution under the **Execution Log**. Click the Execution Log link to view the log details.

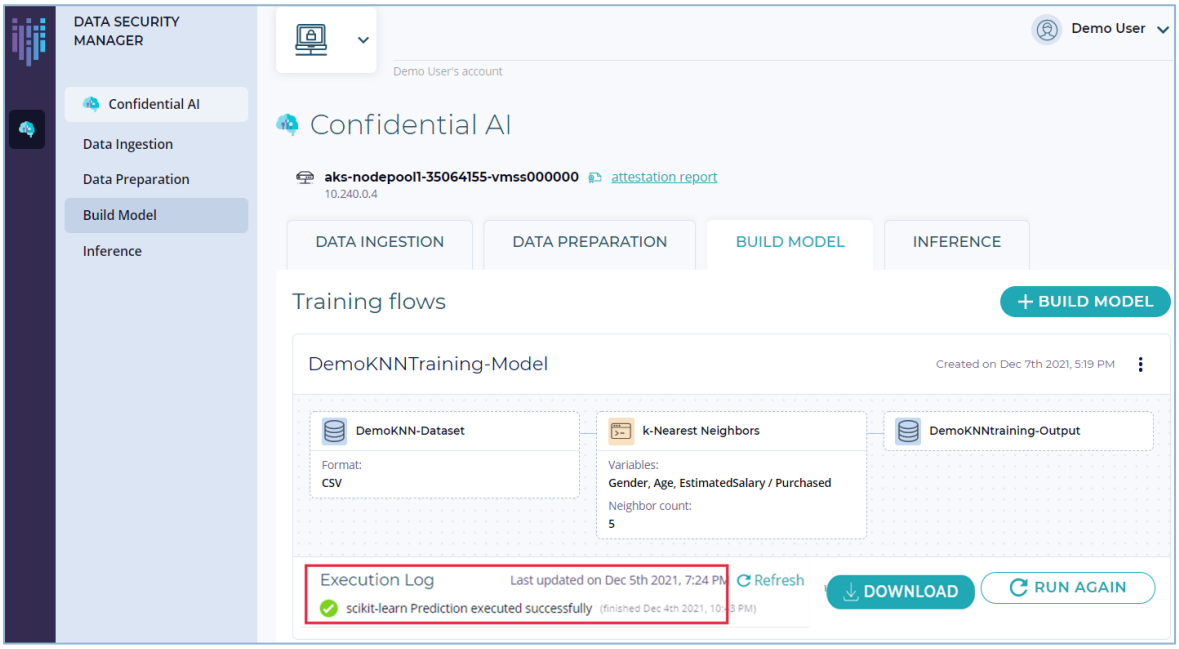

FIGURE 11: INFERENCE SUCCESS

12. After the execution is completed successfully, the output is now predicted and ready to be viewed. To view the output, click the **DOWNLOAD** button.

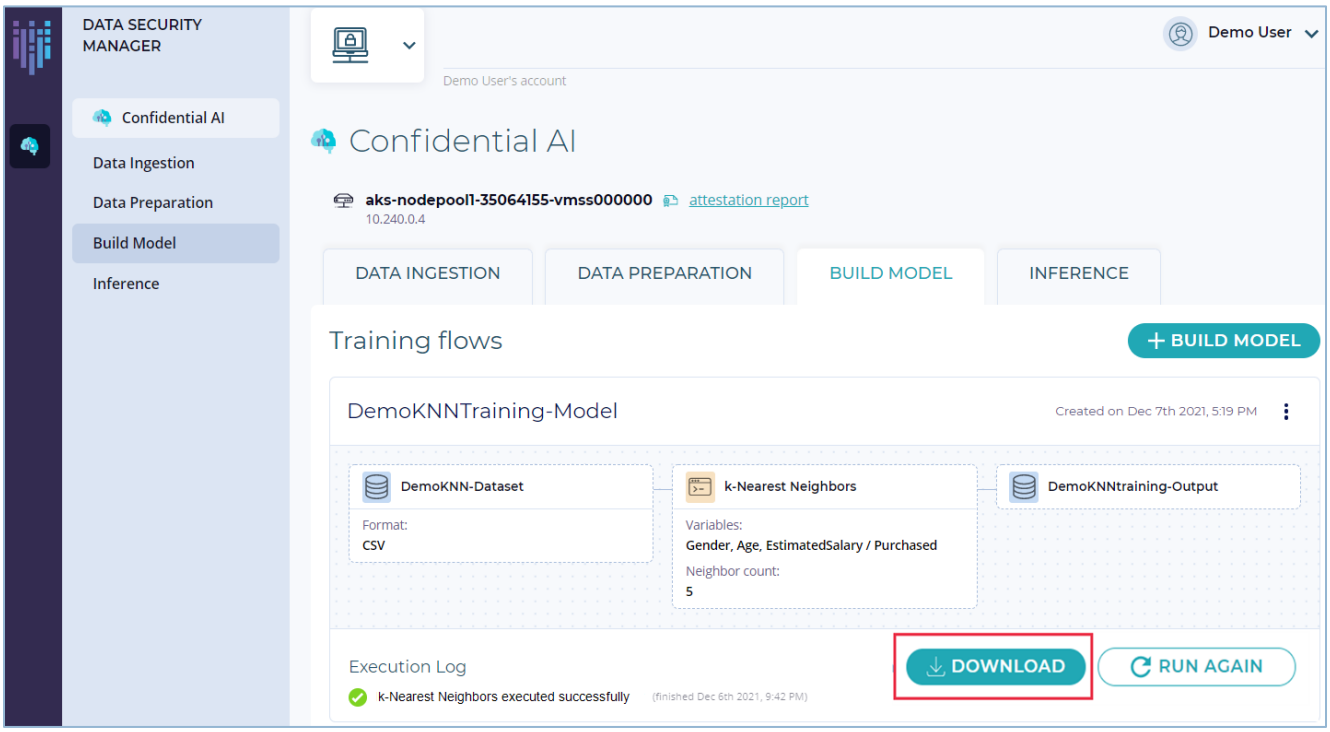

FIGURE 12: DOWNLOAD OUTPUT

13. In the DOWNLOAD dialog box, enter the **Encryption key** to decrypt the output.

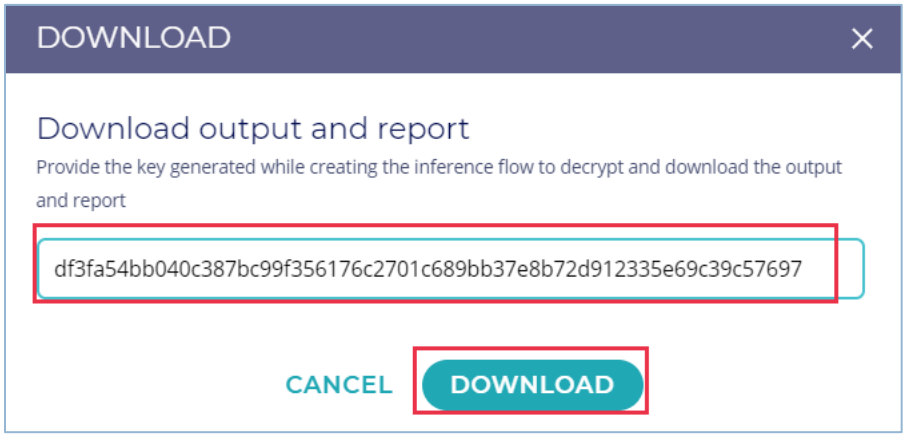

FIGURE 13: DECRYPT OUTPUT

14. A `*.tar.gz` file is generated on your local machine. Extract the contents of the file. A snapshot of the output appears as shown below.

User ID	Gender	Age	Purchased	EstimatedSalary
15624510	0	19	N	19000
15810944	0	35	N	20000
15668575	1	26	N	43000
15603246	1	27	N	57000
15804002	0	19	N	76000
15728773	0	27	N	58000
15598044	1	27	N	84000
15694829	1	32	Y	150000
15600575	0	25	N	33000
15727311	1	35	N	65000
15570769	1	26	N	80000
15606274	1	26	N	52000
15746139	0	20	N	86000
15704987	0	32	N	18000
15628972	0	18	N	82000
15697686	0	29	N	80000
15733883	0	47	Y	25000
15617482	0	45	N	26000
15704583	0	46	Y	28000
15621083	1	48	Y	29000
15649487	0	45	Y	22000
15736760	1	47	N	49000
15714658	0	48	N	41000
15599081	1	45	Y	22000
15705113	0	46	Y	23000
15631159	0	47	N	20000

FIGURE 14: SAMPLE OUTPUT

4.0 DOCUMENT INFORMATION

4.1 DOCUMENT LOCATION

The latest published version of this document is located at the URL:

https://support.fortanix.com/hc/en-us/articles/4412300629652-Classification-and-Learning-Inference-Using-K-Nearest-Neighbor-KNN-

4.2 DOCUMENT UPDATES

This document will typically be updated on a periodic review and update cycle.

For any urgent document updates, please send an email to: support@fortanix.com