



NIKO 101 Hands On Guide: **How to build a binary classification model?**

Last Updated: 2023.05.18





In this document, you will understand basic steps of building Binary Regression model on NIKO

The screenshot shows the NIKO dashboard interface. At the top right, a blue button labeled "Build model >" is circled in yellow. Below it, the "Usage" section displays three metrics for "This Month": 194.2K / 20M Rows trained, 0 / 3 Models deployed, and 20 / 10K Predictions made. To the right, a "Walkthrough" section contains a "Data preparation guide >" link. The main area is a "Model list" table with columns for Starred, Created Date, Name, Exp. Version, Drift Version, Problem Type, Status, and Actions. The table lists four models, including one with a "Training Failed" status.

STARRED	CREATED DATE	NAME	EXP. VERSION	DRIFT VERSION	PROBLEM TYPE	STATUS	ACTIONS
☆	2023-05-16 11:43:21	churn	Version 1.0	Version 1.0	Binary classification	Review	⋮
☆	2023-05-15 17:31:11	german_credit_test	Version 2.0	Version 1.0	Binary classification	Review	⋮
☆	2023-05-15 17:29:18	german_credit_test	Version 1.0	Version 1.0	Binary classification	Training Failed	⋮
★	2023-05-09 16:32:19	credit_scoring_p	Version 3.0	Version 1.0	Binary classification	Review	⋮

See what happens when you click it.



Before proceeding to the model building on NIKO, you have to prepare your **modeling dataset.**

If you have your **OWN DATASET** ready

Proceed next after checking your **dataset** qualifies all the data requirements

If you **DON'T HAVE** your dataset yet

Download sample **OPEN DATA** here.



Open dataset for CREDIT SCORING use case

This is a dataset containing credit card default data of Taiwanese customers and their demographic characteristics and previous loan history.

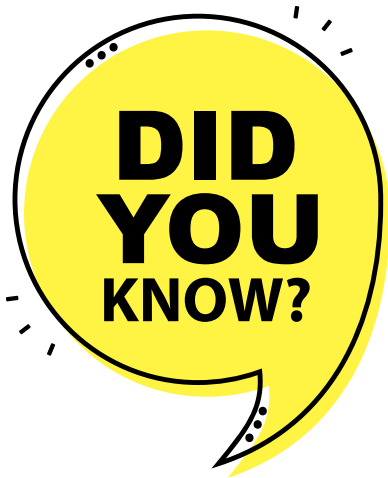
→ If you would like to know the details of the dataset and download it in xls format, [please click here.](#)

→ If you would like to download in NIKO-ready CSV format, [please click here.](#)

First column must contain unique ID values

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	ID	LIMIT_BAL	SEX	EDUCATI	MARRIA	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AM	BILL_AM	BILL_AM	BILL_AM	BILL_AM	BILL_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	PAY_AM	default pay
2	1	20000	2	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	0	0	0	0	1
3	2	120000	2	2	2	26	-1	2	0	0	0	2	2682	1725	2682	3272	3455	3261	0	1000	1000	1000	0	2000	1
4	3	90000	2	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1000	1000	1000	5000	0
5	4	50000	2	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1200	1100	1069	1000	0
6	5	50000	1	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10000	9000	689	679	0
7	6	50000	1	1	2	37	0	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	657	1000	1000	800	0
8	7	500000	1	1	2	29	0	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770	0
9	8	100000	2	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	0
10	9	140000	2	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	0
11	10	20000	1	3	2	35	-2	-2	-2	-2	-1	-1	0	0	0	0	13007	13912	0	0	0	13007	1122	0	0
12	11	200000	2	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	0
13	12	260000	2	1	2	51	-1	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640	0
14	13	630000	2	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0
15	14	70000	1	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3000	3000	1500	0	1
16	15	250000	1	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3000	3000	3000	3000	0
17	16	50000	2	3	3	23	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1100	1200	1300	1100	0
18	17	20000	1	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1500	0	1650	0	1
19	18	320000	1	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	0
20	19	360000	2	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
21	20	180000	2	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0
22	21	130000	2	3	2	39	0	0	0	0	0	-1	38358	27688	24489	20616	11802	930	3000	1537	1000	2000	930	33764	0
23	22	120000	2	2	1	39	-1	-1	-1	-1	-1	-1	316	316	316	0	632	316	316	316	0	632	316	0	1
24	23	70000	2	2	2	26	2	0	0	2	2	2	41087	42445	45020	44006	46905	46012	2007	3582	0	3601	0	1820	1
25	24	450000	2	1	1	40	-2	-2	-2	-2	-2	-2	5512	19420	1473	560	0	0	19428	1473	560	0	0	1128	1

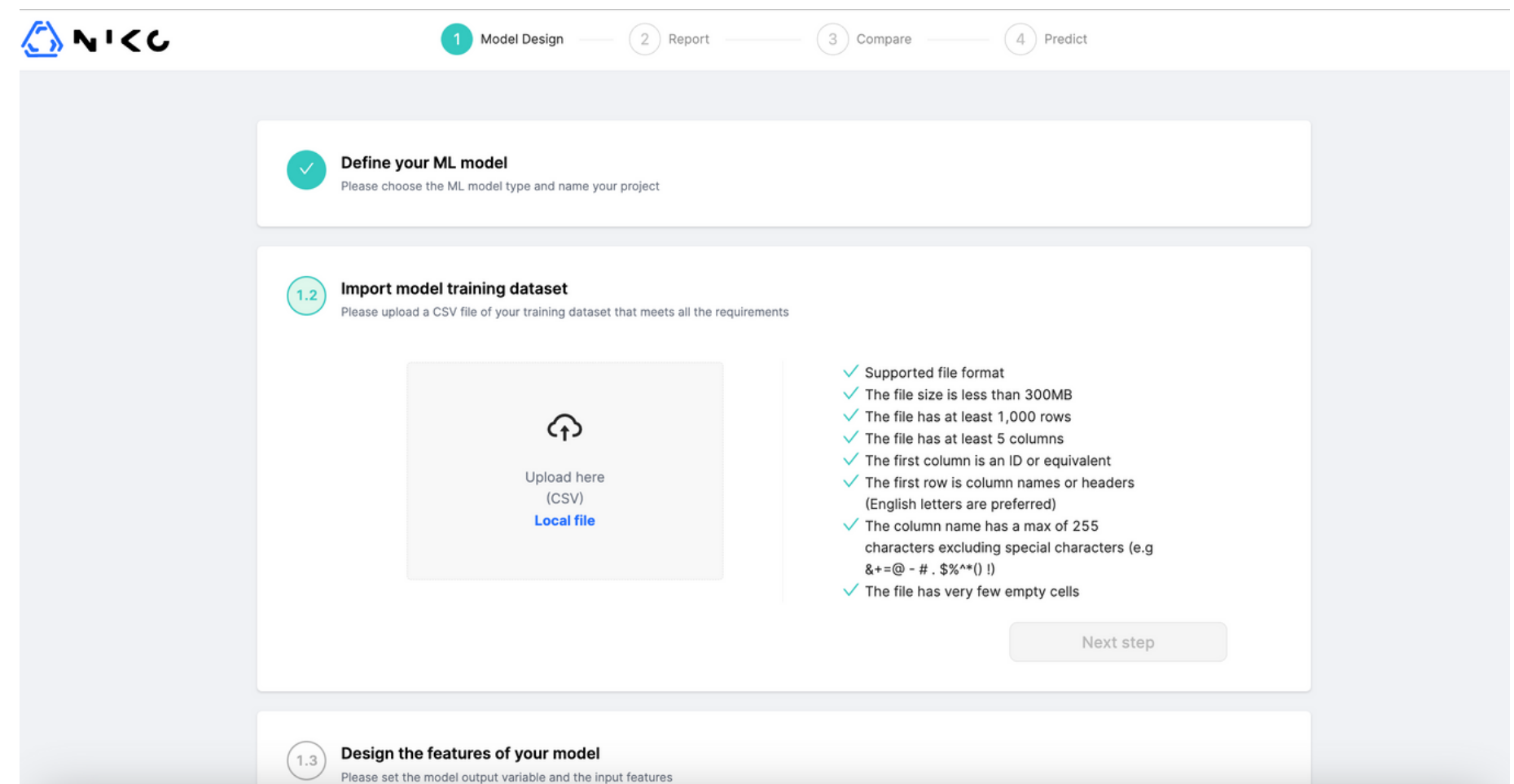
Our target value. The value 1 means "defaulted loan" and 0 means "non-default good loans"

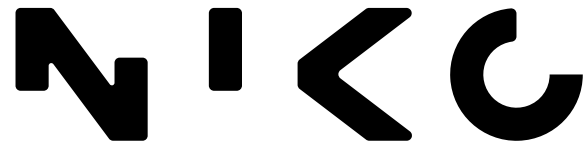


Here is the dataset requirements of NIKO. You can find the same list on the NIKO platform as well.

Data requirements:

- CSV file format
- File size must be less than 300MB
- The file must have at least 1,000 rows (data records)
- The file must have at least 5 columns
- The first column must contain unique IDs or equivalent value
- The first row should be column names
 - English letters are preferred
 - Has a maximum of 255 characters excluding special characters (e.g: &+=@ - # . \$%^*() !)





Build a binary classification model: STEP#1

Click here!

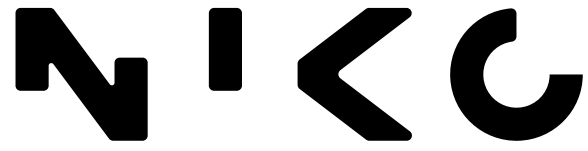
The screenshot shows the NIKE ML platform dashboard. At the top right, a blue button labeled "Build model >" is circled in yellow. The dashboard is divided into several sections:

- Usage:** A section with three cards showing metrics for "This Month":
 - Rows trained: 194.2K / 20M
 - Models deployed: 0 / 3
 - Predictions made: 20 / 10K
- Walkthrough:** A section with a button for "Data preparation guide >".
- Model list:** A table listing models with various filters and columns.

Model list filters: Date, Model name, Experiment Version, Drift Version, By problem type, By status, My Models, By Star.

STARRED	CREATED DATE	NAME	EXP.VERSION	DRIFT VERSION	PROBLEM TYPE	STATUS	ACTIONS
☆	2023-05-16 11:43:21	churn	Version 1.0	Version 1.0	Binary classification	Review	⋮
☆	2023-05-15 17:31:11	german_credit_test	Version 2.0	Version 1.0	Binary classification	Review	⋮
☆	2023-05-15 17:29:18	german_credit_test	Version 1.0	Version 1.0	Binary classification	Training Failed	⋮
★	2023-05-09 16:32:19	credit_scoring_p	Version 3.0	Version 1.0	Binary classification	Review	⋮

User Profile: Sarangerel



Build a binary classification model: STEP#2

1.1 Define your ML model
Please choose the ML model type and name your project

Binary classification Version 1.0

Drift Version 1.0

1.2 Import model training dataset
Please upload a CSV file of your training dataset that meets all the requirements

1.3 Design the features of your model
Please set the model output variable and the input features

1.4 Evaluate the quality of your training dataset
Check the data quality score and correlation matrix to ensure your dataset is suitable for model building

Choose "binary classification" as model type and name your model.



NICK

Build a binary classification model: STEP#3



1.2

Import model training dataset

Please upload a CSV file of your training dataset that meets all the requirements

default of credit card cli...

replace

- ✓ Supported file format
- ✓ The file size is less than 300MB
- ✓ The file has at least 1,000 rows
- ✓ The file has at least 5 columns
- ✓ The first column is an ID or equivalent
- ✓ The first row is column names or headers (English letters are preferred)
- ✓ The column name has a max of 255 characters excluding special characters (e.g &+=@ - # . \$%^*() !)
- ✓ The file has very few empty cells

Next step

1.3

Design the features of your model

Please set the model output variable and the input features

1.4

Evaluate the quality of your training dataset

Check the data quality score and correlation matrix to ensure your dataset is suitable for model building

Upload your data here and make sure it clears all the dataset requirements.



NIKO

Build a binary classification model: STEP#4

1 Model Design

1.3 Design the features of your model
Please set the model output variable and the input features

Target column: default payment next ...

Positive target data: 1

Negative target data: 0

Columns to be used in model building from uploaded data sample: SEX EDUCATION MARRIAGE AGE PAY_0 PAY_2 PAY_...

2	120000	2	2	1	24	2	2	-1
3	90000	2	2	2	34	0	0	0
4	50000	2	2	1	37	0	0	0
5	50000	1	2	1	57	-1	0	-1

Set your target value to predict here.
Please choose a column that has only two values as target variable.
Make sure to set "Positive" and "Negative" labels as well.



N I < C

Build a binary classification model: STEP#5



1 Model Design 2 Report 3 Compare 4 Predict

1.4 Evaluate the quality of your training dataset
Check the data quality score and correlation matrix to ensure your dataset is suitable for model building

1. Missing values	
SEX	0.00%(0/30000)
MARRIAGE	0.00%(0/30000)
EDUCATION	0.00%(0/30000)
ID	0.00%(0/30000)
AGE	0.00%(0/30000)
PAY_AMT4	0.00%(0/30000)
PAY_AMT3	0.00%(0/30000)
PAY_AMT2	0.00%(0/30000)
TOTAL	0.00%(0/750000)

83%
good

DATA QUALITY

2. Class imbalance of target column	
0 - 78%	1 - 22%
(23364/30000)	(6636/30000)

3. Duplicated values
Duplicated values 0.19% (56/30000)

Make sure your data quality is not too bad

1.5 Alter the split ratio of train and test datasets
The default value is set at 30%. The recommended range is between 10%-50%

Next step



N I K O

Build a binary classification model: STEP#6



1 Model Design — 2 Report — 3 Compare — 4 Predict

✓ **Import model training dataset**
Please upload a CSV file of your training dataset that meets all the requirements

✓ **Design the features of your model**
Please set the model output variable and the input features

✓ **Evaluate the quality of your training dataset**
Check the data quality score and correlation matrix to ensure your dataset is suitable for model building

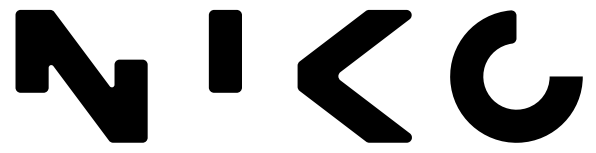
1.5 **Alter the split ratio of train and test datasets**
The default value is set at 30%. The recommended range is between 10%-50%



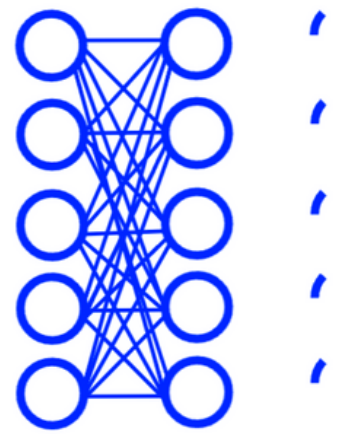
Since it's your first model, leave the split ratio at the default value of 30%.

Build model

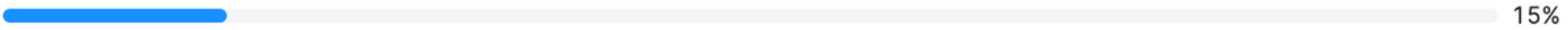
Click here!



Build a binary classification model: STEP#7



TRAINING IS STARTED



Enjoy your coffee, while we do the model building!

NIKO is building your model. Please enjoy your





Build a binary classification model: STEP#8

1 Report 2 Compare 3 Predict [Next step](#)

Model name: my_first_model

☆ Experiment Version 1.0

Drift Version 1.0

0.81
model accuracy

Problem type:
Binary classification

Model method:
Random Forest

Total record count: 30000

Self-validation count: 9000

Marked as positive: 6636 (22%)

Marked as negative: 23364 (78%)

Build time: 17.18 sec

Created date: 2023-05-19 22:37:31

Deploy [Undeployed](#)

[Report](#) Prediction history

[Feature importance & statistics](#) [Details](#) [Decision threshold tuning](#)

FEATURE IMPORTANCE

Feature	Importance
PAY_0:	0.09
AGE:	0.07
BILL_AMT1:	0.06
LIMIT_BAL:	0.06
BILL_AMT2:	0.05
PAY_2:	0.05
BILL_AMT3:	0.05
PAY_AMT1:	0.05
BILL_AMT5:	0.05
BILL_AMT6:	0.05

[See all](#)

FEATURE STATISTICS

PAY_0

Data types

Type	Count	Percentage
Valid	30000	100%
Mismatched	0	0%
Missing	0	0%

Quantiles

Quantile	Value
Min	-2.00
25%	-1.00
50%	0.00
75%	0.00
Max	8.00

Mean: -0.02
Std. Deviation: 1.12

Once you got this screen, TA-DA!!!! you've built your first model with NIKO.

