

# PDC Project: Terminal Output Report

Date: 2026-01-04 04:30

>>> Running cpu\_train.py ...

== Preprocessing Output ==

Full dataframe shape: (5110, 12)

Train/Test distribution:

split

train 4088

test 1022

Name: count, dtype: int64

First 5 rows:

	gender	age	hypertension	...	smoking_status	stroke	split
0	-0.840344	0.205661	-0.327962	...	0.589225	1.0	train
1	1.187594	-1.254901	-0.327962	...	0.589225	1.0	train
2	-0.840344	1.046590	-0.327962	...	0.589225	1.0	train
3	1.187594	0.028623	-0.327962	...	1.522127	1.0	train
4	1.187594	-1.299160	-0.327962	...	-1.276579	1.0	train

[5 rows x 12 columns]

Preprocessed data saved to: C:\Users\RBTG\pdc\results\preprocessed\_data.csv

== Preprocessed Data Snapshot ==

X\_train (first 5 rows):

	gender	age	...	bmi	smoking_status
0	-0.840344	0.205661	...	0.538479	0.589225
1	1.187594	-1.254901	...	-1.020641	0.589225
2	-0.840344	1.046590	...	-0.518114	0.589225
3	1.187594	0.028623	...	-0.531000	1.522127
4	1.187594	-1.299160	...	0.345200	-1.276579

[5 rows x 10 columns]

y\_train distribution:

stroke

0 3889

1 199

Name: count, dtype: int64

X\_test (first 5 rows):

	gender	age	...	bmi	smoking_status
0	1.187594	0.869552	...	0.757529	0.589225
1	-0.840344	-0.015636	...	0.564250	0.589225
2	-0.840344	-0.900825	...	-0.028473	1.522127
3	-0.840344	-0.989344	...	-0.518114	0.589225
4	1.187594	1.046590	...	-0.337720	-0.343677

[5 rows x 10 columns]

```
y_test distribution:  
stroke  
0 972  
1 50  
Name: count, dtype: int64
```

[SUCCESS] Preprocessed data saved to: C:\Users\RBTG\pdc\results\preprocessed\_data.csv

Training on CPU...

```
[LightGBM] [Info] Number of positive: 199, number of negative: 3889  
[LightGBM] [Info] Auto-choosing row-wise multi-threading, the overhead of testing was 0.000114 seconds.  
You can set `force_row_wise=true` to remove the overhead.  
And if memory is not enough, you can set `force_col_wise=true`.  
[LightGBM] [Info] Total Bins 637  
[LightGBM] [Info] Number of data points in the train set: 4088, number of used features: 10  
[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.048679 -> initscore=-2.972603  
[LightGBM] [Info] Start training from score -2.972603
```

[SUCCESS] CPU Training Complete

```
{'Model': 'LightGBM', 'Device': 'CPU', 'Training Time (s)': 1.2930281162261963, 'Accuracy (%)': 94.42270058708415, 'F1 Score': 0.12307692307692308, 'CPU': 'AMD64 Family 25 Model 33 Stepping 0, AuthenticAMD'}  
Results saved to: C:\Users\RBTG\pdc\results\cpu_results.json  
[SUCCESS] CPU model saved to: C:\Users\RBTG\pdc\results\cpu_model.pkl
```

[SUCCESS] cpu\_train.py completed successfully.

>>> Running gpu\_train.py ...

==== Preprocessing Output ===

Full dataframe shape: (5110, 12)

Train/Test distribution:

```
split  
train 4088  
test 1022
```

Name: count, dtype: int64

First 5 rows:

	gender	age	hypertension	...	smoking_status	stroke	split
0	-0.840344	0.205661	-0.327962	...	0.589225	1.0	train
1	1.187594	-1.254901	-0.327962	...	0.589225	1.0	train
2	-0.840344	1.046590	-0.327962	...	0.589225	1.0	train
3	1.187594	0.028623	-0.327962	...	1.522127	1.0	train
4	1.187594	-1.299160	-0.327962	...	-1.276579	1.0	train

[5 rows x 12 columns]

Preprocessed data saved to: C:\Users\RBTG\pdc\results\preprocessed\_data.csv

==== Preprocessed Data Snapshot ===

X\_train (first 5 rows):

	gender	age	...	bmi	smoking_status
0	-0.840344	0.205661	...	0.538479	0.589225
1	1.187594	-1.254901	...	-1.020641	0.589225
2	-0.840344	1.046590	...	-0.518114	0.589225

```
3 1.187594 0.028623 ... -0.531000      1.522127
4 1.187594 -1.299160 ...  0.345200     -1.276579
```

[5 rows x 10 columns]

y\_train distribution:

```
stroke
0    3889
1     199
```

Name: count, dtype: int64

X\_test (first 5 rows):

	gender	age	bmi	smoking_status	
0	1.187594	0.869552	...	0.757529	0.589225
1	-0.840344	-0.015636	...	0.564250	0.589225
2	-0.840344	-0.900825	...	-0.028473	1.522127
3	-0.840344	-0.989344	...	-0.518114	0.589225
4	1.187594	1.046590	...	-0.337720	-0.343677

[5 rows x 10 columns]

y\_test distribution:

```
stroke
0    972
1     50
```

Name: count, dtype: int64

[SUCCESS] Preprocessed data saved to: C:\Users\RBTG\pdc\results\preprocessed\_data.csv

Training on GPU...

[LightGBM] [Info] Number of positive: 199, number of negative: 3889

[LightGBM] [Info] This is the GPU trainer!!

[LightGBM] [Info] Total Bins 637

[LightGBM] [Info] Number of data points in the train set: 4088, number of used features: 10

[LightGBM] [Info] Using requested OpenCL platform 0 device 0

[LightGBM] [Info] Using GPU Device: gfx1031, Vendor: Advanced Micro Devices, Inc.

[LightGBM] [Info] Compiling OpenCL Kernel with 256 bins...

[LightGBM] [Info] GPU programs have been built

[LightGBM] [Info] Size of histogram bin entry: 8

[LightGBM] [Info] 8 dense feature groups (0.03 MB) transferred to GPU in 0.000897 secs. 1 sparse feature groups

[LightGBM] [Info] [binary:BoostFromScore]: pavg=0.048679 -> initscore=-2.972603

[LightGBM] [Info] Start training from score -2.972603

[SUCCESS] GPU Training Complete

{'Model': 'LightGBM', 'Device': 'GPU', 'Training Time (s)': 2.776308059692383, 'Accuracy (%)': 94.52054794520548, 'F1 Score': 0.125, 'GPU': 'gfx1031 | Vendor: Advanced Micro Devices, Inc.'}

Results saved to: C:\Users\RBTG\pdc\results\gpu\_results.json

[SUCCESS] GPU model saved to: C:\Users\RBTG\pdc\results\gpu\_model.pkl

[SUCCESS] gpu\_train.py completed successfully.

>>> Running plot\_results.py ...

All plots saved in results folder.

[SUCCESS] plot\_results.py completed successfully.

>>> Running cpu\_confusion\_matrix.py ...

[SUCCESS] CPU confusion matrix saved to: C:\Users\RBTG\pdc\results\cpu\_confusion\_matrix.png

[SUCCESS] cpu\_confusion\_matrix.py completed successfully.

>>> Running gpu\_confusion\_matrix.py ...

GPU confusion matrix saved at: C:\Users\RBTG\pdc\results\gpu\_confusion\_matrix.png

[SUCCESS] gpu\_confusion\_matrix.py completed successfully.

>>> Running generate\_report.py ...

[SUCCESS] PDF report generated: C:\Users\RBTG\pdc\results\PDC\_Project\_Report.pdf

[SUCCESS] generate\_report.py completed successfully.