# Feature Selection Report

# Best Model: CatBoost (Run 7)

**Reasons:**

# Strong top variable score (0.591).

# Balanced and meaningful variable selections including transaction time windows and state-level aggregations.

# CatBoost is also known to perform better on tabular data with default settings compared to LGBM/RandomForest unless those are finely tuned.

# Inferences from the Runs

# 1. Consistently High Filter Scores

# Across all models, the variable Cardnum\_unique\_count\_for\_card\_state\_1 appears at the top with a filter score of ~0.519. This suggests it's a strong and stable feature across multiple model types.

# 2. Top-Performing Model by Individual Variable Score

# In Run 7 (CatBoost), Cardnum\_total\_1 achieves the highest single filter score: 0.591, which is higher than any score from LGBM or Random Forest runs. This is a strong indication of CatBoost's effectiveness in identifying informative variables.

# 3. Model Sophistication and Feature Selection

# CatBoost can handle categorical variables natively and often performs well with structured data like transaction logs.

# Random Forest identified more diverse features but didn’t surpass the CatBoost filter scores.

# First Run – Baseline Model (LGBM)

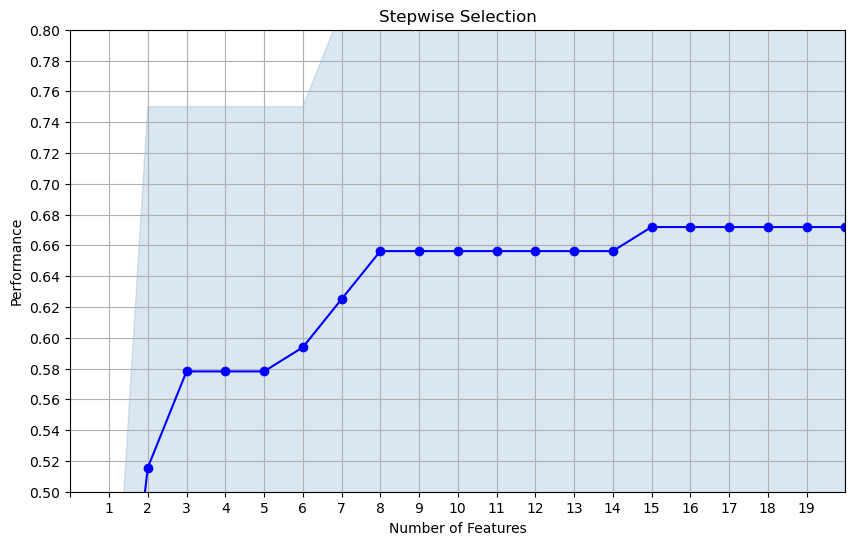
Model: LGBMClassifier

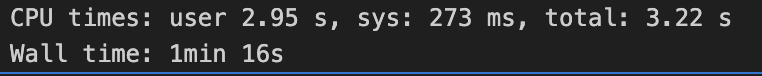
Wrapper Complexity: n\_estimators = 5

Feature Selection: Forward

num\_out\_of\_filter: 200

Wrapper Parameters: n\_estimators = 20, num\_leaves = 4





List of top 20 variables:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **wrapper order** | **variable** | **filter score** |
| 0 | 1 | U\*\_cardnum | 0.436571 |
| 1 | 2 | Cardnum\_total\_3 | 0.484935 |
| 2 | 3 | Cardnum\_max\_14 | 0.371055 |
| 3 | 4 | Cardnum\_total\_7 | 0.515563 |
| 4 | 5 | Cardnum\_count\_14 | 0.464369 |
| 5 | 6 | Cardnum\_count\_7 | 0.457883 |
| 6 | 7 | card\_zip\_day\_since | 0.486312 |
| 7 | 8 | Cardnum\_total\_14 | 0.400785 |
| 8 | 9 | Cardnum\_count\_30 | 0.437424 |
| 9 | 10 | Cardnum\_count\_60 | 0.437424 |
| 10 | 11 | Cardnum\_total\_0 | 0.391105 |
| 11 | 12 | Cardnum\_day\_since | 0.399146 |
| 12 | 13 | Cardnum\_actual/max\_0 | 0.406090 |
| 13 | 14 | Cardnum\_actual/toal\_7 | 0.361752 |
| 14 | 15 | card\_state\_day\_since | 0.263324 |
| 15 | 16 | Cardnum\_max\_3 | 0.357243 |
| 16 | 17 | Dow\_Risk | 0.344416 |
| 17 | 18 | Cardnum\_actual/med\_1 | 0.334802 |
| 18 | 19 | Cardnum\_actual/med\_0 | 0.333421 |
| 19 | 20 | Cardnum\_actual/med\_3 | 0.323105 |

# Second Run – LGBM (Backward)

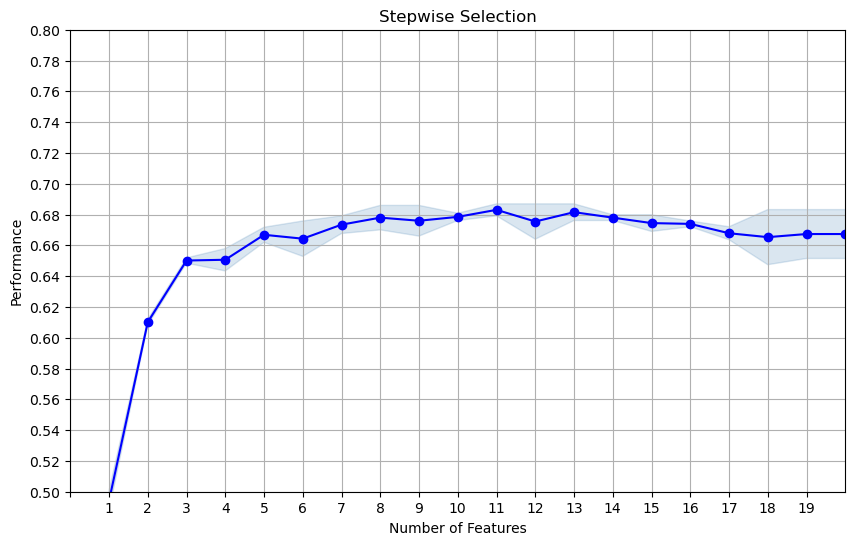
Model: LGBMClassifier

Wrapper Complexity: n\_estimators = 5

Feature Selection: Backward

num\_out\_of\_filter: 200

Wrapper Parameters: n\_estimators = 20, num\_leaves = 4





List of top 20 variables:

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | card\_zip\_total\_amount\_1\_by\_60 | 0.3458850828428253 |
| 3 | Card\_dow\_vdratio\_0by60 | 0.5084477665556563 |
| 4 | Cardnum\_total\_7 | 0.5159873870717485 |
| 5 | Cardnum\_total\_30 | 0.3691095461455587 |
| 6 | Cardnum\_unique\_count\_for\_card\_zip\_60 | 0.3014462545628256 |
| 7 | card\_merch\_vdratio\_1by30 | 0.2733331094751868 |
| 8 | card\_merch\_vdratio\_0by30 | 0.2733331094751868 |
| 9 | card\_merch\_day\_since | 0.2733706479521256 |
| 10 | card\_merch\_vdratio\_0by7 | 0.2733706479521257 |
| 11 | card\_merch\_vdratio\_1by14 | 0.2733706479521257 |
| 12 | card\_merch\_vdratio\_1by7 | 0.2733706479521257 |
| 13 | card\_merch\_vdratio\_0by14 | 0.2733706479521257 |
| 14 | card\_zip\_count\_1\_by\_30\_sq | 0.2737839069856730 |
| 15 | Card\_dow\_unique\_count\_for\_state\_des\_60 | 0.2747260118836998 |
| 16 | Card\_dow\_unique\_count\_for\_Card\_Merchdesc\_60 | 0.2747760631862850 |
| 17 | card\_zip\_vdratio\_0by60 | 0.2762725401131096 |
| 18 | card\_zip\_vdratio\_1by60 | 0.2762725401131096 |
| 19 | card\_zip\_vdratio\_0by30 | 0.2764352068465113 |
| 20 | card\_zip\_vdratio\_1by30 | 0.2764352068465113 |

# Third Run – Random Forest

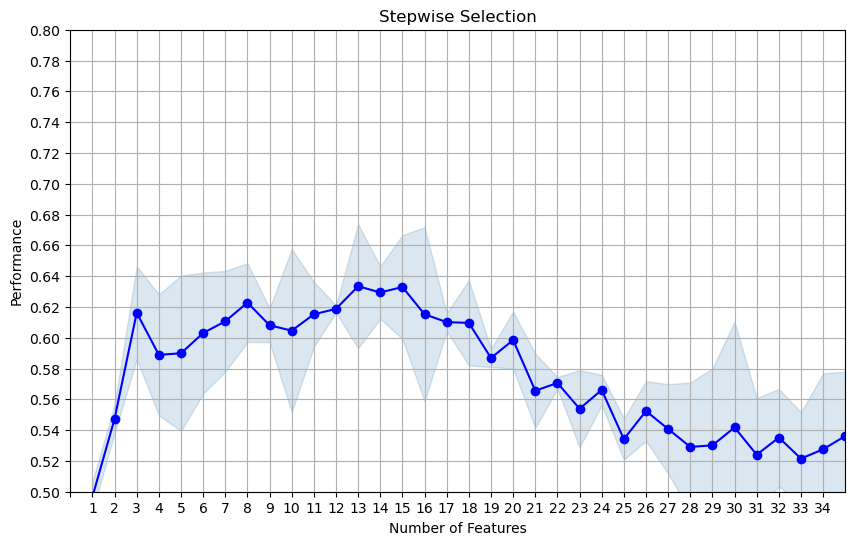
Model: RandomForestClassifier

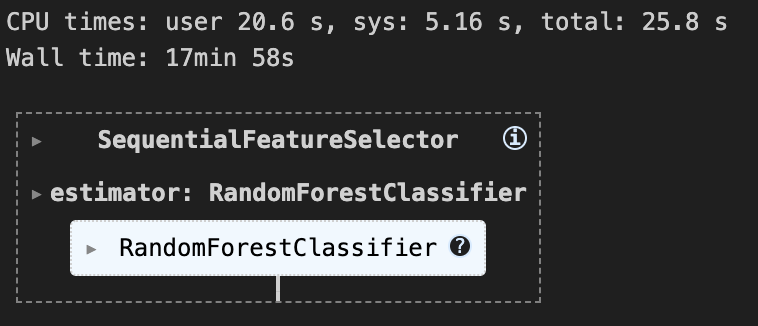
Wrapper Complexity: n\_estimators = 5

Feature Selection: Forward

num\_out\_of\_filter: 400

num\_out\_of\_wrapper: 35





List of top 20 variables:

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | Card\_Merchnum\_desc\_vdratio\_1by60 | 0.2700875169200329 |
| 3 | Card\_Merchnum\_desc\_total\_30 | 0.2543774394070377 |
| 4 | card\_merch\_total\_amount\_0\_by\_60 | 0.2357413374024831 |
| 5 | Card\_Merchnum\_Zip\_count\_1\_by\_30 | 0.2450144626089843 |
| 6 | Card\_Merchdesc\_Zip\_vdratio\_0by60 | 0.2777984459387328 |
| 7 | card\_merch\_count\_1\_by\_60\_sq | 0.3045681000240437 |
| 8 | Card\_Merchnum\_Zip\_vdratio\_1by60 | 0.2725814592869083 |
| 9 | Card\_Merchnum\_Zip\_total\_amount\_1\_by\_60 | 0.3058974499494893 |
| 10 | Card\_Merchnum\_Zip\_count\_1\_by\_60 | 0.3039107363528629 |
| 11 | Card\_Merchnum\_desc\_vdratio\_0by60 | 0.2700875169200329 |
| 12 | Card\_Merchnum\_Zip\_count\_0\_by\_60\_sq | 0.2359292135198072 |
| 13 | Card\_dow\_unique\_count\_for\_Card\_Merchdesc\_1 | 0.4868859821823698 |
| 14 | Card\_Merchdesc\_Zip\_vdratio\_1by60 | 0.2777984459387328 |
| 15 | Card\_Merchdesc\_total\_3 | 0.2598959439754863 |
| 16 | Card\_Merchdesc\_State\_vdratio\_0by60 | 0.2788128654655833 |
| 17 | Card\_Merchnum\_Zip\_total\_amount\_1\_by\_30 | 0.2493830163259113 |
| 18 | Cardnum\_count\_0\_by\_14 | 0.2444695243003826 |
| 19 | Merchdesc\_Zip\_total\_3 | 0.2521105462212693 |
| 20 | Card\_dow\_total\_amount\_0\_by\_60 | 0.3268602849566375 |

# Fourth Run – Random Forest

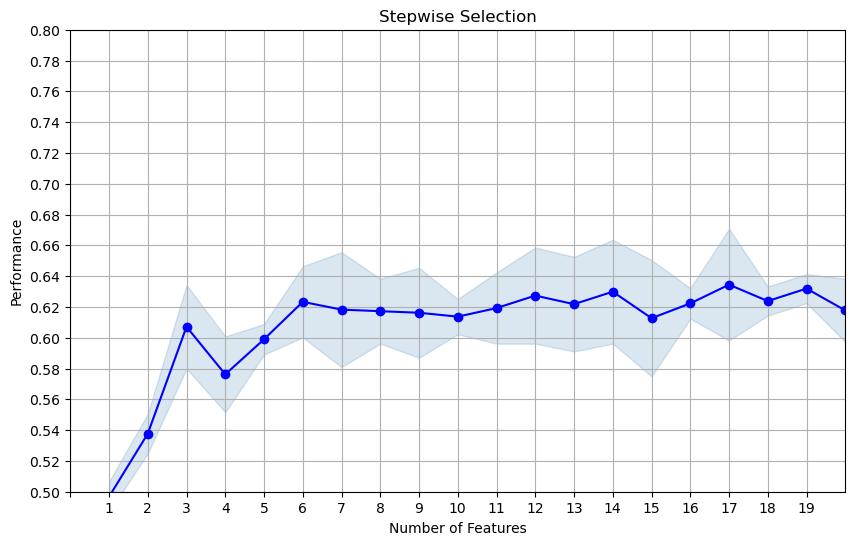
Model: RandomForestClassifier

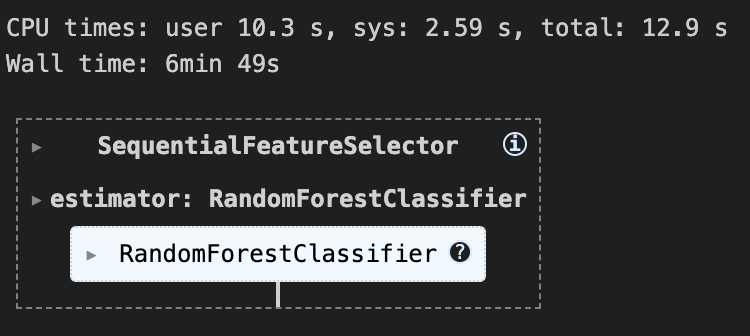
Wrapper Complexity: n\_estimators = 5

Feature Selection: Forward

num\_out\_of\_filter: 400

num\_out\_of\_wrapper: 20





**List of top 20 variables:**

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | Card\_Merchdesc\_vdratio\_1by60 | 0.2791824341468994 |
| 3 | Card\_Merchnum\_desc\_total\_60 | 0.2407418173247832 |
| 4 | Card\_Merchnum\_State\_vdratio\_0by60 | 0.2729702375314444 |
| 5 | Card\_Merchdesc\_Zip\_vdratio\_1by60 | 0.2777984459387328 |
| 6 | Card\_Merchnum\_State\_count\_1\_by\_60 | 0.3043929204649957 |
| 7 | Card\_Merchnum\_State\_count\_1\_by\_30 | 0.2450211593465581 |
| 8 | Card\_Merchdesc\_vdratio\_0by60 | 0.2791824341468994 |
| 9 | Card\_Merchnum\_State\_total\_amount\_0\_by\_60 | 0.2355661578434351 |
| 10 | card\_merch\_vdratio\_1by60 | 0.2732330068700165 |
| 11 | card\_merch\_total\_amount\_1\_by\_30 | 0.2489092902503547 |
| 12 | Merchnum\_total\_1 | 0.2431599668140857 |
| 13 | Card\_Merchnum\_Zip\_total\_amount\_1\_by\_60 | 0.3058974499494893 |
| 14 | Card\_Merchnum\_Zip\_count\_1\_by\_30 | 0.2450144626089843 |
| 15 | card\_merch\_count\_1\_by\_60\_sq | 0.3045681000240437 |
| 16 | Card\_Merchnum\_State\_count\_1\_by\_30\_sq | 0.2450211593465581 |
| 17 | Card\_Merchdesc\_Zip\_vdratio\_0by60 | 0.2777984459387328 |
| 18 | Card\_Merchnum\_State\_total\_amount\_1\_by\_30 | 0.2487841619938918 |
| 19 | Card\_Merchdesc\_total\_1 | 0.2436526997134087 |
| 20 | Card\_dow\_unique\_count\_for\_Card\_Merchdesc\_1 | 0.4868859821823698 |

# Fifth Run – Random Forest

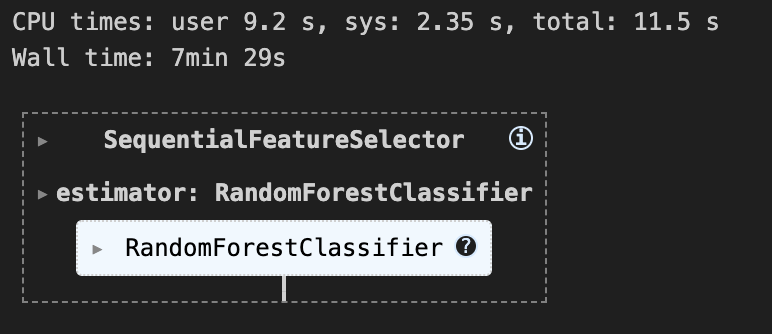
Model: RandomForestClassifier

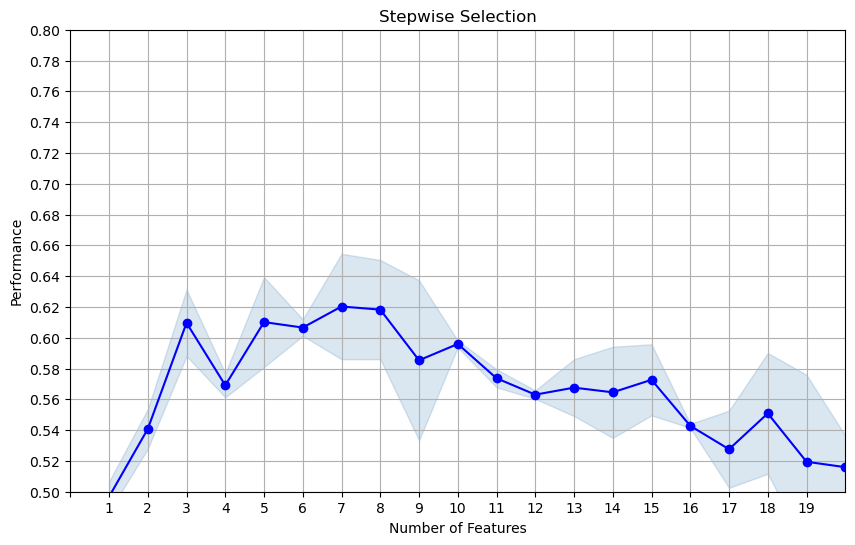
Wrapper Complexity: n\_estimators = 5

Feature Selection: Forward

num\_out\_of\_filter: 350

num\_out\_of\_wrapper: 20





**List of top 20 variables:**

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | Card\_Merchnum\_desc\_vdratio\_1by60 | 0.2700875169200329 |
| 3 | Card\_Merchnum\_Zip\_total\_3 | 0.2546391443634789 |
| 4 | Card\_Merchdesc\_vdratio\_0by60 | 0.2791824341468994 |
| 5 | Card\_Merchnum\_desc\_vdratio\_0by60 | 0.2700875169200329 |
| 6 | Card\_Merchdesc\_State\_vdratio\_0by60 | 0.2788128654655833 |
| 7 | Merchnum\_total\_1 | 0.2431599668140857 |
| 8 | Card\_Merchdesc\_State\_vdratio\_1by60 | 0.2788128654655833 |
| 9 | Card\_Merchnum\_State\_total\_30 | 0.2539895418120028 |
| 10 | Card\_dow\_count\_0\_by\_60\_sq | 0.3738634822211754 |
| 11 | Card\_Merchdesc\_total\_30 | 0.2552448791657018 |
| 12 | Card\_Merchnum\_State\_count\_0\_by\_60 | 0.2364364232832324 |
| 13 | card\_zip\_total\_amount\_1\_by\_30 | 0.2810112010380260 |
| 14 | Card\_dow\_unique\_count\_for\_state\_des\_7 | 0.4232198256377344 |
| 15 | Cardnum\_unique\_count\_for\_card\_state\_14 | 0.4416929948133546 |
| 16 | Card\_Merchnum\_desc\_total\_60 | 0.2407418173247832 |
| 17 | Card\_dow\_count\_0\_by\_60 | 0.3738634822211754 |
| 18 | Cardnum\_max\_3 | 0.4662718780420817 |
| 19 | Cardnum\_actual/max\_0 | 0.4609703807161835 |
| 20 | card\_zip\_max\_1 | 0.2400758879116918 |

# Sixth Run – CatBoost

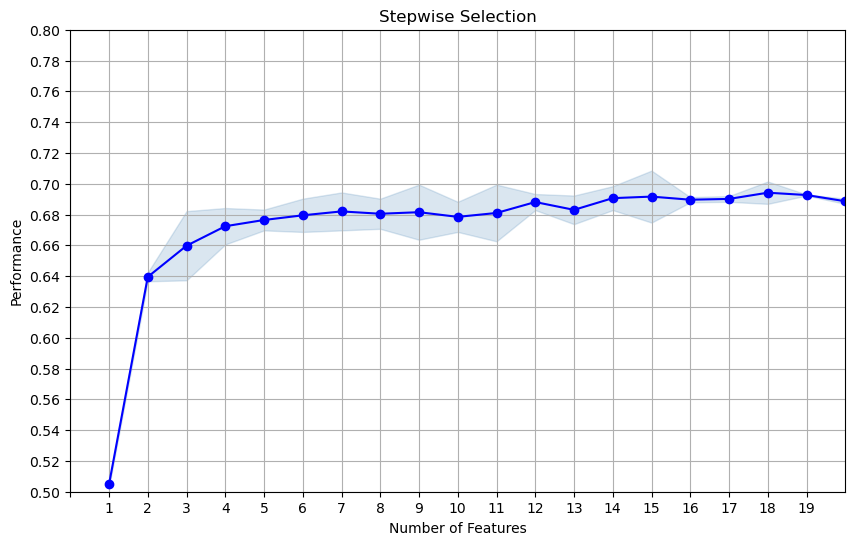
Model: CatBoostClassifier

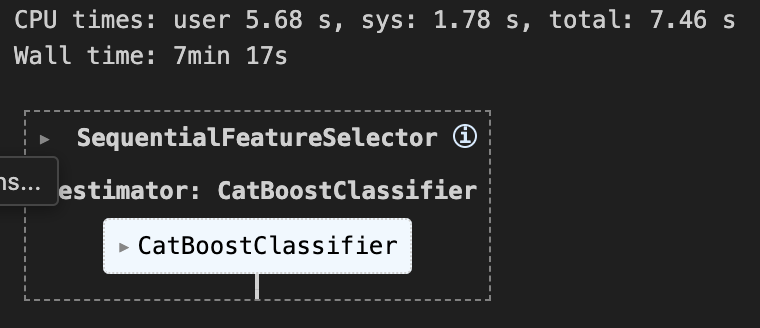
Wrapper Complexity: iterations = 20, depth = 3

Feature Selection: Forward

num\_out\_of\_filter: 200

num\_out\_of\_wrapper: 20





**List of top 20 variables:**

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | Cardnum\_total\_1 | 0.5911137143420107 |
| 3 | Card\_dow\_count\_0\_by\_60 | 0.3738634822211754 |
| 4 | card\_zip\_total\_amount\_1\_by\_30 | 0.2810112010380260 |
| 5 | Cardnum\_vdratio\_0by14 | 0.4017270867195838 |
| 6 | Card\_Merchnum\_Zip\_total\_amount\_1\_by\_60 | 0.3058974499494893 |
| 7 | card\_state\_total\_amount\_1\_by\_60 | 0.3488908916449487 |
| 8 | Cardnum\_avg\_0 | 0.3388579940008128 |
| 9 | Card\_dow\_max\_14 | 0.4319636171379462 |
| 10 | Cardnum\_total\_60 | 0.2876972314343721 |
| 11 | Card\_dow\_actual/max\_7 | 0.3596461246195071 |
| 12 | Cardnum\_actual/toal\_3 | 0.4156378832607220 |
| 13 | card\_state\_total\_3 | 0.2915713416378370 |
| 14 | card\_zip\_count\_1\_by\_30 | 0.2737839069856730 |
| 15 | card\_state\_count\_1\_by\_60\_sq | 0.3443984197726720 |
| 16 | Card\_dow\_actual/toal\_7 | 0.3941931555794056 |
| 17 | Cardnum\_max\_7 | 0.3799071073166451 |
| 18 | card\_state\_count\_1\_by\_30\_sq | 0.3048364256907635 |
| 19 | Cardnum\_total\_amount\_0\_by\_60 | 0.3702963005435002 |
| 20 | card\_state\_total\_amount\_1\_by\_30 | 0.2998429086015698 |

# Seventh Run – CatBoost

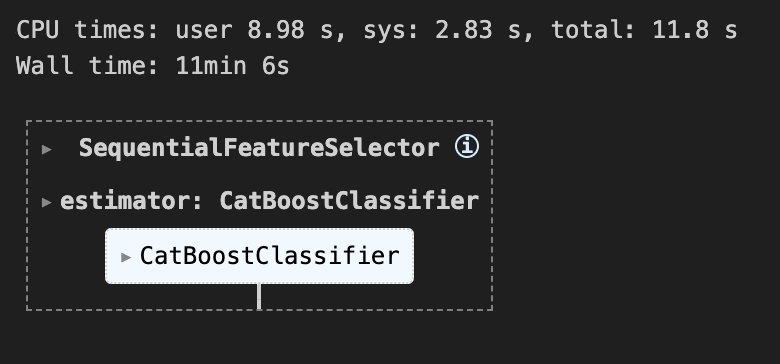
Model: CatBoostClassifier

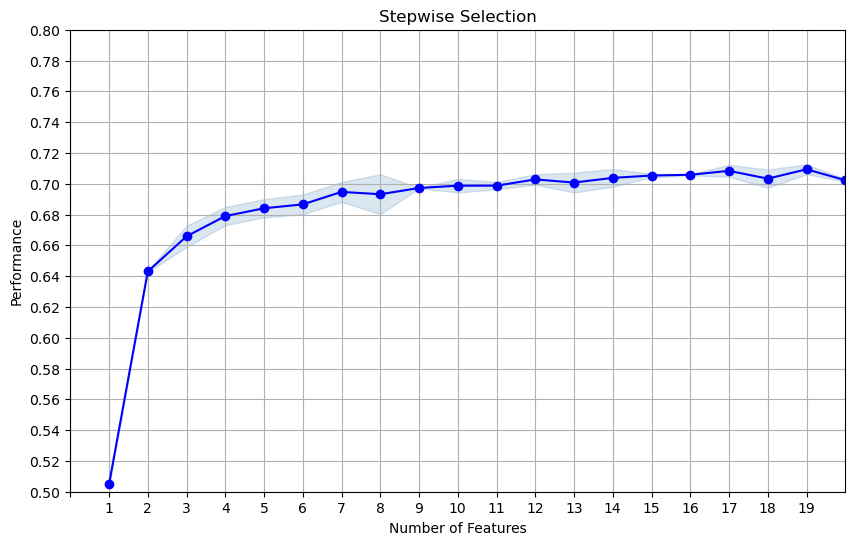
Wrapper Complexity: iterations = 20, depth = 3

Feature Selection: Forward

num\_out\_of\_filter: 300

num\_out\_of\_wrapper: 20





**List of top 20 variables:**

|  |  |  |
| --- | --- | --- |
| Wrapper Order | Variable | Filter Score |
| 1 | Cardnum\_unique\_count\_for\_card\_state\_1 | 0.5186059951957085 |
| 2 | Card\_Merchdesc\_State\_total\_3 | 0.2599209696267789 |
| 3 | Card\_Merchnum\_Zip\_total\_amount\_1\_by\_60 | 0.3058974499494893 |
| 4 | Cardnum\_max\_0 | 0.4923163140953655 |
| 5 | card\_merch\_total\_7 | 0.2598647537776855 |
| 6 | Cardnum\_total\_amount\_0\_by\_60 | 0.3702963005435002 |
| 7 | card\_zip\_total\_amount\_0\_by\_60 | 0.2655482819256827 |
| 8 | card\_state\_max\_7 | 0.2539558237066410 |
| 9 | Card\_dow\_vdratio\_0by14 | 0.5026571349557347 |
| 10 | Card\_dow\_total\_30 | 0.4268030110610211 |
| 11 | High\_Amount\_For\_Merchant | 0.2515712466037181 |
| 12 | Cardnum\_variability\_med\_0 | 0.2691895078531444 |
| 13 | Card\_dow\_avg\_60 | 0.2510064905135359 |
| 14 | card\_state\_total\_14 | 0.2642250521974898 |
| 15 | Card\_dow\_actual/toal\_14 | 0.3370331741934217 |
| 16 | Card\_dow\_unique\_count\_for\_Card\_Merchdesc\_1 | 0.4868859821823698 |
| 17 | Card\_dow\_count\_0\_by\_60\_sq | 0.3738634822211754 |
| 18 | Merchnum\_desc\_State\_total\_3 | 0.2491566197124204 |
| 19 | Cardnum\_unique\_count\_for\_card\_state\_60 | 0.3273076232251982 |
| 20 | Cardnum\_variability\_max\_3 | 0.4159074760490262 |