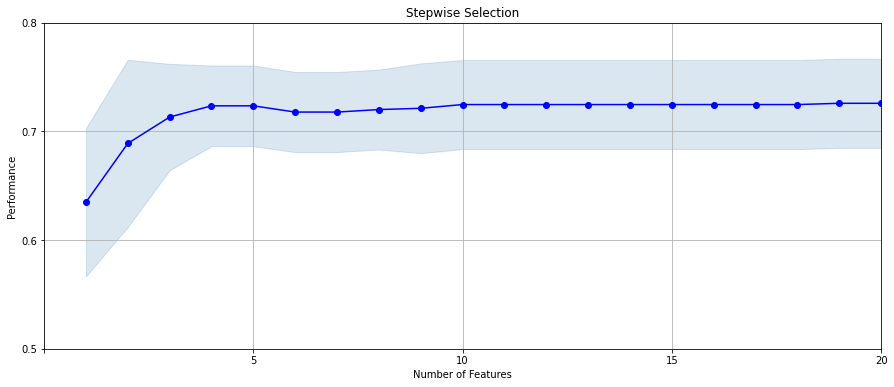
**Forward Selection**

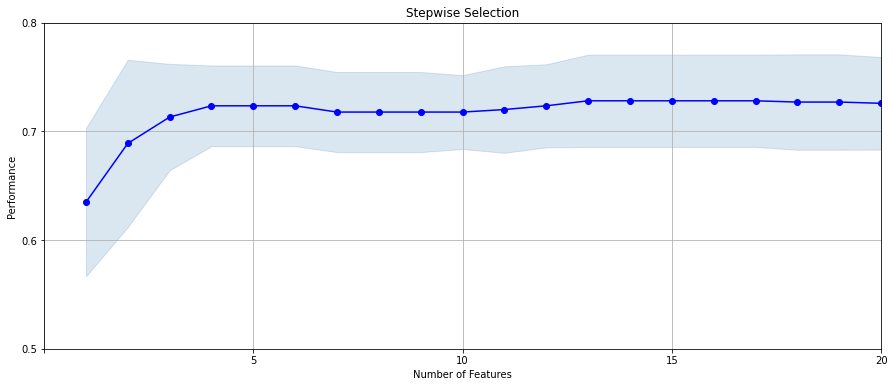
**LGBMClassifier with (n\_estimators=20,num\_leaves=4)**

Filter number= 100

Wrapper number=20

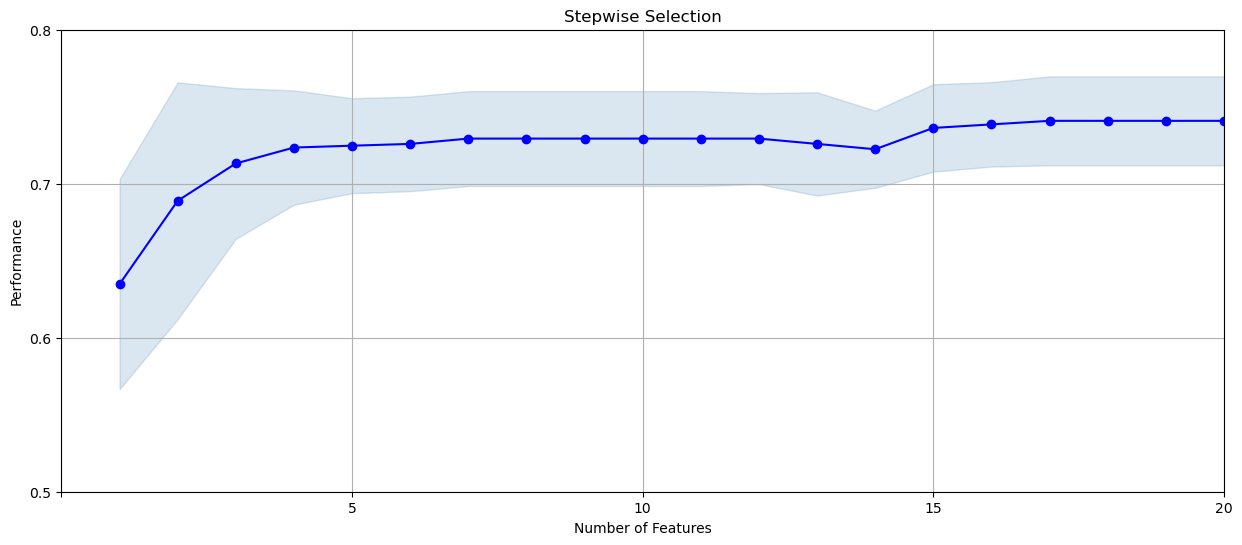
Filter number = 200

Wrapper number = 20



Filter number = 300

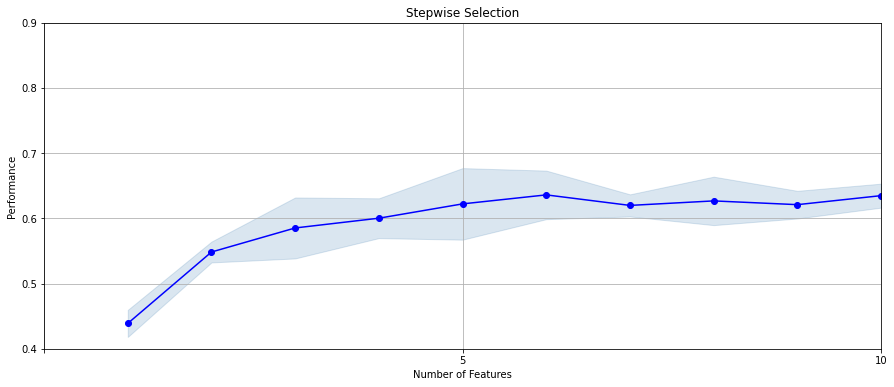
Wrapper number = 20



**Random Forest: (n\_estimators=5)**

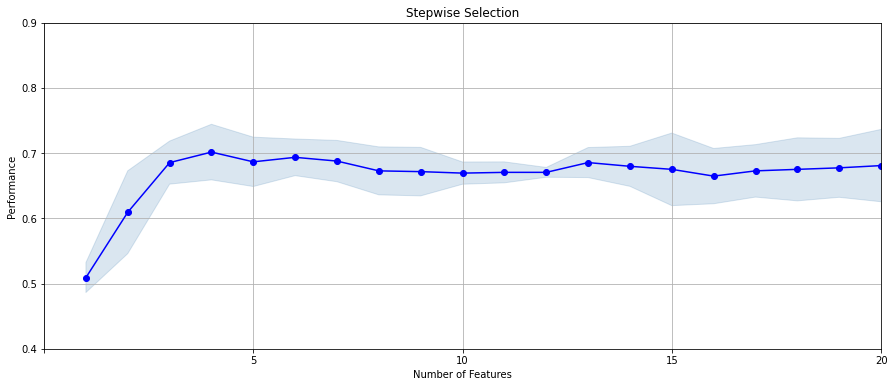
Filter number =100

Wrapper number =10



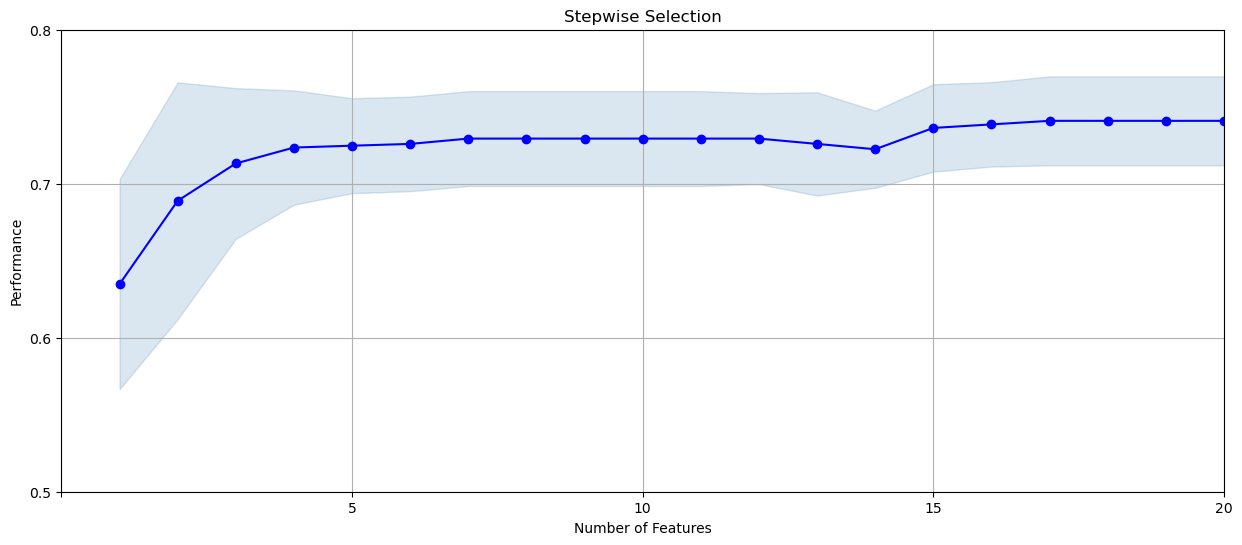
Filter number =200

Wrapper number =20



Filter number: 300

Wrapper number: 20

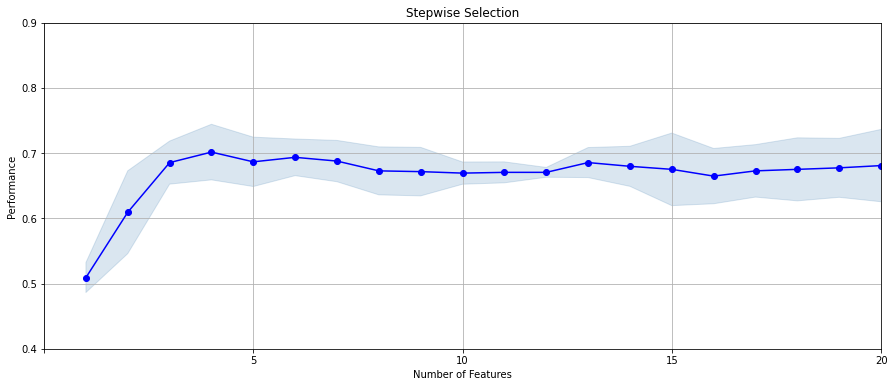


**Backward selection:**

LGBMClassifier with (n\_estimators=20,num\_leaves=4)

Filter number: 150

Wrapper number: 20

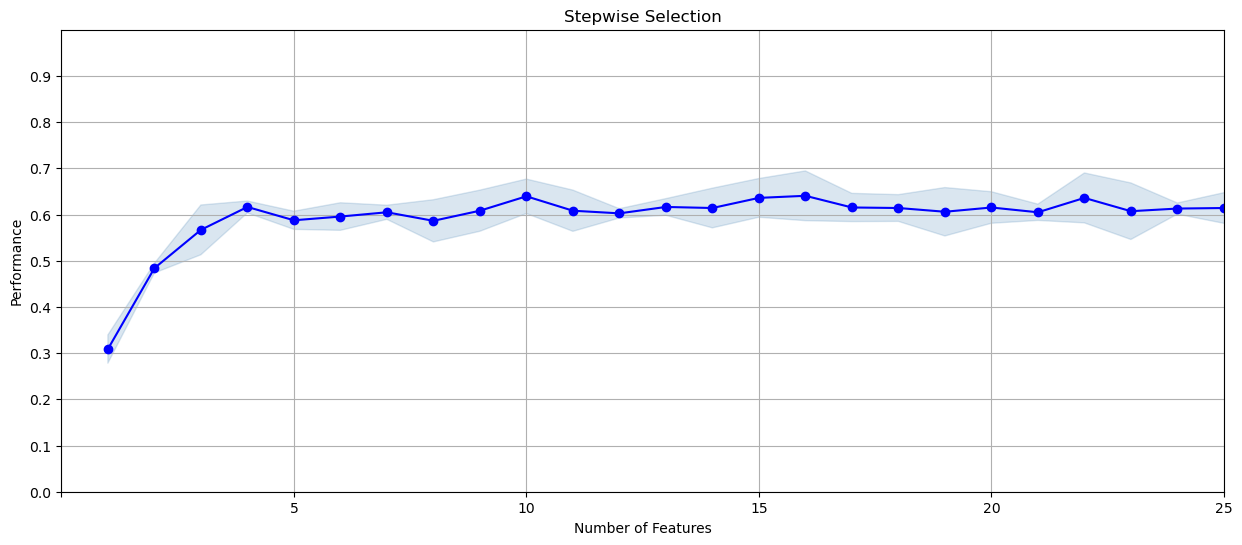


**Backward selection:**

**Random Forest ((n\_estimators=5)**

Filter number: 100

Wrapper number: 20



For the forward selection, I run the model of **LGBMClassifier** and **Random Forest** for three time respectively. It turns out that when I increase the number of filters from 100 to 300, and number of wrappers from 10 to 20, the filter score increases as well. The performance of **LGBMClassifier** isbetter than random forest in most cases when I run the model for several times. In terms of backward selection, I also run LGBMClassifier (filter = 150, wrapper =20) and RF (filter = 100, wrapper= 20) with parameters that are the same as forward selection. Overall, the performance of backward selection is not good as forward selection.

Therefore, the final model I choose is LGBMClassifier with (n\_estimators=20,num\_leaves=4) with filter number: 300 and wrapper number:20 since it has the highest filter score of around 0.73 with relatively low running time.

**Final list of Variables: 20**

|  |  |
| --- | --- |
|  | variable name |
| 1 | **card\_merch\_total\_14** |
| 2 | **['card\_zip3\_max\_14']** |
| 3 | **['card\_merch\_total\_1']** |
| 4 | **['Card\_Merchdesc\_max\_30']** |
| 5 | **['zip3\_actual/med\_14']** |
| 6 | **['card\_zip\_max\_30']** |
| 7 | **['Card\_Merchnum\_desc\_max\_7']** |
| 8 | **['Card\_Merchnum\_desc\_max\_14']** |
| 9 | **['card\_merch\_max\_14']** |
| 10 | **['card\_merch\_max\_7']** |
| 11 | **['amount\_cat']** |
| 12 | **['zip3\_variability\_avg\_3']** |
| 13 | **['card\_zip\_total\_30']** |
| 14 | **['zip3\_actual/avg\_60']** |
| 15 | **['zip3\_variability\_avg\_7']** |
| 16 | **['card\_zip3\_max\_0']** |
| 17 | **['zip3\_variability\_med\_3']** |
| 18 | **['Card\_Merchnum\_desc\_max\_1']** |
| 19 | **['Card\_Merchdesc\_total\_0']** |
| 20 | **['card\_zip\_max\_1']** |