Feature Selection - Credit Card Transaction
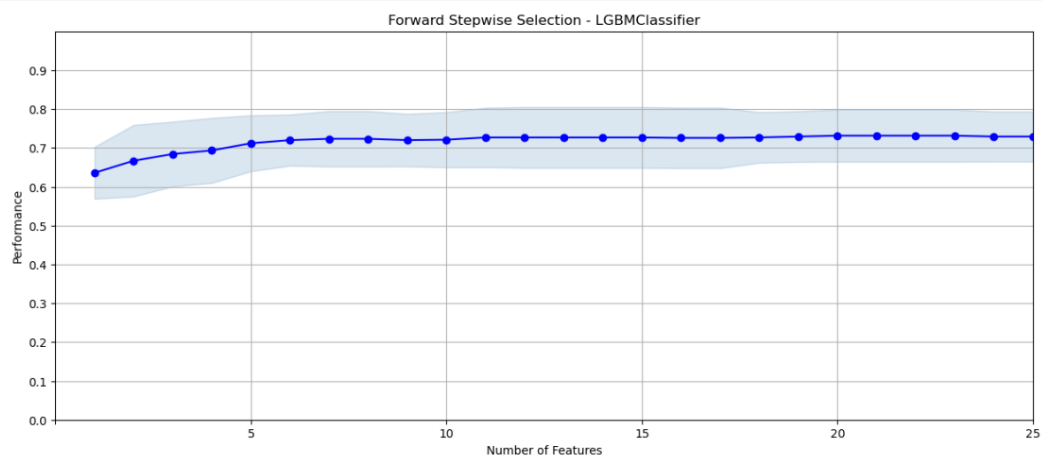
I made my final selection from the three wrappers I have as follows based on the improvement starting from the fifth variable:
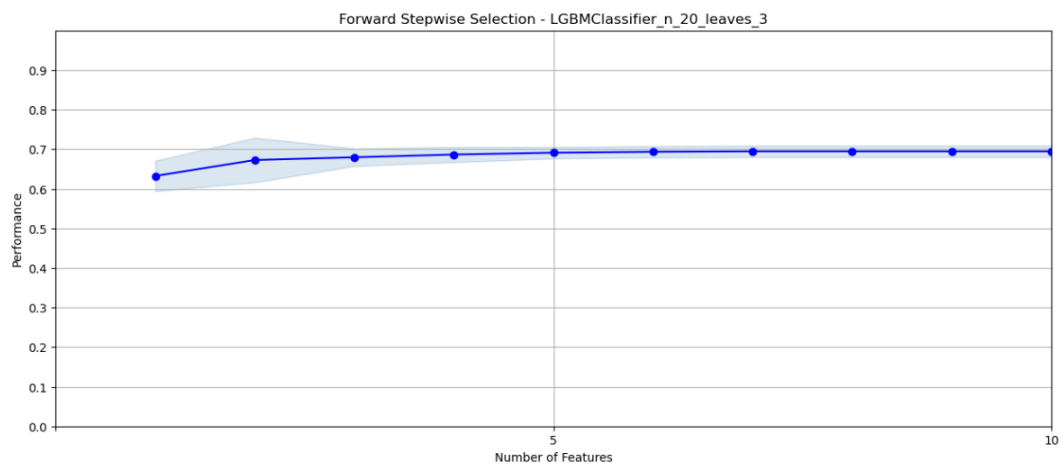
- LGBMClassifier(n_estimators=30, num_leaves=4, cv=3) provides the improvement from 5 to 10 number of features. The maximum score is at 0.73. I choose this wrapper.

```
In [245]:  1  %%time
           2  sfs_forward_LGBM = wrapper_execution(algorithm = LGBMClassifier(n_estimators=30, num_leaves=4),
           3                                       title = 'Forward Stepwise Selection - LGBMClassifier',
           4                                       filename = 'wrapper_LGBMClassifier_n_30_leaves_4_forward.png',
           5                                       forward_flag=True)
           6
           7  print('time to here:', dt.datetime.now() - start_time)
```
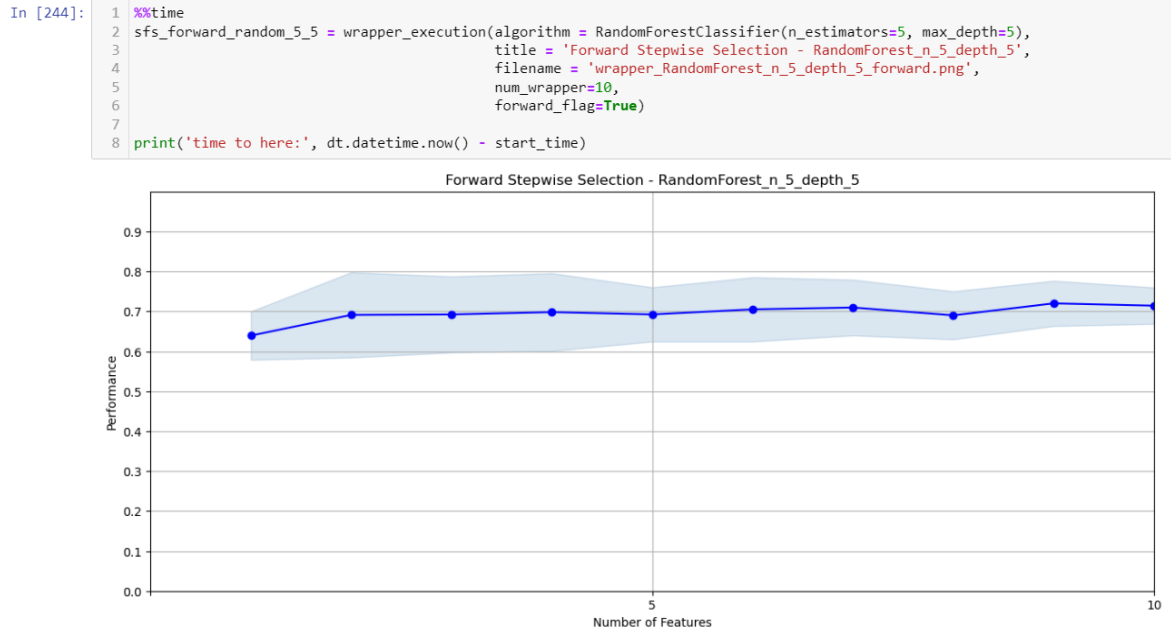

Forward Stepwise Selection - LGBMClassifier

- LGBMClassifier(n_estimators=20, num_leaves=3, cv=3) has no improvement after the fifth feature. The maximum score is at 0.70

```
In [240]:  1  %%time
           2  sfs_forward_LGBM_20_3 = wrapper_execution(algorithm = LGBMClassifier(n_estimators=20, num_leaves=3),
           3                                            title = 'Forward Stepwise Selection - LGBMClassifier_n_20_leaves_3',
           4                                            filename = 'wrapper_LGBMClassifier_n_20_leaves_3_forward_10.png',
           5                                            num_wrapper=10,
           6                                            forward_flag=True)
           7
           8  print('time to here:', dt.datetime.now() - start_time)
```


Forward Stepwise Selection - LGBMClassifier_n_20_leaves_3

- RandomForestClassifier(n_estimator=5, max_depth=5, cv=3) also has no improvement after the fifth feature. The maximum score is at 0.70

```
In [244]:  1  %%time
           2  sfs_forward_random_5_5 = wrapper_execution(algorithm = RandomForestClassifier(n_estimators=5, max_depth=5),
           3                                 title = 'Forward Stepwise Selection - RandomForest_n_5_depth_5',
           4                                 filename = 'wrapper_RandomForest_n_5_depth_5_forward.png',
           5                                 num_wrapper=10,
           6                                 forward_flag=True)
           7
           8  print('time to here:', dt.datetime.now() - start_time)
```


This is the list of variables from LGBMClassifier(n_estimators=30, num_leaves=4, cv=3) that I will use for modeling.

wrapper order	variable	filter score
1	card_Merch_num_state_total_14	0.676008511
2	card_Merchnum_max_30	0.648741321
3	card_Merch_des_zip_total_1	0.634249492
4	Merch_description_total_0	0.581051024
5	card_Merch_state_zip_total_14	0.670751528
6	card_Merch_des_state_max_30	0.654714819
7	card_Merch_des_zip_max_30	0.651810225
8	card_Merch_des_state_zip_max_30	0.651810225
9	card_Merch_num_des_zip_total_0	0.602714009
10	card_Merch_des_state_zip_total_30	0.652168078
11	card_Merch_num_state_zip_total_0	0.607052787
12	card_Merch_num_state_zip_total_30	0.655072672
13	card_Merch_des_zip_total_30	0.652155541
14	card_Merch_num_zip_total_0	0.60704025
15	card_Merch_des_zip_total_0	0.605781583
16	card_Merch_state_zip_total_0	0.61025694
17	card_Merch_num_state_total_30	0.659830082
18	Merch_state_zip_variability_avg_14	0.40467819

wrapper order	variable	filter score
19	card_Merch_num_des_total_0	0.60717186
20	card_Merch_des_state_total_30	0.657050858
21	card_Merch_des_state_total_0	0.610189287
22	card_Merch_description_total_0	0.610164213
23	card_Merch_zip_total_0	0.609993663
24	card_Merch_num_des_zip_total_14	0.65931699
25	card_Merch_description_total_30	0.656837729

This is the wrapper score from LGBMClassifier(n_estimators=30, num_leaves=4, cv=3) for reference.

avg_score	ci_bound	std_dev	std_err	order	variable name
0.6359	0.1508	0.0670	0.0474	1	card_Merch_num_state_total_14
0.6669	0.2077	0.0923	0.0653	2	card_Merchnum_max_30
0.6842	0.1875	0.0833	0.0589	3	card_Merch_des_zip_total_1
0.6934	0.1886	0.0838	0.0593	4	Merch_description_total_0
0.7119	0.1622	0.0721	0.0510	5	card_Merch_state_zip_total_14
0.7200	0.1485	0.0660	0.0467	6	card_Merch_des_state_max_30
0.7234	0.1610	0.0716	0.0506	7	card_Merch_des_zip_max_30
0.7234	0.1610	0.0716	0.0506	8	card_Merch_des_state_zip_max_30
0.7200	0.1525	0.0678	0.0479	9	card_Merch_num_des_zip_total_0
0.7211	0.1602	0.0712	0.0503	10	card_Merch_des_state_zip_total_30
0.7269	0.1726	0.0767	0.0542	11	card_Merch_num_state_zip_total_0
0.7269	0.1762	0.0783	0.0554	12	card_Merch_num_state_zip_total_30
0.7269	0.1762	0.0783	0.0554	13	card_Merch_des_zip_total_30
0.7269	0.1762	0.0783	0.0554	14	card_Merch_num_zip_total_0
0.7269	0.1762	0.0783	0.0554	15	card_Merch_des_zip_total_0
0.7257	0.1759	0.0782	0.0553	16	card_Merch_state_zip_total_0
0.7257	0.1759	0.0782	0.0553	17	card_Merch_num_state_total_30
0.7269	0.1468	0.0653	0.0461	18	Merch_state_zip_variability_avg_14
0.7292	0.1470	0.0653	0.0462	19	card_Merch_num_des_total_0
0.7315	0.1513	0.0672	0.0476	20	card_Merch_des_state_total_30
0.7315	0.1513	0.0672	0.0476	21	card_Merch_des_state_total_0
0.7315	0.1513	0.0672	0.0476	22	card_Merch_description_total_0
0.7315	0.1513	0.0672	0.0476	23	card_Merch_zip_total_0
0.7292	0.1456	0.0647	0.0457	24	card_Merch_num_des_zip_total_14
0.7292	0.1456	0.0647	0.0457	25	card_Merch_description_total_30