1 2022/7/11

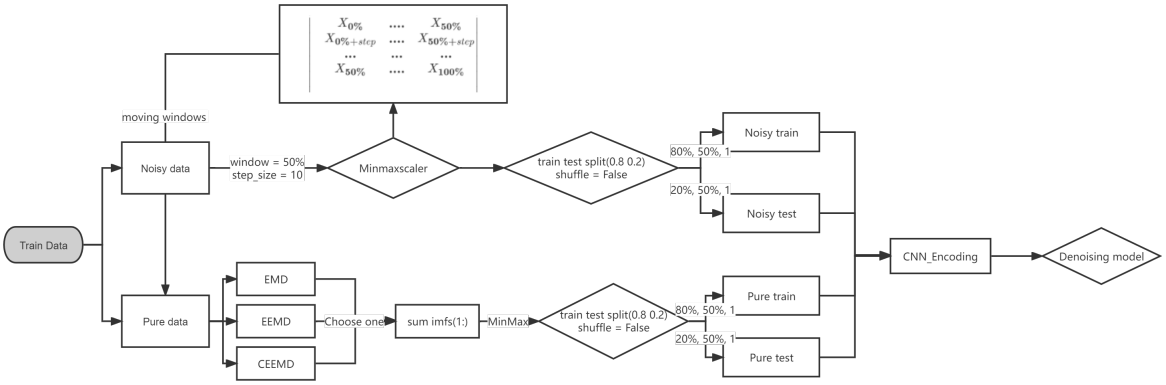

Figure 1: Pretext task

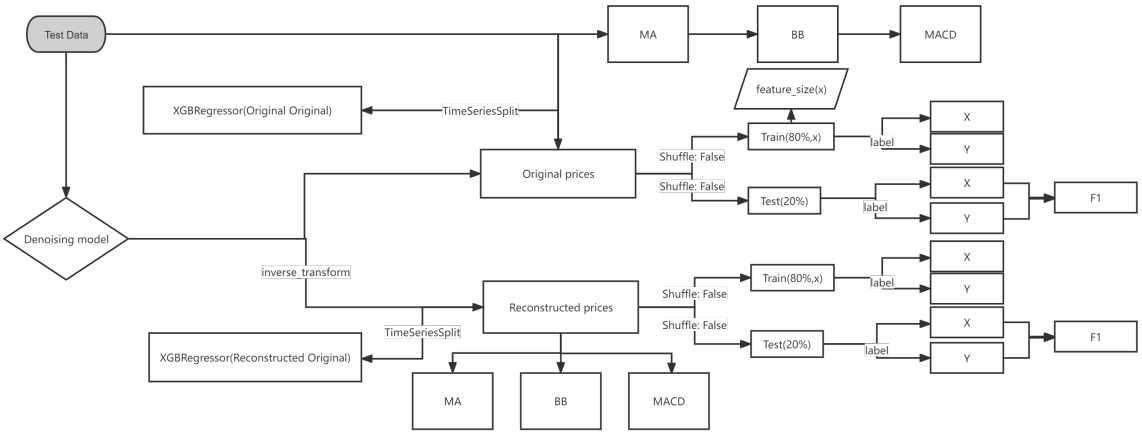

Figure 2: Downstream task

1.1 Close price of S&P 500 index from Yahoo finance (small dataset)

- **Train Data: 2017.01.01-2019.01.01**
  Test Data: 2017.01.01-2019.01.01

  1. 502 observations (step=10)
     Pure data: (26, 251, 1)
     Noise data: (26, 251, 1)

2. Loss plot

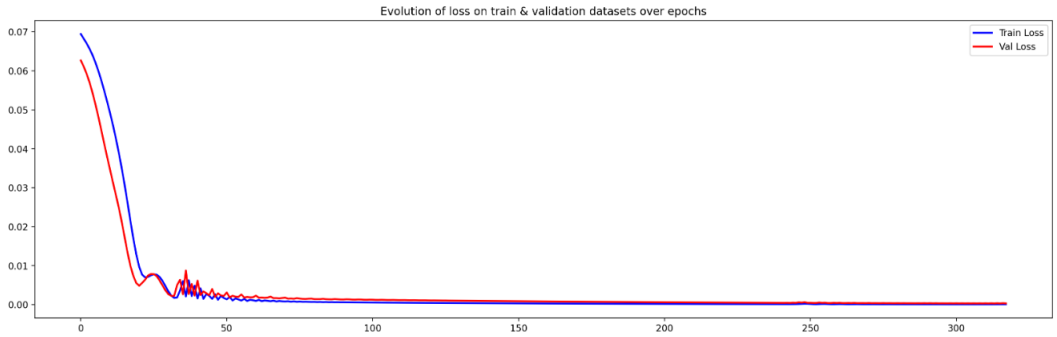

Figure 3: Loss plot(2017-2019)

3. Denoised plot

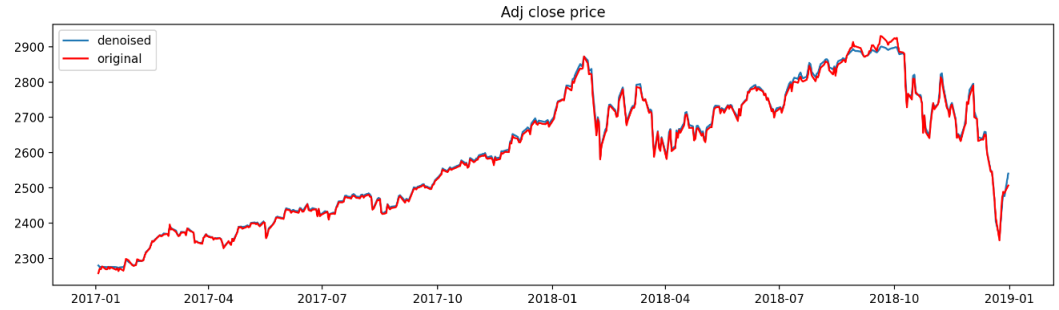

Figure 4: Denoised plot(2017-2019)

4. Regression Forecasting

   (a) Generate features(200 features in total)
   SMA(2-51)
   EMA(2-51)
   STDDEV(2-51)
   LAG(2-51)

(b) TimeSeriesSplit=10

1. Using feature matrix generated by reconstructed data to predict original data

The last test MSE: 3626.6775423724475

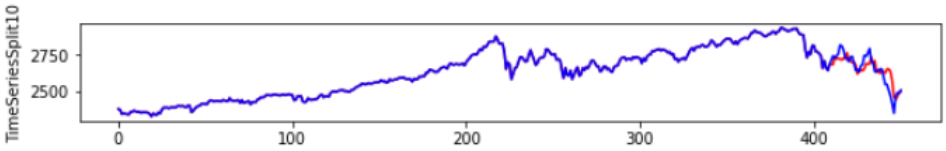

Figure 5: XGBre-RO-1719-19

HalvingGridSearchCV:
Optimal values for hyperparameters: 'learning_rate': 0.1, 'max_depth': 10, 'n_estimators': 500
Best model score: -2585.5981169086917
Test MSE: 3332.947

HalvingRandomSearchCV:
Optimal values for hyperparameters: 'subsample': 0.95, 'scale_pos_weight': 0.4, 'reg_lambda': 0.6, 'reg_alpha': 0.5, 'n_estimators': 700, 'min_child_weight': 0, 'max_depth': 10, 'learning_rate': 0.02, 'gamma': 0.4, 'colsample_bytree': 0.6
Best model score: -2436.3794010297015
Test MSE: 3121.3774

BayesianOptimization:
Test MSE: 2191.956298828125

2. Using feature matrix generated by original data to predict original data

The last test MSE: 4746.077826718005

HalvingGridSearchCV:
Optimal values for hyperparameters: 'learning_rate': 0.05, 'max_depth':

10, 'n_estimators': 700
Best model score: -2811.671274117187
Test MSE: 4087.8457103080864

HalvingRandomSearchCV:
Optimal values for hyperparameters: 'subsample': 0.85, 'scale_pos_weight': 1, 'reg_lambda': 1, 'reg_alpha': 0.5, 'n_estimators': 100, 'min_child_weight': 0, 'max_depth': 10, 'learning_rate': 0.1, 'gamma': 0.1, 'colsample_bytree': 0.7
Best model score: -2955.2431469829
Test MSE: 5002.151511272279

BayesianOptimization:
Test MSE: 2740.978656614699

5. SVC(label-Trend Forecast)

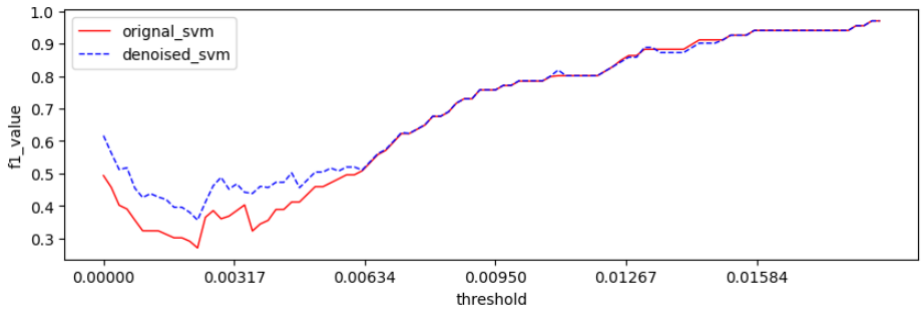

Figure 6: SVC(2017-2019)

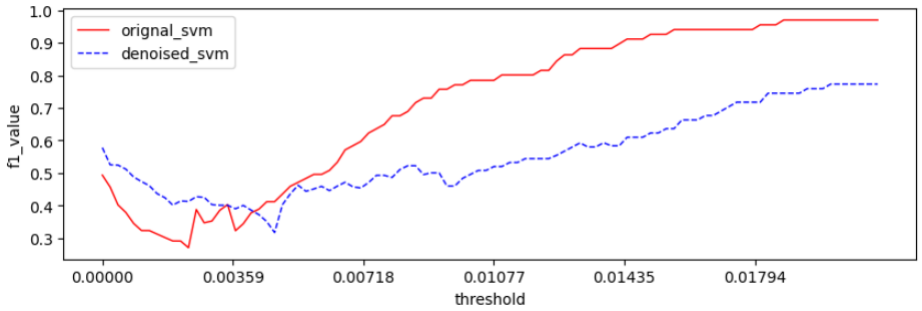

Figure 7: EMDSVC(2017-2019)

6. Strategies

   Moving Average crossover

Table 1: Close Price Comparison via MA Crossover

Date	Buy with Original Signals	Date	Buy with Original Signals
2017-01-18	2271.889893	2017-01-30	2280.899902
2017-04-28	2384.199951	2017-04-28	2384.199951
2017-09-05	2457.850098	2017-09-05	2457.850098
2018-03-16	2752.010010	2018-03-16	2752.010010
2018-05-14	2730.129883	2018-05-14	2730.129883

   Bollinger Bands

Table 2: Close Price Comparison via BB

Date	Buy with Original Signals	Date	Buy with Original Signals
2017-03-21	2344.020020	2017-03-21	2341.590088
2017-04-13	2328.949951		
2017-06-29	2419.699951	2017-06-29	2419.699951
2017-07-06	2409.750000	2017-07-06	2409.750000
2017-08-10	2438.209961	2017-08-10	2438.209961
2017-08-17	2430.010010	2017-08-17	2430.010010
2018-02-05	2648.939941	2018-02-05	2648.939941
2018-02-08	2581.000000		
2018-03-22	2643.689941	2018-03-22	2643.689941
2018-06-27	2699.629883		
2018-10-10	2785.679932	2018-10-09	2880.340088
2018-10-24	2656.100098	2018-10-24	2656.100098
2018-12-17	2545.939941	2018-12-17	2545.939941
2018-12-19	2506.959961	2018-12-19	2506.959961

Moving average convergence diver

Table 3: Close Price Comparison via MACD

Date	Buy with Original Signals	Date	Buy with Original Signals
2017-01-04	2270.750000		
2017-01-11	2275.320068		
2017-01-24	2280.070068	2017-01-20	2271.310059
2017-04-24	2374.149902	2017-04-24	2374.149902
2017-05-25	2415.070068	2017-05-25	2415.070068
2017-06-19	2453.459961		
2017-07-13	2447.830078	2017-07-13	2447.830078
2017-08-31	2471.649902	2017-08-31	2471.649902
		2017-11-07	2590.639893
2017-11-28	2627.040039	2017-11-27	2601.419922
2018-01-04	2723.989990	2018-01-04	2723.989990
2018-02-23	2747.300049	2018-02-22	2703.959961
2018-03-05	2720.939941	2018-03-05	2720.939941
2018-04-10	2656.870117	2018-04-11	2642.189941
2018-05-07	2672.629883	2018-05-07	2672.629883
2018-06-04	2746.870117	2018-06-05	2748.800049
2018-07-09	2784.169922	2018-07-09	2784.169922
2018-08-06	2850.399902	2018-08-06	2850.399902
2018-08-24	2874.689941	2018-08-27	2896.739990
2018-09-20	2930.750000	2018-09-21	2929.669922
2018-11-02	2723.060059	2018-11-02	2723.060059
2018-11-28	2743.790039	2018-11-28	2743.790039
		2018-12-31	2506.850098

- **Train Data: 2017.01.01-2019.01.01**
  Test Data: 2018.06.01-2019.06.01

  1. Same denoised model trained by 2017-2019
  2. Denoised plot

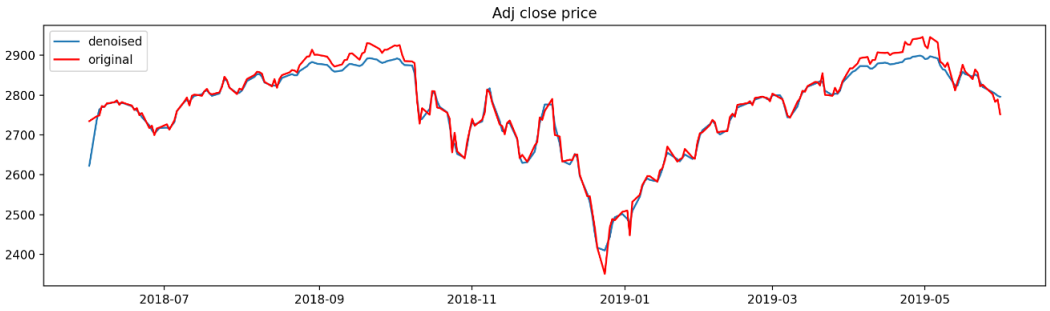

Figure 8: Denoised plot(2018-2019)

  3. Regression Forecasting
     (a) TimeSeriesSplit=10
         1. Using feature matrix generated by reconstructed data to predict original data
         The last test MSE: 2152.6313423679935
         EEMD: 2457.243271258142
         CEEMDAN: 1467.4669466283585

         HalvingGridSearchCV:
         Best model score: -2873.987590862372
         Test MSE: 755.9846
         EEMD: 1124.2512
         CEEMDAN: 2774.193

         HalvingRandomSearchCV:
         Best model score: -4261.291451234443
         Test MSE: 422.973
         EEMD: 676.0305
         CEEMDAN: 1223.3672

BayesianOptimization:
Test MSE: 392.46634048998067
EEMD: 528.8270263671875
CEEMDAN: 262.0235595703125

2. Using feature matrix generated by original data to predict original data
The last test MSE: 1896.106842233075

HalvingGridSearchCV:
Best model score: -5114.213177644038
Test MSE: 2094.160648379061

HalvingRandomSearchCV:
Best model score: -3981.967834908588
Test MSE: 1422.358187529776

BayesianOptimization:
Test MSE: 666.5796529948711

4. SVC(label-Trend Forecast)

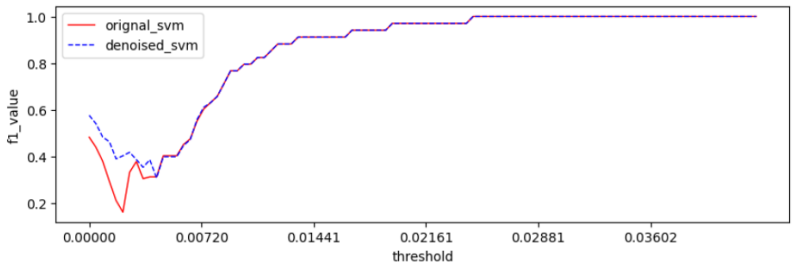

Figure 9: SVC(2017-2019)

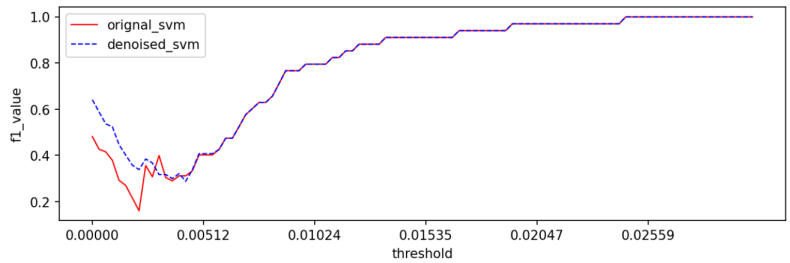

Figure 10: SVC(EEMD)

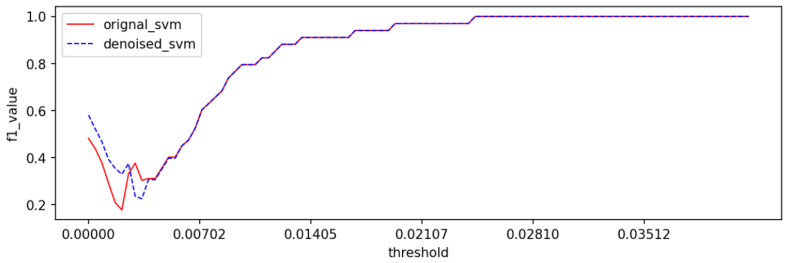

Figure 11: SVC(CEE)

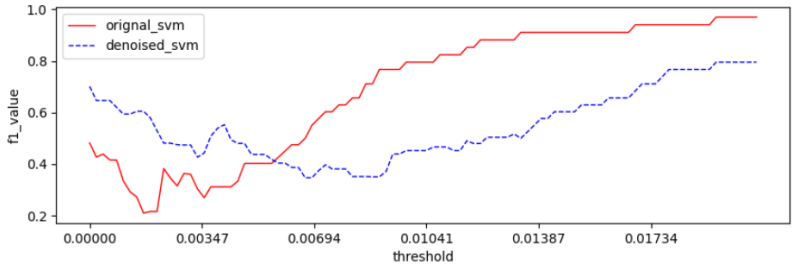

Figure 12: EMDSVC(2017-2019)

5. Strategies

Moving Average crossover

Table 4: Close Price Comparison via MA Crossover

Date	Buy with Original Signals	Date	Buy with Original Signals
2018-06-15	2779.659912	2018-06-15	2779.659912
2018-07-13	2801.310059	2018-07-13	2801.310059
2019-01-24	2642.330078	2019-01-24	2642.330078

Bollinger Bands

Table 5: Close Price Comparison via BB

Date	Buy with Original Signals	Date	Buy with Original Signals
2018-06-25	2717.070068		
2018-06-27	2699.629883		
2018-10-10	2785.679932	2018-10-09	2880.340088
2018-10-24	2656.100098		
2018-12-17	2545.939941		
2018-12-19	2506.959961	2018-12-19	2506.959961
2019-05-13	2811.870117	2019-05-09	2870.719971
2019-05-31	2752.060059		

Moving average convergence diver

Table 6: Close Price Comparison via MACD

Date	Buy with Original Signals	Date	Buy with Original Signals
2018-06-04	2746.870117	2018-06-04	2746.870117
2018-07-06	2759.820068	2018-07-06	2759.820068
2018-08-03	2840.350098	2018-08-03	2840.350098
2018-08-22	2861.820068		
2018-08-24	2874.689941	2018-08-24	2874.689941
2018-09-20	2930.750000	2018-09-21	2929.669922
2018-11-02	2723.060059	2018-11-02	2723.060059
2018-11-28	2743.790039	2018-11-28	2743.790039
2019-01-02	2510.030029	2019-01-03	2447.889893
2019-03-19	2832.570068	2019-03-19	2832.570068
2019-04-02	2867.239990	2019-04-02	2867.239990
2019-04-23	2933.679932		
		2018-12-31	2506.850098

1.2 Minute price of bitcoin (large dataset)

- **Train Data: 2021.01.01-2021.05.12**
  Test Data: 2021.01.01-2021.05.12

  1. 188316 observations (step=2000)
     Pure data: (48, 94158, 1)
     Noise data: (48, 94158, 1)
  2. Loss plot

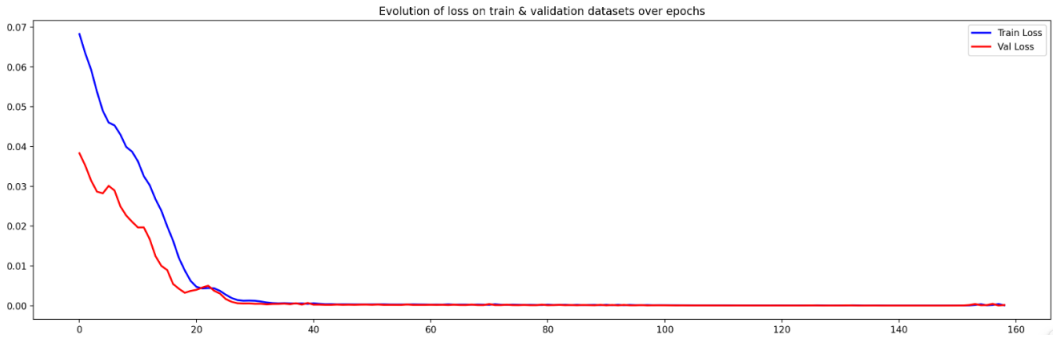

Figure 13: Loss plot(bit)

  3. Denoised plot

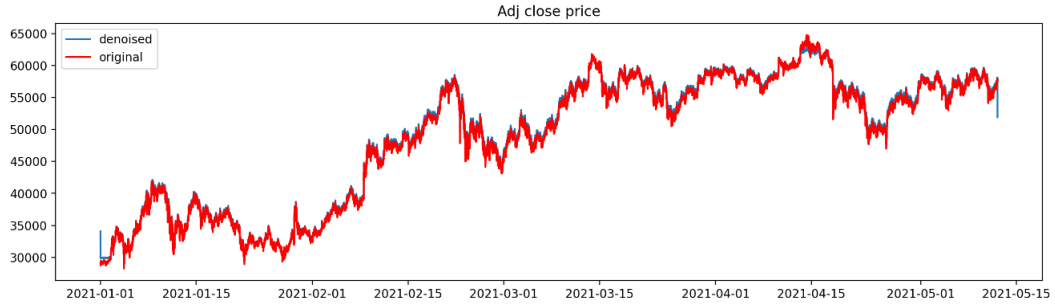

Figure 14: Denoised plot(bit)

  4. Regression Forecasting
     (a) Generate features(200 features in total)
     SMA(2-51)
     EMA(2-51)

STDDEV(2-51)
LAG(2-51)

(b) TimeSeriesSplit=10

1. Using feature matrix generated by reconstructed data to predict original data
The last test MSE: 3775.5785444924863

2. Using feature matrix generated by original data to predict original data
The last test MSE: 4237.80436168226

5. SVC(label-Trend Forecast)

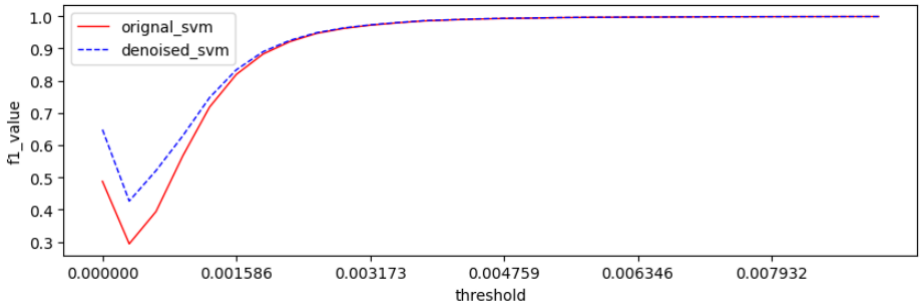

Figure 15: SVC(bit)

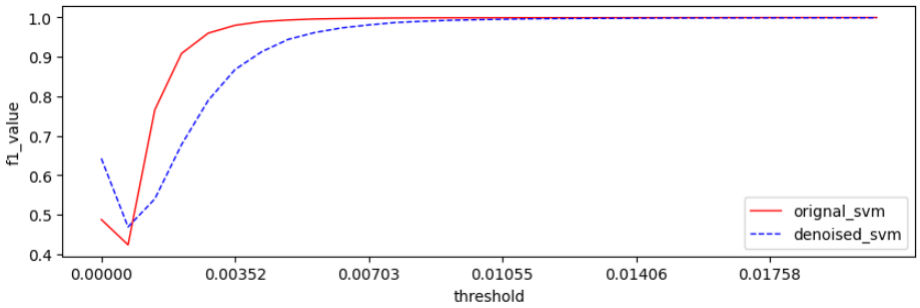

Figure 16: EMDSVC(bit)

- **Train Data: 2021.01.01-2021.05.12(Only using first 80%, 150652 observation)**
  Test Data: 2021.01.01-2021.05.12

  1. Loss plot

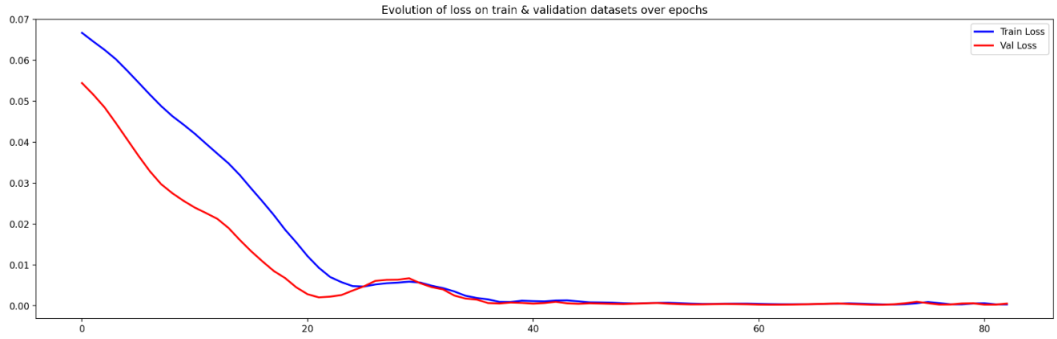

Figure 17: Loss plot(bit)

  2. Denoised plot

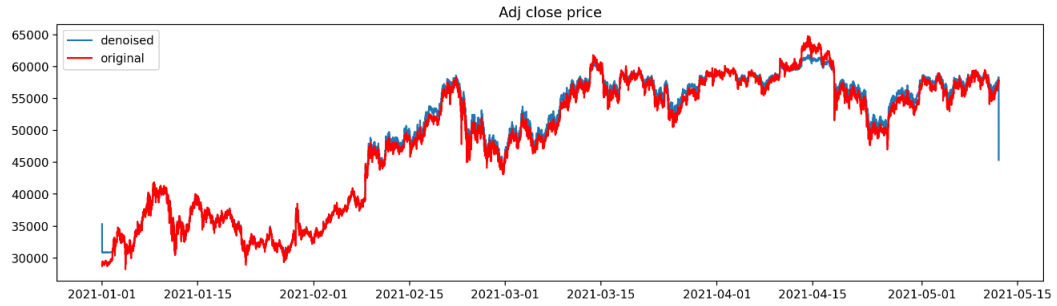

Figure 18: Denoised plot(bit)

  3. Regression Forecasting

Table 7: R to O and O to O

MSE	R to O	O to O
TimeSeriesSplit1	4029.446469436446	6959.288549566705
TimeSeriesSplit2	2528.3044600754783	4641.225435585642
TimeSeriesSplit3	26204805.266945664	28271068.21184866
TimeSeriesSplit4	6048024.076939112	6095684.625164483
TimeSeriesSplit5	8639.155680171423	10902.335757344075
TimeSeriesSplit6	165837.36769624185	198253.59791907933
TimeSeriesSplit7	2121.1523391233663	3662.9754652226393
TimeSeriesSplit8	582471.3196687133	717363.6937461786
TimeSeriesSplit9	4893.151336454887	8267.96235254641
TimeSeriesSplit10	4221.8781678319665	4237.80436168226
The average MSE	3302757.111970282	3532104.1720600342

4. SVC(label-Trend Forecast)

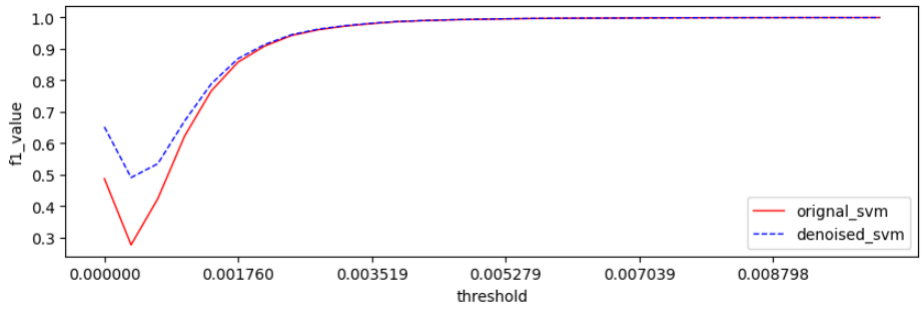

Figure 19: SVC(bit)

1.3 Minute data of road congestion (another type of dataset)

- **Train Data: 1991.04.01-1991.09.30**
  Test Data: 1991.04.01-1991.09.30

  1. 188316 observations (step=100)
     Pure data: (66, 6529, 1)
     Noise data: (66, 6529, 1)
  2. Loss plot

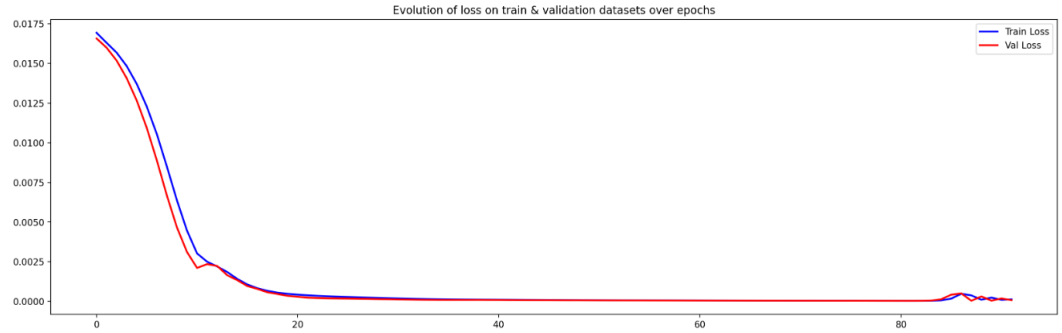

Figure 20: Loss plot(tra)

  3. Denoised plot

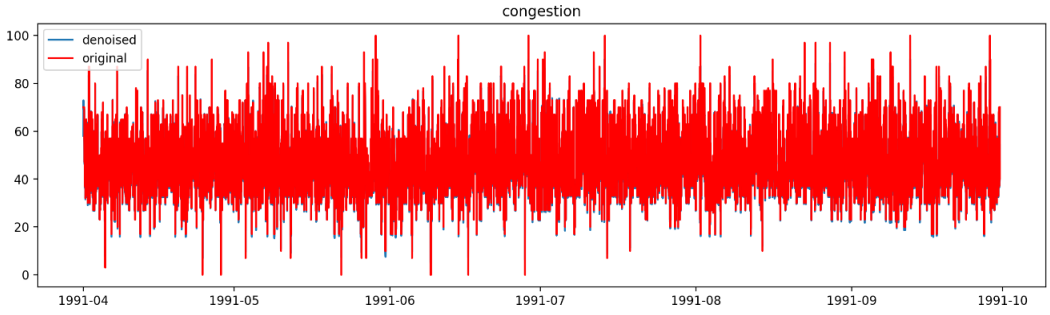

Figure 21: Denoised plot(tra)

  4. Regression Forecasting
     (a) Generate features(200 features in total)
     SMA(2-51)
     EMA(2-51)

STDDEV(2-51)
LAG(2-51)

(b) TimeSeriesSplit=10

1. Using feature matrix generated by reconstructed data to predict original data

The last test MSE: 109.76000038377794

BayesianOptimization:

Test MSE: 99.5308837890625

2. Using feature matrix generated by original data to predict original data

The last test MSE: 124.23597242474737

BayesianOptimization:

Test MSE: 115.01419207394254

5. SVC(label-Trend Forecast)

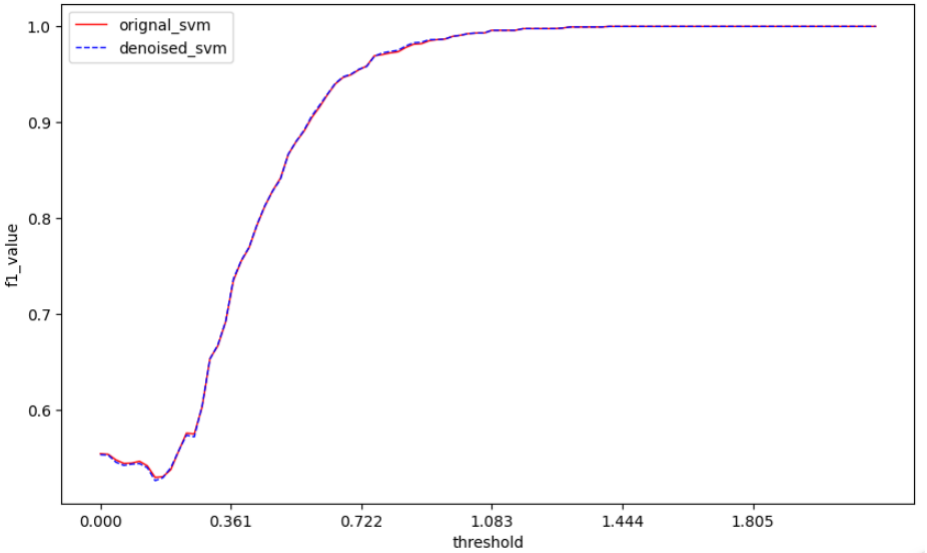

Figure 22: SVC(tra)

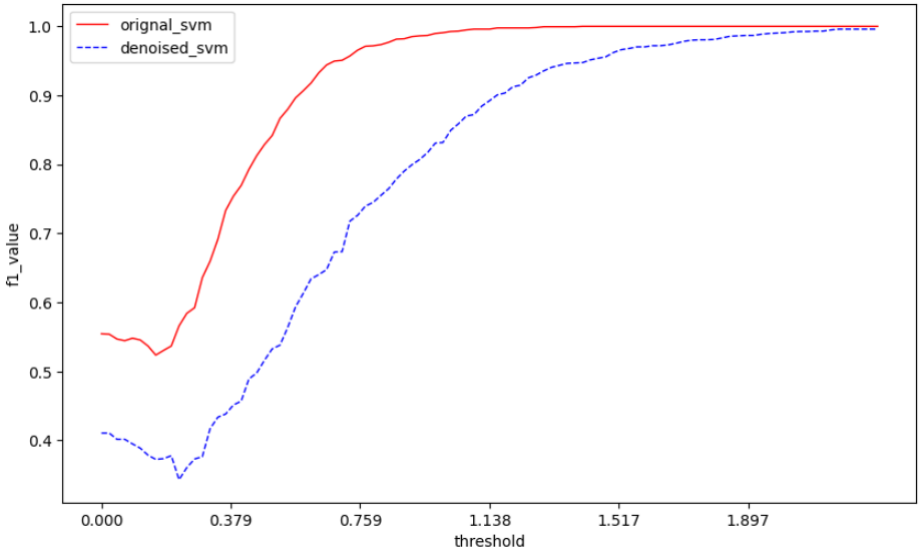

Figure 23: EMDSVC(tra)

- **Train Data: 1991.04.01-1991.09.30(Only using first 80%, 10446 observation)**
  Test Data: 1991.04.01-1991.09.30(Later 20%)

  1. Loss plot

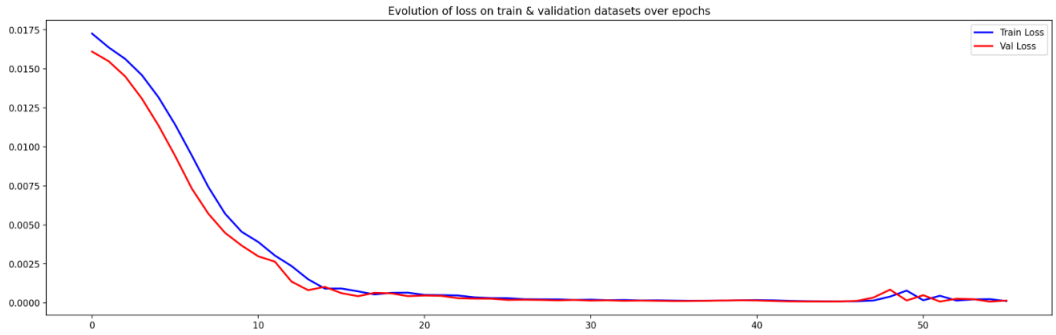

Figure 24: Loss plot(tra)

  2. Denoised plot

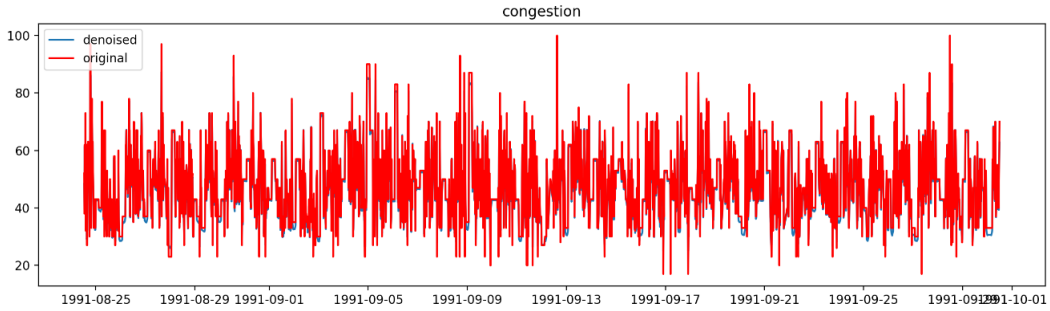

Figure 25: Denoised plot(tra)

  3. Regression Forecasting

Table 8: R to O and O to O

MSE	R to O	O to O
TimeSeriesSplit1	242.68974552124521	290.4062846537116
TimeSeriesSplit2	158.76399623358037	159.50257515128237
TimeSeriesSplit3	219.7605641416004	297.4470299555518
TimeSeriesSplit4	190.1979981669478	200.6940283853277
TimeSeriesSplit5	186.39734349179187	179.50189293799235
TimeSeriesSplit6	123.11959620435692	126.25034154453321
TimeSeriesSplit7	134.58254181552334	137.24418629961175
TimeSeriesSplit8	131.34448558423148	123.13017839965764
TimeSeriesSplit9	117.84628670281049	129.53874957473602
TimeSeriesSplit10	161.7295449938986	155.925421258954
The average MSE	166.64321028559863	179.96406881613584
BayesianOptimization	123.53077697753906	136.86177050502738

4. SVC(label-Trend Forecast)

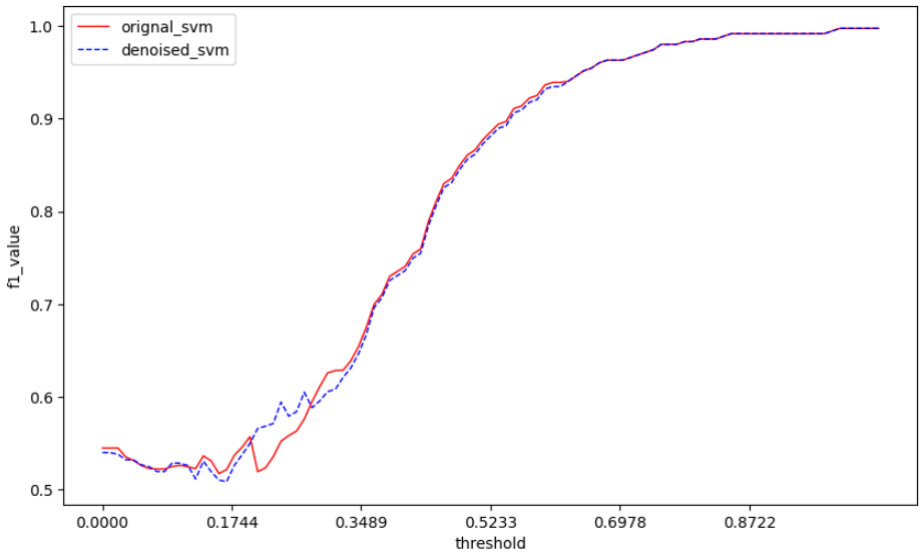

Figure 26: SVC(tra)