

Demonstration that the Use of Feature Selection on High Dimensional South African Macroeconomic Data Results in Improved Performance with Lower Compute Requirements

Rudzani Mulaudzi and Ritesh Ajoodha

The University of the Witwatersrand, Johannesburg, South Africa,
0601737R@students.wits.ac.za,
WWW home page: <https://www.wits.ac.za/>

Abstract. This document shows the tables that show results for a conference paper. The paper is part of the Sixth International Congress on Information and Communication Technology Conference proceedings. The full details being: “Demonstration that the Use of Feature Selection on High Dimensional South African Macroeconomic Data Results in Improved Performance with Lower Compute Requirements,” Sixth International Congress on Information and Communication Technology Conference, ISBN: 2194-5357, 2021.

Keywords: Machine Learning, Feature Selection, Forecasting, Unemployment.

1 Results and Discussions

| Feature Selection | Number of Features |
|--|--------------------|
| No Filter Selection (NO FS) | 147 |
| Removal of Duplicates (Unique) | 145 |
| Pearson Correlation Coefficient (No Cor) | 55 |
| Mutual Information Gain (MIG) | 36 |
| Mutual Information Gain and Pearson Correlation Coefficient (MIG No Corr) | 14 |
| Variance Threshold (Variance) | 145 |
| Variance Threshold and Removal of Duplicates (Variance Unique) | 144 |
| Removal of Duplicates and Pearson Correlation Coefficient (Unique No Corr) | 54 |
| Variance Threshold, Removal of Duplicates, and Pearson Correlation Coefficient (Variance Unique No Corr) | 53 |
| Analysis of Variance (ANOVA) | 36 |
| Recursive Feature Elimination with elastic net (ENET) as Black-box estimator (RFE ENET) | 5 |
| Recursive Feature Elimination with Ridge as Blackbox estimator (RFE Ridge) | 5 |
| Recursive Feature Elimination with LASSO as Blackbox estimator (RFE LASSO) | 5 |
| Random Forest | 8 |
| Extreme Gradient Boosting (XGBoost) | 5 |
| Principal Component Analysis (PCA) | 50 |
| Least Absolute Shrinkage and Selection Operator (LASSO) | 4 |
| Elastic Net (ENET) | 5 |

Table 1: Feature selection techniques used to test the hypothesis H1.

| | NO FS | UNIQUE | VARIANCE | NO CORR | MIG | ANOVA |
|-------------|-------|--------|----------|---------|------|-------|
| ENET | 0,60 | 0,60 | 0,60 | 0,43 | 0,50 | 0,60 |
| Bayes Ridge | 0,43 | 0,44 | 0,43 | 0,48 | 0,75 | 0,70 |
| LASSO | 0,62 | 0,62 | 0,62 | 0,44 | 0,52 | 0,62 |
| LSTM | 0,84 | 0,85 | 0,83 | 0,63 | 0,81 | 0,81 |
| GRU | 0,85 | 0,84 | 0,87 | 0,64 | 0,77 | 0,82 |
| Ridge | 0,65 | 0,65 | 0,65 | 0,55 | 0,47 | 0,72 |
| SVR | 0,88 | 0,90 | 0,88 | 1,58 | 0,58 | 0,49 |
| BiLSTM | 0,86 | 0,85 | 0,85 | 0,77 | 0,79 | 0,80 |
| RFR | 0,72 | 0,70 | 0,73 | 0,69 | 0,67 | 0,70 |
| XGB | 0,75 | 0,73 | 0,74 | 0,68 | 0,70 | 0,67 |
| MLP | 0,80 | 1,04 | 0,88 | 0,74 | 0,79 | 1,09 |

Table 2: The MASE of the univariate filter feature selection methods compared with not applying any feature selection.

| | NO FS | RFE Ridge | RFE ENET | RFE LASSO |
|-------------|--------------|------------------|-----------------|------------------|
| ENET | 0,60 | 0,44 | 0,590 | 0,590 |
| Bayes Ridge | 0,43 | 1,08 | 0,783 | 0,783 |
| LASSO | 0,62 | 0,44 | 0,585 | 0,585 |
| LSTM | 0,84 | 0,47 | 0,450 | 0,454 |
| GRU | 0,85 | 0,72 | 0,463 | 0,484 |
| Ridge | 0,65 | 1,06 | 0,762 | 0,762 |
| SVR | 0,88 | 1,06 | 0,786 | 0,786 |
| BiLSTM | 0,86 | 0,58 | 0,507 | 0,504 |
| RFR | 0,72 | 0,66 | 0,569 | 0,566 |
| XGB | 0,75 | 0,77 | 0,605 | 0,605 |
| MLP | 0,80 | 0,74 | 0,803 | 0,773 |

Table 3: The MASE of the multivariate wrapper feature selection methods compared with not ap-plying any feature selection.

| | NO FS | EM ENET | EM LASSO | XGBoost | Random Forest |
|-------------|--------------|----------------|-----------------|----------------|----------------------|
| ENET | 0,60 | 0,60 | 0,60 | 0,51 | 0,53 |
| Bayes Ridge | 0,43 | 0,78 | 0,78 | 0,71 | 0,66 |
| LASSO | 0,62 | 0,62 | 0,62 | 0,49 | 0,52 |
| LSTM | 0,84 | 0,86 | 0,85 | 0,65 | 0,60 |
| GRU | 0,85 | 0,85 | 0,87 | 0,65 | 0,66 |
| Ridge | 0,65 | 0,63 | 0,63 | 0,71 | 0,72 |
| SVR | 0,88 | 0,77 | 0,77 | 0,85 | 0,71 |
| BiLSTM | 0,86 | 0,87 | 0,85 | 0,83 | 0,80 |
| RFR | 0,72 | 0,77 | 0,76 | 0,78 | 0,80 |
| XGB | 0,75 | 0,77 | 0,77 | 0,70 | 0,67 |
| MLP | 0,80 | 0,74 | 0,75 | 0,82 | 0,95 |

Table 4: The MASE of the multivariate embedded feature selection methods compared with not applying any feature selection.

| | NO FS | Filter | Wrapper | Embedded | PCA |
|-------------|--------------|---------------|----------------|-----------------|------------|
| ENET | 0,00868 | 0,0045 | 0,0022 | 0,0040 | 0,0028 |
| Bayes Ridge | 0,02612 | 0,0132 | 0,0037 | 0,0034 | 0,0064 |
| LASSO | 0,00771 | 0,0042 | 0,0022 | 0,0043 | 0,0026 |
| LSTM | 15,75152 | 15,3353 | 17,5586 | 15,4537 | 12,6019 |
| GRU | 15,31274 | 15,0273 | 17,7250 | 19,4560 | 11,8673 |
| Ridge | 0,00675 | 0,0063 | 0,0041 | 0,0047 | 0,0038 |
| SVR | 0,25344 | 0,2136 | 0,0595 | 0,0637 | 0,1643 |
| BiLSTM | 28,08850 | 44,5586 | 83,3304 | 47,6442 | 16,8032 |
| RFR | 2,56104 | 1,4069 | 0,2141 | 0,8185 | 1,7519 |
| XGB | 0,05348 | 0,0146 | 0,0034 | 0,0092 | 0,0144 |
| MLP | 1,04305 | 0,8435 | 0,3477 | 0,3087 | 0,7577 |

Table 5: The computation time, in milliseconds, for different feature selection methods, PCA, and, original data without applying any feature selection.

| | NO FS | Filter | Wrapper | Embedded | PCA |
|-------------|--------------|---------------|----------------|-----------------|------------|
| ENET | 147 | 75 | 5 | 17 | 50 |
| Bayes Ridge | 147 | 75 | 5 | 6 | 50 |
| LASSO | 147 | 75 | 5 | 6 | 50 |
| LSTM | 93463 | 72528 | 51918 | 35783 | 65084 |
| GRU | 50326 | 39054 | 27956 | 19268 | 35045 |
| Ridge | 147 | 75 | 5 | 6 | 50 |
| SVR | 71177 | 52859 | 34825 | 35337 | 46345 |
| BiLSTM | 163561 | 126925 | 90857 | 62621 | 113897 |
| RFR | 38744 | 38744 | 38744 | 38744 | 38744 |
| XGB | 147 | 75 | 5 | 21 | 50 |
| MLP | 53769 | 40030 | 26505 | 26889 | 35145 |

Table 6: A comparison of the number of parameters in different machine learning methods with feature selection methods, PCA, and, the original data without applying any feature selection.

| | NO FS | PCA |
|-------------|--------------|------------|
| ENET | 0,60 | 0,51 |
| Bayes Ridge | 0,43 | 0,78 |
| LASSO | 0,62 | 0,52 |
| LSTM | 0,84 | 0,65 |
| GRU | 0,85 | 0,65 |
| Ridge | 0,65 | 0,70 |
| SVR | 0,88 | 0,77 |
| BiLSTM | 0,86 | 0,65 |
| RFR | 0,72 | 0,77 |
| XGB | 0,75 | 0,76 |
| MLP | 0,80 | 0,76 |

Table 7: The MASE of the multivariate dimension reduction techniques, PCA, compared with not applying any feature selection.