The Effects of Filter, Wrapper, and Embedded Feature Selection on Forecast Accuracy of the South African Unemployment Rate

Bohlokoa Tilo, Rudzani Mulaudzi, and Ritesh Ajoodha

School of Applied Mathematics and Computer Science
University of the Witwatersrand, Johannesburg, South Africa
1852529@students.wits.ac.za

Abstract. The representation of macroeconomic data usually uses too many features, while only a few are relevant to the target. Forecasting using high-dimensional data without feature selection can be challenging for machine learning models. Thus, this study examines the effects of filter, wrapper, and embedded feature selection techniques on the forecast accuracy of the South African unemployment rate using data from the SARB. The experiments consisted of 15 feature selection techniques and 3 regression models. MAPE served as the evaluation metric. The filter, embedded, and wrapper methods, respectively, increased model performance by 41%, 32%, and 22% on average. Overall, feature selection methods offered a more accurate way to predict the South African unemployment rate with fewer features. Thus, it would be worthwhile for economic forecasters to consider incorporating these methods into future forecasts.

Keywords: feature selection, forecast, unemployment rate, high-dimensional

1 Introduction

Forecasting the unemployment rate with as much accuracy as possible could aid economic decision-making and policy formulation so that the root causes can be identified and reduced. Thus, the purpose of this paper is to study the effects of feature selection methods on the forecast accuracy of the South African unemployment rate versus using all marcoeconomic factors.

The representation of macroeconomic data usually uses high volumes of factors, while only a few are relevant to the target concept. In theory, more features should allow more information to be stored. In practice, however, more features are rarely beneficial because real-world data usually contains noise and redundancy. Feature selection is crucial in macroeconomic forecasting since it aids in comprehending the data, reducing processing requirements, reducing the influence of the curse of dimensionality, and improving predictor performance [3] [9]. Prior studies have shown that using feature selection methods on high-dimensional, macroeconomic data typically produced more accurate unemployment rate forecasts [12] [9] [13]. However, the majority only experimented with one regularization embedded model to evaluate its performance compared to traditional models and not solely as a method for feature selection [9] [13]. Mulaudzi [12] addressed this by demonstrating the benefits of applying feature selection methods within the filter, wrapper, and embedded domains on high-dimensional, macroeconomic data. However, to the best of the authors' knowledge, none of the existing literature assesses the factors prevalent among feature selection techniques and how they collectively affect model performance. By examining the features that are most common across feature selection methods, we can better understand what macroeconomic factors affect the unemployment rate in South Africa. Additinally, finding a benchmark feature set can help to improve the forecasting process.

As such, this study used feature selection methods in the filter, wrapper, and embedded categories on high-dimensional macroeconomic data from the South African Reserve Bank (SARB). Among the feature selection techniques are univariate statistical tests, information-theoretic measures, and greedy search forward elimination. The research evaluated the performance of regression models on the full dataset and compared the results to their performance on the reduced feature sets. Furthermore, the paper examined the features selected from the best-performing methods. The performance was evaluated using a standard forecast measure for time-series forecasting.

The embedded methods exhibited the highest decrease in prediction error. The results also identified three macroeconomic factors that showed significant influence on the South African unemployment rate. While there is no definitive answer to which feature selection technique is best suited for predicting the unemployment rate, a benchmark can be provided.

Therefore, this research (i) shows how feature selection techniques within the filter, wrapper, and embedded domains can help improve the accuracy of the South African unemployment rate forecasts, (ii) demonstrates that only a few features are relevant for forecasting the South African unemployment rate, and (iii) identifies an ideal feature set for each model used.

The remainder of the sections are organized as follows. Section 2 covers the various studies that have done similar research on the extent of feature selection methods for improving the forecast accuracy of the unemployment rate. Section 3 provides details on the data, forecast measure, feature selection approaches, and models used for the research. Section 4 goes through the results and discussion of the experiments. Lastly, Section 5 provides a summary of the research, its contributions, and consideration for future research.

2 Related Work

The purpose of this section is to examine how filter, wrapper, and embedded methods have been previously applied to macroeconomic, timeseries data to predict unemployment rates.

2.1 Filter Methods

There has not been extensive use of filter methods in timeseries prediction. However, in the few kinds of literature that have explored them, it has been shown that filters that identify only the relevant timeseries subset are more efficient in feature selection [5]. Mulaudzi [12] used data from the South African Reserve Bank (SARB) with 147 features and the unemployment rate as the target label to demonstrate how feature selection methods could improve the forecast accuracy of the South African unemployment rate. The data was considered high dimensional due to the large number of features [12]. Using this data, Mulaudzi [12], estimated that the filter methods resulted in 28% more accuracy than not using feature selection. His research revealed how the filter methods could be used to determine which macroeconomic features influence the South African unemployment rate. According to his experiments, the greatest gain in model performance (MASE reduction) was obtained using filter techniques. He established that this is most likely due to the non-stationary character of the macroeconomic factors. Mulaudzi [12], also claims filter feature selection approaches to be excellent for reducing noise. Therefore, the results from this paper indicated that though many variables may be used to predict the South African unemployment rate, only a few were valuable.

2.2 Wrappper Methods

Unlike filter methods, wrapper methods include the model as part of the feature subset search. In this manner, several subsets of features are discovered and assessed by the model. However, the availability of processing capacity limits the use of wrappers. Therefore, similar literature that employed wrapper approaches restricted them to select at most five features [12] [3].
Mulaudzi [12] used the Recursive Feature Elimination (RFE) wrapper method to find macroeconomic variables that were most predictive of the South African unemployment rate. Due to the computing costs of the technique, the study limited the search to only five features and used regression learning models (i.e., SVR, LASSO) as the black-box models. The results from this study indicated that when compared to not using feature selection, using wrapper methods produced predictions that were generally 21% more accurate.
In their paper on which factors influence the prediction of the movements of stock markets, Altinbas and Biskin [3] employed the Sequential Forward Selection (SFS) algorithm on the Turkish stock market: Borsa Istanbul data. The results showed that using the initial 17 macroeconomic variables gave an error of 10.19%, whereas, with those selected by the SFS, the error went down to 9.6%. They found that month-lagged index values were sufficient to predict the future value of a market indicator index without the inclusion of other features. Though the paper does not focus on predicting the unemployment rate, it demonstrated that using a wrapper method on macroeconomic variables can improve the predictive accuracy by selecting only the necessary features.

2.3 Embedded Methods

Several works of literature in South Africa and the USA have examined the effects of using embedded methods for forecasting the unemployment rate. Mulaudzi [12] demonstrated how various embedded methods increased the predictive accuracy of the South African unemployment rate. The approaches consisted of the Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net (ENET), Extreme Gradient Boosting (XGB), and Random Forest. These approaches yielded a 15% average accuracy increase. According to the study, they provided the lowest accuracy improvement compared to the wrapper and filter methods.

Hall [13] wrote a similar paper, which focused on how using machine learning approaches on high-dimensional macroeconomic data improved the forecast accuracy of the USA unemployment rate. The study used Federal Reserve Economic Data (FRED) which consisted of 138 macroeconomic variables from diverse economic categories. The overall data collection included 698 monthly observations from March 1959 to April 2017. Hence, the data was also considered high-dimensional due to the large feature set [12].

However, unlike Mulaudzi [12] who employed several embedded models, Hall [13] only evaluated one regularized machine learning model, ENET, for his research. He showed particular interest in the ability of the model to balance issues regarding bias and variance using regularization. That is, ENET can select groups of correlated variables, as well as perform automated variable selection and continuous shrinkage at the same time [13]. The results of this study demonstrated that ENET was capable of producing forecasts that were more accurate than the traditional time-series models (i.e., the Autoregressive model). The ENET model outperformed the Blue Chip by 0.07% on average. Additionally, 10 out of 138 macroeconomic variables were deemed relevant for forecasting the unemployment rate in the USA using the ENET method.

Kriener and Duca [9] used 588,967 macroeconomic variables from the FRED to predict the unemployment rate in the USA in a similar way to Hall [13]. The data ranges from 1970 to 2018. This study investigated the performance gains of using LASSO and a neural network model against econometric approaches when predicting the USA unemployment rate using high-dimensional data. Due to the enormous number of variables from FRED, the study employed Principal Components Analysis (PCA) to reduce the data dimensionality before training the models on the data. After feature reduction, the resulting FRED feature set had 10,646 variables from the initial 588,967. Despite having the best computation time, when used with the South African macroeconomic data, PCA gave poor accuracy performance compared to not using feature selection in half of the cases [12]. Therefore, it provided no performance benefits and was deemed unnecessary in the case of South African unemployment rate forecasting [12].

Overall, LASSO outperformed the econometric approach, though not to a large extent as the average difference between their Root Mean Square Errors (RSME) was less than 0.2.

A number of approaches were used in previous literature to forecast the unemployment rate. However, due to the possibility that timeseries with similar frequency and domain might show different patterns, it is advisable to develop an automatic, data-driven method for feature evaluation and selection that does not require human input. Thus, this study focused on those papers that employed machine learning models and data-driven feature selection techniques. This section reviewed how filter, wrapper and embedded feature selection methods have been used to improve forecast accuracy of the unemployment rates in various countries.

3 Experiment Design

3.1 Data Preparation

The study used data retrieved from a GitHub repository [1]. Previous, similar papers also used this data [12] [11]. The data initially came from the SARB with 147 macroeconomic predictors, and the unemployment rate served as the target variable. There was a total of 794 observations ranging from January 1970 to January 2020. As a result of the data having different frequencies (monthly versus quarterly), there were considerably more missing data in the consolidated dataset. The method employed to impute the missing data used the last known value before the missing value. Macroeconomics generally refers to GDP as the most recent value, so the imputation technique seemed logical [12]. The data were then resampled to a quarterly frequency since the unemployment rate records are quarterly. The final dataset had 201 observations.
It is imperative to know how well a model generalizes. For this reason, we split the data into training and testing subsets to properly evaluate the models and avoid overfitting. The training set contained 80% of the observations (i.e., 1970 - 2009), and the testing set was the remaining 20% (i.e. 2010 - 2020). Additionally, the data values were scaled and set between 0.000001 and 1 due to the wide range of the original values, which could have skewed the results of some of the models and feature selection methods.

3.2 Performance Measure

This paper used mean absolute percentage error (MAPE) to evaluate model performance. MAPE is a conventional measure for determining the comparative accuracy of time-series forecasts [7]. Moreso, it has the ability to be scaled and compared across various time-series. Given that $\hat{y}_i$ is the forecast value of the i^{th} sample and y_i is the analogous true value, then the MAPE estimated over $n_{samples}$ is defined as:

$$mape(y, \hat{y}) = \frac{1}{n_{samples}} \Sigma_{i=0}^{n_{samples}-1} \frac{|y_i - \hat{y}_i|}{max(\epsilon, |y_i|)} \tag{1}$$

where ϵ represents an arbitrarily small yet strictly positive number to avoid undefined results when $y = 0$. It is worth noting that MAPE values are not expressed as a percentage in the range [0, 100]. Rather, equation 1 represents them in the range [0, 1/eps].

3.3 Methodology

The objective of the experiments was to evaluate the influence of feature selection methods on model performance by comparing the forecast measure (MAPE) of the models when feature selection was used and when it was not used. As well as to examine which features are prevalent among the feature selection methods. For further information on the features see [1]. To achieve the objective, the research employed 15 feature selection methods in conjunction with three regression models. This section provides details on the experiments.

Regression Models The regression models considered for this research were the support vector regression (SVR), Bayesian ridge regression (Bayes Ridge), and multilayer perceptron (MLP). These models are among some of the models commonly used for regression problems [5] [11]. The models were not optimized so as to fully explore the impact of the feature selection methods on their performance. The models were adapted from [2]. The MLP architecture was kept consistent throughout the experiments for reasons of simplicity and replication. Moreover, Bayes Ridge was used only as a regression model as no sources could be found on its capabilities as an independent feature selection method.

Filter Methods The filter techniques used were Pearson Correlation Coefficient Criteria (p_corr), Regression F-Score (f_reg), and Mutual Information (MI). These are among the filter techniques typically used for regression data [12] [6]. However, the p_corr approach selects features that are highly correlated with the target without considering the correlation between the features themselves. The predictive power of each feature is determined by a correlation criterion, which is measured using the Pearson Correlation Coefficient and it is defined as:

$$r(j) = \frac{cov(x_j, y)}{\sqrt{var(x_j)var(y)}} \tag{2}$$

where $cov()$ and $var()$ represent the covariance and variance, respectively. The selected features being those that had a correlation > 0.8 with the unemployment rate (target). Thus, p_corr was applied to the selected features to address the possibility of high intercorrelations among the independent variables where if two variables were highly correlated, one was dropped. Since this method removes multicollinearity (no_colin) it served as the fourth filter technique. Where applicable, the correlation and variance thresholds were 0.8, and 0.01, respectively. MI was set to select the top five ranked features because that was determined to be the optimal number through the process of trial and error.

Filter Method	No. of Selected Features
f_reg	126
p_corr	26
MI	5
no_colin	4

Table 1. The resulting number of features from each filter feature selection technique.

Wrapper Methods The wrapper techniques explored were the Recursive Feature Elimination (RFE), RFE with Cross-Validation (RFECV), Sequential Forward Selection (SFS), and Boruta. SVR, Bayes Ridge, and Random Forest Regression (RFR) were used as the black-box models for the wrapper methods. RFR was used with Boruta as this wrapper method was built around the random forest model [10] and thus, it proved to be the only appropriate black-box model. The Boruta algorithm was designed to eliminate features that are statistically insignificant compared to a random probe (artificial noise variables introduced by the algorithm). It is said to usually yield a decent global optimization for feature selection [10].
Since wrapper methods are computationally heavy, where applicable, they were restricted to select five or fewer features. For wrappers such as RFECV and Boruta, however, there is no direct way to limit the maximum number of features selected. As such, the number of features varied for the models. RFECV and Boruta seem impractical due to their lack of restrictions on the number of selected features, but they still offer significant improvements in accuracy for some models and run at comparable speeds to the other wrappers.

Embedded Methods LASSO, ENET, RFR and XGB served as the embedded feature selection approaches for the experiments. RFR and XGB are decision tree techniques that, as part of their forecasting activity, can provide a feature relevance score [12]. The top performing features were selected by setting a threshold using the variance of the high-ranked features. The features which had an importance value higher than the threshold were selected. The following provides high-level insight into how ENET works. The ENET has ability to perform automatic feature selection and continuous shrinkage, simultaneously [14]. It can also select feature subsets that consist of correlated features. ENET has the following properties: (i) It is a hybrid of the L1 and L2 regularization techniques. (ii) It offers a more efficient regularization technique. (ii) It has two hyperparameters, λ and α. The loss function that ENET minimizes is defined as:

$$Loss(\hat{\beta}) = \frac{\Sigma_{i=1}^{n}(y_i - x_i'\hat{\beta})^2}{2n} + \lambda(\frac{1-\alpha}{2}\Sigma_{j=1}^{m}\hat{\beta}_j^2 + \alpha\Sigma_{j=1}^{m}|\hat{\beta}_j|). \tag{3}$$

Shrinkage refers to how much the weights can be reduced when compared to linear regression. In Equation 3 the quantity of shrinkage is denoted by λ. When

Wrapper Method	Algorithm	No. of Selected Features
RFE	Greedy Backward Elimination	– SVR - 5 – Bayes Ridge - 3
SFS	Greedy Forward Selection	– SVR - 3 – Bayes Ridge - 5
RFECV	Greedy Backward Elimination	– SVR - 10 – Bayes Ridge - 29
Boruta	Random Generation [10]	RFR - 80

Table 2. The resulting number of features from each wrapper feature selection technique with respect to the underlaying models.

Embedded Method	No. of Selected Features
LASSO	3
ENET	9
RFR	2
XGB	4

Table 3. The resulting number of features from the embedded feature selection techniques.

$\lambda = 0$ it suggests that all features were taken into account, and it is equal to linear regression, in which just the RSS is used to form a prediction model. In the case when $\lambda = \infty$, it suggests that no feature was considered, implying that as λ approaches infinity, it excludes more and more features [14]. For the experiments the default sklearn [8] $\alpha = 1.0$ and $\lambda = 0.5$ were used.

4 Results and Discussions

This section provides a discussion on the results attained from the experiments mentioned in Section 3.

4.1 Filter Methods

Table 4 indicates that most of the filter techniques improved the forecast accuracy of the models compared to using all the variables when predicting the South African unemployment rate. SVR and Bayes Ridge showed improvement with each approach, whereas MLP showed improvement with all except p_corr and f_reg. Which seemed to be the case because those approaches do not address multicollinearity. SVR performed best with p_corr indicating a MAPE 70% lower than that of the entire feature set. MLP showed the highest performance gain with no_colin producing a MAPE 75% lower than that of the entire feature set. Bayes performed best with MI showing a 78% decrease from the MAPE of the unfiltered data. According to the results, removing multicollinearity gave the lowest average MAPE of 0.087, which is 77% less than when no feature selection was implemented. Mulaudzi [11] also found this approach to yield the lowest forecast errors in his research. This consistency suggests that recursively dropping highly correlated features from the SARB data is most likely to improve the level of accuracy in predicting the South African unemployment rate. On average the filter feature selection approaches reduced the model forecast errors by 41%.

	All Features	p_corr[a]	f_reg[b]	MI	no_colin[c]
SVR	0.2323	0.0696	0.2255	0.1064	0.0908
MLP	0.3891	0.4604	0.5110	0.2390	0.0374
Bayes Ridge	0.5266	0.1587	0.4876	0.1165	0.1343

[a] Pearson Correlation Coeffiecient criteria
[b] Regression F-Score
[c] No Multicollinearity

Table 4. Comparison of the MAPE of the models with and without filter feature selection.

4.2 Wapper Methods

The impact of wrapper methods on the models varied greatly, and thus none of the approaches consistently improved the models. In terms of performance gain for the models, SVR with SFS, MLP with SVR-SFS, and Bayes with RFECV each showed an improvement of 87%, 91%, and 20%, respectively. Collectively, the wrapper methods had the lowest performance gains compared to the other methods. The poorest results came from Bayes Ridge. Indicating that Bayes Ridge does not train well when used as a black-box model. The wrapper methods exhibited an average decrease in prediction errors of 22%.

	All Features	RFE	RFECV	SFS	Boruta
SVR	0.2323	0.0579	0.0451	0.0296	0.1779
MLP	0.3891	0.1103	0.4749	0.0368	0.4703
Bayes Ridge	0.5266	0.6100	0.4230	1.0580	0.6712

Table 5. Comparison of the MAPE of the models with and without wrapper feature selection.

4.3 Embedded Methods

In the case of embedded methods, all the learning models achieved their best performance improvements with the data subset from the ENET approach. The ENET gave an average decrease in MAPE of approximately 88% across the models. LASSO and XGB exhibited an average reduction in MAPE of 72% and 63%, respectively. RFR was only beneficial to the SVR. The high correlation of the features might explain the poor selection of features in this approach. Decision trees tend to give similar and lowered importance to correlated features compared to what they would if the trees were built without the correlated counterparts [4]. Overall, employing embedded feature selection techniques reduced the error of forecasts by approximately 34% on average. Additinally, ENET gave an average MAPE of 0.045, which was the lowest compared to all 15 feature selection methods employed. Thus, the subset of features from the ENET technique could serve as the baseline for the optimal feature subset for predicting the South African unemployment rate.

	All Features	LASSO	ENET	RFR	XGB
SVR	0.2323	0.1098	0.0291	0.0823	0.0734
MLP	0.3891	0.0880	0.0538	1.0261	0.1234
Bayes Ridge	0.5266	0.1276	0.0534	1.3381	0.2294

Table 6. Comparison of the MAPE of the models with and without embedded feature selection.

4.4 Feature Subsets

We did further analysis on the feature subsets from the feature selection methods that produced the lowest errors. According to Table 7, there are at most five relevant features for predicting the South African unemployment rate, depending on the model run. *Domestic output: All groups, The nominal effective exchange rate, and Total South African population* appear to have a significant impact on the unemployment rate. That is, most subsets that had either one or more of

them resulted in considerably reduced forecast errors. Interestingly, ENET was the only feature selection method that included all three predictors in its subset, which could be the reason for its notable performance. The R^2 score is shown in table 7 to give an indication of goodness of fit and thus a measure of how well unseen data are likely to be predicted by the models, with 1 being the optimal score. Figure 1 depicts how the models performed with their optimal subsets against the valid data, and as can be seen, each model has multiple intersections with the valid data.
To further evaluate how the models perform on their optimal feature sets, we examined the confidence level of the forecasts. The respective means of the predictions from SVR, MLP, and Bayes Ridge are 25.5, 25.3, and 26.4. The actual values have a mean of 25.9. The mean values are close, which indicates that the models closely predict actual values. Accordingly, the standard deviations are 1.46, 1.89, and 2.2 for the SVR, MLP, and Bayes Ridge. Figure 2 illustrates the consistency in the predictions of each model.

Model	Feature Selection Technique	Feature Subset	MAPE	R^2
Support Vector Regression	Union of Sequential Forward Selection & No Multicollinearity features	– Domestic output: All groups – The nominal effective exchange rate – Total South African population – Final consumption expenditure by general government – Foreign exchange rate: SA rand per USA dollar	0.0273	0.6232
Multilayer Preceptron	Custom	– Domestic output: All groups – The nominal effective exchange rate – Total South African population	0.0355	0.3604
Bayesian Ridge Regression	Intersection of ENET& Sequential Forward Selection features	– Domestic output: All groups – Total South African population	0.0426	0.2101

Table 7. Summary of the performance of the models with combinations of the most prevalent features from the ENET, SFS-SVR, and No multicollinearity feature selection methods.

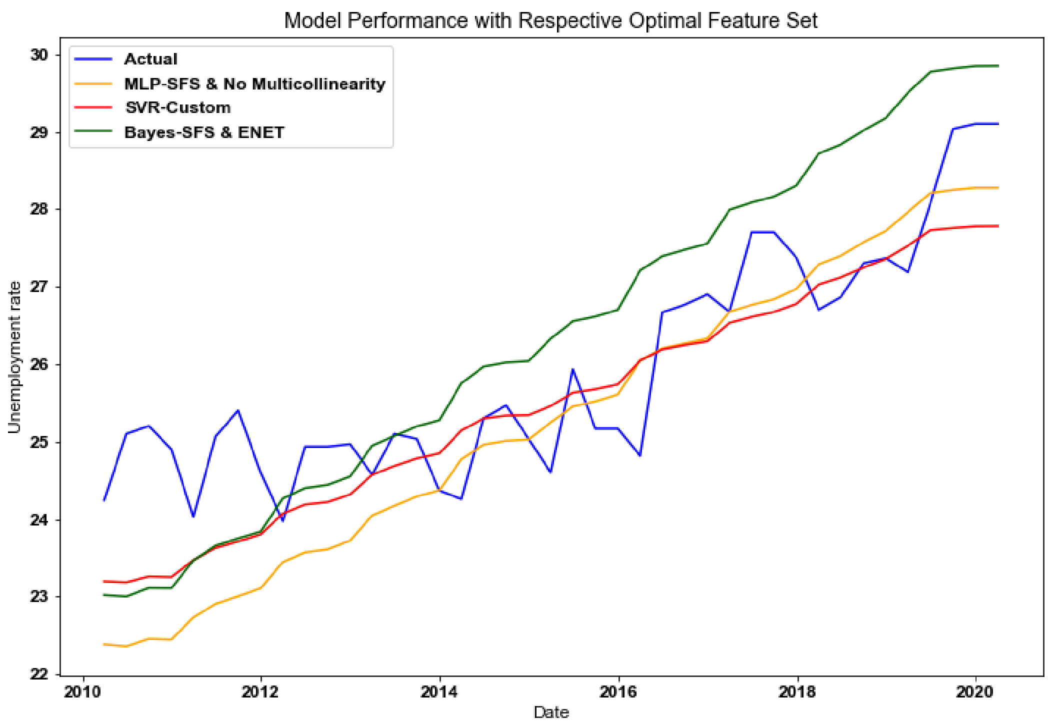

Fig. 1. Graph of the performance of the models with combinations of the most prevalent features from the ENET, SFS-SVR, and No multicollinearity feature selection methods.

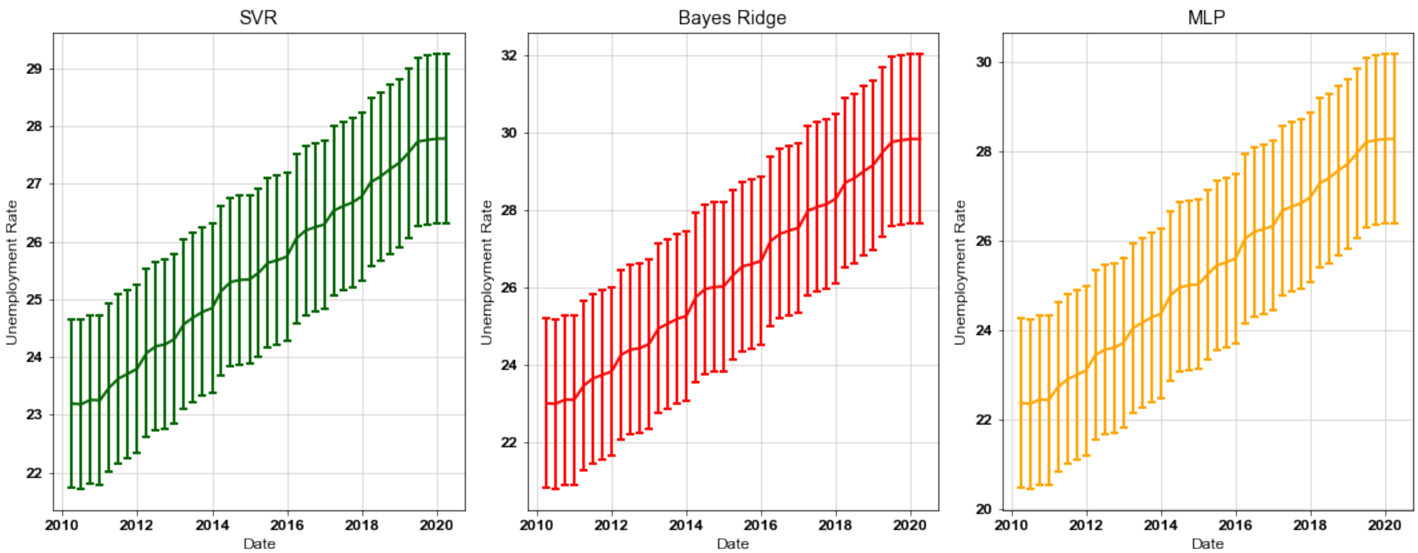

Fig. 2. A depiction of the confidence or precision in the set of predicted values of each model using mean and standard deviation.

5 Conclusion

The purpose of this paper was to evaluate how feature selection techniques based on the data influenced the forecast accuracy of the South African unemployment rate when applied to high-dimensional data. The combinations of the features that were frequently chosen by the feature selection techniques were then studied to find the generally relevant ones.

The results show that feature selection techniques within the filter, wrapper, and embedded domains often result in a lower MAPE. Therefore, using machine learning models with feature selection techniques is likely to produce more accurate unemployment rate forecasts than using all the features. Some of the models showed better improvement than others, but which models perform best depends on the feature selection technique. According to this study, feature selection provides a minimum accuracy gain of 22% on average, which is close to the number indicated in a comparable study [12].

Moreover, there is no one-fits-all solution, but a trend could be established among the macroeconomic factors the feature selection methods typically chose. Mulaudzi [11] hinted at the features that may influence the South African unemployment rate, but much like all prior research, the study did not further evaluate the predictive value of these features. As a result, this paper did an analysis on the most influencial features to the target and indentified *Domestic output: All groups, The nominal effective exchange rate, and Total South African population* as the most prevalent factors across the top performing feature selection

approaches. Our study found that an ideal feature subset for forecasting the South African unemployment rate includes at least two of those features. Thus, applying feature selection techniques to SARB data can aid in accurately predicting the unemployment rate in South Africa with fewer features.
That said, the objective of this paper was to further shed light on the effectiveness of using feature selection methods based on the data when predicting the South African unemployment rate. The study also gave a comprehensive look into which macroeconomic factors were most predictive of the South African unemployment rate. Knowing the factors that influence the unemployment rate could be helpful to South African economists and policymakers. The paper also explored using the Boruta feature selection technique for time-series data which has not been done in any prior literature. Moreover, due to time constraints of the academic year, the research could only explore three models and one performance metric.
Future studies are encouraged to explore employing hybrid approaches to improve the efficiency and accuracy of feature subset selection techniques by combining the best filter and wrapper approaches from the existing literature. A more comparative study should be conducted to obtain more accurate results that account for the influence of the COVID-19 pandemic on South African macroeconomic data.

6 Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant number: 121835).

References

1. South african reserve bank cleaned economic data. https://github.com/rudzanimulaudzi/sarb_feature_set. Accessed: 2021-06-14
2. Supervised learning. https://scikit-learn.org/stable/supervised_learning.html. Accessed: 2021-10-26
3. Altinbas, H., Biskin, O.T.: Selecting macroeconomic influencers on stock markets by using feature selection algorithms. Procedia Economics and Finance **30**, 22–29 (2015)
4. Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M.: Benchmark for filter methods for feature selection in high-dimensional classification data. Computational Statistics and Data Analysis **143**, 106–839 (2020)
5. Crone, S.F., Kourentzes, N.: Feature selection for time series prediction - a combined filter and wrapper approach for neural networks. Neurocomputing **73**, 1923–1936 (2010)
6. Ghojogh, B., Samad, M.N., Mashhadi, S.A.: Feature selection and feature extraction in pattern analysis: A literature review. Tech. rep., Department of Electrical and Computer Engineering, University of Waterloo (2019)
7. Hyndman, R.J., Koehler, A.B.: Business statistics another look at measures of forecast accuracy. International journal of forecasting **22**, 679–688 (2005)

8. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l1-regularized least squares. IEEE Journal on Selected Topics in Signal Processing **1**, 606–617 (2007)
9. Kreiner, A.S.: Can machine learning on economic data better forecast the unemployment rate? Tech. rep., Oberlin College (2019)
10. Kursa, M.B., Rudnicki, W.R.: Feature selection with the boruta package. Journal of Statistical Software **36**, 1–13 (2010)
11. Mulaudzi, R., Ajoodha, R.: Application of deep learning to forecast the south african unemployment rate: A multivariate approach. In: 7th The Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) 2020. IEEE (2020)
12. Mulaudzi, R., Ajoodha, R.: Demonstration that the use of feature selection on high dimensional south african macroeconomic data results in improved performance with lower compute requirements. In: Sixth International Congress on Information and Communication Technology Conference (ICICT) 2021. SPRINGER (2021)
13. Smalter Hall, A.: Machine learning approaches to macroeconomic forecasting. Tech. rep., The Federal Reserve Bank of Kansas City (2018)
14. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology **55** (2020)