Predicting maize crop yields in Eswatini using multiple linear regression and backward elimination

Johan van den Burg
University of Eswatini

Introduction

Why Maize?

- The most produced and consumed grain in the world, however, apart from South Africa commercial production in the rest of Africa is relatively very low (International Grains Council, 2022).
- Research into improving crop yields has been done in other parts of the world but, again, much less has been done in Africa. This is important because we have different soil types, climate patterns, crop varieties etc.
- Over 70% of Eswatini population relies on subsistence farming, most of it maize (World Food Programme 2021; U.S. Department of Commerce 2018).
- 63% of the population lives below the poverty line (United Nations, 2018).

Currently in Eswatini, published research into applying modern technologies to solve problems is in its infancy, particularly in crucial industries such as agriculture and so there is a need to create a method to implement machine learning (ML) with regression analysis to maximize maize crop yields in the region to help the industry progress, and allow small-scale and commercial farmers to benefit from maximized yields.

This research project aimed to design and develop a supervised machine learning model to determine the maximum yield of maize grown in a single season in Eswatini using multiple linear regression and backward elimination.

This was achieved by:

1. Designing a set of machine learnings tasks utilizing regression analysis to determine the yield of crops given various environmental factors.
2. Using open source datasets of the various factors and typical growth data of the crops in order to train the model.
3. Using local data from the region to train and test the model in order to determine the yield for crops in the region.

Methodology

Determining factors (7 features) affecting yield (response variable), and applying regression analysis and supervised ML

- Water input (rainfall and irrigation), Nitrogen, Crop evapotranspiration, Area, Average air temperature, Vapor pressure, Solar radiation.

Data Preprocessing

- Open source data obtained from experiments carried out by the USDA-Agricultural Research Service (Trout & Bausch, 2017; Comas, et al., 2018).
- Local data for crop yield, land use, and fertilizer use variables came from the Ministry of Agriculture in Eswatini. The Centre for Environmental Data Analysis (UK) for satellite data showing precipitation, evapotranspiration, temperature, and vapor pressure, and the free solar radiation data gathered by satellites and provided by SoDa's (www.soda-pro.com) HelioClim-1 web service.

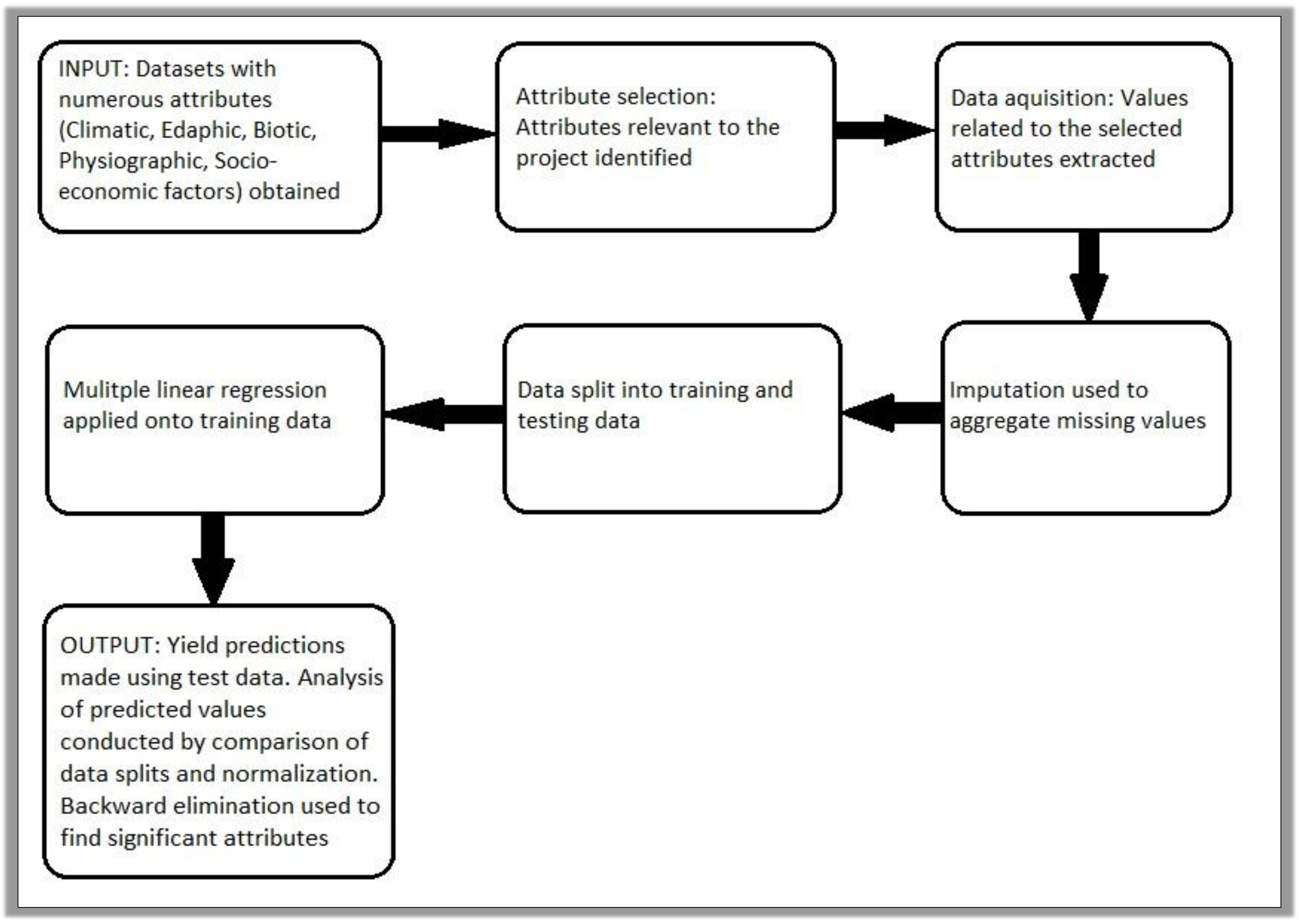
Multiple linear regression modeled using the formula:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Evaluated using:

- $RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$,
- $R^2 = 1 - \frac{SS_{regression}}{SS_{total}}$,
- $Adjusted\ R^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-(k+1)}\right)$.

Framework used:



INPUT: Datasets with numerous attributes (Climatic, Edaphic, Biotic, Physiographic, Socio-economic factors) obtained

Attribute selection: Attributes relevant to the project identified

Data aquisition: Values related to the selected attributes extracted

Imputation used to aggregate missing values

Data split into training and testing data

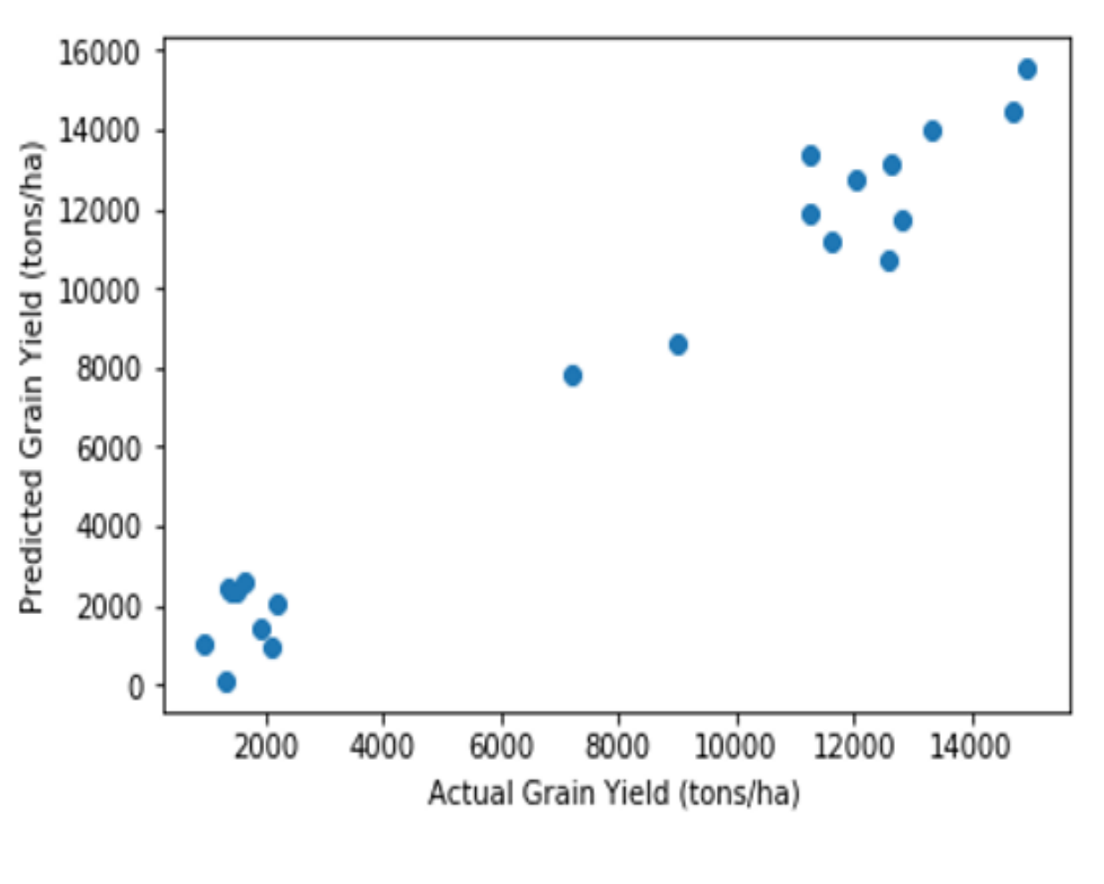
Mulitple linear regression applied onto training data

OUTPUT: Yield predictions made using test data. Analysis of predicted values conducted by comparison of data splits and normalization. Backward elimination used to find significant attributes

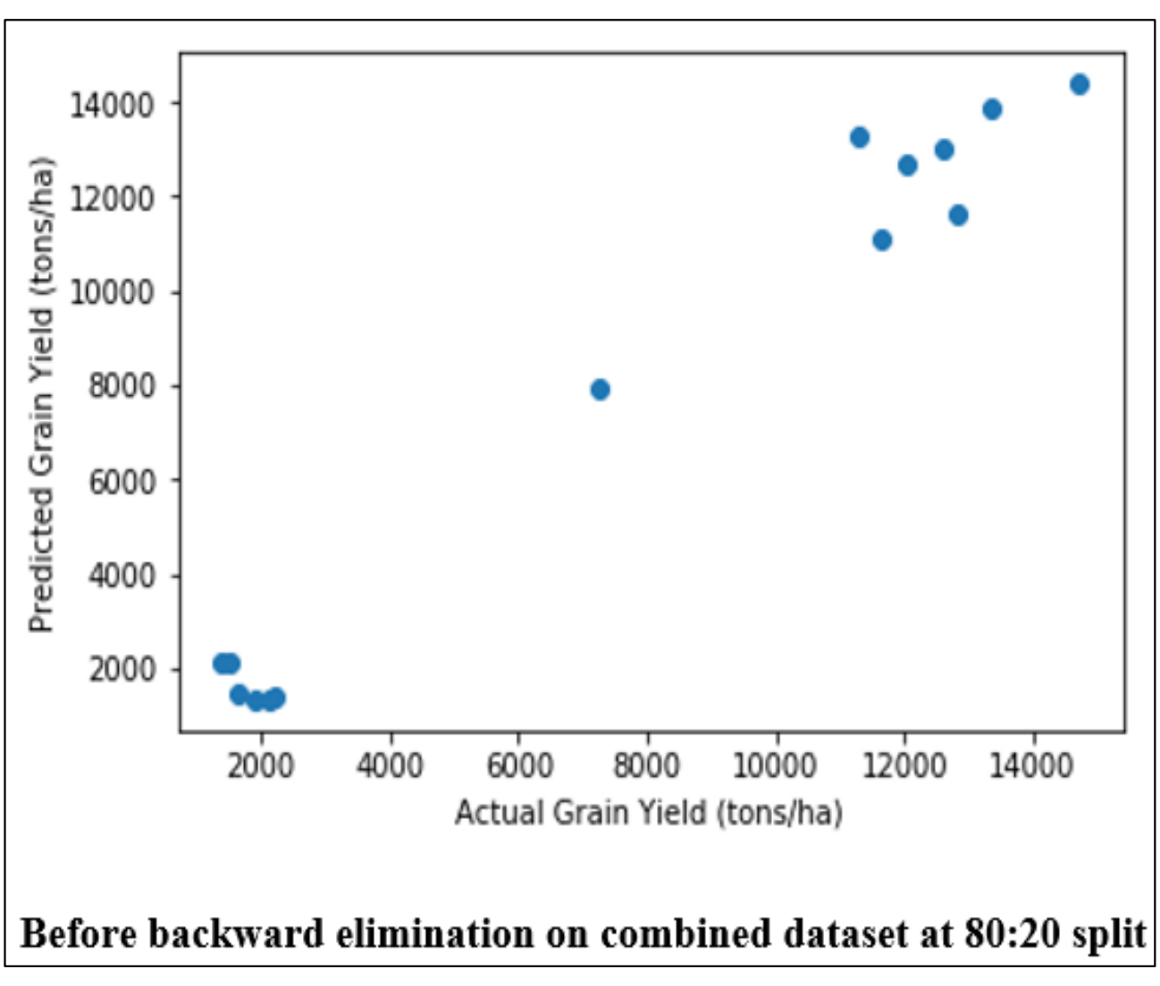
Results

Before normalisation and backward elimination:

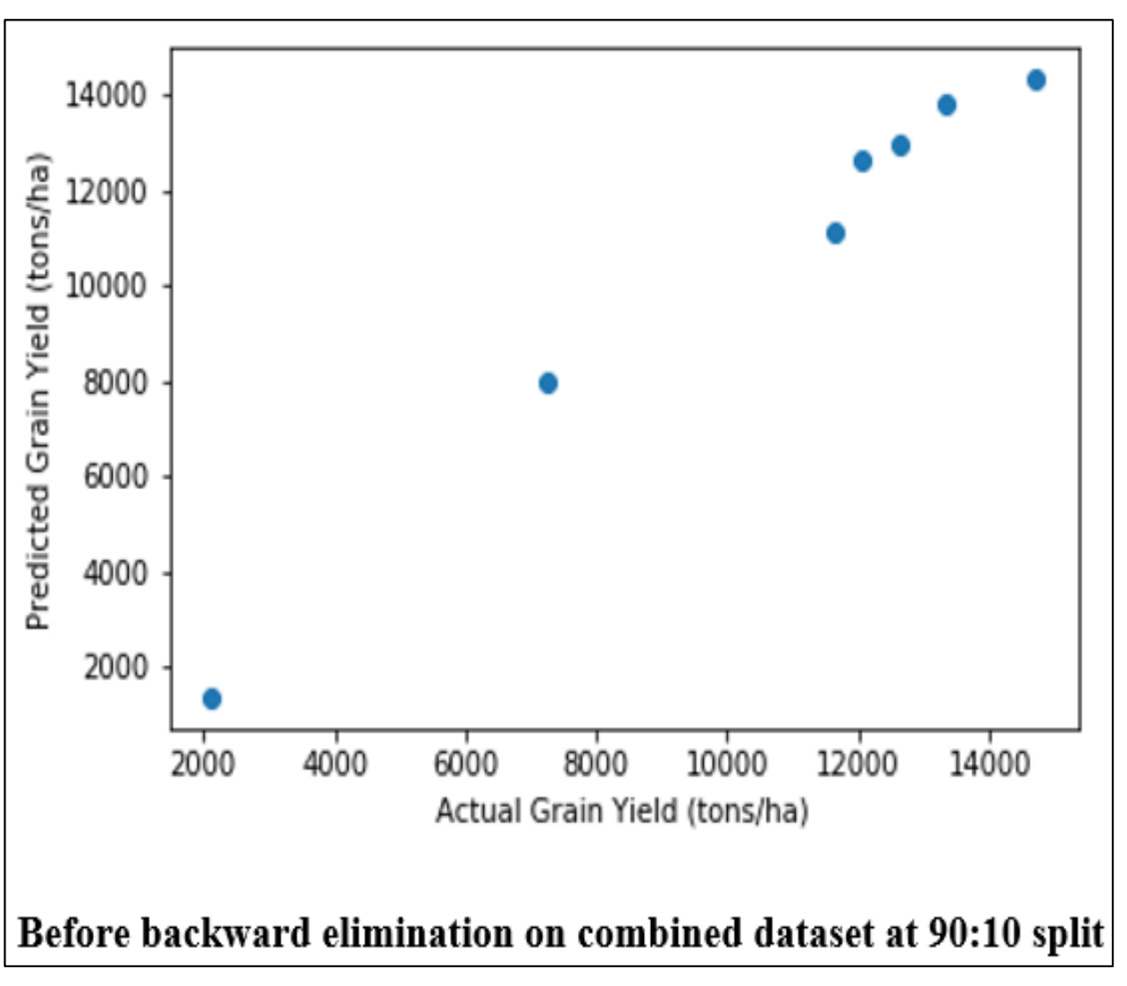
Test size	Train size	R²	Adjusted R²	RMSE
30%	70%	0.970	0.966	934.21
20%	80%	0.980	0.972	828.53
10%	90%	0.980	0.978	569.95



Before backward elimination on combined dataset at 70:30 split



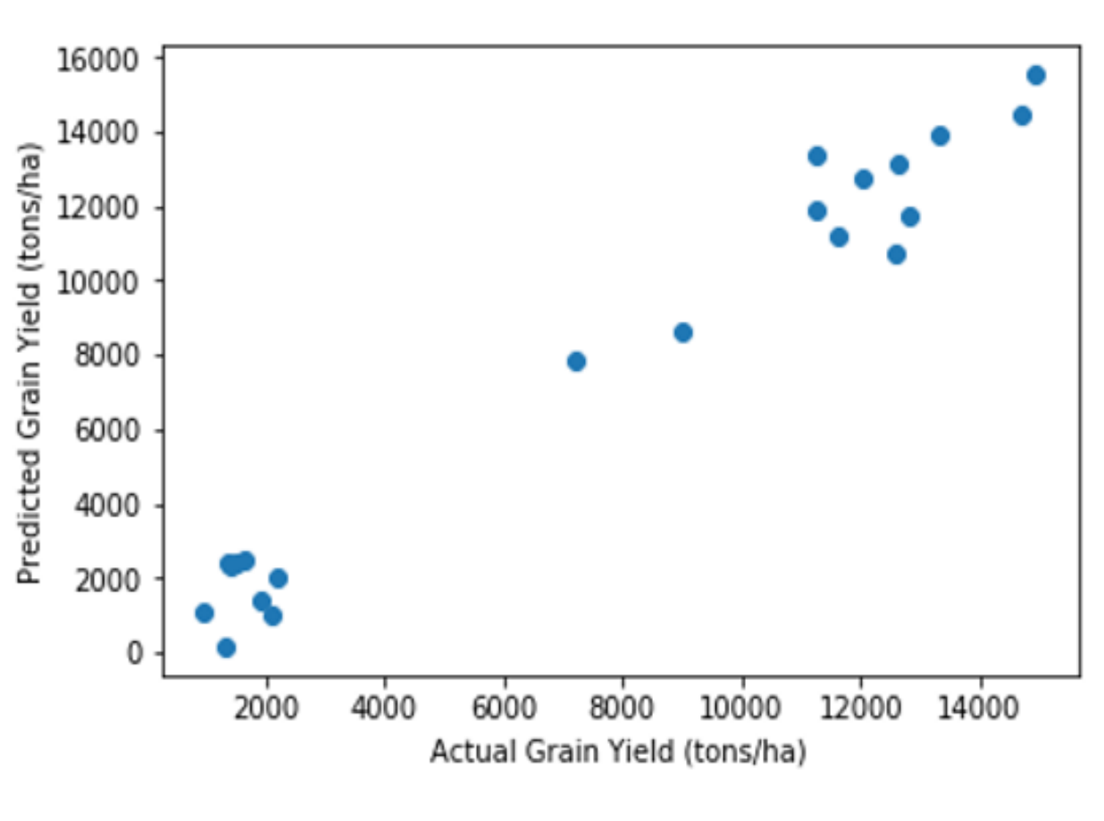
Before backward elimination on combined dataset at 80:20 split



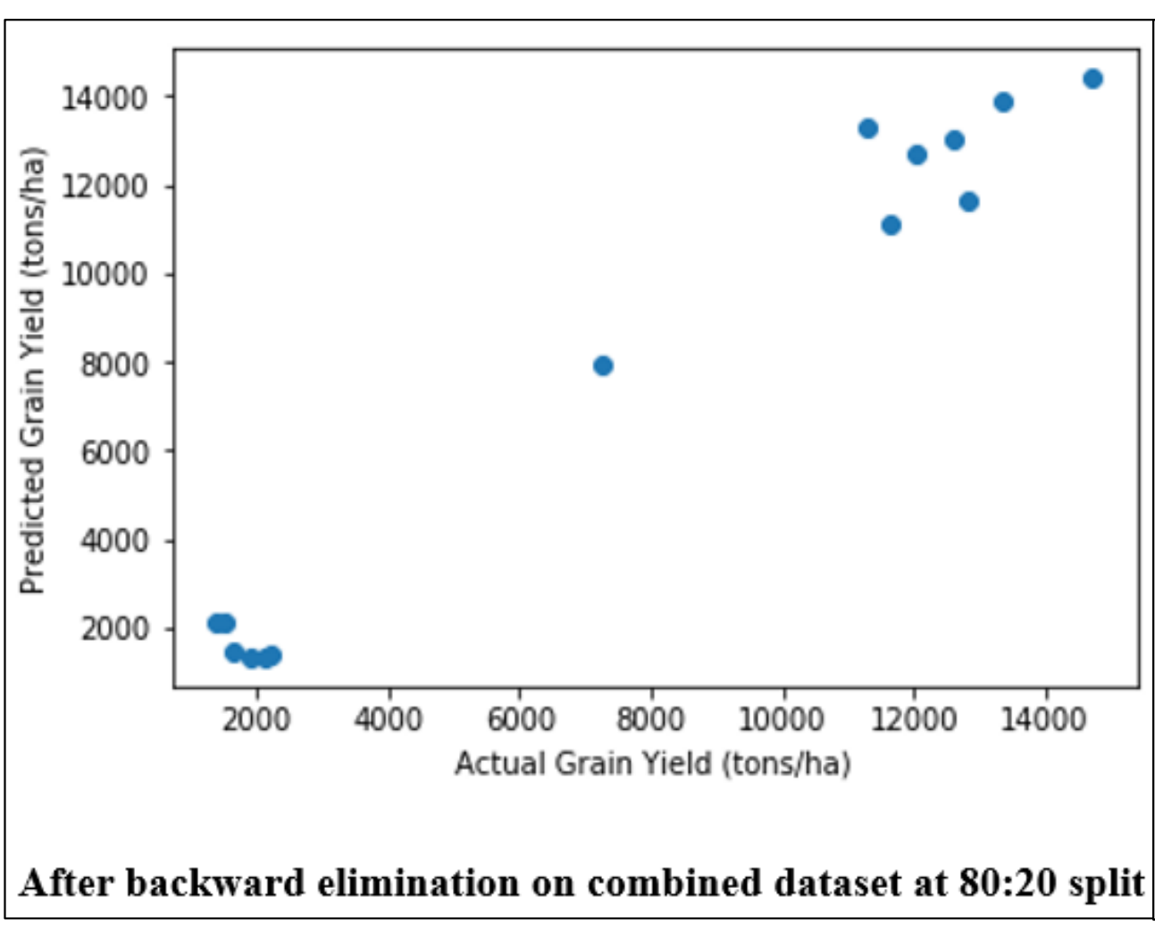
Before backward elimination on combined dataset at 90:10 split

After backward elimination:

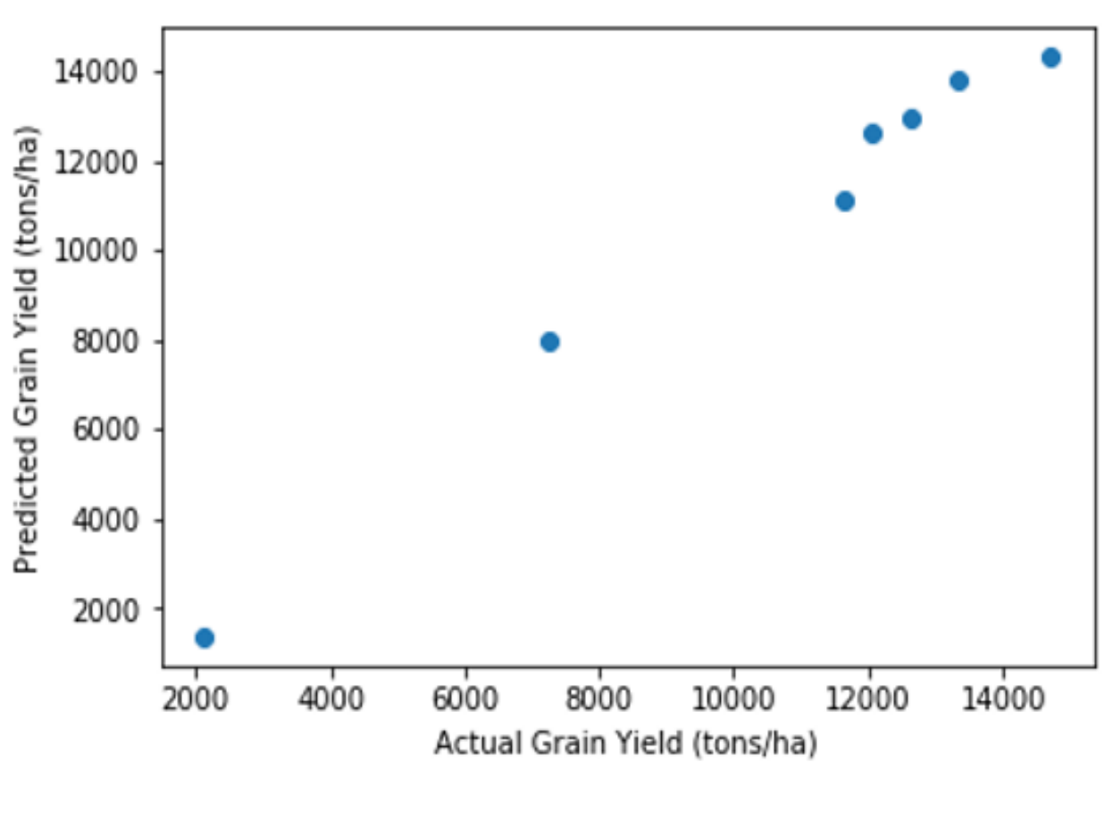
Test size	Train size	R²	Adjusted R²	RMSE
30%	70%	0.970	0.967	925.88
20%	80%	0.980	0.973	826.71
10%	90%	0.980	0.978	568.45



After backward elimination on combined dataset at 70:30 split



After backward elimination on combined dataset at 80:20 split



After backward elimination on combined dataset at 90:10 split

After normalisation:

Test size	Train size	R²	Adjusted R²	RMSE
30%	70%	1.00	1.00	0.00
20%	80%	1.00	1.00	0.00
10%	90%	1.00	1.00	0.00

Conclusions & Recommendations

Conclusions:

- The R² and adjusted R² values found both before and after backward elimination show a strong correlation between the predictors used in the model and the response variables.
- The use of normalization helped to improve the Root Mean Squared Error score, however, the results obtained did not give a clear picture of how well normalization helped.
- From the use of backward elimination, it was shown that the predictor variable that affected the yield prediction the least was the amount of fertilizer used, while the factors that affected the prediction the most were water input, relative humidity, and solar radiation.
- Limitations observed from implementation included a relatively small dataset, and greater processing power required.

Recommendations:

- Larger training and testing datasets
- Using more regionally-specific data to predict yields for certain areas in the country
- Look into the effect of using different maize varieties in different regions of the country
- Incorporating the model into an application which could be used by small-scale and commercial farmers

References

- International Grains Council. (2022). Grain Market Report 528. International Grains Council.
- World Food Programme. (2021). WFP Eswatini Country Brief. World Food Programme.
- U.S. Department of Commerce. (2018, October 26). Swaziland - Agricultrual Sector. Retrieved from Export.gov: https://www.export.gov/article?id=Swaziland-Agricultural-Sector
- United Nations. (2018, October 13). Challenges. Retrieved from United Nations in Swaziland: http://sz.one.un.org/content/unct/swaziland/en/home/about-the-country/challenges.html