# Highlights

**An analysis of interrelations between economic and environmental variables in Australian Winegrowing.**

Author

- ???
- ???
- ????
- ????

# An analysis of interrelations between economic and environmental variables in Australian Winegrowing.

Author[1,1,1]

## 1. Introduction

In the past decade, the Australian winegrowing industry has undergone a variety of pressures, such as changing market demands, disease and drought (Australia, 2021a). Furthermore, natural resources are likely to decrease, as pressures from climate change increase, making it more important than ever to improve the efficiency and sustainability of crops (AGDEE, 2021). It has become crucial for those in the wine industry to address issues relating to environmental sustainability and economic viability, with a growing need for the industry to close the research gaps between sustainable practices and their real and perceived environmental and economic advantages (Montalvo-Falcón et al., 2023; Ouvrard et al., 2020). This paper aims to provide a comprehensive analysis of the intricate relationships between economic and environmental variables within the winegrowing sector. By employing statistical machine learning we show how interconnected vineyard variables are. Utilising a ten year data set that spans Australia we show the predominant elements in classifying region and year, as well as determining a vineyards potential operational costs and profit.

This analysis utilises XGBoosted trees to classify region and year; and to determine operating cost and profit of Australian vineyards. Classification and regression trees are utilised as surrogate models to shed insight into the key partitions used by the XGBoost ensembles. Variable importance is further used to illustrate the interconnectedness of the different predictor variables, and to show the similarity in variable importance between different response variables (particularly between region and operational cost). This study aims to assist in uncovering the complexity of variables that are affected by a variety of vineyard management decisions to illustrate complex interplay of variables. This study endeavours to gain insight into the similarity in predictor variable importance between year, region, operational costs and profit.

## 2. Methods

### *2.1. Data*

Data used in this analysis were obtained from Sustainable Winegrowing Australia. Australia's national wine industry sustainability program, which aims to facilitate grape-growers and winemakers in demonstrating and improving their sustainability (SWA, 2022). Data recorded by the SWA is entered manually by winegrowers using a web based interface tool. A total of 6091 observations were collected from 2012/2013 to 2021/2022 financial years. 23 variables were used for each observation reflecting a vineyards account for the given year (see Table 2.1). The profit variable was additionally transformed and included as another separate variable, profitable; depicting whether a vineyard was profitable or not.

The data originally contained only two multiclass variables: year and region. Variables that measured the same metric from different sources (such

Table 1: Summary of variables used in the analysis. The recorded column indicate values that were either greater than zero or that were not missing.

| Variable | Units | Recorded | Number of Classes |
|---|---|---|---|
| Water Used | Mega Litres | 5846 | |
| Diesel | Litres | 5585 | |
| Biodiesel | Litres | 25 | |
| LPG | Litres | 958 | |
| Herbicide Spray | Times per year | 2026 | |
| Year | Class | 6091 | 10 |
| Disease | Class | 6091 | 2 |
| Region | Class | 6091 | 58 |
| Solar | Kilowatt Hours | 622 | |
| Irrigation Type | Class | 6091 | 20 |
| Petrol | Litres | 4309 | |
| Slashing | Times per year | 2290 | |
| Yield | Tonnes | 5935 | |
| Irrigation Energy | Class | 6091 | 16 |
| Area Harvested | Hectares | 6091 | |
| Electricity | Kilowatt Hours | 1015 | |
| Insecticide Spray | Times per year | 1092 | |
| Fertiliser | Kilograms of Nitrogen | 795 | |
| Fungicide Spray | Times per year | 2260 | |
| Cover Crop | Class | 6091 | 32 |
| Water Type | Class | 6091 | 39 |
| Profit | AUD | 853 | |
| Operating Costs | AUD | 853 | |

as water collected from rivers versus water from dams) were converted into multiclass variables representing the source. The total amount used from these variables was retained as a separate variable. Occurrences of multiple sources were defined as separate classes.

The variable region represented one of the 65 Geographical Indicator Regions (GI Region) used to describe different unique localised traits of vineyards across Australia (Halliday, 2009; Oliver et al., 2013; SOAR et al., 2008). Each region is explicitly defined under the Wine Australia Corporation Act of 1980 (Attorney-General's Department, 2010).

### *2.2. XGBoosted Trees*

XGBoosted (eXtreme Gradient Boosting) trees were created using the XGBoost library (Chen and Guestrin, 2016) in the Python Programming language (G. van Rossum, 1995). They were chosen for this analysis as they provide both a high predictive performance and ability to effectively capture complex relationships. An XGBoosted tree was created for each variable to show how they interacted. Each tree included all but the economic variables (profit and operating cost), which were only included once as response variables.

Following Chen and Guestrin (Chen and Guestrin, 2016), XGboosted trees predict a value $y_i$ from the input $x_i$. The method of prediction is achieved through a tree ensemble model, using $K$ additive functions to predict the output.

$$\hat{y_i} = \phi(x_i) = \sum_{k=1}^{K} f_K(x_i), f_K \in \mathcal{F}, \tag{1}$$

where each function $f_K$ is a classification or regression tree, such that all functions are in the set of all decision trees $\mathcal{F}$, defined by $f(x) = \omega_{q(x)}(q : \mathbb{R}^m \to T, \omega \in \mathbb{R}^T)$. Where, $f_K$ corresponds to an independent tree structure $q$ of $\omega$ weights. Each tree has $T$ leaves, which contain a continuous score, represented by $\omega_i$ for the i-th leaf. The final prediction is determined by the sum of the score of the corresponding leaves, given by $\omega$. The set of functions used by the tree is determined by minimising the regularised objective function, given by:

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i^{t-1} + f_t(x_i)) + \sum_k \Omega(f_K). \tag{2}$$

The difference between the prediction and actual variable is a convex loss function $l$. To optimise $l$, the difference is calculated for the i-th instance at the t-th iteration. The function $f_t$ is selected according to which value minimises (2). The model complexity is penalised by the function $\Omega$, this acts to smooth weights in an attempt to prevent over fitting.

As predictions are made using additive tree functions, XGboosted trees can be used for classification and regression. Due to the mixture of continuous, binary and multiclass variables in this analysis, both classification and regression trees were created. The difference between the trees created for this analysis was the objective function used. XGBoosted regression trees were created for continuous variables, using the root-mean-square as the objective function. Binary class variables utilised the logistic loss function as the objective. And, Multiclass variable used the soft max function. All objective functions are defined within the SKlearn library (Buitinck et al., 2013), linked via an API to the XGBoost library (Chen and Guestrin, 2016).

Chen and Guestrin (Chen and Guestrin, 2016) further illustrate, using Taylor expansions, that for a fixed structure $q(x)$ the optimal weight $\omega_j^*$ for a leaf $j$ can be derived. Furthermore, they show the loss reduction after the split is given by the function:

$$\mathcal{L}_{split} = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \gamma, \tag{3}$$

with the tree structure defined using left $I_L$ and right $I_R$ instance sets of nodes, with $I = I_L \cup I_R$. Instead of enumerating all possible tree structures, a greedy algorithm iteratively adds branches to the tree minimising $\mathcal{L}_{split}$ in (3). The frequency of a variable's occurrence within a tree is directly attributed to the minimisation of the objective function (or loss) through the minimisation of $\mathcal{L}_{split}$.

The frequency of a variable appearing as a node within the ensemble was used as a measure of importance. This measure was chosen as it connected a variable to the minimisation of its associated objective function, translating the value into a simple count metric. Creating XGBoosted trees for each variable allowed the use of importance to show how strongly variables were associated with each other. The importance of predictor variables to economic variables was illustrated through the use of Sankey diagrams constructed using the Holoviews python library (Rudiger et al., 2020). Other variable's interconnectedness was demonstrated through the use of a chord diagram also created using Holoviews.

Each variable utilised 80% of the data to train the XGBoost ensemble, with 20% reserved for testing and validation. Testing was done through the iterative minimisation of the respective objective function for the variables type. For continuous variables 20% was used as testing data, minimising the root-mean-square function. The final model was validated using repeated k-fold cross validation for 10 folds, repeated 10 times. $R^2$ scores were used to determine the best regression models during validation. For binary and multiclass variables data was split into 80% training, 10% testing and 10% validation data. Due to class disparity in multiclass variables (most prominently in region) data was stratified into each subset at the same ratio of class occurrence. Validation was summarised through the accuracy, the proportion of true negatives and positives.

The use of the XGBoost library incorporates regularisation techniques built into the software to mitigate over-fitting and enhance model generalisation. The further use of cross validated grid search functions allowed for the selection of better performing hyperparameters when selecting the final model. The performance measure for model selection was root-mean-square error for continuous variables. The receiver operator characteristic's area under the curve was used for category variables (Hanley and McNeil, 1982). Multiclass variables utilised the one verse one approach to minimise sensitivity to class disparity (Ferri et al., 2009; Hand and Till, 2001).

### *2.3. Classification and Regression Trees*

Classification and Regression Trees were created for region, year, profit and operating cost. These models describe the partitions that are useful in predicting these variables; giving insight into the trees that make up the ensembles created by XGBoost. These trees were created using the rparts and caret packages (Kuhn, 2008; Terry Therneau and Beth Atkinson, 2022) in the R statistical programming language (R Core Team, 2021).

Decision trees were validated using K-fold cross validation. Each model was validated using 10 folds, utilising a random selection of different samples ten separate times to validate each of the decision trees. The same measure of accuracy as the XGBoosted trees was used for comparison.

## 3. Results

### *3.1. Region*

Region classification performed at 32.34% (3.67% standard deviation) and 56.82% accuracy (50.58% validation accuracy), for the classification tree and XGBoosted ensemble respectively. The most prominent feature used to classify regions with the classification tree was water source (see Figure 1). This differed from the variables that illustrated the greatest importance for the XGBoosted ensemble (see Figure (2), with predictor variables being highly interrelated in importance. Area, water, fuel and yield were more determining factors when predicting region using XGBoost. Although water and diesel were two of the three most frequently occurring variables in predicting region, they were not as connected to the other predictor variables as Yield and area harvested were.

It is reasonable that regions, being subjected to different rainfalls and temperatures, would require different amounts of water, and would have access to different water sources. The relation of area harvested and fuel (particularly petrol) is prominent with other predictors. Due to the wide variety of uses of petrol and diesel, it is likely that they are representative of other activities within the vineyard, such as pruning and harvesting. With predictors such as yield and area being highly interconnected as they likely operate as proxy variables to other factors, possibly other present variables.

Many of the regions had significantly lower reporting rates, resulting in much poorer classification performance. The regions with the most samples performed the best. Notably bordering regions were routinely grouped together and misclassified as the same region. Two areas that suffered the most from this, specifically with the classification tree were the Limestone Coast (cool coastal areas in South Australia) and the warmer inland regions along the Murray Darling. The classification tree likely had more difficulty discerning vineyards closer to the river using only water sources due to the greater access to river water in these areas.

### *3.2. Year*

The classification tree and XGBoosted ensemble performed similarly for classifying year with 35.20% (6.28% standard deviation) and 51.81% (42.20% validation accuracy) respectively. Electricity and the type of irrigation were highly influential within the classification tree. Similarly, electricity was the most frequently occurring node in the XGBoost ensemble. However, other variables such as slashing passes, and fungicide and herbicide spraying were more prevalenct than in the classification tree. Weed and disease outbreaks are likely an influential factor when classifying different years, making the decisions to spray and slash unique factors that differ year to year. Climatic differences between years are likely tied to the influence of yield and water use.

Over half of the interrelated importance of the predictor variables is dominated by area harvested, yield and slashing passes. Although all the predictor variables are highly connected, their relative importance is not as prominent

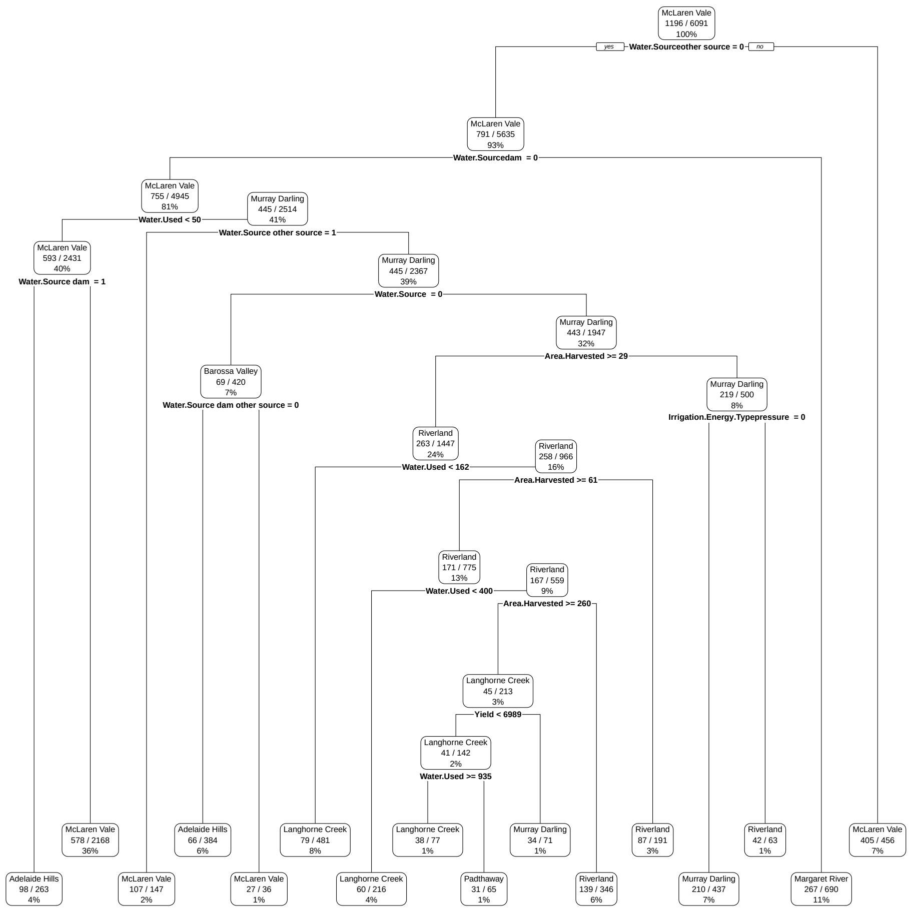


Figure 1: Decision tree predicting Region. Each node indicates the class predicted, and the proportion of elements agreeing with nodes partitioning, with the left direction indicating a yes to the nodes rule.

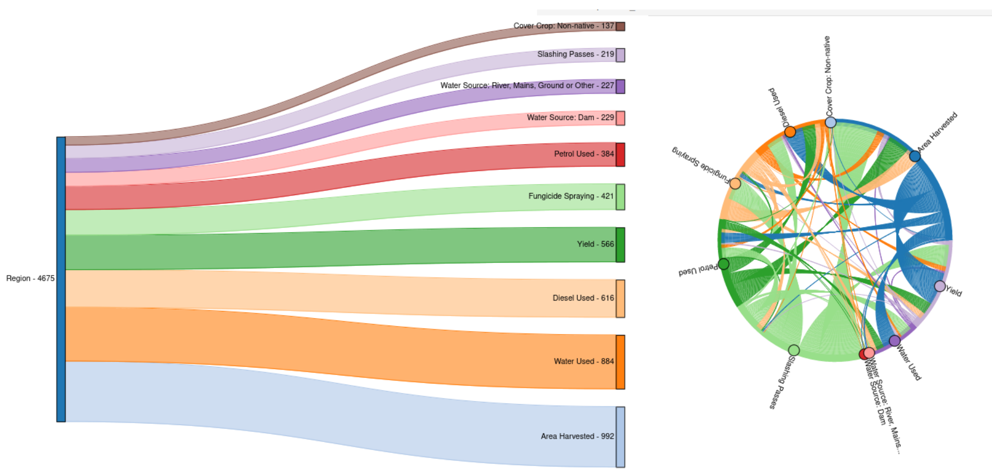


Figure 2: The left-hand side depicts the 10 most important variables in predicting Region using XGBoosted trees as a measure of node occurrence, using a Sankey diagram. The right-hand side depicts the interrelated importance of the ten predictor variables using a chord diagram.

as the three major variables. It is of particular note of the relative importance of slashing to area, fuel and yield; as these are not directly related activities. The connection between slashing and spraying is that those who do a set number of spraying or slashing passes tended to do that many passes for all slashing and spraying activities.

### *3.3. Operating Costs*

There was a pronounced difference in accuracy between the regression tree and the XGBoost model when predicting Operating costs. With the regression tree achieving an $R^2$ of 0.0931 (with a standard deviation of 0.0197) in its cross validation. The XGBoosted regression ensemble achieved an $R^2$ of 0.8025 (with a standard deviation of 0.1033).

Within the XGBoost ensemble's nodes for operating costs (see figure 5)

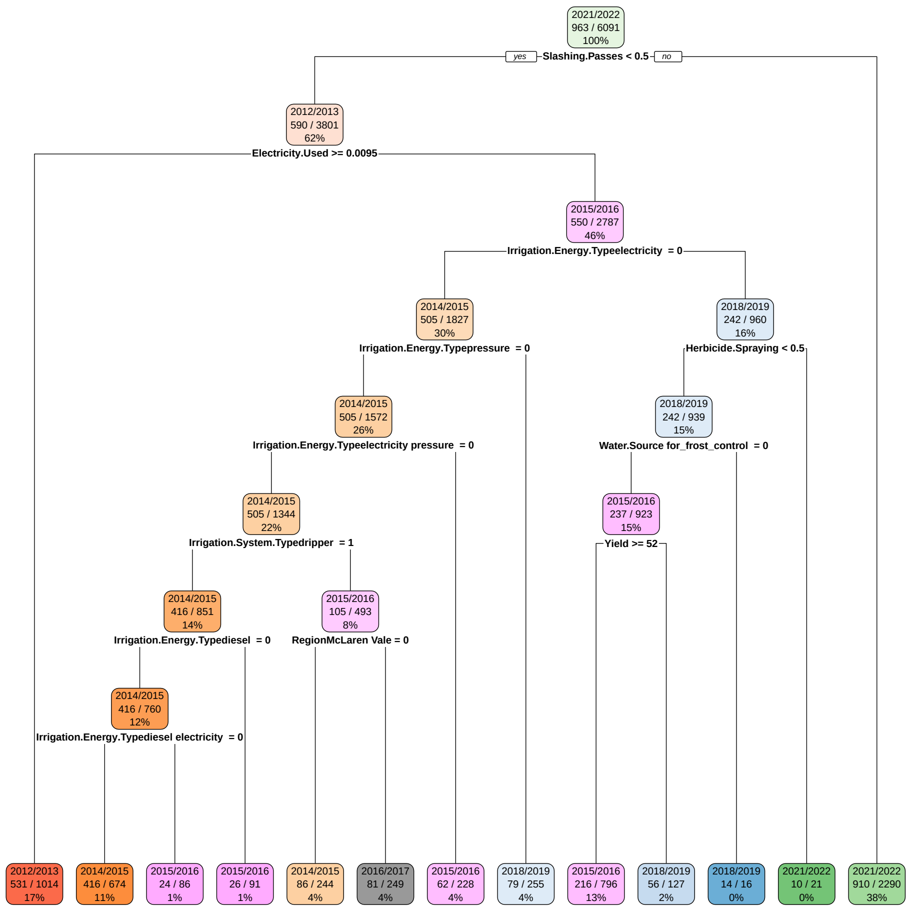


Figure 3: Decision tree predicting Year. Each node indicates the class predicted, and the proportion of elements agreeing with nodes partitioning, with the left direction indicating a yes to the nodes rule.

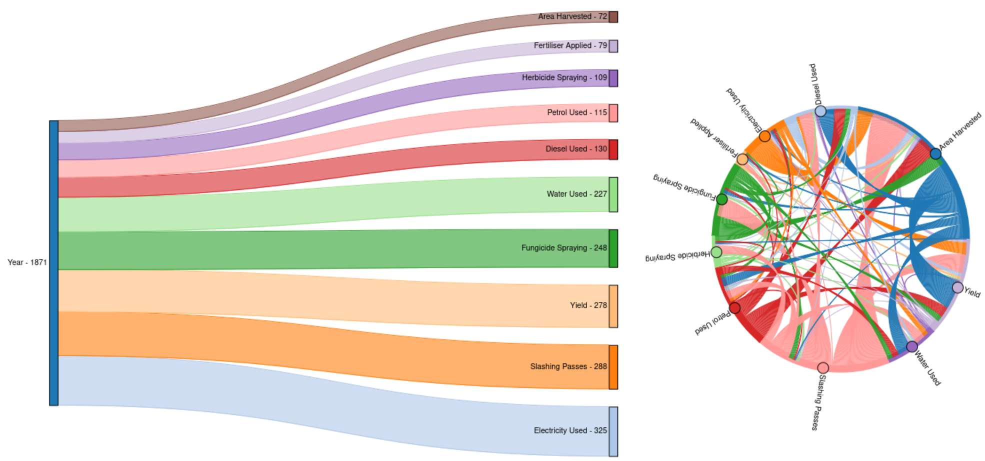


Figure 4: The left-hand side depicts the 10 most important variables in predicting Year using XGBoosted trees as a measure of node occurrence, using a Sankey diagram. The right-hand side depicts the interrelated importance of the ten predictor variables using a chord diagram.

fuel, water, area and yield occurred the most, similarly to region. Both diesel and petrol were of more relative importance (being ranked higher) in operating costs than water was compared with region. It is surprising that electricity, slashing and spraying was not more prominent in operating costs. However, Figure 4 shows that electricity, slashing and spraying are important variables in determining area and yield. Electricity in particular is used predominantly for irrigation and so is related largely to the size of a vineyard. However, slashing and spraying are measured in discrete tractor passes and show a surprising connection to the overall size of a vineyard, despite not being scaled to any measure of size. This would mean that, although measured as the same increment, a slashing or spraying pass in a larger vineyard would consume more fuel and wages than in a smaller

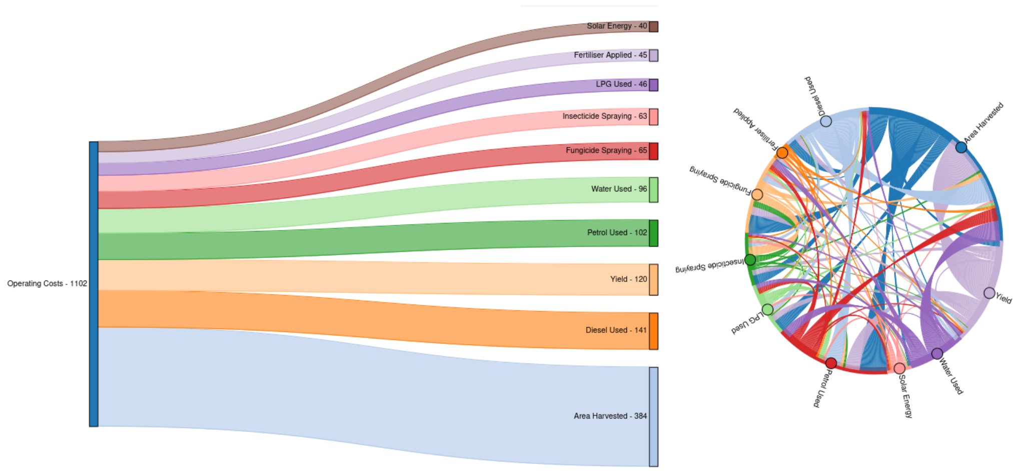


Figure 5: The left-hand side depicts the 10 most important variables in predicting Operating Costs using XGBoosted trees as a measure of node occurrence, using a Sankey diagram. The right-hand side depicts the interrelated importance of the ten predictor variables using a chord diagram.

vineyard.

### *3.4. Profit*

Predictions of profit performed poorly compared to operating costs with the regression tree having an $R^2$ of 0.1873 (with a standard deviation of 0.0522) and the XGBoosted ensemble achieving an $R^2$ of 0.2535 (with a standard deviation of 0.3126). The high standard deviation in the XGBoosted tree was a bias in more accurately predicting vineyards that made profit compared to those that lost money.With much higher $R^2$ values being achieved in k-folds containing only those that made profits (recording a maximum of 0.7634).

There was a disparity of 66.63% of vineyards recording a profit than those that did not. When predicting if a vineyard would be profitable or not the classification tree and XGBoosted ensemble did not perform considerably differently from this proportion. With the regression tree achieving an accuracy of 68.66% (and a standard deviation of 0.01%) and the XGBoost ensemble achieving 71.97% accuracy (with a validation accuracy of 70.59%).

It was surprising that operating costs performed substantially better in $R^2$ compared to profit. Interestingly the important variables when attempting to determine profit were similar to those used to classify region (see Figure 7), with the exception of water used. Both the regression tree and the XGBoosted ensemble used region, specifically the Hunter Valley. The regression tree also used Tasmania when determining profit. Both the Hunter valley and Tasmania are known for the production of high quality grapes used in export wines (Wine Australia, 2022). A major difference between region and profit was the importance given to water use, with water use being a more important variable in predicting region than profit.

## 4. Discussion

Several physical parameters such as climate, geology and soil are predetermined by a vineyard's location; making it a widely considered key determinant of grape yield and quality (Abbal et al., 2016; Agosta et al., 2012; Fraga et al., 2017). The association between yield and region is demonstrated by its position as fourth most occurring variable within the nodes of the XGBoosted ensemble which determined region (see Figure 2). The association with area and region is likely a connection to the change in land costs, with inland Australian areas (particularly of lower rainfall) being substantially cheaper to buy than coastal regions, allowing larger areas to be purchased

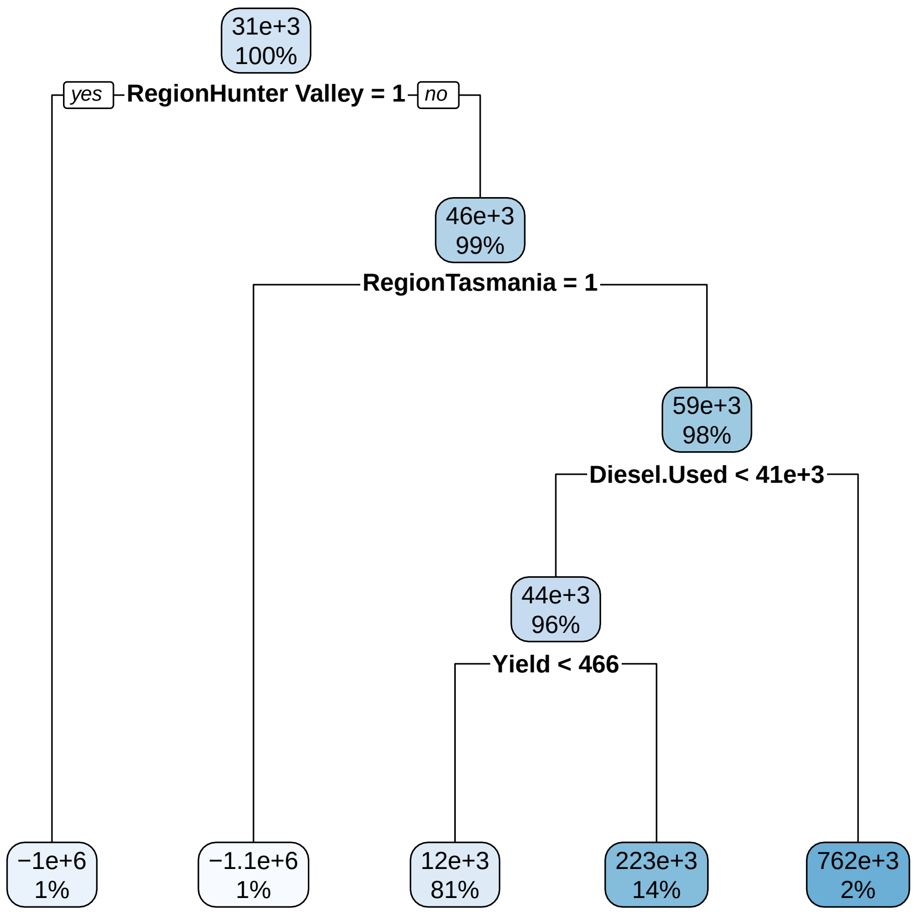


Figure 6: Decision tree predicting Profit. Each node indicates the class predicted, and the proportion of elements agreeing with nodes partitioning, with the left direction indicating a yes to the nodes rule.

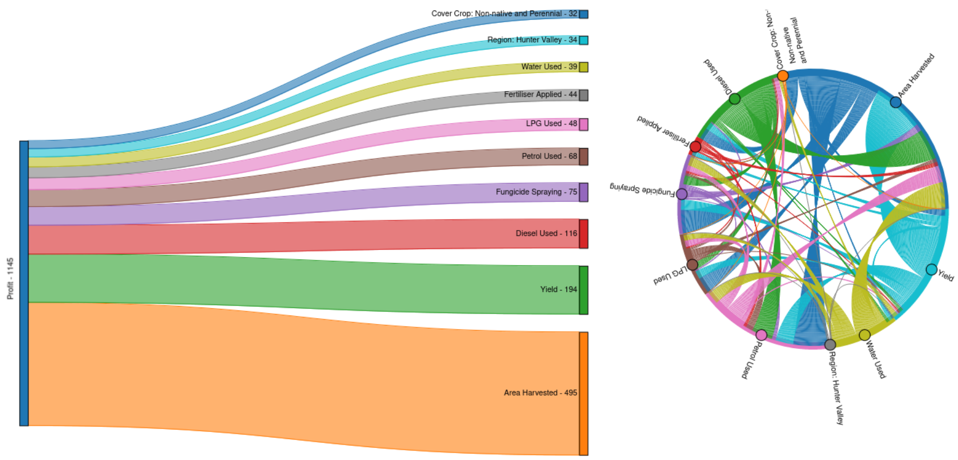


Figure 7: The left-hand side depicts the 10 most important variables in predicting Profit using XGBoosted trees as a measure of node occurrence, using a Sankey diagram. The right-hand side depicts the interrelated importance of the ten predictor variables using a chord diagram.

(Will Chancellor et al., 2019).

Regions with lower land costs are also warmer (Will Chancellor et al., 2019), which is known to be beneficial in hastening the ripening process of winegrapes (WEBB et al., 2011). Warmer regions are also associated with lower quality grapes, caused largely due to this hastened ripening (Botting et al., 1996). In general warmer regions have been associated with lower yields due to their generally lower rainfall, which can be mitigated through applying excess water (Camps and Ramos, 2012). It is likely that the combination of larger vineyards with higher water use is a determining factor in classifying regions which favour larger production of lower quality grapes; reflected through the variables' importance of water use in the XGBoost ensemble. The practice of utilising larger quantities of water for inland Australian wine crops is partly reflected in the prior use of flood style irrigation to saturate soil (BG Coombe and P Iland, 2004). This classification can be contrasted with other warmer regions of higher rainfall that use the warmer climate to concentrate their grapes, increasing the flavour profile (and thus quality) (Goodwin I, Jerie P, 1992; MG McCarthy et al., 1986). This is possibly the connection between the presence of the Hunter Valley within the XGBoost ensemble that determined profit (see Figure 7). With this connection reflecting the restriction of possible strategies employable by winegrowers between different regions.

In part some winegrowing strategies are restricted simply through access to water resources, being reflected through the region classification tree (see Figure 1). Regions are likely to have varying access to different water sources, such as those along the River Murray being able to utilise river water for crops unlike most coastal regions. Similarly, the connection between region and fuel use is likely an indicator of the level of infrastructure within the region. Where, the need to pressurise irrigation systems from river water or to generate power would require larger amounts of diesel and petrol.

Operational costs showed similar importance across fuel, water and tractor use. The dominating factor of area likely played a large part in determining how costly a tractor pass would be, or in defining the ratio of water applied to the amount of vines. The node frequency was high for area but much lower in general across the other variables, which could indicate the need to be more circumstantial in determining operational costs. Although it was attempted to capture the complexity between how variables interacted when determining operational costs (see Figure 5), it is likely yet more complicated still. An example of how interrelated operational costs can be, is the optimisation of tractor passes being shown to reduce energy use in vineyards, decreasing running costs, as well as reducing soil compaction (Capello et al., 2019).

Decisions made on the ground have far-reaching effects and are difficult to completely capture. Greater tractor use as a preventative measure for occurrences such as disease, may incur higher operational costs but could be critical in preventing long term losses. With factors such as erosion and soil health being difficult to capture but also influenced by tractor use (Capello et al., 2019, 2020). Although, performing well in $R^2$, the ability to predict operational costs is limited by the variables incorporated. Reductions in fuel, water and tractor use are obvious methods to reduce operational costs but not necessarily achievable decisions. Without fully capturing more granular activities such as the specifics of what fuel was used for, it is hard to determine what decisions specifically influence the operational costs.

Although less important in the XGBoost ensembles for profit, the variables: cover crops, fungicide spraying and slashing are likely linked to broad environmental properties of regions (see Figure 2 and 7). Rainfall being related to fungal growth and disease, as well as weeds. With cover crops being an effective and sustainable method to alleviate these issues (Delpuech and Metay, 2018). It is difficult to extrapolate findings to these methods and the reason for their use due to the broad and varying definition of the regions. Utilising the Geographical Indicator regions defined by Wine Australia (Australia, 2021b) is a limitation, as it is too broad to fully capture a vineyards location and its influence on more granular variables. The reasoning for using approaches such as cover crops can be widely varying. Where, a cover crop may be employed to help increase soil water retention, reduce erosion, increase biodiversity and reduce weeds (Capello et al., 2019, 2020; Delpuech and Metay, 2018). However, cover crops can introduce competition with grapevines and may reduce yield depending upon the plants used and the density of the cover crop (Gosling and Shepherd, 2005; Monteiro and Lopes, 2007). A more granular definition of region may help to better discern the differences in practices, and the reason for employing them. More sophisticated models, specifically those that utilise expert opinion, may also help to capture and address the decision making process. An example is the optimisation of fungicide sprays using Bayesian models that forecast disease risk (Lu et al., 2020).

The disparity in accuracy between profit and operational costs is reflective of the complexity in trying to address challenges such as climate change, disease and changing market demands (Wine Australia, 2020, 2021, 2022). The difference between turning a profit or loss is dependent on decisions made and chance. The difference between vineyards that make profit and those that do not could be a multitude of factors including differences in farming practices not captured within this study. Some decisions leading to latent effects such as large scale soil deposition in extreme rain events can be caused by soil compaction due to overworking a vineyard (Capello et al., 2020).

## 5. Conclusion

This study has provided valuable insights into the multifaceted dynamics governing operational costs, different yearly effects and vineyard regions. Highlighting the complex interrelatedness of variables within a vineyard. The paper underscores how factors such as water and fuel use intersect to impact operational costs. How different yearly events affect these operations and the significance of context-specific decision-making. While this investigation utilised a broad regional classification, the potential benefits of adopting a more nuanced approach and incorporating expert knowledge have been highlighted. By delving deeper into the complex interplay of variables, further advancements can be made in optimising vineyard management strategies for lowering operational costs and enhancing sustainability.

## References

Abbal, P., Sablayrolles, J.M., Matzner-Lober, É., Boursiquot, J.M., Baudrit, C., Carbonneau, A., 2016. Decision Support System for Vine Growers Based on a Bayesian Network. Journal of agricultural, biological, and environmental statistics 21, 131–151. doi:10.1007/s13253-015-0233-2.

AGDEE, A.G.D.o.t.E.a.E., 2021. National Greenhouse Accounts Factors: 2021.

Agosta, E., Canziani, P., Cavagnaro, M., 2012. Regional climate variability impacts on the annual grape yield in Mendoza, Argentina. Journal of Applied Meteorology and Climatology 51, 993–1009.

Attorney-General's Department, 2010. Wine Australia Corporation Act 1980.

Australia, W., 2021a. Australian Wine: Production, Sales and Inventory 2019–20.

Australia, W., 2021b. Wine Australia-Open Data.

BG Coombe, P Iland, 2004. Grape Berry Development and Winegrape Quality. In: Viticulture – Resources.. volume 1. 2 ed., Winetitles, Adelaide, South Australia.

Botting, D., Dry, P., Iland, P., 1996. Canopy architecture-implications for Shiraz grown in a hot, arid climate .

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API design for machine learning software: Experiences from the scikit-learn project, in: ECML PKDD Workshop: Languages for Data Mining and Machine Learning, pp. 108–122.

Camps, J.O., Ramos, M.C., 2012. Grape harvest and yield responses to interannual changes in temperature and precipitation in an area of north-east Spain with a Mediterranean climate. International Journal of Biometeorology 56, 853–64. doi:10.1007/s00484-011-0489-3.

Capello, G., Biddoccu, M., Cavallo, E., 2020. Permanent cover for soil and water conservation in mechanized vineyards: A study case in Piedmont, NW Italy 15.

Capello, G., Biddoccu, M., Ferraris, S., Cavallo, E., 2019. Effects of Tractor Passes on Hydrological and Soil Erosion Processes in Tilled and Grassed Vineyards. Water 11. doi:10.3390/w11102118.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA. pp. 785–794. doi:10.1145/2939672.2939785.

Delpuech, X., Metay, A., 2018. Adapting cover crop soil coverage to soil depth to limit competition for water in a Mediterranean vineyard. European Journal of Agronomy 97, 60–69. doi:10.1016/j.eja.2018.04.013.

Ferri, C., Hernández-Orallo, J., Modroiu, R., 2009. An experimental comparison of performance measures for classification. Pattern Recognition Letters 30, 27–38. doi:10.1016/j.patrec.2008.08.010.

Fraga, H., Costa, R., Santos, J.A., 2017. Multivariate clustering of viticultural terroirs in the Douro winemaking region. Ciência Téc. Vitiv. 32, 142–153.

G. van Rossum, 1995. Python tutorial, Technical Report CS-R9526. Centrum voor Wiskunde en Informatica (CWI),.

Goodwin I, Jerie P, 1992. Regulated deficit irrigation: Concept to practice. Advances in vineyard irrigation. Australian and New Zealand Wine Industry Journal 7.

Gosling, P., Shepherd, M., 2005. Long-term changes in soil fertility in organic arable farming systems in England, with particular reference to phosphorus and potassium. Agriculture, Ecosystems & Environment 105, 425–432. doi:10.1016/j.agee.2004.03.007.

Halliday, J.C.J.C., 2009. Australian Wine Encyclopedia. Hardie Grant Books, VIC.

Hand, D.J., Till, R.J., 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning 45, 171–186. doi:10.1023/A:1010920819831.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143, 29–36.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software, Articles 28, 1–26. doi:10.18637/jss.v028.i05.

Lu, W., Newlands, N.K., Carisse, O., Atkinson, D.E., Cannon, A.J., 2020. Disease Risk Forecasting with Bayesian Learning Networks: Application to Grape Powdery Mildew (Erysiphe necator) in Vineyards. Agronomy (Basel) 10, 622. doi:10.3390/agronomy10050622.

MG McCarthy, RM Cirami, DG Furkaliev, 1986. The effect of crop load and vegetative growth control on wine quality. .

Montalvo-Falcón, J.V., Sánchez-García, E., Marco-Lajara, B., Martínez-Falcó, J., 2023. Sustainability Research in the Wine Industry: A Bibliometric Approach. Agronomy 13. doi:10.3390/agronomy13030871.

Monteiro, A., Lopes, C.M., 2007. Influence of cover crop on water use and performance of vineyard in Mediterranean Portugal. Agriculture, Ecosystems & Environment 121, 336–342. doi:10.1016/j.agee.2006.11.016.

Oliver, D., Bramley, R., Riches, D., Porter, I., Edwards, J., 2013. Review: Soil physical and chemical properties as indicators of soil quality in Australian viticulture. Australian Journal of Grape and Wine Research 19, 129–139. doi:10.1111/ajgw.12016.

Ouvrard, S., Jasimuddin, S.M., Spiga, A., 2020. Does Sustainability Push to Reshape Business Models? Evidence from the European Wine Industry. Sustainability 12. doi:10.3390/su12062561.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Rudiger, P., Stevens, J.L., Bednar, J.A., Nijholt, B., Andrew, B, C., Randelhoff, A., Mease, J., Tenner, V., maxalbert, Kaiser, M., ea42gh, Samuels, J., stonebig, LB, F., Tolmie, A., Stephan, D., Lowe, S., Bampton, J., henriqueribeiro, Lustig, I., Signell, J., Bois, J., Talirz, L., Barth, L., Liquet, M., Rachum, R., Langer, Y., arabidopsis, kbowen, 2020. Holoviz/holoviews: Version 1.13.3. Zenodo. doi:10.5281/zenodo.3904606.

SOAR, C., SADRAS, V., PETRIE, P., 2008. Climate drivers of red wine quality in four contrasting Australian wine regions. Australian journal of grape and wine research 14, 78–90. doi:10.1111/j.1755-0238.2008.00011.x.

SWA, S.W.A., 2022. Sustainable Wingrowing Australia. https://sustainablewinegrowing.com.au/case-studies/.

Terry Therneau, Beth Atkinson, 2022. Rpart: Recursive Partitioning and Regression Trees.

WEBB, L.B., WHETTON, P.H., BARLOW, E.W.R., 2011. Observed trends in winegrape maturity in Australia. Global change biology 17, 2707–2719. doi:10.1111/j.1365-2486.2011.02434.x.

Will Chancellor, Shiji Zhao, Lucy Randall, Kenton Lawson, Khanh Hoang, 2019. Measuring Australian broadacre farmland value: Phase 1 - Statistical infrastructure.

Wine Australia, 2020. National Vintage Report 2020 .

Wine Australia, 2021. National Vintage Report 2021 .

Wine Australia, 2022. National Vintage Report 2022 .