Cover sheet for submission of work for assessment

SWINBURNE UNIVERSITY OF TECHNOLOGY

UNIT DETAILS

Unit name	Introduction to Data Science			Class day/time	Friday 12:30 – 2:30	Office use only
Unit code	COS10022	Assignment no.	03	Due date	29-05-2021	
Name of lecturer/teacher	Pei-Wei Tsai					
Tutor/marker's name	Pei-Wei Tsai					Faculty or school date stamp

STUDENT(S)

	Family Name(s)	Given Name(s)	Student ID Number(s)
(1)	SYED	OMAIR MAQDOOM MOHIUDDIN	102863768
(2)			
(3)			
(4)			
(5)			

DECLARATION AND STATEMENT OF AUTHORSHIP

1. I/we have not impersonated, or allowed myself/ourselves to be impersonated by any person for the purposes of this assessment.
2. This assessment is my/our original work and no part of it has been copied from any other source except where due acknowledgement is made.
3. No part of this assessment has been written for me/us by any other person except where such collaboration has been authorised by the lecturer/teacher concerned.
4. I/we have not previously submitted this work for this or any other course/unit.
5. I/we give permission for my/our assessment response to be reproduced, communicated, compared and archived for plagiarism detection, benchmarking or educational purposes.

I/we understand that:

6. Plagiarism is the presentation of the work, idea or creation of another person as though it is your own. It is a form of cheating and is a very serious academic offence that may lead to exclusion from the University. Plagiarised material can be drawn from, and presented in, written, graphic and visual form, including electronic data and oral presentations. Plagiarism occurs when the origin of the material used is not appropriately cited.

Student signature/s

I/we declare that I/we have read and understood the declaration and statement of authorship.

(1)	SYED OMAIR	(4)	
(2)		(5)	
(3)		(6)	

WINE PREDICTION MODEL

INTRODUCTION TO DATA SCIENCE

COS10020

SEMESTER – 1 2021

SYED OMAIR MAQDOOM MOHIUDDIN

(102863768)

Introduction

We are going to build a predictive model for predicting the wine type. The goal of the dataset that was given is to build a predictive model by following the best practices of data science (design, training and tuning) and help the users identify type of wine. This dataset was collected from online *[Cortez et al., 2009].* Although, due to privacy and logistics issues, some of the data characteristics (wine extract, brand, retail value ...) are not available. Only physicochemical (inputs) used for preparation and wine type (the output) variables are available. The dataset contains 2000 data records of two different types of wines (red/white) with 12 different attributes which includes where *Wine Type* identifies type of wine (Red/White) and remaining attributes describe the characteristics (pH, Density, …) and proportions of physicochemical (Sulphates, chlorides, alcohol …) for the preparation of the wine.

Data Preparation

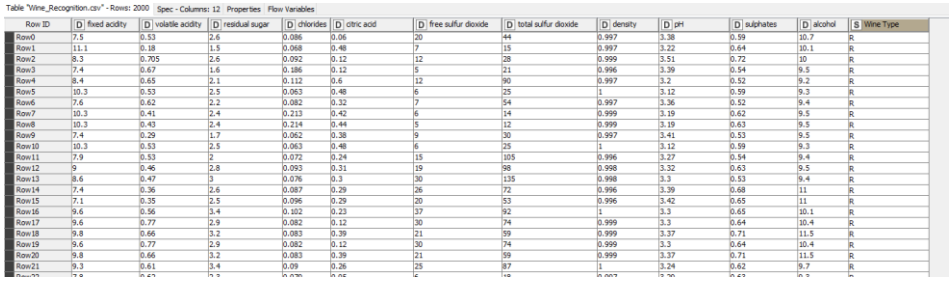
Data Observation

Here, we are going to identify the data types (Nominal/Ordinal/Continuous/Discrete).

Attribute	Nominal	Ordinal	Continuous	Discrete
Fixed Acidity			X	
Volatile Acidity			X	
Citric Acid			X	
Residual Sugar			X	
Chlorides			X	
Free Sulfur dioxide			X	
Total sulfur dioxide			X	
Density			X	
pH			X	
Sulphates			X	
Alcohol			X	
Wine Type	X			

Data Cleaning

Data should be cleaned (no missing values) to achieve the goals by creating a good predictive model. Therefore, removing all the missing values gives best results. The dataset does not contain missing data. Also, the dataset is a balanced dataset, the number of instances (tuples) from both the wine types is identical, i.e., 1000 tuples of red wine and 1000 tuples of white wine. As there is no missing data, we can use whole dataset for our prediction model. We have used KNIME Analytics Platform to recognize the missing values, type of dataset (balanced/unbalanced) and it is a valid dataset.



Row ID	fixed acidity	volatile acidity	residual sugar	chlorides	citric acid	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	Wine Type
Row0	7.5	0.53	2.6	0.086	0.06	20	44	0.997	3.38	0.59	10.7	R
Row1	11.1	0.18	1.5	0.068	0.48	7	15	0.997	3.22	0.64	10.1	R
Row2	8.3	0.705	2.6	0.092	0.12	12	28	0.999	3.51	0.72	10	R
Row3	7.4	0.67	1.6	0.186	0.12	5	21	0.996	3.39	0.54	9.5	R
Row4	8.4	0.65	2.1	0.112	0.6	12	90	0.997	3.2	0.52	9.2	R
Row5	10.3	0.53	2.5	0.063	0.48	6	25	1	3.12	0.59	9.3	R
Row6	7.6	0.62	2.2	0.082	0.32	7	54	0.997	3.36	0.52	9.4	R
Row7	10.3	0.41	2.4	0.213	0.42	6	14	0.999	3.19	0.62	9.5	R
Row8	10.3	0.43	2.4	0.214	0.44	5	12	0.999	3.19	0.63	9.5	R
Row9	7.4	0.29	1.7	0.062	0.38	9	30	0.997	3.41	0.53	9.5	R
Row10	10.3	0.53	2.5	0.063	0.48	6	25	1	3.12	0.59	9.3	R
Row11	7.9	0.53	2	0.072	0.24	15	105	0.996	3.27	0.54	9.4	R
Row12	9	0.46	2.8	0.093	0.31	19	98	0.998	3.32	0.63	9.5	R
Row13	8.6	0.47	3	0.076	0.3	30	135	0.998	3.3	0.53	9.4	R
Row14	7.4	0.36	2.6	0.087	0.29	26	72	0.996	3.39	0.68	11	R
Row15	7.1	0.35	2.5	0.096	0.29	20	53	0.996	3.42	0.65	11	R
Row16	9.6	0.56	3.4	0.102	0.23	37	92	1	3.3	0.65	10.1	R
Row17	9.6	0.77	2.9	0.082	0.12	30	74	0.999	3.3	0.64	10.4	R
Row18	9.8	0.66	3.2	0.083	0.39	21	59	0.999	3.37	0.71	11.5	R
Row19	9.6	0.77	2.9	0.082	0.12	30	74	0.999	3.3	0.64	10.4	R
Row20	9.8	0.66	3.2	0.083	0.39	21	59	0.999	3.37	0.71	11.5	R
Row21	9.3	0.61	3.4	0.09	0.26	25	87	1	3.24	0.62	9.7	R

Figure 1: Whole Dataset with Attributes Rows:2000/ Columns: 12

Attribute Selection and Conversion

As there is no missing data, I have selected all the attributes for my model, every single attribute plays a vital role in wine preparation, so selecting all the attributes gives a good result for the model. Also, selecting all the data attributes gives a diverse selection of attributes based on various scenarios in the ongoing prediction process. We have a well-structured data, and it is enough to deal if any error occurs in the prediction process.

Majority of the attribute types are numerical (double), whereas Wine Type is a string which distinguish between type of wine. Data conversion is not required in the chosen model. Using the attributes, we can build the prediction model.

Pilot Study

Instead of using whole dataset for our research, we will extract subset of the data to run a small-scale test to know whether the chosen model gives the expected results., this study is known as Pilot Study. We use this study to make sure what we are doing is correct and reliable (promising network). We use minimal percentage (5%) of data from the whole dataset and test the model to make sure the followed procedure gives us accurate results which we are willing.

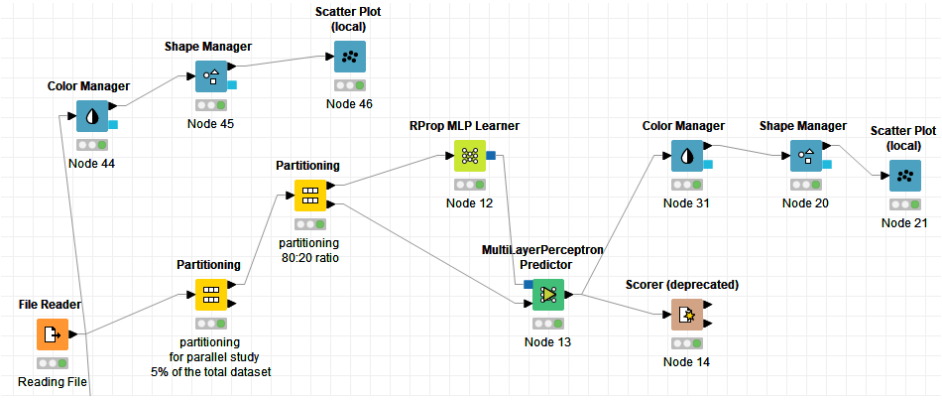


Figure 2: Pilot Study using KNIME.

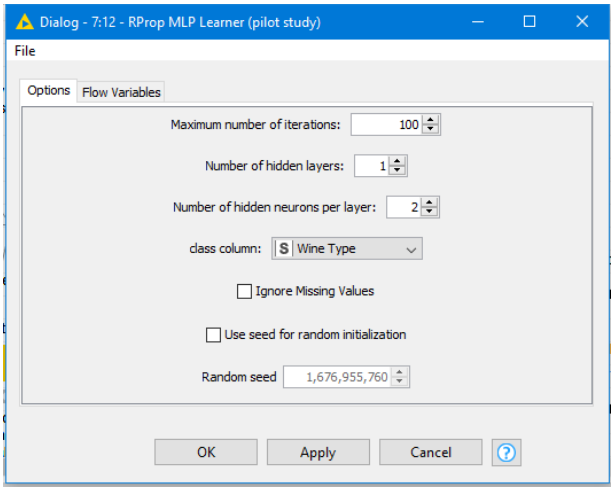


Figure 3: Giving parameters for Learner

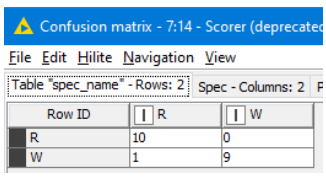


Figure 4: Accuracy Statistics

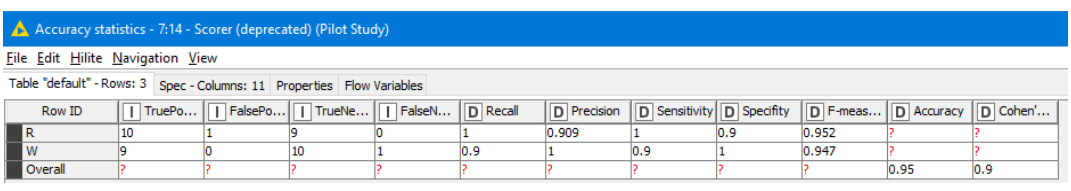


Figure 5: Confusion Matrix

We have used 5% (100 tuples) of the dataset for this study. We have partitioned the data into 2 sets. A Training Set and a Test Set as 8:2 ratio. We have used Rprop MLP Learner method for my prediction model. It uses Rprop algorithm and is an artificial neural network. Using the data from learner and taking test set and using MultiLayer Predictor to predict

results. Using Scorer to compare the predicted result. The model chosen gives the required results, although there was 1 error in the prediction. We are good to go with the chosen method for prediction.

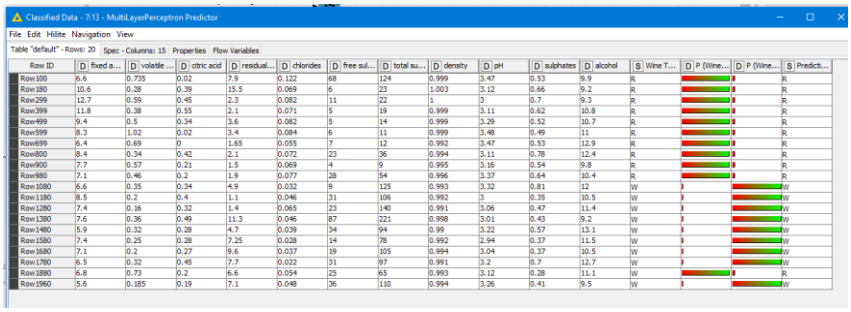


Figure 6: Classified Data from the Predictor

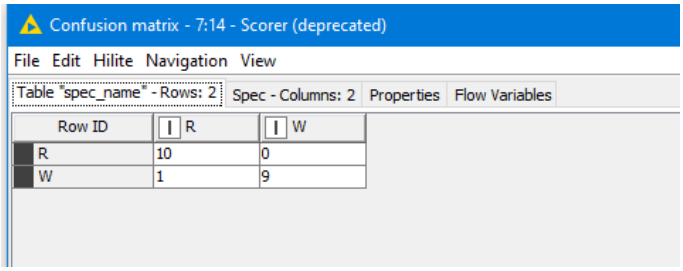


Figure 7: Confusion Matrix

Data Shuffling

In KNIME analytics platform, using shuffling node for shuffling the data.

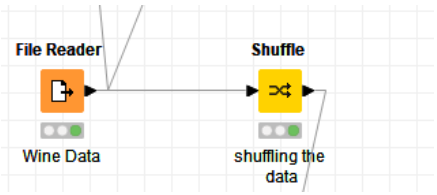


Figure 8: Shuffling of Data

Data Partition

We need to follow best practices for preparing the prediction model. Following more sophisticated cross validation method, partitioning the whole dataset into 3 sets, training set, validation set and test set. As we have chosen Rprop MLP learner (classifier) method, we have parameters for iterations, hidden layers, and hidden neurons. Depending on the provided values for these parameters, they may improve or degrade the performance of the learner. Using validation set, we can determine parameter values that will give the learner its best performance. Once the classifier is properly fine-tuned using the validation set, will evaluate its true performance on the test set.

First partitioning the whole dataset into 2 sets, one with 60% (Training set – 1200 rows) and other with 40% (Dataset1 – 800 rows). Now using the second partitioned data (Dataset1) and partitioning it into 2 sets. A Validation set (50% - 400 rows) and a Test set (50% - 400 rows) as 50:50 ratio.

The Whole Shuffled Dataset		
Training Set (60%)	Validation Set (20%)	Test Set (20%)

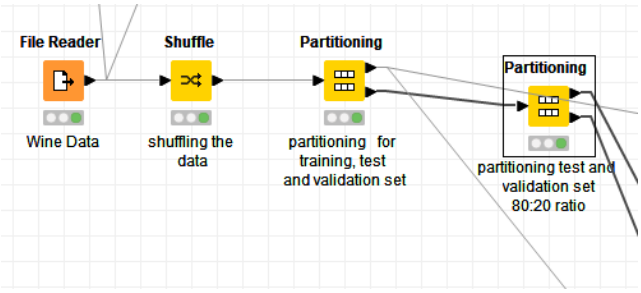


Figure 9: Data Partitioning

Prediction Model

Model Selection

I have selected Rprop (Resilient backpropagation) MLP Learner method for the prediction mode. It is a learning heuristic for supervised learning in feedforward neural networks (FFNN). It is based on Rprop algorithm, a first-order optimization algorithm. It is usually faster when compared to the classic Backpropagation. The data travel direction (connection) is straight forward from the input to resulting output. There is no looping, which means no going back to previous layers. It is very efficient, when we are deploying trained Artificial neural Network (ANN), computation is very quick and thus we can use this in real-time scenarios. The model gives us high reliability and efficiency when dealing with huge datasets (Bigdata). We also have conveniency to adjust the parameters (epochs, hidden layers, neurons, …) which helps us in getting accurate results.

Model

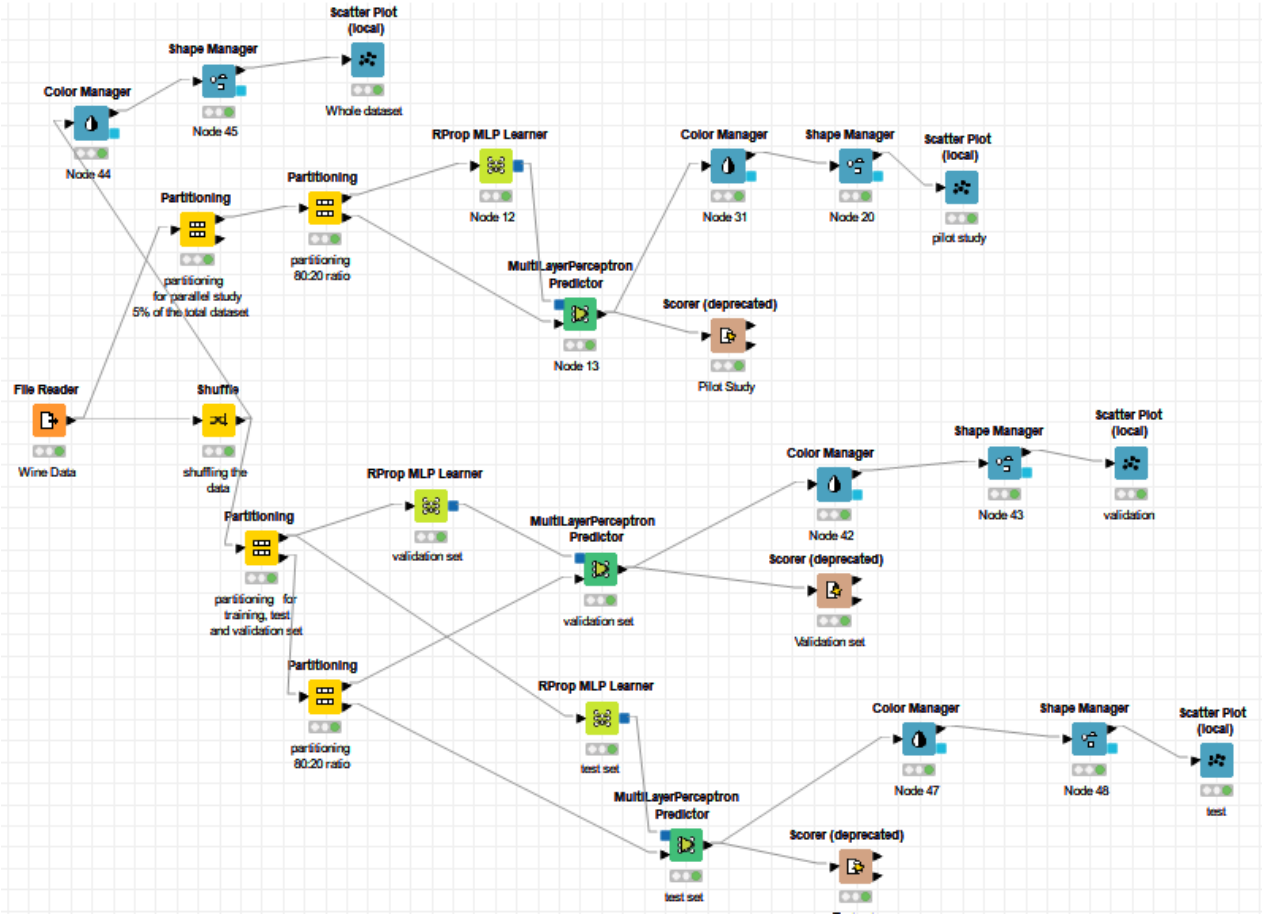


Figure 10: Prediction Model

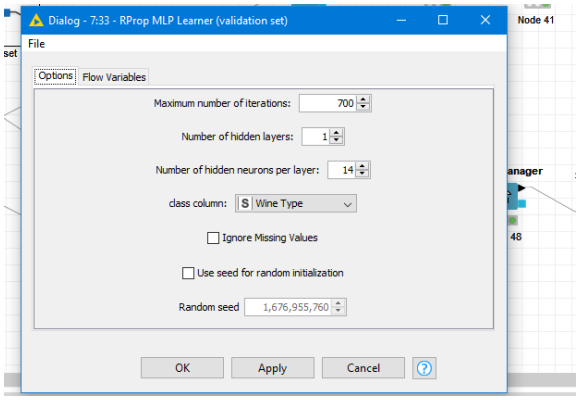
Using Rprop MLP Learner to train the training set. For initial we have set the parameters as follows, **Maximum number of iterations: 500** (recurrently training the model for number of times)**, Number of hidden layers: 1** (number of hidden

layers in between the input and output layer)**, Number of hidden neurons per layer: 10** (number of hidden neurons per hidden layer). Then using the output data from the learner and taking the data from validation set and using Multilayer Perceptron Predictor to validate the results. Finally, we have determined the parameters which will give our learner its best performance. Final parameters are, **Maximum number of iterations: 700, Number of hidden layers: 1, Number of hidden neurons per layer: 14.**

We should always remember, if we have a small model (with small dataset)**,** the training will be faster, if we have a complex model (when dealing with huge data, complex input), then we need to increase the number of iterations to get accurate results, because our model requires more training. Parallelly, when we are increasing the number of iterations, we should also increase the number of neurons in the hidden layer, it changes the probability of the prediction. In simple words, we try to get the answers (result) correctly.

- If our input is complex, then we need a larger neural network structure, i.e., a greater number of hidden layers, so we need a greater number of iterations (epoch) to train our model.

Now coming to our model, using the Scorer (deprecated) to compare the columns (actual and predicted) and generate confusion matrix (using this matrix to identify the errors and change the parameters if required, after this process, we finalise the parameters based on the confusion matrix). Here we are using Wine Type (Red or White) as the target column for the prediction model. Again, using the trained data from the learner and taking the data from test set and using Multilayer Perceptron Predictor to predict the results (we have already finalised the parameters for the learner using validation set). Using the Scorer (deprecated) to compare the two columns and generate confusion matrix.



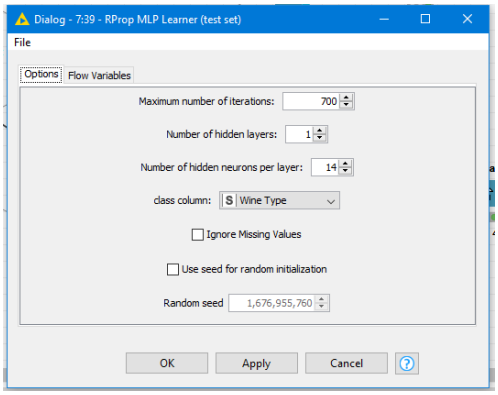


Figure 11: Learner Parameters for Test set

Figure 12: Learner Parameters for Validation set

Table "default" - Rows: 400 | Spec - Columns: 15 | Properties | Flow Variables

Row ID	fixed a...	volatile ...	citric acid	residual...	chlorides	free sul...	total su...	density	pH	sulphates	alcohol	Wine Type	P (Wine Type=R)	P (Wine Type=W)	Prediction (Wine Type)
Row1795	5.5	0.24	0.45	1.7	0.046	22	113	0.992	3.22	0.48	10	W			W
Row940	7.5	0.61	0.2	1.7	0.076	36	60	0.995	3.1	0.4	9.3	R			R
Row1896	7	0.25	0.33	2.1	0.021	17	76	0.99	3.26	0.45	12.3	W			W
Row219	7.3	0.365	0.49	2.5	0.088	39	106	0.997	3.36	0.78	11	R			W
Row217	10.4	0.61	0.49	2.1	0.2	5	16	0.999	3.16	0.63	8.4	R			R
Row837	10.4	0.52	0.45	2	0.08	6	13	0.998	3.22	0.76	11.4	R			R
Row1329	6.4	0.24	0.29	1	0.038	18	122	0.991	3.3	0.42	11.5	W			W
Row1076	6.4	0.15	0.4	1.5	0.042	23	87	0.99	3.11	0.46	12.2	W			W
Row555	7.6	0.735	0.02	2.5	0.071	10	14	0.995	3.51	0.71	11.7	R			R
Row1390	6.6	0.24	0.3	11.3	0.026	11	77	0.994	3.13	0.55	12.8	W			W
Row1029	7.5	0.18	0.37	6.2	0.05	21	138	0.995	3.2	0.55	10.5	W			W
Row1999	6.4	0.24	0.49	5.8	0.053	25	120	0.994	3.01	0.98	10.5	W			W
Row1524	6	0.23	0.15	9.7	0.048	101	207	0.996	3.05	0.3	9.1	W			W
Row1101	6.7	0.14	0.51	4.3	0.028	57	124	0.992	2.91	0.54	10.7	W			W
Row1579	6.7	0.22	0.22	1.2	0.038	5	124	0.991	3.1	0.37	11.2	W			W
Row1367	6.2	0.3	0.21	1.1	0.032	31	111	0.989	2.97	0.42	12.2	W			W
Row773	8	0.62	0.33	2.7	0.088	16	37	0.997	3.31	0.58	10.7	R			R

Figure 13: Prediction Table (classified data) for Validation set

Bars in the above image show the percentage, the probability, how our results belong to group. We can check the predicted data using Prediction column. We can see that there is some drop in the confidence, there is some overlapping. So, the probability of belonging to a specific group starts to decrease. Although there was some data wrongly predicted, we got accurate results.

Table "default" - Rows: 400 | Spec - Columns: 15 | Properties | Flow Variables

Row ID	fixed a...	volatile ...	citric acid	residual...	chlorides	free sul...	total su...	density	pH	sulphates	alcohol	Wine T...	P (Wine...	P (Wine...	Predicti...
Row1917	7.1	0.37	0.3	6.2	0.04	49	139	0.99	3.17	0.27	13.6	W			W
Row1217	7.2	0.13	0.46	1.3	0.044	48	111	0.991	2.97	0.45	11.1	W			W
Row1403	7.7	0.46	0.18	3.3	0.054	18	143	0.994	3.12	0.51	10.8	W			W
Row303	13.2	0.46	0.52	2.2	0.071	12	35	1.001	3.1	0.56	9	R			R
Row125	6.6	0.84	0.03	2.3	0.059	32	48	0.995	3.52	0.56	12.3	R			R
Row702	9.1	0.29	0.33	2.05	0.063	13	27	0.995	3.26	0.84	11.7	R			R
Row1830	7.1	0.44	0.23	5.8	0.035	24	100	0.991	3.15	0.57	13.2	W			W
Row1746	7	0.15	0.28	14.7	0.051	29	149	0.998	2.96	0.39	9	W			W
Row419	7.1	0.66	0	3.9	0.086	17	45	0.998	3.46	0.54	9.5	R			R
Row927	9	0.58	0.25	2	0.104	8	21	0.998	3.27	0.72	9.6	R			R
Row118	11.9	0.38	0.51	2	0.121	7	20	1	3.24	0.76	10.4	R			R
Row1837	8	0.27	0.33	1.2	0.05	41	103	0.99	3	0.45	12.4	W			W
Row1166	6.1	0.68	0.52	1.4	0.037	32	123	0.99	3.24	0.45	12	W			W
Row1106	7.7	0.26	0.51	2.6	0.045	26	159	0.991	3	0.5	11.2	W			W

Figure 14: Prediction Table (classified data) for Test set

As we have discussed above, bars in the image tells the percentage, the probability, how our results belong to a particular group. We can check the predicted data using Prediction column. Although the image does not show all the results, there is very low drop in the confidence, very minimal overlapping. So, the probability of belonging to a particular group starts to increase. In contrast, there was very less amount of data whose prediction was wrong. We can use confusion matrix to know which and how much of data was wrongly predicted.

Confusion Matrix

Confusion matrix helps us to understand the accuracy (metric) of the predicted data. How much amount pf data is correct, if there is some data which is incorrectly predicted, then what was the actual group of the data. We can gather all the crucial information regarding our results using confusion matrix.

Confusion matrix - 7:34 - Scorer (deprecated) (Validation set)

File Edit Hilite Navigation View

Table "spec_name" - Rows: 2 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	R	W
R	196	4
W	6	194

Figure 15: Confusion Matrix for Validation set

As it can be seen from the above confusion matrix, the left column is for the actual type and the right other two columns are predicted types. Out of all the data, 196 red wine data predictions were correct and 194 white wine data predictions were correct. Our model predicted 6 white wine data as red wine data and 4 red wine data as white wine data; these were some errors in predicting the data. These errors were minimised in the test set.

Confusion matrix - 7:40 - Scorer (deprecated) (Test set)

File Edit Hilite Navigation View

Table "spec_name" - Rows: 2 | Spec - Columns: 2 | Properties | Flow Variables

Row ID	R	W
R	193	2
W	1	204

Figure 16: Confusion Matrix for Test set

As it can be seen from the above confusion matrix, the left column is for the actual type and the right other two columns are predicted types. The prediction of data was almost correct with some minimal errors. Out of all the data, 193 red wine data predictions were correct and 204 white wine data predictions were correct. Our model predicted 1 white wine data as red wine data and 2 red wine data as white wine data; these are some errors in predicting the data. Altogether, the data prediction was accurate with efficiency and effectiveness.

Having some errors in the prediction of data is normal. There are very minimal errors in the prediction of data, so the chosen method gives us the best prediction of the data with accuracy and precision. We will consider different accuracy statistics when deciding the effectiveness of the prediction model.

Accuracy Statistics

Accuracy statistics - 7:34 - Scorer (deprecated) (Validation set)

File Edit Hilite Navigation View

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
R	196	6	194	4	0.98	0.97	0.98	0.97	0.975	?	?
W	194	4	196	6	0.97	0.98	0.97	0.98	0.975	?	?
Overall	?	?	?	?	?	?	?	?	?	0.975	0.95

Figure 17: Accuracy Statistics for Validation set

Accuracy statistics - 7:40 - Scorer (deprecated) (Test set)

File Edit Hilite Navigation View

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen'...
R	193	1	204	2	0.99	0.995	0.99	0.995	0.992	?	?
W	204	2	193	1	0.995	0.99	0.995	0.99	0.993	?	?
Overall	?	?	?	?	?	?	?	?	?	0.993	0.985

Figure 18: Accuracy Statistics for Test set

Accuracy

We can draw the Accuracy from the Accuracy statistics table provided by Scorer (deprecated). A good model should have a high accuracy % (80% and above).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

Using the formula,

Accuracy for Validation set is 97.5%.

Accuracy for Test set is 99.3% (Approx).

We consider the accuracy of the test set, by observing the above accuracy of the Test set, we can conclude that model has 99.3% Accuracy.

True Positive Rate

We can draw True Positive Rates (TPR) from the Accuracy statistics table provided by Scorer (deprecated). A well-performed model should have a high TPR (ideally 1). It is also called as “Recall”, it shows what percentage of positive instances a classifier correctly identified.

$$\text{TPR} = \frac{TP}{TP+FN}$$

Using the formula,

TPR for Validation set is 0.98 (Approx).

TPR for Test set is 0.99 (Approx).

We consider the TPR of the Test set, by observing the above TPR for the Test set, we can conclude that the TPR of model is 0.99.

False Positive Rate

We can draw False Positive Rates (FPR) from the Accuracy statistics table provided by Scorer (deprecated). A well-performed model should have a low FPR (ideally 0). It is also called as “False Alarm Rate”, or “Type I error” it shows what percentage of negatives that a classifier marks as positive.

$$\text{FPR} = \frac{FP}{FP+TN}$$

Using the formula,

FPR for Validation set is 0.03 (Approx).

FPR for Test set is 0.01 (Approx).

We consider the of the Test set, by observing the above FPR for the Test set, we can conclude that the FPR of model is 0.01 (Approx).

False Negative Rate

We can draw False Negative Rates (FNR) from the Accuracy statistics table provided by Scorer (deprecated). A well-performed model should have a low FNR (ideally 0). It is also called as “Miss Rate”, or “Type II error” it shows what percentage of positives that a classifier marks as negative. $$\text{FNR} = \frac{FN}{FN + TP}$$

Using the formula,

FNR for Validation set is 0.03 (Approx).

FNR for Test set is 0.01 (Approx).

We consider the of the Test set, by observing the above FNR for the Test set, we can conclude that the FNR of model is 0.01 (Approx).

Precision

We can draw Precision from the Accuracy statistics table provided by Scorer (deprecated). A good model should have a high precision. It is the percentage of instances marked positive that are really positive. $$\text{Precision} = \frac{TP}{TP + FP}$$

Using the formula,

Precision for Validation set is 0.98 (Approx).

Precision for Test set is 0.99 (Approx).

We consider the precision of the test set, by observing the above precision of the Test set, we can conclude that model has 0.99 Precision.

Visualization

Scatter Plot was used to visualize the data. Scatter plot gives us diverse options to choose the attributes for the different axis. We can change the attributes from the plot if required after executing the plot. IT also tells us the row from which the data point is taken using the tooltip. Using this visualization, we can show our predictons and findings to support our prediction model. I have used Total Sulfur dioxide on X-axis and Volatile acidity on the Y-axis for the scatter plot. We have used Color Manager node for selecting the color for the wine type and Shape manager node for distinguishing the shape between the type of wine.

As we can see from the below vislualizations, we can easily disntinguish the wine type looking at the color and shape, red wine is green colored and white wine is red colored. If we carefully observe the plot from the our pilot study, we can easily distinguish the data. Like we can say that usually red wine has lower percentage of total sulfuric acid and higher volatile acidity and vice verse for white wine. Although, there are some outliers in the data which lie in other region of the type. We can determine a linear regression for this model and distinguish the wine type based on it.

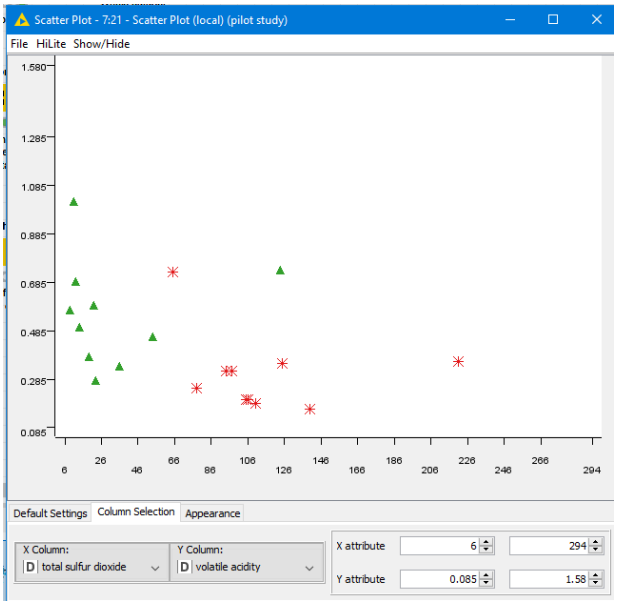


Figure 19: Scatter Plot for Pilot Study shuffled data

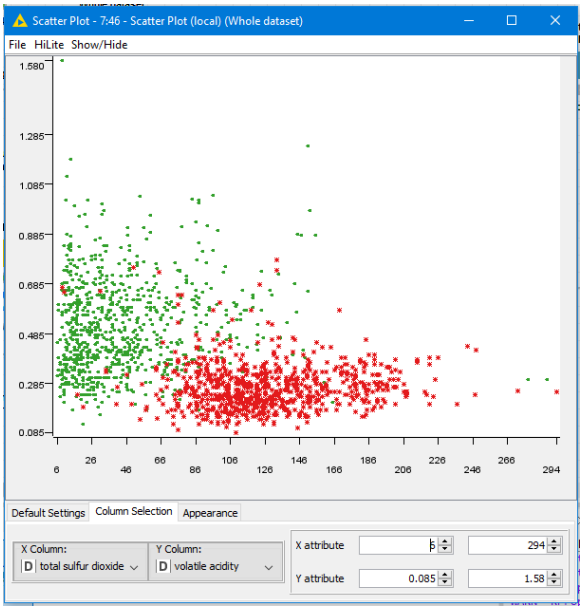


Figure 20: Scatter Plot for Whole

If we have a look at the scatter plot generated from the whole shuffled data, we can disntiguish them into two groups. One group at the top left side which is red wine and bottom right side which is white wine. We can reate a cluster for these two groups and distinguish them.

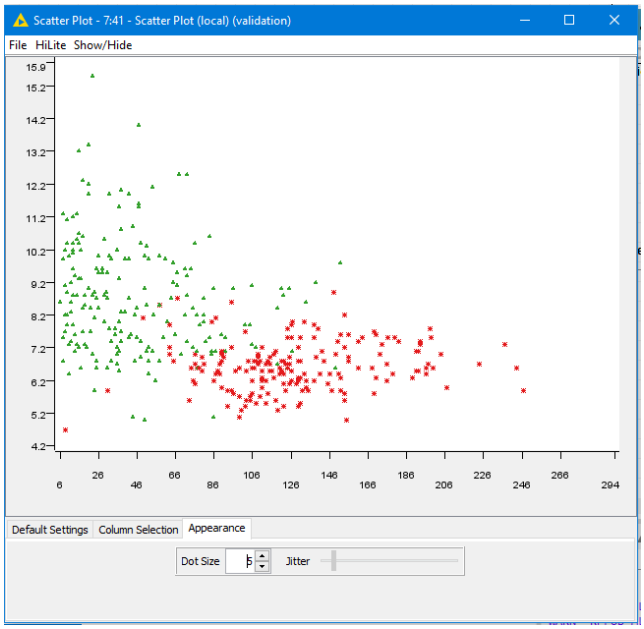


Figure 21: Scatter Plot for Test data

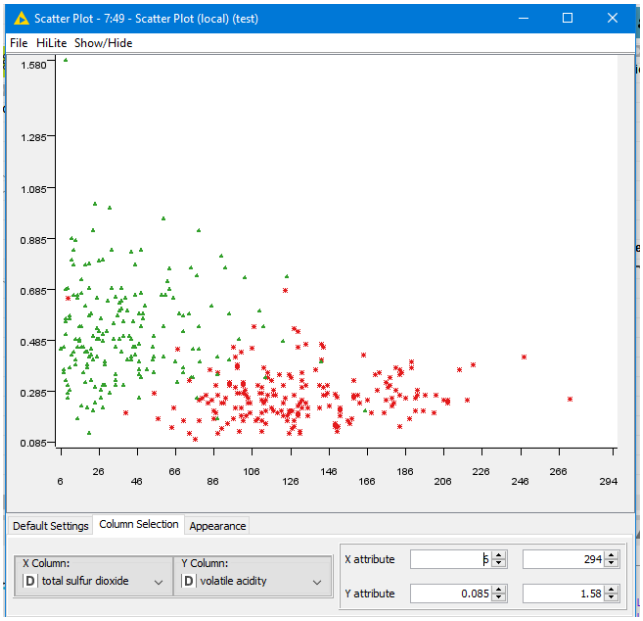


Figure 22: Scatter Plot for Validation data

Any outliers in the data, can be suspecting, i.e., if there a data from the other group them either its prediction was wrong, or it is an outlier. By distinguish them into two groups and determine the linear regression we can easily distinguish the wine type.

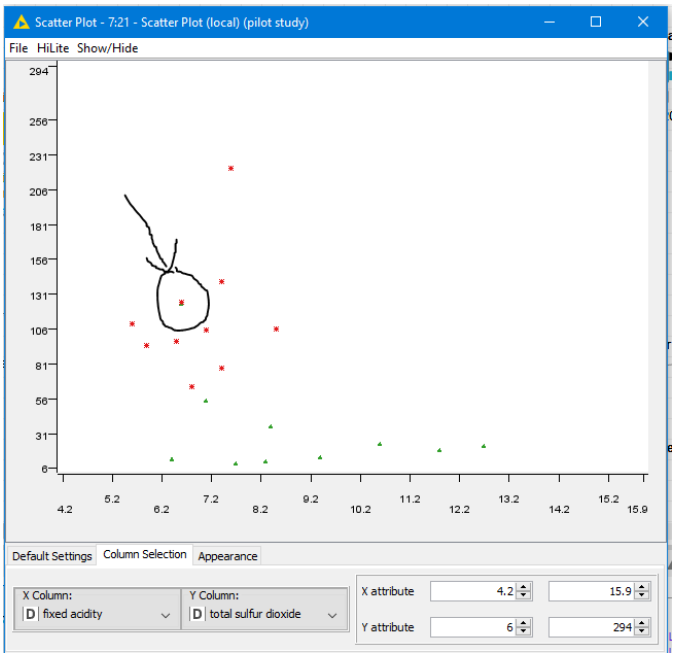


Figure 23: Scatter plot with different attributes

If we remember our pilot study, in our confusion matrix we have one data value whose prediction was incorrect. It was predicted as red wine but was white wine. When we carefully observe the scatter plot, a green coloured dot is placed in the red group, it is the data value which was wrongly predicted. Due to the characteristics of the data value, its was predicted as red wine but actually was white wine.

Conclusion

Prediction model was built by considering all the best practices to be followed for supervised models. All the phases in the model were implemented without any errors. The proposed model has high Precision, high Accuracy, high True Positive Rate, very low False Positive Rate and very low False Negative Rate. This model will help in predicting the wine type. Proposed model deals with complex data, is highly accurate, effective with complex data, highly efficient, greatly reliable and durable in dealing with missing data. Visualizations are provided to get deeper and clear understanding of the model. It can be used to identify and understand the trends in the data, distinguish the wine groups based on data values and predict the upcoming data (Red/White Wine).

References

Dietrich, D. ed., 2015. Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. EMC Education Services, Pagesg 359-396. [22 May 2021]

Graupe, D 2013, Principles Of Artificial Neural Networks (3rd Edition), World Scientific Publishing Company, Singapore. Available from: ProQuest Ebook Central. [22 May 2021].

M. Riedmiller and H. Braun, "A direct adaptive method for faster backpropagation learning: the RPROP algorithm," IEEE International Conference on Neural Networks, 1993, pp. 586-591 vol.1, doi: 10.1109/ICNN.1993.298623. [22 May 2021]

T. Kathirvalavakumar, S. Jeyaseeli Subavathi, Neighborhood based modified backpropagation algorithm using adaptive learning parameters for training feedforward neural networks, Neurocomputing, Volume 72, Issues 16–18, 2009, Pages 3915-3921, ISSN 0925-2312. [23 May 2021]

Tsai, P 2021, 'Advanced Predictive Model: Artificial Neural Network', Introduction to Data Science, Learning material via canvas, Swinburne University of Technology, 1 March. [21 May 2021]

Data Source

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009