**APPAREL PRICE PREDICTION USING ORANGE TOOL**

**Koustav moitra , Ankeeta kumari , Tanmay pramanik**

\*1 [koustavmoitra711@gmail.com](mailto:koustavmoitra711@gmail.com) , \*2 [upadhay.ankeeta@gmail.com](mailto:upadhay.ankeeta@gmail.com) , \*3 [pramanick15@gmail.com](mailto:pramanick15@gmail.com)

**MCA 2nd semester**

**Department of Computer Application, Siksha ‘O’ Anusandhan Deemed to be University,**

**Bhubaneswar, India-751030**

**Abstract:**-

Different machine learning techniques are helpful to extract the knowledge from the raw and plain data from database. So we have also created a dataset which will guide a customer with the help of different factors of the apparel (dress) whether the customer should pay less or more for that apparel.

Keeping and analysing all the other factors such as rating, size, material etc. we have trained the model and applied some classification techniques to predict the accuracy of our training and testing model in 70:30 ratio.

**Keywords:**-

Apparel data set, Orange tool, evaluating the factors, Classification techniques, Accuracy.

1. **Overview/Introduction:-**

We have created a dataset “**apparel dataset**” with taking a reference from UCI machine learning repository and by analyzing their dataset we made up our modified dataset which will predict the apparel price on the basis of various factors. We have used some classification techniques in this project and we have used the “orange **tool**” for making the model work.

**Orange Tool:-**

Orange is a open software for data mining.It helps in building and analysis workflows visually because if its diverse toolbox Orange tool comprises some components in it and these components are called widget and using it we have done the evaluation and predictive modelling

1. **Methods/Techniques:-**

For our dataset we have used classification techniques which will help our dataset to predict best accuracy.

* svm
* Neural networks
* Naïve bayes
* Random forest

**svm:** svm is a supervised machine learning technique that helps to separate the attribute spaces with hyper lane

**Random Forest:** Random forest is learning method also used for classification, regression tasks.In this technique the forest is built is an ensemble of decision tree.

**Naive Bayes:** Naive Bayes classifier are the collection of classification algorithm based on bayes theorem.It is not a single algorithm but a family of algorithm where all of them share a common principle every pair is independent of each other

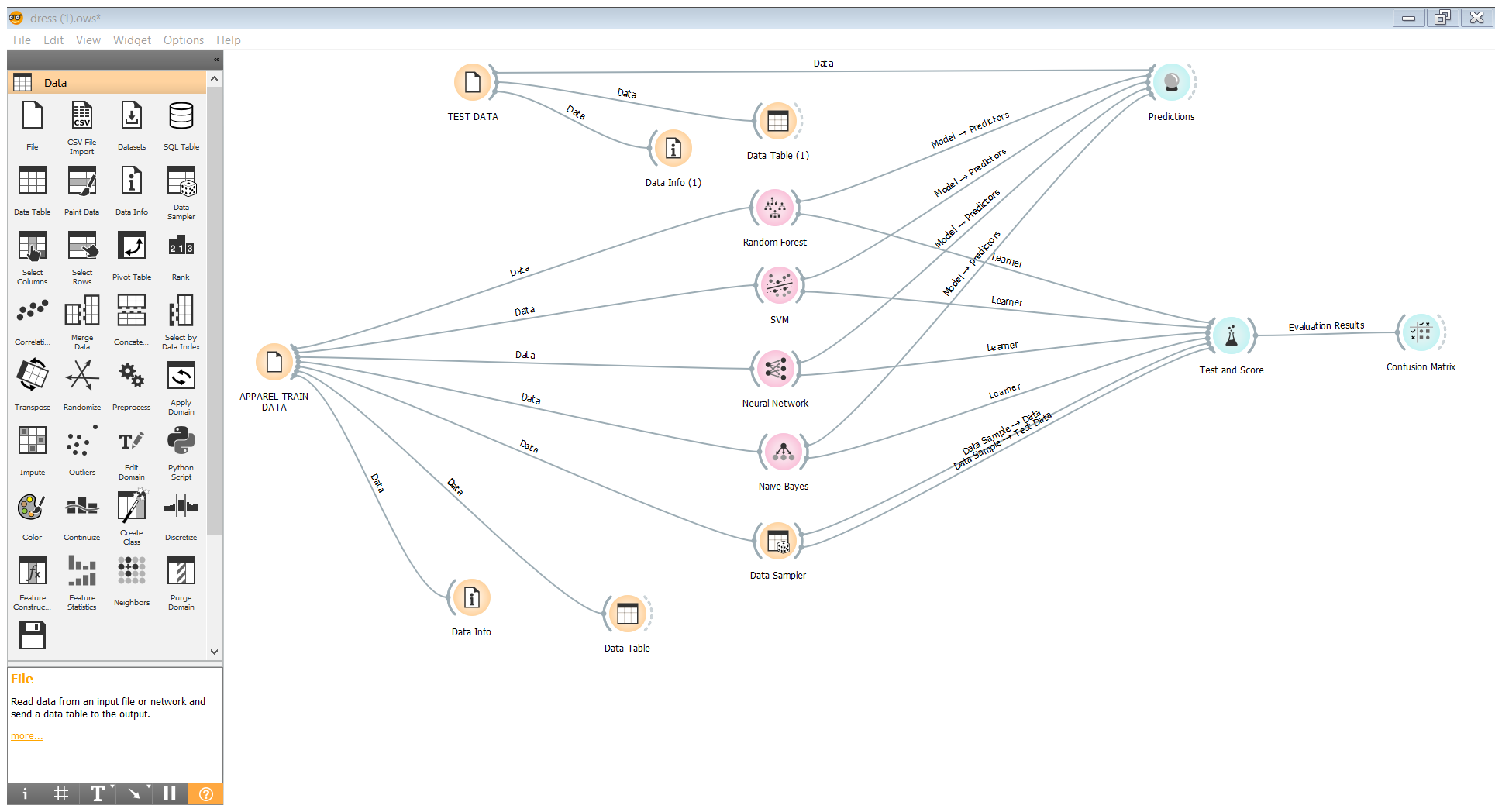
**Neural Network:**  The neural network widget uses sklearn’s Multi-layer algorithm that can learn non-linear models as well as linear. It executes them in the following order removes instances with unknown target valuescontinuizes categorical variables (with one-hot-encoding) removes empty columns

1. **Discussion:-**

**Dataset 🡪**

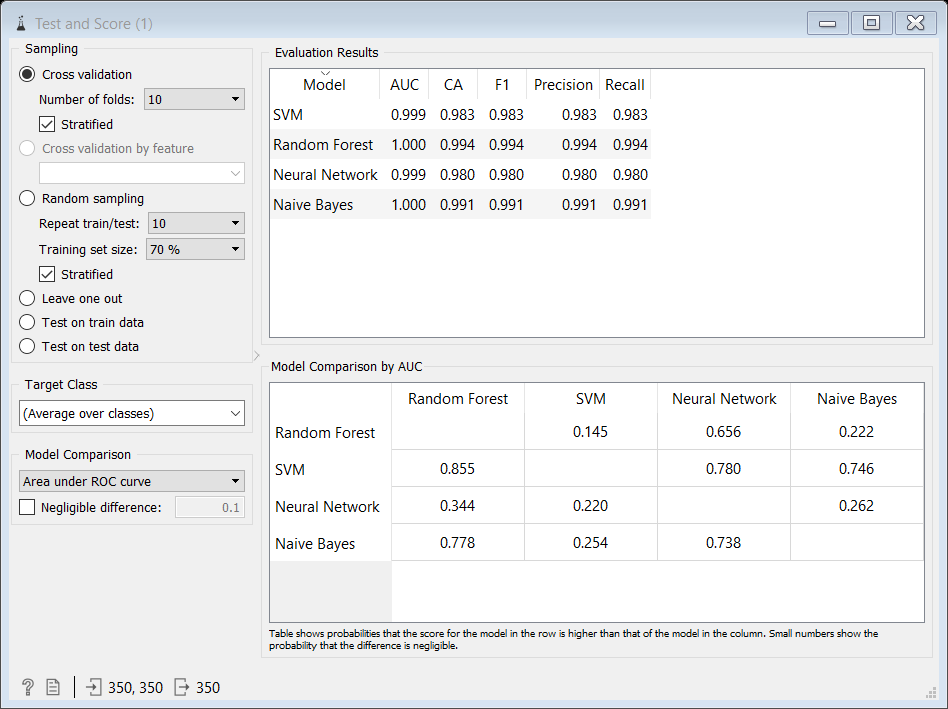
|  |  |  |
| --- | --- | --- |
| **Serial number** | **Attributes** | **datatype** |
| 1. | Style | Categorical |
| 2. | Rating | Numerical |
| 3. | Size | Categorical |
| 4. | Season | Categorical |
| 5. | Neck-line | Categorical |
| 6. | Sleeve-length | Categorical |
| 7. | Fabric type | Categorical |
| 8. | Material | Categorical |
| 9. | Pattern type | Categorical |
| 10. | Price | Categorical |

**Workflow 🡪**



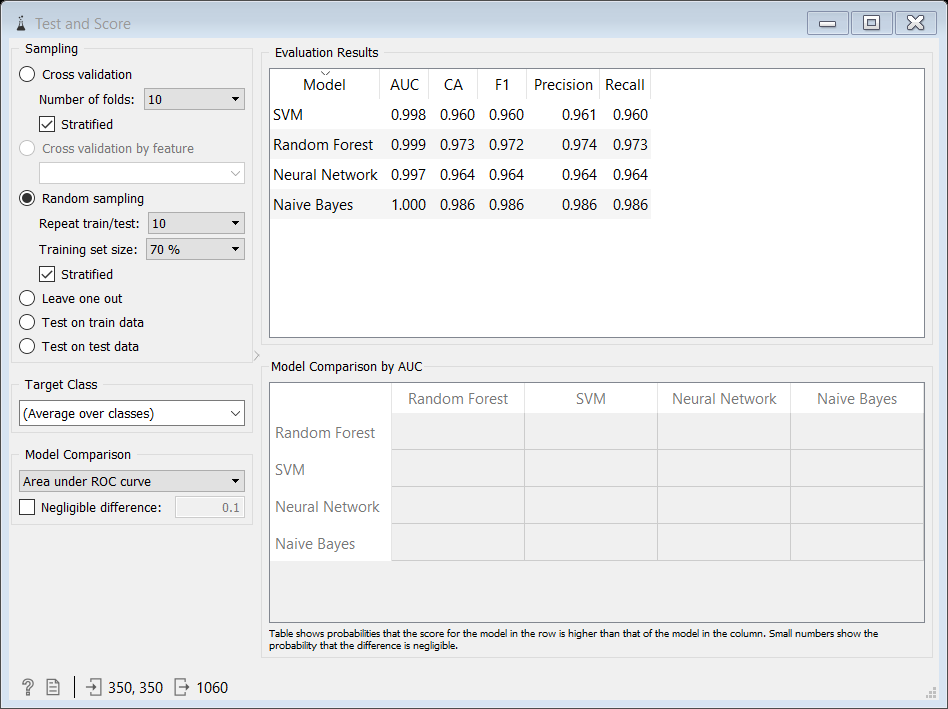
As per the workflow we have used Svm , Neural networks , Naïve bayes and Random forest models to train our dataset after connecting these models we have checked the accuracy of these 4 models using the TEST AND SCORE widget and their accuracy are as follows :-

**Accuracy analysis on the basis of cross-validation:-**



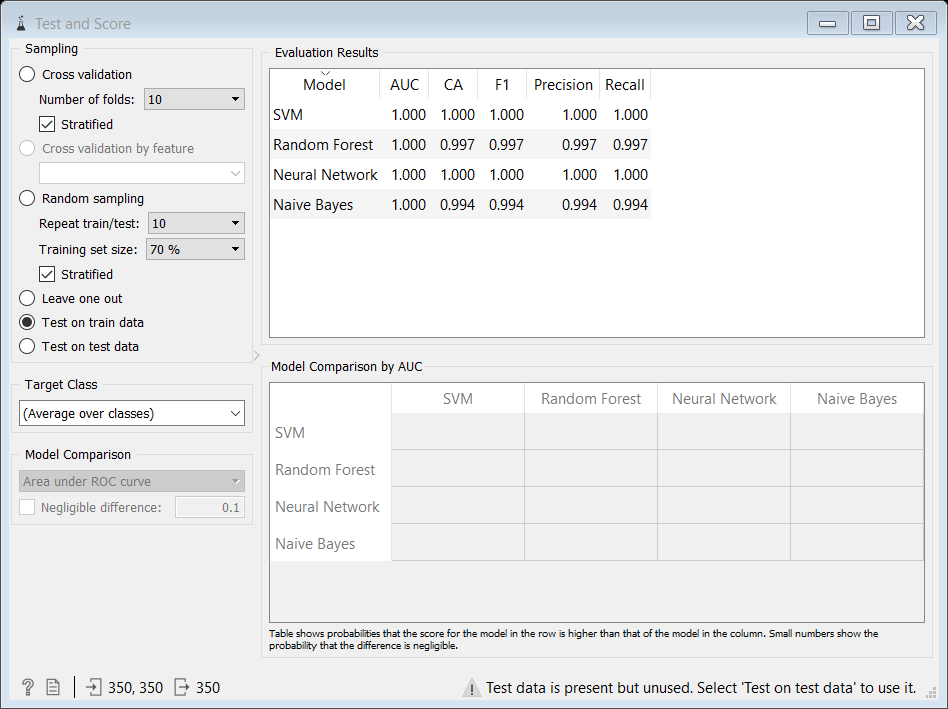
|  |  |
| --- | --- |
| **Models** | **Classification Accuracy** |
| Svm | 98.3% |
| Random forest | 99.4% |
| Neural network | 98.0% |
| Naïve bayes | 99.1% |

**Accuracy analysis on the basis of random-sampling:-**



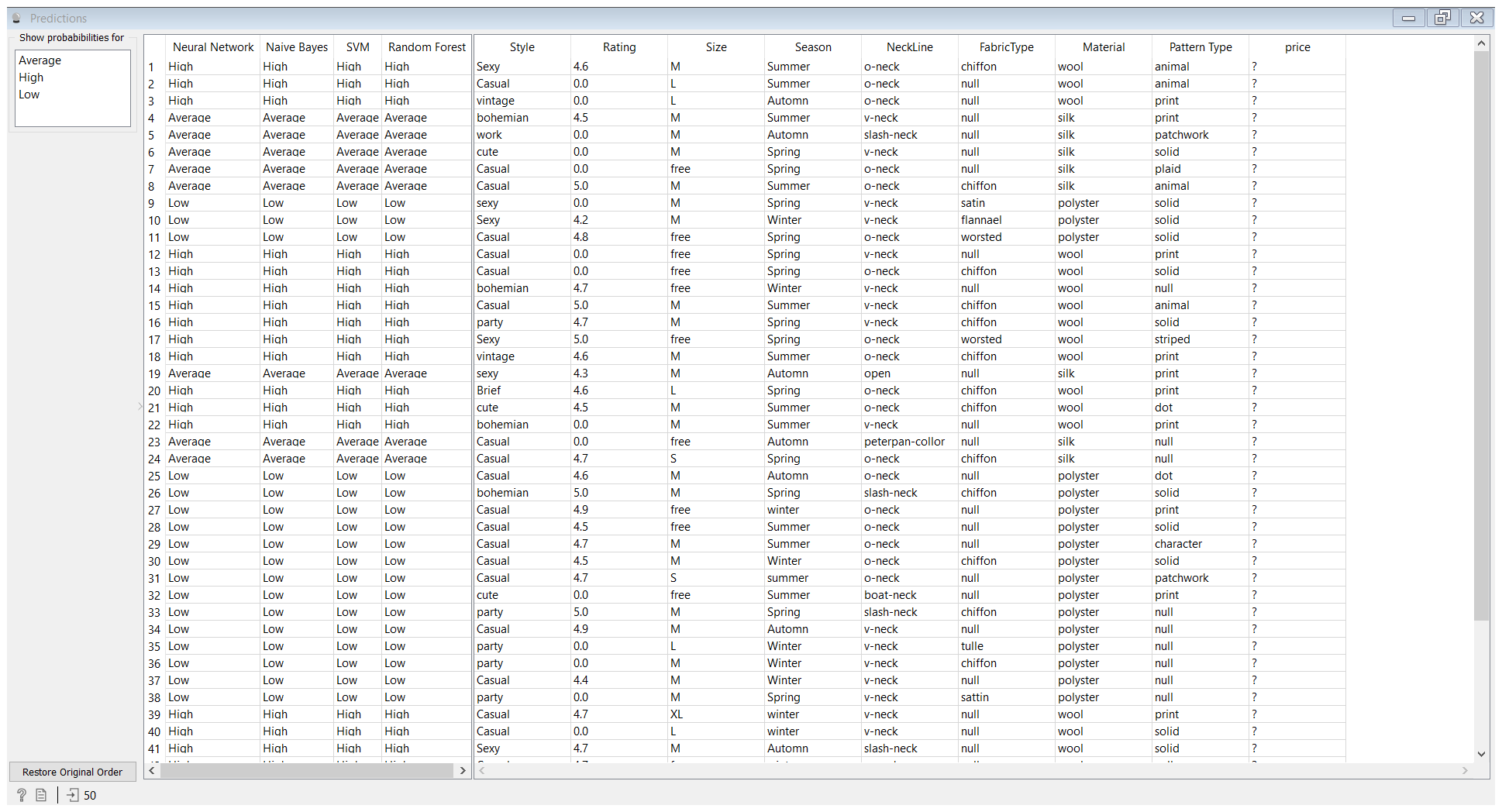
|  |  |
| --- | --- |
| **Models** | **Classification Accuracy** |
| Svm | 96% |
| Random forest | 97.3% |
| Neural network | 96.4% |
| Naïve bayes | 98.6% |

**Accuracy analysis on the basis of test on train data :-**

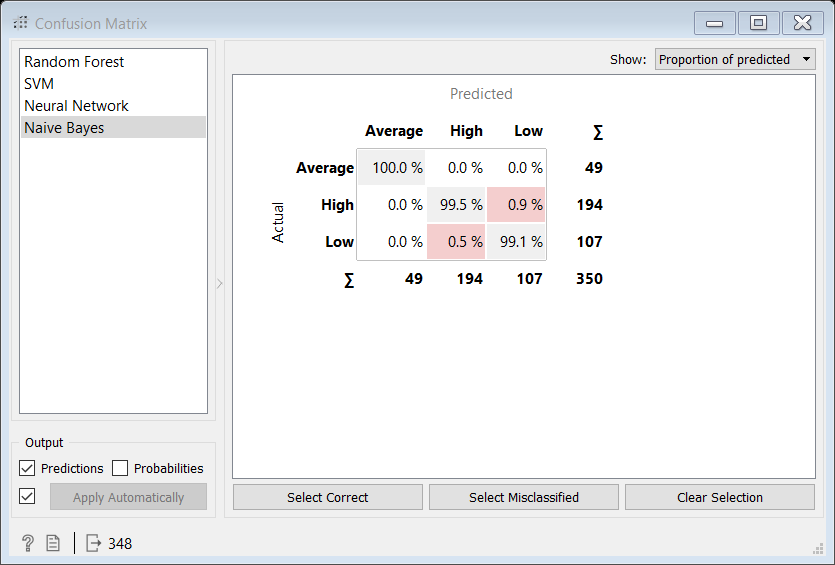


|  |  |
| --- | --- |
| **Models** | **Classification Accuracy** |
| Svm | 100% |
| Random forest | 99.7% |
| Neural network | 100% |
| Naïve bayes | 99.4% |

**Prediction of test data :-**



**Confusion matrix:-**

****

as per the dataset all the classification techniques have shown 100% accuracy but naïve bayes have shown a margin of 0.6% error rate ie, 99.4% CA , which was again checked by confusion matrix widget to know the exact error % . We found that our class attribute **price** having three values – **High, Average, Low**

**High = prediction was 100% accurate**

**Average = 0.5% instances predicted wrong**

**Low = 0.9% instances predicted wrong**

1. **Conclusion :-**

By analyzing different sampling methods via cross-validation , random sampling and test on train data in **test and score** widget it is seen that neural networks and svm models have best accuracy among others models , so for our apparel dataset we can say that the svm model and neural networks have the best accuracy rates

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Cross-validation** | **Random – sampling** | **Test on train data** |
| **Svm** | 98.3% | 96% | 100% |
| **Neural networks** | 98.0% | 96.4% | 100% |