

Introduction to Artificial Intelligence

Assignment #4

Syed Hassaan Murtaza Naqvi | 11554 | May 13, 2018

# Problem and Data Description

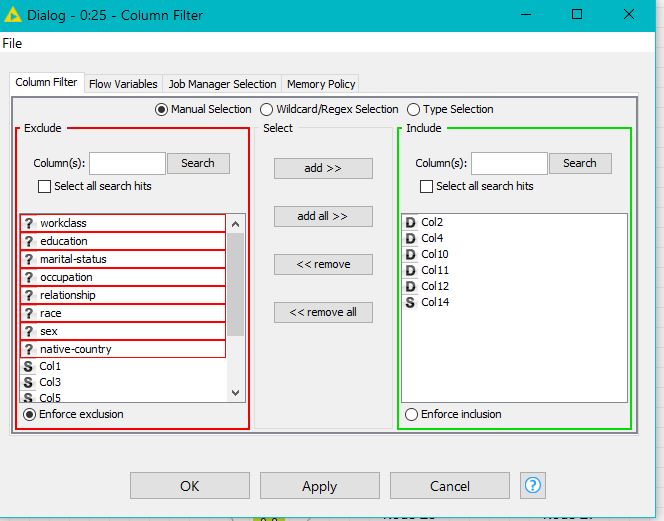
## Extraction was done by Barry Becker from the 1994 Census database.

## Prediction task is to determine whether a person makes over 50K a year based on the following attributes:-

* Age: continuous.
* Work class: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
* Fnlwgt: continuous.
* education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
* Education-num: continuous.
* Marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
* Occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
* Relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
* Race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
* Sex: Female, Male.
* Capital-gain: continuous.
* Capital-loss: continuous.
* Hours-per-week: continuous.
* native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc.), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holland-Netherlands.

## Details of Pre-Processing

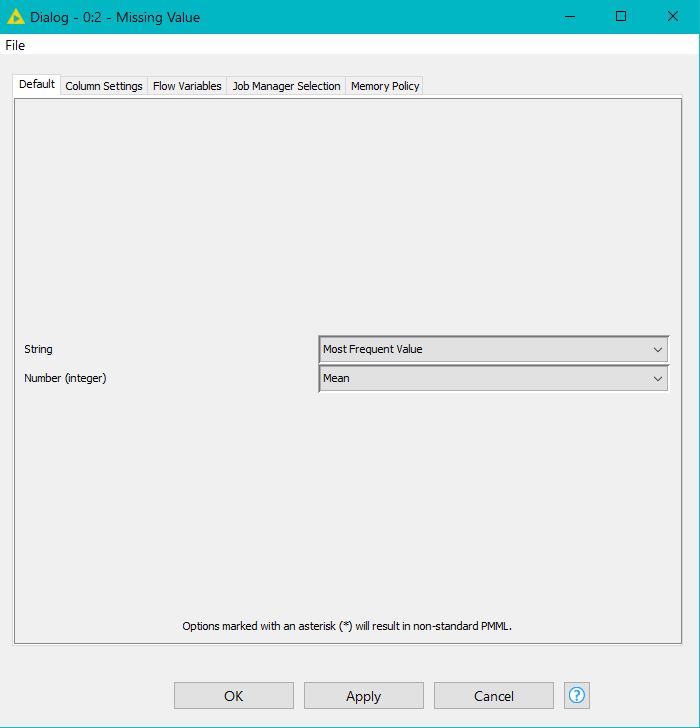
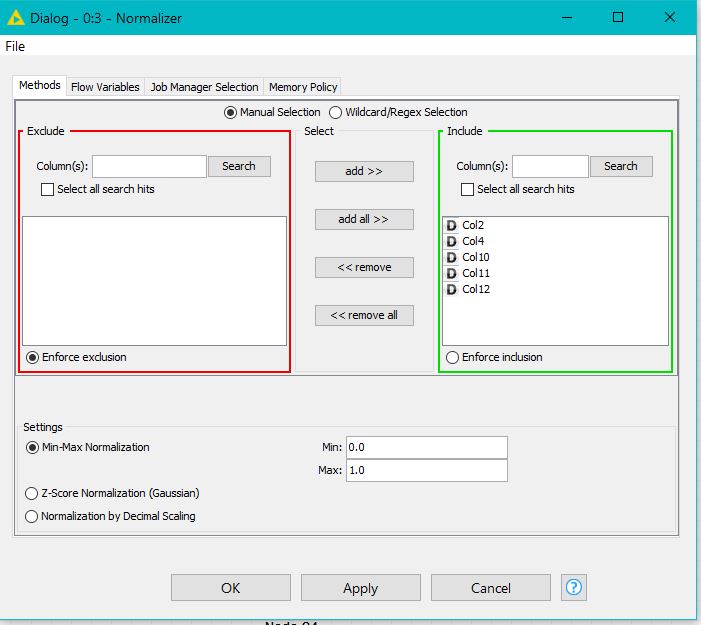
Column filter for filtering out ID



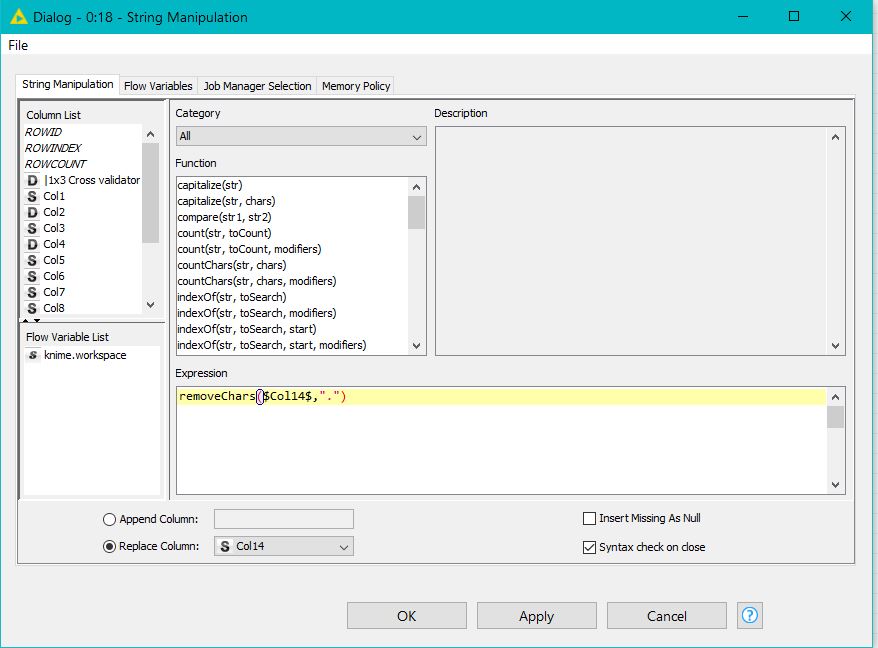
Missing values handling

Column filter for numeric values for neural network

Normalsing values numeric



String manipulation



# Specific Details of each Technique and KNIME Workflows

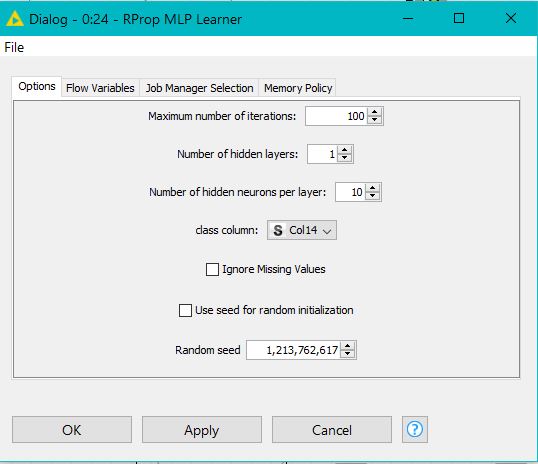
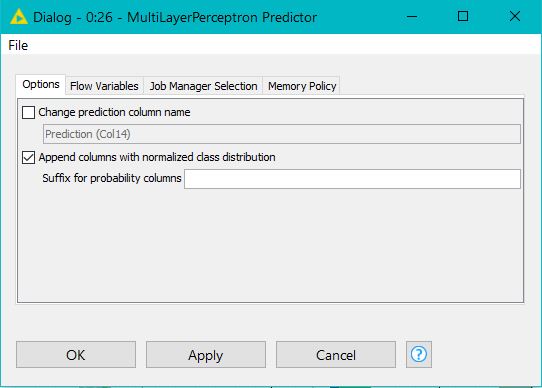
Naive Bayes Predictor Config

Naive Bayes Learner Config

Decision Tree Predictor Configuration

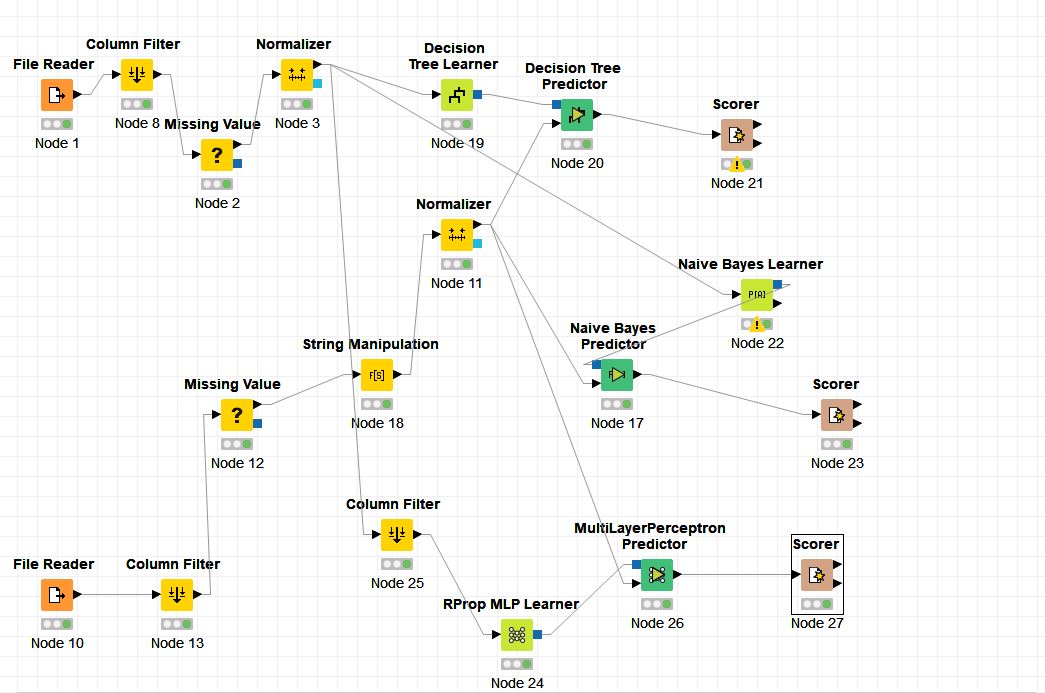
Decision Tree Configuration

Neural Network Config



Neural Network Predictor

KNIME Workflow

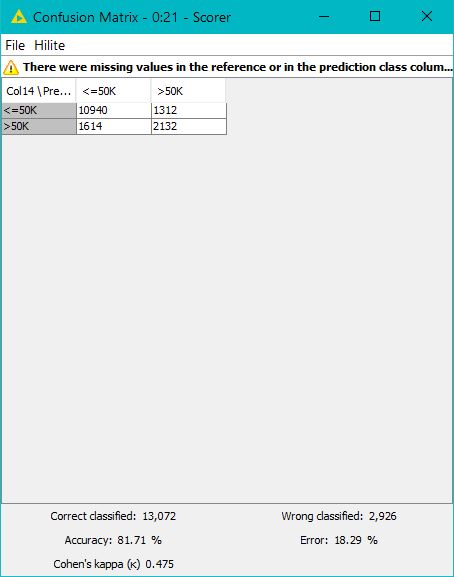
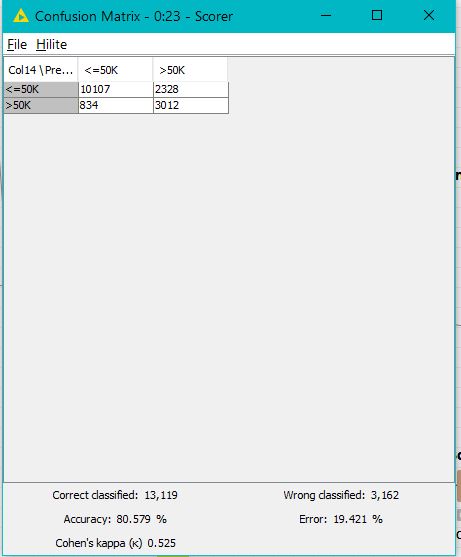


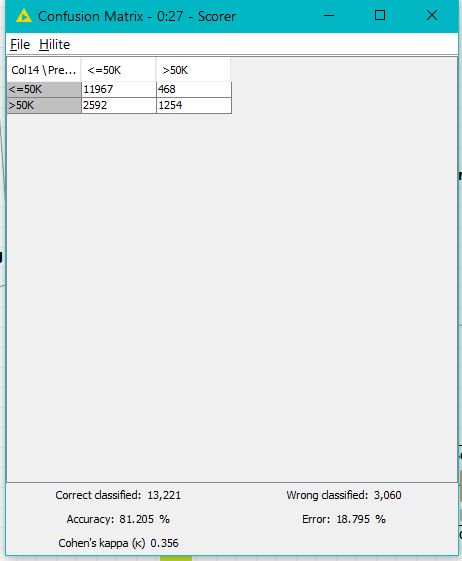
# Results

Confusion Matrix Neural Network

Confusion Matrix Decision Tree

Confusion Matrix Naïve Bayes





Naive Bayes Score

Neural Network Score

Decision Tree Score

# CONCLUSION

My attempts yielded the result that the Decision Tree classifier has yielded the highest accuracy for predicting the data, with an accuracy of 81.71%, compared to the Naïve Bayes and Neural Network classifiers, which yielded accuracies of 80.579% and 81.205% respectively.