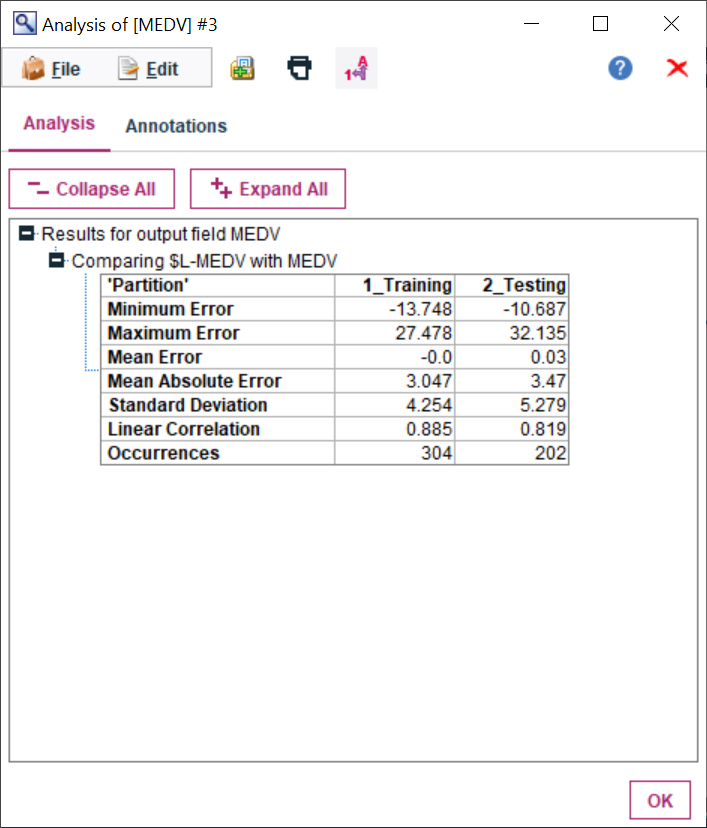
**Boston House Price Prediction using SPSS modeler**

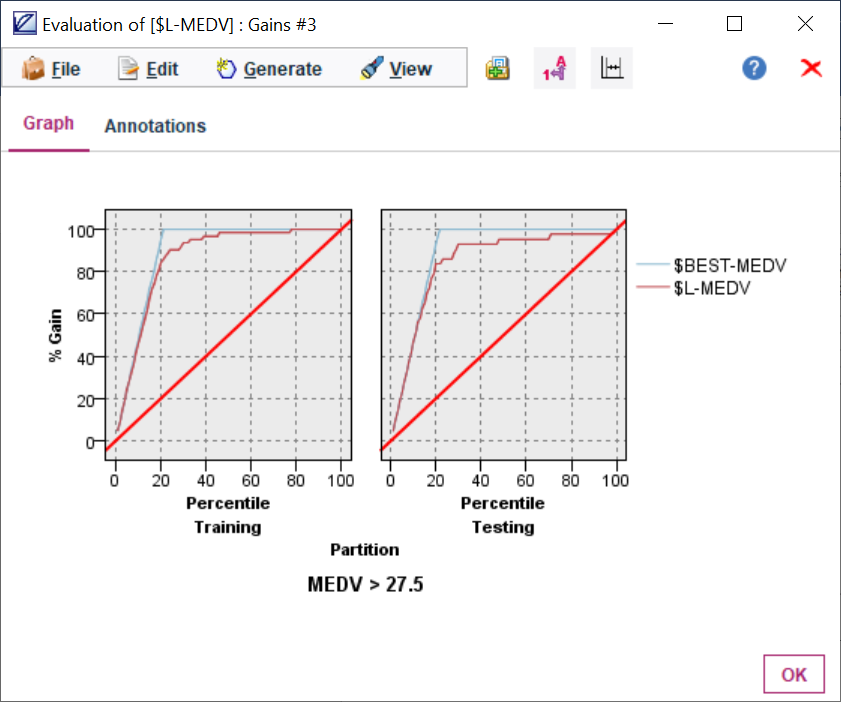
**Question: What is the best Model for predicting Boston house prices (variable MEDV)?**

**Steps:**

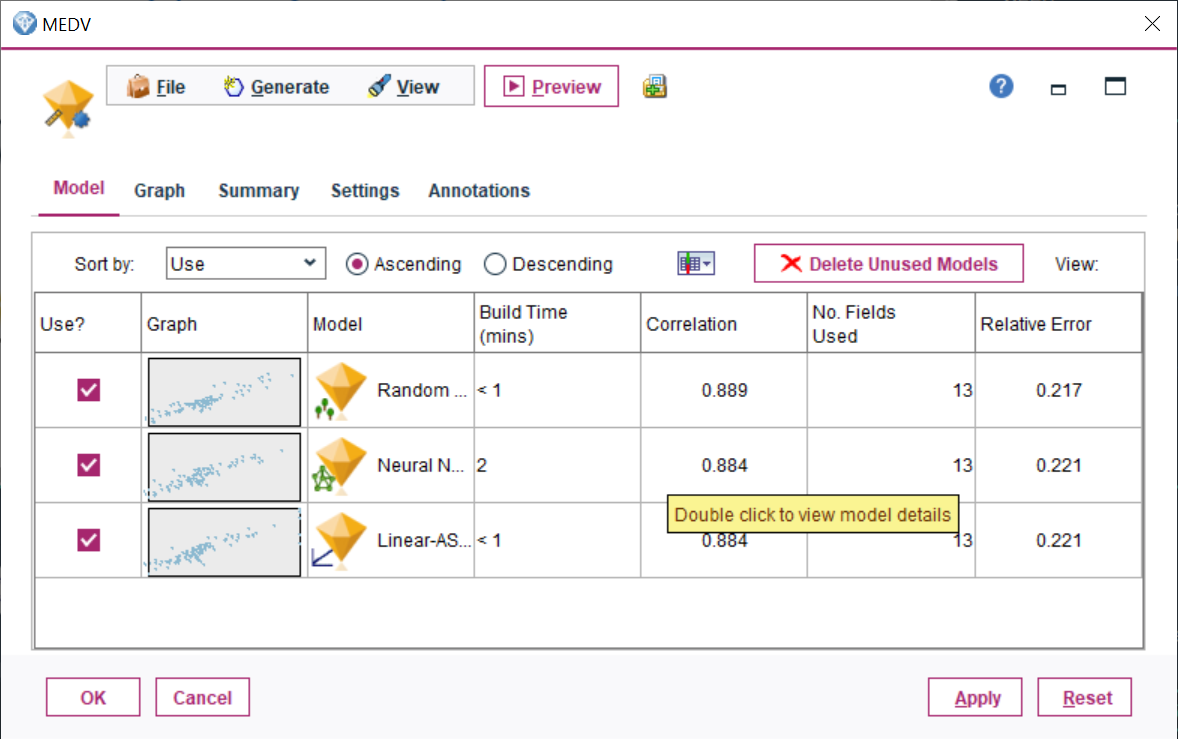
1. Open SPSS Modeler->File>New Stream
2. Insert ‘Excel’ from Source. Import the excel sheet Boston\_Housing.xls
3. Insert ‘Type’ from Field Ops. Change the Measurement (data type) of CHAS and CAT.MEDV to categorical. Change the role of CAT.MEDV to None and MEDV to target.
4. Insert ‘Partition’ from Field Ops. Change training partition size to 60 and testing partition size to 40.
5. Insert ‘Linear’ from Model. Change Model selection method to ‘Include all predictors’. Click on Apply and then Run. A Model summary is generated.
6. Insert ‘Analysis’ from Output. Click on Run. Linear Correlation = 0.885, which shows that the actual price and predicted price values are strongly correlated. Occurrences is the total number of records.



1. Insert ‘Evaluation’ from Graphs. Tick ‘Include best line’ to see how good is our model compared to the best model. Apply and Run.



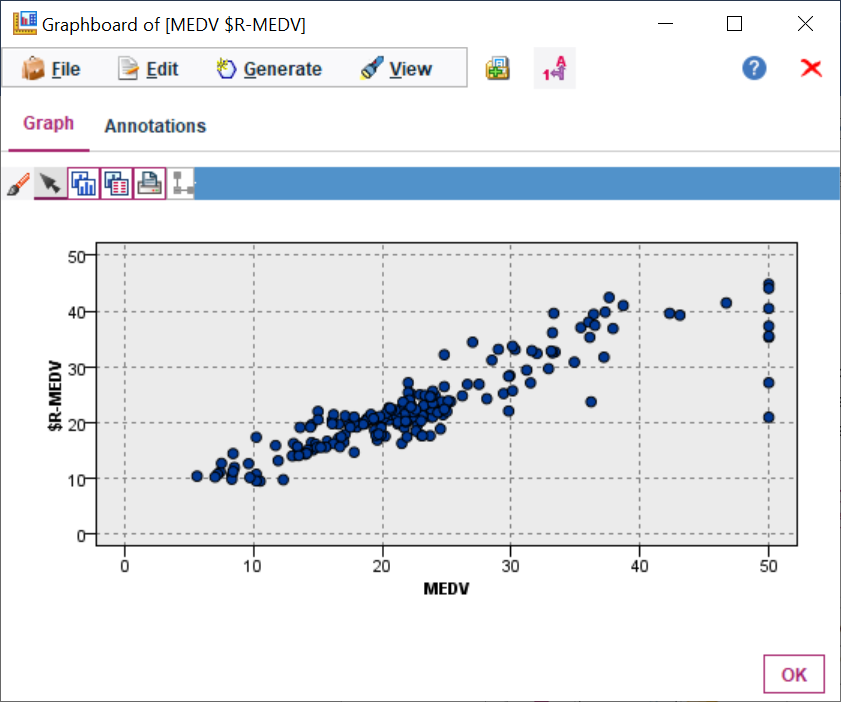
1. Insert ‘Auto Numeric’ from Modelling. Change number of models to 3. Select KNN, SVM as well. Apply and Run. The correlation is highest for Random trees, 0.889 which shows high correlation between predicted price and actual price values.



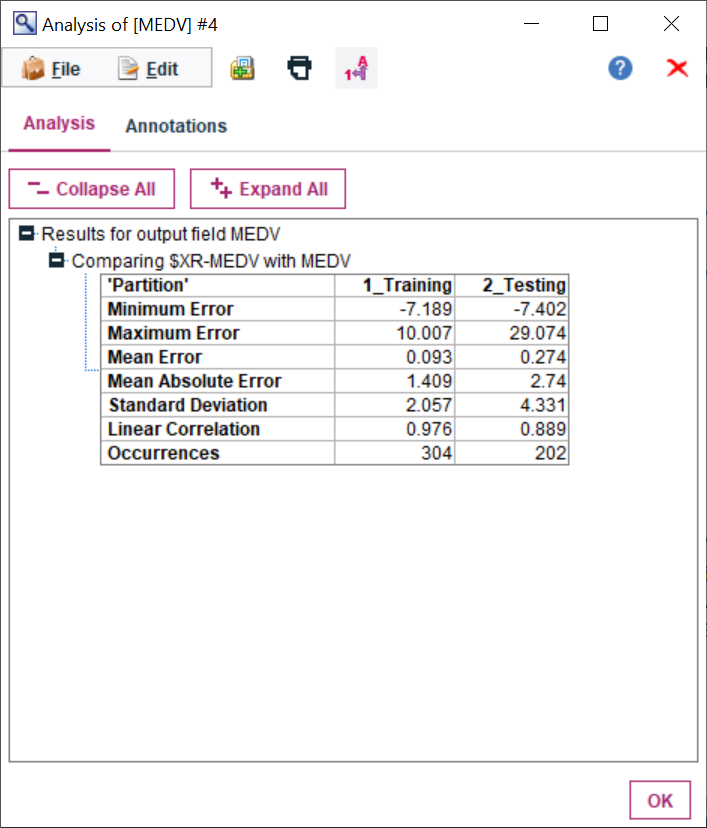
Scatter plot for Random trees:

X axis: actual price value

Y axis: predicted price value

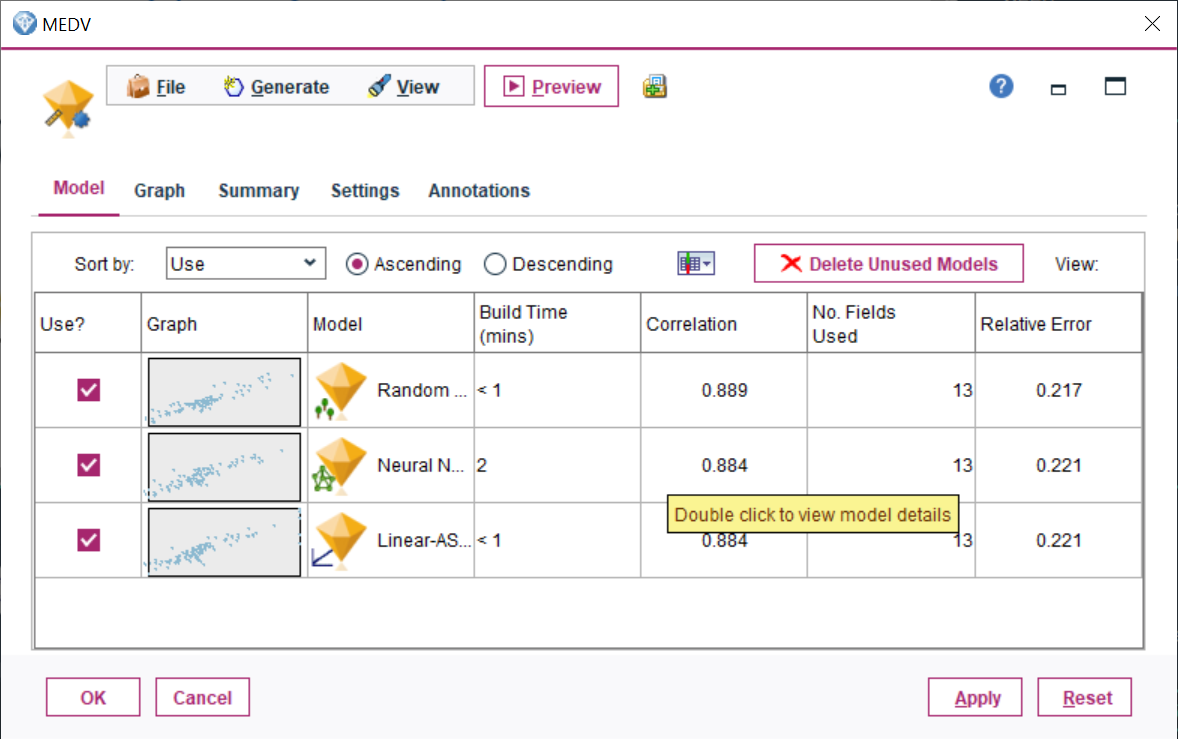


1. Insert ‘Analysis’ from Output. Click on MEDV and delete the other 2 models except Random trees. Click on Analysis, Apply and Run.

’

**Answer: I have created three models and evaluated the Relative Error and Correlation of each of them with train-test split (60-40) dataset:**

* **Random Trees**
* **Neural Network**
* **Linear Regression**



**We can see that the best model for predicting price (MEDV) is Random Trees as it has the lowest Relative Error and highest Correlation.**