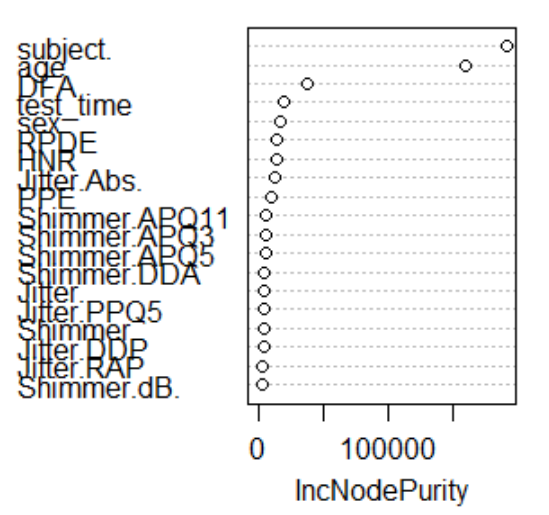
1. Which predictors are statistically significant in predicting UPDRS scores?

Ans:



From the above feature importance diagram, we could see that Subject, Age and DFA are the most important variables in predicting the UPDRS score. Below is the analysis of important variables.

1. The most important variable is the Subject. As a result, total UPDRS varies greatly from subject to subject.
2. Age is the second important factor; this answer greatly relates to the box plots of different subjects and age groups which we have done earlier.
3. Also, the Sexuality and HNR values have a great impact on total\_UPDRS score prediction.
4. Which machine learning model works the best for predicting the scores?

Ans: From the final performance measures of different Machine Learning Models Random Forest performs the best compared to SVM and Neural Nets. Since our dataset is greatly affected by outliers, ensemble methods performed best (which are robust to outliers) compared to other models. We have got the best R2 value of 99.84% variance explained when mtry = 19 (which is equal to number of predictors). This indicates that Bagging Tree performed best compared to regular RF with mtry = 6 (~ p/3). If we select R2 as a measure there is no significant change with mtry = 6 and mtry =19 as with mtry = 6, we get 97.4% variance explained. So, we could say that Random Forest performed best on this data. Neural Nets performed the worst with R2 value of just 2%, but the RMSE value is lower compared to RF model. This indicates that Neural Nets are not robust to outliers and since our dataset has a significant outlier issue, it performed the worst. We have not removed the outliers as, there will be a significant change in the distribution of data and we might not get satisfactory results.

1. Which dysphonia measure is most significant in predicting UPDRS?

Ans: Dysphonia measures are the noise measures which are useful in predicting UPDRS. From the feature importance diagram, we could see that variables DFA, RPDE, HNR, Jitter (Absolute value) and PPE are the noise measures which are most important in predicting UPDRS. Among the Shimmer variables Shimmer.APQ11 is the most important measure for UPDRS. So, a persons UPDRS scores can be predicted based on the HNR, Jitter (Absolute value), Shimmer.APQ11, RPDE and DFA values. This gives us useful info on whether a person has a severe effect or a mild effect of this disease.