SESSION 13: Decision Tree Based Models Assignment 3

5. Problem Statement
1. Use the given link below:
https://archive.ics.uci.edu/ml/machine-learning-databases/00304/
Problem- prediction of the number of comments in the upcoming 24 hours on those blogs, the train data was generated from different base times that may temporally overlap. Therefore, if you simply split the train into disjoint partitions, the underlying time intervals may overlap. Therefore, you should use the provided, temporally disjoint train and test splits to ensure that the evaluation is fair.

a. Interpret the final model coefficients.
b. Plot the model result and compare it with assumptions of the model.

Q1- interpret the final model coefficients

summary(final_model)

coef(final_model) # coefficients of the model

#comments in H Hrs has slope with Independent variables as below:

talking d5 d7 d8 d10 d11

-1.858115e-05 -4.759496e-01 8.609203e-01 1.675394e-01 -1.239555e-01 -2.236221e-03

d12 d13 d16 d17 d19 d20 d22

1.612318e-01 1.276223e-01 1.114969e-02 1.085186e-01 -1.165972e-01 4.201675e-01 -8.837498e-01

d23 cc1 cc2 cc3 cc4 basetime postshre

-2.159461e-01 4.338324e-02 2.196493e-01 -2.272725e-02 -6.728051e-02 -1.933110e-01 2.921963e-03

Hhrs

3.880629e-01

 # Q2- plot the model result and compare it with assumptions of the model

par(mfrow=c(2,2))

plot(final_model)

Model does not pass the test of normality

the data is heteroscadatic

Observations 3528,30608,16432 may have the leverage or potential for influencing the model

##