

ML WITH R PROJECT

Graduate Admission Problem



ATTRIBUTES PROVIDED

- CHANCE OF ADMIT
- SOP
- LOR
- CGPI
- UNIVERSITY RATING
- GRE SCORE
- TOEFL SCORE
- RESEARCH

OUR UNDERSTANDING OF THE PROBLEM

- CONTINUOUS DATA
- LINEAR REGRESSION
- DATA NEEDS TO BE CLEANED



CLEANING THE DATA

ARRANGED DATA

- REMOVED NA VALUES, UNWANTED ELEMENTS
- CONVERTED INTO FACTORS
- RANDOMIZED THE DATA



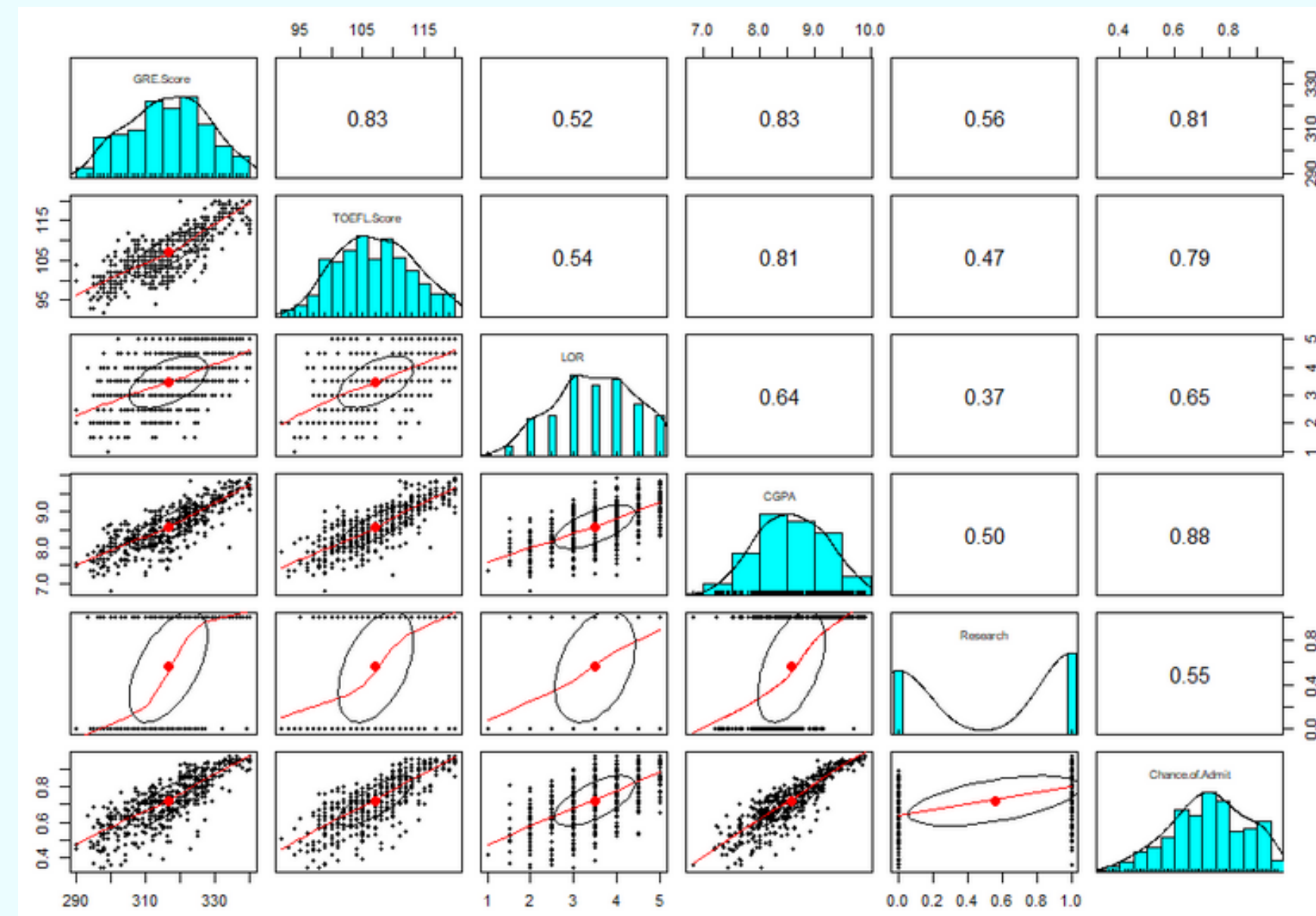
FITTING A MODEL

LINEAR MODEL IS USED TO FIT LINEAR MODELS. IT CAN BE USED TO CARRY OUT REGRESSION, SINGLE STRATUM ANALYSIS OF VARIANCE AND ANALYSIS OF CO-VARIANCE



CORRELATION MATRIX

EXPLORING RELATIONSHIP AMONG VARIOUS FEATURES
HIGHER THE CORRELATION BETTER
IS THE FEATURE PREDICTING THE TARGET VARIABLE





TRAINING THE MODEL

WE USED 70% DATA TO TRAIN THE MODEL AND
30% FOR TESTING PURPOSES.





EVALUATING THE MODEL

ACCURACY: 88%

ERROR CHANCE: 11%

NUMBER OF PREDICTORS USED: 8





IMPROVISING THE MODEL

DIMENSION REDUCED MODEL

removed variables that gave lower beta coefficients
removed

ADDING HIGHER ORDER VARIABLES

adding new variables based on existing variables

USING DIFFERENT ALGORITHMS

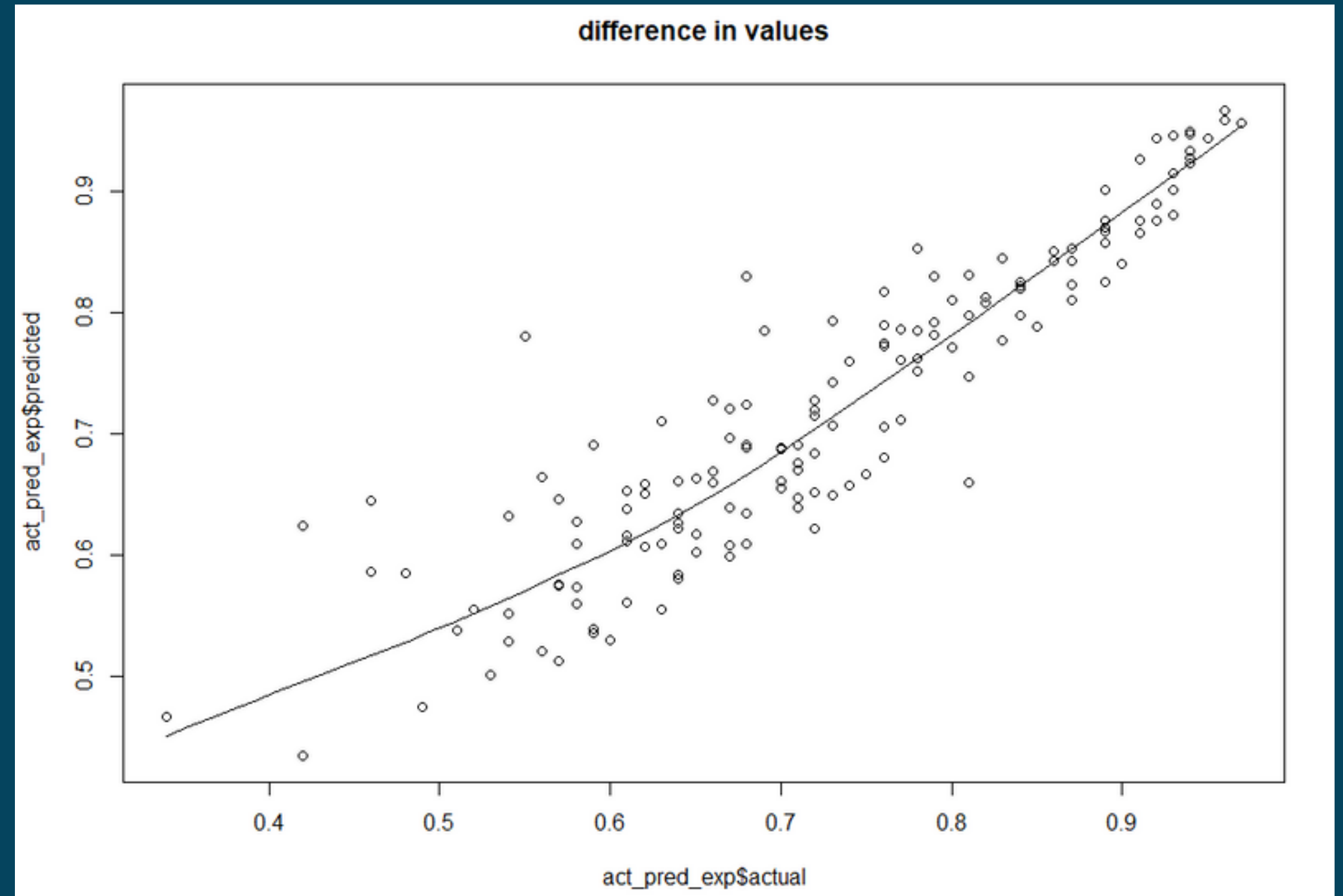
K-Nearest Neighbours
Random forest

Maximized Efficiency

ACCURACY :94%

MAPE:6%

NUMBER OF VARS USED :6



Thank you
ANY QUESTIONS

TEAM MEMBERS:
SAI PRASAD
SHWETA
PAVITHRA
SOORAJ