

SUMMARY OF HOW WE DID THE PROJECT AND WHAT UNDERSTANDING WE GOT

1.DATA CLEANING

- a) FIRST, WE DID THE UNDERSTANDING OF THE DATA
- b) THEN WE PROCEEDED WITH DATA CLEANING AND REMOVED THE REDUNDANT VARIABLES
- c) WE SAW THAT SOME COLUMNS HAVE LABEL "SELECT" WHICH MEANS THEY LEFT IT UNATTENDED SO BETTER TO REPLACE THOSE WITH NULL VALUES
- d) AFTER THAT WE REMOVED THE COLUMNS WHICH HAVE MORE THAN 30 PERCENT OF THE NULL VALUES
- e) AND FOR THE REST OF THE MISSING VALUES, WE HAVE IMPUTED IT BY MAXIMUM OCCURENCES
- f) WE FOUND OUT "GOOGLE WAS PRESENT TWICE ONCE IN CAPITAL LETTERS AND ONCE IN SMALL LETTERS
- g) SO, WE FIXED THIS ISSUE BY CHANGING THE LABELS IN ONE FORMAT

2.TRANSFORMATION OF DATA

- a) NOW WE CHANGED THE MULTICATEGORY LABELS INTO DUMMY VARIABLES AND BINARY VARIABLES INTO "0" AND "1"
- b) WE CHECKED FOR THE OUTLIERS AND THEN CREATED BINS FOR THEM
- c) CHECKED FOR ALL REDUNDANT AND REPEATED COLUMNS AND THEN REMOVED THEM

3.DATA PREPARATION

- a) WE HAVE NOW SPLIT THE DATA SET INTO TEST AND TRAIN DATA SET
- b) AFTER THAT FOR CHECKING THE CORRELATIONS BETWEEN THE VARIABLES WE DREW THE HEATMAP
- c) WHEN WE FOUND SOME CORRELATIONS THEN THEY WERE DROPPED

4. BUILDING A MODEL

- a) WE CHOSE OUR FINAL MODEL WITH RFE 19 VARIABLE AS IT HAS MORE STABILITY AND ACCURACY THAN THE OTHER
- b) FOR OUR FINAL MODEL WE CHECKED THE OPTIMAL PROBABILITY CUTOFF BY FINDING POINTS AND CHECKING THE ACCURACY, SENSITIVITY AND SPECIFICITY
- c) WE CHOSE ONE CONVERGENT POINT AND SELECTED THAT POINT FOR CUTOFF AND FOR PREDICTION OF FINAL OUTCOMES
- d) WE ALSO CHECKED PRECISION AND RECALL WITH ACCURACY, SENSITIVITY, AND SPECIFICITY FOR OUR FINAL MODEL AND THE TRADEOFFS
- e) THE PREDICTION MADE IN TEST SET AND PREDICTED VALUE WAS RECORDED
- f) WE DID MODEL EVALUATION ON TEST SET LIKE CHECKING ITS ACCURACY, RECALL/SENSIVITY TO FIND HOW THE MODEL IS
- g) AFTER THAT WE FOUND THAT SCORE OF SENSITIVITY AND ACCURACY FROM OUR FINAL TEST MODEL IS IN ACCEPTABLE RANGE
- h) WE HAVE GIVEN LEAD SCORE TO TEST DATA SET CONCLUDING THAT HIGH LEAD SCORE ARE HOT LEADS AND LOW LEAD SCORE ARE NOT HOT LEADS

CONCLUSION

- a) THE TEST SET IS HAVING ACCURACY, RECALL/SENSITIVITY IN AN ACCEPTABLE RANGE
- b) OUR MODEL IS HAVING ACCURACY AND IT WILL ALSO ADJUST WITH BUSINESS REQUIREMENT IN COMING FUTURE
- c) TOP FEATURES FOR GOOD CONVERSION RATE:
 - LAST NOTABLE ACTIVITY_ HAD A PHONE CONVERSATION
 - LEAD ORIGIN_LEAD ADD FORM
 - WHAT IS YOUR CURRENT OCCUPATION_ WORKING PROFESSIONAL