

SUMMARY OF HOW WE DID AND WHAT UNDERSTANDING WE GOT

1.DATA CLEANING

1. FIRST WE DID THE UNDERSTANDING OF THE DATA
2. THEN WE PROCEEDED WITH DATA CLEANING AND REMOVED THE REDUNDANT VARIABLES
3. WE SAW THAT SOME COLUMNS HAVE LABEL "SELECT" WHICH MEANS THEY LEFT IT UNATTENDED SO BETTER TO REPLACE THOSE WITH NULL VALUES
4. AFTER THAT WE REMOVED THE COLUMNS WHICH HAVE MORE THAN 30 PERCENT OF THE NULL VALUES
5. AND FOR THE REST OF THE MISSING VALUES WE HAVE IMPUTED IT BY MAXIMUM OCCURENCES
6. WE FOUND OUT "GOOGLE" WAS PRESENT TWICE ONCE IN CAPITAL LETTERS AND ONCE IN SMALL LETTERS
7. SO WE FIXED THIS ISSUE BY CHANGING THE LABELS IN ONE FORMAT

2.TRANSFORMATION OF DATA

1. NOW WE CHANGED THE MULTI CATEGORY LABELS INTO DUMMY VARIABLES AND BINARY VARIABLES INTO "0" AND "1"
2. WE CHECKED FOR THE OUTLIERS AND THEN CREATED BINS FOR THEM
3. CHECKED FOR ALL REDUNDANT AND REPEATED COLUMNS AND THEN REMOVED THEM

3.DATA PREPARATION

1. WE HAVE NOW SPLIT THE DATA SET INTO TEST AND TRAIN DATA SET
2. AFTER THAT FOR CHECKING THE CORRELATIONS BETWEEN THE VARIABLES WE DREW THE HEATMAP
3. WHEN WE FOUND SOME CORRELATIONS THEN THEY WERE DROPPED

4. BUILDING A MODEL

1. WE CHOSE OUR FINAL MODEL WITH RFE 19 VARIABLE AS IT HAS MORE STABILITY AND ACCURACY THAN THE OTHER
2. FOR OUR FINAL MODEL WE CHECKED THE OPTIMAL PROBABILITY CUTOFF BY FINDING POINTS AND CHECKING THE ACCURACY , SENSITIVITY AND SPECIFICITY
3. WE CHOSE ONE CONVERGENT POINT AND SELECTED THAT POINT FOR CUTOFF AND FOR PREDICTION OF FINAL OUTCOMES

4. WE ALSO CHECKED PRECISION AND RECALL WITH ACCURACY , SENSITIVITY , AND SPECIFICITY FOR OUR FINAL MODEL AND THE TRADEOFFS

5.THE PREDICTION MADE IN TEST SET AND PREDICTED VALUE WAS RECORDED

6. WE DID MODEL EVALUATION ON TEST SET LIKE CHECKING ITS ACCURACY , RECALL/SENSITIVITY TO FIND HOW THE MODEL IS

7.AFTER THAT WE FOUND THAT SCORE OF SENSITIVITY AND ACCURACY FROM OUR FINAL TEST MODEL IS IN ACCEPTABLE RANGE

8.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

1 . THE TEST SET IS HAVING ACCURACY , RECALL/SENSITIVITY IN AN ACCEPTABLE RANGE

2. OUR MODEL IS HAVING ACCURACY AND IT WILL ALSO ADJUST WITH BUSINESS REQUIREMENT IN COMING FUTURE

3. 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