WHAT ARE SOME GOOD PREDICTORS OF WHETHER YOU'LL STAY IN THIS WEEKEND?

THE WEATHER

DO YOU HAVE SOMETHING URGENT TO DO AT WORK?

IS YOUR SPOUSE/SIGNIFICANT OTHER IN TOWN?

IS THERE AN IMPORTANT GAME ON TV?

USUALLY, THESE WOULD HAVE A CERTAIN ORDER - HOW IMPORTANT ARE EACH OF THESE PREDICTORS TO YOUR DECISION OF STAYING IN?

THE WEATHER

DO YOU HAVE SOMETHING URGENT TO DO AT WORK?

IS THERE AN IMPORTANT GAME ON TV?

IS YOUR SPOUSE/SIGNIFICANT OTHER IN TOWN?

SO, WHAT DOES THIS ORDER IMPLY?

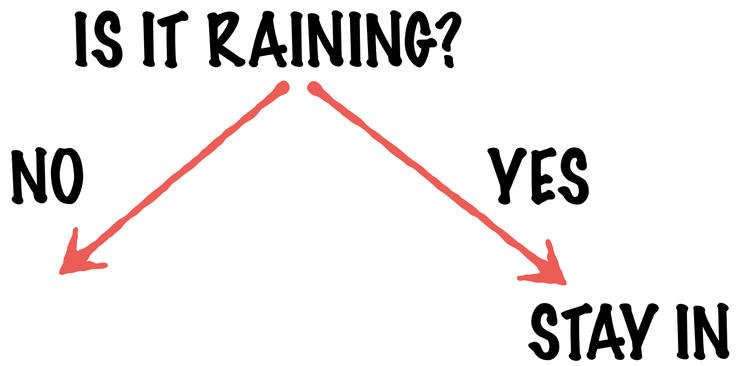
IF YOU ARE THINKING OF GOING OUT, FIRST YOU WOULD LOOK AT THE WEATHER

USUALLY, THESE WOULD HAVE A CERTAIN ORDER - HOW IMPORTANT ARE EACH OF THESE PREDICTORS TO YOUR DECISION OF STAYING IN?

SO, WHAT DOES THIS ORDER IMPLY?

IF YOU ARE THINKING OF GOING OUT, FIRST YOU WOULD LOOK AT THE WEATHER

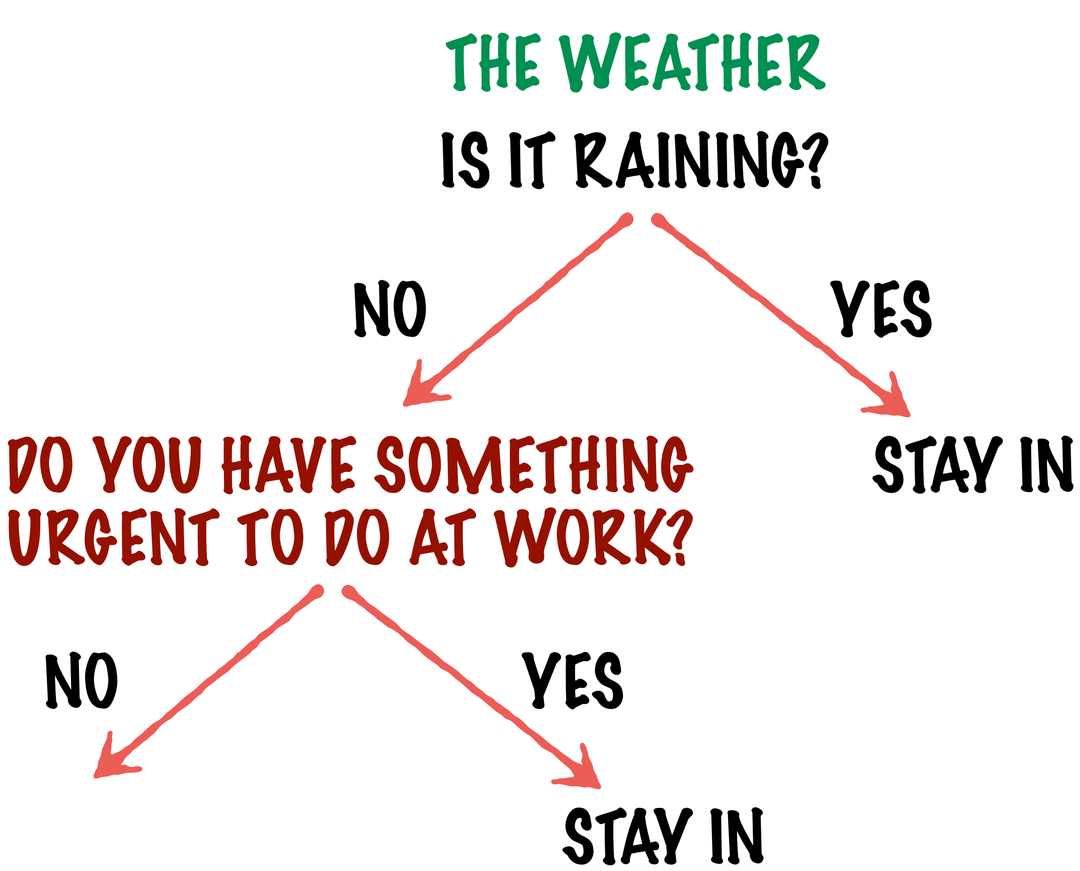
THE WEATHER



DO YOU HAVE SOMETHING URGENT TO DO AT WORK?

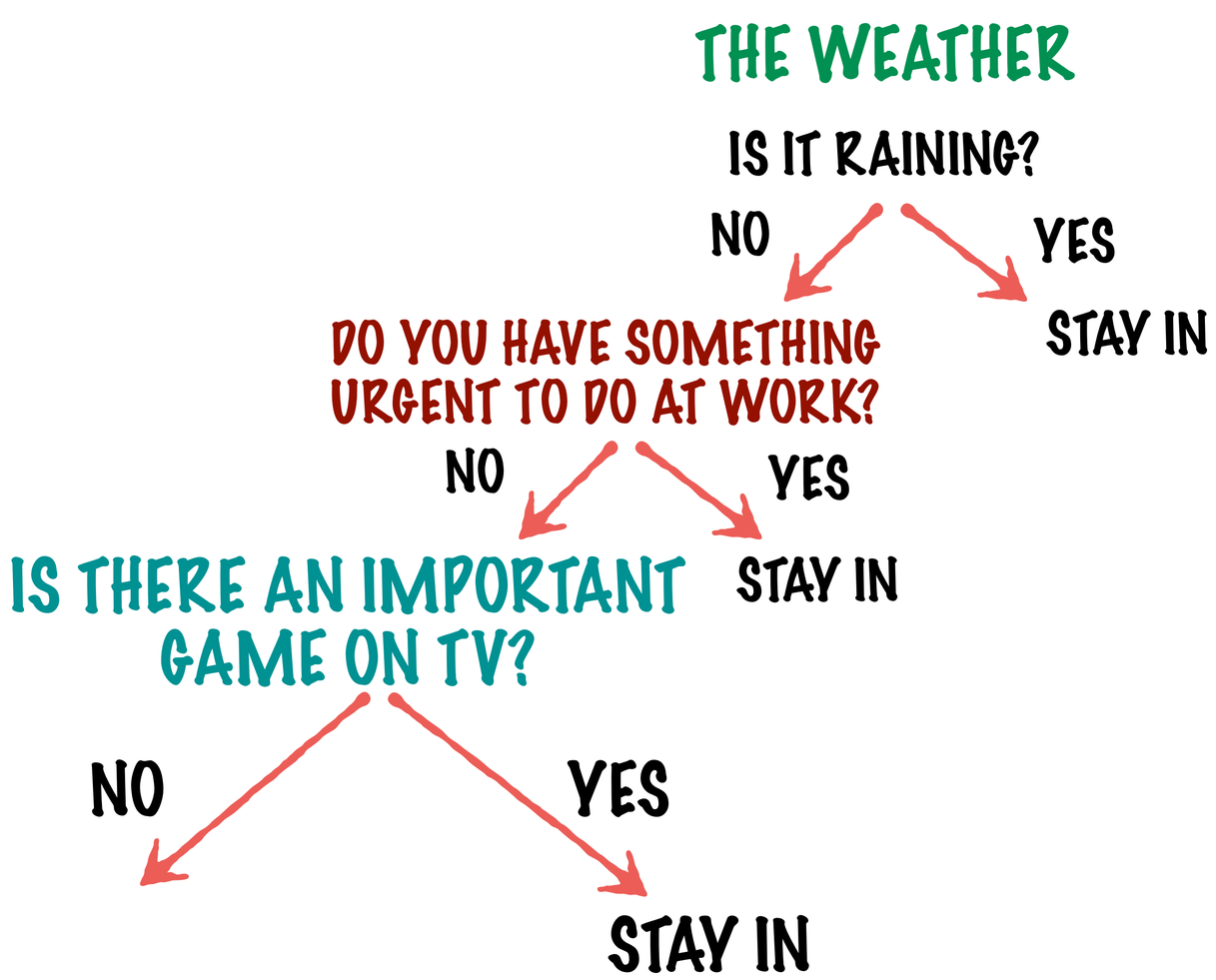
IS THERE AN IMPORTANT GAME ON TV?

IS YOUR SPOUSE/SIGNIFICANT OTHER IN TOWN?



IS THERE AN IMPORTANT GAME ON TV?

IS YOUR SPOUSE/SIGNIFICANT OTHER IN TOWN?

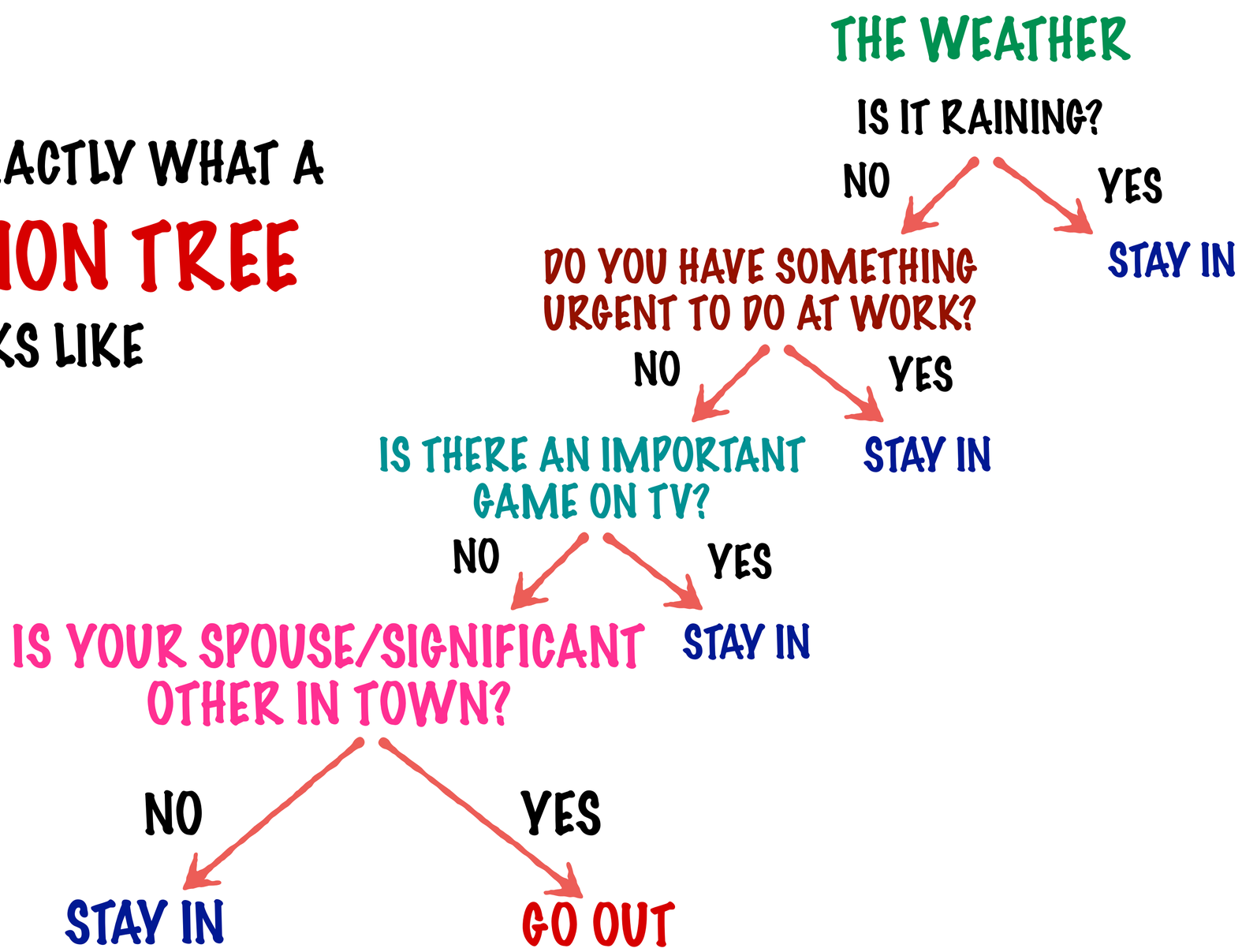


IS YOUR SPOUSE/SIGNIFICANT OTHER IN TOWN?

THIS IS EXACTLY WHAT A

DECISION TREE

LOOKS LIKE



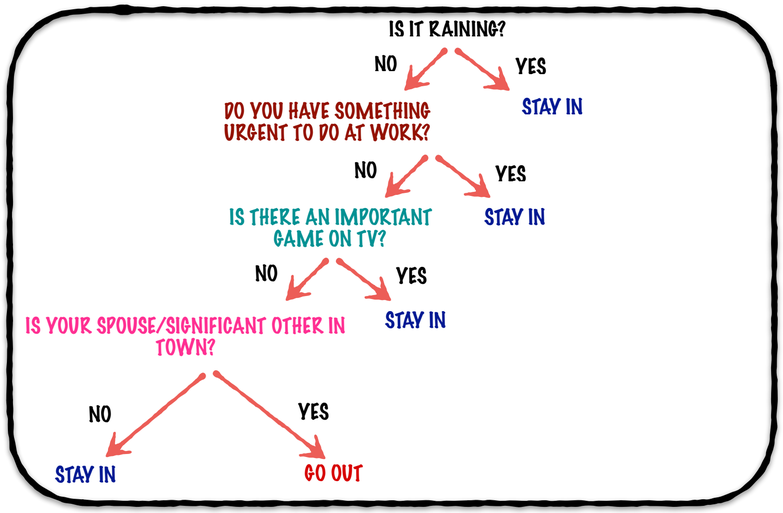
THIS IS EXACTLY WHAT A

DECISION TREE

LOOKS LIKE

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

INPUT VARIABLES/ PREDICTORS

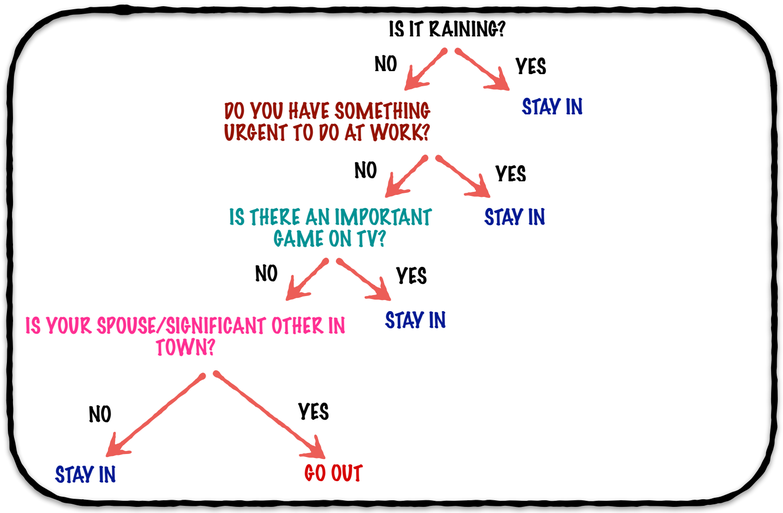


OUTCOME/OUTPUT VARIABLES

A DECISION TREE PREDICTS THE OUTCOME
GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

INPUT VARIABLES/
PREDICTORS



OUTCOME/OUTPUT
VARIABLES

INPUT VARIABLES/PREDICTORS

THE INPUT VARIABLES COULD BE CATEGORICAL OR CONTINUOUS

THE WEATHER

IS IT RAINING?

YES

NO

HOW HOT IS IT?

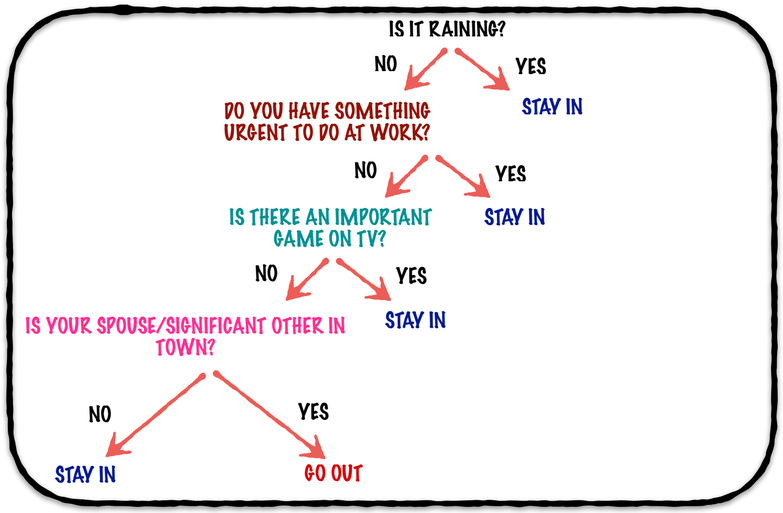
TEMP. BETWEEN 15^0 C AND 35^0 C

TEMP > 35^0 C, OR TEMP< 15^0 C

A DECISION TREE PREDICTS THE OUTCOME
GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

INPUT VARIABLES/
PREDICTORS



OUTCOME/OUTPUT
VARIABLES

OUTCOME/OUTPUT VARIABLES

THE OUTPUT VARIABLES COULD BE CATEGORICAL OR CONTINUOUS

WILL YOU BE STAYING IN OR GOING OUT?

STAYING IN GOING OUT

IS THIS EMAIL SPAM OR HAM?

SPAM HAM

IS THIS CUSTOMER GOING TO DEFAULT ON THEIR CARD PAYMENT?

YES NO

OUTCOME/OUTPUT VARIABLES

CLASSIFICATION TREES

THE OUTPUT VARIABLES COULD BE CATEGORICAL OR CONTINUOUS

WHAT WILL THE SALE PRICE OF A HOUSE BE?

HOW MANY DAYS WILL A PATIENT STAY IN THE HOSPITAL?

REGRESSION TREES

CATEGORICAL

WILL YOU BE STAYING IN OR GOING OUT?

STAYING IN GOING OUT

IS THIS EMAIL SPAM OR HAM?

SPAM HAM

IS THIS CUSTOMER GOING TO DEFAULT ON THEIR CARD PAYMENT?

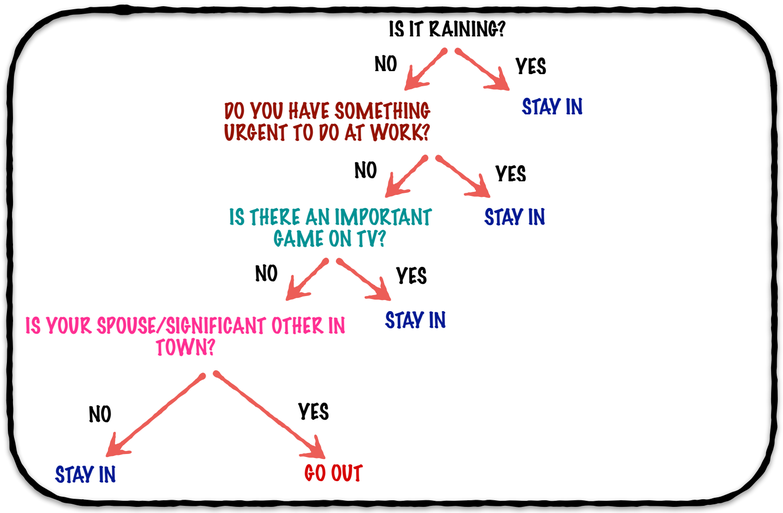
YES NO

DECISION TREES CAN BE USED TO SOLVE CLASSIFICATION OR REGRESSION PROBLEMS

A DECISION TREE PREDICTS THE OUTCOME
GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

INPUT VARIABLES/
PREDICTORS



OUTCOME/OUTPUT
VARIABLES

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

THE LEAVES OF THE TREE ARE

THE OUTCOMES

IN A CLASSIFICATION PROBLEM THESE ARE THE

CLASS LABELS

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

THE LEAVES OF THE TREE ARE THE OUTCOMES

THE BRANCHES ARE THE COMBINATIONS OF THE PREDICTOR VALUES THAT LEAD TO A PARTICULAR OUTCOME

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

THE LEAVES OF THE TREE ARE THE OUTCOMES

THE BRANCHES ARE THE COMBINATIONS OF THE PREDICTOR VALUES THAT LEAD TO A PARTICULAR OUTCOME

IF YOUR FEATURE VECTOR IS
RAINING (1/0), WORK (1/0), GAME(1/0) , SPOUSE (1/0)

THIS BRANCH REPRESENTS

(1,X,X,X)

(0,1,X,X)

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

A DECISION TREE USUALLY ALSO GIVES US THE CONDITIONAL PROBABILITY OF THE OUTCOME GIVEN THE VALUES OF THE INPUT VARIABLES

P(STAYING IN/IT IS RAINING)

P(STAYING IN/
(IT IS NOT RAINING AND
YOU HAVE TO WORK))

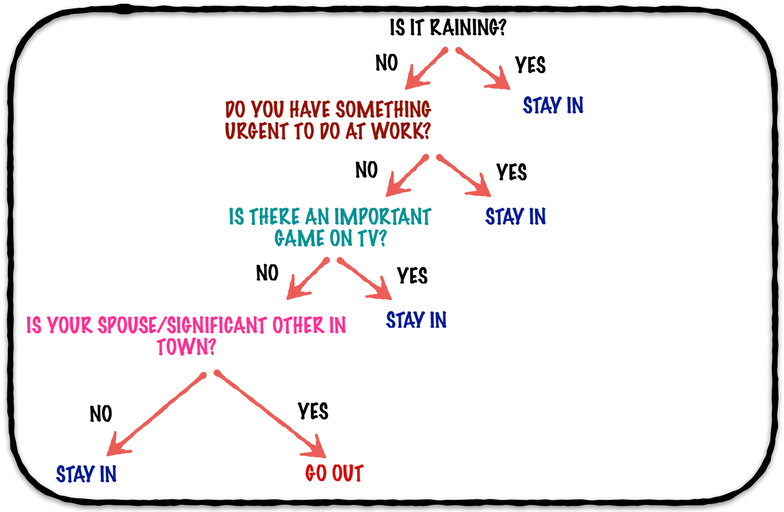
GIVEN THE INPUT VALUES, THE OUTCOMES ARE NOT DETERMINISTIC

THE LEAVES OF THE DECISION TREE REPRESENT THE MOST LIKELY OUTCOME - THE ONE WITH THE HIGHEST CONDITIONAL PROBABILITY GIVEN THE VALUE OF THE VARIABLE

A DECISION TREE PREDICTS THE OUTCOME GIVEN THE VALUES OF INPUT VARIABLES

DECISION TREE

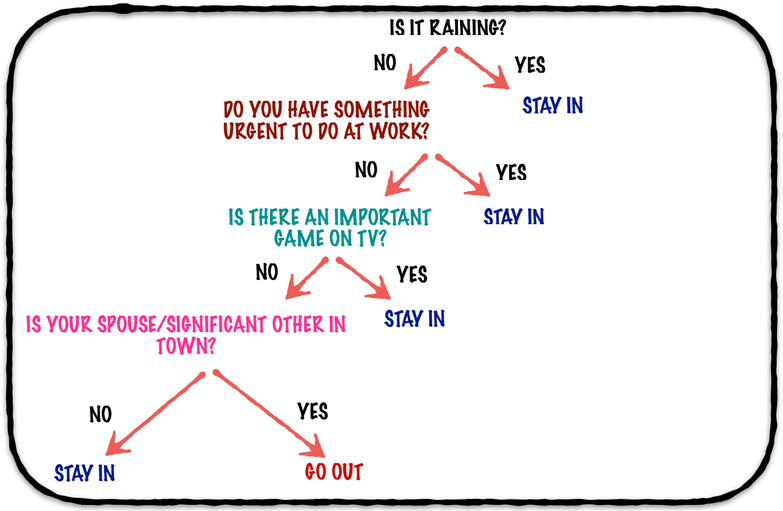
INPUT VARIABLES/ PREDICTORS



OUTCOME/OUTPUT VARIABLES

DECISION TREE

INPUT VARIABLES/ PREDICTORS



OUTCOME/OUTPUT VARIABLES

DECISION TREE LEARNING

IS THE PROCESS OF CREATING/LEARNING A DECISION TREE FROM TRAINING DATA

RECURSIVE PARTITIONING

IS THE MOST COMMON STRATEGY FOR DECISION TREE LEARNING

ID3 C4.5

CART CHAID

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

DECISION TREE LEARNING

IS THE PROCESS OF CREATING/ LEARNING A DECISION TREE FROM TRAINING DATA

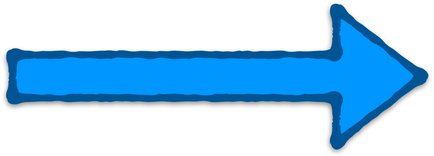
START WITH TRAINING DATA
IN THE FORM OF
FEATURE VECTOR, LABEL/OUTCOME

IF YOUR FEATURE VECTOR IS
RAINING (1/0), TEMPERATURE

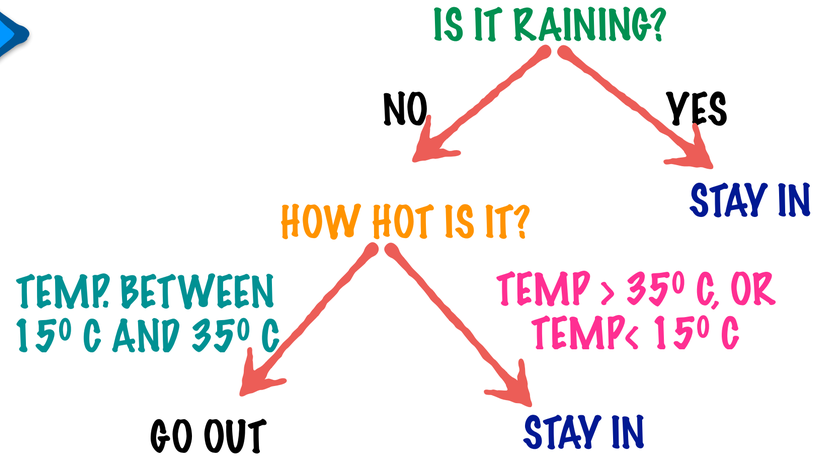
YOUR TRAINING DATA WILL LOOK LIKE

(1, 20^0 C) , STAY IN

(0, 25^0 C) , GO OUT



LEARN A DECISION TREE
WHICH IS USED TO
CLASSIFY/PREDICT A
NEW INSTANCE



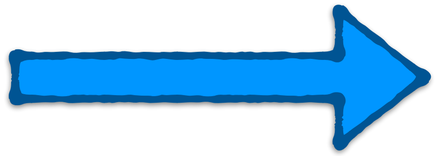
DECISION TREE LEARNING

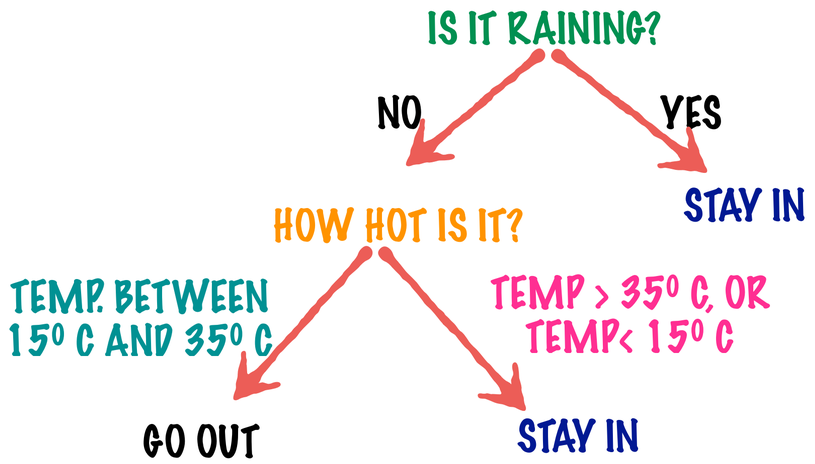
IF YOUR FEATURE VECTOR IS
RAINING (1/0), TEMPERATURE

YOUR TRAINING DATA WILL LOOK LIKE

(1, 20° C) , STAY IN

(0, 25° C) , GO OUT





THE TREE TELLS US THE ORDER IN WHICH THE PREDICTORS NEED TO BE LOOKED AT

FIRST, WHETHER IT IS RAINING
THEN THE TEMPERATURE

DECISION TREE LEARNING

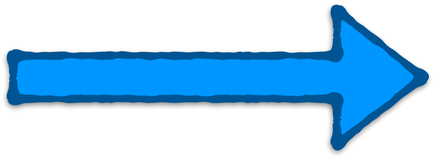
IF YOUR FEATURE VECTOR IS
RAINING (1/0), TEMPERATURE

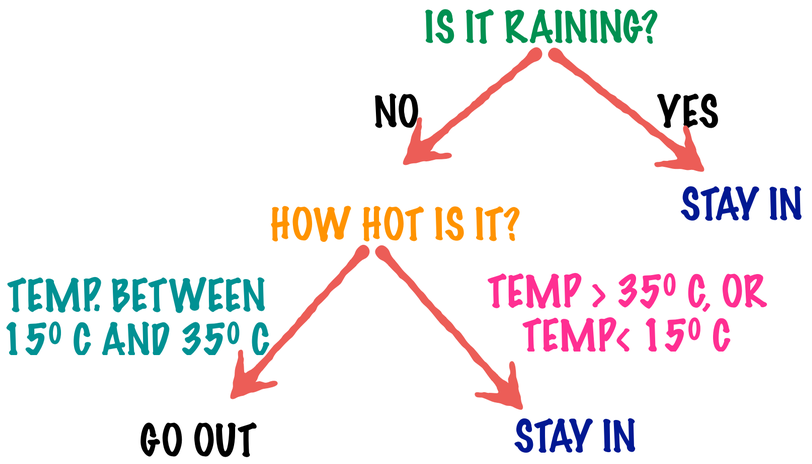
YOUR TRAINING DATA WILL LOOK LIKE

(1, 20° C) , STAY IN

(0, 25° C) , GO OUT

THE TREE TELLS US THE ORDER IN WHICH THE PREDICTORS NEED TO BE LOOKED AT





IF THE PREDICTOR VARIABLE IS CONTINUOUS, THE TREE TELLS US THE RANGES WHICH ARE IMPORTANT TO THE DECISION

THE TEMPERATURE VALUE IS SPLIT INTO THE RANGES - 15-35; <15; >35

EACH RANGE LEADS TO A DIFFERENT OUTCOME

DECISION TREE LEARNING

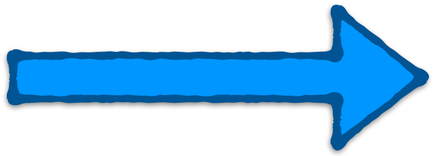
IF YOUR FEATURE VECTOR IS
RAINING (1/0), TEMPERATURE

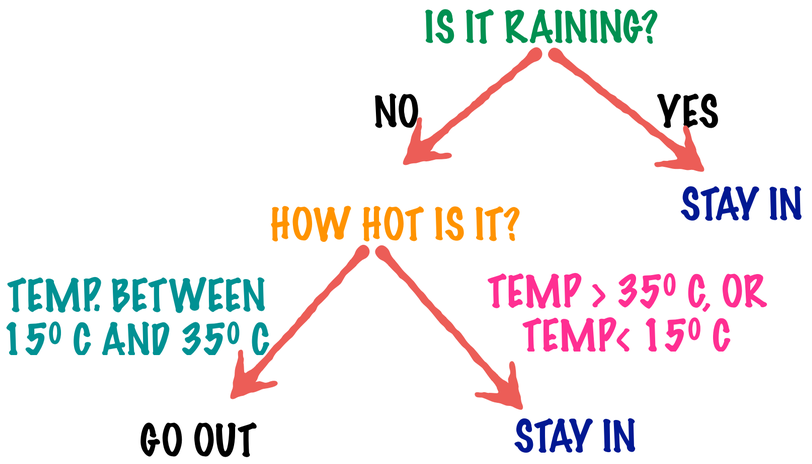
YOUR TRAINING DATA WILL LOOK LIKE

(1, 20° C) , STAY IN

(0, 25° C) , GO OUT

THE TREE TELLS US THE ORDER IN WHICH THE PREDICTORS NEED TO BE LOOKED AT

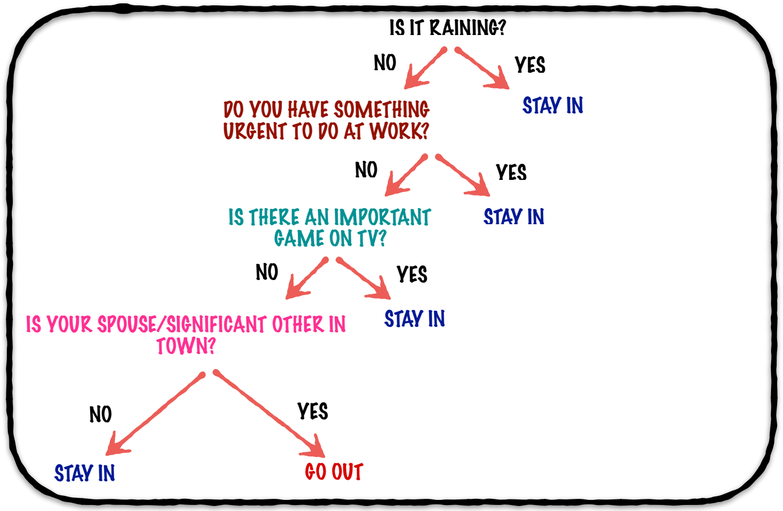




IF THE PREDICTOR VARIABLE IS CONTINUOUS, THE TREE TELLS US THE RANGES WHICH ARE IMPORTANT TO THE DECISION

DECISION TREE

INPUT VARIABLES/ PREDICTORS



OUTCOME/OUTPUT VARIABLES

DECISION TREE LEARNING

IS THE PROCESS OF CREATING/LEARNING A DECISION TREE FROM TRAINING DATA

RECURSIVE PARTITIONING

IS THE MOST COMMON STRATEGY FOR DECISION TREE LEARNING

ID3 C4.5

CART CHAID

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

RECURSIVE PARTITIONING

IS THE MOST COMMON STRATEGY FOR DECISION TREE LEARNING

IT INVOLVES SPLITTING THE TRAINING DATA INTO SUBSETS BASIS THE INPUT VARIABLES/ATTRIBUTES

FOR EACH SPLIT, ONE ATTRIBUTE IS CHOSEN TO BE THE BASIS OF THE SPLIT

LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

GIVEN THE TYPE OF HOUSING, RENT/BEDROOM AND THE YEAR IT WAS BUILT

PREDICT THE CITY TO WHICH A RESIDENCE BELONGS

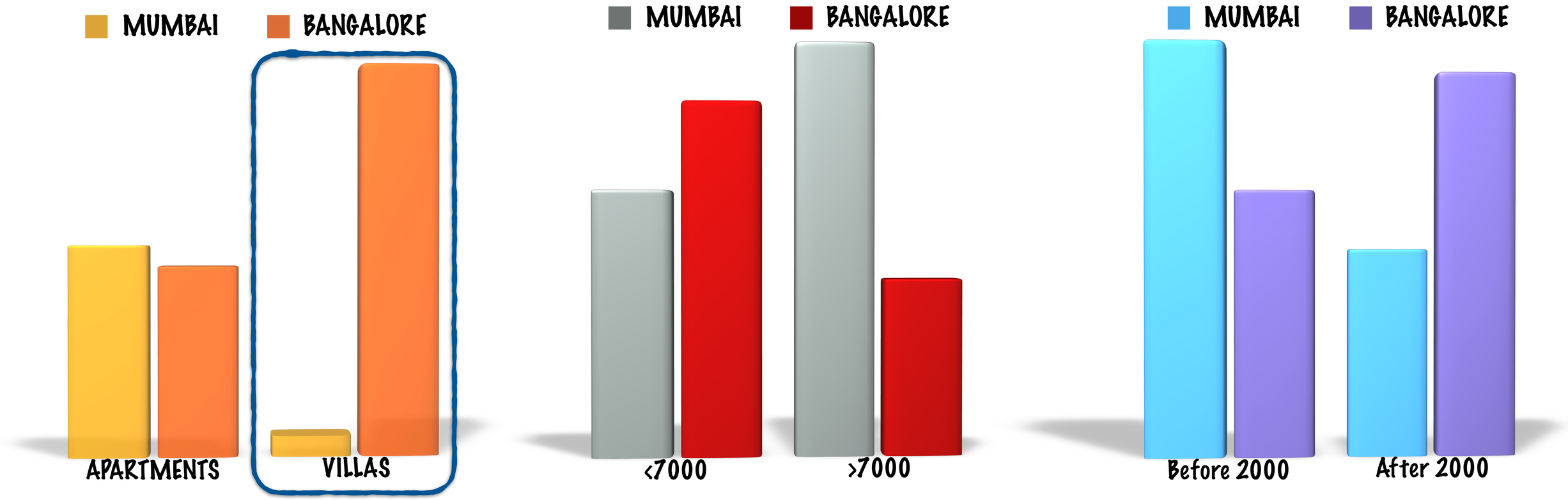
MUMBAI (OR) BANGALORE

LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

GIVEN THE TYPE OF HOUSING, RENT/BEDROOM AND THE YEAR IT WAS BUILT

PREDICT THE CITY TO WHICH A RESIDENCE BELONGS

DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY

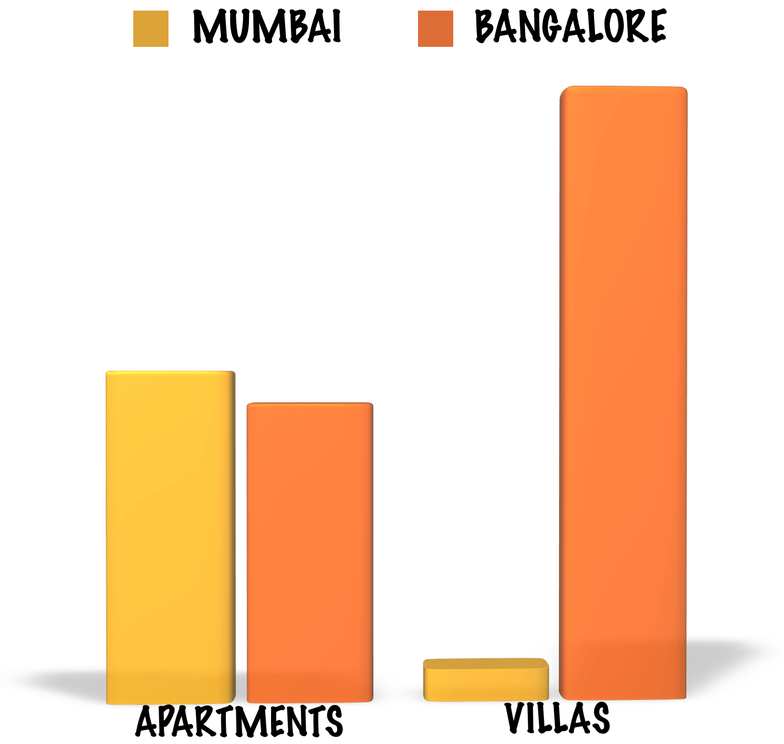


LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

GIVEN THE TYPE OF HOUSING, RENT/BEDROOM AND THE YEAR IT WAS BUILT

PREDICT THE CITY TO WHICH A RESIDENCE BELONGS

DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY



THE TYPE OF HOUSING SEEMS TO BE THE CLEAREST INDICATOR OF WHETHER A RESIDENCE BELONGS TO MUMBAI OR BANGALORE

IF IT IS A VILLA, THEN IT MOST LIKELY BELONGS TO BANGALORE

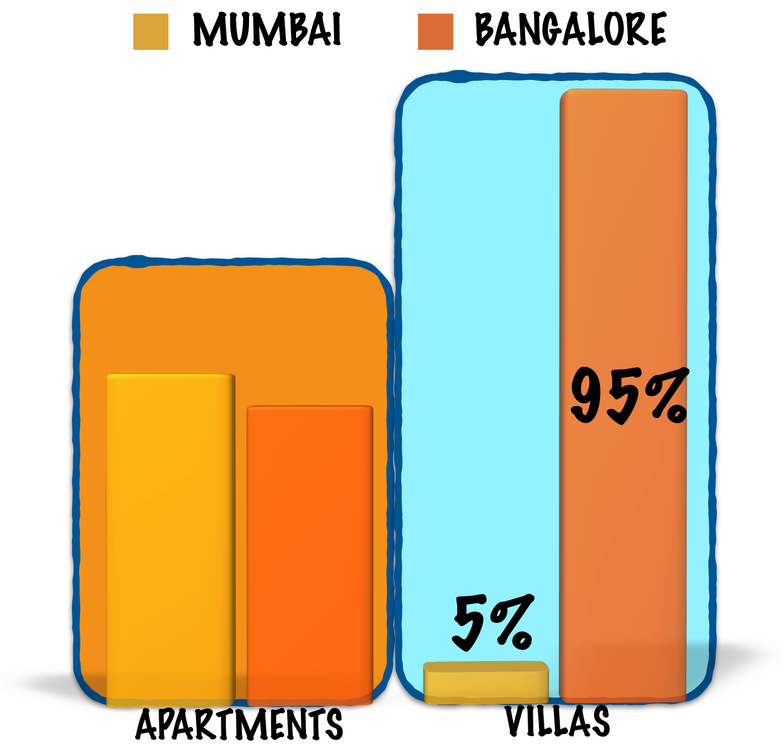
IF IT IS AN APARTMENT, WE ARE STILL NOT SURE WHICH CITY IT BELONGS TO

LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

GIVEN THE TYPE OF HOUSING, RENT/BEDROOM AND THE YEAR IT WAS BUILT

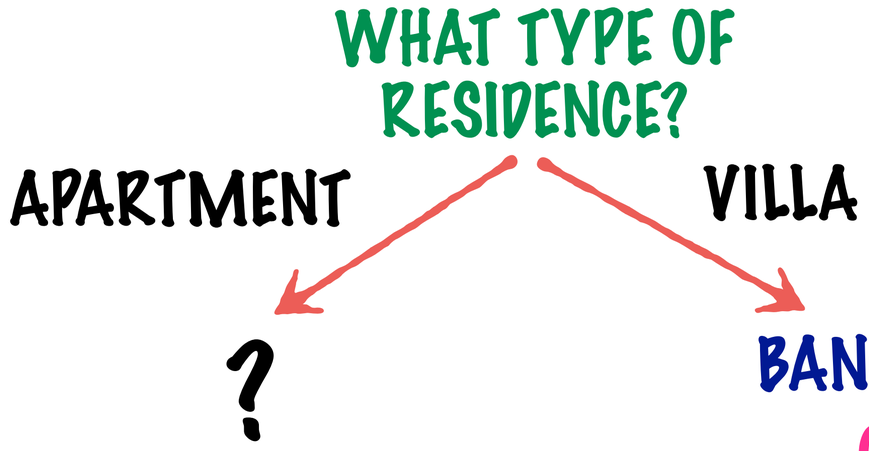
PREDICT THE CITY TO WHICH A RESIDENCE BELONGS

DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY



WE'VE DIVIDED OUR DATA INTO 2 SUBSETS - APARTMENTS AND VILLAS

WE NOW HAVE THE FIRST NODE OF OUR DECISION TREE

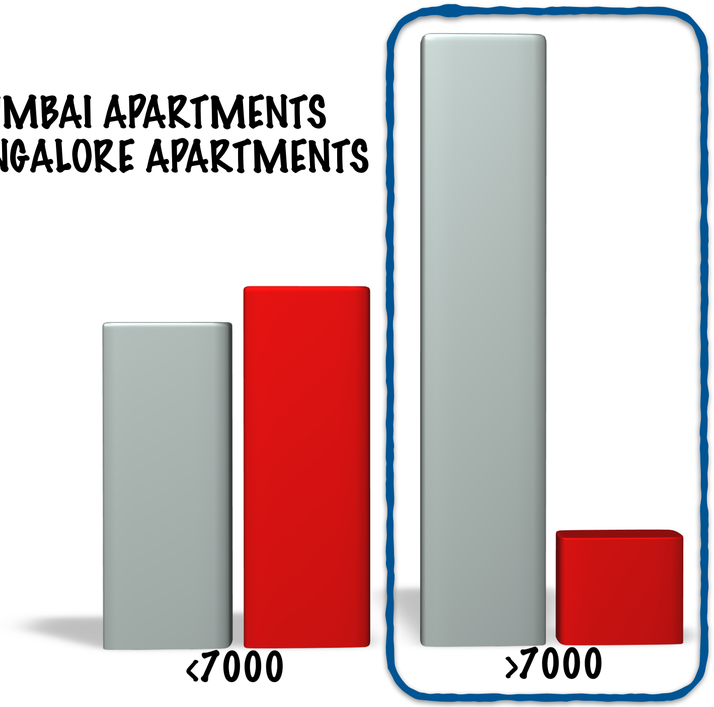


LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

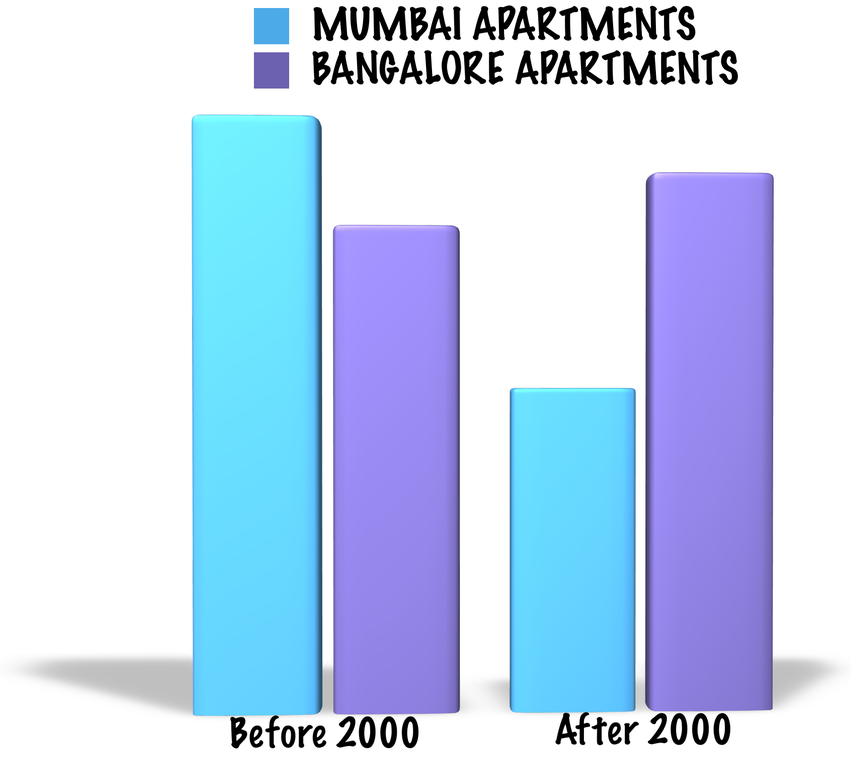
DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY

WE'VE DIVIDED OUR DATA INTO 2 SUBSETS - APARTMENTS AND VILLAS

DRAW A HISTOGRAM FOR EACH OF THE REMAINING ATTRIBUTES FOR ONLY THE APARTMENTS IN EACH CITY



REPEAT THIS PROCESS RECURSIVELY, FOR EACH SUBSET

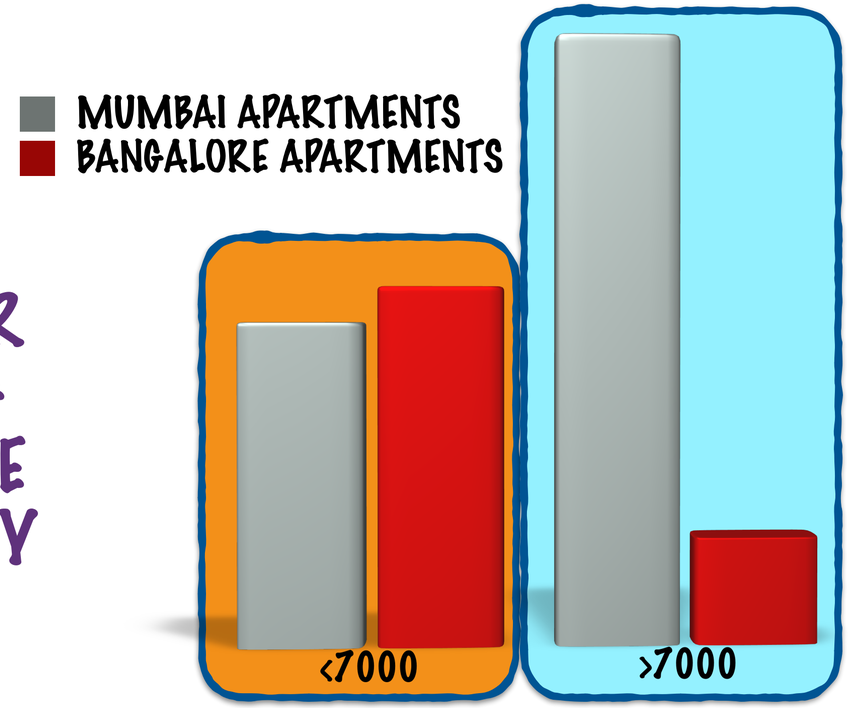


LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY

WE'VE DIVIDED OUR DATA INTO 2 SUBSETS - APARTMENTS AND VILLAS

DRAW A HISTOGRAM FOR EACH OF THE REMAINING ATTRIBUTES FOR ONLY THE APARTMENTS IN EACH CITY



REPEAT THIS PROCESS RECURSIVELY, FOR EACH SUBSET

WHEN YOU LOOK ONLY AT APARTMENTS, THE RENT/ BEDROOM SEEMS TO BE A GOOD PREDICTOR OF THE CITY

NOW, WE'LL DIVIDE THE APARTMENTS INTO TWO SUBSETS, RENT >7000, RENT <7000

LET'S LOOK AT A GREEDY ALGORITHM FOR LEARNING A DECISION TREE

DRAW A HISTOGRAM FOR EACH ATTRIBUTE, FOR RESIDENCES IN EACH CITY

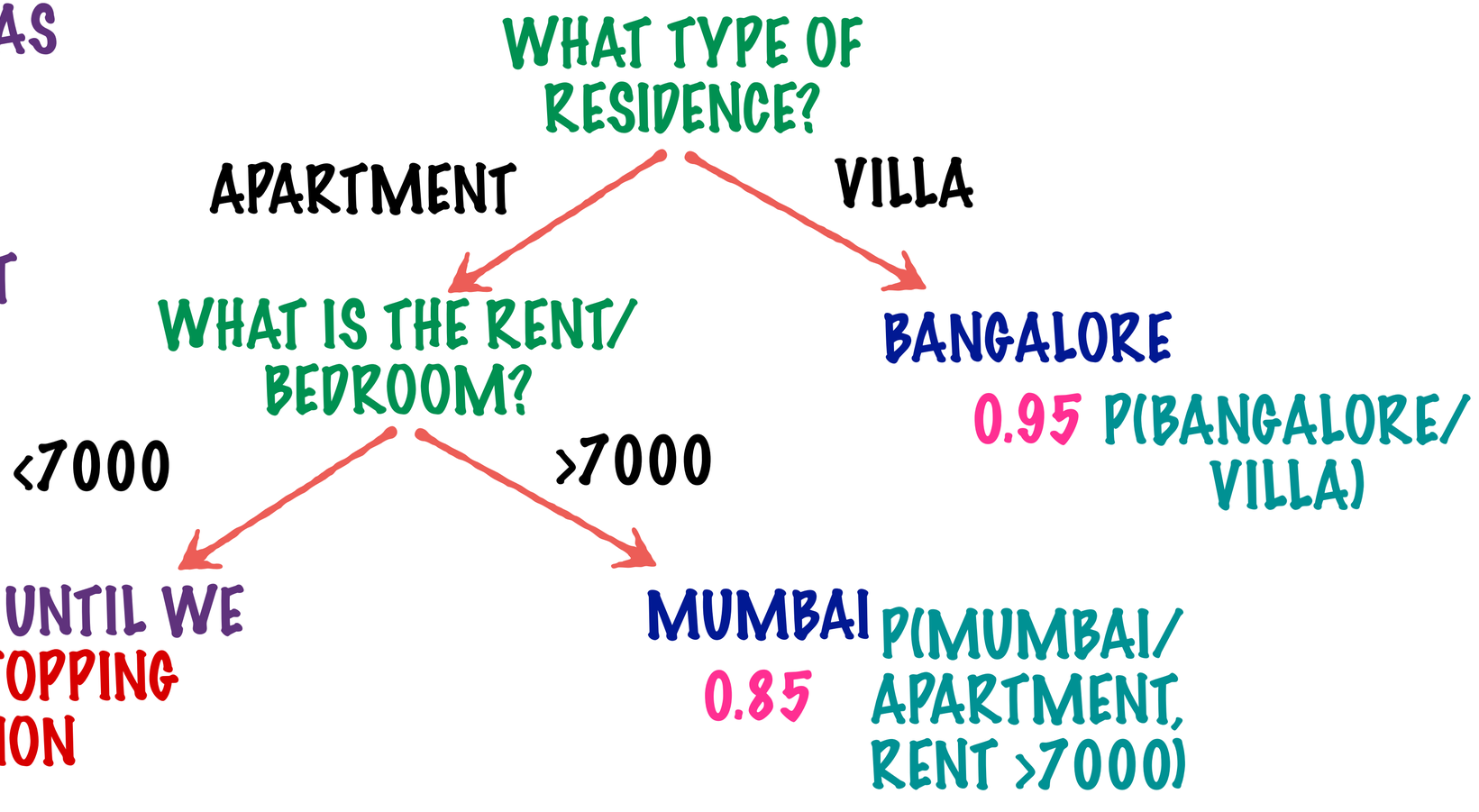
WE'VE DIVIDED OUR DATA INTO 2 SUBSETS - APARTMENTS AND VILLAS

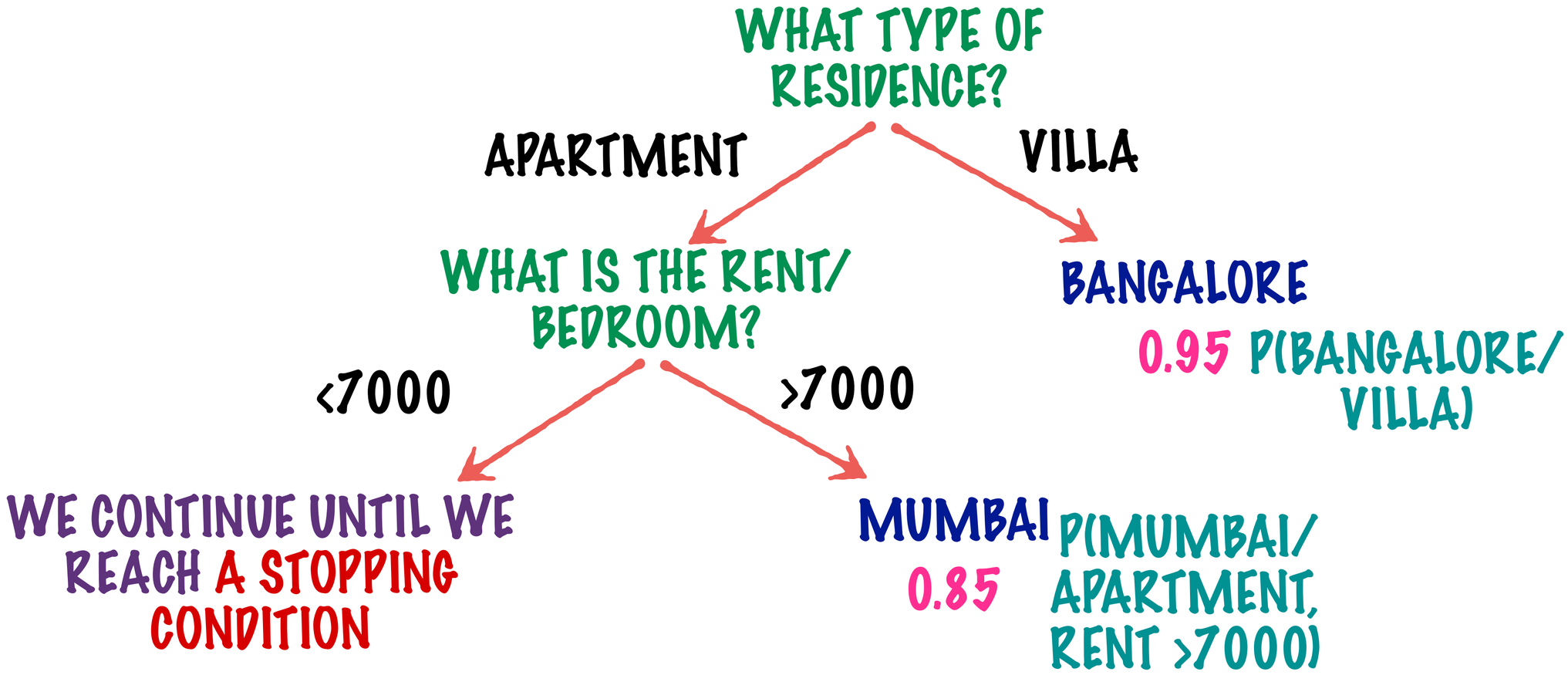
NOW, WE'LL DIVIDE THE APARTMENTS INTO TWO SUBSETS, RENT >7000, RENT <7000

WE CONTINUE UNTIL WE REACH A STOPPING CONDITION

REPEAT THIS PROCESS RECURSIVELY, FOR EACH SUBSET

SO, WE NOW INCLUDE THIS CRITERION IN OUR DECISION TREE



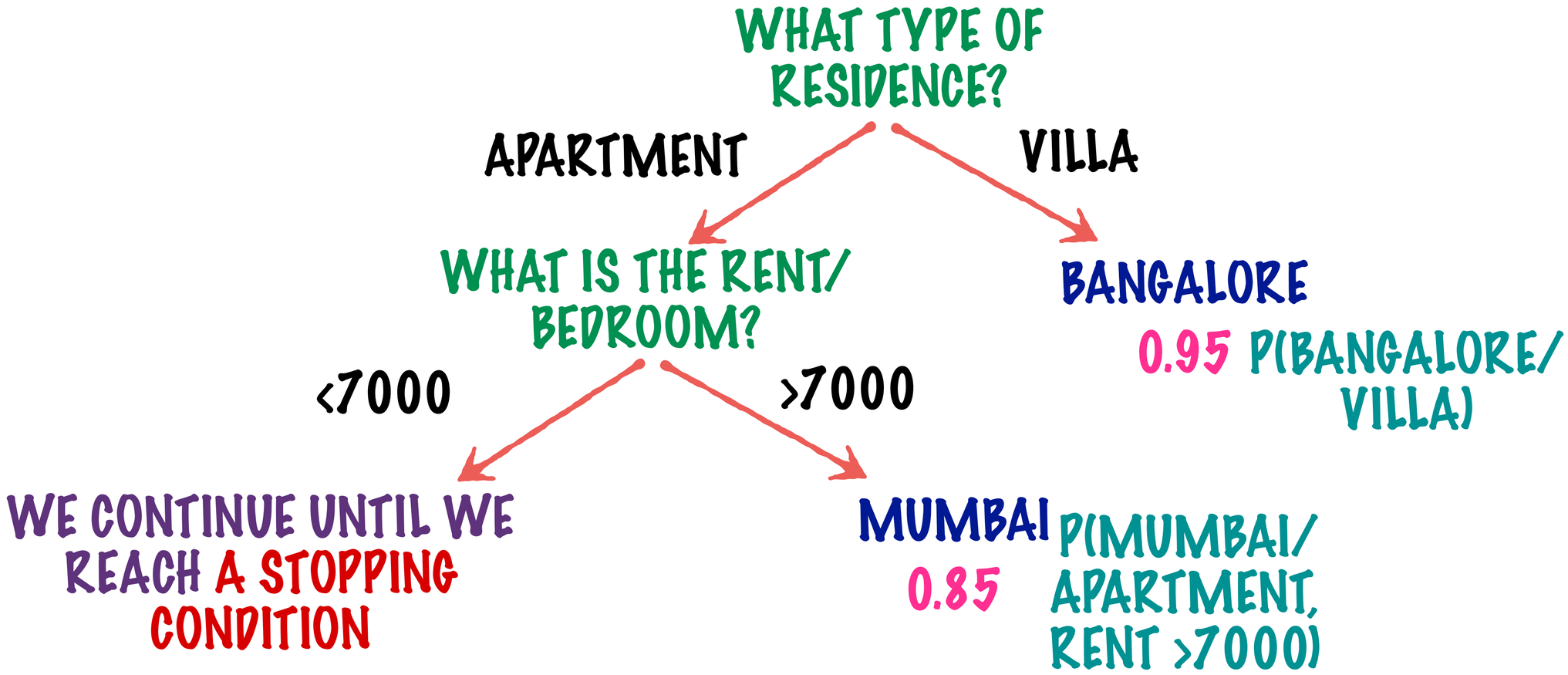


THE **STOPPING CONDITION** COULD BE

ALL OUR SUBSETS ARE MOSTLY HOMOGENOUS

WE RUN OUT OF ATTRIBUTES

THE TREE IS TOO BIG



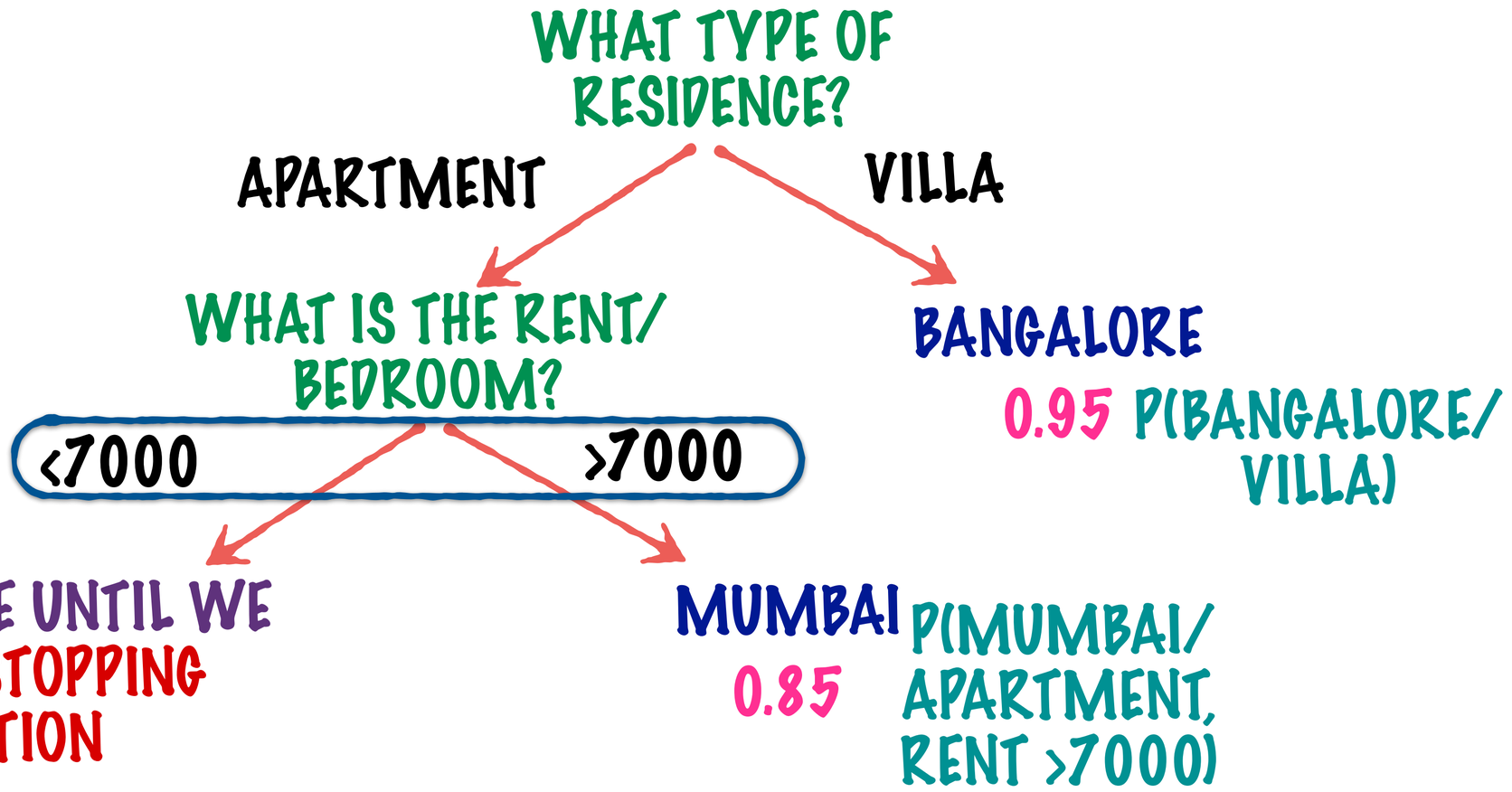
THE PROCESS OF SPLITTING DATA INTO SUBSETS RECURSIVELY IS CALLED **RECURSIVE PARTITIONING**

THE PROCESS OF SPLITTING DATA INTO SUBSETS RECURSIVELY IS CALLED **RECURSIVE PARTITIONING**

BEFORE WE MOVE ON TO SPECIFIC DECISION TREE ALGORITHMS, LET'S TALK ABOUT FINDING **THE BEST SPLIT** FOR A CONTINUOUS INPUT VARIABLE

BEFORE WE MOVE ON TO SPECIFIC DECISION TREE ALGORITHMS,

LET'S TALK ABOUT FINDING **THE BEST SPLIT** FOR A CONTINUOUS INPUT VARIABLE



WE CONTINUE UNTIL WE REACH **A STOPPING CONDITION**

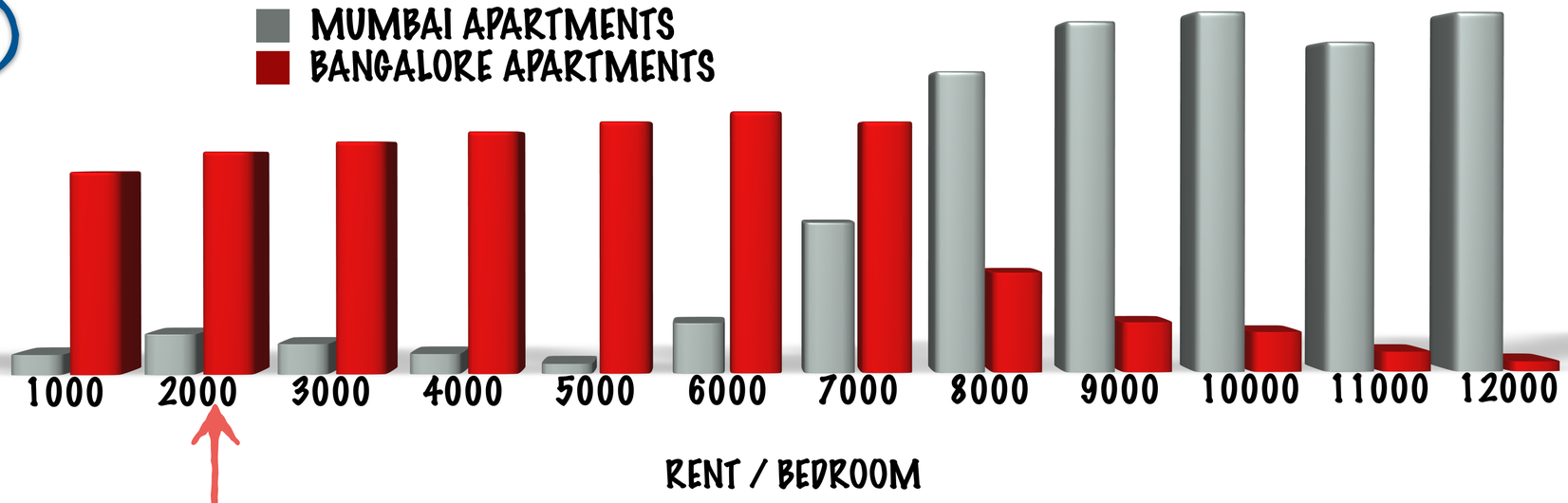
WHEN WE ARE SPLITTING OUR DATA INTO SUBSETS BASED ON A CONTINUOUS VARIABLE - HOW DO WE FIND THE BEST POINT TO SPLIT?

THE BEST SPLIT

<7000 >7000

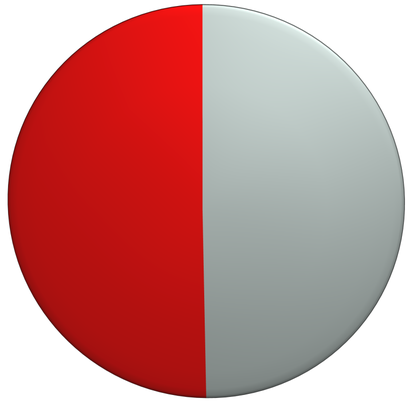
WHEN WE ARE SPLITTING OUR DATA INTO SUBSETS BASED ON A CONTINUOUS VARIABLE - HOW DO WE FIND THE BEST POINT TO SPLIT?

FIRST WE DRAW A HISTOGRAM FOR RENT FOR APARTMENTS IN EACH CITY



<2000

>2000

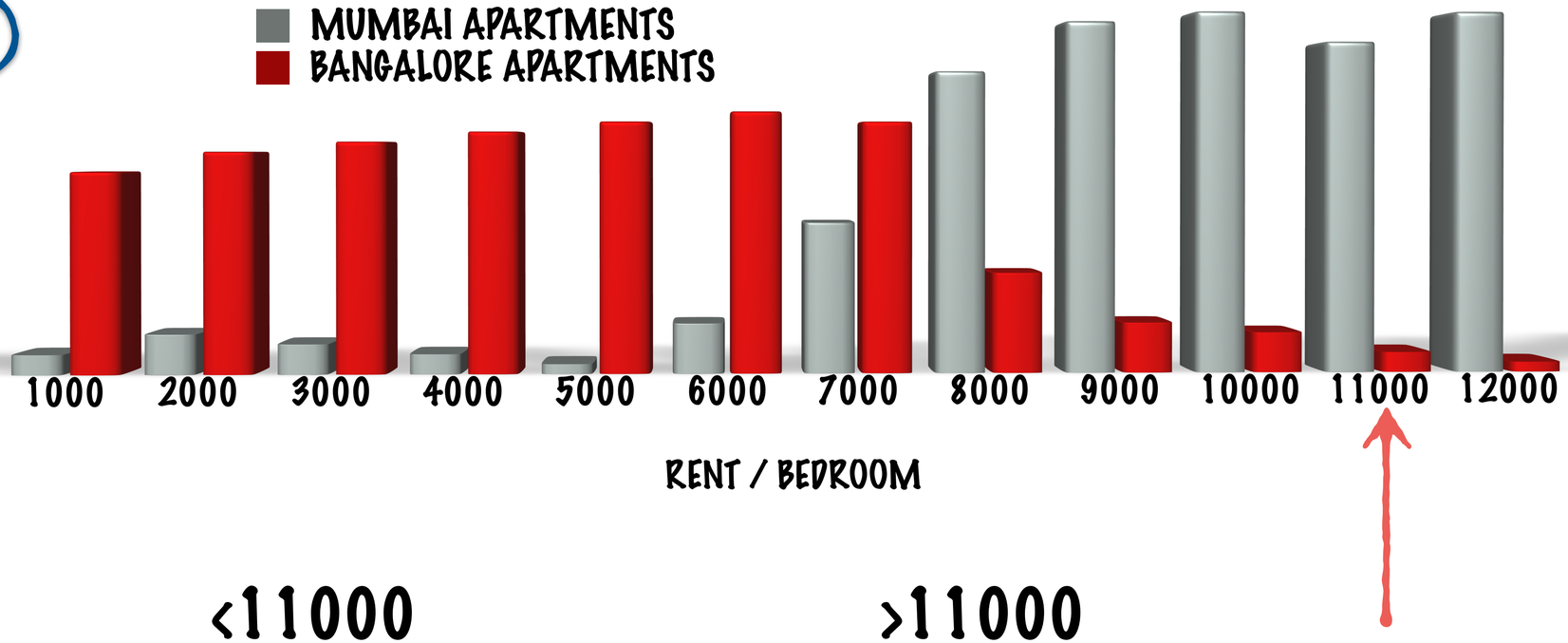


THE BEST SPLIT

<7000 >7000

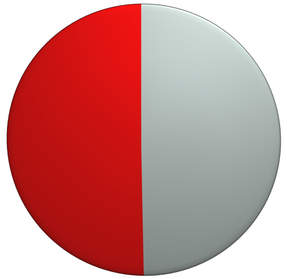
WHEN WE ARE SPLITTING OUR DATA INTO SUBSETS BASED ON A CONTINUOUS VARIABLE - HOW DO WE FIND THE BEST POINT TO SPLIT?

FIRST WE DRAW A HISTOGRAM FOR RENT FOR APARTMENTS IN EACH CITY

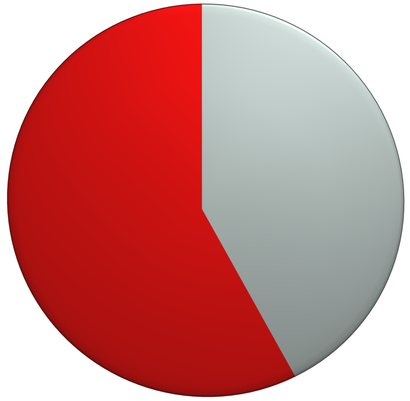


<2000

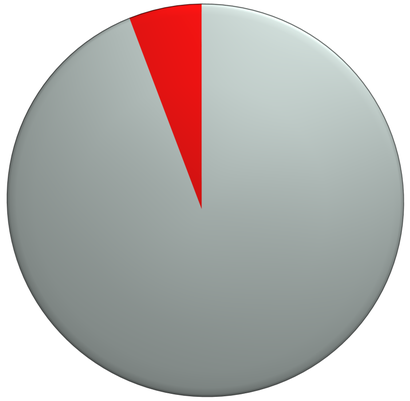
>2000



<11000



>11000

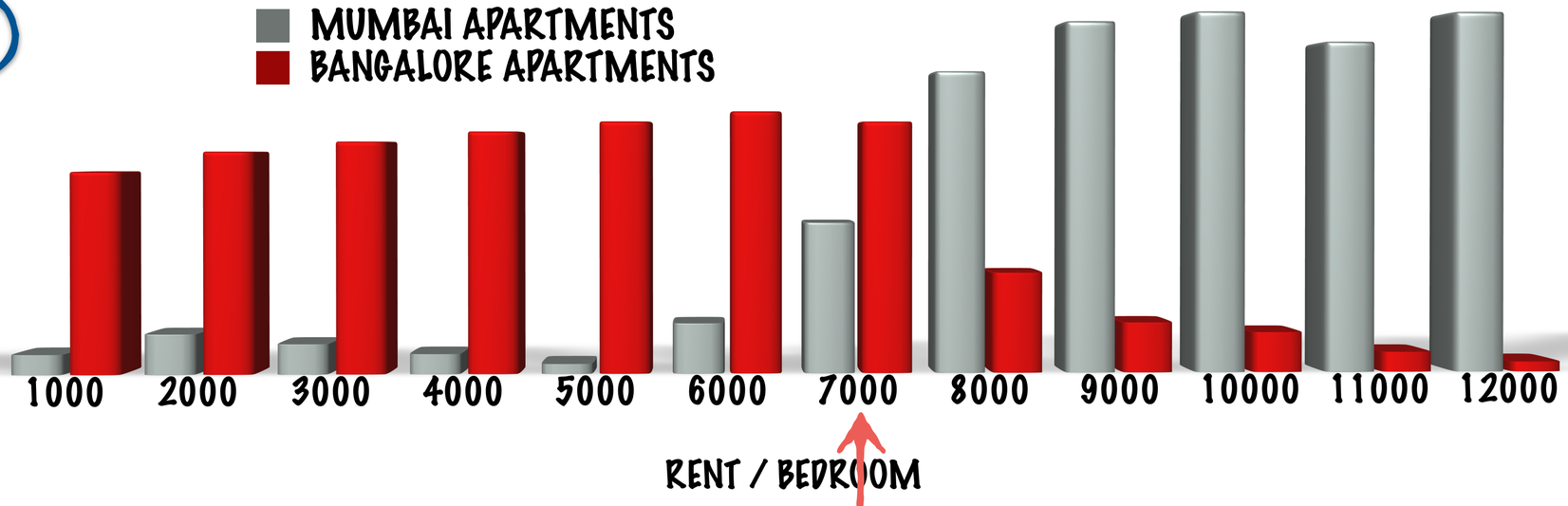


THE BEST SPLIT

<7000 >7000

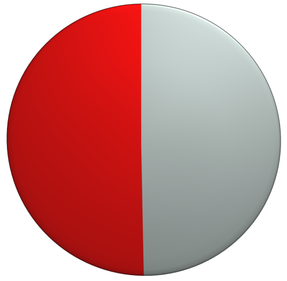
WHEN WE ARE SPLITTING OUR DATA INTO SUBSETS BASED ON A CONTINUOUS VARIABLE - HOW DO WE FIND THE BEST POINT TO SPLIT?

FIRST WE DRAW A HISTOGRAM FOR RENT FOR APARTMENTS IN EACH CITY

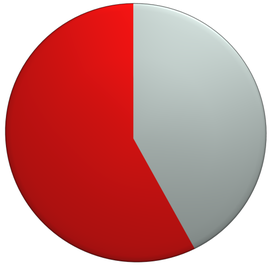


<2000

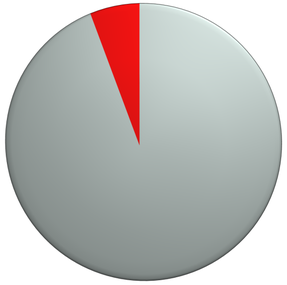
>2000



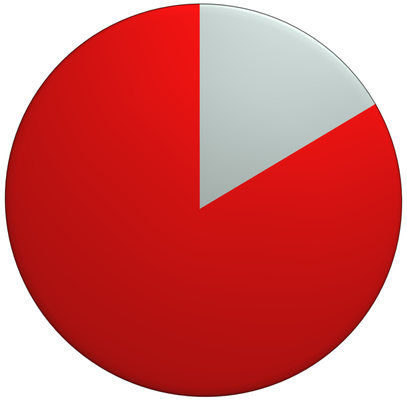
<11000



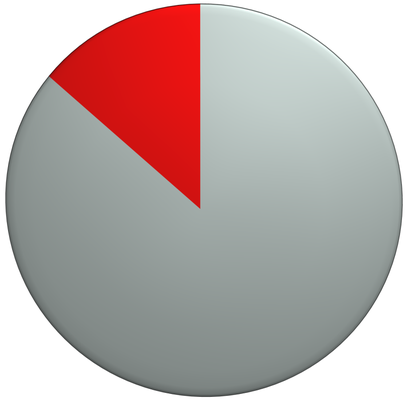
>11000



<7000



>7000



WE CHOOSE THE POINT WHERE BOTH THE SUBSETS WE GET ARE MOSTLY HOMOGENOUS

WE ALREADY MENTIONED THAT -

THE PROCESS OF SPLITTING DATA INTO SUBSETS RECURSIVELY IS CALLED **RECURSIVE PARTITIONING**

THE BASIC IDEA IS THAT AT EACH STEP PARTITIONING IS DONE BASED ON THE VALUE OF 1 ATTRIBUTE

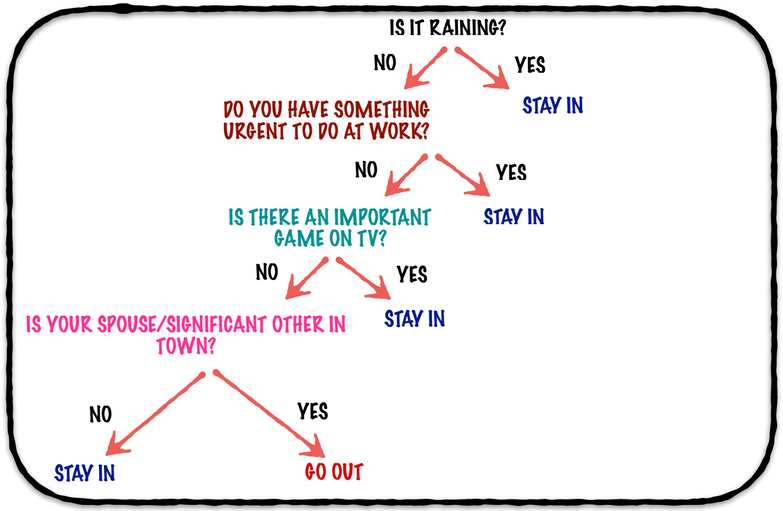
AT EACH STEP , YOU CHOOSE THE 'BEST' ATTRIBUTE

JUST LIKE THE IDEA OF **BEST SPLIT**

THE BEST ATTRIBUTE IS THE ONE WHICH ALLOWS YOU TO CREATE MOSTLY **HOMOGENOUS SUBSETS**

DECISION TREE

INPUT VARIABLES/ PREDICTORS



OUTCOME/OUTPUT VARIABLES

DECISION TREE LEARNING

IS THE PROCESS OF CREATING/LEARNING A DECISION TREE FROM TRAINING DATA

RECURSIVE PARTITIONING

IS THE MOST COMMON STRATEGY FOR DECISION TREE LEARNING

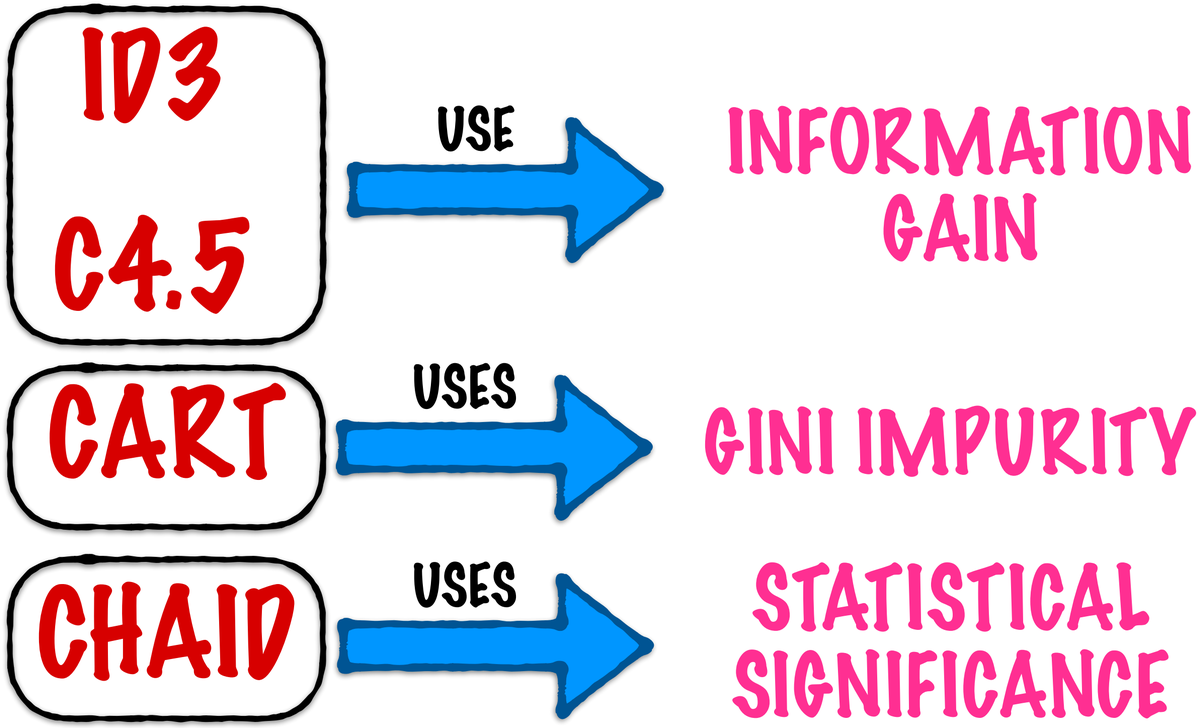
ID3 C4.5

CART CHAID

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

EACH HAS A SLIGHTLY DIFFERENT WAY OF ARRIVING AT THE BEST ATTRIBUTE (OR) MEASURING THE HOMOGENEITY OF A SUBSET



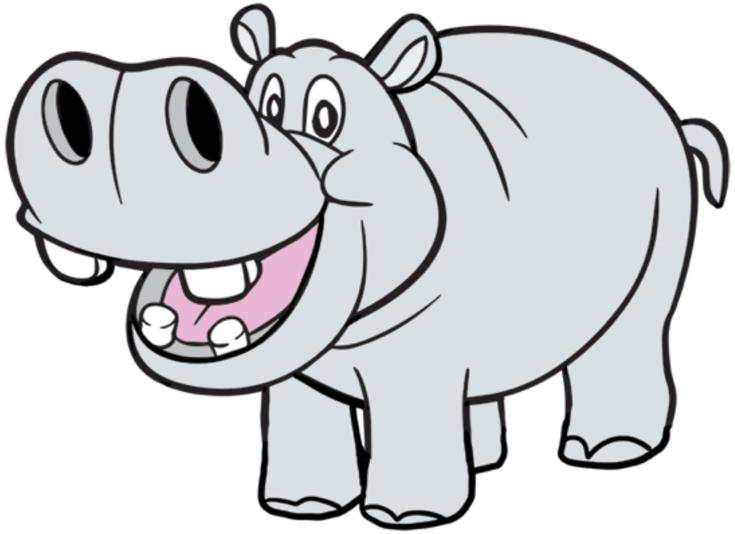
INFORMATION GAIN

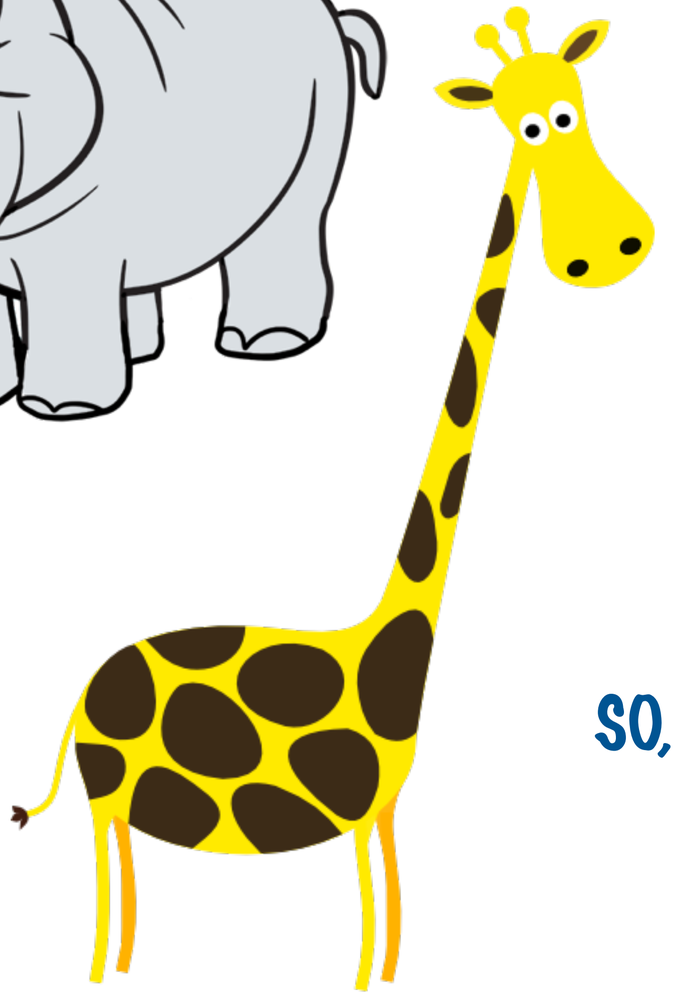
ANY STATEMENT , NEWS OR MESSAGE CONTAINS INFORMATION

SOME HAVE MORE INFORMATION AND SOME LESS

THE IDEA OF INFORMATION GAIN IS TO REDUCE ENTROPY AND MAXIMIZE INFORMATION

LET'S SAY YOU HAVE TO CLASSIFY AN ANIMAL AS A GIRAFFE OR A HIPPO





IF YOU WERE TOLD, THIS ANIMAL HAS 4 LEGS

THIS IS BASICALLY USELESS! BOTH GIRAFFES AND HIPPOS HAVE 4 LEGS, SO THIS STATEMENT GIVES US NO INFORMATION

BUT IF YOU WERE TOLD, THIS ANIMAL IS 10 FEET TALL

THIS IS USEFUL INFORMATION!

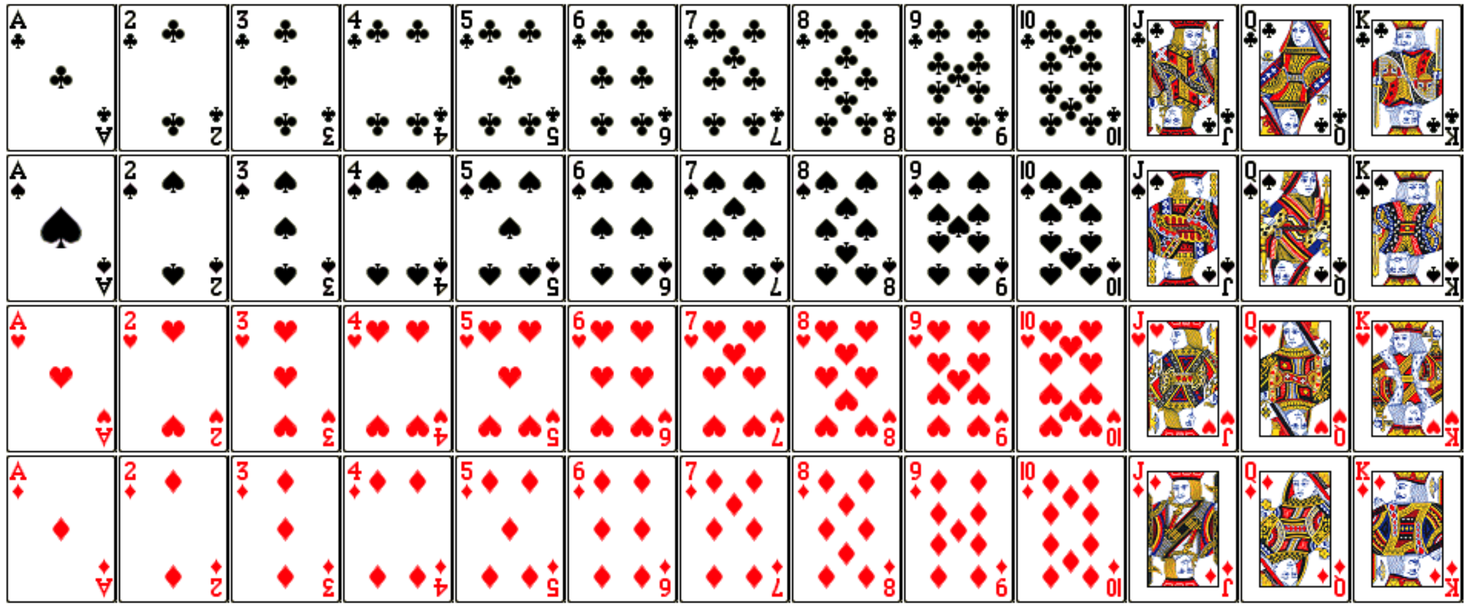
IT TELLS YOU THAT THE ANIMAL IS VERY LIKELY A GIRAFFE

SO, CLEARLY - THE VALUES OF SOME ATTRIBUTES GIVE US MORE INFORMATION THAN OTHERS

AND THERE IS A MATHEMATICAL WAY TO MEASURE THIS INFORMATION

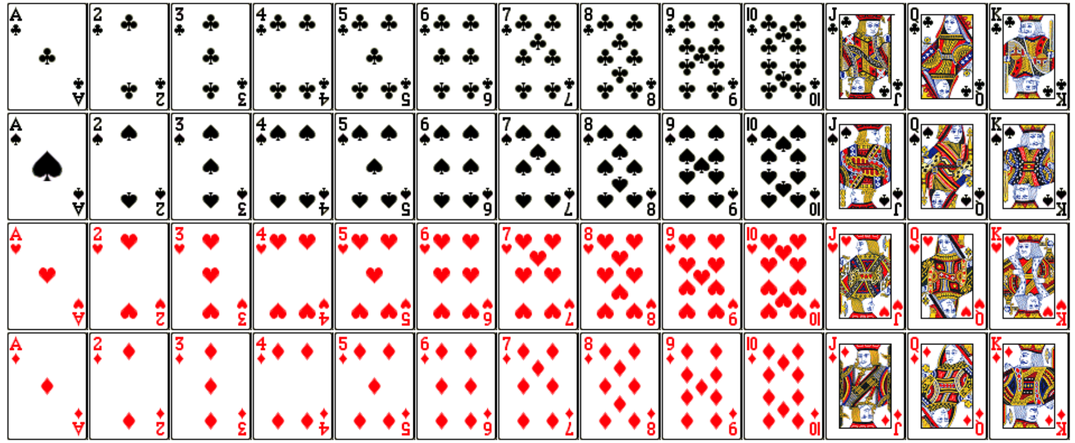
GUESS THE CARD YOUR OPPONENT HOLDS

YOU ARE ALLOWED TO ASK THEM YES/NO QUESTIONS



INITIALLY, THERE ARE 52 POSSIBLE OUTCOMES IN ALL

INITIALLY, THERE ARE 52 POSSIBLE OUTCOMES IN ALL

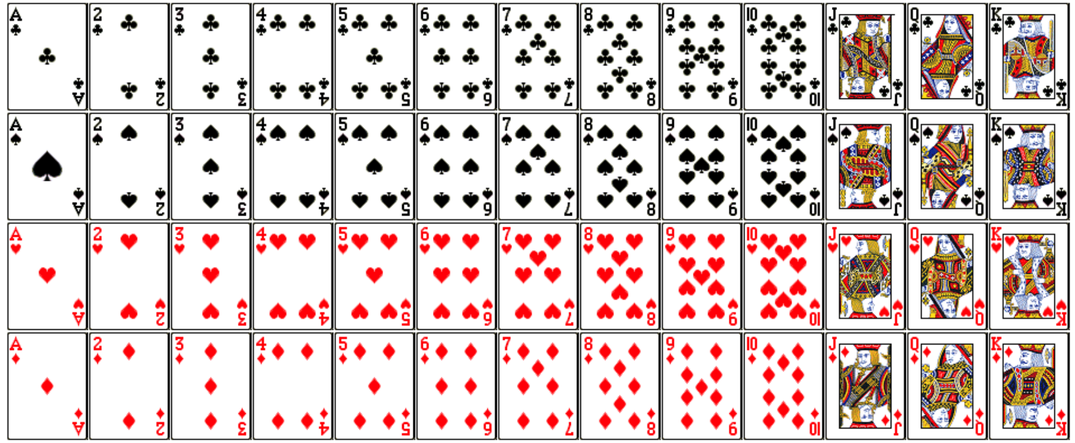


YES

YOU ASK

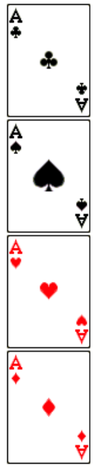
IS THE CARD AN ACE?

INITIALLY, THERE ARE 52 POSSIBLE OUTCOMES IN ALL



YES

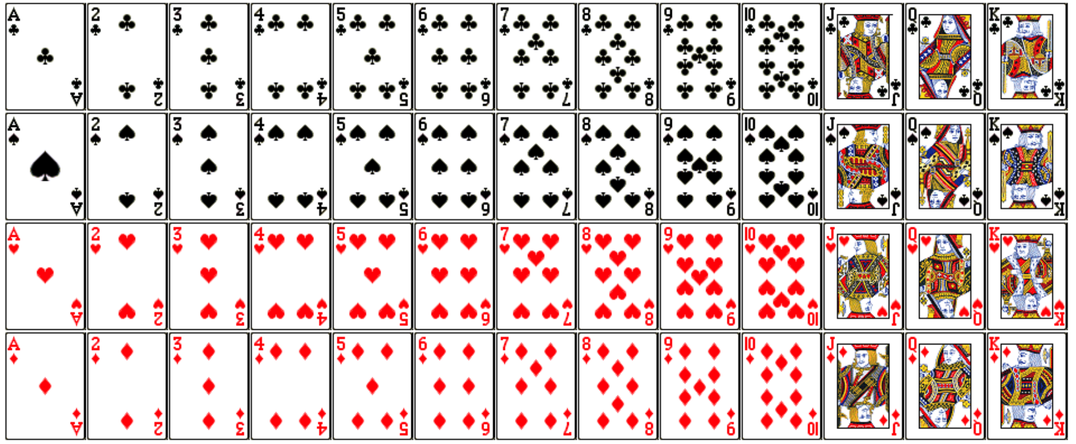
NO



YOU ASK

IS THE CARD AN ACE?

INITIALLY, THERE ARE 52 POSSIBLE OUTCOMES IN ALL



YOU ASK

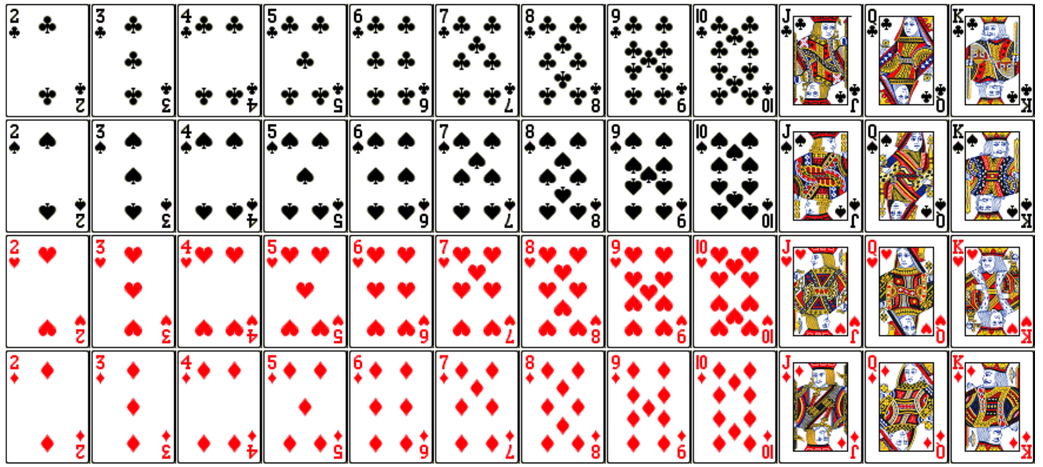
IS THE CARD AN ACE?

YES

NO

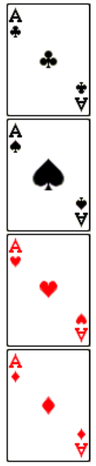
LEFT WITH 4 POSSIBLE OUTCOMES

LEFT WITH 48 POSSIBLE OUTCOMES

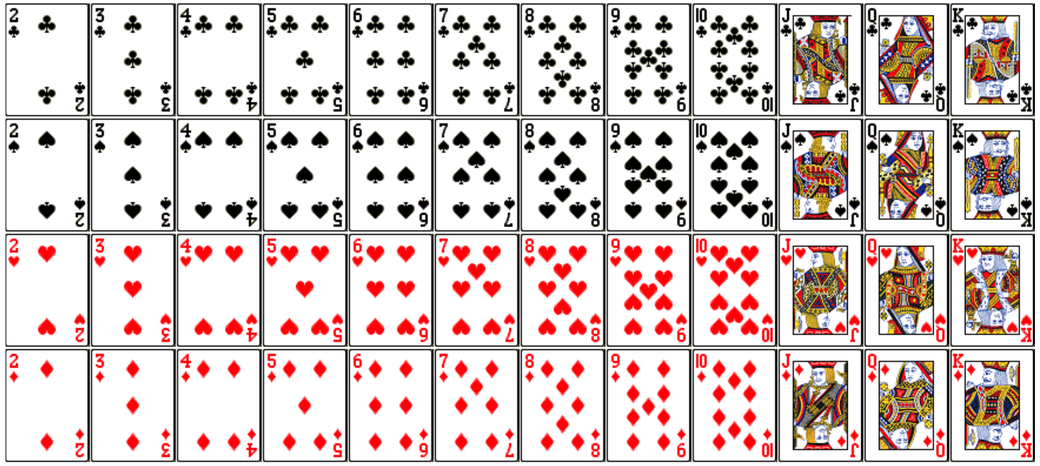


THE ANSWER "YES" GIVES US MORE INFORMATION THAN THE ANSWER "NO"

YES



NO



P(YES)=4/52

P(NO)=48/52

IS THE CARD AN ACE?

THE ANSWER "YES" GIVES US **MORE** **INFORMATION** THAN THE ANSWER "NO"

THE ANSWER "YES" HAS A LOWER PROBABILITY

THE **LOWER THE PROBABILITY** OF THE ANSWER, THE **MORE** **INFORMATION** YOU GET

IF X IS A RANDOM VARIABLE THAT REPRESENTS THE ANSWER TO OUR QUESTION

INFORMATION CONTENT OF (X=YES) = -LOG(P(X=YES))

INFORMATION CONTENT OF (X=NO) = -LOG(P(X=NO))

INFORMATION CONTENT OF (X=x) = -LOG(P(X=x))

INFORMATION GAIN

ANY STATEMENT , NEWS OR MESSAGE CONTAINS INFORMATION

SOME HAVE MORE INFORMATION AND SOME LESS

THE IDEA OF INFORMATION GAIN IS TO REDUCE ENTROPY AND MAXIMIZE INFORMATION

IF YOU WERE TO GUESS, WHAT NEWS YOU'LL HEAR, BEFORE IT HAPPENS

SOME GUESSES ARE EASY

SOME GUESSES ARE HARD

WHAT TIME WILL THE SUN RISE?

WHAT WILL BE THE RESULT OF A COIN TOSS?

IF X IS A RANDOM VARIABLE THAT REPRESENTS THE ANSWER TO OUR QUESTION

INFORMATION CONTENT OF (X=x) = $-\log(P(X=x))$

AVERAGE VALUE OF THE INFORMATION CONTENT (ALSO CALLED THE EXPECTED VALUE)=

$$\sum P(X=x)\,(-\log(P(X=x)))$$

ENTROPY H(X)

ENTROPY IS THE AMOUNT OF UNCERTAINTY/UNPREDICTABILITY THERE IS IN THE ANSWER

ENTROPY INCREASES WITH

1) NUMBER OF POSSIBLE ANSWERS
2) THE EVENNESS OF THE PROBABILITY DISTRIBUTION

P(YES) = 0 => THERE IS NO UNCERTAINTY => ENTROPY =0

YES AND NO HAVE EQUAL PROBABILITY => VERY HIGH ENTROPY

IF X IS A RANDOM VARIABLE THAT REPRESENTS THE ANSWER TO OUR QUESTION

INFORMATION CONTENT OF (X=x) = -LOG(P(X=x))

ENTROPY H(X) = AVERAGE VALUE OF THE INFORMATION CONTENT (ALSO CALLED THE EXPECTED VALUE)= $\Sigma P(X=x)\,(-\log(P(X=x))$

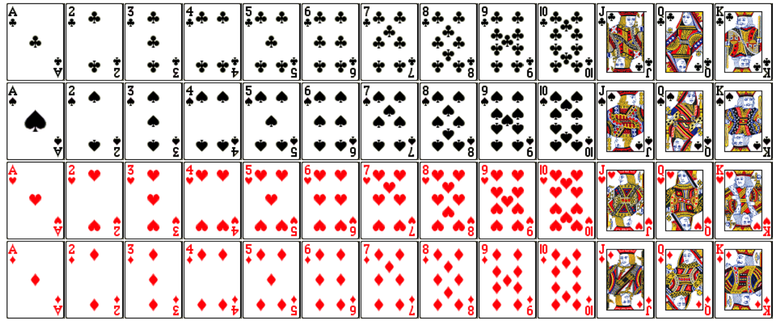
THE GAME IS TO ANSWER THE QUESTION

WHICH CARD DO YOU HOLD?

BEFORE WE HAVE ASKED ANY YES/NO QUESTIONS, THE UNCERTAINTY (ENTROPY) IN OUR GUESS IS VERY HIGH

INITIALLY, THERE ARE 52 POSSIBLE OUTCOMES IN ALL

EACH HAS SAME PROBABILITY =1/52



ENTROPY=H(X)=

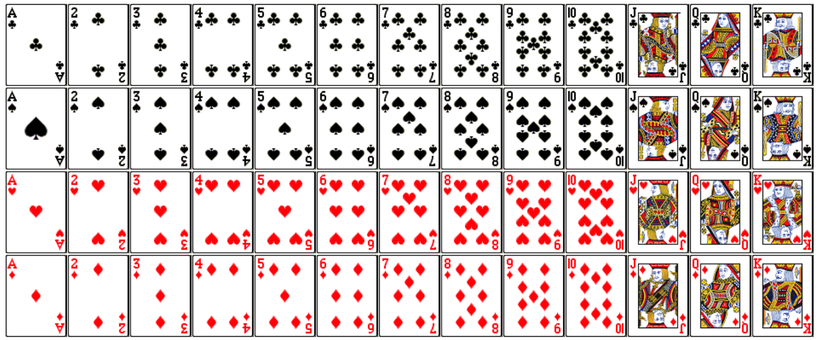
$\Sigma(1/52)(-\log(1/52)=$

LOG(52)

BEFORE WE HAVE ASKED ANY YES/NO QUESTIONS, THE UNCERTAINTY (ENTROPY) IN OUR GUESS IS VERY HIGH

THERE ARE 52 POSSIBLE OUTCOMES IN ALL

EACH HAS PROBABILITY =1/52



ENTROPY=H(X)= LOG(52)

ONCE YOU ASK

IS THE CARD AN ACE?

YES

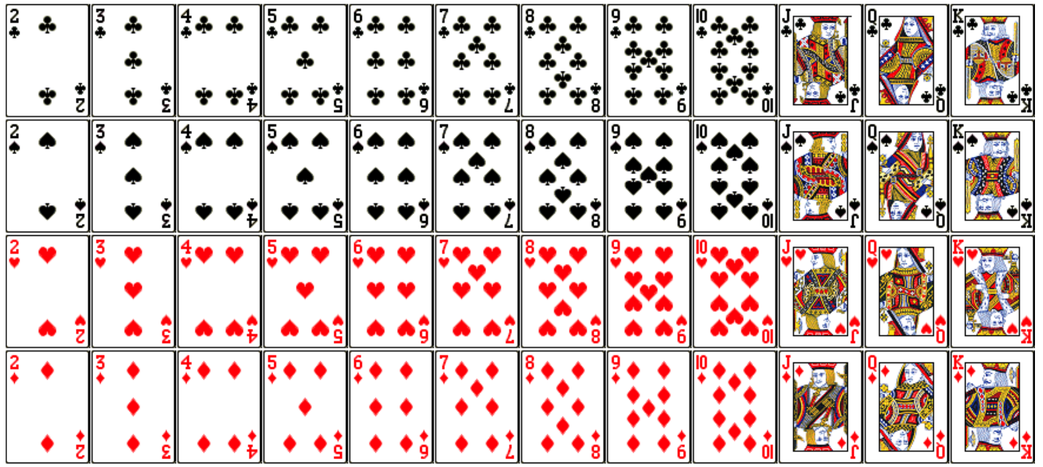
NO

ENTROPY = H(X/Q1 =YES)= LOG(4)

P(YES)=4/52

ENTROPY = H(X/Q1 =NO)= LOG(48)

P(NO)=48/52



WITHIN EACH GROUP, THE ENTROPY HAS DECREASED

THE MORE HOMOGENOUS EACH GROUP IS, THE LOWER THE ENTROPY

BEFORE WE HAVE ASKED ANY YES/NO QUESTIONS, THE UNCERTAINTY (ENTROPY) IN OUR GUESS IS VERY HIGH

ENTROPY = H(X)=LOG(52)

ONCE YOU ASK IS THE CARD AN ACE?

ENTROPY = H(X/Q1 =YES)= LOG(4)

P(YES)=4/52

ENTROPY = H(X/Q1 =NO)= LOG(48)

P(NO)=48/52

ENTROPY AFTER Q1 = H(X/Q1) =

P(YES)*H(X/Q1 =YES)+P(NO)*H(X/Q1 =NO) =

4/52*LOG(4) + 48/52 * LOG(48)

BEFORE WE HAVE ASKED ANY YES/NO QUESTIONS, THE UNCERTAINTY (ENTROPY) IN OUR GUESS IS VERY HIGH

ENTROPY = H(X)

ONCE YOU ASK IS THE CARD AN ACE?

ENTROPY AFTER Q1 = H(X/Q1)

INFORMATION GAIN = REDUCTION IN ENTROPY OVERALL = H(X) - H(X/Q1)

AS YOU SAW, WHENEVER YOU ASK A QUESTION, SUBSETS ARE FORMED

WHEN EACH OF THOSE SUBSETS ARE HOMOGENOUS, THE INFORMATION GAIN IS MAXIMUM

ENTROPY = H(X)

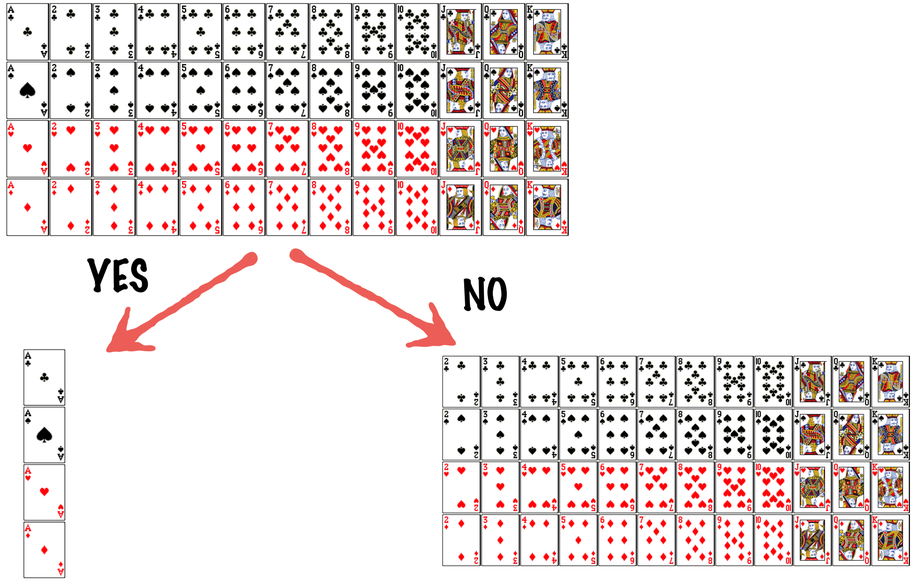
ENTROPY AFTER Q1 = H(X/Q1)

INFORMATION GAIN = REDUCTION IN ENTROPY = H(X) - H(X/Q1)
OVERALL

AS YOU SAW, WHENEVER
YOU ASK A QUESTION,
SUBSETS ARE FORMED

WHEN EACH OF THOSE SUBSETS
ARE HOMOGENOUS, THE
INFORMATION GAIN IS MAXIMUM

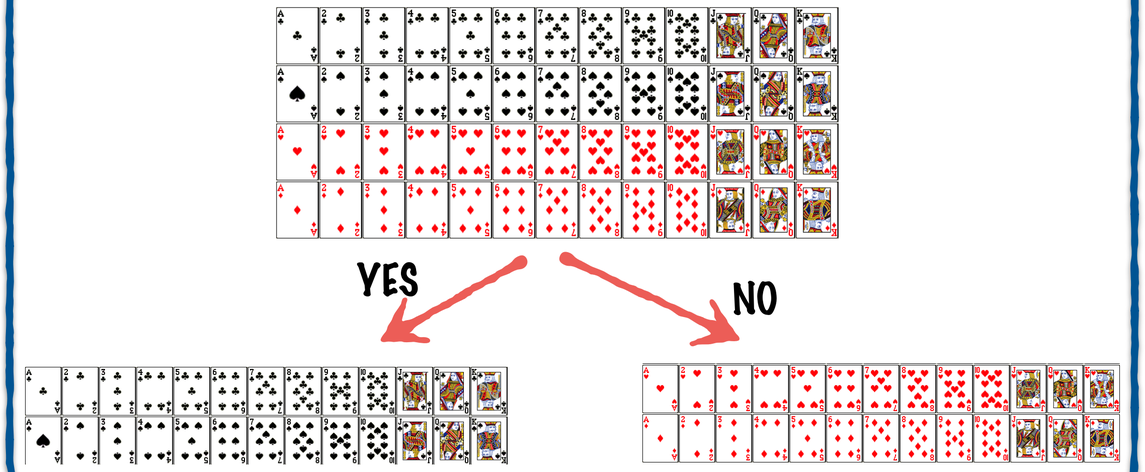
IS IT AN ACE?



H(X/Q1) = 4/52*LOG(4) + 48/52 * LOG(48)

IG= H(X)-H(X/Q1) = 0.12

IS IT A BLACK?



H(X/Q1) = LOG(26)

IG= H(X)-H(X/Q1) = LOG(52)-LOG(26) =0.30

INFORMATION GAIN

ANY STATEMENT , NEWS OR MESSAGE CONTAINS INFORMATION

SOME HAVE MORE INFORMATION AND SOME LESS

THE IDEA OF INFORMATION GAIN IS TO REDUCE ENTROPY AND MAXIMIZE INFORMATION

IF YOU WERE TO GUESS, WHAT NEWS YOU'LL HEAR, BEFORE IT HAPPENS

SOME GUESSES ARE EASY

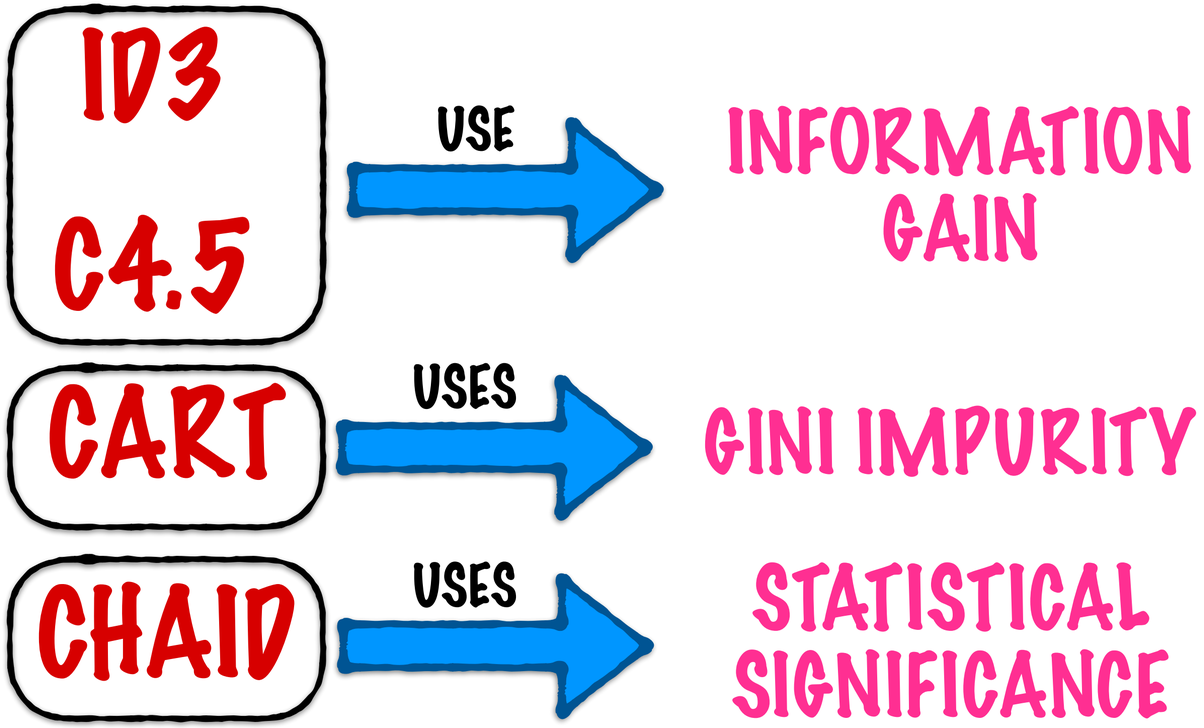
SOME GUESSES ARE HARD

WHAT TIME WILL THE SUN RISE?

WHAT WILL BE THE RESULT OF A COIN TOSS?

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

EACH HAS A SLIGHTLY DIFFERENT WAY OF ARRIVING AT THE BEST ATTRIBUTE (OR) MEASURING THE HOMOGENEITY OF A SUBSET



ID3

C4.5

ARE DECISION TREE LEARNING METHODS

BOTH ID3 AND C4.5 WORK BY MAXIMIZING INFORMATION GAIN AT EACH STEP

COMPUTE THE INFORMATION GAIN FOR EACH ATTRIBUTE AND CHOOSE THE ATTRIBUTE WITH MAX INFORMATION GAIN AS THE DECISION TREE NODE

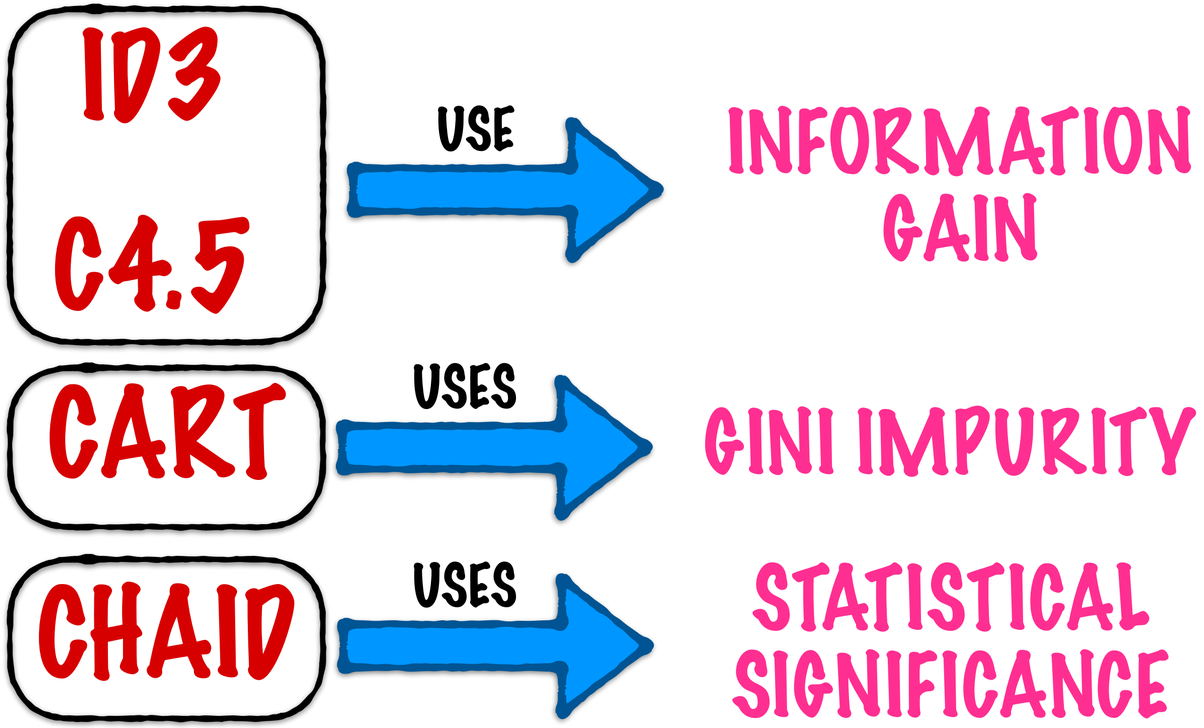
A NOTE : INFORMATION GAIN BASED METHODS ARE BIASED TO CHOOSE ATTRIBUTES WITH MORE LEVELS

- MORE LEVELS INHERENTLY IMPLIES MORE ENTROPY AND HENCE KNOWING THEIR VALUE GIVES US MORE INFORMATION

C4.5 IMPROVES UPON ID3 - IT SUPPORTS CONTINUOUS VARIABLES AS WELL AS CATEGORICAL VARIABLES

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

EACH HAS A SLIGHTLY DIFFERENT WAY OF ARRIVING AT THE BEST ATTRIBUTE (OR) MEASURING THE HOMOGENEITY OF A SUBSET



CART IS ANOTHER DECISION TREE LEARNING METHOD

(CLASSIFICATION AND REGRESSION TREES)

IT USES A DIFFERENT WAY TO CHOOSE AN ATTRIBUTE

MINIMIZING GINI IMPURITY

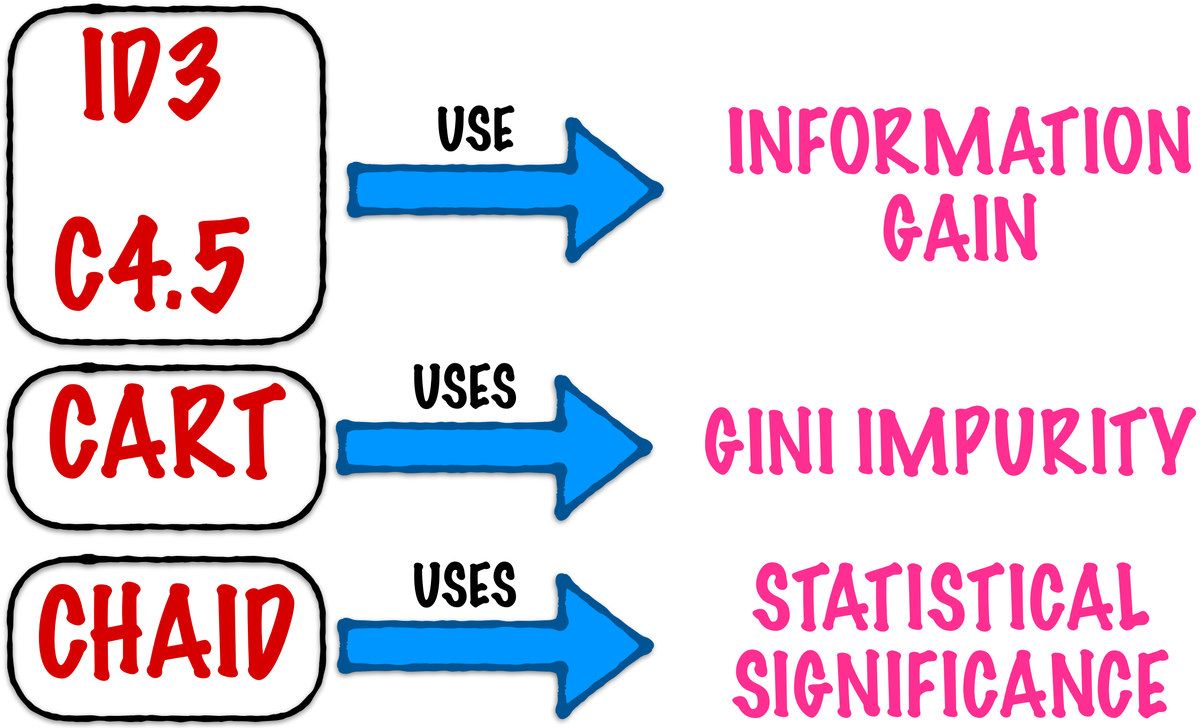
THE IDEA BEHIND GINI IMPURITY IS SIMPLE

CHOOSE THE ATTRIBUTE SUCH THAT - IF YOU STOP THE DECISION TREE WITH THAT ATTRIBUTE AND GO NO FURTHER

THE PROBABILITY OF A FALSE LABEL IS MINIMIZED

DECISION TREE LEARNING ALGORITHMS BASED ON RECURSIVE PARTITIONING

EACH HAS A SLIGHTLY DIFFERENT WAY OF ARRIVING AT THE BEST ATTRIBUTE (OR) MEASURING THE HOMOGENEITY OF A SUBSET



(CHI-SQUARED AUTOMATIC INTERACTION DETECTOR) **CHAID** IS SLIGHTLY DIFFERENT FROM THE OTHER METHODS WE'VE SEEN

CHAID CHECKS WHETHER THE ATTRIBUTES/VARIABLES WE ARE USING ARE CORRELATED

IF THEY ARE CORRELATED, IT MERGES THEM TO CREATE ONE VARIABLE

IT ALSO PERFORMS STATISTICAL SIGNIFICANCE TESTS BEFORE SPLITTING THE DATA INTO SUBSETS

ONE OF THE RISKS INHERENT WITH DECISION TREES

RISK OF OVERFITTING

YOU END UP WITH A TREE THAT IS TOO LARGE

IT IS TOO SPECIFIC TO THE TRAINING DATA AND WHEN A NEW INSTANCE COMES IN - IT DOESN'T PERFORM WELL

PRUNING

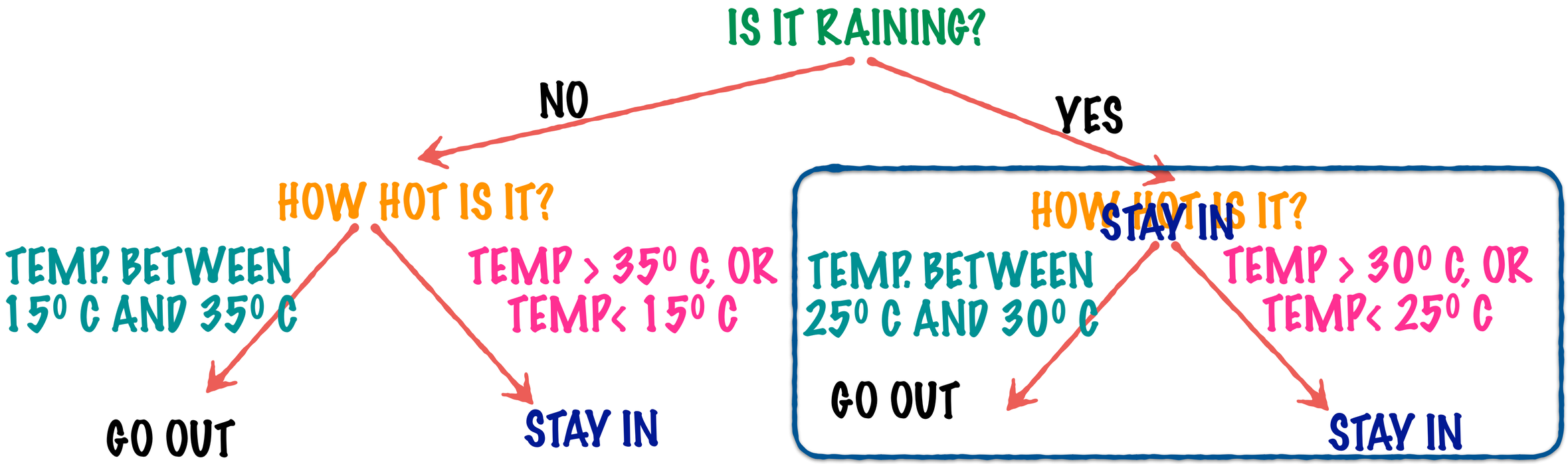
IS THE MOST POPULAR WAY TO AVOID THIS PROBLEM

THIS INVOLVES REMOVING SOME OF THE NODES OF YOUR DECISION TREE AND REPLACING THEM WITH A LEAF

PRUNING

IS THE MOST POPULAR WAY TO AVOID THIS PROBLEM

THIS INVOLVES REMOVING SOME OF THE NODES OF YOUR DECISION TREE AND REPLACING THEM WITH A LEAF



PRUNING

IS THE MOST POPULAR WAY TO AVOID THIS PROBLEM

THIS INVOLVES REMOVING SOME OF THE NODES OF YOUR DECISION TREE AND REPLACING THEM WITH A LEAF

IN GENERAL PRUNING IS PERFORMED BY REMOVING A NODE/SUB-TREE AND CHECKING WHETHER THE ACCURACY OF PREDICTION IS AFFECTED

IF THE ACCURACY IF NOT AFFECTED, THAT NODE/SUB-TREE IS PRUNED