RANDOM FOREST (CLASSIFICATION)

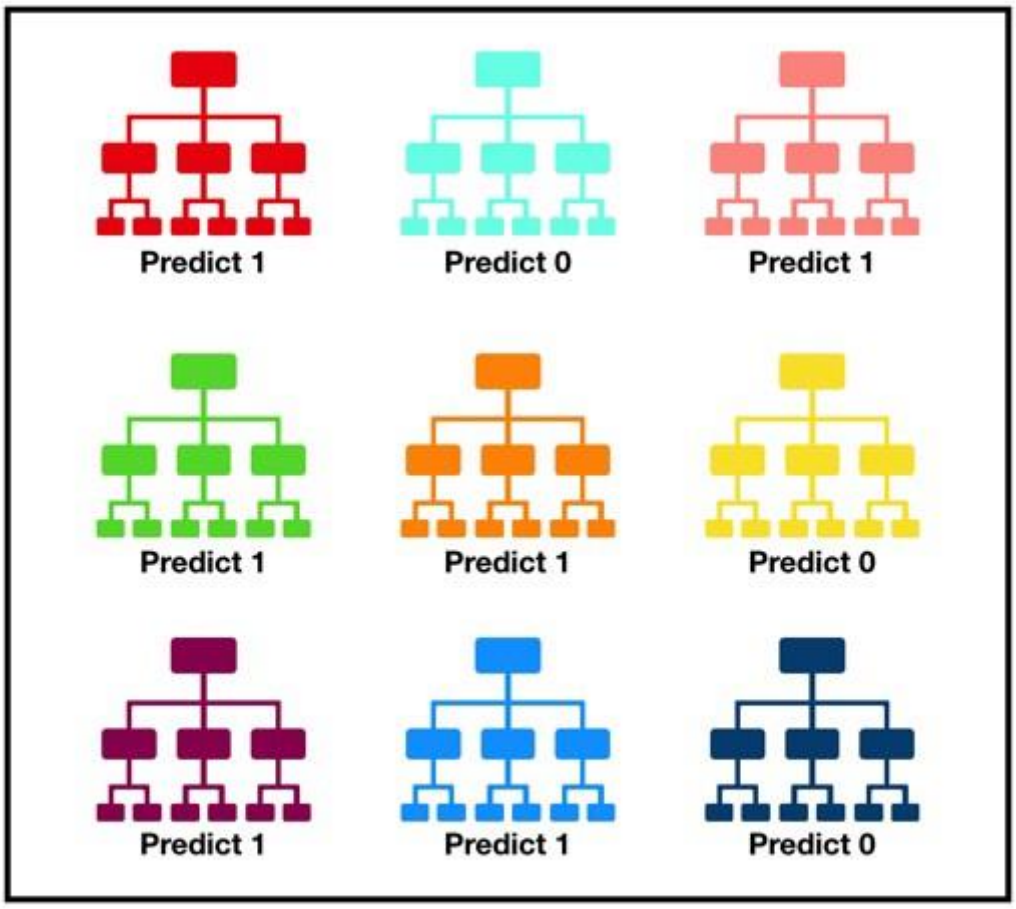
USING WEKA

DR. ALVIN ANG

CONTENTS

I. HOW DO RANDOM FOREST WORK?

- Random Forest = an Ensemble of "Trees" (Decision Trees).
- https://www.alvinang.sg/s/Decision-Tree-Classification-using-WEKA-by-Dr-Alvin-Ang.pdf → we've already learnt how Decision Trees are formed.
- Picture below shows an example of a "Random Forest"…. Made up of multiple Decision Trees.
- Ultimately, the Prediction is made from "Aggregating" all the Trees together.

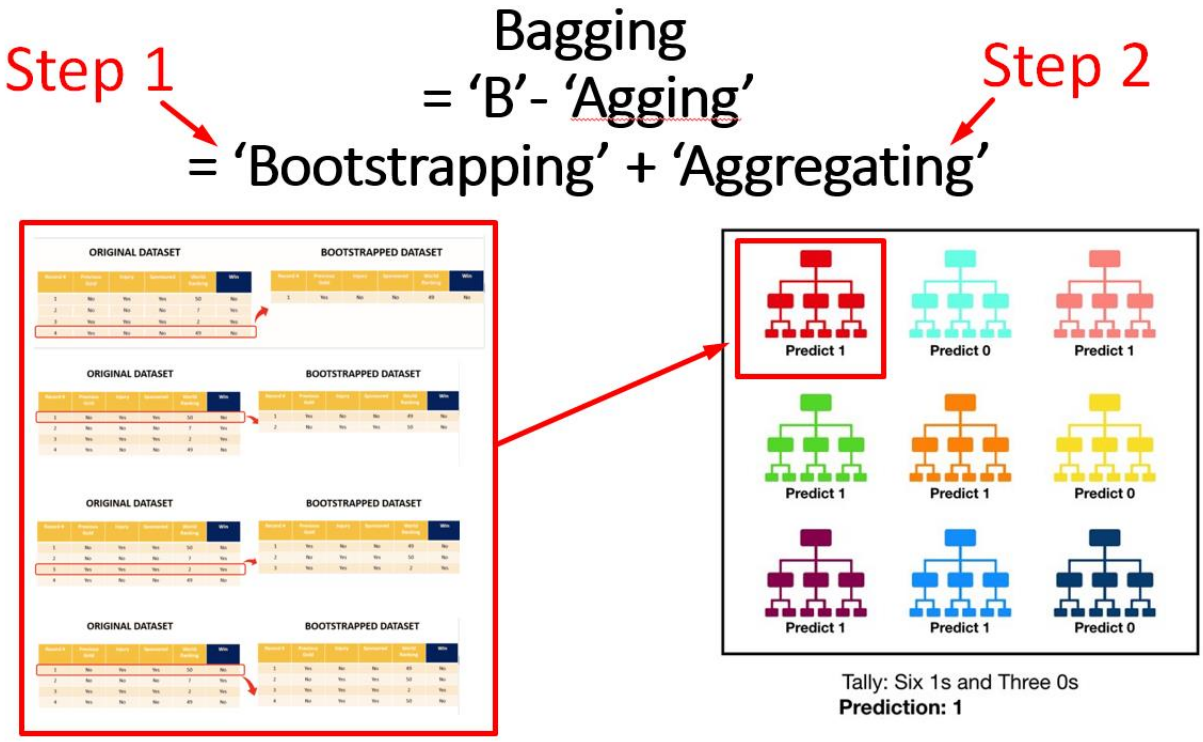


Tally: Six 1s and Three 0s
Prediction: 1

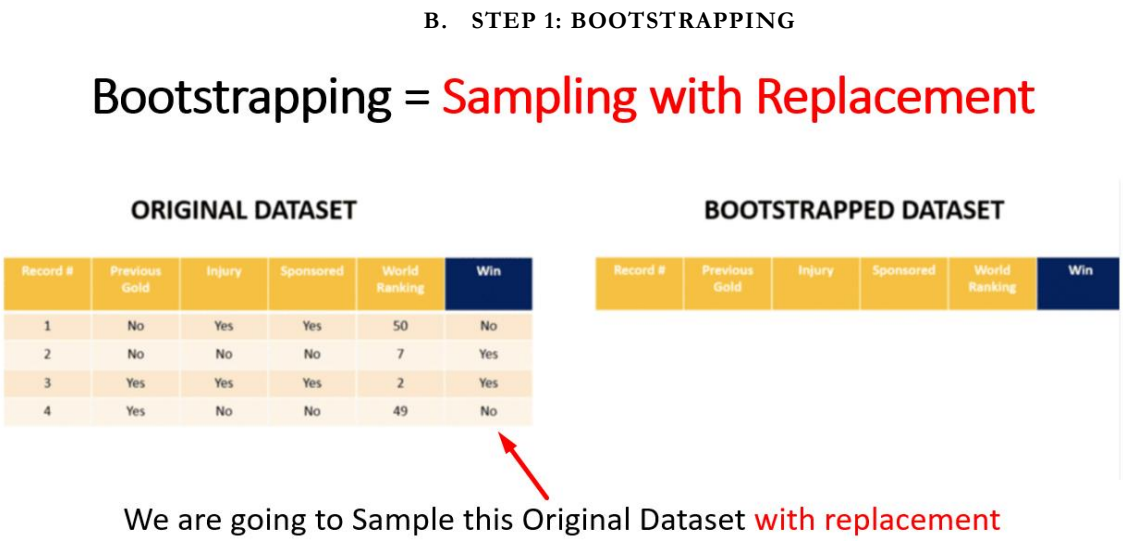
II. HOW ARE RANDOM FORESTS CREATED?

A. WHAT IS BAGGING?

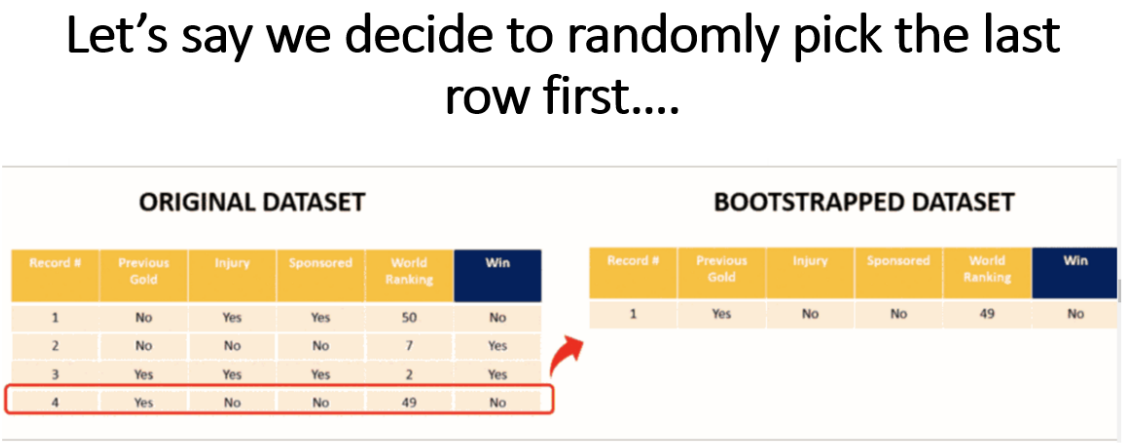
- Random Forests are created through a process called Bagging…



B. STEP 1: BOOTSTRAPPING



- Bootstrapping (I believe) means the normal way of "strapping your boots" with your shoelaces.
- You sequentially take out one row of data (from the left side) and insert / strap it to the right side.
- But this could be done 'randomly' as opposed to an ordered / sequential manner.



Subsequently, we randomly choose the 1st row to "bootstrap" to the 2nd sample

ORIGINAL DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	No	Yes	Yes	50	No
2	No	No	No	7	Yes
3	Yes	Yes	Yes	2	Yes
4	Yes	No	No	49	No

BOOTSTRAPPED DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	Yes	No	No	49	No
2	No	Yes	Yes	50	No

Next, we randomly choose the 3rd row to "bootstrap" to the 3rd sample

ORIGINAL DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	No	Yes	Yes	50	No
2	No	No	No	7	Yes
3	Yes	Yes	Yes	2	Yes
4	Yes	No	No	49	No

BOOTSTRAPPED DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	Yes	No	No	49	No
2	No	Yes	Yes	50	No
3	Yes	Yes	Yes	2	Yes

Somehow, lastly, we randomly still choose back the 1st row to "bootstrap" to the 4th sample

ORIGINAL DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	No	Yes	Yes	50	No
2	No	No	No	7	Yes
3	Yes	Yes	Yes	2	Yes
4	Yes	No	No	49	No

BOOTSTRAPPED DATASET

Record #	Previous Gold	Injury	Sponsored	World Ranking	Win
1	Yes	No	No	49	No
2	No	Yes	Yes	50	No
3	Yes	Yes	Yes	2	Yes
4	No	Yes	Yes	50	No

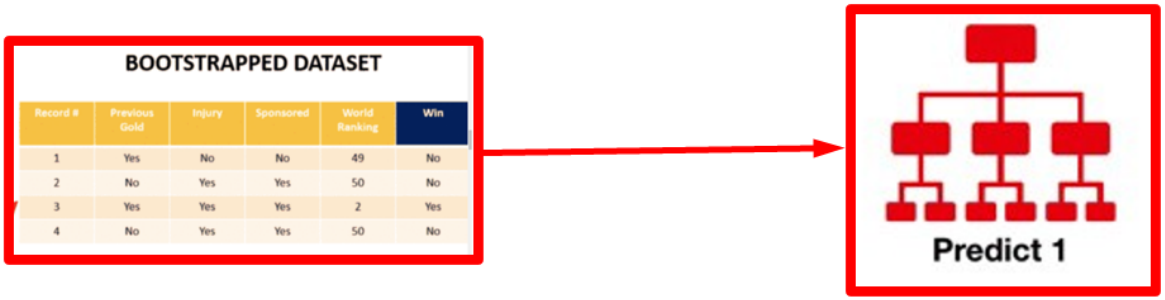
Notice that we didn't touch the 2nd row at all!

This is possible in Bootstrapping since its purely random!

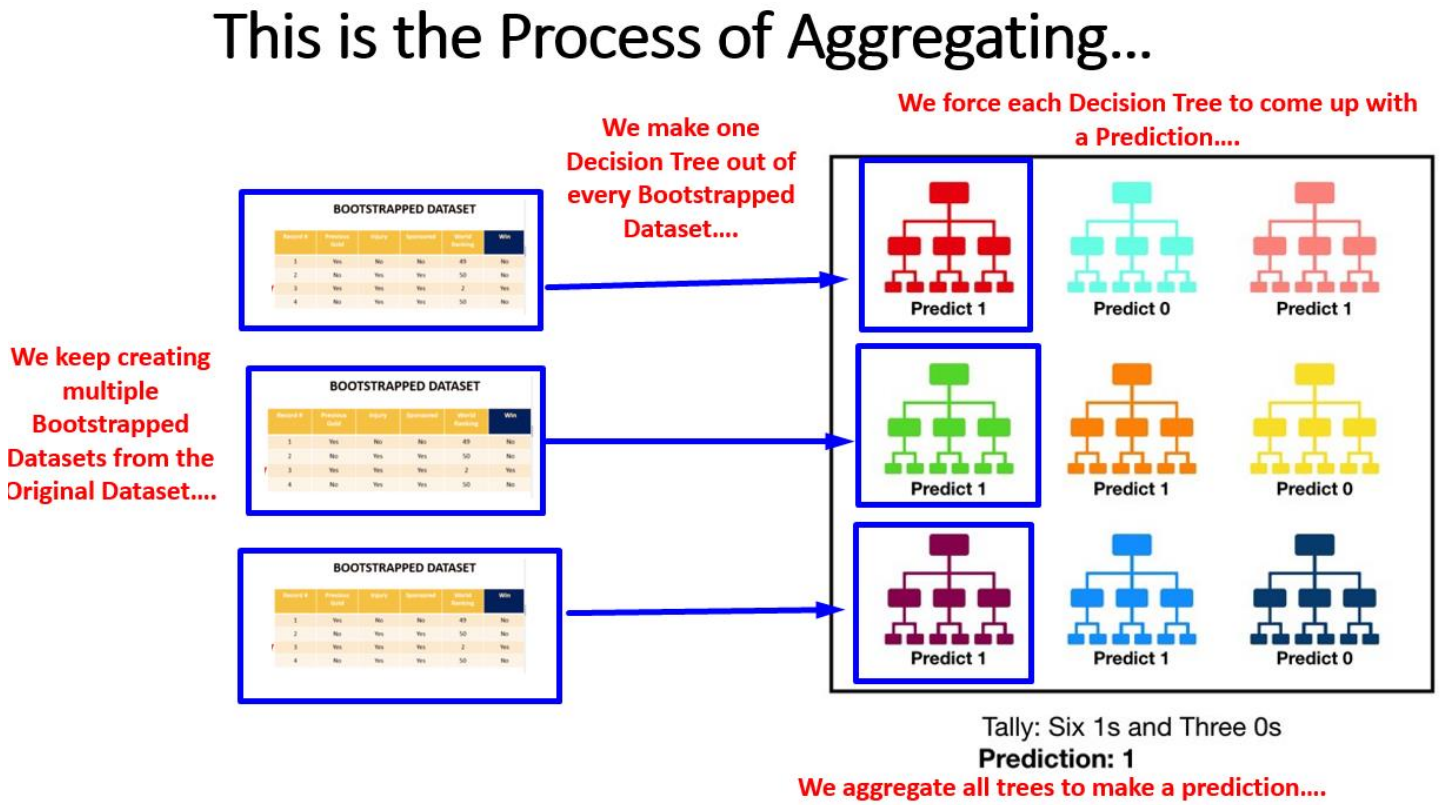
In essence, we have created a "Bootstrapped" Dataset FROM the "Original" Dataset….

Creating ONE Decision Tree from a Bootstrapped Dataset….

- Now that we have created our 1st Bootstrapped Dataset, we use it to identify the IMPORTANT KEY Features
- Then, we create ONE Decision Tree out of it.
- This process of Creating a Decision Tree out from a dataset has already been explained here….https://www.alvinang.sg/s/Decision-Tree-Classification-using-WEKA-by-Dr-Alvin-Ang.pdf



C. STEP 2: AGGREGATING



Note: Each "Tree" is also known as ONE "Bag" or ONE "Classifier"

III. RANDOM FOREST WITH WEKA

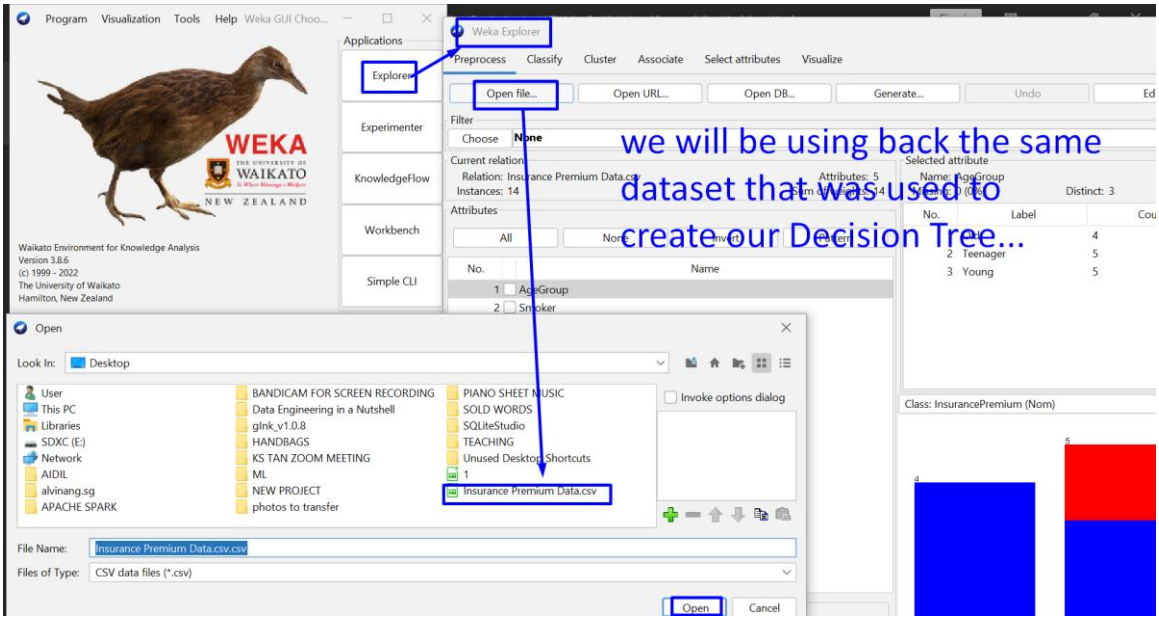
https://www.alvinang.sg/s/Insurance-Premium-Datacsv.csv

https://www.cs.waikato.ac.nz/ml/weka/

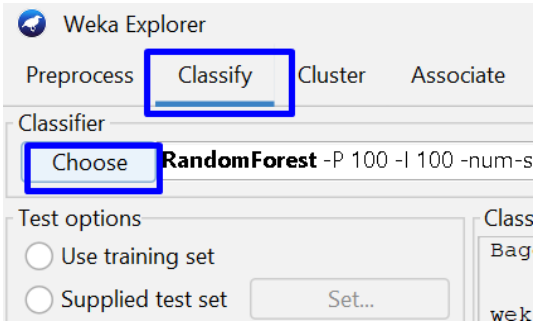
Random Forest continues off from Decision Tree….. here….

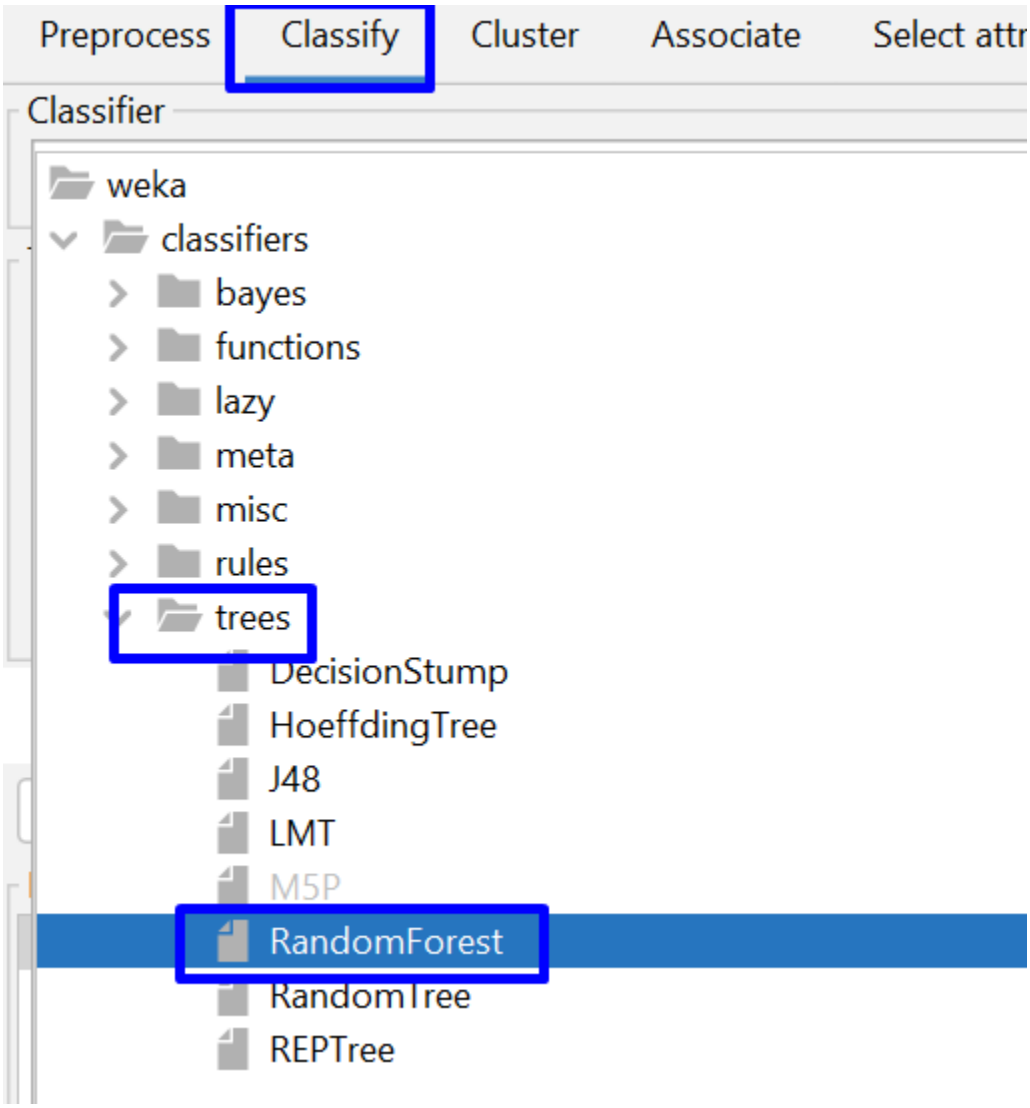
https://www.alvinang.sg/s/Decision-Tree-Classification-using-WEKA-by-Dr-Alvin-Ang.pdf

A. STEP 1: GETTING THE DATASET

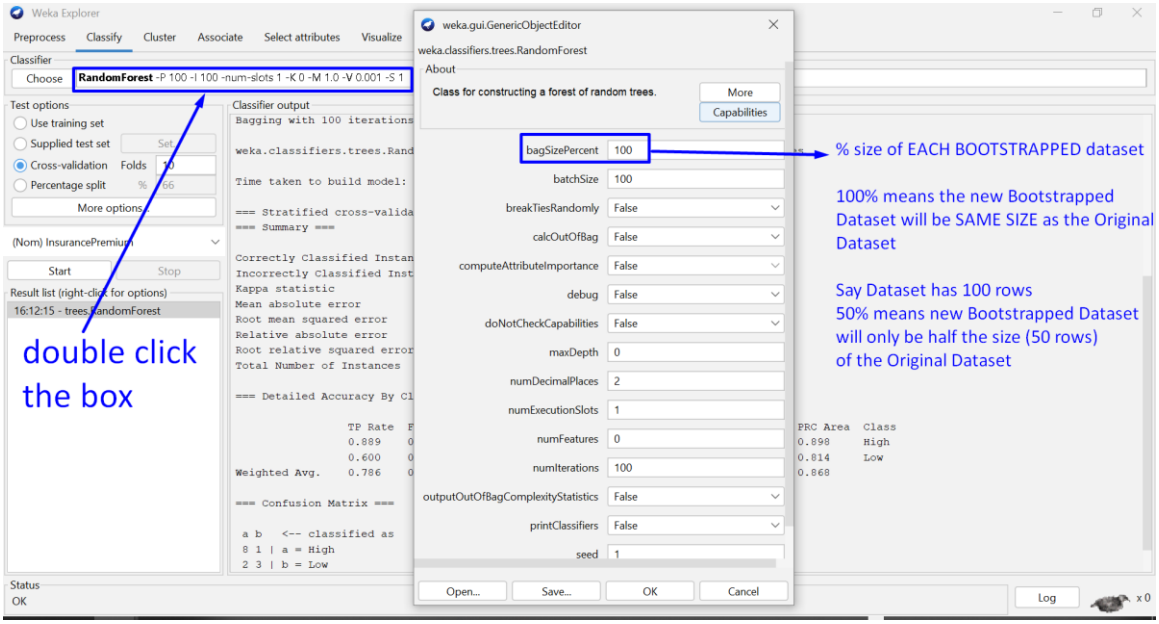


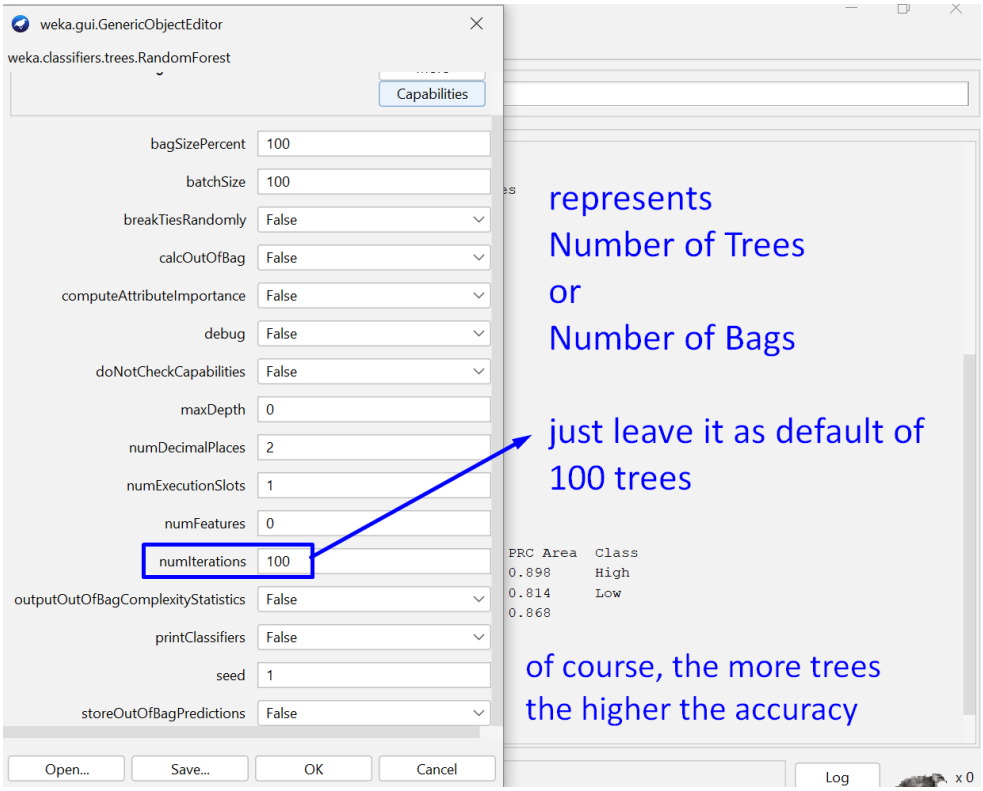
B. STEP 2: CHOOSING RANDOM FOREST CLASSIFIER

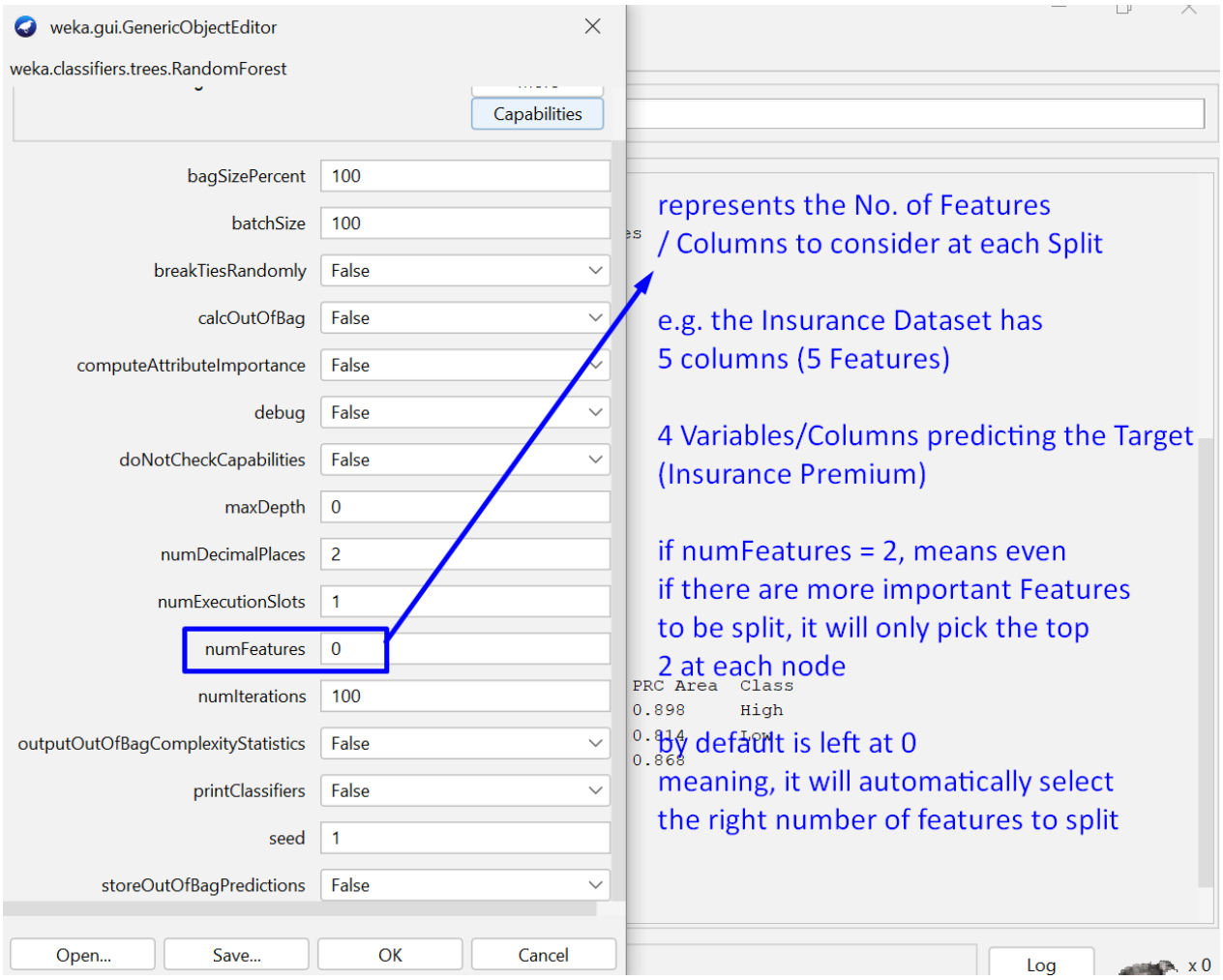




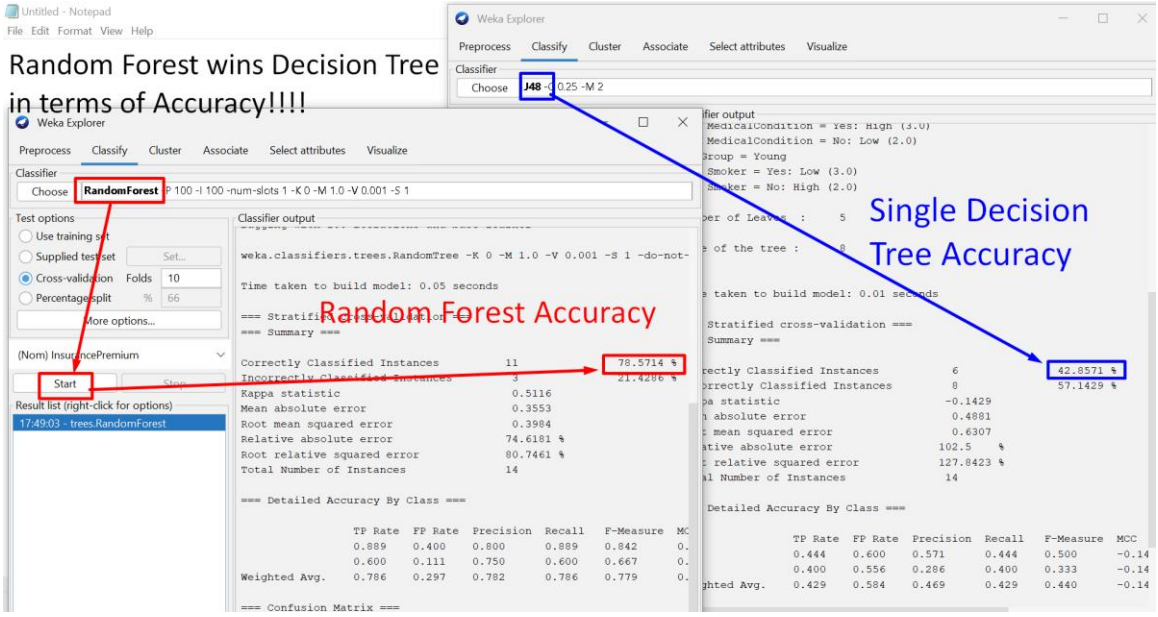
C. STEP 3: TUNING THE HYPERPARAMETERS

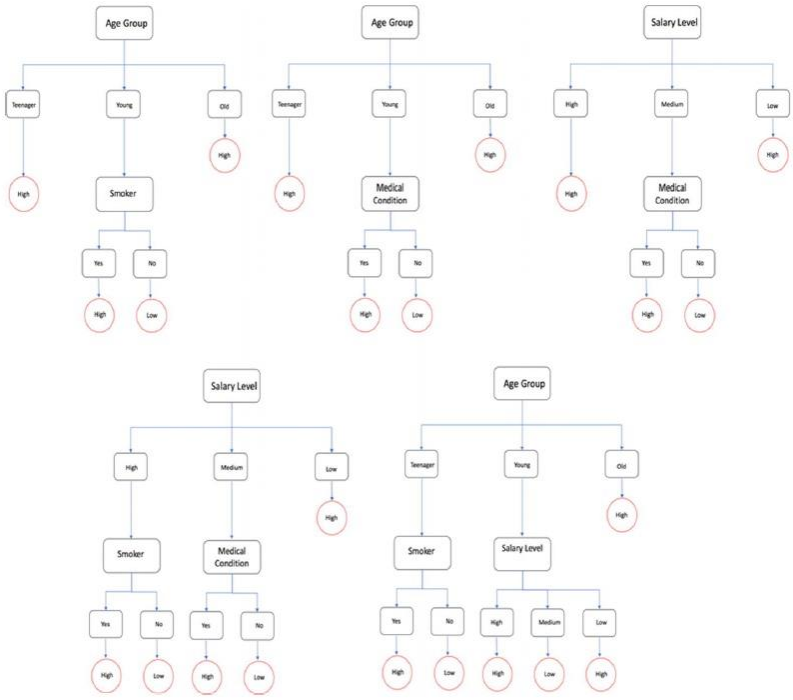






D. STEP 4: CHECKING THE RANDOM FOREST ACCURACY



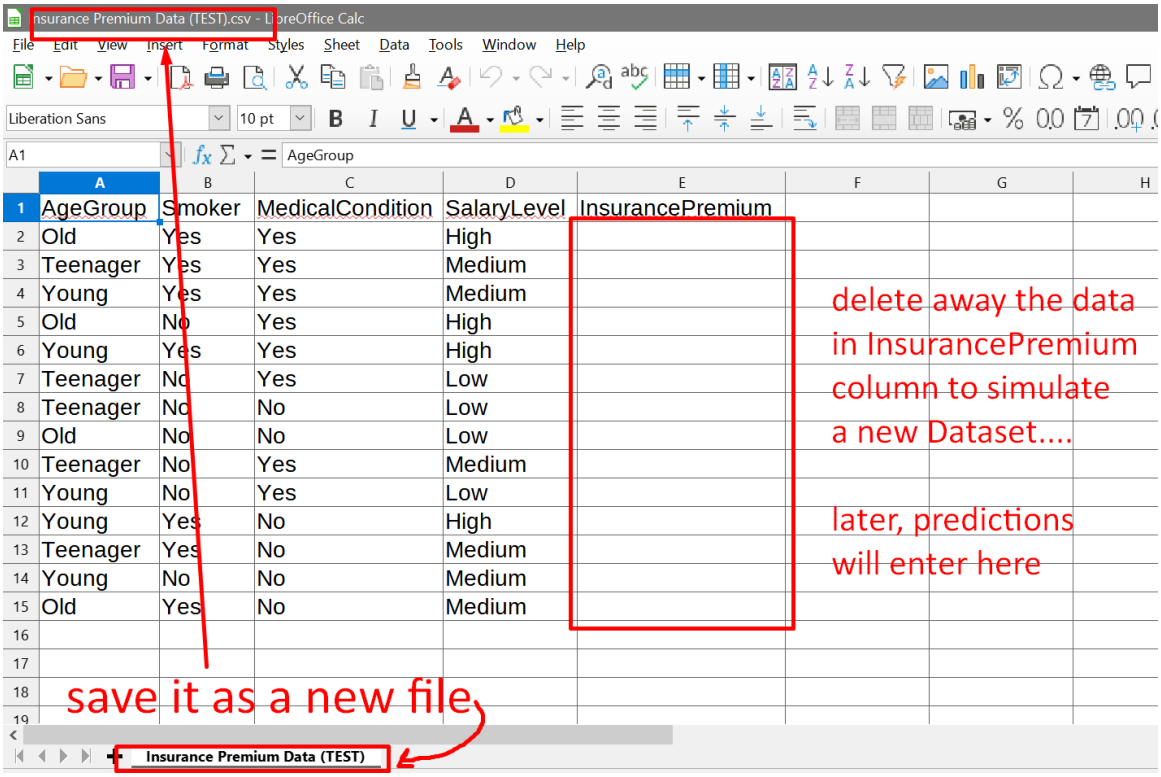


Above shows a possible Random Forest made up of multiple trees.

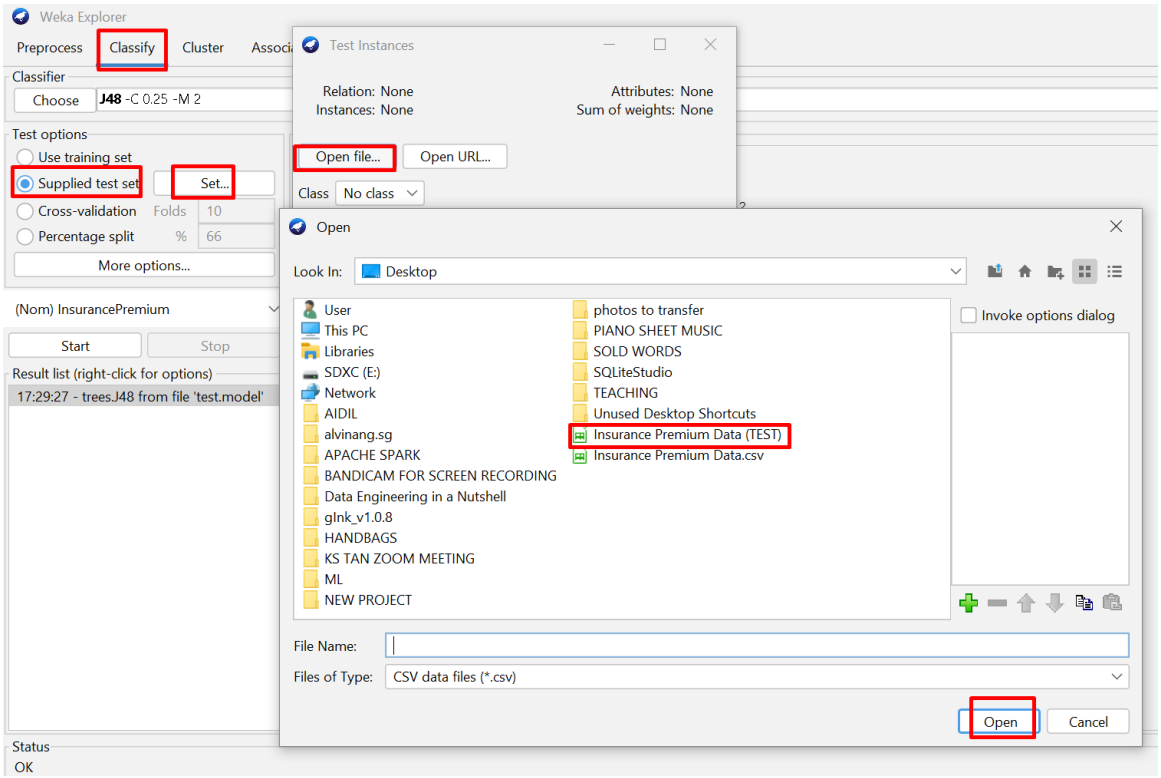
E. STEP 5: PREDICT WITH A NEW DATASET

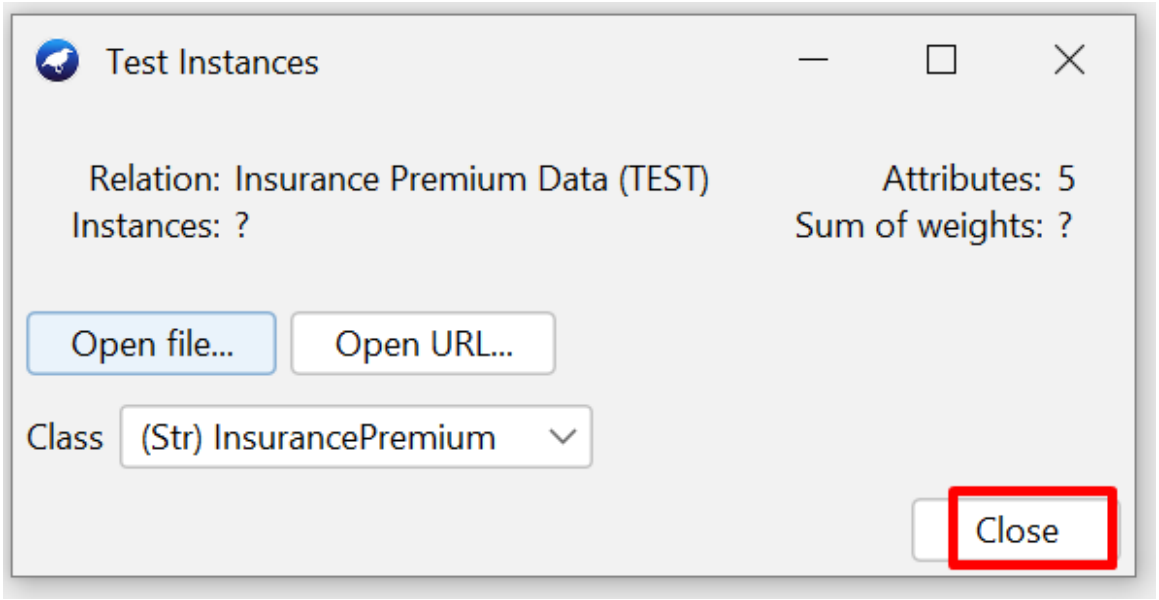
Go here and download the data again:

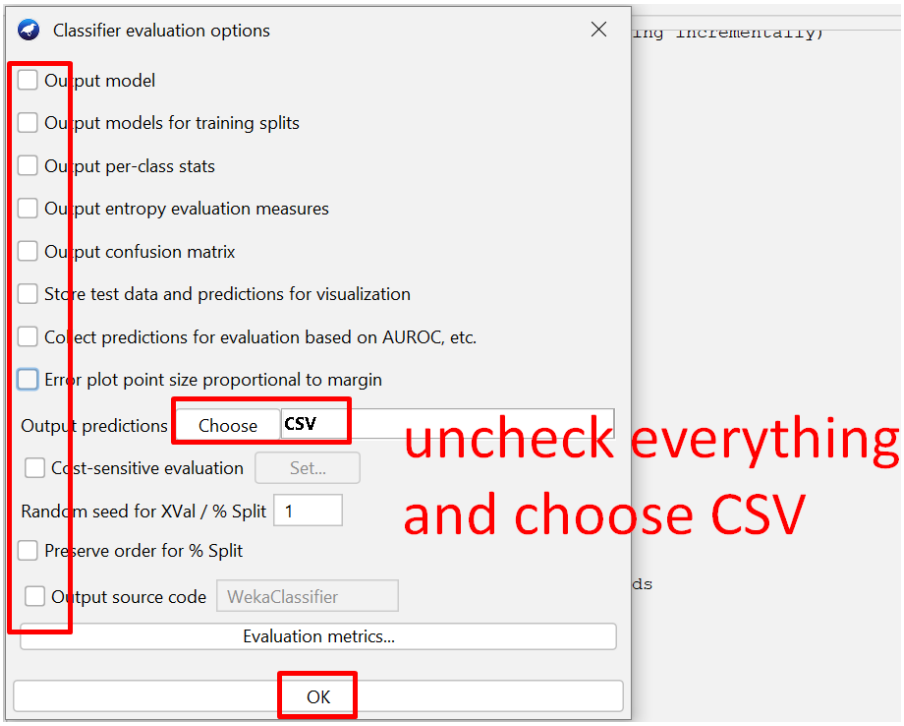
https://www.alvinang.sg/s/Insurance-Premium-Datacsv.csv

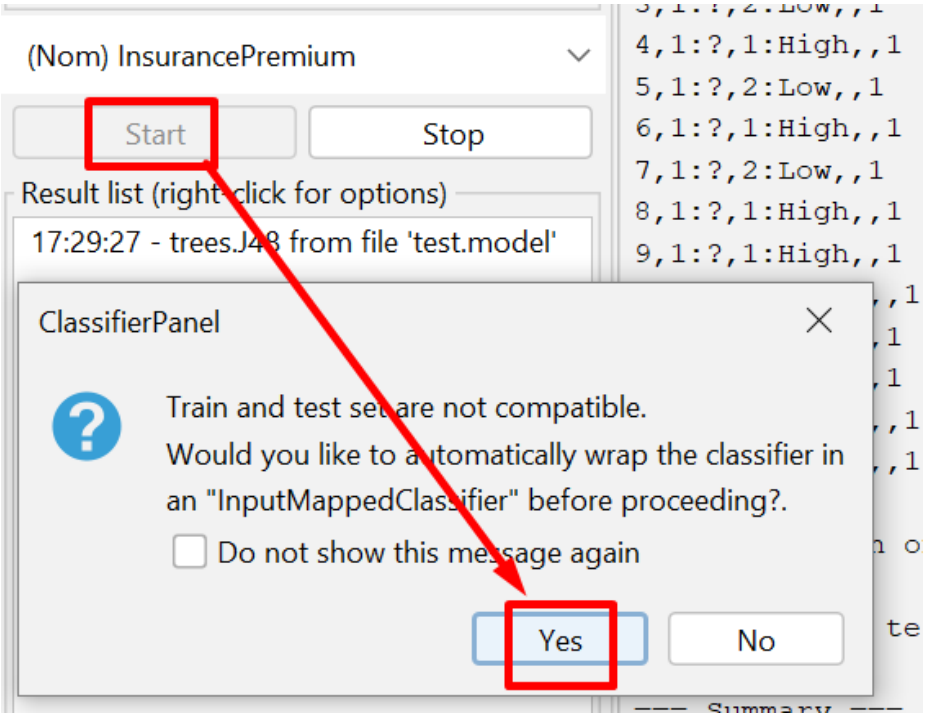


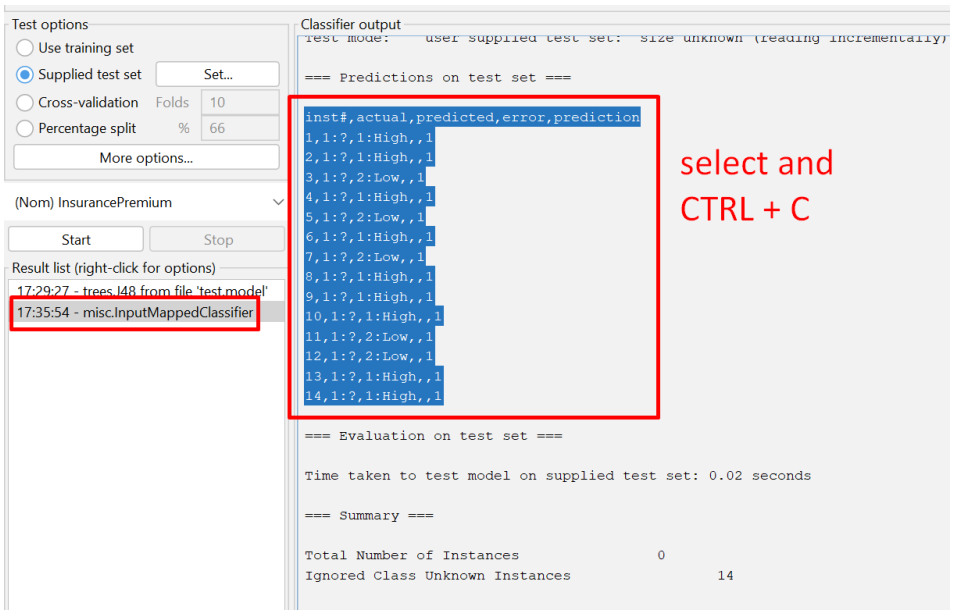
	A	B	C	D	E	F	G	H
1	AgeGroup	Smoker	MedicalCondition	SalaryLevel	InsurancePremium			
2	Old	Yes	Yes	High				
3	Teenager	Yes	Yes	Medium				
4	Young	Yes	Yes	Medium				
5	Old	No	Yes	High				
6	Young	Yes	Yes	High				
7	Teenager	No	Yes	Low				
8	Teenager	No	No	Low				
9	Old	No	No	Low				
10	Teenager	No	Yes	Medium				
11	Young	No	Yes	Low				
12	Young	Yes	No	High				
13	Teenager	Yes	No	Medium				
14	Young	No	No	Medium				
15	Old	Yes	No	Medium				
16								
17								
18								

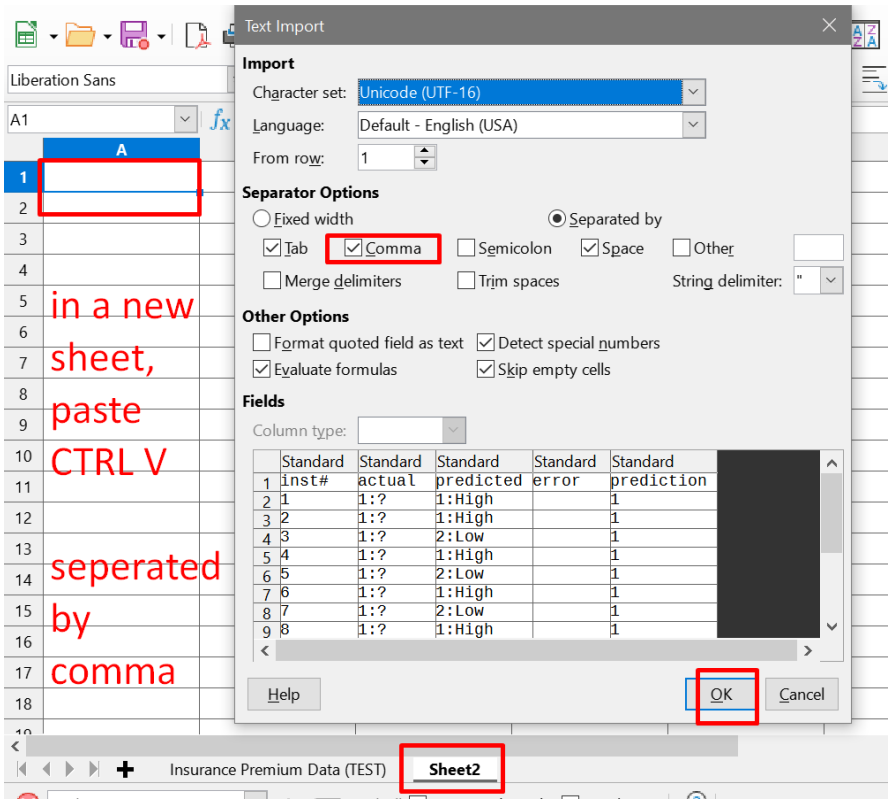


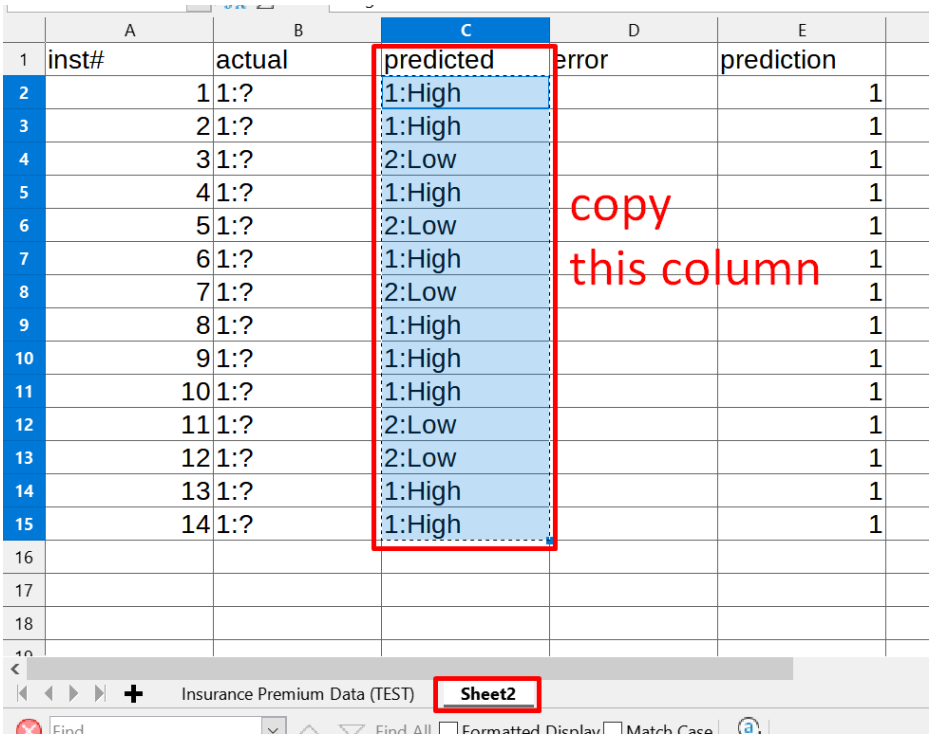


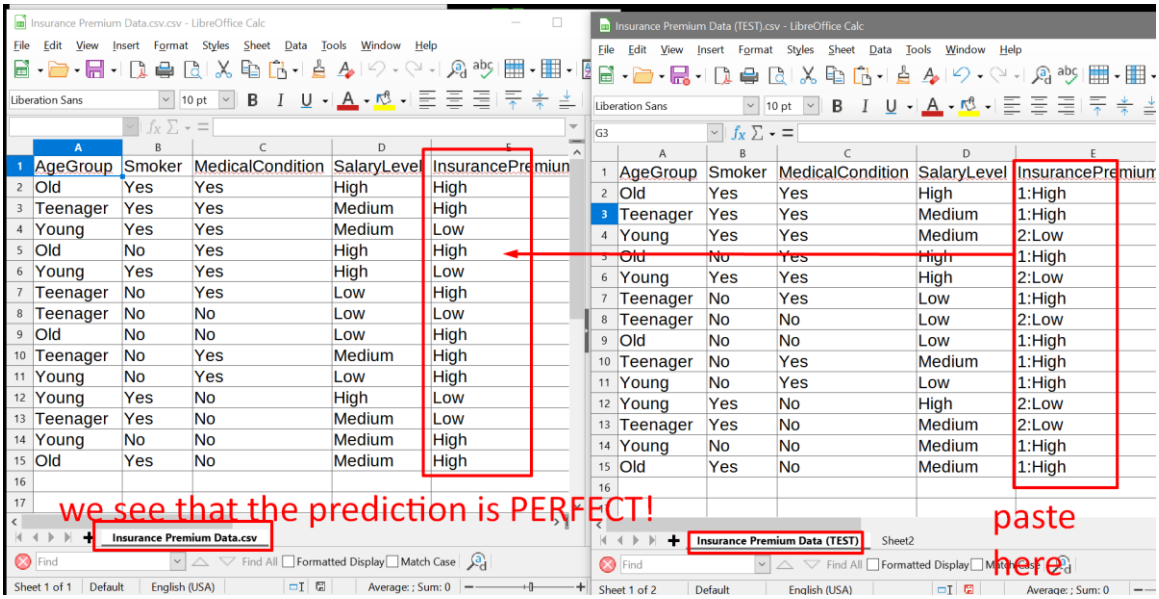












ABOUT DR. ALVIN ANG

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.