109.1 IOTBigData Final Form-AI (Decision Tree)	
Your works: Your team needs to implement the Decision Tree algorithms mentioned in this course in which the data set at your works is 500M at least. The motivation and purposes of your implementation are decided by your team, rather than from somebody else. However, you have to make them reasonable for your works in order to get a nice grade. Your Chinese writing is connected to your grade. ---Every team should have different works.---	
Title of your works: **政府對於新冠肺炎數據建立防疫措施**	
Chinese Name/ ID	蘇宇祥/4070E021
Chinese Name/ ID	沈明楷/4070E022
Code submission	■ R code is submitted as attached
Url shorten for your dataset	https://reurl.cc/Ag1O6Z
a. What are the variables you use for your model creation. b. Please write down the column names and their corresponding Chinese names.	a. Case_Type,People_Total_Tested_Count, Cases,Difference,Date,Combined_Key, Country_Region,Province_State,Admin2,iso2, iso3,FIPS,Lat,Long,Population_Count, People_Hospitalized_Cumulative_Count, Data_Source,Prep_Flow_Runtime b.個案類型、人數總計、確診人數、區別、日期、組合鍵、國家地區、省州、管理員、iso2、iso3、處理技術、年份、長期、人數統計、確診民眾住院累積人數、資料來源、流動時間
What is the goal/ target variable in your project?	日期、確診人數
a. Any **data cleaning** works in your project? b. If yes, what are api statements you use for these?	a.有 b.無
Please analyze your data on your **EDA stage**. Notice: please describe the each variables and their data distribution for your works.	a. b.
a. What kind of api	

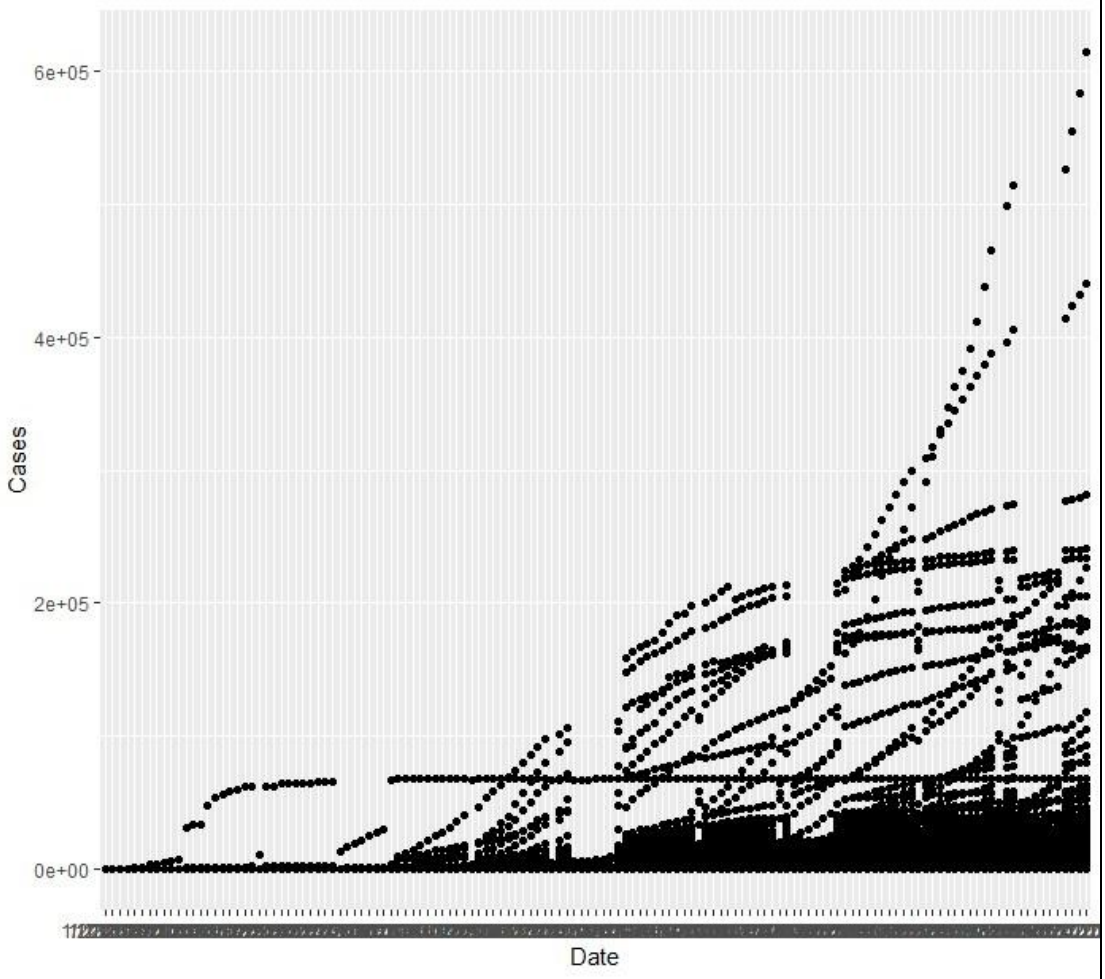
statements you use for your **model creation**? b. What is algorithm you use for Decision Tree model creation?	
a. What is the method your algorithm uses to **split up** a tree node? b. What are the **pruning skills** you use for your model creation? Why? c. Also, please provide the corresponding api statements.	
a. How do you train your **training data**? ex. Method, data proportion to training data, possible outputs and so on. b. Show your api statements for these.	
a. What are the **estimations** you use for verifying your model is all right? b. What are the corresponding values for the estimations?	
Motivation	目前新冠肺炎在全世界造成極大影響，病毒侵略速度越來越快，以致全世界各國都有確診病例，而希望藉由數據分析來查看各國確診狀況，以掌握病毒路徑，較好控制疫情。

Purposes	透過新冠肺炎數據分析，得到各國確診人數與日期比例，可得知確診人數增加或減少，能提供給政府做政策上的調配與整頓，建立完善防疫措施，好讓疫情能快速過去。

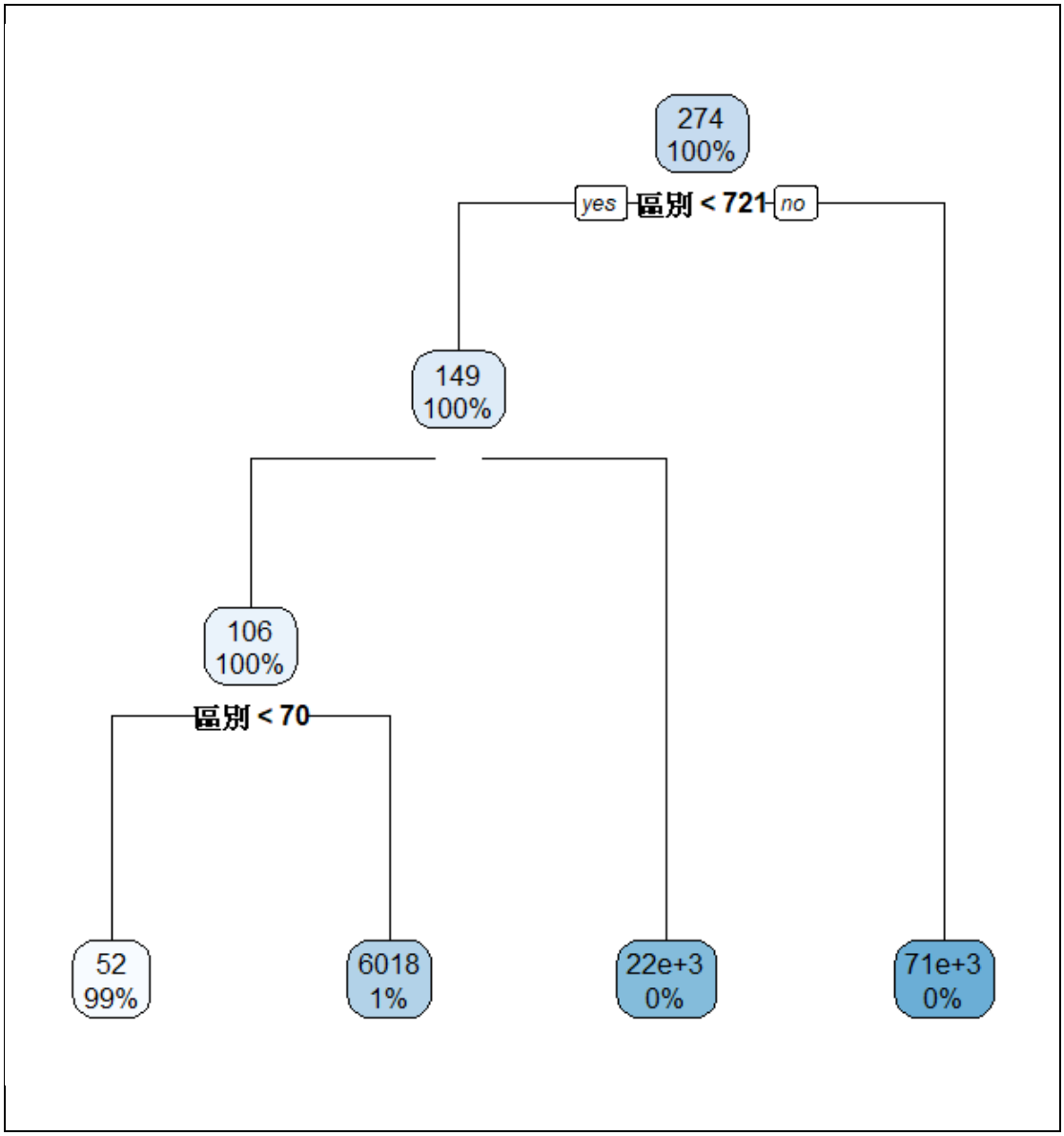
Outputs for your analysis and related works:

(Each photo for your output needs to be attached a corresponding description.)

結果呈現



從圖可知，當日期越往前走時，確診人數越多，所以在未來的日子，各國政府需自訂更好的防疫措施，以降低疫情持續擴散和確診人數增加的機率。


274
100%
yes
區別 < 721
no
149
100%
106
100%
區別 < 70
52
99%
6018
1%
22e+3
0%
71e+3
0%