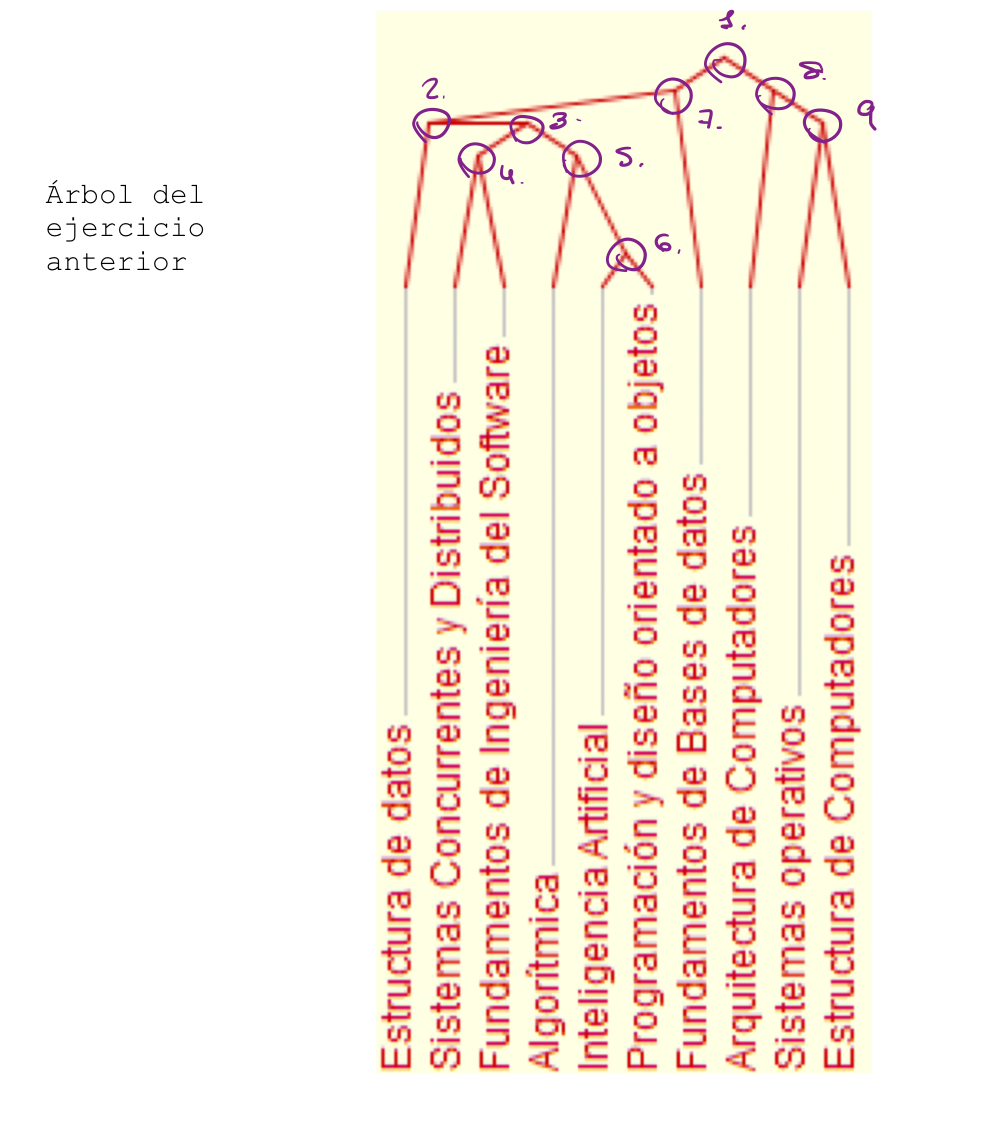
**PARTE 1: CLASIFICACIÓN DE ASIGNATURAS DE PRIMERO EN LA TAXONOMÍA OBTENIDA UTILIZANDO LA REJILLA DE REPERTORIO SOBRE LAS ASIGNATURAS DE SEGUNDO**

*Del ejercicio de clase de la semana pasada habréis obtenido clases de asignaturas de acuerdo a vuestro propio criterio utilizando la técnica de la rejilla de repertorio sobre las asignaturas de segundo curso, jerarquizadas en forma de taxonomía simple. Además para cada grupo habréis obtenido los criterios que definen esos grupos*

*Se trata de aplicar esos criterios para decidir para cada asignatura de primero a que grupos de los obtenidos pertenece (su clasificación en la taxonomía). Si no os convence la clasificación, replantearos los criterios, e incluso en su caso la coherencia de los grupos obtenidos, indicando las asignaturas de primero para las que no os convence la clasificación y por qué.*

*Se entrega la taxonomía, los criterios de cada grupo, y la clasificación para cada una de estas asignatura:*

*Algebra Lineal y Estructuras Matemáticas; Cálculo; Fundamentos de Programación; Fundamentos de Software; Fundamentos Físicos y Tecnológicos; Estadística; Ingeniería, Empresa y Sociedad; Lógica y Métodos Discretos; Metodología de la Programación; y Tecnología y Organización de Computadores.*

*Añadir comentarios sobre si os parece lógica la clasificación.*

Partiendo del árbol obtenido en la aplicación, tenemos nueve categorías diferentes en las que clasificar nuestras asignaturas de primero:

1. Todas las asignaturas.
2. Asignaturas orientadas a la programación.
3. Concurrencia y cálculos.
4. Software.
5. Procesamiento de datos.
6. Inteligencia artificial.
7. Programación + BD.
8. Bajo nivel.
9. Orientados a hardware.

Con más constructores, podríamos haber obtenido grupos más refinados. De partida, hay muchas asignaturas de primero relacionadas con la física , las matemáticas y las empresariales, como ALEM, Cálculo, FFT, Estadística, IES y LMD. Estas asignaturas no encuadran objetivamente en ninguna de nuestras categorías, ya que en ningún constructor hemos tenido en cuenta la carga matemática, física o de empresariales que tiene una asignatura, ni hemos hecho un constructor relacionado. Lo más lógico sería añadir constructores y realizar de nuevo todo el proceso.

Otras asignaturas sin embargo podrían estar en alguno de los grupos formados:

* FP: la incluiría en el 2. Si refinamos más, debe formar un grupo con ED.
* FS: en el 9. Preparatorio de SO.
* MP: ídem a FP.
* TOC: en el grupo 8. Se trabaja con puertas lógicas y álgebra de Boole, que es con lo que trabaja el hardware a bajo nivel.

**PARTE 2: CONOCIMIENTO PARA DESARROLLAR SBC PARA ACONSEJAR QUE RAMA ELEGIR UTILIZANDO APRENDIZAJE DE ÁRBOLES DE DECISIÓN**

Seguramente ya habréis comprobado que asesorar en qué asignaturas matricularse es un problema complejo con muchas opciones distintas. Por ello nos vamos a centrar en un subproblema concreto, aconsejar que rama o mención elegir.

Supongamos que inicialmente vamos a intentar aconsejar que rama elegir a partir de las siguientes características del alumno que pide consejo:

* Si le gusta las matemáticas
* En qué quiere trabajar
* La nota media que ha obtenido en las asignaturas que ha cursado hasta ahora
* Cómo de trabajador se considera
* Si le gusta programar
* Si prefiere asignaturas teóricas o prácticas

Para ello vamos a recoger una serie de alumnos y preparamos una tabla para que el experto rellene cual sería la respuesta o respuestas posibles para ese caso:

| Caso | Gusta Matemáticas | Quiere trabajar | Nota media | Gusta hardware | Es trabajador | Gusta  Programar | Prefiere teóricas o prácticas | Rama(s) aconsejada |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Alumno 1 | Si | Docencia | Alta | No | Mucho | Si | teóricas | **CSI** |
| Alumno 2 | No | Empresa  Pública | Media | No | Normal | No | prácticas | **SI** |
| Alumno 3 | Si | Empresa  Privada | Media | Si | Normal | Si | ambas | **CSI** |
| Si | Empresa  Privada | Media | Si | Normal | Si | ambas | **IS** |
| Alumno 4 | No | Empresa  Privada | Baja | No | Poco | Si | ambas | **IS** |
| No | Empresa  Privada | Baja | No | Poco | Si | ambas | **TI** |
| Alumno 5 | No | le da igual | Alta | Si | Mucho | Si | prácticas | **IC** |
| Alumno 6 | Si | Docencia | Media | No | Poco | Si | teóricas | **TI** |
| Si | Docencia | Media | No | Poco | Si | teóricas | **IS** |
| Alumno 7 | No | Docencia | Alta | No | Normal | Si | prácticas | **IS** |
| Alumno 8 | No | Empresa  Pública | Baja | Si | Normal | No | ambas | **IC** |
| Alumno 9 | Si | Empresa  Privada | Alta | No | Normal | Si | ambas | **CSI** |
| Si | Empresa  Privada | Alta | No | Normal | Si | ambas | **SI** |
| Si | Empresa  Privada | Alta | No | Normal | Si | ambas | **IS** |
| Alumno 10 | No | Empresa  Pública | Baja | No | Poco | Si | prácticas | **IS** |
| Alumno 11 | Si | le da igual | Baja | No | Poco | Si | ambas | **TI** |
| Alumno 12 | Si | Empresa  Privada | Alta | Si | Normal | No | prácticas | **SI** |
| Alumno 13 | No | Empresa  Privada | Baja | Si | Poco | Si | prácticas | **IC** |
| No | Empresa  Privada | Baja | Si | Poco | Si | prácticas | **IS** |
| Alumno 14 | No | Empresa  Privada | Alta | Si | Normal | Si | ambas | **SI** |
| No | Empresa  Privada | Alta | Si | Normal | Si | ambas | **IC** |
| Alumno 15 | Si | Docencia | Media | No | Mucho | Si | teóricas | **CSI** |
| Alumno 16 | No | Docencia | Media | No | Normal | No | teóricas | **SI** |
| Alumno 17 | No | Empresa  Pública | Media | Si | Poco | No | prácticas | **IC** |
| Alumno 18 | Si | le da igual | Media | No | Normal | Si | ambas | **CSI** |
| Si | le da igual | Media | No | Normal | Si | ambas | **IS** |
| Alumno 19 | No | Empresa Pública | Alta | Si | Mucho | Si | prácticas | **IS** |
| Alumno 20 | Si | le da igual | baja | No | Poco | No | teóricas | **CSI** |

**Ejercicio propuesto:**

0) Opcional, antes de empezar a aplicar la técnica del aprendizaje de árboles de decisión, añade o modifica características, modifica los valores posibles (adaptando la tabla), y añade ejemplos de alumnos que consideres que serían interesantes, obteniendo una nueva tabla adaptada a tu criterio sobre la que trabajar.

1. Haz tú de experto y rellena el campo de las Ramas Aconsejadas indicando la rama o las ramas que aconsejarias en cada caso (en muchos casos pensarás le puedes aconsejar más de una)
2. Aplicar un algoritmo de aprendizaje para obtener un árbol de clasificación a partir de esa tabla rellena (Ojo cuando tengas dos o más ramas aconsejadas para un alumno, el árbol deberás obtenerlo creando un ejemplo para cada rama, además tendrá que decidir cómo manejar los atributos donde algún valor no es incompatible con otros )
3. Obtener las reglas de decisión asociadas al árbol obtenido
4. Analizar las reglas obtenidas:
   1. En el caso de que no se consideren válidas, proponer modificaciones a las mismas o añadir ejemplos que complementen las posibles situaciones y que son las que harían no válidas esa regla. Si se añaden nuevos casos revisar de nuevo las reglas a obtener.
   2. En el caso donde el algoritmo de aprendizaje no sea capaz de decidir, pero si hay muchas más posibilidades de que ocurra una de las posibles opciones, proponer esa como regla por defecto y analizar bajo qué condiciones habría que descartar la opción por defecto.
   3. En el caso de varias opciones sin que una sea claramente preponderante, analizar cómo podría tratarse esa situación.

Se entrega una descripción del proceso seguido: la tabla propia inicial, las reglas iniciales obtenidas por el algoritmo, y para cada regla inicial si se considera adecuado o si se debe modificar, en cuyo se indica las modificaciones y el por qué.

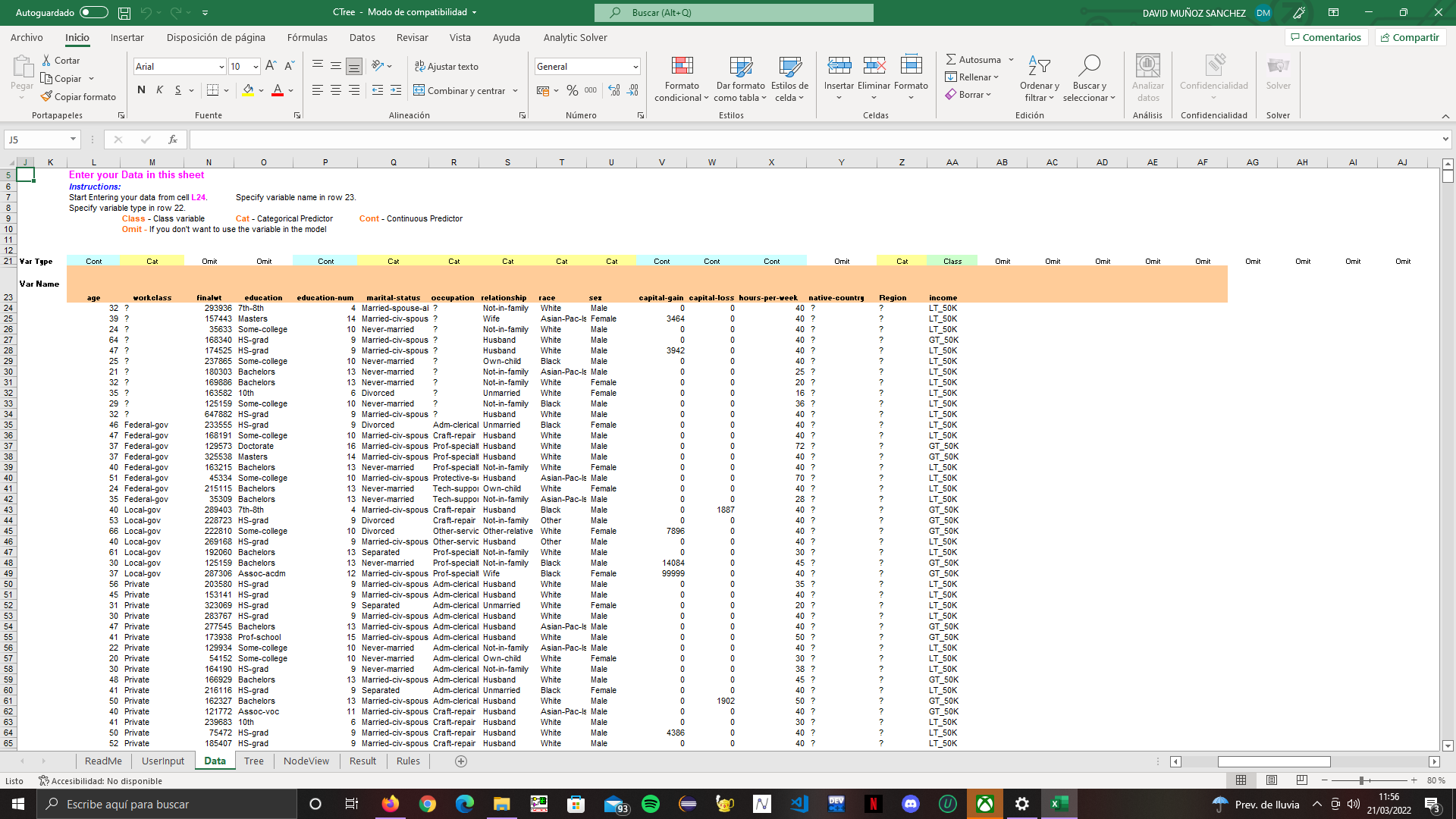
La tabla en el enunciado está rellena. En muchos casos se puede recomendar más de una rama. Las ramas son:

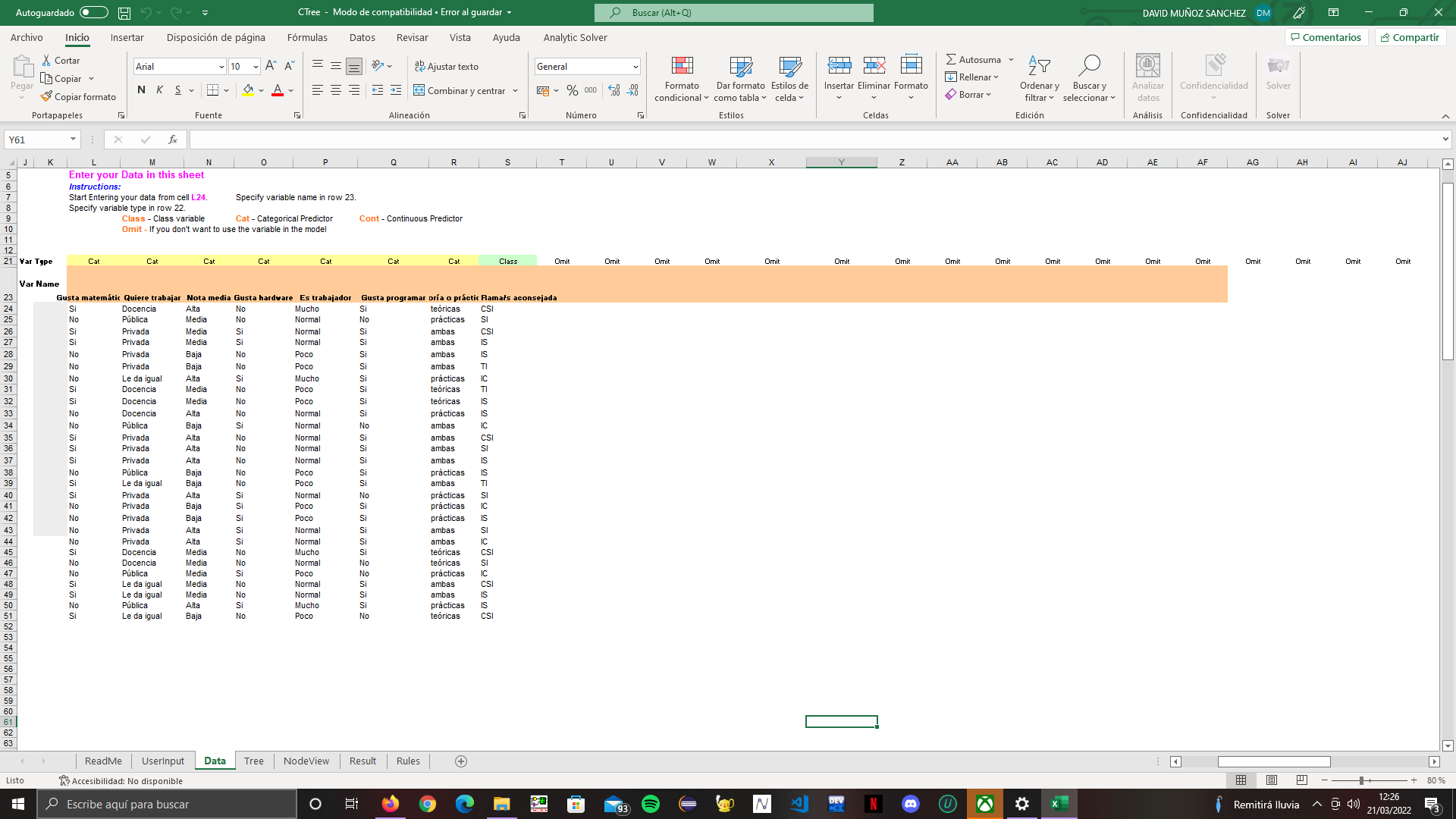
* CSI (Computación y Sistemas Inteligentes).
* IC (Ingeniería de Computadores).
* IS (Ingeniería del Software).
* SI (Sistemas de Información).
* TI (Tecnologías de la Información).

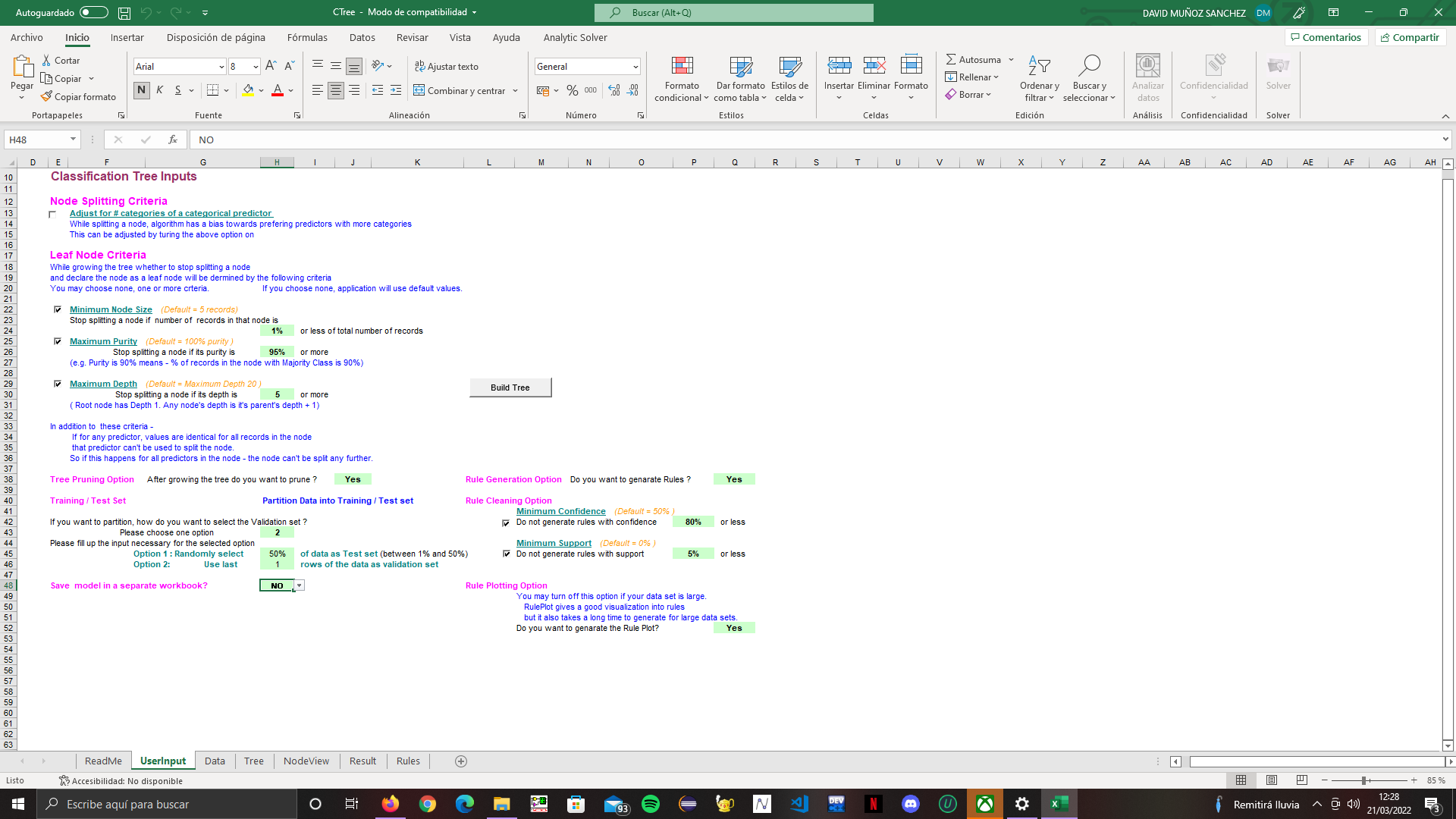
El software de árbol de decisión que voy a utilizar será CTree. Es una hoja de cálculo, adjunto captura de pantalla del programa en la hoja siguiente (número 7).

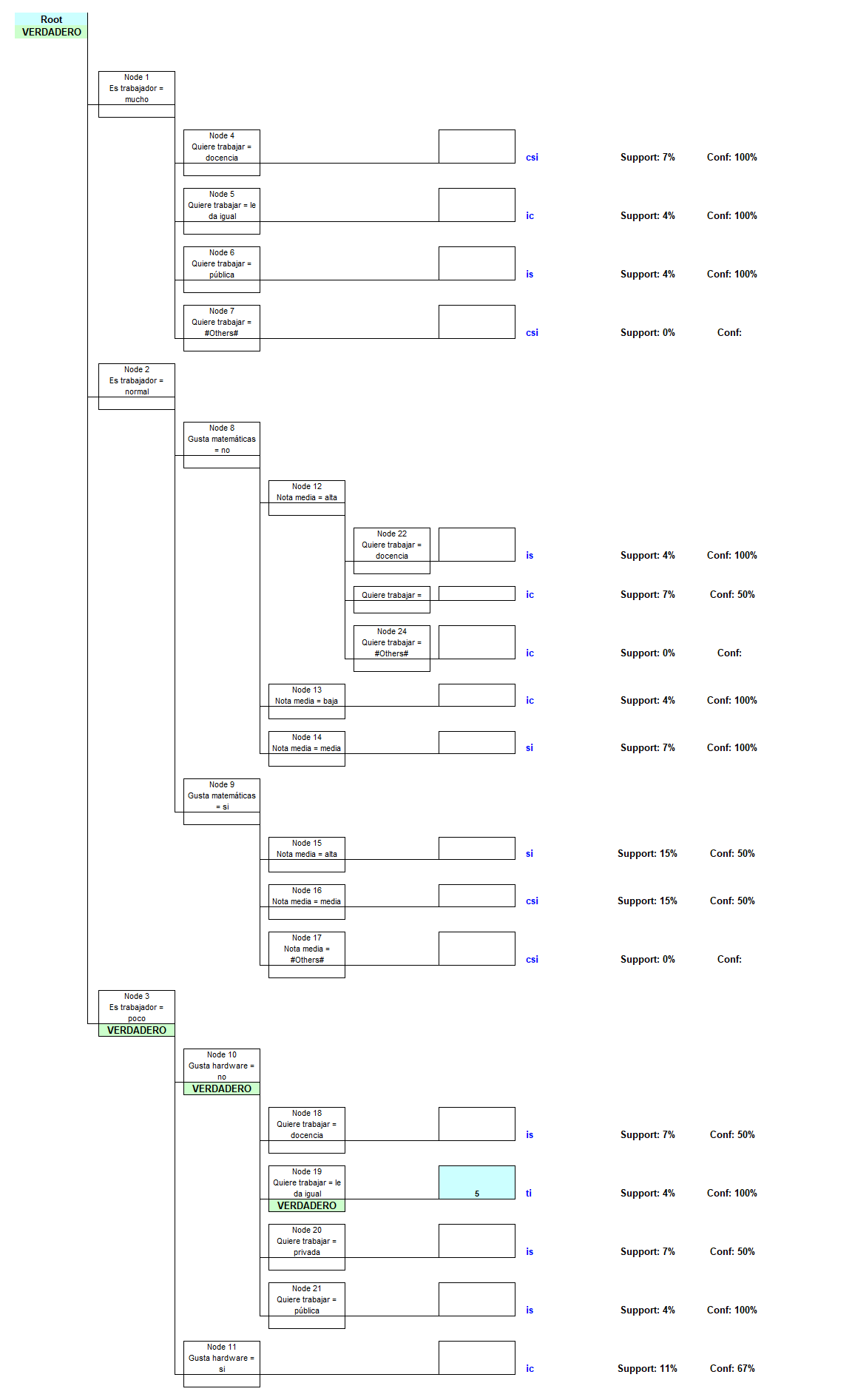
En la zona de datos que es la que se adjunta en la pantalla, debemos introducir nuestra tabla, todo son categorías menos la rama elegida, que es una clase.

En la página 8 se adjunta captura de pantalla de los datos introducidos y continúa el documento (se adjuntan capturas de la configuración para generar el árbol y del árbol resultante, así como otra información relativa al árbol de decisión).









La foto no se aprecia bien a simple vista, pero adjunto la foto para que se pueda observar mejor.

A partir del árbol generado obtenemos las siguientes reglas:

1. Si es muy trabajador y quiere trabajar en la empresa privada o en la docencia recomienda CSI.
2. Si es muy trabajador y le da igual donde trabajar recomienda IC.
3. Si es muy trabajador y quiere trabajar en la pública, recomienda IS.
4. Si es normal trabajando, no le gustan las matemáticas, su nota media es alta y quiere trabajar en cualquier sitio menos la docencia, recomienda IC.
5. Ídem a la anterior pero quiere trabajar en la docencia, entonces recomienda IS.
6. Igual que la cuatro pero, su nota media es media y no tenemos en cuenta donde quiere trabajar, recomienda SI.
7. Igual a la anterior pero la nota media es baja, recomienda IC.
8. Si es normal trabajando, le gustan las matemáticas y tiene nota media alta, SI.
9. Igual que antes pero con nota media media, recomienda CSI.
10. Igual pero la nota media es baja, recomienda CSI.
11. Si es poco trabajador, no le gusta el hardware y quiere trabajar en la privada, en la pública o en docencia, IS.
12. Igual pero si no especifica dónde trabajar, TI.
13. Igual que la 11 pero si le gusta el hardware recomienda IC.

El árbol de decisión presenta varios problemas. La opción primera de si trabaja mucho, poco o normal debería cambiarse por hardware y software, ya que después tendríamos menos reglas incoherentes, como en la 11 y la 12, que recomienda TI e IS indistintamente, o en la rama de muy trabajador, que no se ha tenido en cuenta la carga matemática, siendo determinante para recomendar CSI.

El árbol se debería hacer con más casos y más columnas de información en la tabla. En mi opinión, el resultado que he obtenido, está demasiado condicionado por el lugar dónde quiere trabajar el alumno y lo que él piensa que trabaja. La regla por defecto podría ser IS, que es mayoritaria (un 33,33% de proporción con respecto a las demás).

# Fuentes

Saha, Angshuman. “Data Mining in Excel - SayHello2Angshu.” *Google Sites*, <https://www.sites.google.com/site/sayhello2angshu/dminexcel>. Accessed 21 March 2022.