# Activity: Introducing Decision Trees

Eckel, TJHSST AI2, Spring 2024

Background & Explanation

For our first example of **supervised learning**, we'll be programming our code to construct **decision trees** – one intuitive way of classifying observations. We want to play with this concept a bit before we get to coding though – a mathematically rigorous treatment of this idea is complex and involves several new ideas!

But to start: we're just going to discuss the basic concept, and then you're just going to **draw your own decision tree** using **your brain**.

The concept of a decision tree is well known - a flowchart with no loops. I suspect the following figure, a decision tree representing a process to decide to grant someone a loan or not, will make intuitive sense to you:
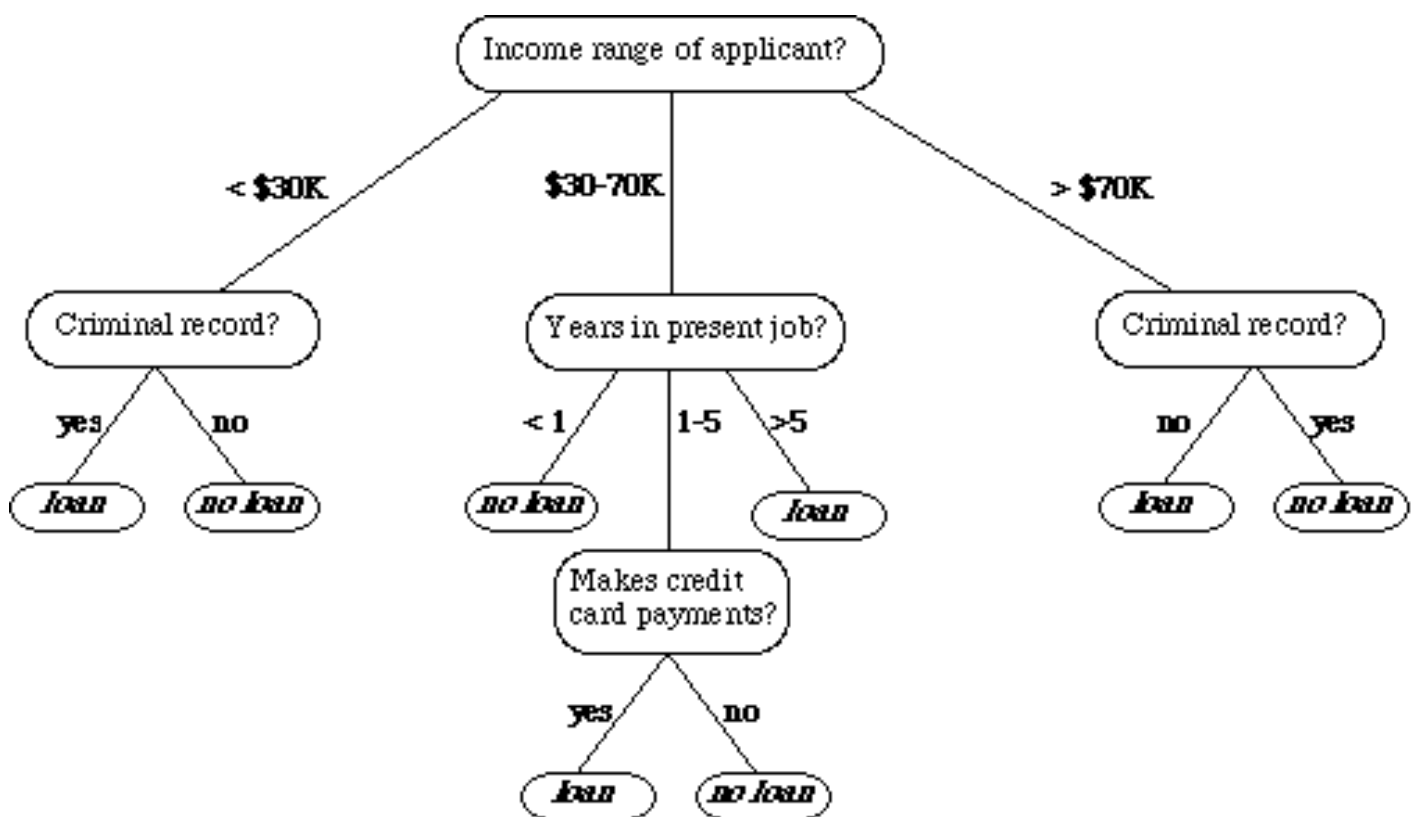
The decision tree considers one variable at a time, proceeding in a series of decisions until it reaches a final classification, in this case a binary choice between "Loan" or "No Loan". Note that we have an input vector for each observation - the variables are "Income Range", "Criminal Record", "Years in Present Job", and "Makes Credit Card Payments". So, this is a length four input vector. **Not every variable in an input vector needs to be considered in order to correctly classify each input vector!**

It can be difficult to generate images like the one on the prior page; for purposes of sharing decision trees easily, and later on in this assignment for purposes of printing out decision trees that your program generates, we can use an outline form instead that uses indentation to show levels. Consider the following text recreation of the flowchart above:

```
* Income Range?
   * < 30K
      * Criminal Record?
         * Yes --> Loan
         * No --> No Loan
   * 30-70K
      * Years in present job?
         * < 1 --> No Loan
         * 1-5
            * Makes credit card payments?
               * Yes --> Loan
               * No --> No Loan
         * > 5 --> Loan
   * > 70K
      * Criminal record?
         * No --> Loan
         * Yes --> No Loan
```

Once the visual and text descriptions of the flow chart make sense, continue on to the next page.

Build a 100% correct decision tree by hand

The first activity to build understanding here is not a coding activity but instead a sort of logic puzzle to do by hand. Your goal in this activity is to **build a correct decision tree by hand that has zero error on the following data set.** This is a common data set used for AI instruction called "Play_Tennis"; it represents a four-variable input vector classifier deciding whether today's conditions are right to play tennis or not.

Since this data set is well-known, if you google it you will find many examples of decision trees. To maximize your learning from this activity, DO THIS YOURSELF FIRST without seeking any help (aside from help from me). It will lead to a deeper understanding in the long run, I promise. AFTER you've made your own, feel free to compare with others!

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

YOUR TASKS:

1. Think carefully about what the *first* variable should be that you use for the *first* decision in your tree.

2. Continue from there to create a decision tree on the above data set that achieves zero error.

3. Either write it out LEGIBLY by hand and take a picture OR type out a text-outline version of your flowchart.

4. Write **a brief explanation** of why you made the choice you did for the *first* variable in #1.

5. If your creation is digital, **send it to me on Mattermost.** If it is on paper, **hand it to me in class.**