Medical Prescription Prediction

Joshua A. Martínez González, B.S. Digital Systems and Robotics
Tecnológico de Monterrey, Campus Querétaro, México
A01703723@tec.mx

Abstract Throughout time, humanity has experienced different diseases, some of which have led to the death of millions of people. However, medicine and the treatment of these and new diseases have also advanced over time along with technology. At present, there are already specific robots for surgery and other electronic devices that are of great help in the medical area, considering the number of inhabitants in the world, applying Artificial Intelligence for the detection and allocation of the necessary treatment to treat a disease. It begins to be a necessity to satisfy the demand in the shortest possible time.

Introduction

Since ancient times, doctors have treated patients with various techniques, which initially lacked therapeutic value, resulting in harm to the patient. With the passage of time medicine evolved, in such a way that now there is a procedure to be able to prescribe a treatment. We went from carrying out treatments based on the word of the "doctor" to clinical trials that assure us almost 100% that the established treatment will be adequate for the disease. [1]
Even with all the evolution of medicine, there are still erroneous prescriptions by doctors.
On the other hand, technology has advanced greatly, in such a way that it is part of our daily lives, now machines are capable of performing human tasks and acting taking into account certain variables, thus having a certain autonomy, keeping this in mind, the use of technology (AI) for the prescription of medical treatments can be a great help, creating a great social and economic impact for good

Framing the problem

The main objective and motivation of this work is to establish a precedent that shows the ability of an AI to make correct medical prescriptions, since this can greatly help to reduce the saturation of patients in hospitals that only come for a minor diagnosis, with this, doctors could devote most of their time to seriously ill patients. Likewise, people would not have to spend a lot of time waiting in the hospital to be treated.

Solution

As a possible solution to the problem described, a decision tree is proposed, since this is a "diagram" that determines a course of action, through which it is possible to observe the steps that must be taken and the different choices with which count and that can affect the process. At the end of this "map" are the results of the possible decisions. [2]
A decision tree is made up of:

- Innternal nodes: represent each one of the characteristics or properties to consider to make a decision.
- Branches: represent the decision based on a certain condition (eg probability of occurrence).
- Final nodes: represent the result of the decisión [3].

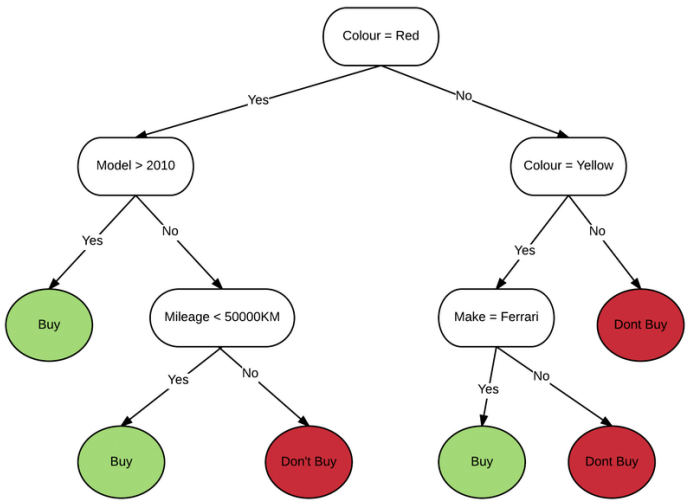

Image 1: Decision Tree diagram

Explaining Data

For our problem we will test a hypothetical case. This hypothetical case was taken from Kaggle, by Pablo M. Gómez. Suppose we have collected information on different patients who suffered from the same type of disease, each of them responded to one of 5 different drugs Drug A, Drug B, Drug C, Drug X and Drug Y.
A model will be built to find the appropriate medication for the patient.
The features available are: Age, Sex, Blood Pressure and Cholesterol.

Data
Age
Sex
Blood Pressure
Choresterol

Table 1: Dataset Labels

Data Preparation

The algorithm written for the creation of the decision tree is capable of dividing the information provided in the best way, likewise, it distinguishes if the information is continuous or categorical, since for continuous information it is possible to ask if the value is greater, less or equal, however, for categorical data we can only ask if the value is equal to another. However, 20% of the information will correspond to the test set to validate the model.
The "Drugs" column will be replaced by "label" as good programming practice.
The process followed by the algorithm for creating the tree is shown below.

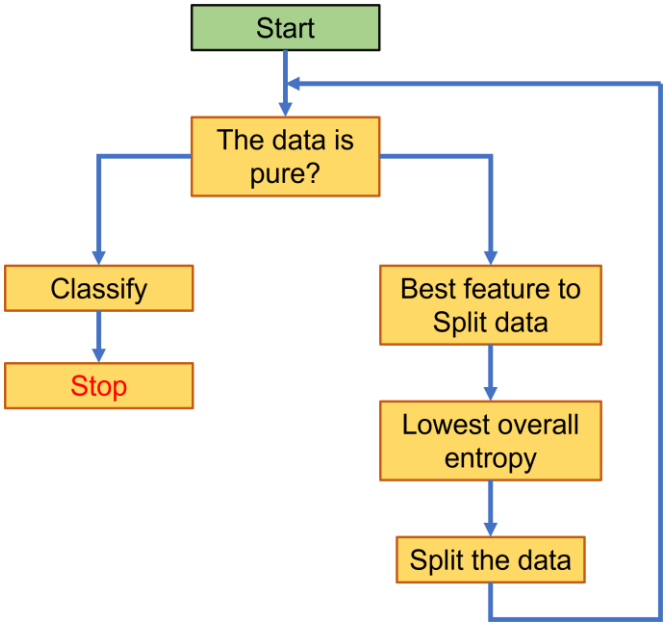

Image 2: Decision Tree diagram flow

As can be seen in the diagram, what is sought is to split the data, in such a way that in the end we have pure data, that is, a single class, after obtaining this we can create the final leaves of the tree, which are those that we will give the final prediction.

Entropy

Entropy is an information theory metric that measures the impurity or uncertainty in a group of observations. It determines how a decision tree chooses to split data. [4]

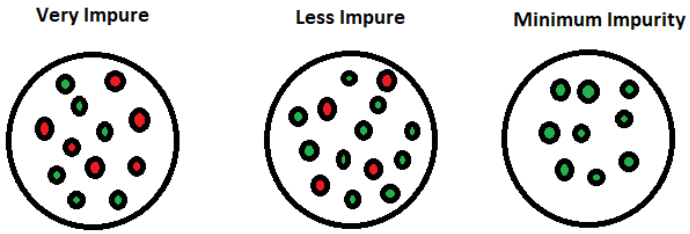

Image 2: Data purity

The entropy may be calculated using the formula below:

$$E = -\sum_{i=1}^{N} p_i \log_2 p_i$$

p_i is the probability of randomly selecting an example in class i.

Information Gain

The information gain is the measure of how much information a feature provides about a class. Information gain helps to determine the order of attributes in the nodes of a decision tree.

The main node is referred to as the parent node, whereas sub-nodes are known as child nodes. We can use information gain to determine how good the splitting of nodes in a decision tree.[4]

$$Gain = E_{parent} - E_{children}$$

Predictive Modelling

An algorithm was made that creates a decision tree, this will be compared with the one created by the sklearn framework. The decision trees created are shown below.

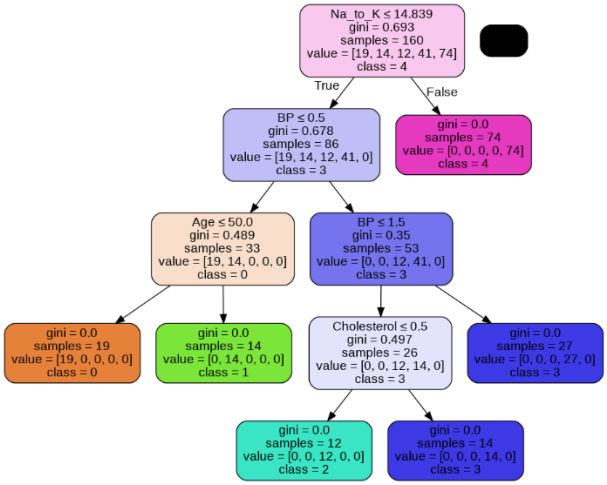

Image 3: Decision Tree (1)

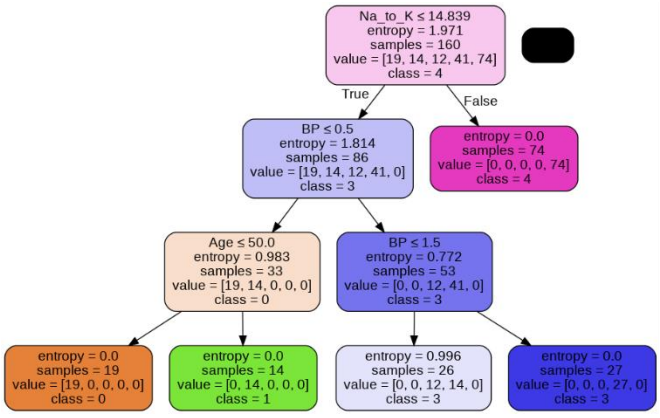

Image 4: Decision Tree (2)

As can be seen in the previous images, in each node it is detected whether or not the condition is met, which determines the route that it will follow until it reaches the final sheet, the prediction created.

Results

The results obtained were very good, the accuracy percentage is between 85% and 95% depending on the tree created. Seeing this project in the future in real treatments, I consider that it will be a great tool if we feed the AI with sufficient data, since it depends on this that it can make the greatest number of possible relationships in order to generate a more accurate prediction, which is what is search in the field of medicine.

```
{'Na_to_K <= 14.642': [{'BP = HIGH': [{'Age <= 50': ['drugA', 'drugB']},
                                      'drugX']},
                       'drugY']}
{'Na_to_K <= 14.642': [{'BP = HIGH': [{'Age <= 50': ['drugA',
                                                     'drugB']},
                                      'drugX']},
                       'drugY']}
Accuracy: 0.95
```

Image 4: Decision Tree (2)

```
Accuracy of the model is 100.0
              precision    recall  f1-score   support

       drugA       1.00      1.00      1.00         4
       drugB       1.00      1.00      1.00         2
       drugC       1.00      1.00      1.00         4
       drugX       1.00      1.00      1.00        13
       drugY       1.00      1.00      1.00        17

    accuracy                           1.00        40
   macro avg       1.00      1.00      1.00        40
weighted avg       1.00      1.00      1.00        40
```

Image 9: Sklearn predictions

Future work

For the continuation of this research, information will be sought that contains different treatments for a specific disease and thus be able to test this decision tree with real data, so that it can be a useful tool for hospitals and clinics.

Conclusion

A simple decision tree can be very useful in human life, since it gives a computer the ability to make relationships and find patterns in information that can take a human a long time to find. For the medical field, this is of vital importance, since we are dealing with people and their lives.

References

[1] Díaz J. (2008). El tratamiento médico: experiencia, base teórica y método. Recuperado de: **http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1729-519X2008000400009**

[2] Artyco (S.F). Qué es un árbol de decisión y su importancia en el Data Driven. Recuperado de: **https://artyco.com/que-es-un-arbol-de-decision-y-su-importancia-en-el-data-driven/**

[3] Javatpoint (S.F). Decision Tree Classification Algorithm. Recuperado de: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

[4] Ogola W. (2021). Entropy and Information Gain to Build Decision Trees in Machine Learning. Recuperado de: https://www.section.io/engineering-education/entropy-information-gain-machine-learning/