# A Contribution to the Time Series Machine Learning Open Source Java Toolkit: TSML – The J48SS Classifier

What is a Classifier?

In the field of Machine Learning, a classifier is an algorithm which accepts data into a model it has formed of a particular *domain* and gives back a classification, a piece of data that acts as a prediction for what the user might like to know. For example, in the weather domain, a user may wish to predict the weather tomorrow. So the user can feed in certain data such as recent temperature, wind speeds and air pressure, from nearby and in the local region, and the classifier can return whether it thinks it is most likely to be sunny, or rainy, or foggy, etc. In order to create this model, the classifier must first be trained on the data, being told which temperature, wind speeds, air pressures etc. result in which weather conditions.
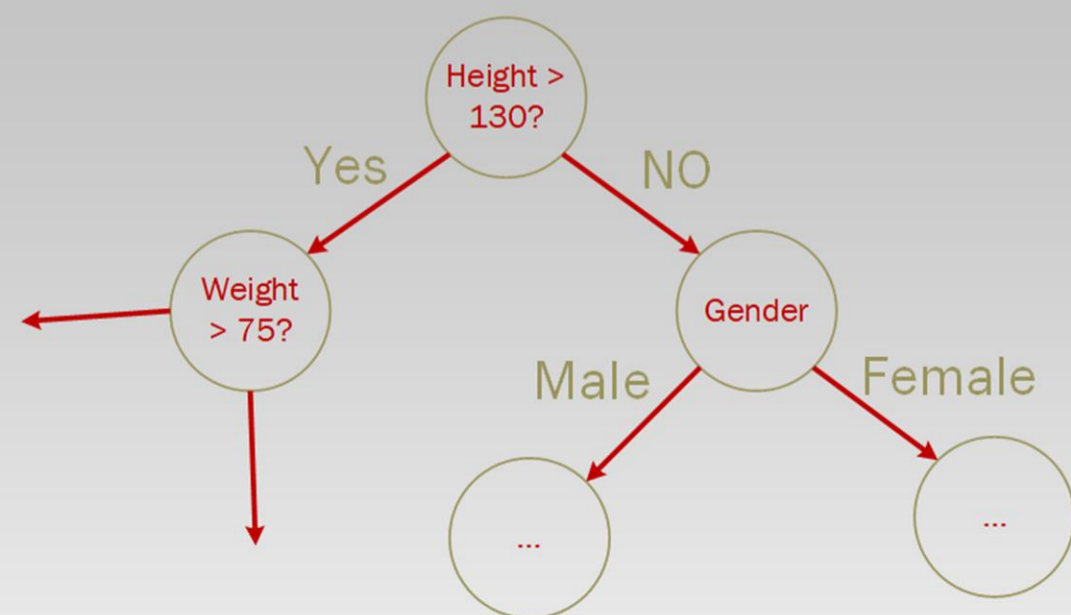
An Example of a Decision Tree Classifier algorithm

What is special about J48SS?

J48SS is able to handle a number of different data types in one data set, and process them together. It can handle discrete (categorical) data, such if a person is a smoker or not; numeric data, like the age or weight of a person; time series data, such as a person's blood pressure measured over several days; and sequence data, such as a person's symptoms or the order of drugs someone has been given. As you might imagine from this scenario, there is potential for this algorithm to be useful to help predict the health risk to a person in order for a doctor to better treat them. J48SS can work on just one type of data if needed, however, and is very efficient at doing so.

What is the TSML Toolkit?

The TSML Toolkit is an ongoing open source project from the TSML research group within UEA. The toolkit is a collection of algorithms designed to perform predictions on time series data, or manipulate data in different ways to allow information to be extracted, or to evaluate the effectiveness of different algorithms. Code within the toolkit help our researchers to find more effective algorithms, and form more complex models to make accurate predictions. For more information, see timeseriesclassification.com
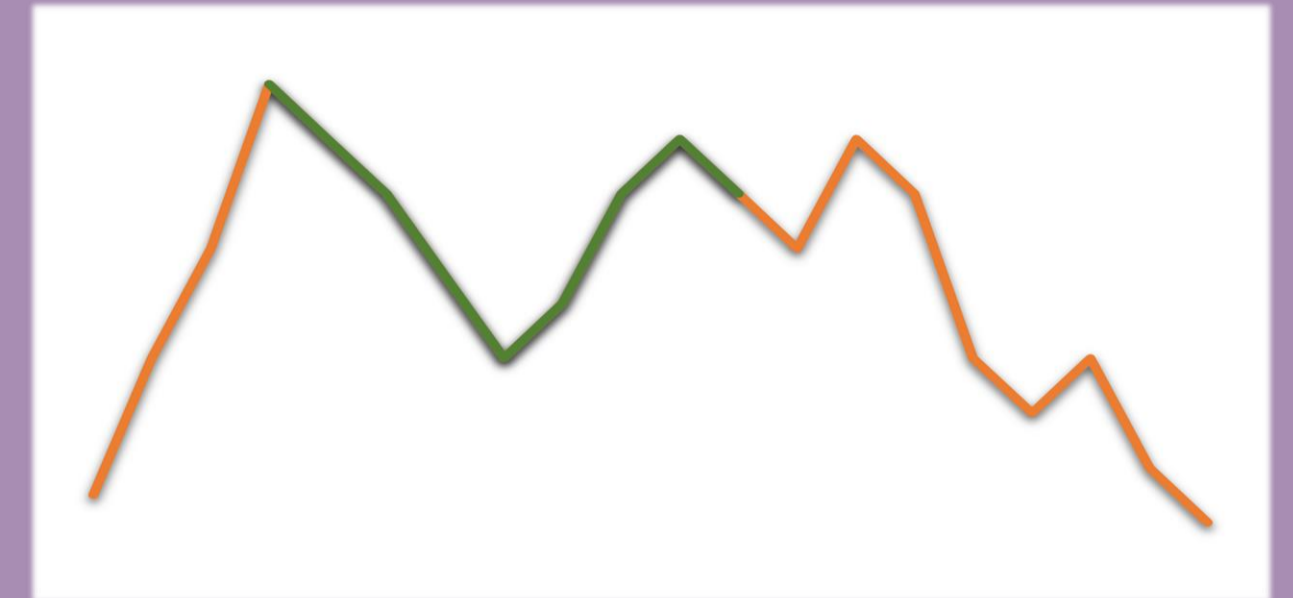
An example segment of time series data. The highlighted green section is an example of a *wavelet*

How does it do that?

Discrete and numeric data is processed in the same way as it's predecessor, J48 – a well known and well used algorithm in Machine Learning.

For time series data, the classifier treats the data as a series of points along a time axis. It then isolates *wavelets* – small subsections of this data, which it can manipulate in different ways to find other matching wavelets, which likely indicates that the same event in different points in time created that signal.

Sequential data is processed through a *pattern generator* which attempts to match a certain patten in the data with other, similar patterns, which would indicate a similarity in the two pieces of data.