DECISION TREE REGRESSION (ID3)

**Decision Tree:**

* Decision tree is a supervised learning technique that can be used for both Classification and Regression.
* It is a tree structured classifier, where internal node represents the feature of datasets, branches represents the decision rules and each leaf node represents the outcome.
* Decision tree contain categorical data as well as numerical data



**Steps for Implementation:**

**Step 1:** Data Pre-Processing step.

**Step 2:** Fitting a Decision Tree algorithm to the training set.

**Step 3:** Predicting the test result.

**Step 4:** Test accuracy of the result (Creation of confusion matrix).

**Step 5:** Visualizing the test set result.

**Advantages of Decision tree:**

* It is simple to understand.
* It can be very useful for solving decision related problem.
* There is less requirement of data cleaning compared to other algorithms.

**Disadvantages of Decision tree:**

* It contains lots of layer which makes it complex.
* It may have an overfitting issue.

**Decision Tree Regression**

Decision Tree Regression is a type of regression analysis in which a decision tree model is used to predict a continuous dependent variable from a set of independent variables.

**Types of Decision Tree Regression:**

1. CART (Classification and Regression Trees)
2. ID3 (Iterative Dichotomiser 3)
3. CHAID (Chi-Squared Automatic Interaction Detection)
4. MARS (Multivariate Adaptive Regression Splines)

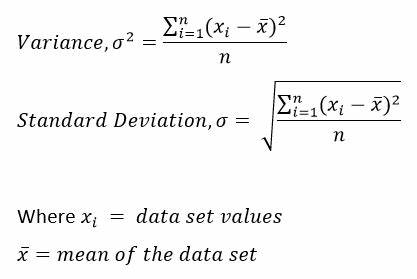
***ID3 (Iterative Dichotomiser 3):***

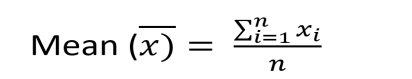
* ID3 stands for Iterative Dichotomiser 3, It was first introduced by Ross Quinlan in 1986 and is named such because the algorithm iteratively (repeatedly) dichotomizes (divides) features into two or more groups at each step.
* The ID3 algorithm recursively splits the data into smaller subsets based on the attribute values until all samples belong to the same class or until no further split is possible. The final tree can be used to make predictions on new, unseen data.
* Most generally ID3 is only used for Classification problems but we can use for Regression by replacing information gain with **Standard Deviation Reduction (SDR)**
* Here, standard deviation is used to calculate the homogeneity of a numerical sample (target variable). If the numerical sample is completely homogeneous w.r.t independent variable (we check for each independent variable separately), then its standard deviation is **zero**.
* We use standard deviation and will check on each split how much reduction in standard deviation is there, and the node (independent variable) which has more reduction in standard deviation that will be declared as a **decision node**. (In first iteration this node will be called as root node).

**Reduction in Variance or Standard Deviation Reduction (SDR):**

|  |  |  |
| --- | --- | --- |
| The standard deviation reduction is based on the decrease in standard deviation after a dataset is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest standard deviation reduction (i.e., the most homogeneous branches). |  |  |
|  |  |  |

**Formulas :**

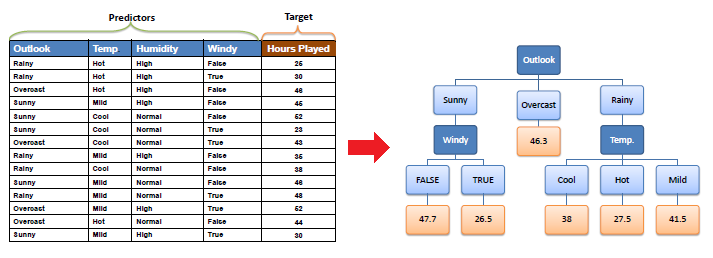




First we calculate the standard deviation of the target variable. And then calculate the weighted standard deviation of target w.r.t each independent variable, Then take a difference and this is known as **Reduction in Standard Deviation.**

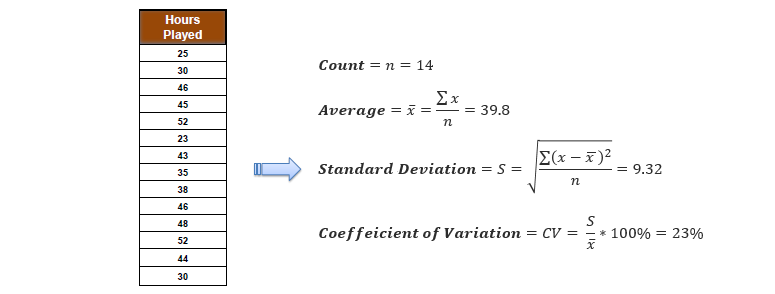
**Example:**

A sport hosting company would like to decide to host a cricket match between India and South Africa based on whether data. Weather data that is available has attributes like outlook, temperature, humidity and wind. And has a decision variable how many hours were played. We need to build a decision tree model to predict based on weather data how many hours will be played?

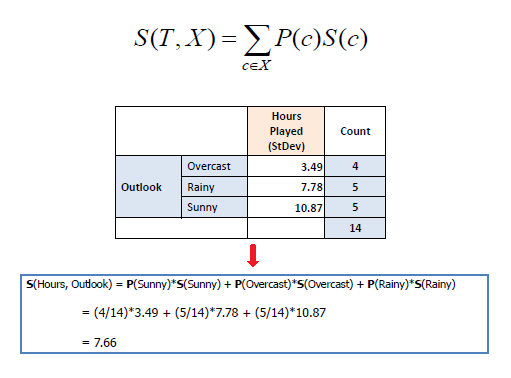
**Data:** 

***Solution:***

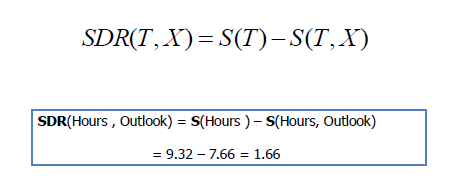
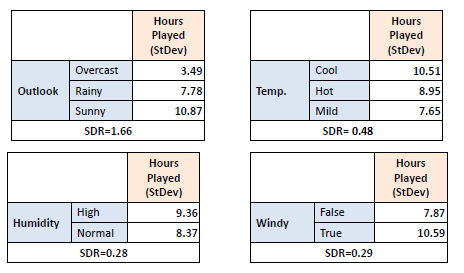
**Step 1: First will calculate the total standard deviation with respect to target variable**



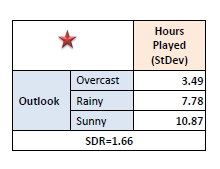
**Step 2:** Standard deviation for two attributes (target and predictor)



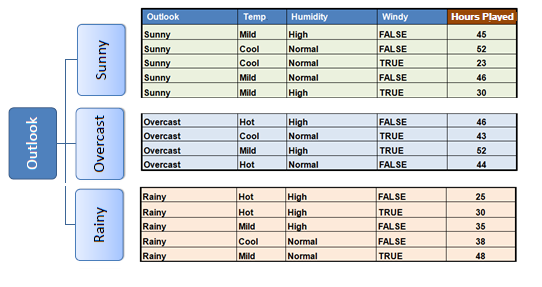
**Step 3:** The dataset is then split on the different attributes. The standard deviation for each branch is calculated. The resulting standard deviation is subtracted from the standard deviation before the split. The result is the standard deviation reduction.



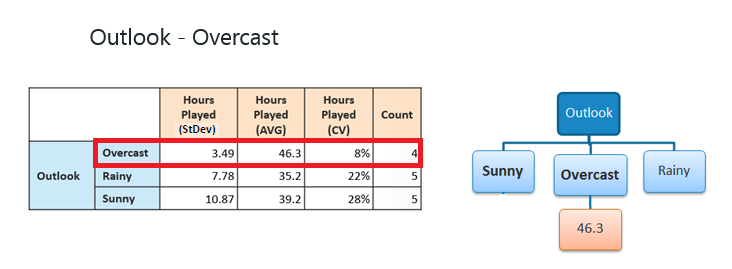
**Step 3**: The attribute with the largest standard deviation reduction is chosen for the decision node.



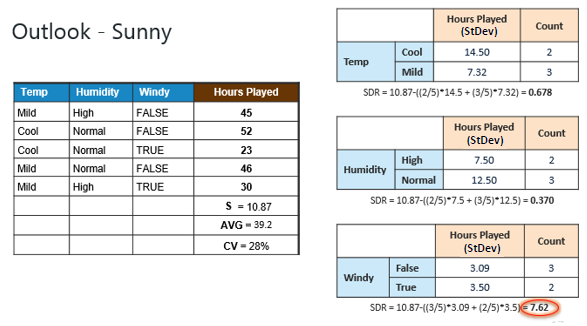
**Step 4**: The dataset is divided based on the values of the selected attribute. This process is run recursively on the non-leaf branches, until all data is processed.

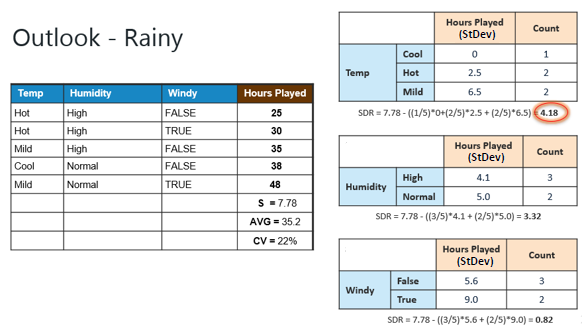


|  |  |  |
| --- | --- | --- |
| "Overcast" subset does not need any further splitting because its CV (8%) is less than the threshold (10%). The related leaf node gets the average of the "Overcast" subset. |  |  |
|  |  |  |

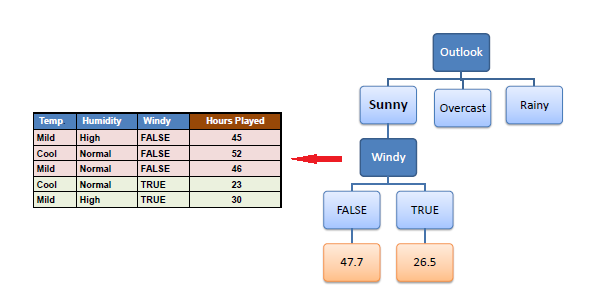


The "Sunny", ”rainy” branches have an CV (28%), (22%) respectively ,more than the threshold (10%) which needs further splitting. We select "Temp" as the best node after "Outlook" because it has the largest SDR.

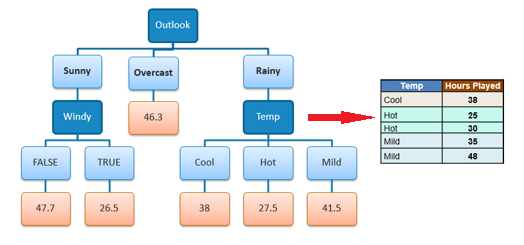




The number of data points for both branches (FALSE and TRUE) is equal or less than 3 we stop further branching and assign the average of each branch to the related leaf node.



The number of data points for all three branches (Cool, Hot and Mild) is equal or less than 3 we stop further branching and assign the average of each branch to the related leaf node.



**Decision Tree Interpretation:**

1. If outlook condition is Rainy and temperature is mild then prediction on number of hours match can be played is 41.5 hours irrespective of other conditions.
2. If outlook is overcast then irrespective of other conditions, prediction is 46.3 hours.
3. If outlook is Sunny then if it is windy then prediction is 26.5 hours and if it is not windy then prediction is 47.7 hours.

**Advantages of ID3 Algorithm:**

* Understandable prediction rules are created from the training data.
* Inexpensive to construct
* Extremely fast at classifying unknown records Easy to interpret for small-sized trees.
* Can easily handle redundant or irrelevant attributes (unless the attributes are interacting)

**Disadvantages of ID3 Algorithm:**

* The space of possible decision trees is exponentially large. Greedy approaches are often unable to find the best tree.
* Does not take into account interactions between attributes
* Each decision boundary involves only a single attribute

**Applications of ID3 Algorithm:**

1. On Food Database.
2. Web Attack Detection.
3. In Diabetes.
4. Cancer identification.
5. For Breast Tumour Diagnosis
6. Predicting Heater Outlet Temperature
7. To reduce Cost sensitive Decision Tree, etc…

**REFERENCES:**

1. Chen Jin, Luo De-lin, Mu Fen-xiang, “An Improved ID3 Decision Tree Algorithm” Proceedings of 2009 4th International Conference on Computer Science & Education 978-1-4244-3521-0/09/$25.00 ©2009 IEEE

2.L.Surya Prasanthi, R.Kiran Kumar,” ID3 and Its Applications in Generation of Decision Trees across Various Domains- Survey” Research Scholar, Department of Computer Science, Krishna University, Machilipatnam, India.

3. <https://ashutoshtripathi.com/2021/10/11/decision-tree-for-regression-models-in-machine-learning/>

4. <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>

5. <https://ashutoshtripathi.com/2021/10/11/decision-tree-for-regression-models-in-machine-learning/>

6. <https://towardsdatascience.com/decision-trees-for-classification-id3-algorithm-explained-89df76e72df1>

7. <https://chat.openai.com/chat/cd4fac4d-2693-47e2-a9ec-dfb8b19412b5>

8.<https://www.saedsayad.com/decision_tree_reg.htm#:~:text=The%20ID3%20algorithm%20can%20be,Gain%20with%20Standard%20Deviation%20Reduction.&text=A%20decision%20tree%20is%20built,with%20similar%20values%20(homogenous)>