Foreground/Background Pixel Classification in Brain Scans Using Random Forest

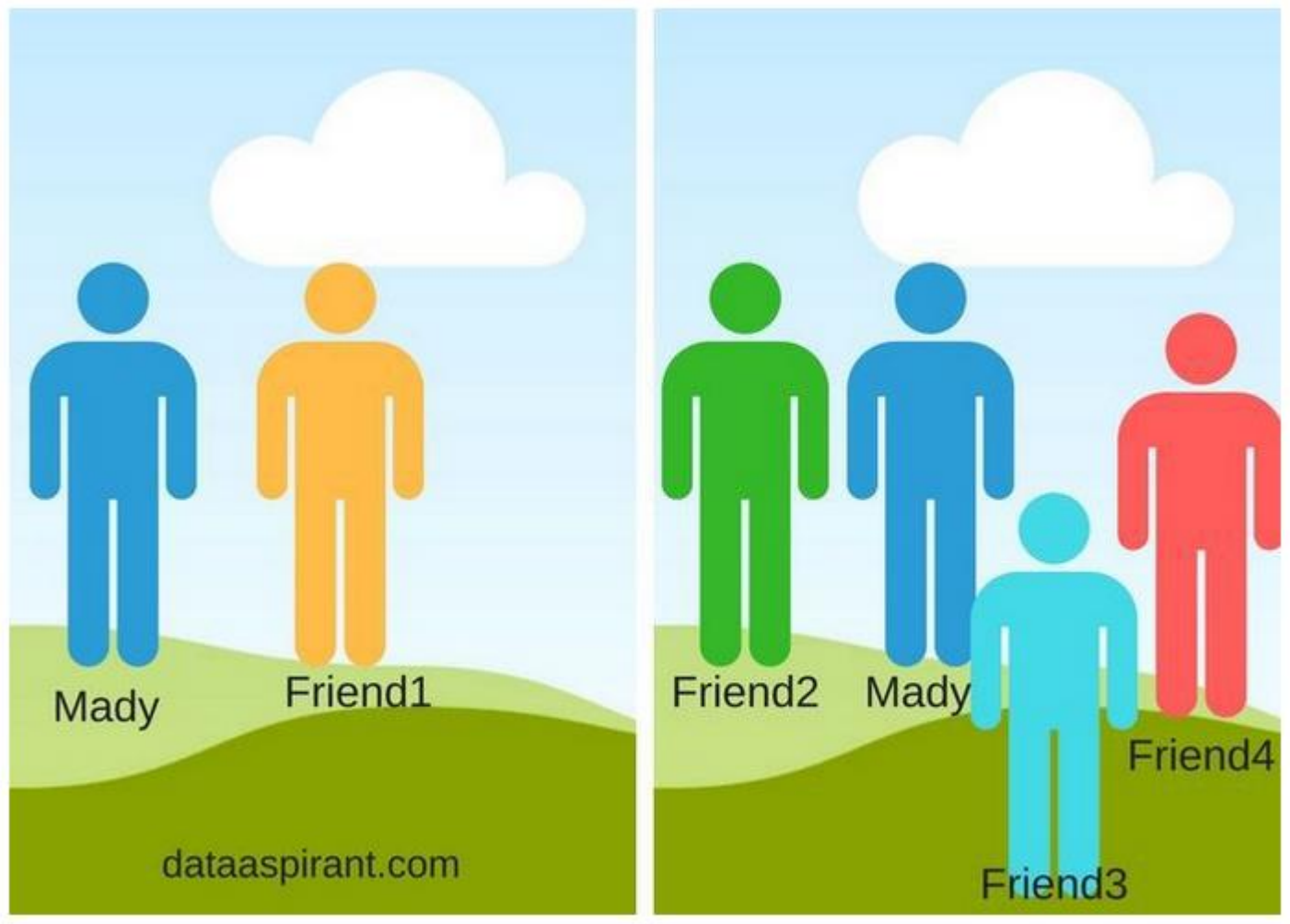
Presented By:
Shabbir Saifee
Shubham Deb

What is Random Forest?

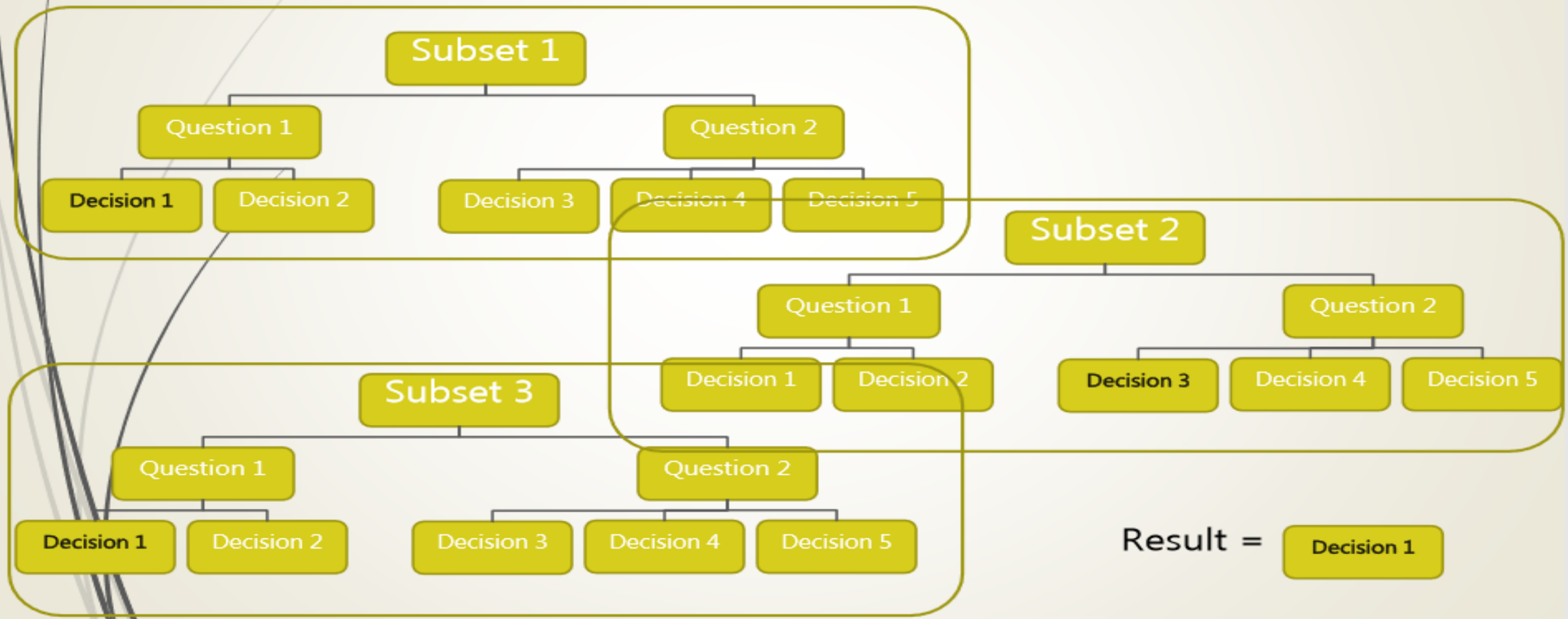
A random forest is an example of an ensemble, which is a combination of predictions from different models. In an ensemble, predictions could be combined either by majority-voting or by taking averages.

In the random forest classifier, the **higher the number** of trees in the forest gives **the high accuracy** results.

Random forest algorithm real life example



An Example of Random Forests



Why Random Forest?

As a random forest is an ensemble of multiple decision trees, it leverages "wisdom of the crowd", and is often more accurate than any individual decision tree.

This is because each individual model has its own strengths and weakness in predicting certain outputs.

Models included in the ensemble must not make the same kind of mistakes. In other words, the models must be *uncorrelated*. This is achieved via a technique called ***Bagging***.

Takes advantage of parallel computing by generating more number of decision trees. Thus, increasing the probability of getting correct prediction.

Practical Consideration

Number Of Trees: The more the merrier.

- Increasing the number of trees will decrease the variance in predictions.
- Training time increases roughly linearly in the number of trees

Depth of the Tree: Increasing the depth makes the model more expressive and powerful. However, deep trees take longer to train and are also more prone to overfitting.

Observations

- On increasing the number of trees to 50 from 25(keeping the other parameters constant), the accuracy decreased from 99.76% from 99.68%.
- The running time increased from 20 minutes to 45 minutes on increasing number of trees.
- On increasing the depth of the tree from 4 to 8 (keeping the other parameters constant), running time increased by 20 minutes but the accuracy was hovering around 99.6%.. Hence to reduce training time we kept the depth as 4.

Continue…

- Studies have shown that the choice of impurity measure has little effect on the performance of decision tree classification algorithms. This is because many impurity measures are quite consistent with each other

- We used Gini impurity than entropy, as it doesn't require to compute logarithmic functions (which is calculated in entropy), which are computationally intensive.

Results

Depth & No. of Trees	Running Time (minutes)	Accuracy
4 &10	26.45	99.76%
10 &10	38.23	99.6%
4 & 50	42.44	99.68%

Questions & Answers

Thank You