

# Glioma Classification

Eric L. Dias

## Glioma Classification Using Decision Tree

In this project, I am attempting to classify glioma diagnoses via mutated genes using a Decision Tree.

```
g <- read.csv("glioma_data.csv")
```

There are two types of diagnoses here. The first is LGG(Lower Grade Glioma), and the second is GBM(Glioblastoma Multiforme). The task is to develop methods to classify patients as either LGG or GBM.

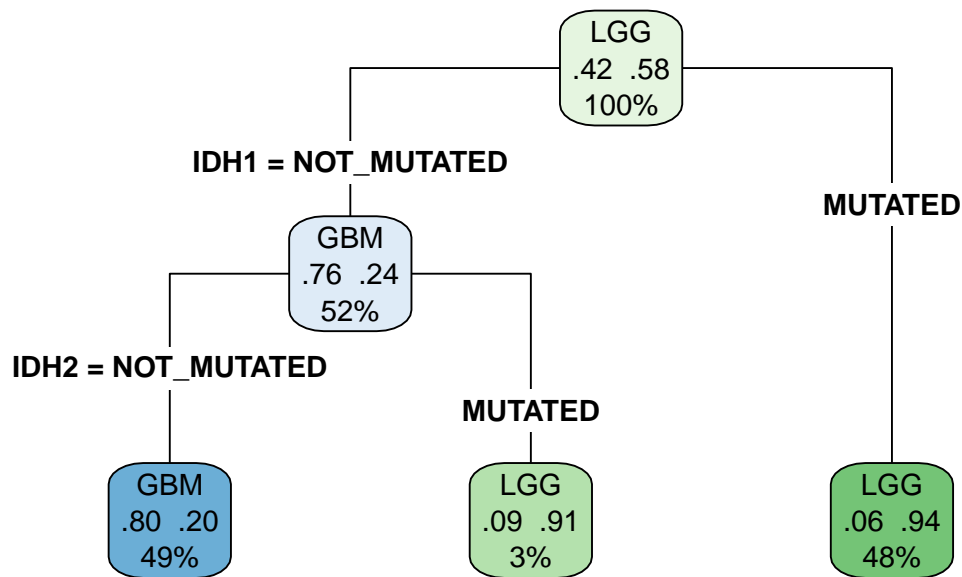
## Using Decision Tree to Determine Significant Features

```
library(rpart)
library(rpart.plot)

g <- g[, -c(2, 3, 5, 6)]#want to classify only on gene mutations

tree_model <- rpart(Grade ~ ., data = g, method = "class")

rpart.plot(tree_model, type = 4, extra = 104, box.palette = "auto", compress = FALSE)
```



The significant features are IDH1 and IDH2.

### Classification Using Decision Tree with 10-Fold Cross Validation

```
df <- g[c("Grade", "IDH1", "IDH2")]
```

```
acc_vec <- rep(0, 10)
```

```
for(i in 1:10){
  train_index <- sample(seq_len(nrow(df)), size = 0.8 * nrow(df))
  train_data <- df[train_index, ]
  test_data <- df[-train_index, ]
```

```
fit <- rpart(Grade ~ ., data=train_data, method="class")
predicted <- predict(fit, newdata=test_data, type="class")
```

```
cm <- table(Predicted = predicted, Actual = test_data$Grade)
```

```
#Let GBM be positive and LGG be Negative.
```

```
#Thus focus must be on Accuracy since we care about predicting both
```

```
TP<- cm["GBM", "GBM"]
TN <- cm["LGG", "LGG"]
FP <- cm["GBM", "LGG"]
FN <- cm["LGG", "GBM"]

acc_vec[i] <- (TP+TN)/(TP+TN+FP+FN)
}
```

```
cat("Average Accuracy: ", mean(acc_vec))
```

Average Accuracy: 0.8774566

### Conclusion:

By first fitting a Decision Tree on the entire dataset, it is possible to determine the significant features, in this case IDH1 and IDH2. Since, the two classes are both different types of glioma: (LGG, and GBM), I had to use accuracy as my metric. I was able to get an accuracy on 0.87 in predicting LGG and GBM.