# Decision Tree

We decided to build a decision tree for predicting the students' grades since it is easily interpretable, the data has nonlinear interactions between variables, and the decision tree model is easy to train. We still used the features selected from our feature selection section: failures, mom's education, the desire to pursue higher education, etc. The decision tree models ultimately performed better on the Portuguese dataset than the Math dataset.

## Evaluation

To evaluate the models for the decision trees, we used **Mean Squared Error (MSE)** and **$R^2$** scores. Additionally, we included a **regularization risk**, which penalizes over-complex trees. Below is the formula for calculating the regularized risk:

$$R_\alpha(T) = R(T) + \alpha|T|$$

## Results

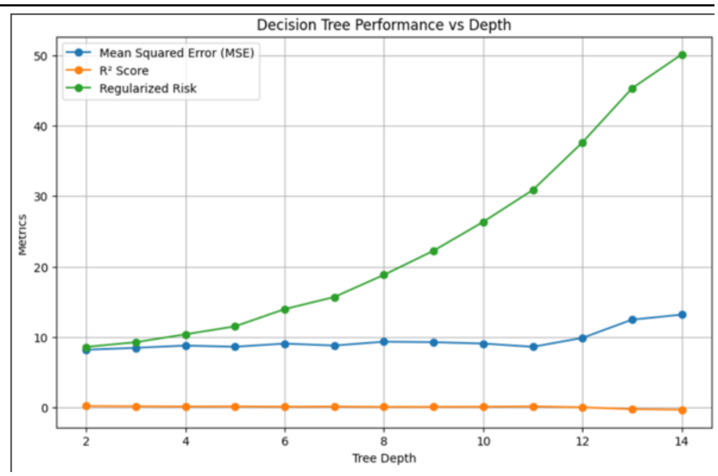
We adjusted the depth of the decision trees since larger depths mean a higher chance of overfitting. Here are the results:

| Depth | MSE | $R^2$ | Regularized Risk |
|---|---|---|---|
| 2 | 8.172335 | 0.161959 | 8.572335 |
| 3 | 8.442874 | 0.134217 | 9.242874 |
| 4 | 8.764858 | 0.101198 | 10.364858 |
| 5 | 8.595407 | 0.118575 | 11.495407 |
| 6 | 9.046504 | 0.072317 | 13.946504 |

The clear winner here is a tree with **3 layers**, which has the lowest MSE and an $R^2$ score closest to 1.

## Decision Tree Visualization

```
|--- failures <= 0.50
|   |--- higher_yes <= 0.50
|   |   |--- Fedu <= 3.50
|   |   |   |--- goout <= 3.50
|   |   |   |   |--- romantic_yes <= 0.50
|   |   |   |   |   |--- value: [12.00]
|   |   |   |   |--- romantic_yes >  0.50
|   |   |   |   |   |--- value: [10.29]
```




### Conclusion Decision Tree

From the analysis, we found that the decision tree with **3 layers** performed the best. This balance between model complexity and predictive performance ensures minimal overfitting while maintaining high interpretability. In addition to that, there is a clear trend of regularization risk going up as the depth of the tree goes up and the model is slowly overfitting, gaining a larger and larger mse as we add more layers to the decision tree. The best Mse and R squared for decision trees is 8.44 and 0.134