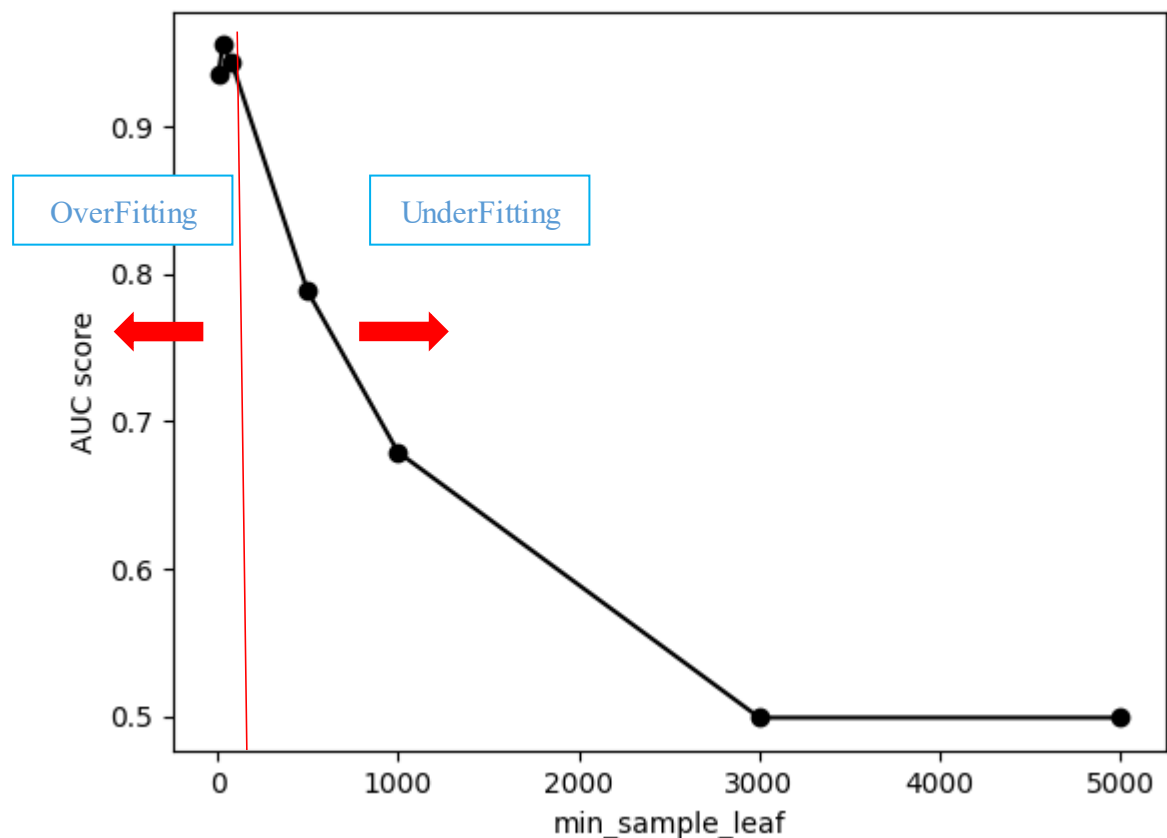


## DATA1 banana ID:1460

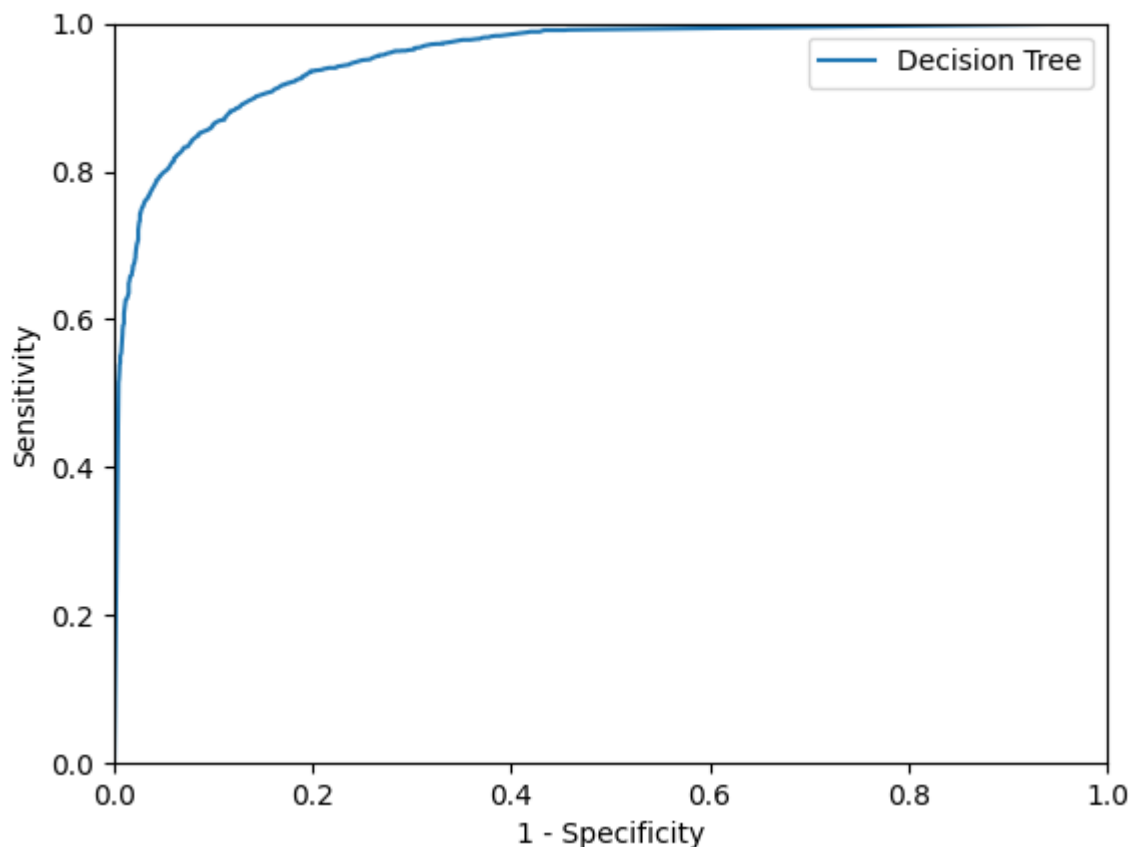
1.A brief description of the dataset (what is the task, what are the features and the target) :

The dataset consists of instances shaped into two distinct banana-like clusters, each labeled either -1 or 1. The primary task is to classify these instances into their respective classes based on their geometric shape. The features, labeled At1 and At2, represent coordinates on the x and y axes respectively. The target is the class label, which identifies to which banana shape an instance belongs. This dataset is typically used to test classification algorithms' ability to handle non-linearly separable data.

2.Graph of AUC values, describing the regions of overfitting and underfitting :



3. ROC curve



#### 4. Discussion of the Result

The decision tree model performs well at low `min_sample_leaf` values but loses predictive accuracy as this parameter increases. This suggests that a finer granularity benefits the model's performance.

The ROC curve supports this conclusion, indicating that the model is quite capable of discriminating between the classes effectively when tuned with appropriate parameters. It is crucial to balance the `min_sample_leaf` to prevent the model from becoming too complex (overfitting) or too general (underfitting).

## DATA2 Diabetes130US ID:45022

1. A brief description of the dataset (what is the task, what are the features and the target) :

### Features in the Dataset:

**time\_in\_hospital:** The duration of a patient's hospital stay (in days).

**num\_lab\_procedures:** Number of lab tests performed during the encounter.

**num\_medications:** Number of distinct medications prescribed during the encounter.

**number\_outpatient:** Number of outpatient visits by the patient in the year preceding the hospitalization.

**number\_emergency:** Number of emergency visits by the patient in the year preceding the hospitalization.

**number\_inpatient:** Number of inpatient visits by the patient in the year preceding the hospitalization.

**number\_diagnoses:** Total number of diagnoses entered to the system during the encounter.

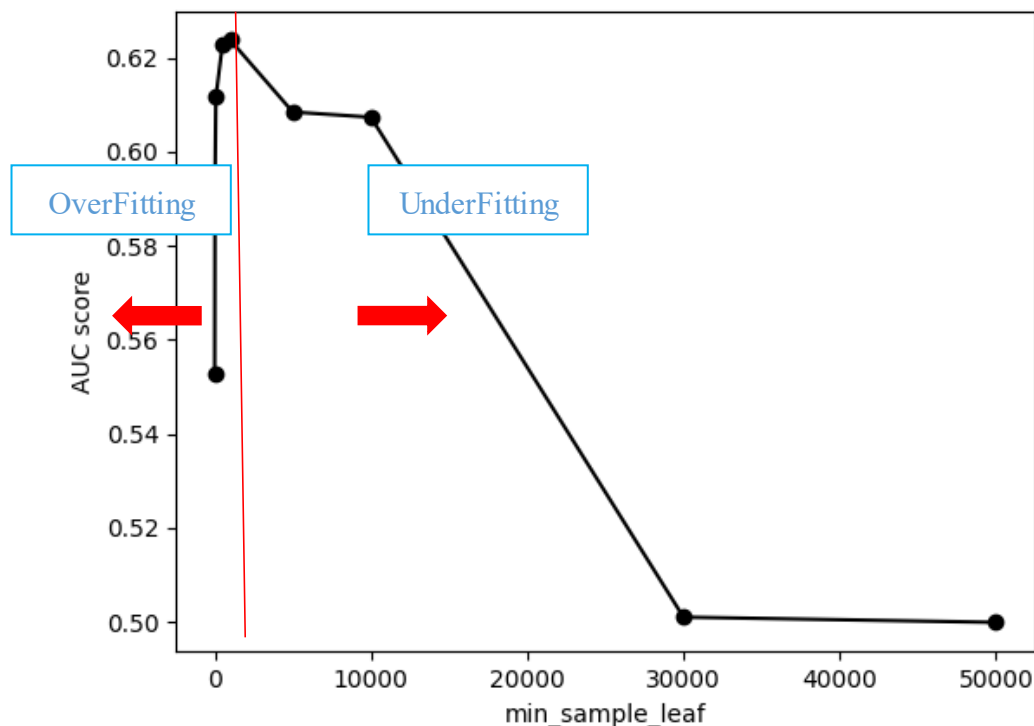
**Target Variable:**

**readmitted:** This is the target variable, indicating whether the patient was readmitted to the hospital after their initial discharge. It has two categories:

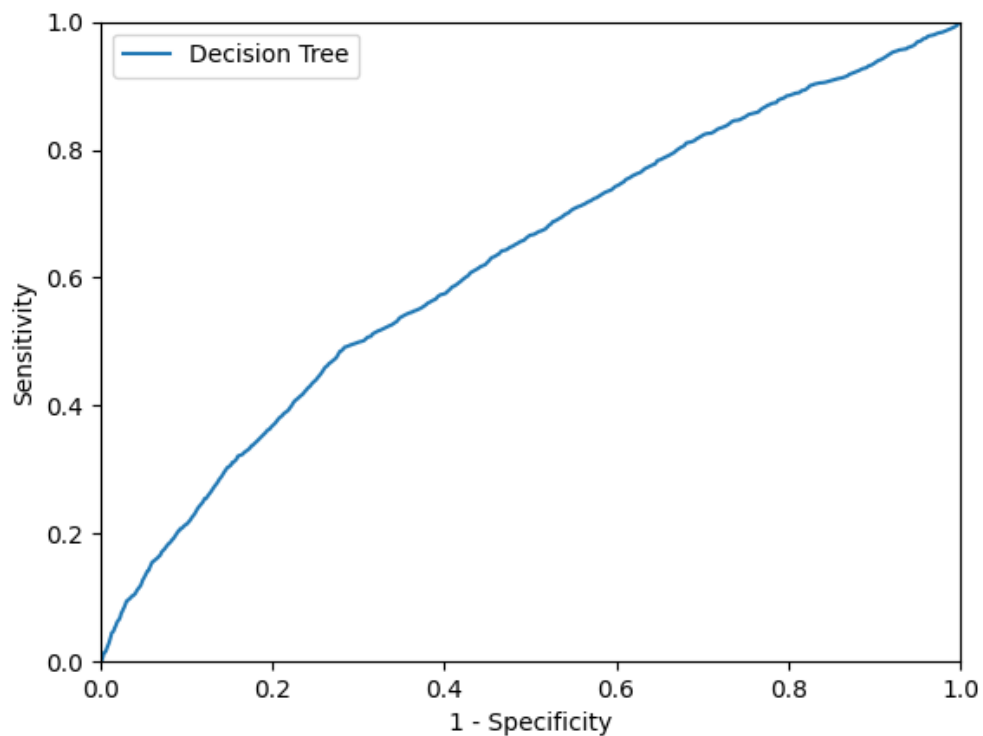
'0': The patient was not readmitted.

'1': The patient was readmitted.

2. Graph of AUC values, describing the regions of overfitting and underfitting :



3. ROC curve



#### 4. Discussion of the Result

The first graph illustrates how the AUC score of a model varies with different values for the `min_samples_leaf` parameter in a decision tree classifier. It shows that a smaller `min_samples_leaf` results in higher AUC scores, indicating better model performance at smaller leaf sizes. However, the performance decreases sharply as the `min_samples_leaf` value increases beyond 10,000.

The second graph is a ROC curve for the decision tree model. This curve shows the trade-off between sensitivity (true positive rate) and 1-specificity (false positive rate). The gradual curve towards the top right corner suggests that the model has a fair degree of discriminative ability, although it is not perfect. Ideally, a ROC curve closer to the top left corner indicates a more effective model in distinguishing between classes.