Classification of Irrigated/Rainfed In Madison Country using DecisionTreeClassifier

In this project, we are trying to solve the problem of classifying the fields as irrigated or rainfed using DecisionTreeClassifier.

Libraries used:

1. Numpy
2. Pandas
3. Matplotlib for visualization and plotting

The classification problem was solved following the steps given below:

1. Data Preprocessing and Cleaning
2. Split data into 3 chunks using stratified sampling
3. Handle skewness using smote algorithm in the training set
4. Apply decision tree algorithm
5. Model selection using validation set and accuracy-depth plot
6. Evaluation of result on test dataset.

1. Data Preprocessing and Cleaning:

The following steps were carried out for data preprocessing and cleaning:
i. Missing values were handled using linear interpolation.
ii. The data was standardized using z-score normalization.
iii. PCA was used for dimensionality reduction. (As using principal components with more than 95% resulted in overfitting from even a small decision tree, 91% variance were preserved while performing the dimensionality reduction with only 10 principal components.)
iv. As the dataset was imbalanced (containing {1: 1288, 0: 5053}), it was handled using an oversampling algorithm: Smote on training set.

2. Split data into 3 chunks using stratified sampling

After data preprocessing and cleaning, the data were splitted into three chunks: train, validation and test containing 70%, 20% and 10% of total data respectively using stratified sampling. Thereafter, the data were shuffled in each of these sets.

3. Apply decision tree algorithm:

```
import numpy as np
class DecisionTree:
	"""
	A class to construct decision tree
```

```
    Attributes
    ----------
        min_sample_size: int
            Minimum sample size required to split the data
        max_depth: int
            Maximum possible depth
        mode: string
            To compute "gini" or "entropy" as a measure of impurity
    """

    def __init__(self, min_sample_size=2, max_depth=1000, mode="gini"):
        self.min_sample_size = min_sample_size
        self.max_depth = max_depth
        self.mode = mode

    def check_purity(self, data):
        """Check if a data contains only a single class label.
        Parameters
        ----------
            data: DataFrame
                Dataset with features and target
        """

        if len(np.unique(data[:, -1])) == 1:
            return True
        else:
            return False

    def classify_data(self, data):
        """Classify data into classes based on the maximum occurence
        Parameters
        ----------
            data: DataFrame
                Dataset with features and target
        Returns
        --------
            class label
        """

        unique_classes, count_unique_classes = np.unique(data[:, -1],
return_counts = True)
        return unique_classes[count_unique_classes.argmax()]

    def get_potential_splits(self, data):
        """Get all the unique value of a particular column of a
        dataframe
        Parameters
        ----------
            data: DataFrame
                Dataset with features and target
        Returns
        -------
            Dictionary of potential splits
        """
```

```
        potential_splits = {}
        _, n_columns = data.shape
        for column in list(range(n_columns -1)):
            values = data[:, column]
            unique_values = np.unique(values)
            if len(unique_values) == 1:
                potential_splits[column] = unique_values
            else:
                potential_splits[column] = []
                for i in range(len(unique_values)):
                    if i != 0:
                        current_value = unique_values[i]
                        previous_value = unique_values[i - 1]
                        potential_splits[column].append((current_value
+ previous_value) / 2)
        return potential_splits

    def split_data(self, data, split_column, split_value):
        """Split data in left and right branch

        Parameters
        ----------
            data: DataFrame
                Dataset with features and target
            split_column: str
                Column to split
            split_value: float
                Value on the basis of which to split
        """

        split_column_values = data[:, split_column]
        return data[split_column_values <= split_value], data[split_column_values
> split_value]

    def calculate_impurity_parent(self, data):
        """Calculate impurity of a parent class

        Parameters
        ----------
            data: DataFrame
                Dataset with features and target
        Returns
        -------
            Impurity of a parent class (for both gini and entropy)
        """

        _, unique_classes = np.unique(data[:, -1], return_counts=True)
        probability = unique_classes/unique_classes.sum()
        if self.mode == "entropy":
            return np.sum(probability * -np.log2(probability))
        else:
            return 1- np.sum(np.square(probability))
```

```
	def calculate_impurity_children(self, data_below, data_above):
		"""Calculate impurity of a parent class


		Parameters
		----------
			data: DataFrame
				Dataset with features and target
		Returns
		--------
			Impurity of a child class (for both gini and entropy)
		"""
		prob_data_below = len(data_below) / (len(data_below) + len(data_above))
		prob_data_above = len(data_above) / (len(data_below) + len(data_above))
		return prob_data_below * self.calculate_impurity_parent(data_below) +
prob_data_above * self.calculate_impurity_parent(data_above)

	def calculate_information_gain(self, data, data_below, data_above):
		"""Calculate information gain by subtracting the impurity of a parent
class
		with that of a child class.

		Parameters
		----------
			data: DataFrame
				Dataset with features and target
			data_below: DataFrame
				Child in the left position of tree
			data_above: DataFrame
				Child in the right position of tree
		Returns
		--------
			Information gain
		"""
		return self.calculate_impurity_parent(data) -
self.calculate_impurity_children(data_below, data_above)

	def determine_best_split(self, data, potential_splits):
		"""Determine best split column and split value based on information gain
		Parameters
		----------
			data: DataFrame
				Dataset with features and target
			potential_splits: dict
				Dictionary of all potential splits
		Returns
		-------
			best_split_column, best_split_value
		"""

		max_info_gain = -float("inf")
		best_split_column = 0
		best_split_value = 0
		for split_column in potential_splits:
			for split_value in potential_splits[split_column]:
```

```
                        data_below, data_above = self.split_data(data,
split_column, split_value)
                        current_info_gain = self.calculate_information_gain(data,
data_below, data_above)
                        if current_info_gain >= max_info_gain:
                            max_info_gain = current_info_gain
                            best_split_column = split_column
                            best_split_value = split_value
            return best_split_column, best_split_value

    def build_decision_tree(self, dataframe, current_depth=0):
        """Build decision tree by checking purity of data, calculating potential
splits and
        determining best split

        Parameters:
        -----------
            dataframe: pd.DataFrame
                Dataset with features and target
            current_depth: int
                Depth of a decision tree

        Returns:
        --------
            Decision Tree
        """
        if current_depth == 0:
            global COLUMNS
            COLUMNS = dataframe.columns
            data = dataframe.values
        else:
            data = dataframe
        if self.check_purity(data) or len(data) < self.min_sample_size or
current_depth == self.max_depth:
            return self.classify_data(data)
        else:
            current_depth += 1
            potential_splits = self.get_potential_splits(data)
            split_column, split_value = self.determine_best_split(data,
potential_splits)
            data_below, data_above = self.split_data(data, split_column,
split_value)
            if len(data_below) == 0 or len(data_above) == 0:
                return self.classify_data(data)
            else:
                question = str(COLUMNS[split_column]) + " <= " +
str(split_value)
                decision_tree = {question: []}
                positive = self.build_decision_tree(data_below,
current_depth)
                negative = self.build_decision_tree(data_above,
current_depth)
                if positive == negative:
                    decision_tree = positive
                else:
```

```
                        decision_tree[question].append(positive)
                        decision_tree[question].append(negative)
                return decision_tree

def classify(self, sample, tree):
        """Classify the sample
        Parameters:
        -----------
        sample: pd.DataFrame
                Features
        tree: Decision tree which is to be used during classification

        Returns
        -------
                classification result
        """

        if not isinstance(tree, dict):
                return tree
        question = list(tree.keys())[0]
        attribute, value = question.split(" <= ")
        if sample[attribute] <= float(value):
                answer = tree[question][0]
        else:
                answer = tree[question][1]
        return self.classify(sample, answer)

def predictions(self, dataframe, tree):
        """Predict the result of entire dataframe
        Parameters
        -----------
        dataframe: pd.DataFrame
                Features
        tree: dict
                Decision tree which is to be used during classification

        Returns
        -------
                prediction result
        """

        return dataframe.apply(self.classify, axis = 1, args = (tree,))

def calculate_accuracy(self, predicted, actual):
        """Calculates Accuracy
        Parameters
        ----------
                predicted: numpy.ndarray
                        Prediction result
                acutal: numpy.ndarray
                        Actual result
        """
        correct = predicted == actual
        return correct.mean()
```

```
def calculate_error(self, predicted, actual):
        """Calculates Error
        Parameters
        ----------
                predicted: numpy.ndarray
                        Prediction result
                acutal: numpy.ndarray
                        Actual result
        """
        incorrect = predicted != actual
        return incorrect.mean()
```

Evaluation of Result:

While applying the above decision tree algorithm, we obtained the following result on different values of maximum depth:

```
maxDepth = 1: accTest = 61.04%, accTrain = 61.99%, accVal = 61.75%, buildTime = 68.03s
maxDepth = 2: accTest = 64.35%, accTrain = 65.40%, accVal = 65.30%, buildTime = 119.99s
maxDepth = 3: accTest = 66.09%, accTrain = 68.33%, accVal = 65.77%, buildTime = 163.53s
maxDepth = 4: accTest = 70.50%, accTrain = 71.59%, accVal = 67.51%, buildTime = 189.19s
maxDepth = 5: accTest = 70.19%, accTrain = 73.21%, accVal = 68.69%, buildTime = 219.62s
maxDepth = 6: accTest = 72.56%, accTrain = 75.38%, accVal = 67.74%, buildTime = 233.68s
maxDepth = 7: accTest = 75.08%, accTrain = 79.34%, accVal = 68.69%, buildTime = 245.81s
maxDepth = 8: accTest = 76.97%, accTrain = 82.27%, accVal = 69.32%, buildTime = 240.14s
maxDepth = 9: accTest = 77.13%, accTrain = 85.02%, accVal = 69.16%, buildTime = 251.42s
maxDepth = 10: accTest = 79.97%, accTrain = 87.44%, accVal = 69.32%, buildTime = 268.94s
maxDepth = 11: accTest = 80.91%, accTrain = 89.91%, accVal = 69.24%, buildTime = 269.37s
maxDepth = 12: accTest = 80.76%, accTrain = 92.09%, accVal = 68.14%, buildTime = 289.53s
maxDepth = 13: accTest = 82.33%, accTrain = 94.22%, accVal = 67.74%, buildTime = 284.75s
maxDepth = 14: accTest = 84.07%, accTrain = 95.92%, accVal = 68.45%, buildTime = 292.29s
maxDepth = 15: accTest = 85.02%, accTrain = 97.39%, accVal = 68.14%, buildTime = 300.45s
maxDepth = 16: accTest = 86.28%, accTrain = 98.42%, accVal = 68.22%, buildTime = 282.71s
maxDepth = 17: accTest = 86.44%, accTrain = 99.04%, accVal = 68.22%, buildTime = 299.08s
maxDepth = 18: accTest = 86.59%, accTrain = 99.50%, accVal = 68.22%, buildTime = 301.04s
```

```
maxDepth = 19: accTest = 86.28%, accTrain = 99.76%, accVal = 68.14%, buildTime =
289.62s
maxDepth = 20: accTest = 86.59%, accTrain = 99.91%, accVal = 67.98%, buildTime =
293.78s
maxDepth = 21: accTest = 86.75%, accTrain = 99.96%, accVal = 67.82%, buildTime =
270.45s
maxDepth = 22: accTest = 86.75%, accTrain = 99.99%, accVal = 67.82%, buildTime =
252.66s
maxDepth = 23: accTest = 86.75%, accTrain = 100.00%, accVal = 67.82%, buildTime =
277.94s
```

Then, the accuracy-depth and error-depth graphs were plotted from the results obtained:

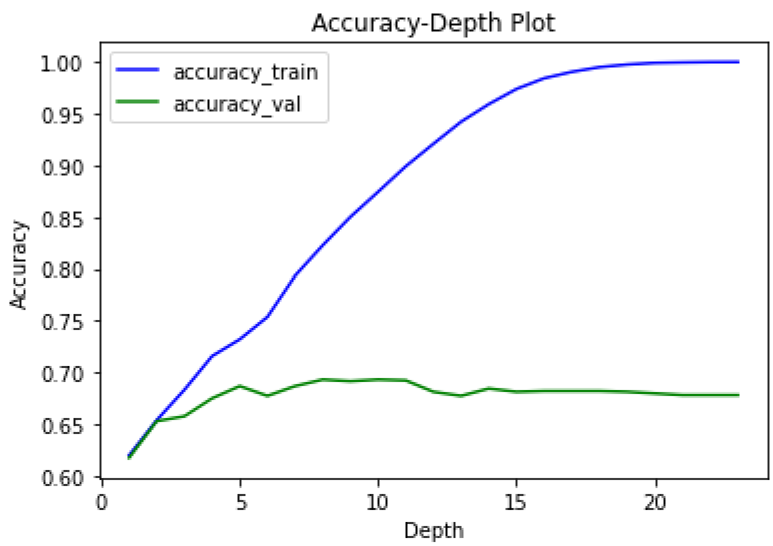

Figure 1a. Plot between accuracy and depth

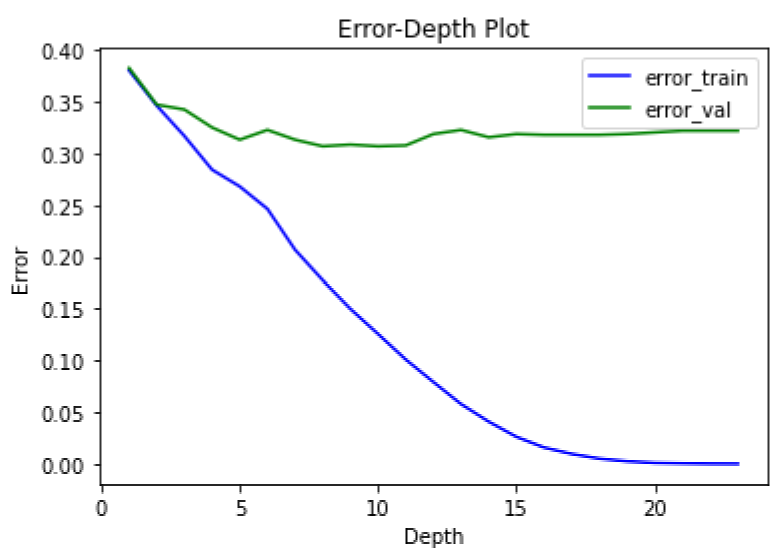

Figure 1b. Plot between error and depth

Therefore, visualizing the training plot, max-depth=15 seems to be the correct model whereas while looking at the validation plot, max-depth=5 seems to be the correct model.

A 2-dimensional decision tree boundary plot around max-depth=5 and max-depth=15 was plotted to help select the best model.

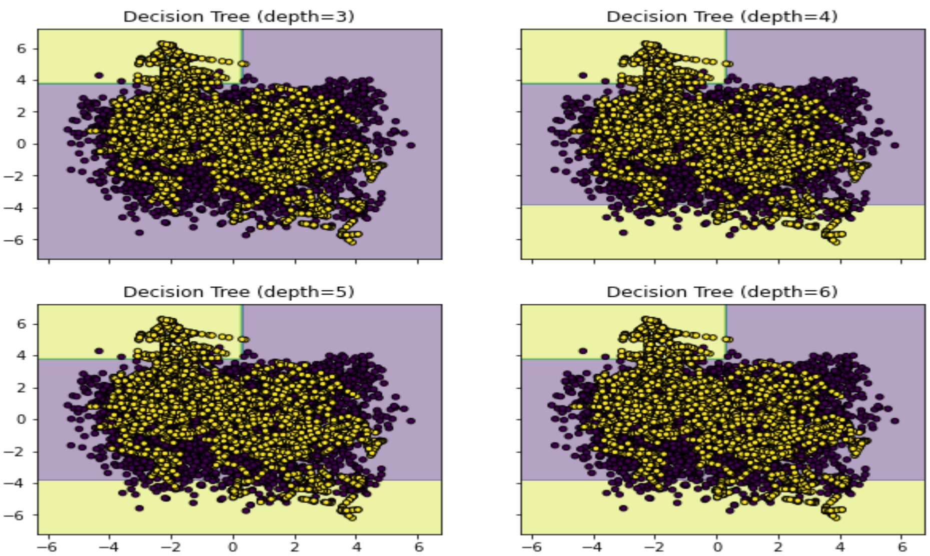

Figure 2a. Decision boundary for max-depth (3,4,5 and 6) in training set

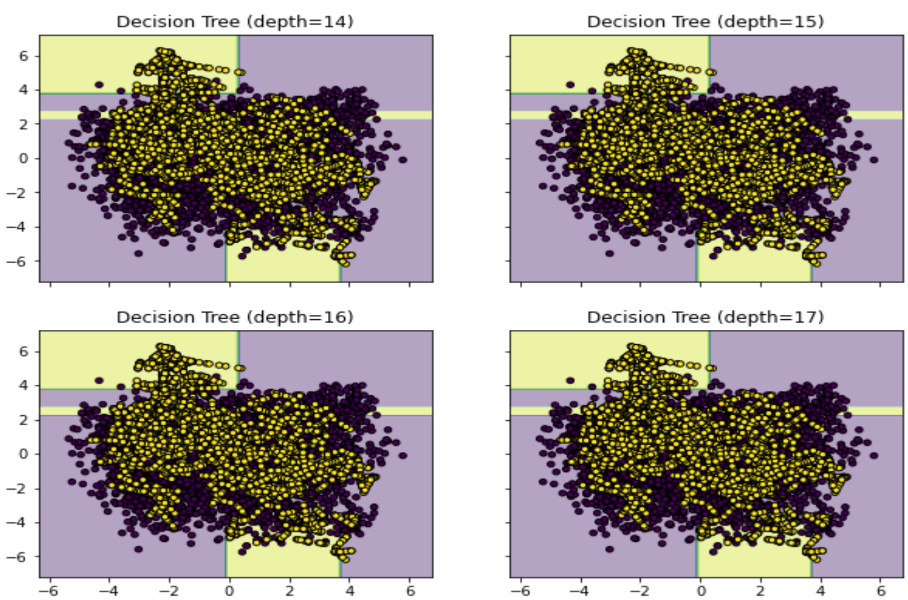

Figure 2b. Decision boundary for max-depth (14,15,16 and 17) in training set

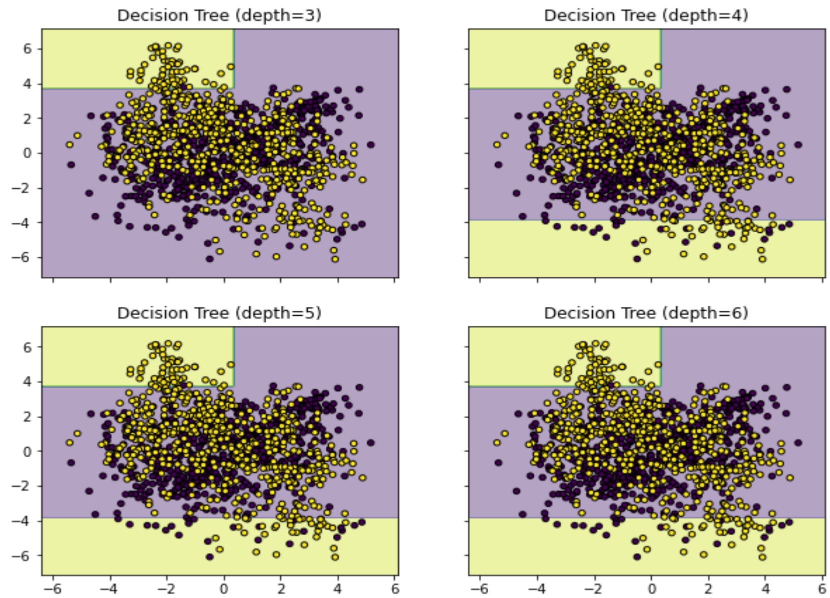

Figure 2c. Decision boundary for max-depth (3, 4, 5 and 6) in validation set

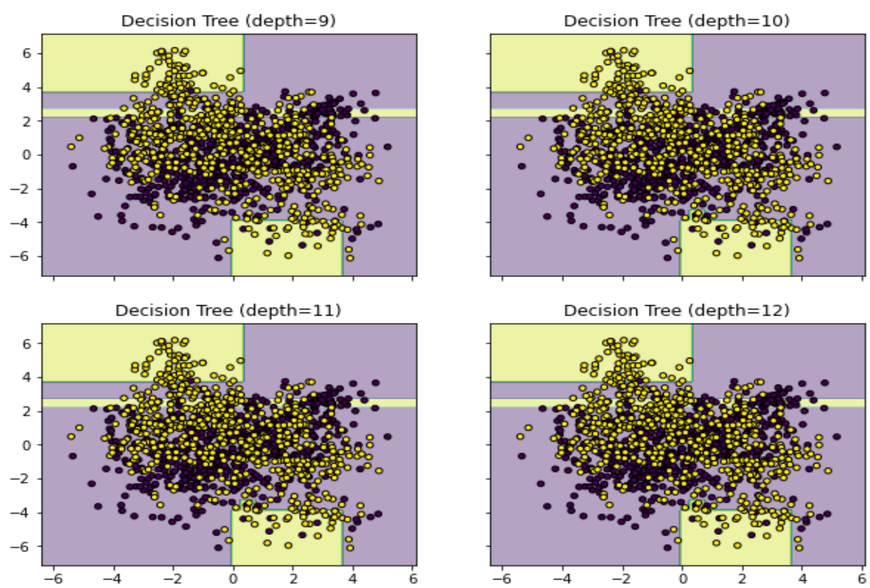

Figure 2d. Decision boundary for max-depth (9,10,11 and 12) in validation set

Pruning the decision tree:

For solving the problem of overfitting, we also pruned the decision tree trained at depth=5 and depth=15. The decision boundaries obtained are as follows:

When depth=5:

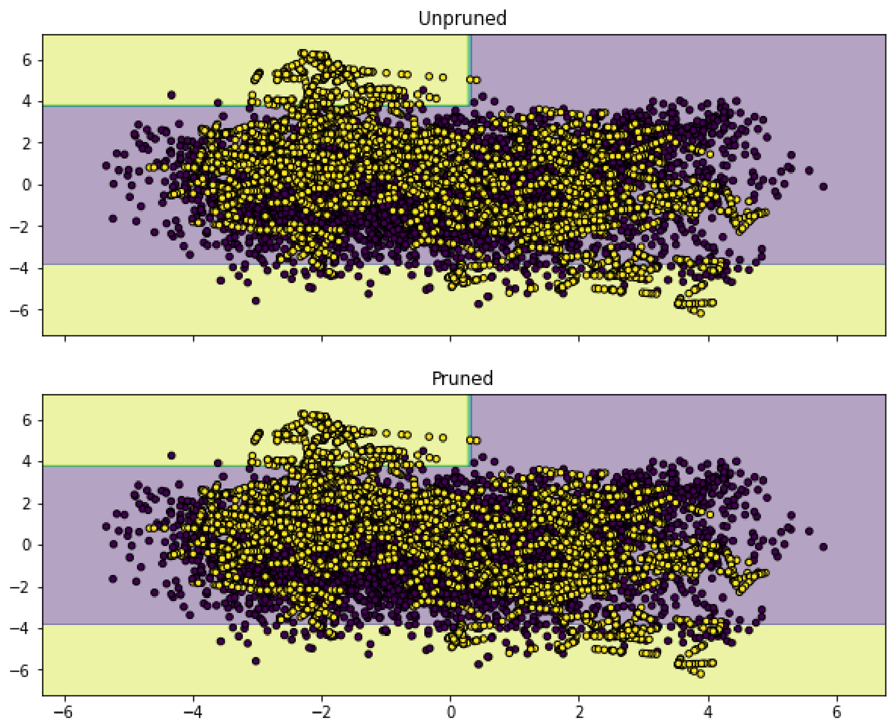

Figure 3a. Effect of pruning on training dataset when max_depth = 5

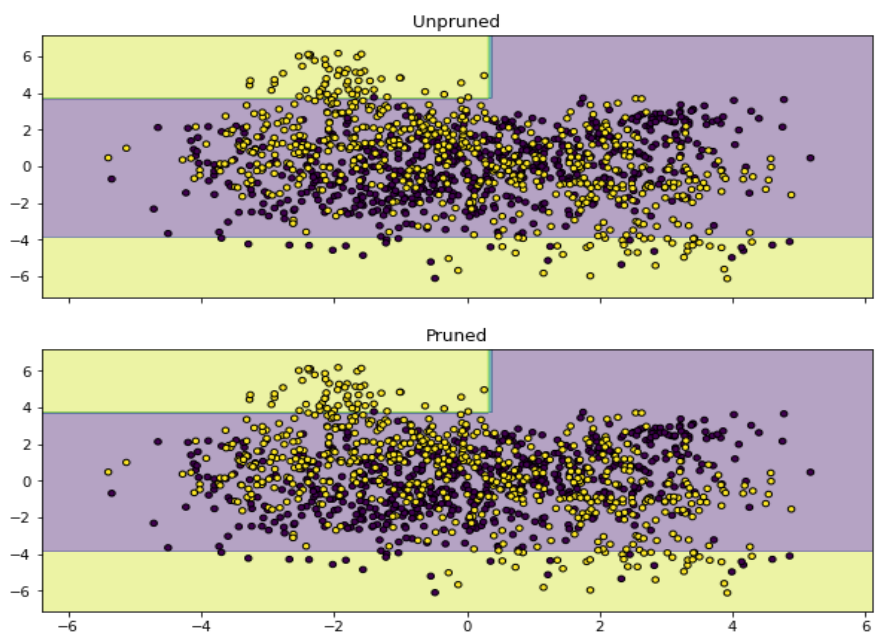

Figure 3b. Effect of pruning on validation dataset when max_depth = 5

When depth = 15

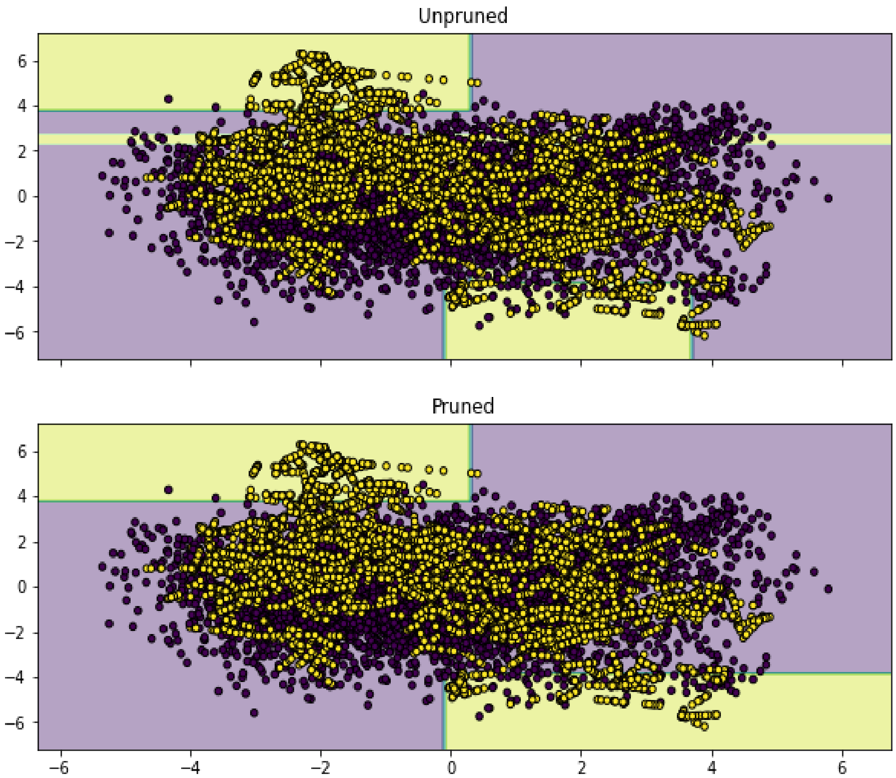

Figure 3c. Effect of pruning on training dataset when max_depth = 15

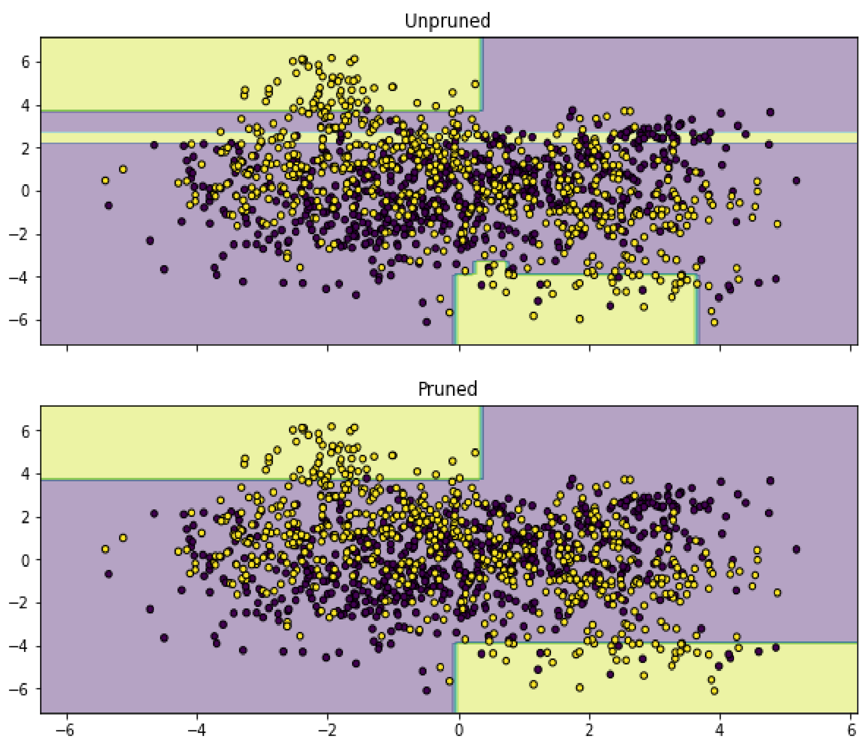

Figure 3d. Effect of pruning on training dataset when max_depth = 15

<u>Decision tree trained at max_depth = 5</u>

Unpruned	Pruned
{'PC3 <= 2.01698085272447': [{'PC2 <= 3.724320189213132': [{'PC8 <= 0.34713819851525285': [{'PC2 <= -3.830829144110229': [{'PC6 <= 0.9642426768212335': [1.0, 0.0]}, 0.0]}, {'PC6 <= 0.24935772115142918': [{'PC9 <= -0.006163800662947843': [1.0, 0.0]}, 1.0]}]}, {'PC1 <= 0.3448380525363134': [{'PC9 <= -0.7465826694517675': [0.0, {'PC10 <= -1.270883359710674': [0.0, 1.0]}]}, {'PC8 <= -0.9429530372139716': [1.0, 0.0]}]}]}, {'PC6 <= -0.5923209052091465': [{'PC7 <= 0.21233068981024117': [{'PC1 <= 0.5357596542629881': [{'PC8 <= 1.8917507396556816': [0.0, 1.0]}, {'PC9 <= 0.20795693271261803': [0.0, 1.0]}]}, {'PC9 <= 0.02811107358035988': [{'PC5 <= -1.1215649710916407': [0.0, 1.0]}, {'PC2 <= 0.9553639428688796': [0.0, 1.0]}]}]}, {'PC8 <= 0.19000952050089764': [{'PC3 <= 2.538896711131528': [1.0, {'PC5 <= -1.965029922021348': [0.0, 1.0]}]}, {'PC3 <= 3.1007115509230196': [{'PC7 <= 0.20041911249044375': [0.0, 1.0]}, {'PC5 <= 1.6874049755011886': [1.0, 0.0]}]}]}]}]}	{'PC3 <= 2.01698085272447': [{'PC2 <= 3.724320189213132': [{'PC8 <= 0.34713819851525285': [{'PC2 <= -3.830829144110229': [{'PC6 <= 0.9642426768212335': [1.0, 0.0]}, 0.0]}, {'PC6 <= 0.24935772115142918': [{'PC9 <= -0.006163800662947843': [1.0, 0.0]}, 1.0]}]}, {'PC1 <= 0.3448380525363134': [1, 0]}]}, {'PC6 <= -0.5923209052091465': [{'PC7 <= 0.21233068981024117': [{'PC1 <= 0.5357596542629881': [{'PC8 <= 1.8917507396556816': [0.0, 1.0]}, {'PC9 <= 0.20795693271261803': [0.0, 1.0]}]}, {'PC9 <= 0.02811107358035988': [{'PC5 <= -1.1215649710916407': [0.0, 1.0]}, {'PC2 <= 0.9553639428688796': [0.0, 1.0]}]}]}, {'PC8 <= 0.19000952050089764': [1, {'PC3 <= 3.1007115509230196': [{'PC7 <= 0.20041911249044375': [0.0, 1.0]}, 1]}]}]}]}

<u>Decision tree trained at max_depth = 15</u>

Unpruned	Pruned
{'PC3 <= 2.01698085272447': [{'PC2 <= 3.724320189213132': [{'PC8 <= 0.34713819851525285': [{'PC2 <= -3.830829144110229': [{'PC6 <= 0.9642426768212335': [{'PC7 <=	{'PC3 <= 2.01698085272447': [{'PC2 <= 3.724320189213132': [{'PC8 <= 0.34713819851525285': [{'PC2 <= -3.830829144110229': [{'PC6 <= 0.9642426768212335': [{'PC7 <=

```
0.19597315411066354': [{'PC1 <=
-0.07320763987337418': [{'PC7 <=
0.11430868848403934': [0.0,
                1.0]},
              {'PC10 <= 0.03822132422578686':
[{'PC3 <= -0.1891952718960509': [{'PC10 <=
0.010918690604387422': [0.0,
                    1.0]},
                  {'PC4 <=
2.6454286340273905': [{'PC1 <=
3.676693079184104': [{'PC5 <=
-1.921512009948625': [0.0,
                        {'PC2 <=
-4.009802435706431': [{'PC10 <=
-0.05203930168273435': [1.0,
                            {'PC10 <=
-0.03496330666029368': [0.0, 1.0]}]},
                          {'PC4 <=
-0.5013251802143533': [0.0,
                            {'PC6 <=
-0.6790370236839172': [0.0, 1.0]}]}]}]},
                      0.0]},
                    0.0]}]},
                {'PC8 <=
-0.5578957178255834': [1.0, 0.0]}]}]},
            {'PC8 <= -0.8714505392907765':
[1.0,
              {'PC3 <= 1.9028497414096885':
[0.0, 1.0]}]}]},
          {'PC2 <= -3.8428036468623206':
[0.0, 1.0]}]},
        {'PC2 <= -0.08690633927528024':
[{'PC1 <= 0.2758981575591521': [{'PC5 <=
1.4613139151014336': [{'PC10 <=
-0.9112841306868809': [{'PC7 <=
-0.1648923865801566': [0.0,
                  {'PC10 <=
-1.5681321571328095': [0.0,
                    {'PC5 <=
1.0772614897865957': [{'PC1 <=
-1.5030411679007827': [1.0,
                        {'PC1 <=
-0.2404482578214358': [{'PC10 <=
-0.9364330919774901': [0.0,
                            1.0]},
                          {'PC10 <=
-1.1695863534766315': [{'PC9 <=
1.0366663626633577': [0.0,
                              1.0]},
                            1.0]}]}]},
                      0.0]}]}]},
                {'PC6 <= 0.9970124977789625':
[{'PC3 <= 1.9286088731434838': [{'PC2 <=
-0.12413471140508514': [{'PC5 <=
1.0847540646783702': [{'PC6 <=
0.5228352327491523': [{'PC10 <=
-0.833695082454498': [{'PC9 <=
0.6186361813662937': [0.0,
                              1.0]},
                            0.0]},
```

```
0.19597315411066354': [{'PC1 <=
-0.07320763987337418': [0,
              {'PC10 <= 0.03822132422578686':
[1,
                {'PC8 <=
-0.5578957178255834': [1.0, 0.0]}]}]},
            {'PC8 <= -0.8714505392907765':
[1.0,
              {'PC3 <= 1.9028497414096885':
[0.0, 1.0]}]}]},
          0]},
        {'PC2 <= -0.08690633927528024':
[{'PC1 <= 0.2758981575591521': [{'PC5 <=
1.4613139151014336': [{'PC10 <=
-0.9112841306868809': [0,
                {'PC6 <= 0.9970124977789625':
[{'PC3 <= 1.9286088731434838': [{'PC2 <=
-0.12413471140508514': [{'PC5 <=
1.0847540646783702': [{'PC6 <=
0.5228352327491523': [{'PC10 <=
-0.833695082454498': [{'PC9 <=
0.6186361813662937': [0.0,
                              1.0]},
                            0.0]},
                          0]},
                        0]},
                      0]},
                    0]},
                  0]}]},
              {'PC9 <= -1.5365227220211755':
[1.0,
                {'PC5 <= 2.533623423506117':
[0,
                  {'PC4 <=
0.4978480881337619': [0.0,
                    {'PC7 <=
-0.270978107883008': [{'PC1 <=
-0.7854283939063409': [0.0,
                        1.0]},
                      0.0]}]}]}]}]},
            {'PC4 <= 1.0333497311248814':
[{'PC7 <= 0.9042197331611117': [{'PC7 <=
0.08817175453346884': [{'PC1 <=
0.7205825860471172': [{'PC7 <=
0.0010635509787784905': [{'PC10 <=
1.1540578718416075': [{'PC3 <=
1.60444976292392': [0,
                      {'PC7 <=
-0.464794996958687': [0.0,
                        {'PC10 <=
0.36660345589609': [1.0, 0.0]}]}]},
                    1.0]},
                  0]},
                {'PC6 <=
-1.2530138423315134': [0.0,
                  {'PC3 <=
0.942943692395166': [{'PC10 <=
0.04961878966139391': [{'PC8 <=
-0.26879662240072283': [{'PC7 <=
-0.03970085690273884': [0.0,
                            1.0]},
                          0]},
                        0]},
                      {'PC5 <=
3.4640865851636535': [1, 0.0]}]}]}]},
                {'PC5 <=
0.6737505410596957': [{'PC4 <=
-2.040266707315133': [0,
                      {'PC7 <=
0.13526115658296334': [{'PC7 <=
```

```
                          {'PC1 <=
-3.5802806975955295': [{'PC7 <=
0.7679849091151234': [0.0,
                              1.0]},
                            {'PC7 <=
-1.4606723191178785': [1.0, 0.0]}]}]},
                        {'PC10 <=
-0.49819223415752756': [{'PC3 <=
0.05465968805808203': [0.0,
                            {'PC9 <=
0.07853838637067326': [1.0, 0.0]}]},
                          0.0]}]},
                      {'PC10 <=
0.8757640647341589': [0.0, 1.0]}]},
                    {'PC5 <=
-1.1194333402621484': [0.0, 1.0]}]},
                  {'PC3 <=
0.4652804470137028': [{'PC2 <=
-2.6749363303878653': [{'PC1 <=
-0.3800000356157136': [0.0,
                        {'PC9 <=
-0.3367760697385835': [0.0, 1.0]}]},
                      0.0]},
                    {'PC1 <=
-0.2703034716215092': [{'PC1 <=
-0.7710761555351999': [{'PC5 <=
-1.0562915637804333': [1.0,
                          {'PC9 <=
-0.6508758071262197': [1.0, 0.0]}]},
                        {'PC2 <=
-0.8525759390979456': [1.0, 0.0]}]},
                      0.0]}]}]}]},
              {'PC9 <= -1.5365227220211755':
[1.0,
                {'PC5 <= 2.533623423506117':
[{'PC3 <= 0.7371023445184302': [{'PC5 <=
1.4653840117793253': [1.0,
                      0.0]},
                    {'PC5 <=
2.1318866022195095': [{'PC1 <=
-2.4488792645133186': [0.0,
                        {'PC1 <=
-1.7413003884818794': [1.0,
                          {'PC9 <=
0.6004504385306488': [0.0, 1.0]}]}]},
                      0.0]}]},
                  {'PC4 <=
0.4978480881337619': [0.0,
                    {'PC7 <=
-0.270978107883008': [{'PC1 <=
-0.7854283939063409': [0.0,
                        1.0]},
                      0.0]}]}]}]}]}]},
            {'PC4 <= 1.0333497311248814':
[{'PC7 <= 0.9042197331611117': [{'PC7 <=
0.08817175453346884': [{'PC1 <=
0.7205825860471172': [{'PC7 <=
0.0010635509787784905': [{'PC10 <=
1.1540578718416075': [{'PC3 <=
1.60444976292392': [{'PC2 <=
0.12702706544857606': [{'PC10 <=
-0.9723183325759405': [1.0,
                            {'PC5 <=
-2.1241389849751764': [1.0, 0.0]}]},
                          1.0]},
                        0.0]}]},
                    0]}]},
                0]},
              0]}]},
          {'PC5 <= -0.9218038037571261':
[{'PC2 <= 1.6932554902054897': [{'PC4 <=
-1.6204935507018048': [{'PC2 <=
0.25109331526462864': [{'PC8 <=
0.051406072850849444': [1,
                    0.0]},
                  0]},
                {'PC8 <=
-0.4546142529982261': [0,
                  {'PC3 <=
0.7872630893096046': [0,
                    {'PC4 <=
0.9138515569396832': [0, 1]}]}]}]},
              {'PC3 <= 0.8300279948471145':
[{'PC7 <= -0.6573975196954583': [0,
                  {'PC4 <=
1.9452595035562505': [{'PC3 <=
-1.252094860976261': [0.0,
                      {'PC8 <=
-1.1652951051286127': [1.0,
                        {'PC10 <=
0.15765411745835667': [{'PC4 <=
-0.5782815970593735': [0,
                            {'PC3 <=
0.7691992306673174': [1.0, 0.0]}]},
                          0]}]}]},
                    0.0]}]},
                1]}]},
            {'PC3 <= -3.3567552993283476':
[{'PC2 <= 0.46073789147014904': [0.0,
                {'PC7 <=
-0.0592749332304799': [0,
                  {'PC5 <=
2.651707337968143': [{'PC1 <=
-3.7060343640691586': [0.0,
                      1]},
                    0.0]}]}]},
              {'PC5 <= 0.74944814407033':
[{'PC3 <= 0.2652807795880745': [0,
                  {'PC2 <=
1.344745879252146': [0,
                    {'PC1 <=
2.025792094416495': [{'PC7 <=
0.4695112299781634': [0,
                        {'PC4 <=
4.553282419027422': [1, 0.0]}]},
                      0]}]}]},
                {'PC4 <= 1.218106743831214':
[{'PC7 <= -0.6749427324310695': [{'PC4 <=
-0.6263513339620621': [0.0,
                      {'PC6 <=
-1.1512181147478189': [0.0,
                        {'PC2 <=
0.1888089897194844': [0.0,
                          {'PC10 <=
0.27058585227704846': [1.0, 0.0]}]}]}]},
                    0]},
                  {'PC8 <=
-0.4521511880240696': [0,
                    {'PC6 <=
-1.593230288050811': [0.0,
```

```
-3.249639303247088': [{'PC8 <=
-0.17571683837998636': [0.0,
                              1.0]},
                            0.0]},
                          {'PC7 <=
-0.464794996958687': [0.0,
                            {'PC10 <=
0.36660345589609': [1.0, 0.0]}]}]},
                        1.0]},
                      {'PC8 <=
-0.20412359080756046': [0.0, 1.0]}]},
                    {'PC6 <=
-1.2530138423315134': [0.0,
                      {'PC3 <=
0.942943692395166': [{'PC10 <=
0.04961878966139391': [{'PC8 <=
-0.26879662240072283': [{'PC7 <=
-0.03970085690273884': [0.0,
                              1.0]},
                            {'PC5 <=
-2.6683451730585297': [1.0, 0.0]}]},
                          {'PC9 <=
-0.11866377013023795': [{'PC8 <=
-0.17438966497560637': [0.0,
                              1.0]},
                            {'PC6 <=
1.234853917410088': [0.0, 1.0]}]}]},
                        {'PC5 <=
3.4640865851636535': [{'PC1 <=
3.357266663592208': [{'PC2 <=
-0.19472181519902135': [1.0,
                              0.0]},
                            0.0]},
                          0.0]}]}]}]},
                  {'PC5 <=
0.6737505410596957': [{'PC4 <=
-2.040266707315133': [{'PC9 <=
0.12303899525635453': [{'PC10 <=
0.4717863793274999': [0.0,
                          1.0]},
                        1.0]},
                      {'PC7 <=
0.13526115658296334': [{'PC7 <=
0.12702706544857606': [{'PC10 <=
-0.9723183325759405': [1.0,
                            {'PC5 <=
-2.1241389849751764': [1.0, 0.0]}]},
                          1.0]},
                        0.0]}]},
                    {'PC8 <=
0.1056629418579166': [{'PC1 <=
2.6390626530007686': [{'PC10 <=
0.7076901117721082': [0.0,
                          {'PC10 <=
0.7433812752532543': [1.0, 0.0]}]},
                        {'PC10 <=
-0.14202751640293806': [0.0, 1.0]}]},
                      {'PC3 <=
-1.5171095054298978': [0.0, 1.0]}]}]}]},
                      {'PC3 <=
0.9007376404409113': [{'PC7 <=
-1.1220195569168347': [0.0,
                          {'PC9 <=
0.9163988801744124': [1, 0.0]}]},

0.0]}]}]}]}]}]}]}]}]}]},
      {'PC6 <= 0.24935772115142918': [{'PC9
<= -0.006163800662947843': [{'PC5 <=
0.8694395756851128': [{'PC2 <=
-3.4149067804908726': [{'PC4 <=
1.453419656620854': [1,
                0.0]},
              {'PC7 <= 1.131036307120978':
[{'PC4 <= 1.2044741244120192': [0,
                  {'PC3 <=
0.8371154102035026': [{'PC5 <=
-0.31352167486131305': [0.0,
                      {'PC3 <=
-0.5502713287454856': [1, 0.0]}]},
                    1]}]},
                1]}]},
            {'PC4 <= -1.2716039374846106':
[{'PC7 <= -1.2823932031241039': [1.0,
                0.0]},
              {'PC10 <= 0.8291164807825707':
[{'PC8 <= 1.4357823556989922': [{'PC2 <=
2.642037886058202': [{'PC9 <=
-0.18653885854110586': [{'PC10 <=
0.05893334601552237': [1,
                        0]},
                      {'PC2 <=
1.4805508322223657': [1.0, 0.0]}]},
                    0.0]},
                  0.0]},
                0.0]}]}]},
          {'PC8 <= 0.845351997187648': [0,
            {'PC1 <= 0.22252380398772548':
[{'PC5 <= -0.1646041865628528': [0,
                {'PC2 <= -1.090223389354784':
[0.0,
                  {'PC8 <=
0.9429492551761164': [0, 1]}]}]},
              0]}]}]},
        {'PC10 <= -0.5229065720498982':
[{'PC5 <= -0.6825654447731452': [0,
            {'PC8 <= 0.7711746643736423':
[{'PC7 <= 0.07076265475851237': [{'PC9 <=
0.881638053229489': [{'PC1 <=
-2.65190765647448': [{'PC10 <=
-0.6252388431405691': [1.0,
                      0.0]},
                    1]},
                  0.0]},
                {'PC4 <= 0.4651696692590329':
[0.0,
                  {'PC9 <=
0.29582023579621436': [{'PC10 <=
-0.9655026762666499': [0.0,
                      1]},
                    0.0]}]}]},
              1]}]},
          {'PC1 <= 2.0594016535599566':
[{'PC1 <= -1.6133090857814083': [{'PC1 <=
-4.000567138397632': [0.0,
                {'PC10 <=
0.4166380014026776': [{'PC9 <=
0.8647120586854768': [{'PC7 <=
0.9108370566012731': [{'PC4 <=
2.3440491617514647': [{'PC7 <=
```

```
                {'PC7 <= 1.3213210240277187':
[{'PC8 <= 0.19569467763186432': [{'PC9 <=
-0.823376823806667': [0.0,
                      {'PC6 <=
1.5248850895183041': [{'PC10 <=
0.6930992709341928': [1.0,
                          {'PC10 <=
0.6971094489109289': [0.0, 1.0]}]},
                        0.0]}]},
                    0.0]},
                  {'PC10 <=
-1.2051184770411172': [1.0, 0.0]}]}]},
              {'PC1 <= 0.3161719370087579':
[{'PC10 <= -0.037683807999995114': [0.0,
                  1.0]},
                {'PC8 <=
-0.5533905078966137': [{'PC2 <=
-1.8298299425122482': [{'PC8 <=
-0.7482716259072058': [0.0,
                      {'PC10 <=
-0.5834176186065494': [0.0, 1.0]}]},
                    0.0]},
                  {'PC9 <=
-1.0651086593661436': [{'PC10 <=
-0.12497812867099176': [0.0,
                      1.0]},
                    {'PC1 <=
0.41993349160604554': [{'PC9 <=
0.5930368485272812': [0.0,
                        1.0]},
                      {'PC3 <=
2.0049150019829667': [{'PC2 <=
-3.7031333776207385': [{'PC10 <=
-0.024830186595475187': [1.0,
                            0.0]},
                          0.0]},
                        {'PC10 <=
0.8533663242857688': [1.0,

0.0]}]}]}]}]}]}]}]}]},
          {'PC5 <= -0.9218038037571261':
[{'PC2 <= 1.6932554902054897': [{'PC4 <=
-1.6204935507018048': [{'PC2 <=
0.25109331526462864': [{'PC8 <=
0.051406072850849444': [{'PC7 <=
0.8988225922919348': [1.0,
                      0.0]},
                    0.0]},
                  {'PC3 <=
1.002145955050452': [0.0,
                    {'PC9 <=
0.42690882273707165': [0.0, 1.0]}]}]},
                {'PC8 <=
-0.454614252998261': [{'PC6 <=
-1.2849332279407153': [{'PC7 <=
-0.08467626302482704': [0.0,
                      {'PC10 <=
0.5329149577689328': [{'PC10 <=
-0.9134715578569079': [0.0,
                          1.0]},
                        0.0]}]},
0.14293876476154835': [{'PC3 <=
0.7255706252429494': [1.0,
                            0.0]},
                          {'PC9 <=
0.6558895663421556': [1, 0.0]}]},
                        0.0]},
                      0.0]},
                    0.0]},
                  {'PC4 <=
-0.7525276445570384': [0,
                    {'PC3 <=
-2.4099956557217244': [1, 0.0]}]}]}]},
              {'PC4 <= 1.0011408963148045':
[{'PC9 <= -0.8310754979957748': [0,
                  {'PC5 <=
0.4133716858614151': [{'PC3 <=
1.856425063260704': [{'PC2 <=
0.76225002644307': [0,
                      {'PC3 <=
-1.841017895877334': [1,
                        {'PC4 <=
-1.125134497216548': [{'PC6 <=
0.8731406911030524': [1.0,
                              0.0]},
                            0.0]}]}]},
                      1.0]},
                    0]}]},
                {'PC3 <=
-0.3875667393330866': [0,
                  {'PC2 <=
1.1018523963005573': [{'PC5 <=
-1.1851352139123574': [0.0,
                      {'PC2 <=
-2.542044789329183': [0.0,
                        {'PC7 <=
1.0906059411394815': [1, 0.0]}]}]},
                    0]}]}]}]},
            {'PC9 <= 0.47522603013853754':
[{'PC4 <= -0.06753590635731142': [1,
                0]},
              0.0]}]}]}]}]},
    {'PC1 <= 0.3448380525363134': [1, 0]}]},
  {'PC6 <= -0.5923209052091465': [{'PC7 <=
0.21233068981024117': [{'PC1 <=
0.5357596542629881': [{'PC8 <=
1.8917507396556816': [0,
          1]},
        {'PC9 <= 0.20795693271261803': [0,
1]}]},
      {'PC9 <= 0.02811107358035988': [{'PC5
<= -1.1215649710916407': [0, 1]},
        {'PC2 <= 0.9553639428688796': [0.0,
1]}]}]},
    {'PC8 <= 0.19000952050089764': [1,
      {'PC3 <= 3.1007115509230196': [{'PC7 <=
0.20041911249044375': [0,
          {'PC2 <= 1.317546980926192': [{'PC2
<= -0.9672343818444014': [{'PC5 <=
-0.14403660918309627': [{'PC10 <=
0.6195084736158839': [0.0,
                  1.0]},
                1.0]},
              0.0]},
            1]}]},
         1]}]}]}]}
```

```
                    0.0]},
                  {'PC3 <=
0.7872630893096046': [{'PC6 <=
0.2673539887374736': [{'PC2 <=
1.47608444595242': [0.0,
                        {'PC10 <=
-0.48345421780640285': [1.0, 0.0]}]]},
                      {'PC4 <=
0.13754929895537474': [{'PC1 <=
-2.552433859504294': [0.0,
                          {'PC1 <=
3.7471063995366585': [{'PC3 <=
0.40152490783126565': [1.0,
                              0.0]},
                            0.0]}]]},
                        {'PC3 <=
0.30420090065934313': [0.0,
                          {'PC10 <=
-0.06897153352414528': [0.0, 1.0]}]}]}]]},
                    {'PC4 <=
0.9138515569396832': [{'PC10 <=
-0.7485299643303847': [1.0,
                        {'PC1 <=
0.7439493933004999': [0.0,
                          {'PC10 <=
-0.18526771366727512': [0.0, 1.0]}]}]]},
                      {'PC8 <=
0.07210400035226742': [1.0, 0.0]}]}]}]}]]},
              {'PC3 <= 0.8300279948471145':
[{'PC7 <= -0.6573975196954583': [{'PC9 <=
0.7780495012817084': [0.0,
                    {'PC10 <=
0.2847590257172494': [1.0, 0.0]}]]},
                  {'PC4 <=
1.9452595035562505': [{'PC3 <=
-1.252094860976261': [0.0,
                      {'PC8 <=
-1.1652951051286127': [1.0,
                        {'PC10 <=
0.15765411745835667': [{'PC4 <=
-0.5782815970593735': [{'PC3 <=
-0.7830021265546403': [1.0,
                              0.0]},
                            {'PC3 <=
0.7691992306673174': [1.0, 0.0]}]]},
                          {'PC5 <=
-2.4532891546765314': [1.0,
                            {'PC8 <=
-1.1086560642754226': [1.0, 0.0]}]}]}]}]]},
                    0.0]}]]},
                {'PC9 <= 0.2745442851588624':
[{'PC3 <= 1.195629741496341': [0.0,
                    {'PC9 <=
0.22399424711048127': [{'PC10 <=
0.11764721450672373': [{'PC10 <=
-0.591792707451088': [1.0,
                          0.0]},
                        {'PC3 <=
1.2581535582207772': [{'PC10 <=
0.4819744666358118': [0.0,
                            1.0]},
```

```
                        1.0]}]},
                    0.0]}]},
                1.0]}]}]},
          {'PC3 <= -3.3567552993283476':
[{'PC2 <= 0.46073789147014904': [0.0,
                {'PC7 <=
-0.0592749332304799': [{'PC4 <=
1.8180585300957457': [0.0,
                  1.0]},
                {'PC5 <=
2.651707337968143': [{'PC1 <=
-3.7060343640691586': [0.0,
                    {'PC9 <=
0.2902517417977556': [{'PC6 <=
-1.1837506178129031': [0.0,
                        {'PC10 <=
1.2080803677840106': [{'PC8 <=
0.34296563919553935': [1.0,
                            0.0]},
                          0.0]}]},
                      0.0]}]},
                  0.0]}]}]},
            {'PC5 <= 0.74944814407033':
[{'PC3 <= 0.2652807795880745': [{'PC8 <=
-0.6269277605900321': [0.0,
                  {'PC2 <=
2.218321552577549': [{'PC3 <=
-1.1200102722517546': [{'PC7 <=
0.27138515646018047': [{'PC5 <=
0.5771803542897344': [0.0,
                          {'PC10 <=
0.3765573264585844': [0.0, 1.0]}]},
                        {'PC5 <=
0.13578399692127405': [{'PC6 <=
1.3747701527081313': [0.0,
                            1.0]},
                          {'PC9 <=
0.5797503912881374': [1.0, 0.0]}]}]},
                      {'PC7 <=
-0.25556127959018005': [{'PC3 <=
0.033692623252732846': [0.0,
                          {'PC9 <=
-0.02725897570609783': [0.0, 1.0]}]},
                        {'PC3 <=
-0.9606623861318556': [{'PC6 <=
1.198952084647933': [0.0,
                            1.0]},
                          0.0]}]}]},
                    {'PC2 <=
2.7242449372006146': [{'PC7 <=
0.21302881675016608': [{'PC4 <=
0.3860691850164887': [{'PC10 <=
0.2925296158137348': [1.0,
                            0.0]},
                          0.0]},
                        {'PC10 <=
1.19459363152717': [0.0, 1.0]}]},
                      {'PC1 <=
-1.4430457809686414': [{'PC10 <=
-0.37413387385329583': [1.0,
                          0.0]},
```

```
                         0.0]}]}]}]},
                  {'PC2 <=
1.344745879252146': [{'PC4 <=
-1.8680838013615395': [{'PC7 <=
-0.14360466953676412': [{'PC9 <=
0.35769172539023447': [1.0,
                          0.0]},
                        0.0]},
                      {'PC3 <=
0.3608049107574125': [{'PC6 <=
-0.36422380209829197': [{'PC6 <=
-0.4663959837586975': [{'PC10 <=
0.7644603280963258': [0.0,
                              1.0]},
                            1.0]},
                          0.0]},
                        {'PC7 <=
-2.052335586266967': [1.0,
                          {'PC4 <=
-1.6770363930262737': [{'PC10 <=
0.33855659682316863': [0.0,
                              1.0]},
                            0.0]}]}]}]},
                    {'PC1 <=
2.025792094416495': [{'PC7 <=
0.4695112299781634': [{'PC8 <=
-0.02931563156994246': [0.0,
                          {'PC8 <=
0.03220543214064046': [1.0,
                            {'PC6 <=
1.0994369655569916': [0.0, 1.0]}]}]},
                        {'PC4 <=
4.553282419027422': [{'PC4 <=
-1.3620183216142623': [0.0,
                            {'PC7 <=
1.003451279591781': [1.0, 0.0]}]},
                          0.0]}]},
                      {'PC2 <=
1.374636575003965': [1.0, 0.0]}]}]}]},
                {'PC4 <= 1.218106743831214':
[{'PC7 <= -0.6749427324310695': [{'PC4 <=
-0.6263513339620621': [0.0,
                      {'PC6 <=
-1.1512181147478189': [0.0,
                        {'PC2 <=
0.1888089897194844': [0.0,
                          {'PC10 <=
0.27058585227704846': [1.0, 0.0]}]}]}]},
                    {'PC5 <=
1.0975904626788016': [{'PC10 <=
1.0798323836506927': [{'PC9 <=
-1.266609726272758': [{'PC10 <=
-1.1622020544649174': [1.0,
                            0.0]},
                          0.0]},
                        {'PC9 <=
-0.2589691044779974': [0.0, 1.0]}]},
                      {'PC5 <=
1.1040133953666613': [{'PC10 <=
-0.31327654155771933': [1.0,
                          0.0]},
```

```
                         {'PC7 <=
0.8917821215633577': [{'PC10 <=
-0.2861082135089891': [{'PC6 <=
-0.3868780314206521': [0.0,
                               1.0]},
                             0.0]},
                           {'PC10 <=
-0.03244222406750417': [0.0,
                             {'PC9 <=
0.6957928831226078': [1.0,

0.0]}]}]}]}]}]},
                   {'PC8 <=
-0.4521511880240696': [{'PC10 <=
-0.8988808557900911': [1.0,
                       0.0]},
                     {'PC6 <=
-1.593230288050811': [0.0,
                       {'PC3 <=
0.9007376404409113': [{'PC7 <=
-1.1220195569168347': [0.0,
                           {'PC9 <=
0.9163988801744124': [{'PC9 <=
-1.2111045376275127': [0.0,
                               1.0]},
                             0.0]}]},

0.0]}]}]}]}]}]}]}]}]},
      {'PC6 <= 0.24935772115142918': [{'PC9
<= -0.006163800662947843': [{'PC5 <=
0.8694395756851128': [{'PC2 <=
-3.4149067804908726': [{'PC4 <=
1.453419656620854': [{'PC10 <=
-0.03195053083590062': [{'PC8 <=
0.886060205947294': [0.0,
                     1.0]},
                   {'PC1 <=
4.247855234109455': [1.0, 0.0]}]},
                0.0]},
              {'PC7 <= 1.131036307120978':
[{'PC4 <= 1.2044741244120192': [{'PC9 <=
-0.38135017656956105': [{'PC1 <=
1.3970913466771804': [{'PC9 <=
-1.182466180682558': [{'PC9 <=
-1.4421922174833286': [0.0,
                           1.0]},
                         {'PC4 <=
0.21957226922014247': [0.0,
                           {'PC3 <=
-1.5948330898561132': [1.0,
                             {'PC4 <=
0.3014125494276207': [1.0, 0.0]}]}]}]},
                       {'PC3 <=
-1.0111497987096252': [0.0,
                         {'PC6 <=
-1.1587194352801118': [0.0,
                           {'PC2 <=
-2.406459802961279': [0.0,
                             {'PC4 <=
-2.2657615852548654': [0.0, 1.0]}]}]}]}]},
```

```
                    {'PC1 <=
-1.4621691628608593': [{'PC6 <=
-0.31896877186146777': [0.0,
                        {'PC10 <=
-0.1525417802352137': [1.0,
                          {'PC8 <=
0.8418589643683747': [0.0, 1.0]}]}]},
                      {'PC1 <=
4.435622295565008': [{'PC3 <=
-1.8355856463988212': [{'PC1 <=
-0.8334294522448529': [1.0,
                            0.0]},
                          {'PC2 <=
-2.7990678592770637': [{'PC5 <=
-1.0865487747060332': [0.0,
                              1.0]},
                            0.0]}]},
                        {'PC6 <=
-0.08854898903923486': [0.0, 1.0]}]}]}]},
                  {'PC3 <=
0.8371154102035026': [{'PC5 <=
-0.31352167486131305': [0.0,
                      {'PC3 <=
-0.5502713287454856': [{'PC8 <=
1.034554622454857': [{'PC3 <=
-1.585617775641861': [{'PC9 <=
-1.2623616870056455': [0.0,
                              1.0]},
                            0.0]},
                          1.0]},
                        0.0]}]},
                    {'PC4 <=
1.7601301751699243': [1.0, 0.0]}]}]},
                {'PC8 <= 1.3936157530047368':
[{'PC6 <= -1.9779541759775652': [0.0,
                    1.0]},
                  0.0]}]}]},
            {'PC4 <= -1.2716039374846106':
[{'PC7 <= -1.2823932031241039': [1.0,
                0.0]},
              {'PC10 <= 0.8291164807825707':
[{'PC8 <= 1.4357823556989922': [{'PC2 <=
2.642037886058202': [{'PC9 <=
-0.18653885854110586': [{'PC10 <=
0.05893334601552237': [{'PC3 <=
0.7694755476052427': [{'PC5 <=
2.2246024624369847': [1.0,
                            {'PC9 <=
-1.3089390723878247': [1.0, 0.0]}]},
                          {'PC7 <=
0.6485253125361187': [0.0, 1.0]}]},
                        {'PC3 <=
-0.47860028831698975': [{'PC8 <=
0.4498522421825361': [1.0,
                            {'PC8 <=
1.0836123502266135': [0.0, 1.0]}]},
                          {'PC8 <=
1.319198786011384': [{'PC5 <=
0.9935943304431896': [0.0,
                              1.0]},
                            0.0]}]}]},
```

```
                      {'PC2 <=
1.4805508322223657': [1.0, 0.0]}]},
                    0.0]},
                  0.0]},
                0.0]}]}]},
          {'PC8 <= 0.845351997187648': [{'PC3
<= -1.5998942205185713': [{'PC3 <=
-1.7900663338065135': [{'PC7 <=
0.7559790664568797': [0.0,
                  1.0]},
                1.0]},
              {'PC2 <= 2.1100596137447676':
[{'PC5 <= 2.810703601285847': [{'PC4 <=
-2.6388780226542297': [{'PC5 <=
-0.832188675314921': [1.0,
                      0.0]},
                    {'PC2 <=
-4.732475657394023': [{'PC10 <=
-0.5585367491347013': [0.0,
                        1.0]},
                      {'PC2 <=
1.6882567516114686': [{'PC8 <=
0.36014729146490904': [{'PC8 <=
0.35963592280622403': [0.0,
                            1.0]},
                          0.0]},
                        {'PC2 <=
1.7477288073673312': [1.0, 0.0]}]}]}]}]},
                  {'PC10 <=
0.10505876464606892': [0.0, 1.0]}]},
                {'PC1 <= -1.678629136737166':
[1.0, 0.0]}]}]}]},
            {'PC1 <= 0.22252380398772548':
[{'PC5 <= -0.1646041865628528': [{'PC4 <=
-0.7545410169081102': [{'PC3 <=
-1.5691029418863351': [0.0,
                    1.0]},
                  0.0]},
                {'PC2 <= -1.090223389354784':
[0.0,
                  {'PC8 <=
0.9429492551761164': [{'PC9 <=
0.25467738520817595': [{'PC5 <=
1.9292338427552038': [1.0,
                        0.0]},
                      {'PC5 <=
0.19546166871605353': [{'PC10 <=
-0.319354202868487': [1.0,
                          0.0]},
                        0.0]}]},
                    {'PC9 <=
0.05985051309498146': [0.0,
                      {'PC8 <=
1.980429799391147': [{'PC7 <=
-0.7735533545570621': [0.0,
                          {'PC3 <=
1.3313070879169804': [1.0,
                            {'PC10 <=
0.37624743099405117': [1.0, 0.0]}]}]},
                        0.0]}]}]}]}]}]},
```

```
              {'PC2 <= -3.215520084521584':
[1.0, 0.0]}]}]}]}]},
        {'PC10 <= -0.5229065720498982':
[{'PC5 <= -0.6825654447731452': [{'PC9 <=
-0.49901319063878524': [{'PC10 <=
-1.076138262465162': [0.0,
                1.0]},
              0.0]},
            {'PC8 <= 0.7711746643736423':
[{'PC7 <= 0.07076265475851237': [{'PC9 <=
0.881638053229489': [{'PC1 <=
-2.65190765647448': [{'PC10 <=
-0.6252388431405691': [1.0,
                      0.0]},
                    {'PC8 <=
0.671041466232702': [{'PC8 <=
0.3824283710249512': [{'PC10 <=
-0.7405984173693816': [{'PC10 <=
-0.9048732948178362': [1.0,
                            0.0]},
                          1.0]},
                        1.0]},
                      {'PC6 <=
1.1771350568493113': [0.0, 1.0]}]}]}]},
                  0.0]},
                {'PC4 <= 0.4651696692590329':
[0.0,
                  {'PC9 <=
0.29582023579621436': [{'PC10 <=
-0.9655026762666499': [0.0,
                      {'PC3 <=
1.9108672144441705': [1.0, 0.0]}]},
                    0.0]}]}]}]},
              {'PC10 <= -1.6020224131268144':
[0.0,
                {'PC9 <=
-0.6547727046694745': [0.0,
                  {'PC4 <=
-2.568151899203631': [0.0,
                    {'PC2 <=
2.7966616795985244': [{'PC7 <=
-1.8109221795204535': [{'PC9 <=
-0.1280177700904333': [1.0,
                          0.0]},
                        {'PC5 <=
0.0716058374582683': [{'PC5 <=
0.06158892587594299': [{'PC1 <=
-2.807519757159563': [0.0,
                              1.0]},
                            0.0]},
                          1.0]}]},
                      0.0]}]}]}]}]}]}]},
          {'PC1 <= 2.0594016535599566':
[{'PC1 <= -1.6133090857814083': [{'PC1 <=
-4.000567138397632': [0.0,
                {'PC10 <=
0.4166380014026776': [{'PC9 <=
0.8647120586854768': [{'PC7 <=
0.9108370566012731': [{'PC4 <=
2.3440491617514647': [{'PC7 <=
```

```
0.14293876476154835': [{'PC3 <=
0.7255706252429494': [1.0,
                            0.0]},
                          {'PC9 <=
0.6558895663421556': [{'PC5 <=
1.337011501660243': [1.0,
                              0.0]},
                            0.0]}]},
                        0.0]},
                      0.0]},
                    0.0]},
                  {'PC4 <=
-0.7525276445570384': [{'PC7 <=
0.058627538913018654': [1.0,
                      0.0]},
                    {'PC3 <=
-2.4099956557217244': [{'PC10 <=
0.7443964269254488': [0.0,
                        1.0]},
                      0.0]}]}]}]},
              {'PC4 <= 1.0011408963148045':
[{'PC9 <= -0.8310754979957748': [{'PC1 <=
0.29958908503120524': [0.0,
                    {'PC10 <=
0.2257291880231684': [1.0, 0.0]}]},
                  {'PC5 <=
0.4133716858614151': [{'PC3 <=
1.856425063260704': [{'PC2 <=
0.76225002644307': [{'PC10 <=
1.0409524290176104': [{'PC8 <=
1.901324921083847': [0.0,
                            1.0]},
                          {'PC9 <=
-0.32639621368183447': [1.0, 0.0]}]},
                        {'PC3 <=
-1.841017895877334': [{'PC10 <=
0.1629315480095975': [{'PC9 <=
0.09156094153819991': [1.0,
                              0.0]},
                            0.0]},
                          {'PC4 <=
-1.125134497216548': [{'PC6 <=
0.8731406911030524': [1.0,
                              0.0]},
                            0.0]}]}]},
                      1.0]},
                    {'PC10 <=
1.329861877312175': [0.0, 1.0]}]}]},
                {'PC3 <=
-0.3875667393330866': [{'PC6 <=
0.4854192065297662': [{'PC10 <=
-0.21418351599662033': [1.0,
                      0.0]},
                    0.0]},
                  {'PC2 <=
1.1018523963005573': [{'PC5 <=
-1.1851352139123574': [0.0,
                      {'PC2 <=
-2.542044789329183': [0.0,
```

```
                        {'PC7 <=
1.0906059411394815': [{'PC1 <=
-1.3684464540818482': [0.0,
                            {'PC10 <=
-0.4790251813455355': [0.0, 1.0]}]},
                          0.0]}]}]},
                    {'PC8 <=
0.409894350907712': [1.0, 0.0]}]}]}]}]}]},
            {'PC9 <= 0.47522603013853754':
[{'PC4 <= -0.06753590635731142': [{'PC7 <=
0.469125667447828': [{'PC7 <=
-1.2995547728894845': [0.0,
                    {'PC10 <=
-0.5016351615227813': [0.0,
                      {'PC8 <=
1.8256264070378676': [{'PC6 <=
1.5033666115480022': [{'PC2 <=
-1.96756812067383': [{'PC10 <=
0.8352739805149486': [0.0,
                              1.0]},
                            1.0]},
                          0.0]},
                        0.0]}]}]},
                  0.0]},
                {'PC5 <= 0.9644530036766057':
[{'PC1 <= 2.177486198829087': [1.0,
                    {'PC5 <=
-2.061011939386846': [1.0, 0.0]}]},
                  {'PC7 <=
1.3914079533418753': [1.0, 0.0]}]}]},
              0.0]}]}]}]}]},
    {'PC1 <= 0.3448380525363134': [{'PC9 <=
-0.7465826694517675': [0.0,
        {'PC10 <= -1.270883359710674': [0.0,
          {'PC8 <= 1.2865280169147266':
[{'PC9 <= -0.5980152764583135': [{'PC7 <=
-0.19470240025229743': [0.0,
                1.0]},
              1.0]},
            0.0]}]}]},
      {'PC8 <= -0.9429530372139716': [1.0,
0.0]}]}]},
  {'PC6 <= -0.5923209052091465': [{'PC7 <=
0.21233068981024117': [{'PC1 <=
0.5357596542629881': [{'PC8 <=
1.8917507396556816': [{'PC2 <=
2.0402301691851616': [0.0,
            {'PC7 <= -0.2503407053425144':
[0.0,
              {'PC10 <= 0.26533154155950645':
[1.0, 0.0]}]}]},
          {'PC10 <= 0.4478827718187208':
[1.0, 0.0]}]},
        {'PC9 <= 0.20795693271261803': [{'PC6
<= -1.053081422251128': [{'PC10 <=
-0.2880914190369364': [0.0,
              1.0]},
            0.0]},
          {'PC5 <= 1.2068834281689282': [0.0,
1.0]}]}]},
```

```
      {'PC9 <= 0.02811107358035988': [{'PC5
<= -1.1215649710916407': [{'PC5 <=
-1.9223177966628961': [1.0,
            0.0]},
          {'PC4 <= 1.1452132496692908': [1.0,
0.0]}]},
        {'PC2 <= 0.9553639428688796': [0.0,
          {'PC5 <= 0.18992699667062474':
[1.0, 0.0]}]}]}]},
    {'PC8 <= 0.19000952050089764': [{'PC3 <=
2.538896711131528': [{'PC2 <=
0.8600717950716231': [{'PC5 <=
-0.026273884762807235': [{'PC3 <=
2.5003204562094723': [{'PC9 <=
-0.9578557095269371': [1.0,
                {'PC1 <= 1.919859134303179':
[{'PC10 <= 0.6890114616281802': [0.0,
                    {'PC10 <=
0.905001478365709': [1.0, 0.0]}]},
                  1.0]}]},
              {'PC10 <= 0.22876642198898925':
[1.0, 0.0]}]},
            {'PC7 <= -0.5925458152083929':
[{'PC10 <= 0.9024471639366771': [0.0,
                1.0]},
              {'PC10 <= 0.7876457905817245':
[{'PC1 <= 0.5342148785357579': [{'PC7 <=
0.1889621171012836': [{'PC9 <=
0.45769198316361753': [0.0,
                      1.0]},
                    {'PC6 <=
0.2941828785936907': [0.0, 1.0]}]},
                  1.0]},
                {'PC6 <=
-0.24487900828102094': [1.0, 0.0]}]}]}]},
          {'PC3 <= 2.5345944179985236':
[{'PC6 <= -0.5201045706669156': [{'PC7 <=
-0.18142157579449464': [0.0,
                1.0]},
              {'PC10 <= 0.38937169162874835':
[1.0,
                {'PC2 <= 2.239787025753483':
[1.0, 0.0]}]}]},
            0.0]}]},
        {'PC5 <= -1.965029922021348': [0.0,
          {'PC9 <= 0.6155674921110397':
[{'PC10 <= -1.2359200436819604': [{'PC10 <=
-1.2606156103625696': [1.0,
                0.0]},
              {'PC3 <= 3.098818905568819 6':
[{'PC6 <= 1.6663945346703646': [{'PC3 <=
3.092542572850069': [{'PC5 <=
-0.961900778728928': [{'PC10 <=
0.3566347197605006': [0.0,
                        1.0]},
                      {'PC8 <=
-0.1813800463668585': [{'PC6 <=
-0.2520569239109438': [{'PC8 <=
-0.681673935958152': [0.0,
                            1.0]},
                          1.0]},
```

```
                        {'PC8 <=
-0.17327016732817593': [0.0,
                          {'PC1 <=
2.76668267783438': [{'PC7 <=
-0.6091975564457992': [0.0,
                               1.0]},
                             0.0]}]}]}]},
                     0.0]},
                   0.0]},
                 1.0]}]},
            {'PC6 <= -0.07821562974125348':
[0.0,
              {'PC8 <= -1.1804545752835924':
[0.0, 1.0]}]}]}]}]}]},
      {'PC3 <= 3.1007115509230196': [{'PC7 <=
0.20041911249044375': [{'PC5 <=
1.8253719349355553': [{'PC10 <=
-0.24121591296079464': [{'PC7 <=
-0.37097318569513904': [{'PC6 <=
-0.38403889625353194': [1.0,
                   0.0]},
                 1.0]},
               0.0]},
             1.0]},
          {'PC2 <= 1.317546980926192': [{'PC2
<= -0.9672343818444014': [{'PC5 <=
-0.14403660918309627': [{'PC10 <=
0.6195084736158839': [0.0,
                   1.0]},
                 1.0]},
               0.0]},
             {'PC3 <= 2.2812953321779226':
[0.0, 1.0]}]}]},
        {'PC5 <= 1.6874049755011886': [{'PC7
<= 0.3648760828774664': [{'PC2 <=
1.4817830485021184': [0.0,
              {'PC5 <= -0.11043136295732768':
[0.0, 1.0]}]},
            {'PC5 <= 1.433522804529145':
[1.0,
              {'PC9 <= -0.5374634840876182':
[1.0, 0.0]}]}]},
          0.0]}]}]}]}]}
```


Evaluation of model on validation dataset

	Unpruned (depth=5)	Pruned (depth=5)	Unpruned (depth=15)	Pruned (depth=15)
Confusion matrix	[[506 128] [269 365]]	[[506 128] [254 380]]	[[537 97] [307 327]]	[[575 59] [262 372]]
Accuracy	0.69	0.7	0.68	0.75

Precision (1)	0.74	0.75	0.77	0.86
Precision (0)	0.65	0.67	0.64	0.69
Recall (1)	0.58	0.6	0.52	0.59
Recall (0)	0.8	0.8	0.85	0.91
F1-score (1)	0.65	0.67	0.62	0.7
F1-score (0)	0.72	0.73	0.73	0.78

As decision tree with max-depth 15 produced better f1-score on validation dataset, we have selected decision-tree with max-depth 15 as the final model.

Result on test dataset:

Confusion matrix:

```
[[297  20]
 [102 215]]
```

Accuracy: 0.81
Precision (1): 0.91
Precision (0): 0.74
Recall (1): 0.68
Recall (0): 0.94
F1-score (1): 0.78
F1-score (0): 0.83

Conclusion:

Firstly, the data was pre-processed and cleaned, standardized and pca was applied. After training with the decision tree classifier, we plotted the graph and found that there was a wide gap between training and validation accuracy. To handle that, we pruned the decision tree, and evaluated the result on a test (unseen) dataset.