**Problem 1**

**1. Depth with the highest Recall and Reason**

**The depth value of 2 results in the highest recall. A higher depth allows the decision - tree to capture more complex patterns in the data. By learning these complex patterns, the model can better identify positive cases, which leads to an increase in recall.**

**2. Depth with the lowest Precision and Reason**

**The depth value of 1 has the lowest precision. A very deep tree might overfit the training data, leading to false positives. Although a depth of 1 is not a deep tree, in this context, it may not be able to capture enough patterns in the data, and the resulting model may not be accurate enough, which could also lead to relatively more false positives and thus lower precision.**

**3. Depth with the best F1 score**

**The depth value of 2 results in the best F1 score. The F1 score is a harmonic mean of precision and recall. Since a depth of 2 provides a good balance between the ability to identify positive cases (recall) and the accuracy of those identifications (precision), it achieves the best F1 score.**

**4. Difference between micro, macro, and weighted methods of score calculation**

**Micro: This method calculates metrics globally. It counts the total number of true positives, false negatives, and false positives across all classes. Each individual prediction is given equal weight. This is useful when dealing with imbalanced datasets because it takes into account the overall performance of the model on all samples.**

**Macro: Metrics are calculated separately for each class, and then the unweighted average of these metrics is taken. It treats all classes equally, regardless of their prevalence in the dataset. This can be affected by minority classes, as each class contributes equally to the final score.**

**Weighted: Similar to the macro method, metrics are calculated for each class separately. However, the average is weighted by the number of samples in each class. This is beneficial for imbalanced datasets, as it gives more importance to classes with a larger number of samples.**

**Problem 2**

**1. Entropy, Gini, and Misclassification Error of the first split**

**Entropy of the first split: 0.5889187667244618**

**Gini of the first split: 0.32550820073530584**

**Misclassification Error of the first split: 0.2767203513909224**

**2. Information Gain**

**The information gain is 0.5889187667244618, which is equal to the entropy gain of the first split.**

**3. Feature selected for the first split and the decision boundary value**

**The feature selected for the first split is uniformity\_of\_cell\_size, and the value determining the decision boundary is 2.5.**

**Problem 3**

**1. Comparison of F1 score, Precision, and Recall of PCA - based single - factor model and original data**

**Original data (without PCA)**

**F1 score: 0.9047619047619048**

**Precision: 0.9047619047619048**

**Recall: 0.9047619047619048**

**PCA - based single - factor model (using only the first principal component)**

**F1 score: 0.9243697478991596**

**Precision: 0.9821428571428571**

**Recall: 0.873015873015873**

**Using only the first principal component, the F1 score and precision are higher than those of the original data, while the recall is slightly lower.**

**2. Using the first and second principal components**

**F1 score: 0.9243697478991596**

**Precision: 0.9821428571428571**

**Recall: 0.873015873015873**

**The results are the same as when using only the first principal component.**

**3. Values from the Confusion Matrix**

**False Positives (FP): 1**

**True Positives (TP): 55**

**False Positive Rate (FPR): 0.009259259259259259**

**True Positive Rate (TPR): 0.873015873015873**

**4. Is using continuous data beneficial for the model? How?**

**Using continuous data can be beneficial for the model as it contains more information. In this case, although the PCA - based models (using the first or the first two principal components) have higher F1 scores and precision, the recall is slightly lower compared to the original continuous data. The continuous data provides a more comprehensive view of the dataset, allowing the model to potentially learn more complex relationships. However, PCA helps reduce the dimensionality of the data, which may lead to faster training and potentially better generalization. To determine if PCA is actually beneficial, we need to consider the specific requirements of the application. If precision is more important, PCA - based models may be a better choice. If recall is the key metric, the original continuous data may be more suitable.**