

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR

Report Summary

Instructor: Sudeshna Sarkar



Krishna Biswakarma

INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR.....	1
Performance Metric.....	3
1. Noiseless Dataset (`diabetes.csv`).....	3
2. Noisy Dataset (`diabetes_noise.csv`).....	3
Performance Comparison.....	4
Accuracy Improvement:.....	4
Precision and Recall Improvement:.....	4
Impact of Noise.....	4
Key Findings and Implications.....	5

Performace Metric

1. Noiseless Dataset (`**diabetes.csv**`)

- Before Pruning:
 - Accuracy: 73.38%
 - Macro Precision: 71.05%
 - Macro Recall: 71.21%
- After Pruning:
 - Accuracy: 76.62%
 - Macro Precision: 75.41%
 - Macro Recall: 77.37%

2. Noisy Dataset (`**diabetes_noise.csv**`)

- Before Pruning:
 - Accuracy: 50.81%
 - Macro Precision: 48.67%
 - Macro Recall: 48.70%
- After Pruning:
 - Accuracy: 58.38%
 - Macro Precision: 57.71%
 - Macro Recall: 57.91%

Performance Comparison

Accuracy Improvement:

- For the noiseless dataset, pruning led to an improvement in accuracy from 73.38% to 76.62%, which is a significant gain of 3.24%.
- For the noisy dataset, pruning improved accuracy from 50.81% to 58.38%, a gain of 7.56%.

Precision and Recall Improvement:

- Noiseless Dataset:
 - Precision improved from 71.05% to 75.41%, and recall improved from 71.21% to 77.37%.
- Noisy Dataset:
 - Precision improved from 48.67% to 57.71%, and recall improved from 48.70% to 57.91%.

Impact of Noise

The presence of noise in the dataset significantly impacts model performance:

- Accuracy: The model performs considerably worse on the noisy dataset compared to the noiseless dataset, both before and after pruning. The accuracy increase due to pruning is more pronounced in the noisy dataset.
- Precision and Recall: Both metrics are lower in the noisy dataset, indicating that the model struggles more with noise. However, pruning helps improve these metrics in noisy conditions, though not to the same extent as in noiseless conditions.

Key Findings and Implications

1. Pruning Benefits: Pruning consistently improves model performance in both noiseless and noisy datasets. However, the improvements are more substantial in the context of noisy data, suggesting that pruning helps in reducing the overfitting caused by noise.

2. Impact of Noise: Noise has a detrimental effect on model accuracy, precision, and recall. The model's performance degrades when faced with noisy data, emphasizing the importance of data quality.

3. Practical Implications: While pruning is beneficial, efforts to reduce noise in datasets should also be considered to achieve optimal performance. For real-world applications, it is crucial to balance data quality and model tuning to enhance predictive accuracy and reliability.