

Post Analysis:

| Feature Name | Information Gain | Metric Type | Inference |
|---------------------|------------------|--------------------|---|
| num_clauses_walksat | 0.192 | Instance Size | The single most important predictor is the total number of clauses. |
| num_clauses_dpII | 0.189 | Instance Size | Identical to the above, confirming the importance of clause count. |
| ratio_walksat | 0.118 | Instance Structure | The Clause-to-Literal Ratio (L/N) is clearly an important differentiating factor. |
| clause_density_dpII | 0.110 | Instance Structure | The Clause Density (L/N^2) is another critical structural metric. |
| ratio_dpII | 0.102 | Instance Structure | The DPLL version of the Clause-to-Literal Ratio. |

Other feature attributes either contribute minimally, or do not contribute at all.

Train-Test Split used : 70:30

The most significant inference is the class imbalance in the model:

- Actual Imbalance: Out of 1220 instances, 1206 (98.85%) are solved better(faster) by DPLL (Class 0), while only 13 (1.15%) are better solved by WalkSAT (Class 1).

Model Accuracy: 0.932

Precision : 0.88

WalkSAT Precision (0.88)

Implying, when the model predicts WalkSAT, it's correct only 88% of the time. Error rate : 12%.

False Positives (DPLL instances misclassified as WalkSAT) is approximately $266 - 234 = 32$ instances.

The model is still misclassifying 32 instances where DPLL was better as instances where WalkSAT should be chosen.