**k-Nearest Neighbours (kNN) Classifier Model**

**Original Dataset - Results Overview:**

| **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC** |
| --- | --- | --- | --- | --- | --- |
| **10 Features** | 0.90 | 0.32 | 0.11 | 0.17 | 0.68 |
| **20 Features** | 0.90 | 0.36 | 0.13 | 0.19 | 0.70 |
| **37 Features** | 0.90 | 0.33 | 0.08 | 0.13 | 0.67 |

**Analysis of Metrics**

1. **Accuracy**: All datasets show a high accuracy of 90%, which might initially suggest strong classifier performance. However, accuracy can be misleading for imbalanced datasets, as it doesn’t account for the imbalance between classes (especially in binary classification with skewed class distributions).
2. **Precision and Recall**:
   * **Precision**: The precision values across the datasets are relatively low (0.32–0.36). This indicates a high rate of false positives (the classifier often misclassifies negative samples as positive)
   * **Recall**: The recall scores are also low (0.08–0.13), reflecting a high rate of false negatives where positive samples are classified as negative. This indicates the model struggles to identify true positives accurately.
3. **F1-Score**: The F1-scores are low across all datasets, especially with the 37-feature set. The F1-score, as the harmonic mean of precision and recall, is heavily influenced by the low recall values. This suggests that neither precision nor recall is strong, and the classifier’s balance between identifying true positives and avoiding false positives is suboptimal.

A graph of a line

Description automatically generated with medium confidence

1. **AUC (Area Under the ROC Curve)**:
   * The AUC values (0.67–0.70) show some variability but remain below 0.75, suggesting that the classifier has limited discriminative power for this problem. Ideally, higher AUC values (above 0.80) would indicate stronger performance. These AUC scores suggest that the classifier performs only slightly better than random chance.

**Conclusions Based on Evaluation Metrics**

1. **Most Suitable Metrics**:
   * **Precision and Recall** are more reliable metrics in this case due to the likely class imbalance. Accuracy alone would be insufficient for evaluation, as it does not capture the classifier's struggles with true positive and false positive rates. The low recall values are particularly concerning, as they indicate many positive cases are being missed, which might be critical depending on the application.
   * **AUC** is also useful as it provides insight into the classifier’s ability to differentiate between classes. However, given the low AUC scores, the k-Nearest Neighbors (kNN) model may not be well-suited for this problem without further adjustments.
2. **Insights on Feature Selection**:
   * The 20-feature dataset provides slightly better precision, recall, F1-score, and AUC compared to the 10- and 37-feature datasets. This could suggest that the 20-feature set strikes a better balance between including relevant information and avoiding unnecessary noise.
   * The 37-feature set, which includes more attributes, shows a slight decline in recall and F1-score, possibly due to the inclusion of irrelevant or redundant features. This outcome reinforces the importance of careful feature selection and dimensionality reduction to improve classifier performance.

**Undersampled Dataset:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Metric** | **10 Features (Undersampled)** | |  | | --- | | **20 Features (Undersampled)** |  |  | | --- | |  | | | **37 Features (Undersampled)** | | --- |  |  | | --- | |  | |
| **Accuracy** | 0.69 | 0.72 | 0.69 |
| **Precision** | 0.68 | 0.71 | 0.69 |
| **Recall** | 0.71 | 0.74 | 0.70 |
| **F1-Score** | 0.70 | 0.72 | 0.69 |
| **AUC** | 0.73 | 0.77 | 0.74 |

**Now we compare and analyse each metric in the context of undersampling:**

1. **Accuracy**:
   * **Original Datasets**: Accuracy was consistent around 0.90 for all feature sets.
   * **Undersampled Datasets**: Accuracy dropped to around 0.69-0.72.
   * **Interpretation**: Accuracy has decreased, this is expected with undersampling, as it often reduces the dominance of the majority class and focuses the model on both classes more equally. This reduction may actually indicate a more balanced model rather than one biased towards the majority class.
2. **Precision**:
   * **Original Datasets**: Precision values were relatively low (0.32 to 0.36).
   * **Undersampled Datasets**: Precision has increased substantially to around 0.68-0.71.
   * **Interpretation**: The improvement in precision suggests that undersampling helped the model reduce false positives by training it to focus more on the minority class. This is a positive impact of undersampling.
3. **Recall**:
   * **Original Datasets**: Recall values were very low (0.08 to 0.13).
   * **Undersampled Datasets**: Recall values increased significantly to around 0.70-0.74.
   * **Interpretation**: The increase in recall indicates that the model is now better at identifying the minority class (heart disease cases). This suggests undersampling effectively helped the model recognize the minority class more reliably, likely due to a balanced dataset.
4. **F1-Score**:
   * **Original Datasets**: F1-Scores were low (0.13 to 0.19).
   * **Undersampled Datasets**: F1-Scores improved to around 0.69-0.72.
   * **Interpretation**: F1-score balances both precision and recall, and this increase reflects that the model has become better at managing both false positives and false negatives. The improved F1-score suggests that the model’s overall performance is more balanced with undersampling.
   * A graph of a number of data

     Description automatically generated
5. **AUC**:
   * **Original Datasets**: AUC values ranged from 0.67 to 0.70.
   * **Undersampled Datasets**: AUC values improved to around 0.73-0.77.
   * **Interpretation**: An increase in AUC indicates better discrimination between classes with the undersampled datasets. The undersampled model's ROC curves (as shown in the image) stay further from the diagonal random line, suggesting better classification ability.

**Impact of Undersampling**

* **Recall**: Undersampling significantly increased recall, indicating that the model has improved its ability to detect the minority class (heart disease cases). This shift is crucial, as higher recall generally means the model is less likely to miss true positives, an important factor in medical diagnoses.
* **Precision**: Precision also increased, which is a good sign. Often, undersampling can lead to a higher number of false positives, reducing precision, but our results show that precision was enhanced. This suggests that the model improved its ability to correctly predict positive cases even with a smaller data set.
* **F1-Score**: With higher F1-scores across the undersampled datasets, the model demonstrates a better balance between precision and recall, making it more suitable for applications where both true positives and false negatives are crucial.

**ROC and AUC Analysis**

The ROC curves (as shown in the image) for the undersampled datasets show a clear improvement, with AUC values now ranging from 0.73 to 0.77. This higher AUC indicates that the undersampling technique allowed the kNN classifier to achieve better separation between the classes, effectively reducing overlap between false positives and true positives.