**Logistic Regression Classifier Model**

**Original Dataset - Results Overview:**

| **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC** | **Confusion Matrix** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TN** | **FP** | **FN** | **TP** |
| **10 Features** | 0.91 | 0.52 | 0.09 | 0.15 | 0.80 | 54440 | 454 | 4959 | 491 |
| **20 Features** | 0.91 | 0.54 | 0.10 | 0.17 | 0.83 | 54413 | 481 | 4895 | 554 |
| **37 Features** | 0.91 | 0.53 | 0.11 | 0.18 | 0.84 | 54387 | 507 | 4868 | 582 |

**Metrics Analysis:**

1. **Accuracy:** While accuracy is a straightforward metric that indicates the overall proportion of correct predictions, it can be misleading in imbalanced datasets. In our case, the accuracy remained consistently high (around 0.91) across the three encoded datasets. However, this high accuracy was somewhat deceptive, as it masked poor performance in identifying heart disease cases. The model could predict the majority class (no heart disease) correctly, leading to inflated accuracy while failing to capture the minority class effectively.
2. **Precision and Recall:**

**Precision**: With values between 0.52 and 0.54, the model shows a moderate ability to reduce false positives but struggles to accurately identify true positives. This suggests that, while the model may correctly classify negative instances, it has a higher chance of misclassifying actual heart disease cases as non-heart disease.

**Recall**: Low recall values (ranging from 0.09 to 0.11) indicate that a substantial number of positive cases are missed (high FN), a significant concern in a medical context where undiagnosed cases could have severe consequences. Recall should be improved to make the model more sensitive to true heart disease cases.

1. **F1-Score:** The low F1-scores across all datasets (0.15–0.18) reflect the poor balance between precision and recall. This low score points to a significant limitation in the model’s predictive capability for the minority class. Improvements in either precision or recall are essential to increase the F1-score and make the model more reliable.
2. **AUC (Area Under the ROC Curve):** With AUC values increasing slightly with the feature count (0.80 to 0.84), there is a slight gain in the model's ability to differentiate between classes. However, these values are below 0.85, suggesting that the model has limited discriminatory power and may require further enhancements, such as more complex feature engineering or resampling, to effectively separate positive and negative classes.

A line graph with different colored lines

Description automatically generated

1. Confusion Matrix: The confusion matrix for each dataset indicates high values for true negatives (TN) but significantly lower true positives (TP). False negatives (FN) are particularly high, underscoring the model’s struggle to identify heart disease cases. Each matrix illustrates the impact of class imbalance, where a focus on maximizing TN and minimizing FP has led to an overlooked positive class.

A blue square with white text

Description automatically generated A blue square with white text

Description automatically generated A blue square with white text

Description automatically generated

**Conclusions Based on Evaluation Metrics**

1. **Most Suitable Metrics:**

Precision and Recall are essential in this case, with recall being particularly important given the model’s high FN rate. A high recall would ensure fewer missed heart disease cases, which is critical in a medical application. Accuracy is less meaningful due to the class imbalance, while AUC provides a broader view of model discrimination, though still below desired levels.

1. **Insights on Feature Selection:**

The 37-feature dataset performed marginally better, suggesting that additional features contributed useful information without overfitting. However, the marginal improvement also indicates that model improvements are likely needed beyond just feature count; resampling or more advanced techniques could further enhance the model’s sensitivity.

**Undersampled Dataset - Results Overview:**

| **Dataset** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **AUC** | **Confusion Matrix** | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TN** | **FP** | **FN** | **TP** |
| **10 Features**  **(Undersampled)** | 0.71 | 0.74 | 0.67 | 0.70 | 0.79 | 4127 | 1311 | 1821 | 3646 |
| **20 Features**  **(Undersampled)** | 0.76 | 0.75 | 0.78 | 0.76 | 0.83 | 3999 | 1439 | 1190 | 4277 |
| **37 Features**  **(Undersampled)** | 0.76 | 0.75 | 0.79 | 0.77 | 0.84 | 3983 | 1455 | 1131 | 4336 |

**Metrics Analysis:**

1. **Accuracy:** Accuracy decreased to around 0.71–0.76 due to undersampling, as expected. This drop reflects a more balanced model that focuses on both classes rather than prioritizing the majority class. Although lower, the accuracy here may represent a more honest assessment of model performance.
2. **Precision and Recall:** Precision improved substantially to 0.74–0.75, indicating a lower rate of false positives and suggesting the model is better at identifying true positives. With recall rising to 0.67–0.79, the model now successfully captures more heart disease cases. This improvement in recall is critical for health-related applications, as it reduces the number of missed diagnoses, contributing to more reliable predictions.
3. **F1-Score:** F1-scores also saw significant improvements, now around 0.70–0.77, indicating that the model’s balance between precision and recall is more optimal with undersampling. This improvement demonstrates the positive impact of undersampling on balancing class prediction.
4. **AUC:** The AUC values improved to around 0.79–0.84, with the best results for the 37-feature dataset. This increase confirms that the model’s discriminative power was enhanced by undersampling, allowing it to more accurately differentiate between positive and negative classes.

A line graph with different colored lines

Description automatically generated

1. **Confusion Matrix:** The confusion matrices reveal a notable improvement in TP values, with a simultaneous reduction in FN. This change indicates that the model is now more balanced, correctly identifying both positive and negative classes with greater accuracy. This balanced confusion matrix outcome confirms that undersampling mitigated the imbalance effects, yielding a more robust model for real-world applications.

A blue squares with white text

Description automatically generated A blue squares with white text

Description automatically generated A blue squares with white text

Description automatically generated

**Impact of Undersampling**

Recall: The model’s recall improvement suggests it is now better suited to detect true positive cases of heart disease, reducing the risk of missed cases.

Precision: Increased precision indicates fewer false positives, suggesting that undersampling helped the model refine its focus on true cases.

F1-Score: Higher F1-scores across all undersampled datasets signify an enhanced balance, making the model more applicable in contexts where both true positives and true negatives are important.