**Model Comparison Report**

Compare different Models:

| Model | Accuracy | Precision | Recall | F1 Score | ROC-AUC |
| --- | --- | --- | --- | --- | --- |
| RF | 99.97% | 1 | 1 | 1 | 1 |
| GB | 99.91% | 1 | 1 | 1 | 0.9023 |
| SVM | 99.86% | 1 | 1 | 1 | 0.9990 |
| LR | 99.86% | 1 | 1 | 1 | 0.9973 |
| XGB | 99.97% | 1 | 1 | 1 | 0.9023 |
| ANN | 99.92% | 1 | 1 | 1 | 0.9992 |

**Insights Report**

1. Random Forest (RF)

Achieved the highest accuracy and ROC-AUC, making it one of the top performers. Random Forest is known for its robustness and ability to handle large datasets with high-dimensional spaces effectively. This model is ideal when interpretability and performance are both critical.

2. XGBoost (XGB)

XGBoost also delivered top-notch performance, closely trailing Random Forest. Its high accuracy and ROC-AUC indicate that it is exceptionally well-suited for complex datasets like the KDD Cup 1999. XGBoost is favoured in scenarios where model performance is paramount, and computational resources are available for hyperparameter tuning. It is highly effective in competitions and real-world applications due to its superior predictive power.

3. Gradient Boosting (GB)

Gradient Boosting performed very well, with only a slight drop in accuracy and ROC-AUC compared to RF and XGB. This model excels in handling unbalanced datasets and can capture intricate patterns.GB is suitable for applications where high accuracy is required, and the data might be unbalanced. However, it may require more tuning compared to Random Forest.

4. Artificial Neural Network (ANN)

The ANN model showed excellent performance, with a high accuracy and ROC-AUC. Neural networks are particularly powerful for capturing non-linear relationships in the data.ANNs are ideal for applications where the data has complex patterns, and interpretability is less of a concern.

5. Support Vector Machine (SVM)

SVM performed well, with high accuracy and ROC-AUC, but slightly below the ensemble methods and ANN. It is effective in high-dimensional spaces and is robust to overfitting, especially in cases with fewer samples. SVM is useful in scenarios with high-dimensional data where the number of features exceeds the number of samples. It works best with well-separated classes.

6. Logistic Regression (LR)

Logistic Regression, while simpler, performed on par with SVM. Its performance was robust, but it may not capture complex non-linear relationships as effectively as the other models. Logistic Regression is suitable for situations where interpretability is critical, and the relationships between features and the target variable are expected to be linear.