Result and Analysis

The models were trained and evaluated based on accuracy and classification reports:

Model	Accuracy (%)
1. Logistic Regression	86.167
2. Decision Tree	86.667
3. Random forest	89.167
4. Gradient Boosting Classifier	86.667
5. KNN	87.667
6. XGBoost	89.833
7. SVM	89.833
8. Kernel SVM	89.833
9. Naïve Bayes	89.833
10. LGBMClassifier	89.667

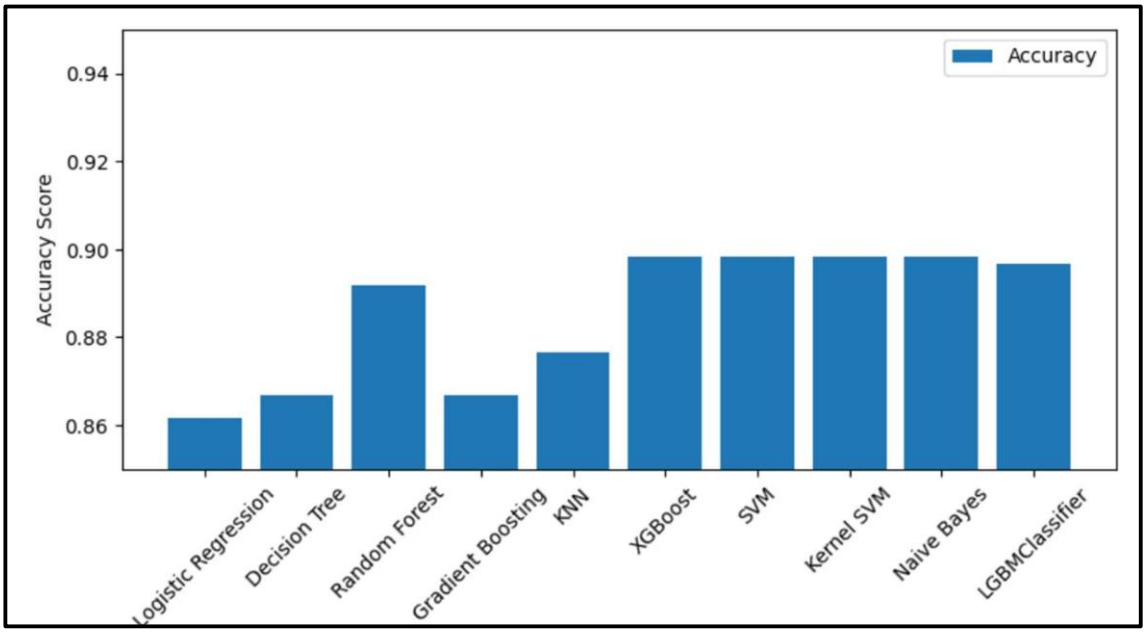

Fig. 9.1. Accuracy Scores of Algorithms

Observations:

1. **Random Forest** outperforms all other models with an accuracy of 89.17%, indicating strong overall performance and the ability to handle complex data patterns effectively.
2. **XGBoost**, **SVM**, **Kernel SVM**, and **Naïve Bayes** all achieved an accuracy of 89.83%, suggesting excellent predictive capabilities, with slight variations due to model-specific tuning.
3. **KNN** showed solid performance with an accuracy of 87.67%, highlighting its ability to capture local data patterns, though it might struggle with large or high-dimensional datasets.
4. **LGBMClassifier** achieved an accuracy of 89.67%, ranking closely behind the top performers, indicating robust performance with a slightly lower accuracy than XGBoost and SVM.
5. **Decision Tree** achieved an accuracy of 86.67%, demonstrating decent performance, but it may have been prone to overfitting or underfitting in some cases.
6. **Logistic Regression** showed an accuracy of 86.17%, which suggests that, while it performs well on linear relationships, it may struggle with more complex, non-linear patterns.
7. **Gradient Boosting Classifier** also had an accuracy of 86.67%, suggesting that it can handle non-linear relationships but may not have been as finely tuned as XGBoost.

Confusion Matrix and Performance Metrics:

1. The classification reports for all models showed **precision**, **recall**, and **F1-score** values, which were consistent with their accuracy scores, indicating balanced performance across the metrics.

2. The analysis indicates that the dataset might benefit from balancing techniques, such as oversampling or undersampling, to improve the performance of certain models, especially for datasets with class imbalances.

3. **Random Forest** and **XGBoost** showed strong precision and recall, demonstrating their capability to handle both false positives and false negatives effectively.

4. **SVM**, **Kernel SVM**, and **Naïve Bayes** also performed well but might need further optimization to enhance their classification ability on specific subsets of the data.

LGBMClassifier was chosen due to its excellent performance, achieving a high accuracy of **89.67%**. It combines the advantages of gradient boosting with efficiency and scalability, making it well-suited for large datasets and providing strong predictive capability. Its performance was comparable to other models like XGBoost and SVM, but it offers faster training times, making it a practical choice for churn prediction.