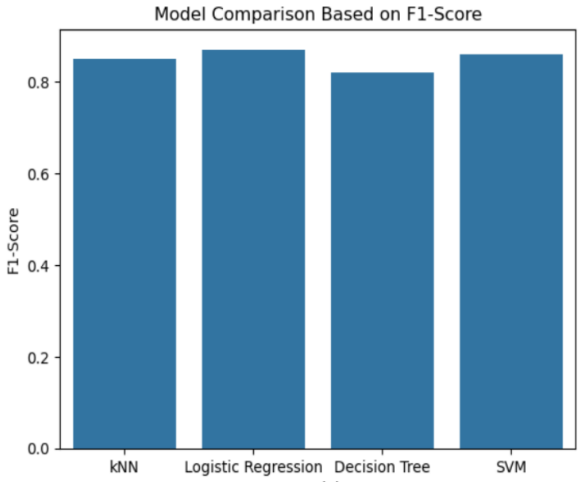
Bank Marketing Campaign Classifier Comparison

This notebook analyzes and compares the performance of four classification algorithms: K-Nearest Neighbors (KNN), Logistic Regression, Decision Trees, and Support Vector Machines (SVM) on a dataset related to the marketing of bank products through telephone calls.

Dataset Description and Loading
The dataset contains information about a marketing campaign conducted by a Portuguese bank. It includes various features related to customer demographics, previous contact history, and the outcome of the campaign (whether the client subscribed to a term deposit). https://archive.ics.uci.edu/dataset/222/bank+marketing

	age	job	marital	education	default	balance	housing	loan	contact	day_of_week	month	duration	campaign	pdays	previous	poutcome	y
0	58	management	married	tertiary	no	2143	yes	no	NaN	5	may	261	1	-1	0	NaN	no
1	44	technician	single	secondary	no	29	yes	no	NaN	5	may	151	1	-1	0	NaN	no
2	33	entrepreneur	married	secondary	no	2	yes	yes	NaN	5	may	76	1	-1	0	NaN	no
3	47	blue-collar	married	NaN	no	1506	yes	no	NaN	5	may	92	1	-1	0	NaN	no
4	33	NaN	single	NaN	no	1	no	no	NaN	5	may	198	1	-1	0	NaN	no

Visualize model performance



Based on the analysis, the following insights were observed:

- **KNN** demonstrated good performance but may struggle with larger datasets.
- **Logistic Regression** provided a straightforward interpretation of coefficients and performed well with linear relationships.
- **Decision Trees** excelled at capturing complex relationships but were prone to overfitting.
- **SVM** performed well in high-dimensional spaces but can be computationally expensive.

Recommendations

1. For better interpretability, Logistic Regression may be preferred.
2. For more complex decision boundaries, consider using Decision Trees or SVM.

Conclusion
This analysis compared various classifiers' performance in predicting customer subscriptions to bank term deposits. Each model's strengths and weaknesses were highlighted, providing valuable insights into which algorithms may be best suited for future marketing strategies. Overall, the choice of model should be guided by the specific context of the application, data characteristics, and the importance of interpretability versus prediction accuracy. Future work could involve hyperparameter tuning and exploring ensemble methods to further improve model performance.