20 August 2025

ML Classification algorithm

Best model selection

Ilango P
pailango@gmail.com

Model Performance Summary

The classification results for Support Vector Machine (SVM), Decision Tree, and Random Forest models are presented below. Each model was evaluated using key metrics: **accuracy**, **precision**, **recall**, and **F1-score**.

Evaluation scores:

The performance of the SVM, Decision Tree, and Random Forest models across key metrics are given below:

SVM

	precision	recall	f1-score	support
0	0.77	**0.97**	0.86	79.00
1	**0.90**	0.44	0.59	41.00
accuracy			0.79	120.00
Macro avg	0.83	0.71	0.73	120.00
Weighted avg	0.81	0.79	0.77	120.00

Datatree

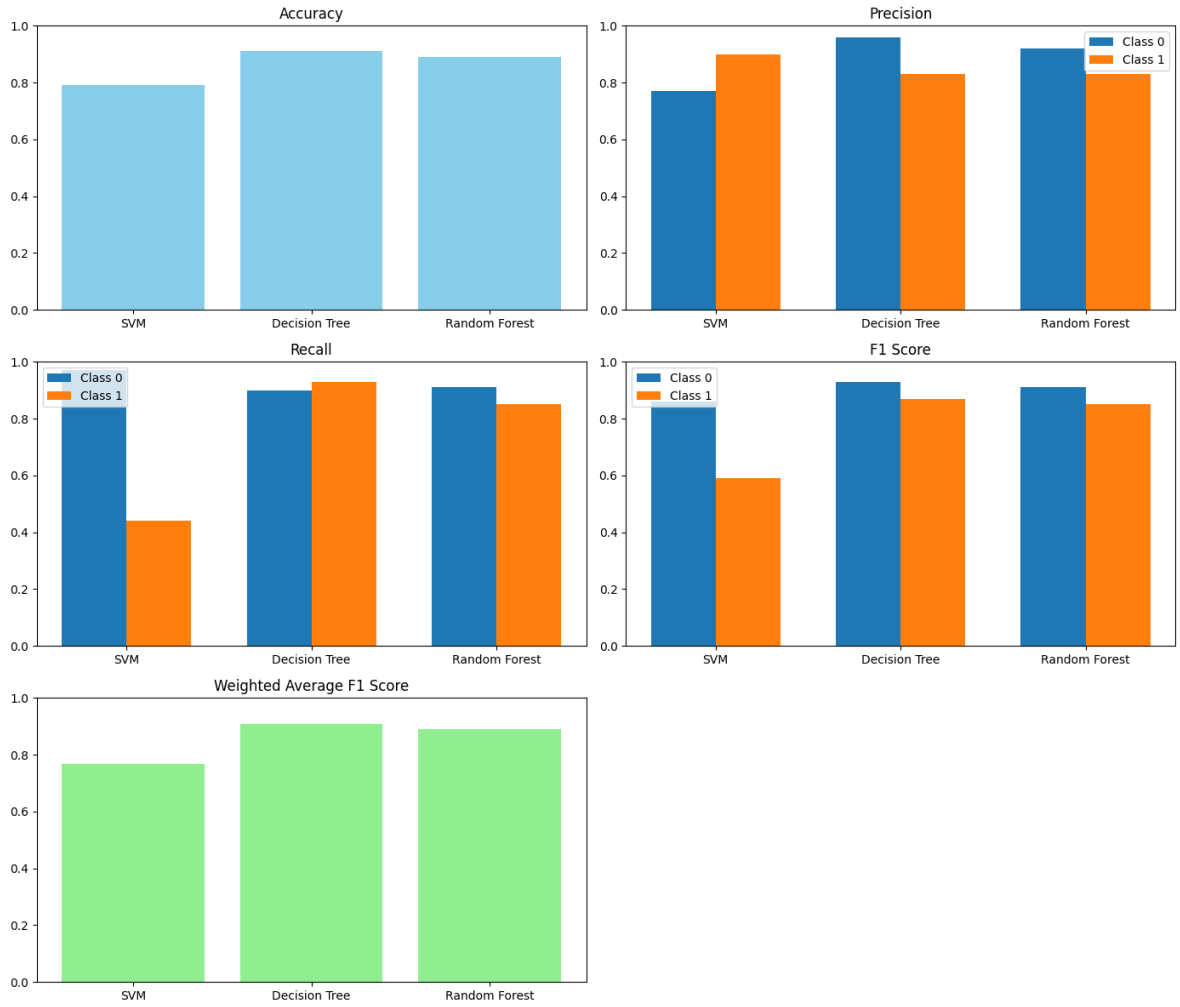
	precision	recall	f1-score	support
0	**0.96**	0.90	**0.93**	79.00
1	0.83	**0.93**	0.87	41.00
accuracy			**0.91**	120.00
Macro avg	0.89	0.91	**0.90**	120.00
Weighted avg	0.91	0.91	**0.91**	120.00

Random forest

	precision	recall	f1-score	support
0	0.92	0.91	0.92	79.00
1	0.83	0.85	0.84	41.00
accuracy			0.89	120.00
Macro avg	0.88	0.88	0.88	120.00
Weighted avg	0.89	0.89	0.89	120.00

Here are the visualizations comparing the performance of the SVM, Decision Tree, and Random Forest models across key metrics:

📊 Model Performance Summary



Accuracy

Which model provides the percentage of correct classification of both [purchased or not purchased] to the total input of the total set?

Accuracy reflects the proportion of correctly classified instances out of the total dataset.

Model	Accuracy
SVM	0.79
Decision Tree	0.91
Random Forest	0.89

Conclusion: The Decision Tree model demonstrates the highest overall accuracy.

Recall

Which model gives the best percentage of correct classification of individual purchased or not to the total input of the item in the test set?

Recall measures the model's ability to correctly identify actual positive cases.

Class	SVM	Decision Tree	Random Forest
0	0.97	0.90	0.91
1	0.44	0.93	0.85

Conclusion: The Decision Tree model excels in identifying positive cases (class 1), making it the most effective in minimizing false negatives.

Precision

Which model gives the best percentage of correct classification of the individual item to sum of prediction [correctly classified and wrongly classified]?

Precision indicates the proportion of correct positive predictions out of all positive predictions made.

Class	SVM	Decision Tree	Random Forest
0	0.77	0.96	0.92
1	0.90	0.83	0.83

Conclusion: The Decision Tree model has the highest precision for class 0, while SVM shows slightly better precision for class 1. However, SVM's low recall undermines its overall effectiveness.

F1-Score

What if the recall value is high and precision value is low. How can we validate the model performance? And which model is the best one?

The F1-score balances precision and recall, offering a more comprehensive view of model performance, especially for imbalanced datasets.

Class	SVM	Decision Tree	Random Forest
0	0.86	0.93	0.91
1	0.59	0.87	0.85

Conclusion: The Decision Tree model achieves the highest F1-score for class 1, indicating a strong balance between precision and recall.

Overall Assessment

The **Decision Tree** model consistently outperforms the others across all key metrics:

- **Highest Accuracy**: 0.91
- **Best Recall for Class 1**: 0.93
- **Strong Precision and F1-Score**: Especially for both classes
- **Balanced Performance**: Across majority and minority classes

While Random Forest is a close contender, the Decision Tree model offers the most reliable and balanced classification performance, making it the preferred choice for this task.
