Recursive Feature Elimination – Report

1.Accuracy Results:

1. RFE with **3 features**

RFE Base Model	Logistic	SVMl	SVMnl	KNN	Naive	Decision	Random
Logistic-RFE	0.94	0.94	0.94	0.94	0.94	0.94	0.94
SVC-RFE	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Random-RFE	0.94	0.94	0.94	0.94	0.90	0.91	0.92
DecisionTree-RFE	0.98	0.98	0.98	0.98	0.79	0.97	0.97

2.RFE with **4 features**

RFE Base Model	Logistic	SVMl	SVMnl	KNN	Naive	Decision	Random
Logistic-RFE	0.95	0.95	0.95	0.95	0.95	0.95	0.95
SVC-RFE	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Random-RFE	0.97	0.97	0.97	0.98	0.87	0.95	0.97
DecisionTree-RFE	0.98	0.98	0.92	0.98	0.81	0.98	0.98

3. RFE with **5 features**

RFE Base Model	Logistic	SVMl	SVMnl	KNN	Naive	Decision	Random
Logistic-RFE	0.98	0.98	0.98	0.98	0.98	0.98	0.98
SVC-RFE	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Random-RFE	0.97	0.97	0.98	0.97	0.91	0.96	0.98
DecisionTree-RFE	0.95	0.98	0.93	0.94	0.85	0.97	0.98

2.Interpretation:

I. **Overall trend**

- Increasing features from **3** → **5** consistently improves accuracy for most classifiers.
- Logistic-RFE improves from **0.94** → **0.98**.
- SVC-RFE improves dramatically from **0.87** → **0.99** (biggest jump).
- RandomForest-RFE stays strong (~0.97–0.98).
- DecisionTree-RFE fluctuates: great with 3 features (0.98), strong with 4, then slightly worse with 5 features in some classifiers.

II. **Naive Bayes is unstable**

- With 3 features, Naive Bayes was okay (0.94 on Logistic-RFE).
- With 4–5 features, accuracy **dropped** under Random/Decision RFE (0.81–0.91).
- This happens because Naive Bayes assumes **feature independence** — the extra features are likely correlated.

III. **Best consistency across classifiers**

- **SVC-RFE with 5 features**: everything hits ~0.99.
- **Logistic-RFE with 5 features**: everything stable at 0.98.
- Both are robust and less sensitive than DecisionTree-RFE.

IV. **Overfitting risk**

- Extremely high accuracy (>0.98 across the board) might indicate possible **overfitting**, especially if the dataset is small.
- Needs validation with **cross-validation or an unseen test set**.

3. Recommendation for Deployment

- **If interpretability matters** (to explain model decisions):
  → Use **Logistic Regression with Logistic-RFE (5 features)**.
  - Accuracy ~0.98.
  - Coefficients easy to explain.
  - Stable across classifiers.
- **If you want best raw accuracy**:
  → Use **SVC-RFE (5 features)** with an **SVM (linear or RBF)**.
  - Accuracy ~0.99.
  - Slightly harder to interpret but excellent generalization.
- **Avoid Naive Bayes** for deployment here → too unstable with more features.
- **Avoid Decision Tree-RFE** as final feature selector → unstable results compared to Logistic/SVC-RFE.

4. Final Recommendation:

For deployment, I'd suggest **SVC-RFE with 5 features** + a **Logistic Regression classifier**.

- You get ~0.99 accuracy.
- Logistic Regression is faster, scalable, and interpretable.
- This balances **performance + interpretability**.