Srinivas University

Name : Adithya Acharya

USN : 01SU24CS011

Branch : CSE

Sem : 4th

Topic : Module 3 Group Task

Algorithm Selection Challenge

1) Spam Detection

- **Most Appropriate: Supervised Learning**

Why?

Spam detection involves classifying emails or messages into:

- **Spam**
- **Not Spam (Ham)**

Since we already have **labeled data** (emails marked as spam or not spam), supervised learning is ideal.

How it Works:

1. Collect previously labeled emails.
2. Extract features (keywords, sender info, links, frequency).
3. Train a classification model.
4. Predict whether a new email is spam.

Common Algorithms:

- Logistic Regression
- Naïve Bayes
- Support Vector Machines
- Random Forest
- Neural Networks

Why Not Others?

- Unsupervised: No labels available (not suitable here).
- Reinforcement: No reward-based interaction.
- Semi-supervised: Used only if labeled data is limited.

Conclusion: Spam detection is a **classification problem with labeled data**, so **Supervised Learning** is best.

---

2) Market Segmentation

- **Most Appropriate: Unsupervised Learning**

Why?

Market segmentation groups customers based on behavior, age, income, purchase patterns, etc., without predefined labels.

There are no fixed categories beforehand — we want the algorithm to discover patterns.

How it Works:

1. Collect customer data.
2. Identify patterns in purchasing behavior.
3. Group customers into clusters.

Common Algorithms:

- K-Means Clustering
- Hierarchical Clustering
- DBSCAN

Why Not Others?

- Supervised: No labeled categories like "Type A customer."
- Reinforcement: No sequential decision-making process.
- Semi-supervised: Only useful if some labeled data exists.

Conclusion: Market segmentation requires grouping without labels → **Unsupervised Learning**.

---

3) Self-Driving Navigation

- **Most Appropriate: Reinforcement Learning**

Why?

Self-driving cars must:

- Make continuous decisions
- Learn from environment feedback
- Optimize actions (speed, steering, braking)

This fits reinforcement learning because:

- The car (agent) interacts with the environment.
- It receives rewards (safe driving, reaching destination).
- It receives penalties (collision, traffic violation).

Companies like Tesla use AI-based learning for autonomous driving systems.

How it Works:

1. Agent observes environment.
2. Takes action.
3. Receives reward/penalty.
4. Updates strategy to maximize long-term reward.

Why Not Others?

- Supervised: Real-time decisions cannot rely only on labeled data.
- Unsupervised: No reward mechanism.
- Semi-supervised: Not suitable for continuous control tasks.

Conclusion: Self-driving navigation requires learning through interaction → **Reinforcement Learning**.

---

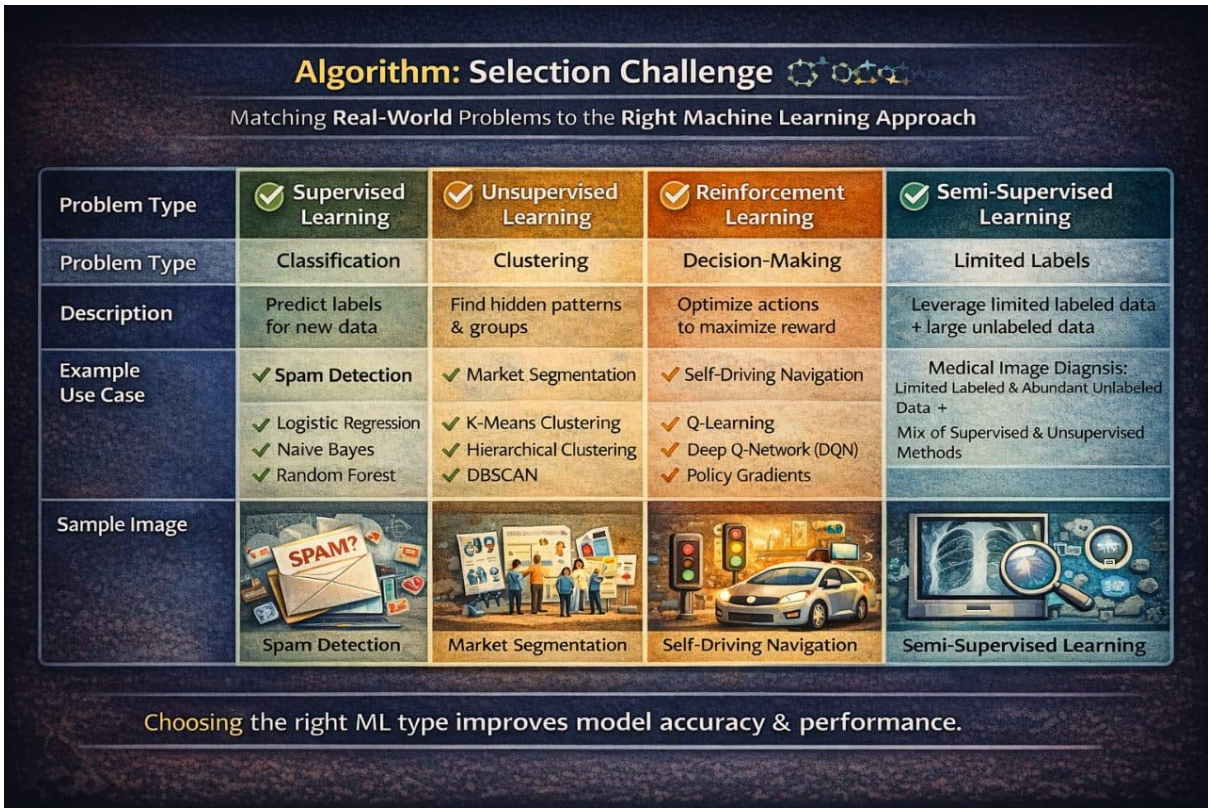
4) Bonus Case: When to Use Semi-Supervised Learning?

Semi-supervised learning is useful when:

- Large amount of unlabeled data
- Small amount of labeled data

Example:
Medical image diagnosis where labeling is expensive.

It combines advantages of supervised and unsupervised learning.



Summary Comparison Table

Problem	**ML Type**	**Reason**
Spam Detection	Supervised	Labeled classification problem
Market Segmentation	Unsupervised	Discover hidden customer groups
Self-Driving Navigation	Reinforcement	Decision-making with reward feedback

Problem	ML Type	Reason
Medical Image Analysis (few labels)	Semi-Supervised	Limited labeled data

---

Final Conclusion

Choosing the correct ML algorithm depends on:

- ✓ Availability of labeled data
- ✓ Type of problem (classification, clustering, decision-making)
- ✓ Nature of environment (static or interactive)

Correct algorithm selection improves:

- Accuracy
- Efficiency
- Model performance