



# Assignment 2 (Group Assignment: Unsupervised Learning - Coding) Rubric

Course: Fall 2025 Machine Learning (DAMO-640-10)

Criteria	Excellent (4) 20 points	Satisfactory (3) 15 points	Needs Improvement (2) 10 points	Poor (1) 5 points	Criterion Score
EDA & Preprocessing	Data loaded correctly; clear stats & distribution; scaling flawless; random_state used.	Data & scaling done; basic commentary.	Minor errors in loading or scaling; sparse notes.	Missing or incorrect preprocessing.	/ 20
Ensemble Methods	RF & GBM trained; both n_estimators settings CV'd; results clearly tabulated & interpreted.	Both methods tested; CV results reported.	One method or one setting missing; table sparse.	Ensemble models missing or nonfunctional.	/ 20
SVM Hyperparameter Comparison	Linear & RBF SVM trained; both C settings CV'd; comparison concise and correct.	Both kernels tested; CV results given.	One kernel or one C missing; limited analysis.	SVM missing or fails to run.	/ 20
Clustering & Evaluation	K-Means & DBSCAN experiments done; silhouette scores computed; PCA plot & cluster analysis insightful.	Clustering applied; scores reported.	Clustering attempted; metrics incomplete.	Clustering missing or incorrect.	/ 20
Test Evaluation & Reporting	Best ensemble & SVM evaluated; all metrics + ROC AUC; ROC curves overlaid; justified choice.	Metrics & ROC curves present; basic discussion.	Some metrics or ROC missing; shallow discussion.	Evaluation/report missing or incoherent.	/ 20

**Dataset:**

**UCI Wholesale Customers Data:**

Total

<https://archive.ics.uci.edu/dataset/292/wholesale+customers>

/ 100

**Features:  
Overall Score**

- Fresh, Milk, Grocery, Frozen, Detergents\_Paper, Delicatessen

**Excellent (4)**

80 points minimum

**Satisfactory (3)**

70 points minimum

**Needs Improvement (2)**

60 points minimum

**Poor (1)**

0 points minimum

- Channel (0 or 1)

**Tasks****1. Data Loading & Preprocessing**

- Load the CSV (headers included).
- Display head, summary statistics, and channel distribution.
- Standardize all six features (use StandardScaler).

**2. Ensemble Methods (Module 5)**

- Bagging: Train a Random Forest classifier with `n_estimators = 50` and `200`; 5-fold CV on the training set to record mean accuracy.
- Boosting: Train a Gradient Boosting classifier with `n_estimators = 50` and `200` (`learning_rate=0.1`); 5-fold CV to record mean accuracy.
- Tabulate and compare CV results.

**3. Support Vector Machine (Module 6)**

- Train an SVM with linear and RBF kernels.
- For each kernel, compare  $C = 0.1$  and  $C = 1.0$  via 5-fold CV accuracy.
- Tabulate results.

**4. Clustering (Module 7)**

- K-Means: Run for k=2,3,4; compute silhouette score for each k; choose best k and plot cluster centers (in 2D via PCA projection).
- DBSCAN: Experiment with eps = 0.5 and 1.0 (`min_samples=5`); report number of clusters (excluding noise) and silhouette score for each.

## 5. Final Evaluation

- Split data 70/30 (`random_state=42`).
- For your best ensemble and best SVM (by CV), evaluate on the test set:
  - Accuracy, precision, recall, F<sub>1</sub>-score, ROC AUC
- Plot ROC curves of both models together and briefly discuss which you'd deploy and why.

---

## Deliverables

- A Jupyter notebook (`.ipynb`) containing all code, outputs, plots, and inline comments.
- A short report ( $\leq 3$  pages PDF) with:
  - **Introduction:** Dataset & objectives
  - **Methods:** Models, hyperparameters, clustering choices
  - **Results:** CV tables, test metrics, ROC plot, clustering evaluation
  - **Discussion:** Insights and final recommendation

Due on Nov 16, 2025 11:59 PM

Available on Nov 2, 2025 12:01 AM. Access restricted before availability starts.

Available until Nov 16, 2025 11:59 PM. Access restricted after availability ends.

▼ Hide Rubrics

**Note:** The activities evaluation will be applied to all members of this group.

**Rubric Name: Assignment 2 (Group Assignment: Unsupervised Learning - Coding) Rubric**

# Submit Assignment

Files to submit

**(0) file(s) to submit**

After uploading, you must click Submit to complete the submission.

Add a File

Record Audio

Record Video

Comments