2024/11/19

1. 比較: RandomForest, GradientBoosting, XGBoost
2. Best model: RandomForest 但分數只有0.7左右
3. Model on the validation set Accruacy=0.6998

Environment setup

[Installing Packages](https://packaging.python.org/en/latest/tutorials/installing-packages/)

[Project Jupyter | Installing Jupyter](https://jupyter.org/install)

[Jupyter Notebook 7.3.0b2 documentation](https://jupyter-notebook.readthedocs.io/en/latest/)

VotingClassifier: majority vote or average predicted probabilities (with weighting or not)

class set={a,b,c,...}

Multiclass(class=[a] vs Multilablel (class=[a,b,c])

Our dataset is multiclass

黃正鵬[RandomForestClassifier — scikit-learn 1.5.2 documentation](https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.RandomForestClassifier.html)

林滋隆[GradientBoostingClassifier — scikit-learn 1.6.dev0 documentation](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html)

XGBoost

康峻瑋[Fpgrowth - mlxtend](https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/fpgrowth/)

許茗鈞[pyECLAT · PyPI](https://pypi.org/project/pyECLAT/)

[1.11. Ensembles: Gradient boosting, random forests, bagging, voting, stacking — scikit-learn 1.5.2 documentation](https://scikit-learn.org/1.5/modules/ensemble.html)