

## Final project 說明

請使用 GDSC cell line gene expression 來建立機器學習模型，預測 16 個 patients 的藥物二元分類問題(resistant, sensitive)，並以 16 個 patients 的 labels 驗證模型的準確率，資料夾包含 2 個檔案 (GDSC\_TCGA\_Docetaxel.csv 和 Patients\_TCGA\_Docetaxel.csv)。

GDSC\_TCGA\_Docetaxel.csv 檔案包含 850 個 GDSC cell lines gene expression levels 及 IC50。TCGA\_Docetaxel 的最大用藥濃度為 0.0125，若 IC50 小於最大用藥濃度，則為 S；若 IC50 大於最大用藥濃度，則為 R。

Test set : Patients\_TCGA\_Docetaxel.csv 檔案包含 16 個 patients' gene expression levels 及 labels。