

## Manufacturing Data Science 製造數據科學

### Assignment 4

Due Date: Dec. 22, 2023

Please solve the following questions and justify your answer. **Show all your analysis result including equation/calculation or Python code in your report.** Upload your “zip” file including MSWord/PDF report and Python code with 檔名: MDS\_Assignment4\_ID\_Name.zip” to NTU COOL by due. The late submission is not allowed.

#### 1. (50%) Data Imbalance Problem

在 Kaggle 開放數據中包含了一個信用卡詐欺數據(Credit Card Fraud Detection, <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>), 信用卡公司能夠識別詐欺性信用卡交易非常重要, 這樣客戶就不會因未購買的商品而被收取費用。此資料集包含歐洲持卡人 2013 年 9 月透過信用卡進行的交易。該資料集顯示了兩天內發生的交易, 其中 284,807 筆交易中有 492 筆詐欺。資料集高度不平衡, 正類(詐欺)佔所有交易的 0.172%。試著參考網路資源學習並撰寫程式, 使用此數據回答下列問題。

請閱讀此文章, ”Best techniques and metrics for Imbalanced Dataset”

<https://www.kaggle.com/code/marcinrutecki/best-techniques-and-metrics-for-imbalanced-dataset/notebook>.

- (a) (15%) 試在該數據分析流程中加入數據平衡的步驟, 使用三種方法來進行數據平衡(e.g. 使用上抽樣、下抽樣或是代價敏感學習等)。
- (b) (10%) 承(a), 試比較三種方法? 建議選用哪種方法最為合適? 為什麼?
- (c) (10%) 對於數據多數群與少數群的比例應當調整至多少? 為什麼? 試透過調整生成比率(i.e. 敏感度分析)來看模型分類結果。(提示: 將敏感度分析以繪圖呈現兩條曲線, x 軸為不同生成比例、y 軸為偽陽性率(false positive rate)與偽陰性率(false negative rate))
- (d) (10%) 試說明特徵挑選步驟應於數據平衡前或後, 這對預測結果有何影響?
- (e) (5%) 就您觀點簡短敘述(不超過 200 字), 您覺得這文章”Best techniques and metrics for Imbalanced Dataset”還有什麼需要補充或改進的地方。

#### 2. (50%) Metaheuristic Algorithm

試求解工單式生產排程(job shop scheduling)問題, 最小化最大完工時間(makespan)。資料來源參考 Taillard's 排程案例

(<http://mistic.heig-vd.ch/taillard/problems.dir/ordonnancement.dir/ordonnancement.html>) , 其 tai20\_15 包含 20 個工件(job)與 15 個機台(machine), 每個工件有 15 個加工作業(operation), 工件加工時間與工件加工機台順序如下表所示。

- (a) (10%) 試以基因演算法求解並說明其設計細節, 包含染色體編碼、交配、突變、選擇、與相關參數(e.g. 母體大小、交配率、突變率等);
- (b) (10%) 若交配突變產生不可行的染色體, 試建議染色體修復方法(repair)。

- (c) (20%) 撰寫相關程式求解，並呈現出最佳解與求解時間；  
(d) (10%) 試繪製演算法收斂圖(e.g. x 軸為迭代、y 軸為適應值)，包含歷史迭代中最佳適應值與目前迭代回合的最佳適應值兩條曲線。(參考文獻 Taillard, E. (1993). “Benchmarks for basic scheduling problems,” *European Journal of Operational Research*, 64(2), 278-285.)

(Hint: <https://github.com/PO-LAB/Operations-Research-Applications> ,

基因演算法於生產排程 1 或基因演算法於生產排程 2)

(Dataset: 在這個連結中

[http://mistic.heig-vd.ch/taillard/problemes.dir/ordonnancement.dir/jobshop.dir/tai20\\_15.txt](http://mistic.heig-vd.ch/taillard/problemes.dir/ordonnancement.dir/jobshop.dir/tai20_15.txt)

最上面的第一個資料集)

時間	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12	O13	O14	O15
J1	25	75	75	76	38	62	38	59	14	13	46	31	57	92	3
J2	67	5	11	11	40	34	77	42	35	96	22	55	21	29	16
J3	22	98	8	35	59	31	13	46	52	22	18	19	64	29	70
J4	99	42	2	35	11	92	88	97	21	56	17	43	27	19	23
J5	50	5	59	71	47	39	82	35	12	2	39	42	52	65	35
J6	48	57	5	2	60	64	86	3	51	26	34	39	45	63	54
J7	40	43	50	71	46	99	67	34	6	95	67	54	29	30	60
J8	59	3	85	6	46	49	5	82	18	71	48	79	62	65	76
J9	65	55	81	15	32	52	97	69	82	89	69	87	22	71	63
J10	70	74	52	94	14	81	24	14	32	39	67	59	18	77	50
J11	18	6	96	53	35	99	39	18	14	90	64	81	89	48	80
J12	44	75	12	13	74	59	71	75	30	93	26	30	84	91	93
J13	39	56	13	29	55	69	26	7	55	48	22	46	50	96	17
J14	57	14	8	13	95	53	78	24	92	90	68	87	43	75	94
J15	93	92	18	28	27	40	56	83	51	15	97	48	53	78	39
J16	47	34	42	28	11	11	30	14	10	4	20	92	19	59	28
J17	69	82	64	40	27	82	27	43	56	17	18	20	98	43	68
J18	84	26	87	61	95	23	88	89	49	84	12	51	3	44	20
J19	43	54	18	72	70	28	20	22	59	36	85	13	73	29	45
J20	7	97	4	22	74	45	62	95	66	14	40	23	79	34	8

機台	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12	O13	O14	O15
J1	4	12	15	2	11	3	5	8	1	13	6	10	7	14	9
J2	6	1	4	9	5	2	13	15	7	8	11	3	10	14	12
J3	3	4	15	1	10	13	6	5	8	11	9	12	14	2	7
J4	9	11	2	14	4	5	15	10	3	6	12	8	1	7	13
J5	15	9	2	3	11	10	13	5	7	6	1	14	4	12	8
J6	4	11	2	6	7	1	9	8	12	14	3	15	13	10	5

J7	3	11	2	13	9	1	8	7	15	14	5	4	6	10	12
J8	2	1	3	5	8	14	12	4	13	6	7	15	10	9	11
J9	5	6	10	11	8	7	3	2	13	4	14	1	9	15	12
J10	2	5	4	11	15	1	7	14	12	9	6	13	8	10	3
J11	4	11	2	1	10	9	15	7	5	8	3	13	6	12	14
J12	3	8	7	9	4	6	15	5	2	1	10	11	14	12	13
J13	1	8	15	9	13	11	10	4	7	2	5	3	12	14	6
J14	13	4	10	5	2	1	11	7	6	3	15	14	8	9	12
J15	4	15	7	6	14	10	2	1	13	8	3	5	11	9	12
J16	6	15	7	13	9	3	5	10	12	14	4	2	8	1	11
J17	4	8	11	15	1	9	2	12	6	14	5	13	7	10	3
J18	11	9	3	12	14	7	15	4	10	8	5	6	13	1	2
J19	4	3	13	14	2	7	15	6	5	9	10	12	1	11	8
J20	12	15	6	7	11	10	14	2	5	9	1	4	13	3	8

**Note**

1. Show all your work in detail. **Innovative idea is encouraged.**
2. If your answer refers to any external source, please “must” give an academic citation. Any “plagiarism” is not allowed.