ONLINE TECHNICAL ASSESSMENT

CLASSIFICATION

BASED ON PERSONALIZATION

SUPERVISED BY: ASSOCIATE PROF. DR. NOR LIYANA BT MOHD SHUIB

PREPARED BY: LIM ZHENG YU (U2102809/1)

INTRODUCTION

The topic is focused on students in these domains: **Computer Science, Technology, Engineering.**

Programming certification examinations such as the Sun's Java Certification Examination and Novell's certification examinations are **multiple-choice** (Roberts et al., 2003), serving no significant purpose to evaluate programming students.

Traditional assessments are **standardized**, making them "indirect and inauthentic" (Bailey, 1998), also meaning that they are effective in measuring students' capabilities at one point in time but **ineffective in informing student progression** (Dikli, 2003).

The **VAK (Visual-Auditory-Kinesthetic) Learning Styles Model** describes that students are **different in showcasing their skills** and **absorbing information**. Hence, it would be fairer to evaluate every student differently too.

INTRODUCTION

VAK Learning Styles (Barbe et al., 1979)

Visual

- Learn by seeing
- Learn using graphs, posters, etc.
- Tend to look up while thinking (Pritchard, 2009)

Auditory

- Learn by listening
- Learn through discussions, lectures, stories, etc.
- Tend to tilt their heads and use eye movements when thinking (Pritchard, 2009)

Kinesthetic

- Learn by doing
- Learn through physical activity and touch
- Hard to sit still and takes lots of breaks when learning

LITERATURE REVIEW

Reference	Domain	Sample Size	Method	Learning Style	Evaluation	Limitations
Garcia et al. (2005)	Learning Object	10 computer science and engineering students	Bayesian Networks	Felder-Silverman	• More than 80% accuracy	• Too few training examples

Ulloa-Cazarez et al. (2018)	Online Students Performance	245 student grades records and log records	Genetic Programming		• GP
					  ◦ MAE = 14.38
					  ◦ MAD = 9.80
					• LR
					  ◦ MAE = 17.49
					  ◦ MAD = 16.53
						• Independent variables unused
						• Hard to acquire relevant data

Cetinkaya et al. (2023)	Programming Test Performance	600 secondary school students	1. SVM 2. Decision Tree 3. KNN 4. Quadratic Discriminant		• 80.8% to 94.8% accuracy • All scores of SVM are above 90% (Kappa, Precision, Recall, F1)	• Findings from study cannot be generalized globally as study was conducted in a certain region

LITERATURE REVIEW

Reference	Domain	Sample Size	Method	Learning Style	Evaluation	Limitations
Dema (2021)	Programming Test Performance	40 Engineering in IT students		VAK/VARK		• No machine learning • Only descriptive statistics • No personalization
Ocepek et al. (2013)	Multimedia Type	272 undergraduates	Multi-target regression tree	Kolb, Rancourt, hemispheric, VAK		• No evaluation on performance of model • Classifcation on multimedia type instead of assessment type
Seyal et al. (2015)	Programming Test Performance	70 Internet Computing undergraduates	Chi-square test	Kolb	• Learning styles have significant influence on academic performance	• No classification model on suitable type of assessment

PROBLEM STATEMENT

There is **limited study** on the **classification of online technical assessments** for students with **technical background**.

Most research focus on **performance** instead of **suitable assessments** based on students' **VAK learning styles**.

OBJECTIVES

Objective 1

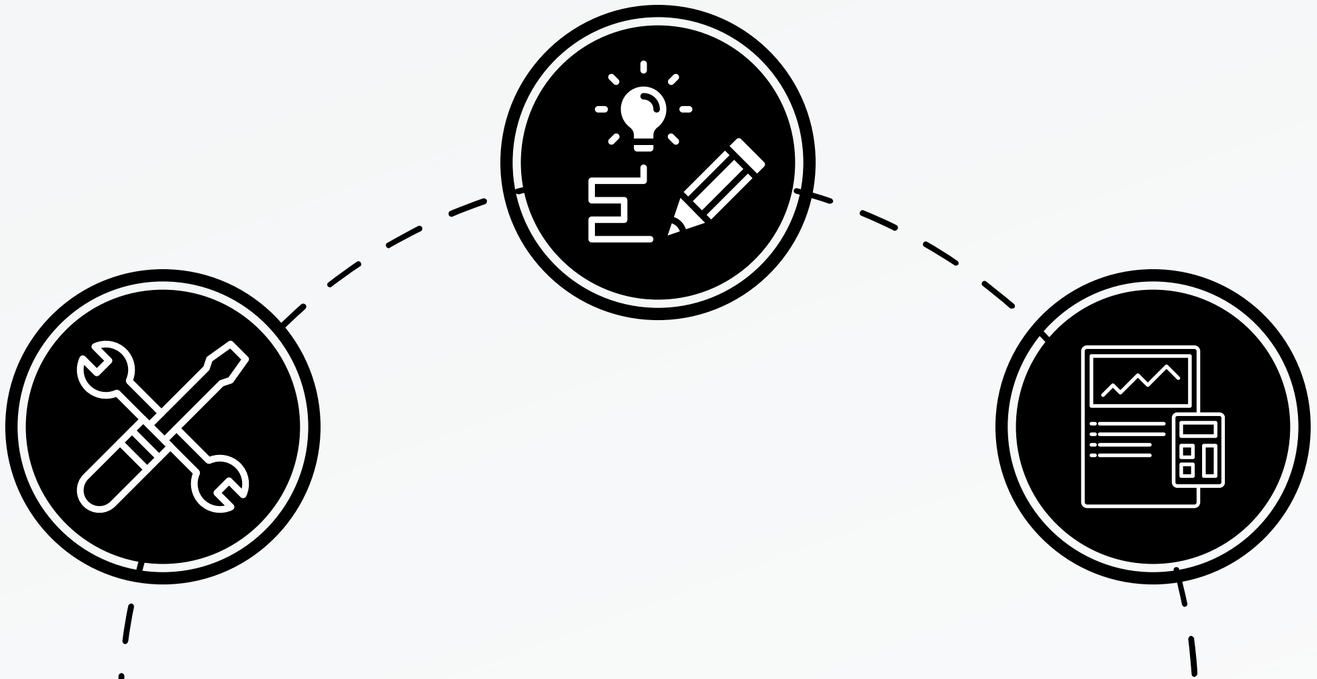
To develop a classification model of online technical assessment based on personalization

Objective 2

To evaluate the performance of the classification model of online technical assessment based on personalization

Objective 3

To develop a data product of the classification model of online technical assessment based on personalization



TOOLS

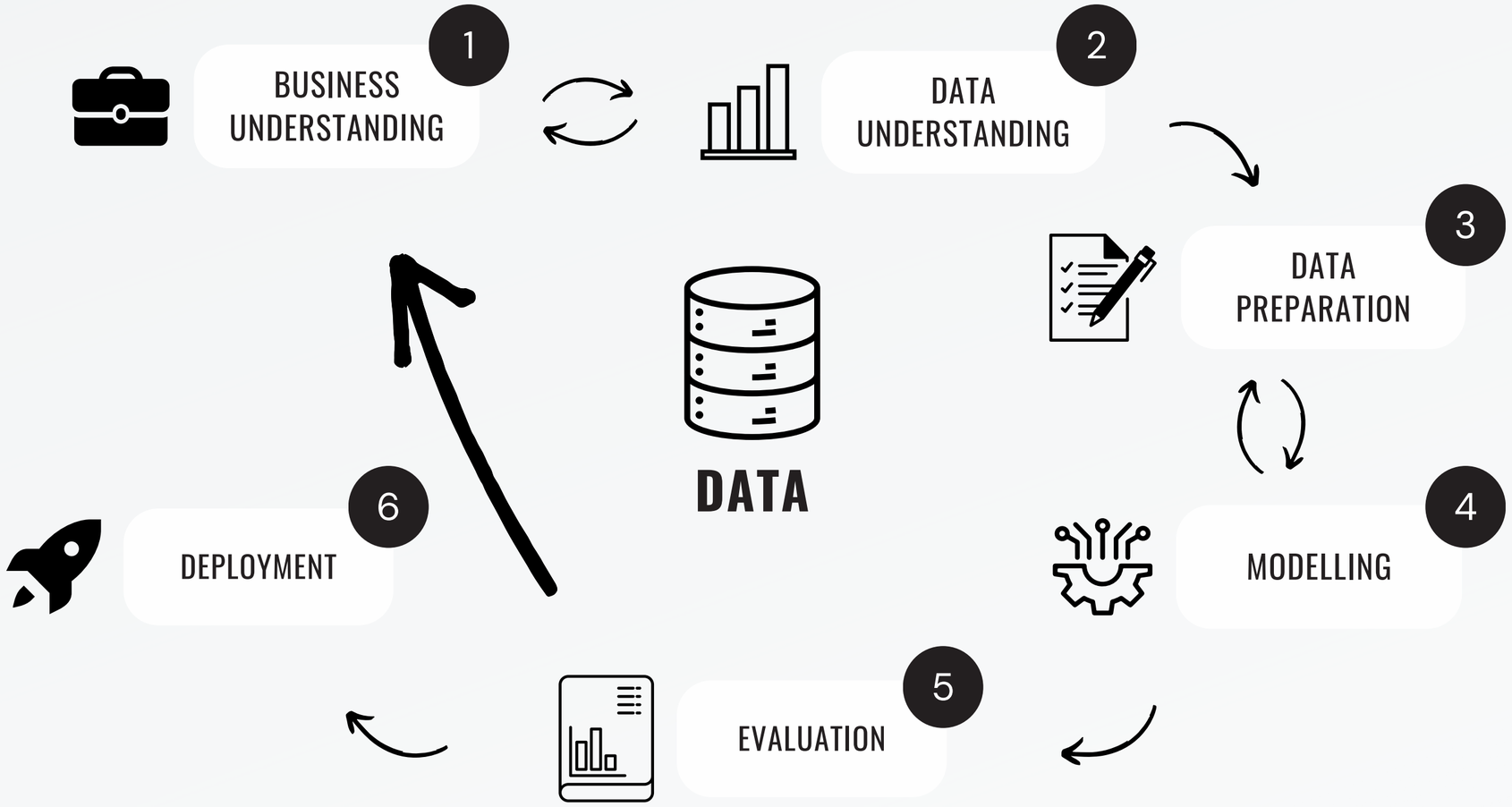
Mainly used for:

- Exploratory Data Analysis
- Data Cleaning
- Data Transformation
- Modelling

and more...

DATA SCIENCE METHODOLOGY

CRISP-DM



1. BUSINESS UNDERSTANDING

- Understanding the **objectives** and **requirements** of the project
  - **Understand the dataset** briefly
  - Determine **business objectives**
    - To develop a classification model for online technical assessments
    - To evaluate said classification model
    - To develop a data product from the classification model
  - **Produce project plan**
    - Plan entire workflow
    - Create schedules

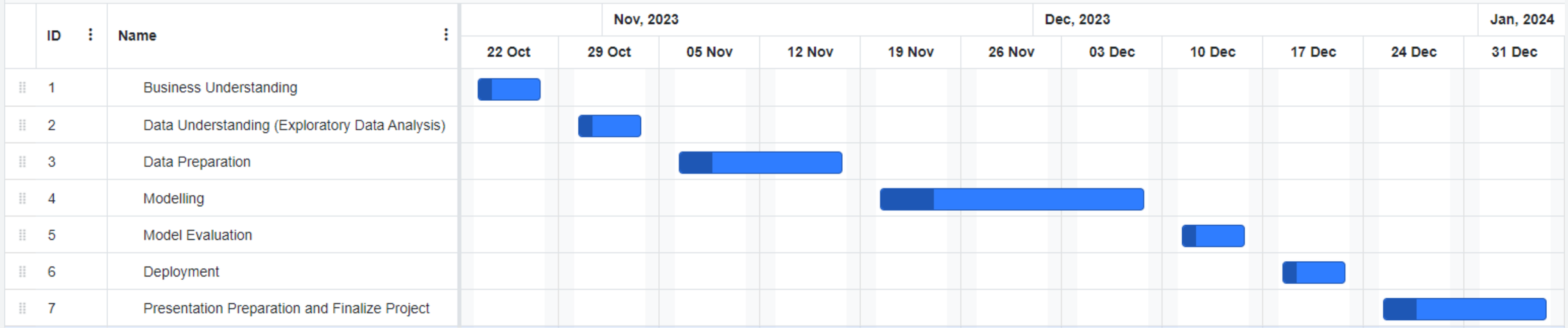
1. BUSINESS UNDERSTANDING

Brief Overview of Dataset

Timestam	Email Adc	I understa	Gender	Level of St	Field of st	Institution	Country	Househol	Preferred I	Preferred S	Preferred (	Difficulties	Learning (	Learning (	Learning (	Learning (	Learning (	Learning (	Learning (	L
########	test@gma	Agree	Female	Undergrac	Veterinary	test	test	RM 3001 -	Face to Fa	Twitter	Whatsapp	Adaptabil	Somewha	Somewha	Somewha	Somewha	Somewha	Somewha	Somewha	S
########	liyanashu	Agree	Female	Postgradu	Computin	UM	Malaysia	RM 3001 -	Face to Fa	Instagram	Whatsapp	Technical	Somewha	Somewha	Somewha	Somewha	Somewha	Somewha	Somewha	S
########	azirasuho	Agree	Female	Postgradu	Computin	UM	Malaysia	RM 3001 -	Face to Fa	Facebook	Email, Un	Technical	Very Much	Somewha	Very Much	Somewha	Undecide	Somewha	Somewha	U
########	haslina_n	Agree	Female	Postgradu	3:00	University	Malaysia	RM 10 001	Face to Fa	Facebook	Email, Wh	Adaptabil	Very Much	Very Much	Very Much	Somewha	Somewha	Undecide	Somewha	U
########	noorain27	Agree	Female	Postgradu	Humanitie	UM	Malaysia	RM 10 001	Face to Fa	Facebook	Whatsapp	Technical	Very Much	Somewha	Very Much	Somewha	Somewha	Somewha	Very Much	S
########	vivianteyO	Agree	Female	Undergrac	Sports	University	Malaysia	RM 3001 -	Face to Fa	Facebook	Email, Un	Adaptabil	Somewha	Undecide	Very Much	Not Really	Somewha	Somewha	Undecide	V
########	norsyazwa	Agree	Female	Undergrac	Sports	Universiti	Malaysia	Less than	Face to Fa	Instagram	University	Adaptabil	Very Much	Somewha	Very Much	Very Much	Very Much	Somewha	Undecide	V
########	helihafisa	Agree	Female	Undergrac	Sports	UM	Malaysia	Less than	Face to Fa	Facebook	Whatsapp	Technical	Very Much	Somewha	Very Much	Undecide	Not Really	Very Much	Very Much	V
########	ainimusta	Agree	Female	Undergrac	Sports	University	Malaysia	Less than	Face to Fa	Instagram	Email, Un	Technical	Very Much	Very Much	Very Much	Very Much	Very Much	Somewha	Somewha	S
########	piaabalqi	Agree	Female	Undergrac	Sports	Univeriti M	Malaysia	Less than	Face to Fa	Facebook	Email, Wh	Adaptabil	Very Much	Undecide	Very Much	Undecide	Very Much	Very Much	Very Much	V
########	a.halili_9	Agree	Female	Undergrac	Sports	University	Malaysia	Less than	Face to Fa	Instagram	Email, Wh	Technical	Somewha	Undecide	Very Much	Somewha	Somewha	Very Much	Very Much	V
########	sitinorash	Agree	Female	Undergrac	Sports	University	Malaysia	RM 3001 -	Face to Fa	Twitter, In	Email, Wh	Technical	Undecide	Not at All	Very Much	Somewha	Somewha	Somewha	Very Much	V
########	kiewan2@	Agree	Male	Undergrac	Sports	University	Malaysia	RM 3001 -	Face to Fa	Instagram	Email, Wh	Adaptabil	Somewha	Not Really	Somewha	Not Really	Somewha	Somewha	Somewha	U
########	izzuwansu	Agree	Male	Undergrac	Sports	University	Malaysia	RM 3001 -	Face to Fa	Instagram	Email, Wh	Technical	Very Much	Very Much	Very Much	Somewha	Very Much	Very Much	Somewha	V
########	nurulaisya	Agree	Female	Undergrac	Sports	Uni malay	Malaysia	Less than	Asynchror	Youtube	Whatsapp	Technical	Somewha	Not at All	Very Much	Undecide	Somewha	Very Much	Very Much	S
########	syafiqiqba	Agree	Male	Undergrac	Sports	University	Malaysia	Less than	Face to Fa	Instagram	Email, Un	Adaptabil	Undecide	Not Really	Not Really	Not Really	Not Really	Not Really	Not Really	N

1. BUSINESS UNDERSTANDING

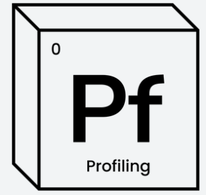
Gantt Chart

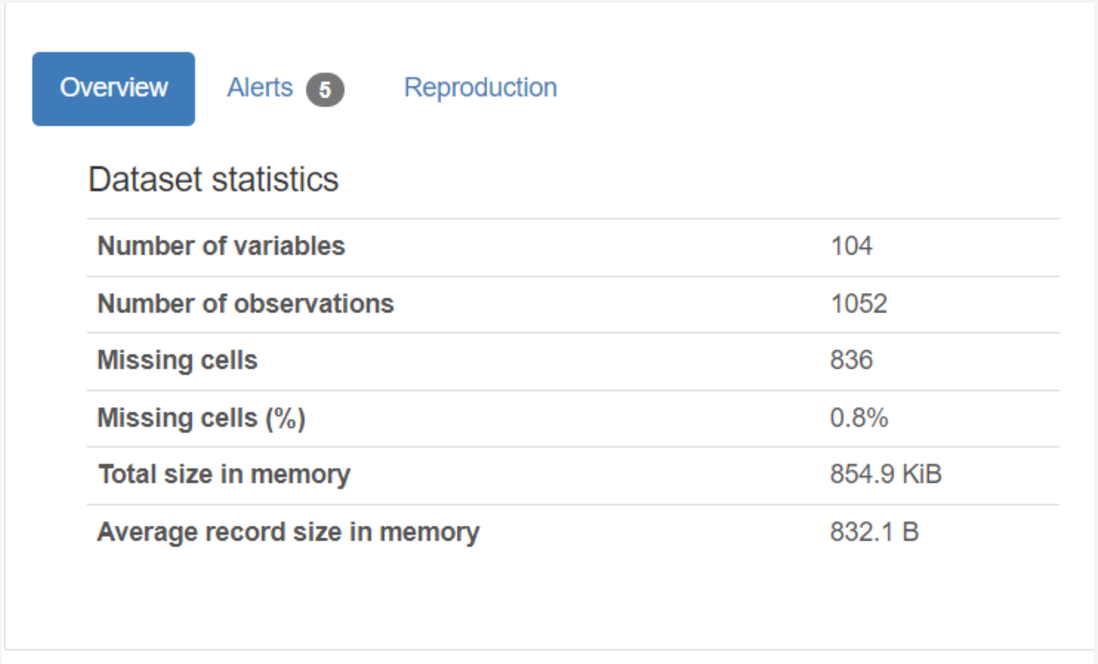


2. DATA UNDERSTANDING

ProfileReport

```
report = ProfileReport(df_, title="Quick EDA", minimal=True, html={"style": {"full_width": True}})
report.to_file("report.html")
```





Overview | Alerts 5 | Reproduction

Dataset statistics

Number of variables	104
Number of observations	1052
Missing cells	836
Missing cells (%)	0.8%
Total size in memory	854.9 KiB
Average record size in memory	832.1 B

Key Takeaways:

- Data from Jan 2021 to Nov 2023
- 104 variables
- 1052 rows
- 836 missing values

2. DATA UNDERSTANDING

ProfileReport

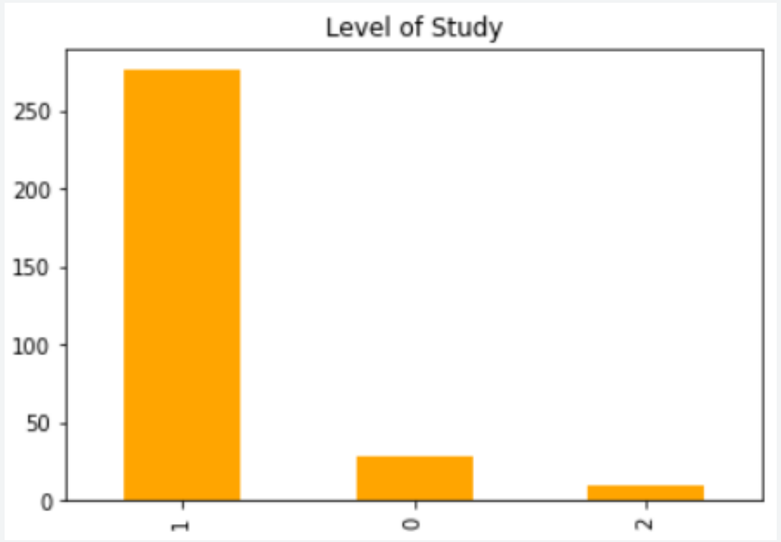
Timestamp

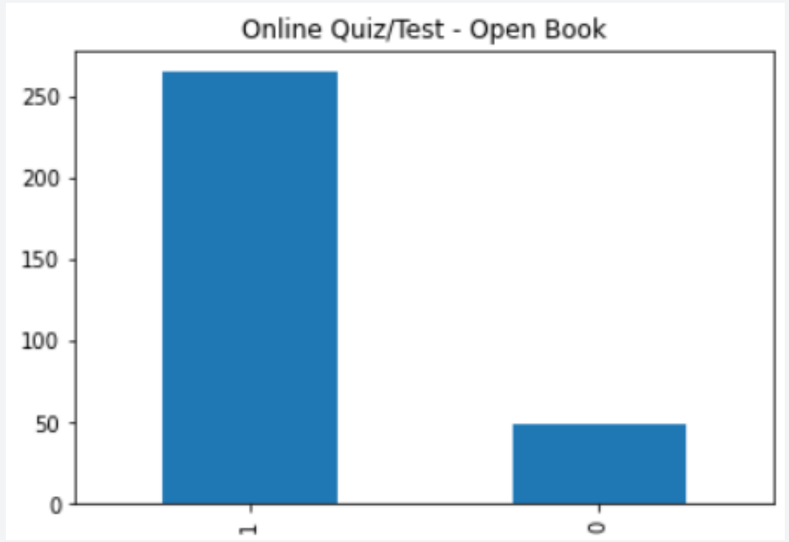
Text

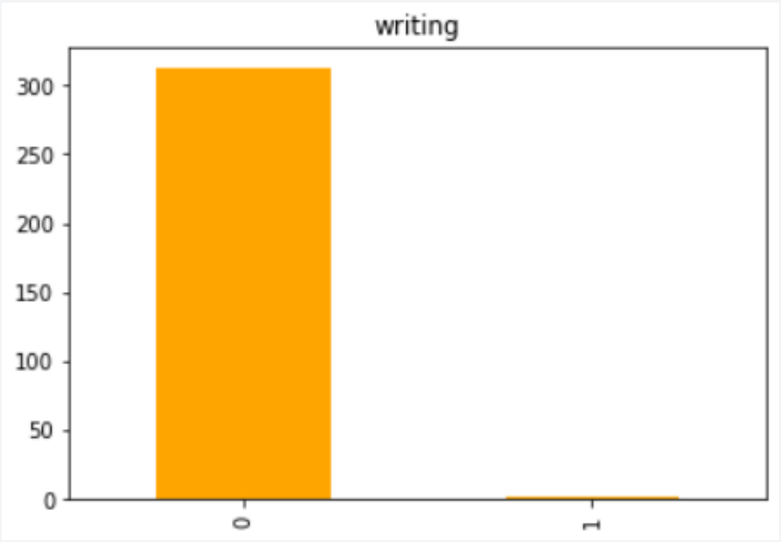
Distinct	756
Distinct (%)	71.9%
Missing	0
Missing (%)	0.0%
Memory size	8.3 KiB

2. DATA UNDERSTANDING

Imbalanced Data







Workaround: Compress minority labels/ variables into one

2. DATA UNDERSTANDING

Dirty Data (Multiple Choice Selections)

```
df["Preferred Social Media Platform "].value_counts()
```

Youtube	59
Facebook, Instagram, Youtube	41
Instagram, Youtube	35
Twitter, Instagram, Youtube	34
Facebook, Youtube	27
Instagram	16
Twitter, Youtube	15
Twitter	10
Twitter, Instagram	10
Facebook	7
Blogger/Wordpress, Youtube	5
Facebook, Instagram	5
Facebook, Twitter, Instagram, Youtube	5
Facebook, Twitter, Instagram, Blogger/Wordpress, Youtube	4

```
df["Preferred Communication Platform"].value_counts()
```

Email, Whatsapp, Telegram	58
Whatsapp	45
Whatsapp, Telegram	42
Email, Whatsapp	38
University eLearning Chat Room, Whatsapp	16
Email, University eLearning Chat Room, Whatsapp, Telegram	13
Email, University eLearning Chat Room, Whatsapp	13
Email, Whatsapp, Call	11
University eLearning Chat Room, Whatsapp, Telegram	10
Email, University eLearning Chat Room, Whatsapp, Call, Telegram	7
Email, Telegram	7
Telegram	6

2. DATA UNDERSTANDING

Solution

```
def one_hot_encode_multiple_choice(col, suffix):
  a = set(', '.join(df[col]).split(', '))
  b = list(a)

  for i in b:
    new_col = str(i) + "_" + suffix
    df[new_col] = df[col].apply(lambda x: 1 if i in x else 0)

  df.drop(col, axis=1, inplace=True)
```

Output

```
one_hot_encode_multiple_choice("Preferred Social Media Platform ", "prefsocmed")
```

Reddit_prefsocmed	Google Classroom _prefsocmed	Tiktok_prefsocmed	google meet_prefsocmed	Twitter_prefsocmed	Google classroom_prefsocmed	Telegram and Google Classroom_prefsocmed
0	0	0	0	0	0	0
0	0	0	0	1	0	0
0	0	0	0	1	0	0
0	0	0	0	0	1	0
0	0	0	0	1	0	0

3. DATA PREPARATION

Finding Missing Values

```
[('I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. I,
hereby agree to take part in the above study.',
  16),
 ('Institutions', 2),
 ('Please share any comments or suggestions related to this issue. Thank You',
  760),
 ('Faculty', 29),
 ('Department', 29)]
```

These columns are unused. They will be dropped before modelling stage.

3. DATA PREPARATION

Label Encoding

```
le = LabelEncoder()
def LabelEncoding(column):
  print(df[column].value_counts())
  print()
  df[column] = le.fit_transform(df[column])
  print(df[column].value_counts())
```


```
LabelEncoding("Gender")
Female    699
Male      352
Name: Gender, dtype: int64

0    699
1    352
Name: Gender, dtype: int64
```

```
LabelEncoding("Household Income")
Less than RM 4,849        158
RM 4,850 â€“ RM10,959     111
More than RM10,960         45
Name: Household Income, dtype: int64

0    158
2    111
1     45
Name: Household Income, dtype: int64
```

```
LabelEncoding("Level of Study")
Undergraduate           900
Certificate/Diploma     134
Master                    9
Postgraduate              5
PhD                       3
Name: Level of Study, dtype: int64

4    900
0    134
1      9
3      5
2      3
Name: Level of Study, dtype: int64
```

3. DATA PREPARATION

Extracting Relevant Study Fields


```
mask = df["Field of study"].str.contains('|'.join(relevant_fields), case=False)
df = df[mask]
```

relevant_fields is a list with terms such as “computer”, “engineering”, “tech”, “jurutera”, “komputer”

```
LabelEncoding("Field of study")
```

```
Computer Science/Information Technology    214
Engineering                                 92
Medical Laboratory Technology                2
sijil sistem komputer                        2
Computer and Information Technology          1
Science and Technology Studies               1
medical lab technology                       1
Chemical engineering                         1
Name: Field of study, dtype: int64

1    214
3     92
4      2
7      2
2      1
5      1
6      1
0      1
Name: Field of study, dtype: int64
```

```
df["Field of study"].replace([7, 2, 5, 6, 0], [1, 1, 4, 4, 3], inplace=True)
df["Field of study"].replace([1, 3, 4], [0, 1, 2], inplace=True)
df["Field of study"].value_counts()
```

```
0    217
1     93
2      4
Name: Field of study, dtype: int64
```

3. DATA PREPARATION

Cleaning Manual Inputs

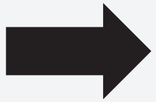
```
nltk.download("wordnet")
nltk.download("stopwords")
stop_words = set(stopwords.words("english"))
```

```
def remove_stopwords(text):
    return ' '.join([word for word in text.split() if word not in stop_words])
```

```
lemmatizer = WordNetLemmatizer()
```

```
def lemmatize_text(text):
    return ' '.join([lemmatizer.lemmatize(word) for word in text.split()])
```

Youtube Video and Live Lecture	5
live lecture and youtube video	4
Games	4
Youtube video and live lecture	4
YouTube video	4
YOUTUBE VIDEO	4
Face to face	4
Procedural Demonstration	4
YouTube	3
Youtube Video & Live Lecture	3
Debugging	3
youtube video	3
Youtube and Live	3
Guided learning	3
face to face	3
Live Lecture or Youtube	3



```
live lecture
71
youtube video
71
youtube
21
youtube video live lecture
5
live
3
live lecture youtube video
3
debugging
3
problem based learning
3
pre-recorded lecture
3
```

NLTK

- Keywords from “Online Instructional Strategies/ Assessment” are stored in a list “tech_pref”
- A lambda function is “applied” onto df (similar to one-hot encoding)
- If any word in “tech_pref” exists in the initial output, a new column is introduced with value 1, else 0.

4. MODELLING

Identifying Target Variables

- **Begin with "6. Online Instructional Strategies/Assessment"**
  - **_Written assignment_**
  - **_Case Study_**
  - **_Real Time Online Exam_**
  - **_Individual Project/Assignment_**
  - **_Group Project/Assignment_**
  - **_Online Quiz/Test - MCQ_**
  - **_Online Quiz/Test - Essay_**
  - **_Online Quiz/Test - Open Book_**
  - **_Peer Review Assessment Live Presentation_**
  - **_Recorded Presentation_**
  - **_Portfolio_**

```
targets = [0, 1, 2, 3, 4]
replacements = [0, 0, 0, 1, 1]
```

- Initial preprocessing was defined as below:
  - Not at All – 0
  - Not Really – 1
  - Undecided – 2
  - Somewhat – 3
  - Very Much – 4
- For these targets to have binary outputs, "Somewhat" and "Very Much" are converted to 1, others to 0
- e.g.

```
df[["6. Online Instructional Strategies/Assessment [Case Study]"]].replace(targets, replacements)
```

4. MODELLING

Removing Unused Columns

- Dropped columns are:
  - Institutions
  - Country
  - Faculty
  - Department
  - Please share any comments or suggestions…
  - For technical or hands-on subjects… (Online Technical Assessment Preference)

```
df.drop(df.columns[59], inplace=True, axis=1)
df.drop(cols_to_drop, inplace=True, axis=1)
```

4. MODELLING

Train Test Split


```
X2 = df.loc[:, ~df.columns.isin(targets)]
y2 = df[targets]
print(f"X2.shape: {X2.shape}")
print(f"y2.shape: {y2.shape}")
```

```
X2.shape: (314, 115)
y2.shape: (314, 11)
```

```
X2_train, X2_test, y2_train, y2_test = train_test_split(X2, y2, test_size=0.2, random_state=2024)
```

4. MODELLING

Models

GradientBoostingClassifier

```
def GBC(X_train, y_train, X_test, learning_rate):
  gbc = GradientBoostingClassifier(learning_rate=learning_rate, random_state=2024)
  model = MultiOutputClassifier(gbc).fit(X_train, y_train)
  prediction = model.predict(X_test)
  return model, prediction
```

XGBClassifier

```
def XGB(X_train, y_train, X_test):
  classifier = MultiOutputClassifier(XGBClassifier(random_state=2024))
  clf = Pipeline([("classify", classifier)])
  clf.fit(X_train, y_train)
  prediction = clf.predict(X_test)
  return clf, prediction
```

RandomForestClassifier

```
def RFC(X_train, y_train, X_test):
  rfc = RandomForestClassifier(random_state=2024)
  model = MultiOutputClassifier(rfc).fit(X_train, y_train)
  prediction = model.predict(X_test)
  return model, prediction
```

4. MODELLING

Initialize Models

```
gbc2, gbc_pred2 = GBC(X2_train, y2_train, X2_test, 0.1)
gnb2, gnb_pred2 = GNB(X2_train, y2_train, X2_test)
svm2, svm_pred2 = SVM(X2_train, y2_train, X2_test)
rfc2, rfc_pred2 = RFC(X2_train, y2_train, X2_test)
lr2, lr_pred2 = LR(X2_train, y2_train, X2_test)
xgb2, xgb_pred2 = XGB(X2_train, y2_train, X2_test)
```

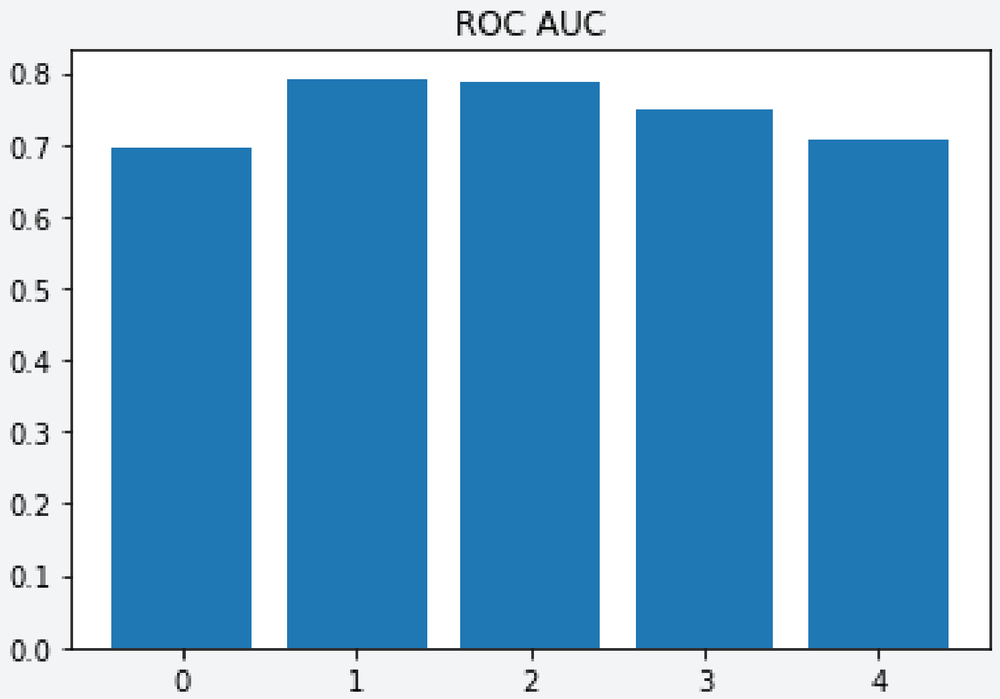
5. EVALUATION

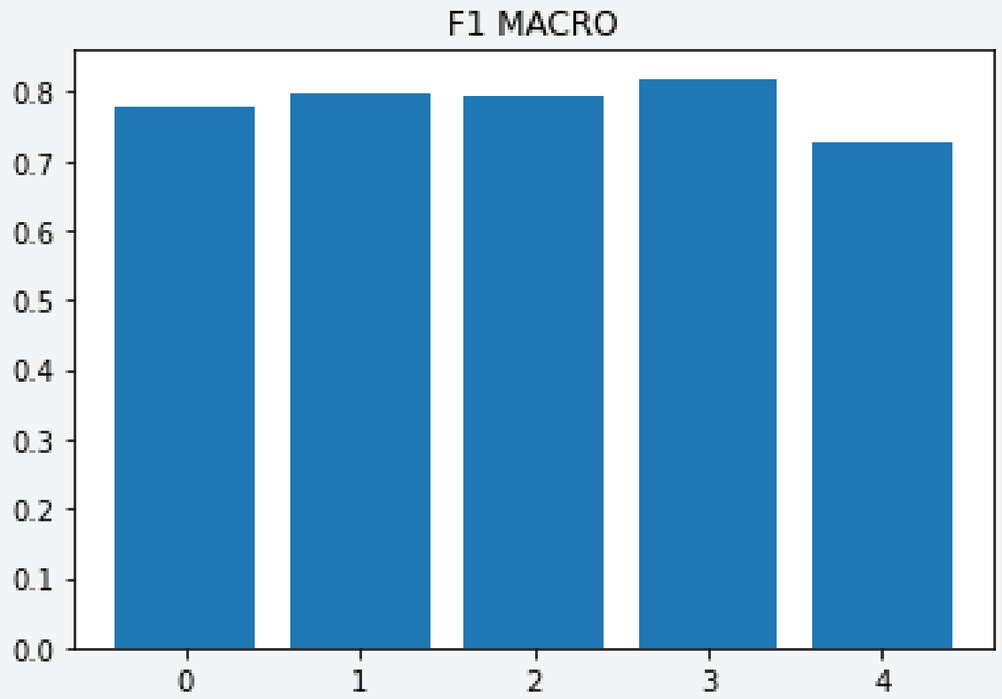
	Hamming Loss	ROC AUC	F1 Score (Macro)
GBC2	0.228	0.693	0.83
RFC2	0.227	0.687	0.83
XGB2	0.258	0.655	0.81
LR2	0.276	0.669	0.79
SVM2	0.242	0.636	0.82
GNB2	0.553	0.576	0.34

5. EVALUATION

Cross Validation with GBC2

```
metrics = ["roc_auc", "f1_macro", "precision_macro", "recall_macro", "accuracy"]
gbc2_scores = cross_validate(gbc2, X, y, cv=5, scoring=metrics)
```

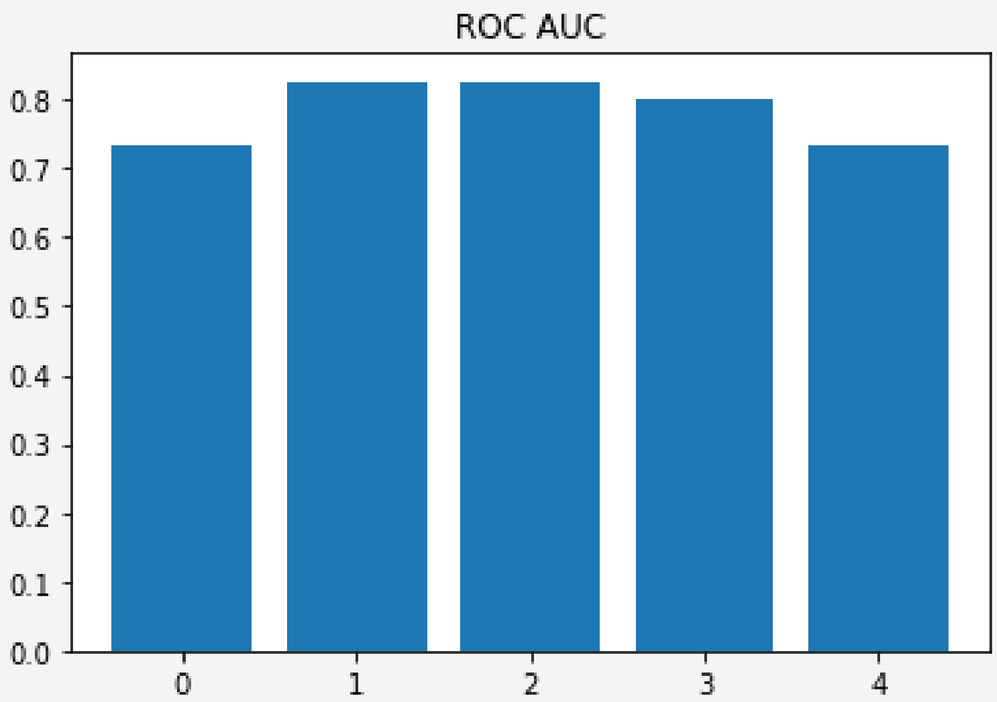


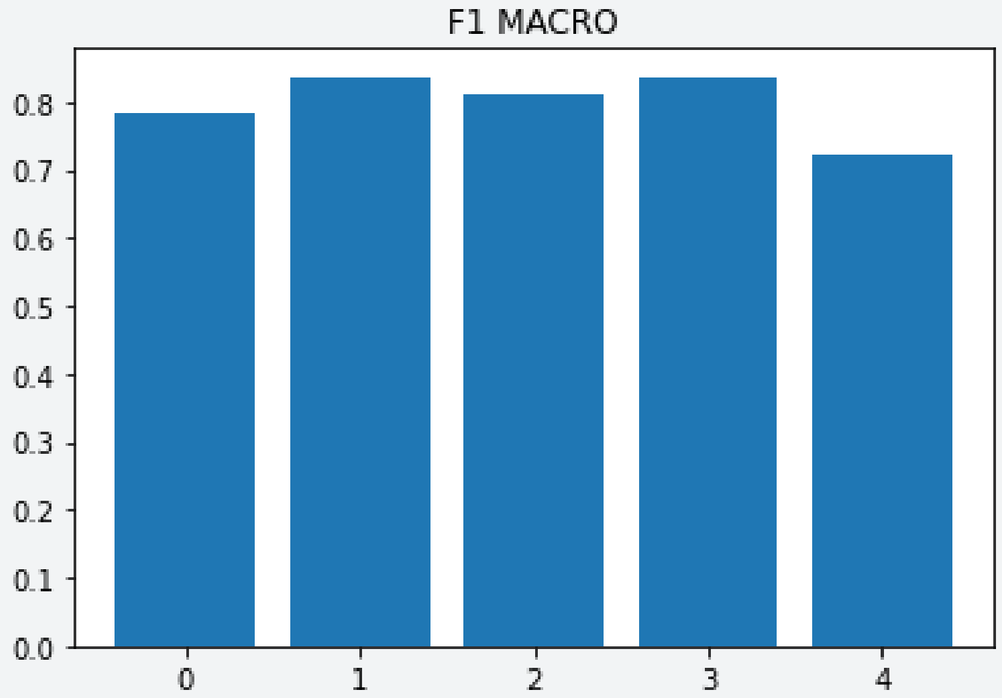


5. EVALUATION

Cross Validation with RFC2

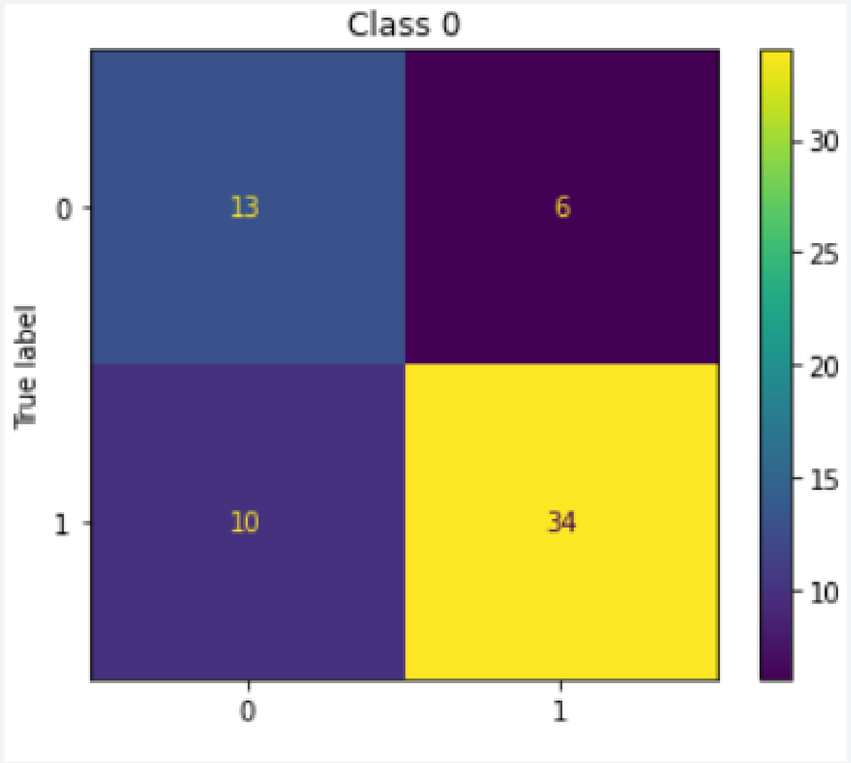
```
metrics = ["roc_auc", "f1_macro", "precision_macro", "recall_macro", "accuracy"]
rfc2_scores = cross_validate(rfc2, X, y, cv=5, scoring=metrics)
```



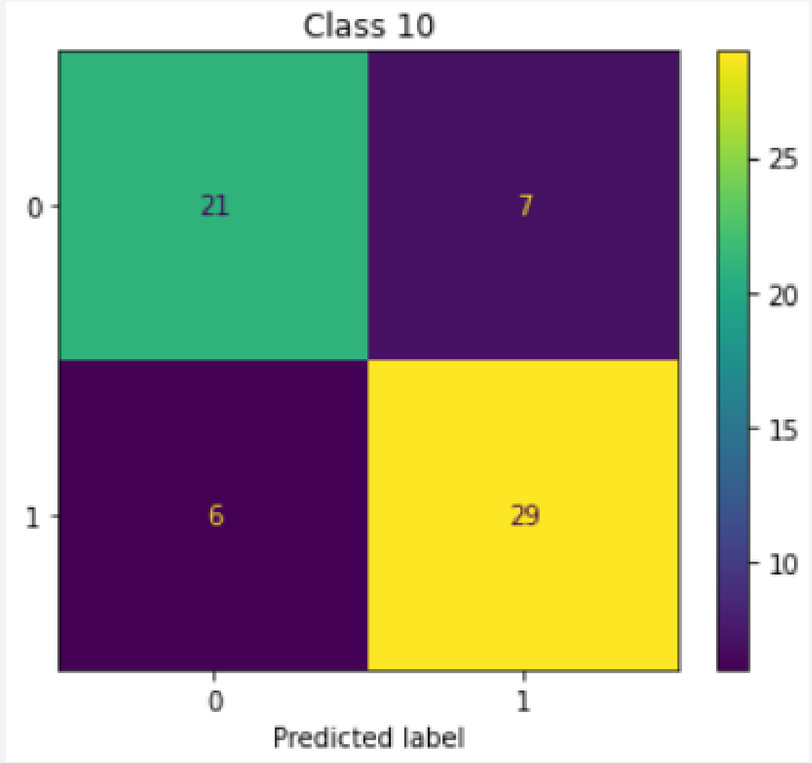


5. EVALUATION

Confusion Matrices of GBC2



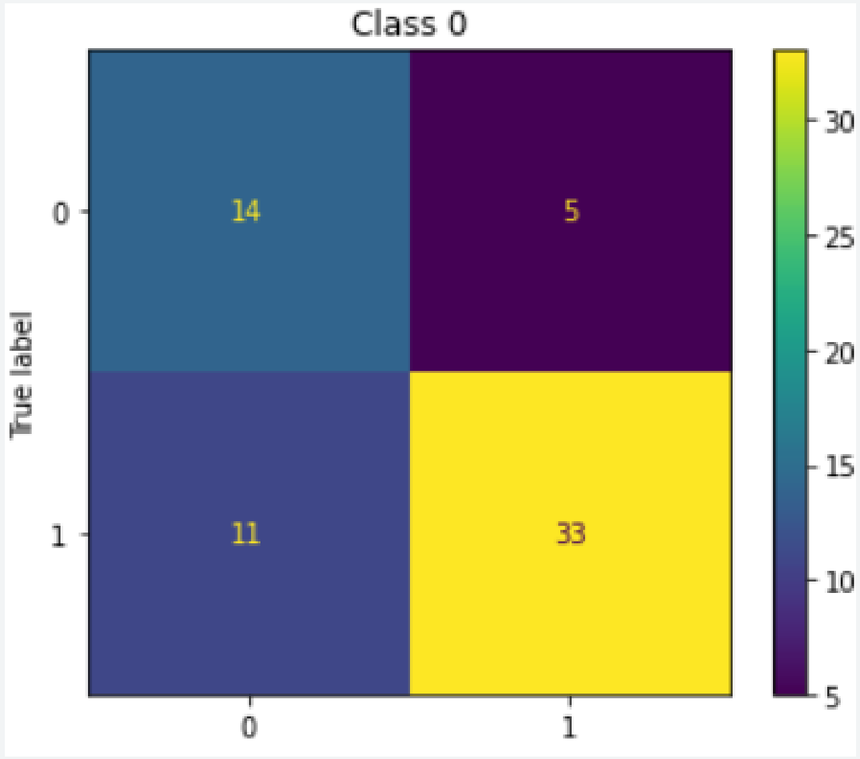
Accuracy: 47/63 = 0.75
Precision: 34/40 = 0.85
Recall: 34/44 = 0.77



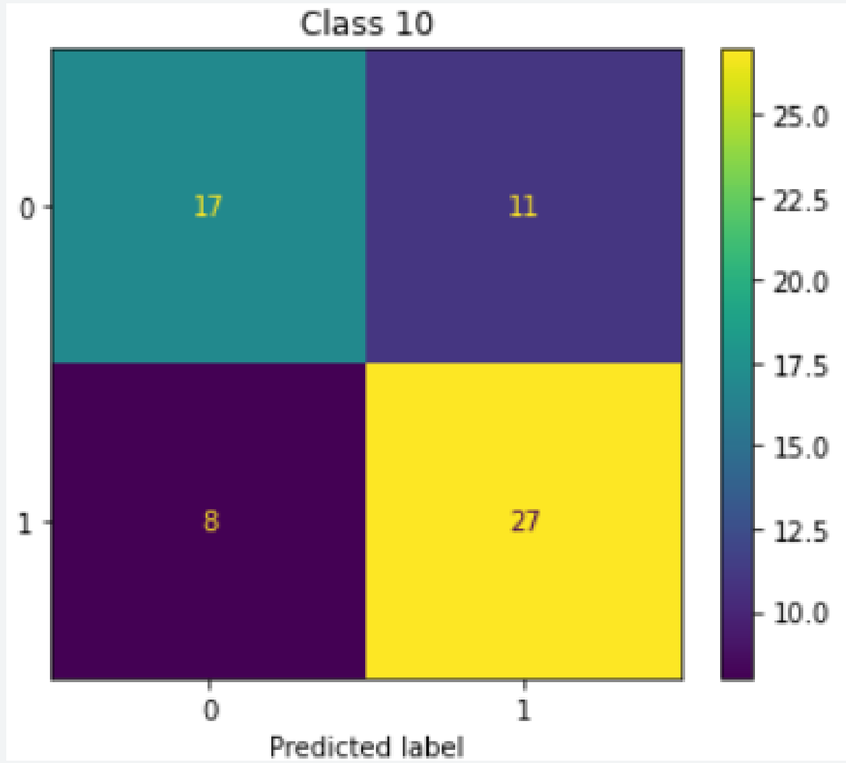
Accuracy: 50/63 = 0.79
Precision: 29/36 = 0.81
Recall: 29/35 = 0.83

5. EVALUATION

Confusion Matrices of RFC2



Accuracy: 47/63 = 0.75
Precision: 33/38 = 0.87
Recall: 33/44 = 0.75



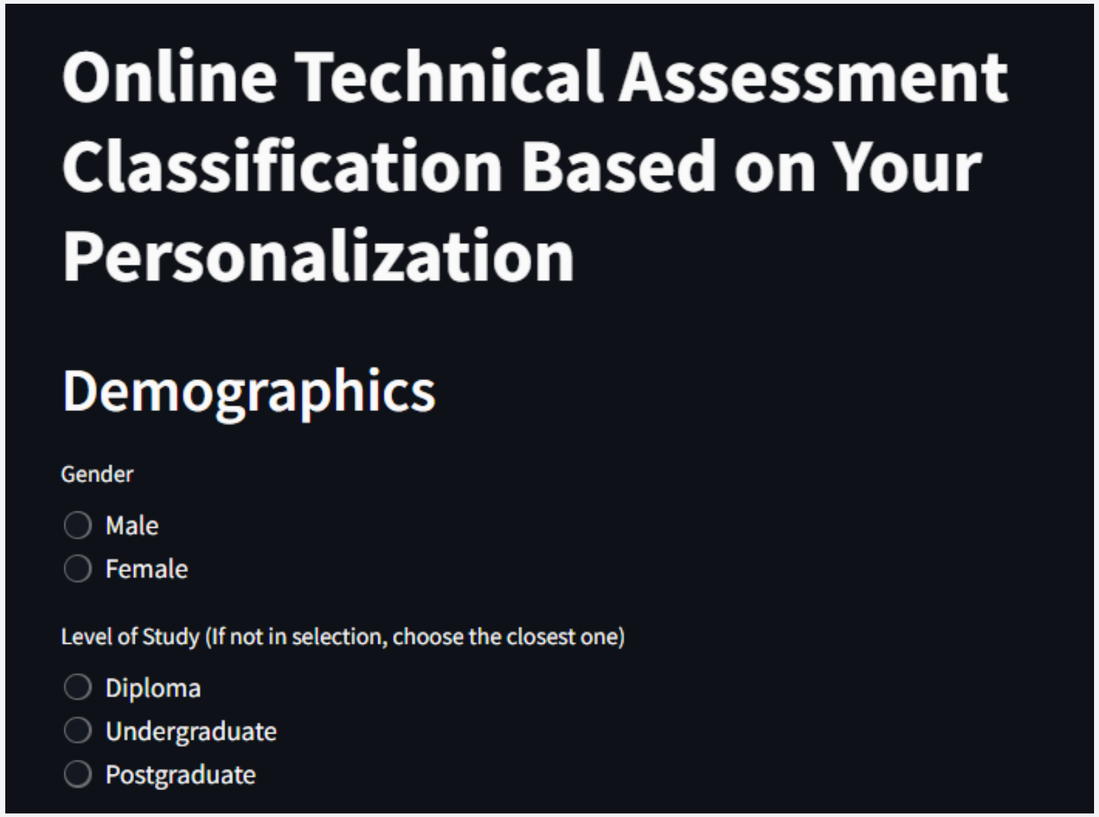
Accuracy: 44/63 = 0.70
Precision: 27/38 = 0.71
Recall: 27/35 = 0.77

6. DEPLOYMENT

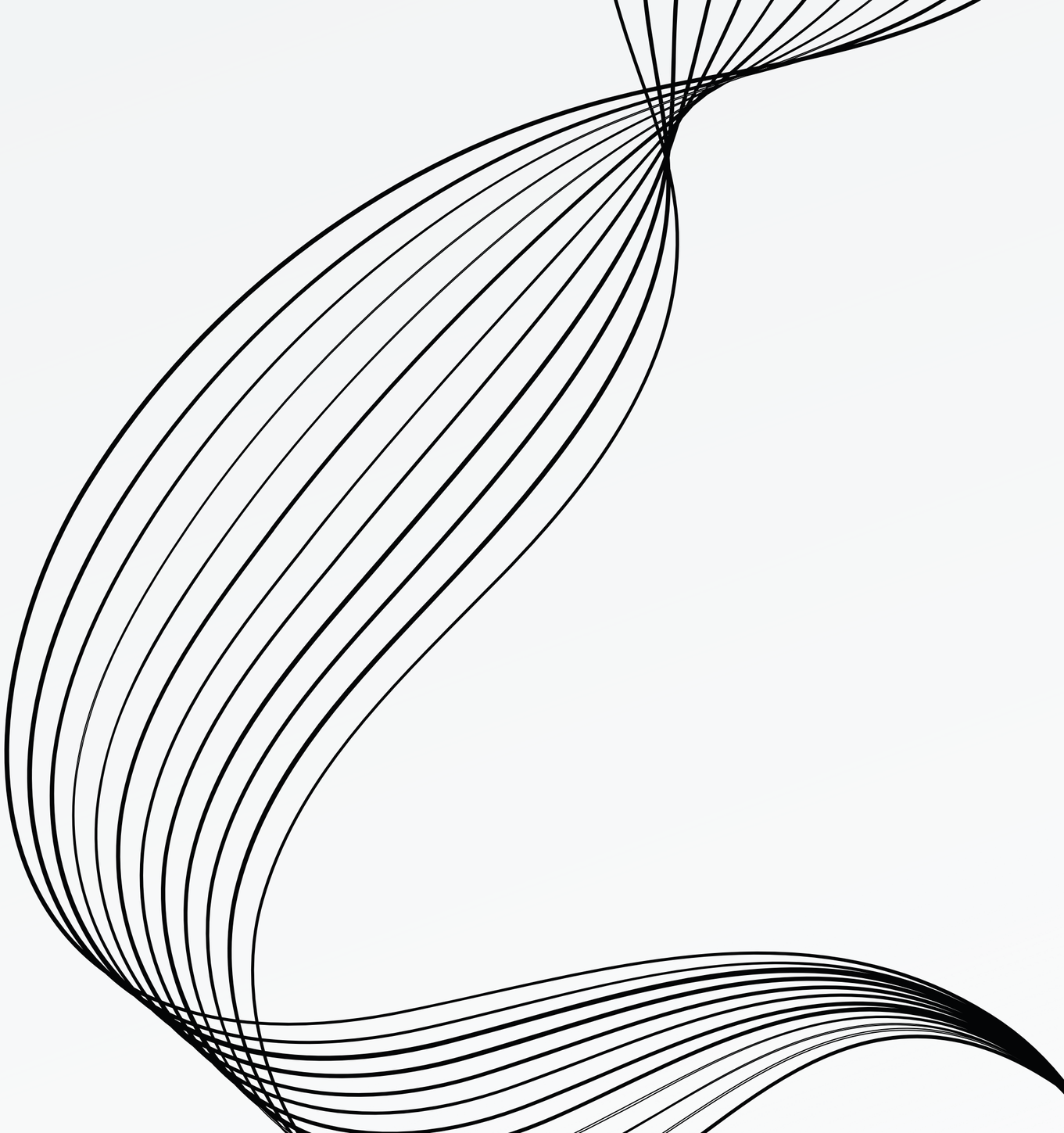
Streamlit Application

- Screenshot

Streamlit



THANK YOU



REFERENCES

- Abduh, M. Y. M. (2021). Full-time online assessment during COVID-19 lockdown: EFL teachers' perceptions. Asian EFL Journal, 28(1.1), 26-46.
- Bailey, K. M. (1998). Learning about language assessment: dilemmas, decisionjs, and directions. Heinle& Heinle: US.
- Barbe, W. B., Swassing, R. H., & Milone, M. N. (1979). Teaching through modality strengths : concepts and practices. Zaner-Bloser. https://cir.nii.ac.jp/crid/1130282272992258816
- Çetinkaya, A., Baykan, Ö. K., & Kırgız, H. (2023). Analysis of Machine Learning Classification Approaches for Predicting Students' Programming Aptitude. Sustainability, 15(17), 12917.
- Chrysafiadi, K., Troussas, C., & Virvou, M. (2018, 7-10 Oct. 2018). A Framework for Creating Automated Online Adaptive Tests Using Multiple-Criteria Decision Analysis. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
- Dema, K. (2021). Understanding Students' C Programming Language Learning Styles: A Case Study in College of Science and Technology. Journal of Information Engineering and Applications, 11(1), 7-14.
- Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. Turkish Online Journal of Educational Technology-TOJET, 2(3), 13-19.
- García, P., Amandi, A., Schiaffino, S., & Campo, M. (2005). Using Bayesian networks to detect students' learning styles in a web-based education system. Proc of ASAI, Rosario, 115, 126.
- Kim, N., Smith, M. J., & Maeng, K. (2008). Assessment in online distance education: A comparison of three online programs at a university. Online Journal of Distance Learning Administration, 11(1), 1-16.
- Maennel, O. M. (2019). Predicting student's success using technical labs as part of university admission to a cyber security program.
- Ocepek, U., Bosnić, Z., Nančovska Šerbec, I., & Rugelj, J. (2013). Exploring the relation between learning style models and preferred multimedia types. Computers & Education, 69, 343-355. https://doi.org/https://doi.org/10.1016/j.compedu.2013.07.029
- Papadakis, S. (2023). MOOCs 2012-2022: An overview. Advances in Mobile Learning Educational Research, 3, 682-693. https://doi.org/10.25082/AMLER.2023.01.017
- Parsons, D., Wood, K., & Haden, P. (2015, January). What are we doing when we assess programming?. In Proceedings of the 17th Australasian Computing Education Conference (ACE 2015) (Vol. 27, p. 30).
- Pritchard, A. (2009). Ways of learning: Learning theories and learning styles in the classroom (2nd ed.). New York, NY:Routledge.
- Roberts, G. H., & Verbyla, J. L. (2003, January). An online programming assessment tool. In Proceedings of the fifth Australasian conference on Computing education-Volume 20 (pp. 69-75).
- Robinson, P. E., & Carroll, J. (2017, 25-28 April 2017). An online learning platform for teaching, learning, and assessment of programming. 2017 IEEE Global Engineering Education Conference (EDUCON) (pp. 547-556). IEEE. https://10.1109/EDUCON.2017.7942900
- Seyal, A. H., Mey, Y. S., Matusin, M. H., Siau, N. H., & Rahman, A. A. (2015). Understanding students learning style and their performance in computer programming course: Evidence from Bruneian technical institution of higher learning. International Journal of Computer Theory and Engineering, 7(3), 241.
- Sievertsen, H. H., & Burgess, S. (2020, April 1). Schools, skills, and learning: The impact of covid-19 on Education. CEPR. https://cepr.org/voxeu/columns/schools-skills-and-learning-impact-covid-19-education
- Ulloa-Cazarez, R. L., López-Martín, C., Abran, A., & Yáñez-Márquez, C. (2018). Prediction of Online Students Performance by Means of Genetic Programming. Applied Artificial Intelligence, 32(9-10), 858-881. https://doi.org/10.1080/08839514.2018.1508839