Bank Loan Defaulter Prediction

Presented by:

- Ahmad Alharthi
- Faisal Alsufyani
- Yahya Alyoubi

Instructor:
Dr.Mejdal Alqahtani

Introduction

- Financing is a credit method that enables the client to borrow a certain amount of money to achieve his financial goals, and the client may stumble in paying the installments as a result of a sudden change in his financial situation or mismanagement of benefiting from financing and not allocating part of the income to pay the installments.

Goals

The goals of the project:

- This project aims to classify a customer whether a loan defaulter or not based on multiple factors like loan amount and Interest Rate.
- Finding the factors that affect clients' failure to pay the loan before obtaining it with high accuracy.

Penalties

Penalty for defaulting on loans:

- Difficulty obtaining another financing shortly.
- Legal accountability and prosecution by creditors.
- Freezing bank accounts.

Data Structure

- This data set was collected from Kaggle.com website
- This is a simple overview of the data structure, in which the names of the columns and some rows are explained Columns 35 Rows: 67463 Before cleaning the data :

	ID	Loan Amount	Funded Amount	Funded Amount Investor	Term	Batch Enrolled	Interest Rate	Grade	Sub Grade	Employment Duration	...	Recoveries	Collection Recovery Fee	Collection 12 months Medical	Applicatior Type
0	65087372	10000	32236	12329.36286	59	BAT2522922	11.135007	B	C4	MORTGAGE	...	2.498291	0.793724	0	INDIVIDUAL
1	1450153	3609	11940	12191.99692	59	BAT1586599	12.237563	C	D3	RENT	...	2.377215	0.974821	0	INDIVIDUAL
2	1969101	28276	9311	21603.22455	59	BAT2136391	12.545884	F	D4	MORTGAGE	...	4.316277	1.020075	0	INDIVIDUAL
3	6651430	11170	6954	17877.15585	59	BAT2428731	16.731201	C	C3	MORTGAGE	...	0.107020	0.749971	0	INDIVIDUAL
4	14354669	16890	13226	13539.92667	59	BAT5341619	15.008300	C	D4	MORTGAGE	...	1294.818751	0.368953	0	INDIVIDUAL
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
67458	16164945	13601	6848	13175.28583	59	BAT3193689	9.408858	C	A4	MORTGAGE	...	564.614852	0.865230	0	INDIVIDUAL
67459	35182714	8323	11046	15637.46301	59	BAT1780517	9.972104	C	B3	RENT	...	2.015494	1.403368	0	INDIVIDUAL
67460	16435904	15897	32921	12329.45775	59	BAT1761981	19.650943	A	F3	MORTGAGE	...	5.673092	1.607093	0	INDIVIDUAL
67461	5300325	16567	4975	21353.68465	59	BAT2333412	13.169095	D	E3	OWN	...	1.157454	0.207608	0	INDIVIDUAL
67462	65443173	15353	29875	14207.44860	59	BAT1930365	16.034631	B	D1	MORTGAGE	...	1.856480	0.366386	0	INDIVIDUAL

67463 rows × 35 columns

Exploratory Data Analysis

- Delete columns
- Duplicate values
- Check for NaN values

	Loan Amount	Funded Amount	Funded Amount Investor	Interest Rate
0	10000	32236	12329.36286	11.135007
1	3609	11940	12191.99692	12.237563
2	28276	9311	21603.22455	12.545884
3	11170	6954	17877.15585	16.731201
4	16890	13226	13539.92667	15.008300
...	...	...	...	...
67458	13601	6848	13175.28583	9.408858
67459	8323	11046	15637.46301	9.972104
67460	15897	32921	12329.45775	19.650943
67461	16567	4975	21353.68465	13.169095
67462	15353	29875	14207.44860	16.034631

67463 rows × 27 columns

Exploratory Data Analysis

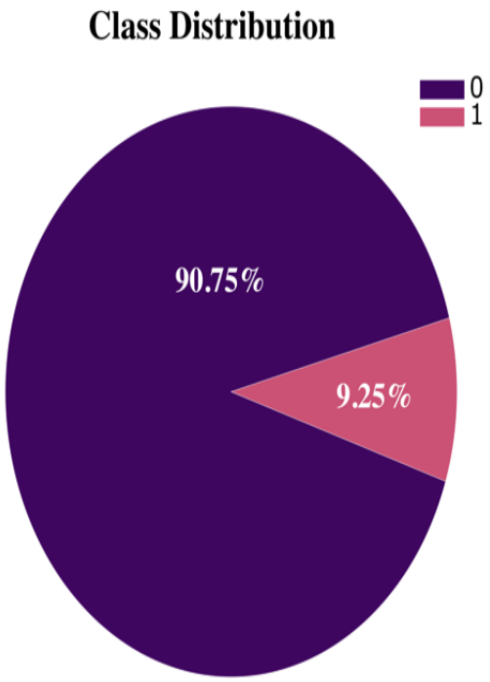


Fig. 1 A plot of imbalanced class distribution

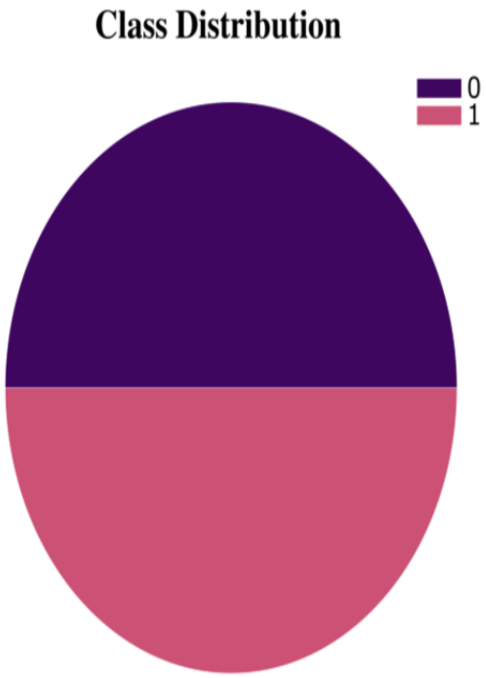
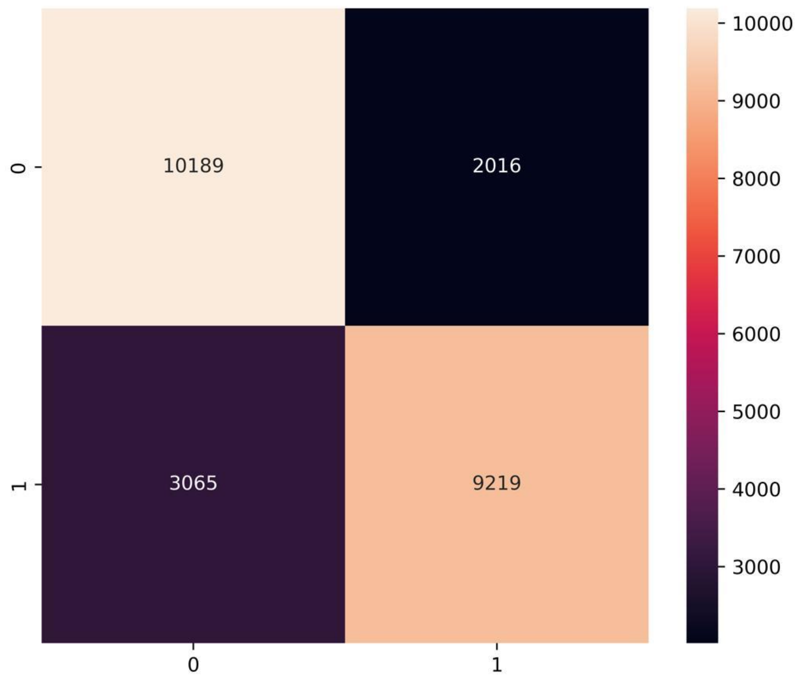


Fig. 2 A plot of a balanced class distribution after applying SMOTE method

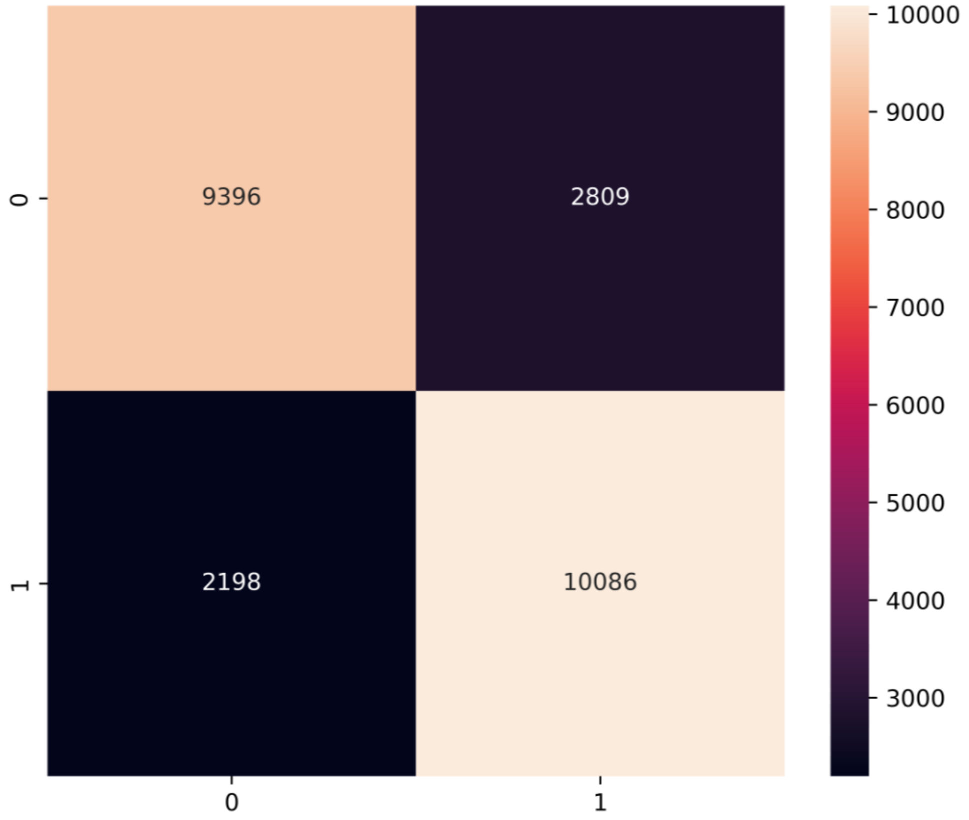
Logistic Regression




```
[13] # Logistic Regression
     from sklearn.metrics import classification_report
     labels = ['Not loan defaulter','loan defaulter']
     print(classification_report(y_test, y_pred, target_names=labels))
```

```
                    precision    recall  f1-score   support

Not loan defaulter       0.77      0.83      0.80     12205
    loan defaulter       0.82      0.75      0.78     12284

          accuracy                           0.79     24489
         macro avg       0.79      0.79      0.79     24489
      weighted avg       0.79      0.79      0.79     24489
```

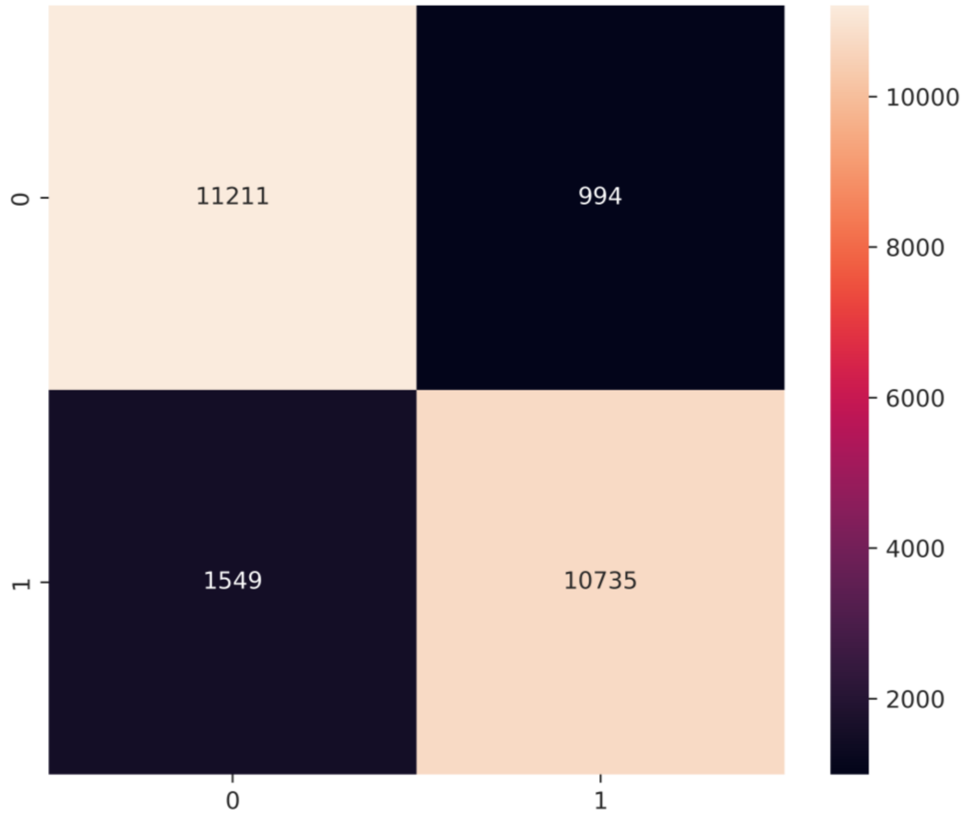
Decision Tree Classifier




```
[14] # DecisionTreeClassifier
     from sklearn.metrics import classification_report
     labels = ['Not a loan defaulter','A loan defaulter']
     print(classification_report(y_test, y_pred, target_names=labels))
```

```
                      precision    recall  f1-score   support

Not a loan defaulter       0.81      0.77      0.79     12205
    A loan defaulter       0.78      0.82      0.80     12284

            accuracy                           0.80     24489
           macro avg       0.80      0.80      0.80     24489
        weighted avg       0.80      0.80      0.80     24489
```

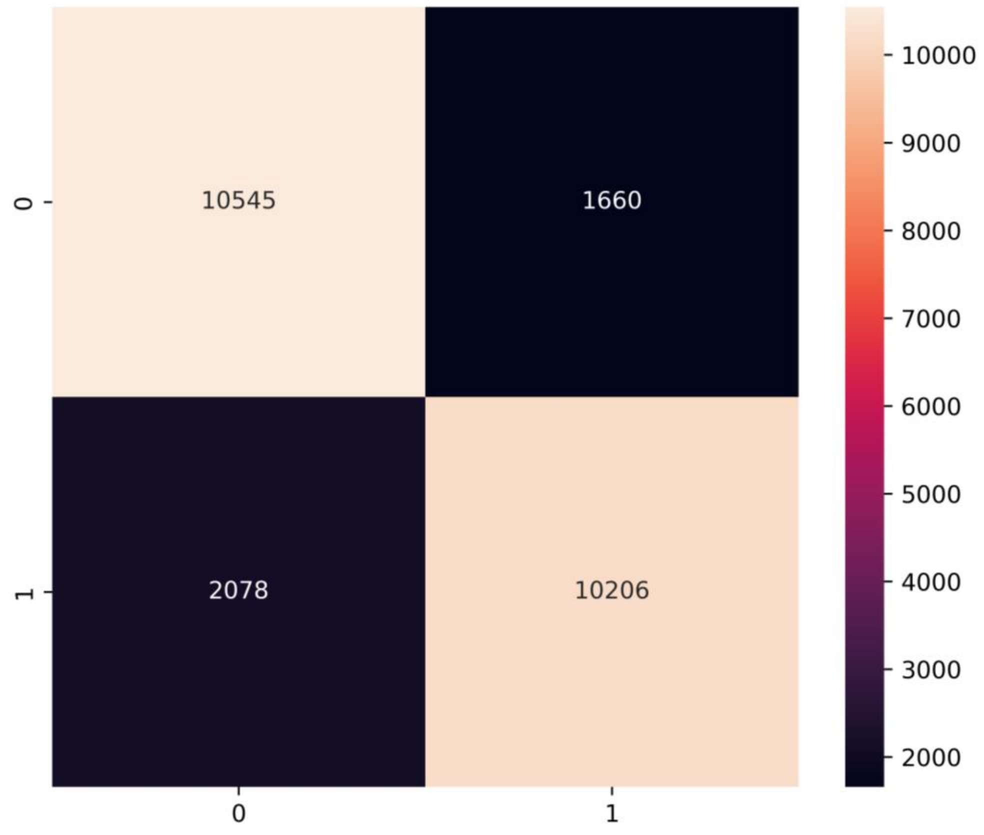
Random Forest Classifier




```
[9] # RandomForestClassifier
    from sklearn.metrics import classification_report
    labels = ['Not loan defaulter','loan defaulter']
    print(classification_report(y_test, y_pred, target_names=labels))
```

```
                    precision    recall  f1-score   support

Not loan defaulter       0.88      0.92      0.90     12205
    loan defaulter       0.92      0.87      0.89     12284

          accuracy                           0.90     24489
         macro avg       0.90      0.90      0.90     24489
      weighted avg       0.90      0.90      0.90     24489
```

XGBoost Classifier



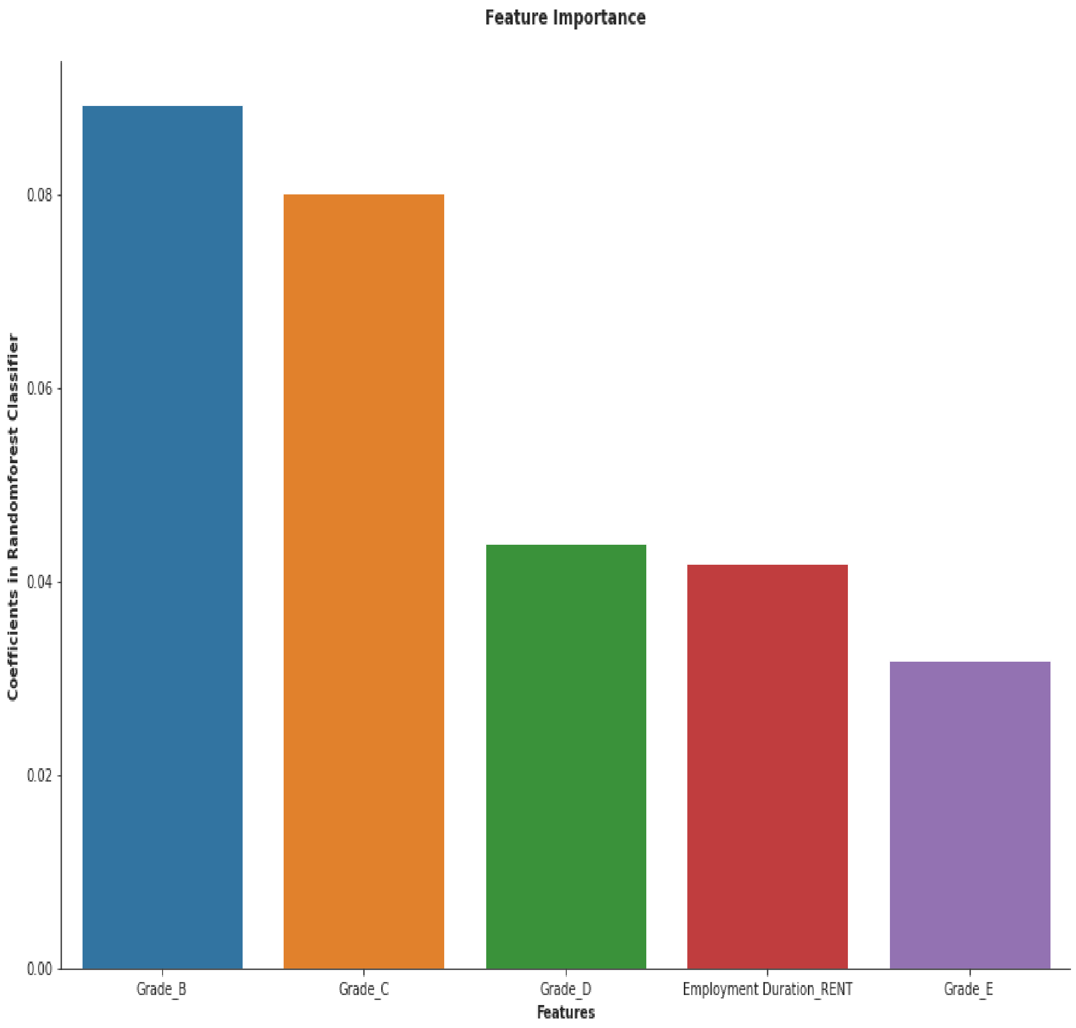

```
[39] # XGBClassifier
     from sklearn.metrics import classification_report
     labels = ['Not a loan defaulter','A loan defaulter']
     print(classification_report(y_test,xgb_classifier.predict(X_test),target_names=labels))
```

```
                      precision    recall  f1-score   support

Not a loan defaulter       0.84      0.86      0.85     12205
    A loan defaulter       0.86      0.83      0.85     12284

            accuracy                           0.85     24489
           macro avg       0.85      0.85      0.85     24489
        weighted avg       0.85      0.85      0.85     24489
```

Conclusion & Future Work

- Model Selection
- Random Forest Classifier

  - with precision which 92% and recall 87%
- Able to predict the person who might be a loan defaulter
- Feature generation needed to improve the accuracy



Thank You

Any Question ?