

LOAN APPROVAL PREDICTION

GROUP 08D

M RANJITH
GOKUL KRISHNAN G
KARTHIK NARAYANAN



LOAN



APPROVED



REJECTED

Problem Statement



In this project, our task is to predict whether a customer is eligible for loan or not based on given customer details.

INTRODUCTION

With the enhancement in the banking sector, there is an increase in the applications for bank loans.

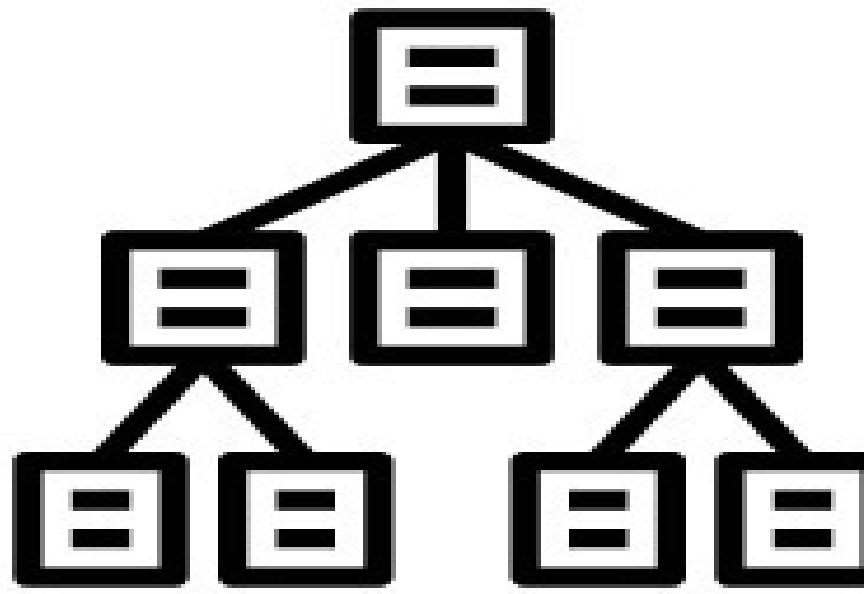
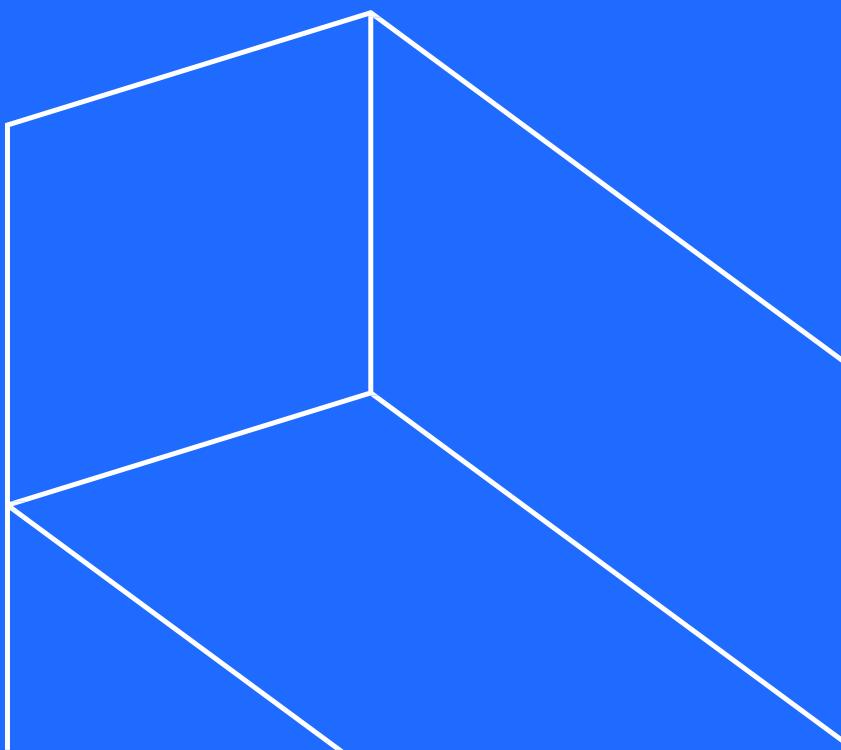
The bank has its limited assets which it has to grant to limited people only.

So finding out to whom the loan can be granted which will be a safer option for the bank is a typical process.

BENEFITS

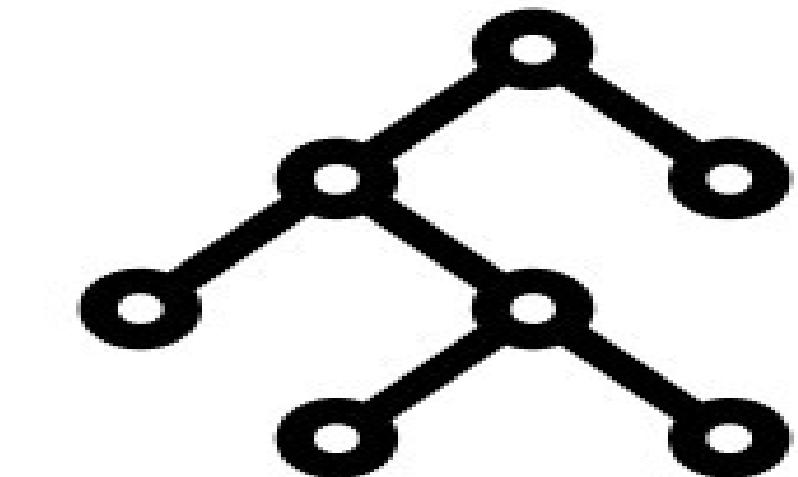
- *Less burdensome if we automate this loan approval process.*
- *Reduces the approval time.*
- *Very helpful for bank employees as well as for the applicant also.*

ALGORITHMS USED



Decision Tree (With and Without Sklearn)

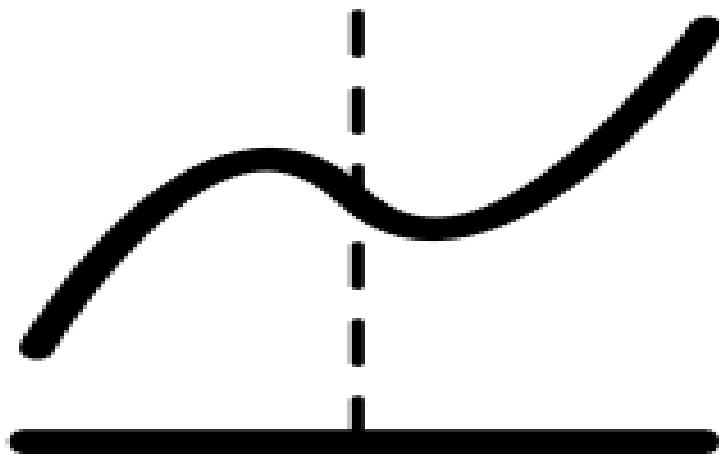
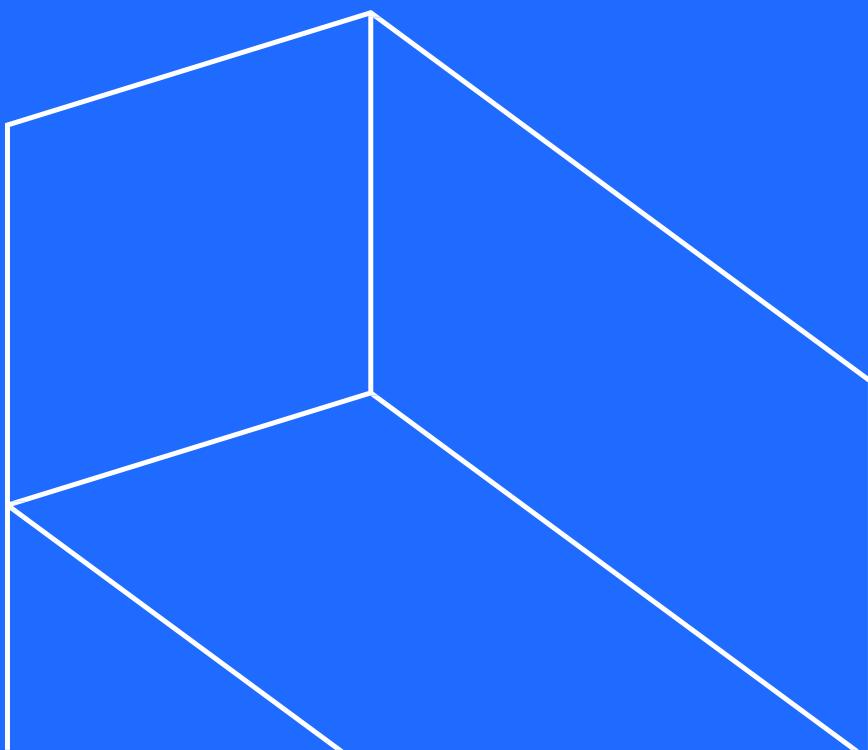
A tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node holds a class label.



Random Forest (Sklearn)

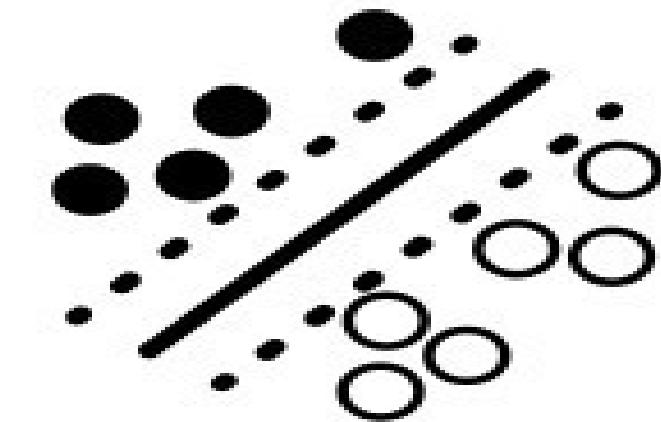
A tree-based machine learning algorithm that leverages the power of multiple decision trees for making decisions.

ALGORITHMS USED



Logistic Regression (Sklearn)

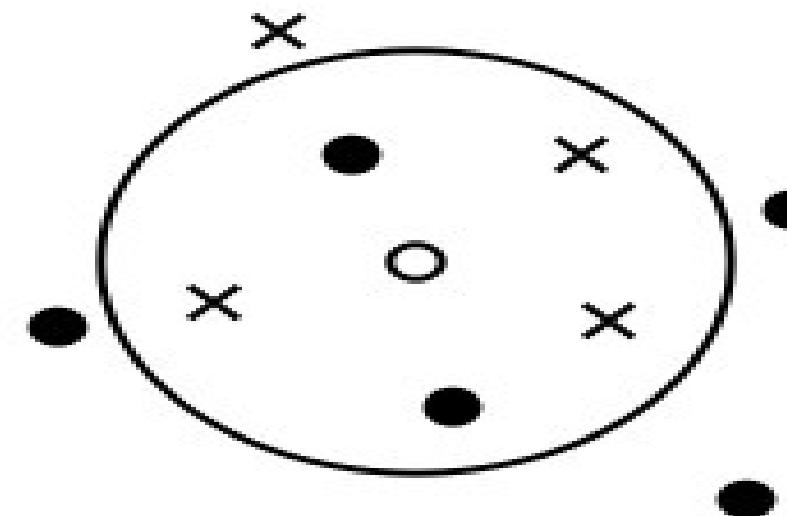
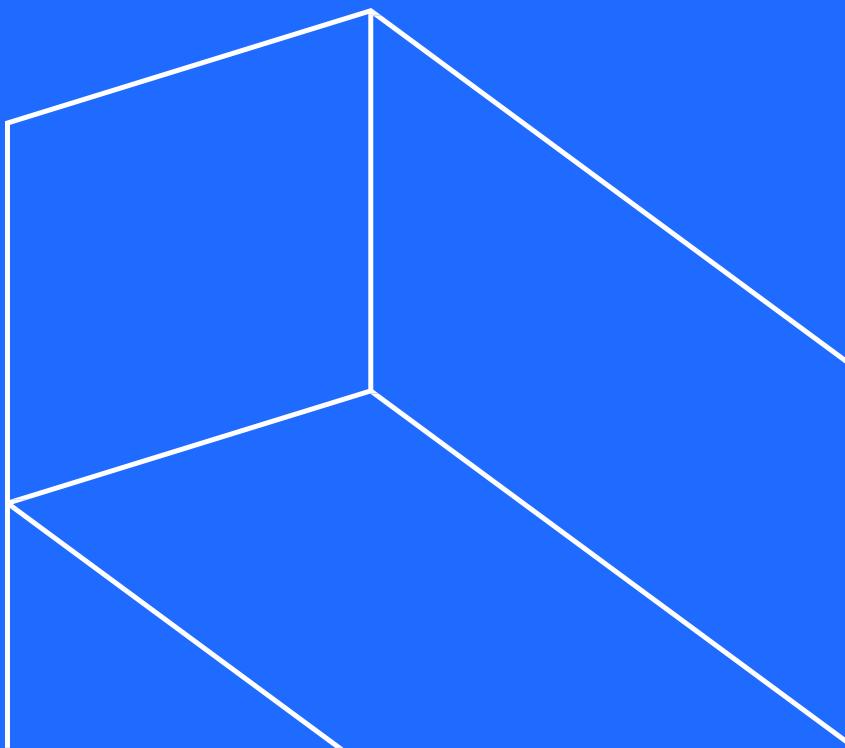
A model predicts a dependent data variable by analyzing the relationship between one or more existing independent variables.



SVM (Sklearn)

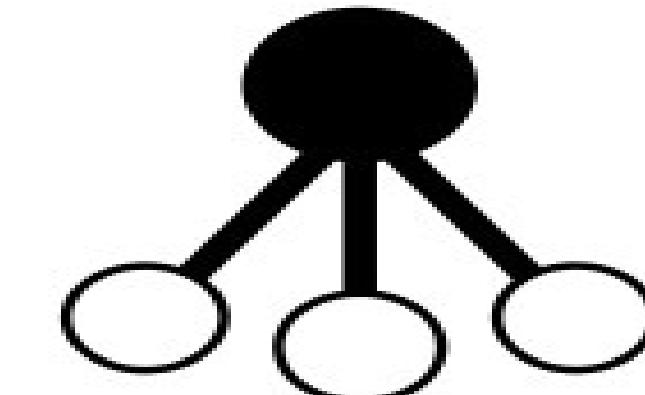
An algorithm that creates a line or a hyperplane which separates the data into classes.

ALGORITHMS USED



KNN (Sklearn)

Algorithm which helps us find which class the new input belongs to when k nearest neighbors are chosen, and distance is calculated between them.



Naïve Bayes (Sklearn)

It is a supervised classification technique used to classify future objects by assigning class labels to instances/records using conditional probability

DATASET-1

M Ranjith

AM.EN.U4CSE19331

This dataset consists of 11 columns and 500 rows.

Features

- Education
- Principal
- Age
- Term
- Gender

Target - Loan Status

- PAIDOFF
- COLLECTION
- COLLECTION_PAIDOFF

Links:

Dataset

Files

```
RangeIndex: 500 entries, 0 to 499
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Loan_ID          500 non-null    object 
 1   loan_status       500 non-null    object 
 2   Principal         500 non-null    int64  
 3   terms             500 non-null    int64  
 4   effective_date   500 non-null    datetime64[ns]
 5   due_date          500 non-null    datetime64[ns]
 6   paid_off_time     392 non-null    object  
 7   past_due_days    208 non-null    float64
 8   age               500 non-null    int64  
 9   education         500 non-null    object  
 10  Gender            500 non-null    object  
dtypes: datetime64[ns](2), float64(1), int64(3), object(5)
memory usage: 43.1+ KB
```

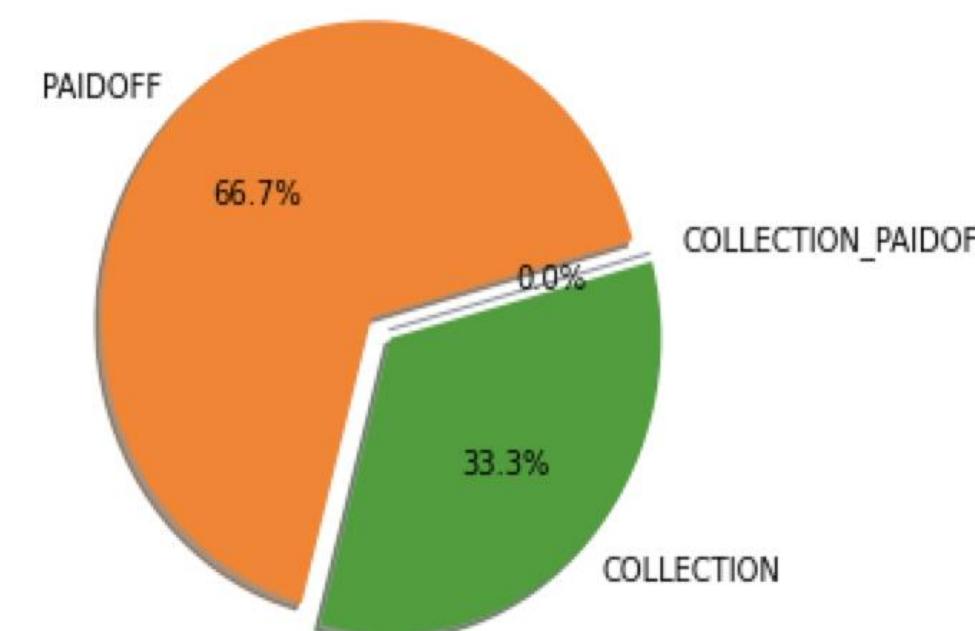
	Loan_ID	loan_status	Principal	terms	effective_date	due_date	paid_off_time	past_due_days	age	education	Gender
0	xqd20160343	PAIDOFF	1000	30	2016-09-12	2016-10-11	10/11/2016 16:00	NaN	31	Bechelor	male
1	xqd20161058	COLLECTION	1000	30	2016-09-11	2016-10-10	NaN	59.0	26	Bechelor	male
2	xqd20161145	COLLECTION	800	30	2016-09-11	2016-10-10	NaN	59.0	38	Bechelor	male
3	xqd20160126	PAIDOFF	1000	30	2016-09-11	2016-10-10	10/10/2016 13:01	NaN	26	college	male
4	xqd20160523	PAIDOFF	1000	7	2016-09-11	2016-09-17	9/15/2016 20:36	NaN	29	college	male

```
PAIDOFF                           292
COLLECTION                        108
COLLECTION_PAIDOFF                 100
Name: loan_status, dtype: int64
```

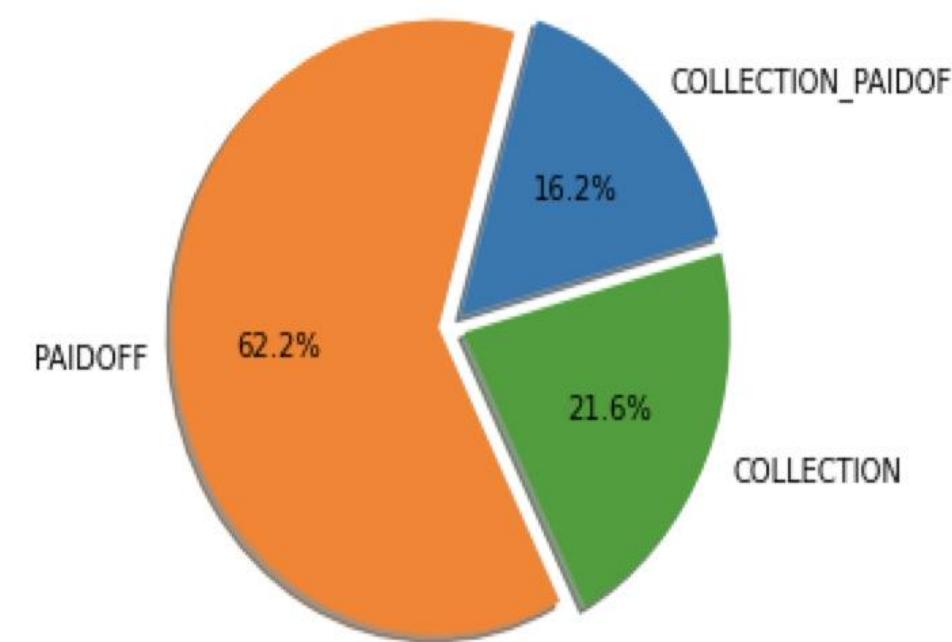
Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Of
Different Education Groups

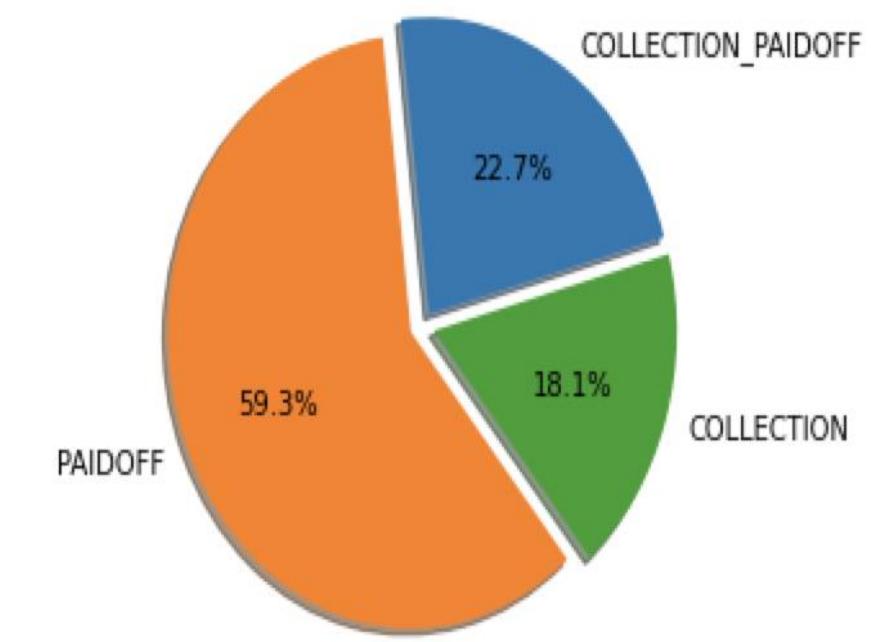
Master or Above



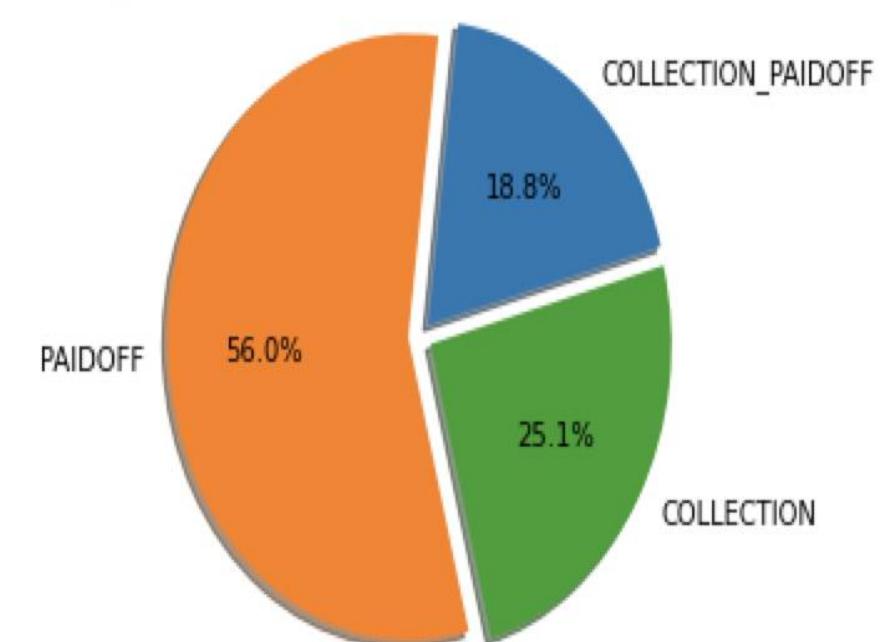
Bachelor



college

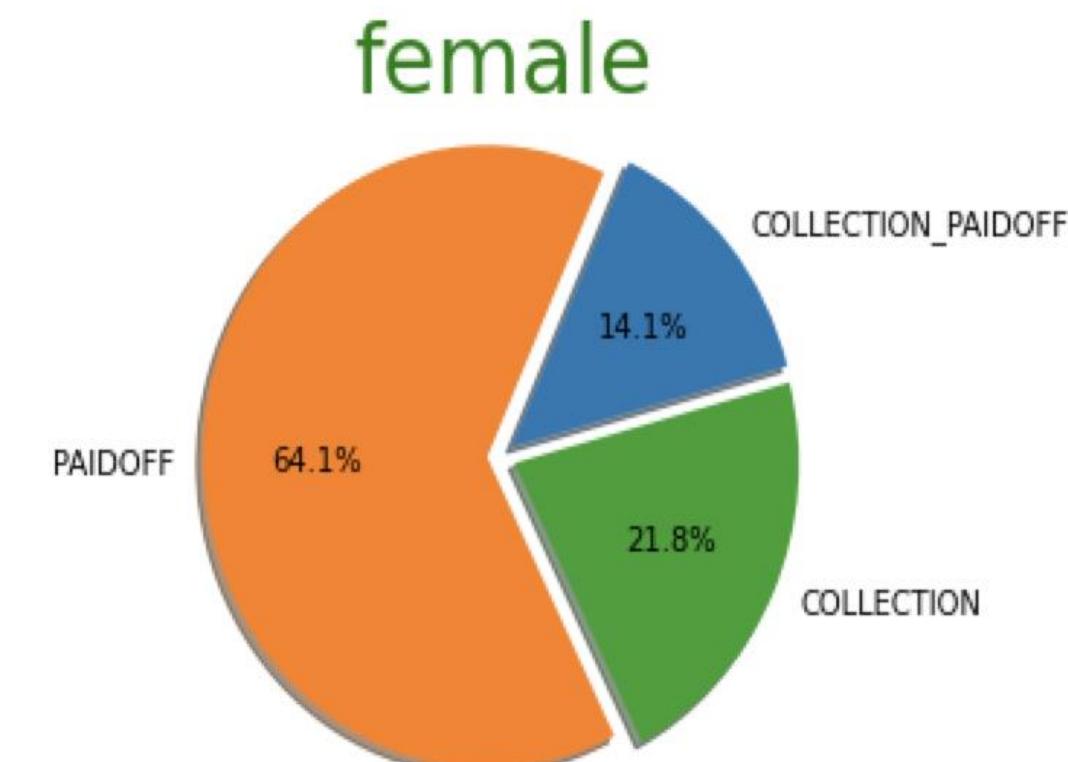
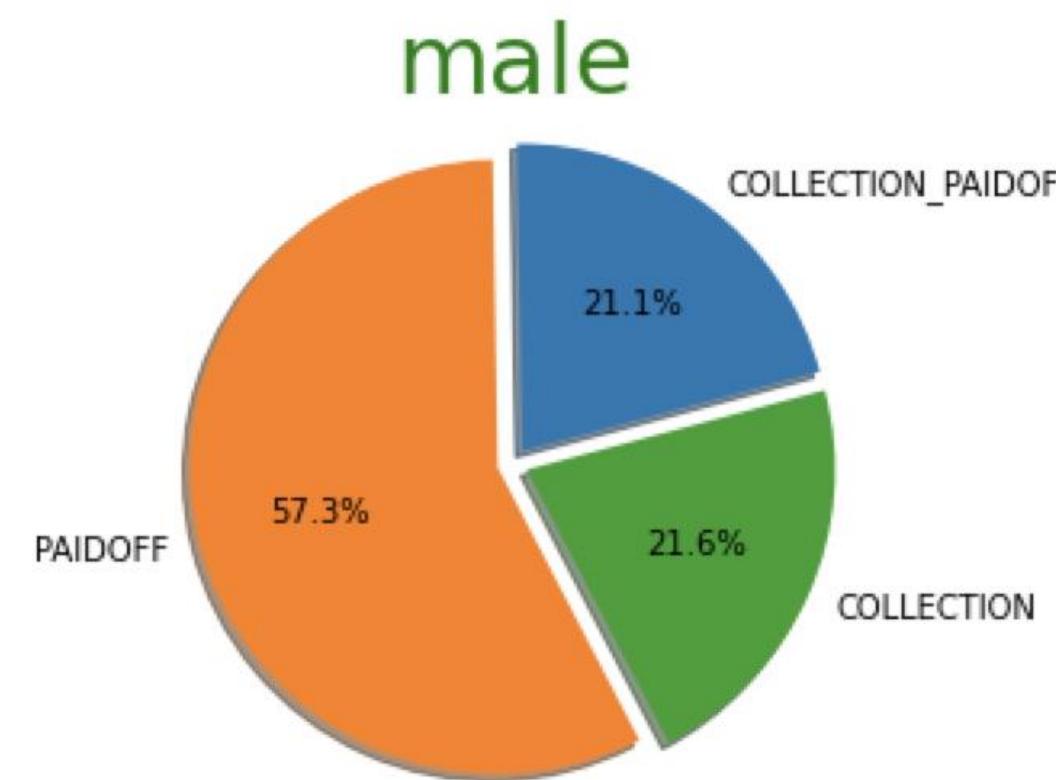


High School or Below



Data Visualization And Interpretation

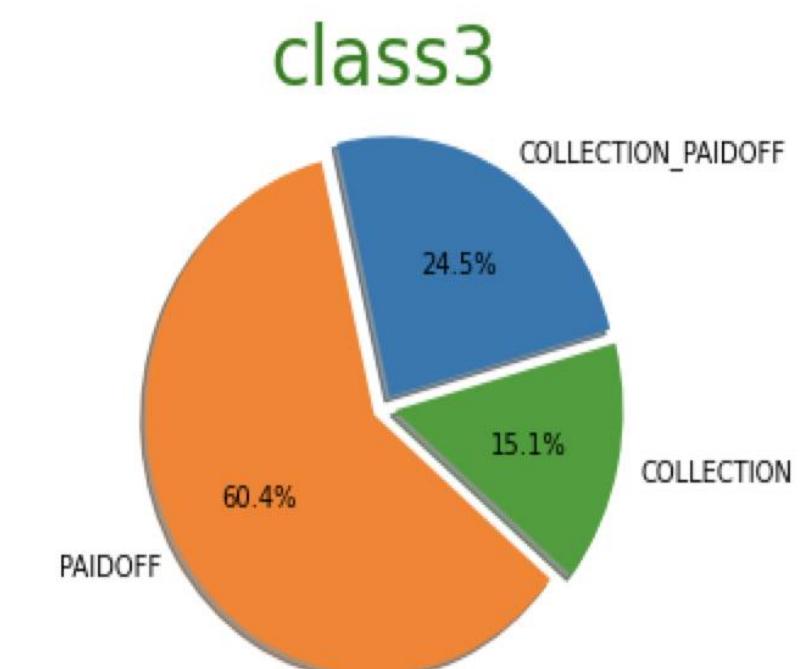
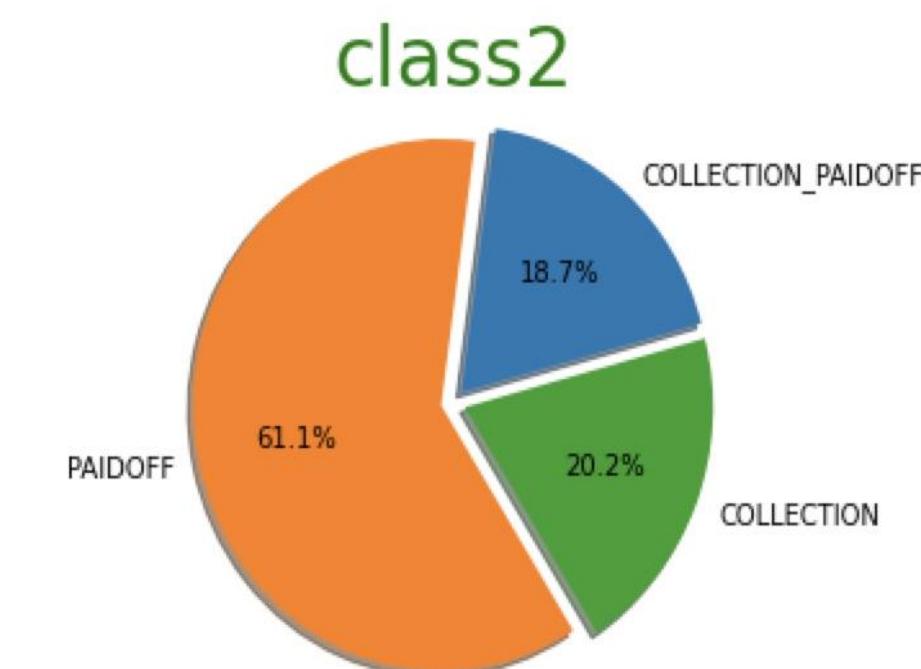
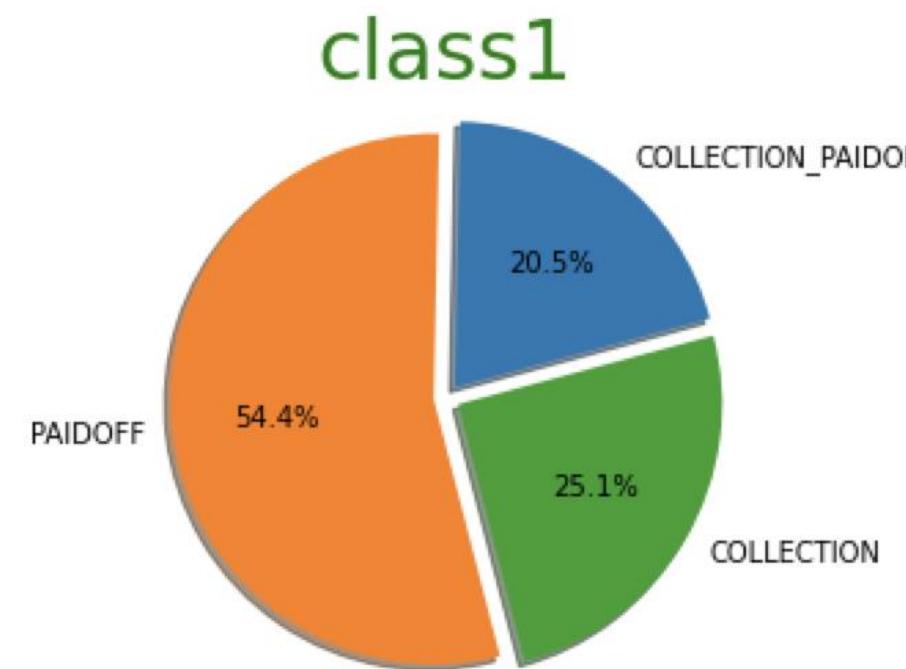
Pie Chart To Represent The Loan Status Based On Gender



Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Based On Age

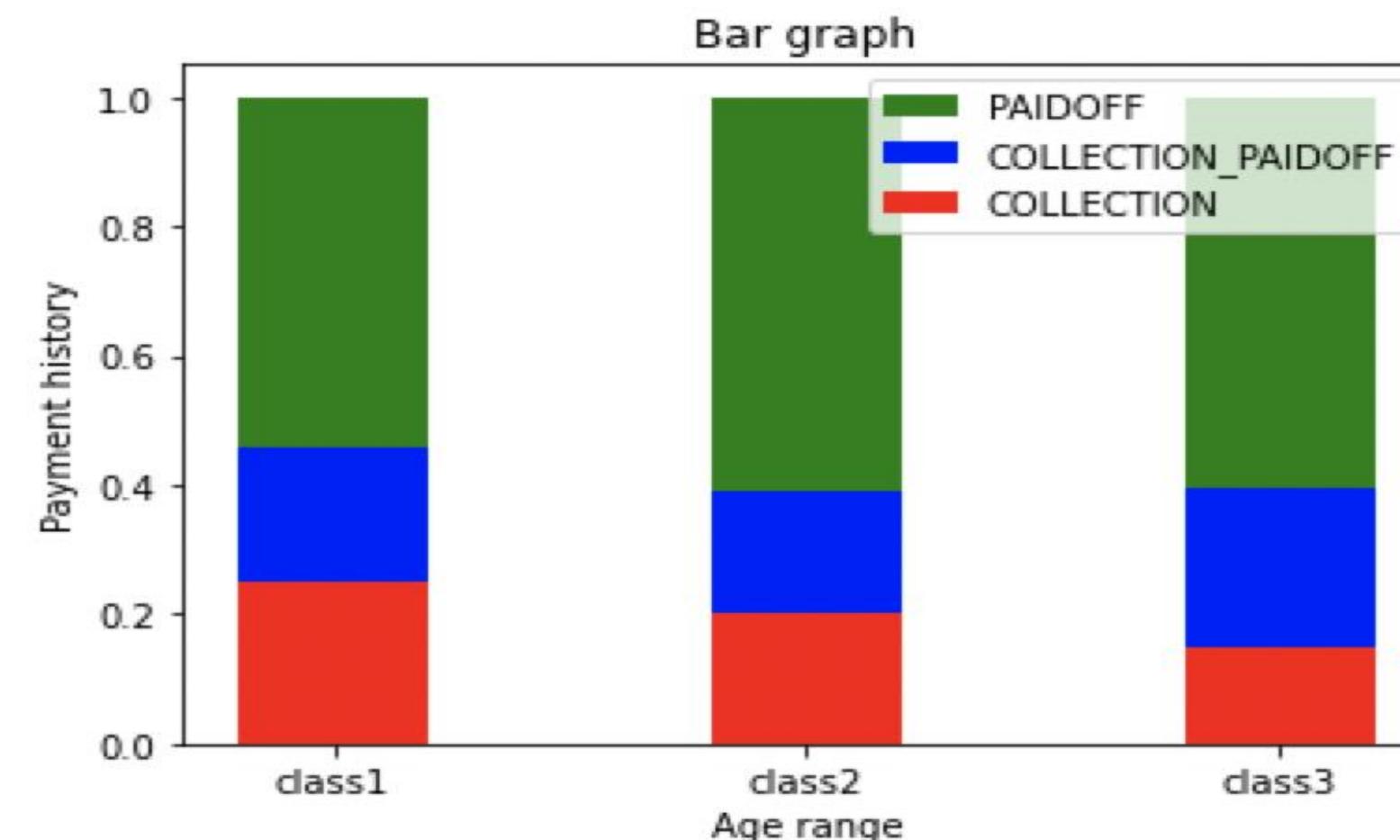
Class 1: [18, 29]
Class 2: [29, 40]
Class 3: [40, 52]



Data Visualization And Interpretation

Bar Chart Between Loan Status And Age

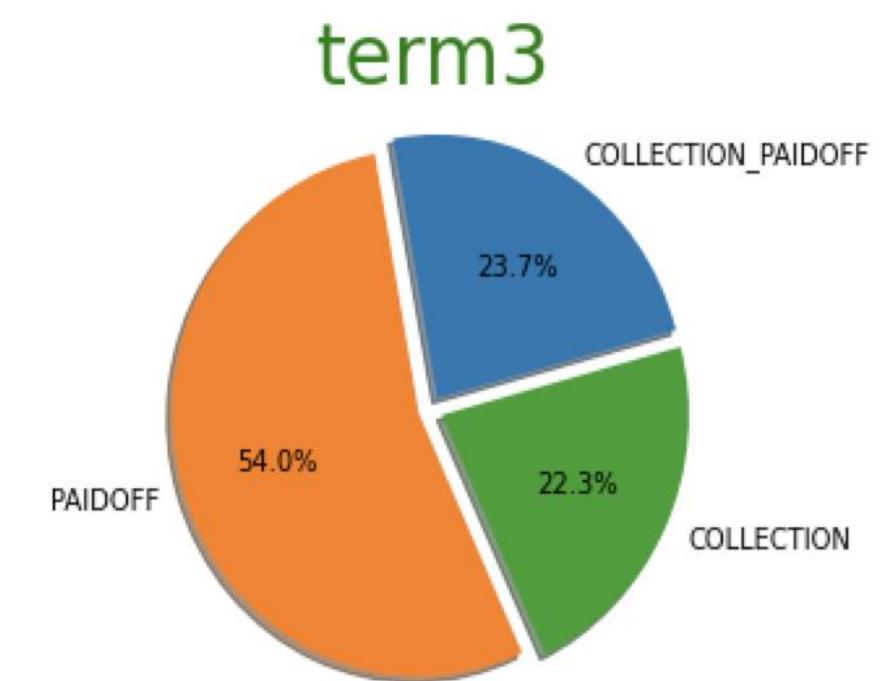
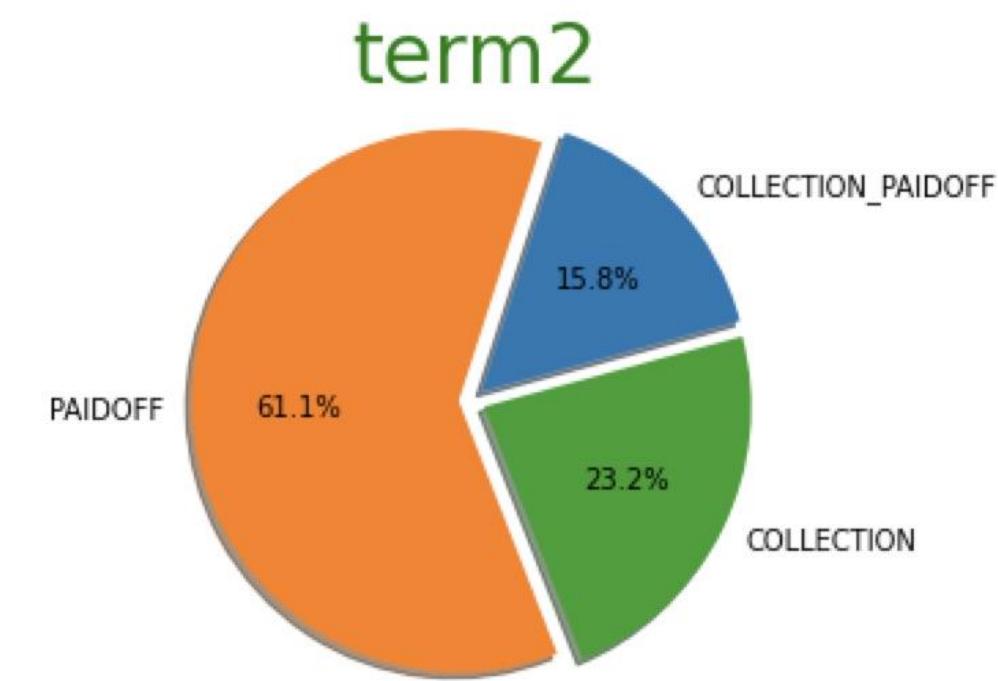
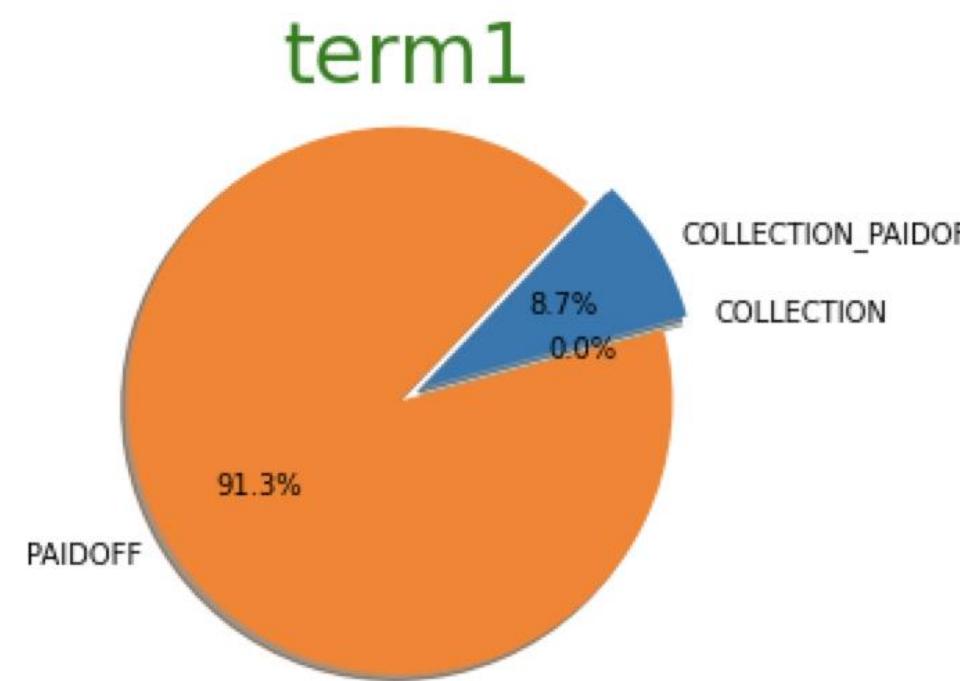
Class 1: [18, 29]
Class 2: [29, 40]
Class 3: [40, 52]



Data Visualization And Interpretation

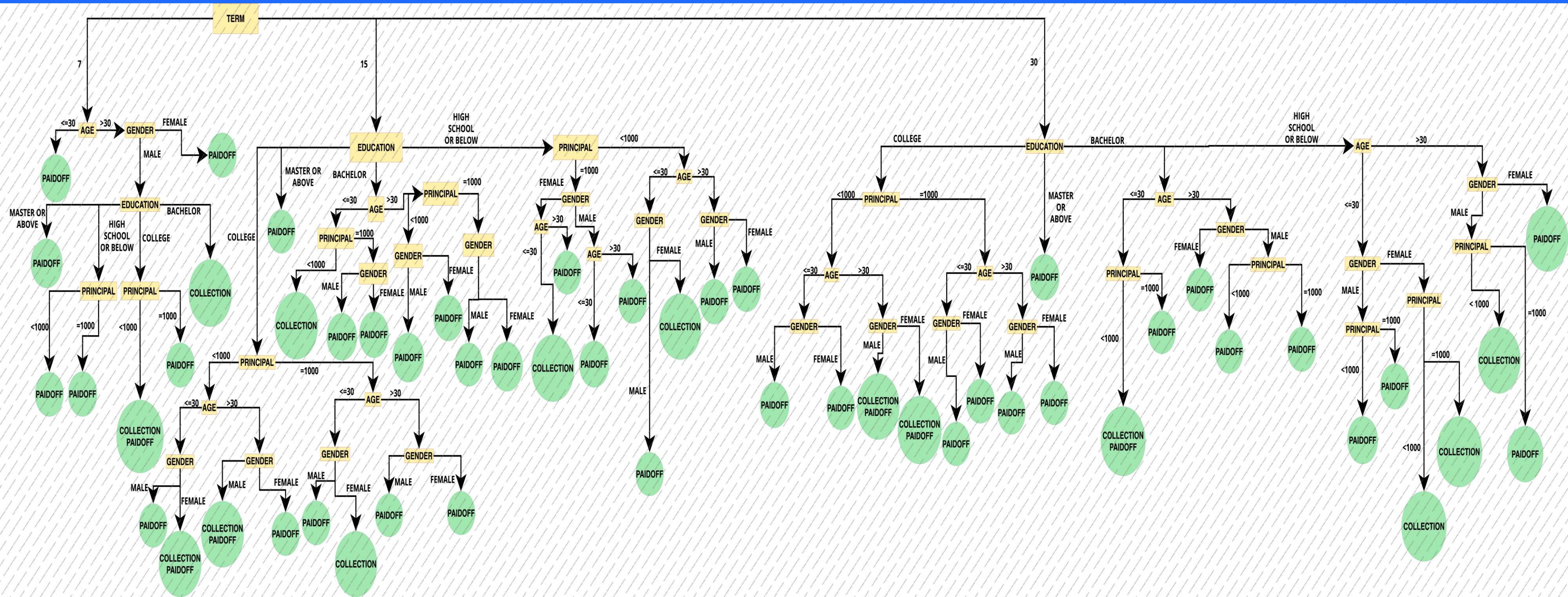
Pie Chart To Represent The Loan Status Based On Term

Class 1: [7, 14]
Class 2: [14, 21]
Class 3: [21, 31]



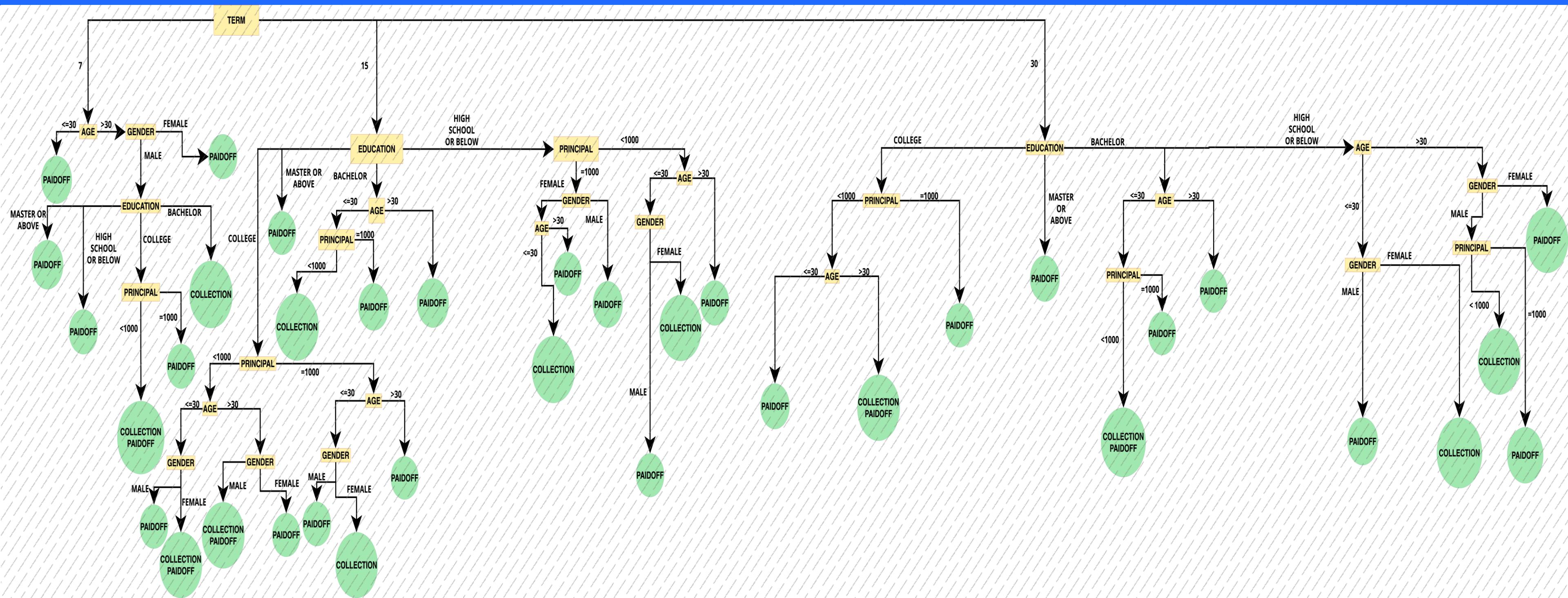
Data Visualization And Interpretation

Decision Tree



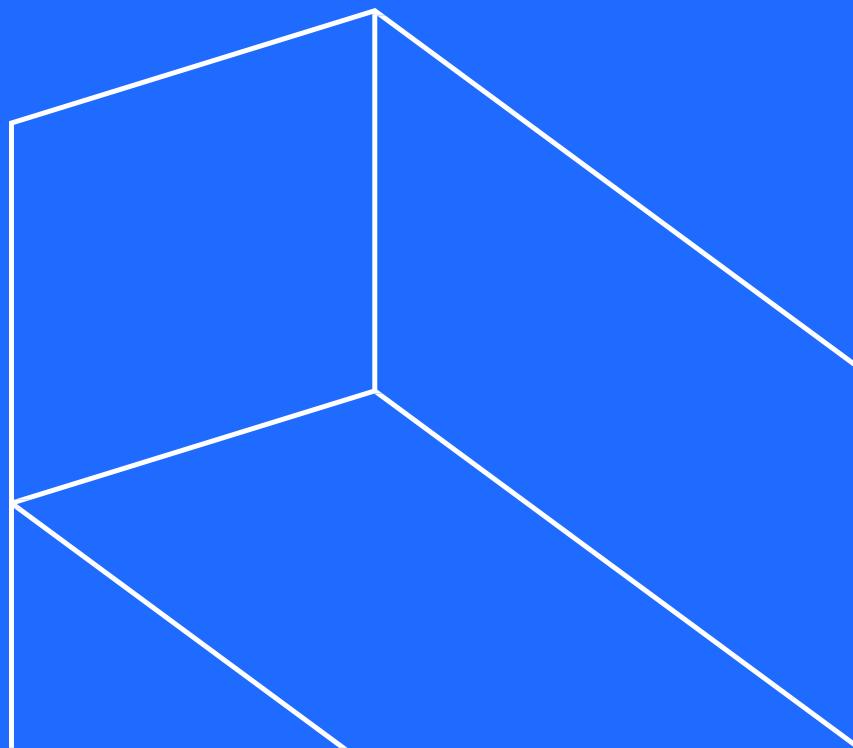
Data Visualization And Interpretation

Pruned Decision Tree



RESULTS

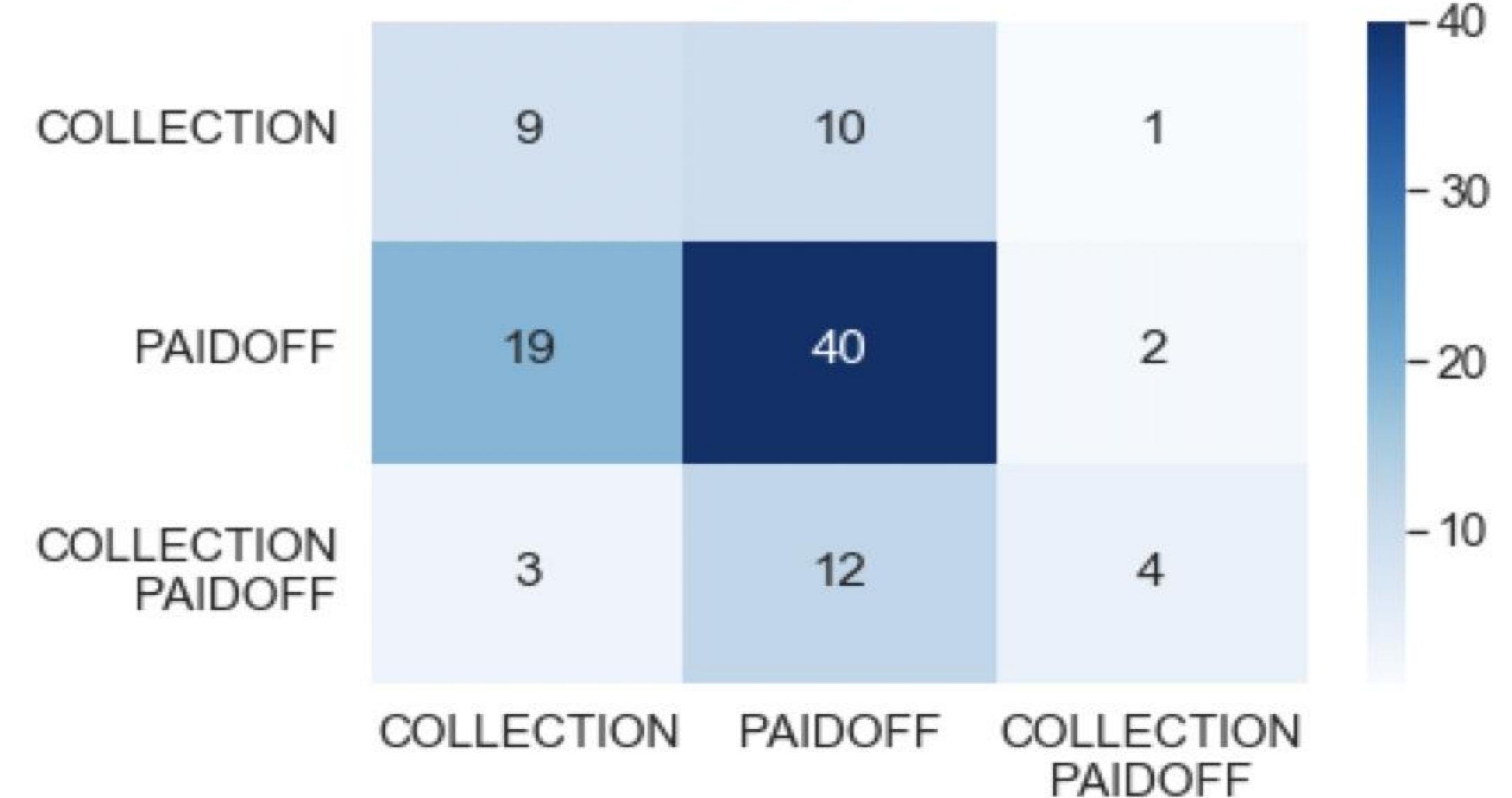
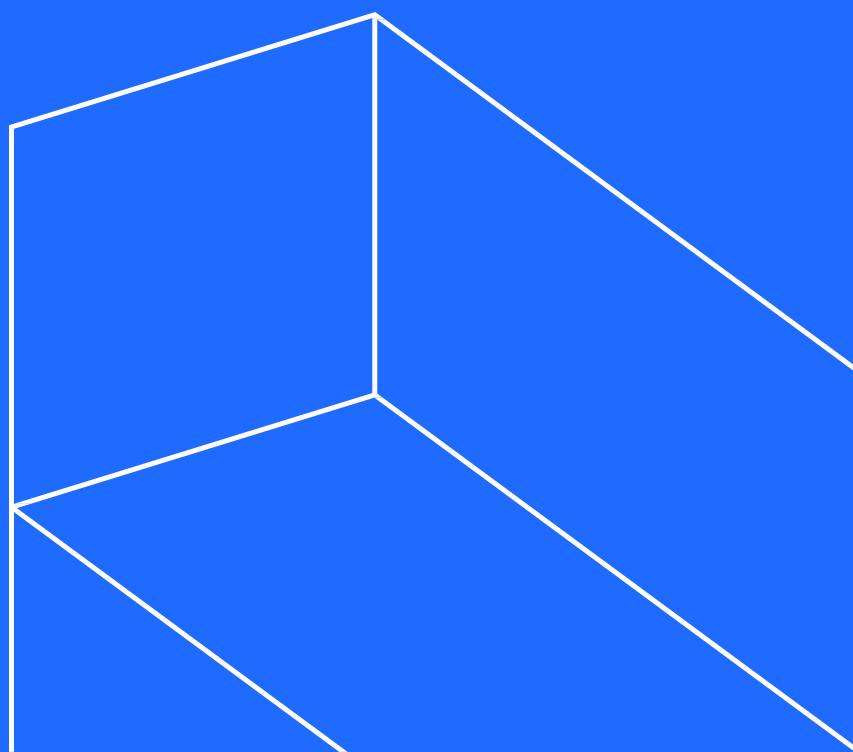
DECISION TREE
(WITHOUT SKLEARN)



	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.67	0.20	0.31	20
COLLECTION	0.50	0.16	0.24	19
PAIDOFF	0.65	0.93	0.77	61
accuracy			0.64	100
macro avg	0.60	0.43	0.44	100
weighted avg	0.62	0.64	0.57	100

RESULTS

DECISION TREE
(SKLEARN)

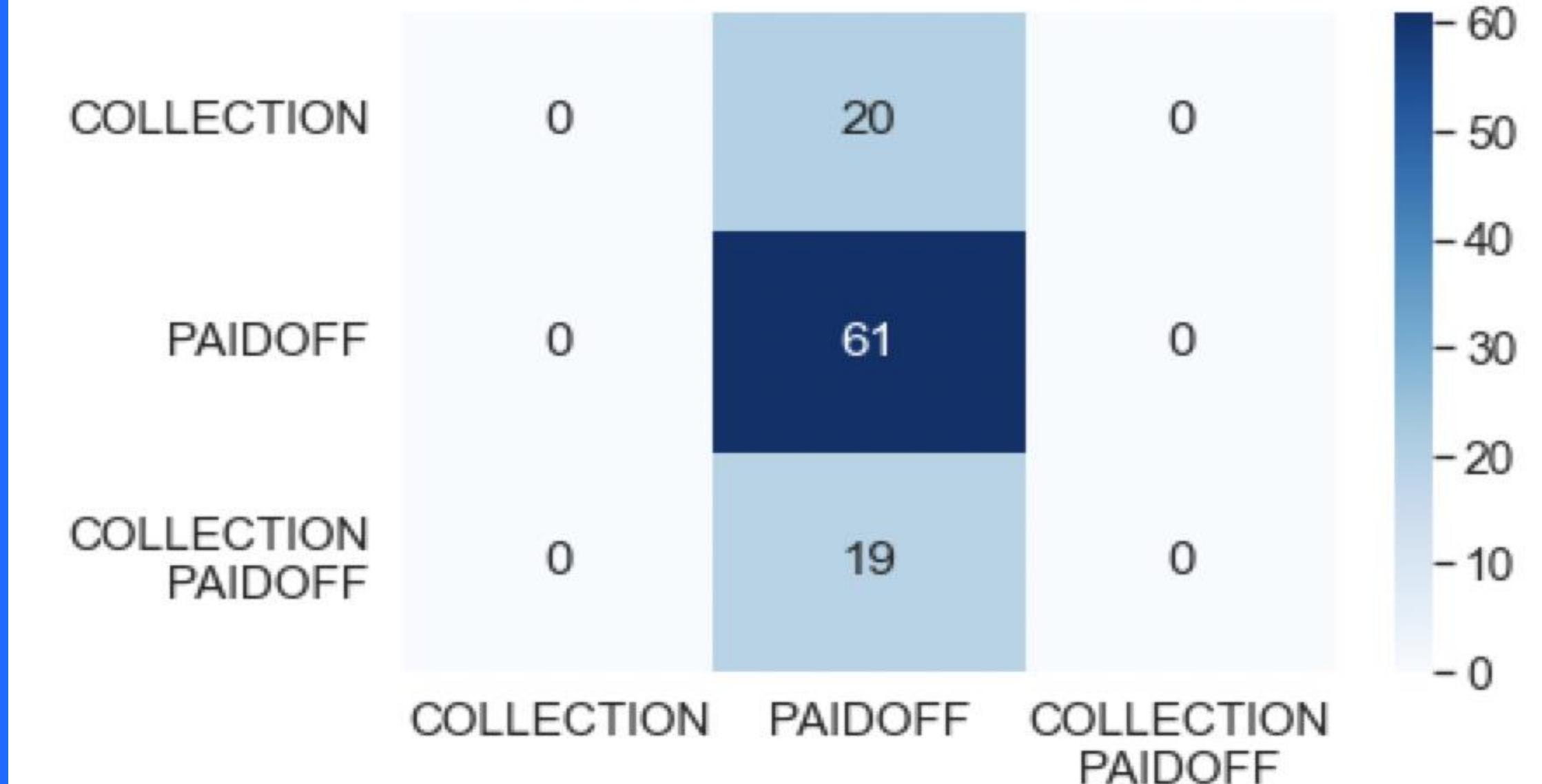


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.29	0.45	0.35	20
COLLECTION	0.65	0.66	0.65	61
PAIDOFF	0.57	0.21	0.31	19
accuracy			0.53	100
macro avg	0.50	0.44	0.44	100
weighted avg	0.56	0.53	0.53	100

RESULTS

SVM (SKLEARN)

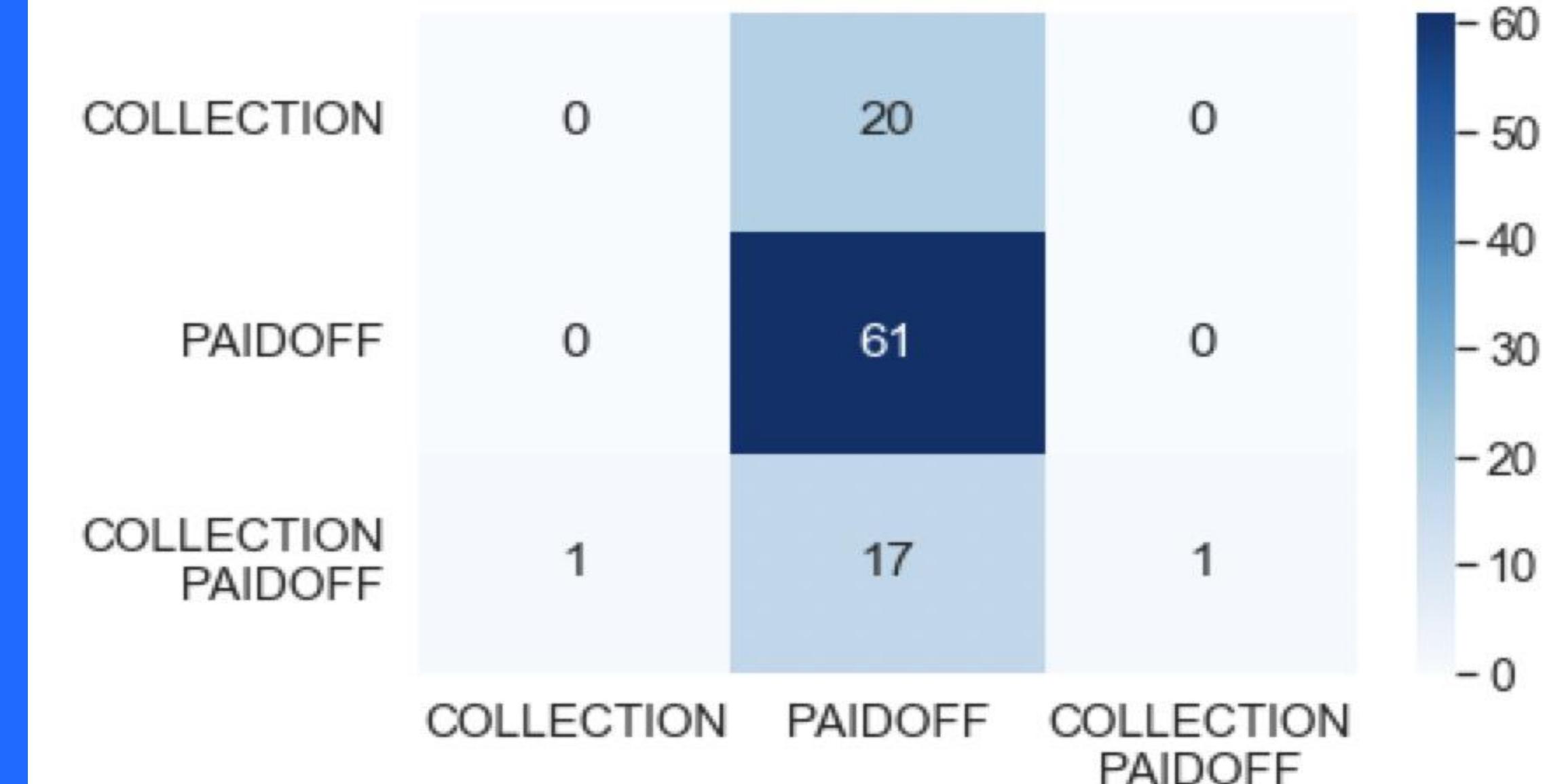


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	20
COLLECTION	0.61	1.00	0.76	61
PAIDOFF	0.00	0.00	0.00	19
accuracy			0.61	100
macro avg	0.20	0.33	0.25	100
weighted avg	0.37	0.61	0.46	100

RESULTS

KNN (SKLEARN)

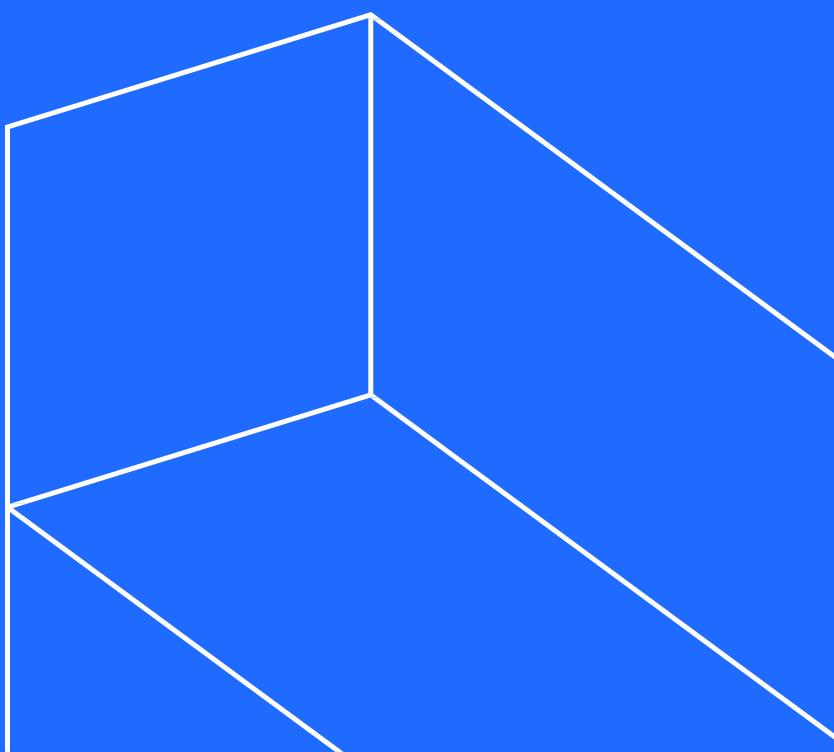


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	20
COLLECTION	0.62	1.00	0.77	61
PAIDOFF	1.00	0.05	0.10	19
accuracy			0.62	100
macro avg	0.54	0.35	0.29	100
weighted avg	0.57	0.62	0.49	100

RESULTS

R A N D O M F O R E S T
(S K L E A R N)

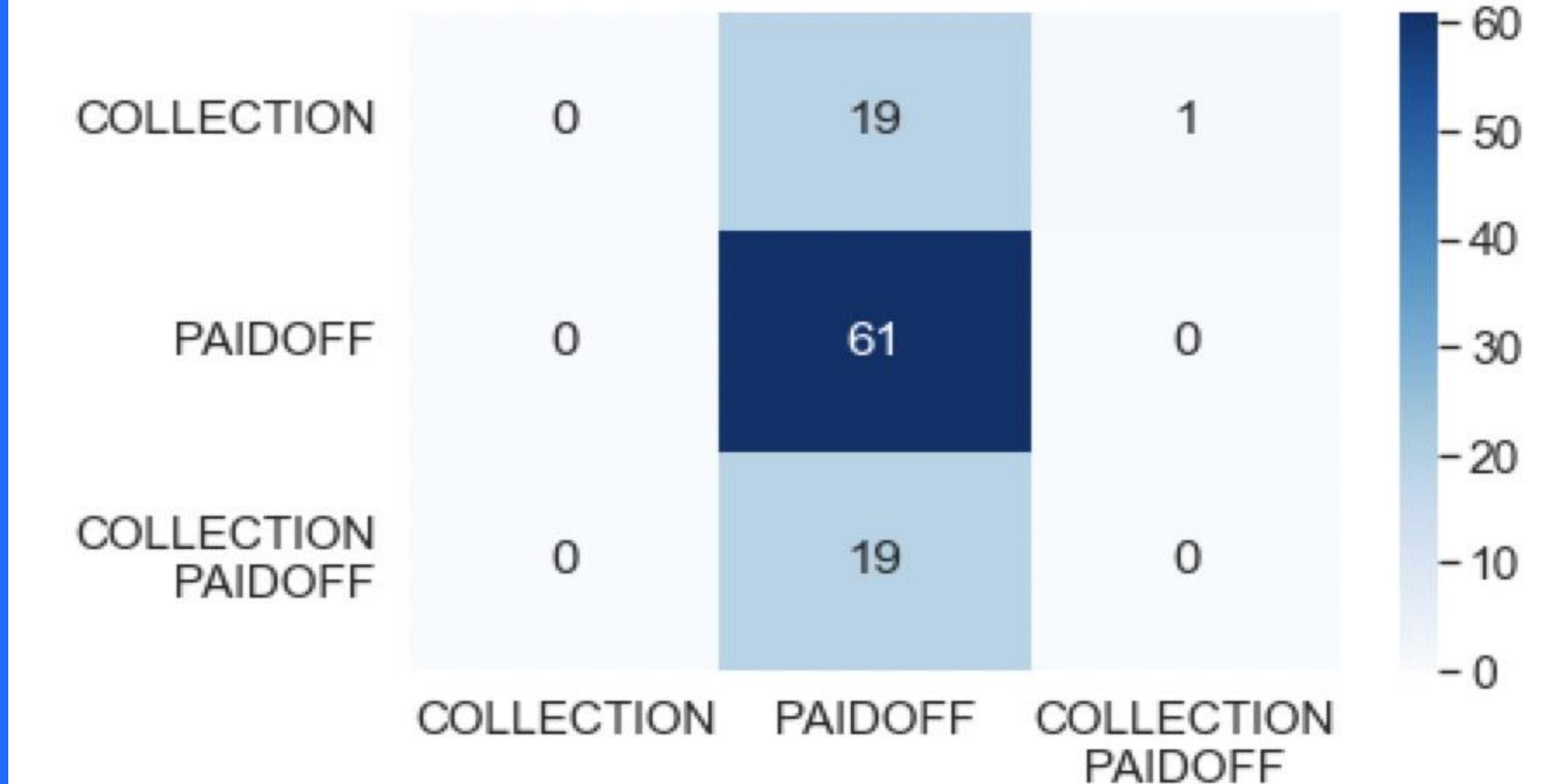


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.32	0.35	0.33	20
COLLECTION	0.65	0.72	0.68	61
PAIDOFF	0.40	0.21	0.28	19
accuracy			0.55	100
macro avg	0.46	0.43	0.43	100
weighted avg	0.53	0.55	0.54	100

RESULTS

NAIVE BAYES
(SKLEARN)

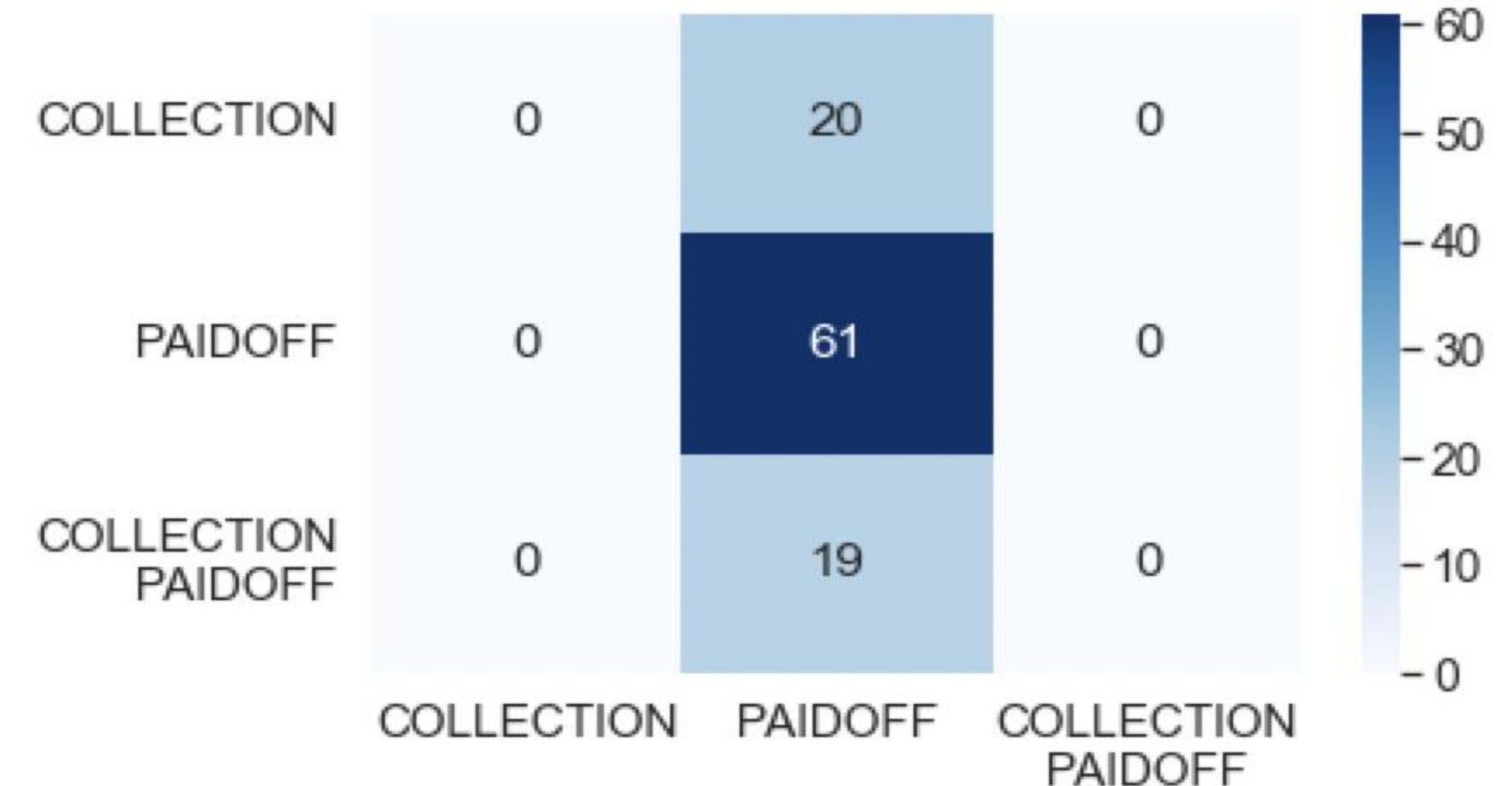
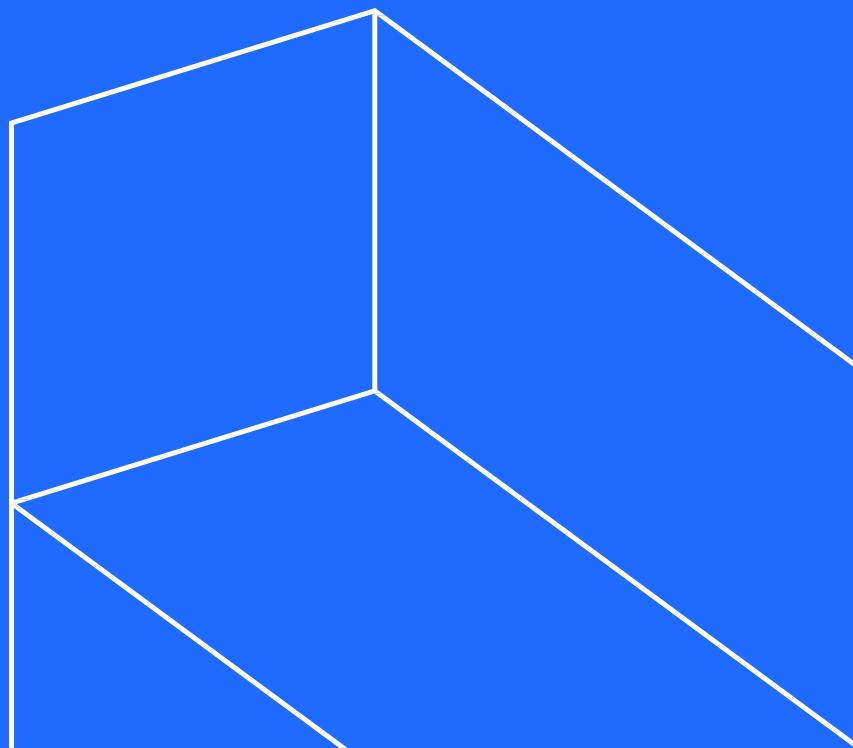


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	20
COLLECTION	0.62	1.00	0.76	61
PAIDOFF	0.00	0.00	0.00	19
accuracy			0.61	100
macro avg	0.21	0.33	0.25	100
weighted avg	0.38	0.61	0.47	100

RESULTS

LOGISTIC REGRESSION
(SKLEARN)



Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	20
COLLECTION	0.61	1.00	0.76	61
PAIDOFF	0.00	0.00	0.00	19
accuracy			0.61	100
macro avg	0.20	0.33	0.25	100
weighted avg	0.37	0.61	0.46	100

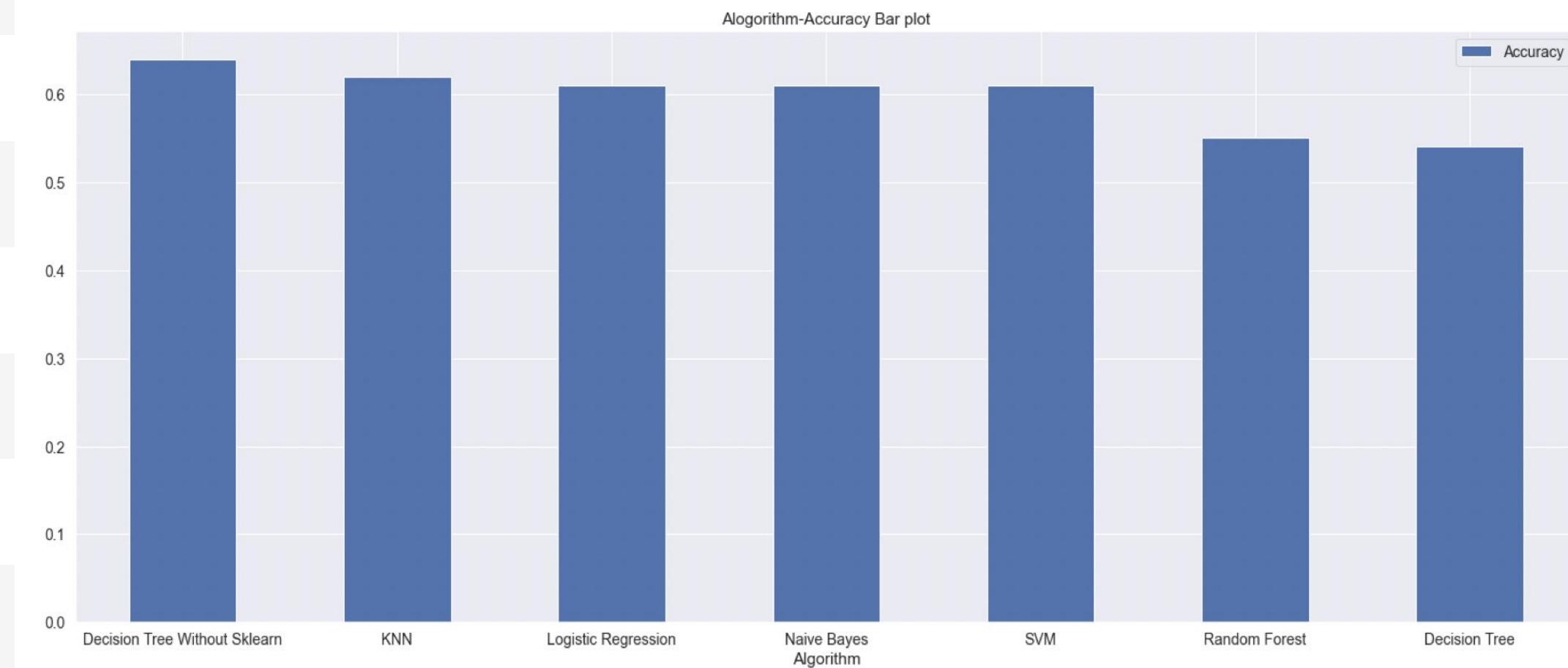
CONCLUSION

Best accuracy obtained is 64% using the decision tree implemented from scratch.

Out of 100 loan requests 57 were approved.

Algorithm Accuracy

	Algorithm	Accuracy
0	Decision Tree Without Sklearn	0.64
1	KNN	0.62
2	Logistic Regression	0.61
3	Naive Bayes	0.61
4	SVM	0.61
5	Random Forest	0.55
6	Decision Tree	0.54



DATASET-2

Gokul Krishnan G

AM.EN.U4CSE19321

This dataset consists of 11 columns and 500 rows.

Features

- Education
- Principal
- Age
- Term
- Gender

Target - Loan Status

- PAIDOFF
- COLLECTION

Links:

[Dataset Files](#)

```
RangeIndex: 500 entries, 0 to 499
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Loan_ID          500 non-null    object  
 1   loan_status       500 non-null    object  
 2   Principal         500 non-null    int64  
 3   terms             500 non-null    int64  
 4   effective_date   500 non-null    datetime64[ns]
 5   due_date          500 non-null    datetime64[ns]
 6   paid_off_time    399 non-null    object  
 7   past_due_days    206 non-null    float64 
 8   age               500 non-null    int64  
 9   education         500 non-null    object  
 10  Gender            500 non-null    object  
dtypes: datetime64[ns](2), float64(1), int64(3), object(5)
memory usage: 43.1+ KB
```

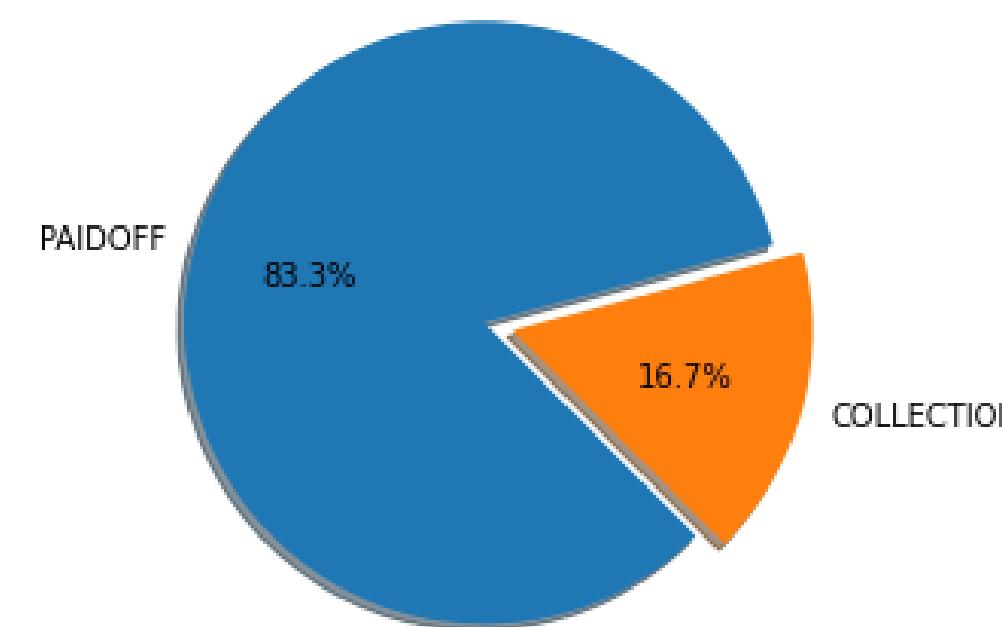
	Loan_ID	loan_status	Principal	terms	effective_date	due_date	paid_off_time	past_due_days	age	education	Gender
0	xqd20161038	COLLECTION	1000	30	9/11/2016	10/10/2016	NaN	59.0	23	High School or Below	male
1	xqd20160436	PAIDOFF	1000	30	9/12/2016	10/11/2016	10/9/2016 10:00	NaN	30	High School or Below	male
2	xqd20161050	COLLECTION	800	15	9/11/2016	9/25/2016	NaN	74.0	29	college	male
3	xqd20160693	PAIDOFF	800	30	9/14/2016	11/12/2016	11/12/2016 9:00	NaN	38	High School or Below	male
4	xqd20161262	COLLECTION	1000	30	9/11/2016	10/10/2016	10/11/2016 11:29	1.0	38	High School or Below	male
...
495	xqd20160479	PAIDOFF	1000	30	9/13/2016	10/12/2016	10/9/2016 21:57	NaN	21	High School or Below	male
496	xqd20160097	PAIDOFF	1000	15	9/11/2016	9/25/2016	9/20/2016 5:38	NaN	26	High School or Below	male
497	xqd20160064	PAIDOFF	1000	30	9/11/2016	10/10/2016	9/12/2016 11:40	NaN	45	college	male
498	xqd20161438	COLLECTION	1000	30	9/11/2016	10/10/2016	10/11/2016 16:13	1.0	34	High School or Below	male
499	xqd20160657	PAIDOFF	1000	15	9/12/2016	9/26/2016	9/26/2016 23:00	NaN	27	High School or Below	male

PAIDOFF 294
COLLECTION 206
Name: loan_status, dtype: int64

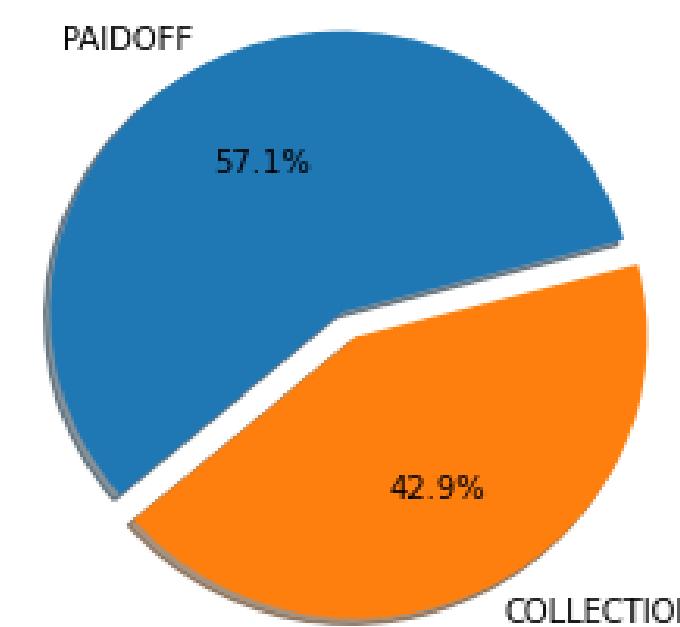
Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Of
Different Education Groups

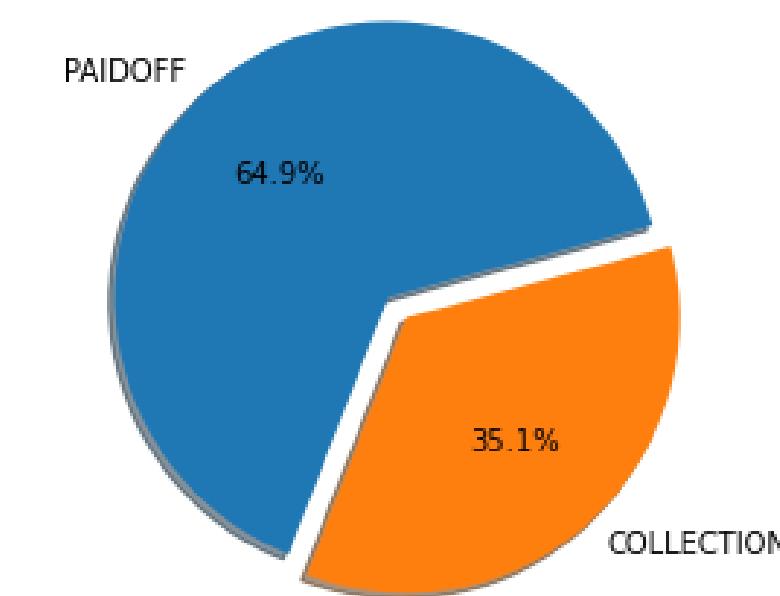
Master or Above



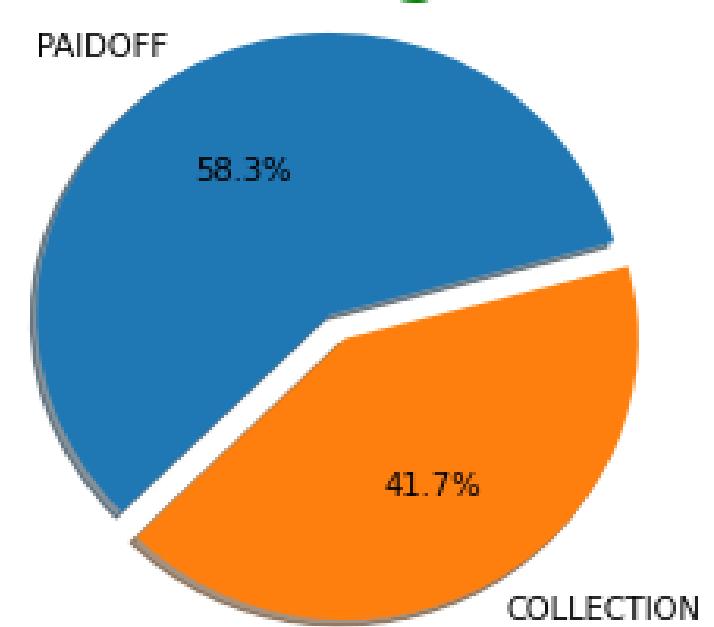
High School or Below



Bechelor



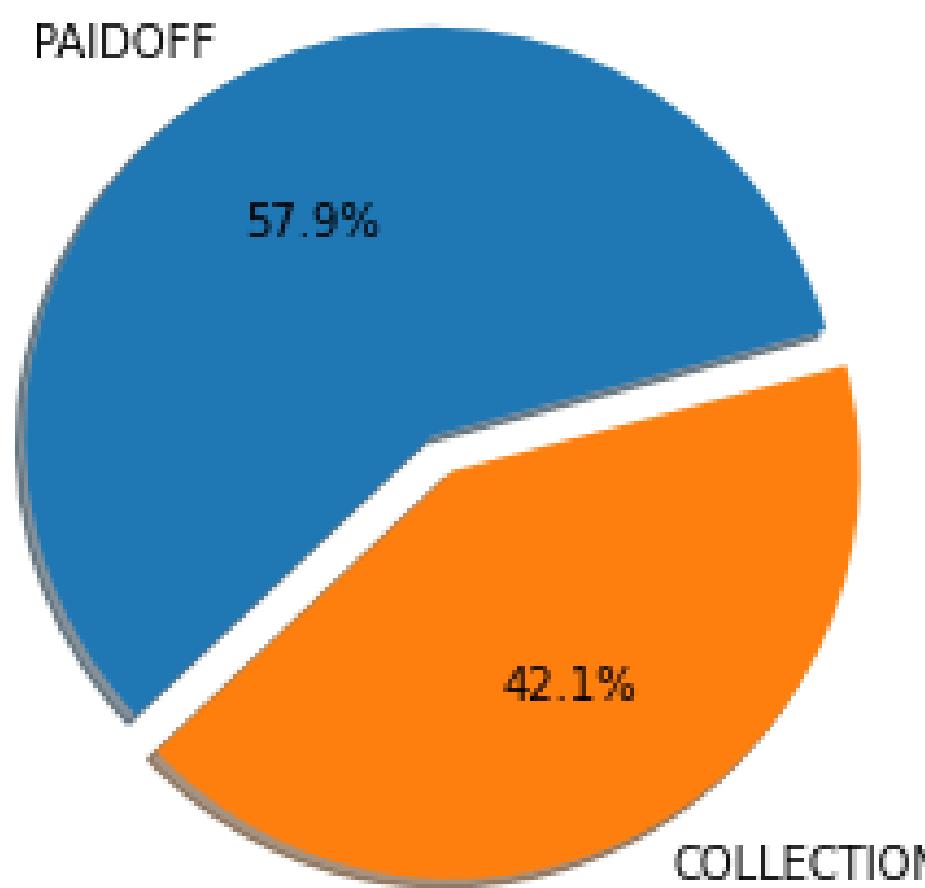
college



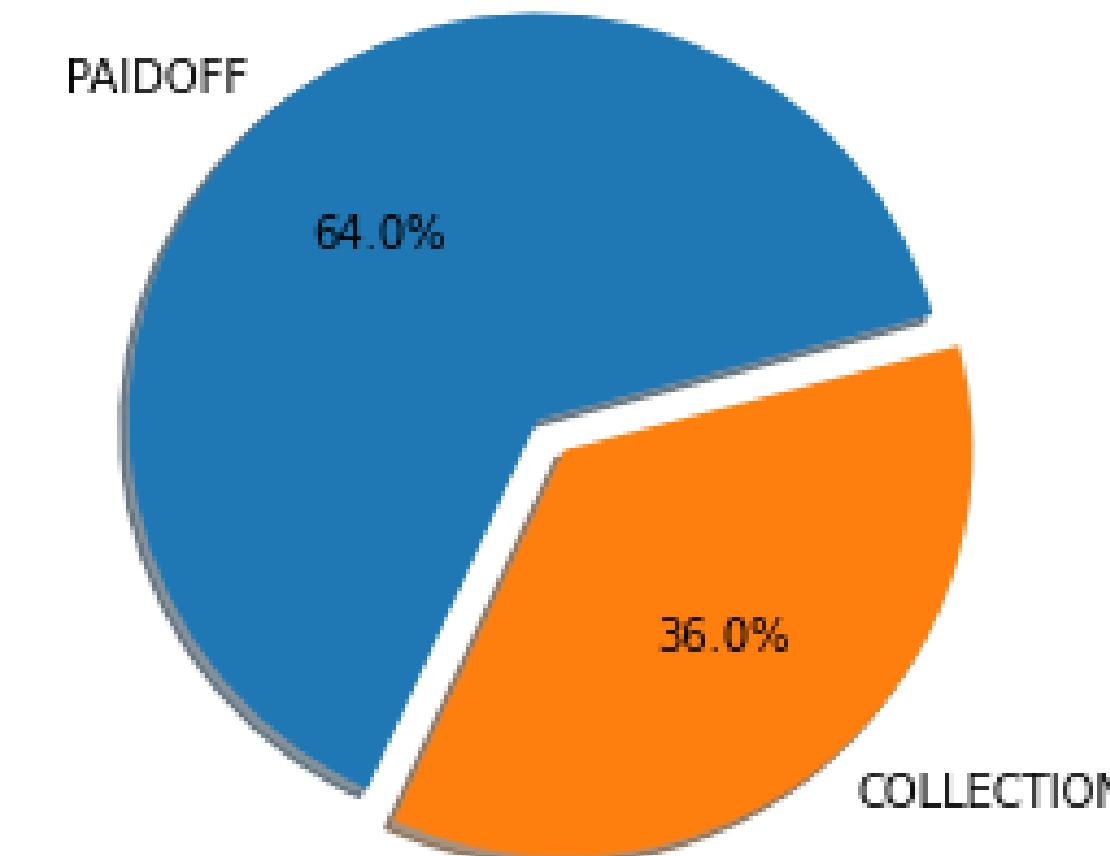
Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Based On Gender

male



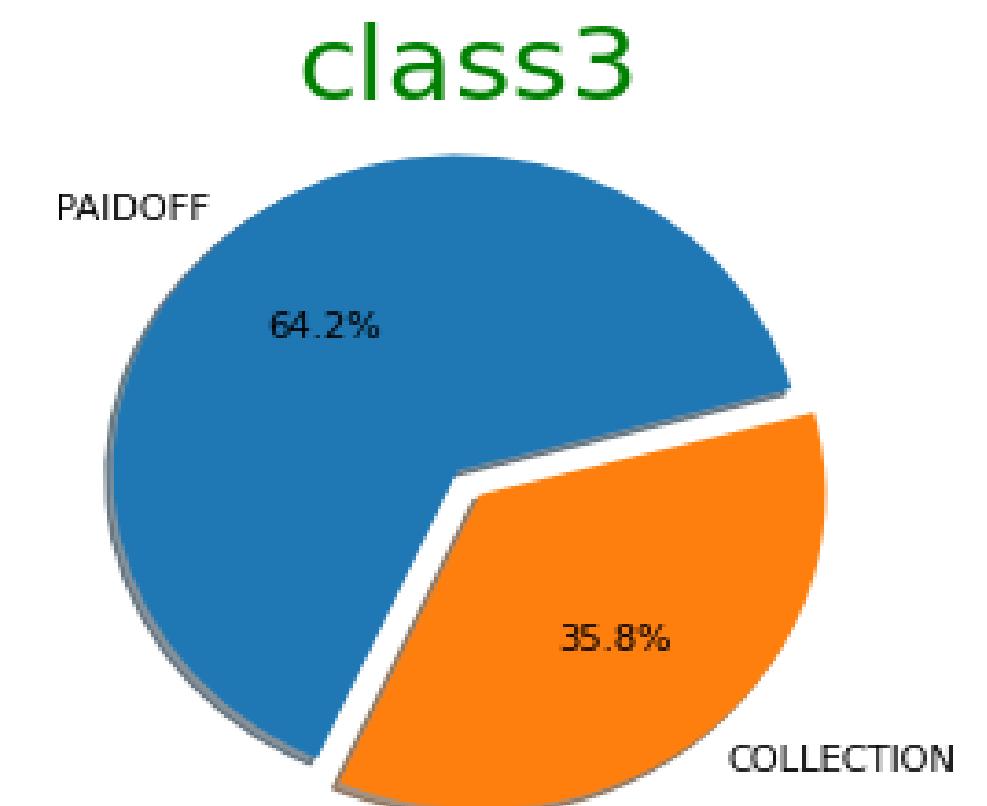
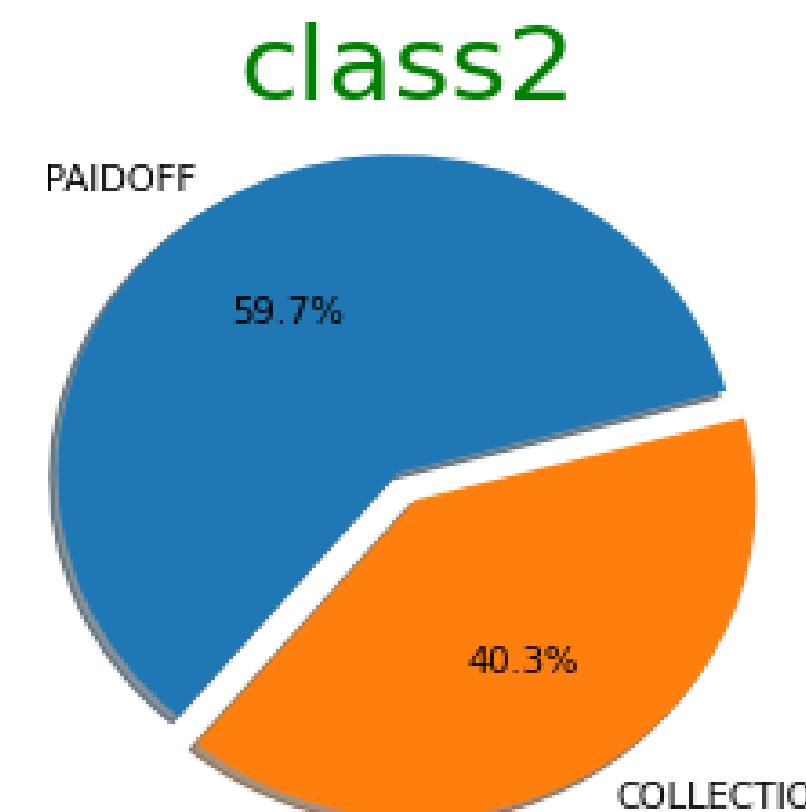
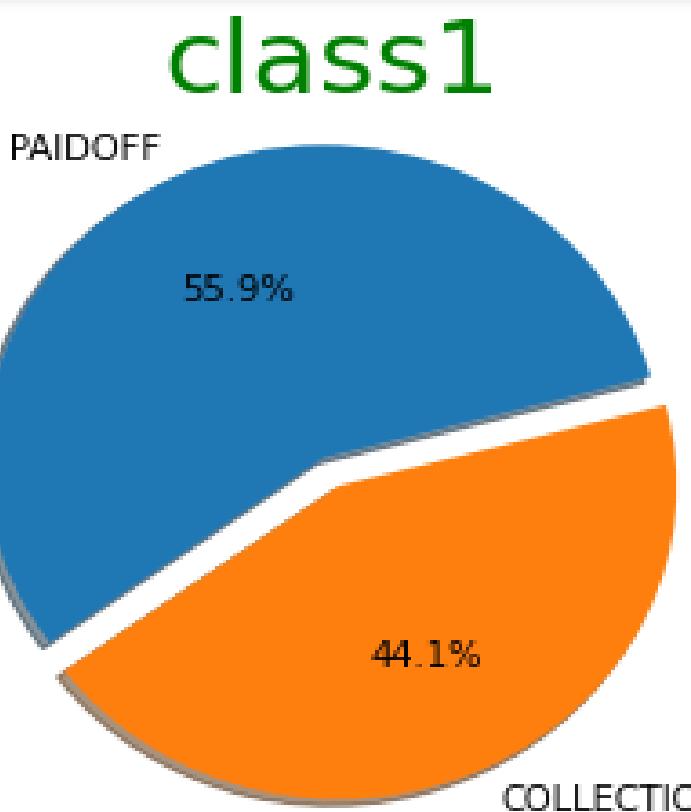
female



Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Based On Age

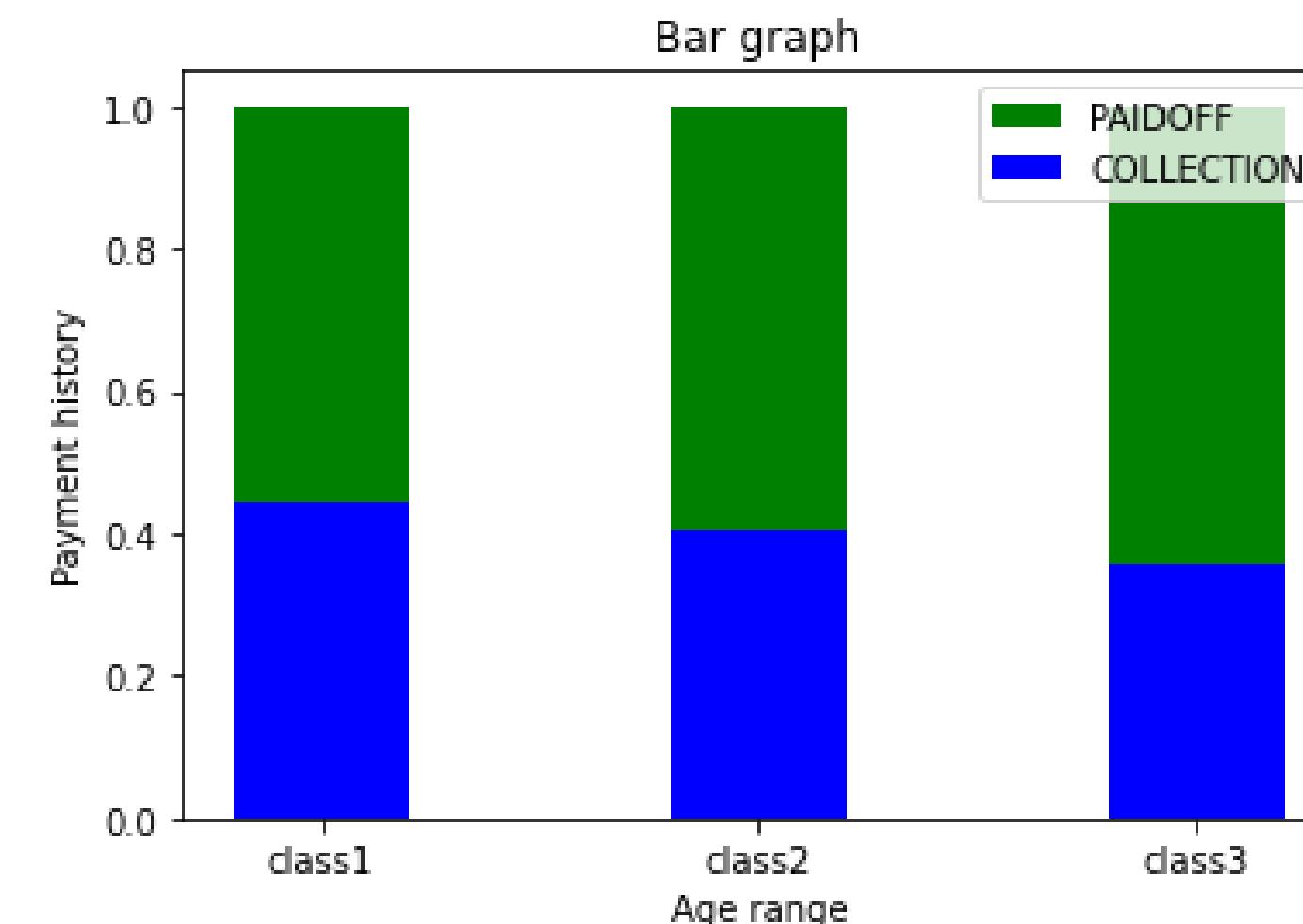
Class 1: [19, 29]
Class 2: [29, 39]
Class 3: [39, 52]



Data Visualization And Interpretation

Bar Chart Between Loan Status And Age

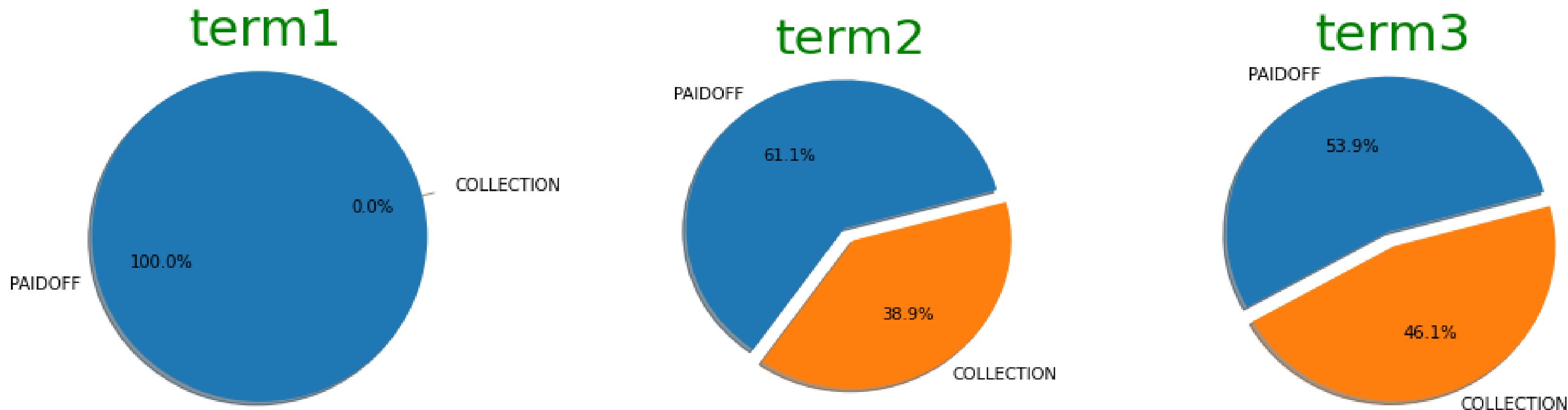
Class 1: [19, 29]
Class 2: [29, 39]
Class 3: [39, 52]



Data Visualization And Interpretation

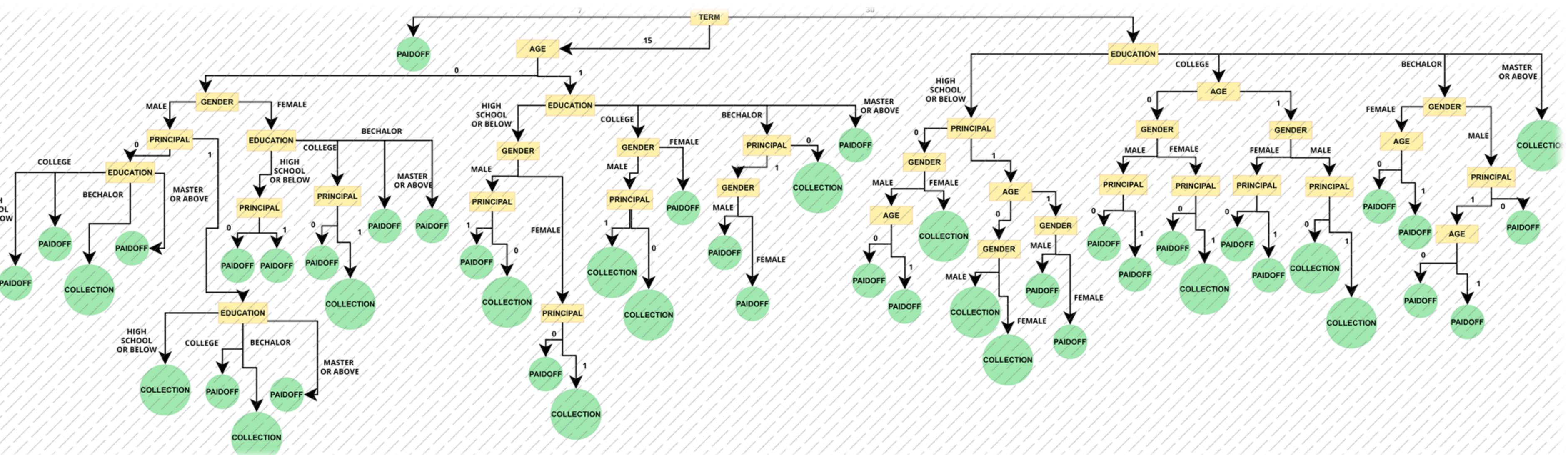
Pie Chart To Represent The Loan Status Based On Term

Class 1: [7, 14]
Class 2: [14, 21]
Class 3: [21, 31]



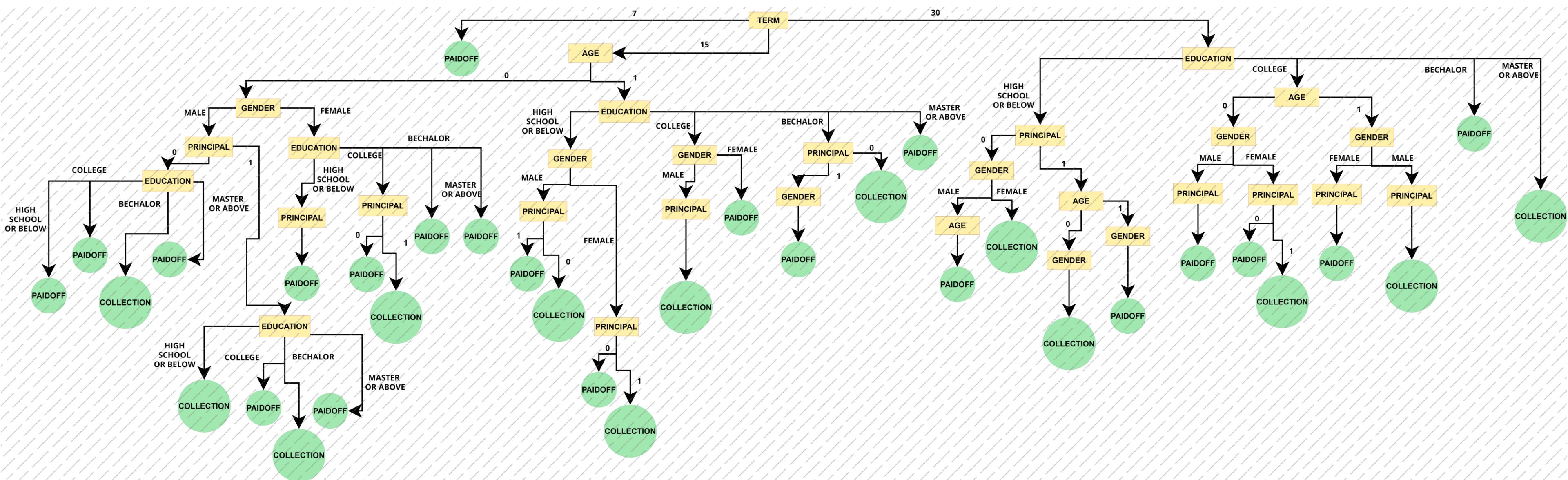
Data Visualization And Interpretation

Decision Tree



Data Visualization And Interpretation

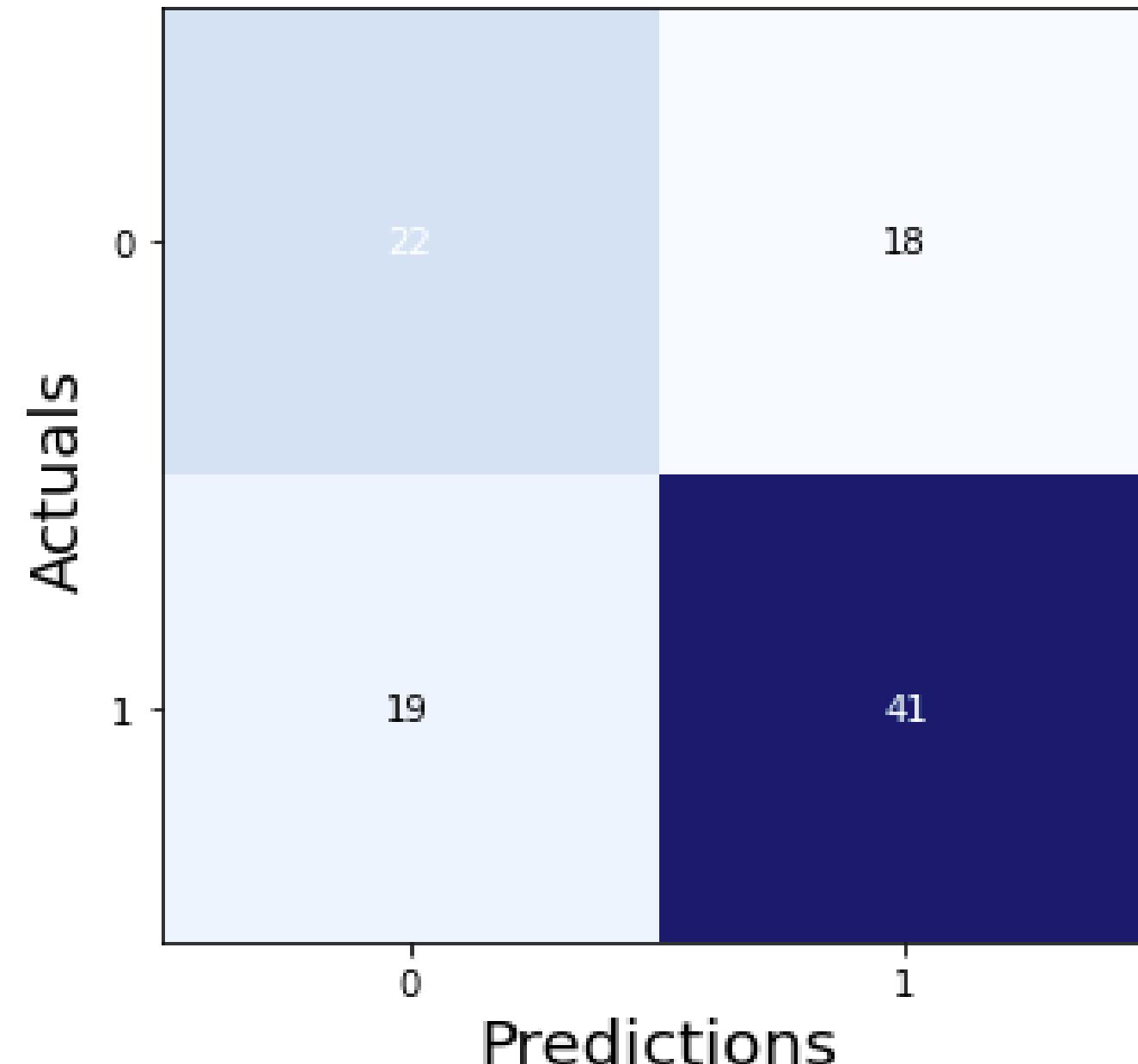
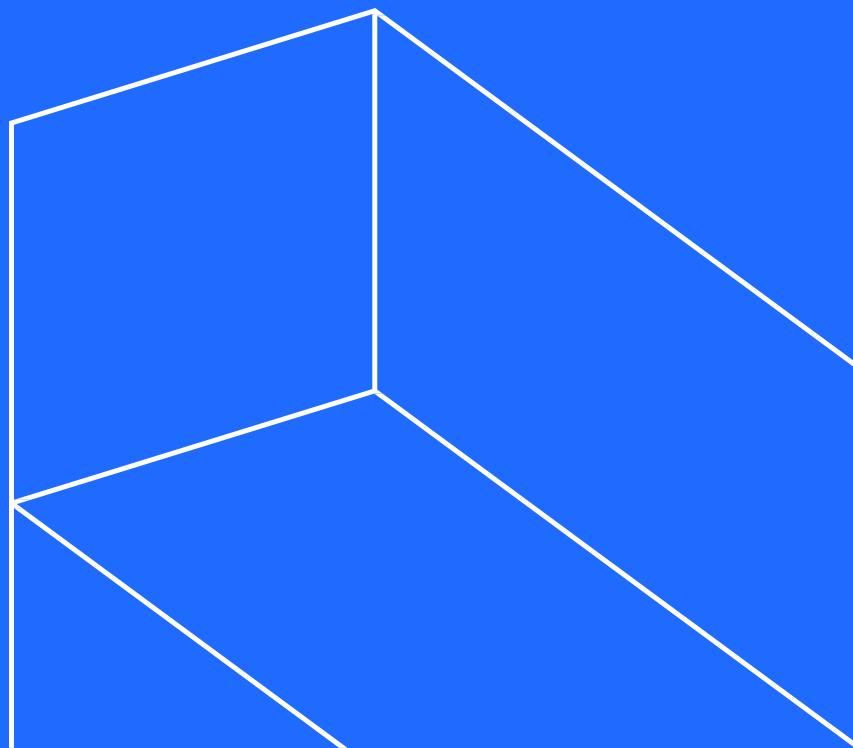
Pruned Decision Tree



RESULTS

DECISION TREE

(WITHOUT SKLEARN)

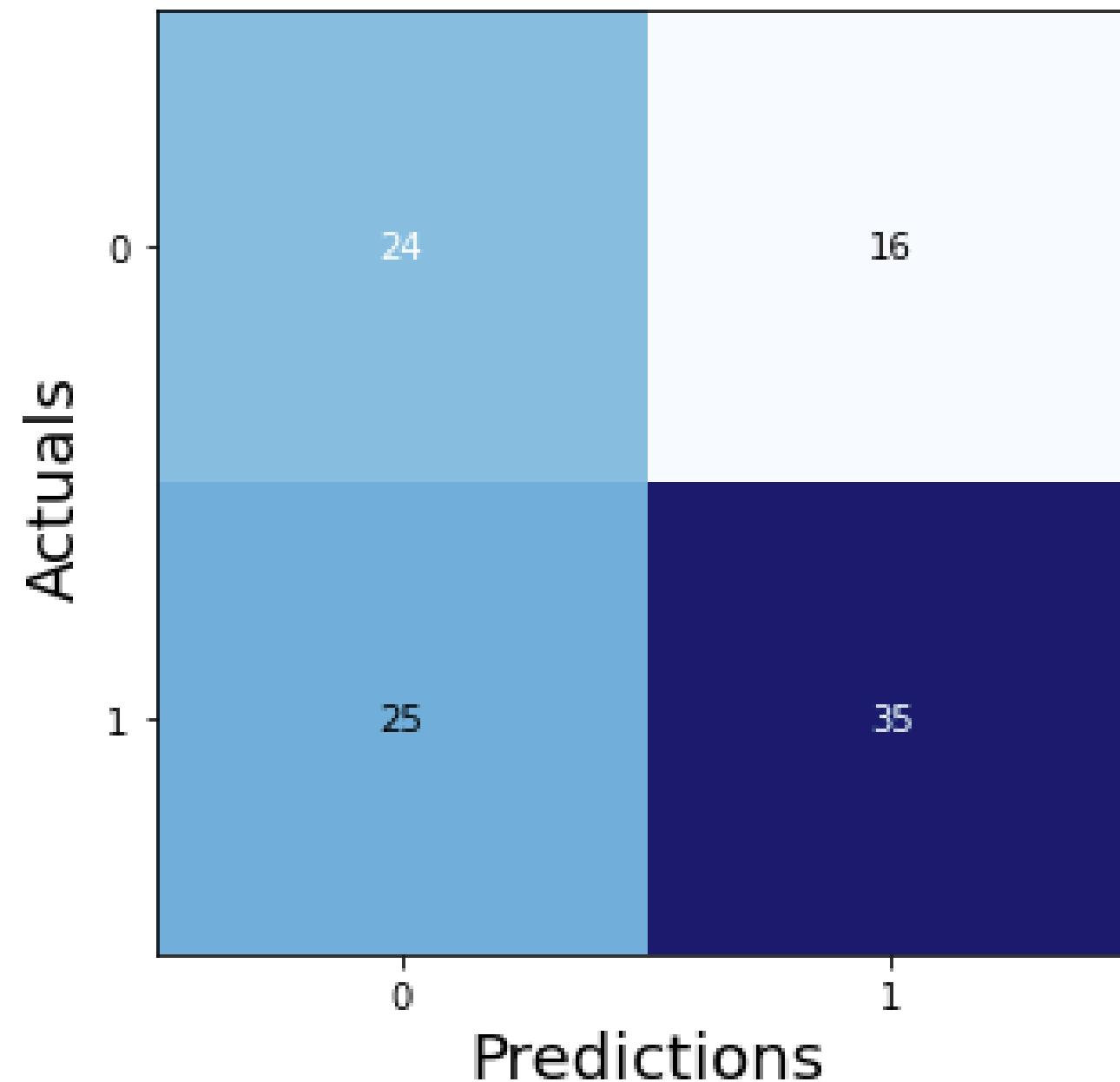
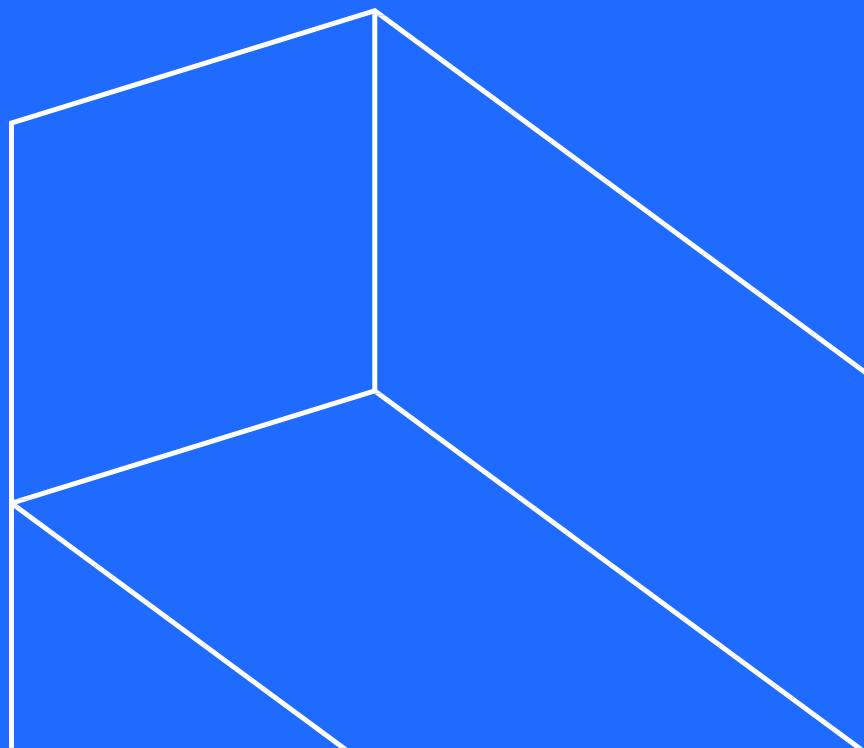


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.54	0.55	0.54	40
PAIDOFF	0.69	0.68	0.69	60
accuracy			0.63	100
macro avg	0.62	0.62	0.62	100
weighted avg	0.63	0.63	0.63	100

RESULTS

DECISION TREE
(SKLEARN)

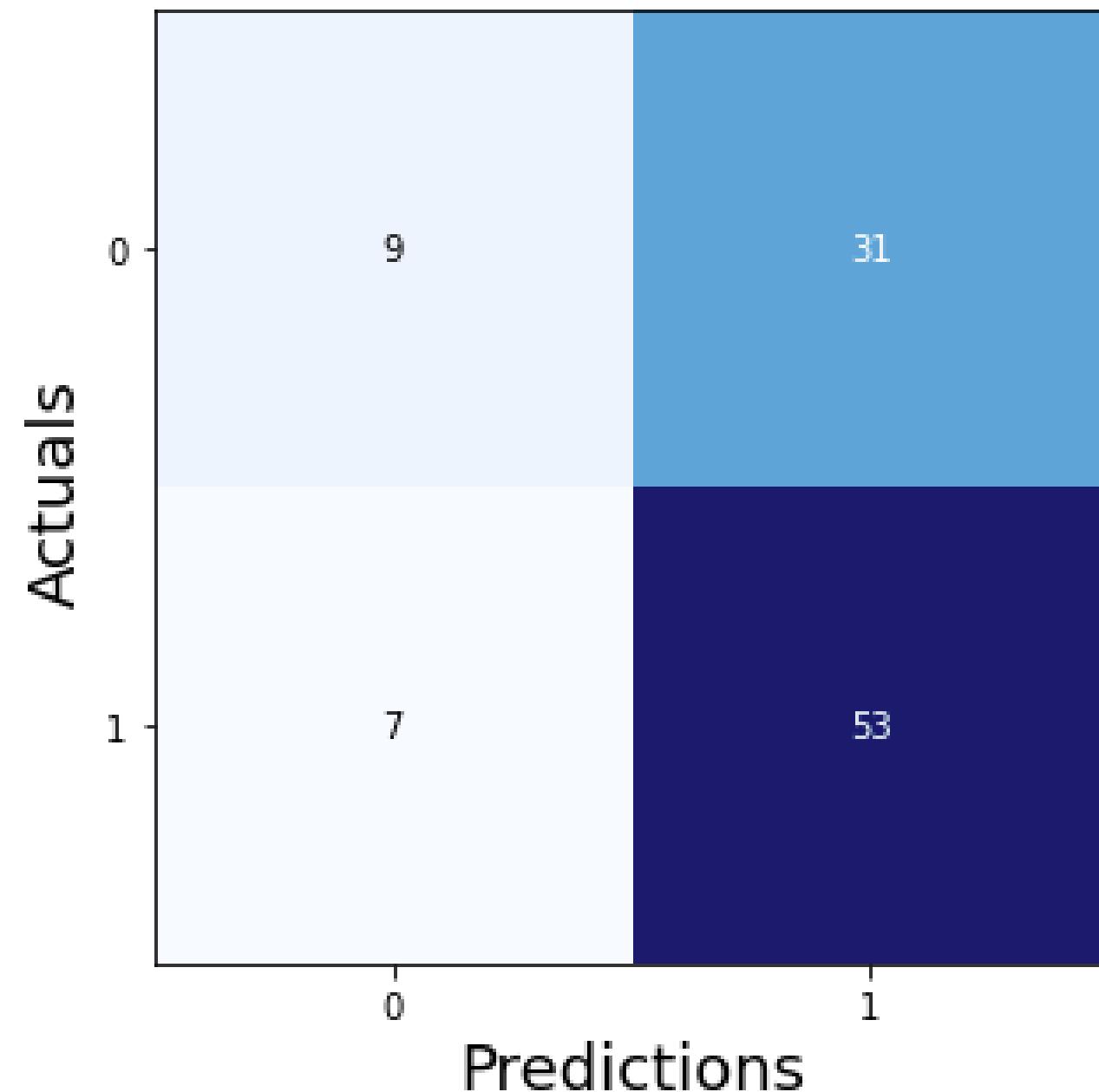
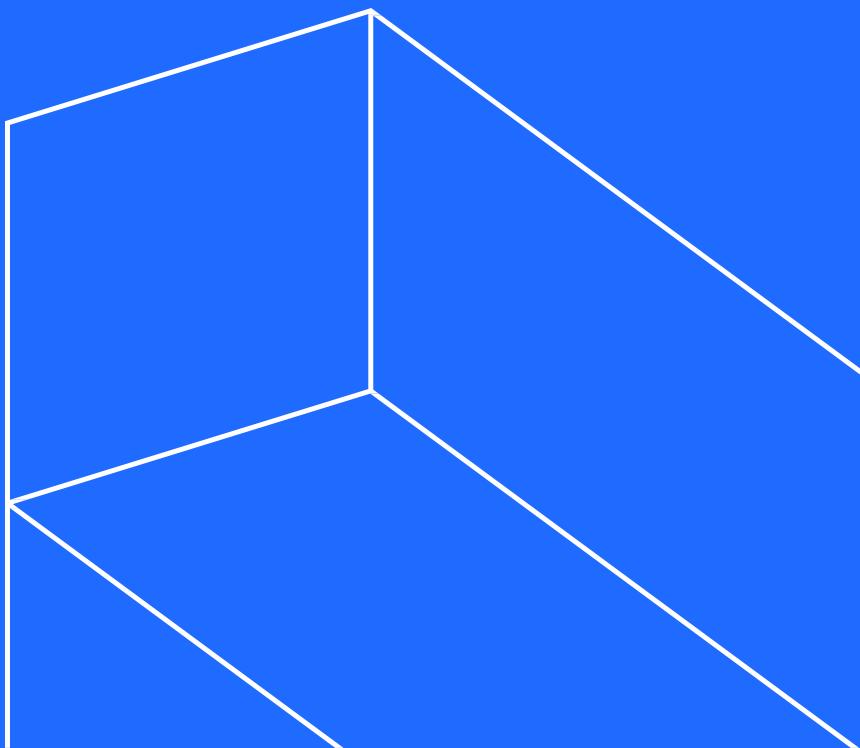


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.49	0.60	0.54	40
PAIDOFF	0.69	0.58	0.63	60
accuracy			0.59	100
macro avg	0.59	0.59	0.58	100
weighted avg	0.61	0.59	0.59	100

RESULTS

SVM (SKLEARN)

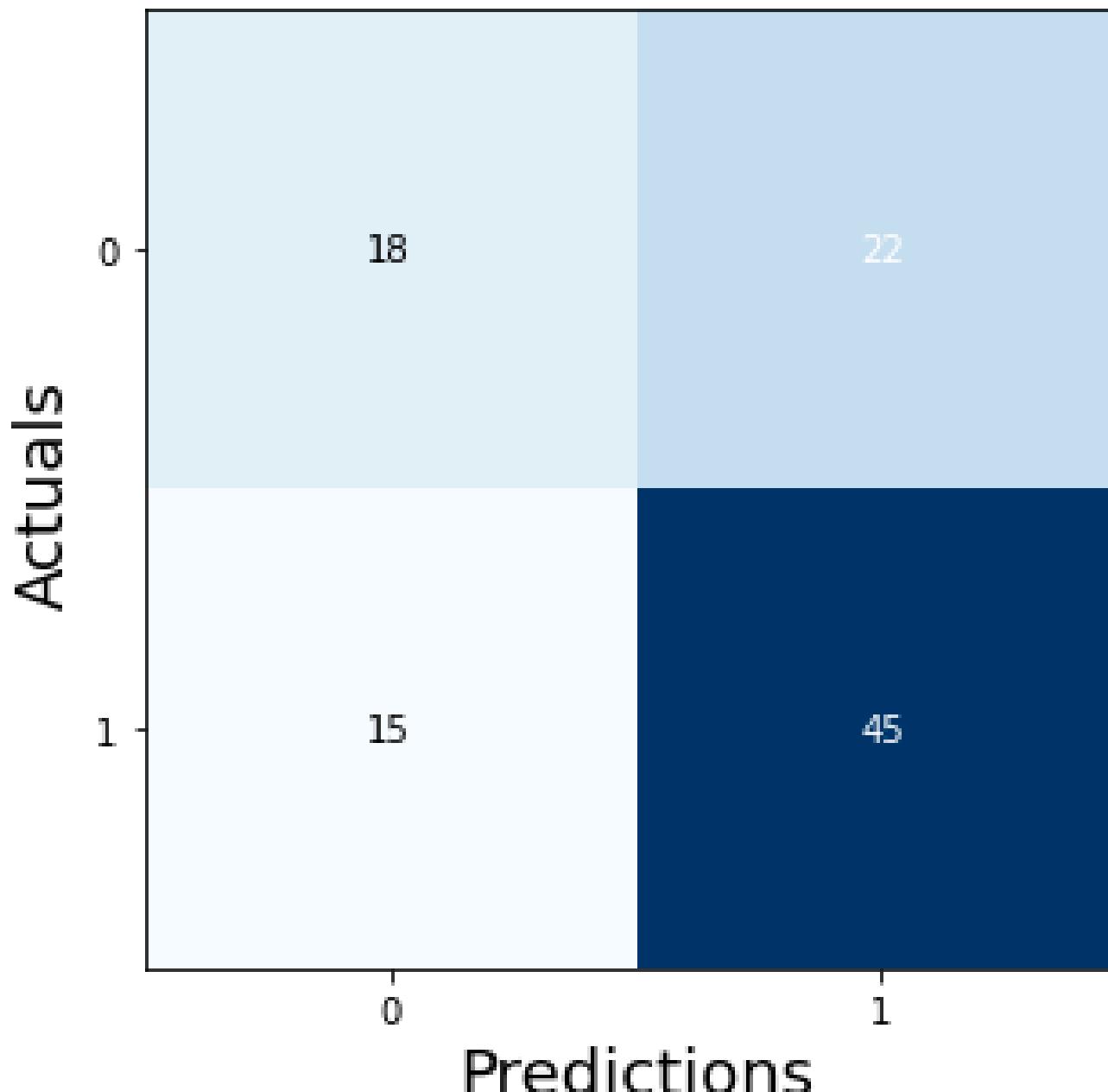
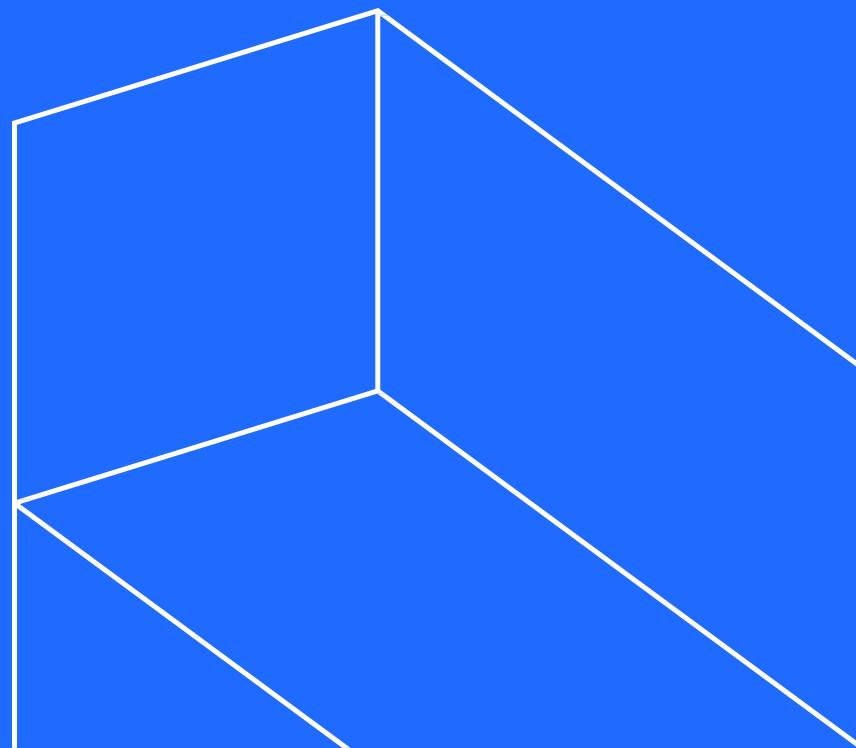


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.56	0.23	0.32	40
PAIDOFF	0.63	0.88	0.74	60
accuracy			0.62	100
macro avg	0.60	0.55	0.53	100
weighted avg	0.60	0.62	0.57	100

RESULTS

KNN (SKLEARN)

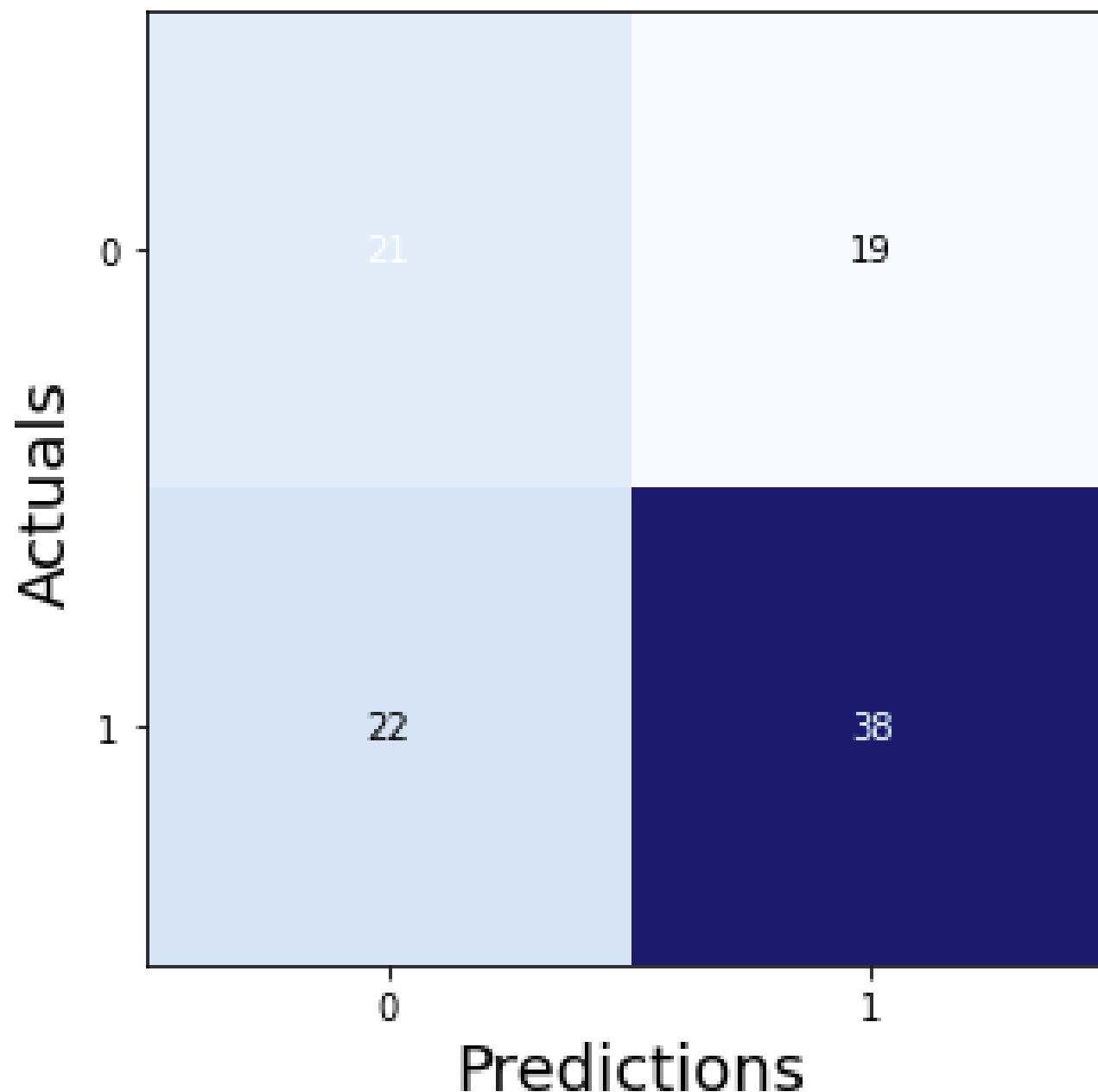
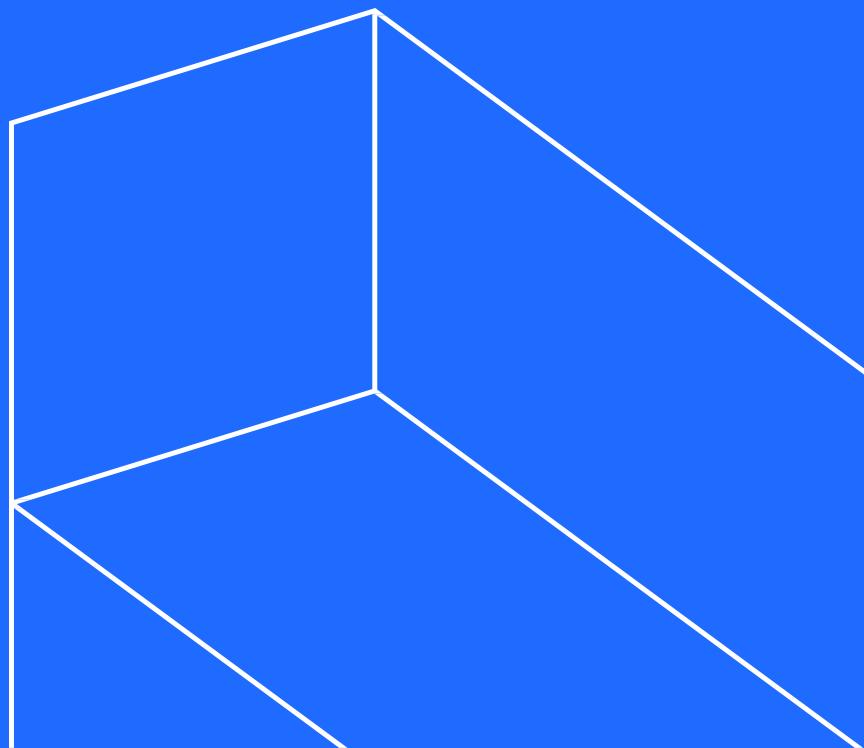


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.55	0.45	0.49	40
PAIDOFF	0.67	0.75	0.71	60
accuracy			0.63	100
macro avg	0.61	0.60	0.60	100
weighted avg	0.62	0.63	0.62	100

RESULTS

RANDOM FOREST
(SKLEARN)

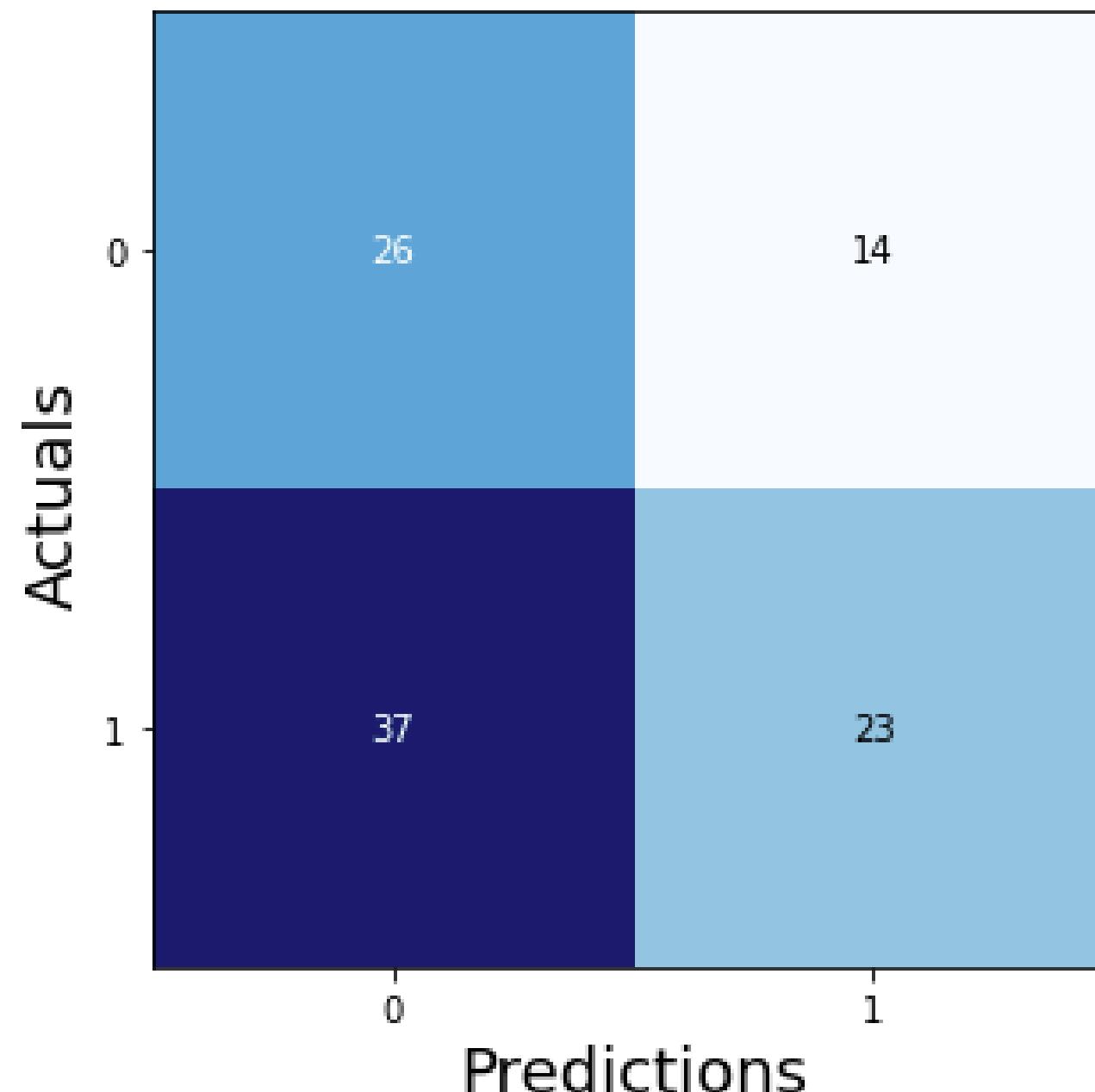
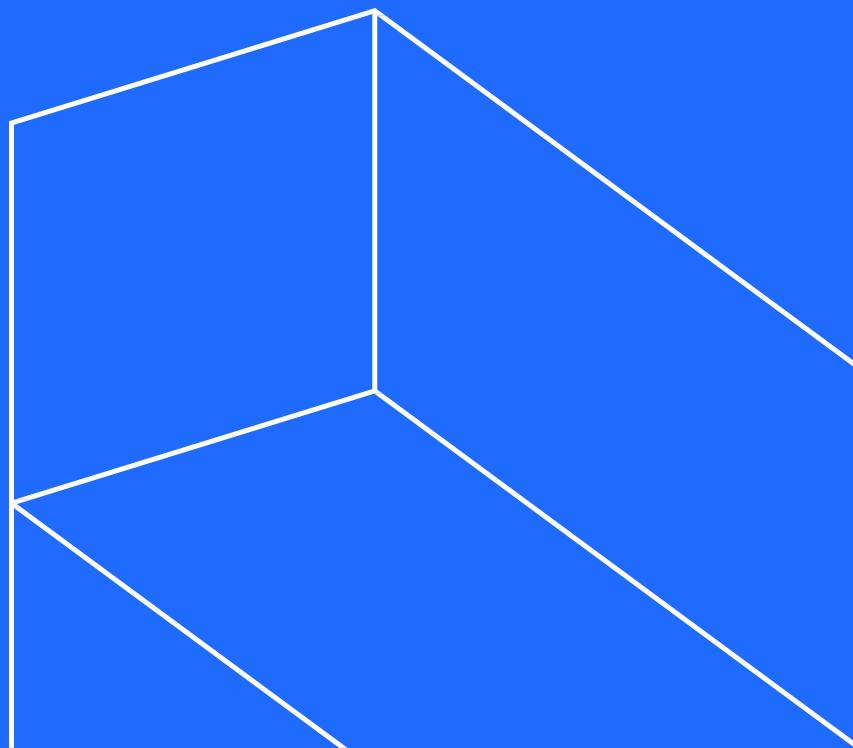


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.49	0.53	0.51	40
PAIDOFF	0.67	0.63	0.65	60
accuracy			0.59	100
macro avg	0.58	0.58	0.58	100
weighted avg	0.60	0.59	0.59	100

RESULTS

NAIVE BAYES
(SKLEARN)

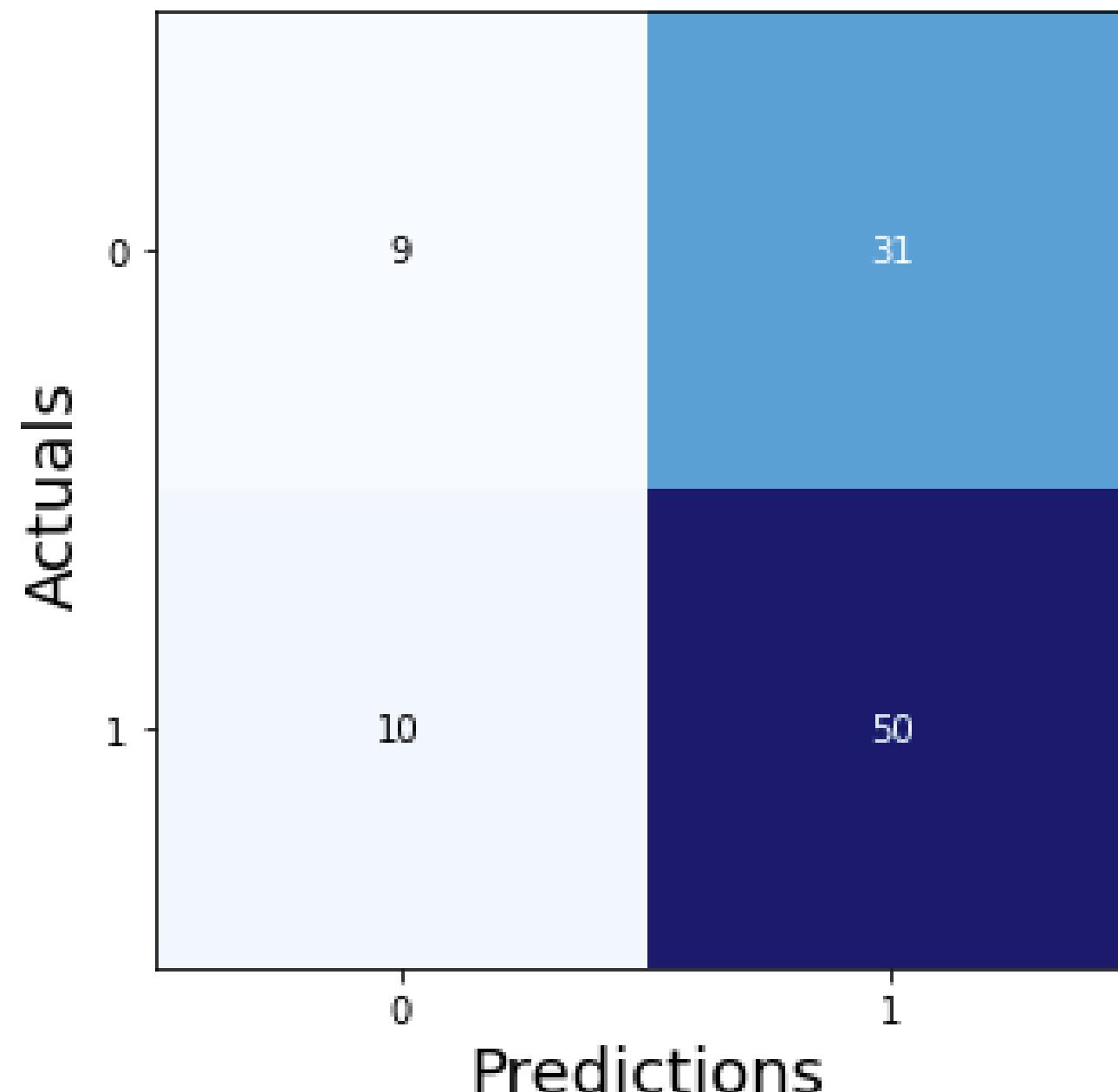
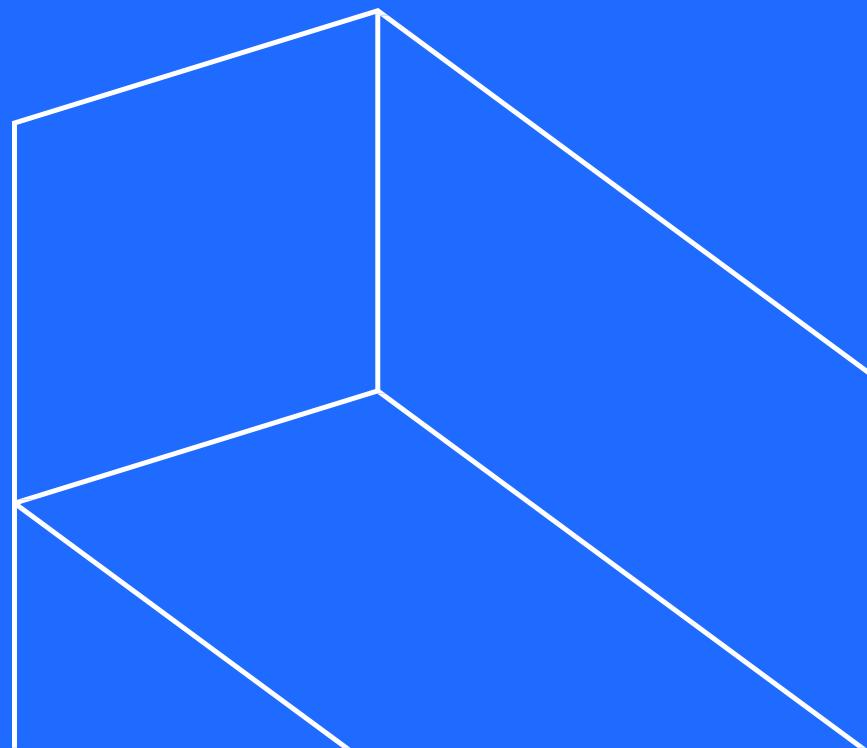


Classification Report

	precision	recall	f1-score	support
COLLECTION	0.41	0.65	0.50	40
PAIDOFF	0.62	0.38	0.47	60
accuracy			0.49	100
macro avg	0.52	0.52	0.49	100
weighted avg	0.54	0.49	0.49	100

RESULTS

LOGISTIC REGRESSION
(SKLEARN)



Classification Report

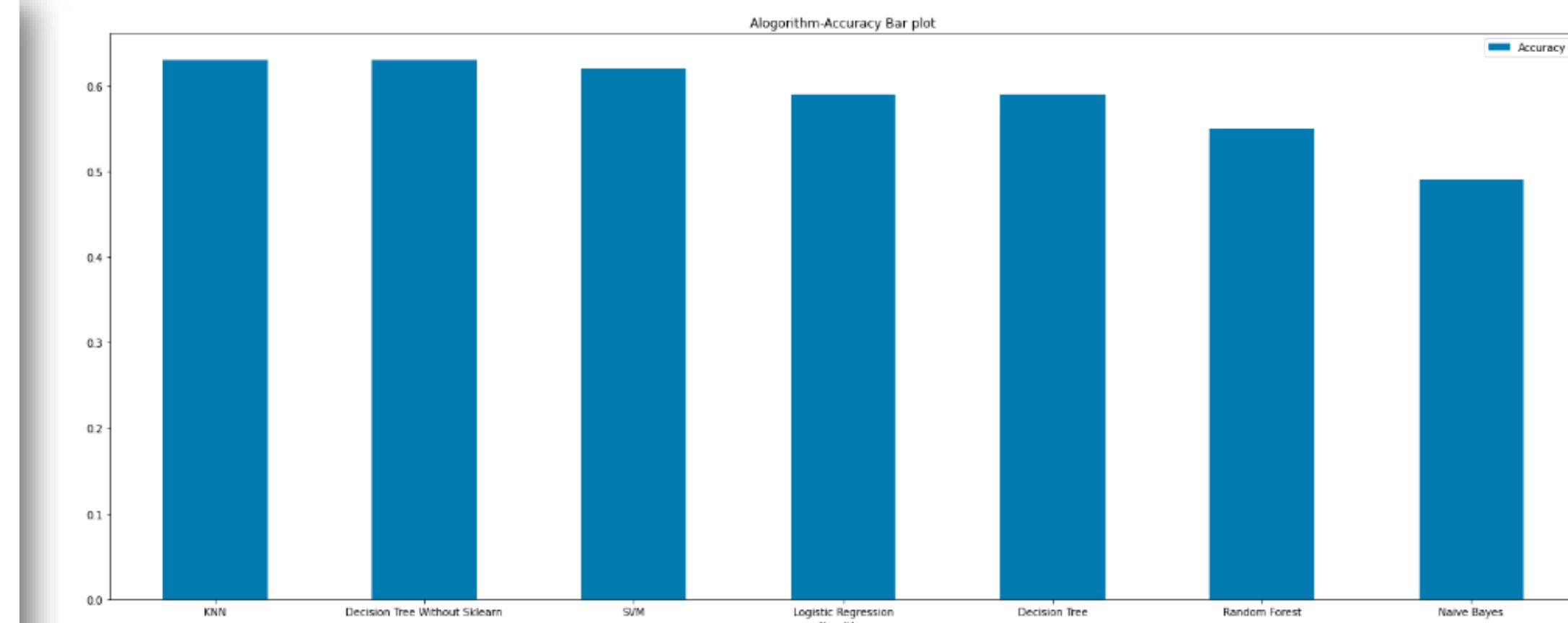
	precision	recall	f1-score	support
COLLECTION	0.47	0.23	0.31	40
PAIDOFF	0.62	0.83	0.71	60
accuracy			0.59	100
macro avg	0.55	0.53	0.51	100
weighted avg	0.56	0.59	0.55	100

CONCLUSION

Best accuracy obtained is 63% using the decision tree implemented from scratch and knn with k=7.

Out of 100 loan requests 41 were approved.

	Algorithm	Accuracy
0	KNN	0.63
1	Decision Tree Without Sklearn	0.63
2	SVM	0.62
3	Logistic Regression	0.59
4	Decision Tree	0.59
5	Random Forest	0.55
6	Naive Bayes	0.49



DATASET-3

Karthik Narayanan

AM.EN.U4CSE19327

This dataset consists of 11 columns and 500 rows.

Features

- Education
- Principal
- Age
- Term
- Gender

Target - Loan Status

- PAIDOFF
- COLLECTION
- COLLECTION_PAIDOFF

```
RangeIndex: 500 entries, 0 to 499
Data columns (total 11 columns):
 #   Column           Non-Null Count   Dtype  
 ---  -- 
 0   Loan_ID          500 non-null     object  
 1   loan_status       500 non-null     object  
 2   Principal         500 non-null     int64  
 3   terms             500 non-null     int64  
 4   effective_date   500 non-null     datetime64[ns]
 5   due_date          500 non-null     datetime64[ns]
 6   paid_off_time    409 non-null    object  
 7   past_due_days    186 non-null    float64 
 8   age               500 non-null     int64  
 9   education         500 non-null     object  
 10  Gender            500 non-null     object  
 dtypes: datetime64[ns](2), float64(1), int64(3), object(5)
 memory usage: 43.1+ KB
```

	Loan_ID	loan_status	Principal	terms	effective_date	due_date	paid_off_time	past_due_days	age	education	Gender	
0	xqd20161085	COLLECTION	1000	30	2016-09-11	2016-10-10			59.0	29	High School or Below	male
1	xqd20160435	PAIDOFF	800	15	2016-09-12	2016-09-26	9/25/2016 13:29		NaN	24	college	male
2	xqd20160944	COLLECTION	1000	15	2016-09-11	2016-09-25			74.0	23	High School or Below	male
3	xqd20160085	PAIDOFF	1000	30	2016-09-11	2016-10-10	10/10/2016 13:00		NaN	37	High School or Below	female
4	xqd20160134	PAIDOFF	1000	30	2016-09-11	2016-10-10	9/30/2016 7:12		NaN	33	college	female

PAIDOFF 314
COLLECTION_PAIDOFF 95
COLLECTION 91
Name: loan_status, dtype: int64

Links:

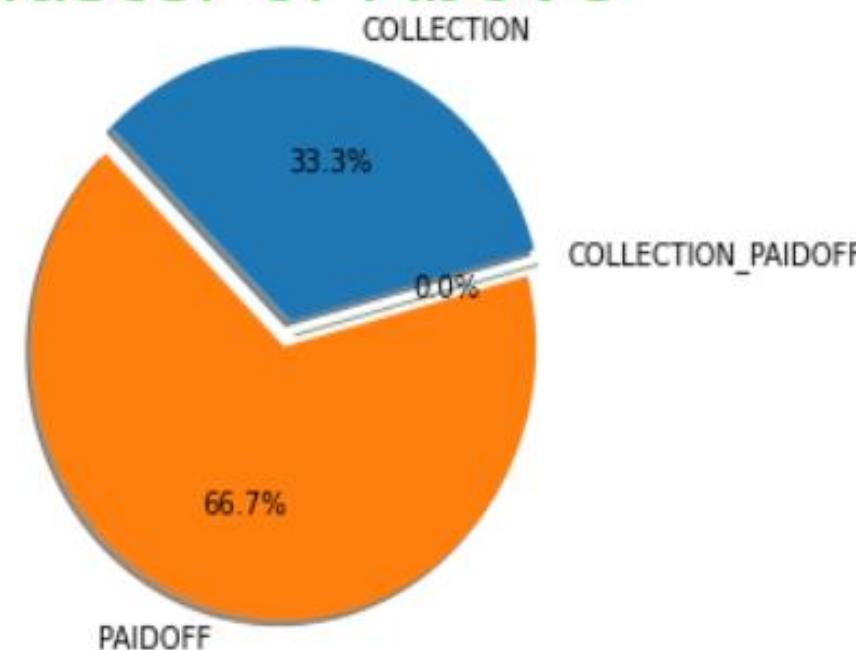
[Dataset](#)

[Files](#)

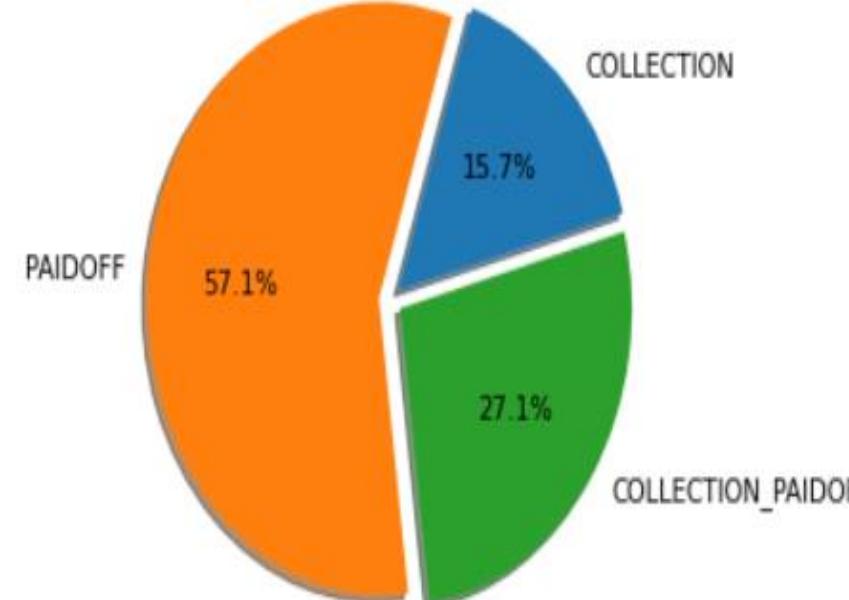
Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Of
Different Education Groups

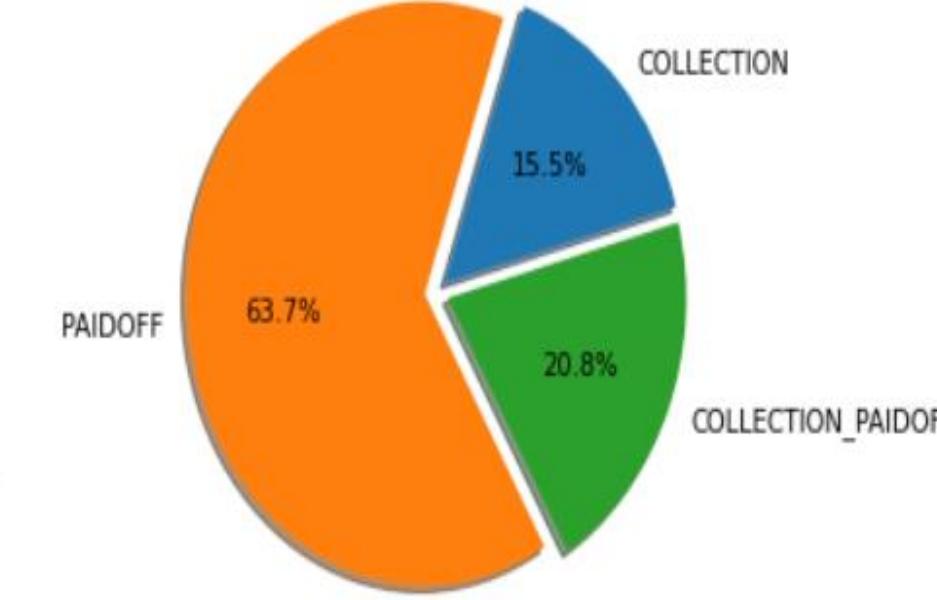
Master or Above



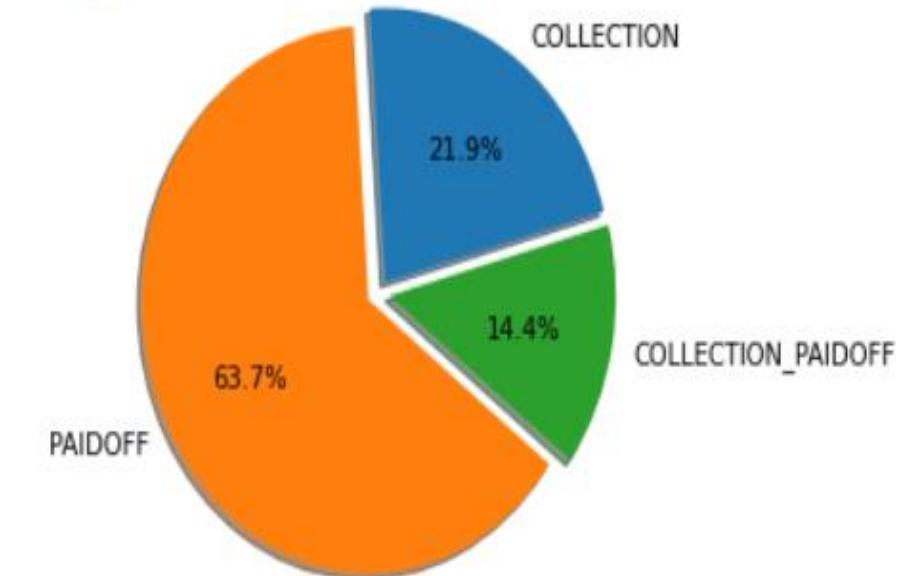
Bechelor



college

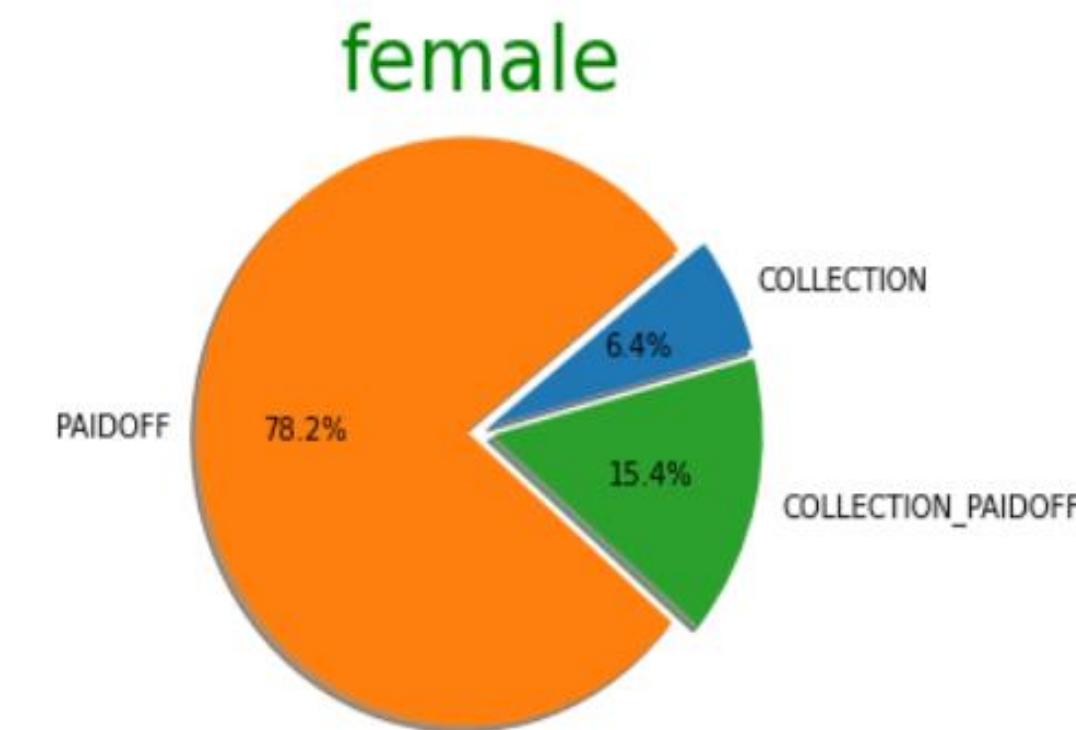
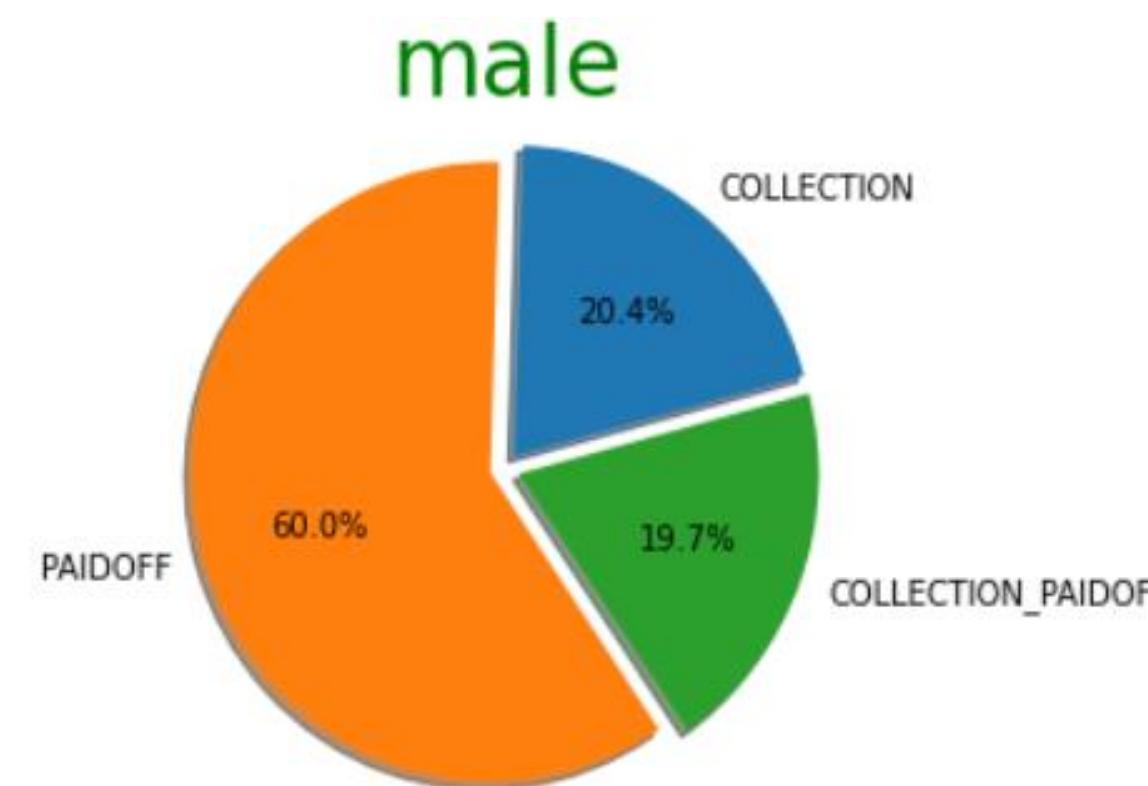


High School or Below



Data Visualization And Interpretation

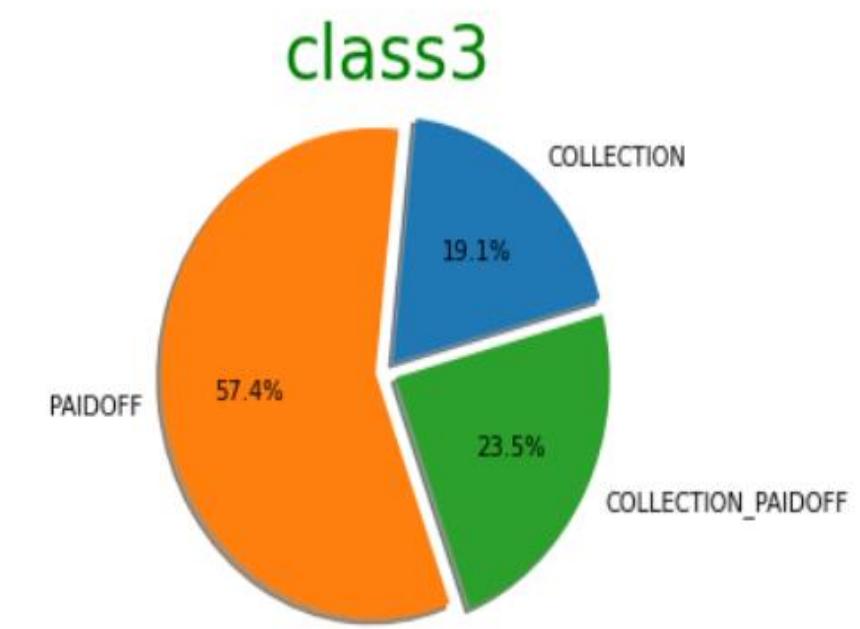
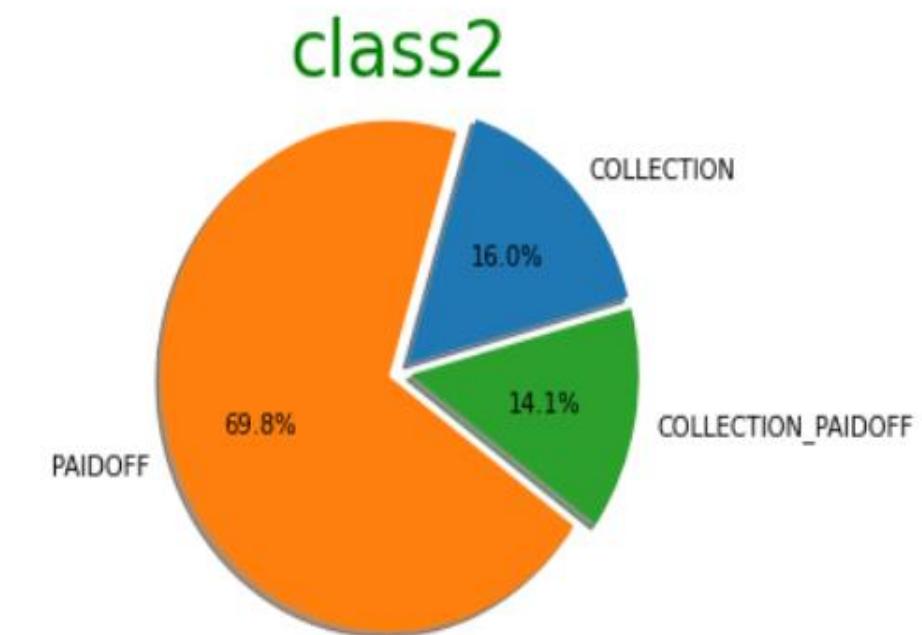
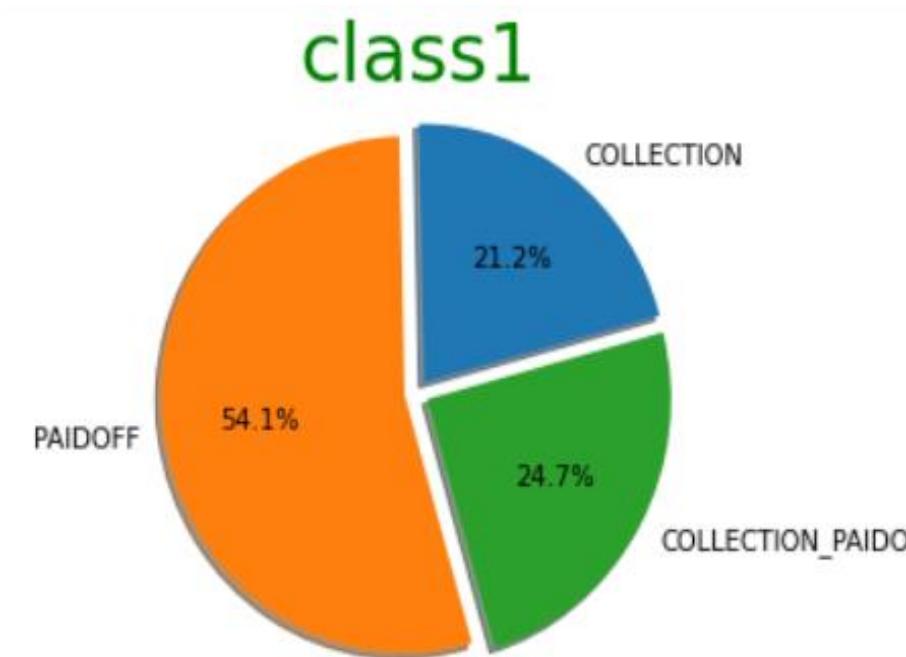
Pie Chart To Represent The Loan Status Based On Gender



Data Visualization And Interpretation

Pie Chart To Represent The Loan Status Based On Age

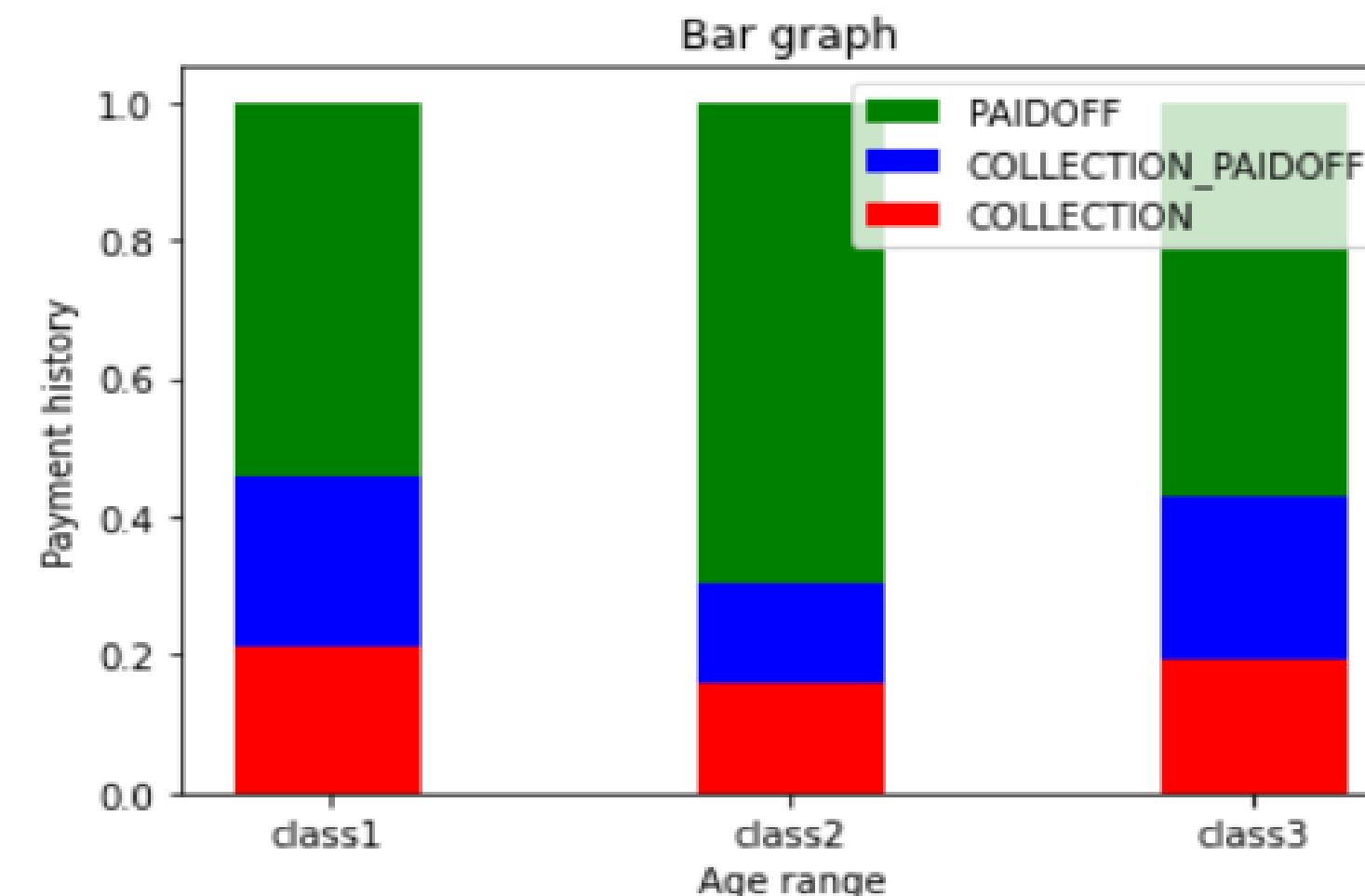
class 1: [18, 28]
class 2: [28, 38]
class 3: [38, 51]



Data Visualization And Interpretation

Bar Chart Between Loan Status And Age

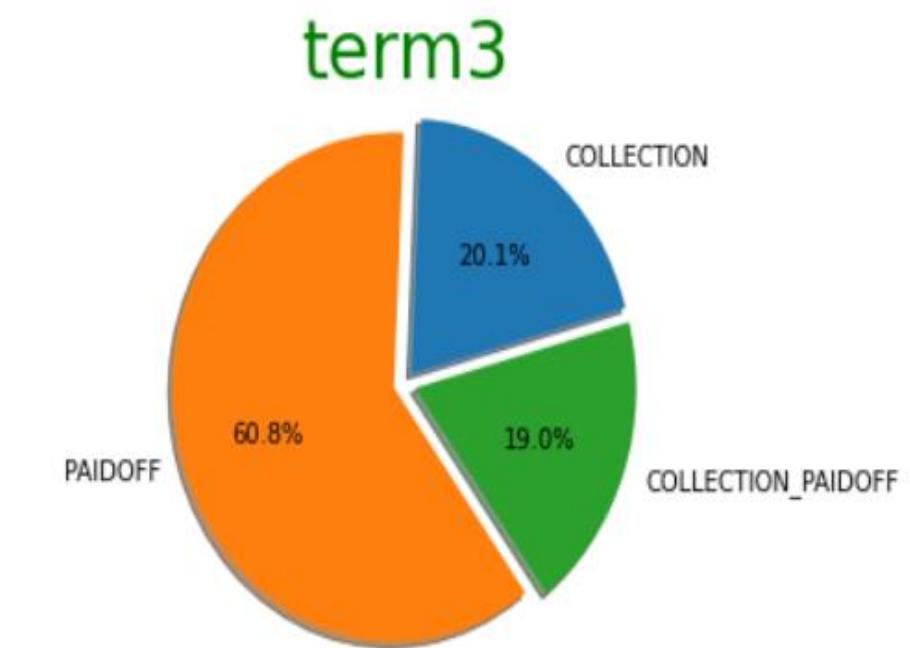
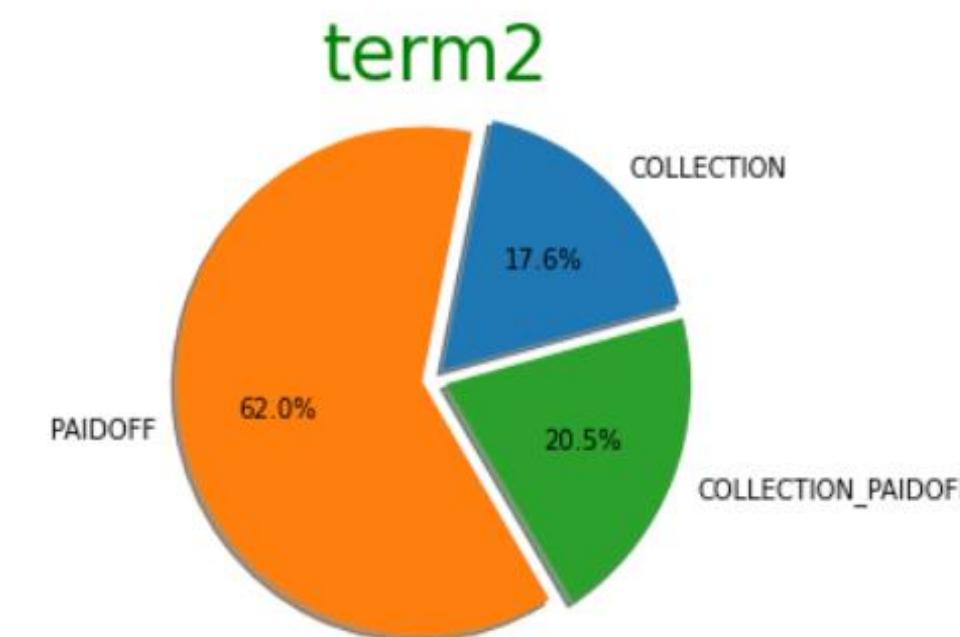
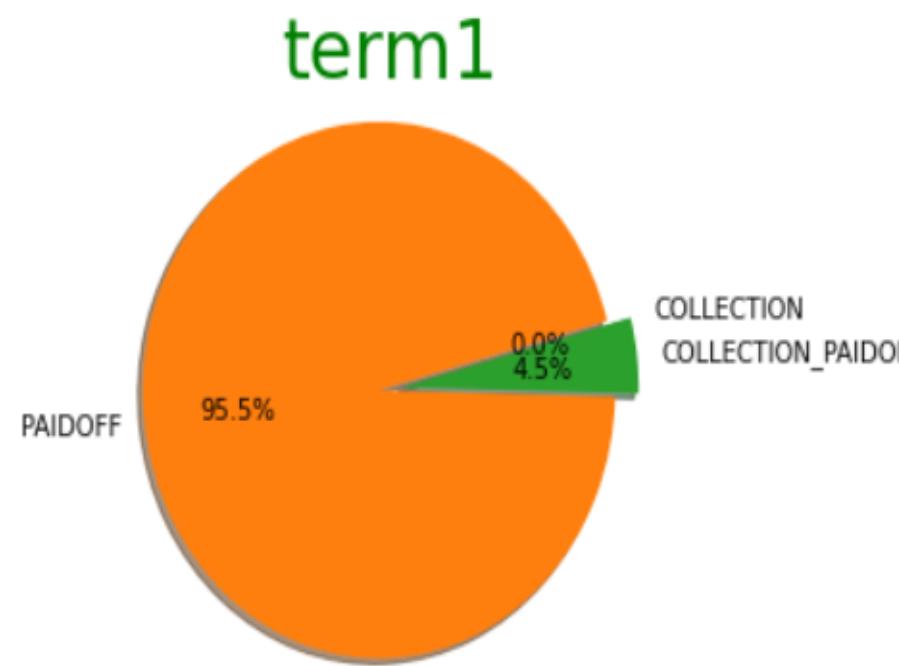
Class 1: [18, 28]
Class 2: [28, 38]
Class 3: [38, 51]



Data Visualization And Interpretation

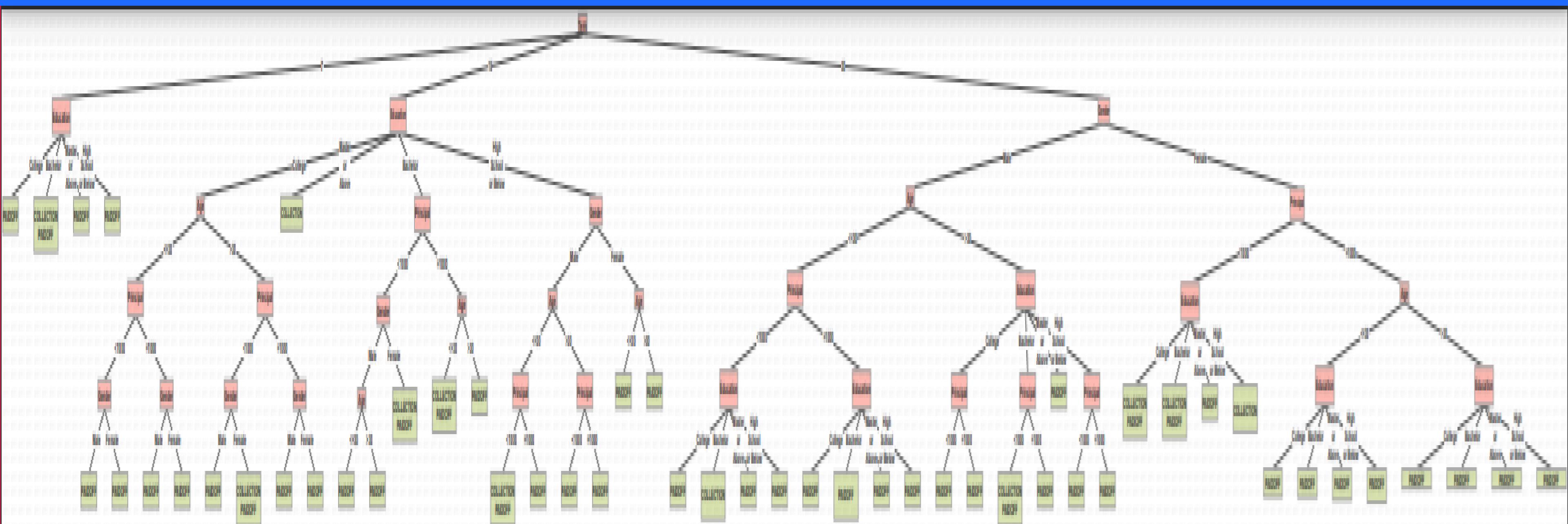
Pie Chart To Represent The Loan Status Based On Term

class 1: [7, 14]
class 2: [14, 21]
class 3: [21, 31]



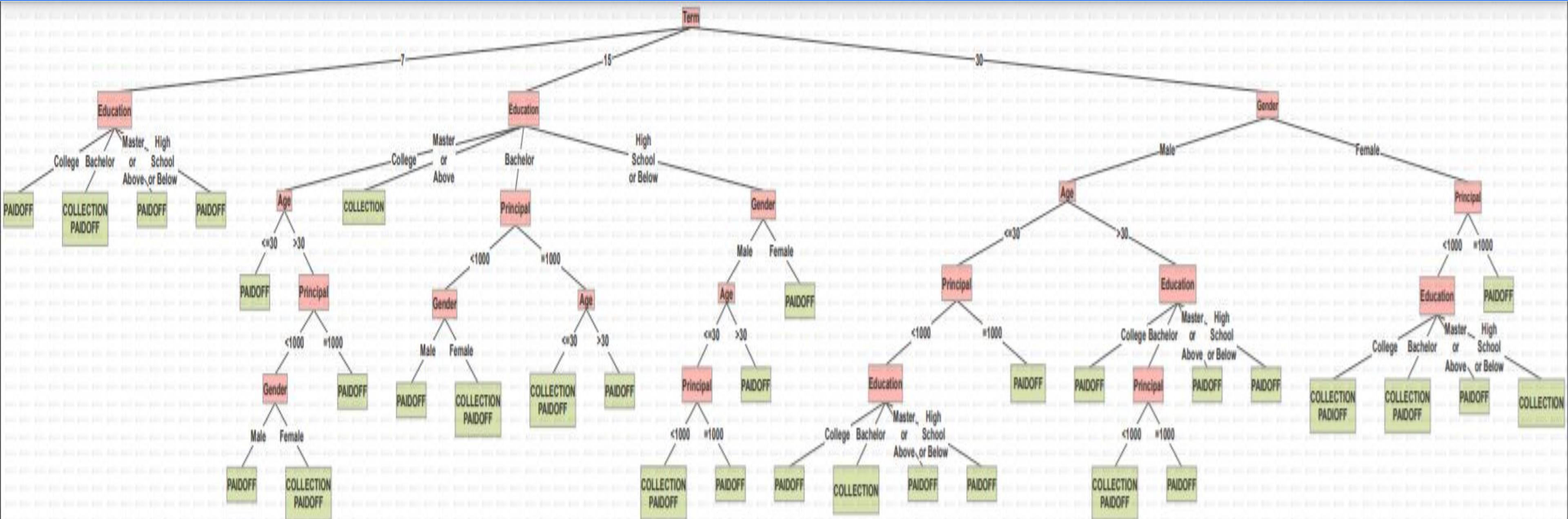
Data Visualization And Interpretation

Decision Tree



Data Visualization And Interpretation

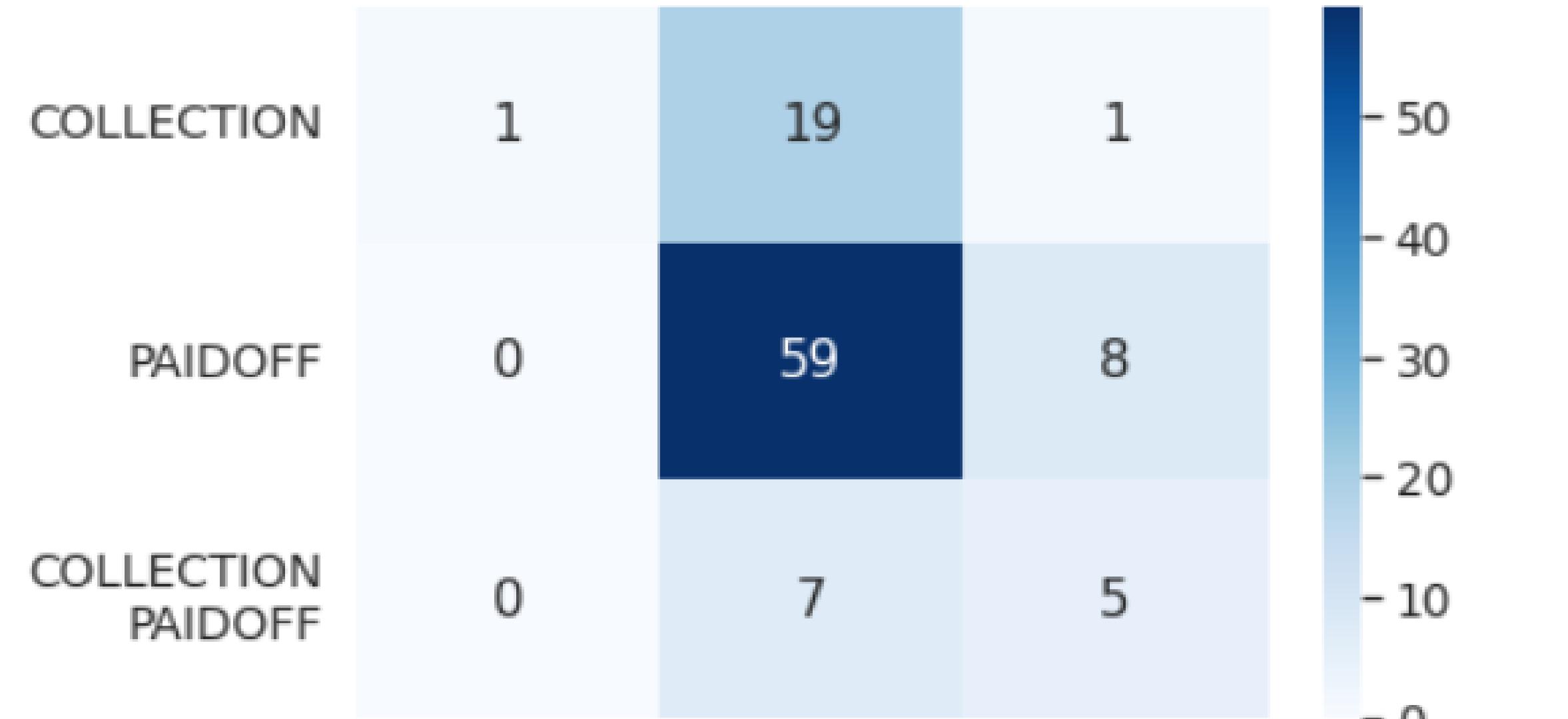
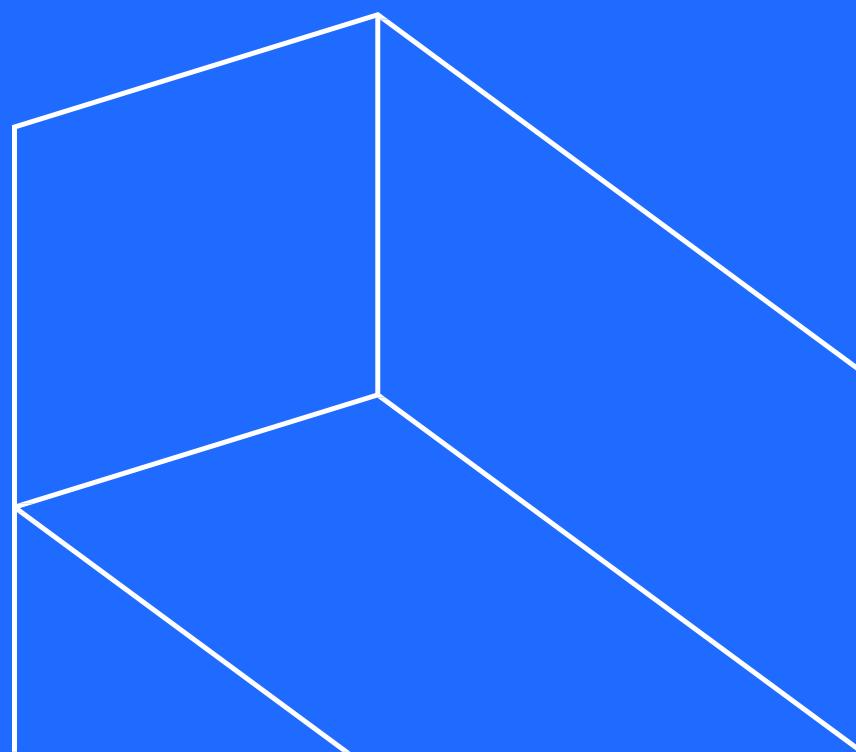
Pruned Decision Tree



RESULTS

DECISION TREE

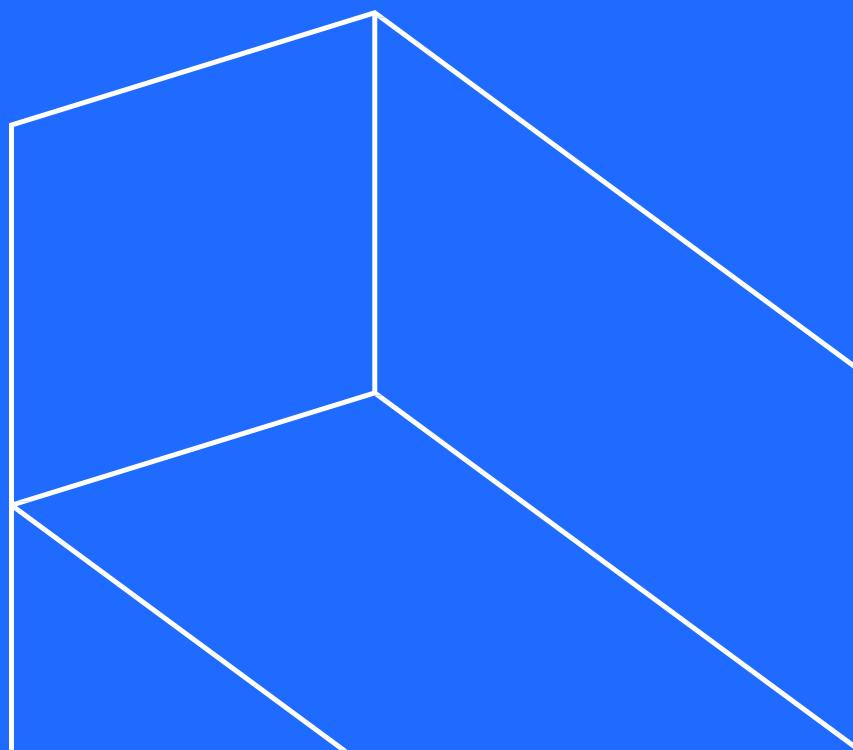
(WITHOUT SKLEARN)



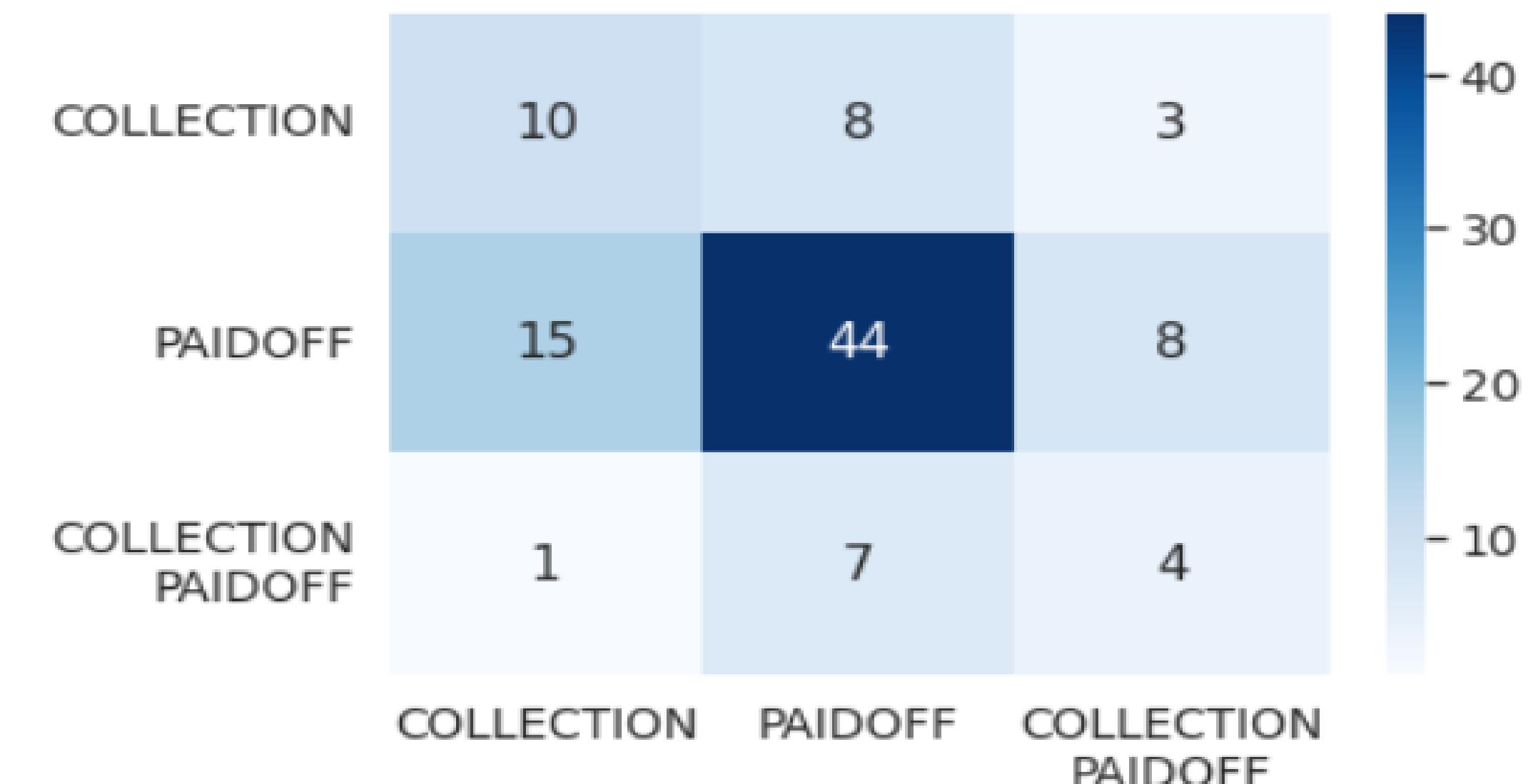
	precision	recall	f1-score	support
COLLECTION_PAIDOFF	1.00	0.05	0.09	21
COLLECTION	0.36	0.42	0.38	12
PAIDOFF	0.69	0.88	0.78	67
accuracy			0.65	100
macro avg	0.68	0.45	0.42	100
weighted avg	0.72	0.65	0.59	100

RESULTS

DECISION TREE
(SKLEARN)



Model - Decision Tree

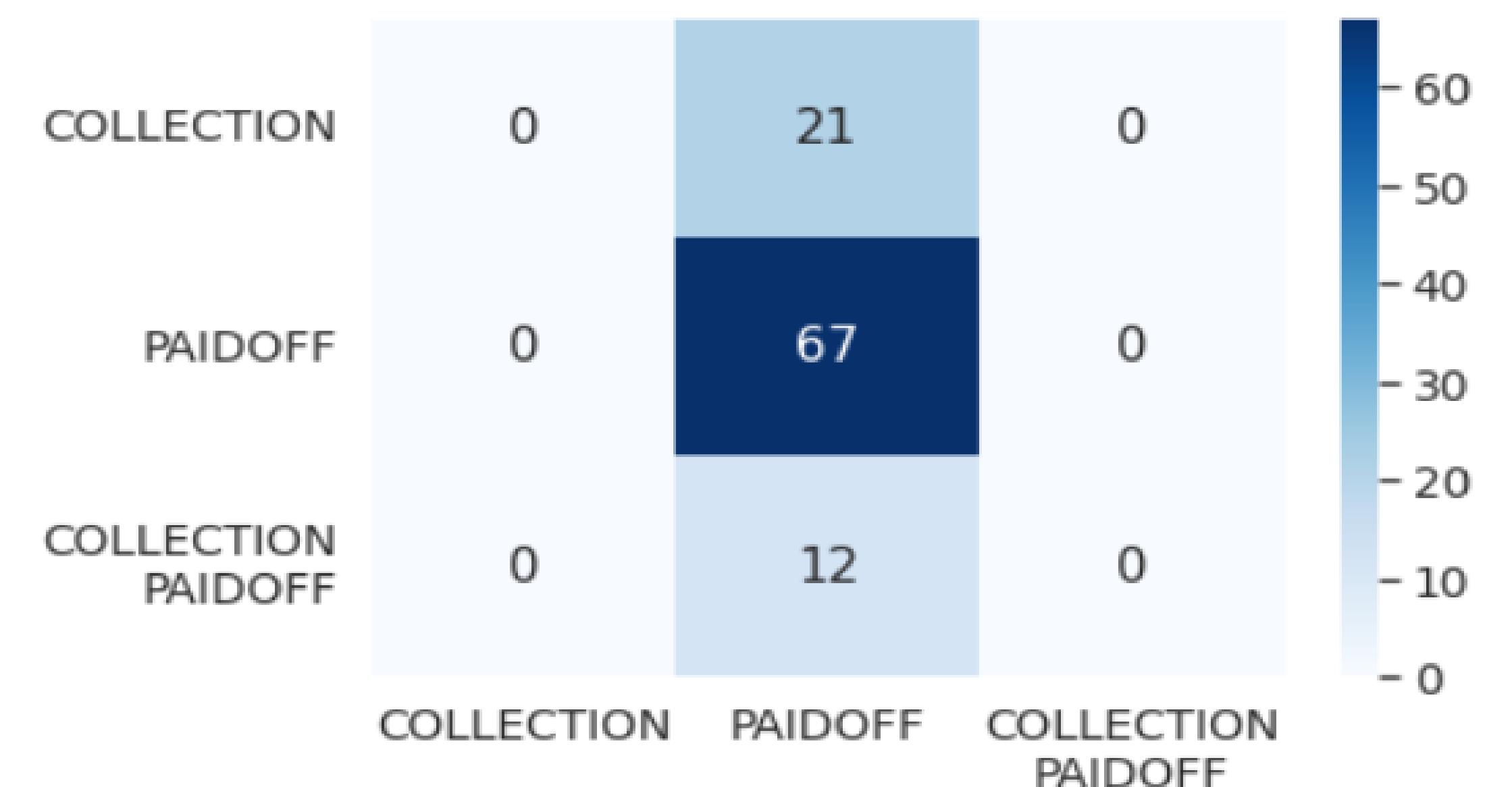
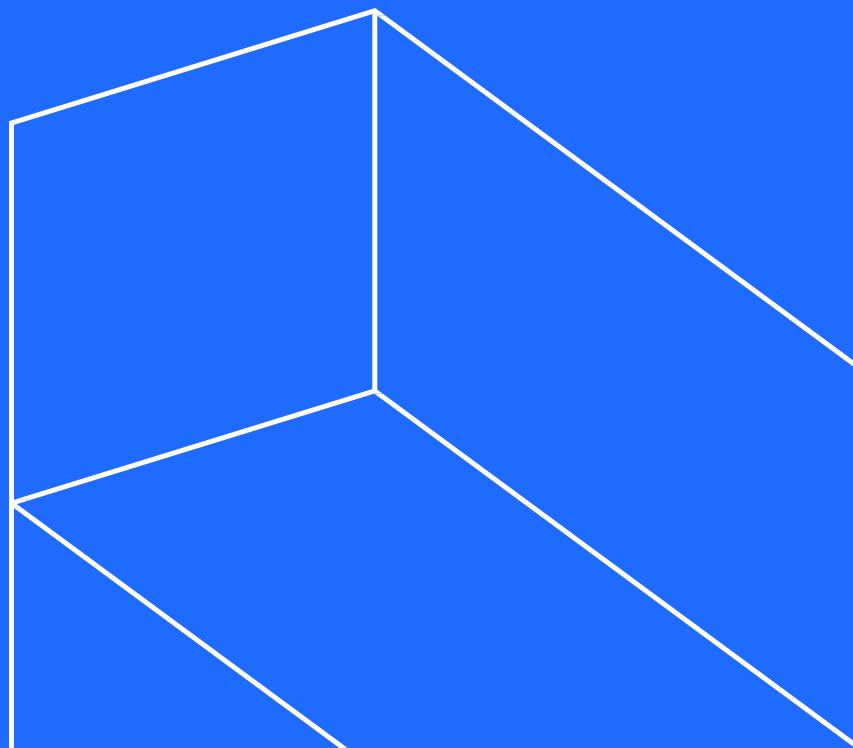


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.38	0.48	0.43	21
COLLECTION	0.75	0.66	0.70	67
PAIDOFF	0.27	0.33	0.30	12
accuracy			0.58	100
macro avg	0.47	0.49	0.47	100
weighted avg	0.61	0.58	0.59	100

RESULTS

SVM (SKLEARN)

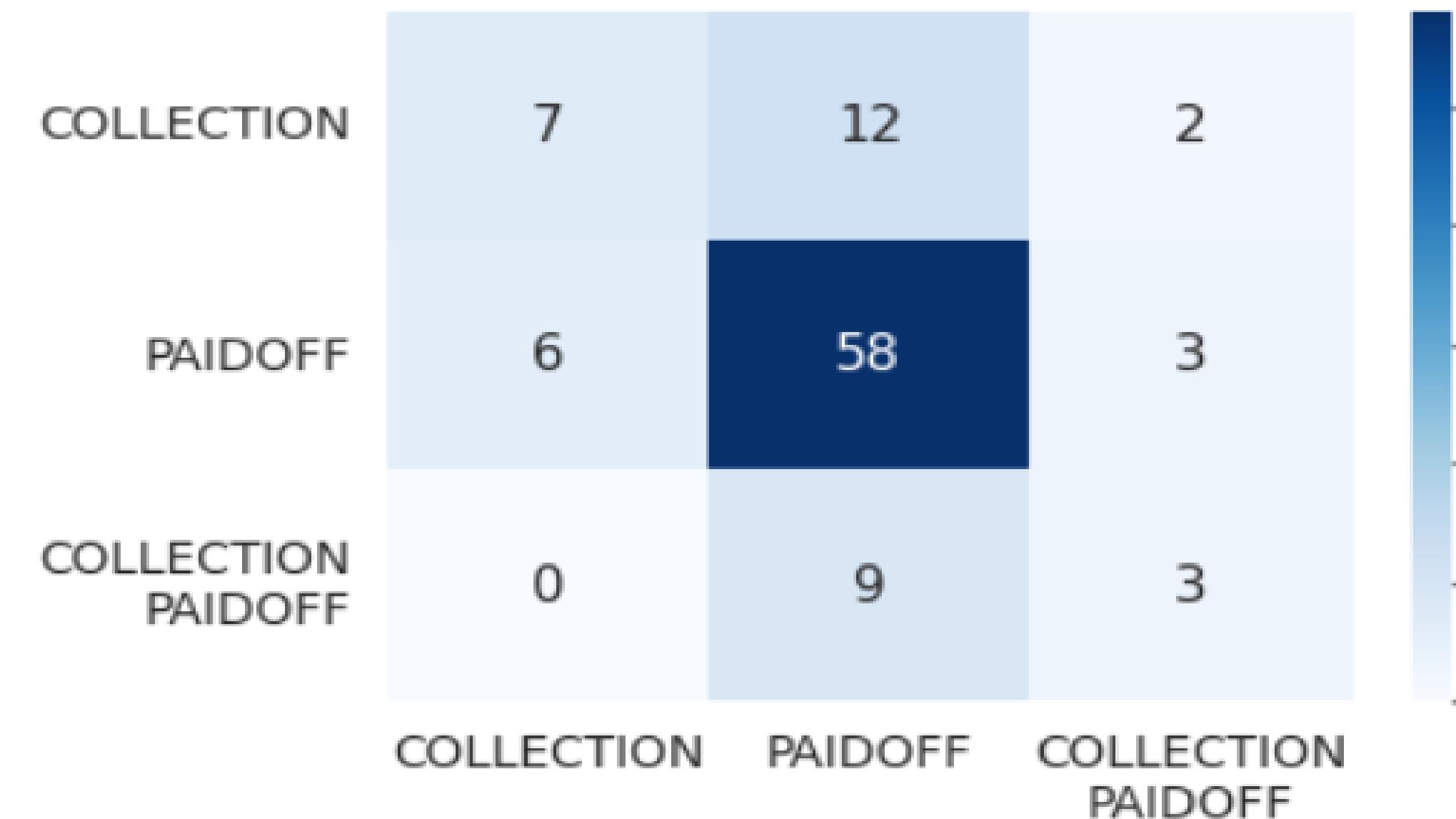
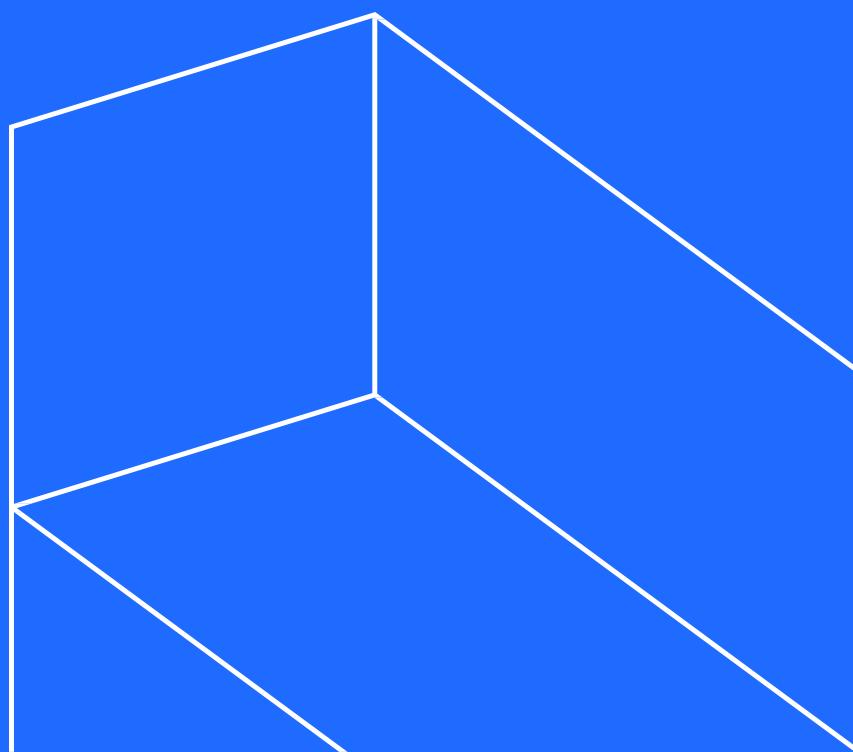


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	21
COLLECTION	0.67	1.00	0.80	67
PAIDOFF	0.00	0.00	0.00	12
accuracy			0.67	100
macro avg	0.22	0.33	0.27	100
weighted avg	0.45	0.67	0.54	100

RESULTS

KNN (SKLEARN)

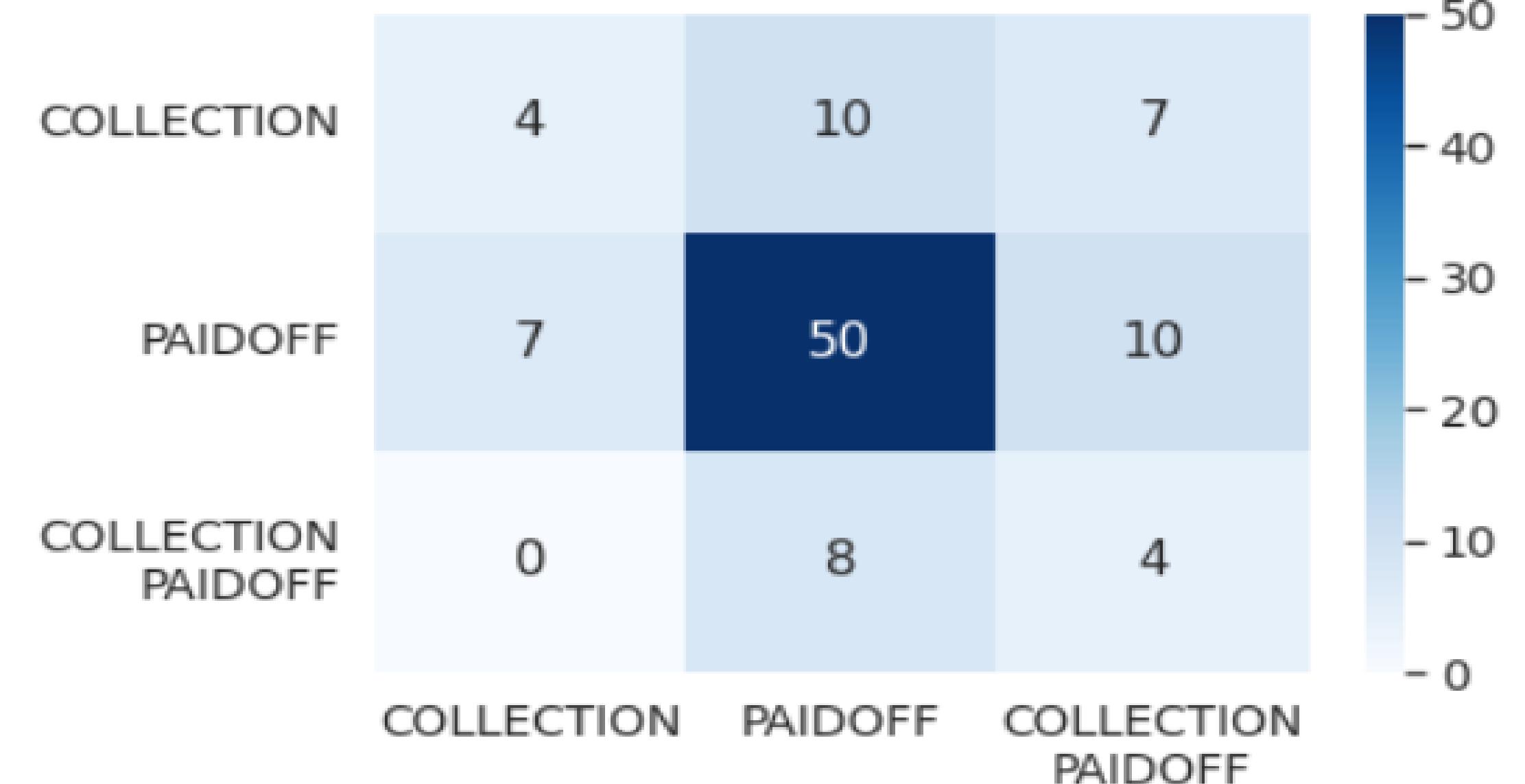
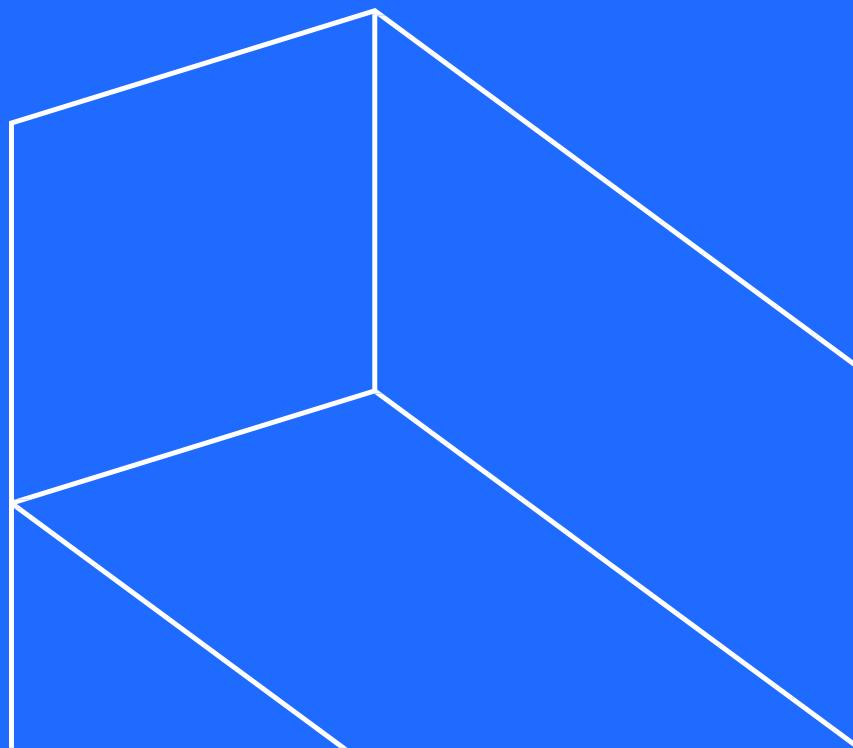


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.54	0.33	0.41	21
COLLECTION	0.73	0.87	0.79	67
PAIDOFF	0.38	0.25	0.30	12
accuracy			0.68	100
macro avg	0.55	0.48	0.50	100
weighted avg	0.65	0.68	0.65	100

RESULTS

R A N D O M F O R E S T
(S K L E A R N)

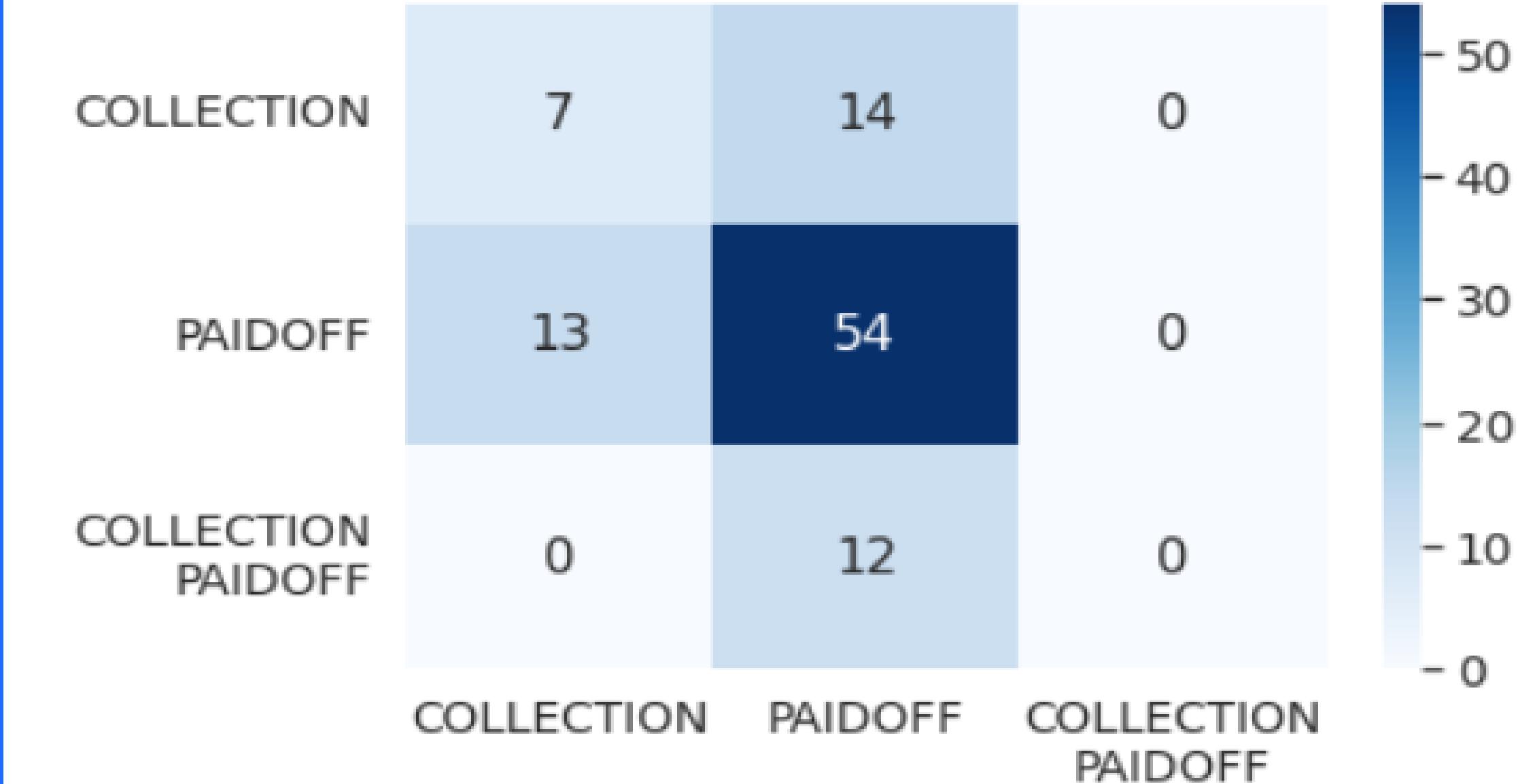
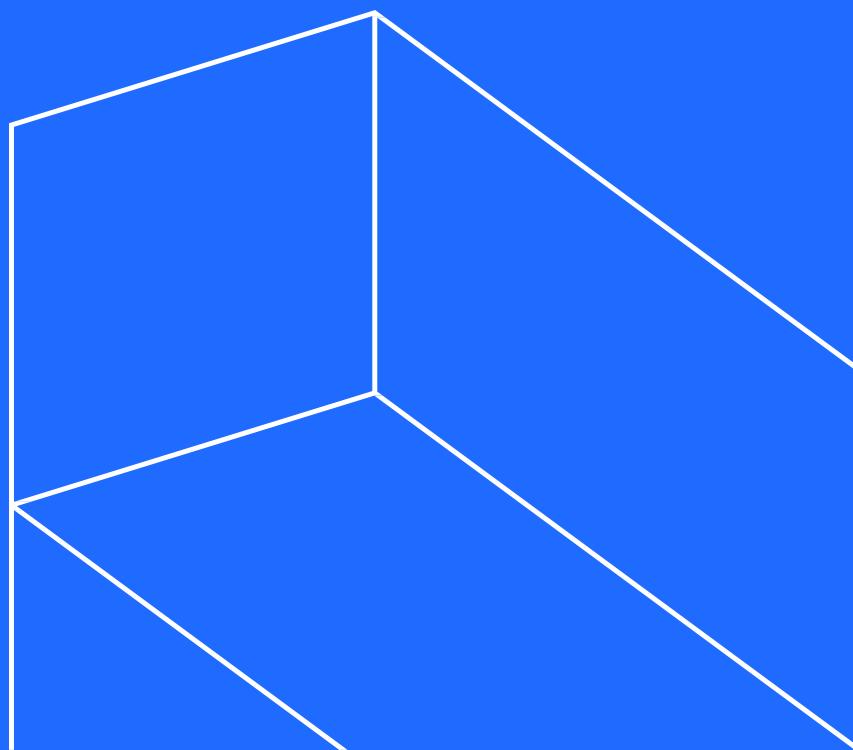


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.36	0.19	0.25	21
COLLECTION	0.74	0.75	0.74	67
PAIDOFF	0.19	0.33	0.24	12
accuracy			0.58	100
macro avg	0.43	0.42	0.41	100
weighted avg	0.59	0.58	0.58	100

RESULTS

NAIVE BAYES
(SKLEARN)

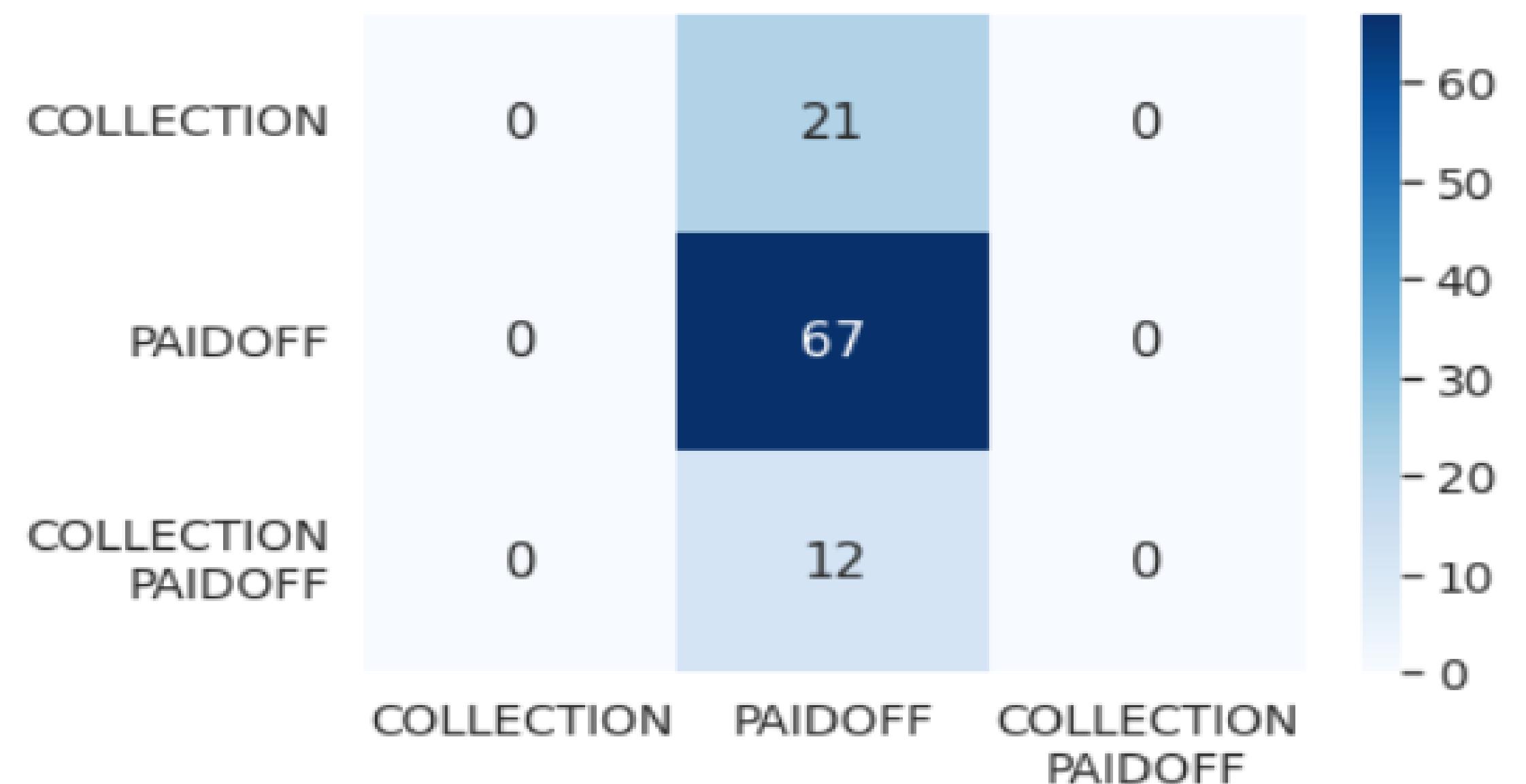
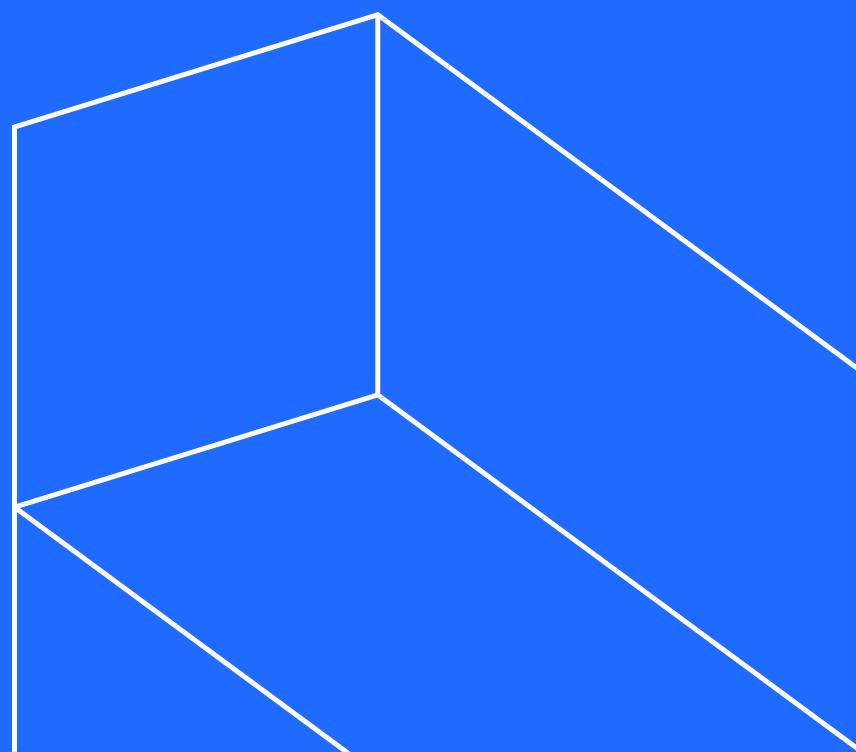


Classification Report

	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.35	0.33	0.34	21
COLLECTION	0.68	0.81	0.73	67
PAIDOFF	0.00	0.00	0.00	12
accuracy			0.61	100
macro avg	0.34	0.38	0.36	100
weighted avg	0.53	0.61	0.56	100

RESULTS

LOGISTIC REGRESSION
(SKLEARN)



Classification Report

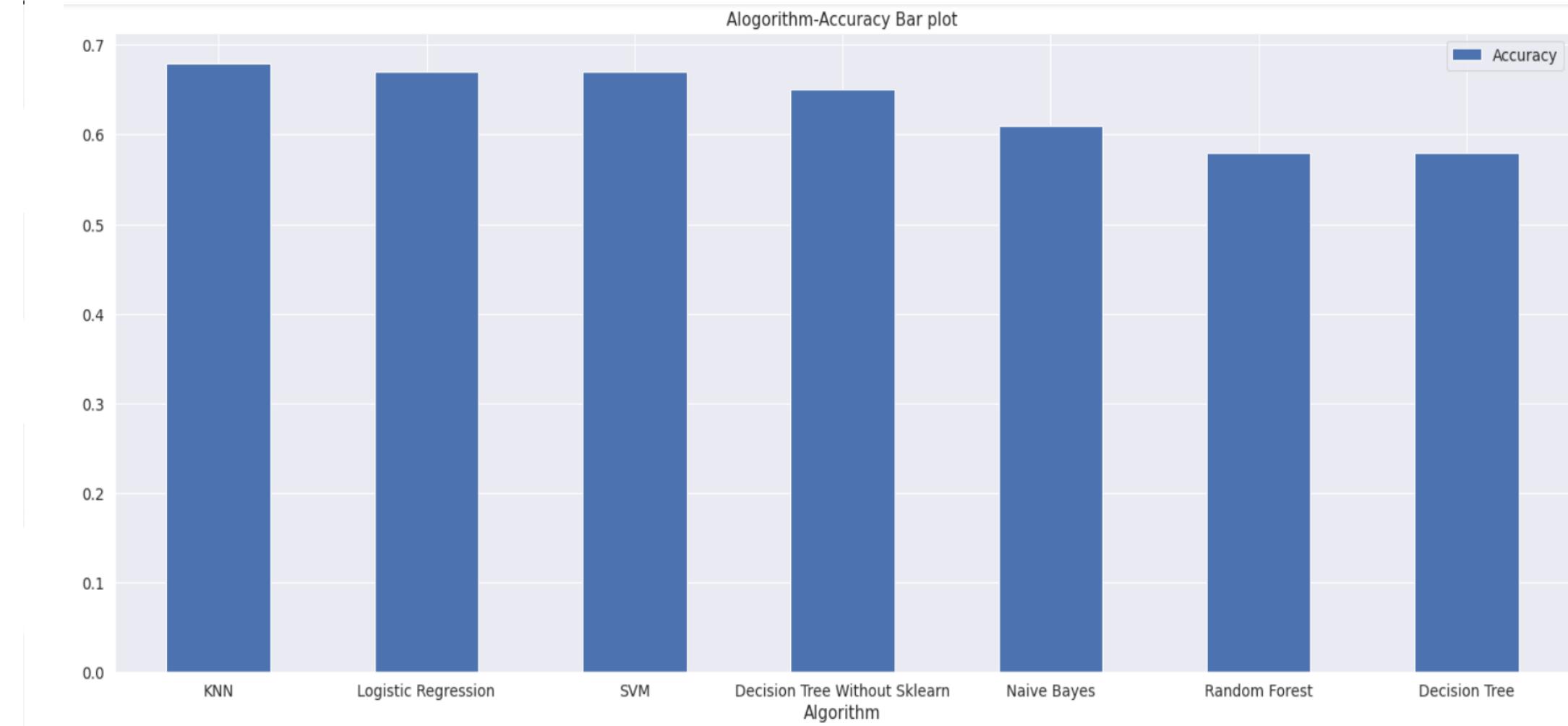
	precision	recall	f1-score	support
COLLECTION_PAIDOFF	0.00	0.00	0.00	21
COLLECTION	0.67	1.00	0.80	67
PAIDOFF	0.00	0.00	0.00	12
accuracy			0.67	100
macro avg	0.22	0.33	0.27	100
weighted avg	0.45	0.67	0.54	100

CONCLUSION

Best accuracy obtained is 68% using KNN implemented from sklearn.

Out of 100 loan requests 59 were approved.

	Algorithm	Accuracy
0	KNN	0.68
1	Logistic Regression	0.67
2	SVM	0.67
3	Decision Tree Without Sklearn	0.65
4	Naive Bayes	0.61
5	Random Forest	0.58
6	Decision Tree	0.58





Thank you!!