VISA FOR LISA

Loan Acceptance Prediction for GALAXY BANK

Introduction

Importance of Data Science for insightfull decision making

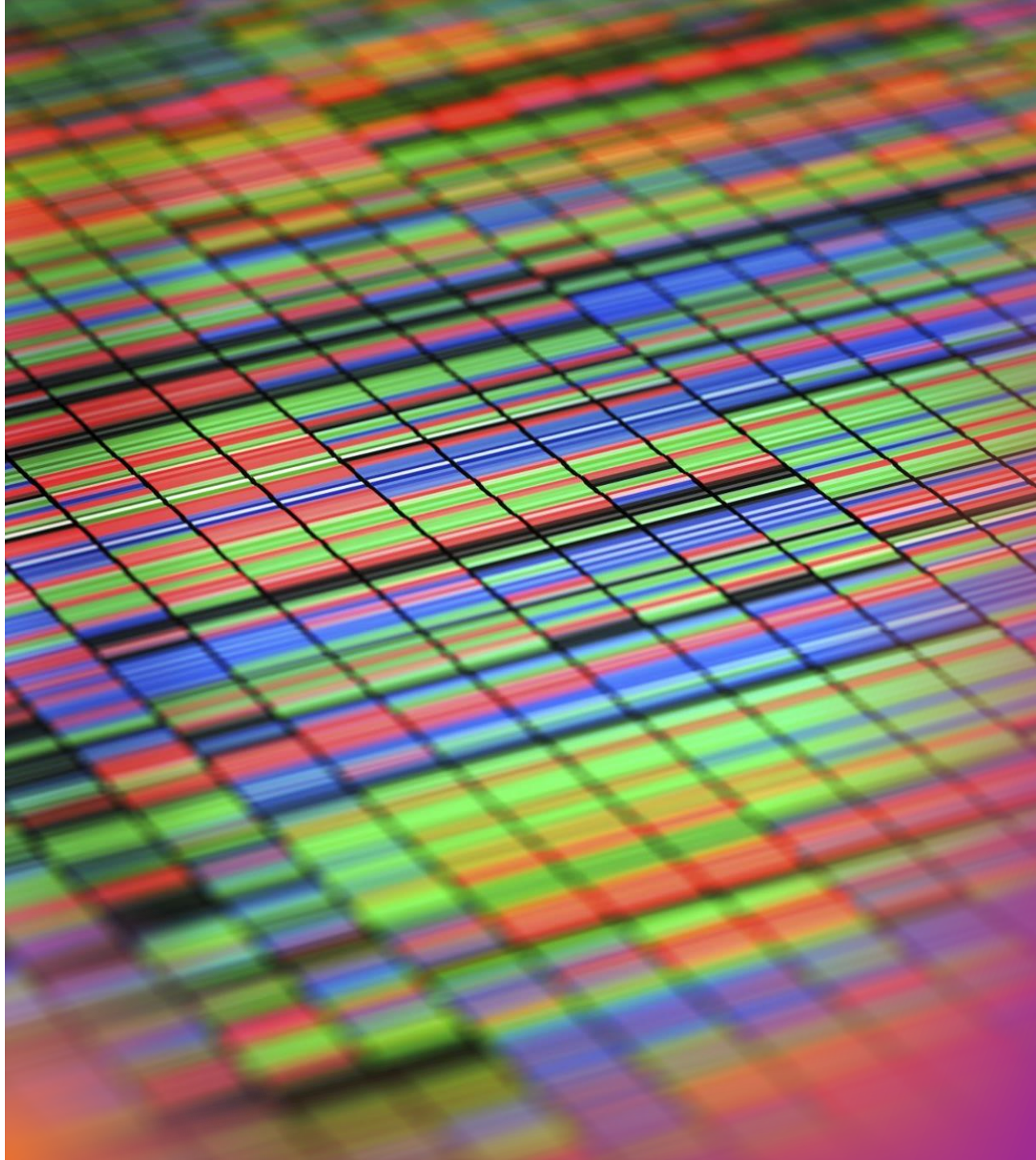
Data collection and exploration

The data we used in this project is provided by Glaxy Bank.

Here is a brief explanation of each variable in the dataset:
1.ID: A unique identifier for each record.
2.Age: The age of the individual.
3.Experience: The number of years of professional experience the individual has.
4.Income: The annual income of the individual (in thousands of dollars).
5.ZIP Code: The ZIP Code where the individual resides.
6.Family: The size of the individual's family (number of family members).
7.CCAvg: The average monthly spending on the credit card (in thousands of dollars).
8.Education: The education level of the individual, typically encoded as a numerical value.
9.Mortgage: The value of the mortgage (in thousands of dollars).
10.Personal Loan: A binary variable indicating whether the individual has taken a personal loan (1 if yes, 0 if no).
11.Securities Account: A binary variable indicating whether the individual has a securities account (1 if yes, 0 if no).
12.CD Account: A binary variable indicating whether the individual has a certificate of deposit (CD) account (1 if yes, 0 if no).
13.Online: A binary variable indicating whether the individual uses online banking (1 if yes, 0 if no).
14.CreditCard: A binary variable indicating whether the individual uses a credit card issued by the bank (1 if yes, 0 if no).

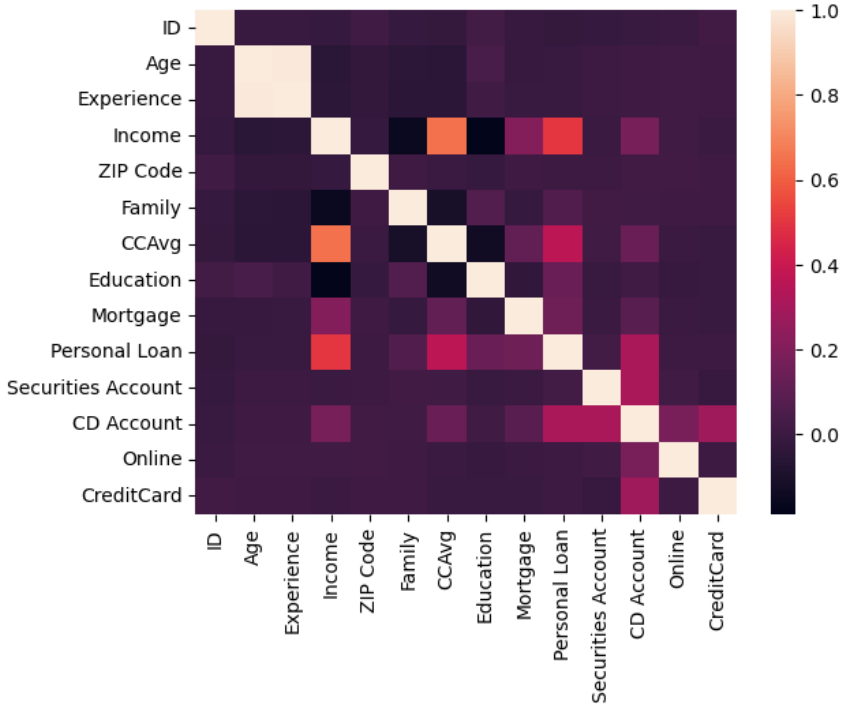
	ID	Age	Experience	Income	ZIP Code	Family	CCAvg	Education	Mortgage	Personal Loan	Securities Account	CD Account	Online	CreditCard
0	1	25	1	49	91107	4	1.6	1	0	0	1	0	0	0
1	2	45	19	34	90089	3	1.5	1	0	0	1	0	0	0
2	3	39	15	11	94720	1	1.0	1	0	0	0	0	0	0
3	4	35	9	100	94112	1	2.7	2	0	0	0	0	0	0
4	5	35	8	45	91330	4	1.0	2	0	0	0	0	0	1

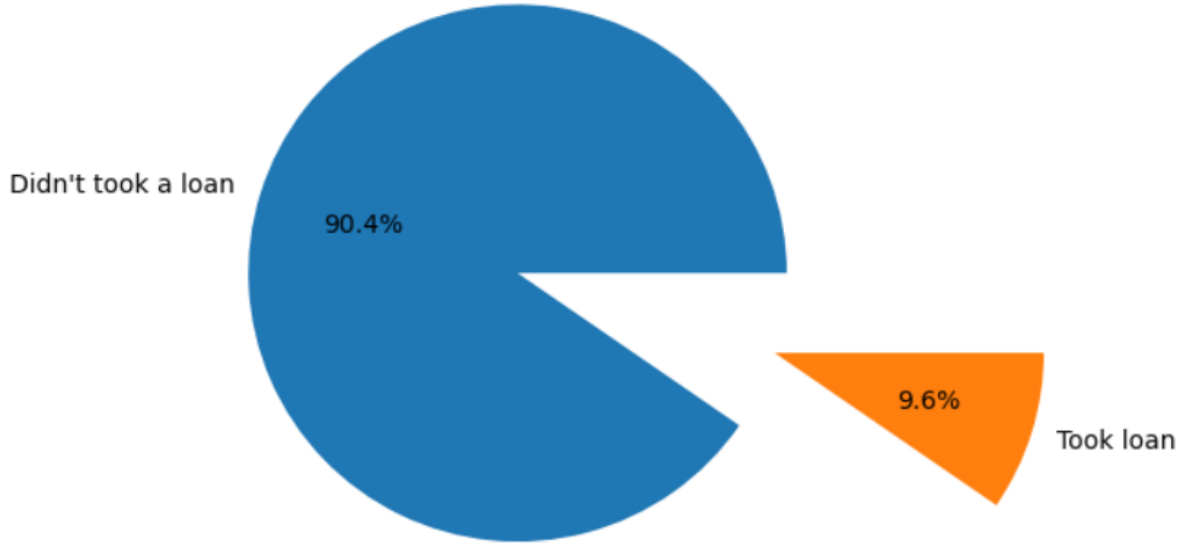
Data Visualization



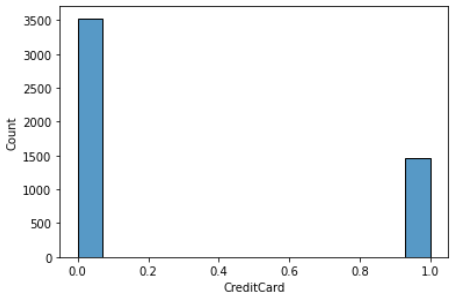
Target variable distribution and Correlation Heatmap

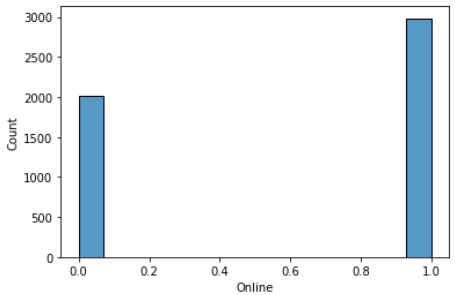
: 'The coorelation matrix informe us that Personal Loan variable I most cooralated with Income and CCAvg'

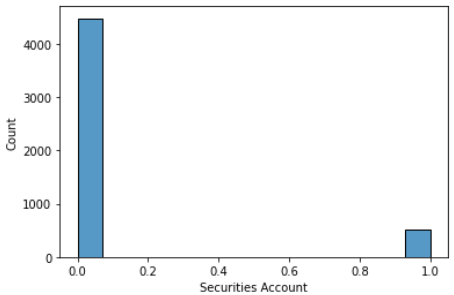


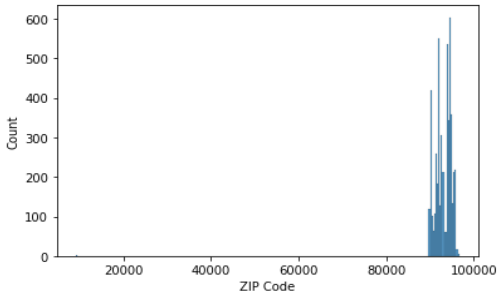


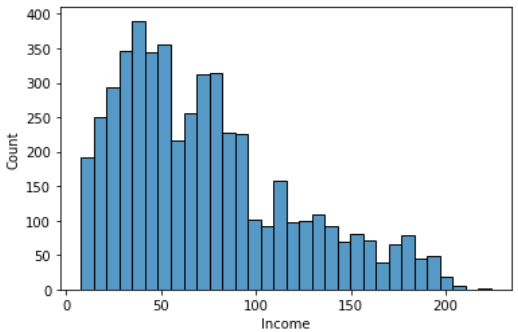
Histogram distribution of variable

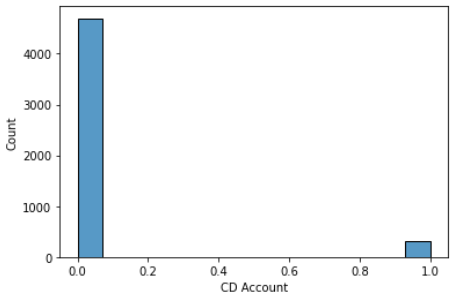


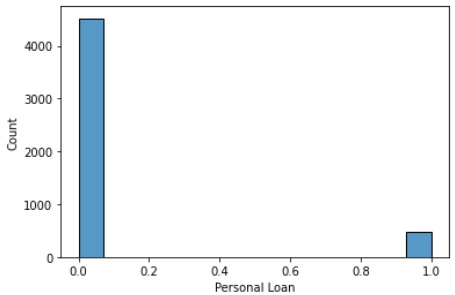


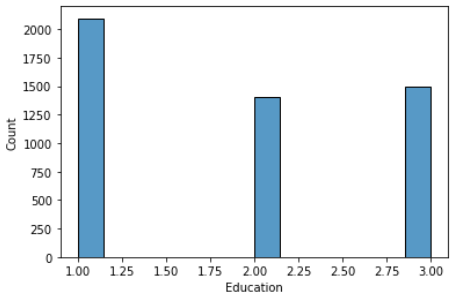


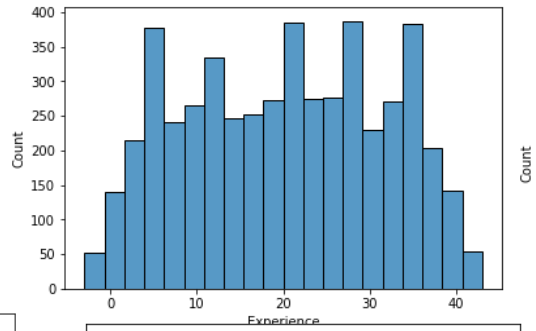


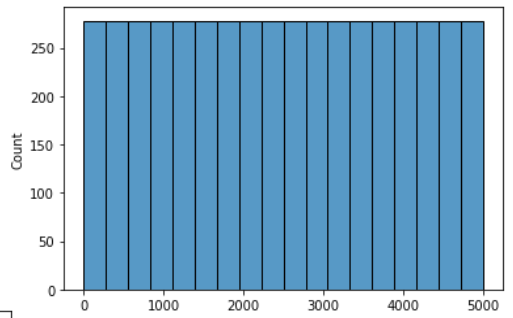


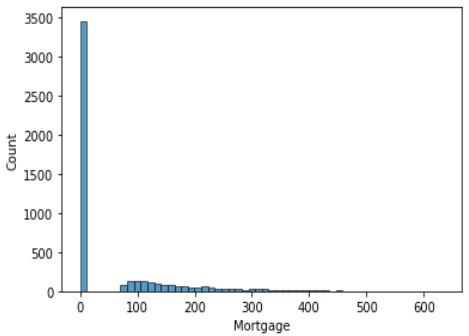


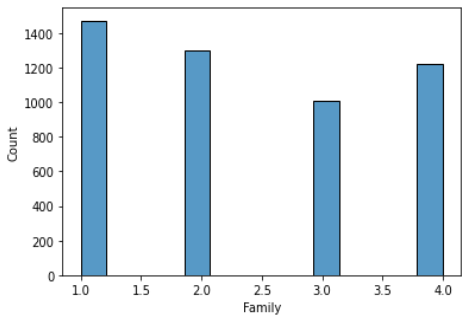


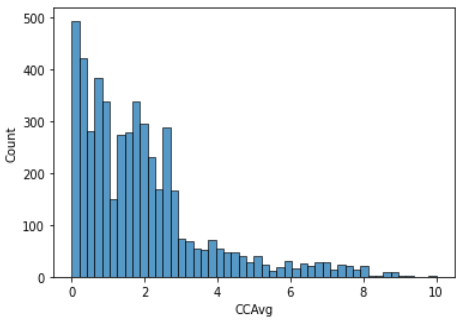


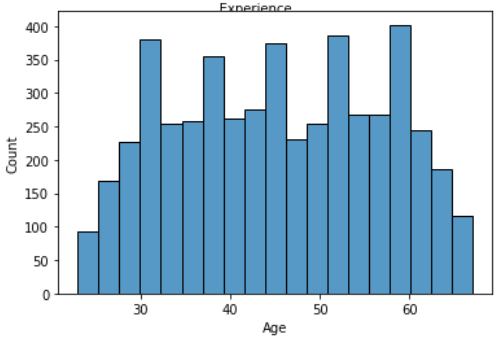




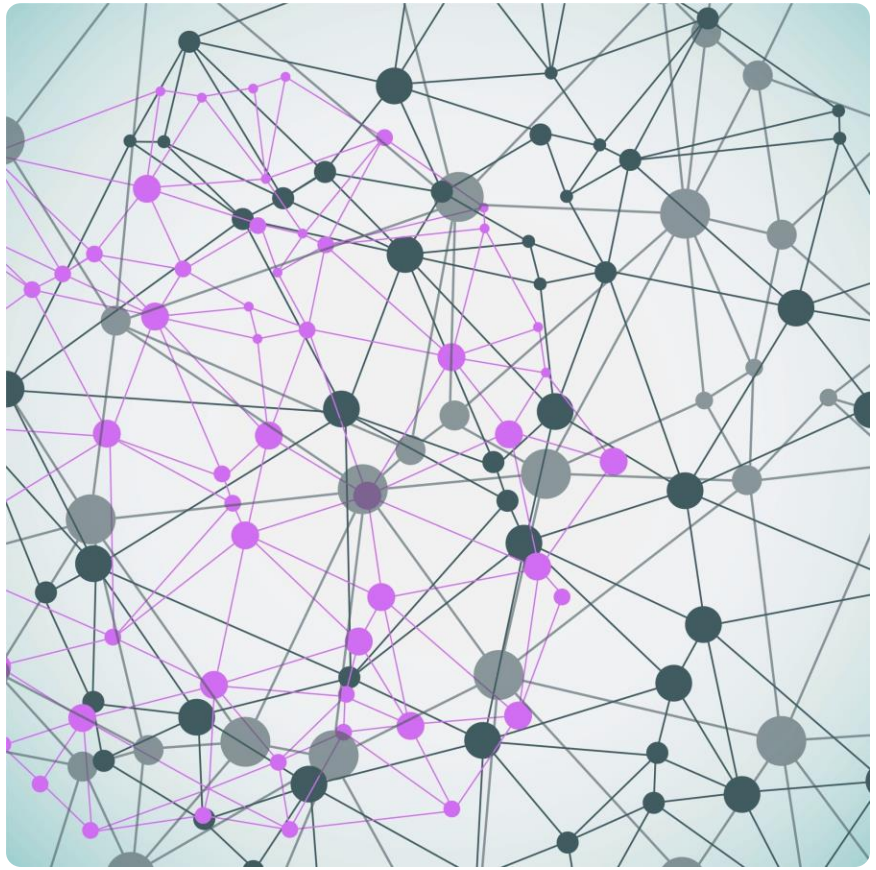








Machine Learning Model



Best ML Model

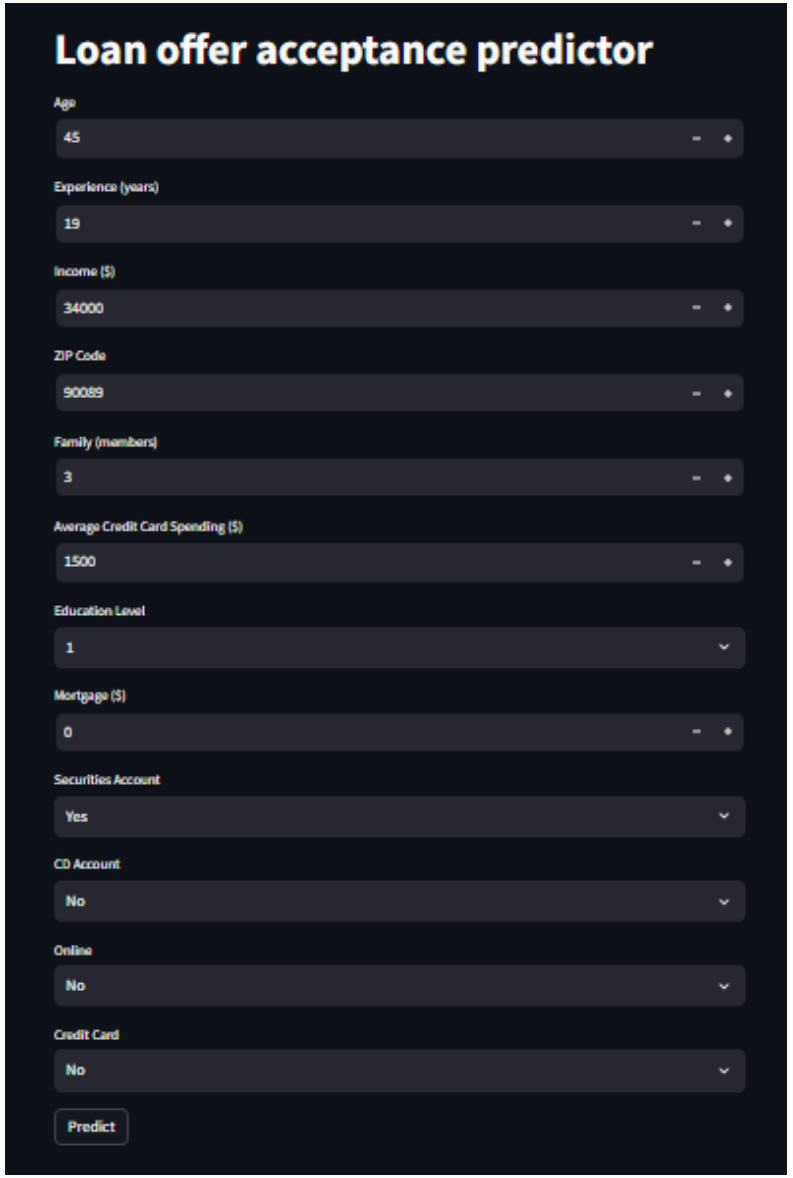
- In order to perform a classification task to predict whether a customer is going to accepte a loan, we use tree classification model from scickit-learn.
- In addition on traning the models we perform some evaluations using the following metrics Accuracy , Precision ,Recall and F1.
- Finally, the **Decision Tree Classifier** reveal itself as the best model performing this task.

	Accurracy	Precision	Recall	F1
SGDClassifier	0.916	0.000000	0.000000	0.000000
Decision Tree Classifier	0.980	0.910256	0.845238	0.876543
Logistic Regression	0.908	0.437500	0.333333	0.378378

Deployment

Deployment with streamlit

- We deploy the Decision tree Classifier we have trained . The marketing segment can then use this interface to predict if a potential customer will accept or not a loan.



COMMUNICATION

Conclusion

With the ML model the marketing segment of Galaxy Bank can pass all the customer to the model and target only those that are predicted positif for a loan acceptance

By targeting a specifique group of customer Galaxy Bank Marketing segment can consequently reduce marketing budget cost

The web app that host the Machine Learning model is user friendly and easy to use. So even non-tech people like marketers can easly use the AI we build