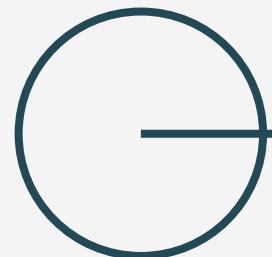


Forage : Standard Bank Tasks

Enhancing Home Loan Approvals with ML



→ Rhydhun Wate



TABLE OF CONTENT



PROJECT OVERVIEW

Introduction Part



PROCESS OVERVIEW

Data Science & ML
Lifecycle



QNA

Answering the Home
Loans Manager's
Questions



ML MODEL

Comparing AutoML vs
BespokeML



DEPLOYMENT

How the selected model
is deployed?



FUTURE SCOPE

Final Summary and
Future Development



PROJECT OVERVIEW



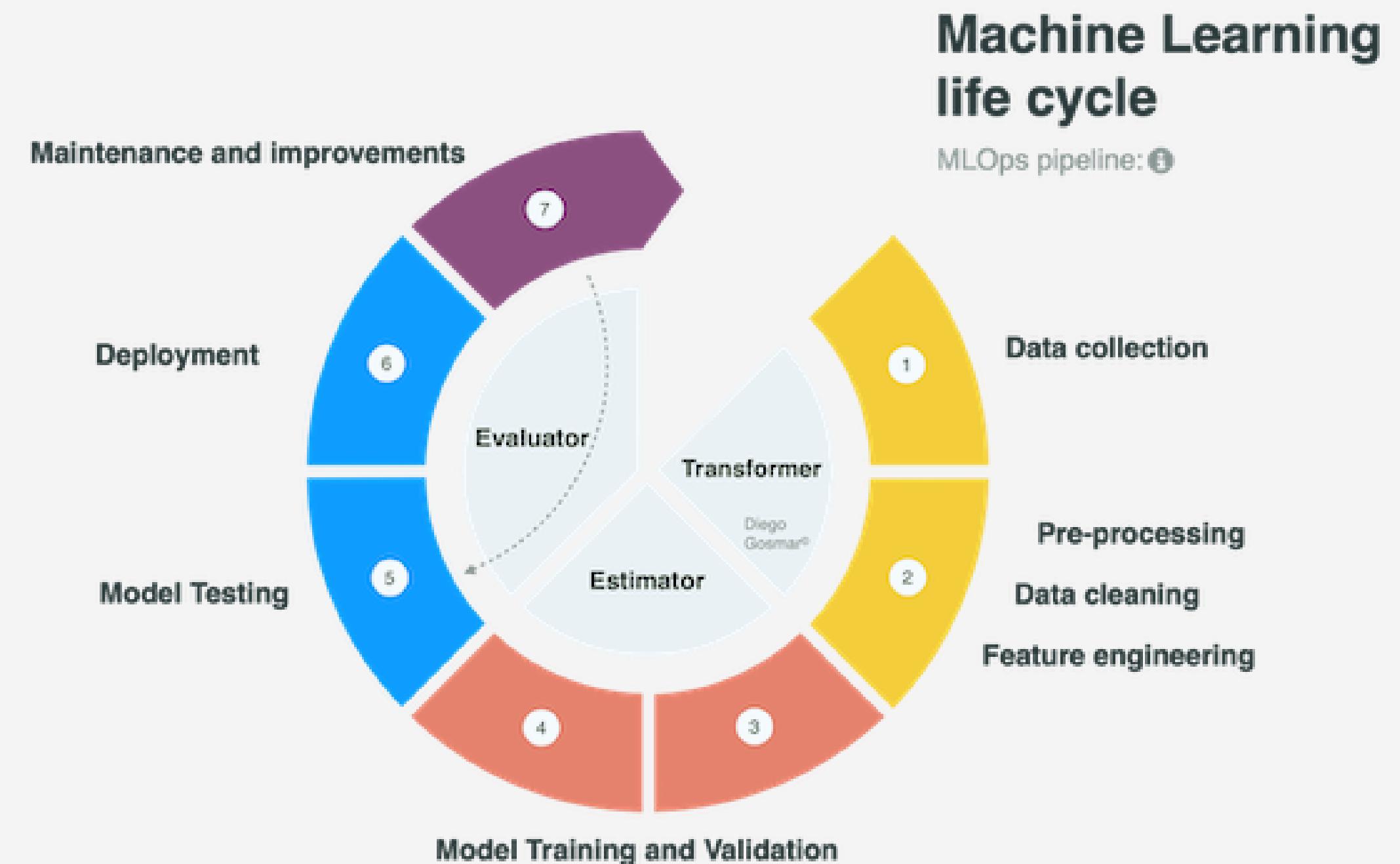
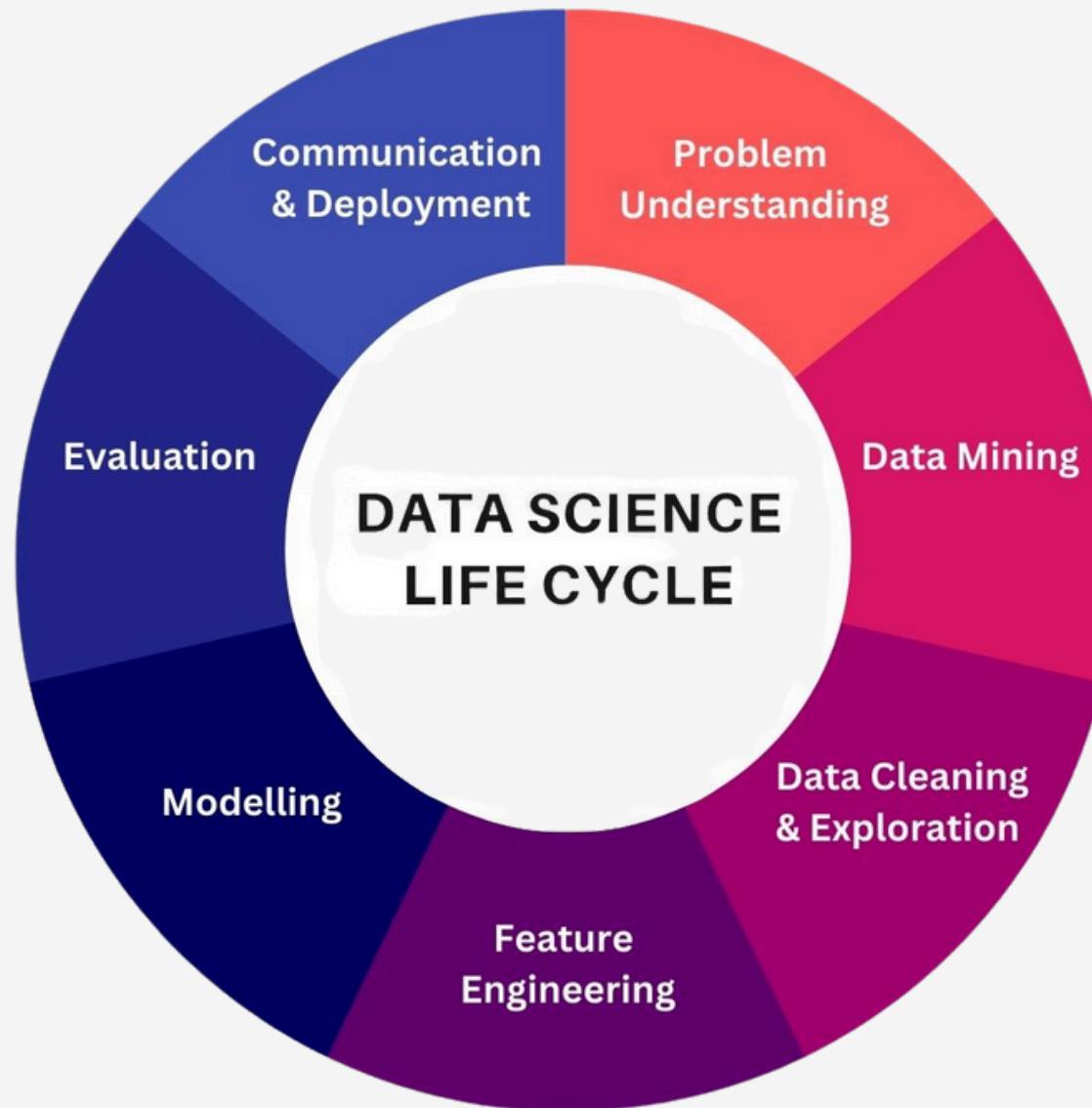
01

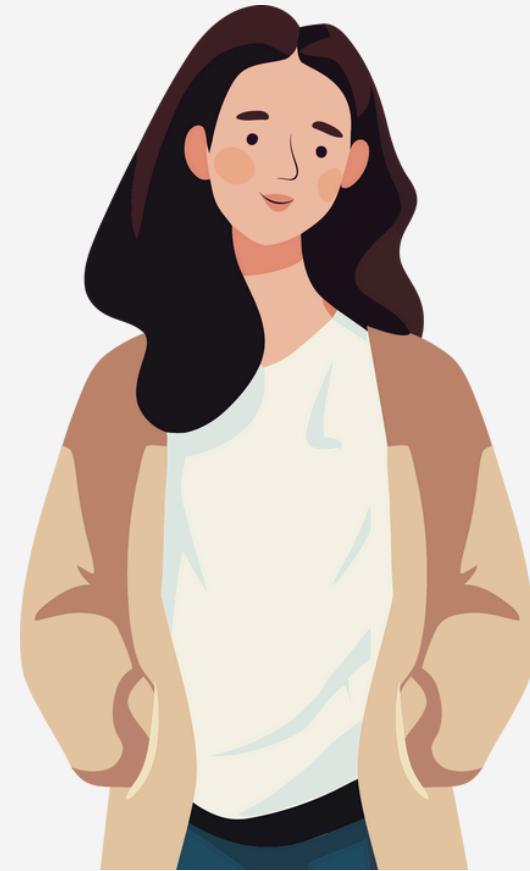
BUSINESS UNDERSTANDING:

"STANDARD BANK'S GOAL IS TO ENHANCE THE APPROVAL PROCESS FOR HOME LOANS BY PREDICTING LOAN APPROVAL LIKELIHOOD BASED ON APPLICANT DATA."

OBJECTIVE:

"WITH MACHINE LEARNING (ML), WE CAN PREDICT LOAN APPROVALS TO STREAMLINE THE DECISION-MAKING PROCESS AND IMPROVE CUSTOMER SATISFACTION."





I had 7 Questions for the Data Scientist working on the given dataset

Manager, Home Loans Department

They are answered in the following slides...



OVERVIEW OF THE DATASET

```
train.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.0	1.0	
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.0	1.0	
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.0	1.0	
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.0	1.0	
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.0	1.0	

```
test.head()
```

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property
0	LP001015	Male	Yes	0	Graduate	No	5720	0	110.0	360.0	1.0	
1	LP001022	Male	Yes	1	Graduate	No	3076	1500	126.0	360.0	1.0	
2	LP001031	Male	Yes	2	Graduate	No	5000	1800	208.0	360.0	1.0	
3	LP001035	Male	Yes	2	Graduate	No	2340	2546	100.0	360.0	NaN	
4	LP001051	Male	No	0	Not Graduate	No	3276	0	78.0	360.0	1.0	

Q1) OVERVIEW OF THE DATASET

COLUMN NAMES:



Loan_ID



Gender



Marital Status



No. of Dependents



Self-Employed or not



Applicant's & Co-applicant's Income



Urban/Rural /Semiurban



Loan Amt.



Loan Amt.



Credit History



Graduate or not

Q2) TREATING NULL VALUES

Used SimpleImputer

Q3) DISTRIBUTION OF APPROVED LOANS

422 approved loans vs 192 rejected

Q4) HOW MANY APPLICANTS HAVE DEPENDENTS?

254

Q5) EARNINGS:

On an average, Self-Employed people earn > Employed people

Q6) CREDIT_HISTORY AND LOAN_STATUS:

No significant association

Q7) 'APPLICANTINCOME' & 'LOANAMOUNT':

55% Positive Correlation



ML MODEL

The Machine Learning Model was chosen only after:

- ✓ Missing Value Imputation
- ✓ Label Encoding Categorical Data

Further, we tried to compare an Auto ML Model vs a Bespoke Model to check whose accuracies were highest.



AUTO ML

AutoML is ideal for quick, scalable, and user-friendly solutions with minimal intervention.

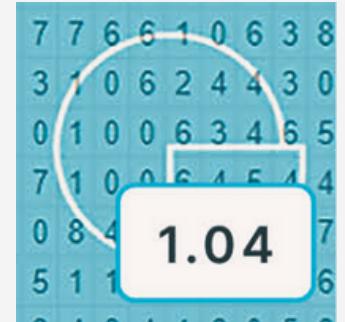
The Score by TPOTClassifier

Model came as 76%

We got this model after running several other ml models and comparing their accuracies



BESPOKE ML



Bespoke ML offers tailored, highly customized models with greater control and potential performance.

The Accuracy Score by

histgradientboostingclassifier

Model came as 91%

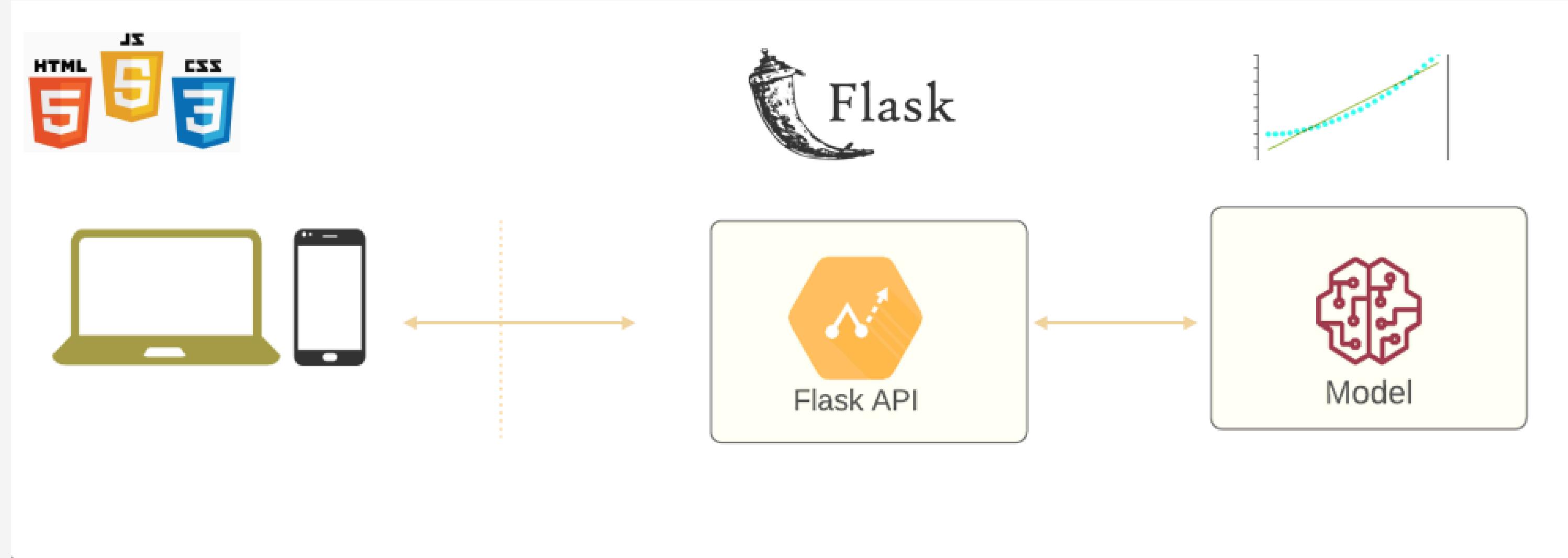
HISTGRADIENTBOOSTINGCLASSIFIER

- Efficiently handles large datasets and complex relationships.
- Uses an ensemble of decision trees to capture intricate patterns.
- Histogram-based approach optimizes speed and memory usage.
- Leads to superior classification performance.



05
**DEPLOY-
MENT**

THE PROCESS OF DEPLOYMENT



THE UI OF THE PREDICTION

Approval of Loan Application (1 for Yes/0 for No)

Gender (1 for Male / 0 for Female)

Married (1 for Yes / 0 for No)

Dependents (Number of Dependents)

Education (0 for Graduate / 1 for Not Graduate)

Self Employed (1 for Yes / 0 for No)

Applicant's Income

Coapplicant's Income

Loan Amount

Loan Amount Term

Credit History (1 for Yes / 0 for No)

Property Area (Urban-2 / Semiurban-1 / Rural-0)

TESTING IF IT WORKS

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	Loan_Status
CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area			
0	0	1	0	0	0	5849.0	1
0.0	146.412162	360.0	1	2			

Gender (1 for Male / 0 for Female)

Married (1 for Yes / 0 for No)

Dependents (Number of Dependents)

Education (0 for Graduate / 1 for Not Graduate)

Self Employed (1 for Yes / 0 for No)

Applicant's Income

Coapplicant's Income

Loan Amount

Loan Amount Term

Credit History (1 for Yes / 0 for No)

Property Area (Urban-2 / Semiurban-1 / Rural-0)

Loan Status 1

IT DOES !!



FUTURE SCOPE



Feature Engineering

Continuously explore and engineer new features from existing data sources to improve model accuracy.



Performance Monitoring

Establish robust monitoring systems to continuously track model performance and accuracy.



Model Maintenance

Regularly update and retrain models to adapt to changing data patterns and maintain high accuracy.

**THANK YOU
SO MUCH!**