

PREDICTION FOR LENDING

LOANS

PREPARED BY:
ANNA GEORGIEVA
NICK McNAMARA
OCTAVIO RIEU



INTRODUCTION



'Dream Housing
Finance Company'

'Loan Default
Prediction'

- **Last semester hypothesis**

'Credit History' main predictor for loan approvals; Research confirmed by 'Lee and Lee';

Second dataset to validate this hypothesis;

- **European Central Bank**

Current mortgage climate impact the purchasing capability of borrowers

- **New hypothetical question**

Can the findings of Lee and Lee (2018) be considered universally applicable and serve as foundational principles across diverse banking systems?



BUSINESS UNDERSTANDING

- **CRISP methodology**

Gaining an understanding of the challenges within the lending industry.

- **Hypothesis:**

'Credit History'

Were our assumption biased?

- **General goal:**

Unbiased outcomes, implement various ML models and cross-validate them; Prove or refute hypothesis;



'Credit History'

or

'Interest Rate'

&

'Income'





TECHNOLOGIES

- **Preprocessing:**

SimpleImputer
KNNImputer
LabelEncoder
StandardScaler
SMOTE

- **Machine learning Models**

Random Forest Classification
Support Vector Machine
Linear Regression
Artificial Neural Networks
Model explainability: SHAP

- **Visualisations**

Plotly express
Bokeh
Matplotlib
Seaborn
Confusion matrix

- **Hypertuning ML models**

GridSearchCV
RandomSearchCV



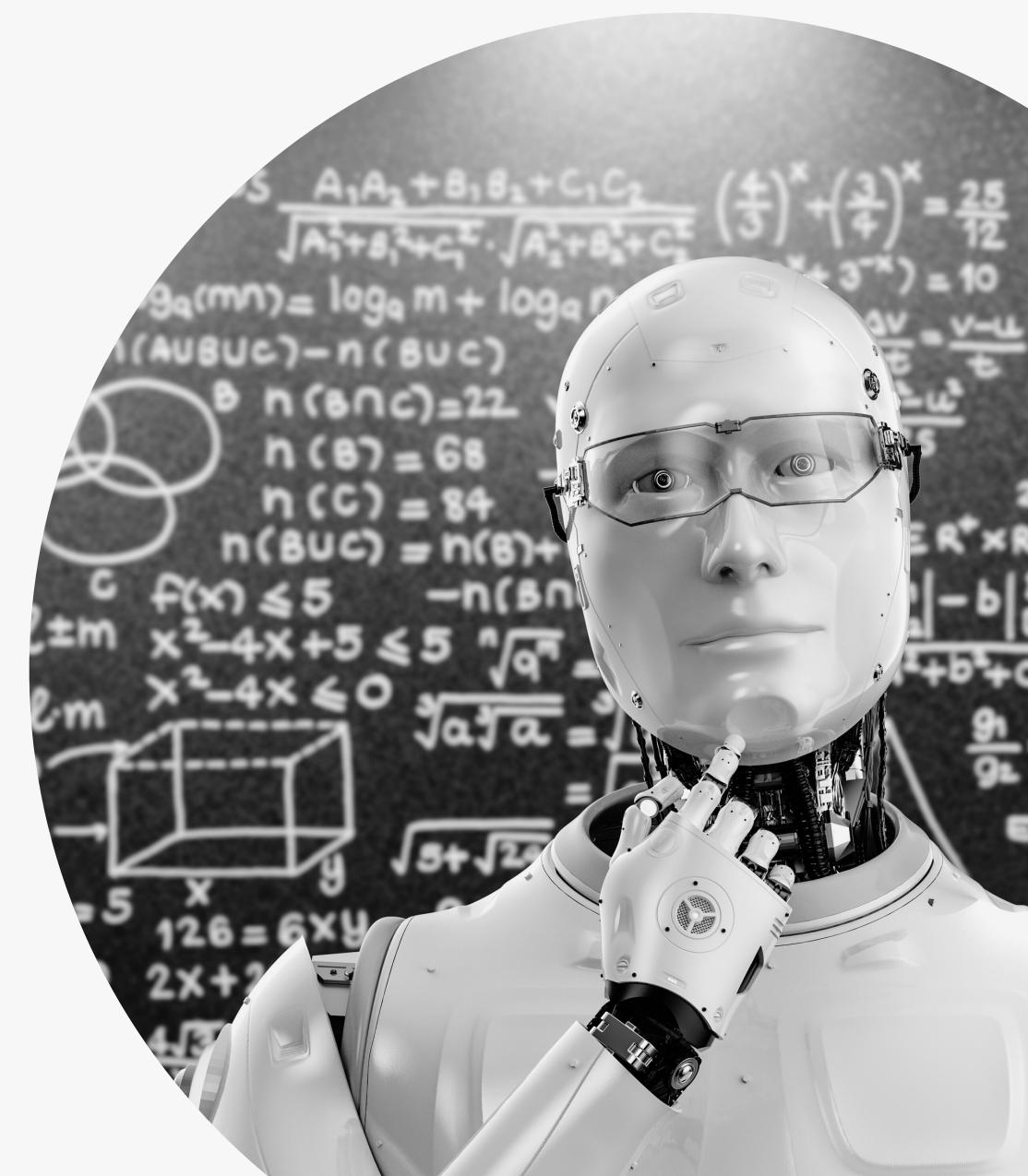
- **Jupiter Notebook** project development



- **Google Colabs** collaboration between team members



- **Github Desktop** version control

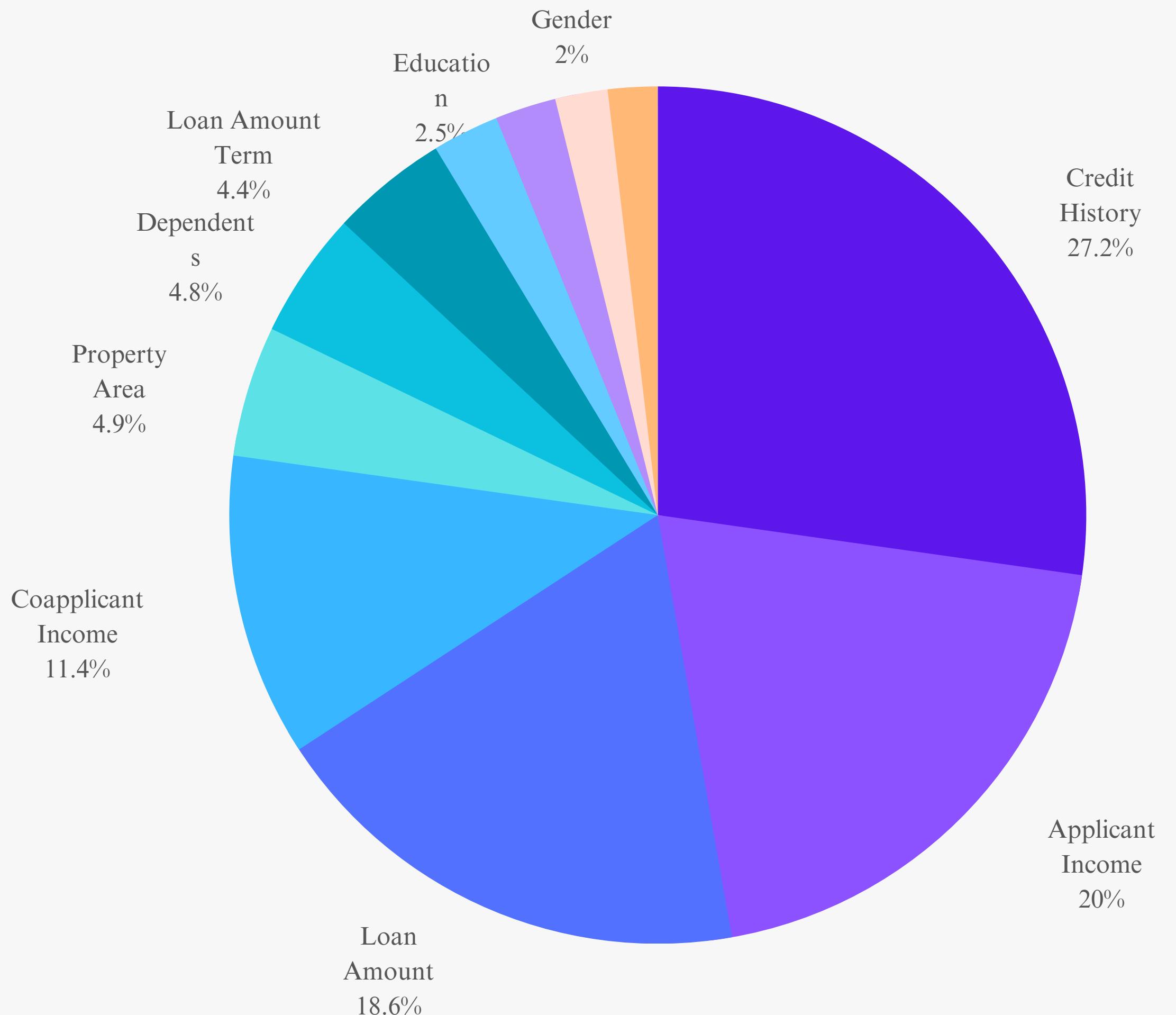


RANDOM FOREST FEATURE SELECTION

Loan Approval Dataset

X Credit History, Applicant Income,
Loan Amount

y Loan Status





BEST ML MODEL FOR LOAN APPROVAL

Support Vector Machine

True Positives

True...

ANN

True Positives

Tru...

Best performing ML model

- **SVM**

SMOTE

Balanced the minority sample

GridSearchCV

Hyperparameter tuning

Metrics

Accuracy: 1.00

Precision: 1.00

Recall: 1.00

Cross-validation: 1

- **ANN**

Training Accuracy: 100.00% - 0s 1ms/step - loss:
1.2141e-05 - accuracy: 1.0000

Testing Accuracy: 99.50% - 0s 2ms/step - loss:
0.0319 - accuracy: 0.9950

Metrics

Accuracy: 0.99

Precision: 1.00

Recall: 0.99



SVM DEPLOYED ON TEST DATASETS



Loan Approval Test Datasets

```
Prediction for instance 97: 0
Prediction for instance 98: 0
Prediction for instance 99: 0
Prediction for instance 100: 0
Prediction for instance 101: 0
Prediction for instance 102: 1 Applicant, whose mortgage loan application can be approved
Prediction for instance 103: 0
Prediction for instance 104: 0
Prediction for instance 105: 0
Prediction for instance 106: 0
Prediction for instance 107: 1 Applicant, whose mortgage loan application can be approved
Prediction for instance 108: 0
Prediction for instance 109: 0
Prediction for instance 110: 0 Applicant, whose mortgage loan application should be rejected
Prediction for instance 111: 0
Prediction for instance 112: 0
Prediction for instance 113: 0
Prediction for instance 114: 0
```

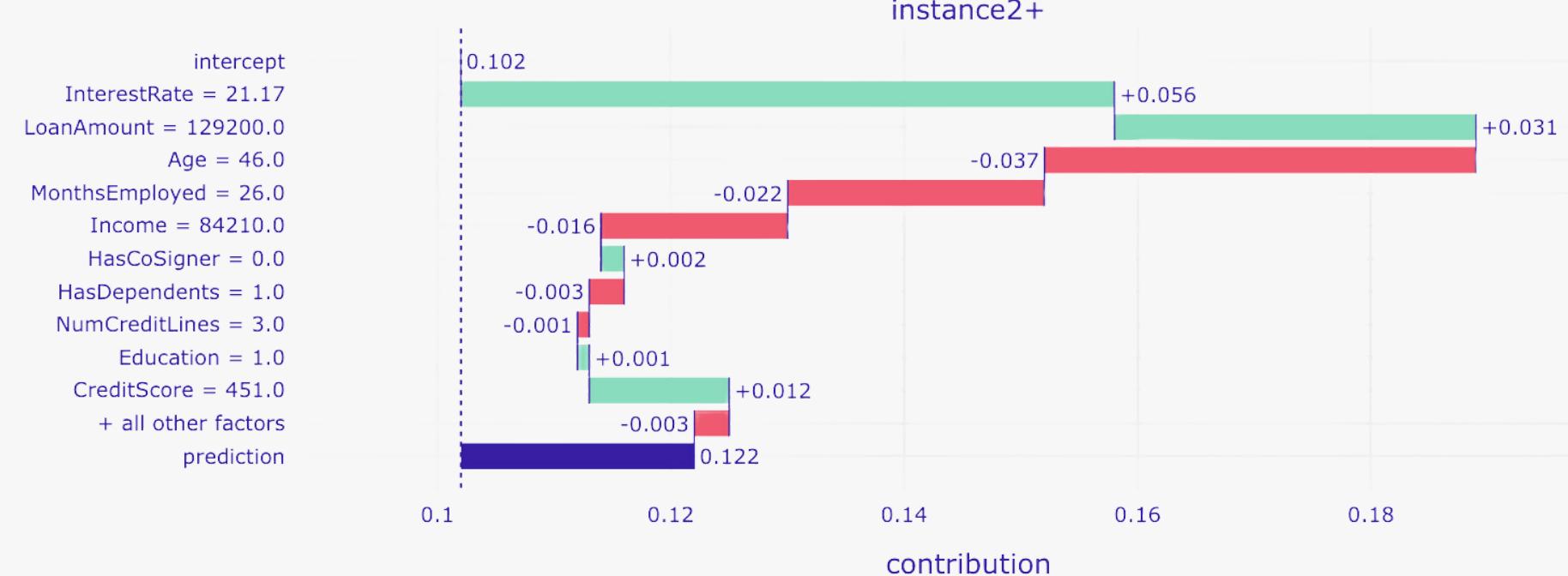
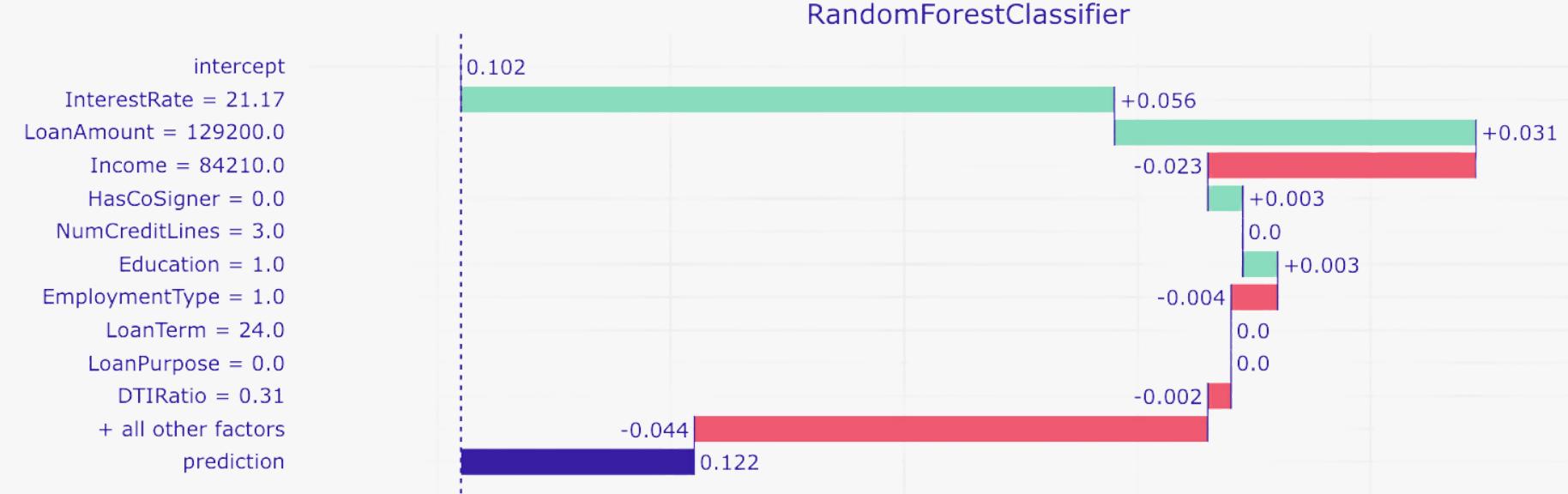




RANDOM FOREST CLASSIFICATION LOAN DEFAULT

- Random Forest Classifier
- Hyper-tuned model
- Cross-validation - 89.78%
- Instance 2 from dataset - predicted default 12%

Break Down



RESULTS AND DISCUSSIONS

- **Loan Approval Dataset**

Feature importance: Credit History, Loan Amount and Applicant Income
SVM model deployed
Prediction 6% on test dataset for approved Loan applications

- **Loan Default Dataset**

SHAP Explainer - Feature importance: Age, Interest Rate, Income
Cross-validation: 89.78%
Prediction on instance 2: 12%



Refute

Or

Validate
hypothesis



CONCLUSIONS



Refute

Or

Validate
hypothesis

- **Loan Approval Dataset**

Has Credit History, but it doesn't have Credit Score, Debt-to-Income Ratio, or Interest rate.

Relatively small dataset - SMOTE

- **Loan Default Dataset**

Big dataset of 51k observations

Proves that Age, Interest rate and Income are the most important predictors, not Credit score

Fluctuations of interest rates and risk for mortgage borrowers



THANK YOU

