

oooo



SMALL BUSINESS CREDIT-WORTHINESS

oooo

MEET THE TEAM



Brian Njau

Data Scientist



Grace Nyongesa

Data Scientist



Josiah Okumu

Data Analyst



Leo Kariuki

Data Scientist



Lynne Mutwiri

Data Storyteller/
ML Engineer



Stella Kitur

Data/ML Engineer



BUSINESS OVERVIEW

- SBA supports small businesses with loan guarantees, promoting job creation, and minimizing defaults.
- SBA also offers mentorship, training, contracting opportunities, and disaster recovery aid for small business success.

PROBLEM STATEMENT

- According to SBA Data, 18% of loans end up being charged off, that is, not paid in full. This percentage, although seemingly small, translates to losses for the organization and reduced bank reputation



PROJECT OBJECTIVES

Analyze SBA dataset
for loan patterns.

Develop a precise
Machine Learning
model for loan
approvals.

Deploy and refine
the model in the
SBA's process.

Data Source

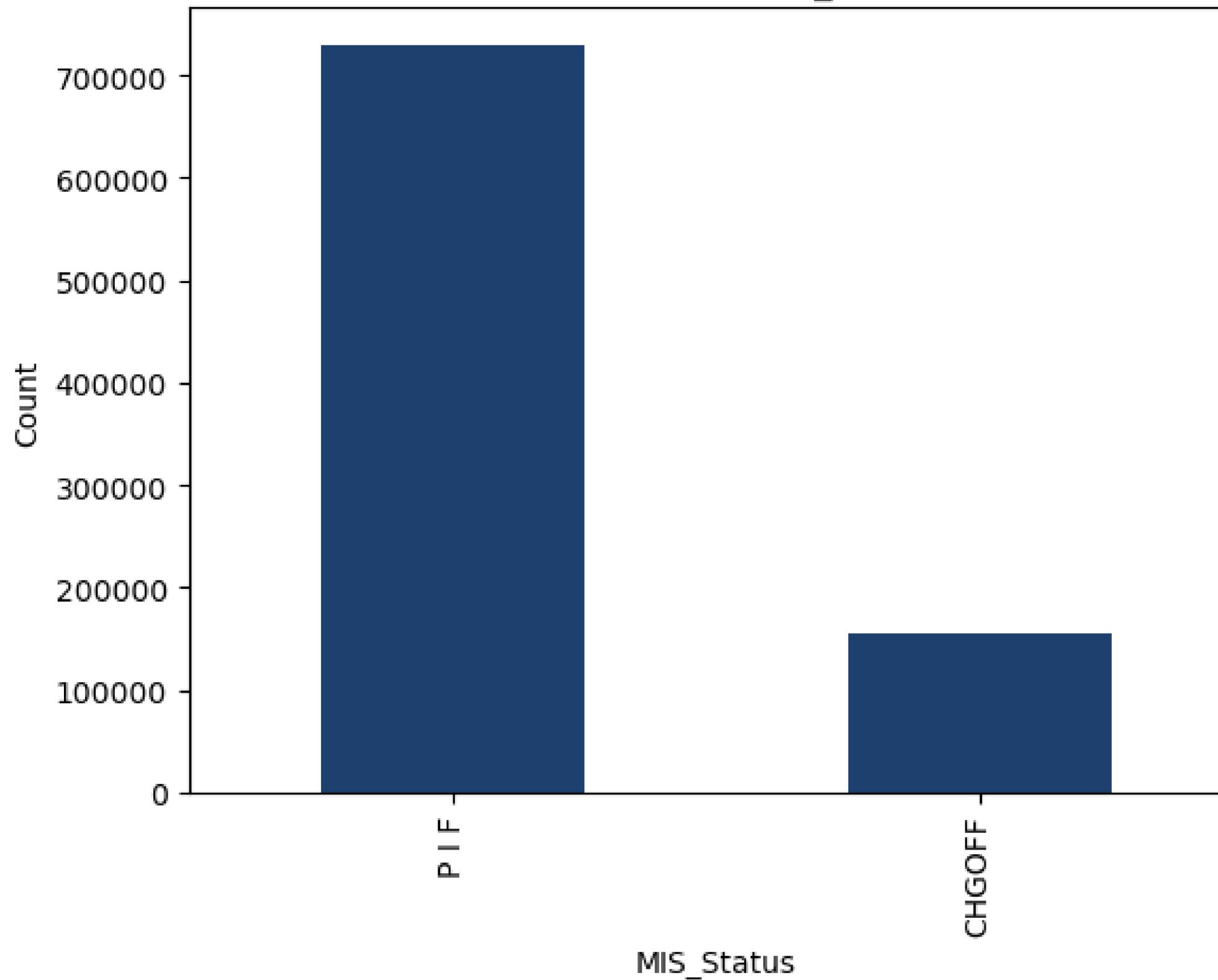
Our data was sourced from the official SBA data portal.



Data Shape

The data had
899164 rows and
27 columns

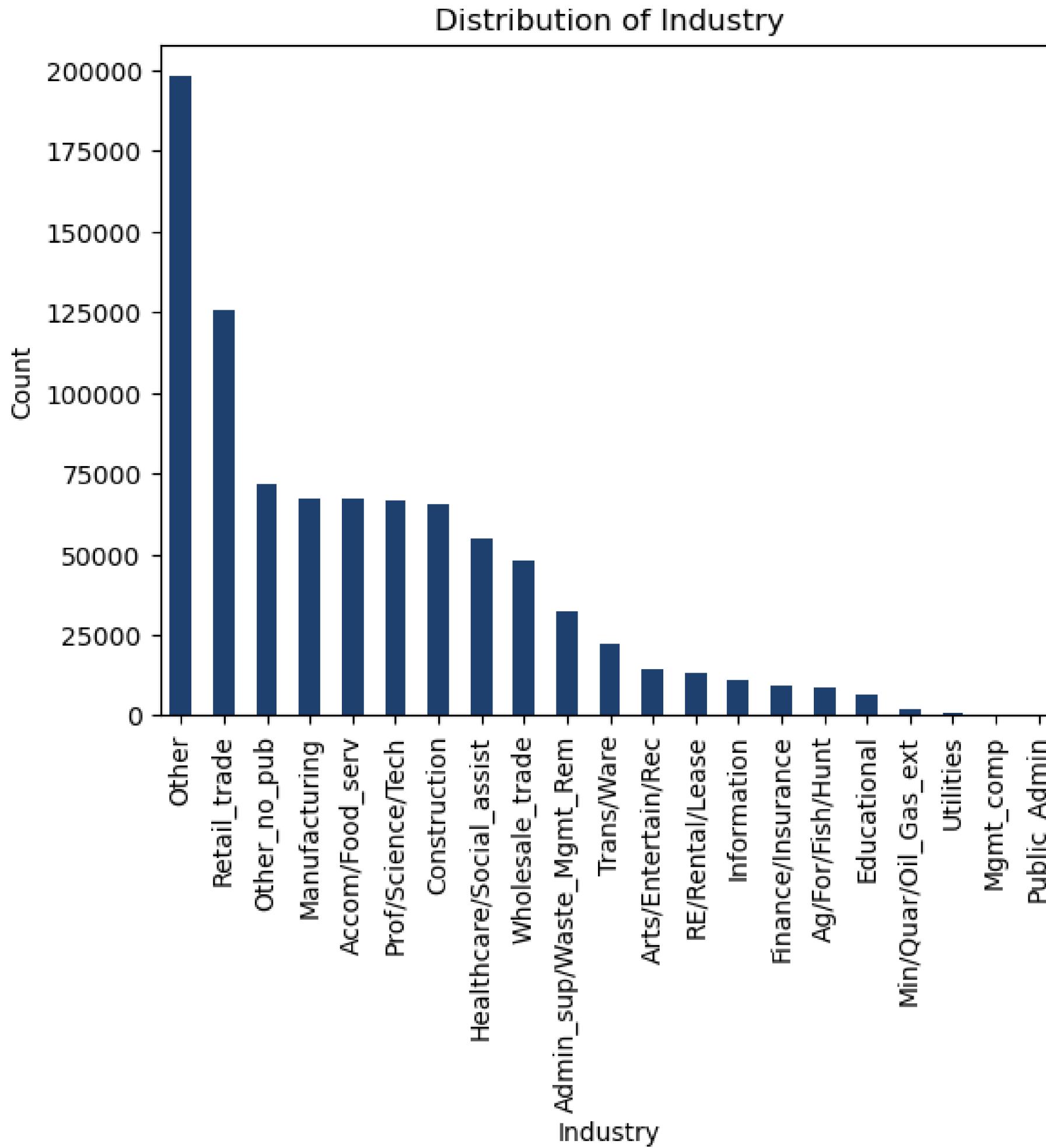
Distribution of MIS_Status



MIS Status was the target variable.

82 % of the loans that were approved, were *Paid in Full (PIF)*

18 % of the loans approved were *Charged Off (CHGOFF)*



The top industries were :

1. Other/Miscellaneous
2. Retail trade
3. Manufacturing
4. Accommodation and Foodservice

MODELING

MODEL	TRAIN ACCURACY	TEST ACCURACY
Logistic Regression	78%	75%
Decision Tree	98%	89%
Random Forest	87%	88%
Support Vector Machine	77%	78%
Neural Network	89%	89%
XGBoost	91%	90%

○ ○ ○ ○

DEPLOYMENT

- The app was deployed using Streamlit.

The input features were:

- Gross Approval Amount
- Industry
- Loan Term
- Number of Employees
- Existence Period
- Region

○ ○ ○ ○

App



CONCLUSION

- Enhanced predictive capabilities of the SBA have automated loan filtration, simplifying and expediting the lending process for entrepreneurs and small businesses.



oooo

RECOMMENDATIONS

- Consider automating the initial filtration of loan applications using this data-driven algorithm, so as to improve the accuracy and efficiency of loan approvals.
- Conduct regular retraining of the loan classifier model using the latest data to maintain its relevancy and adaptability.



oooo



THANK YOU

Project collaborators

Brian Njau
Grace Nyongesa
Josiah Okumu
Leo Kariuki
Lynne Mutwiri
Stella Kitur



GitHub Repository

