Enhancing Loan Approval Processes through Predictive Modeling

Maddie Lee, Alexia Wells, Leah Ekblad, Whitney Holt

HOME CREDIT

Introduction

HOME CREDIT

Introduction

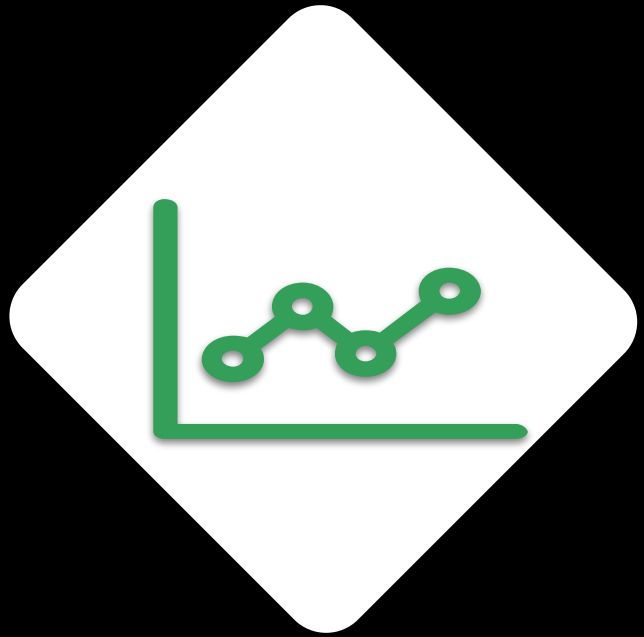
HOME CREDIT

Introduction

HOME CREDIT

Ensemble Model

- Models
  - Random Forest
  - Bayesian Additive Trees
  - Logistic Regression
  - Extra Trees
- Kaggle AUC: 0.70311
- Runtime: 1.05 hours

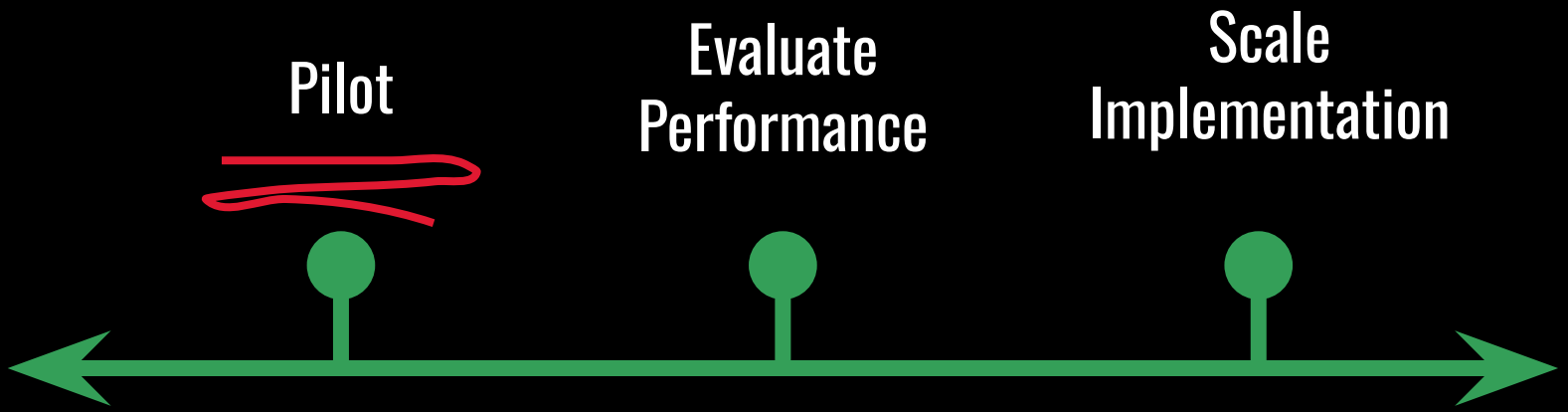


HOME CREDIT

Recommendation

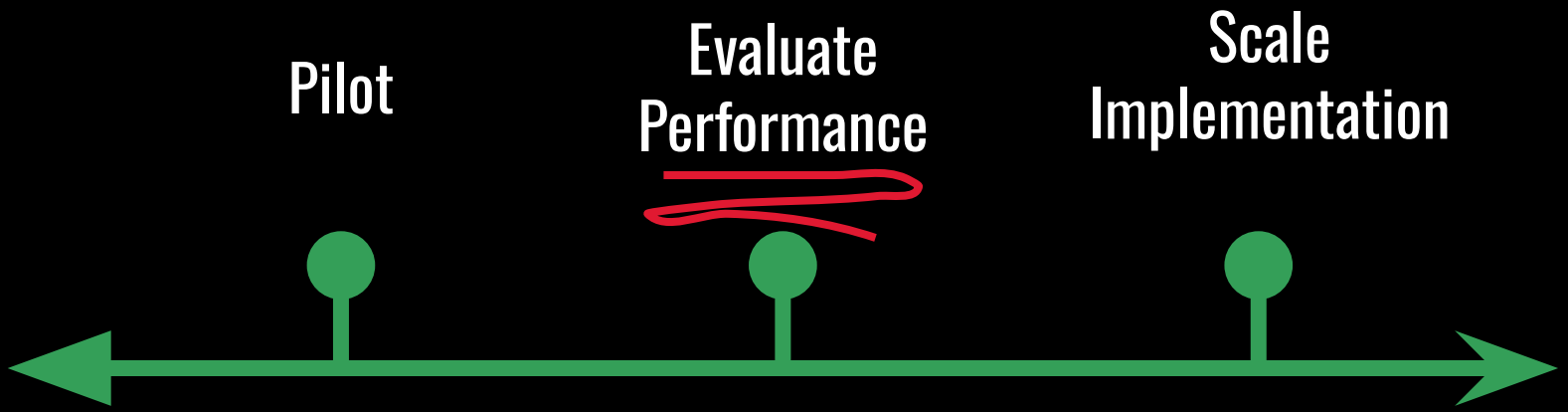
HOME CREDIT

Recommendations



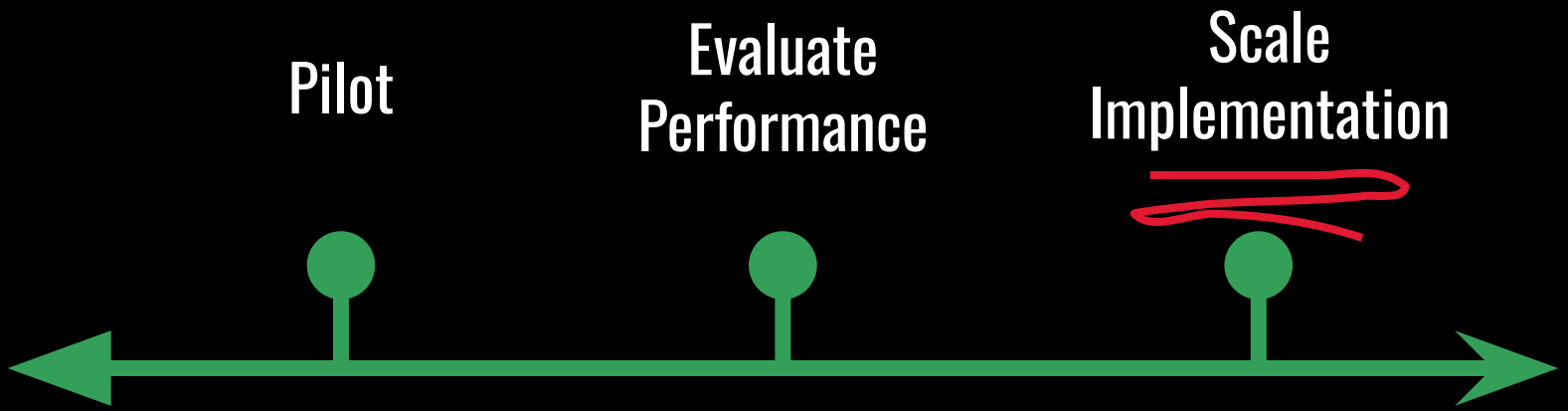
HOME CREDIT

Recommendations



HOME CREDIT

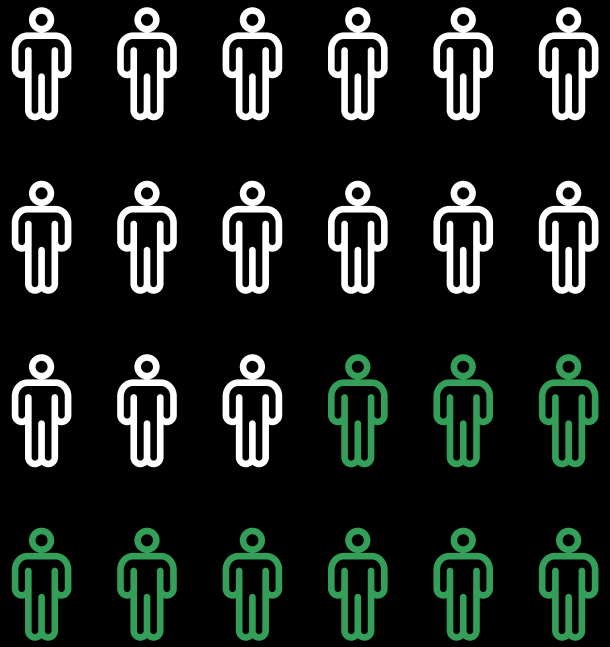
Recommendations



HOME CREDIT

Business Implications

HOME CREDIT



Expand Access to Credit

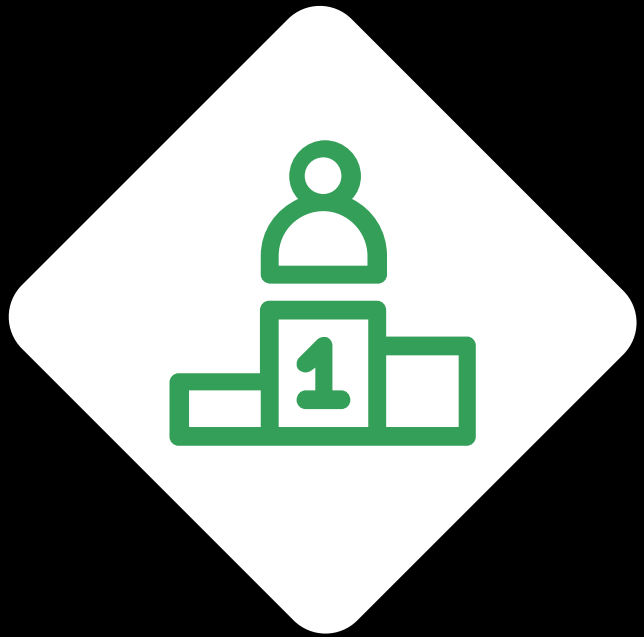
HOME CREDIT

Drive Revenue Growth

HOME CREDIT

Maintain Risk Control

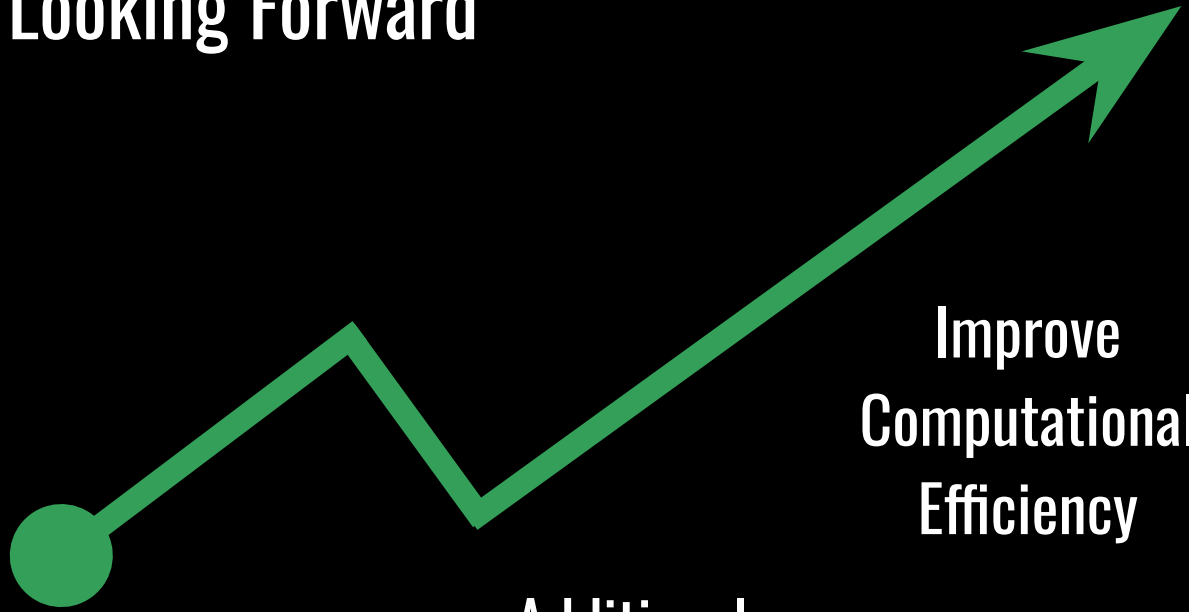
HOME CREDIT



Enhance Competitive Advantage

HOME CREDIT

Looking Forward



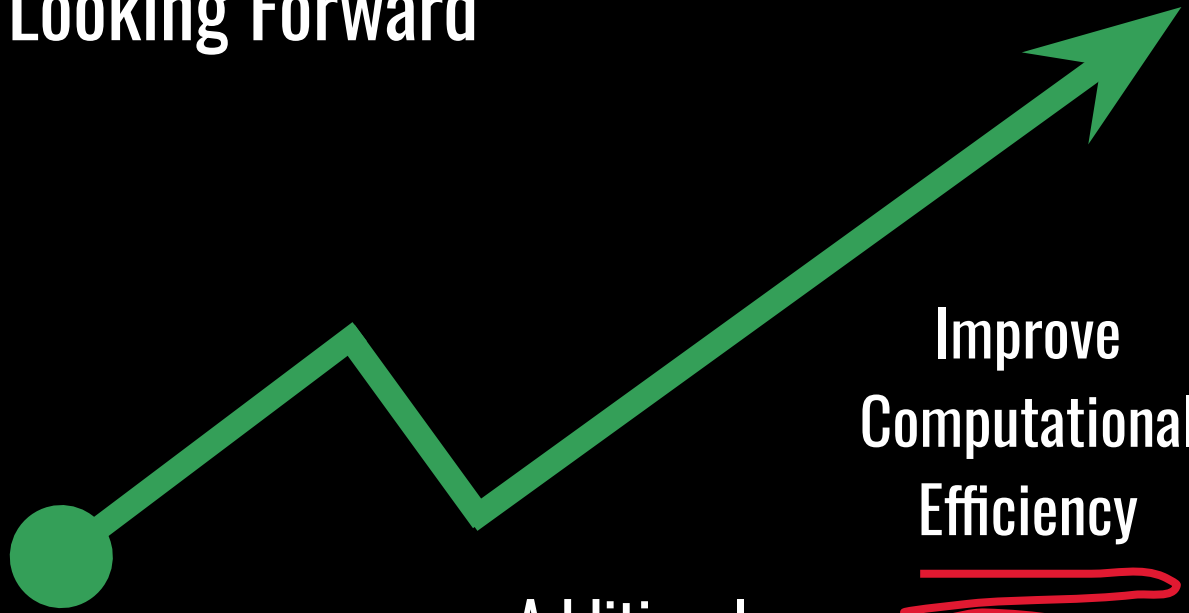
Expand Application

Improve Computational Efficiency

Additional Features

HOME CREDIT

Looking Forward



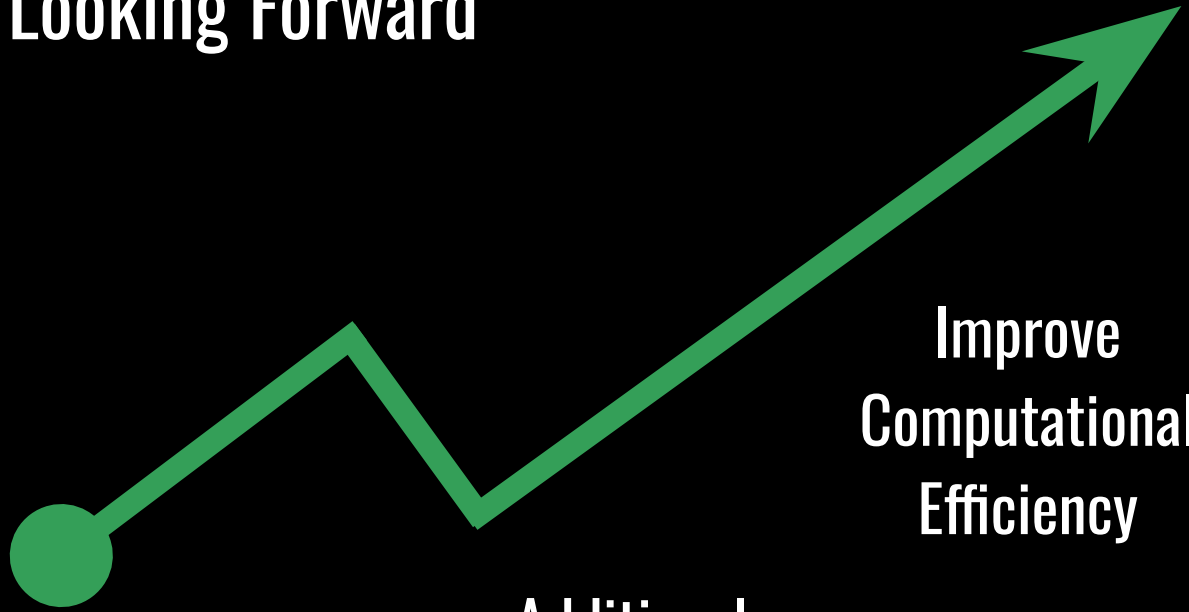
Expand Application

Improve Computational Efficiency

Additional Features

HOME CREDIT

Looking Forward



Additional Features

Improve Computational Efficiency

Expand Application

HOME CREDIT

QUESTIONS?

HOME CREDIT

Appendix

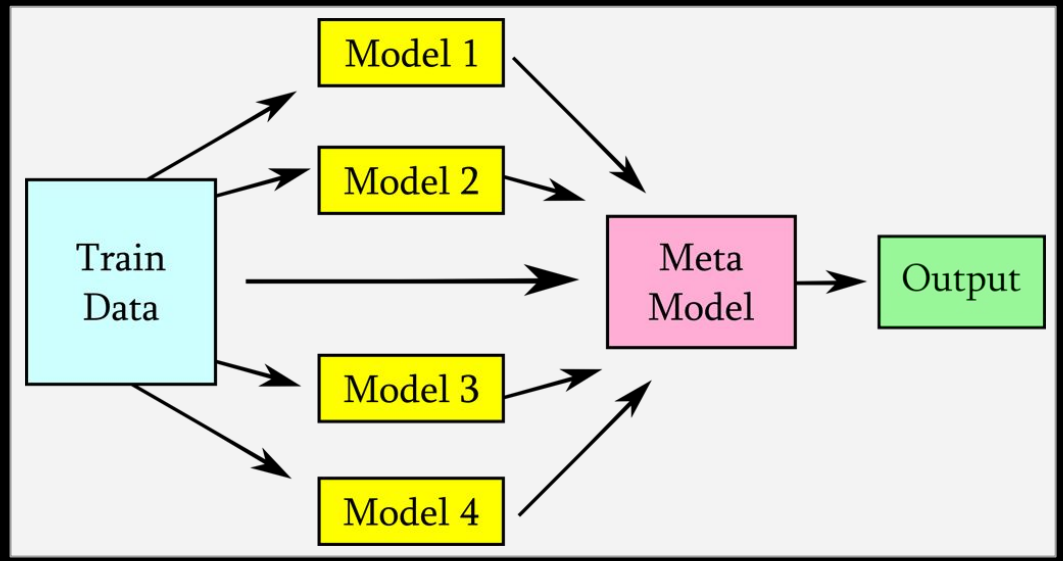
Model Performance Comparison

Model Details		Performance Metrics		
Model Type	**Model Description**	**In-Sample AUC**	**Out-of-Sample AUC**	**Kaggle Score**
Majority Class Classifier	Baseline Model	0.500	0.50	NA
Logistic Regression	Feature engineering, all predictors	0.770	0.66	NA
Logistic Regression	All predictors	0.680	0.65	NA
Logistic Regression	All predictors, stepwise	0.740	0.67	0.677230
Logistic Regression	Feature engineering, all predictors	0.691	NA	0.678960
Logistic Regression	Feature engineering, attribute selection	NA	NA	0.639271
Naive Bayes	Feature engineering	0.680	0.66	0.582870
LASSO Regression	Feature engineering	0.720	0.68	0.676590
Random Forest	500 trees	0.720	0.68	NA
Random Forest	Feature engineering, 500 trees	NA	NA	0.673620
Random Forest	Feature engineering, 100 trees	NA	NA	0.663660
Random Forest	Top 10 features, 500 trees	NA	NA	0.663660
Random Forest	Using LASSO-selected predictors, 500 trees	0.710	0.68	0.676240
Random Forest	MTRY Adjustment, 500 trees	NA	NA	0.673840
Random Forest	Feature engineering, all predictors, auto hyperparameter tuning	0.690	NA	0.698770
Extra Trees	Feature engineering, all predictors, 1000 trees, auto hyperparameter tuning for min_n()	0.690	NA	0.693390
LightGBM Boosted Trees	Feature engineering, all predictors, auto hyperparameter tuning	0.690	NA	0.686760
Bayesian Additive Regression Trees (BART)	Feature engineering, all predictors, auto hyperparameter tuning	0.693	NA	0.701130
Ensemble Model	LightGBM + BART + Logistic Regression	NA	NA	0.691690
Ensemble Model	Random Forest, Logistic Regression, and BART	NA	NA	0.700500
Ensemble Model	LightGBM + Random Forest + Extra Trees + Logistic Regression	NA	NA	0.697650
Ensemble Model	Random Forest + BART + Logistic Regression + Extra Trees	NA	NA	0.703110

HOME CREDIT

Appendix - Stacking Model Predictions

- Goal - Reduce bias and improve predictions
  - Combines predictions of diverse models
  - Meta-learner that makes the final predictions



HOME CREDIT

Appendix

```
# Stacked predictions/ensemble method

# Best performing
my_stack <- stacks() %>%
  add_candidates(rf_CV_results) %>%
  add_candidates(bart_CV_results) %>%
  add_candidates(log_CV_results) %>%
  add_candidates(extra_trees_CV_results)

## Fit the stacked model
stack_mod <- my_stack %>%
  blend_predictions() %>% # LASSO penalized regression meta-learner
  fit_members() ## Fit the members to the dataset

## Use the stacked data to get a prediction
stacked_predictions <- stack_mod %>%
  predict(new_data=test, type = "prob")
```

HOME CREDIT

Appendix - Ensemble Model

- Models Used
  - Random Forest - Bagging/boost strapping aggregating
  - Bayesian Additive Trees - Essentially looped boosting algorithm
  - Logistic Regression
  - Extra Trees - Bagging
- Kaggle AUC: 0.70311
- Runtime: 1.05 hours

HOME CREDIT