SMART UNDERWRITING

DATA-DRIVEN HOME INSPECTION INSIGHTS

PURDUE UNIVERSITY

Mitchell E. Daniels, Jr.
School of Business

INDIANA FARM BUREAU INSURANCE

TEAM

Ashvin Raj

Chu-Yun Hsiao

Raghav Chadha

Xiaoyu Guan

Yi-Chun Huang

PURDUE UNIVERSITY | Mitchell E. Daniels, Jr. School of Business

AGENDA

01 BUSINESS PROBLEM

02 ANALYTICAL PROBLEM FRAMING

03 METHODOLOGY

04 DATASET

05 MODEL BULIDING

06 DEPLOYMENT AND BENEFITS

INTRODUCTION

House insurance inspections are vital evaluations of a property's condition, structure, systems, and risks. These results guide decisions on property insurance suitability, coverage levels, and pricing.

PROJECT OVERVIEW

MISSION:

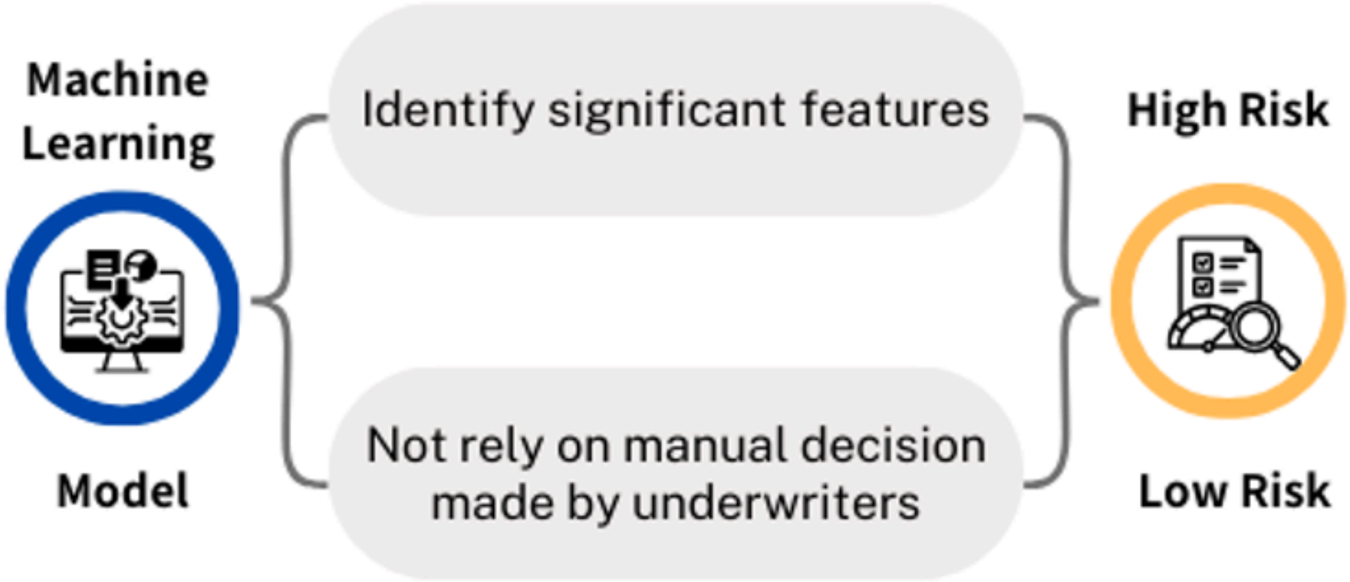
1. Develop an enhanced predictive model that more accurately identifies properties requiring inspections.
2. Determine the key risk factors influencing inspection needs.
3. Recommend strategic actions to mitigate risks associated with high-risk properties.

GOAL:

Develop effective machine learning models to gain insights on house inspection prediction. Propose business strategies to proactively screen high-risk houses at the quote stage.

BUSINESS PROBLEM

ANALYTICS PROBLEM FRAMING



SUCCESS METRICS

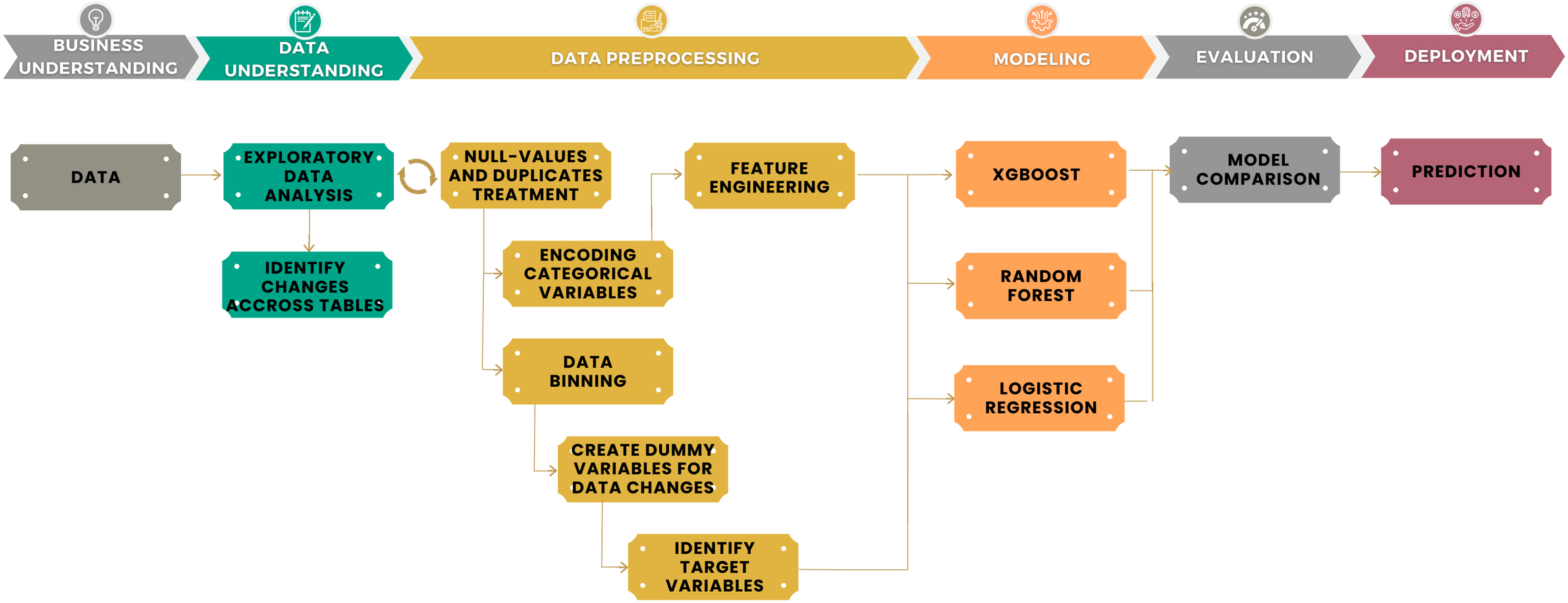
ROC Curve

Confusion Matrix

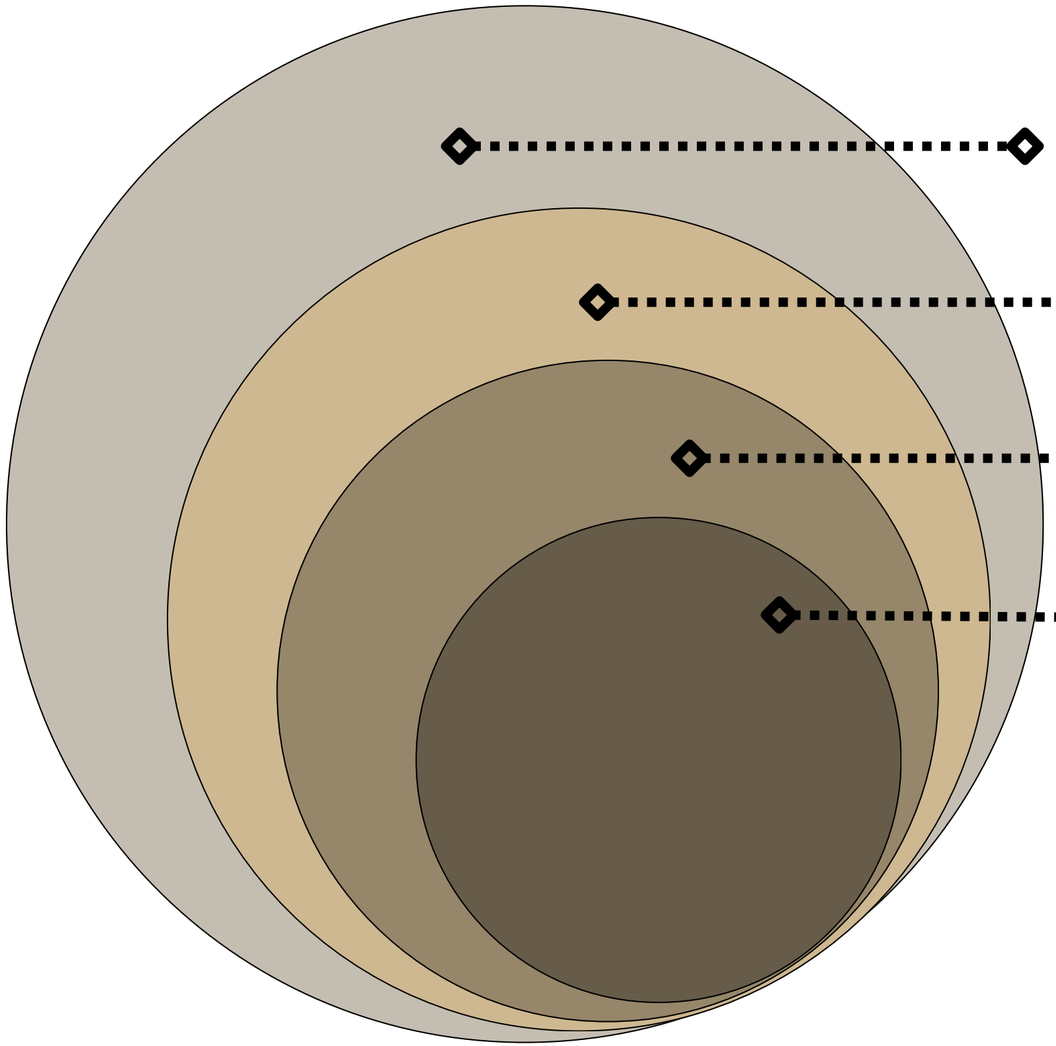
F1 Score

TECHNOLOGY FRAMEWORK

METHODOLOGY



DATASET
OVERVIEW



106K rows of different cases

Duration of *9 years* (2016-2024)

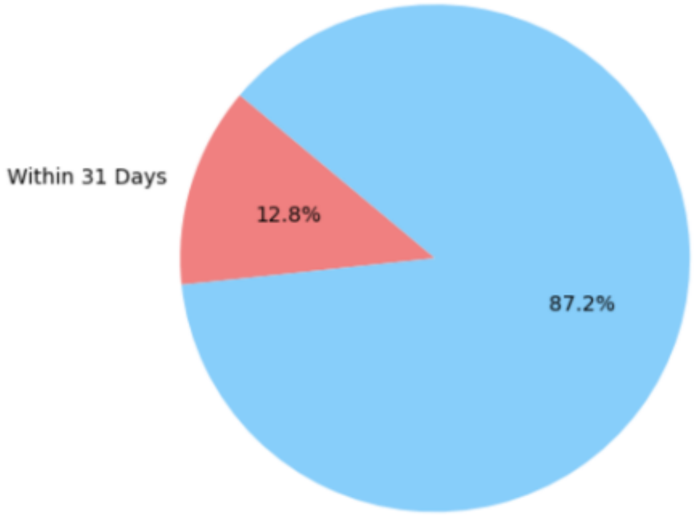
Quote stage inspection data merged with SCD and Policy table

Target Variable composed of 3 key variables

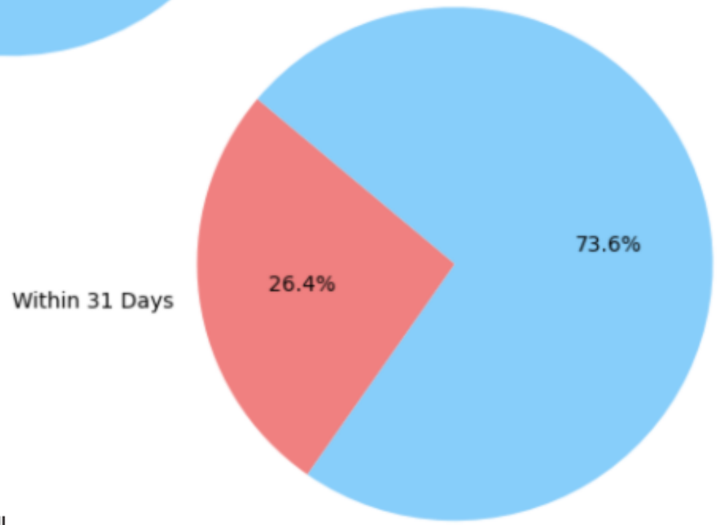
Cleaned Dataset -> **94K** rows

DATASET
EDA

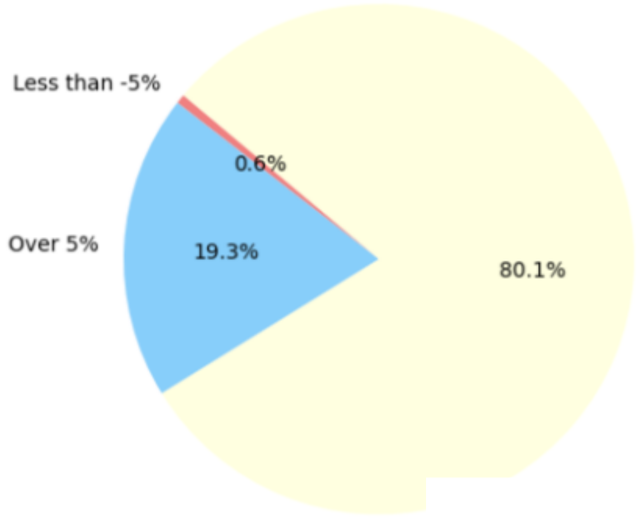
Coverage A



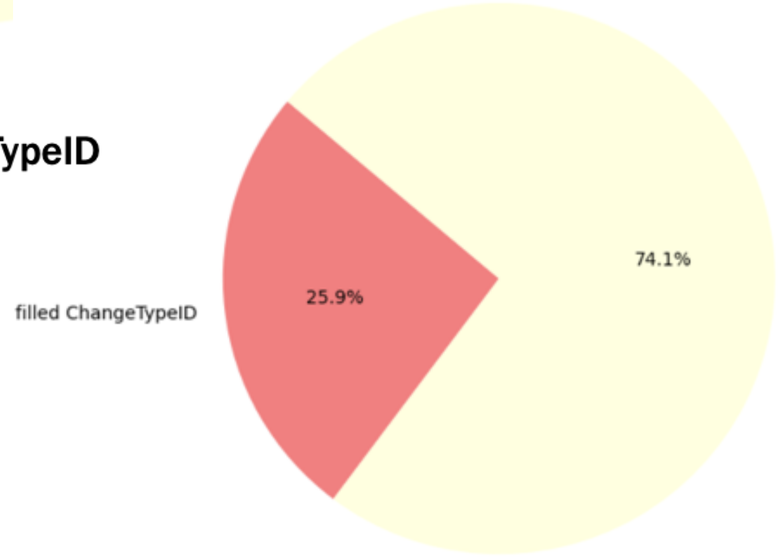
Dwelling



Coverage A change %

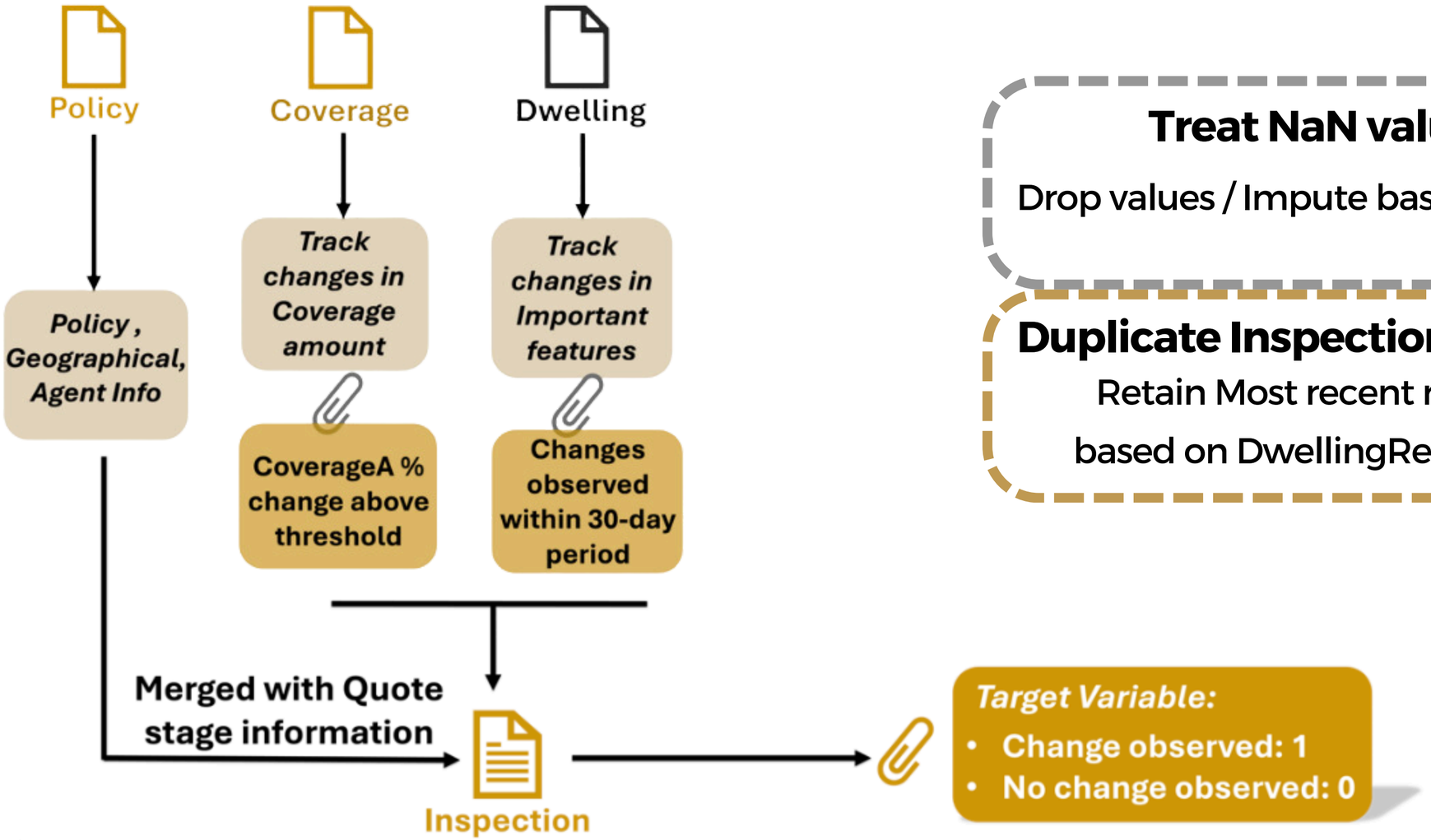


Inspection — ChangeTypeID



PURDUE UNIVERSITY | Mitchell School of Business

DATASET
PREPROCESSING



Treat NaN values

Drop values / Impute based on logic

Duplicate Inspection records

Retain Most recent record based on DwellingRequestID

DATASET
PREPROCESSING

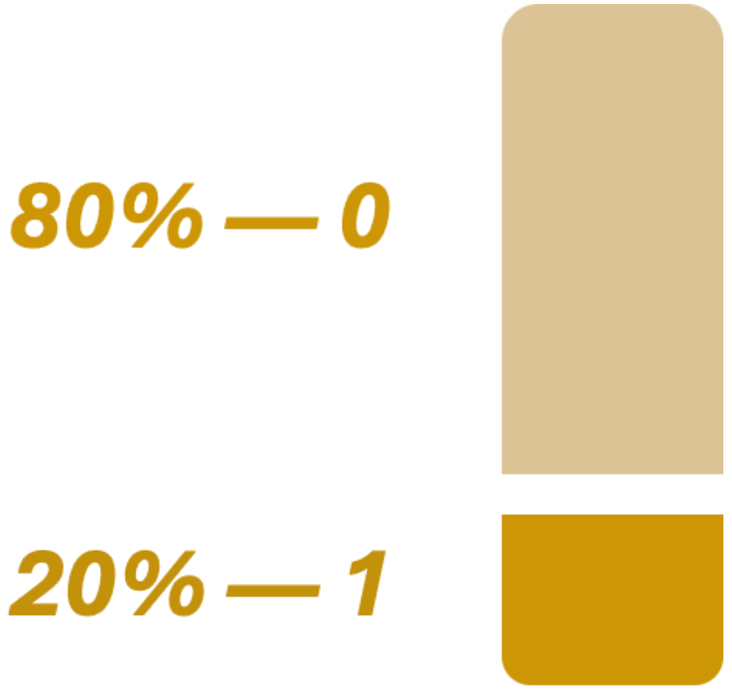
UPDATED

Isolated changes which happened over a window of 30-day for each policy

Target variable:
- ***Change observed: 1***
- ***No change observed: 0***

KEY Variables

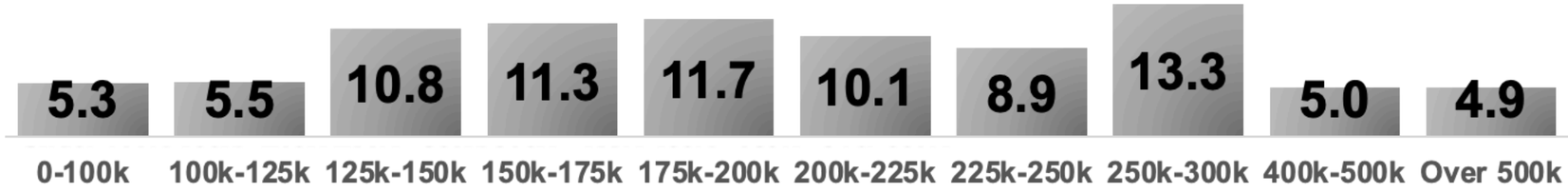
- CovA_SCD_Change
- Dwelling_Change
- ChangeTypeID
- **Target_Changes**



FEATURE ENGINEERING

COVERAGE AMOUNT

Categorizing houses based on insurance coverage value



DEDUCTIBLE AMOUNT

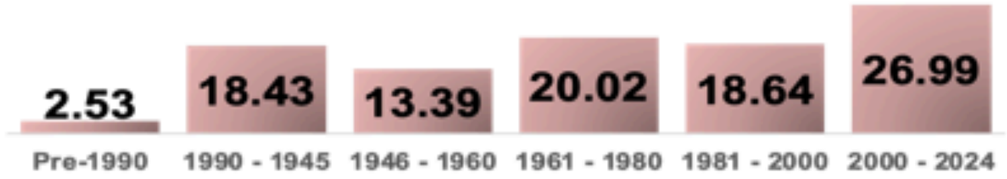
Simplifying model's understanding of risk associated with varying deductibles

REPLACEMENT COST

Risk assessment based on potential replacement expenses

CONSTRUCTION YEAR

Accounting for variability in building standards & materials across years



MODEL BUILDING

One-Hot Encoded Variables:

CoverageAAmount_Binned

Deductible_Binned

ReplacementCost_Binned

ConstructionYear_Era

AgentType

PopulationClassification

HouseholdStatus

RCTMethodID

DwellingUnderConstruction

ConditionID

Label (Ordinal) Encoded Variables

ConstructionType

PolicyTypeID

FireProtectionID

RoofingMaterialId

County

Numerical Variables

RoofAge

AnimalCount

YearsSince

InsuranceScoreLevel

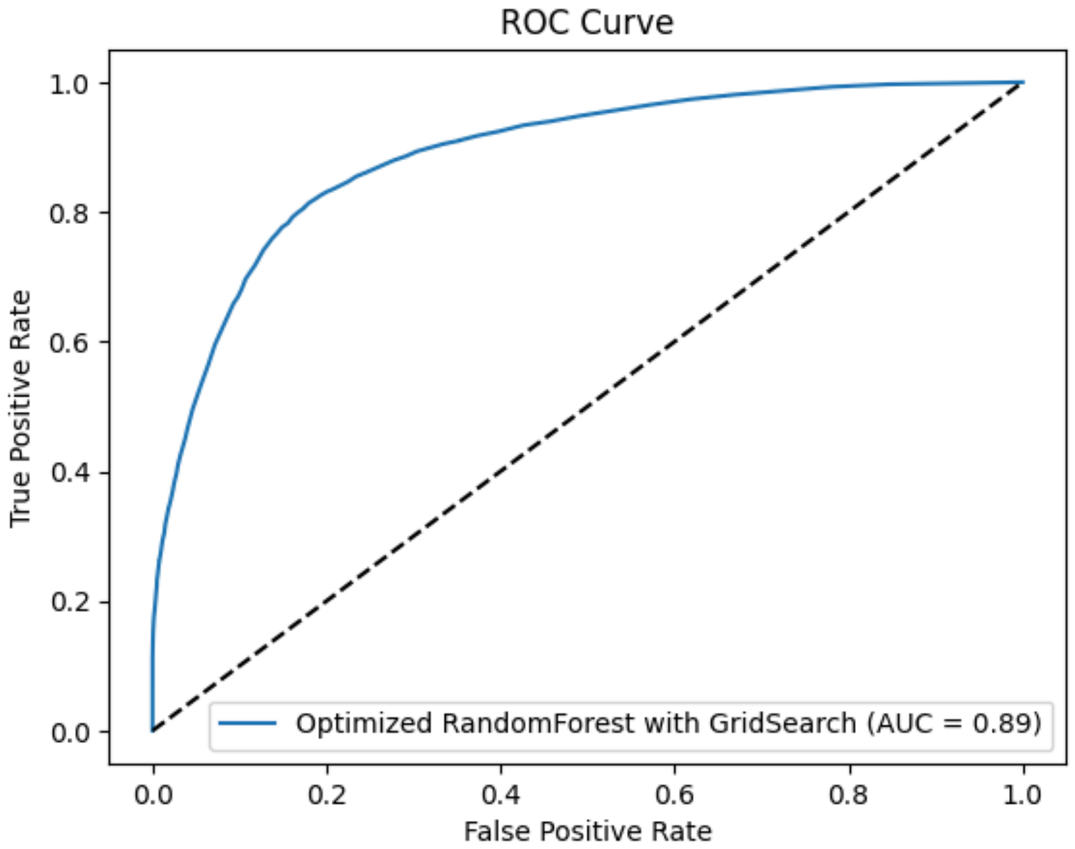
MODEL BUILDING

MODEL TRAINING AND EVALUATION

Data Splitting: Train (80%) test (20%) split ratio allows for a sufficient amount of data to train the models while still reserving a sizable portion for independent evaluation

Cross-validation and Hyper-parameter Tuning: Employed K-Fold cross-validation (5 folds) techniques to ensure model robustness
Hyperparameters tuned to optimize performance

Precision	Recall	F1 Score
0.74	0.58	0.65



MODEL EVALUATION

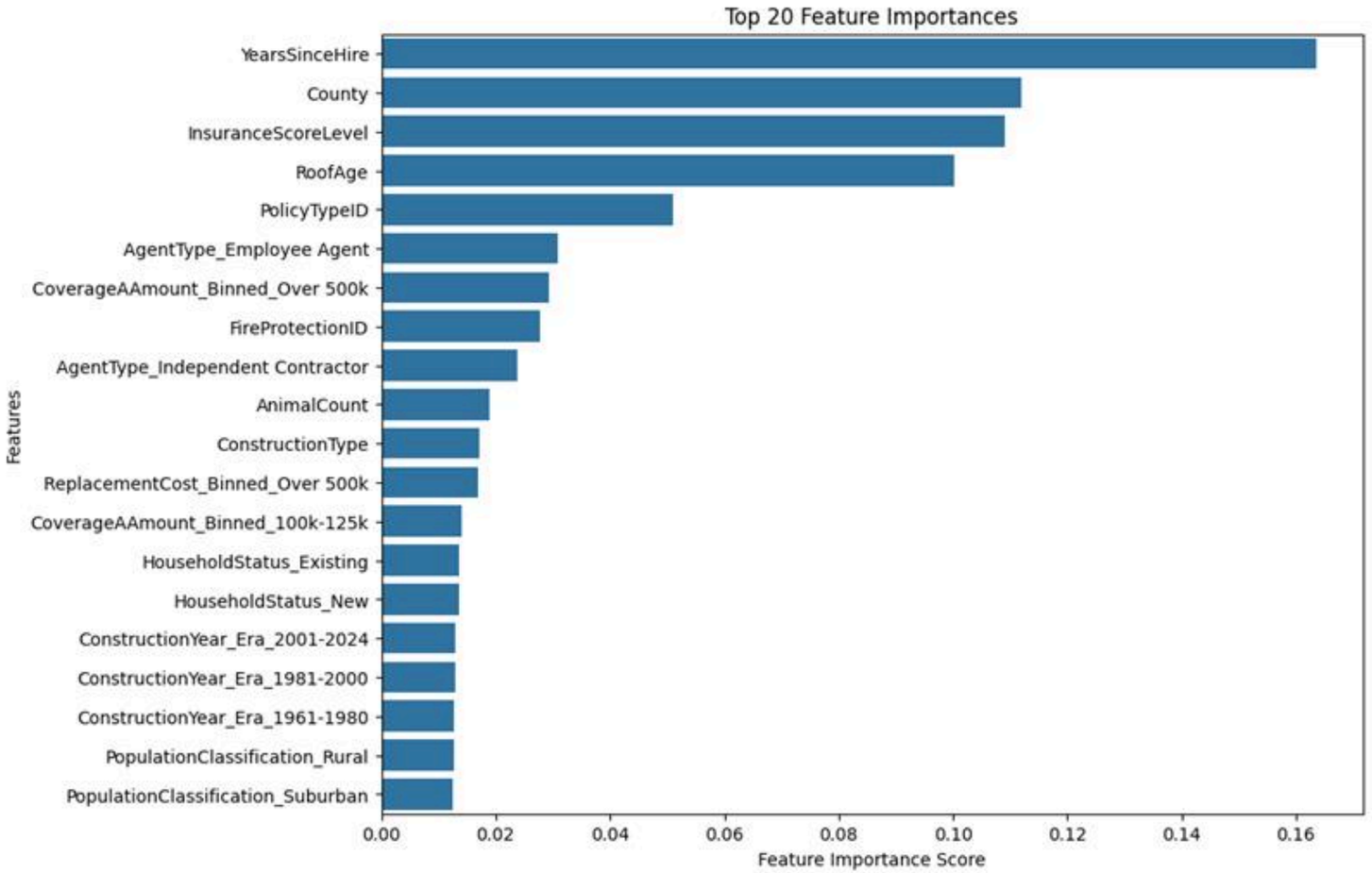
MODEL SELECTION

Models with Best Result:

- Identified **RandomForest** and **XGBoost** as the top-performing models

Important Features Identified:

- Highlighted features that contribute most to detect high-risk houses



PURDUE UNIVERSITY | Mitchell E. Daniels, Jr. School of Business

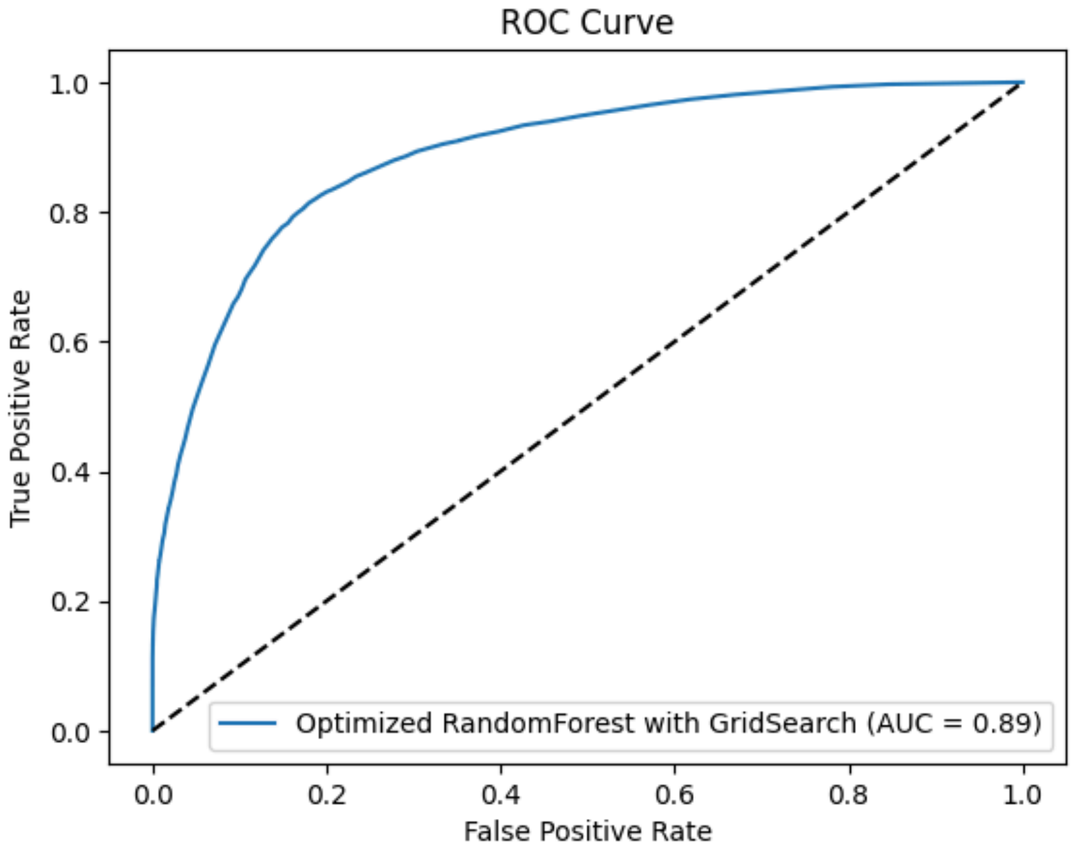
MODEL BUILDING

MODEL TRAINING AND EVALUATION

Data Splitting: Train (70%) test (30%) split ratio allows for a sufficient amount of data to train the models while still reserving a sizable portion for independent evaluation

Cross-validation and Hyper-parameter Tuning: Employed K-Fold cross-validation (5 folds) techniques to ensure model robustness
Hyperparameters tuned to optimize performance

Precision	Recall	F1 Score
0.74	0.58	0.65



BENEFITS

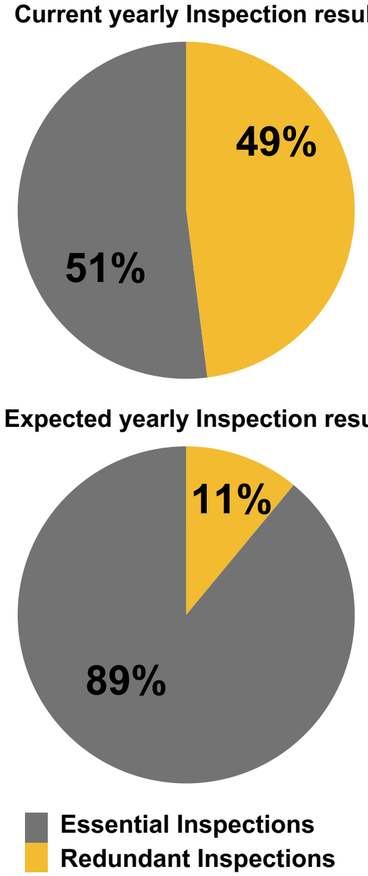

Reduced Risk
Increased Accuracy
Lower Costs
Improved Efficiency
Higher Customer Satisfaction
BENEFITS

DEPLOYMENT

Current	# of redundant inspections (Yearly Average) 1324
Updated	# of redundant inspections (Yearly Average) 270

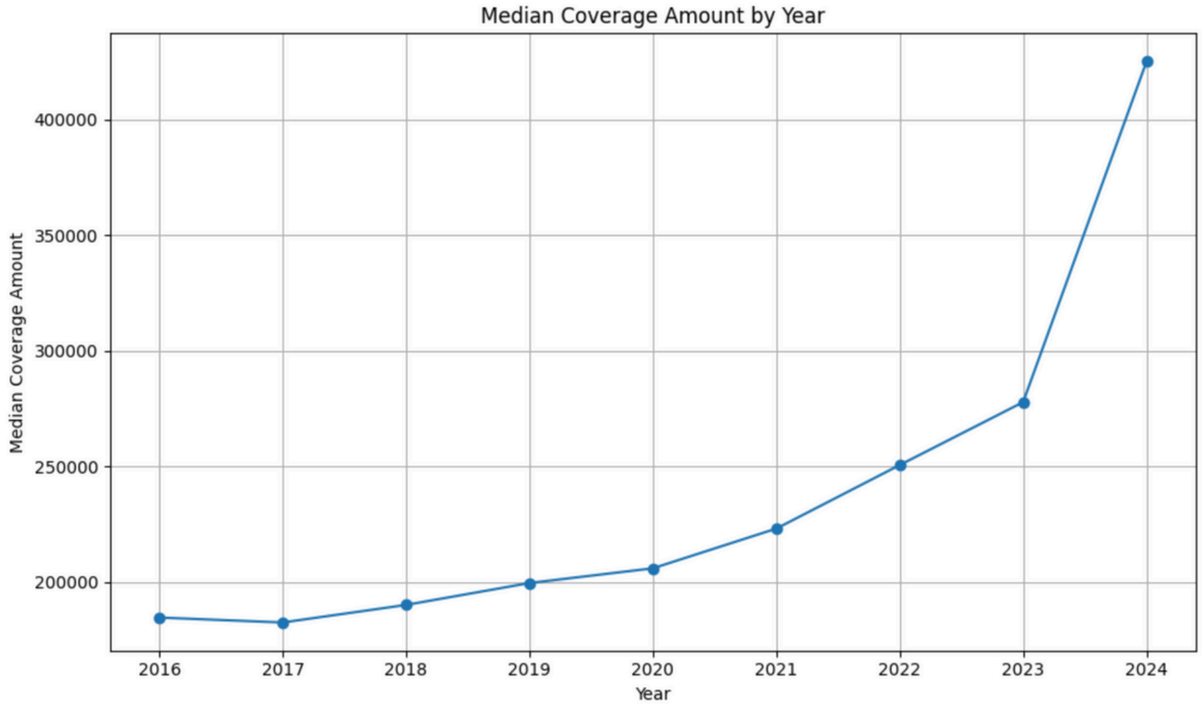
Average Cost Per Inspection
~$400

Estimated Average Savings
421K/Year



FUTURE SCOPE

1. Adjust model for market instability and Inflation



FUTURE SCOPE

1. Adjust model for market instability and Inflation

2. Continuous Improvement/Continuous Deployment based on tests conducted on updated data

THANK YOU

We would like to thank our industry partner, for your guidance and support on this project as well as Purdue MS-BAIM program faculty mentors.

PURDUE UNIVERSITY | Mitchell E. Daniels, Jr. School of Business

APPENDICES

MODEL EVALUATION

AGGREGATED

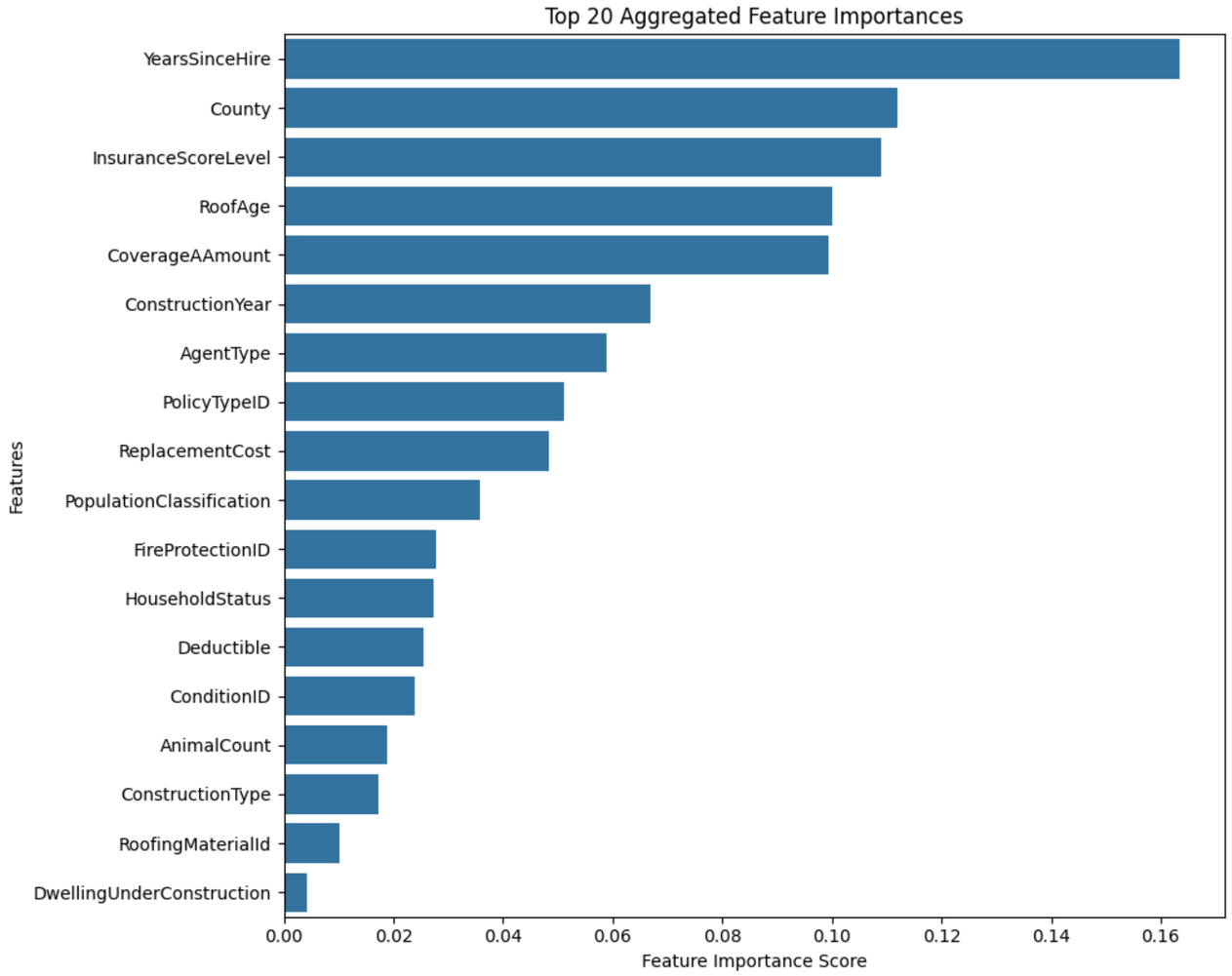
MODEL SELECTION

Models with Best Result:

- Identified **RandomForest** and **XGBoost** as the top-performing models

Important Features Identified:

- Highlighted features that contribute most to detect high-risk houses



COST ANALYSIS

Inspection column

20:00 ⟵ inspection time

Redundant Inspection

Inspection was filled
ChangeTypeID show no change (0)

23661 cases

current (total) | updated (total)

11914 **2430**

Total inspection*model accuracy(0.897)