Annuity Suitability Peer Review Session 2

Enterprise Analytics Office

Shruti Pandey
Julia Chen

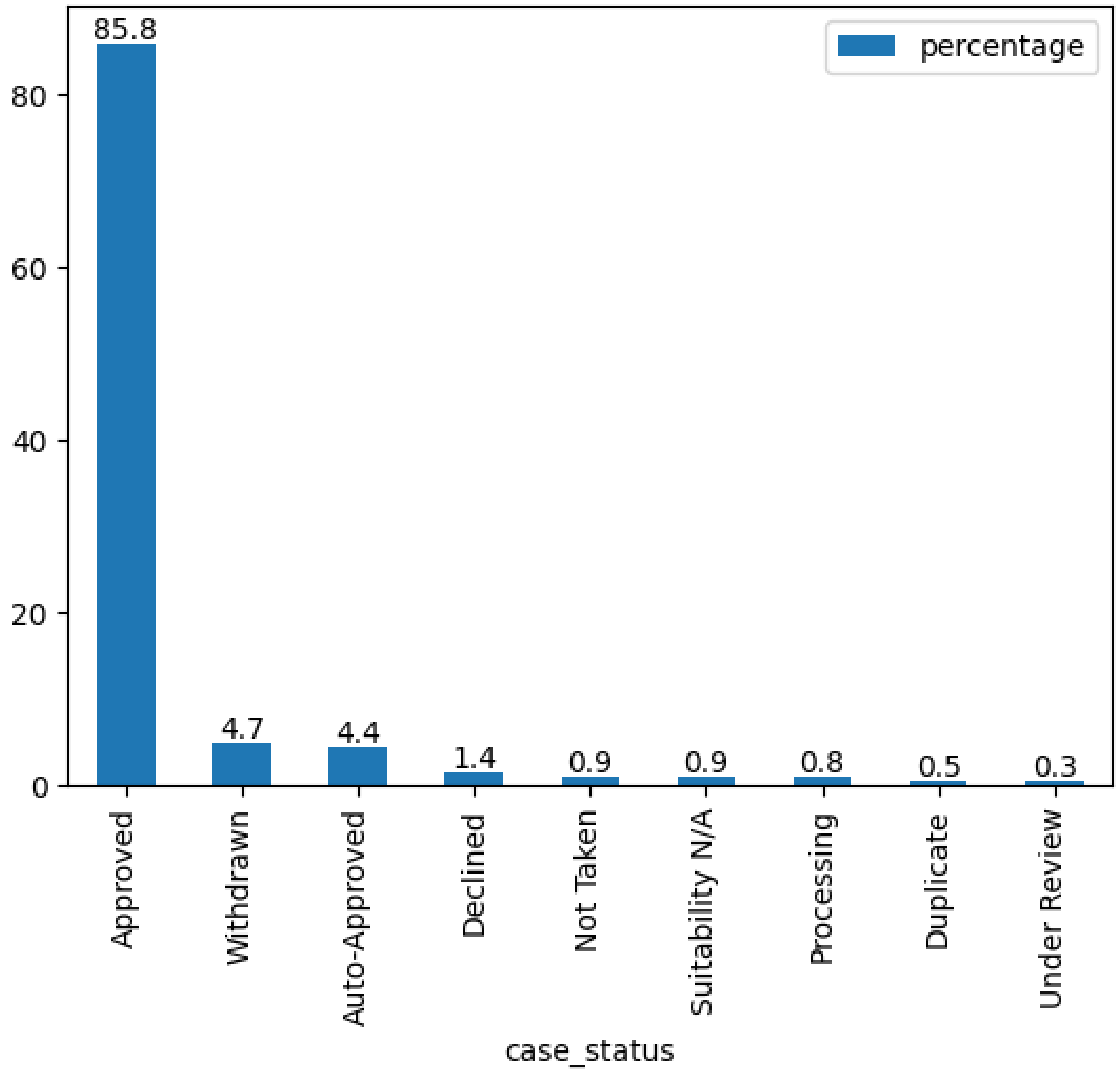
05/09/2024

Nationwide®

Table of Contents

Introduction

- Annuity sales are approved more than 90% of the time, there are no auto-declines, and the process of suitability decision is different for different applicants
  - The auto-approved cases in the graph are special cases in place in Q1’23 when the application volume was high
- Manual involvement of principals in evaluating and approving an application
- The risk assessment method is rules based (OpenL) and require a more adaptable method to evaluate risk associated with each application



Business Problem and Model Goals

Problem Statement

Annuity sales are approved 98% of the time, there are no auto-declines, and the process of suitability decision is for different applicants. These factors cause friction for financial professionals and negatively impact customer experience.

Benefits for annuities

- Automation of suitability decision with a target of 50% by the end of 2024
- Better customer experience by ensuring good financial decision for the customer, consistent application process and automation of suitability decision
- Reduce friction for financial professionals and help them make a data informed decision on an application with automation, past data analysis and third-party data on financial health of applicant

Financial Benefits

- Expense saving will be around $225,000 at 50% automation level
- Reduction by ~7000 hours of work by suitability principals at $32 an hour, saving $225,000
- Reduce the cost of incorrect decision

Session 1 Review

Nationwide®

Overview

In session 1, following topics were covered:

- Business problem and model goals
- Metrics and regulatory considerations for the model
- Model in production
- Data description
- Data preprocessing

Feedback and follow up questions

<u>Auto Approval Rate</u>

- Increasing the auto-approval rate over time?
  - The plan is to keep the approval rate to 50% to account for cases that would necessitate a manual review or require further communication with the applicant to obtain key information

- Approval in 98% cases but 60-65% easy cases, how to reconcile?
  - As answered above, all the cases cannot be automated because even for easier cases, there are several cases that require further communication when either key information is missing, or the information provided is insufficient to reach a decision

<u>Refining suitability criteria</u>

- Will model refine the guidelines of determining suitability?
  - Currently, the target is to create a model that can auto-approve based on current criteria. As an extension of this project, the suitability criteria will be revisited to update the decision rule

Feedback and follow up questions

Declined rate over the years

- 2020: 2.2% cases declined
- 2021: 2.48% cases declined
- 2022: 1.345% cases declined
- 2023: 1.44% cases declined

Using current approved and declined cases as ground truth

- Business partners are confident with the business process and the criteria used to determine suitability. Although the rules are used to determine the final decision, there is still a lot of subjectivity involved from principals which is better suited to be modeled using machine learning

Data Review

Nationwide®

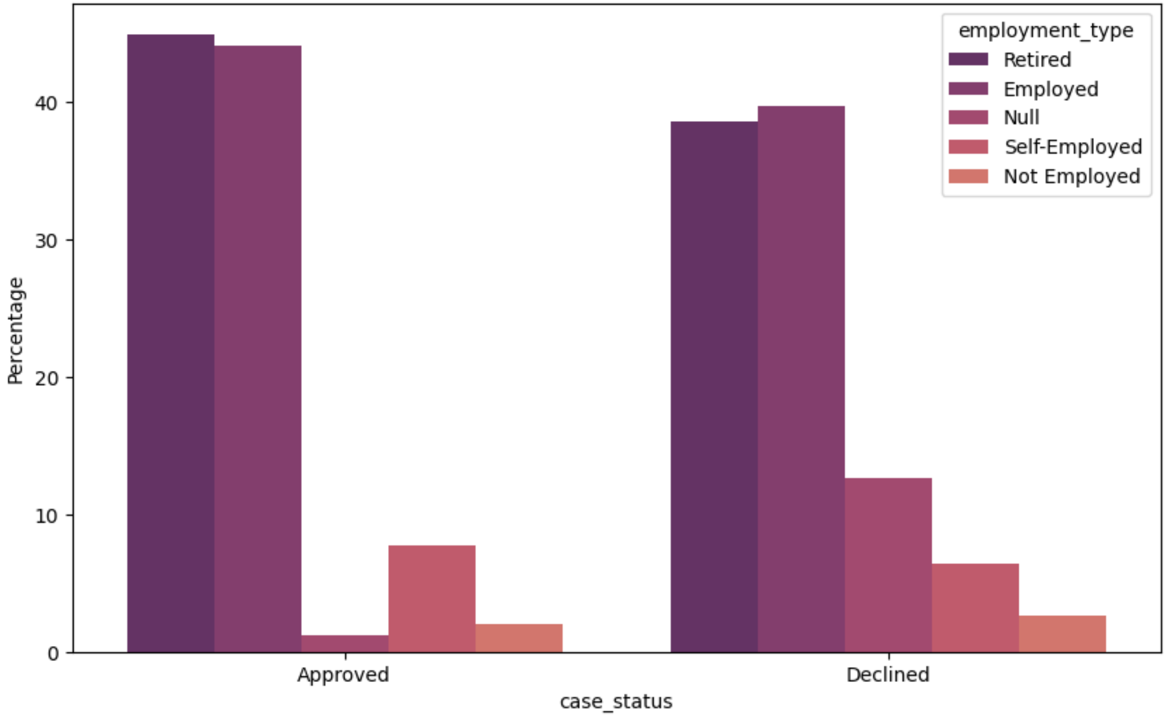
Data Description

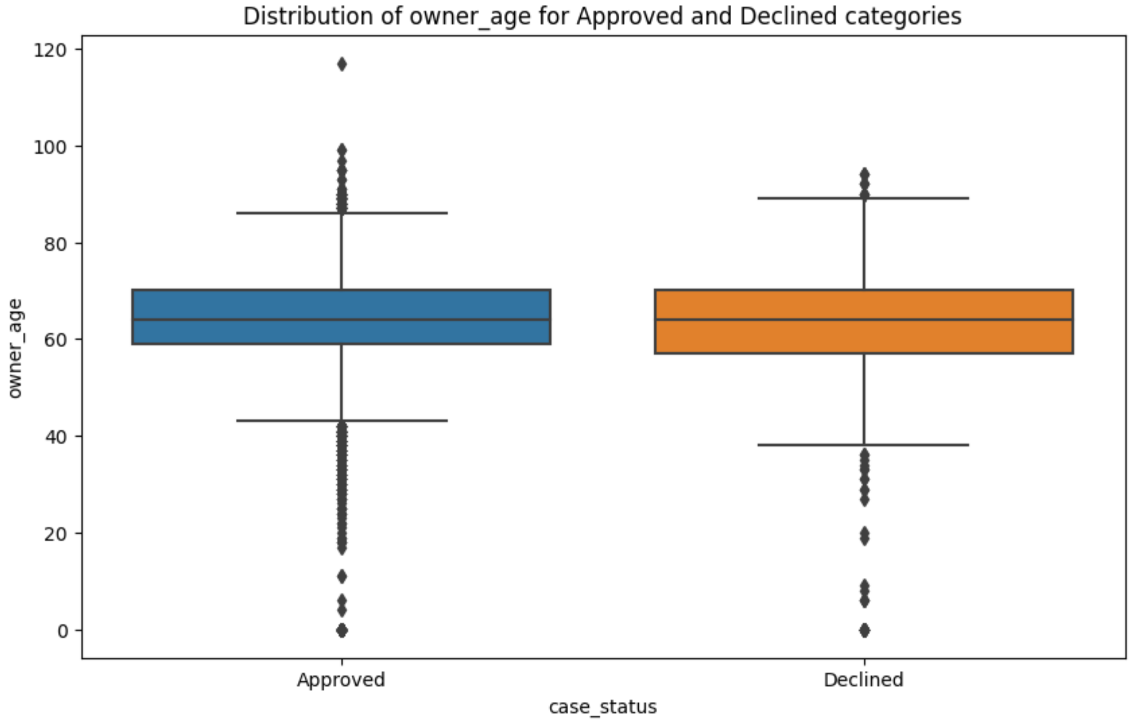
The data points in the combined dataset fall under these categories:

- Individual’s demographic characteristics
  - Age, Employment, etc.
- Individual’s financial history
  - Income, expenses, liquid assets, net worth, etc.
- Product related questions
  - Product group, product name, surrender period, etc.
- Joint applicant information
  - Rider, living benefit, etc.
- Anticipated use of annuity
  - Distribution type

Exploratory Data Analysis

The distribution of target variable is imbalanced and the predictive value of predictors in the data is low individually. More details in the GitHub repository: Link





Data Preprocessing

- Cleaned the data steps to fix data quality issues like inconsistencies, inaccuracies, incompleteness, ambiguity, and human error. Some examples:
  - Negative age
  - Zero expenses
- While in production, such data points would be tagged for further review
- Data Imputation of missing values after consultation with business partners. This is done for the key predictors, and they have <2% missing values usually
  - Mean in case of continuous variable
  - Mode in case of categorical variable
- Dropped the features
  - That have more than 80% missing values and aren't key predictor
  - That have only one unique value (e.g. all true) for a predictor variable
- Outcome Variable
  - Binary values: Approved or Declined
  - Imbalanced data, addressed in model collab

Feature Engineering

- Encoded the categorical variables
  - Binary encoding performed on binary categorical variables (usually True and False values)
  - One hot encoding performed on variables with more than 2 unique values.

- New features were created by combining two or more features
  - Premium amount / Net Worth
  - Premium amount / Liquid Assets
  - Premium amount / (Income – Expenses)

Method Collab

Nationwide®

Data and Model Overview

<u>Training and Validation Data:</u>

- The model will be trained using approved and declined cases from April 1st to October 31st, 2023.
- Data split was done using train_test_split() with function with the stratify method to ensure balanced representation of classes.
- The split resulted in a 70-30 ratio for training and validation data respectively.

<u>Testing Data:</u>

- Testing set of data is from Sep 1st, 2023, to Apr 23rd, 2024.
- Considering some updates after October 31st, cases associated with watch list producers after then will be managed separately, excluding them from the testing set and future production.

<u>Model:</u>

- Two models are under consideration based on replacement questions, with the current model focusing on replacement question "False" only.
- Model training utilizes StratifiedKFold to preserves the percentage of the sample in each class.

Predictor Variable Selection

1. Shap Values from baseline model
2. Chi-Square Test of Independence:
   - A hypothesis test that determines whether two categorical variables are related.
3. Variance Inflation Factor:
   - Tree-based models are capable of handling complex relationships among variables and not affected by multicollinearity. But it till important to understand the distinction between then and why multicollinearity might not be a concern for decision trees.

Class Imbalance

Data splitting involved stratified sampling to maintain proportional representation.

Model training using StratifiedKFold, the class label in each fold is proportional and preserves the percentage of the sample in each class

Method 1: Oversampling with SMOTE-NC

- A variant of SMOTE designed for datasets containing numerical and categorical features.
- Finds K nearest neighbors from minority samples to overcome overfitting issues posed by random oversampling.

Method 2: Tuned weight hyperparameters

- In Random Forest model set class_weight as [total_negative_examples: total_positive_examples]
- In XG Boost model set scale_pos_weight as total_negative_examples / total_positive_examples

Model Technique

<u>Model</u>

Tree based classification model

- Random Forest
- XG boost

Anomaly detection Model

- IsolationForest

<u>Model Selection</u>

- **Recall**: True Positive / (True Positive + False Negative)
- **Precision**: True Positive / (True Positive + False Positive)
- **Recall score from the minority group**: Focus on the recall score from the minority group to ensure the model performs well in detecting cases of interest.
- **AUC**: Measures the ability of the model to distinguish between the positive and negative classes.

Attached spread contains all the results, including metrics such as recall, precision, recall score from the minority group, and AUC.

Next Steps

1. Finalized the features selection and DPD
2. Set the performance thresholds and finalized model result.

Reference

- SMOTE: Synthetic Minority Over-sampling Technique, https://arxiv.org/abs/1106.1813
- The Chi-square test of independence. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900058/

Nationwide®
is on your side