

# Redefining Claim Severity: A Data-Driven Approach to Measuring the True Impact of Vehicle Accidents



Artak Kamalyan, Lead Deep Learning Scientist  
Natali Gzraryan, Deep Learning Scientist

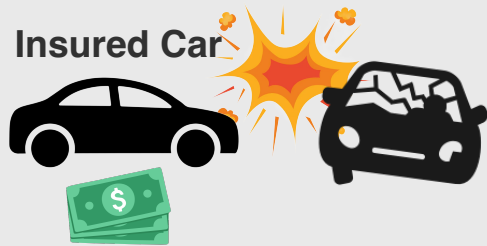


- Working at Plat.ai — a U.S.-based Data Science company operating in Armenia
- Main project: Premium Prediction Model for a leading Armenian insurance company
- Today, we will present a key component of the personalized system we built for the company.

## Types of Car Insurance

### Collision insurance

Covers damage to your own car



Insurance pays for the insured (guilty) car

### Liability Insurance

Covers damage you cause to other people

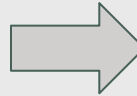


Insurance pays for the damaged car

# Data Science in Insurance: Features

## Data Used

- Driver history and accident records
- Age and driving experience
- Vehicle make and model
- Claim history
- Location and driving patterns



## Features

Number of claims occurred in last contract

Age group (18-26,26-40,40-70,70+)

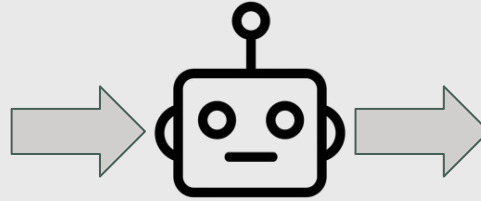
Target Encoded Mark,Model

...

etc

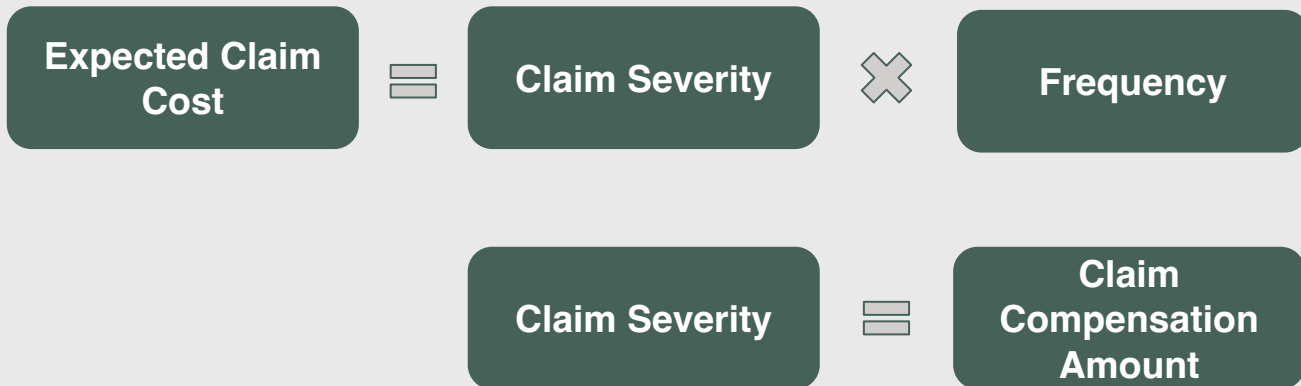
# Data Science in Insurance

Features
Number of claims occurred in last contract
Age group (18-26,26-40,40-70,70+)
Target Encoded Mark,Model
...
etc



Model Output
<ul style="list-style-type: none"><li>• Probability of accident</li><li>• Expected claim cost</li><li>• Frequency</li></ul>

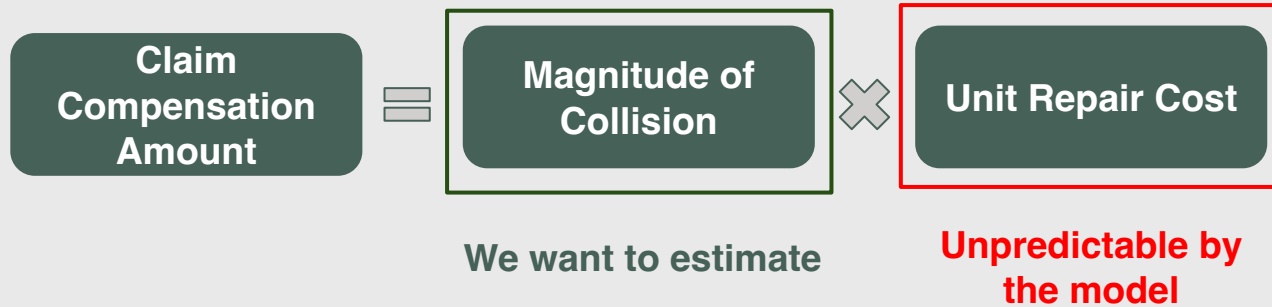
# Traditional Approach: Target



- Strong dependence on vehicle price and repair cost rather than true accident impact
- Poor comparability between different vehicle categories and brands
- Highly skewed target with extreme outliers causing unstable model behavior

Claim Severity is the key component we aim to address.

## Understanding Claim Compensation



- Repair costs are primarily determined by the characteristics of the damaged vehicle.
- The magnitude of the damage, however, is largely influenced by the driving behavior and skill of the insured driver.

# Traditional Claim Severity

Traditional claim severity = *Compensation Amount*  
Fails to reflect *true damage impact*



**Claim amount 5,000 €**



**Claim amount 5,000 €**

Identical payouts can imply *very different damage levels*

# Problem Definition

**We propose a normalized claim severity metric that normalizes compensation amounts relative to the expected repair-cost level of specific vehicle**

- The proposed severity aims to reflect relative accident impact and driver behavior rather than the economic value of the vehicle itself
- This transformation reduces vehicle-specific cost bias and creates a more standardized and comparable severity representation across heterogeneous vehicle categories

# BrainStorming

Normalized  
Claim Severity

=

Repair Amount

Average Compensation of the car group

1. Scrape all prices for car models  
**Problem: The challenge is that while a car itself may be affordable, its spare parts might be expensive.**
2. Scrape spare parts prices  
**Problem: Not all car models' parts prices are available**
3. We have claim data where we can see what was the compensation amount for different car models  
**Problem: Perfect Idea let's develop it**

# What We Have

Accident Date	Contract Id	Mark/Model	Claim Compensation
05/05/2024	1	Toyota Camry	2,300 €
03/04/2025	2	BMW X5	8,700 €
04/04/2026	3	Mercedes C class	5,400 €
05/06/2025	4	Honda Civic	1,700 €
05/06/2025	5	Mazda 3	4,200 €
04/12/2024	6	Avatr 07	6,000 €

Example data with a limited number of columns for demonstration purposes only.



Accident Compensation: 500 €



Average compensation amount for Toyota Corolla vehicles that were 10-years-old at the time of the accident: €1,000

Claim Compensation	Normalized Severity
500 €	$500/1000 = 0.5$



Accident Compensation: 2000 €



Average compensation in our dataset for 3-years-old Mercedes G-Class (G-Wagen) vehicles at the time of the accident: €4000

Claim Compensation	Normalized Severity
2000 €	$2000/4000 = 0.5$

# Severity Custom Approach

Step 1

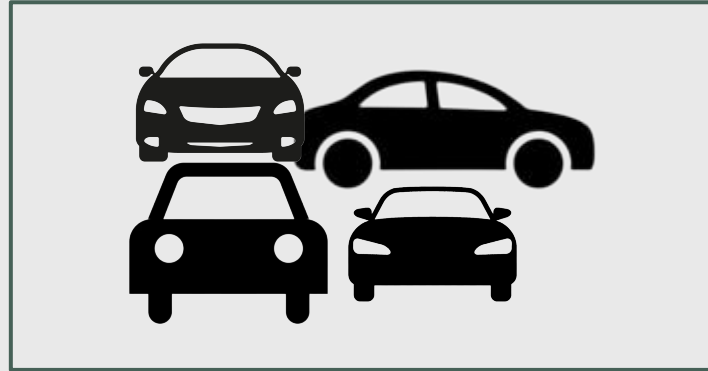


**Mark:** BMW  
**Model:** 7 Series  
**Age:** 3 years old

Cluster vehicles based on their claim severity using features like make, model, body type, and age. Estimate *average compensation* per group

Group 1

Group 2

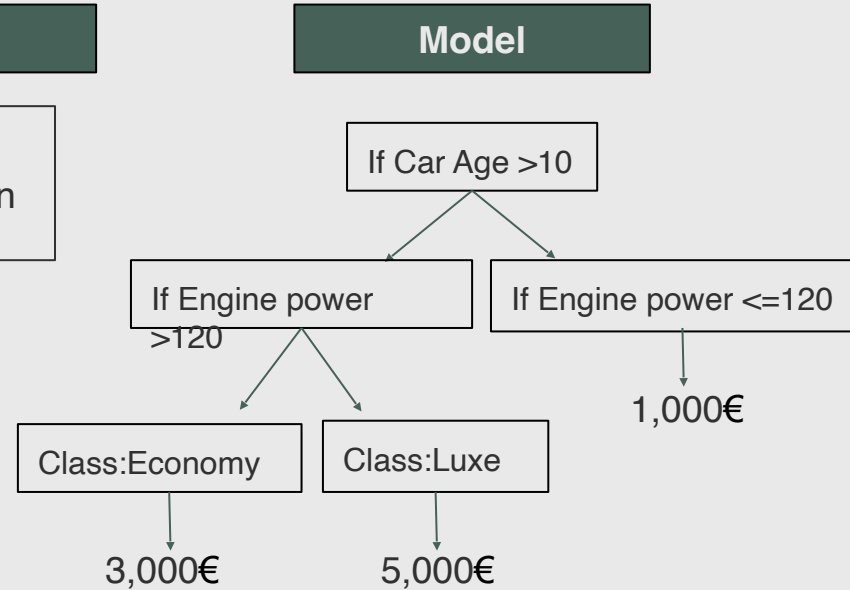


# Grouping: Decision Tree

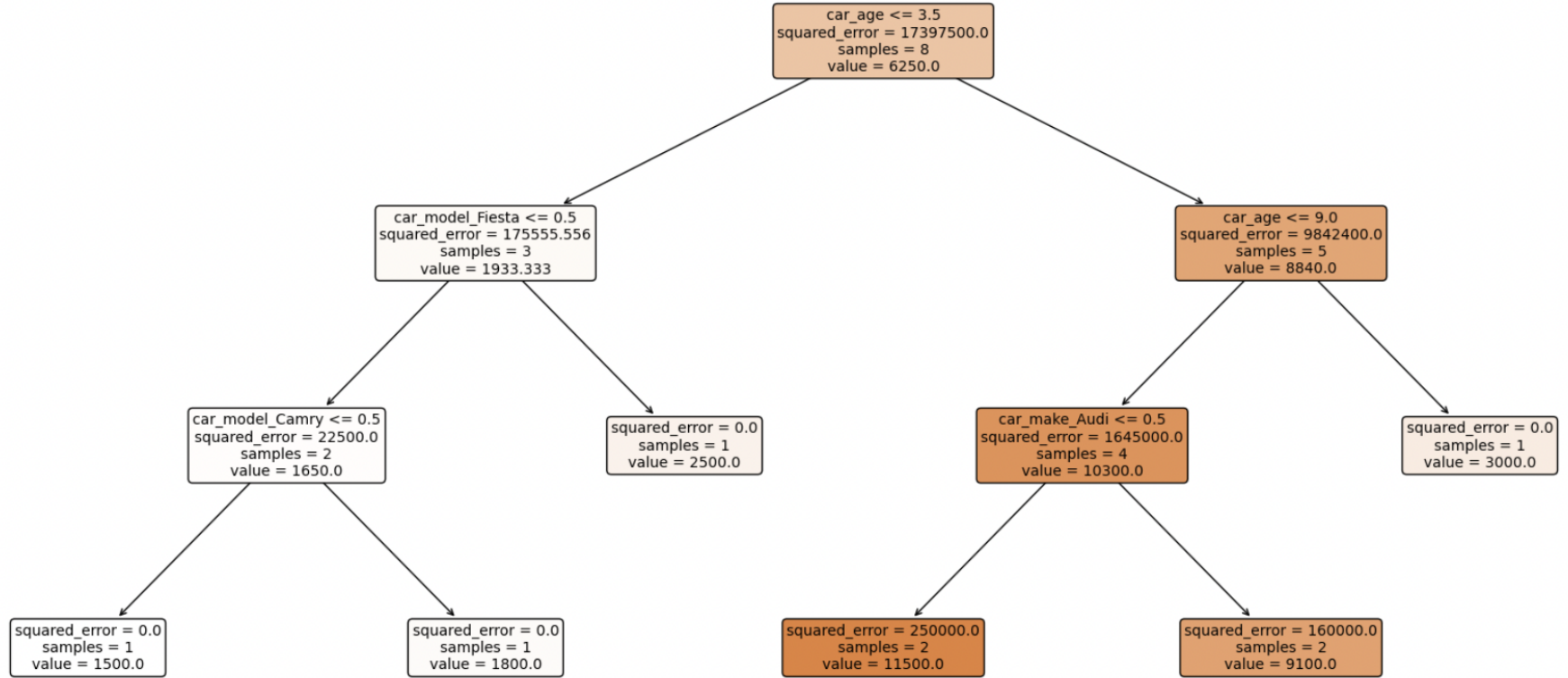
Given information about vehicles, we built a decision tree model targeting claim compensation in order to group cars with similar expected repair-cost levels.

Features
Mark
Model
Body Type
Age
Class ( Economy,Luxe, etc.)
Engine Power

Target
Claim Compensation Amount

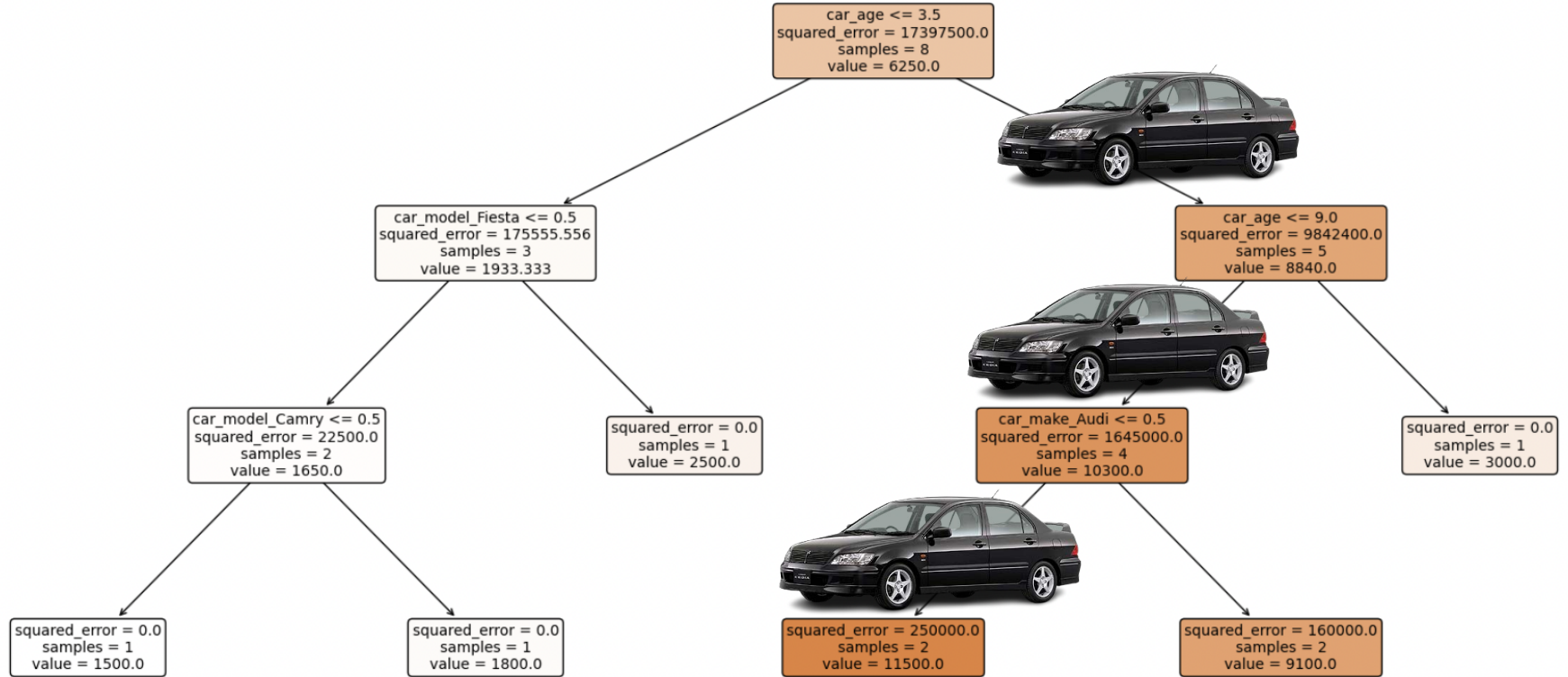


# Severity Custom Approach



This is not the real trained decision tree.

# Severity Custom Approach

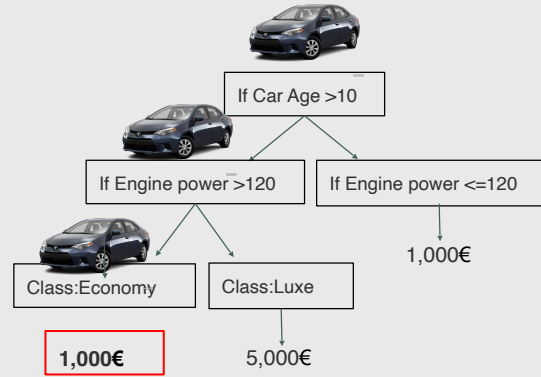


This is not the real trained decision tree.

# Summary



Accident Compensation: 500 €



**Normalized  
Severity**

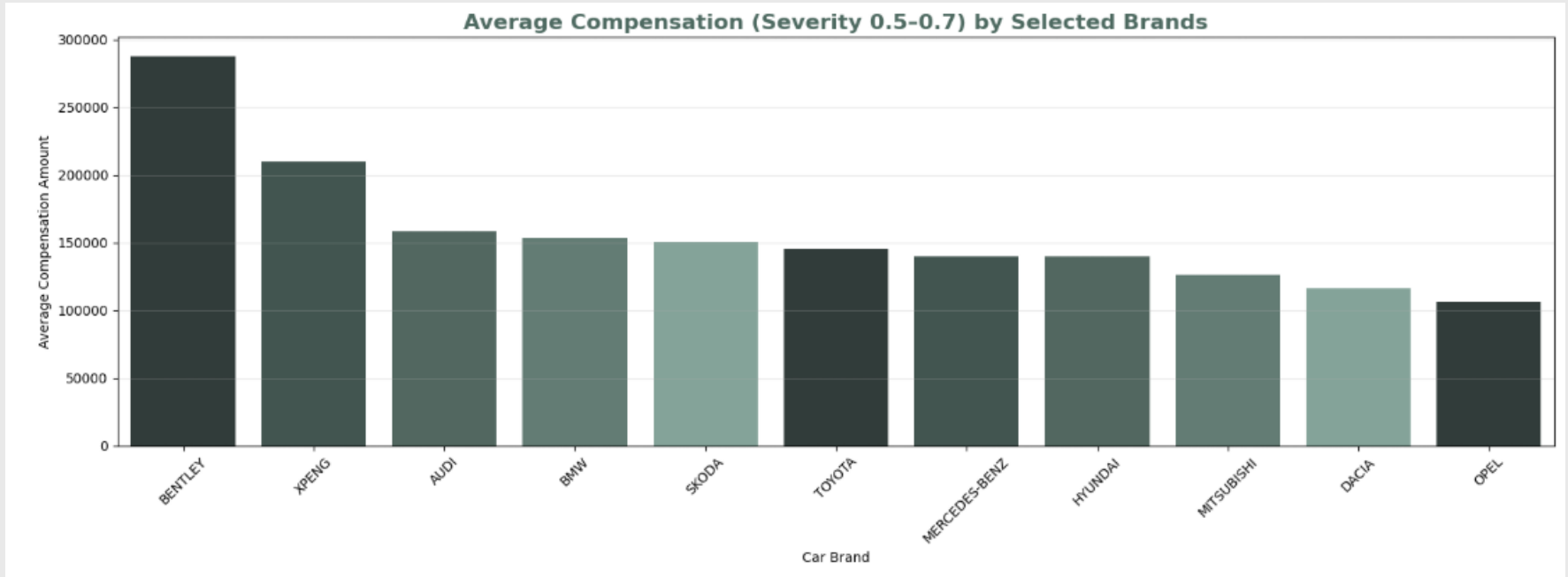
$$500/1000 = 0.5$$

# Compensation vs Severity

Mark	Model	compensation_amount	severity
TOYOTA	AVENSIS 2.0	20000.0	0.598372
IVECO	NaN	20000.0	0.671459
OPEL	ASTRA 1.8	20000.0	0.631725
TOYOTA	CAMRY 2.4	20000.0	0.588129
VAZ	21063	20000.0	0.502361
TOYOTA	VITZ 1.0	20000.0	0.593892
VAZ	21102	20000.0	0.558553
CADILLAC	NaN	475000.0	0.677097
MERCEDES-BENZ	GLE 43AMG 4MATIC	480000.0	0.684224
MERCEDES-BENZ	G 63AMG	480000.0	0.684224
BMW	X6 XDRIVE 40I	480000.0	0.684224
MERCEDES-BENZ	GLE 43AMG 4MATIC	481160.0	0.685878
LEXUS	LX 570 5.7P	483000.0	0.688501

For the same severity level, compensation amounts vary significantly — from approximately €46 to €1,100(20,000 AMD to 483,000 AMD).

# Severity by Car Brands



## Severity Usage

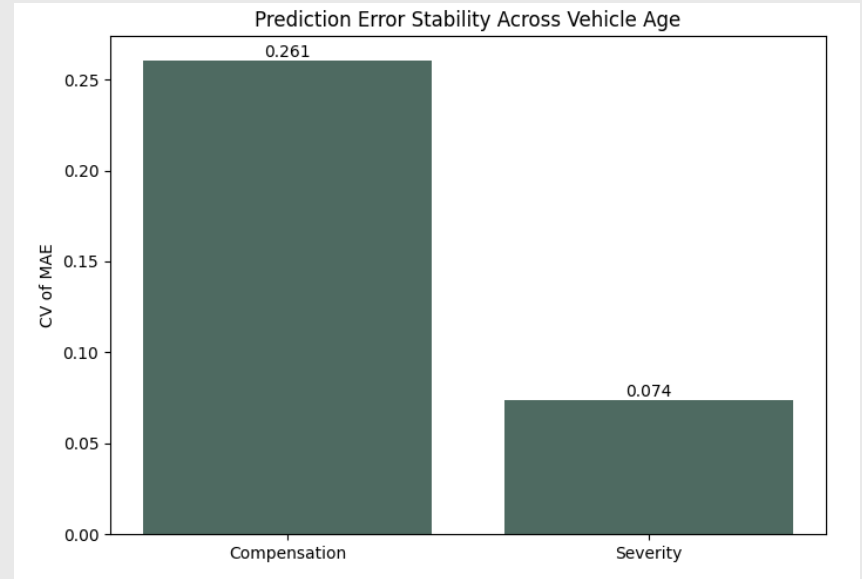
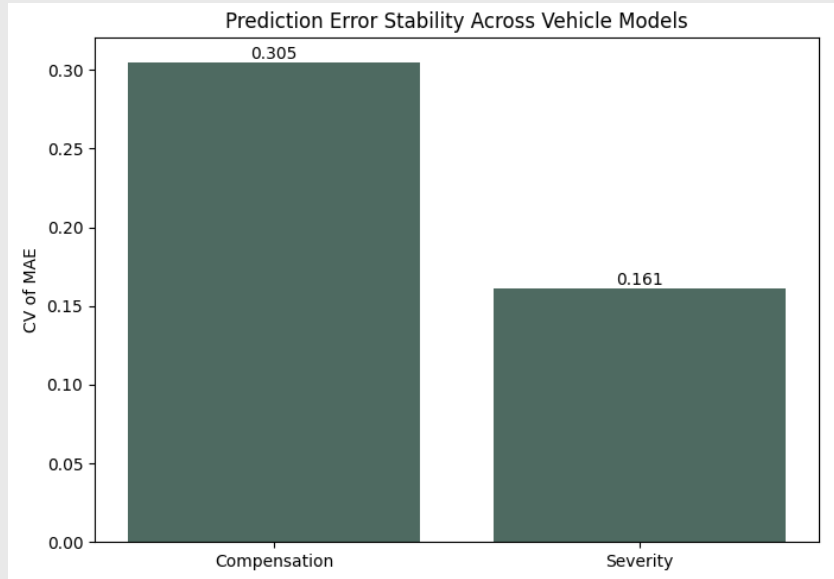
### Severity as Target

We can target severity besides the frequency and compensation as more standard approaches.

### Severity as Feature

We can use historical severity of the driver to better describe his/her behavior

# Severity: Target



$$CV = \frac{\text{Std}(MAE)}{\text{Mean}(MAE)}$$

Prediction errors were calculated separately for each vehicle model/age group. The variability of these errors was measured using the Coefficient of Variation (CV). Lower CV indicates greater stability and lower dependence on vehicle-specific characteristics.

## Severity Usage

### Severity as Target

We can target severity besides the frequency and compensation as more standard approaches.

### Severity as Feature

We can use historical severity of the driver to better describe his/her behavior

# Severity: Feature

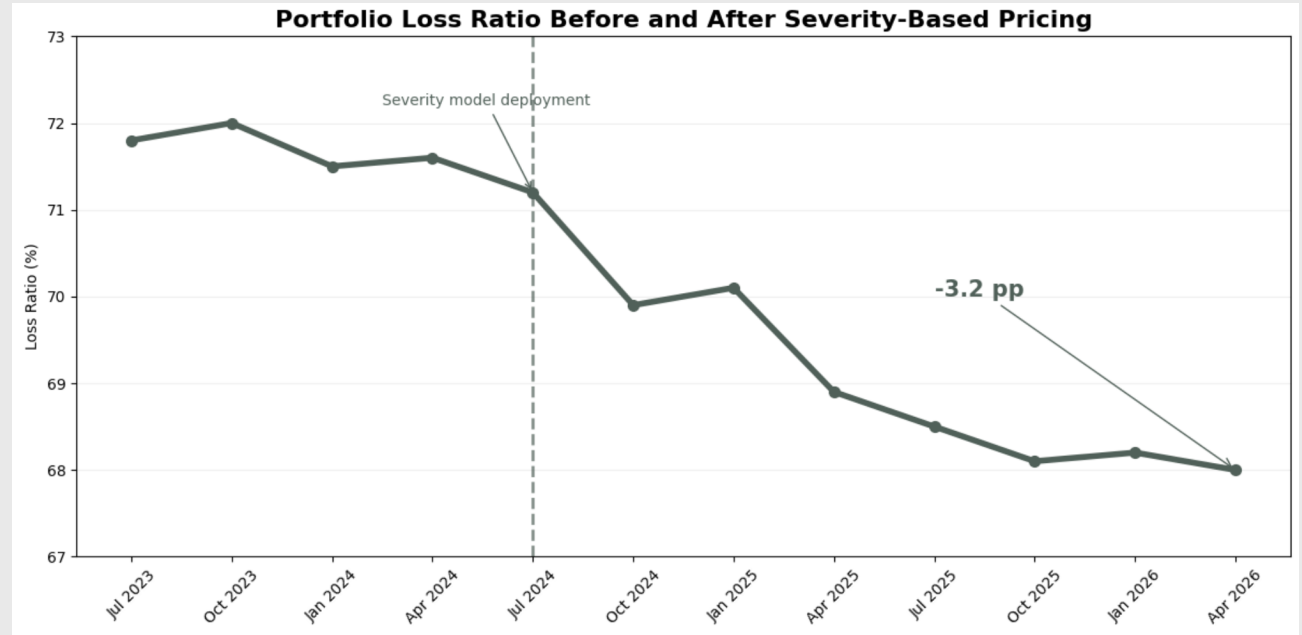
Feature name	Feature Description
Latest Vehicle Claim Severity	Measures the severity level of the most recent contract identified by the VIN. Higher values indicate more costly or serious incidents.
Driver Average Accident Severity	Represents the average severity of accidents involving the driver across their historical records. Higher scores reflect a history of more severe accidents or claims.

AUC without new features	0.75
AUC with new features	0.81

New features are in top 5 important features according to catboost feature importance.

# Results: Loss Ratio

Based on observed portfolio performance, our approach contributed to an estimated 3% reduction in loss ratio.



Thank You