

HOW WE TOOK A LEAP INTO MACHINE LEARNING AS ACTUARIES

STACKING GLM

November 24, 2021

Guy Carpenter Japan | Global Strategic Advisory

Suguru Fujita
Sara Sakimura

A business of Marsh McLennan

Introduction

About the team member – we are colleagues!



Suguru Fujita, FIAJ, ASA, CERA

- Actuary
- Life (3yr) -> Non-Life (1.5yr) -> Reinsurance (4yr)
- M.S./B.E. – Applied Mathematics



Sara Sakimura

- Actuary
- Accounting (1yr) -> Reinsurance (3yr)
- BSc. – Applied Statistics



Kotaro Kuge

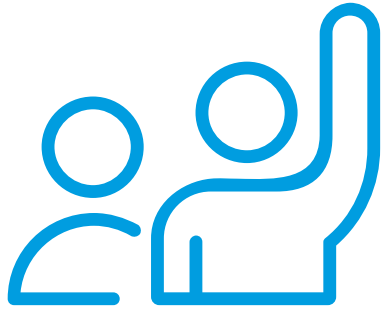
- Catastrophe Risk Engineer
- Non-Life (7yr) -> Reinsurance (4.5yr)
- ME – Structural Reliability Engineering

Guy Carpenter Japan, Inc.

- Reinsurance Broking
- Reinsurance Pricing
- Natural Catastrophe Risk Analysis



We always welcome new challenges to apply ML methodologies into our tasks!!!



First...Poll Time!



What We're Here to Talk About

1. Overview of the competition (*10mins*)
2. Our submitted model & Final results (*15mins*)
3. Our findings (*5mins*)



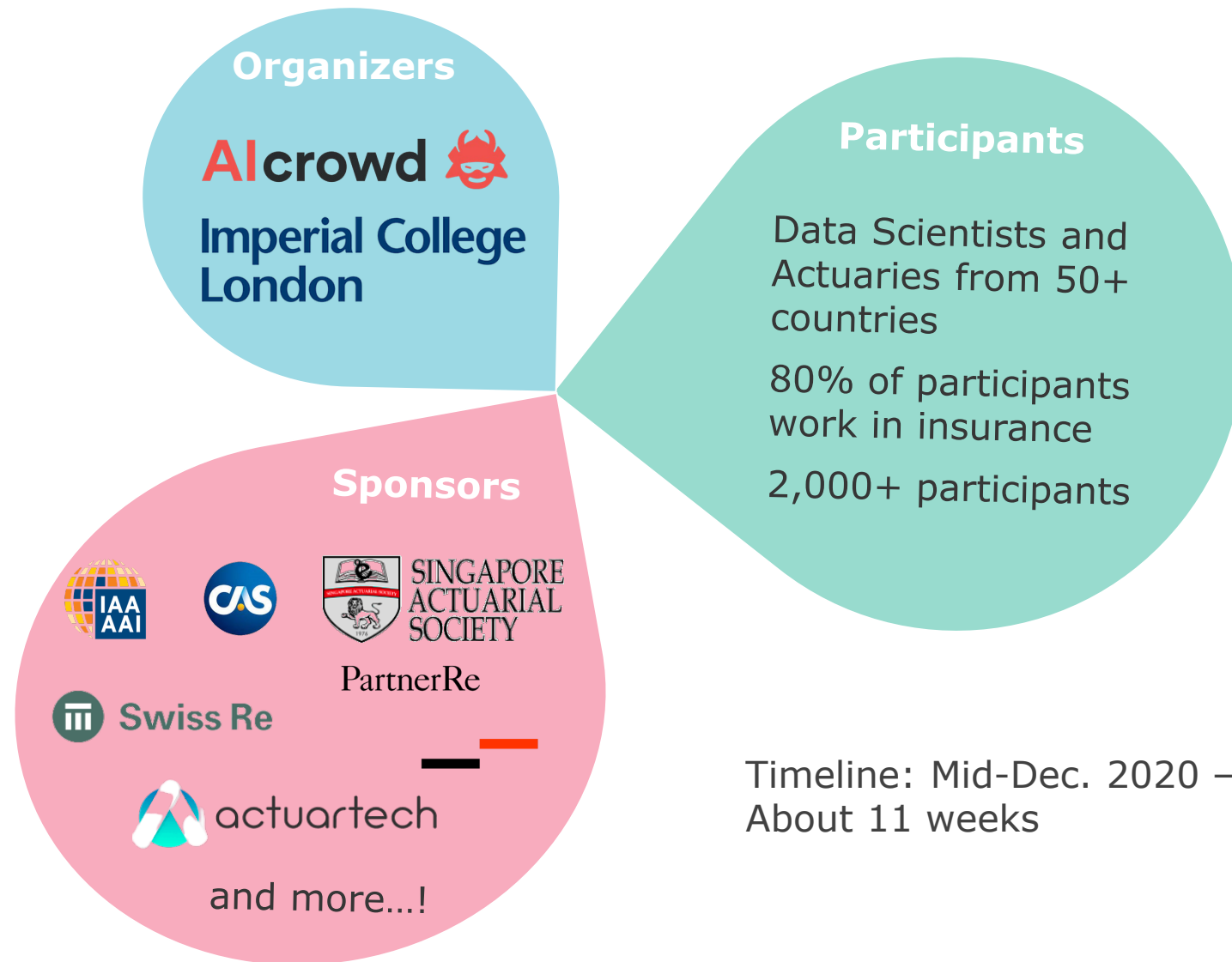
Insurance Pricing Game - Motor Insurance market simulation



<https://www.aicrowd.com/challenges/insurance-pricing-game>

Overview of the Competition

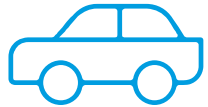
Competition Overview



Timeline: Mid-Dec. 2020 – Mid-Mar. 2021
About 11 weeks

Competition Rules and Challenges

RULES:



We act as a motor insurance company, where we build a pricing model and compete against other players for profit

- In this simulated market, customer always buys **the cheapest** insurance
- For policies you sell, you will earn premiums and you are also responsible for any claims that occur

EVALUATION METRICS:



Prediction Accuracy

- RMSE (Root Mean Squared Error) for claims
- Instant feedback



Profit Leaderboard

- Final Ranking is based on the profit
- Weekly feedback

Competition Rules and Challenges

Cheapest-wins-market Example

	Company 1	Company 2	Claim amount
Policy 1	120	90	100
Policy 2	30	25	10
Policy 3	10	15	0
Policy 4	5	10	0
Total revenue	15	115	
	0	110	Total loss
Total profit	15	5	
Prediction Accuracy Measurement (Lower the better)	15.2	12.7	

Diagram illustrating the Cheapest-wins-market Example. The table shows the predicted claim amounts for four policies across two companies, along with the actual claim amounts. The total revenue, total loss, total profit, and prediction accuracy measurement (lower is better) are calculated for each company. Company 1's total profit is highlighted as the 'Most profit' (15), and Company 2's prediction accuracy measurement is highlighted as the 'Most Accurate' (12.7).

- Having the most accurate prediction **DOES NOT** mean it can generate the most profit
- Ideally, we want to price as close to the actual claim amount as possible, but slightly above

Data and Feedback

Training Data: table format / 60k records



Policy Info

- Coverage
- Type
- Payment Frequency



Driver Info

- Age
- Sex
- Driving History Indicator



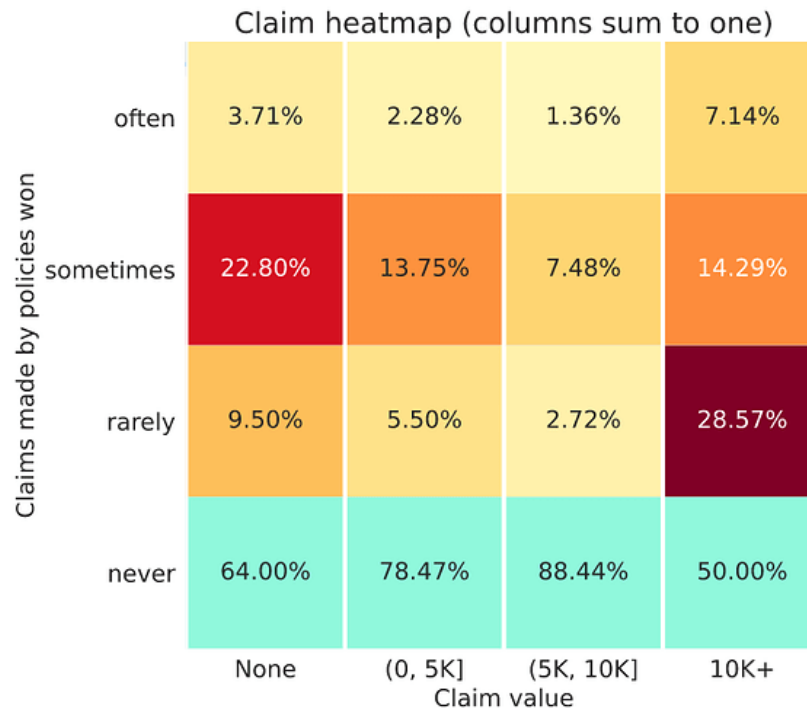
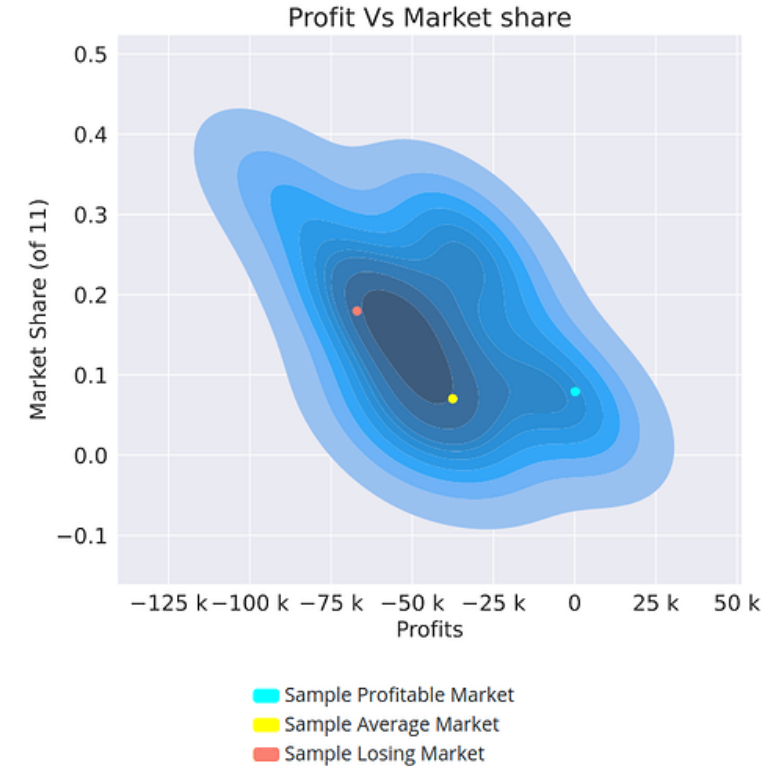
Vehicle Info

- Value
- Weight
- Max Speed



Location Info

- Population
- Town Surface Area



Leaderboard and Discussion Board

Discussion

Access to the test data?
By simon_coulombe 19 days ago

15 **Uncertainty quantification type models**
♥ By mikeoma 21 days ago

9 **Ideas for an eventual next edition**
♥ By simon_coulombe 21 days ago

1 **Townhall Recording**
♥ By aalok_devkota 20 days ago

24 **Sharing of industrial practice**
♥ By davidlkl 28 days ago

91 **Announcing a bonus round while you wait**
By alfarzan About 1 month ago

39 **[Final Week] Announcement timeline, duplicate accounts**
By alfarzan About 1 month ago

40 **You are probably wondering how much profit loading to select?**
♥ By michael_bordeleau About 1 month ago

2 **Time of town hall**
♥ By jocelyn 24 days ago

105 **It's (almost) over! sharing approaches**
♥ By simon_coulombe About 1 month ago

4 **A thought for people with frequency-severity models**
♥ By simon_coulombe 26 days ago

28 **Legal / ethical aspects and other obligation**
♥ By moto 27 days ago

13 **Who else got a call from Ali?**
♥ By simon_coulombe 28 days ago

19 **"Asymmetric" loss function?**
♥ By simon_coulombe About 1 month ago

6

3

2

11

10

16

15

4

47

5

15

5

16

Leaderboard

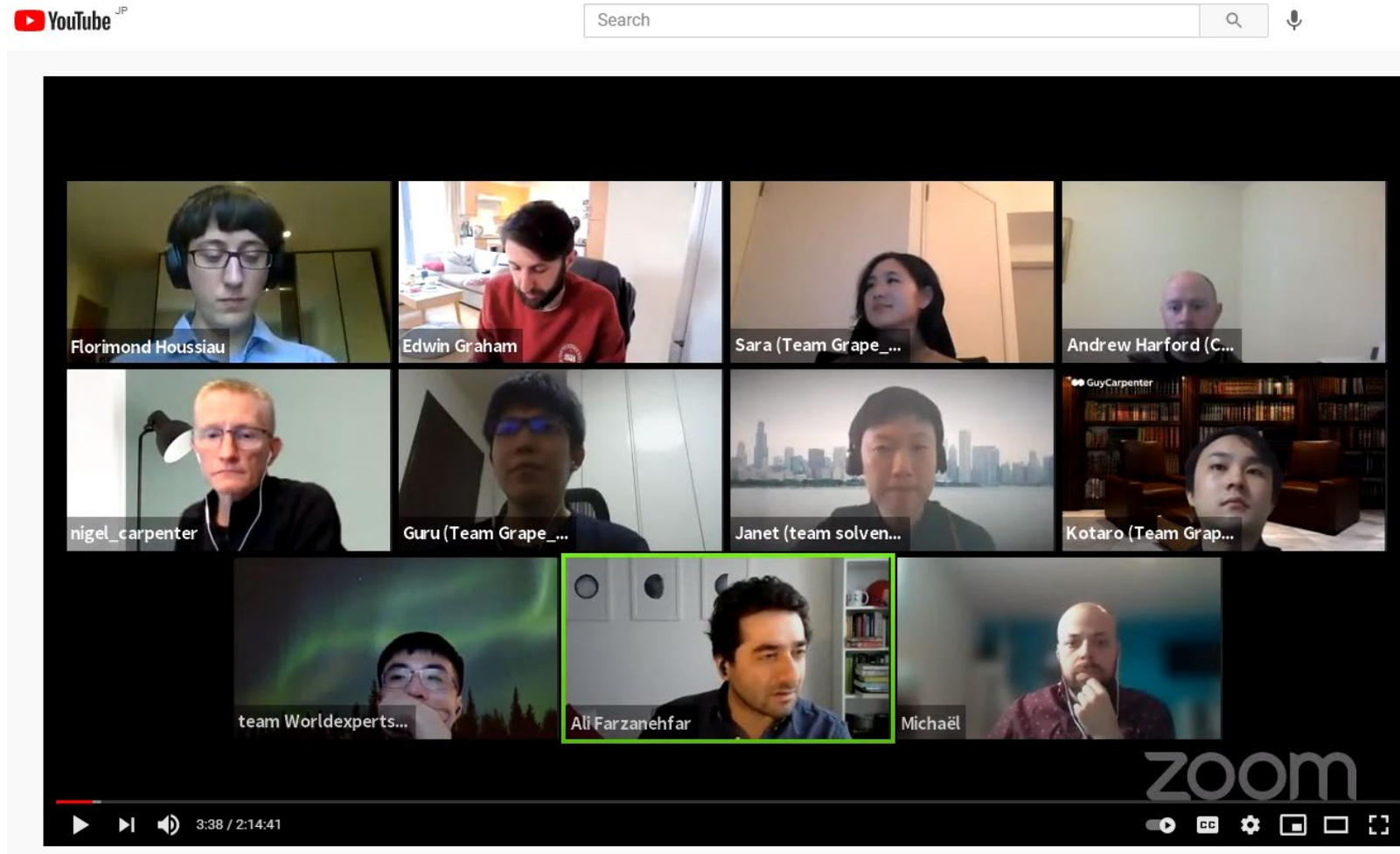
Profit

		Average Profit		Market Share		Submission Trend	
Δ	#	Participants					
01		jocelyn					View
02		solve_fx	185919.311	0.161			View
03		edwin_graham					View
04		glep	90108.871	0.190			View
05		moto	Yes				View
06		tom_snowdon	Yes	9932.590	0.094	1	View
07		Calico	Yes	9816.430	0.015	1	View

Accuracy

		RMSE		Entries		Submission Trend	
Δ	#	Participants					
01		Insolvent	499.168	542			View
02		NetSim	499.431	234			View
03		WorldExpe					View
04		solve_fx	499.442	319			View
05		michael_borde...					View
06		MCAF-VLC	499.505	63			View
07		jocelyn	499.544	31			View

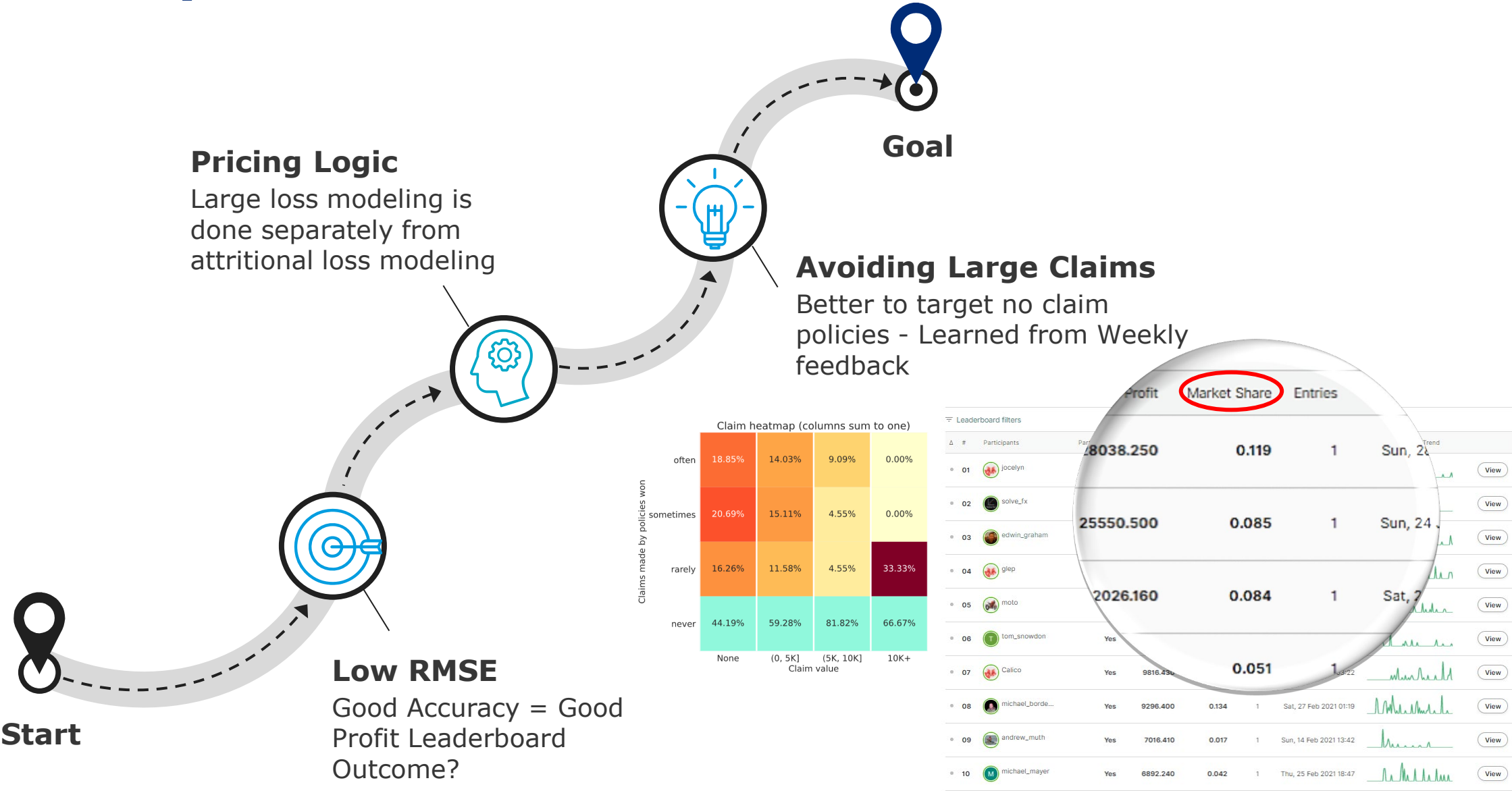
Townhall and Final Results



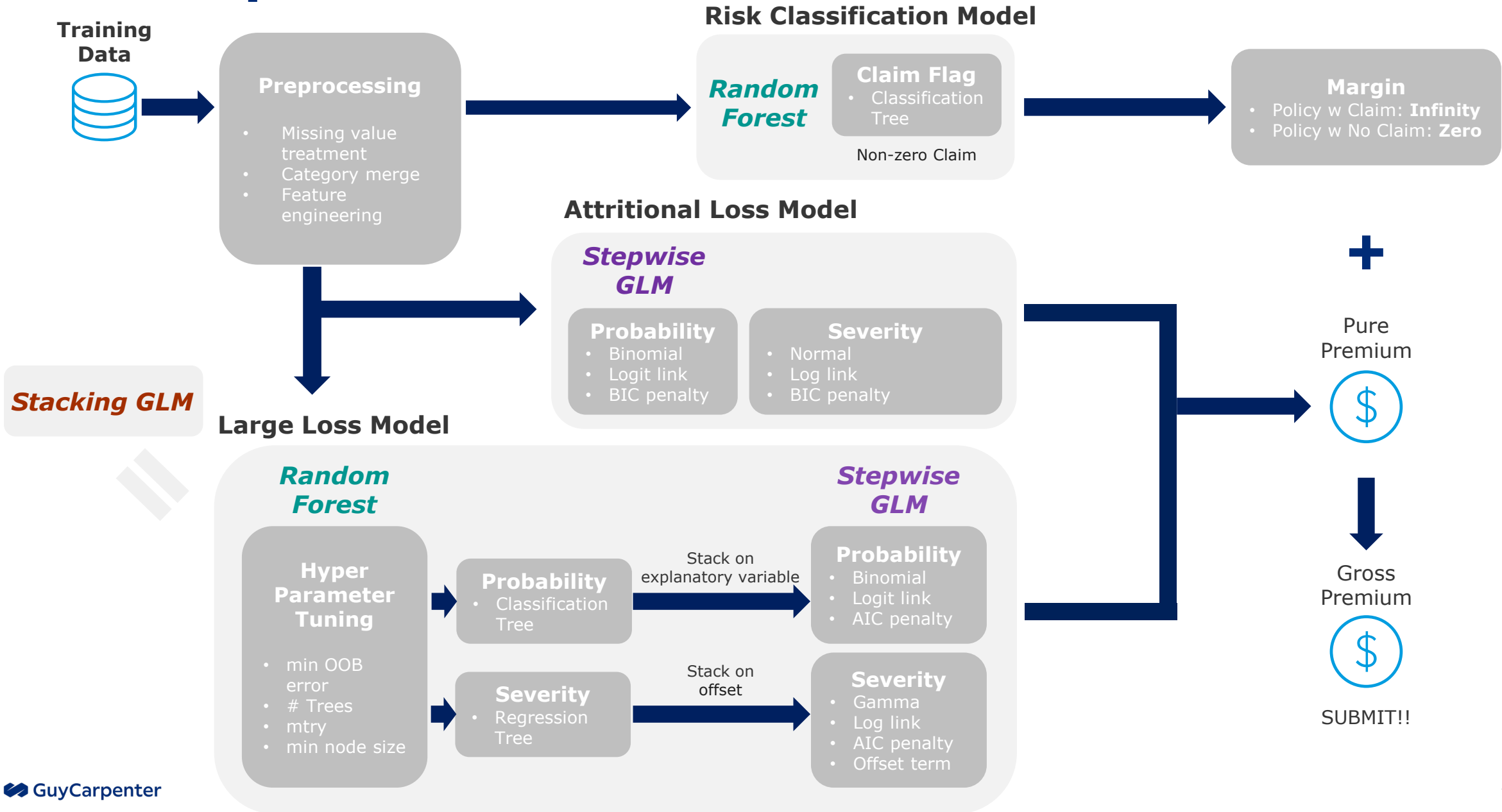
<https://www.youtube.com/watch?v=GkU2IqZu1gA>

Our Submitted Model & Final Results

Roadmap

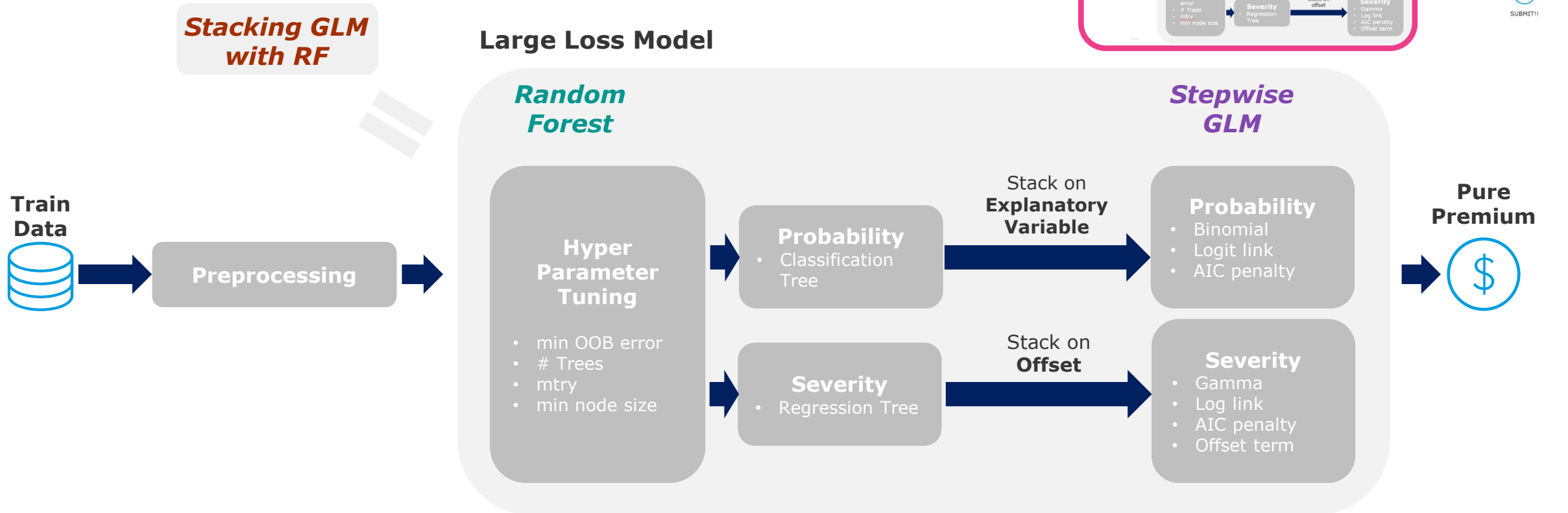


Model Pipeline



Ensemble Approach (Stacking)

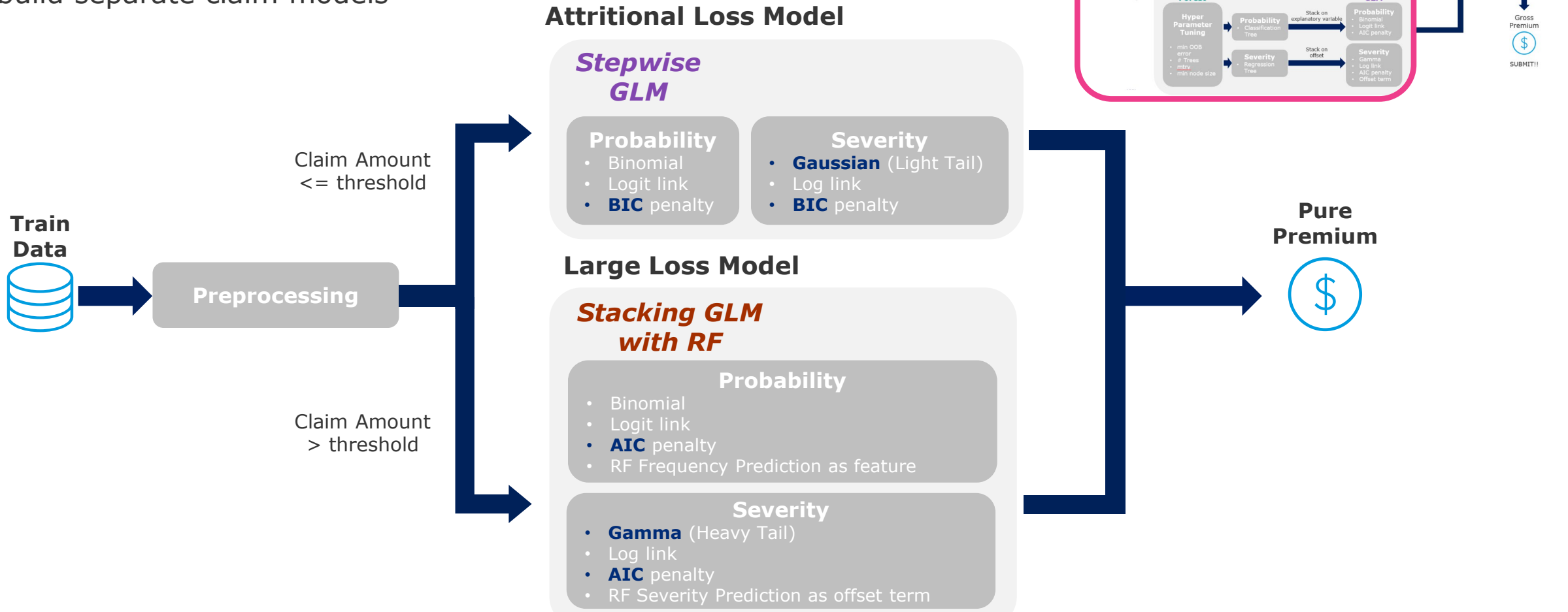
Using Random Forest's prediction value as a feature/offset term



Random Forest can capture large losses & Relatively easy hyper-parameter tuning!

Loss Element Decomposition

Split train data into “Attritional” and “Large” losses and build separate claim models

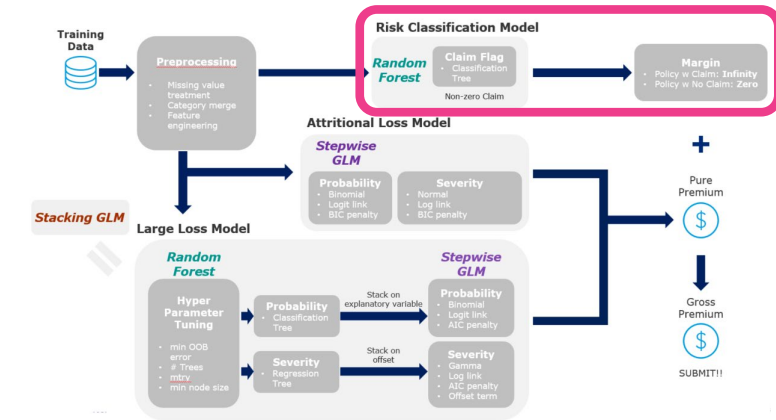
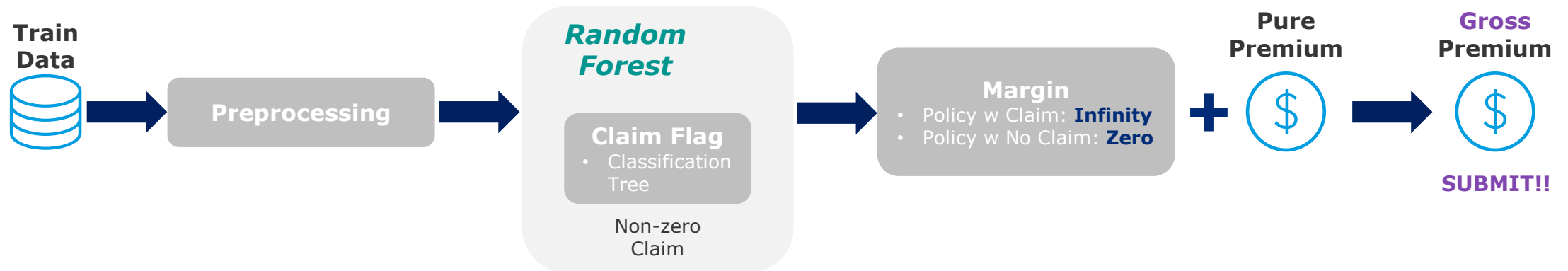


Capturing different characteristics of Attritional (Light Tail) and Large (Heavy Tail) losses by separating the two!

Pricing Model for Profit Leaderboard

Accurate model does not always win in the profit leaderboard

- Do NOT accept policies that potentially cause losses
- On the other hand, underwrite good risks actively which lead to profits



Using Random Forest Flag to select low risk policies to win in the profit leaderboard

Result



OUR ACHIEVEMENT

10th Place
In the Final Profit Leaderboard

Top Models
selected by the organizer

- Invited to present our model along with top 4 participants at the townhall

#	Participants	Participation	Average Profit	Market Share	Entries	Last Submission	Submission Trend	
01	WorldExperts	Yes	185919.311	0.161	1	Sun, 7 Mar 2021 14:11		View Code
02	glen...			0.190	1	Mon, 8 Mar 2021 07:16		View
		Yes			1	Mon, 8 Mar 2021 03:14		View
						Thu, 4 Mar 2021 00:58		View
09	majestic-kernel	Yes	5803			7 Mar 2021 12:27		View
						Mar 2021 04:09		View
10	Grape_Century	Yes	1453.2			Mar 2021 23:27		View
						Mar 2021 05:42		View
11	mohith_k_sakt...	Yes	487			Mar 2021 00:31		View
						Mon, 8 Mar 2021 12:54		View
	tmon01	Yes			1	Tue, 2 Feb 2021 18:04		View
12			0.000		1	Mon, 8 Mar 2021 04:00		View
13	iminokhi	Yes	10.161	0.003	1	Sat, 6 Mar 2021 21:18		View
14	jeremiedb	Yes	8.803	0.000	1	Sun, 17 Jan 2021 05:57		View
15	dmitri	Yes	-33.749	0.001	1	Tue, 2 Mar 2021 19:56		View
16	themathaguy	Yes	-107.520	0.000	1	Mon, 1 Mar 2021 05:21		View
17	arun_k	Yes	-168.406	0.000	1	Sun, 7 Mar 2021 01:36		View
18	AIISlarz623	Yes	-193.940	0.000	1	Sat, 6 Mar 2021 02:23		View
19	AmianG_itian	Yes	-280.122	0.001	1	Sun, 28 Feb 2021 04:10		View
20	alpertemel	Yes	-569.087	0.000	1	Mon, 15 Feb 2021 00:20		View

Our Findings

Summary

We used the combination of GLM and Random Forest, "Stacking GLM", for our model
→ More likely to be accepted in the insurance industry due to a high degree of **interpretability & accountability**

Through Our First Machine Learning Competition...



Firsthand experience of the strength of the common task frame



Amazing learning opportunity



Challenges of aspects other than prediction accuracy

Japanese Participants Achievement

	Player 1	Player 2	Our Team
Individual or Team	Individual	Individual	Team
Participation Period	Week 3 – (9 weeks)	Week 3 – (9 weeks)	Week 6 – (6 weeks)
Submitted Model	AGLM	NGBoost	GLM + RF (Stacking)
RMSE Leaderboard	18 th place	126 th place	43 rd place
Profit Leaderboard	52 nd place	12 th place	10 th place

Utilizing Machine Learning Methods

- We hope to develop machine learning skills that align with insurance industry needs such as **uncertainty, accountability and model governance**
- As a reinsurance broker, we are keen to apply machine learning methods to evaluate **low-frequency & high-severity risks** for realistic market, despite our strategy to win profit leaderboard by opting to sell only low risk policies in this competition
- Looking forward to project/competition focusing on optimization of **insurance risk diversification** (multi-year, geographical, markets, etc.)



Suguru Fujita

suguru.fujita@guycarp.com

Sara Sakimura

sara.sakimura@guycarp.com

Guy Carpenter Japan, Inc.
GC Analytics Tokyo

<https://www.guycarp.com/>

Appendix. Validation Process

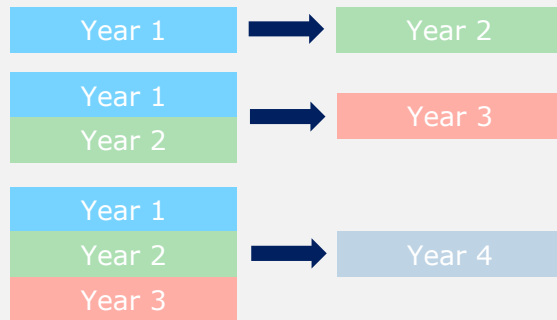
Validation Process

- Two patterns
- Chronological validation
- RMSE

Pattern 1

Training

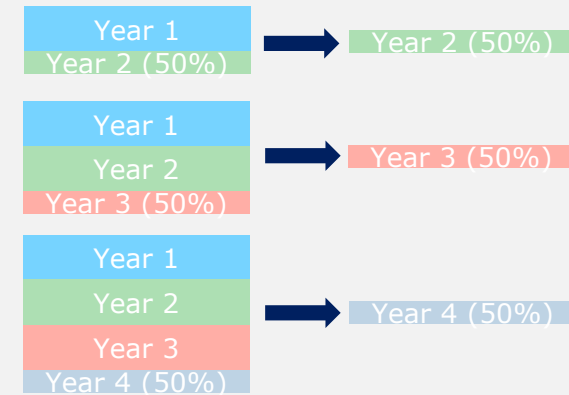
Testing



Pattern 2

Training

Testing



Appendix. Poll

1. Are you interested in participating in data analytics competition?

- Yes or No

2. Have you participated in data analytics competition before?

- Yes or No

3. To people who are interested in data analytics competition but not ready to participate, what is the biggest barrier for you?

- No time
- Concerns about knowledge and programming skills
- Don't want to do it alone
- Other

Appendix. Poll

1. Are you interested in participating in data analytics competition?

- Yes or No

2. Have you participated in data analytics competition before?

- Yes or No

3. To people who are interested in data analytics competition but not ready to participate, what is the biggest barrier for you?

- No time
- Concerns about knowledge and programming skills
- Don't want to do it alone
- Other