Towards Open Collaboration in Insurance Innovation

Kevin Kuo @kevinykuo
RStudio, Kasa AI (kasa.ai)

TL;DR: Get folks **together** to do research in the open, create software and workflows, and develop educational materials.

Wouldn't it be nice if...

We had open source state-of-the-art individual claims forecasting models that people can easily adopt for their own data so they can do reserving *and* claims analytics better?

Wouldn't it be nice if...

We addressed the interpretability issues of ML models for ratemaking and get regulators on board so we can charge more accurate premiums?

Wouldn't it be nice if...

We had end-to-end tutorials for pricing/reserving with realistic messy data for onboarding new actuaries/data scientists?

Wouldn't it be nice if...

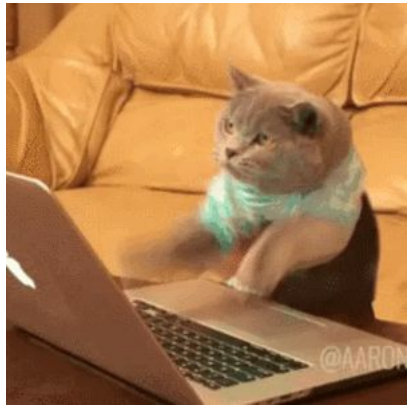
We had a standard set of packages/workflows for common problems so we reduce friction in audits and knowledge transfer?

Or, **solving common problems together** so we have more resources to focus on our own unique problems

Together?
Who?

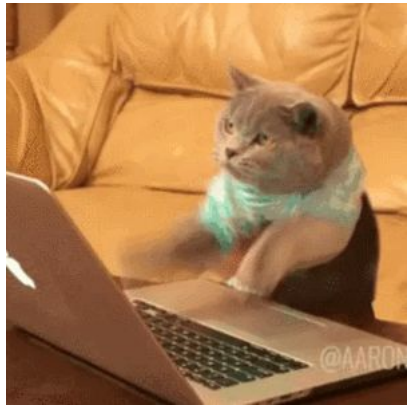
Actuary

Data scientist / software engineer



Actuary

Data scientist / software engineer

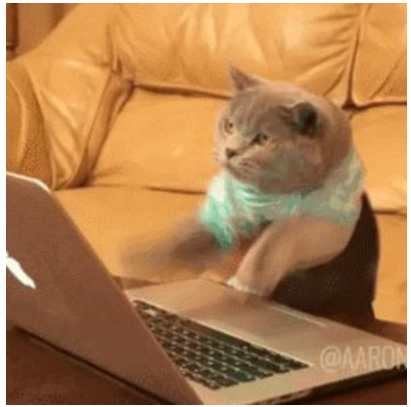


Actuary

“Business stakeholders”

How?

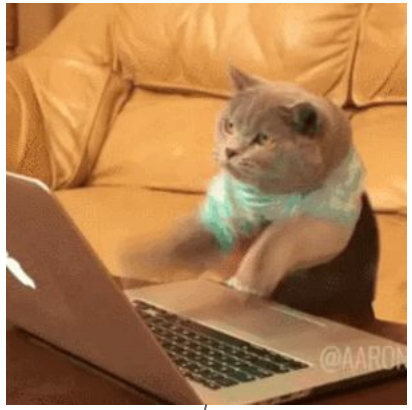
Data scientist / software engineer



Actuary

“Business stakeholders”

Data scientist / software engineer



Actuary

GitHub

slack

“Business stakeholders”

Git what?

```r
make_keras_model <- function() {
  input_gender <- layer_input(shape = 2, name = "gender")
  input_issue_age_group <- layer_input(shape = 1, name = "avg_issue_age")
  input_face_amount_band <- layer_input(shape = 1, name = "face_amount")

  # input_avg_premium_jump_ratio <- layer_input(shape = 1, name = "avg_premium_jump_ratio")
  input_premium_jump_ratio <- layer_input(shape = 1, name = "premium_jump_ratio")
  # input_risk_class_mapped <- layer_input(shape = 1, name = "risk_class_mapped")
  input_risk_class <- layer_input(shape = 1, name = "risk_class")

  input_premium_mode <- layer_input(shape = 1, name = "premium_mode")
  input_duration <- layer_input(shape = 3, name = "duration")

  embedding_risk_class <- input_risk_class %>%
    layer_embedding(9, 2, name = "embedding_risk_class") %>%
    layer_flatten()
  embedding_premium_jump_ratio <- input_premium_jump_ratio %>%
    layer_embedding(24, 2, name = "embedding_premium_jump_ratio") %>%
    layer_flatten()
  embedding_face_amount_band <- input_face_amount_band %>%
    layer_embedding(4, 2, name = "embedding_face_amount_band") %>%
    layer_flatten()
  embedding_premium_mode <- input_premium_mode %>%
    layer_embedding(6, 2, name = "embedding_premium_mode") %>%
    layer_flatten()
```

Home for code

Commits on Feb 6, 2019

Added init documentation for download functions
actuarialvoodoo committed on Feb 6 ✓ e394275

Moved download functionality to a function int he pricingtutorial pac... ···
actuarialvoodoo committed on Feb 6 6468476

Updated download_file to check for presence of download folder before... ···
actuarialvoodoo committed on Feb 6 eac9244

Commits on Feb 2, 2019

Added script to fetch data from web and add to external_data folder. ···
actuarialvoodoo committed on Feb 2 ✓ bd97b93

Commits on Jan 3, 2019

Merge pull request #72 from kasaai/bugfix/update-distill ···
kevinykuo committed on Jan 3 ✓ Verified 75b0fd9

rebuild blog post
kevinykuo committed on Jan 3 ✗ 6ecfb88

add a couple library() calls to prevent cached content from breaking ... ···
kevinykuo committed on Jan 3 8df716d

fix citation for naic whitepaper
kevinykuo committed on Jan 3 9067c39

Version control

Decide on model type #45

Open **kevinykuo** opened this issue on Dec 14, 2018 · 7 comments

kevinykuo commented on Dec 14, 2018 — Member

Let's keep it simple with GLM.

- Regularization?
- Frequency/severity vs pure premium
- By-peril or not
  - Whether to account for dependency if by-peril

kevinykuo added the modeling label on Dec 16, 2018

RonRichman commented on Dec 18, 2018

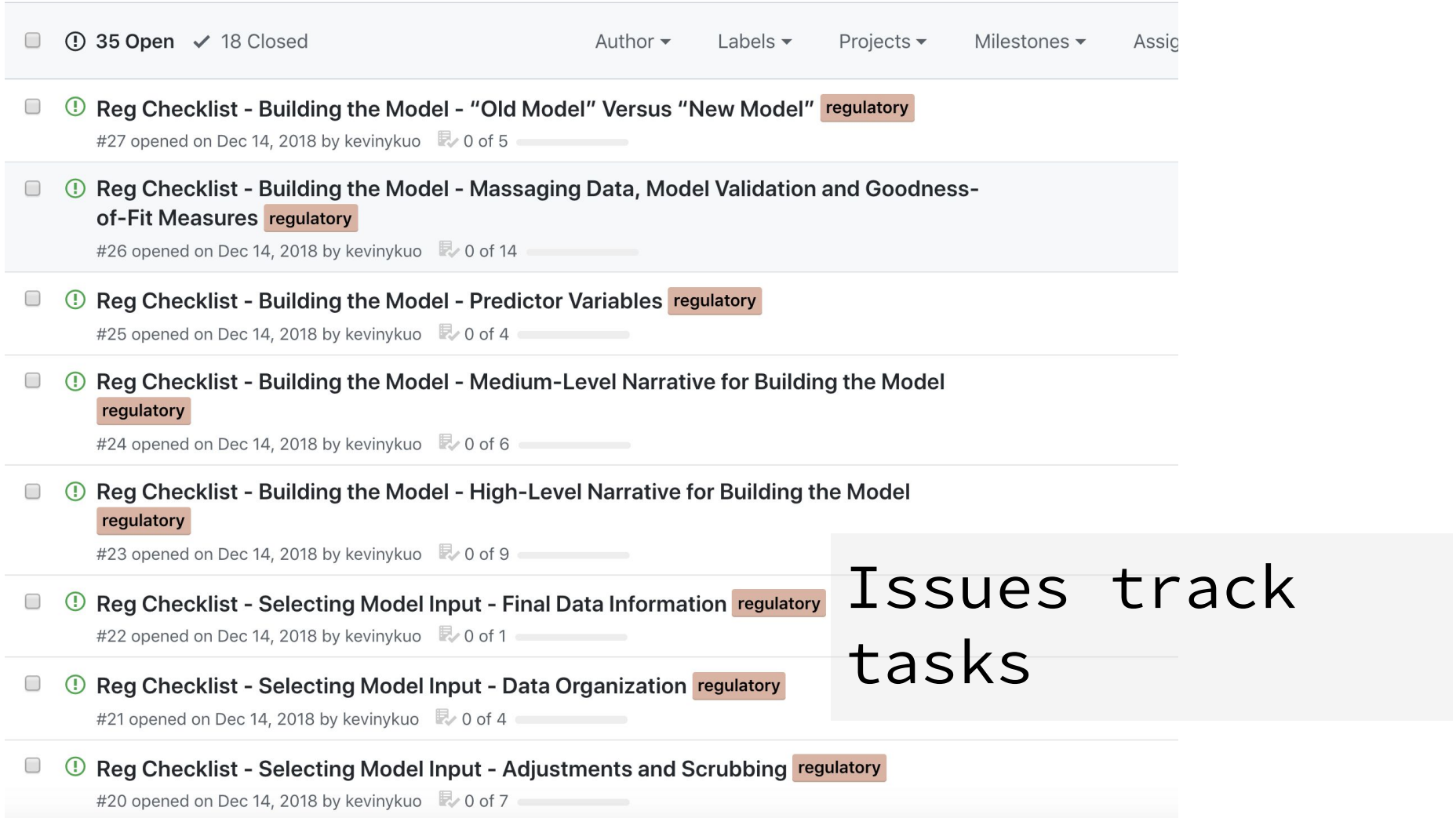
My thoughts would be that it makes sense to start of with GLM, ideally freq/sev for all perils together. With that baseline, it then becomes easier to show the benefits of regularization, or ML/DL methods.

EKtheSage commented on Jan 16

What if you have little to no loss history and you want to use industry loss cost to supplement your analysis?

In our company we used ISO rating structure as the starting point and built a residual model off of that. The model was a LR model since our exposure was not reliable.

Discuss approaches in issues


35 Open
18 Closed
Author
Labels
Projects
Milestones
Assig
Reg Checklist - Building the Model - "Old Model" Versus "New Model" regulatory
#27 opened on Dec 14, 2018 by kevinykuo 0 of 5
Reg Checklist - Building the Model - Massaging Data, Model Validation and Goodness-of-Fit Measures regulatory
#26 opened on Dec 14, 2018 by kevinykuo 0 of 14
Reg Checklist - Building the Model - Predictor Variables regulatory
#25 opened on Dec 14, 2018 by kevinykuo 0 of 4
Reg Checklist - Building the Model - Medium-Level Narrative for Building the Model regulatory
#24 opened on Dec 14, 2018 by kevinykuo 0 of 6
Reg Checklist - Building the Model - High-Level Narrative for Building the Model regulatory
#23 opened on Dec 14, 2018 by kevinykuo 0 of 9
Reg Checklist - Selecting Model Input - Final Data Information regulatory
#22 opened on Dec 14, 2018 by kevinykuo 0 of 1
Reg Checklist - Selecting Model Input - Data Organization regulatory
#21 opened on Dec 14, 2018 by kevinykuo 0 of 4
Reg Checklist - Selecting Model Input - Adjustments and Scrubbing regulatory
#20 opened on Dec 14, 2018 by kevinykuo 0 of 7
Issues track tasks

This conversation was marked as resolved by **actuarialvoodoo** Hide conversation

kevinykuo on Feb 5 **Member**

Can we wrap this mechanism into a function that takes the time periods to be downloaded as an argument? I think it can live in the package and we can provide example usage in `README`.

actuarialvoodoo on Feb 5 Author **Member**

If I were just doing this analysis for myself, I would (ideally) use a makefile to automate/orchestrate the analysis. Is that a direction we want to consider for this project?

kevinykuo on Feb 6 **Member**

I think GNU Make is a little too intimidating for newer users. That said, we do need to stay organized as we scale up. Maybe we can look at https://github.com/ropensci/drake?

actuarialvoodoo on Feb 6 Author **Member**

I added an argument to the function `download_data` which will allow users to specify the datasets to download. I will file an issue to use drake tomorrow.

Reply...

Unresolve conversation

```
2  + library(tidyverse)
3  + library(fs)
```

Discuss changes line by line

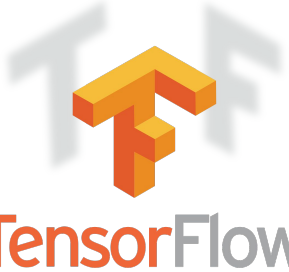
Leverage resources across the industry

Make the world a better place

Sounds good but...


Yeah if you could stop divulging trade secrets
That would be great

Or, solving common problems together so we have more resources to focus on our own unique problems



Open source? Yeah, it's a thing

KASA AI is a new open source community dedicated to making insurance analytics even more fun!

* Kasa = Japanese for ***umbrella***. Because insurance.

Kasa.ai

Blog.kasa.ai

@kevinykuo