MICRORESERVING FOR WORKER'S COMP

A case study from a Kaggle competition

Boosted Goose (Nelvis Fornasin, Attila Gulyas)

DNA Seminar "Advanced methods for reserving, claims processing, and fraud"

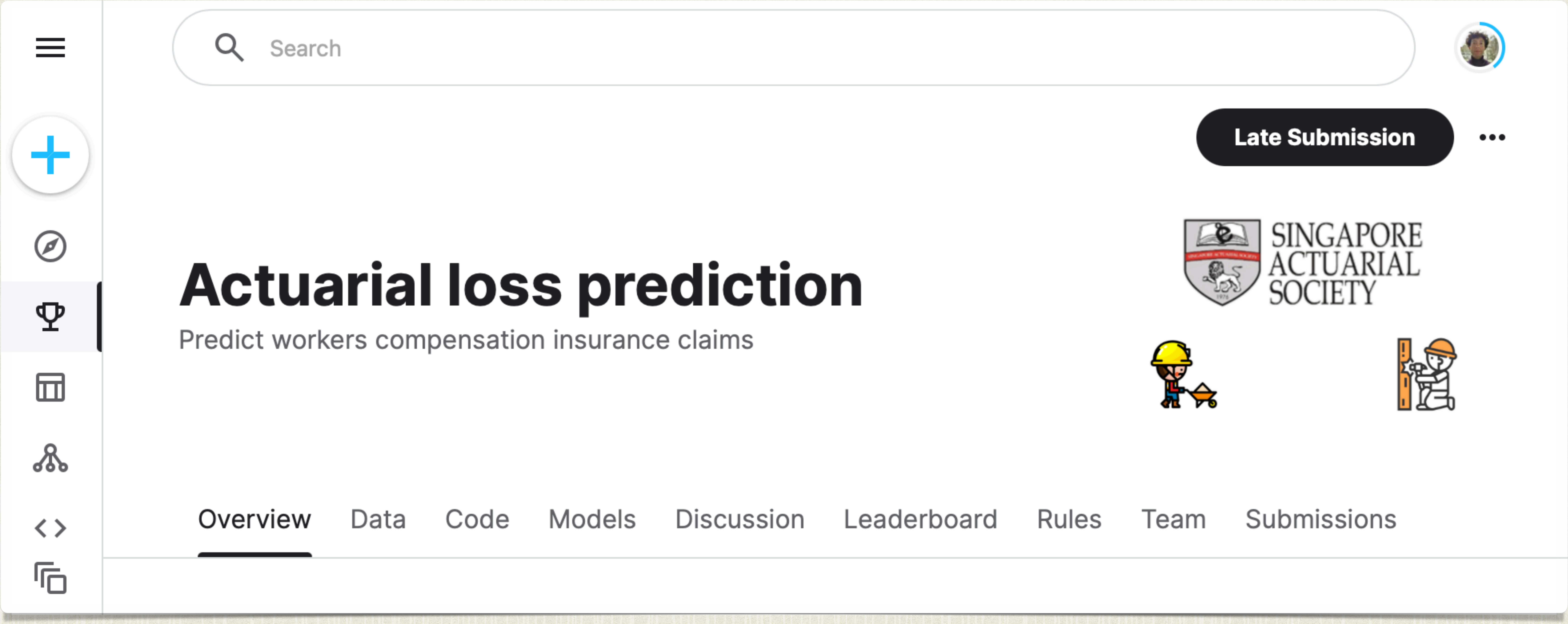
Oslo, 30.05.2024

INTRODUCTION

- At the beginning of 2021, Attila and I took part to the Kaggle competition "Actuarial Loss Prediction". Aim of the competition was to estimate Workers' Compensation claims on a single claim basis. Our team ("Boosted Goose") won second place with an xgboost based solution.
- Today I will present the model!

AGENDA

- The Kaggle competition
- Data & Context
    - Claim descriptions
- Modelling with xgboost
- Final remarks



THE KAGGLE COMPETITION

Actuarial Loss Prediction

ACTUARIAL LOSS PREDICTION

- **Duration**: December 2020 to April 2021
- **Organisers**: Actuaries Institute of Australia, Institute and Faculty of Actuaries and Singapore Actuarial Society
- **Platform**: Kaggle
- **Objective**: Prediction of workers' compensation reserves on an individual claim basis, i.e., development of an individual claim reserving model for IBNeR (Incurred But Not enough Reported)
- **Data**: Synthetically generated without reference to a specific jurisdiction or country. The dataset included, among other things, the insured person's anagraphic data, a description of the damage, and an initial estimate of the ultimate.
- **Evaluation of submitted solutions**: Mean squared error (MSE)
- **Our result**: 2nd place among the 140 participating teams/individuals.

DATA & CONTEXT

Workers' Comp and Data Exploration

WORKERS' COMPENSATION

- "The real grievance of the worker is the insecurity of his existence; he is not sure that he will always have work, he is not sure that he will always be healthy, and he foresees that he will one day be old and unfit to work. If he falls into poverty, even if only through a prolonged illness, he is then completely helpless, left to his own devices, and society does not currently recognise any real obligation towards him." [Otto von Bismarck, 1884]

- **Purpose:** Provides wage replacement and medical benefits to employees injured in the course of employment. It limits employer liability in the event of workplace injuries, preventing lawsuits.

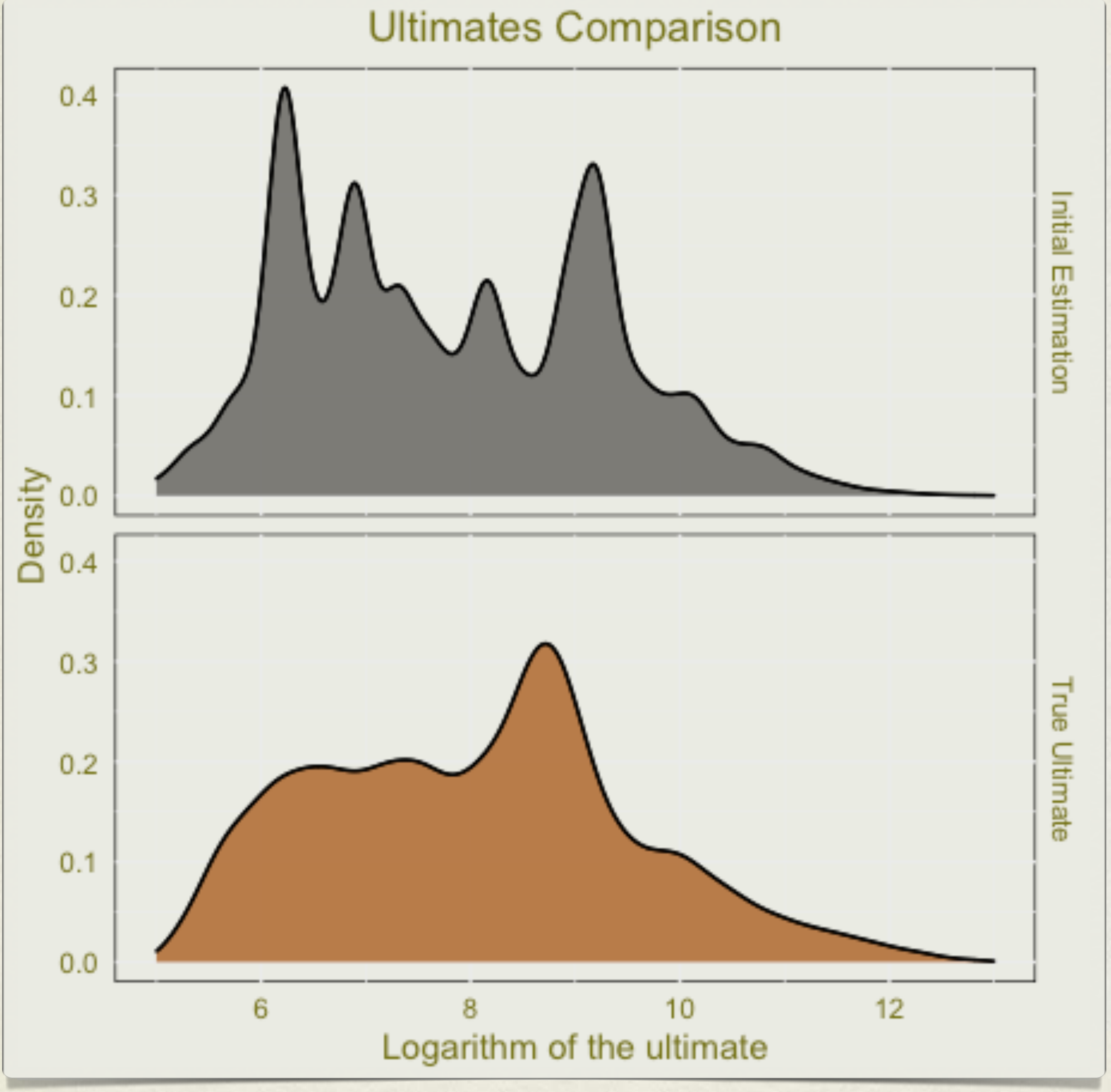
- **Coverage:** Includes medical expenses, rehabilitation costs, and lost wages. Covers temporary and permanent disability.

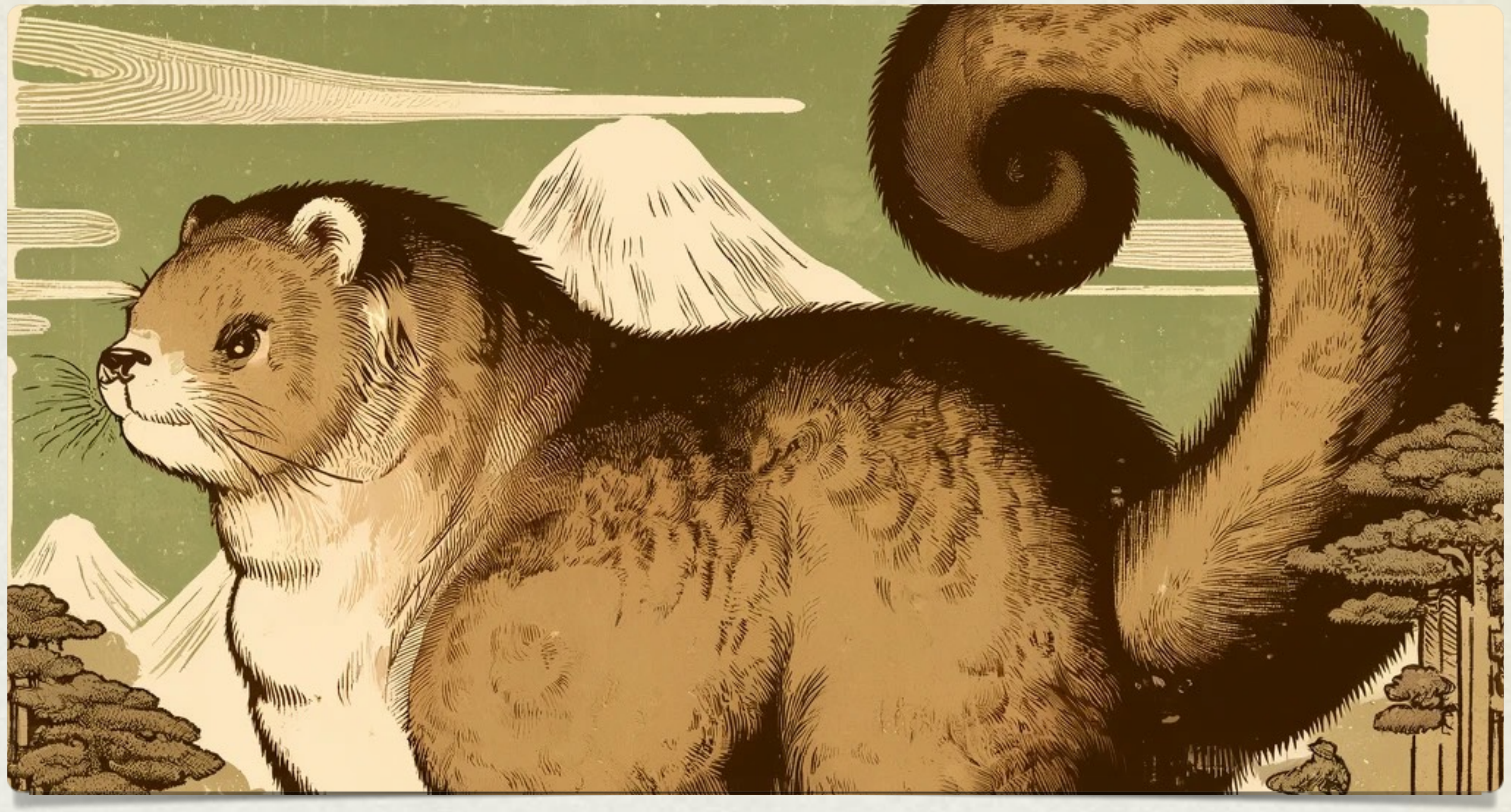
OVERVIEW OF THE DATA

Age	Gender	Marital Status	Dependent Children
42	F	S	0

Weekly wages	Part time/Full time	Hrs worked per week	Days worked per week
1000	F	40	5

Datetime of accident	Date reported	Claim description	Initial incurred costs
09/04/2002 7:00	05/07/2022	Cut on sharp edge cut left thumb	1500

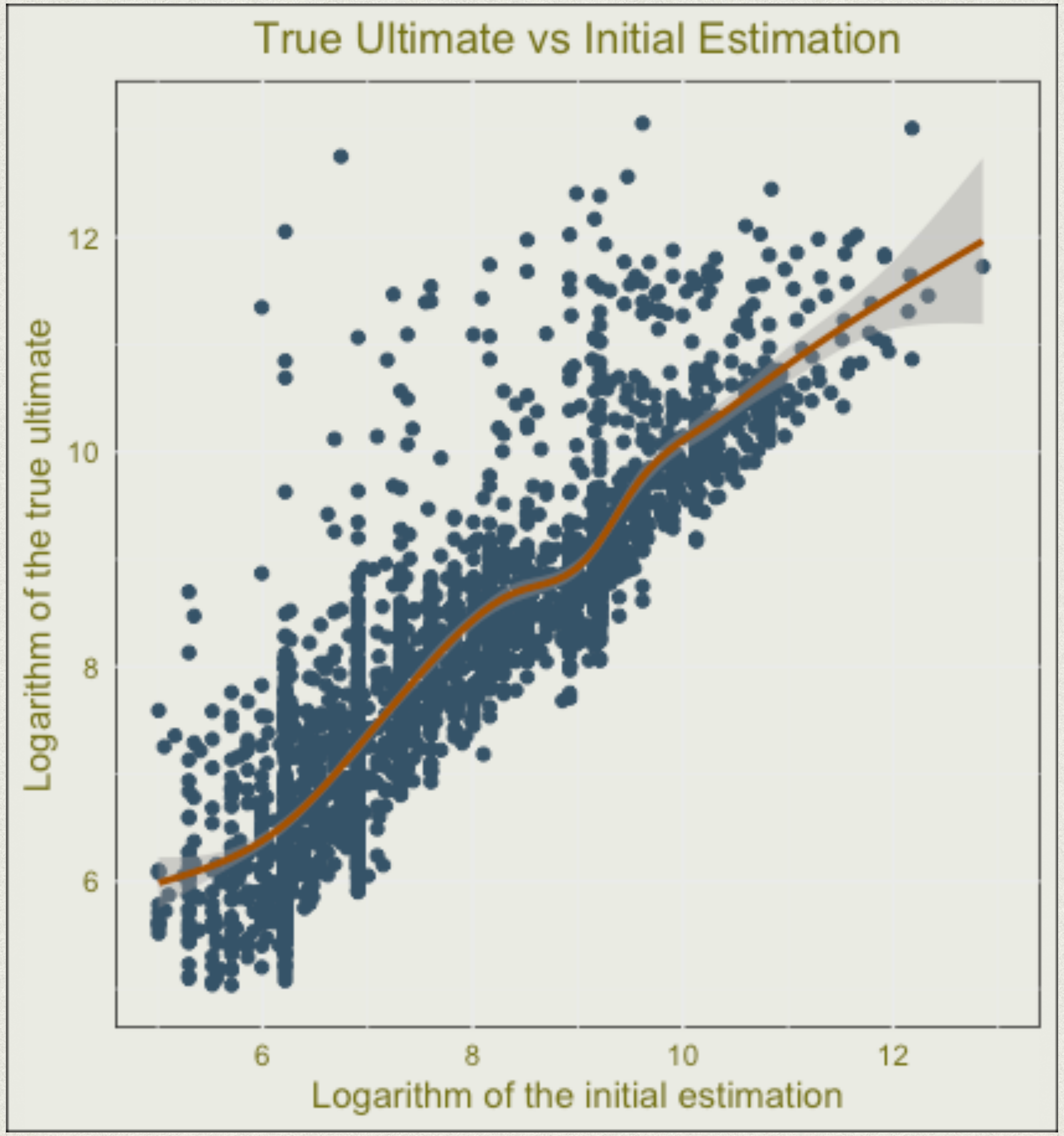




	Mean	Median	Max
Initial Estimation	7.803	2.000	872.980
True Ultimate	10.919	3.369	865.770

EXPLANATORY VARIABLES

- The initial estimation is a good indicator of the ultimate, but it fails spectacularly in single cases.
- We need to create a model that builds up on the initial estimates using the remaining explanatory variables.
- The most interesting one is the claim description!



CLAIM DESCRIPTION ANALYSIS

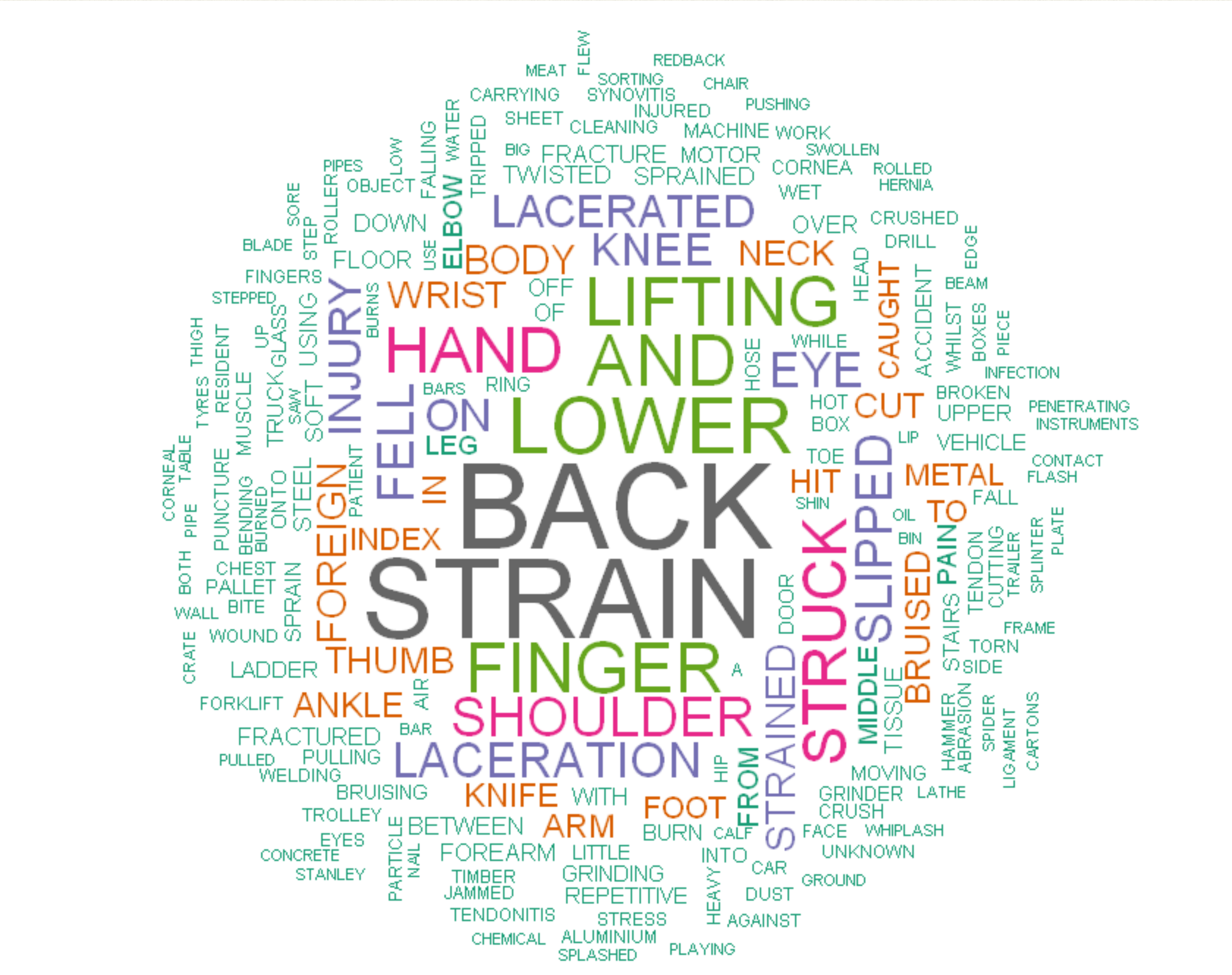
How to analyse text?

Claim Description	Frequency
SLIPPED ON ROLLER TENDONITIS RIGHT SHOULDER	289
REDBACK SPIDER BITE RIGHT FOOT RIGHT FRACTURE	256
WHILE DEALING CARDS RIGHT TENDON SYNOVITIS RIGHT WRIST	197
HOT METALLIC COFFEE MUG ON STAIRS TENDON STRAIN	93
MOTOR VEHICLE ACCIDENT WHIPLASH	52
DEFAMATION AND VICTIMISATION ANXIETY AND DEPRESSION	5
NAIL FROM NAIL GUN FOREIGN BODY LEFT CORNEA	1
COW HIT GATE GATE HIT TEETH BROKEN TEETH AND KNEE	1

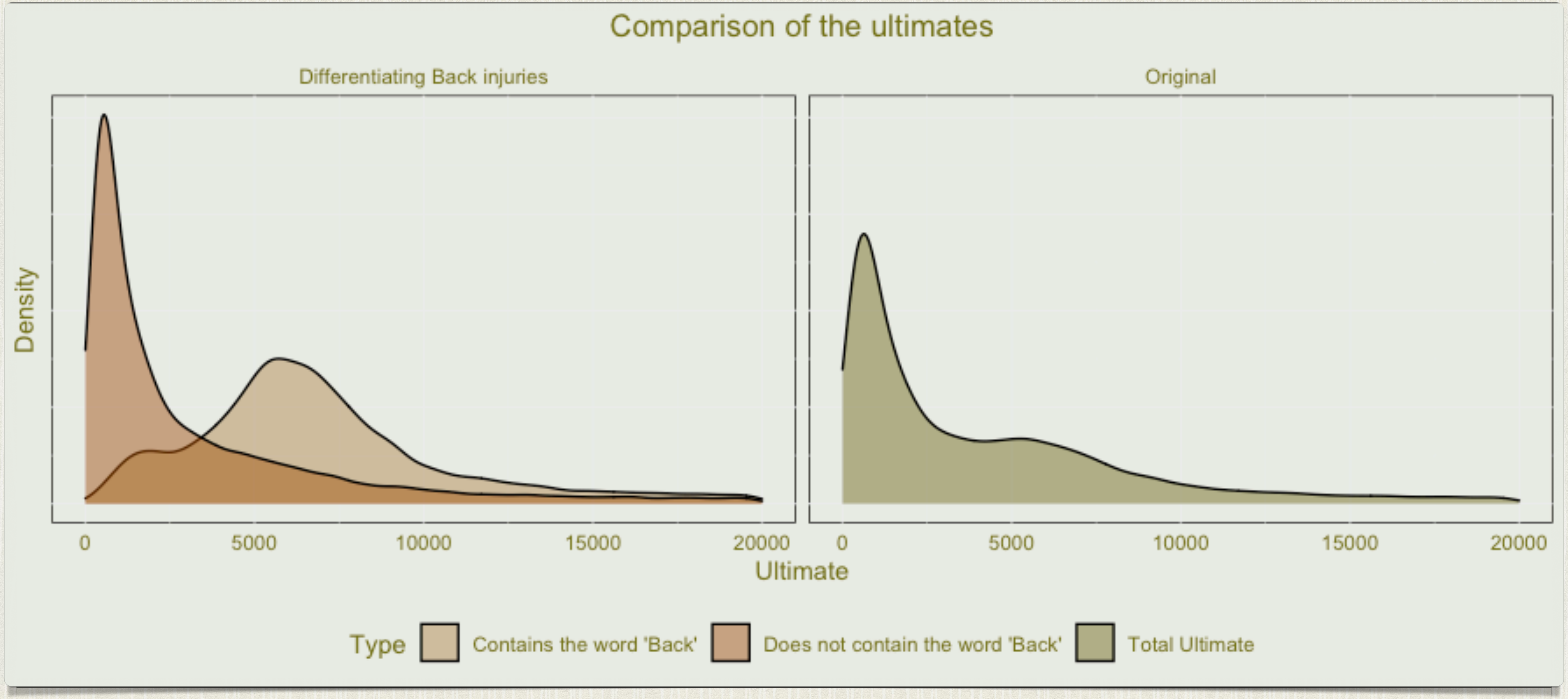
In the test dataset:

- 54.000 rows;
- 28.112 different claim descriptions;
- 21.435 appear only once;
- 3.641 appear twice;
- 31 appear at least 100 times.

WORDCLOUD



DUOLCDROW



BACK INJURIES

Claims whose description contains the word "back" behave differently

REPETITIVE BENDING OVER BRUISED RIGHT FOOT

- What is important in this claim description? How to simplify it?
- **Lemmatization**: bending => bend
- **Stemming:** repetitive => repet
- Drop unimportant words
- Get: “REPET BEND BRUISE FOOT”
- Now e.g. one hot encoding.

SAMPLE ANALYSIS

Claim Descriptions
PULLING PIECES OF STEEL LACERATED FINGER
BITTEN BY DOG THUMB LACERATION AND UPPER ARM

pull	lacer	finger	bite	thumb	arm	cluster-hand
1	1	1	0	0	0	1
0	1	0	1	1	1	1

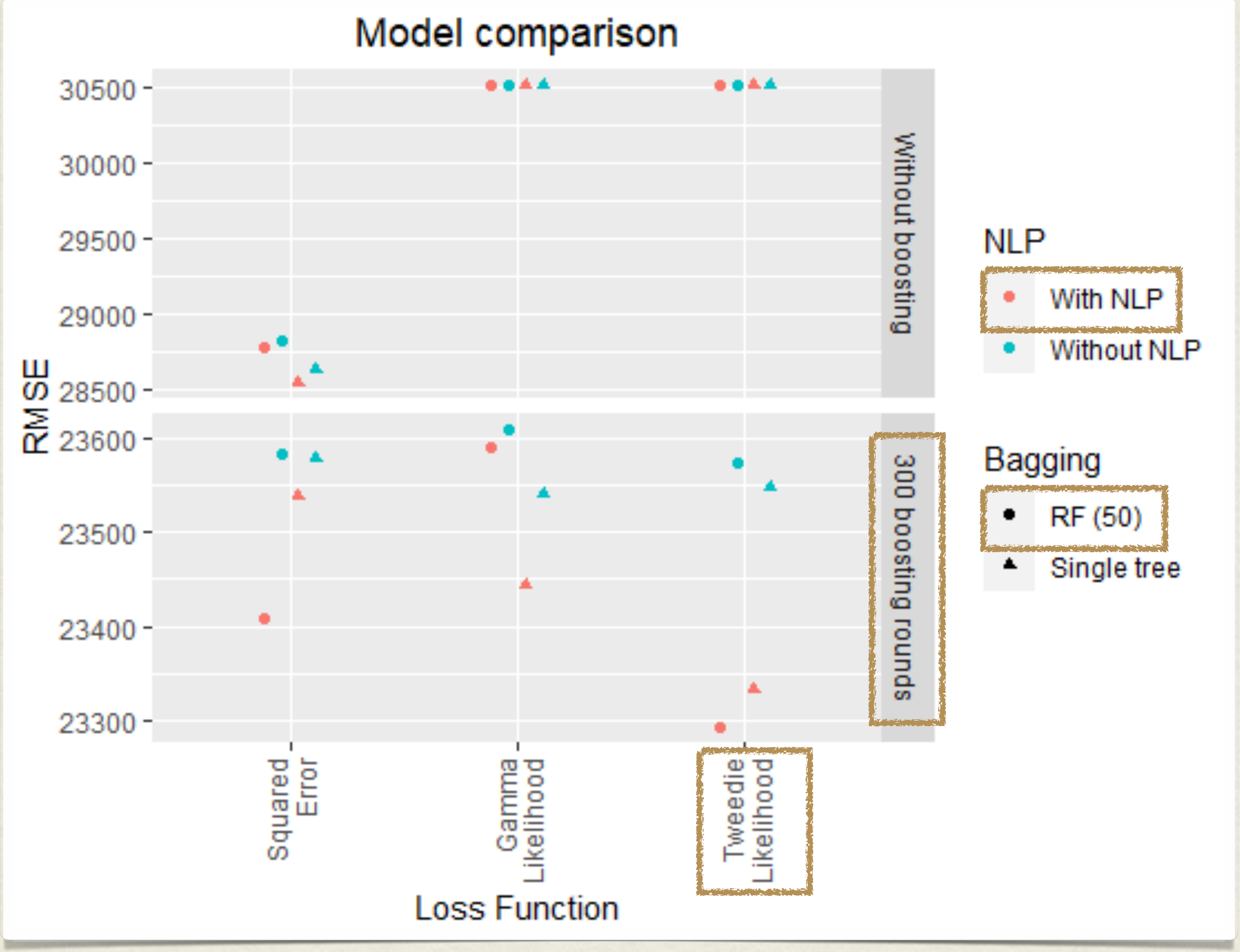
ONE-HOT ENCODING

- One-Hot Encoding a variable with N unique values generates N columns.
- In our case we ended up with a dataset with 230ish features!
- So we needed a model that can handle this amount of information without overfitting.

MODELLING WITH XGBOOST

Details on the model and a quick primer on the xgboost algorithm

THE MODEL



Main ingredients of our model:

- Text analysis (NLP)
- Random forest
- Boosting
- Likelihood loss function

WHAT IS BOOSTING?

- **Boosting:** training multiple models sequentially to improve the accuracy of the overall system.
- Each model (e.g. decision tree) feeds on the prediction of the previous one and tries to refine it.
- The final prediction is the sum of the initial prediction plus all the iterative refinements.

Sample algorithm:

1. Predict target y: $(x, y) \rightarrow \hat{y}_0$
2. Adapt target to $y_1 = y_1(y, \hat{y}_0)$
3. Predict y_1: $(x, y_1) \rightarrow \hat{y}_1$
4. Final prediction: $\hat{y} = \hat{y}_0 + \hat{y}_1$

Key question: how to define y_1?

- **Loss function:** quantity $l(y, \hat{y})$ to be minimised by the model, e.g. MSE.

- How to find a minimum? With some regularity of the loss function there are several classical methods available.

- In the following we are going to assume that the following quantities are well defined:

$$g_y(\hat{y}) := \frac{\partial}{\partial \hat{y}} l(y, \hat{y}), \quad h = \frac{\partial^2}{\partial \hat{y}^2} l(y, \hat{y})$$

Squared Error:

$$l_{SE}(y, \hat{y}) := (\hat{y} - y)^2$$

then

$$g_{SE} = 2(\hat{y} - y)$$

$$h_{SE} = 2$$

GRADIENT DESCENT

- **1st order approximation:** if a first prediction $\hat{y}_0$ is given, we can approximate in a neighbourhood of $\hat{y}_0$:

$$l(y, \hat{y}) \sim l(y, \hat{y}_0) + g_y(\hat{y}_0)(\hat{y} - \hat{y}_0)$$

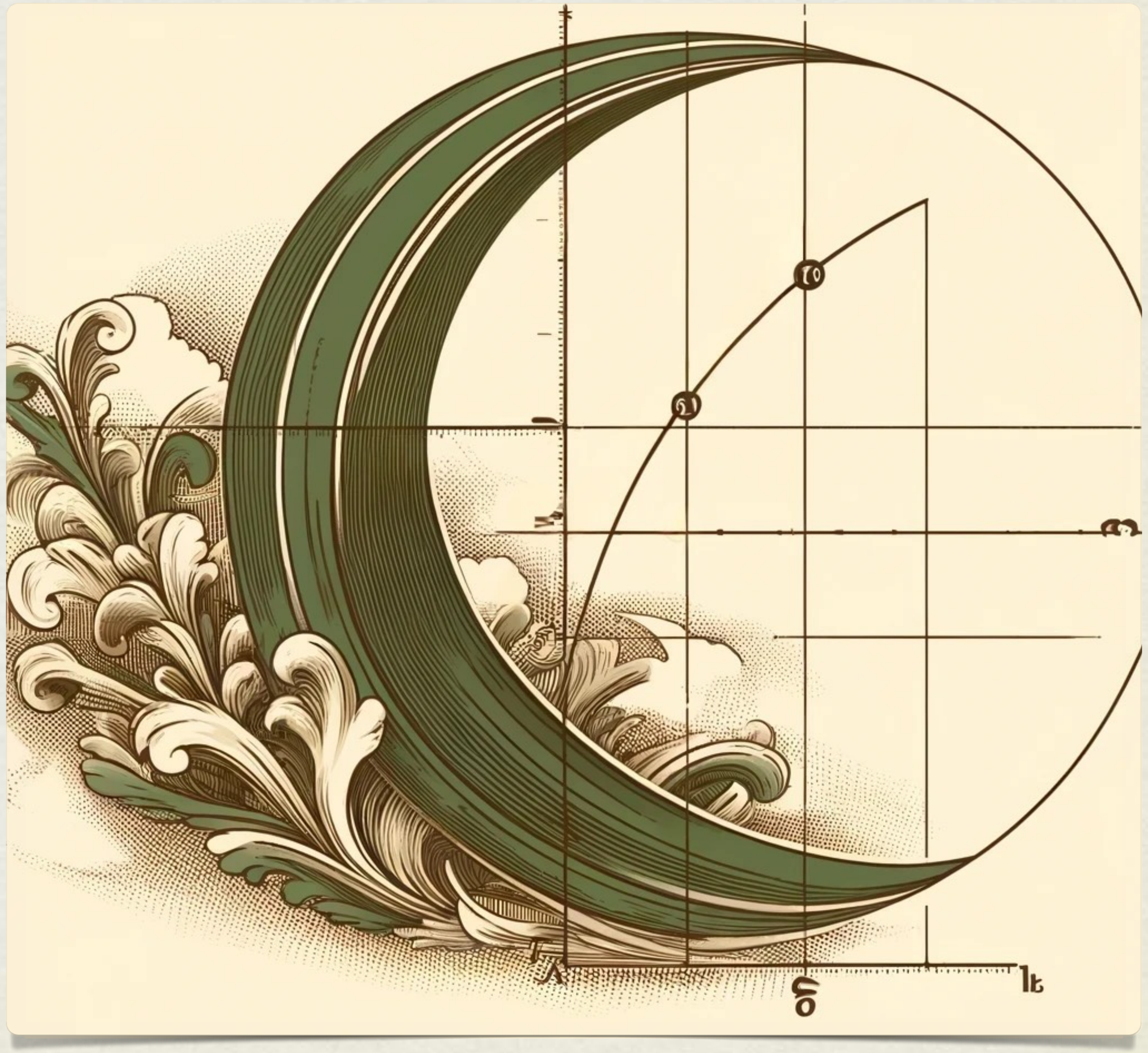
- Picking $\hat{y}_1 = \hat{y}_0 - \eta g_y(\hat{y}_0)$ with η small enough yields:

$$l(y, \hat{y}_1) \sim l(y, \hat{y}_0) + g_y(\hat{y}_0)(\hat{y}_1 - \hat{y}_0) = l(y, \hat{y}_0) - \eta g_y^2(\hat{y}_0) \leq l(y, \hat{y}_0)$$

- This is the basic idea behind **gradient descent**! Remarks:

  1. Equality only holds if $g_y(\hat{y}_0) = 0$, i.e. $l(y, \hat{y}_0)$ is already a minimum.

  2. The quantity η is called the learning rate.



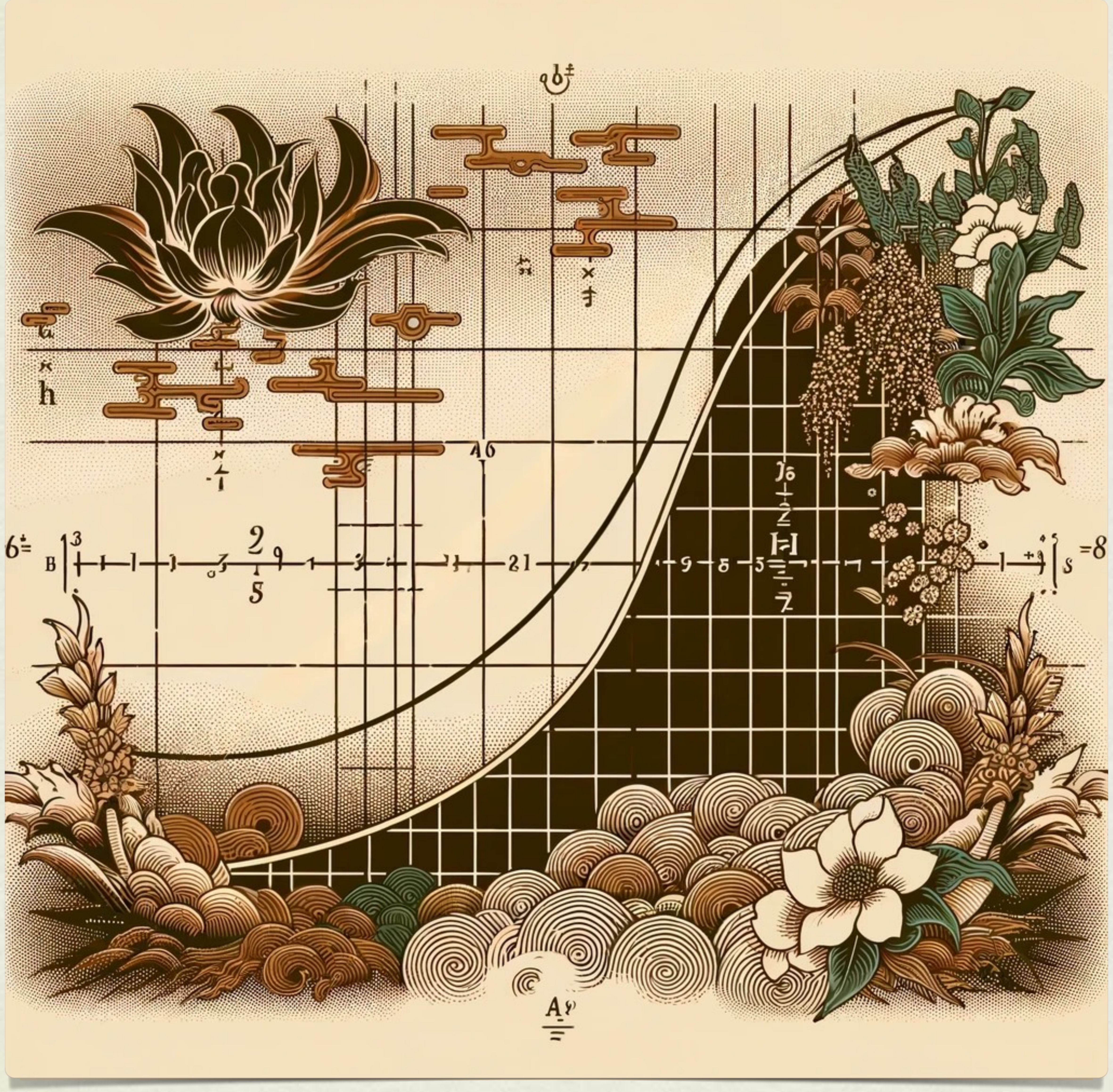
THE NEWTON-RAPHSON METHOD

- **2nd order approximation:** if a first prediction $\hat{y}_0$ is given, we can approximate further in a neighbourhood of $\hat{y}_0$: $l(y, \hat{y}) \sim l(y, \hat{y}_0) + g_y(\hat{y}_0)(\hat{y} - \hat{y}_0) + h_y(\hat{y}_0)\dfrac{(\hat{y} - \hat{y}_0)^2}{2}$

- Assume $h_y(\hat{y}_0) > 0$. Picking $\hat{y}_1 = \hat{y}_0 - g_y(\hat{y}_0)/h_y(\hat{y}_0)$ yields

$$l(y, \hat{y}_1) \sim l(y, \hat{y}_0) - \frac{g_y(\hat{y}_0)^2}{h_y(\hat{y}_0)} + h_y(\hat{y}_0)\frac{g_y^2(\hat{y}_0)}{2h_y^2(\hat{y}_0)} = l(y, \hat{y}_0) - \frac{g_y(\hat{y}_0)^2}{2h_y(\hat{y}_0)} \leq l(y, \hat{y}_0)$$

- This method amounts to fitting a parabola (rather than line) tangent to the loss function in $\hat{y}_0$ and computing its minimum.



Simplified xgboost algorithm:

1. Predict target y: $(x, y) \rightarrow \hat{y}_0$
2. New target: $y_1 = -\,g_y(\hat{y}_0)/h_y(\hat{y}_0)$
3. Predict y_1: $(x, y_1) \rightarrow \hat{y}_1$
4. Prediction: $\hat{y} = \hat{y}_0 + \hat{y}_1$
5. ...reiterate process

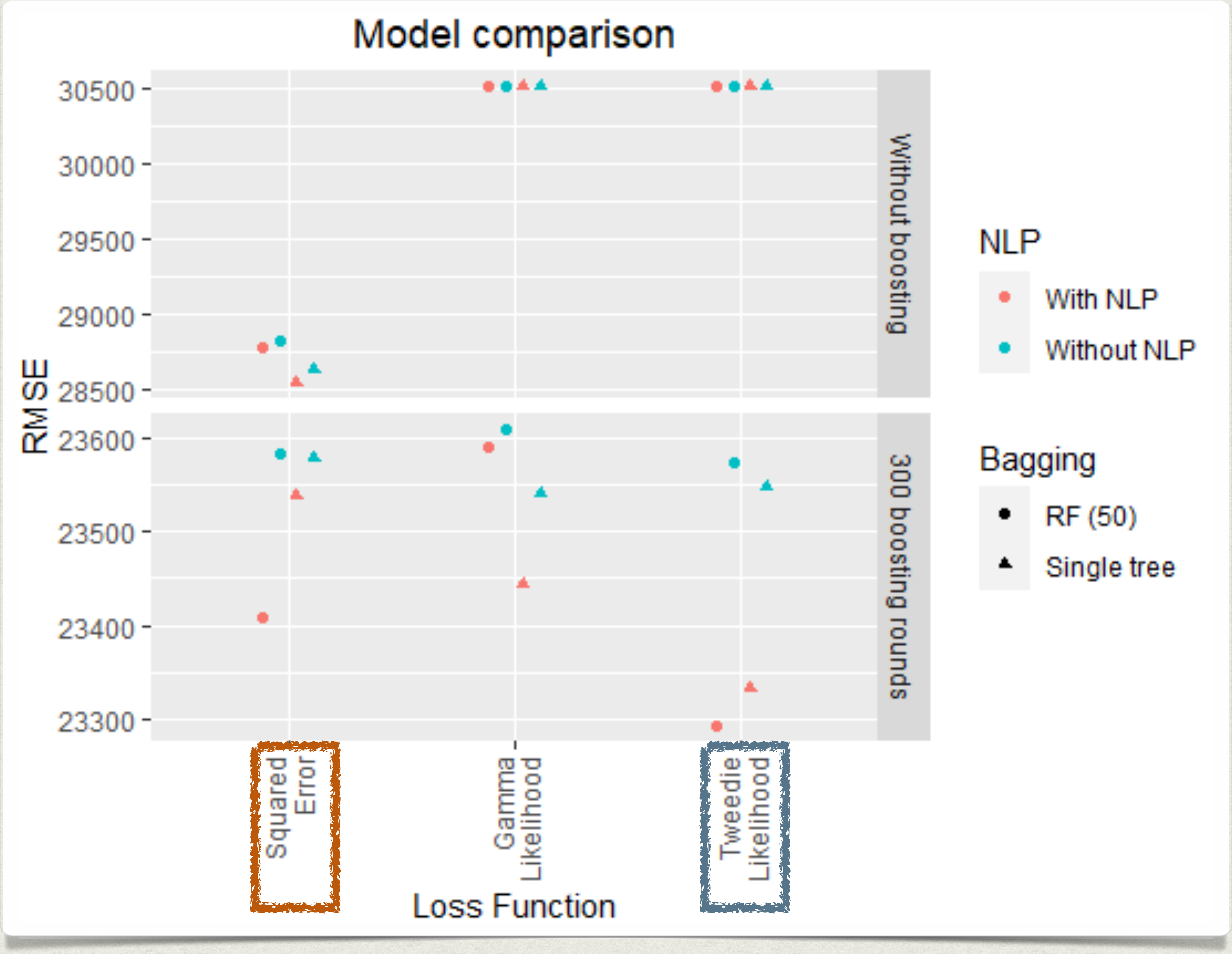
LIKELIHOOD AS A LOSS FUNCTION

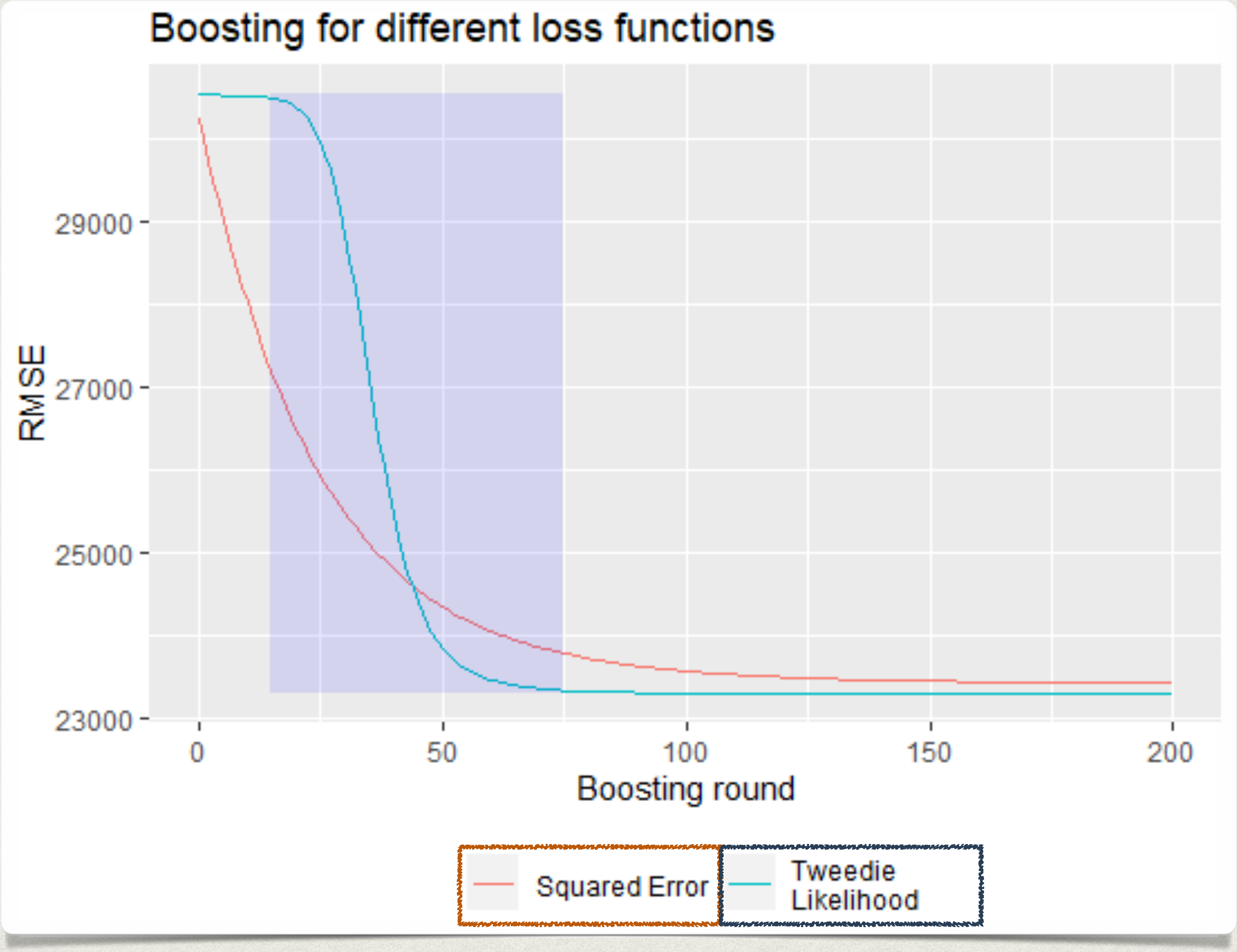
- We can apply the method described to different loss functions than MSE: for example, we can minimise the negative likelihood (=> maximise likelihood) to a given distribution.

- For the squared error we saw $g_{SE} = 2(\hat{y}_0 - y)$ and $h_{SE} = 2$. Then
$$y_1 = -g_{SE}/h_{SE} = y - \hat{y}_0$$

- For a Gamma likelihood we obtain a slightly different equation:
$$y_1 = -g_\gamma/h_\gamma = -(1 - e^{y-\hat{y}_0})/e^{y-\hat{y}_0} = 1 - e^{\hat{y}_0 - y} \sim y - \hat{y}_0 \text{ for } \hat{y}_0 \text{ close to } y$$

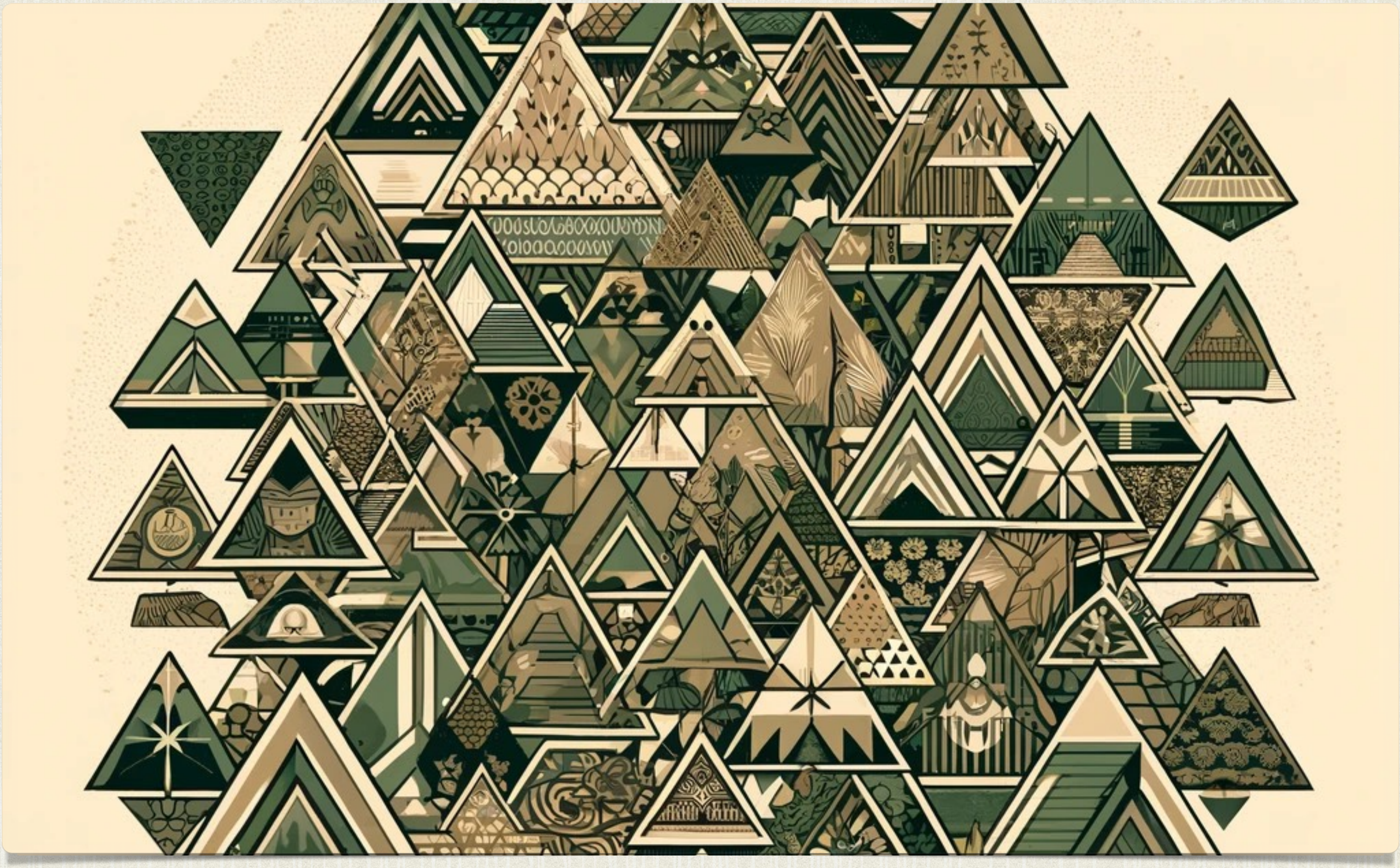
DIFFERENT LEARNING BEHAVIOURS





```python
def train_xgb_cv(config: dict):

    data=xgb.DMatrix(X_xgb, label=y)
    params = {'objective':config['objective'],
              'eval_metric':config['eval_metric'],
              'eta': config['eta'],
              'max_depth':config['max_depth'],
              'subsample':config['subsample'],
              'min_child_weight':config['min_child_weight'],
              'colsample_bylevel': config['colsample_bylevel'],
              'colsample_bynode' : config['colsample_bynode'],
              'colsample_bytree': config['colsample_bytree'],
              'gamma' : config['gamma'],
              'alpha' : config['alpha'],
              'monotone_constraints' : monotone_constraints,
              'tweedie_variance_power' : config['tweedie_variance_power']
              }

    results = xgb.cv(config,data,verbose_eval=False,num_boost_round=200,early_stopping_rounds=20,nfold=5,as_pandas=False)

    tune.report(rmse_mean=results['test-rmse-mean'][-1])

search_space = {
    'objective':'reg:tweedie',
    'eval_metric':'rmse',
    'eta': tune.uniform(1e-2, 2e-1),
    'max_depth': tune.randint(3, 6),
    'min_child_weight': tune.randint(4, 8),
    'subsample': tune.uniform(0.5, 0.8),
    'colsample_bytree' : tune.uniform(0.8,0.99),
    'colsample_bylevel': tune.uniform(0.8,1.0),
    'colsample_bynode' : tune.uniform(0.6,0.8),
    'gamma' : tune.loguniform(1e+1, 1e+5),
    'alpha' : tune.loguniform(1e+0, 5e+4),
    'tweedie_variance_power' : tune.uniform(1.1,1.6)
    }

analysis = tune.run(train_xgb_cv,config=search_space,metric='rmse_mean',mode="min",num_samples=50,verbose=1,
                    keep_checkpoints_num=1, checkpoint_score_attr="rmse_mean")
```

X
G
B
O
O
S
T



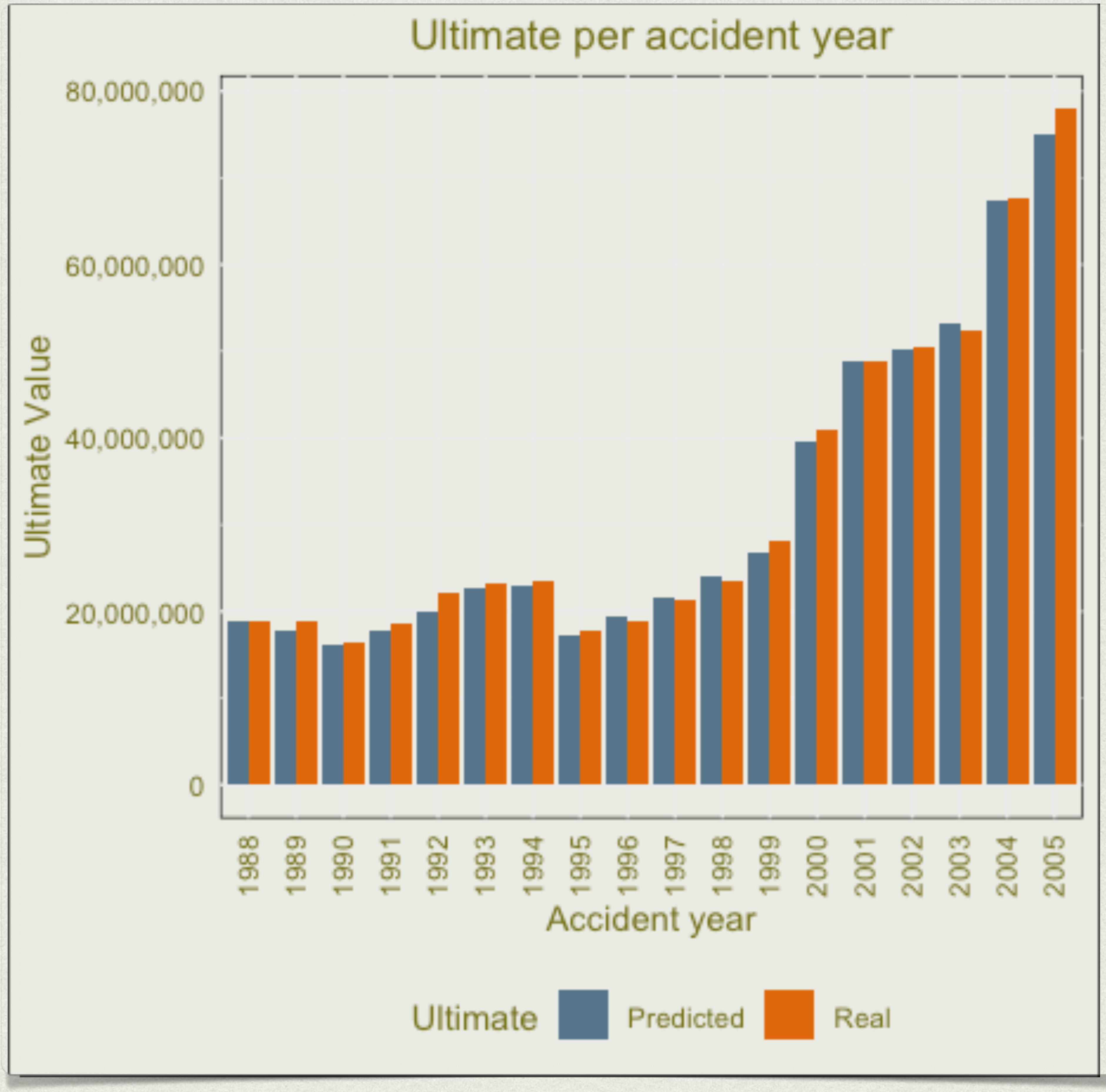
FINAL REMARKS

Or, what's next?

ON EXPLAINABILITY

Remarks on explainability:

- Are we holding traditional reserving models to the same standard?
- What does explainable really mean?



ON MODELLING CHOICES

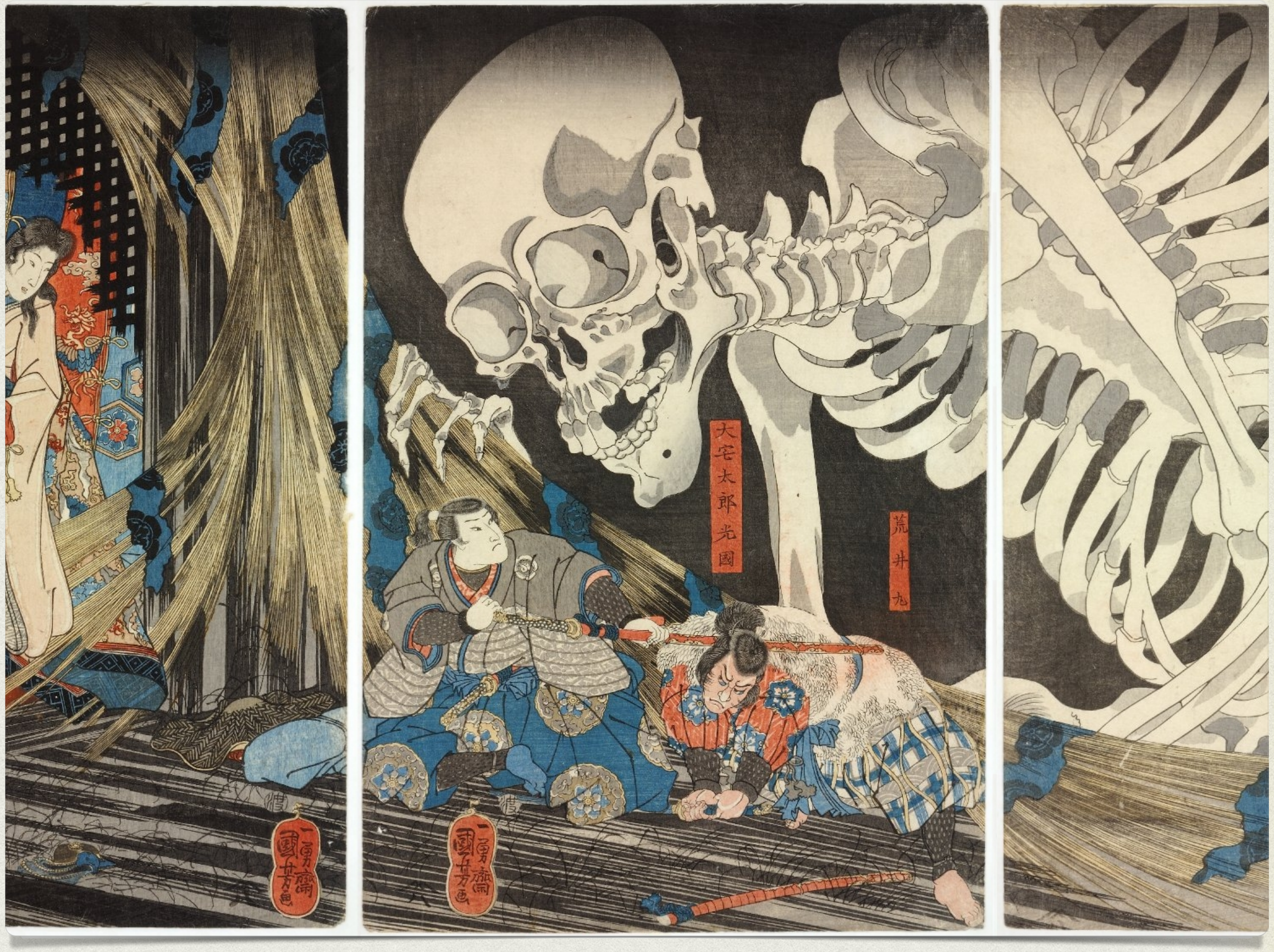
Things to think about:

- Boosting neural networks;
- Improve text analysis;
- Modelling of IBNyR.

ON UKIYO-E

"Living only for the moment, savouring the moon, the snow, the cherry blossoms, and the maple leaves, singing songs, drinking sake, and diverting oneself just in floating, unconcerned by the prospect of imminent poverty, buoyant and carefree, like a gourd carried along with the river current: this is what we call ukiyo."

Asai Ryōi , Ukiyo Monogatari



Takiyasha the Witch and the Skeleton Spectre, U. Kuniyoshi