2023 INTERNATIONAL CONGRESS OF ACTUARIES

BRIDGE TO TOMORROW

28 MAY – 1 JUNE 2023 · SYDNEY

Actuaries Institute Australia

IAA AAI International Actuarial Association
Association Actuarielle Internationale

Micro-reserving for Workers' Compensation claims

a case study from Kaggle

© Boosted Goose (Nelvis Fornasin, Attila Gulyas)

This presentation has been prepared for the 2023 International Congress of Actuaries.
The Actuaries Institute Australia Council wishes it to be understood that opinions put forward herein are not necessarily those of the Institute and the Council is not responsible for those opinions.

The "Actuarial Loss Prediction" Competition

- **Duration:** December 2020 to April 2021
- **Organizers:** 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

A quick overview of the data

- Details on the **insured person**:

Age	Gender	Marital Status	Dependent Children	Dependents Other
43	M	F	1	0

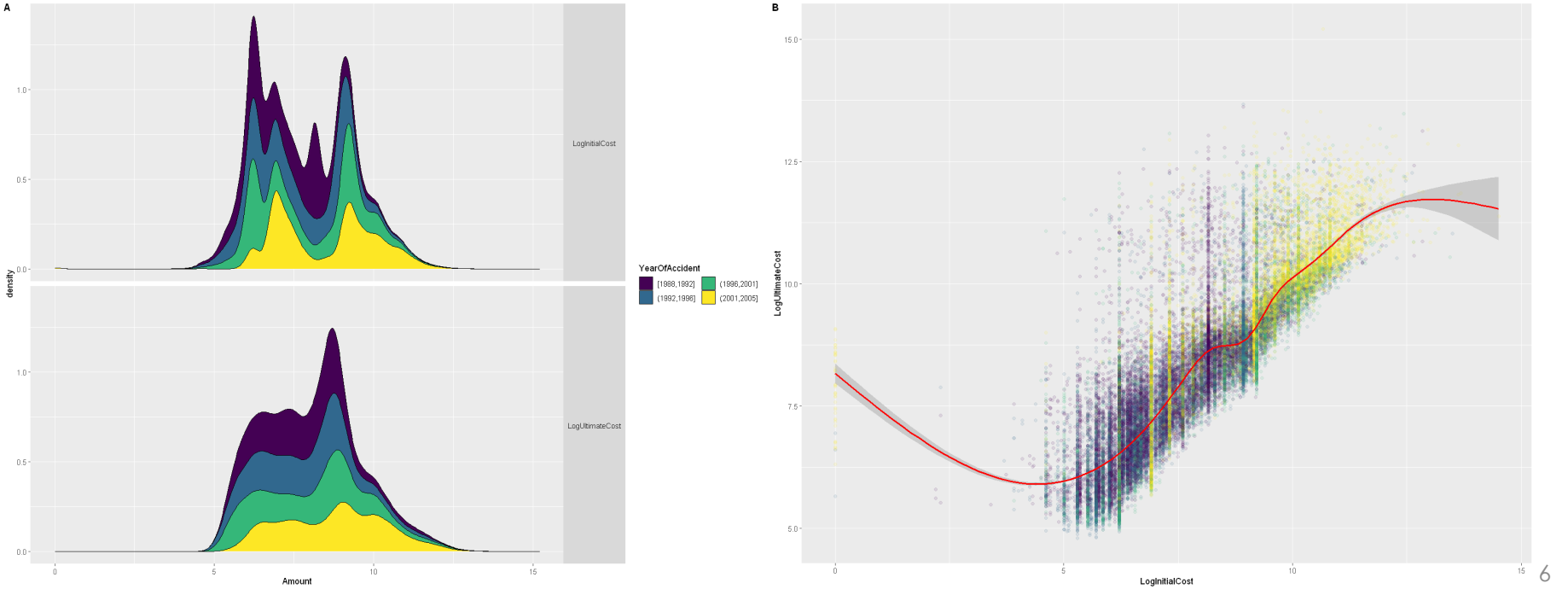
- Details on the **profession**:

Weekly Wage	Part Time/Full Time	Hours Worked per Week	Days Worked per Week
43	M	F	1

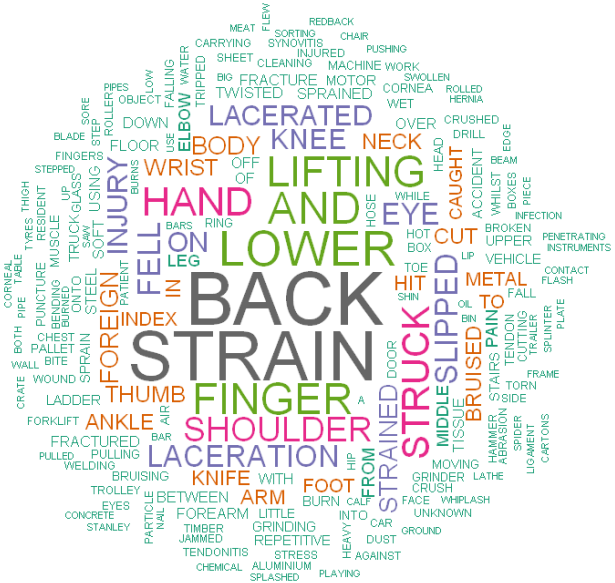
- Details on the **claim**:

Date and Time of Accident	Date Reported	Claim Description	Initial Incurred Costs
43	M	CUT ON SHARP EDGE CUT LEFT THUMB	1

Initial vs. Ultimate Claim Cost



Claim Descriptions - Wordcloud



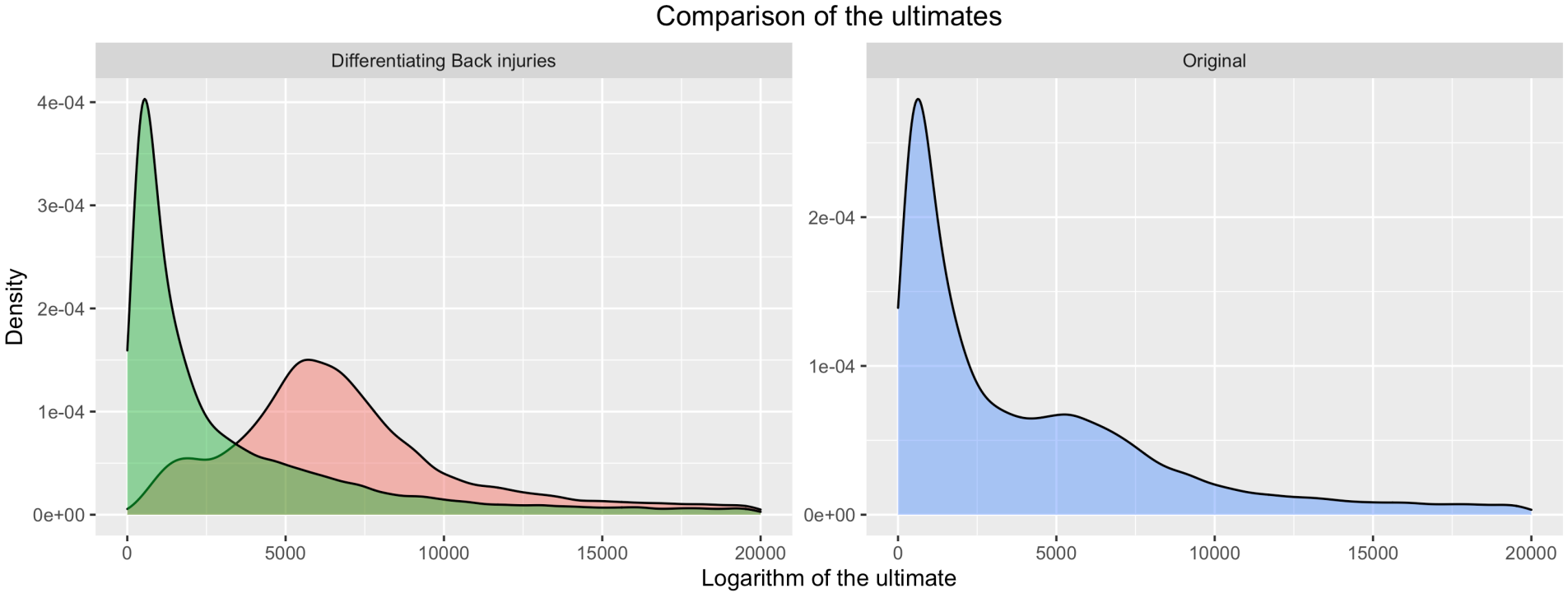
Remarks:

- Many different words
  Synovitis, Table, Particle, Ring...
- Body parts, types of accidents
  e.g. Back, Shoulder, Finger, Strain, Laceration

Important for the analysis:

- Remove stop words
  e.g. and, to, with, on
- Stem remaining words
  e.g. fall, falling, fell => fall
- Cluster the stemmed words
  e.g. eye, cornea => "eye" cluster

Claim Descriptions - Effects on the ultimate



Claim Descriptions - Analysis

The (synthetically generated) claim descriptions don't offer much grammatical structure, sometimes simply nonsensical, e.g. "TO RIGHT LEG RIGHT KNEE".

- Our approach to analysis:

– Remove stop words ("in", "on", ...);

– Lemmatization and stemming of words ("Feet" and "foot" are both mapped to "foot", "laceration" and "lacerated" to "lacer");

– Clustering of words according to context and ultimate;

– OHE for the most common words.

- In the end, we OHEncoded about 100 words and created 30 clusters.

Some details on our model

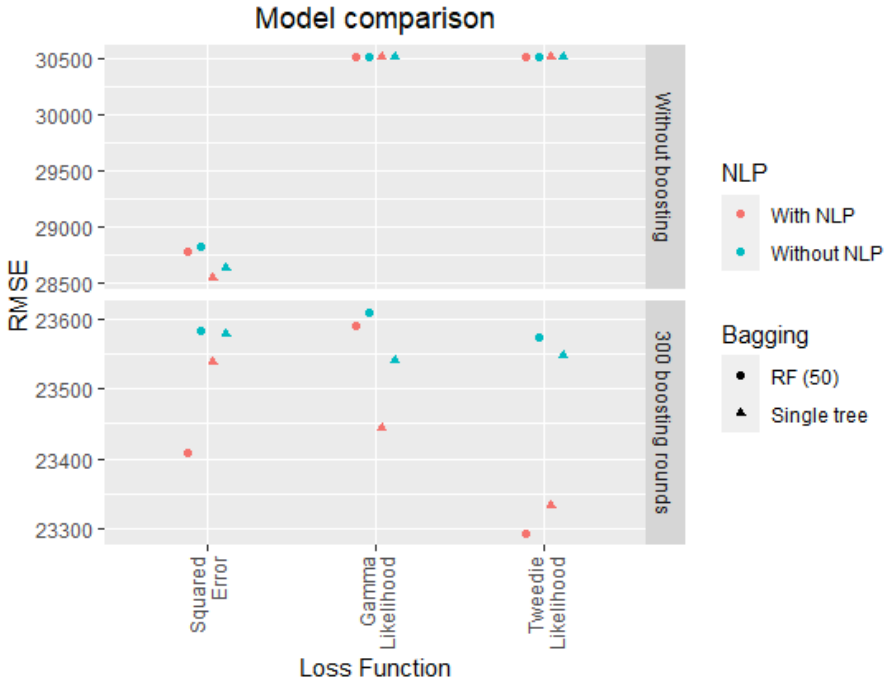
Our algorithm consisted of the following ensemble methods:

- **Boosting**: Gradient boosting with xgboost
- **Bagging**: Random forest as base learner
- **Voting**: Combination of models based on expert estimates

Adjusting the following model parameters significantly improved our position on the leaderboard:

- **num_parallel_tree**: Setting this parameter to a number greater than 1 allows the use of Random Forest as the base model;
- **monotone_constraints**: This parameter can be used, for example, to enforce a positive relationship between the number of children and the ultimate;
- **objective**: Setting to reg:gamma and reg:tweedie.

Model comparisons



Takeaways:

- Greatest impact is given by boosting
- Without boosting:
  - NLP does not influence the results
  - Distribution-based loss functions perform significantly worse
- Bagging does not consistently improve the result

Conclusion

Lessons learned:

- Feature engineering was more important than hyperparameter tuning;
- There is a Human Learning process which goes hand in hand with Machine Learning;
- Stacking models led to overfitting;
- Neural networks don't solve every problem.
- Large claims had a disproportionate impact on predictions. (MSE)

Thank you for your attention!