2023 Blue Shield of California Case Competition

Team 11
Yueran Yu, Ribhav Mittal, Danh Tran, James Tinajero

Agenda

1 Introduction

2 Manual Trend

3 Experience Trend

4 Credibility

5 XGBoost vs Our Model

6 Conclusion

1 Introduction

1 Introduction - Trend Prediction Logic

- **Weighted Average**
  - Weight of Dollars (Allowed) and Utilization
  - For markets / LOBs
  - Event-Modified
- **Linear Model**
- **Limited Fluctuation Method**

1 Introduction - Major Assumption

- Monthly Weights (Utilization, Allowed) are similar in 2021 and 2022.
- Historical Trends are fully credible.
- Half-Year Trends between 2 near years are similar to their Whole-Year Trends

1 Introduction - Major Adjustment

Manual Trend:

- Taking out 2019 Prof LOB because of extremely high irregular trends.
- Affected by COVID-19, taking out entire 2020 year data.

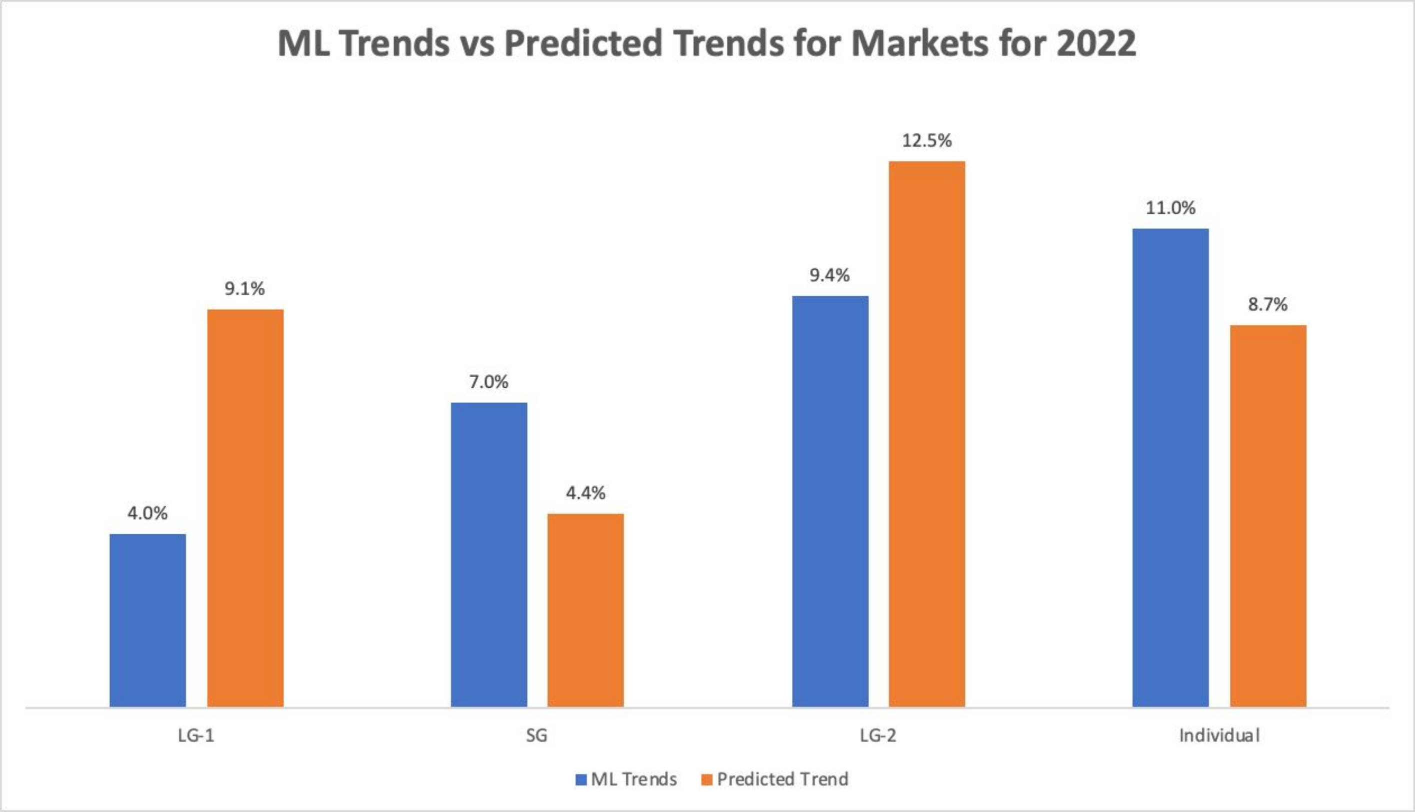
1

Introduction - Major Adjustment, Examples

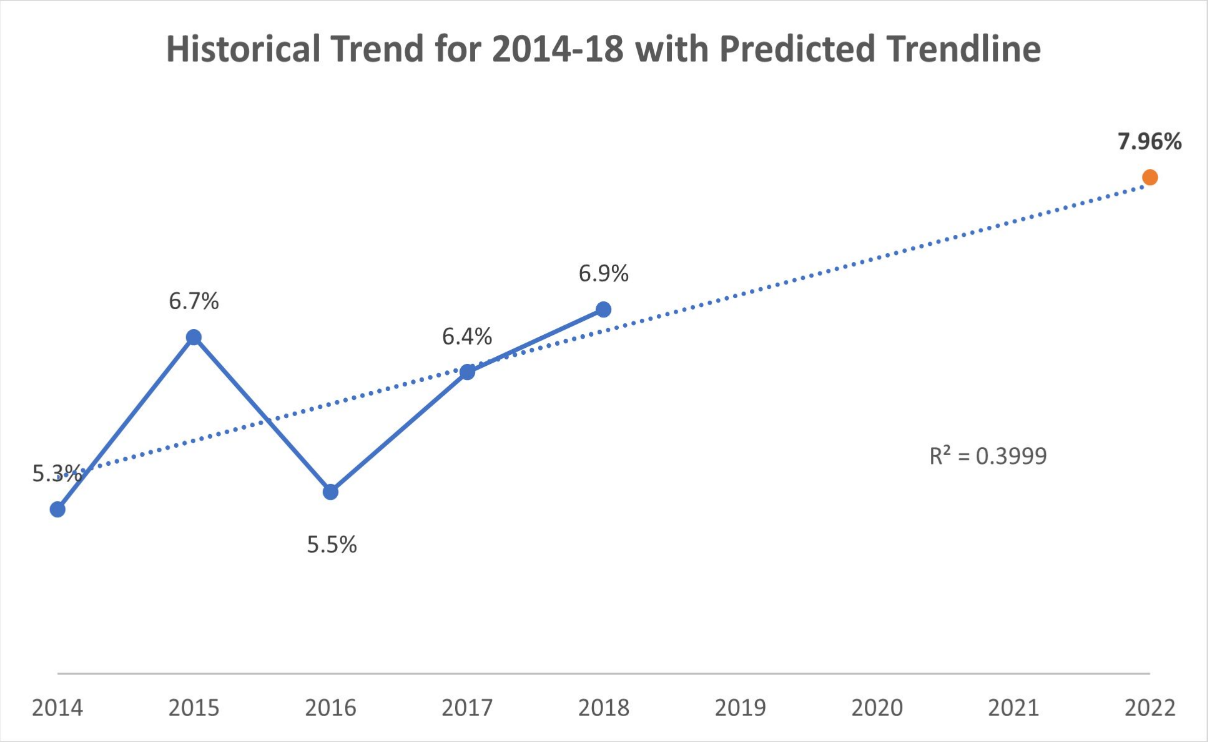
Year	Market	Type of Service	Unit Cost	Utilization
2020	LG-1	Prof	-41.0%	113.3%
2020	LG-2	Prof	-42.0%	90.6%
2020	Individual	Prof	-14.6%	38.1%
2020	Individual	Brand	16.0%	-2.1%
2019	LG-1	Prof	98.1%	3.9%
2019	LG-2	Prof	195.5%	240.5%
2019	SG	Prof	20.1%	110.7%
2019	Individual	Prof	33.6%	-16.7%

1

Introduction - Major Adjustment, Results



1 Introduction - Why Linear Model



R-Squared = 40%

Raw Predicted Trend:
7.96%

1

Introduction - R ANOVA Table and Residuals Assumptions proved

```r
model <- list("x" = c(2014, 2015, 2016, 2017, 2018), "y" = c(5.3, 6.7, 5.5, 6.4, 6.9))
model
linear_reg <- lm(y~x, data = model)
summary(linear_reg)
anova(linear_reg)

### Check Residuals Assumptions
# Check independence
plot(linear_reg$residuals)
abline(h=0, col = "red")
# Check mean of residual approaches 0
sum(linear_reg$residuals)/5
# Check residuals are normally distributed
hist(linear_reg$residuals, breaks = c(-1, -0.3, 0.4, 1.1))
# Check variance of residual approaches regression variance
y_bar <- mean(model$y)
sigma_squared <- sum((model$y - y_bar)^2) / 5
  # Residual Variance from anova: 0.41033, compare variances.
abs((0.41033 - sigma_squared) / sigma_squared) < 0.01 # difference is very small
```

1

Introduction - Event Trend

3 affecting events

- **2021 deferred claims event:**
  underlying units = data / (1 - deferred rate)

- **2022 events:**
  Using monthly weights and half-year weights to annualize given event trends.

1

Introduction - 2022 Event Trend

LOBs	Trend Type	Annualized Event Trend
Generic	Utilization	0.00%
	Allowed	-3.67%
	Unit Cost	-3.67%
Brand	Utilization	0.78%
	Allowed	-0.30%
	Unit Cost	-1.07%

2 Manual Trend

2 Manual Trend

$$\sum\left(\text{Market Weights} \times \text{Raw Predicted Trend}\right) = \text{Predicted Underlying Trend}$$

2 Manual Trend

Predicted 2022 Underlying Trend		
Market	**Unit Cost**	**Utilization**
Ancillary	0.08	-0.08
Brand	0.10	-0.01
Generic	0.00031	0.0056
IP	0.07	0.02
OP	0.04	0.02
Professional	0.03	0.09

3 Experience Trend

3

Experience Trend

Acquired Event-modified data → **Predicted annual trend** → **Predicted 2022 second half Underlying units**

Predicted 2022 PMPM using ratio → 2021, 2022 Underlying & Trend Units → Experience Trend

3 Experience Trend

Predicted 2022 Second Half Underlying Units		
Benefit Type	**Sum of Utilization**	**Sum of Allowed**
Ancillary	4,456,449	324,381,572
Brand	1,062,376	795,455,098
Generic	6,774,547	158,384,870
Inpatient	166,685	1,103,747,852
Outpatient	3,609,366	1,249,756,295
Professional	11,984,069	1,147,198,884
Total	**28,059,099**	**4,773,219,413**

3 Experience Trend

Acquired Event-modified sums → Predicted annual trend → Predicted 2022 second half Underlying units

Predicted 2022 PMPM using ratio → **2021, 2022 Underlying & Trend Units** → **Experience Trend**

3 Experience Trend

$$\text{2022 first half member months} \times \frac{\text{2021 whole year member months}}{\text{2021 first half member months}}$$

2022 Estimated Member Months	
Ancillary	15,609,074
Brand	15,609,074
Generic	15,609,074
IP	15,609,074
OP	15,609,074
Professional	15,609,074
Total	**93,654,444**

3

Experience Trend

Utilization Trend			
Benefit Type	**Underlying Trend**	**Event Trend**	**Adjusted Core Trend**
Ancillary	6.5%	0.0%	6.5%
Brand	-2.4%	0.4%	-2.1%
Generic	3.5%	0.0%	3.5%
Inpatient	10.5%	0.0%	10.5%
Outpatient	13.4%	0.0%	13.4%
Professional	7.8%	0.0%	7.8%
Total	**6.8%**	**0.0%**	**6.8%**

Util/K Calculation:

$$\frac{\text{(Sum of Utilization)} * 12{,}000}{\text{Sum of Member Months}}$$

3

Experience Trend

Unit Cost Trend			
Benefit Type	**Underlying Trend**	**Event Trend**	**Adjusted Core Trend**
Ancillary	0.3%	0.0%	0.3%
Brand	15.7%	-0.7%	14.9%
Generic	10.6%	-3.7%	6.5%
Inpatient	-9.6%	0.0%	-9.6%
Outpatient	-4.9%	0.0%	-4.9%
Professional	0.9%	0.0%	0.9%
Total	**0.1%**	**-0.2%**	**-0.1%**

Unit Cost Calculation:

$$\frac{\text{Sum of Allowed Dollars}}{\text{Sum of Utilization}}$$

3 Experience Trend

PMPM Trend			
Benefit Type	**Underlying Trend**	**Event Trend**	**Adjusted Core Trend**
Ancillary	6.8%	0.0%	6.8%
Brand	12.9%	-0.3%	12.5%
Generic	14.4%	-3.7%	10.2%
Inpatient	-0.1%	0.0%	-0.1%
Outpatient	7.9%	0.0%	7.9%
Professional	8.7%	0.0%	8.7%
Total	**6.9%**	**-0.2%**	**6.7%**

PMPM Calculation:

(1 + Utilization Trend) * (1 + Unit Cost Trend) - 1

4 Credibility

4

Limited Fluctuation Method

Credibility-Weighted Rate = (Z) * (Observed Rate) + (1 - Z) * (Prior Rate)

Observed Rate = Experience Trend

Prior Rate = Manual Trend

4

Square Root Formula

$$Z = \sqrt{\frac{N_{AC}}{N_{FC}}}$$

N_{AC} = sum of member months for 2022
N_{FC} = sum of member months for 2014-2019

$$Z = \sqrt{\frac{93{,}654{,}445.86}{595{,}079{,}473.6}} = \mathbf{0.3967}$$

4

Weighted-Credibility Trend

Weighted-Credibility Trend = (0.39) * (Experience Trend) + (1 - 0.39) * (Manual Trend)

4 Weighted-Credibility Trend

Utilization

Benefit Type	WCT
Ancillary	-2.0%
Brand	-1.0%
Generic	1.7%
Inpatient	5.3%
Outpatient	6.4%
Professional	8.6%
Total	**4.7%**

Unit Cost

Benefit Type	WCT
Ancillary	5.2%
Brand	11.0%
Generic	0.4%
Inpatient	0.4%
Outpatient	0.6%
Professional	1.9%
Total	**3.1%**

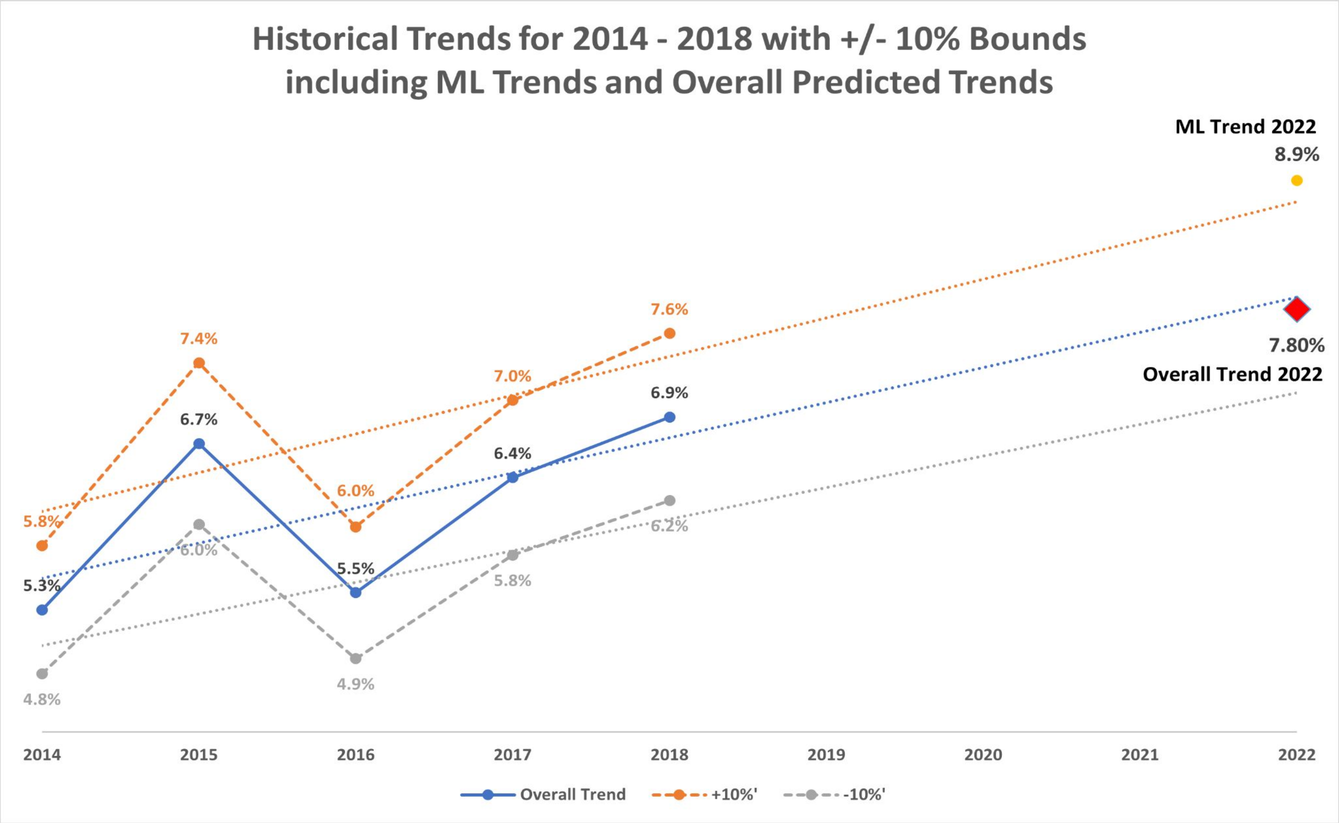
PMPM

Benefit Type	WCT
Ancillary	2.8%
Brand	9.8%
Generic	2.2%
Inpatient	5.4%
Outpatient	6.9%
Professional	10.6%
Total	**7.8%**

5 XGBoost vs Our Trend

5

XGBoost ML Trend vs Weighted Credibility Trend



ML Trend 2022

8.86 %

Linear Trendline

7.96 %

Weighted Credibility Trend

7.80 %

5 Advantages & Disadvantages of XGBoost

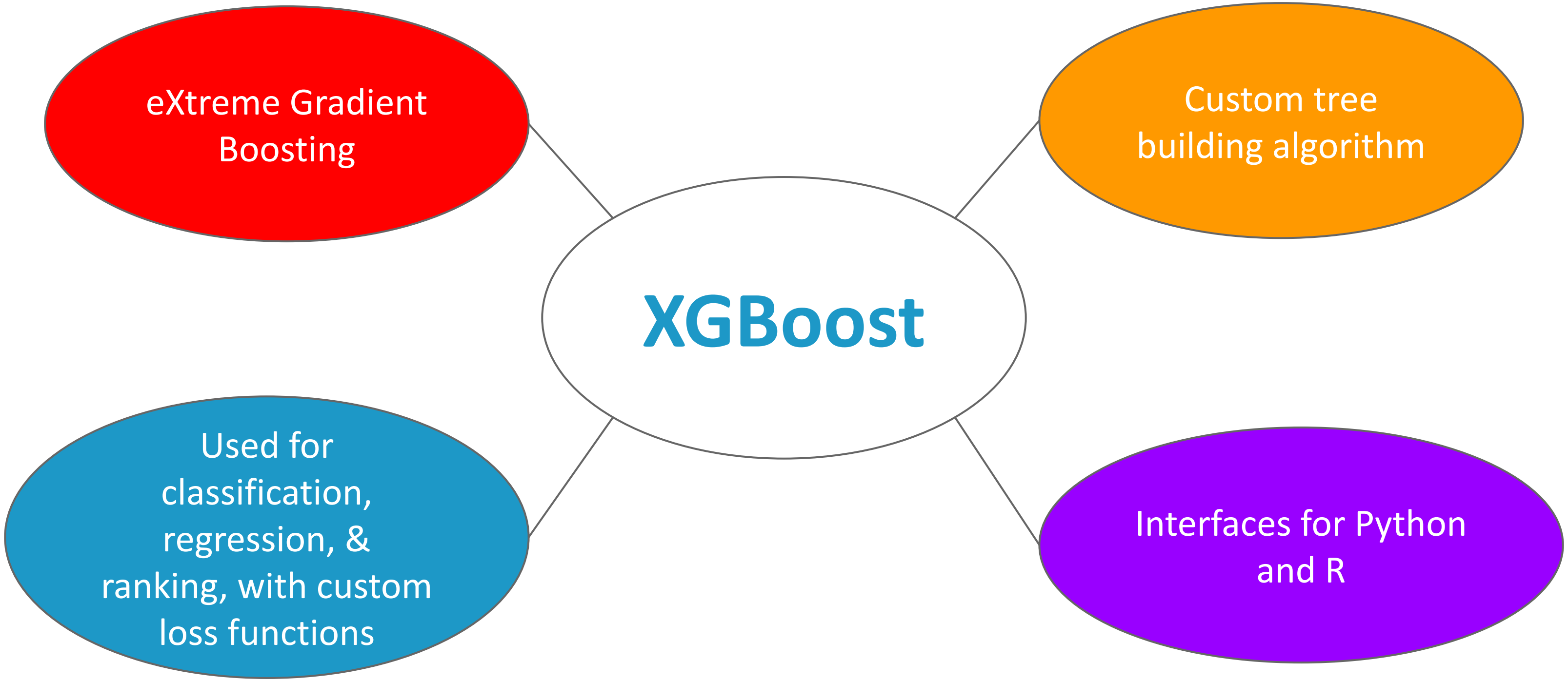
Pros

- Works well with large data sets containing more than thousands of rows
- Efficient handling of missing data
- Regularization to avoid overfitting the data

Cons

- Doesn't perform well on unstructured data
- Sensitive to outliers

5 Future of XGBoost in Actuarial Work



6

Conclusion

6 Conclusion

- XGBoost has limitation in prediction because of its high sensitivity, but it performs well when taking care of missing data

- **Advice**:
    - XGBoost should become an assistance for actuaries, however, data selection, analysis, and scrutinization by actuaries are necessary for a better prediction.

Thank You!

Team 11