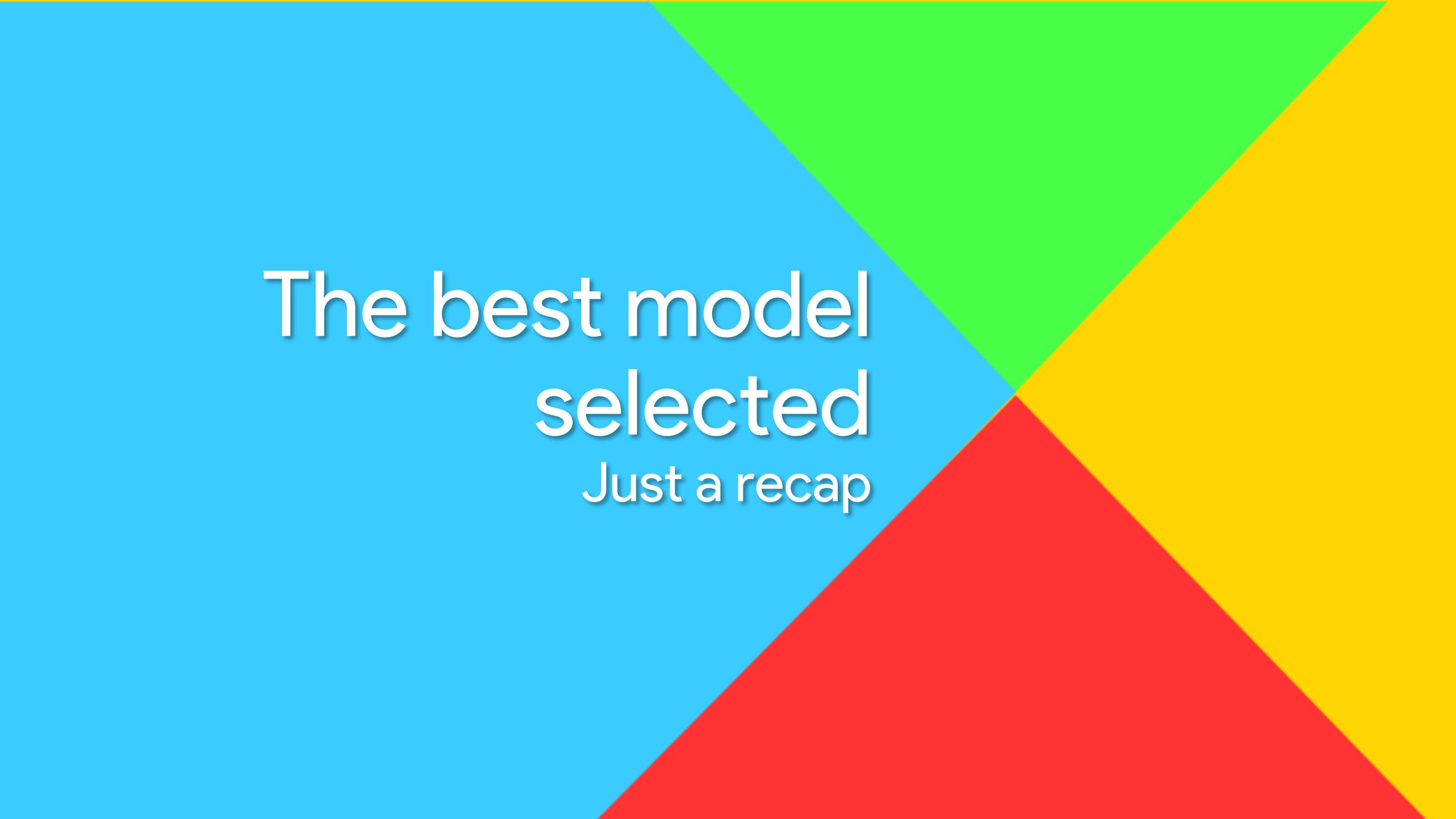


# Google Play

*Store Apps Rating*

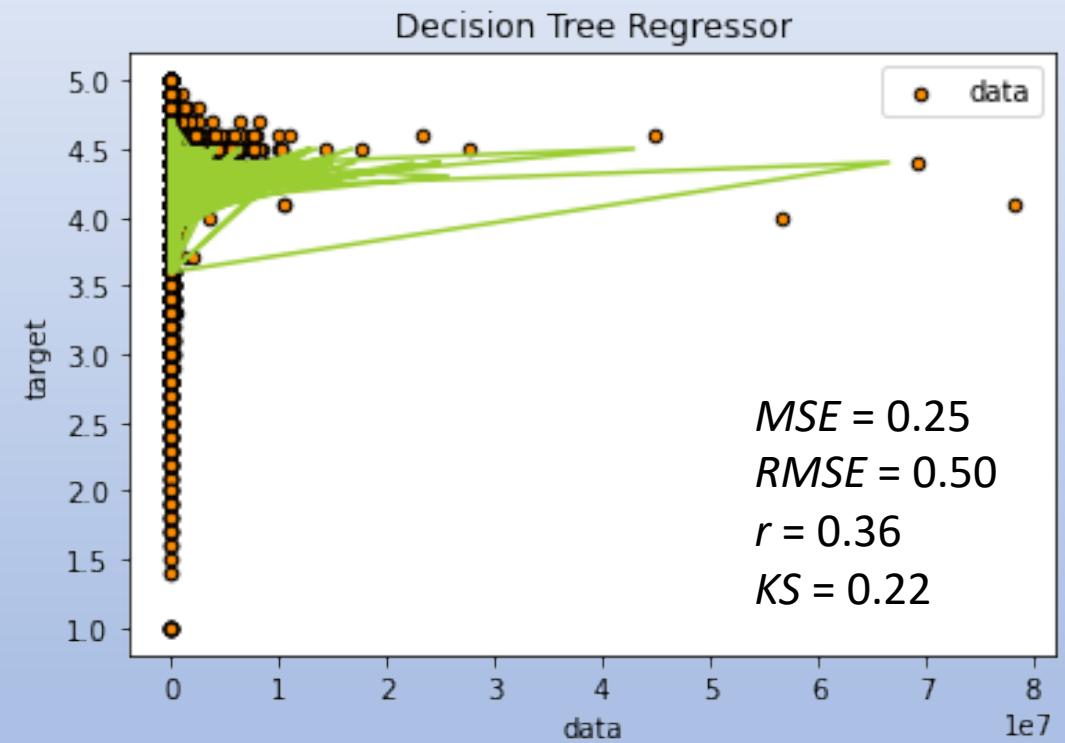
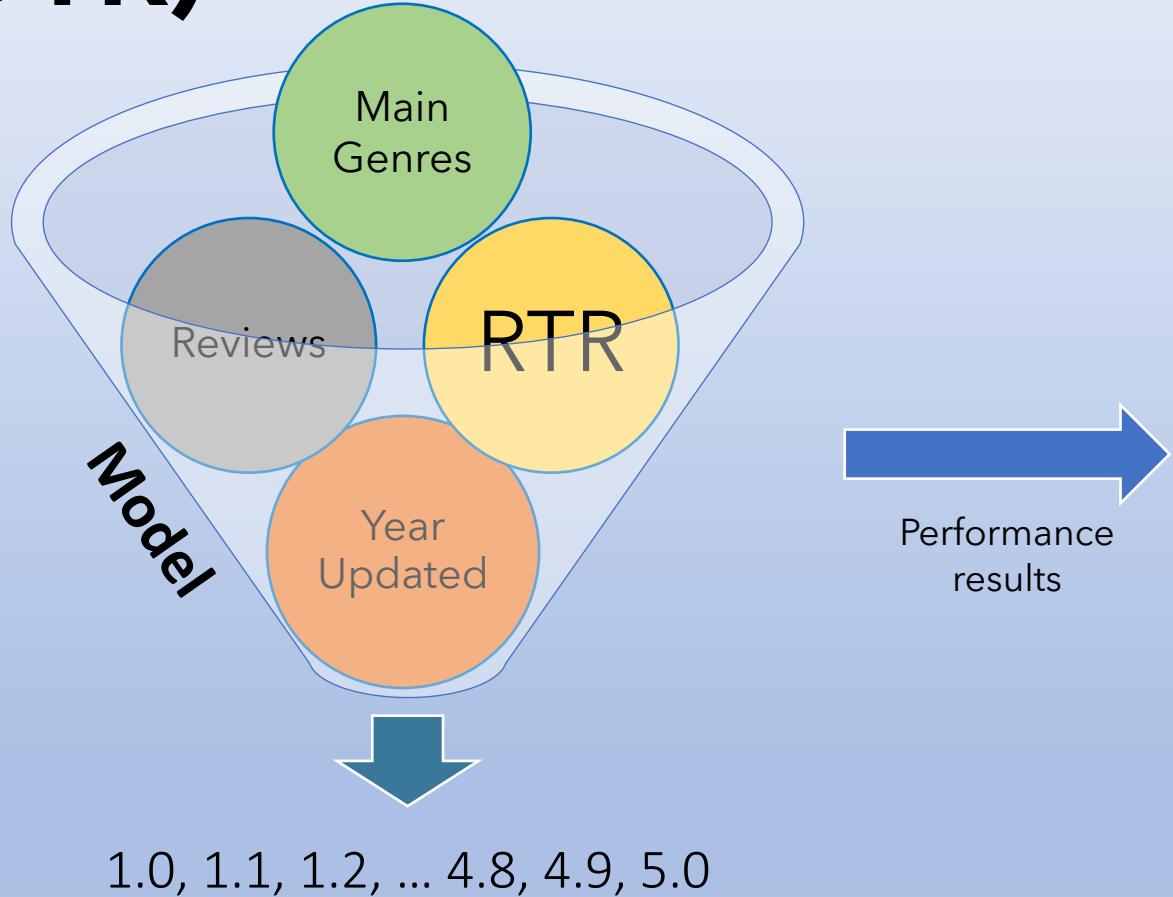




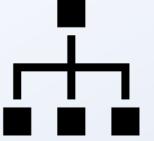
# The best model selected

Just a recap

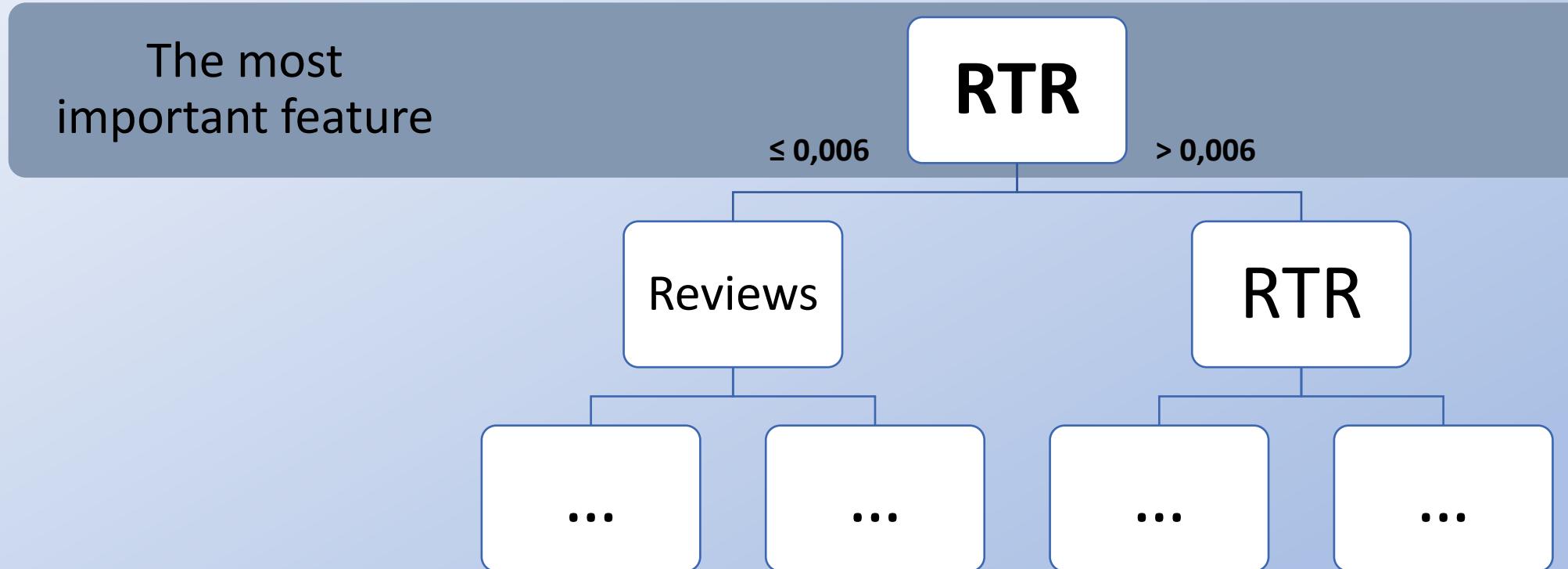
# *Decision Tree Regressor* (DTR)



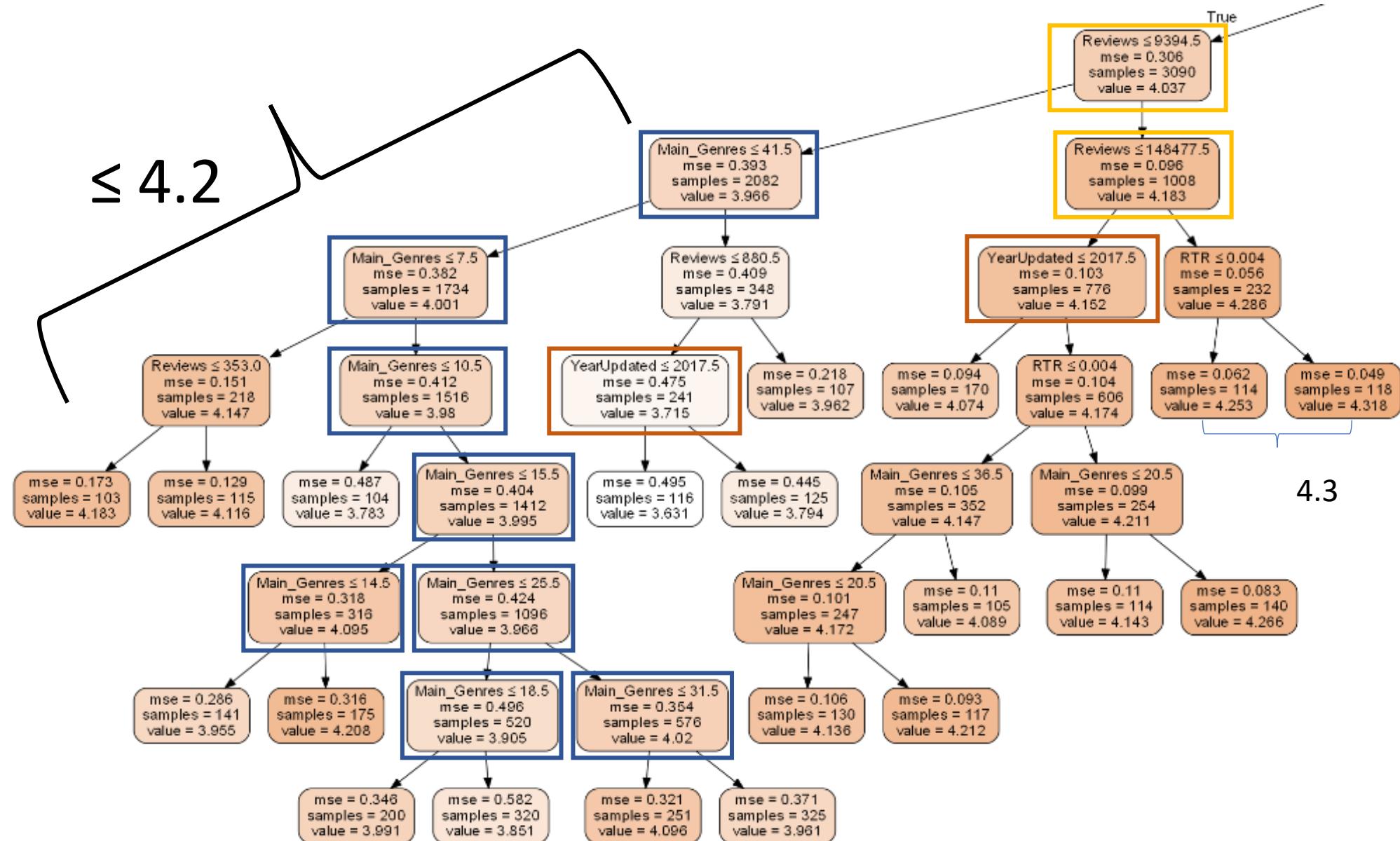
# Global explanation



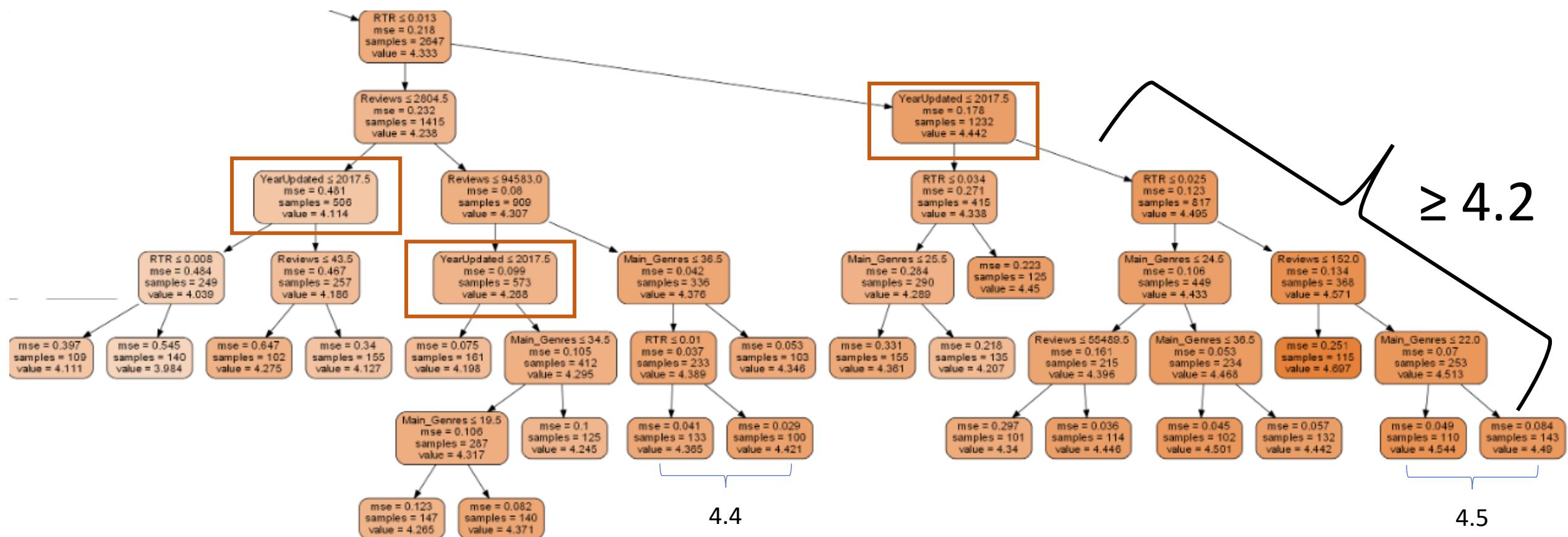
# How the tree is made?

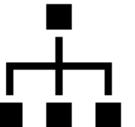


# Left main branch

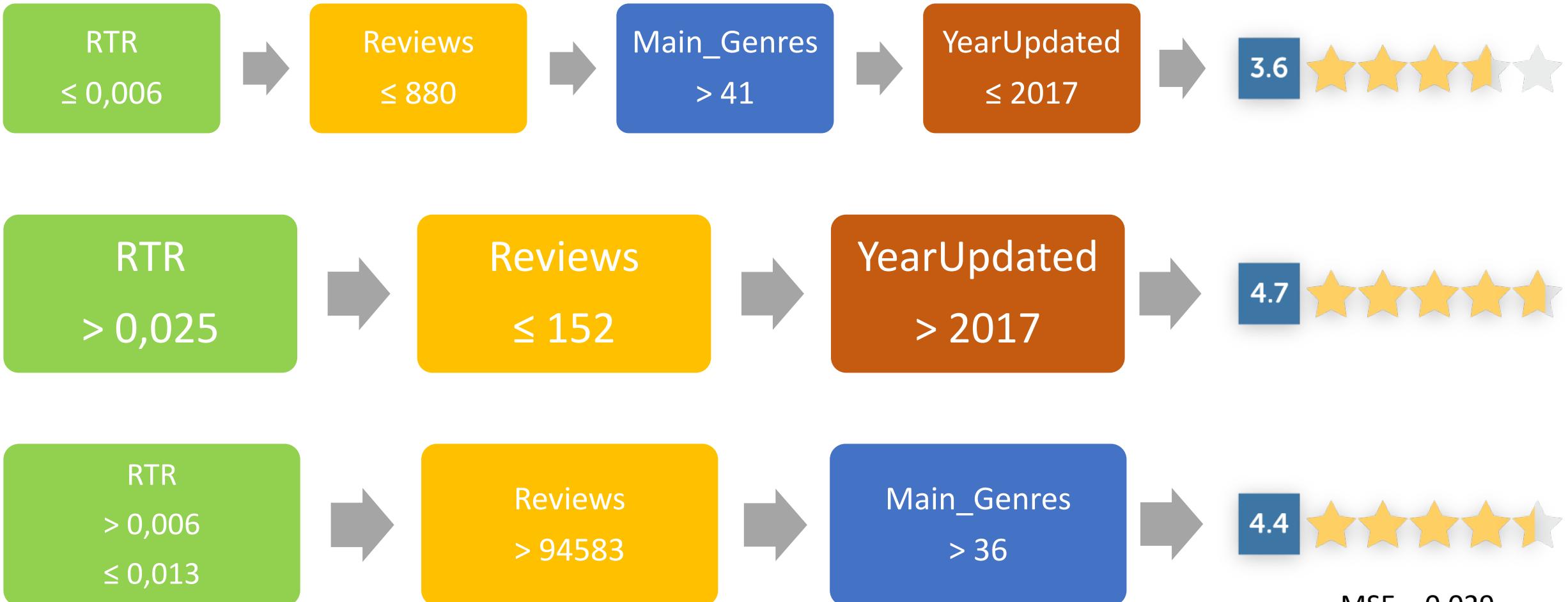


# Right main branch

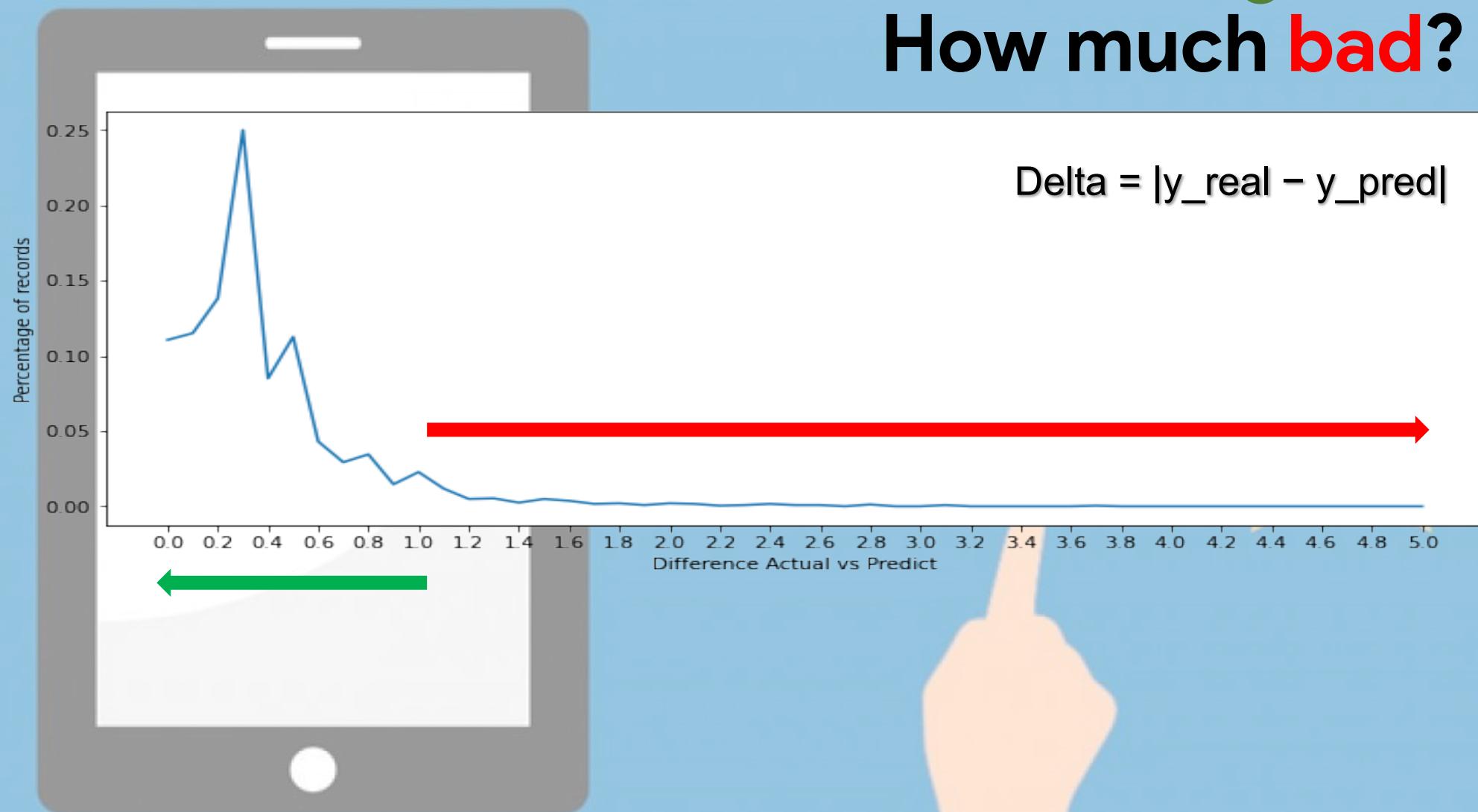




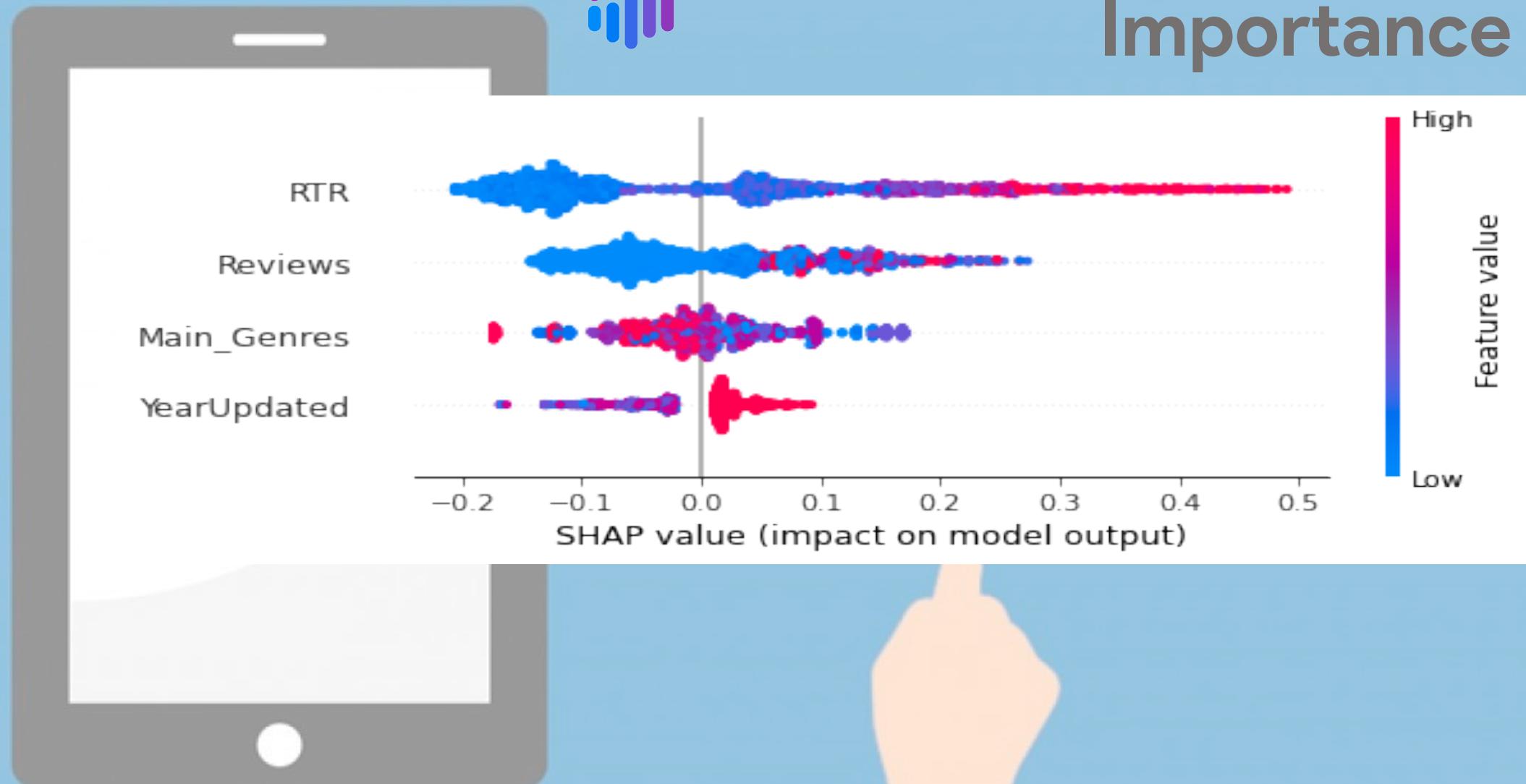
# Some prediction rules



# How much **good**? How much **bad**?



# SHAP Feature Importance

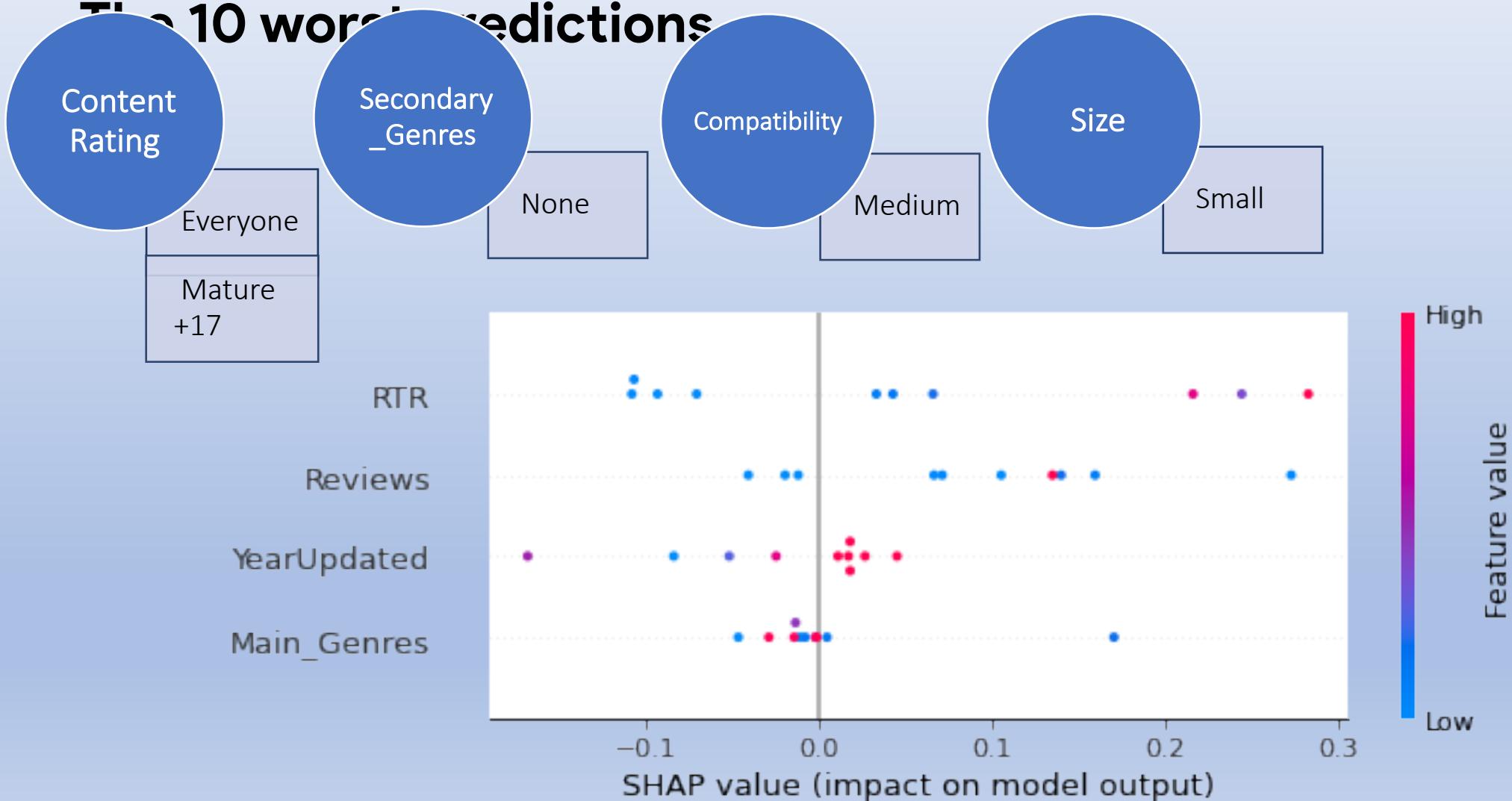


# Local explanation



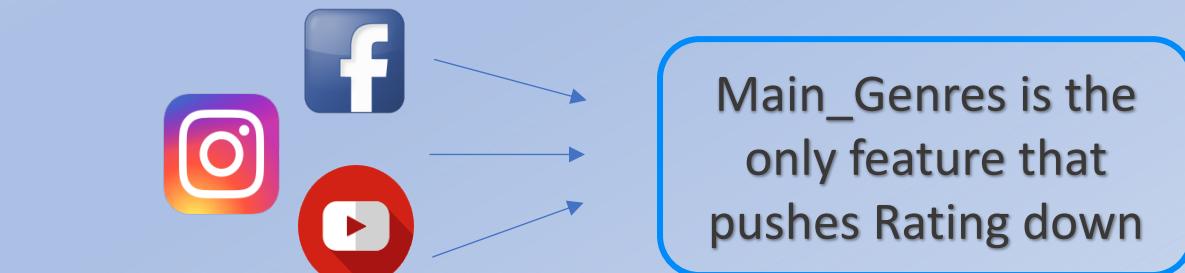
# SHAP for Local explanation

## The 10 worst predictions



# Some of the most popular Apps of 2018

<https://trak.in/tags/business/2018/07/05/which-are-the-most-downloaded-apps-games-in-2018-check-full-list/>



base value

4.175

4.225

4.250

4.275

4.300

4.325

4.350

4.375

4.400

higher ↛ lower  
f(x)  
**4.42**

YearUpdated = 2018.0

Main\_Genres = 29.0

RTR = 0.01166577224852421

higher ↛ lower

f(x)

**4.25**

Reviews = 11667403.0



4.05

4.10

4.15

4.20

4.25

4.30

4.35

YearUpdated = 2018.0

Reviews = 909226.0 RTR = 0.0022714187478409336



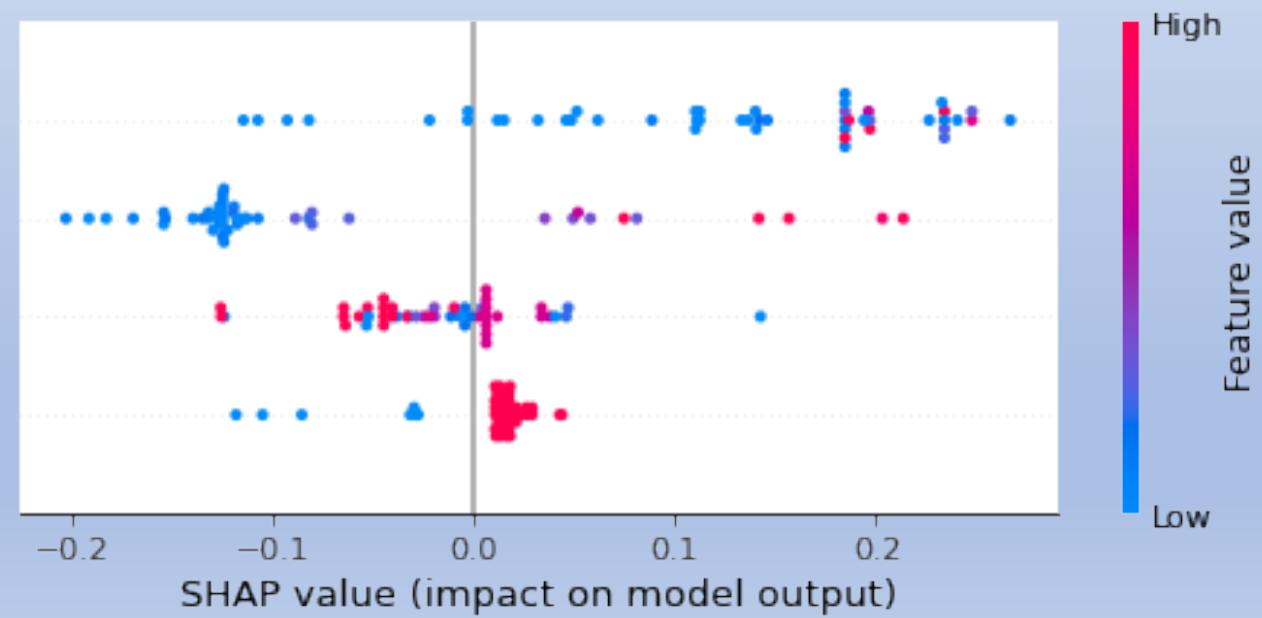
- ✓ These kinds of apps are very well-known to be very useful and among the best tools, so **Reviews** may don't affect much their ratings;
- ✓ **Tools** and **Travel & Local** apps push rating down.

Reviews

RTR

Main\_Genres

YearUpdated



# Our Best Apps of 2018

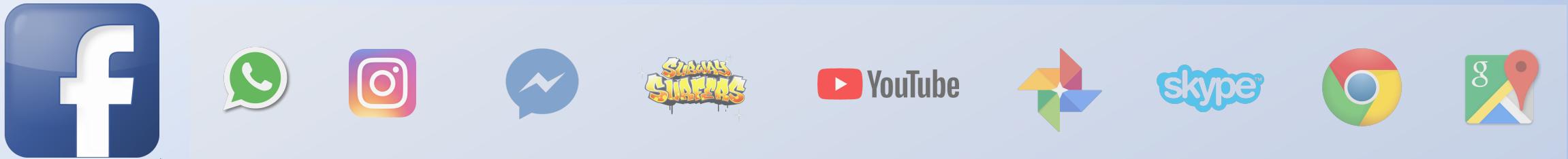


The most  
downloaded  
ranked by  
RTR



# LOcal Rule-based Explanations (LORE)

The top 10 apps

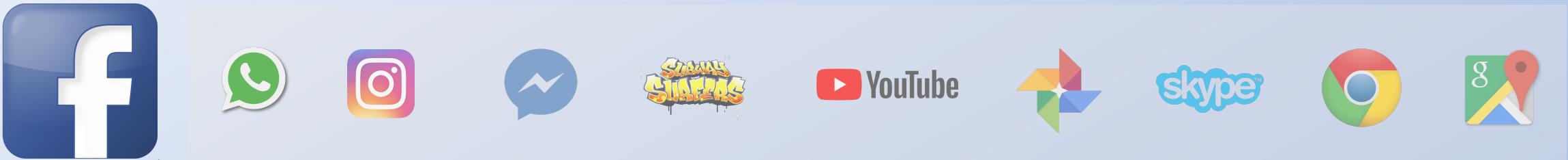


***Decision rule***

$r = \{RTR \leq 0.06, RTR > 0.01, Main\_Genres > 35.50, YearUpdated > 2017.38, Reviews > -551430.47\}$

# LLocal Rule-based Explanations (LORE)

The top 10 apps



**Decision rule**

$$r = \{RTR \leq 0.06, RTR > 0.01, \text{Main\_Genres} > 35.50, \text{YearUpdated} > 2017.38, \text{Reviews} > -551430.47\}$$

**Counter rules**





# *LOcal Rule-based Explanations* (LORE)

## The top 10 apps



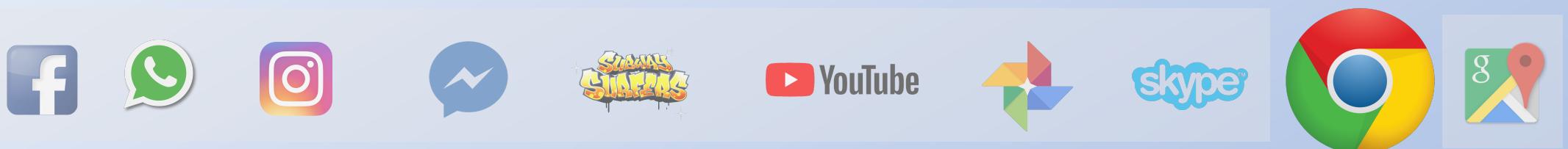
***Decision rule***

$$r = \{RTR \leq 0.00, \text{Reviews} > 131709,89\}$$



# LOcal Rule-based Explanations (LORE)

The top 10 apps



**Decision rule**

$$r = \{RTR \leq 0.00, \text{Reviews} > 131709,89\}$$

**Counter rules**





# Local Rule-based Explanations (LORE)

## The 10 worst predictions

Delta =  $|y_{\text{real}} - y_{\text{pred}}|$

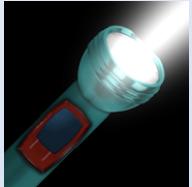


Rank	Explanatory Rule	Pred.	Real
1	RTR > 0.02, Reviews $\leq$ 144.00, YearUpdated > 2017.52	4.7	1.0
2	RTR $\leq$ 0.01, Reviews $\leq$ 20377.50, Reviews > -3345422.75, Main_Genres $\leq$ 31.48, Main_Genres > 25.42	4.1	1.0
3	RTR $\leq$ 0.00, Reviews $\leq$ 20377.50, Main_Genres $\leq$ 31.41, Main_Genres > 25.50	4.1	1.0
4	RTR > 0.03, Reviews $\leq$ 152.50, YearUpdated > 2017.52	4.7	1.9
5	RTR $\leq$ 0.02, RTR > 0.01, Main_Genres $\leq$ 25.50, Reviews $\leq$ 54144.00	4.4	1.7
6	RTR $\leq$ 0.01, RTR > 0.01, Reviews $\leq$ 2254.50, YearUpdated $\leq$ 2017.55	4.0	1.4
7	RTR $\leq$ 0.01, Main_Genres $\leq$ 42.50, Main_Genres > 41.87, YearUpdated > 2017.83, Reviews $\leq$ 882.00	3.8	1.0
8	RTR $\leq$ 0.01, Reviews $\leq$ 10084.00, Main_Genres $\leq$ 8.89, Main_Genres > 7.50	3.8	1.0
9	RTR > 0.03, YearUpdated $\leq$ 2017.55	4.5	2.0
10	RTR $\leq$ 0.00, Main_Genres $\leq$ 15.49, Main_Genres > 14.52, Reviews $\leq$ 20556.00, YearUpdated > 2014.50	4.3	1.8



# LOcal Rule-based Explanations (LORE)

## The 10 worst prediction



**DS FLASHLIGHT**



**CR TRACKER FOR CHESTS**





# LOcal Rule-based Explanations (LORE)

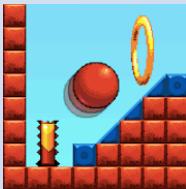
## The 10 worst prediction



**SHARED CARE CO-PARENTING**

**RTR**  $\leq 0.01$

**4.1**



**BOUNCE CLASSIC**

**RTR**  $> 0.01$

**4**

# Conclusions

# Task proposal



Predict App Rating





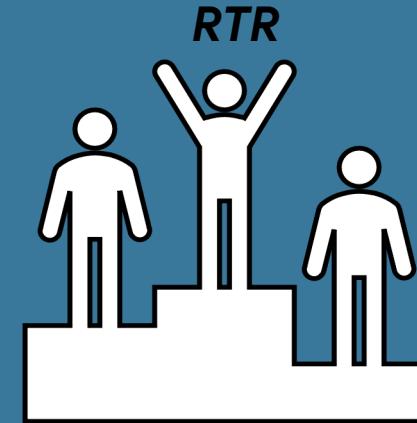
# New features engineered

Both Stores

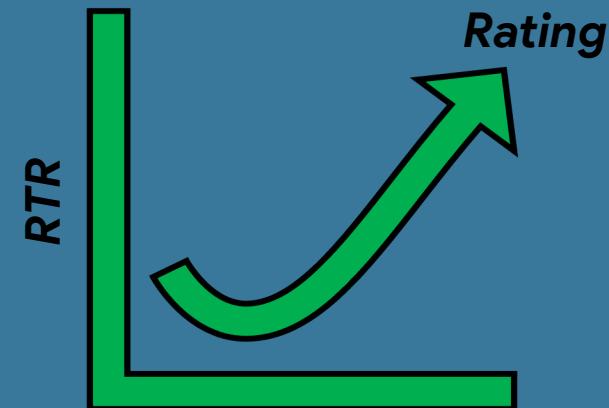


does not  
influence the  
Rating prediction

RTR



The most  
important



# *Is it a reliable model?*



It tends to predict Rating values perfectly or by slightly deviating from the real values

- Inability to predict **lower classes** is probably due to the scarcity of these values in the dataset and the lack of variables that allow better discrimination between high and low values.
- **Ordinal Regression:** In the middle between a classification and a regression.

*Thanks for the  
attention!*

