

# Google Play

*Store Apps Rating*



# Model Creation & Evaluation Metodology

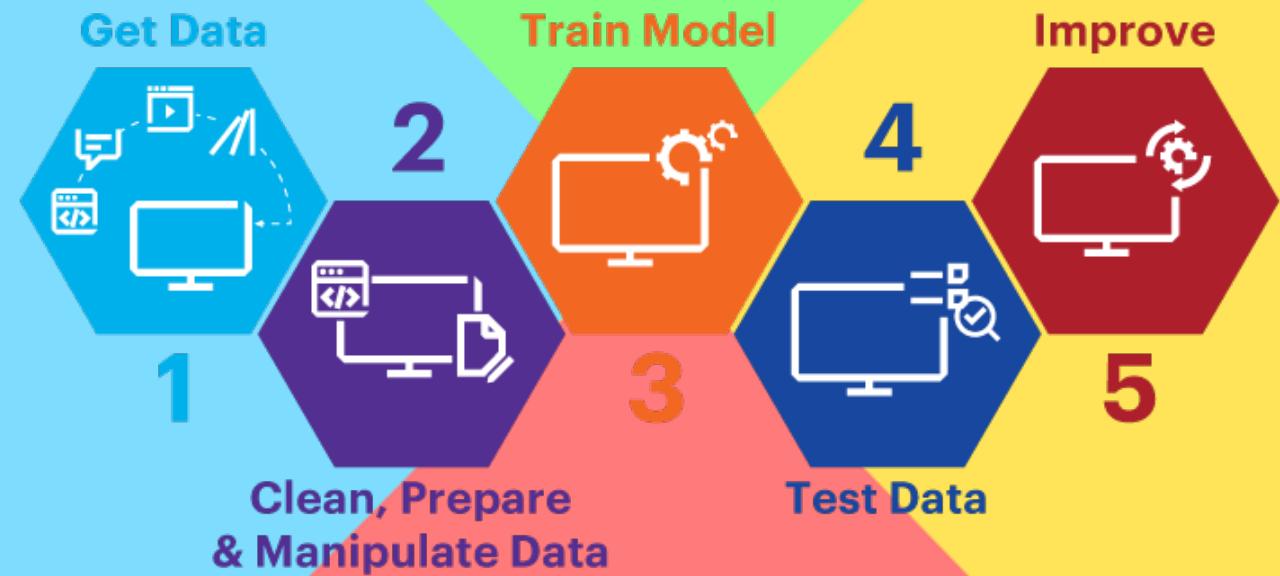
# Model Creation & Evaluation Metodology

## Models implemented:

- Decision Tree Regression
- Random Forest Regression
- Ordinal Regression
- Support Vector Machine
- Convolutional Neural Network

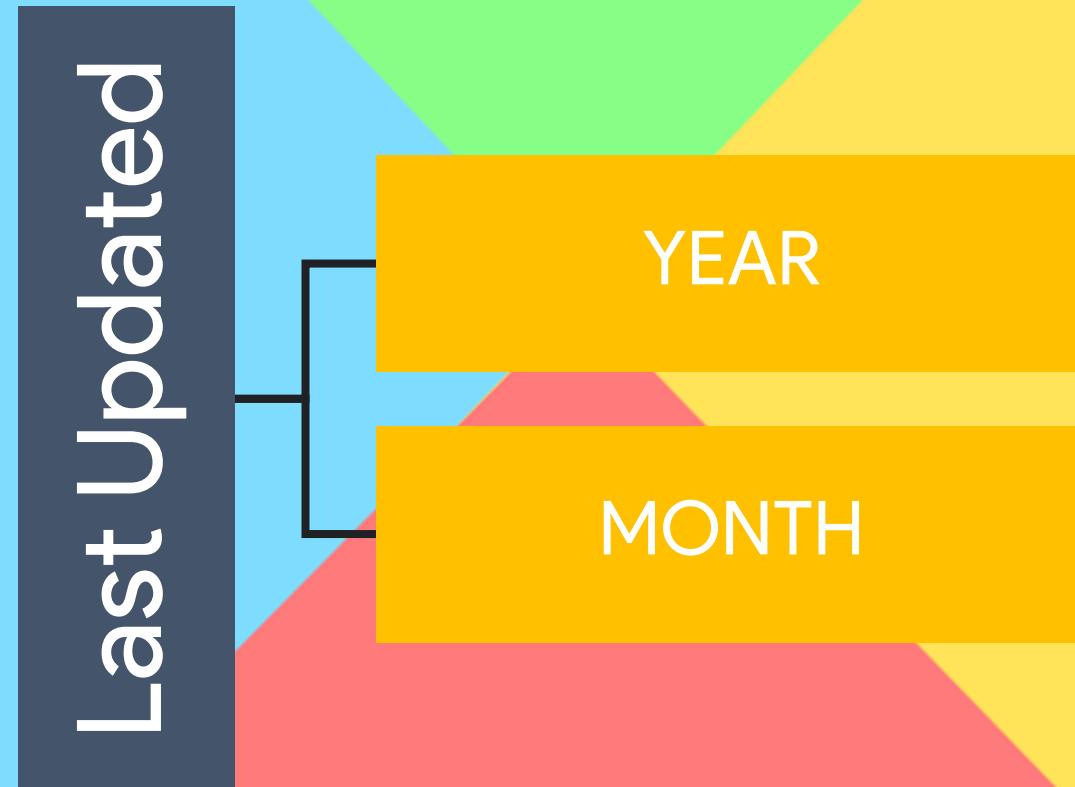
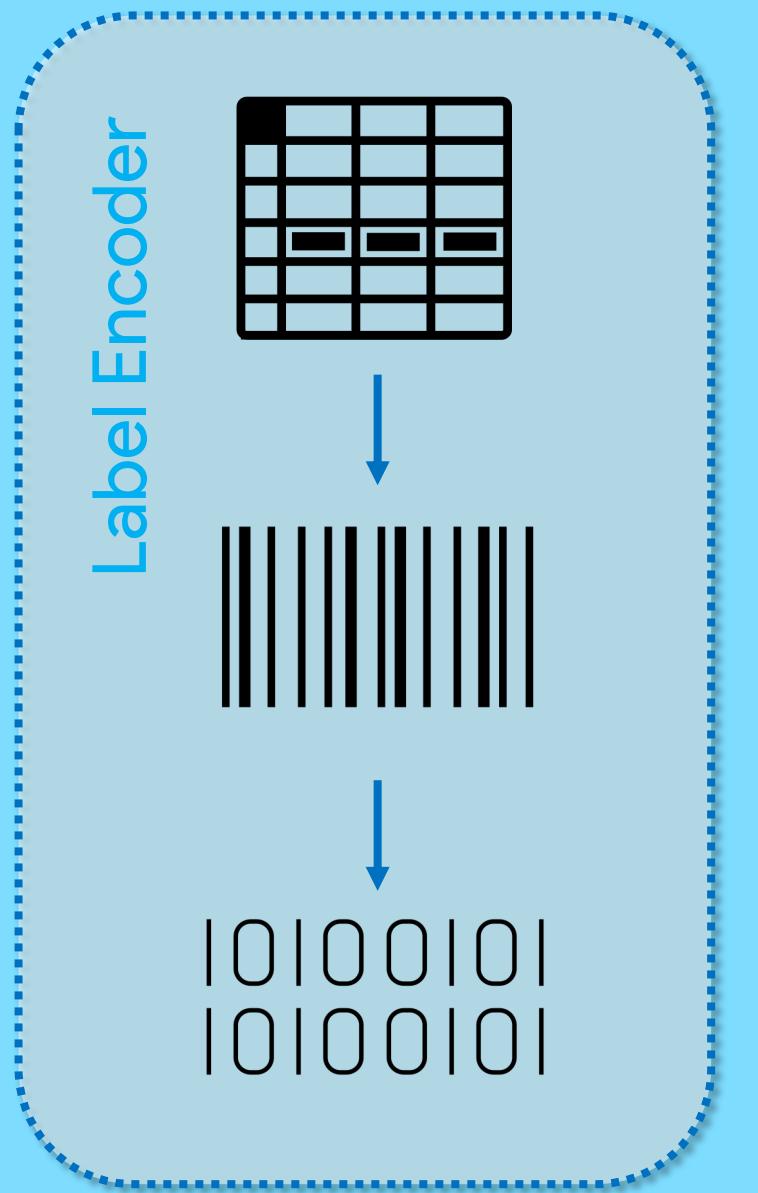
## Metrics For Evaluation:

- Mean Square Error (**MSE**)
- Mean Absolute Error (**MAE**)
- Root Mean Squared Error (**RMSE**)
- Kolmogorov - Smirnov statistics (**KS**)
- Pearson Correlation Coefficient (**r**)



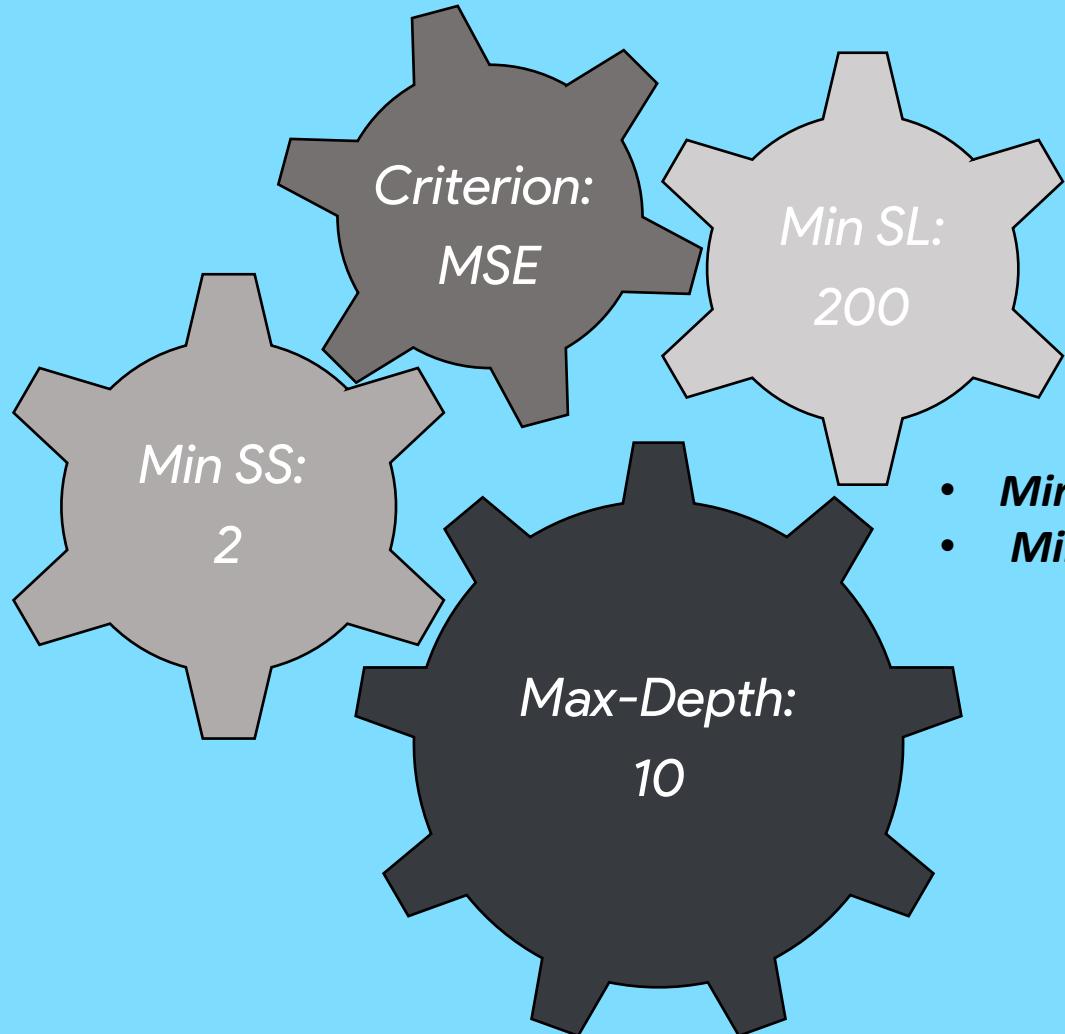
# Decision Tree Regressor

# Data preprocessing



# Hyper-parameters tuning:

## Grid search approach



- ***MinSS*:** Min Samples Split
- ***MinSL*:** Min Samples Leaf

# G Features Importance

1		RTR 4,4★	8		Price	14		Secondary_Genres
2		Reviews	9		MonthUpdated	15		Content Rating
3		Main_Genres	10		Both_Stores	16		Installs
4		YearUpdated	11		Compatibility	17		App
5		Android Ver	12		Is_Game			
6		Current Ver	13		Is_Family			
7		Size						



# Features Selection



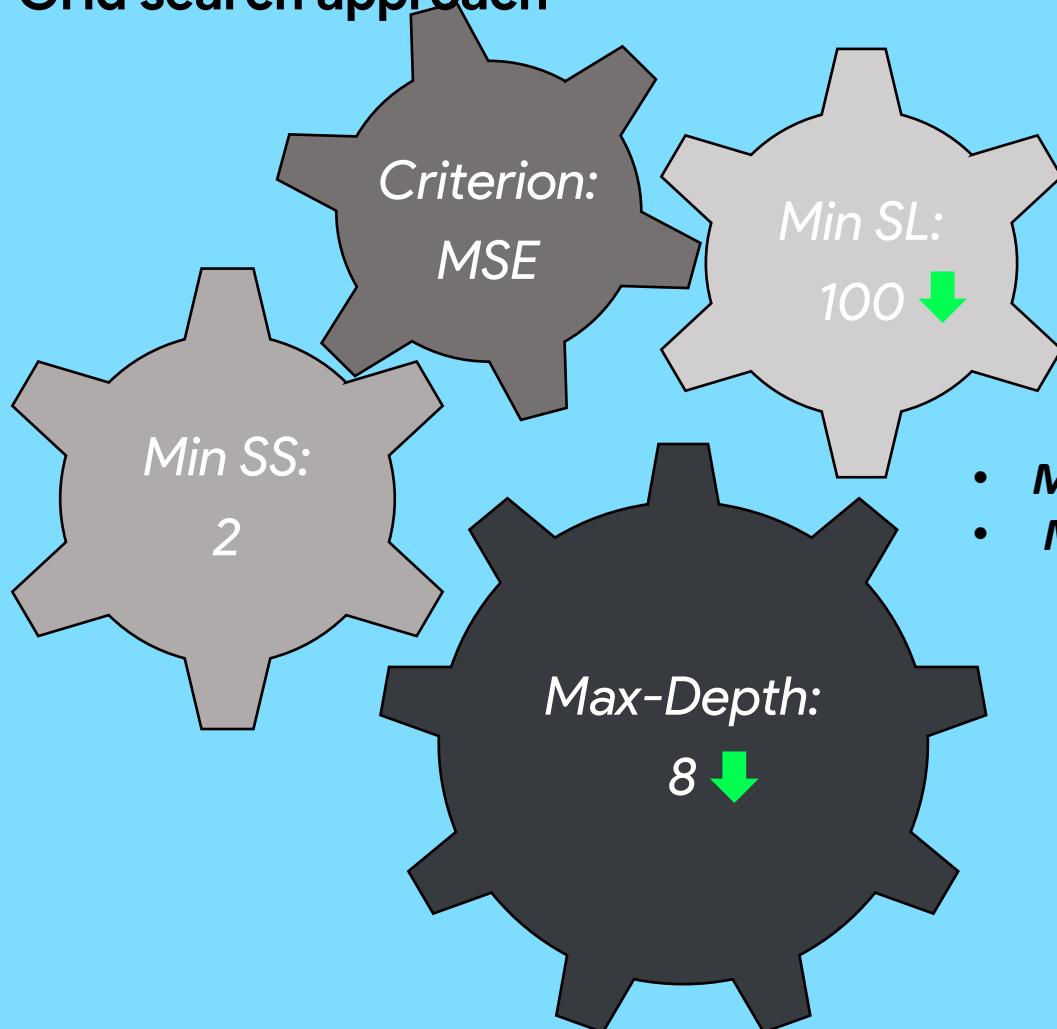
1		RTR	
2		Reviews	
3		Main_Genres	
4		YearUpdated	

8		Price		Secondary_Genres
9		MonthUpdated		Content Rating
10		Both_Stores		Installs
11		Compatibility		App
12		Is_Game		
13		Is_Family		

5		Android Ver
6		Current Ver
7		Size

# Hyper-parameters tuning after features selection:

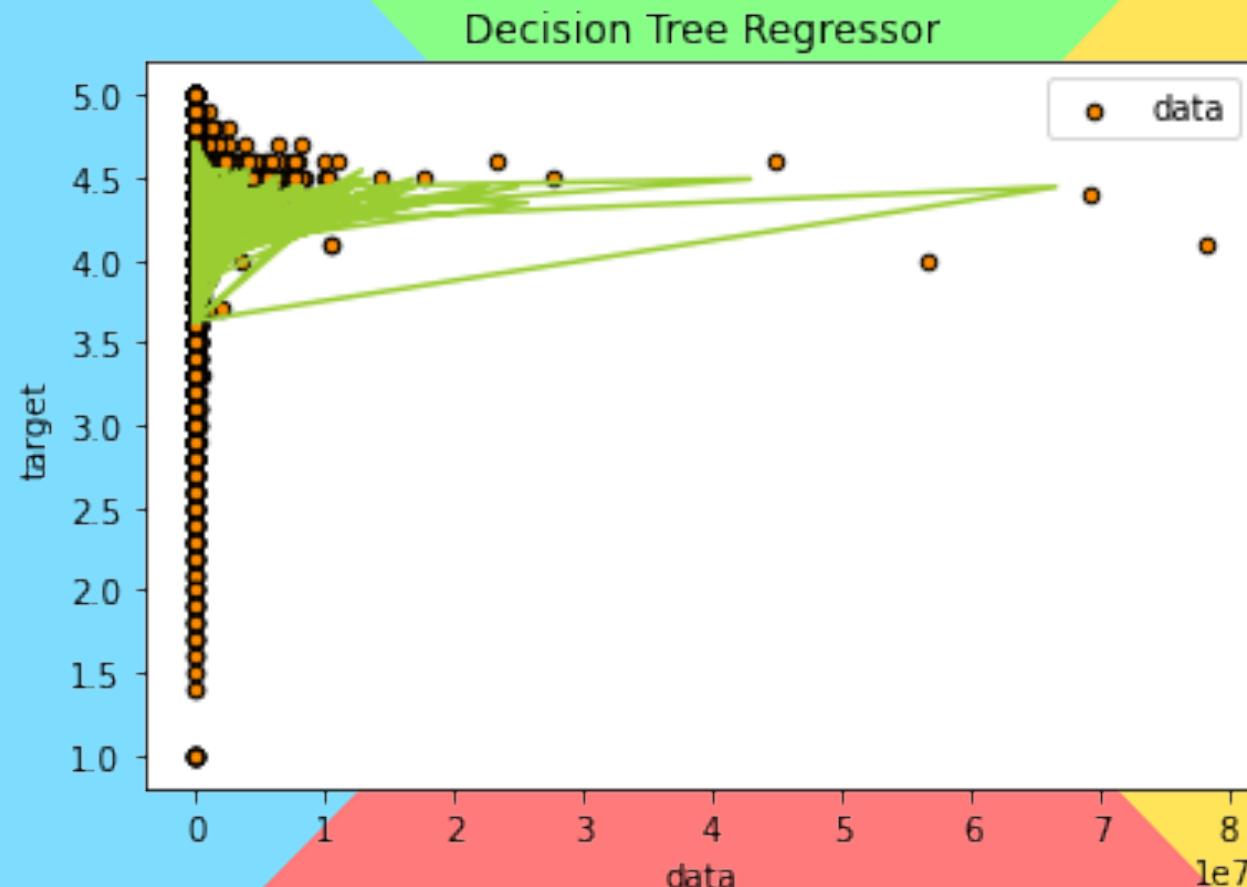
Grid search approach



- ***MinSS*:** Min Samples Split
- ***MinSL*:** Min Samples Leaf

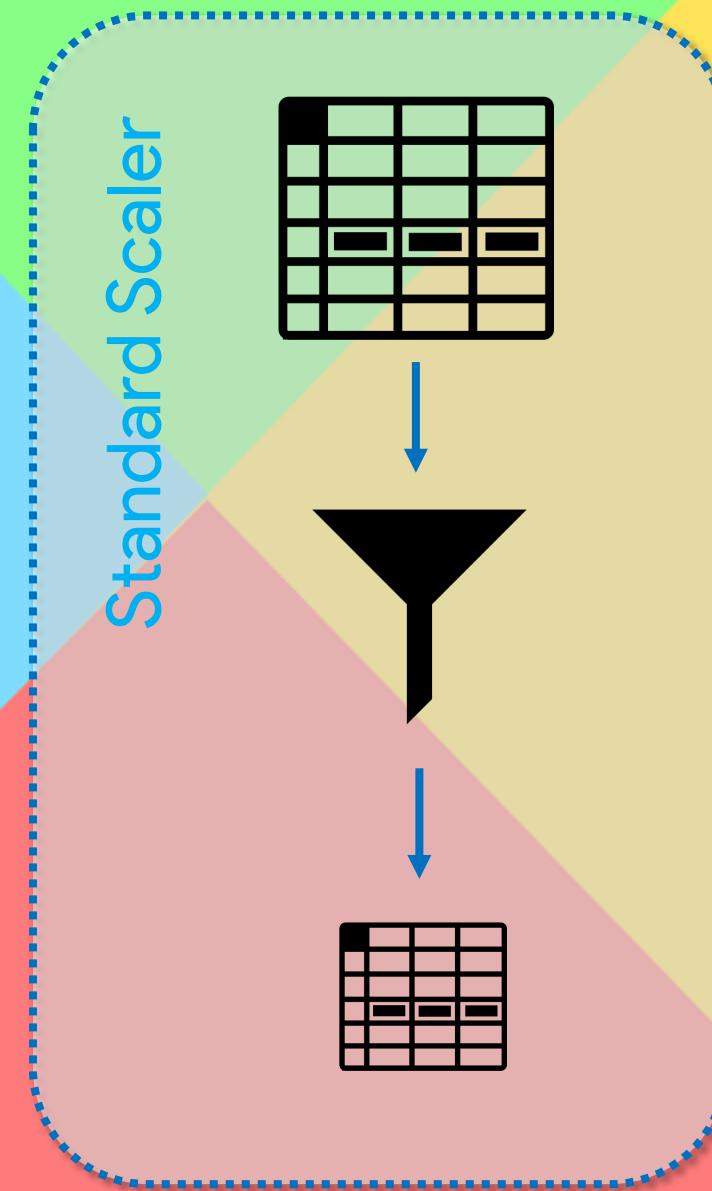
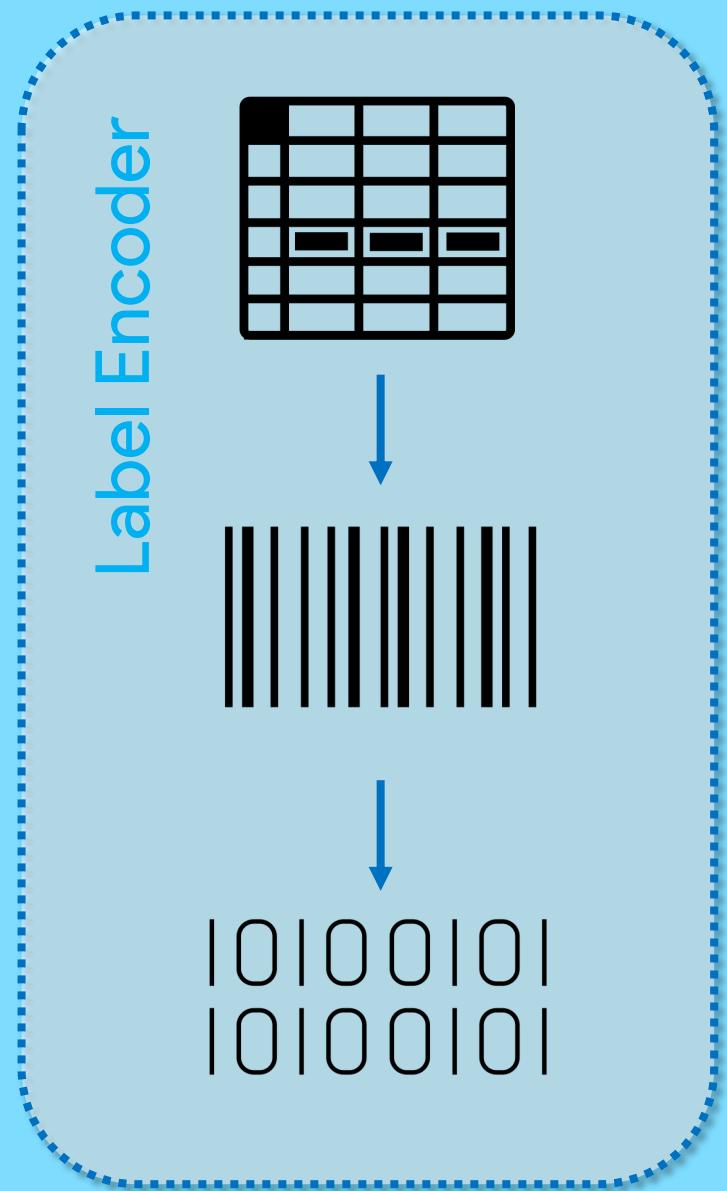
# Final results and model evaluation

PERFORMANCES before FS	
MSE	0.25
RMSE	0.50
MAE	0.34
Pearson (r)	0.37
KS Test	0.30
PERFORMANCES after FS	
MSE	0.25
RMSE	0.50
MAE	0.34
Pearson (r)	0.37
KS Test	0.25 



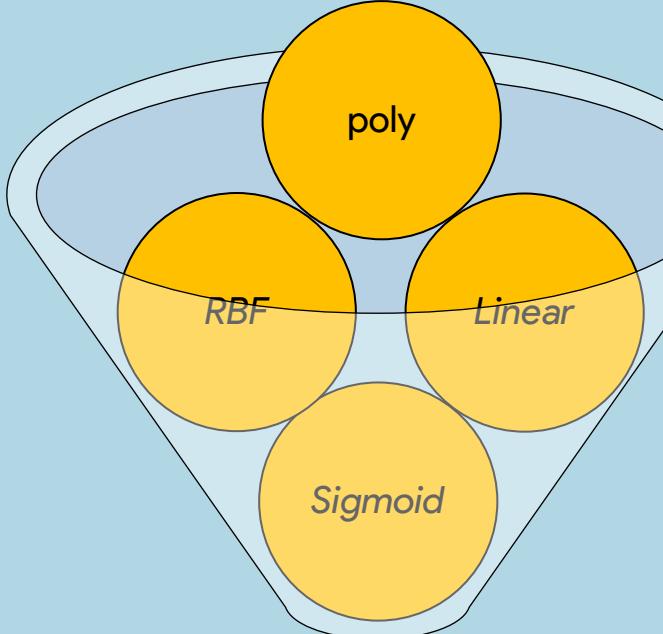
# Support Vector Machine

# Preparing the dataset for the model

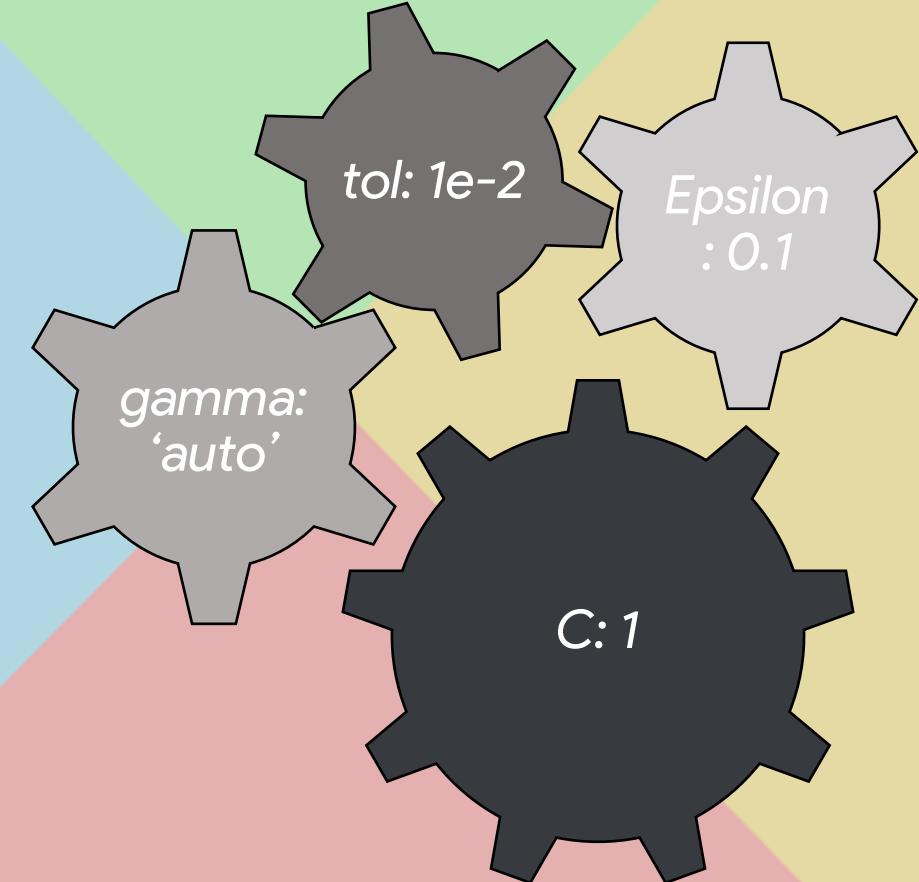


# Kernel function selection and Hyper-parameters tuning

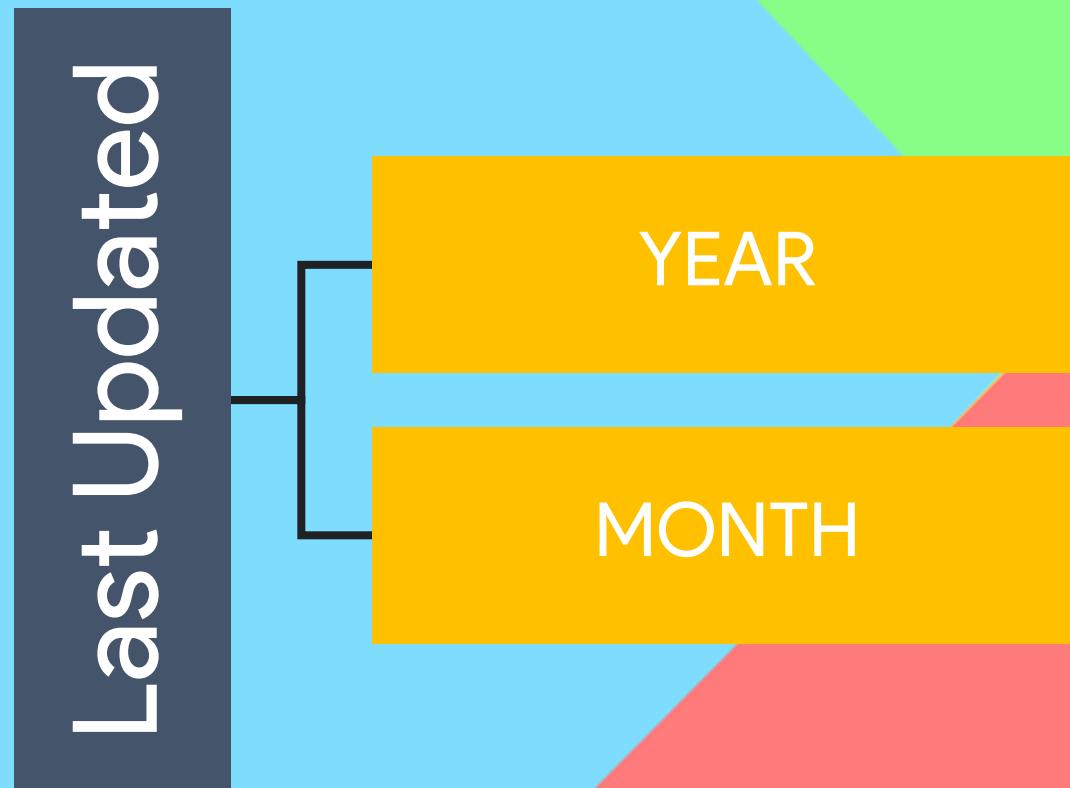
## Kernel function Selection



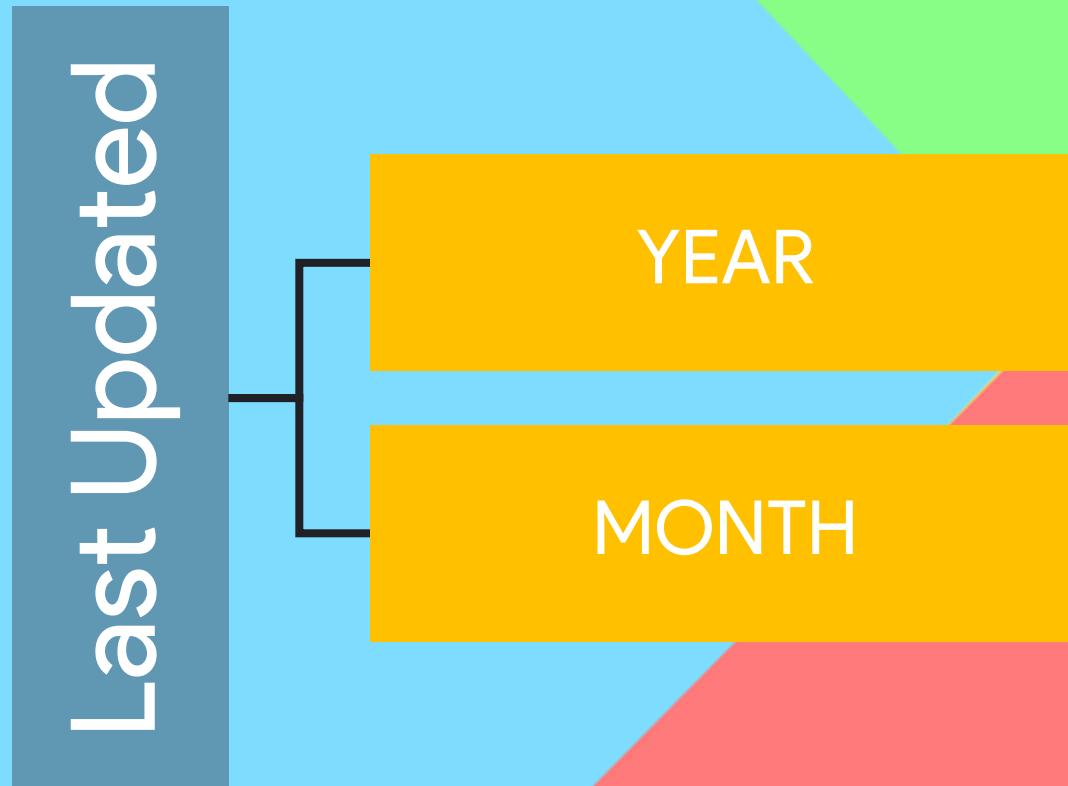
## Hyperparameter tuning



# RFEcv Feature Importance



# RFE CV Feature Importance





# Features Importance

1	 RTR, Year, Month	8	 Secondary_Genres	14	 Is_Game
2	 Content Rating	9	 Is_Family		
3	 Price	10	 Installs		
4	 Compatibility	11	 Size		
5	 Current Ver	12	 Both_Stores		
6	 Reviews	13	 Android Ver		
7	 Main_Genres				



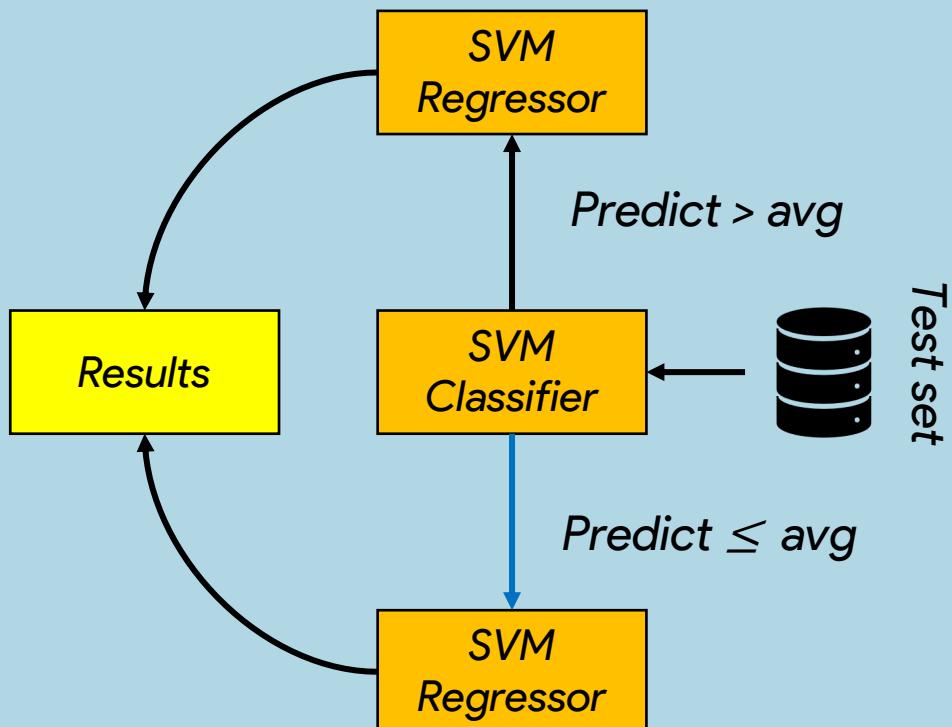
# Features Importance



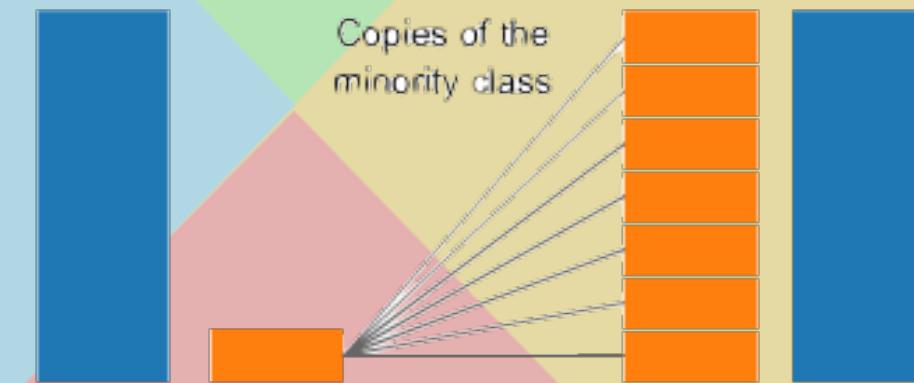
1		RTR, Year, Month
2		Content Rating
3		Price
4		Compatibility
5		Current Ver
6		Reviews
7		Main_Genres
8		Secondary_Genres
9		Is_Family
10		Installs
11		Size
12		Both_Stores
13		Android Ver
14		Is_Game

# Other experiments

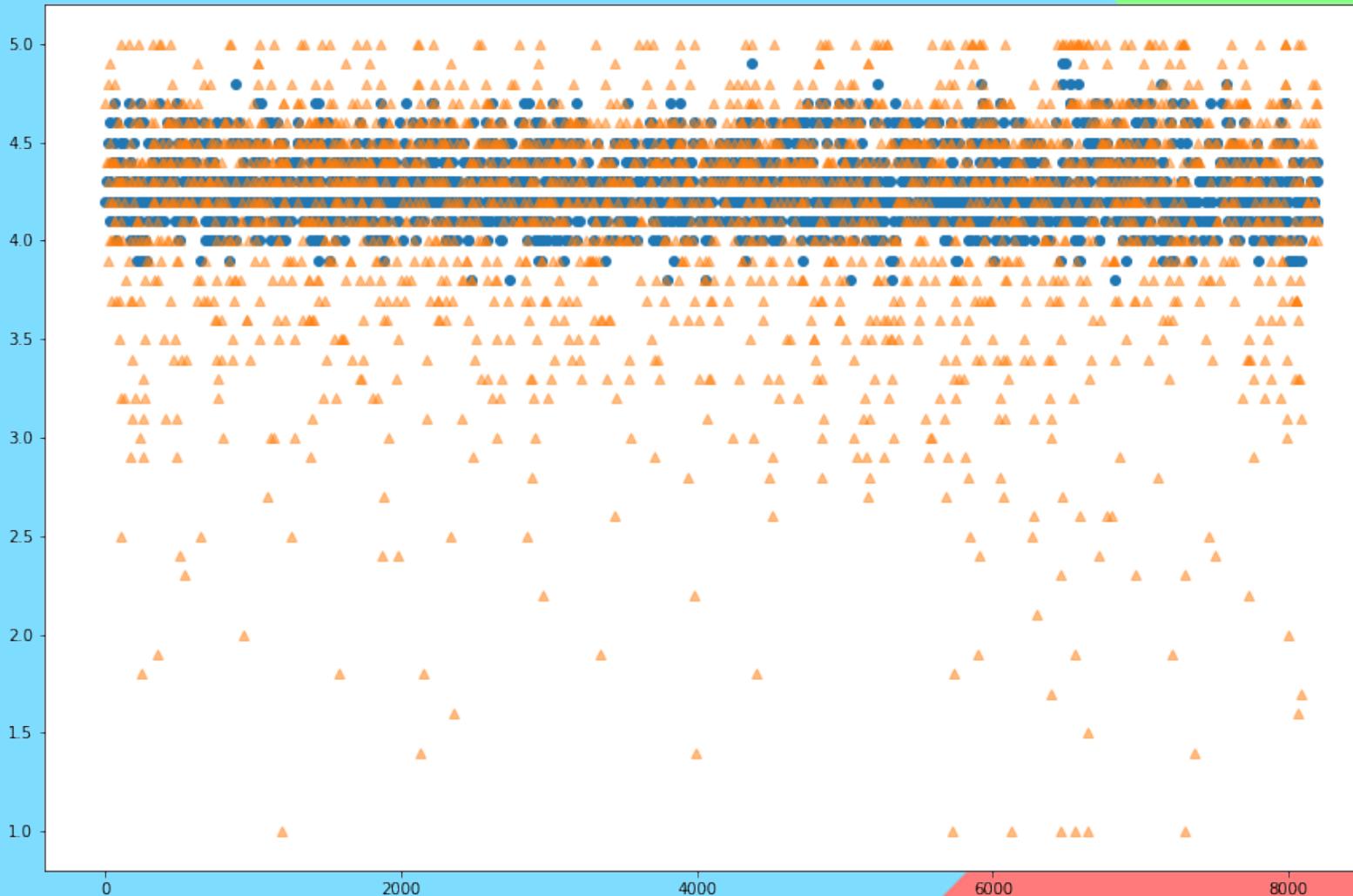
## Splitting the problem



## Oversampling



# Final results and evaluation

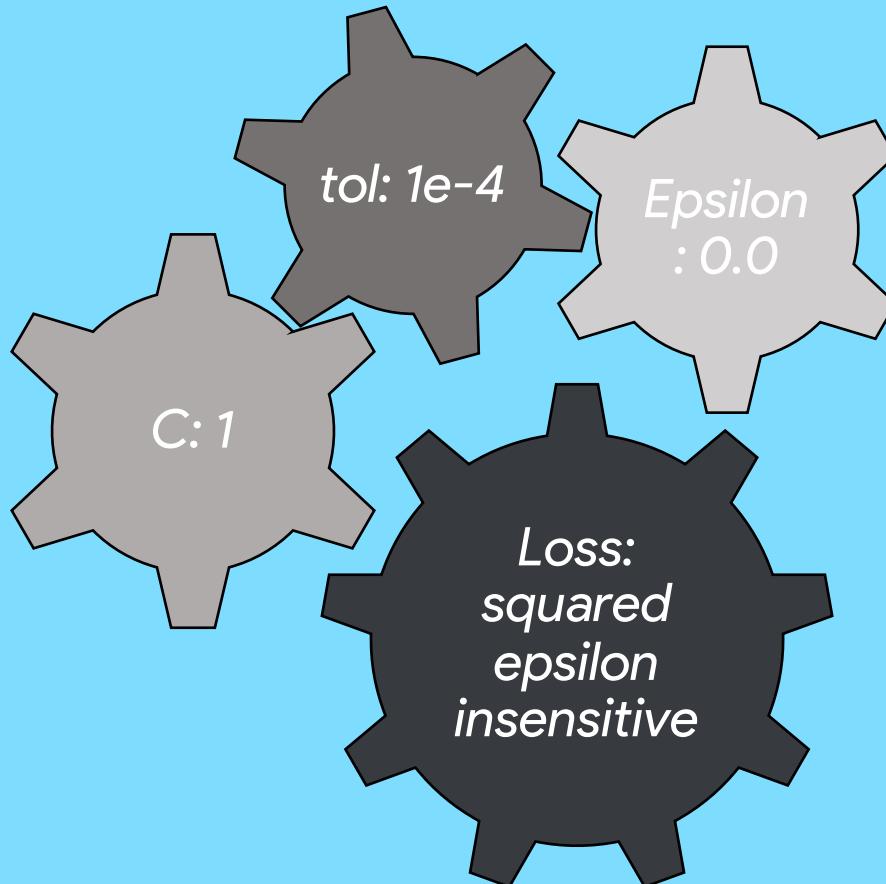


ORIGINAL DATASET	
MSE	0.275
RMSE	0.524
MAE	0.343
Pearson (r)	0.35
KS Test	0.256
RFECV	
MSE	0.271
RMSE	0.520
MAE	0.341
Pearson (r)	0.376
KS Test	0.285
RFECV MONTH - YEAR	
MSE	0.269
RMSE	0.518
MAE	0.338
Pearson (r)	0.384
KS Test	0.258
PROBLEM SPLIT	
MSE	0.325
RMSE	0.570
MAE	0.404
Pearson (r)	0.265
KS Test	0.355

# Ordinal Regressor

# LAD

## Grid search approach

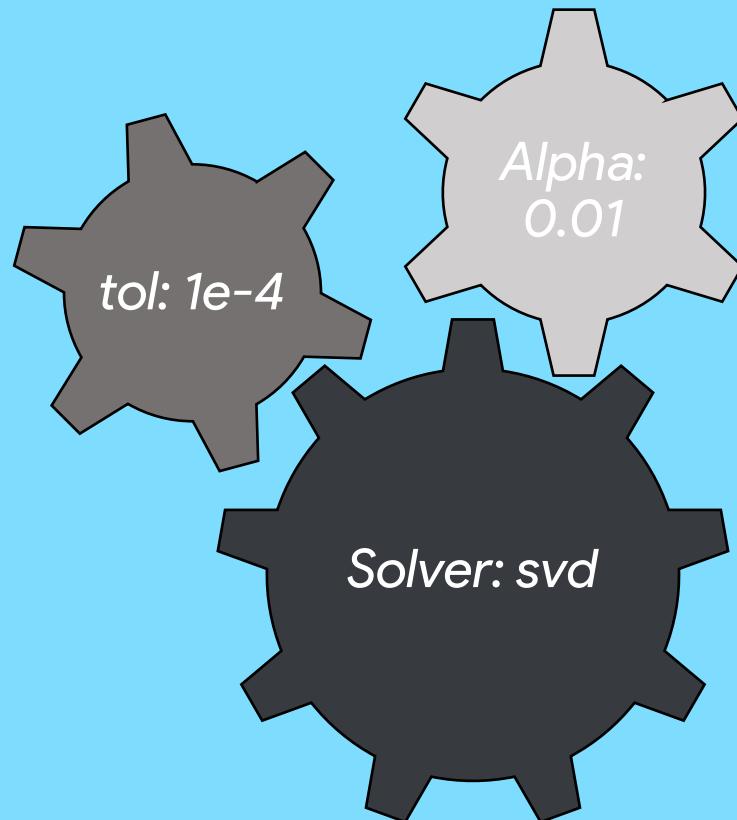


## Metrics evaluation

LAD	
MSE	1.39
RMSE	1.17
MAE	0.75
Pearson (r)	-0.143
KS Test	0.704

# Ordinal Ridge

Grid search approach



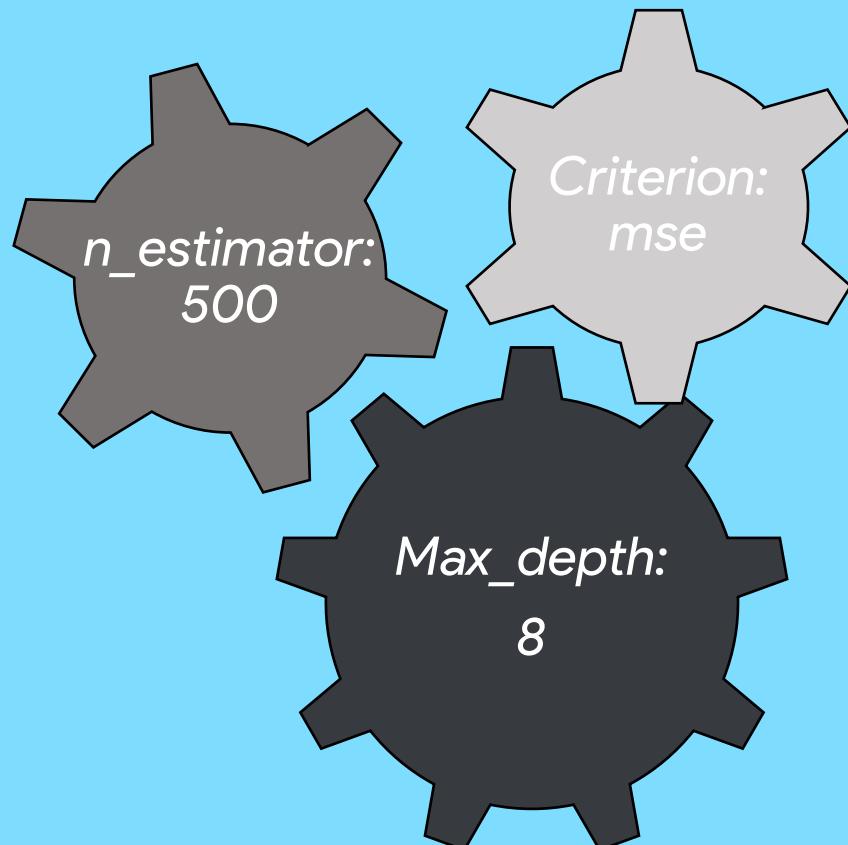
Metrics evaluation

ORDINAL RIDGE	
MSE	0.32
RMSE	0.66
MAE	0.44
Pearson (r)	0.095
KS Test	0.69

# RandomForest Regressor

# Random Forest Regressor

## Grid search approach



## Metrics evaluation

RFR	
MSE	0.24
RMSE	0.48
MAE	0.33
Pearson (r)	0.435
KS Test	0.284



# Features Importance

1		RTR	8		Year
2		Reviews	9		Price
3		Current Ver	10		Installs
4		Size			
5		Main_Genres			
6		Android Ver			
7		Month			

# Convolutional Neural Network

1. Multi - Class Classification
2. Ordinal Regression
3. Stacked-Ensemble

# Preparing the dataset for the model

Last Updated

YeraUpdated

MonthUpdated

DayUpdated

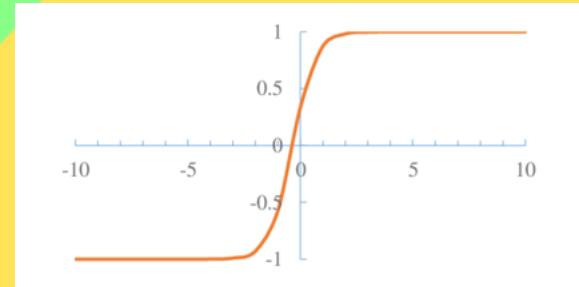
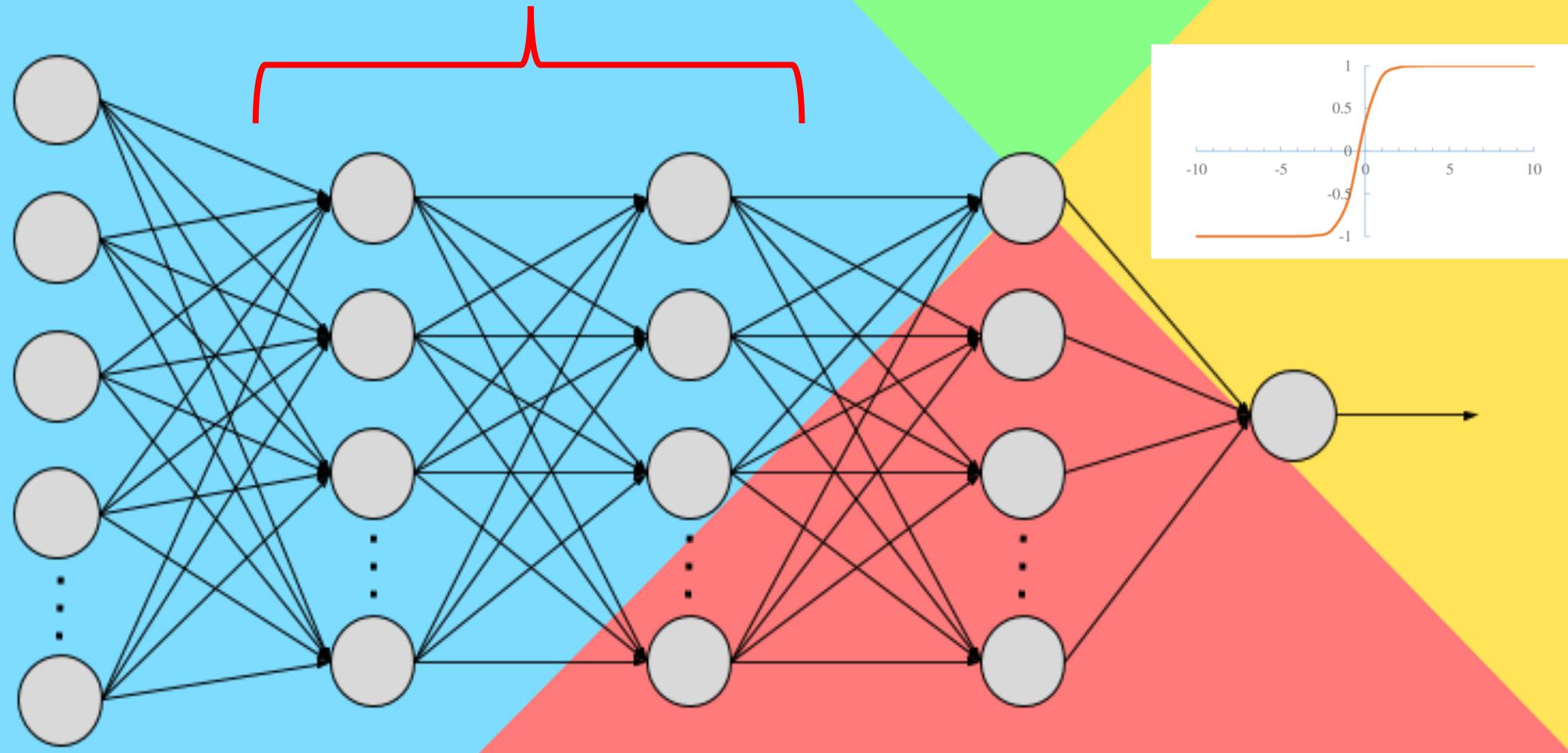
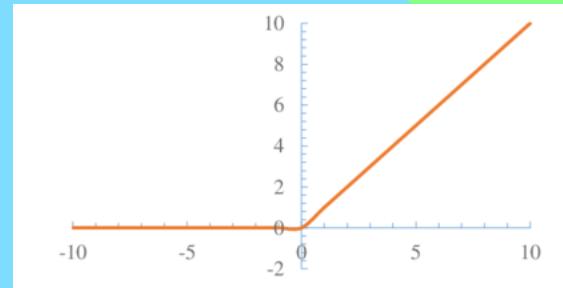
Google playstore user reviews dataset



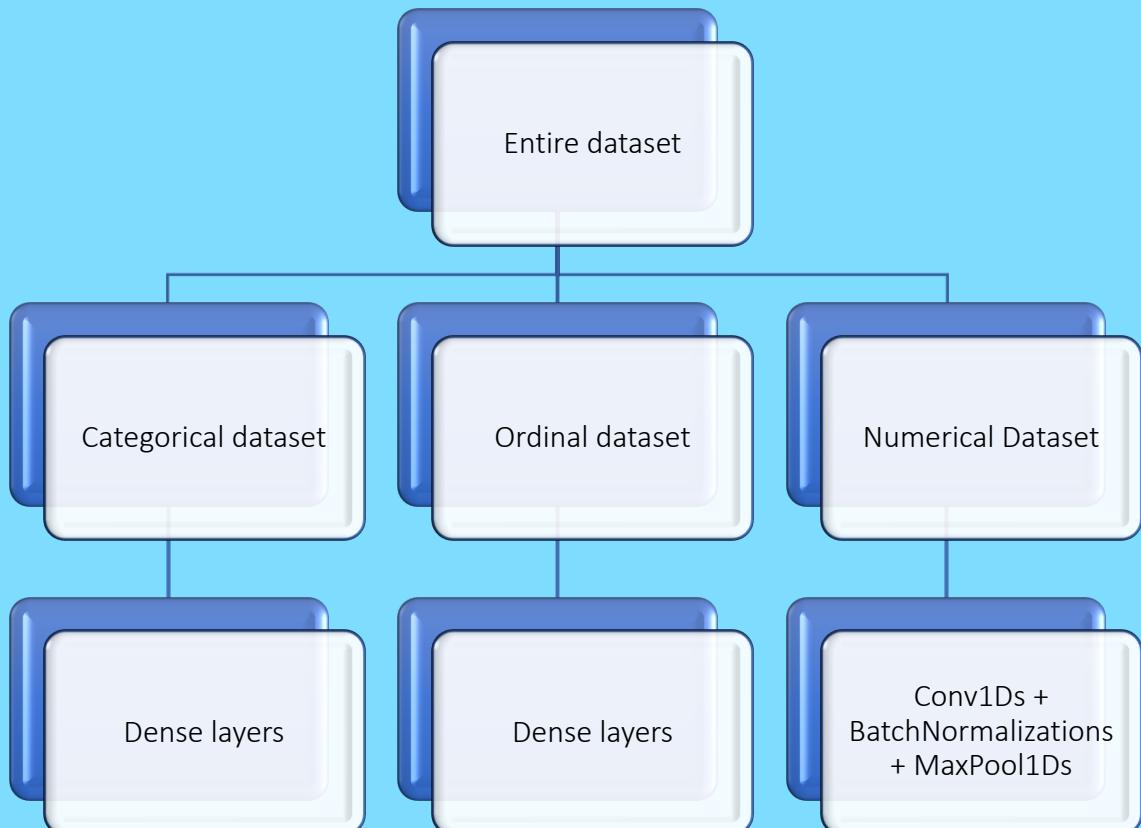
SentimentPolarityMean

# 1. Multi - Class Classification

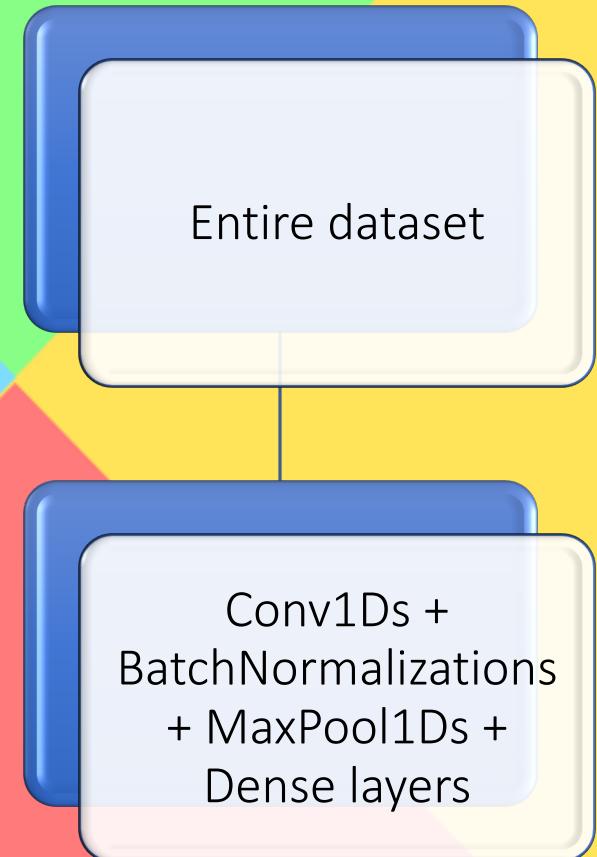
# Activation functions



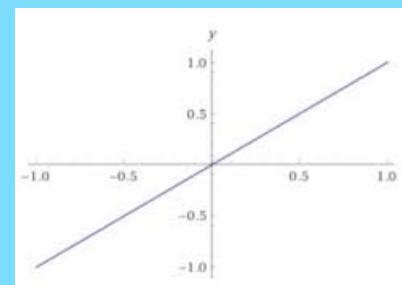
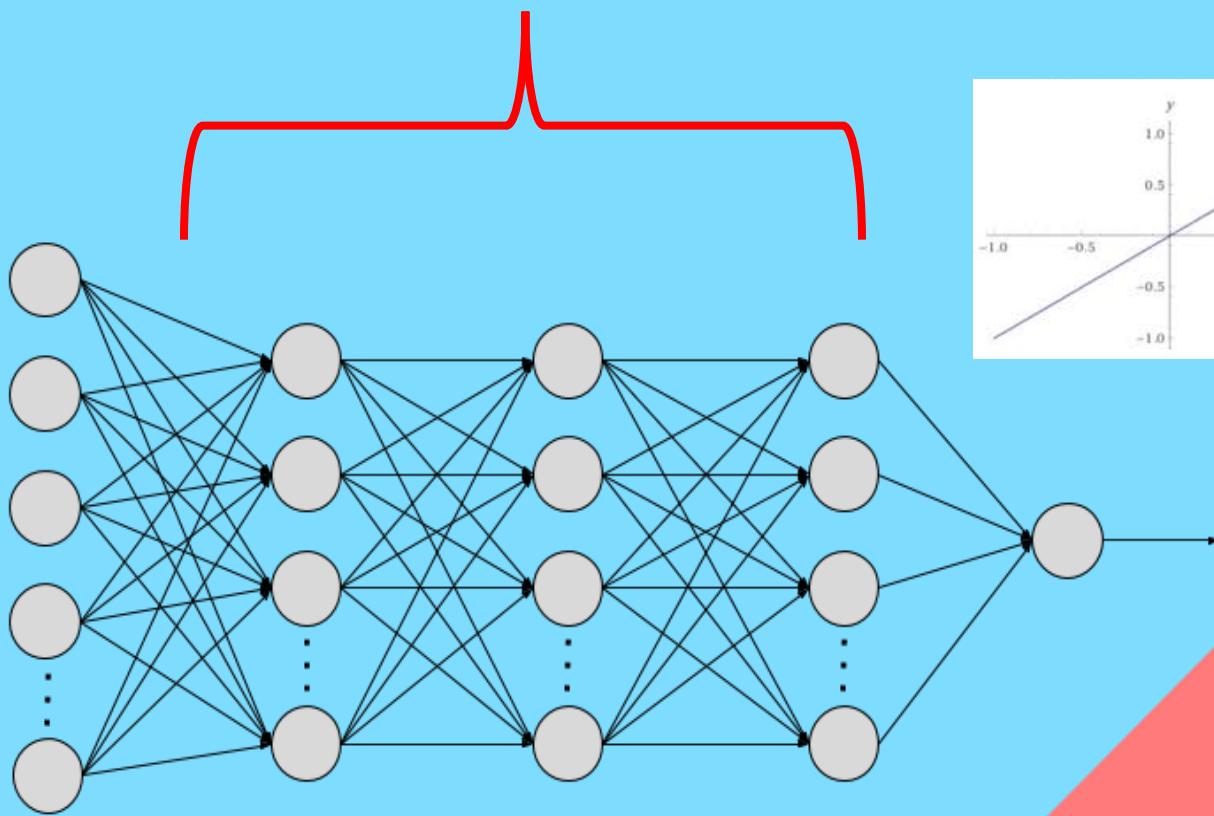
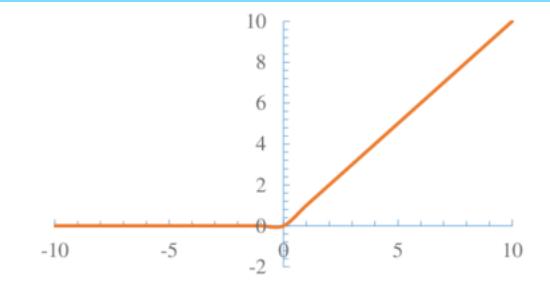
## *Model\_1 & Model\_4 (SMOTE)*



## *Model\_2 & Model\_3 (SMOTE)*



## 2. Ordinal Regression



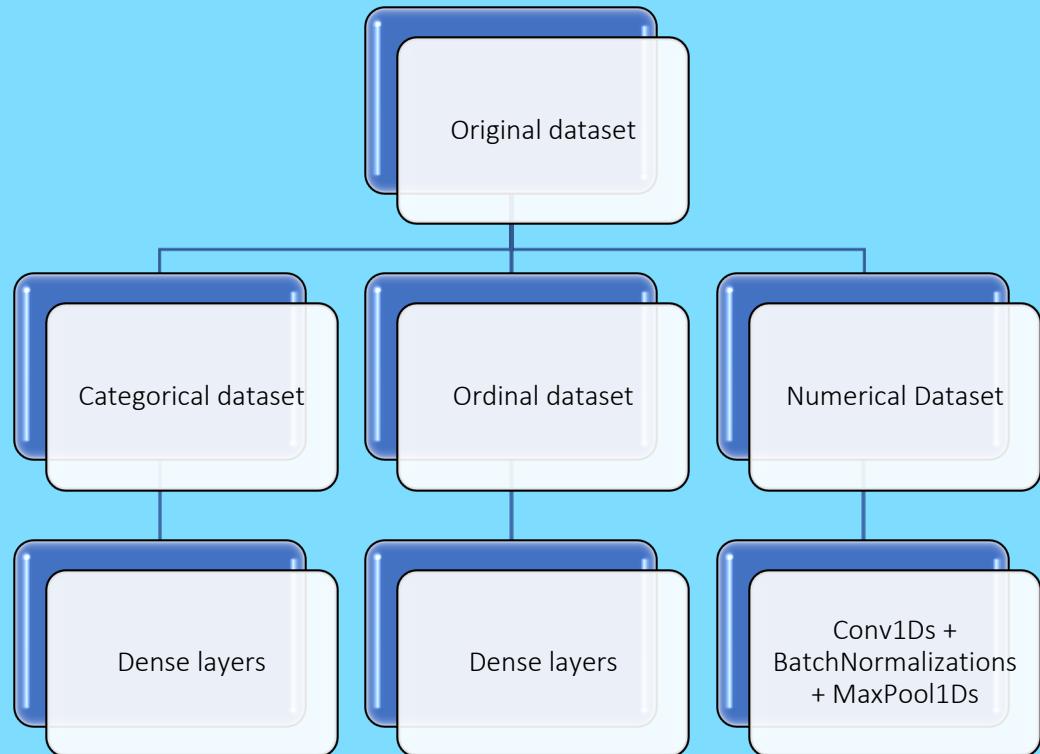
***Model\_5 – Model\_6 – Model\_7***

Entire dataset

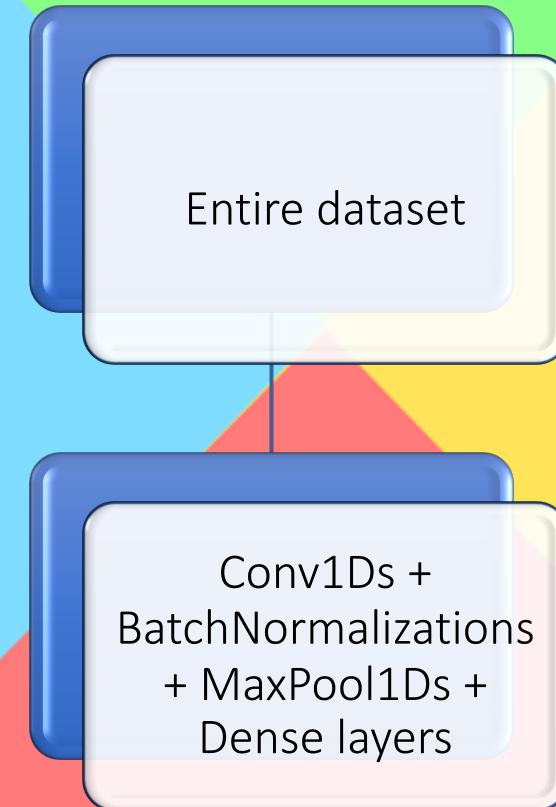
Conv1Ds +  
BatchNormalizations  
+ MaxPool1Ds +  
Dense layers

### 3. Stacked- Ensemble

*Model\_1*



*Model\_7*



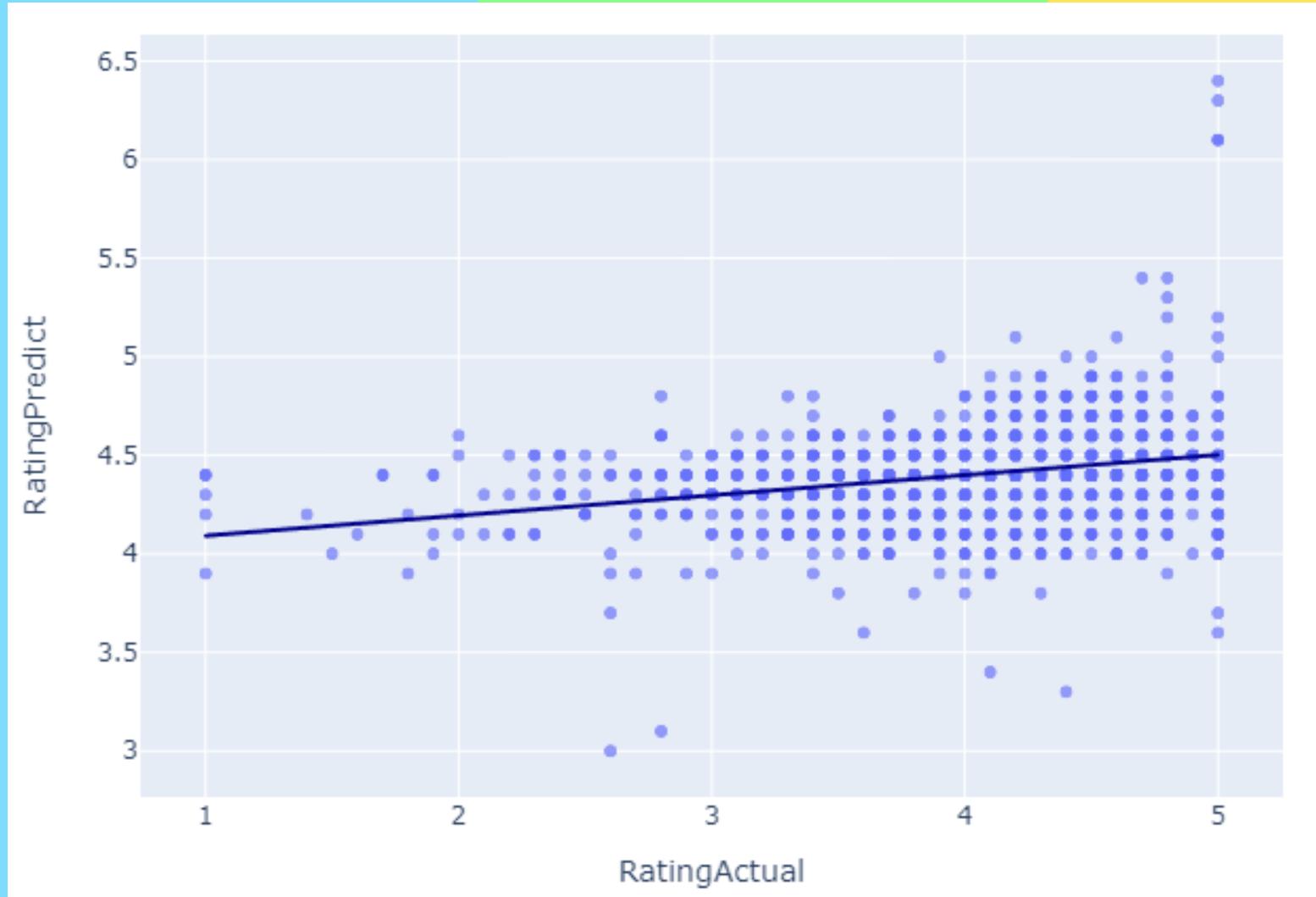
# Final results

MULTI - CLASS	
MODEL 1	
RMSE	0.57
MAE	0.37
Pearson (r)	0.29
KS Test	0.33
MODEL 2	
RMSE	0.73
MAE	0.52
Pearson (r)	0.01
KS Test	0.04
MODEL 3 (SMOTE)	
RMSE	0.84
MAE	0.59
Pearson (r)	0.01
KS Test	0.06
MODEL 4 (SMOTE)	
RMSE	0.65
MAE	0.59
Pearson (r)	0.17
KS Test	0.07

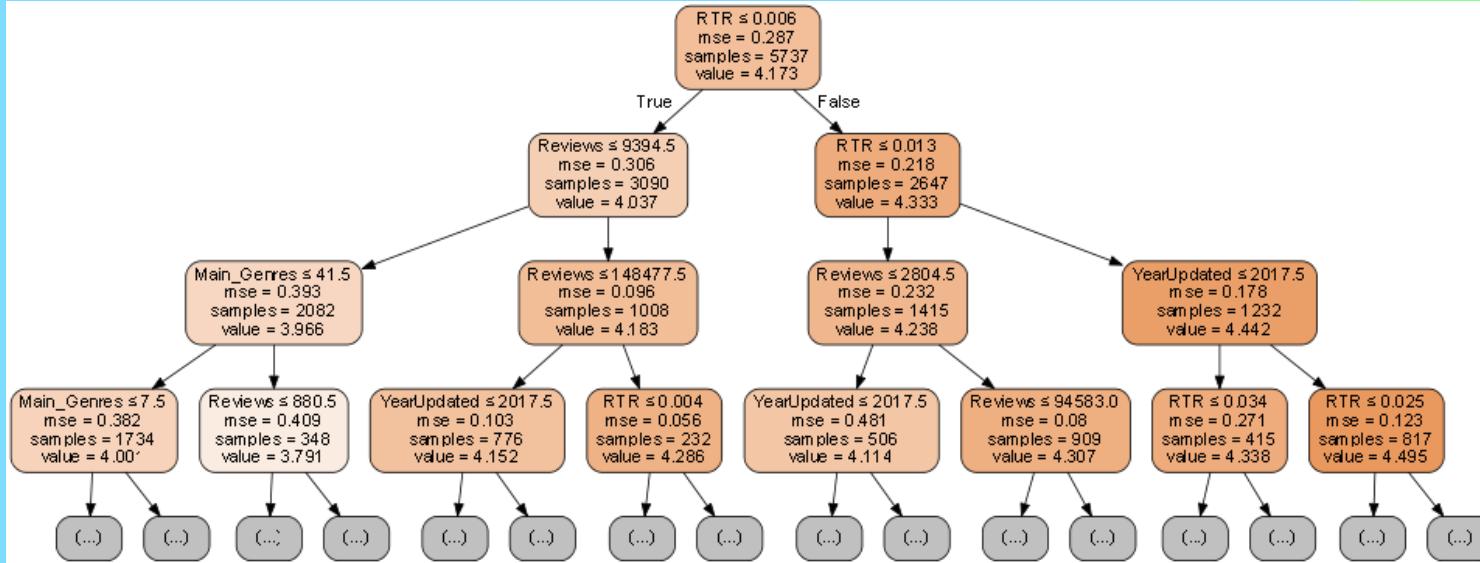
ORDINAL REGRESSION	
MODEL 5	
RMSE	0.54
MAE	0.41
Pearson (r)	0.20
KS Test	0.56
MODEL 6	
RMSE	0.50
MAE	0.34
Pearson (r)	0.34
KS Test	0.24
MODEL 7	
RMSE	0.50
MAE	0.24
Pearson (r)	0.34
KS Test	0.24

STACKED ENSEMBLE	
MODEL 8	
RMSE	0.57
MAE	0.39
Pearson (r)	0.27
KS Test	0.41

# Stacked-Ensemble



# Best classification model



		Score
<b>RMSE</b>		
DRT	0,48	
SVMR	0,52	
OR	0,66	
RFR	0,48	
NN	0,57	
<b>MSE</b>		
DRT	0,23	
SVMR	0,27	
OR	0,32	
RFR	0,24	
NN	0,32	
<b>MAE</b>		
DRT	0,34	
SVMR	0,34	
OR	0,44	
RFR	0,33	
NN	0,39	
<b>r</b>		
DRT	0,43	
SVMR	0,38	
OR	0,095	
RFR	0,44	
NN	0,27	
<b>KS</b>		
DRT	0,26	
SVMR	0,26	
OR	0,69	
RFR	0,29	
NN	0,41	