

Google Play App

Predict the rating on Google Play Store Apps for the Android market with machine learning

Xiongfeng Wang, Brown DSI, 10/16/2020
<https://github.com/XiongfengWang/1030project>





01

Introduction



02

EDA



03

Preprocessing



04

Q&A

Introduction



- Predict the rating for Google Play Store Apps:
- Rating affect App's success and visibility.
- Good prediction on new Apps is beneficial.
- Regression: rating scales from 1 to 5 with decile level between each integer.
- Dataset: Kaggle
<https://www.kaggle.com/lava18/google-play-store-apps>

Dataset Overview

- 10,358 samples, 13 columns
- Drop duplicate samples
- Drop samples missing rating
- 8,893 samples
- missing values in 4 columns
- Group structure
- Imbalance

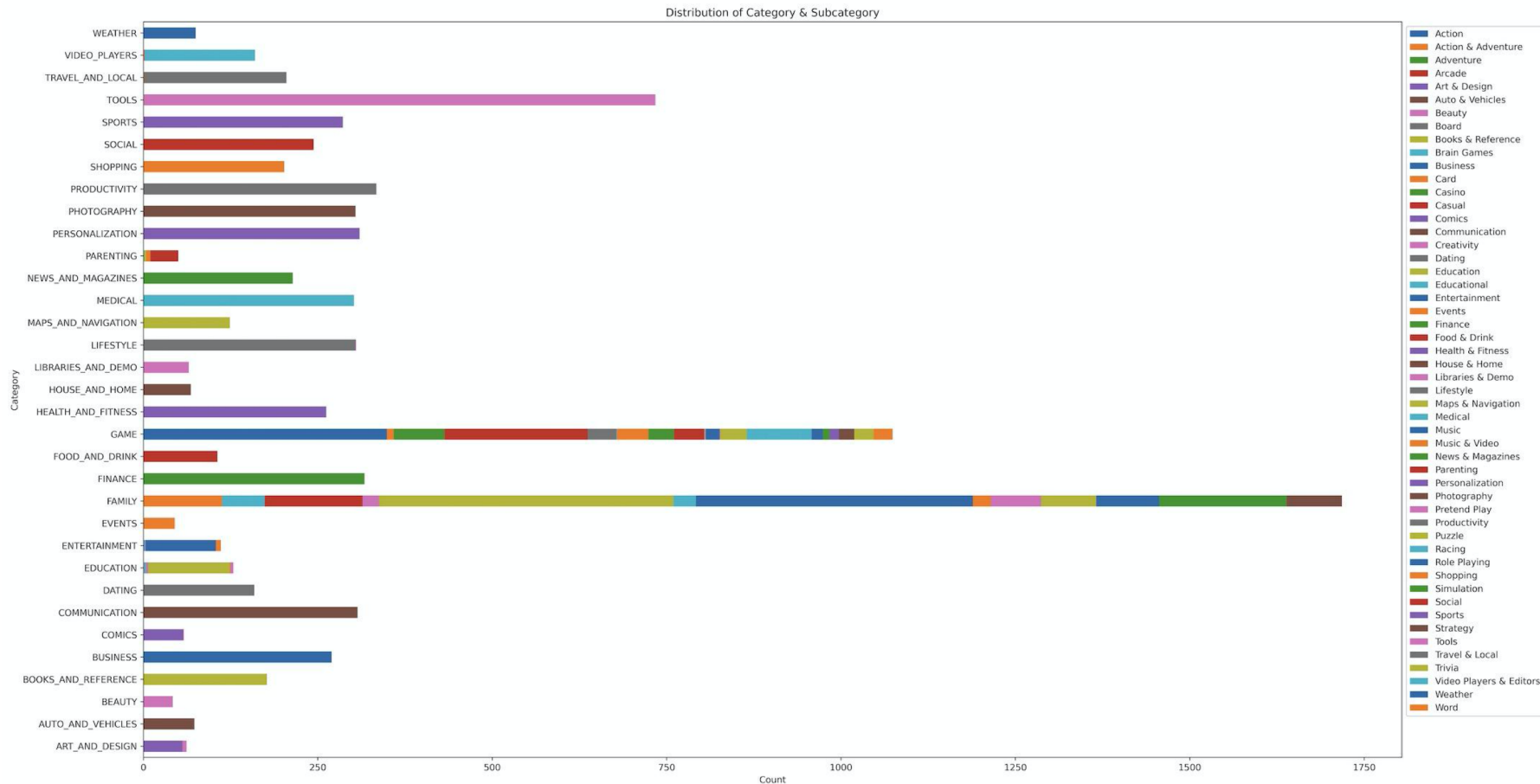


EDA

	Variable	Classification ^a	Note
Key	App		
Features	Category	Categorical	Imbalance
	Reviews	Numerical	'Varies with device' kB, MB to Byte
	Size	Numerical	
	Installs	Categorical	
	Type	Categorical	
	Price	Numerical	Remove \$ sign
	Content Rating	Categorical	

	Variable	Classification	Note
Features	Genres	Categorical	Transfer to Subcategory
	Last Updated	Numerical	Transfer to Days
	Current Ver	Categorical	Unique format
	Android Ver	Categorical	
	Subcategory	Categorical	8446 null value
	Days	Numerical	
Target	Rating	Numerical	

Group Structure: Category and Subcategory

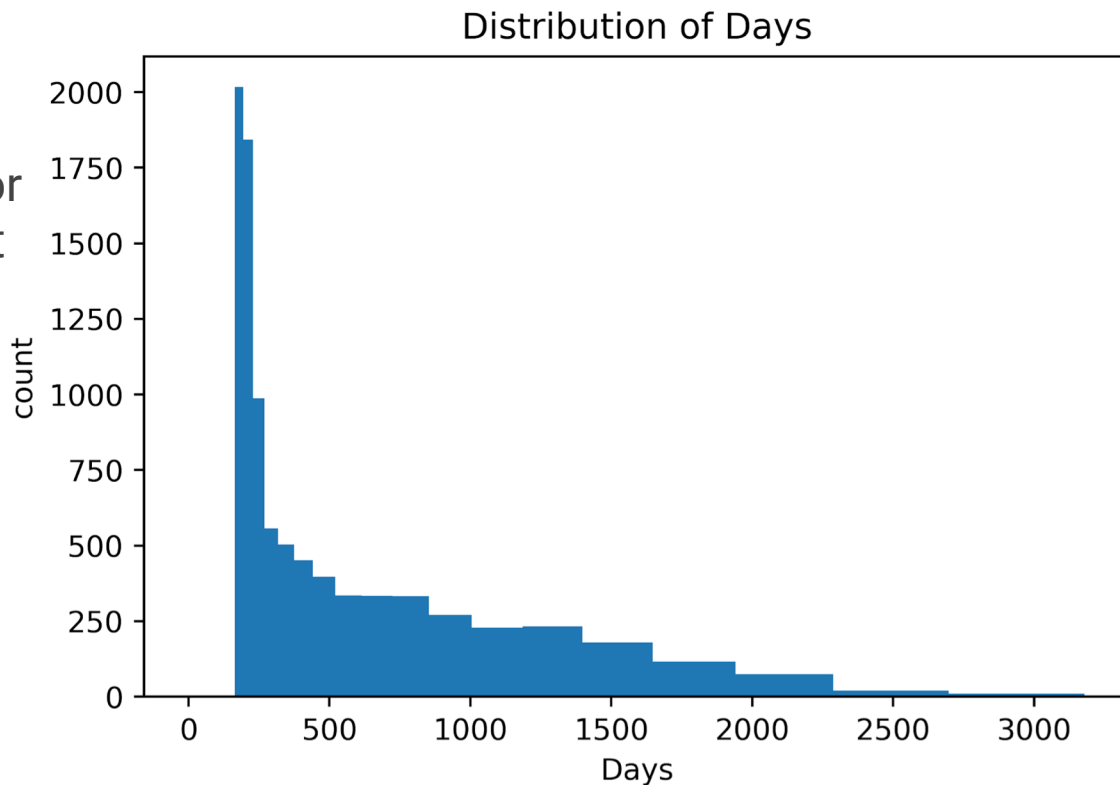


Last Updated -> Days

Difference between
last updated date for
App and for Dataset

Gap: Dataset last
update date is
02/03/2019

Potential new
feature: Month

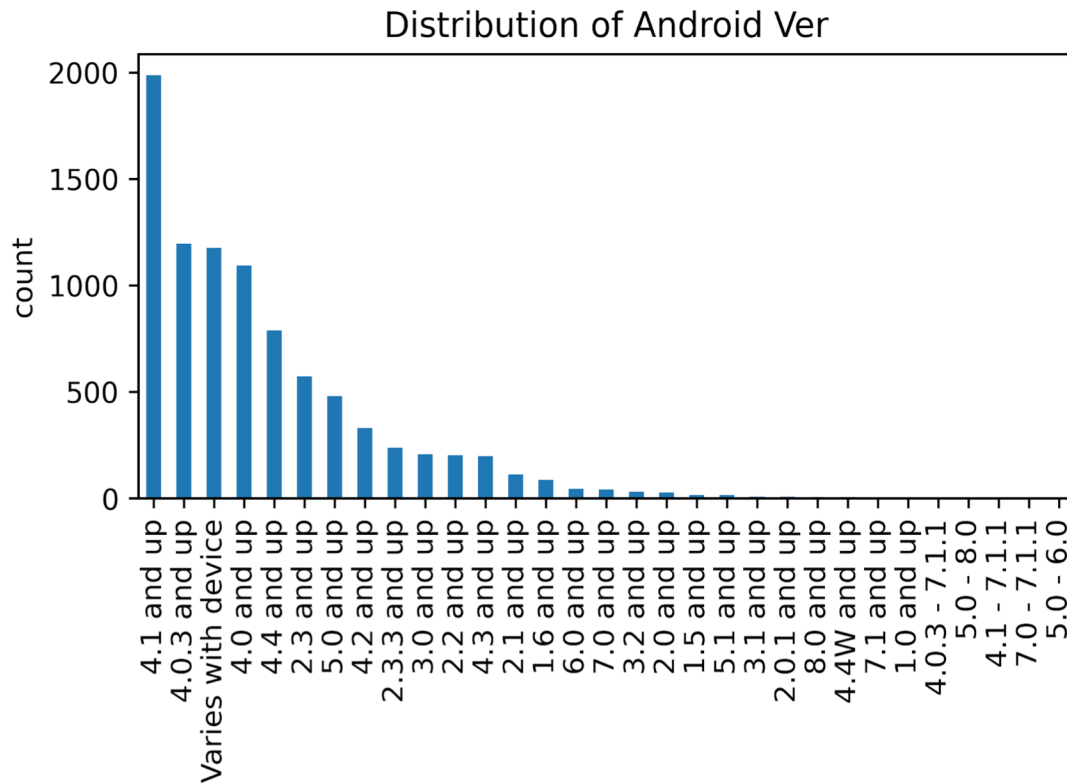


Android Ver

Whether to make it numerical

Different format for newly added samples

Varies with device:
potential null value,
more missing values
to fill



Target: Rating

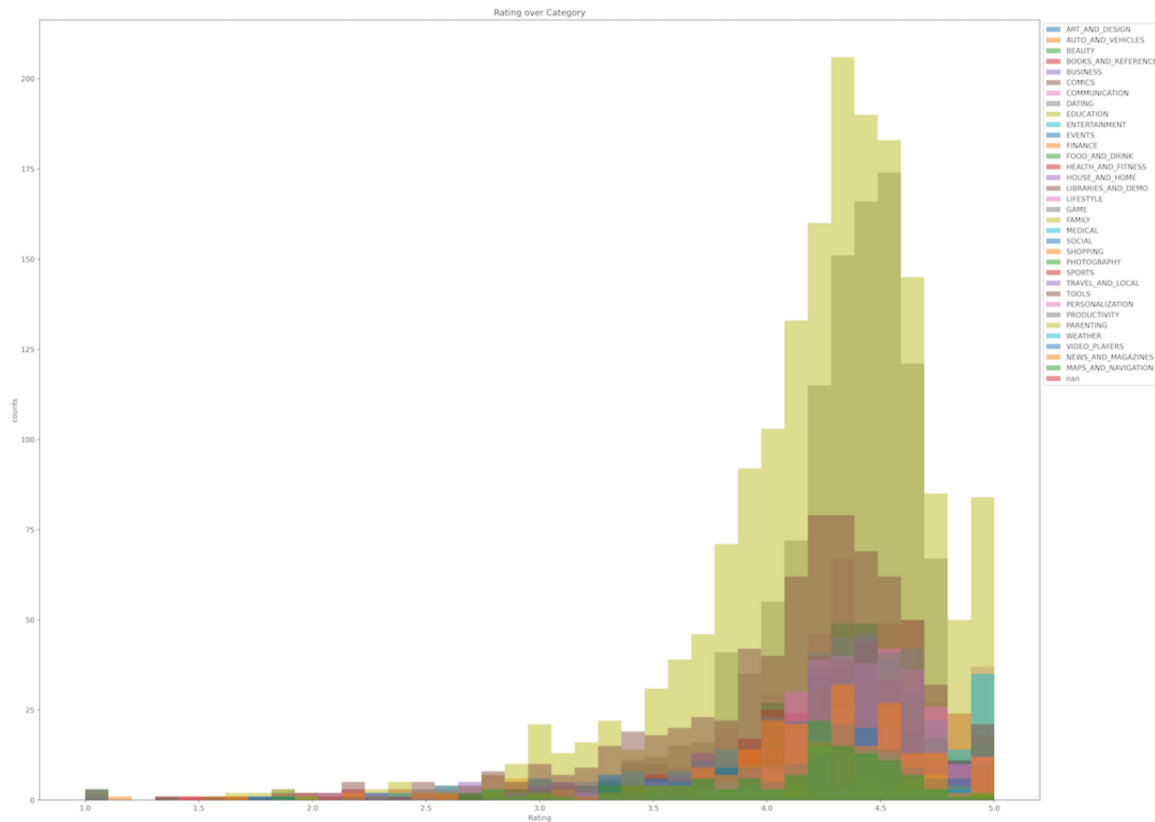
Normal distribution,
left skewed

mean : 4.19

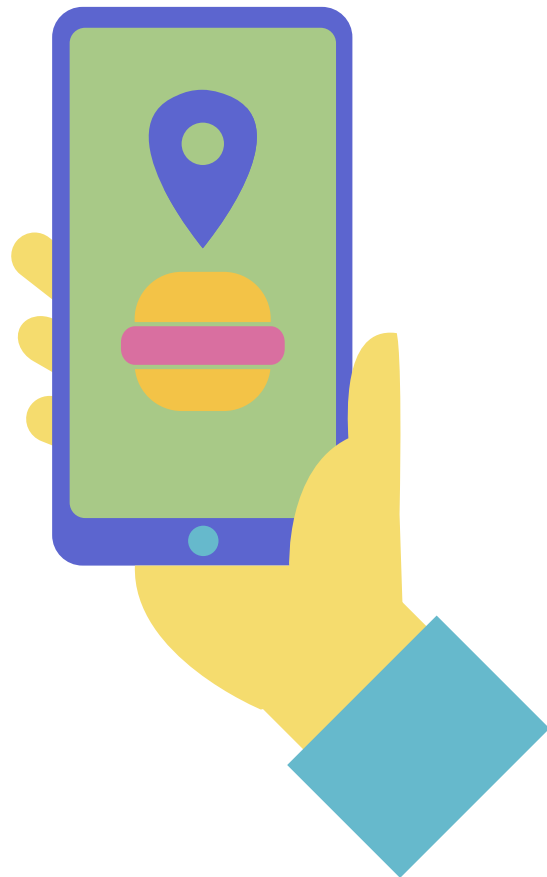
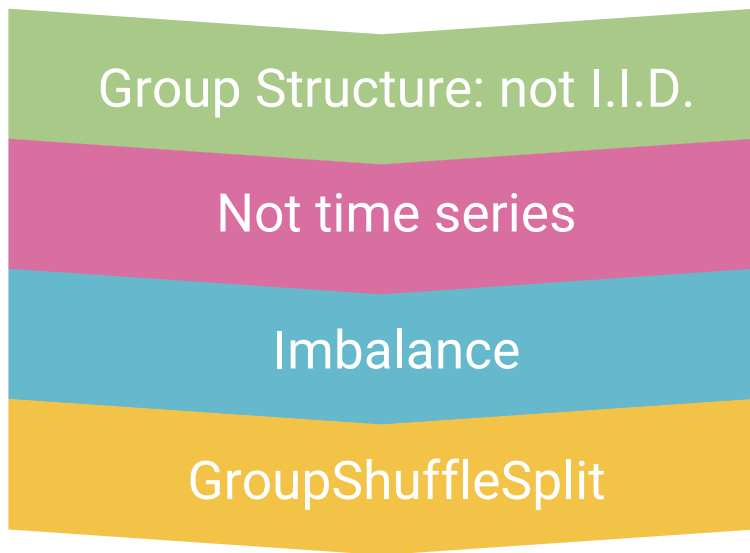
Std: 0.52

25% - 75%: 4 - 4.5

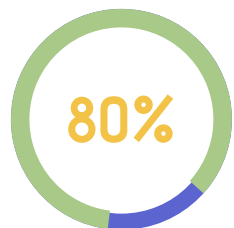
bin = 49



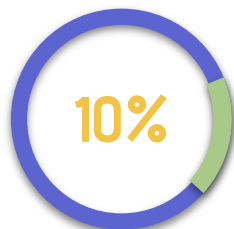
Preprocessing



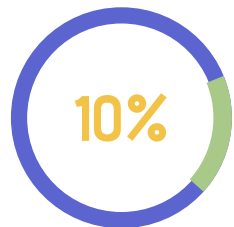
How to split



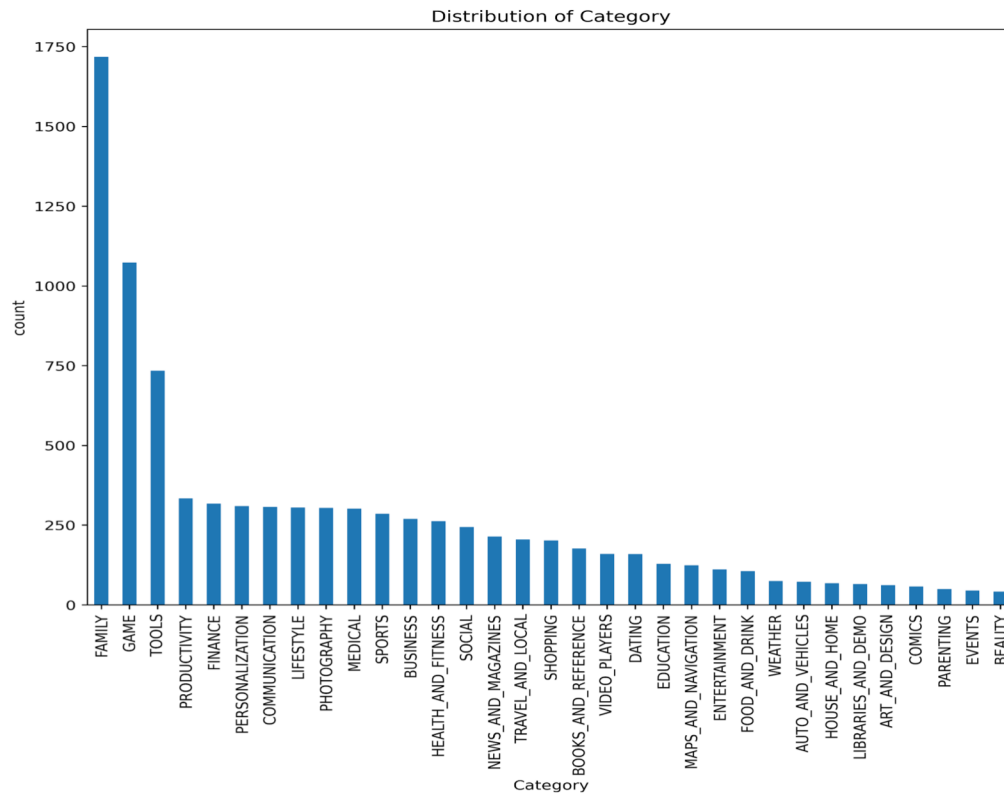
Train



Validation



Test



Encoding

	Variable	Classification	Encoder
Key	App		
Features	Category	Categorical	OneHotEncoder
	Reviews	Numerical	StandardScaler
	Size	Numerical	StandardScaler
	Installs	Categorical	OrdinalEncoder
	Type	Categorical	OneHotEncoder
	Price	Numerical	StandardScaler
	Content Rating	Categorical	OneHotEncoder

	Variable	Classification	Encoder
Features	Genres	Categorical	
	Last Updated	Numerical	
	Current Ver	Categorical	
	Android Ver	Categorical	OneHotEncoder
	Subcategory	Categorical	OneHotEncoder
	Days	Numerical	StandardScaler
Target	Rating	Numerical	StandardScaler



Q&A

Thanks!



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.