A Google App Store Educational Apps Rating Analysis

By Hongling Yang

Objective

- Educational tech apps, committed to the development of a new way of generating and delivering knowledge to motivate learning, is in great demand. Efforts should be made to develop more effective and efficient educational apps, to complement traditional learning channels.

- What features of Educatioinal Apps would users value more and can boost Apps' Rating Score?

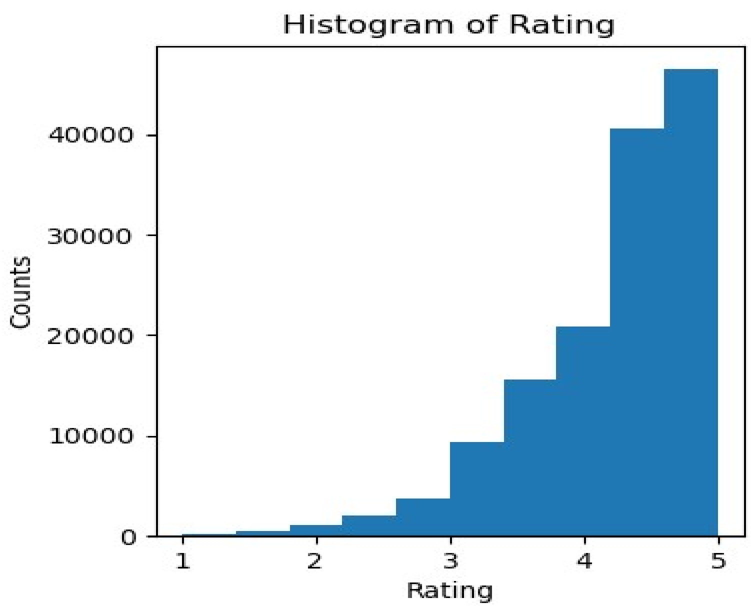
- Features to be examined (defined in Table 1)

  - Installations
  - Price
  - Size
  - Price
  - Year of Release
  - last update
  - Ad_Supported
  - Editors_Choice
  - Content_Rating
  - In_App_Purchases

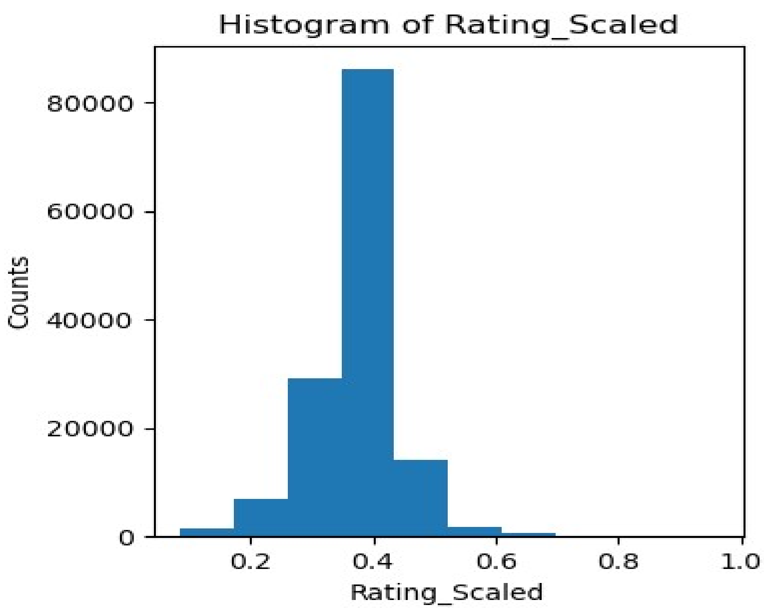
- Data Source (Kaggle):
  https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps

Distribution of Raw & Scaled Ratings

(1) Raw Ratings



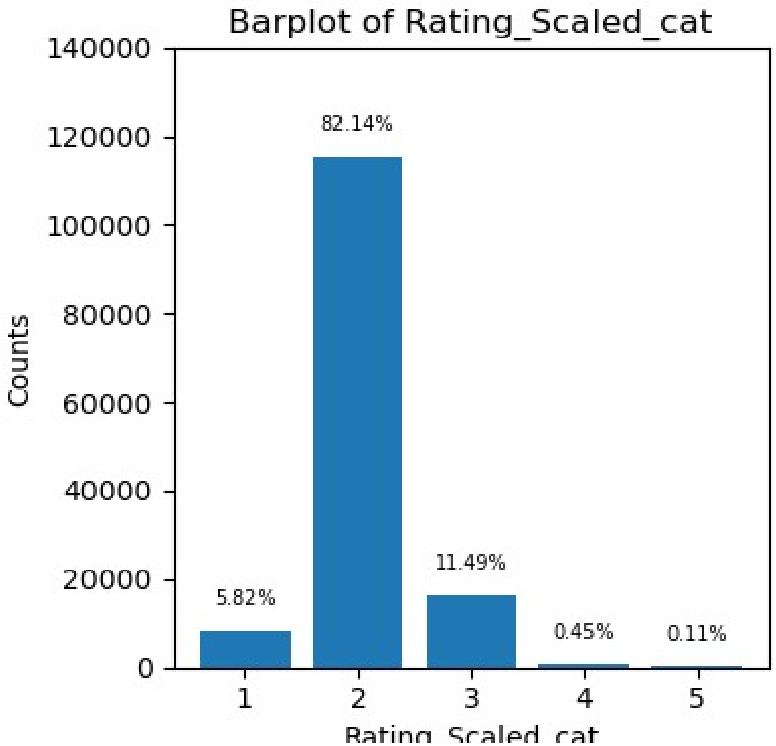
(2) Scaled Ratings



- Scaled Rating = Raw Rating/log (Rating Counts)
- Scaled Tating is right -skewed,

Distribution of Scaled & Categorized Ratings

(3) Scaled_Categorized Ratings

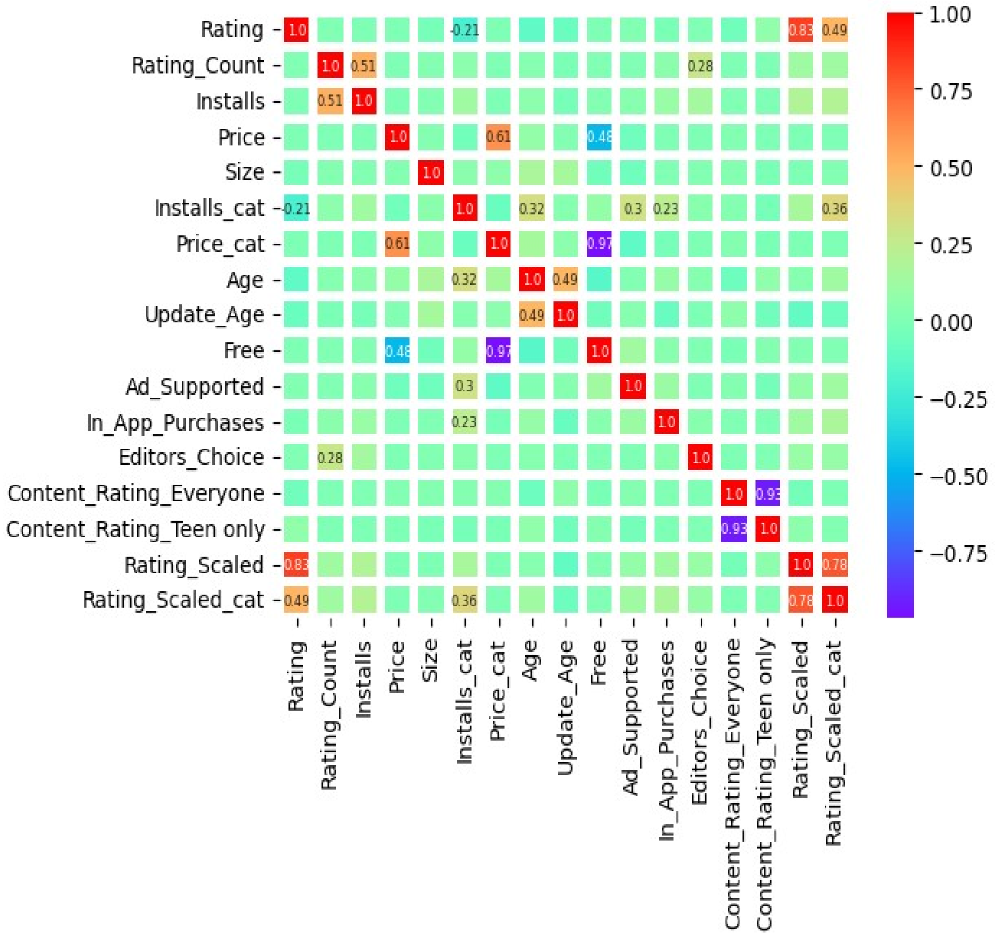


Class Imbalance

Binary	Multi-class	Percent
Low (~88%)	1	~5.8%
	2	~82%
High (~12%)	3	~11.5%
	4	~0.45%
	5	~0.11%

- Severe Class imbalance in Scaled & Transformed Ratings
- Binary Classification (“1”, ”2”, ”3”-- “Low”
“4”, “5”--”High”)

Correlation Heatmap Among Features



Only strnegth of Corerlation > 0.2 is annotated

- 1) A weak correlation between rating and installations, more frequently installed Apps get higher ratings
- 2) Features that can boost installations: In-App purchase feature, Ads_Supported feature, Age of Apps

Raw & Transformations of Features in Classification

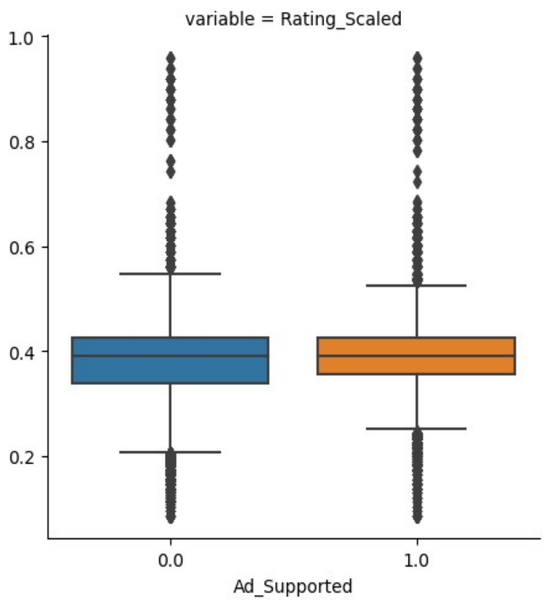
Features	Old Values	New Values
Install	Categorical	"under 1,000", "1,000 to 10,000","10,000 to 100,000", "above 100,000"
Content_Rating	Categorical	”Everyone” , “Adult Only”, "Teen Only"
Size	Categorical	“K”“M”“V”
Price	Categorical	“Free”“<= $10”,“> $10”
Age*	Numerical	Time in Years from “Released” to Jun. 15th 2021** Standardized by StandardScaler()
Updated_Age*	Numerical	Time in Years from “Last_Updated” to Jun. 15th 2021** Standardized by StandardScaler()
Rating*	Categorical	Rating_Scaled= Raw Rating/*log* (Rating_Counts) Re-map Rating_Scaled into standard rating scales of 1-5
Editors_Choice	Boolean	Yes/No
In_App_Purchase	Boolean	Yes/No
Ads_Supported	Boolean	Yes/No

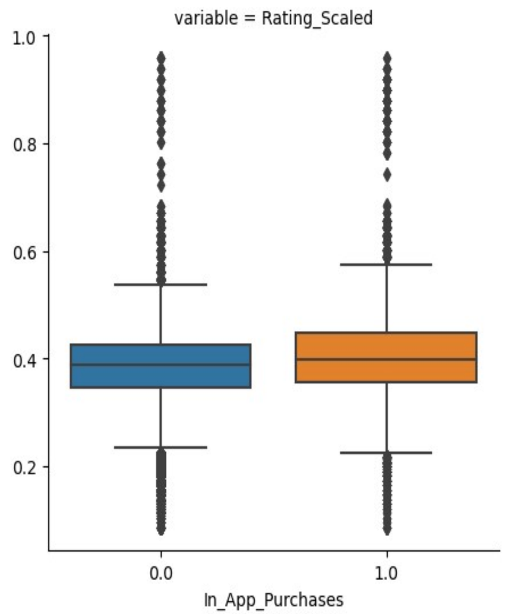
*New Calculated Features

**Jun 15th 2021 is when the data was scraped

Expanatry Data Analysis

- 3)However, In-App purchase & Ads_Supported each by itself does not show effect on App's rating score

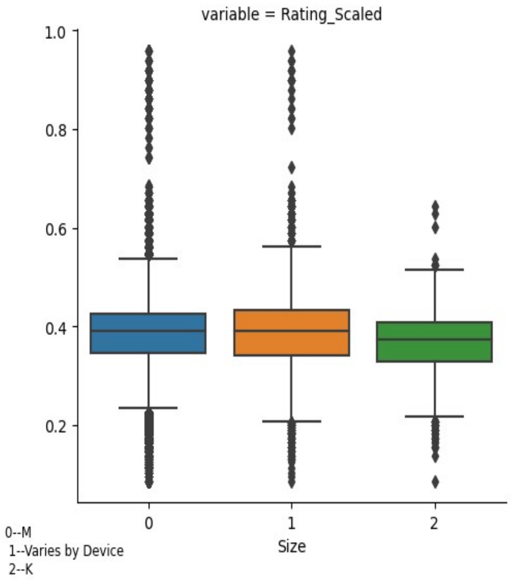


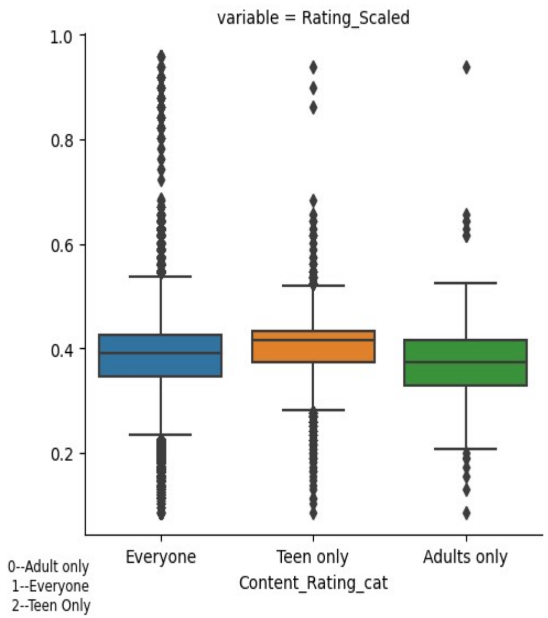


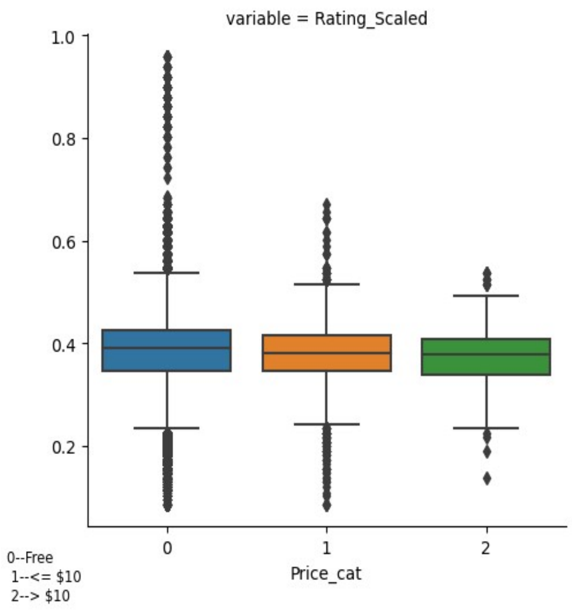
0--False
1--True

Expanatry Data Analysis

- 4) Size, Price & Content_Rating do not affect Rating

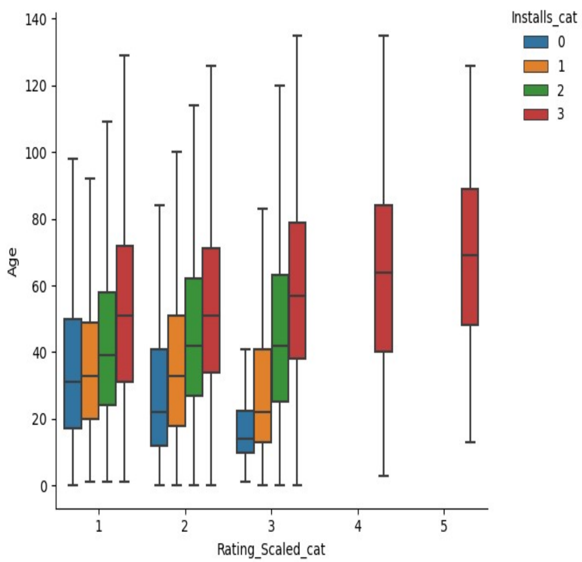


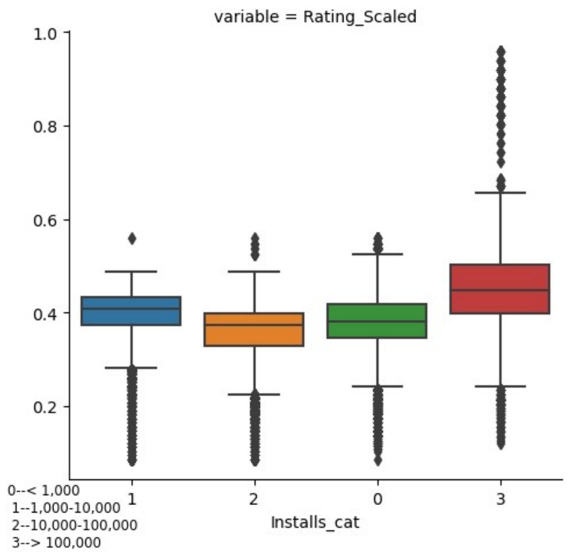




Expanatry Data Analysis

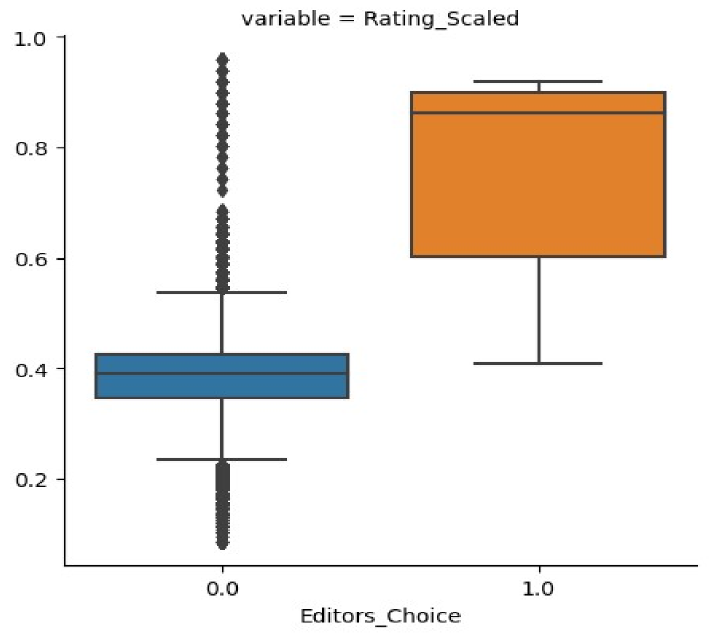
- 5) After controlling for installations, Age shows a postive relationship with Ratings, Older Apps have higher rating scores, which is especially true for Apps with installations above 100,000.





Expanatry Data Analysis

- 6) Being on the Editor’s Choice list is positively related to rating counts, which in turn entails higher rating scores.
- 7) Older Apps are updated less frequently and have longer duration since last updates.



0--False
1--True

Training and Model Selection

Features:

- *Ordinal Categorical Factors*: Installstion, Size, Price
- *Nominal Categorical Factors*: In_App_Purcahse, Ads_Supported,
  Editors_Choice, Content_Rating
- *Numerical Features*: Age, Updated_Age
- *Target Feature*: Scaled/Categorized/Binary Rating

Classifiers:

- *DecisionTreeClassifier*
- *RandomForestClassifier*
- *GradientBoostingClassifier*

Resampling:

- *RandomOverSampler*
- *RandomUnderSampler*

Hyperparameter Tuning:

- *RandomizedSearchCV*

Modeling

- ***Final Model***:

  GradientBoostingClassifier

- ***Best Parameters***:

  {'under__sampling': {Calss “Low”: 55578},
  'over__sampling': {Class “High”: 15225},
  'n_estimators': 297,   'max_depth': 22,
  'criteria': 'log_loss', 'learning_rate': 0.25}

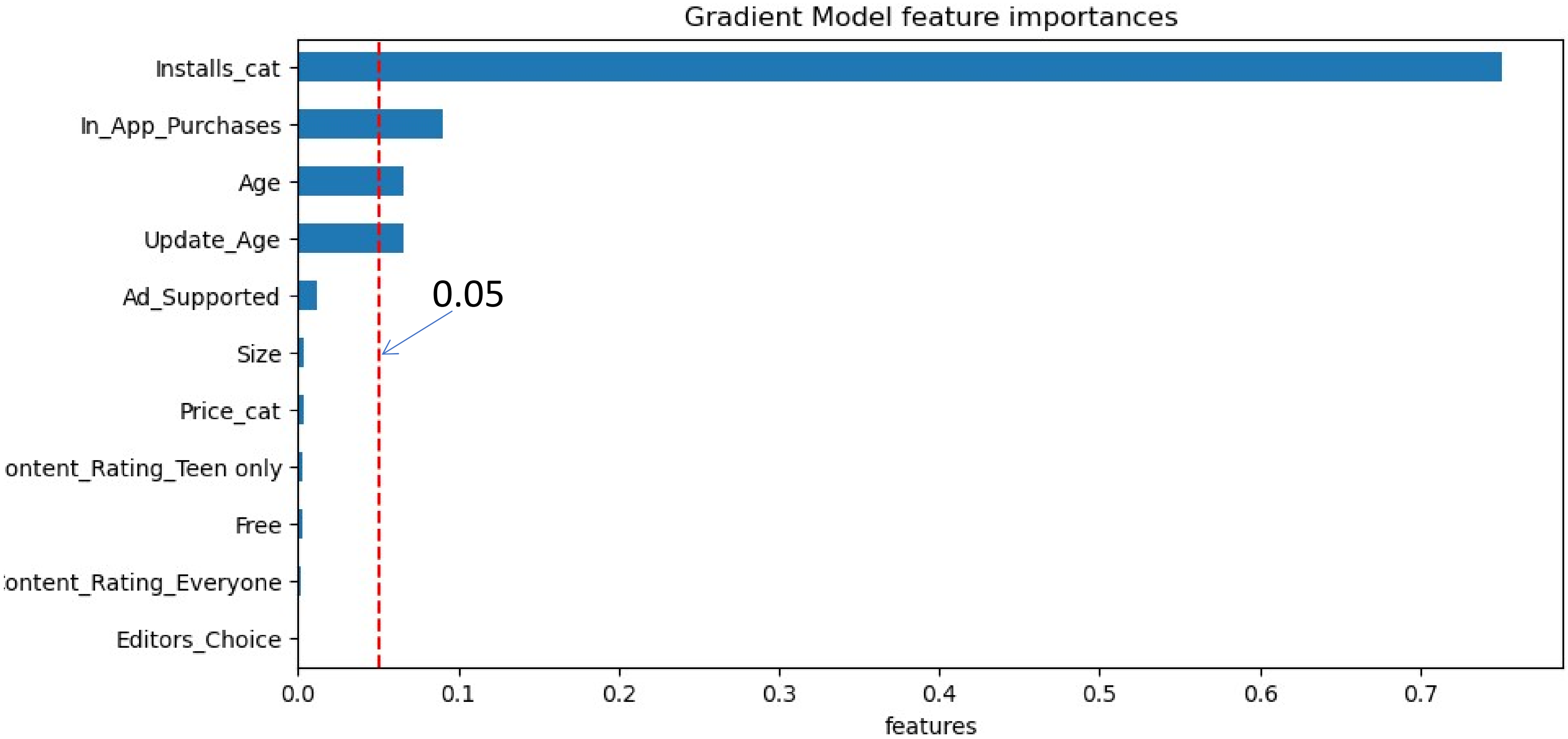
- ***Classification Results***:

  accuracy=0.89

  Low scores for Class “High”

	precision	recall	f1-score
Class “Low”	0.93	0.94	0.94
Class “High”	0.54	0.51	0.52
macro avg	0.74	0.73	0.73
weighted avg	0.89	0.89	0.89

Importance of Factors in Determing Salary



Decreasing Order:

Installation, In_App_Purchases, Age, Updated_Age

Conclusion

➢ *App's counts of "Installatons" is the dominant feature to forecast Rating*

More installation means higher rating

➢ *The following can be manipulated to boosting Installations*

- Having in-App-Purchase Feature
- Supporting Ads
- Allowing time for Apps to grow and having longer history

➢ *Having In-App_Purchases, Having a longer history can also bring up Ratings but not as significant*

➢ *Price, Size, Content_Rating have little or no effect on Ratings*

➢ *Alternatives need to be explored to improve classification performance* :

Discussion

- *Feature Engineering can be explored for possibly missed-out important features*

- *Alternative metrics that could better measure quality of Apps than Ratings should be looked into*

- *Finging alternatives as Target Feature*:

  e.g., Installation instead

- *Study other geners of Apps to gain some insights*