

Raw input columns

1. Ratings(label). → label\_indexer
2. Developer → split → extract selector
3. Publisher → split → extract selector
4. Platforms → split → extract selector → vectorAssembler
5. Categories → split → extract selector
6. Tags → split → extract selector
7. Price → doubleToVee → minMaxScaler → selector

Price:

8. In minMaxScaler (price\_features) Trans: minmaxscaler
- In selector (selected\_price\_features) Trans: ChiSqSelector(fpr)
- Minmaxscaler → ChiSqSelector(fpr)

Platform:

9. In extract (platform\_features) Trans: CountVectorizer
10. In selector (selected\_platform\_features) Trans: ChiSqSelector(fdr)
- CountVectorizer → ChiSqSelector(fdr)

Categories:

11. In extract (categories\_features) Trans: CountVectorizer
12. In selector (selected\_categories\_features) Trans: ChiSqSelector(fdr)
- CountVectorizer → ChiSqSelector(fdr)

Tags:

13. In extract ... Trans: CountVectorizer
14. In selector ... Trans: ChiSqSelector(fdr)
- CountVectorizer → ChiSqSelector(fdr)

Developer:

15. In extract ... Trans: CountVectorizer
16. In selector ... Trans: ChiSqSelector(fdr)
- CountVectorizer → ChiSqSelector(fdr)

Publisher:

17. In extract ... Trans: CountVectorizer
18. In selector .. Trans: ChiSqSelector(fdr)
- CountVectorizer → ChiSqSelector(fdr)

19. vectorAssembler (features) Trans: VectorAssembler

stages

(minmaxscaler → ChiSqSelector(fpr)  
→ Trans:CountVectorizer → ChiSqSelector(fdr)  
→ VectorAssembler. → RandomForest

columns count:20

developer

publisher

platforms

categories

tags

ratings

price

price\_features

selected\_price\_features

platforms\_features

selected\_platforms\_features

categories\_features

selected\_categories\_features

tags\_features

selected\_tags\_features

developer\_features

selected\_developer\_features

publisher\_features

selected\_publisher\_features

features