

Chat Application Success Prediction

Data Mining Project

Review 2

Slot: E1+TE1



FACULTY: PROF. THIPPA REDDY





Group Info



Data Mining Project

Created on 3 Mar '17



MEMBERS

3/1000



VIT IT Dhanashri Patil 15BIT0337



VIT IT Arnav Bhutani 15BIT0044



VIT IT Sunayna Ray 15BIT0225



Introduction

Sunayna

This is the era of internet and mobile phones.

Chat Applications are a major part of this digital revolution.

Gone are the days when people typed long emails to each other. People today want fast, short and easy communication methods. With this background, chat applications have proven to be of immense success.

Thus there is a large pool of customers for using chat applications.

Subsequently, there is a large pool of developers (and apps) too.

Hence we find that predicting the success rate of a chat application is a very relevant topic today and hence we took up this topic.

10:30 AM

Tool

Weka:

Weka contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions.

Advantages of WEKA

- Free availability under the GNU General Public License.
- Portability, since it is fully implemented in the Java programming language and thus runs on almost any modern computing platform.
- A comprehensive collection of data preprocessing and modeling techniques.
- Ease of use due to its graphical user interfaces.

Dataset

Viewer

Relation: app-weka.filters.unsupervised.attribute.Reorder-R1,2,3,4,5,6-weka.filters.un...

No.	1: Age	2: Top Developer	3: Ads	4: Size (MB)	5: Rating	6: Downloads (M)
	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal
1	18+	No	Yes	21.54	4	1
2	12+	No	No	8.31	4.5	0.1
3	18+	No	Yes	2.14	3.3	0.1
4	18+	Yes	Yes	14.37	4.4	10
5	18+	Yes	Yes	6.86	4.2	10
6	3+	Yes	No	34.09	4.2	100
7	12+	No	Yes	5.14	3.9	1
8	12+	No	Yes	16.88	4.4	10
9	18+	Yes	Yes	3.69	4.2	10
10	18+	No	No	4.65	4.2	0.5
11	18+	Yes	Yes	28.59	4.2	10
12	3+	No	Yes	19.42	3.8	0.5
13	18+	Yes	Yes	34.17	4.1	5
14	18+	No	Yes	3.96	3.5	0.5
15	18+	No	Yes	21.9	4.6	1
16	18+	Yes	Yes	29.09	4.3	100
17	18+	No	Yes	11.47	4	1
18	18+	Yes	Yes	18.76	4.2	10
19	12+	No	Yes	3.22	3.9	0.1
20	12+	No	Yes	4.32	3.7	0.1
21	12+	No	No	20.55	4.7	1
22	18+	Yes	Yes	8.03	4.1	1
23	18+	Yes	Yes	14.64	4	10
24	18+	No	No	15.3	4	1
25	18+	Yes	Yes	13.19	4.1	10
26	18+	No	Yes	33.89	4.3	10
27	18+	Yes	Yes	7.51	4.1	10
28	18+	No	Yes	9.14	4	0.5
29	18+	No	Yes	20.94	4.2	10
30	18+	No	Yes	2.56	4.2	1
31	3+	No	No	20.88	4.3	10
32	3+	No	Yes	15.79	5	10
33	12+	Yes	Yes	36.44	4.3	100
34	3+	No	No	35.43	4.3	50
35	18+	Yes	Yes	57	4.4	1000
36	18+	Yes	No	61.62	3.9	1000
37	18+	Yes	Yes	25.08	4.3	500
38	18+	Yes	Yes	4.83	4.3	100

Add instance Undo OK Cancel

Attribute: Age

Selected attribute

Name: Age
Missing: 0 (0%)

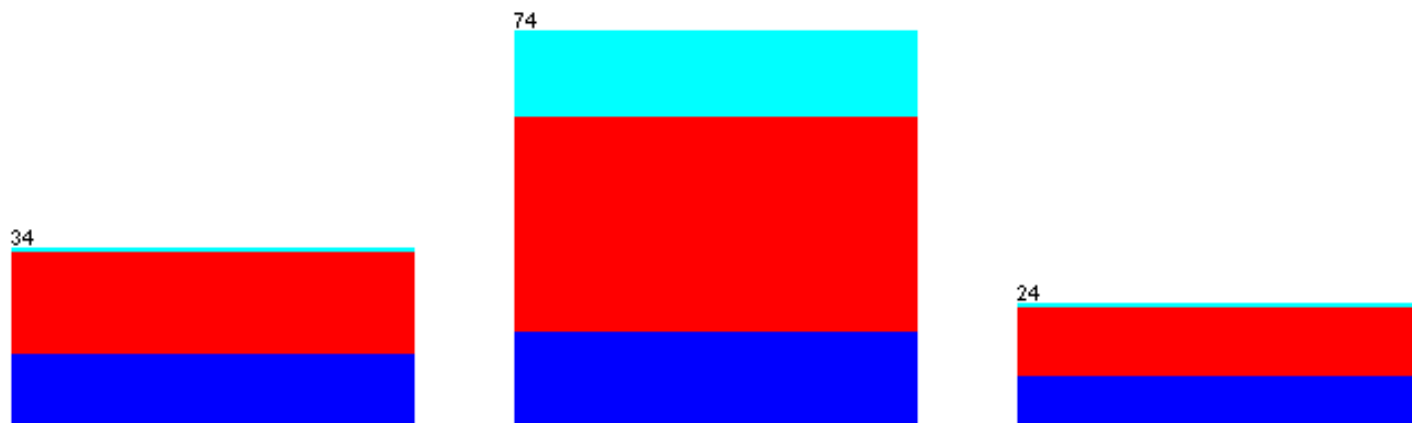
Distinct: 3

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	3+	34	34.0
2	18+	74	74.0
3	12+	24	24.0

Class: Downloads (M) (Nom)

Visualize All



Attribute: Top Developer

Selected attribute

Name: Top Developer
Missing: 0 (0%)

Distinct: 2

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Yes	44	44.0
2	No	88	88.0

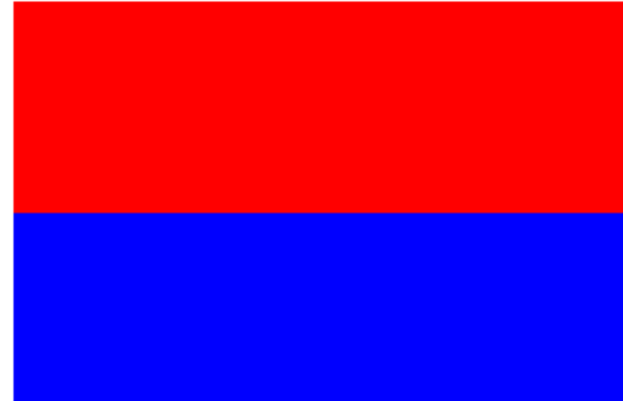
Class: Downloads (M) (Nom)

Visualize All

44



88



Attribute: Ads

Selected attribute

Name: Ads
Missing: 0 (0%)

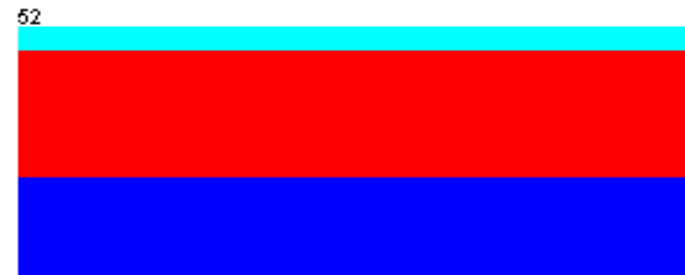
Distinct: 2

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Yes	80	80.0
2	No	52	52.0

Class: Downloads (M) (Nom)

Visualize All



Attribute: Size

Selected attribute

Name: Size (MB)

Missing: 0 (0%)

Distinct: 124

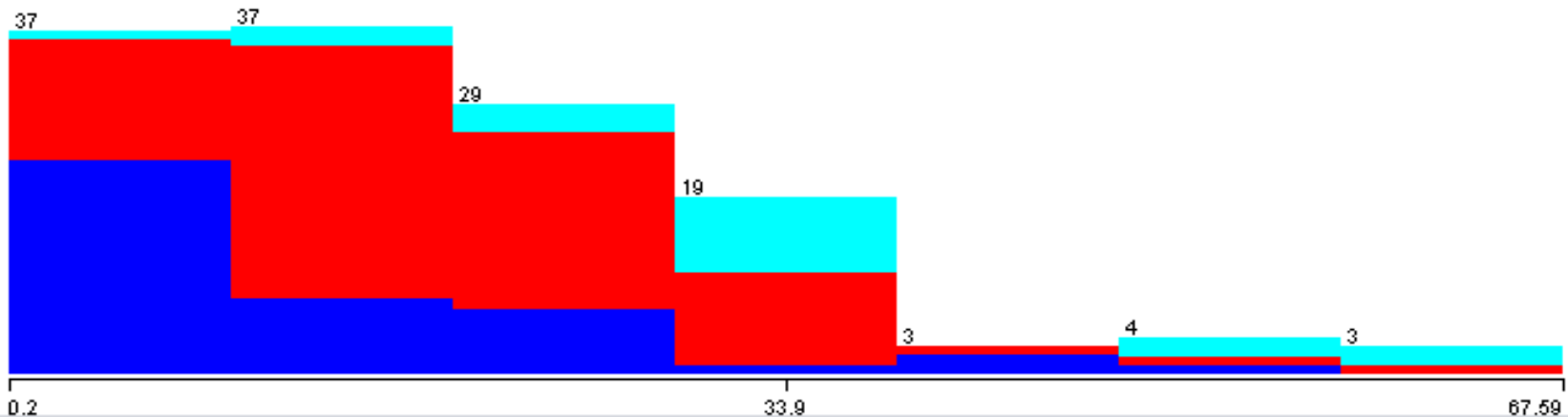
Type: Numeric

Unique: 116 (88%)

Statistic	Value
Minimum	0.2
Maximum	67.59
Mean	19.688
StdDev	14.551

Class: Downloads (M) (Nom)

Visualize All



Attribute: Rating

Selected attribute

Name: Rating
Missing: 0 (0%)

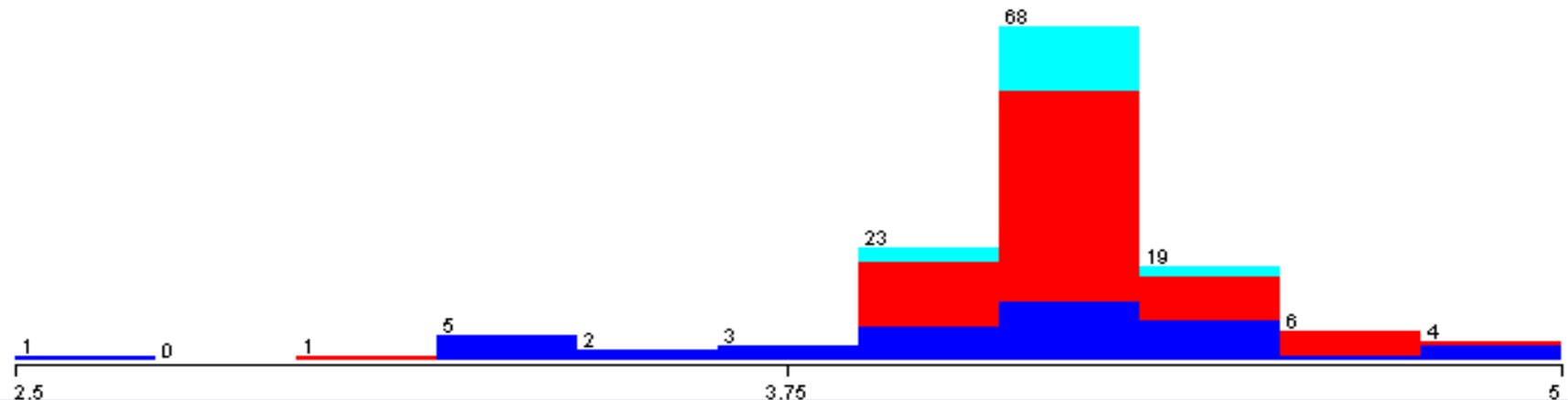
Distinct: 19

Type: Numeric
Unique: 7 (5%)

Statistic	Value
Minimum	2.5
Maximum	5
Mean	4.16
StdDev	0.341

Class: Downloads (M) (Nom)

Visualize All



Class Attribute: Downloads (M)

Selected attribute

Name: Downloads (M)

Missing: 0 (0%)

Distinct: 3

Type: Nominal

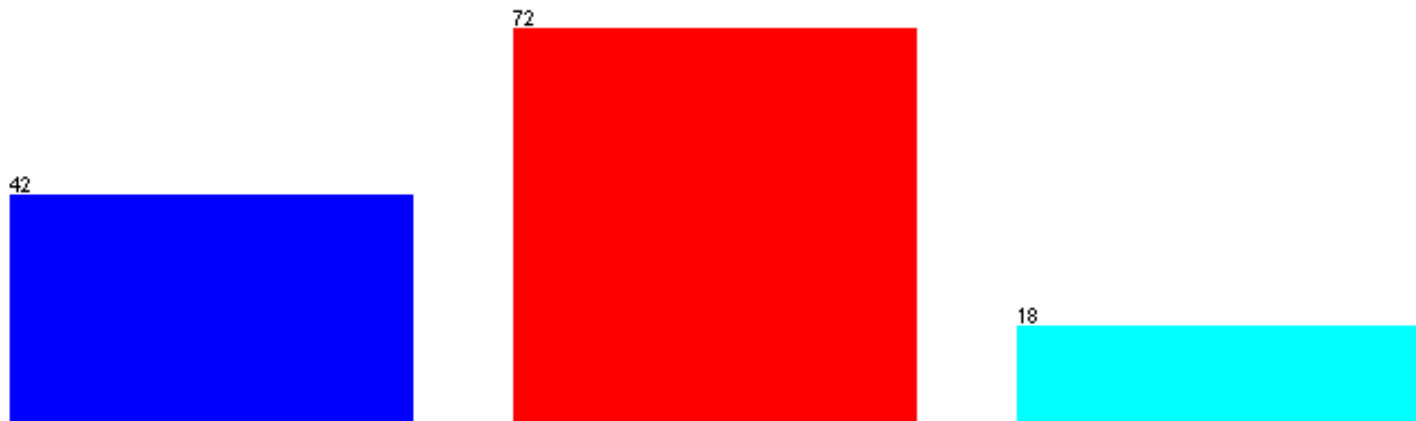
Unique: 0 (0%)

No.	Label	Count	Weight
1	Low	42	42.0
2	Average	72	72.0
3	High	18	18.0

Class: Downloads (M) (Nom)



Visualize All



Data Preprocessing

❖ We made the dataset by referring to play store. So for preprocessing we did the following:

- ❑ 1. Removing useless attributes: Different value for each tuple using mean and mode
- ❑ 2. Filter Examples: Removes the columns which were unnecessary.
(Name)
- ❑ 3. Replace missing values: In it missing values are replaced by the average value of that attribute. (Few of the apps had no specified age limit)
- ❑ 4. Discretization: Convert attribute: Downloads (M) values to 3 discrete classes:
Low, Average, High.
- ❑ 5. Normalisation



File Edit View

appbasic.arff *

Relation: Databasic

No.	1: Sr.No	2: Name	3: Age	4: Top Developer	5: Ads	6: Size (MB)	7: Rating	8: 5 Stars	9: Downloads (M)
	Numeric	Nominal	Nominal	Nominal	Nominal	String	Numeric	Nominal	String
5	21.0	LIVE ...	12	NO	NO	20.55	4.7		1
6	22.0	Meet4U	18+	Yes	Yes	8.03	4.1		1
7	24.0	Asian...	18+	No	No	15.3	4.0		1
8	30.0	4 Chat	18+	No	Yes	2.56	4.2		1
9		DISA		no	no	25.67	4.2	.1+	1
10		Maaii		no	no	47.92	4.4	70k	1
11		Voxer		yes	no	19.67	4.3	.15+	1
12		Soma...		no	yes	20.06	4.4	.3+	1
13		Singl...		yes	no	23.29	4.6	88k+	1
14		VMS v...		no	yes	13.75	4.1	4k+	1
15		SLIQ-...		no	no	10.84	4.2	5k+	1
16		Chat...	3+	No	No	14.85	4.1	15k	1
17		Primo	12+	No	Yes	35.83	4.1	24k	1
18		TalkU	3+	No	Yes	29	4.4	54k	1
19		Wire-...	3+	No	No	17	4.1	13k	1
20		Vona...	3+	No	No	20.74	4.1	17k	1
21		Chea...	3+	No	No	19.77	4.0	21k	1
22		trueC...	3+	No	No	9.78	4.0	4k	1
23		Near...	12+	No	Yes	5.14	3.9	11k	1
24		fiesta ...	12+	No	Yes	32.41	4.2	66k	1
25		mood...	3+	No	No	18.27	4.5	25k	1
26		MyDa...	12+	No	Yes	12.42	4.1	5k	1
27		Talk2	3+	No	Yes	14.85	3.9	19k	1
28	2.0	Glynk	12+	No	No	8.31	4.5		0.1
29	3.0	Stran...	18+	No	Yes	2.14	3.3		0.1
30	19.0	Girls ...	12+	No	Yes	3.22	3.9		0.1
31	20.0	India...	12+	No	Yes	4.32	3.7		0.1
32		imes...		no	yes	7	4.2	5k+	0.1
33		ChatOn		no	yes	598kb	3.9	1301	0.1
34		Just s...		no	no	20.47	4.3	2k+	0.1
35		Meec...		no	no	1.95	4.4	3k+	0.1
36		MeeM...		no	no	30.52	4.4	357only	0.1
37	4.0	Jaumo	18+	Yes	Yes	14.37	4.4		10
38	5.0	Hi	18+	Yes	Yes	6.86	4.2		10
39	8.0	Insta...	12+	No	Yes	16.88	4.4		10
40	9.0	Moco	18+	Yes	Yes	3.69	4.2		10
41	11.0	MeetMe	18+	Yes	Yes	28.59	4.2		10
42	18.0	Skout	18+	Yes	Yes	18.76	4.2		10

Initial Dataset



Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

RemoveUseless -M 99.0

Apply

Current relation

Relation: Databasic
Instances: 132Attributes: 9
Sum of weights: 132

Selected attribute

Name: Sr.No

Missing: 98 (74%)

Distinct: 34

Type: Numeric
Unique: 34 (26%)

Attributes

All

None

Invert

Pattern

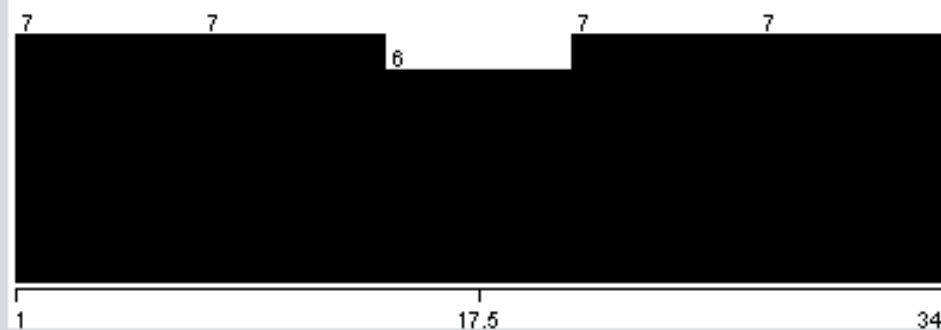
No.		Name
1	<input checked="" type="checkbox"/>	Sr.No
2	<input type="checkbox"/>	Name
3	<input type="checkbox"/>	Downloads (M)
4	<input type="checkbox"/>	Age
5	<input type="checkbox"/>	Top Developer
6	<input type="checkbox"/>	Ads
7	<input type="checkbox"/>	Size (MB)
8	<input type="checkbox"/>	Rating
9	<input type="checkbox"/>	5 Stars

Remove

Statistic	Value
Minimum	1
Maximum	34
Mean	17.5
StdDev	9.958

Class: 5 Stars (Nom)

Visualize All



Status

OK

Log

 x 0



Remove useless attribute:
Removes Name
And Sr. No.



Preprocess Classify Cluster Associate Select attributes Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

RemoveUseless -M 99.0

Apply

Current relation

Relation: Databasic-weka.filters.un...
Instances: 132

Attributes: 7
Sum of weights: 132

Selected attribute

Name: Downloads (M) Type: String
Missing: 0 (0%) Distinct: 21 Unique: 3 (2%)

Attributes

All

None

Invert

Pattern

No.		Name
1	<input checked="" type="checkbox"/>	Downloads (M)
2	<input type="checkbox"/>	Age
3	<input type="checkbox"/>	Top Developer
4	<input type="checkbox"/>	Ads
5	<input type="checkbox"/>	Size (MB)
6	<input type="checkbox"/>	Rating
7	<input type="checkbox"/>	5 Stars

Remove

Class: 5 Stars (Nom)

Visualize All

Attribute is neither numeric nor nominal.

Status

OK

Log

 x 0



Fill in missing data:

1. Few apps did not have age specification:
Default: 18+
2. Few of us noted down no. of 5 star ratings, few didn't: fill missing values.





Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose

ReplaceMissingValues

Apply

Current relation

Relation: appbasicv1-weka.filters.u...
Instances: 132

Attributes: 7
Sum of weights: 132

Attributes

AllNoneInvertPattern

No.		Name
1	<input checked="" type="checkbox"/>	Age
2	<input type="checkbox"/>	Top Developer
3	<input type="checkbox"/>	Ads
4	<input type="checkbox"/>	Size (MB)
5	<input type="checkbox"/>	Rating
6	<input type="checkbox"/>	5 Stars
7	<input type="checkbox"/>	Downloads

Remove

Selected attribute

Name: Age
Missing: 0 (0%)

Distinct: 3

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	18+	26	26.0
2	12+	24	24.0
3	3+	82	82.0

Class: Downloads (Num)

Visualize All



Status

OK

Log

x 0



Convert Numeric
value of Downloads
attribute to nominal
(Discretisation)





Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter

Choose Discretize -B 10 -M -1.0 -R first-last Apply

Current relation

Relation: appbasicv1-weka.filters.u... Attributes: 7
Instances: 132 Sum of weights: 132

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> Age
2	<input type="checkbox"/> Top Developer
3	<input type="checkbox"/> Ads
4	<input type="checkbox"/> Size (MB)
5	<input type="checkbox"/> Rating
6	<input type="checkbox"/> 5 Stars
7	<input checked="" type="checkbox"/> Downloads

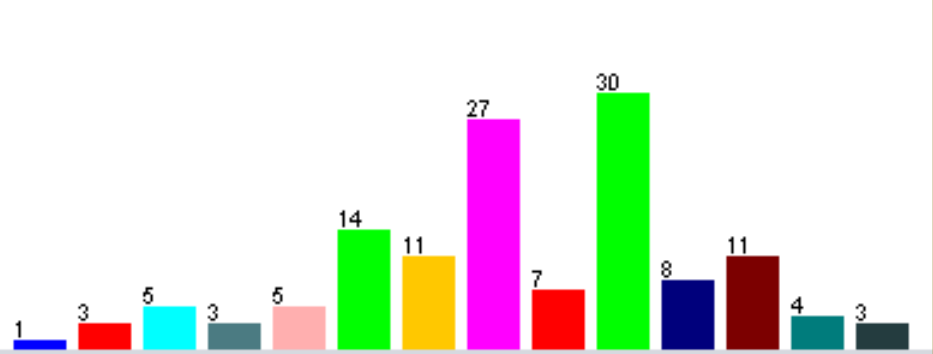
Remove

Selected attribute

Name: Downloads
Missing: 0 (0%) Distinct: 14 Type: Nominal
Unique: 1 (1%)

No.	Label	Count	Weight
1	0.0005	1	1.0
2	0.001	3	3.0
3	0.005	5	5.0
4	0.01	3	3.0

Class: Downloads (Nom) Visualize All



Status

OK Log x 0



Normalize all
attributes
except class
attribute





Preprocess Classify Cluster Associate Select attributes Visualize

Open file...

Open URL...

Open DB...

Generate...

Undo

Edit...

Save...

Filter

Choose **Normalize** -S 5.0 -T 0.0 -unset-class-temporarily Apply

Current relation

Relation: appbasicv1-weka.filters.u... Attributes: 7
Instances: 132 Sum of weights: 132

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> Age
2	<input type="checkbox"/> Top Developer
3	<input type="checkbox"/> Ads
4	<input type="checkbox"/> Size (MB)
5	<input checked="" type="checkbox"/> Rating
6	<input type="checkbox"/> 5 Stars
7	<input type="checkbox"/> Downloads

Remove

Selected attribute

Name: Rating
Missing: 0 (0%)
Distinct: 19
Type: Nominal
Unique: 7 (5%)

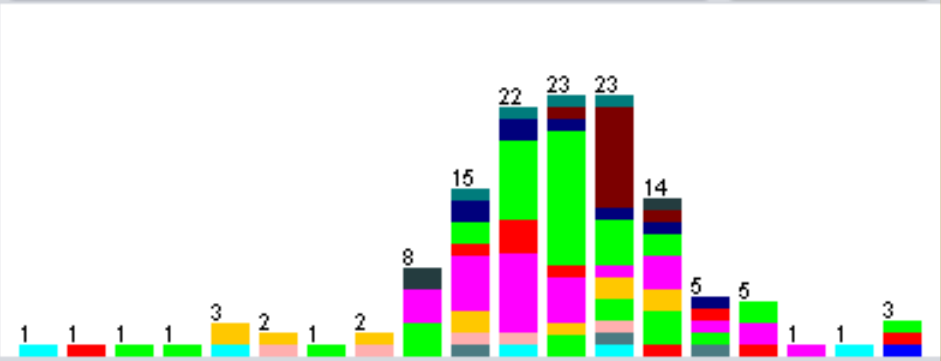
No.	Label	Count	Weight
1	2.5	1	1.0
2	3.1	1	1.0
3	3.2	1	1.0
4	3.3	1	1.0

Class: Downloads (Nom) Visualize All

Status

OK

Log  x 0





Merge
infrequent
nominal values
in class attribute





Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter

Choose Apply

Current relation

Relation: app.symbolic Attributes: 6
Instances: 132 Sum of weights: 132

Attributes

All | None | Invert | Pattern

No.	Name
1	<input type="checkbox"/> Age
2	<input type="checkbox"/> Top Developer
3	<input type="checkbox"/> Ads
4	<input type="checkbox"/> Size (MB)
5	<input type="checkbox"/> Rating
6	<input checked="" type="checkbox"/> Downloads (M)

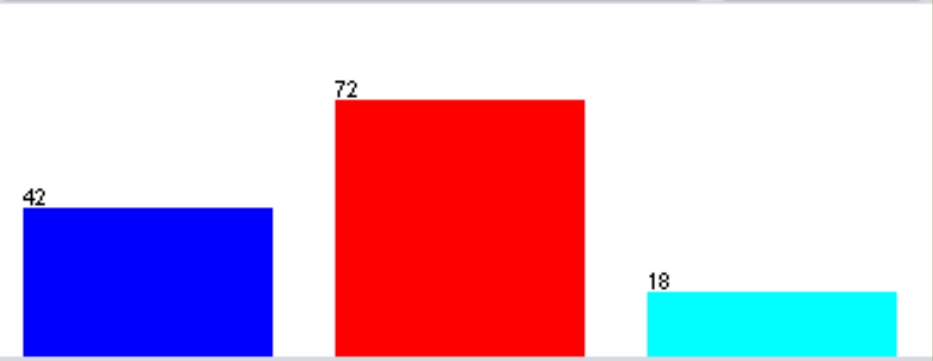
Remove

Selected attribute

Name: Downloads (M) Type: Nominal
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

No.	Label	Count	Weight
1	Low	42	42.0
2	Average	72	72.0
3	High	18	18.0

Class: Downloads (M) (Nom) ▼ Visualize All



Status

OK | Log x 0

