

PERTEMUAN 11

TEKNIK ASSOCIATION RULE MINING DALAM R

TUJUAN PRAKTIKUM

Mahasiswa akan dapat menggunakan teknik-teknik dasar *association rule mining* dengan *tools R*

TEORI PENUNJANG

Analisis asosiasi berguna untuk menemukan hubungan penting yang tersembunyi di antara set data yang sangat besar. Hubungan yang sudah terbuka direpresentasikan dalam bentuk aturan asosiasi (*association rule*) atau set aturan item yang sering muncul. Aturan asosiasi adalah pernyataan implikasi bentuk $X \rightarrow Y$, dimana X dan Y adalah *itemset* yang lepas (*disjoint*), yang memenuhi persyaratan $X \cap Y = \{\}$. Kekuatan aturan asosiasi dapat diukur dengan *support* dan *confidence*. *Support* digunakan untuk menentukan seberapa banyak aturan dapat diterapkan pada set data, sedangkan *confidence* digunakan untuk menentukan seberapa sering item di dalam Y muncul dalam transaksi yang berisi X .

LAPORAN PENDAHULUAN

1. Apa yang anda ketahui tentang *Association Rule Mining*?
2. Sebutkan algoritme asosiasi yang anda ketahui!

MATERI PRAKTIKUM

Algoritme Apriori dengan *tools R*

Titanic dataset adalah data 4 dimensi dengan informasi nasib penumpang di Titanic yang diringkaskan menurut *social class*, *sex*, *age*, dan *survival*. Untuk membuat data ini cocok untuk *association rule mining*, kita perlu melakukan praproses data, di mana setiap baris mewakili 1 orang dengan langkah sebagai berikut:

```
> str(Titanic)
> df <- as.data.frame(Titanic)
> head(df)
> titanic.raw <- NULL
> for(i in 1:4) { titanic.raw <- cbind(titanic.raw,
rep(as.character(df[,i]), df$Freq)) }
> titanic.raw <- as.data.frame(titanic.raw)
> names(titanic.raw) <- names(df)[1:4]
> dim(titanic.raw)
> str(titanic.raw)
> head(titanic.raw)
> summary(titanic.raw)
```

Setelah data selesai di praproses, selanjutnya dilakukan tahapan *association rule mining* dengan menggunakan algoritme apriori:

```
> library(arules)
> # find association rules with default settings
> rules <- apriori(titanic.raw)
> rules
> quality(rules) <- quality(rules)
> inspect(rules)

# rules with rhs containing "Survived" only
> rules <- apriori(titanic.raw, parameter = list(minlen=2,
supp=0.005, conf=0.8), appearance =
list(rhs=c("Survived=No", "Survived=Yes"), default="lhs"),
control = list(verbose=F))
> rules.sorted <- sort(rules, by="lift")
> inspect(rules.sorted)

> # find redundant rules
> subset.matrix <- is.subset(rules.sorted, rules.sorted)
> subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
> redundant <- colSums(subset.matrix, na.rm=T) >= 1
> which(redundant)

> # remove redundant rules
> rules.pruned <- rules.sorted[!redundant]
> inspect(rules.pruned)

> # interpreting rules
> rules <- apriori(titanic.raw, parameter = list(minlen=3,
supp=0.002, conf=0.2), appearance =
list(rhs=c("Survived=Yes"), lhs=c("Class=1st",
"Class=2nd", "Class=3rd", "Age=Child", "Age=Adult"),
default="none"), control = list(verbose=F))
> rules.sorted <- sort(rules, by="confidence")
> inspect(rules.sorted)
```

| |
|-----------------------|
| DAFTAR PUSTAKA |
|-----------------------|

- Han J, Kamber M, and Pei J. 2012. *Data Mining: Concept and Techniques*. Massachusetts: Morgan Kaufmann.
- Yanchang Zhao. 2012. R and Data Mining: Examples and Case Studies. <http://www.rdatamining.com/docs/RDataMining.pdf>