# Laporan LKP 2 Data Mining

#### Kelompok 1 :

Abdul Aziz Nurachmady G64160077

Muhammad Fakhri 'Alauddin Hidayat G64170015

Fakhoor Izaaz Wildhanrahman G64170017

1. Jelaskan persamaan dan perbedaan terkait statistika ringkasan dan visualisasi data dengan menggunakan Weka dan R!

|  |
| --- |
| Pada Statistika Ringkasan, Weka dan R sama-sama dapat memberikan nilai minimum, maksimum dan mean. Perbedannya adalah Weka memuat nilai standar deviasi dan R bisa memberikan nilai Q1, median, Q3.  Pada Visualisasi, Weka hanya bisa menampilkan histogram dan scatter plot sedangkan R mampu menampilkan histogram, pie chart, line chart, density, boxplot, scatter plot, dan juga mampu menvisualisasikan untuk kasus pada ​multiple variables dengan melalui ​barchart, boxplot, scatterplot, pairsplot, serta dengan menghitung nilai covariance​ dan ​correlation ​antar variabel. |

1. Bukalah dataset ​ New York Air Quality Measurements yang terdapat pada R! Ada berapakah jumlah atribut dan data yang ada pada dataset tersebut? Sebutkan tipe data dari masing-masing atribut pada dataset tersebut!

|  |  |  |
| --- | --- | --- |
| Ada 6 atribut yaitu   1. Ozone bertipe int 2. Solar.R bertipe int 3. Wind bertipe num 4. Temp bertipe int 5. Month bertipe int 6. Day bertipe int   Terdapat 153 objek data   |  |  | | --- | --- | | *Kode untuk nomor 2* | *Hasil Console dari kode untuk nomor 2* | |

1. Bukalah ​ adult\_census\_income.csv di R! Jelaskan hasil dari statistika ringkasan dari setiap atribut (​ workclass​ , education​ , occupation​ , marital.status​ , capital.gain​ , capital.loss​ , ​ hours.per.week​ , dan​ income​ )!

|  |  |
| --- | --- |
| Statistika Ringkasan Atribut | Penjelasan |
| Atribut Workclass | Pada atribut workclass,  ada sebanyak 22696 orang yang bekerja di sektor Private,  960 orang di sektor Federal-gov,  1298 orang di sektor State-gov,  2093 orang di sektor Local-gov,  7 orang tidak pernah bekerja,  1116 orang di sektor Self-emp-inc,  2541 orang di sektor Self-emp-not-inc,  14 orang bekerja tanpa dibayar,  1836 orang tidak diketahui data workclassnya.  Paling banyak orang bekerja di sektor private. |
| Atribut Education | Sebanyak  51 orang berpendidikan di Preschool,  168 orang berpendidikan di SD kelas 1-4,  333 orang berpendidikan di SD kelas 5 - 6,  646 orang berpendidikan di SMP kelas 7 - 8,  514 orang berpendidikan di SMP kelas 9, 933 orang berpendidikan di SMA kelas 10,  1175 orang berpendidikan di SMA kelas 11,  433 orang berpendidikan SMA kelas 12,  5355 orang berpendidikan Sarjana (S1),  413 orang berpendidikan Doktor (S2),  1723 orang berpendidikan Master (S3),  1067 orang berpendidikan Assoc-acdm,  1382 orang berpendidikan Assoc-voc,  10501 orang lulusan SMA,  7291 orang berpendidikan lainnya.  Paling banyak adalah orang yang lulusan SMA. |
| Atribut Occupation | Pada atribut Occupation sebanyak  3770 orang bekerja di bidang adm-clerical,  9 orang bekerja sebagai tentara,  4099 orang bekerja di bidang craft-repair,  4066 orang bekerja sebagai executive managerial,  994 orang bekerja di bidang pertanian dan perikanan,  1370 orang bekerja di bidang kebersihan,  2002 orang bekerja di bidang machine operator inspector,  149 orang bekerja di bidang private house service,  4140 orang bekerja di bidang profesion specialty.  649 orang bekerja di bidang layanan perlindungan,  3650 orang bekerja di bidang penjualan,  928 orang bekerja di bidang tech-support,  1597 orang bekerja di bidang transportasi,  3295 orang bekerja di bidang jasa lainnya |
| Atribut marital.status | Sebanyak 4443 orang bercerai, 993 orang berstatus janda, 10683 orang belum pernah menikah, 1025 orang berpisah, 418 orang yang menikah tapi tidak tinggal bersama dengan pasangannya, 23 orang yang berstatus Married-AF-spouse, 14976 orang berstatus Married-civ-spouse |
| Atribut capital.gain | Pada atribut capital.gain, rata-ratanya sebesar 1078. Nilai minimumnya adalah 0 dan nilai maksimalnya adalah 99999 dengan nilai tengah (median) sebesar 0. |
| Atribut capital.loss | Pada atribut capital.loss, rata-ratanya adalah sebesar 87.3. Nilai minimumnya adalah 0 dan nilai maksimalnya adalah 4356.0 dengan nilai tengah (median) sebesar 0.0 |
| Atribut hours.per.week | Pada atribut hours per week, Rata-rata orang bekerja selama 40.44 jam per-minggu nya. Selain itu, waktu terendah orang bekerja selama per-minggu nya adalah satu jam dan waktu terlama adalah 99 jam per-minggu. |
| Atribut Income | Pada atribut income, terdapat 24.720 orang yang memiliki penghasilan kurang dari 50K dan terdapat 7.841 yang memiliki penghasilan lebih dari 50K |

1. Masih berkaitan dengan nomor 3, bandingkan hasil visualisasi data histogram dan scatter plot pada atribut ​ education.num dan ​ hour.per.week dengan menggunakan weka dan R! Jelaskan hasil masing-masing diagram!

|  |
| --- |
| **# Using R**        **#Using Weka**  Histogram of education.num    Histogram of hours per week    Scatter plot of education.num vs hours per week    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Dalam visualisasi data menggunakan Histogram, implementasi dengan R dapat menghasilkan visualisasi data histogram lebih jelas dibandingkan dengan menggunakan Weka, data histogram dengan R terdapat frequency sebagai sumbu Y, sedangkan pada Weka hanya menampilkan sumbu X saja.Namun dalam visualisasi data menggunakan scatterplot, implementasi dengan Weka lebih unggul dalam memvisualisasikan data dibandingkan implementasi menggunakan R. Data scatterplot yang dihasilkan dapat dibaca denga mudah karena data telah dibedakan dengan warna dan bentuk plot dengan 2 jenis, yakni dot dan cross. Dalam segi kemudahan pula implementasi dengan Weka lebih unggul karena dapat memvisualisasikan seluruh atribut dengan scatterplot secara otomatis, jadi pengguna tinggal memilih perbandingan atribut pada scatterplot yang diinginkan. |