

## DAFTAR ISI

|                                                    |       |
|----------------------------------------------------|-------|
| PERNYATAAN.....                                    | iii   |
| PRAKATA .....                                      | vi    |
| DAFTAR ISI .....                                   | viii  |
| DAFTAR GAMBAR .....                                | xi    |
| DAFTAR TABEL .....                                 | xiv   |
| DAFTAR PERSAMAAN .....                             | xv    |
| DAFTAR ALGORITMA.....                              | xvi   |
| DAFTAR LAMPIRAN .....                              | xvii  |
| INTISARI .....                                     | xviii |
| ABSTRACT .....                                     | xix   |
| BAB 1. PENDAHULUAN .....                           | 1     |
| 1.1 Latar Belakang.....                            | 1     |
| 1.2 Rumusan Masalah.....                           | 3     |
| 1.3 Batasan Masalah .....                          | 3     |
| 1.4 Tujuan Penelitian .....                        | 3     |
| 1.5 Manfaat Penelitian .....                       | 4     |
| 1.6 Metode Penelitian .....                        | 4     |
| 1.7 Sistematika Penulisan .....                    | 5     |
| BAB 2. TINJAUAN PUSTAKA .....                      | 7     |
| BAB 3. LANDASAN TEORI.....                         | 12    |
| 3.1 Prediksi .....                                 | 12    |
| 3.1.1 Prediksi kualitatif.....                     | 12    |
| 3.1.2 Prediksi kuantitatif.....                    | 12    |
| 3.2 Pengembalian Produk atau Retur.....            | 13    |
| 3.3 Belanja Online ( <i>Online Shopping</i> )..... | 13    |
| 3.4 Bagan Alir ( <i>Flowchart</i> ).....           | 14    |
| 3.4.1 Program <i>flowchart</i> .....               | 15    |
| 3.5 UML (Unified Modelling Language) .....         | 17    |
| 3.5.1 Use case diagram .....                       | 18    |
| 3.5.2 Activity diagram .....                       | 18    |

|        |                                                                            |    |
|--------|----------------------------------------------------------------------------|----|
| 3.6    | Data Mining .....                                                          | 19 |
| 3.7    | Klasifikasi .....                                                          | 20 |
| 3.8    | Algoritma k-NN ( <i>k-Nearest Neighbor</i> ).....                          | 21 |
| 3.8.1  | Similaritas atribut numerik .....                                          | 22 |
| 3.8.2  | Similaritas atribut nominal.....                                           | 23 |
| 3.8.3  | Similaritas tipe data campuran.....                                        | 23 |
| 3.9    | <i>F-measure</i> .....                                                     | 24 |
| 3.10   | Pengujian Tingkat Akurasi .....                                            | 25 |
| BAB 4. | ANALISIS DAN RANCANGAN SISTEM.....                                         | 27 |
| 4.1    | Analisis Sistem.....                                                       | 27 |
| 4.1.1  | Deskripsi sistem .....                                                     | 27 |
| 4.1.2  | Analisis kebutuhan sistem.....                                             | 28 |
| 4.2    | Rancangan Sistem.....                                                      | 34 |
| 4.2.1  | Rancangan algoritma .....                                                  | 36 |
| 4.2.2  | Rancangan penyimpanan data.....                                            | 40 |
| 4.2.3  | Rancangan pengujian sistem.....                                            | 40 |
| 4.2.4  | Rancangan sistem.....                                                      | 42 |
| 4.2.5  | Rancangan <i>user interface</i> .....                                      | 55 |
| BAB 5. | IMPLEMENTASI.....                                                          | 62 |
| 5.1    | Spesifikasi Sistem .....                                                   | 62 |
| 5.1.1  | Spesifikasi perangkat keras.....                                           | 62 |
| 5.1.2  | Spesifikasi perangkat lunak .....                                          | 62 |
| 5.2    | Implementasi Proses .....                                                  | 63 |
| 5.2.1  | Implementasi proses pembacaan data <i>training</i> .....                   | 63 |
| 5.2.2  | Implementasi proses inisialisasi data <i>testing</i> .....                 | 64 |
| 5.2.3  | Implementasi proses perhitungan similaritas.....                           | 68 |
| 5.2.4  | Implementasi proses perankingan.....                                       | 74 |
| 5.2.5  | Implementasi proses perhitungan <i>F-measure</i> dan <i>Accuracy</i> ..... | 75 |
| 5.3    | Implementasi <i>User Interface</i> .....                                   | 77 |
| 5.3.1  | Implementasi jendela utama sistem .....                                    | 77 |
| 5.3.2  | Implementasi tab Training data .....                                       | 80 |
| 5.3.3  | Implementasi tab Testing data .....                                        | 81 |
| 5.3.4  | Implementasi tab Class testing data.....                                   | 82 |
| 5.3.5  | Implementasi tab Similarity .....                                          | 84 |

|                                       |                                                                              |     |
|---------------------------------------|------------------------------------------------------------------------------|-----|
| 5.3.6                                 | Implementasi tab Ranking .....                                               | 85  |
| 5.3.7                                 | Implementasi kotak dialog <i>running time</i> .....                          | 86  |
| BAB 6. PENGUJIAN DAN PEMBAHASAN ..... |                                                                              | 87  |
| 6.1                                   | Pengujian Sistem.....                                                        | 87  |
| 6.1.1                                 | Proses pengecekan data <i>training</i> .....                                 | 87  |
| 6.1.2                                 | Proses penginputan data <i>testing</i> .....                                 | 89  |
| 6.1.3                                 | Proses pemilihan data <i>random</i> .....                                    | 91  |
| 6.1.4                                 | Proses klasifikasi satu set kumpulan <i>record</i> data <i>testing</i> ..... | 93  |
| 6.1.5                                 | Proses klasifikasi satu record data <i>testing</i> .....                     | 100 |
| 6.2                                   | Pengujian Algoritma .....                                                    | 110 |
| 6.2.1                                 | Pengujian parameter algoritma <i>k-Nearest Neighbor</i> .....                | 110 |
| 6.3                                   | Pengujian Hasil Prediksi Retur .....                                         | 113 |
| 6.3.1                                 | Pengujian hasil prediksi data <i>testing</i> .....                           | 113 |
| 6.3.2                                 | Perhitungan <i>F-measure</i> .....                                           | 119 |
| 6.3.3                                 | Perhitungan akurasi.....                                                     | 119 |
| BAB 7. PENUTUP.....                   |                                                                              | 122 |
| 7.1                                   | Kesimpulan .....                                                             | 122 |
| 7.2                                   | Saran .....                                                                  | 123 |
| DAFTAR PUSTAKA .....                  |                                                                              | 124 |
| 9. LAMPIRAN.....                      |                                                                              | 127 |

## DAFTAR GAMBAR

|             |                                                                                                                                                     |    |
|-------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Gambar 4.1  | <i>Flowchart</i> sistem prediksi .....                                                                                                              | 35 |
| Gambar 4.2  | <i>Flowchart</i> dari algoritma <i>k</i> -NN .....                                                                                                  | 38 |
| Gambar 4.3  | <i>Use case diagram</i> sistem prediksi .....                                                                                                       | 43 |
| Gambar 4.4  | <i>Activity diagram</i> proses prediksi .....                                                                                                       | 45 |
| Gambar 4.5  | <i>Activity diagram</i> proses prediksi berdasarkan input satu set <i>record</i><br>data <i>testing</i> .....                                       | 47 |
| Gambar 4.6  | <i>Activity diagram</i> proses prediksi berdasarkan input data <i>random</i><br>yang dipilih dari satu set <i>record</i> data <i>testing</i> .....  | 49 |
| Gambar 4.7  | <i>Activity diagram</i> proses prediksi berdasarkan input satu <i>record</i><br>yang dipilih dari satu set <i>records</i> data <i>testing</i> ..... | 51 |
| Gambar 4.8  | <i>Activity diagram</i> proses prediksi berdasarkan input satuan yang<br>dimasukkan pengguna tiap fitur satu per satu .....                         | 54 |
| Gambar 4.9  | Jendela utama sistem prediksi .....                                                                                                                 | 56 |
| Gambar 4.10 | Tab Training data .....                                                                                                                             | 57 |
| Gambar 4.11 | Tab Testing data .....                                                                                                                              | 58 |
| Gambar 4.12 | Tab Class testing data .....                                                                                                                        | 59 |
| Gambar 4.13 | Tab Similarity .....                                                                                                                                | 60 |
| Gambar 4.14 | Tab Ranking .....                                                                                                                                   | 60 |
| Gambar 4.15 | Kotak dialog <i>running time</i> .....                                                                                                              | 61 |
| Gambar 5.1  | Kode proses pembacaan data <i>training</i> yang kemudian diisikan ke<br>dalam tabel yang ada di tab Training data .....                             | 63 |
| Gambar 5.2  | Kode proses input data <i>testing</i> yang kemudian diisikan ke dalam<br>tabel yang ada di tab Testing data .....                                   | 65 |
| Gambar 5.3  | Kode program pemilihan beberapa data <i>random</i> dari data <i>testing</i> . 66                                                                    |    |
| Gambar 5.4  | Kode program pemilihan data <i>random</i> sesuai <i>radio button</i> yang<br>dipilih .....                                                          | 67 |
| Gambar 5.5  | Kode program <i>Euclidean distance</i> .....                                                                                                        | 68 |
| Gambar 5.6  | Kode program <i>nominal similarity</i> .....                                                                                                        | 69 |
| Gambar 5.7  | Kode program pemilahan fitur .....                                                                                                                  | 69 |

|             |                                                                                                       |    |
|-------------|-------------------------------------------------------------------------------------------------------|----|
| Gambar 5.8  | Kode program similaritas Length date .....                                                            | 70 |
| Gambar 5.9  | Kode program similaritas Item ID, Size, Color, dan Manufacturer<br>ID .....                           | 71 |
| Gambar 5.10 | Kode program similaritas Price .....                                                                  | 71 |
| Gambar 5.11 | Kode program similaritas Customer ID dan Salutation .....                                             | 72 |
| Gambar 5.12 | Kode program similaritas Date of Birth, State, dan Creation Date .....                                | 72 |
| Gambar 5.13 | Kode program similaritas total dari variabel heterogen .....                                          | 73 |
| Gambar 5.14 | Kode program <i>sorting</i> nilai similaritas secara <i>descending</i> .....                          | 74 |
| Gambar 5.15 | Kode program pemilihan label mayoritas .....                                                          | 75 |
| Gambar 5.16 | Kode program perhitungan nilai <i>Precision</i> dan <i>Recall</i> .....                               | 76 |
| Gambar 5.17 | Kode program perhitungan <i>F-measure</i> dan <i>Accuracy</i> .....                                   | 77 |
| Gambar 5.18 | Antarmuka jendela utama sistem .....                                                                  | 78 |
| Gambar 5.19 | Antarmuka tab Training data .....                                                                     | 81 |
| Gambar 5.20 | Antarmuka tab Testing data .....                                                                      | 81 |
| Gambar 5.21 | Kotak dialog pemilihan file data testing .....                                                        | 82 |
| Gambar 5.22 | Antarmuka tab Class testing data .....                                                                | 83 |
| Gambar 5.23 | Antarmuka tab Similarity .....                                                                        | 84 |
| Gambar 5.24 | Antarmuka tab Ranking .....                                                                           | 85 |
| Gambar 5.25 | Antarmuka kotak dialog <i>running time</i> .....                                                      | 86 |
| Gambar 6.1  | Proses pengecekan data .....                                                                          | 88 |
| Gambar 6.2  | Proses input data melalui tab Testing data .....                                                      | 89 |
| Gambar 6.3  | Kotak dialog untuk memasukkan data <i>testing</i> .....                                               | 90 |
| Gambar 6.4  | Tabel pada tab Testing data yang berisi data <i>testing</i> .....                                     | 90 |
| Gambar 6.5  | Proses penginputan parameter <i>random data</i> dengan optional<br><i>without missing value</i> ..... | 91 |
| Gambar 6.6  | Tabel berisi output data <i>random</i> tanpa <i>missing value</i> .....                               | 92 |
| Gambar 6.7  | Data <i>random</i> tanpa <i>missing value</i> .....                                                   | 92 |
| Gambar 6.8  | Tabel berisi output data <i>random</i> dengan <i>missing value</i> .....                              | 92 |
| Gambar 6.9  | Data <i>random</i> dengan <i>missing value</i> .....                                                  | 93 |
| Gambar 6.10 | Memasukkan nilai <i>k</i> untuk proses klasifikasi .....                                              | 94 |
| Gambar 6.11 | Kotak dialog <i>running time</i> hasil klasifikasi satu set data <i>testing</i> .....                 | 94 |

|             |                                                                                          |     |
|-------------|------------------------------------------------------------------------------------------|-----|
| Gambar 6.12 | Pemilihan 1 <i>record</i> data tanpa <i>missing value</i> untuk klasifikasi ..           | 101 |
| Gambar 6.13 | Klasifikasi 1 <i>record</i> data tanpa <i>missing value</i> .....                        | 102 |
| Gambar 6.14 | Output similaritas klasifikasi 1 <i>record</i> data tanpa <i>missing value</i>           | 103 |
| Gambar 6.15 | Output ranking klasifikasi 1 <i>record</i> data tanpa <i>missing value</i> ....          | 104 |
| Gambar 6.16 | Pengujian data <i>record</i> dari penginputan tiap fitur .....                           | 106 |
| Gambar 6.17 | Output similaritas klasifikasi 1 <i>record</i> data dari penginputan tiap<br>fitur ..... | 107 |
| Gambar 6.18 | Output ranking klasifikasi 1 <i>record</i> data dari penginputan tiap fitur<br>.....     | 108 |

## DAFTAR TABEL

|           |                                                                                                                                     |     |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabel 2.1 | Perbandingan penelitian mengenai <i>k</i> -NN .....                                                                                 | 10  |
| Tabel 3.1 | Simbol-simbol pada program <i>flowchart</i> (Andriani, 2009) .....                                                                  | 15  |
| Tabel 3.2 | <i>Similarity</i> dan <i>dissimilarity</i> untuk atribut tunggal .....                                                              | 22  |
| Tabel 4.1 | Fitur-fitur pada data pembelian .....                                                                                               | 29  |
| Tabel 6.1 | Output hasil klasifikasi satu set data <i>testing</i> .....                                                                         | 96  |
| Tabel 6.2 | Daftar ranking 21 data dengan nilai similaritas tertinggi pada<br>klasifikasi 1 <i>record</i> data tanpa <i>missing value</i> ..... | 104 |
| Tabel 6.3 | Daftar ranking 21 data dengan nilai similaritas tertinggi pada<br>klasifikasi 1 record data dari penginputan tiap fitur .....       | 108 |
| Tabel 6.4 | Hasil pengujian 100 data acak tanpa <i>missing value</i> .....                                                                      | 111 |
| Tabel 6.5 | Hasil pengujian 100 data acak dengan <i>missing value</i> .....                                                                     | 112 |
| Tabel 6.6 | Ringkasan hasil klasifikasi data <i>testing</i> .....                                                                               | 114 |

## **DAFTAR PERSAMAAN**

|                                                                |    |
|----------------------------------------------------------------|----|
| Persamaan (3.1) Euclidean distance.....                        | 23 |
| Persamaan (3.2) Konversi dissimilarity ke similarity.....      | 23 |
| Persamaan (3.3) Similaritas global atau similaritas total..... | 24 |
| Persamaan (3.4) Precision, Recall, F-measure.....              | 25 |
| Persamaan (3.5) F-measure dengan bobot sama.....               | 25 |
| Persamaan (3.6) Akurasi .....                                  | 25 |

## DAFTAR ALGORITMA

|                                                          |    |
|----------------------------------------------------------|----|
| Algoritma 3.1 Algoritma <i>k-Nearest Neighbor</i> .....  | 21 |
| Algoritma 3.2 Similaritas obyek beratribut nominal ..... | 23 |
| Algoritma 3.3 Similaritas obyek tipe heterogen .....     | 24 |

## DAFTAR LAMPIRAN

|                                                                      |     |
|----------------------------------------------------------------------|-----|
| A. Gambaran data <i>training</i> setelah <i>pre-processing</i> ..... | 127 |
| B. Gambaran data <i>testing</i> setelah <i>pre-processing</i> .....  | 133 |
| C. Data acak tanpa <i>missing value</i> .....                        | 138 |
| D. Data acak dengan <i>missing value</i> ..                          | 146 |