

# Akurasi Algoritma C4.5 dalam Klasifikasi Data 5K2S

Des Suryani\*, Ause Labellapansa, Cahaya Fitri

Jurusan Teknik Informatika, Fakultas Teknik

Universitas Islam Riau

Jl. Kaharuddin Nasution No. 113 Marpoyan Pekanbaru

des.suryani@eng.uir.ac.id\*, ause.labella@eng.uir.ac.id, cahaya.uchiha69@gmail.com

**Abstrak**— PT. Kimia Tirta Utama merupakan salah satu perusahaan yang mengelola perkebunan jenis kelapa sawit dan memproduksi salah satu bahan makanan yaitu minyak goreng. Agar mendapatkan kualitas minyak goreng yang baik dan layak untuk dikonsumsi, perusahaan ini perlu memperhatikan setiap standar produksi, kebersihan, serta kelayakan mesin produksi yang digunakan pada pabrik pengolahannya dengan menerapkan konsep produksi bersih dan Ketertiban, Kerapian, Kebersihan, Kedisiplinan, Kelestarian, Semangat Kerja dan *Safety* (5K2S). Konsep ini digunakan untuk mendukung tercapainya proses produksi yang memenuhi standar dan menjadi acuan untuk melakukan penilaian terhadap hasil akhir sebagai evaluasi bagi perusahaan. Dalam sistem penilaian 5K2S, perusahaan menggunakan peringkat warna, yaitu: emas, hijau, biru, merah dan hitam. Hitam merupakan peringkat paling rendah, sedangkan emas adalah paling tinggi. Penilaian 5K2S dilakukan oleh *Assesor Health, Safety, and Environment* (HSE) dengan meninjau langsung ke lapangan. Untuk membantu perusahaan dalam peningkatan kualitas keputusan penilaian 5K2S, perlu dilakukan kajian terhadap data penilaian 5K2S tersebut berdasarkan pada kriteria penilaian 5K2S di setiap stasiun pabrik pengolahan kelapa sawit. Dalam kajian ini dilakukan proses *data mining* menggunakan metode klasifikasi dengan algoritma Decision Tree C4.5. Kriteria/atribut yang digunakan dalam penilaian untuk menghasilkan pola nilai akhir terdiri dari 26 atribut, yaitu: pengelolaan sampah, lantai dan area kerja, jalan, mesin, crane atau alat angkat, dinding, jendela dan atap, tiang dan struktur, rak dan lemari, got/saluran air, penempatan barang/benda kerja, penyimpanan bahan mudah terbakar, material B3 (bahan kimia), alat pemadam kebakaran, sistem pemadam kebakaran, gambar petunjuk (poster-poster), instruksi kerja (IK), pakaian kerja, alat pelindung diri, panel listrik, lampu penerangan, toilet/wc/kamar mandi, sertifikat izin operasional (khusus pabrik sawit), penanganan limbah, tempat istirahat/makan, penyimpanan barang, sanitasi dan *higiene* dengan atribut *label/class* yaitu nilai akhir. Berdasarkan tahapan dalam proses *Data Mining* 5K2S yang telah dilakukan, dihasilkan pengetahuan berupa pohon keputusan sebagai hirarki aturan keputusan dengan tingkat akurasi 82.05%. Hasil ini dapat digolongkan dalam klasifikasi baik (*good classification*) sehingga dapat dijadikan acuan bagi perusahaan dalam menentukan keputusan perolehan nilai akhir.

**Kata kunci**—*data mining*; 5K2S; klasifikasi; decision tree.

## I. PENDAHULUAN

Sumber daya manusia dan sumber daya alam merupakan hal terpenting dalam sebuah perusahaan. Disamping itu dengan dukungan teknologi dan *engineering* yang sesuai standar dapat

mempengaruhi terkelolanya perusahaan dengan baik. Dalam hal ini, perusahaan perlu memperhatikan aspek-aspek tersebut agar kualitas produk menjadi tetap terjaga, khususnya pada perusahaan yang bergerak di bidang produksi bahan makanan.

Salah satu perusahaan yang bergerak dalam bidang produksi bahan makanan tersebut adalah PT. Kimia Tirta Utama. Perusahaan ini mengelola perkebunan jenis kelapa sawit yang memproduksi salah satu bahan makanan yaitu minyak goreng. Untuk mendapatkan kualitas minyak goreng yang baik dan layak untuk dikonsumsi, perusahaan tersebut perlu memperhatikan setiap standar produksi, kebersihan, serta kelayakan mesin produksi yang digunakan pada pabrik pengolahannya.

Untuk mengelola dan melestarikan lingkungan demi keberlanjutan usahanya, PT. Kimia Tirta Utama menerapkan Kebijakan Lingkungan, Keselamatan, dan Kesehatan Kerja (LK3). Disamping itu LK3 juga dijadikan pedoman untuk menerapkan konsep produksi bersih dan Ketertiban, Kerapian, Kebersihan, Kedisiplinan, Kelestarian, Semangat Kerja dan *Safety* (5K2S) untuk mendukung tercapainya proses produksi yang memenuhi standar. 5K2S dijadikan acuan untuk melakukan penilaian terhadap hasil akhir pengelolaan dan jumlah kecelakaan kerja pada perusahaan. Hasil dari penilaian 5K2S tersebut akan dijadikan bahan untuk melakukan evaluasi bagi perusahaan.

Sistem penilaian 5K2S menggunakan peringkat nilai akhir dengan warna, yaitu: emas, hijau, biru, merah dan hitam. Hitam merupakan peringkat paling rendah, sedangkan emas adalah yang paling tinggi. Penilaian 5K2S dilakukan oleh *Assesor Health, Safety, and Environment* (HSE) dengan meninjau langsung ke lapangan. *Assesor* akan memberikan penilaian berupa *point*, khususnya di pabrik pengolahan pada bagian stasiun-stasiun pabrik setiap bulannya. Kemudian seluruh *point* tersebut akan diakumulasikan untuk mendapatkan hasil akhir penilaian dalam bentuk peringkat warna.

Jumlah data yang banyak sebenarnya dapat bermanfaat untuk menjadi sumber data historis untuk menemukan suatu pola dan pengetahuan baru yang dapat bermanfaat di masa depan. Dalam pencarian pola dan pengetahuan baru dari data tersebut memerlukan suatu teknologi *Data Mining* yang dapat memecahkan masalah “*too much data, less information*” [1].

Berdasarkan uraian tersebut, perlu dilakukan kajian terhadap himpunan data penilaian 5K2S pada PT. Kimia Tirta Utama untuk membantu perusahaan dalam menghasilkan

pengetahuan (*knowledge*) berupa pola dalam bentuk pohon keputusan (*decision tree*). Pohon keputusan ini dapat menggambarkan aturan-aturan keputusan untuk melakukan penilaian dalam memperoleh nilai akhir. Kajian terhadap himpunan data 5K2S ini akan dilakukan beberapa tahapan dalam proses *Data Mining* dan proses pemodelan penilaian menggunakan metode klasifikasi dengan algoritma C4.5. Algoritma C4.5 merupakan metode berbasis pohon keputusan yang menggunakan pendekatan induksi dimana dalam pendekatan ini, algoritma C4.5 membagi data berdasarkan kriteria yang dipilih untuk membuat sebuah pohon keputusan yang menggunakan pendekatan secara *top-down*[2]. Hasil kajian ini akan diperoleh pengetahuan berupa pola dalam bentuk pohon keputusan (*decision tree*) berdasarkan kriteria-kriteria penilaian 5K2S di setiap stasiun pabrik pengolahan.

## II. METODE

Penelitian adalah usaha menemukan dengan proses yang metodis untuk memperkaya pengetahuan itu sendiri serta dengan yang lainnya, oleh penemuan fakta dan wawasan tidak biasa. Ada empat metode penelitian yang umum digunakan, diantaranya: *Experiment, Action Research, Survey dan Case Study*[3].

Penelitian ini menggunakan metode eksperimen yaitu penelitian yang menyangkut tindakan pada parameter atau variabel terkait dan memakai tes yang dikendalikan oleh peneliti itu sendiri.

### A. Data Mining

Data tidak dapat dipisahkan dari kehidupan sehari-hari. Suatu organisasi dapat dibanjiri dengan berbagai macam data. Sangatlah tidak berguna data yang terkumpul dan tersimpan jika tidak dimanfaatkan. Permasalahan yang muncul adalah bagaimana memproses data tersebut sehingga dapat menampilkan suatu pengetahuan (*knowledge*) yang berguna bagi manajemen dalam mengambil keputusan. Dalam proses penemuan pengetahuan dapat digunakan algoritma *Data Mining* untuk menemukan pola yang bermanfaat.

*Data mining* didefinisikan sebagai suatu proses yang menggunakan berbagai perangkat analisis data untuk menemukan pola dan relasi data agar dapat digunakan untuk membuat prediksi yang tepat. Tahapan proses penemuan pengetahuan adalah pembersihan data, integrasi data, seleksi data, transformasi data, proses *Data mining* dan evaluasi pola. Sebelum proses *Data mining* dilakukan, data melewati proses pembersihan, integrasi, seleksi, dan transformasi terlebih dahulu dalam data *warehouse*. Pada *Data mining* dilakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui dari suatu data[4].

### B. Decision Tree

*Decision tree* merupakan suatu pendekatan yang sangat populer dan praktis dalam *machine learning* untuk menyelesaikan permasalahan klasifikasi. Konsep *decision tree* pada dasarnya adalah mengubah data menjadi pohon keputusan dan aturan hirarki (aturan-aturan keputusan) yang pada perkembangan selanjutnya dapat disederhanakan dengan menghilangkan cabang-cabang atau aturan-aturan yang tidak

perlu. Sedangkan atribut-atribut yang menyertai data tersebut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan pohon. Salah satu atribut merupakan atribut yang menyatakan data penyelesaian per *item* data yang disebut dengan klasifikasi atau kelas. Atribut memiliki nilai-nilai yang dinamakan dengan *instance*. Misalkan atribut pendapatan mempunyai *instance* berupa rendah, menengah, dan tinggi.

Pohon keputusan terdiri dari *node* yang membentuk pohon berakar, yang berarti pohon tersebut diarahkan oleh *node* yang disebut akar (*root*) yang tidak memiliki masukan (*input*). Semua *node* lain memiliki tepat satu masukan (*input*). Sebuah *node* tanpa keluaran (*output*) disebut *internal node*. Dan semua selain *node* disebut daun (*leaf*) atau seringkali dikenal sebagai *decision node*. Untuk setiap *leaf*, sebuah keputusan telah dibuat dan diaplikasikan ke semua observasi di dalam *leaf*. *Leaf* tersebut merupakan hasil akhir pengelompokan data dari pohon keputusan yang membagi ke dalam kelompok klasifikasi.

### C. Algoritma C4.5

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

- 1) Pilih atribut sebagai akar
- 2) Buat cabang untuk masing-masing nilai
- 3) Bagi kasus dalam cabang
- 4) Ulangi proses untuk masing-masing cabang sampai
- 5) semua kasus pada cabang memiliki kelas yang sama

Untuk memilih atribut akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk mendapatkan nilai *Gain* (2), harus ditentukan terlebih dahulu nilai *Entropy* seperti pada (1).

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

- $S$  = Himpunan Kasus
- $n$  = Jumlah Partisi  $S$
- $p_i$  = Proporsi dari  $S_i$  terhadap  $S$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

- $S$  = Himpunan Kasus
- $A$  = Atribut
- $n$  = Jumlah Partisi Atribut  $A$
- $|S_i|$  = Jumlah Kasus pada partisi ke- $i$
- $|S|$  = Jumlah Kasus dalam  $S$

### D. Langkah-Langkah Penelitian

Sebagai suatu rangkaian proses, *Data Mining* dapat dibagi menjadi beberapa tahap proses. Tahap-tahap dalam penelitian adalah sebagai berikut:

#### 1) Himpunan data (Dataset)

Pada tahap ini perlu dilakukan pemahaman dan pengolahan data dengan cara:

#### a) Pembersihan data (*data cleaning*).

Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan.

#### b) Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Seluruh data 5K2S digabungkan mulai dari tahun 2004 sampai 2016.

c) Seleksi data (*data selection*)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Dalam hal ini, atribut yang diperlukan pada data 5K2S dengan 26 atribut kriteria, yaitu : 1) pengelolaan sampah, 2) lantai dan area kerja, 3) jalan, 4) mesin, 5) crane atau alat angkat, 6) dinding, jendela dan atap, 7) tiang dan struktur, 8) rak dan lemari, 9) got/saluran air, 10) penempatan barang/benda kerja, 11) penyimpanan bahan mudah terbakar, 12) material B3(bahan kimia), 13) alat pemadam kebakaran, 14) sistem pemadam kebakaran, 15) gambar petunjuk(poster-poster), 16) instruksi kerja (IK), 17) pakaian kerja, 18) alat pelindung diri, 19) panel listrik, 20) lampu penerangan, 21) toilet/wc/kamar mandi, 22) sertifikat izin operasional (khusus pabrik sawit), 23) penanganan limbah, 24) tempat istirahat/makan, 25) penyimpanan barang, 26) sanitasi dan hygiene dengan label/class berupa nilai akhir.

d) Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining. Format atribut data yang digunakan adalah 1) sampah, 2) area kerja, 3) jalan, 4) mesin, 5) crane, 6) DJA, 7) tiang, 8) RL, 9) got, 10) pb\_kerja, 11) BMT, 12) materialB3, 13) APAR, 14) hydran, 15) poster, 16) IK, 17) pakaian, 18) APD, 19) panel listrik, 20) lampu, 21) toilet, 22) SIO, 23) limbah, 24) t\_istirahat, 25) gudang, 26) sanitasi dan nilai akhir (nilai) sebagai atribut label/class. Dan data yang akan dianalisa sebagai data *training* sebanyak 156 record. Adapun nilai akhir untuk penilaian 5K2S tersebut terdiri dari warna merah = 47 data, hijau=68 data, biru 32 data dan merah 9 data.

2) Proses mining

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data. Dalam hal ini menggunakan metode klasifikasi dengan algoritma Decision Tree C4.5 yang akan menghasilkan pengetahuan berupa pola/model/tree/rule.

3) Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Evaluasi yang digunakan adalah *Accuracy*.

4) Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

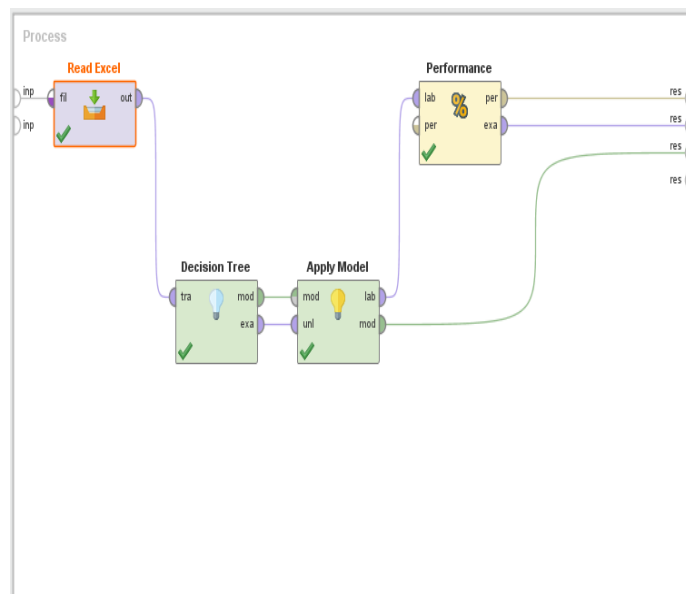
### III. HASIL DAN DISKUSI

#### A. Proses Pemodelan Penilaian 5K2S

Dalam proses pemodelan ini dibutuhkan dataset untuk dilakukan analisa. Dataset yang digunakan bersifat *private* dataset yang diperoleh dari PT. Kimia Tirta Utama. Dalam hal ini data disimpan dalam sebuah file 5K2S.xls.

Untuk mengolah dataset tersebut menggunakan Rapidminer Studio Version 7.0 untuk mendapatkan pengetahuan. Rapidminer ini merupakan *software tool Open Source* untuk *Data Mining*. Rapidminer menyediakan prosedur *Data Mining* dan *Machine Learning* termasuk *extraction, transformation, loading* (ETL), *data preprocessing, visualisasi, modeling* dan evaluasi.

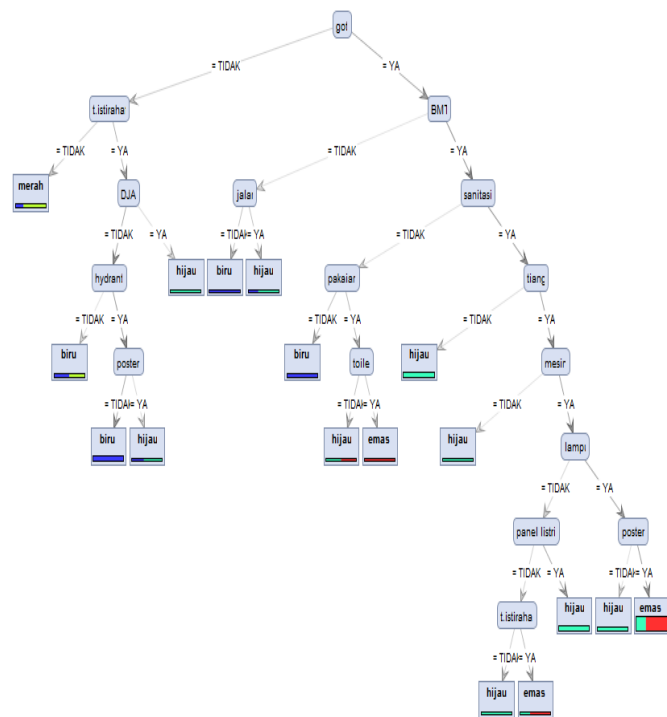
Sebelum melakukan proses pemodelan, 26 atribut sebagai parameter yang menyebabkan *class/label/target* terjadi masing-masingnya ditentukan tipe nilai berupa binominal karena target adalah polynominal karena bernilai lebih dari dua nilai yaitu : emas, hijau, biru dan merah. Model rancangan (*design*) dari proses yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Proses model penilaian 5K2S

#### B. Pohon Keputusan Penilaian 5K2S

Berdasarkan jenis dataset 5K2S tersebut, metode yang cocok/sesuai digunakan adalah klasifikasi dengan algoritma Decision Tree (C4.5). Pemodelan ini akan menghasilkan sebuah pohon keputusan seperti Gambar 2. Hasil nilai akhir dari klasifikasi tersebut ada empat nilai, yaitu emas, hijau, biru dan merah. Dan atribut got/saluran air merupakan atribut akar tertinggi dari penilaian 5K2S tersebut.



Berdasarkan Tabel 1, dari 156 jumlah *record* terdapat 26 *record* diprediksi secara benar memperoleh nilai akhir biru, 48 *record* diprediksi secara benar memperoleh nilai akhir hijau, 8 *record* diprediksi secara benar nilai akhir merah dan 46 *record* diprediksi secara benar memperoleh nilai akhir emas sehingga dapat ditentukan tingkat akurasi =  $(26+48+8+46)/156=82.05\%$ . Tingkat akurasi tersebut dapat diklasifikasikan baik (*good classification*)[5].

TABEL 1. TINGKAT AKURASI HASIL PENILAIAN 5K2S

	true biru	true hijau	true merah	true emas	class precision
pred biru	26	0	1	0	96.30%
pred hijau	3	48	0	1	92.31%
pred merah	3	0	8	0	72.73%
pred emas	0	20	0	46	69.70%
class recall	81.25%	70.59%	88.89%	97.87%	

Gambar 2. Pohon keputusan (*decision tree*) penilaian 5K2S

### C. Rule Penilaian 5K2S

Berdasarkan pohon keputusan pada Gambar 2, model pohon tersebut dapat diubah menjadi aturan (*rule*):

```

got = TIDAK
  Istirahat = TIDAK: merah {biru=3, hijau=0, merah=8, emas=0}
  Istirahat = YA
    DJA = TIDAK
      hydrant = TIDAK: biru {biru=1, hijau=0, merah=1, emas=0}
      hydrant = YA
        poster = TIDAK: biru {biru=19, hijau=0, merah=0, emas=0}
        poster = YA: hijau {biru=2, hijau=3, merah=0, emas=0}
      DJA = YA: hijau {biru=0, hijau=4, merah=0, emas=0}
    got = YA
      BMT = TIDAK
        jalan = TIDAK: biru {biru=3, hijau=0, merah=0, emas=0}
        jalan = YA: hijau {biru=1, hijau=2, merah=0, emas=0}
      BMT = YA
        sanitasi = TIDAK
          pakalari = TIDAK: biru {biru=3, hijau=0, merah=0, emas=0}
          pakalari = YA
            toilet = TIDAK: hijau {biru=0, hijau=1, merah=0, emas=1}
            toilet = YA: emas {biru=0, hijau=0, merah=0, emas=2}
          sanitasi = YA
            tiang = TIDAK: hijau {biru=0, hijau=11, merah=0, emas=0}
            tiang = YA
              mesin = TIDAK: hijau {biru=0, hijau=5, merah=0, emas=0}
              mesin = YA
                lampu = TIDAK
                  panel listrik = TIDAK
                    istirahat = TIDAK: hijau {biru=0, hijau=2, merah=0, emas=0}
                    istirahat = YA: emas {biru=0, hijau=1, merah=0, emas=2}
                  panel listrik = YA: hijau {biru=0, hijau=13, merah=0, emas=0}
                lampu = YA
                  poster = TIDAK: hijau {biru=0, hijau=7, merah=0, emas=0}
                  poster = YA: emas {biru=0, hijau=19, merah=0, emas=42}

```

### D. Performance Vector Penilaian 5K2S

Tahap evaluasi dari proses mining adalah pengukuran dari seberapa baik model mengkorelasikan antara hasil dengan atribut dalam data yang telah disediakan. Evaluasi terhadap klasifikasi data 5K2S ini menggunakan *Confusion Matrix: Accuracy* yang dapat digambarkan pada Tabel 1.

## IV. KESIMPULAN

Berdasarkan hasil kajian dan proses pemodelan menggunakan algoritma Decision Tree C4.5 yang dilakukan terhadap data 5K2S pada PT. Kimia Tirta Utama1. Data 5K2S dapat dijadikan sebagai data *training* untuk menghasilkan pengetahuan dengan menggunakan metode klasifikasi.

Hasil kajian data *training* perolehan nilai akhir 5K2S dapat membentuk pengetahuan berupa model pohon keputusan yang mengubah data menjadi pohon keputusan yang merepresentasikan aturan.

Tingkat akurasi yang diperoleh dari pengolahan data mencapai 82,05% yang dapat diklasifikasikan Baik sehingga hasil ini dapat menjadi acuan bagi perusahaan dalam menentukan keputusan perolehan nilai akhir.

## DAFTAR PUSTAKA

- [1] J.Han and Micheline Kamber, Data Mining: Concepts and Techniques Third Edition, Elsevier, 2012.
- [2] C.C. Wei, Jiing-Yun You, C4.5 Classifier For Solving The Problem Of Water Resources Engineering, Proceeding Of The International Conference On Advanced Science, Engineering And Information Technology, isbn 978-983-42366-4-9, January 2011.
- [3] Dawson, C. Project in Computing and Information Systems. England: Pearson Education. 2009.
- [4] I. H. Witten, Frank Eibe, Mark A. Hall, Data mining: Practical Machine Learning Tools and Techniques 3rd Edition, Elsevier, 2011.
- [5] F.Gorunescu, Data Mining: Concepts, Models and Techniques, Springer, 2011.