

## OPTIMASI KLASIFIKASI PENILAIAN AKREDITASI LEMBAGA KURSUS MENGGUNAKAN METODE K-NN DAN NAIVE BAYES

**MUHAMMAD AMIN,S.Kom, M.Kom**  
([maminbjm58@gmail.com](mailto:maminbjm58@gmail.com))

### ABSTRACT

*Salah satu tugas pokok dan fungsi Badan Akreditasi Nasional Pendidikan Non Formal (BAN-PNF) adalah melaksanakan akreditasi terhadap Lembaga Kursus dan Pelatihan (LKP). Akreditasi adalah kegiatan penilaian kelayakan satuan beserta program PNF berdasarkan atas kriteria yang telah ditetapkan. Untuk menilai kelayakan tersebut disusun instrumen akreditasi yang mengacu pada Standar Nasional Pendidikan (SNP) sebagaimana ditetapkan melalui Peraturan Pemerintah Republik Indonesia Nomor 32 Tahun 2013 tentang Perubahan atas Peraturan Pemerintah Republik Indonesia Nomor 19 Tahun 2005, yang mencakup delapan standar.*

*Algoritma K-Neares Naighbor (k-NN) dan Naive Bayes adalah suatu metode adalah suatu metode yang menggunakan algoritma supervised, dimana pengujian yang baru diklasifikasikan berdasarkan dari katagori K-NN dan Naive Bayes. Penelitian ini dilakukan untuk mangkaji hasil akurasi yang lebih tinggi dan kemudian mengklikasipikasikan kedua algoritma tersebut dalam klasifikasi data hasil Akreditasi yang mempunyai 3 katagori yaitu nilai A,B dan C.*

*Klasifikasi menggunakan dua Algoritma k-NN dan Naive Bayes perhitungan menggunakan Aplikasi Rapid Miner dimana hasil hitungan nilai akurasi yang paling tinggi adalah*

*Kata kunci : Algoritma klasifikasi K- Nearest neighbor (k-NN), Naive Bayes,Akreditasi*

### PENDAHULUAN

Sistem pendidikan nasional sebagai tersebut dalam Undang-Undang Republik Indonesia No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional (selanjutnya : UU RI No.20/ 2003), adalah keseluruhan komponen pendidikan yang saling terkait secara terpadu untuk mencapai tujuan pendidikan nasional dimana jalur pendidikan terdiri atas pendidikan formal, nonformal, dan informal yang dapat saling melengkapi dan memperkaya serta dapat diselenggarakan dengan sistem terbuka melalui tatap

muka dan/atau melalui jarak jauh. Salah satu jalur pendidikan adalah pendidikan

non formal ( ( selanjutnya disebut PNF ) yang dapat dilaksanakan secara terstruktur dan berjenjang. Pendidikan nonformal diselenggarakan bagi warga masyarakat yang memerlukan layanan pendidikan yang berfungsi sebagai pengganti, penambah, dan / atau pelengkap pendidikan formal dalam rangka mendukung pendidikan sepanjang hayat dan berfungsi mengembangkan potensi peserta didik

dengan penekanan pada penguasaan pengetahuan dan keterampilan fungsional serta pengembangan sikap dan kepribadian profesional.

**TUJUAN PENELITIAN**

Berdasarkan latar belakang dan rumusan masalah diatas, maka penelitian ini bertujuan melakukan klasifikasi untuk penilaian Akreditasi dengan menggunakan algoritma k-NN dan algoritma Naive Bayes sehingga didapatkan tingkat akurasi pada Akreditasi Lembaga.

**TARGET LUARAN DAN MANFAATT PENELITIAN**

Manfaat praktis Manfaat hasil penelitian ini adalah agar dapat menghasilkan dengan menggunakan dua algoritma k-NN dan Naive Bayes Logic menghasilkan nilai A, B, dan C.

**PENENTUAN KLASIFIKASI K-NN**

Nobertus Krisandi, *sekumpulan informasi yang dapat digali dan dikembangkan demi kemajuan industri tersebut dengan menggunakan metode Data Mining. Data mining dikelompokkan dalam dua kategori, yakni supervised dan unsupervised. Algoritma k-Nearest Neighbor (k-NN) adalah suatu metode yang menggunakan algoritma supervised, dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada k-NN. Penelitian ini dilakukan untuk mengkaji tentang Algoritma k-NN dan kemudian mengaplikasikan Algoritma k-NN dalam klasifikasi data. Data yang digunakan adalah data hasil.*

**PENERAPAN NAIVE BAYES**

Setelah Dilakukan penilaian dengan Algoritma k-NN mendapatkan Label A, B, C, digabung kan dengan nilai Naive

Bayes Logic digabungkan sehingga mendapat nilai Akhir yang lebih akurat.

**KLASIFIKASI AKREDITASI LEMBAGA KURSUS AKREDITASI**

Adalah kegiatan penilaian kelayakan suatu program dalam satuan pendidikan berdasarkan kriteria yang telah ditetapkan. Akreditasi dilakukan atas dasar kriteria yang bersifat terbuka.

**PENGERTIAN LEMBAGA KURSUS**

Diselenggarakan bagi masyarakat yang memerlukan bekal pengetahuan, keterampilan, kecakapan hidup, dan sikap untuk mengembangkan diri, mengembangkan profesi, bekerja, usaha mandiri, dan/atau melanjutkan pendidikan ke jenjang yang lebih tinggi.

**NILAI AKREDITASI**

Penilaian Akreditasi lembaga :  
Menjumlah Varibel – varibel dan memberikan batasan nilai menjadi skor A.B, dan C

**ALGORITMA K-NN**

Algoritma *K-NN* adalah sebagai berikut: Algoritma k-NN menggunakan klasifikasi ketetangaan sebagai nilai prediksi dari sampel uji yang baru Jarak yang digunakan adalah jarak Euclidean Distance. Jarak Euclidean adalah jarak yang paling umum digunakan pada data numerik Euclidean distance didefinisikan sebagai berikut :

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2}$$

1. Tentukan:
  - Jarak Euclidean =  $d(x_i, x_j)$  ;
  - Record ke i =  $(x_i)$
  - Record ke j =  $(x_j)$

- Data ke r =  $(a_r)$
- $i, j = 1, 2, 3, \dots, n$ ;

**ALGORITMA NAIVE BAYES**

Menurut Larose pendekatan Bayesian digunakan untuk menentukan kemungkinan terhadap asumsi sekitarnya. Dalam statistik Bayesian, Parameter dipertimbangkan terhadap variabel yang acak dan data dipertimbangkan terhadap hasil kemungkinan [ ] *Naive Bayaes* adalah salah satu metode Klasifikasi yang dapat memprediksi probabilitas sebuah *Class*, sehingga dapat menghasilkan keputusan berdasarkan data dengan memberikan akurasi klasifikasi yang kompetitif dan Efisien komputasi. Hal ini menyebabkan *Naive bayes* banyak diterapkan dalam Praktek.

Persamaan dan teori Bayes adalah  

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots (2)$$

- Keterangan  
 X : Data dengan Class yang belum diketahui  
 H: Hipotesis data X data merupakan suatu Class spesifik  
 P(H|X): Probabilitas hipotesis H berdasar kondisi X (Posteriori probality)  
 P(H) : Probabilitas hipotesis H (prior probability)  
 P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H  
 P(X) : Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *bayes* di atas disesuaikan sebagai berikut:

$$P(C | F1 \dots Fn) = \frac{P(C) P(F1 \dots Fn | C)}{P(F1 \dots Fn)} \dots\dots\dots(3)$$

Dimana Variabel *C* merepresentasikan kelas, sementara variabel *F1 ... Fn*

merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas *C* (*Posterior*) adalah peluang munculnya kelas *C* (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik karakteristik sampel pada kelas *C* (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior \times likelihood}{evidence}$$

- Keterangan :  
 P : Peluang  
 Xi : Atribut ke i  
 xi : Nilai atribut ke i  
 Y : Kelas yang dicari  
 yj : Sub kelas Y yang dicari  
 $\mu$  : Mean, menyatakan rata-rata dari seluruh atribut  
 $\sigma$  : Deviasi standar, menyatakan varian dari seluruh atribut.

**METODE PENGOLAHAN**

**DATA AWAL**

Data yang didapatkan dari inforkursus dan diambil data hanya untuk data lembaga kursus kalimantan selatan tahun 2015.

| Validasi  | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Accuracy  | 65,62 | 65,31 | 67,50 | 65,64 | 68,76 | 67,19 | 68,37 | 70,34 | 69,66 |
| Precision | 64,66 | 64,10 | 65,53 | 64,36 | 65,84 | 65,18 | 65,84 | 67,08 | 66,66 |
| Recall    | 91,49 | 93,09 | 94,15 | 93,58 | 97,88 | 95,21 | 96,78 | 98,41 | 98,42 |
| AUC       | 0,723 | 0,755 | 0,774 | 0,790 | 0,832 | 0,805 | 0,825 | 0,850 | 0,839 |

**EKSPRIMEN DAN PENGUJIAN MODEL/METODE**

Metode yang telah dikembangkan dalam penelitian ini akan diterapkan pada data Data Kursus 2015 melalui suatu model simulasi menggunakan RapidMiner. Sebanyak 90 % data akan digunakan untuk membangun struktur keputusan melalui metode k-NN dan Naive Bayes.

**EKSPERIMEN DAN PENGUJIAN MODEL/METODE**

Melakukan perbandingan pengujian dengan tehnik folds cross validasi, K-Nearest Neighbor (KNN). Tahap pertama melakukan pengujian akurasi menggunakan tehnik folds cross validation data yang diuji adalah data original. Pengujian mulai dari 2,3,4,5,6,7,8,9, dan 10 sehingga didapatkan hasil akurasi dan dapat memastikan hasil kategori miskin dan sangat miskin, tetapi hasil akurasinya masih kurang tinggi kemudian tahap kedua dilakukan kembali pengujian dengan menggunakan metode K-Nearest Neighbor (KNN), data yang di uji adalah data original. Pengujian mulai dari 1,2,3,4,5,6,7,8,9 dan 10 namun akurasinya tidak terlalu meningkat tahap ketiga dilakukan kembali pengujian dengan metode K- Nearest Neighbor data yang di uji menggunakan data original. Pengujian mulai dari 5,6,7,8,9,10,11,12,13,14 dan 15.

**PENGUJIAN MEGGUNAKAN FOLDS CROSS VALIDATION**

Hasil validasi percobaan tehnik *folds cross validation* terhadap data Akreditasi teknik folds cross validation dengan pengujian data mulai 2,3,4,5,6,7,8,9 dan 10 kemudian dievaluasi dan dibandingkan dengan beberapa algoritma lain dengan hasil sebagai berikut :

Tabel 4,1,1 Hasil Pengujian folds Cross validation K- nn

Dari hasil percobaan seperti tabel 4.1.1 hasil pengujian sebanyak 9 percobaan dengan hasil pengukuran berupa nilai rata-rata 9 kali pengujian memiliki tingkat akurasi tertinggi sebesar 70.34 %

Table 4.1.2 Hasil Pengujian folds Cross validation Naive Bayes

**Hasil akurasi K-Nearest Neighbor**

| Validasi  | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Accuracy  | 74,06 | 60,62 | 60,94 | 60,63 | 60,62 | 60,63 | 60,60 | 60,63 | 60,62 |
| Precision | 71,97 | 59,94 | 60,16 | 59,97 | 59,95 | 59,95 | 59,96 | 59,99 | 59,96 |
| Recall    | 91,49 | 99,47 | 99,47 | 99,47 | 99,46 | 99,45 | 99,46 | 99,47 | 99,47 |
| AUC       | 0,728 | 0,362 | 0,316 | 0,422 | 0,286 | 0,379 | 0,364 | 0,370 | 0,293 |

**K= 1**

Tahap pertama pengujian dilakukan dengan data original nomer akurasi validasi 1 menggunakan *K- Nearest Neighbor* dengan validasi model klasifikasi dilakukan terhadap data original Ketika di implementasikan menghasilkan data sebagai berikut.

Tabel 0.1 Hasil Akurasi K-Nearest k= 1

| Akurasi: 68.82% +/- 7.11% (mikro: 68.83%) |         |        |         |
|---|---------|--------|---------|
|   | True B  | True C | True A  |
| Pred. B                                   | 76      | 0      | 2       |
| Pred. C                                   | 55      | 188    | 0       |
| Pred. A                                   | 1       | 0      | 2       |
| Class recall                              | 57,58 % | 100 %  | 50,00 % |

Dari hasil pengujian pada Tabel 4.2.10 menggunakan *K-Nearest Neighbor* dengan data original tersebut didapatkan akurasi 68,82 %.

**Hasil akurasi K-Nearset Neighbor**

**K= 2**

Tahap kedua pengujian dilakukan dengan data original nomer akurasi validasi 2 menggunakan *K- Nearest Neighbor* dengan validasi model klasifikasi dilakukan terhadap data original Ketika di implementasikan menghasilkan data sebagai berikut.

**Tabel 0.2 Hasil Akurasi K-Nearest k= 2**

|   |         |        |         |
|---|---------|--------|---------|
| Akurasi: 83.67% +/- 7.33% (mikro: 83.64%) |         |        |         |
|   | True B  | True C | True A  |
| Pred. B                                   | 82      | 0      | 3       |
| Pred. C                                   | 49      | 188    | 0       |
| Pred. A                                   | 1       | 0      | 1       |
| Class recall                              | 57,58 % | 100 %  | 50,00 % |

Dari hasil pengujian pada Tabel 4.2.11 menggunakan *K-Nearest Neighbor* dengan data original tersebut didapatkan akurasi 83,57 %.

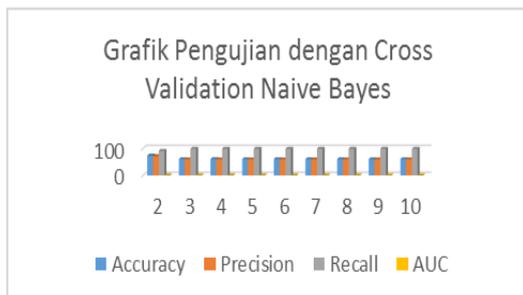
**Hasil akurasi K-Nearset Neighbor**

**K= 3**

Tahap ketiga pengujian dilakukan dengan data original nomer akurasi validasi 3 menggunakan *K- Nearest Neighbor* dengan validasi model klasifikasi dilakukan terhadap data original Ketika di implementasikan menghasilkan data sebagai berikut.

**Tabel 0.3 Hasil Akurasi K-Nearest k= 3**

|   |
|---|
| Akurasi: 76.25% +/- 6.58% (mikro: 76.23%) |
|---|



|              |         |        |         |
|--------------|---------|--------|---------|
|              | True B  | True C | True A  |
| Pred. B      | 56      | 0      | 1       |
| Pred. C      | 75      | 188    | 0       |
| Pred. A      | 1       | 0      | 3       |
| Class recall | 42,42 % | 100 %  | 75,00 % |

Dari hasil pengujian pada Tabel 4.2.12 menggunakan *K-Nearest Neighbor* dengan data original tersebut didapatkan akurasi 76.25%.

**Hasil akurasi K-Nearset Neighbor**

**K = 4**

Tahap keempat pengujian dilakukan dengan data original nomer akurasi validasi 4 menggunakan *K- Nearest Neighbor* dengan validasi model klasifikasi dilakukan terhadap data original Ketika di implementasikan menghasilkan data sebagai berikut.

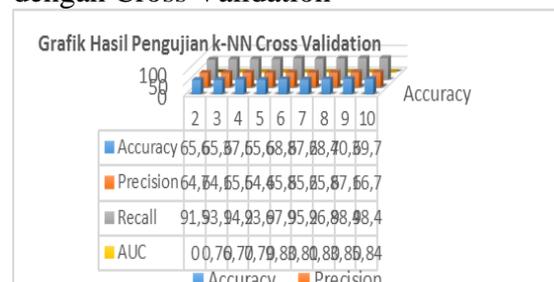
**Tabel 0.4 Hasil Akurasi K-Nearest k= 4**

|   |         |        |        |
|---|---------|--------|--------|
| Akurasi: 78.42% +/- 4.84% (mikro: 78.40%) |         |        |        |
|   | True B  | True C | True A |
| Pred. B                                   | 66      | 0      | 4      |
| Pred. C                                   | 66      | 188    | 0      |
| Pred. A                                   | 0       | 0      | 0      |
| Class recall                              | 50,00 % | 100 %  | 0,00 % |

Dari hasil pengujian pada Tabel 4.2.12 menggunakan *K-Nearest Neighbor* dengan data original tersebut didapatkan akurasi 78,42%.

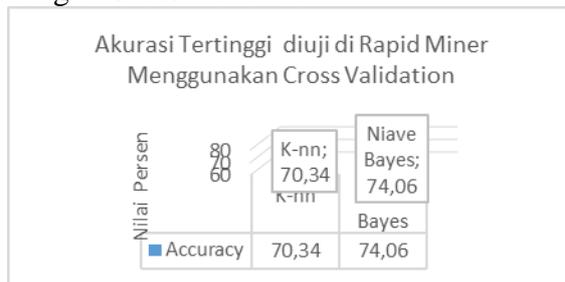
**IMPLEMENTASI GRAFIK**

**Grafik 4.2.1 Hasil Pengujian k-NN dengan Cross Validation**



**Grafik 4.2.2 Hasil Pengujian Naive Bayes dengan Cross Validation**

Grafik 4.2.3 Hasil Pengujian Tertinggi dengan Cross Validation



Dari Hasil Grapik dapat disimpulkan bahwa Naive Bayes Akurasi lebih tinggi dalam akurasi keseluruhan dari beberapa kali pengujian tetapi dalam akurasi Naive Bayes hasil satu dan yang lain jarak cukup jauh dibandingkan dengan pengujian dengan K-NN yang cenderung stabil.

### IMPLEMENTASI PENELITIAN

Data hasil perbandingan tersebut disimpulkan Metode *K-Nearest Neighbor* memprediksi lebih akurat dari pada *K-Nearest Neighbor* tanpa seleksi atribut dan *Folds Cross Validation*, Atribut yang terpilih adalah jumlah keluarga, jumlah individu status penguasaan bangunan, tempat tinggal, jenis dinding terluas, kualitas dinding dan jenis lantai

perbandingan pada tabel 4.2.31 tersebut disimpulkan bahwa *K-Nearest Neighbor* (K-NN) berbasis  $K = 2$  status akreditasi lebih akurat dan juga melihat hasil.

- 1 Nilai C 100 %
- 2 Nilai B 57,58%
- 3 Nilai A 0 %

Dengan demikian, adanya penerapan *K-Nearest Neighbor* (K-NN) mampu memberikan solusi dan dapat membantu pihak terkait, yang dalam hal ini untuk mengidentifikasi status kesejahteraan rumah tangga miskin sesuai dengan kategori akreditasi A, B dan C.

### PENUTUP

#### KESIMPULAN

Berdasarkan hasil klasifikasi dan pembahasan, maka diperoleh kesimpulan sebagai berikut.

1. Penerapan seleksi atribut dapat meningkatkan akurasi *K-Nearest Neighbor* menjadi lebih baik meskipun kenaikan yang dihasilkan tidak terlalu besar namun secara umum hasil penerapan *K-Nearest Neighbor* dengan seleksi atribut menggunakan  $K$  lebih baik dari pada *K-Nearest Neighbor* tanpa menggunakan seleksi atribut.
2. Secara umum *K-Nearest Neighbor* meningkat dari pada tanpa menggunakan atribut namun menerapkan tanpa seleksi atribut masih lebih baik dibandingkan dengan pengujian seperti *folds Cross validation*. Berdasarkan keunggulan tingkat akurasi sebagai tingkat kedekatan antara nilai klasifikasi dengan nilai aktual.

### DAFTAR PUSTAKA

- [1] BADAN AKREDITASI NASIONAL PENDIDIKAN NON FORMAL “Instrumen Akreditasi Kursus dan Pelatihan “ Jakarta 2014
- [2] Nobertus Krisandi, "*K-Nearest Neighbor Dalam Klasifikasi Data* ," Pontianak : Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster) Volume 02, No.1(2013), hal. 33-38..
- [3] T.Sutojo, S.Si, "*Kecerdasan Buatan*" Semarang Andi Yogyakarta 2010
- [4] Eko Prasetyo “Data Mining Mengolah Data Menjadi Informasi “ Gresik Fuzzy 2014

- [5] Adhitya Yoga Yudianto, “*Optimalisasi Lampu Lalu Lintas dengan Fuzzy Logic*”; Jakarta, Indonesia, 2013.
- [6] Ali Mohammad Ahmadvand, A *hybrid dataMining For effective citizen relationship Management : a case study on Tehran municipality* “. Tehran, 2010.
- [7] Nazwar Fahmi, *Analisis Dan Komparasi Algoritma Klasifikasi Data Mining Untuk Penentuan Resiko Kredit*. Semarang, Indonesia, 2012.
- [8] Krisandi, N., Prihandono, B., & Bayes, N. (2013). Algoritma k-nearest neighbor dalam klasifikasi data hasil produksi kelapa sawit pada pt . *Minamas*, 02(1), 33–38.
- [9] Kepemilikan, K., & Bemotor, K. (2013). Penerapan algoritma k-nearest neighbor untuk penentuan resiko kredit kepemilikan kendaraan bemotor, *I*(1), 65–76.
- [10] Subbalakshmi, G., et.al., 2011. Decision Support in Heart Disease Prediction System using *Naive Bayes*. *Indian Journal of Computer Science and Engineering*, 2 (2), p. 170.
- [11] Han, J., Kamber, M., 2006. *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann Publishers..
- [12] Hidayah, N., 2013. *Klasifikasi Penjurusan Program Studi Sekolah Menengah Atas dengan Algoritma Naive Bayes Classifier pada SMAN 1 Subah*. UDINUS Semarang.
- [13] Indu Indah Purnomo, 2015, *Klasifikasi Status Rumah Tangga Miskin dan Paling Miskin Menggunakan Algoritma K-Nearest Neighbor dan Seleksi Fituri Berbasis Chi Squared*, UDINUS Semarang.
- [14] Bramer, M. 2007. *Principles Of Data Mining*. London: Springer-Verlag London Limited.
- [15] Gorunescu, F. 2010. *Data Mining Concepts, Models, and Techniques*. Berlin: Springer.
- [16] Larose, *Data Mining Methods And Models*, Canada: John Wiley & Sons, Inc, 2006.
- [17] M. Dunham, *Data Mining Introductory and Advanced Topics*, New Jersey: Prentice Hall, 2003.
- [18] V. Moertini, *Data Mining Sebagai Solusi Bisnis*, Integral Vol. 7 No. 1, 2002.
- [19] R. Bellazzi dan B. Zupanb, *Predictive Data Mining In Clinical Medicine: Current Issues And Guidelines*, International Journal Of Medical Informatics, 2008.
- [20] W. Ian H. dan E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2 penyunt., San Francisco: Morgan Kaufmann Publishers is an imprint of Elsevier, 2005.
- [21] F. Gorunescu, *Data Mining: Concept, Models and Techniques*, Romania: Springer, 2010.
- [22] Undang-Undang Dasar Negara Republik Indonesia 1945 [Amandemen], surabaya: pustaka agung harapan, 2002.
- [23] J. Han dan M. Kamber, *Data Mining : Concepts and Techniques*, Second penyunt., M. R. Jim Gray, Penyunt., San Francisco: Morgan Kaufmann Publishers, 2007.
- [25] Kemensos, *Petunjuk Pelaksanaan Kelompok Usaha Bersama (KUBE)*, jakarta: direktorat penanggulangan kemiskinan pedesaan, 2014.