PENGENALAN KATA AKSARA JAWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Zaky Mukhoyyar
Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang
Jl. Nakula I No. 5-11 Semarang Telp (024) 3517261
e-mail: 111201105854@mhs.dinus.ac.id, zackeinherjar@gmail.com

Abstrak

Seiring cepatnya perkembangan teknologi informasi dan komunikasi, lambat laun kebudayaan pun akan semakin bergeser. Banyak sekali tulisan-tulisan peninggalan nenek moyang yang hampir terlupakan oleh generasi masa kini. Untuk mengantisipasi hal tersebut, maka muncullah teknik pengenalan huruf atau tulisan agar peninggalan nenek moyang tidak habis termakan zaman. Pemanfaatan pengolahan citra digital dapat digunakan untuk mengenali tulisan-tulisan peninggalan nenek moyang. Maka dari itu, penulis mendapatkan ide untuk melakukan pelestarian sekaligus melakukan penelitian terhadap metode dalam pengolahan citra digital agar dapat digunakan sebagai media pengenalan tulisan-tulisan peninggalan nenek moyang. Ide tersebut menggunakan metode K-Nearest Neighbor sebagai metode klasifikasinya.

Kata kunci : Pengolahan Citra Digital, Optical Character Recognition, K-Nearest Neighbor, Manhattan Distance.

Abstract

As the rapid development of information and communication technology, gradually culture will increasingly shift. Lots of writings ancestral almost forgotten by the present generation. To anticipate this, then comes the letter or word recognition techniques in order for ancestors is not eroded by time. Utilization of digital image processing can be used to identify the writings of the ancestor. Therefore, the author got the idea to do the preservation and conduct research on digital image processing methods in order to be used as a medium for the introduction of the writings of the ancestor. The idea of using K-Nearest Neighbor method as a method of classification.

Keywords : Digital Image Processing, Optical Character Recognition, K-Nearest Neighbor, Manhattan Distance.

1. PENDAHULUAN

Seiring perkembangan zaman, seiring pula berkembangnya teknologi. Tak terkecuali pada teknologi pengolahan citra. Perkembangan dibidang pengolahan citra sering digunakan untuk riset dan pengembangan aplikasi dan teknologi. Banyak sekali metode dan algoritma yang diciptakan dan digunakan untuk mempermudah kinerja citra pada suatu obyek maupun media. OCR (Optical Character Recognition) merupakan salah satu area studi dalam bidang pengenalan pola yang menarik untuk dieksplorasi[6]. Sistem pengenalan huruf atau sering disebut OCR merupakan solusi yang

efektif untuk proses konversi dari dokumen cetak ke dalam bentuk dokumen digital[6].

Kemajuan pengolahan citra juga dapat digunakan dalam melestarikan ilmu dan budaya. Salah satu penerapannya dapat digunakan dalam melestarikan aksara jawa yang hingga kini hampir punah karena perkembangan zaman yang semakin modern. Aksaara jawa merupakan bagian dari Bahasa jawa yang melekat pada kebudayaan masyarakat jawa, untuk saat ini hampir mengalami kepunahan. Usaha pemerintah dalam melestarikan aksara jawapun hanya mencakup tentang pendidikan di sekolah dasar dan menengah. Minimnya penggunaan aksara jawa dikalangan masyarakat jawa, membuat aksara jawa jarang digunakan untuk kehidupan sehari-hari. Padahal pengguna bahasa jawa semakin bertambah banyak dan sedikit dari kalangan remaja yang mengenal aksara jawa dengan jelas. Lokasi bersejarah di jawa beberapa diantaranya masih menggunakan aksara jawa untuk menyampaikan pesan sejarah yang terkandung dalam lokasi tersebut. Banyak wisatawan lokal yang diantaranya orang jawa susah dalam membaca aksara jawa yang tertera pada lokasi bersejarah yang dikunjungi sehingga mengakibatkan kurangnya pemahaman wisatawan lokal terhadap pesan tersebut.

Pengenalan Aksara Jawa Tulisan Tangan telah dilakukan (Rizkina Muhammad Syam, 2013) dengan menggunakan Ekstraksi Fitur Zoning dan Klasifikasi K-Nearest Neighbour guna mengatasi masalah pada pengenalan pola (Suryo Hartanto, Aris Sugiharto, Sukmawati Nur Endah, Optical Character Recognition Menggunakan Algoritma Template Matching Correlation, 2012) (Oskar Ika Adi Nugroho, Pranowo, Yudi Dwiyandianta, Segmentasi Citra Aksara Jawa Menggunakan Algoritma Particle Swarm Optimization, 2014) (Ni Made Ari Pratiwi, Widi Hapsari, Theresia Herlina R, Pengenalan Aksara Bali Dengan Pendekatan Metode Direction Feature dan Area Binary Object Feature, 2013) (I Kadek Agus Dwi Putra dan Luh Putu Ayu Prapitasari, Segmentasi Karakter Pada Skrip Bahasa Bali Menggunakan Metode Canny Edge Detection, 2011). K-Nearest Neighbor dapat melakukan klasifikasi dengan hasil yang cukup memuaskan.

2. METODE PENELITIAN

Penelitian merupakan penelitian eksperimental, yaitu penelitian yang pengumpulan datanya dilakukan melalui pencatatan langsung dari hasil percobaan yang dilakukan.

2.1 Preprosessing

Preprocessing adalah tahap awal pengolahan citra asli sebelum citra tersebut diolah. Dalam tahap ini, umumnya preprocessing mengubah citra masukan menjadi citra abu-abu (grayscale) dan untuk memperbaiki kualitas citra. Selain mengubah citra masukan, preprocessing juga bertujuan untuk memanipulasi citra agar sesuai dengan hasil yang diinginkan. Metode dalam melakukan preprocessing menggunakan metode Thresholding. Thresholding adalah citra yang memiliki dua nilai tingkat keabuan yaitu hitam dan putih. Thresholding memiliki proses pengembangan hasil dari grayscaling untuk menghasilkan citra biner.

$$g(x,y) = \begin{cases} 1 \; if \; f(x,y) \geq T \\ 0 \; if \; f(x,y) < T \end{cases}$$

g(x,y) merupakan citra biner dari citra grayscale f(x,y) yang diproses melalui thresholding dengan nilai T. Hasil

thresholding bergantung pada nilai T yang memegang peranan penting untuk menentukan kualitas hasil citra biner.

2.2 Segmentasi

Segmentasi digunakan untuk memotong suatu teks pada gambar yang diproses menjadi beberapa karakter. Selain itu, segmentasi juga bertujuan untuk menyederhanakan, mengelompokkan atau mengubah penyajian dari suatu gambar menjadi sesuatu yang lebih mudah untuk menganalisanya.

2.3 Algoritma K-Nearest Neighbor

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-Nearest Neighbor. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan ciri dan data latih.

Algoritma K-Nearest Neighbor bekerja berdasarkan pada jarak terpendek dari data uji ke data latih untuk menentukan nilai K-Nearest Neighbor nya. Setelah menentukan nilai kemiripannya, maka nilai tersebut dikelompokkan pada kelas-kelas tertentu. Salah satu cara untuk menentukan jarak terdekat atau jarak terjauhnya dengan menggunakan metode pendekatan Manhattan Distance.

$$d(x,y) = \sum_{i=1}^{m} |x_i - y_i|$$

xi = nilai X pada data latih.
yi = nilai Y pada data uji.
m = batas jumlah banyaknya data

Jika angka hasil dari rumus tersebut besar, maka tingkat kemiripan antara kedua objek akan semakin kecil dan sebaliknya, jika angka hasil rumus tersebut kecil, maka tingkat kemiripan antara kedua objek akan semakin besar. Objek yang dimaksud adalah data latih dan data uji yang akan dihitung tingkat kemiripannya.

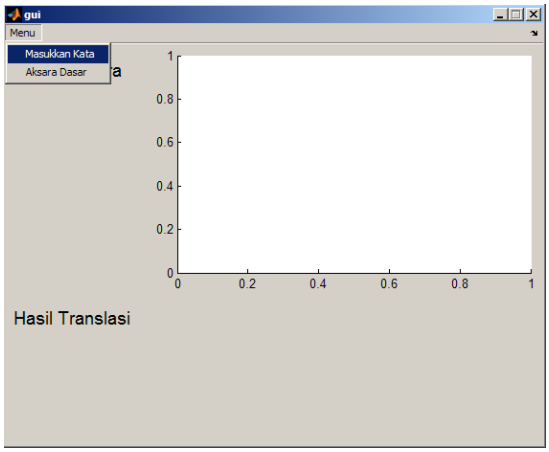
Langkah-langkah untuk menghitung algoritma K-Nearest Neighbor :

1. Menentukan nilai *k*.
2. Menghitung jarak masing-masing objek terhadap data latih yang diberikan dengan menggunakan metode *Manhattan Distance*.
3. Urutkan hasil perhitungan nilai kemiripan dari yang terkecil hingga yang terbesar.

Tentukan kelas dari data uji berdasarkan kelas yang paling kecil jaraknya.

3. HASIL DAN PEMBAHASAN

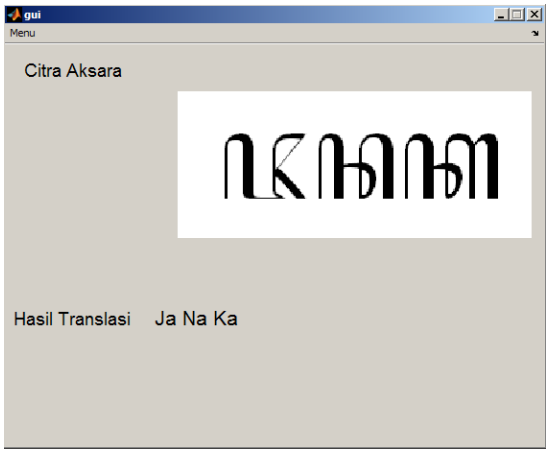
Sistem ini dibuat menggunakan MATLAB 2013b. Form terdapat menu masukkan kata dimana user dapat meng-input-kan gambar yang didapat dari sebuah huruf dalam microsoft word dengan font aksara jawa dan ukuran font 48. Hasil akan ditampilkan tepat di sebelah kanan tulisan Hasil Translasi.



Gambar 1. Tampilan Menu

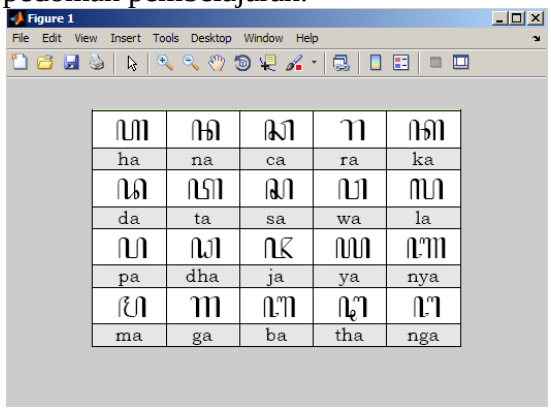
Berikut adalah tampilan saat user meng-input-kan contoh citra aksara jawa

dan hasilnya keluar tepat di kanan tulisan hasil translasi.



Gambar 2. Tampilan Contoh

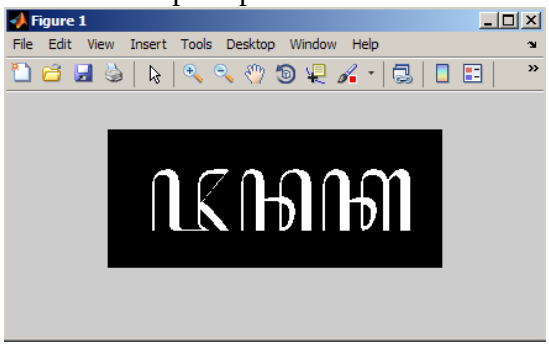
Selain user dapat meng-input-kan gambar untuk diubah menjadi tulisan alfabet yang dapat dibaca, sistem juga menyediakan daftar aksara dasar untuk pedoman pembelajaran.



Gambar 3. Tampilan Aksara Dasar

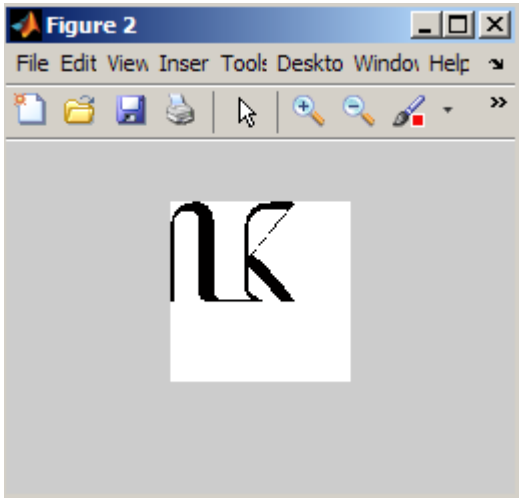
Pada langkah awal proses, proses yang terjadi adalah thresholding. Tahapan dalam preprocessing digunakan karena citra awal yang berupa citra grayscale dan citra hasil yang berupa citra biner tidak berbeda jauh dan mengurangi kebutuhan memori sehingga proses penghitungan mampu berjalan lebih cepat dan ringan.

Berikut ini adalah hasil konversi citra grayscale menjadi citra biner yang dinegasikan dengan ukuran matriks 112x271 sama seperti pada Gambar 2.

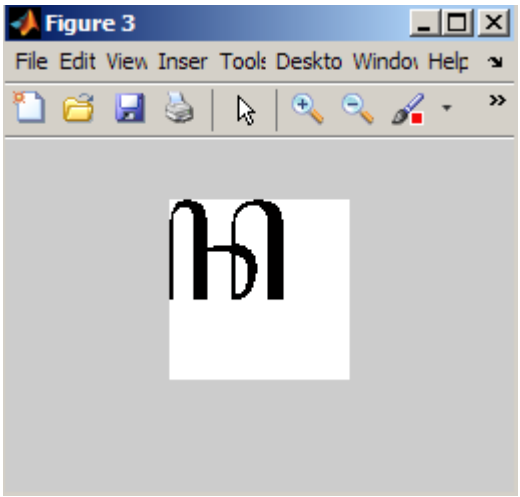


Gambar 4. Hasil Thresholding

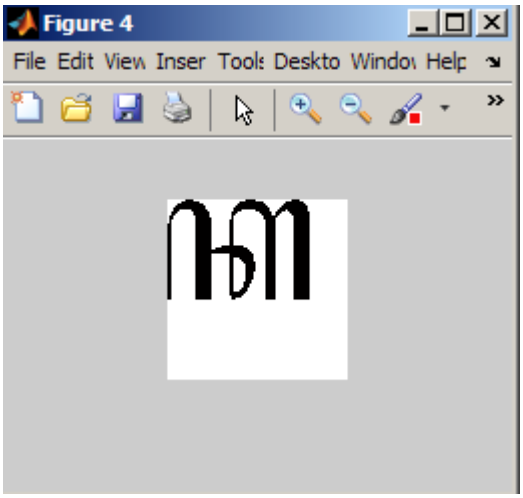
Tahap selanjutnya adalah segmentasi.Pada proses ini, citra yang terdapat 2 atau lebih karakter aksara jawa akan dibagi menjadi 1 citra tiap 1 karakter dengan menggunakan metode histogram agar dapat dilakukan proses klasifikasi. Pada segmentasi citra ini, matriks citra yang sudah melalui tahap thresholding dijumlah secara vertikal. Berikut adalah contoh segmentasi dari Gambar 2.



Gambar 5. Segmentasi karakter pertama



Gambar 6. Segmentasi karakter kedua



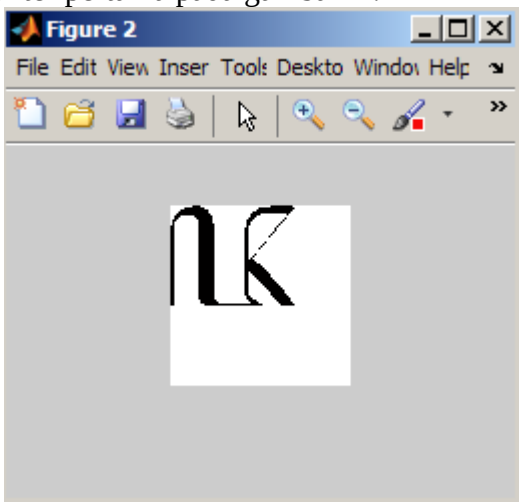
Gambar 7. Segmentasi karakter ketiga

Satu citra huruf aksara jawa pada awalnya yang terdapat 3 karakter atau huruf aksara jawa, ketiga karakter tersebut dilakukan segmentasi dan menghasilkan satu citra tiap satu karakter atau huruf yang selanjutnya akan dilakukan klasifikasi.

Tahap terakhir adalah klasifikasi, klasifikasi menggunakan algoritma K-Nearest Neighbor dan dengan menggunakan pendekatan Manhattan Distance. Langkahnya adalah sebagai berikut :

1. Huruf yang akan diuji dan huruf training akan dilakukan reduksi dimensi pada matriksnya, yang awalnya 90x90 menjadi matriks 1x8100.
2. Matriks huruf uji dan matriks huruf training akan diuji dengan menggunakan algoritma K-Nearest Neighbor dengan nilai k=20 lalu dimasukkan pada rumus Manhattan Distance untuk mendapatkan matriks pendekatan.
3. Matriks pendekatan tersebut dijumlahkan secara horizontal untuk mengetahui total kedekatan yang didapatkan dan dibandingkan dengan kedekatan huruf lainnya yang total berjumlah 20 huruf.

Berikut adalah proses klasifikasi karakter pertama pada gambar 2 :

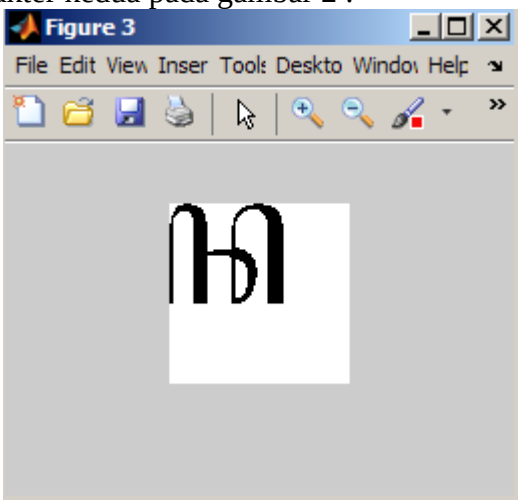


Gambar 8. Klasifikasi karakter pertama

Huruf	Nilai Kedekatan
HA	1093
NA	1020
CA	957

RA	771
KA	1092
DA	777
TA	1165
SA	1125
WA	775
LA	1327
PA	783
DHA	1074
JA	**2**
YA	1353
NYA	1672
MA	847
GA	1509
BA	1010
THA	1110
NGA	1035

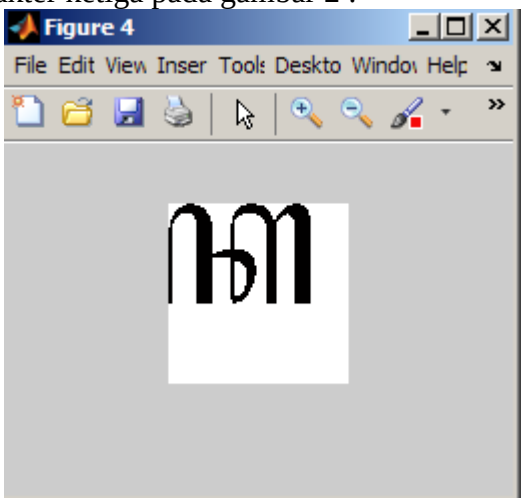
Berikut adalah proses klasifikasi karakter kedua pada gambar 2 :



Gambar 9. Klasifikasi karakter kedua

Huruf	Nilai Kedekatan
HA	1421
NA	**250**
CA	849
RA	989
KA	902
DA	987
TA	1141
SA	1119
WA	571
LA	1259
PA	529
DHA	994
JA	978
YA	1349
NYA	1756
MA	769
GA	809
BA	1122
THA	862
NGA	611

Berikut adalah proses klasifikasi karakter ketiga pada gambar 2 :



Huruf	Nilai Kedekatan
HA	1453
NA	990
CA	1465
RA	1363
KA	**0**
DA	1493
TA	1177
SA	1545
WA	1165
LA	1937
PA	1235
DHA	1734
JA	1092
YA	1145
NYA	2092
MA	1241
GA	1591
BA	1634
THA	1612
NGA	1315

Data dari tabel diatas menunjukkan nilai kedekatan huruf yang diuji terhadap nilai huruf training. Semakin kecil nilainya, tingkat kedekatan antara huruf yang diuji dan huruf training semakin dekat.

Huruf pertama dengan nilai kedekatan 2 terletak pada huruf ke-13 yaitu huruf JA, maka huruf tersebut dikenali sebagai huruf JA.

Huruf kedua dengan nilai kedekatan 250 terletak pada huruf ke-2 yaitu huruf NA, maka huruf tersebut dikenali sebagai huruf NA.

Huruf ketiga dengan nilai kedekatan 0 terletak pada huruf ke-5 yaitu huruf KA, maka huruf tersebut dikenali sebagai huruf KA.

Huruf ketiga pendekatannya sempurna karena menunjukkan angka 0 sehingga mutlak tak ada perbedaan diantara huruf uji dan huruf training, sedangkan huruf kedua yang nilai kedekatannya paling banyak diantara kedua huruf lainnya menunjukkan bahwa ada cukup banyak perbedaan dan masih bisa dikenali dengan tepat.

4. KESIMPULAN DAN SARAN

4.1 KESIMPULAN

Salah satu penerapan algoritma K-Nearest Neighbor yaitu Manhattan Distance mampu melakukan klasifikasi terhadap citra aksara jawa dengan tingkat keberhasilan pengenalan 100%. Implementasi segmentasi berbasis histogram memiliki tingkat keberhasilan 90% berdasarkan data yang digunakan. Sistem translasi aksara jawa menjadi alfabet telah berhasil diimplementasikan

dan dapat digunakan oleh pemula yang ingin belajar untuk mengenal aksara jawa.

4.2 SARAN

Penggunaan data training yang berbeda agar tidak terjadi kesalahan pada segmentasi maupun klasifikasi. Meningkatkan tingkat keberhasilan pada segmentasi dengan menggunakan metode lain.

DAFTAR PUSTAKA

[1] Bahri, Raden Sofian., Maliki, Irfan. PERBANDINGAN ALGORITMA TEMPLATE MATCHING DAN FEATURE EXTRACTION PADA OPTICAL CHARACTER RECOGNITION, Jurnal Komputer dan Informatika, Edisi. I Volume. 1, Maret 2012

[2] Amit, Yali. (2002). 2D object recognition and detection: models, algorithms and networks. Cambridge: MIT Press.

[3] Nugroho, Oskar Ika Adi., Pranowo, Dwiyandianta. Yudi. (2014). SEGMENTASI CITRA AKSARA JAWA MENGGUNAKAN ALGORITMA PARTICLE SWARM OPTIMIZATION, ISSN: 2089-9813.

[4] Pratiwi, Ni Made Ari., Hapsari, Widi., R., Theresia Herlina., PENGENALAN AKSARA BALI DENGAN PENDEKATAN METODE DIRECTION FEATURE DAN AREA BINARY OBJECT FEATURE, Informatika Vol.9, No. 1,April 2003

[5] Putra, I Kadek Agus Dwi., Prapitasari. Luh Putu Ayu,. (2011). SEGMENTASI KARAKTER PADA SKRIP BAHASA BALI MENGGUNAKAN METODE CANNY EDGE DETECTION, KNS&I11-063

[6] Sutojo, T, Mulyanto, E, Suhartono, V, Nurhayati, O, D, Wijanarto (2009). “Teori Pengolahan Citra Digital”. Yogyakarta: Andi.

[7] Munir, Rinaldi (2004). "Pengolahan Citra Digital Dengan Pendekatan Algoritmik". Bandung: Informatika.

[8] Sianipar. R.H.,S.T,M.T,M.Eng,Ph.D (2013). "Pemrograman MATLAB dalam contoh dan penerapan". Bandung: Informatika.

[9] Putra, D. (2010). "Pengolahan Citra Digital". Yogyakarta : Andi.

[10] Hartanto, Suryo., Sugiharto, Aris., Endah, Sukmawati Nur(2012). "Optical Character Recognition Menggunakan Algoritma Template Matching Correlation"

[11] Prasetyo, Eko(2012). "Data Mining Konsep dan Aplikasi menggunakan MATLAB". Yogyakarta: Andi