

Kombinasi Algoritma K-NN dan Manhattan Distance untuk Menentukan Pemenang Lelang

Khoiriya Latifah

Program Studi Informatika, Fakultas TEKNIK, Universitas PGRI Semarang

Gedung B Lantai 3, Kampus 1 Jl. Sidodadi Timur 24, Semarang

E-mail : khoiriyalatifah@upgrismg.ac.id

***Abstract**— Determination of the auction is non-linear problem (the problem is heavily influenced by natural and environmental factors) and the managerial decisions (a qualitative decision) so it takes knowledge to make an assessment of the winning bidder.*

In the construction or infrastructure industry there is a set of information that can be explored and developed for the advancement of the industry by using Data Mining. Data mining is grouped into two categories, namely supervised and unsupervised. Algorithm k-Nearest Neighbor (k-NN) is a method that uses a supervised algorithm, where the results of the new test samples are classified based on the majority of categories on the k-NN. This study was conducted to assess of the k-NN algorithm and then apply the k-NN algorithm in data classification.

Variable judging from the results obtained questionnaires to the winning bidder and the data of copyrighted works. From the experiment results obtained accuracy and precision recal F. Measure all predictions above 0.8 means that the system worked well to predict the winner of the auction with method of prequalification

Keywords: *Data Mining, K-NN Algorithm, Manhattan Distance*

Abstrak—Penentuan pemenang lelang adalah masalah non linier (yang banyak dipengaruhi oleh faktor alam dan lingkungan) dan merupakan keputusan manajerial (alamiah merupakan keputusan kualitatif)sehingga memerlukan pengetahuan untuk melakukan penilaian terhadap pemenang lelang. Di dalam Industri konstruksi atau infrastruktur terdapat 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.

Variabel penilaian diperoleh dari hasil kuisisioner kepada para pemenang lelang dan data dari cipta karya. Dari hasil experiment diperoleh precision recal accuracy dan F.Measure semua diatas 0,8 artinya system prediksi berhasil dengan baik untuk memprediksi pemenang lelang dengan metode prakualifikasi

Kata Kunci - *Data Mining , Algoritma K-NN, Manhattan Distance*

PENDAHULUAN

Keberhasilan proyek pembangunan sarana dan prasarana *public* yang dilakukan oleh pemerintah atau swasta sangat ditentukan oleh peran dari para pelaku konstruksi yang terlibat didalamnya [1].

Pemilihan penyedia jasa yang berkualitas dan memenuhi syarat memberikan kepercayaan pada klien karena penyedia jasa yang terpilih dapat mencapai tujuan proyek, menguntungkan bagi negara dan tanpa ada intervensi kepada panitia pengadaan atau pelelangan dan pemimpin proyek sehingga tidak terjadi dampak negative pada hasil pelelangan.

Ketepatan pemilihan pelaku konstruksi sangat berpengaruh pada hasil project yang dikerjakan. karena pemenang lelang yang berkualitas yang dapat melaksanakan pekerjaan dengan tepat mutu, tepat waktu dan tepat biaya sehingga tujuan proyek bisa tercapai [2]. Tetapi pemilihan penyedia jasa masih menemui banyak kendala dalam penerapannya. Hal ini umumnya dikarenakan kriteria evaluasi yang digunakan menimbulkan subyektifitas panitia pengadaan. Penilaiannya lebih bersifat kualitatif dan subjektif, sehingga *Subyektif Judgment* yang didasarkan pengalaman sangatlah menentukan [3]. Parameter-parameter yang bersifat subjectif ini biasanya noninstitutional sehingga sulit untuk dinilai. Parameter subjektif memerlukan perhatian khusus sehingga evaluasi pemilihan kontraktor merupakan masalah yang menarik.

Dari permasalahan tersebut diatas maka penelitian ini membahas analisis kelayakan pemilihan pemenang lelang dengan metode klasifikasi *data mining*. Upaya pengolahan data hasil project yang dihasilkan oleh pemenang lelang terdahulu di kalsifikasikan dengan menggunakan salah satu metode data mining, yakni dengan

menggunakan metode *k-NN*. *International Conference on Data Mining (ICDM)* mencantumkan Algoritma *k-NN* sebagai salah satu algoritma terbaik dalam *Top10 algorithms in data mining*. Algoritma-algoritma tersebut antara lain *C4.5*, *kMeans*, *SVM*, *Apriori*, *EM*, *PageRank*, *AdaBoost*, *k-NN*, *Naïve Bayes*, and *CART* [4]. Penelitian ini mengkaji tentang Algoritma *k-NN* dan kemudian mengaplikasikan Algoritma *k-NN* dalam klasifikasi data. Data yang digunakan adalah data sekunder berupa data pemenang lelang dan hasil project yang telah diselesaikan oleh pemenang lelang pada tahun 2011-2013 pada salah satu instansi pemilik project.

Algoritma *nearest neighbor* merupakan pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Algoritma *Nearest Neighbor dan Manhattan Distance* sebagai alat untuk pengambilan keputusan pada penentuan pemenang lelang dengan menggunakan variabel/parameter-parameter yang mempengaruhi pemilihan penyedia jasa. Penggunaan *Manhattan Distance* untuk pengukuran jarak memperoleh akurasi yang tinggi dibandingkan dengan *Euclidean Distance* sehingga dalam penelitian ini menggunakan klasifikasi metode *k-Nearest Neighbor* dengan menggunakan *Manhattan Distance* untuk menghitung jarak terdekat dalam metode klasifikasi.[5].

TINJAUAN PUSTAKA

Data Mining

Data Mining didefinisikan sebagai proses penemuan pola dalam data [6]. Berdasarkan tugasnya, *data mining* dikelompokkan menjadi deskripsi, estimasi, prediksi, klasifikasi, *clustering* dan asosiasi

[7]. Proses dalam tahap *data mining* terdiri dari tiga langkah Utama [8], yaitu :

a. *Data Preparation*

Pada langkah ini, data dipilih, dibersihkan, dan dilakukan preprocessed mengikuti pedoman dan knowledge dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

b. Algoritma *data mining*

Penggunaan algoritma data mining dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

c. Fase analisa data

Keluaran dari data mining dievaluasi untuk melihat apakah knowledge domain ditemukan dalam bentuk rule yang telah diekstrak dari jaringan.

Klasifikasi

Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Model itu sendiri bisa berupa aturan “jika-maka”, berupa *decision tree*, *formula matematis* atau *neural network*. Metode-metode klasifikasi antara lain C4.5, *RainForest*, *Naïve Bayesian*, *neural network*, *genetic algorithm*, *fuzzy*, *case-based reasoning*, dan *k-Nearest Neighbor* [9].

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi

[10]. Proses klasifikasi didasarkan pada empat komponen :

a. Kelas

Variabel dependen yang berupa kategorikal yang merepresentasikan ‘label’ yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa.

b. *Predictor*

Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.

c. *Training dataset*

Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.

d. *Testing dataset*

Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Nearest Neighbor Retrieval

Dalam penelitian ini penulis menggunakan Retrieval K-Nearest Neighbor (KNN). *K-Nearest Neighbor* (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah *k* obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari *k* obyek.. algoritma KNN menggunakan klasifikasi

ketetapan sebagai nilai prediksi dari *query instance* yang baru.

Nearest Neighbor menghitung tingkat kemiripan (jarak) suatu kasus terhadap kasus lain berdasarkan beberapa atribut yang didefinisikan berdasar pembobotan tertentu dan kemudian tingkat kemiripan (jarak) dari keseluruhan atribut akan dijumlahkan. *Nearest Neighbor* didefinisikan berdasar persamaan sebagai berikut :

$$\text{Similarity}(T, S) = \sum_{i=1}^n f(T_i, S_i) W_i$$

- T : Kasus target / baru
- S : Kasus sumber / lama / pembanding
- n : Jumlah atribut dalam setiap kasus
- i : Atribut individu dari 1 sampai n
- f : Fungsi kemiripan untuk atribut I dalam kasus T dan S
- w : Bobot atribut i

Algoritma metode KNN bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada ruang ini ditandai kelas *c* jika kelas *c* merupakan klasifikasi yang paling banyak ditemui pada *k* buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *jarak*.

Manhattan Distance

Sedangkan untuk menghitung jarak ada beberapa metode yang terkenal adalah Manhattan Distance [11], Euclidean Distance, Mahalanobis Distance, geometric similarity measures [13], and probabilistic similarity measures [14]. Penulis

menggunakan Manhattan atau City Distance digunakan untuk mengambil kasus yang cocok dari basis kasus dengan menghitung jumlah bobot absolute dari perbedaan antara kasus yang sekarang dan kasus yang laen dalam basis kasus.

Untuk menghitung bobot digunakan persamaan berikut :

$$d_{ij} = \sum W_k |x_{ik} - c_{jk}|$$

Dimana diketahui d_{ij} adalah jarak antara kasus antara i_{th} dan j_{th} dengan semua parameternya. W merepresentasikan jumlah dari bobot. X adalah kasus yang baru dikurangi dengan C yaitu history (kasus yang ada dalam Casse Base). Manhattan Distance adalah pengukuran similarity / kemiripan yang paling cocok untuk approval project yang merepresentasikan kasus yang relevant dengan angka yang natural atau dengan data yang bersifat kuantitative. [14]

Dengan menganalisa data hasil perhitungan dengan K-NN, panitia lelang akan memiliki acuan untuk pengambilan keputusan persetujuan terhadap pemenang dan membuat prediksi hasil keputusan menjadi lebih akurat *.security clearances* apapun yang dibutuhkan.

METODE PENELITIAN

Dalam penelitian ini digunakan metode eksperimen, Eksperimen adalah metode penelitian yang bertujuan untuk meneliti hubungan (bisa berupa hubungan sebab akibat atau bentuk hubungan lainnya) antar dua variabel atau lebih pada satu atau lebih kelompok eksperimental, serta membandingkan hasilnya dengan kelompok yang tidak mengalami manipulasi yakni yang disebut dengan kelompok kontrol.

Metode Pengumpulan Data

Dalam penelitian ini untuk mendapatkan data-data yang diperlukan penulis menggunakan beberapa metode

pengumpulan data dengan studi pustaka dan wawancara serta kuisisioner.

Jenis Data

Menurut jenisnya maka data yang digunakan dalam penelitian ini terdiri dari data primer yaitu data yang diperoleh langsung dari sumbernya, dengan melakukan wawancara terhadap para expert system, serta dokumen kualifikasi dan data sekunder yaitu data yang diperoleh dari buku-buku literatur, journal, dokumen-dokumen, laporan.

Klasifikasi Data Berdasarkan Jenis

Datanya

Data kuantitatif adalah data yang dipaparkan dalam bentuk angka-angka. Data yang penulis dari dokumen dan arsip panitia berupa angka-angka hasil penilaian pemenang lelang dan data kualitatif adalah data yang disajikan dalam bentuk kata-kata yang mengandung makna. Data yang penulis peroleh dari para expert bagaimana menentukan pemenang lelang.

Nearest Neighbor (K-NN)

Model yang peneliti hasilkan adalah variable-variable yang bersifat kualitatif dalam hal ini pendapat para expert yang peneliti dapat dari kuisisioner . Untuk mengatasi masalah-masalah yang terjadi dalam menentukan pemenang lelang peneliti menggabungkan data kuantitatif dan kualitatif. Mengklasifikasikan data pemenang lelang dengan data mining menggunakan algoritma K-NN dan Manhattan Distance.

Dalam penelitian ini penulis menggunakan Retrieval K-Nearest Neighbor (KNN). *K-Nearest Neighbor* (KNN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari

algoritma ini adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik *query*, akan ditemukan sejumlah *k* obyek atau (titik *training*) yang paling dekat dengan titik *query*. Klasifikasi menggunakan *voting* terbanyak diantara klasifikasi dari *k* obyek.. algoritma KNN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru.

Nearest Neighbor menghitung tingkat kemiripan (jarak) suatu kasus terhadap kasus lain berdasarkan beberapa atribut yang didefinisikan berdasar pembobotan tertentu dan kemudian tingkat kemiripan (jarak) dari keseluruhan atribut akan dijumlahkan. *Nearest Neighbor* didefinisikan berdasar persamaan sebagai berikut [2] :

$$Similarity(T, S) = \sum_{i=1}^n f(T_i, S_i) W_i$$

- T : Kasus target / baru
- S : Kasus sumber / lama / pembandingan
- n : Jumlah atribut dalam setiap kasus
- i : Atribut individu dari 1 sampai n
- f : Fungsi kemiripan untuk atribut I dalam kasus T dan S
- w : Bobot atribut i

Prinsip K-Nearest Neighbor

K-nearest neighbor merupakan suatu metode untuk mengklasifikasikan suatu data baru berdasarkan kesamaan dengan labeled tertentu. Dalam hal ini kesamaan dinyatakan dalam metric jarak, dan satuan jarak dalam penelitian ini menggunakan Manhattan / City Distance.

Ada 2 jenis *Nearest Neighbor*, yaitu;

1. 1-NN

1-NN merupakan jenis algoritma *Nearest Neighbor* yang pengklasifikasian dilakukan terhadap 1 labeled terdekat. Urutan Algoritma 1-NN sebagai berikut:

- a. Hitung jarak antara data baru ke setiap labeled data
- b. Tentukan 1 labeled data yang mempunyai jarak yang paling minimal
- c. Klasifikasikan data baru ke dalam labeled data tersebut

2. K-NN

K-NN merupakan algoritma *Nearest Neighbor* yang pengklasifikasian dilakukan terhadap K labeled data terdekat, dimana $K > 1$. Urutan algoritma K-NN sebagai berikut:

- a. Tentukan nilai K
- b. Hitung jarak antara data baru ke setiap labeled data
- c. Tentukan K labeled data yang mempunyai jarak yang paling minimal
- d. Klasifikasikan data baru kedalam labeled data yang mayoritas

Algoritma metode KNN bekerja berdasarkan jarak terpendek dari *query instance* ke *training sample* untuk menentukan KNN-nya. *Training sample* diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi *training sample*. Sebuah titik pada

ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat dari titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan *jarak*.

Manhattan Distance

Sedangkan untuk menghitung jarak ada beberapa metode yang terkenal adalah Manhattan Distance [11], Euclidean Distance, Mahalanobis Distance, geometric similarity measures [12], and probabilistic similarity measures [13]. Manhattan atau City Distance digunakan untuk mengambil kasus yang cocok dari basis kasus dengan menghitung jumlah bobot absolute dari perbedaan antara kasus yang sekarang dan kasus yang laen dalam basis kasus.

Untuk menghitung bobot digunakan persamaan berikut :

$$d_{ij} = \sum W_k |x_{ik} - c_{jk}|$$

Dimana diketahui d_{ij} adalah jarak antara kasus antara i_{th} dan j_{th} dengan semua parameternya. W merepresentasikan jumlah dari bobot. X adalah kasus yang baru dikurangi dengan C yaitu history. Manhattan Distance adalah pengukuran similarity / kemiripan yang paling cocok untuk approval project yang merepresentasikan kasus yang relevant dengan angka yang natural atau dengan data yang bersifat kuantitative. [13]

Dengan menganalisa data, panitia lelang akan memiliki acuan untuk pengambilan keputusan persetujuan terhadap pemenang. Membuat prediksi hasil keputusan menjadi lebih akurat.

Tabel 1 : Data Pemenang Lelang Terdahulu dengan hasil project

No Kasus	VARIABEL							Y = Klasifikasi
	x1	x2	x3	x4	x5	x6	x7	
A	5	6	9	6	12	5	7	Good
B	11	6	7	9	8	6	7	Good

C	4	8	5	7	9	6	8	Bad
D	8	9	5	6	8	9	4	Bad
E	9	7	8	9	8	10	8	Good
F	6	5	7	4	3	8	8	Bad
G	6	7	9	5	7	8	9	Good
H	6	5	8	9	8	9	11	Good
I	4	8	8	9	6	5	8	Bad
J	3	4	6	7	9	10	7	Bad

1. Tetapkan nilai k = number of nearest neighbors

Dalam hal ini k = 3

2. Hitung jarak antara query dan semua sampel kasus

Dengan kasus baru dimana :

X1=10 X2=9 X3=8 X4=9 X5=4 X6=3 X7=9

maka dapat dihitung jarak antara kasus baru dengan kasus lama

$$d_{ij} = \sum W_k |x_{ik} - c_{jk}|$$

Matriks baris pertama kolom pertama : $|5-10|+|6-9|+|9-8|+|6-9|+|12-9|+|5-5|+|7-9| = 5$

Tabel 2 : Tabel Hasil Perhitungan Dengan Manhattan Distance

No Kasus	VARIABEL							Klasi fikasi	Y = Klasi fikasi
	x1	x2	x3	x4	x5	x6	x7		
A	5	3	1	3	8	2	2	24	Good
B	1	3	1	0	4	3	2	14	Good
C	6	1	3	2	5	3	1	21	Bad
D	2	0	3	3	4	6	5	23	Bad
E	1	2	0	0	4	7	1	15	Good
F	4	4	1	5	1	5	1	21	Bad
G	4	2	1	4	3	5	0	19	Good
H	4	4	0	0	4	6	2	20	Good
I	6	1	0	0	2	2	1	12	Bad
J	7	5	2	2	5	7	2	30	Bad

a. Perancangan Pemodelan Dengan UML (Unified Modeling Language)

Metode perancangan sistem menggunakan *use case diagram* dan *sequence diagram*. *Use case diagram* menggambarkan interaksi antara aktor dengan proses atau sistem yang dibuat, dan *sequence diagram* menggambarkan interaksi antara objek-objek dalam suatu sistem.

T5	BAGUS	BAGUS	1			
T6	JELEK	JELEK				1
T7	BAGUS	BAGUS	1			
T8	JELEK	JELEK				1
T9	JELEK	JELEK				1
T10	JELEK	JELEK				1
T11	JELEK	JELEK				1
T12	BAGUS	BAGUS	1			
T13	BAGUS	BAGUS	1			
T14	BAGUS	BAGUS	1			
T15	BAGUS	BAGUS	1			
T16	BAGUS	BAGUS	1			
T17	JELEK	JELEK				1
T18	BAGUS	BAGUS	1			
T19	BAGUS	BAGUS	1			
T20	BAGUS	BAGUS	1			
T21	JELEK	JELEK				1
T22	BAGUS	BAGUS	1			
T23	BAGUS	JELEK		1		
T24	BAGUS	JELEK		1		
T25	BAGUS	BAGUS	1			
T26	BAGUS	BAGUS	1			
T27	BAGUS	BAGUS	1			
T28	BAGUS	BAGUS	1			
T29	BAGUS	JELEK		1		
T30	BAGUS	JELEK		1		
			19	4	0	7

Berdasarkan pengukuran pemenang lelang dengan metode Case Base Reasoning didapat hasil sebagai berikut :

$$\begin{aligned} \text{Accuracy} &= (TP + TN) / (TP + FP + TN + FN) \\ &= (19 + 7) / (19 + 4 + 0 + 7) \\ &= 0,8667 \end{aligned}$$

$$\begin{aligned} \text{Precision} &= TP / (TP + FP) \\ &= 19 / (19 + 4) \\ &= 0,8261 \end{aligned}$$

$$\begin{aligned} \text{Recall} &= TP / (TP + FN) \\ &= 19 / (19 + 0) \\ &= 19 / 19 \\ &= 1 \end{aligned}$$

$$\text{F.Measure} = \frac{(2 \cdot \text{Recall} \cdot \text{Precision})}{(\text{Recall} + \text{Precision})}$$

$$= 1,6522 / 1,8261$$

$$= 0,904$$

Dari pengujian data tersebut diatas dapat diperoleh hasil sebagai berikut :

Tabel 17 : Tabel Hasil Pengujian Data

No	Precision	Recall	F.Measure	Accuracy
1	0,8261	1	0,904	0,8667

Tabel ini menunjukkan bahwa pengukuran kinerja system diatas diperoleh Precision = 0,8261 ; Recall = 1; Accuracy =

0,8667; F.Measure = 0,904. Ini menunjukkan bahwa system prediksi ini bekerja dengan bagus dan bisa digunakan sebagai alat bantu panitia lelang untuk memprediksi pemenang yang akurat.

Kesimpulan

Dari hasil penelitian yang dilakukan mulai dari tahap awal hingga pengujian, penerapan sistem prediksi pemenang lelang dengan algoritma k-NN dan manhattan distance yang diimplementasikan berupa DSS, didapatkan kesimpulan bahwa system prediksi pemenang lelang dapat membantu proses penentuan keputusan pemenang lelang lebih akurat dengan menggunakan Algoritma K-NN dan Manhattan Distance dengan Accuracy 0,8.

Saran

Penelitian akan lebih akurat apabila training set lebih banyak dan untuk penelitian yang akan datang system prediksi untuk menentukan pemenang lelang bisa menggabungkan dengan metode Fuzzy Logic sehingga hasil akan menjadi lebih akurat.

DAFTAR PUSTAKA

- [1] Departemen Pekerjaan Umum (2007), Peraturan Menteri Pekerjaan Umum Nomor 43 Tahun 2007 tentang Standar dan Pedoman Pengadaan Jasa Konstruksi, Departemen Pekerjaan Umum, Jakarta.
- [2]. Nabil El-Sawalhi , David Eaton , Rifat Rustom ; Contractor pre-qualification model: State-of-the-art; International Journal of Project Management 25 (2007) 465–474., Science Direct 2007.
- [3]. J.S. Russell, M.J. Skibniewski, Decision criteria in contractor prequalification, Journal of Management in Engineering, ASCE 4 (2). 1988. 148–164.
- [4]. Wu X, Kumar V. *The Top Ten Algorithms in Data Mining*. New York: CRC Press; 2009.
- [5]. Johanes Widagdho Yodha , Achmad Wahid Kurniawan, Pengenalan Motif Batik Menggunakan Deteksi Tepi Canny dan K-Nearest Neighbor, Techno Com , 2014.
- [6]. Witten, I. H., Frank, E., & Hall, M. A., *Data Mining: Practical Machine Learning and Tools*. Burlington Morgan Kaufmann Publisher. , 2011.
- [7]. Larose, D. T. , *DiscoverData*. New Jersey Sons, Inc. , 2005.
- [8]. Sumathi, & S., Sivanandam, S.N. , *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer, 2006.
- [9]. Larose D. *Discovering Knowledge in Data*. USA: John Wiley's and Son; 2005
- [10]. Han, J., & Kamber, M., *Data Mining Concept and Techniques*. San Francisco: Morgan Kaufman. , 2006.
- [11]. Sai On Cheung a,*, Peter Shek Pui Wong a, Ada S.Y. Fung b, W.V. Coffey, Predicting project performance through neural networks, International Journal of Project Management 24 (2006) 207–215 , Scient Direct, 2006.
- [12]. Lin H., Zhang, G. J., Li, S., Qiu, Y. C., Design and implementation of examination and approval system based on CBR, Machine Learning and Cybernetics, 2003 International Conference on Publication Date: 2-5 Nov. 2003 Vol.3 1477- 1481
- [13]. Muhammad K. Farooq, Malik Jahan Khan, Shafay Shamail and Mian M. Awais, Intelligent Project Approval Cycle for Local Government – Case-Based Reasoning Approach, CEGOV2009, November 10-13, Bogota, Colombia, ACM 2009.