

## Data Mining untuk Estimasi Biaya Produksi pada Industri Kecil dengan Sistem Produksi *Job order*

Uuf Brajawidagda

Jurusan Teknik Informatika  
Politeknik Negeri Batam  
Parkway, Batam Centre, Batam 29461, Indonesia  
E-mail: uuf@polibatam.ac.id

### Abstrak

Sistem produksi *job order* yang sering digunakan oleh industri kecil adalah sistem produksi berbasis pesanan yang dicirikan dengan variasi produk banyak, skala produksi berukuran kecil dan daur hidup produk yang pendek. Pesanan yang datang dari konsumen harus ditanggapi dengan segera dalam bentuk penawaran harga dan waktu produksi. Proses penawaran harga menuntut kecepatan proses estimasi biaya produksi sehingga respon kepada konsumen dapat dilakukan lebih cepat. Dengan metode K-Means dan Multiple Linear Regression (MLR) secara bersama-sama, estimasi biaya dengan metode analogi dapat dilakukan sehingga dapat mempercepat proses estimasi biaya. Metode K-Means digunakan untuk melakukan pengelompokan produk sedangkan metode MLR digunakan untuk membangkitkan model estimasi biaya untuk setiap kelompok. Akurasi model diukur dengan menggunakan Mean Absolute Percentage Error (MAPE) dan digunakan untuk menentukan kemampuan model estimasi biaya yang dibangkitkan.

**Kata Kunci:** Job Order, Estimasi Biaya, K-Means, Multiple Linear Regression

### 1 Pendahuluan

Dalam lingkungan bisnis manufaktur *job order*, *order response* dilakukan dengan menyampaikan penawaran harga dan waktu produksi kepada konsumen setelah konsumen menyampaikan pesanan. Estimasi biaya produksi dibuat berdasarkan gambar rancangan produk manufaktur yang dimodelkan dalam bentuk file Computer Aided Design (CAD). Proses estimasi biaya yang didasarkan pada rancangan produk membutuhkan waktu dan sumber daya yang besar, sehingga *order response* menjadi lama. Karena itu diperlukan metode estimasi biaya lain sehingga *order response* dapat dipercepat.

Seiring perkembangan teknologi di bidang pengelolaan data dan data mining, data historis dapat digunakan untuk mendukung kegiatan pengambilan keputusan. Data Mining adalah sebuah tahapan dalam *Knowledge Discovery in Database (KDD)* yang diawali dengan kegiatan data cleaning dan data selection (tahap persiapan data) untuk menghasilkan pattern dan rule. Selanjutnya *pattern* dan *rule* akan dievaluasi oleh manajemen organisasi untuk digunakan sebagai salah satu pendukung pengambilan keputusan.

Rancangan produk manufaktur berisi informasi tentang fitur geometri dan fitur pemesinan, dimana kedua informasi tersebut berkorelasi dengan proses produksi yang diperlukan untuk membentuk sebuah

produk. Informasi geometri dan pemesinan yang terdapat pada data historis rancangan produk manufaktur mengandung informasi yang berharga yang dapat membantu proses estimasi biaya produksi manufaktur. Karena itu dalam penelitian ini membahas ekstraksi informasi dari data historis rancangan produk manufaktur yang selanjutnya digunakan sebagai dasar estimasi biaya produksi manufaktur melalui teknik *clustering* dan *predictive data mining*.

### 2 Estimasi Biaya pada Manufaktur

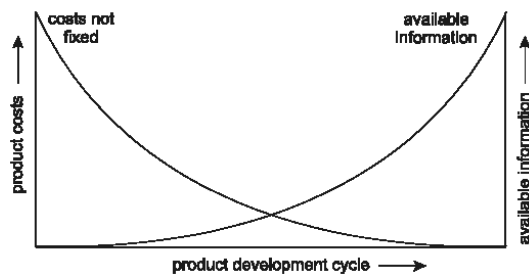
Menurut Chisholm (1990), manufaktur adalah serangkaian aktifitas dan operasi yang melibatkan kegiatan perancangan, pemilihan material, perencanaan produksi, proses produksi dan penjaminan kualitas produk. Biaya, secara umum dapat didefinisikan sebagai sumber daya ekonomi yang diperlukan untuk menyelesaikan aktifitas pekerjaan dan biasanya dinyatakan dalam satuan mata uang.

TABEL 1 PENERAPAN METODE PERKIRAAN BIAYA DALAM DAUR HIDUP PRODUK (LONG, 2000)

Tahap	Analitis	Analogi	Parametrik
Konsepsi		▲	▲
Pengembangan	▲	▲	▲
Produksi	▲		▲
Operasi	▲		▲
Purna Jual	▲	▲	▲

Dalam manufaktur, perkiraan biaya adalah prosedur untuk memperkirakan biaya manufaktur dari sebuah produk sebelum seluruh tahapan dalam siklus pengembangan produk dilakukan. Long (2000) mengelompokkan siklus pengembangan produk menjadi 5 (lima) tahap, yaitu: konsepsi, pengembangan, produksi, operasi dan purna jual seperti pada Tabel 1. Pada awal daur hidup produk, informasi yang tersedia sangat terbatas dan akan semakin lengkap seiring pertambahan tahapan daur hidup produk tersebut. Berdasarkan ketersediaan informasi, metode perkiraan biaya yang sesuai untuk tahap konsepsi adalah metode analogi dan parametrik.

Kegiatan *order response* merupakan salah satu kegiatan dalam tahap konsepsi, dimana informasi detail produk belum tersedia. Metode perkiraan biaya produksi yang sesuai dalam tahap ini adalah metode analogi dan parametrik seperti pada Tabel 1.



Gambar 1 Paradok Estimasi Biaya (Bode, 1998)

Besaran biaya dalam manufaktur yang diperlukan untuk menyelesaikan sebuah produk dapat diperkirakan lebih akurat pada saat informasi detail produk sudah tersedia. Dalam tahap perancangan produk, 70% informasi detail produk sudah tersedia sehingga diharapkan perkiraan biaya yang dilakukan pada tahap ini lebih akurat. Namun demikian, pada tahap perancangan produk, informasi tentang detail informasi produk belum tersedia, sehingga sulit untuk membuat perkiraan biaya pada tahap ini. Fenomena ini disebut sebagai *cost estimation paradox* seperti pada Gambar 1.

Menurut Weustink (2000), perkiraan biaya yang dilakukan berdasarkan pada rancangan produk ditentukan oleh cost driver yang saling terkait, terdiri dari geometri, material, proses produksi dan perencanaan produksi. Pada tahap perancangan, cost driver yang sangat berpengaruh pada perkiraan biaya adalah geometri dan material.

Secara umum menurut Camargo (2003) terdapat 3 metode dalam penentuan cost estimation pada industri manufaktur:

1. Metode analitis, yaitu perkiraan biaya berdasarkan data perancangan. Rancangan produk dirinci menjadi bill of material dan proses-proses yang dibutuhkan untuk menyelesaikan produk tersebut. Metode ini

sesuai untuk diterapkan pada tahap produksi pengembangan sebuah produk baru. Kelemahan dari metode ini adalah a) hanya dapat diterapkan jika data rancangan dan proses produksi sudah diketahui b) membutuhkan data dan sumber daya yang besar.

2. Metode analogi, yaitu perkiraan biaya berdasarkan kemiripan produk terhadap produk sejenis yang pernah diproduksi. Aspek kemiripan yang dibandingkan dalam metode ini di antaranya adalah kemiripan fungsi, kemiripan fitur dan kemiripan morfologi fisik. Metode ini menuntut kepakaran dan pengalaman dalam melakukan penentuan tingkat kemiripan antar produk.
3. Metode parametrik, yaitu metode dengan perkiraan biaya dengan pemodelan matematis dari sejumlah *Cost Estimation Relationships* (CER), aturan dasar (*ground rules*), asumsi-asumsi, variabel dan konstanta untuk mendeskripsikan dan mendefinisikan kondisi tertentu. Metode ini dapat dilakukan dengan cepat, namun menuntut pemutakhiran model matematis sesuai perkembangan variasi produk. Metode ini sangat berguna terutama pada tahap awal perancangan produk dimana informasi detail belum tersedia. CER merupakan ekspresi matematis yang memberikan biaya produksi sebagai fungsi dari satu atau lebih variabel cost driver. Kelebihan penggunaan CER adalah: memungkinkan estimasi yang cepat tanpa perlu informasi detail dan CER dilakukan berdasarkan ongkos produksi historis yang nyata.

### 3 Algoritma K-Means dan MLR

Algoritma k-means merupakan salah satu teknik clustering dengan *partitioning method*, dimana inialisasi titik pusat cluster secara random sangat mempengaruhi hasil proses *clustering*. Menurut Deeters (2007), perbaikan proses inialisasi awal titik pusat *cluster* dapat dilakukan dengan cara melakukan partisi data berdasar pada atribut yang mempunyai variansi tertinggi.

Sedangkan *Multiple Linear Regression* (MLR) merupakan salah satu teknik dalam predictive data mining. Menurut Fayyad (1996), tujuan akhir dari data mining adalah untuk membuat prediksi terhadap data baru berdasarkan data historis, karena itu *predictive data mining* banyak diterapkan dalam berbagai aktivitas bisnis. *Predictive data mining* prediksi sebuah label data baru berdasarkan aturan yang telah dibangkitkan dari data historis.

## 4 Metode

Penelitian ini berencana melakukan data mining untuk data historis rancangan produk manufaktur. Rancangan produk manufaktur biasanya disimpan dalam bentuk gambar. Karena itu dalam penelitian ini perlu melakukan proses ekstraksi gambar menjadi informasi berbentuk teks agar pemrosesan data dapat dilakukan lebih lanjut. Proses ekstraksi dan pengolahan data dalam penelitian ini dilakukan oleh perangkat lunak yang secara khusus dikembangkan untuk mendukung penelitian ini.

Terkait teknik data mining yang digunakan, maka perlu dilakukan kegiatan penyiapan data, yang dalam hal ini terdiri dari: 1) *Data cleaning*, terdiri dari kegiatan untuk menghilangkan *noise* dan mengelola *missing value*. 2) *Data integration and transformation*, yaitu integrasi dari beberapa sumber data berupa database, file atau *data cube*.

Sedangkan untuk teknik data mining yang digunakan, digunakan kombinasi antara *clustering* dan *classification*. *Clustering* ini digunakan untuk mengelompokkan produk yang mirip. Kemiripan dapat diformalisasikan menjadi tiga bagian, yaitu:

1. Kemiripan sebagai predikat, menyatakan apakah dua buah objek A dan B mempunyai kemiripan atau tidak
2. Kemiripan sebagai tingkat relasi, menyatakan bahwa objek A lebih mirip kepada objek B dibandingkan objek C
3. Kemiripan sebagai ukuran, menyatakan tingkat kuantitatif kemiripan dari dua buah objek A dan B

Layer (2003) menyatakan bahwa tidak ada cara untuk mengukur kemiripan secara umum, karena kemiripan selalu terkait dengan tujuan spesifik yang ingin dicapai.

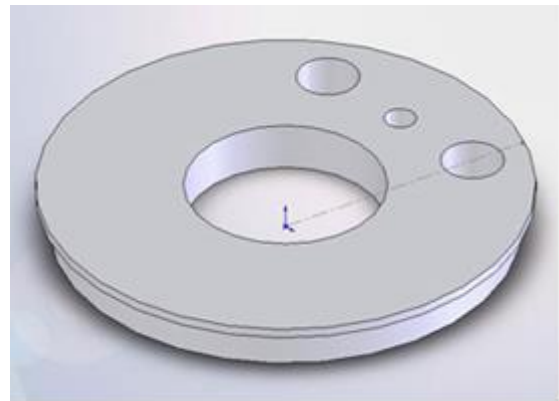
Dalam penelitian ini, *clustering* dengan teknik *k-means* digunakan sebagai pengolahan awal terhadap data historis rancangan produk dengan pertimbangan kemudahan perhitungan dan kemudahan visualisasi untuk keperluan interpretasi. *K-means* membagi koleksi dari  $n$  vector  $x_j, j=1, \dots, n$  menjadi  $c$  kelompok  $G_i, i=1, \dots, c$  dan menentukan pusat *cluster* dalam setiap kelompok dengan berdasarkan *cost function* yang minimum. Selanjutnya, teknik *multiple linear regression* (MLR) digunakan untuk menentukan rumus estimasi biaya produksi untuk setiap kelompok produk yang mempunyai kemiripan fitur geometri dan fitur pemesinan.

Untuk akurasi prediksi, ukuran yang digunakan adalah Minimum Average Percentage of Error (MAPE). MAPE adalah rata-rata persentase absolut dari penyimpangan hasil estimasi, yang didapat dengan

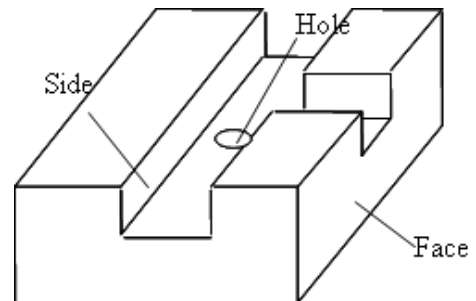
menghitung jumlah kesalahan absolut tiap data dan selanjutnya dibagi dengan jumlah data.

## 5 Data Set

Data yang digunakan adalah 44 buah file solidwork (\*.sldprt) yang merupakan data rancangan produk manufaktur di sebuah industri kecil PT X pada periode bulan Januari-Mei 2007. Jumlah dataset yang digunakan sangat sedikit untuk keperluan data mining, dikarenakan tingkat ketersediaan file rancangan dalam bentuk solidwork yang terbatas. Contoh file yang akan diekstrak disajikan pada Gambar 2.



Gambar 2 Contoh File Gambar Rancangan Produk Manufaktur



Gambar 3 Fitur Pemesinan

## 6 Ekstraksi Gambar Rancangan Manufaktur

Proses ekstraksi ini akan mengenali jenis objek komponen gambar pada setiap sumbu gambar  $\{x+, x-, y+, y-, z+, z-\}$ . Informasi yang diekstrak dari gambar rancangan manufaktur berisi informasi terkait fitur pemesinan. Fitur adalah sekumpulan bentuk permukaan produk manufaktur yang mempunyai arti bentuk rekayasa khusus seperti pada Gambar 3. Knoblach (1998) meneliti tentang perkiraan biaya produksi berbasis pada biaya pembuatan fitur untuk mass production.

orderID	productName	MatLength
7030	MOLD RING HINO LAF	300
7030	ORING HINO M	13.7
7030	plate cavity	250
7030	plate core	250
7030	RUBBER RING HINO L	40
7030	RUBBER RING HINO S	26.5
7030	striper	250

Gambar 4 Hasil Ekstraksi Fitur Produk Manufaktur

Secara rinci, informasi yang diekstrak dalam penelitian ini terdiri dari OrderID, ProductName, Fitur Geometri dan Fitur Pemesinan seperti disajikan pada Gambar 4. Informasi Fitur Geometri berisi tentang Panjang, Lebar, Tinggi dan Volume Produk, sedangkan Fitur Pemesinan menyimpan informasi mengenai jumlah Hole, Face, Pocket, Boss, Surface dan Side.

## 7 Penyiapan Data

Kegiatan penyiapan data yang digunakan dalam penelitian ini terdiri dari 2, yaitu pemilihan atribut dan penanganan nilai kosong. Seleksi atribut digunakan untuk melakukan pemilihan atribut yang akan digunakan untuk data mining. Proses ini mempergunakan data masukan pengguna berdasarkan skema data hasil ekstraksi seperti disajikan pada Gambar 5.

Sedangkan penanganan nilai kosong digunakan untuk mengisi nilai kosong yang mungkin didapatkan sehingga data menjadi lengkap. Nilai kosong diisi dengan nilai mean dari sebuah atribut tempat dimana data kosong berada.

Feature Name	Data Type
MatLength	double
MatWidth	double
MatThick	double
MatVolume	double
NumFace	int(11)
NumBoss	int(11)
NumSurface	int(11)
OrderID	varchar(4)
ProductName	varchar(50)
NumHole	int(11)
NumPocket	int(11)
NumSide	int(11)

Gambar 5 Pemilihan Fitur Produk untuk Clustering

## 8 Clustering

Clustering dimulai dengan kegiatan inialisasi titik pusat *cluster* sesuai langkah pertama algoritma K-Means. Pada algoritma K-Means, titik pusat cluster

awal ditentukan dengan cara memilih sejumlah data secara acak. Dengan proses ini, diharapkan proses pengelompokan produk dapat dilakukan secara lebih baik dibandingkan dengan penggunaan algoritma K-Means biasa. Proses ini dilakukan dengan menentukan jumlah cluster seperti disajikan pada Gambar 6 dan selanjutnya hasil dari proses ini adalah sejumlah *centroid* sesuai dengan jumlah *cluster*.

OrderID	ProductName	NumFac
7038	15. Upper Table	6
7038	16. Lower plate Mold ur	6
7038	handlae	2
7038	limit switch	4
7038	lower plate mold	6
7038	UPPER PLATE MOULD	6
7042	MOLD RING HINO LAF	6
7042	MOLD RING HINO SM	6
7042	ORING HINO M	2
7042	RUBBER RING HINO L	2
7042	RUBBER RING HINO S	2

Number of cluster(s): 3

Gambar 6 Proses Clustering

Selanjutnya dilakukan pengelompokan data ke dalam sejumlah kelompok data yang disebut *cluster*. Proses ini dilakukan dengan cara menghitung jarak Euclidian data terhadap *centroid*. Jarak terkecil terhadap titik pusat cluster dijadikan sebagai dasar pengalokasian data ke dalam *cluster*.

## 9 Pembuatan Aturan Estimasi dan Pengujian

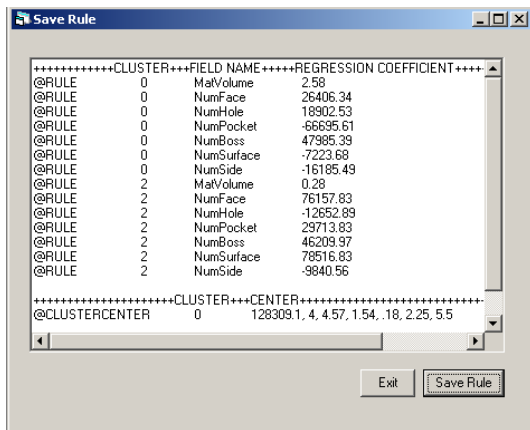
Pembuatan aturan estimasi dilakukan dengan menggunakan algoritma MLR. Proses pembuatan algoritma dilakukan dengan melakukan langkah-langkah berikut:

1. Membuat matrik persamaan linier untuk membuat dua buah matrik, yaitu matrik yang berisi variabel dependen dan matrik biaya. Matrik variabel dependen diperoleh dari proses *clustering*, sedangkan matrik biaya diperoleh dari data pesanan. Matrik variabel dependen selanjutnya digunakan sebagai data masukan pada proses penghitungan koefisien regresi.
2. Membuat koefisien regresi dilakukan untuk menghitung koefisien nilai regresi.

Dengan dua proses tersebut, maka aturan estimasi pada masing-masing *cluster* diperoleh seperti disajikan pada Gambar 7.

Setelah aturan estimasi terbentuk, dilakukan pengujian

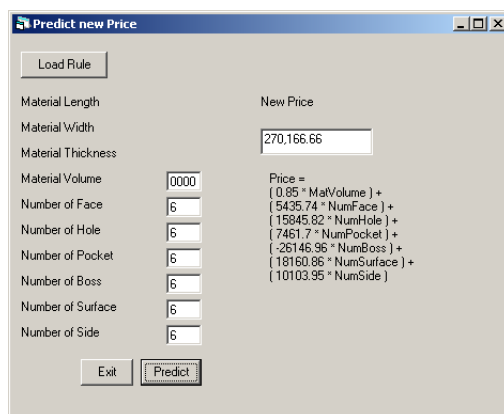
dengan menggunakan data historis. Pengujian dilakukan dengan membandingkan antara nilai estimasi dengan biaya produksi historis. Dari 44 data yang diuji, menghasilkan nilai MAPE yang bervariasi dengan kecenderungan semakin banyak jumlah cluster, maka nilai MAPE lebih kecil. Artinya, semakin banyak jumlah cluster, maka akurasi estimasi semakin baik. Namun demikian, ketika jumlah cluster semakin banyak, akurasi tidak linear turun terus tetapi cenderung stabil pada kisaran 34% seperti disajikan pada Tabel 2.



Gambar 7 Aturan Estimasi setiap Cluster

TABEL 2 HASIL PENGUJIAN

Jumlah Data	Jumlah Cluster	MAPE	Akurasi
44	2	34.5%	65.5%
44	3	32.5%	67.5%
44	4	32.1%	67.9%
44	5	31.5%	68.5%
44	10	32.0%	68.0%
44	20	36.0%	64.0%
44	30	33.5%	66.5%
44	40	34.0%	66.0%



Gambar 8 Estimasi terhadap Produk Baru

## 10 Estimasi Harga Terhadap Produk Baru

Terdapat dua langkah yang dilakukan dalam menghitung estimasi harga baru ini. Pertama adalah menentukan keanggotaan sebuah data baru terhadap kelompok data yang dihasilkan dari aturan estimasi. Jarak Euclidian terkecil antara data baru dengan titik pusat masing-masing cluster dijadikan sebagai acuan.

Setelah keanggotaan dalam sebuah cluster ditentukan, maka estimasi biaya produk berdasarkan koefisien regresi dapat dilakukan dengan cara mengalikan parameter prediksi dengan masing-masing koefisien regresi akan menghasilkan prediksi biaya seperti disajikan pada Gambar 8.

## 11 Pembahasan

Ekstraksi data fitur pemesinan sudah mampu mengekstraksi informasi umum produk seperti: panjang, lebar dan tinggi material serta jumlah Hole, Face, Pocket, Boss, Surface dan Side. Namun demikian, ekstraksi belum dapat merepresentasikan proses pemesinan karena informasi tidak dilakukan secara lebih rinci. Informasi seperti volume massa yang dihasilkan dari sebuah fitur tentu akan sangat berpengaruh terhadap proses pemesinan yang diperlukan. Data fitur pemesinan yang lebih rinci diduga akan menghasilkan proses pengelompokan produk dan prediksi yang lebih baik.

Pada teknik penyiapan data, sebenarnya terdapat banyak pilihan yang digunakan. Dalam penelitian ini hanya dapat disediakan pilihan terbatas kepada pengguna. Terkait dengan teknik data mining yang digunakan, deskripsi kelompok produk dapat ditambahkan sehingga memudahkan pengguna untuk memahami batasan kuantitatif dari sebuah kelompok produk. Selain itu, teknik data mining lain untuk keperluan prediksi produk baru dapat diterapkan sehingga dapat memberikan pertimbangan yang lebih baik bagi pengguna perangkat lunak dalam menentukan harga prediksi.

Terhadap hasil estimasi, dengan nilai MAPE berkisar 24%, akurasi estimasi dipastikan tidak baik.

## 12 Kesimpulan

Estimasi biaya selalu menjadi pekerjaan penting dalam system produksi *job order*. Penelitian ini telah berhasil membuktikan bahwa sebuah teknik estimasi berdasarkan data historis dapat dilakukan. File rancangan produk masa lalu dapat dijadikan sebagai basis estimasi. Namun demikian, identifikasi fitur geometri sebagai dasar estimasi biaya pemesinan masih merupakan tantangan. Berbagai penelitian telah dilakukan pada bidang ini dan terus berkembang. Pada akhirnya, penelitian ini dapat dijadikan untuk

penelitian lanjutan pada area estimasi biaya, *manufacturing feature identification* ataupun data mining.

### Referensi

- [1] Betts, M., "The Almanac: Hot Tech", ComputerWorld 52, 2003
- [2] Berkhin, P., "Survey of Clustering Data Mining Techniques", Accrue Software, 1045 Forest Knoll Dr., San Jose, 2002
- [3] Bode, J., 'Decision support with neural networks in the management of research and development: concepts and application to cost estimation', Information & Management 34, pp. 33-40, 1998
- [4] Brinke, E.T., "Costing Support and Cost Controlling in Manufacturing: A Cost Estimation Tool Applied in The Sheet Metal Domain", PrintPartners Ipskamp, Enschede, The Netherlands, 2002
- [5] Camargo M., A.M Jolly Dessodt, J.M Castelain, "Application of Parametric Cost Estimation in the Textile Supply Chain", Journal of Textile Apparel, Volume 3, Issue 1, 2003
- [6] Chisholm, A.W.J, 'Nomenclature and Definitions for Manufacturing Systems', Annals of the CIRP 39/2, pp. 735-742, 1990.
- [7] Deelters, S., "Enhancing K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance", Proceedings of World Academy of Science, Engineering And Technology Volume 26 ISSN 1307-6884, 2007
- [8] Fayyad, Usama, M., "Data-Mining and Knowledge Discovery: Making Sense Out of Data", Microsoft Research IEEE Expert, 11:5. (1996), pp. 20-25, 1996
- [9] Giudici, P., "Applied Data-Mining: Statistical Methods for Business and Industry", West Sussex, England: John Wiley and Sons, 2003.
- [10] Godswill, C Nsofor, "A Comparative Analysis of Predictive Data Mining Techniques", Tesis, University of Tennessee, 2006
- [11] Han Jiawei and Micheline Kamber, "Data Mining Concept and Techniques", Morgan Kaufmann, 2001
- [12] Knoblach, Jurgen, "Cost Estimation for Large Scale Production of Sheet Metal Parts Using Artificial Neural Networks", Production Engineering Journal, Vol. V/2, 1998
- [13] Layer, Alexander, "Case-Based Cost Estimation, A Building Block for Product Cost Management and Design-For-X", University of Twente, 2003
- [14] Long J., "Parametric Cost Estimating in The New Millenium", Price Sistem, 2000
- [15] Weiss, Sholom M. et al., "Predictive Data-Mining: A Practical Guide", San Francisco, Morgan Kaufmann, 1998
- [16] Weustink, I.F., Brinke, E.T., Streppel, A.H., Kals, H.J.J., 'A generic framework for cost estimation and cost control in product design', Journal of Materials Processing Technology 103, pp. 141-148, 2000