

NV Kuznietsova, Cand. Sc. (Eng.), Ass. Prof.

## TECHNOLOGIES INFORMASI UNTUK MENGANALISIS KEUANGAN PENINDASAN DI PROZORRO PLATFORM

*Kertas menunjukkan kemungkinan menganalisis pembelian on-line di platform yang ProZorro dengan metode data mining dengan tujuan untuk mengidentifikasi perilaku perusahaan dan karakteristik dimana adalah mungkin untuk mengungkapkan kolusi dan kegiatan yang melanggar hukum selama keikutsertaan mereka dalam perdagangan on-line.*

**Kata kunci:** regresi logistik, jaringan saraf, pembelian tender, ProZorro, teknologi informasi.

### **pengantar**

Dengan tujuan untuk memberikan pembelian yang efektif dan transparan, untuk menciptakan lingkungan yang kompetitif di bidang pengadaan publik, untuk mencegah korupsi di bidang ini, untuk mengembangkan persaingan yang adil, Verkhovna Rada Ukraina mengadopsi Hukum "Pada Pengadaan Publik" [1]. adopsi dan pengembangan Hukum ini sendiri merupakan langkah yang signifikan dalam memerangi perjanjian non-transparan dan kolusi, yang alasan untuk menyia-nyiakan keuangan anggaran dan biaya yang terlalu tinggi dari karya-karya. Undang-undang ini memandang bahwa semua layanan dan barang untuk jumlah yang melebihi 50 ribu UAH harus dilakukan dengan menggunakan sistem pengadaan elektronik untuk memilih pemasok barang, penyedia jasa dan pemain karya untuk penandatanganan kontrak, pelanggan harus mematuhi dengan prinsip-prinsip pelaksanaan pengadaan publik dinyatakan dalam UU. Untuk tujuan ini "ProZorro" sistem diciptakan, yang merupakan database elektronik untuk pelaksanaan pengadaan negara publik dalam modus on-line [2]. Sistem itu sendiri menyediakan pengguna portal dengan kemampuan akses real-time ke semua pembelian yang dilakukan dan dengan demikian memverifikasi transparansi pengadaan, akses yang sama dari semua peserta pasar serta memeriksa langsung bagaimana pajak warga Ukraina dihabiskan untuk layanan pembelian sektor negara.

**Kertas tujuan di** menganalisis dan memverifikasi transparansi tender pengadaan menggunakan metode data mining, diimplementasikan dalam bentuk teknologi informasi pada platform yang SAS, untuk mengungkapkan kemungkinan pelanggaran dan pelanggaran. Berdasarkan hasil analisis, adalah mungkin untuk menguraikan rekomendasi tentang perbaikan lebih lanjut dari sistem monitoring pengadaan online dan untuk mengurangi kerugian keuangan wajib pajak yang disebabkan oleh non-transparansi perdagangan dan tidak dapat diaksesnya pelaku pasar kompetitif yang nyata kepada mereka.

### **Pemantauan e-procurement**

Untuk memverifikasi kewajaran perdagangan yang dilakukan dan pembelian, Hukum Ukraina [1] membayangkan kemungkinan untuk menyatakan pelanggaran tersedia melalui media massa atau melalui asosiasi publik dan untuk memberikan informasi resmi dari otoritas negara atau dari badan-badan pemerintahan lokal yang mandiri dalam hal pelanggaran, diidentifikasi oleh tubuh kontrol keuangan, atau dalam hal indikator risiko otomatis terdeteksi di pembelian dilakukan.

Makalah ini akan menunjukkan kemungkinan untuk meningkatkan indikator risiko otomatis sebagai kriteria khusus dengan yang telah ditentukan parameter, yang memungkinkan deteksi otomatis dari tanda-tanda pelanggaran hukum dan pelanggaran dalam prosedur pengadaan.

Untuk saat ini, 22 platform perdagangan berwenang dalam basis ProZorro: Zakupki.Prom.ua, lembut e-, Newtend, SmartTender, «Держзакупівлі онлайн», PublicBid dan «ПриватМаркет», dll daftar mereka terus diperbarui di situs [2]. Satu-satunya cara untuk terhubung ke sistem ProZorro adalah melalui salah satu platform.

Untuk mendeteksi tanda-tanda pelanggaran undang-undang di bidang pengadaan publik, data berikut dapat digunakan: informasi yang diungkapkan dalam sistem pengadaan elektronik, informasi yang terkandung dalam register negara kesatuan, informasi dalam database membuka kepada otoritas eksekutif pusat, yang menerapkan ini kebijakan negara di bidang pengawasan keuangan publik, data negara

badan otoritas dan badan-badan lokal pemerintahan sendiri, perusahaan, lembaga, organisasi, pelanggan dan peserta dari prosedur pengadaan, yang dapat diperoleh oleh badan pengawasan keuangan negara dengan cara yang ditentukan oleh UU.

### Analisis e-procurement dalam rangka membangun keteraturan dan elaborasi tertentu rekomendasi

Untuk meningkatkan prosedur monitoring dan sistem deteksi otomatis dan penolakan perusahaan yang tidak adil, sistem mengimplementasikan satu set tertentu dari algoritma klasifikasi. Ada tidak ada akses terbuka untuk pengaturan dan parameter, dimana klasifikasi dilakukan, untuk mencegah kemungkinan manipulasi dan penyediaan data statistik palsu dari perusahaan mengajukan permohonan partisipasi dalam tender. Oleh karena itu, mengingat tujuan simulasi mengidentifikasi hubungan sebab-akibat dan karakteristik statistik untuk memastikan klasifikasi kualitatif serta memperoleh informasi untuk prediksi pembelian mencurigakan diperlukan untuk aplikasi resmi dalam urutan, yang ditentukan oleh UU [1], untuk melakukan pemeriksaan.

Asumsi dibuat tentang adanya hubungan tertentu antara durasi partisipasi perusahaan dalam kesepakatan dan karakteristiknya, yaitu, apakah durasi partisipasi perusahaan dalam kesepakatan tergantung pada kecurigaan dalam kegiatan melanggar hukum yang mungkin pada platform (yaitu di kolusi dengan perusahaan lain). Sampel Berikut ini data yang empiris dibentuk:

1. Menang - jumlah kemenangan dari perusahaan tertentu;
2. Kerugian - jumlah kerugian perusahaan di penawaran;
3. Sum\_of\_deals - Jumlah total penawaran won;
4. Partisipasi - jumlah partisipasi dalam penawaran;
5. Keberatan - jumlah keberatan yang diajukan oleh perusahaan yang diberikan;
6. DATE\_START - tanggal mulai partisipasi dalam sistem penawaran;
7. Date\_finish - tanggal partisipasi terakhir di penawaran;
8. IdTenderer - nomor unik dari peserta lelang;
9. Diduga - variabel, yang menunjukkan jika sebuah perusahaan diduga dalam kolusi ilegal dengan lainnya peserta;
10. Churn out - variabel target, yang sama dengan 1 jika perusahaan berhenti keikutsertaannya dalam penawaran dalam jangka waktu singkat (diasumsikan bahwa perusahaan tiba-tiba berhenti partisipasi atau itu sebuah perusahaan fiktif untuk satu transaksi saja).

Perusahaan ini dianggap sebagai salah satu yang terus trading, jika periode antara memulai perdagangan pada platform dan waktu dari kesepakatan terakhir adalah lebih dari 60 hari (rata-rata durasi statistik siklus bisnis dari perusahaan diberikan pada platform menurut data resmi dari ProZorro [2]). Hal ini juga diperhatikan bahwa perusahaan yang berpartisipasi dalam kompetisi di platform kurang dari tiga kali, disaring keluar dari sampel.

sampel masukan yang terkandung 3966 kasus: 1757 perusahaan, yang berhenti partisipasi mereka dalam jangka waktu yang sangat singkat, dan 2209 perusahaan, yang terus berpartisipasi dalam tender.

Sampel dibagi menjadi sampel pelatihan dan verifikasi menggunakan metode stratifikasi sehubungan dengan variabel target dengan rasio 70/30.

Sejak pernyataan tugas yang terlibat memecahkan masalah klasifikasi, diusulkan untuk menggunakan metode seperti analisis data intelektual sebagai jaringan saraf, pohon keputusan, regresi logistik dan klasifikasi Bayesian [3 - 5]. Dalam rangka untuk memilih model terbaik, kriteria berikut dapat digunakan [6]:

1. MSE - rata kuadrat error:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

dimana  $y$   $\in$  dependen nilai-nilai variabel, diperkirakan dengan menggunakan model matematika yang dibangun;

$y$

nilai yang sebenarnya dari variabel dependen;

2) jumlah kesalahan kuadrat (SSE):

$$\sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (2);$$

3) kriteria informasi Akaike ini:

$$AIC = -2 \ln \left( \frac{1}{N} \sum_{k=1}^N \frac{1}{\sigma^2} (y_k - \hat{y}_k)^2 \right) + 2p \quad (3);$$

4) Bayesian - kriteria Schwarz

$$BIC = -2 \ln \left( \frac{1}{N} \sum_{k=1}^N \frac{1}{\sigma^2} (y_k - \hat{y}_k)^2 \right) + p \ln(N) \quad (4),$$

dimana  $p$  - jumlah parameter model, diestimasi dengan menggunakan data statistik ( $p$  - nomor

parameter dari bagian auto-regresi dari model;  $q$  - jumlah parameter rata-rata bergerak; "1" muncul ketika offset diperkirakan (atau persimpangan, yaitu

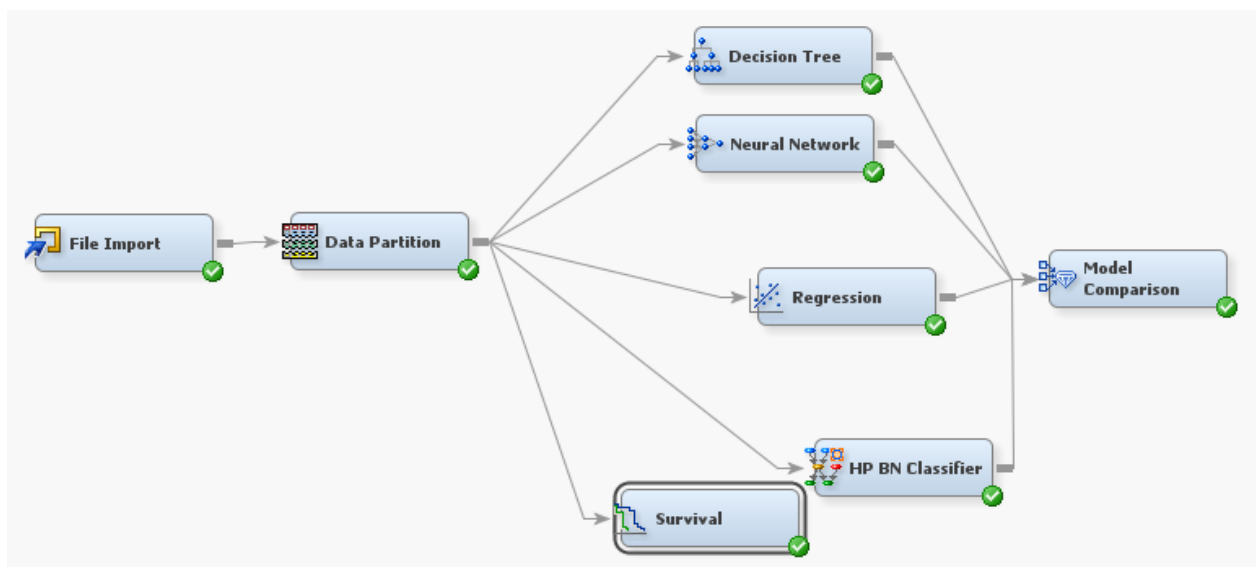
sebuah - contoh

panjangnya;

5) Kesalahan klasifikasi Rate dihitung sebagai rasio antara jumlah keliru diprediksi parameter dan jumlah total  $N$  nilai-nilai:

$$\text{Misclassification Rate} = \frac{\text{prediksi keliru yang nomor parameter}}{N} \quad (5).$$

Pembangunan model dijelaskan dan perhitungan kriteria statistik dilakukan berdasarkan pada teknologi informasi Perusahaan Miner [7]. Pemilihan model terbaik yang bisa dilakukan secara otomatis atau berdasarkan kriteria yang telah ditentukan kualitas. Urutan analisis tender pengadaan disajikan pada Gambar. 1.



Gambar. 1. Urutan analisis tender pengadaan berdasarkan pada teknologi informasi SAS Enterprise Miner

### Jaringan syaraf

Kami membangun berbagai jenis jaringan saraf [5] dengan nomor yang berbeda dari lapisan, fungsi aktivasi yang berbeda, dll kriteria Akaike ini terpilih sebagai kriteria kualitas (agar tidak membuat model terlalu

kompleks dan untuk mencapai keseimbangan antara parameter model dan kualitas). Sebuah sederhana perceptron jaringan saraf dengan 20 lapisan tersembunyi dan standarisasi input, didasarkan pada penyimpangan, fungsi mengkombinasikan radial, fungsi aktivasi logistik dan kriteria pelatihan kesalahan kesalahan klasifikasi ternyata menjadi model terbaik untuk data masukan. kriteria kualitas statistik terbaik jaringan saraf diberikan pada Tabel 1.

Tabel 1

**karakteristik statistik yang terbaik jaringan saraf**

| Target    | Fit Statistik   | statistik Label                   | Melatih  | pengesahan |
|-----------|-----------------|-----------------------------------|----------|------------|
| churn_out | _DFT_           | Total Derajat Kebebasan           | 2774     | -          |
| churn_out | _DFE_           | Derajat Kebebasan untuk Kesalahan | 2743     | -          |
| churn_out | _DFM_           | Derajat model Freedom             | 31       | -          |
| churn_out | _NW_            | Jumlah Berat Perkiraan            | 31       | -          |
| churn_out | _AIC_           | Kriteria Informasi Akaike ini     | 3227.794 | -          |
| churn_out | _SBC_           | Bayesian Criterion Schwarz        | 3411.564 | -          |
| churn_out | _ASE_           | Rata-rata Kesalahan Squared       | 0.197105 | 0.193835   |
| churn_out | _MAX_           | Kesalahan maksimum Absolute       | 0.986715 | 0.984846   |
| churn_out | _DIV_           | Pembagi untuk ASE                 | 5548     | 2384       |
| churn_out | _NOBS_          | Sum Frekuensi                     | 2774     | 1192       |
| churn_out | _MERUNTUHKAN_   | Kesalahan akar rata-rata Squared  | 0.443965 | 0.440267   |
| churn_out | _SSE_           | Sum of Squared Kesalahan          | 1093.537 | 462.1033   |
| churn_out | _SUMW_          | Sum Berat Kasus Kali Freq         | 5548     | 2384       |
| churn_out | _FPE_           | Akhir Prediksi Kesalahan          | 0,20156  | -          |
| churn_out | _MSE_           | Berarti Kesalahan Squared         | 0.199332 | 0.193835   |
| churn_out | _RFPE_          | Akar Akhir Prediksi Kesalahan     | 0.448954 | -          |
| churn_out | _RMSE_          | Kesalahan Root Mean Squared       | 0.446467 | 0.440267   |
| churn_out | _AVERR_         | Rata-rata Kesalahan Fungsi        | 0.570619 | 0.567532   |
| churn_out | _BERBUAT SALAH_ | kesalahan Fungsi                  | 3165.794 | 1352.996   |
| churn_out | _MISC_          | kesalahan klasifikasi Tingkat     | 0.317231 | 0.305369   |
| churn_out | _SALAH_         | Jumlah Klasifikasi Salah          | 880      | 364        |

*model regresi*

Sejak masalah klasifikasi diselesaikan, sebagai model berikutnya yang cocok untuk menganalisis apakah perusahaan akan menghentikan partisipasi dalam tender berikutnya (output biner: 0 - "tidak" atau 1 - "gencatan regresi") logistik dipilih, berdasarkan metode konstruksi model stepwise dengan berpasangan memasukkan dan keluaran dari karakteristik dari model. Untuk model ini karakteristik sebagai berikut diperoleh:

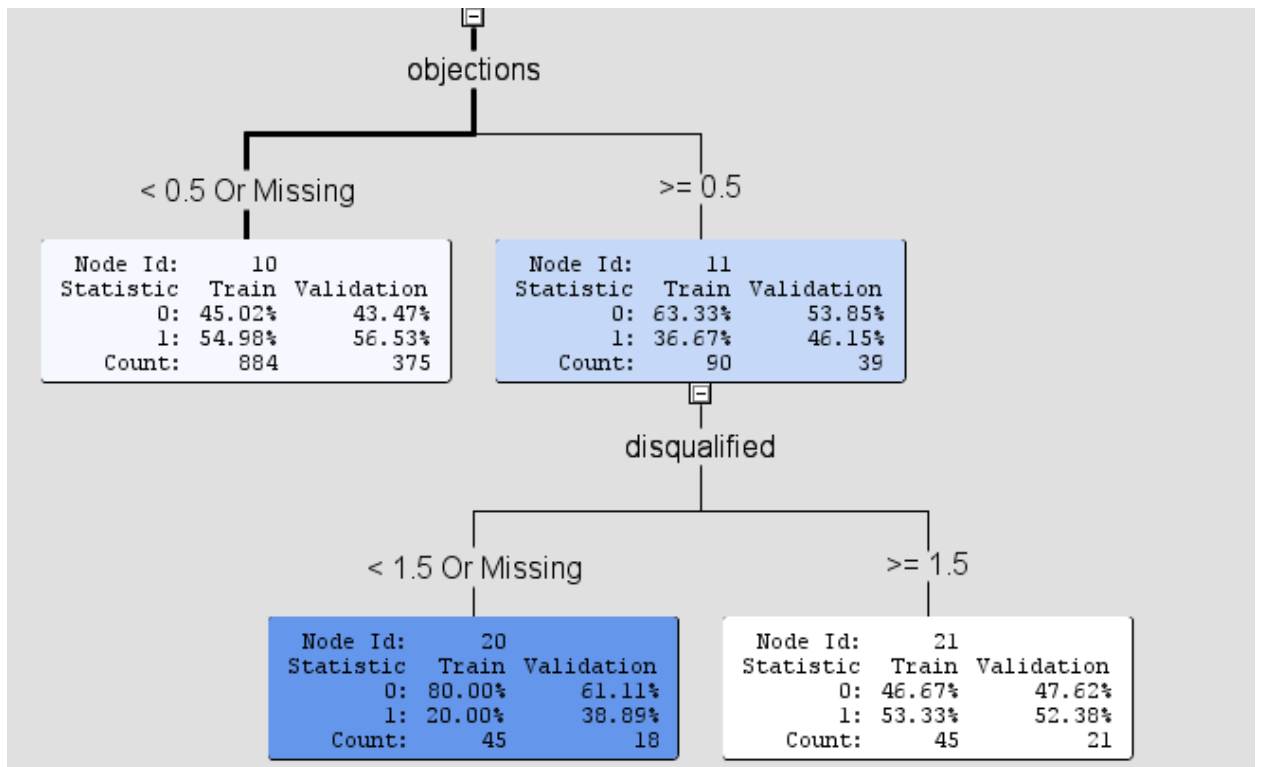
- 3301 AIC dan  $e$  1068,76310

*pohon keputusan*

Kemudian simulasi dilakukan, berdasarkan metode pohon keputusan [3], dengan pengaturan yang berbeda dari aturan cut-off, jumlah keturunan dan algoritma pembentukan sub-pohon. Pohon dengan tingkat kesalahan klasifikasi minimum ternyata menjadi yang terbaik:

$e$  1068,76310

Struktur pohon ditunjukkan pada Gambar. 2.



Gambar. 2. Struktur pohon keputusan dibangun

### *Naif classifier Bayesian*

Dalam aplikasi langsung dari classifier Bayesian naif tanpa pengaturan tambahan, persentase misclassifications adalah sekitar 50%, yang tidak memungkinkan untuk menggunakan model tersebut untuk analisis perusahaan. Setelah penyesuaian dan meningkatkan jumlah partisi, hasil agak lebih baik diperoleh (Tabel 2), tapi masih model tersebut tidak dianjurkan untuk digunakan dalam praktek.

Meja 2

**kriteria statistik untuk classifier Bayesian naif**

| Fit Statistik | statistik Label                  | Melatih  | pengesahan |
|---------------|----------------------------------|----------|------------|
| _ASE_         | Rata-rata Kesalahan Squared      | 0.240782 | 2.41E-01   |
| _DIV_         | Pembagi untuk ASE                | 5548     | 2384       |
| _MAX_         | Kesalahan maksimum Absolute      | 0.849634 | 8.18E-01   |
| _NOBS_        | Sum Frekuensi                    | 2774     | 1192       |
| _MERUNTUHKAN_ | Kesalahan akar rata-rata Squared | 0.490695 | 4.91E-01   |
| _SSE_         | Sum of Squared Kesalahan         | 1335.859 | 575.7147   |
| _DISF_        | Frekuensi Kasus Baris            | 2774     | 1192       |
| _MISC_        | kesalahan klasifikasi Tingkat    | 0.460707 | 0.463926   |
| _SALAH_       | Jumlah Klasifikasi Salah         | 1278     | 553        |

### *analisis perbandingan hasil dan pemilihan model terbaik*

Sejak masalah klasifikasi dipecahkan, pilihan model terbaik dilakukan berdasarkan kriteria jumlah contoh kesalahan klasifikasi, pada sampel validasi. Hal ini disebabkan kekhususan metode tertentu dan kecenderungan untuk beradaptasi dengan sampel pelatihan dan, karena itu, perbandingan dengan penggunaan sampel validasi lebih dibenarkan. Hasil simulasi disajikan dalam

Tabel 3.

tabel 3

Hasil klasifikasi dengan metode yang berbeda

| Model                    | kesalahan klasifikasi Tingkat |
|--------------------------|-------------------------------|
| Jaringan syaraf          | 0.305369                      |
| Regresi logistik         | 0.307047                      |
| pohon keputusan          | 0.313758                      |
| Naif classifier Bayesian | 0.463926                      |

Dengan demikian, jaringan saraf, yang memungkinkan untuk memprediksi dengan akurasi 70% apakah perusahaan akan terus berpartisipasi dalam perdagangan umum, ternyata menjadi model terbaik untuk menganalisis data dari sistem ProZorro. Tidak adanya mekanisme yang efektif untuk menghilangkan peserta yang tidak adil dari sistem perdagangan dikonfirmasi dan dua kelompok potensi perusahaan-hantu - satu hari dan permanen - ditemukan.

### kesimpulan

Analisis pembelian lembut dan memastikan kemungkinan untuk berpartisipasi dalam perdagangan untuk semua peserta terlepas dari hubungannya dengan badan-badan negara tertentu atau kolusi akan memungkinkan mencapai terobosan signifikan dalam memerangi korupsi, akan menjadi stimulus yang kuat untuk perusahaan kecil - produsen Ukraina - untuk berpartisipasi dalam tender dan memungkinkan menerima pendapatan tambahan untuk anggaran Ukraina. Model yang dibangun memungkinkan untuk mengklasifikasikan aplikasi lembut seperti dalam sistem, yang berisi karakteristik kolusi, untuk mendeteksi perusahaan yang mencurigakan

- peserta perdagangan, yang hanya boneka sehingga perdagangan bisa terjadi. Komplikasi dalam konstruksi dan klasifikasi dan, masing-masing, dalam memperoleh hasil yang akurat adalah karena ketidakmampuan untuk menentukan apakah perusahaan yang mencurigakan yang nyata karena mereka tidak dikecualikan dari partisipasi dalam perdagangan, yaitu model klasifikasi yang ada di platform yang on-line tidak mendeteksi mereka dan diklasifikasikan sebagai peserta nyata perdagangan. Namun, hasil analisis dan aplikasi dari yang sesuai asosiasi publik, media massa bisa menjadi alasan untuk pemantauan dan pemeriksaan tambahan untuk mengkonfirmasi atau menolak kecurigaan tersebut. Analisis sistem itu sendiri berguna dalam hal mendapatkan informasi statistik untuk rata-rata jumlah peserta dalam perdagangan, efektivitas pembatasan normatif pada akses ke perdagangan situs, kemudahan dan transparansi pengadaan publik. Penelitian lebih lanjut akan bertujuan penyempurnaan dari analisis yang diusulkan dengan membangun model perilaku untuk memprediksi perilaku peserta nyata perdagangan dan mendeteksi peserta perdagangan ilegal dan untypical.

### REFERENSI

1. Hukum Ukraina "Tentang Pengadaan Publik" [Elektronik sumber daya] / Verkhovna Rada Ukraina. - Modus akses: <http://zakon2.rada.gov.ua/laws/show/922-19>. (UKR).
2. ProZorro: pengadaan publik [sumber daya Elektronik] / Mode akses: <https://prozorro.gov.ua/>. (UKR).
3. Chubukova IA Data Mining / Chubukova IA - M.: Binom LBZ, 2008. - 384 p. (Rus).
4. Bruno GR Compendium Ramping. Pengantar modern Teori Manufacturing / Bruno GR - Springer International Publishing, 2018.
5. Zaichenko YP Dasar-dasar Intelektual Desain Sistem / Zaichenko YP - K.: Slovo, 2006. - 352 p. (UKR).
6. Bidiuk PI Analisis time series / PI Bidiuk, VD Romanenko, OL Timoshchuk. - Kyiv: Polytechnica, 2013. - 600 p. (Rus).
7. Kuznietsova NV Teknologi Informasi Pengolahan Data dan Analisis Manajemen Risiko Keuangan / NV Kuznietsova // Teknologi Informasi dan Keamanan Khusus. - IPRI, 2015. - №1. - P. 86 - 98. (UKR).

kantor redaksi menerima kertas 2018/03/06. kertas  
ditinjau 2018/03/12.

***Kuznietsova Nataliya*** - Cand. Sc. (Eng.), Ass. Prof. Departemen Metode Matematika Sistem

Analisis, e-mail: natalia-kpi@ukr.net.

Institut Analisis Sistem Terapan dari Universitas Teknik Nasional Ukraina "Igor Sikorsky Kyiv Polytechnic Institute".