**IDENTIFICAREA DATELOR ABERANTE PROVENITE DE LA CONTOARELE INTELIGENTE PRINTR-O METODĂ STATISTICĂ BAZATĂ PE “DATA MINING”**

 **Autori: stud. Maria Relea**

**Drd. ing. Claudia Bondar**

**Coordornator: Prof. univ. dr. ing. Gheorghe Grigoraș**

**Universitatea Tehnică „Gheorghe Asachi” din Iasi**

**Facultatea de Inginerie Electrică, Energetică si Informatică Aplicată**

**Departamentul de Energetică**

**Rezumat.** Lucrarea abordează o problemă de mare actualitate, respectiv evaluarea eficienței sau suficienței energetice. Într-o interpretare mai simplă ne referim la calitatea datelor provenite de la sistemele de contorizare, a sistmelor SCADA sau de monitorizare integrate în nodurile rețelelelor electrice de distribuție de joasă tensiune. Pentru a identifica și valida modalitatea coerectă de calcul/estimare a consumului de energie electrică trebuie lucrat cu baze de date mari aferente posturilor de transformare. Pentru a rezolva această problemă din ce în ce mai stringentă tehnicile de explorare a datelor au răspuns foarte bine, în special conceptul de „data mining”. Analiza probabilistică împreună cu posibilitățile oferite de inteligența artificială conduc la modele matematice simple, dar cu încărcătură științifică deosebită, de determinare a curbelor de sarcină care prezintă consumuri neconforme, aleatoare sau anormale (outliers). În acest context, în lucrare se propune o metodă statistică originală care are la bază procesul de data mining sau de extragere a informațiilor provenite de la contoarele inteligente. Astfel, rezultatele au demonstrat faptul că prin eliminarea datelor anormale un operator de distribuție poate evalua în mod pragmatic principalele caracteristici ale curbelor de sarcină, cu ajutorul cărora să determine pierderile de energie..

**1. INTRODUCERE**

 Modelarea consumului de energie electrică este o acţiune căreia trebuie să i se acorde o deosebită şi continuă atenţie, întrucât constituie un instrument de bază pentru fundamentarea strategiilor de dezvoltare a infrastructurii companiilor de distribuţie și a optimizării funcţionării reţelelor de alimentare cu energie electrică. Caracteristicile sarcinilor electrice constituie elementele de bază în scopul rezolvării problemelor tehnico-economice complexe legate de analiza rețelelor electrice de distribuție care alimentează diferite categorii de consumatori. Într-o altă ordine de idei, ca definiție o *informație anormală* sau o *valoare aberantă* dintr-un set de date este observată ca fiind foarte diferită faţă de restul valorilor şi, de aceea, se poate considera ca fiind generată pur întâmplător sau în urma apariţeii unui regim scurt necorespunzător, respectiv o nefuncţionare corespunzătoare a contorului inteligent. Valorile aberante (extreme) pot influența esențial rezultatele, deoarece sunt generate de factori neobișnuiți. Distribuția normală este cea mai importantă distribuție continuă, deoarece în practică multe variabile aleatoare sunt variabile aleatoare normale, sunt aproximativ variabile aleatoare normale, sau pot fi transformate în variabile aleatoare normale [1]. Trebuie subliniat faptul că dacă valorile aberante nu sunt eliminate, acestea participă la valoarea medie de calcul a indicatorilor curbelor de sarcină care, ulterior vor fi folosite la determinarea consumului propriu tehnologic (pierderi de energie). Problema aceasta s-a accentuat în ultima perioadă din cauza citirii datelor de saricină de la distanță fie de la contoare inteligente, fie direct din sistmele SCADA [2].

 Planificarea dezvoltării unei rețele de alimentare cu energie electrică are la bază predeterminarea sarcinii electrice, informații ce pot fi extrase dintr-o bază de date care are un număr impresionant de valori efective provenite de la contoarele inteligente [3]. Noua era a eficienței și suficienței energetice creează noi provocări sociale și tehnice pentru rețeaua electrică modernă. Interacțiunea și compatibilitatea dintre informațiile de sarcină în timp real și rețeaua este un subiect care necesită cercetare și dezvoltare. Există studii cu metode tradiționale privind modul în care datele aberante sunt identificate [4]. În literatura de specialitate găsim cercetări limitate despre deazavantajele calculelor de proiectare și planificare de rețea având valori aberante, însă găsim o mulțime de studii asupra modului în care idnetificarea acestora se poate realiza. În acest context, metode bazate pe inteligenta artificială sunt utilizate respectiv algorimii genetici în [1], tehnicile de clustering în [5], tehnicile fuzzy și rețelele neuronale artificiale în [6] și algortimul roiului de particule în [7]. Toate metodele utilizate pentru identificarea și eliminarea datelor aberante prin procedee complexe.

 Analiza probabilistică împreună cu posibilitățile oferite de inteligența artificială conduc la modele matematice simple, dar cu încărcătură științifică deosebită, de determinare a curbelor de sarcină care prezintă consumuri neconforme, aleatoare sau anormale (outliers). În acest context, în lucrare se propune o metodă statistică care are la bază procesul de data mining sau de extragere a informațiilor provenite de la contoarele inteligente. Rezultatele au demonstrat faptul că prin eliminarea datelor anormale un operator de distribuție poate evalua în mod pragmatic principalele caracteristici ale curbelor de sarcină, cu ajutorul cărora să determine pierderile de energie. În continuare, lucrarea este structurată astfel: stadiul actial al problematicii abordate în capitolul 2, în următorul capitol se propune o metodă statistică de identificare a valorilor improprii (outliers), prezentând și modalitatea de calcul a pierderilor de energie la nivelul transformatoarelor de putere. În final este realizat un studiu de caz, care să valideze modelul matematic.

**2. CONCEPTUL DE EXPLORARE A DATELOR „DATA MINING”**

Explorarea datelor (engl. Data Mining) are, ca și alte concepte, mai multe definiţii. În esenţă, acestea converg spre ideea unui proces de extragere de informații noi din colecțiile de date existente. Termenul de dată este utilizat cu semnificaţia de descriere a unui eveniment precis, produs în lumea reală şi verificabil prin raportare la aceasta. Informaţia constituie descrierea unei categorii abstracte, ce acoperă mai multe evenimente sau exemple concrete. Principiul de funcționare în data mining este următorul: se prelucrează datele referitoare la perioadele trecute, examinând o varietate de situaţii care s-au produs şi ale căror rezultate sau consecinţe sunt deci, bine cunoscute, pentru a evidenţia caracteristicile acestora şi a permite elaborarea unui model. Odată construit, modelul poate fi aplicat situaţiilor noi de acelaşi tip. Deoarece calculatoarele au permis posibilitatea stocării unor cantități mari de date, este normal să se recurgă la tehnici moderne de prelucrare pentru a descoperi tipare/modele caracteristice. În timpul acestor procese, abordările folosite pentru a reprezenta datele și a măsura similitudinea dintre ele în scopul obținerii de tipare/modele, pot fi confuze în ceea ce privește alegerea celor mai potrivite tehnici. În aceste condiții, este necesară automatizarea explorării bazelor de informații (cunoștințe), pentru a ajuta operatorul uman în luarea deciziilor privind exploatarea optimală a rețelelor electrice de distribuție. Etapele ce trebuie parcurse în procesul de „Data mining” sunt indicate în Fig. 1 și prezentate sintetic în continuare [8], [9]:

* Stabilirea elementelor supuse procesului de explorare a datelor – constă în alegerea din baza de date a elementelor care corespund cel mai bine obiectivului problemei ce trebuie rezolvată.
* Extragerea atributelor elementelor supuse procesului de explorare a datelor – constă în identificarea celor mai utile și reprezentative atribute ale elementelor supuse procesului de explorare a datelor; în această etapă pot fi efectuate una sau mai multe transformări ale elementelor pentru obâinerea de noi caracteristici dominante.
* Definirea unei măsuri de similitudine – constă, în mod obișnuit, în calcularea distanței între perechi de elemente, care să reflecte similitudinea dintre două elemente; cea mai utilizată fiind distanța euclidiană.
* Procesul de „data minig” – utilizând diferite metode, care în final să conducă la obținerea unui set de informații suficient pentru orice set de date de intrare. În cadrul studiilor efectuate, s-a utilizat metoda K-Medii [8].
* Extragerea rezultatelor – constă într-o descriere succintă a fiecărui set de informații prin intermediul unor elemente reprezentative.
* Evaluarea rezultatelor – constă într-o evaluare a procesului de explorare a datelor și utilizează de obicei un criteriu de optimizare. Ca și în cazulul procesului de grupare, (clustering), această analiză este obiectivă și determină dacă rezultatele obținute sunt corecte. Un cluster este validat dacă nu apare accidental.



Fig. 1.Etapele procesului de explorare a datelor, [10].

**3. METODĂ STATISTICĂ DE IDENTIFICARE A VALORILOR IMPROPRII (OUTLIERS)**

Metoda statistică propusă pornește de la legea normală de distribuție - distribuția Gauss, iar apoi se ține seama de ipotezele implicite, adică informațiile neaberante sunt caracterizate de primele două momente statistice [-2σ, 2σ]. Densitatea probabilă a datelor normale se poate estima printr-o simplă observare asupra Fig. 2 [9].



Fig. 2.Legea normală de distribuție Gauss

În cazul curbei ideale sau normale de variație, valoarea reală și corectă se află în intervalul [-σ, σ] acoperind 68,26 % din totalul valorilor unei serii de variaţie. Între medie şi 2σ sunt cuprinse 95,45 % din totalul valorilor seriei, iar între medie şi 3σ se găsesc 99,73 % din valorile serie statistice. În calculele obişnuite, nu este nevoie de o exactitate mai mare în măsurarea abaterii decât de 3σ, căreia îi corespunde, într-o repartiţie normală, aproape totalitatea cazurilor, cu excepţia a 0,27 % dintre acestea, ceea ce este neglijabil. Chiar dacă am calcula abaterea de la medie cu 4σ, nu am obţine o exactitate cu mult mai mare: 4σ = 99,993 %, iar 5σ = 99,999 %. În cercetările experimentale, care necesită o exactitate ridicată, se calculează deviaţia dublă sau triplă. În cercetările clinice este suficientă calcularea deviaţiei duble, care cuprinde, cum s-a arătat, 95,45 % din totalul valorilor seriei statistice.

Deviaţia standard caracterizează cel mai corect gradul de dispersie faţă de media aritmetică, deoarece ţine seama de toate observaţiile repartiţiei de frecvenţe şi, totodată, este afectată mult mai puţin de fluctuaţiile de selecţie, comparativ cu celelalte măsurători ale gradului de dispersie. Un aspect practic ce trebuie luat în considerare este faptul că, atunci când populaţia din care a fost extras eşantionul are o distribuţie aproximativ normală (gaussiană), atunci deviaţia standard oferă o interpretare utilă a datelor din punctul de vedere al probabilităţilor. Trebuie remarcat faptul că, într-o populaţie cu o distribuţie normală, media, mediana şi modulul sunt egale. Totodată, distribuţia normală constituie punctul de pornire pentru o serie de teste analitice. Distribuţia Gaussiană poate fi generată dintr-o bază de date mare cu măsurători, fiind perfect echilibrată dacă valoarea medie (mediana) se află chiar în mijloc. Acesta este adesea numită curba lui Gauss, deoarece graficul densității de probabilitate arată ca un clopot. Acest tip de distribuţie se întâlneşte adesea în problemele de măsurare a datelor, notate cu X.

Din punct de vedere matematic, o abatere standard este μ ± σ (68% din datele totale), în care *μ* este media aritmetică. Conform Fig. 2 deoarece două abateri standard conţin peste 95 % din informaţiile provenite din baza de date, valorile din afara acestui interval se consideră valori aberante (outliers).

Ţinând seama că abaterea standard, notată cu *σ* determinată cu relaţia:

 (1)

reprezintă, de fapt, distanţa specifică de la valoarea medie notată cu *μ*:

 (2)

valorile improprii sau outliers notate cu *Λ*, pot fi detectate utilizând o combinaţie a indicatorilor statistic calculaţicu relaţiile (1) şi (2), utilizând următoarea aproximaţie:

 (3)

Termenul ψ poate fi evaluat după o analiză atentă, conform [11].

Pentru determinarea pierderilor de energie activă în fiecare element de reţea de tip transformator de putere și pentru fiecare regim caracteristic se utilizează relaţiile următoare:

a) *în cazul încărcării constante pe durata regimului caracteristic*:

 (4)

unde: – CPT constant de energie electrică activă în transformator; – CPT variabil de energie electrică activă în transformator; D*P*o - pierderea de mers în gol a transformatorului; D*P*sc - pierderea în sarcină în înfăşurările transformatorului; *Sn* - puterea aparentă nominală a transformatorului; *T* – durata regimului caracteristic; *P, Q* – puterea active și reactivă în regimul caracteristic.

b) *în cazul încărcării conform curbei de sarcină*:

 (5)

unde: *Smax*– puterea aparentă maximă calculată pentru curba de sarcină; *τ* – timpul de pierderi determinat pentru curba de sarcină.

Pierderile de putere și energie totale aferente rețelei electrice sau zonei de rețea delimitate de conturul propus pentru o perioadă de calcul se determină prin însumarea pierderilor determinate pentru fiecare element de rețea inclus în contur, obținut totodată prin însumarea pierderilor de putere activă aferente fiecărui regim caracteristic stabilit, multiplicat cu numărul de repetări ale fiecărui regim caracteristic în cursul perioadei de calcul.

**4. STUDIU DE CAZ**

În scopul realizării unei analize privind variațiile sarcinilor electrice din nodurile de joasă tensiune ale transformatoarelor de putere rurale care alimentează un număr semnificativ de consumatori, s-a realizat un studiu de caz în care sunt indicate principalele valori care se abat de la componenta de tendinţă a curbelor zilnice de sarcină ale posturilor de transformare, subliniind astfel influenţa acestora asupra calculelor privind pierderile de putere şi energie activă prin transformatoarele de putere. Având la dispoziție o bază de date cu înregistrările privind curbele de sarcină provenite de la contoarele inteligente implementate de către operatorul de distribuţie din zona Moldovei, în 298 de posturi de transformare rurale, pentru efectuarea analizei propuse au fost considerate curbele de sarcină aparentă (kVA) pentru o săptămână din regimul rece de iarnă (22 – 28 ianuarie 2024), iar analiza este efectuată pentru ziua standard de miercuri 24 ianuarie 2024. Studiul de caz fiind realizat pentru rețelele electrice de distribuție de joasă tensiune aferente celor șase județe (Iași, Neamț, Botoșani, Vaslui, Suceava și Bacău). Caracteristicile nominale ale transformatoarelor de putere necesare calculului pierderilor de putere sunt prezentate în *Tab. 1*.

Tab. 1 Caracteristicile transformatoarelor de putere analizate

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nr. crt. | Numărul de transformatore | Putere aparentă[kVA] | Pierderi active | Rezistența[Ω] |
| ΔP0 [kW] | ΔPsc[kW] |
| 1 | 9 | 40 | 0.230 | 1.0 | 0.100 |
| 2 | 47 | 63 | 0.300 | 1.5 | 0.060 |
| 3 | 84 | 100 | 0.350 | 2.3 | 0.037 |
| 4 | 92 | 160 | 0.525 | 3.1 | 0.019 |
| 5 | 56 | 250 | 0.680 | 4.4 | 0.011 |
| 6 | 9 | 400 | 0.940 | 6.0 | 0.006 |
| 7 | 1 | 630 | 1.250 | 8.2 | 0.003 |

Toate posturile de transformare erau echipate cu contoare inteligente care pot înregistra curbele de sarcină, tensiunile pe cele trei faze, curenții pe cele trei faze și pe nul, factorul de putere etc., la nivelul barelor de joasă tensiune. Măsurătorile efectuate pentru cele 298 de posturi de transformare rurale au fost efectuate pentru o săptămână cu un pas de eșantionare de 10 minute. Pentru extragerea caracteristicilor curbelor de sarcină a fost utilizat un proces de data mining, care avea la bază metodologia de calcul indicată în §4. Pentru a evidenția datele anormale, care au fost propuse pentru eliminare, în Fig. 3 sunt indicate curbele de sarcină pentru trei posturi de transformare: PTA 5 Cabești, PTA Șut Onești și PTA 1 Dragomir, pentru ziua de 24 ianuarie 2024. Pentru baza de date avută la dispoziție, în *Tab. 2* sunt indicate numărul datelor anormale reieșite din utilizarea metodologiei indicate în §4, pentru o 30 posturi de transformare analizate. În *Tab. 3.* au fost indicaţi principalii indicatori ai curbelor zilnice de sarcină, valorile anormale sub formă de limită minimă şi maximă, numărul acestora şi abaterea standard şi dispersia.

Valoarea paramentrului ψ din (3) este 2, determinat în urma unui proces statistic asupra bazei de date [10]. În *Fig. 4* sunt indicate sub formă grafică valorile tuturor indicatorilor amintiţi anterior, în care prin linia întreruptă se delimitează limitele valorilor fiecărui indicator în parte, iar prin culoarea roşie valorile fiecărei date anormale, calculate cu ajutorul relaţiei (3) care utilizează indicatorii statistici.

 

Fig. 3. Curbele de sarcină înregistrate în trei posturi de transformare

Tab. 2 Numărul de date aberante (outliers) la nivelul curbelor de sarcină ale PT

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nr.****Crt.** | **Nume PT** | **S****[kVA]** | **Outlier** | **Nr.****Crt.** | **Nume PT** | **S****[kVA]** | **Outlier** | **Nr.****Crt.** | **Nume PT** | **S****[kVA]** | **Outlier** |
| 1 | PT1 Balesti | 160 | 2 | 11 | PTA1 Padureni | 100 | 1 | 21 | PT1 Mitesti | 63 | 1 |
| 2 | PT1 Baltateni | 160 | 11 | 12 | PTA1 Zaicesti | 160 | 21 | 22 | PT1 Negresti | 250 | 29 |
| 3 | PT1 Balteni | 250 | 6 | 13 | PTA2 Supitica | 100 | 57 | 23 | PT1 Pic Lupului | 100 | 39 |
| 4 | PT1 Cepesti | 160 | 25 | 14 | PT1 Marasesti | 100 | 16 | 24 | PT1 Prajast | 250 | 5 |
| 5 | PT1 Cetatuia | 100 | 17 | 15 | PT1 Plotonesti | 100 | 39 | 25 | PT1 Sar Dobreni | 250 | 3 |
| 6 | PT1 Corodesti | 250 | 21 | 16 | PT1 Poi. Deleni | 100 | 48 | 26 | PT1 Scheia | 250 | 0 |
| 7 | PT1 Cristesti | 160 | 24 | 17 | PT1 Poiana | 63 | 13 | 27 | PT1 Scobinti | 100 | 4 |
| 8 | PT1 Iana | 160 | 1 | 18 | PT2 Draxeni | 63 | 16 | 28 | PT1 Slobozia | 100 | 30 |
| 9 | PT1 Iezere | 160 | 16 | 19 | PT5 Pastraveni | 400 | 12 | 29 | PT1 Soci | 160 | 31 |
| 10 | PT1 Girceni | 100 | 17 | 20 | PTA2 Catamarasti | 160 | 12 | 30 | PT1 Traian | 250 | 58 |

Tab. 3 Detectarea valorilor anormale pentru o serie de indicatori ai curbelor de sarcină

|  |  |
| --- | --- |
| Caracteristici curbe sarcină | Date aberante (outliers), *Λ* |
| min | max | No. | µ | σ |
| DSM | 3.142 | 18.729 | 12 | 10.936 | 3.897 |
| FS | 0.154 | 0.796 | 13 | 0.505 | 0.176 |
| DP | 0.578 | 13.908 | 13 | 6.665 | 3.621 |
| FP | 0.021 | 0.637 | 6 | 0.308 | 0.165 |
| FN | 0.009 | 0.611 | 13 | 0.301 | 0.155 |
| CU | 0.154 | 0.856 | 12 | 0.505 | 0.175 |

Prin eliminarea datelor anormale şi în urma procesului de curăţare (cleansing), opt posturi de transformare au fost îndepărtate fiind utilizate doar 290 de posturi. În continuare analiza a fost realizată pe grupe de transformatoare. Prin utilizarea metodologiei propuse, s-a impus și o analiză a posturilor de transformare în funcție de puterile lor nominale. Astfel, baza de date a fost împărțită în șapte categorii de transformatoare, astfel: 40 kVA (9), 63 kVA (45), 100 kVA (80), 160 kVA (90), 250 kVA (56), 400 kVA (9) și 630 kVA (1).

****

Fig. 4 Valorile indicatorilor caracteristici pentru cele 290 de posturi de transformare

Pentru a valida o acuratețe a valorilor indicatorilor din baza de date, în *Tab. 4* sunt indicate rezultatele (fără valori anormale) pentru toate categoriile de consumatori. De exemplu, în Fig. 5sunt indicate valorile medii ale puterii pentru toate categoriile de transformatoare, cu și fără date anormale (outliers). Aceste valori au pot fi utilizate la efectuarea postcalcul a pierderilor de energie electrică în transformatoarele de putere de către operatorii de distribuție, pentru o planificare corespunzătoare a rețelelor electrice de distribuție.

Tab. 4 Rezultatele statistice medii ale indicatorilor analizați pe categorii de transformatoare

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Indicatori  | DSM | FS | DP | FP | FN | CU |
| 40 kVA | µ | 16.176 | 0.674 | 12.445 | 0.519 | 0.512 | 0.744 |
| σ | 1.424 | 0.059 | 1.318 | 0.055 | 0.066 | 0.031 |
| 63 kVA | µ | 15.023 | 0.626 | 11.794 | 0.491 | 0.510 | 0.743 |
| σ | 2.457 | 0.102 | 2.432 | 0.101 | 0.203 | 0.121 |
| 100 kVA | µ | 14.765 | 0.615 | 11.582 | 0.483 | 0.519 | 0.747 |
| σ | 2.362 | 0.098 | 2.332 | 0.097 | 0.187 | 0.110 |
| 160 kVA | µ | 15.001 | 0.625 | 11.748 | 0.489 | 0.519 | 0.749 |
| σ | 2.705 | 0.113 | 2.641 | 0.110 | 0.174 | 0.099 |
| 250 kVA | µ | 15.041 | 0.627 | 11.824 | 0.493 | 0.528 | 0.752 |
| σ | 2.164 | 0.090 | 1.916 | 0.080 | 0.172 | 0.099 |
| 400 kVA | µ | 15.907 | 0.663 | 12.120 | 0.505 | 0.472 | 0.731 |
| σ | 1.444 | 0.060 | 1.540 | 0.064 | 0.145 | 0.078 |
| 630 kVA | µ | 17.271 | 0.720 | 13.237 | 0.552 | 0.469 | 0.740 |



Fig. 5 Valorile medii ale sarcinii aparente cu sau fără date anormale (outliers)

De asemenea, pentru a evidenția influența acestor valori anormale (outliers) asupra pierderilor de energie electrică, în Fig. 6sunt indicate pierderile de energie pentru câte două transformatoare de putere din cele șapte categorii, calculate cu ajutorul relației (5), dar a căror indicatori au fost determinați în urma metodologiei propuse, respectiv cazul de bază și cel în care au fost eliminate toate informațiile anormale provenite de la contoarele inteligente.



Fig. 6 Valorile pierderilor de energie activă pentru două posturi de transformare de puteri nominale diferite, cu sau fără date anormale (outliers)

**5. CONCLUZII**

 În prezent din ce în ce mai mult efort este alocat reducerea cantității de energie utilizată, motiv pentru care devine interesant să detectăm automat un consum „exterior” de energie în timpul unei anumite zile. Capacitatea de a detecta anomalii într-un consum de zi cu zi este utilă pentru acele firme responsabile de producerea și distribuția energiei, deoarece momentan aceasta nu poate fi stocată. Pe de altă parte, cererea excesivă de energie într-un anumit moment al zilei poate deveni o problemă. În lucrare propunem o combinație a două modele nesupravegheate pentru a detecta datele aberante din curbele de sarcină. Scopul analizei nu este doar a datelor aberante, ci și identificarea „neregularităților” de calcul al indicatorilor care caracterizează curbele de sarcină, respectiv pierderile de putere. În acest context, s-a propus o metodă statistică de identificare a datelor aberante (outlier). Rezultatele arată că există erori majore de calcul a indicatorilor curbelor de sarcină care, la rândul lor, sunt utilizați în calcule de sarcină, estimări de consum, determinări de pierderi de energie etc. Instrumentul matematic propus poate fi un instrument eficient în strategiile de planificare ale rețelelor electrice de distribuție. Datorită rezultatelor obținute, in viitor ne propunem să dezvoltăm metoda pein implementarea algoritmilor specifici antrenării datelor multiple, respectiv tehnicile de clustering și evaluarea „datelor lipsă – valori nule”.

**6. BIBLIOGRAFIE**

1. Rabie, A.H., Ali, S.H., Saleh, A.I. and Ali, H.A., 2020. A new outlier rejection methodology for supporting load forecasting in smart grids based on big data. *Cluster Computing*, *23*, pp.509-535.
2. Mitra, S., Chakraborty, B. and Mitra, P., 2024. Smart meter data analytics applications for secure, reliable and robust grid system: Survey and future directions. *Energy*, *289*, p.129920.
3. Räsänen T., Voukantsis D., Niska H., Karatzas K., Kolehmainen M., *Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data*. Applied Energy, 87(11), 3538-3545, 2010.
4. Neagu B.C., Grigoraş G., Scarlatache F., Schreiner C., Ciobanu R., "Patterns discovery of load curves characteristics using clustering based data mining," 2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWER ENG), Cadiz, Spain, 2017, pp. 83-87, doi: 10.1109/CPE.2017.7915149.
5. Hurst, W., Montañez, C.A.C. and Shone, N., 2020. Time-pattern profiling from smart meter data to detect outliers in energy consumption. *IoT*, *1*(1), p.6.
6. de Andrade, P.H.M., Villanueva, J.M.M. and de Macêdo Braz, H.D., 2020. Outliers treatment in power curves using hybrid artificial intelligence technique. *International Journal of Innovative Computing, Information and Control*, *16*(2), pp.513-525.
7. Fang, J., Wang, Z., Liu, W., Lauria, S., Zeng, N., Prieto, C., Sikström, F. and Liu, X., 2022. A new particle swarm optimization algorithm for outlier detection: Industrial data clustering in wire arc additive manufacturing. *IEEE Transactions on Automation Science and Engineering*, *21*(2), pp.1244-1257.
8. Grigoraș G., *Contribuții privind utilizarea tehnicilor de inteligență artificială în supravegherea și conducerea sistemelor electroenergetice*, Teza de abilitare, Universitatea Tehnică Gheorghe Asachi, 2019, Iași.
9. Neagu B.C., Grigoraş G.,"*A Data-Mining-Based Methodology to Identify the Behavioural Characteristics of Prosumers within Active Distribution Networks*," 2020 International Symposium on Fundamentals of Electrical Engineering (ISFEE), Bucharest, Romania, 2020, pp. 1-6, doi: 10.1109/ISFEE51261.2020.9756166.
10. Neagu B.C., Grigoras G., Scarlatache F., "Influence of Outliers on Transformer Power Losses Estimation Using a Statistical Based Data Mining Approach," 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 2018, pp. 1-4, doi: 10.1109/ECAI.2018.8679002.
11. Neagu B.C., Grigoras G. "Data Mininig Tools in Electricity Distribution Systems." Acta Electr. J. 56.3 (2015): 209-212.