

## IZAZOVI KORIŠĆENJA VEŠTAČKE INTELIGENCIJE U OTKRIVANJU NETEHNIČKIH GUBITAKA

### CHALLENGES IN USING ARTIFICIAL INTELLIGENCE FOR DETECTING NON-TECHNICAL LOSSES

Miloš KOSTIĆ, Elektrodistribucija Srbije d.o.o. Beograd, Srbija  
Dejana STEFANOVIĆ-KOSTIĆ, Agencija za Energetiku Republike Srbije, Beograd, Srbija

#### KRATAK SADRŽAJ

ODS Elektrodistribucija Srbije se priprema za upotrebu veštačke inteligencije u delu poslovanja koji se odnosi na analizu informacija o potrošačima i mernim mestima. Prikazana je generalna metodologija koja se koristi za izbor algoritma veštačke inteligencije koji će se koristiti u procesu otkrivanja netehničkih gubitaka isporučene električne energije. U radu su prezentovani konkretni izazovi koji se uzimaju u obzir, a koji se odnose na specifičnosti primene veštačke inteligencije nad podacima koji su na raspolaganju. Date preporuke treba da obezbede visok učinak u operativnoj primeni tehnologija veštačke inteligencije u Elektrodistribuciji Srbije.

**Cljučne reči:** veštačka inteligencija, netehnički gubici, gubici električne energije

#### ABSTRACT

DSO Elektrodistribucija Srbije is exploring the use of artificial intelligence in the process of data analysis of consumer and meter data. The general methodology for choosing a specific artificial intelligence algorithm that will be used in the process of detection of non-technical losses out of total supplied electrical energy is shown. Concrete challenges that are being considered are presented in this paper, related to specifics of applying artificial intelligence on the available data. Recommendations for overcoming enumerated challenges should ensure better results in operative usage of artificial intelligence in Elektrodistribucija Srbije.

**Key words:** artificial intelligence, non-technical losses, losses of electrical energy

#### 1. UVOD

Elektrodistribucija Srbije se priprema za uvođenje alata koji koriste veštačku inteligenciju za identifikaciju potencijalnih netehničkih gubitaka. Pojam i značaj netehničkih gubitaka je dat u 2. poglavlju. U procesu pripreme se sprovodi istraživanje kojim se analiziraju moderne metodologije za selekciju konkretnih algoritama, koji bi se koristili u radu sa podacima. Najbolja praksa za selekciju je predstavljena u 3. poglavlju. Struktura i osobine podataka o potrošačima i mernim mestima postavljaju specifične izazove za primenu algoritma veštačke inteligencije. Ovi izazovi, i preporuke za njihovo prevazilaženje su dati u 4. poglavlju.

#### 2. NETEHNIČKI GUBICI ELEKTRIČNE ENERGIJE

Gubici u distributivnom elektroenergetskom sistemu predstavljaju razliku između preuzete električne energije na višem naponu (od transportnog sistema), i očitane električne energije kod potrošača na nižem naponu. Ova razlika se javlja usled više razloga, koji se mogu podeliti u dve grupe: tehnički i netehnički. Prema tome se i gubici dele na tehničke i netehničke gubitke.

Uzroci tehničkih gubitaka su generalno povezani sa elektromagnetnim karakteristikama sistema za distribuciju, te se mogu proceniti kao funkcija količine preuzete energije.

Osnovni uzroci netehničkih gubitaka su: 1) dotrajala infrastruktura 2) ilegalno korišćenje električne energije 3) umanjeno očitavanje usled manipulacije brojilima 4) neispravna brojila 5) nelegalno priključenje na vodove, i 6) manipulacije periodom očitavanja i tarifnim stavom[1].

Netehnički gubici se često označavaju kao „ekonomski“, ili „komercijalni“ gubici, čime se aludira na činjenicu da bi se adekvatnim poslovnim aktivnostima mogli smanjiti do zanemarljivog nivoa (za razliku od tehničkih gubitaka). U ovom radu će se koristiti izraz netehnički gubici (engl. Non-Technical Losses, skr. NTL), koji preovlađuje u svetskoj literaturi.

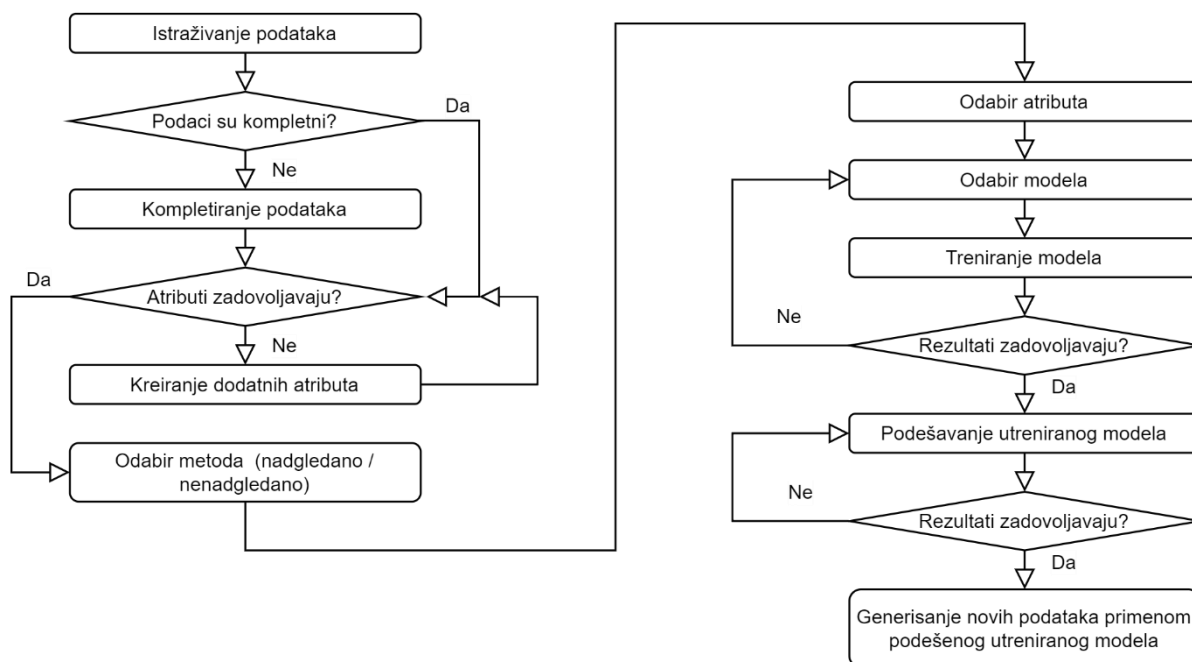
Netehnički gubici predstavljaju veliko finansijsko opterećenje za elektrodistributivna preduzeća. U 2022. godini Elektrodistribucija Srbije je za potrebe pokrivanja svih gubitaka preuzela preko 3.3 miliona MWh energije, što je preko 11% ukupne energije koja je isporučena potrošačima[2]. Ukoliko imamo u vidu da od ove količine manje od polovine ide na pokrivanje tehničkih gubitaka (tačan procenat zavisi od dela mreže), jasno je da je finansijski trošak Elektrodistribucije Srbije prouzrokovan netehničkim gubicima iznosio preko 120 miliona evra samo u toj godini (po prosečnoj ceni MWh od EUR 66).

Elektrodistributivna preduzeća se sistematski bore sa netehničkim gubicima. Jedan od elemenata ove poslovne aktivnosti je detekcija potrošača koji su potencijalno odgovorni za NTL. Za to se koriste različite metode, od tehničkih mera do analiza prijava potrošača. Nakon identifikacije potencijalnih NTL, sprovode se akcije kontrole mernih mesta, i na kraju se analiziraju rezultati. Fokus ovog rada je na analizi mogućnosti poboljšanja kvaliteta detekcije NTL korišćenjem algoritama veštačke inteligencije.

### 3. METODOLOGIJA ODABIRA MODELA MAŠINSKOG UČENJA

Oblast veštačke inteligencije koja se može upotrebiti za detekciju netehničkih gubitaka je mašinsko učenje (engl. machine learning, skr. ML). Primena ML se može podeliti u dve faze: priprema modela i upotreba modela u eksploataciji. U ovom radu ćemo se fokusirati na proces pripreme modela.

Tokom pripreme modela se bira model koji će se upotrebljavati u procesu ML. Metodologija odabira počinje analizom dostupnih podataka u smislu upotrebljivosti za neki od ML modela. Nakon odabira atributa iz podataka, treniraju se različiti modeli i proverava se njihov učinak. Na kraju se bira onaj model koji ima najbolje karakteristike. Uopštena metodologija pripreme alata za primenu ML je data na slici:



Slika 1 - Uopštena metodologija pripreme ML alata

#### 3.1 Istraživanje podataka

Tokom istraživanja podataka (engl. exploratory data analysis) se stiče slika o upotrebljivosti raspoloživih podataka, kao i o eventualnim koracima za dovođenje kvaliteta podataka na nivo koji zahtevaju modeli.

### 3.2 Kompletiranje podataka

Često se dešava da nekim kolonama nedostaju podaci. Dostupni ML algoritmi uglavnom ne dozvoljavaju da ulazni podaci nisu kompletni. U ovoj fazi se vrši popunjavanje nepostojećih podataka.

### 3.3 Odabir atributa

U ovoj fazi se određuju atributi (engl. feature) koji će se koristiti u samom ML modelu. Selekcija se vrši procenom korelacije sa ciljnim atributom (engl. target feature, label). Selekciju je moguće vršiti i nakon treniranja modela kada su dostupni rezultati, korišćenjem Šenonove funkcije entropije [3].

### 3.4 Odabir metoda i modela

Metode ML se dele na nadgledano (engl. supervised) i nenadgledano (engl. unsupervised) mašinsko učenje. Generalno se bira nadgledano učenje ukoliko je ciljni atribut dostupan za treniranje modela. Izbor modela zavisi od željenog cilja – ako je ciljni atribut diskretna vrednost biraju se klasifikatorski modeli, a ukoliko se traži numerička vrednost biraju se regresioni ili klaster modeli. Nakon toga se izabrani modeli treniraju i njihove karakteristike se porede da bi se odabrao najbolji model.

### 3.5 Podešavanje odabranog modela

Nakon odabira modela, vrši se iterativno treniranje sa podešavanjem hiperparametara izabranog modela. Ovaj proces se odvija sve dok je moguće poboljšati model.

## 4. TEHNOLOŠKI IZAZOVI U UPOTREBI ML ZA DETEKCIJU NTL

Prilikom upotrebe ML alata za detekciju NTL dolazi do kombinacije tehnoloških izazova koji su specifični za ovu oblast primene. Neki od ovih izazova, kao i predložena rešenja su:

### 4.1 Neuravnoteženost raspodele klasa ciljnog atributa u podacima

Problem neuravnoteženosti raspodele klasa ciljnog atributa u podacima (engl. imbalanced dataset) kod primene mašinskog učenja je poznat u literaturi [4][5]. Javlja se kada klase nisu ravnomerno raspodeljene u podacima, odnosno kada se veliki deo podataka klasifikuje na jedan način. U slučaju prepoznavanja netehničkih gubitaka u podacima iz računima radi se o klasi koja ima ispod 1% pozitivnih reizenata. Kao takva, predstavlja veliki problem klasifikatorskim algoritmima da je tretiraju na odgovarajući način. Često se dešava da ovi podaci budu tretirani kao šum, i algoritmi mašinskog učenja ih ignorišu. Postoji više načina za prevazilaženje ovog problema, i ovde ćemo predstaviti najčešće korišćene.

#### 4.1.1 Korišćenje odgovarajuće metrike za evaluaciju modela

Jedan od koraka u procesu mašinskog učenja je izbor modela koji se koristi. U praksi se to realizuje treniranjem različitih modela i zatim evaluacijom utreniranih modela korišćenjem test podataka. Kod neuravnotežene raspodele klasa (mogućih vrednosti atributa) ne treba koristi metrike za evaluaciju modela koje ravnomerno tretiraju klase. Prema tome treba izbegavati standardno korišćene metrike poput tačnosti (engl. accuracy) i F1 rezultata (engl. F1 score), jer one favorizuju klase prema frekvenciji. Samim tim manjinska klasa ne utiče dovoljno na metriku, i može dovesti do pogrešne evaluacije modela [1].

Umesto njih treba koristiti složenije metrike poput ROC krive (engl. skr. Receiver Operator Characteristics), koja je namenjena za analizu binarnih klasifikatora. Zajedno sa ROC krivom se koristi i AUC skor (engl. skr. Area Under roc Curve). Pored ROC/AUC metrike koristi se preciznost (engl. precision) i osetljivost (engl. recall, u literaturi prevedeno i kao „odziv“) [6]. Izbor između preciznosti i osetljivosti zavisi od upotrebe modela. Ukoliko je namena modela sticanje slike o mogućim krađama, onda se favorizuje metrika koja favorizuje smanjenje klasifikacije mogućih NTL kao uredne kupce (minimizacija pogrešno negativnih uz povećanje pogrešno pozitivnih). U tom slučaju treba koristiti osetljivost kao metriku. Ukoliko je pak namena modela jasno određivanje mogućih NTL, onda treba koristiti metriku koja favorizuje smanjenje broja klasifikacija urednih kupaca kao NTL (minimizacija pogrešno pozitivnih uz povećanje pogrešno negativnih). Za to se koristi preciznost kao metrika.

#### 4.1.2 Neravnomerno uzorkovanje

Najčešće korišćen metod za umanj enje problema neuravnoteženosti klasa je neravnomerno uzorkovanje. Ovaj metod koristi preuzorkovanje manjinske klase ili poduzorkovanje većinske klase (ili oba zajedno u ekstremnim slučajevima). Iako se na ovaj način ne dolazi do kvalitativno novog znanja, umanjuje se mogućnost zanemarivanja manjinske klase pretreniranjem (engl. overfitting).

Preuzorkovanje manjinske klase se oslanja na metod kojim se bira način određivanja nove vrednosti manjinske klase. Često se koristi SMOTE algoritam, koji kreira nove uzorke manjinske klase pomoću blisko lociranih uzoraka (engl. nearest neighbor). Pored samog SMOTE, koriste se i njegove varijante Borderline-SMOTE, SMOTE-NC, SMOTE, SVM i ADASYN [4].

Poduzorkovanje većinske klase je složeniji problem, te se preporučuje eksperimentisanje sa različitim algoritmima, kako bi se pronašao optimum za konkretan slučaj. Najbolje rezultate daju sledeći algoritmi za poduzorkovanje: Condensed Nearest Neighbor rule (CNN), Near miss, Tomek Links, Edited Nearest Neighbors rule (ENN), One-Sided Selection (OSS), kao i Neighborhood Cleaning Rule (NCR) [4].

#### 4.1.3 Redefinicija problema

U literaturi se pominje moguće umanj enje problema neuravnoteženosti „redefinicijom problema“ [5]. Kako se radi o postupku koji je strogo vezan za domen konkretnog problema, ne postoji precizan postupak „redefinicije“. U slučaju NTL, ova „redefinicija“ se može ostvariti podelom podataka prema atributima koji utiču na klasu. Na primer, podaci se mogu podeliti prema atributu TIP\_GREJANJA, što smanjuje neuravnoteženost jer iskustvo pokazuje manje NTL kod domaćinstava sa centralnim grejanjem.

#### 4.2 Nepostojeći podaci

Postoji više načina za kompletiranje podataka, kojima se umanjuje problem nedostajućih podataka [7]. Najprostiji je brisanje redova koji nisu kompletni. Ovaj metod treba generalno izbegavati jer dovodi do gubitka dragocenih informacija.

Drugi način je kreiranje indikatorskog atributa koji opisuje nedostatak informacije. Ovaj metod ima smisla kada se nedostatak podataka nekog atributa može dovesti u vezu sa ciljnim atributom. Uobičajeno je da se originalni atribut izbaci prilikom korišćenja indikatora.

Treći način je tzv. imputiranje vrednosti, čime se neka vrednost upisuje na mesto nedostajuće vrednosti atributa. Korišćenje imputiranja se razlikuje u zavisnosti da li se imputira numerička vrednost ili neki diskretni atribut. Generalno, u slučaju numeričkih vrednosti se može koristiti neka srednja vrednost za taj atribut, dok se kod diskretnih može koristiti najčešća vrednost atributa. Ipak, najbolje je koristiti neki algoritam za generisanje nedostajuće vrednosti na osnovu ostalih atributa tog uzorka.

#### 4.3 Slaba korelacija između podataka i ciljnog atributa

Jedan od glavnih ograničavajućih faktora koji utiče na kvalitet rezultata primene algoritama ML je slaba korelacija između podataka i ciljnog atributa. To se najčešće ilustruje tzv. korelacionom mapom (engl. heatmap), iz koje se vidi da su vrednosti polja uglavnom bliske nuli [8].

Ovaj problem se donekle umanjuje postupkom inženjeringa atributa (engl. feature engineering). Inženjeringom atributa se dodaju novi atributi u postojeći skup podataka, tako što se generišu na osnovu postojećih podataka. Pretpostavka je da se na taj način povećava korelacija sa ciljnim atributom, što se proverava korelacionom mapom. Generisanje novih podataka može biti neka matematička funkcija poput multivarijantne normalne raspodele ili neke složenija analiza poput analize periodičnosti podataka. Takođe je moguće koristiti SQL agregacione funkcije za transformaciju postojećih podataka [9].

Pored inženjeringa atributa iz osnovnog skupa podataka, moguće je i korišćenje prepoznavanja teksta iz dokumenata povezanih sa mernim mestom (npr. zapisnici o kontroli mernih mesta i sl.), kako bi se dobio atribut koji ima značajnu korelaciju sa ciljnim atributom [10].

## 5. ZAKLJUČAK

Višedecenijski napori da se pronađu načini za smanjenje NTL ne prestaju. Poslednjih nekoliko godina fokus je na primeni veštačke inteligencije u ovu svrhu. Obimna istraživanja i rezultati ukazuju da ima potencijala za upotrebu ML algoritama, vodeći računa o specifičnosti domena problema. U ovom radu su prezentovani neki od izazova, uključujući tri najveća: neuravnotežena raspodela klasa, nepostojeći podaci i slaba korelacija sa ciljnim atributom. Dati su načini i preporuke za njihovo prevazilaženje.

## LITERATURA

- [1] Sushmita Poudel, Udaya Raj Dhungana, 2022, "Artificial intelligence for energy fraud detection: a review", "International Journal of Applied Power Engineering (IJAPE)", "Vol. 11, No. 2", 109
- [2] Elektrodistribucija Srbije, 2023, „Izveštaj o stepenu usklađenosti planiranih i realizovanih aktivnosti iz trogodišnjeg programa poslovanja za period 01.01. – 31.12.2022.“, [www.elektrodistribucija.rs](http://www.elektrodistribucija.rs)
- [3] Rubaidi Z.S., Ammar B.B., Aouicha M.B., 2022, "Fraud Detection Using Large-scale Imbalance Dataset", "International Journal on Artificial Intelligence Tools", "Vol. 31, No. 8"
- [4] Fernández A., García S., Galar M., Prati R., Krawczyk B., Herrera F., 2018, „Learning from Imbalanced Data Sets“, „Springer“
- [5] Haibo H., Yunqian M., „Imbalanced learning“, „IEEE Press“, 2013., 26
- [6] Branco P., Torgo L., Ribeiro R., 2015, „A Survey of Predictive Modelling under Imbalanced Distributions“, „CoRR“
- [7] Kelleher J., Namee B., D’Arcy A., 2015, "Fundamentals of machine learning for predictive data analytics : algorithms, worked examples, and case studies", "MIT Press", 121
- [8] Oprea, SV., Bâra A., 2021, "Machine learning classification algorithms and anomaly detection in conventional meters and Tunisian electricity consumption large datasets", "Computers & Electrical Engineering", "94"
- [9] Oprea, SV., Bâra, A., 2022, "Feature engineering solution with structured query language analytic functions in detecting electricity frauds using machine learning.", "Nature Scientific Reports", "12", članak br.3257
- [10] Leon C., Biscarri F., Monedero I., Guerrero J.I., Biscarri J., Millan R., 2011, "Integrated expert system applied to the analysis of non-technical losses in power utilities", Expert systems with applications, vol. 38, no. 8, pp. 10 274–10 285