

Prognoziranje cijene električne energije na tržištu različitim metodama

Nikolić, Andrea

Master's thesis / Diplomski rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Electrical Engineering, Computer Science and Information Technology Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:200:255773>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-11-30**

Repository / Repozitorij:

[Faculty of Electrical Engineering, Computer Science and Information Technology Osijek](#)



**SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU
FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I
INFORMACIJSKIH TEHNOLOGIJA**

Sveučilišni studij

**PROGNOZIRANJE CIJENE ELEKTRIČNE ENERGIJE
NA TRŽIŠTU RAZLIČITIM METODAMA**

Diplomski rad

Andrea Nikolić

Osijek, 2021.

SADRŽAJ

1. UVOD.....	3
1.1. Zadatak diplomskog rada	4
2. METODE ZA PROGNOZU TRŽIŠNE CIJENE	5
3. METODE UMJETNE INTELIGENCIJE KORIŠTENE ZA PREDVIĐANJE CIJENE	8
3.1. Umjetna neuronska mreža (ANN)	8
3.1.1. Struktura ANN učenja.....	12
3.2. Support vector machine (SVM)	14
3.2.1. Struktura SVM učenja.....	16
4. USPOREDBA ANN I SVM	16
4.1. Općenito o klasifikaciji	16
4.2. Aproksimiranje granice s SVM-om	18
4.3. Aproksimiranje granice s ANN-om	19
4.4. Sličnosti ANN-a i SVM-a.....	19
4.5. Razlike ANN-a i SVM-a	20
5. OPIS KORIŠTENOG MODELA ZA PREDVIĐANJE CIJENE ELEKTRIČNE ENERGIJE	21
5.1. Model.....	21
5.2. Faze modela	23
5.2.1. Priprema podataka	23
5.2.2. Odabir atributa.....	26
5.2.3. Validacija atributa.....	29
5.2.4. Prognoza.....	31
5.3. Ulazni podaci	32
6. REZULTATI.....	39
6.1. Rezultati NN metode.....	39
6.2. Rezultati SVM metode	41
6.3. Usporedba rezultata NN metode i SVM metode	43
7. ZAKLJUČAK	45
LITERATURA.....	46
SAŽETAK	48
ABSTRACT	48
ŽIVOTOPIS	49

**FERIT**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA
I INFORMACIJSKIH TEHNOLOGIJA OSIJEK

Obrazac D1: Obrazac za imenovanje Povjerenstva za diplomski ispit

Osijek, 19.09.2021.

Odboru za završne i diplomske ispite

Imenovanje Povjerenstva za diplomski ispit

Ime i prezime studenta:	Andrea Nikolić
Studij, smjer:	Diplomski sveučilišni studij Elektrotehnika
Mat. br. studenta, godina upisa:	D-1254, 06.10.2019.
OIB studenta:	99229404356
Mentor:	Izv. prof. dr. sc. Krešimir Fekete
Sumentor:	
Sumentor iz tvrtke:	Slaven Kaluđer
Predsjednik Povjerenstva:	Izv. prof. dr. sc. Zvonimir Klaić
Član Povjerenstva 1:	Izv. prof. dr. sc. Krešimir Fekete
Član Povjerenstva 2:	Zorislav Kraus
Naslov diplomskog rada:	Prognoziranje cijene električne energije na tržištu različitim metodama
Znanstvena grana rada:	Elektroenergetika (zn. polje elektrotehnika)
Zadatak diplomskog rada:	U diplomskom radu potrebno je statistički obraditi set podataka o cijeni električne energije na dostupnoj burzi (npr. CROPEX). Na temelju prikupljenih podataka, kreirati set za trening metoda koje predviđaju cijenu električne energije. Opisati nekoliko metoda koje se koriste za predviđanje cijene električne energije. Na temelju obrađenih podataka, izvršiti prognozu cijene električne energije sa nekoliko različitih metoda i analizirati dobivene rezultate. Tema rezervirana za: Andrea Nikolić Sumentor iz tvrtke: Dr.sc. Slaven Kaluđer (HEP ODS, Elektroslavonija)
Prijedlog ocjene pismenog dijela ispita (diplomskog rada):	Vrlo dobar (4)
Kratko obrazloženje ocjene prema Kriterijima za ocjenjivanje završnih i diplomskih radova:	Primjena znanja stečenih na fakultetu: 3 bod/boda Postignuti rezultati u odnosu na složenost zadatka: 2 bod/boda Jasnoća pismenog izražavanja: 2 bod/boda Razina samostalnosti: 2 razina
Datum prijedloga ocjene mentora:	19.09.2021.
Potpis mentora za predaju konačne verzije rada u Studentsku službu pri završetku studija:	Potpis:
	Datum:

**FERIT**FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA
I INFORMACIJSKIH TEHNOLOGIJA OSIJEK**IZJAVA O ORIGINALNOSTI RADA**

Osijek, 08.10.2021.

Ime i prezime studenta:	Andrea Nikolić
Studij:	Diplomski sveučilišni studij Elektrotehnika
Mat. br. studenta, godina upisa:	D-1254, 06.10.2019.
Turnitin podudaranje [%]:	1 %

Ovom izjavom izjavljujem da je rad pod nazivom: **Prognoziranje cijene električne energije na tržištu različitim metodama**

izrađen pod vodstvom mentora izv. prof. dr. sc. Krešimir Fekete

i sumentora

moj vlastiti rad i prema mom najboljem znanju ne sadrži prethodno objavljene ili neobjavljene pisane materijale drugih osoba, osim onih koji su izričito priznati navođenjem literature i drugih izvora informacija. Izjavljujem da je intelektualni sadržaj navedenog rada proizvod mog vlastitog rada, osim u onom dijelu za koji mi je bila potrebna pomoć mentora, sumentora i drugih osoba, a što je izričito navedeno u radu.

Potpis studenta:

1. UVOD

Energija je ključan temelj društvenog napretka i gospodarskog razvoja. Cijene energije predstavljaju srž energetskeg tržišta. Fluktuacije cijena utječu na protok i raspodjelu resursa na energetskeg tržištu i imaju bitno ulogu u ekonomiji. Kako bi se prognoza dobro napravila potrebno je poznavati tržište, osnovni zakon tržišta je da u svakom trenutku ponuda mora biti jednaka potražnji. Svako tržište uključuje prognoziranje cijene električne energije, što ne predstavlja jednostavan posao obzirom da na cijenu utječe dosta vanjskih čimbenika koji nisu lako predvidivi. Kao alat za previđanje cijena električne energije mogu se koristiti razne metode, tj. modeli, danas se više teži korištenju metoda umjetne inteligencije, obzirom da one rade na principu ljudskog mozga, primjer tih metoda su: umjetna neuronska mreža (ANN), metoda strojnog učenja (SVM) i metoda stabla odluke (DT). Svaki od modela se može koristiti za polusatno predviđanje, satno predviđanje, te dan unaprijed tržište.

Autori u sljedećim literaturama koriste već spomenute metode. Za primjer ANN metode za predviđanje cijene električne energije su autori u literaturi [8], tj. oni koriste DFNN (eng. Deep Feedforward Neural Network) model za satno predviđanje cijene električne energije na Njemačkom i Austrijskom tržištu. Jedna vrsta SVM metode predstavljena je u literaturi [9], u njoj se koristi SSA-CS-SVM (eng. Singular Spectral Analysis-Cuckoo Search-Support vector machine) model za predviđanje polusatne cijene električne energije na Novom Južnom Walesu. DT model ima 2 vrste koje se najčešće koriste za predviđanje cijene električne energije, a to su: RF (eng. Random Forest) i XG Boost (eng. Extreme gradient boosting). Kao primjer, RF model autori u literaturi [10] koriste za satno predviđanje cijene električne energije na Španjolskom tržištu.

U diplomskom radu su opisane različite literature koje se bave metodama za prognoziranje cijene električne energije. U radu je stavljen naglasak na dvije vrste metoda, tj. ANN i SVM metode. Nakon teorijske podloge o metodama, napravljen je model za svaku pojedinu metodu, model je detaljno analiziran, te se na osnovu toga pokušala predvidjeti buduća moguća cijena električne energije.

Diplomski rad se sastoji od sedam poglavlja. U drugom poglavlju ugrubo su opisane sve vrste metoda za prognozu tržišne cijene. U trećem poglavlju detaljno je razrađena problematika umjetne neuronske mreže i strojnog učenja jer su to ujedno i dvije metode na kojima se bazira rad. Četvrto poglavlje predstavlja presjek te dvije metode, tj. usporedbu kako bi se materija lakše razumjela. U petom poglavlju je opisan model koji je napravljen u programu „Orange“, isti model je korišten za predviđanje cijene električne energije, također su opisane i faze modela koje su ključne za dobru

prognozu te ulazni podaci. Poglavlje šest prikazuje rezultate prognoze, za svaku metodu posebno, te usporedbu prognoziranih cijena obje metode sa stvarnim cijenama sa CROPEX-a. U zadnjem, sedmom poglavlju dan je osvrt na cijeli rad i napisana su zaključna razmatranja.

1.1. Zadatak diplomskog rada

U ovom će se diplomskome radu razraditi tema prognoze cijene električne energije na tržištu. Potrebno je statistički obraditi set podataka o cijeni električne energije na dostupnoj burzi (u ovom slučaju će se uzeti podaci sa CROPEX-a). Na temelju prikupljenih podataka, kreirati će se set za trening metoda koje predviđaju cijenu električne energije. Opisati će se nekoliko metoda koje se koriste za predviđanje cijene električne energije. Na temelju obrađenih podataka, izvršiti će se prognoza cijene električne energije koristeći program „Orange“ sa dvije različite metode i analizirati dobiveni rezultati.

2. METODE ZA PROGNOZU TRŽIŠNE CIJENE

Glavna tema ovog rada je prognoziranje cijene električne energije na tržištu. Točno predviđanje cijene električne energije ključno je za orijentaciju energetskog tržišta, ono može osigurati referencu za političare i sudionike na tržištu. U praksi, na cijenu energije utječu vanjski čimbenici te je njihovo točno predviđanje vrlo izazovno. Prema autoru: „Predviđanje cijene energije se odnosi na skupljanje povijesnih podataka i uspostavljanje matematičkog modela kako bi se istražili unutarnji odnosi i propisi između cijene električne energije i faktora koji na to utječu, te onda na osnovi toga predvidjeti buduću cijenu tako da bude zadovoljena određena točnost i brzina.“ [1]

U literaturi [2] je definirano da čimbenici koji se inače razmatraju uključuju ponudu i potražnju na tržištu, sudionike tržišta, troškove, strukturu energetskog sustava te na kraju socio-ekonomsku situaciju.

Prema autoru u literaturi [1] modeli predviđanja povezani s energijom mogu se grubo podijeliti na konvencionalne metode i metode umjetne inteligencije. U osnovi s brzim razvojem metoda umjetne inteligencije mnogi znanstvenici koriste razne metode umjetne inteligencije za predviđanje cijene energije.

Konvencionalne metode su: TS (*eng. Time series forecasting* – Predviđanje vremenskih serija) i regresijski modeli. TS predviđanje se odnosi na predviđanje budućnosti na osnovi prošlih trendova na tržištu. TS modeli uključuju AR (*eng. Autoregressive*) model, MA (*eng. Moving Average* – Pokretni Prosjek) model, ARMA (*eng. Autoregressive Moving Average*) model, ARIMA (*eng. Autoregressive Integrated Moving Average*) i GARCH (*eng. Generalized Autoregressive Conditional Heteroskedasticity*) model. ARMA model kombinira AR i MA modele i može češće dati jednostavniji model dok je s druge strane ARIMA bazirana na ARMA modelu i ta metoda rješava nestacionarne slijedove problema [3]. GARCH model je regresijski koji je izričito prilagođen financijskim podacima [4]. Regresijsko predviđanje se odnosi na uspostavljanje jednadžbe između varijable i korištenja kao model predviđanja koji se temelji na analizi tržišta. [5] Najčešće korišteni regresijski modeli za cijenu energije uključuju LR (*eng. Linear Regression* – Linearna Regresija), RR (*eng. Ridge Regression*) i LASSO (*eng. Least Absolute Shrinkage and Selection Operator*). S druge strane postoje metode umjetne inteligencije od kojih autori ovog rada uzimaju u obzir ANN (*eng. Artificial Neural Network* – Umjetna Neuronska Mreža) metodu, SVM (*eng. Support vector machine*) i DT (*eng. decision tree*- stablo odluke) metodu. Postoji još metoda, ali se oni ograničavaju na ove. [1] ANN metoda je najpopularniji model u kategoriji strojnog

učenja, ona ne predstavlja neki specifični model nego opći okvir. [6] ANN metoda ima više vrsti i varijanti te može biti određena kao dinamični ili statički model te se naravno metoda koristi za prognoziranje cijene električne energije. SVM model predstavlja strojno učenje koje se temelji na načelu statistike, ima jedinstvene prednosti u rješavanju malih uzoraka i nelinearnih problema. [7] Decision tree metoda je model koji koristi strukturu stabla koja se temelji na obilježjima podataka. Može se zaključiti da ovisno o danom problemu se gleda koju metodu primijeniti te je biranje pogodnog, osnovnog modela vrlo važno u predviđanju cijene električne energije. Prednosti i nedostaci već gore navedenih metoda se nalaze u tablici 2.1. kako bi se olakšao odabir modela za istraživače i praktikante. [1]

Tablica 2.1. Prednosti i nedostaci pojedinih modela

VRSTA MODELA	PREDNOSTI	NEDOSTACI
<i>TS model</i>	Jednostavnost, dobri rezultati i stabilni podaci, dobri rezultati za kratka i srednja predviđanja.	Velika ograničenja kod dugoročnog planiranja.
<i>Regresijski model</i>	Model je jednostavan te je brzina modeliranja velika. Može pokazati odnose utjecaja između neovisne i ovisne varijable.	U nekim slučajevima izbor i izričaj čimbenika je spekulativan i ograničen u primjeni.
<i>ANN</i>	Ima jaku nelinearnu sposobnost uklapanja, jednostavna pravila učenja te se lako realizira sa računalom. Velika robusnost, mogućnost memorije, nelinearna sposobnost mapiranja i velika sposobnost samoučenja.	Interpretabilnost modela je slaba, te lako dođe do neuklapanja.
<i>SVM</i>	Dobro ga je primjenjivati kod malih uzoraka problema s jednostavnim algoritmom i robusnošću.	Teško je provesti uzorke osposobljavanja velikih razmjera. Osjetljiv na podatke koji nedostaju te izbor parametara.
<i>Decision tree</i>	Model ima veliku mogućnost generalizacije, brzi trening i nije osjetljiv na podatke koji nedostaju.	Ako postoje nejasnoće u podacima lako će doći do neuklapanja.

Što se tiče ANN modela, razne države koriste drugačije tipove te metode za predviđanje cijene električne energije. Windler i suradnici koriste DFNN (eng. Deep Feedforward Neural Network) model za satno predviđanje cijene električne energije na Njemačkom i Austrijskom tržištu [8]. Zhang i suradnici koriste vrstu SVM modela, točnije SSA-CS-SVM (eng. Singular Spectral Analysis-Cuckoo Search-Support vector machine) model za predviđanje polusatne cijene električne energije na Novom Južnom Walesu [9]. DT model ima 2 vrste koje se najčešće koriste za predviđanje cijene električne energije, a to su: RF (eng. Random Forest) i XG Boost (eng. Extreme gradient boosting). Na primjer, Romero i suradnici koriste RF model za satno predviđanje cijene električne energije na Španjolskom tržištu. [10]

3. METODE UMJETNE INTELIGENCIJE KORIŠTENE ZA PREDVIĐANJE CIJENE

3.1. Umjetna neuronska mreža (ANN)

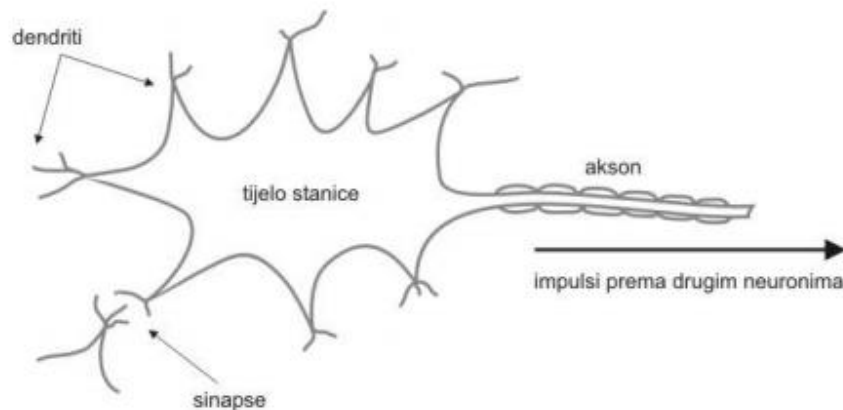
Dugi tijek evolucije dao je ljudskom mozgu mnoge poželjne karakteristike koje nisu prisutne ni u Neumann-u niti u modernim paralelnim računalima kao što su: masivni paralelizam, raspodijeljeno predstavljanje i računanje, sposobnost učenja, sposobnost generalizacije, prilagodljivost, svojstvena kontekstualna obrada informacija, tolerancija kvarova i mala potrošnja energije. Bilo bi poželjno da uređaji temeljeni na biološkim neuronskim mrežama posjeduju neke od ovih karakteristika. Biološka neuronska arhitektura sustava potpuno se razlikuje od von Neumannove arhitekture (Tablica 3.1.). [11] Ova razlika značajno utječe na vrstu funkcija koje svaki računski model može najbolje izvršiti.

Tablica 3.1. Razlike između biološke neuronske i Neumannove arhitekture

	<i>Neumann</i>	<i>Biološko neuronska</i>
Procesor	Jedan ili nekoliko, kompleksan, velika brzina,	Veliki broj, jednostavan, mala brzina
Memorija	Odvojena od procesora, lokalizirana, bez sadržaja adrese	Integriran u procesor, raspodijeljena, postoje sadržaji adresa
Računanje, prebrojavanje	Centralizirano, sekvencijalno, pohranjeni programi	Raspodijeljeno, paralelno, samoučenje
Pouzdanost	Vrlo osjetljiv	Snažan
Stručnost	Brojčane i simbolične manipulacije	Perceptivni problemi
Radno okruženje	Dobro definirano i ograničeno	Siromašno definirano i neobuzdano

Model umjetne neuronske mreže (ANN) je uređen po uzoru na ljudski mozak i predstavlja masivno paralelni računalni sustav koji se sastoji od izuzetno velikog broja jednostavnih procesora s mnogo međusobnih veza. Potrebno je također spomenuti neuron, od koga sve i počinje, neuron predstavlja posebnu biološku stanicu koja obrađuje informacije. Sastavljen je od staničnog tijela i dvije vrste produžetaka: aksona i dendrita kao što je prikazano na slici 3.1. Stanično tijelo ima jezgru koja sadrži informacije o nasljednim osobinama i plazmu u kojoj se nalazi molekularna oprema za

proizvodnju materijala potrebnog neuronu. Neuron prima signale od drugih neurona kroz njegove dendrite i prenosi signale koje generira njegovo tijelo stanice uz akson. Na terminalima ovih strana nalaze se sinapse. Sinapsa je elementarna struktura i funkcionalna jedinica između dva neurona. Kada impuls dolazi do terminala sinapse tada dolazi do oslobađanja neurotransmitera, koji ustvari predstavljaju kemijske tvari koje prenose živčani impuls od stanice do stanice.



Slika 3.1. Građa neurona [11]

➤ Zadaci koje neuronska mreža može obavljati:

- 1) Klasifikacija uzoraka
 - Njegov zadatak je dodijeliti obrazac unosa koji je predstavljen vektorom obilježja jednoj od mnogih unaprijed specificiranih klasa
- 2) Kategorizacija
 - Poznata je i kao klasifikacija uzoraka bez nadzora, algoritmi kategorizacije istražuju sličnosti između uzoraka i organiziraju slične uzorke u grupe
- 3) Aproksimacija funkcije
 - Cilj je pronaći procjenu nepoznate funkcije
- 4) Predviđanje
 - Bitno je samo reći da na osnovu skupa n uzoraka u nekakvom određenom vremenu je potrebno predvidjeti radnju u budućem vremenu, primjer takvog predviđanja je tržište dionica te vremensko predviđanje
- 5) Optimizacija
 - Zadatak optimizacijskih algoritama je pronaći rješenje koje zadovoljava skup ograničenja tako da je ciljana funkcija maksimizirana ili minimizirana
- 6) Dohvaćanje sadržaja

- Jako je poželjno pravljenju multimedijских baza podataka

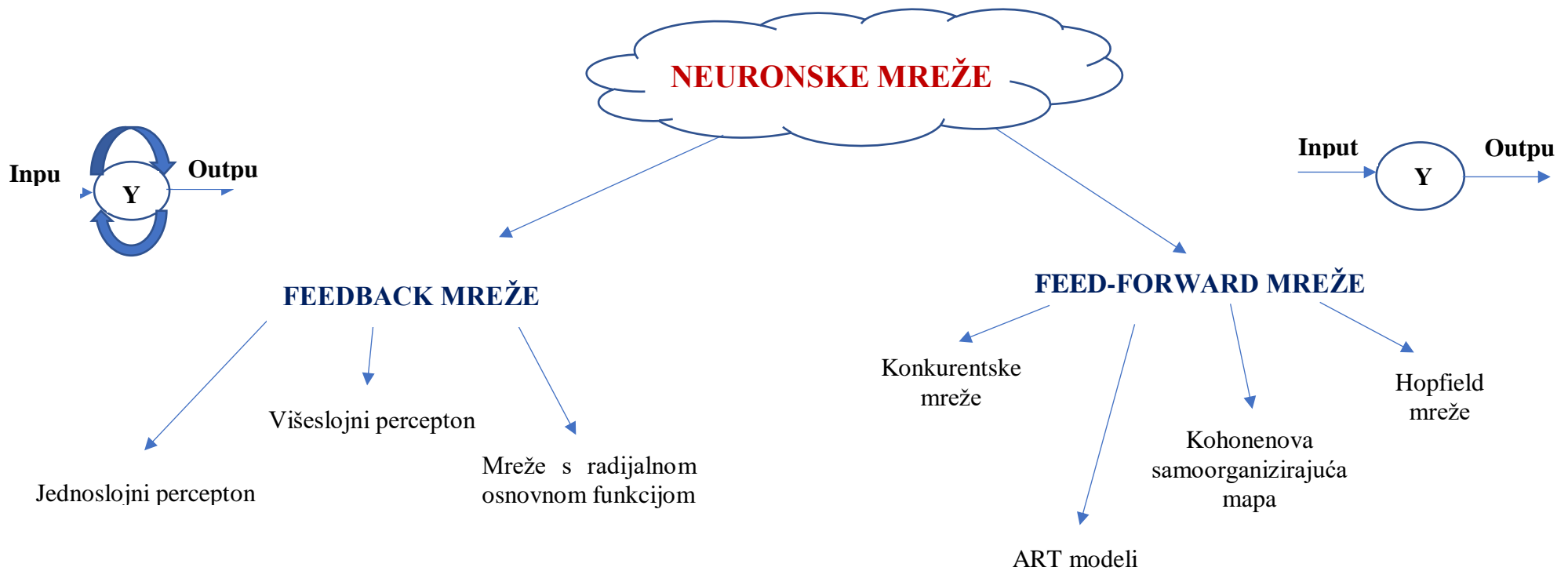
7) Kontrola

- Cilj je generirati upravljački ulaz takav da sustav slijedi željenu putanju određenu referentnim modelom

Kao što je vidljivo na slici 3.2., na temelju uzoraka veze ANN se može podijeliti u 2 grupe:

- 1) Povratne mreže (*eng. feedback*) u kojima postoje petlje te one nastaju zbog povratnih veza
- 2) *eng.feed-forward* mreže u kojima ne postoje petlje i povratne veze

Općenito govoreći, feed-forward mreže su statične, odnosno one proizvode samo jedan skup izlaznih vrijednosti umjesto niza vrijednosti. Također one su „slabe memorijom“ u smislu da je njihov odgovor na ulazu neovisan od prethodnog stanja mreže. Dok s druge strane, povratne mreže su dinamični sustavi. Kada je novi obrazac unosa predstavljen izračunavaju se izlazi neurona. Zbog povratnih puteva ulazi u svaki neuron su modificirani što dovodi do toga da mreža ulazi u novo stanje. [11]



Slika 3.2. Podjela neuronskih mreža

3.1.1. Struktura ANN učenja

Sposobnost učenja je temeljna crta inteligencije. Iako je preciznu definiciju učenja teško odrediti, proces učenja u ANN-u se može promatrati kao problem ažuriranja mrežne arhitekture i težine mreže tako da mreža može učinkovito izvršiti određeni zadatak. Mreža obično mora učiti iz težine veze iz dostupnih obrazaca treninga, izvedba se s vremenom poboljšava iterativnim ažuriranjem težina u mreži. Sposobnost ANN-a da automatski uči na primjerima čini ih atraktivnim i uzbudljivim. Umjesto da slijedi skup pravila navedenih od strane ljudskih stručnjaka ANN-ovci uče osnovna pravila iz dane zbirke reprezentativnih primjera. Ovo je jedna od glavnih prednosti neuronskih mreža u odnosu na tradicionalne ekspertne sustave. Da bi se razumio ili dizajnirao proces učenja potrebno je prvo imati model okoline u kojoj neuronska mreža djeluje, odnosno potrebno je znati koje informacije su dostupne mreži. Takav model učenja naziva se paradigma učenja. [12] Drugo, obavezno je razumjeti način na koji se mrežne težine ažuriraju odnosno koja pravila učenja upravljaju postupkom ažuriranja. Algoritam učenja se odnosi na postupak u kojemu se za podešavanje težine koriste pravila učenja. Tri su glavne paradigme učenja: nadzirano, bez nadzora i hibridno. U učenju pod nadzorom mreža ima točan odgovor za svaki obrazac unosa. Težine su presudne da mreža daje odgovore što bliže poznatim točnim odgovorima. Nasuprot tome, učenje bez nadzora ne zahtijeva ispravan odgovor povezan sa svakim uzorkom unosa. Kod takvog učenja se istražuje temeljna struktura podataka ili korelacije između obrazaca u podacima, i organizira uzorke u kategorije iz tih korelacija. Na kraju, hibridno učenje kombinira nadzirano i nenadzirano učenje. Obično se određuju težine kroz nadzirano učenje, dok su ostali dobiveni nenadziranim učenjem.

Također teorija učenja se mora baviti trima temeljnim i praktičnim pitanjima koja su povezana s učenjem na uzorcima: kapacitet, složenost uzorka i računaska složenost. Kapacitet obuhvaća koliko se uzoraka može pohraniti i koje funkcije i granice odluka mreža može oblikovati. Složenost uzorka određuje broj treninga tj. obrazaca koji su potrebni za osposobljavanje mreže koja jamči ispravnu generalizaciju. Na kraju, računaska složenost, koja se odnosi na potrebno vrijeme za algoritam učenja za procjenu rješenja iz obrasca treninga. Mnogi postojeći algoritmi učenja imaju visoku razinu složenosti računanja. Potrebno je spomenuti da postoje 4 osnovna tipa pravila učenja, a to su: pravilo ispravka pogreške (*eng. Error correction*), Boltzmann-ovo i Hebbian-ovo učenje te na kraju natjecateljsko učenje.

Pravilo ispravka pogreške pripada nadziranom učenju, princip na kojem se temelji pravilo je da se koristi pogrešan signal kako bi se ispravila težina veze te kako bi se pogreška postupno smanjivala.

Kod Boltzmannovog pravila cilj je prilagoditi težine veze tako da stanja nevidljivih jedinica zadovoljavaju određenu raspodjelu vjerojatnosti. Također se Boltzmannovo učenje može gledati kao poseban slučaj pravila ispravka pogreške u kojem greška nije mjerena kao razlika između željenog i stvarnog izlaza nego ona predstavlja razliku između korelacija dvaju neurona pod stegnutim i slobodnim uvjetima rada. Hebbianovo pravilo, čije je najvažnije svojstvo da se učenje radi lokalno, i promjena težine sinapse ovisi samo o aktivnostima dvaju neurona koji su povezani. Kod konkurentnog učenja se izlazne jedinice natječu međusobno radi aktivacije, te kao rezultat se dobije to da je samo jedna izlazna jedinica aktivna u bilo kojem trenutku, ovaj fenomen je poznat kao „pobjednik koji sve uzima“. Najjednostavnija natjecateljska mreža za učenje se sastoji od jednoslojne izlazne jedinice.

Tablica 3.2. prikazuje vrste paradigmi, te određeno pravilo učenja i arhitekturu za nadzirano, nenadzirano i hibridno učenje. Oba nadzirano i nenadzirano učenje se temelji na Hebbianovom, natjecateljskom i pravilu ispravka pogreške. Pravilo učenja koje se temelji na ispravku pogreške se može koristiti za osposobljavanje feed-forward mreža, dok se s druge strane Hebbianovo pravilo učenja može koristiti za sve vrste mrežnih arhitektura. Naravno svaki algoritam učenja je dizajniran za osposobljavanje specifične arhitekture.

Tablica 3.2. Algoritmi učenja

<i>PARADIGMA</i>	<i>PRAVILO UČENJA</i>	<i>ARHITEKTURA</i>	<i>ALGORITAM UČENJA</i>	<i>CILJ</i>
NADZIRANO	Ispravak pogreške	Jednoslojni ili višeslojni perceptron	Perceptron algoritam i algoritam povratnog širenja (<i>eng. Back propagation</i>)	Klasifikacija uzorka, predviđanje, kontrola, aproksimacija funkcije
	Boltzmann	Povratna	Boltzmann algoritam	Klasifikacija uzorka
	Hebbian	Višeslojni feed-forward	Linearna diskriminantna analiza	Analiza podataka, klasifikacija uzoraka

	Natjecateljsko	Natjecateljska	Kvantizacijski vektor	Kategorizacija unutar klase, kompresija podataka
NENADZIRANO	Ispravak pogreške	Višeslojni feed-forward	Sammon's projection	Analiza podataka
	Hebbian	Feed-forward ili natjecateljska, Hopfield mreža	Analiza glavnih komponenata, asocijativna memorija	Analiza podataka, kompresija podataka, asocijativno pamćenje
	Natjecateljsko	Natjecateljska	Kvantizacijski vektor	Kategorizacija, kompresija podataka
HIBRIDNO	Ispravak pogreške i natjecateljsko	RBF mreže	RBF algoritam učenja	Klasifikacija podataka, aproksimacija funkcije, predviđanje, kontrola

Višeslojni perceptron (*eng. Multilayer perceptron*) predstavlja najpopularniju vrstu višeslojnih mreža za prosljeđivanje kod kojega svaka računski jedinica koristi ili funkciju praga ili sigmoidnu funkciju. Također, već navedeni perceptoni mogu oblikovati proizvoljno složene granice odluke i predstavljati bilo koju Booleovu funkciju. Svaki višeslojni perceptron se sastoji od minimalno 3 sloja: ulazni sloj, skriveni sloj i izlazni sloj. Razvoj algoritma učenja povratnog širenja se koristi za određivanje težina u višeslojnom perceptonu tako da se računa gradijent funkcije gubitka s obzirom na težine mreže za jedan primjer ulaz-izlaz.

3.2. Support vector machine (SVM)

Idući najpoznatiji algoritam za rudarenje podacima koji je ujedno i najtočniji i najsnažniji je *eng. Support Vector Machine*, koji sadrži klasifikator i regresor vektorske potpore. Sam algoritam utemeljio je Vapnik devedesetih godina prošloga stoljeća, zasniva se na teoriji statističkog učenja te SVM nije osjetljiv na broj dimenzija. Općenito govoreći SVM se u zadnje vrijeme razvija vrlo velikom brzinom u teoriji i praksi.

Prema autoru u literaturi [13], zadatak klasifikatora vektorske potpore (*eng. Support Vector Classifier*) je pronaći hiperravninu koja je beskonačne dimenzije te može odvojiti 2 klase zadanih uzoraka s maksimalnom marginom koja se pokazala sposobna pružiti najbolju generalizacijsku sposobnost. Hiperravnina predstavlja ustvari količinu prostora između dvije klase. Generalizacijska sposobnost se odnosi na to da klasifikator ima dobre ne samo klasifikacijske performanse nego jamči i visoku točnost predviđanja za buduće podatke iz iste distribucije.

Što se tiče regresora vektorske potpore (*eng. Support Vector Regressor*), on se koristi za rješavanje problema nelinearne regresije. SVR je sličan algoritmu klasifikacije, tj. očekuje se istraživanje glavne karakteristike metodama maksimalne marže iskorištavanjem nelinearne funkcije. Također algoritmi moraju biti učinkoviti u velikim dimenzijama. [14] Obzirom da se razvija robusan procjenitelj koji nije osjetljiv na male promjene u modelu znači da tražimo takozvanu ϵ neosjetljivu funkciju gubitka.

Kako bi se odredio početak točaka SVM algoritma, potrebno je pronaći maksimalnu maržu SVC-a i sljedeći SVR. Međutim, u mnogim realnim problemima to može biti prevelik zahtjev da se sve točke linearno odvoje, posebno u kompliciranim slučajevima nelinearne klasifikacije. Kada se uzorci ne mogu potpuno odvojiti, marže mogu biti negativne. U tim slučajevima ostvarljivo područje prvobitnog problema je prazno, pa je stoga odgovarajući dualni problem neograničen cilj funkcija. To onemogućava rješenje problema optimizacije. [14] Kako bi se riješili ovi problemi koriste se dva pristupa, prvi je opuštanje krutih nejednakosti u jednadžbi koje će dovesti do tzv. meke optimizacije marže, a drugi pristup je primjena trika jezgre za linearizaciju onih nelinearnih problema. Kod drugog pristupa se prvo provodi optimizacija mekih marži. Slijedom toga, u odnosu na SVC s mekom maržom, ovaj se obično naziva SVC s tvrdom maržom. Kada bi se dogodio slučaj u kojemu se pomiješa nekoliko točaka suprotnih klasa u podacima. Točke predstavljaju pogreške u treningu koje čak postoje i za maksimum marže hiperravni. Ideja „meke marže“ želi proširiti SVC algoritam tako da hiperravnina omogućuje postojanje nekoliko takvih bučnih podataka.

Trik s jezgrom je još jedna uobičajena tehnika za rješavanje linearno nerazdvojnih problema. Problem je definirati odgovarajuću funkciju jezgre na temelju internih umnožaka danih podataka, kao nelinearna transformacija podataka s ulaza u prostor obilježja s višom (čak i beskonačnom) dimenzijom kako bi se problemi linearno odvojili. Osnovno opravdanje može se naći u Coverovu teoremu o odvojenosti uzoraka, odnosno složeni problem klasifikacije uzoraka koja je bačena u visokodimenzionalni prostor nelinearno je vjerojatnije da će biti linearno odvojiva nego u prostoru

niske dimenzije. [15] Također značaj jezgre je što ju možemo koristiti za konstrukciju optimalne hiperravnine u prostoru značajki bez potrebe za razmatranjem konkretnog oblika transformacija, koja ne mora biti izričito formulirana u višem (čak i beskonačnom) prostoru obilježja. Kao rezultat, primjena jezgre može učiniti algoritam neosjetljivim na dimenziju, tako da se osposobljava linearni klasifikator u prostor veće dimenzije za učinkovito rješavanje linearno nerazdvojivih problema. Prednost istoimene metode je što se s njom pojednostavi izračun, točnije, pomoću nje se može izbjeći izravno računanje složenog prostora značajki ne samo u proračunu unutarnjih proizvoda nego i u dizajnu klasifikatora. Međutim vrlo je bitno prije primjene trika jezgre razmotriti kako bi trebalo konstruirati funkciju jezgre, tj. koje bi karakteristike trebala zadovoljiti. Tako se prvo uvodi Mercerov teorem koji karakterizira svojstvo funkcije K za kada se smatra istinskom funkcijom jezgre.

3.2.1. Struktura SVM učenja

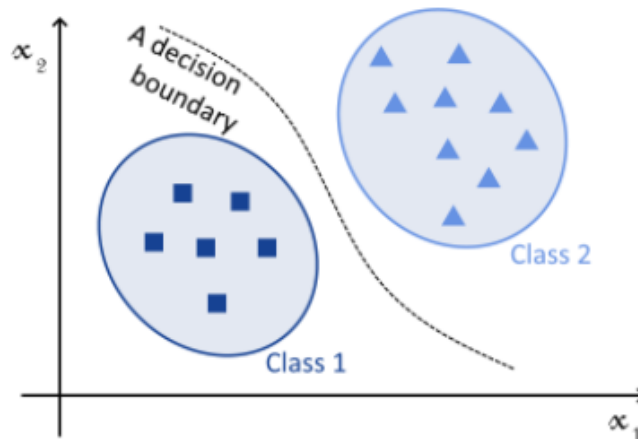
Početna motivacija SVM učenja je maksimizacija marže. S toga ta vrsta učenja usmjerava pozornost na razdvajanje između klase uzoraka, ali ne koristi u dovoljnoj mjeri prethodne podatke o njihovoj raspodjeli unutar klasa. Poznati teorem „Nema besplatnog ručka“ ukazuje na to da ne postoji metoda klasifikacije uzoraka koja je u osnovi superiornija od drugih. [16] To je vrsta problema, početna informacija i količina uzoraka i treninga koji određuju oblik klasifikatora koji se primjenjuje, što zapravo odgovara različitim stvarnim problemima, različite klase mogu imati različite temeljne strukture podataka. Klasifikator bi trebao prilagoditi diskriminirajuće granice kako bi odgovarale strukturama koje su ključne za klasifikaciju, posebno za generalizacijsku sposobnost klasifikatora. Međutim tradicionalni SVM ne razlikuje strukture. Razvijeni su također algoritmi koji daju više brige o strukturnim informacijama nego SVM. Oni pružaju novi pogled na dizajn klasifikatora, gdje klasifikator može biti osjetljiv na strukturu distribucije podataka. Takvi algoritmi su podijeljeni na 2 vrste pristupa, prvi pristup je kroz višestruko učenje, dok je drugi pristup iskorištavanje algoritma klasteriranja, pretpostavljajući da podaci sadrže nekoliko klastera koji sadrže prethodne informacije o distribuciji. [17]

4. USPOREDBA ANN I SVM

4.1. Općenito o klasifikaciji

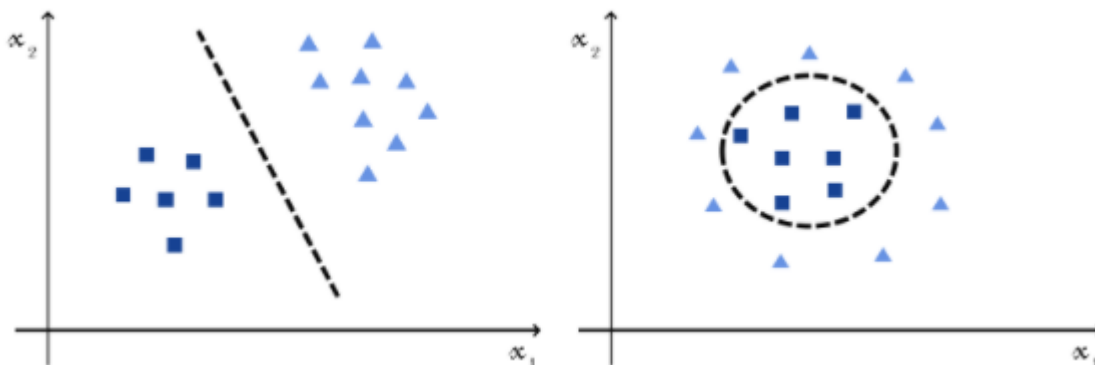
U ovom kratkom pregledu će se proučiti sličnosti i razlike između dvije najpopularnije vrste učenja: support vector machine (SVM) i umjetne neuronske mreže (ANN). Na kraju ovog rada će

se znati razlika između te dvije vrste, karakteristike te slučajevi u kojima je bolje primijeniti jedno ili drugo učenje. Način određivanja granica, točnije problem klasifikacije sastoji se u učenju funkcije oblika $y = f(x)$, gdje je x vektor obilježja, a y vektor koji odgovara klasama povezanim s opažanjima što je prikazano na slici 4.1. Vidljivo je da postoje dvije vrste klasa te su one odijeljene granicom koju je potrebno odrediti.



Slika 4.1. Klasifikacija [18]

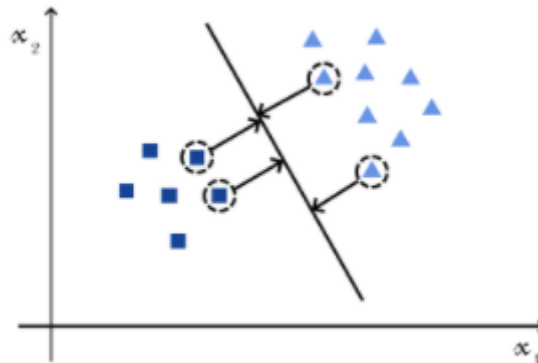
Oba učenja obavljaju ovaj zadatak, naravno s odgovarajućim izborom jezgre. Što bi ukratko značilo da razlika između učenja nije u vrstama zadataka koje obavljaju nego u drugim karakteristikama njihove teorijske osnove i provedbe. Također vrlo bitna stavka je da i SVM i ANN mogu aproksimirati linearne i nelinearne funkcije što je prikazano grafički na sljedećoj slici 4.2. radi lakšeg razumijevanja, u slučaju linearne granice klase su vidno odijeljene, tj. granica je linearna, dok je kod nelinearne funkcije i granica nelinearna. To znači da se oba algoritma mogu jednako nositi sa svim vrstama klasifikacijskih problema, stoga odluka o korištenju jedne nad drugom ne ovisi o samom problemu.



Slika 4.2. Razlike između linearne (lijevi graf) i nelinearne (desni graf) granice [18]

4.2. Aproximiranje granice s SVM-om

Što se tiče SVM-a oni rade na principu identificiranja hiperravnine koja odgovara najboljim mogućim razdvojenostima među najbližim opažanjima koja pripadaju različitim klasama. Ova zapažanja se nazivaju „vektori potpore“. SVM ima hiperravninu koja najbolje razdvaja već spomenute vektore potpore te je ona prikazana na slici 4.3., točnije njezino određivanje.



Slika 4.3. Određivanje granice [18]

Ako hiperravnina ili granica odlučivanja ne postoje u izvornom prostoru obilježja, tada SVM može projicirati taj prostor u novi vektorski prostor veće dimenzionalnosti. Ravnina odvajanja tada se traži u novom, dimenzionalnijem vektorskom prostoru. U vektorskom prostoru gdje na kraju SVM pronađe granicu odluke, ta granica odluke je kontinuirano područje tog vektorskog prostora. Točnije, to je hiperravnina tog prostora. Razlog zašto će se uvijek pronaći rješenje problema klasifikacije je taj što, kao što je ovdje objašnjeno, SVM nisu ograničeni na prostor značajki u kojem je ulaz definiran. Umjesto toga, oni mogu povećati dimenzionalnost problema do prostora u kojem rješenje postoji. To konkretno znači da mogu koristiti onoliko parametara koliko je potrebno sve dok veličina SVM -a ne omogući pronalaženje rješenja.

Što se tiče SVM-a, postoje dvije vrste, točnije, linearni i nelinearni SVM. Linearni SVM je najjednostavniji i slijedi jednostavno pravilo. Kad god se točkasti proizvod izračunava između dvije značajke ulaza, ovaj je proizvod jednak linearnoj kombinaciji unosa:
 $f(ax_n, bx_m) = f(ax_n) + f(bx_m)$

Nelinearni SVM je, umjesto toga, SVM za koji ovo pravilo ne vrijedi. Prilikom izračunavanja izlaza točkastog proizvoda između dvije značajke ulaza, nelinearni SVM koristi jezgru. Riječ jezgra, u strojnom učenju, ima drugačije značenje od jezgre za operativne sustave. U

umjetnoj inteligenciji jezgra odgovara metodi za smanjenje dimenzionalnosti ulaza u klasifikator. Jezgra je, dakle, funkcija koja se koristi umjesto točkastog proizvoda između dva vektora, kad god je potrebno izračunavanje jednog. Primjeri jezgri uključuju polinomsko jezgro, hiperboličnu tangentu i eksponencijalnu jezgru. [18]

4.3. Aproximiranje granice s ANN-om

Ovdje sve počinje od toga da teorem univerzalne aproksimacije govori da neuronska mreža s jednim skrivenim slojem i nelinearnom aktivacijom može aproksimirati bilo koju kontinuiranu funkciju, uz naravno odgovarajući odabir težina.

Ako se granica odluke klasifikacijskog problema može definirati kao kontinuirana funkcija, što je uvijek slučaj, tada se može definirati i kao kontinuirano preslikavanje prostora obilježja, što znači da univerzalni aproksimacijski teorem jamči da ga se može aproksimirati pomoću ANN-a.

Neuronske mreže imaju drugačiji način rada i, posebno, ne zahtijevaju jezgre. Naravno, s izuzetkom konvolucijskih neuronskih mreža. Primjeri nelinearnih aktivacijskih funkcija za ANN su: hiperbolična tangenta i logističke funkcije. Sve ove funkcije uzimaju kao ulaz linearnu kombinaciju vektora značajki x i vektora težine w . Zatim vraćaju izlaz koji se nalazi u nekom konačnom intervalu, obično $(0,1)$ ili $(-1,1)$. Jedini problem ANN je što on ne osigurava konvergenciju kao SVM.

4.4. Sličnosti ANN-a i SVM-a

Postoji nekoliko sličnosti, a to su:

- 1) Oboje su parametarski, u slučaju SVM-a tipični parametri su: tzv. parametar „meke margine“ i parametar funkcije jezgre. ANN mreže zahtijevaju više parametara, te se najvažniji parametri odnose na broj slojeva i njihovu veličinu ali i na broj epoha učenja te brzinu učenja. Ukratko ta dva modela su slična parametarski, ali su različita s obzirom na vrstu i broj parametara koji su im potrebni.
- 2) Oba modela mogu aproksimirati nelinearnost, u slučaju SVM-a to se radi korištenjem jezgre, dok kod neuronskih mreža se nelinearnost ugrađuje pomoću funkcija nelinearne aktivacije.
- 3) Oba modela rade s usporedivom točnošću u odnosu na isti skup podataka, naravno ako im se da usporediva obuka. Međutim, ako im se da što veća obuka i računalna snaga, NN -ovi imaju tendenciju nadmašiti SVM -ove. Vrijeme potrebno za obuku ta dva algoritma se uvelike razlikuje.

4.5. Razlike ANN-a i SVM-a

1) Struktura, SVM posjeduje niz parametara koji se linearno povećavaju s linearnim povećavanjem veličine ulaza, dok s druge strane ANN ne.

2) Količina potrebnih podataka za obuku, SVM koriste samo podskup podataka kao podatke za obuku, to je zato što pouzdano identificiraju granicu odlučivanja na temelju vektora potpore. Kao posljedica toga, broj zapažanja potrebnih za obuku SVM-a nije velik.

Kod neuronskih mreža obuka se odvija na temelju hrpe podataka koji se u nju unose. Što točnije znači da granica odluke uvelike ovisi o redosljed u kojem se prezentiraju skupine podataka, to sve zahtijeva obradu cijelog skupa podataka o obuci, jer u suprotnom mreža može djelovati loše.

3) Vrijeme potrebno za obučavanje algoritma, SVM-ovci općenito brzo treniraju, dok s druge strane neuronske mreže su dosta sporije. Ponekad im je potrebno i nekoliko tjedana da bi se mogle obučiti.

4) Optimizacija parametara, kod neuronskih mreža se optimizacija odvija niz gradijent. Korištenje gradijentnog silaska također je jedan od razloga zašto neuronske mreže ponekad ne mogu naučiti funkciju ako ih početna konfiguracija postavlja na lokalni minimum funkcije. Umjesto toga, SVM koristi metodu koja se naziva kvadratno programiranje. Kvadratno programiranje sastoji se od optimizacije funkcije prema linearnim ograničenjima njezinih varijabli.

5) Osjetljivost na početnu randomizaciju težina.

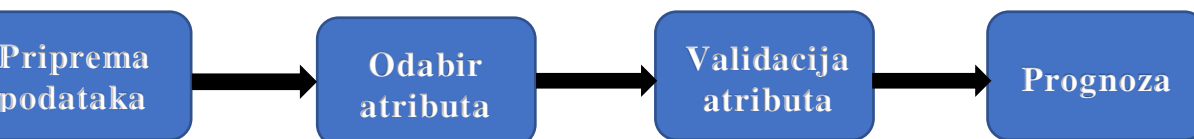
Budući da ANN koriste gradijentno spuštanje, to ih čini osjetljivim na početnu randomizaciju njegove matrice težine. To je zato što, ako početna randomizacija postavi neuronsku mrežu blizu lokalnog minimuma optimizacijske funkcije, točnost se nikada neće povećati iznad određenog praga. SVM -i su umjesto toga pouzdaniji i jamče konvergenciju na globalni minimum bez obzira na njihovu početnu konfiguraciju.

Na kraju, može se zaključiti da se svi zadaci u načelu mogu riješiti s obje klase algoritama. Međutim, kako uvodimo ograničenja koja nemaju veze samo sa zadatkom strojnog učenja, već s procesom razvoja sustava strojnog učenja, tada će jedan algoritam sam biti prikladan za rješavanje određenog problema zajedno s njegovim ograničenjima.

5. OPIS KORIŠTENOG MODELA ZA PREDVIĐANJE CIJENE ELEKTRIČNE ENERGIJE

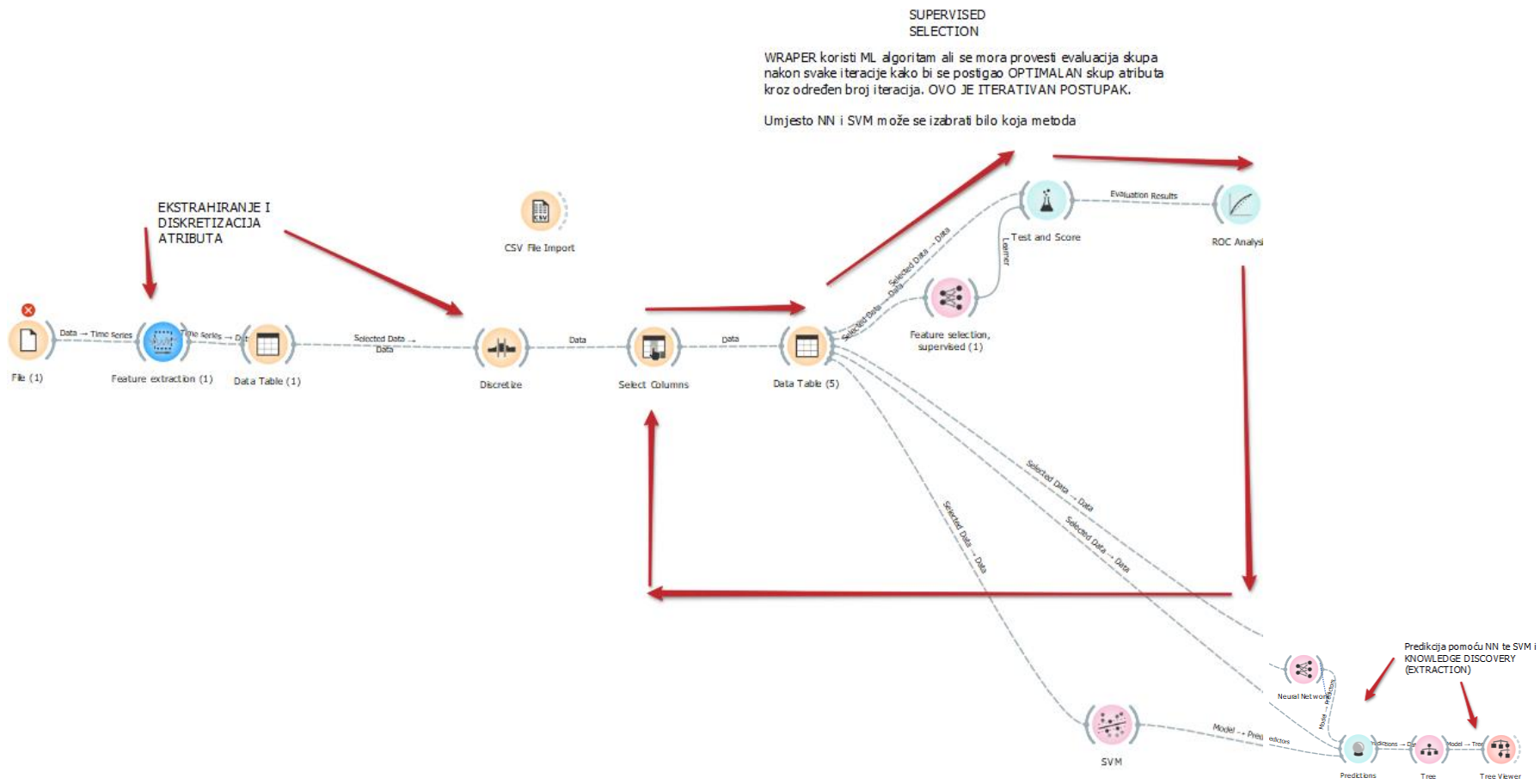
5.1. Model

Simulacijski dio diplomskog rada će se raditi u programu „Orange“ [19] koji je nastao na sveučilištu u Ljubljani, te je pisan u Python programskom jeziku. Već spomenuti program uključuje brojne postupke strojnog učenja te se koristi za pripremu, analizu, prognozu i klasifikaciju podataka. Sama dinamika sustava se može opisati na 3 načina, a to je: pomoću diferencijalnih jednadžbi, koristeći funkcionalno i strukturno znanje, te pomoću atributa. Razvijeni korišteni model u programu se sastoji od 4 faze: priprema podataka, odabir atributa, validacija atributa i na kraju prognoze, koje su prikazane na slici 5.1.



Slika 5.1. Faze za predviđanje cijene električne energije

Priprema podataka obuhvaća čišćenje, organiziranje, diskretizaciju, ekstrahiranje atributa, interpoliranje i redukciju podataka. Atributi kao pojam u ovim slučajevima predstavljaju statističke vrijednosti (npr. minimum, maximum, srednja vrijednost itd.). Što postoji veći broj atributa unutar modela to će i dinamika sustava biti bolje opisana. Broj atributa koji je optimalan za zadovoljavajuće rezultate rijetko je veći od 10. Cijeli ovaj postupak odabira atributa je iterativni postupak, kombiniraju se atributi te se traže oni koji će najviše odgovarati, to je ustvari metoda pokušaja i pogrešaka te nije uvijek jednako vrijeme koje se utroši na to, ponekad to može trajati kraće, ali naravno postoje slučajevi gdje nije lako pronaći kompatibilne attribute, što bi značilo da je cijeli proces poluautomatski. Omotač (*eng. wrapper*) provjerava kompatibilnost ulaznih atributa sa odabranom metodom. Metoda koja se koristi za prognozu mora se koristiti i za validaciju podataka. Nije bitno koja će se metoda koristiti u modelu, može bilo koja te se također može kombinirati dvije metode. U modelu koji se koristi u ovom radu će se primjenjivati metoda strojnog učenja (SVM) i neuronska mreža (NN). Na slici 5.2. je označeno što obuhvaća pripremu podataka, to uključuje ekstrahiranje i diskretizaciju podataka. Odabir atributa predstavlja ustvari iteraciju u krug, koje su označene crvenim strelicama. Validacija atributa je dobivena na osnovi kriterija dobivenih ROC analizom. Prognozu predstavljaju zadnje tri ikone: predictions, tree i tree viewer. Tree i tree viewer služe za vizualizaciju podataka, oni sami poslože podatke, točnije sortiraju vrijednosti.

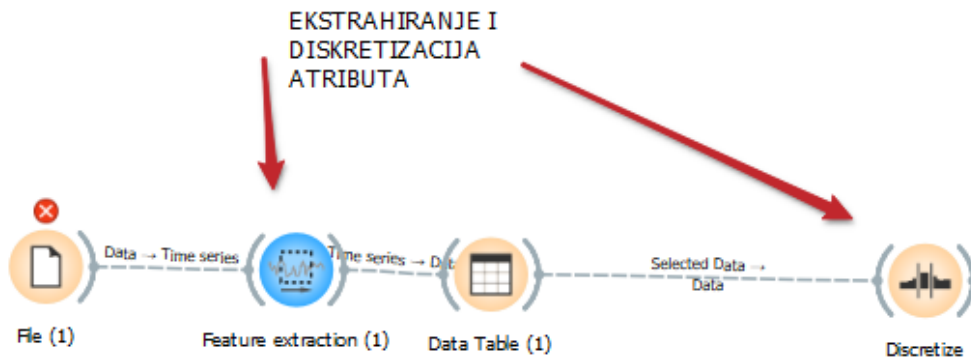


Slika 5.2. Model za prognozu pomoću SVM i NN

5.2. Faze modela

5.2.1. Priprema podataka

Kao što je već navedeno, sama priprema obuhvaća čišćenje, organiziranje, ekstrahiranje atributa, diskretizaciju, interpoliranje i redukciju podataka. Na modelu prema slici 5.3. su to prve četiri ikone.



Slika 5.3. Priprema podataka

Priprema podataka je početak procesa rudarenja podataka te se rudarenje dosta oslanja na kvalitetu pripremljenih podataka. Rudarenje je proces koji uključuje mnogo različitih zadataka i koji ne može biti automatiziran. Ukoliko je loša kvaliteta pripremljenih podataka to onda rezultira netočnim i nepouzdanim rezultatima rudarenja. „Odabir atributa, poznat i kao odabir varijabli, točnije promjenjivog podskupa je tehnika odabira podskupa relevantnih atributa za izgradnju robusnih modela učenja“. [20] Odabir tih atributa također pomaže ljudima da steknu bolje razumijevanje o njihovim podacima tako što će im reći koji su podaci važni i kako su oni međusobno povezani. Jednostavni algoritmi odabira značajki su ad hoc, ali ima i više metodičke pristupe. Ako je dostupan veliki broj značajki, to je nepraktično. Algoritmi odabira značajki obično spadaju u dvije kategorije: rangiranje značajki i odabir podskupa. Rangiranje značajki značajke mjernim podatkom i uklanja sve značajke koje ne postižu odgovarajuću ocjenu. Odabir podskupa pretražuje skup mogućih značajki za optimalni podskup. U statistici, najpopularniji oblik odabira značajki je korak po korak regresije. To je pohlepni algoritam koji dodaje najbolju značajku (ili briše najgoru značajku) u svakom krugu. Glavno pitanje kontrole je odlučivanje kada zaustaviti algoritam. U strojnom učenju to obično čini unakrsna provjera valjanosti. U statistici su neki kriteriji optimizirani. Algoritmi odabira podskupova mogu se obavljati na dva načina, a to je pristup preko omotača (*eng. wrapper approach*) i pristup preko filtera (*eng. filter approach*). Vrlo bitan dio pripreme podataka je diskretizacija, to uključuje podjelu raspona mogućih vrijednosti u

podskupove koji se nazivaju grupe (npr. podjela ljudi po godinama: 0-12=> dijete; 12-17=> tinejdžer; 18-35=>mladi; 36-59=> srednje godine; 60-..=> stariji).

Nadalje, neki oblici podataka, kao što su neobrađeni dnevници, često nisu izravno upotrebljivi zbog svoje nestrukturirane prirode. Drugim riječima, potrebno je izdvojiti korisne značajke iz tih izvora podataka. Stoga je potrebna faza pripreme podataka. Faza pripreme podataka višestupanjski je proces koji obuhvaća nekoliko pojedinačnih koraka, od kojih se neki ili svi mogu koristiti u određenoj aplikaciji. Ovi koraci su sljedeći:

- 1) Izdvajanje značajki i prenosivost - neobrađeni podaci često su u obliku koji nije prikladan za obradu . Osim toga, neki algoritmi mogu raditi samo s određenom vrstom podataka, dok podaci mogu sadržavati heterogene vrste. U takvim slučajevima prenosivost vrste podataka postaje važno gdje se atributi jedne vrste pretvaraju u drugu. To rezultira homogeniji skup podataka koji se može obraditi postojećim algoritmima
- 2) Čišćenje podataka - u fazi čišćenja podataka nedostaju, pogrešni i nedosljedni unosi uklanjaju se iz podataka. Osim toga, neke stavke koje nedostaju mogu se procijeniti i procesom poznatim kao imputacija.
- 3) Smanjenje podataka, odabir i transformacija - u ovoj fazi veličina podataka se smanjuje odabirom podskupa podataka, odabirom podskupa značajki ili transformacijom podataka. Dobici dobiveni u ovoj fazi su dvostruki. Prvo, kada je veličina podataka algoritmi su općenito učinkovitiji. Drugo, ako su nevažne značajke ili uklanjaju se nevažni zapisi, poboljšava kvaliteta procesa rudarenja podataka.

Izdavanje i prenosivost značajki => „Prva faza ekstrakcije značajki je ključna, iako je vrlo specifična za primjenu. U nekim je slučajevima ekstrakcija značajki usko povezana s konceptom prenosivosti vrste podataka, ako se značajke niske razine jedne vrste mogu pretvoriti u značajke više razine druge vrste“ [21]. Priroda ekstrakcije značajki ovisi o domeni iz koje se podaci izvlače:

- a) Podaci senzora koji se često prikupljaju kao velike količine signala niske razine, što su masivni. Signali niske razine ponekad se pretvaraju u značajke više razine koristeći valne ili Fourierove transformacije. U drugim slučajevima, vremenska serija se koristi izravno nakon nekog čišćenja. Područje obrade signala ima opsežnu literaturu posvećenu takvim metodama. Te su tehnologije korisne i za prijenos podataka iz vremenskih serija na višedimenzionalnih podataka.

- b) Slikovni podaci, u najprimitivnijem obliku slikovni podaci predstavljani su kao pikseli. Na nešto viša razina, histogrami boja mogu se koristiti za predstavljanje značajki u različitim segmentima slike. U novije vrijeme upotreba vizualnih riječi postala je više popularna.
- c) Web-zapisnici, obično su predstavljeni kao tekstni nizovi u unaprijed određenom obliku. Budući da su polja u tim zapisnicima jasno određena i odvojena, relativno je jednostavno za pretvaranje zapisnika web-pristupa u višedimenzionalni prikaz (relevantnih) kategorički i numerički atributi.
- d) Mrežni promet, u mnogim primjenama za otkrivanje neovlaštenog ulaska, karakteristike mrežni paketi koriste se za analizu upada ili drugih zanimljivih aktivnosti. Ovisno na temeljnoj primjeni, iz tih se značajki mogu izdvojiti različite značajke pakete, kao što su broj prenesenih bajta, korišteni mrežni protokol i slično.
- e) Podaci dokumenta, često su dostupni u neobrađenom i nestrukturiranom obliku te podaci mogu sadržavati bogate jezične odnose među različitim subjektima. Jedan pristup je ukloniti zaustavne riječi, zaustaviti podatke i koristiti prikaz vrećice riječi. Drugi metode koriste ekstrakciju entiteta za određivanje jezičnih odnosa.

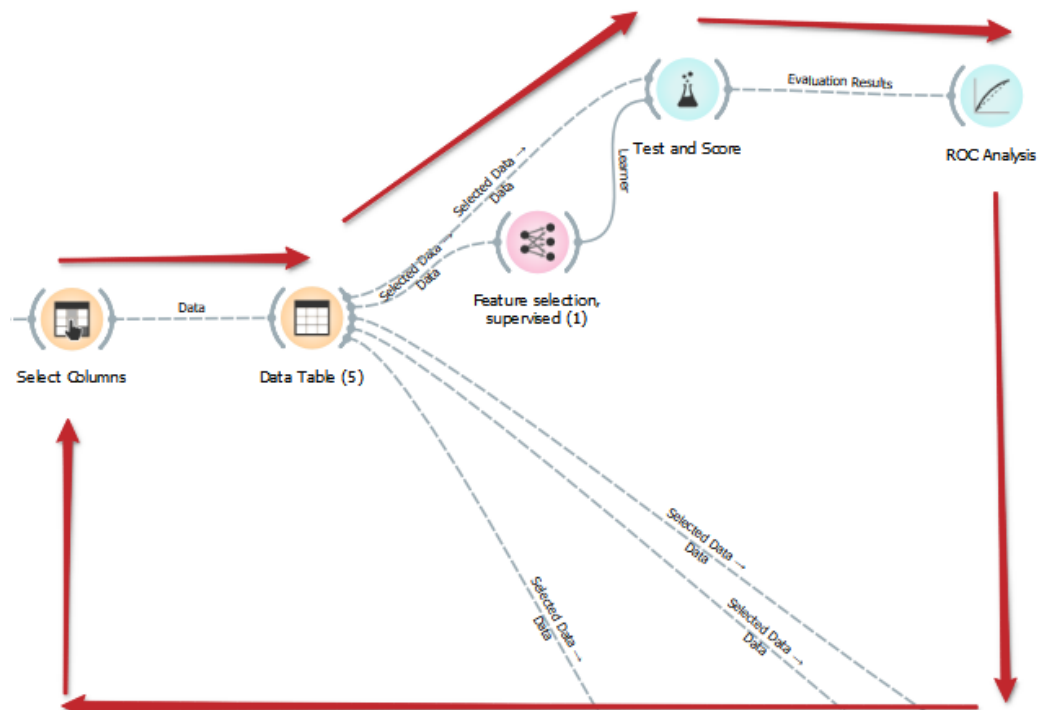
Da se sažme ekstrakcija značajki je umjetnička forma koja uvelike ovisi o vještini analitičara da odaberite značajke i njihov prikaz koji najbolje odgovaraju zadatku.

Postupak čišćenja podataka važan je zbog pogrešaka povezanih s podacima postupka prikupljanja. Tijekom podataka može se pojaviti nekoliko izvora nedostajućih unosa i pogrešaka postupka prikupljanja. Potrebne su metode za uklanjanje ili ispravljanje nedostajućih i pogrešnih unosa iz podaci. Postoji nekoliko važnih aspekata čišćenja podataka, a to su: rukovanje stavkama koje nedostaju, rukovanje neispravnim unosima, skaliranje i normalizacija.

Cilj smanjenja podataka je kompaktnije predstavljanje. Kada je veličina podataka manja, mnogo je lakše primijeniti sofisticirane i računalno skupe algoritme, smanjenje podataka može biti u smislu broja redaka (zapisa) ili u smislu broj stupaca (dimenzija). Smanjenje podataka rezultira gubitkom informacija. Upotreba sofisticiranijeg algoritma ponekad može nadoknaditi gubitak informacija koji proizlazi iz smanjenja podataka. Različite vrste smanjenja podataka koriste se u različitim aplikacijama kao npr. uzorkovanje podataka, odabir značajke, smanjenje podataka s rotacijom osi, smanjenje podataka s transformacijom vrste.

5.2.2. Odabir atributa

Glavni cilj odabira atributa je odabir podskupa relevantnih atributa, od svih dostupnih atributa koji su ponuđeni. Slika 5.4. predstavlja postupak odabira atributa.



Slika 5.4. Odabir atributa

Kada se atributi odaberu onda se oni prosljeđuju algoritmu za rudarenje podataka. Veličina prostora za pretraživanje eksponencijalno raste s brojem atributa. „Odabir atributa vrlo je važan zadatak u rudarenju podataka, posebno u otkrivanju pravila predviđanja. Uostalom, ako se žele otkriti pravila predviđanja vrijednosti određenog atributa cilja, ključno je da atributi prediktora dani algoritmima za otkrivanje pravila budu relevantni za predviđanje tog atributa cilja“. [22] Postoji nekoliko mogućih motivacija za odabir atributa. Prvo, dobro odabran podskup atributa često dovodi do otkrića pravila s većom prediktivnom točnošću od cijelog skupa atributa. Osnovna ideja je da nevažni atributi mogu zavarati ili "zbuniti" algoritam za rudarenje podataka, tako da uklanjanje tih atributa zapravo se može poboljšati kvaliteta otkrivenih pravila. Drugo, ako je metoda odabira atributa brza, ukupno vrijeme potrebno za pokretanje načina odabira atributa, a zatim pokretanje algoritma klasifikacije koristeći samo odabrani atribut mogu biti kraći od vremena potrebnog za pokretanje algoritma klasifikacije pomoću cijelog skupa atributa. Međutim, u praksi atribut odabira može biti prilično spor, tako da odabir atributa zapravo može povećati ukupno vrijeme obrade. Treće, odabrani podskup atributa može dovesti do jednostavnijih (kraćih)

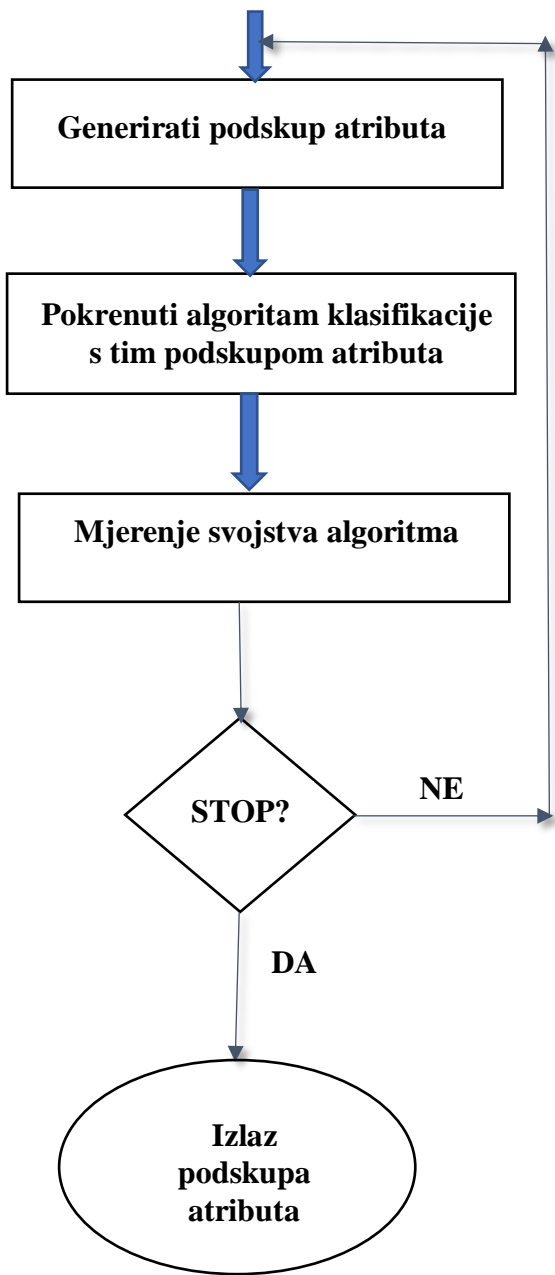
otkrivenih u usporedbi s cijelim skupom izvornih atributa. Takva jednostavnost pravila je u kontekstu rudarenja podataka.

Metode odabira atributa mogu se kategorizirati u dva široka pristupa: filtarski pristup i pristup omotača, oba pristupa su prikazana na slici 5.5. Ključna razlika između ova dva pristupa su sljedeći. U pristupu filtra provodi se odabir atributa uzimajući u obzir algoritam klasifikacije koji će se primijeniti na odabrani atributi. Nasuprot tome, u pristupu omotača odabir atributa provodi se uzimanjem u obzir algoritam klasifikacije koji će se primijeniti na odabrane attribute. U ovom pristupu cilj je odabrati podskup atributa koji je "optimiziran" za određeni algoritam klasifikacije. Također je bitno napomenuti da se pristup omotača pokreće mnogo puta, ali svaki put s različitim podskupom atributa.

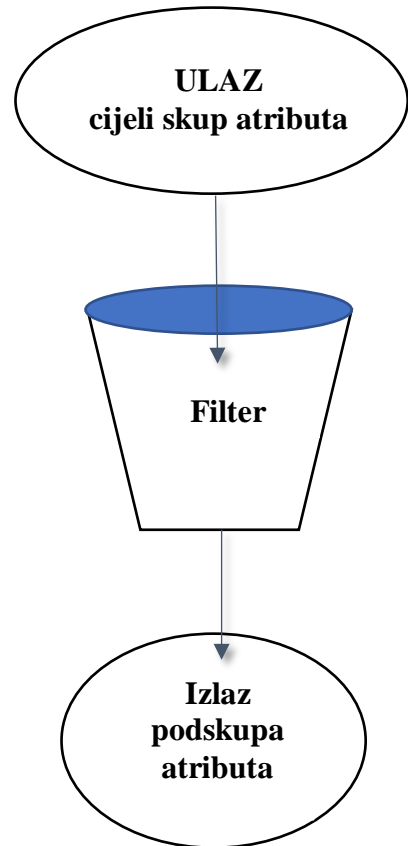
„Kod pristupa omotača koji se koristi u modelu ovog rada važnu točku predstavlja to da se performanse algoritma klasifikacije, koji se koristi za procjenu kvalitete trenutnog podskupa atributa, ne može se procijeniti na testni set. Ako bi se to učinilo, to bi bio "fatalni grijeh" korištenja testnog seta za obuku. Da bi se to izbjeglo, mora se rezervirati dio skupa osposobljavanja u svrhu procjene učinkovitosti algoritma klasifikacije unutar petlje“. [22] Jedan jednostavan način da se to učini sastoji se od nasumičnog dijeljenja treninga postavljenog na dva podskupa, koji se zovu set "trening-trening" i set "trening-test". Prvi se koristi za obuku klasifikacijskog algoritma. Nakon što se algoritam obuči, njegova će se učinkovitost mjeriti na skupu "testa osposobljavanja". Ovaj skup podataka igra ulogu dosad neviđenog testa za algoritam klasifikacije unutar petlje pristupa omotača, ali on je uključen u skup osposobljavanja koji će algoritam klasifikacije koristiti nakon što se pokrene metoda odabira atributa. Kada se metoda odabira atributa prekine, najbolji podskup atributa pronađen tom metodom daje se algoritmu klasifikacije, koji se zatim konačno pokreće na cijelom skupu vježbanja. Strategije pretraživanja atributa koje se koriste kod pristupa omotača su sekvencijalni odabir naprijed i sekvencijalni odabir unatrag. Oba pristupa su ista samo se kod odabira unaprijed iterativno dodaju atributi dok se kod odabira unatrag iterativno uklanjaju atributi, dok ne dođe do poboljšanja neke mjere uspješnosti. Uspješnost se kod prvog pristupa postiže dodavanjem atributa, dok se u drugom postiže uklanjanjem atributa.



ULAZ
cijeli skup atributa



(1) Pristup preko omotača
(eng. wrapper approach)



(2) Pristup preko filtera
(eng. filter approach)

Slika 5.5. Razlika između dva osnovna pristupa (1) i (2)

5.2.3. Validacija atributa



Slika 5.6. Validacija podataka

Validaciju podataka ustvari predstavlja ROC analiza (*eng. receiver operating characteristic-krivulja radnih karakteristika*), čija se ikona nalazi na slici 5.6., ta analiza se odnosi na pitanje kvantifikacije. Postoje definirana dva izazova koja je potrebno riješiti, a to su:

- 1) Metodološka pitanja => njihov cilj je podijeliti na odgovarajući način označene podatke u segmente osposobljavanja i ispitivanja za evaluaciju. Odabir metodologije izravno utječe na postupak evaluacije, kao što je podcjenjivanje ili precjenjivanje točnosti klasifikatora. Moguće je nekoliko pristupa, kao što su zadržavanje i unakrsna validacija.
- 2) Pitanje kvalifikacije => Pitanja kvantifikacije povezana su s pružanjem brojčane mjere za kvalitetu metode nakon odabira posebne metodologije za evaluaciju. Primjeri takvih mjera mogli bi uključivati točnost, točnost osjetljivosti na troškove ili krivulja radnih karakteristika prijemnika (ROC analiza).

Definirat će se kako se provodi kvantifikacija točnosti klasifikatora nakon što se fiksiraju obuka i testni set za klasifikator. Nekoliko mjera točnosti ovisno o prirodi razrednog izlaza:

- a) U većini klasifikatora izlaz se predviđa u obliku oznake povezane s testnom instancom. U takvim slučajevima uspoređuje se oznaka temeljne istine ispitnog slučaja s predviđenom oznakom kako bi se stvorila ukupna vrijednost točnosti klasifikatora.
- b) „U mnogim slučajevima izlaz je prikazan kao brojčana ocjena za svaku mogućnost označavanja za testnu instancu. Primjer je Bayes klasifikator u kojem se prijavljuje vjerojatnost za testnu instancu. Kao konvencija, pretpostavit će se da su više vrijednosti rezultata imaju veću vjerojatnost pripadanja određenoj klasi“. [21]

Izlaz kao numerički rezultat predstavlja prednost jer pruža veću fleksibilnost u ocjenjivanju ukupnog kompromisa između označavanja različitog broja točaka podataka kao pozitivnih. To se postiže korištenjem praga brojčane ocjene za pozitivnu klasu za definiranje binarne naljepnice. Ako je prag odabran preagresivno kako bi se smanjio broj deklarirane pozitivne instance klase, algoritam će propustiti istinito pozitivne instance klase. S druge strane, ako se prag odabere na opušteniji način, to će dovesti do previše lažno pozitivnih rezultata. To dovodi do kompromisa

između lažno pozitivnih i lažno negativnih. Problem je u tome što se "ispravan" prag za uporabu nikada ne zna točno u stvarnom scenarij. Međutim, cijela krivulja kompromisa može se kvantificirati primjenom različitih mjera, i dva algoritma mogu se usporediti u cijeloj krivulji kompromisa. Dva primjera takvih krivulje su krivulja preciznog opoziva i krivulja radne karakteristike prijemnika (ROC).

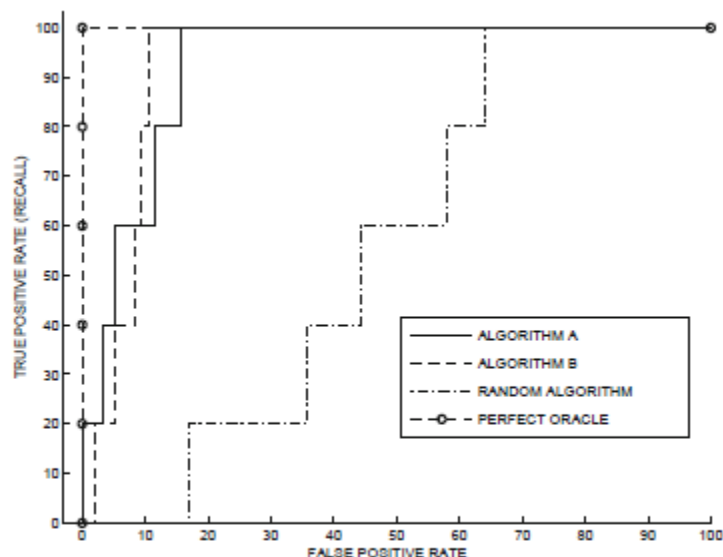
ROC krivulja, prikazana na slici 5.7., definirana je iscrtavanjem $FPR(t)$ na osi X i $TPR(t)$ na osi Y za različite vrijednosti t . Treba imati na umu da su krajnje točke ROC krivulje uvijek na $(0, 0)$ i $(100, 100)$, a očekuje se da će nasumična metoda pokazati izvedbu duž dijagonalne linije povezujući ove točke. Područje ispod krivulje ROC-a pruža konkretnu kvantitativnu procjenu učinkovitosti određene metode. Lažno pozitivna stopa $FPR(t)$ postotak je lažno prijavljenih pozitivnih rezultata iz negativnih na temelju istine. Stoga, za skup podataka D s pozitivnim na temelju istine G , ovaj mjera se definira na sljedeći način:

$$FPR(t) = 100 * \frac{|S(t) - G|}{|D - G|}$$

Istinski pozitivna stopa, koja je jednaka stopi opoziva, definira se kao postotak pozitivnih na temelju istine koji su predviđeni kao pozitivni slučajevi na prag t .

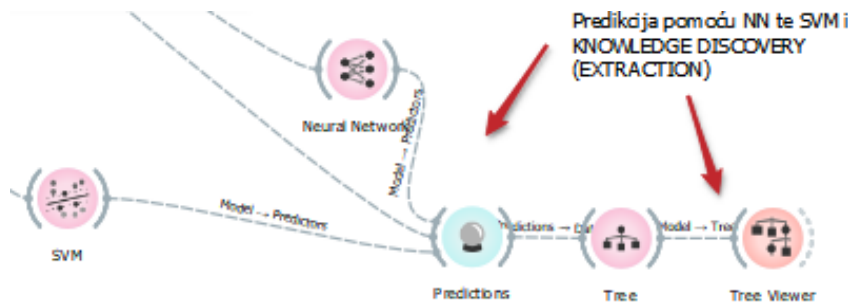
$$TPR(t) = 100 * \frac{|S(t) \cap G|}{|G|}$$

$S(t) \Rightarrow$ deklarirani pozitivni skup klasa, kako se mijenja t mijenja se i $S(t)$



Slika 5.7. ROC krivulja [21]

5.2.4. Prognoza



Slika 5.8. Prognoza

Slika 5.8. prikazuje proces prognoze, ona sadrži predikciju koja služi za predviđanje jednog koraka, npr. ako se predviđa cijena nekog energenta za jedan dan unaprijed, onda na kraju kao rezultat se dobije cijena energenta za idući dan, a ne za nekoliko idućih dana. Za svaki idući dan je potrebno učitati nove podatke. Predikcija se vrši na temelju unaprijed odabrane metode, u ovom diplomskog radu su to SVM i NN metode. Neuronske mreže koriste ML algoritam koji može učitati kako linearne tako i nelinearne modele, ovaj model je nelinearan. Neuronska mreža koristi zadanu predobradu kada nema drugih predprocesora. Izvršava ih sljedećim redoslijedom, također sve to navedeno vrijedi i za SVM:

- Uklanja instance s nepoznatim ciljnim vrijednostima
- Kontinuirira kategoričke varijable
- Uklanja prazne stupce
- Imputira vrijednosti koje nedostaju srednjim vrijednostima
- Normalizira podatke centriranjem na srednju vrijednosti i skaliranjem na standardnu devijaciju od 1

Za regresijske zadatke SVM izvodi linearnu regresiju u prostoru značajke visoke dimenzije pomoću ϵ (neosjetljivog gubitka). Njegova točnost procjene ovisi o dobroj postavci parametara jezgre, te C (troškovi) i ϵ . Program donosi predviđanja klase na temelju SVM regresije, te on radi klasifikacijske i regresijske zadatke i zadatak im je da minimiziraju pogreške

Tree je jednostavan algoritam koji može raditi s kontinuiranim i diskretnim skupovima podataka, te je bitno naglasiti da on ne koristi nikakvu predobradu. Kraće rečeno „Tree“ i „Tree Viewer“ služe za vizualizaciju podataka, oni sami sortiraju vrijednosti tj. poslože podatke.

5.3. Ulazni podaci

Za ulazne podatke su se uzete cijene električne energije [€/MWh]. Podaci su preuzeti sa hrvatske burze električne energije, za tržište dan unaprijed (CROPEX). [23] CROPEX je osnovan 2014.godine te je 50% u vlasništvu Hrvatskog operatora prijenosnog sustava d.o.o. (HOPS) te preostalih 50% je u vlasništvu hrvatskog operatora tržišta energije d.o.o. (HROTE). Hrvatsko tržište dan unaprijed za električnu energiju je pokrenuto u veljači 2016.godine. Burza ustvari predstavlja „mjesto“ na kojemu se susreću prodavači i kupci električne energije, trgovanje je potpuno anonimno, sigurno i organizirano, te burza preuzima rizike kupnje i prodaje električne energije zaključenih transakcija. Kao ulazni podaci za ovaj diplomski rad su se uzimali svi sati karakterističnih srijeda u mjesecu rujnu, kao najoptimalniji dan u tjednu. Podaci se protežu od rujna 2016.godine do 08.rujna 2021.godine, što bi sveukupno bilo 23 podatka. Prognoza se radi za 15. rujna 2021.godine. Cilj rada je prognozirati cijenu električne energije po satima za navedeni datum. Model koji se koristi za to je definiran u poglavlju 4.1. Prognoza se vrši pomoću metode strojnog učenja (SVM) i metode neuronske mreže (NN). U dobivenim podacima će biti vidljivo koja metoda je točnija na osnovi prognozirane vrijednost i njene vjerojatnosti za ovu prognozu.. Svi ulazni podaci su smješteni u excel tablicama po satima radi lakše preglednosti te razumijevanja. Tablica 5.1. prikazuje cijene električne energije za prva četiri sata svih srijeda mjeseca rujna, u prvom stupcu se nalaze datumi, a u drugom stupcu se nalazi cijena električne energije za definirane datume, s tim da treba uzeti u obzir da se pod prvi sat podrazumijeva vremenski period od ponoći do jedan sat ujutro, drugi sat se gleda kao raspon od jedan sat ujutro do dva sata ujutro, i isti princip vrijedi i za sve ostale sate. Tablica 5.2. prikazuje cijene električne energije od 05-08h. U tablici 5.3. se nalaze podaci o cijenama za deveti, deseti, jedanaesti i dvanaesti sat. Tablica 5.4.prikazuje cijene za raspon sati od 13-16h, dok tablica 5.5 definira cijene električne energije od 17-20h, i na kraju tablica 5.6. prikazuje cijene električne energije od 21-24h.

Tablica 5.1. Cijene električne energije tj. ulazni podaci za određene datume po satima (01-04h)

01h		02h		03h		04h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	23,30	07.09.2016	22,07	07.09.2016	26,56	07.09.2016	23,90
14.09.2016.	26,02	14.09.2016.	27,40	14.09.2016.	24,40	14.09.2016.	24,22
21.09.2016.	28,42	21.09.2016.	27,62	21.09.2016.	22,40	21.09.2016.	25,15
28.09.2016.	41,83	28.09.2016.	32,95	28.09.2016.	34,90	28.09.2016.	38,86
06.09.2017.	40,44	06.09.2017.	40,40	06.09.2017.	33,56	06.09.2017.	32,45
13.09.2017.	23,08	13.09.2017.	20,01	13.09.2017.	19,75	13.09.2017.	17,14
20.09.2017.	36,55	20.09.2017.	32,75	20.09.2017.	31,90	20.09.2017.	29,10
27.09.2017.	32,25	27.09.2017.	31,05	27.09.2017.	30,05	27.09.2017.	29,05
05.09.2018.	55,75	05.09.2018.	54,47	05.09.2018.	53,36	05.09.2018.	52,66
12.09.2018.	59,65	12.09.2018.	51,61	12.09.2018.	51,11	12.09.2018.	49,52
19.09.2018.	67,00	19.09.2018.	48,13	19.09.2018.	46,65	19.09.2018.	43,09
26.09.2018.	44,05	26.09.2018.	42,14	26.09.2018.	34,04	26.09.2018.	29,84
04.09.2019.	42,75	04.09.2019.	39,00	04.09.2019.	38,60	04.09.2019.	37,52
11.09.2019.	44,33	11.09.2019.	42,99	11.09.2019.	41,82	11.09.2019.	41,64
18.09.2019.	46,35	18.09.2019.	39,30	18.09.2019.	33,90	18.09.2019.	33,90
25.09.2019.	46,15	25.09.2019.	45,54	25.09.2019.	40,66	25.09.2019.	39,93
02.09.2020.	41,67	02.09.2020.	38,02	02.09.2020.	36,35	02.09.2020.	34,72
09.09.2020.	35,42	09.09.2020.	34,62	09.09.2020.	32,70	09.09.2020.	31,03
16.09.2020.	50,12	16.09.2020.	46,30	16.09.2020.	45,10	16.09.2020.	43,12
23.09.2020.	43,51	23.09.2020.	41,08	23.09.2020.	39,71	23.09.2020.	38,45
30.09.2020.	41,70	30.09.2020.	42,18	30.09.2020.	40,39	30.09.2020.	37,95
01.09.2021.	94,78	01.09.2021.	92,61	01.09.2021.	89,53	01.09.2021.	87,45
08.09.2021.	115,00	08.09.2021.	110,00	08.09.2021.	106,62	08.09.2021.	104,80

Tablica 5.2. Cijene električne energije tj. ulazni podaci za određene datume po satima (05-08h)

05h		06h		07h		08h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	23,65	07.09.2016	25,90	07.09.2016	30,66	07.09.2016	37,80
14.09.2016.	24,52	14.09.2016.	30,67	14.09.2016.	35,07	14.09.2016.	41,46
21.09.2016.	26,97	21.09.2016.	31,70	21.09.2016.	41,77	21.09.2016.	45,26
28.09.2016.	38,81	28.09.2016.	38,81	28.09.2016.	42,67	28.09.2016.	45,60
06.09.2017.	40,43	06.09.2017.	46,62	06.09.2017.	52,75	06.09.2017.	57,02
13.09.2017.	20,00	13.09.2017.	32,31	13.09.2017.	51,80	13.09.2017.	60,70
20.09.2017.	29,75	20.09.2017.	35,80	20.09.2017.	40,95	20.09.2017.	50,60
27.09.2017.	29,08	27.09.2017.	31,07	27.09.2017.	42,70	27.09.2017.	53,00
05.09.2018.	53,18	05.09.2018.	55,32	05.09.2018.	72,00	05.09.2018.	75,65
12.09.2018.	50,05	12.09.2018.	56,10	12.09.2018.	72,16	12.09.2018.	76,04
19.09.2018.	44,26	19.09.2018.	56,06	19.09.2018.	63,15	19.09.2018.	75,99
26.09.2018.	32,10	26.09.2018.	43,01	26.09.2018.	57,08	26.09.2018.	64,30
04.09.2019.	37,67	04.09.2019.	38,35	04.09.2019.	47,15	04.09.2019.	47,00
11.09.2019.	41,82	11.09.2019.	41,99	11.09.2019.	52,00	11.09.2019.	52,00
18.09.2019.	39,21	18.09.2019.	47,93	18.09.2019.	55,55	18.09.2019.	56,50
25.09.2019.	40,69	25.09.2019.	52,00	25.09.2019.	150,05	25.09.2019.	90,08
02.09.2020.	35,70	02.09.2020.	38,35	02.09.2020.	50,13	02.09.2020.	62,23
09.09.2020.	31,41	09.09.2020.	34,07	09.09.2020.	47,84	09.09.2020.	50,20
16.09.2020.	44,00	16.09.2020.	46,35	16.09.2020.	59,98	16.09.2020.	74,55
23.09.2020.	38,41	23.09.2020.	41,80	23.09.2020.	51,92	23.09.2020.	53,21
30.09.2020.	37,79	30.09.2020.	41,89	30.09.2020.	51,18	30.09.2020.	62,37
01.09.2021.	88,43	01.09.2021.	98,80	01.09.2021.	120,43	01.09.2021.	135,12
08.09.2021.	107,38	08.09.2021.	114,26	08.09.2021.	145,00	08.09.2021.	151,16

Tablica 5.3. Cijene električne energije tj. ulazni podaci za određene datume po satima (09-12h)

09h		10h		11h		12h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	43,52	07.09.2016	39,52	07.09.2016	32,47	07.09.2016	32,62
14.09.2016.	46,00	14.09.2016.	49,80	14.09.2016.	49,90	14.09.2016.	53,85
21.09.2016.	46,93	21.09.2016.	44,84	21.09.2016.	43,90	21.09.2016.	43,05
28.09.2016.	47,57	28.09.2016.	42,30	28.09.2016.	35,92	28.09.2016.	35,20
06.09.2017.	62,85	06.09.2017.	65,15	06.09.2017.	60,72	06.09.2017.	60,20
13.09.2017.	60,75	13.09.2017.	61,90	13.09.2017.	56,60	13.09.2017.	58,01
20.09.2017.	55,30	20.09.2017.	53,80	20.09.2017.	52,10	20.09.2017.	51,00
27.09.2017.	53,05	27.09.2017.	50,95	27.09.2017.	46,45	27.09.2017.	44,90
05.09.2018.	76,76	05.09.2018.	75,35	05.09.2018.	71,96	05.09.2018.	74,20
12.09.2018.	76,04	12.09.2018.	73,36	12.09.2018.	59,08	12.09.2018.	64,17
19.09.2018.	77,02	19.09.2018.	84,01	19.09.2018.	80,33	19.09.2018.	78,97
26.09.2018.	62,98	26.09.2018.	52,92	26.09.2018.	46,11	26.09.2018.	71,91
04.09.2019.	55,01	04.09.2019.	50,62	04.09.2019.	47,00	04.09.2019.	47,00
11.09.2019.	54,00	11.09.2019.	52,36	11.09.2019.	47,59	11.09.2019.	45,00
18.09.2019.	64,47	18.09.2019.	63,61	18.09.2019.	59,47	18.09.2019.	59,47
25.09.2019.	75,01	25.09.2019.	72,04	25.09.2019.	75,06	25.09.2019.	70,70
02.09.2020.	71,11	02.09.2020.	63,51	02.09.2020.	54,51	02.09.2020.	49,50
09.09.2020.	49,85	09.09.2020.	54,42	09.09.2020.	51,92	09.09.2020.	50,36
16.09.2020.	84,35	16.09.2020.	75,00	16.09.2020.	65,25	16.09.2020.	59,94
23.09.2020.	58,32	23.09.2020.	58,02	23.09.2020.	54,14	23.09.2020.	57,73
30.09.2020.	72,80	30.09.2020.	65,76	30.09.2020.	55,16	30.09.2020.	52,29
01.09.2021.	140,65	01.09.2021.	130,00	01.09.2021.	116,99	01.09.2021.	115,03
08.09.2021.	149,96	08.09.2021.	138,00	08.09.2021.	124,17	08.09.2021.	109,46

Tablica 5.4. Cijene električne energije tj. ulazni podaci za određene datume po satima (13-16h)

13h		14h		15h		16h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	30,14	07.09.2016	34,32	07.09.2016	42,43	07.09.2016	45,03
14.09.2016.	52,15	14.09.2016.	50,91	14.09.2016.	43,27	14.09.2016.	56,24
21.09.2016.	39,10	21.09.2016.	37,55	21.09.2016.	36,20	21.09.2016.	33,20
28.09.2016.	39,35	28.09.2016.	38,80	28.09.2016.	38,97	28.09.2016.	29,68
06.09.2017.	52,85	06.09.2017.	50,45	06.09.2017.	54,03	06.09.2017.	55,26
13.09.2017.	54,10	13.09.2017.	52,20	13.09.2017.	51,75	13.09.2017.	52,45
20.09.2017.	48,90	20.09.2017.	47,25	20.09.2017.	46,01	20.09.2017.	43,45
27.09.2017.	42,85	27.09.2017.	42,10	27.09.2017.	38,07	27.09.2017.	38,07
05.09.2018.	71,99	05.09.2018.	72,05	05.09.2018.	72,69	05.09.2018.	72,10
12.09.2018.	65,00	12.09.2018.	65,00	12.09.2018.	65,00	12.09.2018.	63,06
19.09.2018.	72,47	19.09.2018.	72,61	19.09.2018.	79,99	19.09.2018.	88,02
26.09.2018.	73,52	26.09.2018.	74,99	26.09.2018.	42,90	26.09.2018.	42,90
04.09.2019.	54,46	04.09.2019.	50,09	04.09.2019.	47,00	04.09.2019.	52,37
11.09.2019.	52,00	11.09.2019.	44,93	11.09.2019.	50,02	11.09.2019.	48,00
18.09.2019.	54,06	18.09.2019.	54,35	18.09.2019.	57,63	18.09.2019.	62,61
25.09.2019.	88,05	25.09.2019.	88,06	25.09.2019.	70,10	25.09.2019.	70,10
02.09.2020.	44,95	02.09.2020.	44,11	02.09.2020.	43,96	02.09.2020.	44,02
09.09.2020.	50,00	09.09.2020.	50,00	09.09.2020.	50,09	09.09.2020.	51,00
16.09.2020.	55,00	16.09.2020.	55,00	16.09.2020.	57,61	16.09.2020.	54,00
23.09.2020.	53,39	23.09.2020.	51,90	23.09.2020.	51,00	23.09.2020.	50,90
30.09.2020.	49,17	30.09.2020.	47,40	30.09.2020.	45,04	30.09.2020.	45,19
01.09.2021.	117,99	01.09.2021.	117,72	01.09.2021.	118,83	01.09.2021.	121,17
08.09.2021.	109,1	08.09.2021.	109,08	08.09.2021.	109,07	08.09.2021.	109,08

Tablica 5.5. Cijene električne energije tj. ulazni podaci za određene datume po satima (17-20h)

17h		18h		19h		20h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	36,65	07.09.2016	38,67	07.09.2016	43,80	07.09.2016	47,02
14.09.2016.	45,77	14.09.2016.	42,88	14.09.2016.	44,80	14.09.2016.	50,75
21.09.2016.	35,12	21.09.2016.	37,28	21.09.2016.	41,14	21.09.2016.	45,40
28.09.2016.	42,25	28.09.2016.	42,35	28.09.2016.	49,87	28.09.2016.	55,83
06.09.2017.	55,35	06.09.2017.	55,35	06.09.2017.	50,35	06.09.2017.	63,95
13.09.2017.	51,30	13.09.2017.	50,19	13.09.2017.	57,30	13.09.2017.	72,95
20.09.2017.	44,15	20.09.2017.	44,85	20.09.2017.	50,35	20.09.2017.	56,16
27.09.2017.	38,95	27.09.2017.	43,80	27.09.2017.	49,75	27.09.2017.	61,55
05.09.2018.	72,02	05.09.2018.	75,36	05.09.2018.	76,76	05.09.2018.	79,92
12.09.2018.	68,79	12.09.2018.	71,99	12.09.2018.	74,98	12.09.2018.	78,78
19.09.2018.	88,11	19.09.2018.	87,63	19.09.2018.	88,10	19.09.2018.	110,38
26.09.2018.	42,99	26.09.2018.	52,09	26.09.2018.	64,90	26.09.2018.	116,39
04.09.2019.	54,39	04.09.2019.	57,06	04.09.2019.	58,00	04.09.2019.	71,43
11.09.2019.	53,61	11.09.2019.	53,93	11.09.2019.	55,00	11.09.2019.	55,00
18.09.2019.	70,00	18.09.2019.	67,52	18.09.2019.	67,61	18.09.2019.	71,48
25.09.2019.	84,99	25.09.2019.	84,99	25.09.2019.	72,00	25.09.2019.	100,09
02.09.2020.	46,46	02.09.2020.	52,35	02.09.2020.	60,40	02.09.2020.	68,89
09.09.2020.	53,02	09.09.2020.	54,07	09.09.2020.	65,00	09.09.2020.	64,37
16.09.2020.	57,83	16.09.2020.	60,00	16.09.2020.	63,40	16.09.2020.	68,32
23.09.2020.	51,94	23.09.2020.	54,93	23.09.2020.	56,43	23.09.2020.	60,82
30.09.2020.	48,94	30.09.2020.	53,98	30.09.2020.	60,14	30.09.2020.	72,43
01.09.2021.	106,57	01.09.2021.	115,32	01.09.2021.	131,00	01.09.2021.	145,32
08.09.2021.	117,52	08.09.2021.	136,03	08.09.2021.	148,68	08.09.2021.	165,78

Tablica 5.6. Cijene električne energije tj. ulazni podaci za određene datume po satima (21-24h)

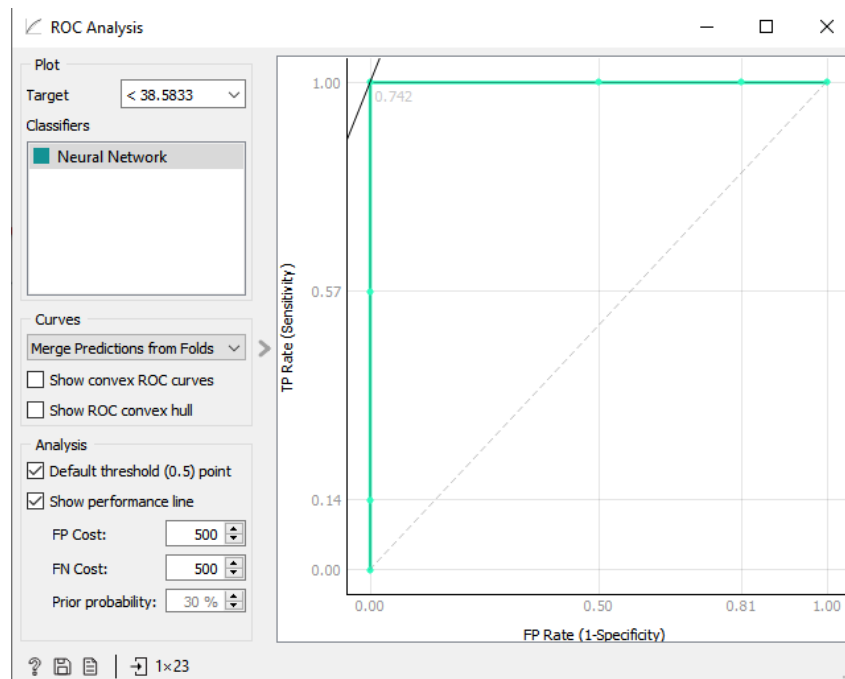
21h		22h		23h		24h	
DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]	DATUM	CIJENA [€/MWh]
07.09.2016	45,27	07.09.2016	40,70	07.09.2016	33,62	07.09.2016	28,79
14.09.2016.	52,47	14.09.2016.	42,22	14.09.2016.	36,87	14.09.2016.	34,05
21.09.2016.	45,05	21.09.2016.	35,27	21.09.2016.	34,27	21.09.2016.	26,62
28.09.2016.	52,15	28.09.2016.	38,47	28.09.2016.	38,70	28.09.2016.	33,55
06.09.2017.	72,72	06.09.2017.	56,25	06.09.2017.	52,55	06.09.2017.	47,5
13.09.2017.	74,40	13.09.2017.	53,60	13.09.2017.	43,95	13.09.2017.	43,11
20.09.2017.	53,65	20.09.2017.	47,05	20.09.2017.	40,45	20.09.2017.	36,85
27.09.2017.	52,35	27.09.2017.	39,56	27.09.2017.	40,09	27.09.2017.	32,80
05.09.2018.	79,33	05.09.2018.	74,20	05.09.2018.	68,82	05.09.2018.	58,84
12.09.2018.	80,08	12.09.2018.	73,42	12.09.2018.	70,91	12.09.2018.	65,61
19.09.2018.	96,79	19.09.2018.	66,15	19.09.2018.	75,00	19.09.2018.	72,59
26.09.2018.	67,46	26.09.2018.	48,08	26.09.2018.	43,08	26.09.2018.	40,30
04.09.2019.	71,93	04.09.2019.	60,00	04.09.2019.	49,00	04.09.2019.	47,00
11.09.2019.	57,09	11.09.2019.	78,00	11.09.2019.	65,00	11.09.2019.	52,75
18.09.2019.	63,61	18.09.2019.	61,00	18.09.2019.	52,82	18.09.2019.	50,62
25.09.2019.	72,02	25.09.2019.	84,99	25.09.2019.	50,00	25.09.2019.	41,00
02.09.2020.	66,05	02.09.2020.	55,46	02.09.2020.	49,34	02.09.2020.	43,94
09.09.2020.	65,00	09.09.2020.	52,82	09.09.2020.	49,09	09.09.2020.	34,47
16.09.2020.	60,20	16.09.2020.	51,98	16.09.2020.	48,77	16.09.2020.	44,18
23.09.2020.	53,80	23.09.2020.	50,90	23.09.2020.	45,00	23.09.2020.	38,5
30.09.2020.	55,34	30.09.2020.	49,95	30.09.2020.	43,13	30.09.2020.	35,67
01.09.2021.	144,45	01.09.2021.	125,30	01.09.2021.	114,39	01.09.2021.	102,73
08.09.2021.	152,36	08.09.2021.	137,11	08.09.2021.	125,01	08.09.2021.	110,54

6. REZULTATI

Prvo će se prikazati rezultati koji su dobiveni metodom neuronskih mreža (NN), a zatim rezultati koji su dobiveni SVM metodom, u tablici se može vidjeti koja metoda predstavlja veću točnost. Ulazni podaci tj. cijene električne energije koje su prikazane u prethodnim tablicama i podešenja korištena za NN i za SVM metodu su identični po satima, te je cijeli algoritam isti, samo se metoda razlikuje. Nakon odabira ulaznih podataka tj. cijena električne energije po satima koji trebaju biti u excel-u, potrebno je odabrati attribute s kojima će algoritam raditi, nakon što su atributi odabrani isti se mogu tijekom provedbe dodati ili oduzeti kako bi se kombiniranjem atributa došlo do najboljeg rezultata. Također kod diskretizacije, gdje se prave kategorije, je potrebno odabrati broj intervala, maksimalan broj je 10, dok u ovom algoritmu kod skoro svih podataka je to broj 6, jedino kod nekih sati se taj broj morao modificirati kako bi se dobila pravilna ROC krivulja te dobra vjerojatnost. U pravilu program raditi točnije s neparnim brojem intervala, ali naravno postoje izuzetci kao u ovom slučaju. Kako bi se približili dobiveni rezultati tj. konačna cijena električne energije za prvi sat će se oni interpretirati i detaljnije objasniti, za ostale sate nema potrebe da se radi detaljna analiza jer je princip isti, kao i analiza. Prognoza u ovom radu se radi za 15. rujna 2021. godine na osnovu prethodnih srijeda od 2016.godine do danas.

6.1. Rezultati NN metode

Prvo će se pojasniti cijena električne energije koja je dobivena metodom neuronske mreže (NN) za prvi sat. Odabir ulaznih podataka u obliku cijene i svi ostali koraci su definirani u prethodnom tekstu. Ikona *Test and Score* prikazuje točnost metode, u ovom slučaju je ta točnost 78,9%. Nakon toga se nalazi ikona *ROC Analysis*, gdje se za te ulazne podatke i točnost dobije ROC krivulja, ista je prikazana na slici 6.1.



Slika 6.1. ROC krivulja za prvi sat (NN metoda)

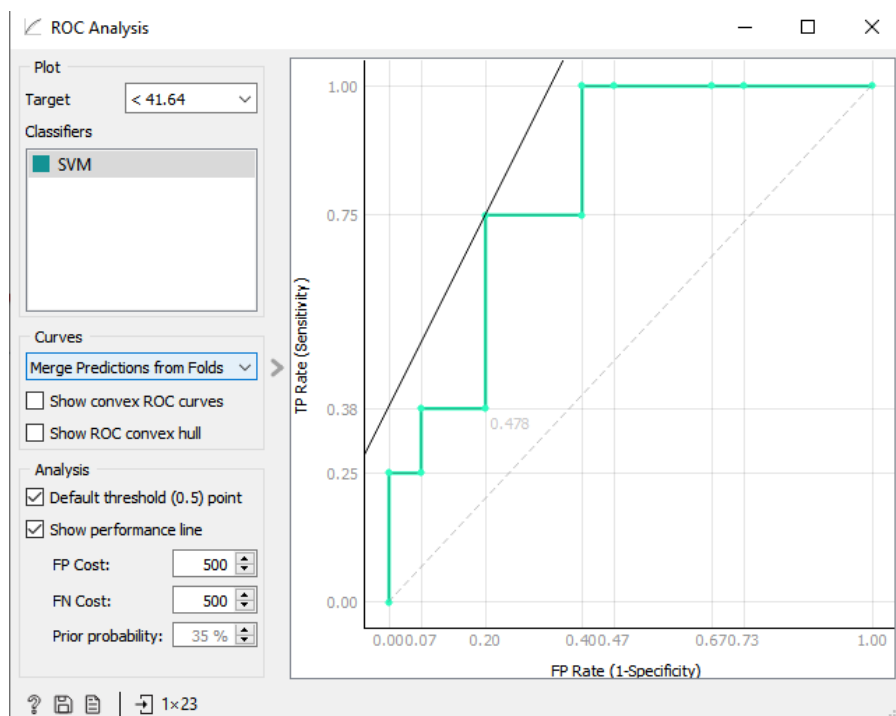
Promatrajući ROC krivulju može se vidjeti da se na Y osi nalazi istinski pozitivna stopa (eng. *True positive rate*-TPR), dok se na X osi nalazi lažno pozitivna stopa (eng. *False positive rate*-FPR). Vrlo je bitno da krivulja ide prema gore pa onda desno do točke (100,100), svaka krivulja mora završiti u toj točki. Točnije, cilj je da krivulja ne pređe iscrtkani dio pravca koji se nalazi na 45°. Ukoliko krivulja ide prvo desno pa se tek onda penje prema gore, krivulja onda ne valja, te onda ni točnost neće biti dobra, što bi značilo da je potrebno promijeniti parametre kao što je npr. broj intervala kod diskretizacije. Nakon što se dobije dobra krivulja i točnost (koja maksimalno može ići do 97%), u ikoni *Tree Viewer* se kao izlazni podaci dobiju razni intervali. Sam program radi na shemu intervala, što znači da su svi dobiveni rezultati tj. intervali točni, na analitičaru je da odabere onaj interval koji po njemu predstavlja najrealniji rezultat obzirom na povijesne podatke.

U ovom slučaju su moguća 3 ishoda, vjerojatnost svakog od njih je ista, pa postoji nekoliko mogućih scenarija u vezi buduće cijene električne energije. Prvi ishod je da cijena električne energije bude od 53-68 €/MWh, drugi ishod je od 68-83 €/MWh, a treći od 83-98 €/MWh te četvrti ishod sa istom vjerojatnošću kao i prva tri je da će cijena biti ≥ 98 €/MWh. Kada se u obzir uzmu cijene koje su bile tijekom prethodnih godina, te da je u ove godine cijena električne energije porasla za 100%, interval koji će se uzeti je da će buduća cijena biti ≥ 98 €/MWh. Čisto za primjer cijena električne energije u rujnu 2019.godine (prva srijeda u mjesecu rujnu, cijena za 15h) je iznosila 47€/MWh, dok u rujnu 2021.godine (također za prvu srijedu u mjesecu rujnu, za 15h) je to bilo 118,23 €/MWh. Vidljivo je da je porast cijene u ovom slučaju čak veći od 100%, postoje

brojni razlozi za to, neki od njih su inflacija, pandemija uzrokovana Covidom-19, smanjenje zaliha plina itd. Da je bila normalna „situacija“, točnije da cijene električne energije nisu naglo skočile zadnje dvije godine svi ovi intervali bi bili točniji, također i sama prognoza. Kada se sve to uzme u obzir neophodno je prilikom prognoze uzeti veću cijenu tj. da će ona biti preko 100€/MWh.

6.2. Rezultati SVM metode

Sam postupak i algoritam je isti kao i za NN metodu. Također je bitno obratiti pozornost na ROC krivulju, osi su već definirane u prethodnom tekstu, sve što je rečeno za ROC krivulju kod NN metode vrijedi i za ovu krivulju. Razlika između te dvije krivulje je što krivulja dobivena SVM metodom ima više „stepenica“, i sama pouzdanost metode je 65,7%. Točnost je manja nego kod NN metode i to za više od 10%, što je i očekivano, krivulja ima više „stepenica“ jer ima veći broj intervala, bez obzira na to krivulja je točna, bitno je da ona ne prelazi liniju od 45° i da završava u točki (100,100), što je vidljivo na slici 6.2. Ulazni podaci su cijene električne energije u obliku excel tablice.



Slika 6.2. ROC krivulja za prvi sat (SVM metoda)

Kod SVM metode se također kao rezultatna cijena dobije nekoliko ishoda sa istom vjerojatnošću. Prvi mogući rezultat je da cijena električne energije bude od 59-77 €/MWh, drugi ishod obuhvaća interval od 77-95 €/MWh, te treći ishod koji ima istu vjerojatnost kao prva dva je da cijena može biti ≥ 95 €/MWh. Interval koji je izabran kao konačni rezultat je ovaj koji kaže da će cijena električne energije biti ≥ 95 €/MWh. Razlozi tome su već naveden prilikom objašnjenja odabira cijene kod NN metode, skok cijena ove godine itd.

Kada se usporede konačne odabrane vrijednosti obje metode, može se vidjeti da se vjerojatnost ne razlikuje puno, točnije za 10%. Konačna odabrana cijena električne energije je za NN metodu ≥ 98 €/MWh, a kod SVM metode ≥ 95 €/MWh. Cijene ne razlikuju previše, obje cijene su približne te i točne vjerojatno obzirom da u zadnje vrijeme postoji veliki skok u cijeni, to može biti također razlog da prognoza ne bude točna. Teže je prognozirati nešto kad postoje takve oscilacije. Rasponi koji su dobiveni kao moguća konačna rješenja su malo veći razlog tome je što je baza podataka koja je korištena za program dosta mala, pa program nije precizan onoliko koliko je maksimalno moguće, tj. program točnije radi s većim bazama podataka.

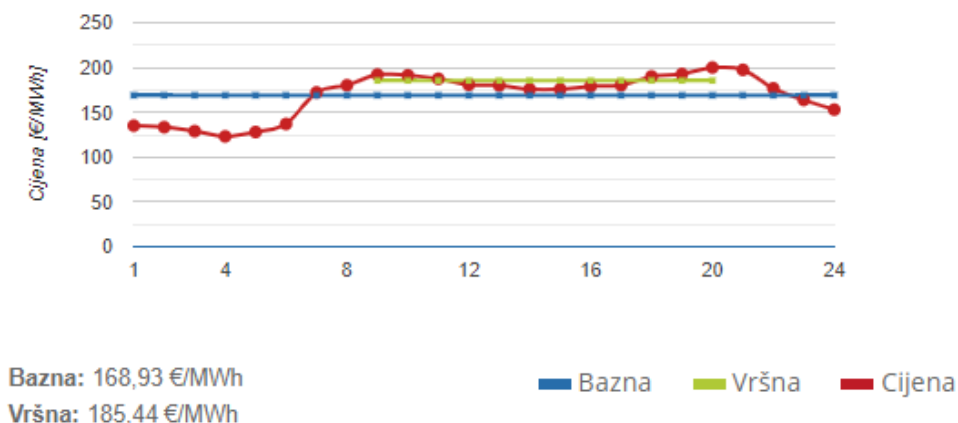
6.3. Usporedba rezultata NN metode i SVM metode

Tablica 6.1. Usporedba predviđenih cijena sa stvarnim cijenama (s CROPEX-a)

SATI (h)	ANN		SVM		STVARNA CIJENA (CROPEX) [€/MWh]
	PROGNOZIRANA CIJENA [€/MWh]	VJEROJATN- OST [%]	PROGNOZIRANA CIJENA [€/MWh]	VJEROJATN- OST [%]	
01	≥ 97,65	78,9	≥ 94,67	65,7	135,08
02	≥ 93,51	64,6	≥ 90,65	76,5	133,40
03	≥ 90,59	60,5	≥ 87,81	81,5	128,95
04	≥ 88,65	56,9	≥ 85,85	67,9	123,10
05	≥ 91,21	67,7	≥ 91,21	52,5	128,01
06	≥ 97,83	81,2	≥ 97,83	84,1	137,01
07	≥ 123,39	70,6	≥ 123,39	57,8	172,23
08	≥ 130,39	67,2	≥ 130,39	66,6	180,24
09	≥ 130,94	72,4	≥ 130,94	71,9	192,11
10	≥ 120,44	80,2	≥ 105,17	83,5	191,32
11	≥ 107,83	72,9	≥ 93,60	83,5	187,42
12	≥ 94,08	72,3	≥ 90,24	65,2	180,80
13	≥ 97,69	62,7	≥ 97,69	69,6	180,00
14	≥ 98,05	57,7	≥ 96,87	63,8	175,74
15	≥ 99,13	55,5	≥ 99,12	53,3	175,76
16	≥ 97,00	55,2	≥ 97,00	57,0	179,09
17	≥ 102,61	76,3	≥ 85,99	80,4	180,25
18	≥ 117,49	52,8	≥ 117,49	52,8	190,00
19	≥ 128,82	75,2	≥ 128,82	76,3	192,78
20	≥ 143,56	73,9	≥ 143,56	57,9	200,00
21	≥ 133,19	74,9	≥ 133,19	61,5	197,41
22	≥ 118,71	46,3	≥ 118,71	48,4	176,78
23	≥ 108,36	61,0	≥ 102,16	74,5	163,82
24	≥ 95,52	57,2	≥ 78,79	74,7	153,02

Kao

Kao što je već rečeno tablica 6.1. prikazuje prognozu po satima za 15.09.2021. godine (srijeda), u prvom stupcu su sati (0-24h), u drugom stupcu se nalaze rezultati predviđenih cijena po NN metodi koji su napravljeni u programu „Orange“, dok se u trećem nalaze predviđene cijene po SVM metodi, za svaki sat. U zadnjem stupcu je konačna cijena električne energije za taj dan, koja je uzeta sa Cropex-a. Tablica je napravljena radi bolje preglednosti svih podataka te kako bi se lakše shvatile promjene i razlike među metodama. Proučavajući podatke može se zaključiti da vrijednosti predviđenih cijena metoda su vrlo približne, u neki satima čak i jednake u obje metode, što bi značilo da niti jedna metoda nije netočna, dapače podaci su kompatibilni sa metodom. Početne postavke programa su podešene i usklađene za NN model, atributi, broj intervala, razreda itd. Iz modela koji je napravljen za NN metodu je izvučen maksimum, tj. najveća moguća točnost, te je to razlog zašto se u tablici 6.1. dobiju točnije i bliže predviđene cijene za NN metodu. Ne mora uvijek biti da je to bolja metoda nego je u ovom slučaju tako ispalo obzirom na dosta mali set ulaznih podataka i ostalo. Atributi koji su se koristili za NN metodu su se uzeli i za SVM, prilagodili su se samo razredi, intervali, točnije diskretizacija. Model nije bio posebno prilagođen postavkama SVM tako da nije čudno što se pokazalo da on ima manju točnost, razlog tome je što iz tog modela nije dobiven maksimum. Kada se to sve pogleda može se reći da se u ovom radu radi paralela između SVM i NN metode, ali s postavkama koje su bazirane na NN metodi, što rezultira većom točnošću NN-a. Konačne cijene električne energije koje su bile za 15. rujna 2021.godine su za svaki sat iznad prognoziranih cijena, to je bio i cilj. Što znači da je prognoza uspješno napravljena, niti u jednom satu nije stvarna cijena bila ispod prognozirane. Kao što se može primijetiti, cijene su vrlo visoke zbog već poznatih razloga, ali bitno je da su iznad prognoziranih cijena. Slika 6.3. prikazuje krivulju cijene električne energije za prognozirani dan po satima.



Slika 6.3. Cijene električne energije za 15. rujna 2021. [23]

7. ZAKLJUČAK

Prognoziranje cijena električne energije je kompleksan posao, potrebno je vrlo dobro poznavati materiju te analitiku za uspješnu prognozu. Postoji dosta čimbenika koji utječu na samu cijenu te ona ima skokovite promjene, što je posebno naglašeno ove godine. Nije bilo očekivano da će cijena doći do 200 €/MWh, međutim to se ostvarilo u kratkom roku. Nakon teorijske podloge o prognoziranju cijene električne energije, metodama kojima je to moguće realizirati, cilj rada je bio predvidjeti cijenu električne energije po satima, jedan dan unaprijed, koristeći NN i SVM metodu, te program „Orange“. Program radi na principu da sortira krajnje rezultate u intervale, svako od rješenja je važeće, na analitičaru je da odabere onaj interval koji bi po njemu najviše imao izgleda kao krajnji rezultat, u slučaju ovog rada, krajnja predviđena cijena. Baza podataka za previđanje cijene je dosta mala pa je to jedan od razloga što program nije skroz precizan, program bolje i točnije radi kada ima veću bazu podataka. Prognoza koja se radila u ovom radu je prognoza tržišta za dan unaprijed. Povijesni podaci tj. cijene električne energije u prethodnim godinama su uzete sa Cropex-a, sa tržišta dan unaprijed. Krajnje predviđene cijene koje su izabrane u oba slučaja nisu znatno odstupale jedna od druge. Previđene cijene koje su dobivene sa NN metodom su točnije i bliže stvarnim cijenama, razlog tome je što su postavke programa usklađene za NN metodu pa program u tom slučaju daje rezultate veće točnosti. Prognoza je uspješno napravljena te niti jedna stvarna cijena električne energije nije ispod predviđene cijene.

LITERATURA

- [1] H. Lu, X. Ma, M. Ma, S. Zhu, Review article Energy price prediction using data-driven models: A decade review, *Computer Science Review*, Vol.39, 2021.
- [2] J. Nowotarski, R. Weron, Recent advances in electricity price forecasting: A review of probabilistic forecasting, *Renewable Sustainable Energy Reviews*, Vol.81, pp. 1548–1568, January 2018.
- [3] M. Valipour, M.E. Banihabib, S.M.R. Behbahani, Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, *Journal of Hydrology*, Vol.476, pp. 433–441, January 2013.
- [4] J.C. Duan, The GARCH option pricing model, *Mathematical Finance*, No.1, Vol.5, pp. 13-32, January 1995.
- [5] A.D. Papalexopoulos, T.C. Hesterberg, A regression-based approach to short-term system load forecasting, *IEEE Transactions on Power Systems*, No.4, Vol.5, pp. 1535–1547, November 1990.
- [6] X. Yao, Evolving artificial neural networks, *Proceeding of the IEEE*, No.9, Vol.87, pp. 1423–1447, September 1999.
- [7] W.S. Noble, What is a support vector machine?, *Nature Biotechnology*, No.12, Vol.24, pp. 1565–1567, December 2006.
- [8] T. Windler, J. Busse, J. Rieck, One month-ahead electricity price forecasting in the context of production planning, *Journal of Cleaner Production*, Vo. 238, November 2019.
- [9] X. Zhang, J. Wang, Y. Gao, A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM, *Energy Economics*, Vol. 81, pp. 899-913, June 2019.
- [10] Á. Romero, J.R. Dorronsoro, J. Díaz, Day-ahead price forecasting for the spanish electricity market, *International Journal of Interactive Multimedia Artificial Intelligence*, No.4, Vol. 5, pp. 42-50, April 2018.
- [11] A. K. Jain, J. Mao, K. M. Mohiuddin, *Artificial Neural Networks: A Tutorial*, IEEE Computer Society, No. 3, Vol. 29, pp. 31-44, March 1996.
- [12] S. Haykin, *Neural Networks: A Comprehensive Foundation*, MacMillan Publishing Company, New York, 1994.

- [13] V. Kumar, X. Wu, The Top Ten Algorithms in Data Mining, CRC Press, United States of America, 2009.
- [14] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.
- [15] S. Haykin. Neural Networks: A Comprehensive Foundation, Cambridge University Press, 2001.
- [16] R.O. Duda, P.E. Hart, and D.G. Stork. Pattern Classification, John Wiley & Sons, New York, 2001.
- [17] P. Rigollet. Generalization error bounds in semi-supervised classification under the cluster assumption. Journal of Machine Learning Research, vol. 8, pp. 1369– 1392, July 2007.
- [18] G.De Luca, SVM Vs Neural Network, Baeldung, 2020, URL: <https://www.baeldung.com/cs/svm-vs-neural-network#classification---a-problem-of-boundary-detection> [30.08.2021.]
- [19] Program: Orange, <https://orangedatamining.com/>
- [20] Guandong Xu, Yu Zong, Zhenglu Yang, Applied Data Mining, CRC tisak, Florida, 2013.
- [21] Charu C. Aggarwal, Data Mining, Springer, Švicarska, 2015.
- [22] Alex A. Freitas, Data Mining and Knowledge Discovery with Evolutionary Algorithms, Springer, Njemačka, 2002.
- [23] Hrvatska burza električne energija, <https://www.cropex.hr/hr/>

SAŽETAK

U ovom diplomskom radu bilo je potrebno pronaći različite izvore koji se bave metodama za prognozu cijene električne energije. Kao glavne metode za prognozu cijene električne energije u ovom radu su metoda umjetne neuronske mreže i metoda strojnog učenja. Svaka od metoda je detaljno analizirana i objašnjena, zatim je napravljena paralela između te dvije metode. Nakon toga je osmišljen model za simulaciju i definirane su njegove osnovne faze provedbe. Simulacijski dio je napravljen u programu „Orange“, te su analizirane dobivene predviđene cijene električne energije u obje metode, cijene su uspoređene međusobno i još sa stvarnim cijenama koje se nalaze na hrvatskoj burzi električne energije. Na temelju rezultata i teorijskog dijela izveden je zaključak o točnosti obje metode i o samoj problematici prognoze koja je na kraju uspješno odrađena.

ABSTRACT

In this thesis it was required to find different sources that deal with methods for predicting the price of electricity. The main methods for predicting the price of electricity in this paper are the artificial neural network method and the machine learning method. Each of the methods was studied and explained in detail, then a parallel was made between the two methods. After that, a simulation model was designed and its basic implementation phases were defined. The simulation part was created in the "Orange" program, and the obtained predicted electricity prices in both methods were analyzed, the prices were compared with each other and with the actual prices on the Croatian electricity exchange. Based on the results and the theoretical part, a conclusion was made about the accuracy of both methods and the very problem of forecasting, which was eventually successfully completed.

ŽIVOTOPIS

Andrea Nikolić rođena je u Slavonskom brodu 8. ožujka 1998. Osnovnu školu završila je u Gradištu. Svoje školovanje nastavlja u Općoj gimnaziji u Županji, te srednjoškolsko obrazovanje završava s odličnim uspjehom. Nakon srednjoškolskog obrazovanja upisuje sveučilišni preddiplomski studij elektrotehnike na Fakultetu elektrotehnike, računarstva i informacijskih tehnologija u Osijeku. Nakon završenog preddiplomskog studija upisuje diplomski sveučilišni studij elektroenergetike, izborni blok elektroenergetski sustavi. Diplomirala je na temu „Prognoziranje cijene električne energije na tržištu različitim metodama“ 2021. godine.