

Badanjak Domagoj  
Fakultet elektrotehnike i računarstva  
[domagoj.badanjak@fer.hr](mailto:domagoj.badanjak@fer.hr)

Ivan Pavić  
Fakultet elektrotehnike i računarstva  
[ivan.pavic@fer.hr](mailto:ivan.pavic@fer.hr)

Doc.dr.sc. Tomislav Capuder  
Fakultet elektrotehnike i računarstva  
[tomislav.capuder@fer.hr](mailto:tomislav.capuder@fer.hr)

## ALAT ZA STATISTIČKU OBRADU VELIKOG SKUPA PODATAKA S ENERGETSKIH TRŽIŠTA S CILJEM STOHAŠTIČKOG MODELIRANJA AGREGATORA ODZIVA POTROŠNJE

### SAŽETAK

Razvijeni programski alat definira korelacije između cijena unutardnevnog tržišta vezanih za uravnoteženje sustava te cijena dan-unaprijed tržišta (ili njihovih prognoza) te tomu sličnih korelacija. Dobiveni rezultati su pridonijeli razvoju histograma te modeliranju frekvencija pojavljivanja određenih cijena u različitim uvjetima na tržištu. Kako se istraživanje temelji na primjeru Danske, tako je njena struktura tržišta električne energije detaljno istražena i u najbitnijim crticama opisana s naglaskom na veleprodaju i pomoćne usluge. Za uspješnu statističku obradu podataka bitan preduvjet predstavljaju kvalitetni ulazni podatci koji po prikupljanju moraju proći proces predobrade kako bi bili u prikladnom obliku za daljnju analizu. Uočivši razliku koja se javlja između cijena unutardnevnog i dan-unaprijed tržišta, metode strojnog učenja su korištenje za predviđanje budućih cjenovnih trendova i značajnih događaja poput velikih razlika u cijeni.

**Ključne riječi:** analiza podataka; modeliranje tržišta; operacijsko planiranje; optimizacija; statistička analiza

## TOOL FOR STATISTICAL ANALYSIS OF ENERGY MARKET DATA WITH THE AIM OF STOCHASTIC MODELLING OF THE DEMAND RESPONSE AGREGATOR

### SUMMARY

The developed tool defines correlations between intra-day prices related to the system balancing with the day-ahead market prices (or its forecast) and similar correlations. Based on these statistical results histograms are created, modelling the frequency of certain price occurrences in different market conditions. As the research is based on the example of Denmark, Danish electricity market structure is thoroughly investigated and described with the focus on the wholesale market and ancillary services. In order to successfully perform the statistical analysis, good quality input data is of the utmost importance. The pre-processing process is conducted so that the data is compatible with the further analysis requirements. As the spread between intraday and da-ahead price is observed, machine learning (ML) algorithms are used to predict future price trends

**Key words:** Data Analytics; Market Modelling; Operational Planning; Optimization; Statistical Analysis

## 1. UVOD

U prošlosti je elektroenergetski sustav gotovo svake zemlje bio jedan veliki monopol. Otvaranjem tržišta u Engleskoj i Walesu [1] devedesetih godina prošlog stoljeća, proces deregulacije u Europi počinje. Od polovice 90-ih godina, *EU Electricity Act* uvodi postepeno otvaranja tržišta električne energije za sve članice Europske unije. Danas je obveza svake države članice Europske unije osigurati liberalizirano i deregularizirano tržište. Trend otvaranja tržišta se nastavlja i dalje, a jedna od težnji je stvaranje jedinstvenog elektroenergetskog tržišta na području Europske unije. Takav razvoj događaja otvara nove mogućnosti, kako proizvođačima električne energije, tako i potrošačima, špekulantima te ostalima koji vide svoju priliku.

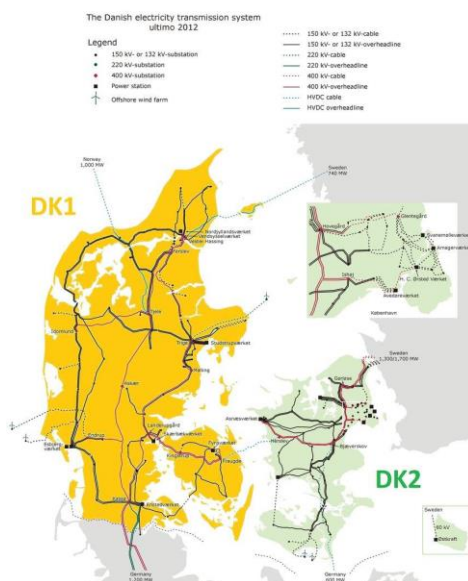
Električna struja je strogo normirana roba kojom se trguje na burzama kao i ostalim dobrima uz bitnu razliku što se i u današnje vrijeme teško skladišti zbog tehničko-ekonomskih razloga. U tu svrhu tržišni mehanizmi osiguravaju da proizvodnja i potrošnja uvijek budu istog iznosa. To se prvenstveno odnosi na pomoćne usluge kojima je za cilj održavanje frekvencije i napona u dozvoljenim intervalima. Upravo su veleprodaja te pomoćne usluge temeljne niše elektroenergetskog tržišta na kojima se bazira istraživanje opisano u ovom referatu.

Kraljevina Danska je odabrana za zemlju čije će tržište biti promatrano. Stoga referat počinje s kratkim opisom danskog elektroenergetskog sustava s naglaskom na veleprodaju i pomoćne usluge. Referat nastavlja s opisom prikupljanja podataka i načinom njihove predobrade. Iako na prvu to ne izgleda kao tema vrijedna posvećivanju pažnje, to je možda najosjetljiviji i vremenski najintenzivniji dio istraživanja. Nastavak donosi statističku analizu predobrađenih podataka te predstavljanje najvećeg doprinosa ovog rada – interpretaciju odnosa cijena različitih aspekata elektroenergetskog tržišta, predviđanje budućih trendova uz pomoć strojnog učenja te prijedlog mogućih daljnjih istraživačkih koraka.

## 2. DANSKA

### 2.1. Struktura elektroenergetskog tržišta

Dansko tržište je podijeljeno u dva područja (DK1 i DK2) kao što je prikazano slikom 1. DK1 je povezano s europskom kontinentalnom mrežom, a DK2 s nordijskom mrežom. Od jeseni 2010. godine dva područja su povezana s 600 MW DC vodom.



Slika 1 - Regije danskog elektroenergetskog sustava [2]

## 2.2. Nord Pool

Energinet je danski operator prijenosnog sustava. Većina veleprodaje se obavlja putem Nord Poola. Nord Pool je kompanija koja, među ostalim, pruža usluge trgovanja na intraday (Elbas) i day-ahead (Elspot) tržištu. Koristi koncept implicitne aukcije. Što znači da se prijenosni kapacitet alocira istovremeno s trgovinom električne energije. Djeluje na području baltičke i nordijske regije te u zemljama poput Njemačke, Francuske, Nizozemske, Belgije, Austrije i Ujedinjenog kraljevstva. U zajedničkom je vlasništvu operatora prijenosnih sustava Švedske, Norveške, Finske, Danske te Baltičkih zemalja.

### 2.2.1. Dan-unaprijed tržište

Zvano i Elspot tržište. Omogućuje trgovanje električnom energijom od 12 do 36 sati prije trenutka isporuka ugovorene energije.

Do 10:00 sati bude poznato koliki je prijenosni kapacitet među pojedinim zonama trgovanja dostupan. Trgovanje završava u 12:00 sati. Tada se formira krivulja ponude i potražnje te su cijene po satu izračunate i objavljene (od 12:42h najranije). U trenutku pisanja rada, najmanja dopuštena cijena je iznosila -500 €/MWh, a najviša 3000 €/MWh. Ukoliko nema zagušenja među različitim zonama, tada su svuda iste cijene. U suprotnom pojedine zone imaju različite cijene.

### 2.2.2. Intraday market

Zvano i Elbas. To je kontinuirano tržište na kojem igrači postavljaju ponude s količinom energije i cijenom za koju su istu spremni prodati/kupiti. Spajaju se pojedinačne ponude i potražnje koje zadovoljavaju dane uvjete dolazi do kupoprodaje električne energije. Elspot ima prednost nad Elbasom u slučaju zagušenja mreže.

### 2.2.3. Tržište uravnoteženja

Tržište koje predstavlja zadnju priliku za ujednačenje proizvodnje i potrošnje bez vanjske intervencije zaduženog regulatora. Često nosi naziv i *real-time* tržište jer se trguje vrlo blizu samog trenutka isporuke. Nudi trgovanje regulacijom snage prema gore i prema dolje (*up* i *down regulation*). Prvo označava povećanu potrebu za energijom, dok potonje označava potrebu za smanjenjem proizvodnje (ili povećanjem potrošnje).

Samo tržište uravnoteženja razlikuje dvije podvrste tržišta prema vremenu trgovanja u odnosu na isporuku: a) u satu isporuke, b) poslije sata isporuke

## 2.3. Pomoćne usluge

Glavna zadaća pomoćnih usluga jest održavanje stabilnosti prijenosnog sustava i kvalitete isporučene energije. Odnosno održavanje frekvencije i napona u dozvoljenim granicama. DK1 regija, koja u fokusu ovog istraživanja, razlikuje tri vrste pomoćnih usluga: a) primarna rezerva (FCR), sekundarna rezerva (aFRR) i tercijarna rezerva (mFRR).

### 2.3.1. FCR

Primarna rezerva je automatska regulacija koja se aktivira odmah po detekciji anomalije i ima za zadatak stabilizirati odstupanje frekvencije blizu 50 Hz. Aktivacija minimalno pola tražene snage mora biti do 15 sekundi nakon naredbe, a ostatak maksimalno do 30 sekundi nakon naredbe te mora moći pružati uslugu minimalno 15 minuta. Po završetku pružanja usluge, ista mora opet biti dostupna u punom kapacitetu nakon maksimalno 15 minuta. Kao i u tržištu uravnoteženja, razlikujemo *up* i *down* regulaciju. Aukcija kapaciteta se odvija za 4-satne blokove dan prije operacije.

### 2.3.2. aFRR

Sekundarnoj rezervi je zadaća vratiti frekvenciju na traženih 50 Hz i rasteretiti primarnu rezervu te ukloniti neravnotežu na interkonekcijama [3]. Kao i primarna, i sekundarna rezerva se sastoji od *up* i *down* regulacije. Mora moći odgovoriti na naredbu u punom kapacitetu nakon ne više od 15 minuta i pružati

uslugu kontinuirano do opoziva. Kapaciteti se dodjeljuju po *pay-as-bid* principu, dok se naknada za pruženu energiju formira u ovisnosti o spot cijeni te je li bila *up* ili *down* regulacija.

### 2.3.3. mFRR

Tercijarna, ili manualna, rezerva pruža pomoćnu uslugu u oba smjera (kao i prethodne dvije). Kao i sekundarna rezerva, mora biti dostupna u punom kapacitetu nakon maksimalno 15 minuta od aktivacijskog signala. Aukcije se odvijaju jednom dnevno u dvije kategorije: *up* i *down* regulacija. Sve prihvaćene ponude rezervacije kapaciteta u pojedinoj kategoriji za naknadu dobivaju najveću prihvaćenu ponudu. Naknada za isporuku energije se formira prema pravilima za formiranje cijena na tržištu uravnoteženja.

## 3. PRIKUPLJANJE PODATAKA I PREDOBRAĐA

Prikupljanje podataka je prvi korak za svaku daljnju analizu. Ne samo da je bitan korak, već može biti i vremenski vrlo intenzivan. Zanimljivo je spomenuti kako je u zadnje dvije godine stvoreno više podataka nego u cijelo ljudskoj povijesti do tada [4]. Imajući to na umu, jasno je kako veliki problem može predstavljati potraga za kvalitetnim i adekvatnim podacima te onda njihova pohrana i daljnja obrada. I u slučaju ovog istraživanja za dohvaćanje pojedinih podataka i njihovu obradu trebalo se dobrano potruditi. Iako su prikupljanje i obrada podatka dva uvelike povezana procesa, ovom prilikom će biti opisani u dva odlomka radi veće jasnoće.

### 3.1. Prikupljanje podataka

Prije no što se krene u sam proces prikupljanja podataka, valja definirati koji se podatci traže. U slučaju ovog referata naglasak je na povijesnim cijenama različitih niša elektroenergetskog tržišta. Pritom su dohvaćene cijene sljedećih niša:

- Tržište dan unaprijed
- Unutardnevno tržište
- Tržište uravnoteženja
- Primarna rezerva
- Sekundarna rezerva
- Manualna (tercijarna) rezerva

Ciljani period je bio od 01.01.2015. do 31.12.2018., no za sve podatke nije bilo moguće pokriti navedeno razdoblje u cijelosti. Taj problem će biti pobliže razjašnjen u dijelu o obradi podataka. Uz podatke o cijenama, dohvaćeni su i drugi relevantni podatci poput: prognoze proizvodnje vjetrom, potrošnja, prognoza proizvodnje, prognoza potrošnje, ukupni planirani tok snage,...

Podatci su bili dohvaćeni iz dva izvora. Prvi jest specijalizirana web-stranica danskog operatora prijenosnog sustava ([www.energidataservice.dk](http://www.energidataservice.dk)), dok je druga web-sjedište Nord Poola ([www.nordpoolgroup.com](http://www.nordpoolgroup.com)).

#### 3.1.1. Energidataservice.dk

Izrađena i vođena od strane Energineta, besplatna je i otvorena za sve posjetitelje željne podataka o danskom elektroenergetskom sustavu. Nudi široku lepezu podataka, od CO2 emisija, do statistika poput podataka o proizvodnji i potrošnji. Nudi pristup podacima putem aplikacijskog programskog sučelja (API) ili direktno preuzimanje u više ponuđenih formata (JSON, CSV, XLSX, XML) uz mogućnost filtriranja podataka prema raznim kategorijama.

#### 3.1.2. Nordpoolgroup.com

Ono što nije bilo moguće dohvatiti putem web-stranice Energineta, dohvaćeno je putem Nord Poola. Iako nudi široku lepezu podataka, sam proces dohvaćanja podataka nije jednostavan i intuitivan kao u prethodnom slučaju. Podatci su već raspoređeni u formirane XLS datoteke i nema mogućnosti pretrage po više kriterija. Štoviše, poneki podatci su dostupni samo po principu učitavanja prikaza tablice na samome web-mjestu i to dan po dan. Zbog toga nije bilo druge doći razviti *web scraping&crawling* alat koji će dohvatiti tražene podatke i spojiti ih u jednu datoteku. Dostični alat je oponašao ljudske pokrete te učitavao prikaz

tablice dan po dan, onda dohvatio tablicu, spojio ju s ostalim tablicama i krenuo u novi dan. Putem HTML tagova alat je „znao“ koji gumb treba stisnuti da bi došao na željeno mjesto. Tu radnju je trebalo ponoviti za svaki dan u četverogodišnjem periodu. Kako je uvijek trebalo pričekati da se stranica nanovo učita za svaki naredni dan, trajanje procesa se mjerilo u satima.

### 3.2. Predobrada podataka

Prilikom prihvaćanja podataka, poglavito kada se oni dohvaćaju iz različitih izvora, može doći do različitih nedosljednosti i problema. Tako mogu nedostajati podatci za određene periode, mogu biti drugačije sortirani i drugi slični razlozi. Stoga bi predobradu opravdano bilo definirati kao transformaciju sirovih podataka u smislen i razumljiv skup podataka pogodan za daljnju analizu. S time da predobrada ne uključuje samo detekciju pogrešaka i slične akcije, već i uređivanje podataka upravo u onu specifičnu formu koja će izvući najviše bitnih informacija iz podataka i biti najpogodnija za konkretni slučaj.

#### 3.2.1. Detekcija i uklanjanje pogrešaka

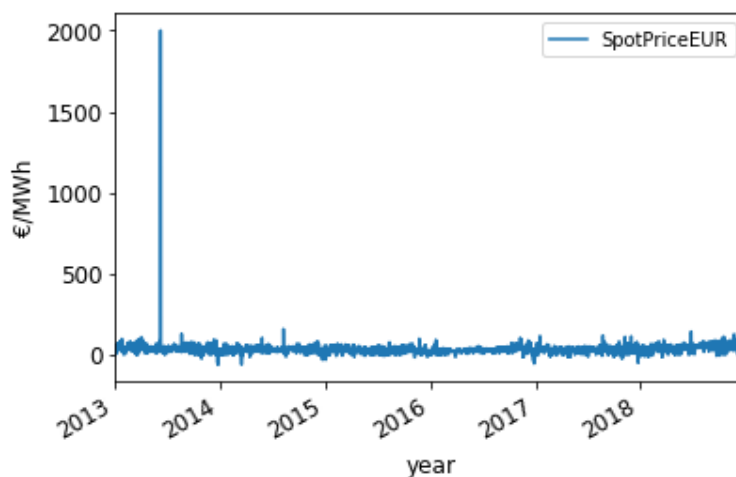
Ovisno o slučaju, razne se greške mogu javljati. U kontekstu istraživanja opisanog u ovom referatu, prvo je valjalo uskladiti podatke dohvaćene iz različitih izvora. Dva su glavna uzročnika neusklađenosti. Prvi je bio taj što su za pojedine kategorije nedostajali podatci za određene periode. To nije predstavljalo veliku prepreku, trebalo je samo otkriti koji periodi su problematični i onda uskladiti sve podatke. No tada se javio i drugi problem. Za razmatrane, valjane, periode nisu sve kategorije imale isti broj unosa. Nakon malo istraživanja kao uzročnik je otkrivena različita interpretacija prelaska na zimsko računanje vremena, odnosno na ljetno (neki bi dva puta ponovili neki sat, neki bi ga izbacili,...). Po otklanjanju tog problema podatci su postali međusobno usklađeni.

#### 3.2.2. Uređivanje podataka

Po otklanjanju svih pogrešaka posao predobrade nije gotov. Sirove podatke valja urediti u pogodnu formu za daljnju analizu. Mogli bismo to nazvati i kao fino ugađanje izgleda podataka jer su sve naredne akcije specifične upravo za potrebe konkretne analize, Za neki drugi slučaj bi pristup možda krenuo u drugom smjeru. Drugim riječima, ovo je proces kojim se ističu implicitno sadržane informacije nekog podatka, a relevantne za nastavak istraživanja.

Prvi korak je bio podijeliti podatke prema sezonama: proljeće, ljeto, jesen, zima. Tako je uz svaki zapis stajala i oznaka kojem godišnjem dobu pripada. Nadalje, podatci su bili dalje kategorizirani i prema vrsti dana: subota, nedjelja (i praznici), radni dan (7:00-18:00) te radne noći (od ponedjeljka navečer do petka ujutro). Tako je iz mnoštva međusobno nepovezanih podataka stvoren mali broj skupova koji kategorizira i objedinjuje veliku količinu podataka s nekim zajedničkim svojstvima.

Po kategorizaciji podataka, još je valjalo istražiti ima li nekih podataka koji svojom vrijednošću jako odskaku i samim time su vrijedni pažnje jer su ili greška prilikom dohvaćanja ili neki specifičan događaj. Tako je uočeno (slika 2) da je spot cijena u jednom trenutku u 2013. godini bila neuobičajeno velika.



Slika 2 - Spot cijene između 2013. i 2019.

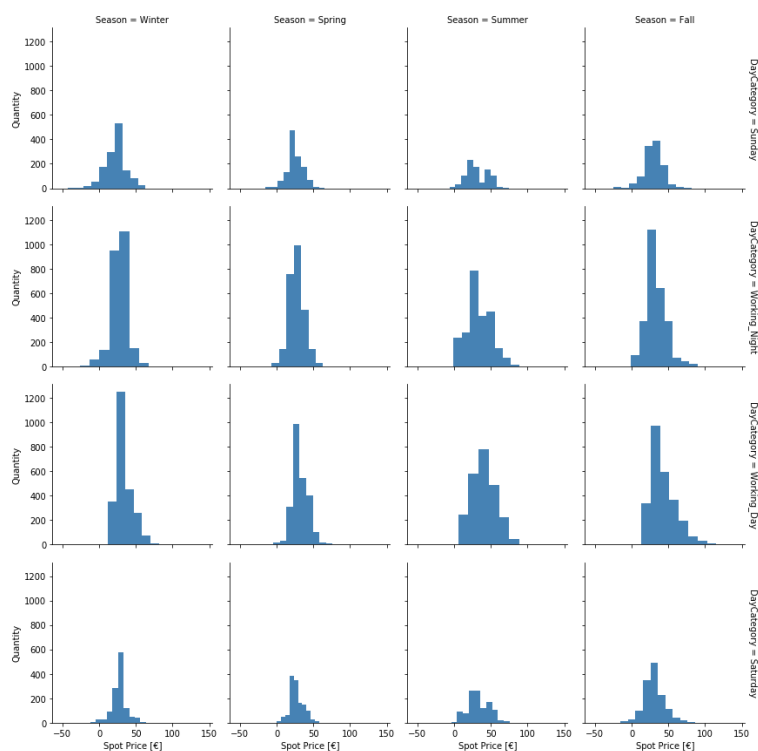
Nakon kratkog istraživanja ustanovljeno je da to nije bila greška u dohvaćanju već su tog dana dvije velike proizvodne jedinice bile van pogona uz još neke okolnosti (poput vjetera slabijeg od predviđenog) koje su se poklopile te je tako u periodu od 5 sati struja bila preko 1900€, dosežući čak i tada zakonom ograničeni maksimum od 2000€.

## 4. STATISTIČKA ANALIZA

Ovo poglavlje posvećeno je detaljnoj statističkoj analizi kojoj je za cilj ispitati eventualne korelacije između dan unaprijed tržišta i ostalih tržišta koja su dio elektroenergetskog sustava.

### 4.1. Distribucija

Kako bi promatrani set podataka bio što jasniji, prvo je potrebno odrediti njegovu distribuciju. Tada parametri pojedine distribucije također mogu puno pomoći u daljnjoj analizi.



Slika 3 - Histogram spot cijena

Graf na slici 3 sugerira možebitnu aproksimaciju normalnom (Gaussovom) distribucijom. Daljnji statistički testovi su pokazali kako je za potrebe ovog referata dovoljno točno opisati distribuciju cijena Gaussovom krivuljom. Pritom su izračunati parametri poput srednje vrijednosti, standardne devijacije i sl.

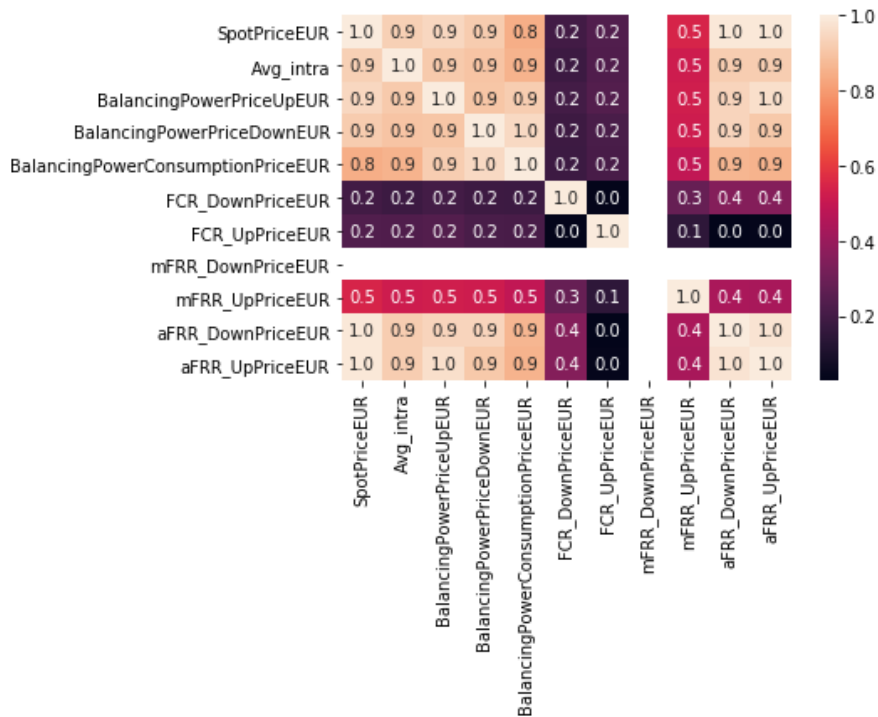
### 4.2. Korelacije

Daljnja analiza za cilj ima potvrditi ili opovrgnuti pretpostavljenu zavisnost kretanja cijena pojedinih niša elektroenergetskog tržišta. Više je formula/metoda kojom se provjeravaju korelacije među pojedinim skupovima podataka. Kako je promatrani skup podataka nalik Gaussovoj distribuciji, no s ipak ne zanemarivim odstupanjem, pregled literature je prednost dao Spearmanovoj metodi [5].

$$\rho_{rank_x, rank_y} = \frac{cov(rank_x, rank_y)}{\sigma_{rank_x} \sigma_{rank_y}} \quad (1)$$

Koeficijent može biti između -1 (potpuna negativna korelacija) i 1 (potpuna pozitivna korelacija), a 0 označava nepostojanje korelacije.

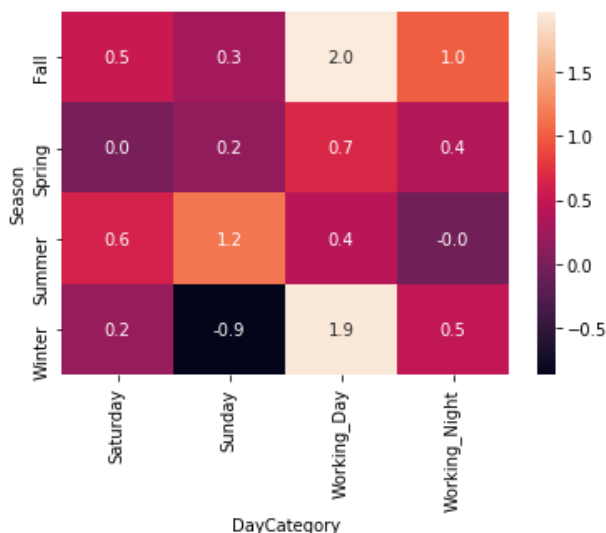
Promotrivši sliku 4, može se uočiti vrlo visoki koeficijent korelacije između pojedinih niša. Među prikazanim odnosima najveću pozornost ovaj referat obraća na povezanost unutardnevnog i dan unaprijed tržišta. Kako je koeficijent korelacije 0.9, s velikom sigurnošću se može tvrditi da se s informacijom o kretanju dan



Slika 4 - Prikaz korelacija

unaprijed tržišta, vrlo precizno može odrediti i trend kretanja unutardnevnog tržišta. Ta spoznaja će biti korištena dalje u istraživanju. Svakako vrijedi napomenuti da pronađene korelacije predstavljaju vrh sante leda u istraživačkom smislu. Potrebna su daljnja podrobna istraživanja koja bi istražila uzročno-posljedične veze međuovisnosti pojedinih niša elektroenergetskog tržišta.

### 4.3. Spread (razlika)



Slika 5 - Srednja vrijednost spreada između spot i unutardnevnih cijena

U prethodnom odlomku je ustanovljena korelacija između unutardnevnih i spot cijena. Stoga se s velikom sigurnošću može modelirati cijena na unutardnevnom tržištu znajući spot cijene ( pa i obrnuto). Promatrajući malo ponnije odnose tih dviju niša, uočava se spread (razlika) u njihovim cijenama. Postojanje takve razlike je velika prilika igračima na elektroenergetskom tržištu za postizanje ekstra profita.

Donekle logičnim i očekivanim se čini uočena razlika u cijeni, no igrači bi tu razliku mogli iskoristiti u svoju korist ako bi imali točne pretpostavke kada će biti viša jedna, odnosno druga cijena. O tome više u narednim poglavljima. Još valja ustanoviti kolike su najveće razlike i koja je vjerojatnost pojavljivanja značajnijih razlika. Promatranjem i analizom zabilježenih spreadova, dolazi se do zaključka kako se u 30% slučajeva dogodio spread veći od 5.2 €/MWh, što predstavlja nezanemarivu količinu ponavljanja. Iz toga proizlazi zaključak kako valja dublje istražiti mogućnost predikcije pojavljivanja značajnih spreadova i adekvatnog strateškog planiranja ponašanja na tržištu, posebice uzevši u obzir da ekstremne razlike sežu i do preko 50 €/MWh.

Jedan od načina određivanja kada će se dogoditi značajni spread jest određivanje nekog arbitrarnog zadovoljavajućeg iznosa razlike koji će zadovoljavati potrebe strateškog planiranja (npr. +/- 10€/MWh), te analiza kada je najvjerojatnije pojavljivanje takvih slučajeva. Dok je drugi, ponešto kompleksniji i vjerodostojniji način korištenje strojnog učenja u svrhu otkrivanja budućeg pojavljivanja značajnih spreadova prema naučenim indikatorima.

## 5. BUDUĆI TRENDOVI KRETANJA CIJENA

Prijašnji dio referata bavio se predobradom podataka te statističkom analizom. To su sve nužni preduvjeti za dobar model strojnog učenja koji će točno predviđati nekakve buduće trendove. Štoviše, statistička analiza je pokazala da se dobro provedenim predviđanjem budućih trendova putem algoritama strojnog učenja daje vrlo dobro iskoristiti uočena razlika između spot i unutardnevne cijene.

### 5.1. Strojno učenje – kratki uvod

Razlikujemo nadzirano i nenadzirano strojno učenje. Nadzirano je tip strojnog učenja u kojemu je dan trening set s ulaznim podacima, ali i izlaznim. Učenje se odvija na temelju trening seta te kasnije algoritam obrađuje nove ulazne podatke čiji izlazni nisu poznati, već ih algoritam predviđa. Dvije glavne potkategorije su regresija i klasifikacija. Regresija predviđa kontinuirane varijable, primjerice cijenu dionice. Dok klasifikacija ulazni podatak svrstava u neku od grupa, dakle diskretna stanja. Za to je dobar primjer klasifikacija ulaznih e-maileova na korisne i na spam. Za potrebe ovog referata korišteno je nadzirano učenje i to klasifikacija. Nenadzirano učenje ni u trening setu nema poznate izlazne podatke, nego samo ulazne. Takvi algoritmi imaju za zadatak pronalazak nekakvih pravilnosti i grupiranje u grupe sa zajedničkim karakteristikama.

### 5.2. Klasifikacija spreada

U prijašnjim poglavljima uočeno je nezanemarivo pojavljivanje razlike između unutardnevnih i spot cijena. Veliku priliku igračima na elektroenergetskom tržištu predstavlja ukoliko takav događaj mogu uz određenu dozu sigurnosti predvidjeti. Tu na scenu stupa nadzirano strojno učenje. Kao značajni spread odabran je onaj od +/- 10 €/MWh te je na algoritmu da uz dobar trening razvije zadovoljavajuću preciznost klasifikacije takvih događaja. Drugim riječima, izlazni podatak će biti zastavica hoće li u određenom trenutku biti značajan spread (veći od +/- 10 €/MWh) ili neće. Ulazni podatci su spot cijena, kategorija dana i sezona. Dok je izlazni podatak u obliku:

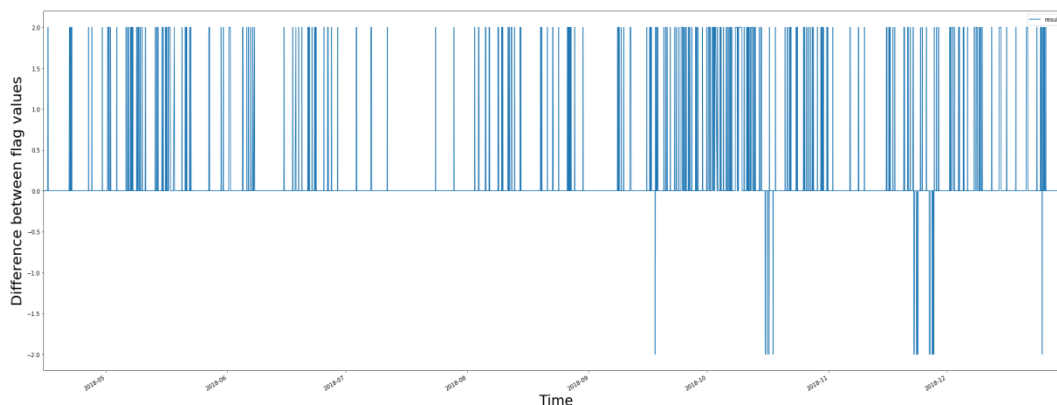
- 1 ako se značajan spread javlja
- -1 ako se ne javlja značajan spread

Po završetku treniranja algoritma strojnog učenja, na red je došla i validacija. Validacija se vršila tako što je na temelju ulaznih podataka algoritam vraća izlazne u obliku zastavice koje su se onda uspoređivale sa stvarnom realizacijom događaja u pojedinom trenutku. Grafički gledano oduzimale su se te dvije zastavice i mogla su se dogoditi sljedeća tri događaja na grafu (slika 6):

- Zastavice se podudaraju, dakle rezultat oduzimanja je „0“
- Zastavice se ne podudaraju, dakle rezultat je -2, ili 2.

Iako grafički rezultati ne izgledaju bajno, to je mala optička varka zbog mnoštva uzorka, a jedinični stupac sam po sebi mora imati neku debljinu da bi imalo bio vidljiv, koja u biti onda izgleda deblje no što bi trebala biti. Numerički gledano, u 90% slučajeva su se zastavice poklapale. Kako su svi ulazni podatci korišteni u ovoj klasifikaciji poznati barem 12h prije trenutka isporuke, ovaj rezultat može poslužiti igračima u elektroenergetskom sustavu i akademskoj zajednici kao temelje za buduće istraživanje u svrhu iskorištavanja ovakvih događaja.





Slika 6 - Validacija klasifikacije velikog spreada

## 6. ZAKLJUČAK

Cilj ovog rada je bio statistički obraditi javno dostupne podatke o tržištima elektroenergetskog sustava Danske te novonastala saznanja upotrijebiti kao podlogu daljnje istraživanje i razvijanje inovativnih poslovnih strategija i modela na tržištima električnom energijom.

Prvo je trebalo dohvatiti sve potrebne podatke. To nije bio trivijalan zadatak, no podatci su uspješno dohvaćeni te zatim predobrađeni kako bi bili u pogodnom obliku za daljnju statističku analizu.

Detaljna statistička analiza dala je pregled odnosa između različitih niša elektroenergetskog tržišta. Među ostalima, zabilježen je velik koeficijent korelacije između unutardnevnih i spot cijena. To saznanje dalo motiva za pomniju analizu međusobne ovisnosti te dvije niše. Uočeno je pojavljivanje nezanemarivih spreadova (i po visini, a i po učestalosti) te je razvijen algoritam strojnog učenja koji ima za zadatak predvidjeti kada je vrlo vjerojatno pojavljivanje značajnog spreada (proizvoljno odabrano da je granica  $\pm 10\text{€/MWh}$ ), odnosno kada nije vjerojatno. Visoka preciznost predviđanja daje za pravo produbiti istraživanje i razviti inovativne strategije ponašanja igrača na tržištu električne energije koristeći ovakve i srodne spoznaje u svrhu maksimizacije profita.

Vrijedi zaključiti kako detaljna statistička analiza i precizni modeli s kvalitetno predobrađenim javno dostupnim podacima može optimizirati operativno planiranje i otvoriti nove vidike za tržišne igrače u svrhu maksimizacije profita.

## 7. LITERATURA

- [1] J. Shikoski and V. Katic, "Deregulation of the Power Industry in Europe," *Emo.Org.Tr*, 1999.
- [2] J. Saez-Gallego, "Inverse Optimization and Forecasting Techniques Applied to Decision-making in Electricity Markets," no. June, 2016, doi: 10.13140/RG.2.2.20193.92000.
- [3] T. Kj, "Ancillary services to be delivered in Denmark," *Energinet*, no. December, pp. 1–49, 2017.
- [4] "Big Data: 20 Mind-Boggling Facts Everyone Must Read." [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/>. [Accessed: 23-Jun-2019].
- [5] "Introduction to Correlation." [Online]. Available: <https://www.datascience.com/learn-data-science/fundamentals/introduction-to-correlation-python-data-science>. [Accessed: 12-Jun-2019].
- [6] L. M. Rathke, M. J. Arentsen, and L. M. Rathke, "The Effects of Electricity Market Liberalisation in the European Union," 2015.
- [7] G. Pepermans, "European energy market liberalization: experiences and challenges," *Int. J. Econ. Policy Stud.*, vol. 13, no. 1, pp. 3–26, 2018.

- [8] P. Sorknæs, M. Henning, W. Thomas, and A. E. Anders N., “Overview of the Danish Power System and RES Integration,” *Store Proj.*, 2013.
- [9] H. E. P. Oct, H. W. A. Oct, H. E. P. Dec, and L. S. O. Dec, “Regulation C2 : The balancing market and balance settlement,” *ReVision*, no. 5452, pp. 1–17, 2008.
- [10] P. Pinson, “&quot;Ancillary services and regulation markets&quot;,” 2018.
- [11] “About - Welcome to Energi Data Service.” [Online]. Available: <https://www.energidataservice.dk/en/about>. [Accessed: 23-Jun-2019].
- [12] “API-fication,” *www.hcltech.com*, 2014.
- [13] *Comma Separated Values (CSV) Standard File Format*. Edoceo, Inc.
- [14] “Nord Pool Group.” [Online]. Available: <https://www.nordpoolgroup.com/>. [Accessed: 26-Jun-2019].
- [15] E. Vargiu and M. Urru, “Exploiting web scraping in a collaborative filtering- based approach to web advertising,” *Artif. Intell. Res.*, vol. 2, no. 1, Nov. 2012.
- [16] G. Boeing and P. Waddell, “New Insights into Rental Housing Markets across the United States: Web Scraping and Analyzing Craigslist Rental Listings,” *J. Plan. Educ. Res.*, vol. 37, no. 4, pp. 457–476, Dec. 2017.
- [17] “Beautiful Soup: We called him Tortoise because he taught us.” [Online]. Available: <https://www.crummy.com/software/BeautifulSoup/>. [Accessed: 24-Jun-2019].
- [18] “Python Data Analysis Library — pandas: Python Data Analysis Library.” [Online]. Available: <https://pandas.pydata.org/>. [Accessed: 24-Jun-2019].
- [19] “Selenium - Web Browser Automation.” [Online]. Available: <https://www.seleniumhq.org/>. [Accessed: 24-Jun-2019]. 40
- [20] “National Holidays in Denmark in 2019 | Office Holidays.” [Online]. Available: <https://www.officeholidays.com/countries/denmark/2019>. [Accessed: 24-Jun-2019].
- [21] E. M. Prices, “The Decreasing Wholesale Electricity Prices,” no. November, pp. 9–12, 2015.
- [22] T. M. (Tom M. Mitchell, *Machine Learning*. McGraw Hill, 1997.
- [23] “An introduction to machine learning with scikit-learn — scikit-learn 0.21.2 documentation.” [Online]. Available: <https://scikit-learn.org/stable/tutorial/basic/tutorial.html>. [Accessed: 25-Jun-2019].
- [24] “scikit-learn: machine learning in Python — scikit-learn 0.21.2 documentation.” [Online]. Available: <https://scikit-learn.org/stable/index.html>. [Accessed: 25-Jun-2019].
- [25] F. Santosa and W. W. Symes, “Linear Inversion of Band-Limited Reflection Seismograms,” *SIAM J. Sci. Stat. Comput.*, vol. 7, no. 4, pp. 1307–1330, Oct. 1986.
- [26] “1.1. Generalized Linear Models — scikit-learn 0.21.2 documentation.” [Online]. Available: [https://scikit-learn.org/stable/modules/linear\\_model.html#lasso](https://scikit-learn.org/stable/modules/linear_model.html#lasso). [Accessed: 25-Jun-2019].
- [27] “Gentle Introduction to Vector Norms in Machine Learning.” [Online]. Available: <https://machinelearningmastery.com/vector-norms-machine-learning/>. [Accessed: 25-Jun-2019].