Istraživanje materijala primjenom strojnog učenja

Sitar, Mia

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:559686

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-28**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

Mia Sitar

ISTRAŽIVANJE MATERIJALA PRIMJENOM STROJNOG UČENJA

Diplomski rad

Zagreb, 2019.

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

INTEGRIRANI PREDDIPLOMSKI I DIPLOMSKI SVEUČILIŠNI
STUDIJ FIZIKA I INFORMATIKA; SMJER NASTAVNIČKI

Mia Sitar

Diplomski rad

Istraživanje materijala primjenom strojnog učenja

Voditelj diplomskog rada: izv. prof. dr. sc. Goranka Bilalbegović

Ocjena diplomskog rada: _____

Povjerenstvo: 1. _____

2. _____

3. _____

Datum polaganja:_____

Zagreb, 2019.

Zahvaljujem mentorici koja mi je svojim idejama i savjetima omogućila izradu ovog diplomskog rada. Također bih željela zahvaliti kolegama i prijateljima koji su mi uljepšali i olakšali čitavo razdoblje studiranja, kao i dečku na pruženom strpljenju i razumijevanju tijekom pisanja ovog rada. Najveću zahvalu dugujem čitavoj svojoj obitelji, a posebno roditeljima, bez čije podrške ništa od ovog ne bi bilo moguće.

Sažetak

Iskorištavanje sunčeve energije predstavlja jedan od ključnih elemenata prelaska na čistu energiju. Perovskitni spojevi su medju najistraživanijim i najperspektivnijim materijalima za izradu solarnih ćelija. U ovom radu su algoritmima nadziranog strojnog učenja predviđeni energijski procjep, standardna toplina stvaranja i najstabilnija faza kristalne rešetke perovskita. Korišteni su algoritmi L_2 regularizirane linearne regresije i stabla odlučivanja. Pronađeni su perovskiti koji bi mogli biti pogodni kao gradivni materijal za solarne ćelije te su uspoređeni s onima koji se trenutno najčešće koriste u tu svrhu. U metodičkom dijelu rada opisano je nekoliko pokusa sa solarnim ćelijama koji bi se mogli izvesti u nastavi s ciljem naglašavanja važnosti solarne energije i informiranja učenika o temeljnim značajkama solarnih ćelija.

Ključne riječi: solarne ćelije, perovskiti, strojno učenje, linearna regresija, stabla odlučivanja

Study of materials using machine learning

Abstract

Exploitation of solar energy is one of the key elements for a clean energy transition. Perovskite compounds are among the most researched and promising materials for the solar cell production. In this diploma thesis supervised machine learning algorithms have been used to predict the band gap, heat of formation and the most stable phase of the crystal lattice for perovskites. Predictions were made with the L_2 regularized linear regression and decision tree algorithms. Perovskites which might be suitable for the production of solar cells were found. They have been compared with ones currently used for this purpose. In the methodical section of this thesis a few solar cell experiments have been decribed, with the goal of emphasizing the importance of solar energy and introducing students to basic features of solar cells.

Keywords: solar cells, perovskites, machine learning, linear regression, decision trees

Sadržaj

1 Uvod **1**

2 Solarne ćelije i znanost o materijalima **2**

2.1 O solarnim ćelijama 2

2.2 Perovskiti 8

3 Osnove strojnog učenja **13**

3.1 O strojnom učenju 13

3.2 Nadzirano strojno učenje 15

3.3 Linearni modeli: linearna regresija i L_2 regularizacija 17

3.4 Stabla odlučivanja 21

4 Istraživanje perovskita primjenom strojnog učenja **26**

4.1 Korišteni alati i metode 26

4.1.1 Python 26

4.1.2 Jupyter bilježnica 26

4.1.3 NumPy 27

4.1.4 Matplotlib 27

4.1.5 ASE 27

4.1.6 Scikit-learn 27

4.1.7 Graphviz 27

4.1.8 GPAW 28

4.2 Uvid u bazu podataka 28

4.3 Strojno učenje za perovskite 30

4.4 Testiranje i procjena modela 41

5 Zaključak **43**

Dodaci **45**

A Solarne ćelije u osnovnoj i srednjoj školi **45**

Literatura **49**

1 Uvod

Prošle je godine, prema istraživanju Međunarodne energetske agencije (engl. *International Energy Agency*, IEA), globalna potražnja za energijom porasla za 2,3% [1], što odgovara gotovo dvostrukom tempu rasta zabilježenom u ovom desetljeću. Jedna je četvrtina ovog porasta zastupljena potražnjom za obnovljivim izvorima energije. Poznato je da se potražnja za solarnom energijom u zadnjih nekoliko godina udvostručila. S obzirom na to da u samo 90 minuta Zemljinu površinu dosegne više sunčeve energije nego što čovječanstvo potroši tijekom jedne godine [2], nije ni čudo da stalno rastuću potrebu za energijom pokušavamo riješiti iskorištavanjem ovog neiscrpnog izvora. Na žalost, prikupljanje sunčeve energije se kroz prošlost pokazalo kao zahtjevan zadatak. Prve solarne ćelije korištene u praksi bile su građene od silicija i imale su korisnost od svega 6% [3]. Silicij je do posljednjeg desetljeća ostao najzastupljeniji materijal za proizvodnju solarnih ćelija [4]. Korisnost silicijskih solarnih ćelija je u međuvremenu porasla na 15% - 20%. S ciljem povećanja korisnosti i smanjenja troškova izgradnje solarnih ćelija istražuju se različite tehnike i materijali [5]. Perovskiti su materijali koji su se u posljednje vrijeme pokazali obećavajućima i na području cijene i na području korisnosti [6]. Prema Masačusetskom institutu za tehnologiju (MIT) perovskiti pokazuju ”virtualno neograničeni” broj kombinacija s određenom strukturom [7]. Zbog tako velikog broja mogućih kombinacija, proces otkrivanja onih koje imaju najpovoljnije karakteristike za izradu solarnih ćelija mogao bi potrajati i do 20 godina. Jedan od načina na koji bi se mogle predvidjeti kombinacije materijala koje su od interesa za daljnje istraživanje su metode strojnog učenja, polja računalne znanosti na granici sa statistikom koje u zadnja dva desetljeća napreduje zapanjujućom brzinom [8].

U ovom radu korišteni su modeli linearne regresije i stabla odlučivanja na podacima o perovskitima kako bi se otkrilo postoje li unutar tih podataka određene ovisnosti. Zatim je pomoću dobivenih modela predviđeno nekoliko kombinacija materijala koje bi mogle rezultirati perovskitom pogodnim za izradu učinkovitih i stabilnih solarnih ćelija.

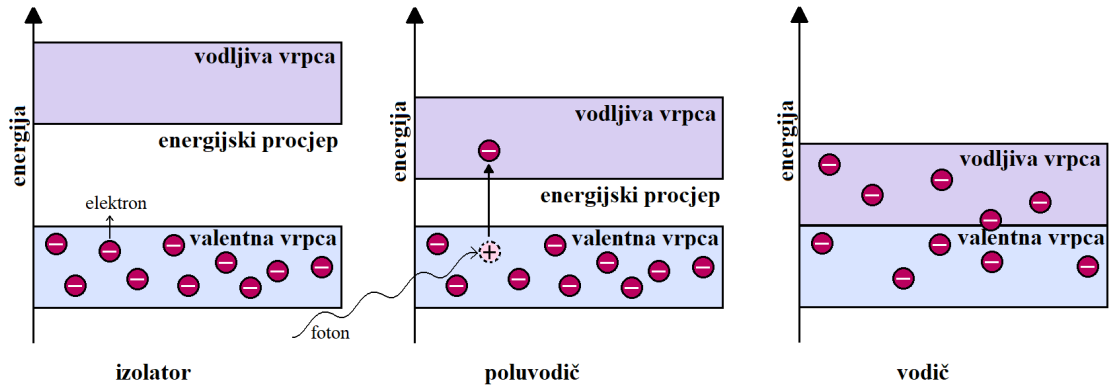
U drugom poglavlju predstavljen je princip rada solarnih ćelija i perovskitnih solarnih ćelija te njihove važnije karakteristike. Treće poglavlje posvećeno je osnovnim pojmovima i idejama strojnog učenja i nadziranog strojnog učenja te metodama linearne regresije i stabla odlučivanja. U četvrtom poglavlju dan je postupak primjene metoda strojnog učenja za istraživanje perovskita te rezultati istraživanja. Dodatak A opisuje nekoliko načina upoznavanja učenika u srednjim i osnovnim školama s osnovnim karakteristikama solarnih ćelija.

2 Solarne ćelije i znanost o materijalima

2.1 O solarnim ćelijama

Solarne ćelije su uređaji koji pomoću fotonaponskog (engl. *photovoltaic*, PV) efekta pretvaraju svjetlosnu energiju Sunca u električnu energiju [9], zbog čega ih se naziva i fotonaponskim ćelijama. Do fotonaponskog efekta dolazi kada svjetlost obasja površinu materijala u kontaktu s drugim materijalom te između njih nastaje razlika potencijala. Kao i fotoelektrični efekt, fotonaponski efekt javlja se pri sudaru fotona i elektrona u metalu, pri čemu foton predaje svoju energiju elektronu. Razlika je u tome što kod fotoelektričnog efekta elektroni napuštaju materijal i postaju slobodni, dok pri fotonaponskom efektu oni odmah ulaze u drugi materijal [10].

Fotonaponski efekt se u solarnim ćelijama događa u sloju građenom od poluvodiča, materijala koji se po svojoj sposobnosti da vode električnu struju nalaze između vodiča i izolatora. Vodiči, poluvodiči i izolatori razlikuju se po elektronskoj strukturi. Elektroni u atomu se nalaze na diskretnim energijskim razinama, koje su vrlo blizu jedna drugoj. Kada je više takvih atoma u neposrednoj blizini, njihove energijske razine su toliko blizu da ih možemo smatrati kontinuiranim pojasom. Najudaljeniji pojas od jezgre koji na temperaturi od 0 K još sadrži elektrone zove se valentni pojas ili valentna vrpca (engl. *valence band*) [11], a elektroni u njemu valentni elektroni. Pošto se valentni elektroni nalaze najdalje od jezgre, njih jezgra najmanje privlači pa im je potrebno najmanje energije da napuste atom [12]. Pojas iznad valentnog zove se vodljivi pojas (engl. *conduction band*) i to je pojas koji se nalazi najbliže jezgri, a da na temperaturi od 0 K u njemu nema elektrona. Između valentnog i vodljivog pojasa nalazi se *energijski procjep* (engl. *band gap*, E_g), odnosno zabranjeni pojas energija u kojem se ne mogu nalaziti elektroni [11]. Njegova širina jednaka je razlici donje energijske granice vodljivog pojasa (E_G) i gornje energijske granice valentnog pojasa (E_V), odnosno $E_g = E_C - E_V$ [11]. Upravo po širini ovog procjepa možemo razlikovati vodiče, poluvodiče i izolatore, što se vidi na slici 2.1. Kod vodiča energijski procjep ne postoji pa se valentna i vodljiva vrpca ili preklapaju, ili su tik jedna do druge, što znači da valentnim elektronima treba izrazito malo energije kako bi prešli u vodljivu vrpcu gdje se mogu slobodno kretati. S druge strane, energijski procjep izolatora je toliko velik da je elektronima praktički nemoguće zadobiti dovoljnu količinu energije da prijeđu iz valentne u vodljivu vrpcu. Također, kod izolatora je valentna vrpca u potpunosti popunjena elektronima i oni se unutar nje ne mogu gibati [13]. Ove dvije činjenice onemogućuju bilo kakvo gibanje elektrona (a samim



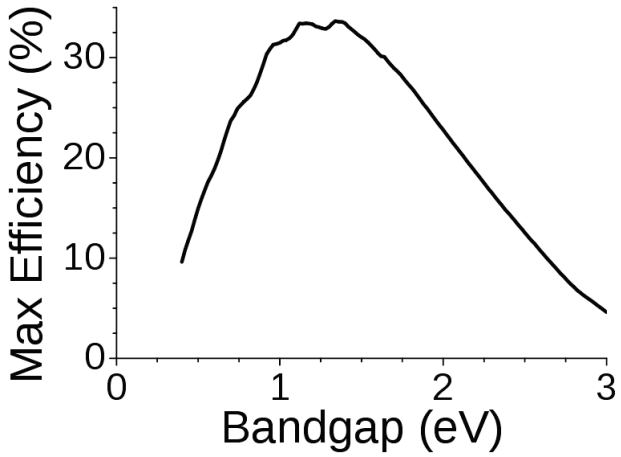
Slika 2.1: Usporedba energijskih procjepa izolatora (lijevo), poluvodiča (sredina) i vodiča (desno) te prikaz stvaranja para elektron-šupljina u poluvodiču.

tim i vođenje struje) u izolatoru. U slučaju poluvodiča energijski procjep postoji, ali je veoma malen. Na niskim temperaturama nema elektrona u vodljivoj vrpci pa se poluvodiči ponašaju kao izolatori, no s povećanjem temperature valentni elektroni počnu dobivati dovoljno energije da prijeđu u vodljivu vrpcu i da poluvodič može voditi struju, zbog čega možemo reći da električna vodljivost poluvodiča raste s porastom temperature [13].

Vrpce poluvodiča nisu ravne, već se mijenjaju ovisno o k-vektoru recipročne rešetke kristala poluvodiča [14]. S obzirom na vrijednosti k-vektora na kojima se nalaze maksimum valentne vrpce i minimum vodljive vrpce, mogu se razlikovati dva tipa energijskih procjepa. Kod direktnog energijskog procjepa se maksimum valentne vrpce nalazi na istoj vrijednosti k-vektora kao i minimum vodljive vrpce, zbog čega elektron može direktno prijeći iz valentne u vodljivu vrpcu ako mu se dovede energija E_g. Ako se pak maksimum valentne vrpce i minimum vodljive vrpce nalaze na različitim vrijednostima k-vektora, elektron i dalje treba energiju kako bi prešao iz valentne u vodljivu vrpcu, ali mora doći i do promjene k-vektora. U tom slučaju radi se o indirektnom energijskom procjepu.

Kada površinu poluvodiča obasja svjetlost, ta svjetlost može proći kroz njega, reflektirati se, ili biti apsorbirana. Ako foton koji upadne na površinu poluvodiča ima energiju manju od one energijskog procjepa, taj foton neće biti apsorbiran već će proći kroz poluvodič [14]. U slučaju u kojem dolazi do apsorpcije, foton se sudara s elektronom unutar poluvodiča, a posljedica tog sudara je prijenos energije s fotona na elektron. Ako se foton sudari s valentnim elektronom i preda mu energiju koja je jednaka onoj energijskog procjepa, valentni će elektron zadobiti dovoljno energije da prijeđe u vodljivu vrpcu. U situaciji u kojoj je energija fotona veća od energije energijskog procjepa, razlika energije pretvorit će se u toplinu, što

nije dobro za solarne ćelije zbog gubitka energije ali i utjecaja zagrijavanja na samu ćeliju. Očito će materijali užeg energijskog procjepa biti bolji za iskorištavanje svjetlosti niže energije, a materijali šireg procjepa za iskorištavanje svjetlosti više energije. Zbog ovog korisnost solarne ćelije ovisi o širini energijskog procjepa materijala od kojeg je ćelija građena. Ovisnost korisnosti o energijskom procjepu prikazana je grafom na slici 2.2. Vrh krivulje grafa definiran je kao Shockley–Queisserova (SQ) granica [15] koja iznosi 33,7%, a nalazi se na energijskom procjepu od 1,34 eV. Zbog toga se materijali čiji energijski procjepi iznose 1,5 eV smatraju najpovoljnijima za izradu solarnih ćelija.



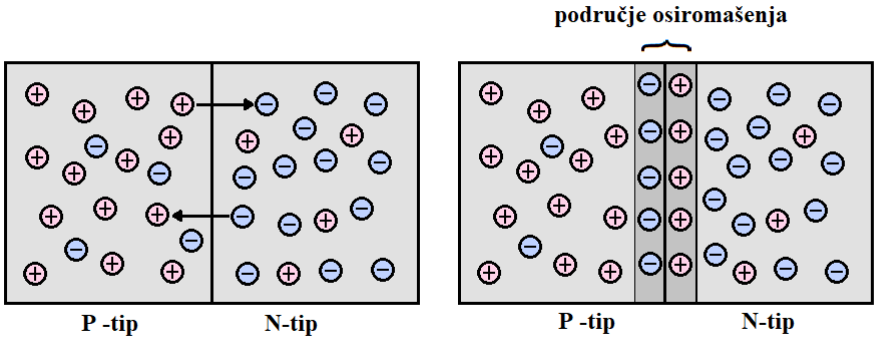
Slika 2.2: Predviđanje maksimalne teorijske korisnosti solarne ćelije kao funkcije energijskog procjepa materijala solarne ćelije [15].

Nakon što foton preda energiju elektronu te elektron prijeđe iz valentne u vodljivu vrpcu, u valentnoj vrpci će se nalaziti manjak elektrona u odnosu na stanje prije prijelaza elektrona. To možemo zamisliti kao pozitivno nabijenu ”šupljinu” na mjestu gdje je prije bio elektron. Dakle, apsorpcija fotona u poluvodiču rezultira nastankom šupljine i slobodnog elektrona, što je vidljivo na slici 2.1. Ovako nastale šupljine mogu se kretati po poluvodiču zamjenom mjesta elektrona i šupljina: ako neki elektron blizu šupljine dođe na njezino mjesto, nastat će nova šupljina na mjestu s kojeg je taj elektron došao, što se može smatrati pomicanjem šupljine. Elektron koji je napustio vodljivu vrpcu nalazi se u pobuđenom stanju [*], u kojem obično ne ostaje dugo, već se ubrzo vraća na mjesto šupljine u valentnoj vrpci, pri čemu ili emitira foton, ili prenosi energiju na druge elektrone ili šupljine. Ovaj proces nazivamo rekombinacijom elektrona i šupljine. Potencijalna energija para elektron-šupljina može biti iskorištena za dobivanje električne struje, no kako bi se to postiglo rekombinacija elektrona i

[*] Stanje u kojem se elektron ne nalazi na najmanjoj mogućoj energijskoj razini zove se pobuđeno stanje [16].

šupljine mora biti spriječena. To se postiže ugradnjom potencijalne barijere koja funkcionira poput dvije polupropusne membrane od kojih jedna propušta samo elektrone, a druga samo šupljine, na suprotnim stranama solarne ćelije [14].

Promotrimo nastanak potencijalne barijere na primjeru silicija, koji se koristi za izradu većine solarnih ćelija. Atom silicija ima četiri valentna elektrona. Kristal silicija formira se na način da svaki atom silicija stvara vezu sa svakim od svoja četiri najbliža susjeda, pri čemu u jednoj vezi sudjeluju dva valentna elektrona, po jedan od svakog atoma. Zamislimo da je u takav kristal ubačena nečistoća u obliku jednog atoma fosfora, koji ima pet valentnih elektrona. Taj bi atom zauzeo mjesto atoma silicija i bio povezan s još četiri atoma silicija, pri čemu bi četiri njegova valentna elektrona bila dio veze. Jedan preostali elektron bio bi slobodan. Ovakav elektron može na sobnoj temperaturi zadobiti dovoljno energije da prijeđe u vodljivu vrpcu. Postupak ubacivanja nečistoća u kristal zovemo dopiranjem kristala, a za nečistoće kojima se u kristal donosi po jedan valentni elektron možemo reći da doniraju slobodni elektron kristalu te ih zovemo donorima. Kristal koji je dopiran određenim brojem donora ima slobodne negativne naboje i zovemo ga N-tipom kristala. Na isti način može se definirati P-tip kristala sa slobodnim pozitivnim nabojima, u kojem su nečistoće atomi s jednim valentnim elektronom manjka. Recimo da je u kristal silicija ubačen atom bora s tri valentna elektrona. On će formirati tri potpune veze, a za četvrtu će mu nedostajati elektron, odnosno nastat će šupljina. Pošto šupljine prihvaćaju elektrone, ovakve nečistoće zovu se akceptori. U N-tipu materijala elektroni se smatraju većinskim nosiocima naboja, a šupljine manjinskim, dok je za P-tip obratno [9]. Spoj N-tipa i P-tipa poluvodiča zove se PN spoj. U PN spoju će na području oko spojnice P-tipa i N-tipa slobodni elektroni iz N-tipa prijeći u P-tip i tamo se rekombinirati sa slobodnim šupljinama, što je prikazano na slici 2.3. Elektroni iz kristalnih veza N-tipa također se mogu rekombinirati sa šupljinama P-tipa, što je ekvivalentno prijelazu šupljine iz P-tipa u N-tip. Dakle, u P-tipu će nastati višak negativnog naboja, a u N-tipu višak pozitivnog naboja. Posljedica ovih prijelaza je da se nosioci naboja koji su prešli u drugi tip materijala tamo više ne mogu slobodno kretati jer su sada dio veze u kristalu. Što se više nosioca naboja jedne vrste skuplja u određenom tipu materijala, to se na spojnici stvara jače električno polje. Ovo električno polje predstavlja potencijalnu barijeru koja se protivi daljnjem prolazu elektrona u P-tip i šupljina u N-tip (slika 2.3), a zove se područje osiromašenja (engl. *depletion region*) jer u njemu nema pokretljivih nosioca naboja [14]. Područje osiromašenja ne sprječava prolazak šupljina iz N-tipa u P-tip, kao ni prolazak elektrona iz P-tipa u N-tip. No, pokretnih šupljina u N-tipu i elektrona u P-tipu

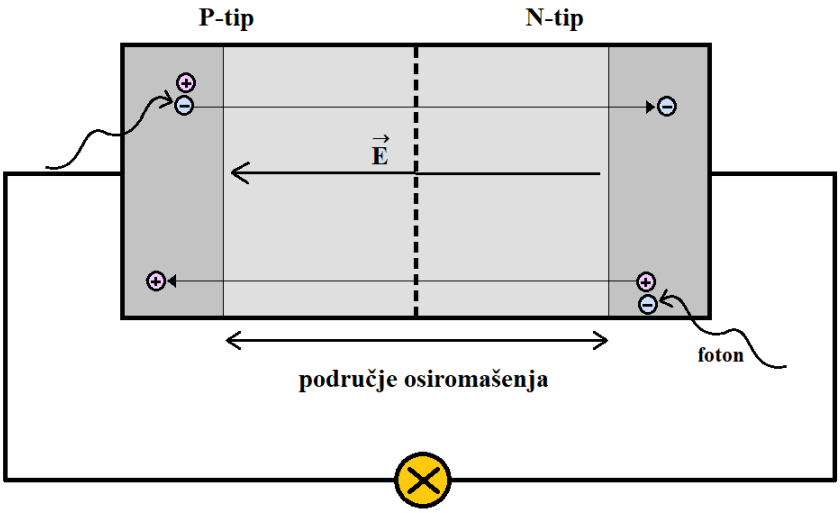


Slika 2.3: Nastanak područja osiromašenja u PN spoju.

ima toliko malo da to možemo zanemariti. Kada je područje osiromašenja dovoljno veliko, sustav je u ravnoteži i nosioci naboja više ne mogu prijeći s jedne na drugu stranu barijere.

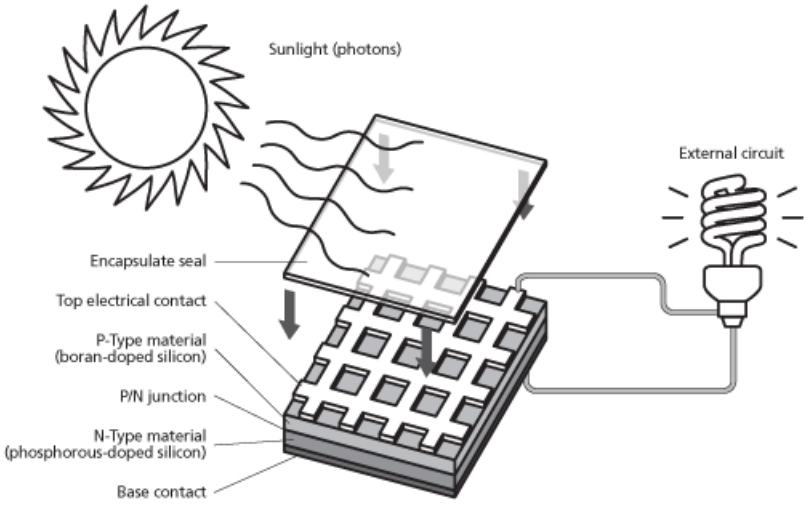
Zamislimo da PN spoj obasja svjetlost i na strani N-tipa poluvodiča foton preda elektronu dovoljno energije za njegovo oslobađanje i tvorbu para elektron-šupljina (slika 2.4). Elektron će ostati u N-tipu jer ne može proći kroz područje osiromašenja, a šupljina će doći do područja osiromašenja, kroz koje će biti ubrzana u P-tip poluvodiča. Kada se jednom nađe u P-tipu s izrazito malo pokretnih elektrona, postoji minimalna vjerojatnost za rekombinaciju, a zbog potencijalne barijere šupljina se više ne može vratiti u N-tip. Analogno tome, kada do tvorbe para elektron-šupljina dođe u P-tipu poluvodiča, elektron prolazi u N-tip i ostaje tamo, a šupljina mora ostati u P-tipu. Na ovaj se način u N-tipu skuplja višak negativnih naboja, a u P-tipu višak pozitivnih naboja, što stvara razliku potencijala između P-tipa i N-tipa poluvodiča. Da su ove dvije vrste materijala povezane vanjskim strujnim krugom kroz krug bi potekla električna struja, a elektroni bi se nakon prolaska kroz strujni krug rekombinirali sa šupljinama.

Najveći broj solarnih ćelija koje su trenutno u uporabi temelji se upravo na PN-spoju poluvodiča. Solarne ćelije imaju i neke dodatne slojeve. PN-spoj mora biti smješten između kontaktnih slojeva kako bi se mogla provoditi struja. U idealnom slučaju, kontaktni slojevi pokrivali bi cijelu površinu ćelije kako bi otpor bio minimalan. To je moguće realizirati na donjoj strani ćelije, koja nije okrenuta prema Suncu, no pokrijemo li cijelu gornju površinu ćelije tamnim kontaktnim slojem, apsorpcija sunčeve svjetlosti neće biti moguća. Zbog toga je gornji kontaktni sloj dizajniran kao rešetka čije su trake dovoljno široke da dobro provode električnu energiju, ali i dovoljno tanke da ne blokiraju svjetlost. Kako bi ćelija bila zaštićena od vanjskih uvjeta, gornji se kontaktni sloj obično prekriva zaštitnim slojem od stakla, koje



Slika 2.4: Fotonaponski efekt u PN spoju.

je najčešće premazan antirefleksnim slojem zbog povećanja korisnosti ćelije. Što se manje svjetlosti reflektira, to će se više svjetlosti apsorbirati i pretvoriti u električnu energiju. Osnovni slojevi tipične solarne ćelije prikazani su na slici 2.5.



Slika 2.5: Konstrukcija solarne ćelije [17].

Napon koji je moguće dobiti jednom silicijevom solarnom ćelijom iznosi između 0.5 i 0.6 V [18], što se može primijeniti za napajanje manjih igračaka, satova i džepnih kalkulatora, no za većinu uređaja koji se koriste u današnje vrijeme potreban je veći napon i struja. Kako bi se to postiglo, više se ćelija spaja u seriju, ili paralelno. U paralelnom spoju solarnih ćelija spajaju se istoimeni terminali ćelija: pozitivan terminal jedne ćelije s pozitivnim terminalnom druge i obratno. Svaka ćelija u paralelnom spoju ima isti napon i taj je napon

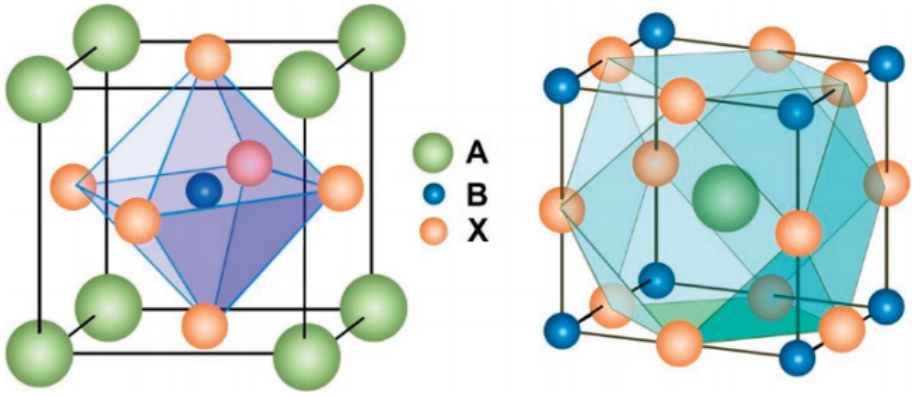
jednak ukupnom naponu strujnog kruga, dok je jakost struje u strujnom krugu jednaka zbroju jakosti struja pojedinih ćelija. Ako se solarne ćelije povezuju serijski, to znači da su im povezani raznoimeni terminali: negativan terminal jedne s pozitivnim terminalom druge ćelije. U tom slučaju kroz svaku ćeliju prolazi struja iste jakosti, a ukupni napon strujnog kruga jednak je zbroju napona svih povezanih ćelija. Više solarnih ćelija povezanih zajedno čini solarni modul (engl. *solar module*). Moduli mogu imati između 10 i 100 solarnih ćelija, a tipični moduli su sastavljeni od 36 serijski spojenih ćelija i imaju napon od 12 V. Ako se više modula poveže paralelno i/ili serijski, dobije se solarna ploča (engl. *solar panel*), a njihovom kombinacijom dolazi se do solarnih polja (engl. *solar array*), koja za većinu kućanstava sadrže 10 do 20 solarnih ploča [9].

Preko 80% solarnih panela korištenih u svijetu radi na principu solarnih ćelija od silicija, kojima se može postići visoka korisnost pretvorbe energije, ali im je i cijena visoka. Zbog toga su se počeli tražiti alternativni načini proizvodnje solarnih ćelija, koji uključuju različite tehnike i materijale. Prema načinu proizvodnje, solarne ćelije možemo podijeliti u tri skupine, odnosno generacije [19]. U prvu generaciju spadaju već spomenute silicijeve ćelije, od kojih razlikujemo monokristalne, polikristalne, amorfne i hibridne [20]. Ćelije druge generacije često se nazivaju i tankoslojnim (engl. *thin film*) jer su im slojevi puno tanji od onih prve generacije. One su najčešće rađene od amorfnog silicija, kadmij telurida (CdTe) i bakar indij galij selenida (CIGS). Troškovi izrade su im puno manji, no nažalost i korisnost, zbog čega je nastala i treća generacija solarnih ćelija, čije je istraživanje i razvoj trenutno u tijeku. U treću generaciju ubrajaju se ćelije sulfida bakra, cinka i kositra, bojom senzibilizirane, organske i polimerne ćelije, ćelije s kvantnim točkama i perovskitne solarne ćelije (engl. *Perovskite Solar Cell*, PSC). S obzirom na broj istraživanja rađen na pojedinom tipu ćelija, perovskitne solarne ćelije mogu se smatrati najbrže rastućom tehnologijom treće generacije solarnih ćelija [19].

2.2 Perovskiti

Izraz ”perovskiti” originalno se odnosio na vrstu minerala koji su otkriveni na planini Ural u Rusiji, a ime su dobili po ruskom mineralogu Levu Alekseevichu Perovskom. Minerali perovskita građeni su od kisika, kalcija i titanija, s formulom $CaTiO_3$ [21]. Za bilo koji kemijski spoj čija je kristalna struktura jednaka onoj minerala perovskita i ima kemijsku formulu oblika ABX_3, gdje A i B predstavljaju pozitivno nabijene ione (katione), a X negativno nabijeni ion (anion), možemo reći da ima perovskitnu strukturu. U ovom se radu

izraz perovskit odnosi na svaki spoj koji ima strukturu perovskita, a ne na mineral. Kristalnu strukturu perovskita, prikazanu na slici 2.6, najjednostavnije je zamisliti kao kocku u čijem je središtu kation B, dok je u svakom vrhu kocke kation A, a u središtu svake strane kocke nalazi se anion X [22]. Ovu strukturu nazivamo oktaedralnom BX_6 strukturom jer šest X aniona čini oktaedar u čijem je središtu kation B. Promatranjem A kationa kao onog u središtu kocke može se dobiti kuboktaedralna AX_{12} struktura. U tom se slučaju u svakom vrhu kocke nalazi kation B, dok je po jedan anion X smješten na središtu svakog brida kocke. Ove su dvije strukture ekvivalentne. Za većinu perovskitnih materijala korištenih za solarne ćelije

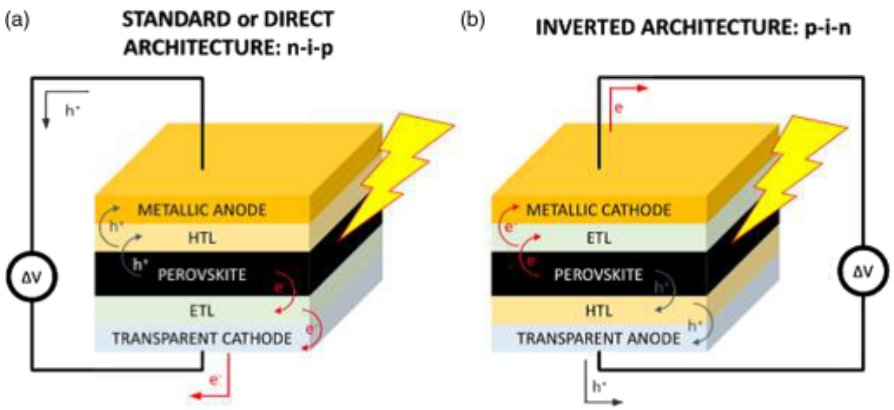


Slika 2.6: Struktura kristala perovskita [23].

upotrebljava se kombinacija organskog kationa (najčešće metilamonij, CH_3NH_3 ili formamidinij, $NH_2CHNH_2{}^+$) kao A, velikog anorganskog kationa (najčešće olovo, $Pb_2{}^+$) kao B i manjeg halogenskog aniona (najčešće klor, Cl^- ili jod, I^-) kao X_3 [22].

Struktura perovskitnih solarnih ćelija razlikuje se od strukture temeljene na PN spoju poluvodiča koji smo do sada razmatrali, no princip rada je vrlo sličan. Dvije osnovne strukture perovskitnih ćelija zasnivaju se na NIP i PIN spojevima, u kojima je između poluvodiča N-tipa i P-tipa smješten intrinzičan poluvodič (poluvodič I-tipa). To je nedopiran poluvodič karakteriziran jednakim brojem slobodnih elektrona i šupljina. Kod NIP strukture sloj N-tipa poluvodiča najbliže je sunčevoj svjetlosti, ispod njega je sloj I-tipa, a na kraju dolazi P-tip. Kod PIN strukture slojevi su posloženi obrnuto, dakle prvo sloj P-tipa, zatim I-tipa te na kraju sloj N-tipa poluvodiča. Zbog toga se NIP struktura često naziva standardnom, a PIN invertiranom strukturom. Intrinzični sloj obično je puno deblji od slojeva N-tipa i P-tipa. Pošto u N-tipu postoji višak slobodnih elektrona u odnosu na šupljine, a u P-tipu višak slobodnih šupljina u odnosu na elektrone, elektroni iz N-tipa i šupljine iz P-tipa prelaze u I-tip i tamo

se međusobno rekombiniraju. Time se stvaraju dva područja osiromašenja: u P-tipu poluvodiča oko spojnice s I-tipom nastaje područje osiromašenja s viškom negativnih naboja, dok u N-tipu oko spojnice s I-tipom nastaje područje osiromašenja s viškom pozitivnih naboja. Ovo uzrokuje nastanak električnog polja u spoju poluvodiča. Prelazak pokretljivih nosioca naboja iz P-tipa i N-tipa u I-tip odvija se tako dugo dok električno polje ne postane dovoljno jako da ga zaustavi, čime dolazi do ravnoteže. Kada se I-tip osvijetli i apsorpcijom fotona u njemu dođe do tvorbe para elektron-šupljina, elektron će biti privučen u N-tip, a šupljina u P-tip poluvodiča, gdje se više ne mogu rekombinirati. Zbog toga se sloj N-tipa naziva i slojem za prijenos elektrona (engl. *Electron Transporting Layer*, ETL), a sloj P-tipa slojem za prijenos šupljina (engl. *Hole Transporting Layer*, HTL). Kod perovskitnih solarnih ćelija, perovskiti imaju ulogu I-tipa poluvodiča. Slojevi NIP i PIN strukture perovskitnih ćelija prikazani su na slici 2.7. Struktura na slici 2.7 zove se planarna struktura, a osim nje postoji još



Slika 2.7: Standardna planarna (a) i invertirana planarna (b) struktura perovskitne solarne ćelije [24].

i mezoskopska struktura u kojoj se između ETL-a i HTL-a nalazi sloj mezopornog [**] metalnog oksida koji sadrži perovskit. Izrađene su i perovskitne ćelije bez ETL-a, ili bez HTL-a. Možemo razlikovati šest struktura perovskitnih solarnih ćelija: planarna NIP, planarna PIN, mezoskopska NIP, mezoskopska PIN, struktura bez ETL-a i struktura bez HTL-a [19].

Razlog naglog interesa za perovskitne solarne ćelije može se pronaći u činjenici da je njihova korisnost pretvorbe energije u zadnjih nekoliko godina doživjela ogroman porast u usporedbi s većinom ostalih tehnologija. Prema podacima Nacionalnog laboratorija za obnovljive izvore energije (engl. *National Renewable Energy Laboratory*, NREL) [26], korisnost perovskitnih ćelija trenutno nadilazi korisnost svih tehnologija tankoslojnih ćelija,

[**] Mezoporni materijali su materijali koji sadrže pore promjera između 2 i 50 nm [25].

kao i silicijskih ćelija. Važno je napomenuti da su ti rezultati dobiveni u laboratoriju, gdje se testiranja vrše na 1 cm^2 uzorka, a na većoj površini uzorka korisnost se smanjuje, što nije slučaj za silicijske ćelije [6]. Ipak, s obzirom na to da je prva perovskitna solarna ćelija izrađena tek 2009. godine, ovi su rezultati obećavajući. Također, proizvodnja perovskitnih solarnih ćelija jednostavnija je i jeftinija od proizvodnje silicijskih ćelija. Najveći nedostatak perovskitnih ćelija je njihova nestabilnost, koja se javlja kao rezultat vanjskih, ali i unutarnjih faktora. Predviđeni životni vijek komercijalnih solarnih ćelija je 25 godina, dok je najduži zabilježeni životni vijek perovskitnih ćelija jedna godina [27]. U vodeće vanjske faktore koji utječu na stabilnost ćelija ubrajaju se osjetljivost na vlagu i kisik, što se djelomično može riješiti zaštitnim slojevima na ćeliji. Veći je problem u unutrašnjoj nestabilnosti koja je posljedica same strukture i svojstava perovskita. Za materijale koji se na sobnoj temperaturi neće spontano raspasti kažemo da su termodinamički stabilni [28]. Termodinamička stabilnost može se povezati sa *standardnom toplinom stvaranja* (engl. *heat of formation*), koja se definira kao količina energije koja je apsorbirana ili proizvedena pri stvaranju jednog mola materijala iz elemenata u slobodnom stanju na sobnoj temperaturi. Standardna toplina stvaranja za neki materijal označava se s ΔH_f°, a jednaka je promjeni entalpije pri stvaranju spoja iz kemijskih elemenata. Ako je standardna toplina stvaranja pozitivna, to znači da je pri stvaranju spoja potrebno elementima dovesti energiju, a da se pri njegovom raspadu energija oslobađa. Takav će materijal biti nestabilan. S druge strane, ako je standardna toplina stvaranja negativna, za raspad spoja potrebna je energija, pa je taj spoj više ili manje stabilan, ovisno o količini energije koja je potrebna za raspad. Dakle, što je standardna toplina stvaranja nekog materijala manja, to je materijal termodinamički stabilniji. Termodinamička stabilnost perovskita, između ostalog, ovisi o njihovoj strukturi [29]. Kao što je prije spomenuto, perovskitima se smatraju spojevi oblika ABX_3 čija je kristalna struktura kubična. Različite veličine kationa A i B te aniona X mogu u kristalnoj rešetci perovskita uzrokovati tzv. oktaedralno naginjanje (engl. *octaedral tilting*), čime se može promijeniti oblik kristalne rešetke. Ovo je opisano faktorom tolerancije t:

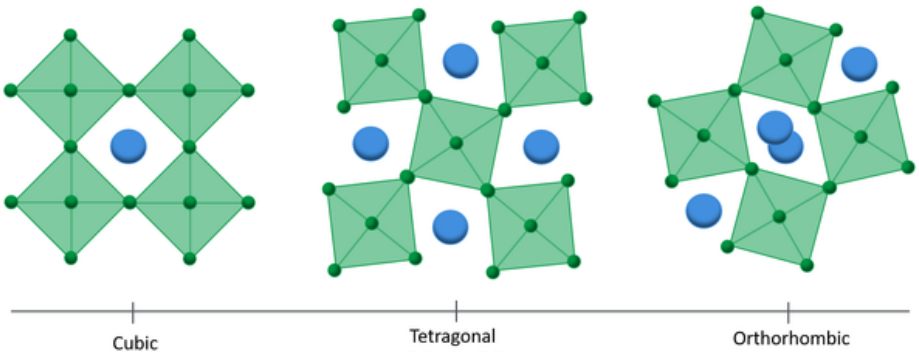
$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}, \tag{2.1}$$

pri čemu su r_A, r_B i r_X radijusi kationa A i B te aniona X. Prema iznosu faktora t moguće je razlikovati tri faze perovskita, prikazane na slici 2.8 [30, 31]:

- $t \approx 1$: perovskit se nalazi u kubičnoj (α) fazi, duljine stranica kristalne rešetke su

$a = b = c$,

- $0,9 > t > 0,7$: perovskit se nalazi u tetragonskoj (β) fazi s duljinama stranica kristalne rešetke $a = b \neq c$, ili u ortorompskoj (γ) fazi s duljinama stranica $a \neq b \neq c$.



Slika 2.8: Oblik kristalne rešetke u tri različite faze perovskita. Plavom bojom označeni su kationi A, a zelenom anioni X [30].

Oktaedralno naginjanje može biti uzrokovano i promjenom temperature na kojoj se perovskit nalazi, pa je tako za perovskit $CsSnI_3$ zamijećena ortorompska faza na temperaturi nižoj od 352 K, dok na temperaturi od 352 K dolazi do prijelaza faze u tetragonsku, te u kubičnu na temperaturi od 431 K [31]. Promjena faze utječe i na energijski procjep: uočeno je da se za isti perovskit širina energijskog procjepa smanjuje povećanjem simetrije kristala pa tako perovskit u kubičnoj fazi ima najmanji energijski procjep, dok u ortorompskoj fazi ima najveći [32]. Perovskit u različitim fazama ima i različitu termodinamičku stabilnost: za $CsSnI_3$ je utvrđeno da je najstabilniji (ima najnižu toplinu stvaranja) u ortorompskoj fazi, dok je najnestabilniji u kubičnoj fazi [33]. Uz prijelaze faza, kod perovskita je moguć i prijelaz energijskog procjepa iz indirektnog u direktni, što je zamijećeno pri promjeni hidrostatskog tlaka na kojem se perovskit nalazi [34]. Dakle, perovskiti imaju i direktan i indirektan energijski procjep, između kojih nema velike razlike: zabilježena razlika indirektnog i direktnog procjepa za $MAPbI_3$ je ≈ 60 meV [31].

Dodatni problem za korištenje perovskita kao materijala za solarne ćelije je činjenica da je jedan od najzastupljenijih elemenata u perovskitima olovo, čije ispuštanje može imati dugoročne negativne posljedice za okoliš, ali i za ljude [6]. Zbog toga bismo pri potrazi za perovskitima povoljnim za solarne ćelije trebali analizirati materijale koji ne sadrže olovo.

3 Osnove strojnog učenja

3.1 O strojnom učenju

U posljednjih nekoliko godina suočavamo se s ogromnim porastom količine podataka (engl. *Big Data*) [35]. Prema IBM-ovom istraživanju, u 2015. godini su 90% podataka u svijetu sačinjavali podaci nastali između 2014. i 2015. godine, te je procijenjeno da svaki dan u svijetu nastane 2.4 kvintilijuna bajtova podataka. Obrada golemih brojeva podataka od velike je važnosti u mnogim područjima poput biologije, astronomije i društvenih znanosti [36], no zbog kontinuiranog povećanja broja podataka za obradu, pokazala se gotovo nemogućim zadatkom za ljude. Na sreću, računalima to ne predstavlja problem. Pri računalnoj obradi velikog broja podataka često se koriste metode strojnog učenja (engl. *Machine Learning*, ML) [8] u kojima algoritmi uče na podacima koje obrađuju te su na temelju naučenog sposobni samostalno donijeti određene zaključke ili predviđanja. Strojno je učenje implementirano u mnoge alate koje često koristimo u svakodnevnom životu, kao što su preporuke filmova na Netflixu, pjesama na Youtubeu ili prijatelja na Facebooku, procjena cijene vožnje taksijem te virtualni asistenti poput Siri, Alexe ili Google asistenta.

Kad kažemo da algoritam uči iz podataka, možemo se zapitati kako je moguće da algoritam uči i što to znači u računalnom svijetu. Prema definiciji koju je 1997. dao Mitchell [37], "računalni program uči iz iskustva E, uzimajući u obzir klasu zadataka T i mjeru rezultata P, ako se njegovi rezultati na zadacima u T, izmjereni pomoću P, povećavaju s iskustvom E." Za primjer uzmimo program koji za dani nalaz tumora prepoznaje radi li se o zloćudnom ili dobroćudnom tumoru (zadatak T). Kako bismo to postigli moramo primijeniti algoritam strojnog učenja kojem ćemo dati određeni broj nalaza te ćemo za svaki od njih definirati prikazuje li zloćudni ili dobroćudni tumor (iskustvo E). Koliko je učenje bilo uspješno vidjet ćemo prema točnosti s kojom program na novim, njemu do tada nepoznatim nalazima određuje vrstu tumora (rezultat P).

Metode strojnog učenja primjenjive su na mnogim računalnim problemima, ali često zahtijevaju puno vremena i resursa. Važno je znati procijeniti je li strojno učenje uistinu najbolji način za rješavanje zadanog problema, ili je praktičnije poslužiti se nekom drugom metodom. Primjena strojnog učenja preporuča se u situacijama u kojima radimo s kompliciranim pravilima ili jednadžbama, kada se uvjeti zadataka često mijenjaju, ili pak kada se mijenja priroda podataka za obradu. Neki od najčešćih tipova problema koji se rješavaju strojnim učenjem su [38]:

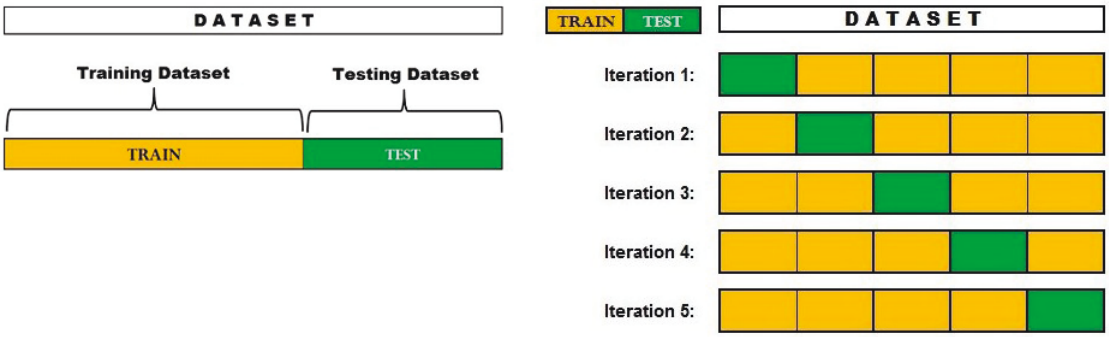
- Klasifikacija: predviđanje diskretne varijable, odnosno kategorije kojoj podatak pripada. Neki od primjera klasifikacije su određivanje pripada li pristigla e-mail poruka u neželjenu poštu ili ne, kao i već spomenuta kategorizacija tumora.
- Regresija: predviđanje kontinuirane varijable, recimo predviđanje količine oborina na temelju temperature, relativne vlažnosti i tlaka zraka.
- Klasteriranje: pronalaženje grupa, to jest klastera u zadanom skupu podataka. Koristi se kod preporuke proizvoda kao što su filmovi slični onom koji smo nedavno gledali, pjesme koju možda želimo slušati ili proizvoda koji se često kupuju uz onaj koji smo upravo naručili.
- Redukcija dimenzije: smanjenje broja varijabli za obradu. Može se ostvariti kroz dva pristupa: selekciju varijabli ili ekstrakciju varijabli.

U primjeni strojnog učenja izrazito je važan način na koji su podaci predstavljeni računalu [39]. To je najzgodnije raditi u obliku tablice u kojoj svaki redak predstavlja jedan podatkovni uzorak, a svaki stupac jedno svojstvo, odnosno varijablu tog podatkovnog uzorka. Cilj svakog algoritma strojnog učenja je za zadani skup ulaznih varijabli jednog podatkovnog uzorka moći predvidjeti izlaznu varijablu tog podatkovnog uzorka. Moguće je razlikovati dvije tehnike kojima se to postiže [38]: nadzirano strojno učenje (engl. *Supervised Machine Learning*) i nenadzirano strojno učenje (engl. *Unsupervised Machine Learning*). Pri nadziranom strojnom učenju predviđanje izlazne varijable za zadane ulazne temelji se na parovima ulaznih i izlaznih varijabli koji su prethodno dani algoritmu. U nadzirano strojno učenje spadaju tehnike klasifikacije i regresije. Tehnike klasteriranja i redukcije dimenzije spadaju u nenadzirano strojno učenje, kod kojeg su algoritmu dane samo ulazne varijable, o kojima on donosi određene zaključke te na temelju tih zaključaka predviđa izlazne varijable. Postoji i polunadzirano strojno učenje (engl. *Semi-Supervised Machine Learning*) [40] koje se nalazi između nadziranog i nenadziranog. Najčešće su to problemi u kojima su algoritmu za dio podatkovnih uzoraka dani parovi ulaznih i izlaznih varijabli, a za dio samo ulazne varijable. U ovom radu korištene su metode nadziranog strojnog učenja.

3.2 Nadzirano strojno učenje

Kod metoda nadziranog strojnog učenja za jedan se podatkovni uzorak definiraju ulazne varijable x_1, x_2, ..., x_n i izlazna varijabla y. Pomoću algoritma se aproksimira funkcija koja ulaznim varijablama pridružuje zadane izlazne, $y = f(\boldsymbol{x})$ [41], pri čemu vektor $\boldsymbol{x}$ označava skup ulaznih varijabli, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$. U primjeru predviđanja količine oborina iz temperature, vlažnosti i tlaka, količina oborina predstavlja izlaznu varijablu (y), dok temperatura, vlažnost i tlak označavaju ulazne (x_1, x_2, x_3). Želimo da funkcija pridruživanja bude aproksimirana toliko dobro da pomoću nje možemo točno predvidjeti izlaznu varijablu za nove, nepoznate ulazne varijable. Ovaj proces može se usporediti s procesom učenja uz pomoć učiteljskog nadzora, po čemu i nosi ime nadziranog učenja: učitelj zna točne odgovore, dok ih učenik predviđa na temelju onog što je prije naučio. Učitelj ispravlja učenikova predviđanja tako dugo dok nije zadovoljan njegovim rezultatima.

Kako bismo znali hoće li model biti uspješan u predviđanju izlaznih varijabli, moramo na neki način provjeriti njegovu točnost. To se može napraviti podjelom podataka za koje su nam poznate i ulazne i izlazne varijable na dva dijela: skup za učenje (engl. *training set*) i ispitni skup (engl. *test set*) [39]. Funkcija pridruživanja određuje se pomoću parova ulaznih i izlaznih varijabli iz skupa za učenje te na temelju toga donosi predviđanja izlaznih varijabli za ulazne iz ispitnog skupa. Ta predviđanja možemo usporediti sa stvarnim izlaznim varijablama skupa za testiranje kako bismo doznali točnost modela. Ova se metoda zove metodom izdvajanja (engl. *holdout method*). Podjela podataka se često vrši na način da 75% podataka budu u skupu za učenje, a 25% u ispitnom skupu. Na taj se način zapravo "gubi" velika količina podataka, jer podatke na kojima se model ispituje ne možemo koristiti za izradu modela, a cilj nam je izgraditi model na što većem broju podataka kako bi bio što točniji. Također, rezultat ove metode ovisi o načinu na koji se podaci dijele na skup za učenje i ispitni skup. Nešto bolja metoda provjere točnosti modela je k-struka unakrsna provjera (engl. *k-fold cross-validation*). Za korištenje ove metode podaci moraju biti podijeljeni na k dijelova, odnosno particija približno jednake veličine, pri čemu je k broj koji korisnik sam određuje, najčešće 5 ili 10. Pri prvom prolazu se prva particija koristi kao ispitni skup, a ostale kao skup za učenje, dok se pri drugom prolazu druga particija koristi kao ispitni skup, a ostale kao skup za učenje. Postupak se ponavlja ukupno k puta (pri čemu je kod i-tog prolaza i-ta particija ispitni skup, a ostale skup za učenje), te se za svaki prolaz dobije točnost modela. Prosječna točnost modela računa se kao srednja vrijednost dobivenih k točnosti. Usporedba metode izdvajanja i k-struke unakrsne provjere prikazana je na slici 3.1.



Slika 3.1: Usporedba provjere točnosti modela pomoću metode izdvajanja (lijevo) [42] i k-struke unakrsne provjere (desno) [43].

Za model koji donosi točne pretpostavke na prethodno neviđenim ulaznim varijablama kaže se da generalizira sa skupa za učenje na ispitni skup [44]. Generalizacijska ili ispitna pogreška predstavlja pogrešku modela na ispitnom skupu, odnosno razliku u pretpostavkama modela i stvarnim podacima ispitnog skupa. Pogreška učenja je, s druge strane, pogreška modela na skupu za učenje. Cilj strojnog učenja je da obje pogreške budu što manje. Ako je pogreška učenja prevelika, dolazi do podnaučenosti modela (engl. *underfitting*), što znači da model na podacima za učenje nije naučio dovoljno te da neće moći donijeti dovoljno dobru pretpostavku na ispitnim podacima. Ako je pak razlika između pogreške učenja i ispitne pogreške prevelika, to će rezultirati prenaučenošću modela (engl. *overfitting*). U tom će slučaju model biti previše prilagođen podacima za učenje, zbog čega će izvrsno funkcionirati na njima, ali neće moći generalizirati na ispitni skup.

Nadziranim strojnim učenjem rješavaju se dvije već spomenute skupine problema: problemi regresije i problemi klasifikacije. Regresija je metoda čije su izlazne varijable realni brojevi u nekom zadanom rangu. Recimo da želimo predvidjeti cijenu kuće u rangu od 50000 € do 200000 € iz njezine lokacije, kvadrature i godine izgradnje. Uzet ćemo podatke o nekim kućama za koje uz lokaciju, kvadraturu i starost znamo i cijenu, te ćemo dio tih podataka dati algoritmu kao skup za učenje. Algoritam će prema određenom modelu učenja odrediti funkciju kojom se za zadanu lokaciju, kvadraturu i starost kući pridružuje cijena. Taj ćemo algoritam zatim pokrenuti na ostatku podataka koji predstavlja ispitni skup te po potrebi, ako se dobiveni rezultati previše razlikuju od očekivanih, prilagoditi model. Neki od modela koji se najčešće koriste za rješavanje problema regresije su linearna regresija, metoda potpornih vektora (engl. *support vector machines*) i metoda slučajnih šuma (engl. *random*

forest regression) [38].

Klasifikacijom se predviđa oznaka klase iz skupa dviju ili više zadanih oznaka. Ako se podatak pridjeljuje jednoj od dvije moguće klase govorimo o binarnoj klasifikaciji, kao kod određivanja pripada li neka e-mail poruka u neželjenu poštu ili ne. S druge strane, kod višeklasne klasifikacije postoji više od dvije moguće klase, kao u prepoznavanju rukom pisanih brojeva. Moguće je kao rezultat klasifikacije dobiti kontinuiranu varijablu ako nam ona označava vjerojatnost da podatak pripada određenoj klasi. U toj se situaciji podatku pridjeljuje klasa koja ima najveću vjerojatnost. Na primjer, ako kod klasifikacije rukom napisanih brojeva određenom broju bude pridjeljena vjerojatnost od 0.8 da se radi o broju 1 i vjerojatnost od 0.2 da se radi o broju 7, taj će se broj klasificirati kao 1. Za klasifikaciju se najčešće koriste naivan Bayesov model (engl. *Naive Bayes Classification*) metoda potpornih vektora te modeli stabla odlučivanja (engl. *Decision Trees*) i slučajnih šuma [38].

3.3 Linearni modeli: linearna regresija i L_2 regularizacija

U ovom radu kao metoda rješavanja problema regresije korištena je linearna regresija u kojoj se, kako joj i ime kaže, izlazne varijable predviđaju kao linearne funkcije ulaznih. Neka predviđena vrijednost izlazne varijable y bude $\hat{y}$. Tada $\hat{y}$ ima oblik [45]

$$\hat{y} = w_0 + w_1\, x_1 + w_2\, x_2 + \ldots + w_n\, x_n, \tag{3.1}$$

gdje x_1, x_2, ..., x_n označavaju ulazne varijable jednog podatka, a w_0, w_1, w_2, ..., w_n su parametri koje želimo da model nauči. Kada bi podatak imao samo jednu ulaznu varijablu x, predviđena vrijednost izlazne varijable bila bi $\hat{y} = w_0 + w\, x$, što možemo prepoznati kao jednadžbu pravca s nagibom w i odsječkom na y osi w_0. U situacijama u kojima ovisnost izlaznih varijabli o ulaznima nije linearna, izraz 3.1 može se prilagoditi na način da se predviđanje izlazne varijabla izrazi kao

$$\hat{y} = w_0 + w_1\, \varphi_1(x) + w_2\, \varphi_2(x) + \ldots + w_n\, \varphi_n(x), \tag{3.2}$$

gdje su $\varphi_i(x)$ proizvoljne funkcije ulaznih varijabli. Funkcije $\varphi_i(x)$ zovu se bazne funkcije, a ovako dobiveni modeli i dalje se smatraju linearnim modelima. Ako su izlazne varijable polinomijalne funkcije ulaznih, odnosno vrijedi $\hat{y} = w_0 + w_1\, x_1 + w_2\, x_2^2 + \ldots + w_n\, x_n^n$, tada će bazna funkcija biti $\varphi_i(x) = x^i$. Takvu regresiju nazivamo polinomijalnom regresijom.

Parametri w_0, w_1, w_2, ..., w_n mogu se zapisati u obliku vektora $\boldsymbol{w} = (w_0, w_1, w_2, ..., w_n)$, a ako se ulazne varijable prikažu kao vektor $\boldsymbol{x} = (1, x_1, x_2, ..., x_n)$, $\hat{y}$ se može izraziti kao [46]

$$\hat{y} = \sum_{i=0}^{n} w_i x_i = \boldsymbol{w} \cdot \boldsymbol{x} = \boldsymbol{w}^\top \boldsymbol{x}. \tag{3.3}$$

Vektor $\boldsymbol{w}$ možemo opisati kao skup težina ulaznih varijabli [44]. Ako je težina w_i varijable x_i pozitivna vrijednost, to znači da se će predviđena vrijednost izlazne varijable $\hat{y}$ povećati s povećanjem x_i, odnosno smanjiti sa smanjenjem x_i. Ako pak varijabla x_i ima negativnu težinu w_i, onda će se povećanjem x_i smanjiti $\hat{y}$ i obratno. Što je težina neke varijable veća, to veći utjecaj ona ima na predviđenu vrijednost izlazne varijable.

Izrazi 3.1, 3.2 i 3.3 odnose se na razmatranje jednog podatkovnog zapisa ili jednog retka u tablici s podacima, no realizacija modela nadziranog strojnog učenja zahtijeva da skup sadrži varijable mnogo podatkovnih zapisa jer će u protivnom gotovo sigurno doći do podučenosti modela. Neka $\boldsymbol{X}$ bude matrica čiji svaki redak predstavlja ulazne varijable jednog podatka [47]. Takvu matricu zovemo matricom dizajna (engl. *design matrix*). Za primjer predviđanja cijene kuće iz njezine lokacije, kvadrature i starosti, matrica dizajna za dvije kuće imala bi oblik $\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \end{pmatrix}$, pri čemu x_{11}, x_{12} i x_{13} označavaju lokaciju, kvadraturu i starost prve kuće, a x_{21}, x_{22} i x_{23} lokaciju, kvadraturu i starost druge kuće. Izlazne varijable svih podataka u skupu za učenje prikažimo kao vektor $\boldsymbol{y}$. U našem primjeru vektor izlaznih varijabli imao bi oblik $\boldsymbol{y} = (y_1, y_2)$ gdje y_1 simbolizira cijenu prve, a y_2 cijenu druge kuće. Na ovaj način predviđanje izlaznih varijabli za više podatkovnih zapisa možemo izraziti kao

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{w}. \tag{3.4}$$

Za regresiju se koriste razni linearni modeli [39], koji se razlikuju po načinu na koji se kontrolira složenost modela i načinu na koji se dolazi do težina $\boldsymbol{w}$. Jedan od najpopularnijih je metoda najmanjih kvadrata (engl. *ordinary least squares*, OLS), u kojoj se nastoje pronaći parametri $\boldsymbol{w}$ takvi da srednja kvadratna pogreška (engl. *mean squared error*, MSE) između y i $\hat{y}$ bude minimalna. Srednju kvadratnu pogrešku definiramo kao omjer zbroja svih kvadriranih razlika između predviđenih i stvarnih izlaznih varijabli i broja podatkovnih zapisa. Za m

podatkovnih zapisa vrijedi [44]:

$$MSE = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2 = \frac{1}{m} \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2. \tag{3.5}$$

U izraz 3.5 uvrstimo izraz 3.4 i sredimo [48], [49]:

$$\begin{aligned} MSE &= \frac{1}{m} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|^2 \\ MSE &= \frac{1}{m} (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^\top (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) \\ MSE &= \frac{1}{m} (\boldsymbol{y}^\top - \boldsymbol{w}^\top \boldsymbol{X}^\top)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) \\ MSE &= \frac{1}{m} (\boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{w} - \boldsymbol{w}^\top \boldsymbol{X}^\top \boldsymbol{y} + \boldsymbol{w}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{w}) \end{aligned} \tag{3.6}$$

Srednja kvadratna pogreška je minimalna kad je njena derivacija s obzirom na vektor težine jednaka nuli, odnosno kad vrijedi $\nabla_{\boldsymbol{w}} MSE = 0$:

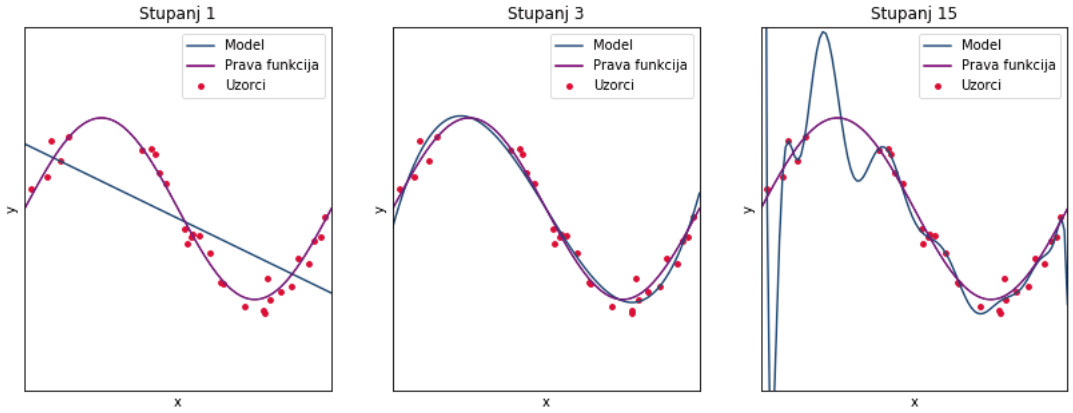
$$\begin{aligned} -\boldsymbol{y}^\top \boldsymbol{X} - \boldsymbol{y}^\top \boldsymbol{X} + \boldsymbol{w}^\top (\boldsymbol{X}^\top \boldsymbol{X} + \boldsymbol{X}^\top \boldsymbol{X}) &= 0 \\ -2\boldsymbol{y}^\top \boldsymbol{X} + 2\boldsymbol{w}^\top \boldsymbol{X}^\top \boldsymbol{X} &= 0 \\ -\boldsymbol{X}^\top \boldsymbol{y} + \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{w} &= 0 \end{aligned} \tag{3.7}$$

Jednadžbe 3.7 zovemo normalnim jednadžbama [44]. Njihovo rješenje je izraz za vektor težina uz koji će srednja kvadratna pogreška biti minimalna:

$$\boldsymbol{w} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}. \tag{3.8}$$

Uspješnost metode najmanjih kvadrata ovisi o broju ulaznih varijabli koje su uključene u model. Vratimo se opet na primjer predviđanja cijene kuće. Zamislimo da smo cijenu predviđali samo na temelju kvadature kuće. Tada bi se u skupu za učenje nalazili podaci iz kojih bi bilo vrlo teško zaključiti nešto o ovisnosti cijene o kvadraturi, primjerice da kuća veličine 70 m^2 i kuća veličine 150 m^2 imaju jednaku cijenu. Model koji je učio na tom skupu će donositi pogrešne pretpostavke o cijenama kuća jer je podnaučen, no ako bismo kvadraturi dodali lokaciju kuće, točnost modela bi se povećala. Kad bismo tome dodali i godinu izgradnje i recimo, stanje kuće, model bi bio sposobniji donositi još bolje pretpostavke o cijeni kuće. Ovakvim razmatranjem mogli bismo doći do zaključka da će model biti točniji što više ulaznih varijabli imamo, no to možemo tvrditi samo do određene razine. Što se broj ulaznih varijabli više približava broju podatkovnih uzoraka u skupu za učenje,

to je veća mogućnost da će doći do prenaučenosti modela. Ovo je vidljivo na slici 3.2, koja prikazuje polinomijalnu regresiju prvog, trećeg i petnaestog stupnja na skupu točaka koje opisuju sinusnu funkciju. Funkcija predviđena modelom regresije trećeg stupnja poprilično



Slika 3.2: Usporedba polinomijalnih regresija prvog (lijevo), trećeg (sredina) i petnaestog stupnja (desno) na skupu točaka koje opisuju sinusnu funkciju.

točno opisuje i zadane točke i zadanu funkciju. S druge strane, regresija prvog stupnja loše opisuje i skup zadanih točaka i zadanu funkciju jer je model podnaučen, dok regresija petnaestog stupnja relativno dobro opisuje skup zadanih točaka, ali ne i zadanu funkciju jer je model prenaučen. Do prenaučenosti modela ne dolazi samo zbog prevelikog broja ulaznih varijabli, nego i zbog prevelikih težina pojedinih varijabli. Kako bi se to izbjeglo koristi se metoda regularizacije (engl. *regularization*) [39] koja nastoji regulirati rast vrijednosti težina varijabli. To se postiže "kažnjavanjem" visokih težina tako što se u izraz za srednju kvadratnu pogrešku dodaje još jedan član [38]. U ovom se radu koristi L_2 regularizacija (engl. *Ridge regularization*) u kojoj je taj član $\alpha\|\boldsymbol{w}\|^2$. Izraz za tako dobivenu srednju kvadratnu pogrešku MSE' je

$$MSE' = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}\|^2 + \alpha\|\boldsymbol{w}\|^2 \tag{3.9}$$

Normalne jednadžbe i izraz za vektor težina u tom slučaju glase:

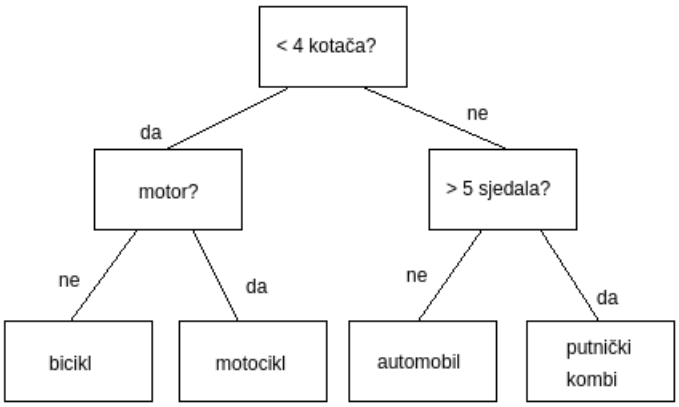
$$\begin{aligned} -\boldsymbol{y}^\top\boldsymbol{X} - \boldsymbol{y}^\top\boldsymbol{X} + \boldsymbol{w}^\top(\boldsymbol{X}^\top\boldsymbol{X} + \boldsymbol{X}^\top\boldsymbol{X}) + 2\alpha\boldsymbol{w}^\top &= 0 \\ -\boldsymbol{y}^\top\boldsymbol{X} + \boldsymbol{w}^\top(\boldsymbol{X}^\top\boldsymbol{X} + \alpha\boldsymbol{I}) &= 0 \\ \boldsymbol{w} = (\boldsymbol{X}^\top\boldsymbol{X} + \alpha\boldsymbol{I})^{-1}\boldsymbol{X}^\top\boldsymbol{y} & \end{aligned} \tag{3.10}$$

Na ovaj način može se postići uspješnost modela čak i u situaciji u kojoj radimo s velikim brojem ulaznih varijabli. No kada je broj ulaznih varijabli malen, često se linearnim mode-

lima ne mogu postići zadovoljavajući rezultati.

3.4 Stabla odlučivanja

Stabla odlučivanja su modeli kojima se mogu rješavati i problemi regresije i problemi klasifikacije [39], a u ovom su radu korištena za klasifikaciju. Realizaciju stabla možemo usporediti s igrom ”Pogodi tko”, u kojoj se ispituje niz da/ne pitanja s ciljem pogađanja osobe. Uzmimo primjer klasifikacije prijevoznog sredstva na bicikle, motocikle, automobile i putničke kombije. Prvo bismo mogli pitati ima li vozilo manje od 4 kotača. Ako je odgovor potvrdan, možemo pitati ima li vozilo motor kako bismo odredili radi li se o biciklu ili motociklu. Ako vozilo nema manje od 4 kotača, možemo raspoznati je li to automobil ili putnički kombi pitajući ima li vozilo više od 5 sjedala. Pripadno stablo odlučivanja prikazano je na slici 3.3. Korijen stabla i njegovi unutarnji čvorovi predstavljaju ulazne varijable, dok grane predstavljaju moguće vrijednosti tih varijabli. Krajnji čvorovi, odnosno listovi stabla označavaju konačne odluke. Dubina stabla jednaka je najvećem broju grana od korijena stabla do lista. Dakle, dubina stabla na slici 3.3 je 2.



Slika 3.3: Primjer stabla odlučivanja za klasifikaciju četiri vrste vozila.

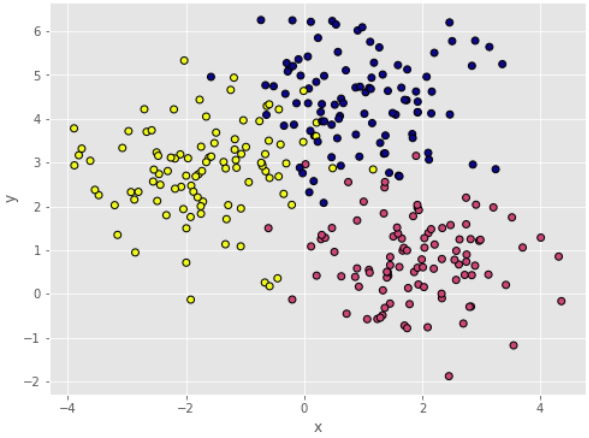
Pri korištenju stabla odlučivanja za klasifikaciju, model iz podataka u skupu za učenje uči koje će uvjete koristi za grananje stabla. Varijable obično nisu diskretne kao one u primjeru klasifikacije vozila, gdje broj kotača i broj sjedala moraju biti prirodni brojevi, a vozilo ili ima ili nema motor, već su kontinuirane. Na te se varijable uvjet postavlja kao if/else pitanje u obliku usporedbe varijable i i vrijednosti a [39]. Ta se pitanja u kontekstu strojnog učenja nazivaju testovima. No, kako algoritam zna koje pitanje postaviti? Odgovor se nalazi u

smanjenju entropije sa svakom razinom stabla [51]. Entropija je mjera neuređenosti sustava. Usporedimo li korijen stabla i njegove listove jasno je da u korijenu još nemamo nikakve informacije o podacima (sustav je potpuno neuređen), dok u listu znamo točno kojoj kategoriji pripada koji podatak (sustav je potpuno uređen). Dakle, iz korijena prema listovima smanjuje se entropija sustava. To se radi na način da ukupno smanjenje entropije iz jedne razine stabla u drugu bude maksimalno. Izraz za entropiju H je [52]

$$H = -\sum p(x)\, log_2 p(x), \tag{3.11}$$

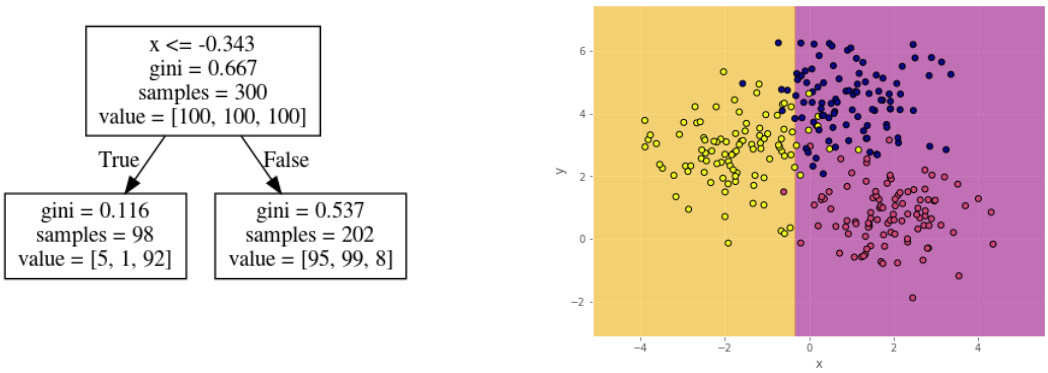
pri čemu $p(x)$ predstavlja vjerojatnost događaja x, a sumira se po svim mogućim događajima.

Pretpostavimo da se u skupu za učenje nalaze točke prikazane na slici 3.4, koje su podijeljene u tri klase na slici označene žutom, plavom i ružičastom bojom. Ulazne varijable točaka su njihove x i y koordinate, a izlazna varijabla svake točke je oznaka klase kojoj ta točka pripada. Kad na ovom skupu točaka pokrenemo algoritam stabla odlučivanja, kao rezultat ćemo dobiti stablo odlučivanja kojim se može za koordinate neke nove točke odrediti kojoj bi klasi pripadala ta točka. Na slici se nalazi ukupno 300 točaka, koje su raspoređene tako da svakoj



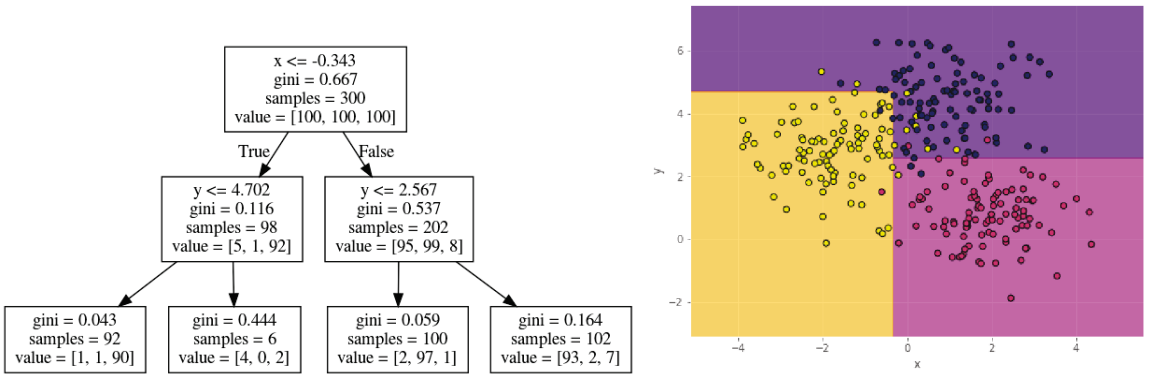
Slika 3.4: Primjer skupa za učenje algoritma stabla odlučivanja.

klasi pripada 100 točaka, što znači da za nasumično odabranu točku vjerojatnost da ona pripada jednoj od klasa iznosi 1/3. Izraz za entropiju ovog sustava je $H_0 = -3\ (\frac{1}{3}\ \log_2(\frac{1}{3})) = -\log_2(\frac{1}{3}) = 1,58$. Prvi test biti će napravljen tako da se ovaj broj smanji što je više moguće. Stablo dubine 1, odnosno stablo nakon prvog testa, prikazano je na slici 3.5. Podaci su podijeljeni u dvije skupine: one točke čija je varijabla x manja ili jednaka -0.343 (lijevi list), i one točke čija je varijabla x veća (desni list). Za svaku podjelu naveden je i njen gini koeficijent, koji predstavlja vjerojatnost da ćemo pogrešno klasificirati nasumično odabranu točku



Slika 3.5: Stablo odlučivanja dubine 1 (lijevo) i klasifikacija točaka na temelju tog stabla (desno).

kojoj smo nasumično dodijelili jednu od klasa zastupljenih u toj podjeli, a izražen je kao $G = \sum p(x)(1-p(x))$, pri čemu se sumira po svim klasama iz podjele [50]. Lijevom listu pripada 98 točaka, od kojih je njih pet iz "plave" klase, jedna iz "ružičaste", a 98 iz "žute". I bez računanja može se zaključiti da je ovo puno uređeniji sustav jer će nasumično odabrana točka iz tog lista imati vjerojatnost 92/98 da pripada "žutoj" klasi. Entropija za lijevi list iznosi $H_l = -\frac{5}{98}\log_2(\frac{5}{98}) - \frac{1}{98}\log_2(\frac{1}{98}) - \frac{92}{98}\log_2(\frac{2}{98}) = 0,37$. Desnom listu pripadaju 202 točke, od kojih 95 pripada "plavoj", 99 "ružičastoj", a osam "žutoj" klasi. Entropija desnog lista iznosi $H_d = -\frac{95}{202}\log_2(\frac{95}{202}) - \frac{99}{202}\log_2(\frac{99}{202}) - \frac{8}{202}\log_2(\frac{8}{202}) = 1,2$. Ukupna entropija stabla dubine 1 računa se kao suma umnoška entropije svakog lista i vjerojatnošću da će se nasumično odabrana točka nalaziti u tom listu [51]: $H_1 = \frac{98}{300}H_l + \frac{202}{300}H_d = 0,93$. Pad entropije jednak je dobitku informacije između početnog stanja sustava i stanja nakon prvog testa i iznosi $\Delta H_{21} = H_1 - H_2 = 0,65$. Uvjet prvog testa, $x \leq -0.343$ odabran je zato što je za taj uvjet dobitak informacije bio najveći. Za ovakvo stablo klasificirat ćemo sve točke lijevog lista kao "žute", a sve točke desnog lista kao "ružičaste" jer su upravo te klase najzastupljenije u pojedinom listu. No, svaki list i dalje sadrži točke svih triju klasa. Kako bismo poboljšali taj rezultat, možemo svaki list promatrati kao zasebno podstablo na kojem ćemo ponoviti postupak traženja najboljeg testa te tako dobiti stablo dubine 2, prikazano na slici 3.6. Ovo stablo ima 4 lista i puno je manji broj točaka pogrešno klasificiran. Unatoč tome, svakom listu i dalje pripadaju točke više od jedne klase. List koji sadrži podatke samo jedne klase smatra se čistim listom (engl. *pure leaf*). Korištenjem metode stabla odlučivanja za klasifikaciju podataka, svaki će se list koji nije čist smatrati novim podstablom i na njemu će se opet raditi podjela. Na taj se način dolazi do rezultata klasifikacije: stabla čiji su svi listovi čisti. Dio stabla dobivenog kao rezultat klasifikacije točaka sa slike 3.4 nalazi se na

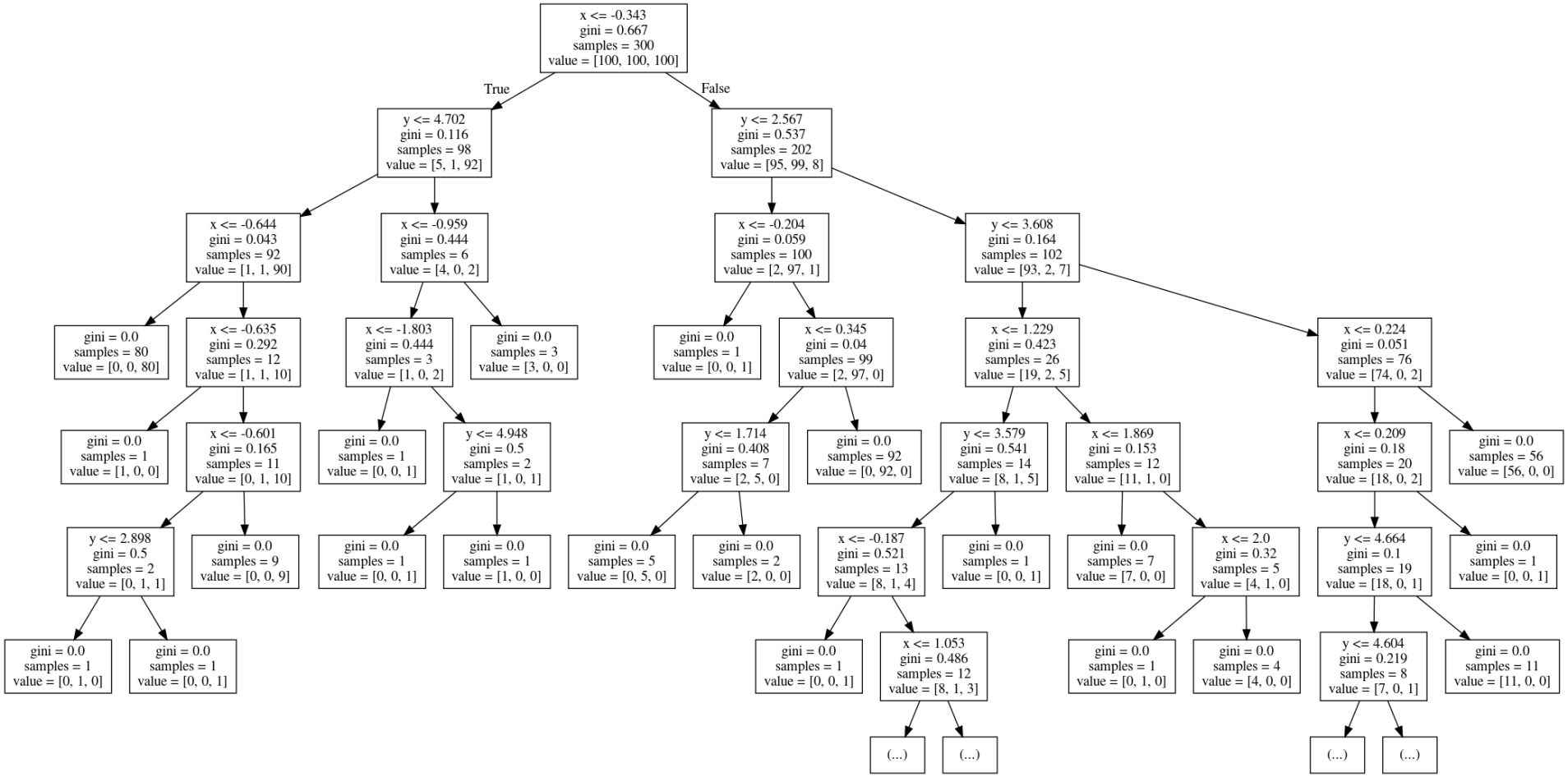


Slika 3.6: Stablo odlučivanja dubine 2 (lijevo) i klasifikacija točaka na temelju tog stabla (desno).

slici 3.7. Dubina čitavog stabla je 10, a prikazan je dio dubine 6.

Ovo je stablo vrlo složeno te nijedan list ne sadrži točke iz dviju različitih klasa, što znači da će ono točno predvidjeti klasu svake točke iz skupa za učenje. Na ispitnom skupu će uspješnost stabla biti puno manja jer je ono prenaučeno. Prenaučenost stabla odlučivanja može se spriječiti zaustavljanjem rasta stabla (engl. *prepruning*), ili rezanjem podstabla koja ne poboljšavaju točnost modela (engl. *postpruning*). Zaustavljanje rasta stabla je češće korištena metoda, a može se realizirati postavljanjem granice na maksimalnu dubinu stabla ili broj listova u stablu, kao i zahtijevanjem da se pojedino podstablo više ne dijeli ako se u njemu nalazi zadani minimalni broj podataka [39].

Stabla odlučivanja su jednostavna za interpretaciju i razumijevanje čak i za nekog tko se prvi put susreće s njima te za razliku od mnogih drugih modela, ne zahtijevaju neki posebni oblik obrade ulaznih varijabli prije korištenja modela. Također, ovi modeli mogu raditi s varijablama u različitim mjernim jedinicama te sa kombiniranim binarnim i kontinuiranim vrijednostima. S druge strane, i uz korištenje spomenutih metoda stabla odluke i dalje mogu biti presložena i sklona prenaučenosti, čega moramo biti svjesni pri njihovom korištenju.



Slika 3.7: Stablo dobiveno kao rezultat klasifikacije stablom odlučivanja.

4 Istraživanje perovskita primjenom strojnog učenja

Ovo istraživanje je rađeno po uzoru na projektni zadatak iz strojnog učenja CAMD ljetne škole za doktorske studente o teoriji elektronske strukture i dizajnu materijala, održane 2018. godine [53]. Cilj rada je pomoću nekoliko metoda strojnog učenja i baze podataka o organskim i anorganskim perovskitima predvidjeti određena svojstva ovih materijala te pronaći perovskitni spoj koji bi mogao biti povoljan za izradu solarnih ćelija.

4.1 Korišteni alati i metode

U ovom potpoglavlju dan je pregled svih programskih alata i metoda korištenih u radu: programski jezik Python, aplikacija Jupyter bilježnica te Python biblioteke NumPy, Matplotlib, ASE, Scikit-learn, Graphviz i Gpaw.

4.1.1 Python

Python [54] je objektno orijentirani programski jezik visoke razine. Može ga se besplatno preuzeti sa službene stranice, a moguće ga je pokrenuti na brojnim platformama kao što su Windows, Linux i Macintosh. Zbog čiste sintakse, jednostavnosti korištenja i lakoće razumijevanja vrlo je pogodan za početnike. Rad u Pythonu olakšava i činjenica da se radi o interpreterskom jeziku, što znači da se kod izvršava liniju po liniju, zbog čega je lakše zamijetiti i ispraviti pogreške kod programiranja. Uz bogatu standardnu biblioteku i veliki skup modula i funkcija koji dolaze s Pythonom, postoji još ogroman broj biblioteka koje korisnik može instalirati i koristiti u razne svrhe, što Python čini jednim od najpopularnijih programskih jezika, korištenim i od strane početnika i od strane naprednih programera. Verzija Pythona korištena u ovom radu je Python 3.7.

4.1.2 Jupyter bilježnica

Jupyter [55] je projekt namijenjen interaktivnom programiranju i razvoju slobodnog softvera koji je ime dobio po programskim jezicima Julia, Python i R, iako podržava i brojne druge jezike. Jedan od ”proizvoda” Jupyter projekta su i Jupyter bilježnice, web aplikacije koje korisnicima omogućuju pisanje koda, njegovo pokretanje, vizualizaciju podataka i prikaz rezultata u istom okruženju te olakšavaju dijeljenje kodova i dokumenata. Zbog ovog su se Jupyter bilježnice pokazale vrlo praktičnim za određene postupke korištene u znanosti o podacima, kao što su čišćenje podataka, analiza podataka, strojno učenje i mnogi drugi.

4.1.3 NumPy

NumPy [56] je Python biblioteka čije metode omogućuju korisnicima rad s objektima poput višedimenzionalnih polja, uključujući osnovne operacije linearne algebre, statistike i numeričke analize, zbog čega se NumPy i smatra temeljnom bibliotekom za znanstveno računanje u Pythonu.

4.1.4 Matplotlib

Kao alat za vizualizaciju podataka u Pythonu najčešće se koristi Matplotlib biblioteka [57], u kojoj se podaci mogu vizualizirati pomoću raspršenih, linijskih i stupčastih grafikona te histograma. Sve karakteristike grafikona mogu se oblikovati programskim naredbama, a grafikone je moguće spremiti u raznim formatima poput PNG, PDF i SVG.

4.1.5 ASE

ASE [58] je okolina za atomske simulacije (engl. *Atomic Simulation Environment*) čiji moduli omogućuju vizualizaciju i analizu atomskih simulacija te manipulaciju istim. ASE moduli mogu se koristiti u razne svrhe. Primjeri su kalkulatori za računanje različitih energija i sila u atomima i molekulama, moduli za kontrolu kretanja, ili pak moduli za vizualizaciju atoma i molekula.

4.1.6 Scikit-learn

Scikit-learn, ili kraće sklearn [59], je jedna od najpoznatijih biblioteka. Sadrži razne algoritme strojnog učenja. Karakterizira je jednolikost sintakse, odnosno činjenica da je za svaki model strojnog učenja sintaksa gotovo jednaka. Zbog toga je vrlo jednostavno raditi s raznim modelima čak i nakon što smo savladali samo jedan od njih.

4.1.7 Graphviz

Još jedan od alata za vizualizaciju podataka u Pythonu je Graphviz [60], čijim se modulima podaci mogu prikazati pomoću raznih dijagrama. Dijagrame je moguće spremiti kao slikovne podatke te u SVG ili PDF formatu, a kod pojedinog dijagrama mogu se mijenjati boje, stilovi linija, fontovi te željeni grafički oblici. Graphviz je u ovom radu korišten za vizualizaciju stabala odlučivanja.

4.1.8 GPAW

GPAW [61] je paket temeljen na teoriji funkcionala gustoće (engl. *density-functional theory*, DFT), PAW metodi (engl. *projector-augmented wave*) i ASE okolini. Paket je namijenjen raznim proračunima vezanim uz elektronsku strukturu.

4.2 Uvid u bazu podataka

Baza podataka o određenim perovskitima preuzeta je sa stranice CAMD ljetne škole [53]. Koristimo podatke za strukture tipa "*organometal*". Svaki red u bazi podataka predstavlja jedan perovskitni materijal. Ukupan broj perovskita u bazi je 140, a za svaki od njih navedene su razne značajke, neke od kojih su:

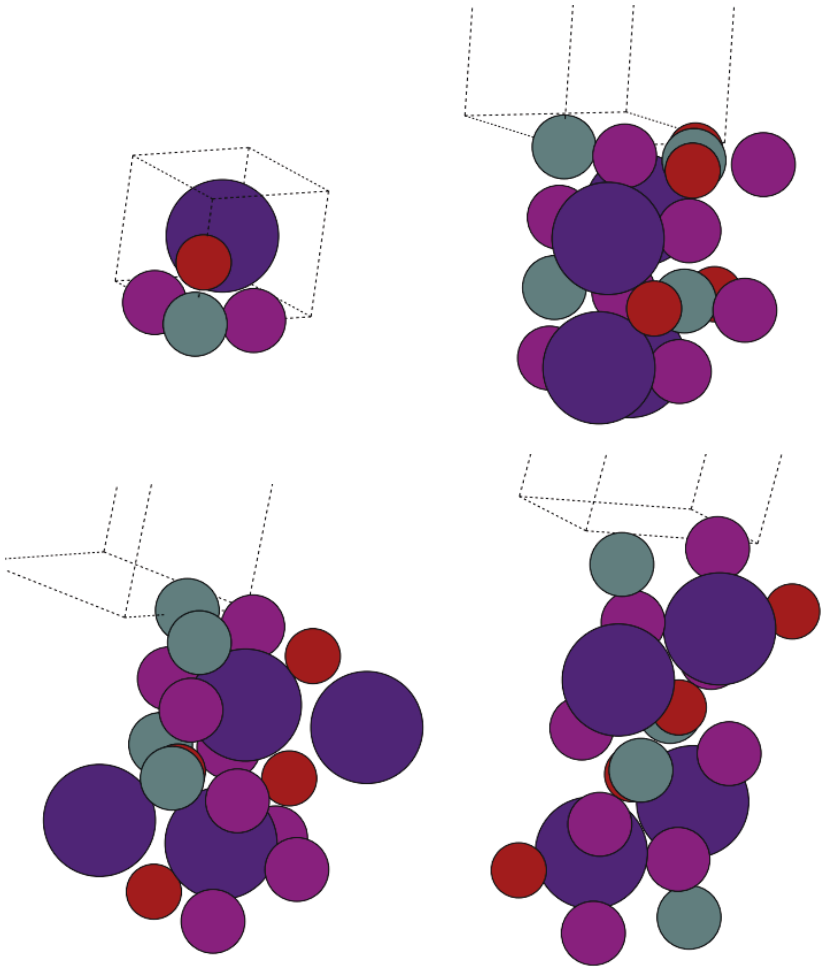
- *project*: ime projekta, u ovom slučaju je to "organometal"
- *name*: ime spoja u obliku kemijske formule
- *symmetry*: vrsta kristalne rešetke
- *space_group*: kristalografska prostorna grupa [***]
- *gllbsc_ind_gap*: indirektan energijski procjep u elektronvoltima [****]
- *gllbsc_dir_gap*: direktan energijski procjep u elektronvoltima
- *numbers*: lista atomskih brojeva elemenata u spoju
- *positions*: lista x,y i z koordinata svakog atoma iz *numbers*
- *cell*: jedinična ćelija kristala opisana listom 3 vektora od kojih svaki ima 3 komponente
- *energy*: potencijalna energija spoja

U bazi podatka postoji i potprojekt imena "*references*", u kojem se nalaze svi elementi korišteni za perovskite u projektu "*organometal*". To su: brom (Br), jod (I), klor (Cl), olovo (Pb), kositar (Sn), cezij (Cs) te organske molekule metilamonij (MA, formula CH_6N) i formamidinij (FA, formula CH_5N_2). Ispisom formula svih perovskita u bazi te njihovih kristalnih struktura i prostornih grupa može se uočiti da ovi spojevi imaju očekivanu općenitu formulu ABX_3, pri čemu se od elemenata iz potprojekta "*references*" na mjestu kationa A nalaze Cs, MA ili FA, na mjestu kationa B su Sn ili Pb, a na mjestu aniona X je neka od kombinacija 3 halogena atoma Cl, Br i I (recimo I_3, $BrCl_2$, IBrCl...). Također, javljaju se četiri vrste kristalne rešetke: kubična, tetragonska, te dvije ortorompske rešetke s različitim prostornim grupama (prva ortorompska i druga ortorompska). Neki od spojeva u bazi se nalaze

[***] Kristalografska prostorna grupa je skup svih simetrijskih operacija neke rešetke [62].

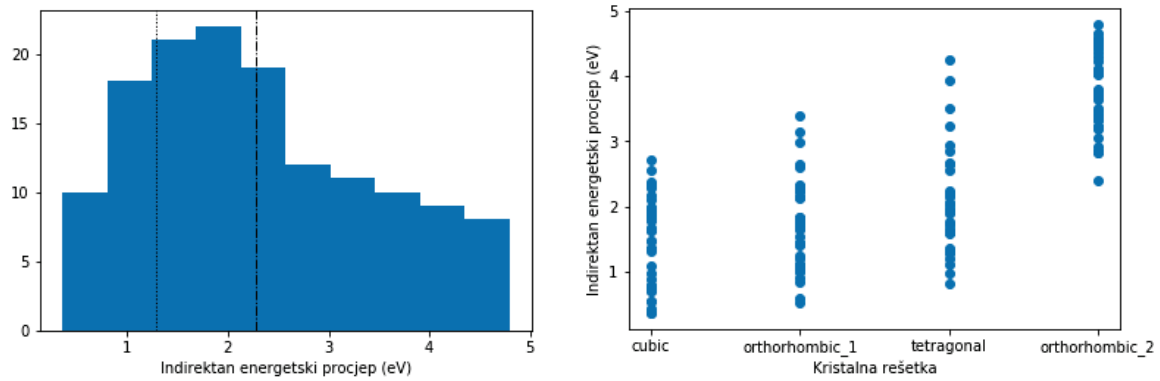
[****] GLLB-sc je jedan od funkcionala u teoriji funkcionala gustoće koji daje dobro slaganje s rezultatima mjerenja energijskog procjepa [63].

samo jednom, dok se neki u bazi nalaze dva, tri ili četiri puta, svaki put s drugačijom kristalnom rešetkom. Za spojeve koji se u bazi nalaze sa samo jednom fazom kristalne rešetke pretpostavlja se da im je to najstabilnija faza. Sve četiri faze kristalne rešetke perovskita $CsSnI_2Br$ vizualizirane u ASE okolini prikazane su na slici 4.1.

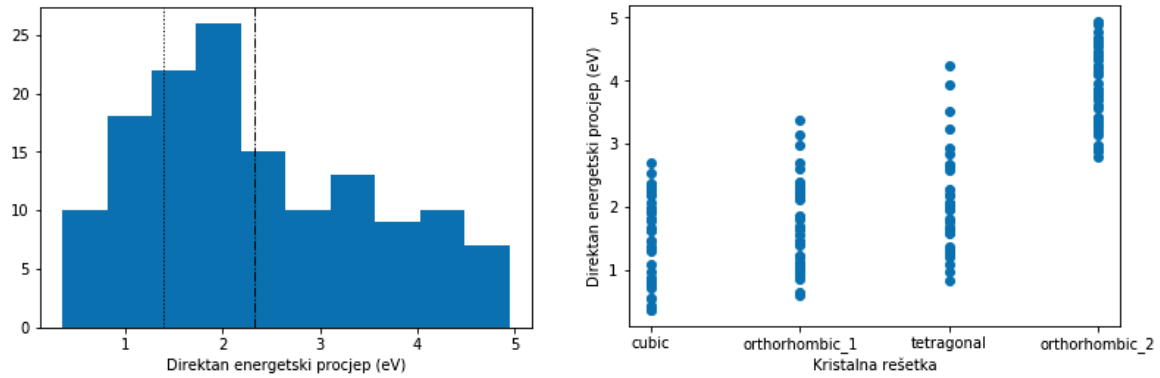


Slika 4.1: Četiri faze kristalne rešetke perovskita $CsSnI_2Br$: kubična (gore lijevo), tetragonska (gore desno), prva ortorompska (dolje lijevo) i druga ortorompska (dolje desno). Na slikama je Cs označen plavom, Sn sivo-zelenom, I crvenom, a Br ljubičastom bojom.

Nakon strukture i faza kristalne rešetke mogli bismo proučiti indirektne i direktne energijske procjepe perovskita u bazi podataka. Njihovi histogrami i grafovi ovisnosti o kristalnoj rešetci vidljivi su na slici 4.2 za indirektan i 4.3 za direktan energijski procjep. Na svakom je histogramu uključena i srednja vrijednost procjepa, prikazana isprekidanom linijom, te varijanca, prikazana točkastom linijom. Oba se procjepa za sve perovskite nalaze u intervalu od 0 do 5 eV te su najmanji za perovskite u kubičnoj fazi, a zatim rastu prema ortorompskoj fazi. To je očekivano jer se ipak u bazi nalazi mnogo perovskita s nekoliko faza kristalne rešetke, čiji energijski procjepi ovise o fazi rešetke na ovaj način.



Slika 4.2: Histogram indirektnog energijskog procjepa (lijevo) i njegova ovisnost o fazi kristalne rešetke (desno).

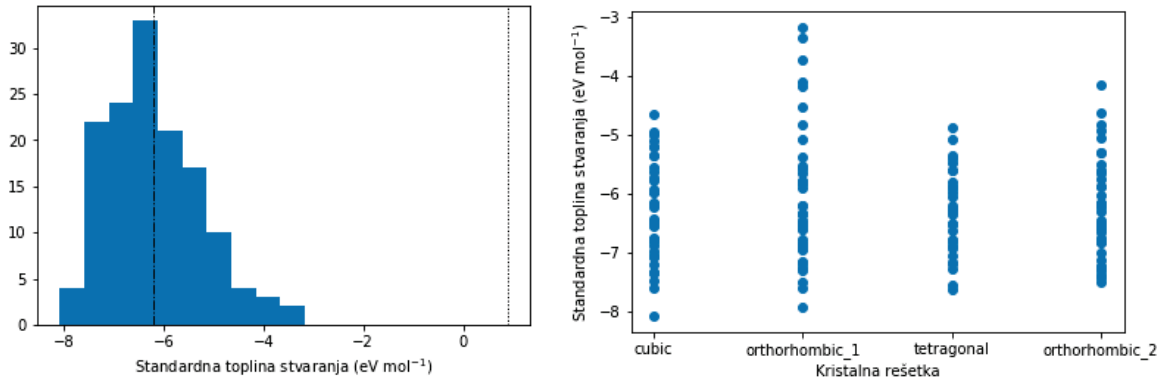


Slika 4.3: Histogram direktnog energijskog procjepa (lijevo) i njegova ovisnost o fazi kristalne rešetke (desno).

Potencijalna energija perovskita navedena u njegovim značajkama je energija u odnosu na neku proizvoljnu referentnu točku koja je postavljena u izvornom kodu ASE kalkulatora. Zgodnije je za svaki perovskit izračunati standardnu toplinu stvaranja kao razliku energije perovskita i sume energija pojedinih aniona i kationa od kojih je perovskit građen. Histogram standardne topline stvaranja s njenom srednjom vrijednosti i varijancom te ovisnost standardne topline stvaranja o kristalnoj rešetci perovskita prikazani su na slici 4.4. Sve standardne topline stvaranja su negativne, što znači da su svi materijali termodinamički stabilni.

4.3 Strojno učenje za perovskite

Jedan od ciljeva ovog rada je metodama nadziranog strojnog učenja predvidjeti energijski procjep perovskitnih materijala na temelju faze njihove kristalne rešetke te aniona i kationa od kojih je građen. Pošto se radi o nadziranom strojnom učenju, biti će nam potrebna matrica

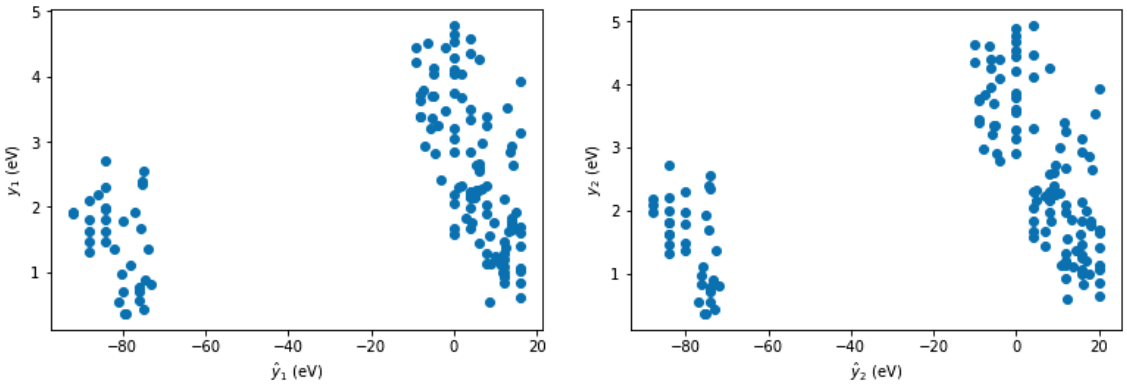


Slika 4.4: Histogram standardne topline stvaranja (lijevo) i njena ovisnost o fazi kristalne rešetke (desno).

dizajna (matrica ulaznih varijabli) $\boldsymbol{X}$ te vektor izlaznih varijabli $\boldsymbol{y}$. Matrica dizajna izrađena je kao matrica sa 140 redova i 13 stupaca, pri čemu svaki red matrice označava jedan perovskitni spoj. Prva četiri stupca matrice predstavljaju vrstu kristalne rešetke, redom kubičnu, tetragonsku, prvu ortorompsku i drugu ortorompsku. U svakom redu matrice će se nalaziti broj 1 u onom od prva četiri stupca koji predstavlja fazu rešetke tog perovskita, a na ostalim mjestima će biti 0. Recimo, ako je materijal u tetragonskoj fazi, prva četiri stupca njegovog reda će izgledati ovako: [0 1 0 0 ...]. Iduća tri stupca odnose se na kation A u perovskitu, pa označavaju redom Cs, FA i MA, po istom principu kao i za strukturu - na mjestu onog kationa koji se nalazi u perovskitu nalazi se broj 1, a na mjestu preostala dva kationa broj 0. Taj princip vrijedi i za sljedeća dva stupca koji se odnose na B kation u perovskitu, odnosno Pb i Sn. Nakon njih slijede tri stupca koja predstavljaju X_3 anion, odnosno jednu od kombinacija tri atoma Cl, I i Br, tim redoslijedom. U tim će se stupcima za svaki od atoma nalaziti njegova brojnost u spoju, recimo za perovskit građen od aniona Br_3 ta će tri stupca izgledati ovako: [... 0 0 3 ...], a za perovskit građen od aniona I_2Cl ovako: [... 1 2 0 ...]. Posljednji stupac u matrici jednak je za sve redove i iznosi 1 kako bi se predviđanje energijskog procjepa moglo izraziti pomoću jednadžbe 3.4. Vektor izlaznih varijabli definiran je kao vektor dimenzije 140, čija je svaka komponenta energijski procjep jednog perovskita iz baze podataka, pri čemu se *i*-ti red matrice dizajna i *i*-ta komponenta vektora izlaznih varijabli odnose na isti perovskit. Linearnom regresijom modelirani su i indirektan i direktan energijski procjep, za što je korištena ista matrica dizajna $\boldsymbol{X}$, dok su za vektor izlaznih varijabli $\boldsymbol{y_1}$ rabljeni direktni energijski procjepi perovskita, a za $\boldsymbol{y_2}$ indirektni. Primjerice, materijal $FAPbBrCl_2$ u tetragonskoj fazi je petnaesti perovskit u bazi podataka te ima indirektan energijski procjep od 3,238 eV, a direktan energijski procjep od 3,250 eV. To znači da će petnaesti red matrice

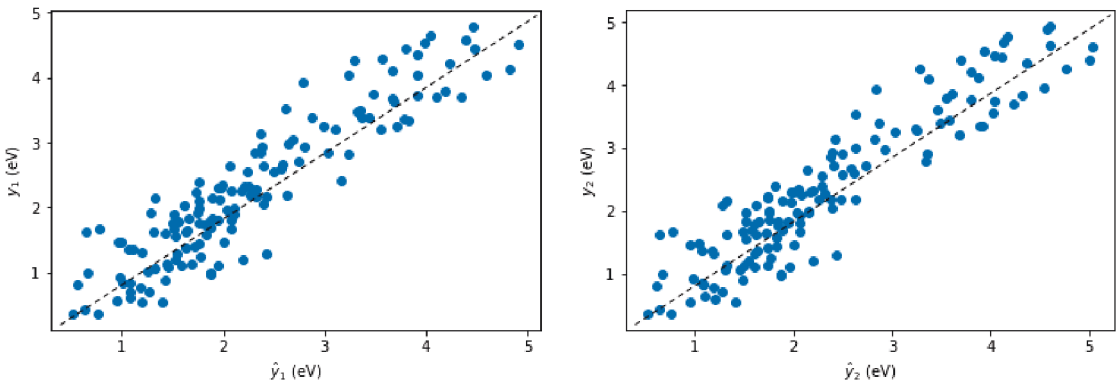
$\boldsymbol{X}$ izgledati ovako: [0, 1, 0, 0, 0, 1, 0, 1, 0, 2, 0, 1, 1], a petnaeste koordinate vektora $\boldsymbol{y_1}$ i $\boldsymbol{y_2}$ će iznositi redom 3,238 i 3,250.

Prvobitne predviđene vrijednost izlaznih varijabli $\hat{\boldsymbol{y_1}}$ i $\hat{\boldsymbol{y_2}}$ dobivene su pomoću izraza $\hat{\boldsymbol{y_i}} = \boldsymbol{X}\boldsymbol{w_i}$, pri čemu je vektor težina $\boldsymbol{w_i}$ izračunat kao $\boldsymbol{w_i} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y_i}$ za $i = 1,2$. Za obje predviđene vrijednosti energijskih procjepa izračunata je i srednja kvadratna pogreška, $MSE_i = \frac{1}{m}\,\|\boldsymbol{y_i} - \boldsymbol{X}\boldsymbol{w_i}\|^2$, pri čemu je m broj redaka matrice $\boldsymbol{X}$, što je u ovom slučaju 140. Dobivene srednje kvadratne pogreške iznose $MSE_1 = 1767,5286$ i $MSE_2 = 1709,9626$, što je jako loš rezultat pošto je cilj linearne regresije da dobivena srednja kvadratna pogreška bude što manja. Na slici 4.5 su prikazane očekivane i stvarne vrijednosti energijskih procjepa: $\boldsymbol{y_1} = f(\hat{\boldsymbol{y_1}})$ i $\boldsymbol{y_2} = f(\hat{\boldsymbol{y_2}})$. U slučaju potpuno točnih predviđanja energijskih procjepa



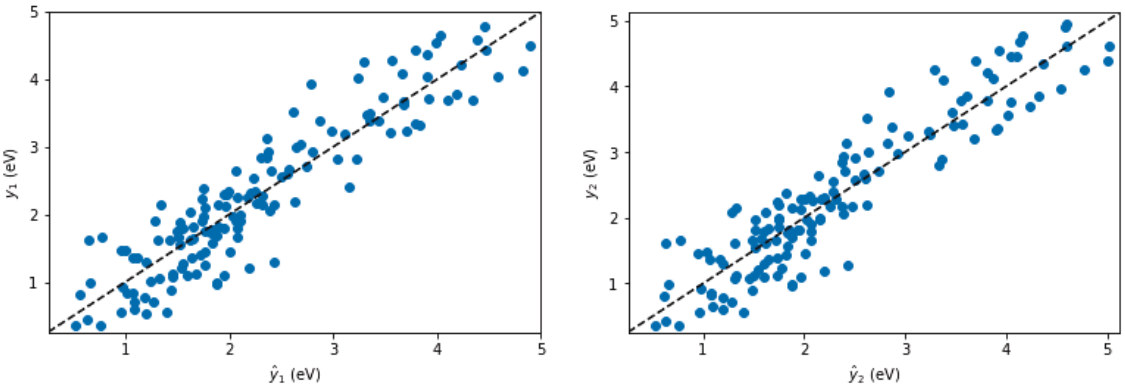
Slika 4.5: Usporedba stvarnih vrijednosti i vrijednosti predviđenih linearnom regresijom za indirektan (lijevo) i direktan (desno) energijski procjep.

koordinate vektora $\hat{\boldsymbol{y_i}}$ i $\boldsymbol{y_i}$ bile bi jednake te bi za grafove $\boldsymbol{y_i} = f(\hat{\boldsymbol{y_i}})$ funkcija $f(\hat{\boldsymbol{y_i}})$ iznosila $f(\hat{\boldsymbol{y_i}}) = \hat{\boldsymbol{y_i}}$, a grafovi bi tada prikazivali odnos $\boldsymbol{y_i} = \hat{\boldsymbol{y_i}}$. Ovdje se ne očekuju potpuno točna, već približno točna predviđanja, ali i u tom slučaju od grafova $\boldsymbol{y_i} = f(\hat{\boldsymbol{y_i}})$ očekujemo da barem približno prikazuju odnos $\boldsymbol{y_i} = \hat{\boldsymbol{y_i}}$. Dobiveni rezultat nije ni blizu očekivanog. Do ovog je došlo zato što se pri izvodu izraza 3.8 pretpostavlja da je matrica $(\boldsymbol{X}^\top \boldsymbol{X})$ invertibilna, a u ovom slučaju ona to nije. To se može riješiti L_2 regularizacijom gdje se vektor težina računa kao $\boldsymbol{w_i} = (\boldsymbol{X}^\top \boldsymbol{X} + \alpha \boldsymbol{I})^{-1}\boldsymbol{X}^\top \boldsymbol{y_i}$, pri čemu je matrica $(\boldsymbol{X}^\top \boldsymbol{X} + \alpha \boldsymbol{I})$ invertibilna. Srednja kvadratna pogreška u tom slučaju iznosi $MSE_i = \|\boldsymbol{y_i} - \boldsymbol{X}\boldsymbol{w_i}\|^2 + \alpha\|\boldsymbol{w_i}\|^2$. Grafovi stvarnih i očekivanih energijskih procjepa $\boldsymbol{y_1} = f(\hat{\boldsymbol{y_1}})$ i $\boldsymbol{y_2} = f(\hat{\boldsymbol{y_2}})$ za $\hat{\boldsymbol{y_1}}$ i $\hat{\boldsymbol{y_2}}$ dobivene L_2 regularizacijom prikazani su na slici 4.6. Ovi su rezultati već na prvi pogled puno bolji od prethodnih. Faktor α u oba slučaja iznosi $\alpha = 0,001$ jer su njime postignute najmanje srednje kvadratne pogreške, $MSE_1 = 0,1977$ i $MSE_2 = 0,1980$ koje dodatno potvrđuju zadovoljavajući rezultat modela L_2 regularizacije.



Slika 4.6: Usporedba stvarnih vrijednosti i vrijednosti predviđenih L_2 regularizacijom za indirektan (lijevo) i direktan (desno) energijski procjep.

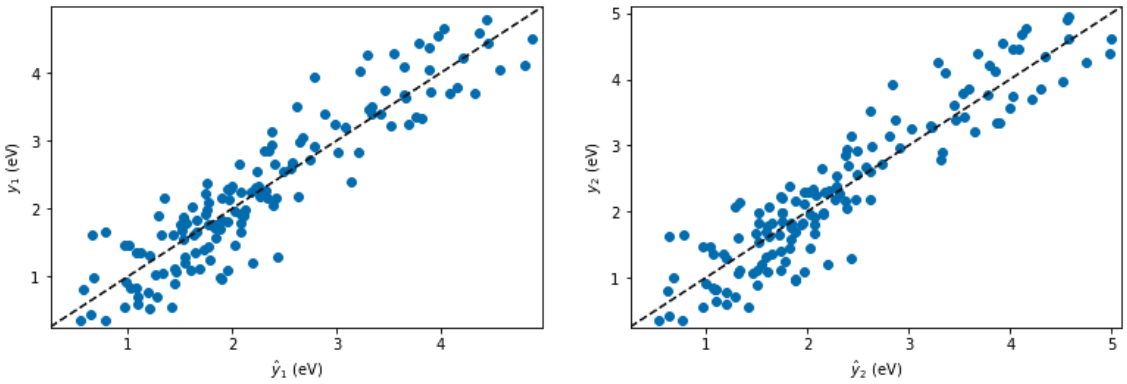
Regresiju je u Pythonu moguće raditi i alatima za strojno učenje iz scikit-learna. U sklearnu postoje razni modeli strojnog učenja, a svaki od njih ima metodu *fit* koja kao ulazne varijable prima matricu $\boldsymbol{X}$ i vektor $\boldsymbol{y}$ te pomoću njih obavlja učenje modela. Nakon što je model naučen može se pozvati metoda *predict*, također dostupna kod svih modela, koja kao ulaznu varijablu prima samo matricu $\boldsymbol{X}$ te pomoću naučenog modela kao izlaznu varijablu daje vektor predviđenih izlaznih vrijednosti $\hat{\boldsymbol{y}}$. Grafovi predviđenih i stvarnih vrijednosti energijskih procjepa dobiveni nakon pozivanja *fit* i *predict* metoda modela linearne regresije prikazani su na slici 4.7. Kako se rezultati ne bi svodili na našu vizualnu procjenu, modeli



Slika 4.7: Usporedba stvarnih vrijednosti i vrijednosti predviđenih sklearn modelom linearne regresije za indirektan (lijevo) i direktan (desno) energijski procjep.

sklearna imaju i metodu *score* koja nam daje omjer predviđenih i stvarnih izlaznih varijabli. Što je rezultat *score* metode bliže broju 1, to je predviđanje modela bliže stvarnim rezultatima. Rezultat score metode na sklearn modelu linearne regresije za indirektan energijski procjep iznosi 0,841, a za direktan energijski procjep 0,858. Za razliku od napisane pro-

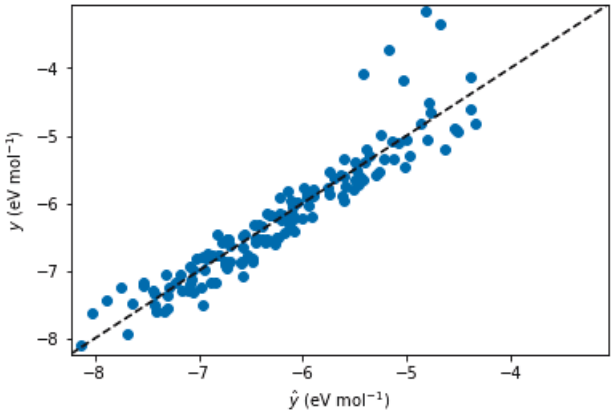
gramske funkcije za linearnu regresiju, metoda sklearna daje dobre pretpostavke izlaznih vrijednosti čak i kada je matrica ($\boldsymbol{X}^\top \boldsymbol{X}$) singularna. Unatoč tome, u ovakvim je situacijama ipak bolje za svaki slučaj koristiti regularizaciju. Grafovi predviđenih i stvarnih vrijednosti energijskih procjepa dobiveni nakon pozivanja *fit* i *predict* metode L_2 regularizacije s faktorom α od 0,5 prikazani su na slici 4.8. Pogledom na grafove vidljivo je da nema prevelike razlike u rezultatima ova dva tipa regresije, što je potvrđeno rezultatima *score* metode na sklearn modelu L_2 regularizacije, koji također iznose 0,841 za indirektan i 0,858 za direktan energijski procjep.



Slika 4.8: Usporedba stvarnih vrijednosti i vrijednosti predviđenih sklearn modelom L_2 regularizacije za indirektan (lijevo) i direktan (desno) energijski procjep.

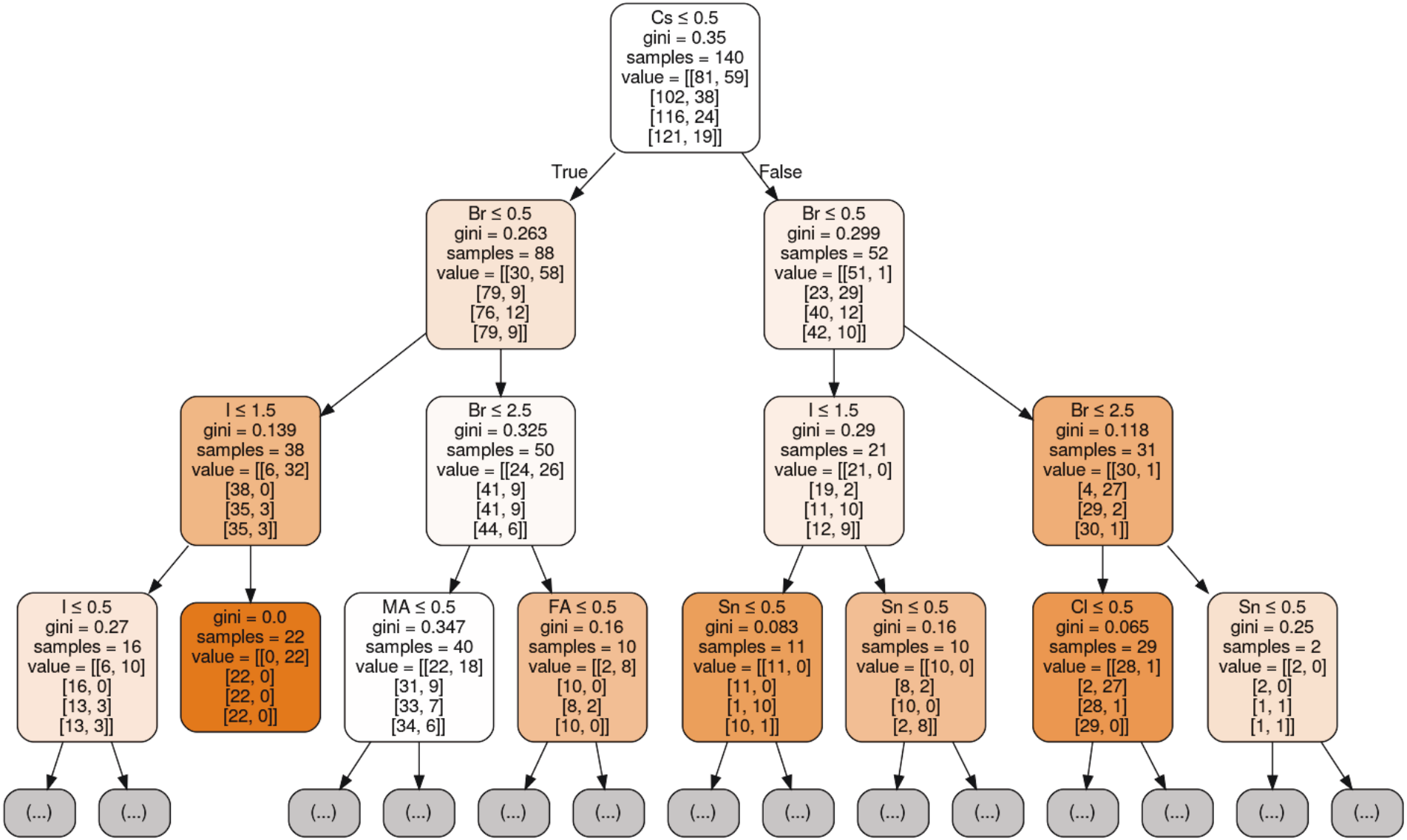
Rezultati dobiveni na ovaj način prikazuju točnost s kojom model predviđa izlazne varijable za isti skup na kojem je i učio, što nije realna situacija. Kako bismo provjerili s kolikom bi točnošću naš model predvidio izlazne varijable za neki skup nepoznatih ulaznih varijabli, korištena je metoda *k*-struke validacije s četiri particije. Rezultat validacije je broj u rangu 0-1, pri čemu je točnost modela to veća što je rezultat validacije bliže broju 1. Dobiveni rezultati *k*-struke validacije za sklearn model linearne regresije su 0,82 za indirektan energijski procjep i 0,84 za direktan. Za sklearn model L_2 regularizacije rezultati *k*-struke validacije su također 0,82 za indirektan i 0,84 za direktan energijski procjep.

Po uzoru na modele za predviđanje energijskih procjepa perovskita može se pokušati predvidjeti i standardna toplina stvaranja pojedinog perovskita na temelju kationa i aniona od kojih je građen te faze njegove kristalne rešetke. U tu je svrhu korištena ista matrica dizajna kao i u prethodnim modelima, dok je za izlazni vektor korišten vektor prethodno izračunatih standardnih toplina stvaranja za perovskite. Korišten je sklearn model L_2 regularizacije. Graf predviđenih i stvarnih standardnih toplina stvaranja prikazan je na slici 4.9. Rezultat *k*-struke validacije s četiri particije za L_2 regularizaciju standardne topline formacije iznosi 0,85.



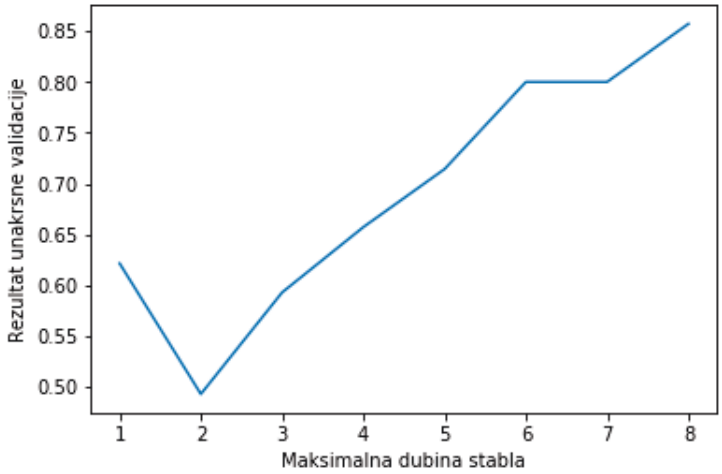
Slika 4.9: Usporedba stvarnih vrijednosti i vrijednosti predviđenih sklearn modelom L_2 regularizacije za standardnu toplinu stvaranja perovskita.

Perovskiti koji se u bazi podataka nalaze više puta s različitim oblicima kristalne rešetke također imaju i različite standardne topline stvaranja, koje određuju i njihovu stabilnost - manja toplina stvaranja označava veću termodinamičku stabilnost perovskita i obratno. Na temelju kationa i aniona od kojih je perovskit građen moguće je predvidjeti koja će od njegovih faza kristalne rešetke biti najstabilnija, odnosno imati najmanju standardnu toplinu stvaranja. Za to su iz prethodne matrice dizajna uklonjena prva četiri stupca koja su se odnosila na strukturu kristalne rešetke, tako da je matrica dizajna za ovaj model sastavljena od devet stupaca od kojih prvih osam predstavlja anione i katione od kojih je građen pojedini perovskit, a deveti stupac je u svim redovima konstanta, odnosno broj 1. Izlazne varijable dane su u obliku matrice sa 140 redova i 4 stupca u kojoj se u svakom redu nalazi kristalna rešetka za koju je perovskit na kojeg se taj red odnosi najstabilniji. Ta je matrica napisana na način da je za svaki perovskit iz baze podataka pronađeno s koliko se faza kristalne rešetke on nalazi u bazi. Ako se nalazi s jednom ta je faza upisana u izlaznu matricu, a ako se radi o više faza za svaku od njih je pronađena pripadajuća standardna toplina stvaranja te je faza s najmanjom toplinom stvaranja upisana u izlaznu matricu. Faza kristalne rešetke je u izlaznoj matrici prikazana kao i u prethodnim modelima, s četiri broja koja se odnose na kubičnu, tetragonsku, prvu ortorompsku i drugu ortorompsku rešetku te koji iznose 1 ako je najstabilnija ta struktura rešetke na koju se oni odnose, a 0 ako ta struktura nije najstabilnija. Pošto su faze kristalne rešetke diskretne varijable, predviđanje kristalne rešetke za koju je perovskit najstabilniji modelirano je stablom odlučivanja. Rezultantno stablo ima dubinu 8, a njegov dio dubine 3 vidljiv je na slici 4.10.



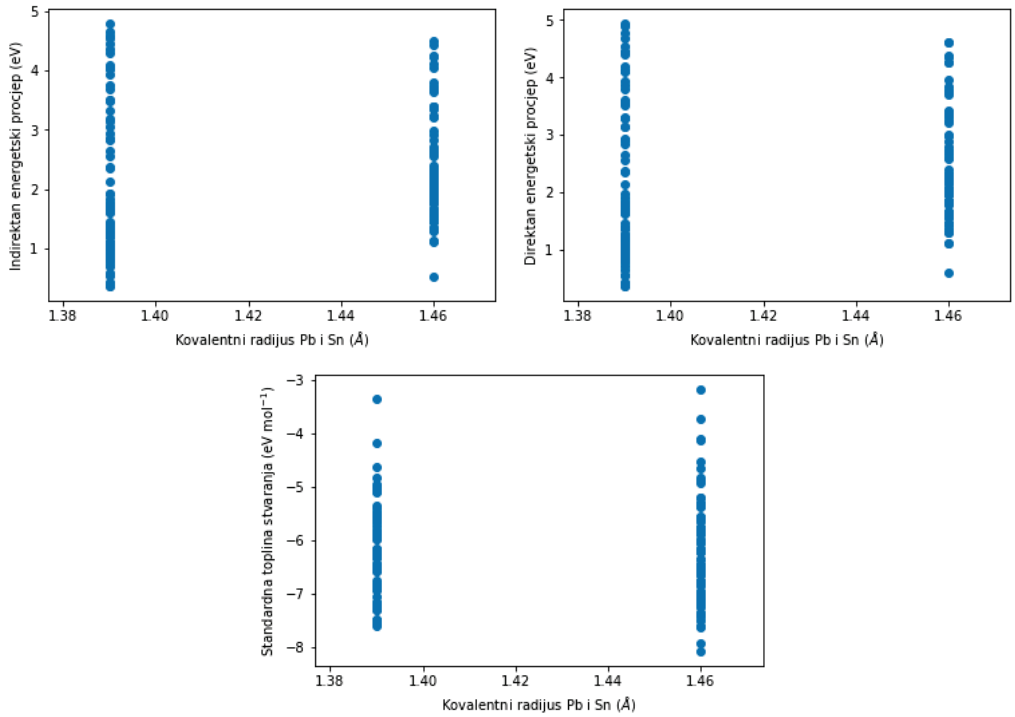
Slika 4.10: Stablo odlučivanja za klasifikaciju strukture kristalne rešetke perovskita na temelju aniona i kationa od kojih je građen.

Kako bismo provjerili da ne dolazi do prenaučenosti stabla odlučivanja, dobro je usporediti točnosti stabla odlučivanja različitih dubina. Graf rezultata k-struke validacije s četiri particije za dubine stabla od 1 do 8 prikazan je na slici 4.11. Da dolazi do prenaučenosti modela, točnost bi nakon određene vrijednosti prestala rasti (ili bi neznatno rasla), što ovdje nije slučaj. Najbolji postignuti rezultat je 0,85 i postignut je za stablo dubine 8. Također je vidljivo da stabla dubine 6 i 7 postižu rezultat od 0,8. To znači da bismo, ako želimo jednostavnije stablo i dovoljna nam je točnost od 0,8 mogli zaustaviti stablo kad dosegne dubinu 6.



Slika 4.11: Usporedba rezultata unakrsne validacije za dubine stabla od 1 do 8.

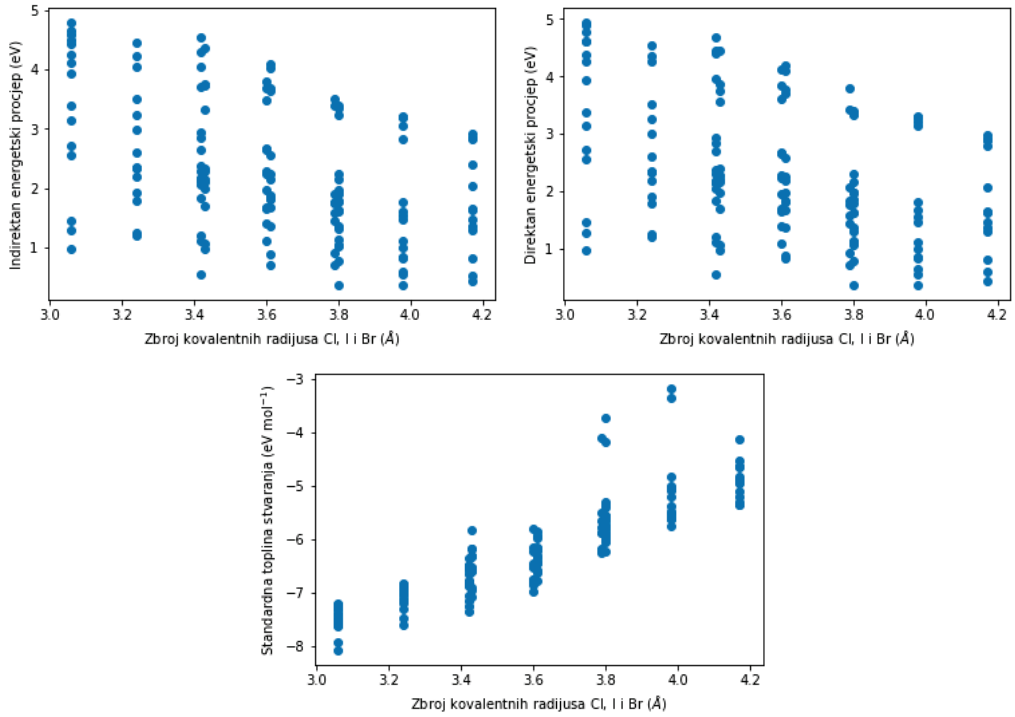
Točnost modela linearne regresije za previđanje energijskih procjepa i standardne topline stvaranja te modela stabla odlučivanja za predviđanje strukture perovskita je između 0,82 i 0,85 za sve modele. To je zadovoljavajući rezultat, no mogli bismo ga pokušati poboljšati tako da u matricu dizajna dodamo još neke vrijednosti koje bi mogle imati utjecaj na energijski procjep ili standardnu toplinu stvaranja, recimo kovalentni polumjer određenih elemenata u perovskitu. Kovalentni polumjer atoma jednak je polovini udaljenosti između jezgara dvaju jednakih atoma povezanih kovalentnom vezom te može utjecati na konstantu kristalne rešetke, a time i na energiju kristala. Prvo ćemo razmotriti kovalentne polumjere kationa olova i kositra u perovskitima. Grafovi ovisnosti energijskih procjepa i topline stvaranja svih perovskita u bazi o kovalentnom polumjeru olova ili kositra prikazani su na slici 4.12. Iz grafova je vidljivo da se i za kovalentni polumjer olova i za kovalentni polumjer kositra postižu energijski procjepi i standardne topline stvaranja u otprilike jednakom rangu, no to ne mora nužno značiti da kovalentni polumjeri olova i kositra nemaju utjecaj na energijske procjepe i na toplinu stvaranja pojedinog perovskita. Kad bismo promatrali energijske procjepe i to-



Slika 4.12: Grafovi ovisnosti energijskih procjepa (indirektni - prvi red lijevo, direktni - prvi red desno) i topline stvaranja (drugi red) o kovalentnom polumjeru olova ili kositra u perovskitu.

plinu stvaranja nekoliko parova perovskita koji se razlikuju samo po tome što jedan u sebi sadrži olovo, a drugi umjesto olova sadrži kositar, tada bismo mogli uočiti postoji li uistinu određena povezanost između kovalentnih polumjera olova i kositra te energijskih procjepa ili standardne topline stvaranja. U ovoj se bazi nalaze svi mogući spojevi perovskita, što znači da se za svaki perovksit koji sadrži olovo u bazi nalazi i njegov partner koji umjesto olova sadrži kositar, ali oni nemaju nužno jednake faze kristalne rešetke, a i moguće je da se jedan od njih u bazi nalazi više puta nego drugi. Zbog toga ćemo ipak dodati kovalentni polumjer olova ili kositra u matricu dizajna kako bismo vidjeli hoće li to imati ikakvog utjecaja na rezultate naših modela. Kada se u matricu dizajna doda još jedan stupac u kojem se nalazi kovalentni polumjer olova ili kositra, ovisno o tome koji se od ova dva elementa nalazi u spoju, rezultat k-struke validacije s četiri particije modela L_2 regularizacije je za indirektan energijski procjep 0,82, a za direktan 0,83, dok za standardnu toplinu stvaranja iznosi 0,86. Točnost klasifikacije kristalne rešetke perovskita stablom odlučivanja dubine 8 iznosi 0,85. Vidljivo je da je jedini utjecaj dodavanja kovalentnog polumjera olova i kositra u ulazne podatke modela smanjenje točnosti predviđanja direktnog energijskog procjepa za 0,01 i povećanje točnosti predviđanja standardne topline stvaranja za 0,01. Osim olova i ko-

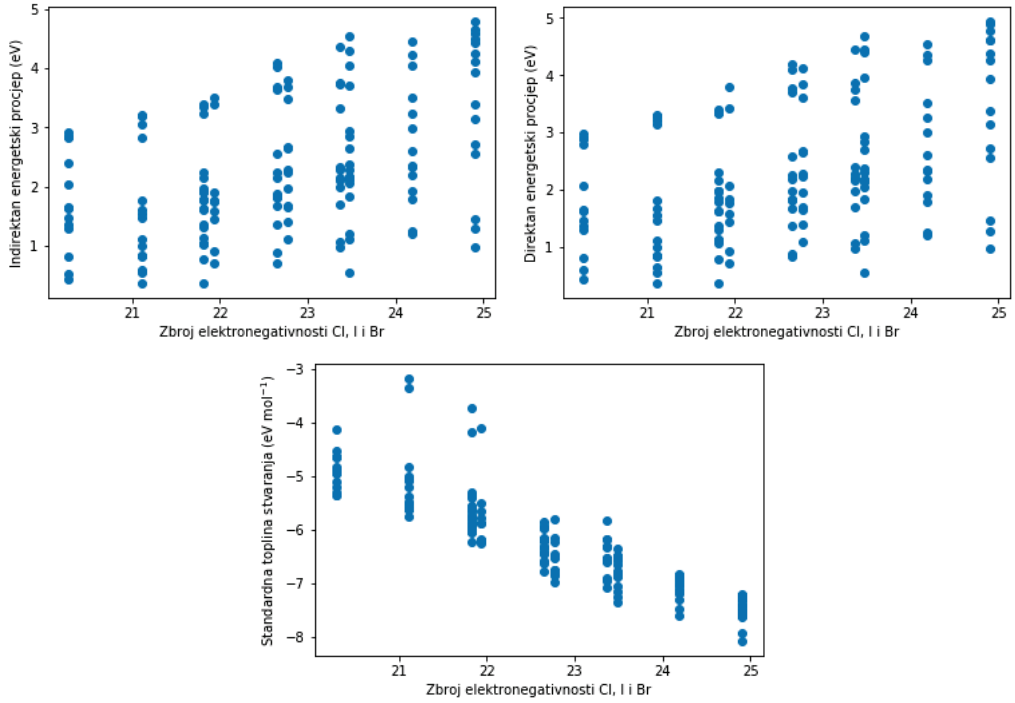
sitra, razmotrit ćemo i kovalentne polumjere klora, broma i joda u perovskitima. Pošto jedan perovskit ne mora sadržavati samo jedan od tih elemenata, kao što je bio slučaj sa olovom i kositrom, već se može raditi o kombinaciji elemenata klora, broma i joda, promatrat ćemo zbroj kovalentnih polumjera svih halogenih elemenata u spoju. Grafovi ovisnosti energijskih procjepa i topline stvaranja o zbrojevima kovalentnih polumjera halogenih elemenata u perovskitu prikazani su na slici 4.13. Vidljivo je da spojevi sa većim zbrojem kovalent-



Slika 4.13: Grafovi ovisnosti energijskih procjepa (indirektni - prvi red lijevo, direktni - prvi red desno) i topline stvaranja (drugi red) o zbrojevima kovalentnih polumjera halogenih elemenata u perovskitu.

nih polumjera halogenih atoma imaju nešto niže energijske procjepe od onih s manjim, kao i da standardna toplina stvaranja raste s porastom zbroja kovalentnih polumjera halogenih elemenata u perovskitu. Kada se u matricu dizajna dodaju još tri stupca u kojima se nalaze kovalentni polumjeri svakog od tri halogena elementa u spoju, rezultat *k*-struke validacije s četiri particije modela L_2 regularizacije je za indirektan energijski procjep 0,82, a za direktan 0,83, dok za standardnu toplinu stvaranja iznosi 0,86. Točnost klasifikacije kristalne rešetke perovskita stablom odlučivanja dubine 8 iznosi 0,86. Opet nema znatne promjene u rezultatu modela kod kojeg su kovalentni polumjeri halogenih elemenata uzeti u obzir pri razmatranju ulaznih podataka, iako polumjeri sami po sebi utječu na energijske procjepe i toplinu stvaranja. Razlog zbog kojeg nema promjene u rezultatu je taj da je i u prethodnoj matrici dizajna

bio uključen broj halogenih elemenata koji se nalaze u perovskitu, tako da je utjecaj njihovih kovalentnih polumjera na neki način već bio uračunat u procjenu modela. Nakon kovalentnih polumjera možemo razmotriti i elektronegativnost halogenih elemenata. Elektronegativnost je sposobnost nekog atoma da privuče elektronski par koji unutar molekule dijeli s drugim atomom u kovalentnoj vezi, zbog čega bi različite elektronegativnosti halogenih elemenata u spoju mogle imati utjecaj na karakteristike veza u kristalu. Grafovi ovisnosti energijskih procjepa i topline stvaranja o zbroju elektronegativnosti halogenih elemenata prikazani su na slici 4.14. Iz grafova se može vidjeti da su energijski procjepi perovskita nešto viši s



Slika 4.14: Grafovi ovisnosti energijskih procjepa (indirektni - prvi red lijevo, direktni - prvi red desno) i topline stvaranja (drugi red) o zbroju elektronegativnosti halogenih elemenata.

porastom zbroja elektronegativnosti halogenih elemenata, dok toplina stvaranja pada kako zbroj elektronegativnosti raste. Kada se u matricu dizajna dodaju još tri stupca u kojima se nalaze elektronegativnosti svakog od tri halogena elementa u spoju, rezultat k-struke validacije s četiri particije modela L_2 regularizacije je za indirektan energijski procjep 0,81, a za direktan 0,82, dok za standardnu toplinu stvaranja iznosi 0,86. Točnost klasifikacije kristalne rešetke perovskita stablom odlučivanja dubine 8 iznosi 0,85. Ni ovdje nema znatne promjene u odnosu na originalne podatke, također zbog toga što je broj pojedinih halogenih elemenata, pa tako i njihovih elektronegativnosti već bio uključen u ulazne podatke modela.

4.4 Testiranje i procjena modela

Kako bismo pronašli perovskite najpovoljnije za izradu solarnih ćelija prvo moramo definirati kakve perovskite tražimo. Želimo spoj koji ima negativnu standardnu toplinu stvaranja, odnosno koji je termodinamički stabilan. Također želimo da mu energijski procjep bude oko 1.5 eV i da mu kristalna rešetka bude u fazi u kojoj je spoj termodinamički najstabilniji, to jest ima najmanju standardnu toplinu stvaranja u usporedbi s ostalim fazama kristalne rešetke tog spoja. Standardne topline stvaranja svih perovskita u bazi su negativne. Što se tiče energijskog procjepa, pronađeni su perovskiti čija se oba energijska procjepa (indirektan i direktan) nalaze u rasponu od 1,3 do 1,6 eV pri čemu se pazilo i da faza njihove kristalne rešetke bude ona najstabilnija. To je učinjeno zahtjevom da faza rešetke perovskita bude ona koja je navedena u izlaznom vektoru za klasifikaciju najstabilnije faze rešetke pojedinog perovskita. Spojevi dobiveni na ovaj način su: $CsPbI_2Br$ u kubičnoj, MASnBr3 u prvoj ortorompskoj, $FASnI_2Cl$ u kubičnoj, CsSnIBrCl u tetragonskoj, $CsSnCl_3$ u prvoj ortorompskoj, $FAPbIBr_2$ prvoj ortorompskoj, $FAPbI_3$ u kubičnoj, $CsPbIBr_2$ u tetragonskoj i $MAPbI_3$ u kubičnoj fazi. Perovskiti koji se trenutno najviše koriste za izradu solarnih ćelija su $MAPbI_3$ i $FAPbI_3$ [64], što je u skladu s ovim rezultatima. Svi bi navedeni spojevi mogli biti povoljni za izradu solarnih ćelija, no možemo izdvojiti one koji ne sadrže olovo te ne predstavljaju opasnost za okoliš. Takvi perovskiti su: MASnBr3 u prvoj ortorompskoj, $FASnI_2Cl$ u kubičnoj, CsSnIBrCl u tetragonskoj i $CsSnCl_3$ u prvoj ortorompskoj fazi.

Provjeru točnosti modela bilo bi najpraktičnije raditi za neki perovskit koji se ne nalazi u bazi podataka, kako bi se vidjelo snalaženje modela s podacima koji mu još nisu poznati. U tu je svrhu korišten perovskit $CsSnBr_3$ koji se u bazi podataka nalazi u drugoj ortorompskoj fazi, a provjere su rađene za kubičnu fazu koje nema u bazi. Kako bismo za kubičnu fazu CsSnBr3 saznali širinu energijskog procjepa i standardnu toplinu stvaranja koje su nam potrebne za kasniju usporedbu s predviđenim vrijednostima modela, uzeta je struktura kubične faze CsPbI2Br te je pomoću nje formiran $CsSnBr_3$ zamjenom atoma olova s atomom kositra i dva atoma joda s dva atoma bora. Promjenom atoma u strukturi, promijenila se i sila na pojedine atome. Zbog toga se ovakav spoj ne nalazi u relaksiranom stanju, već ga treba relaksirati. To je urađeno pomoću GPAW kalkulatora. Nakon relaksacije izračunata je potencijalna energija i energijski procjep ovog perovskita. Izračunati direktan energijski procjep iznosi 1,123 eV. Potencijalna energija iznosi -16.913 eV, a pomoću nje može se izračunati standardna toplina stvaranja na isti način kao i za perovskite u bazi podataka. Dobivena standardna toplina stvaranja iznosi -6.319 eV mol^{-1}.

Za predviđanje vrijednosti energijskog procjepa i standardne topline stvaranja perovskita $CsSnBr_3$ u kubičnoj fazi korišteni su sklearn modeli L_2 regularizacije, dok je za predviđanje najstabilnije faze rešetke korišten model stabla odluke. Svi modeli su dobiveni učenjem na podacima o perovskitima iz baze podataka. Nakon učenja modela iskorištena je njihova *predict* metoda na ulaznom vektoru za $CsSnBr_3$, koji ima oblik [1,0,0,0,1,0,0,0,1,0,0,3,1] za modele predviđanja energijskog procjepa i standardne topline stvaranja te [1,0,0,0,1,0,0,3,1] za predviđanje najstabilnije faze kristalne rešetke. Rezultat *predict* metode za direktan energijski procjep je 1,186 eV, s pogreškom aproksimacije od 0,063 eV i relativnom pogreškom aproksimacije od 0,056. Rezultat *predict* metode za standardnu toplinu stvaranja je -6.469 eV mol^{-1}, s pogreškom aproksimacije od 0,15 eV mol^{-1} i relativnom pogreškom aproksimacije od 0,024. Faza kristalne rešetke dobivena kao rezultat metode *predict* za stablo odlučivanja je druga ortorompska, koja se i nalazi u bazi podataka kao najstabilnija. Važno je napomenuti da je cjelokupni proces za dobivanje vrijednosti energijskog procjepa i standardne topline stvaranja GPAW kalkulatorom trajao otprilike jedan sat, a da je odabrana faza kristalne rešetke bila složenija od kubične, trajao bi i nekoliko sati. S druge strane, rezultati modela strojnog učenja dobiveni su u svega nekoliko trenutaka.

5 Zaključak

Ovo je istraživanje provedeno s ciljem predviđanja energijskog procjepa, standardne topline stvaranja i najstabilnije faze perovskita te pronalaska perovskitnih spojeva koji bi mogli poslužiti kao pogodan materijal za solarne ćelije.

Za predviđanje energijskog procjepa i standardne topline stvaranja perovskita korištena je L_2 regresija, kojom smo uspjeli predvidjeti i energijski procjep i standardnu toplinu stvaranja na temelju kemijske formule i faze kristalne rešetke perovskita. Stablom odlučivanja na temelju kemijske formule perovskita predvidjeli smo koja će faza njegove kristalne rešetke biti najstabilnija. Može se zaključiti da je za predviđanje najstabilnije faze kristalne rešetke perovskita dovoljno znati njegovu kemijsku formulu, dok energijski procjep i standardna toplina stvaranja ovise i o kemijskoj formuli, i o fazi kristalne rešetke perovskita.

Dodavanje kovalentnih polumjera kationa B i aniona X_3, ili pak elektronegativnosti aniona X_3 u skup za učenje nije rezultiralo očekivanim poboljšanjem točnosti modela, što ne mora značiti da ove karakteristike nemaju utjecaj na energijske procjepe, standardnu toplinu stvaranja ili najstabilniju fazu perovskita. Točnost modela u ovim se slučajevima nije poboljšala zbog toga što se i u prvobitnom skupu za učenje nalazila informacija o kationu B i anionu X_3 od kojih je perovskit građen. U procesu učenja modela, svakom od tih aniona i kationa pridjeljene su određene težine kojima je opisan njihov utjecaj na veličinu koju smo modelom pokušavali predvidjeti. Dodavanjem recimo, kovalentnog radijusa kationa B u skup za učenje, naše se predviđanje nije znatno poboljšalo jer smo i već prije imali informaciju o načinu na koji određeni kation B (a time i njegov kovalentni radijus) utječe na veličinu koju predviđamo.

Kako bi određeni perovskit bio povoljan za solarne ćelije, potrebno je da ima energijski procjep oko 1,5 eV, da mu standardna toplina stvaranja bude negativna te da se nalazi u najstabilnijoj fazi kristalne rešetke. Pronađeni perovskiti koji zadovoljavaju ove uvjete su: $CsPbI_2Br$ u kubičnoj, $FAPbIBr_2$ u prvoj ortorompskoj, $FAPbI_3$ u kubičnoj, $CsPbIBr_2$ u tetragonskoj i $MAPbI_3$ u kubičnoj fazi, a od onih koji su sigurniji za okoliš jer nisu građeni od olova pronađeni su $MASnBr_3$ u prvoj ortorompskoj, $FASnI_2Cl$ u kubičnoj, CsSnIBrCl u tetragonskoj i $CsSnCl_3$ u prvoj ortorompskoj fazi. Među pronađenim perovskitima nalaze se i oni trenutno korišteni u solarnim ćelijama te možemo zaključiti da su uvjeti pretrage bili dobro postavljeni i da bi perovskiti izdvojeni u ovom radu mogli biti potencijalni kandidati za izradu solarnih ćelija.

Izračunavanje energijskog procjepa i standardne topline stvaranja za perovskit koji se ne nalazi u bazi podataka može trajati i do nekoliko sati, dok se vrlo dobra procjena istih vrijednosti uz pomoć modela iz ovog rada može dobiti za nekoliko sekundi. Modeli dobiveni u ovom radu stoga su zgodna alternativa za uvid u približni energijski procjep i standardnu toplinu stvaranja perovskita koji se ne nalaze u bazi podataka u nedostatku vremena potrebnog za računanje istih.

Dodaci

Dodatak A Solarne ćelije u osnovnoj i srednjoj školi

Stalan rast potrebe za obnovljivim izvorima energije u svijetu za sobom povlači i rast potrebe za edukacijom članova društva o važnosti korištenja obnovljivih izvora i njihovim značajkama. S ciljem što ranijeg podizanja svijesti o ovoj temi kod budućih nosioca našeg društva, edukacija može započeti već u školskoj i predškolskoj dječjoj dobi.

U Hrvatskoj se u posljednjem desetljeću informiranost učenika o obnovljivim izvorima energije pokušala postići kroz razne projekte pojedinih škola. U Srednjoj školi Oroslavlje i Srednjoj školi Ruđer Bošković Zagreb krajem 2010. proveden je projekt Povećanje znanja i informacija o obnovljivim izvorima energije [65]. Projekt je trajao godinu dana, a u tom je razdoblju po 10 profesora iz obje škole prošlo edukaciju o obnovljivim izvorima te su zatim stečeno znanje prenijeli na profesore ostalih strukovnih škola u za to organiziranim radionicama. U posljednjem koraku projekta u obje je škole odabrana skupina učenika koji su u edukaciji sudjelovali u sklopu izbornog predmeta "Obnovljivi izvori energije". U drugoj polovici 2018. u Elektrotehničkoj i prometnoj školi Osijek proveden je projekt pod imenom Festival obnovljivih izvora energije: srednjoškolci za osnovnoškolce [66], tijekom kojeg su učenici škole osmislili brojne projekte i radionice za demonstraciju tehnologije obnovljivih izvora energije i energetske učinkovitosti, koje su zatim izvodili pred učenicima osnovnih škola. Provedeno je još projekata sličnih ovima, a u nekim su školama i poduzeti prvi koraci prema prelasku na čistu energiju, poput korištenja solarnih panela. U kurikulumu iz fizike obnovljivi izvori energije spominju se i za osnovne i za srednje škole u sklopu početnih lekcija o energiji, najčešće u ishodima za ocjenu vrlo dobar i odličan. Uz opis neobnovljivih i obnovljivih izvora energije te raspravu o njihovim učincima, u nastavi bi se na ovu temu mogao izvesti i kratki pokus ili čak njih nekoliko, kako bi se učenicima dočarale temeljne ideje i principi barem jednog od mehanizama za iskorištavanje obnovljivih izvora energije. Pošto su solarne ćelije jedan od mehanizama kojeg svatko može koristiti u svom vlastitom domu, bilo bi zgodno s učenicima eksperimentalno istražiti neka od njihovih svojstava.

Jedna od karakteristika koja se na nastavi može eksperimentalno odrediti je korisnost solarne ćelije. Za pokus su potrebne solarne ćelije, nekoliko žica, mikrometar (ili ampermetar i voltmetar), trošilo, izvor svjetlosti i ravnalo. Većina škola od tog pribora obično ima sve osim solarnih ćelija, koje se mogu vrlo povoljno naručiti. Za frontalno izvođenje pokusa dovoljna

je i jedna ćelija, dok je za rad u grupama potrebno je 5-6 ćelija. Može se i kupiti set opreme za izvođenje pokusa sa solarnim ćelijama [67]. U slučaju nedovoljne Sunčeve svjetlosti kod izvođenja pokusa može se kao izvor svjetlosti koristiti stolna lampa ili čak opcija svjetiljke na mobitelu. Na početku pokusa mikrometar se spoji na solarnu ćeliju i postavi na opciju za mjerenje napona. Nakon toga osvijetli se solarna ćelija i iščita te zabilježi vrijednost napona. To je unutarnji napon V_U, koji još zovemo i naponom otvorenog strujnog kruga ili naponom praznog hoda, a predstavlja napon na polovima izvora kad kroz strujni krug ne teče struja. Potom se mikrometar prebaci na opciju za mjerenje jakosti struje te se izmjeri i zabilježi jakost struje. Ta vrijednost predstavlja struju kratkog spoja I_{KS}, odnosno maksimalnu struju koju solarna ćelija može proizvesti. Kad su nam poznati napon praznog hoda i struja kratkog spoja, može se izračunati maksimalna teorijska snaga solarne ćelije P_{max} po formuli

$$P_{max} = V_U \cdot I_{KS}. \tag{A.1}$$

Zatim se spoji strujni krug s trošilom i solarnom ćelijom te se mjeri napon na trošilu V, pri čemu se usput ponavlja da se voltmetar u strujni krug spaja paralelno. Nakon toga mjeri se struja u strujnom krugu I, pri čemu se ponavlja da se ampermetar u strujni krug spaja serijski. Kad su poznati napon V i struja I, može se izračunati stvarna snaga koju ćelija daje trošilu, $P = V \cdot I$. Posljednji dio pokusa uključuje računanje korisnosti ćelije, za što je prvo potrebno izračunati sunčevu snagu koja dolazi na ćeliju. Ona je jednaka $P_S = S_i \cdot A$, pri čemu je S_i intenzitet Sunca na površini Zemlje koji iznosi 1000 W/m^2, dok je A površina solarne ćelije koja se mora izračunati mjerenjem duljine i širine solarne ćelije i množenjem ta dva broja. Korisnost solarne ćelije je omjer snage ćelije i snage koja dolazi na ćeliju. Maksimalna teorijska korisnost je prema tome

$$\eta_{max} = P_{max}/P_S, \tag{A.2}$$

a stvarna korisnost $\eta = P/P_S$. Nakon pokusa se prokomentiraju rezultati i donosi zaključak: maksimalna i stvarna korisnost su različite jer se dio svjetlosti koja upada na ćeliju reflektira, a dio blokira [68]. Ovaj je pokus prikladan za izvođenje u srednjim školama.

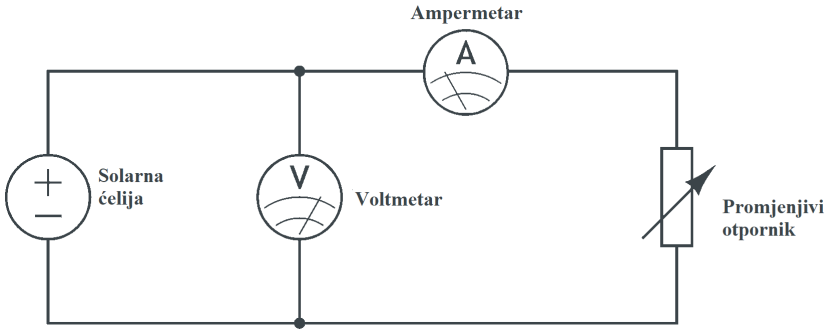
Može se i eksperimentalno ispitati utjecaj intenziteta svjetlosti, kuta upada na solarnu ćeliju i boje svjetlosti na maksimalnu teorijsku korisnost solarne ćelije. Za istraživanje utjecaja intenziteta, mikrometar se spaja na solarnu ćeliju, mijenja se intenzitet svjetlosti te se mjere struja kratkog spoja I_{KS} i napon praznog hoda V_U za nekoliko razina intenziteta. Pri

tome se kut upada, udaljenost izvora od ćelije i boja svjetlosti drže stalnima. Intenzitet se može mijenjati korištenjem žarulja različitih snaga u izvoru svjetlosti, za što se s učenicima prvo mora ponoviti da intenzitet ovisi o snazi žarulje i njenoj udaljenosti od solarne ćelije, a kako udaljenost ne mijenjamo, promjena intenziteta proporcionalna je promjeni snage. Nakon mjerenja se za svaki intenzitet računaju maksimalna teorijska snaga P_{max} prema formuli A.1 te maksimalna teorijska korisnost ćelije η_{max} prema formuli A.2. Nakon toga se komentiraju rezultati i donosi zaključak: pri većem intenzitetu svjetlosti veća je i maksimalna teorijska korisnost solarne ćelije. Za provjeru ovisnosti korisnosti o kutu upada svjetlosti na ćeliju odabire se jedan intenzitet koji se koristi tijekom cijelog pokusa i mijenja se kut pod kojim svjetlost pada na ravninu solarne ćelije. Sve se ostale varijable drže stalnima. Mjere se struja kratkog spoja I_{KS} i napon praznog hoda V_U za kutove od 90, 45 i 0 stupnjeva. Nakon mjerenja računaju se maksimalna teorijska snaga P_{max} prema formuli A.1 te maksimalna teorijska korisnost ćelije η_{max} prema formuli A.2 za svaku vrijednost kuta upada. Zatim se rasprave rezultati i donosi zaključak: maksimalna teorijska korisnost je najveća kad svjetlost upada okomito na ravninu ćelije, a najmanja kad svjetlost upada paralelno s ravninom ćelije. Poslijednji dio pokusa uključuje mjerenje korisnosti za različite boje filtera koji prekrivaju solarnu ćeliju pri stalnom intenzitetu i kutu upada svjetlosti. Ovdje se solarna ćelija prekrije filterom u boji i zatim se mjeri struja kratkog spoja I_{KS} i napon praznog hoda V_U za različite boje filtera. Nakon mjerenja računaju se maksimalna teorijska snaga P_{max} prema formuli A.1 te maksimalna teorijska korisnost ćelije η_{max} prema formuli A.2 za svaku boju filtera. Zatim se komentiraju rezultati i donosi zaključak: korisnost ćelije ovisi o boji filtera. Tu ne treba raspravljati o valnim duljinama pojedine boje jer se to uči tek kasnije. Kad se završi sa sva tri dijela pokusa, s učenicima se može raspraviti što nam dobiveni rezultati govore o položaju i boji solarne ćelije, odnosno da je solarnu ćeliju najpovoljnije postaviti na osvjetljeno mjesto tako da kut upada svjetlosti bude okomit na njenu ravninu te da je bolje imati neobojane ćelije. Ovaj se pokus može izvoditi u srednjoj školi, no moguće ga je i prilagoditi osnovnoj školi ako se izbaci računanje snage i korisnosti te se samo promatraju I_{KS} i V_U.

Zgodno je istražiti i svojstva paralelnog i serijskog spoja solarnih ćelija. Prvo se na solarnu ćeliju spoji mikrometar te se izmjere struja kratkog spoja I_{KS} i napon praznog hoda V_U za jednu ćeliju, a zatim se izračuna maksimalna teorijska snaga P_{max} prema formuli A.1 te maksimalna teorijska korisnost η_{max} prema formuli A.2. Nakon toga se serijski spoje dvije solarne ćelije te se izmjere struja kratkog spoja I_{KS} i napon praznog hoda V_U i opet izračunaju teorijska snaga P_{max} te maksimalna teorijska korisnost spoja η_{max}. Ovaj rezultat

može se usporediti s rezultatom za jednu ćeliju: iznos struje serijskog spoja je jednak iznosu struje koja prolazi kroz jednu solarnu ćeliju, a napon je jednak zbroju napona dviju ćelija. Ukupna snaga n serijski spojenih ćelija jednaka je zbroju n snaga jedne ćelije. Isto pravilo vrijedi i za korisnost. Zatim se ćelije spoje paralelno i opet se izmjere struja kratkog spoja I_{KS} i napon praznog hoda V_U te izračunaju teorijska snaga P_{max} i maksimalna teorijska korisnost spoja η_{max}. Rezultati dobiveni za paralelni spoj usporede se sa strujom i naponom za jednu ćeliju - napon je jednak naponu jedne ćelije, a iznos struje zbroju iznosa struja dvaju ćelija. Ukupna snaga i korisnost n paralelno spojenih ćelija mogu se dobiti zbrajanjem n snaga ili korisnosti jedne ćelije. Na kraju se donese zaključak: snaga i korisnost spojenih ćelija ne ovise o tome jesu li ćelije spojene serijski ili paralelno, ali struja i napon ovise. Za prilagodbu pokusa izvođenju u osnovnoj školi opet se može izostaviti računanje snage i korisnosti.

U srednjim se školama može eksperimentalno odrediti i strujno-naponska karakteristika solarne ćelije. Za taj pokus potreban je promjenjivi otpornik, solarna ćelija, nekoliko žica, voltmetar i ampermetar. Spoji se strujni krug kao na slici A.1 te se otpor promjenjivog otpornika prvo postavi na najmanju vrijednost otpora. Očitaju se struja i napon. Zatim se mijenja vrijednost otpora i za svaku promjenu se očitaju struja i napon. Nakon pokusa slijedi prikaz krivulje strujno-naponske karakteristike te rasprava - za niske vrijednosti otpora jakost struje je konstantna preko određenog ranga vrijednosti napona. Kod većih vrijednosti otpora jakost struje počinje opadati, a napon rasti. Ako su prije izvođenja ovog pokusa izmjereni struja kratkog spoja i napon praznog hoda za solarnu ćeliju, njihove se vrijednosti mogu usporediti s maksimalnim dobivenim vrijednostima struje i napona.



Slika A.1: Strujni krug za određivanje $I-V$ karakteristike solarne ćelije [69].

Ovo je samo nekoliko prijedloga pokusa koji bi se mogli izvesti na nastavi. Može ih se izvesti i pri obrađivanju lekcija iz električne struje, ako se smatra da ih učenici prije toga neće moći u potpunosti shvatiti.

Literatura

[1] Global Energy and CO2 Status Report, http://www.iea.org/geco/, 16.08.2019.

[2] Solar FAQs, http://www.sandia.gov/~jytsao/Solar%20FAQs.pdf, 16.08.2019.

[3] A History of Solar Cells: How Technology Has Evolved, http://tiny.cc/2xsiaz, 16.08.2019.

[4] Crystalline Silicon Photovoltaics Research, http://tiny.cc/uszhbz, 16.08.2019.

[5] Luceño-Sánchez, J. A.; Díez-Pascual A. M.; Capilla, R. P.; Materials for Photovoltaics: State of Art and Recent Developments. // International Journal of Molecular Sciences. Vol. 20(4), 976 (2019), http://tiny.cc/lkhpbz, 16.08.2019.

[6] Extance, A.; Perovskites on trial. // Nature. Vol. 570 (2019), str. 429-432.

[7] How to speed up the discovery of new solar cell materials, http://tiny.cc/rqhpbz, 16.08.2019.

[8] Jordan, M .I.; Mitchell, T. M.; Machine learning: Trends, perspectives, and prospects. // Science. Vol. 349 (2015), str. 255-260.

[9] Hersch, P.; Zweibel, K.; Basic Photovoltaic Principles and Methods. Golden, Colorado: Solar Energy Research Institute, 1982. http://tiny.cc/uuhpbz, 16.07.2019.

[10] Difference Between Photoelectric Effect and Photovoltaic Effect, http://tiny.cc/czwiaz, 16.07.2019.

[11] Energy Band Diagrams, http://tiny.cc/2c6iaz, 16.07.2019.

[12] Valence and core electrons, http://tiny.cc/qq6iaz, 16.07.2019.

[13] Conductors – Insulators – Semiconductors, http://tiny.cc/jx4lbz, 16.07.2019.

[14] A Student Introduction to Solar Energy, http://tiny.cc/5yhpbz, 20.07.2019.

[15] Shockley–Queisser limit, http://tiny.cc/q1hpbz, 20.07.2019.

[16] Atoms and Light Energy, http://tiny.cc/l2ylaz, 20.07.2019.

[17] The Difference between Solar Cell, Module & Array, http://tiny.cc/aj7qaz, 20.07.2019.

[18] The Output Watts of Solar Panels, http://tiny.cc/ex7qaz, 20.07.2019.

[19] Hussain, I.; Tran, H. P.; Jaksik, J.; Moore, J.; Islam, N.; Uddin M. J.; Functional materials, device architecture, and flexibility of perovskite solar cell // Emergent Materials. Vol. 1, 3-4 (2018), str. 133-154, http://tiny.cc/0aipbz, 22.07.2019.

[20] Understand the Advantages, Disadvantages of Different Solar Cells and Who the Market Leaders Are, http://tiny.cc/4bipbz, 22.07.2019.

[21] Perovskite: Mineral information, data and localities, http://tiny.cc/rdipbz, 07.08.2019.

[22] Perovskites and Perovskite Solar Cells: An Introduction, http://tiny.cc/aeipbz, 07.08.2019.

[23] Liu, W; Liu, Y; Wang, J; Wu, C; Liu, C; Xiao, L; Chen, Z; Wang, S; Gong, Q.; Twin Domains in Organometallic Halide Perovskite Thin-Films. // Crystals. Vol 8, Special Issue: Metal Halide Perovskite Crystals: Growth Techniques, Properties and Emerging Applications (2018), str 141-158. http://tiny.cc/xgipbz, 10.08.2019.

[24] Krishnan, U.; Kaur, M.; Kumar, M.; Kumar, A.; Factors affecting the stability of perovskite solar cells: a comprehensive review. // Journal of Photonics for Energy. Vol. 9,2 (2019), http://tiny.cc/vc7uaz, 10.08.2019.

[25] Mesoporous Material, http://tiny.cc/ce29az

[26] Best Research-Cell Efficiency Chart, http://tiny.cc/esipbz, 12.08.2019.

[27] Meng, L.; You, J.; Yang, Y.; Addressing the stability issue of perovskite solar cells for commercial applications. // Nature Communications. Vol 9 (2018), http://tiny.cc/kyipbz, 12.08.2019.

[28] Enthalpy of Formation, http://tiny.cc/0zipbz, 12.08.2019.

[29] Nagabhushanaa, G. P.; Shivaramaiaha, R.; Navrotskya, A.; Direct calorimetric verification of thermodynamic instability of lead halide hybrid perovskites. // PNAS. Vol. 113, 28 (2016), str. 7717-7721. http://tiny.cc/05ipbz, 12.08.2019.

[30] PSC: Methods of Increasing Stability and Durability, http://tiny.cc/dq4nbz, 12.08.2019.

[31] Shojaei, F.; Yin, W. J.; Stability Trend of Tilted Perovskites // The Journal of Physical Chemistry C. Vol. 122,27 (2018) str. 15214-15219, http://tiny.cc/2gjpbz, 14.08.2019.

[32] Tao, S. X.; Cao, X.; Bobbert, P. A.; Accurate and efficient band gap predictions of metal halide perovskites using the DFT-1/2 method: GW accuracy with DFT expense. // Scientific Reports. Vol. 7. (2017), http://tiny.cc/nljpbz, 14.08.2019.

[33] Silva, E.; Silva, D.; Skelton, J.; Parker, S.; Walsh, A.; Phase stability and transformations in the halide perovskite CsSnI3 // Physical Review B. Vol. 91 (2015), http://tiny.cc/txjpbz, 14.08.2019.

[34] Wang, T.; Daiber, B.; Frost, J. M.; Mann, S. A.; Garnett, E. C.; Walsh, A.; Ehrler, B.; Indirect to direct bandgap transition in methylammonium lead halide perovskite. // Energy & Environmental Science. Vol. 10 (2017) str. 509-515. http://tiny.cc/1xjpbz, 14.08.2019.

[35] What's the Big Deal With Data?, http://data.bsa.org, 05.06.2019.

[36] Appenzeller, T.; The scientists' apprentice. // Science. Vol. 357, 6346 (2017), str. 16-17.

[37] Mitchell, T. M. Machine Learning. New York: McGraw-Hill, 1997.

[38] VanderPlas, J.; Python Data Science Handbook. Sebastopol: O'Reilly Media, 2017. http://tiny.cc/24jpbz, 05.06.2019.

[39] Müller, A. C.; Guido, S.; Introduction to Machine Learning with Python. Sebastopol: O'Reilly Media, 2017. http://tiny.cc/f6jpbz, 05.06.2019.

[40] Chapelle, O.; Schölkopf, B.; Zien, A.; Semi-Supervised Learning. Cambridge, Massachusetts: MIT Press, 2006.

[41] Supervised and Unsupervised Machine Learning Algortihms, http://tiny.cc/lu4l9y, 05.06.2019.

[42] Holdout cross-validation, http://tiny.cc/258r9y, 05.06.2019.

[43] K-fold cross-validation, http://tiny.cc/478r9y, 05.06.2019.

[44] Goodfellow, I.; Bengio Y.; Courville A.; Deep Learning. MIT Press, 2016.

[45] Linear Regression, http://tiny.cc/0v6v9y, 10.06.2019.

[46] Dot Products, Transposes, and Orthogonal Projections, http://tiny.cc/gdbw9y, 10.06.2019.

[47] How to Solve Linear Regression Using Linear Algebra, http://tiny.cc/emmw9y, 10.06.2019.

[48] Matrix Differentiation, http://tiny.cc/fvfz9y, 10.06.2019.

[49] OLS in Matrix Form, http://tiny.cc/fwfz9y, 10.06.2019.

[50] A Simple Explanation of Gini Impurity, https://victorzhou.com/blog/gini-impurity/, 29.08.2019.

[51] How Are Decision Trees Constructed In Machine Learning, http://tiny.cc/nla49y, 12.06.2019.

[52] Cover T. M.; Thomas J. A.; Elements of Information Theory. New York: John Wiley & Sons, 1991.

[53] Machine Learning Exercise, http://tiny.cc/qa29az, 20.08.2019.

[54] O Pythonu, http://tiny.cc/tqkpbz, 20.08.2019.

[55] O Jupyter bilježnici, http://tiny.cc/v719az, 20.08.2019.

[56] O NumPy paketu, https://www.numpy.org/, 20.08.2019.

[57] O Matplotlib biblioteci, https://matplotlib.org, 20.08.2019.

[58] O ASE biblioteci, http://tiny.cc/lpkpbz, 20.08.2019.

[59] O scikit-learn bibilioteci, https://scikit-learn.org/, 20.08.2019.

[60] O Graphvizu, https://www.graphviz.org/, 20.08.2019.

[61] O GPAW-u, https://wiki.fysik.dtu.dk/gpaw/, 20.08.2019.

[62] Space group, https://www.britannica.com/science/space-group, 18.08.2019.

[63] Kuisma, M.; Ojanen, J; Enkovaara, J.; and Rantala, T. T.; Kohn-Sham potential with discontinuity for band gap materials // Physical Review B. Vol. 82, 115106 (2010), https://arxiv.org/abs/1003.0296, 18.08.2019.

[64] Targhi, F. F.; Jalili, Y. S.; Kanjouri, F. $MAPbI_3$ and $FAPbI_3$ perovskites as solar cells: Case study on structural, electrical and optical properties. // Results in Physics. Vol. 10. (2018) str. 616-627, http://tiny.cc/3on1bz, 18.08.2019.

[65] Povećanje znanja i informacija o obnovljivim izvorima energije, http://tiny.cc/139ybz, 25.08.2019.

[66] Zimska škola obnovljivih izvora energije, http://tiny.cc/scazbz, 25.08.2019.

[67] Oprema za učionice: solarna ćelija, http://tiny.cc/wdbvbz, 25.08.2019.

[68] Output of a Solar Cell, http://tiny.cc/p282bz, 25.08.2019.

[69] PV infinity: Teacher guides, http://tiny.cc/e382bz, 25.08.2019.