# Upotreba umjetne inteligencije u klasifikaciji analitičkih podataka i prepoznavanju uzoraka ponašanja

**Gajdek, Dorotea**

**Undergraduate thesis / Završni rad**

**2016**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Zagreb, Faculty of Chemical Engineering and Technology / Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije**

*Permanent link / Trajna poveznica:* https://urn.nsk.hr/urn:nbn:hr:149:030488

*Rights / Prava:* In copyright / Zaštićeno autorskim pravom.

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

FKITMCMXIX

*Repository / Repozitorij:*

Repository of Faculty of Chemical Engineering and Technology University of Zagreb

zir.nsk.hr

dabar
DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI PREDDIPLOMSKI STUDIJ

Dorotea Gajdek

# ZAVRŠNI RAD

Zagreb, srpanj 2016.

SVEUČILIŠTE U ZAGREBU
FAKULTET KEMIJSKOG INŽENJERSTVA I TEHNOLOGIJE
SVEUČILIŠNI PREDDIPLOMSKI STUDIJ

Dorotea Gajdek

# UPOTREBA UMJETNE INTELIGENCIJE U KLASIFIKACIJI ANALITIČKIH PODATAKA I PREPOZNAVANJU UZORAKA PONAŠANJA

**ZAVRŠNI RAD**

Voditelj rada: doc. dr. sc. Šime Ukić

Članovi ispitnog povjerenstva: doc. dr. sc. Šime Ukić
dr. sc. Lidija Furač, viši predavač
dr. sc. Suzana Sopčić, viši znanstveni suradnik

Zagreb, srpanj 2016.

*Zahvaljujem svome mentoru doc. dr. sc. Šimi Ukiću na vodstvu, nesebičnoj pomoći i strpljenju tijekom izrade završnoga rada.*

*Također, zahvaljujem svojim roditeljima i bratu, prijateljima i dečku na njihovoj iznimnoj podršci i ljubavi za vrijeme mog studija i izrade ovoga rada.*

## Sažetak

Umjetna inteligencija (AI) je skup metoda kojima uređaji obavljanju inteligentne radnje. Glavni problemi, odnosno ciljevi AI, su: razumijevanje, znanje, planiranje, učenje, komunikacija, percepcija te sposobnost manipuliranja te pomicanja objekata.

AI koristi veliki broj alata kao što su: metode matematičke optimizacije, logičko zaključivanje, metode bazirane na vjerojatnosti... Područje AI je interdisciplinarno te uključuje računalne znanosti, matematiku, psihologiju, lingvistiku, neuroznanost, kao i specijalizirane znanosti kao što je umjetna psihologija.

U ovom radu fokus je na primjeni AI u analitičkoj kemiji. Primjena AI u analitičkoj kemiji ima dugu povijest, no tek je ubrzani razvoj informatičkih znanosti u zadnjih nekoliko desetljeća povećao intenzitet i raznolikost njene upotrebe i doveo je do razine da je postala uobičajeni alat za modeliranje ili obradu analitičkih podataka.



**Ključne riječi:** umjetna inteligencija, klasifikacija podataka, prepoznavanje uzoraka ponašanja, analitička kemija

## Summary

Artificial intelligence (AI) is the intelligence exhibited by machines. The main problems of AI research include: reasoning, knowledge, planning, learning, natural language processing (communication), perception, and the ability to move and manipulate objects.

There is large number of tools applied in AI, including mathematical optimization techniques, logic, methods based on probability, and many others. The AI field is interdisciplinary, including computer science, mathematics, psychology, linguistics, neuroscience, as well as other specialized fields such as artificial psychology.

In this specific work, the focus will be on AI in analytical chemistry. The application of AI in analytical chemistry has a long history. Yet, the benefits provided by significant development of informatics in recent decades generated diversity of AI applications and significantly increased its importance.

**Key words:** artificial intelligence, data classification, pattern recognition, analytical chemistry

# Sadržaj

1. Uvod ..... 1

2. Umjetna inteligencija ..... 3

2.1. Povijesni razvoj umjetne inteligencije ..... 4

2.2. Umjetne neuronske mreže ..... 5

2.2.1. Vrste neuronskih mreža ..... 10

2.3. Neizrazita logika ..... 14

2.3.1. Usporedba binarne i neizrazite logike ..... 14

2.3.2. Neizraziti skupovi ..... 16

2.4. Genetički algoritam ..... 17

2.4.1. Operator selekcije ..... 18

2.4.2. Operator križanja ..... 19

2.4.3. Operator mutacije ..... 20

2.5. Hibridne tehnike umjetne inteligencije ..... 21

3. Prepoznavanje uzoraka ..... 22

3.1. Podjela metoda prepoznavanja uzoraka ..... 24

3.1.1. Eksploracijska analiza podataka ..... 24

3.1.2. Nenadgledano prepoznavanje uzoraka ..... 27

3.1.3. Nadgledano prepoznavanje uzoraka ..... 32

4. Analiza analitičkih podataka ..... 36

4.1. Određivanje odstupajućih podataka ..... 37

4.1.1. Dixonov *Q*-test ..... 37

4.1.2. Grubbsov test ..... 39
4.2. Linearna regresija ..... 40
4.2.1. Model jednostavne linearne regresije ..... 40
4.2.2. Model višestruke linearne regresije ..... 43
4.3. Nelinearna regresija ..... 44
4.3.1. Polinomni regresijski model ..... 44
4.3.2. Jednostavni eksponencijalni regresijski model ..... 45

5. Primjena ..... 46
5.1. Praktične primjene neuronskih mreža ..... 47
5.2. Primjena metode prepoznavanja uzoraka ..... 48

6. Zaključak ..... 50

7. Literatura ..... 52

Životopis ..... 58

# 1. Uvod

Umjetna inteligencija je izum ljudskog roda čiji je primarni cilj objasniti način djelovanja biološkog svijeta te primijeniti najbolje karakteristike u načinu djelovanja ljudskoga života. Život kakav danas poznajemo stvar je biologije, a život kakav bi mogao biti je stvar umjetne inteligencije.[1] Područje umjetne inteligencije je spoj bioloških znanja o ljudskom mozgu i njegovom živčanom sustavu te računalnih znanosti i tehnika. To je dio znanosti koji se svakim danom razvija te otkriva nešto novo.

Samo znanje o ljudskom mozgu je poprilično ograničeno, znanstvenici još uvijek ne uspijevaju u potpunosti shvatiti kako i zašto naš mozak nešto radi. Umjetna inteligencija proučavanjem djelovanja našeg mozga nastoji kreirati i razviti algoritme i metode koje će pomoći da upoznamo sami sebe, ljudsku rasu, te pomoći u rješavanju raznih medicinskih, biomedicinskih te ostalih problema sa kojima se ljudi susreću.

Iako je baza umjetne inteligencije u statistici i matematici te računalnim algoritmima, ona nije samo to – matematika. Primjenom neizrazite logike nastoji se umjetne neuronske mreže naučiti da razmišljaju „izvan okvira“ binarne logike. Umjetne neuronske mreže uče kako primati te analizirati podatke koji nisu linearni čime se približavaju ljudskome mozgu, a znanstvenicima koji se njima koriste uvelike pomažu u obradi i analizi velikog broja podataka koji vode ka rješavanju raznih medicinskih te ostalih znanstvenih pitanja. Jedna od metoda koja sigurno tome doprinosi je i prepoznavanje uzoraka o čemu će biti riječi. Prepoznavanje uzoraka omogućuje umjetno stvorenom neuronu da zapamti obilježja promatranih situacija i podataka čime su se znanstvenici još više približili poboljšanju kvalitete samog života.

# 2. Umjetna inteligencija

## 2.1. Povijesni razvoj umjetne inteligencije

Godine 1943., neuroznanstvenik Warren S. McCulloch, i logičar Walter Pitts osmislili su matematički model kojim su pokušali simulirati rad središnjeg živčanog sustava živih bića. Time su razvili prvi model umjetne neuronske mreže (engl. *artificial neural network*). Svoj opis umjetnog neurona kao tvorevine koji prima podatke, obrađuje ih te prosljeđuje drugim neuronima, objavili su u *The Bulletin of Mathematical Biophysics*[2].

Nakon prvog modela neurona, 1958. godine je Rosenblatt predstavio novi model umjetnog neurona kojeg je nazvao *perceptron*[3]. Perceptron je prvotno zamišljen kao uređaj, no kako se razvijao, napredovao je do modela koji omogućava kategoriziranje ulaznih signala.

Ubrzo nakon predstavljanja perceptrona, Widrow i Hoff su 1960. godine predstavili novi model umjetnog neurona nazvanog *ADALINE*[4]. ADALINE je jednoslojni linearni umjetni neuron koji je baziran na neuronu McCullocha i Pittsa te za učenje koristi metodu najmanjih kvadrata (engl. *least mean square*).

Razvitkom modela umjetnih neurona, počela se razvijati i neizrazita logika (engl. *fuzzy logic*). Njen začetnik, L. A. Zadeh, dao je 1965.[5] teoriju prema kojoj neizrazita logika proizlazi iz međusobnog povezivanja, percepcije, zaključivanja i spoznaje. Njegova teorija je karakterizirana kao struktura kojoj nedostaju jasno definirane granice. Dakle, neizrazita logika služi za opisivanje elemenata čije granice nisu jasno definirane, te problema koji su nejasno opisani. Neizrazita logika se smatra alatom koji omogućuje manipulaciju i rad sa neodređenim podacima.[6]

Paralelno s time, 50-ih godina XX. stoljeća, Alan Turing predložio je tzv. „uređaj za učenje“ (engl. *learning machine*) koji bi bio temeljen na načelima evolucije[7]. Njegove ideja uskoro je postala intenzivno proučavana, a možda ključan korak uradio je desetak godina nakon američki znanstvenik John Henry Holland osmišljavanjem genetičkog algoritma (engl. *genetic algorithm*) kojeg je detaljno predstavio u svojoj knjizi iz 1975.[8] Njegov prvotni cilj nije bio kreirati algoritam koji će rješavati samo određeni problem nego promatrati fenomen prilagodbe svake iduće generacije koji se javlja u prirodi. Krajnji pak cilj bio je ostvariti način na koji će se mehanizmi prirodne adaptacije usaditi u računalne sustave.[9]

## 2.2. Umjetne neuronske mreže

Umjetne neuronske mreže (ANN) se mogu najbolje opisati kao računalni modeli sa određenim karakteristikama kao što su sposobnost prilagodbe i učenja, sposobnost generaliziranja te prikupljanja i organizacije podataka.[1] Njihova je uloga da pokušaju simulirati djelovanje središnjeg živčanog sustava živih bića, odnosno mozga. Sastoje se od jednostavnih procesnih jedinica čiji je zadatak međusobna komunikacija slanjem signala preko velikog broja poveznica. Neuroni i živčane stanice su procesne jedinice te je njihov zadatak primanje, obrada i slanje signala do idućeg primatelja.

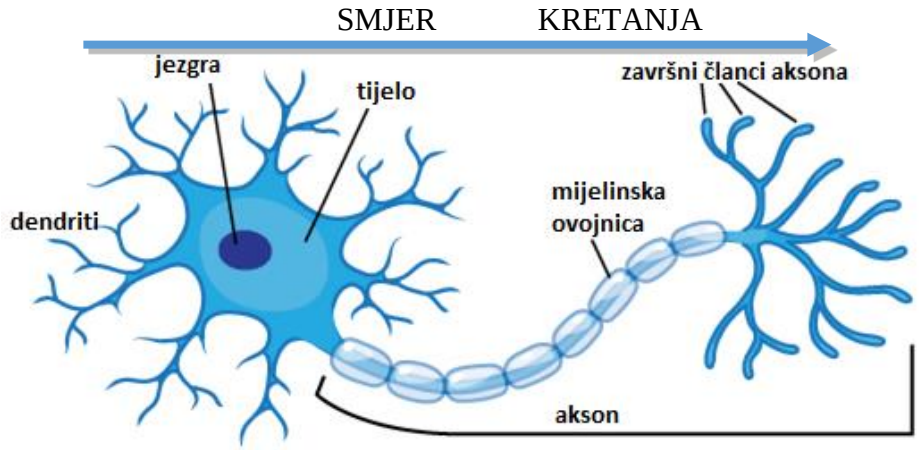
Neuronske mreže rade na principu spajanja neurofizioloških spoznaja o biološkim neuronima te mrežnom sustavu koji povezuje biološke neurone. Umjetne neuronske mreže su simulacija bioloških neuronskih mreža[10], te kao takve koriste vrlo jednostavne matematičke operacije poput zbrajanja i množenja kako bi riješile kompleksne matematičke probleme. Tradicionalni algoritam će upotrijebiti kompleksni set jednadžbi te ih iskoristiti samo za rješavanje danog problema dok će umjetne neuronske mreže biti računski i algoritamski vrlo jednostavne te će imati samoorganizacijske značajke koje će omogućiti primjenu algoritama i jednadžbi na razne vrste problema.

Umjetne neuronske mreže su zamišljene tako da uz svoju jednostavnost te široku primjenu, zadrže otpornost na oštećenja. Navedena karakteristika je preuzeta iz poznavanja bioloških neurona za koje se zna da posjeduju ovu karakteristiku, npr. veliki broj živčanih stanica prirodno odumre u ljudskom mozgu svake godine, no ljudski mozak nastavlja dalje funkcionirati. Otpornost se narušava onda kada u vrlo kratkom vremenu odumre veliki broj živčanih stanica, npr. veliki moždani udar. Informacije koje su vezane za živčane stanice koje su pogođene oštećenjem se gube trajno ili ih je vrlo teško vratiti. Stoga, umjetne neuronske mreže nastoje razviti otpornost na oštećenja tako da čak i velike poteškoće ne uzrokuju gubitak podataka. Trenutne umjetne neuronske mreže su simulirane na računalima te njihova otpornost na oštećenja nije dobro izražena.[10]

Da bi se bolje razumio način djelovanja umjetnih neuronskih mreža, potrebno je detaljno objasniti sastav te djelovanje bioloških neuronskih mreža.

Biološke neuronske mreže se sastoje od živčanih stanica (neurona). Neuroni su sastavni dio ljudskog mozga koji posjeduje oko $10^{11}$ neurona od čega je svaki neuron povezan sa $10^4$

neurona.[11] Unatoč tome što je svaki neuron povezan sa velikim brojem drugih neurona, to ne znači da je svaki signal upućen prema neuronu jednake važnosti. Svaki biološki neuron se sastoji od tijela neurona okruženog nitastim tvorevinama zvanim dendriti čija je funkcija prikupljanje podražaja (informacija) iz okoline (slika 1). Na tijelo neurona nastavlja se cilindrično dugo vlakno zvano akson koje završava nizom završnih članaka.



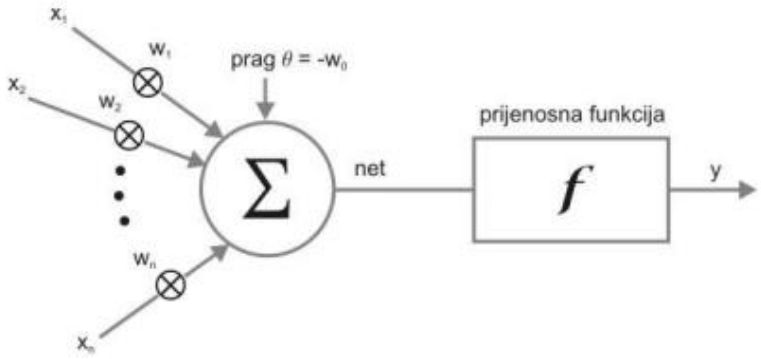

**Slika 1.** Prikaz građe biološkog neurona.

Aksoni služe za prijenos živčanih signala s tijela neurona na drugi neuron ili organ. Signal s jednoga neurona na drugi prelazi na način da prvi neuron odašilje živčani signal u obliku tzv. post-sinaptičkog potencijala[12]. Dakle, na sinapse, koje predstavljaju spojna mjesta dva susjedna neurona dolazi post-sinaptički potencijal prvoga neurona. Taj signal uzrokuje smanjenje ili povećanje potencijala stanice promatranog neurona. U tijelu neurona se nalazi informacija o potencijalu stanice te se tamo prikupljaju post-sinaptički potencijali svih neurona koji u tome trenutku šalju signale na promatrani neuron. U trenutku kada potencijal stanice promatranog neurona prijeđe granični, tj. kritični potencijal stanice, odašilje se akcijski potencijal.[12] Akcijski potencijal se zatim prenosi preko aksona do njegovih završetaka gdje se, ovisno o jačini emitiranog akcijskog potencijala, stvaraju i otpuštaju kemikalije nazvane neurotransmiteri koje zatim na idućem neuronu izazovu jednaki slijed događaja kao na promatranom neuronu. Vrlo je bitno naglasiti kako se signal širi samo u jednom smjeru preko neurona te da ne postoji mogućnost vraćanja signala kroz neuron.

U početku, umjetne neuronske mreže nisu bile zamišljene kako bi se kopirao način djelovanja bioloških neuronskih mreža nego kao način rješavanja određenih problema i zadataka.

Gledani zasebno, dijelovi umjetnog neurona su vrlo jednostavni, određeni elementi primaju ulazne signale, određeni ih obrađuju te ih određeni prosljeđuju dalje do drugog umjetnog neurona. Unatoč jednostavnosti djelovanja svakog pojedinog dijela umjetnog neurona, mreža neurona je u mogućnosti primati te obrađivati veliki broj kompleksnih signala te problema gdje se vidi povezanost sa biološkim neuronom.[13] Umjetne neuronske mreže se razlikuju od klasičnih računalnih algoritama po tome što nisu linearne te se pomoću nelinearnosti približavaju načinu rada bioloških neuronskih mreža, nelinearnost je važna značajka umjetnih neuronskih mreža.

Kao što je ranije navedeno, jedna od značajki koja razlikuje umjetne neurone od klasičnih računalnih algoritama je sposobnost učenja koja omogućava umjetnom neuronu da prilagodi svoju strukturu jačini odnosno vrsti signala kojeg prima. Stoga svaka poveznica između dva neurona ima određenu važnost, tj. težinu, *w* (engl. *weight*, slika 2), kojom množimo ulazni signal *x*.




**Slika 2**. Shema umjetnog neurona.

Kod bioloških neurona signal prelazi na drugi neuron u onome trenutku kada se formira akcijski potencijal. Kod umjetnih neurona računa se suma umnožaka ulaznih signala i pripadajućih težina, *S*:

$$S = \sum_{i=1}^{n} w_i x_i + \theta \qquad (1)$$

Za daljnji prijenos informacije koristi se potom aktivacijska funkcija, *f*, koja se stoga često naziva i prijenosna funkcija (engl. *transfer function*).

$$y = f\left( \sum_{i=1}^{n} w_i x_i + \theta \right) \qquad (2)$$

U jednadžbama (1) i (2) $\theta$ predstavlja tzv. prag (engl. *bias*), tj. vrijednost koju je potrebno prijeći da dođe do aktivacije sume *S*.

Postoji više vrsta aktivacijskih funkcija, a među najučestalijima su:

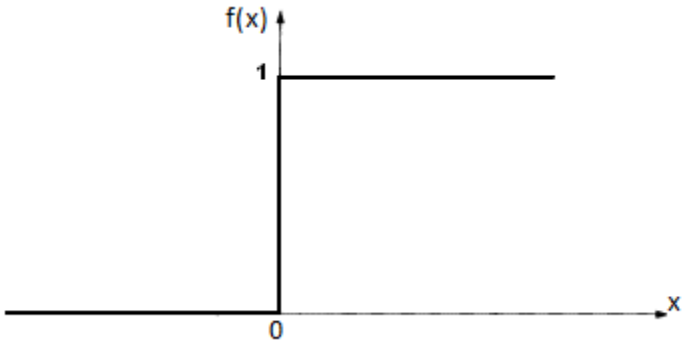
1. skokovita aktivacijska funkcija (engl. *threshold function*)
2. aktivacijska funkcija linearna po odsječcima (engl. *piecewise linear function*)
3. sigmoidna aktivacijska funkcija (engl. *sigmoid function*)

*Skokovita aktivacijska funkcija*

Skokovita aktivacijska funkcija, prikazana na slici 3, dana je jednadžbom:

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

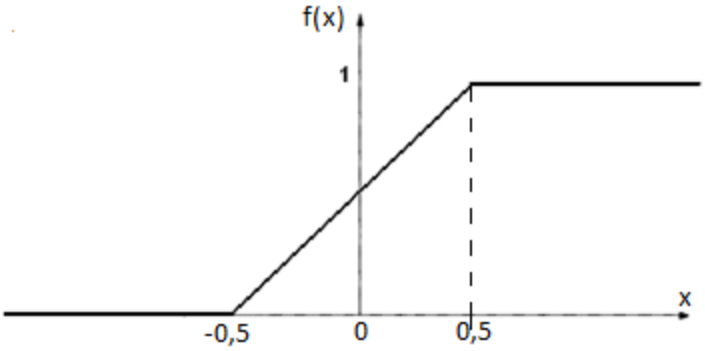
Ova funkcija sve ulazne vrijednosti koje se nalaze iznad praga, preslikava u vrijednost 1, a ostale vrijednosti, koje se nalaze ispod praga, preslikava u nulu. Uobičajeno se kao prag uzima vrijednost 0, te ovakva skokovita funkcija daje binarni izlaz iz neurona.[14]




**Slika 3.** Skokovita aktivacijska funkcija.

*Aktivacijska funkcija linearna po odsječcima*

Aktivacijska funkcija linearna po odsječcima dobiva se kada je izlaz iz umjetnog neurona ograničen na određeni interval. Linearni dio ove funkcije je onaj koji je unutar zadanih granica koje su najčešće -0,5 i 0,5. Sve ostale vrijednosti se preslikavaju u nulu ako su manje od 0,5 te u jedinicu ako su veće od 0,5.[14]




**Slika 4.** Funkcija linearna po odsječcima.

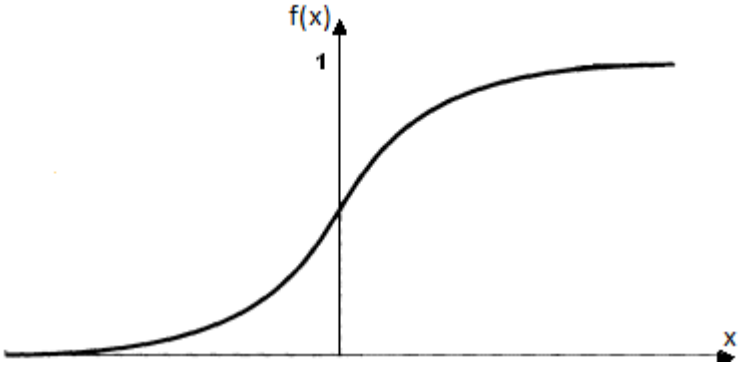
Kao što je prikazano na slici 4, aktivacijska funkcija linearna po odsječcima se definira pomoću jednadžbe:

$$f(x)=\begin{cases}1, x \geq 0,5\\ x, -0,5 < x < 0,5\\ 0, x \leq 0,5\end{cases} \tag{4}$$

*Sigmoidna aktivacijska funkcija*

Najčešće korištena aktivacijska funkcija jest sigmoidna funkcija:

$$f(x)=\frac{1}{1+e^{-x\cdot a}} \tag{5}$$




**Slika 5.** Sigmoidna funkcija.

Ova funkcija sadrži parametar nagiba *a* čijim se povećanjem smanjuje područje „linearnosti“ (slika 5).[14]

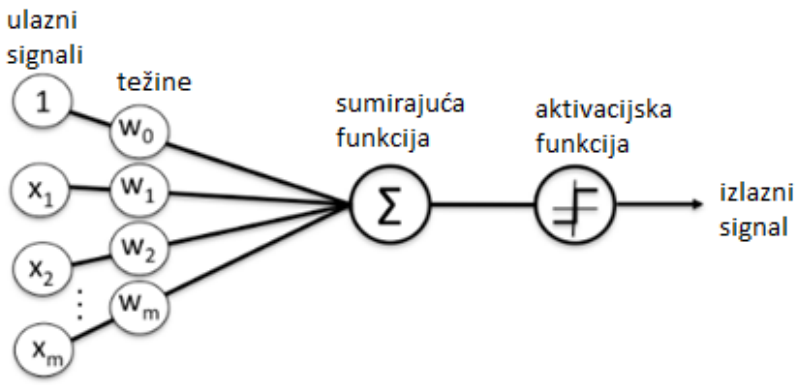
### 2.2.1. Vrste neuronskih mreža

*Perceptron*

Perceptron, prikazan na slici 6 je prvi model unaprijedne umjetne neuronske mreže. Perceptron koristi skokovitu aktivacijsku funkciju te je prvotno namijenjen obradi linearnih ulaznih signala. Ovaj umjetni neuron djeluje na način da preuzima ulazne signale u obliku vektora te ih pretvara u izlazni signal koji je aktivacijska funkcija praga:

$$f(x) = \begin{cases} 1, & w \cdot x + \theta > 0 \\ 0, & w \cdot x + \theta \leq 0 \end{cases} \tag{6}$$

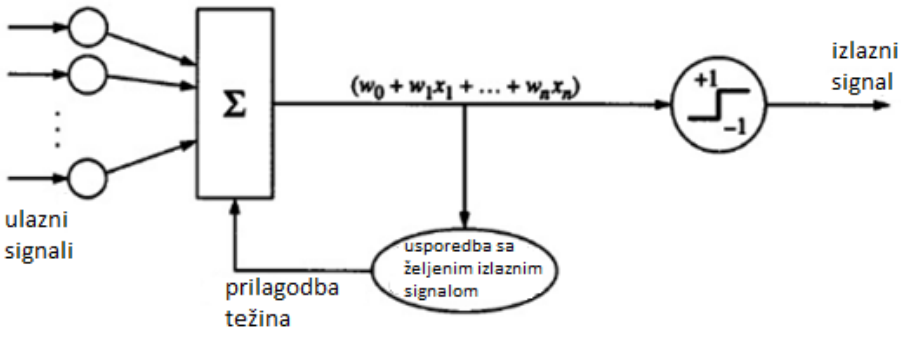
Pri tome je $\theta$ *bias*, odnosno faktor koji pomiče granicu dalje od zadane vrijednosti i nije ovisan o ulaznim vrijednostima[15], $w$ vektor težina, a $w \cdot x$ skalarni produkt definiran kao $\sum_i w_i \cdot x_i$ .




**Slika 6.** Shematski prikaz Rosenblattovog perceptrona.

*ADALINE*

Idući model umjetnog neurona je *ADALINE* (slika 7). Ovaj model je precizniji od perceptrona zato što nakon svake interakcije provjerava da li težina odgovara svim ulaznim signalima, dok perceptron to ne radi. ADALINE također koristi skokovitu aktivacijsku funkciju te obrađuje linearne ulazne signale.

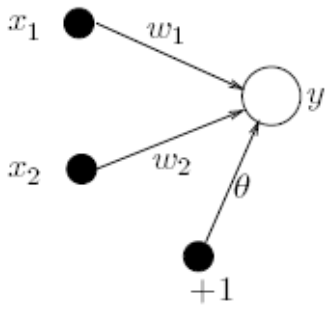



**Slika 7.** Shematski prikaz ADALINE umjetnog neurona.

ADALINE radi na principu da nakon svakog zaprimljenog ulaznog signala provjerava sve dosadašnje ulazne signale, dobivene izlazne signale te zadane željene izlazne signale pri čemu prikupljene informacije koristi kako bi podesio težinu signala u korist željenim izlaznim vrijednostima. Widrow i Hoff su postavili pravilo koje osigurava minimiziranje pogreške u LMS-u putem niza težina te se zato potencijalni lokalni problemi brzo ukljanjaju.[16]

*Unaprijedne neuronske mreže*

Kod unaprijednih neuronskih mreža (enlg. *feed-forward network*, FF) signal putuje isključivo u jednom smjeru: od ulaza prema izlazu. Promotrimo jednostavan slučaj mreže sastavljene od dva ulazna signala i jednim izlaznim signalom kao što je prikazano na slici 8.




**Slika 8.** Jednostruka neuronska mreža sa dva ulazna i jednim izlaznim signalom.

Ulazni signali, *x*, množe se s odgovarajućim težinama, *w*, prenose na sljedeći neuron, te tamo sumiraju. Dobivena suma aktivira se odgovarajućom funkcijom, *f* (može biti linearna ili nelinearna), dajući izlazni signal, *y*:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + \theta\right) \tag{7}$$

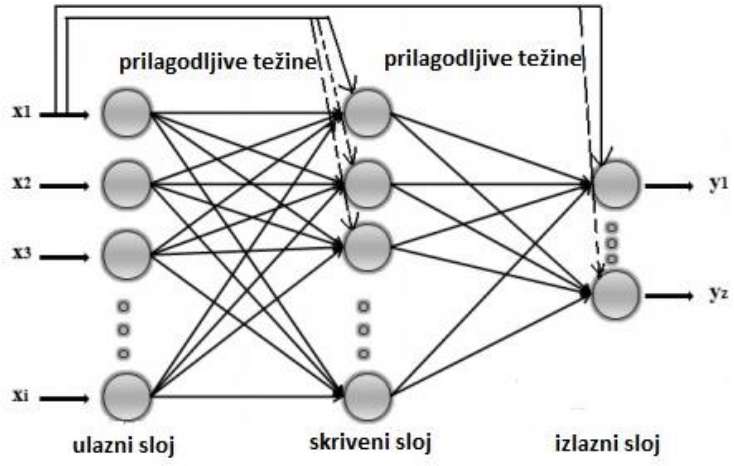
Kao što je navedeno, ova vrsta neuronskih mreža koristi aktivacijsku funkciju praga koja glasi:

$$y = \begin{cases} +1, & \sum_{i=1}^{n} w_i x_i + \theta > 0 \\ -1, & \sum_{i=1}^{n} w_i x_i + \theta \leq 0 \end{cases} \tag{8}$$

Mreža koja je ovako postavljena može ulazne podatke svrstati u dvije zasebne kategorije: u slučaju kada je zbroj ulaznih signala pozitivan, ulazni podatak će biti svrstan u kategoriju +1; u obrnutom slučaju, biti će svrstan u kategoriju -1.

*Kaskadne unaprijedne mreže*

Kaskadna unaprijedna mreža (engl. *cascade-forward neural network,* CF) je neuronska mreža slična običnoj unaprijednoj. Neuroni ovakvih neuronskih mreža nisu povezani samo s neuronima iz prethodnog sloja nego su povezani s neuronima iz više prethodnih slojeva. Između svih slojeva se prilagođavaju težine kako bi izlazni sloj dobio informacije što točnije željenima te ih proslijedio do idućeg neurona u obliku izlaznih signala $y_1$ i $y_2$ (slika 9)[18].




**Slika 9.** Shematski prikaz kaskadne unaprijedne mreže.[18]

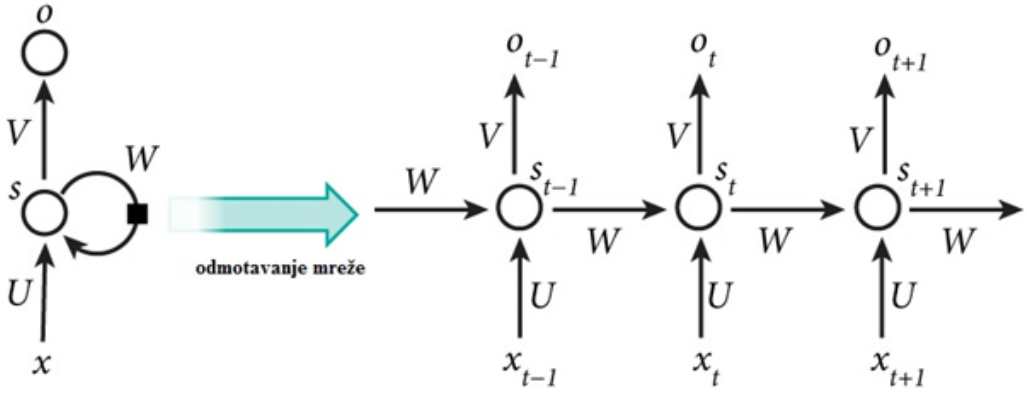
*Povratne neuronske mreže*

Za razliku od ranije navedenih neuronskih mreža, koje obrađuju ulazne signale ne uzimajući u obzir prethodno obrađene ulazne signale, povratne neuronske mreže (engl. *feedback* ili *recurrent neural networks*, RNN) djeluju tako da obrađuju skup ulaznih podataka uzimajući u obzir svaki od njih te njihovu povezanost. To bi značilo da one „pamte“ ranije obrađene informacije te prema tome obrađuju nove ulazne signale. U teoriji, ove mreže bi trebale pamtiti vrlo duge nizove ranije obrađenih informacija, no nažalost je to u praksi ograničeno na samo nekoliko koraka unazad.

Na slici 10 je prikazana povratna neuronska mreža te njeno „odmotavanje“ radi lakšeg poimanja rada ove mreže. Dakle, RNN mreža će se podijeliti u onoliko slojeva koliko je podataka u ulaznom skupu (npr. ako je ulazni skup rečenica od tri riječi, RNN mreža će sadržavati tri sloja). Kao što je prikazano na slici 10, $x$ označava ulazne signale, tj. njihov skup. Shodno tome, $x_t$ označava ulazni signal primljen u vremenu $t$ što u primjeru rečenice sa tri riječi može označavati drugu riječ. Isto tako, $x_{t-1}$ i $x_{t+1}$ označavaju pojedinačne ulazne signale u vremenima $t$-1 i $t$+1. Oznaka $s_t$ predstavlja skriveno stanje koraka $t$, ono označava „memoriju“ neuronske mreže, a izračunava se na temelju prethodnog skrivenog stanja te trenutnog ulaznog signala:

$$s_t = f\left(U_{x_t} + W_{s_{t-1}}\right) \tag{9}$$

gdje $U$ i $W$ predstavljaju parametre koji su ranije naučeni te su isti za svaki korak, čime pokazuju da se svaka informacija iz ulaznog skupa obrađuje na jednak način. Naposljetku, izlazni signal, $o$, se dobiva iz svakog sloja što znači da će izlazni signal $o_t$ pripadati koraku $t$ pri čemu je $V$ parametar koji se koristi za izračunavanje izlaznog signala.[19]




**Slika 10.** Prikaz RNN mreže i njenog „odmotavanja“.[19]

## 2.3. Neizrazita logika

Ideja neizrazite logike je približiti način funkcioniranja ljudskog mozga i matematičkih modela. Neizrazita logika omogućava elementu da djelomično pripada skupu. Dakle, u neizrazitoj logici tvrdnje nisu ograničene na istinite (1) i neistinite (0) nego mogu biti i djelomično istinite što znači da poprimaju vrijednosti između 0 i 1. Za razliku od neizrazite logike, klasična binarna logika omogućava elementu da pripada, ili ne pripada skupu, tj. da vrijednost elementa bude ili 1 ili 0. Vidljivo je kako klasična binarna logika ima jasno definirane granice te kao takva nije u potpunosti primjenjiva za opisivanje i karakteriziranje funkcioniranja ljudskog mozga. Upravo se iz tog razloga, L. A. Zadeh zalagao za to da neizrazita logika postane jedan od osnovnih načina istraživanja.

### 2.3.1. Usporedba binarne i neizrazite logike

Binarna logika koristi Booleanove operatore za rješavanje problema. Booleanovi operatori su logičko I (AND), logičko ILI (OR) te logičko NE (NOT). Njihovo djelovanje je takvo da, ako se promatra logičko I (AND), postoje tri moguća slučaja:

$$\begin{aligned} \{1 \quad 1\} &\rightarrow 1 \\ \{1 \quad 0\} &\rightarrow 0 \\ \{0 \quad 0\} &\rightarrow 0 \end{aligned} \tag{10}$$

Prvi slučaj je kada su obje ulazne varijable 1, tada je izlazna vrijednost petlje 1. Drugi slučaj je kada su varijable koje ulaze u petlju 0 i 1. Tada će izlazna varijabla biti 0. U trećem slučaju su obje ulazne varijable 0, te je izlazna varijabla također 0. Vidljivo je, dakle, da ukoliko su obje ulazne varijable jednake, izlazna varijabla će biti jednaka ulaznima, a kada su ulazne varijable različite, izlazna varijabla će uvijek biti 0. Kada 1 promatramo kao istinitu tvrdnju, a 0 kao neistinitu, logičko I će kao izlaznu varijablu dati istinitu samo u onome slučaju kada su obje ulazne varijable istinite.

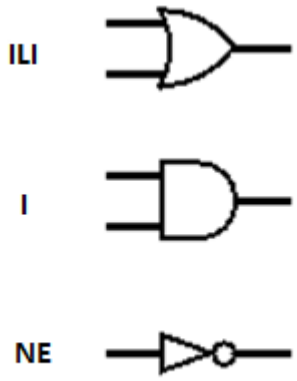
Logičko ILI (OR) će, za razliku od logičkog I (AND), vratiti istinu (1) u slučaju kada je barem jedna ulazna varijabla istinita (1). Dakle:

$$\begin{aligned} \{1 \quad 0\} &\rightarrow 1 \\ \{1 \quad 1\} &\rightarrow 1 \\ \{0 \quad 1\} &\rightarrow 1 \\ \{0 \quad 0\} &\rightarrow 0 \end{aligned} \tag{11}$$

Vidljivo je da će izlazna varijabla biti lažna (0), samo onda kada su obje ulazne varijable lažne (0).

Treći Booleanov operator je logičko NE (NOT). Ovo je linearni operator sa samo jednom ulaznom varijablom i jednom izlaznom varijablom. On pretvara istinitu izjavu u lažnu i obrnuto pri čemu izlazna varijabla postaje nova izjava:

$$\begin{aligned} 1 &\rightarrow 0 \\ 0 &\rightarrow 1 \end{aligned} \tag{12}$$




**Slika 11.** Prikaz logičkih ILI, I i NE operatora.

U neizrazitoj logici postoje operatori koji su ekvivalentni onima u binarnoj logici. Operator *union* je ekvivalent logičkom ILI[20], no naravno njihovo se djelovanje razlikuje. Tako će operator *union* od dvije ulazne varijable kao izlaznu dati onu koja ima veću vrijednost. Dakle, ako je $A$=0,1 i $B$=0,6, te ako je izlazna varijabla $C$=$A$ OR $B$, onda će njen iznos biti 0,6. Matematički zapisano:

$$union = max\left(x_1, x_2, x_3, \ldots..., x_n\right) \tag{13}$$

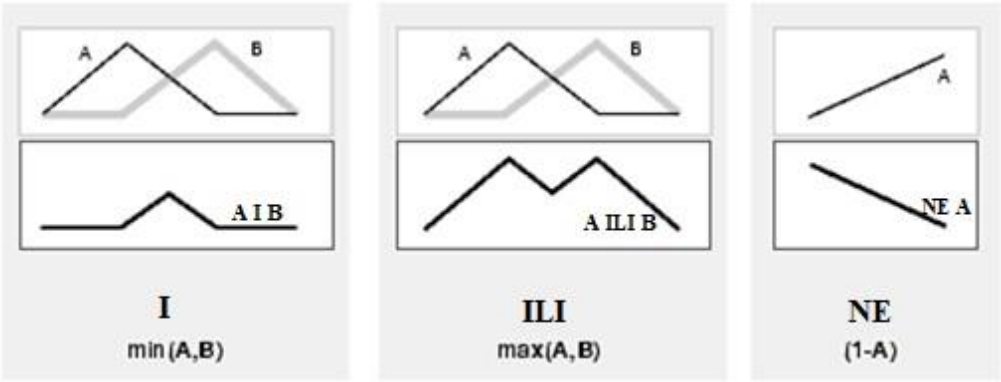
Operator *intersection* ekvivalent je logičkom I[20]; izlazna će varijabla biti jednaka najmanjoj vrijednosti ulaznih varijabli:

$$intersection = min\left(x_1,\ x_2,\ x_3,\ldots..,x_n\right) \tag{14}$$

Na primjeru koji je ranije naveden vrijednost izlazne varijable *C=A* AND *B* biti će 0,1.

Posljednji operator, *complement*, ekvivalent je logičkom NE.[20] On djeluje tako da od broja 1 oduzme vrijednost ulazne varijable:

$$complement = 1 - x \tag{15}$$




**Slika 12.** Prikaz neizrazitih operatora.[20]

### 2.3.2. Neizraziti skupovi

Jedno od obilježja neizrazitih skupova je sposobnost odvajanja logike i neizrazitosti. Klasična binarna logika, odnosno binarni skupovi nisu u mogućnosti izvršavati ove radnje te se sukladno tome prilagođavaju pravila. Teorija klasičnih tj. binarnih skupova se može promatrati kao dio teorije neizrazitih skupova iz razloga što je teorija neizrazitih skupova proširenje, odnosno nadogradnja teorije klasičnih skupova. Razlika između navedene dvije vrste skupova je prikazana na idućem primjeru.

Kao što je ranije navedeno, binarna logika omogućava elementu da pripada ili ne pripada skupu. Pripadnost elementa *x* u skupu *C* se može izraziti funkcijom pripadnosti, $\mu_C(x)$ koja za klasičan (binarni) skup poprima samo dvije vrijednosti:

$$\mu_c\left(x\right) = \begin{cases} 1, & x \in C \\ 0, & x \notin C \end{cases} \tag{16}$$

Neizraziti skupovi omogućuju elementima djelomičnu pripadnost skupovima te u tom slučaju elementi mogu poprimiti vrijednosti između 0 i 1. Pripadnost elementa *x* neizrazitim skupovima *F* se izražava funkcijom $\mu_F(x)$ koja, dakle, za neizraziti skup može poprimati vrijednosti između 0 i 1:

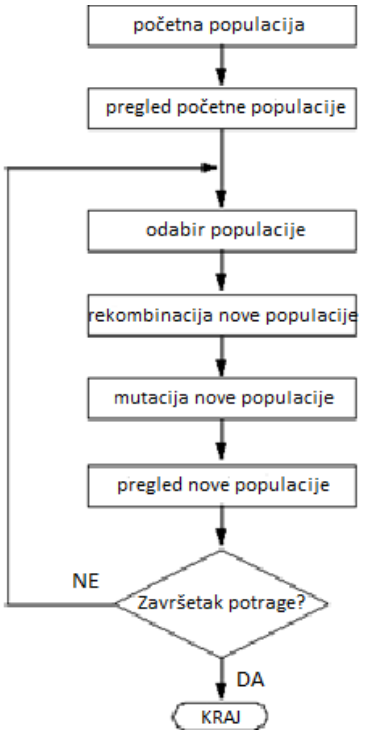
$$\mu_F(x) = \begin{cases} \in \langle 0,1], & x \in F \\ 0, & x \notin F \end{cases} \tag{17}$$

Definicija neizrazitih skupova omogućava uvođenje lingvističkih (neizrazitih) varijabli, što je temelj neizrazite logike. Riječi ili rečenice poprimaju vrijednosti u obliku lingvističkih varijabli. Neke od tih riječi su: visok, nizak, topao, hladan, vjetrovito te je vidljivo kako su te riječi subjektivne, što znači da one donose subjektivnost u definiranje varijabli te je upravo iz tog razloga varijable poželjno predstavljati neizrazitim skupovima.[21]

## 2.4. Genetički algoritam

Pod pojmom *genetički algoritam* (GA) podrazumijevaju se postupci traženja i optimiranja koji pripadaju većoj skupini *evolucijskih algoritama* (EA). Genetički algoritmi se zasnivaju na procesima biološke evolucije. Ovi algoritmi služe se operacijama rekombinacije i mutacije kako bi došli do optimalnog rješenja nekog problema.

Genetički algoritam je zamišljen kao alat koji će skenirati više generacijskih ciklusa te nastojati, putem križanja i mutacija u populaciji, od početne jedinke razviti jedinku sa najboljim karakteristikama prethodnih generacija.[22]




**Slika 13.** Shematski prikaz djelovanja genetičkog algoritma.

Prije nego što započnu procesi traženja i optimiranja, potrebno je odrediti početna rješenja što znači da je potrebno imati mehanizam za postavljanje *početne populacije.* Da bi se odredila sposobnost neke jedinke da preživi u raznim okolnostima, potrebno je uvesti *evaluacijske funkcije* koje definiraju na koji će način teći proces optimiranja. *Mehanizam selekcije* će se odabrati sukladno zahtjevima evaluacijske funkcije te će on birati jedinke sa većom sposobnošću i upućivati ih na *reprodukciju*.

U procesu reprodukcije jedinke će biti podvrgnute raznim transformacijama kako bi od generacije roditelja nastala generacija djece. Navedene transformacije su *križanje* odnosno *rekombinacija* i *mutacija jedinki.* Križanje je jednostavan operator za čiju provedbu je potrebno definirati niz parametara specifičnih za neki promatrani problem. Ti parametri su: duljina niza, veličina populacije te različite vjerojatnosti izbora u procesu selekcije jedinki.[22]

### 2.4.1. Operator selekcije

Operator selekcije se koristi na početku generacijskog ciklusa za odabir jedinki koje će sudjelovati u reprodukciji (često se naziva i *operatorom reprodukcije*). Postoji nekoliko različitih operatora selekcije.

*Selekcija temeljena na slučajnim brojevima*

Selekcija temeljena na slučajnim brojevima provodi se na način da se u svakom selekcijskom koraku biraju dva nasumična cijela broja, $z_1, z_2 \in [1, 2, ...., N]$, iz početnog skupa svih jedinki *P*. Cijeli brojevi označavaju jedinke, a skup *P* označava populaciju. Između dva nasumično odabrana broja, u skup *P'*, koji označava međupopulaciju, uvrstiti će se onaj broj, odnosno jedinka, koja ima veću sposobnost $F(x_i)$, $x_i \in \{z_1, z_2\}$.

*Selekcija temeljena na igri ruleta*.

Prvi korak u selekciji temeljenoj na igri ruleta je izračunavanje sposobnosti $F(x_i)$ svih jedinki u populaciji *P*:

$$F(x_i), \quad i = 1, 2, .., N, \quad F(x_i) > 0 \tag{18}$$

Sumiranjem vrijednosti sposobnosti po svim jedinkama populacije, *N*, dobije se ukupna sposobnost populacije:

$$F_{\Sigma} = \sum_{i=1}^{N} F\left(x_i\right) \tag{19}$$

Diskretna vrijednost selekcije, *p*, se dobije kada se $F(x_i)$ normira prema $F_{\Sigma}$:

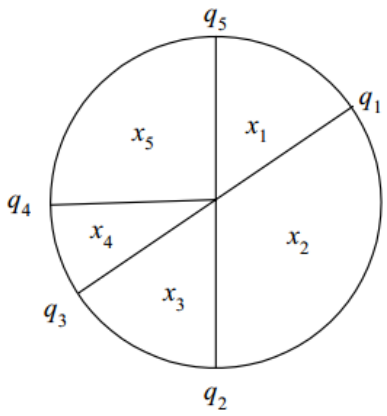
$$p_i = p\left(x_i\right) = \frac{F\left(x_i\right)}{F_{\Sigma}} \tag{20}$$

Promatrajući izraz (20) jasno je da $p_i$ poprima vrijednosti između 0 i 1. Iz *N* vrijednosti gornje jednadžbe dobiju se kumulativne vrijednosti:

$$q_i = \sum_{j=1}^{i} p_j, \; j = 1, 2, ..., N \tag{21}$$

pri čemu se uzima da je $q_0$=0.

Kotač ruleta se zavrti *N* puta te se nakon svakog zaustavljanja očita vrijednost $x_i$ koja pripada sektoru $q_i$. Ovaj selekcijski postupak daje šansu i „osrednjim“ jedinkama, dok slabe uvijek potiskuje.[22]




Slika 14. Prikaz kotača ruleta.

### 2.4.2. Operator križanja

Pomoću operatora selekcije se odabire međupopulacija *P'*. Iz međupopulacije se odabiru potencijalni roditelji za križanje te se svrstavaju u skup *P''*. Izbor roditelja se obavlja u skladu sa vjerojatnošću $p_c$. Za svaku jedinku $x_i$ iz međupopulacije *P'* se generira realni slučajni broj $z \in [0,1]$ te ukoliko je $z<p_c$, jedinka $x_i$ će se podvrgnuti parenju. Očekivani broj parova je $p_c(N/2)$.

Križanje može biti provedeno na jednoj ili više točaka jedinke. Jednotočkasto križanje (engl. *one-point crossover*), sukladno imenu, sadrži samo jednu točku križanja: $N_c=1$. U višetočkastom križanju (engl. *multiple- point crossover*) vrijedi da je $N_c>1$.

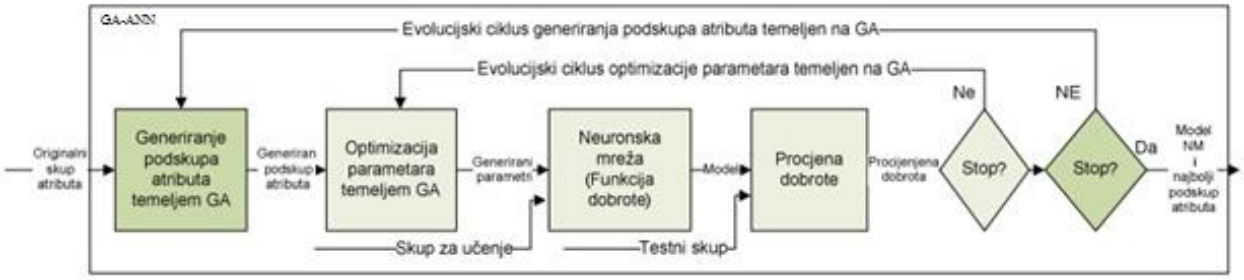
### 2.4.3. Operator mutacije

Operator mutacije osigurava različitost budućih generacija što znači da uvodi nove karakteristike. Jednako kao u biološkoj mutaciji, operator mutacije uzima jedinke te ih „mijenja". Dakle, iz skupa $P''$ koji se sastoji od jedinki koje su izabrane za parenje, biraju se one jedinke koje će se podvrgnuti mutaciji. To bi značilo da se iz skupa $P''$ generira realni slučajni broj $z$, $z \in [0,1]$ iz kojega će se mutacijom dobiti nova jedinka $b_j$ ukoliko je zadovoljen uvjet $z<p_M$, gdje $p_M$ predstavlja vjerojatnost mutacije. Korak će se provesti $N\cdot(m+1)$ puta, pri čemu $m$ označava duljinu niza. Stoga će očekivani broj novih mutiranih jedinki biti $p_M\cdot N\cdot(m+1)$.[22]

## 2.5. Hibridne tehnike umjetne inteligencije

Hibridne tehnike umjetne inteligencije, kao što im i samo ime naznačava, kombiniraju različite AI tehnike. Primjerice GA i ANN ili kombinacija GA i ANFIS sustava (engl. *adaptive neuro-fuzzy inference systems*).[21] Hibridne tehnike bazirane na GA omogućuju podešavanje i prilagodbu klasifikacijskih algoritama podacima i obrnuto. Svrha hibridnih tehnika je preuzeti ono najbolje od kombiniranih tehnika te razviti novi algoritam koji će biti u mogućnosti rješavati teške i kompleksne probleme.[23]

Na slici 15 je shematski prikazan hibridni *GA-ANN* model koji se eksperimentalno pokazao dobrim alatom za istovremeno obavljanje zadataka selekcije i klasifikacije.[23]




**Slika 15.** Dijagram toka *GA-ANN* hibrida.[23]

# 3. Prepoznavanje uzoraka

Prepoznavanje uzoraka karakteristika je ljudskog mozga koja omogućava da sve što čovjek vidi i čuje oko sebe svrsta u određeni kontekst. Na primjer, malom djetetu odrasli pokažu drvo i ukažu mu na to da se radi o drvetu. Do toga trenutka dijete je vidjelo smeđe-zelenu tvorevinu koju nije moglo svrstati u nikakav kontekst. Nakon toga dijete zna da se radi o drvetu te kako ono izgleda. Dakle, kada dijete idući puta vidi drvo, prepoznati će uzorak te smeđe-zelenu tvorevinu svrstati u kontekst. Također, prepoznavanje uzoraka će djetetu, odnosno ljudskome mozgu, omogućiti da prepozna drvo čak i onda kada je bez lišća ili djelomično zaklonjeno. Prepoznavanje uzoraka omogućava ljudskome mozgu da podatke, odnosno uzorke prikupljene iz okoline, bilo vidom, sluhom, njuhom ili opipom, poveže sa već prikupljenim podacima te ih razumije.

Počeci teorije prepoznavanja uzoraka bila su statistička istraživanja velikih grupa podataka. Danas je prepoznavanje uzoraka sastavni dio svakog umjetno stvorenog inteligentnog sustava. Za realizaciju prepoznavanja uzoraka u inteligentnim sustavima koriste se razne matematičke tehnike koje se i dan danas razvijaju. Svrha prepoznavanja uzoraka je analizirati situacije iz stvarnoga svijeta te stvoriti opis određene situacije koji će pomoći rješavanju nekog problema. Informacije iz stvarnoga svijeta primaju se preko senzora te zatim klasificiraju i opisuju pomoću programa prepoznavanja uzoraka.

Postoji veliki broj definicija prepoznavanja uzoraka od kojih bih htjela izdvojiti dvije. Satosi Watanabe, japanski teorijski fizičar je dao definiciju uzorka koja kaže da je uzorak „suprotno od kaosa, entitet koji je nejasno opisan, a kojemu se može dati ime“.[24]Anil K. Jain, indijsko-američki inženjer računarstva, prepoznavanje uzoraka je definirao kao generalizirani pojam koji opisuje široki spektar problema kao što su prepoznavanje, opisivanje, klasifikacija te grupiranje uzoraka.[24]

## 3.1. Podjela metoda prepoznavanja uzoraka

Prepoznavanje uzoraka se dijeli obzirom na metode koje se koriste pri analizi ulaznih podataka. Prva metoda koja se koristi je *eksploracijska analiza podataka* (engl. *exploratory data analysis*) koja se dijeli na *analizu glavnih komponenata* (engl. *principal components analysis*) i *faktorska analiza* (engl. *factor analysis*). Preostale dvije metode su *nenadgledano prepoznavanje uzoraka* (engl. *unsupervised pattern recognition*) i *nadgledano prepoznavanje uzoraka* (engl. *supervised pattern recognition*).

### 3.1.1. Eksploracijska analiza podataka

*Analiza glavnih komponenata*

U literaturi, prvi je metodu analize glavnih komponenata (PCA) spomenuo fizičar Cauchy davne 1829. godine, a njeno prvo povezivanje s kemijom zabilježeno je 1878. godine. Međutim, kao početak razvoja ove metode koja koristi dvije ili više varijabli (engl. *multivariable*) uzimaju se 30-e i 40-e godine 20. stoljeća, a prvi veliki korak u primjeni u kemiji dogodio se 30-ak godina kasnije.[25] Tradicionalna istraživanja u kemiji su koristila samo jednu varijablu, što znači da se na jednome uzorku moglo provesti samo jedno mjerenje. PCA koristi više varijabli, omogućujući da se na jednom uzorku provede više mjerenja te time osigurava da se bitne informacije o uzorku ne gube. Poznavanjem više informacija o uzorku lakše je i točnije kategoriziranje odnosno prepoznavanje uzorka.

Prvi korak PCA metode je provesti eksperiment te prikupiti podatke. Promotrimo hipotetski slučaj u kojem je prikupljeno *n* setova podataka u dvije dimenzije: {*x*, *y*}. Prvi korak u obradi je dobiti skup podataka kojima je srednja vrijednost nula.[25] Skup se dobije tako da se svakome podatku iz dimenziji *x* oduzme srednja vrijednost te dimenzije, a isti postupak provodi se i za dimenziju *y*. Srednja vrijednost se izračunava prema jednadžbi:

$$\bar{x} = \frac{\sum_{i}^{n} x_i}{n} \tag{22}$$

$$\bar{y} = \frac{\sum_{i}^{n} y_i}{n} \tag{23}$$

Provedbom navedenog dobiva se skup podataka prikazan u tablici 1.

**Tablica 1.** Prikaz skupa podataka čija je srednja vrijednost nula.

| $x$ | $y$ |
|---|---|
| $x_1 - \bar{x}$ | $y_1 - \bar{y}$ |
| $x_2 - \bar{x}$ | $y_2 - \bar{y}$ |
| ... | ... |
| $x_n - \bar{x}$ | $y_n - \bar{y}$ |

Idući korak je računanje *kovarijance* koja označava koliko su dvije varijable slične. Kovarijanca se uvijek računa između samo dvije dimenzije te formula za njeno računanje glasi:

$$\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{(n-1)} \tag{24}$$

Kovarijanca postaje više pozitivnom za svaki par vrijednosti koji se razlikuje od njihovih srednjih vrijednosti u istom smjeru, te postaje više negativna za svaki par vrijednosti koji se razlikuje od njihovih srednjih vrijednosti u suprotnim smjerovima. Što je brojčana vrijednost kovarijance manja, to su uspoređivane varijable manje slične, a kada je jednaka 0 možemo kazati da su dvije varijable nezavisne.

Između dimenzija se može izračunati matrica kovarijance čije su dimenzije jednake broju uzoraka.[25] Dobivena matrica kovarijance će biti simetrična po dijagonali te će matrica kovarijance za dvije dimenzije biti:

$$C = \begin{pmatrix} \operatorname{cov}(x, x) & \operatorname{cov}(x, y) \\ \operatorname{cov}(y, x) & \operatorname{cov}(y, y) \end{pmatrix} \tag{25}$$

Nakon što je izračunata matrica kovarijance, računaju se svojstveni vektori (engl. eigenvector) i njihove svojstvene vrijednosti (engl. *eigenvalue*) za matricu kovarijance. Svojstveni vektori su vektori koji ne mijenjaju smjer nakon što ih se pomnoži sa matricom. Dakle, ako se matrica A pomnoži sa vektorom $x$, dobiti će se vektor A$x$ čija je vrijednost broj $\lambda$ pomnožen sa vektorom $x$:

$$\mathrm{A}x = \lambda x \tag{26}$$

gdje je broj $\lambda$ svojstvena vrijednost matrice A. Svojstvena vrijednost govori o tome da li se vektor $x$ promijenio nakon množenja matricom A. Vrijednost $\lambda$ može iznositi 0, 1, -1, ½, 2.[26] Nakon izračuna svojstvenih vektora i svojstvenih vrijednosti, određuje se koji je podatak glavna komponenta. To će biti svojstveni vektor sa najvećom svojstvenom vrijednošću. Nakon toga, sve svojstvene vrijednosti slažu se po veličini ne bi li se odredila njihova važnost, tj. u ovom koraku se odlučuje ide li se u daljnju analizu sa svim podacima ili će se neki podaci izbaciti. Nadalje, formira se *vektor varijabli* (engl. *feature vector*) koji se sastoji od matrice zadržanih svojstvenih vektora. Posljednji korak dobivanja novoga skupa podataka nakon provedbe PCA metode je množenje vektora varijabli sa skupom podataka koji je prikazan u tablici 1.

*Faktorska analiza*

Faktorska analiza[27] je metoda kojom se velikom skupu varijabli pridružuje manji skup faktora pri čemu se svakoj varijabli pridodaje jedan ili više faktora koji ju opisuje. Postoje dvije vrste faktora: *zajednički faktori* ($F_1, F_2, ..., F_m$) koji se pridodaju većem broju varijabli i *specifični faktori* ($U_1, U_2, ..., U_n$) koji su jedinstveni za svaku varijablu. Na taj način je reduciran veliki broj početnih varijabli te su one svrstane u grupe ovisno o opisu koji su im pružili pridodani faktori.

Prikažimo skup varijabli $\{x_1, x_2, ..., x_n\}$ kao linearnu funkciju zajedničkih $\{F_1, F_2, \ldots, F_m\}$ i specifičnog faktora $\{U_1, U_2, ..., U_n\}$:

$$x_1 = a_{11}F_1 + a_{12}F_2 + \ \ldots \ + a_{1m}F_m + U_1 \tag{27}$$

$$x_2 = a_{21}F_1 + a_{22}F_2 + \ \ldots \ + a_{2m}F_m + U_2 \tag{28}$$

$$\ldots$$

$$x_n = a_{n1}F_1 + a_{n2}F_2 + \ \ldots \ + a_{nm}F_m + U_n \tag{29}$$

Koeficijenti $a$ u jednadžbama (27)–(29) označavaju „težinu“ pojedinog faktora, odnosno njegovu važnost.

Kada faktori nisu korelirani suma kvadrata težina varijable $x_i$:

$$\sum_{j=1}^{m} a_{ij}^2 \tag{30}$$

koju ovdje nazivam *združenom vrijednošću* (engl. *communality*) pokazuje udio varijance varijable $x_i$ koja se može objasniti zajedničkim faktorima. Što je združena vrijednost neke

varijable veća to je faktorska analiza uspješnija. Slično tome, zbroj kvadrata koeficijenata za općeniti faktor $F_i$, dakle:

$$\sum_{j=1}^{n} a_{ji}^2 \tag{31}$$

pokazuje udio varijance svih varijabli koji se može objasniti tim promatranim faktorom.

### 3.1.2. Nenadgledano prepoznavanje uzoraka

Postupak nenadgledanog prepoznavanja uzoraka naziva se klaster analiza (engl. *cluster analysis*). Postupak je primarno korišten u biologiji kako bi se pronašla veza između različitih životinjskih vrsta. Ovakav način povezivanja uzoraka koristan je i u kemiji. Na primjer ukoliko je potrebno odrediti da li su dvije sorte vina povezane podrijetlom i/ili okusom. Određivanje ovih informacija provest će se upravo metodom prepoznavanja uzoraka kemijskih karakteristika loze.

Klaster analiza se, dakle, koristi da bi se različiti podaci grupirali korištenjem raznih metoda. Promotrimo primjer (tablica 2) u kojem su prikupljeni podaci šest analitičkih uzoraka od čega je na svakom uzorku provedeno mjerenje pet parametara (A-E).

**Tablica 2**. Hipotetski slučaj prikupljenog seta mjerenja parametara A-E na 6 analitičkih uzoraka (1-6).

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | 0.9 | 0.5 | 0.2 | 1.6 | 1.5 |
| 2 | 0.3 | 0.2 | 0.6 | 0.7 | 0.1 |
| 3 | 0.7 | 0.2 | 0.1 | 0.9 | 0.1 |
| 4 | 0.1 | 0.4 | 1.1 | 1.3 | 0.2 |
| 5 | 1.0 | 0.7 | 2.0 | 2.2 | 0.4 |
| 6 | 0.3 | 0.1 | 0.3 | 0.5 | 0.1 |

Postoji velik broj načina na koji se podaci mogu međusobno usporediti. Tri najčešće korištena su:

1. izračun koeficijenta korelacije,
2. izračun Euklidove udaljenost,
3. izračun manhattanske udaljenosti.

*Izračun koeficijenta korelacije*

Koeficijent korelacije[28] (engl. *correlation coefficient*), *r*, među skupovima podataka, *x* i *y*, je broj koji pokazuje koliko su promatrani skupovi slični.

$$r(x,y)=\frac{\mathrm{cov}(x,y)}{s_x \cdot s_y} \tag{32}$$

Pri tome su $s_x$ i $s_y$ standardna odstupanja unutar skupova *x* i *y*. Što je apsolutna vrijednost koeficijenta korelacije bliže jedinici, to su skupovi sličniji; za koeficijent korelacije jednak jedinici uspoređivani skupovi su potpuno jednaki. Shodno tome, što je koeficijent korelacije bliže 0 uzorci su različitiji.[25]

**Tablica 3.** Prikaz korelacijskih koeficijenata za analitičke uzorke iz tablice 2.

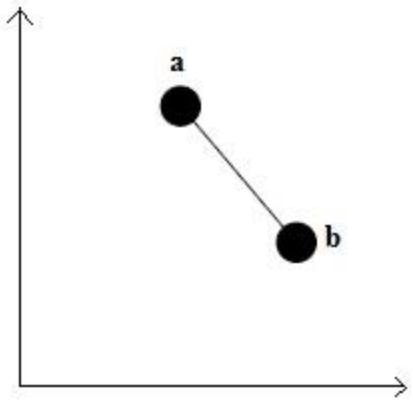
| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1.000 | | | | | |
| 2 | −0.041 | 1.000 | | | | |
| 3 | 0.503 | 0.490 | 1.000 | | | |
| 4 | −0.018 | 0.925 | 0.257 | 1.000 | | |
| 5 | −0.078 | 0.999 | 0.452 | 0.927 | 1.000 | |
| 6 | 0.264 | 0.900 | 0.799 | 0.724 | 0.883 | 1.000 |

*Euklidova udaljenost*

Euklidova udaljenost (engl. *Euclidean distance*) je direktna udaljenost između dva uzorka. Ako se uzme da su ti uzorci **a** i **b** tada će udaljenost biti definirana jednadžbom:

$$d_{\mathbf{ab}}=\sqrt{\sum_j \left(x_{\mathbf{a}j}-x_{\mathbf{b}j}\right)^2} \tag{33}$$

gdje *j* predstavlja broj mjerenja, a $x_{\mathbf{a}j}$ je *j*-to mjerenje na uzorku **a**. Očitanjem iz tablice 2, $x_{53}$ iznosi 2,0 (treće mjerenje na petom uzorku). Što je manja vrijednost udaljenosti između dva uzorka, to su oni sličniji. Euklidova udaljenost nema ograničenja što se tiče njenih brojčanih vrijednosti i kao što se vidi iz jednadžbe (33) može imati samo pozitivne vrijednosti.[25]



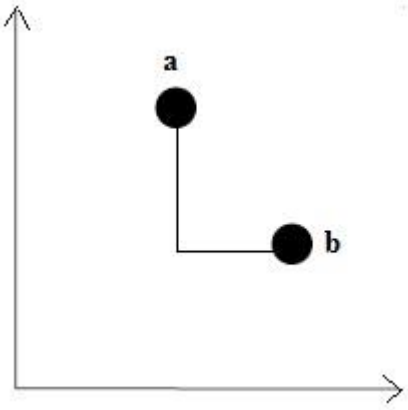

**Slika 16.** Grafički prikaz Euklidove udaljenosti između uzoraka **a** i **b**.

**Tablica 4.** Prikaz Euklidovih udaljenosti za podatke iz tablice 3.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0.000 | | | | | |
| 2 | 1.838 | 0.000 | | | | |
| 3 | 1.609 | 0.671 | 0.000 | | | |
| 4 | 1.800 | 0.837 | 1.253 | 0.000 | | |
| 5 | 2.205 | 2.245 | 2.394 | 1.600 | 0.000 | |
| 6 | 1.924 | 0.374 | 0.608 | 1.192 | 2.592 | 0.000 |

*Manhattanska udaljenost*

Manhattanska udaljenost (engl. *Manhattan distance*, slika 17) je jedan od načina izražavanja udaljenosti između dvaju uzoraka.




**Slika 17.** Grafički prikaz manhattanske udaljenosti između uzoraka **a** i **b**.

Manhattanska udaljenost između uzoraka **a** i **b** definirana je jednadžbom:

$$d_{\mathbf{ab}} = \sum_j \left| x_{\mathbf{a}j} - x_{\mathbf{b}j} \right| \tag{34}$$

Brojčana vrijednost manhattanske udaljenosti će uvijek biti veća od Euklidove, a samo u rijetkim slučajevima identična njoj. Jednako kao i kod Euklidove udaljenosti, što je udaljenost manja to su uzorci sličniji. Razlika između ove dvije udaljenosti najbolje je vidljiva na grafičkim prikazima na slikama 16 i 17.[25]

**Tablica 5.** Prikaz manhattanske udaljenosti za podatke iz tablice 3.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 0 | | | | | |
| 2 | 3.6 | 0 | | | | |
| 3 | 2.7 | 1.1 | 0 | | | |
| 4 | 3.4 | 1.6 | 2.3 | 0 | | |
| 5 | 3.8 | 4.4 | 4.3 | 3.2 | 0 | |
| 6 | 3.6 | 0.6 | 1.1 | 2.2 | 5 | 0 |

Nakon što je odabran način izračuna udaljenosti između uzoraka, te su sve udaljenosti određene, potrebno je povezati uzorke međusobno. Najčešće korišteni pristup je uzimanje dva uzorka koja su najsličnija te njihovo grupiranje. Na primjeru podataka iz tablice 2., korištenjem koeficijenta korelacije za računanje udaljenosti između uzoraka, uzimaju se uzorci 2 i 5 zbog toga što su oni najsličniji, što je pak vidljivo iz tablice 3. Formiraju se grupe podataka od kojih će jedna grupa sadržavati dva najsličnija uzorka (2 i 5) a ostale četiri grupe će sadržavati ostale uzorke (1, 3, 4 i 6). Nakon što su uzorci grupirani, potrebno je povezati nove grupe uzoraka. Postupak povezivanja grupa se sastoji od toga da se svaki pojedini uzorak povezuje sa grupom koja sadrži uzorke koji su najsličniji. Postoji više načina na koji to može biti odrađeno, a neki od uobičajenih su:

1. metoda najbližeg susjeda,
2. metoda najudaljenijeg susjeda,
3. metoda prosječnog povezivanja.

*Metoda najbližeg susjeda*

Metoda najbližeg susjeda (engl. *nearest neighbour method*) uspoređuje svaki preostali originalni uzorak sa oba elementa u novoj grupi te ona udaljenost čija je brojčana vrijednost veća opisuje novu vezu između originalnog uzorka i grupe. Iz tablice 3. je vidljivo da udaljenost uzorka 6 od uzorka 2 iznosi 0,900, a od uzorka 5 iznosi 0,883. Stoga, nova udaljenost uzorka 6 od grupe će biti 0,900.

*Metoda najudaljenijeg susjeda*

Metoda najudaljenijeg susjeda (engl. *furthest neighbour method*) je suprotna metodi najbližeg susjeda. U ovom slučaju će nova udaljenost uzorka 6 od grupe biti 0,883 zbog toga što se nakon usporedbe uzima ona udaljenost koja je manja. Važno je napomenuti da se obavezno prije provedbe ove metode uzorci grupiraju na ranije navedeni način, dva najbliža uzorka u jednu grupu, a ostali u druge grupe. Nakon toga se provodi metoda najudaljenijeg susjeda kako je opisano.[25]

*Metoda prosječnog povezivanja*

Kao nova udaljenost u metodi prosječnog povezivanja (engl. *average linkage*) uzima se srednja vrijednost udaljenosti samostalnog uzorka od svakog uzorka u grupi. U ranije navedenom primjeru to će biti 0,892. Izračun nove udaljenosti se može provesti na dva načina, ovisno o tome koliko uzoraka se nalazi u kojoj grupi. U navedenom primjeru, obje grupe sadrže isti broj uzoraka (prva grupa sadrži udaljenosti između uzorka 2 i 6 te uzoraka 5 i 6, a druga grupa sadrži udaljenosti uzoraka 6 i 2 te 6 i 5). Tada se nova udaljenost računa pomoću jednadžbe:

$$s_{AB} = \frac{\left(s_A + s_B\right)}{2} \tag{35}$$

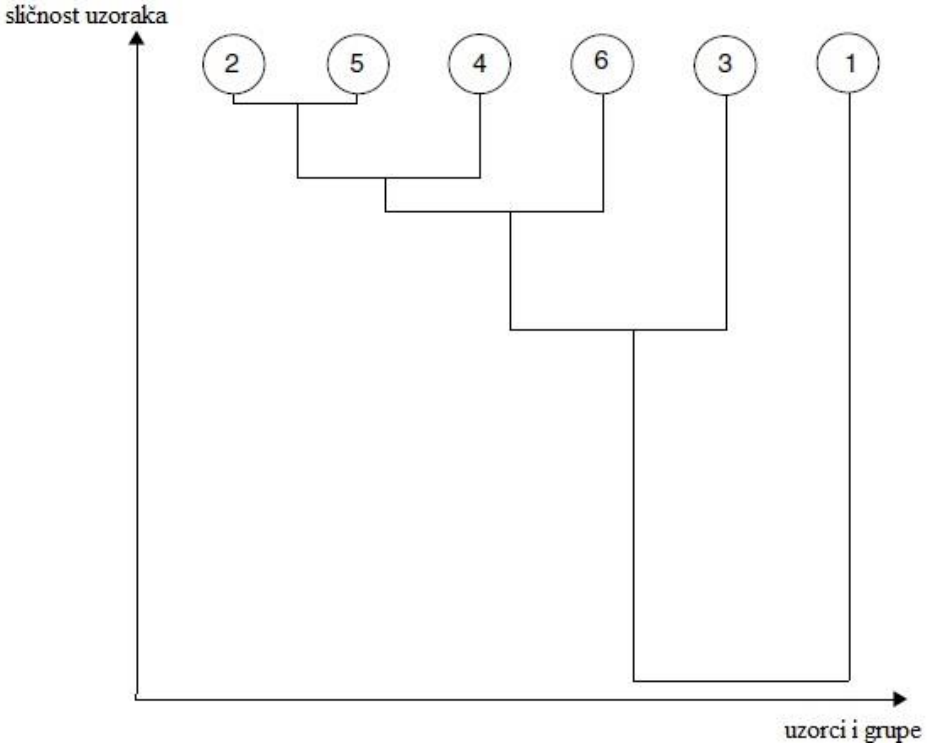
gdje A i B predstavljaju grupe, a *s* udaljenost.

U slučaju kada dvije uspoređivane grupe sadrže različiti broj uzoraka, nova udaljenost se računa pomoću jednadžbe:

$$s_{AB} = \frac{\left(N_A s_A + N_B s_B\right)}{\left(N_A + N_B\right)} \tag{36}$$

gdje $N_A$ i $N_B$ predstavljaju broj uzoraka u grupama A i B, a *s* udaljenost.[25]

Postupci povezivanja uzoraka unutar grupa se ponavljaju sve dok se svi uzorci ne povežu u jednu veliku grupu podataka. U svakom koraku se prvo definira najsličniji par uzoraka koji se zatim uspoređuje sa ostalima nekom od ranije navedenih metoda. Uobičajeno se rezultat analize i povezivanja uzoraka prikazuje *dendrogramom*. Vrijednosti na ordinati pokazuju koliko su uzorci slični. Na dendrogramu prikazanom na slici 18, koji je dobiven primjenom metode koeficijenta korelacije i metode najbližeg susjeda, vidljivo je da je uzorak 1 najmanje sličan ostalima te da su, kako je već i navedeno, uzorci 2 i 5 najsličniji. Preporuča se da se na istim uzorcima provede nekoliko različitih metoda klaster analize. Ukoliko se krajnji rezultati dobiveni različitim metodama ne razlikuju puno, tada se rezultat bilo koje metode vodi kao valjan.




**Slika 18.** Dendogram dobiven obradom podataka iz tablice 2.[25]

### 3.1.3. Nadgledano prepoznavanje uzoraka

Metoda prepoznavanja uzoraka koja koristi nadzirano učenje naziva se i klasifikacija. Klasifikacija je vrlo važna metoda u kemiji: bilo organskoj, anorganskoj ili analitičkoj. Klasifikacija podrazumijeva grupiranje analiziranih podataka prema kriteriju koji je modelu ranije zadan, tj. prema kriteriju koji je model ranije naučio. Razlika između ove metode i klaster

analize je prvenstveno u tome što ovdje nije bitno kakva je međusobna povezanost uzoraka, nego samo krajnje kategorije u koje će uzorci biti smješteni.

Kao primjer koristi se skup snimljenih spektara soka od naranče. Snimljeno je 10 spektara čije su naranče podrijetlom iz Brazila, 10 iz Španjolske te je posljednjih 10 spektara snimljeno na sokovima koji su proizvedeni od zastarjelih naranči.[25] Cilj je dobiti matematički model koji će predvidjeti kojoj skupini naranči pripada snimljeni spektar. Na početku je, dakako, potrebno povezati mjerenja na zadanim uzorcima sa poznatim skupovima. Kreirani matematički model se sastoji od seta podataka koji se koriste za učenje (engl. *training set*) te on sadrži informacije kako su podaci klasificirani. Ovaj skup podataka neće imati 100%-tnu točnost, što znači da neće sve spektre snimljene na narančama iz Brazila svrstati u grupu „Brazil" nego će neke od tih 10 spektara klasificirati kao da su snimljeni na narančama iz Španjolske ili na zastarjelim narančama. Parametar koji govori koliko je informacija točno klasificirano je *CC* postotak. Što je vrijednost *CC* postotka veća, to je veća preciznost metode. Dobiveni set podataka, tj. metodu, treba provjeriti. Provjera metode se vrši validacijom. Dvije vrste validacije se koriste za provjeru metode: *standardna* i *unakrsna validacija.*

*Standardna validacija*

Da bi se provela validacija, potrebno je kreirati set podataka za testiranje prepoznatosti uzorka/klasifikacije (engl. *test set*) koja se sastoji od nasumično odabranih podataka. Na ranije navedenom primjeru to bi značilo da se određeni broj od ukupnog broja snimljenih spektara upotrijebi za treniranje mreže, a preostalih dio poslužiti će za testiranje. Na testnoj grupi podataka se provodi jednaka analiza odnosno klasifikacija podataka kao i na skupu za treniranje. Vrijednost *CC* postotka će za testnu skupinu uvijek biti manja, što znači da će manje ulaznih podataka biti ispravno klasificirano. Ukoliko su dobivene vrijednosti *CC* postotka setova za treniranje i testiranje dovoljno slične, metoda se smatra valjanom te se može primjenjivati na raznim uzorcima. Ukoliko su vrijednosti značajno različite, metoda nije valjana i potrebno je provesti novi trening.

*Unakrsna validacija*

Metoda unakrsne validacije se provodi na način da se iz početne grupe podataka oduzima po jedan uzorak koji služi za testiranje, a ostatak grupe se koristi za treniranje. Postupak se

ponavlja dok svaki od uzoraka iz početne grupe ne bude iskorišten za testiranje. Testiranjem se dobiva *CC* postotak koji u biti govori o preciznosti metode. *CC* postotak metode unakrsne validacije će biti manji od *CC* postotka početne grupe, no ukoliko su te dvije vrijednosti dovoljno slične metoda se smatra valjanom i može primjenjivati na raznim uzorcima. Ukoliko su vrijednosti značajno različite, metoda nije valjana te ju je potrebno korigirati.

U slučajevima kada metoda nije zadovoljavajuća, provode se postupci koji na kraju dovode do veće preciznosti metode. Preciznost metode je moguće povećati korištenjem drugačijeg algoritma ili modifikacijom postojeće metode. Postojeća metoda može biti modificirana odabirom samo važnijih podataka što bi značilo da se, na ranije navedenom primjeru spektara, ne koriste sve valne duljine tj. cijeli spektar nego samo one relevantnije valne duljine čime se grupa podataka smanjuje ali se i njena točnost povećava. Naposljetku, ukoliko navedeni postupci nisu povećali preciznost metode, potrebno je provjeriti da li je primjenjivana analitička metoda pogodna za analizu danih podataka.[25]

Postoji više različitih vrsta klasifikacije. Ovdje ću se dotaknuti dvije: *jednovarijantne* i *multivarijantne klasifikacije*.

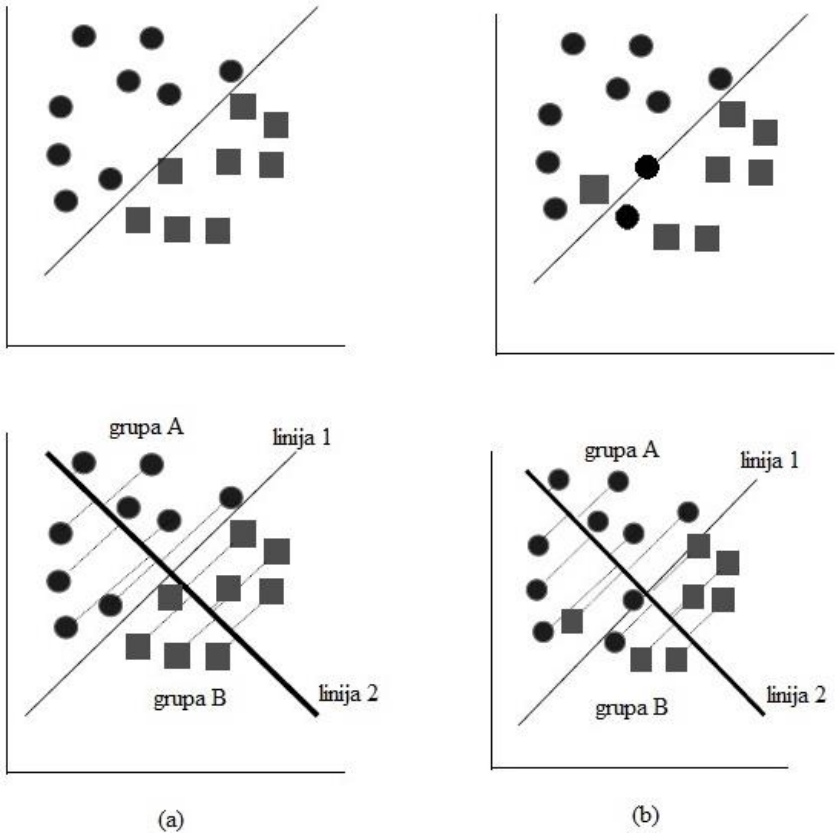
*Jednovarijantna klasifikacija*

Jednovarijantna klasifikacija (engl. *univariate clasiffication*) je najjednostavniji oblik klasificiranja dobivenih mjerenja. Ova metoda se sastoji od samo jednog mjerenja ili jedne varijable koja služi da bi se objekti raspodijelili u grupe.

*Multivarijantna klasifikacija*

Multivarijantna klasifikacija (engl. *multivariate clasiffication*) koristi dvije ili više varijabli ili mjerenja kako bi analizirane uzorke smjestila u odgovarajuće grupe. Dvije ili više varijabli je potrebno iz razloga što samo jedna zasebno ne može u potpunosti opisati te klasificirati analizirane podatke. Dobivena mjerenja te njihove grupe se mogu predočiti grafički te kao što je vidljivo na slici 19, postoje dva slučaja. U slučaju (a), grupirani podaci su jasno odvojeni. Linijama 1 i 2 se prostor dijeli na četiri dijela nakon čega se zaključuje da sva mjerenja koja se nalaze iznad linije 1 pripadaju u grupu A, dok sva mjerenja ispod linije 1 pripadaju u grupu B. Slučaj (b) prikazuje dvije grupe koje nisu jasno podijeljene. U ovome slučaju podjela

na četiri dijela omogućuje točnije klasificiranje mjerenja u grupe. U slučaju kada postoji veliki broj varijabli odnosno mjerenja te kada krajnji odgovor metode nije „da“ ili „ne“, ovakva mala odstupanja neće naškoditi krajnjem rezultatu. Na primjer, ako se klasificiraju esteri i ketoni, u uzorku se mogu naći spojevi koji nisu esteri ni ketoni, oni će dakle biti pogrešno grupirani no neće uvelike naštetiti krajnjem rezultatu. No, u slučaju kada krajnji rezultat mora pokazati da li, na primjer, u uzorku postoji neka otrovna tvar, u dobivenim grupama podataka ne smije biti krivo klasificiranih jedinki. Ako se to dogodi, to je znak da se trebaju provesti detaljnije analize s manjim količinama početnih uzoraka.[25]




**Slika 19.** Grafički prikaz multivarijantne klasifikacije.

# 4. Analiza analitičkih podataka

## 4.1. Određivanje odstupajućih podataka

Svaka analiza počinje skupom prikupljenih podataka. Prikupljeni skup se sastoji od velikog broja podataka (informacija) od kojih su neki ispravni, a neki pak odstupaju od očekivanog. Za te odstupajuće podatke često se u hrvatskom koristi izvorni engleski termin *outlier*. Očekivalo bi se da su ti podaci netočni i nepoželjni te ih je potrebno ukloniti iz analiziranog skupa, no to nije uvijek točno. Odstupajuće podatke je potrebno detektirati te provedbom raznih testova odlučiti da li će se oni isključiti iz daljnje analize ili ne. Za analizu odstupajućih podataka postoji veliki broj metoda, a u analitičkoj kemiji među zastupljenijima su *Dixonov Q-test* i *Grubbsov test.*

### 4.1.1. Dixonov *Q*-test

Dixonov *Q*-test je statistički test koji se primjenjuje u situacijama kada na skupu podataka želimo provjeriti da li je samo jedan podatak odstupajući. Ovo je najjednostavniji test koji se ne može primijeniti na skupu podataka u kojemu su prva dva ili zadnja dva podatka u rastućem nizu jednaki. Početak provedbe Q – testa je slaganje podataka u rastući niz $x_1$, $x_2$, …, $x_n$ gdje je $x_1$ najmanji, a $x_n$ najveći podatak u skupu. Nakon toga se određuju $Q$ vrijednosti prema izrazu:

$$Q = \begin{cases} \dfrac{|x_1 - x_2|}{x_n - x_1}; & \text{za minimalnu } x \text{ vrijednost} \\ \dfrac{|x_n - x_{n-1}|}{x_n - x_1}; & \text{za maksimalnu } x \text{ vrijednost} \end{cases} \qquad (37)$$

Izračunate $Q$ vrijednosti uspoređuju se s kritičnom vrijednosti, $Q_{krit}$[29], koja se može očitati u statističkim tablicama (tablica 6) uz odgovarajuću pouzdanost, odnosno rizik $\alpha$.[30] .Ukoliko je $Q > Q_{krit}$ radi se o odstupajućem podatku koji se izuzima iz daljnje analize.

**Tablica 6.** Kritične vrijednosti za Dixonov test.

| N \ α | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 | 0,005 |
|---|---|---|---|---|---|---|---|
| 3 | 0,684 | 0,781 | 0,886 | 0,941 | 0,976 | 0,988 | 0,994 |
| 4 | 0,471 | 0,560 | 0,679 | 0,765 | 0,846 | 0,889 | 0,926 |
| 5 | 0,373 | 0,451 | 0,557 | 0,642 | 0,729 | 0,780 | 0,821 |
| 6 | 0,318 | 0,386 | 0,482 | 0,560 | 0,644 | 0,689 | 0,740 |
| 7 | 0,281 | 0,344 | 0,434 | 0,507 | 0,596 | 0,637 | 0,680 |
| 8 | 0,318 | 0,385 | 0,479 | 0,554 | 0,631 | 0,683 | 0,725 |
| 9 | 0,288 | 0,352 | 0,441 | 0,512 | 0,587 | 0,635 | 0,677 |
| 10 | 0,265 | 0,325 | 0,409 | 0,477 | 0,551 | 0,597 | 0,639 |
| 11 | 0,391 | 0,442 | 0,571 | 0,576 | 0,638 | 0,679 | 0,713 |
| 12 | 0,370 | 0,419 | 0,490 | 0,546 | 0,605 | 0,642 | 0,675 |
| 13 | 0,351 | 0,399 | 0,467 | 0,521 | 0,578 | 0,615 | 0,649 |
| 14 | 0,370 | 0,421 | 0,429 | 0,546 | 0,602 | 0,641 | 0,674 |
| 15 | 0,353 | 0,402 | 0,472 | 0,525 | 0,579 | 0,616 | 0,647 |
| 16 | 0,338 | 0,386 | 0,454 | 0,507 | 0,559 | 0,595 | 0,624 |
| 17 | 0,325 | 0,373 | 0,438 | 0,490 | 0,542 | 0,577 | 0,605 |
| 18 | 0,314 | 0,361 | 0,424 | 0,475 | 0,527 | 0,561 | 0,589 |
| 19 | 0,304 | 0,350 | 0,412 | 0,462 | 0,514 | 0,547 | 0,575 |
| 20 | 0,295 | 0,340 | 0,401 | 0,450 | 0,502 | 0,535 | 0,562 |
| 21 | 0,287 | 0,331 | 0,391 | 0,440 | 0,491 | 0,524 | 0,551 |
| 22 | 0,280 | 0,323 | 0,382 | 0,430 | 0,481 | 0,514 | 0,541 |
| 23 | 0,274 | 0,316 | 0,374 | 0,421 | 0,472 | 0,505 | 0,532 |
| 24 | 0,268 | 0,310 | 0,367 | 0,413 | 0,464 | 0,497 | 0,524 |
| 25 | 0,262 | 0,304 | 0,360 | 0,406 | 0,457 | 0,489 | 0,516 |

### 4.1.2. Grubbsov test

Određivanje odstupajućih podatka korištenjem Grubbsovog testa zahtjeva izračunavanje srednje vrijednosti podataka, $\bar{x}$, te njihova standardnog odstupanja, $s$:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{38}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} \tag{39}$$

Pomoću vrijednosti dobivenih iz izraza (38)i (39) računa se Grubbsov parametar, $G$, pomoću kojeg se određuje da li je podatak $x_i$ odstupajući ili ne:

$$G = \frac{\left|x_i - \bar{x}\right|}{s} \tag{40}$$

Vrijednosti Grubbsova parametra uspoređuje se potom s kritičnom vrijednošću, $G_{krit}$. Ukoliko je:

1. $G \leq G_{krit}$ uz $\alpha$=0,05; s 95 %-tnom vjerojatnošću tvrdimo da testirani podatak nije odstupajuć te stoga ulazi u daljnju analizu,
2. $G > G_{krit}$ uz $\alpha$=0,05 i $G \leq G_{krit}$ uz $\alpha$=0,01; podatak može i ne mora biti odstupajuć, analitičar odlučuje da li koristi podatak u daljnjoj analizi ili ne,
3. $G > G_{krit}$ uz $\alpha$=0,05; s 95 %-tnom vjerojatnošću tvrdimo da testirani podatak odstupa od ostalih podataka te stoga ne ulazi u daljnju analizu.[29]

**Tablica 7.** Kritične vrijednosti za Grubbsov test.

| N \ α | 0,05 | 0,025 | 0,01 | N \ α | 0,05 | 0,025 | 0,01 |
|---|---|---|---|---|---|---|---|
| 3 | 1,15 | 1,15 | 1,15 | 15 | 2,41 | 2,55 | 2,71 |
| 4 | 1,46 | 1,48 | 1,49 | 16 | 2,44 | 2,59 | 2,75 |
| 5 | 1,67 | 1,71 | 1,75 | 17 | 2,47 | 2,62 | 2,79 |
| 6 | 1,82 | 1,89 | 1,94 | 18 | 2,50 | 2,65 | 2,82 |
| 7 | 1,94 | 2,02 | 2,10 | 19 | 2,53 | 2,68 | 2,85 |
| 8 | 2,03 | 2,13 | 2,22 | 20 | 2,56 | 2,71 | 2,88 |
| 9 | 2,11 | 2,21 | 2,32 | 21 | 2,58 | 2,73 | 2,91 |
| 10 | 2,18 | 2,29 | 2,41 | 22 | 2,60 | 2,76 | 2,94 |
| 11 | 2,23 | 2,36 | 2,48 | 23 | 2,62 | 2,78 | 2,96 |
| 12 | 2,29 | 2,41 | 2,55 | 24 | 2,64 | 2,80 | 2,99 |
| 13 | 2,33 | 2,46 | 2,61 | 25 | 2,66 | 2,82 | 3,01 |
| 14 | 2,37 | 2,51 | 2,66 | 30 | 2,75 | 2,91 | / |

## 4.2. Linearna regresija

Linearna regresija je metoda analize podataka koja se temelji na proučavanju utjecaja neke nezavisne varijable, *x*, na zavisnu varijablu, *y*. Rezultat regresijske analize je regresijski pravac dobiven na način da su ukupne udaljenosti zavisne varijable od pravca minimalne. Postoje dva modela linearne regresije: *model jednostavne linearne regresije* i *model višestruke linearne regresije*.[31]

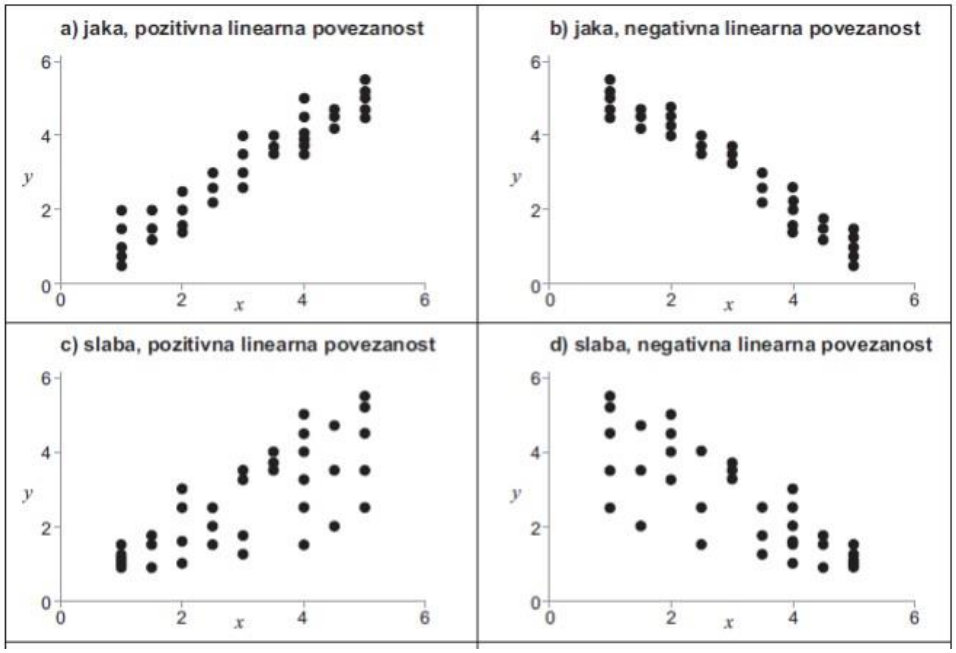
### 4.2.1. Model jednostavne linearne regresije

Model jednostavne linearne regresije ima jednu zavisnu i jednu nezavisnu varijablu te se njime opisuje odnos analiziranih podataka u kojem porastu jedne varijable odgovara približna

linearna promjena druge varijable. Stoga, jednadžba koja opisuje jednostavnu linearnu regresiju glasi:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon \tag{41}$$

Pri tome je $x$ nezavisna, a $y$ zavisna varijabla, $\beta_0$ i $\beta_1$ parametri modela, a $\varepsilon$ slučajna pogreška. Varijabla $\varepsilon$ predstavlja sve parametre koji ne utječu direktno na zavisnu varijablu.




**Slika 20.** Prikaz dijagrama rasipanja za četiri slučaja linearne regresije.

Prije početka regresijske analize potrebno je napraviti *dijagram rasipanja* (slika 20) koji prikazuje sve parove varijabli na kojima se vrši analiza. Dakle, postoji $n$ parova $\{x_i, y_i\}$ koji čine točke u koordinatnom sustavu. Ovisno o broju parova, model se opisuje jednakim brojem jednadžbi od kojih svaka predstavlja mogući pravac koji opisuje odnos među varijablama:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \tag{42}$$

Dakle, dobiva se $n$ linearnih jednadžbi od kojih svaka predstavlja regresijski pravac. Od svih dobivenih jednadžbi potrebno je odrediti onu koja će najbolje opisivati analizirane podatke, tj. koja će imati najmanje odstupanje zavisnih varijabli od pravca. Označimo ju kao:

$$y = b_0 + b_1 \cdot x \tag{43}$$

gdje je $y$ predviđena vrijednost zavisne varijable $y$, $x$ nezavisna varijabla, $b_0$ i $b_1$ procjenjeni parametri modela (43).

Budući da su sva odstupanja od potpune linearnosti utjecaj slučajnih pogrešaka, definira se tzv. *rezidualno odstupanje*, *e*, koje predstavlja procjenu nepoznatih vrijednosti varijable $\varepsilon$:

$$e_i = y_i - y_i \tag{44}$$

*Ukupno rezidualno odstupanje*, *E* pokazuje kolika je razlika između regresijskog pravca i analiziranih podataka. Što je *E* manja vrijednost, to su pravac i podaci sličniji.[32] Ukupno rezidualno odstupanje definirano je jednadžbom:

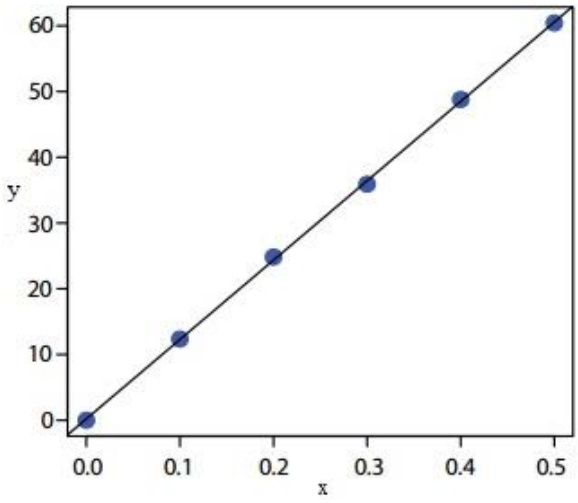
$$E = \sum_i \left( y_i - y_i \right)^2 \tag{45}$$

U linearnoj regresiji je potrebno odrediti parametre $b_0$ i $b_1$ koji daju najmanje ukupno rezidualno odstupanje. Parametri $b_0$ i $b_1$ se računaju pomoću slijedećih jednadžbi:

$$b_0 = \frac{\sum_i x_i - b_i \sum_i x_i}{n} \tag{46}$$

$$b_1 = \frac{n\sum_i x_i y_i - \sum_i x_i \sum_i y_i}{n\sum_i x_i^2 - \left( \sum_i x_i \right)^2} \tag{47}$$

Naposljetku se dobiva linearni regresijski model opisan jednadžbom (43). Na grafu regresijski pravac neće prolaziti kroz sve točke (slika 21). Da bi se odredila reprezentativnost dobivenog regresijskog modela, potrebno je procijeniti *varijancu regresije*, *standardnu devijaciju regresije* te *koeficijent varijacije regresije*.




**Slika 21.** Linearna regresija

Varijanca regresije se procjenjuje pomoću jednadžbe:

$$s_y^2 = \frac{\sum_i \left(y_i - y_i\right)^2}{n-p} = \frac{E}{n-p} = \frac{E}{n-(k+1)} \tag{48}$$

gdje *n-p*, odnosno *n*-(*k*+1), predstavlja broj stupnjeva slobode[32]. Standardna devijacija regresije se procjenjuje pomoću varijance regresije:

$$s_y = \sqrt{s_y^2} \tag{49}$$

Koeficijent varijacije regresije definiran je jednadžbom:

$$V_y = \frac{s_y}{\bar{y}} \cdot 100 \tag{50}$$

te prikazuje relativno prosječno odstupanje empirijskih od regresijskih vrijednosti zavisne varijable. Regresijski model je prihvatljiv u onom slučaju kada su procjene standardne devijacije i koeficijenta varijacije male.

### 4.2.2. Model višestruke linearne regresije

Model višestruke linearne regresije se koristi za analiziranje problema gdje postoji jedna zavisna (*y*) te više nezavisnih, promjenjivih varijabli (*x*). Model višestruke linearne regresije je tada dan jednadžbom: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .... + \beta_k x_k + \varepsilon$ (51)

Procijenjeni model višestruke linearne regresije glasi:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + .... + \beta_k x_k \tag{52}$$

u kojem $\beta_i$ predstavlja procijenjene parametre koji se računaju kao parcijalna derivacija $\partial y / \partial x_j$ gdje je *j*=1, 2, …, *k*. Procijenjeni parametri se nazivaju *regresijskim koeficijentima* te opisuju promjenu regresijske vrijednosti zavisne varijable za jedinični porast nezavisne varijable.

Kao i kod jednostavne linearne regresije, procjena varijance je bitna za određivanje reprezentativnosti modela višestruke linearne regresije. Jednadžbe su slične, uz promjenu broja parametara. Višestruka linearna regresija ima *p* parametara, pa stoga jednadžba glasi:

$$s_y^2 = \frac{\sum_i \left(y_i - y_i\right)^2}{n-p} = \frac{E}{n-p} \tag{53}$$

gdje *E* također predstavlja ukupno rezidualno odstupanje.

## 4.3. Nelinearna regresija

### 4.3.1. Polinomni regresijski model

Polinomni regresijski model se koristi u slučajevima kada nezavisna varijabla nelinearna, tj. u slučajevima kada se odnos između zavisne i nezavisne varijable ne može opisati linearnom regresijom.[33] Ukoliko je model polinomne regresije drugog stupnja sa jednom nezavisnom varijablom, tada će njegova jednadžba glasiti:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon \tag{54}$$

gdje je opet $y$ zavisna, a $x$ nezavisna varijabla, $\beta_i$ parametri koje treba procijeniti, a $\varepsilon$ slučajna pogreška.

Za procjenu parametara $\beta_i$ koristi se metoda najmanjih kvadrata. Rezultat su jednadžbe metode najmanjih kvadrata kojih ima $k$+1, ovisno o tome koliko ima nepoznatih regresijskih koeficijenata (parametara).[33] Jednadžbe koje za rezultat imaju procijenjene vrijednosti najmanjih kvadrata regresijskih koeficijenata glase:

$$n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i + \beta_2 \sum_{i=1}^{n} x_i^2 + \ldots + \beta_k \sum_{i=1}^{n} x_i^k = \sum_{i=1}^{n} y_i \tag{55}$$

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \beta_2 \sum_{i=1}^{n} x_i^3 + \ldots + \beta_k \sum_{i=1}^{n} x_i^{k+1} = \sum_{i=1}^{n} x_i y_i \tag{56}$$

...

$$\beta_0 \sum_{i=1}^{n} x_i^k + \beta_1 \sum_{i=1}^{n} x_i^{k+1} + \ldots + \beta_k \sum_{i=1}^{n} x_i^{2k} = \sum_{i=1}^{n} x_i^k y_i \tag{57}$$

gdje su $\beta_i$ procjene parametara $\beta_i$.

### 4.3.2. Jednostavni eksponencijalni regresijski model

Jednostavni eksponencijalni regresijski model ima nelinearne parametre, a njegov je oblik dan jednadžbom:

$$y = \alpha \cdot \beta^{x} \tag{58}$$

Prije upotrebe metode najmanjih kvadrata za određivanje parametara potrebno je logaritmirati jednadžbu kako bi parametri modela postali linearni uz nelinearne varijable pri čemu nezavisna varijabla *x* ostaje nepromijenjena.[33] Logaritmirana jednadžba glasi:

$$\log y = \log \alpha + x \log \beta \tag{59}$$

Jednadžbe za određivanje parametara $\alpha$ i $\beta$ su:

$$\alpha = \overline{\log y} - \log \beta \bar{x} \tag{60}$$

$$\beta = \frac{\sum_{i=1}^{n} x_i \log y_i - n\overline{x \log y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^{-2}} \tag{61}$$

Kao i kod linearne regresije, reprezentativnost metode se provjerava procjenom varijance, $\hat{s}_{\log y}$, i koeficijentom varijacije regresije, *V*:

$$\hat{s}_{\log y} = \sqrt{\frac{\sum_{i=1}^{n} \log^2 \varepsilon_i}{n - p}} \tag{62}$$

$$V = \frac{\hat{s}_{\log y}}{\log y} \cdot 100 \tag{63}$$

# 5. Primjena

## 5.1. Praktične primjene neuronskih mreža

Neuronske mreže se koriste u raznim područjima kao što su biomedicina[34–37], prometno inženjerstvo[38,39], meteorološke primjene[40], poslovne primjene[41], održivi razvoj[42] te energetika[43].

Primjena se u biomedicini ostvaruje na razne načine od kojih bih htjela spomenuti upotrebu ANN mreža za pronalaženja podudarnosti između donora i primatelja u slučajevima transplantacije jetre te za usporedbu dobivenih rezultata sa potvrđenim rezultatima za odbacivanje presadka. Istraživanje su proveli J. Briceño i suradnici[34] u bolnicama diljem Španjolske. Dobiveni su rezultati testiranja 64 donor-primatelj para koji predviđaju prihvaćanje presadka u 90,79% slučajeva. Nadalje, C. Manna i suradnici[35] su upotrijebili AI za odabir najpogodnijih jajnih stanica i embrija prilikom postupka umjetne oplodnje. Postupak je proveden na grupi jajnih stanica te na grupi embrija dobivenih iz već analiziranih jajnih stanica. U Kini su C. Zhang i suradnici[36] proveli kliničko istraživanje na pacijentima koji boluju od reumatoidnog artritisa. AI je korištena za utvrđivanje povezanosti određenih simptoma bolesti te uspješnosti biomedicinske terapije. Grupa znanstvenika iz Češke Republike i Španjolske predvođen F. Amatom je upotrijebio AI za klasificiranje simptoma i medicinskih podataka u procesu postavljanja dijagnoza pacijenta[37].

Primjena AI vidljiva je i u prometnom inženjerstvu gdje se koristi za predviđanje rada strojeva pri iskapanju tunela. Ovo istraživanje proveo je A. Salimia sa suradnicima[38] na Geotehničkom institutu u Njemačkoj. Rezultati su pokazali da se AI može učinkovito koristiti za predviđanje rada strojeva te radnih uvjeta. A. Król je primijenio AI, točnije genetički algoritam, na prometnicama svih većih gradova u Poljskoj[39]. Njegov rad se temelji na prikupljanju podataka sa svih prometnica uključujući broj vozila, njihovu prosječnu brzinu, zastoje u prometu, broj pješačkih prijelaza itd. Pomoću AI kreirao je model koji služi za optimizaciju već postojećeg prometnog sustava te je njegov model predložio prometni sustav u kojem će putovanje pojedinca biti ubrzano za 10%.

Nadalje, AI se upotrebljava u meteorologiji za predviđanje kretanja ciklona i anticiklona. Kreirani modeli prikupljaju širok spektar podataka na temelju kojih zatim predvide vremenske uvjete. Istraživanje upotrebe AI u meteorologiji proveli su Z. M. Yaseena i suradnici[40] u Maleziji.

Ekonomsko tržište, kao što znamo, vrvi podacima. Svaki dan pristižu velike količine podataka u razne sustave u bankama i ostalim financijskim institucijama. R. Iqbala i suradnici[41] su iskoristili AI kao alat za klasifikaciju ovih podataka. Njihov model korišten je za prepoznavanje ulaznih podataka te njihovo grupiranje.

U području održivog razvoja, S.W. Tsang i C.Y. Jim upotrijebili su AI za optimizaciju navodnjavanja zelenih krovova[42]. Njihov model je korišten za predviđanje vlažnosti tla na temelju podataka o vremenskim uvjetima tokom perioda od četiri dana. Također su upotrijebili neizrazite ANN za predviđanje vremena potrebnog za navodnjavanje kao i potrebnog volumena vode. Na termalnoj elektrani u Kocaeli (Turska) Y. Tunckayaa i suradnici[43] proveli su istraživanje te predložili model za predviđanje otpuštenje štetnih plinova u atmosferu korištenjem AI.

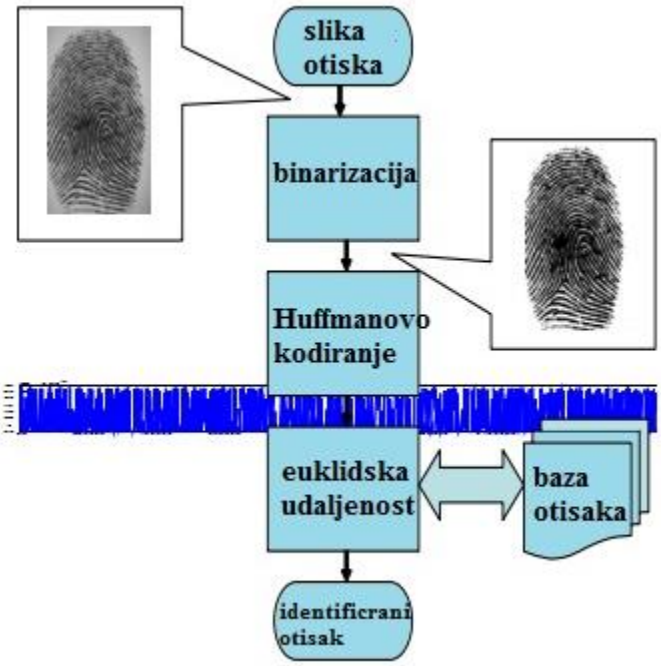
Neuronske mreže se također mogu koristiti kao alat za prepoznavanje prirodnog govora, pretvaranje napisanog teksta u govor, korištenje glasovnih naredbi za upravljanje strojevima, automatsko prevođenje. Neke navedene primjene neuronskih mreža mogu uvelike pomoći gluhim osobama i osobama koje su fizički nepokretne. Ograničenost neuronskih mreža je u tome što su u mogućnosti prepoznati govor samo one osobe koja ih je uvježbavala odnosno trenirala. U tim slučajevima dolazi do smanjenja obujma riječi koje će neuronska mreža prepoznati kada se njome koristi osoba koja ju nije uvježbavala.[44]

## 5.2. Primjena metode prepoznavanja uzoraka

Primjer primjene metode prepoznavanja uzoraka je prepoznavanje otisaka prsta. Prepoznavanje otisaka prsta ljudskim okom vrši se usporedbom linija na koži, njihovog oblika i udaljenosti. *AFIS* (engl. *automated fingerprint identification systems*) je sustav za raspoznavanje otisaka prstiju koji koristi metodu prepoznavanja uzoraka tj. umjetnu inteligenciju. AFIS pretvara sliku otiska prsta u tablicu dvodimenzionalnih vektora nazvanu *minutiae* pri čemu svaka minutia ima točno određene koordinate ($x$,$y$) koje definiraju položaj vektora i njegov smjer. u prepoznavanju otisaka prste postoje dvije metode. Prva metoda je verifikacija otiska prsta koja podrazumijeva istovremenu usporedbu otiska prsta sa samo jednim otiskom, dok identifikacija kao druga metoda podrazumijeva istovremenu usporedbu otiska sa velikim brojem drugih otisaka. Verifikacija se susreće sa mnogobrojnim problemima od kojih je najčešća poteškoća

uzrokovana djelomičnim otiskom prsta ili njegovom nejasnom slikom. U tom slučaju, teško je usporediti otisak sa onim otiskom koji je dokumentiran u sustavu na ispravan način.[45]

Na slici 24. je prikazan algoritam za prepoznavanje uzoraka otiska prsta. Prvo se uzima otisak koji u algoritam ulazi u obliku slike nakon čega se prevodi u binarni oblik u sklopu pripreme za analizu. Binarni podaci se potom kompresiraju korštenjem *Huffmanovog* kodiranja[45], nakon čega se generira jedinstveni vektor za taj otisak prsta. Generirani vektor se zatim uspoređuje metodom euklidske udaljenosti sa svim vektorima otisaka prstiju koji su dostupni u bazi podataka. Najidealniji slučaj je, naravno, da je udaljenost 0, no u većini slučajeva to nije moguće pa se prihvaća najmanja euklidska udaljenost. Nakon što je vektro uspoređen, te je nađen njegov par, sustav javlja rješenje te vlasnika ispitivanog otiska.




**Slika 22.** Prikaz postupka prepoznavanja otiska prsta.[45]

# 6. Zaključak

Umjetna inteligencija je područje koje postoji već dugi niz godina no još uvijek se intenzivno razvija. Svojom primjenom su neuronske mreže, zajedno sa prepoznavanjem uzoraka uvelike doprinijele obradi velikih količina podataka te ubrzanju rješavanja mnogih procesa. Velika prednost neuronskih mreža i umjetne inteligencije je obrada onih informacija koje nije moguće obraditi tradicionalnim sekvencijalnim postupcima.

Mogućnosti umjetne inteligencije su nevjerojatne, čak i sada, zapravo na početku razvitka te primjene obrađenih metoda u medicini, kemiji, fizici, te ostalim disciplinama, vidljivo je koliko umjetna inteligencija može biti moćan alat.

U ovom radu su obrađene metode umjetne inteligencije te njihova primjena. Vidljivo je da u analitičkoj kemiji umjetna inteligencija i neuronske mreže imaju veliki utjecaj te se učestalo koriste. Korištenjem ovih metoda u kemijskim analizama skraćeno je vrijeme dobivanja željenih rezultata, reduciran eksperimentalni rad te povećana točnost analitičkih metoda.

Jasno je da će se u budućnosti umjetna inteligencija i dalje razvijati, osobito prateći nezadrživ razvoj informatičkih znanosti. Postoji čitav niz novih pravaca istraživanja, poput hibridnih sustava, čiji potencijal se tek naslućuje i nemoguće je procijeniti gdje će nas buduća istraživanja dovesti kao ni koje su krajnje granice iskoristivosti umjetne inteligencije.

# 7. Literatura

1. B. Kröse, P. van der Smagt, An Introduction to Neural Networks, University of Netherlands, Amsterdam, 1996.
2. W. S. McCulloch, W. Pitts, *Bull. Math. Biophys.* **5** (1943) 115–133.
3. F. Rosenblatt, *Psychol. Rev.* **65** (1958) 386–408.
4. B. Widrow, M.E. Hoff, Jr., Adaptive Switching Circuits, u: IRE WESCON Convention Record, Vol. 4, pp. 96–104, The Institute of Radio Engineers, Inc., New York, SAD, 1960.
5. L.A. Zadeh, *Inform. Control* **8** (1965) 338–353.
6. D. Li, Y. Du, Artificial Intelligence with Uncertainty, Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL, SAD, 2008.
7. A.M. Turing, *Mind* **LIX** (1950) 433–460.
8. J.H. Holland, Adaptation in Natural and Artifcial Systems, University of Michigan Press, Ann Arbor, MI, SAD, 1975.
9. M. Mitchell, An Introduction to Genetic Algorithms, Massachusetts Institute of Technology, Massachusetts, MA, SAD, 1999.
10. D. Graupe, Principles of artificial neural networks, 2. izd., World Scientific Publishing, University of lllinois, Chicago, SAD, 2007.
11. B. Dalbelo Bašić, M. Čupić, J. Šnajder, Umjetne neuronske mreže, Zagreb, 2008. (https://www.fer.unizg.hr/_download/repository/UmjetneNeuronskeMreze.pdf, pristupljeno 1. srpnja 2016.)
12. D. Purves, G.J. Augustine, D. Fitzpatrick, L.C. Katz, A. LaMantia, J.O. McNamara, S. M. Williams, Neuroscience, 2 izd., Sinauer Associates, Sunderland, MA, SAD, 2001.
13. M. Khanmohammadi, Current Applications od Chemometrics, Nova Science Publishers, New York, NY, SAD, 2015.
14. D. Klemen, Raspoznavanje uzoraka primjenom umjetne kolonije pčela, diplomski rad, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2011.
15. D. Kriesel, A brief Introduction to Neural Networks, 2007. (http://www.dkriesel.com, pristupljeno 1. srpnja 2016.)
16. J.E. Dayhoff, Neural Networks Architectures: An Introduction, Van Nostrand Reinhold Co., New York, NY, SAD, 1990.
17. V. Schetinin, *Neural. Process. Lett.* **17** (2003) 21–31.

18. O.N.A. AL-Allaf, Cascade-Forward vs. Function Fitting Neural Network for Improving Image Quality and Learning Time in Image Compression System, u: Proceedings of the World Congress on Engineering 2012, Vol II (ur. S.I. Ao, L. Gelman, D.W.L. Hukins, A. Hunter, A.M. Korsunsky), Newswood Limited, London, Velik Britanija, 2012.
19. D. Britz, Recurrent Neural Networks, part 1 – Introduction to RNNs, WILDML, 2015. (http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/, pristupljeno 1. srpnja 2016.)
20. S.V. Kartalopoulos, Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications, Wiley-IEEE Press, New York, NY, SAD, 1997.
21. M. Novak Stankov, Molekulsko modeliranje i umjetna inteligencija u razvoju ionskih kromatografskih metoda, doktorski rad, Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije, Zagreb, 2015.
22. I. Petrović, N. Perić, Inteligentno upravljanje sustavima, skripta, Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb, 2007. (https://www.fer.unizg.hr/_download/repository/IUS_2.dio_SKRIPTA.pdf, pristupljeno 1. srpnja 2016.)
23. S. Oreški, Hibridne tehnike kombinatorne optimizacije temeljene na genetskim algoritmima s primjenom na odabir atributa u ocjenjivanju kreditnog rizika građana, doktorski rad, Sveučilište u Zagrebu, Fakultet organizacije i informatike, Varaždin, 2014.
24. V.K. Jain, Information Technology Issues & Challenges, Pioneer Institute of Professional Studies, Indore, 2009.
25. R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons Ltd, West Sussex, Velika Britanija, 2003.
26. G. Strang, Introduction to Linear Algebra, 5. izd., Wellesley-Cambridge Press, Wellesley, MA, SAD, 2016.
27. N. Narula, Evaluation of Card Sorting Experiments Using Factor Analysis, M.Sc. Thesis, Paderborn University, Paderborn, 2014.
28. T. Bolanča, Š. Ukić, Kemometrika u analitici okoliša, u: Analitika okoliša (ur. M. Kaštelan-Macan i M. Petrović), HINUS i Fakultet kemijskog inženjerstva i tehnologije, Zagreb, 2013.

29. B. Runje, Predavanja iz kolegija Teorija i tehnika mjerenja, Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, Zagreb, 2014. (https://bib.irb.hr/datoteka/764412.Predavanja_TTM.pdf, pristupljeno 1. srpnja 2016.)
30. M. Benšić, N. Šuvak, Primijenjena statistika, Sveučilište J.J. Strossmayera u Osijeku, Osijek, 2013.
31. I. Ličina, D. Lisjak, Primjena statističkih metoda s naglaskom na trend, korelaciju i regresiju, seminarski rad iz kolegija Revizija financijskih institucija, Sveučilište J.J. Strossmayera u Osijeku, Ekonomski fakultet, Osijek, 2014. (http://www.efos.unios.hr/imijoc/wp-content/uploads/sites/71/2013/09/Primjena-statisti%C4%8Dkih-metoda-s-naglaskom-na-trend-korelaciju-i-regresiju.pdf, pristupljeno 1. srpnja 2016.)
32. D. Harvey, Analytical Chemistry 2.0, David Harvey, Greencastle, 2009.
33. I. Lulić, Uporaba metode regresijske analize u rješavanju problema vezanih za inženjersku praksu, završni rad, Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, Zagreb, 2014.
34. J. Briceño, M. Cruz-Ramírez, M. Prieto, M. Navasa, J. Ortiz de Urbina, R.Orti, M. Á. Gómez-Bravo, A. Otero, E. Varo, S. Tomé, G. Clemente, R. Bañares, R. Bárcena, V. Cuervas-Mons, G. Solórzano, C. Vinaixa, Á. Rubín, J. Colmenero, A. Valdivieso, R. Ciria, C. Hervás-Martínez, M. de la Mata, Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: Results from a multicenter Spanish study, u: Journal of Hepatology, Vol. 61, pp. 1020-1028, Elsevier B.V., Córdoba, Španjolska, 2014.
35. C. Manna, L. Nannic, A. Luminid, S. Pappalardoa, Artificial intelligence techniques for embryo and oocyte classification, u: Reproductive BioMedicine Online, Vol. 26, pp. 42-49, Elsevier Ltd., Amsterdam, Nizozemska, 2013.
36. C. Zhang, Q. Zha, Y. He, M. Jiang, C. Lü, A. Lü, The Extraarticular Symptoms Influence ACR Response in the Treatment of Rheumatoid Arthritis with Biomedicine: A Single-blind, Randomized, Controlled, Multicenter Trial in 194 patients, u: Journal of Traditional Chinese Medicine, Vol. 31, pp. 50-55, Elsevier Pte. Ltd., Amsterdam, Nizozemska, 2011.

37. F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, J. Havel, Artificial neural networks in medical diagnosis, u: Journal of Applied Biomedicine, Vol. 11, pp. 47-58, Elsevier Urban & Partner Sp. z.o.o., České Budějovice, Češka Republika, 2013.
38. A. Salimia, J. Rostamib, C. Moormanna, A. Delisioc, Application of non-linear regression analysis and artificial intelligence algorithms for performance prediction of hard rock TBMs, u: Tunnelling and Underground Space Technology, Vol. 58, pp. 236-248, Elsevier Ltd., Amsterdam, Nizozemska, 2016.
39. A. Król, The Application of the Artificial Intelligence Methods for Planning of the Development of the Transportation Network, u: Transportation Research Procedia, Vol. 14, pp. 4532-4541, Faculty of Transport, University of Technology, Katowice, Poljska, 2016.
40. Z. M. Yaseena, A. El-shafiea, O. Jaafara, H. A. Afana, K. N. Sayl, Artificial intelligence based models for stream-flow forecasting: 2000–2015, u: Journal of Hydrology, Vol. 530, pp. 829-844, Elsevier B.V., Amsterdam, Nizozemska, 2015.
41. R. Iqbala, F. Doctora, B. Morea, S. Mahmuda, U. Yousuf, Big Data analytics: Computational intelligence techniques and application areas, u: International Journal of Information Management, Elsevier Ltd., Amsterdam, Nizozemska, 2016.
42. S.W. Tsang, C.Y. Jim, Applying artificial intelligence modeling to optimize green roof irrigation, u: Energy and Buildings, Vol. 127, pp. 360-369, Elsevier Ltd., Amsterdam, Nizozemska, 2016.
43. Y. Tunckayaa, E. Koklukaya, Comparative analysis and prediction study for effluent gas emissions in a coal-fired thermal power plant using artificial intelligence and statistical tools, u: Journal of the Energy Institute, Vol 88, pp. 118-125, Elsevier Ltd., Amsterdam, Nizozemska, 2015.
44. J. Popčević, I. Varga, P. Žuvela, Izračunavanje temperature vrelišta uz pomoć umjetnih neuronskih mreža, seminarski rad iz kolegija Uvod u matematičke metode u inženjerstvu, Sveučilište u Zagrebu, Fakultet kemijskog inženjerstva i tehnologije, Zagreb, 2012. (http://matematika.fkit.hr/novo/izborni/referati/Popcevic_Varga_Zuvela_Neuronske_mreze.pdf, pristupljeno 1. srpnja 2016.)

45. A.A. Aburas, S.A. Rehiel, Fingerprint Patterns Recognition System Using Huffman Coding, u: Proceedings of the World Congress on Engineering 2008, Vol III (ur. S.I. Ao, L. Gelman, D.W.L. Hukins, A. Hunter, A.M. Korsunsky), Newswood Limited, London, Velika Britanija, 2008.

# Životopis

Dorotea Gajdek rođena je 18. kolovoza 1993. godine u Zagrebu. 2000. je započela sa svojim osnovnoškolskim obrazovanjem pohađajući školu "Mate Lovrak" u Zagrebu. Srednjoškolsko obrazovanje stekla je u Petoj gimnaziji u Zagrebu gdje je maturirala 2012. godine. Nakon mature upisuje studij Primijenjene kemije na Fakultetu kemijskog inženjerstva i tehnologije Sveučilišta u Zagrebu.

Tijekom studiranja odradila je stručnu praksu u farmaceutskoj kompaniji Pliva Hrvatska d.o.o. u odjelu Istraživanje i razvoj.