

Označavanje podataka i metrike

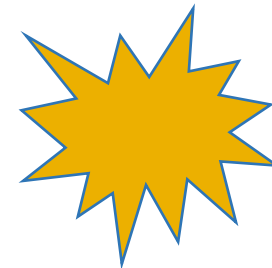
Uvod u znanost o podacima

8. predavanje

doc. dr. sc. Ana Sović Kržić

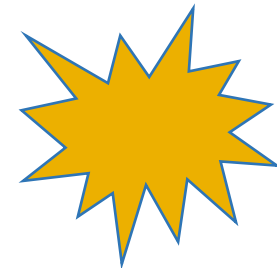
2021./2022.

Sadržaj



- Označavanje podataka
- Metrike

Data labeling – označavanje podataka

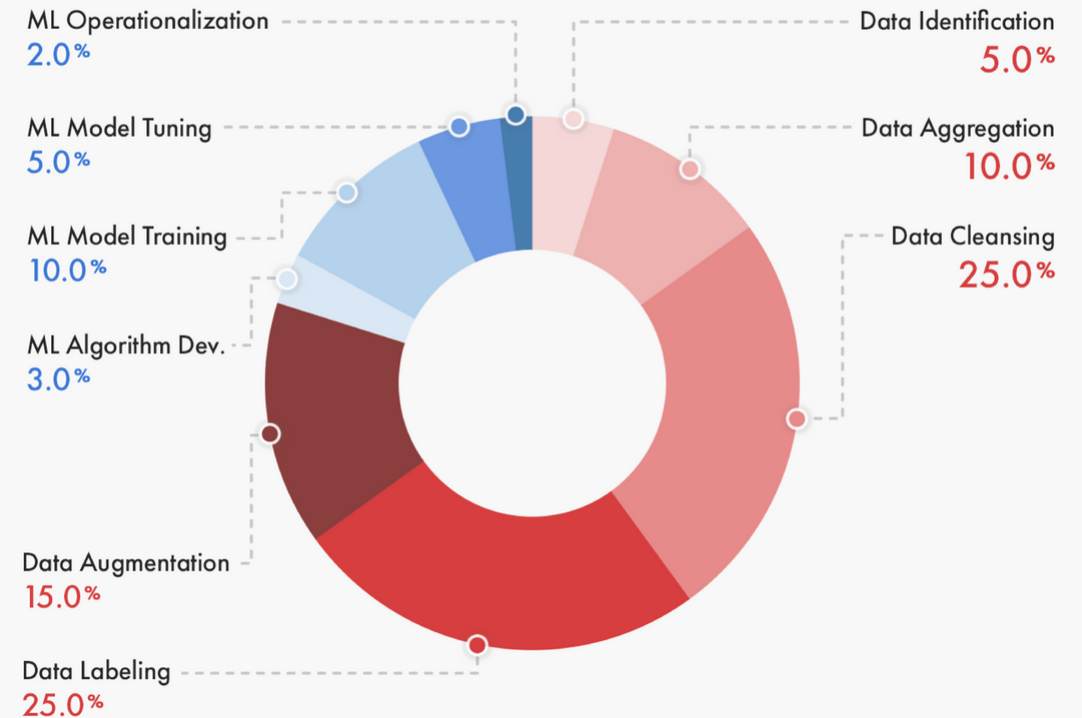


- proces **dodavanja oznaka** na sirove podatke (npr. slike, video, tekst, audio)
- ove oznake obilježavaju kojoj klasi (skupini) objekata promatrani podaci pripadaju i pomažu algoritmima strojnog učenja identificirati određenu klasu (skupinu) objekata kada se nađu u podacima bez oznaka

Označavanje podataka

„Salaries for data scientists can cost up to \$190,000/year. It's expensive to have some of your highest-paid resources wasting time on basic, repetitive work.”

Percentage of Time Allocated to Machine Learning Project Tasks



Source: Cognilytica

Podaci za učenje (training data)

- podaci koji su prikupljeni i pomoću kojih model strojnog učenja uči o podacima
- mogu biti različitih oblika: slike, glas, tekst, značajke – ovisno o korištenom modelu strojnog učenja i cilju koji se želi postići
- podaci mogu biti **anotirani ili neanotirani** – kada su anotirani uzimaju se kao „ground truth“ tj. „**referentna vrijednost**“

Referentna vrijednost

- koristi se za informacije za koje se unaprijed zna da su **istinite**
- npr. algoritam za izoštravanje slike – uzmemo oštru sliku (ground truth) – zamutimo ju – takvo zamućenu sliku izoštrimo našim algoritmom – i usporedimo s ground truth kako bismo vidjeli uspješnost algoritma

Vrste učenja (1)

supervised learning

Input data



Annotations

These are
apples



Model



Prediction

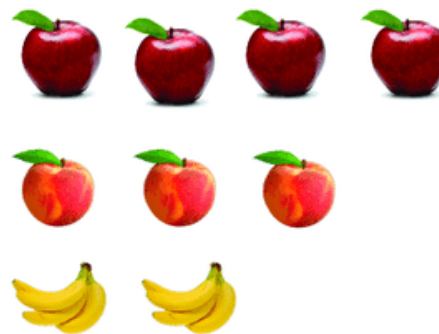
Its an
apple!

unsupervised learning

Input data



Model

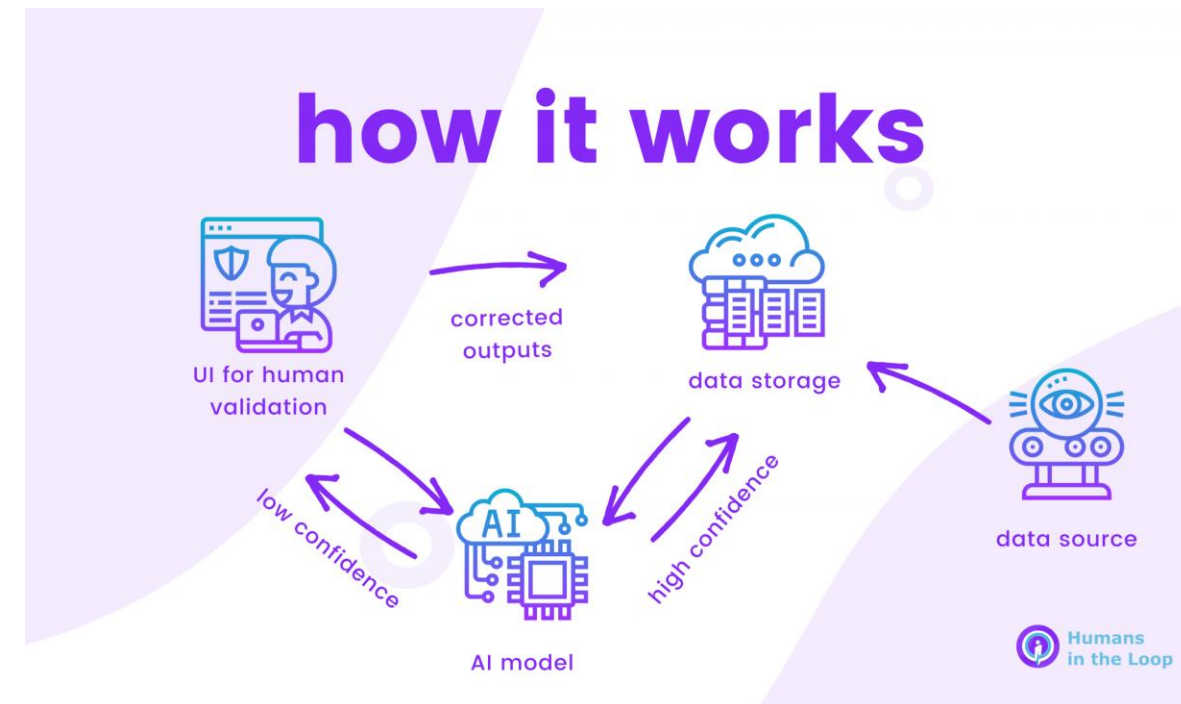


Vrste učenja (2)

Supervised learning Nadzirano učenje	Unsupervised learning Nenadzirano učenje	Semi-supervised learning Polunadzirano učenje
anotirani podaci	neanotirani podaci	anotirani i neanotirani podaci
koristi se za predikciju	koristi se za analizu	
ima mehanizam povratne informacije	nema mehanizam povratne informacije	
regresija, klasifikacija, segmentacija	autoenkoderi koji imaju izlaze jednake ulazima, klastering	klasificiranje sekvenci proteina, analiza sadržaja Interneta

Human-in-the-loop HITL

- stalni **nadzor i provjera rezultata AI modela od strane čovjeka**: provjera radi li predikcija ispravno, identifikacija praznina u podacima za učenje, povratna informacija modelu
- Upotreba:
 - **označavanje podataka** za učenje: ljudski anotatori označavaju podatke za učenje koji se unose u (nadzirane/polu-nadzirane) modele strojnog učenja
 - **učenje modela**: ljudi uče model stalnim nadzorom detalja modela kao što su funkcija gubitka i predviđanja. S vremena na vrijeme ljudi validiraju izvedbu modela i predviđanja, a rezultati provjere se vraćaju modelu



Skala

- omogućiti **fleksibilan proces** označavanja koji omogućuje skaliranje kako se potrebe i slučajevi upotrebe budu razvijali
- npr. 1h videa – oko 800 čovjek-sata za anotiranje
- 10 min video (30-60 fps) = 18.000-36.000 slika (frame)

Pristupi označavanju podataka (1)

- ovisi o problemu, vremenskom okviru projekta, raspoloživom broju ljudi

1. In-house data labeling

- osigurava **najveću kvalitetu**
- obično označavaju **znanstvenici ili osobe u organizaciji** (mogu biti i zaposleni na dio radnog vremena / osobe kojim obilježavanje podataka nije u opisu posla)
- ispravno obilježavanje je bitno u npr. osiguranju ili zdravstvu
- često zahtijeva konzultiranje sa stručnjacima u području za ispravno obilježavanje podataka
- za veću kvalitetu oznaka – vrijeme raste drastično → cijeli proces je **vrlo spor**

Pristupi označavanju podataka (2)

2. Crowdsourcing

- **uz pomoć velikog broja freelancera registriranih na crowdsourcing platformi** (imaju desetke tisuća registriranih anotatora podataka)
- obilježeni skupovi podataka uglavnom se sastoje od trivijalnih podataka, npr. slika životinja, biljaka i prirodnog okoliša i ne zahtijevaju dodatnu stručnost
- https://www.youtube.com/watch?v=6E_IJR22oXk

Pristupi označavanju podataka (3)

3. Outsourcing

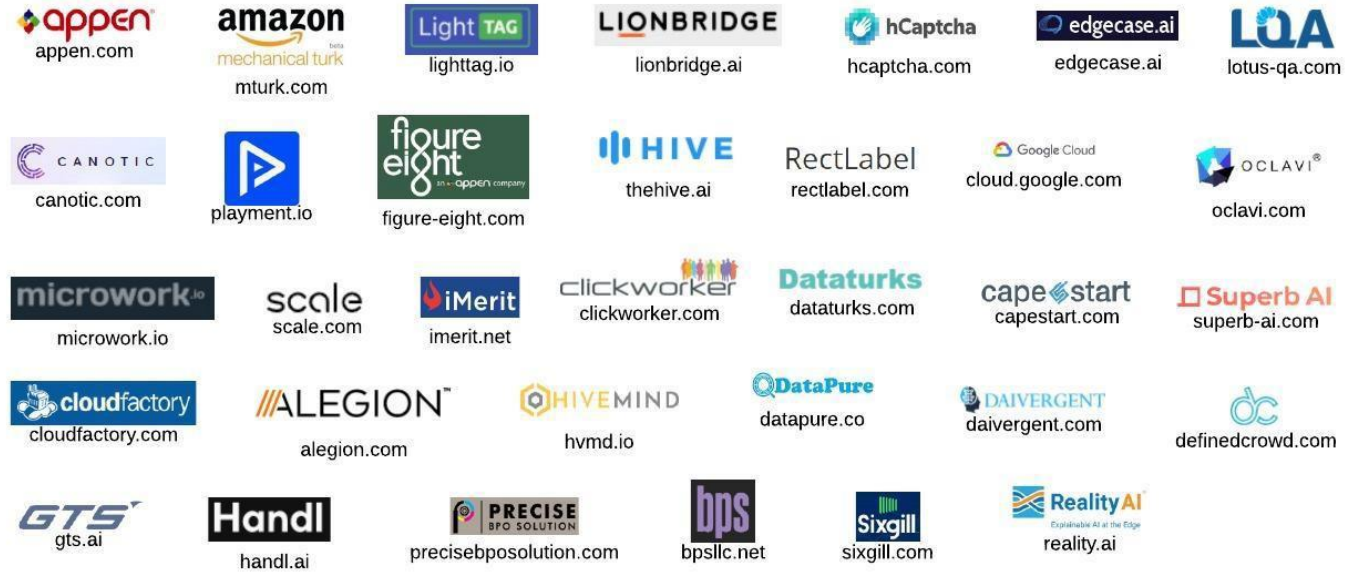
- sredina između crowdsourcinga i in-house označavanja
- zadatak označavanja podataka prepušta se **pojedincu ili organizaciji koja ima trenirane anotatore**
- jedna od prednosti outsourcinga pojedincima je da oni mogu biti procijenjeni za određenu temu prije nego što im se posao preda
- za projekte koji nemaju puno financiranja, ali zahtijevaju značajnu kvalitetu obilježavanja podataka
- <https://www.bbc.com/news/technology-46055595>

Pristupi označavanju podataka (4)

4. Machine-based annotation

- korištenjem **alata za anotiranje i automatizacije** što drastično može povećati brzinu anotiranja bez smanjivanja kvalitete
- automatizacija koristi nenadzirane (klastering) i polu-nadzirane metode strojnog učenja

Managed Labeling
Solutions / Platforms



Active Learning API
support



Labeling Functions
(Weak supervision)



Osiguravanje kvalitete

- **izuzetno bitno zbog točnosti strojnog učenja koji će koristiti te podatke**, npr. obilježavanje prolaznika, znakova i drugih vozila za samovozeći automobil
- dva pogleda:
 - **Točnost (accuracy)** – mjeri koliko je dobro stavka obilježena u usporedbi s uvjetima u stvarnom svijetu
 - **Kvaliteta (quality)** – točnost za cijeli skup podataka – da li posao svih anotatora izgleda jednako

Metode mjerenja kvalitete

1. **Zlatni standard** – postoji točan odgovor za zadatak – kvaliteta se mjeri na temelju točnih i netočnih zadataka
2. **Pregled uzorka** – odabere se nasumični uzorak dovršenih zadataka, iskusniji radnik (npr. voditelj tima ili projekta) pregledava uzorak
3. **Konsenzus** – nekoliko ljudi izvrši isti zadatak, a točan odgovor je onaj koji dolazi od većine anotatora
4. **Intersection over union (IoU)** – model konsenzusa koji se često koristi u otkrivanju objekata unutar slika – kombinira ljude i automatizaciju za usporedbu „graničnih okvira” ručno označenih slika s predviđenim graničnim okvirima iz modela

Cronbach alpha

- mjera prosječne korelacije ili konzistentnosti stavki u skupu podataka
- ovisno o karakteristikama istraživanja (na primjer, homogenosti), može pomoći u brzom pristupu sveukupnoj pouzdanosti oznaka
- mjera pouzdanosti – **govori koliko je ljestvica dosljedna iznutra**
- uz N – broj stavki (pitanja), \bar{r} – srednja korelacija između stavki

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}}$$

Cronbach Alpha	unutarnja dosljednost
$0.9 \leq \alpha$	odlično
$0.8 \leq \alpha < 0.9$	dobro
$0.7 \leq \alpha < 0.8$	prihvatljivo
$0.6 \leq \alpha < 0.7$	upitno
$0.5 \leq \alpha < 0.6$	slabo
$\alpha < 0.5$	neprihvatljivo

Cronbach alpha primjer

- vlasnik restorana želi izmjeriti opće zadovoljstvo posjetitelja
- 10 posjetitelja je zamolio da odgovore na 3 pitanja (Q1, Q2, Q3 → N = 3 pitanja) s ocjenama 1 – 2 – 3

	Q1	Q2	Q3
0	1	1	1
1	2	1	1
2	2	1	2
3	3	2	1
4	2	3	2
5	2	3	3
6	3	2	3
7	3	3	3
8	2	3	2
9	3	3	3

izračunati korelacijsku matricu: r_{xy}

	Q1	Q2	Q3
Q1	1.000000	0.429945	0.507630
Q2	0.429945	1.000000	0.662842
Q3	0.507630	0.662842	1.000000

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$\bar{r} = \frac{0.429945 + 0.507630 + 0.662842}{3} = 0.5335$$

$$\alpha = \frac{N \cdot \bar{r}}{1 + (N - 1) \cdot \bar{r}} = \frac{3 \cdot 0.5335}{1 + (3 - 1) \cdot 0.5335} = 0.774$$

Vrste oznaka - primjeri

- ovisi o tome što želimo da bude rezultat strojnog učenja
- najčešća područja:
 - Računalni vid
 - obilježeni vizualni podaci u obliku slika
 - Obrada prirodnog jezika
 - analiza ljudskog jezika i njihovih oblika tijekom interakcije s drugim ljudima i strojevima
 - Govor / zvuk

Klasifikacija slika

- **dodavanje oznake slici**
- broj jedinstvenih oznaka u cijeloj bazi podataka je broj klasa koje model može klasificirati
- klasifikacijski problem se može podijeliti na:
 - **Binarna klasifikacija** (koja se sastoji od samo dvije oznake)
 - **Višeklasna klasifikacija** (koja sadrži više oznaka)
- moguća i klasifikacija s **više oznaka**, npr. kod otkrivanja bolesti → svaka slika ima više od jedne oznake

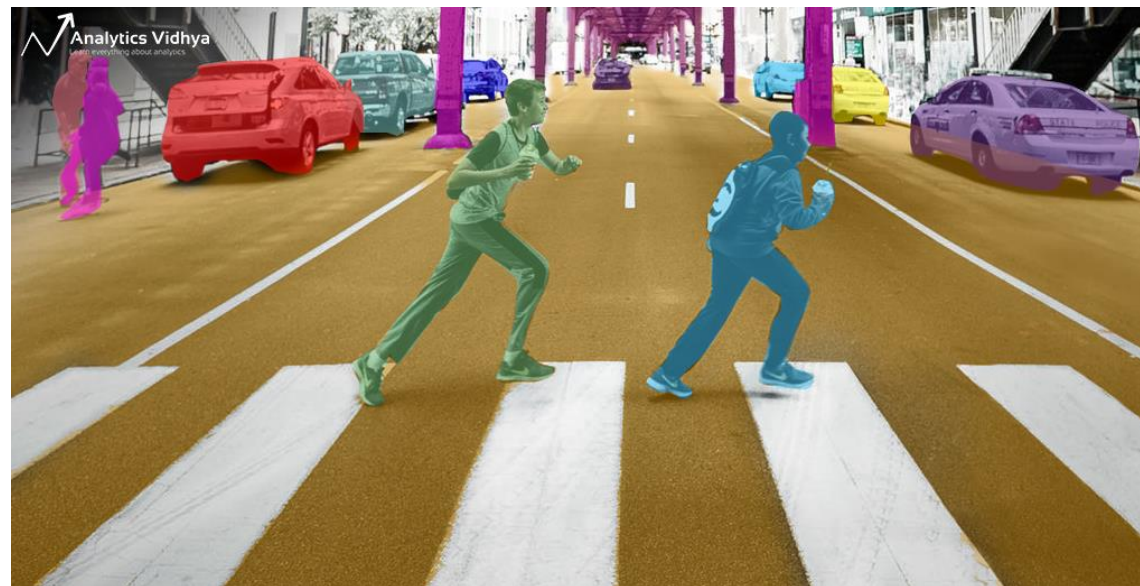
Classification



CAT

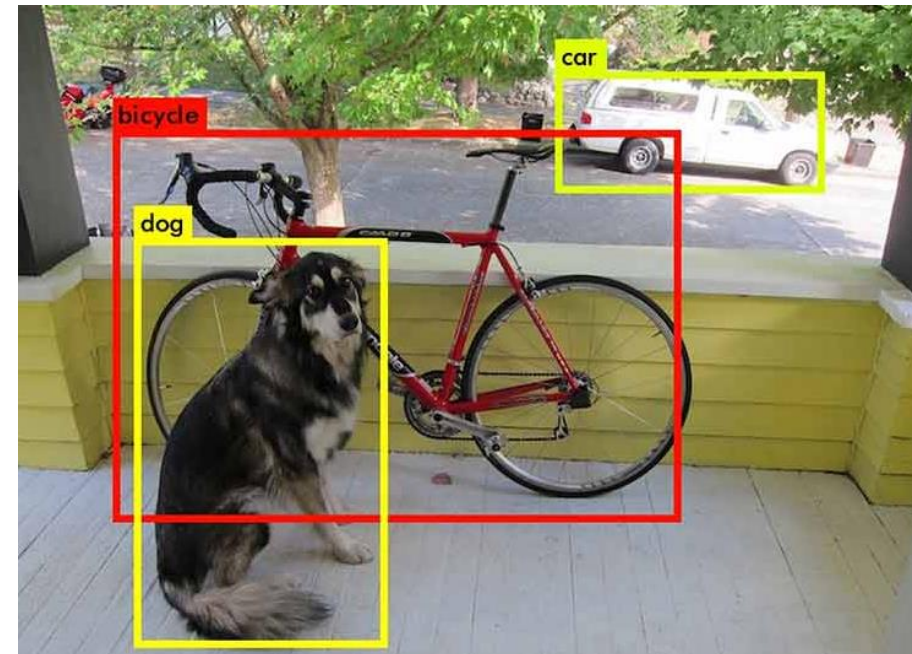
Segmentacija slika

- **odvojiti objekte slike od njihove pozadine i drugih objekata** na slici
→ obično je rezultat slika jednake veličine kao i originalna slika, koja sadrži 1 gdje je objekt prisutan i 0 inače
- ako se unutar jedne slike segmentira više objekata, svaki objekt je obilježen na posebnom kanalu → zbroj svih kanala je referentna slika



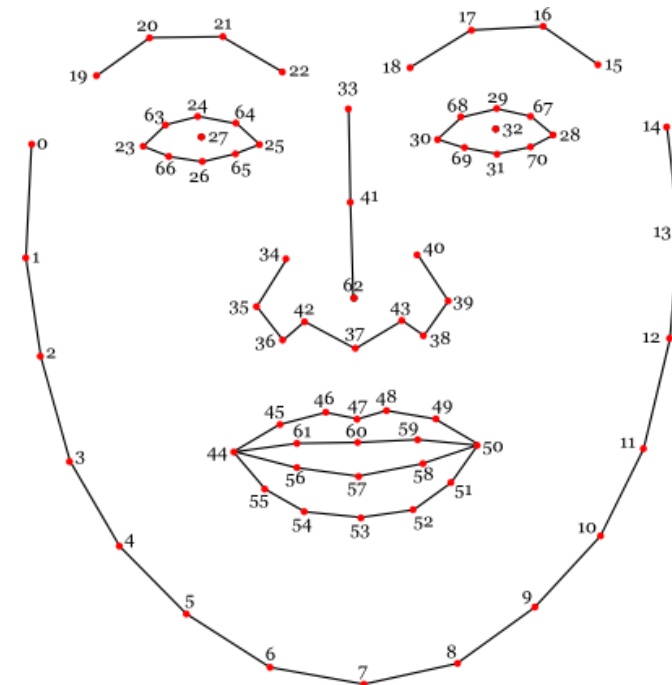
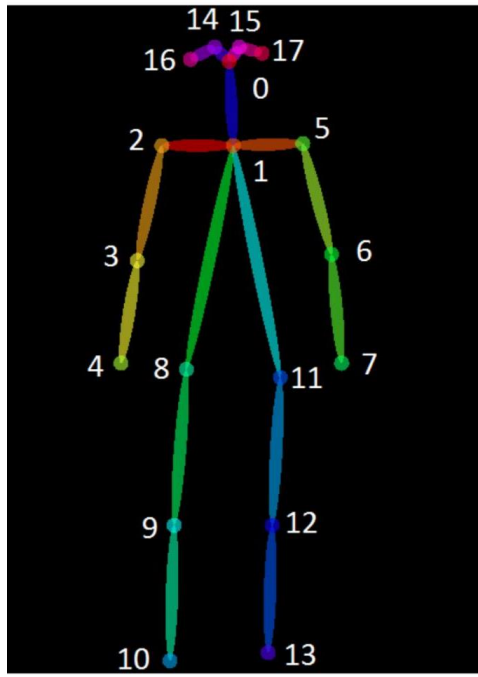
Detekcija objekata

- **detekcija objekata i njihovih lokacija**
- svaki objekt je obilježen najmanjim mogućim pravokutnikom koji ga okružuje (**bounding box**)
- obično se svakom pravokutniku pridružuje i **oznaka**
- pamte se **koordinate pravokutnika** i pripadajuća oznaka te se pohranjuje u JSON datoteku u formatu rječnika gdje je ključ rječnika broj ili ID slike



Estimacija poze čovjeka

- detekcija **ključnih točaka** na tijelu i njihovo povezivanje s pozom
- referentna vrijednost: koordinate i pripadajuće oznake ključnih točaka



Obilježavanje entiteta

- lociranje, izdvajanje i označavanje entiteta u tekstu
- anotatori čitaju tekst, lociraju ciljne entitete i obilježavaju ih pomoću unaprijed definiranih oznaka
- riječ „entitet” može imati različite oblike ovisno o zadatku:
 - **vlastite imenice** → obilježavanje entiteta se odnosi na identifikaciju i označavanje imena u tekstu
 - za analizu **fraza** → obilježavaju se ključne riječi ili ključne fraze
 - za analizu i označavanje **funkcionalnih elemenata** bilo kojeg teksta kao što su glagoli, imenice, prijedlozi → označavanje dijelova govora (**Parts of Speech tagging, POS**) → koristi se za raščlanjivanje, strojno prevođenje i generiranje jezičnih podataka



Povezivanje entiteta

- označavanje entiteta prati povezivanje entiteta, gdje se označeni entiteti **povezuju sa spremištima podataka** o njima kako bi se svakom od tih entiteta dodijelio jedinstveni identitet
- važno kada tekst sadrži podatke koji mogu biti dvosmisleni, a moraju biti nedvosmisleni
- povezivanje entiteta često se koristi za semantičku anotaciju, gdje se semantičke informacije entiteta dodaju kao oznake
- za poboljšanje pretraživanja ili korisničkog iskustva

Klasifikacija teksta

- dodavanje jedne ili više **oznaka blokovima teksta**
- tekst se promatra kao cjelina (sadržaj, predmet, namjera, osjećaj unutar teksta) i oznaka se dodjeljuje toj cjelini (na temelju poznatog popisa oznaka)
- vrste klasifikacije teksta:
 - klasifikacija na temelju **osjećaja ili mišljenja** (za analizu osjećaja),
 - klasifikacija na temelju **teme** koju tekst želi prenijeti (za kategorizaciju tema)
 - klasifikacija **dokumenata** – za sortiranje i dohvaćanje dokumenata na temelju sadržaja



Anotacija osjećaja

- detekcija skrivenih konotacija, sarkazma, duhovitosti – pravih emocija ispod teksta
- označavanje: **emocija, mišljenja, sentimenta**
- npr. analiza recenzija gostiju – na temelju danih recenzija anotatori trebaju izabrati oznaku: pozitivni / neutralno / negativno

I love this product! It's great.

This product is alright. Not great, but functional.

This product is horrible! Switch products for sure.

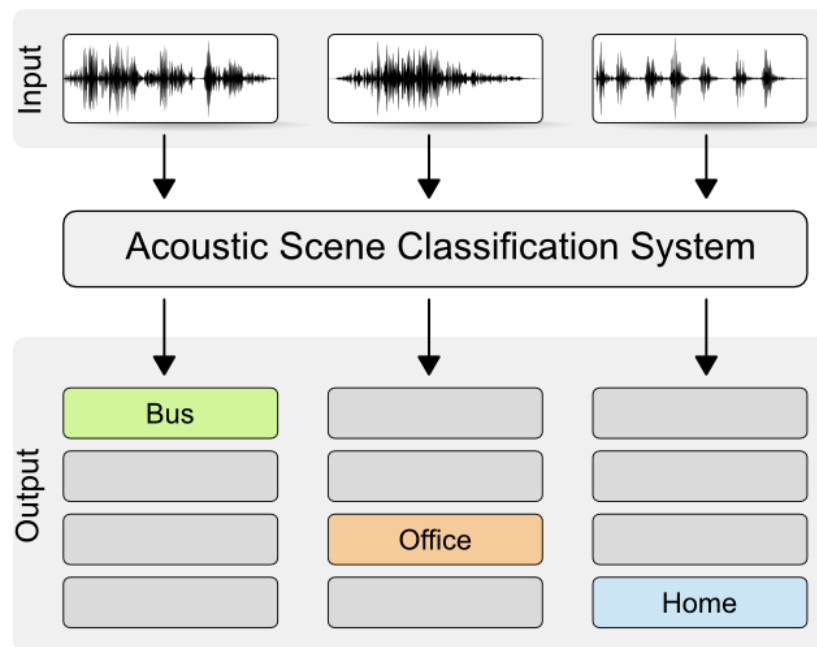
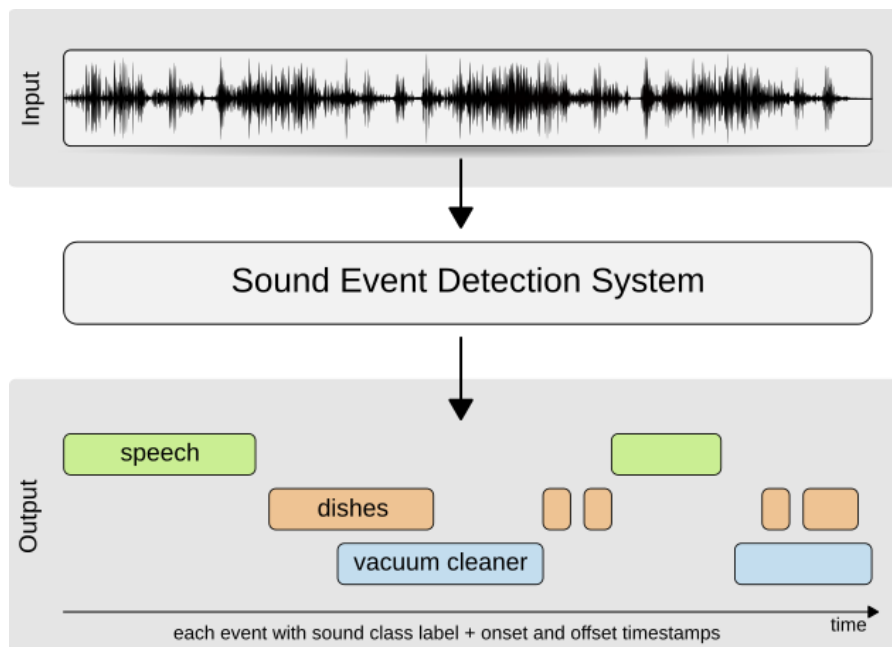
Sentiment



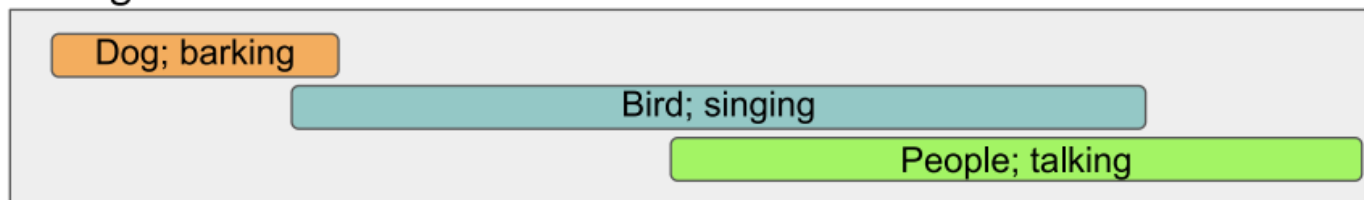
Lingvistička anotacija

- identifikacija i označavanje **gramatičkih, semantičkih ili fonetskih elemenata** u tekstu ili govoru
- posebno se koristi u chatbotovima, virtualnim asistentima, pretraživačima, prevoditeljima
- Vrste:
 - Anotacija **diskursa** – povezivanje anafora i katafora s njihovim prethodnim ili slijedećim subjektima. Primjer: Ivan je razbio stolicu. Osjećao se jako loše zbog toga.
 - Označavanje **dijela govora** (POS) – bilješka o različitim funkcijskim riječima unutar teksta
 - **Fonetska** anotacija – označavanje intonacije, naglasaka i prirodnih pauza u govoru
 - **Semantička** anotacija – anotacija definicija riječi

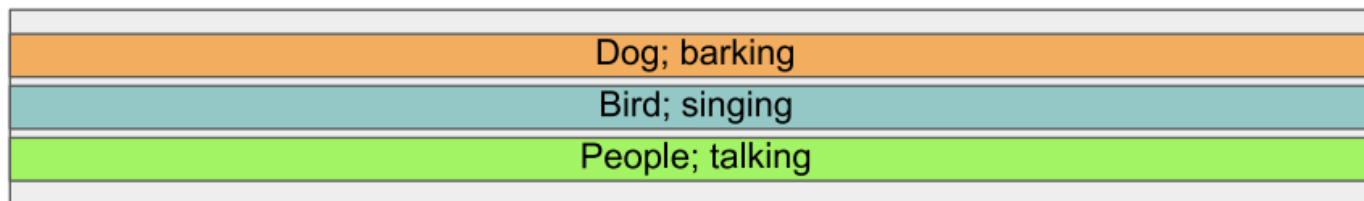
Zvuk



Strong labels

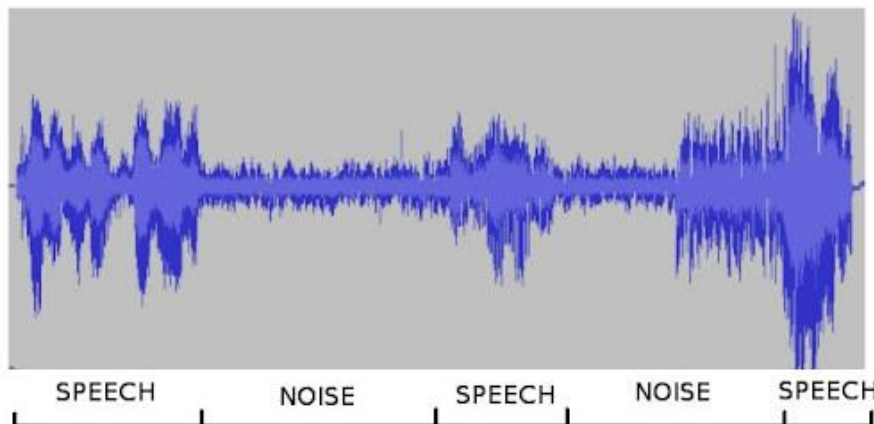


Weak labels

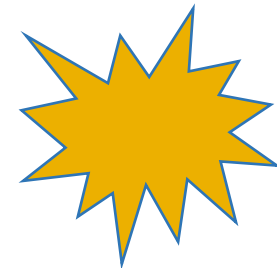


Govor

- identifikacija govornika
 - dodavanje oznake audio datoteci
- lingvistički podaci
 - prvo se obilježavaju jezične regije jer se ne očekuje da zvuk sadrži 100 posto govora
 - označavaju se okolni zvukovi i stvara se prijepis govora za daljnju obradu uz pomoć NLP algoritama

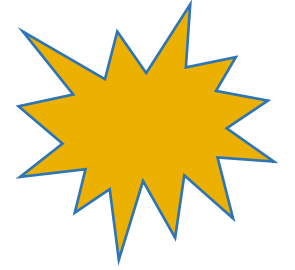


Metrike



- koriste se za praćenje i mjerenje izvedbe modela (za vrijeme učenja i testiranja)
- Različite metrike za različite zadatke:
 - Klasifikacija
 - Regresija
 - Rangiranje
 - Obrada slike
 - Duboko učenje
 - NLP

Klasifikacija



- **primjena:** prepoznavanje lica, kategorizacija Youtube videa, moderacija sadržaja, medicinske dijagnoze, klasifikacija teksta, detekcija govora mržnje
- **popularni modeli:** support vector machine (SVM), logička regresija, stabla odluke, slučajne šume, XGboost, konvolucijske neuronske mreže, recurrent neural network

Matrica zabune

- confusion matrix – konfuzijska matrica – matrica zabune
- nije metrika, već tablična vizualizacija dobivenih predikcija u odnosu na anotirane oznake
- dijagonalni elementi – točne predikcije za različite klase
- elementi van dijagonala – broj uzoraka koji su krivo klasificirani

		Stvarne klase	
		Klasa I	Klasa II
Predviđene klase	Klasa I	TP	FP
	Klasa II	FN	TN

Matrica zabune - primjer

- binarna klasifikacija je li na slici pas ili nije pas
- set za testiranje ima 1100 slika: 100 slika predstavlja pse, a 1000 ne sadrži pse
- 80 slika koje sadrže pse je dobro predviđeno (true-positive TP), 20 nije (false negative FN)
- od 1000 slika koje nemaju pse: 950 je predviđeno ispravno (true-negative TN), 60 je predviđeno pogrešno (false-positive FP)

		Stvarne klase	
		Psi	Nisu psi
Predviđene klase	Psi	80	50
	Nisu psi	20	950

Točnost (1)

- accuracy
- omjer **točno pogođenih u odnosu na ukupan broj predviđanja**

$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

- primjer:

$$CA = \frac{80 + 950}{80 + 950 + 20 + 50} = 93.6\%$$

Točnost (2)

- nije dobra mjera ako je distribucija klasa nebalansirana = jedne klase ima puno više od drugih → može doći do slučaja da se svi uzorci predvide da su iz najčešće klase → točnost će biti visoka, a model ništa ne predviđa, već jednostavno sve stavlja u najčešću klasu
- npr. sve slike su predviđene kao "nisu psi"

$$CA = \frac{0 + 1000}{0 + 1000 + 0 + 100} = 90.9\%$$

		Stvarne klase	
		Psi	Nisu psi
Predviđene klase	Psi	0	0
	Nisu psi	100	1000

Preciznost

- precision
- mjera **izvedbe pojedine klase** (po retcima)

$$PR = \frac{TP}{TP + FP}$$

$$PR = \frac{TN}{TN + FN}$$

- primjer:
 - PR za točno predviđene pse: $PR = \frac{80}{80+50} = 61.5\%$
 - PR za točno predviđanje da na slici nisu psi: $PR = \frac{950}{950+20} = 97.9\%$
 - puno veća preciznost za predviđanje da na slici nisu psi u odnosu na predviđanje da su na slici psi → zato što ima puno više slika na kojima nisu psi

Recall

- omjer uzoraka klase koja je točno predviđena (po stupcima)

$$RE = \frac{TP}{TP + FN}$$

$$RE = \frac{TN}{TN + FP}$$

- Primjer:

- RE za točno predviđene pse: $RE = \frac{80}{80+20} = 80\%$
- RE za točno predviđanje da na slici nisu psi: $RE = \frac{950}{950+50} = 95\%$

F1 score

- kombinacija PR i RE – **harmonijska sredina PR i RE**

$$F1 = \frac{2 \cdot PR \cdot RE}{PR + RE}$$

- trade-off između PR i RE nekog modela – ako je PR prevelika, RE postaje jako mala i obratno

- Primjer: $F1 = \frac{2 \cdot 0.615 \cdot 0.8}{0.615 + 0.8} = \frac{0.984}{1.415} = 69.5\%$

- Generalizirana F-vrijednost

$$F_{\beta} = (1 + \beta^2) \frac{PR \cdot RE}{\beta^2 \cdot PR + RE}$$

ROC krivulja

- receiver operating characteristic curve
- grafički prikaz izvedbe **binarnog klasifikatora kao funkcija različitih pragova** korištenih pri klasifikaciji
- true positive rate (TPR) u odnosu na false positive rate (FPR) za različite pragove
- **true positive rate (TPR)** (= recall)

$$TPR = \frac{TP}{TP + FN}$$

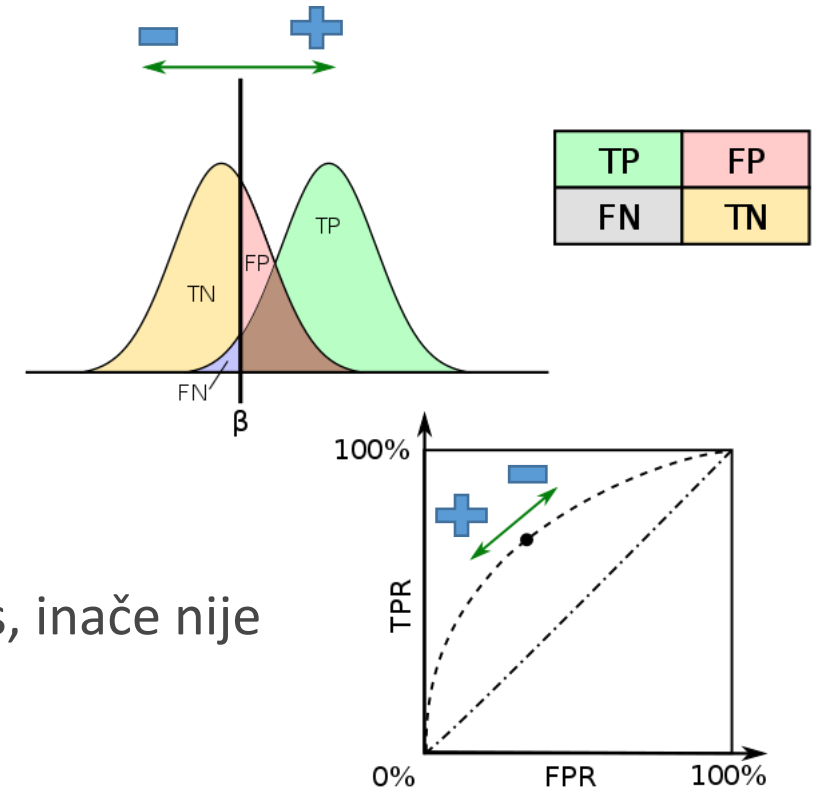
- **false positive rate (FPR)**

$$FPR = \frac{FP}{FP + TN}$$



ROC krivulja - primjer

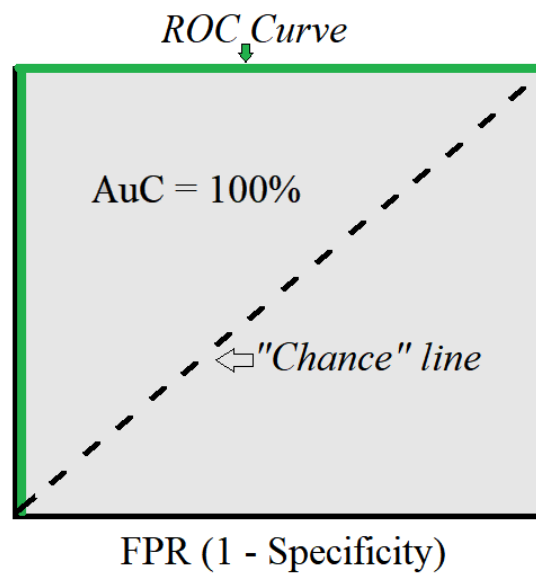
- većina klasifikacijskih modela radi pomoću **vjerojatnosti** – predviđaju vjerojatnost da je na slici pas
- vjerojatnost koju dobiju **usporede sa zadanim pragom** – ukoliko je dobivena vjerojatnost veća od praga, na slici je pas, inače nije
- npr. dobivene vjerojatnosti za 4 slike: [0.4, 0.6, 0.3, 0.8].
- Uz različite pragove, dobiva se:
 - prag = 0.2 → predviđanje = [1, 1, 1, 1]
 - prag = 0.5 → predviđanje = [0, 1, 0, 1]
 - prag = 0.7 → predviđanje = [0, 0, 0, 1]
- u ovisnosti o pragu – kao rezultat se dobiva različiti PR, RE, TPR, FPR
- **ROC krivulja prikazuje TPR i FPR za različite pragove**



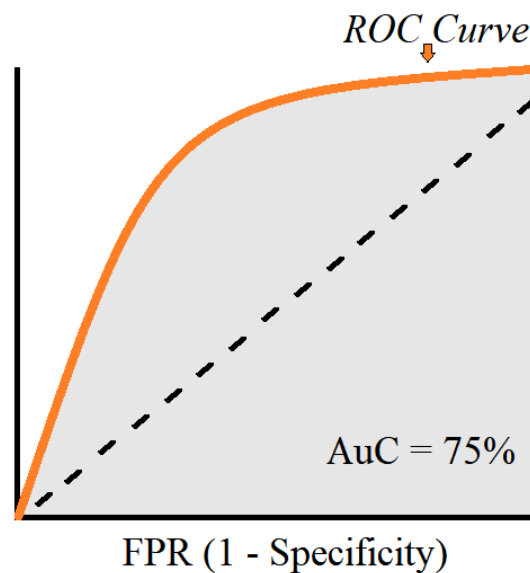
AUC

- area under the curve (AUC)
- agregirana mjera izvedbe binarnog klasifikatora na svim mogućim vrijednostima praga (i stoga je invarijantna pragu)
- vjerojatnost da će model više rangirati slučajni pozitivan primjer od slučajnog negativnog primjera
- **površina ispod ROC krivulje, iznosi između 0 i 1**
- veći AUC – bolji model
 - model čije su predikcije 100% krive \rightarrow AUC = 0.0
 - model čije su predikcije 100% točne \rightarrow AUC = 1.0

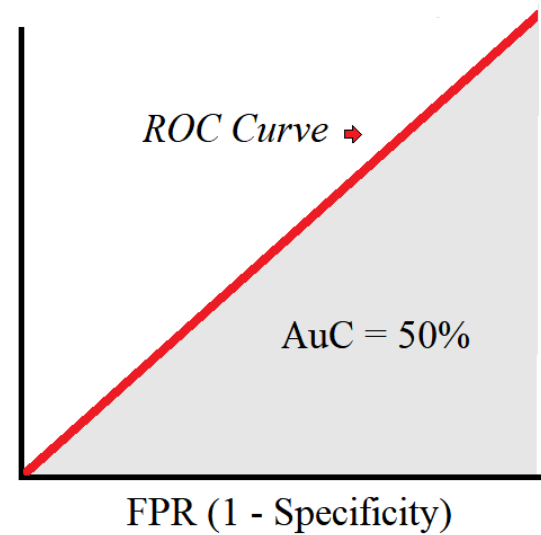
TPR (Sensitivity)



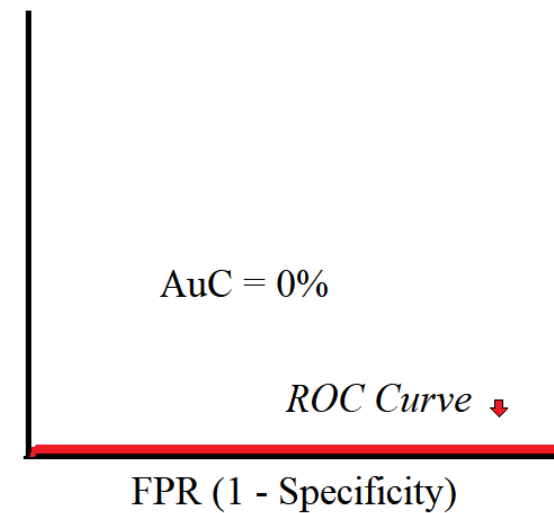
TPR (Sensitivity)



TPR (Sensitivity)



TPR (Sensitivity)



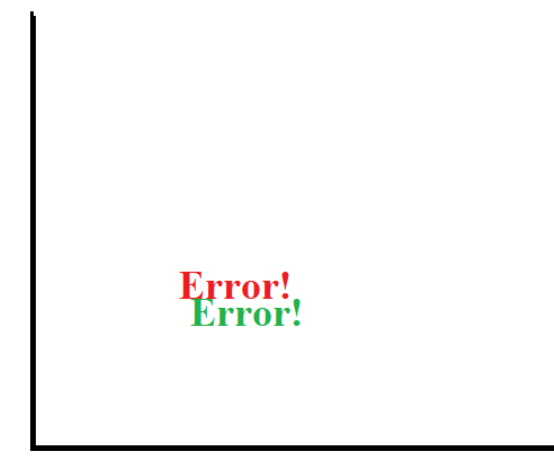
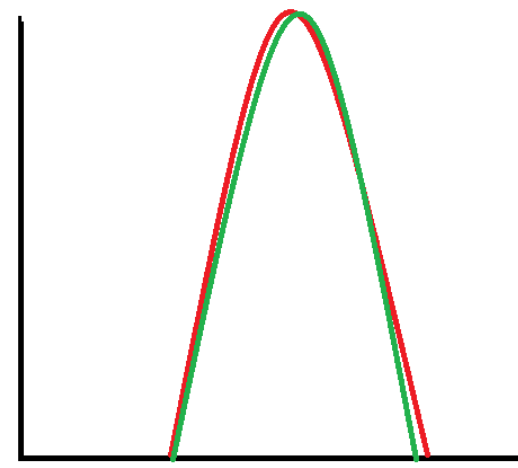
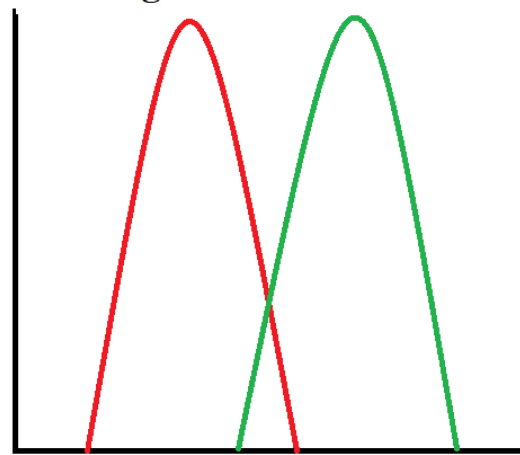
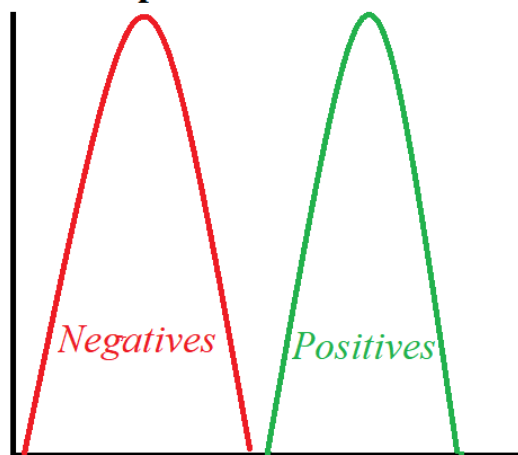
Excellent

Good

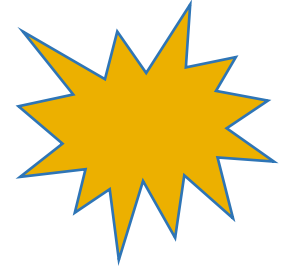
No Separability

Problematic

Overlap = How well the model separates Negatives and Positives



Regresija



- regresija se koristi za predviđanje kontinuiranih vrijednosti
- primjene: predviđanje cijene kuća, predviđanje cijene dionica, predviđanje vremenske prognoze, superrezolucija slika, kompresija slika
- korišteni modeli: linearna regresija, slučajne šume, XGboost, konvolucijske neuronske mreže, recurrent neuronske mreže

Srednja kvadratna pogreška MSE

- mean square error (MSE)
- traži **srednju kvadratnu pogrešku** između predviđene i stvarne vrijednosti

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

RMSE

- **korijen od MSE**
- kako bi mjera pogreške imala istu jedinicu kao i promatrane vrijednosti
- prosječna devijacija u modelu od ciljane vrijednosti

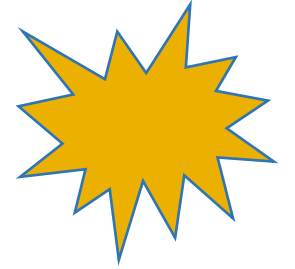
Srednja apsolutna pogreška MAE

- mean absolute value (MAE)
- traži **srednju apsolutnu udaljenost** između predviđene i stvarne vrijednosti

$$MSE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

- robusnija na outliere od MSE – jer MSE kvadrira pogrešku, a outlieri imaju veliku pogrešku, pa se kvadriranjem još povećava

Rangiranje



- rangiranje nekih stavki na temelju njihove relevantnosti u određenom zadatku
- npr. rangiranje stranica na Google na temelju njihove relevantnosti za dani upit, predlaganje filmova (Netflix, Youtube), rangiranje proizvoda (Amazon), automatsko nadopunjavanje upita, pretraživanje slika (Vimea), traženje hotela (Booking)
- algoritmi za rangiranje:
 - **point wise modeli** (točkovni modeli) – pokušavaju predvidjeti (podudarni) rezultat za svaki par upit-dokument u skupu podataka i koriste ga za rangiranje stavki
 - **pair-wise modeli** (modeli u paru) – pokušavaju naučiti binarni klasifikator koji će reći koji je dokument relevantniji za upit, za zadani par dokumenata.
 - **list-wise modeli** (modeli s popisom) – pokušavaju izravno optimizirati vrijednost jedne od navedenih mjera evaluacije, prosječne po svim upitima u podacima za učenje
- za vrijeme evaluacije, uspoređuju se **referentna vrijednost redoslijeda stavki s predviđenim redoslijedom** tih istih stavki

Srednji recipročni rang MRR

- mean reciprocal rank
- **srednja vrijednost recipročnih rangova „prve relevantne stavke“ za dani upit Q**

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rang_i}$$

- **nedostatci:** uzima u obzirom rang samo jedne stavke (najrelevantnije), ignorira sve ostale – nije dobra mjera ukoliko želimo listu relevantnih stavki

Srednji recipročni rang – primjer

- model predviđa množinu riječi na temelju tri ponuđene množine

upit	ponuđeni rezultati	ispravan odgovor	rang	recipročni rang
dijete	dijeca, djeca , dijeta	djeca	2	1/2
podatak	podataki, podatki, podaci	podaci	3	1/3
bor	borovi , bori, bora	borovi	1	1/1

$$MRR = \frac{1}{3} \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{1} \right) = \frac{11}{18}$$

Preciznost pri k ($P@k$)

- broj relevantnih dokumenata između top k dokumenata

$$P@k = \frac{\text{broj predviđenih stavki @k koje su relevantne}}{\text{broj predviđenih stavki @k}}$$

- **primjer:** tražite „dezinficijens“ na Google, na prvoj stranici 8 od 10 linkova su relevantni uz „dezinficijens“, $P@10 = 0.8$
- **preciznost pri k za skup upita Q:** srednja vrijednost $P@k$ za sve upite u Q
- **nedostatci:** ne uzima u obzir poziciju relevantnih dokumenata među top k
- jednostavno ručno evaluirati model – top k rezultata treba provjeriti kako bi se utvrdilo jesu li relevantni ili ne
- postoji i $\text{recall}@k$

Kumulativna dobit CG

- cumulative Gain CG
- **kumulativna dobit skupa dohvaćenih dokumenata** je suma njihovih ocjena relevantnosti na upit, uz pretpostavku da je poznata ocjena relevantnosti za svaki dokument upita

$$CG_p = \sum_{i=1}^p rel_i$$

Težinska kumulativna dobit DCG

- discounted Cumulative Gain (DCG)
- logaritamski redukcijski faktor se koristi za **zmanjenje ocjene relevantnosti proporcionalno poziciji rezultata** – **želimo dati veći prioritet prvih nekoliko stavki nego onima kasnije**

$$DCG_p = \sum_{i=1}^p \frac{rel_i}{\log_2(i + 1)}$$

- Drugi oblik:

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

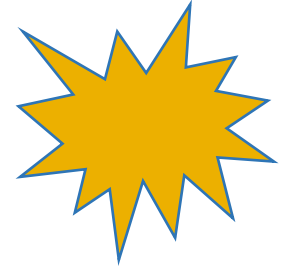
Normalizirana težinska kumulativna dobit NDCG

- unaprijeđenje DCG kako bi mjera bila bolje prilagođena primjeni iz stvarnog svijeta
- **dijeli DCG s DCG idealnog sustava (IDCG)** – jer skup dohvaćenih stavki može varirati u veličini između različitih upita ili sustava
- sortira dokumente rezultatne liste po relevantnosti, pronalazi najveći DCG (dobiven koristeći idealni sustav) za poziciju p i koristi ju za normaliziranje DCG

$$NDCG_p = \frac{DCG_p}{IDCG_p}$$

- **nedostatci:** ne penalizira loše dokumente u rezultatu, možda nije primjenjiva za mjerenje performanse upita koji mogu imati više jednako dobrih rezultata (posebno je ovo istina ukoliko smo zainteresirani za prvih nekoliko rezultata kao što se to događa u praksi)

Literatura



<https://www.v7labs.com/blog/data-labeling-guide>

<https://appen.com/blog/data-labeling/>

<https://www.cloudfactory.com/data-labeling-guide>

<https://labelyourdata.com/articles/data-labeling-quality-and-how-to-measure-it>

<https://towardsdatascience.com/cronbachs-alpha-theory-and-application-in-python-d2915dd63586>

<https://www.statology.org/cronbachs-alpha-in-python/>

<https://hackernoon.com/introduction-5-different-types-of-text-annotation-in-nlp-78523www>

<https://towardsdatascience.com/20-popular-machine-learning-metrics-part-1-classification-regression-evaluation-metrics-1ca3e282a2ce>

<https://towardsdatascience.com/20-popular-machine-learning-metrics-part-2-ranking-statistical-metrics-22c3e5a937b6>