

Analiza Medicinske Slike  
- nedelja 4 -

# Procena učinka ML modela

# Danas ćemo proći

1. Definicija modela za regresiju, klasifikaciju i segmentaciju
2. Uticaj podataka na model
3. Merenje učinka modela
4. Razumevanje učinka ML modela

Nedelja 1	Nedelja 2	Nedelja 3	Nedelja 4	Nedelja 5	Nedelja 6	Nedelja 7	Nedelja 8
Uvod u predmet	Primena ML u analizi medicinske slike	Konvoluciski modeli	Procena učinka ML modela				Rad na projektima

“All models are wrong, but some are useful.” G. Box

“All models are wrong, but some are useful.” G. Box

Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

# “All models are wrong, but some are useful.” G. Box

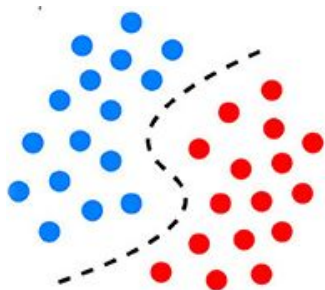
Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.

# “All models are wrong, but some are useful.” G. Box

Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja**.\*

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.



## Nadgledano

Podaci: Označeni (*Labeled*)

Svrha: Regresija, klasifikacija...

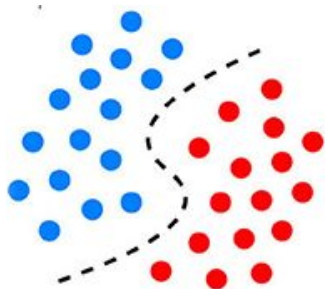
Primer korišćenja: klasifikacija slika  
(npr. ćelija raka), filtriranje spam  
mejlova,...

# “All models are wrong, but some are useful.” G. Box

Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.

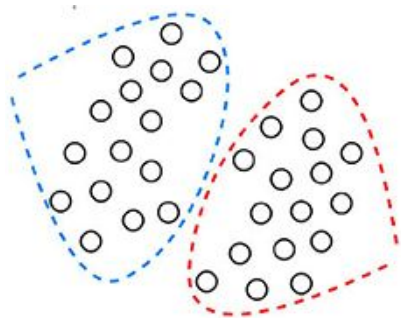
**Nadgledano**



# “All models are wrong, but some are useful.” G. Box

Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja**.\*

Sušтина mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.



## Nenadgledano

Podaci: Neoznačeni (*Unlabeled*)

Svrha: Smanjenje dimenzionalnosti,  
grupisanje, izdvajanje obeležja...

Primer korišćenja: preporuka  
reklama, detekcija anomalija  
(uklanjanje šuma)...

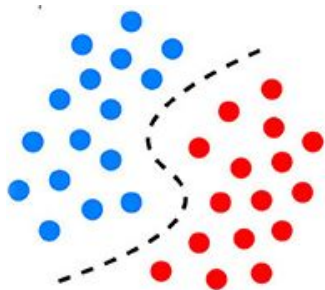


# “All models are wrong, but some are useful.” G. Box

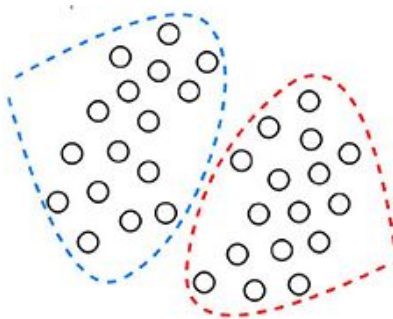
Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.

**Nadgledano**



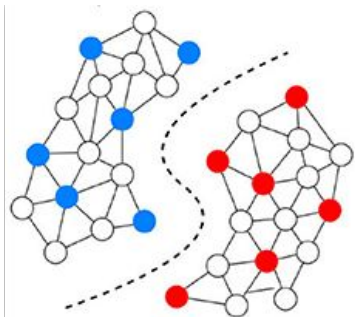
**Nenadgledano**



# “All models are wrong, but some are useful.” G. Box

Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.



## Polunadgledano

Podaci: Označeni i neoznačeni

Svrha: Traženje granica parametara  
označavanje neoznačenih...

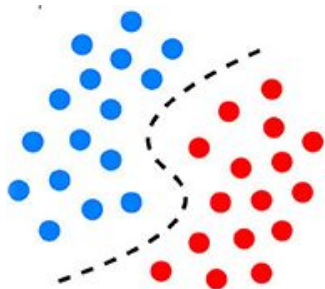
Primer korišćenja: analiza govora,  
sekvenciranje u genetici...

# “All models are wrong, but some are useful.” G. Box

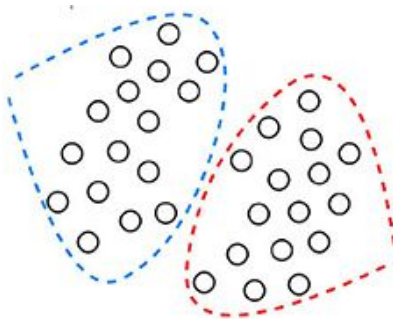
Mašinsko učenje daje računarima sposobnost da **uče bez eksplicitnog programiranja.\***

Suština mašinskog učenja je da koristi podatke da izgradi model koji će vršiti predikcije.

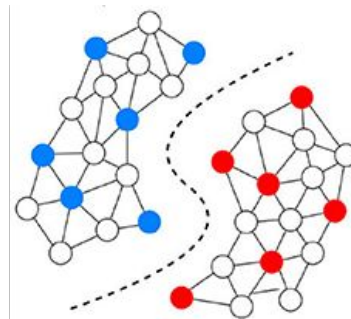
**Nadgledano**



**Nenadgledano**



**Polunadgledano**



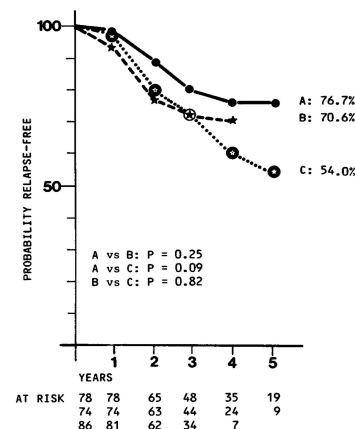
Za šta koristimo ML modele

# Za šta koristimo ML modele



Regresija

Utvrđivanje veze (fit) između dve (ili više) promenljive.



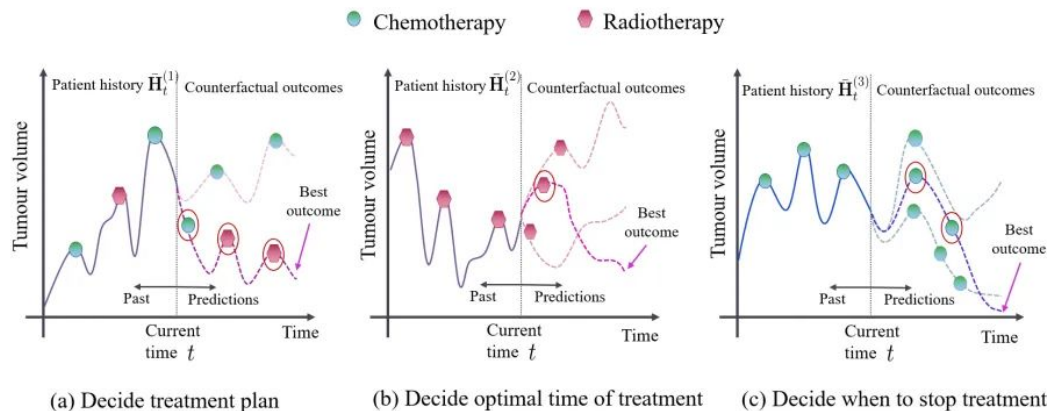
Bonadonna et al. 1981, [Dose-Response Effect of Adjuvant Chemotherapy in Breast Cancer](#)

# Za šta koristimo ML modele



Regresija

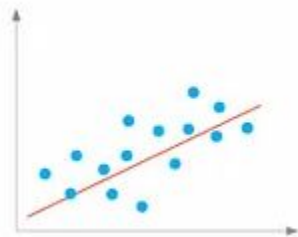
Predikcija vremenske serije



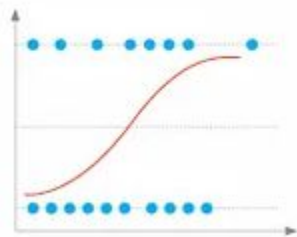
# Za šta koristimo ML modele



Regresija



Linearna



Logistička



Polinomijalna



Vremenska serija

[What Is Regression in Machine Learning?](#)

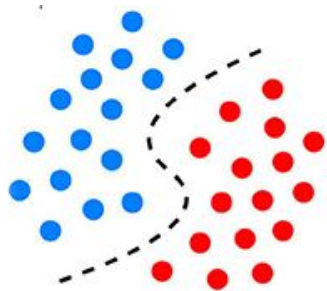
# Za šta koristimo ML modele



Regresija

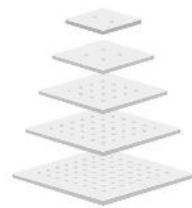


# Za šta koristimo ML modele

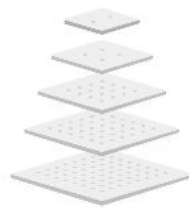


Klasifikacija

Full CT volume model



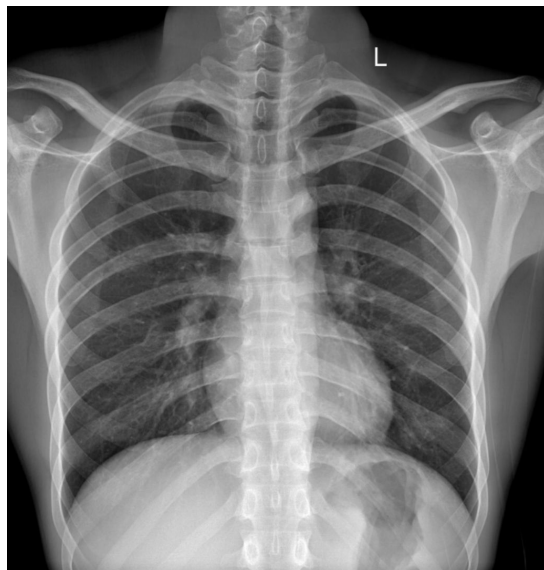
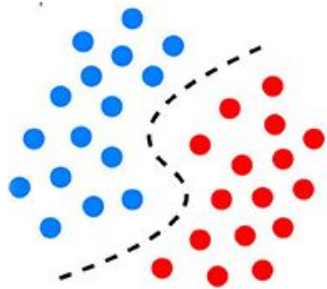
Malignant lesion detection



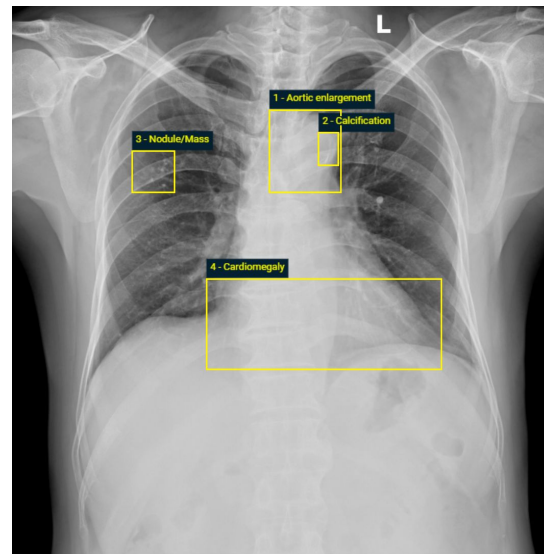
Izvor: [A promising step forward for predicting lung cancer \(blog.google\)](https://blog.google/health/ai/google-deepmind-ai-for-lung-cancer/)

# Za šta koristimo ML modele

Binarna



No finding



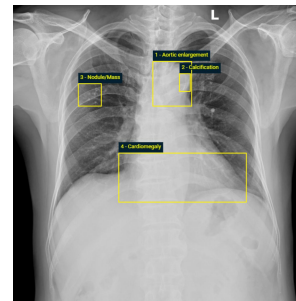
Tuberculosis

# Za šta koristimo ML modele

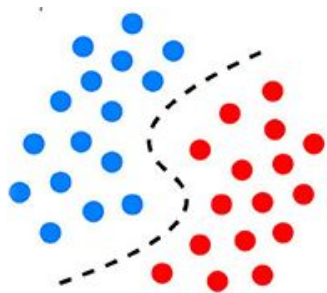
Binarna



No finding



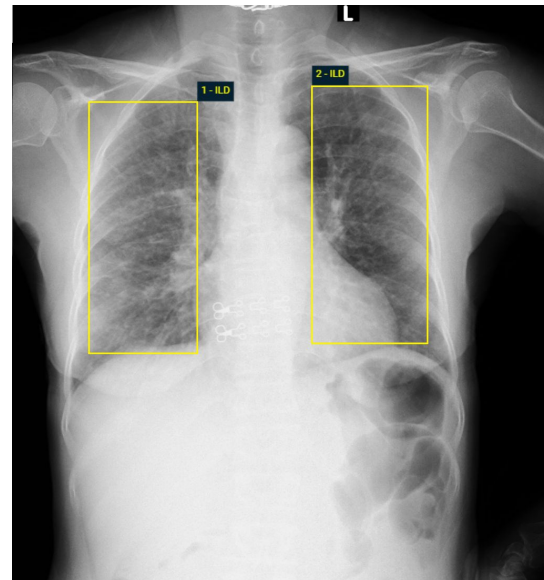
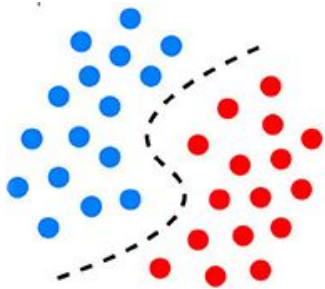
Tuberculosis



Klasifikacija

# Za šta koristimo ML modele

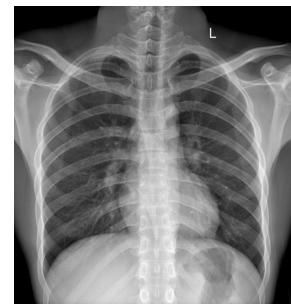
Višeklasna



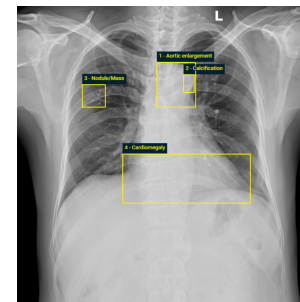
Pneumonia

# Za šta koristimo ML modele

Binarna

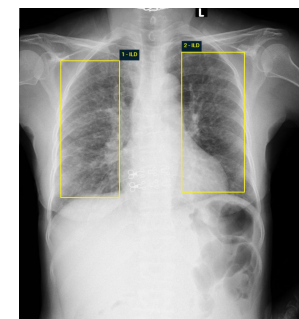


No finding

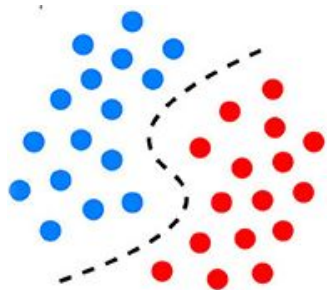


Tuberculosis

Višeklasna



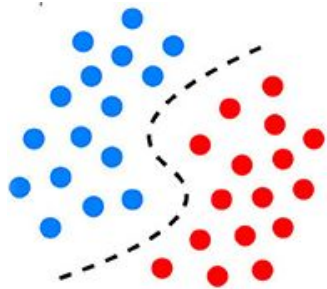
Pneumonia



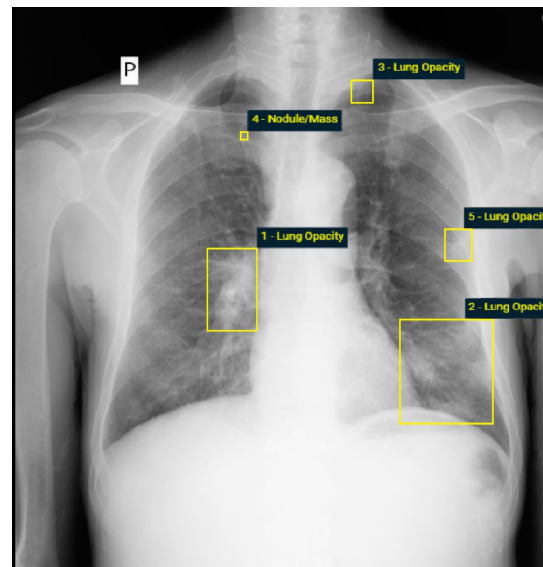
Klasifikacija

# Za šta koristimo ML modele

Višeznačna



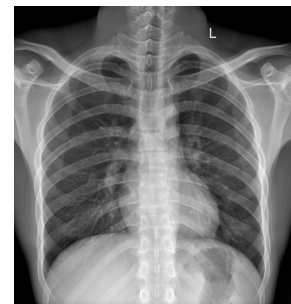
Klasifikacija



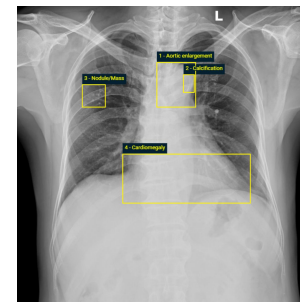
Pneumonia, Lung tumor

# Za šta koristimo ML modele

Binarna

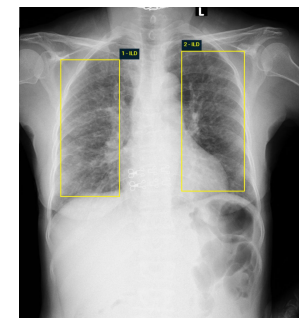


No finding



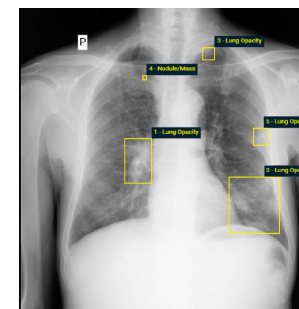
Tuberculosis

Višeklasna

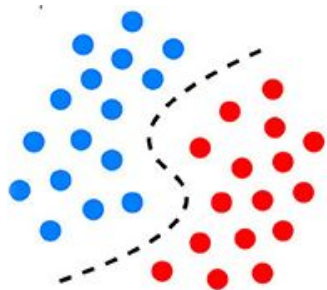


Pneumonia

Višeznačna

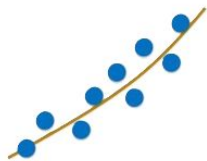


Pneumonia, Lung tumor

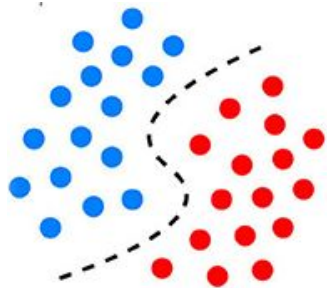


Klasifikacija

# Za šta koristimo ML modele



Regresija

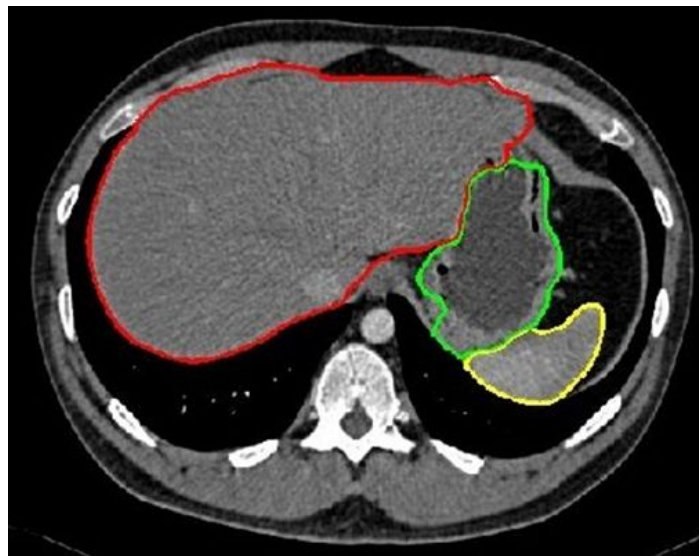


Klasifikacija



# Za šta koristimo ML modele

Segmentacija



# Za šta koristimo ML modele



(a) Image



(b) Semantic segmentation



(c) Instance segmentation



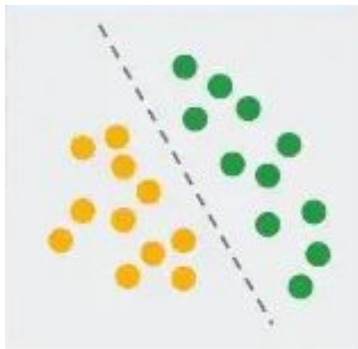
(d) Panoptic segmentation

# Za šta koristimo ML modele

Gen AI

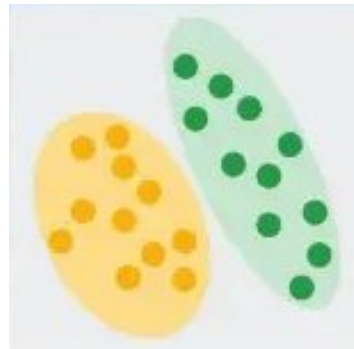
# Tipovi Modela

## Diskriminativni



Klasifikacija  
Predikcija...

## Generativni

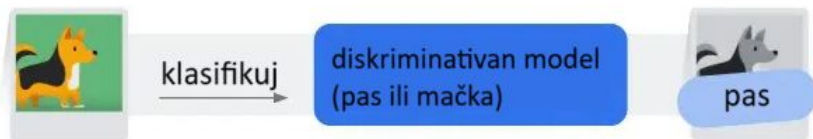


Generisanje novih  
podataka na osnovu datih

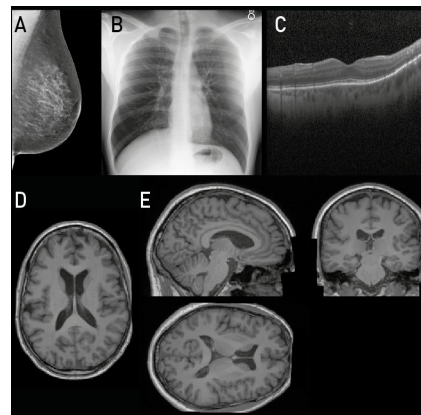
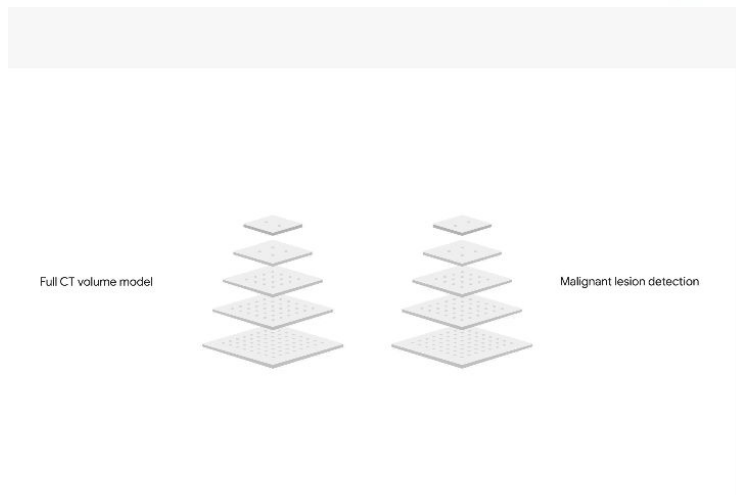
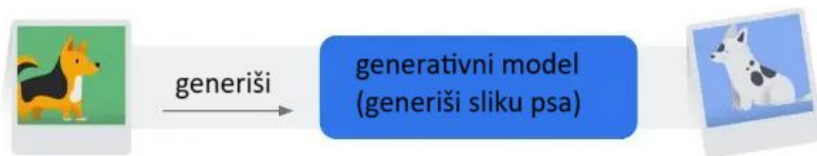
Izvor: [Introduction to Generative AI](#)

# Za šta koristimo ML modele

## Diskriminativni



## Generativni



Izvor: [A promising step forward for predicting lung cancer \(blog.google\)](https://blog.google/technology/ai/google-deepmind-disco-ai/)

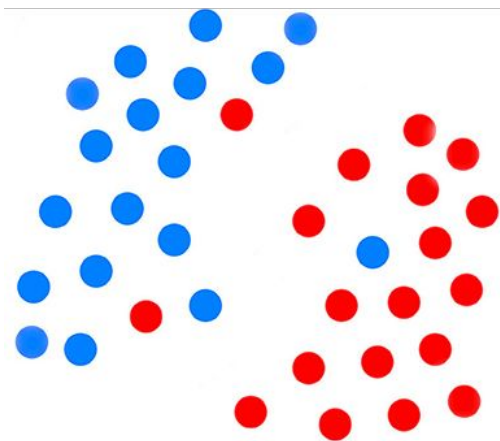
# Primer Generativnog Modela



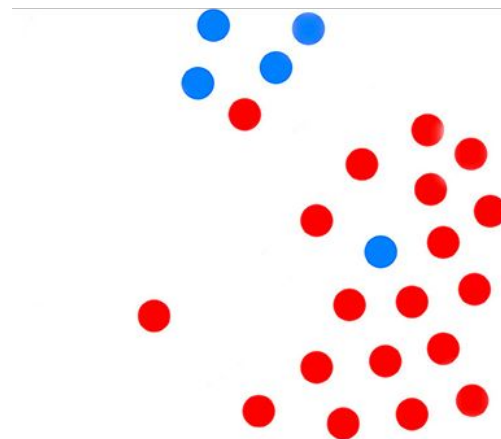
Nasumično izabrani sintetički snimci treniranog LDM (na MIMIC-CXR bazom). Nakon 200 DDIM koraka. Izvor: [Pinaya et al. 2023. Generative AI for Medical Imaging: extending the MONAI Framework](#)

# Koliko verno podaci prikazuju realnost?

Realna slika podataka



Podaci koje posedujemo

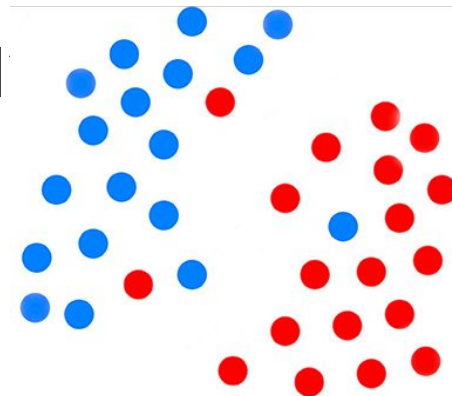


Postoji "pristrasnost" (*bias*) u podacima koja se može preneti na model.

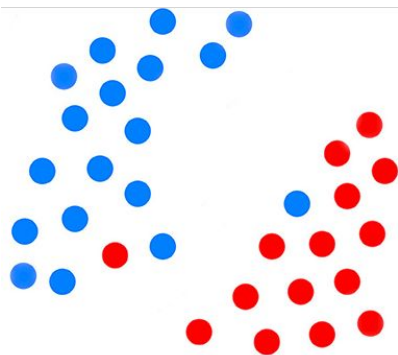
# Kako podaci utiču na model

Podela na train i test

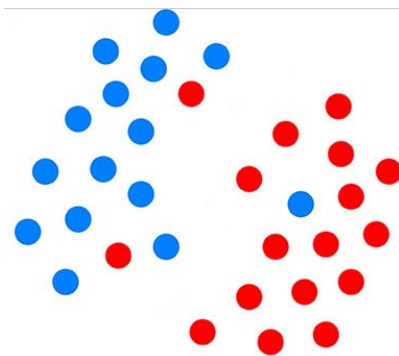
Koji train skup bi bio najbolji izbor i zašto?



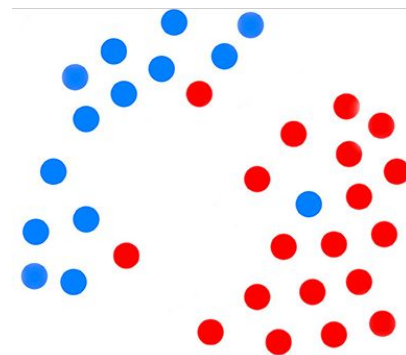
A



B



C





# Mere sličnosti

# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

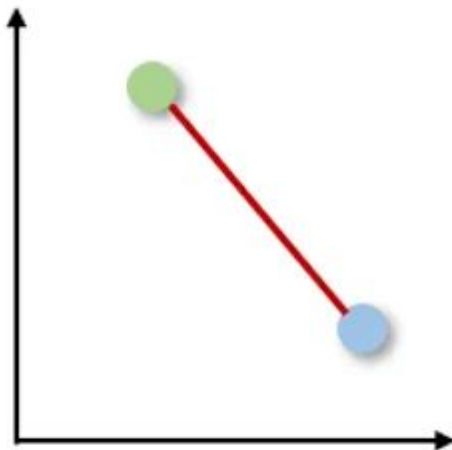
# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

**Euklidsko**  
(*Euclidean*)



$$d = \sqrt{\sum_i (x_i - y_i)^2}$$

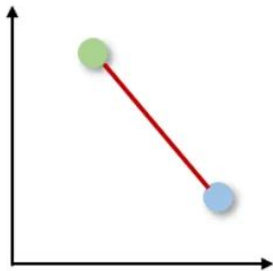
# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

**Euklidsko**  
(*Euclidean*)



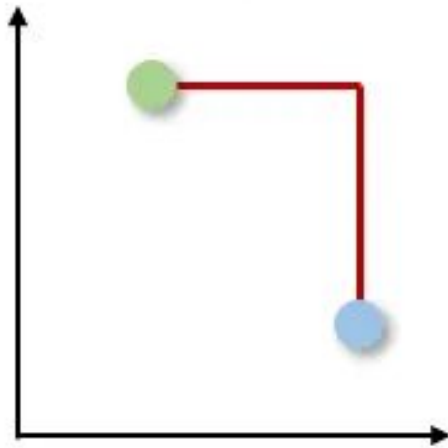
# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

**Menhetn**  
(*Manhattan*)



$$d = \sum_i |x_i - y_i|$$

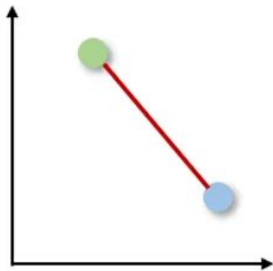
# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

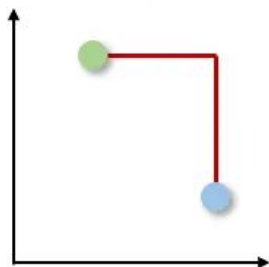
Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

**Euklidsko**  
(*Euclidean*)



**Menhetn**  
(*Manhattan*)





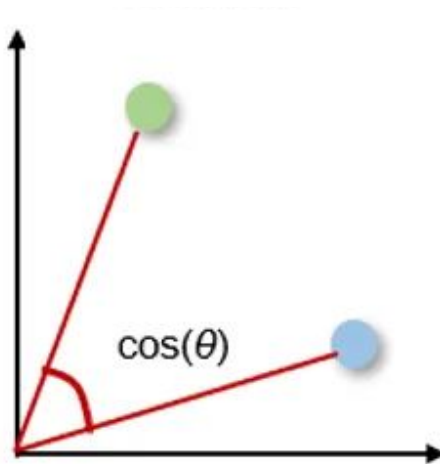
# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

**Kosinus**  
(cosine)



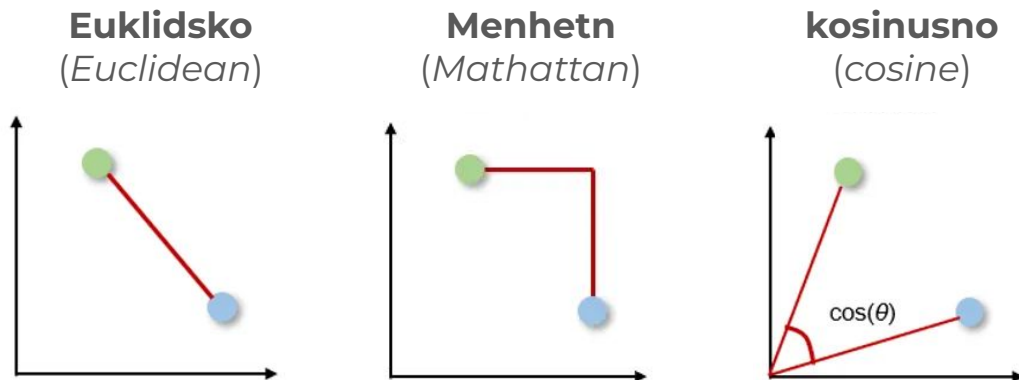
$$d = \cos(\theta) = \frac{xy}{\|x\|_2 \|y\|_2}$$

# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:

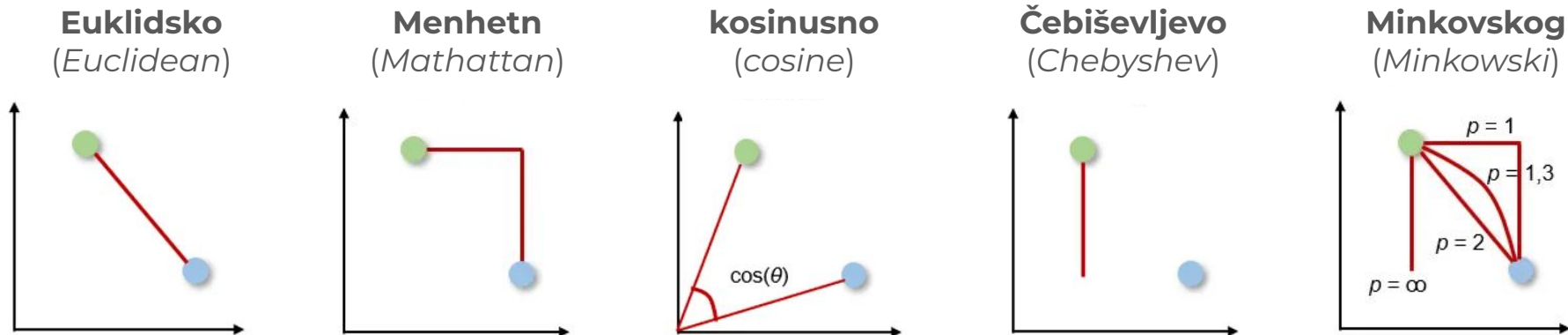


# Mere sličnosti

Pretpostavka: Elementi iste klase bi trebalo da dele **slične** karakteristike (obeležja).

Šta je sličnost? Jedna od često korišćenih mera sličnosti je zapravo rastojanje između obeležja.

Zavisno od vrste podataka možemo koristiti različite funkcije rastojanja:



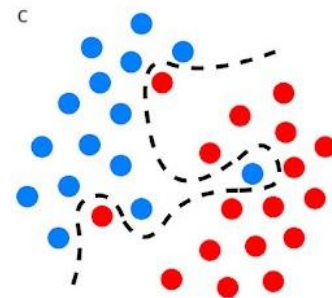
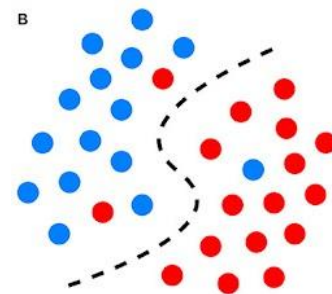
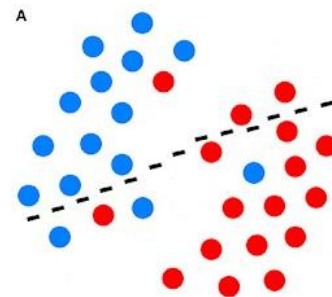
# Klasifikacija

Da bismo napravili dobar klasifikator, moramo da definišemo šta znači “dobar klasifikator”.

# Klasifikacija

Da bismo napravili dobar klasifikator, moramo da definišemo šta znači “dobar klasifikator”.

Da li je dobra klasifikaciona podela pod A, B ili C?



# Klasifikacija

Da bismo napravili dobar klasifikator, moramo da definišemo šta znači “dobar klasifikator”.

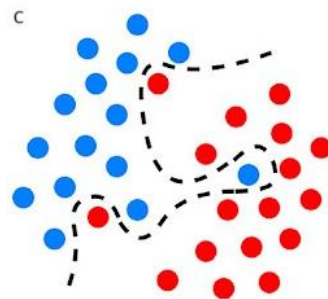
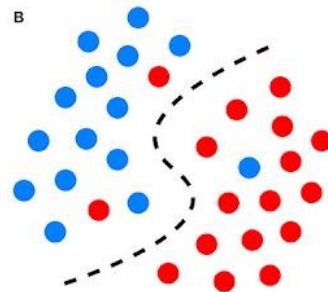
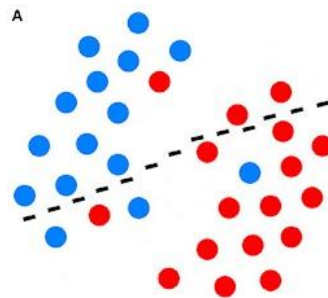
Da li je dobra klasifikaciona podela pod A, B ili C?

Koraci:

- Podeliti bazu podataka koju imamo na trening i test skup\*

- Odrediti metriku za merenje performansi klasifikatora

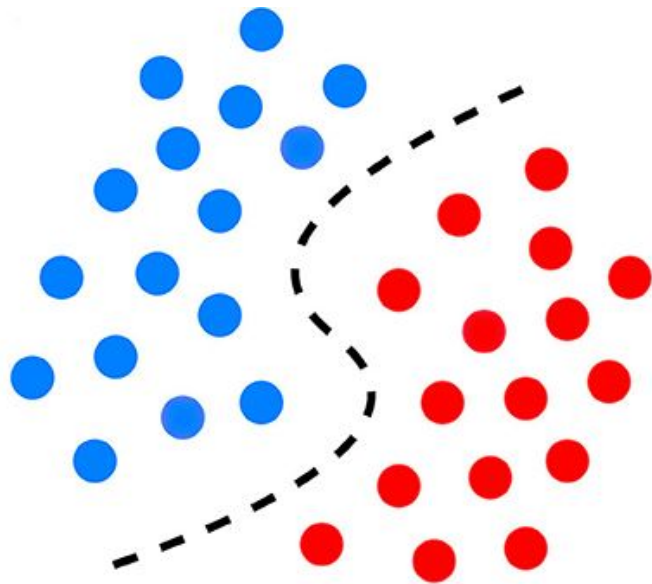
- Pitanje: Da li je bitnije da savršeno naučimo trening podatke ili da model bude sposoban da generalizuje tako da može da predviđa elemente iz test skupa?



# Performanse klasifikacije

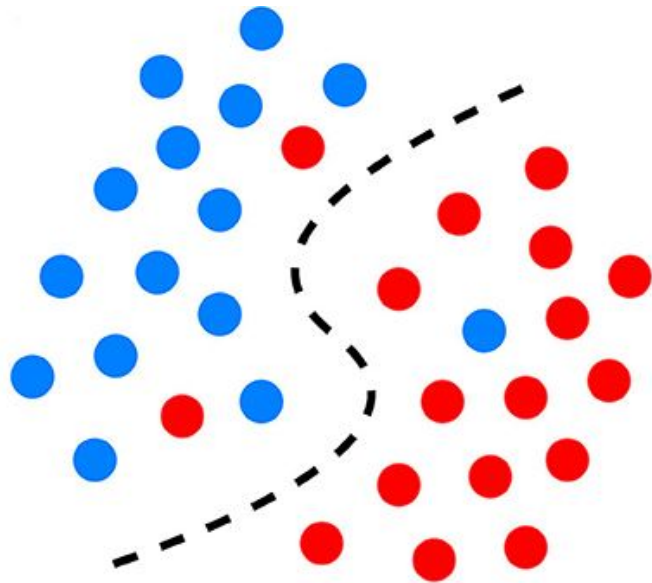


# Performanse klasifikacije

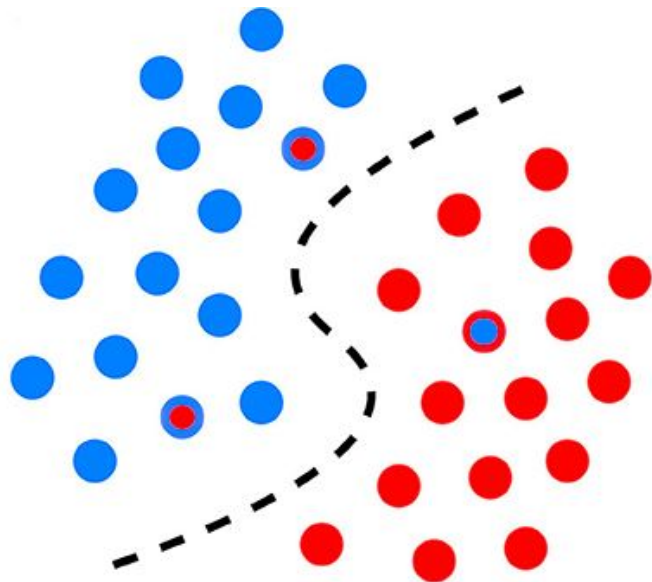




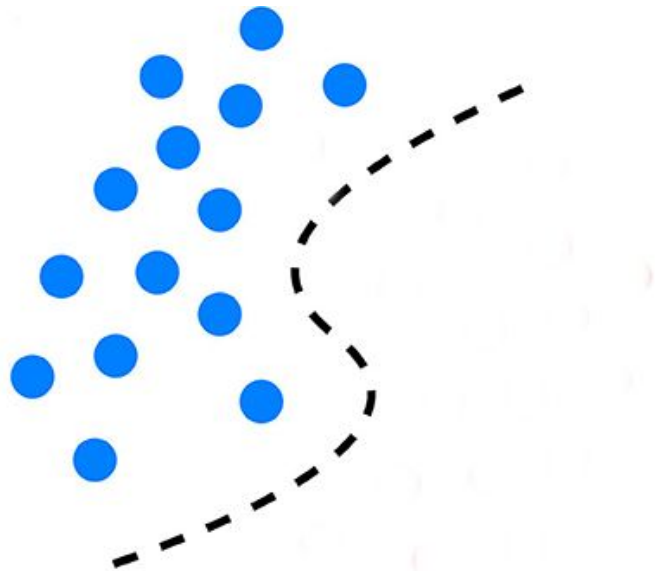
# Performanse klasifikacije



# Performanse klasifikacije



# Performanse klasifikacije



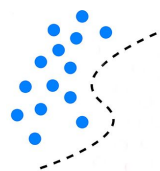
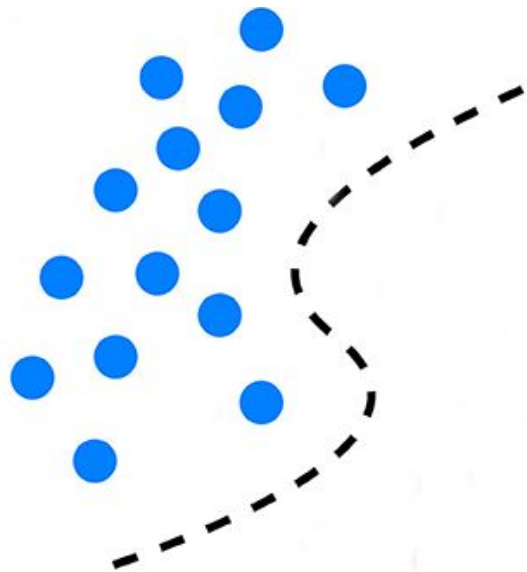
Istinski pozitivno

Lažno pozitivno

Lažno negativno

Istinski negativno

# Performanse klasifikacije



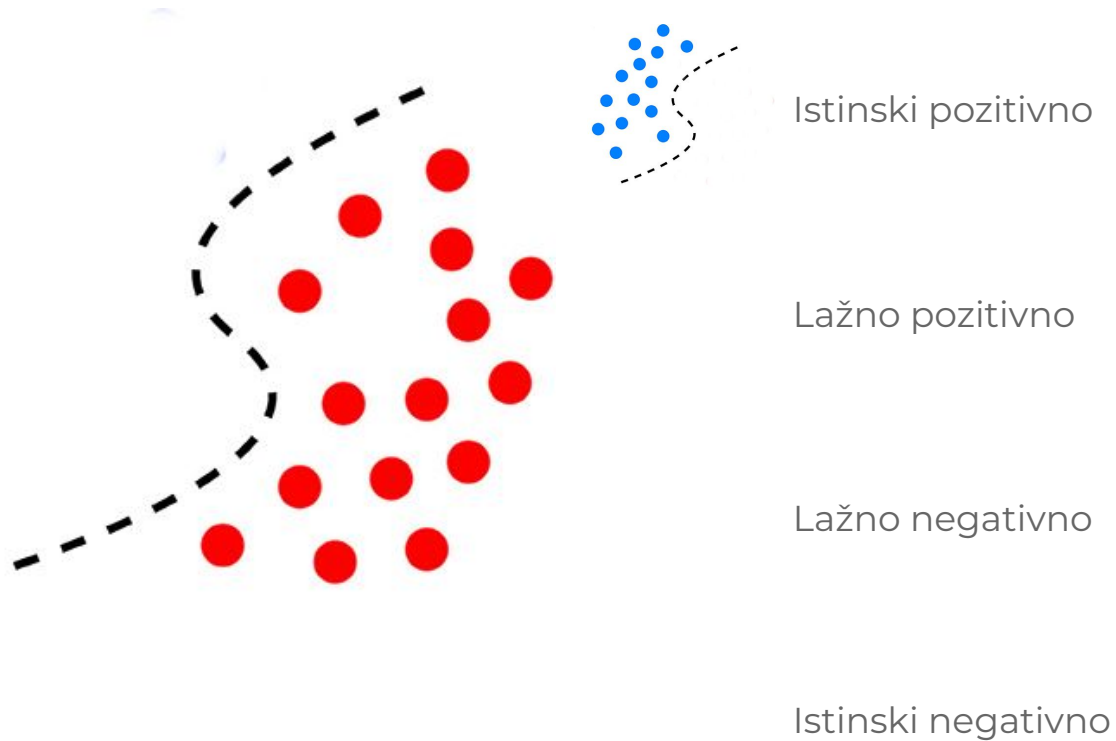
Istinski pozitivno

Lažno pozitivno

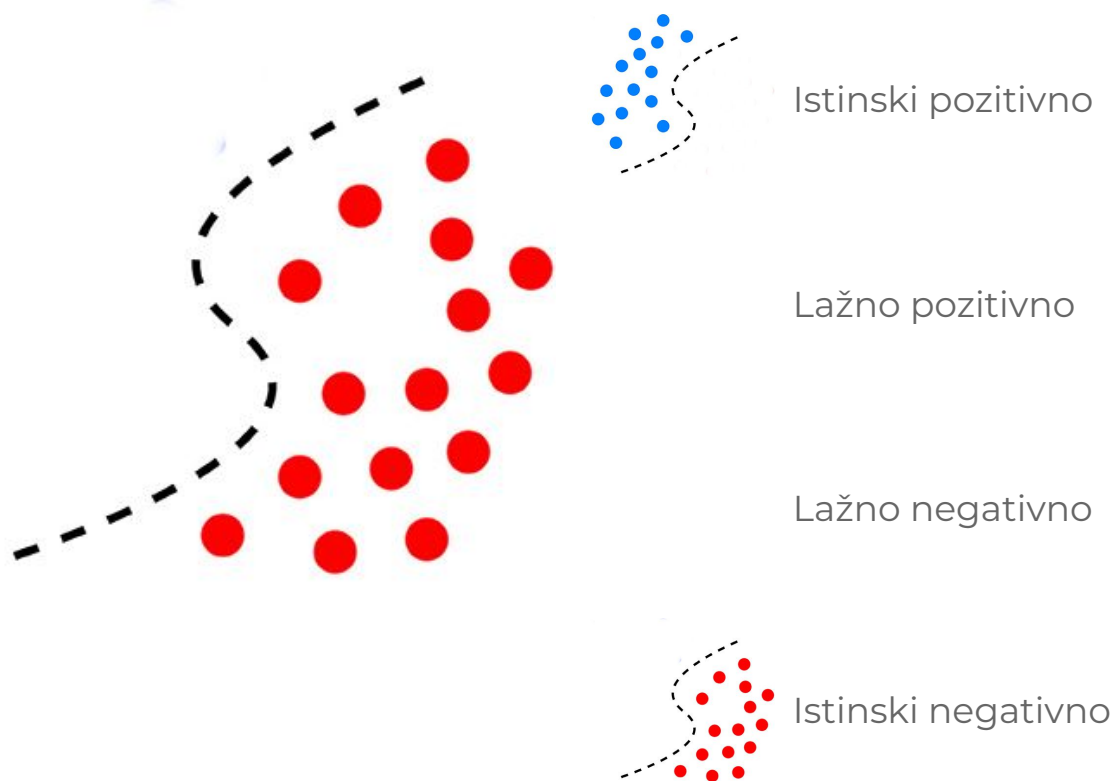
Lažno negativno

Istinski negativno

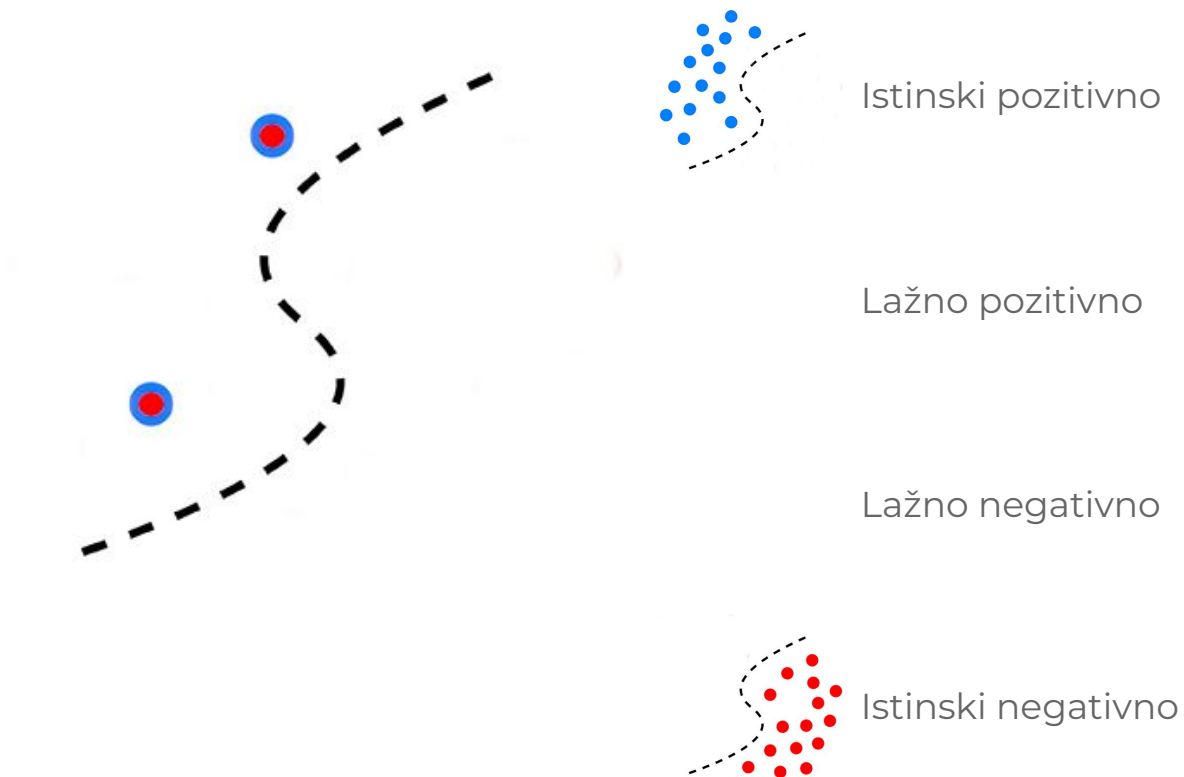
# Performanse klasifikacije



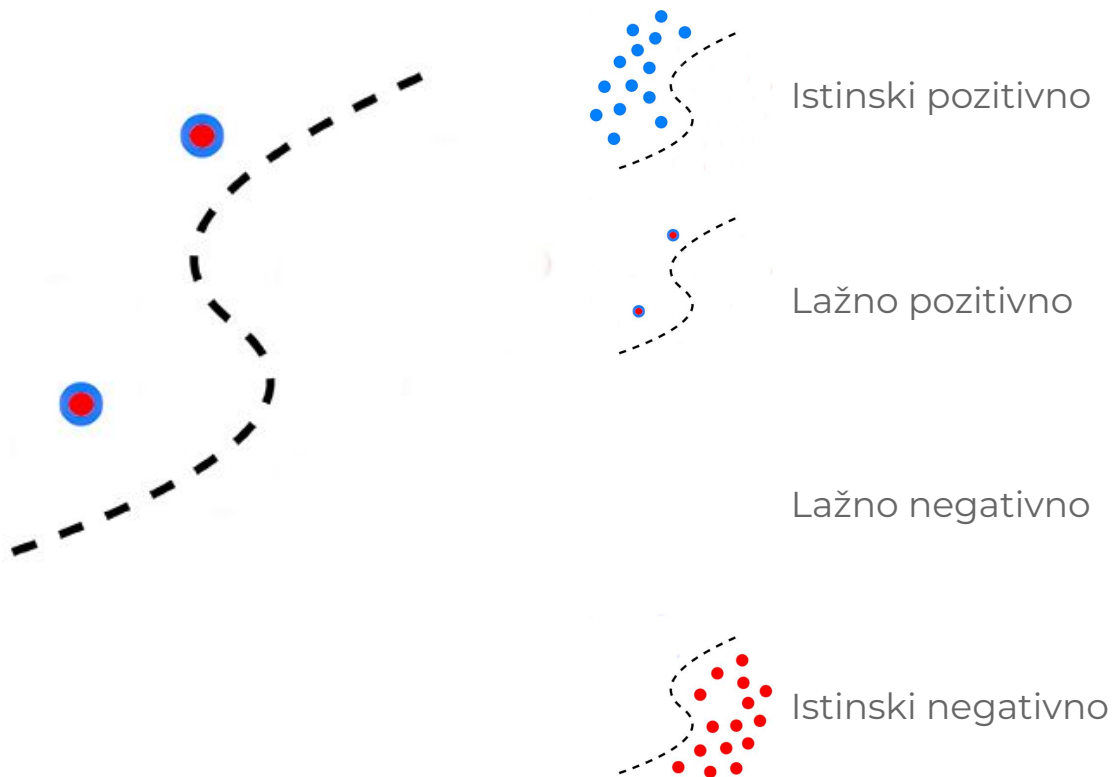
# Performanse klasifikacije



# Performanse klasifikacije

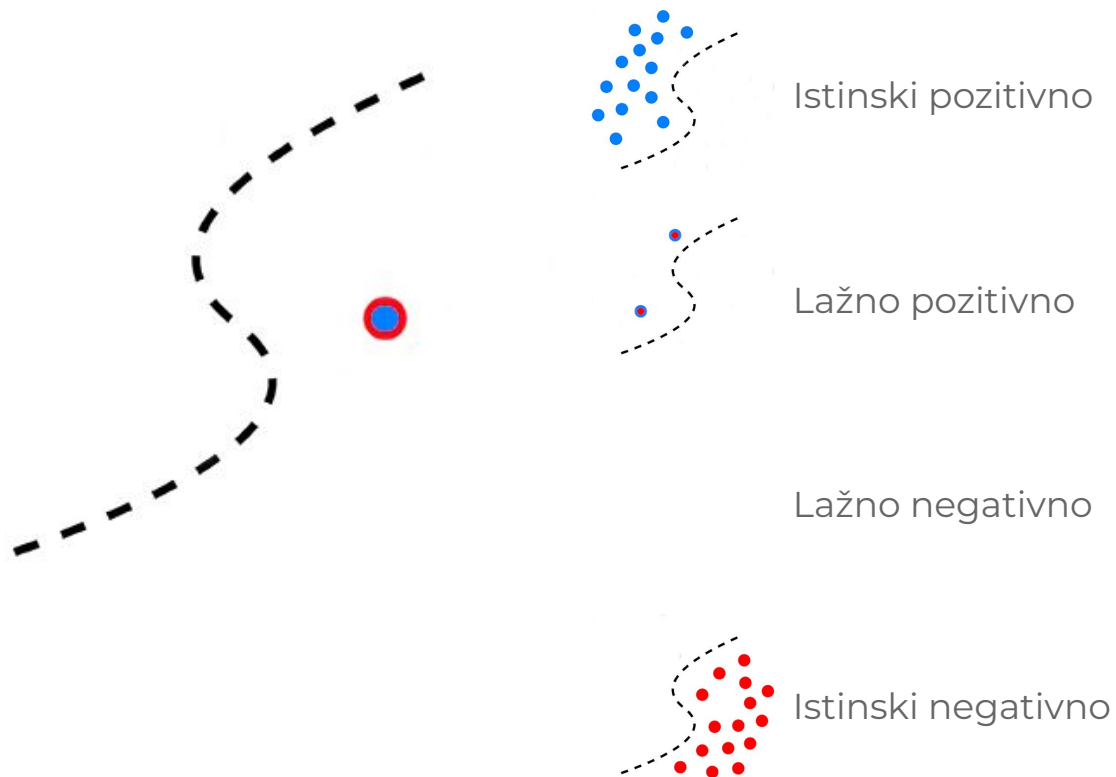


# Performanse klasifikacije

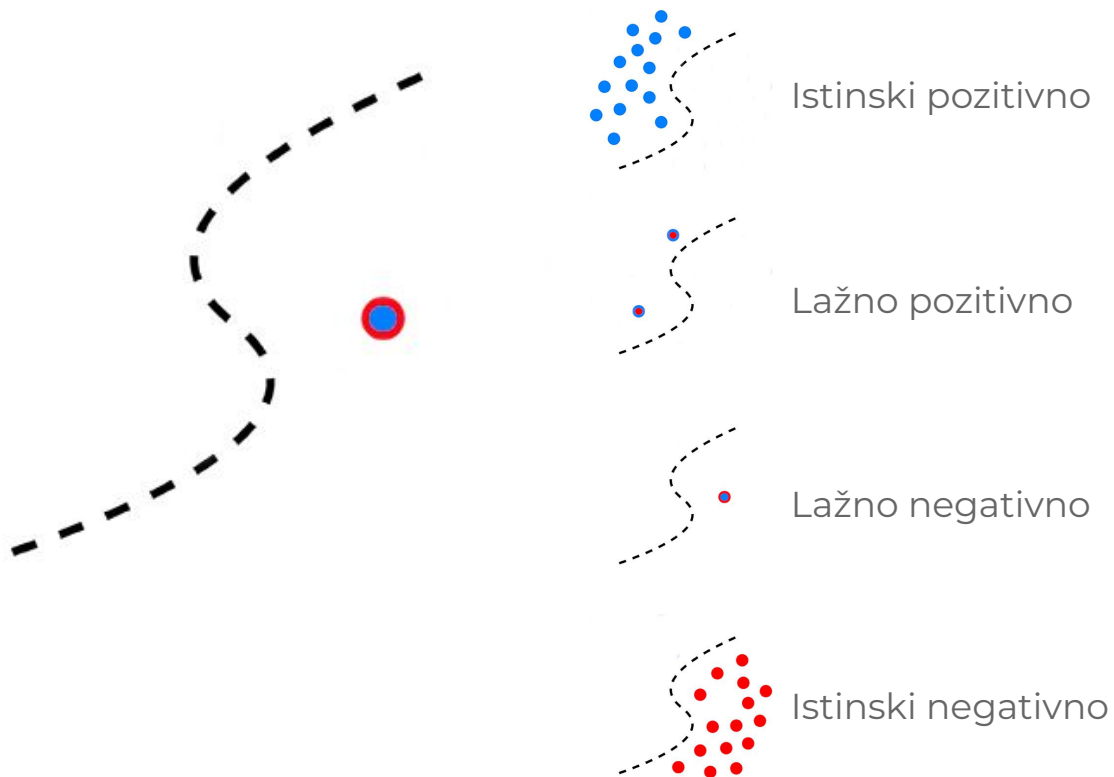




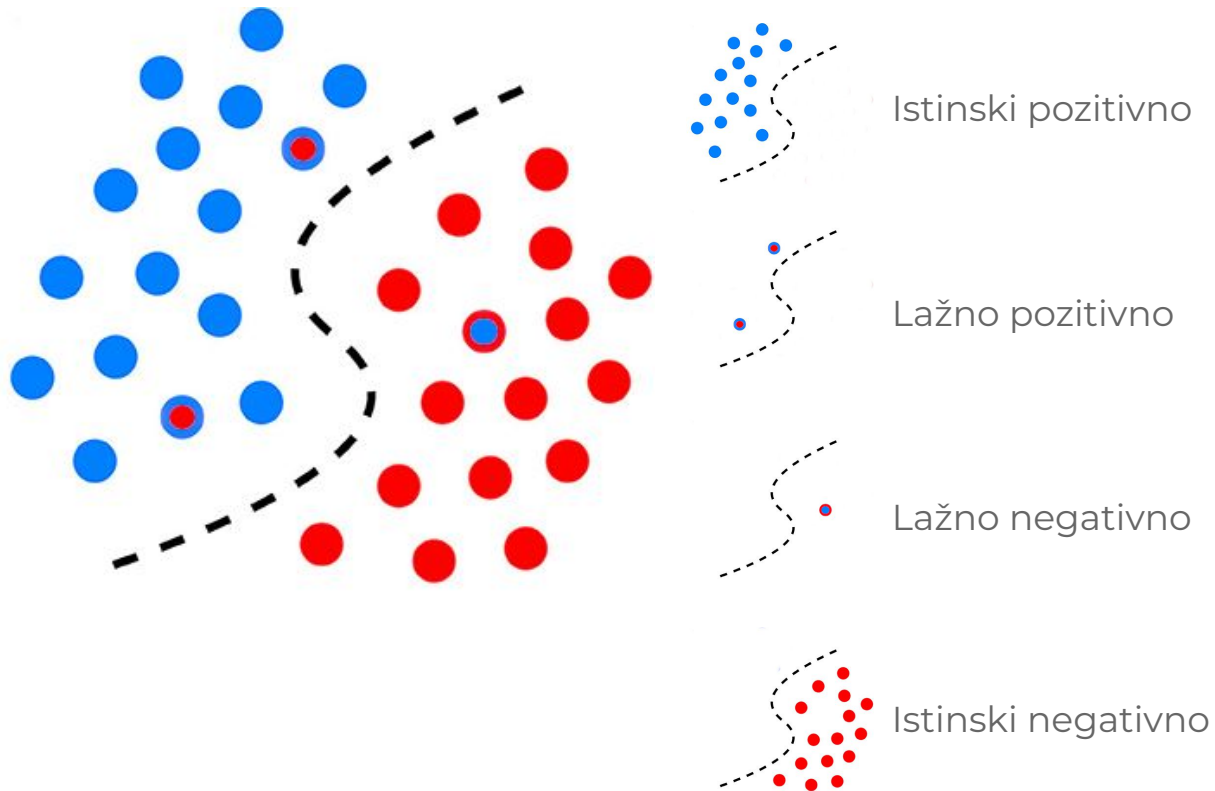
# Performanse klasifikacije



# Performanse klasifikacije

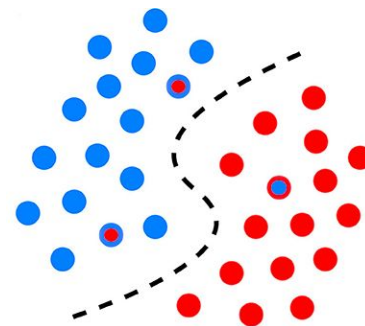
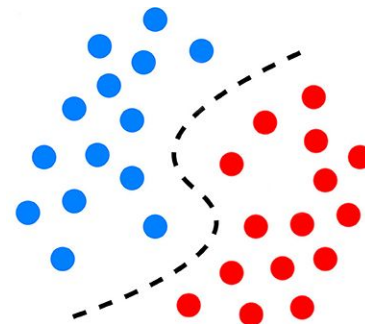


# Performanse klasifikacije

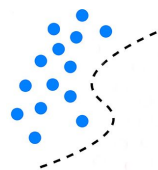
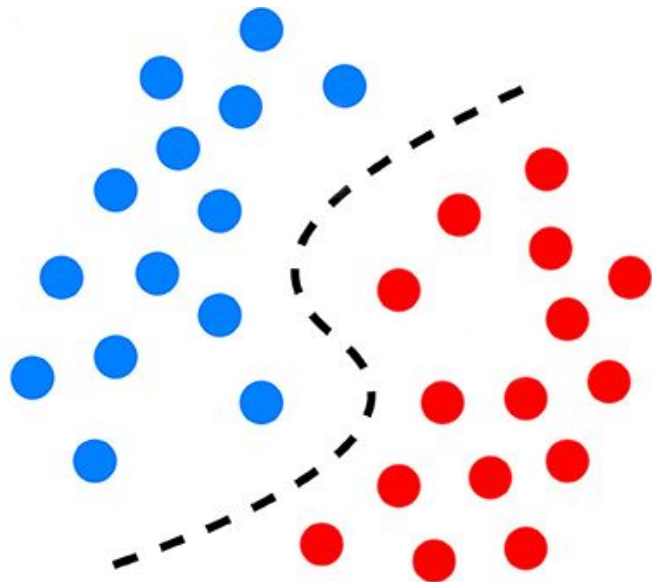


# Performanse klasifikacije

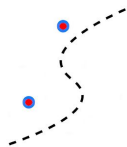
$$\text{Tačnost} = \frac{\text{Istinski pozitivno} + \text{Istinski negativno}}{\text{Istinski pozitivno} + \text{Istinski negativno} + \text{Lažno pozitivno} + \text{Lažno negativno}} =$$



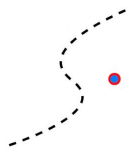
# Performanse klasifikacije



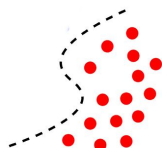
Istinski pozitivno



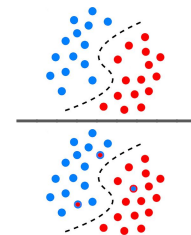
Lažno pozitivno



Lažno negativno



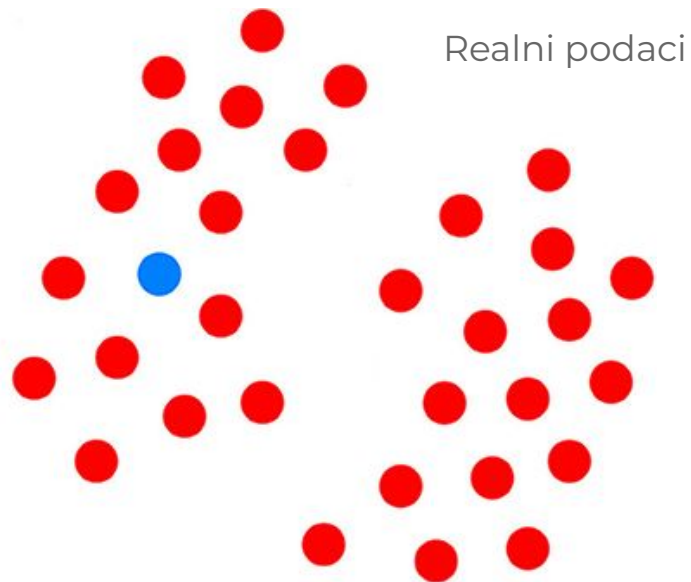
Istinski negativno



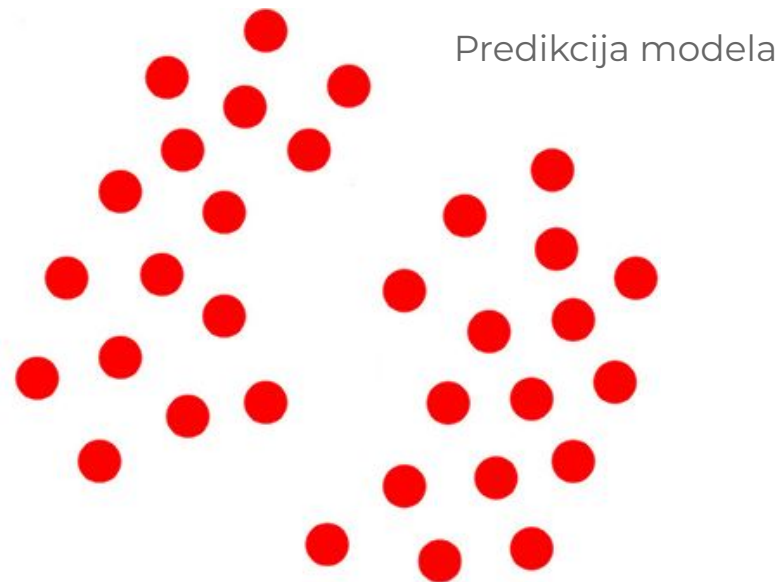
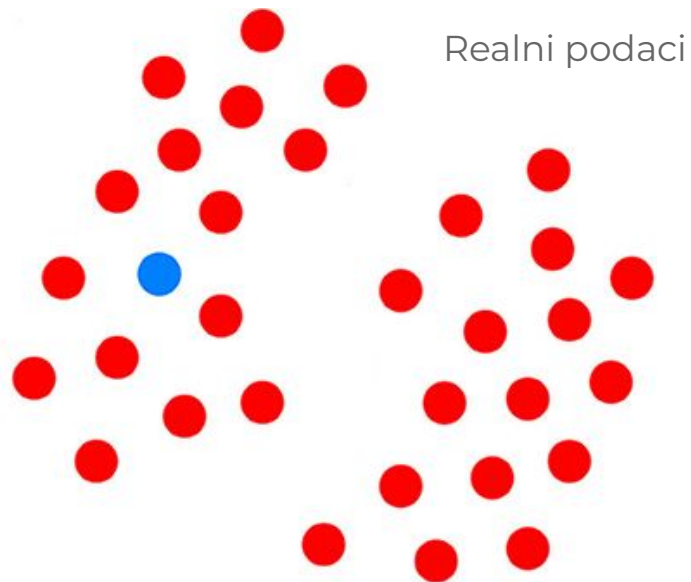
Tačnost

# Performanse klasifikacije - Nebalansirane klase

# Performanse klasifikacije - Nebalansirane klase

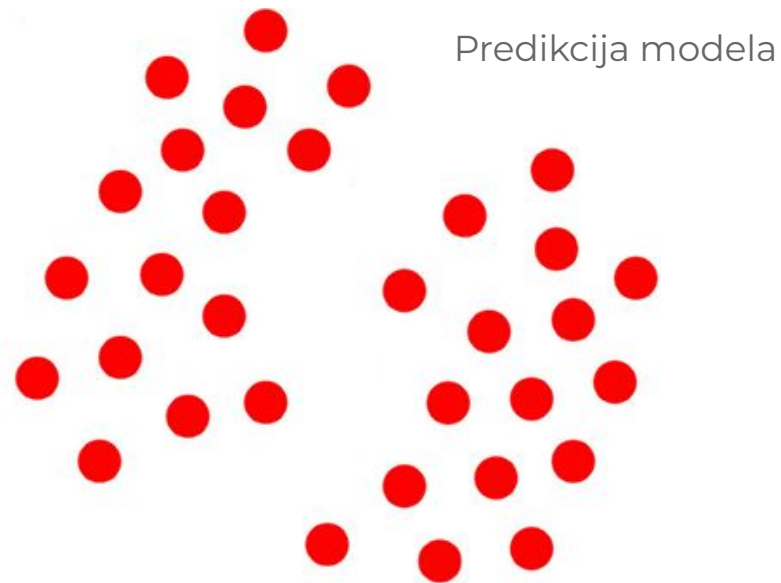
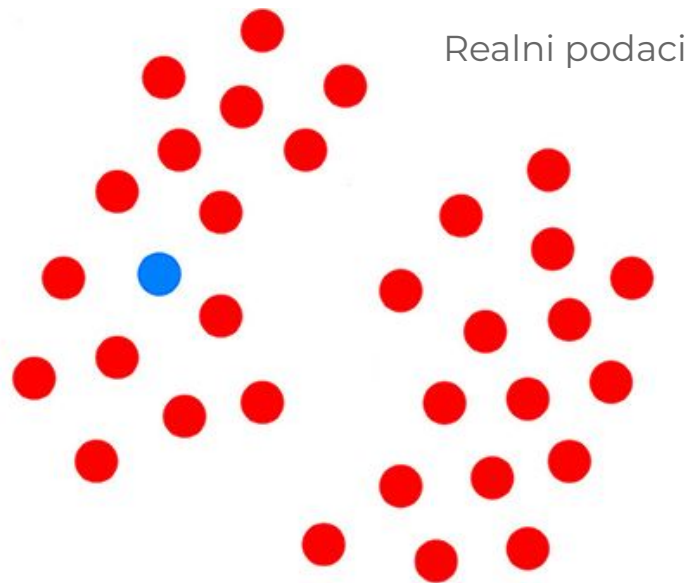


# Performanse klasifikacije - Nebalansirane klase



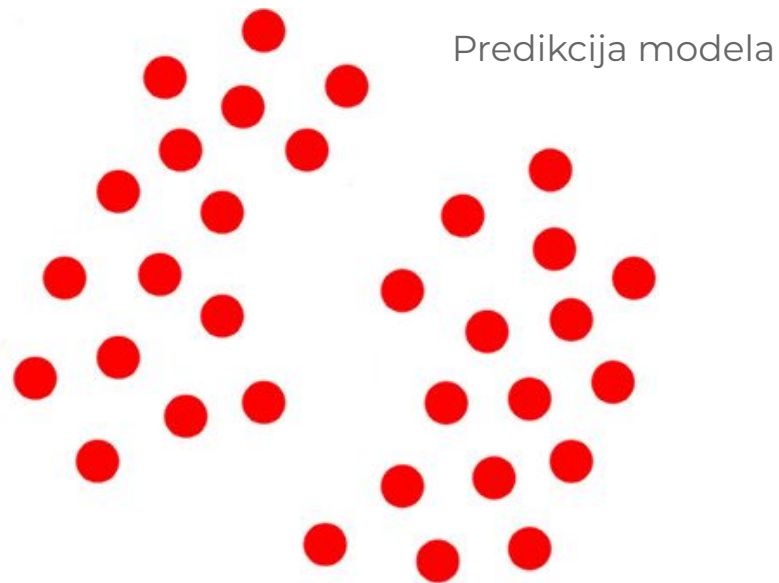
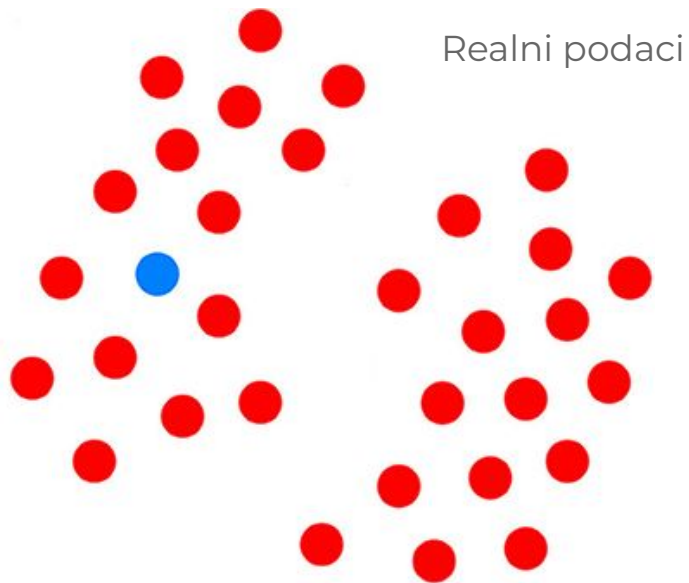


# Performanse klasifikacije - Nebalansirane klase



Tačnost = 96.88%

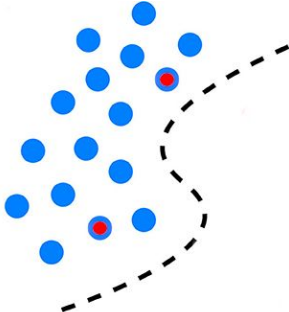
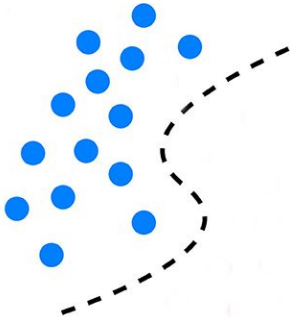
# Performanse klasifikacije - Nebalansirane klase



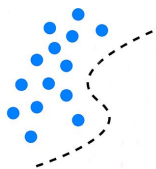
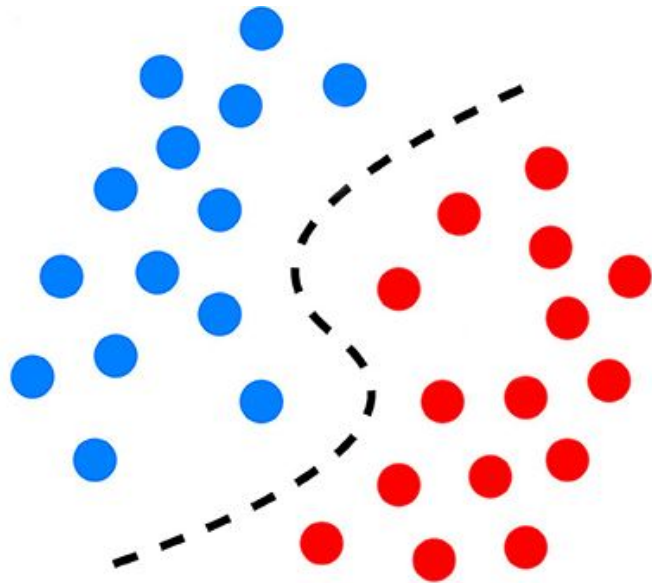
Tačnost = 96.88%

Pogođeni elementi plave klase 0%

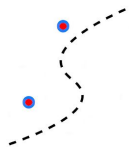
# Performanse klasifikacije

$$\text{Preciznost} = \frac{\text{Istinski pozitivno}}{\text{Istinski pozitivno} + \text{Lažno pozitivno}} = \frac{\text{Diagram 1}}{\text{Diagram 2}}$$


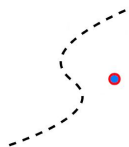
# Performanse klasifikacije



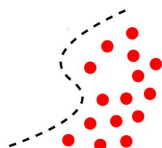
Istinski pozitivno



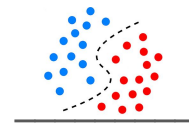
Lažno pozitivno



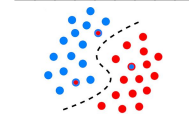
Lažno negativno



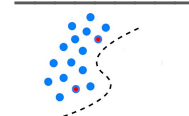
Istinski negativno



Tačnost



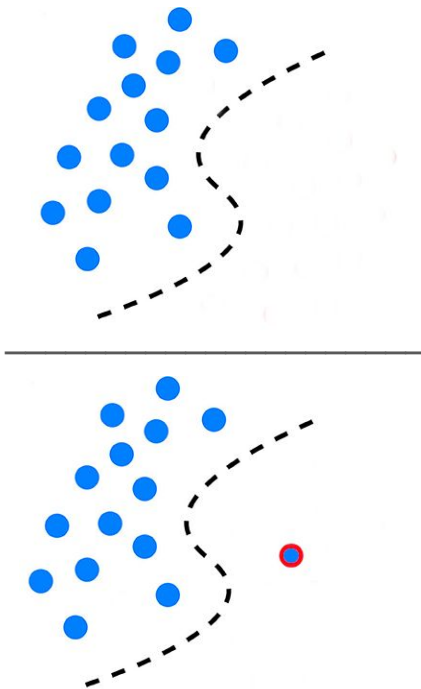
Preciznost



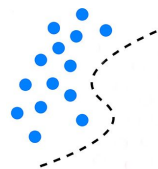
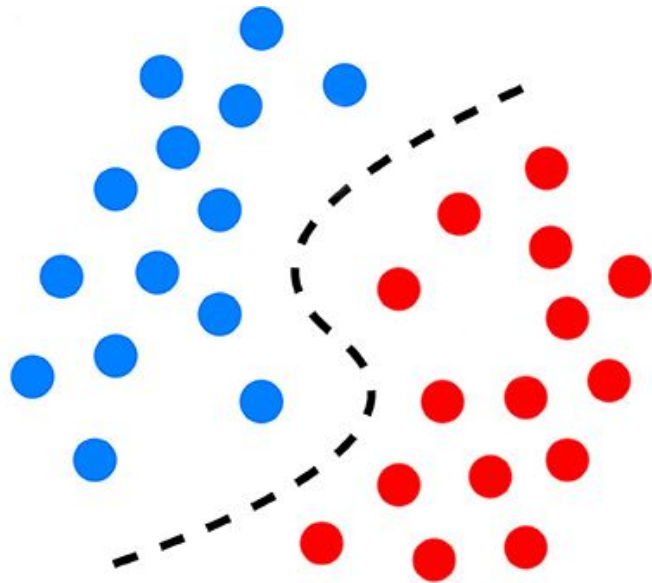
# Performanse klasifikacije

$$\text{Odziv} = \frac{\text{Istinski pozitivno}}{\text{Istinski pozitivno} + \text{Lažno negativno}}$$

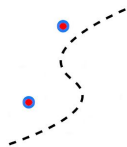
=



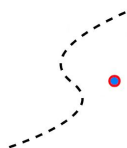
# Performanse klasifikacije



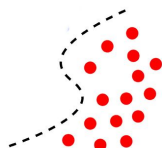
Istinski pozitivno



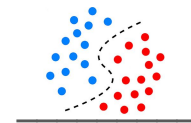
Lažno pozitivno



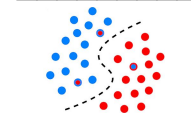
Lažno negativno



Istinski negativno



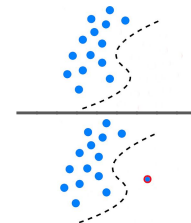
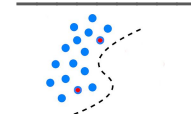
Tačnost



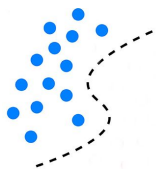
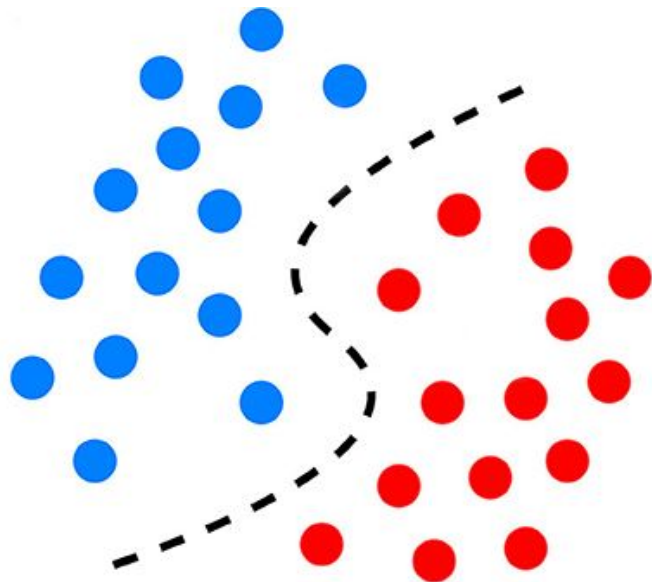
Preciznost



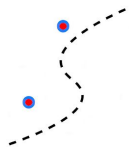
Odziv



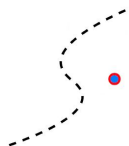
# Performanse klasifikacije



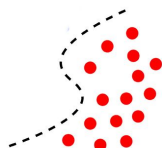
Istinski pozitivno



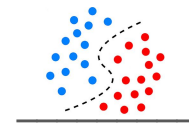
Lažno pozitivno



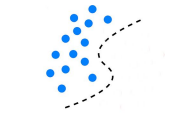
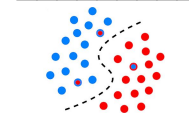
Lažno negativno



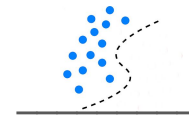
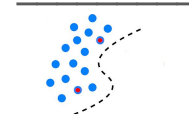
Istinski negativno



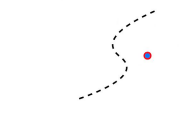
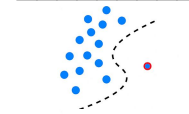
Tačnost



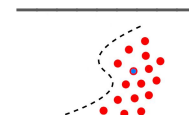
Preciznost



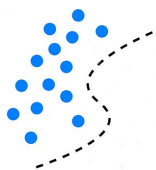
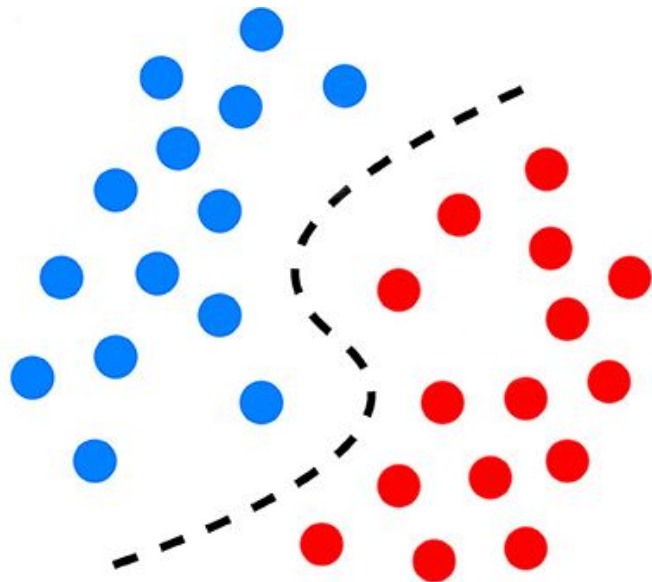
Odziv



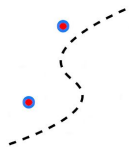
Mera lažno pozitivnih



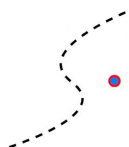
# Performanse klasifikacije



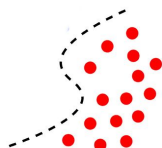
True positive



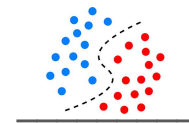
False positive



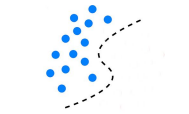
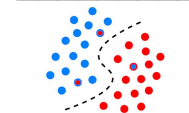
False negative



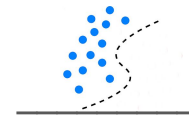
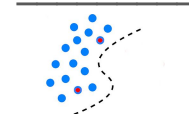
True negative



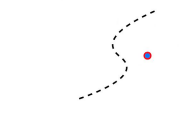
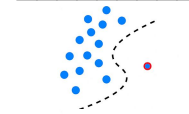
Accuracy



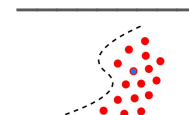
Precision



Recall



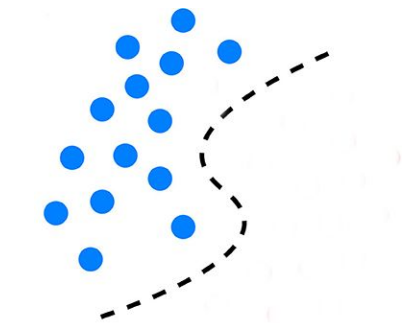
False negative rate



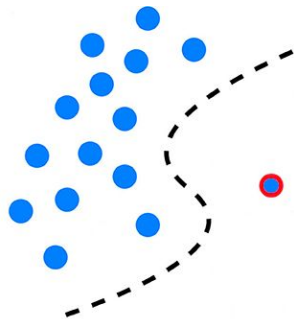


# Senzitivnost i specificnost

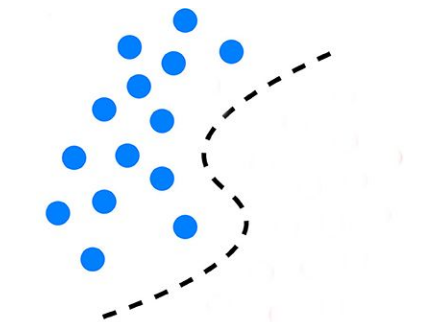
# Senzitivnost i specifičnost



Senzitivnost

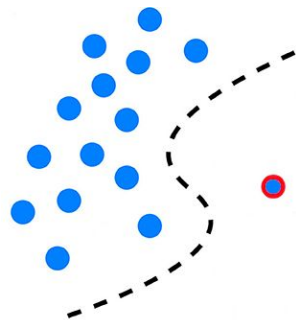


# Senzitivnost i specifičnost

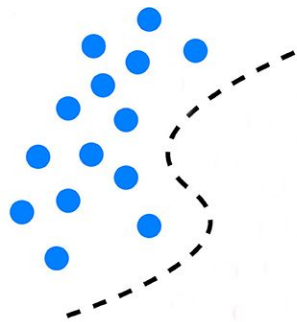


Senzitivnost

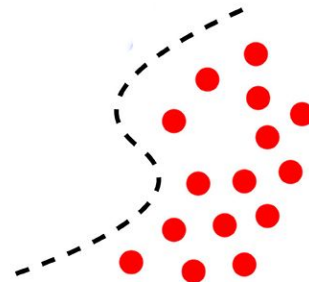
poznato?



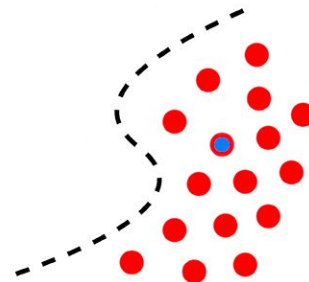
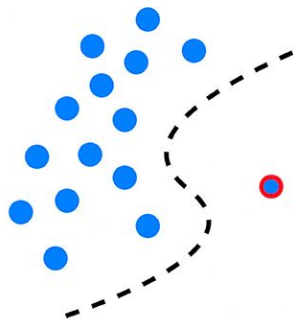
# Senzitivnost i specifičnost



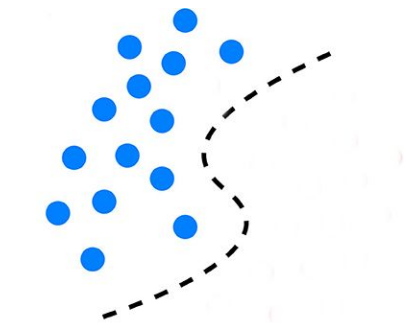
Senzitivnost



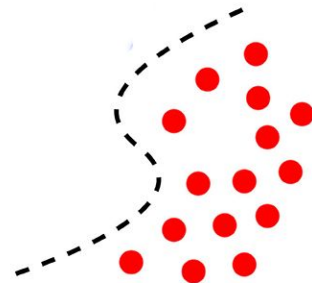
Specifičnost



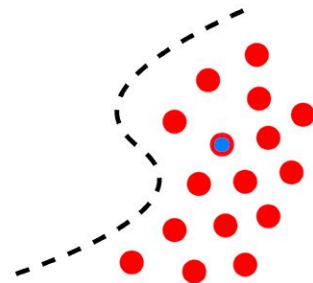
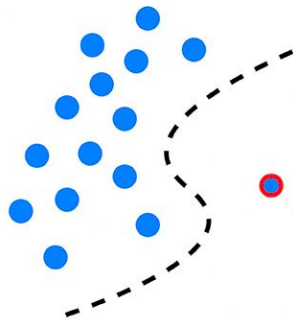
# Senzitivnost i specifičnost



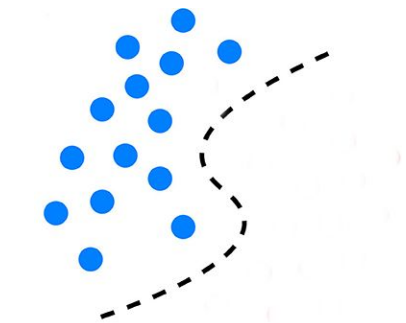
Sensitivity



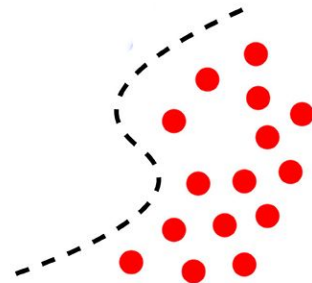
Specificity



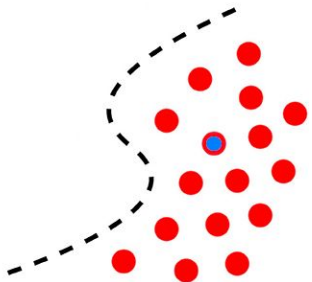
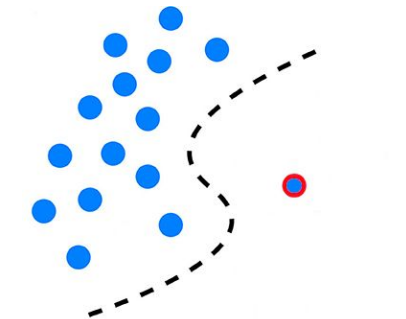
# Senzitivnost i specifičnost



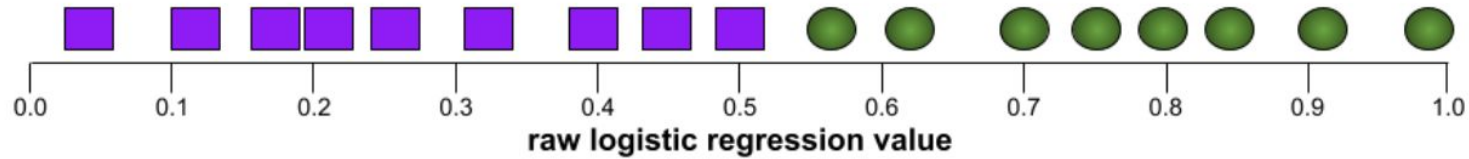
Sensitivity  
TPR



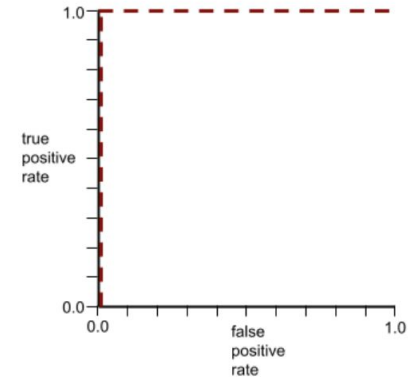
Specificity  
FPR



# ROC (Receiver operating characteristic)

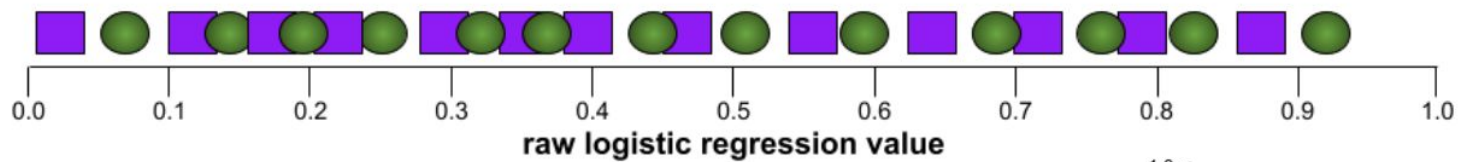


■ negative class      ● positive class

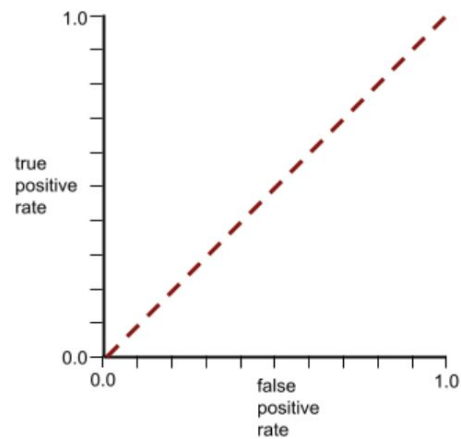


[Machine Learning Glossary](#) | [Google for Developers](#)

# ROC

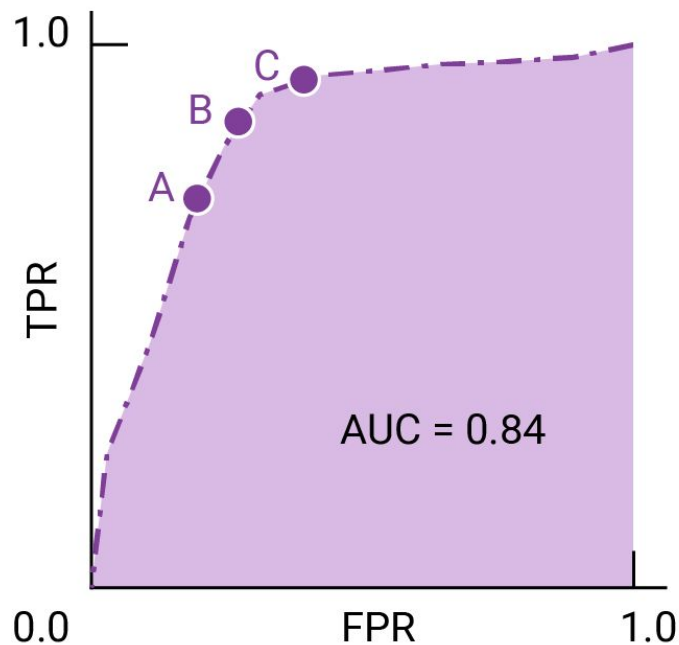


negative class      positive class



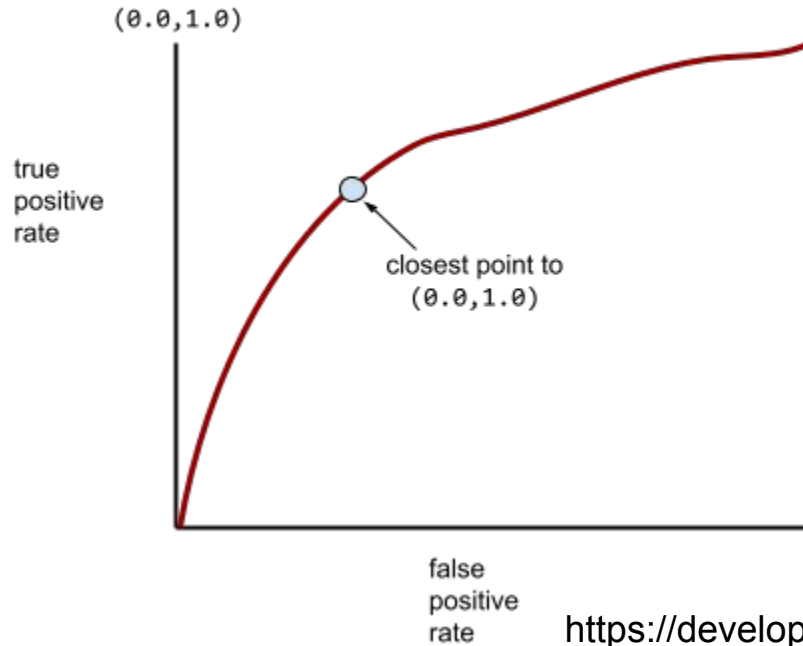


# ROC - AUC (Area Under the Curve)



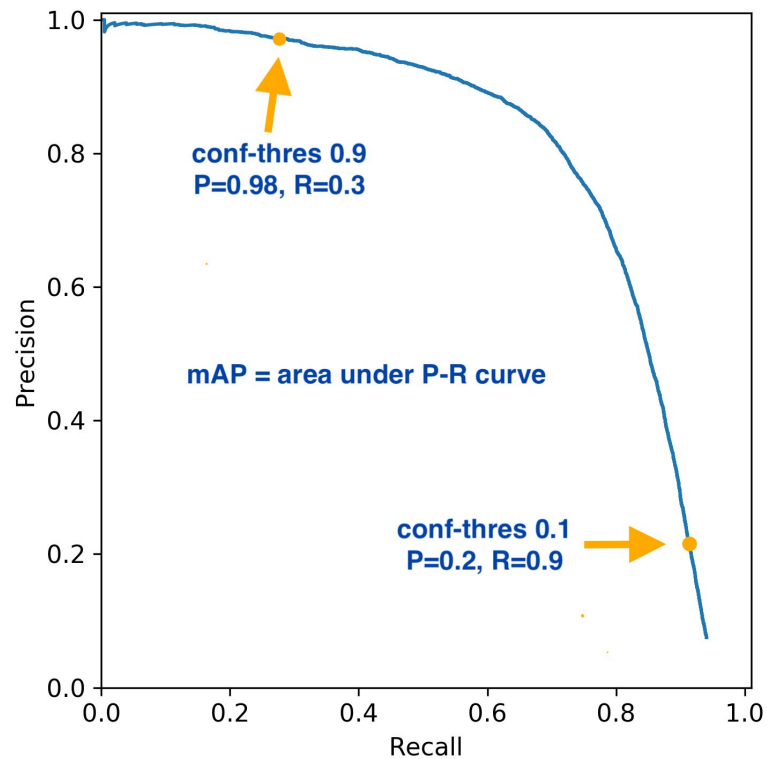
<https://developers.google.com/machine-learning/crash-course/classification>

# ROC - AUC



<https://developers.google.com/machine-learning/crash-course/classification>

# PR (Precision-Recall) kriva



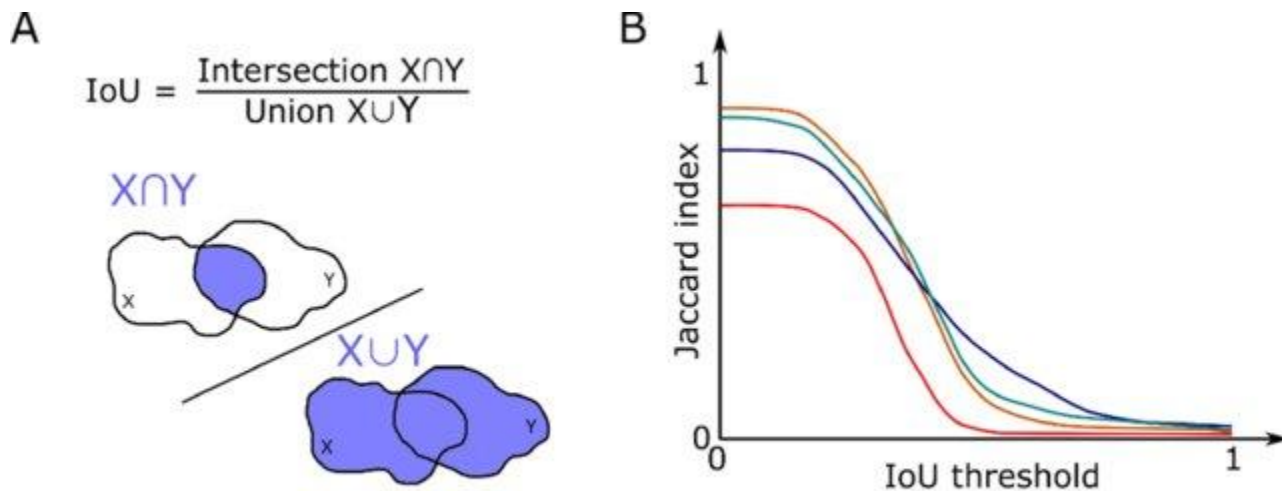
## Merenje greške modela za regresiju

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^N (x_i - m_i)^2$$

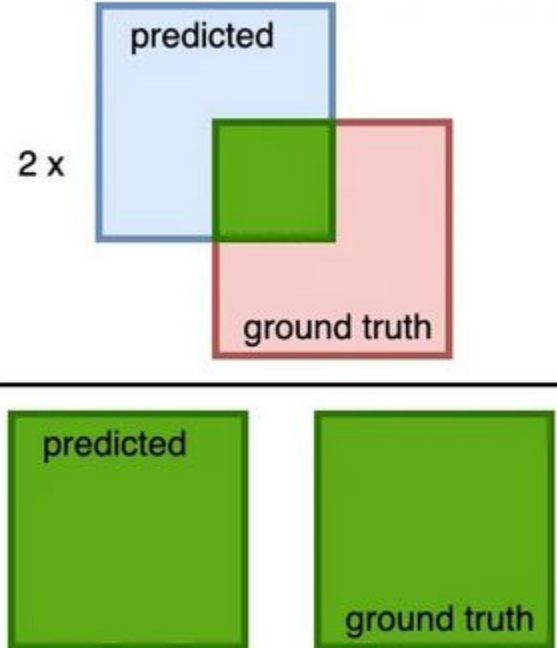
$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (x_i - m_i)^2}$$

$$MAE = \frac{1}{N} \cdot \sum_{i=1}^N |x_i - m_i|$$

# Merenje greške modela za segmentaciju

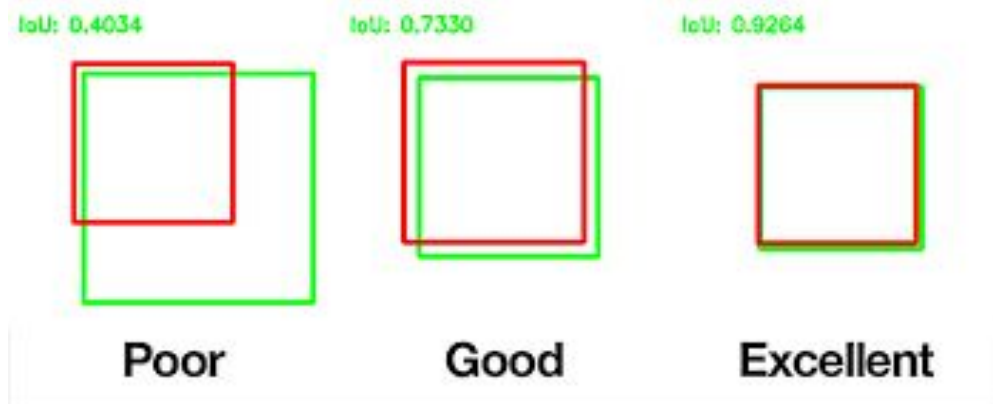


# Merenje greške modela za segmentaciju

$$\text{Dice coefficient} = \frac{2 \times \text{area of overlapped (green)}}{\text{total area (green)}} =$$


The diagram illustrates the Dice coefficient calculation for image segmentation. It shows two overlapping rectangles: a blue one labeled "predicted" and a red one labeled "ground truth". The overlapping region is shaded green. A label "2 x" points to this green area. Below the equation, two separate green rectangles are shown, one labeled "predicted" and one labeled "ground truth", representing the total area of the green regions.

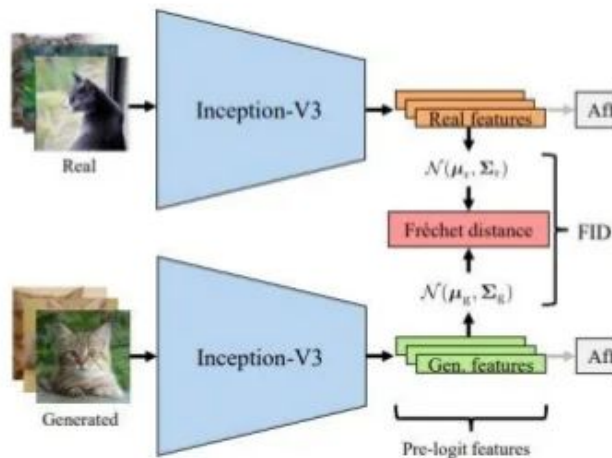
# Merenje greške modela za segmentaciju



# Merenje greške generativnih modela

FID (Fréchet Inception Distance)

Fréchet Inception Distance



“perceptual distance between two distribution”

is it true?

$$\text{FID}(\mu_r, \Sigma_r, \mu_g, \Sigma_g) = \|\mu_r - \mu_g\|_2^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}), \quad (1)$$



# Preciznost i prediktivnost

Scenario:

50-godišnja žena, bez simptoma, rutinski mamografski pregled. Rezultat je pozitivan. Žena, uznemirena, želi da zna da li to znači da zasigurno ima rak i koje su joj šanse.

Informacije koje imamo:

Prevalenca: 1%

Senzitivnost skrininga: 90%

Specifičnost skrininga: 91%

Ako je osoba dijagnostifikovana pozitivno na mamografski skrining, koja je verovatnoća da ima rak dojke?

- A) 9/10
- B) 8/10
- C) 1/10
- D) 1/100

# Naivni Bajesov algoritam

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

Želimo da proverimo hipotezu, na osnovu nekog događaja.

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

Želimo da proverimo hipotezu, na osnovu nekog događaja.

Iskaz:

Stevan je vrlo stidljiv i povučen, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

Želimo da proverimo hipotezu, na osnovu nekog događaja.

Iskaz:

Stevan je vrlo stidljiv i povučen, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

Pitanje:

Da li verovatnije da je Stevan bibliotekar ili poljoprivrednik?

# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

Želimo da proverimo **hipotezu**, na osnovu nekog **dogadjaja**.

Iskaz:

Stevan je vrlo stidljiv i povučen, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

Pitanje:

Da li verovatnije da je Stevan **bibliotekar** ili **poljoprivrednik**?



# Naivni Bajesov algoritam

Osnova: **Bajesova teorema**. Jedan od ključnih principa u matematici.

Želimo da proverimo hipotezu, na osnovu nekog događaja.

Iskaz:

Stevan je vrlo stidljiv i povučen, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

Pitanje:

Da li verovatnije da je Stevan bibliotekar ili poljoprivrednik?

$$P(H|E)$$

# Naivni Bajesov algoritam

Pretpostavka.

Neka **40%** od svih bibliotekara može odgovarati u većoj ili manjoj meri opisanom iskazu, dok se za tek **10%** od svih poljoprivrednika može reći isto.

**Kaneman i Tverski**, serija eksperimenata 1960ih:

**90%** ispitanika je reklo da je veća verovatnoća da je Steva **bibliotekar**.

**10%** reklo da je **poljoprivrednik**.

Većina ispitanika nije uzela u obzir:

(u to vreme) **20** puta više poljoprivrednika u Americi (danas i do **60**).

# Naivni Bajesov algoritam

Što je značilo da je verovatnoća da je Stevan **poljoprivrednik** **83.33%**.

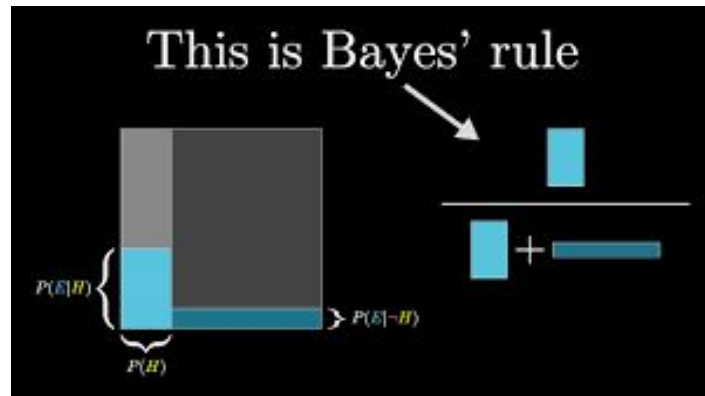
**Kaneman** (*Kahneman*) je za ovo i ostala istraživanja dobio **Nobelovu nagradu** iz ekonomije 2002. (**Tverski** (*Tversky*) je preminuo 1996., i nagrada se ne dodeljuje posthumno).

U Srbiji ima 628 hiljada poljoprivrednih gazdinstava, naspram 2700 biblioteka.\*

$$p = 20 * b$$

$$0.4 * b / (0.4 * b + 0.1 * p) = 0.4 * b / (0.4 * b + 2 * b) = 0.167$$

$$0.1 * p / (0.4 * b + 0.1 * p) = 0.1 * p / (0.02 * p + 0.1 * p) = 0.833$$



# Naivni Bajesov algoritam

**Bajesova teorema:**

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

# Naivni Bajesov algoritam

**Bajesova teorema:**

$$P(H|E) = \frac{P(H)P(E|H)}{P(E)}$$

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.

# Naivni Bajesov algoritam

Iskaz:

Stevan je vrlo stidljiv i povučen, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.

# Naivni Bajesov algoritam

Iskaz:

Stevan je vrlo **stidljiv** i **povučen**, **neizostavno pomaže svima**, ali s druge strane **nema interes** u druge ljude i **živi u svom svetu**. **Blage** i **uredne naravi**. Ima **potrebu da sve bude na svom mestu**, i **veliku strast prema detaljima**.

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.

# Naivni Bajesov algoritam

Iskaz:

Stevan je vrlo stidljiv i **povučen**, neizostavno pomaže svima, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.



# Naivni Bajesov algoritam

## Iskaz:

Stevan je vrlo stidljiv i povučen, **neizostavno pomaže svima**, ali s druge strane nema interes u druge ljude i živi u svom svetu. Blage i uredne naravi. Ima potrebu da sve bude na svom mestu, i veliku strast prema detaljima.

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.

# Naivni Bajesov algoritam

Iskaz:

Stevan je vrlo **stidljiv** i **povučen**, **neizostavno pomaže svima**, ali s druge strane **nema interes u druge ljude** i **živi u svom svetu**. **Blage** i **uredne naravi**. Ima **potrebu da sve bude na svom mestu**, i **veliku strast prema detaljima**.

**"Naivna" pretpostavka o nezavisnosti:**  $E_1, E_2, E_3, \dots$  nezavisni s obzirom na ciljni atribut  $H$ , što pojednostavljuje izraz za verovatnoću posmatranja atributa:

$$P(E|H) = P(E_1|H) \cdot P(E_2|H) \dots = \prod_i P(E_i|H)$$

Mnogo olakšava računanje i ubrzava proces treniranja.

# Kako da se odlučimo koji klasifikator da koristimo?

Svaki od nabrojanih metoda ima svoje prednosti i mane.

Najbolji izbor zavisi:

- od svakog problema ponaosob
- od veličine i prirode skupa podataka
- željene interpretibilnosti modela.

Primer dobre prakse je probati nekoliko metoda i uporediti njihove performanse, nakon čega sledi odabir modela koji ima najbolje performanse za zadat problem.

# Sažetak lekcije

Kako klasifikator koristi labelirane podatke u obuci modela

Uticaj podataka, i njihova podela na trening i test skup na klasifikator

Mere sličnosti

Kako se meri uspešnost klasifikatora i zašto nije dovoljna samo tačnost

Naivne Bajesove metode.

# Literatura

1. Predavanje o klasifikaciji (SAIS radionica 2023)
2. Kurs Mašinskog Učenja [CS102](#)
3. Fast.ai [kurs](#) o Mašinskom učenju
4. [Članak](#) o merama rastojanja
5. Deo Google kursa o Mašinskom učenju u kome je objašnjena [tačnost](#), [preciznost...](#)
6. 3blue1brown [video o Bajesovoj teoremi](#) i [medicinskom paradoksu](#)