🎯 ML FUNDAMENTALS

1. Sta je overfitting?

Overfitting je kada se model previse prilagodi trening podacima, pa lose generalizuje na nove primere.

Kako se sprecava:

- regularizacija (L2, dropout)
- vise podataka
- data augmentation
- jednostavniji model
- early stopping

---

2. Sta je underfitting?

Underfitting je kada je model preslab i ne uci dovoljno, pa ima losu performansu i na trening i na test podacima.

---

3. Razlika izmedju overfitting i underfitting?

- Overfitting → model uci previse
- Underfitting → model uci premalo

---

4. Sta je bias-variance tradeoff?

- Bias = koliko model pojednostavljuje problem (underfit)
- Variance = koliko je model osetljiv na podatke (overfit)

Ne mozes smanjiti oba u isto vreme — uvek balans.

---

5. Sta je regularizacija?

Tehnike koje sprecavaju overfitting ogranicavanjem slozenosti modela.

Primeri:

- L1
- L2
- Dropout
- Data augmentation

---

6. L1 vs L2 regularizacija?

- L1 → sparsity (gura parametre ka 0)
- L2 → smanjuje vrednosti parametara ali ih ne ponistava

---

7. Sta je dropout?

Tehnika koja nasumicno iskljucuje neurone tokom treninga da bi smanjila overfitting.

---

8. Sta je learning rate?

Koliko brzo gradient descent menja parametre.

- previsok → eksplodira
- prenizak → sporo ucenje

---

9. Sta je gradient descent?

Algoritam koji optimizuje model tako sto ide u smeru opadanja loss funkcije.

---

10. Batch, mini-batch i full-batch?

- Full batch: koristi ceo dataset
- Mini batch: najcesci, balans brzine i stabilnosti

- Stochastic: jedan primer

11. Sta je loss funkcija?

Mera koliko je predikcija daleko od cilja.

Primeri:

- MSE za regresiju
- Cross-entropy za klasifikaciju

12. Precision vs Recall?

- Precision = od predvidjenih pozitivnih, koliko je stvarno pozitivnih
- Recall = od svih stvarno pozitivnih, koliko smo pogodili

13. Kada je bitan recall, kada precision?

- Recall = medicina, fraud detection (ne sme propustiti pozitivne)
- Precision = reklame, preporuke (ne sme davati pogresne preporuke)

14. Sta je F1 score?

Harmonijska sredina precision i recall.

Koristi se kada su klase nebalansirane.

15. Sta je confusion matrix?

Tabela koja prikazuje TP, FP, FN, TN — osnovna za analizu modela.

16. Sta je ROC kriva i AUC?

ROC = odnos TPR i FPR pri razlicitim threshold-ovima.

AUC = povrsina ispod ROC krive → sto blize 1, to bolje.

17. Razlika: regresija vs klasifikacija?

- Regresija → predvidja broj
- Klasifikacija → predvidja klasu

---

18. Sta je normalizacija podataka?

Skaliranje features-a u slican opseg (npr. 0–1).

Neophodno za modele osetljive na skalu (SVM, KNN, NN).

---

19. Sta je standardizacija?

Pretvaranje features-a u mean=0, std=1.

Koristi se kada su podaci normalno rasporedjeni.

---

20. Zasto je podela podataka na train/validation/test bitna?

Da bismo merili:

- trening performansu
- generalizaciju
- finalnu procenu modela bez leakage-a

---

21. Sta je feature engineering?

Proces pripreme i transformacije podataka kako bi model imao sto vise korisnih informacija.

Ukljucuje:

- encoding
- scaling
- kreiranje novih feature-a
- biranje najbitnijih feature-a

22. Sta je one-hot encoding?

Pretvaranje kategorickog podatka u binarne kolone (0/1).

Primer: “boja = crvena” → [1,0,0].

23. Sta je label encoding?

Pretvaranje kategorija u brojeve (0, 1, 2...).

Koristi se za ordinalne podatke.

24. Sta je data leakage?

Kada informacije iz test seta “procure” u trening set, pa model izgleda bolje nego sto jeste.

Primer: skaliranje pre train/test podele.

25. Sta je class imbalance?

Kada jedna klasa ima mnogo vise primera od druge.

Primer: 95% healthy, 5% diseased.

Rešenja:

- SMOTE
- undersampling
- metric fokus: F1, recall
- class weights

26. Sta je cross-validation?

Tehnika kojom se dataset deli na vise fold-ova da bi se dobila stabilnija evaluacija.

Najcesci: **k-fold**.

27. Kada koristimo k-fold cross validation?

- kada imamo malo podataka
- kada zelimo stabilne metrike
- kada radimo model selection

---

28. Sta je hyperparameter tuning?

Proces biranja vrednosti hiperparametara (learning rate, depth, batch size...) koji najvise poboljsavaju model.

Alati:

- grid search
- random search
- Bayesian optimization

---

29. Sta je early stopping?

Prekida trening kada validation loss pocne da raste (znak overfittinga).

---

30. Sta je batch normalization?

Normalizacija izlaza slojeva unutar neuralne mreze da bi se ubrzao trening i stabilizovao model.

---

31. Zasto se neuralne mreze tesko treniraju bez normalizacije?

Zbog:

- exploding gradients
- vanishing gradients
- nestabilnog ucenja

---

32. Sta je activation funkcija?

Funkcija koja daje nelinearnost modelu.

Primeri:

- ReLU
- Sigmoid
- Tanh
- Softmax

---

33. Zasto se koristi ReLU?

- brza
- jednostavna
- ne saturise kao sigmoid/tanh
- resava problem vanishing gradients delom

---

34. Sta je softmax funkcija?

Pretvara vektor u verovatnosne raspodele (sabiraju se u 1).

Koristi se u multi-class klasifikaciji.

---

35. Sta je embedding?

Ucenje kompaktnog, gustog, nizodimenzionalnog predstavljanja podataka.

Klasicno u NLP-u: pretvara reci u vektore.

---

36. Sta je bag-of-words?

Stari NLP pristup gde se tekst reprezentuje kao brojanje reci bez konteksta.

Slabost: gubi redosled i semantiku.

---

37. Sta je TF-IDF?

Poboljsani bag-of-words koji kaznjava reci koje se preterano cesto pojavljuju u dokumentima.

---

38. Razlika izmedju parametarskih i neparametarskih modela?

- Parametarski: fiksiran broj parametara (logistic regression, linear regression).
- Neparametarski: broj parametara raste sa podacima (kNN, decision tree).

---

39. Da li je KNN isto sto i CNN?

Ne!

KNN = k-nearest neighbors → neparametarski algoritam.

CNN = convolutional neural network → duboka neuronska mreza za slike.

Ovo cesto pitaju da vide da li se zbunis.

---

40. Kako radi decision tree?

- deli podatke prema feature-u koji najvise smanjuje impurity
- uci strukturu “if–else” pravila
- lako se pretreniraju ako se ne ogranicavaju (depth, min_samples_leaf)

41. Kako radi Random Forest?

Ensemble od vise decision tree stabala gde svako stablo uci na random podskupu podataka i feature-a.

Prednost: bolja generalizacija, manji overfitting.

---

42. Zasto Random Forest manje overfituje od Decision Tree?

Jer kombinuje mnogo razlicitih stabala → smanjuje varijansu.

43. Sta je Boosting?

Metoda gde se modeli treniraju jedan za drugim i svaki ispravlja greske prethodnog.

Primeri:

- AdaBoost
- XGBoost
- LightGBM
- CatBoost

44. Razlika izmedju Bagging i Boosting?

- Bagging (npr. Random Forest) → modeli uce *paralelno*
- Boosting → modeli uce *sekvencijalno* i fokusiraju se na greske

45. Sta je XGBoost?

Napredni boosting algoritam koji je vrlo efikasan, brz, i cesto najbolji izbor za tabular podatke.

46. Kada koristiti neuralne mreze, a kada klasicne modele (RF, XGBoost)?

- Tabular: XGBoost/Random Forest su cesce bolji
- Slike: CNN
- Tekst: Transformers
- Kada ima puno podataka: neuralne mreze dobijaju prednost

47. Sta je sigmoid funkcija i kada se koristi?

Pretvara vrednost u opseg (0,1).

Koristi se kod binary klasifikacije.

48. Sta je tanh funkcija i kada se koristi?

opseg (-1,1) → stabilnije od sigmoid.

Koristi se u nekim starijim RNN modelima.

49. Sta je softmax i kada se koristi?

U multi-class klasifikaciji za dobijanje verovatnoca po klasama.

50. Sta je epoch?

Jedan prolaz kroz ceo dataset tokom treniranja neuralne mreze.

51. Overfitting u neuralnim mrezama – glavni uzroci?

- previse parametara
- premalo podataka
- bez regularizacije
- predugo treniranje

52. Kako smanjiti overfitting u neuralnim mrezama?

- dropout
- weight decay (L2)
- early stopping
- data augmentation
- manja mreza

53. Sta je convolution (konvolucija)?

Operacija gde filter klizi preko slike i detektuje lokalne obrasce (ivice, oblike...).

54. Zasto se koriste CNN?

Jer uce lokalne i hijerarhijske reprezentacije slika → najbolji su za vizuelne zadatke.

55. Sta je pooling u CNN?

Smanjenje dimenzija slike (najcesce max pooling) → smanjuje kompleksnost i hvata najbitnije karakteristike.

56. Sta je transfer learning?

Koriscenje pretreniranih modela (ResNet, MobileNet, BERT...) kao start, pa ih fino doteras (fine-tuning) na manjem datasetu.

Velika prednost kad je dataset mali.

57. Kada biramo fine-tuning, a kada zamrzavamo slojeve?

- Mali dataset → zamrzni vecinu slojeva, treniraj samo poslednje
- Veliki dataset → full fine-tuning

58. Sta je RNN?

Recurrent Neural Network → model za sekvencijalne podatke (tekst, audio, time series).

Ima problem long-term dependencies (resava se LSTM/GRU).

59. Sta je LSTM?

Long Short-Term Memory → napredni RNN koji bolje pamti kontekst i izbegava problem dugih sekvenci.

60. Razlika izmedju LSTM i GRU?

- GRU je jednostavniji i brzi, sa manjim brojem parametara
- LSTM je mocniji, ali sporiji

Oba sluze za sekvencijalne zadatke.

---

61. Sta je attention mehanizam?

Attention je mehanizam koji modelu omogucava da “obrati paznju” na najbitnije delove ulaza, posebno u NLP zadacima.

Daje kontekst bolje od RNN-a.

---

62. Sta je Transformer arhitektura?

Model baziran samo na attention mehanizmu (nema RNN/CNN).

Koristi se u skoro svim modernim NLP modelima (BERT, GPT, T5...).

---

63. Razlika izmedju BERT i GPT?

- **BERT**: bidirekcioni → super za klasifikacije, embeddinge
- **GPT**: autoregresivni → super za generisanje teksta

---

64. Sta je tokenizacija?

Pretvaranje teksta u tokene (rec, podrec, karakter).

Primer: WordPiece tokenizacija u BERT-u.

---

65. Sta je padding i zasto se koristi?

Dodavanje specijalnih tokena da bi sve sekvence bile iste duzine.

Neophodno kod batch treniranja.

---

66. Sta je mask attention u BERT-u?

BERT maskira deo teksta i trenira da pogodi koji token nedostaje.

To mu omogucava ucenje konteksta.

---

67. Sta je perplexity?

Mera kvaliteta language modela.

Manja perplexity = bolji model.

---

68. Sta je embedding dimenzija?

Velicina vektora koji predstavlja rec/segment.

Veca dimenzija → vise informacija, ali sporije.

---

69. Sta je classical NLP feature extraction?

Pre Transformers ere:

- bag-of-words
- TF-IDF
- n-grams
- word2vec
- GloVe

---

70. Razlika: word2vec vs TF-IDF?

- TF-IDF: frekvencija reci, nema semantiku
- word2vec: uci semanticki slicne reci u bliske vektore

---

71. Sta je unsupervised learning?

Model uci strukturu podataka bez labela.

Primeri: clustering, PCA, autoencoders.

---

72. Sta je supervised learning?

Model uci na podacima sa labelama.

Primeri: klasifikacija, regresija.

73. Sta je semi-supervised learning?

Kombinacija malog broja labelovanih podataka i mnogo nelabelovanih.

74. Sta je reinforcement learning?

Model uci kroz nagrade i kazne.

Koristi se u igrama, robotici, optimizaciji odluka.

75. Sta je clustering?

Grupisanje podataka prema slicnosti bez labela.

Primer: K-means.

76. Kako radi K-means clustering?

1. izabere k centara
2. dodeli podatke najblizem centru
3. azurira centre
4. ponavlja dok se ne stabilizuje

77. Sta je PCA?

Principal Component Analysis → smanjuje dimenzionalnost podataka tako sto trazi pravce najvece varijanse.

78. Kada koristiš PCA?

- kada imas mnogo feature-a

- kada hoces brzi i jednostavniji model
- kada imas redundantne feature-e

---

79. Sta je autoencoder?

Neuralna mreza koja uci kompresovanu reprezentaciju (encoding) i rekonstruise ulaz (decoding).

Koristi se za:

- smanjenje dimenzija
- anomaly detection
- generativne zadatke

---

80. Sta je anomaly detection?

Prepoznavanje neobicnih podataka koji odskacu od normalnih obrazaca.

Koristi se u:

- fraud detection
- industrija
- sigurnost sistema

81. Sta je feature importance?

Mera koja pokazuje koliko je svaki feature doprinosio odlukama modela.

Koristi se u tree modelima i interpretabilnosti.

---

82. Sta je SHAP?

SHAP vrednosti objasnjavaju koliko svaki feature doprinosi pojedinacnoj predikciji modela.

Najbolja metoda za interpretabilnost ML modela.

---

83. Sta je Lime?

Lokalna interpretabilnost modela → objasnjava predikciju u malom lokalnom okruzenju oko primera.

84. Sta je model drift?

Kada se promeni odnos u podacima tokom vremena, pa model pocne da radi lose.

Primer: fraud se menja tokom meseci.

85. Sta je data drift?

Statistika ulaznih podataka se menja tokom vremena (npr. nove navike korisnika).

86. Kako detektovati drift?

- pracenje metrike kroz vreme
- grafovi distribucija
- statisticki testovi (KS-test)
- monitoring u produkciji

87. Sta je kalibracija modela?

Proces koji osigurava da verovatnoce modela zaista odgovaraju realnim verovatnocama.

Primer: ako model kaze 70% → da li je stvarno 70% tacnosti?

88. Sta je ensembling?

Kombinovanje vise modela radi boljeg performansa.

Primeri:

- bagging
- boosting

- stacking

89. Kada ensembling ne pomaze?

- kada su svi modeli slicni
- kada dataset nema mnogo varijacije
- kada je problem linearan i jednostavan

90. Sta je stacking?

Kombinovanje vise modela gde finalni model (meta-model) uci na njihovim izlazima.

91. Sta je AUCPR i kada se koristi?

Area Under Precision-Recall Curve

Koristi se kod ekstremno nebalansiranih podataka (fraud detection!).

92. Sta je log-loss?

Loss funkcija za klasifikaciju koja kaznjava pogresno predvidjene verovatnoce.

93. Zasto accuracy nije dobar metric kod nebalansiranih podataka?

Jer mozes imati 99% accuracy-a sa modelom koji nikada ne prepozna pozitivnu klasu.

94. Sta raditi kada imas malo podataka?

- data augmentation
- transfer learning
- regularizacija

- k-fold cross-validation
- jednostavniji model

---

95. Sta je ROC-AUC limitation?

Moze da bude visok i kada model lose radi na minornoj klasi.

Zato se kod fraud-a uvek gleda PR-AUC.

---

96. Sta je threshold tuning?

Biranje praga za klasifikaciju da bi se optimizovala neka metrika (npr. recall).

---

97. Sta je label smoothing?

Tehnika gde se ciljne verovatnoce malo “omekšaju” (npr. 1 → 0.9) da bi model bio stabilniji.

---

98. Sta je gradient clipping?

Ogranicavanje velicine gradijenta da se izbegne exploding gradients.

---

99. Sta je vanishing gradient problem?

Kod dubokih mreza gradijenti postaju mali → model ne moze da uci.

Resenja:

- ReLU
- batch norm
- bolja arhitektura (ResNet)

---

100. Kako objasnjavas ML model nekome ko nije tehnicki?

Ovako:

“Model uci obrasce iz primera i koristi ih da predvidi nove situacije, kao sto i ljudi prepoznaju slicnosti i donose odluke na osnovu iskustva.”

Ovo Microsoft obozava — jednostavno, jasno, bez ponizavanja sagovornika.