



# بررسی الگوریتم **XGBoost** روی سرطان

ارائه دهندگان:

حورا محمودیان

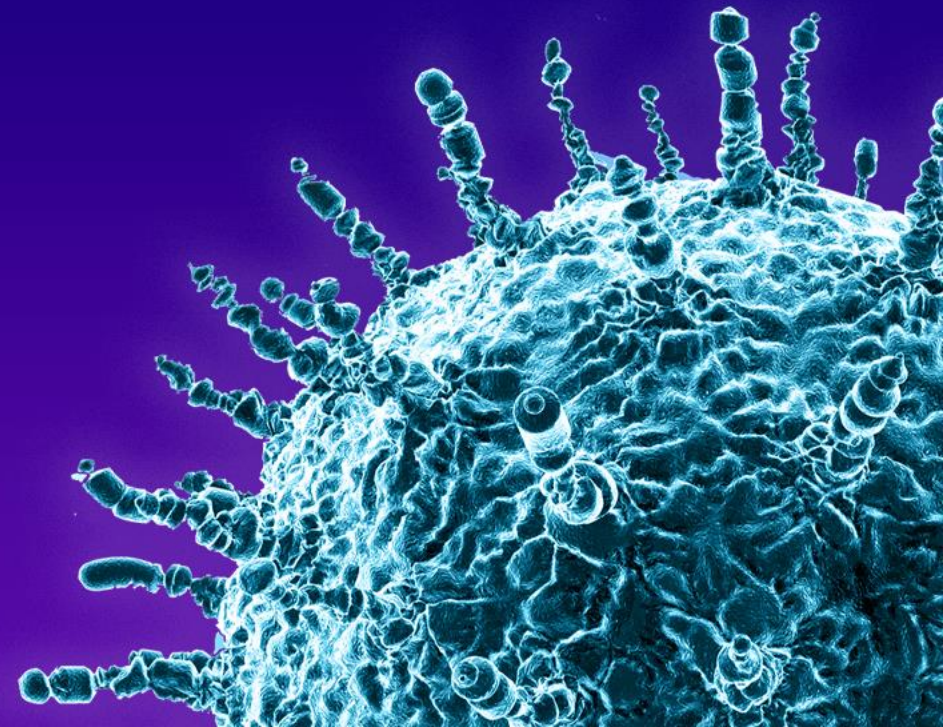
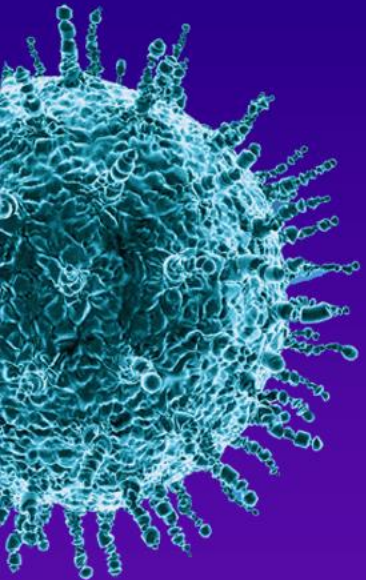
زهرا مستاجران

حوری دهش

استاد:

سرکار خانم دکتر شیرین بقولی زاده

خرداد ۱۴۰۱



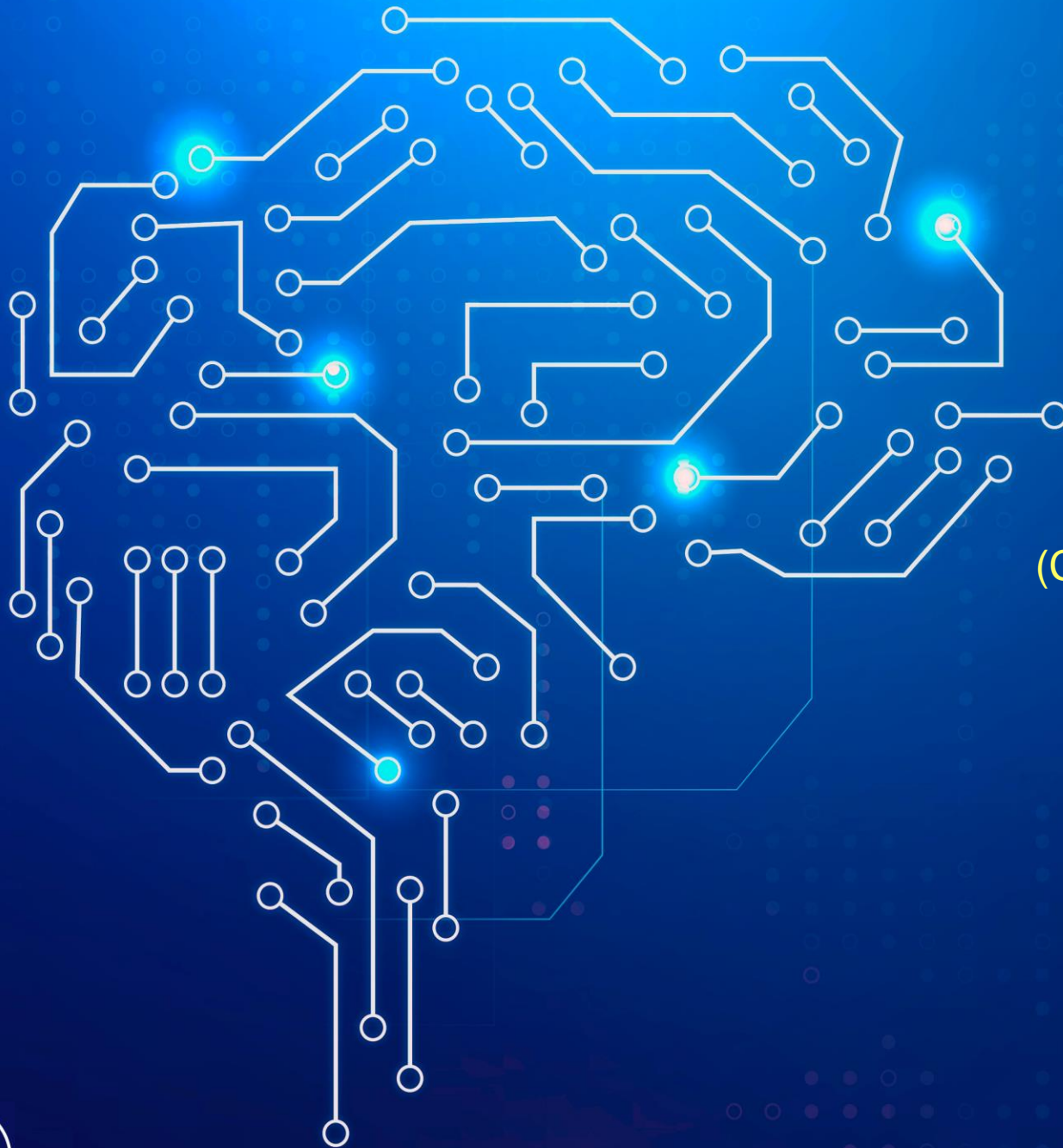
# رئوس مطالب

## ○ یادگیری ماشین

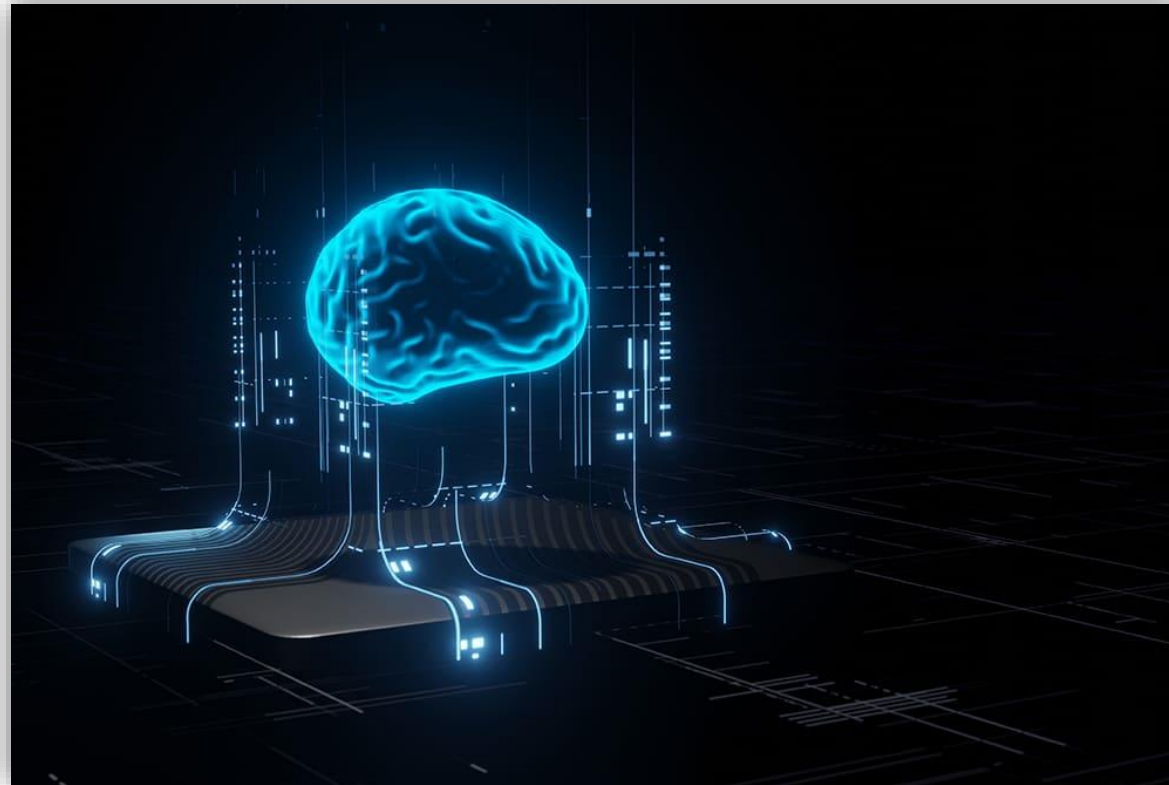
- انواع یادگیری ماشین
- یادگیری بانظارت
- طبقه‌بندی
- الگوریتم تقویتی (Boosting)
- الگوریتم XGBoost

## ○ معماری (CRISP) Cross Industry Standard Process

- درک کسب‌وکار
- درک داده‌ها
- پیش‌پردازش داده‌ها
- مدل‌سازی
  - بهینه‌سازی مدل
- ارزیابی مدل
- استقرار
- جمع‌بندی

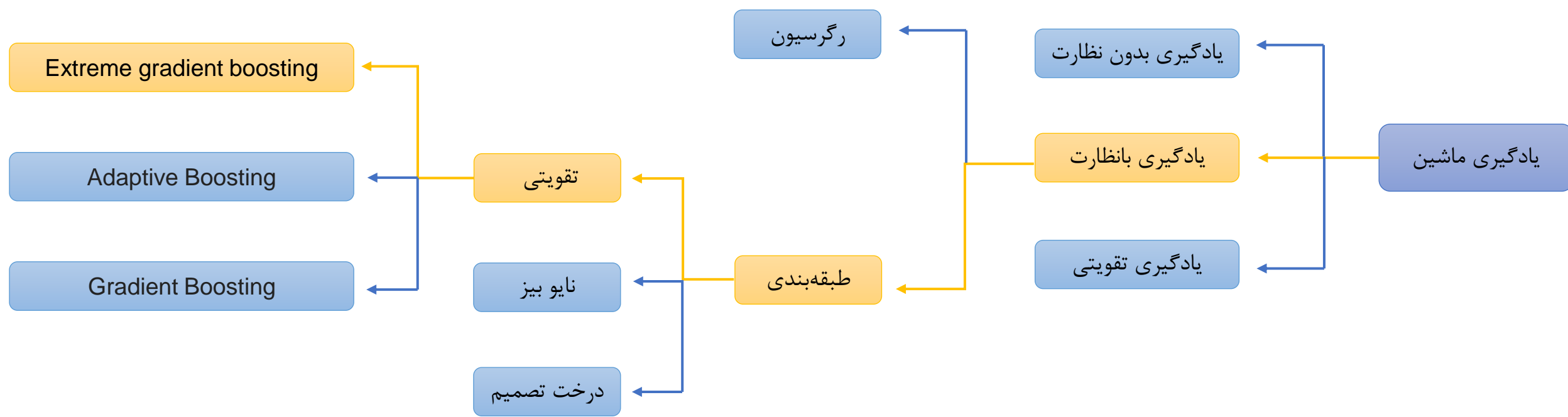






[2]

# انواع یادگیری ماشین





[3]

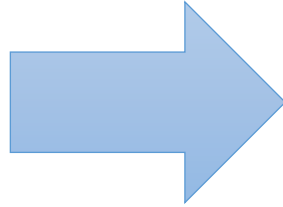


[4]

- K-نزدیک‌ترین همسایگی
- درخت تصمیم
- نایو بیز
- رگرسیون لجستیک
- ماشین بردار پشتیبانی
- **تقویتی (Boosting)**



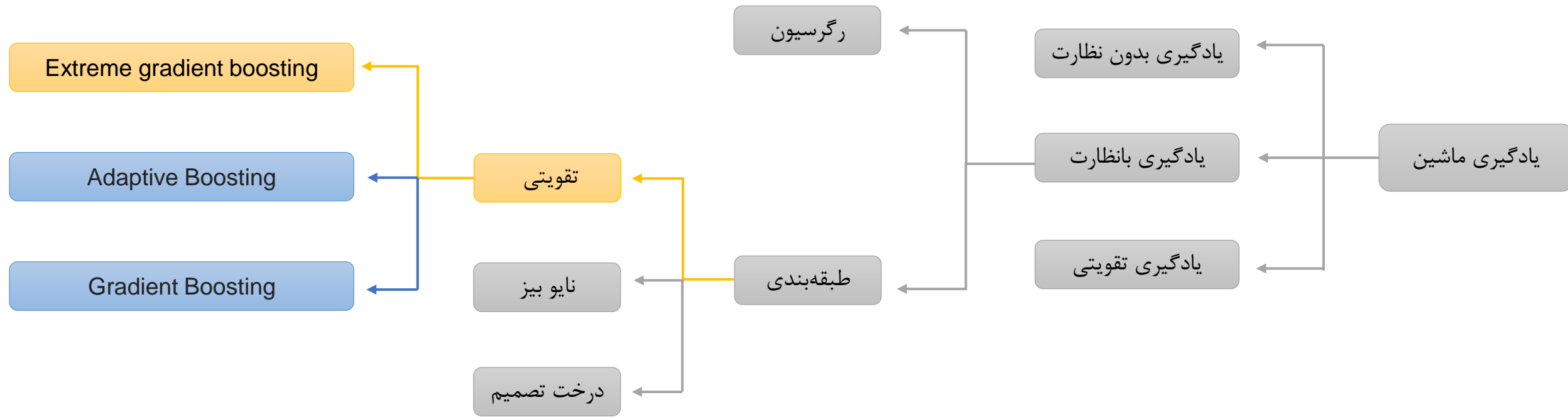
# الگوریتم تقویتی (Boosting)



[5]

[5]

# انواع الگوریتم تقویتی







***XGBoost***

[6]

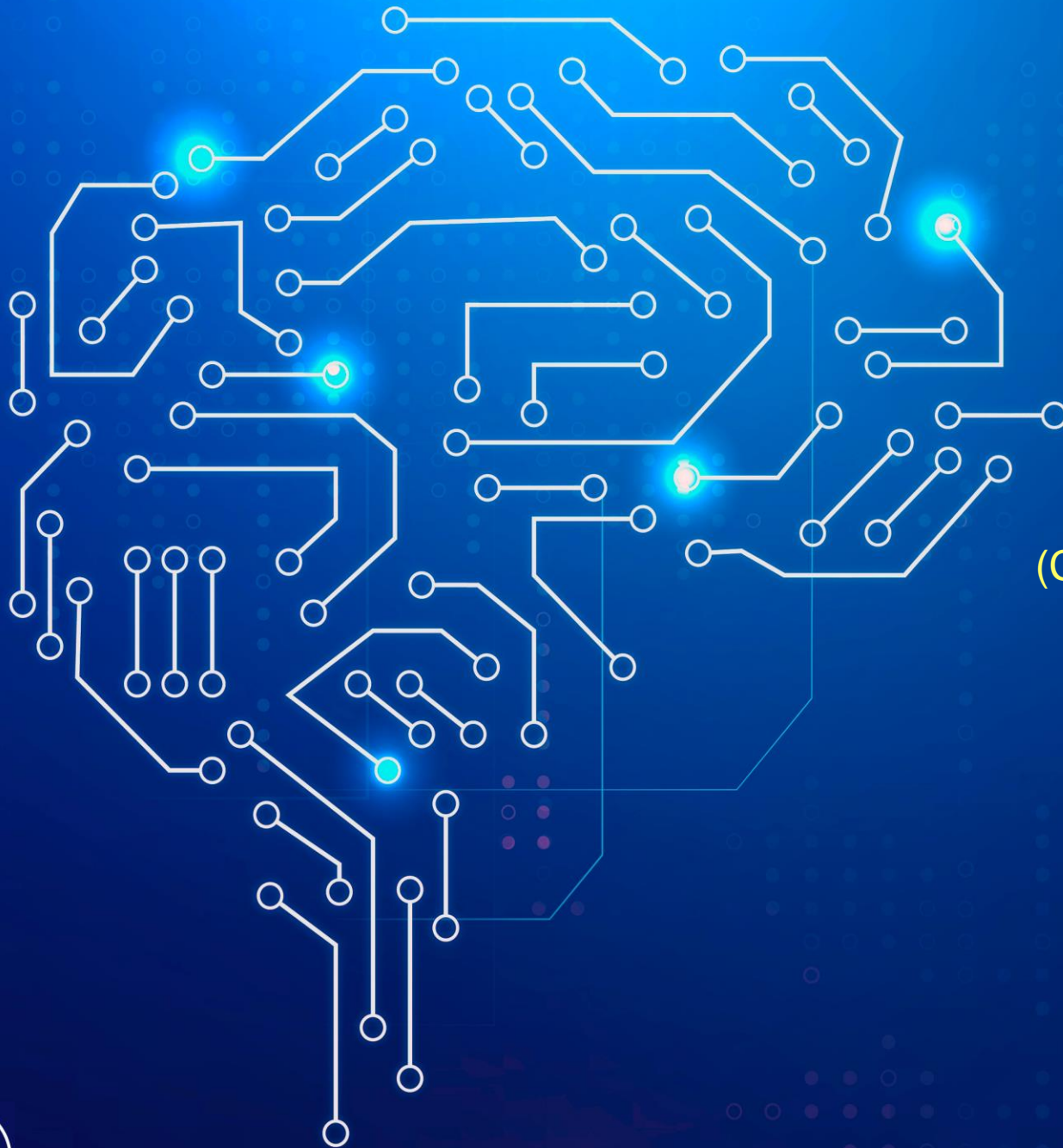
# رئوس مطالب

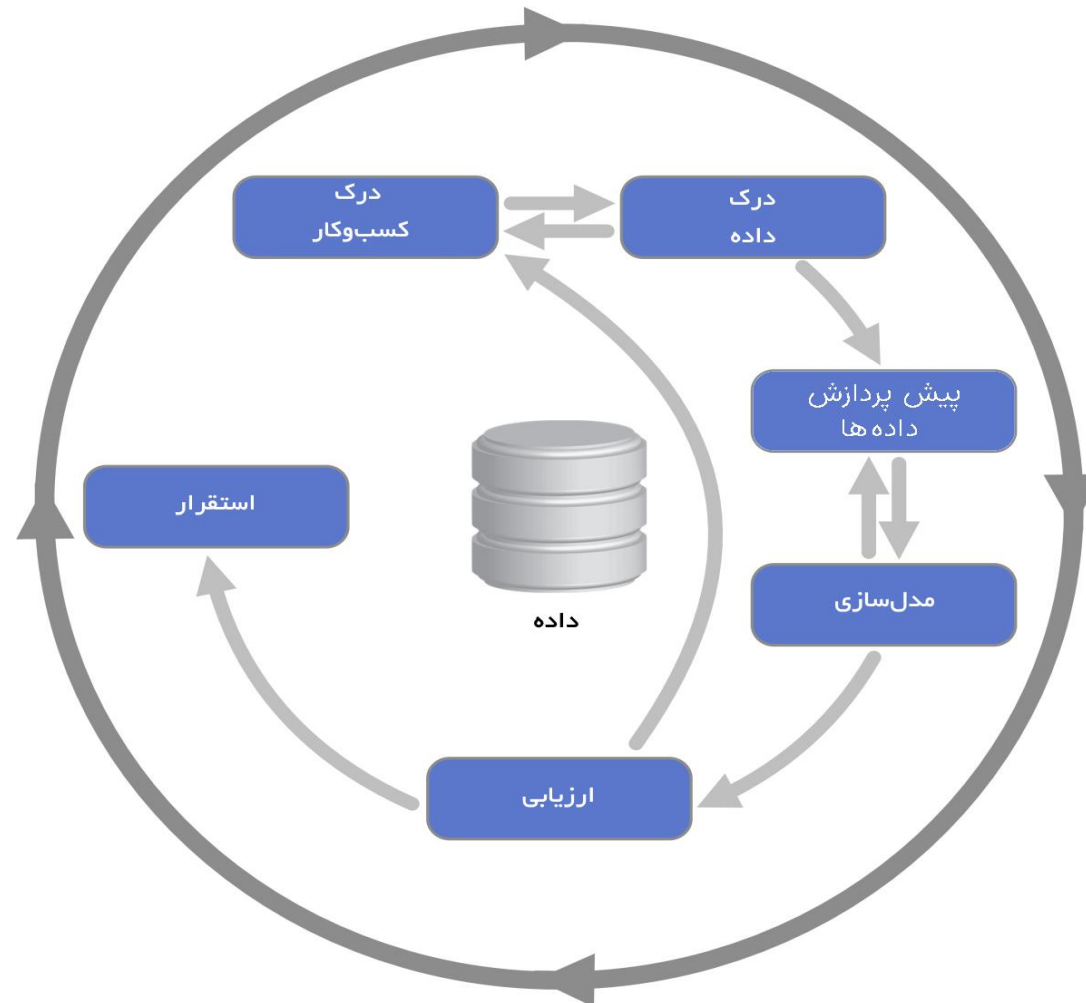
## ○ یادگیری ماشین

- انواع یادگیری ماشین
- یادگیری بانظارت
- طبقه‌بندی
- الگوریتم تقویتی (Boosting)
- الگوریتم XGBoost

## ○ معماری (CRISP) Cross Industry Standard Process

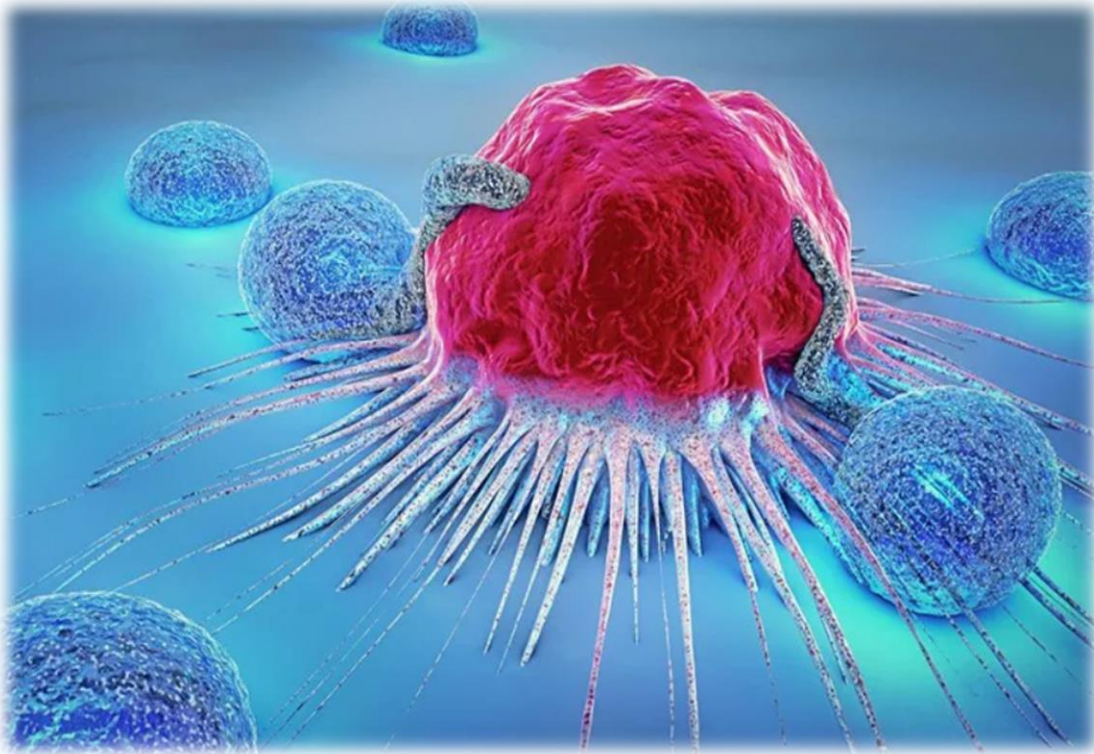
- درک کسب‌وکار
- درک داده‌ها
- پیش‌پردازش داده‌ها
- مدل‌سازی
  - بهینه‌سازی مدل
- ارزیابی مدل
- استقرار
- جمع‌بندی







هدف: جداسازی سرطان‌های مرحله اولیه و آخر



- توصیف محل قرارگیری سرطان
- توصیف محل گسترش
- آیا سرطان بر سایر قسمت‌های بدن تاثیر میگذارد؟

[8]

جمع آوری داده‌ها از اطلس ژنوم سرطان (TCGA)

ارائه متمرکز بر ۴ نوع سرطان:

نوع سرطان	تعداد نمونه
کارسینوم سلول شفاف کلیه	۵۳۷
کارسینوم سلول پاپیلاری کلیه	۲۹۱
کارسینوم سلول سنگفرشی ریه	۵۰۴
کارسینوم سلول سنگفرشی سر و گردن	۵۲۸

اطلاعات موجود در هر مجموعه داده : داده های mRNA، miRNA، DNA و اطلاعات بالینی

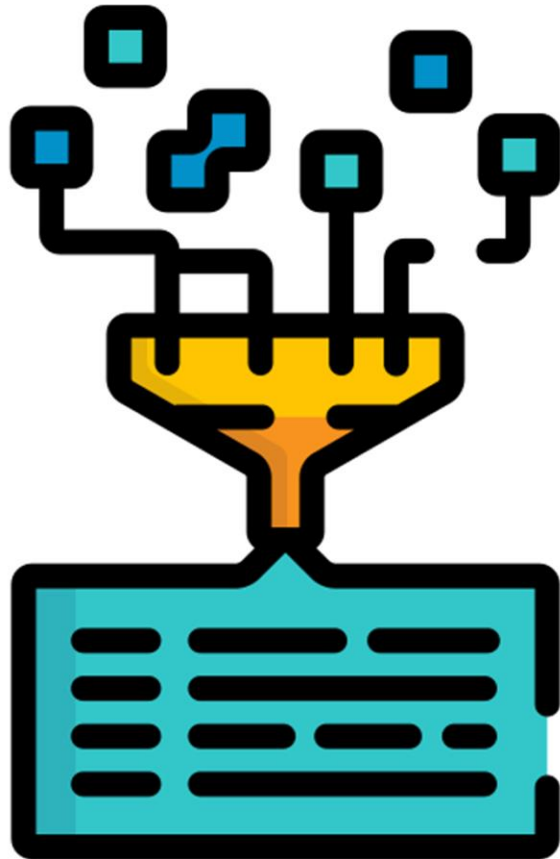
نمونه های رایج پیش پردازش عبارتند از:

- پیدا کردن مقادیر گم شده و حذف آنها
- حذف هر نمونه با بیش از ۲۰٪ گمشدگی
- تبدیل مقادیر
- رمز گذاری متغیرهای طبقه بندی شده



[9]





[10]

- پیدا کردن مقادیر گمشده و حذف آنها

- حذف هر نمونه با بیش از ۲۰٪ گمشدگی

- تبدیل مقادیر

- رمز گذاری متغیرهای طبقه بندی شده

# رئوس مطالب

## ○ یادگیری ماشین

- انواع یادگیری ماشین
- یادگیری بانظارت
- طبقه‌بندی
- الگوریتم تقویتی (Boosting)
- الگوریتم XGBoost

## ○ معماری (CRISP) Cross Industry Standard Process

- درک کسب‌وکار
- درک داده‌ها
- پیش‌پردازش داده‌ها

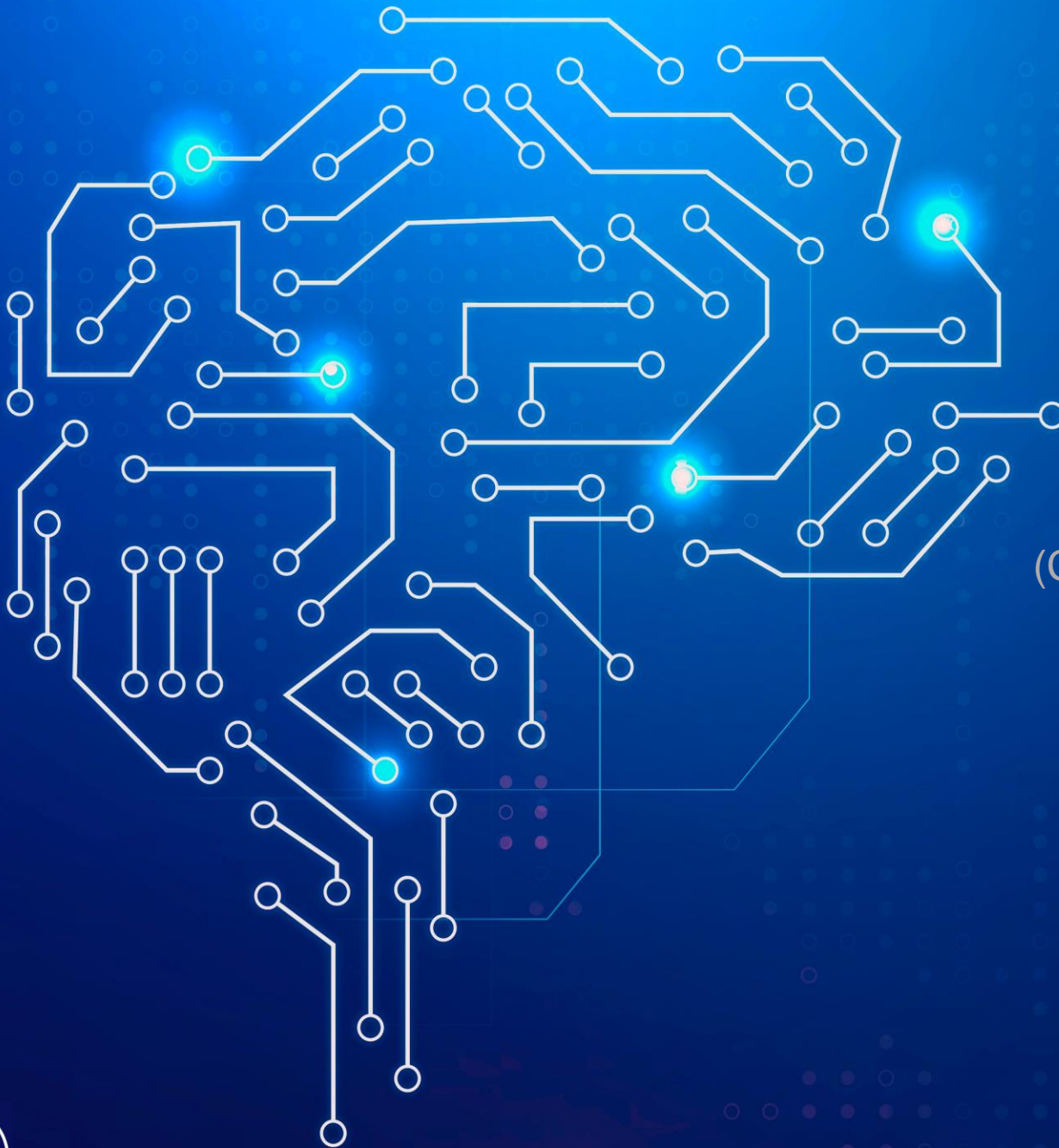
## ○ مدل‌سازی

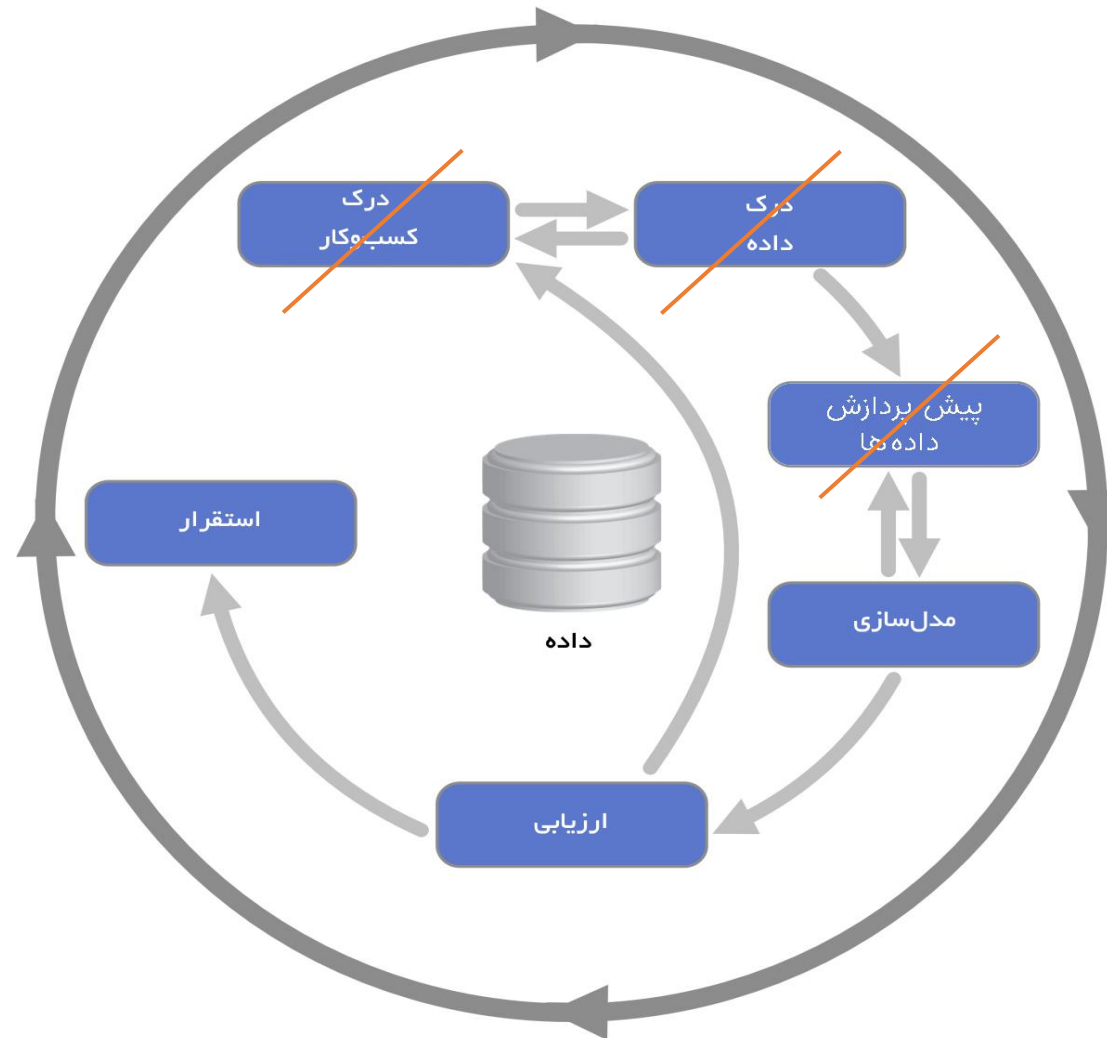
- بهینه‌سازی مدل

## ○ ارزیابی مدل

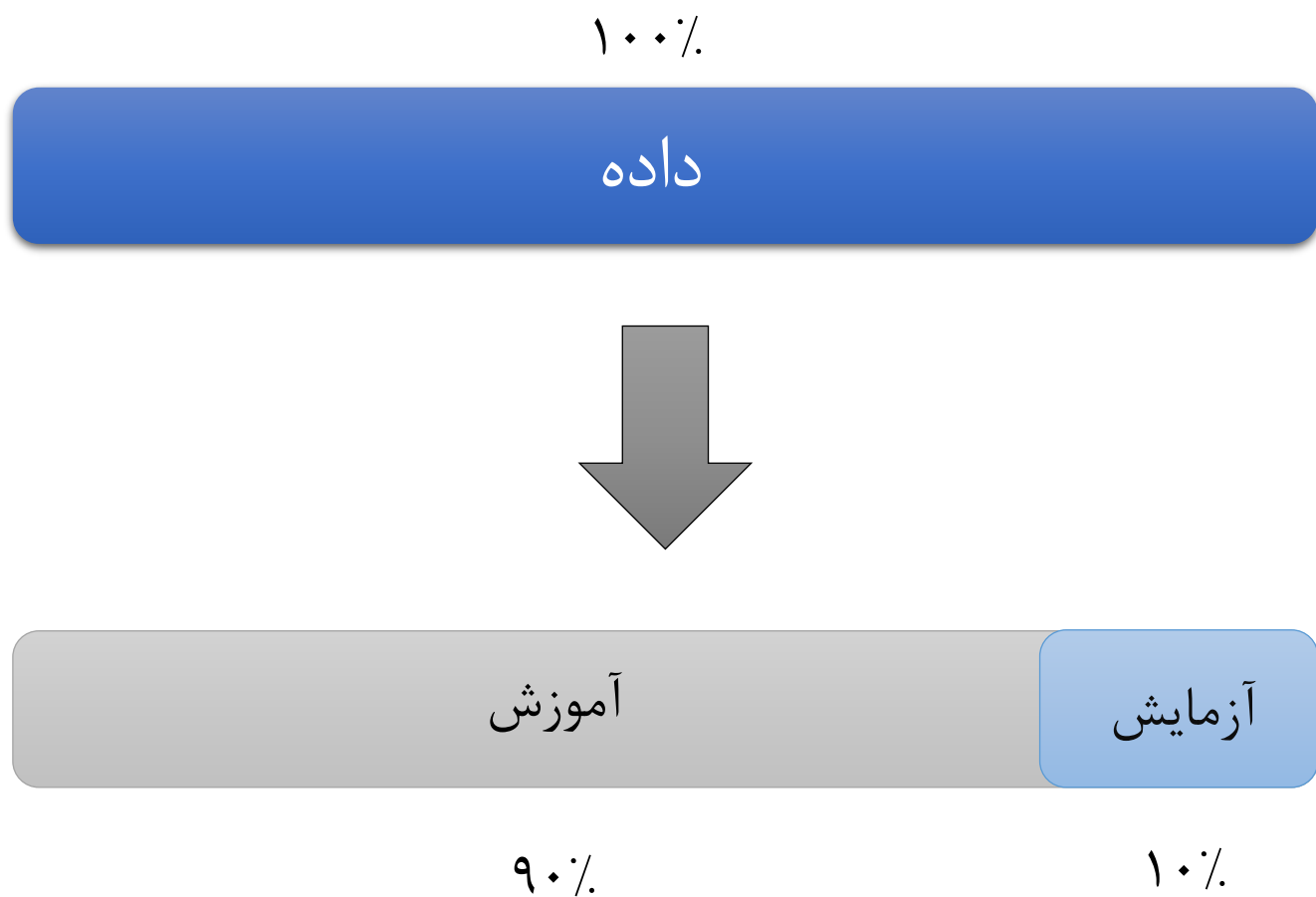
## ○ استقرار

## ○ جمع‌بندی



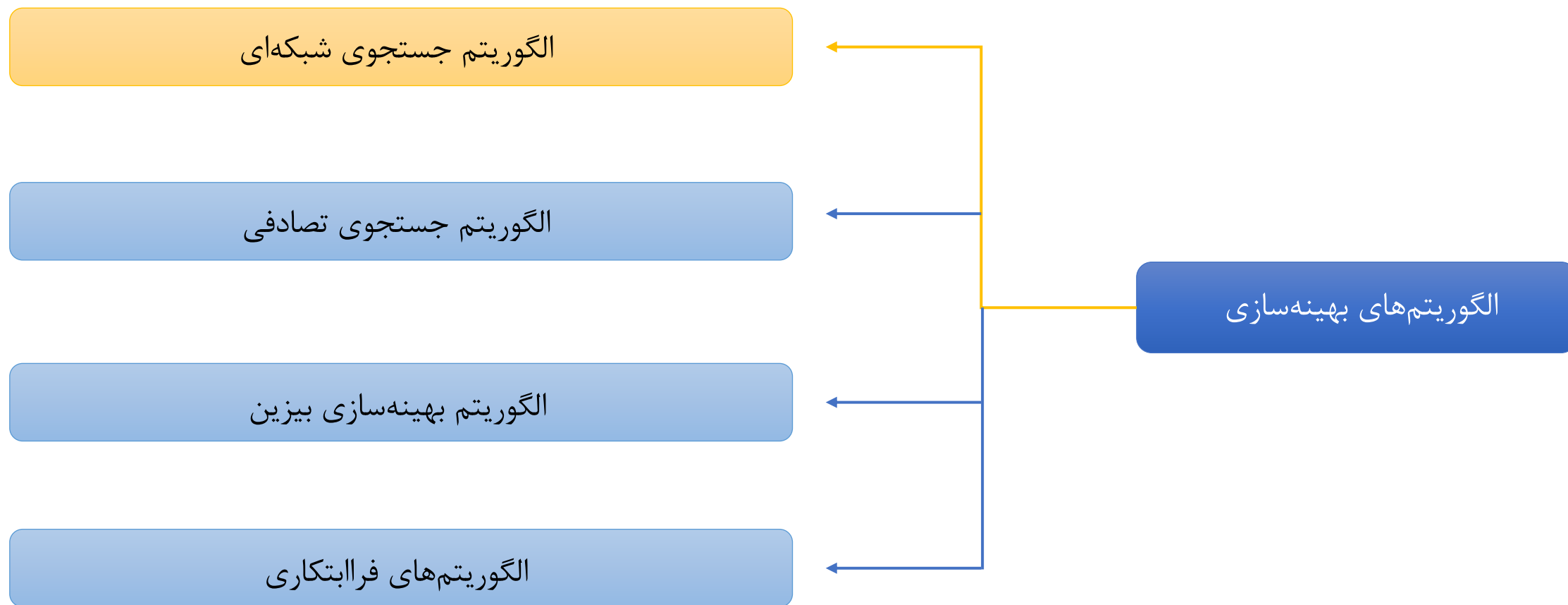






- داده‌ی آموزش (train)

- داده‌ی آزمایش (test)



# آیا این مدل واقعا کمک کننده بود؟



[11]

- Extreme gradient boosting (XGBoost)
- ماشین بردار پشتیبان (SVM)
- جنگل تصادفی (RF)
- شبکه عصبی عمیق (DNN)
- K-نزدیک ترین همسایه (KNN)
- نایو بیز (NB)
- Elastic Net

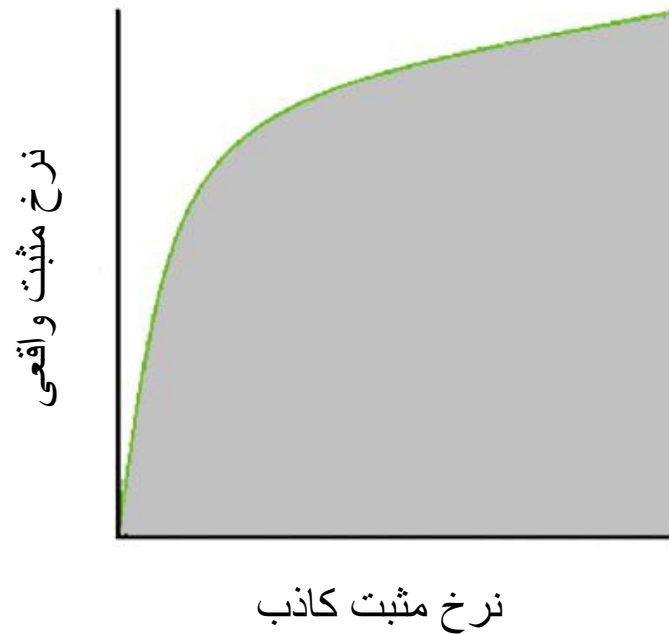


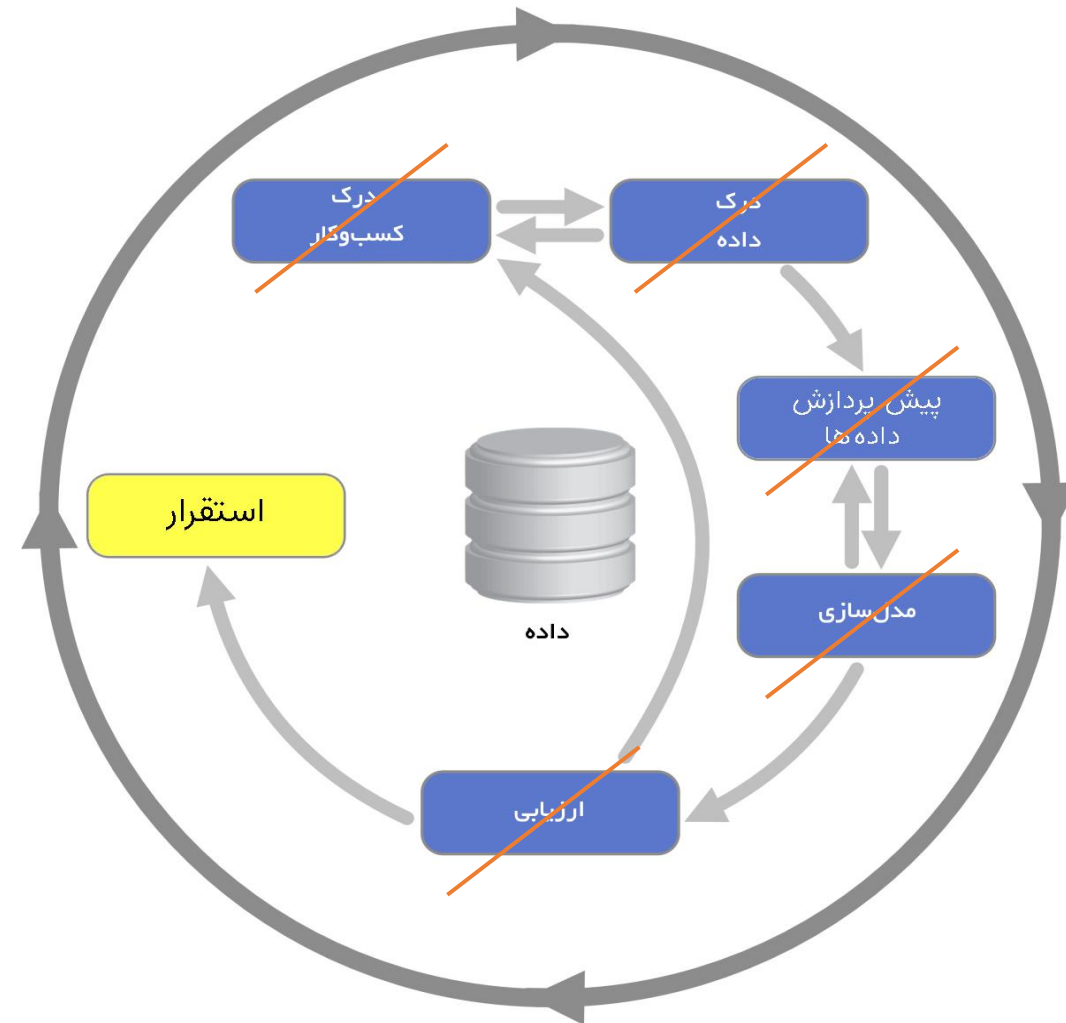
- مساحت زیر منحنی (AUC)

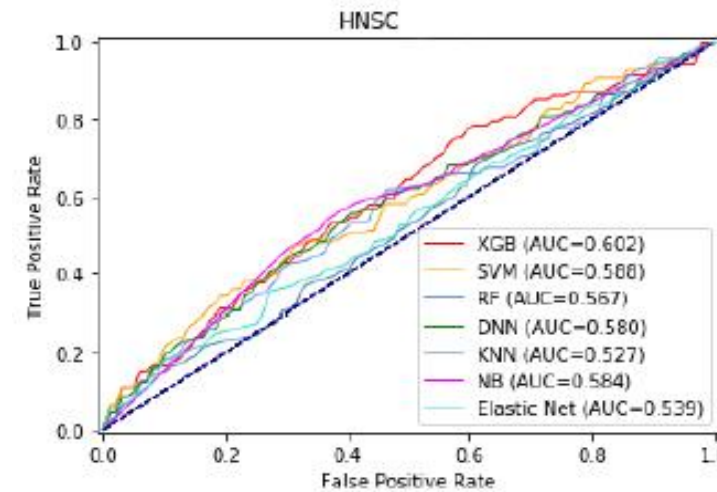
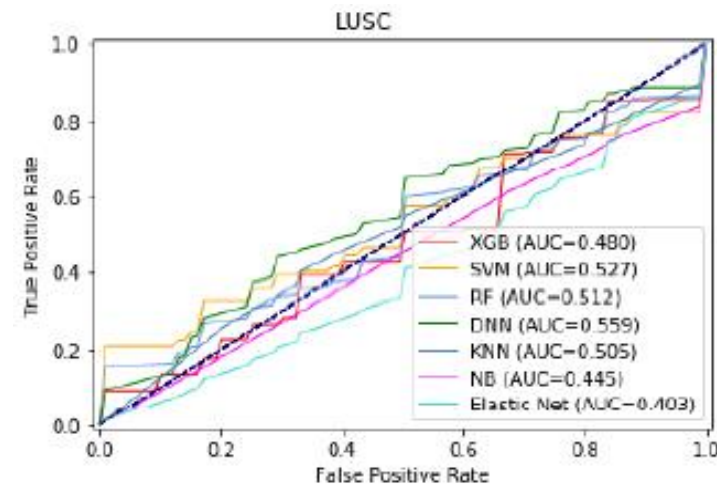
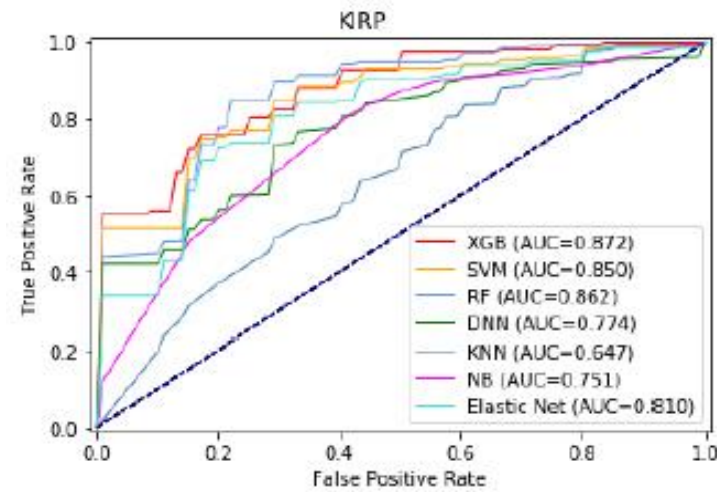
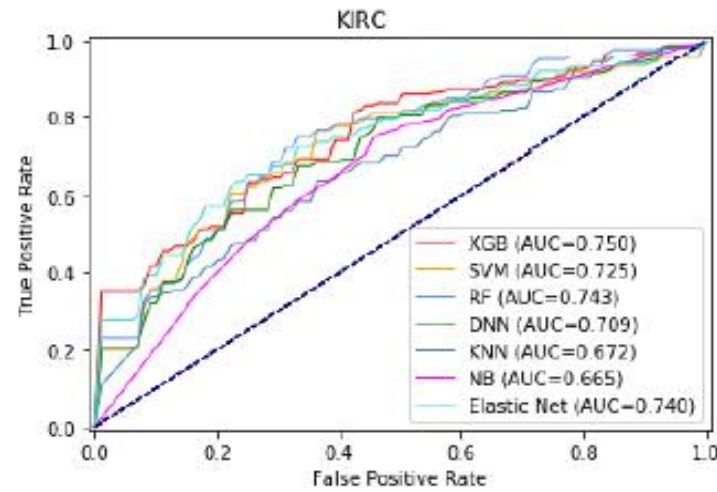
- دقت

- پوشش

- صحت





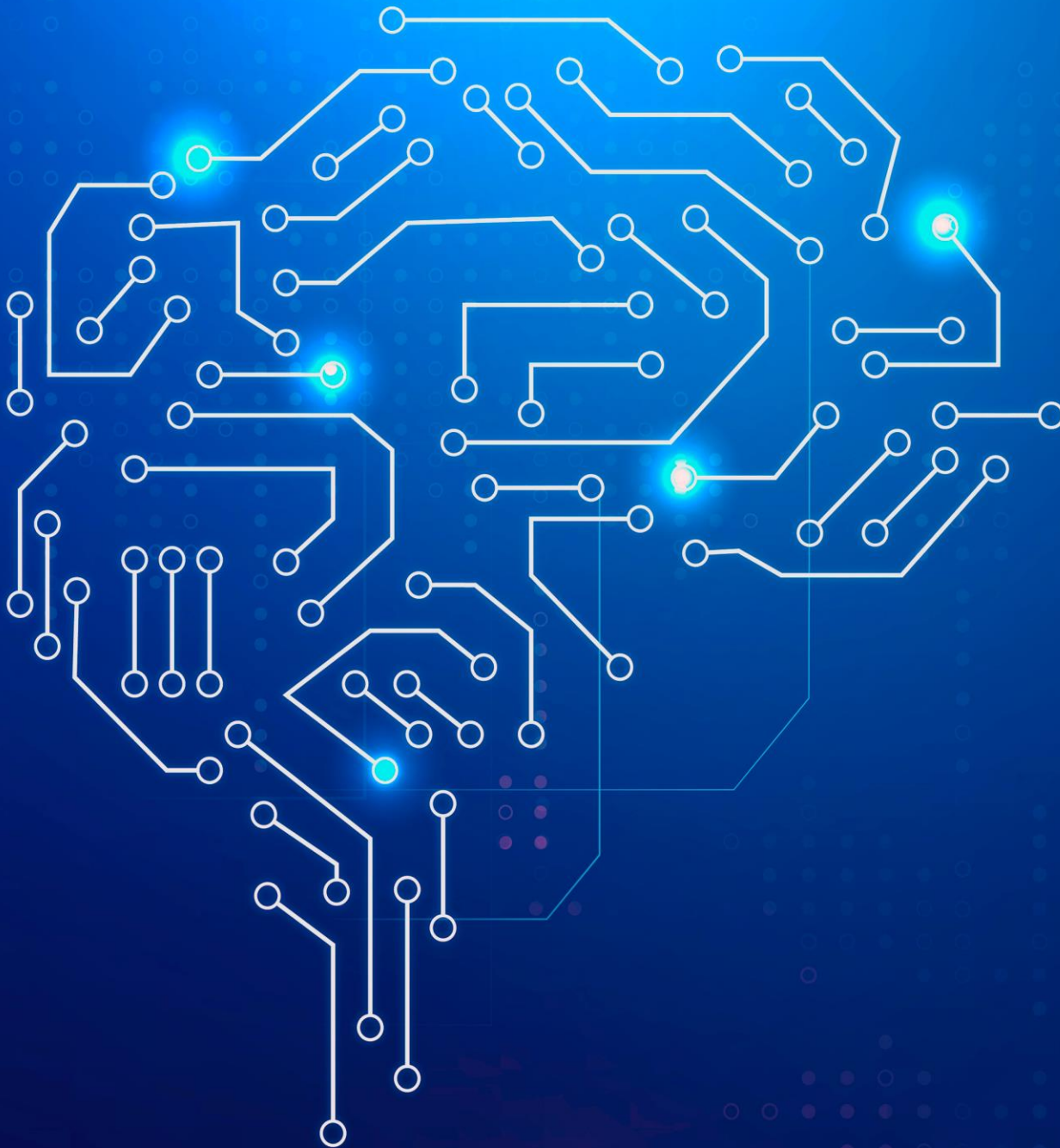


١. B. Ma, F. Meng, G. Yan, H. Yan, B. Chai, and F. Song, "Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data," Computers in Biology and Medicine, vol. 121, p. 103761, Jun. 2020, doi: 10.1016/j.combiomed.2020.103761.
٢. <https://bernardmarr.com/should-i-choose-machine-learning-or-big-data/>
٣. <https://dehkhodaedu.com/%DB%8C%D8%A7%D8%AF%DA%AF%DB%8C%D8%B1%DB%8C-%D8%B2%D8%A8%D8%A7%D9%86-%D8%A8%D8%A7-%D9%81%D9%84%D8%B4-%DA%A9%D8%A7%D8%B1%D8%AA/>
٤. <http://www.dadehara.com/%D8%B7%D8%A8%D9%82%D9%87-%D8%A8%D9%86%D8%AF%DB%8C-%DA%86%DB%8C%D8%B3%D8%AA.html>
٥. <http://wikichejoor.ir/how-to-win-at-tug-of-war/>
٦. <https://blockgeni.com/introduction-to-xgboost-loss-functions/>
٧. [https://blog.faradars.org/wp-content/uploads/2018/09/CRISPDM\\_Process\\_Diagram.png](https://blog.faradars.org/wp-content/uploads/2018/09/CRISPDM_Process_Diagram.png)
٨. <https://cdn01.zoomit.ir/2018/6/66139b7f-c46c-42bb-bb27-3e562c8af0e8.jpg>
٩. [https://www.data4logic.net/images/data\\_preprocessing.png](https://www.data4logic.net/images/data_preprocessing.png)



۱۰. <https://www.flaticon.com/premium-icon/icons/svg/901/901030.svg>
۱۱. [https://www.freepik.com/free-vector/thoughtful-woman-with-laptop-looking-big-question-mark\\_13330330.htm#query=confused%20woman&position=10&from\\_view=keyword](https://www.freepik.com/free-vector/thoughtful-woman-with-laptop-looking-big-question-mark_13330330.htm#query=confused%20woman&position=10&from_view=keyword)
۱۲. [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)
۱۳. <https://amanjacademy.com/types-of-machine-learning-with-example/>
۱۴. [https://amanjacademy.com/complete-supervised-machine-learning-guide/#\\_Supervised\\_Learning](https://amanjacademy.com/complete-supervised-machine-learning-guide/#_Supervised_Learning)
۱۵. <https://www.matlabdl.com/classification-and-classification-algorithms/>
۱۶. <https://fa.wikipedia.org/wiki/%D8%A8%D9%88%D8%B3%D8%AA%DB%8C%D9%86%DA%AF>
۱۷. <https://www.analyticsvidhya.com/blog/2021/04/best-boosting-algorithm-in-machine-learning-in-2021/#:~:text=Boosting%20can%20be%20referred%20to,machine%20learning%20community%20for%20years.>
۱۸. <https://shahaab-co.com/mag/edu/ml/xgboost-tutorial/>
۱۹. <https://chistio.ir/%D9%81%D8%B1%D8%A2%DB%8C%D9%86%D8%AF-%DA%A9%D8%B1%DB%8C%D8%B3%D9%BE-crisp-%D9%BE%D8%B1%D9%88%DA%98%D9%87-%D8%AF%D8%A7%D8%AF%D9%87-%DA%A9%D8%A7%D9%88%DB%8C/>

- ٢٠. <https://towardsdatascience.com/pre-processing-and-training-data-d16cc12dfbac>
- ٢١. [https://en.wikipedia.org/wiki/Training,\\_validation,\\_and\\_test\\_sets](https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets)
- ٢٢. <https://www.analyticsvidhya.com/blog/2021/06/tune-hyperparameters-with-gridsearchcv/#:~:text=Grid%20Search%20uses%20a%20different,the%20number%20of%20hyperparameters%20involved.>
- ٢٣. <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>



با تشکر از حسن توجه شما