1) In the **iris dataset**, you have the iris feature (sepal-length, sepal width, petal length and petal width), and you want to determine which the species (target) be: setosa, vigrinica or versicolor? Then, what kind of the ML task should you choose for this?

   A. Multi-class regression
   B. Binary classification
   C. Multi-class classification
   D. Clustering.

2) Look at the following chart
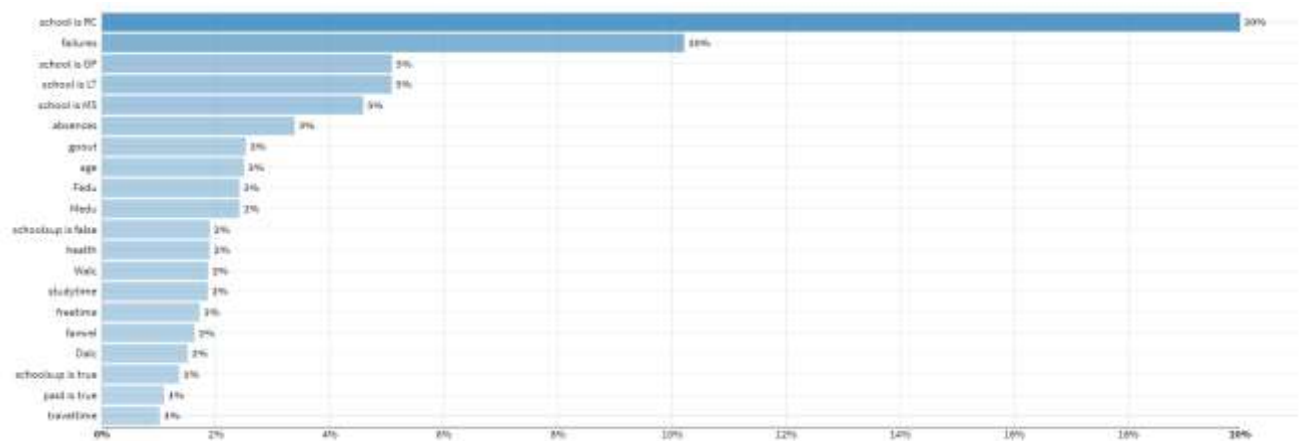
   Which of the following ML algorithms can generate the chart above:

   A. Extra Tree Classifier
   B. Gradient Boosting Tree Classifier
   C. Logistic Regression
   D. Random Forest
   E. SVM (Support Vector Machine)
   F. Knn (k nearest neighbors)

3) The charts in the question 2 is called _____, the top 3 features have the strongest relationship to the respone is _____, _____, and _____; while the least important is _____.

4) This formula

$$\frac{1}{n} \sum_{k=1}^{n} \left| \frac{y_{true}(k) - y_{pred}(k)}{y_{true}(k)} \right|$$

   named the _____.

5) The MAE and MAPE can be used to evaluate the _ problem.

   A. Regression
   B. Classification
   C. Time series forecasting
   D. Clustering

6) Named at least 6 ML algorithms that can be used in the classification problem.

   ....
   ....
   ....
   ....
   ....

7) Named at least 4 ML algorithms that can be used in segmentation / clustering.

....

....

....

8) Named at least 5 algorithms that can be used in time-series forecasting.

....

....

....

....

9) Which of the following score is defined be the harmonic mean of precision and recall?

A. f1-score   C. accuracy

B. ROC-AUC   D. Cross-entropy.

10) Look at the "box-plot of the grade by shool_id", which of the following statements be True
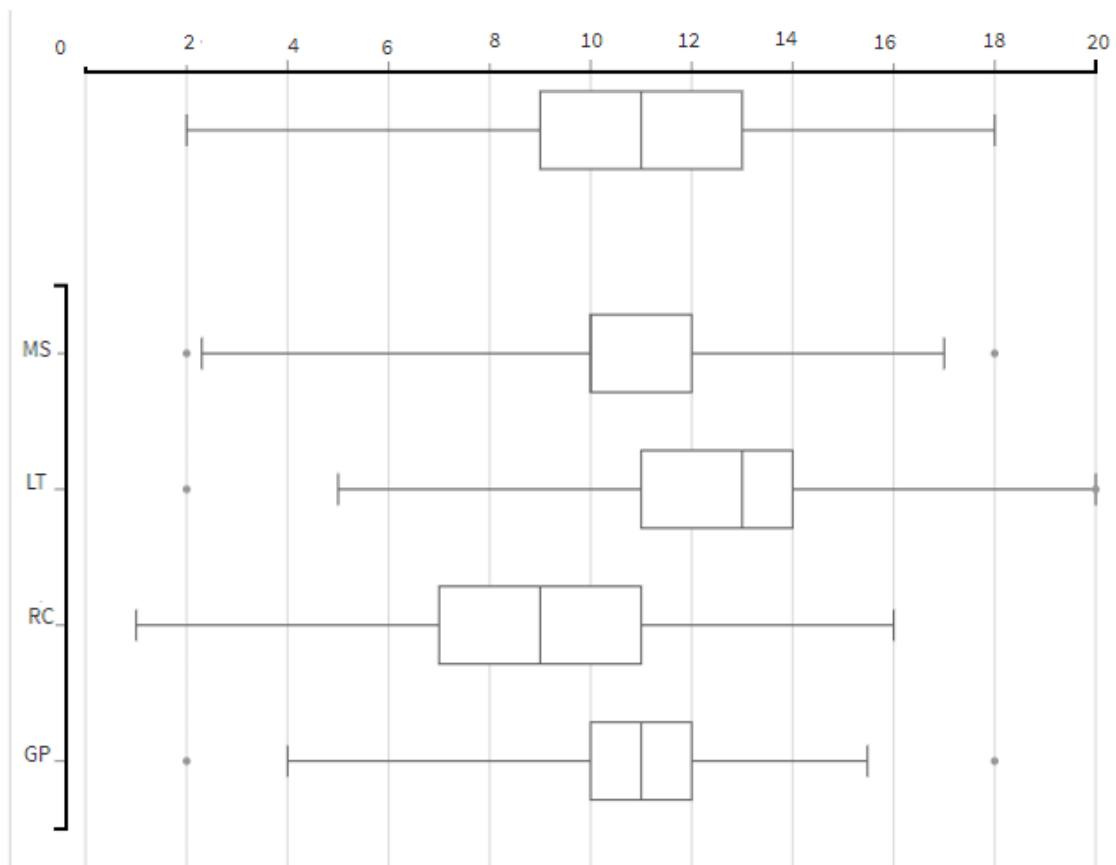
A. The school_id has the best grade is LT.
B. The median of grade at LT is about 13, the highest be 20 while the minumum is zero.
C. 50% student's grade at RC is in (6.5, 11), lower than the 50%_lowest grade at LT.
D. None of above
E. All of above

11) What are some reasons that make you generate the "polynomial combination of features" (like $ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2$) or "pairwise linear combinations" (like ax1 + bx2).?
   A. Improve the model peformance
   B. Recover the unseen feature
   C. Uncover the new relationships between the features and the target.
   D. Reduce the resulting number of features
   E. Ensure that the numerical features are properly rescaled.
   F. Reduced training time and doing better handling of missed_values.

12) Which of the following algorithms is belong to suppervised learning?
   A. Classification
   B. Principal Component Analysis
   C. K-Mean clustering
   D. Regression

13) Which of the following algorithms is belong to suppervised learning?
   A. Classification
   B. Principal Component Analysis
   C. K-Mean clustering
   D. Regression

14) The difference between "suppervised learning" and "un-suppervised learning" is that
   A. Supervised learng used labeled data as input, while the others used the un-labeled data.
   B. Un-supervised learng used labeled data as input, while the others used the un-labeled data.
   C. Both of them are True
   D. Both of them are False

15) Which of the following statements is True in linear regression, with the regression line defined as $y.hat = wx + b$
   A. The term w is called slope of the line while the point b is called y-intercept, where the line across the y-axis. The $y.hat$ is the predicted value of y (the dependent variable)
   B. The term w is called slope of the line while the point b is called y-intercept, where the line across the y-axis. The $y.hat$ is the predicted value of y (the independent variable)
   C. The term w is called intercept of the line while the point b is called y-slope, where the line across the y-axis. The $y.hat$ is the predicted value of x (the independent variable)
   D. The term w is called slope of the line while the point b is called y-intercept, where the line across the y-axis. The $y.hat$ is the predicted value of x (the dependent variable)

16) Which statements is True when discuss the difference between the K-Means and hierarchical clustering ?
   A. In hierarchical clustering, we aim to first understand the relationship between the clusters visually, and then determine the number of clusters, or hierarchy level, that best portrays the different groupings
   B. In hierarchical clustering, we aim to set K to an optimal number, creating just the right number of clusters, where adding more clusters would no longer provide a sufficient decrease in variation.
   C. They are the same.
   D. Both of A and B are True.

17) Which of the following algorithms can be used in both classification and regression?
   A. ExtraTree: ExtraTreeClassifier and ExtraTreeRegressor.
   B. RandomForest: RandomForestClassifier and RandomForestRegressor
   C. Ridge-Regression and Lasso-Regression

18) Which examples be True when mention the data-leakage?
   A. Predict the passed / failed results based on the grade, which directly inferred the passed or failed of a student.
   B. Predict the price of Wednesday by using the window-functions of the average-price of the previous 3 days: Tuesday, Monday and last Sunday.
   C. Predict the number of customers on Tuesday by using the number of sales on Tuesday and weather on Tuesday.
   D. All of above are True.
   E. Only A and C are True.

19) This formula

$$\frac{1}{n} \sqrt{\sum_{k=1}^{n} \left(\log(1 + y_{\text{true}}(k)) - \log(1 + y_{\text{pred}}(k))\right)^2}$$
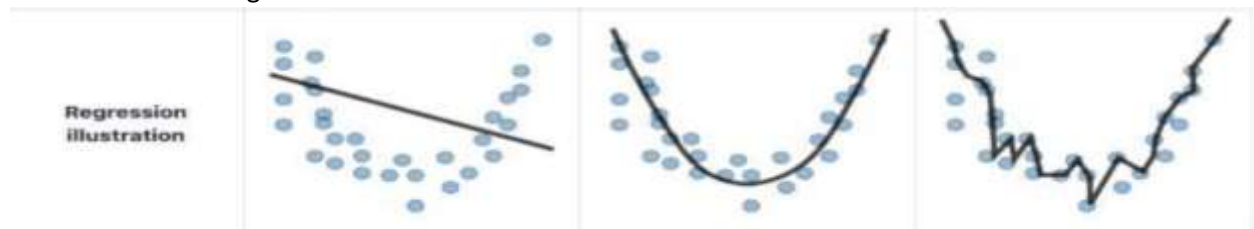
named the _____.

20) In the regression problem, the _____ R2 score, also known as the _____ coeficience, the better the model performance. The highest R2 score is equal to ___.

21) The word IC in the evaluation metrics, which used in time-series forecasting and regression: AIC and BIC; stands for _____ _____.

22) In regression problem, which of the following concepts is the squared differences between the observed dependent variable and its mean?
   A. SST (Sum of Square Total)
   B. SSR (Sum of Square due to Regression)
   C. MSE (Mean Square Error)
   D. MAE (Mean Absolute Error)
   E. RMSLE (Root Mean Squared Logarithm Error)
   F. BIC (Bayesian Information Criterion)

23) Look at the following chart

Regression illustration

(name) :   ....

(symptom):   ....

⇨ Mapped the correct names and its symptoms in the 2 following sets:

Group of name :
i)   Overfiting
ii)   Underfiting
iii)   Just-right

Group of symptom:
A. The training error is slightly lower than the test error.
B. Very low training error, high variance. The training error is much lower than the test error
C. High training error, high bias. Training error is close to the test error.

24) In the `sklearn.decomposition.PCA`, which statements is True about the parameter `n_components`:
   A. This can be a integer
   B. This can be a float
   C. If this is a float, this must be in (0, 1) which reflect the threshold that expained the percentage of the variance of the number of principal components must be exceeded.
   D. This can be set at "mle" and hence you must set the "svd_solver" be "full" or "auto".
   E. `n_components == min(n_samples, n_features) - 1`
   F. All of them are True
   G. Only A, B, C and E are True.