

Question 1
Answer saved
Marked out of 3.00
Flag question

Time left 0:31:40

There are many reasons for reducing the number of attributes in a dataset; some of them are. (Select All Correct Answers)

- ☒ a. Removing irrelevant attributes improves performance of algorithms
- ☐ b. Reducing the number of table reduces execution time of algorithms
- ☒ c. Reducing the number of attributes reduces execution time of algorithms
- ☐ d. Reducing the number of attributes reduces space requirement for algorithms

Next page

Question 2
Answer saved
Marked out of 2.00
Flag question

Time left 0:31:19

The selling price of a house depends on many factors. For example, it depends on the number of bedrooms, number of kitchen, number of bathrooms, the year the house was built, and the square footage of the lot. Given these factors, predicting the selling price of the house is an example of _____ task.

- ☐ a. Simple Linear Regression
- ☒ b. Multiple Linear Regression
- ☐ c. Multilabel Classification
- ☐ d. Binary Classification

Previous page

Next page

Question 3
Answer saved
Marked out of 10.00
Flag question

Time left 0:30:39

Match the question with the answers

Divides the dataset into partitions without changing the order of the examples, i.e., subsets with consecutive examples are created

Stratified sampling

Builds random subsets and ensures that the class distribution in the subsets is the same as in the whole dataset. For example, in the case of a binominal classification

Linear sampling

Builds random subsets of the dataset, i.e., examples are chosen randomly for making subsets.

Shuffled sampling

Previous page

Next page

→ Tugas 1 CD & ML Data Preprocessing

Jump to...

Question 4
Answer saved
Marked out of 5.00
Flag question

Time left 0:30:06

The performance vector delivered by this operator has numerous classification performance criteria for example (Select ALL correct answers)

- ☐ a. Kappa and spearman rho
- ☒ b. Accuracy
- ☒ c. Classification Error

Previous page

Next page

→ Tugas 1 CD & ML Data Preprocessing

Jump to...

Time left 0:29:54

Question 5

Answer saved
Marked out of
2.00

Flag question

Which of the following statements about model bias and variance are true? (Select ALL correct answers)

- ☐ a. Testing error more reliably indicates the optimal model compared to training error
- ☐ b. High bias models typically are underfit
- ☒ c. It is usually possible to minimize error from both bias and variance simultaneously
- ☒ d. Overfitting tends to lead to models with high variance

Previous page

Next page

Time left 0:29:46

Question 6

Answer saved
Marked out of
2.00

Flag question

Which algorithm is most sensitive to data that is not on the same scale and has not been normalized? (Select one)

- ☐ a. Neural Network
- ☒ b. Naive Bayes
- ☐ c. k-NN
- ☐ d. Linear Regression

Clear my choice

Previous page

Next page

Time left 0:29:32

Question 7

Answer saved
Marked out of
10.00

Flag question

Lakukanlah klasifikasi data `X_test` dari model yang telah terbentuk dan simpanlah hasil prediksi ke dalam variable `y_pred`

- ☐ a. `y_pred = decisiontree.fit(X_train, y_train)`
- ☐ b. `y_pred = decisiontree.fit(X_test)`
- ☐ c. `y_pred = decisiontree.fit(X_test)`
- ☐ d. `y_pred = model.predict(X_test, y_test)`
- ☒ e. `y_pred = model.predict(X_test)`

Pilihan jawaban ialah :

- ☐ a.

```
decisiontree = DecisionTreeClassifier(criterion="entropy",
                                     random_state=0, max_depth=100,
                                     min_samples_split=5, min_samples_leaf=2,
                                     min_weight_fraction_leaf=0, max_leaf_nodes=50,
                                     min_impurity_decrease=0)
```
- ☐ b.

```
decisiontree = DecisionTreeClassifier(criterion="entropy",
                                     random_state=0, max_depth=5,
                                     min_samples_split=50, min_samples_leaf=2,
                                     min_weight_fraction_leaf=0, max_leaf_nodes=100,
                                     min_impurity_decrease=0)
```
- ☐ c.

```
decisiontree = DecisionTreeClassifier(criterion="gini",
                                     random_state=0, max_depth=100,
                                     min_samples_split=5, min_samples_leaf=2,
                                     min_weight_fraction_leaf=0, max_leaf_nodes=50,
                                     min_impurity_decrease=0)
```
- ☐ d.

```
decisiontree = DecisionTreeClassifier(criterion="gini",
                                     random_state=0, max_depth=5,
                                     min_samples_split=2, min_samples_leaf=50,
                                     min_weight_fraction_leaf=0, max_leaf_nodes=100,
                                     min_impurity_decrease=0)
```
- ☒ e.

```
decisiontree = DecisionTreeClassifier(criterion="entropy",
                                     random_state=0, max_depth=5,
                                     min_samples_split=2, min_samples_leaf=50,
                                     min_weight_fraction_leaf=0, max_leaf_nodes=100,
                                     min_impurity_decrease=0)
```

Question 9
Answer saved
Marked out of
10.00
Flag question

Lakukan splitting dataset dengan rasio 30% dari total keseluruhan data X dan y menjadi data uji dan 70% menjadi data latih

- ☐ a.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```
- ☐ b.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.7, random_state=42)
```
- ☒ c.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```
- ☐ d.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.7, random_state=42)
```
- ☐ e.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Previous page

Next page

Time left 0:28:30

Question 10
Answer saved
Marked out of
2.00
Flag question

What are some possible uses of Naive Bayes? (Select ALL correct answers)

- ☒ a. Fast-running spam filter based on text of e-mails
- ☐ b. Simple recommendation system
- ☐ c. Precise demand-forecasting for electricity

Clear my choice

Previous page

Next page

100%

Flag question

Time left 0:28

| | |
|---|-------------------------------|
| 1 | 1,5,1,3,5,1,4,0,2,Iris-setosa |
| 2 | 2,4,9,3,0,1,4,0,2,Iris-setosa |
| 3 | 3,4,7,3,2,1,3,0,2,Iris-setosa |
| 4 | 4,4,6,3,1,1,5,0,2,Iris-setosa |
| 5 | 5,5,0,3,6,1,4,0,2,Iris-setosa |
| 6 | 6,5,4,3,9,1,7,0,4,Iris-setosa |

Baca dan simpanlah file tersebut ke dalam variable irisDataset menggunakan pandas

- ☐ a. `irisDataset = pd.read_csv("iris.csv", sep='\t')`
- ☐ b. `irisDataset = pd.read_csv("iris.csv", sep=',')`
- ☐ c. `irisDataset = pd.read_csv("iris.csv", sep=';')`
- ☒ d. `irisDataset = pd.read_csv("iris.csv", sep=',', skiprows=0)`
- ☐ e. `irisDataset = pd.read_csv("iris.csv", sep=';', skiprows=0)`

Question 12

Answer saved
Marked out of 20.00

Flag question

Match the questions and the answers!

The set of examples that the model

The training data set

The Discretize By Binning operator

Used for converting the type of all the numeric attributes (A2, A3, A8, A11, A14, A15) to nominal

Principal components are artificial attributes

Represent (or try to represent as closely as possible) the original attributes by accounting for most of the variance in the original attributes

Learned on the training set and is then applied on the testing set

The Model

Operator provides the testing dataset through the test set port of the Testing sub-process

The Split validation

Operator can be used for performance evaluation of all types of learning tasks

Performance (Classification) operator or Performance (Binominal Classification) operator

The size of two subsets can be adjusted by parameters like

split, split ratio, training set size, and test set size parameters

Operator should be used for Binominal Classification tasks

Performance (Binominal Classification) operator

The Filter Examples operator is

Used in this process to simply remove the examples with missing values

Principal component analysis (PCA)

An attribute reduction procedure

Time left 0:27:11

Question 13

Answer saved
Marked out of 2.00

Flag question

Which of the following methods do we use, to find the best fit line for data in Linear Regression?

- ☐ a. Logarithmic Loss
- ☒ b. Least Square Error
- ☐ c. Least Square Error & Logarithmic Loss
- ☐ d. Maximum Likelihood

Clear my choice

Previous page

Next page

Answer saved
Marked out of
10.00
Flag question

Pertahankan potongan kode berikut ini.

Time left 0:26:54

```
irisDataset = irisDataset.to_numpy()
print("Ukuran Matrix : ", irisDataset.shape)
print(irisDataset)
```

```
Ukuran Matrix : (150, 5)
[[5.1 3.5 1.4 0.2 0. ]
 [4.9 3.  1.4 0.2 0. ]
 [4.7 3.2 1.3 0.2 0. ]
```

Masukan 3 kolom pertama dari 5 kolom dan seluruh baris data pada matrix irisDataset ke dalam variable X

- ☒ a. `X = irisDataset[:, 0:3]`
- ☐ b. `X = irisDataset[0:3, 0:3]`
- ☐ c. `X = irisDataset[:, 0:4]`
- ☐ d. `X = irisDataset[0:4, :]`

Question 15
Answer saved
Marked out of
2.00
Flag question

You are given reviews of movies marked as positive, negative, and neutral. Classifying reviews of a new movie is an example of..

- ☒ a. Supervised learning
- ☐ b. Unsupervised Learning
- ☐ c. Reinforcement Learning
- ☐ d. None of these

Clear my choice

Previous page

Finish attempt ...