ML4SecQuiz#2-PreMidSemCoverageExceptHE-3rdApril2023

p22cs013@coed.svnit.ac.in Switch account

Your email will be recorded when you submit this form

ML4SecQuiz#2-PreMidSemCoverageExceptHE-3rdApril2023

As mentioned earlier in Sec 1

Predicting how much a used car would sell for given historical data on recent used car sales in the area is an example of ML task ___________

- ○ principal component analysis
- ○ classification
- ○ regression
- ○ clustering

__________ approaches to security anticipate and eliminate vulnerabilities in the cyber system, while remaining prepared to defend effectively and rapidly against attacks, and needs ______________.

- none of these options
- Reactive, higher-level adaptive cyber defense systems
- Reactive
- Reactive, firewalls and IDSs
- Proactive, higher-level adaptive cyber defense systems
- Proactive, firewalls and IDSs
- Other:

In anonymization technique for privacy preservation, ____________ data concerns with what data needs to be removed from the anonymized view because it would lead to identification? For example, names or unique identification numbers.

- sensitive
- identifier
- none of these
- quasi-identifier

Consider an anonymization design here that shows the data anonymised to achieve k-anonymity of k = ________, achieved by generalising some quasi-identifier attributes.

Name	Postcode	Age	Gender	Disease
*	SW1 *	22	Male	Cardiovascular
*	SW1 *	23	Male	Respiratory
*	SW1 *	18	Male	No Illness
*	NW10 *	47	Female	Cancer
*	NW10 *	42	Female	No Illness
*	NW10 *	56	Female	Cardiovascular
*	E17 *	23	*	Respiratory
*	E17 *	29	*	Liver
*	E17 *	18	*	Cancer

- 4
- 2
- 3
- 5

Fig shows a typical data of a medical application published while devising anonymization approach for PPML. Here, the downside is that ___________

Published Data

#	Zip	Age	Nationality	Condition
1	13053	28	Indian	Heart Disease
2	13067	29	American	Heart Disease
3	13053	35	Canadian	Viral Infection
4	13067	36	Japanese	Cancer

#	Name	Zip	Age	Nationality
1	John	13053	28	American
2	Bob	13067	29	American
3	Chris	13053	23	American

Voter List

- there is a data leak because sensitive data "Nationality" can be inferred from the <zip, age, nationality> if there is a single tuple pertaining to the latter
- there is NO data leak because sensitive data "condition" cannot be inferred from the <zip, age, nationality> if there is a single tuple pertaining to the latter
- there is NO data leak because sensitive data "Age" cannot be inferred from the <zip, age, nationality> if there is a single tuple pertaining to the latter
- there is a data leak because sensitive data "condition" can be inferred from the <zip, age, nationality> if there is a single tuple pertaining to the latter.

______________ allows many privacy-enhancing strategies to allow multiple input sources to train ML models cooperatively without exposing their private data in its original form.

- Homomorphic encryption
- Zero-knowledge proofs
- Federated learning
- Ensembling learning

__________ concerns with how a company protects the data from un-authorized access or corruption, whereas __________ concerns with controlling extent, timing, and circumstances of sharing one's own data with others.

- Data privacy, Data security
- Data privacy, Data privacy
- Data security, Data security
- Data security, Data privacy

Consider that in an application data was collected for an ML algorithm. This data was for example of the kind as follows: Input could be anything, for example, *email messages, pictures, or sensor measurements*. Outputs were supposed to be usually *real numbers, or labels (e.g. "spam", "not_spam", "cat", "dog", "mouse", etc). In some cases, outputs are vectors (e.g., four coordinates of the rectangle around a person on the picture), sequences (e.g. ["adjective", "adjective", "noun"] for the input "big beautiful car")*, or have some other structure. Then the ML algorithm must be ____________

- Principle Component Analysis
- Basic Apriori algorithm.
- KNN
- Decision Tree

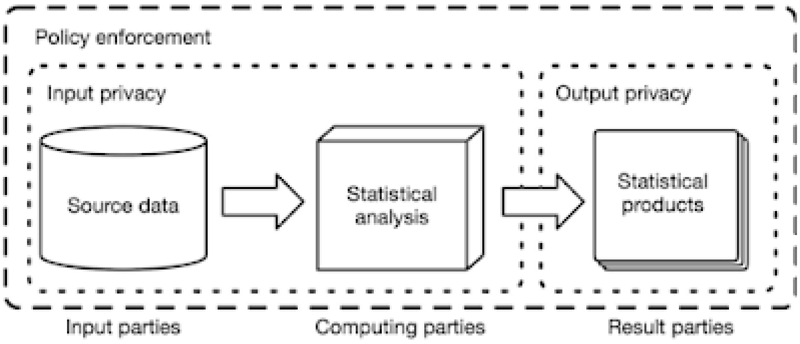
In ____________, typically there will not be any false positives and gives instant results (time to value), whereas in ________________ there can be false positives and requires training.

- pattern recognition, anomaly detection
- anomaly detection, pattern recognition,
- pattern recognition, pattern recognition,
- anomaly detection, anomaly detection

The purpose of k-anonymity is to ensure the two categories of data viz. _________ data (*e.g. name, zip code, gender, etc.*) and ___________ data (*e.g. health records, prescriptions, financial information, passwords, etc.*) **cannot be connected** to one another, to protect against hackers or malicious parties using 're-identification.'

- identifying, sensitive
- identifying, identifying
- sensitive, sensitive,
- sensitive, identifying

Consider the figure shown here. One of the inferences from the figure is that _______________.



- input privacy guarantees output privacy
- multiple privacy goals co-exist in a system, with four stakes, typically.
- input privacy guarantees privacy of statistical analysis
- privacy of statistical analysis is the core of data privacy
- none of these

Anomaly detection focusses on ________________ with the observation that there can be an _________ including even those derived from hypothetical data that do not exist in the training or testing datasets.

- ◯ tracking dis-similarities to identify patterns, infinite number of anomalous patterns as patterns
- ◯ tracking similarities to identify patterns, infinite number of anomalous patterns
- ◯ tracking similarities to identify outliers, infinite number of anomalous patterns as anomalies
- ◯ tracking dis-similarities to identify anomalies, infinite number of anomalous data
- ◯ Other:

The use of federated learning in applications involving machine learning represents the following approach to privacy preservation viz. ____________.

- ◯ none of these
- ◯ designing ML specific approaches for privacy preservation.
- ◯ augmenting conventional ML with different strategies} that protect data privacy.
- ◯ both of these

The use of homomorphic encryption algorithms in applications involving machine learning represents the following approach to privacy preservation viz. ____________.

- ◯ none of these
- ◯ both of these
- ◯ augmenting conventional ML with different strategies} that protect data privacy.
- ◯ designing ML specific approaches for privacy preservation.

The use of zero knowledge proofs in applications involving machine learning represents the following approach to privacy preservation viz. _____________.

- both of these
- augmenting conventional ML with different strategies} that protect data privacy.
- designing ML specific approaches for privacy preservation.
- none of these

Model built using just _________ gets highly biased to the dataset and may ______ the training dataset; whereas model built with ___________; though performs much better than the model trained using entire dataset; (however,) when trained for long time, ___________

- training dataset, underfit, training & validation data set both, does not affect the model
- training dataset, underfit, validation data set, the model gets biased.
- training dataset, overfit, validation data set, does not affect the model
- training dataset, overfit, training & validation data set both, the model gets biased.
- Other:

The focus in k-anonymization is to change data in such a way that for each tuple in the resulting table there are atleast ___________ other tuples with the same value for the quasi-identifier.

- k-2
- k+1
- k
- k-1

Consider an anonymization design here. This is an example of _____ -anonymization.

#	Zip	Age	Nationality	Condition
1	130**	< 40	*	Heart Disease
2	130**	< 40	*	Heart Disease
3	130**	< 40	*	Viral Infection
4	130**	< 40	*	Cancer

? -anonymized

- 2
- 5
- 4
- 3

______________ is an example of Probability density and mass function estimation problems and use _____________ ML algorithm.

- Malware detection, BIRCH
- Email Spam Detection, SVM
- Market Basket Analysis, DBSCAN

In anonymization technique for privacy preservation, ______________ data concerns with what data could lead to people being **re-identified**, **even if identifiers are removed because of individuals' unique combination of attributes - e.g.** , age, zip code, start year, education, marital status, location.

- identifier
- none of these
- sensitive
- quasi-identifier

Normally , there is a split of ________ for training and _______ for testing dataset.

- 50%, 50%
- 40%, 60%
- 80%, 20%
- 20%, 80%

_____________ and __________ represent non-cryptographic approaches to achieve privacy preservation.

- Perturbation, Anonymization
- Homomorphic encryption, Federated learning
- Zero Knowledge Proofs, Ensemble learning
- Secure Multi-party Computation, Zero knowledge proofs

The use of ensemble learning in applications involving machine learning represents the following approach to privacy preservation viz. _____________.

- none of these
- augmenting conventional ML with different strategies} that protect data privacy.
- designing ML specific approaches for privacy preservation.
- both of these

__________ focusses on identifying similarities, that is, patterns extracted through pattern recognition ____________ the observed data used to train the algorithm.

- anomaly detection, must NOT be strictly derived from
- pattern recognition, must be strictly derived from
- pattern recognition, must NOT be strictly derived from
- anomaly detection, must be strictly derived from

Threats due to data sets in Privacy -Preserving Machine Learning is due to ____________

(a) probability of large sets of data - used for training - becoming available publicly
(b) criticality of data privacy in domains like healthcare or intrusion detection systems
(c) probability of profit making by identifying people or other valuable information based on the stolen data
(d) the ML models themselves pose a vulnerability since sensitive data may be extracted from them

- (d)
- (c)
- (b) and (c)
- (a) and (c)
- (a)
- (a) and (b)
- (b)
- (a), (b), (c), (d)
- (b) and (d)

Helping with when one is looking for a particular product online but couldn't find it through traditional search methods OR similarity matching to present present other relevant products are examples of __________ and could use ______________ algorithm

- Classification, SVM
- Regression, LASSO/Ridge
- Clustering, KMeans
- Similarity Matching, KNN

The goal of ____________ is to prevent a situation where even if one removes the direct uniquely identifying attributes from a table, there are some fields that may still uniquely identify some individual.

- Federated learning
- Homomorphic Encryption
- Anonymization-based approaches
- Zero-knowledge proofs

An _________ for banks and financial institutions is a _____________ ML based application to develop credit rating for those who do not have a credit cards and hence no formal credit score.

- Smart Data Labelling, Supervisory ML-based
- Smart Data Labelling, Un-Supervisory ML-based
- Ethical credit scoring system, Supervisory ML-based
- Ethical credit scoring system, Un-Supervisory ML-based

__________ is the assurance that a malicious party will not reverse-engineer the training data - although gathering information about training data and model is more difficult than that for the data.

- Privacy of the input data
- Privacy of the model
- Privacy of the output data
- Data privacy in training

In choosing k, in k-anonymization, k=1 and k=n are __________ (for a data set of size n). This is so, because the former (i.e. k=1) provides _____________, whereas the latter (i.e. k=n) provides _____________ but, does not retain any utility - other than about very basic info like the size of the data set.

- generally useless, no anonymity, highest security
- generally useful, highest anonymity, highest security
- generally useful, highest security, highest anonymity
- generally useless, no security, highest anonymity

- Option 1

In traditional computer programming, outputs or decisions are ___________, whereas machine learning (also) __________ as input to build a decision model.

- uses data, pre-defined by the programmer,
- pre-defined by the programmer, uses data
- pre-defined by the programmer, pre-defined by the programmer,
- uses data, uses data

Fig shows a typical data of a medical application published while devising anonymization approach for PPML. Here, the sensitive data attribute(s) is/are ___________

#	Zip	Age	Nationality	Name	Condition
1	13053	28	Indian	Kumar	Heart Disease
2	13067	29	American	Bob	Heart Disease
3	13053	35	Canadian	Ivan	Viral Infection
4	13067	36	Japanese	Umeko	Cancer

- <zip, age, Nationality>
- <Age, Nationality, Name>
- <Name, Condition>
- <Name. Nationality>
- <Nationality>

The use of secure multi-party computing in applications involving machine learning represents the following approach to privacy preservation viz. _____________.

- both of these
- augmenting conventional ML with different strategies} that protect data privacy.
- designing ML specific approaches for privacy preservation.
- none of these

In anonymization technique for privacy preservation, ____________ data concerns with what data should be analyzed but must not be associated with individuals? For example, salaries, health status, property...

- identifier
- none of these
- quasi-identifier
- sensitive

If an insurer receives an average MRI check for Rs 2500 / from patients and suddenly gets a Rs 25000/- check for the same procedure. This is an example of __________ [This question carries only one mark]

- a pattern recognition problem
- anomaly detection
- none of these two

A computer program is said to learn from experience **E** with respect to some task **T** and some performance measure **P** if its performance on **T**, as measured by **P**, improves with experience **E.** Suppose we feed a learning algorithm a lot of historical weather. data, and have it learn to predict weather. Then, a reasonable choice for P would be ___________.

- The probability of it correctly predicting a future date's weather.
- The process of the algorithm examining a large amount of historical weather data.
- The weather prediction task.
- None of these.

_________________ refers to the critical process of performing **initial investigations** on data so as to discover patterns,to spot anomalies,to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

- Exploratory Data Analysis
- Feature Engineering
- Data gathering
- Model Training

Page 2 of 2

Back Submit Clear form