

Task1: Hacker Rank-Done

Task2: SQL Injection

SQL injection is a type of cybersecurity vulnerability that occurs when an attacker can manipulate or inject malicious SQL code into a query. This can happen in situations where user input is not properly validated or sanitized before being incorporated into SQL statements. SQL injection can lead to unauthorized access, data manipulation, and even data exfiltration from a database. A successful SQL injection attack can result in unauthorized access to sensitive data, such as:

- Passwords.
- Credit card details.
- Personal user information.

SQL injection attacks have been used in many high-profile data breaches over the years. These have caused reputational damage and regulatory fines. In some cases, an attacker can obtain a persistent backdoor into an organization's systems, leading to a long-term compromise that can go unnoticed for an extended period.

Task3: Semi-supervised machine learning

Semi-supervised machine learning is an approach that combines both labeled and unlabeled data for training a model. In traditional supervised learning, a model is trained on a dataset where each example is paired with a corresponding label. On the other hand, unsupervised learning deals with unlabeled data, where the algorithm identifies patterns and relationships without explicit guidance.

Semi-supervised learning falls between these two paradigms. It leverages a smaller set of labeled data along with a larger set of unlabeled data to build a model. This is particularly useful in scenarios where obtaining labeled data is expensive or time-consuming.

Here are some key points about semi-supervised learning:

1. Labelled Data:

- A small portion of the dataset contains labeled examples with corresponding ground truth labels.
- The labeled data is used to guide the learning process and fine-tune the model.

2. Unlabeled Data:

- The majority of the dataset consists of unlabeled examples.
- The model uses these unlabeled examples to learn patterns, relationships, and structures within the data.

3. Advantages:

- **Cost-Efficiency:** Labeling data is often expensive and time-consuming. Semi-supervised learning allows leveraging the benefits of labeled data without the need for an extensive labeling effort.
- **Scalability:** It's easier to scale up a semi-supervised model by adding more unlabeled data, as opposed to obtaining more labeled data.

4. Methods in Semi-Supervised Learning:

- **Self-training:** The model is initially trained on the labeled data. It then makes predictions on unlabelled data, and the high-confidence predictions are added to the training set as if they were labelled. The process iterates.
- **Co-training:** Two or more models are trained on different subsets of features or views of the data. They share information during training, enhancing the learning process.
- **Multi-view learning:** Data is represented in multiple ways, and models are trained on each representation. Combining the outputs from different models can improve performance.

5. Applications:

- **Natural Language Processing (NLP):** Semi-supervised learning is often used for tasks like sentiment analysis or text classification where labelled data might be limited.
- **Computer Vision:** Image recognition and object detection can benefit from semi-supervised learning, especially when labelled images are scarce.
- **Speech Recognition:** Training a speech recognition system with a combination of labeled and unlabelled audio data.

6. Challenges:

- **Quality of Unlabeled Data:** The effectiveness of semi-supervised learning depends on the quality and representativeness of the unlabeled data.
- **Model Selection:** Choosing the right semi-supervised learning algorithm and configuration can be challenging and may require experimentation.