CLASS 19 16-06-2021

**QUESTIONS**

👉 What is K-fold cross validation?  
👉 How K-fold cross validation is implemented?  
👉 What is K Nearest Neighbor machine learning?  
👉 Why KNN is called lazy?  
👉 What is web scraping?  
👉 Why python is good for web scraping?  
👉 Why is Web scraping used?  
👉 How do you scrape data from a website?

**ANSWERS**

1. **K-fold cross validation:** This technique involves randomly dividing the dataset into k groups or folds of approximately equal size. The first fold is kept for testing and the model is trained on k-1 folds. The process is repeated K times and each time different fold or a different group of data points are used for validation.

2. **K-fold cross validation implementation:** When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation. If k=5 the dataset will be divided into 5 equal parts and the below process will run 5 times, each time with a different holdout set.

3. **K-Nearest Neighbour** is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

4. **KNN is called lazy because** it does no training at all when you supply the training data. At training time, all it is doing is storing the complete data set but it does not do any calculations at this point.

5. **Web scraping** is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web.

6. **Python** is mostly known as **the best web scraper language**. It's more like an all-rounder and can handle most of the web crawling related processes smoothly. Beautiful Soup is one of the most widely used frameworks based on Python that makes scraping using this language such an easy route to take.

7. **Web scraping** can **help** you extract any kind of data that you want. You would then be able to retrieve, analyze and use the data the way you want. So web scraping simplifies the process of extracting data, speeds it up by automating it and creates easy access to the scrapped data by providing it in a CSV format.

8. **Steps to Scrape Data From A Website:**

1. Find the URL that you want to scrape.
2. Inspecting the Page.
3. Find the data you want to extract.
4. Write the code.
5. Run the code and extract the data.
6. Store the data in the required format.