## Artificial Intelligence and

## Machine Learning

Project Report

Semester-IV (Batch-2022)

Case Study: String Dataset

<https://drive.google.com/file/d/19DZ1yvrGE1Gdl8U7jTeAkRaX9pT-xq9A/view?usp=drive_link>

A red and white sign

Description automatically generated with low confidence

**Supervised By: Submitted By:**

Rajeev Bhardwaj Taksh Joshi

2210990884

G-13

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**Library Description: -**

1. In this file, we used the **pandas** library. Pandas is a powerful and popular open-source Python library for data analysis and manipulation. It provides data structures such as Series (one-dimensional) and Data Frame (two-dimensional) to handle various types of data. Pandas simplify data wrangling, reading/writing data from various file formats, filtering, sorting, merging, handling missing data, reshaping, aggregating, and time series analysis.
2. NumPy, which stands for Numerical Python, is a fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these elements efficiently. NumPy is a cornerstone library for data science, machine learning, and scientific computing in Python.

**Method Description**

1. **str.lower()** method in Pandas is used to convert all characters in the strings of a Series to lowercase.
2. **apply()** method in Pandas is used to apply a function along the axis of a Pandas Series.
3. **map()** method in Pandas is used to substitute each value in a Series with another value, based on a provided dictionary or a function.
4. **str.upper()** method in Pandas is used to convert all characters in the strings of a Series to uppercase.
5. **str.len()** method in Pandas is for obtaining the length of strings.
6. **str.strip()** method in Pandas removes leading and trailing whitespaces from each string in a Series.
7. **str.split()** method in Pandas is used to split each string in a Series into a list of substrings based on a specified delimiter.
8. **str.contains()** method in Pandas is used to test if each element in a Series contains a specific substring. It returns a Boolean Series indicating whether the specified substring is found in each element.
9. **str.replace()** method in Pandas replaces a specified substring or pattern with another substring.
10. **str.startswith()** method in Pandas is used to check whether each element in a Series starts with a specified prefix. It returns a Boolean Series indicating whether the elements start with the specified prefix.
11. **str.endswith()** method in Pandas is used to check whether each element in a Series ends with a specified suffix. It returns a Boolean Series indicating whether the elements end with the specified suffix.
12. **str.cat()** method in Pandas is used to concatenate strings in a Series along a particular axis. It allows you to concatenate elements of one or more Series into a new Series.
13. **str.get()** method in Pandas is used to get the element at a specified position in each string of a Series. It allows you to extract a specific character from each string.
14. **str.slice()** method in Pandas is used to slice substrings from each element in a Series. It allows you to extract a portion of the string based on start and end positions.
15. **str.find()** method in Pandas is used to find the first occurrence of a substring in each element of a Series. It returns the position (index) of the first occurrence of the substring, or -1 if the substring is not found.