

Data_Preprocessing_NLP_Removing Extraspace [white Space]

By Venkatesh_Mungi venkateshmungi1247@gmail.com

1.Remove Leading and Trailing Spaces Using the strip() Method

The Python String strip() method removes leading and trailing characters from a string. The default character to remove is space.

The following string has different types of whitespace and newline characters, such as space (), tab (\t), newline (\n), and carriage return (\r).

```
In [1]: ven = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
ven.strip()
```

```
Out[1]: 'Hi! How are you \t\n\r, My name is Venkatesh  Mungi'
```

```
In [2]: def strip(ven):
        v = ven.strip()
        return v
ven = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
strip(ven)
```

```
Out[2]: 'Hi! How are you \t\n\r, My name is Venkatesh  Mungi'
```

2.Remove All Spaces Using the replace() Method

You can use the replace() method to remove all the whitespace characters from the string, including from between words.

```
In [3]: ve = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
ve.replace(" ", "")
```

```
Out[3]: 'Hi!Howareyou\t\n\r,MynameisVenkateshMungi'
```

```
In [4]: def rmv_space(ve):
        m = ve.replace(" ", "")
        return m
ve = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
rmv_space(ve)
```

```
Out[4]: 'Hi!Howareyou\t\n\r,MynameisVenkateshMungi'
```

3.Remove Duplicate Spaces and Newline Characters Using the join() and split() Methods

You can remove all of the duplicate whitespace and newline characters by using the join() method with the split() method. In this example, the split() method breaks up the string into a list, using the default separator of any whitespace character. Then, the join() method joins the list back into one string with a single space (" ") between each word.

```
In [5]: strn = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
        " ".join(strn.split())
```

```
Out[5]: 'Hi! How are you , My name is Venkatesh Mungi'
```

```
In [6]: def rm_sp(strn1):
        sp = " ".join(strn1.split())
        return sp

strn1 = "  Hi! How are you \t\n\r, My name is Venkatesh  Mungi  "
rm_sp(strn1)
```

```
Out[6]: 'Hi! How are you , My name is Venkatesh Mungi'
```

```
In [ ]:
```