**Head+Tail method**

Title and highlight:

Before training:

Graphical user interface, text

Description automatically generated

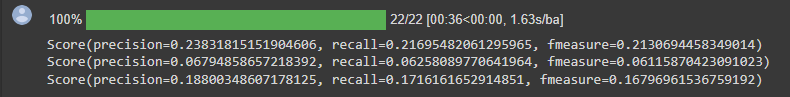
After training: Graphical user interface, text

Description automatically generated

Conclusion: bad

Only Highlight

Before Training:



**First 512 tokens method**

Title + highlight

Before training:

Text

Description automatically generated