Sagar Tripathi Individual Project
NLP Fall 2022
December 11, 2022

Introduction:

The purpose of the project was to identify the relationship between summarized/text ratio against the model and evaluating the model performance. To achieve this evaluation, we used a variety of rogue metrics and created a text ratio metric that measures the comparison between summary and text length.

The division of the work was:

- EDA: Ricardo Diaz
- Models: Alexis Kaldany
- Results: Sagar Tripathi

Description of your individual work:

In the group, my responsibility is to assess the metrics findings that my teammate has presented. to ascertain whether there is any relationship between the feature summary-to-text ratio and the metric score of the models employed for summarization. For the purpose of evaluating the association between the independent variable Summary to text ratio and the rouge1, rouge2, and rougeL metrics (scores from 3 separate models), I utilized a linear regression model.

Results:

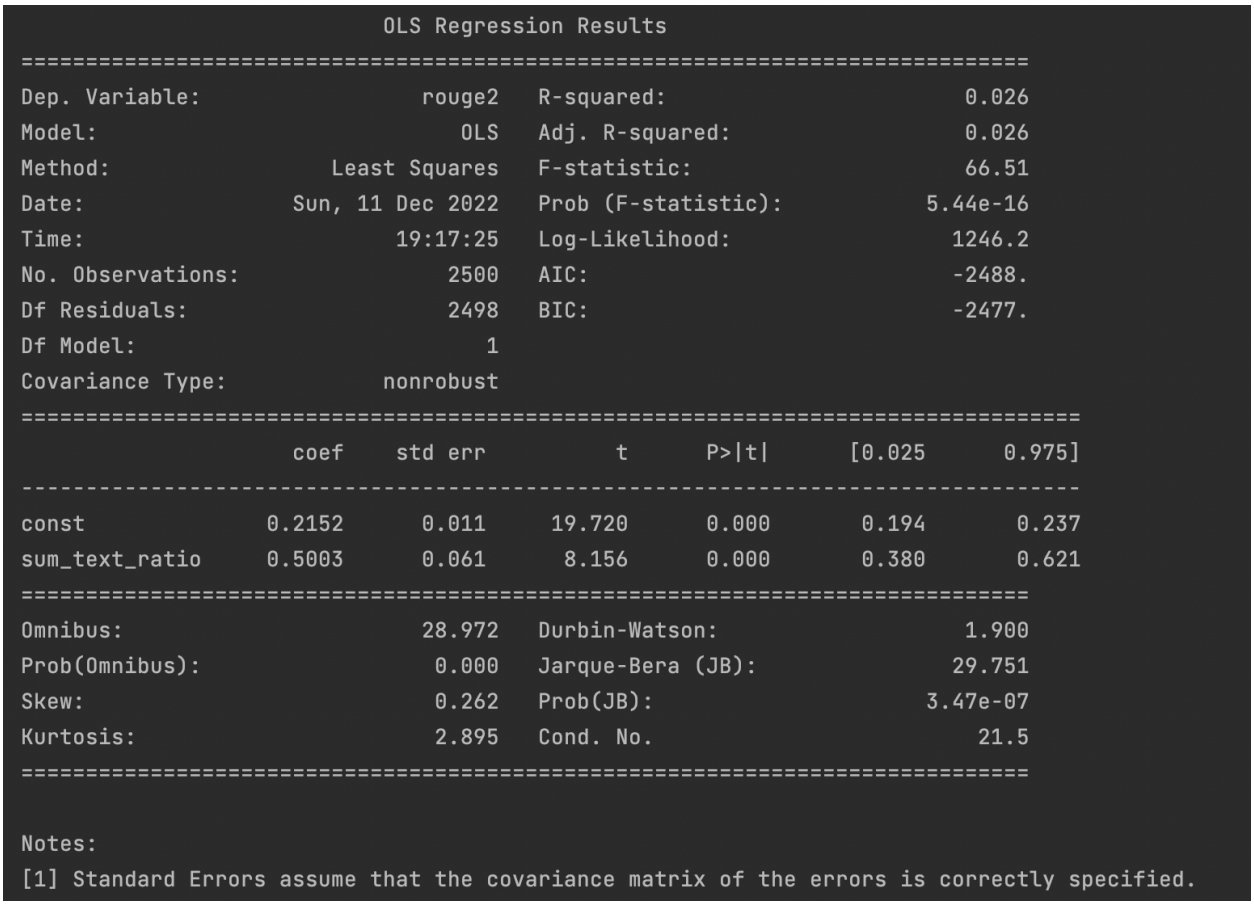
Regression models use independent variables to forecast the results of the dependent variables. Regression analysis takes significance into account to address the most challenging issues. How can we properly analyze statistical evidence for relationships between the observed variables while adjusting for the existence of additional factors is a key challenge in regression analysis? If you are not an expert statistician, regression analysis can be abused. It is a powerful instrument for explaining complicated phenomena and a very persuasive technique to show links between them. Additionally, there are numerous traps that can befall the execution and interpretation of linear regression analysis. In this project, the purpose of statistical evaluation of the Dialog Sum dataset is often to describe relationships between two variables or among several variables. For example, we would like to know whether the Summary text ratio has any influence on any of the scores interpreted by the models. The variables to be explained (Rouge1, Rouge2, and RougeL) are called the dependent variables, or,

alternatively, the response variable; the variable that explains it (Summary text ratio) is called independent variables or predictor variables.

To ascertain the connection between the scores obtained from the models and the summary to text ratio, we used linear regression. First, we did linear regression on the relevant variables, using Rouge1 as the target (dependent variable) and summary to text ratio as the feature (independent variable). Secondly, we did linear regression using Rouge2 as the target (dependent variable) and summary to text ratio as the feature (independent variable) and finally, we did linear regression using RougeL as the target (dependent variable) and summary to text ratio as the feature (independent variable). The outputs of the regression analysis are mentioned below.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 rouge1   R-squared:                       0.018
Model:                            OLS   Adj. R-squared:                  0.018
Method:                 Least Squares   F-statistic:                     45.84
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           1.60e-11
Time:                        19:17:25   Log-Likelihood:                1718.3
No. Observations:                2500   AIC:                            -3433.
Df Residuals:                    2498   BIC:                            -3421.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.5253      0.009     58.133      0.000       0.508       0.543
sum_text_ratio 0.3438      0.051      6.770      0.000       0.244       0.443
==============================================================================
Omnibus:                       17.859   Durbin-Watson:                   1.807
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.056
Skew:                          -0.199   Prob(JB):                     0.000120
Kurtosis:                       2.881   Cond. No.                         21.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Img 1a. Rouge1 and Summary to text ratio



```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 rouge2   R-squared:                       0.026
Model:                            OLS   Adj. R-squared:                  0.026
Method:                 Least Squares   F-statistic:                     66.51
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           5.44e-16
Time:                        19:17:25   Log-Likelihood:                 1246.2
No. Observations:                2500   AIC:                            -2488.
Df Residuals:                    2498   BIC:                            -2477.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              0.2152      0.011     19.720      0.000       0.194       0.237
sum_text_ratio     0.5003      0.061      8.156      0.000       0.380       0.621
==============================================================================
Omnibus:                       28.972   Durbin-Watson:                   1.900
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               29.751
Skew:                           0.262   Prob(JB):                     3.47e-07
Kurtosis:                       2.895   Cond. No.                         21.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

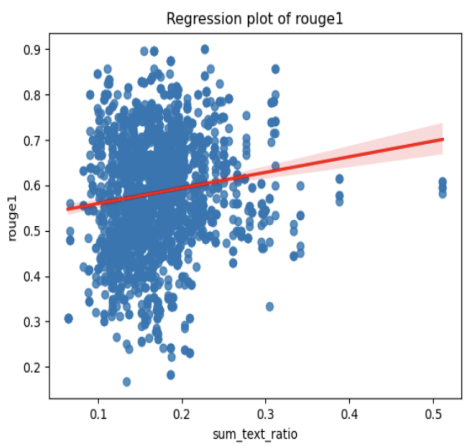
Img 1b. Rouge2 and Summary to text ratio

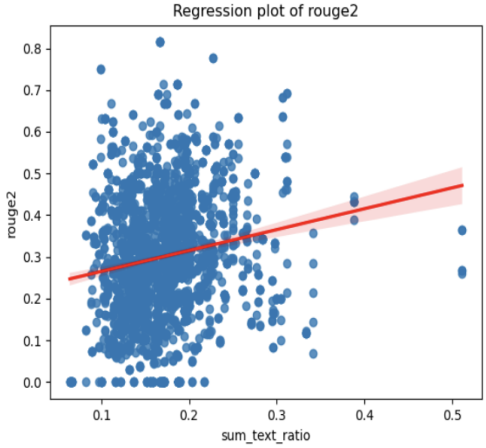
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 rougeL   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     41.84
Date:                Sun, 11 Dec 2022   Prob (F-statistic):           1.19e-10
Time:                        19:17:25   Log-Likelihood:                 1625.1
No. Observations:                2500   AIC:                            -3246.
Df Residuals:                    2498   BIC:                            -3235.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.4979      0.009     53.086      0.000       0.480       0.516
sum_text_ratio 0.3410      0.053      6.468      0.000       0.238       0.444
==============================================================================
Omnibus:                        5.747   Durbin-Watson:                   1.855
Prob(Omnibus):                  0.057   Jarque-Bera (JB):                5.639
Skew:                          -0.094   Prob(JB):                       0.0596
Kurtosis:                       2.863   Cond. No.                         21.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Img 1c. RougeL and Summary to text ratio

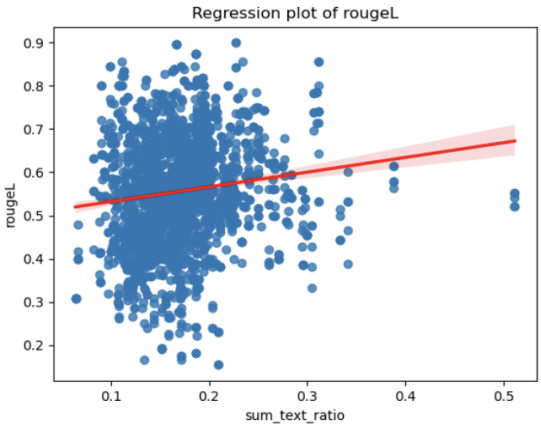
As per image Img1a, Img1b, Img1c R2 (Rouge1: 0.018, Rouge2: 0.026, and RougeL: 0.016) and Adjusted R2 (Rouge1: 0.018, Rouge2: 0.026, and RougeL: 0.016) of the all target variables are near to zero but not zero. Hence, we can conclude that the response variables can be explained by the predictor variable, so there is some relationship between target variables and feature variables.

The Regression plots of the above mentioned three linear regression are show in below mentioned graphs 3a,3b,3c:



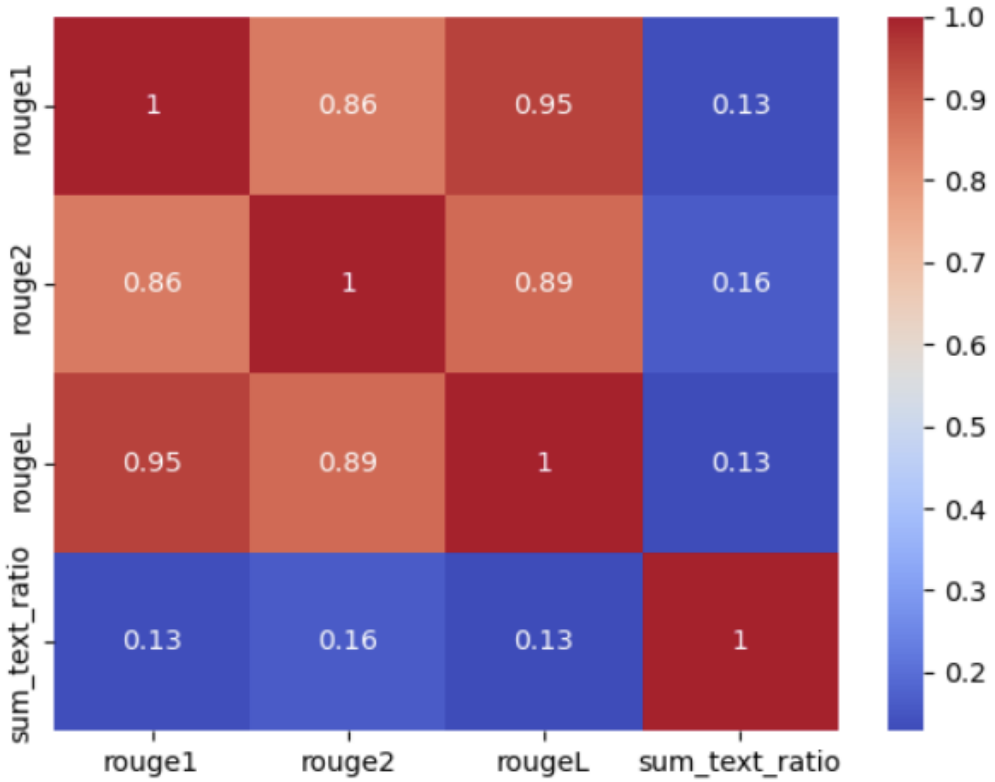


Img 3a. Regression plot of summary text ratio vs Rouge1 Img 3b. Regression plot of summary text ratio vs Rouge2



Img 3c. Regression plot of summary text ratio vs RougeL

We used the heatmap correlation between the target and feature variables to confirm the relationship hypothesis.



Img 3d correlation plot between target and feature

Conclusion:

We reject the null hypothesis and accept the alternative hypothesis as $p<0.05$. So, we can say that there is a relationship between Targets (Rouge1, Rouge2 and RougeL) and feature (Summary to text ratio). As R2 (Rouge1: 0.018, Rouge2: 0.026, and RougeL: 0.016) and Adjusted R2 (Rouge1: 0.018, Rouge2: 0.026, and RougeL: 0.016) of the all target variables are near to zero but not zero. Hence, we can conclude that the response variables can be explained by the predictor variable, so there is some relationship between target variables and feature variables. Seeing that the models trained on the DialogSum dataset for fine tuning it to text summarization, we built an easy text summarization Machine Learning model from facebook/blenderbot_small-90M",” bert-base-uncased”, “t5” and “t5-small” to compare predicted summaries to correct summaries to generate metrics which will be used for hypothesis testing. The examples above illustrate that it works well, which is impressive! In future we can apply different NLP models for better predictions of summary via text and can generate better score

Code:

172 , 100% in PEA.py file

References: stat.yale.edu/Courses/1997-98/101/linreg.htm