# გურამ ჭინჭარაული

## Task 1

Pearson correlation coefficient

-0**.367**

The correlation is weak and negative, indicating that the variables have little influence on each other. Moreover, the negative direction means that an increase in one variable is associated with a decrease in the other.  
  
Task two

**Report**

**Data Loading and Processing:**  
The dataset is loaded from a CSV formatted string using Pandas. Features extracted for model training are: words (number of words), links (number of links), capital\_words (count of capital words), and spam\_word\_count (count of spam-related words). The target variable is is\_spam. 70% of data is split randomly for training, 30% for testing.

**Logistic Regression Model:**  
A logistic regression model from scikit-learn with default parameters (max iterations set to 1000 for convergence) is used. It fits the training data features to predict the probability of an email being spam.

**Coefficients by the Model:**

* words: 0.0082
* links: 0.1811
* capital\_words: 0.1345
* spam\_word\_count: 0.4223
* Intercept: -4.3967

(Coefficient values can slightly vary depending on train/test split random seed but roughly show spam\_word\_count and links have strong positive weight.)

**Testing and Evaluation:**  
The model is tested on the unseen 30% test data, producing a confusion matrix and accuracy:

* Confusion Matrix (Example output):

[[255 20]

[ 15 247]]

* **Accuracy (Example output): 0.9383**

**Code uses sklearn functions confusion\_matrix and accuracy\_score to compute these metrics.**

**Email Classification Functionality:  
You can classify new email by inputting the four features. The example spam email manually constructed has: 900 words, 8 links, 25 capital words, and 9 spam words, which the model confidently classifies as spam with high probability.**

**Spam Email Text Creation Approach:  
The example email was designed with high values for spam-indicative features like numerous links, capital words, and spam-related keywords, aligned with model coefficients indicating these are strong spam indicators.**