

# **Classifier: Logistic Regression**

Master Seminar: Classification and Clustering Lecturer: Stefan Langer Group E: Anna Minaeva Ivo Júnior Lin Shui Shijia Zhou

05.12.2022



#### Content

#### Machine Learning System for Classification

- A feature representation of the input.
- A classification function that computes the estimated class.
- A cost function for learning and an algorithm for optimizing the cost function.
- An evaluation and analysis for the model.

#### Recap: Linear Regression

$$y(\mathbf{x},\mathbf{w}) = w_0 + w_1 x_1 + \ldots + w_D x_D$$



- linear function of the parameters w0, ..., wD.
- regression: predict the numerical value for "new" objects

Recap: Linear Regression → Perceptron

- threshold function  $f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > \mathbf{0}, \\ 0 & \text{otherwise} \end{cases}$
- X: a real-valued vector to represent the input
- W: a vector of real-valued weights
- b: bias

#### New: Perceptron → Logistic Regression

sigmoid function: y= 1/(1+e^(-z)) creates a probability



$$P(y=1) = \sigma(w \cdot x + b)$$
$$= \frac{1}{1 + e^{-(w \cdot x + b)}}$$
$$P(y=0) = 1 - \sigma(w \cdot x + b)$$

$$= 1 - \frac{1}{1 + e^{-(w \cdot x + b)}}$$
$$= \frac{e^{-(w \cdot x + b)}}{1 + e^{-(w \cdot x + b)}}$$

#### *New: Perceptron*→*Logistic Regression*

• advantage: less sensitive to outliers



#### Generalization: Multinomial Logistic Regression (Softmax Regression)

- separate weight vectors for each of the K classes
- softmax function: create a probability distribution

softmax
$$(z_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \ 1 \le i \le k$$
  
softmax $(z) = \left[\frac{e^{z_1}}{\sum_{i=1}^k e^{z_i}}, \frac{e^{z_2}}{\sum_{i=1}^k e^{z_i}}, ..., \frac{e^{z_k}}{\sum_{i=1}^k e^{z_i}}\right]$ 

$$p(y = cx) = \frac{e^{w_c} \cdot x + b_c}{\sum_{j=1}^k e^{w_j} \cdot x + b_j}$$

#### **Generalization: Multinomial Logistic Regression**

• Property of softmax regression

$$\begin{aligned} \textit{overparameterized!} \\ P(y^{(i)} = k | x^{(i)}; \theta) &= \frac{\exp((\theta^{(k)} - \psi)^{\top} x^{(i)})}{\sum_{j=1}^{K} \exp((\theta^{(j)} - \psi)^{\top} x^{(i)})} \\ &= \frac{\exp(\theta^{(k)^{\top}} x^{(i)}) \exp(-\psi^{\top} x^{(i)})}{\sum_{j=1}^{K} \exp(\theta^{(j)^{\top}} x^{(i)}) \exp(-\psi^{\top} x^{(i)})} \\ &= \frac{\exp(\theta^{(k)^{\top}} x^{(i)})}{\sum_{j=1}^{K} \exp(\theta^{(j)^{\top}} x^{(i)})}. \end{aligned}$$

## **Cost Function**

#### Mean Squared Error (MSE) for Linear Regression

- loss function captures the difference between the actual and predicted values for a single record
- cost function aggregates the difference for the entire training dataset
  - error representation, it shows how the model
     predicts compared to the original dataset

 $ext{MSE} = rac{1}{n}\sum_{i=1}^n (Y_i - \hat{Y_i})^2$ 

 $\mathbf{MSE}$  = mean squared error

- n = number of data points
- $Y_i$  = observed values
- $\hat{Y}_i$  = predicted values



### **Logistic Regression Cost Function**

Maximum Likelihood Estimation

$$\begin{split} L(\mathbf{w}) &= P(\mathbf{y} \mid \mathbf{x}; \mathbf{w}) \\ &= \prod_{i=1}^{n} P(y^{(i)} \mid x^{(i)}; \mathbf{w}) \\ &= \prod_{i=1}^{n} \left( \sigma(z^{(i)}) \right)^{y^{(i)}} \left( 1 - \sigma(z^{(i)}) \right)^{1 - y^{(i)}} \end{split}$$

Log Likelihood Function 📫 maximize

$$l(\mathbf{w}) = \log L(\mathbf{w}) = \sum_{i=1}^{n} \left[ y^{(i)} \log \left( \sigma(z^{(i)}) \right) + (1 - y^{(i)}) \log \left( 1 - \sigma(z^{(i)}) \right) \right]$$

Negative Log Likelihood 
$$\implies$$
 minimize $\mathcal{L}(\mathbf{w}) = -l(\mathbf{w}) = -\sum_{i=1}^{n} \left[ y^{(i)} \log \left( \sigma(z^{(i)}) \right) + (1 - y^{(i)}) \log \left( 1 - \sigma(z^{(i)}) \right) \right]$ 

$$P(y = 1|x) \approx 1$$
 if  $y = 1$   
 $P(y = 0|x) = 1 - P(y = 1|x) \approx 1$  if  $y = 0$ 

#### Bernoulli distribution

$$p(y|x) = \hat{y}^{y} (1-\hat{y})^{1-y}$$

$$log p(y|x) = log [\hat{y}^{y} (1 - \hat{y})^{1}|^{-y}]$$
  
=  $y log \hat{y} + (1 - y) log (1 - \hat{y})$ 

$$\operatorname{Cost}(\sigma(z), y) = \begin{cases} -\log(\sigma(z)) & \text{if } y = 1\\ -\log(1 - \sigma(z)) & \text{if } y = 0 \end{cases}$$

### Loss for Single Training Example



$$P(y = 1|\mathbf{x}) \approx 1 \qquad \text{if } y = 1$$
$$P(y = 0|\mathbf{x}) = 1 - P(y = 1|\mathbf{x}) \approx 1 \qquad \text{if } y = 0$$

$$\operatorname{Cost}(\sigma(z), y) = \begin{cases} -\log(\sigma(z)) & \text{if } y = 1\\ -\log(1 - \sigma(z)) & \text{if } y = 0 \end{cases}$$

$$-\left[y^{(i)}\log\left(\sigmaig(z^{(i)}ig)
ight)+ig(1-y^{(i)}ig)\log\left(1-\sigmaig(z^{(i)}ig)
ight)
ight]$$

#### **Gradient Descent**







### Regularization

$$\hat{w} = \operatorname{argmax}_{w} \sum_{1=1}^{m} \log P(y^{(i)}|x^{(i)}) - \alpha R(w)$$

- is a technique that tries to achieve the generalization of the model. It means that our model works well not only with training or test data, but also with the data it'll receive in the future
- used for tuning the function by adding an additional penalty term in the error function
- *L1(Lasso)* helps to reduce the overfitting in the model as well as feature selection (Manhattan distance)

$$R(W) = ||W||_1 = \sum_{i=1}^{N} |w_i|$$

• *L2(Ridge)* is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients (Euclidean distance)

$$R(W) = ||W||_2^2 = \sum_{j=1}^N w_j^2$$

#### **Loss Function for Softmax Regression**

loss function for a single training sample:

$$L_{CE}(\hat{y}, y) = -\sum_{k=1}^{K} y_k log(\hat{y}_k)$$
  
=  $-log \ \hat{y}_c$   
=  $-log[\frac{e^{(w_c.x+b_c)}}{\sum_{j=1}^{K} e^{(w_j.x+b_j)}}]$ 

loss function for all training samples:

$$L_{CE}(\hat{Y}, Y) = -\sum_{i=1}^{N} \sum_{k=1}^{K} 1[y_i = k] log[\frac{e^{(w_c.x+b_c)}}{\sum_{j=1}^{K} e^{(w_j.x+b_j)}}]$$

$$\mathcal{L} = \sum_{i=1}^n \sum_{j=1}^h -y_j^{[i]} \log\left(a_j^{[i]}
ight)$$



- Feature Extraction:
  - CountVectorizer
  - TfidfVectorizer
- Tokenization: character 4-grams
- No padding
- min\_df, max\_df: Default

# **Classification Model**

- Division between:
  - Smaller datasets:
    - Author recognition
    - Language recognition
  - Larger datasets
    - Category recognition
    - Sentiment analysis

## **Classification Model**

- Common hyperparameters:
  - C (Inverse of regularization strength): 1e5
  - multi\_class: "ovr" faster, but "multinomial" had better results

## **Classification Model**

- For smaller datasets:
  - solver: "lbfgs" (Limited-memory BFGS)
  - tol: 1e-3
- For larger datasets:
  - solver: "saga" (Stochastic Average Gradient)
  - tol: 1e-2

- 0.8

0.6

0.4

- 0.2

0.0



| Author Classific                | cation (  | Count  | Vector   | izer)   |
|---------------------------------|-----------|--------|----------|---------|
| <pre># Training time: 18s</pre> |           |        |          |         |
| #Classification report:         |           |        |          |         |
|                                 | precision | recall | f1-score | support |
| Franz Kafka                     | 0.94      | 0.94   | 0.94     | 280     |
| Friedrich Schiller              | 0.82      | 0.85   | 0.84     | 266     |
| Henrik Ibsen                    | 0.99      | 0.99   | 0.99     | 897     |
| James Joyce                     | 0.95      | 0.96   | 0.95     | 682     |
| Johann Wolfgang von Goethe      | 0.79      | 0.76   | 0.78     | 228     |
| Virginia Woolf                  | 0.99      | 0.98   | 0.98     | 1901    |
| Wilhelm Busch                   | 0.97      | 0.96   | 0.97     | 627     |
| accuracy                        |           |        | 0.96     | 4881    |
| macro avg                       | 0.92      | 0.92   | 0.92     | 4881    |
| weighted avg                    | 0.96      | 0.96   | 0.96     | 4881    |
|                                 |           |        |          |         |



| Author Classifi                 | cation (  | TfidfV | ectoria  | zer)    |
|---------------------------------|-----------|--------|----------|---------|
| <pre># Training time: 22s</pre> |           |        |          |         |
| #Classification report:         | precision | recall | f1-score | support |
| Franz Kafka                     | 0.96      | 0.97   | 0.96     | 280     |
| Friedrich Schiller              | 0.83      | 0.89   | 0.86     | 266     |
| Henrik Ibsen                    | 1.00      | 1.00   | 1.00     | 897     |
| James Joyce                     | 0.96      | 0.97   | 0.96     | 682     |
| Johann Wolfgang von Goethe      | 0.84      | 0.77   | 0.80     | 228     |
| Virginia Woolf                  | 0.99      | 0.98   | 0.99     | 1901    |
| Wilhelm Busch                   | 0.98      | 0.97   | 0.98     | 627     |
| accuracy                        |           |        | 0.97     | 4881    |
| macro avg                       | 0.93      | 0.94   | 0.93     | 4881    |
| weighted avg                    | 0.97      | 0.97   | 0.97     | 4881    |
|                                 |           |        |          |         |



#### Language Classification (CountVectorizer)

| <pre># Training ti #Classificati</pre> | me: 8s<br>on report: |        |          |         |
|--|----------------------|--------|----------|---------|
|  | precision            | recall | f1-score | support |
| da                                     | 1.00                 | 1.00   | 1.00     | 838     |
| de                                     | 1.00                 | 1.00   | 1.00     | 1443    |
| en                                     | 1.00                 | 1.00   | 1.00     | 2517    |
| fr                                     | 1.00                 | 0.97   | 0.99     | 36      |
| it                                     | 0.93                 | 0.90   | 0.92     | 31      |
| unknown                                | 0.82                 | 0.56   | 0.67     | 16      |
|  |                      |        |          |         |
| accuracy                               |                      |        | 1.00     | 4881    |
| macro avg                              | 0.96                 | 0.91   | 0.93     | 4881    |
| weighted avg                           | 1.00                 | 1.00   | 1.00     | 4881    |
|  |                      |        |          |         |

- 1.0

- 0.8

- 0.6

- 0.4

- 0.2

0.0



|   | Language                                   | Classifi         | cation | (Tfidf)  | /ectoriz | er) |
|---|--|------------------|--------|----------|----------|-----|
|   | <pre># Training time #Classification</pre> | : 16s<br>report: | rocall | f1 ccoro | cupport  |     |
|   | þ  | recision         | recall | 11-50016 | Support  |     |
|   | da   | 1.00             | 1.00   | 1.00     | 838      |     |
| I | de   | 1.00             | 1.00   | 1.00     | 1443     |     |
|   | en   | 1.00             | 1.00   | 1.00     | 2517     |     |
|   | fr   | 0.95             | 1.00   | 0.97     | 36       |     |
|   | it   | 0.93             | 0.90   | 0.92     | 31       |     |
|   | unknown                                    | 0.67             | 0.25   | 0.36     | 16       |     |
| I | accuracy                                   |                  |        | 1.00     | 4881     |     |
|   | macro avg                                  | 0.92             | 0.86   | 0.87     | 4881     |     |
|   | weighted avg                               | 1.00             | 1.00   | 1.00     | 4881     |     |



#### **Sentiment Classification (CountVectorizer)**

| <pre># Training time #Classification p</pre> | : 33s<br>report:<br>recision | recall       | f1-score     | support      |
|--|------------------------------|--------------|--------------|--------------|
| negative<br>positive                         | 0.90<br>0.89                 | 0.89<br>0.90 | 0.89<br>0.89 | 4985<br>5015 |
| accuracy                                     |                              |              | 0.89         | 10000        |
| macro avg                                    | 0.89                         | 0.89         | 0.89         | 10000        |
| weighted avg                                 | 0.89                         | 0.89         | 0.89         | 10000        |



#### Sentiment Classification (TfidfVectorizer)

| <pre># Training time: #Classification</pre> | 12s<br>report: |        |          |         |
|---|----------------|--------|----------|---------|
| pr  | ecision        | recall | f1-score | support |
|   |                |        |          |         |
| negative                                    | 0.91           | 0.90   | 0.90     | 4985    |
| positive                                    | 0.90           | 0.91   | 0.91     | 5015    |
|   |                |        |          |         |
| accuracy                                    |                |        | 0.90     | 10000   |
| macro avg                                   | 0.90           | 0.90   | 0.90     | 10000   |
| weighted avg                                | 0.90           | 0.90   | 0.90     | 10000   |
|   |                |        |          |         |



| <pre># Training time:</pre> | : 14s     |        |          |         |
|-----------------------------|-----------|--------|----------|---------|
| #Classification             | report:   |        |          |         |
|                             | precision | recall | f1-score | support |
|                             | 0 47      | 0 25   | 0 10     | 292     |
|                             | 0.47      | 0.35   | 0.40     | 202     |
|                             | 0.44      | 0.40   | 0.45     | 392     |
| COLLECE                     | 0.59      | 0.42   | 0.41     | 212     |
| COMEDY                      | 0.55      | 0.45   | 0.49     | 212     |
| COMEDI                      | 0.55      | 0.47   | 0.50     | 399     |
|                             | 0.50      | 0.02   | 0.59     | 409     |
|                             | 0.45      | 0.40   | 0.40     | 410     |
| EDUCATION                   | 0.70      | 0.05   | 0.70     | 419     |
|                             | 0.49      | 0.40   | 0.44     | 190     |
|                             | 0.42      | 0.40   | 0.44     | 207     |
|                             | 0.47      | 0.40   | 0.47     | 300     |
|                             | 0.37      | 0.30   | 0.37     | 273     |
|                             | 0.42      | 0.37   | 0.39     | 202     |
|                             | 0.20      | 0.20   | 0.27     | 300     |
|                             | 0.73      | 0.74   | 0.75     | 300     |
|                             | 0.52      | 0.52   | 0.52     | 400     |
| MONEY                       | 0.00      | 0.00   | 0.00     | 395     |
|                             | 0.49      | 0.55   | 0.51     | 324     |
|                             | 0.45      | 0.49   | 0.47     | 391     |
|                             | 0.44      | 0.44   | 0.44     | 420     |
|                             | 0.00      | 0.00   | 0.05     | 415     |
| CCTENCE                     | 0.02      | 0.50   | 0.59     | 440     |
|                             | 0.00      | 0.57   | 0.50     | 414     |
|                             | 0.00      | 0.00   | 0.00     | 410     |
|                             | 0.52      | 0.50   | 0.54     | 415     |
|                             | 0.00      | 0.00   | 0.00     | 207     |
|                             | 0.02      | 0.00   | 0.04     | 397     |
|                             | 0.50      | 0.00   | 0.00     | 410     |
|                             | 0.55      | 0.54   | 0.00     | 405     |
| WEDDINGS                    | 0.75      | 0.77   | 0.70     | 400     |
| WEIRD NEWS                  | 0.35      | 0.37   | 0.30     | 408     |
| WELLINESS                   | 0.31      | 0.33   | 0.32     | 407     |
| WOPLD                       | 0.41      | 0.41   | 0.41     | 400     |
| WUKLD                       | 0.50      | 0.00   | 0.58     | 404     |
| accuracy                    |           |        | 0.51     | 12824   |
| macro avg                   | 0.51      | 0.50   | 0.51     | 12824   |
| weighted avg                | 0.51      | 0.51   | 0.51     | 12824   |

|                                  |       | <pre># Training time:</pre> | 11s       |        |          |         |
|----------------------------------|-------|-----------------------------|-----------|--------|----------|---------|
|                                  |       |                             | precision | recall | f1-score | support |
| uatio                            | n     | ARTS & CULTURE              | 0.46      | 0.35   | 0.39     | 282     |
|                                  |       | BLACK VOICES                | 0.45      | 0.44   | 0.45     | 392     |
|                                  |       | BUSINESS                    | 0.37      | 0.40   | 0.38     | 380     |
|                                  |       | COLLEGE                     | 0.52      | 0.46   | 0.49     | 212     |
| 4                                |       | COMEDY                      | 0.49      | 0.51   | 0.50     | 399     |
| a                                |       | CRIME                       | 0.57      | 0.60   | 0.59     | 409     |
|                                  |       | CULTURE & ARTS              | 0.43      | 0.48   | 0.46     | 391     |
| )                                | - 0.7 | DIVORCE                     | 0.78      | 0.68   | 0.72     | 419     |
| 01 0.00 0.01 0.00 0.06 0.01      |       | FDUCATTON                   | 0.53      | 0.38   | 0.44     | 198     |
| 01 0.01 0.01 0.01 0.04 0.00      |       | ENTERTATIMENT               | 0.43      | 0.45   | 0.44     | 387     |
| 02 0.00 0.00 0.02 0.05 0.00      |       | ENVIRONMENT                 | 0.43      | 0.47   | 0.45     | 355     |
| 01 0.01 0.04 0.02 0.02 0.01      | - 0.6 | FIFTY                       | 0.44      | 0.41   | 0.42     | 273     |
| 02 0.00 0.06 0.01 0.01 0.01      |       | GOOD NEWS                   | 0.43      | 0.38   | 0.42     | 305     |
| 03 0.01 0.04 0.01 0.02 0.02      |       |                             | 0 24      | 0.28   | 0.40     | 386     |
| 00 0.06 0.01 0.02 0.03 0.00      |       |                             | 0.24      | 0.20   | 0.20     | 386     |
| 01 0.01 0.02 0.01 0.04 0.01      |       |                             | 0.72      | 0.75   | 0.72     | 400     |
| 01 0.00 0.06 0.02 0.01 0.04      | - 0.5 |                             | 0.55      | 0.54   | 0.55     | 400     |
| 03 0.01 0.00 0.08 0.03 0.00      |       | MEDIA                       | 0.00      | 0.00   | 0.59     | 395     |
| 02 0.01 0.09 0.02 0.01 0.01      |       |                             | 0.47      | 0.50   | 0.49     | 324     |
| 01 0.02 0.01 0.20 0.04 0.01      |       | PARENTING                   | 0.43      | 0.45   | 0.44     | 391     |
| 02 0.01 0.01 0.02 0.00 0.00      |       | POLITICS                    | 0.48      | 0.45   | 0.47     | 420     |
| 00 0.00 0.01 0.01 0.02 0.02      | - 0.4 | QUEER VOICES                | 0.65      | 0.59   | 0.62     | 415     |
| 01 0.02 0.02 0.02 0.01 0.01      |       | RELIGION                    | 0.64      | 0.57   | 0.60     | 440     |
| 02 0.01 0.02 0.03 0.05 0.00      |       | SCIENCE                     | 0.58      | 0.55   | 0.57     | 414     |
| 01 0.00 0.02 0.01 0.04 0.06      |       | SPORTS                      | 0.62      | 0.64   | 0.63     | 410     |
| 01 0.02 0.01 0.01 0.01 0.01 0.01 | - 0.3 | STYLE                       | 0.53      | 0.54   | 0.54     | 413     |
| 01 0.00 0.05 0.05 0.00 0.01      |       | STYLE & BEAUTY              | 0.64      | 0.64   | 0.64     | 391     |
| 01 0.01 0.03 0.01 0.01 0.01      |       | TASTE                       | 0.66      | 0.67   | 0.66     | 397     |
| 01 0.00 0.02 0.01 0.03 0.00      |       | TECH                        | 0.57      | 0.58   | 0.58     | 416     |
| 02 0.01 0.01 0.01 0.01 0.00      |       | TRAVEL                      | 0.58      | 0.53   | 0.55     | 405     |
| 02 0.00 0.03 0.02 0.00 0.01      | - 0.2 | WEDDINGS                    | 0.75      | 0.76   | 0.75     | 400     |
| 53 0.00 0.02 0.02 0.00 0.02      |       | WEIRD NEWS                  | 0.33      | 0.33   | 0.33     | 408     |
| 01 0.76 0.01 0.01 0.01 0.00      |       | WELLNESS                    | 0.29      | 0.31   | 0.30     | 407     |
| 02 0.01 0.33 0.01 0.02 0.01      |       | WOMEN                       | 0.39      | 0.41   | 0.40     | 400     |
| 01 0.00 0.00 0.31 0.01 0.01      |       | WORLD                       | 0.57      | 0.59   | 0.58     | 404     |
| 01 0.01 0.02 0.04 0.41 0.01      | - 0.1 |                             |           |        |          |         |
|                                  |       | accuracy                    |           |        | 0.51     | 12824   |
| EDDIN<br>RD NEV<br>WOMI<br>WORI  |       | macro avg                   | 0.51      | 0.50   | 0.51     | 12824   |
| WEIF W                           |       | weighted avg                | 0.51      | 0.51   | 0.51     | 12824   |
|                                  | 0.0   | weighted avg                | 0151      | 0.51   | 0151     | 12027   |

#### Category Classification based on headline (TfidfVectorizer)

| ARTS & CULTURE | 0.35           | 0.04         | 0.01     | 0.00    | 0.03   | 0.00  | 0.16           | 0.00    | 0.00      | 0.05          | 0.01        | 0.02  | 0.01      | 0.02           | 0.01          | 0.02   | 0.01  | 0.00  | 0.02      | 0.01     | 0.03         | 0.00     | 0.02    | 0.00   | 0.02  | 0.00           | 0.01  | 0.02 | 0.01   | 0.00     | 0.01 0     | 0.00     | 0.06 0 | .01   |
|----------------|----------------|--------------|----------|---------|--------|-------|----------------|---------|-----------|---------------|-------------|-------|-----------|----------------|---------------|--------|-------|-------|-----------|----------|--------------|----------|---------|--------|-------|----------------|-------|------|--------|----------|------------|----------|--------|-------|
| BLACK VOICES   | 0.01           | 0.44         | 0.01     | 0.03    | 0.02   | 0.06  | 0.03           | 0.00    | 0.01      | 0.05          | 0.01        | 0.00  | 0.01      | 0.01           | 0.00          | 0.03   | 0.02  | 0.01  | 0.02      | 0.04     | 0.01         | 0.02     | 0.00    | 0.05   | 0.01  | 0.03           | 0.00  | 0.01 | 0.01   | 0.01     | 0.01 0     | 0.01     | 0.04 0 | .00   |
| BUSINESS       | 0.01           | 0.02         | 0.40     | 0.01    | 0.01   | 0.02  | 0.01           | 0.00    | 0.01      | 0.00          | 0.02        | 0.02  | 0.01      | 0.02           | 0.01          | 0.03   | 0.03  | 0.11  | 0.02      | 0.03     | 0.01         | 0.00     | 0.01    | 0.01   | 0.01  | 0.01           | 0.01  | 0.07 | 0.01   | 0.00     | 0.02 0     | 0.03     | 0.03 0 | .02   |
| COLLEGE        | 0.00           | 0.02         | 0.03     | 0.46    | 0.01   | 0.03  | 0.03           | 0.01    | 0.06      | 0.00          | 0.01        | 0.00  | 0.01      | 0.01           | 0.00          | 0.02   | 0.02  | 0.02  | 0.01      | 0.03     | 0.01         | 0.01     | 0.00    | 0.03   | 0.00  | 0.01           | 0.01  | 0.01 | 0.02   | 0.00     | 0.00 0     | 0.02     | 0.05 0 | .00   |
| COMEDY         | 0.01           | 0.01         | 0.01     | 0.01    |        | 0.02  | 0.01           | 0.01    | 0.00      | 0.06          | 0.01        | 0.01  | 0.00      | 0.02           | 0.02          | 0.02   | 0.05  | 0.00  | 0.02      | 0.02     | 0.01         | 0.02     | 0.01    | 0.02   | 0.02  | 0.01           | 0.01  | 0.01 | 0.01   | 0.01     | 0.04 0     | 0.02     | 0.02 0 | .01   |
| CRIME          | 0.00           | 0.07         | 0.01     | 0.01    | 0.00   | 0.60  | 0.00           | 0.00    | 0.00      | 0.01          | 0.02        | 0.00  | 0.01      | 0.00           | 0.00          | 0.01   | 0.01  | 0.02  | 0.02      | 0.01     | 0.01         | 0.02     | 0.00    | 0.02   | 0.00  | 0.00           | 0.00  | 0.01 | 0.02   | 0.00     | 0.06 0     | 0.01     | 0.01 0 | .01   |
| CULTURE & ARTS | 0.08           | 0.03         | 0.02     | 0.01    | 0.02   | 0.00  | 0.48           | 0.00    | 0.00      | 0.03          | 0.02        | 0.01  | 0.02      | 0.01           | 0.02          | 0.01   | 0.01  | 0.00  | 0.01      | 0.00     | 0.01         | 0.02     | 0.02    | 0.01   | 0.01  | 0.01           | 0.01  | 0.02 | 0.03   | 0.01     | 0.04 0     | 0.01     | 0.02 0 | .02   |
| DIVORCE        | 0.00           | 0.01         | 0.01     | 0.00    | 0.00   | 0.00  | 0.00           | 0.68    | 0.00      | 0.01          | 0.01        | 0.02  | 0.01      | 0.02           | 0.01          | 0.01   | 0.00  | 0.02  | 0.02      | 0.00     | 0.01         | 0.00     | 0.00    | 0.00   | 0.00  | 0.00           | 0.00  | 0.00 | 0.00   | 0.06     | 0.01 0     | 0.02     | 0.03 0 | .00   |
| EDUCATION      | 0.01           | 0.01         | 0.04     | 0.08    | 0.01   | 0.01  | 0.03           | 0.00    | 0.38      | 0.00          | 0.01        | 0.01  | 0.01      | 0.04           | 0.01          | 0.08   | 0.01  | 0.04  | 0.04      | 0.03     | 0.01         | 0.03     | 0.02    | 0.01   | 0.01  | 0.01           | 0.00  | 0.01 | 0.00   | 0.01     | 0.02       | 0.05     | 0.02 0 | .02   |
| ENTERTAINMENT  | 0.02           | 0.06         | 0.01     | 0.00    | 0.06   | 0.00  | 0.02           | 0.01    | 0.00      | 0.45          | 0.01        | 0.01  | 0.02      | 0.00           | 0.00          | 0.01   | 0.03  | 0.01  | 0.02      | 0.03     | 0.03         | 0.01     | 0.01    | 0.02   | 0.05  | 0.01           | 0.00  | 0.02 | 0.01   | 0.01     | 0.02 (     | 0.01     | 0.04 0 | .01   |
| ENVIRONMENT    | 0.01           | 0.01         | 0.04     | 0.01    | 0.01   | 0.02  | 0.02           | 0.00    | 0.01      | 0.00          | 0.47        | 0.01  | 0.06      | 0.01           | 0.01          | 0.03   | 0.01  | 0.02  | 0.01      | 0.03     | 0.01         | 0.01     | 0.04    | 0.01   | 0.01  | 0.01           | 0.01  | 0.00 | 0.01   | 0.00     | 0.06 0     | 0.02     | 0.01 0 | .04   |
| FIFTY          | 0.01           | 0.01         | 0.02     | 0.01    | 0.03   | 0.01  | 0.01           | 0.02    | 0.00      | 0.00          | 0.01        | 0.41  | 0.01      | 0.08           | 0.01          | 0.01   | 0.01  | 0.03  | 0.05      | 0.00     | 0.02         | 0.03     | 0.00    | 0.00   | 0.01  | 0.01           | 0.01  | 0.01 | 0.03   | 0.01     | 0.00       | 0.08     | 0.03 0 | .00   |
| GOOD NEWS      | 0.02           | 0.01         | 0.02     | 0.01    | 0.01   | 0.03  | 0.01           | 0.00    | 0.00      | 0.01          | 0.06        | 0.01  | 0.38      | 0.03           | 0.01          | 0.05   | 0.01  | 0.01  | 0.05      | 0.01     | 0.02         | 0.01     | 0.01    | 0.03   | 0.01  | 0.01           | 0.02  | 0.02 | 0.02   | 0.01     | 0.09 0     | 0.02     | 0.01 0 | .01   |
| HEALTHY LIVING | 0.01           | 0.02         | 0.05     | 0.01    | 0.01   | 0.01  | 0.02           | 0.01    | 0.01      | 0.00          | 0.02        | 0.04  | 0.00      | 0.28           | 0.01          | 0.04   | 0.01  | 0.01  | 0.03      | 0.01     | 0.02         | 0.01     | 0.04    | 0.01   | 0.02  | 0.01           | 0.03  | 0.02 | 0.01   | 0.02     | 0.01       | 0.20     | 0.04 0 | .01   |
| HOME & LIVING  | 0.00           | 0.01         | 0.01     | 0.00    | 0.02   | 0.01  | 0.03           | 0.00    | 0.00      | 0.00          | 0.01        | 0.01  | 0.00      | 0.01           | 0.73          | 0.01   | 0.01  | 0.01  | 0.01      | 0.00     | 0.00         | 0.00     | 0.01    | 0.01   | 0.01  | 0.03           | 0.02  | 0.01 | 0.02   | 0.01     | 0.01 0     | 0.02     | 0.00 0 | .00   |
| IMPACT         | 0.01           | 0.02         | 0.04     | 0.01    | 0.01   | 0.01  | 0.01           | 0.01    | 0.01      | 0.01          | 0.05        | 0.01  | 0.05      | 0.06           | 0.01          | 0.34   | 0.01  | 0.02  | 0.06      | 0.03     | 0.02         | 0.01     | 0.01    | 0.01   | 0.02  | 0.00           | 0.01  | 0.02 | 0.01   | 0.00     | 0.00 0     | 0.06     | 0.04 0 | .04   |
| AIDEM P        | 0.01           | 0.02         | 0.02     | 0.00    | 0.04   | 0.01  | 0.01           | 0.00    | 0.01      | 0.02          | 0.01        | 0.01  | 0.01      | 0.00           | 0.00          | 0.02   | 0.60  | 0.01  | 0.01      | 0.06     | 0.01         | 0.01     | 0.01    | 0.03   | 0.01  | 0.01           | 0.00  | 0.02 | 0.00   | 0.00     | 0.01 0     | 0.01     | 0.02 0 | .02   |
| MONEY .        | 0.00           | 0.01         | 0.10     | 0.02    | 0.00   | 0.02  | 0.01           | 0.01    | 0.02      | 0.01          | 0.00        | 0.03  | 0.01      | 0.04           | 0.02          | 0.02   | 0.00  | 0.50  | 0.01      | 0.01     | 0.01         | 0.00     | 0.01    | 0.01   | 0.01  | 0.01           | 0.01  | 0.04 | 0.01   | 0.02     | 0.02 (     | 0.02     | 0.01 0 | .01   |
| PARENTING      | 0.01           | 0.01         | 0.02     | 0.02    | 0.02   | 0.01  | 0.01           | 0.02    | 0.03      | 0.02          | 0.02        | 0.02  | 0.02      | 0.05           | 0.01          | 0.04   | 0.01  | 0.01  | 0.45      | 0.02     | 0.02         | 0.00     | 0.01    | 0.00   | 0.02  | 0.01           | 0.02  | 0.02 | 0.02   | 0.01     | 0.02 0     | 0.03     | 0.05 0 | .00   |
| POLITICS       | 0.00           | 0.02         | 0.05     | 0.00    | 0.03   | 0.03  | 0.00           | 0.00    | 0.01      | 0.02          | 0.02        | 0.00  | 0.01      | 0.02           | 0.00          | 0.02   | 0.05  | 0.02  | 0.01      | 0.45     | 0.02         | 0.02     | 0.00    | 0.02   | 0.00  | 0.00           | 0.00  | 0.01 | 0.01   | 0.00     | 0.02 (     | 0.01     | 0.04 0 | .06   |
| QUEER VOICES   | 0.01           | 0.02         | 0.01     | 0.00    | 0.02   | 0.01  | 0.02           | 0.01    | 0.01      | 0.04          | 0.00        | 0.00  | 0.01      | 0.01           | 0.01          | 0.01   | 0.02  | 0.00  | 0.02      | 0.01     | 0.59         | 0.03     | 0.00    | 0.01   | 0.01  | 0.00           | 0.01  | 0.01 | 0.01   | 0.02     | 0.01 0     | 0.01     | 0.01 0 | .01   |
| RELIGION       | 0.00           | 0.02         | 0.01     | 0.00    | 0.01   | 0.02  | 0.03           | 0.01    | 0.00      | 0.01          | 0.00        | 0.00  | 0.00      | 0.03           | 0.00          | 0.05   | 0.01  | 0.01  | 0.01      | 0.02     | 0.02         | 0.57     | 0.01    | 0.01   | 0.01  | 0.00           | 0.01  | 0.00 | 0.01   | 0.01     | 0.01 (     | 0.02     | 0.02 0 | .05   |
| SCIENCE        | 0.01           | 0.01         | 0.01     | 0.00    | 0.01   | 0.01  | 0.02           | 0.00    | 0.00      | 0.00          | 0.06        | 0.01  | 0.00      | 0.06           | 0.01          | 0.02   | 0.00  | 0.00  | 0.02      | 0.01     | 0.01         | 0.01     | 0.55    | 0.00   | 0.01  | 0.01           | 0.01  | 0.01 | 0.01   | 0.00     | 0.05 0     | 0.05     | 0.00 0 | .01   |
| SPORTS         | 0.00           | 0.02         | 0.01     | 0.01    | 0.01   | 0.02  | 0.00           | 0.00    | 0.00      | 0.02          | 0.02        | 0.00  | 0.01      | 0.02           | 0.00          | 0.01   | 0.01  | 0.01  | 0.01      | 0.00     | 0.01         | 0.00     | 0.02    | 0.64   | 0.02  | 0.01           | 0.00  | 0.00 | 0.01   | 0.01     | 0.03 0     | 0.01     | 0.01 0 | .01   |
| STYLE          | 0.01           | 0.01         | 0.01     | 0.01    | 0.03   | 0.00  | 0.01           | 0.00    | 0.00      | 0.06          | 0.01        | 0.02  | 0.01      | 0.01           | 0.02          | 0.00   | 0.01  | 0.01  | 0.00      | 0.00     | 0.00         | 0.00     | 0.01    | 0.01   | 0.54  | 0.08           | 0.01  | 0.01 | 0.01   | 0.00     | 0.02 0     | 0.01     | 0.03 0 | .00   |
| STYLE & BEAUTY | 0.01           | 0.01         | 0.01     | 0.00    | 0.02   | 0.00  | 0.02           | 0.00    | 0.00      | 0.01          | 0.00        | 0.01  | 0.00      | 0.01           | 0.02          | 0.01   | 0.01  | 0.01  | 0.01      | 0.00     | 0.01         | 0.00     | 0.00    | 0.00   | 0.12  | 0.64           | 0.01  | 0.01 | 0.02   | 0.01     | 0.01 0     | 0.01     | 0.01 0 | .00   |
| TASTE          | 0.00           | 0.01         | 0.02     | 0.00    | 0.02   | 0.00  | 0.02           | 0.00    | 0.01      | 0.02          | 0.01        | 0.00  | 0.01      | 0.01           | 0.02          | 0.02   | 0.01  | 0.01  | 0.01      | 0.01     | 0.00         | 0.00     | 0.01    | 0.02   | 0.02  | 0.02           | 0.67  | 0.01 | 0.02   | 0.00     | 0.03 (     | 0.02     | 0.00 0 | .01   |
| TECH           | 0.01           | 0.01         | 0.05     | 0.00    | 0.02   | 0.00  | 0.01           | 0.00    | 0.00      | 0.01          | 0.02        | 0.01  | 0.01      | 0.02           | 0.00          | 0.02   | 0.02  | 0.02  | 0.01      | 0.02     | 0.00         | 0.01     | 0.01    | 0.00   | 0.01  | 0.00           | 0.00  | 0.58 | 0.01   | 0.00     | 0.03 (     | 0.00     | 0.01 0 | .02   |
| TRAVEL         | 0.00           | 0.00         | 0.01     | 0.01    | 0.01   | 0.00  | 0.05           | 0.00    | 0.00      | 0.01          | 0.03        | 0.01  | 0.01      | 0.02           | 0.02          | 0.02   | 0.01  | 0.00  | 0.02      | 0.00     | 0.01         | 0.01     | 0.01    | 0.01   | 0.03  | 0.02           | 0.03  | 0.01 | 0.53   | 0.00     | 0.02 (     | 0.02     | 0.00 0 | .02   |
| WEDDINGS       | 0.00           | 0.00         | 0.01     | 0.00    | 0.00   | 0.00  | 0.01           | 0.05    | 0.00      | 0.01          | 0.00        | 0.01  | 0.00      | 0.01           | 0.01          | 0.00   | 0.00  | 0.01  | 0.01      | 0.00     | 0.01         | 0.00     | 0.00    | 0.01   | 0.01  | 0.03           | 0.01  | 0.01 | 0.01   | 0.76     | 0.01 0     | 0.01     | 0.01 0 | .00   |
| WEIRD NEWS     | 0.01           | 0.02         | 0.00     | 0.00    | 0.06   | 0.10  | 0.02           | 0.00    | 0.00      | 0.02          | 0.04        | 0.01  | 0.04      | 0.02           | 0.00          | 0.00   | 0.01  | 0.02  | 0.03      | 0.01     | 0.00         | 0.01     | 0.04    | 0.02   | 0.03  | 0.01           | 0.04  | 0.03 | 0.02   | 0.01     | 0.33 0     | J.01     | 0.02 0 | .01   |
| WELLNESS       | 0.00           | 0.00         | 0.03     | 0.00    | 0.01   | 0.00  | 0.02           | 0.02    | 0.00      | 0.01          | 0.02        | 0.03  | 0.00      | 0.21           | 0.02          | 0.05   | 0.00  | 0.01  | 0.02      | 0.00     | 0.00         | 0.02     | 0.05    | 0.01   | 0.01  | 0.00           | 0.05  | 0.01 | 0.01   | 0.00     | 0.00       | 3.31     | 0.01 0 | .01   |
| WOMEN          | 0.03           | 0.01         | 0.02     | 0.02    | 0.03   | 0.01  | 0.01           | 0.01    | 0.00      | 0.04          | 0.00        | 0.02  | 0.00      | 0.06           | 0.01          | 0.04   | 0.02  | 0.01  | 0.03      | 0.03     | 0.01         | 0.01     | 0.01    | 0.02   | 0.01  | 0.01           | 0.00  | 0.00 | 0.01   | 0.01     | 0.02 0     | 3.04     | 0.41 0 | .01   |
| WORLD          | 0.00           | 0.00         | 0.02     | 0.00    | 0.00   | 0.04  | 0.02           | 0.00    | 0.00      | 0.01          | 0.03        | 0.00  | 0.01      | 0.01           | 0.00          | 0.03   | 0.02  | 0.00  | 0.00      | 0.05     | 0.01         | 0.04     | 0.01    | 0.01   | 0.00  | 0.00           | 0.00  | 0.01 | 0.01   | 0.00     | 0.01 0     | J.01     | 0.01 0 | .59   |
|                | ARTS & CULTURE | BLACK VOICES | BUSINESS | COLLEGE | COMEDY | CRIME | CULTURE & ARTS | DIVORCE | EDUCATION | ENTERTAINMENT | ENVIRONMENT | FIFTY | GOOD NEWS | HEALTHY LIVING | HOME & LIVING | IMPACT | WEDIA | MONEY | PARENTING | POLITICS | QUEER VOICES | RELIGION | SCIENCE | SPORTS | STYLE | STYLE & BEAUTY | TASTE | TECH | TRAVEL | WEDDINGS | WEIRD NEWS | WELLNESS | WOMEN  | MORLD |

05

0.2

- 0.1

#### Category Classification based on short description (CountVectorizer)

BLACK VOICES 0.01 0.22 0.01 0.01 0.02 0.10 0.03 0.01 0.02 0.07 0.01 0.01 0.00 0.01 0.00 0.01 0.03 0.02 0.01 0.02 0.03 0.02 0.03 0.01 0.05 0.09 0.04 0.01 0.01 0.01 0.01 0.03 0.01 COLLEGE 0.01 0.08 0.04 0.23 0.00 0.04 0.01 0.00 0.07 0.00 0.01 0.00 0.02 0.01 0.04 0.03 0.04 0.03 0.04 0.00 0.02 0.00 0.01 0.00 0.02 0.21 0.00 0.02 0.01 0.00 0.02 0.01 0.00 0.02 0.01 0.03 0.01 0.02 0.01 0.00 0.19 0.02 0.01 0.02 0.00 0.05 0.02 0.01 0.01 0.01 0.04 0.01 0.03 0.01 0.03 0.05 0.02 0.01 0.02 0.03 0.15 0.02 0.03 0.02 0.03 0.02 0.01 0.07 0.02 0.02 0.01 COMEDY CRIME CULTURE & ARTS DIVORCE EDUCATION 00 0.01 0.03 0.09 0.02 0.02 0.03 0.01 .36 0.00 0.01 0.02 0.00 0.01 0.04 0.07 0.02 0.03 0.04 0.02 0.01 0.01 0.01 0.02 0.09 0.01 0.01 0.03 0.01 0.01 0.00 0.02 0.01 0.02 ENTERTAINMEN ENVIRONMEN' 0.00 0.00 0.01 0.00 0.01 0.00 0.01 0.03 0.01 0.00 0.01 0.07 0.00 0.04 0.03 0.02 0.01 0.03 0.07 0.01 0.01 0.01 0.01 0.00 0.26 0.01 0.01 0.01 0.01 0.01 0.06 0.01 0.00 0.06 0.05 0.00 FIET GOOD NEWS HEALTHY LIVING HOME & LIVING 0.01 0.02 0.04 0.01 0.01 0.02 0.02 0.01 0.01 0.00 0.04 0.02 0.01 0.03 0.01 0.23 0.01 0.03 0.03 0.02 0.03 0.01 0.01 0.09 0.03 0.02 0.02 0.02 0.03 0.01 0.01 0.06 0.04 0.04 MEDIA MONEY 0.01 0.01 0.01 0.01 0.02 0.01 0.04 0.02 0.02 0.02 0.03 0.01 0.03 0.03 0.04 0.01 0.02 0.36 0.01 0.01 0.01 0.01 0.01 0.04 0.03 0.03 0.02 0.01 0.03 0.00 0.06 0.03 0.01 PARENTING POLITICS · RELIGION SCIENCE SPORTS STYLE STYLE & BEAUTY TASTE TECH · TRAVEL WEDDING WEIRD NEWS WELLNESS 0.01 0.01 0.03 0.01 0.00 0.00 0.02 0.03 0.01 0.00 0.03 0.03 0.00 0.09 0.01 0.05 0.00 0.02 0.06 0.00 0.02 0.01 0.03 0.01 0.00 0.01 0.04 0.01 0.03 0.00 0.03 0.02 0.01 0.01 0.04 0.01 0.01 0.02 0.03 0.02 0.02 0.00 0.02 0.01 0.06 0.01 0.02 0.01 0.04 0.03 0.01 0.03 0.01 0.03 0.01 0.01 0.12 0.04 0.02 0.02 0.01 0.01 0.01 0.08 0.20 0.0 WOMEN -

| <pre># Training time:</pre> | 50s       |        |          |         |
|-----------------------------|-----------|--------|----------|---------|
| #Classification             | report:   |        | 64       |         |
|                             | precision | recall | T1-score | support |
|                             | 0 30      | 0 12   | 0 18     | 282     |
|                             | 0.30      | 0.12   | 0.10     | 302     |
| BUSTNESS                    | 0.25      | 0.22   | 0.24     | 380     |
| COLLEGE                     | 0.20      | 0.23   | 0.24     | 212     |
| COMEDY                      | 0.24      | 0.19   | 0.21     | 300     |
| CRIME                       | 0.24      | 0 11   | 0.41     | 409     |
|                             | 0.30      | 0.36   | 0 38     | 301     |
|                             | 0.59      | 0.50   | 0.50     | /10     |
| EDUCATION                   | 0.00      | 0.36   | 0.50     | 108     |
|                             | 0.21      | 0.50   | 0.41     | 397     |
| ENVIRONMENT                 | 0.21      | 0.10   | 0.10     | 355     |
| ETETY                       | 0.33      | 0.30   | 0.35     | 272     |
|                             | 0.21      | 0.17   | 0.19     | 305     |
|                             | 0.22      | 0.05   | 0.00     | 386     |
|                             | 0.20      | 0.15   | 0.15     | 386     |
|                             | 0.37      | 0.40   | 0.41     | 100     |
| MEDIA                       | 0.20      | 0.23   | 0.23     | 305     |
| MONEY                       | 0.30      | 0.20   | 0.32     | 333     |
|                             | 0.42      | 0.40   | 0.45     | 301     |
|                             | 0.32      | 0.30   | 0.34     | 420     |
|                             | 0.32      | 0.30   | 0.31     | 420     |
|                             | 0 11      | 0.31   | 0.30     | 415     |
| SCIENCE                     | 0 43      | 0.27   | 0.35     | 440     |
| SPORTS                      | 0.43      | 0.33   | 0.37     | 414     |
|                             | 0.45      | 0.33   | 0.37     | 410     |
| STVLE & REALITY             | 0.11      | 0.44   | 0.17     | 301     |
|                             | 0.47      | 0.54   | 0.49     | 397     |
| TECH                        | 0.42      | 0.30   | 0.40     | A16     |
| TRAVEL                      | 0.40      | 0.39   | 0.40     | 410     |
| WEDDINGS                    | 0.61      | 0.64   | 0.62     | 405     |
| WETRD NEWS                  | 0.18      | 0.12   | 0.15     | 400     |
| WELLNESS                    | 0.20      | 0.12   | 0.15     | 400     |
| WOMEN                       | 0.26      | 0.20   | 0.22     | 407     |
| WORLD                       | 0.10      | 0.20   | 0.22     | 400     |
| WORLD                       | 0.40      | 0.50   | 0.59     | 404     |
| accuracy                    |           |        | 0.33     | 12824   |
| macro avg                   | 0.35      | 0.32   | 0.33     | 12824   |
| weighted avg                | 0.35      | 0.33   | 0.33     | 12824   |

|   | <pre># Training time: 2 #Classification re</pre> | 2s<br>port: |
|---|--|-------------|
|   | pr   | ecisio      |
| Evaluation  | ARTS & CULTURE                                   | 0.2         |
|   | BUSINESS   | 0.2         |
|   | COLLEGE  | 0.3         |
|   | COMEDY   | 0.2         |
| n based on 🛛 🗖  | CRIME  | 0.4         |
| (Veeterizer)  | CULTURE & ARTS                                   | 0.4         |
| vectorizer  | DIVORCE  | 0.6         |
| 0.03 0.02 0.04 0.00 0.02 0.03 0.01 0.00 0.06 0.00 0.06 0.01 | EDUCATION  | 0.4         |
| 0.00 0.04 0.09 0.02 0.01 0.01 0.01 0.01 0.01 0.01 0.03 0.01 | ENTERTAINMENT                                    | 0.2         |
| 0.01 0.01 0.14 0.01 0.02 0.05 0.01 0.01 0.01 0.04 0.02 0.02 | ENVIRONMENT                                      | 0.3         |
| 0.00 0.02 0.21 0.00 0.00 0.02 0.01 0.00 0.01 0.01           | FIFTY  | 0.2         |
|   | GOOD NEWS  | 0.1         |
|   |  | 0           |

- 0.3

- 0.2

- 0.1

#### **Results & Evaluation**

#### Category Classification based on short description (TfidfVectorizer)

| ARTS & CULTU | RE - 0.1               | 6 0.05       | 0.01     | 0.01    | 0.03   | 0.01  | 0.11           | 0.01    | 0.01      | 0.09          | 0.01        | 0.01  | 0.01      | 0.02           | 0.04          | 0.01   | 0.03  | 0.00  | 0.02      | 0.02     | 0.04         | 0.03 (   | 0.03    | 0.02   | 0.04 0 | 0.00           | 0.02  | 0.03 | 0.01   | 0.00     | 0.06       | 0.00     | 0.06  | ).01  |
|--------------|------------------------|--------------|----------|---------|--------|-------|----------------|---------|-----------|---------------|-------------|-------|-----------|----------------|---------------|--------|-------|-------|-----------|----------|--------------|----------|---------|--------|--------|----------------|-------|------|--------|----------|------------|----------|-------|-------|
| BLACK VOIC   | ES - 0.0               | 1 0.27       | 0.01     | 0.02    | 0.03   | 0.07  | 0.02           | 0.01    | 0.02      | 0.06          | 0.02        | 0.01  | 0.01      | 0.02           | 0.01          | 0.03   | 0.04  | 0.00  | 0.02      | 0.04     | 0.02         | 0.03 (   | 0.00    | 0.04   | 0.09 0 | 0.02           | 0.01  | 0.01 | 0.01   | 0.01     | 0.01       | 0.01     | 0.03  | 0.01  |
| BUSINE       | ss - 0.0               | 1 0.01       | 0.21     | 0.01    | 0.01   | 0.01  | 0.01           | 0.01    | 0.01      | 0.01          | 0.03        | 0.01  | 0.01      | 0.04           | 0.01          | 0.04   | 0.02  | 0.11  | 0.02      | 0.04     | 0.02         | 0.00 (   | 0.01    | 0.01   | 0.14 0 | 0.01           | 0.02  | 0.05 | 0.01   | 0.01     | 0.01       | 0.04     | 0.02  | ).02  |
| COLLE        | GE - 0.0               | 1 0.07       | 0.03     | 0.24    | 0.00   | 0.01  | 0.00           | 0.00    | 0.05      | 0.00          | 0.02        | 0.00  | 0.00      | 0.02           | 0.00          | 0.02   | 0.03  | 0.06  | 0.01      | 0.04     | 0.00         | 0.02 (   | 0.00    | 0.02   | 0.21 0 | 0.00           | 0.00  | 0.02 | 0.01   | 0.00     | 0.01       | 0.01     | 0.03  | 0.02  |
| COME         | DY - 0.0               | 1 0.04       | 0.02     | 0.00    | 0.20   | 0.01  | 0.01           | 0.01    | 0.00      | 0.06          | 0.02        | 0.01  | 0.02      | 0.02           | 0.02          | 0.01   | 0.04  | 0.02  | 0.03      | 0.03     | 0.04         | 0.01 0   | 0.02    | 0.03   | 0.16 0 | 0.01           | 0.03  | 0.02 | 0.00   | 0.00     | 0.05       | 0.02     | 0.04  | 0.01  |
| CRI          | ME - 0.0               | 0 0.07       | 0.00     | 0.01    | 0.01   | 0.37  | 0.01           | 0.00    | 0.00      | 0.01          | 0.02        | 0.01  | 0.01      | 0.01           | 0.00          | 0.02   | 0.01  | 0.00  | 0.01      | 0.03     | 0.01         | 0.01 (   | 0.00    | 0.01   | 0.22 0 | 0.00           | 0.00  | 0.01 | 0.01   | 0.00     | 0.04       | 0.01     | 0.01  | ).02  |
| CULTURE & AR | стs - <mark>0.0</mark> | 3 0.02       | 0.01     | 0.00    | 0.01   | 0.01  | 0.34           | 0.01    | 0.01      | 0.01          | 0.01        | 0.01  | 0.00      | 0.01           | 0.02          | 0.02   | 0.01  | 0.00  | 0.02      | 0.00     | 0.02         | 0.01     | 0.03    | 0.00   | 0.27 0 | 0.02           | 0.01  | 0.02 | 0.02   | 0.00     | 0.00       | 0.02     | 0.01  | 0.02  |
| DIVOR        | CE - 0.0               | 0 0.01       | 0.01     | 0.00    | 0.01   | 0.01  | 0.00           | 0.58    | 0.00      | 0.01          | 0.00        | 0.01  | 0.00      | 0.03           | 0.02          | 0.01   | 0.00  | 0.02  | 0.05      | 0.00     | 0.01         | 0.01     | 0.00    | 0.01   | 0.00 0 | 0.01           | 0.02  | 0.01 | 0.01   | 0.07     | 0.00       | 0.02     | 0.02  | 0.00  |
| EDUCATIO     | 0.0 - NC               | 1 0.02       | 0.02     | 0.09    | 0.00   | 0.02  | 0.02           | 0.00    | 0.33      | 0.00          | 0.01        | 0.01  | 0.00      | 0.01           | 0.03          | 0.08   | 0.02  | 0.04  | 0.08      | 0.04     | 0.02         | 0.01 (   | 0.02    | 0.02   | 0.09 0 | 0.00           | 0.00  | 0.02 | 0.01   | 0.00     | 0.00       | 0.02     | 0.01  | 0.01  |
| ENTERTAINME  | NT - 0.0               | 3 0.05       | 0.03     | 0.01    | 0.06   | 0.03  | 0.02           | 0.01    | 0.00      | 0.18          | 0.01        | 0.01  | 0.02      | 0.01           | 0.01          | 0.01   | 0.03  | 0.00  | 0.02      | 0.01     | 0.03         | 0.01 0   | 0.01    | 0.03   | 0.21 0 | 0.02           | 0.02  | 0.02 | 0.01   | 0.00     | 0.05       | 0.00     | 0.04  | 0.01  |
| ENVIRONMEI   | NT - 0.0               | 1 0.01       | 0.02     | 0.00    | 0.02   | 0.01  | 0.01           | 0.01    | 0.01      | 0.01          | 0.37        | 0.01  | 0.02      | 0.00           | 0.03          | 0.02   | 0.02  | 0.01  | 0.02      | 0.03     | 0.01         | 0.01     | 0.06    | 0.01   | 0.14 0 | 0.00           | 0.01  | 0.02 | 0.04   | 0.00     | 0.01       | 0.01     | 0.01  | 0.03  |
| FIF          | TY - 0.0               | 1 0.00       | 0.02     | 0.01    | 0.01   | 0.00  | 0.01           | 0.04    | 0.00      | 0.01          | 0.00        | 0.18  | 0.00      | 0.04           | 0.03          | 0.01   | 0.00  | 0.04  | 0.07      | 0.01     | 0.01         | 0.00     | 0.01    | 0.00   | 0.26 0 | 0.02           | 0.01  | 0.01 | 0.04   | 0.02     | 0.00       | 0.07     | 0.05  | 0.00  |
| GOOD NEV     | ws - 0.0               | 2 0.02       | 0.01     | 0.01    | 0.03   | 0.02  | 0.01           | 0.01    | 0.01      | 0.04          | 0.06        | 0.02  | 0.09      | 0.04           | 0.02          | 0.03   | 0.01  | 0.01  | 0.04      | 0.02     | 0.02         | 0.01 (   | 0.02    | 0.03   | 0.27 0 | 0.01           | 0.02  | 0.01 | 0.00   | 0.01     | 0.05       | 0.02     | 0.03  | 0.01  |
| HEALTHY LIVI | NG - 0.0               | 1 0.01       | 0.02     | 0.01    | 0.02   | 0.00  | 0.01           | 0.02    | 0.00      | 0.01          | 0.02        | 0.04  | 0.01      | 0.14           | 0.02          | 0.03   | 0.00  | 0.01  | 0.02      | 0.02     | 0.01         | 0.02 (   | 0.03    | 0.02   | 0.19   | 0.01           | 0.03  | 0.01 | 0.01   | 0.01     | 0.02       | 0.18     | 0.04  | 0.01  |
| HOME & LIVI  | NG - 0.0               | 2 0.01       | 0.02     | 0.00    | 0.03   | 0.00  | 0.02           | 0.02    | 0.01      | 0.03          | 0.02        | 0.02  | 0.01      | 0.02           | 0.48          | 0.01   | 0.00  | 0.02  | 0.02      | 0.01     | 0.01         | 0.01 (   | 0.02    | 0.01   | 0.03 0 | 0.03           | 0.04  | 0.02 | 0.03   | 0.01     | 0.02       | 0.01     | 0.01  | 0.00  |
| IMPA         | ст - 0.0               | 0 0.02       | 0.03     | 0.01    | 0.01   | 0.01  | 0.01           | 0.02    | 0.01      | 0.01          | 0.04        | 0.01  | 0.03      | 0.05           | 0.01          | 0.22   | 0.01  | 0.03  | 0.04      | 0.02     | 0.03         | 0.02 (   | 0.02    | 0.01   | 0.08 0 | 0.02           | 0.01  | 0.03 | 0.02   | 0.01     | 0.01       | 0.03     | 0.04  | 0.03  |
| Iabel label  | DIA - 0.0              | 2 0.04       | 0.05     | 0.01    | 0.05   | 0.02  | 0.01           | 0.00    | 0.01      | 0.03          | 0.01        | 0.01  | 0.00      | 0.01           | 0.00          | 0.02   | 0.27  | 0.01  | 0.01      | 0.08     | 0.01         | 0.01 (   | 0.01    | 0.02   | 0.18 0 | 0.01           | 0.01  | 0.03 | 0.00   | 0.01     | 0.02       | 0.01     | 0.02  | ).02  |
| MOM MON      | EY - 0.0               | 1 0.01       | 0.10     | 0.02    | 0.00   | 0.02  | 0.01           | 0.02    | 0.01      | 0.00          | 0.02        | 0.01  | 0.00      | 0.03           | 0.02          | 0.02   | 0.01  | 0.44  | 0.02      | 0.02     | 0.00         | 0.00     | 0.01    | 0.02   | 0.01 0 | 0.01           | 0.01  | 0.02 | 0.04   | 0.02     | 0.01       | 0.03     | 0.01  | ).01  |
| PARENTI      | vg - 0.0               | 1 0.02       | 0.00     | 0.01    | 0.03   | 0.01  | 0.01           | 0.04    | 0.02      | 0.02          | 0.02        | 0.05  | 0.02      | 0.04           | 0.03          | 0.04   | 0.01  | 0.02  | 0.31      | 0.01     | 0.03         | 0.01 (   | 0.02    | 0.01   | 0.04 0 | 0.02           | 0.03  | 0.02 | 0.01   | 0.02     | 0.01       | 0.05     | 0.05  | 0.01  |
| POLITI       | cs - 0.0               | 1 0.04       | 0.03     | 0.00    | 0.04   | 0.04  | 0.01           | 0.00    | 0.01      | 0.03          | 0.03        | 0.00  | 0.00      | 0.02           | 0.00          | 0.03   | 0.05  | 0.01  | 0.00      | 0.29     | 0.01         | 0.04 (   | 0.01    | 0.02   | 0.12 0 | 0.00           | 0.01  | 0.02 | 0.00   | 0.00     | 0.02       | 0.01     | 0.02  | 0.05  |
| QUEER VOIC   | ES - 0.0               | 2 0.03       | 0.01     | 0.00    | 0.02   | 0.01  | 0.02           | 0.01    | 0.01      | 0.03          | 0.01        | 0.00  | 0.02      | 0.01           | 0.01          | 0.02   | 0.01  | 0.00  | 0.03      | 0.03     | 0.35         | 0.02 0   | 0.01    | 0.02   | 0.15   | 0.01           | 0.01  | 0.01 | 0.01   | 0.02     | 0.01       | 0.01     | 0.03  | 0.01  |
| RELIGIO      | 0.0 - NC               | 1 0.03       | 0.01     | 0.01    | 0.00   | 0.02  | 0.02           | 0.00    | 0.01      | 0.01          | 0.01        | 0.02  | 0.01      | 0.02           | 0.00          | 0.03   | 0.02  | 0.00  | 0.01      | 0.03     | 0.02         | 0.30     | 0.01    | 0.02   | 0.25   | 0.01           | 0.00  | 0.01 | 0.02   | 0.00     | 0.01       | 0.01     | 0.02  | ).05  |
| SCIEN        | CE - 0.0               | 0 0.01       | 0.03     | 0.00    | 0.01   | 0.01  | 0.01           | 0.01    | 0.00      | 0.02          | 0.06        | 0.01  | 0.01      | 0.03           | 0.01          | 0.01   | 0.01  | 0.01  | 0.02      | 0.01     | 0.02         | 0.01     | 0.33    | 0.01   | 0.17 0 | 0.01           | 0.01  | 0.03 | 0.02   | 0.00     | 0.02       | 0.03     | 0.01  | ).02  |
| SPOR         | TS - 0.0               | 0 0.03       | 0.01     | 0.00    | 0.06   | 0.03  | 0.01           | 0.01    | 0.00      | 0.02          | 0.02        | 0.01  | 0.01      | 0.01           | 0.00          | 0.01   | 0.02  | 0.00  | 0.02      | 0.02     | 0.03         | 0.00 (   | 0.01    | 0.34   | 0.16 0 | 0.01           | 0.01  | 0.01 | 0.02   | 0.00     | 0.02       | 0.01     | 0.03  | ).02  |
| STY          | 'LE - 0.0              | 1 0.02       | 0.02     | 0.00    | 0.04   | 0.00  | 0.00           | 0.00    | 0.00      | 0.04          | 0.00        | 0.01  | 0.01      | 0.02           | 0.04          | 0.01   | 0.01  | 0.00  | 0.02      | 0.02     | 0.01         | 0.01 0   | 0.01    | 0.02   | 0.43 0 | 0.09           | 0.02  | 0.03 | 0.01   | 0.01     | 0.02       | 0.01     | 0.02  | 0.00  |
| STYLE & BEAU | TY - 0.0               | 0 0.02       | 0.01     | 0.00    | 0.03   | 0.00  | 0.03           | 0.01    | 0.01      | 0.02          | 0.00        | 0.02  | 0.00      | 0.01           | 0.05          | 0.01   | 0.00  | 0.01  | 0.03      | 0.01     | 0.02         | 0.00 (   | 0.01    | 0.01   | 0.07   | 0.51           | 0.02  | 0.02 | 0.02   | 0.01     | 0.01       | 0.01     | 0.05  | ).01  |
| TAS          | TE - 0.0               | 1 0.00       | 0.02     | 0.01    | 0.02   | 0.00  | 0.02           | 0.01    | 0.00      | 0.02          | 0.01        | 0.02  | 0.01      | 0.02           | 0.05          | 0.02   | 0.00  | 0.01  | 0.02      | 0.02     | 0.01         | 0.01 (   | 0.01    | 0.02   | 0.04 0 | 0.01           | 0.54  | 0.02 | 0.03   | 0.00     | 0.03       | 0.01     | 0.01  | ).01  |
| TE           | сн - 0.0               | 1 0.02       | 0.07     | 0.01    | 0.04   | 0.02  | 0.01           | 0.00    | 0.01      | 0.02          | 0.00        | 0.01  | 0.01      | 0.01           | 0.01          | 0.02   | 0.04  | 0.03  | 0.01      | 0.03     | 0.01         | 0.01 0   | 0.04    | 0.01   | 0.02 0 | 0.00           | 0.02  | 0.38 | 0.02   | 0.01     | 0.04       | 0.02     | 0.02  | ).03  |
| TRAV         | 'EL - 0.0              | 2 0.01       | 0.01     | 0.00    | 0.02   | 0.01  | 0.03           | 0.00    | 0.00      | 0.02          | 0.01        | 0.02  | 0.01      | 0.01           | 0.04          | 0.02   | 0.00  | 0.01  | 0.02      | 0.01     | 0.01         | 0.01 (   | 0.02    | 0.02   | 0.05 0 | 0.01           | 0.05  | 0.01 | 0.45   | 0.02     | 0.01       | 0.02     | 0.00  | ).02  |
| WEDDIN       | GS - 0.0               | 1 0.01       | 0.01     | 0.01    | 0.01   | 0.01  | 0.01           | 0.07    | 0.00      | 0.03          | 0.00        | 0.03  | 0.01      | 0.00           | 0.02          | 0.01   | 0.00  | 0.01  | 0.03      | 0.00     | 0.01         | 0.01 (   | 0.01 (  | 0.01   | 0.01 0 | 0.03           | 0.00  | 0.01 | 0.03   | 0.59     | 0.00       | 0.01     | 0.01  | ).01  |
| WEIRD NEV    | NS - 0.0               | 1 0.01       | 0.01     | 0.00    | 0.04   | 0.07  | 0.01           | 0.00    | 0.00      | 0.05          | 0.03        | 0.01  | 0.02      | 0.02           | 0.03          | 0.03   | 0.02  | 0.01  | 0.02      | 0.03     | 0.02         | 0.01 (   | 0.02    | 0.05   | 0.22 0 | 0.00           | 0.03  | 0.02 | 0.02   | 0.00     | 0.15       | 0.00     | 0.01  | ).01  |
| WELLNE       | SS - 0.0               | 0 0.00       | 0.02     | 0.01    | 0.01   | 0.01  | 0.02           | 0.03    | 0.01      | 0.00          | 0.02        | 0.03  | 0.00      | 0.12           | 0.02          | 0.04   | 0.00  | 0.02  | 0.08      | 0.00     | 0.02         | 0.02 (   | 0.03    | 0.00   | 0.01 0 | 0.01           | 0.03  | 0.02 | 0.02   | 0.01     | 0.00       | 0.32     | 0.04  | ).00  |
| WOM          | EN - 0.0               | 2 0.04       | 0.01     | 0.01    | 0.02   | 0.03  | 0.01           | 0.02    | 0.00      | 0.03          | 0.01        | 0.02  | 0.01      | 0.04           | 0.01          | 0.05   | 0.03  | 0.01  | 0.04      | 0.02     | 0.03         | 0.02 (   | 0.01    | 0.02   | 0.14 0 | 0.03           | 0.01  | 0.02 | 0.00   | 0.01     | 0.01       | 0.04     | 0.17  | ).01  |
| WOR          | LD - 0.0               | 1 0.02       | 0.03     | 0.00    | 0.01   | 0.04  | 0.01           | 0.00    | 0.00      | 0.01          | 0.03        | 0.00  | 0.01      | 0.00           | 0.00          | 0.05   | 0.03  | 0.00  | 0.01      | 0.07     | 0.01         | 0.02 (   | 0.02    | 0.01   | 0.16   | 0.00           | 0.00  | 0.01 | 0.02   | 0.00     | 0.01       | 0.00     | 0.00  | 0.34  |
|              | ARTS & CULTURE         | BLACK VOICES | BUSINESS | COLLEGE | COMEDY | CRIME | CULTURE & ARTS | DIVORCE | EDUCATION | ENTERTAINMENT | ENVIRONMENT | FIFTY | GOOD NEWS | HEALTHY LIVING | HOME & LIVING | IMPACT | WEDIA | d lab | PARENTING | POLITICS | QUEER VOICES | RELIGION | SCIENCE | SPORTS | STYLE  | STYLE & BEAUTY | TASTE | TECH | TRAVEL | WEDDINGS | WEIRD NEWS | WELLNESS | WOMEN | WORLD |

| #Classification | : ZZS     |           |          |         |
|-----------------|-----------|-----------|----------|---------|
| #CLASSIFICATION | report:   | ma aa 1 1 | £1       | aunnant |
|                 | precision | recall    | TI-Score | support |
| ARTS & CUILTURE | 0.24      | 0.16      | 0,19     | 282     |
|                 | 0.24      | 0.10      | 0.15     | 302     |
| BUSTNESS        | 0.2/      | 0.21      | 0.27     | 380     |
|                 | 0.24      | 0.21      | 0.23     | 212     |
|                 | 0.32      | 0.24      | 0.27     | 200     |
| COMEDI          | 0.21      | 0.20      | 0.20     | 400     |
|                 | 0.41      | 0.3/      | 0.39     | 201     |
|                 | 0.42      | 0.54      | 0.50     | 410     |
|                 | 0.01      | 0.00      | 0.39     | 419     |
|                 | 0.44      | 0.33      | 0.30     | 190     |
|                 | 0.20      | 0.10      | 0.19     | 207     |
|                 | 0.37      | 0.37      | 0.37     | 200     |
|                 | 0.21      | 0.10      | 0.19     | 2/3     |
|                 | 0.18      | 0.09      | 0.12     | 305     |
|                 | 0.10      | 0.14      | 0.15     | 380     |
| HUME & LIVING   | 0.44      | 0.48      | 0.40     | 380     |
|                 | 0.24      | 0.22      | 0.23     | 400     |
| MEDIA           | 0.34      | 0.27      | 0.30     | 395     |
| MONEY           | 0.43      | 0.44      | 0.44     | 324     |
| PARENTING       | 0.27      | 0.31      | 0.29     | 391     |
| POLITICS        | 0.30      | 0.29      | 0.29     | 420     |
| QUEER VOICES    | 0.40      | 0.35      | 0.37     | 415     |
| RELIGION        | 0.45      | 0.30      | 0.36     | 440     |
| SCIENCE         | 0.40      | 0.33      | 0.36     | 414     |
| SPORTS          | 0.41      | 0.34      | 0.37     | 410     |
| STYLE           | 0.10      | 0.43      | 0.17     | 413     |
| STYLE & BEAUTY  | 0.53      | 0.51      | 0.52     | 391     |
| TASTE           | 0.51      | 0.54      | 0.53     | 397     |
| TECH            | 0.40      | 0.38      | 0.39     | 416     |
| TRAVEL          | 0.47      | 0.45      | 0.46     | 405     |
| WEDDINGS        | 0.66      | 0.59      | 0.63     | 400     |
| WEIRD NEWS      | 0.21      | 0.15      | 0.17     | 408     |
| WELLNESS        | 0.31      | 0.32      | 0.31     | 407     |
| WOMEN           | 0.18      | 0.17      | 0.18     | 400     |
| WORLD           | 0.42      | 0.34      | 0.37     | 404     |
| accuracy        |           |           | 0.33     | 12824   |
| macro avg       | 0.35      | 0.32      | 0.33     | 12824   |
| weighted avg    | 0.35      | 0.33      | 0.33     | 12824   |
|                 |           |           |          |         |

0.7

0.6

0.5

0.4

0.3

- 0.2

0.1

#### Category Classification based on headline and short description (CountVectorizer)

| ARTS & CULTURE | 0.41          | 0.04        | 0.00    | 0.00   | 0.04  | 0.00 | 0.12          | 0.00   | 0.00     | 0.08         | 0.02       | 0.02 | 0.00     | 0.00          | 0.00         | 0.00    | 0.01 | 0.01 | 0.02     | 0.02    | 0.04        | 0.02    | 0.01   | 0.01  | 0.02    | 0.00          | 0.01 | 0.02 | 0.01  | 0.01    | 0.02      | 0.01    | 0.02 | 0.01 |
|----------------|---------------|-------------|---------|--------|-------|------|---------------|--------|----------|--------------|------------|------|----------|---------------|--------------|---------|------|------|----------|---------|-------------|---------|--------|-------|---------|---------------|------|------|-------|---------|-----------|---------|------|------|
| BLACK VOICES   | 0.01          |             | 0.01    | 0.03   | 0.02  | 0.08 | 0.02          | 0.00   | 0.02     | 0.08         | 0.00       | 0.01 | 0.01     | 0.01          | 0.01         | 0.04    | 0.03 | 0.00 | 0.02     | 0.04    | 0.01        | 0.03    | 0.01   | 0.05  | 0.01    | 0.03          | 0.01 | 0.01 | 0.01  | 0.01    | 0.01      | 0.00    | 0.02 | 0.00 |
| BUSINESS       | 0.00          | 0.01        |         | 0.01   | 0.01  | 0.02 | 0.00          | 0.01   | 0.00     | 0.00         | 0.03       | 0.01 | 0.00     | 0.03          | 0.01         | 0.04    | 0.03 | 0.09 | 0.01     | 0.03    | 0.02        | 0.00    | 0.01   | 0.02  | 0.01    | 0.01          | 0.01 | 0.06 | 0.01  | 0.01    | 0.01      | 0.04    | 0.02 | 0.02 |
| COLLEGE        | 0.00          | 0.06        | 0.02    | 0.50   | 0.00  | 0.03 | 0.01          | 0.00   | 0.06     | 0.00         | 0.00       | 0.00 | 0.00     | 0.01          | 0.01         | 0.02    | 0.04 | 0.02 | 0.00     | 0.05    | 0.01        | 0.01    | 0.00   | 0.01  | 0.01    | 0.00          | 0.00 | 0.00 | 0.00  | 0.00    | 0.00      | 0.02    | 0.04 | 0.00 |
| COMEDY         | 0.02          | 0.01        | 0.02    | 0.01   |       | 0.01 | 0.01          | 0.01   | 0.00     | 0.06         | 0.03       | 0.01 | 0.00     | 0.01          | 0.02         | 0.01    | 0.03 | 0.00 | 0.02     | 0.06    | 0.01        | 0.01    | 0.02   | 0.02  | 0.02    | 0.01          | 0.01 | 0.02 | 0.01  | 0.01    | 0.07      | 0.00    | 0.02 | 0.01 |
| CRIME          | 0.00          | 0.06        | 0.01    | 0.00   | 0.00  | 0.66 | 0.00          | 0.00   | 0.01     | 0.01         | 0.02       | 0.00 | 0.02     | 0.00          | 0.00         | 0.01    | 0.00 | 0.00 | 0.01     | 0.02    | 0.01        | 0.01    | 0.00   | 0.01  | 0.00    | 0.00          | 0.00 | 0.01 | 0.01  | 0.00    | 0.05      | 0.00    | 0.02 | 0.02 |
| CULTURE & ARTS | 0.08          | 0.02        | 0.01    | 0.00   | 0.02  | 0.01 | 0.55          | 0.01   | 0.00     | 0.03         | 0.01       | 0.01 | 0.00     | 0.01          | 0.02         | 0.01    | 0.00 | 0.00 | 0.00     | 0.00    | 0.01        | 0.01    | 0.03   | 0.01  | 0.02    | 0.01          | 0.01 | 0.01 | 0.03  | 0.00    | 0.02      | 0.02    | 0.01 | 0.02 |
| DIVORCE        | 0.00          | 0.00        | 0.01    | 0.00   | 0.00  | 0.00 | 0.00          | 0.75   | 0.00     | 0.00         | 0.00       | 0.01 | 0.01     | 0.01          | 0.01         | 0.00    | 0.00 | 0.01 | 0.03     | 0.00    | 0.01        | 0.00    | 0.00   | 0.00  | 0.00    | 0.00          | 0.00 | 0.01 | 0.00  | 0.04    | 0.00      | 0.02    | 0.02 | 0.00 |
| EDUCATION      | 0.01          | 0.02        | 0.03    | 0.08   | 0.00  | 0.02 | 0.03          | 0.00   | 0.46     | 0.00         | 0.01       | 0.01 | 0.01     | 0.02          | 0.01         | 0.07    | 0.02 | 0.04 | 0.06     | 0.03    | 0.01        | 0.01    | 0.01   | 0.02  | 0.01    | 0.00          | 0.00 | 0.01 | 0.00  | 0.00    | 0.00      | 0.03    | 0.01 | 0.02 |
| ENTERTAINMENT  | 0.01          | 0.04        | 0.01    | 0.00   | 0.06  | 0.02 | 0.03          | 0.01   | 0.00     | 0.42         | 0.00       | 0.00 | 0.03     | 0.01          | 0.00         | 0.00    | 0.04 | 0.00 | 0.03     | 0.02    | 0.03        | 0.01    | 0.01   | 0.03  | 0.05    | 0.01          | 0.01 | 0.02 | 0.01  | 0.01    | 0.04      | 0.01    | 0.03 | 0.01 |
| ENVIRONMENT    | 0.01          | 0.01        | 0.03    | 0.00   | 0.01  | 0.01 | 0.01          | 0.00   | 0.00     | 0.01         | 0.50       | 0.01 | 0.05     | 0.00          | 0.02         | 0.05    | 0.02 | 0.01 | 0.01     | 0.02    | 0.00        | 0.01    | 0.05   | 0.00  | 0.01    | 0.01          | 0.01 | 0.01 | 0.02  | 0.00    | 0.04      | 0.02    | 0.00 | 0.03 |
| FIFTY          | 0.01          | 0.00        | 0.01    | 0.01   | 0.01  | 0.00 | 0.01          | 0.02   | 0.01     | 0.01         | 0.01       | 0.33 | 0.01     | 0.06          | 0.01         | 0.03    | 0.01 | 0.03 | 0.07     | 0.01    | 0.02        | 0.01    | 0.00   | 0.01  | 0.03    | 0.01          | 0.02 | 0.03 | 0.04  | 0.02    | 0.01      | 0.08    | 0.05 | 0.01 |
| GOOD NEWS      | 0.00          | 0.02        | 0.01    | 0.00   | 0.01  | 0.02 | 0.01          | 0.00   | 0.02     | 0.02         | 0.05       | 0.01 | 0.36     | 0.01          | 0.02         | 0.04    | 0.01 | 0.01 | 0.07     | 0.00    | 0.01        | 0.01    | 0.01   | 0.04  | 0.04    | 0.00          | 0.02 | 0.01 | 0.01  | 0.02    | 0.11      | 0.02    | 0.01 | 0.01 |
| HEALTHY LIVING | 0.00          | 0.01        | 0.01    | 0.01   | 0.01  | 0.01 | 0.01          | 0.02   | 0.01     | 0.00         | 0.01       | 0.04 | 0.01     | 0.28          | 0.01         | 0.03    | 0.00 | 0.01 | 0.03     | 0.02    | 0.01        | 0.03    | 0.03   | 0.01  | 0.02    | 0.00          | 0.03 | 0.02 | 0.01  | 0.01    | 0.02      | 0.23    | 0.03 | 0.02 |
| HOME & LIVING  | 0.01          | 0.00        | 0.00    | 0.00   | 0.01  | 0.01 | 0.02          | 0.00   | 0.00     | 0.00         | 0.01       | 0.01 | 0.00     | 0.02          | 0.76         | 0.01    | 0.00 | 0.01 | 0.02     | 0.00    | 0.00        | 0.00    | 0.00   | 0.01  | 0.01    | 0.03          | 0.02 | 0.01 | 0.03  | 0.01    | 0.00      | 0.02    | 0.01 | 0.00 |
| IMPACT         | 0.01          | 0.03        | 0.02    | 0.00   | 0.00  | 0.00 | 0.01          | 0.01   | 0.02     | 0.01         | 0.06       | 0.02 | 0.03     | 0.05          | 0.01         | 0.38    | 0.01 | 0.04 | 0.05     | 0.03    | 0.03        | 0.02    | 0.01   | 0.01  | 0.01    | 0.01          | 0.01 | 0.02 | 0.01  | 0.00    | 0.00      | 0.04    | 0.04 | 0.04 |
| AIDEM MEDIA    | 0.00          | 0.02        | 0.01    | 0.01   | 0.07  | 0.02 | 0.01          | 0.00   | 0.01     | 0.02         | 0.01       | 0.00 | 0.01     | 0.01          | 0.00         | 0.01    | 0.56 | 0.01 | 0.01     | 0.08    | 0.01        | 0.01    | 0.01   | 0.03  | 0.00    | 0.01          | 0.00 | 0.03 | 0.01  | 0.00    | 0.02      | 0.01    | 0.01 | 0.02 |
| P MONEY        | 0.00          | 0.00        | 0.07    | 0.02   | 0.00  | 0.02 | 0.01          | 0.00   | 0.00     | 0.01         | 0.02       | 0.04 | 0.00     | 0.01          | 0.02         | 0.02    | 0.00 | 0.56 | 0.01     | 0.01    | 0.00        | 0.00    | 0.00   | 0.01  | 0.01    | 0.01          | 0.01 | 0.03 | 0.02  | 0.02    | 0.02      | 0.02    | 0.00 | 0.00 |
| PARENTING      | 0.01          | 0.01        | 0.01    | 0.01   | 0.02  | 0.01 | 0.01          | 0.01   | 0.02     | 0.02         | 0.01       | 0.05 | 0.03     | 0.02          | 0.01         | 0.05    | 0.01 | 0.01 | 0.49     | 0.01    | 0.01        | 0.01    | 0.01   | 0.01  | 0.01    | 0.02          | 0.02 | 0.01 | 0.01  | 0.01    | 0.02      | 0.03    | 0.04 | 0.00 |
| POLITICS       | -0.01         | 0.03        | 0.03    | 0.00   | 0.03  | 0.03 | 0.00          | 0.00   | 0.00     | 0.02         | 0.02       | 0.00 | 0.00     | 0.01          | 0.00         | 0.02    | 0.05 | 0.02 | 0.00     | 0.52    | 0.01        | 0.03    | 0.01   | 0.01  | 0.01    | 0.00          | 0.00 | 0.00 | 0.01  | 0.00    | 0.01      | 0.00    | 0.03 | 0.05 |
| QUEER VOICES   | 0.02          | 0.01        | 0.00    | 0.00   | 0.02  | 0.02 | 0.00          | 0.01   | 0.01     | 0.04         | 0.00       | 0.01 | 0.01     | 0.01          | 0.00         | 0.02    | 0.01 | 0.00 | 0.02     | 0.03    | 0.61        | 0.01    | 0.00   | 0.02  | 0.01    | 0.00          | 0.00 | 0.01 | 0.00  | 0.02    | 0.01      | 0.01    | 0.02 | 0.00 |
| RELIGION       | 0.01          | 0.03        | 0.00    | 0.00   | 0.01  | 0.02 | 0.02          | 0.00   | 0.00     | 0.00         | 0.01       | 0.02 | 0.02     | 0.02          | 0.01         | 0.03    | 0.01 | 0.01 | 0.02     | 0.02    | 0.03        | 0.56    | 0.01   | 0.01  | 0.00    | 0.01          | 0.00 | 0.00 | 0.01  | 0.01    | 0.01      | 0.01    | 0.02 | 0.05 |
| SCIENCE        | 0.00          | 0.01        | 0.01    | 0.01   | 0.02  | 0.01 | 0.01          | 0.00   | 0.00     | 0.01         | 0.06       | 0.01 | 0.01     | 0.03          | 0.01         | 0.00    | 0.00 | 0.01 | 0.01     | 0.01    | 0.01        | 0.01    | 0.59   | 0.01  | 0.00    | 0.00          | 0.02 | 0.02 | 0.02  | 0.00    | 0.04      | 0.04    | 0.01 | 0.01 |
| SPORTS         | - 0.00        | 0.03        | 0.01    | 0.00   | 0.02  | 0.04 | 0.01          | 0.00   | 0.00     | 0.02         | 0.01       | 0.01 | 0.01     | 0.00          | 0.00         | 0.00    | 0.01 | 0.00 | 0.00     | 0.01    | 0.00        | 0.01    | 0.01   | 0.63  | 0.02    | 0.01          | 0.00 | 0.01 | 0.01  | 0.00    | 0.03      | 0.00    | 0.02 | 0.02 |
| STYLE          | -0.00         | 0.02        | 0.01    | 0.00   | 0.01  | 0.00 | 0.02          | 0.00   | 0.00     | 0.05         | 0.01       | 0.01 | 0.01     | 0.01          | 0.02         | 0.01    | 0.02 | 0.00 | 0.00     | 0.01    | 0.00        | 0.00    | 0.01   | 0.02  | 0.54    | 0.07          | 0.02 | 0.01 | 0.02  | 0.01    | 0.02      | 0.00    | 0.02 | 0.00 |
| STYLE & BEAUTY | - 0.00        | 0.01        | 0.01    | 0.00   | 0.01  | 0.00 | 0.01          | 0.00   | 0.01     | 0.02         | 0.01       | 0.01 | 0.00     | 0.01          | 0.03         | 0.01    | 0.01 | 0.01 | 0.00     | 0.00    | 0.00        | 0.00    | 0.00   | 0.01  | 0.11    | 0.68          | 0.00 | 0.01 | 0.02  | 0.01    | 0.00      | 0.02    | 0.02 | 0.00 |
| TASTE          | - 0.00        | 0.01        | 0.01    | 0.00   | 0.02  | 0.00 | 0.01          | 0.00   | 0.00     | 0.02         | 0.01       | 0.01 | 0.00     | 0.02          | 0.03         | 0.02    | 0.00 | 0.01 | 0.01     | 0.01    | 0.00        | 0.01    | 0.01   | 0.01  | 0.03    | 0.01          | 0.70 | 0.00 | 0.03  | 0.00    | 0.02      | 0.01    | 0.00 | 0.00 |
| TECH           | -0.00         | 0.01        | 0.05    | 0.00   | 0.02  | 0.01 | 0.00          | 0.00   | 0.00     | 0.02         | 0.01       | 0.00 | 0.00     | 0.01          | 0.01         | 0.02    | 0.02 | 0.01 | 0.01     | 0.02    | 0.00        | 0.01    | 0.02   | 0.00  | 0.01    | 0.00          | 0.01 | 0.62 | 0.01  | 0.00    | 0.02      | 0.01    | 0.01 | 0.03 |
| TRAVEL         | -0.00         | 0.00        | 0.01    | 0.00   | 0.02  | 0.00 | 0.04          | 0.00   | 0.00     | 0.01         | 0.02       | 0.02 | 0.00     | 0.00          | 0.03         | 0.01    | 0.00 | 0.01 | 0.01     | 0.00    | 0.01        | 0.01    | 0.00   | 0.00  | 0.02    | 0.01          | 0.05 | 0.00 | 0.62  | 0.01    | 0.02      | 0.00    | 0.00 | 0.01 |
| WEDDINGS       | 0.00          | 0.00        | 0.01    | 0.00   | 0.01  | 0.00 | 0.01          | 0.05   | 0.00     | 0.01         | 0.00       | 0.01 | 0.01     | 0.01          | 0.01         | 0.00    | 0.00 | 0.01 | 0.01     | 0.00    | 0.01        | 0.00    | 0.00   | 0.01  | 0.01    | 0.02          | 0.00 | 0.01 | 0.01  | 0.76    | 0.01      | 0.02    | 0.01 | 0.00 |
| WEIRD NEWS     | 0.01          | 0.01        | 0.00    | 0.00   | 0.04  | 0.10 | 0.01          | 0.00   | 0.00     | 0.03         | 0.03       | 0.00 | 0.05     | 0.01          | 0.01         | 0.01    | 0.02 | 0.02 | 0.02     | 0.01    | 0.00        | 0.01    | 0.03   | 0.02  | 0.00    | 0.00          | 0.05 | 0.02 | 0.02  | 0.02    | 0.35      | 0.00    | 0.00 | 0.01 |
| WELLINESS      | 0.01          | 0.00        | 0.02    | 0.00   | 0.01  | 0.00 | 0.02          | 0.01   | 0.01     | 0.00         | 0.02       | 0.04 | 0.01     | 0.10          | 0.01         | 0.03    | 0.01 | 0.01 | 0.03     | 0.00    | 0.01        | 0.02    | 0.03   | 0.01  | 0.00    | 0.00          | 0.02 | 0.01 | 0.02  | 0.00    | 0.01      | 0.92    | 0.03 | 0.00 |
| WORLD          | 0.01          | 0.03        | 0.01    | 0.01   | 0.00  | 0.01 | 0.02          | 0.00   | 0.00     | 0.01         | 0.00       | 0.00 | 0.01     | 0.00          | 0.01         | 0.04    | 0.03 | 0.00 | 0.00     | 0.04    | 0.00        | 0.02    | 0.01   | 0.02  | 0.00    | 0.01          | 0.01 | 0.01 | 0.00  | 0.01    | 0.01      | 0.00    | 0.01 | 0.63 |
| WORLD          | <u>u</u>      | s.or        | 5       | ш      | ~     | ш    | 5.02          |        | z        | -            | -          | ×    | 5.01     | 5             | 5.00         | -       | 4    | ×    | .00      | 5.04    | 5.00        | 2       | ш.     | s.oz  | ш.<br>ш | ~             | ш.   | Ť    | 1     | 5.00    | s         | s.oo    | z    | 0.05 |
|                | ARTS & CULTUR | BLACK VOICE | BUSINES | COLLEG | COMED | CRIM | CULTURE & ART | DIVORC | EDUCATIO | ENTERTAINMEN | ENVIRONMEN | FIFT | GOOD NEW | HEALTHY LIVIN | HOME & LIVIN | g IMPAC | WEDI | MONE | PARENTIN | POLITIC | QUEER VOICE | RELIGIO | SCIENC | SPORT | STYL    | STYLE & BEAUT | TAST | TEC  | TRAVE | WEDDING | WEIRD NEW | WELLNES | WOME | WORL |

| <pre># Training time #Classification</pre> | : 54s<br>report: |        |          |         |
|--|------------------|--------|----------|---------|
|  | precision        | recall | f1-score | support |
| ARTS & CULTURE                             | 0.52             | 0.41   | 0.46     | 282     |
| BLACK VOICES                               | 0.44             | 0.40   | 0.42     | 392     |
| BUSINESS                                   | 0.48             | 0.41   | 0.44     | 380     |
| COLLEGE                                    | 0.59             | 0.50   | 0.54     | 212     |
| COMEDY                                     | 0.48             | 0.45   | 0.47     | 399     |
| CRIME                                      | 0.56             | 0.66   | 0.61     | 409     |
| CULTURE & ARTS                             | 0.54             | 0.55   | 0.55     | 391     |
| DIVORCE                                    | 0.82             | 0.75   | 0.79     | 419     |
| EDUCATION                                  | 0.56             | 0.46   | 0.50     | 198     |
| ENTERTAINMENT                              | 0.40             | 0.42   | 0.41     | 387     |
| ENVIRONMENT                                | 0.47             | 0.50   | 0.49     | 355     |
| FIFTY                                      | 0.35             | 0.33   | 0.34     | 273     |
| GOOD NEWS                                  | 0.44             | 0.36   | 0.40     | 305     |
| HEALTHY LIVING                             | 0.31             | 0.28   | 0.29     | 386     |
| HOME & LIVING                              | 0.68             | 0.76   | 0.71     | 386     |
| IMPACT                                     | 0.37             | 0.38   | 0.37     | 400     |
| MEDIA                                      | 0.56             | 0.56   | 0.56     | 395     |
| MONEY                                      | 0.55             | 0.56   | 0.56     | 324     |
| PARENTING                                  | 0.45             | 0.49   | 0.47     | 391     |
| POLITICS                                   | 0.48             | 0.52   | 0.50     | 420     |
| QUEER VOICES                               | 0.65             | 0.61   | 0.63     | 415     |
| RELIGION                                   | 0.63             | 0.56   | 0.59     | 440     |
| SCIENCE                                    | 0.63             | 0.59   | 0.61     | 414     |
| SPORTS                                     | 0.59             | 0.63   | 0.61     | 410     |
| SIYLE                                      | 0.50             | 0.54   | 0.52     | 413     |
| SIYLE & BEAULY                             | 0.68             | 0.68   | 0.68     | 391     |
| TASTE                                      | 0.64             | 0.70   | 0.67     | 397     |
|  | 0.59             | 0.02   | 0.00     | 410     |
|  | 0.00             | 0.02   | 0.01     | 405     |
|  | 0.75             | 0.70   | 0.74     | 400     |
| WELLNESS                                   | 0.30             | 0.33   | 0.33     | 400     |
| WELLINESS                                  | 0.37             | 0.42   | 0.39     | 407     |
| WORLD                                      | 0.42             | 0.42   | 0.42     | 400     |
| WURLD                                      | 0.35             | 0.03   | 0.01     | 404     |
| accuracy                                   |                  |        | 0.53     | 12824   |
| macro avg                                  | 0.53             | 0.53   | 0.53     | 12824   |
| weighted avg                               | 0.53             | 0.53   | 0.53     | 12824   |
|  |                  |        |          |         |

0.6

04

0.3

0.2

0.1

#### Category Classification based on headline and short description (TfidfVectorizer)



| <pre># Training time</pre> | : 27s     |        |          |         |
|----------------------------|-----------|--------|----------|---------|
| #Classification            | report:   |        |          |         |
|                            | precision | recall | f1-score | support |
| ARTS & CULTURE             | 0.50      | 0.41   | 0.45     | 282     |
| BLACK VOICES               | 0.48      | 0.46   | 0.47     | 392     |
| BUSINESS                   | 0.41      | 0.42   | 0.41     | 380     |
| COLLEGE                    | 0.50      | 0.49   | 0.49     | 212     |
| COMEDY                     | 0.49      | 0.49   | 0.49     | 399     |
| CRIME                      | 0.61      | 0.62   | 0.62     | 409     |
| CULTURE & ARTS             | 0.56      | 0.58   | 0.57     | 391     |
| DIVORCE                    | 0.80      | 0.76   | 0.78     | 419     |
| EDUCATION                  | 0.56      | 0.47   | 0.51     | 198     |
| ENTERTAINMENT              | 0.42      | 0.49   | 0.45     | 387     |
| ENVIRONMENT                | 0.51      | 0.56   | 0.53     | 355     |
| FIFTY                      | 0.38      | 0.38   | 0.38     | 273     |
| GOOD NEWS                  | 0.47      | 0.36   | 0.41     | 305     |
| HEALTHY LIVING             | 0.26      | 0.29   | 0.28     | 386     |
| HOME & LIVING              | 0.74      | 0.74   | 0.74     | 386     |
| IMPACT                     | 0.37      | 0.39   | 0.38     | 400     |
| MEDIA                      | 0.60      | 0.58   | 0.59     | 395     |
| MONEY                      | 0.57      | 0.54   | 0.55     | 324     |
| PARENTING                  | 0.46      | 0.49   | 0.48     | 391     |
| POLITICS                   | 0.47      | 0.50   | 0.48     | 420     |
| QUEER VOICES               | 0.70      | 0.64   | 0.67     | 415     |
| RELIGION                   | 0.66      | 0.61   | 0.63     | 440     |
| SCIENCE                    | 0.61      | 0.59   | 0.60     | 414     |
| SPORTS                     | 0.67      | 0.70   | 0.69     | 410     |
| STYLE                      | 0.55      | 0.57   | 0.56     | 413     |
| STYLE & BEAUTY             | 0.72      | 0.71   | 0.71     | 391     |
| TASTE                      | 0.68      | 0.71   | 0.69     | 397     |
| TECH                       | 0.63      | 0.62   | 0.62     | 416     |
| TRAVEL                     | 0.64      | 0.62   | 0.63     | 405     |
| WEDDINGS                   | 0.79      | 0.73   | 0.76     | 400     |
| WEIRD NEWS                 | 0.39      | 0.42   | 0.40     | 408     |
| WELLNESS                   | 0.36      | 0.39   | 0.37     | 407     |
| WOMEN                      | 0.41      | 0.42   | 0.41     | 400     |
| WORLD                      | 0.64      | 0.66   | 0.65     | 404     |
| accuracy                   |           |        | 0.55     | 12824   |
| macro avg                  | 0.55      | 0.54   | 0.54     | 12824   |
| weighted avg               | 0.55      | 0.55   | 0.55     | 12824   |
|                            |           |        |          |         |

# Conclusion

**Logistic Regression:** 

- supervised machine learning classifier
- binary classification or multiclass classification
- Iearning weights from a labelled training set via loss function
- minimizing the loss function using gradient descent
- using regularization to avoid overfitting

# Summary

Inear regression vs perceptron vs logistic regression

- linear regression predicts continuous value
- perceptron = linear regression + threshold function
- logistic regression = perceptron + sigmoid function
- cost fuction
  - cross entropy loss
- gradient descent

# Summary

- regularization
  - L1 (Lasso) vs L2 (Ridge) regularization
- feature extraction & parameter adjustment
  - tfidfvectorizer vs countvectorizer
  - min\_df, max\_df, C, multi\_class, solver, tol
- results & evaluation

### Reference

- Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4. No. 4. New York: springer, 2006.
- 2. Jurafsky, Dan and Martin, James H.(2021): Speech and Language Processing
- 3. Unsupervised Feature Learning and Deep Learning Tutorial. Retrieved December 3, 2022, from <a href="http://ufldl.stanford.edu/tutorial/supervised/SoftmaxRegression/">http://ufldl.stanford.edu/tutorial/supervised/SoftmaxRegression/</a>
- 4. Sklearn.linear\_model.logisticregression. scikit. (n.d.). Retrieved December 3, 2022, from <a href="https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LogisticRegression.html">https://scikit-learn.org/stable/modules/generated/sklearn.linear\_model.LogisticRegression.html</a>
- 5. Sklearn.feature\_extraction.text.CountVectorizer. scikit. (n.d.). Retrieved December 3, 2022, from <a href="https://scikit-">https://scikit-</a>

<u>learn.org/stable/modules/generated/sklearn.feature\_extraction.text.CountVectorizer.html?highlig</u> <u>ht=countvectorizer#sklearn.feature\_extraction.text.CountVectorizer</u>

Sklearn.feature\_extraction.text.TfidfVectorizer. scikit. (n.d.). Retrieved December 3, 2022, from <a href="https://scikit-">https://scikit-</a>

<u>learn.org/stable/modules/generated/sklearn.feature\_extraction.text.TfidfVectorizer.html#sklearn.f</u> <u>eature\_extraction.text.TfidfVectorizer</u>

#### **THANK YOU FOR LISTENING**

Group E: Logistic Regression