

Финальный проект. Задание в R

Марина Филатова

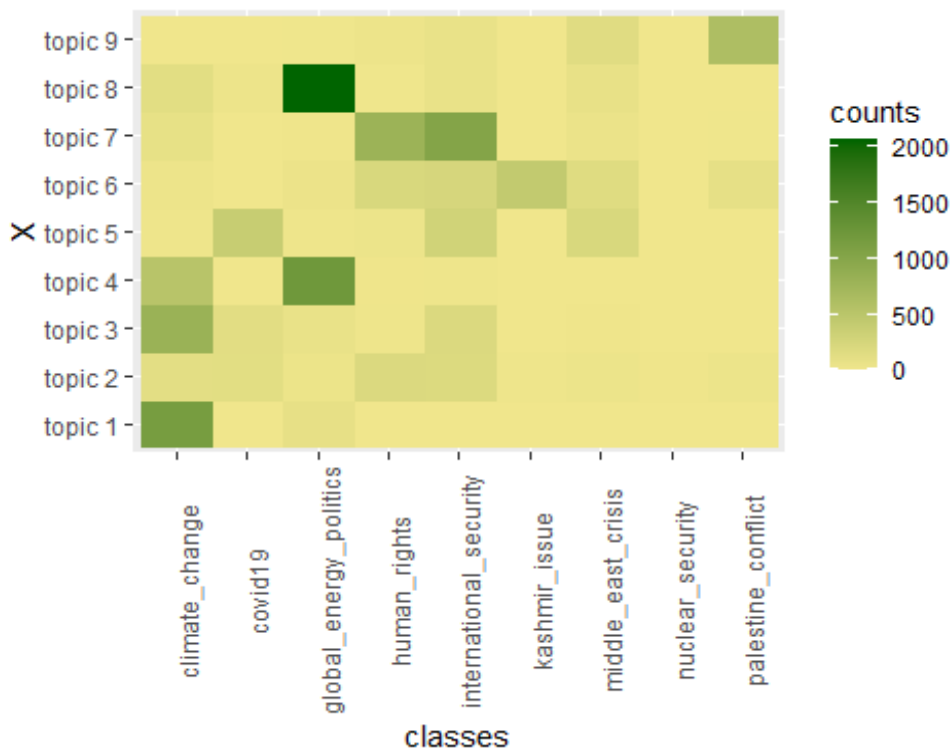
05 07 2022

Представленные графики моделей Латентного размещения Дирихле LDA, Анализа главных компонент PCA, Факторного анализа, k средних, Берч Birch, иерархической агломеративной кластеризации Feature Agglomeration для тематического моделирования показывают распределение известных категорий (classes) в определенных моделью темах (topics). Таким образом, видно, что лучше отработали модель LDA и PCA, хуже остальных иерархическая агломеративная кластеризация.

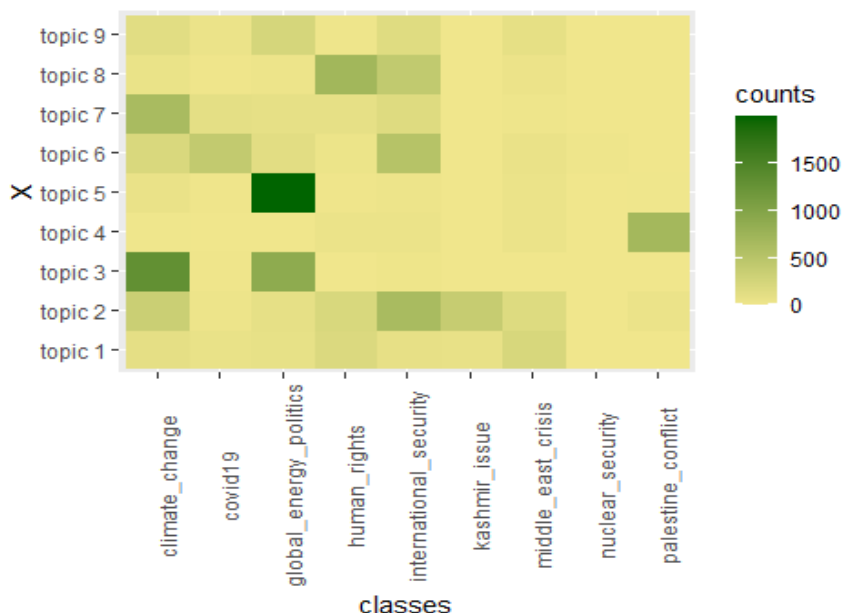
```
library(tidyverse)
```

```
library(ggplot2)
```

```
lda_df <- read.csv("C:/Users/rigel/Desktop/ВШЭ/финальный_проект/lda_df.csv")
lda_df %>%
  pivot_longer(-c(X), names_to = "classes", values_to = "counts") %>%
  ggplot(aes(x=classes, y=X, fill=counts)) +
  geom_raster() +
  scale_fill_gradient(low = "khaki", high = "dark green") +
  theme(axis.text.x = element_text(angle = 90), legend.position = "right")
```



```
pca_df <- read.csv("C:/Users/rigel/Desktop/ВШЭ/финальный_проект/pca_df.csv")
pca_df %>%
  pivot_longer(-c(X), names_to = "classes", values_to = "counts") %>%
  ggplot(aes(x=classes, y=X, fill=counts)) +
  geom_raster() +
  scale_fill_gradient(low = "khaki", high = "dark green") +
  theme(axis.text.x = element_text(angle = 90), legend.position = "right")
```



```
fa_df <- read.csv("C:/Users/rigel/Desktop/ВШЭ/финальный_проект/fa_df.csv")
fa_df %>%
  pivot_longer(-c(X), names_to = "classes", values_to = "counts") %>%
  ggplot(aes(x=classes, y=X, fill=counts)) +
  geom_raster() +
  scale_fill_gradient(low = "khaki", high = "dark green") +
  theme(axis.text.x = element_text(angle = 90), legend.position = "right")
```

