

Information Age - Homework2

Load the packages

```
library(ggplot2)
library(tidyverse)
library(Metrics)
library(irr)
```

Load the data

```
df <- read_csv("C:/Users/Ines/OneDrive - FSV/CU1/Navigating the Information Age/HW2/jpm444_hw_data2.csv")
```

```
## Rows: 982 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (3): id, label_true, label_assigned
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
print(df)
```

```
## # A tibble: 982 x 3
##   id          label_true label_assigned
##   <chr>         <chr>         <chr>
## 1 id33333334 politics      economy
## 2 id34847143 human_interest economy
## 3 id35222578 economy      human_interest
## 4 id36325222 politics      human_interest
## 5 id3695654  economy      politics
## 6 id37828973 economy      economy
## 7 id38173592 politics      politics
## 8 id38849595 economy      economy
## 9 id39831316 economy      human_interest
## 10 id39865428 politics      human_interest
## # i 972 more rows
```

Create three separate dfs for each category -> In the label_true assign 0 if it has the value of that category and 1 if it is any of the other categories.

```
df_human_interest <- df
df_economy <- df
df_politics <- df
```

```
df_human_interest$label_true <- ifelse(df$label_true=="human_interest", 0, 1)
df_human_interest$label_assigned <- ifelse(df$label_assigned=="human_interest", 0, 1)
df_human_interest
```

```
## # A tibble: 982 x 3
##   id          label_true label_assigned
##   <chr>          <dbl>          <dbl>
## 1 id33333334         1            1
## 2 id34847143         0            1
## 3 id35222578         1            0
## 4 id36325222         1            0
## 5 id3695654          1            1
## 6 id37828973         1            1
## 7 id38173592         1            1
## 8 id38849595         1            1
## 9 id39831316         1            0
## 10 id39865428        1            0
## # i 972 more rows
```

```
df_economy$label_true <- ifelse(df$label_true=="economy", 0, 1)
df_economy$label_assigned <- ifelse(df$label_assigned=="economy", 0, 1)
df_economy
```

```
## # A tibble: 982 x 3
##   id          label_true label_assigned
##   <chr>          <dbl>          <dbl>
## 1 id33333334         1            0
## 2 id34847143         1            0
## 3 id35222578         0            1
## 4 id36325222         1            1
## 5 id3695654          0            1
## 6 id37828973         0            0
## 7 id38173592         1            1
## 8 id38849595         0            0
## 9 id39831316         0            1
## 10 id39865428        1            1
## # i 972 more rows
```

```
df_politics$label_true <- ifelse(df$label_true=="politics", 0, 1)
df_politics$label_assigned <- ifelse(df$label_assigned=="politics", 0, 1)
df_politics
```

```
## # A tibble: 982 x 3
##   id          label_true label_assigned
##   <chr>          <dbl>          <dbl>
## 1 id33333334         0            1
## 2 id34847143         1            1
## 3 id35222578         1            1
## 4 id36325222         0            1
## 5 id3695654          1            0
## 6 id37828973         1            1
## 7 id38173592         0            0
```

```
## 8 id38849595      1      1
## 9 id39831316      1      1
## 10 id39865428     0      1
## # i 972 more rows
```

Calculate precision

```
precision_human_interest <- precision(df_human_interest$label_true, df_human_interest$label_assigned)
precision_human_interest
```

```
## [1] 0.7856174
```

```
precision_economy <- precision(df_economy$label_true, df_economy$label_assigned)
precision_economy
```

```
## [1] 0.9375951
```

```
precision_politics <- precision(df_politics$label_true, df_politics$label_assigned)
precision_politics
```

```
## [1] 0.9192982
```

Calculate Recall

```
recall_human_interest <- recall(df_human_interest$label_true, df_human_interest$label_assigned)
recall_human_interest
```

```
## [1] 0.9633943
```

```
recall_economy <- recall(df_economy$label_true, df_economy$label_assigned)
recall_economy
```

```
## [1] 0.9361702
```

```
recall_politics <- recall(df_politics$label_true, df_politics$label_assigned)
recall_politics
```

```
## [1] 0.7432624
```

Calculate F1-score

```
f1_human_interest <- fbeta_score(df_human_interest$label_true, df_human_interest$label_assigned)
2*recall(df_human_interest$label_true, df_human_interest$label_assigned)*precision(df_human_interest$label_assigned)
```

```
## [1] 0.8654709
```

```
f1_human_interest
```

```
## [1] 0.8654709
```

```
f1_economy <- fbeta_score(df_economy$label_true,df_economy$label_assigned)
2*recall(df_economy$label_true,df_economy$label_assigned)*precision(df_economy$label_true,df_economy$label_assigned)
```

```
## [1] 0.9368821
```

```
f1_economy
```

```
## [1] 0.9368821
```

```
f1_politics <- fbeta_score(df_politics$label_true,df_politics$label_assigned)
2*recall(df_politics$label_true,df_politics$label_assigned)*precision(df_politics$label_true,df_politics$label_assigned)
```

```
## [1] 0.8219608
```

```
f1_politics
```

```
## [1] 0.8219608
```

Create table to hold all the values

```
final_table <- data.frame(
  Category = c("human_interest", "economy", "politics"),
  Precision = c(precision_human_interest, precision_economy, precision_politics),
  Recall = c(recall_human_interest, recall_economy, recall_politics),
  F1_Score = c(f1_human_interest, f1_economy, f1_politics)
)
```

```
final_table
```

```
##      Category Precision    Recall  F1_Score
## 1 human_interest 0.7856174 0.9633943 0.8654709
## 2      economy 0.9375951 0.9361702 0.9368821
## 3      politics 0.9192982 0.7432624 0.8219608
```