# Emotion detection in song lyrics

Simone Quadrelli (938667)

# Agenda

- Introduction
- Objectives
- Dataset
- Methodology
- Results
- Conclusions

# Introduction

Introduction Objectives Dataset Methodology Results Conclusions

# Introduction

01 Detect emotions and their intensity from song lyrics

02 Study the correlation between genre and sentiment

03 Compute personalized playlists

# Objectives

Introduction Objectives Dataset Methodology Results Conclusions

# Objectives

01 Detect emotions and their intensity in song lyrics exploiting predictors trained with labelled tweets

02 Analyze the correlation between genre and sentiment

03 Compute personalized playlists given genre and sentiment

# Datasets

Introduction Objectives Dataset Methodology Results Conclusions

# Dataset: tweets

01 Tweet annotated dataset

02 Emotions: anger, joy, fear, sadness

03 Intensity $\in [0,1]$

# Dataset: Song lyrics

01 Song lyrics

02 Title

03 Genre

04 Artist

# Methodology

Introduction

Objectives

Dataset

Methodology

Results

Conclusions

# Methodology

# Preprocessing and feature extraction

01 Punctuation removal

02 Stop words removal

03 Lemmatization

04 Part of speech (POS) selection of adjectives and nouns for raw text and lemmatized text

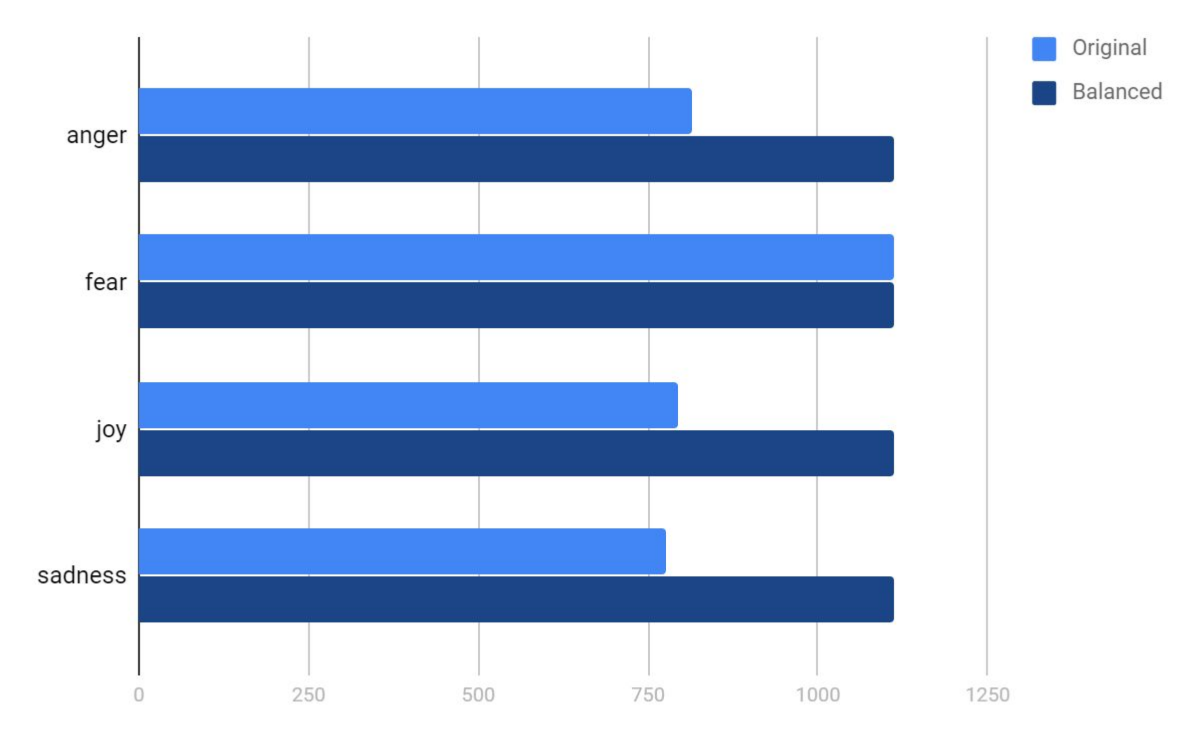
# Independence

Spearman's correlation test between intensity and emotions

| | sentiment | intensity |
|---|---|---|
| sentiment | 1.000000 | -0.012154 |
| intensity | -0.012154 | 1.000000 |

Since sentiment and intensity are uncorrelated it is possible to train two independent predictors

# Unbalancedness



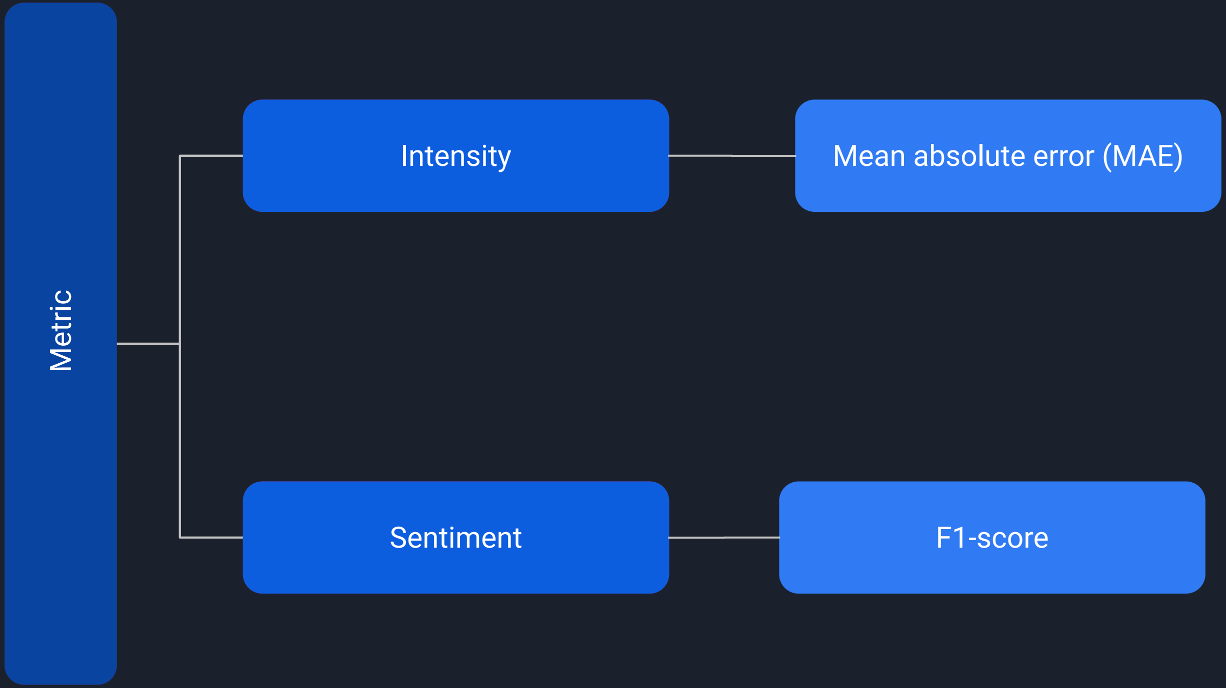
# Weighting

01 Term frequency - inverse document frequency (tf-idf)

02 Assign high weight to very rares word in the corpus which are repeated in a document

03 Assign low weight to very common words

# Metrics


Metric
Intensity
Mean absolute error (MAE)
Sentiment
F1-score

# Predictors: sentiment and intensity

01 K-nearest neighbours

02 Random forests

03 Support vector machines

# Predictors: sentiment

| Predictor | Features | F1-score |
|---|---|---|
| KNN | POS lemma | 0.66 |
| KNN | POS raw | 0.61 |
| RF | POS lemma | 0.70 |
| RF | POS raw | 0.70 |
| SVM | POS lemma | 0.87 |
| SVM | POS raw | 0.89 |

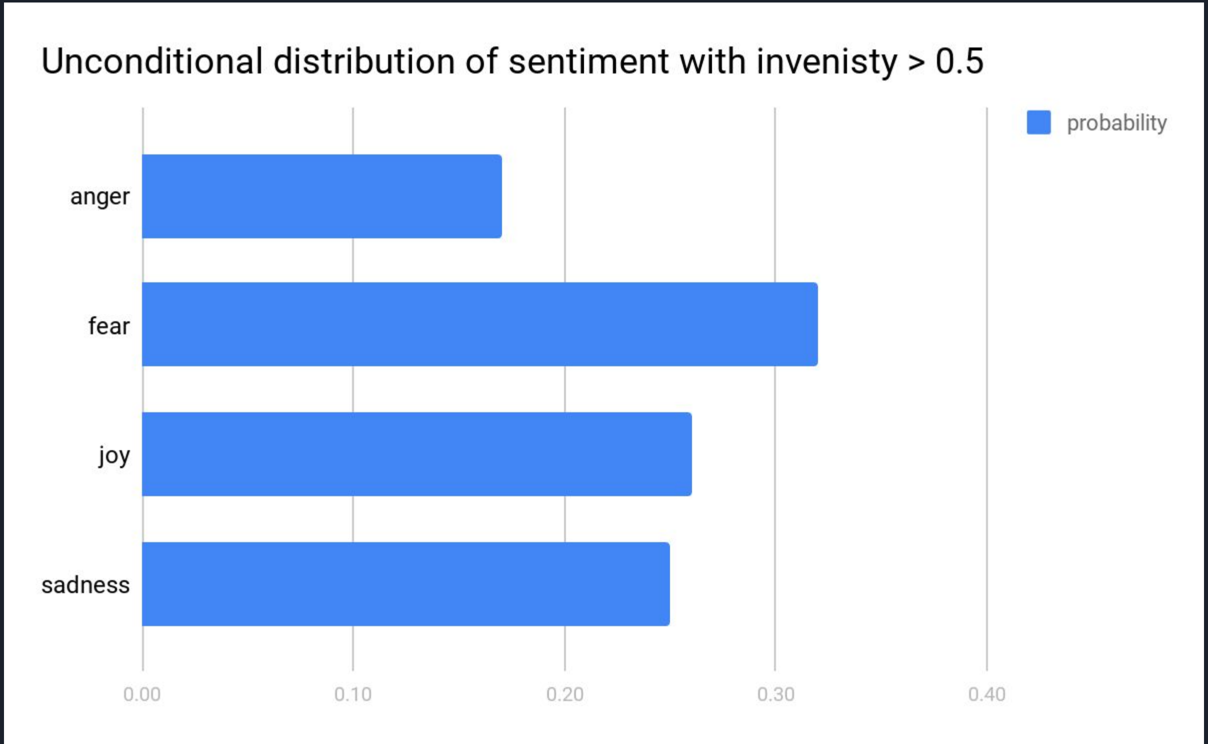
Save the best classifier

# Predictors: intensity

| Predictor | Features | MAE |
|---|---|---|
| KNN | POS lemma | 0.14 |
| KNN | POS raw | 0.14 |
| RF | POS lemma | 0.15 |
| RF | POS raw | 0.15 |
| SVM | POS lemma | 0.11 |
| SVM | POS raw | 0.11 |

Save the best predictor

# Predictions



# $\chi$-square test

Test for independence of genre and sentiment

```
Power_divergenceResult(statistic=59236142.0970242, pvalue=0.0)
Power_divergenceResult(statistic=11203428.310288608, pvalue=0.0)
Power_divergenceResult(statistic=869650.506747599, pvalue=0.0)
Power_divergenceResult(statistic=221392022.56083325, pvalue=0.0)
Power_divergenceResult(statistic=2198070.1291956482, pvalue=0.0)
Power_divergenceResult(statistic=12938333.560142726, pvalue=0.0)
Power_divergenceResult(statistic=159398734.2544964, pvalue=0.0)
Power_divergenceResult(statistic=84988573.01230262, pvalue=0.0)
Power_divergenceResult(statistic=4968962.380629653, pvalue=0.0)
Power_divergenceResult(statistic=310096282.6461709, pvalue=0.0)
Power_divergenceResult(statistic=2732474.551130772, pvalue=0.0)
Power_divergenceResult(statistic=2563877910.467037, pvalue=0.0)
```

Genre and sentiments are correlated

# Results

Introduction

Objectives

Dataset

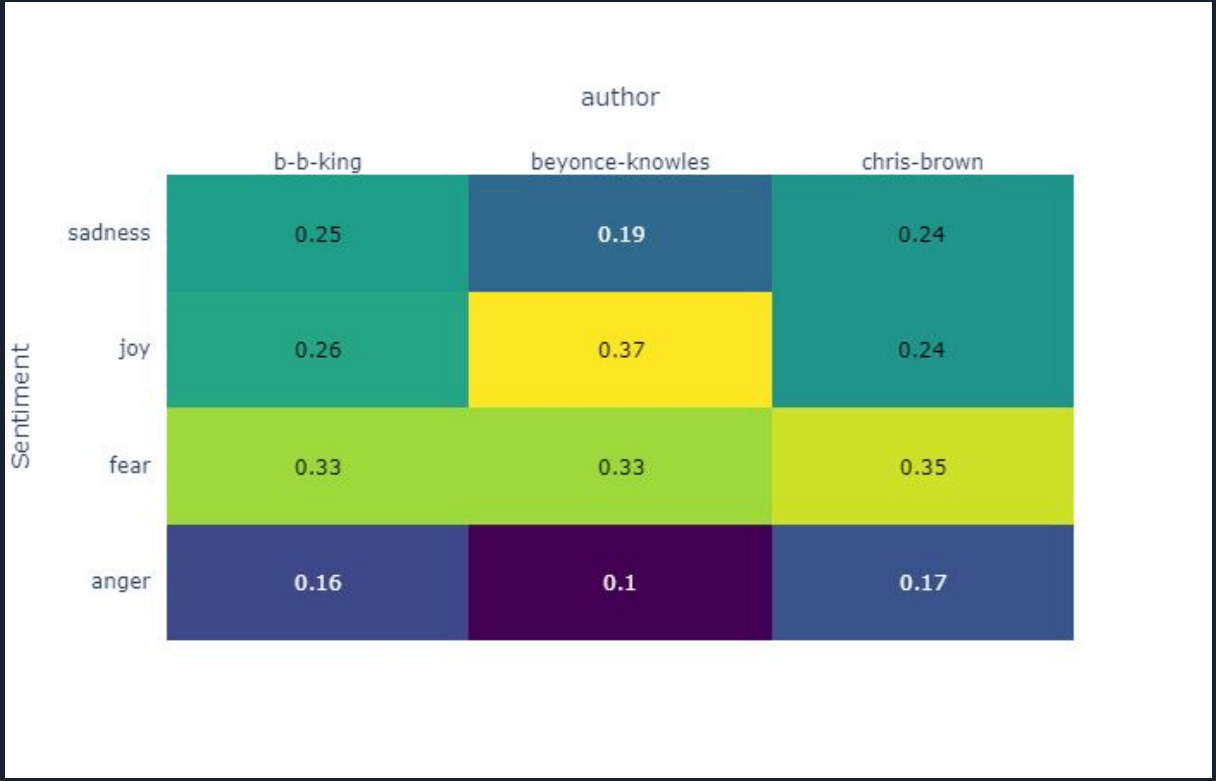
Methodology

Results

Conclusions

# Genre and sentiment relationships

# Author and sentiment relationships



# Playlist computation

```
Enter a gerne: Jazz
Enter a mood: joy
september song
body and soul
i am loved
body soul
sunny
i didn t know about you
it s a whistling kinda morning
what game shall we play today
```

```
Enter a gerne: Metal
Enter a mood: fear
nervous heart
unleashed upon mankind
from beyond the grave
seeds of mans destruction
intro unleashed upon mankind
stigmata
spellbound by the devil
```

# Conclusions

Introduction | Objectives | Dataset | Methodology | Results | Conclusions

# Conclusions

01 Genre and sentiments are correlated

02 Authors may have a sentiment distribution very far from that of their genre

03 Song provided by the playlist seem pertinent

# Issues

01 The high frequence of fear may depend on the bias of the umbalancedness of the original twitts dataset

02 Melody may convey sentiment

# Future work

01 Compare the results with a dataset of annotated songs

02 Use melody to construct a predictor

# Thanks for your attention

Q&A