Silesian University of Technology

CELEBRITIES PICTURES CLASSIFICATION

Magdalena Ługowska, Agata Muszyńska
WYDZIAŁ AUTOMATYKI, ELEKTRONIKI I INFORMATYKI

European Funds
Knowledge Education Development

AIDA
APPLIED INTEGRATIVE DATA ANALYSIS

European Union
European Social Fund

The data set

- Source : kaggle.com
- 202 599 face images of various celebrities
- 10,177 unique identities
- 40 binary attribute anotations per image

Project outline

1. Data preprocessing
   - Check for missing values in attributes data set
   - Check for correlations
   - Perform feature selection

2.Perform simple classification for both attributes and images

3.Build a model for celebrities face recognition using neural networks
- See if there are any biases in good performance (it is easier to classify woman)
- Check if it is possible to correctly classify childhood pictures

Project outline

1. Data preprocessing
   - Check for missing values in attributes data set
   - Check for correlations
   - Perform feature selection

2.Perform simple classification for both attributes and images

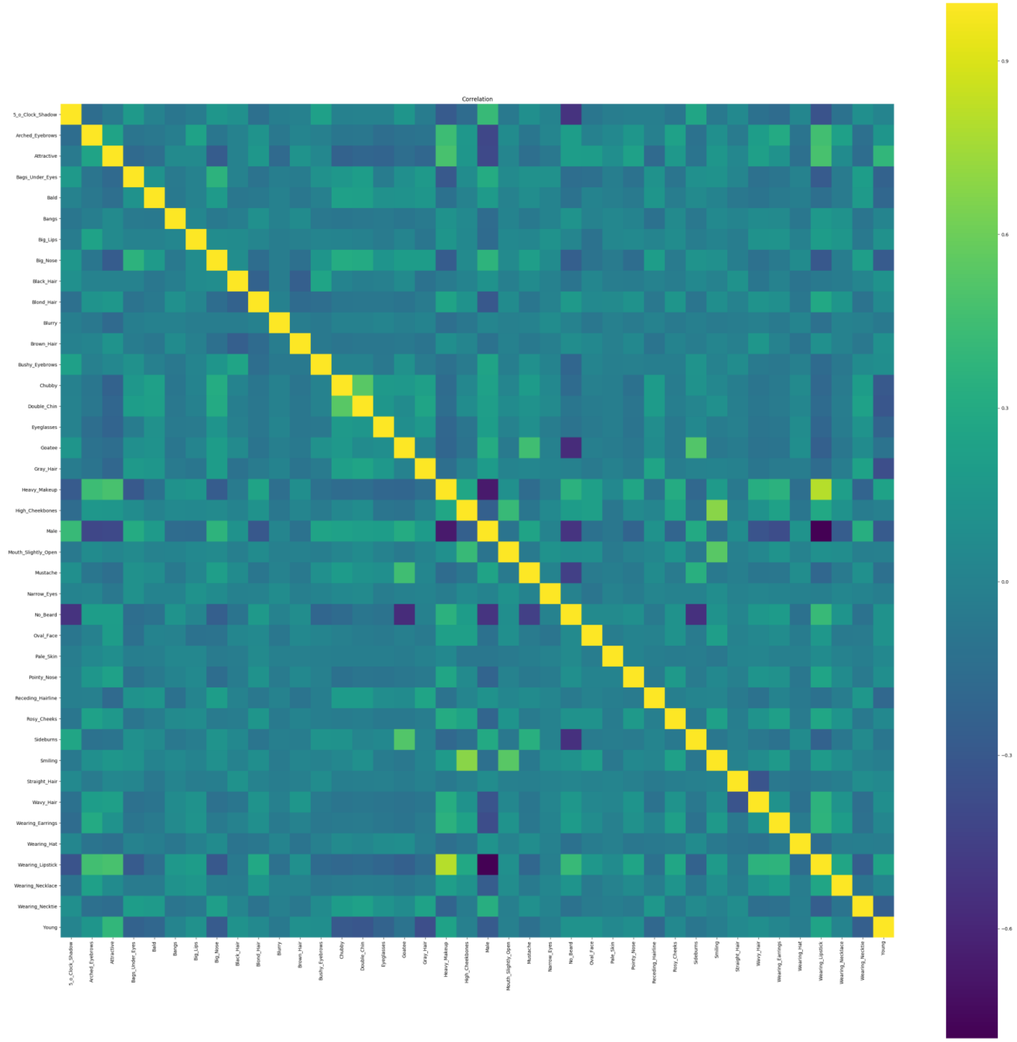
DONE

Also:
- Add additional metrics
- Use different method of feature selection

3.Build a model for celebrities face recognition using neural networks
- See if there are any biases in good performance (it is easier to classify woman)
- Check if it is possible to correctly classify childhood pictures

TO DO

Correlation plot



- Goatee is negatively correlated with no beard
- As expected Male is negatively correlated with features such as wearing neklace, heavy makeup
- Smiling is positively correlated with high cheekbones.

Feature extraction

- Images:
  - Extracting the mean and standard deviation for each color channel in the image. 6 features in total.

- Atrributes:
  - Feature selection using PCA.
  - Features, which explained less than 0.001 variance were removed. 26 features in total.

Accuracy results obtained from predicting gender based on 10 000 images or attributes

	Knn	Naive Bayes	Logistic Regression	SVM	Decision tree	Rando forest	MLP
Images	0.572	0.586	0.577	0.565	0.520	0.536	0.582
Attributes	0.910	0.864	0.933	0.927	0.920	0.933	0.936

Thank you for your attention!