

pipeline__dataset__v3

August 24, 2025

1 PIPELINE WITH MODEL 4, DATASET V3 (https://universe.roboflow.com/sidsproject/back_stomach_123_merged-xxmxb/dataset/1), DIFFERENT TYPES OF EMBEDDINGS

1.1 PIPELINE

1.1.1 1. load model face__detection

1.1.2 2. create embeddings

1.1.3 3. retrieval to evaluate embeddings goodness

1.1.4 4. train model classification

1.2 1. load the model face__detection

Step 1 is the same for each pipeline, so we do it once at the beginning.

Import dependencies

```
[37]: %load_ext autoreload
      %autoreload 2

      from embeddings import EmbeddingBuilder
      import os
      import ipynbname

      from retrieval import ImageRetrieval
      from classifier import Classifier
      import file_manager

      project_dir = f"{os.getcwd()}.
      ↪split('SIDS_revelation_project')[0]}SIDS_revelation_project/"
      image_dataset_path = f"{project_dir}datasets/onback_onstomach_v3"
      model_path = f"{project_dir}/models/4.weights/best.pt"
```

The autoreload extension is already loaded. To reload it, use:

```
%reload_ext autoreload
```

Choose how to process the dataset: 1. “extract_features”: extracting features and labels
2. “extract_features_imageswithinference”: extracting features and labels and saving predicted

images with bboxes 3. “load”: loading features and labels

```
[38]: emb_builder = EmbeddingBuilder(model_path, image_dataset_path, "load")
```

```
Extracting dataset info from .coco.json
file:-----
Dataset contains 4158 valid samples, and labels are {'baby_on_back': 1,
'baby_on_stomach': 2}
-----

Loading features from
.csv-----
Features loaded succesfully, in particular there are 4158 files in the dataset
-----

Embedding builder initialized
successfully-----
Face detection model: 4 (YOLOv8)
Dataset: /home/terra/Documents/AI_engineering/SIDS-
project/python_project/SIDS_revelation_project/datasets/onback_onstomach_v3
Dataset dimension: 4158
Dataset labels: {'baby_safe': 0, 'baby_unsafe': 1}
-----
```

1.3 2. Extract embeddings from dataset

Create embeddings

```
[39]: embeddings = emb_builder.embedding_flags()
```

```
Creation of flags features
embedding-----
4158 embedding created
-----
```

```
[40]: embeddings.head()
```

```
[40]:   flag_eye1  flag_eye2  flag_nose  flag_mouth
0         0         0         0         0
1         0         0         0         0
2         0         0         0         0
3         1         0         1         0
```

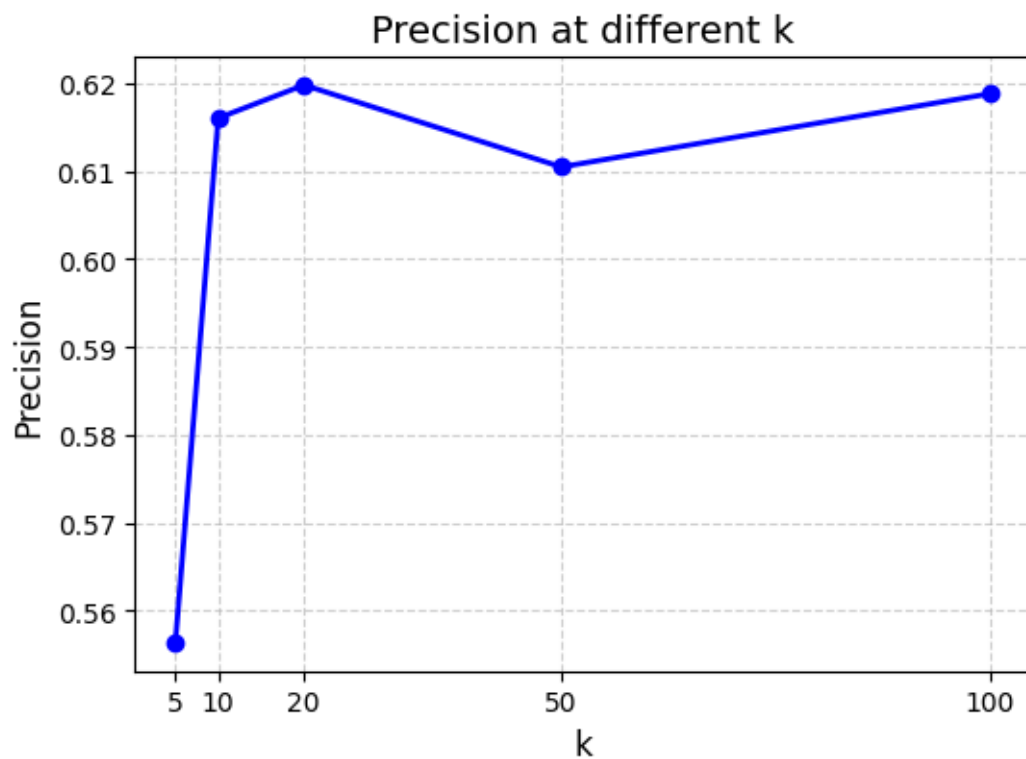
4 1 1 1 1

1.4 3. Retrieval to evaluate embedding goodness

```
[41]: ret = ImageRetrieval(embeddings, emb_builder.y, emb_builder.image_paths,
    ↪ image_dataset_path, emb_builder.classes_bs)
    ret.report("euclidean")
```

Precision at different

k:-----



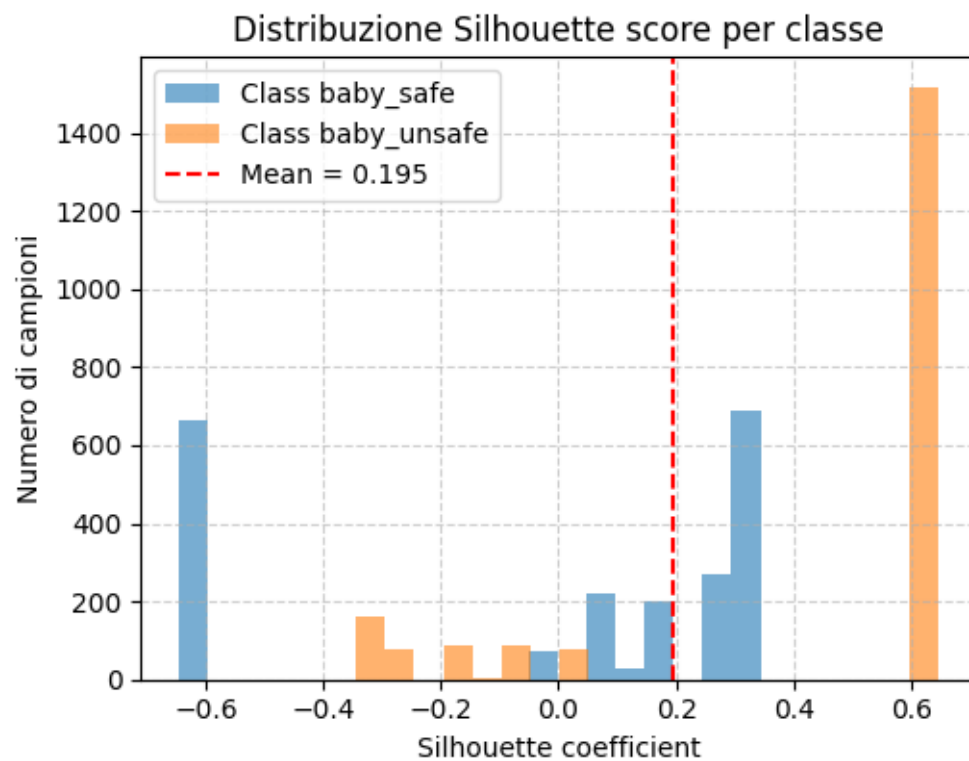
Recall at

R-----

0.5850676155785931

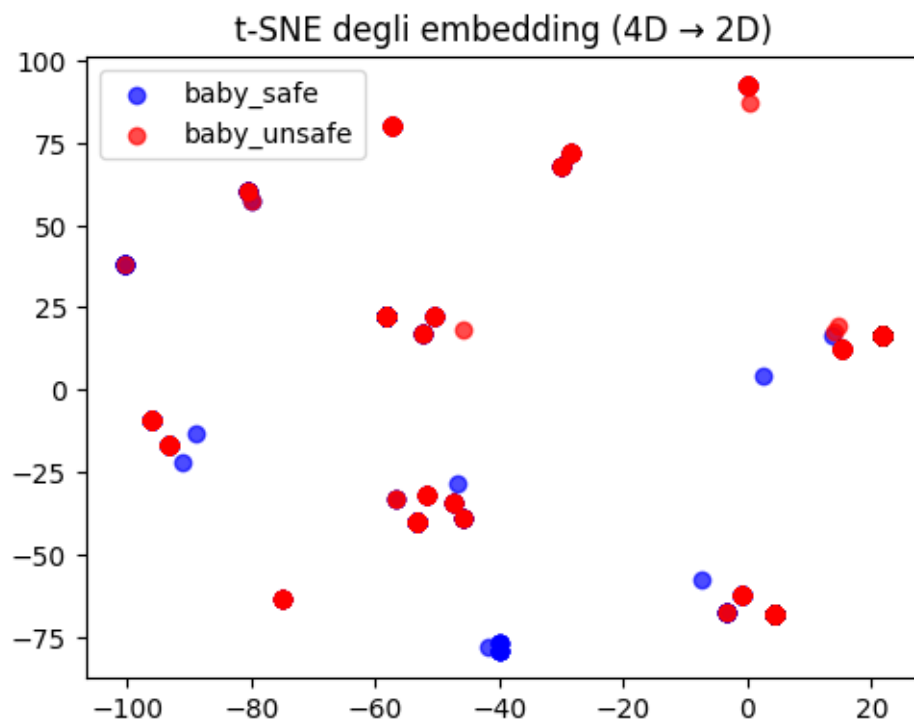
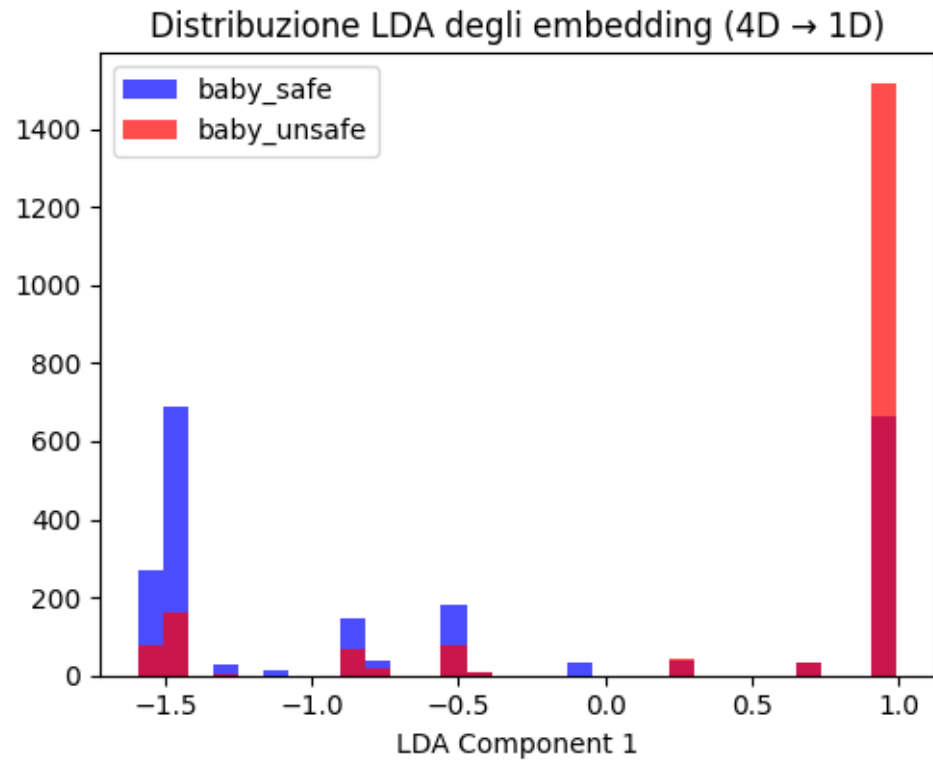
Silhouette

score-----

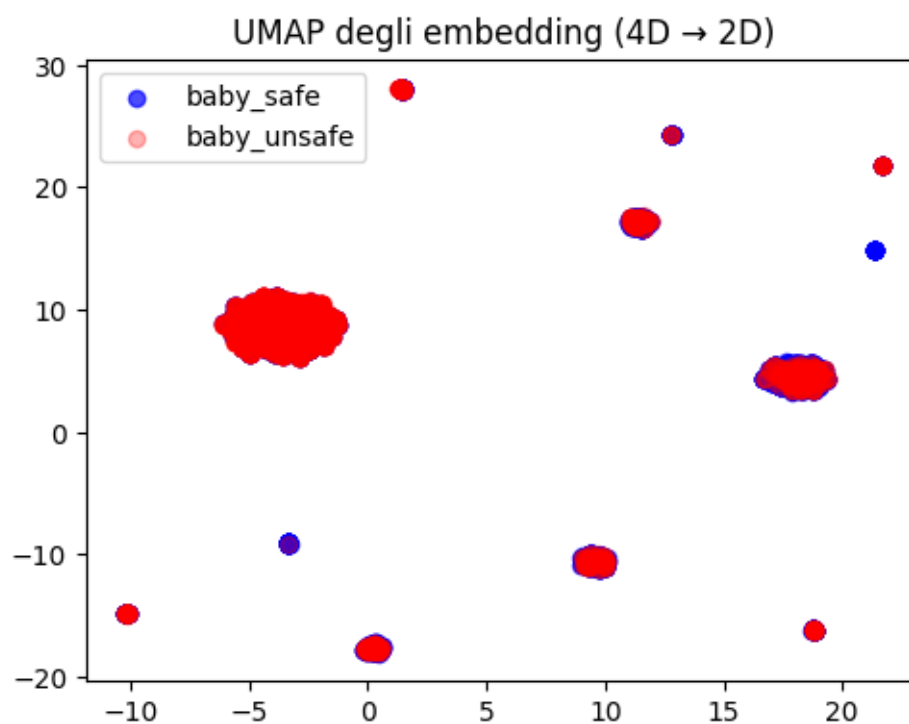


Embeddings

distributions-----



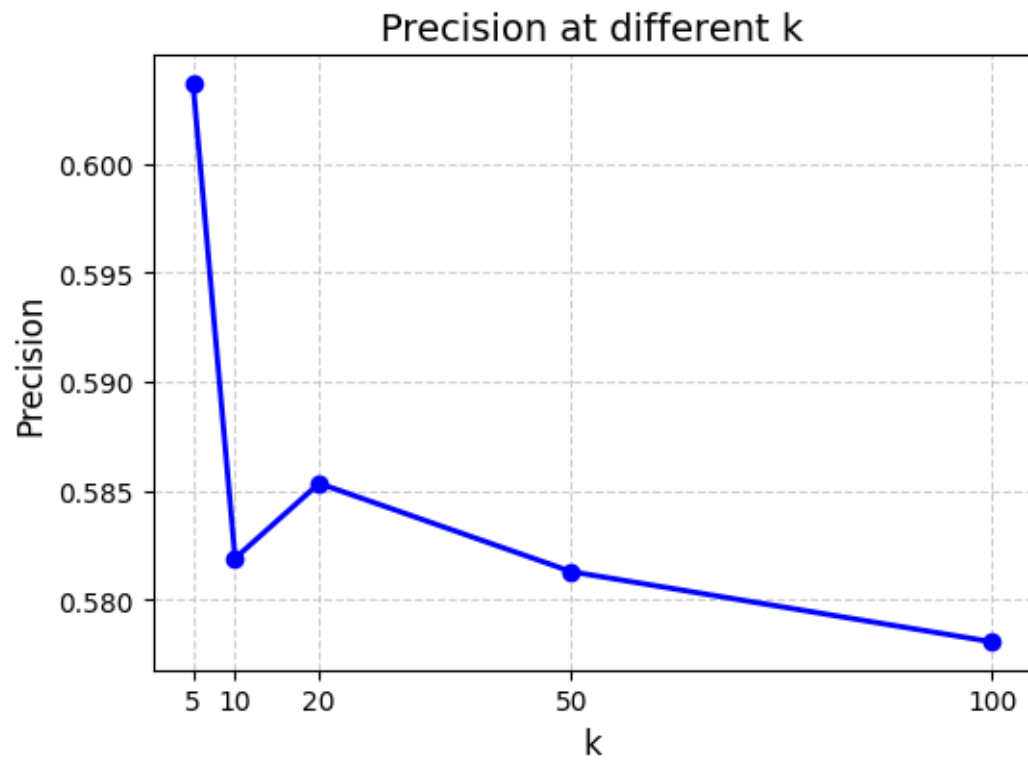
```
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-  
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by  
setting random_state. Use no seed for parallelism.  
warn(
```



```
[42]: ret.report("cosine")
```

Precision at different

k:-----

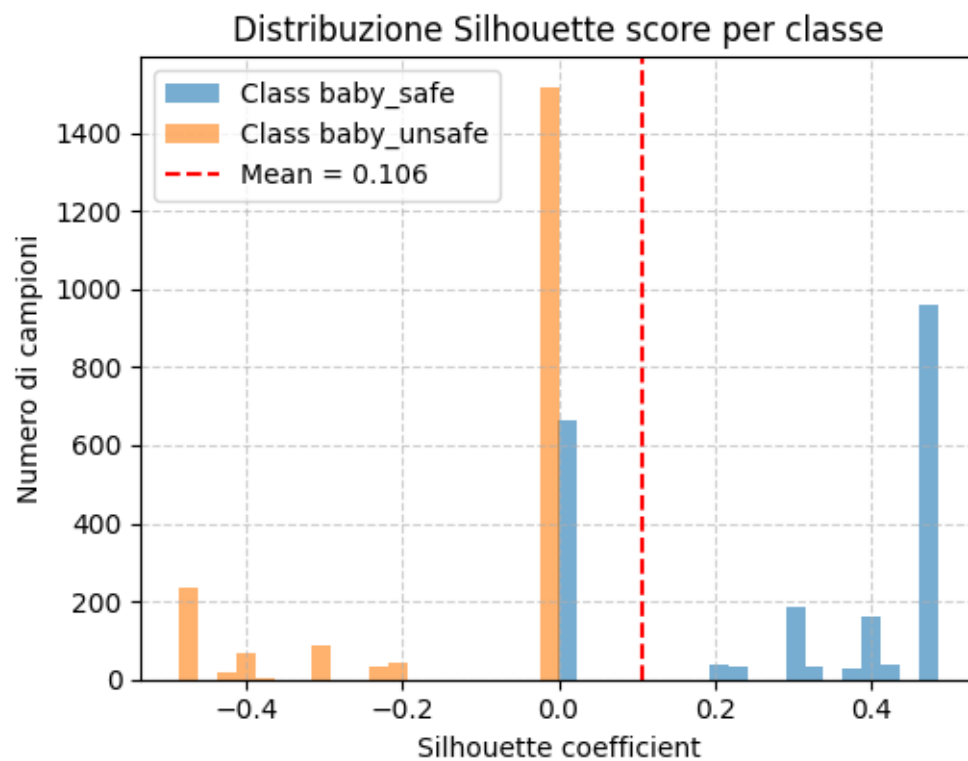


Recall at

R-----
0.5428515074643188

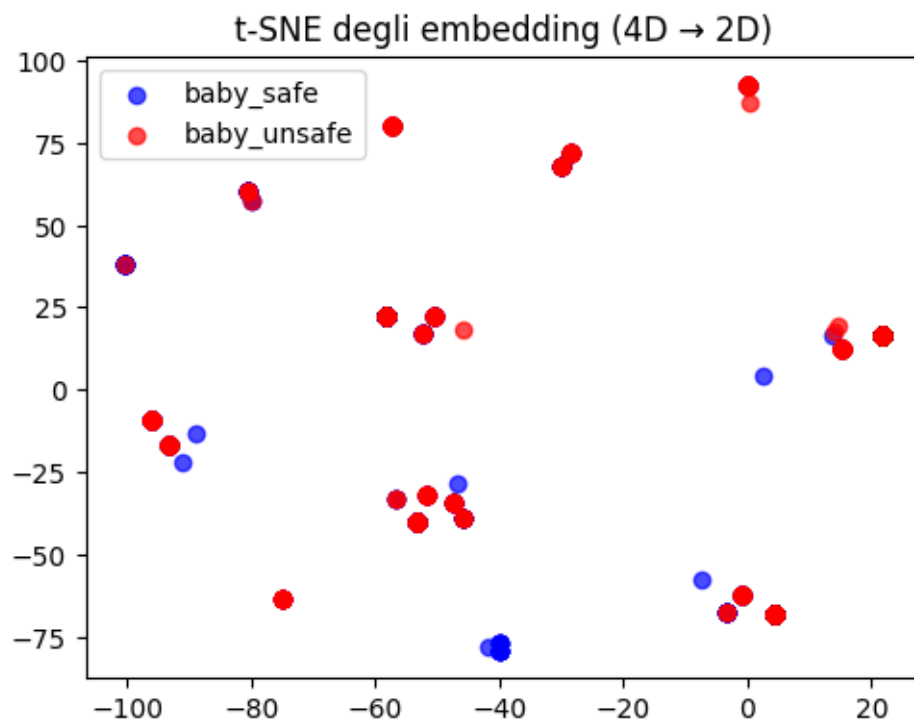
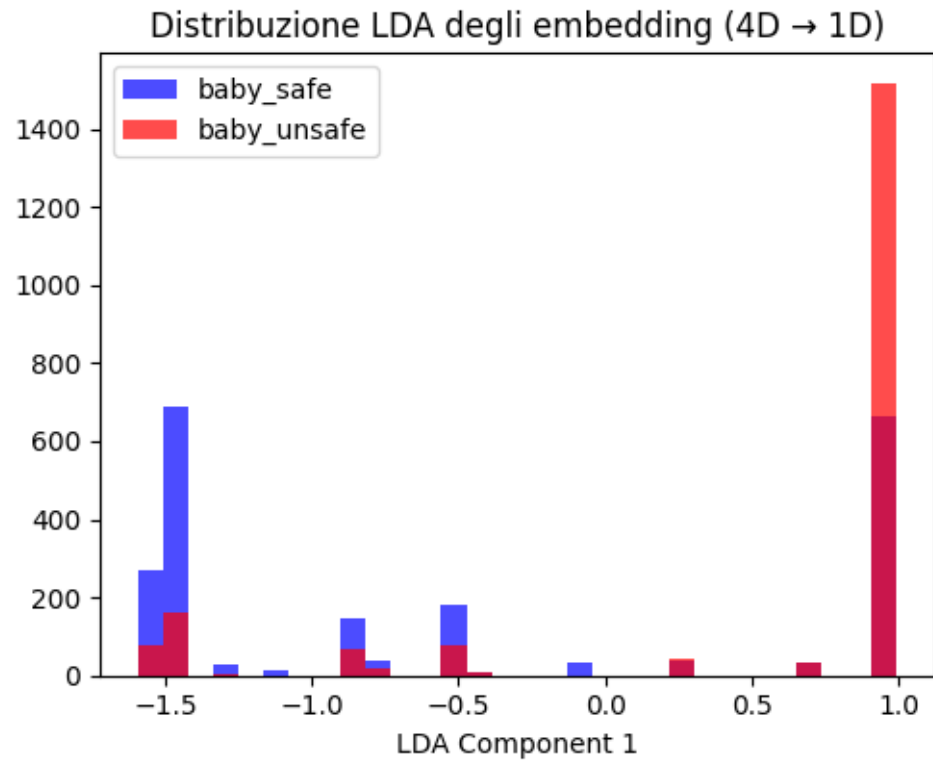
Silhouette

score-----



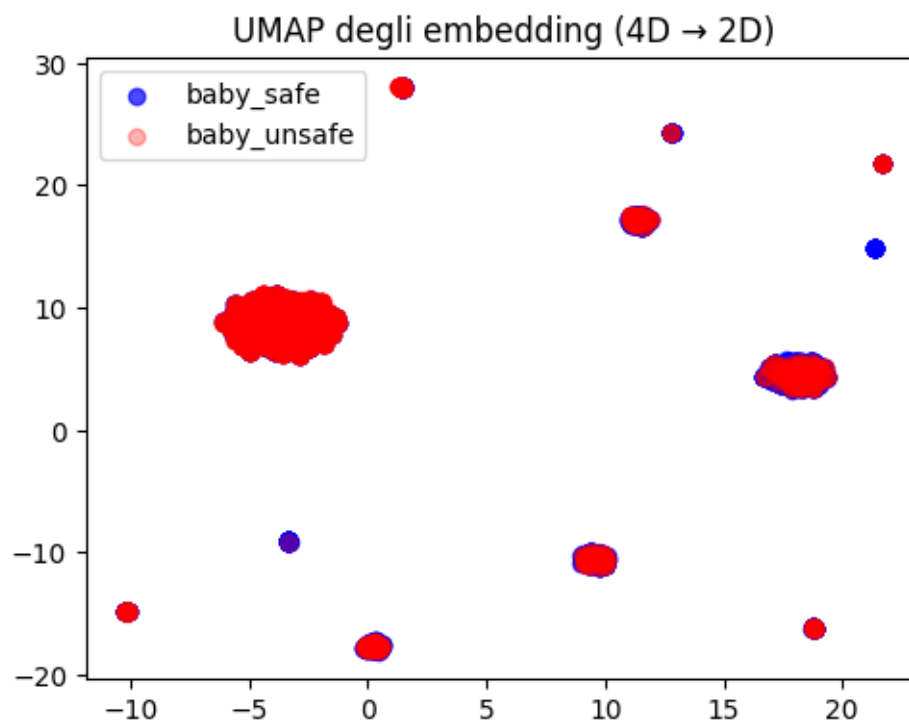
Embeddings

distributions-----



```
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-  
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by  
setting random_state. Use no seed for parallelism.  
warn(  

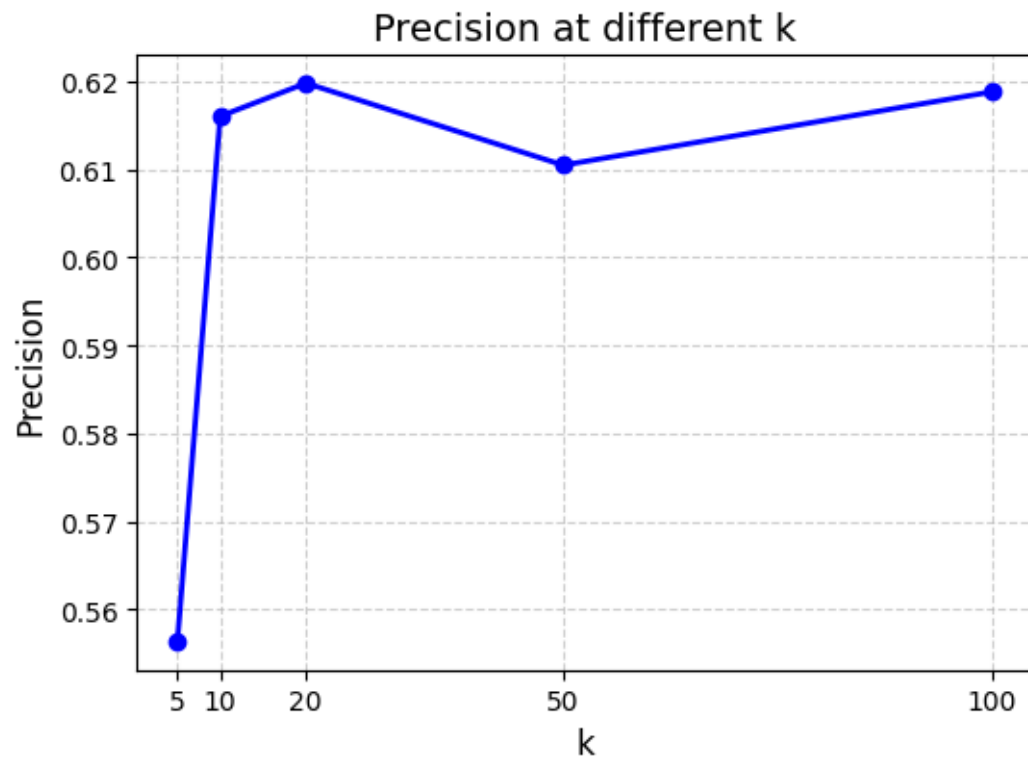
```



```
[43]: ret.report("minkowski")
```

Precision at different

k:-----

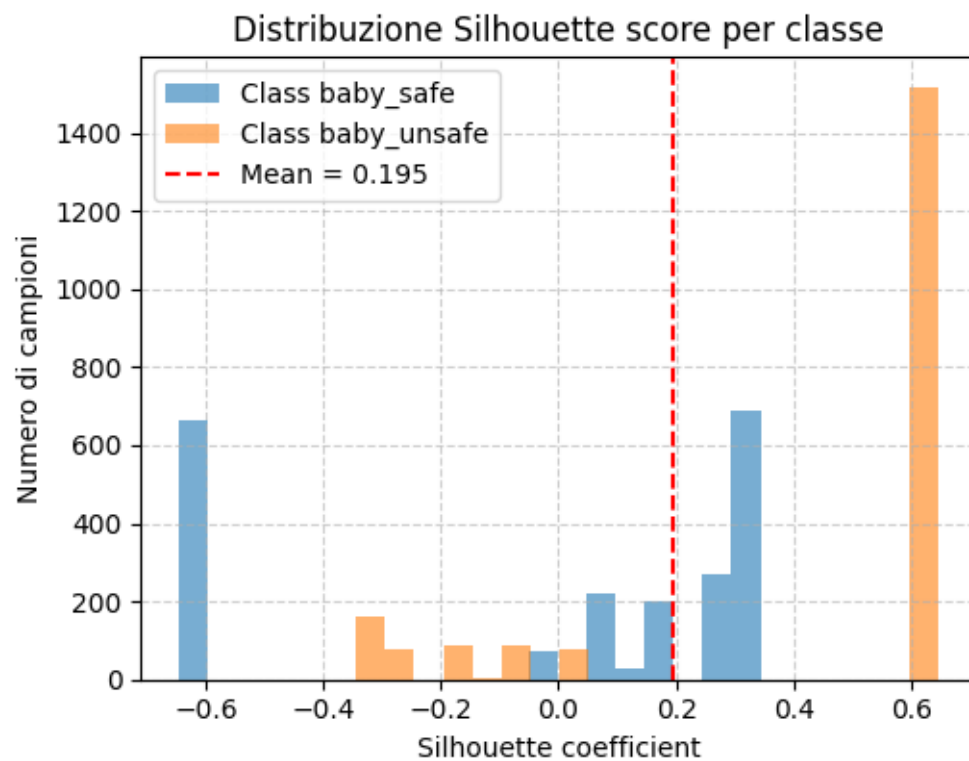


Recall at

R-----
0.5850676155785931

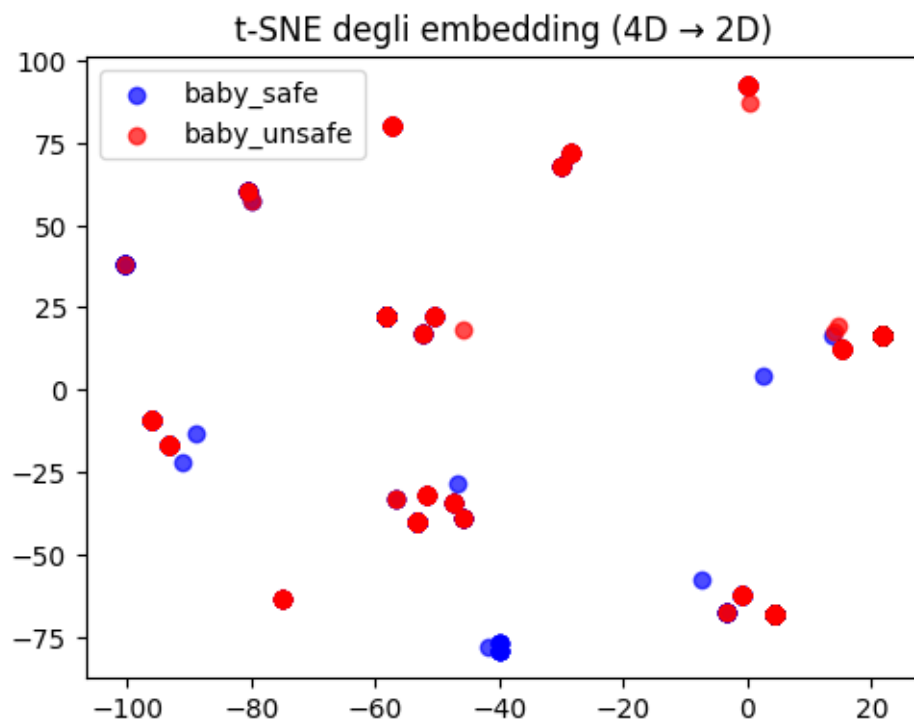
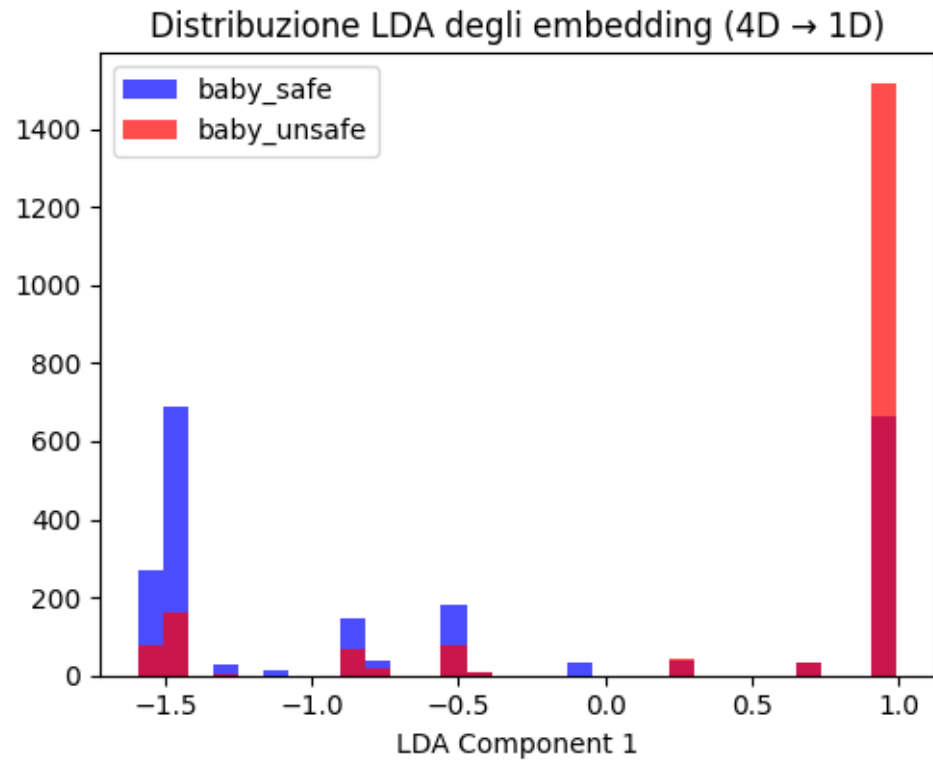
Silhouette

score-----



Embeddings

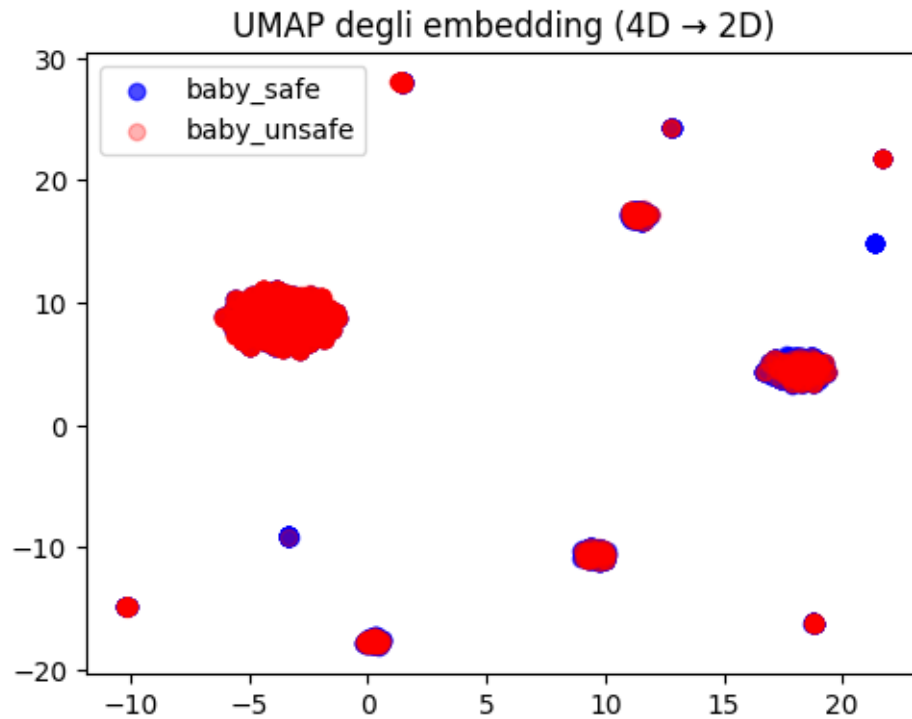
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
warn(

```



1.5 4. Train model classification

```

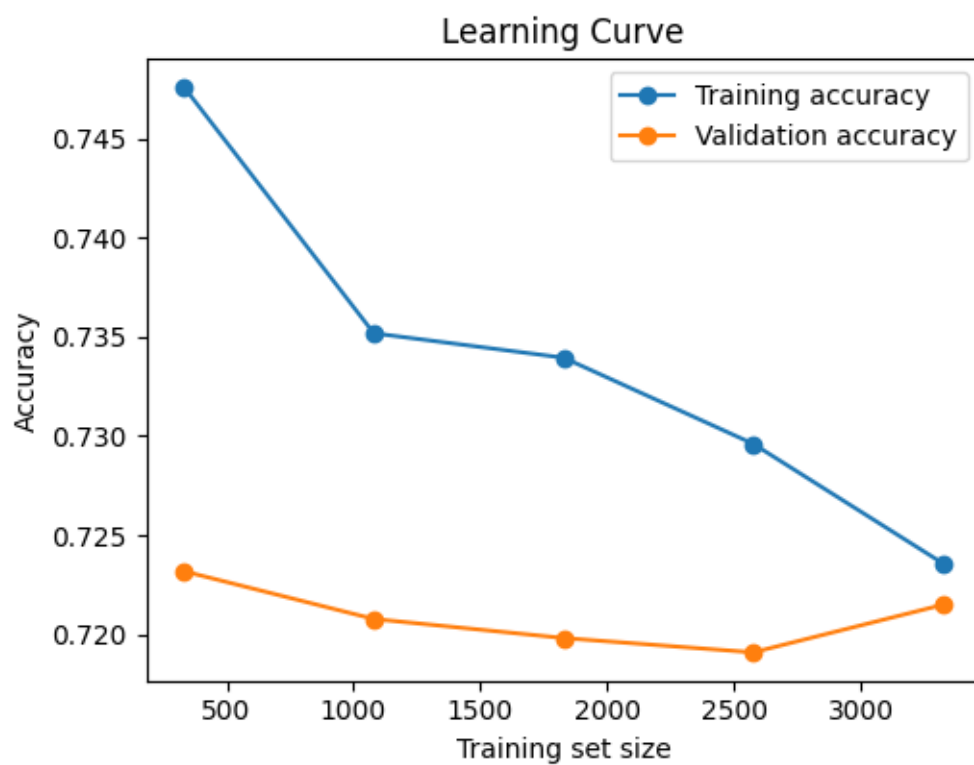
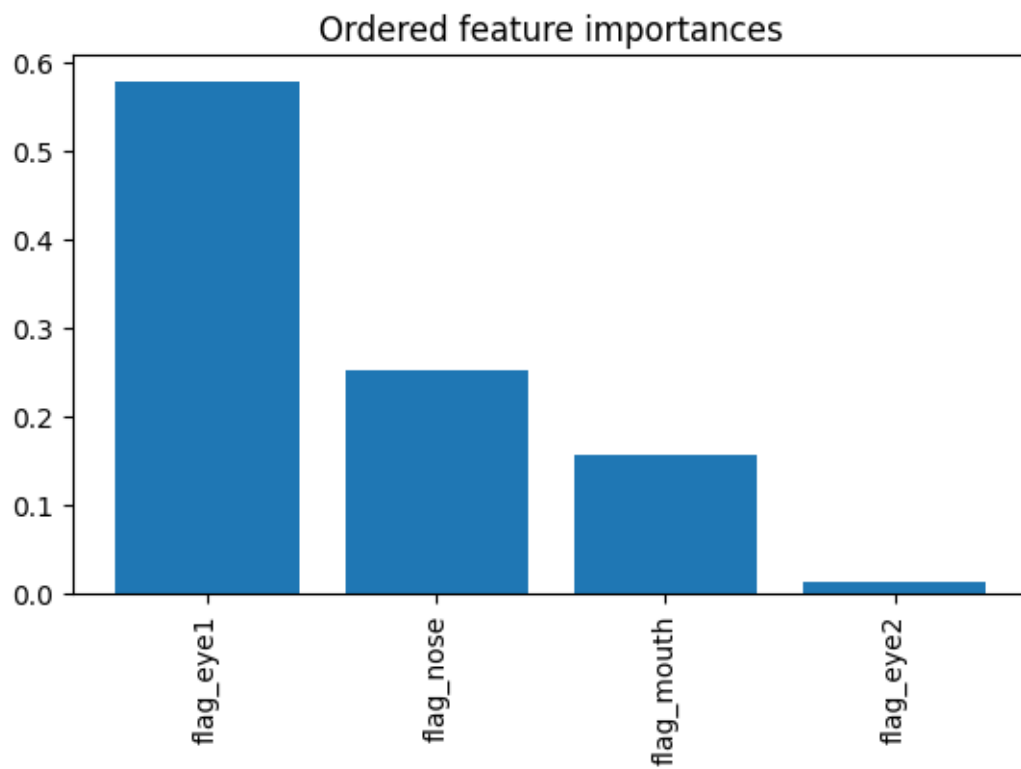
[44]: clf = Classifier(embeddings, emb_builder.y, emb_builder.classes_bs)
      clf.random_forest()

```

```

-----
-----
-----FIRST
ANALYSIS-----
-----
-----

```

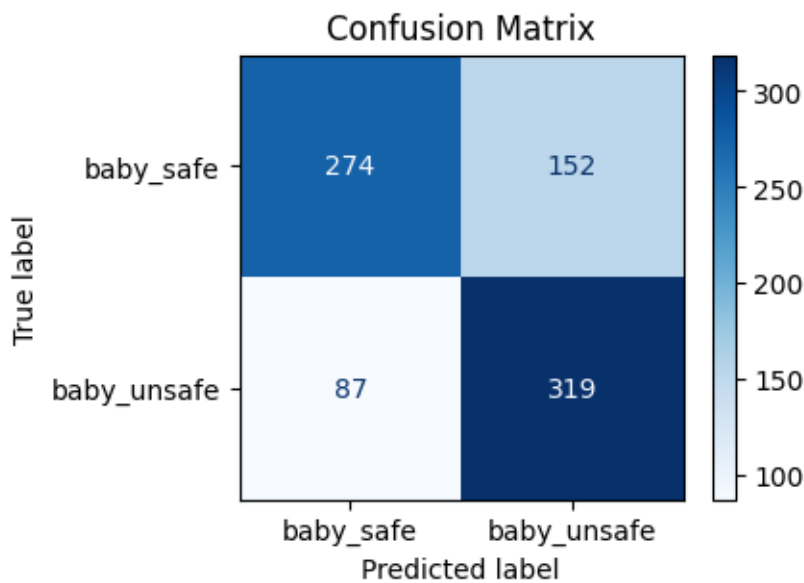


Dataset labels:-----
{'baby_safe': 0, 'baby_unsafe': 1}

Report-----

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| baby_safe | 0.76 | 0.64 | 0.70 | 426 |
| baby_unsafe | 0.68 | 0.79 | 0.73 | 406 |
| accuracy | | | 0.71 | 832 |
| macro avg | 0.72 | 0.71 | 0.71 | 832 |
| weighted avg | 0.72 | 0.71 | 0.71 | 832 |

Confusion matrix-----



1.6 2. Extract embeddings from dataset

Create embeddings

```
[45]: embeddings = emb_builder.embedding_all_features()
```

Creation of all features

embedding-----
4158 embedding created


```
-----
-----
```

```
[46]: embeddings.head()
```

```
[46]:   flag_eye1  flag_eye2  flag_nose  flag_mouth   x_eye1   y_eye1   x_eye2  \
0         0         0         0         0 -1.000000 -1.000000 -1.000000
1         0         0         0         0 -1.000000 -1.000000 -1.000000
2         0         0         0         0 -1.000000 -1.000000 -1.000000
3         1         0         1         0  0.199751  0.875817 -1.000000
4         1         1         1         1  0.747866  0.955937  0.746319

      y_eye2   x_nose   y_nose   x_mouth  y_mouth  eye_distance  \
0 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000      -1.000000
1 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000      -1.000000
2 -1.000000 -1.000000 -1.000000 -1.000000 -1.000000      -1.000000
3 -1.000000  0.246237  0.891980 -1.000000 -1.000000      -1.000000
4  0.680452  0.709361  0.853581  0.650598  0.84086      0.27549

      face_vertical_length  face_angle_vertical  face_angle_horizontal  \
0                -1.000000                -1.000000                -1.000000
1                -1.000000                -1.000000                -1.000000
2                -1.000000                -1.000000                -1.000000
3                -1.000000                -1.000000                -1.000000
4                 0.060125                122.830696                147.334481

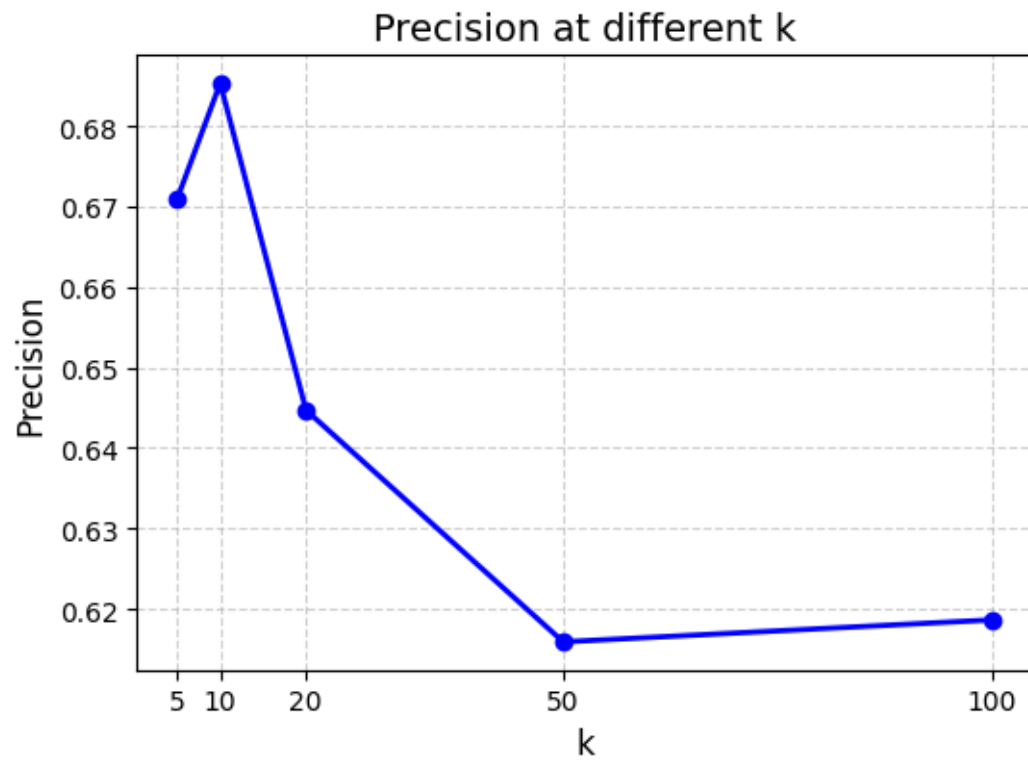
      symmetry_diff
0         0.000000
1         0.000000
2         0.000000
3         0.000000
4         0.085138
```

1.7 3. Retrieval to evaluate embedding goodness

```
[47]: ret = ImageRetrieval(embeddings, emb_builder.y, emb_builder.image_paths,
    ↪ image_dataset_path, emb_builder.classes_bs)
ret.report("euclidean")
```

Precision at different

k:-----

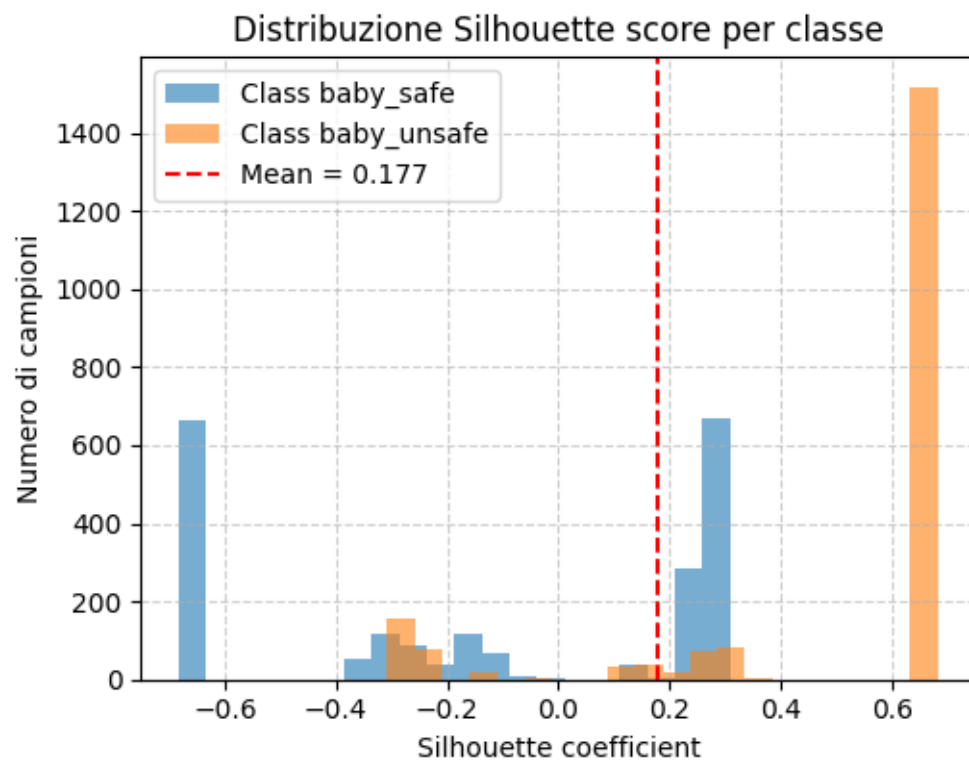


Recall at

R-----
0.5546667579984337

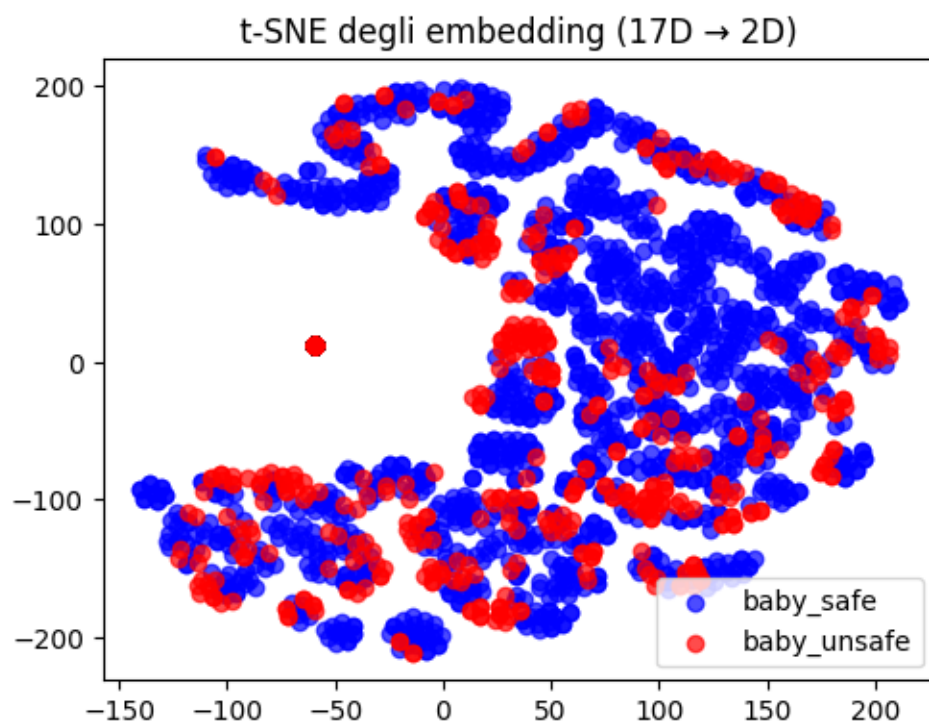
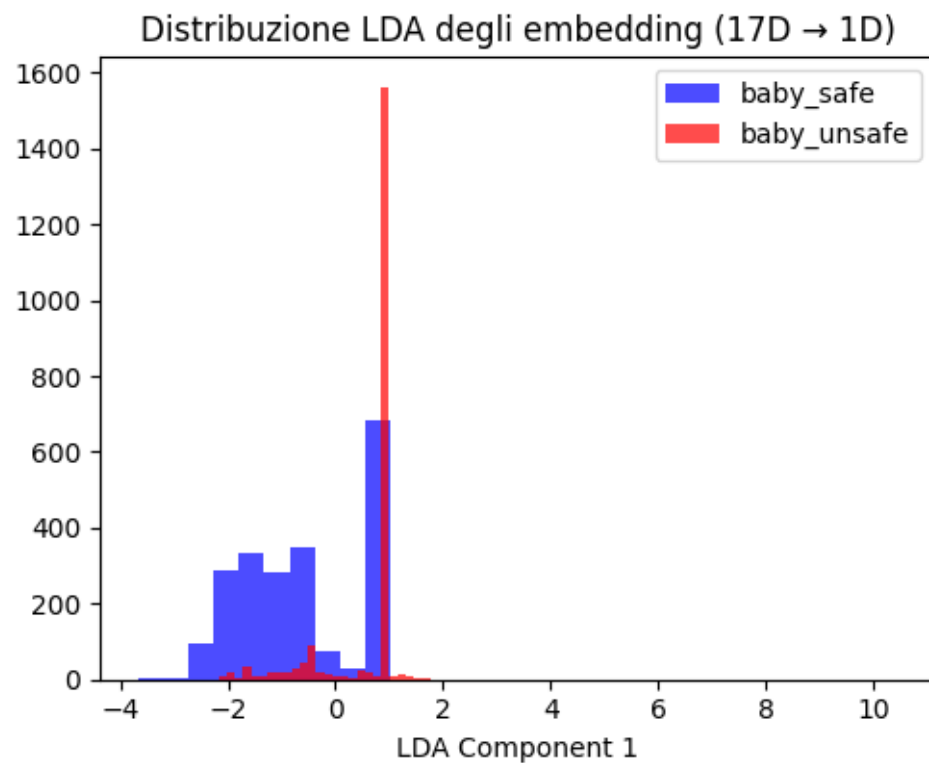
Silhouette

score-----



Embeddings

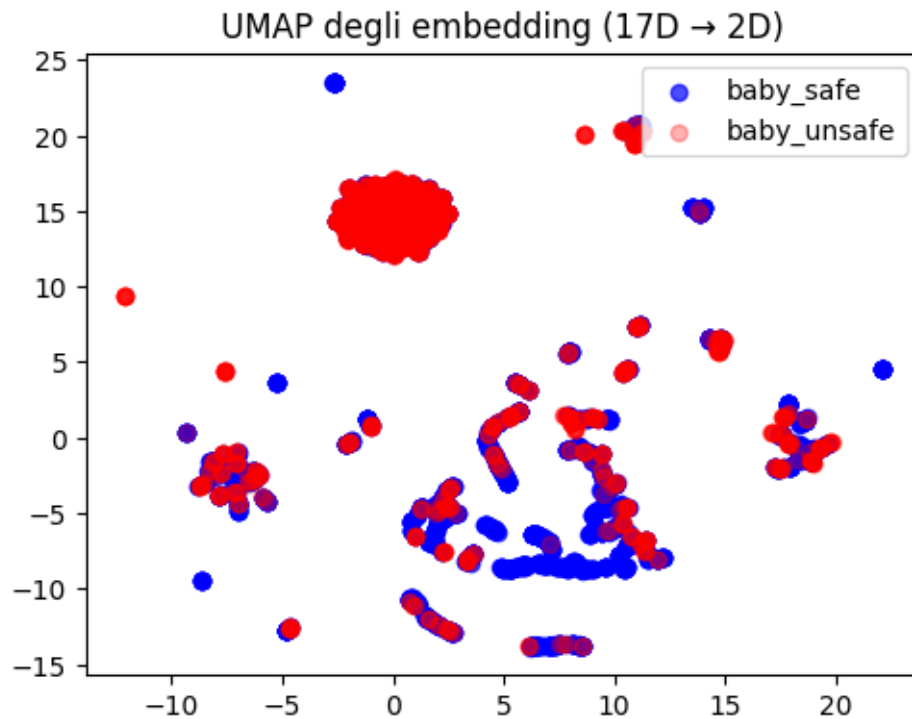
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
    warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
    warnings.warn(

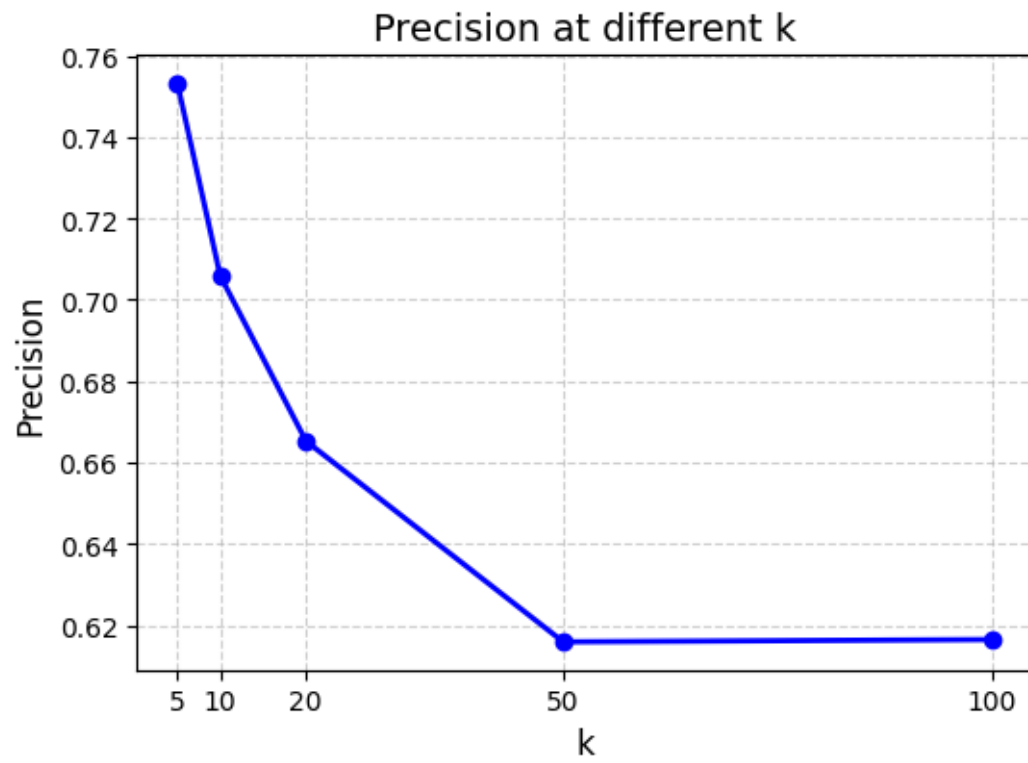
```



```
[48]: ret.report("cosine")
```

Precision at different

k:-----

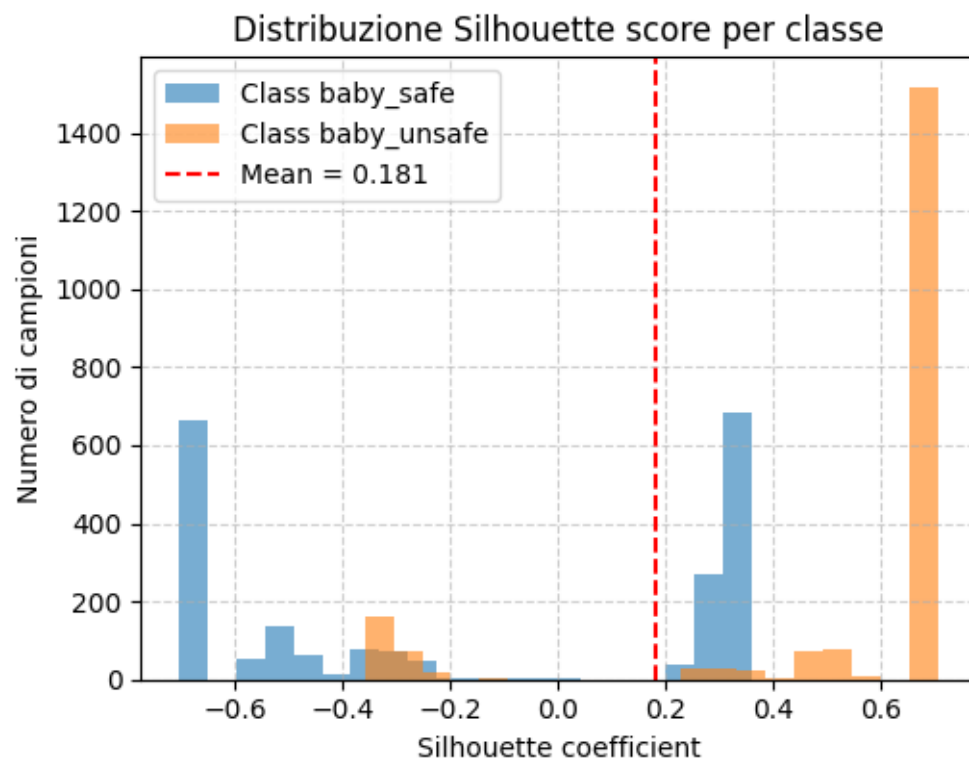


Recall at

R-----
0.5537950154830042

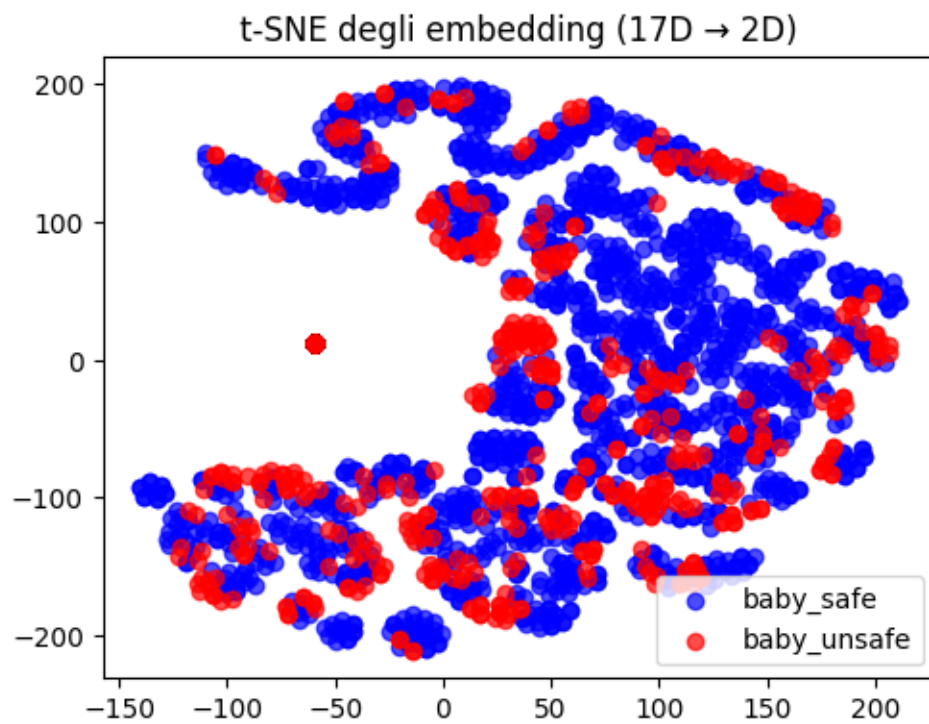
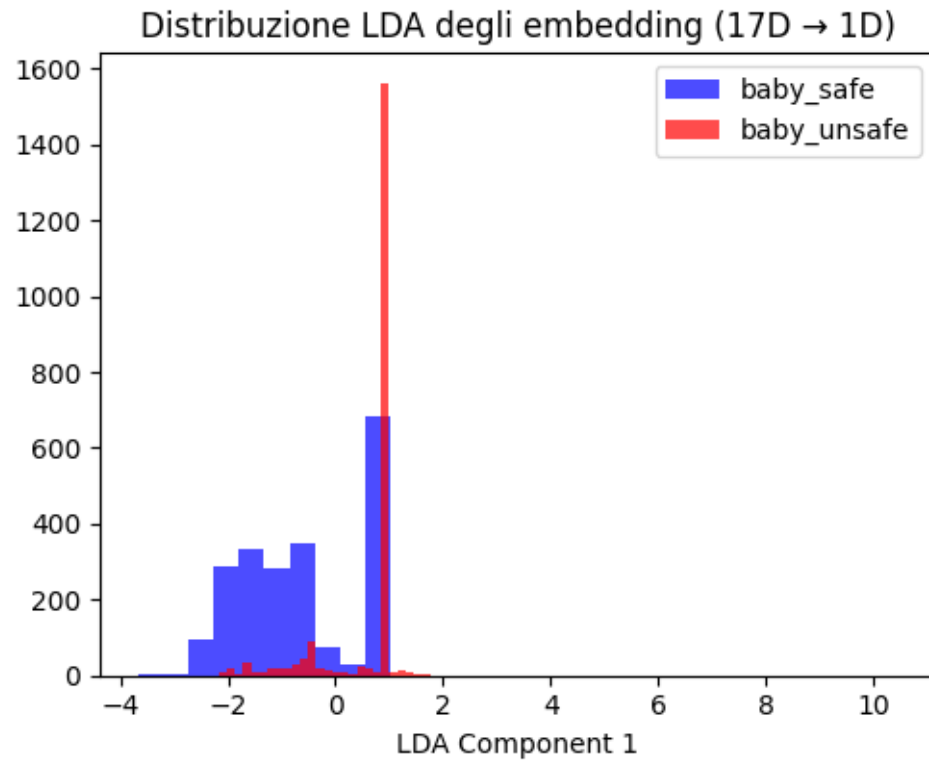
Silhouette

score-----



Embeddings

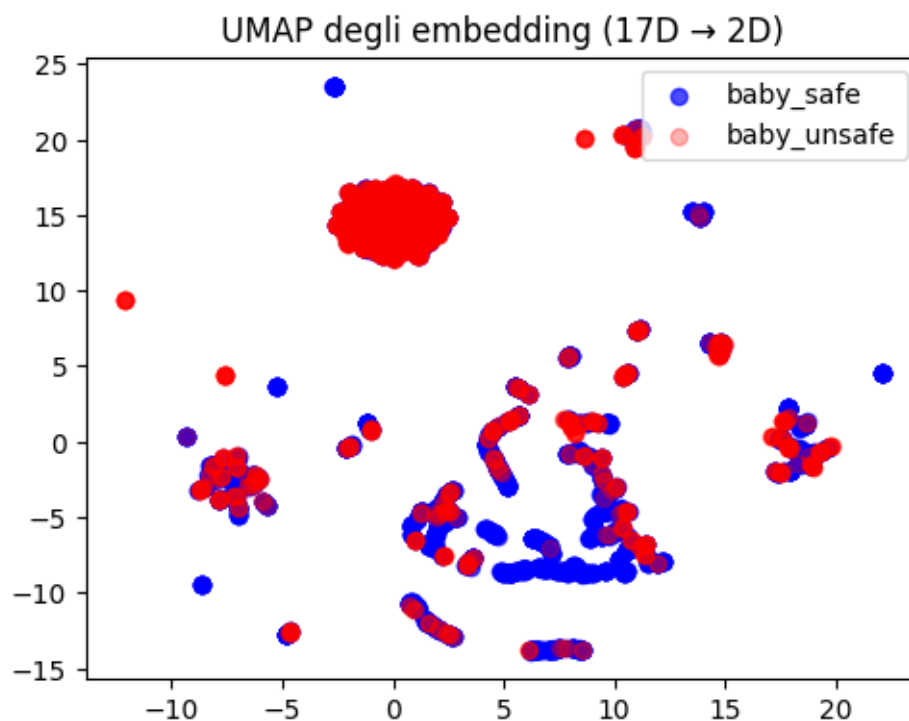
distributions-----




```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
  warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
  warnings.warn(

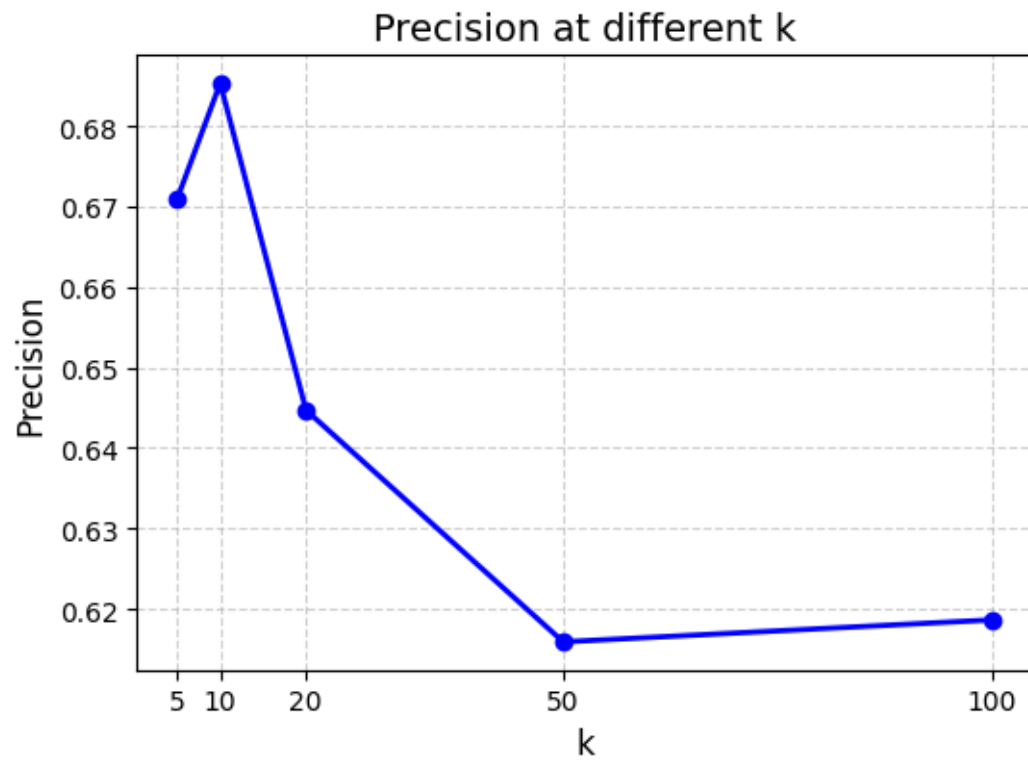
```



```
[49]: ret.report("minkowski")
```

Precision at different

k:-----

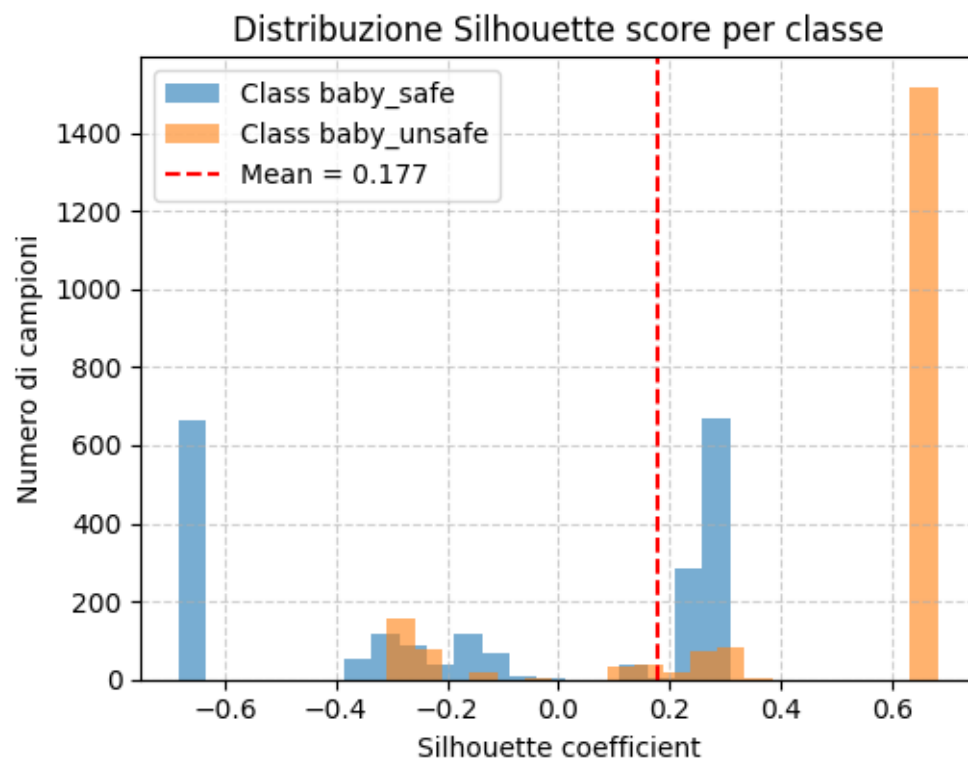


Recall at

R-----
0.5546667579984337

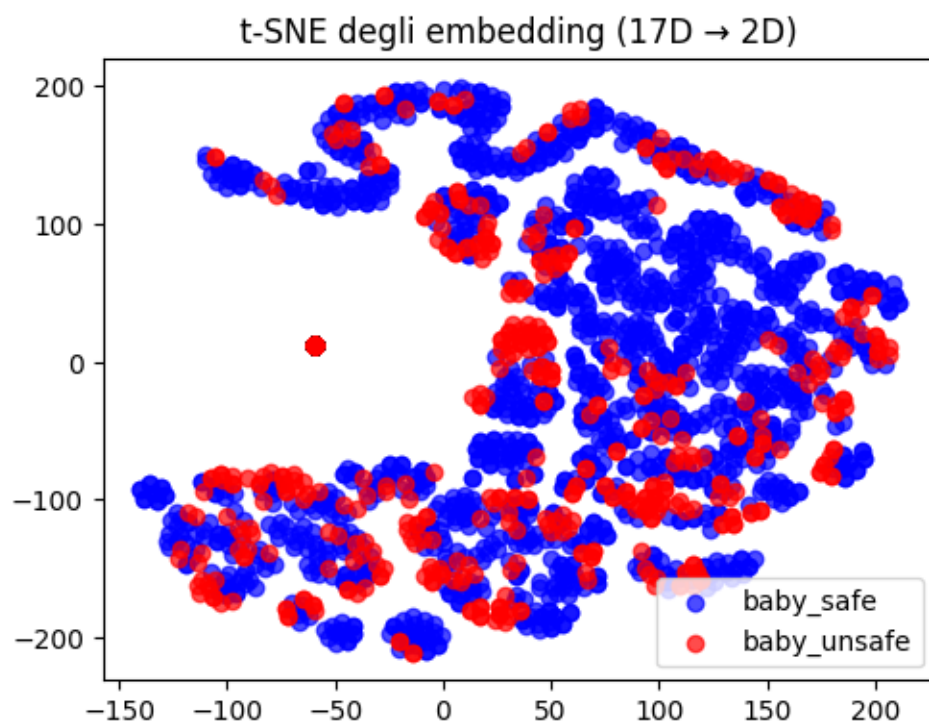
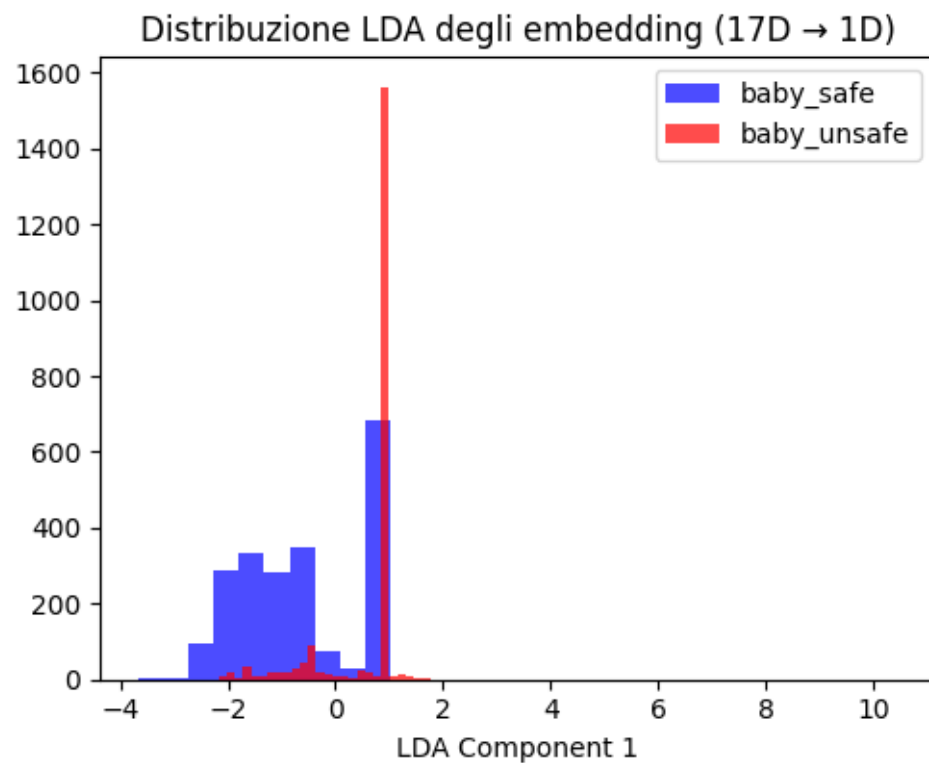
Silhouette

score-----



Embeddings

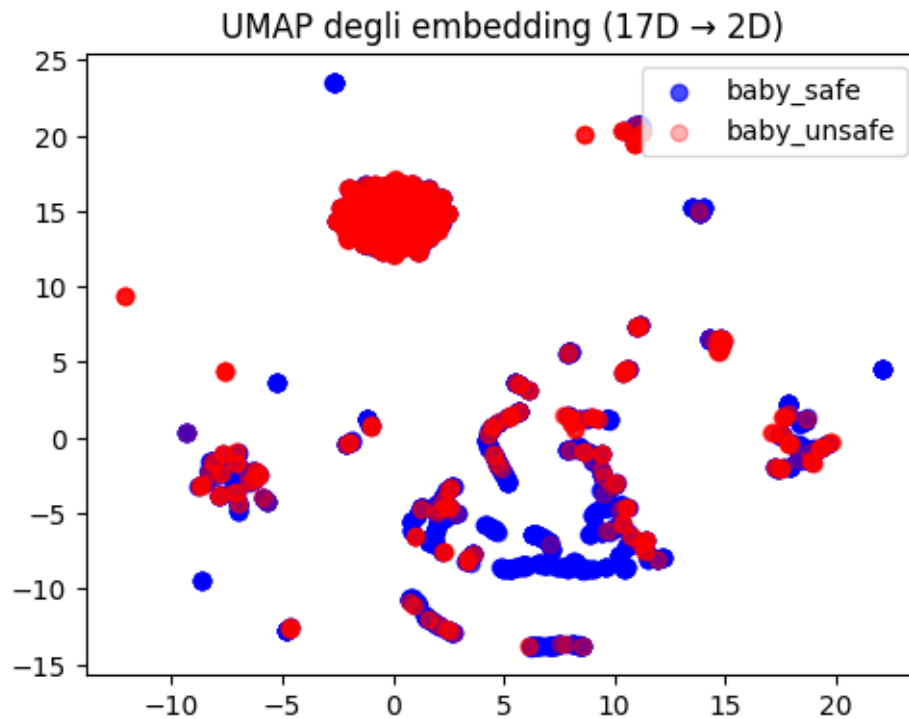
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
    warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
    warnings.warn(

```



1.8 4. Train model classification

```

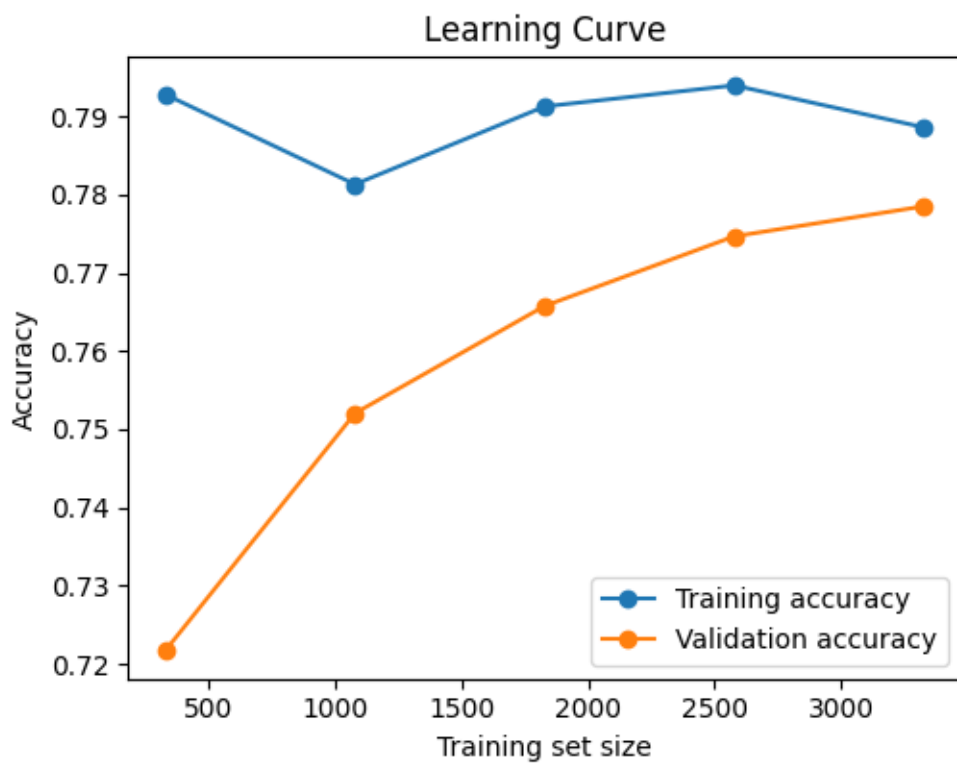
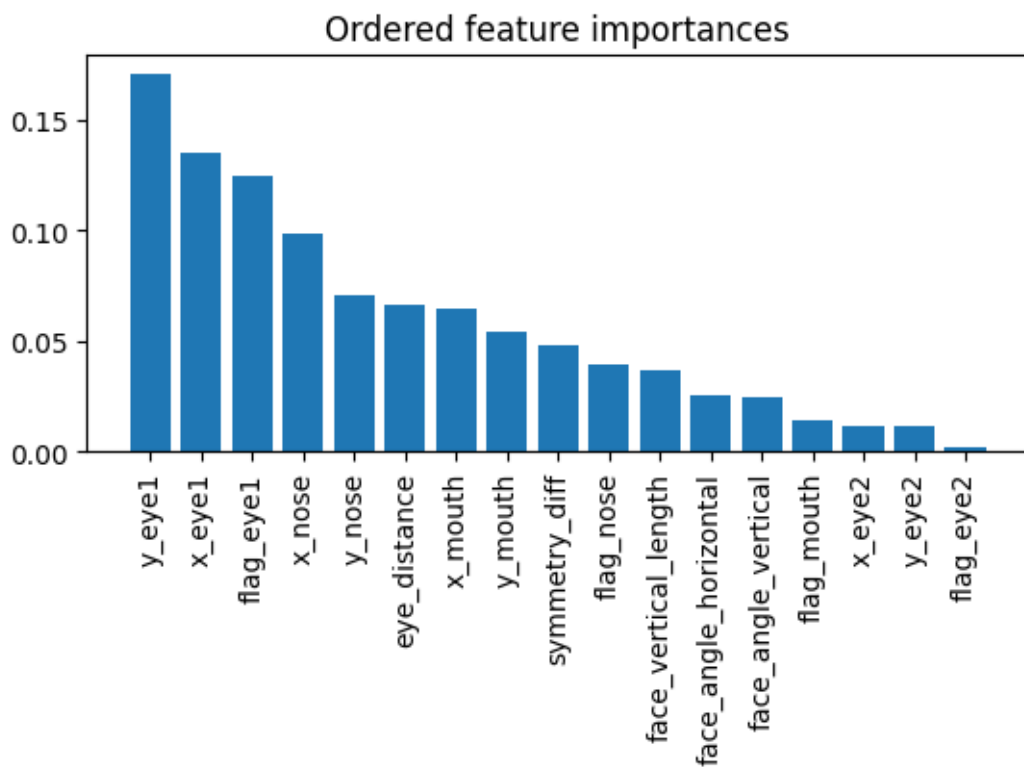
[50]: clf = Classifier(embeddings, emb_builder.y, emb_builder.classes_bs)
      clf.random_forest()

```

```

-----
-----
-----FIRST
ANALYSIS-----
-----
-----

```

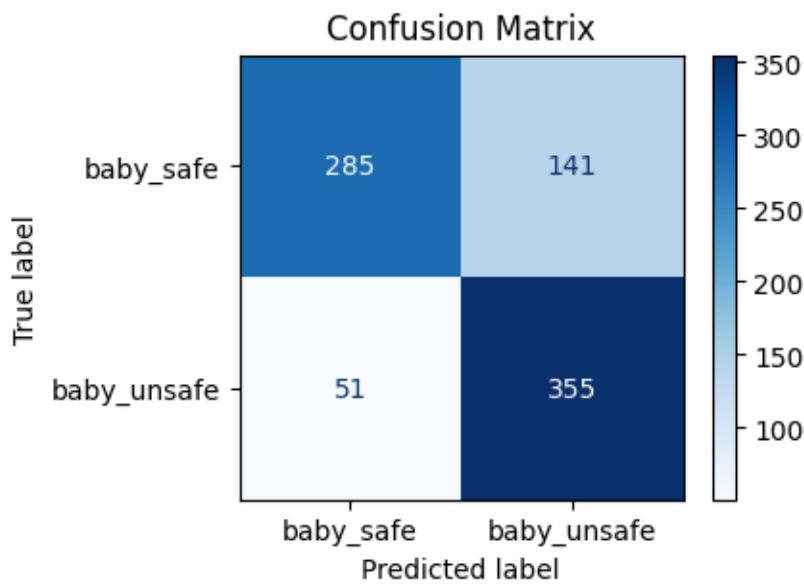


Dataset labels:-----
{'baby_safe': 0, 'baby_unsafe': 1}

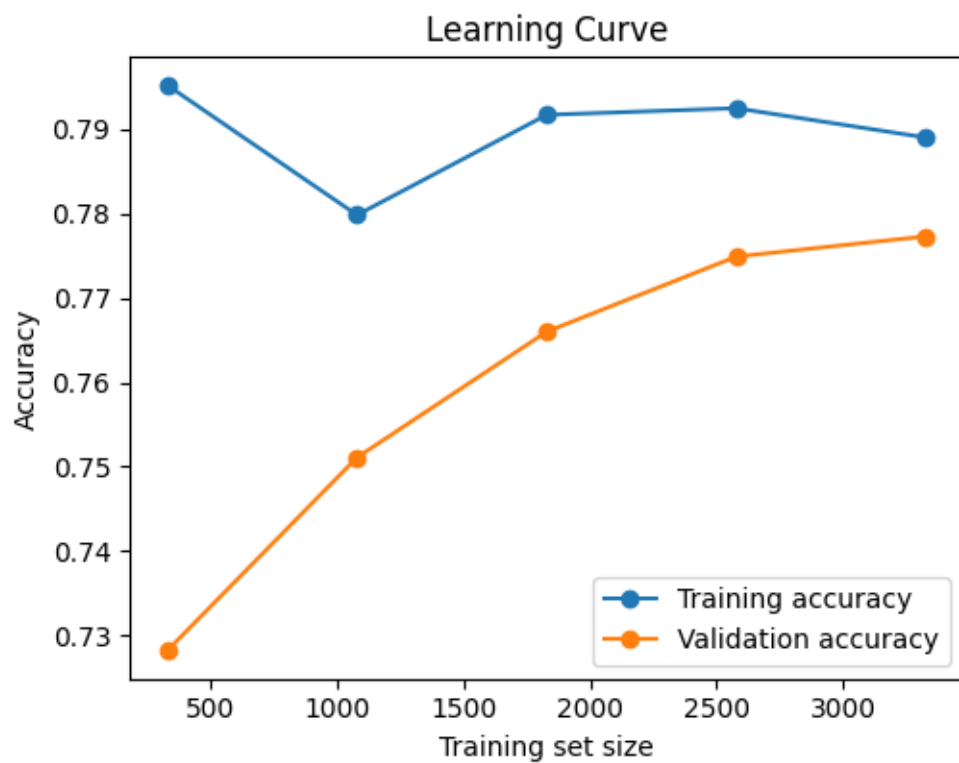
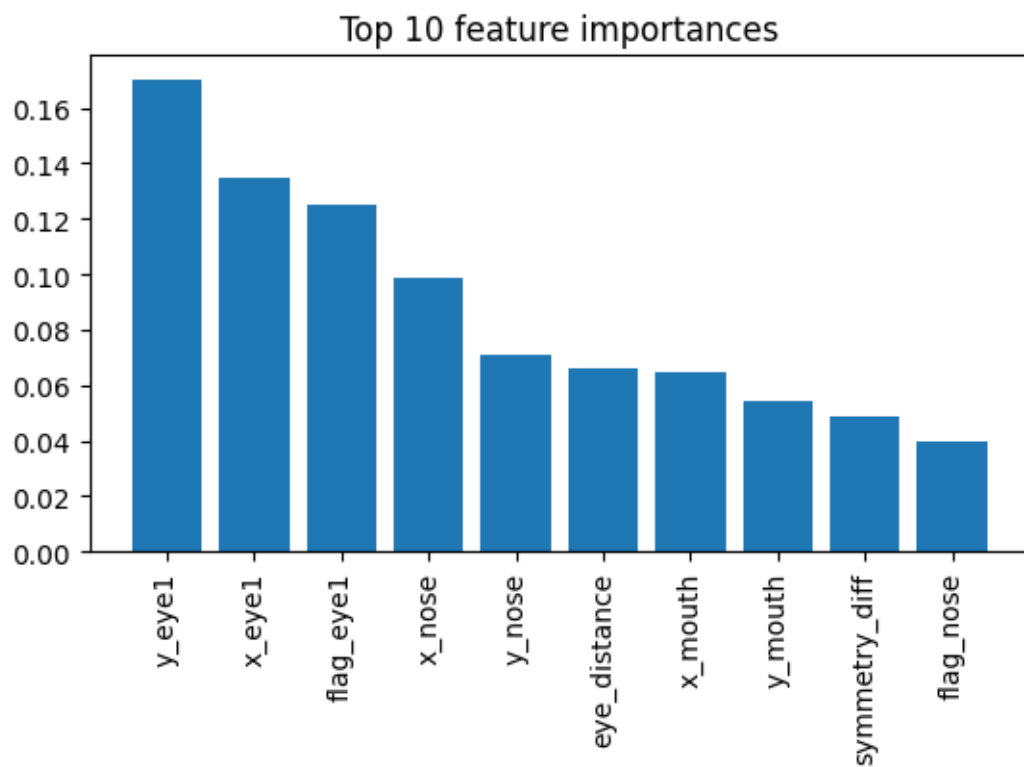
Report-----

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| baby_safe | 0.85 | 0.67 | 0.75 | 426 |
| baby_unsafe | 0.72 | 0.87 | 0.79 | 406 |
| accuracy | | | 0.77 | 832 |
| macro avg | 0.78 | 0.77 | 0.77 | 832 |
| weighted avg | 0.78 | 0.77 | 0.77 | 832 |

Confusion matrix-----



-----TOP 10 FEATURES
ANALYSIS-----

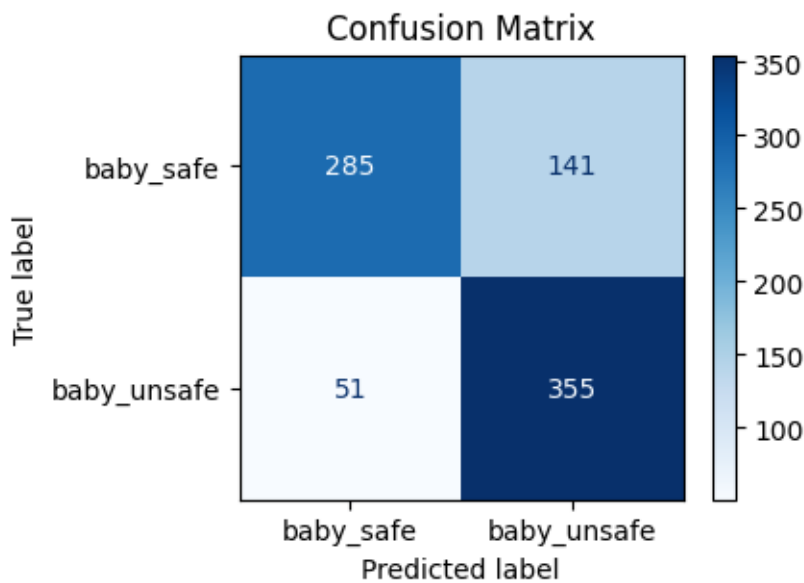


Dataset labels:-----
{'baby_safe': 0, 'baby_unsafe': 1}

Report-----

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| baby_safe | 0.85 | 0.67 | 0.75 | 426 |
| baby_unsafe | 0.72 | 0.87 | 0.79 | 406 |
| accuracy | | | 0.77 | 832 |
| macro avg | 0.78 | 0.77 | 0.77 | 832 |
| weighted avg | 0.78 | 0.77 | 0.77 | 832 |

Confusion matrix-----



1.9 2. Extract embeddings from dataset

Create embeddings

```
[51]: embeddings = emb_builder.embedding_all_features_norm()
```

Creation of all features

embedding-----
FINISHED: 4158 embedding created

```
[52]: embeddings.head()
```

```
[52]:   flag_eye1  flag_eye2  flag_nose  flag_mouth   x_eye1   y_eye1   x_eye2  \
0         0         0         0         0 -1.000000 -1.000000 -1.000000
1         0         0         0         0 -1.000000 -1.000000 -1.000000
2         0         0         0         0 -1.000000 -1.000000 -1.000000
3         1         0         1         0  0.199751  0.875817 -1.000000
4         1         1         1         1  0.747866  0.955937  0.746319

      y_eye2   x_nose   y_nose  ...  x_mouth_norm  y_mouth_norm  \
0 -1.000000 -1.000000 -1.000000  ...    -7.346064    -3.560087
1 -1.000000 -1.000000 -1.000000  ...    -1.519790    -1.998541
2 -1.000000 -1.000000 -1.000000  ...    -1.000000    -1.000000
3 -1.000000  0.246237  0.891980  ...    -6.005907    -1.195665
4  0.680452  0.709361  0.853581  ...     0.814792     1.258686

      eye_distance  eye_distance_norm  face_vertical_length  \
0        -1.00000         -1.000000         -1.000000
1        -1.00000         -1.000000         -1.000000
2        -1.00000         -1.000000         -1.000000
3        -1.00000         -1.000000         -1.000000
4         0.27549         0.685025         0.060125

      face_vertical_length_norm  face_angle_vertical  face_angle_horizontal  \
0                -1.000000         -1.000000         -1.000000
1                -1.000000         -1.000000         -1.000000
2                -1.000000         -1.000000         -1.000000
3                -1.000000         -1.000000         -1.000000
4                 0.090708        122.830696        147.334481

      symmetry_diff  head_ration
0         0.000000     1.292278
1         0.000000     1.395011
2         0.000000    -1.000000
3         0.000000     0.837183
4         0.085138     1.648186
```

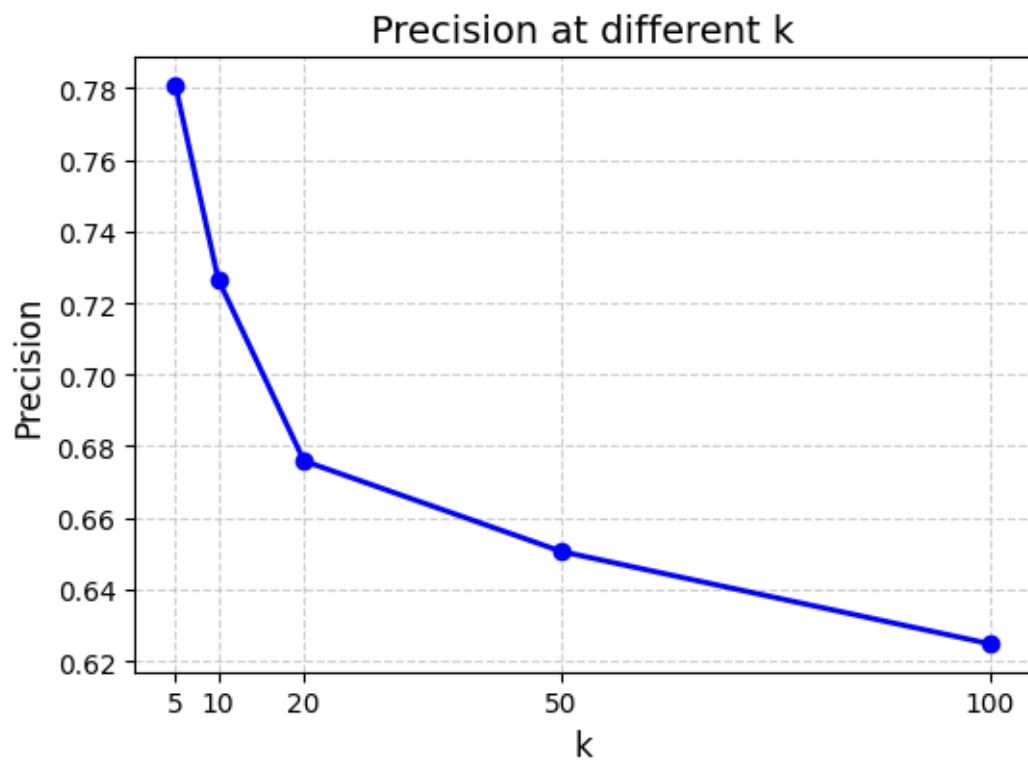
[5 rows x 28 columns]

1.10 3. Retrieval to evaluate embedding goodness

```
[53]: ret = ImageRetrieval(embeddings, emb_builder.y, emb_builder.image_paths,
    ↪ image_dataset_path, emb_builder.classes_bs)
ret.report("euclidean")
```

Precision at different

k:-----

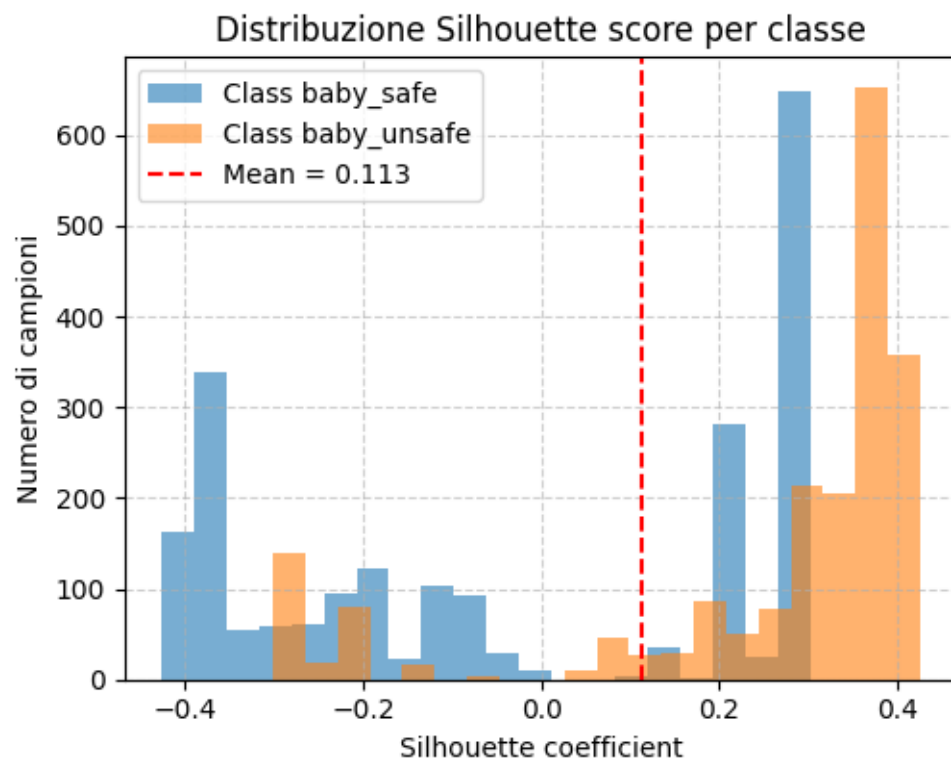


Recall at

R-----
0.5498375106937107

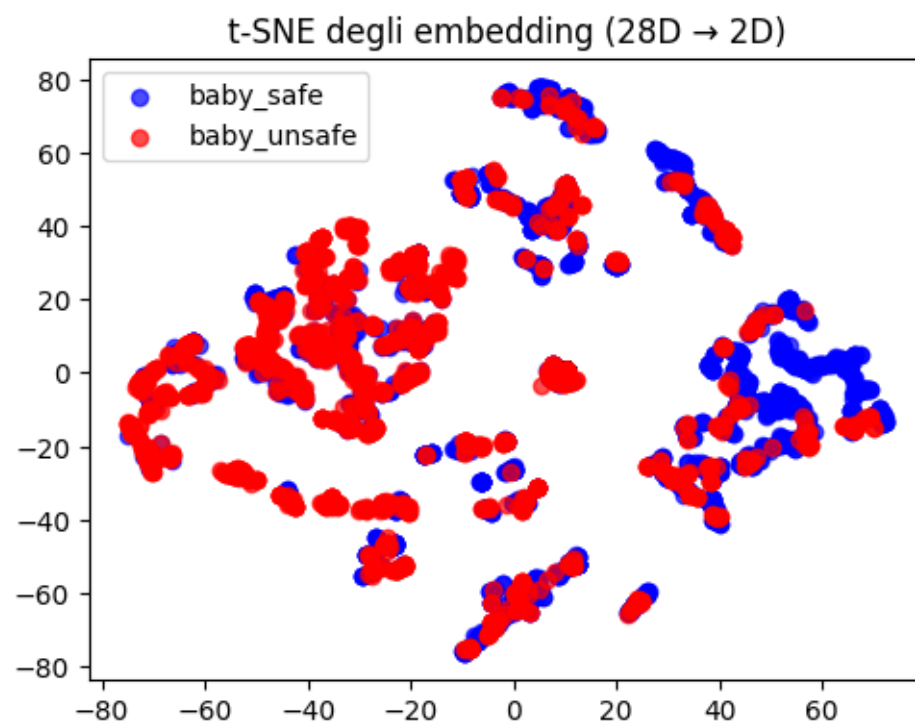
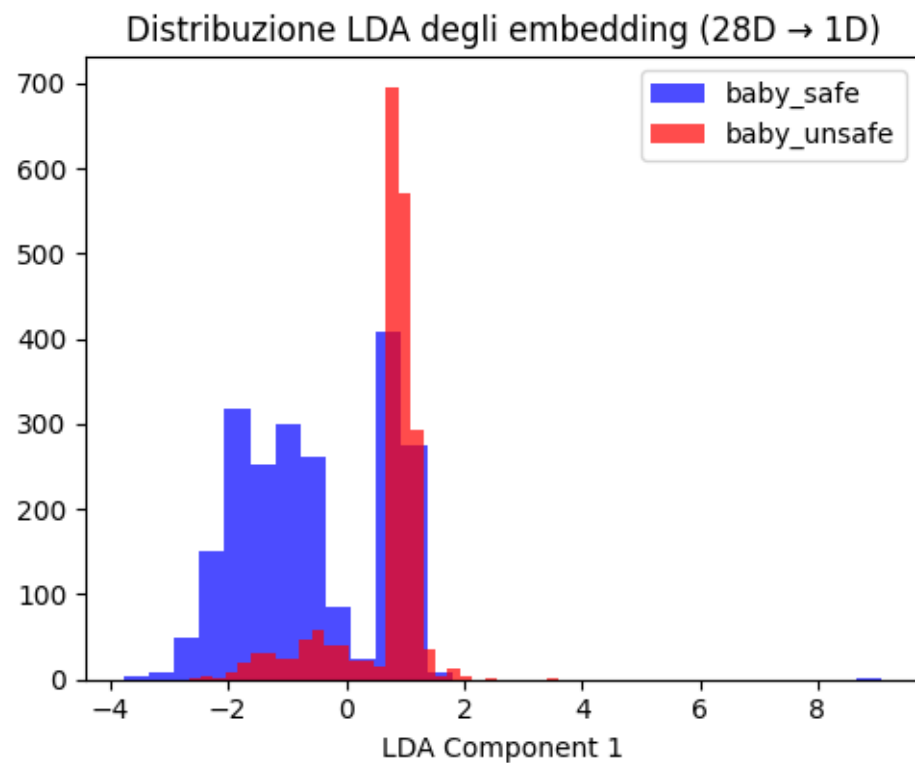
Silhouette

score-----



Embeddings

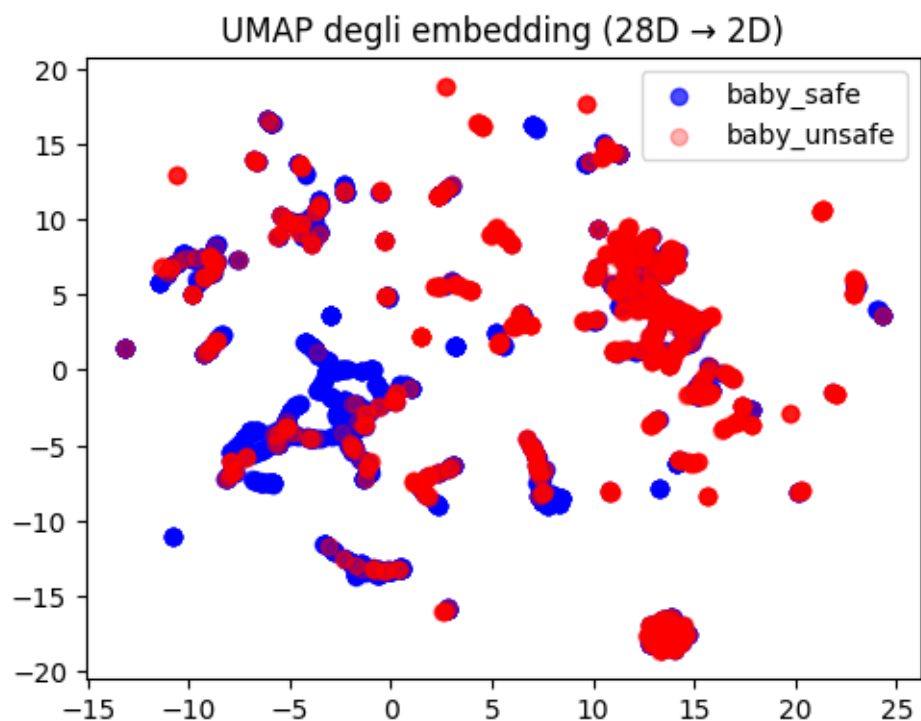
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
    warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
    warnings.warn(

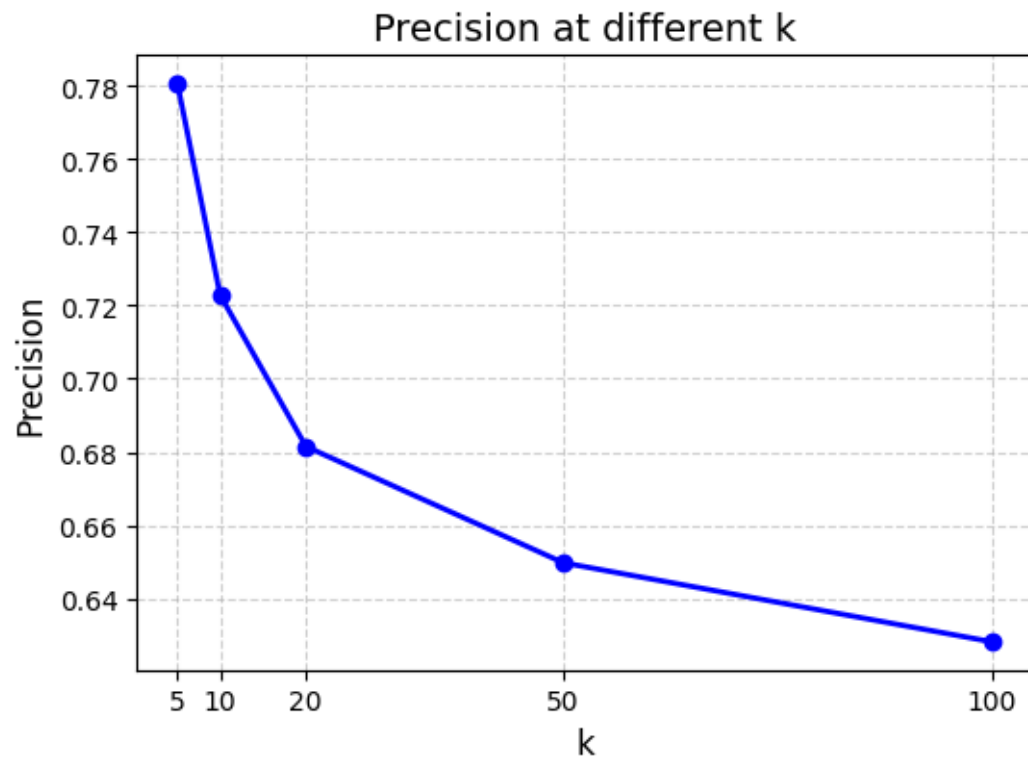
```



```
[54]: ret.report("cosine")
```

Precision at different

k:-----

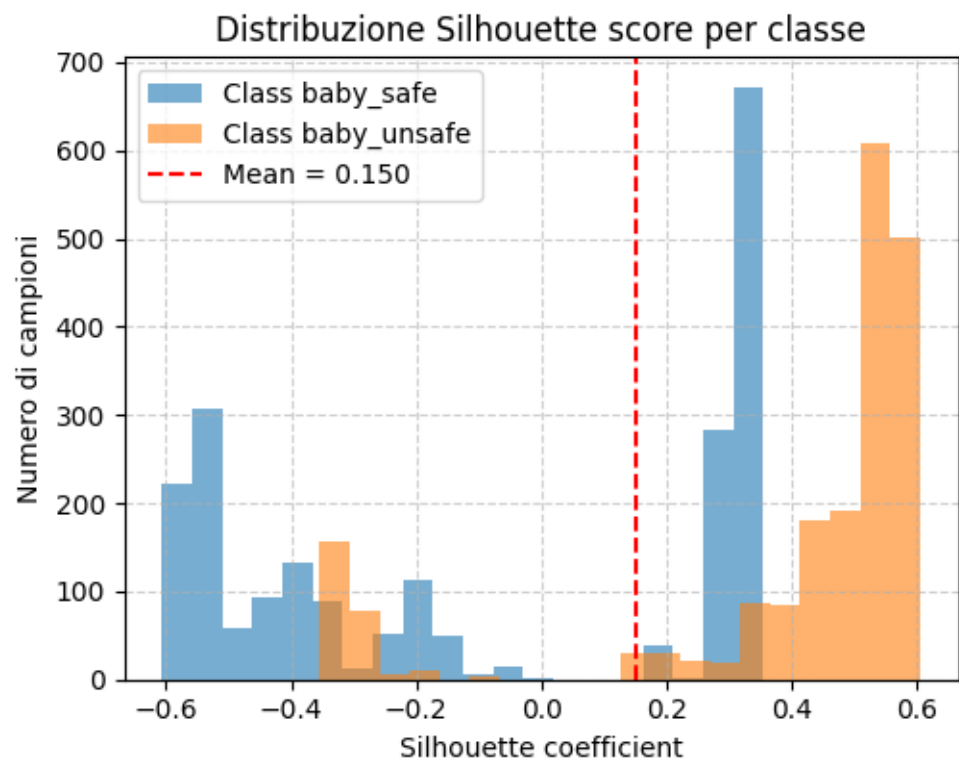


Recall at

R-----
0.5497312241944039

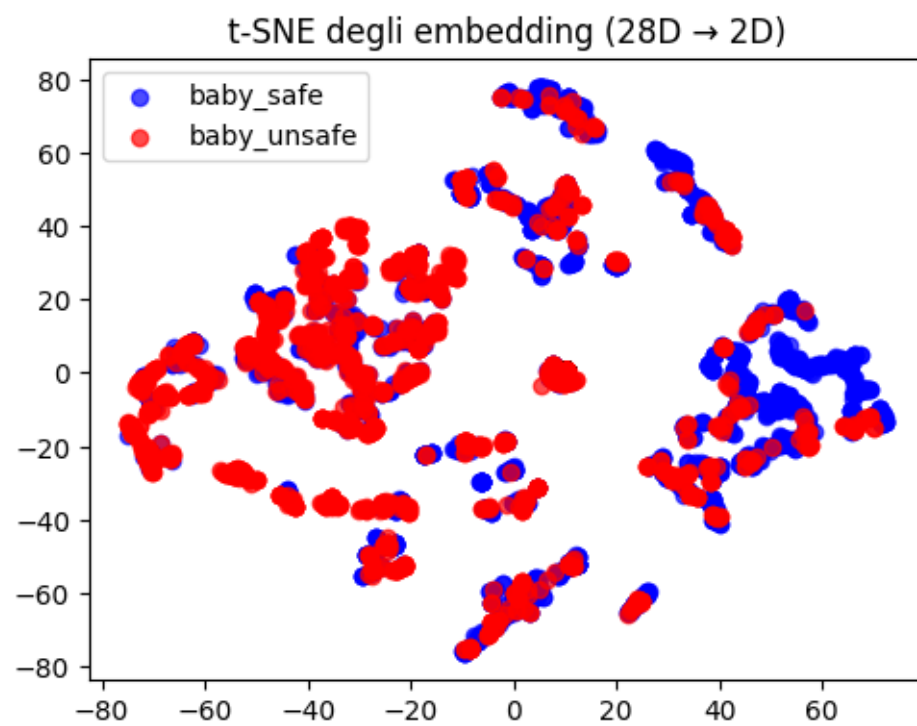
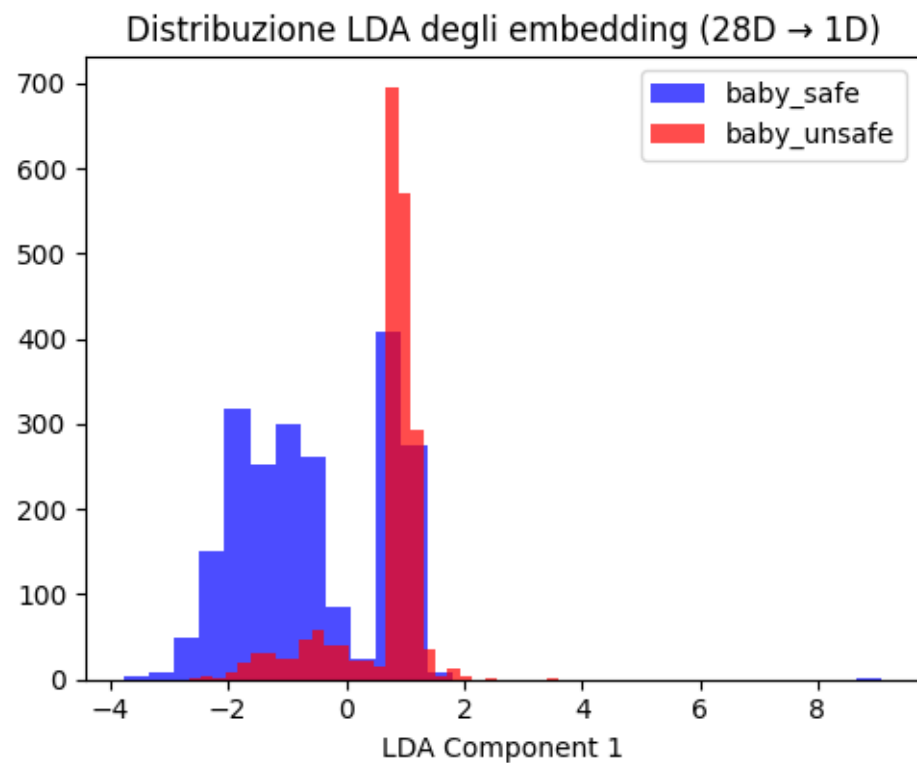
Silhouette

score-----



Embeddings

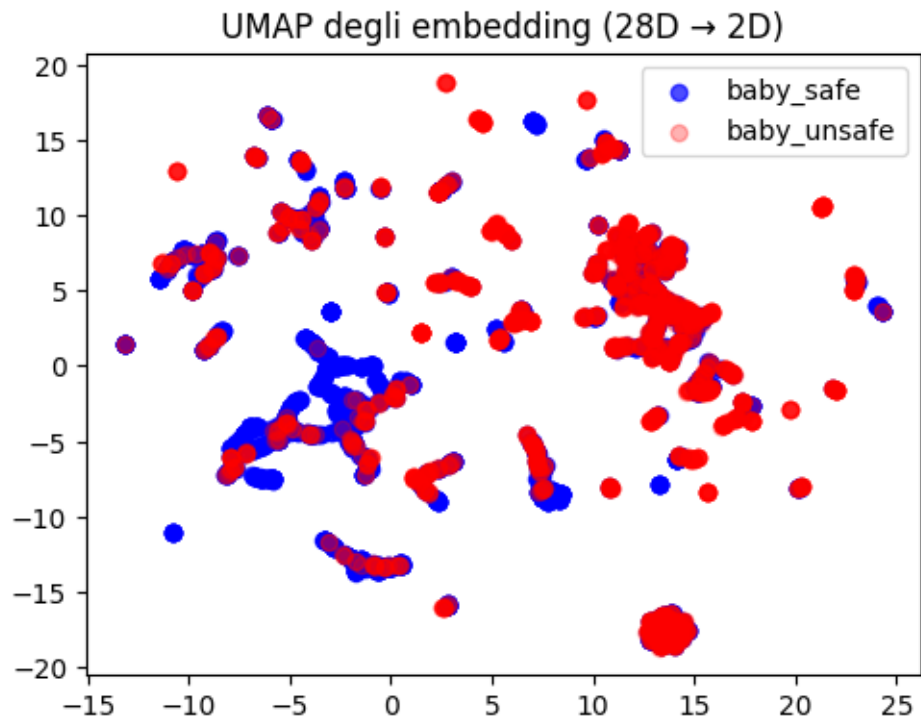
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
    warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
    warnings.warn(

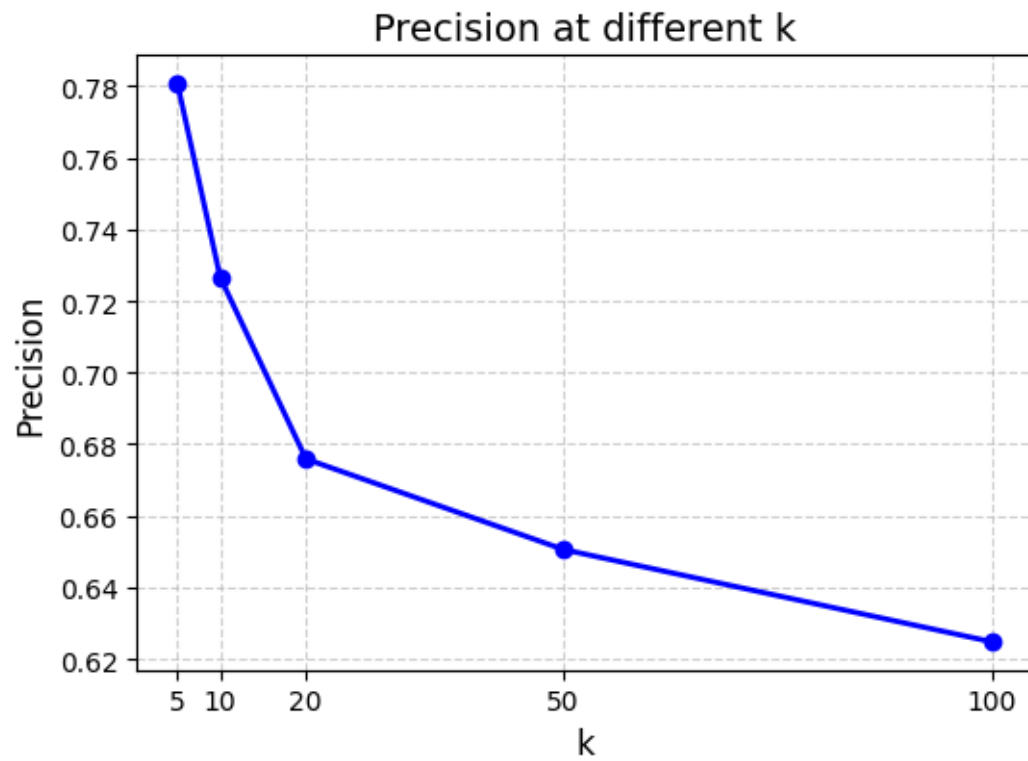
```



```
[55]: ret.report("minkowski")
```

Precision at different

k:-----

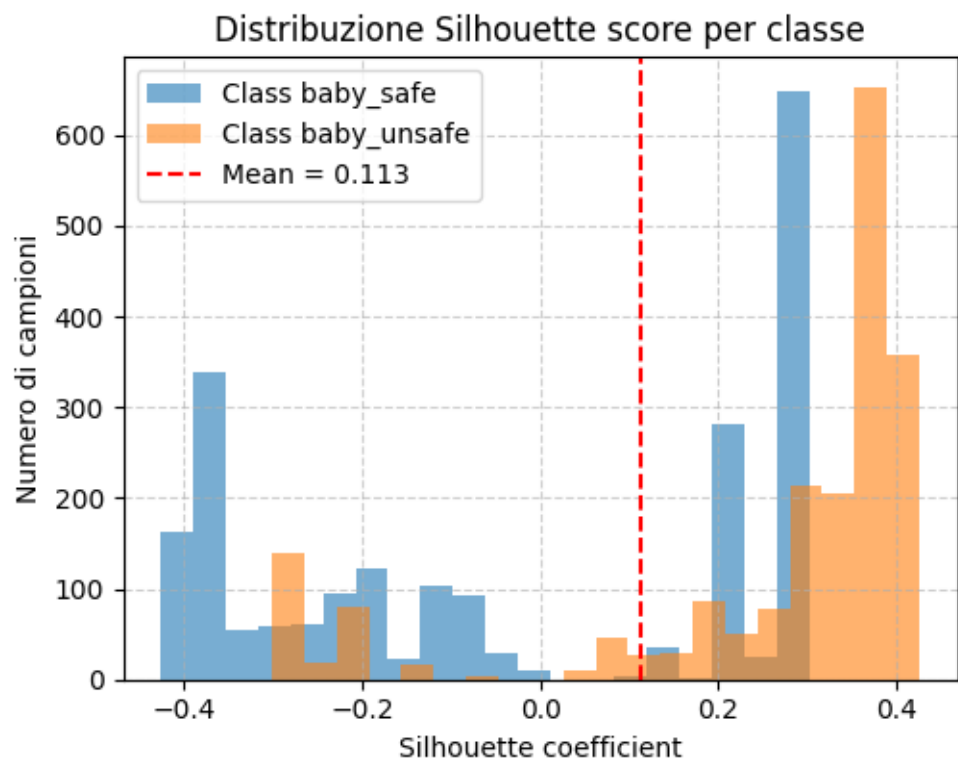


Recall at

R-----
0.5498375106937107

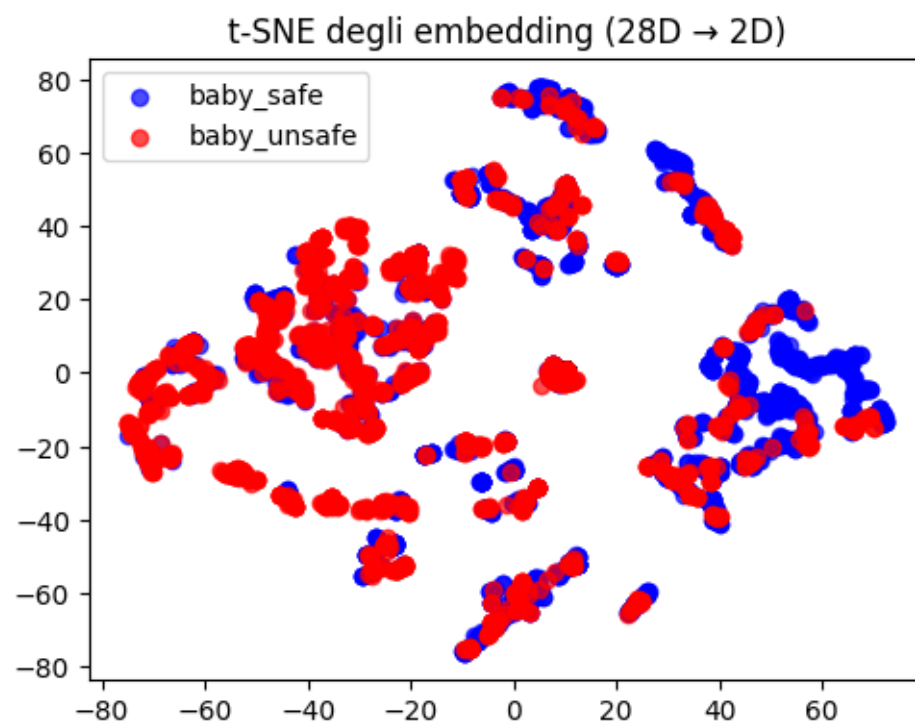
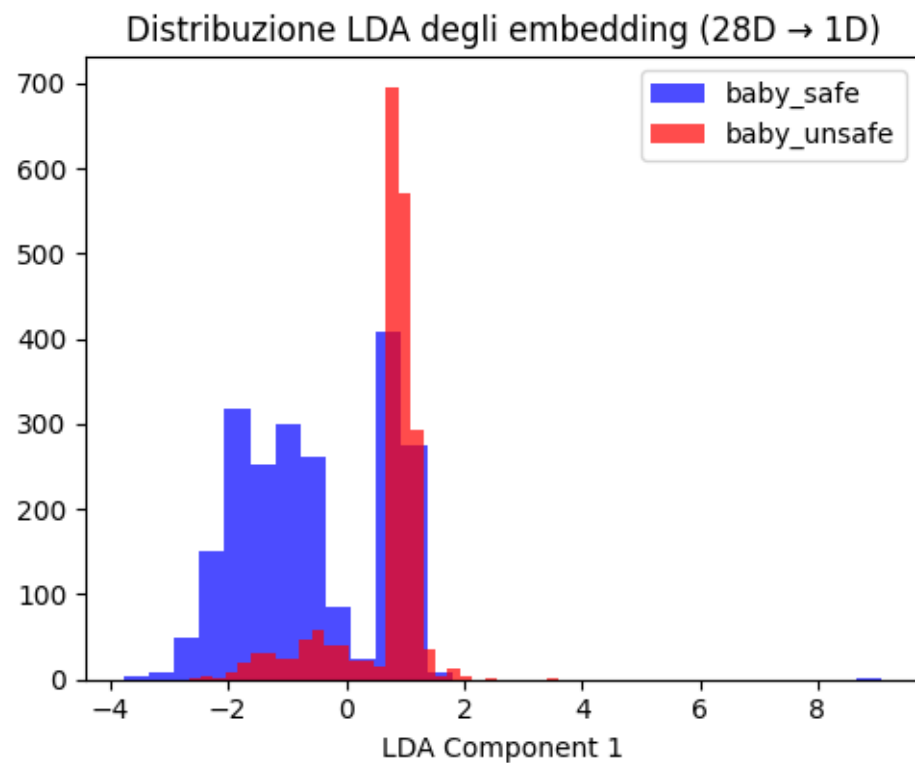
Silhouette

score-----



Embeddings

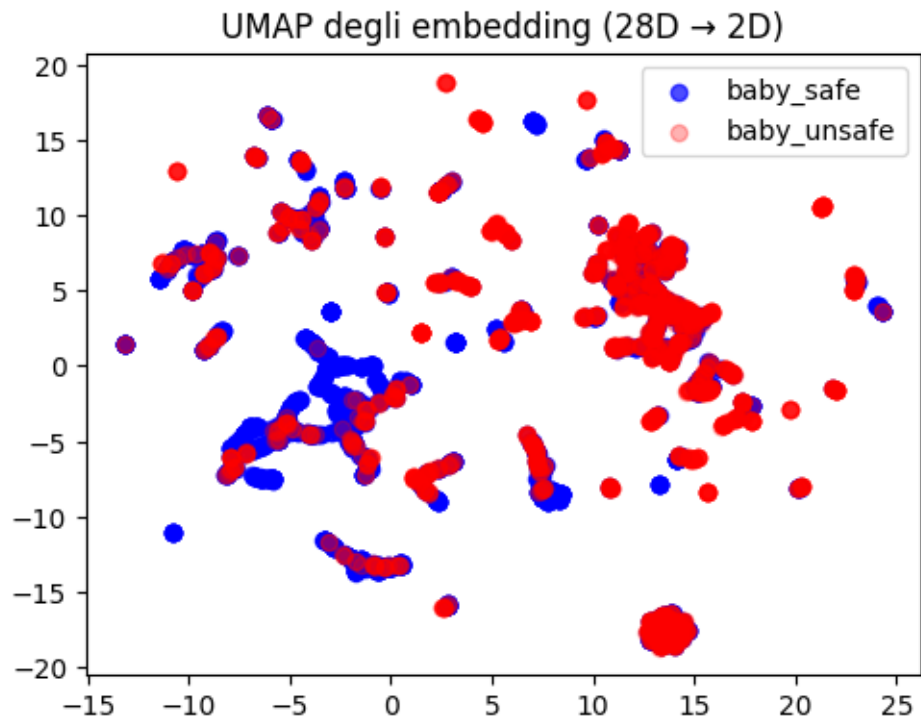
distributions-----



```

/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/umap/umap_.py:1952: UserWarning: n_jobs value 1 overridden to 1 by
setting random_state. Use no seed for parallelism.
    warn(
/home/terra/anaconda3/envs/SIDS_revelation_project/lib/python3.10/site-
packages/sklearn/manifold/_spectral_embedding.py:328: UserWarning: Graph is not
fully connected, spectral embedding may not work as expected.
    warnings.warn(

```



1.11 4. Train model classification

```

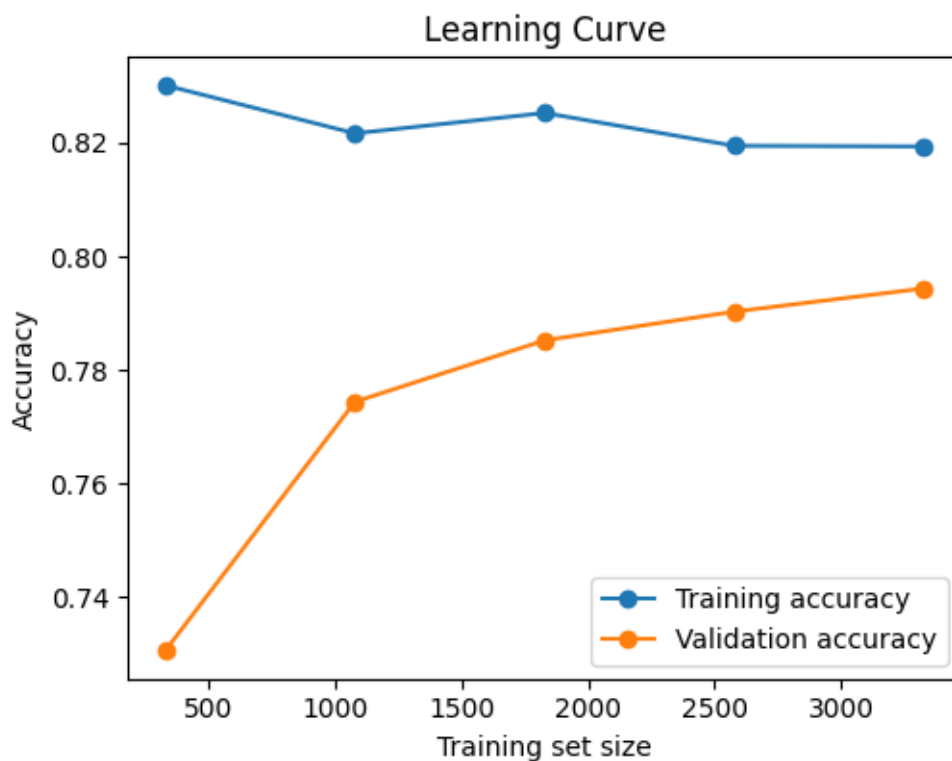
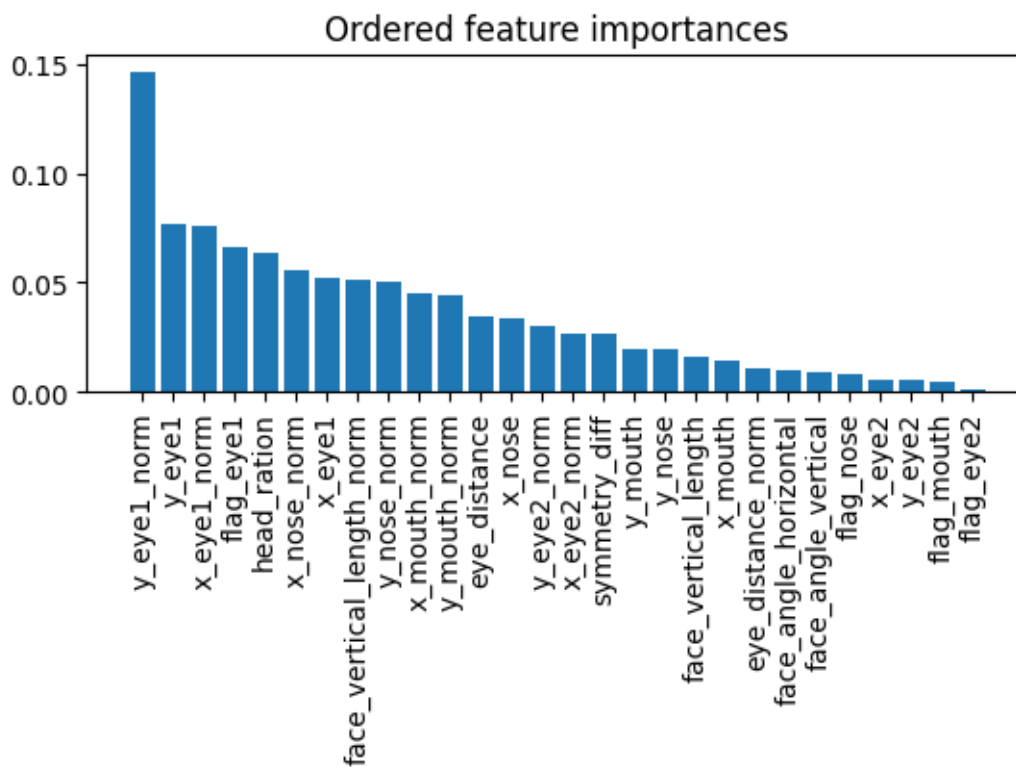
[ ]: clf = Classifier(embeddings, emb_builder.y, emb_builder.classes_bs)
      clf.random_forest()

```

```

-----
-----
-----FIRST
ANALYSIS-----
-----
-----

```

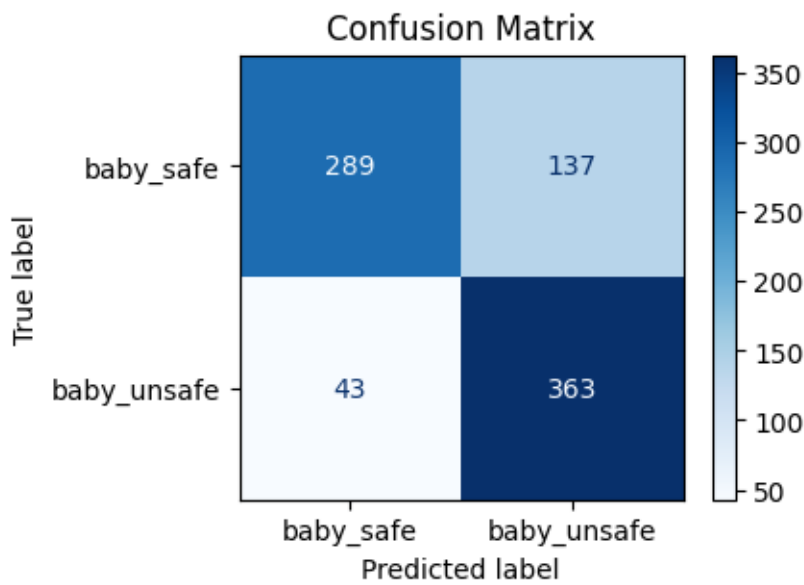


Dataset labels:-----
{'baby_safe': 0, 'baby_unsafe': 1}

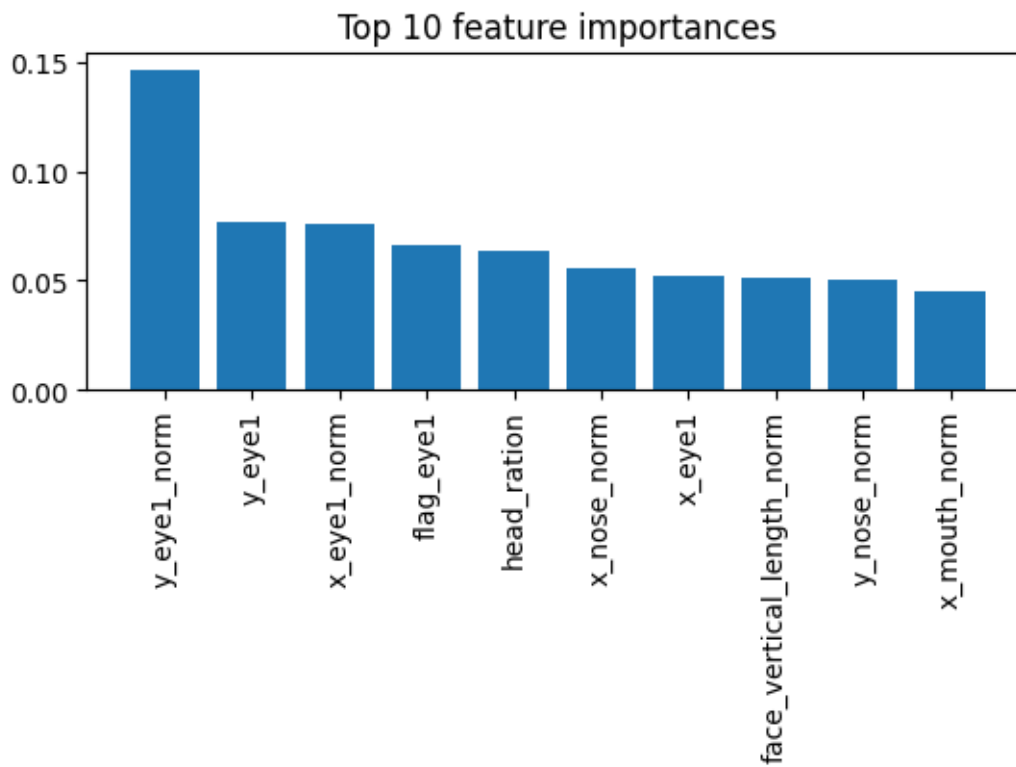
Report-----

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| baby_safe | 0.87 | 0.68 | 0.76 | 426 |
| baby_unsafe | 0.73 | 0.89 | 0.80 | 406 |
| accuracy | | | 0.78 | 832 |
| macro avg | 0.80 | 0.79 | 0.78 | 832 |
| weighted avg | 0.80 | 0.78 | 0.78 | 832 |

Confusion matrix-----



-----TOP 10 FEATURES
ANALYSIS-----



```
[ ]: clf.XGBC()
```

```
[ ]: clf.logistic_regression()
```

1.12 Save notebook

```
[ ]: file_manager.save_as_pdf(ipynbname.path())
```