**Gender Classification using Machine Learning Documentation Report**

**Introduction**

The Gender Classification using Machine Learning project aims to develop and evaluate models for predicting the gender of individuals based on facial images. This report provides a step-by-step overview of the project, including data collection, preprocessing, model selection, training, and evaluation.

**Data Collection**

The dataset used for this project was sourced from Kaggle and can be accessed through the following link: [CelebA Dataset](https://www.kaggle.com/datasets/jessicali9530/celeba-dataset/code). The dataset contains a large number of facial images with corresponding gender labels.

**Preprocessing**

1. Data Reduction

One of the challenges encountered during data preprocessing was the large number of images, approximately 200,000. To address this issue, the dataset was reduced to a more manageable size by selecting 5,000 images for the project.

2. Image Resizing

Another preprocessing challenge was the varying sizes of images. Images were resized to a common dimension of 64x64 pixels to ensure uniformity and reduce computational complexity.

**Model Selection**

To select the best machine learning model for gender classification, a variety of classification algorithms were explored. Each model was trained and evaluated on the preprocessed dataset. Here are the results for model accuracy and precision:

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Name** | **Model accuracy** | **Model precision** | **training\_time** |
| AdaBoostClassifier | 0.55 | 0.429553265 | 226.5435858 |
| BaggingClassifier | 0.566 | 0.441314554 | 464.8457623 |
| BernoulliNB | 0.486 | 0.399617591 | 0.907184124 |
| CalibratedClassifierCV | 0.591 | 0 | 650.2261753 |
| DecisionTreeClassifier | 0.506 | 0.4 | 76.95646644 |
| DummyClassifier | 0.591 | 0 | 0.004359484 |
| ExtraTreeClassifier | 0.505 | 0.39953271 | 0.308481455 |
| ExtraTreesClassifier | 0.573 | 0.428571429 | 16.51778221 |
| GaussianNB | 0.493 | 0.405405405 | 1.015052319 |
| GaussianProcessClassifier | 0.591 | 0 | 227.4593296 |
| GradientBoostingClassifier | 0.559 | 0.394736842 | 1294.621638 |
| HistGradientBoostingClassifier | 0.546 | 0.400881057 | 294.0355227 |
| KNeighborsClassifier | 0.541 | 0.433862434 | 0.055232048 |
| LabelPropagation | 0.409 | 0.409 | 7.818722725 |
| LabelSpreading | 0.409 | 0.409 | 7.577718973 |
| LinearDiscriminantAnalysis | 0.493 | 0.391111111 | 128.5558164 |
| LinearSVC | 0.521 | 0.41745283 | 143.4746962 |
| LogisticRegression | 0.526 | 0.421307506 | 14.23765182 |
| LogisticRegressionCV | 0.572 | 0.405940594 | 476.9548936 |
| MLPClassifier | 0.54 | 0.437037037 | 186.997484 |
| NearestCentroid | 0.481 | 0.396616541 | 0.254918814 |
| NuSVC | 0.534 | 0.413897281 | 225.2420442 |
| PassiveAggressiveClassifier | 0.519 | 0.414691943 | 41.95060253 |
| Perceptron | 0.518 | 0.412887828 | 9.607045889 |
| QuadraticDiscriminantAnalysis | 0.514 | 0.418259023 | 69.20334578 |
| RandomForestClassifier | 0.563 | 0.393939394 | 0.052495241 |
| RidgeClassifier | 0.514 | 0.417910448 | 41.18847489 |
| RidgeClassifierCV | 0.507 | 0.408296943 | 8.572001696 |
| SGDClassifier | 0.533 | 0.431279621 | 28.29031157 |
| SVC | 0.594 | 0.542857143 | 13.57183933 |

The above table provides an overview of model performance metrics. It is important to note that the choice of the final model should consider not only accuracy but also precision and other relevant metrics based on the project's requirements.

**Model Training**

The selected machine learning models were trained on the preprocessed dataset. Training time was measured for each model to assess computational cost and efficiency.

**Model Evaluation**

The models were evaluated using various metrics, including accuracy, precision, recall, F1-score, and confusion matrix. These metrics provide a comprehensive view of each model's performance and help in selecting the most suitable model for gender classification.

**Conclusion**

The Gender Classification using Machine Learning project involved data collection, preprocessing, model selection, training, and evaluation. Through these steps, we identified models that demonstrated promising performance in predicting gender based on facial images. Further refinement and fine-tuning of the selected model will lead to a robust gender classification system.