**Gender Classifier Comparison using Various Machine Learning Algorithms**

**Usama Arif C1 -14**

[**Usamaarifbhatti594@gmail.com**](mailto:Usamaarifbhatti594@gmail.com)

**Introduction:**

The classification of gender based on facial images has gained significant attention due to its potential applications in various fields such as security, marketing, and entertainment. In this report, we explore the performance of different machine learning algorithms for the task of gender classification using facial images. The algorithms considered are k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Random Forest, Logistic Regression, and Decision Tree. We also discuss the preprocessing steps involved in preparing the data for training and testing, including resizing, normalization, label encoding, and random permutation.

**Preprocessing Steps**

1. **Resizing:**

The first step in preprocessing involves resizing the input images to a consistent size. This is crucial to ensure that all images have the same dimensions, which is a requirement for many machine learning algorithms. Resizing also helps to reduce the computational complexity during training.

1. **Normalization:**

Normalization is performed to scale the pixel values of the images to a common range, typically between 0 and 1. This step improves the convergence of training algorithms and makes the model less sensitive to variations in pixel intensity.

1. **Label Encoding:**

Label encoding is used to convert categorical labels (e.g., 'male' and 'female') into numerical values. This is necessary for most machine learning algorithms that require numerical inputs. In this case, 'male' is encoded as 0 and 'female' as 1.

1. **Random Permutation:**

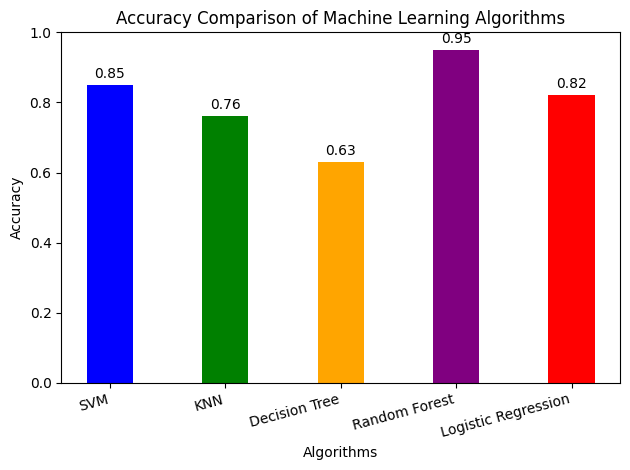
Random permutation of the dataset is important to introduce randomness and prevent any potential bias that might arise from the ordering of data samples. It helps to ensure that the training and testing sets are representative of the entire dataset.

**Algorithm Performance and Analysis:**

The accuracy results for the gender classification task using different machine learning algorithms are as follows:

* Support Vector Machine (SVM): 85%
* Random Forest: 95%
* k- Logistic Regression: 82%
* Nearest Neighbors (k-NN): 76%
* Decision Tree: 63%

**Performance Visualization:**



**Model Evaluation:**

When we use new pictures in the models, they usually give results that are quite similar to their accuracies. This means they're good at telling if someone is male or female.

**Conclusion:**

In conclusion the Random Forest model was the best at making predictions with a very high accuracy of 95%. The SVM and Logistic Regression models did well too, with accuracies of 85% and 82%. KNN was decent with 76% accuracy, but the Decision Tree model didn't perform as well, only achieving 63%. This information helps us choose the right model for different jobs where we need predictions.