20B91A05W0

RIGHT ALGORITHM:

The "best" classifier for resume screening using NLP depends on various factors such as the size and nature of the dataset, the complexity of the classification task, computational resources available, and the desired performance metrics. Different classifiers may perform differently based on these factors. However, I can provide some general insights into each of the classifiers listed:

OneVsRestClassifier with KNeighborsClassifier: KNeighborsClassifier is a simple and intuitive algorithm that works well for certain types of data. It tends to perform better on smaller datasets with lower dimensionality. However, its performance can degrade as the dataset grows larger due to its computational inefficiency and memory requirements.

OneVsRestClassifier with LogisticRegression: Logistic Regression is a widely used linear classification algorithm that works well for binary and multiclass classification tasks. It's computationally efficient, interpretable, and performs well when the data is linearly separable or close to linearly separable. However, its performance may suffer when dealing with complex non-linear relationships in the data.

OneVsRestClassifier with SVC (Support Vector Classifier): SVC is a powerful classification algorithm that is effective in high-dimensional spaces and can capture complex relationships in the data through the use of kernel functions. It tends to perform well in various scenarios, but its training time and computational complexity can be higher, especially with large datasets.

OneVsRestClassifier with RandomForestClassifier: RandomForestClassifier is an ensemble learning method based on decision trees. It's known for its robustness, scalability, and ability to handle high-dimensional data with complex interactions. It typically performs well across a wide range of datasets and can handle noisy data effectively. However, it may not be as interpretable as some other algorithms, and its performance can degrade if hyperparameters are not tuned properly.

**Accuracies of Respective Models:**

Accuracy of OneVsRestClassifier(estimator=KNeighborsClassifier()) on training set : 0.9805825242718447

Accuracy of OneVsRestClassifier(estimator=KNeighborsClassifier()) on test set : 0.966804979253112

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Accuracy of OneVsRestClassifier(estimator=LogisticRegression()) on training set : 1.0

Accuracy of OneVsRestClassifier(estimator=LogisticRegression()) on test set : 0.991701244813278

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Accuracy of OneVsRestClassifier(estimator=SVC()) on training set : 1.0

Accuracy of OneVsRestClassifier(estimator=SVC()) on test set : 0.991701244813278

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Accuracy of OneVsRestClassifier(estimator=RandomForestClassifier()) on training set : 1.0

Accuracy of OneVsRestClassifier(estimator=RandomForestClassifier()) on test set : 0.983402489626556

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

all scores calculated