

ABSTRACT

According to the embodiments of the present disclosure, a method and apparatus for protecting data, and a device and a medium are provided. The method comprises acquiring, by a first device, a feature representation generated by a second device on the basis of sample data and according to a second model, wherein the first device has label information for the sample data; and the first device and the second device are used for jointly training a first model at the first device and the second model at the second device. The method further comprises generating, by the first device, a prediction label for the sample data on the basis of the feature representation and according to the first model. The method further comprises determining, by the first device and on the basis of the feature representation, the label information and the predicted label, a total loss value used for training the first model and the second model. In this way, a second device, which does not have label information of data, can be prevented from predicting the label information of the data, thereby protecting data, especially sensitive data, from being leaked.