**Machine Learning for Portfolio Selection**

Date Pre-processing extension

Before feeding the data set to models, data pre-processing is very important, In the real-world data are generally incomplete. After adding features and selecting the proper features, the data set also needs to handle the missing value and do Feature Scaling. For missing values, in this project, it’s simple to deal with. Here, missing values are filled by previous values, and keep null if no previous value. Because the data is stock price time series, if there is one missing value on a trading date, it’s reasonable to fill it with the pre-trading date stock price. For feature scaling, standardization is used to scale the features, and from the figure 1 below, it works well for scaling-sensitive models like SVM, which performs better after scaling.

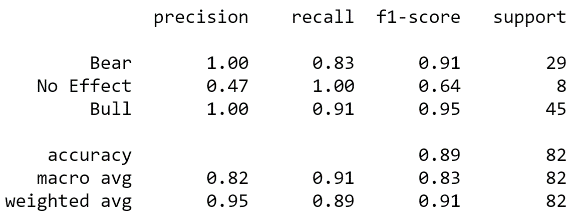


Figure 1. Classification report (Left: Plain SVM Right: SVM With Scaled)