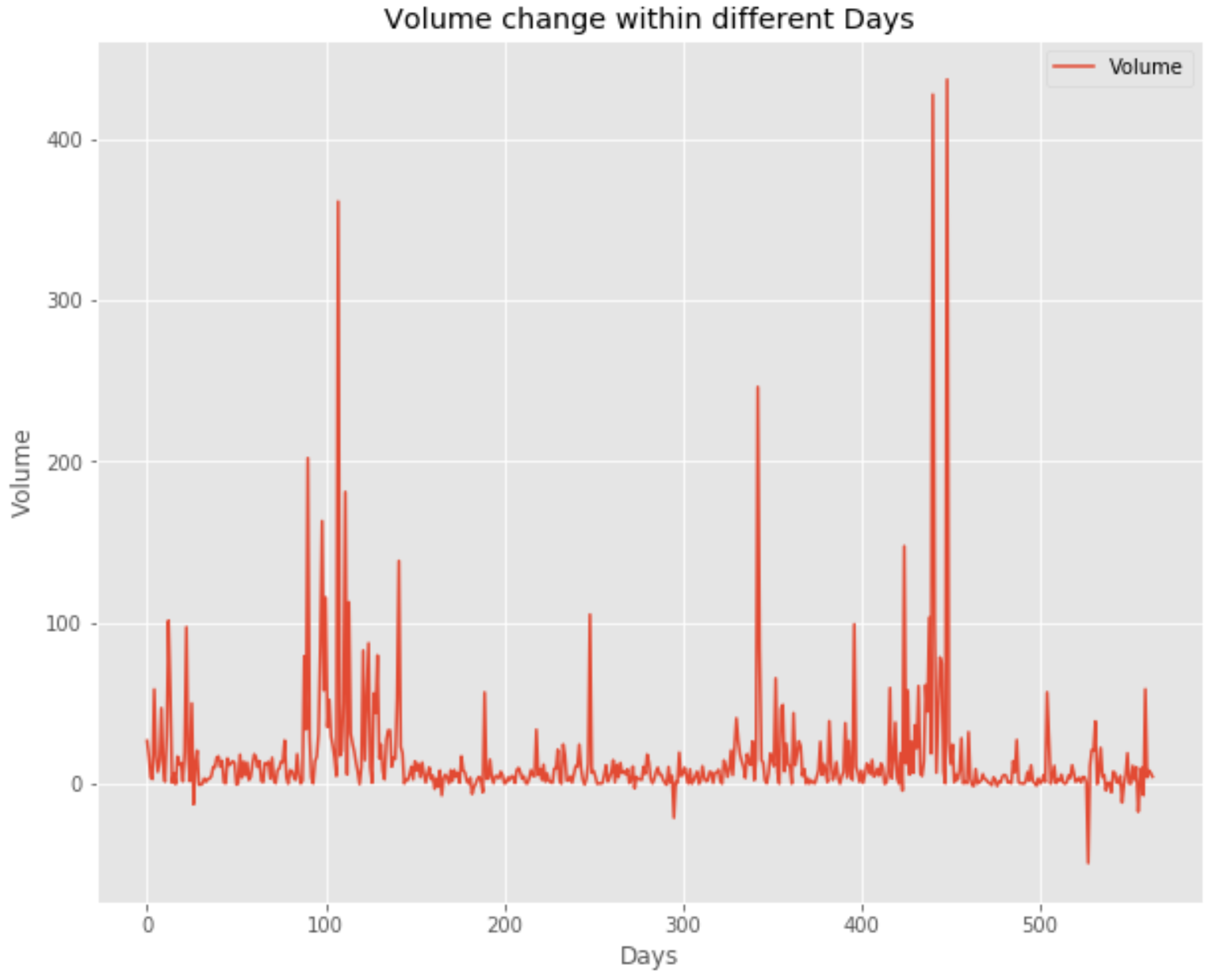
CarreFour销量预测

本次针对CarreFour进行了销量时间序列预测，因本次目标值需要对之后的7天分别进行预测，所以问题为MultiOuputRegression。具体数据情况，过程及结果如下：

1. 数据

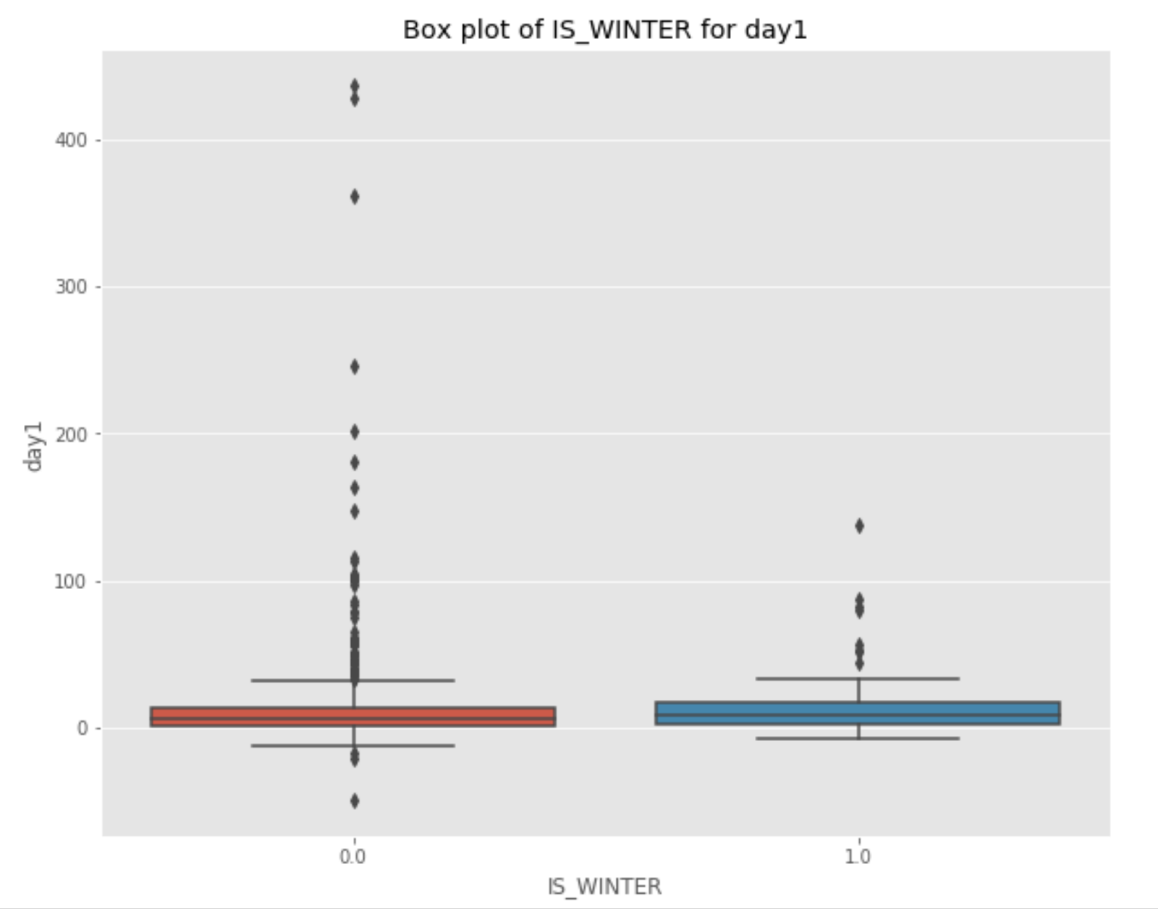
本次主要针对CarreFour进行了时间序列预测，将数据By Day进行聚合，将目标值和特征变量分别进行整合，共有**558**天数据，选取的特征共有**70**列，具体对数据的分析如下：

目标值变化：



可以发现目标值有一定的离群点即异常值，不过因为需要进行连续的时间序列分析，所以对异常值进行了保留。

特征变量：



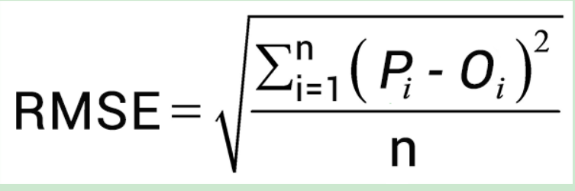
可以发现特征变量is\_winter对目标值的整体区分不大，但对0只有较多的异常值，也能够从一定角度来说，is\_winnter为0时，有较大的可能出现较大的销量。

因时间原因，无法对大部分特征变量对目标值的影响进行分析。

1. 建模结果：

本次对数据进行数据预处理，如standard，Normalization等， 并利用了机器学习算法和深度学习算法分别进行实现，具体结果如下：

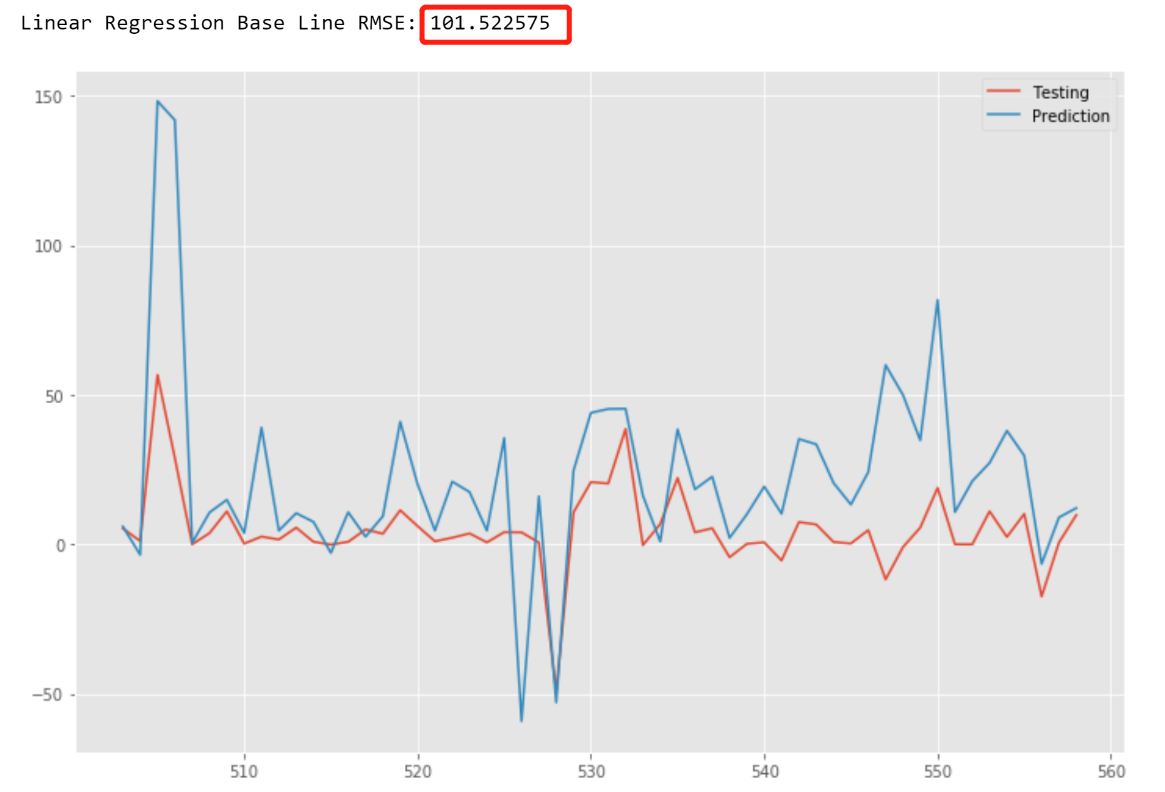
机器学习需要有目标函数，本次为回归问题，所以目标函数为RMSE, 公式如下：



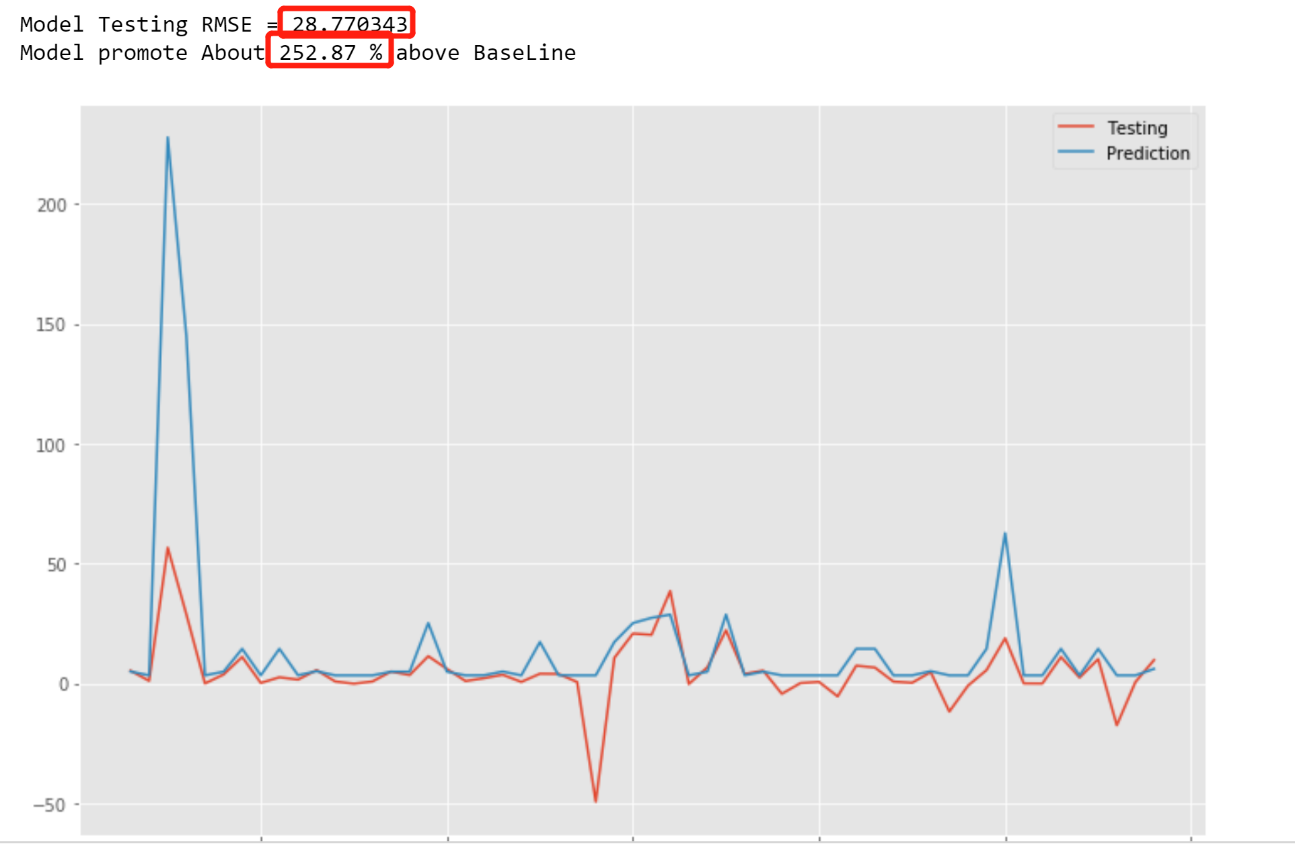
目标为最小化RMSE,其中pi为预测值，oi为真实值。

1. BaseLine

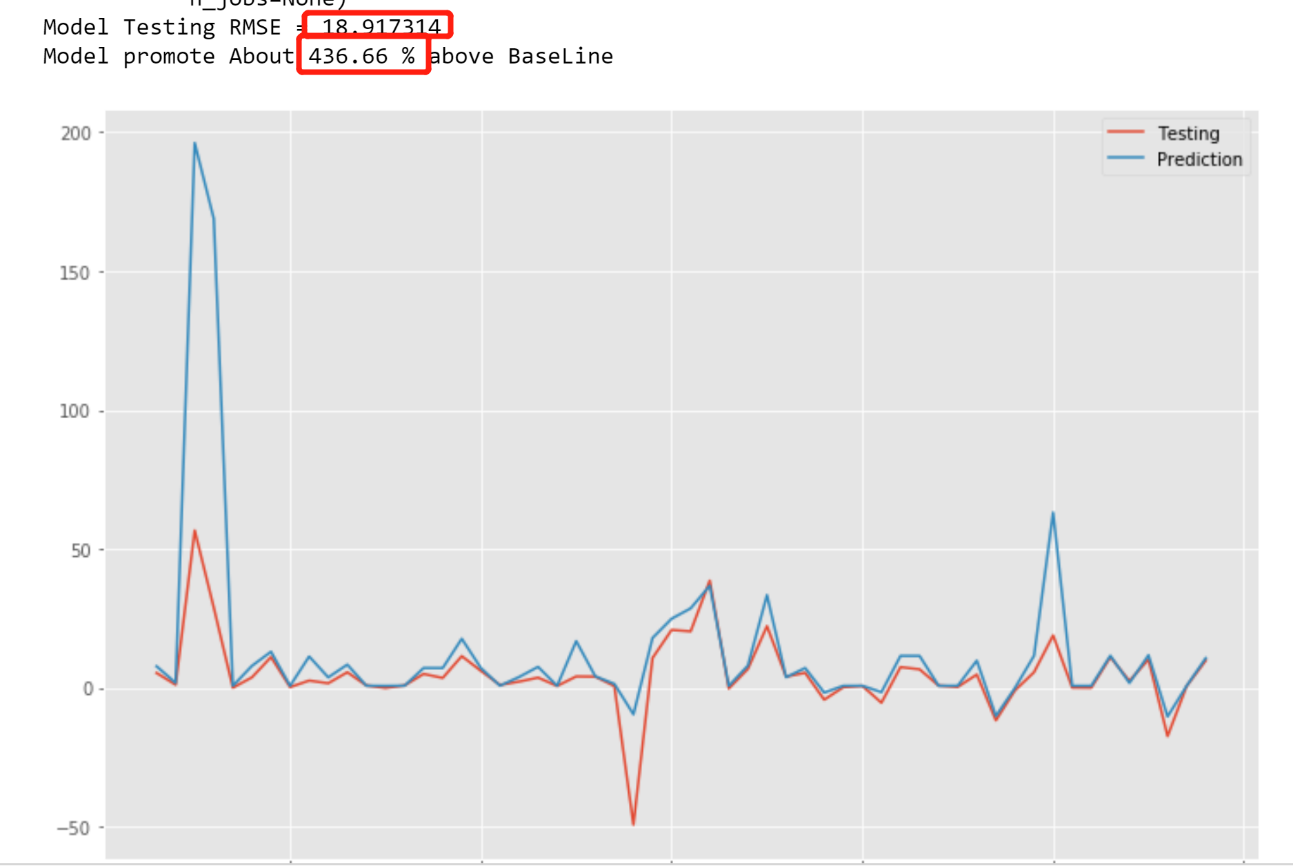
首先利用LinearRegression完成BaseLine，具体结果如下：



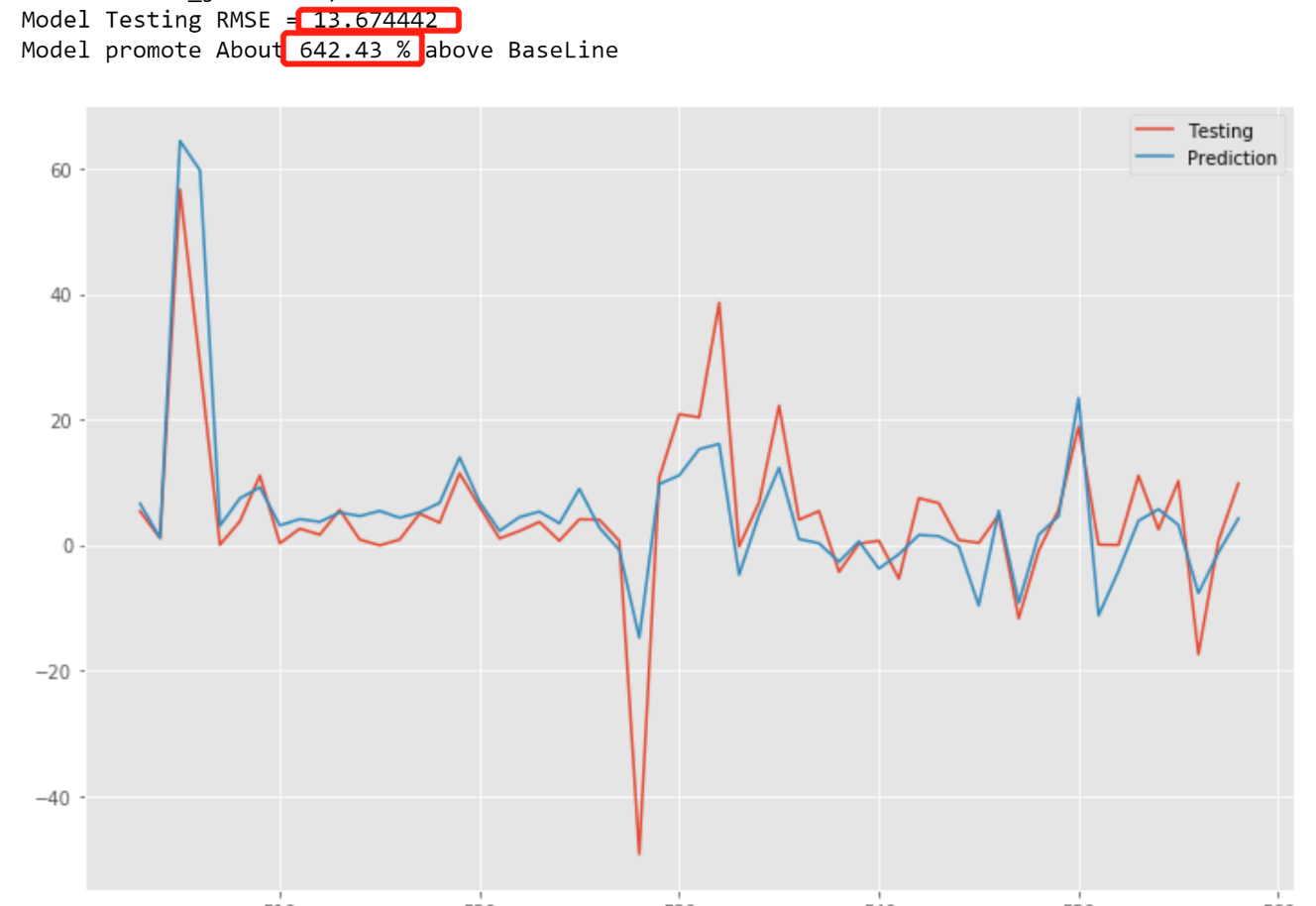
1. Adaboost algorithm



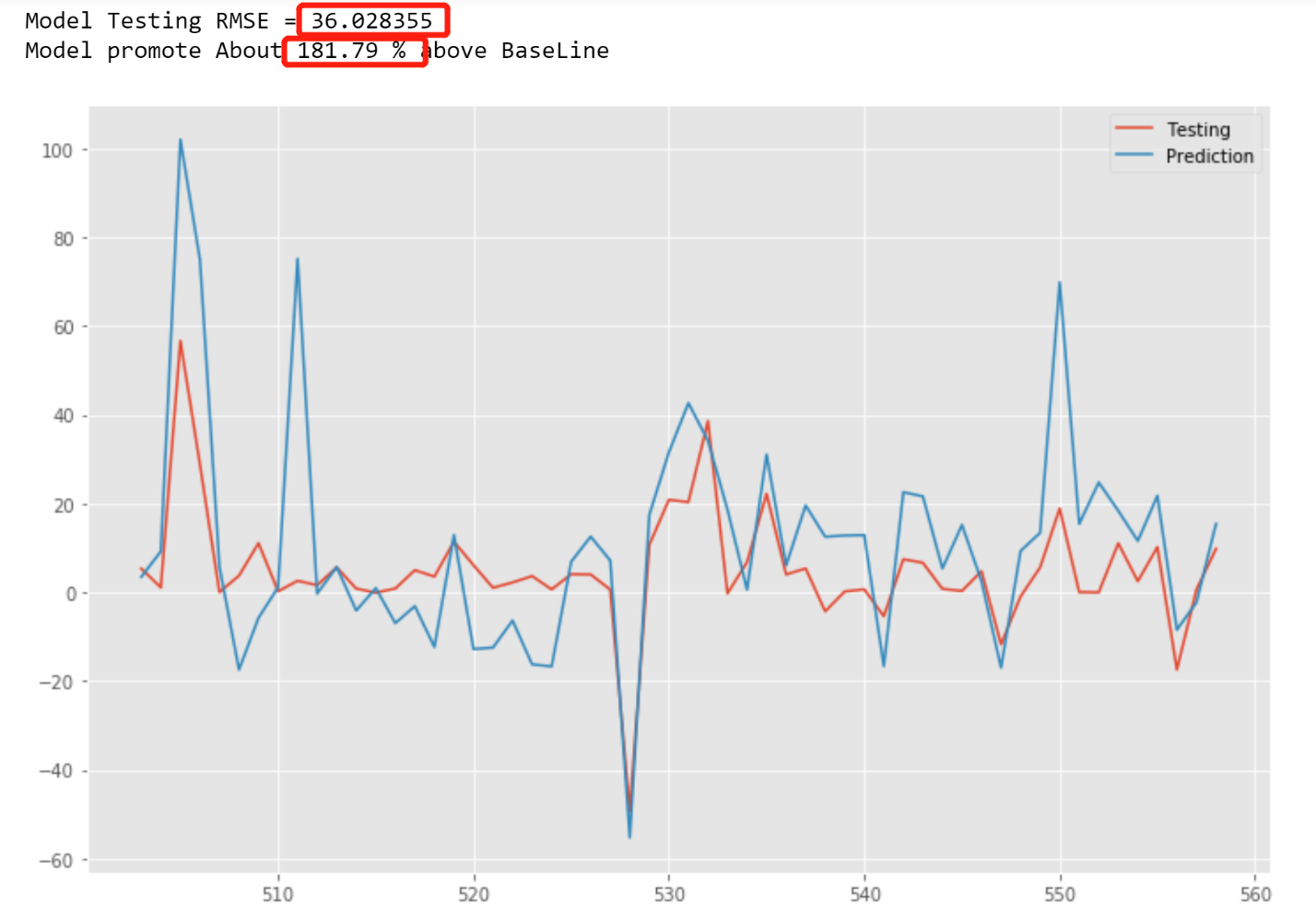
1. RandomForestRegression Algorithm



1. SVR Algorithm



1. Deep Learning algorithm



1. 总结

可以看出利用机器学习算法对比深度学习算法，对CarreFour数据集表现较好，如从目标值优化情况看，SVR RMSE可以达到13.67，超BaseLine约6.4倍！相对全部数据的结果如下图，可以发现对整体的趋势预测较优，预测准确：

