**Rossmann 店铺销售预测**

**项目背景：**

Rossmann在7个欧洲国家运营有超过3000家门店. 本项目的目的是预测Rossmann所有门店在未来6周的销售额，根据他们现在所拥有的数据及市场数据进行数据分析。一个稳定且准确率高的预测分析，能帮助门店经理合理的调配资源包括货物及人力，从而将精力放在提高用户体验及满意度上面来提升品牌价值。通过数据预测数据值，本是一个回归预测任务，用到的模型包括Linear Regression Models, Support Vector Machines(SVM), Regression Tree, Gaussain Process Regression Models, Ensembles of Trees等回归预测模型。

**问题描述：**

本项目需要进行对未来6个月Rossman各门店的销售数据预测，通过日期, 门店性质(门店类型，竞品性质, 门店推广参数)及节假日因素来对数据进行回归预测。是一个supervised regression 的问题。通过回归模型可以通过Mean Absolute Error, Mean Square Error 或者 R2 Score的回归评测方法来判定预测值与真实值之间的差距从而得出该回归模型的表现情况.

**数据集及输入描述:**

本项目数据来自Kaggle的的竞赛项目Rossmann Store Sales. 分别有train.csv, test.csv 还有 stroe.csv. 其中 train.csv 里面包含了各个店铺在过去三年的销售的额及客户总体数，是否店铺正常营业，店铺是否在做促销活动及是否公众假日，和学校假日数据。而test.csv测试集。训练数据有1017209条销售数据，测试集有 41088测试数据，占训练数据4%。店铺数据store.cvs有1115家店铺属性数据其中包括竞争对手属性，包括开业时间及与本店铺距离，是否参与促销活动，促销活动频次及商店总类。一般地，销售数据会受到促销，商店总类，及节假日等因素影响，另外还会受到大经济环境及外部消费数据等因素所影响。