# 机器学习工程师纳米学位

## 开题报告

优达学城

2018-12-26

## 预测Rossmann未来的销售额

### 选题背景

随着经济全球化的发展，企业面临着更加复杂和残酷的市场竞争。能够快速准确的预测出来销售额从而合理的安排生产和库存，用低成本的产品快速满足客户要求成为企业关心的重点。传统的销售预测方法分为定性和定量两类，定性方法主要有市场调研、购买者期望分析、专家小组法等，定量方法主要有平均数趋势预测、因果预测分析、时间序列分析法等统计方法。随着大数据和人工智能技术的兴起，机器学习模型给销售额的预测带来了新的思路。

### 问题描述

问题源自Kaggle竞赛，为欧洲的一家连锁药店Rossmann预测未来的销售情况。Rossmann在欧洲的7个国家拥有3000多家连锁药店。需要帮助他们的管理者，对位于德国的1115家药店提前6周预测日销售额。可靠的预测值能够帮助他们制定有效的员工时间表，从而提高生产效率和积极性。

### 数据说明

Kaggle提供的数据集有三个，包含1115家店铺的基本信息（店铺类型、品类、竞争对手的距离及开业时间、是否连续促销及促销时间）、训练数据集（店铺编号、日期、星期数、当日销售额、客户数、开业状态、假期状态等），测试数据集（店铺编号、星期数、日期、开业状态、促销状态、假期状态等）。需要我们根据训练数据集和店铺的基本信息情况，预测出测试数据集中店铺在给出的日期和促销状态下的销售额。

### 解决方案

### 基准模型

### 评估指标

### 方案设计

第一步：识别问题，第二步：分离数据，第三步：构造提取特征，第四步：组合数据，第五步：分解，第六步：选择特征，第七步：选择算法进行训练