# 价格预测与提前购买天数

## 术语解释

**价格预测**：是指对某天两地之间的机票的普通舱的最低票价的预测

**提前购买天数**：承上，指对某天两地之间的机票的普通舱的达到最低票价的日期预测，也就是购买日期推荐与实际出发日期之差值，天数为单位

## 数据准备

**初始数据**：日期,出发时间,始发城市,目的城市,最低票价,提前天数

**处理方式**：日期字符串转换为三个int值(年,月,日)，出发时间字符串转换为三个int值(时,分,秒)，并分局日期计算出当日是周几；使用平均值填充缺失值； 始发城市跟目的城市进行热编码；最低票数与提前天数均准转换为float型。

**Feature**：年,月,日,时,分,周几, 始发城市,目的城市

(以上除了提前天数之外的其他时间都是指起飞时间)

**预测变量**：最低票价，提前天数

## 模型生成

采用随机森林回归。

**n\_estimators=10：使用了10课子树**

**max\_features=7: 最多使用7个特征**

**max\_depth=10: 最大深度为10，防止过深过拟合。**

使用以上**feature**数据，并分别使用**最低票价**与**提前天数**作为训练标签，训练出两个随机森林模型。

训练后的模型使用Pickle序列化存于文件

## 对外接口

**Python与java调用方式**

通过web service的方式将模型预测接口开放给java EE平台。使用python内置的BaseHTTPRequestHandler搭配HttpServer实现简单的webService服务器。

**访问方法**为：

get

**访问格式**为：

http://ip:port/y-m-d/h:M:s/departCity/destinationCity

如：

<http://115.29.138.254:8088/2017-07-01/20:45:00/BJS/SHA>

**返回格式**为：

{'days': 提前天数, 'result': 成功或失败, 'price': 预测最低价格}

如：

{'days': 11.0, 'result': True, 'price': 1156.7741569507343}