**合作意向信息表**

|  |  |  |  |
| --- | --- | --- | --- |
| 基  本  信  息 | 姓名或英文名 | 闫思明 |  |
|  | 联系方式 | 18321028235 |  |
| 特长 | 擅长软件 | R语言、Python、SQL、Tableau | |
|  | 擅长领域 | 数据采集，数据分析，机器学习 | |
| 代表性的项目介绍 | 项目名称 | 印度二手车市场价格分析 | |
|  | 项目挑战 | 比较多种机器学习方法优劣性，分析二手车价格影响因素，训练模型预测二手车价格。 | |
|  | 解决方案 | **任务/目标**  根据印度二手车交易市场1996-2019年数据，进行清洗，建模，预测。  **数据源准备**  7253笔交易数据包括汽车属性和交易日期、地点等信息。分析数据构成：  将数据分为NA和非NA组，分析缺失值是否均匀分布：    对于的因变量“交易价格”，可见其缺失值基本均匀分布。    其他自变量的缺失值也基本均匀分布。  **特征转换**  对一些因变量进行dummy variable转换。对大数值变量如引擎容量，已行驶的公里数进行log transformation。  **划分训练集和测试集**  75% training data, 25 test data. RMSE作为衡量模型精度的标准。  **建模**  **10 folds Validation when training models** to choose best model tuning parameters**.**   |  |  | | --- | --- | |  |  |  1. **Linear Regression with mixing Lasso & Ridge Penalty:**       包含三种模型的混合预测。    Best tune: Alpha 0.25, lambda .053. RMSE 5.332   1. **Support Vector Machines with Radial Basis Function Kernel**     Best tune: Cost(M) = 10.   |  |  | | --- | --- | |  |  |  1. *Random Forests:*     随着随机选定的因变量数量提高，10 folds Cross Validation所展示的拟合效果也有波折地逐渐提高。   1. *Stochastic Gradient Boosting Machine*     调整的参数为树深，树层数达到6时拟合效果最好。   |  |  | | --- | --- | |  |  |   **模型优化**   1. 各个模型都进行了调参过程。主要依据为10 folds cross validation | |
|  | 项目结果 | 在此案例中，*Stochastic Gradient Boosting Machine* 所得到的RMSE值最小，预测效果最好。  预测结果仅作为参考一个权重值，还需要专家意见，按照一定的权重来计算。  以下为预测集和测试集的部分展示： | |
| 接受的咨询方式 | 驻场项目咨询  在线项目咨询  线下一对一培训  线下多人培训  在线一对一培训  在线多人培训 | | |