

分析流程



分析步骤

1. 通过训练集数据来建立XGBoost回归模型。
2. 通过建立的XGBoost来计算特征重要性。
3. 将建立的XGBoost回归模型应用到训练、测试数据，得到模型评估结果。
4. 由于XGBoost具有随机性，每次运算的结果不一样，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。
5. 注：XGBoost无法像传统模型一样得到确定的方程，通常通过测试数据预测精度来对模型进行评价。

详细结论

输出结果1：模型参数

复制

参数名	参数值
训练用时	0.284s
数据切分	0.7
数据洗牌	是
交叉验证	否
基学习器	gbtree
基学习器数量	100
学习率	0.1
L1正则项	0
L2正则项	1
样本征采样率	1
树特征采样率	1
节点特征采样率	1
叶子节点中样本的最小权重	0

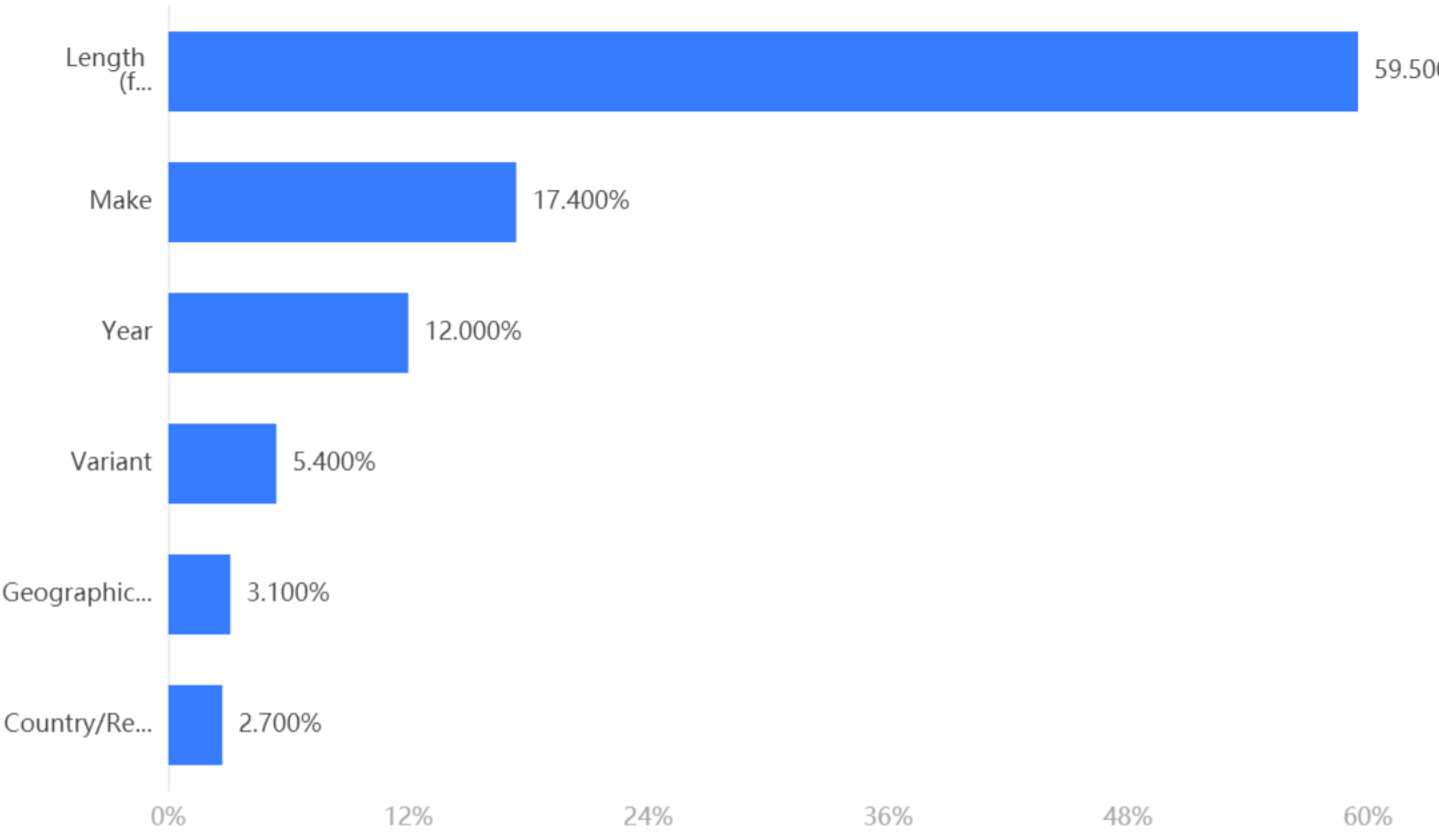
树的最大深度	10
--------	----

图表说明：

上表展示了模型各项参数配置以及模型训练时长。

输出结果2：特征重要性

柱形图



图表说明：

上柱形图或表格展示了各特征（自变量）的重要性比例。

输出结果3：模型评估结果

复制

	MSE	RMSE	MAE	MAPE	R²
训练集	188164022.972	13717.289	8169.746	4.54	0.99
测试集	7477644322.049	86473.373	42959.212	17.207	0.74

图表说明：

上表中展示了交叉验证集、训练集和测试集的预测评价指标，通过量化指标来衡量XGBoost的预测效果。其中，通过交叉验证集的评价指标可以不断调整超参数，以得到可靠稳定的模型。

- MSE（均方误差）： 预测值与实际值之差平方的期望值。取值越小，模型准确度越高。
- RMSE（均方根误差）： 为MSE的平方根，取值越小，模型准确度越高。
- MAE（平均绝对误差）： 绝对误差的平均值，能反映预测值误差的实际情况。取值越小，模型准确度越高。
- MAPE（平均绝对百分比误差）： 是 MAE 的变形，它是一个百分比值。取值越小，模型准确度越高。
- R²： 将预测值跟只使用均值的情况下相比，结果越靠近 1 模型准确度越高。

输出结果4：测试数据预测结果

下载

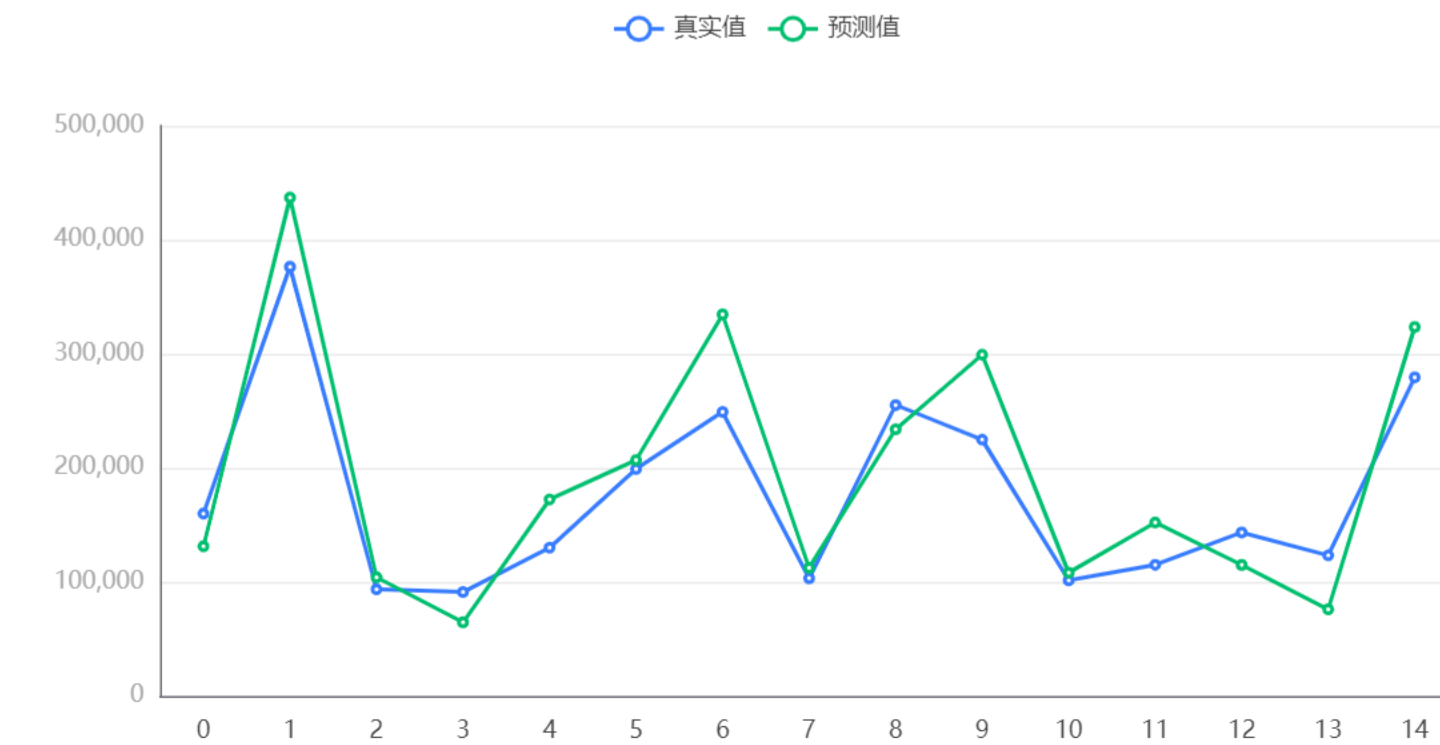
预测结果Y	Listing Price (USD)	Make	Variant	Length (ft)	Geographic Region	Country/Region/State	Year
131250.078125	159900	8	28	38	2	17	2006
437058.40625	376313	3	270	56	1	22	2014
104066.1875	93519	22	160	42	1	2	2008
64569.4765625	91101	2	2	38	1	3	2005
172353.328125	129999	3	359	40	3	36	2018
206825.734375	199000	8	38	38	2	15	2012
334555.125	249013	22	276	50	1	1	2014
112291.9140625	103253	2	7	46	1	3	2005
233835.125	255052	2	205	54	1	22	2010
299222.15625	224719	3	245	49	1	22	2013
108069.0859375	101427	3	12	39	1	10	2005
152107.296875	115000	3	306	38	3	61	2016
114943.3984375	143334	22	256	40	1	4	2011
75975.5625	123292	2	4	39	1	2	2007
323455.9375	279380	2	205	54	1	6	2012

图表说明：

上表格为预览结果，只显示部分数据，全部数据请点击[下载按钮](#)导出。

上表展示了XGBoost对测试数据的预测情况。

输出结果5：测试数据预测图





图表说明：

图表说明：上图中展示了XGBoost对测试数据的预测情况。

输出结果6：模型预测与应用

请选择文件所在路径

模型预测 ?

☐ 数据是否包括实际因变量值Y

图表说明：

- 系统会自动保存模型，需要注意的是：在机器学习中的XGBoost算法保存的模型是非常复杂的，不是类似于线性回归那样可以用一个公式保存，系统以二进制文件方式进行序列化保存。
- 由于XGBoost具有随机性，每次训练的模型可能不一致，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测。
- 若删除本分析报告将会直接删除模型的缓存。

参考文献

- [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from <https://www.spsspro.com>.
- [2] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[J]. ACM, 2016.