**2022-2023学年度第一学期**

**华南师范大学**

**人工智能学院**

**《机器学习课程项目》报告**

|  |  |
| --- | --- |
| **题目：** | 随机森林算法预测车辆接受度 |
| **班级：** | 人工智能一班 |
| **学号：** | 20214001052 |
| **姓名：** | 凌绍彬 |
| **指导老师：** | 刘海 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **总分** | **指导老师评分** | **备注** |
| **一、报告文档** | **30** |  |  |
| 1内容基本要点 | 20 |  |  |
| 2格式 | 10 |  |  |
| **二、系统验收** | **50** |  |  |
| 1功能 | 30 |  |  |
| 2程序结构、编码规范 | 10 |  |  |
| 3代码正确性 | 10 |  |  |
| **三、作品展示** | **20** |  |  |
| 1功能亮点 | 10 |  |  |
| 2专业表达 | 10 |  |  |
| **总评分** | |  |  |

目 录

目录

[2022-2023学年度第一学期 1](#_Toc3734)

[《机器学习课程项目》报告 1](#_Toc8236)

[1课题介绍 2](#_Toc2709)

[2 主要任务 2](#_Toc6282)

[3 实现组件 2](#_Toc18189)

[4 实现效果 3](#_Toc12942)

[5 项目收获 3](#_Toc19039)

[参考资料 4](#_Toc30111)

**1课题介绍**

车辆的研发投入巨大，汽车制造商需要对车辆的各个功能进行重要性评估，以合理运用经费。对于消费者而言，在众多车款中初步筛选出满意度高的汽车也不是简单的事情。

由此，本课题利用随机森林算法对车辆满意度数据集进行训练和分析，得出车辆满意度的预测工具和车辆满意度的重要因素。

**2 主要任务**

为了训练出优秀的模型以及合理探索数据，我做了以下工作：

1. 训练模型时，把树的棵数从1遍历到100，挑选出表现最优的随机森林。

2. 利用最优的随机森林模型，计算数据集中样本的各个因素的重要程度，进而分析厂商应当重视的车辆因素。

**3 实现组件**

本课题采用了常见的Python包实现各个功能：

1. 数据可视化：seaborn、matplotlib。

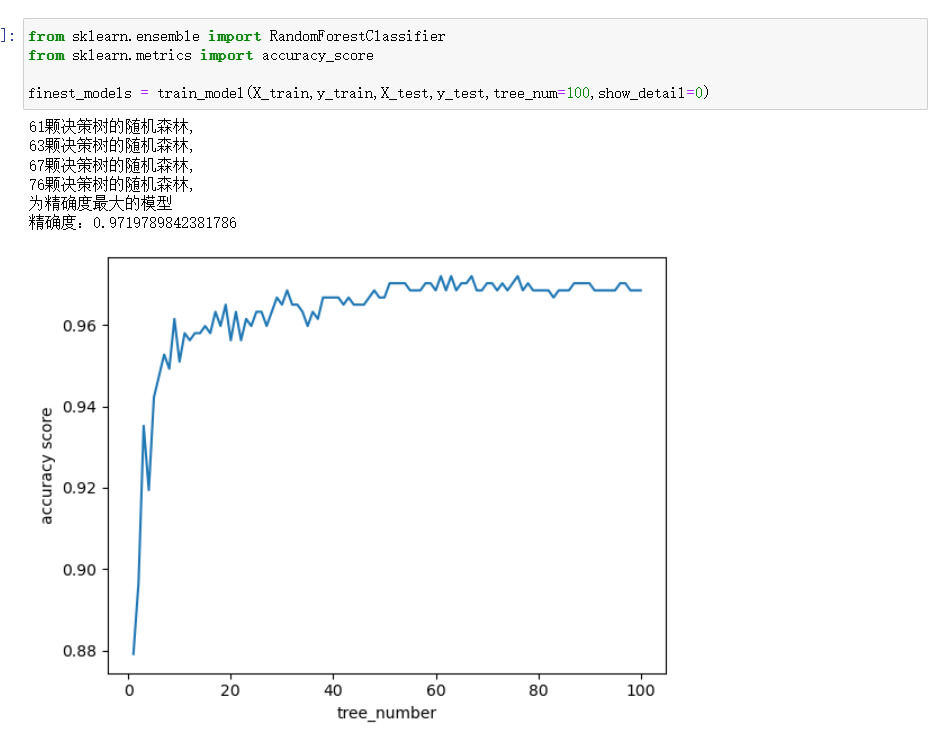
2. 机器学习模型：scikit-learn。

3. 数据处理：numpy、pandas、category\_encoders。

主要实现内容请参见jupyter-notebook的内容。

**4 实现效果**

实现了精确度为0.97的随机森林模型



**5 项目收获**

我感受到传统机器学习方法的魅力。对比现在热度很高的深度学习，传统机器学习的复现程度、可解释性比神经网络好很多。通过完整的走过一边随机森林算法的搭建和分析，我感受到机器学习的一些趣味，也提醒我不能一味地去追热点技术而忘记了这个学科最重要的，最本质的课题：机器学习。通过现实生活中的数据，用数学工具进行分析,制作出一个有用的工具。

我更深入地了解了一个机器学习项目的流程。从算法的理论部分到模型的搭建，我感觉python丰富的第三方库，极大简化了数据处理和模型搭建的代码。结合着数据集，感受到机器学习算法的实用意义。

**参考资料**

[1] https://www.kaggle.com/code/prashant111/random-forest-classifier-tutorial