



湖南工程學院

# 毕业设计（论文）

题目： 基于机器学习及车联网数据的  
车辆碰撞识别预警系统

学院： 计算科学与电子学院

专业： 数据科学与大数据技术

班级： 大数据 2101 班 学号： 2021\*\*\*\*\*

学生姓名： 刘 \*\*

导师姓名： 文 \*\*

完成日期： 2025 年 6 月 2 日

# 诚信声明

本人声明：

1、本人所呈交的毕业设计（论文）是在老师指导下进行的研究工作及取得的研究成果；

2、据查证，除了文中特别加以标注和致谢的地方外，毕业设计（论文）中不包含其他人已经公开发表过的研究成果，也不包含为获得其他教育机构的学位而使用过的材料；

3、我承诺，本人提交的毕业设计（论文）中的所有内容均真实、可信。

作者签名：

日期： 年 月 日

## 目 录

摘要 .....	I
Abstract .....	II
第 1 章 绪论 .....	1
1.1 引言 .....	1
1.2 国内外研究现状 .....	1
1.2.1 国内研究现状 .....	1
第 2 章 预备知识 .....	2
2.1 特征编码 .....	2
2.2 重采样 .....	2
2.3 分类模型评估方法 .....	2
第 3 章 数据预处理 .....	3
3.1 数据探索及清洗 .....	3
3.1.1 异常标签探索及清洗 .....	3
3.1.2 缺失值及异常值探索及清洗 .....	3
参考文献 .....	5
致谢 .....	6

# 基于机器学习及车联网数据的 车辆碰撞识别预警系统

**摘要：**随着新能源电动汽车的飞速发展，各行业与用户对于低成本高效的车辆碰撞识别系统的需求越来越高。然而...

**关键词：**数据探索及清洗；特征工程；贪心随机网格搜索；随机森林；GBDT

# Vehicle Collision Detection and Warning System based on Machine Learning and Vehicle-to-Everything Data

**Abstract:** With the rapid development of new energy electric vehicles, there is an increasing demand across industries and among users for low-cost, high-efficiency vehicle collision detection systems. However ...

**Keywords:** Data exploration and cleaning; Feature engineering; Greedy randomized search; Random forest; GBDT

# 第 1 章 绪论

## 1.1 引言

为贯彻落实新发展理念、科技强国，国家大力推动新能源汽车的发展，电动汽车在国内占比逐渐增大。用户、汽车主机厂、保险等各行各业对于低成本、高效的新能源电动汽车的碰撞预警识别系统的需求越来越多。

然而...

## 1.2 国内外研究现状

### 1.2.1 国内研究现状

民用汽车从 20 世纪开始量产到如今已经成为现代社会重要的交通工具<sup>[1-2]</sup>。截至 2024 年全国机动车保有量达 4.53 亿辆，其中汽车 3.53 亿辆，汽车产业成为国家经济系统的重要组成部分<sup>[3]</sup>。

在智能化...

## 第 2 章 预备知识

### 2.1 特征编码

非纯数字类型的定性特征字段可大致分为定类特征和定序特征两类，定类特征表明该记录所属类别，定序特征表明该记录在某指标下的某种程度，定类特征和定序特征常用的编码方式有两种，一种是直接用自然数进行编码的常规编码方式，另一种是独热指标（One-Hot Encoding）方式<sup>[4]</sup>。

... 原理如图2.1所示：

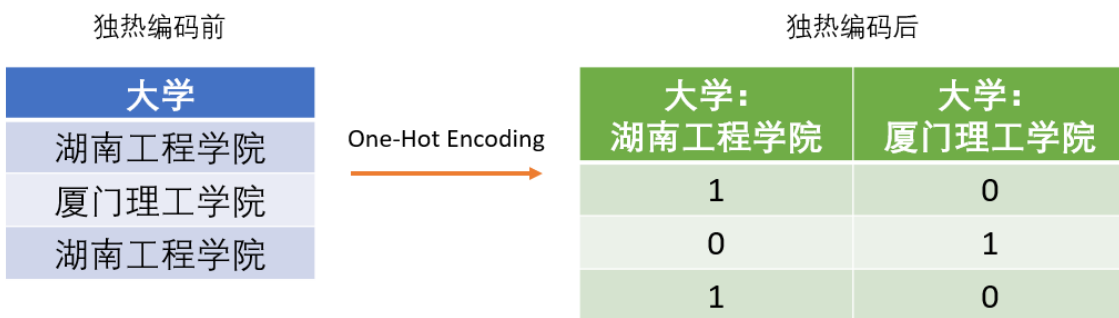


图 2.1 独热编码原理图

### 2.2 重采样

### 2.3 分类模型评估方法

...F1 分数公式如下：

$$F1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.1)$$

## 第3章 数据预处理

### 3.1 数据探索及清洗

#### 3.1.1 异常标签探索及清洗

#### 3.1.2 缺失值及异常值探索及清洗

... 其中各个特征的缺失值数量如表3.1所示。

表 3.1 数据各特征缺失值数

特征名	缺失值数
加速踏板位置	173443
驾驶员需求扭矩值	173321
电池包主负继电器状态	173321
电池包主正继电器状态	173321
制动踏板状态	173321
驾驶员离开提示	173321
主驾驶座占用状态	173420
驾驶员安全带状态	173321
手刹状态	173373
整车钥匙状态	173373
整车当前档位状态	173472
整车当前总电流	173373
整车当前总电压	173321
车辆行驶里程	173321
车速	173321
方向盘转角	173321

本文的核心 Pipeline 框架如下：

代码 3.1 基于 Scikit-learn 的 Pipeline 流程示例

```
from sklearn.compose import ColumnTransformer
```

```
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.pipeline import Pipeline

numeric_transformer = Pipeline(steps=[
    ('imputer', SimpleImputer(strategy='median')),
    ('scaler', StandardScaler())
])
```

## 参考文献

- [1] Xiong X, Chen L, Liang J. A new framework of vehicle collision prediction by combining SVM and HMM[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 19(3): 699-710.
- [2] Mukhtar A, Xia L, Tang T B. Vehicle detection techniques for collision avoidance systems: A review[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2318-2338.
- [3] 公安部. 公安部发布 2024 年度全国机动车、驾驶人数据[J]. 道路交通管理, 2025(01): 9.
- [4] 杨惠, 李文, 郭若礁, et al. 工程造价 SVM 预测的特征分析及编码研究[J]. 内蒙古科技与经济, 2022(18): 114-116+119.
- [5] Liu Xurong. HNIE-thesis-latex-template[EB/OL]. (2026) [2026-05-05]. <https://github.com/XurongLiu/HNIE-Thesis-LaTeX-Template>.

## 致 谢

本论文的完成离不开开源项目<sup>[5]</sup>的支持，在此感谢项目作者 **XurongLiu** 的无私奉献。

...