电子科技大学信息与软件工程学院

**项 目 报 告**

课程名称 大数据分析与智能计算

理论教师

实验教师

学生信息：

|  |  |  |
| --- | --- | --- |
| 序号 | 学号 | 姓名 |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |
| 5 |  |  |

**电子科技大学教务处制表**

**电 子 科 技 大 学**

**项 目 报 告**

**指导教师： 地点：**

1. **项目名称：航空公司****延误和取消分析项目**
2. **项目时间：2022.12.2—2022.1.2**
3. **项目原理**

# 四、 项目内容

# 五、 需求分析与设计

（需求分析描述、用例图、用例规约、数据流图、顺序图、流程图）

# 六、项目计划

# 七、项目环境配置管理

**7.1** **操作系统**：

**7.2 开发工具**：

**7.3 配置过程**

# 八、项目实践过程

（代码示例格式如下：）

建立特征索引，其作用是XXXXXXX，代码如下所示。

代码1 建立特征索引

|  |
| --- |
| # 建立特征索引  from pyspark.ml.feature import VectorIndexer  featureIndexer = VectorIndexer(inputCol="features", outputCol="indexedFeatures", maxCategories=24).fit(dataFrame) |

**九、项目结果与分析（含重要数据结果分析或核心代码流程分析）**

**1. 查看飞机延误时间最长的前10名航班。**

**2. 计算延误的和没有延误的航空公司的比例。**

**3. 分析****一天中延误最严重的飞行时间。**

**4. 分析一周中延误最严重的飞行时间。**

**5. 短途航班和长途航班，哪种航班取消更严重？**

**6. 建立机器学习算法模型，预测未来航班取消情况。**

# 十一、总结及心得体会

# 十二、对本项目过程及方法、手段的改进建议