# 华为AI日志分析模型项目进展报告

## 1. 第一版模型实现思路

第一版模型旨在对华为AI日志数据针对华为提供的动态日志中提取的列表项数据进行误报分类识别，整体流程包括数据预处理、特征提取、模型训练与评估，以及模型推理功能的实现。主要步骤如下：  
1) 数据预处理：  
 - 使用OneHotEncoder对结构化字段（oracle\_name、sut.component、sut.component\_set、sut.module）进行独热编码。  
 - 使用CodeBERT对文本字段（api\_ut、tags）进行语义向量化编码（768维）。  
2) 特征融合：  
 - 将OneHot编码向量与CodeBERT文本向量拼接成完整特征向量。  
3) 数据集处理：  
 - 划分训练集与测试集，使用RandomOverSampler进行上采样以缓解类别不平衡问题。  
4) 模型结构：  
 - 使用两层全连接神经网络（输入层→隐藏层ReLU→输出层）进行二分类。  
5) 训练与评估：  
 - 使用加权交叉熵损失函数（CrossEntropyLoss）平衡类别权重。  
 - 训练过程中记录Loss、Accuracy、Precision、Recall、F1等指标，并使用TensorBoard可视化。  
6) 推理功能：  
 - 提供推理脚本infer.py，支持加载训练时的编码器（encoder.pkl）与模型权重（log\_classifier.pt）对新数据进行预测。

## 2. 项目当前进展

目前已完成：  
- 第一版模型的设计、训练与测试，且推理脚本已实现。  
- 模型在训练集上的指标：Acc=82.24%，Precision=0.84，Recall=0.80，F1=0.82。  
- 模型在测试集上的指标：Acc=81.3%，Precision=0.99，Recall=0.81，F1=0.89。  
- 已保存OneHot编码器（encoder.pkl）、编码列顺序（encoder\_columns.npy）及模型权重（log\_classifier.pt）。  
- 数据输入格式与华为提供的原始日志提取的列表项格式一致。

## 3. 潜在风险与应对措施

潜在风险：  
- 训练数据与实际部署流水线数据分布可能存在较大差异，影响模型泛化性能。  
  
应对方案：  
1) 第二版模型将引入静态分析的警报信息与代码信息，通过主动学习逐步优化模型性能。  
2) 根据上线运行情况进行数据反馈闭环，持续微调模型参数与特征工程。  
3) 在特征提取阶段增加调用栈相关的代码信息，以提升模型在复杂场景下的判别能力。

## 4. 后续数据需求

第二轮迭代模型需要补充的数据包括：  
1) 完整的动态日志。  
2) 报错代码调用链上的完整代码（包括触发错误的代码）。  
3) 报错代码的静态分析结果，包括警报类型、警报优先级、严重程度等多有静态分析工具输出的信息。  
说明：有了调用栈信息后，可以在日志中保留相关函数的运行信息，从而提取更多有价值的运行时特征。