# Malicious Document Detection and Robust ML Model Construction恶意文档检测与强鲁棒性ML模型的构建

**摘要：**随着信息安全技术的迅速发展，对恶意文档检测变得越来越重要，但是，由于文档结构的多样性，使攻击者可以获得更多的攻击机空间，在本文中，我们旨在工业界和学术界中构建一个人工智能（AI）文档分类器。我们一共收集了200,000个样本用于训练AI模型与优化。我们的实验结果表明，经过多次训练与调优，我们的模型准确度高达99.82%，并且误报率仅低于0.01%。此外，通过对抗性学习的研究，该模型有一定的抗逃逸性与良好的鲁棒性。最后，我们展示了模型可以广泛的部署在实际的应用场景中，例如安全产品和终端安全中。

实验部分

## 构建基于机器学习的文档分类器

在这个章节中，我们主要集中在对设计分类器，一种将数据分类成两种预定义的学习系统。分类器的预测是通过计算一些数学或者概率的分数或者计算特定的阈值，我们主要关注于分类器的以下几个方面：

A 通用机器学习分类框架

B 数据集

C 特征工程

D 分类算法选择

E 模型优化

### 通用机器学习分类框架

通用机器学习的分类器框架如图1所示，我们的目标是训练一个鲁棒性的模型用于恶意文档分类，首先，在数据收集阶段，我们需要收集到大量的恶意文件和良性文件