Vulnerability Assessment Report

Company: Swajyot Technologies Pvt Ltd

Project: Machine Learning Model Vulnerability Assessment

Model Name: image_classifier_v2

Report Date: June 18, 2025

Executive Summary

This report outlines the findings from a vulnerability assessment conducted on the machine learning model file `image_classifier_v2`. The assessment was designed to identify security risks such as exposed credentials, model tampering risks, insecure dependencies, and susceptibility to adversarial attacks. This report was prepared by a third-party cybersecurity firm with experience in machine learning system audits.

Assessment Methodology

The model was analyzed using a combination of static code analysis, dependency inspection, model introspection, and simulation of adversarial attack scenarios. Key components of the assessment included:

- Hash integrity verification
- Dependency vulnerability scanning
- API exposure audit
- Metadata and comments inspection
- Adversarial robustness testing

Detailed Findings

1. Insecure Dependency Detected

The model relies on a vulnerable version of the `tensorflow` package (v2.4.0), which has multiple known vulnerabilities including remote code execution (CVE-2021-3177).

2. Embedded Secrets

Hardcoded AWS credentials were found in the model metadata comments. Exposure of these credentials can lead to unauthorized access to cloud storage or compute resources.

3. Susceptibility to Adversarial Attacks

The model shows high susceptibility to fast gradient sign method (FGSM) adversarial examples, leading to misclassification with confidence over 90%.

4. Unverified Model Source

The model file lacks cryptographic signatures to verify its integrity and origin, increasing the risk of tampering or replacement by malicious actors.

Impact Analysis

These vulnerabilities, if exploited, could lead to system compromise, data leakage, or undermining of the model's reliability in production environments. For applications in critical sectors such as healthcare or finance, this poses a significant operational and reputational risk.

Recommendations

- Upgrade all ML dependencies to the latest secure versions.
- Remove all embedded secrets from model files.
- Implement model signing and validation mechanisms.
- Retrain and harden model against common adversarial attack vectors.
- Conduct regular vulnerability scans during the ML lifecycle.

Disclaimer

This report is based on a simulated vulnerability assessment. The data and findings are fictional and intended for demonstration purposes only.