**Software Requirements Specification (SRS)**

**Project Name: Sensitive Information Detection System (SIDS)**

**Revision History**

| **Version** | **Date of Release** | **Pages Affected** | **Reasons for Change** | **Signature** |
| --- | --- | --- | --- | --- |
| 1.0 | YYYY-MM-DD | All | First Release |  |

**Table of Contents**

1.0 Project Code  
2.0 Title of the Project  
3.0 Introduction  
3.1 Purpose  
3.2 Document Conventions  
3.3 Intended Audience and Reading Suggestions  
3.4 Project/Product Scope  
3.5 References  
4.0 Overall Description  
4.1 Project/Product Perspective  
4.2 Project/Product Functions  
4.3 User Classes and Characteristics  
4.4 Operating Environment  
4.5 Design and Implementation Constraints  
4.6 User Documentation  
4.7 Assumptions and Dependencies  
5.0 External Interface Requirements  
5.1 User Interfaces  
5.2 Hardware Interfaces  
5.3 Software Interfaces  
5.4 Communications Interfaces  
6.0 System Features  
7.0 Other Nonfunctional Requirements  
7.1 Performance Requirements  
7.2 Safety Requirements  
7.3 Security Requirements  
7.4 Software Quality Attributes  
7.5 Other Requirements  
8.0 Acceptance Criteria  
9.0 Deliverables

**1.0 Project Code**

SIDS-007

**2.0 Title of the Project**

Sensitive Information Detection System (SIDS)

**3.0 Introduction**

**3.1 Purpose**

The purpose of this project is to develop a machine learning system to automatically detect and flag sensitive information within banking datasets. This system aims to enhance data security and compliance with data protection regulations by identifying potentially sensitive data before it can be accessed or processed further.

**3.2 Document Conventions**

None

**3.3 Intended Audience and Reading Suggestions**

This document is intended for the development team, project managers, business analysts, and stakeholders such as banking compliance officers and IT security personnel.

**3.4 Project/Product Scope**

The scope of the project includes the design, development, testing, and deployment of a machine learning-based sensitive information detection system. The system will be used internally within banking institutions to secure sensitive data and ensure regulatory compliance.

* 1. **References**

[[adobe link](https://blog.developer.adobe.com/using-machine-learning-to-help-detect-sensitive-information-5bfb32eeb34e)]

<https://blog.developer.adobe.com/using-machine-learning-to-help-detect-sensitive-information-5bfb32eeb34e>

**4.0 Overall Description**

**4.1 Project/Product Perspective**

The Sensitive Information Detection System (SIDS) is a standalone application that integrates with existing banking data management systems. It will leverage machine learning algorithms to analyze and detect sensitive information within large datasets.

**4.2 Project/Product Functions**

* Data ingestion from various banking sources
* Preprocessing of data for machine learning analysis
* Training and deployment of machine learning models
* Real-time detection and flagging of sensitive information
* Reporting and visualization of detected sensitive data
* Integration with existing data management and compliance systems

**4.3 User Classes and Characteristics**

* **Data Analysts:** Users responsible for managing and analyzing banking datasets.
* **Compliance Officers:** Users responsible for ensuring data security and regulatory compliance.
* **IT Security Personnel:** Users responsible for maintaining the security of banking systems and data.

**4.4 Operating Environment**

**4.4.1 Server**

* Language: Python
* Database: MySql
* Platform: Windows, Jupiter Notebook
* Tools: Scikit-Learn
* Data Visualization: Matplotlib , Seaborn , SciPy

**4.4.2 Client**

* Browser-based clients either in Microsoft Windows or Linux Environment.
* Android-based mobile client.

**4.5 Design and Implementation Constraints**

* Must comply with banking data security regulations.
* Integration with existing banking data management systems.
* High performance and scalability to handle large datasets.

**4.6 User Documentation**

User manuals and online help documents will be delivered along with the software. Installation instructions will be available in the Installation manual.

**4.7 Assumptions and Dependencies**

* Availability of labeled datasets for training machine learning models.
* Access to existing banking data management systems for integration.

**5.0 External Interface Requirements**

**5.1 User Interfaces**

The targeted browsers include:

* Mozilla Firefox 12.0
* Google Chrome
* Microsoft Edge

**5.2 Hardware Interfaces**

The proposed system configuration is as follows:

| **Item** | **Server 1** | **Server 2** |
| --- | --- | --- |
| Processor | Quad-core Processor (64 bit) | Quad-core Processor (64 bit) |
| No. of Processor | One | One |
| Memory | 8 GB expandable to 24 GB | 8 GB expandable to 24 GB |
| HDD Capacity | 1 TB | 1 TB |
| Network | 2 Gigabit RJ45 connectors | 2 Gigabit RJ45 connectors |
| USB Ports | Four | Four |
| OS | Linux | Linux |

**5.3 Software Interfaces**

* Integration with banking data management systems
* APIs for data ingestion and processing
* Machine learning frameworks (Scikit-learn)

**5.4 Communications Interfaces**

Automated e-mail sending facility for reporting detected sensitive information.

**6.0 System Features**

. The system features include detailed functionalities with flow diagrams and use case diagrams. These are the functionalities to be completed at the end of the project. The functional document has the details of these functionalities

**7.0 Other Nonfunctional Requirements**

**7.1 Performance Requirements**

The system must process large datasets in real-time and detect sensitive information with high accuracy and minimal latency.

**7.2 Safety Requirements**

The system must ensure the confidentiality and integrity of the banking datasets.

**7.3 Security Requirements**

The system must comply with data protection regulations and include features for secure data handling and access control.

**7.4 Software Quality Attributes**

The system should be reliable, maintainable, and scalable to handle increasing volumes of data.

**7.5 Other Requirements**

* Compliance with industry standards and regulations.
* Regular updates and maintenance.

**8.0 Acceptance Criteria**

The system will be considered acceptable if it accurately detects sensitive information within banking datasets and integrates seamlessly with existing data management systems. Performance benchmarks and security tests will be conducted to ensure compliance with requirements.

**9.0 Deliverables**

* Source code and executables of the SIDS application
* User manuals and installation guides
* Training materials and documentation
* System integration and test report