**Unit 8: Data Analytics Security Plan**

Laurence T. Burden

Purdue Global University

IN498: Bachelor’s Capstone in Analytics

Robert Kayl

January 9, 2024

**Objectives and Goals**

The goal of this security plan is to protect the financial data protected by this company. This includes following applicable laws and regulations. Proper ethical control of the data will also be a driving factor in storing this data.

The primary objective of this plan is to prevent any unauthorized access to the data stored in support of data analytics at the company. A robust response plan must also be followed in the event of any breaches of this plan. The trust of the organization’s clients can only be maintained by making these objectives a top priority.

**Applicable Laws and Ethics**

Many laws exist that the company must follow to ensure our data is properly protected. These include the General Data Protection Regulation (GDPR) in the EU. The United States also has the Gramm-Leach-Billey Act and the Privacy Act that must be understood and adhered to.

All three laws require tenants that follow the principles of minimization, storage requirements, and accountability. Minimization refers to collecting only the information that is needed to complete the job and to reject keeping unneeded information. The storage requirements ensure that the data is securely stored and properly deleted when not needed anymore. Finally, accountability refers to the organization putting processes in place that will hold anyone accountable who fails to meet these standards.

**Security Methods**

**Data classifications and separation**

The data will be separated into various classification buckets based on NIST IR 8496, Data Classification Concepts, and Considerations for Improving Data Protection (2023). The organization will classify data based on the applicable laws. Any data sets that contain PII or UFD will be separated from any data sets that contain only GUD. The following labels will be used.

* **PII**: used for all personally identifiable information (protected data)
* **UFD**: used for user financial data (protected data)
* **GUD**: used for general user data (non-protected data)

**Data Flows and Transfer**

Protected data will only flow through an approved data pipeline. This pipeline will be designed to anonymize data and only allow the anonymized data to flow to the final model construction software. The data shall not flow to any other source and the final data model will not be sent back to the protected data server. Instead, the final model should be sent to a separate server for use.

**Data Storage**

All data will be stored on-premise servers. These will be administered by company employees. Contractors will be barred from being granted admin privileges on any server that stores confidential data. The storage servers shall also implement data-at-rest storage security practices, including encryption.

**Encryption**

All data will be encrypted while it is at rest or in transit. This will ensure that data cannot be easily read in the event of an unauthorized copy of the database being exfiltrated by the company. The following policies will cover the two states of data:

* **Data At Rest**: Data will be encrypted using the provided data at rest tools of the DBMS being utilized.
* **Data in Transit**: Data will only be transmitted over secure network lines using HTTPS and SSL/TLS.

**Privacy**

The privacy of our users is a top priority. All information used for analytics must be anonymized before being processed. This will involve removing any PII from the dataset before being sent to the model creation software.

**Access/Roles/Permissions**

The data owner shall be responsible for overseeing the permissions of the various roles. Only employees with a need-to-know will be allowed to have access to data that is considered protected. The following roles shall be used for administering the data: consumer, user, and admin. Consumers shall only have read access. Users are allowed to read and edit permissions. Finally, admins shall be allowed to read, write, and delete the data.

**Account Management**

Accounts will be managed by the IT department. The process of account management will follow the larger company's IT policy.

**Reviews**

The data and policies surrounding it shall be audited every quarter. These audits will review the access logs to ensure that no unauthorized access has occurred. Logs shall also be reviewed on changes made to the data. Finally, the data will be reviewed for conformance with privacy policies.

**Data Destruction**

The data will be retained following company policy and relevant federal laws. The top-level company data policy follows the FISM guidelines. NIST 800-171 outlines retention lengths for various classification levels. Any data that does not fit a data classification definition will be retained for a minimum of three years. The WipeDisk software tool will be used to ensure the data is properly deleted.

Proper destruction of unneeded information is key to preventing the illegal use of said data. For this reason, the hard drives that contain protected data will be destroyed when no longer needed. This involves degaussing hard drives and physically destroying SSDs.

**Data Analysis Tools Allowed**

All hardware used for data analysis and storage will be sourced through reputable companies. The operating system shall be any version of Windows Server or Red Hat Linux still supported by their respective company. The operating system will go through proper hardening processes and have all applicable security updates applied.

The main software used for data analysis will be Python 3.X. NumPy, SciPy, Pandas, and Matplotlib are the allowed packages for Python. The database shall utilize PostgreSQL, version 16.1 or later.

**Data Backups**

All data will follow a regular data backup plan. This involves a rotating schedule of tap backups. A full backup will be performed weekly and incremental backups will be completed daily.

**Physical Security Measures**

All servers that contain protected data will be placed in a secure room. The room will require a badge and pin system with proper access controls. Access logs will be maintained to track who accesses the physical location.

**Training**

All members with access to protected data will be required to maintain their cybersecurity training in line with company policy. Any member with elevated access shall be required to complete advanced training.

**Monitoring**

Proper logs shall be maintained for both digital and physical access to protected data. These logs will be continuously monitored for compliance with all policies. This shall be performed by the data governance compliance team.

**Roll Out**

This data security plan will be implemented immediately. All stated policies will apply to all standing instances of the data. Any processes that currently deviate from this document will be noted and fixed as soon as possible.

**References**

Data Classification Concepts and Considerations for Improving Data Protection (2023, November). https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8496.ipd.pdf