SEDGE THE SOLUTIONS EDGE

APPLICATION OF GDPR RULES TO AI / MACHINE LEARNING APPLICATION

Data Governance and GDPR Rules application

www.sedge.ai

ABSTRACT

Data Governance and the importance of principles and rules governing various types of data. What is GDPR Rule and its application to the AI application and how SEDGE ensures compliance of the process regarding GDPR rules.

SVM Analytics and Solutions Inc.,

A. Data governance and why is it important?

Data governance is the “norms, principles and rules governing various types of data”[1]. A strong and effective data governance ensures that data is trustworthy, consistent and secure. With the advent of new data privacy rules such as GDPR, organisations are complying with regulations and are creating process which perform governance of data with the purpose to

a. Ensuring there are no inconsistencies in the data
b. Maintaining a common data definition across different department within organization.
c. Improving the data quality
d. Decision making which is based on reliable and accurate analytics
e. To ensure compliance with data privacy rules
f. Preventing erroneous data and misuse of data

B. Why do organizations need to apply GDPR rules to any AI application?

Artificial intelligence applications process large amounts of data and therefore have to address the data privacy concerns. One of the fundamental questions is how should application functionalities be controlled, in order to address the data privacy concerns. The data privacy rules, is defined in the General Data Protection Regulation (GDPR). The objective of GDPR is to protect the data privacy of the EU citizen and the individuals more control over their personal data, in terms of transparency and standards of privacy. Many countries around the world have adopted EU GDPR rule framework to their own data privacy rules with variations to suit their standards.

[1] https://web.archive.org/web/20190805083643/https://datagovhub.org/224-2/

Machine learning is an application of artificial intelligence (AI), where the system reads the data, cleans the data, transform the data, learns the data and corrects itself automatically and generates model with which the system can approximately predict numerical values or classify objects. The accuracy of this function is determined based on the %age of the correctly predicted target value.

As the system needs to read large amount of data, which could also involve personal data of EU citizens which fall under the GDPR rules. One of the key question is - when using predictive analytics, can the business or individuals using predictive analytics tools, ensure that personal data protection rights are protected as defined in the GDPR rules?

SEDGE application guides the user to navigate through the fundamentals of GDPR where the users can answer the key questions and makes the user understand the importance of data privacy and the purpose of the analysis. These set of questions are saved along with the model and becomes a part of the document repository associated with the model that is created. Below are some of the key questions which SEDGE engages the user and making the user aware of the importance of data privacy and data protection.

C. How is GDPR rules applied to any machine learning?

The GDPR rules applies to the predictive analytics software, when

a. The model is being built by using variables which are defined as personal data
b. When the target variable arrives at a decision about an individual
c. If the ML produces an outcome about an individual

Following are the GDPR rules which govern predictive analytics software-

Article 22. 1 state that "*The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which*

produces legal effects concerning him or her or similarly significantly affects him or her."[2]

Article 22. 4 state that "*Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place."*

Article 13.2.f - *the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.*

Article 15.1.h states that *"the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject."*

D. How does SEDGE ensure that the process followed in model building, is compliant to GDPR rule?

SEDGE has a robust process with respect to GDPR functionality. The whole process from loading to building and deploying model, the system guides the users through different steps of GDPR compliant process. There are 3 major steps that the application guides the user to handle the personal data in a GDPR compliant manner-

1. Every data that is loaded, a checklist guides the user to ensure that the data which has the personal data, being processed has to comply with

[2] https://gdpr-info.eu/art-22-gdpr/

the GDPR rules or similar data privacy rules of other countries. These details are tagged to every data that the data controller works with.

2. The data controller (person loading data) has the ability to publish the data to the data processor, ensuring that the personal data fields within the data has been Pseudonymized.
3. The model which the system has built, is transparent and the user has the ability to interpret the fields and demonstrate that the model generated does not bias the outcome of the prediction with the use of sensitive fields.
4. Data controllers get the assurance of data protection by-

    a. pseudonymizing personal data as soon as possible
    SEDGE- Pseudonymization feature in GDPR page

    b. transparency with regard to the functions and processing of personal data
    SEDGE- Function to view the pipeline, transformation and data clean steps done on data

    c. Enabling the controller to create and improve security features
    SEDGE- mechanism where data controllers can pseudonymizing the personal data columns and then publish the data to controllers

    d. enabling the data subject to monitor the data processing
    SEDGE - Ability to provide users the login to SEDGE and publishing the data for monitoring purpose

    e. Identifying the source of data
    SEDGE- Logging of information in the SEDGE GDPR page

    f. Identifying the people who will be processing the data
    SEDGE- Logging of information in the SEDGE GDPR page

    g. Ensuring a time validity of every data
    SEDGE- Logging of information in the SEDGE GDPR page and automatically disabling data access once validity date has expired

    h. Purpose of processing data
    SEDGE- logging of information in the SEDGE GDPR page and attaching documentary artifacts to the model for the purpose of audit

    i. Manner in which data subjects are impacted by this analysis
    SEDGE- logging of information in the SEDGE GDPR page and attaching documentary artifacts to the model for the purpose of audit