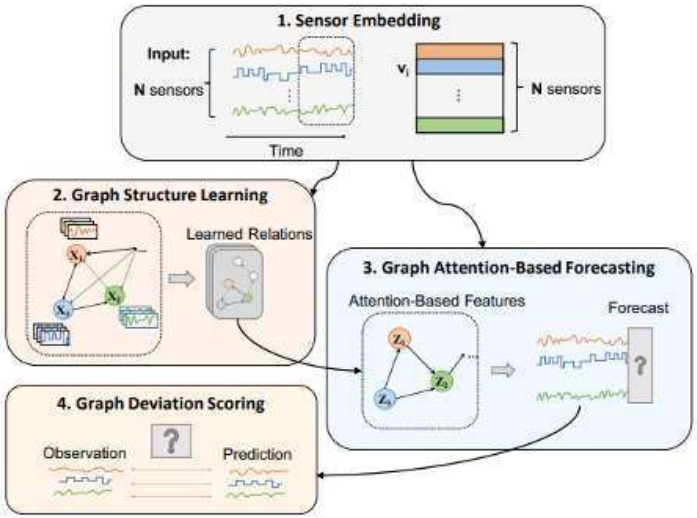
Project Design Phase-II
Technology Stack (Architecture & Stack)

Date	27 October 2023
Team ID	8.1
Project Name	System that uses AI for real-time data classification, leak detection, and prevention to enhance data security.
Maximum Marks	4 Marks

Technical Architecture:
Reference:



Guidelines:

1. **Clearly Define Objectives and Scope:** Begin by defining the specific objectives and scope of your project. Clearly outline what types of data you will classify and protect, what constitutes a data leak or security breach, and what level of enhancement you aim to achieve in terms of data security. A well-defined scope will help in setting clear project goals and expectations.
2. **Select Appropriate AI Models and Tools:** Carefully choose the AI models and tools that are most suitable for your project. Consider the nature of your data, the real-time requirements, and the types of security threats you anticipate. Ensure that the AI models you select are well-suited for data classification, anomaly detection, and prevention.
3. **Data Privacy and Compliance:** Prioritize data privacy and compliance with relevant regulations and standards, such as GDPR, HIPAA, or industry-specific security requirements. Implement robust data encryption, access controls, and auditing mechanisms to ensure that sensitive data is handled in accordance with best practices and legal requirements.
4. **Continuous Monitoring and Response:** Develop a system for continuous monitoring of data and security events. Implement real-time alerts and incident response mechanisms to respond to security threats as they occur. The ability to quickly detect and mitigate security incidents is a critical aspect of your system.
5. **User-Friendly Interface and Training:** Ensure that the system is user-friendly and provides clear information to security analysts. Additionally, provide training and documentation for your team to effectively use the system. Well-designed interface and well-trained personnel will help maximize the system's effectiveness in enhancing data security.

Table-1 : Components & Technologies:

SL.No	COMPONENT	DESCRIPTION	TECHNOLOGY
1	Data Collector	Collects and preprocesses data from various sources.	Apache Kafka, Apache NiFi, Logstash

2	Data Classification	Classifies incoming data in real-time to identify sensitive information.	TensorFlow, PyTorch, scikit-learn
3	Anomaly Detection	Uses AI algorithms to detect anomalies or unauthorized access patterns.	Elasticsearch, Kibana,
4	Alert Generation	Generates alerts for detected anomalous network traffic patterns.	PagerDuty, OpsGenie
5	Model Manager	Manages the lifecycle of machine learning models, including training, deployment, and monitoring.	MLflow, Neptune
6	Model Registry	Stores and manages machine learning models.	TensorFlow Serving, Amazon SageMaker Model Registry
7	Incident Response	Initiates actions like alerts, blocking, and remediation in case of threats.	SIEM Systems, SOAR Platforms, IDS/IPS, SIRPs, Threat Intelligence Platforms
8	Data Encryption	Encrypts sensitive data to protect from unauthorized access.	AES, RSA
9	Logging and Monitoring	Logs all activities and monitors the system for security events.	Splunk, ELK Stack (Elasticsearch, Logstash, Kibana)
10	Data Storage	Securely stores sensitive data and related metadata.	Amazon S3, Azure Blob Storage, Google Cloud Storage

11	API Gateway	Provides controlled external access to the system.	AWS API Gateway, Azure API Management, Apigee
12	External Threat Intelligence	Integrates external threat feeds for enhanced threat detection.	Threat intelligence platforms, OSINT feeds

Table-2: Application Characteristics:

SL.No	Characteristic	Description	Technology
1	Adaptability	The ability to handle large amounts of network traffic data.	Utilizing distributed systems and cloud computing.
2	Real-time Detection	The capacity to identify abnormal network traffic patterns as they occur.	Employing high-performance computing and in-memory data processing.
3	Precision	The skill to accurately pinpoint irregular network traffic patterns while minimizing the	Utilizing machine learning algorithms and advanced feature engineering techniques.

		occurrence of false positives and false negatives.	
4	Resilience	The robustness to operate reliably even in the presence of data noise and errors.	Accomplished through thorough data preprocessing and meticulous model validation.
5	Transparency	The capability to provide clear explanations for detected anomalies to analysts.	Utilizing machine learning models that offer interpretable insights.
6	User-Friendliness	The ease of use that enables analysts to interact effectively with the system.	Achieved through the implementation of user-friendly interfaces and comprehensive documentation.
7	Seamless Integration	The ability to integrate with existing security systems and infrastructures.	Facilitated by open APIs and adherence to standard data formats.
8	Cost Efficiency	The capability to be deployed and maintained without excessive financial burden.	Leveraging cloud computing resources and open-source software solutions.
9	Enhanced Security	The proficiency to protect against unauthorized access and security breaches.	Implemented through encryption, strict access control mechanisms, and vigilant security monitoring.

10	Comprehensive Auditing	The capacity to generate thorough logs for auditing and troubleshooting purposes.	Generated audit logs and integration with Security Information and Event Management (SIEM) systems.
11	Regulatory Compliance	The ability to align with pertinent security regulations and standards.	Attainment of industry-specific security certifications and utilization of compliance reporting tools.

References: