

Machine Learning-Based Diabetes Prediction System Using AWS Cloud

Abstract:

This project focuses on developing a machine learning-based system for predicting diabetes risk using AWS cloud services. The system collects and processes patient health data to build predictive models using AWS SageMaker, which apply supervised learning techniques to determine diabetes risk levels. By leveraging AWS's scalable infrastructure, the framework provides healthcare providers with a reliable and cost-effective tool for diabetes prediction, early diagnosis, and preventive care. With integrated analytics and visualization capabilities, this system empowers healthcare professionals to identify at-risk individuals quickly and recommend timely interventions, ultimately improving patient outcomes.

Existing System:

Traditional diabetes diagnosis relies on time-consuming clinical tests, and many predictive models lack scalability and real-time analysis capabilities. Systems that rely on on-premise infrastructure often require extensive resources to handle large datasets and are challenging to integrate with other healthcare management platforms.

Proposed System:

This machine learning-based system on AWS allows for scalable, real-time diabetes risk prediction. Data from patient records, such as blood glucose levels, BMI, age, and other health metrics, is ingested into AWS for processing. AWS SageMaker trains and deploys predictive models using algorithms like logistic regression and decision trees, which are fine-tuned to improve accuracy. AWS Lambda enables data processing and feature extraction, while DynamoDB stores patient data and prediction results. The system provides healthcare providers with a secure, efficient tool for diabetes screening and risk assessment, offering real-time insights into patient health data. Amazon QuickSight is used to visualize trends and risk factors, aiding doctors in decision-making and patient counseling.

Key Features:

Automated Machine Learning Pipeline: SageMaker automates model training, validation, and deployment for diabetes risk prediction.

Real-Time Data Processing: Lambda functions preprocess patient data and extract features for input into the model.

Scalable Data Storage: DynamoDB stores patient data and prediction outputs, allowing seamless access and integration.

Data Visualization and Insights: QuickSight provides visual dashboards to track risk factors and patient trends.

Data Security and Compliance: AWS services ensure data encryption, access control, and compliance with healthcare regulations.

Software Tools:

AWS S3: The primary storage service for securely holding patient health information.

AWS Lambda: Automates data processing tasks when new data is uploaded to S3.

Amazon RDS: Stores metadata for patient records to support efficient querying and indexing.

AWS CloudWatch: Monitors system performance, logs access, and sets alerts for system health.

AWS IAM: Manages access and permissions to ensure only authorized users can access data.

AWS KMS (Key Management Service): Provides encryption keys to secure patient information.

Hardware Tools:

AWS Cloud Infrastructure: Includes compute power for data processing, storage for historical data, and IoT device connectivity.