INFORMATION SECURITY PROJECT

TEAM

SAYED AMAN SHAH

AHMED RAZA

SALEH AHMAD

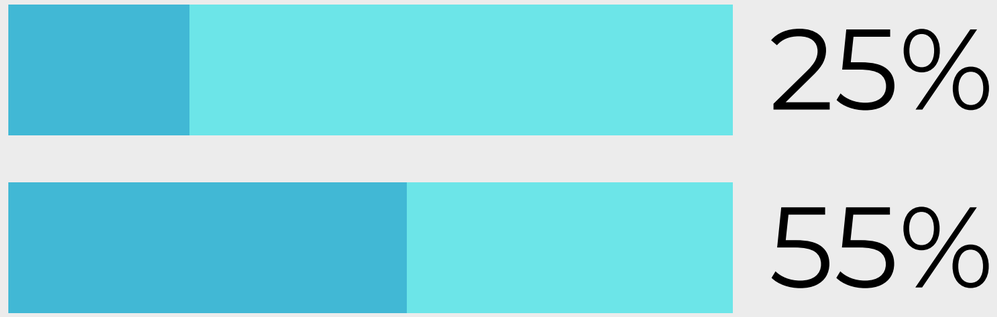
TABLE OF CONTENTS

INTRODUCTION TO DATASET

- Disease prediction from blood analysis
- Parameters such as glucose, cholesterol, blood pressure WBC, RBC , Insulin and BMI, etc
- Disease: This indicates what disease the person has.
- Structured tabular data for analysis.

RELATED WORK

- The dataset is famous kaggle with 1.5k+ downloads.
- A lot of people worked on this dataset.
- Some people created classifiers from scratch.
- Pretrained networks were also used.
- Results were in range 25% to 55% depending upon the scale of the classifier.
- Transfer learning and Domain adaptation techniques for training.



ENCRYPTION ALGORITHM

Security in this project relies on **RSA Encryption**, allowing computations on encrypted data without decryption. This protects model parameters' privacy during distributed training.

- **Client-side Training:** Clients train models on their data, ensuring data privacy and sovereignty.
- **Encryption of Model Parameters:** Trained model parameters are encrypted with RSA Encryption before transmission to the central server, safeguarding sensitive information.
- **Secure Aggregation on Server:** The server performs secure aggregation operations on encrypted parameters, preserving user confidentiality during computation.
- **Transmission of Averaged Parameters:** Encrypted averaged model parameters are sent back to clients, eliminating interception or tampering risks, as encrypted.
- **Decryption and Update on Client-side:** Clients decrypt and update their models with averaged parameters, enabling collaborative knowledge exchange without compromising privacy.

ML / DL ALGORHITHM

Project's deep learning model is a dynamic feedforward neural network with flexible architecture, allowing developers to experiment with data complexity and network models without compromising functionality.

- **Feedforward Structure:** The model uses a feedforward architecture, allowing sequential data flow from input to output layers for feature extraction and decision-making without feedback loops.
- **Dynamic Architecture:** Its adaptable design lets you change the number of layers and neurons dynamically without performance loss, enhancing flexibility for different datasets and optimal performance.
- **Activation Layers:** Incorporates activation layers (ReLU, sigmoid, tanh) for introducing non-linearity, aiding in complex data learning and feature extraction.
- **Preprocessing:** Very basic preprocessing on the data like scaling the numeric values, outlier detection and null removal i.e. data cleaning.

PERFORMANCE OF OUR MODELS

Here are some performance metrics for our approaches:

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Basic ML / DL	**61**	**47**	**52**	**49**

CONTINUING

# Client	Accuracy (%)		Precision (%)		Recall (%)		F1 (%)	
Aggregation	Before	After	Before	After	Before	After	Before	After
Client 1	21	61	25	51	30	52	29	53
Client 2	20	59	29	55	35	59	31	50
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

THANK YOU

Any Questions?