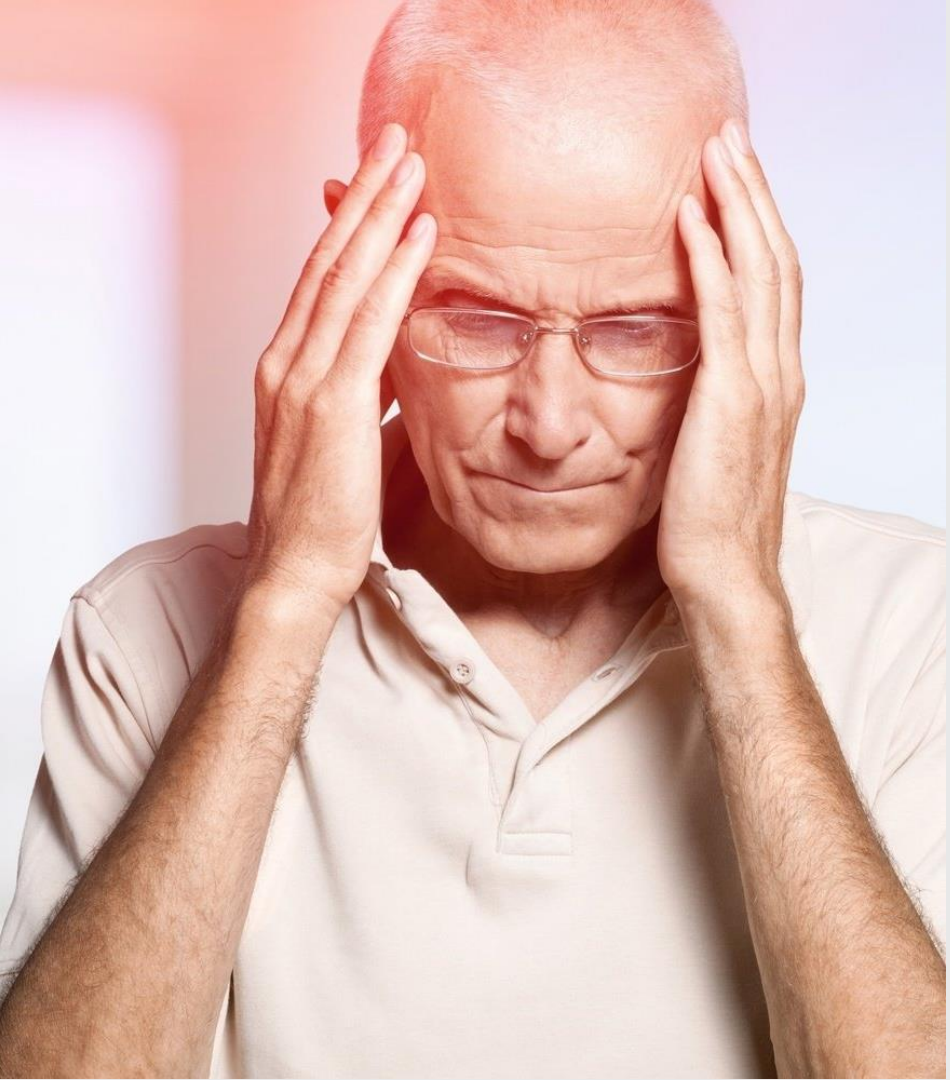
GROUP JIGGLYPUFF

Gamze Aksu - 171180005
Cansu Ayten - 171180010
Ceren Umay Özten - 181180060



MOTIVATION

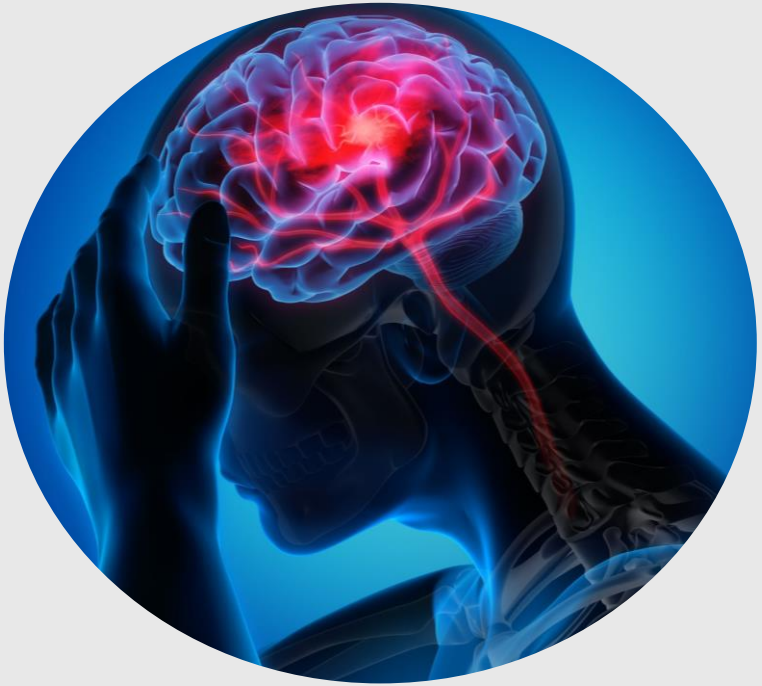
- Stroke is the third leading cause of death and disability worldwide.
- Some special factors are important for the detection of stroke disease throughout people's lives.

DATASET

This dataset includes 11 features.

These are:

- Id
- Gender
- Age
- Hypertension
- Heart Disease
- Marriage Status
- Work Type
- Residence Type
- Average Glucose Level
- BMI (Body Mass Index)
- Smoking Status



PREPROCESSING

There are 5110 records in the dataset.

- id feature not included in the project
- There are 201 null values in the BMI feature. These null values have been changed to be the median of the BMI feature.
- Outlier values were found. Outlier values are also replaced with the median of the feature.

METHOD

- ANN Model
- Dense layers that have sequentially 64-32-16-1 number of neural nodes from input layer - hidden layer - output layer.
- As the activation function, ‘‘relu” in the hidden layer and the ‘‘sigmoid” function in the output layer were used.
- ‘‘Binary cross entropy” function was used as loss function.
- Model is trained 50 epochs.

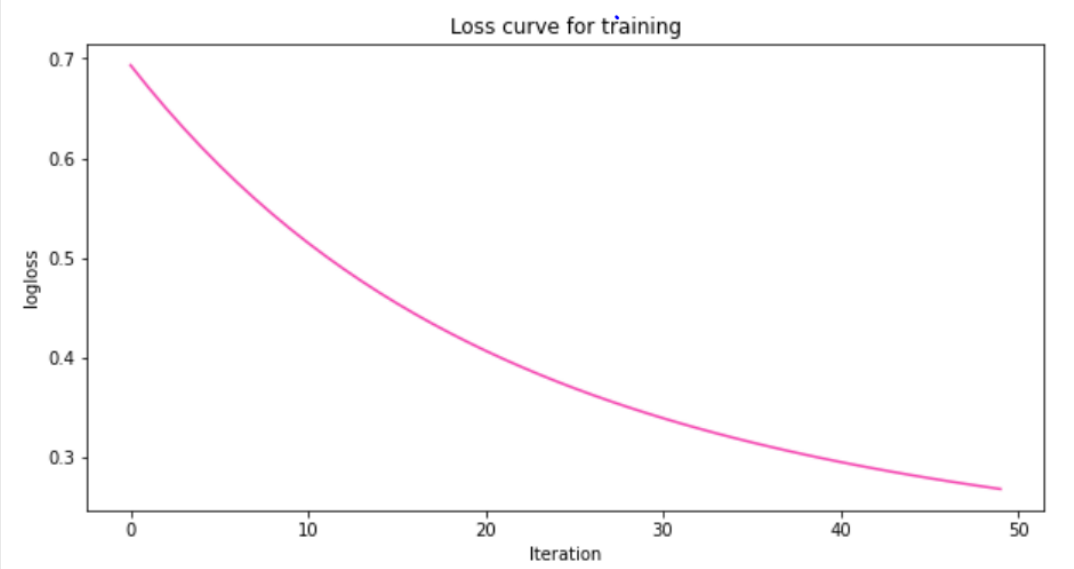
EXPERIMENTS - 1

1. The learning rate values, epoch and initial weights were changed.
   - Learning rate values changed to 0.1, 0.01, 0.001, and 0.0001.
   - Epoch values were tried as 50 and 100.
   - The initial values of the weights are chosen randomly over the normal distribution, and secondly, they are all zero.
   - Then the weights were assigned as initial values.

   Outcomes were observed at the end of these experiments. The best output is obtained in 50 epochs trained model with zero weights and a learning rate of 0.1.

RESULT - 1

- 0.94 accuracy with 50 epochs and learning rate of 0.1



INTERPRETABILITY

- Interpretability is used to get the models created with Deep Learning out of the Black Box state. Thus, the model's decision-making style can be easily understood by associating it with its outputs.
  - A local method is used to understand the decision-making process for a single instance of the model.
  - A global method is used to understand the decision-making process according to the general structure of the model.

INTERPRETABILITY

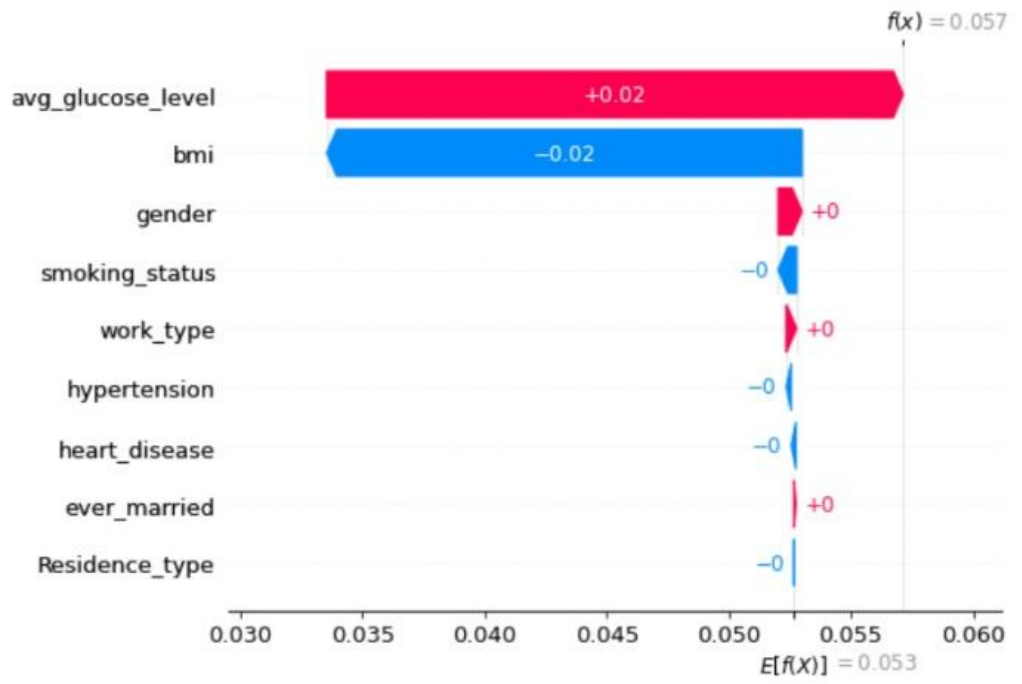
- SHAP
- Contains a local and a global method.
- With the SHAP Values, interpretability is enhanced by associating the model and data with outputs.

EXPERIMENTS - 2

2. It is seen that the age feature contributes the most to the model in the interpretability summary plot. Age feature affects the model a lot as it has much more contribution than others.

- The age feature was removed and the effect of other features on the model was examined.
- After removing the age feature, it is seen that it is the average glucose level feature that contributes the most to the model.

RESULT - 2

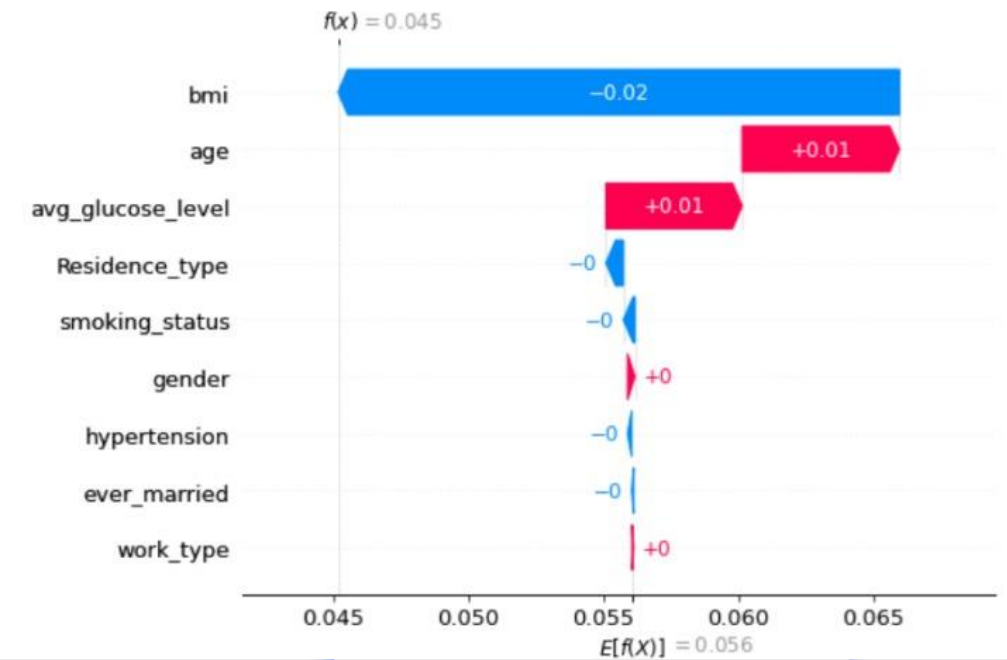

f(x) = 0.057
avg_glucose_level
+0.02
bmi
−0.02
gender
+0
smoking_status
−0
work_type
+0
hypertension
−0
heart_disease
−0
ever_married
+0
Residence_type
−0
0.030
0.035
0.040
0.045
0.050
0.055
0.060
E[f(X)] = 0.053

EXPERIMENTS - 3

3. It is seen that the feature with the least contribution in the summary plot of the model is the heart disease feature.

   - The heart disease feature was removed and the effect of other features on the model was observed.
   - After removing the heart disease feature, the age feature again contributes by far the most.

RESULT - 3


f(x) = 0.045
bmi
age
avg_glucose_level
Residence_type
smoking_status
gender
hypertension
ever_married
work_type
−0.02
+0.01
+0.01
−0
−0
+0
−0
−0
+0
0.045
0.050
0.055
0.060
0.065
E[f(X)] = 0.056

THANKS