rnwfn1dy6

February 2, 2025

[2]:
```
#importing the packages that needed
import pandas as pd
import numpy as np
import tensorflow as tf
from tensorflow import keras
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score, classification_report,␣
 ↪confusion_matrix
import matplotlib.pyplot as plt
```

[4]:
```
from google.colab import files

# Upload file
uploaded = files.upload()
```

```
<IPython.core.display.HTML object>
```

```
Saving alzheimers_prediction_dataset.csv to alzheimers_prediction_dataset.csv
```

[5]:
```
# Load dataset
filename = list(uploaded.keys())[0]  # Get the uploaded filename
data = pd.read_csv(r'alzheimers_prediction_dataset.csv', encoding='latin-1')  ␣
 ↪# Change to read_excel(filename) for Excel files

# Display first few rows
data.head()
```

[5]:
```
        Country  Age  Gender  Education Level   BMI Physical Activity Level  \
0         Spain   90    Male                1  33.0                  Medium
1     Argentina   72    Male                7  29.9                  Medium
2  South Africa   86  Female               19  22.9                    High
3         China   53    Male               17  31.2                     Low
4        Sweden   58  Female                3  30.0                    High

  Smoking Status Alcohol Consumption Diabetes Hypertension  …  \
0          Never        Occasionally       No           No  …
1         Former               Never       No           No  …
```

```
2        Current        Occasionally       No           Yes  ...
3          Never           Regularly      Yes            No  ...
4         Former               Never      Yes            No  ...

  Dietary Habits Air Pollution Exposure  Employment Status Marital Status  \
0        Healthy                   High            Retired         Single
1        Healthy                 Medium         Unemployed        Widowed
2        Average                 Medium           Employed         Single
3        Healthy                 Medium            Retired         Single
4      Unhealthy                   High           Employed        Married

  Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿ Social Engagement Level Income Level  \
0                                    No                     Low       Medium
1                                    No                    High          Low
2                                    No                     Low       Medium
3                                    No                    High       Medium
4                                    No                     Low       Medium

  Stress Levels Urban vs Rural Living Alzheimerï¿½ï¿½ï¿½ï¿½ï¿
0          High                 Urban                      No
1          High                 Urban                      No
2          High                 Rural                      No
3           Low                 Rural                      No
4          High                 Rural                      No

[5 rows x 25 columns]
```

```
[7]: data.columns
```

```
[7]: Index(['Country', 'Age', 'Gender', 'Education Level', 'BMI',
            'Physical Activity Level', 'Smoking Status', 'Alcohol Consumption',
            'Diabetes', 'Hypertension', 'Cholesterol Level',
            'Family History of Alzheimerï¿½ï', 'Cognitive Test Score',
            'Depression Level', 'Sleep Quality', 'Dietary Habits',
            'Air Pollution Exposure', 'Employment Status', 'Marital Status',
            'Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿', 'Social Engagement Level',
            'Income Level', 'Stress Levels', 'Urban vs Rural Living',
            'Alzheimerï¿½ï¿½ï¿½ï¿½ï¿'],
           dtype='object')
```

```
[8]: data.tail()
```

```
[8]:       Country  Age  Gender  Education Level   BMI Physical Activity Level  \
     74278  Russia   60  Female                3  22.6                    High
     74279      UK   58    Male               18  30.6                     Low
     74280   Spain   57  Female               13  28.2                  Medium
     74281  Brazil   73  Female                7  29.0                     Low
```

```
74282  Norway   57  Female                 1  31.7                     Low

      Smoking Status Alcohol Consumption Diabetes Hypertension  …  \
74278         Former               Never       No           No  …
74279          Never        Occasionally      Yes           No  …
74280          Never           Regularly       No           No  …
74281          Never           Regularly       No           No  …
74282        Current           Regularly       No           No  …

      Dietary Habits Air Pollution Exposure  Employment Status Marital Status  \
74278        Average                   High         Unemployed        Widowed
74279        Average                 Medium         Unemployed         Single
74280        Healthy                    Low           Employed         Single
74281        Healthy                    Low           Employed        Widowed
74282        Average                    Low         Unemployed         Single

      Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿ Social Engagement Level  \
74278                                    No                  Medium
74279                                    No                  Medium
74280                                   Yes                    High
74281                                    No                     Low
74282                                    No                     Low

      Income Level Stress Levels Urban vs Rural Living Alzheimerï¿½ï¿½ï¿½ï¿½ï¿
74278         High        Medium                 Rural                      No
74279         High          High                 Rural                      No
74280          Low           Low                 Rural                      No
74281          Low          High                 Rural                      No
74282       Medium        Medium                 Urban                      No

[5 rows x 25 columns]
```

[9]:

```
data.describe()
```

[9]:

```
                Age  Education Level           BMI  Cognitive Test Score
count  74283.000000     74283.000000  74283.000000          74283.000000
mean      71.964703         9.487514     26.780639             64.654241
std       12.980748         5.757020      4.764679             20.153247
min       50.000000         0.000000     18.500000             30.000000
25%       61.000000         4.000000     22.700000             47.000000
50%       72.000000         9.000000     26.800000             65.000000
75%       83.000000        14.000000     30.900000             82.000000
max       94.000000        19.000000     35.000000             99.000000
```

[10]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 74283 entries, 0 to 74282
Data columns (total 25 columns):
 #   Column                                 Non-Null Count  Dtype
---  ------                                 --------------  -----
 0   Country                                74283 non-null  object
 1   Age                                    74283 non-null  int64
 2   Gender                                 74283 non-null  object
 3   Education Level                        74283 non-null  int64
 4   BMI                                    74283 non-null  float64
 5   Physical Activity Level                74283 non-null  object
 6   Smoking Status                         74283 non-null  object
 7   Alcohol Consumption                    74283 non-null  object
 8   Diabetes                               74283 non-null  object
 9   Hypertension                           74283 non-null  object
 10  Cholesterol Level                      74283 non-null  object
 11  Family History of Alzheimerï¿½ï        74283 non-null  object
 12  Cognitive Test Score                   74283 non-null  int64
 13  Depression Level                       74283 non-null  object
 14  Sleep Quality                          74283 non-null  object
 15  Dietary Habits                         74283 non-null  object
 16  Air Pollution Exposure                 74283 non-null  object
 17  Employment Status                      74283 non-null  object
 18  Marital Status                         74283 non-null  object
 19  Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿  74283 non-null  object
 20  Social Engagement Level                74283 non-null  object
 21  Income Level                           74283 non-null  object
 22  Stress Levels                          74283 non-null  object
 23  Urban vs Rural Living                  74283 non-null  object
 24  Alzheimerï¿½ï¿½ï¿½ï¿½ï¿                74283 non-null  object
dtypes: float64(1), int64(3), object(21)
memory usage: 14.2+ MB
```

[11]:
```
data.isnull().sum()
```

[11]:
```
Country                                  0
Age                                      0
Gender                                   0
Education Level                          0
BMI                                      0
Physical Activity Level                  0
Smoking Status                           0
Alcohol Consumption                      0
Diabetes                                 0
Hypertension                             0
Cholesterol Level                        0
Family History of Alzheimerï¿½ï          0
Cognitive Test Score                     0
```

```
Depression Level                            0
Sleep Quality                               0
Dietary Habits                              0
Air Pollution Exposure                      0
Employment Status                           0
Marital Status                              0
Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿    0
Social Engagement Level                     0
Income Level                                0
Stress Levels                               0
Urban vs Rural Living                       0
Alzheimerï¿½ï¿½ï¿½ï¿½ï¿                      0
dtype: int64
```

[12]:
```python
from sklearn.preprocessing import LabelEncoder
object_columns = data.select_dtypes(include=['object']).columns.tolist()
le = LabelEncoder()
for col in object_columns:
    data[col] = le.fit_transform(data[col])
```

[13]:
```python
data.head()
```

[13]:
```
   Country  Age  Gender  Education Level   BMI  Physical Activity Level  \
0       16   90       1                1  33.0                        2
1        0   72       1                7  29.9                        2
2       14   86       0               19  22.9                        0
3        4   53       1               17  31.2                        1
4       17   58       0                3  30.0                        0

   Smoking Status  Alcohol Consumption  Diabetes  Hypertension  ...  \
0               2                    1         0             0  ...
1               1                    0         0             0  ...
2               0                    1         0             1  ...
3               2                    2         1             0  ...
4               1                    0         1             0  ...

   Dietary Habits  Air Pollution Exposure  Employment Status  Marital Status  \
0               1                       0                  1               1
1               1                       2                  2               2
2               0                       2                  0               1
3               1                       2                  1               1
4               2                       0                  0               0

   Genetic Risk Factor (APOE-ï¿½ï¿½ï¿½ï¿  Social Engagement Level  \
0                                         0                        1
1                                         0                        0
2                                         0                        1
```

```
3                                       0                        0
4                                       0                        1

   Income Level  Stress Levels  Urban vs Rural Living  Alzheimerï¿½ï¿½ï¿½ï¿½ï¿
0             2              0                      1                        0
1             1              0                      1                        0
2             2              0                      0                        0
3             2              1                      0                        0
4             2              0                      0                        0

[5 rows x 25 columns]
```

[15]:
```
# Split data into features (X) and target variable (y)
X = data.drop('Alzheimerï¿½ï¿½ï¿½ï¿½ï¿', axis=1)
y = data['Alzheimerï¿½ï¿½ï¿½ï¿½ï¿']
```

[16]:
```
#Split data into training and testing sets (80% for training and 20% for␣
 ↪testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

[17]:
```
# Create StandardScaler instance
scaler = StandardScaler()

# Fit and transform both training and testing data
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

[18]:
```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

[19]:
```
# Define ANN model architecture
model = Sequential()
model.add(Dense(64, activation='relu', input_shape=(X_train_scaled.shape[1],)))
model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
```

```
/usr/local/lib/python3.11/dist-packages/keras/src/layers/core/dense.py:87:
UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When
using Sequential models, prefer using an `Input(shape)` object as the first
layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
```

[20]:
```
# Compile model with binary crossentropy loss and Adam optimizer
model.compile(loss='binary_crossentropy', optimizer='adam',␣
 ↪metrics=['accuracy'])
```

[21]:

```
# Train model on scaled training data
history = model.fit(X_train_scaled, y_train, validation_data=(X_test_scaled,␣
 ↪y_test), epochs=10, batch_size=32)
```

```
Epoch 1/10
1858/1858              10s 3ms/step -
accuracy: 0.6798 - loss: 0.5824 - val_accuracy: 0.7119 - val_loss: 0.5492
Epoch 2/10
1858/1858              6s 3ms/step -
accuracy: 0.7192 - loss: 0.5439 - val_accuracy: 0.7172 - val_loss: 0.5436
Epoch 3/10
1858/1858              7s 4ms/step -
accuracy: 0.7184 - loss: 0.5415 - val_accuracy: 0.7198 - val_loss: 0.5426
Epoch 4/10
1858/1858              8s 3ms/step -
accuracy: 0.7182 - loss: 0.5389 - val_accuracy: 0.7187 - val_loss: 0.5436
Epoch 5/10
1858/1858              7s 4ms/step -
accuracy: 0.7212 - loss: 0.5388 - val_accuracy: 0.7191 - val_loss: 0.5424
Epoch 6/10
1858/1858              5s 3ms/step -
accuracy: 0.7257 - loss: 0.5324 - val_accuracy: 0.7181 - val_loss: 0.5438
Epoch 7/10
1858/1858              5s 3ms/step -
accuracy: 0.7262 - loss: 0.5306 - val_accuracy: 0.7170 - val_loss: 0.5428
Epoch 8/10
1858/1858              6s 3ms/step -
accuracy: 0.7266 - loss: 0.5299 - val_accuracy: 0.7141 - val_loss: 0.5463
Epoch 9/10
1858/1858              10s 3ms/step -
accuracy: 0.7268 - loss: 0.5294 - val_accuracy: 0.7166 - val_loss: 0.5465
Epoch 10/10
1858/1858              10s 3ms/step -
accuracy: 0.7306 - loss: 0.5255 - val_accuracy: 0.7156 - val_loss: 0.5466
```

[23]:

```
#model summary
model.summary()
```

```
Model: "sequential"


 Layer (type)                           Output Shape                        ␣
 ↪Param #

 dense (Dense)                          (None, 64)                            ␣
 ↪1,600
 dense_1 (Dense)                       (None, 32)                          ␣
 ↪2,080

 dense_2 (Dense)                       (None, 1)                             ␣
 ↪ 33
```

```
Total params: 11,141 (43.52 KB)

Trainable params: 3,713 (14.50 KB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 7,428 (29.02 KB)
```

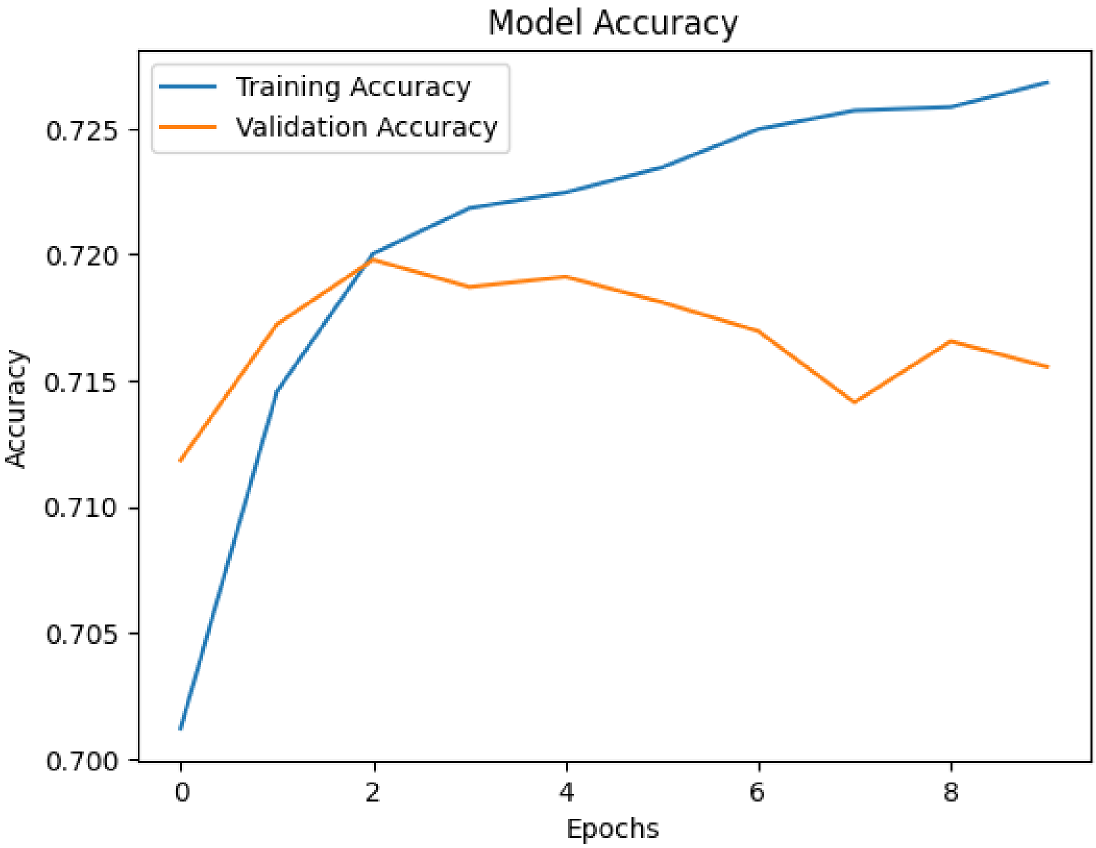
[24]:

```
# Evaluate model on testing data
loss, accuracy = model.evaluate(X_test_scaled, y_test)
print(f'Test Loss: {loss:.3f}, Test Accuracy: {accuracy:.3f}')
```

```
465/465              1s 3ms/step -
accuracy: 0.7191 - loss: 0.5449
Test Loss: 0.547, Test Accuracy: 0.716
```

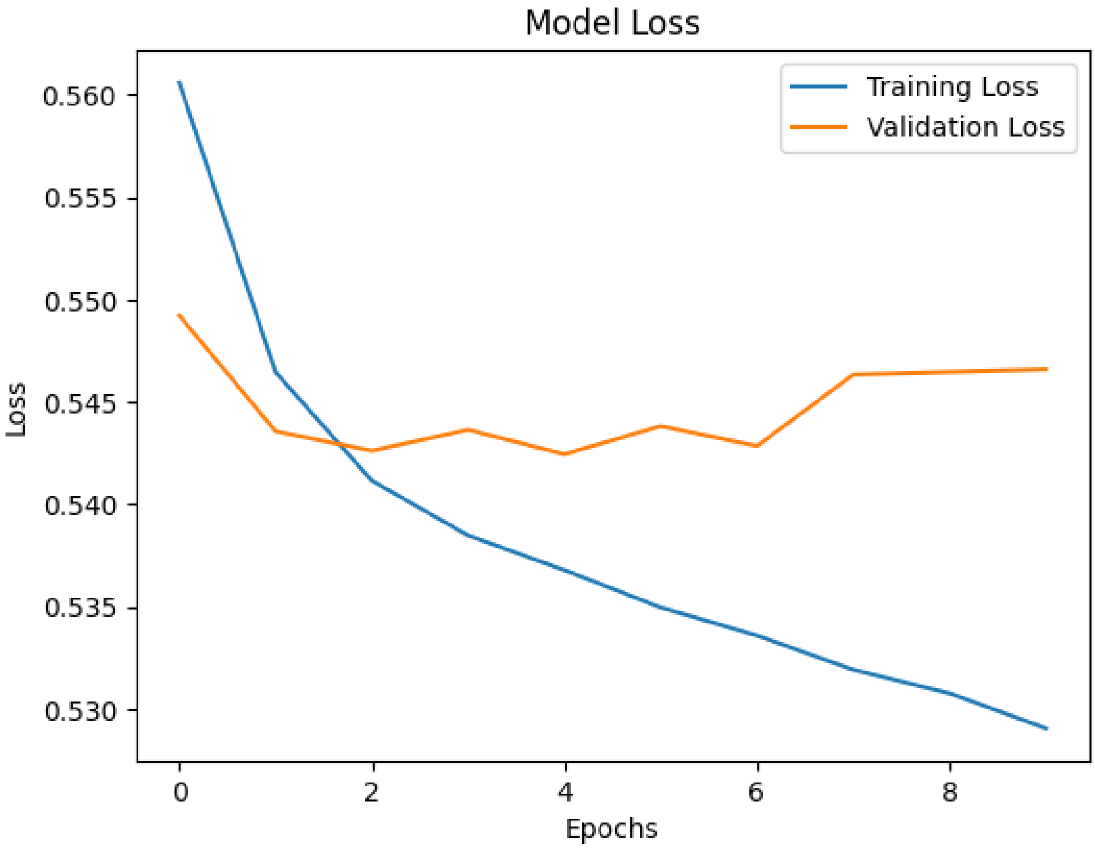
[25]:

```
# Plot training and validation accuracy
plt.plot(history.history['accuracy'], label='Training Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.title('Model Accuracy')
plt.xlabel('Epochs')
plt.ylabel('Accuracy')
plt.legend()
plt.show()
```



[26]:

```
# Plot training and validation loss
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('Model Loss')
plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.legend()
plt.show()
```


Model Loss
Training Loss
Validation Loss
Loss
0.560
0.555
0.550
0.545
0.540
0.535
0.530
Epochs
0
2
4
6
8