Rajalakshmi Engineering College
(An Autonomous Institution)
Rajalakshmi Nagar, Thandalam-602105

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

MINI PROJECT

AI23521 BUILD AND DEPLOY FOR MACHINE LEARNING APPLICATIONS

REGISTRATION NUMBER : **2116231501176**

STUDENT NAME : **UMA VYSHNAVI G**

YEAR : **2023-2027**

Key tools and technologies used include:

- Python: for backend logic and machine learning integration.

- **Streamlit**: for building a web interface to interact with the ML model.

- NumPy: to handle numeric input data.

- Pickle: for loading a pre-trained ML model.

- **Base64:** for safely embedding a local background image.

- VS Code: for writing and debugging code.

Project Overview

PROJECT TITLE : "Medical Insurance Premium Predictor".

OBJECTIVE:

To develop a user-friendly web application that predicts a user's insurance premium based on various input factors like age, BMI, gender, smoking habits, region, and number of children using a pre-trained machine learning model.

Key Features:

1. Age:
- Numerical input influencing health risk and insurance pricing.

2. BMI (Body Mass Index):
- Directly linked to health risks. Provided via slider input.

3. Gender:
- Selected via dropdown. Coded as binary for model input.

4. Smoker:
- Yes/No input that significantly impacts insurance costs.

5. Region:
- Chosen from 4 options, encoded numerically

6. Children:
- Number of dependent children influencing total premium.

7. Local Background Image:
- Custom image used to enhance visual design.

Technologies Used:

- Programming Language: Python
- Libraries: Streamlit, NumPy, Pickle, base64
- IDE: Visual Studio Code
- Model Deployment: Streamlit Web Interface

PROJECT IMPLEMENTATION:

The application was built using Streamlit for UI and a Pickle-loaded ML model for prediction. Inputs were processed and formatted into a NumPy array. The model predicted the insurance premium, which was displayed on the interface using styled HTML.

The background image was embedded using base64 encoding for full compatibility with Streamlit. Labels were made white for better visibility against the image.

PERFORMANCE ANALYSIS:

The prediction accuracy depends on the trained model stored in 'MIPML.pkl'. Though model training is outside this internship's scope, the deployed model was tested with sample data and provided meaningful premium estimates.
The app interface is clean, responsive, and styled for easy use.

CONCLUSION AND REFLECTION:

This project helped me gain practical experience in building and deploying machine learning models into web applications. I learned about data encoding, model loading, UI design using Streamlit, and styling enhancements using HTML and CSS inside Python.

I am now confident in handling similar ML deployment projects and creating interactive user-facing AI tools.

CODE:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
import pickle as pkl


# Loading Dataset
insurance_data = pd.read_csv('insurance.csv')


# Basic Exploration
print(insurance_data.head())
print(insurance_data.info())
print(insurance_data.describe())
```

```
print("Shape of data:", insurance_data.shape)
print("Missing values in each column:\n", insurance_data.isnull().sum())

# Data Distribution for Numerical Columns
sns.set()
plt.figure(figsize=(6,6))
sns.displot(insurance_data['age'])
plt.title('Distribution of Age')
plt.show()

plt.figure(figsize=(6,6))
sns.displot(insurance_data['bmi'])
plt.title('Distribution of BMI')
plt.show()

plt.figure(figsize=(6,6))
sns.displot(insurance_data['children'])
plt.title('Distribution of Children')
plt.show()
# Categorical Columns Visualization
plt.figure(figsize=(6,6))
sns.countplot(x='sex', data=insurance_data)
plt.title('Count of Gender')
plt.show()
```

```
plt.figure(figsize=(6,6))
sns.countplot(x='smoker', data=insurance_data)
plt.title('Count of Smoker')
plt.show()

plt.figure(figsize=(6,6))
sns.countplot(x='region', data=insurance_data)
plt.title('Count by Region')
plt.show()
# Encoding Categorical Data
insurance_data.replace({'sex': {'female': 0, 'male': 1}}, inplace=True)
insurance_data.replace({'smoker': {'no': 0, 'yes': 1}}, inplace=True)
insurance_data.replace({'region': {'southeast': 0, 'southwest': 1, 'northeast':
2, 'northwest': 3}}, inplace=True)

# Splitting Data into Inputs and Output
input_data = insurance_data.drop(columns='charges')
output_data = insurance_data['charges']

# Splitting Dataset into Training and Testing Sets
input_train_data, input_test_data, output_train_data, output_test_data =
train_test_split(
  input_data, output_data, test_size=0.2
```

```
)

print("Input Shapes - Full:", input_data.shape, "Train:", input_train_data.shape,
"Test:", input_test_data.shape)
print("Output Shapes - Full:", output_data.shape, "Train:",
output_train_data.shape, "Test:", output_test_data.shape)

# Model Creation and Training
model = RandomForestRegressor(n_estimators=100, max_depth=7)
model.fit(input_train_data, output_train_data)

# Making Predictions on Test Data
test_data_predictions = model.predict(input_test_data)

# Model Evaluation using R2 Score
print("R2 Score:", r2_score(test_data_predictions, output_test_data))


# Predicting Premium for Custom Input
# Format: age, sex, bmi, children, smoker, region
input_data = (35, 1, 35, 1, 0, 0)
input_data_array = np.asarray(input_data).reshape(1, -1)
insurance_premium = model.predict(input_data_array)
print("Predicted Insurance Premium:", insurance_premium[0])

# Saving the Model
pkl.dump(model, open('MIPML.pkl', 'wb')
```

APP.PY

```
import numpy as np
```

```
import pickle as pkl
import streamlit as st
import base64

# Load trained model
model = pkl.load(open('MIPML.pkl', 'rb'))

# Add local background image using base64
def add_bg_from_local(image_file):
    with open(image_file, "rb") as img_file:
        encoded = base64.b64encode(img_file.read()).decode()
    st.markdown(f"""
        <style>
        .stApp {{
            background: url("{encoded}") no-repeat center
center fixed;
            background-size: cover;
        }}
        .main-title {{
            color: white; font-size: 32px; text-align: center;
            background: rgba(0,0,0,0.6); padding: 10px; border-radius: 10px;
        }}
        label, .css-1cpxqw2, .st-af {{
            color: white !important; font-weight: bold;
        }}
        .result-box {{
            background: rgba(255,255,255,0.85); padding: 15px;
            border-radius: 10px; text-align: center;
            font-size: 22px; font-weight: bold; color: #006400;
            margin-top: 20px;
        }}
        </style>
    """, unsafe_allow_html=True)
```

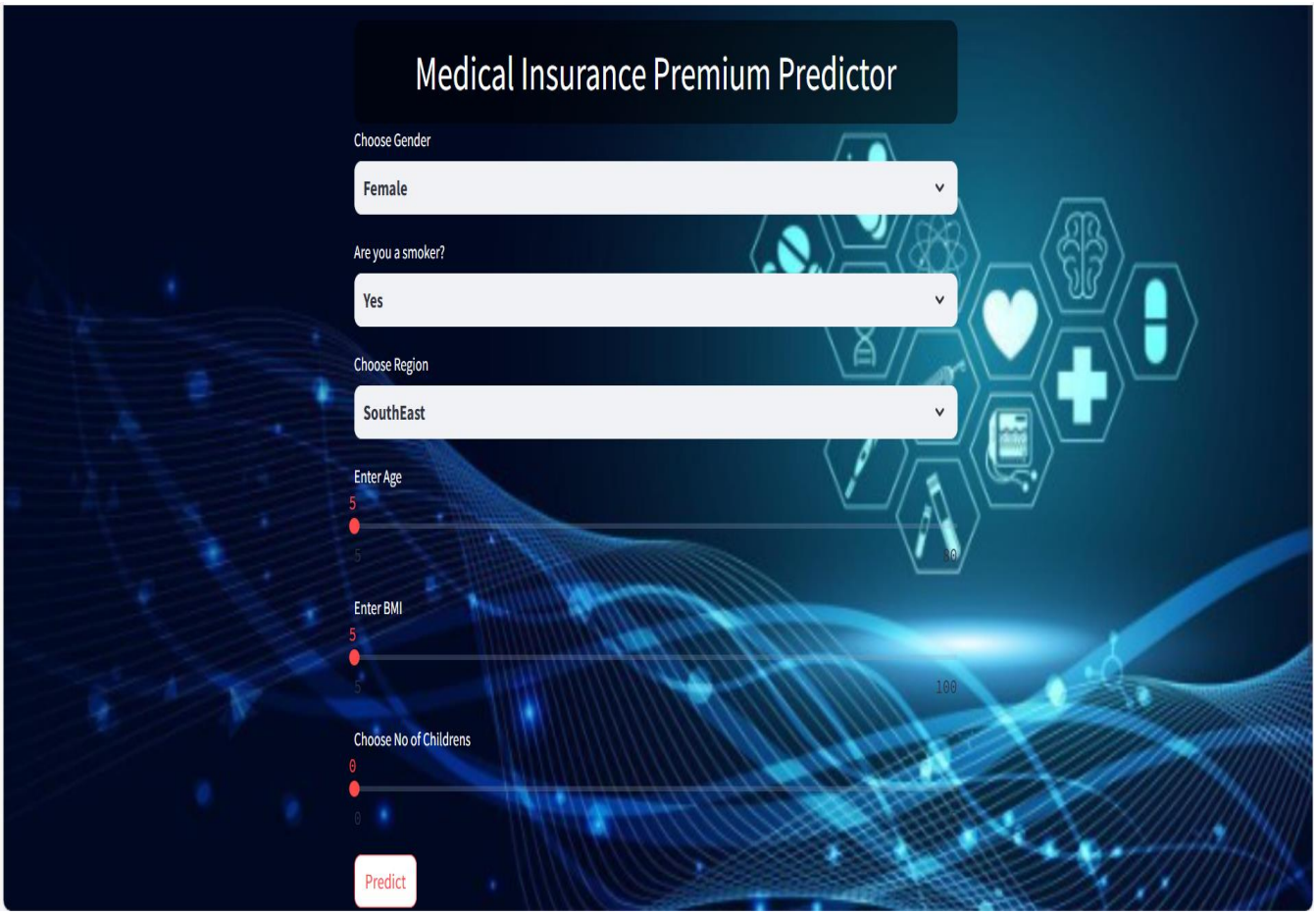
```
# Set background
add_bg_from_local("medical_bg.jpg")

# Title
st.markdown('<div class="main-title">Medical Insurance Premium Predictor</div>', unsafe_allow_html=True)

# Inputs
gender = st.selectbox('Choose Gender', ['Female', 'Male'])
smoker = st.selectbox('Are you a smoker?', ['Yes', 'No'])
region = st.selectbox('Choose Region', ['SouthEast', 'SouthWest', 'NorthEast', 'NorthWest'])
age = st.slider('Enter Age', 5, 80)
bmi = st.slider('Enter BMI', 5, 100)
children = st.slider('Choose No of Childrens', 0, 5)

# Prediction
if st.button('Predict'):
   gender = 0 if gender == 'Female' else 1
   smoker = 1 if smoker == 'Yes' else 0
   region = {'SouthEast': 0, 'SouthWest': 1, 'NorthEast': 2, 'NorthWest': 3}[region]
   input_data = np.asarray((age, gender, bmi, children, smoker, region)).reshape(1, -1)
   predicted = model.predict(input_data)[0]
st.markdown(f'<div class="result-box">Insurance Premium will be {round(predicted, 2)} USD Dollars</div>', unsafe_allow_html=True)
```

OUTPUT:


Medical Insurance Premium Predictor
Choose Gender
Female
Are you a smoker?
Yes
Choose Region
SouthEast
Enter Age
5
5
100
Enter BMI
5
5
100
Choose No of Childrens
0
0
Predict


Insurance Premium will be 15377.88 USD Dollars

CERTIFICATION:

TRIOS TECHNOLOGIES PVT.LTD

Rani Arcade, #24/5, 2nd Floor, 1st Avenue, Ashok Nagar, Chennai - 83.

INTERNSHIP PROGRAM

Certificate Of Completion

This is to Certify that ~~Mr.~~/Ms. UMA VYSHNAVI - G

of RAJALAKSHMI ENGINEERING COLLEGE

has successfully completed the Internship On DATA SCIENCE

..... in our Organization from 17-06-2025 to 01-07-2025

During the Period ~~His~~ / Her Performance was good and we wish for future endeavours.

TRIOS TECHNOLOGIES PVT. LTD. Chennai

Technical Head

Phone : 044 4854 0077

Web: www.triostechnology.com

Email:support@triostechnology.com