**Aiml Mini Project Report**

on

**Insurance Cost Predictor**

by

Manish Shashikant Jadhav

2023301005

Mayur Krishna Solankar

2023301018

Vishesh Bhimji Savani

2022300100

**Guide Name:**

Prof. Swapnali Kurhade



**Computer Engineering Department**

Bharatiya Vidya Bhavan's

Sardar Patel Institute of Technology

Munshi Nagar, Andheri(W), Mumbai-400058

University of Mumbai

Nov 2024

# Introduction

## Problem Statement

To develop a machine learning-based predictive model for accurately estimating health insurance costs by analyzing key variables such as age, BMI, number of children, gender, smoking status, and region, addressing the challenges of manual cost estimation and improving the efficiency of insurance pricing strategies.

Key Challenges:

* Manual insurance cost calculation is time-consuming
* High potential for human error
* Complex interactions between multiple factors affecting insurance charges
* Need for an automated, data-driven approach to cost prediction

## Literature Survey/Market Survey

**Research Paper Link:**

https://www.researchgate.net/publication/348559741\_Predict\_Health\_Insurance\_Cost\_by\_using \_Machine\_Learning\_and\_DNN\_Regression\_Models

* 1. **Research Objectives**
* Analyze factors influencing health insurance costs
* Implement linear regression and random forest models
* Generate predictive model for insurance cost estimation

# Methodology

**3.1 Data Source**

* **Dataset:** Kaggle Medical Cost Personal Dataset
* **Features:** Age, BMI, number of children, gender, smoking status, region
* **Target Variable:** Medical insurance charges

**3.2 Data Preprocessing**

* Cleaned and prepared dataset
* Converted categorical variables to numeric values
* Split dataset into training and testing sets
* Created pickle file for random forest model

**3.3 Machine Learning Models**

1. Linear Regression
2. Random Forest Regression

**3.4 Technology Stack**

* Backend: Django (Python)
* Frontend: React.js
* Machine Learning: scikit-learn, pickle

**3.5 System Architecture**

**3.5.1 Backend Development (Django)**

* Create Django REST Framework for API endpoints
* Implement machine learning model loading from pickle file
* Develop prediction logic for insurance cost estimation
* Handle data validation and preprocessing
* Create secure API for model inference

**3.5.2 Frontend Development (React)**

* Design responsive user interface
* Create input forms for insurance variables
* Implement state management
* Develop prediction result display
* Connect frontend with Django backend API

**3.5.3 Machine Learning Pipeline**

* Data preprocessing
* Linear regression model
* Random forest regression model
* Model serialization using pickle
* Performance evaluation metrics

**3.5.4 Integration Approach**

* Django backend serves machine learning model
* React frontend consumes prediction API
* Seamless communication between frontend and backend
* Secure data transmission

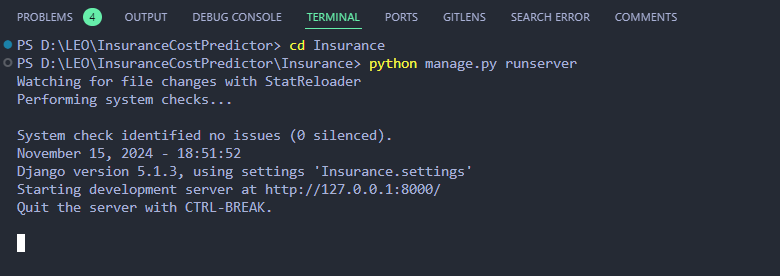
# Implementation

**Github Repo Link:**

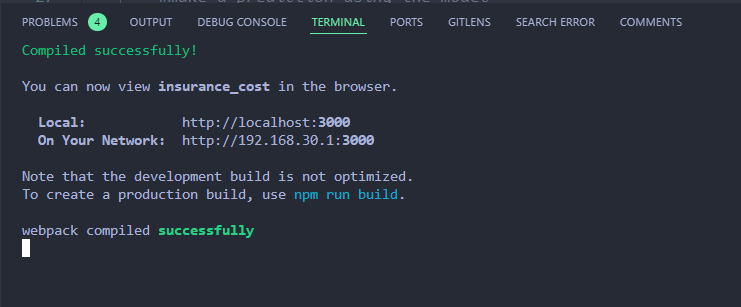
https://github.com/manishjadhav9/InsuranceCostPredictor

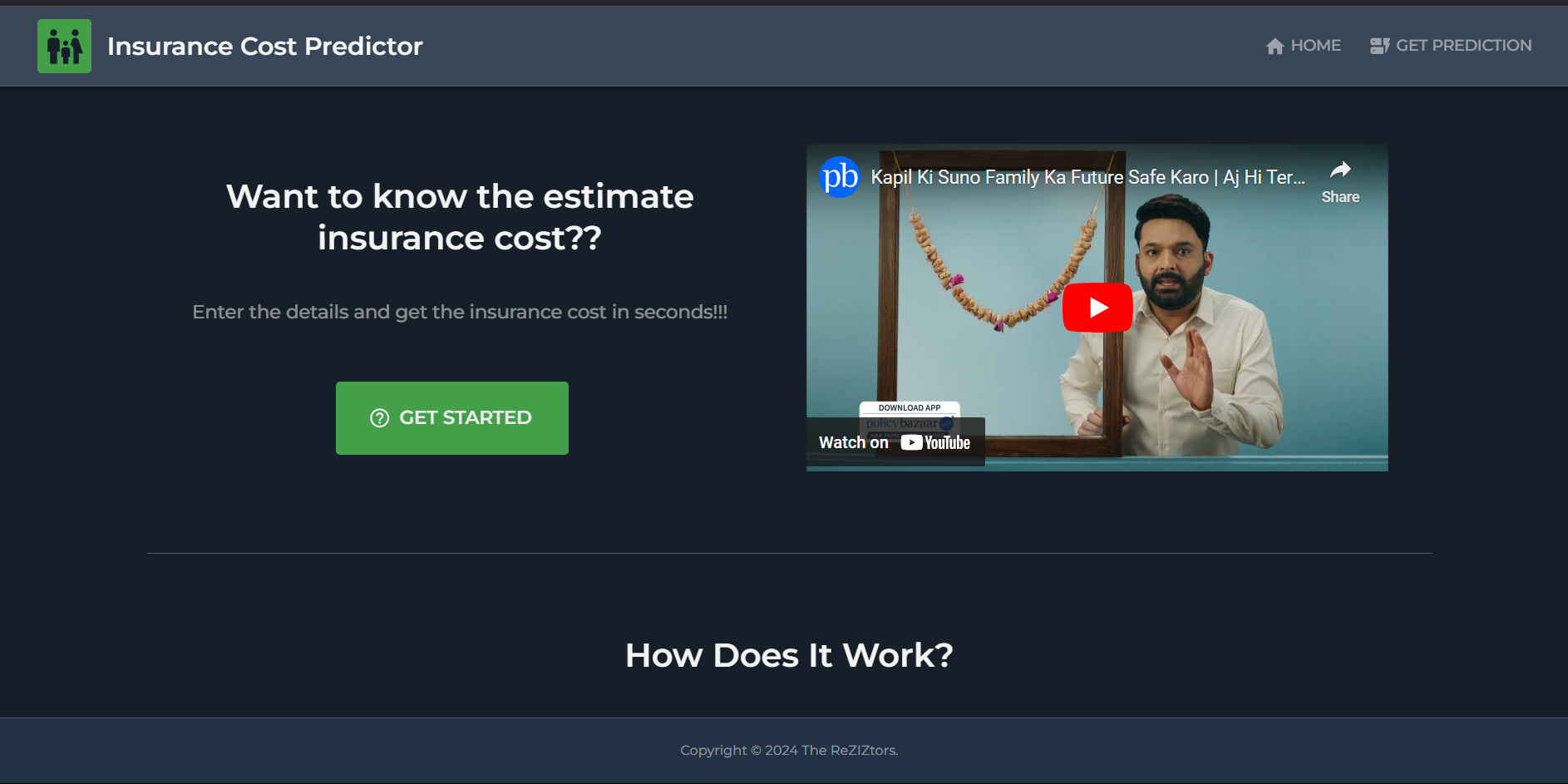
**Collab Link:** <https://colab.research.google.com/drive/1KF87J81hl5vbbIUWB_9pvp9Vj8L0BMGC>

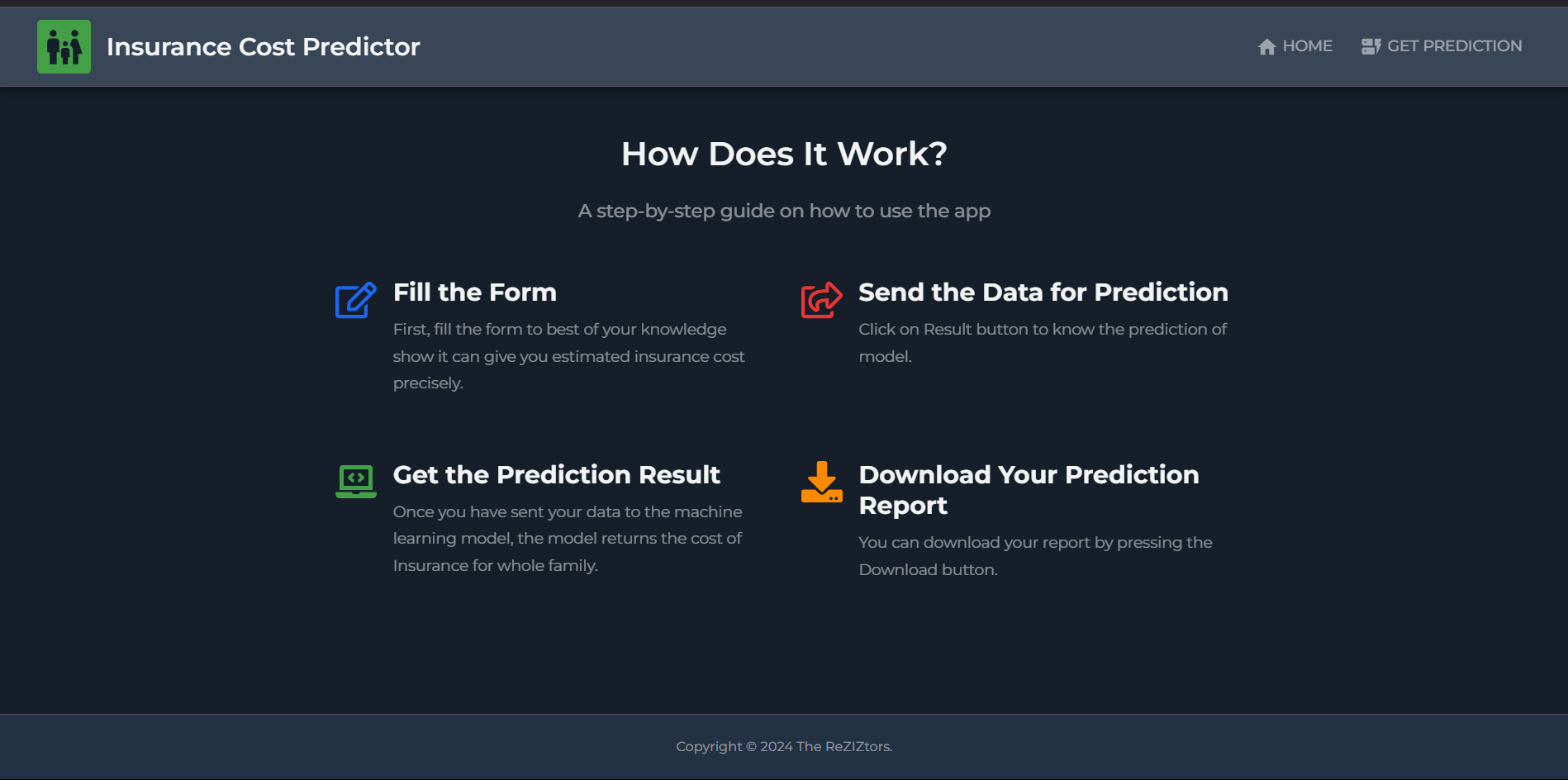
1. **Start Server:**

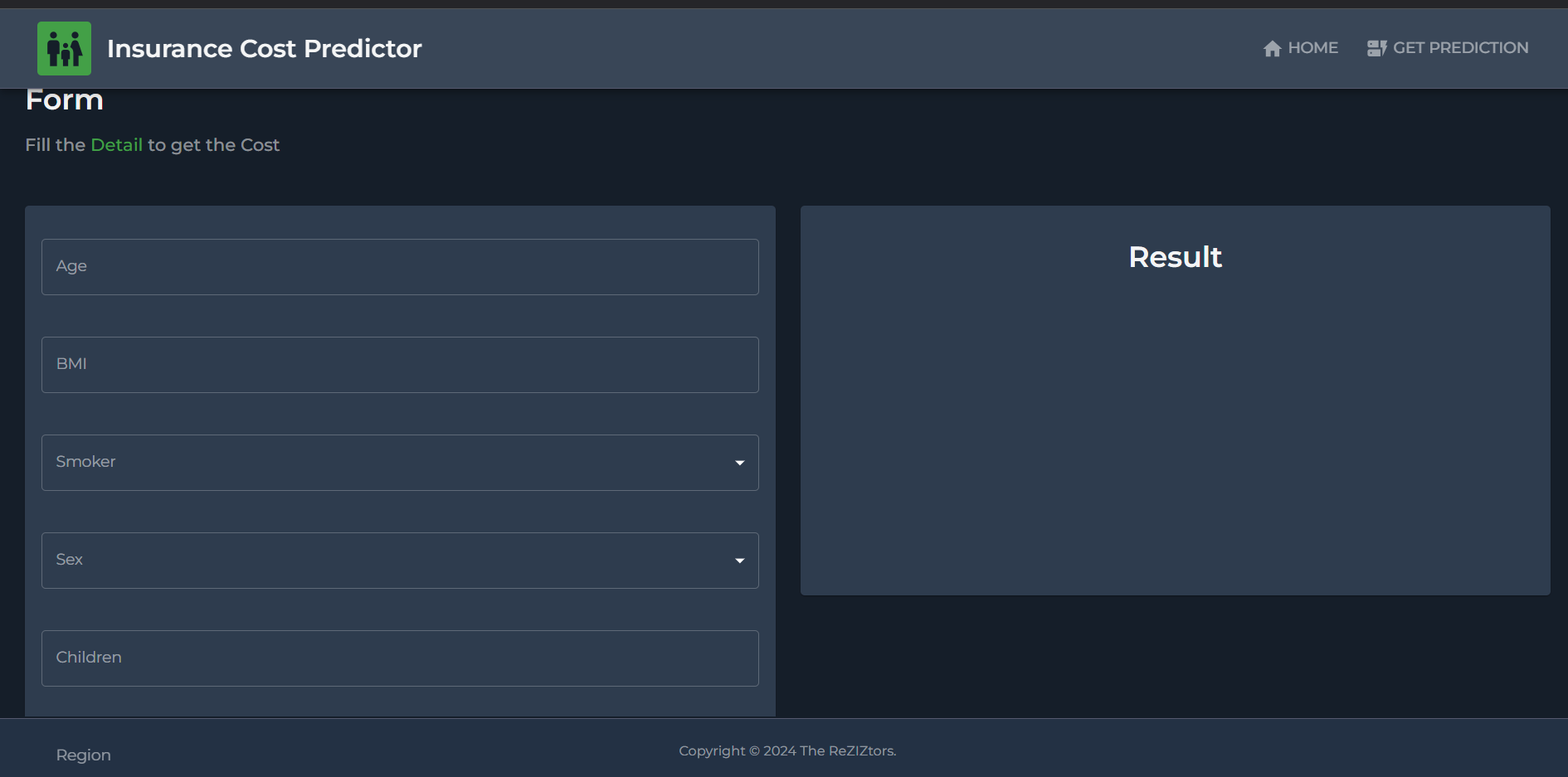
****

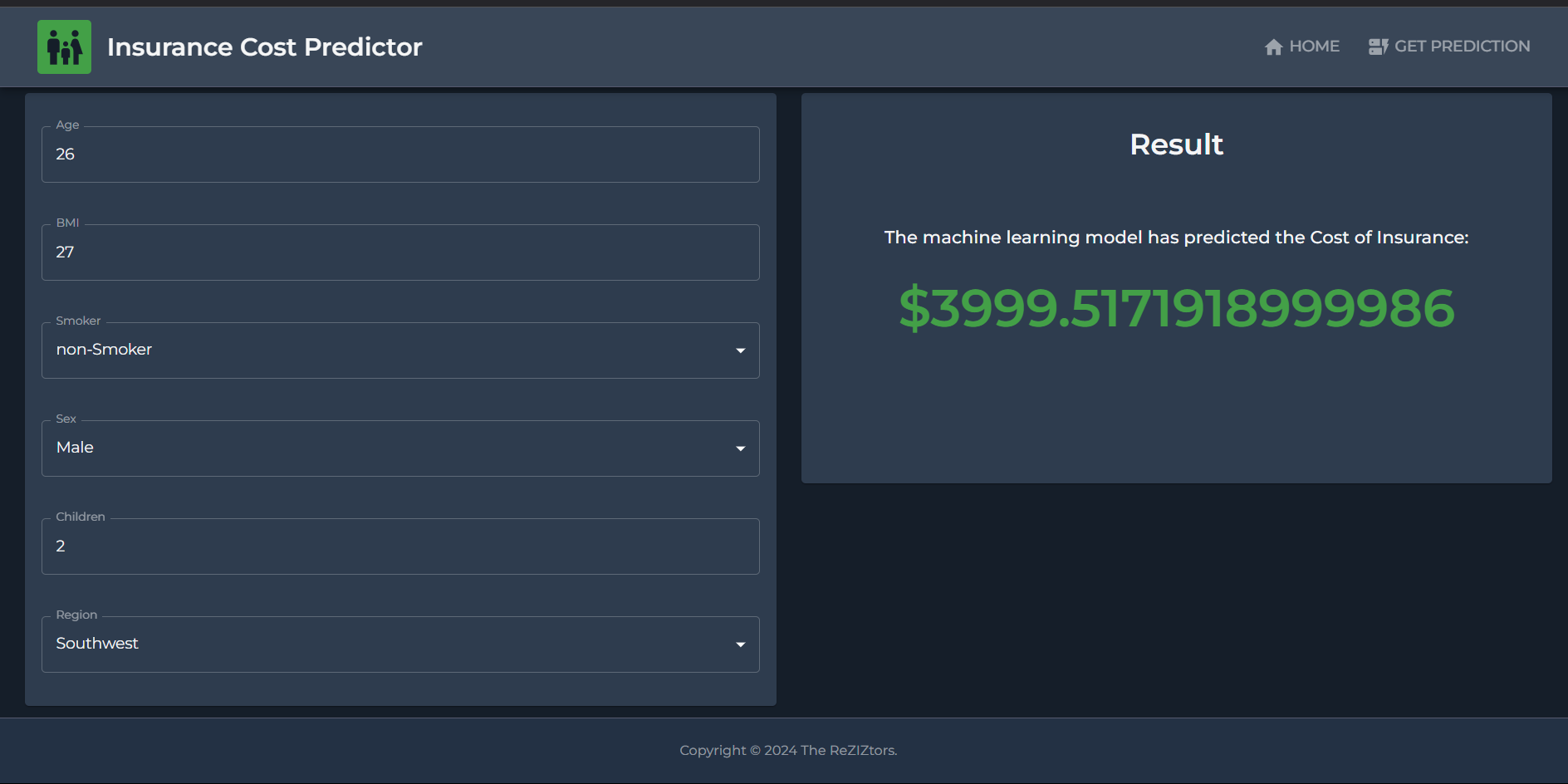
1. **Start frontend:**

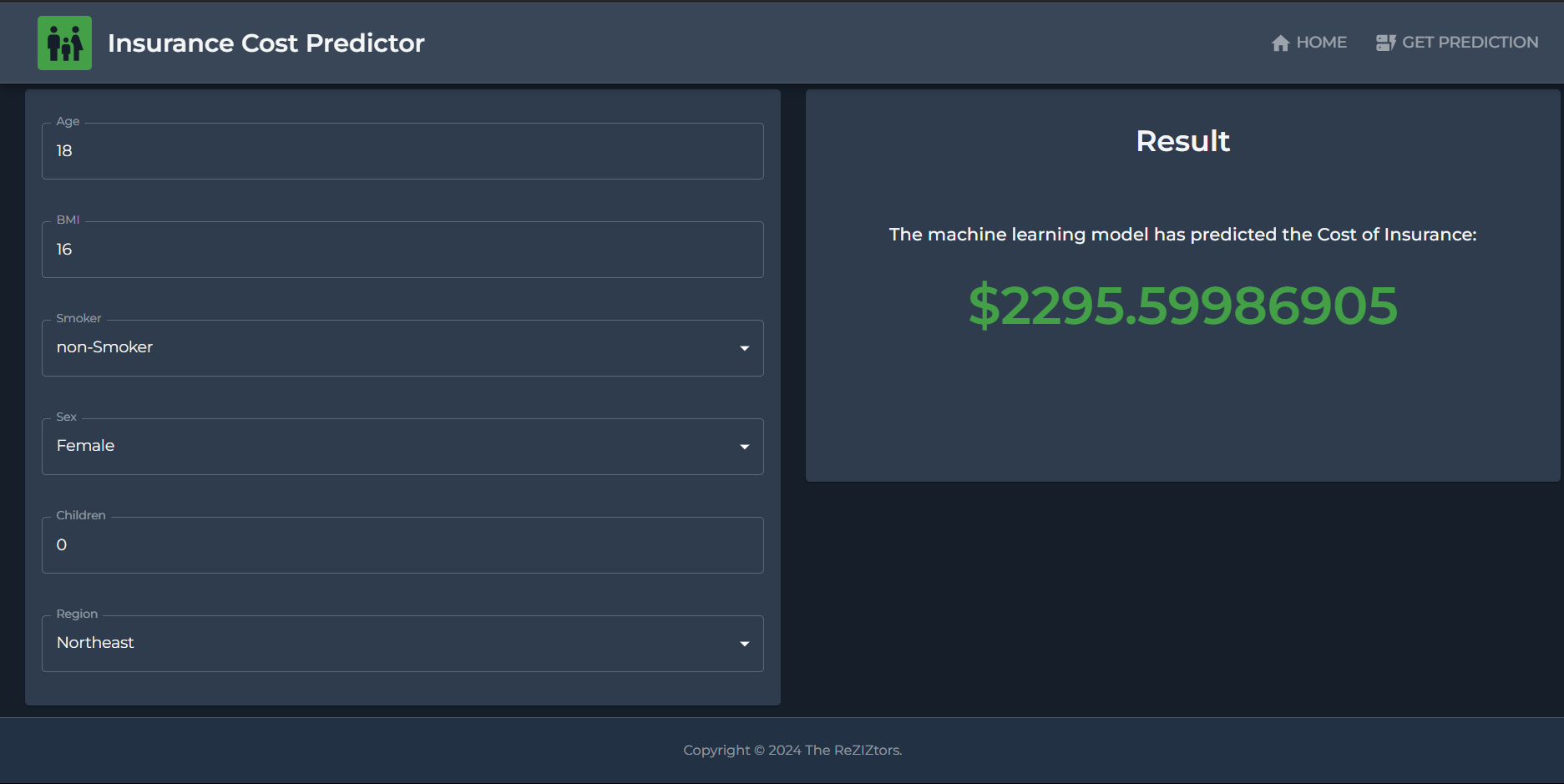
****

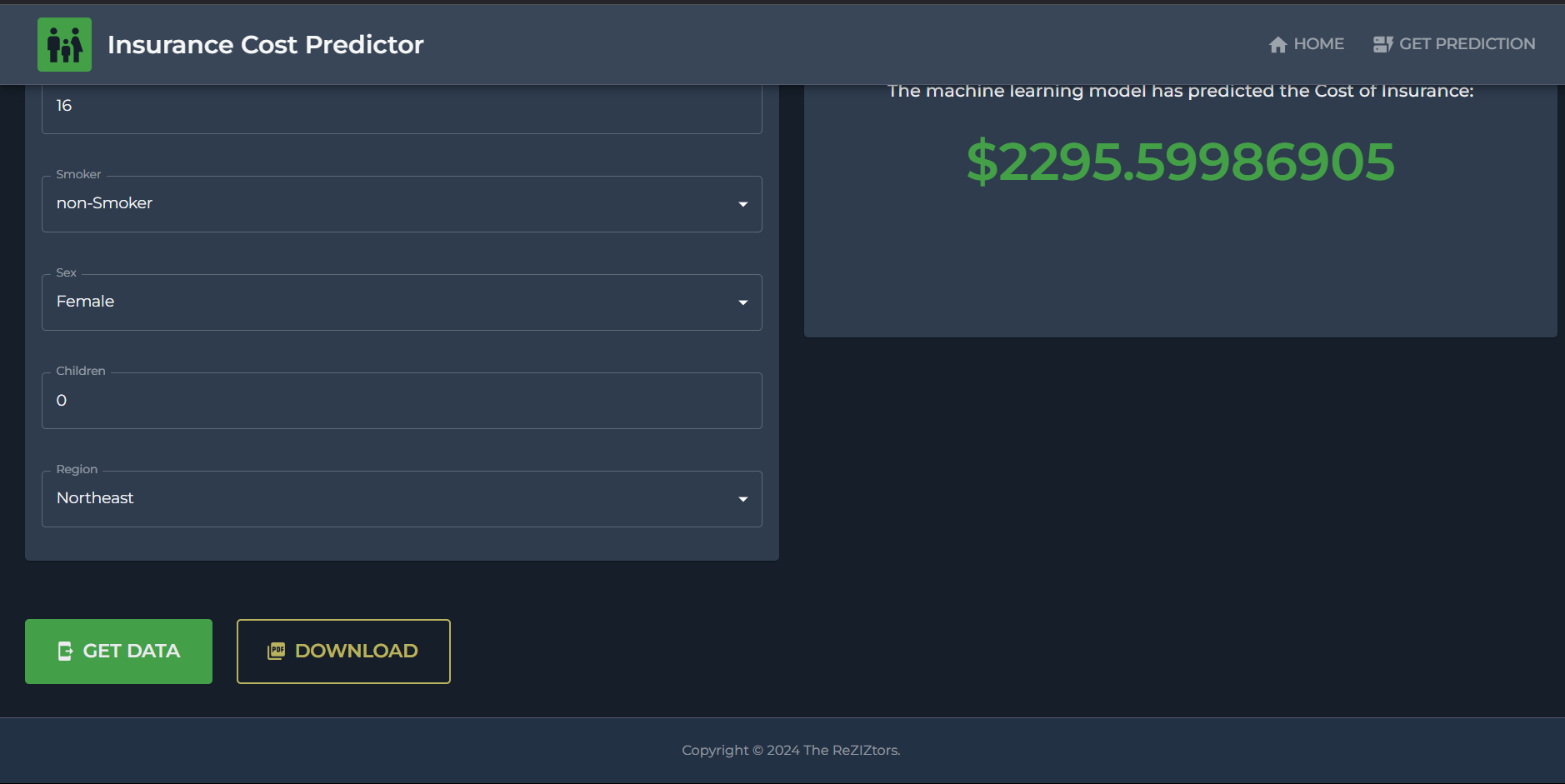
****

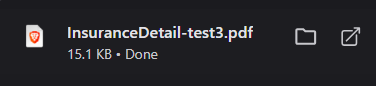
****

****

****

****

1. **Download the result:** ****

****

# Conclusion

The **Insurance Cost Predictor Project** is a testament to how technology can streamline financial planning and decision-making in the insurance sector. By leveraging machine learning algorithms, this project enables accurate predictions of insurance costs based on key factors such as age, gender, BMI, smoking habits, and more.

This tool not only benefits customers by offering transparency in cost estimation but also aids insurance companies in optimizing their pricing strategies, enhancing customer satisfaction, and promoting fairness. The project showcases the immense potential of data science in solving real-world problems, emphasizing the importance of predictive analytics in improving efficiency and decision-making across industries.

# References

* <https://www.kaggle.com/datasets/mirichoi0218/insurance?select=insurance.csv>
* https://www.researchgate.net/publication/348559741\_Predict\_Health\_Insurance\_Cost\_by\_using \_Machine\_Learning\_and\_DNN\_Regression\_Models