**MEDICAL INSURANCE PRICE PREDICTION USING GPU PROGRAMMING**

**PROJECT REPORT**

**INDEX**

Table of Contents

[1.Introduction: 3](#_Toc152935872)

[2. Design: 3](#_Toc152935873)

[3. Functions 5](#_Toc152935875)

[4. Workflow 7](#_Toc152935877)

[5. How to Execute 8](#_Toc152935878)

[6. Dependencies 9](#_Toc152935879)

[7. Conclusion 10](#_Toc152935880)

[8. References 11](#_Toc152935881)

# INTRODUCTION:

In the fast-paced world of medical informatics, predictive analytics can play a pivotal role in understanding and managing healthcare costs. This project presents a cutting-edge approach to forecast medical expenses using machine learning algorithms accelerated by GPU programming. By leveraging the powerful computational capabilities of GPUs, the project aims to achieve rapid data analysis and training of deep learning models, ultimately providing accurate predictions of medical costs and develop a model to predict medical costs for individuals based on various health and demographic factors. By analyzing historical insurance data, the model can learn patterns and relationships between these factors and predict charges for future patients. This information can be valuable for patients, insurance companies, and healthcare providers in making informed decisions.

# 2. DESIGN:

The project utilizes a two-stage approach:

* Exploratory Data Analysis (EDA): This stage involves analyzing the medical insurance dataset to understand the distribution of features like age, sex, BMI, smoker status, and region. It also explores the correlation between these features and the target variable, which is the medical cost (charges).
* Model Building: Several machine learning models, including K-Nearest Neighbors, Linear Regression, Support Vector Regression, Decision Trees, Random Forests, and a Neural Network, are trained on the preprocessed data. The performance of each model is evaluated using Mean Absolute Error (MAE) on a held-out test set. The model with the lowest MAE is chosen as the final prediction model

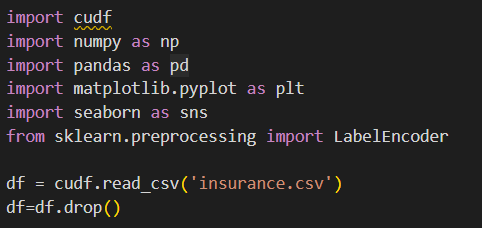
**Phases in the design:**

* Data collection: Collect historical loan information on the following topics: loan characteristics (amount, purpose, length), applicant demographics, job status, and income; and loan outcomes (approved, denied).
* Data Preprocessing: In order to clean up and prepare the data, data preparation includes handling outliers, missing values, and encoding categorical variables. Feature engineering can also be used to create new features out of existing ones.
* Model Selection: Choose the most effective machine learning algorithms based on the project's objectives and the characteristics of the data. Popular choices include Logistic Regression, Decision Trees, Random Forests, Support Vector Machines (SVM), and Gradient Boosting.
* Model Training: Split the data into sets for the model's testing and training. Use the training set to train the model and find the link between characteristics and loan outcomes.
* Model Evaluation: Evaluate the model's performance on the hidden testing data using measures like accuracy, precision, recall, and F1-score.

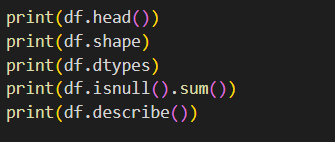
# Model Refinement: Using the evaluation findings as a guide, tweak hyperparameters, try out other algorithms, or use feature engineering techniques to make the model better.

# 3. FUNCTIONS:

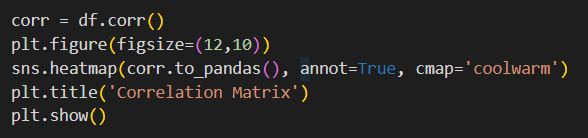
Data Loading and Cleaning:



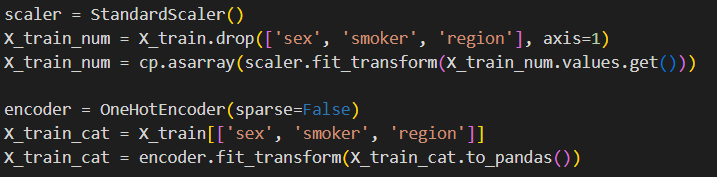
Basic Exploration:



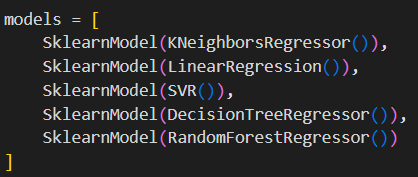
Correlation Matrix:



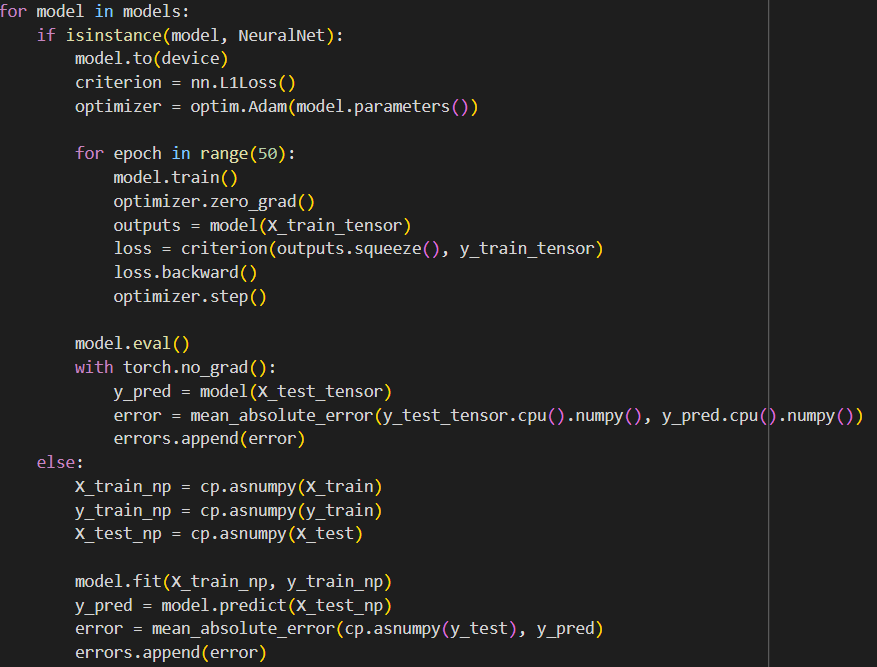
Feature Selection:



Model Selection:



Model Prediction:



# 4. WORKFLOW:

* Data Collection: Obtain a medical insurance dataset containing features like age, sex, BMI, smoker status, region, and medical charges.
* Data Preprocessing: Clean and prepare the data for analysis.
* Exploratory Data Analysis: Gain insights into the data and understand the relationships between features.
* Model Building: Train and evaluate different machine learning models for medical cost prediction.
* Model Selection: Choose the model with the best performance based on the evaluation metric (MAE).
* Prediction: Use the chosen model to predict medical costs for new data points.

# 5. HOW TO EXECUTE:

* Ensure you have the required libraries installed (pandas, scikit-learn, matplotlib, seaborn, PyTorch, etc.).
* Load the project code and specify the path to your medical insurance data (CSV file).
* Run the script sequentially. It will perform data preprocessing, EDA, model training, evaluation, and potentially save the final model for future predictions.
* Have access to an environment with the necessary GPU and CUDA toolkit installed.
* Execute the EDA scripts to perform data analysis and preprocessing.
* Run the model building scripts to train and validate the machine learning models.
* Use the chosen model to generate predictions on new data.

# 6. DEPENDENCIES:

**Software:**

* Programming Language: Python
* Libraries:
  + Data Processing: pandas (manipulation), NumPy (calculations)
  + Machine Learning: scikit-learn (models)
  + Data Visualization: matplotlib/seaborn (creating charts)
  + Deep Learning: CuDf and PyTorch (GPU-accelerated processing)
* Development Environment: Tools like Visual Studio Code, Jupyter Notebook, Anaconda, or Google Colab

**Hardware:**

* Graphics Card (GPU): NVIDIA with CUDA compatibility (check <https://developer.nvidia.com/cuda-gpus> for supported models)

**Software for GPU Acceleration :**

* CUDA Toolkit: Core development environment for creating GPU applications (download from <https://developer.nvidia.com/cuda-toolkit>)
* CUDA Libraries: Additional libraries for specific tasks (e.g., cuDNN for deep learning, cuBLAS for linear algebra)

**Operating System:**

* A compatible version of your chosen operating system

**Drivers:**

* Latest NVIDIA drivers for your specific GPU

# 7. CONCLUSION:

The application of machine learning techniques for medical cost prediction can be very beneficial. By analyzing historical data, the model can provide valuable insights into factors influencing medical expenses. This information can be used by various stakeholders to make informed decisions regarding healthcare costs. The use of GPUs in machine learning projects like Medical Insurance Price Prediction signifies a remarkable leap towards accelerated data processing and sophisticated model training. This project embraces GPU programming to offer substantial improvements in speed and accuracy when forecasting medical costs, ultimately enhancing decision-making in healthcare management.

# 8. REFERENCES:

https://ieeexplore.ieee.org/document/9824201

<https://www.w3schools.com/python/python_reference.asp>

<https://www.python.org/>

<https://realpython.com/python-application-layouts/>

<https://scikit-learn.org/stable/>

<https://pandas.pydata.org>

<https://www.geeksforgeeks.org/getting-started-with-pytorch/>

<https://www.geeksforgeeks.org/medical-insurance-price-prediction-using-machine-learning-python>

https://docs.nvidia.com/cuda/pdf/CUDA\_C\_Programming\_Guide.pdf