

វិទ្យាស្ថានជាតិប្រៃសណីយ៍ ទូរគមនាគមន៍

បច្ចេកវិទ្យាគមនាគមន៍ និងព័ត៌មាន

**NATIONAL INSTITUD OF POSTS, TELECOMS & ICT**

**School of Computer Science**

Major : Machine Learning

Lecturer : Mr. Leang Sotheara

**Topic : Health Insurance Cost Prediction**

Members:

⬩ Un Thaisan

⬩ Sok Saorath

⬩ Khov Manny

⬩ Run Seyha

⬩ Neak Bora

**January 2020**

# **ACKNOWLEDGMENTS**

First of all, we would like to thank teacher Mr. Leang Sotheara that has been teaching us on Machine learning field for the past three months and help us understand clearly about machine learning process as well as coding. With your guidance, we have learnt so much about how machine learning works and how to build models that help predicting data which is very helpful in everyday life problems.

Additionally, we also would like to thank Mr. Keo Saly that helped assisting on our education during this term. When we have problems or obstacles with learning, you open your arms and help us throughout everything.

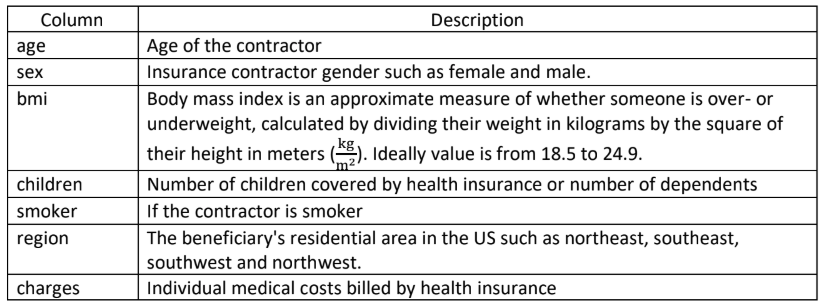
Lastly, we would like to thank all the classmates for helping and learning together as well as supporting with all the kindness.

**I. Problems of the Project**



This project aims to have the students experience in a practical problem of Machine Learning used for predicting the cost of individual health insurance. With this project, they will learn how to analyze and formulate the problem and how to apply Machine Learning techniques in term of preprocessing data, training and evaluation the models.

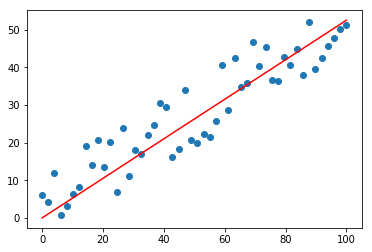
**II**. **Dataset of the Project**



That all of this data we think it is very important for use to predict the cost of individual health insurance.

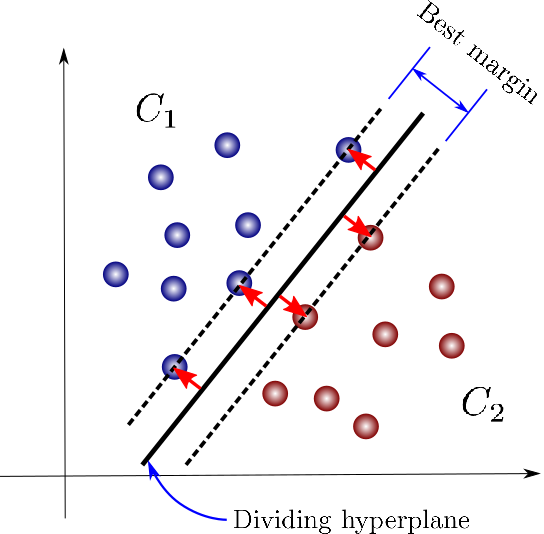
**III. Learning Algorithms**

1. **Linear Regression Algorithm**



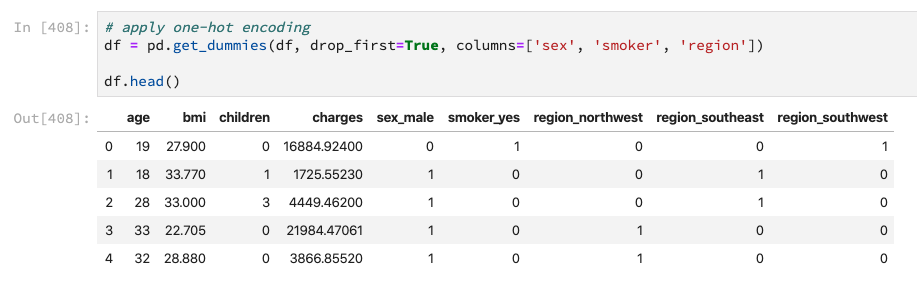
Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting.

1. **Support Vector Machine Algorithm**



Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

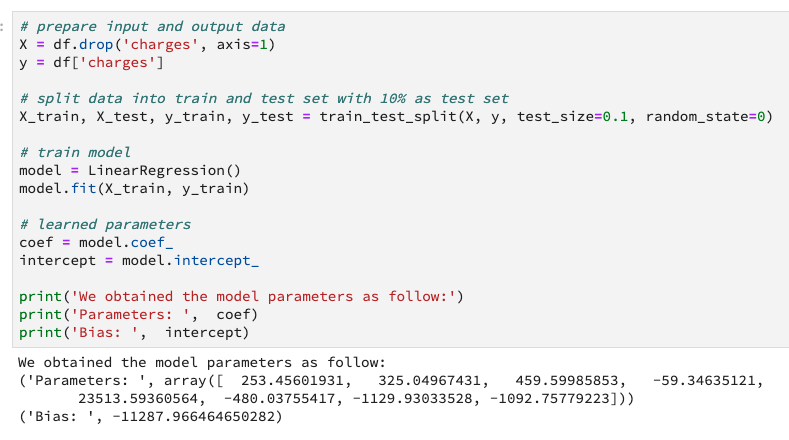
**IV. Data Preprocessing**



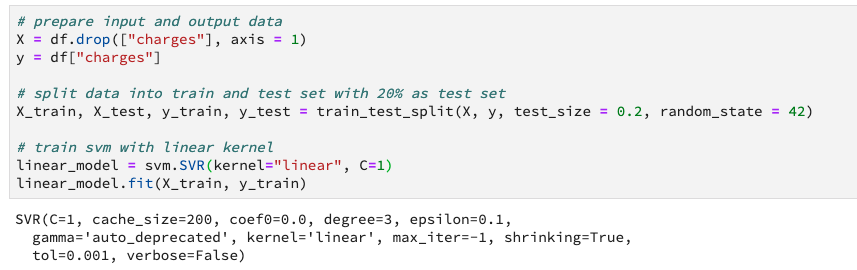
We have to apply one-hot encoding on columns sex, smoker and region.

**V. Data Training**

**1. Using Linear Regression Algorithm**

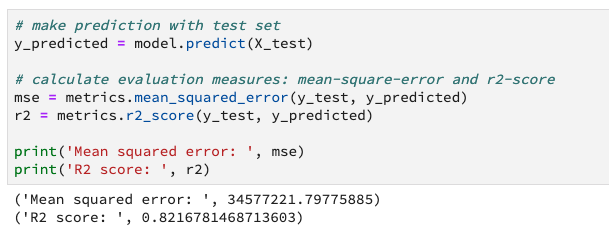


**2. Using Support Vector Machine Algorithm**



**VI. Evaluation the Performance of the Models**

1. **Using Linear Regression Model**



**2. Using Support Vector Machine Model**

