**INTRODUCTION**

* 1. **Overview**

The purpose of this project is to explore the use of machine learning algorithms to predict the prices of annual health insurance premiums given the specifications of the beneficiary. That is, given a health insurance contract and information about a person, can we accurately predict how much it will cost per year?

Building predictive models requires time, effort, and knowledge of algorithms to create effective systems that can predict the outcome accurately. IBM has introduced AutoAI, which automates all of the tasks involved in building predictive models for different requirements.

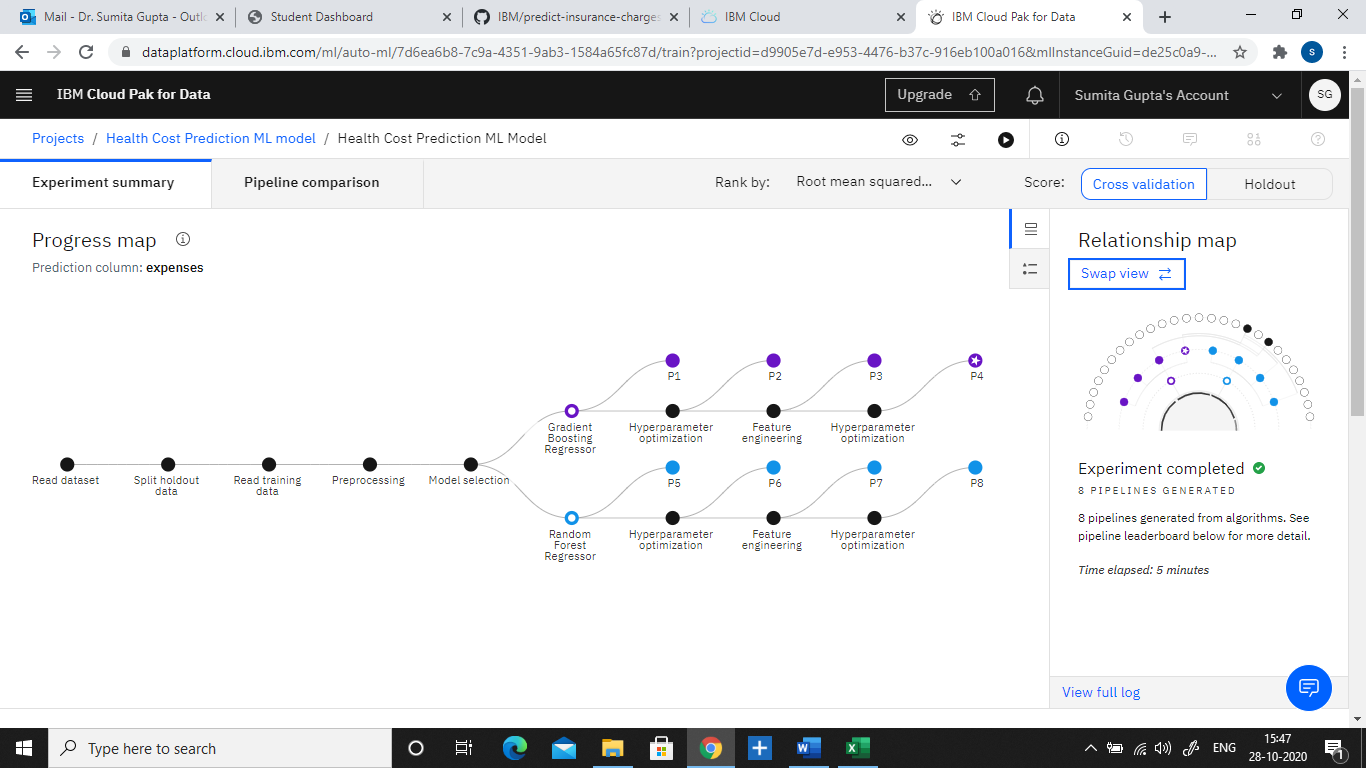
**1.2 Purpose**

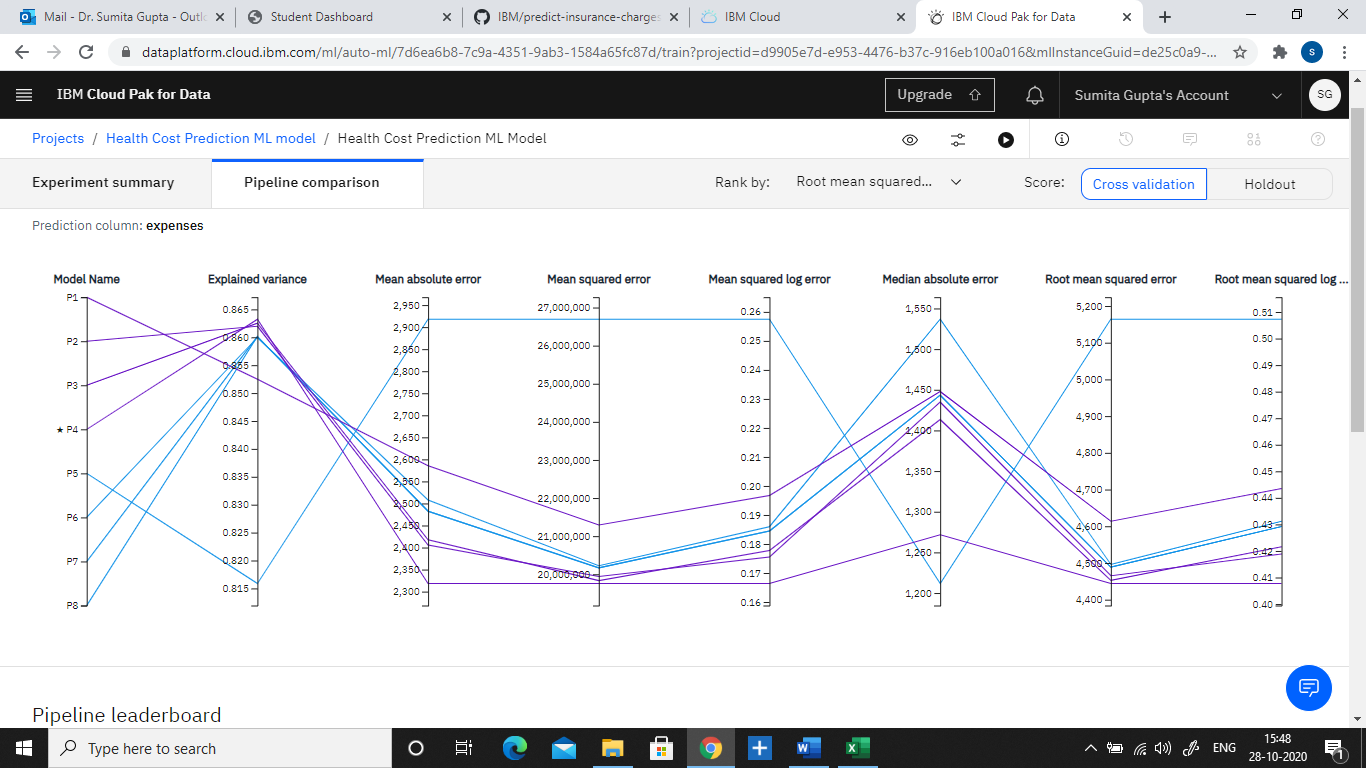
Insurance companies consider several factors when calculating insurance premiums. The purpose of this project is to study the effects of age, smoking, BMI, gender, and region to determine how much of a difference these factors can make on insurance premium. By using our application, customers see the radical difference their lifestyle choices make on their insurance charges. By leveraging AI and machine learning, we help customers understand just how much smoking increases their premium, by predicting how much they will have to pay within seconds.

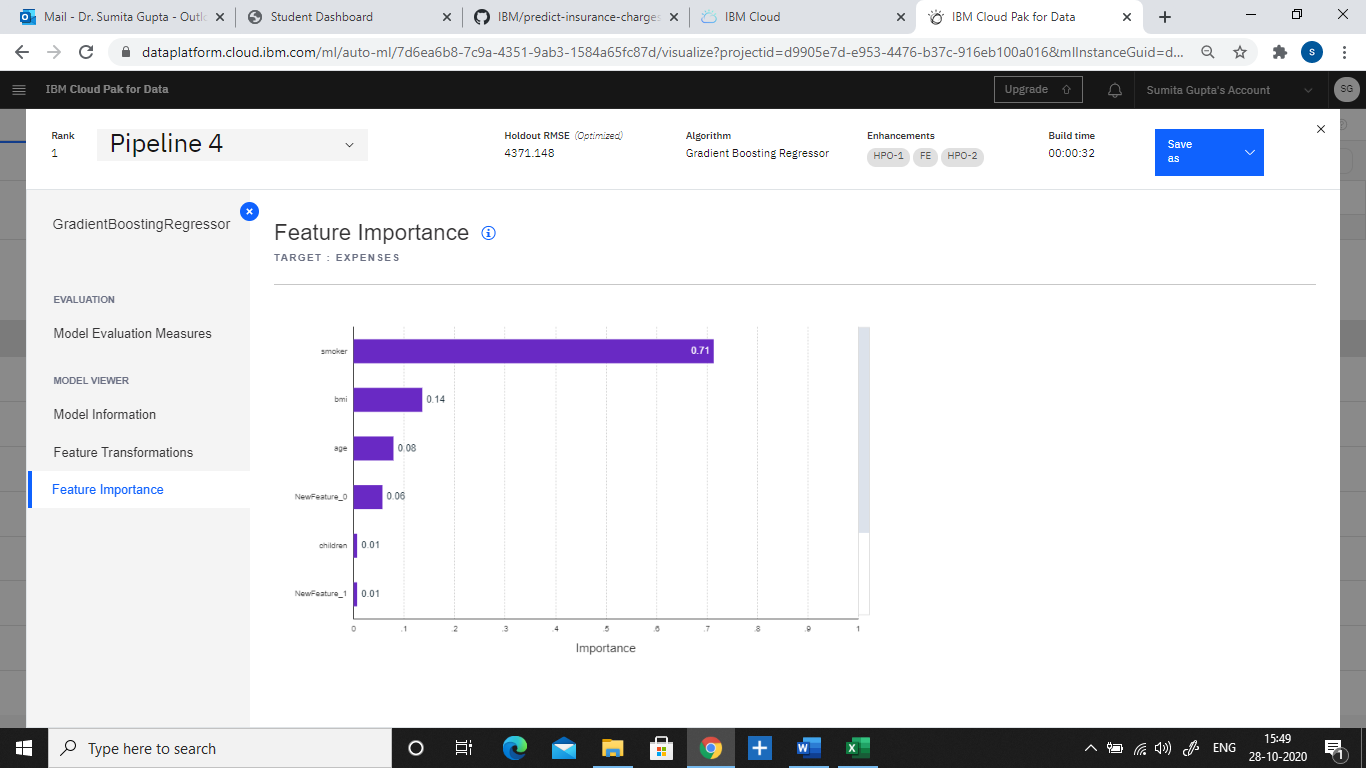
**2. FLOW DESCRIPTION**

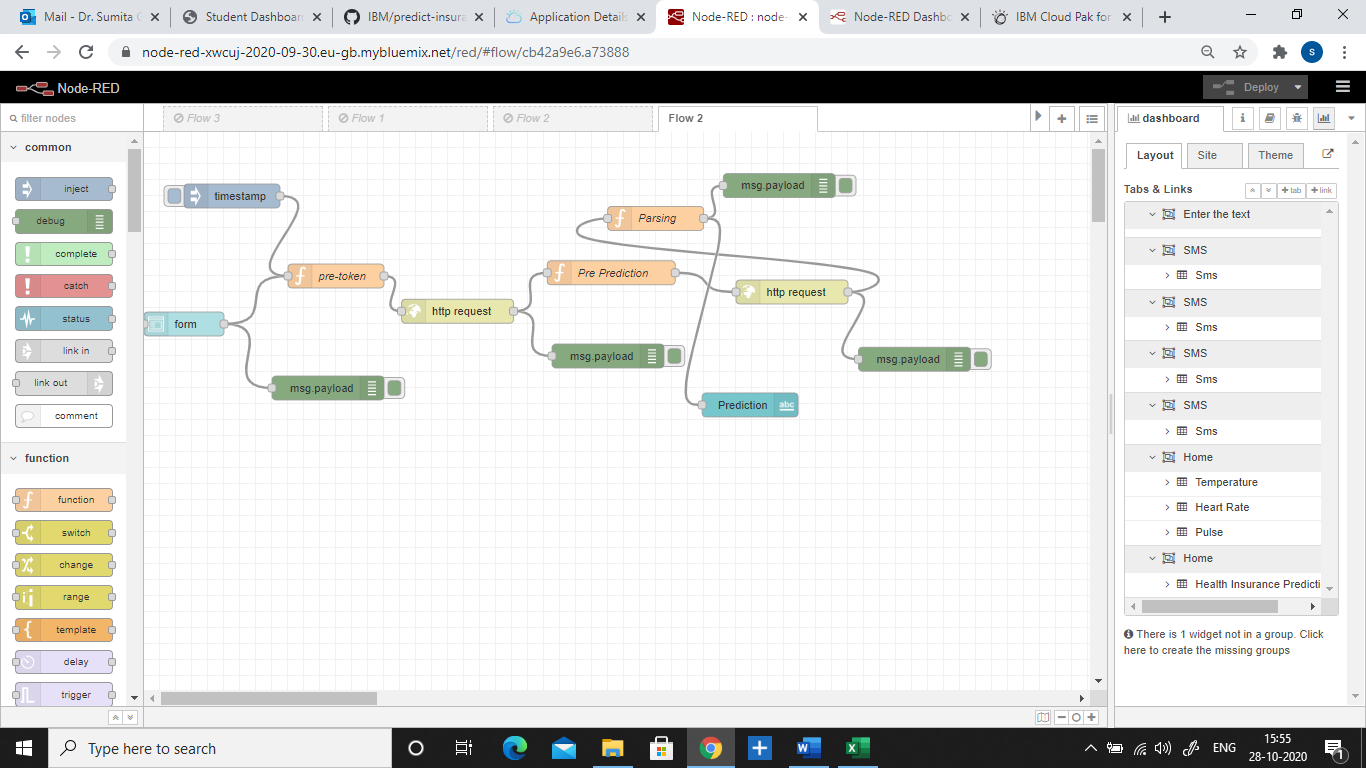
* The user creates an IBM Watson Studio Service on IBM Cloud.
* The user creates an IBM Cloud Object Storage Service and adds that to Watson Studio.
* The user uploads the insurance premium data file i.e. Dataset into Watson Studio.
* The user creates an AutoAI Experiment to predict insurance premium on Watson Studio
* AutoAI uses Watson Machine Learning to create several models, and the user deploys the best performing model.
* The user uses the Flask web-application to connect to the deployed model and predict an insurance charge.

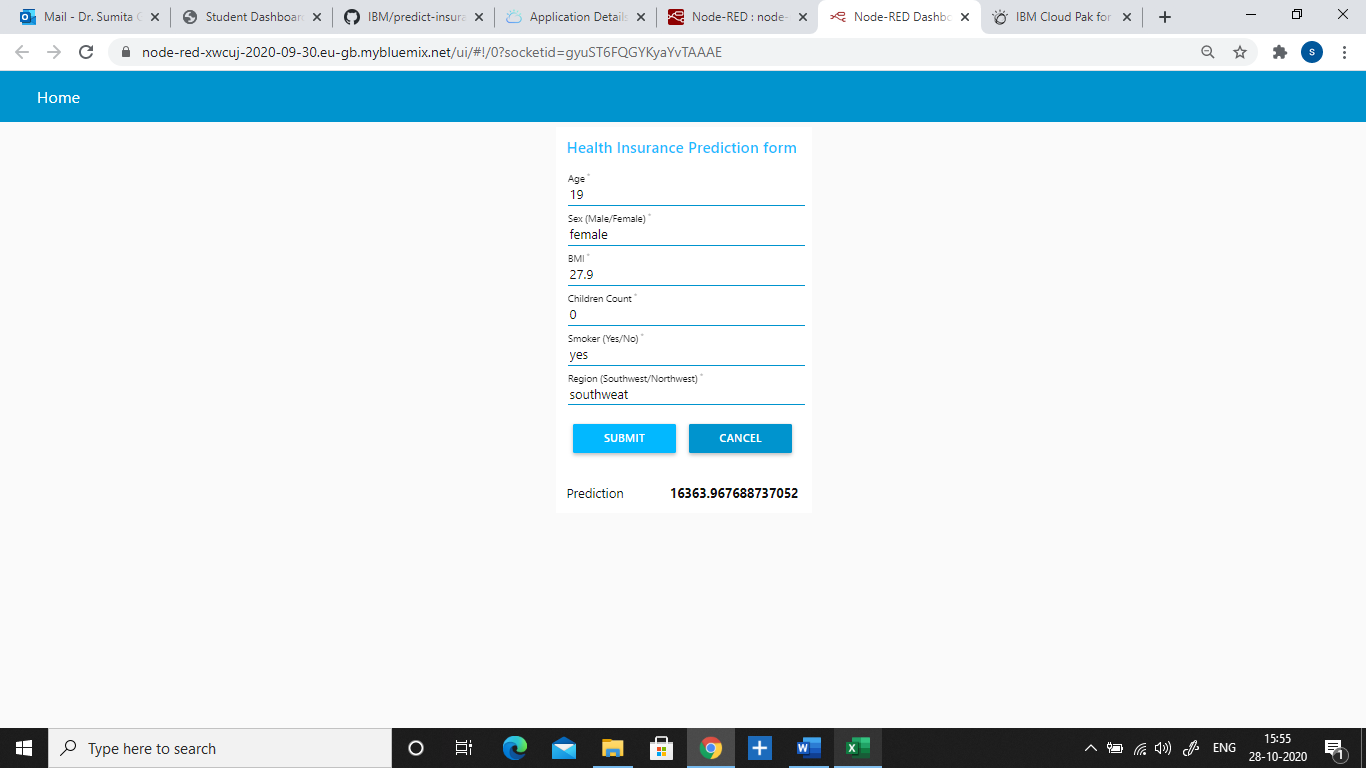
**3. SCREENSHOTS**











<https://dataplatform.cloud.ibm.com/dashboards/11e68cf5-3733-4bd0-b6d2-09f7eb741e4b/view/7427ff2d018b31fe52b6c0e407cd250f7d32705ae6bbd60a85867b490b677797f36a14c2c8264a5c8f430463f3bd1258cf>

**4. EXPERIMENTAL INVESTIGATION**

**4.1 Data Set info:** The data used in this project contains 7 columns and 1338 rows of different users. 90% data is used for training and 10% data is used for testing purpose.

**4.2 Dataset Attributes:**

**1. Age:** Age of primary beneficiary

**2. Sex:** Primary beneficiary’s gender

**3. BMI:** Body mass index (providing an understanding of the body, weights that are relatively high or low relative to height)

**4. Children:** Number of children covered by health insurance / Number of dependents

**5. Smoker:** Smoking (yes, no)

**6. Region:** Beneficiary’s residential area (southwest, northwest)

**7. Expenses:** Individual medical costs billed by health insurance

**4.3 Result Analysis**

By using a Cognos Dashboard feature which is available on IBM cloud from "Add to Project" option in Watson Studio project we can analyse the various relationships between the factors. For example how sex, BMI, number of of children and smoking might influence the insurance premium.

