

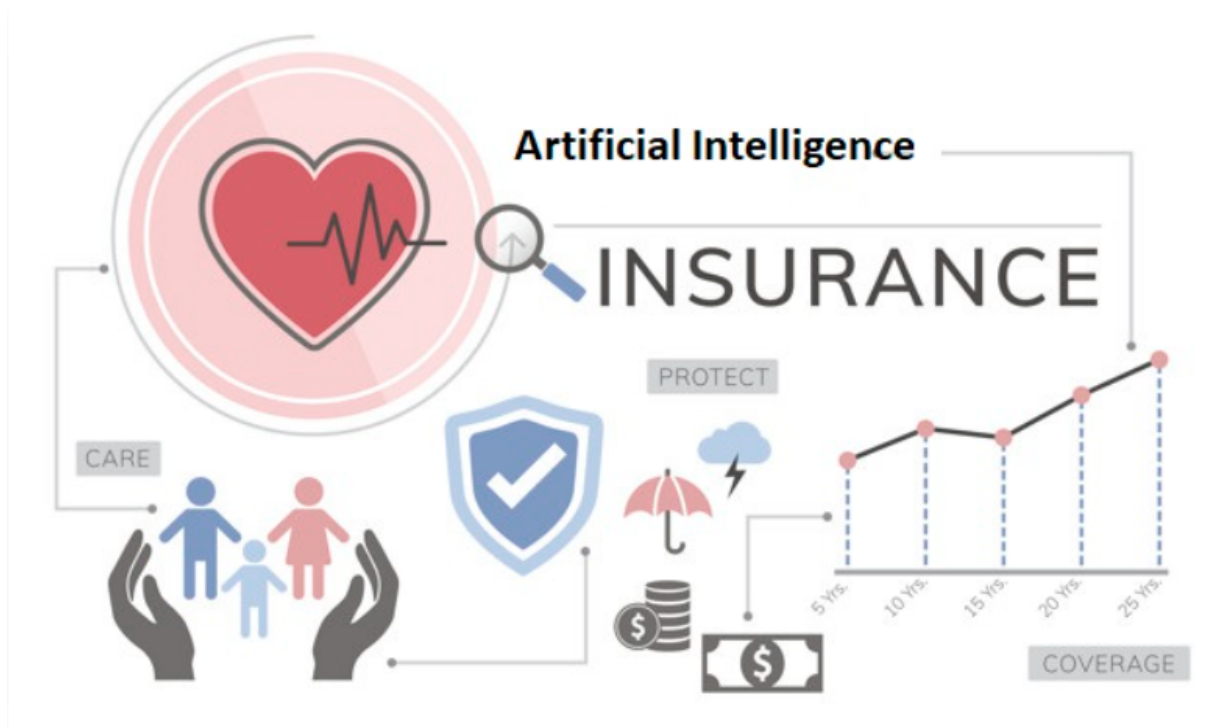
# HEALTH INSURANCE PREMIUM PREDICTION USING IBM AUTO AI SERVICES

## Introduction:

### 1.1 Overview:

Artificial Intelligence is driving significant change in business, and insurance is no exception. AI has the potential to transform the business model of an insurer.

In this project we create a machine learning model to predict the insurance charges, and helps the customer understand how smoking or decreasing your BMI affects insurance premiums.



We use IBM cloud as a host to complete this project and here the ML pipeline workloads be completely automated. The increase in growth in data science and AI, IBM is offering, Auto AI which can easily perform automated data preparation, apply ML algorithms, and build model pipelines best suited for our datasets and calculate the insurance premium.

## **1.2 Purpose:**

The use of artificial intelligence (AI) is being increasingly used within the health care field and related to them. Some AI services has also been provided by the companies like the insurance companies to their customers. Applications of the use of AI in this field are the treatment recommendations, diagnosis and administrative related activities such as the impact of their habits on the insurance premium they pay.

According to the reports, the value of gross insurance premiums worldwide continue to increase past 5-6 trillion dollars, and this increase or growth can be reduced by mainly quit smoking and concentrating on the BMI. Following these measures, we can save a lot of money to the payment of insurance premium and in this project, we study the effects of age, smoking, BMI, gender, and region to determine how much of a difference these factors can make on your insurance premium. By using our project, customers see the huge difference their lifestyle choices make on their insurance charges. By making use of the concept of AI and machine learning, we help customers understand that how much smoking increases their premium, by predicting the premium they has to pay in no time.

## **Literature Survey:**

### **2.1 Exsisting Problem:**

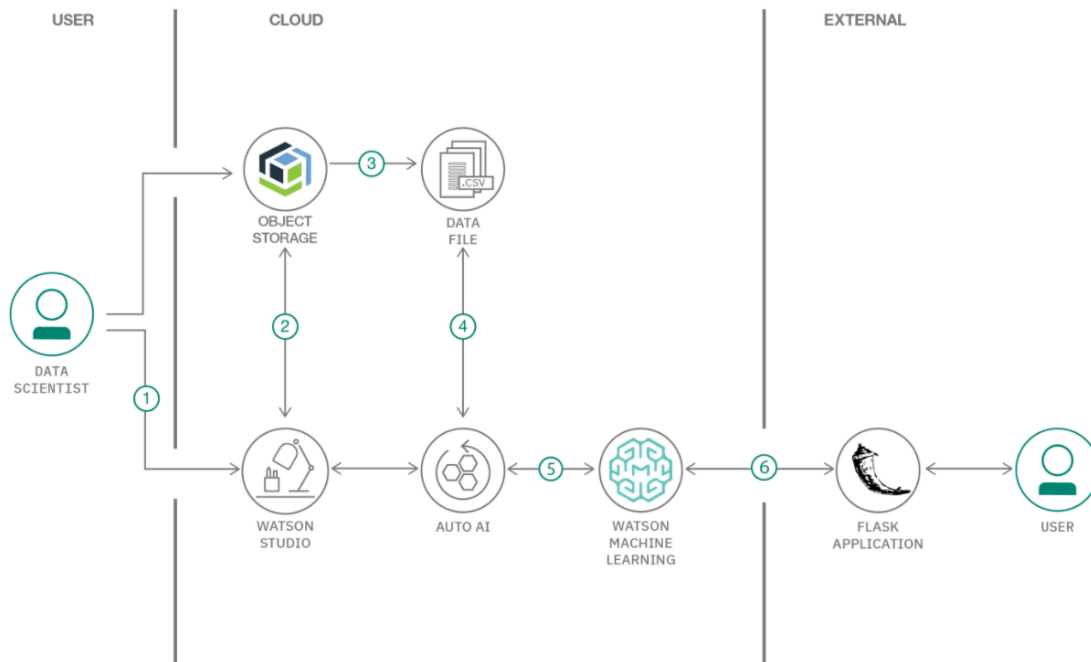
At present the main exsisting problem in the insurance industry is that the compamies are not customer friendly and also the customers don't know exactly the process of calculation of the insurance premium paid by them and also there is inefficiency in the calculation of insurance premium and also many claims are getting wrongly delivered. This is all due to the amount of work required to be processed manually often leads to inefficiency. Furthermore, an so many sum of tasks may cause mistakes that impact customers' lives.

### **2.1 Proposed Solution:**

Using AI we get solutions which are designed to automate, simplify, and speed up the process of claims handling, which leads to increased customer satisfaction and cost savings in operations. AI-based applications are extremely effective in collecting and processing claims data, verifying and analyzing them. With the help of AI, customer experience can be improved to so much extent.

## Theoretical Experience:

### 3.1 Block Diagram:



- We will set up the services on IBM Cloud to build the model
- We will ingest the data and initiate the Auto-AI process
- We will give the column to be predicted that is the expenses and build different models using Auto-AI and evaluate the performance
- We will choose the best model and complete the deployment
- We will generate predictions using the deployed model by making REST calls
- We will visualize the deployed model using a front-end application

## 3.2 HARDWARE & SOFTWARE Designing:

### **Hardware Components used:**

Since we are using the IBM cloud as a platform to execute this project we don't need any hardware components other than our system.

### **Software Components used:**

We will be using Watson studio, Node red service from the IBM cloud to complete the project.

### **WATSON STUDIO:**

Watson Studio is one of the core services in Cloud Pak for Data as a Service.

Watson Studio provides you with the environment and tools to solve your business problems by collaboratively working with data. You can choose the tools you need to analyze and visualize data, to cleanse and shape data, or to build machine learning models.

This illustration shows how the architecture of Watson Studio is centered around the project. A project is a workspace where you organize your resources and work with data.

Watson Studio projects fully integrate with the catalogs and deployment spaces:

- Deployment spaces are provided by the Watson Machine Learning service
  - You can easily move assets between projects and deployment spaces.

### **NODE-RED SERVICE:**

Node-RED provides a browser-based flow editor that makes it easy to wire together flows using the wide range of nodes in the palette. Flows can be then deployed to the runtime in a single-click.

The flows created in Node-RED are stored using JSON which can be easily imported and exported for sharing with others.

These are the two applications used to complete our project.

## Experimental Investigations:

### 3.1 DATA ANALYSIS AND OBSERVING THE TRENDS: (Using Jupyter Notebook)

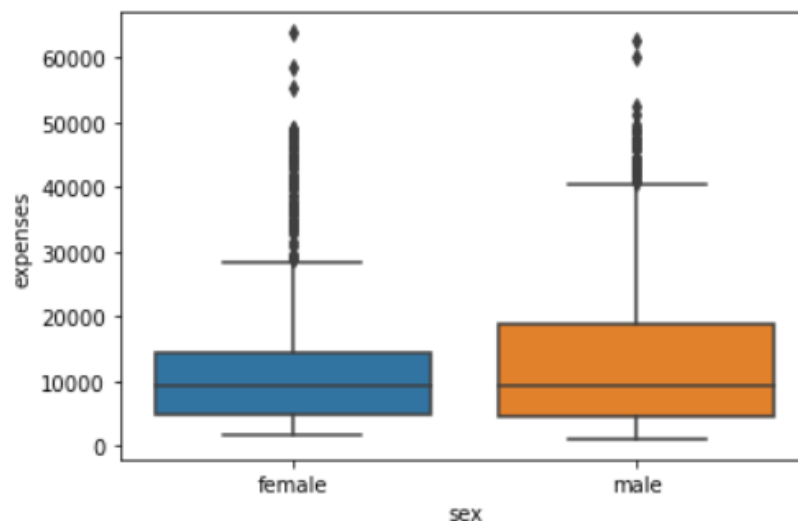
- Once we import, we see the data into a data frame, and call the `premium.head()` function, you see the first 5 rows of the data set. We see the features to be age, sex, bmi, children, smoker, and region.

	age	sex	bmi	children	smoker	region	expenses
0	19	female	27.9	0	yes	southwest	16884.92
1	18	male	33.8	1	no	southeast	1725.55
2	28	male	33.0	3	no	southeast	4449.46
3	33	male	22.7	0	no	northwest	21984.47
4	32	male	28.9	0	no	northwest	3866.86

- To check if there is a strong relationship between sex and expenses we create a box plot. We see that the average claims for males and females are similar, whereas males have a bigger proportion of the higher claims.

```
In [7]: 1 sns.boxplot(x='sex',y='expenses',data=premium)
```

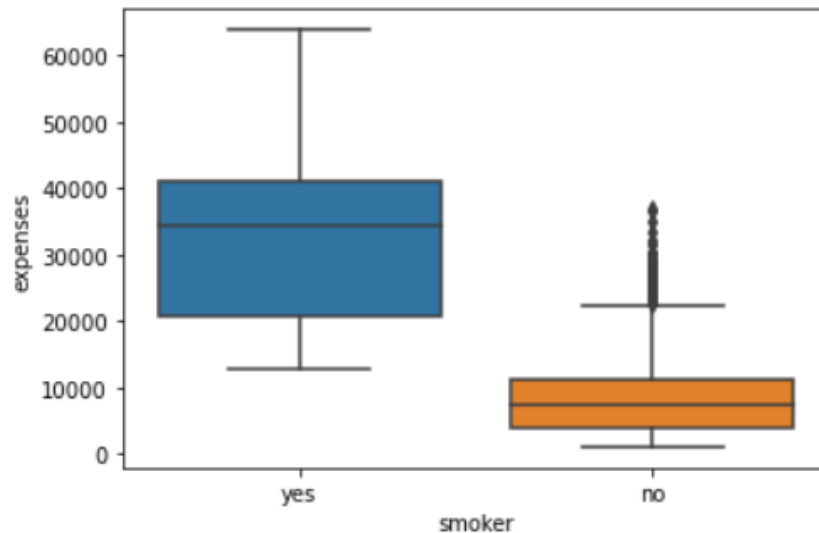
```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x20f9ee71df0>
```



- To check if there is a strong relationship between smoker and expenses we create a box plot. We see that if you are a smoker, our claims are much higher on average.

```
In [8]: 1 sns.boxplot(x='smoker',y='expenses',data=premium)
```

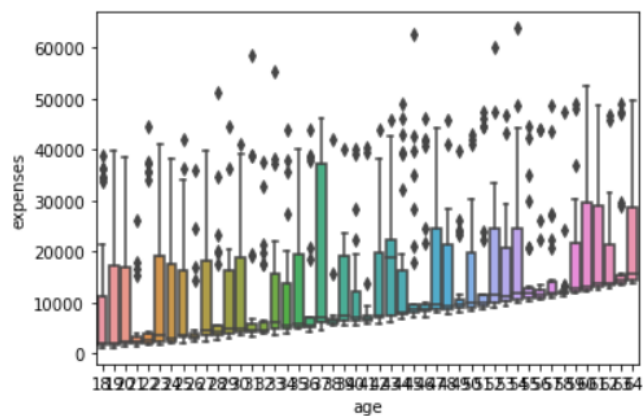
```
Out[8]: <matplotlib.axes._subplots.AxesSubplot at 0x20f9f63e4f0>
```



To check if there is a strong relationship between age and expenses we create a box plot. We see that as age increases, our claims are much higher on average.

```
In [9]: 1 sns.boxplot(x='age',y='expenses',data=premium)
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x20f9f6fe6d0>
```

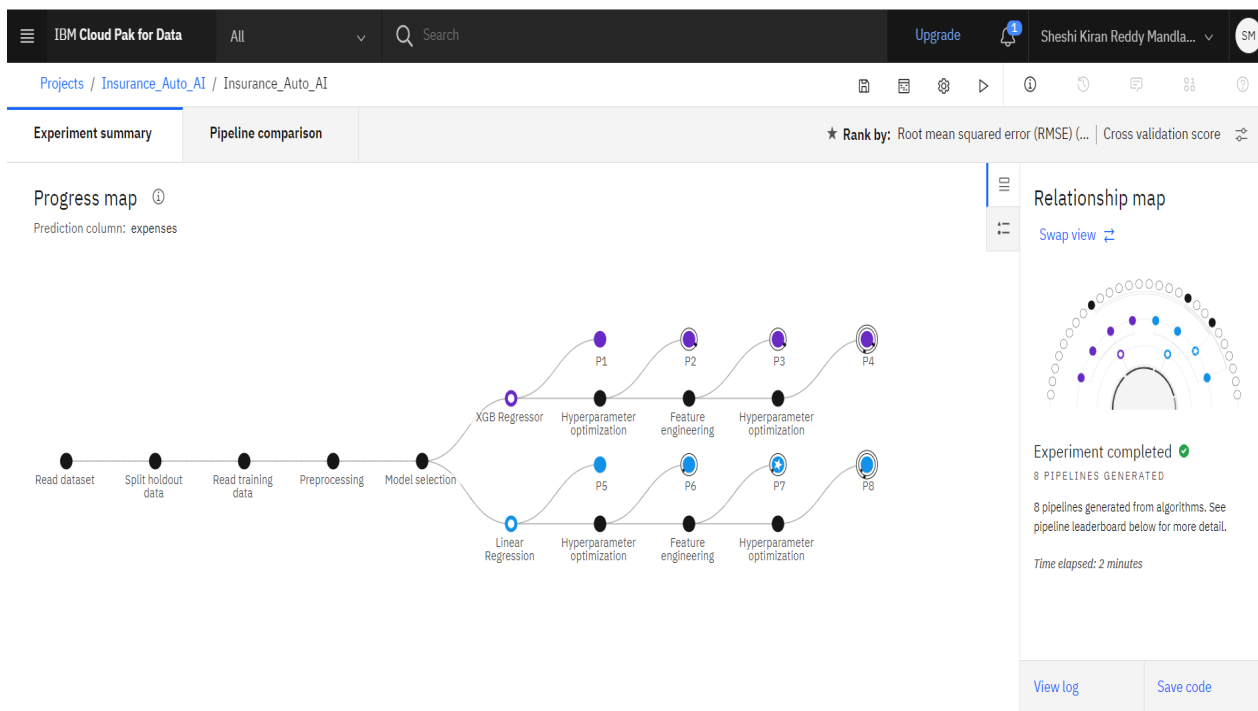
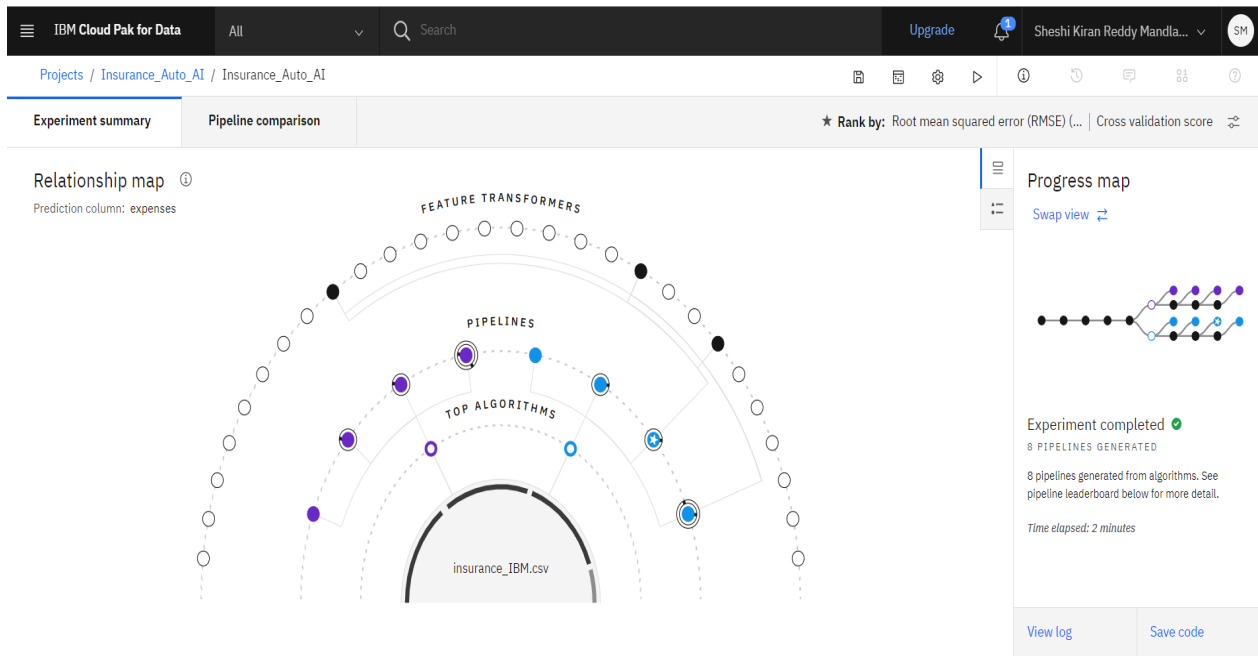


## Flow Chart & Results with Screenshots:

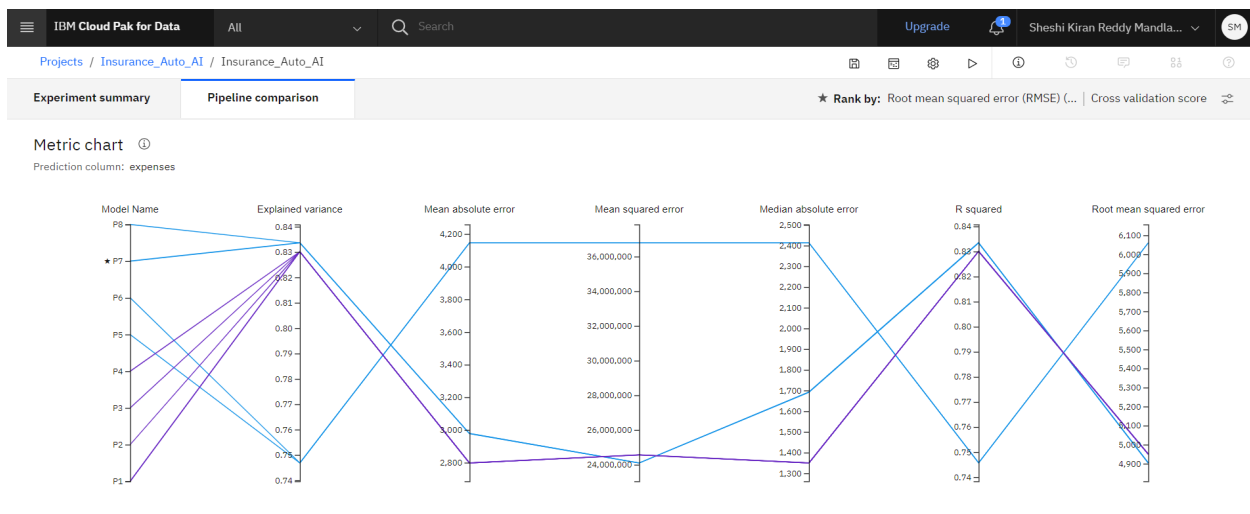
### 5.1 Flow Chart & Results using the Watson Studio Service:

- 1) Sign up/ Login to the IBM cloud account providing your credentials
- 2) Create a Watson studio services choosing the plan of your choice.
- 3) Now under services in the resource list open the watson studio service and you will be redirected to this page.
- 4) Now click on get started and click on "create a project " and enter the name of the project.
- 5) Beside you will find the option "add" to add the CLOUD OBJECT STORAGE .Click on it and choose your plan and click on create. Therefore now a COS is associated to your project.
- 6) Now select "Auto - AI PROJECT" in the asset option and associate a machine learning service to it in the page you were directed.
- 7) Now upload your Data file (Insurance excel file) and select the column which you want to predict ( "expenses" ).
- 8) The watson studio service will automatically display the algorithm after providing the above details. In our case it is "Regression" algorithm.
- 9) Next, your AutoAI experiment runs on its own. You see a progress map on the right side of the screen which shows which stage of the experiment is running. This may be Hyper Parameter Optimization, feature engineering, or some other stage.
- 10) You have different pipelines that are created, and you see the rankings of each model. Each model is ranked based on the metric that you selected. In the specific case that is the RMSE(Root mean squared error). Given that you want that number to be as small as possible, you can see that in the experiment, the model with the smallest RMSE is at the top of the leaderboard.
- 11) Below you can see the metric chart and the graphs related to the experiment

## ScreenShots:







## 12) Pipeline Comparison Chart:

**Pipeline comparison**

★ Rank by: Root mean squared error (RMSE) (...) | Cross validation score

Rank	↑	Name	Algorithm	RMSE (Optimized) Cross Validation	Enhancements	Build time
★ 1		Pipeline 7	Linear Regression	4903.651	HPO-1 FE	00:00:15
2		Pipeline 8	Linear Regression	4903.651	HPO-1 FE HPO-2	00:00:02
3		Pipeline 1	XGB Regressor	4949.683	None	00:00:01
4		Pipeline 2	XGB Regressor	4949.683	HPO-1	00:00:10
5		Pipeline 3	XGB Regressor	4949.683	HPO-1 FE	00:00:19
6		Pipeline 4	XGB Regressor	4949.683	HPO-1 FE HPO-2	00:00:27
7		Pipeline 5	Linear Regression	6060.420	None	00:00:01
8		Pipeline 6	Linear Regression	6060.420	HPO-1	00:00:01

13) Now that AutoAI has successfully generated eight different models, you can rank the models by different metrics, by clicking on the drop-down next to “Rank by:” on the top right corner of the screen, such as explained variance, root mean squared error, R-Squared, and mean absolute error. Each time you select a different metric, the models are re-ranked by that metric.

14) In our case, we have RMSE as the experiment's metric. You see the smallest RMSE value is 4903.651, from Pipeline 7.

15) Now click on "save as" beside pipeline 7 and deploy the model and it is promoted to the deployment space and we are ready to test the model.

16) We will make note of the scoring endpoint and we will test the model.

### Screenshot:

The screenshot displays the IBM Cloud Pak for Data interface. At the top, the navigation bar shows 'IBM Cloud Pak for Data' and 'All' filters. Below the navigation bar, the breadcrumb trail indicates the current location: 'Deployments / Insurance\_Deploy\_AI\_9Aug / Insurance\_Auto\_AI - P7 Linear R... / insurance\_Deploy\_AI\_9Aug'. The main content area shows the 'insurance\_Deploy\_AI\_9Aug' deployment, which is 'Deployed' and 'Online'. The 'Test' tab is active, showing the 'Enter input data' section with fields for 'sex' (other), 'bmi' (Double), 'children' (Integer), 'smoker' (other), and 'region' (other). The 'Input list (1)' section shows the input data: '[18, male, 25, 0, no, southeast]'. The 'Result' section displays the output: 'predictions: [4243.23828125]'. The URL at the bottom of the interface is 'eu-gb.dataplatform.cloud.ibm.com/ml-runtime/deployments/.../test?space\_id=dd7d38de-11...'.

17) Now we are getting our results that are the expenses with the given input.

18) Now we will associate it to a UI using "Node red" services.

For the Node Red Services : (Work done after watson studio)

- Get a IAM access token in the manage section and create a new API key key and make a note of it.
- Integrating with Node Red Service:
  - 1) Create a node red services by choosing the plan of your service.
  - 2) Now click on the Node Red App in the services list and you will be directed node red website.

In the Node Red the steps to be followed to create the Flow for calculating the premium:

- 1) Create a UI
- 2) Grab the values from the UI
- 3) Set global variables for the UI
- 4) Get the access token with the help of the http request node
- 5) Send the input values to scoring end point in json format along with the access token using the http request node
- 6) Get the predicted value
- 7) Parse the output.
- 8) Showcase the output on the UI.

[Note : Paste the correct header and payload messages while giving API key and also the scoring end point with correct columns

```
var age=global.get("age")
var sex=global.get("sex")
var children=global.get("children")
var bmi=global.get("bmi")
var region=global.get("region")
var smoker=global.get("smoker")
var token=msg.payload.access_token
msg.headers={'Content-Type':'application/json',"Authorization":"Bearer "+token,"Accept":"application/json"}
msg.payload={"input_data":[{"fields":["age","sex","bmi","children","smoker","region"]},"values":[[age,sex,bmi,children,smoker,region]]}}
return msg;]
```

**Screenshots:(Flow and the Results)**



In the debug window of Node red service:

```
8/11/2021, 6:39:13 AM node: 872b308f93e96ee1
msg.payload : string[4]
"male"

8/11/2021, 6:39:19 AM node: 37d2672f48f2e963
msg.payload : string[9]
"southeast"

8/11/2021, 6:39:35 AM node: 0aea351ae33d8f8c
msg.payload : Object
▶ { age: "18", children: "0", bmi: "25" }

8/11/2021, 6:39:36 AM node: 1094edb32c11dbe0
msg.payload : Object
▶ { grant_type: "urn:ibm:params:oauth:grant-typ...", apikey:
"2a0ULC97E1eGk7s7s6d_BD1lmVYc47..." }

8/11/2021, 6:39:37 AM node: 10bed46b7f0ddc46
msg.payload : Object
▶ { access_token: "eyJraWQiOiIyMDIxMDcxOTE4MzciLC...",
refresh_token: "not_supported", token_type: "Bearer",
expires_in: 3600, expiration: 1628647701 ... }

8/11/2021, 6:39:37 AM node: fe9c7535ec803d10
msg.payload : Object
▶ { predictions: array[1] }

8/11/2021, 6:39:37 AM node: 2f35130387ad2652
msg.payload : number
4243.23828125
```

## Advantages & Disadvantages:

### 6.1 Advantages:

- The customer will be able to know the premium to be paid on his own and there will be the customer satisfaction increases and there will be no wrong claims either by the customer or the insurance company.
- The sales for the company increases and it is also proved already that 32% stock price increase on an average has been to the companies which adopted the AI application.

### 6.2 Disadvantages:

- Lacks creativity
- Chances Of unemployment
- Chances of company making fraud using the technology

## Applications of AI in Healthcare:

- Support in Clinical Decisions
- Enhance Primary Care and Triage through Chatbots
- Robotic Surgeries
- Virtual nursing assistants
- Aiding in the accurate diagnosis
- Minimizing the burden of EHR use

## Conclusion:

- Machine learning algorithms can effectively scan all the incoming data, interpret instead of insurance agents, and provide faster settlement to end-users.
- So the workload decreases and the customer satisfaction can be improved.
- Customers can also get to know the factors for the increase in the premium to be paid by them and can control them like smoking etc.
- So it is profitable to both company and the customers.

## Future Scope:

- Machine learning algorithms can be used by the customers to select the best insurance policy based on the given data to them and the customers only can choose the best plan.
- The future scope of this project can be the above mentioned point.

## References:

- <https://github.com/IBM/predict-insurance-charges-with-autoai/blob/master/README.md>
- <https://developer.ibm.com/patterns/create-an-application-to-predict-your-insurance-premium-cost-with-autoai/>
- <https://usmsystems.com/ai-innovation-in-the-insurance-sector/>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/>
- <https://blog.re-work.co/challenges-in-implementing-ai-in-the-insurance-industry-and-improve-claims-management-process/>
- <https://developer.ibm.com/patterns/create-an-application-to-predict-your-insurance-premium-cost-with-autoai/>
- <https://theconversation.com/very-risky-business-the-pros-and-cons-of-insurance-companies-embracing-artificial-intelligence-106536>
- <https://www.v7labs.com/blog/ai-insurance>
- <https://www.analyticssteps.com/blogs/artificial-intelligence-healthcare-applications-and-threats>

**THE END**

Name: Mandla Sheshi Kiran Reddy

Email-id: sheshikiranmandla2589@gmail.com