Report

On

Breast Cancer Risk Prediction System

(https://github.com/sonaliBodkhe/Breast-Cancer-Prediction-Using-IBM-WATSON)

Submitted Under

Category of

Machine Learning

for

SmartInternz GuruCool Training Program

based on

IBM WATSON

By

Sonali T. Bodkhe

PhD. in Computer Science & Engineering

2020

INDEX

1 INTRODUCTION

1.1 Overview

Breast cancer is one of the main causes of cancer death worldwide. Early diagnostics significantly increases the chances of correct treatment and survival, but this process is tedious and often leads to a disagreement between pathologists. Computer-aided diagnosis systems showed the potential for improving diagnostic accuracy. But early detection and prevention can significantly reduce the chances of death. It is important to detect breast cancer as early as possible.

1.2 Purpose

Since early detection and prevention can significantly reduce the chances of death the earlier, the better. The purpose here is to build a model in Watson Studio and deploy the model in IBM Watson Machine Learning. To interact with the model, Node-Red and scoring Endpoint will be used.

2 LITERATURE SURVEY

2.1 Existing problem

Doctors use many tests to find, or diagnose, breast cancer. They may also do tests to learn if the cancer has spread to a part of the body other than the breast and the lymph nodes under the arm. If this happens, it is called a metastasis. Doctors may also do tests to learn which treatments could work best.

For most types of cancer, a biopsy is the only sure way for the doctor to know if an area of the body has cancer. In a biopsy, the doctor takes a small sample of tissue for testing in a laboratory.

This section describes options for diagnosing breast cancer. Not all tests listed below will be used for every person. Doctors may consider these factors when choosing a diagnostic test:

- The type of cancer suspected
- Your signs and symptoms
- Your age and general health
- The results of earlier medical tests

The series of tests needed to evaluate a possible breast cancer usually begins when a woman or their doctor discover a mass or abnormal calcifications on a screening mammogram, or a lump or nodule in the breast during a clinical or self-examination. Less commonly, a woman might notice a red or swollen breast or a mass or nodule under the arm.

The following tests may be used to diagnose breast cancer or for follow-up testing after a breast cancer diagnosis.

Imaging tests: Imaging tests show pictures of the inside of the body. The following imaging tests of the breast may be done to learn more about a suspicious area found in the breast during screening. In addition to these, there are other new types of tests that are being studied.

- **Diagnostic mammography. Diagnostic mammography** is similar to screening mammography except that more pictures of the breast are taken. It is often used when a woman is experiencing signs, such as a new lump or nipple discharge. Diagnostic mammography may also be used if something suspicious is found on a screening mammogram.
- **Ultrasound.** An **ultrasound** uses sound waves to create a picture of the breast tissue. An ultrasound can distinguish between a solid mass, which may be cancer, and a fluid-filled cyst, which is usually not cancer.
- **MRI.** An **MRI** uses magnetic fields, not x-rays, to produce detailed images of the body. A special dye called a contrast medium is given before the scan to help

create a clear picture of the possible cancer. This dye is injected into the patient's vein. A breast MRI may be used after a woman has been diagnosed with cancer to find out how much the disease has grown throughout the breast or to check the other breast for cancer. Breast MRI is also a screening option, along with mammography, for some women with a very high risk of developing breast cancer and for some women who have a history of breast .MRI may also be used if locally advanced breast cancer is diagnosed or if chemotherapy or endocrine therapy is being given first, followed by a repeated MRI for surgical planning. Finally, MRI may be used as a surveillance method following a breast cancer diagnosis and treatment.

Biopsy:A biopsy is the removal of a small amount of tissue for examination under a micro scope. Other tests can suggest that cancer is present, but only a biopsy can make a definite diagnosis. A pathologist then analyzes the sample(s).

Using the TNM system, the "T" plus a letter or number (0 to 4) is used to describe the size and location of the tumor. Tumor size is measured in centimeters (cm). A centimeter is roughly equal to the width of a standard pen or pencil. Stage may also be divided into smaller groups that help describe the tumor in even more detail. Specific tumor stage information in listed below.

TX: The primary tumor cannot be evaluated.

T0 (T plus zero): There is no evidence of cancer in the breast.
Tis: Refers to carcinoma in situ**.** The cancer is confined within the ducts of the breast tissue and has not spread into the surrounding tissue of the breast. There are 2 types of breast carcinoma in situ:

- **Tis (DCIS):** DCIS is a non-invasive cancer, but if not removed, it may develop into an invasive breast cancer later. DCIS means that cancer cells have been found in breast ducts and have not spread past the layer of tissue where they began.

- **Tis (Paget's):** Paget disease of the nipple is a rare form of early, non-invasive cancer that is only in the skin cells of the nipple. Sometimes Paget disease is associated with an invasive breast cancer. If there is an invasive breast cancer, it is classified according to the stage of the invasive tumor.

T1: The tumor in the breast is 20 millimeters (mm) or smaller in size at its widest area. This is a little less than an inch. This stage is then broken into 4 substages depending on the size of the tumor:

- T1mi is a tumor that is 1 mm or smaller.
- T1a is a tumor that is larger than 1 mm but 5 mm or smaller.
- T1b is a tumor that is larger than 5 mm but 10 mm or smaller.
- T1c is a tumor that is larger than 10 mm but 20 mm or smaller.

T2: The tumor is larger than 20 mm but not larger than 50 mm.

T3: The tumor is larger than 50 mm.

T4: The tumor falls into 1 of the following groups:

- T4a means the tumor has grown into the chest wall.
- T4b is when the tumor has grown into the skin.
- T4c is cancer that has grown into the chest wall and the skin.
- T4d is **inflammatory breast cancer**. The detailing is huge and can go on and on.
-

Types: There are three main types of tumor:

Benign: These are not cancerous. They either cannot spread or grow, or they do so very slowly. If a doctor removes them, they do not generally return.

Premalignant: In these tumors, the cells are not yet cancerous, but they have the potential to become malignant.

Malignant: Malignant tumors are cancerous. The cells can grow and spread to other parts

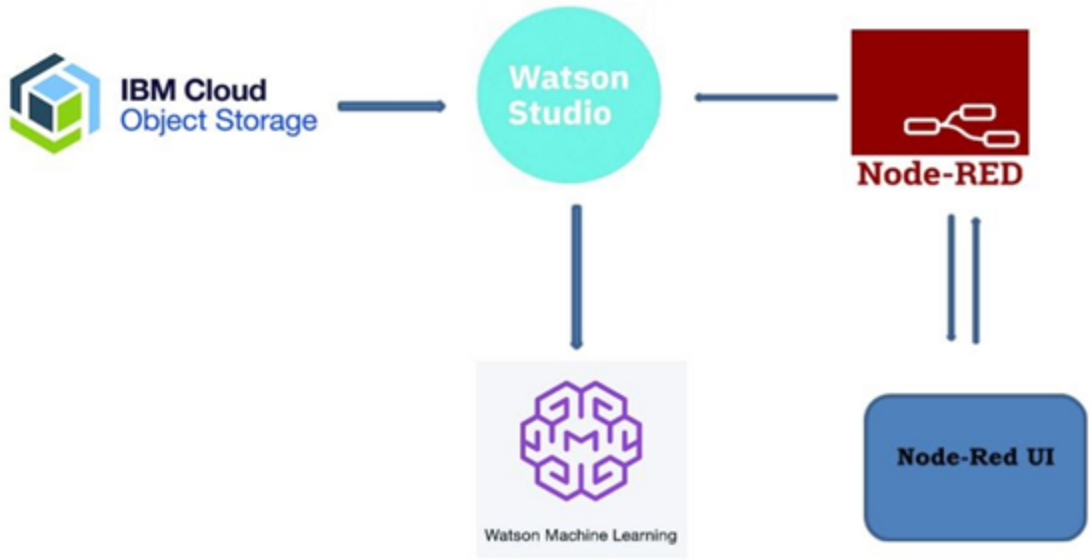
of the body.

2.2 Proposed solution

As can be seen, there is a series of tests and diagnosis to be carried out. A lot of processing and documentation and manual analysis is involved. Instead if all this data is retrieved into a model , early detection and prevention will significantly reduce the chances of death. The purpose here is to build a machine learning and deploy it in Watson Studio by creating an endpoint. To interact with the model, Node-Red and scoring Endpoint will be used.

3. THEORITICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing

The following skillset is used:

- IBM Nodered
- IBM Watson Studio
- IBM Machine Learning,
- IBM Cloud Object Storage

4 EXPERIMENTAL INVESTIGATIONS

1. Create a project and add auto AI experiment.
2. Create ML instance.
3. Associate ML instance to project
4. Import related data and associate to cloud storage object
5. Select prediction parameter in dataset.
6. Train the model(top n models training info displayed based on no. of cores)
7. Deploy the model.
8. Build web application using node-red. UI can be created to input a new unknown data-point and the interface immediately displays the result.

5 RESULT

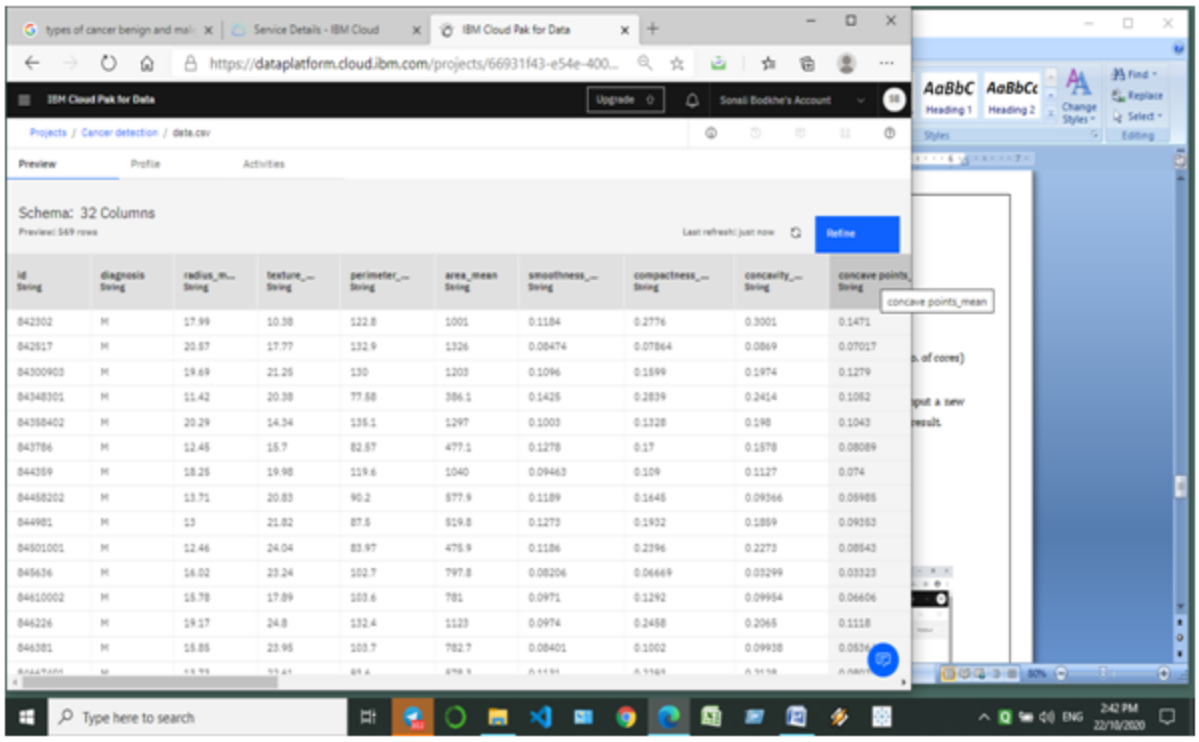


Figure 1: Cancer Dataset

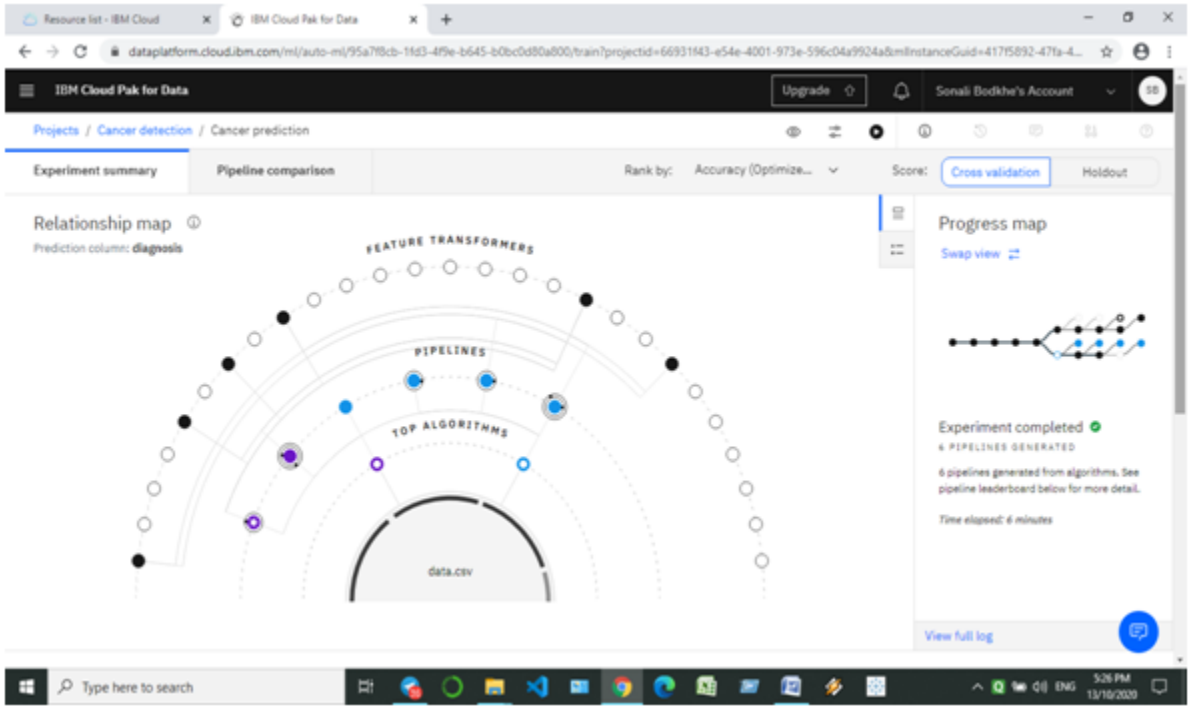


Figure 2: AutoAI Experiment Model Experiment Summary(Relationship Map)

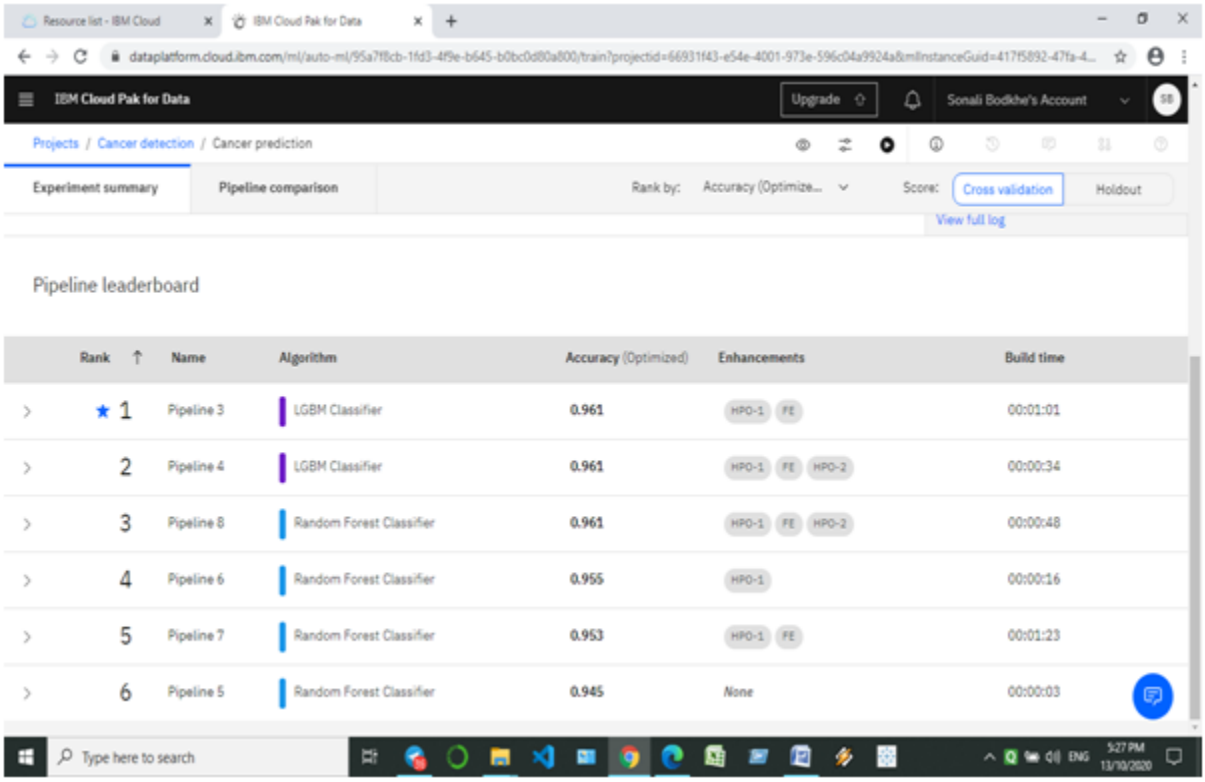


Figure 3: AutoAI Experiment Model Experiment Summary(Pipeline Leaderboard)

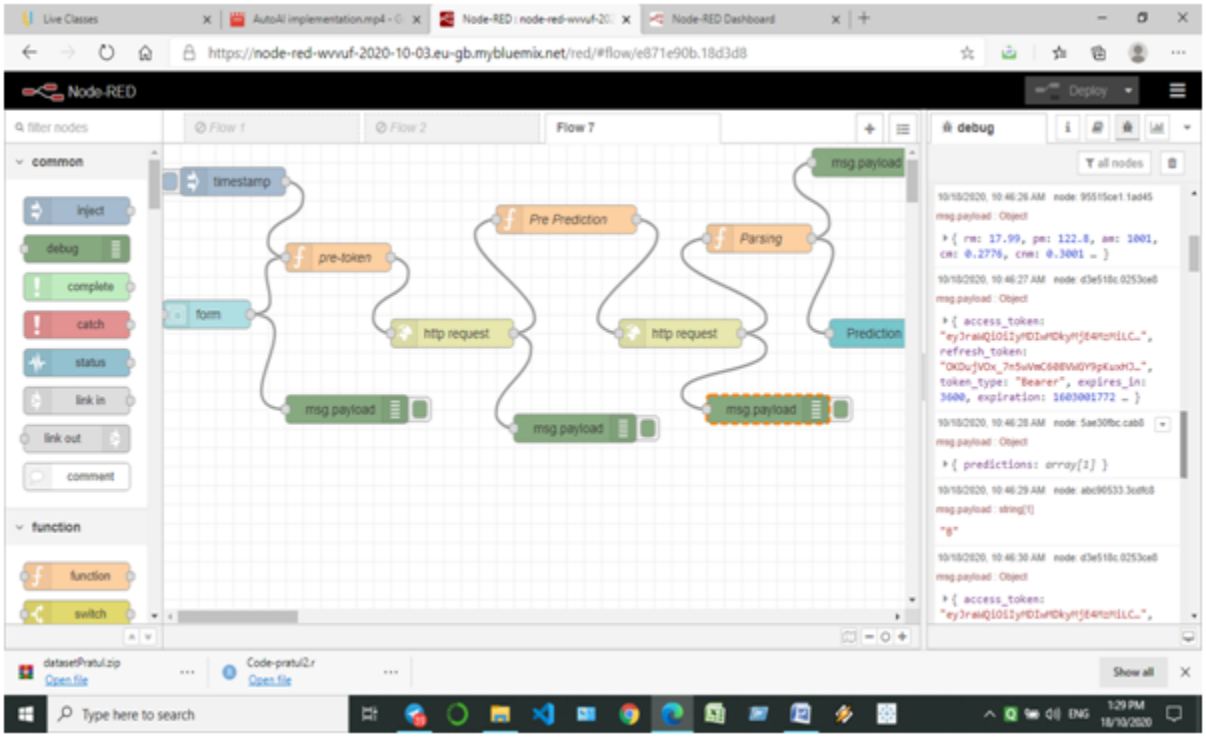


Figure 4: Nodered Flow and deployment

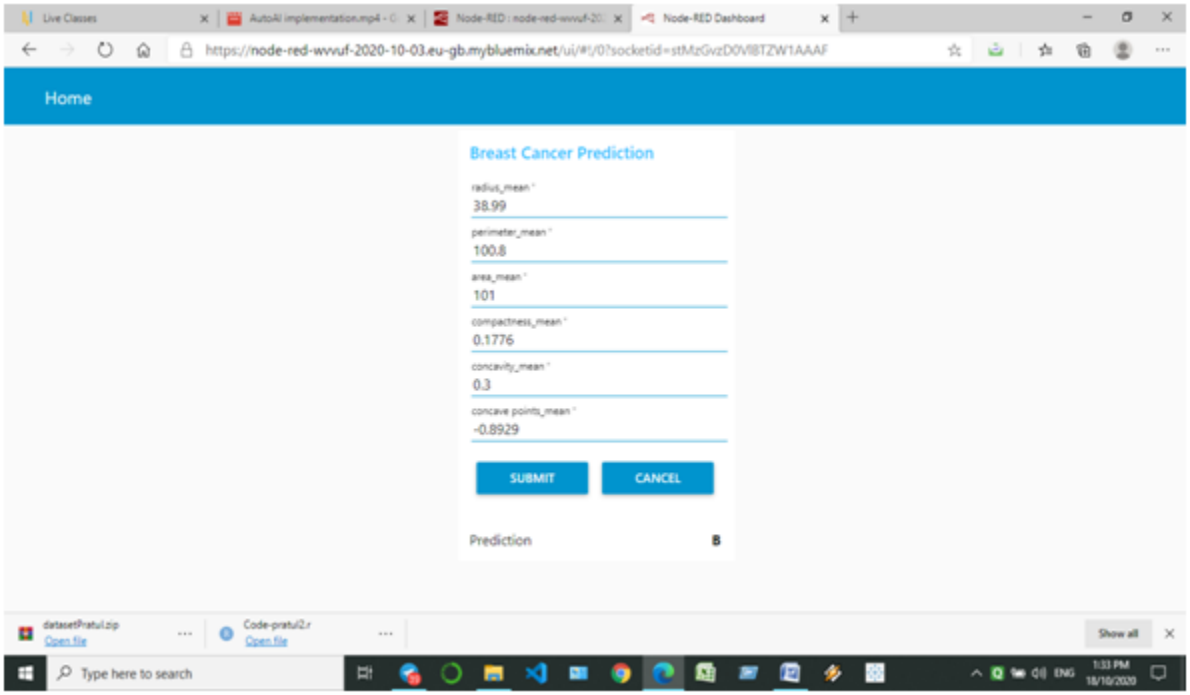


Figure 5:UI for Nodered Flow

6. ADVANTAGES & DISADVANTAGES

Looking into the severity of the disease and its extreme consequences which result into death, the machine leaning model can be used to detect the cancer at an earlier stage and the process of detection also gets speeded up since manual methods/procedures are not involved.

APPENDIX

A. Source code

```
[
    {
        "id": "e871e90b.18d3d8",
        "type": "tab",
        "label": "Flow 7",
        "disabled": false,
        "info": ""
    },
    {
        "id": "a2a06fbe.47d6e",
        "type": "ui_form",
        "z": "e871e90b.18d3d8",
        "name": "",
        "label": "",
        "group": "3bfc3902.f58296",
        "order": 1,
        "width": 0,
        "height": 0,
        "options": [
            {
```

```
                "label": "radius_mean",
                "value": "rm",
                "type": "number",
                "required": true,
                "rows": null
            },
            {
                "label": "perimeter_mean",
                "value": "pm",
                "type": "number",
                "required": true,
                "rows": null
            },
            {
                "label": "area_mean",
                "value": "am",
                "type": "number",
                "required": true,
                "rows": null
            },
            {
                "label": "compactness_mean",
                "value": "cm",
                "type": "number",
                "required": true,
                "rows": null
            },
            {
                "label": "concavity_mean",
```

```
                "value": "cnm",
                "type": "number",
                "required": true,
                "rows": null
            },
            {
                "label": "concave points_mean",
                "value": "cpm",
                "type": "number",
                "required": true,
                "rows": null
            }
        ],
        "formValue": {
            "rm": "",
            "pm": "",
            "am": "",
            "cm": "",
            "cnm": "",
            "cpm": ""
        },
        "payload": "",
        "submit": "submit",
        "cancel": "cancel",
        "topic": "",
        "x": 50,
        "y": 180,
        "wires": [
            [
```

```
                "95515ce1.1ad45",
                "10188a50.c32d06"
            ]
        ]
    },
    {
        "id": "95515ce1.1ad45",
        "type": "debug",
        "z": "e871e90b.18d3d8",
        "name": "",
        "active": true,
        "tosidebar": true,
        "console": false,
        "tostatus": false,
        "complete": "payload",
        "targetType": "msg",
        "statusVal": "",
        "statusType": "auto",
        "x": 210,
        "y": 280,
        "wires": []
    },
    {
        "id": "10188a50.c32d06",
        "type": "function",
        "z": "e871e90b.18d3d8",
        "name": "pre-token",
                                                    "func":
  "global.set(\"rm\",msg.payload.rm)\nglobal.set(\"pm\",msg.p
```

```
ayload.pm)\nglobal.set(\"am\",msg.payload.am)\nglobal.set(\"cm\",msg.payload.cm)\nglobal.set(\"cnm\",msg.payload.cnm)\nglobal.set(\"cpm\",msg.payload.cpm)\nvar apikey=\"RFYhTBDTo4dStEFnlSGtA95D010aAOxg0NH2S6F7y122\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;\n",
        "outputs": 1,
        "noerr": 0,
        "initialize": "",
        "finalize": "",
        "x": 200,
        "y": 120,
        "wires": [
            [
                "d8a176bb.f923e8"
            ]
        ]
    },
    {
        "id": "d8a176bb.f923e8",
        "type": "http request",
        "z": "e871e90b.18d3d8",
        "name": "",
        "method": "POST",
        "ret": "obj",
        "paytoqs": "ignore",
        "url": "https://iam.cloud.ibm.com/identity/token",
        "tls": "",
```

```
        "persist": false,
        "proxy": "",
        "authType": "",
        "x": 330,
        "y": 200,
        "wires": [
            [
                "d3e518c.0253ce8",
                "65185361.a924dc"
            ]
        ]
    },
    {
        "id": "d3e518c.0253ce8",
        "type": "debug",
        "z": "e871e90b.18d3d8",
        "name": "",
        "active": true,
        "tosidebar": true,
        "console": false,
        "tostatus": false,
        "complete": "payload",
        "targetType": "msg",
        "statusVal": "",
        "statusType": "auto",
        "x": 470,
        "y": 300,
        "wires": []
    },
```

```
158     {
159         "id": "65185361.a924dc",
160         "type": "function",
161         "z": "e871e90b.18d3d8",
162         "name": "Pre Prediction",
163             "func": "var rm = global.get(\"rm\")\nvar pm =
   global.get(\"pm\")\nvar am = global.get(\"am\")\nvar cm =
   global.get(\"cm\")\nvar cnm = global.get(\"cnm\")\nvar cpm
   =                                    global.get(\"cpm\")\nvar
   token=msg.payload.access_token\nmsg.headers={'Content-Type'
   :              'application/json',\"Authorization\":\"Bearer
   \"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"i
   nput_data\":[{\"fields\":[\"id\",\"radius_mean\",\"texture_
   mean\",\"perimeter_mean\",\"area_mean\",\"smoothness_mean\"
   ,\"compactness_mean\",\"concavity_mean\",\"concave
   points_mean\",\"symmetry_mean\",\"fractal_dimension_mean\",
   \"radius_se\",\"texture_se\",\"perimeter_se\",\"area_se\",\
   "smoothness_se\",\"compactness_se\",\"concavity_se\",\"conc
   ave
   points_se\",\"symmetry_se\",\"fractal_dimension_se\",\"radi
   us_worst\",\"texture_worst\",\"perimeter_worst\",\"area_wor
   st\",\"smoothness_worst\",\"compactness_worst\",\"concavity
   _worst\",\"concave
   points_worst\",\"symmetry_worst\",\"fractal_dimension_worst
   \",\"Unnamed:
   32\"],\"values\":[[null,rm,null,pm,am,null,cm,cnm,cpm,null,
   null,null,null,null,null,null,null,null,null,null,null,null
   ,null,null,null,null,null,null,null,null,null,null]]}]}\nre
   turn msg;\n",
```

```
        "outputs": 1,
        "noerr": 0,
        "initialize": "",
        "finalize": "",
        "x": 460,
        "y": 80,
        "wires": [
            [
                "e1e9d313.bde21"
            ]
        ]
    },
    {
        "id": "e1e9d313.bde21",
        "type": "http request",
        "z": "e871e90b.18d3d8",
        "name": "",
        "method": "POST",
        "ret": "obj",
        "paytoqs": "ignore",
                                                    "url":
  "https://us-south.ml.cloud.ibm.com/ml/v4/deployments/3edf5e
  e5-9660-48f0-a5a1-accc145a5098/predictions?version=2020-09-
  01",
        "tls": "",
        "persist": false,
        "proxy": "",
        "authType": "",
        "x": 590,
```

```
        "y": 200,
        "wires": [
            [
                "5ae30fbc.cab8",
                "f56b0c6e.02f32"
            ]
        ]
    },
    {
        "id": "5ae30fbc.cab8",
        "type": "debug",
        "z": "e871e90b.18d3d8",
        "name": "",
        "active": true,
        "tosidebar": true,
        "console": false,
        "tostatus": false,
        "complete": "payload",
        "targetType": "msg",
        "statusVal": "",
        "statusType": "auto",
        "x": 690,
        "y": 280,
        "wires": []
    },
    {
        "id": "685e47a7.412ee8",
        "type": "inject",
        "z": "e871e90b.18d3d8",
```

```
        "name": "",
        "props": [
            {
                "p": "payload"
            },
            {
                "p": "topic",
                "vt": "str"
            }
        ],
        "repeat": "",
        "crontab": "",
        "once": false,
        "onceDelay": 0.1,
        "topic": "",
        "payload": "",
        "payloadType": "date",
        "x": 80,
        "y": 40,
        "wires": [
            [
                "10188a50.c32d06"
            ]
        ]
    },
    {
        "id": "abc90533.3cdfc8",
        "type": "debug",
        "z": "e871e90b.18d3d8",
```

```
248         "name": "",
249         "active": true,
250         "tosidebar": true,
251         "console": false,
252         "tostatus": false,
253         "complete": "payload",
254         "targetType": "msg",
255         "statusVal": "",
256         "statusType": "auto",
257         "x": 810,
258         "y": 20,
259         "wires": []
260     },
261     {
262         "id": "f56b0c6e.02f32",
263         "type": "function",
264         "z": "e871e90b.18d3d8",
265         "name": "Parsing",
266                                                     "func":
  "msg.payload=msg.payload.predictions[0].values[0][0]\nretur
  n msg;",
267         "outputs": 1,
268         "noerr": 0,
269         "initialize": "",
270         "finalize": "",
271         "x": 680,
272         "y": 100,
273         "wires": [
274             [
```

```
                "abc90533.3cdfc8",
                "50f4301d.5b382"
            ]
        ]
    },
    {
        "id": "50f4301d.5b382",
        "type": "ui_text",
        "z": "e871e90b.18d3d8",
        "group": "3bfc3902.f58296",
        "order": 2,
        "width": 0,
        "height": 0,
        "name": "",
        "label": "Prediction",
        "format": "{{msg.payload}}",
        "layout": "row-spread",
        "x": 820,
        "y": 200,
        "wires": []
    },
    {
        "id": "3bfc3902.f58296",
        "type": "ui_group",
        "z": "",
        "name": "Breast Cancer Prediction",
        "tab": "47a95329.022c7c",
        "order": 1,
        "disp": true,
```

```
        "width": "6",
        "collapse": false
    },
    {
        "id": "47a95329.022c7c",
        "type": "ui_tab",
        "z": "",
        "name": "Home",
        "icon": "dashboard",
        "disabled": false,
        "hidden": false
    }
]
```