**A REPORT ON**

**BREAST CANCER PREDICTION**

**SUBMITTED BY:**

**Bhavana K**

**Abstract:**

Breast cancer (BC) is one of the most common cancers among women worldwide, representing the majority of new cancer cases and cancer-related deaths according to global statistics, making it a significant public health problem in today’s society.

The early diagnosis of BC can improve the prognosis and chance of survival significantly, as it can promote timely clinical treatment to patients. Further accurate classification of benign tumors can prevent patients undergoing unnecessary treatments. Thus, the correct diagnosis of BC and classification of patients into malignant or benign groups is the subject of much research. Because of its unique advantages in critical features detection from complex BC datasets, machine learning (ML) is widely recognized as the methodology of choice in BC pattern classification and forecast modelling.

We also proposed a conceptual model for the prediction of breast cancer using different machine learning techniques.

**Introduction to machine learning:**

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

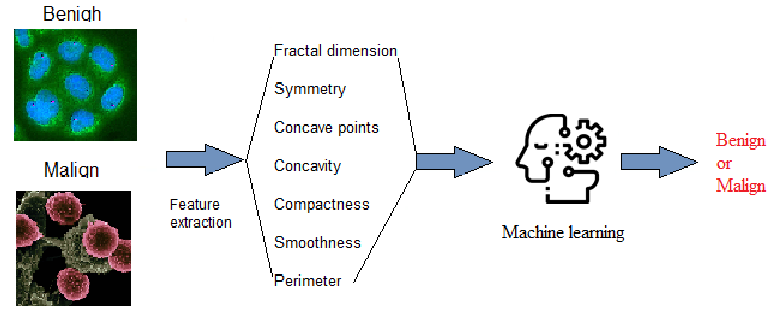
**Survey:**

A survey of breast cancer survivors in the United States was conducted to define what they had been told about their prognosis and the value of adjuvant therapy, what they estimated their prognosis to be with and without adjuvant therapy, and what level of improvement they would have found minimally worthwhile.

Only 39% of the women recalled receiving quantitative estimates of their prognosis, and only 31% of women received a quantitative estimate both with and without adjuvant therapy. Sixty-eight percent of the women were able to provide a quantitative estimate for their outcome at 5 years both with and without adjuvant therapy. From these estimates, we calculated that the median estimated proportional risk reduction for recurrence that women thought they had achieved was 79%. Women were asked what degree of absolute benefit they would have found acceptable. The median acceptable extension of life expectancy was 3 to 6 months, and acceptable reduction in recurrence risk was 0.5% to 1.0%. However, there was considerable variation, with 27% of women not accepting less than 1 year and 26% not accepting a less than 5% reduction in recurrence risks.

In general, American women in the surveyed population (1) do not recall being provided quantitative estimates of outcome during the process of making decisions about adjuvant therapy, (2) overestimate the value of their therapy, and (3) often will accept remarkably low degrees of net benefit. Overall, these observations can be used to support the argument that improvements in doctor/patient communication may be important to truly informed decision-making, and that flexibility for individual patients' preferences should not be superseded by rigid treatment guidelines.

**Methodology:**

****

**Steps:**

**Data set:**

The data set is used in prediction model. This data set contains

The following attributes

ID number 2) Diagnosis (M = malignant, B = benign) Ten real-valued features are computed for each cell nucleus:

1. radius (mean of distances from center to points on the perimeter)
2. texture (standard deviation of gray-scale values)
3. perimeter
4. area
5. smoothness (local variation in radius lengths)
6. compactness (perimeter² / area — 1.0)
7. concavity (severity of concave portions of the contour)
8. concave points (number of concave portions of the contour)
9. symmetry
10. fractal dimension (“coastline approximation” — 1)

**Model Selection**

This is the most exciting phase in Applying Machine Learning to any Dataset. It is also known as Algorithm selection for Predicting the best results.

Usually Data Scientists use different kinds of Machine Learning algorithms to the large data sets. But, at high level all those different algorithms can be classified in two groups : supervised learning and unsupervised learning.

Supervised learning : Supervised learning is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing. Supervised learning problems can be further grouped into **Regression** and **Classification** problems.

A **regression** problem is when the output variable is a real or continuous value, such as “salary” or “weight”.

A **classification** problem is when the output variable is a category like filtering emails “spam” or “not spam”.

We have different types of classification algorithms in Machine Learning :-

1. Logistic Regression

2. Nearest Neighbor

3. Support Vector Machines

4. Kernel SVM

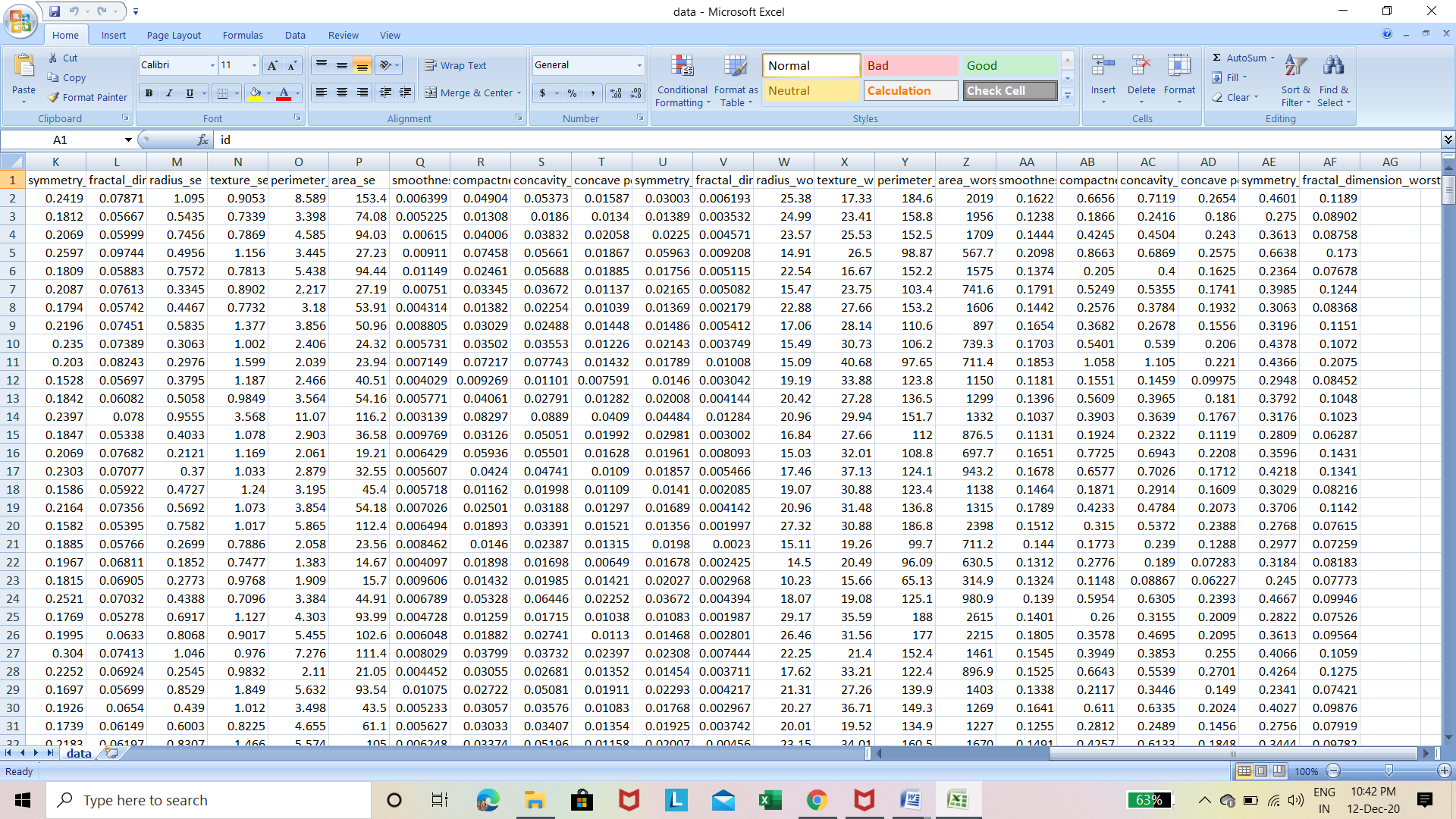
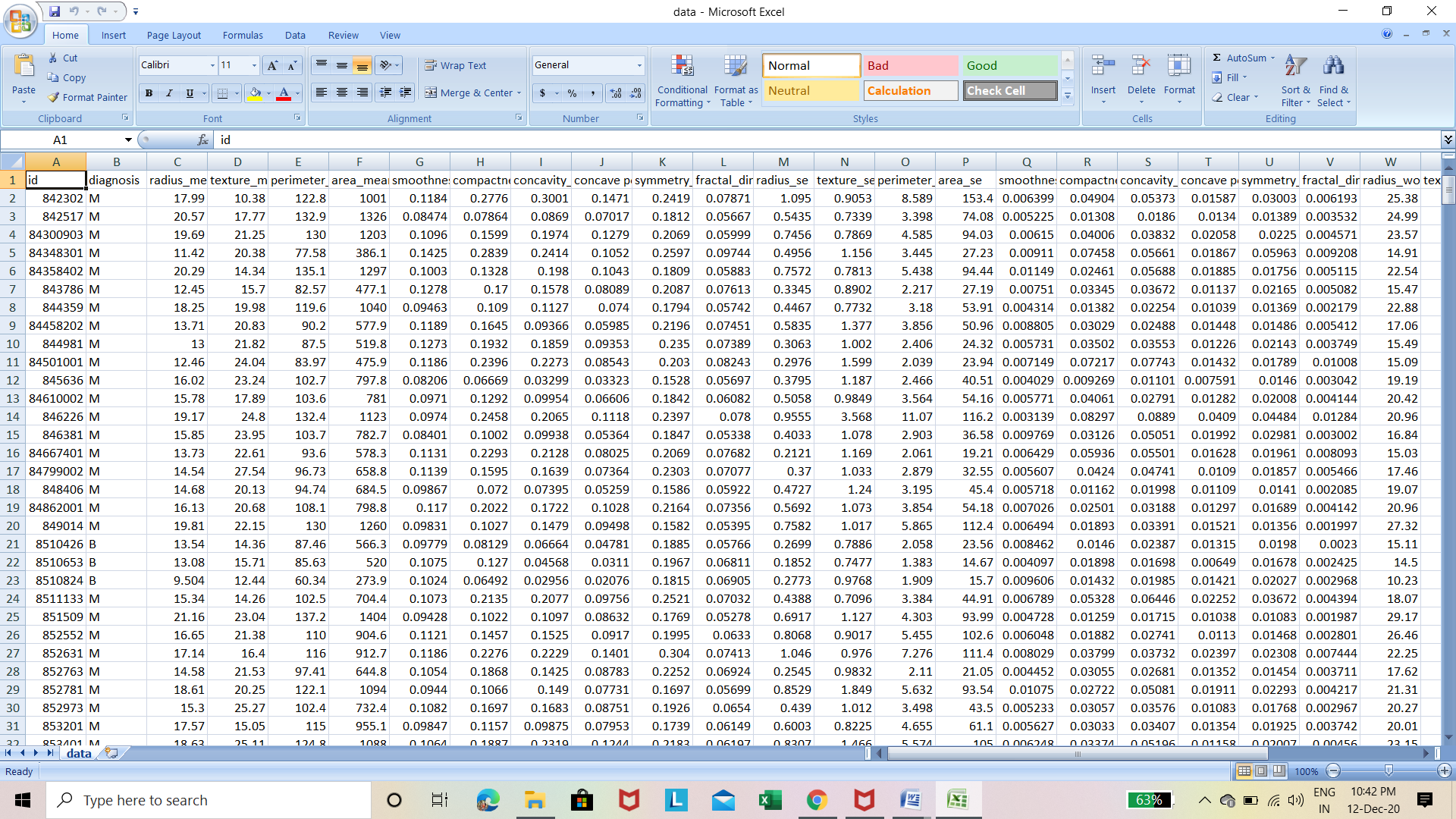
5. Naïve Bayes

6. Decision Tree Algorithm

7. Random Forest Classification

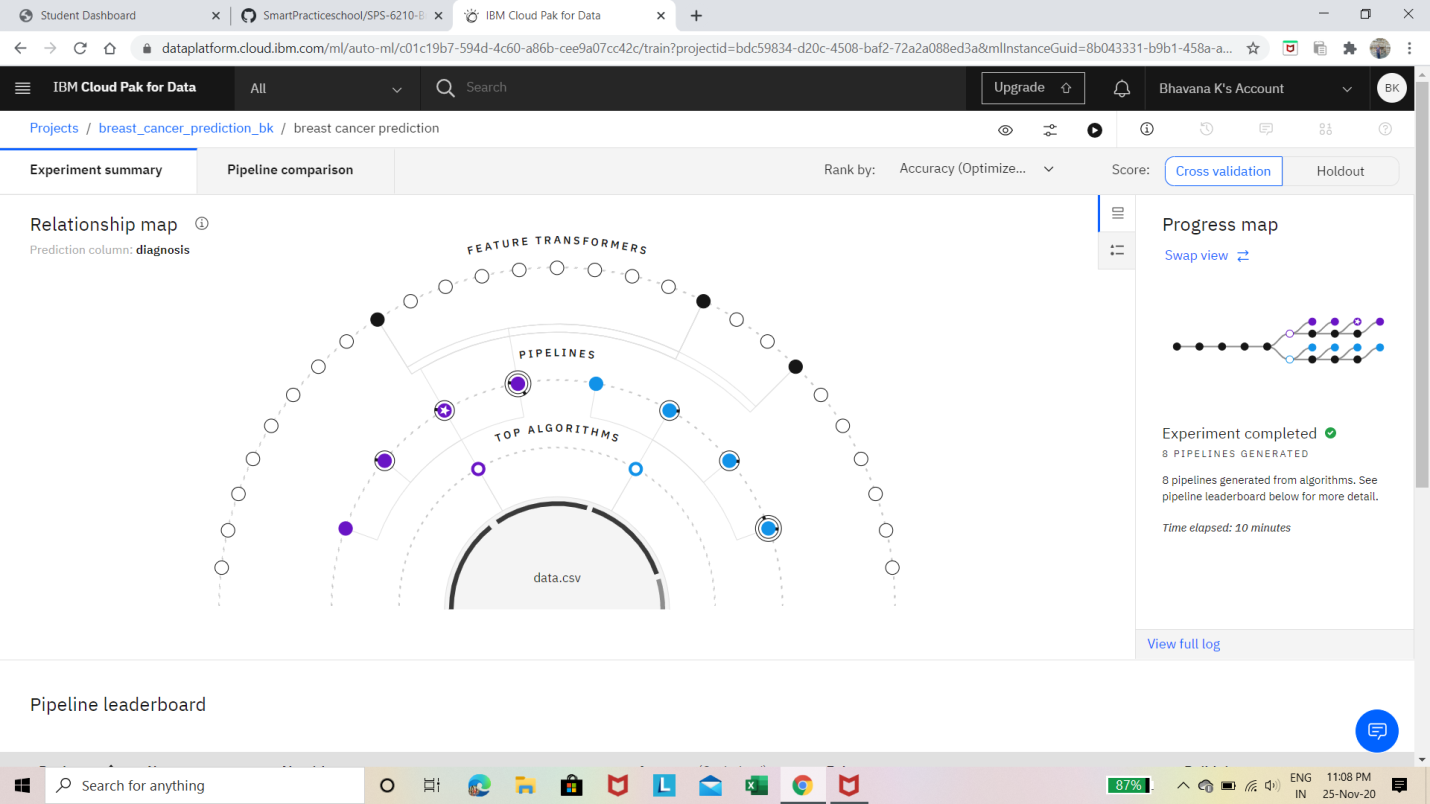
**Data set preview:**

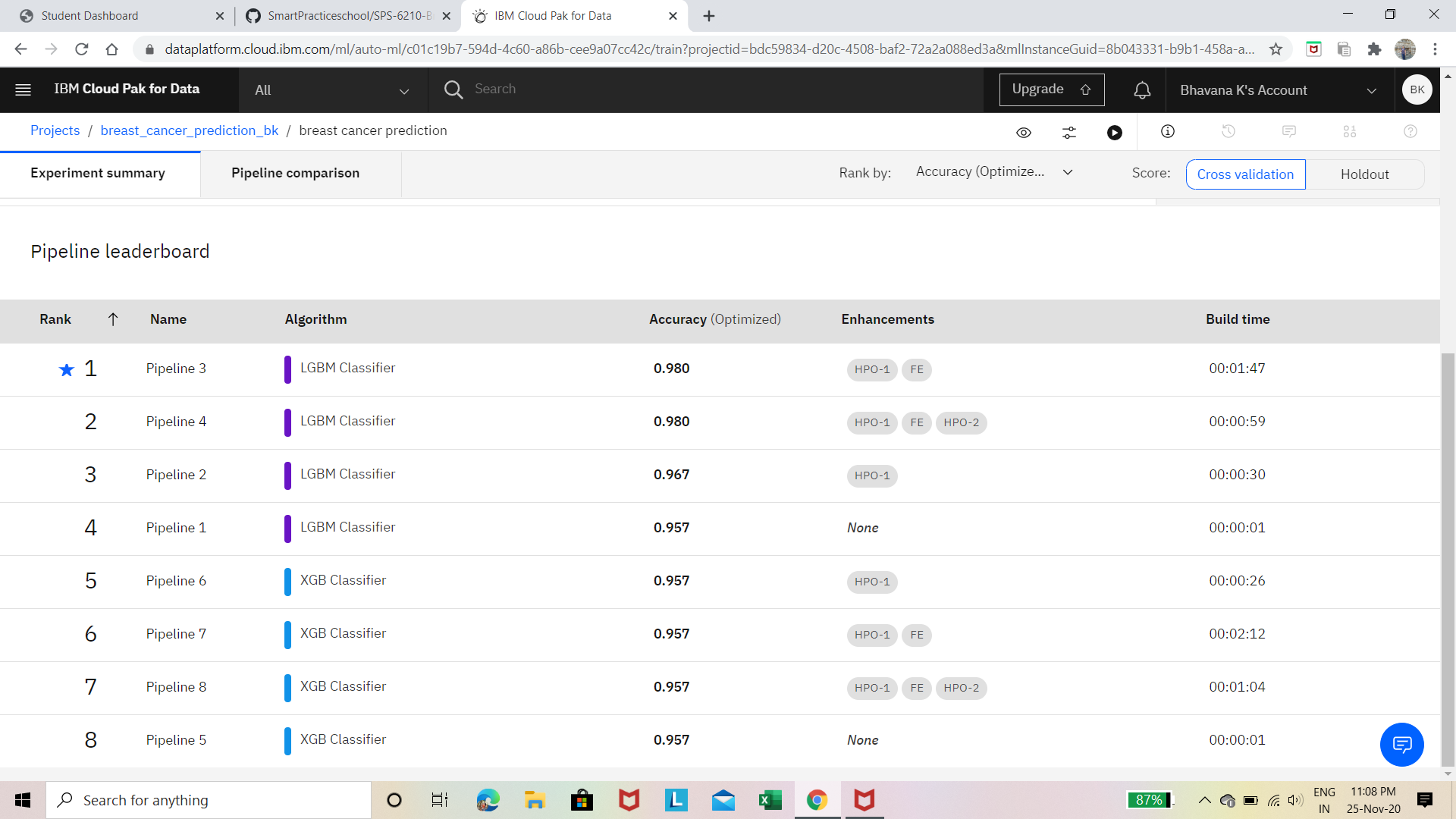
The following is the data which contains the above attributes:



**Auto AI Experiment:**

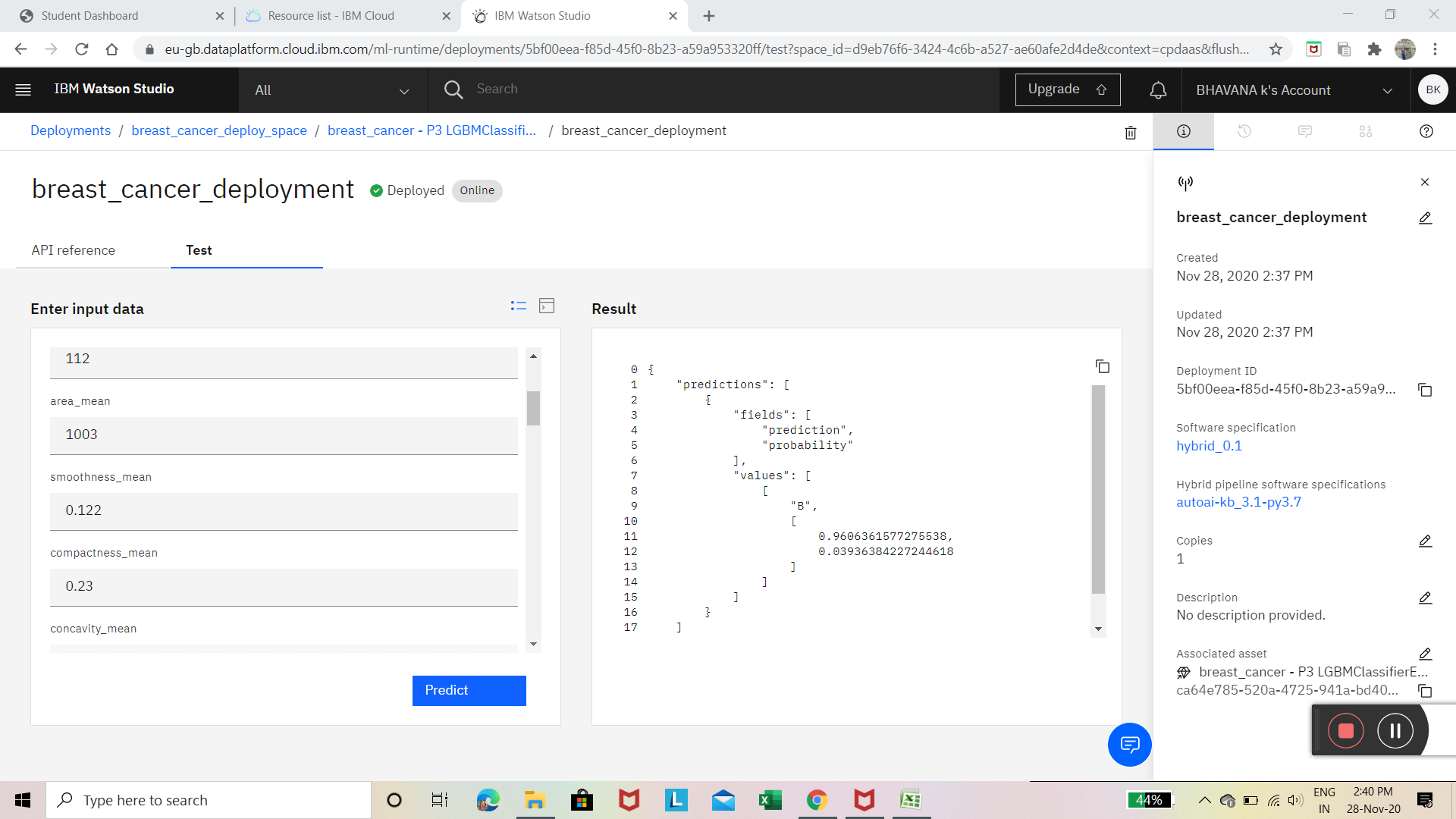
In auto AI experiment the data se is fed into it and machine learning instance is associated which predicts which model is best suitable for prediction.



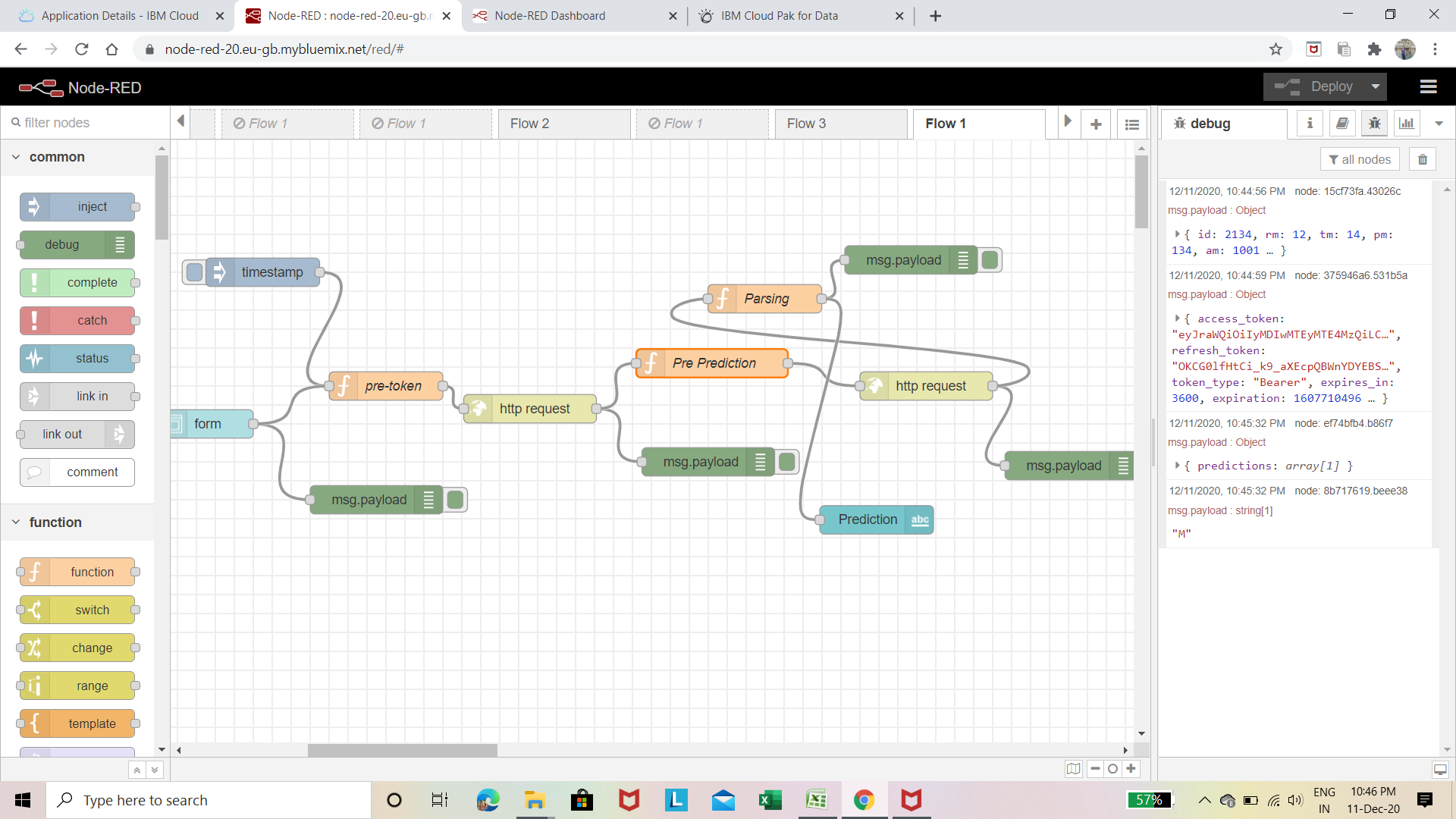


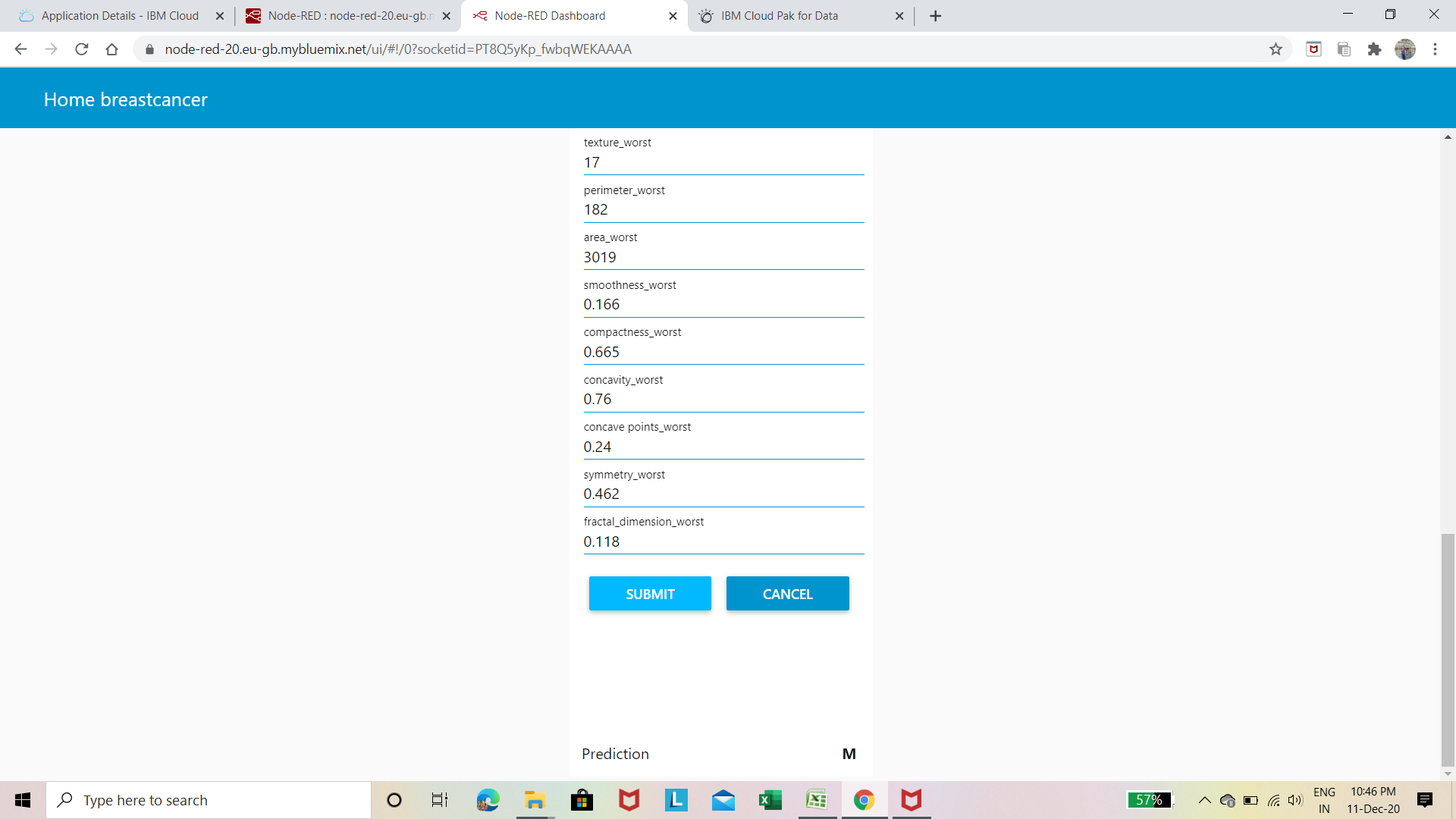
The AI model predicts the algorithm and this model is saved.

Then the model is deployed to get the predictions:



Then the model is connected to node red –UI model which has a form to take up inputs from user and predicts the diagnosis.





**Conclusion:**

In this we have studied different ML algorithms. The main objective is to predict breast cancer and to improve the lives of all people affected by it. The advantage of this system is that, the prediction process is less time consuming. It will help the doctors to start the treatments early for the breast cancer patients.

**Reference link:**

The data set is from is from the following link:

<https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>

\*\*END\*\*