**DISEASE PREDICTOR SYSTEM**

**PROJECT BY:**  52 - Atharva Deore

59 - Roshan Dhadiwal

62 - Ganesh Dhere

71 - Rohan Gaikwad

**ABSTRACT**

In today’s world, health plays a very important role in determining a person’s behavior. Being properly fit and healthy has tremendous advantages and plays a vital role in shaping a person’s career. Physical and Mental Health Affect Your Career in many ways. Both your physical and mental well-being are directly tied into how well you perform at work. When you're physically drained, you're more likely to get sick and miss work. A poor mindset can also impact your job by leading to burnout and decreased productivity.

Often due to changes in weather, or due to some ill habits regarding the diet or cleanliness leads to a person falling ill. Then it affects our work and daily routine. Based on what symptoms we have or whatever effects we are experiencing, the doctor gives us a prescription having certain medications listed that will help us to recover in a short amount of time.

But there are thousands of diseases around the world having symptoms that are almost common in each case. Predicting which disease, the person is suffering from, is quite a tedious task to be performed. What if there is a smart disease predictor system designed for this purpose, to detect the disease based on the symptoms….

**INTRODUCTION TO THE PROJECT**

The aim of this project is to develop a smart disease predictor system which will be able to predict what type of disease a person is suffering from, based on the symptoms entered by the person. This task can be achieved with the help of Machine Learning and Data analysis techniques.

The project consists of a well-designed user interface which enables the user to enter the symptoms he/she feels he is suffering from. Then by the use of certain machine learning algorithms predictions will be made on what disease he might have. The final output regarding the disease type will be consequently displayed.

## The Role of a Data Scientist in Healthcare

The role of a Data Scientist is to implement all techniques of Data Science for integrating it in healthcare software. The Data Scientist extracts useful insights from the data to make predictive models. Overall, the responsibilities of a Data Scientist in healthcare are as follows:

* Collecting data from patients
* Analyzing the needs of hospitals
* Structuring and sorting the data for use
* Performing Data Analytics using various tools
* Implementing algorithms on the data to extract insights
* Building predictive models with the development team

**DATA FLOW DIAGRAM**

Generating the UI and inputting the symptoms from the user

Predicting The disease for custom input from the user

Testing the model on the test dataset

Training the machine learning algorithms on the training dataset and creating a model for making the predictions based on the custom user input provided

Data preprocessing and visualization

Importing the training and testing datasets

**Benefits of Data Science in Healthcare**

Data Science helps in advancing healthcare facilities and processes. It helps boost productivity in diagnosis and treatment and enhance the workflow of healthcare systems. The ultimate goals of the healthcare system are as follows:

* To ease the workflow of the healthcare system
* To reduce the risk of treatment failure
* To provide proper treatment on time
* To avoid unnecessary emergency due to the non-availability of doctors
* To reduce the waiting time of patients

**ALGORITHMS IN USE**

**DECISION TREE ALGORITHM**

Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

**max\_depth is what the name suggests: The maximum depth that you allow the tree to grow to. The deeper you allow, the more complex your model will become. For training error, it is easy to see what will happen. If you increase max\_depth, training error will always go down (or at least not go up).**

* A significant advantage of a decision tree is that it forces the consideration of all possible outcomes of a decision and traces each path to a conclusion. It creates a comprehensive analysis of the consequences along each branch and identifies decision nodes that need further analysis.
* Help determine worst, best and expected values for different scenarios.
* Are easy to understand and interpret

**RANDOM FOREST ALGORITHM**

Random forest is a Supervised Learning algorithm which uses ensemble learning method for classification and regression.

Random forest is a bagging technique and not a boosting technique. The trees in random forests are run in parallel. There is no interaction between these trees while building the trees.

It operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

A random forest is a meta-estimator (i.e. it combines the result of multiple predictions) which aggregates many decision trees, with some helpful modifications:

**KNN ALGORITHM**

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

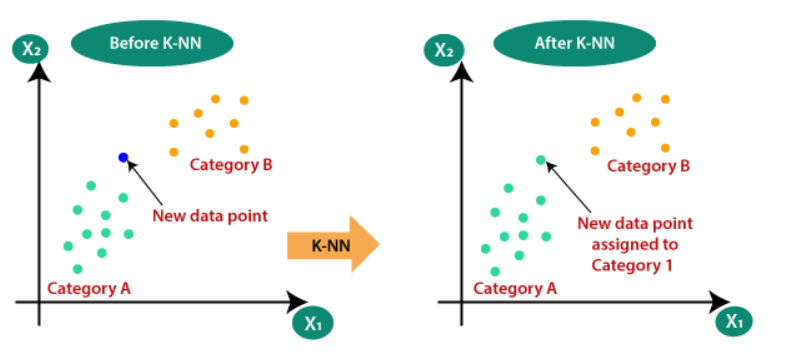
K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x1, so this data point will lie in which of these categories. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset. Consider the below diagram:



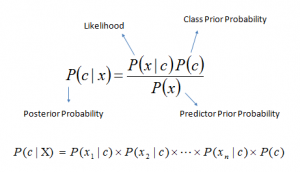
**NAÏVE BAYES ALGORITHM**

It is a classification technique based on Bayes’ Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Definition in simple words :- You know something about the world, and based on what you know, you setup a probability model and you write down probabilities about the different outcomes. Then someone gives you some new information, which changes your beliefs and thus changes the probabilities of your outcomes.

Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Bayes theorem provides a way of calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:

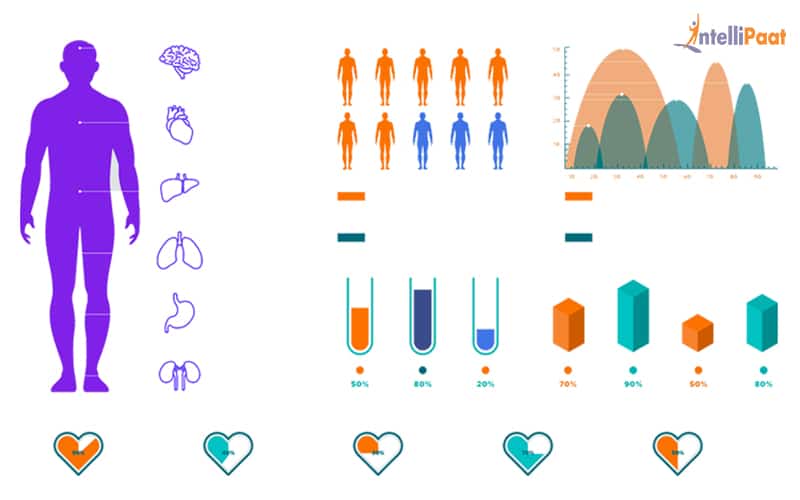


**Why do we use Data Science in Healthcare ?**

According to a study, the data generated by every human body is 2 terabytes per day. This data includes activities of brain, the stress level, heart rate, the sugar level, and many more. To handle such a large amount of data, now, we have more advanced technologies and one of them is Data Science. It helps monitor patients’ health using recorded data.

With the help of the application of Data Science in healthcare, it has now become possible to detect the symptoms of a disease at a very early stage. Also, with the advent of various innovative tools and technologies, doctors are able to monitor patients’ conditions from remote locations.

In earlier days, doctors and hospital managements were not able to handle multiple numbers of patients at the same time. And due to the lack of proper treatment, the patients’ conditions used to get worse.



However, now, the scenario has changed. With the help of Data Science and Machine Learning applications, doctors can be notified about the health conditions of the patients through wearable devices. Then, hospital management can send their junior doctors, assistants, or nurses to these patients’ homes.

Hospitals can further install various equipment and devices for the diagnosis of these patients. These devices built on top of Data Science are capable of collecting data from the patients such as their heart rate, blood pressure, body temperature, etc. Doctors get this real-time data of the patients’ health through updates and notification in mobile applications. They can then diagnose the conditions and assist the junior doctors or nurses to give specific treatments to the patients at home. This is how Data Science helps in caring for patients using technology.

Some Key points for use of DS in Health Care :

* It helps in the management of chronic diseases.
* It efficiently monitors and analyzes the demand for pharmaceutical logistics.
* It predicts a patient’s condition and suggests preventive measures.
* It provides faster documentation of hospital data.
* It helps in efficiently utilizing doctors and other resources for the benefit of the maximum number of patients.
* It predicts the future medical crises of a patient.

**Future of Data Science in Healthcare**

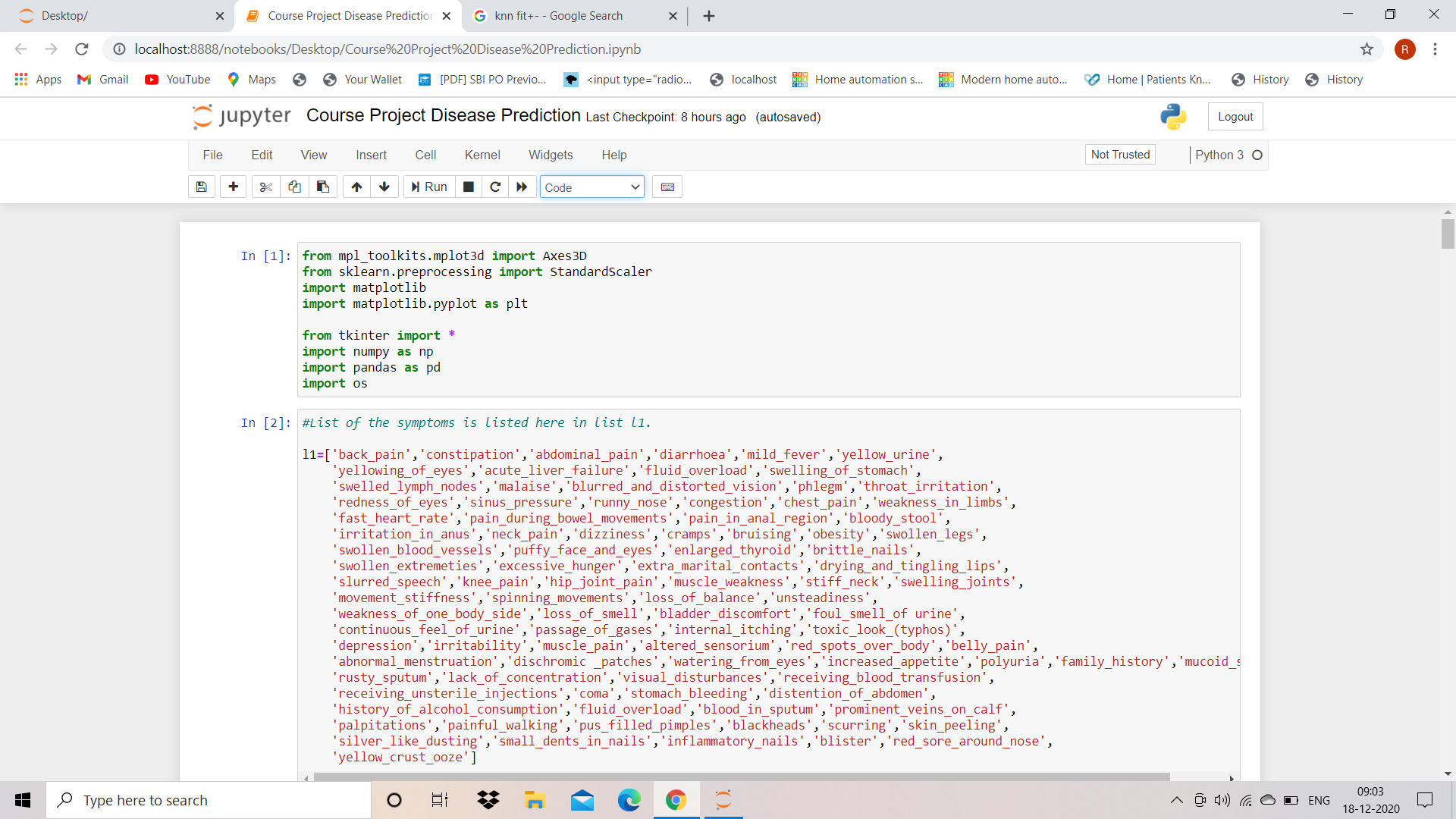
Basically, there are four factors leading to rapid improvement in the healthcare industry:

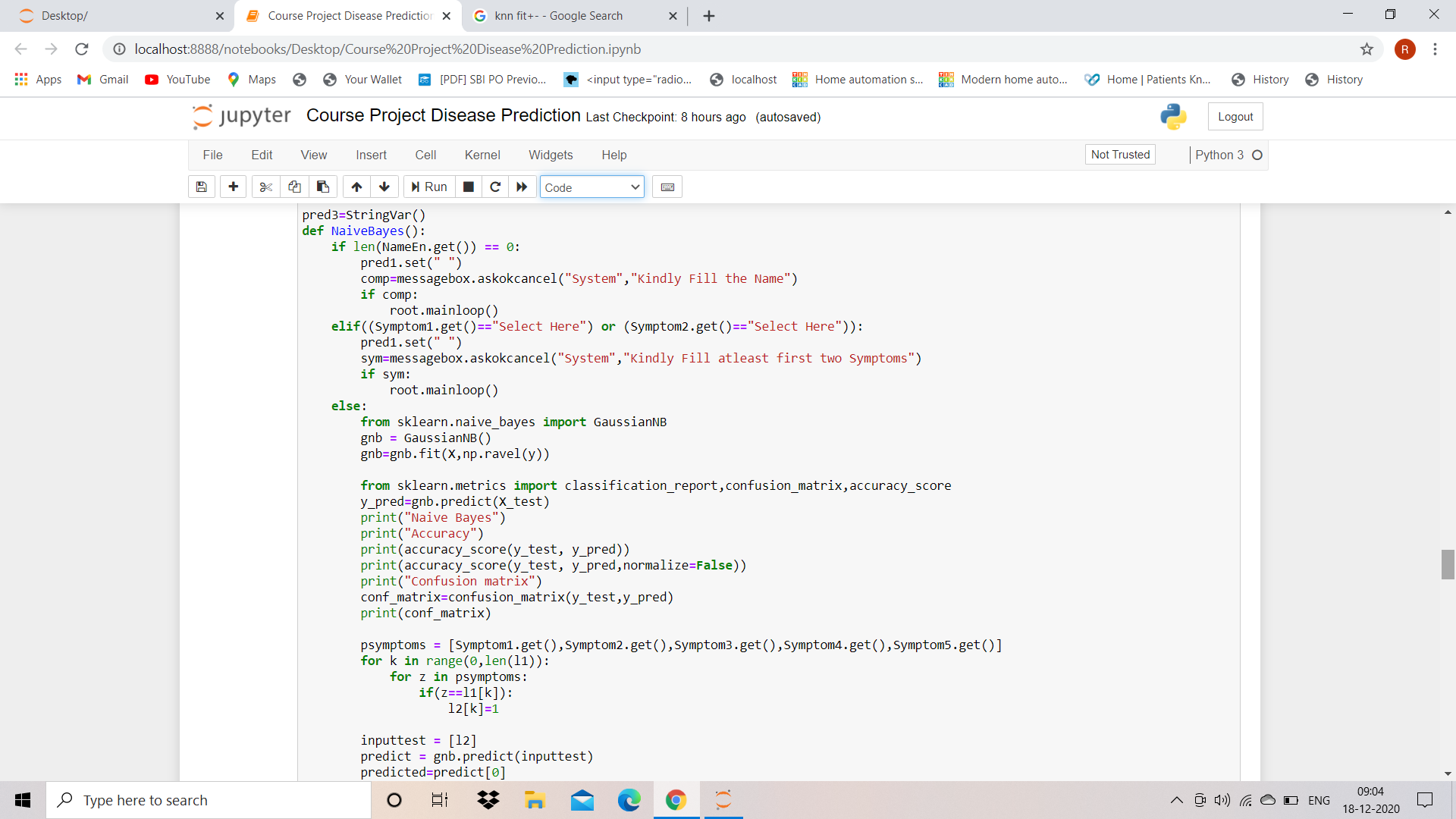
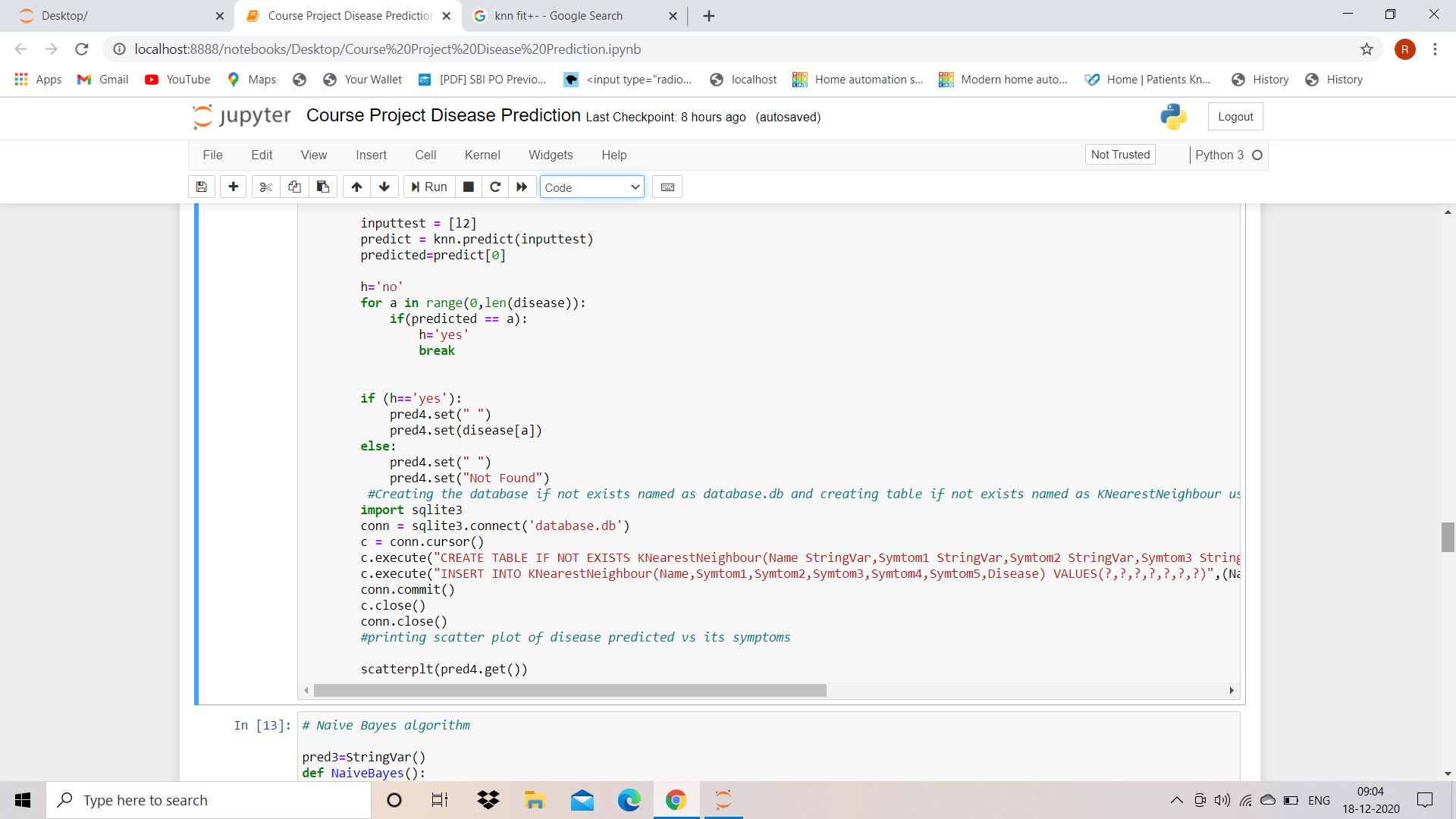
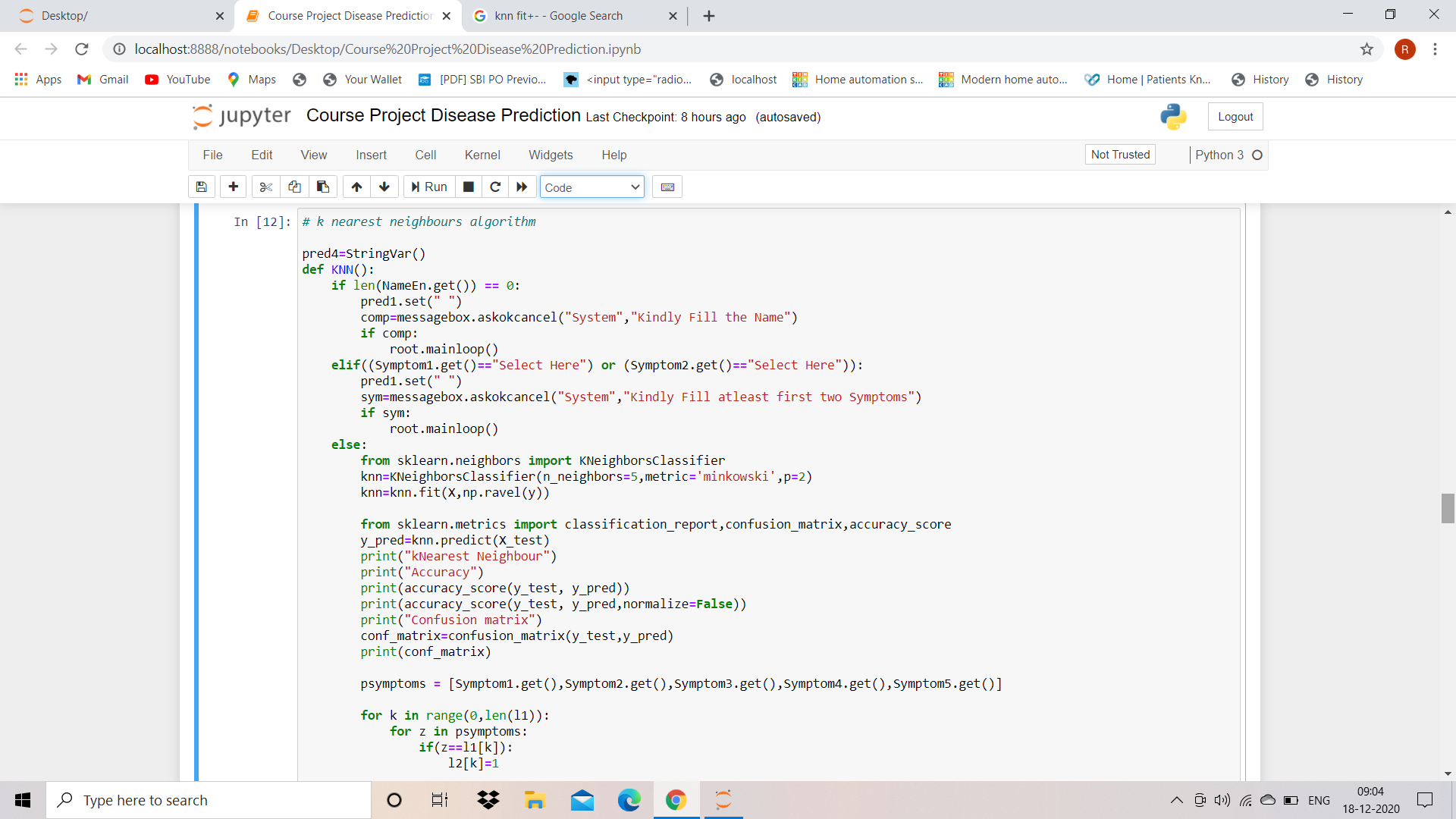
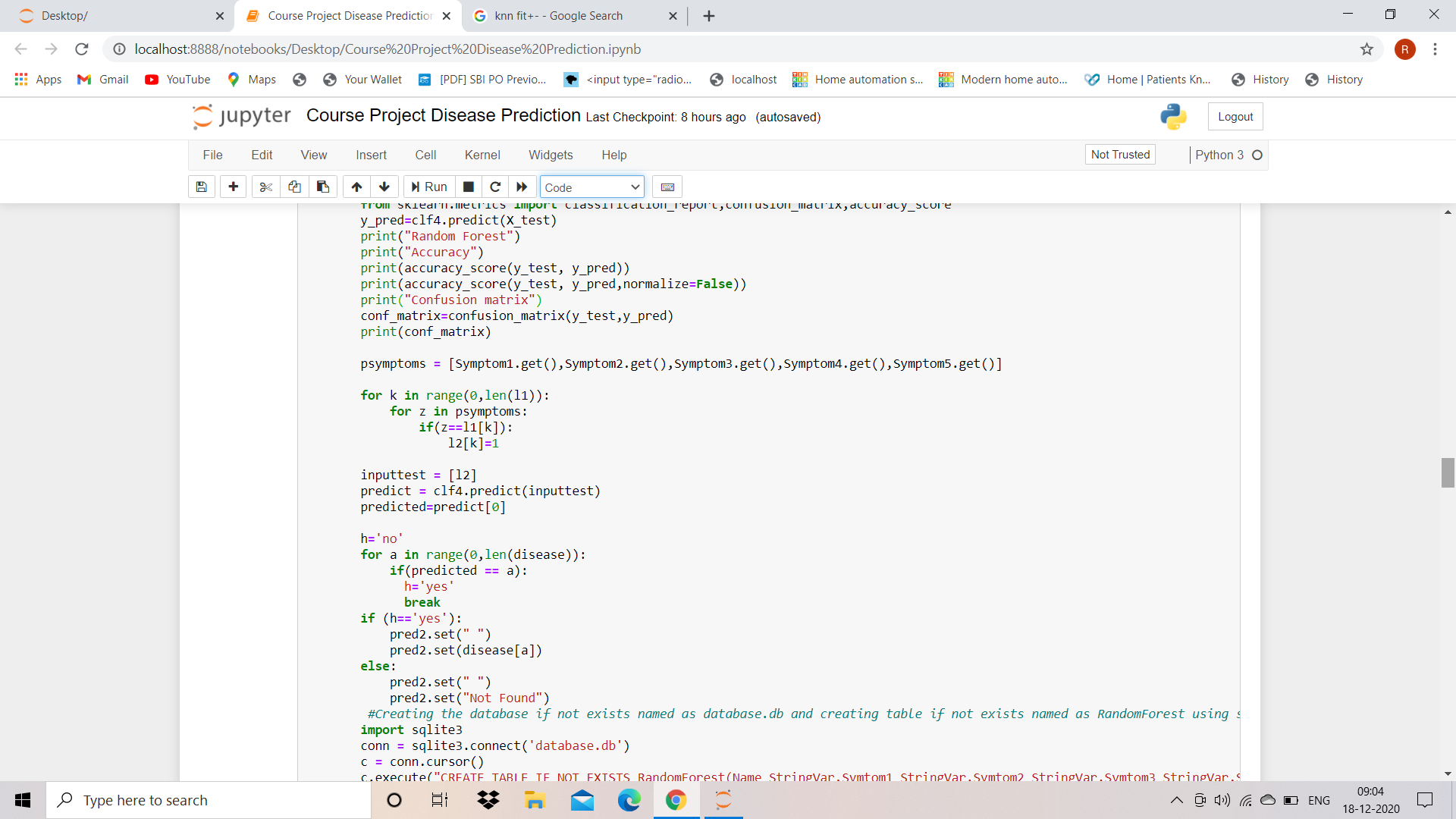
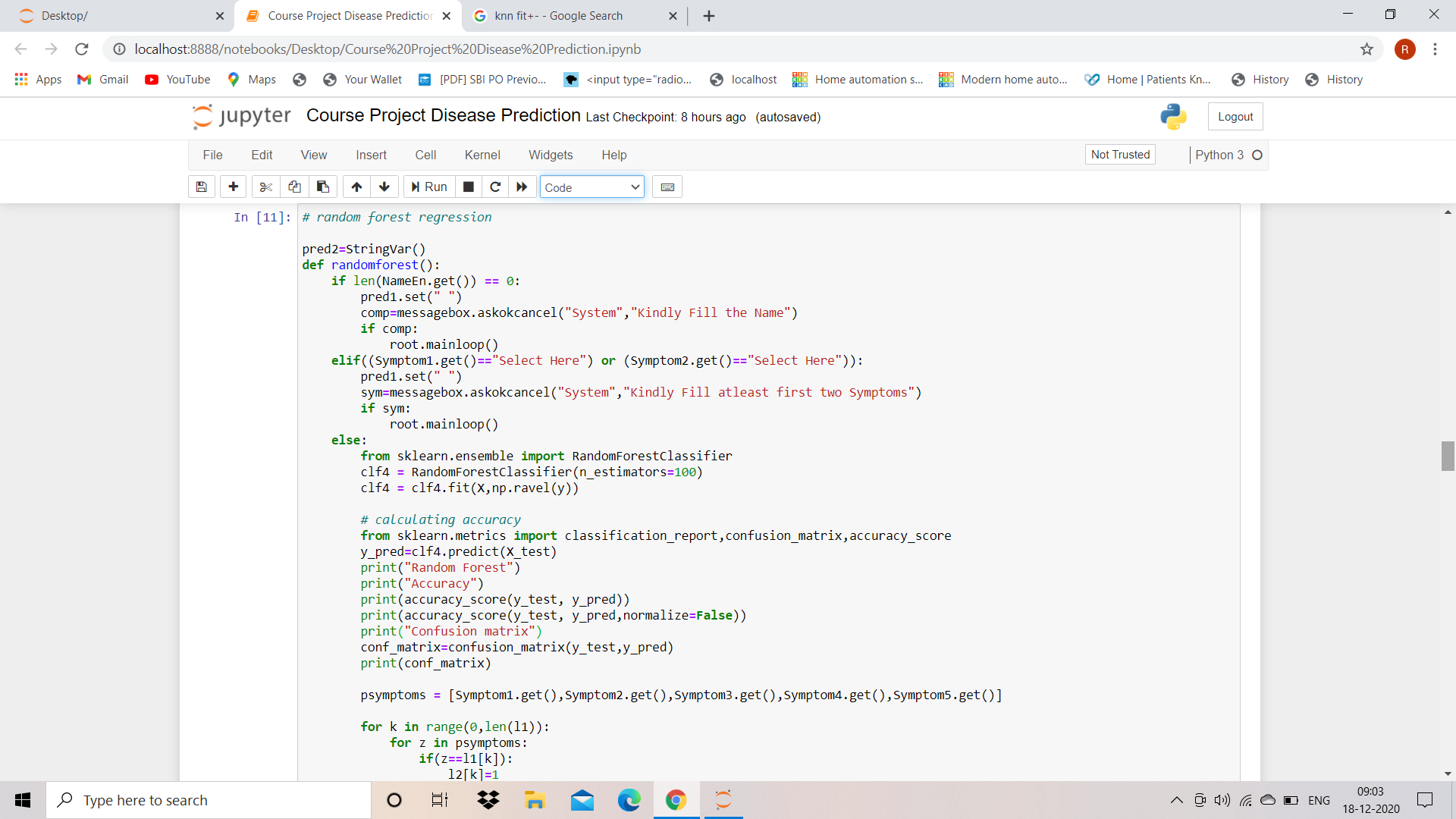
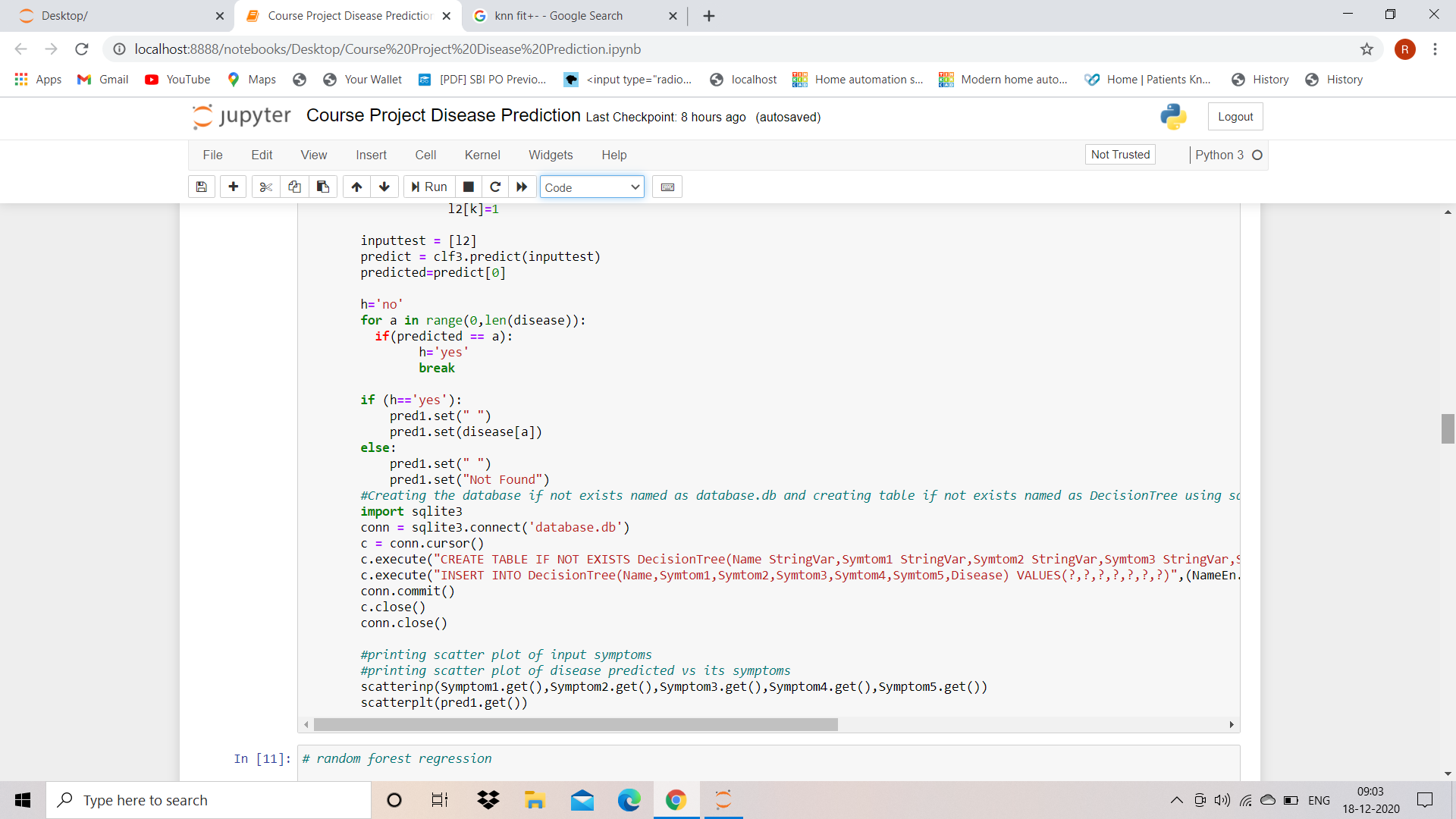
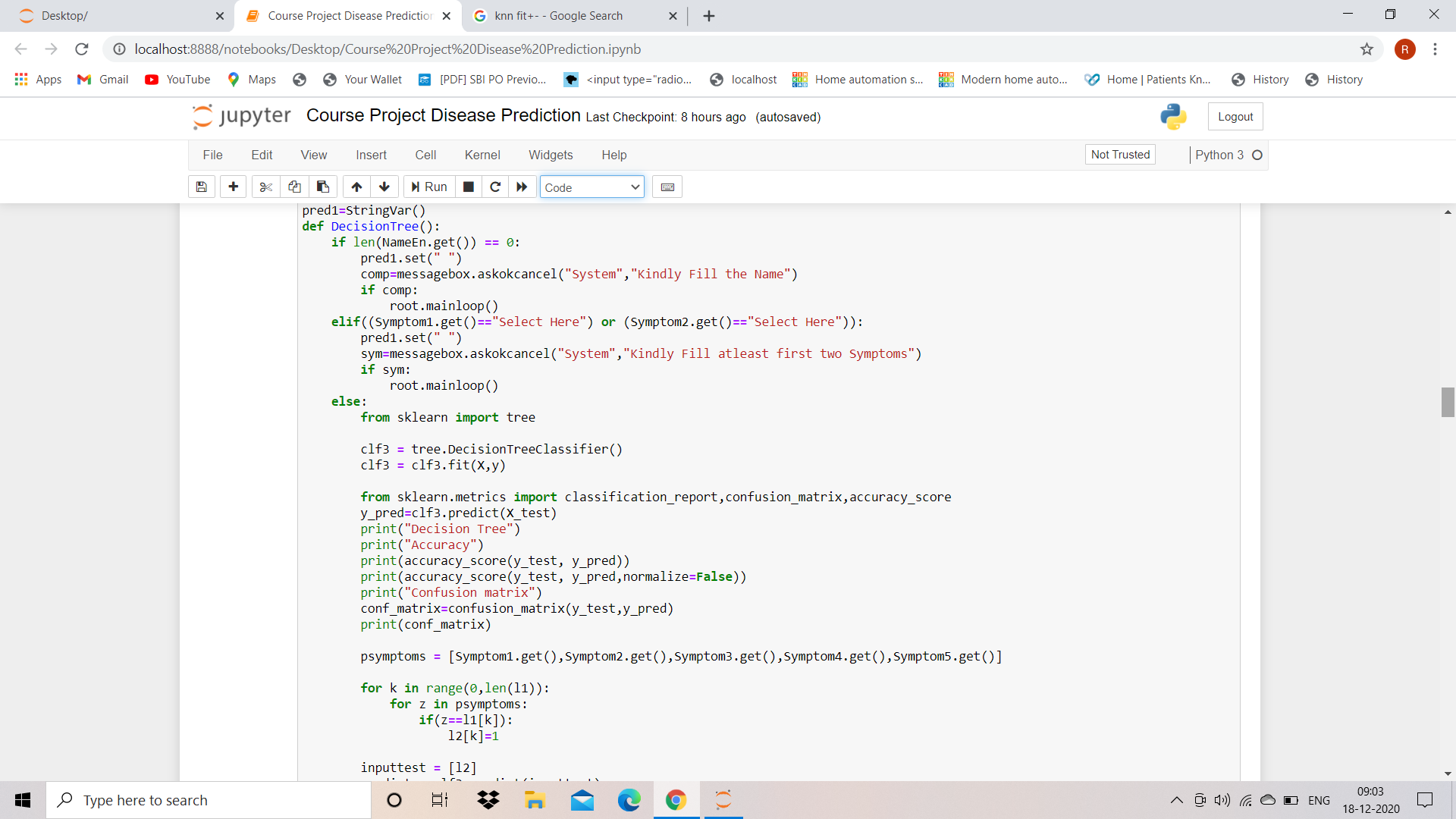
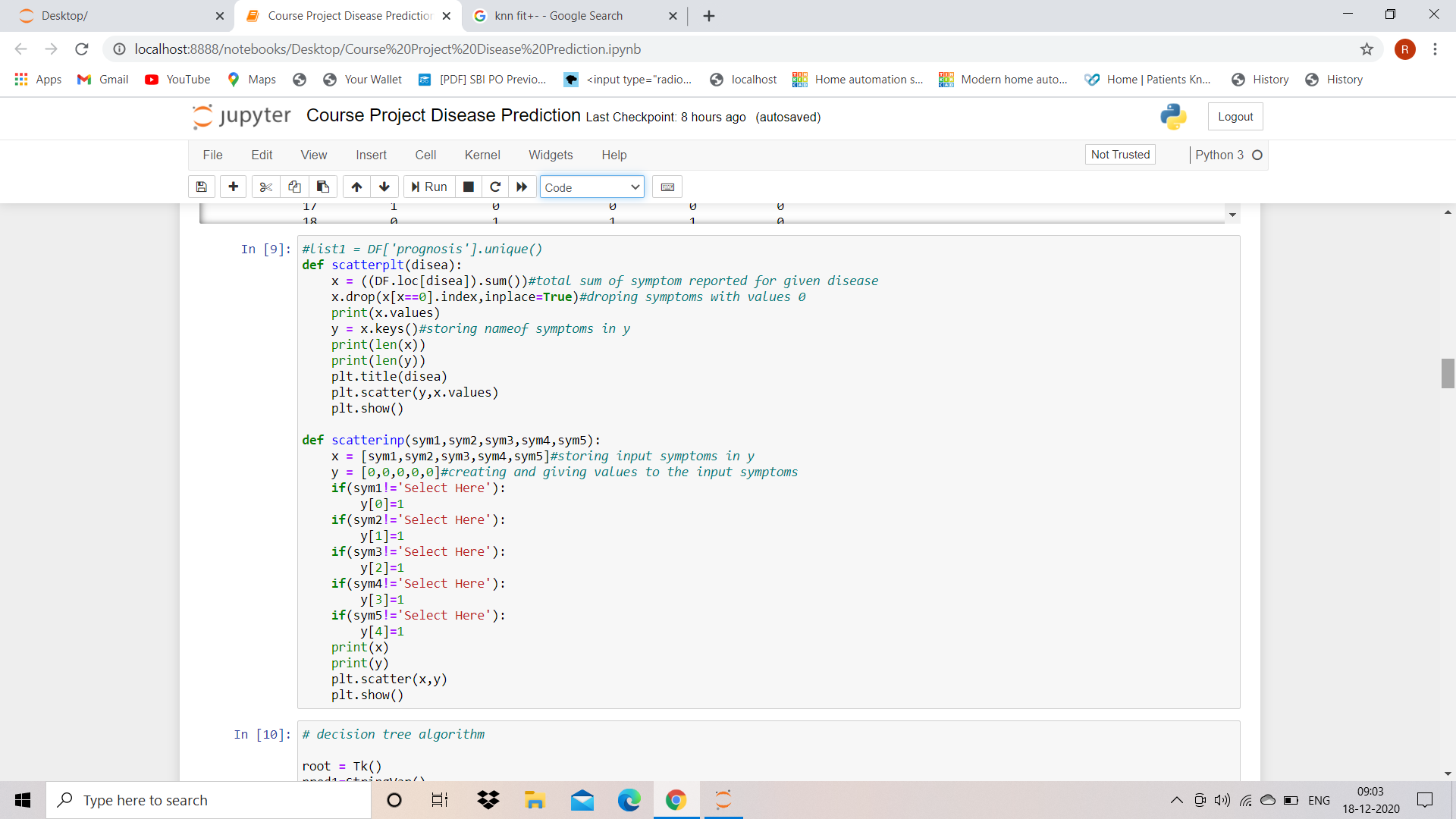
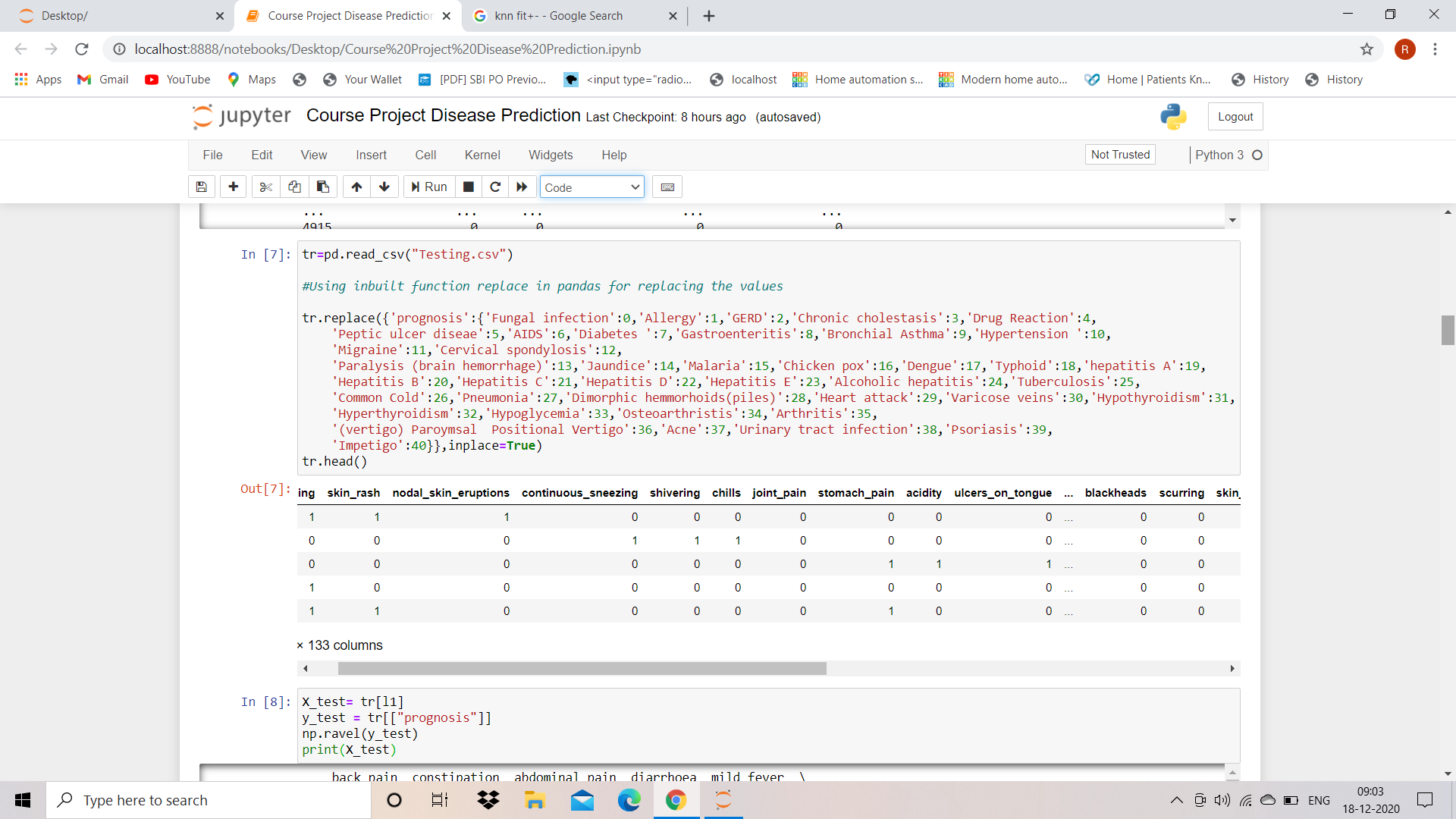
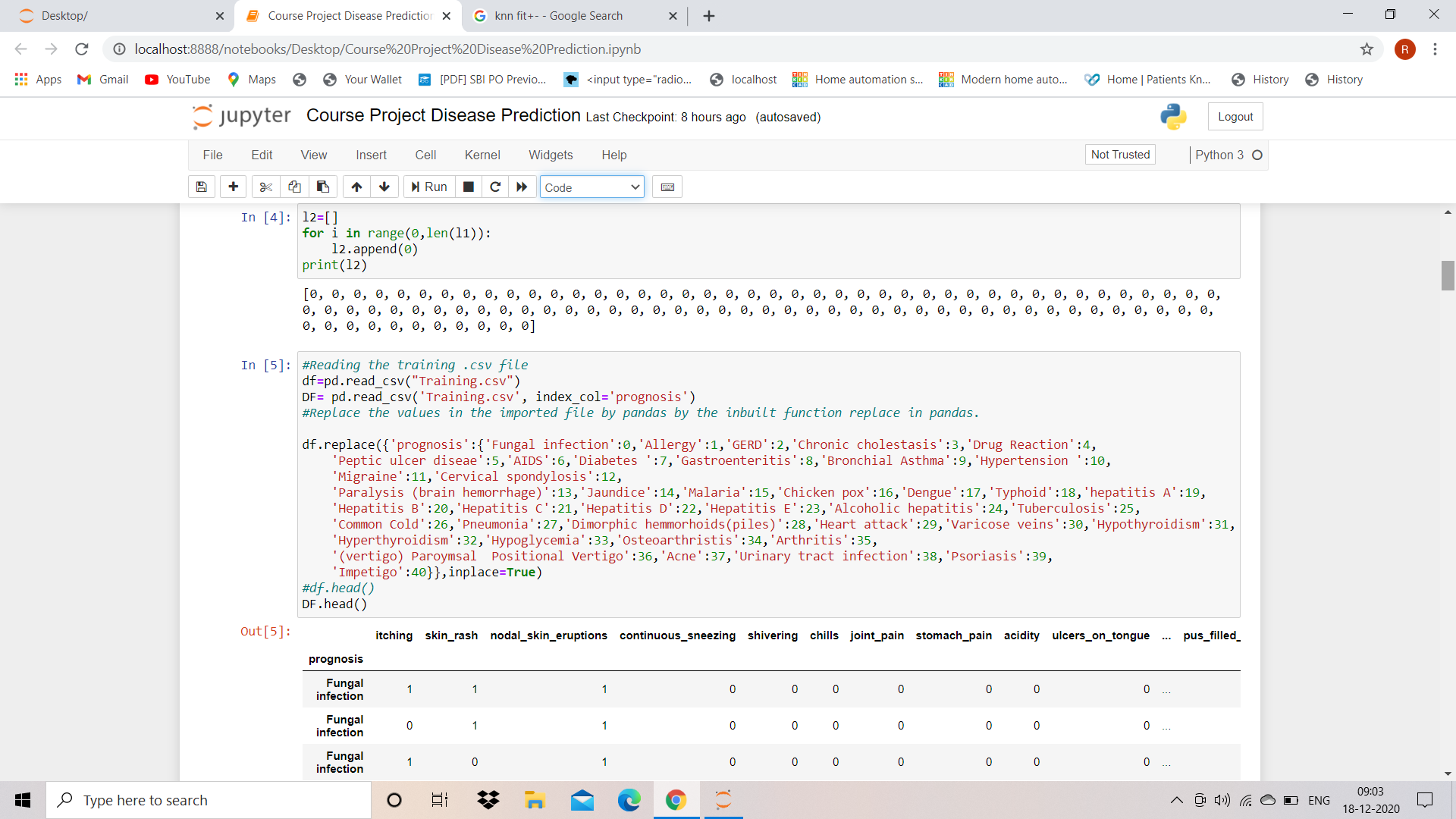
* Technological advancements
* Digitalization
* Need for reducing treatment costs and duration
* Need for handling large population

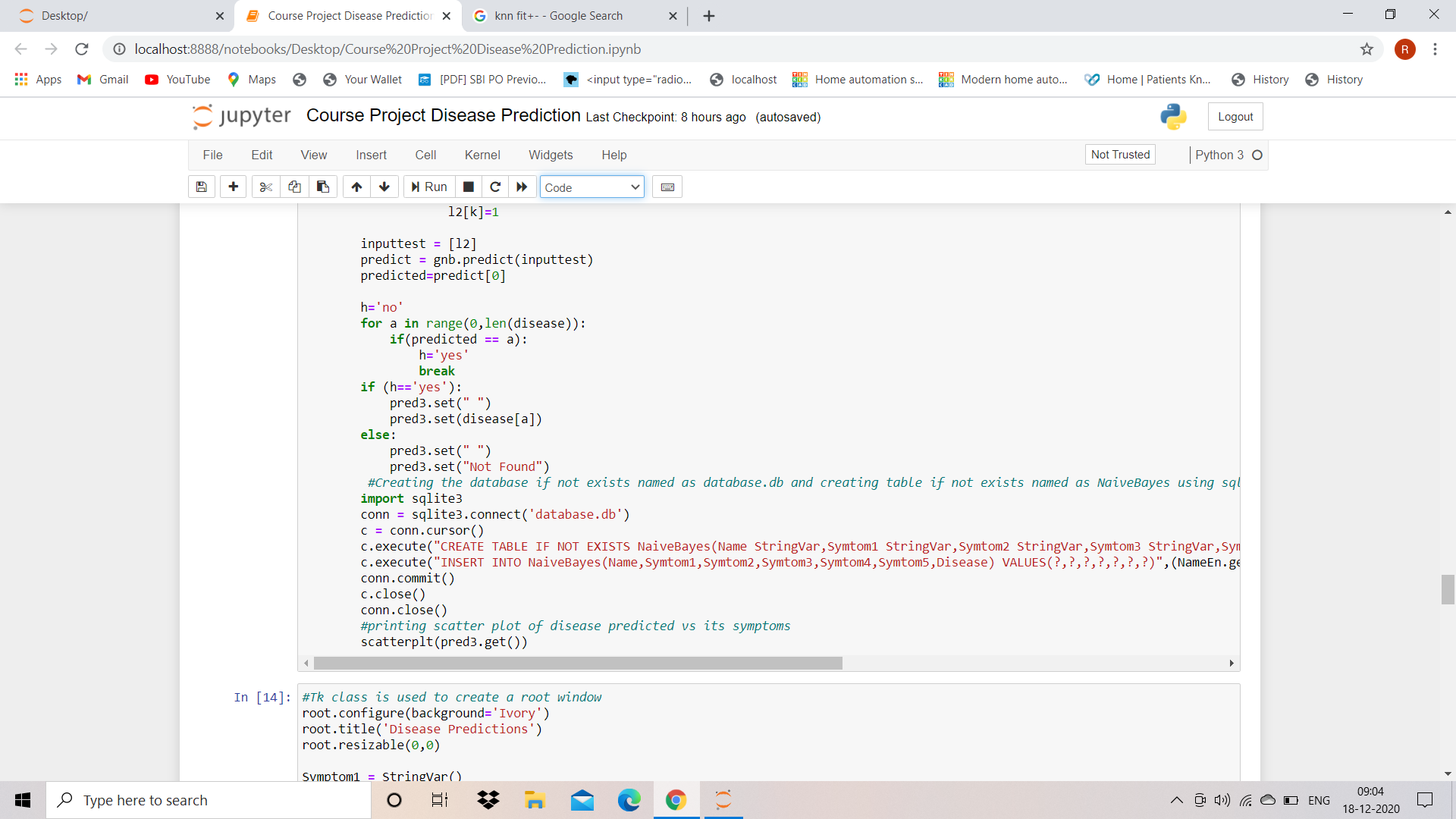
Data Science has already started addressing all these to bring the desired effect. As Data Science is already doing wonders for society, there is no doubt that its application in the future will prove itself to be more invaluable. It will take the healthcare industry to further heights. Doctors will get ample assistance and patients will get a more personalized experience and perfect treatments.

To conclude, the applications of Data Science in healthcare has the potential to enhance the entire healthcare system.

**CODE SNIPPETS**







**OUTPUT SNIPPET**

