Problem statement is to build a model which classifieds whether a wafer is working or not or next question is what a wafer actually is slice of semiconductor material which is used in various electronic circuits electronic devices it is also used with photovoltaic cells so just imagine that the scenario is like our client has a production line and on that production line various wafers are deployed and whenever there is a fault in any wafer they have to manually check all the wafers and see which wafer is faulty stop the entire production line and then replace that faulty wafer so this is the problem that our client is facing and now for a solution of this problem our client has asked us to create a model which successfully identifies at which for wafer is faulty and start their operation so that the entire production line it does not get injured so client at their site based on our predictions whatever we are going to make with the wafer is faulty or not based on that client is going to set an alarm at their end that and it will be triggered whenever we say that a certain wafer is faulty so it will get triggered at that point of time and only that particular section on the production line will be stored there that wafer is deployed it that wafer will be replaced and again the production line will resume without any hindrance so this is what our problem statement is.

For solving this problem statement we are getting a data like this so now if you look at our data so the data consists of various columns so the first column the column it consists of the wafer name or the wafer number so that you know that that row number two let's say this row #2 is the reading for this wafer number 501 and the readings from different sensor sensor 1 to sensor 590 so we have 590 sensors which are getting readings or which are monitoring all the 1500 sensors that have been deployed by the client and based on these readings we have a target column and that target column is this good or bad column and now based on this good or bad column value we have to identify that whether a certain wafer it is in working condition or not so there are two scenarios that can be possible so for bad we'll get one values and if a certain wafer it is not working we are going to get minus one for those values and if a wafer is working or so in that scenario the value is going to be plus one so end goal is to create a model which predicts either minus one or plus one so it is a binary classification problem that we are going to solve so our end goal will be to predict based on the inputs from all these sensors right from sensor 1 to censor #590 we have to predict the minus one or plus one value that will decide whether the wafer is working or not once we return that value the client in their end and they will have a mechanism to identify or to trigger alarm saying that OK wait for numbered this this row for row number this side a certain wafer number is given so they might be able to identify that OK wafer number 501 is not working where it is deployed so some techniqion go and they will only stop that particular part of the production line they do not need to stop the entire production line and then again that wafer is when that wafergets replaced the entire production line or it will start working without any hindrance so this is what we are going to achieve using our solution scratch