Covid Risk Analyzer and Pooled Testing

It's impossible to contain covid-19 without knowing who's infected, until a safe and effective vaccine is widely available. If we combine machine learning with test pooling, large populations can be tested weekly or even daily, for as low as Rs220 to Rs370 per person. Infrequent testing (monthly seems to be the default in many proposals) or haphazard screening allow active cases to spread the virus for weeks before it's caught. And the price is still high at around Rs1000 per test.

Pooled testing, guided by machine-learning algorithms, can fundamentally change this calculus. In pooled testing, many people's samples are combined into one. If no virus is detected in the combined sample, that means no one in the pool is infected. The entire pool can be cleared with just one test. But there's a catch; if anyone in the pool is infected, the test will be positive and more testing will be required to figure out who has the virus. Machine learning can give us the precise individual-level estimates we need to make pooling work, by identifying those likely to test positive and keeping them out of large pools. Such groups of risky and non-risky participants can be made based on the following factors that can be fed into the machine learning algorithm:

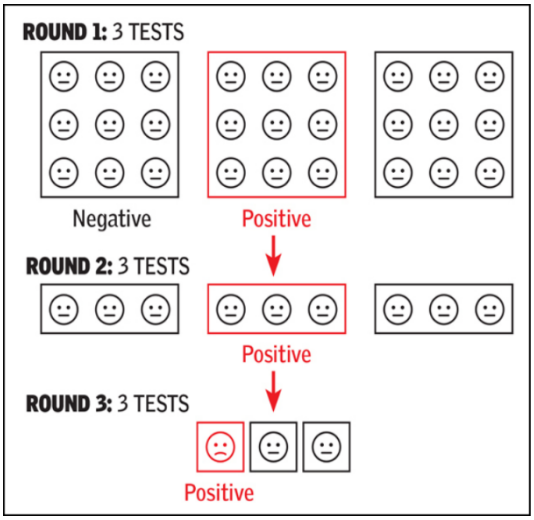
1) Occupation (for example, healthcare workers and policemen are more prone)
2) Residential location
3) Workplace location
4) Patients with obvious symptoms
5) Travel history
6) Whether participant has been in close contact with positively tested person or not.
7) Immunity level from previous hospital records.
8) Blood Type (Once researchers have established evidence that people with type O blood have lower risk)

To understand pooled testing strategies (1-3), we'll make some assumptions:

- Total number of people to be tested: 27
- Number of pools: 3
- People in each pool: 9

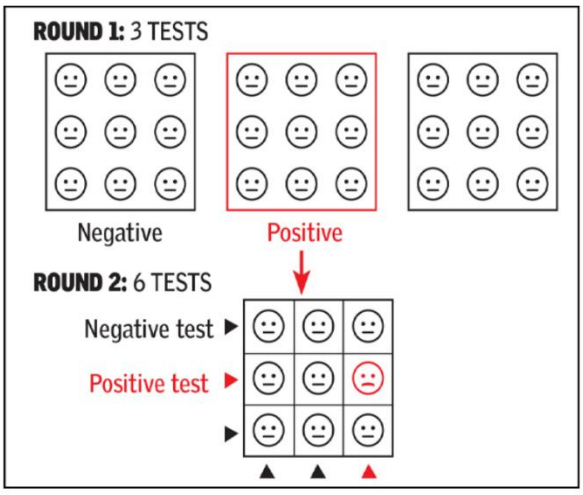
STRATEGY 1:

Now, say, of the 3 pools (let's call them A, B, C), two show no positivity, that is, these groups are clear of any infections and only one group of 9 people (say B) returns a positive result. That would mean that at least one of the 9 people in Pool B is positive. To zero in on which person that is all that the clinic has to do is to test the 9 samples individually. Thus, 27 people were tested using 3 + 9 = 12 kits.



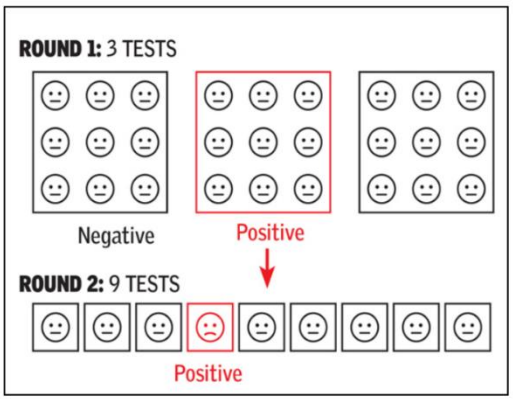
STRATEGY 2:

Again, assuming that only one of the three pools returns a positive test, it means that at least one of the 9 people in that particular group is positive. Now comes a variation on Strategy 1. Instead of testing the 9 samples in the positive pool individually, the clinic breaks them down into 3 smaller pools of 3 samples each and tests them again. When one of these smaller groups returns a positive result, the clinic can individually test just those 3 samples to zero in on the infected sample. In the second method, 27 people were tested using 3 + 3 + 3 = 9 kits



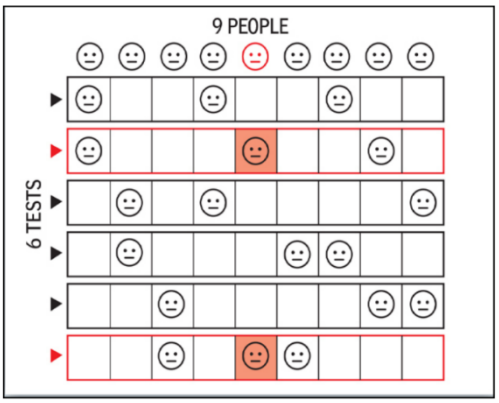
STRATEGY 3:

The above strategies leave room for improvement in terms of time taken to identify positive cases and the number of rounds of testing needed. To solve this, there are two other, slightly complicated, strategies. The first of these is like a grid system. Here's how it works: 27 samples are split into 3 pools of 9 each, just like the previous strategies. Once an infection is found, these 9 samples are split into horizontal (rows) and vertical (columns) grid. Now, the horizontal samples are each tested taking one row as one pool. So, that's three tests. The vertical samples are also tested with one column as one pool, which makes it another three tests. Only one sample can be common in the horizontal and vertical grids, thus leading the clinic to the infected sample. In the third method: 27 people tested using 3 + (3 + 3) = 9 kits.

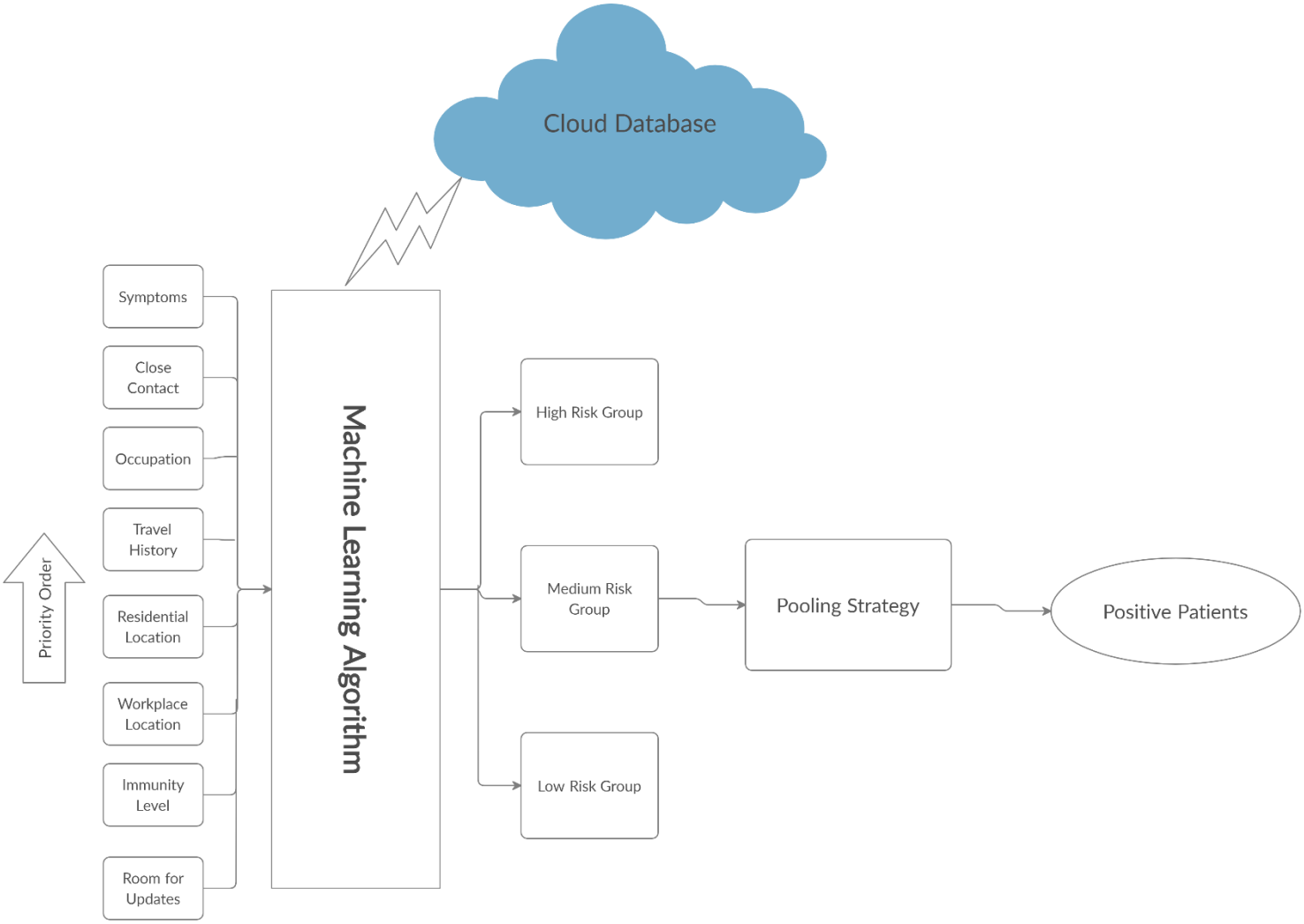


STRATEGY 4:

Time is of the essence and even two rounds of testing could be too much. In this strategy we do pooled testing in one round using "many overlapping groups". While this method increases the number of tests, it saves time. Samples are distributed in different pools (rows) as per mathematical rules. Every sample falls in two pools, and there is only one overlap between any two pools. So when you know which two pools have tested positive, you will also know which is the common sample that is positive. Every test should also include the same number of samples (3 in this case). In unpublished results from clinical trials in India, 5 positive samples were identified out of 320 using only 48 tests.



High Level Architecture Diagram



Our Solution

We use the AI Watson Assistant to ask the participant for the following information and save it on a database on IBM cloud. An additional column calculates the total score gathered by each participant which then decides what risk level group they get admitted to. The score for Travel, Residential and Workplace locations are calculated with the help of machine learning. The AI compares the user's input with that of up to date covid hotspot locations from online sources to decide how risky the location is. The occupations are also given particular score based on risk factor with the help of online data. The three risk level groups can be used by hospitals to apply one of the strategies described earlier and can help save time and cost per test kit.

Services Used

Here are the IBM cloud services we used to implement our idea:

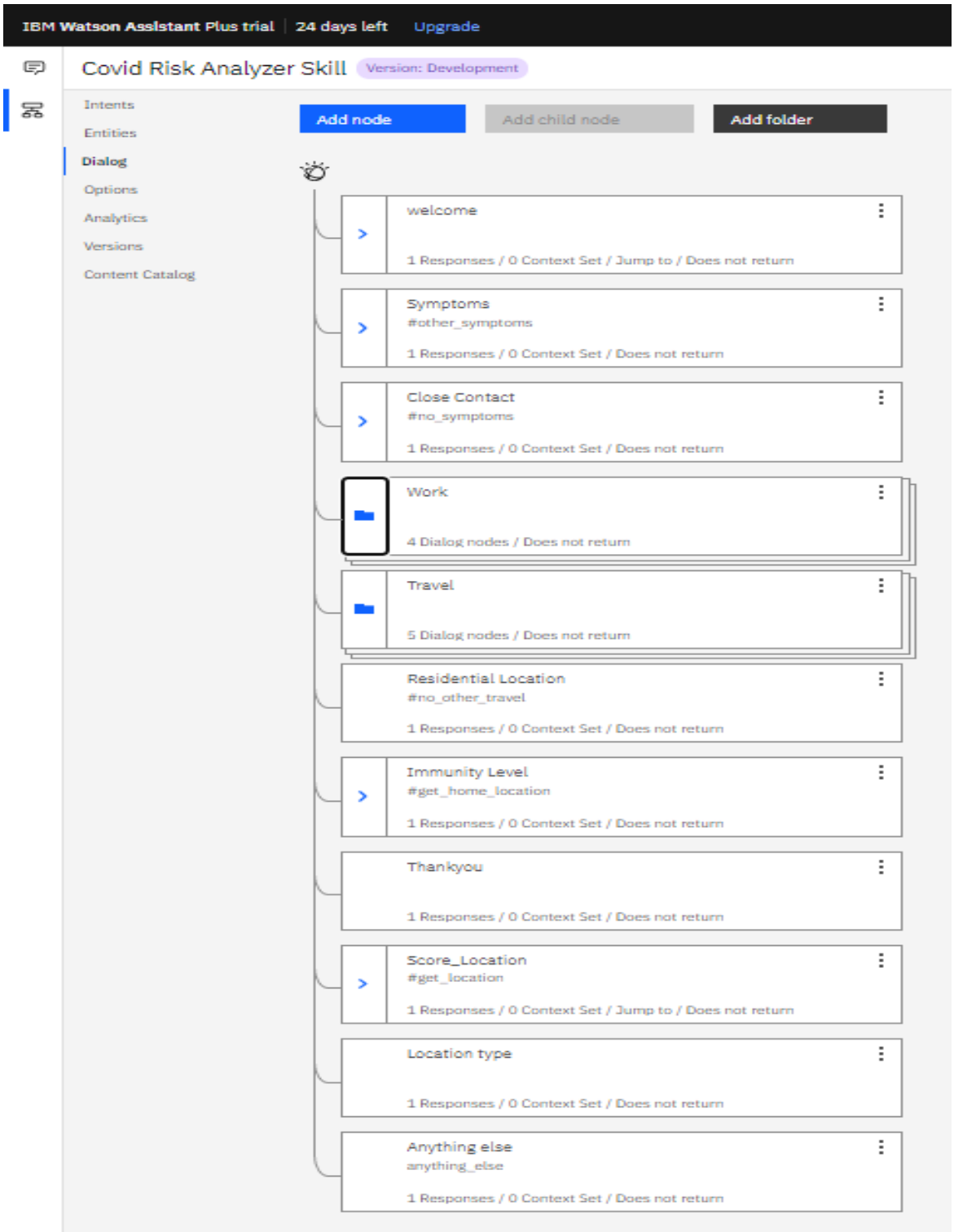
1. Watson Assistant: We made the Covid Risk Chatbot using this API. This chatbot asks the user a series of questions and returns a score for each answer given by the user. Some scores are predetermined while some (like location and occupation) are found by scanning online sources hence the chatbot is always up to date.
2. Watson Discovery: We leveraged this API to the chatbot using a search skill. This helps the chatbot find online data for risk scores of occupation and location.
3. Cloudant: We have made use of IBMs powerful cloud database to store users inputs into a table. This will prove useful in the future for statistical analysis.
4. Node Red: It is a flow based development tool that helps connect our various services together. We have also used it to create our Covid Risk Calculator, which is a form that user can enter their scores into and find out their average score and risk level. Finding out the risk level of several patients can help hospitals use pooled testing techniques more efficiently and save time and cost per test kit.

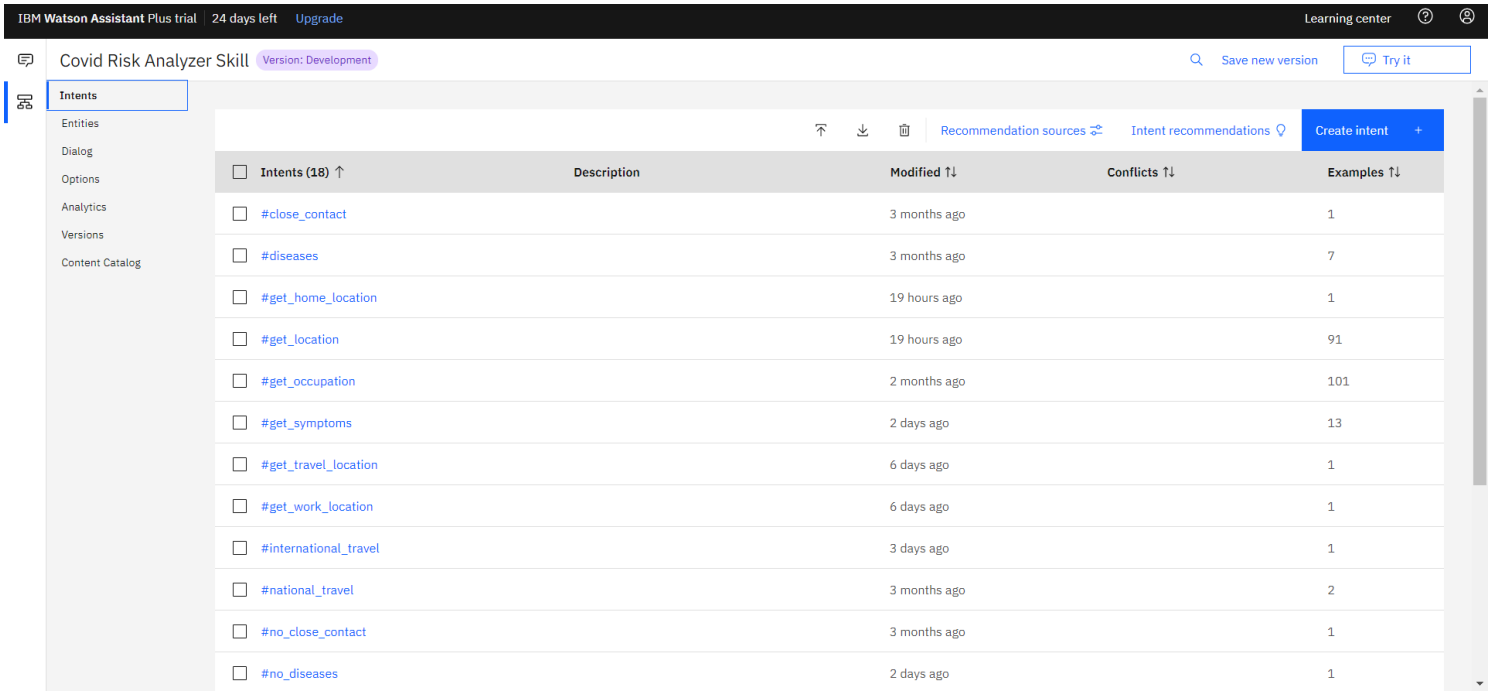
Our Implementation

Covid Risk Chatbot

Here's a link to try out our chatbot:

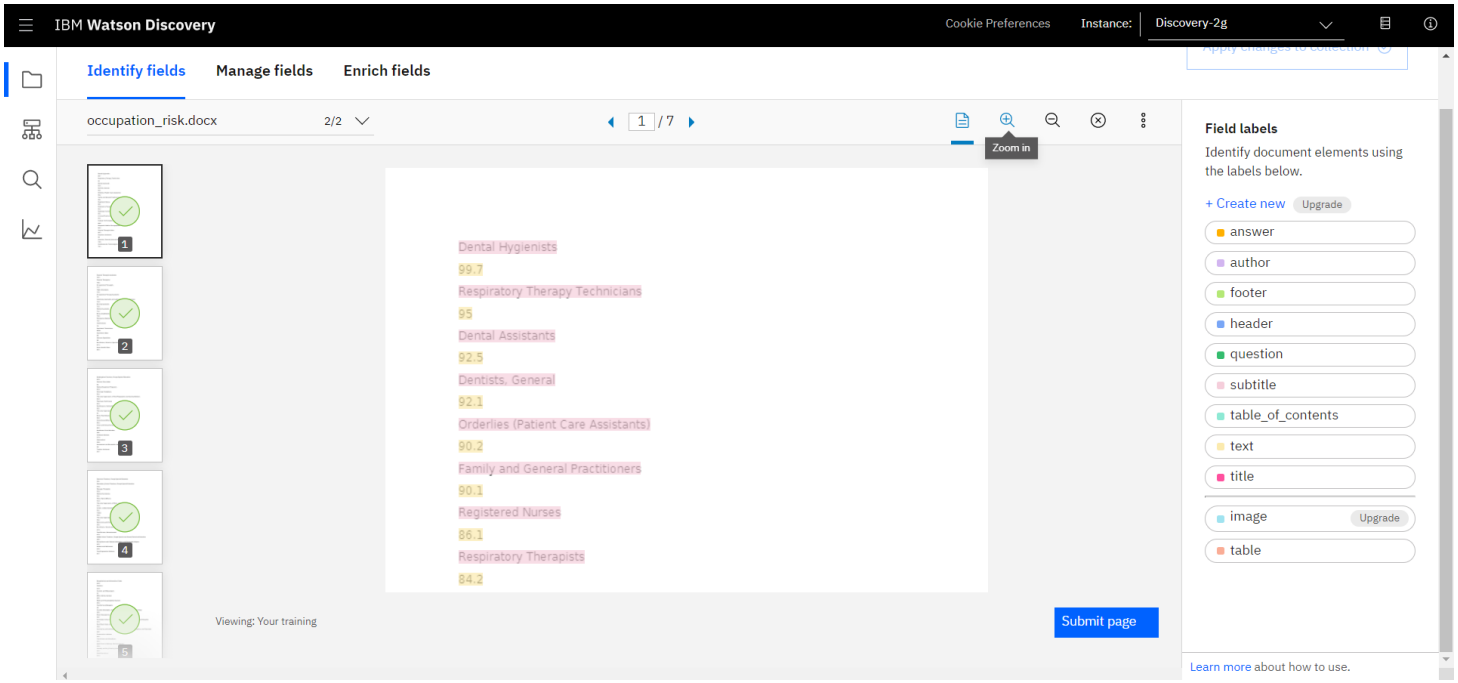
https://web-chat.global.assistant.watson.cloud.ibm.com/preview.html?region=eu-gb&integrationID=f61b8a02-9481-4273-8c67-36870979fb63&serviceInstanceID=75980f2e-aa89-4419-9885-3c9be0c8312b



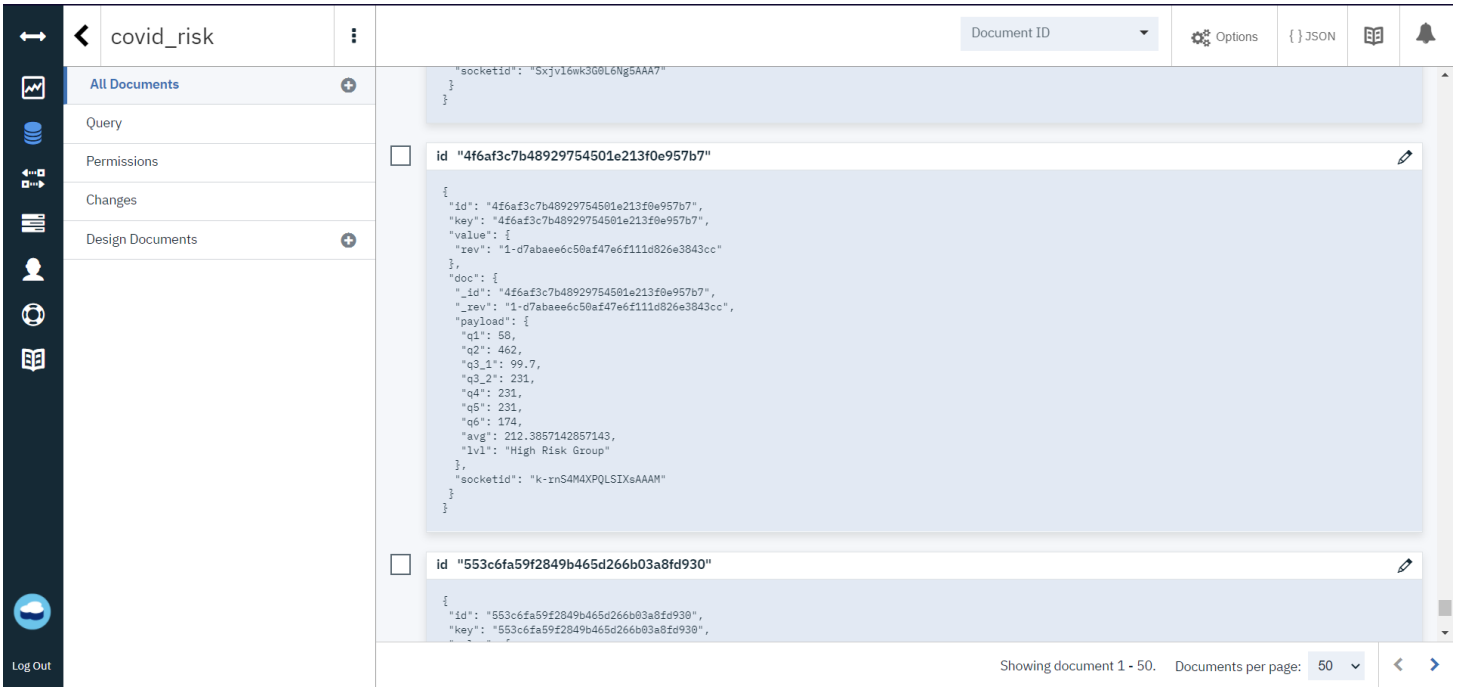


Watson Discovery

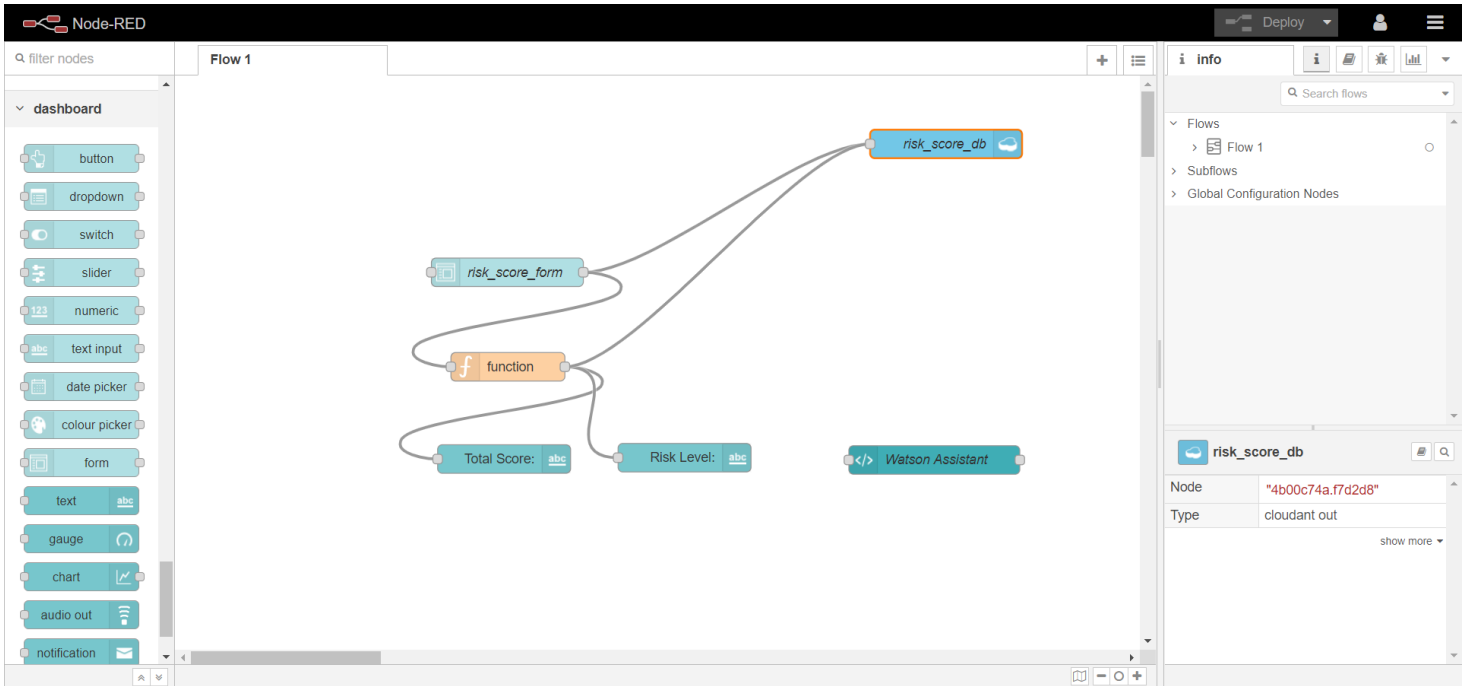
Due to limitations on the IBM Cloud Trial plan we were not able to include a large amount of website data. So, we were only able to use Covid hotspot locations of southern districts of Kerala (Thiruvananthapuram, Kollam, Pathanamthitta, Alappuzha, Kottayam and Idukki).

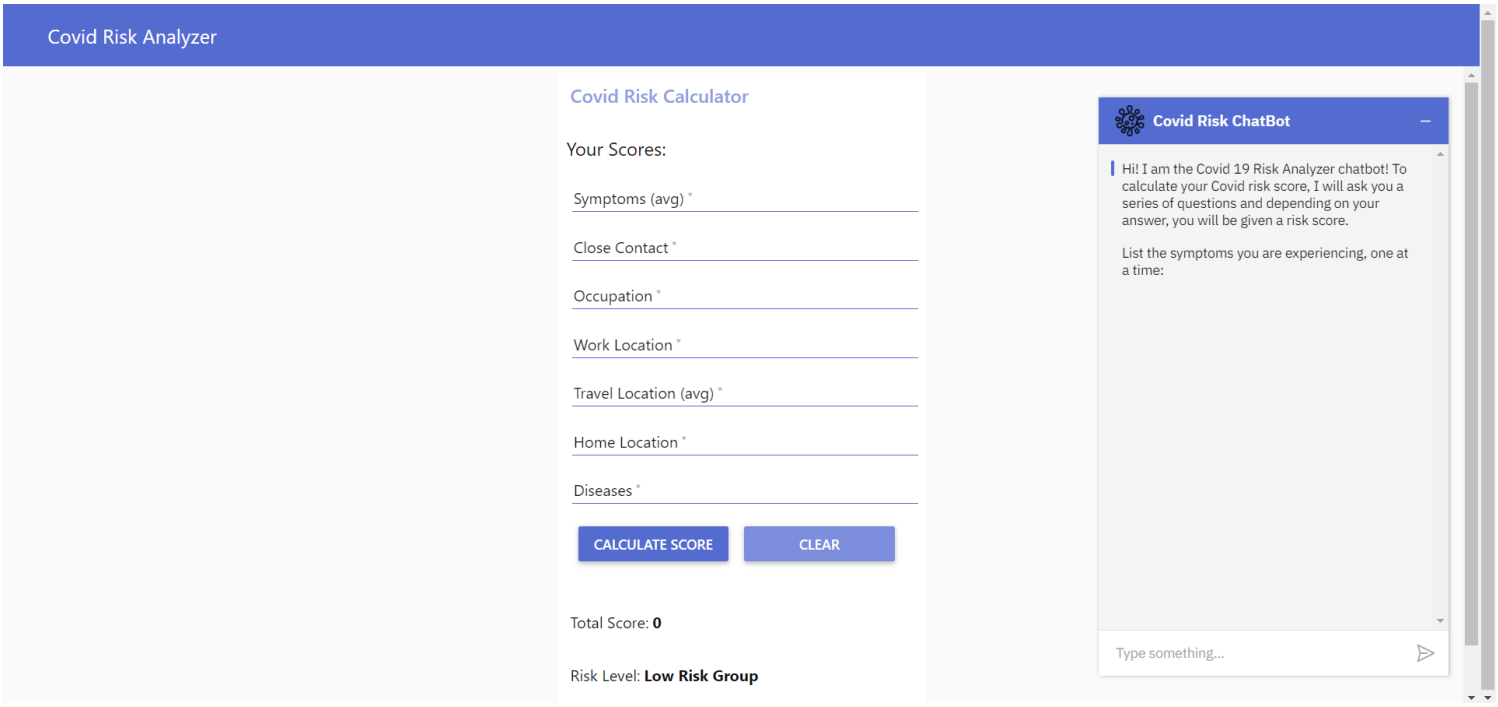


Cloudant Database



Node Red





Covid Risk Analyzer Project link

(Remember to use only locations of districts of Southern Kerala. Due to limitations of IBM Cloud Trial plan we could not outsource to locations around the entire country.)

https://covid-risk-analyzer-node.eu-gb.mybluemix.net/ui/#!/0?socketid=SnyG-fM8KekV8rAPAAAi

Demo Video

https://drive.google.com/file/d/1L8ykMIFcOgfQcIsnoDfz5gYqYDhNgo0y/view?usp=sharing

References

Location intents: https://data.gov.in/resources/complete-town-directory-indiastatedistrictsub-district-level-census-2011-kerala

Occupations & Risk: https://www.visualcapitalist.com/the-front-line-visualizing-the-occupations-with-the-highest-covid-19-risk/

Kerala Hotspots: https://dashboard.kerala.gov.in/hotspots.php

Prepared by Team T853

Cyril Shaji

Kuriachen Sunil

Chandichen Sunil

Meenakshi Sajeev