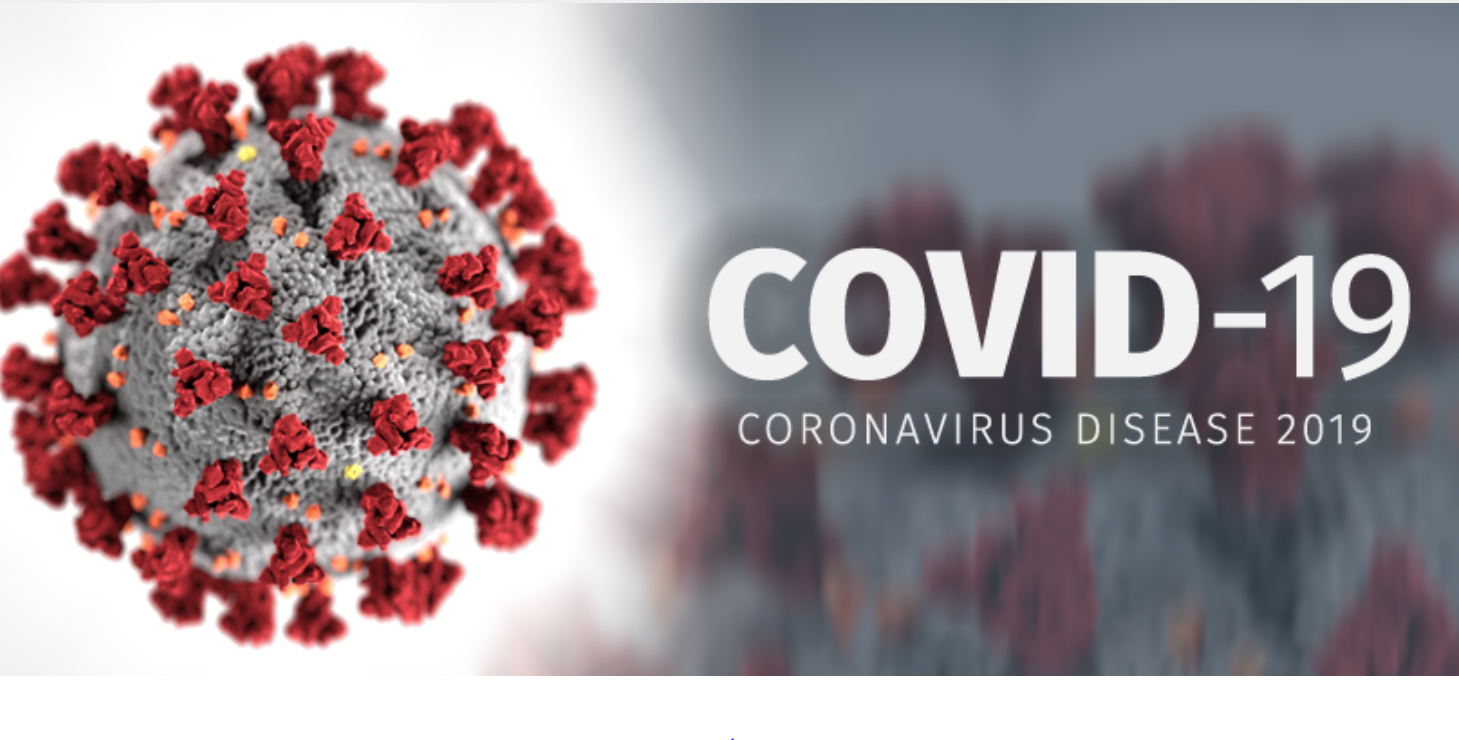
**COVID-19 Data Analysis and Prediction using Machine Learning**

By

Rohini



**Introduction**

In today's world, where the global community continues to combat the COVID-19 pandemic, the development of accurate and efficient tools for disease prediction has become increasingly crucial. In line with this need, our proposal aims to build a machine learning model to predict whether an individual is COVID-19 positive or not based on a dataset of symptoms. By leveraging the power of machine learning algorithms, we aim to enhance the accuracy and speed of COVID-19 diagnosis, thereby contributing to effective medical treatment and reducing the burden on healthcare systems.

**Importance in Today's World:**

The COVID-19 pandemic has overwhelmed healthcare systems worldwide, leading to a pressing need for efficient and reliable diagnostic methods. Traditional diagnostic approaches, such as PCR testing, often require time-consuming laboratory procedures and are limited by testing capacity. By developing a machine learning model for COVID-19 prediction, we offer a promising solution that can provide quick and accurate assessments based on symptoms alone, without the need for extensive laboratory testing. This can greatly improve the early identification of COVID-19 cases, enabling prompt medical intervention and appropriate allocation of healthcare resources.

**Improving Medical Treatment:**

Accurate disease prediction plays a crucial role in improving medical treatment outcomes. By accurately identifying individuals who are likely to be COVID-19 positive, our machine learning model can enable early intervention, leading to timely medical care and the implementation of necessary preventive measures. This early detection can help prevent disease progression, reduce the severity of symptoms, and improve overall patient outcomes.

**Impact on the Medical Field:**

The implementation of an effective screening tool based on our machine learning model can have a profound impact on the medical field. It can significantly reduce the burden on healthcare systems by streamlining the diagnostic process, particularly in areas with limited testing capacities. By efficiently identifying COVID-19 positive individuals, healthcare resources can be optimized, ensuring that critical cases receive immediate attention while preventing the unnecessary utilization of resources for low-risk cases. This targeted allocation of resources can help in managing the influx of patients, improving overall healthcare delivery, and potentially saving lives.

**Future Applications and Knowledge Gap:**

Beyond its immediate application in COVID-19 prediction, our proposed machine learning model can serve as a valuable framework for predicting other diseases based on symptoms. The knowledge gained from developing this model, such as feature selection techniques and algorithmic approaches, can be applied to future disease prediction efforts. This has the potential to fill knowledge gaps in various healthcare domains, allowing for faster and more accurate diagnosis of diseases beyond COVID-19. The ability to predict diseases accurately has far-reaching implications for personalized medicine, public health planning, and the development of targeted treatment strategies, ultimately leading to improved healthcare outcomes for individuals and communities alike.

**Initial Hypothesis**

Based on the dataset provided, which includes variables such as Test\_date, Cough\_symptoms, Fever, Sore\_throat, Shortness\_of\_breath, Headache, Corona, Age\_60\_above, Sex, and Known\_contact, we can form the following initial hypotheses:

Hypothesis 1: Patients who were in direct contact with Confirmed (Covid Positive) Patients are more likely to be Corona positive.

This hypothesis suggests that individuals who have had close contact with confirmed COVID-19 positive patients are at a higher risk of being infected themselves. The assumption is that the virus spreads primarily through close proximity and direct contact with infected individuals. By examining the variable "Known\_contact" in the dataset, we can investigate whether there is a correlation between known contact with COVID-19 positive patients and the likelihood of testing positive for the virus.

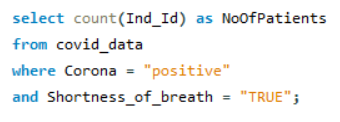
Hypothesis 2: Shortness\_of\_breath, Fever, and Cough\_symptoms are important factors in determining COVID-19 positive or negative cases.

This hypothesis proposes that symptoms such as shortness of breath, fever, and cough are key indicators in identifying COVID-19 positive cases. These symptoms are commonly associated with respiratory infections and have been identified as prominent symptoms of COVID-19. By analyzing the variables "Shortness\_of\_breath," "Fever," and "Cough\_symptoms" in the dataset, we can explore the relationship between these symptoms and the likelihood of a person being COVID-19 positive or negative.

It is important to note that these are initial hypotheses based on the provided dataset, and further analysis and modeling will be required to validate these assumptions and uncover additional insights related to COVID-19 prediction.

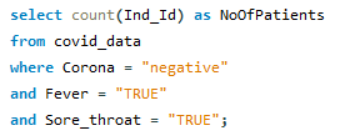
SQL

Q.1 Find the number of corona patients who faced shortness of breath.



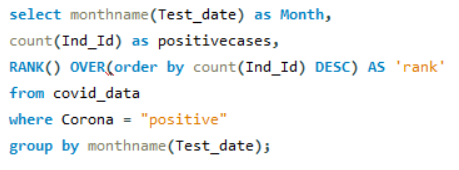
Conclusion: There are 1162 corona patients who faced shortness of breath.

Q.2 Find the number of negative corona patients who have fever and sore\_throat.



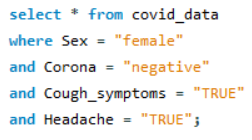
Conclusion: There are 121 corona negative patients who have fever and Sore throat.

Q.3 Group the data by month and rank the number of positive cases.



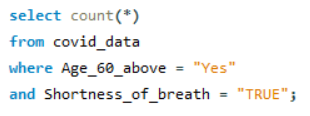
Conclusion: In April we found most number of corona positive cases(8863), And in march 5863 corona positive cases.

Q.4 Find the female negative corona patients who faced cough and headache.



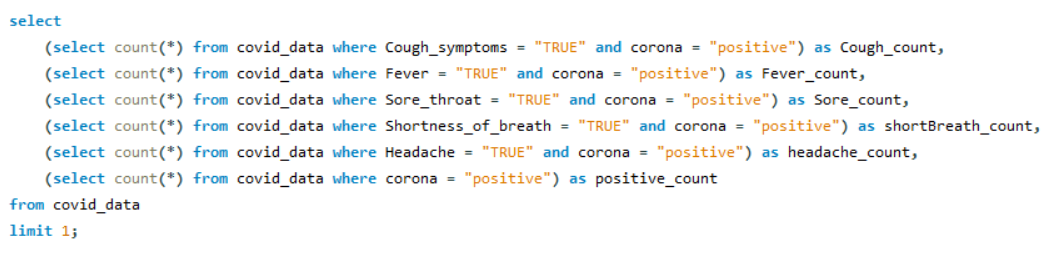
Conclusion: 32 female negative corona patients who faced cough and headache.

Q.5 How many elderly corona patients have faced breathing problems?



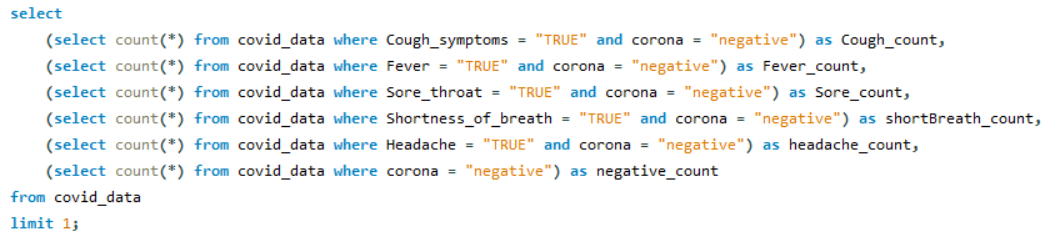
Conclusion: 286 elderly corona patients have faced breathing problems. (In problem it's not mentioned corona positive or negative patients)

Q.6 Which three symptoms were more common among COVID positive patients?



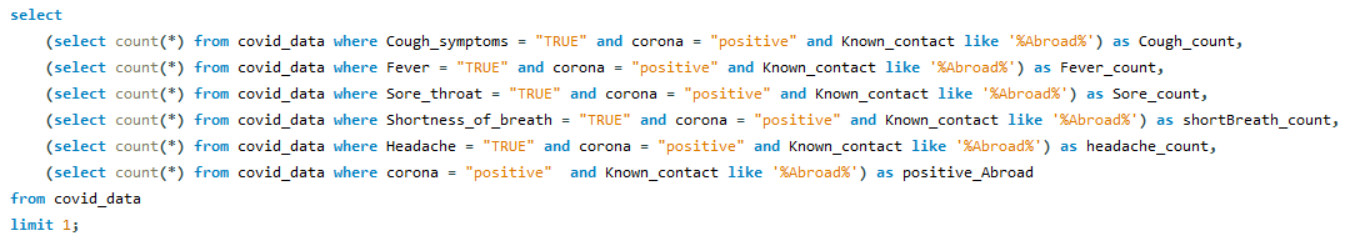
Conclusion: Cough\_Symptoms, Fever and Headache are the most common symptoms among Corona positive patients.

Q.7 Which symptom was less common among COVID negative people?



Conclusion: Headache, Shortness of Breath, Sore throat symptom was less common among COVID negative people.

Q.8 What are the most common symptoms among COVID positive males whose known contact was abroad?



Conclusion: Cough and Fever are the most common symptoms among COVID positive males whose known contact was abroad.