

Evaluating Machine Learning Rules : Project Group 10

The project focuses on creating awareness about the Corona Virus Disease(COVID-19) and understanding about the people affected and other attributes related to the disease with the help of data analysis. The analysis can help a wide range of audience such as government officials, who want to control the spread of these pandemic diseases. Major concerns answered would be the country with the majority of patients and the impact of COVID-19 (Death vs Confirm of patients). For this various machine learning rules are applied on the analysis to get the best results and precision.

Before Machine Learning Phase - First, design and implement metrics : Tracking and gathering as much as possible for the system beforehand so that the model has a scope to expand. The topic we are working on has unpredictability and uncertainty in the data and its behaviour. So it is better to get historical data beforehand for something that might be a concern in the future. By being more liberal about gathering metrics, we can gain a broader picture of the system. For any problems or changes in Covid data, add a metric to track it.

ML Phase first pipeline - Keep the first model simple and get the infrastructure right : Before the use of advanced machine learning models, it's really important to get examples of the affected data of coronavirus for learning algorithms and integrating the model into the application. Choosing simple features makes it easier to ensure that the features reach the learning algorithm correctly, model learns reasonable weights and the features reach the model in the server correctly. This is essential at first in our project which can be enhanced later as the project approaches and we find more data on COVID-19.

Feature Engineering ML phase - Explore with features of content that generalize across contexts : Once the end to end system is ready, we can start exploring it further with this approach. Often a machine learning system is a small part of a much bigger picture. Using the existing features such as coronavirus impacted sex ratio in the country, the age group and blood group and realting it all together would provide statistics to the learner, it can promote new insights that it has no data for in the context it is optimizing. We can also consider the weather conditions in different countries across the globe to estimate the survival ratio of the coronavirus affected people. All of these features allow us to bring new content into the context.

Slowed growth, optimization refinement and complex models phase - When performance plateaus, look for qualitatively new sources of information to add rather than refining existing signals : Even after adding the demographic information, going through template exploration, and tuning the regularization, if there isn't much improvement in the key metrics, it is time to start building the infrastructure for radically different features. So as the data keeps on changing for the coronavirus and its spread, it's uncertainty needs in refactoring of the architecture with the new changing attributes. Adding new features to the existing ones can be really productive and effective.