

TW3 Exploratory Data Analysis on COVID-19 Datasets

Names: Alfin Abraham, Karthik Nalla

GitHub: [\(Alfin\) TW3 GitHub link](#), [\(Karthik\) TW3 GitHub link](#)

Task 1: DS Methodology Case Study 2

1. Work on the use case: DataScienceMethodologyUseCase-2.pdf

A. Define a sub-problem in your own way, which satisfies the goal of the project.

Prediction of asthma risk using non-image data in the absence of high-quality data sets in related fields showing the effectiveness of transfer learning in this domain.

B. Discuss what data we will need to obtain for a solution to the problem.

To solve the problem, we will require data about indoor air quality, general health data of patient and air quality data of the region.

C. Discuss how the obtained data will be used in a framework (DS methodology).

In the DS methodology framework once we have collected the data it will be first explored then a rigorous dataset will be prepared for further modelling and analysis.

D. Have a rough sketch of your plan for data preprocessing.

First, we will clean the data by addressing missing data and any noisy data, then we will select relevant attributes and select a subset of the data if it is too large.

2. Suppose you have some datasets provided by a medical group (your client). You then have cleaned the data and normalized it.

a) Now you plan to conduct an EDA using descriptive statistics. Discuss your objectives from this analysis.

Our objective is to use descriptive statistics techniques to perform EDA on the dataset to learn about five number summary and determine trends that we can use for prediction.

b) Are these descriptive statistics enough to see data insights? Share your thoughts.

Making use of descriptive statistics alone to gain data insights will lead to uninformed decision making. Outliers can be missed. Therefore, performing data visualization will help in gaining insights that statistical methods simply cannot convey.

Task 2 Part 2:

- *A summary of what you learned from the teamwork assignment.*

Upon completion of TW3 we learned the importance of EDA in the data science methodology, and we learned that in order to solve any problem in data science we must approach it from every corner and use all kinds of EDA methods and tools so that we gain proper insights as this will help us with the modelling stage.