# Readme

The project has many components and software used. To run this project, there are some libraries and components must use.

## Hadoop

* Go to website <https://hadoop.apache.org/> and click on download button.
* Click on Binary download. The version used is **version 3.3.6**.

## Apache Spark

* Go to website <https://spark.apache.org/downloads.html> .
* Select Spark released. The version used is **version 3.3.2**.
* Select package type **Pre-built for Apache Hadoop 3.3 and later (Scala 2.13)**.

## Tools and Software

* Tool used for visualization is Tableau. Download Link: <https://www.tableau.com/>
* XAMPP is used for the Apache server and MySQL. Download Link: <https://www.apachefriends.org/>
* The IDE used in this project is **Visual Studio Code**. Download Link: <https://code.visualstudio.com/>

## Front-End Languages

* HTML, CSS, JavaScript

## Back-End Languages

* Python

## Frameworks

* Bootstrap
* Flask

## Libraries

* Flask-Login
* SQLAlchemy
* Pandas
* Scikit-learn
* Scikit-learn – Regression
* Scikit-learn – Mean Squared Error
* Matplotlib
* PySpark

## Links:

* Dashboard: <https://public.tableau.com/app/profile/abdul.samad8639/viz/InfluenzaDashboardFinalBuild_version03/Influenza_Dashboard>
* Data set: <https://www.kaggle.com/datasets/cdc/state-specific-influenza-vaccination-coverage?select=state-specific-influenza-vaccination-coverage-among-women-with-live-birth-prams-2009-10-influenza-season.csv>
* Blog Link: <https://demo1.nrcodex.com/2023/08/14/viroshield/>
* Github: <https://github.com/Faisalkhan12676/msg-clusters>

## Database Name:

Influenza\_db

## Assumptions:

The project is assuming that the user have **Hadoop** and **Apache Spark** configuration in built. The libraries mentioned above are also prerequisite for **prediction**.

To run project run this command in terminal or command line

flask run