DS 4004:Big Data Analytics
Group Project – Proposal

Group Details	Group 03: Naethree Premnath - 15552, Zafra Farhan- 15521, Dilmi Abeytunga - 15509

Title of the Project: Global Leptospirosis Surveillance and Forecasting Dashboard

Introduction: Leptospirosis is a zoonosis most commonly found in the tropics caused by the spirochetes of the genus Leptospira.[1] It is a potentially serious bacterial zoonotic disease affecting many parts of the body.

The global annual incidence of leptospirosis is estimated at around 1.03 million cases, resulting in about 58,900 deaths per year.[2] Most cases and fatalities occur in tropical regions between the Tropic of Cancer and the Tropic of Capricorn.

Monitoring and forecasting leptospirosis cases and deaths is crucial for public health management. This project aims to create a comprehensive database and an interactive dashboard to track leptospirosis cases and deaths globally, provide real-time updates, and forecast future trends.

Objectives:

1. Data Collection: Scrape, find, and copy leptospirosis data from all available sources.
2. Real-Time Data Integration: Connect with available APIs or scrape websites to obtain real-time data and store it in the database.
3. Forecasting: Implement forecasting models to predict leptospirosis trends at least one month into the future.
4. Visualization: Create a user-friendly dashboard to visualize current data and forecasts.

Data:

1. The Surveillance Atlas of Infectious Diseases: This resource provides a comprehensive list of reported leptospirosis cases in different regions from 2007 to 2022.
   Link: https://atlas.ecdc.europa.eu/public/index.aspx?Dataset=27&HealthTopic=31
2. Annual Epidemiological Report for 2022: Leptospirosis reported cases in EU/EEA countries (contains data from 2018-2022).
   Link: https://www.ecdc.europa.eu/en/publications-data/leptospirosis-annual-epidemiological-report-2022
3. Centers for Disease Control and Prevention Nationally Notifiable Infectious Diseases and Conditions, United States - Annual Tables: Includes annual cases of selected infectious national notifiable diseases from the National Notifiable Diseases Surveillance System (NNDSS). NNDSS data reported by the 50

states, New York City, the District of Columbia, and the U.S. territories are collated and published. Specifically, Table 2i contains data on leptospirosis. Data from 2016-2021 can be filtered and used.
Link: https://wonder.cdc.gov/nndss/nndss_annual_tables_menu.asp

4. Weekly Epidemiological Reports - Sri Lanka: The last report for each year includes cumulative cases of leptospirosis reported within Sri Lanka. Data can be extracted from column B under Leptospirosis in the table of these reports.
Link: https://www.epid.gov.lk/weekly-epidemiological-report/weekly-epidemiological-report

Methodology:

Data Collection:

1. Website filters: Configuring filters on websites(E.g. Reported Cases and/or No.of Deaths, Year) to specifically target and extract the necessary information related to leptospirosis. (For Dataset 1)
2. Scrape No.of reported cases of leptospirosis data from the tables. (For Datasets 2, 3 and 4)

Database Creation:

1. Create a database to store Year, Region, Reported Cases and No.of Deaths.
2. For records with missing values (E.g.: missing No.of deaths), NA will be added.

Development of Dashboard:

1. Create a dashboard with the following features:

- Global map showing leptospirosis cases and deaths.
- Time-series graphs for trend analysis.
- Filtering options by location and year.

Forecasting:

1. Use time series modeling techniques to forecast future confirmed cases of leptospirosis based on the collected data. This will provide insights into potential outbreaks in the future.

References:

[1] Epidemiology Unit. (2020). *Epidemiology Unit, Ministry of Health, Sri Lanka: Weekly epidemiological report, Vol. 49, No. 22*. Retrieved from https://www.epid.gov.lk/epid/public/storage/post/pdfs/vol_49_no_22-english_1.pdf

[2] Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., Stein, C., Abela-Ridder, B., & Ko, A. I. (2015). Global morbidity and mortality of leptospirosis: A systematic review. *PLoS Neglected Tropical Diseases, 9*(9), e0003898. https://doi.org/10.1371/journal.pntd.0003898