Epidemiology Model Inference and Machine Learning

Mahi Agrawal and Dr. Peter Beerli

Introduction

Epidemiology is the study of all the aspects of the presence or absence of diseases and disorders. It plays an important role in the prognosis of epidemics of diseases.

The SIS model is a mathematical model used in epidemiology to describe how some infectious diseases are spread. The SIS (Susceptible-Infected-Susceptible) model simulates the spread of diseases in which an individual can become infected, recover, and then become susceptible to infection again. These are diseases which do not provide immunity after recovery from infection, such as the common cold and influenza.

Figure 1 shows a simple diagram of the SIS model. All individuals in a population are considered susceptible initially, as they can become infected.

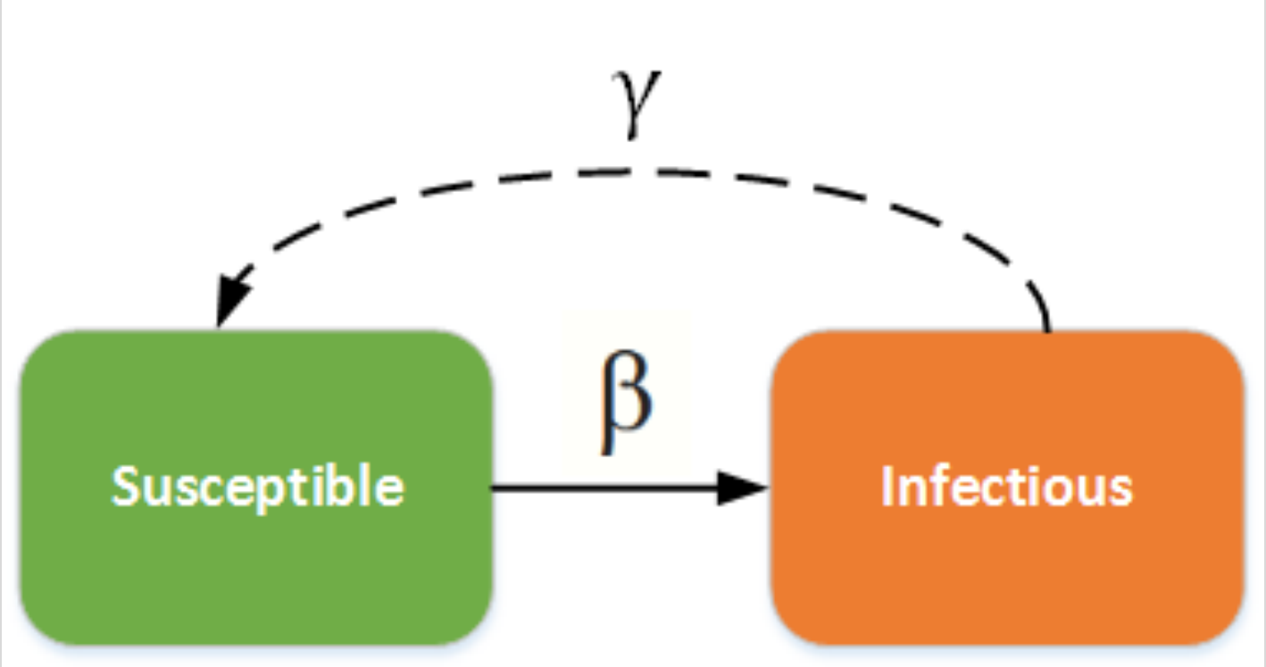

Figure 1: SIS model

The solid line shows the beta value (β), also known as the infectious rate. The dotted line shows the individuals who recover, and thus become susceptible again. This is the gamma value (γ), also known as the recovery rate.

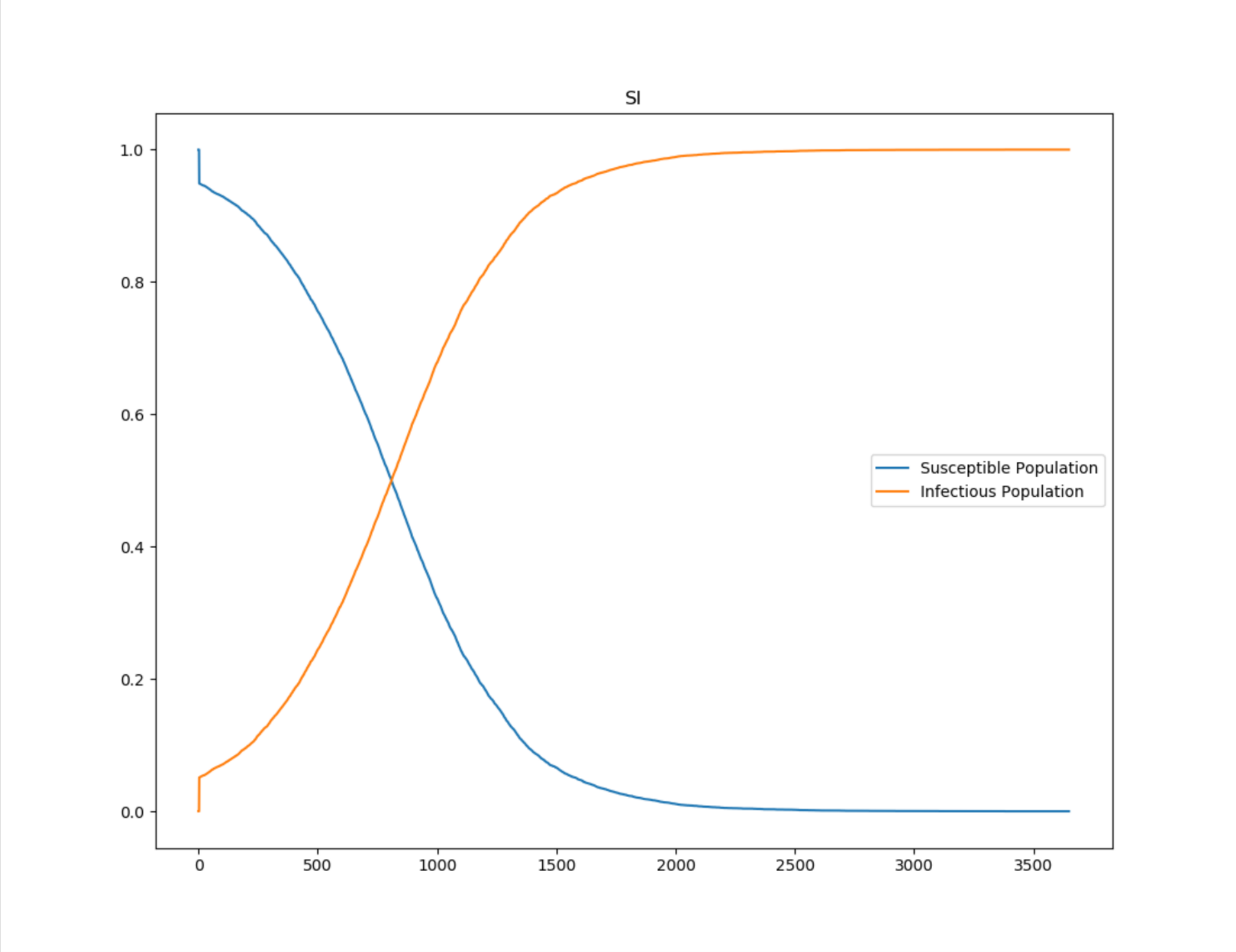

Figure 2: Correlation between the susceptible and infectious population

Figure 2 shows the relationship between susceptible and infectious populations. The rate of infection is proportional to the population size while the rate of recovery is proportional to the size of the infected group.

Methods

We have developed an SIS model using a stochastic approach. Using this model, we have trained a neural network using PyTorch to estimate the infection rate (β) and the recovery rate (γ) based on the number of generations , the sample size, and the number of infected individuals at two time points, 25% of the total generations (t1) and 75% of the total generations (t2).

This stochastic model is tested against an SIS ODE which verifies that this machine-learning approach accurately infers SIS model parameters. The following are the differential equations for the SIS model:

$$\frac{dS}{dt} = -\beta SI + \gamma I$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

Figure 3 shows a graph comparing the results of our stochastic model to an ODE based model. It uses a population of 100, initial infected of 10, β of 0.5, γ of 0.2, and 200 generations.

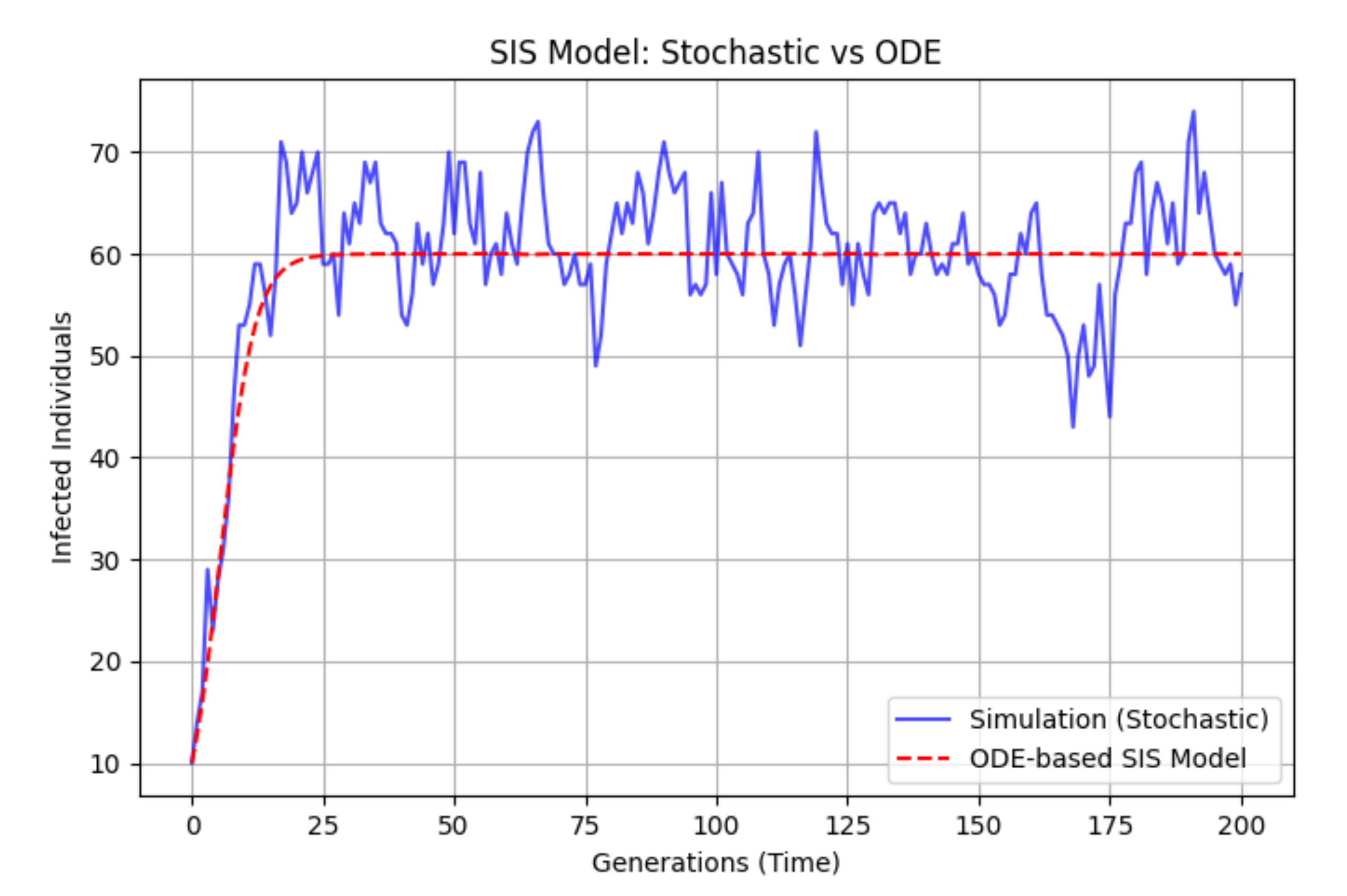

Figure 3: Graph comparing stochastic model results to ODE model results

Results

This model can be used to forecast infection rates for future, real-world diseases. The infection and recovery rates from this model can then be used to calculate the R_0 (R-naught) value, which indicates how contagious a disease is.

$R_0 < 1$: the disease will die out
$R_0 = 1$: will stay alive; no epidemic or outbreak
$R_0 > 1$: may be an epidemic or outbreak

The R_0 value can be calculated using the following formula:

$$R_0 = \frac{\beta}{\gamma}$$

We tested this by using our model to predict the R_0 value of a disease with 200 generations, a sample size of 50, t1 = 50 infected, and t2 = 45 infected. The model returned a predicted β of 0.5997 and a predicted γ of 0.3929.

$$R_0 = \frac{\beta}{\gamma} = \frac{0.5997}{0.3929} = 1.526$$

Based on the calculated R_0 value, we know that this disease will stay alive and may cause an outbreak or epidemic.

Such forecasting can help improve epidemic preparedness and response.

References

Kuhl, E. (1970, January 1). The classical SIS model. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-030-82890-5_2

Ramirez, V. B. (2023, June 14). What is R0? Gauging Contagious Infections. Healthline. https://www.healthline.com/health/r-naught-reproduction-number

SI and SIS models. SI and SIS models - Malaria Model documentation. (n.d.). https://docs.idmod.org/projects/emod-malaria/en/2.20_a/model-si.html

U.S. Department of Health and Human Services. (n.d.). What is Epidemiology?. National Institute of Deafness and Other Communication Disorders. https://www.nidcd.nih.gov/health/statistics/what-epidemiology