Data-driven Identification of Compartmental Model of Covid-19

Aditya Firman Ihsan

Telkom University

Introduction

- Prediction of Covid-19 transmission dynamics is important.
- Some existing predictive models treat the number of infections as a pure time series data.
- More deterministic approach such as dynamical system may be needed.
- Many dynamical systems are formulated by assumption

Methodology: Compartmental Model

- Compartmental model is a dynamical system describing a population in terms of labeled compartments.
- Defining a compartmental model decides the dynamics of corresponding epidemy.
- Usual labels used in Epidemiology are Susceptible (S), Infected (I), and Recovery (R), with following dynamics

$$\frac{dS}{dt} = -\alpha IS, \qquad \frac{dI}{dt} = \alpha IS - \beta I, \qquad \frac{dR}{dt} = \beta I.$$

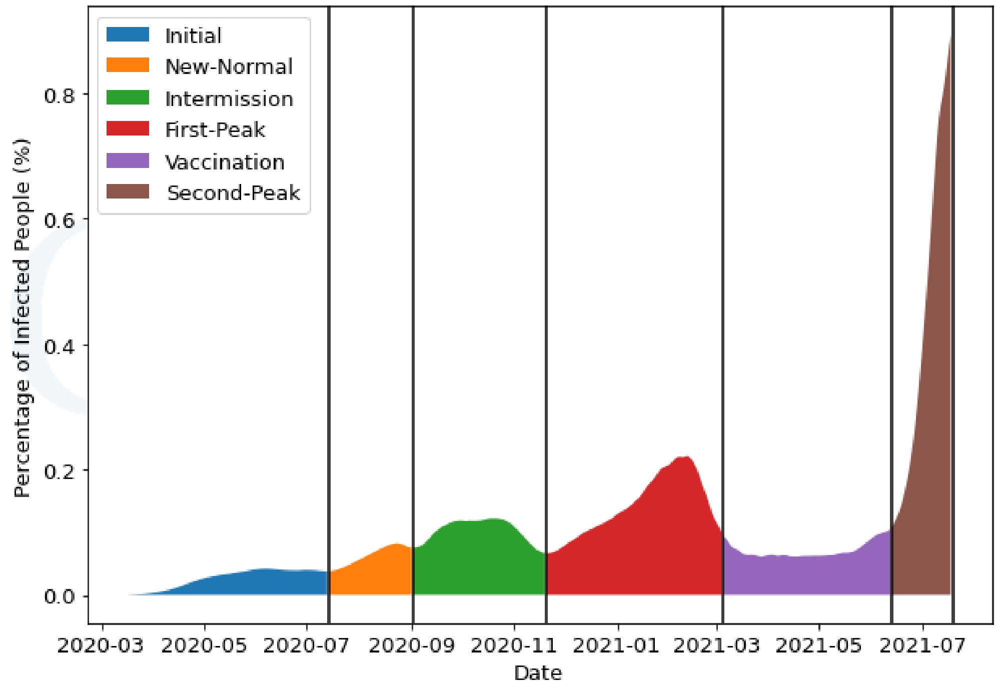
- In this paper, SIRVD model is used. V is for vaccination and D is for deceased.

Methodology: SINDy

- Sparse Identification of Nonlinear Dynamics (SINDy) identifies a dynamical system by data.
- SINDy uses LASSO regression to obtain a set of sparse parameters defining a correlation of variables (and its higher order interactions) with their derivatives.
- With enough relevant data, a nonlinear dynamical system governing a phenomena can be predicted.

Methodology: Data used

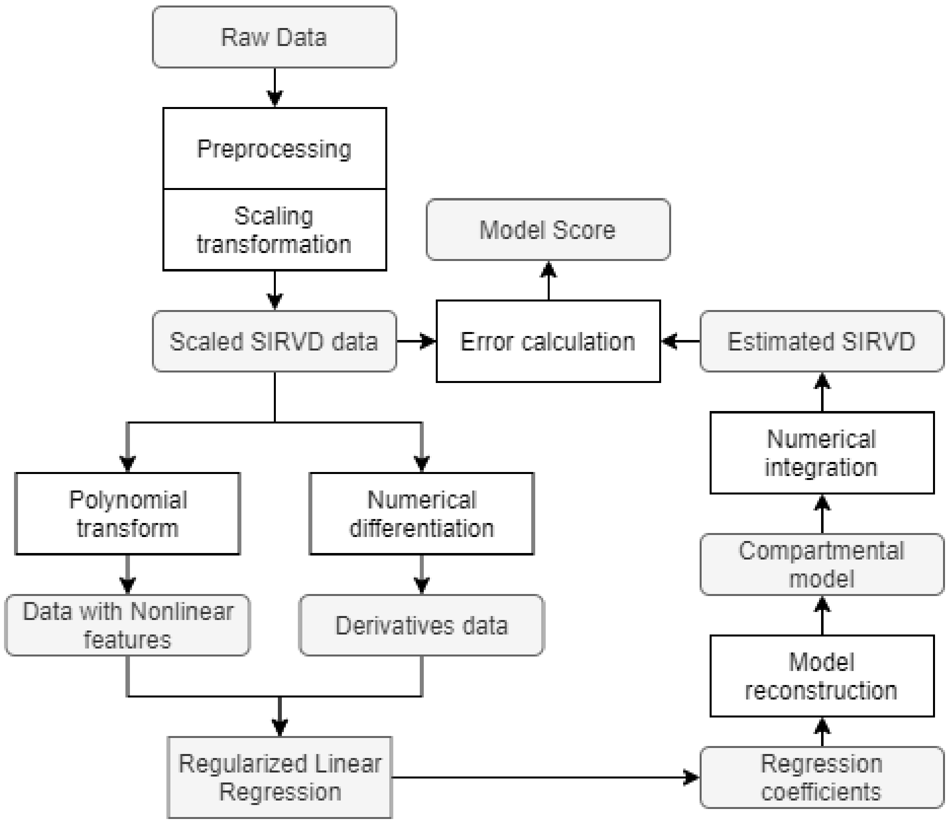
- Covid-19 Data in Jakarta from March 3rd 2020, to July 19th 2021 (489 datapoints).
- Divided to 6 phases.
- Scaled with respect to total Jakarta population



Methodology: System Design

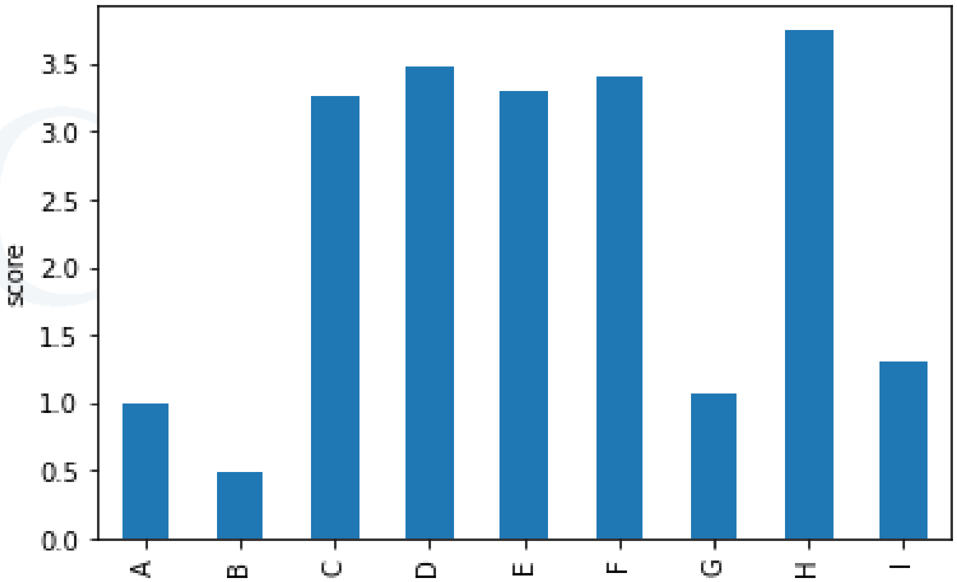
- Second order polynomial
- l_1 and l_2 regularization
- Validation score:

$1 - \log(MSE(\boldsymbol{x}_{pred}, \boldsymbol{x}_{true}))$

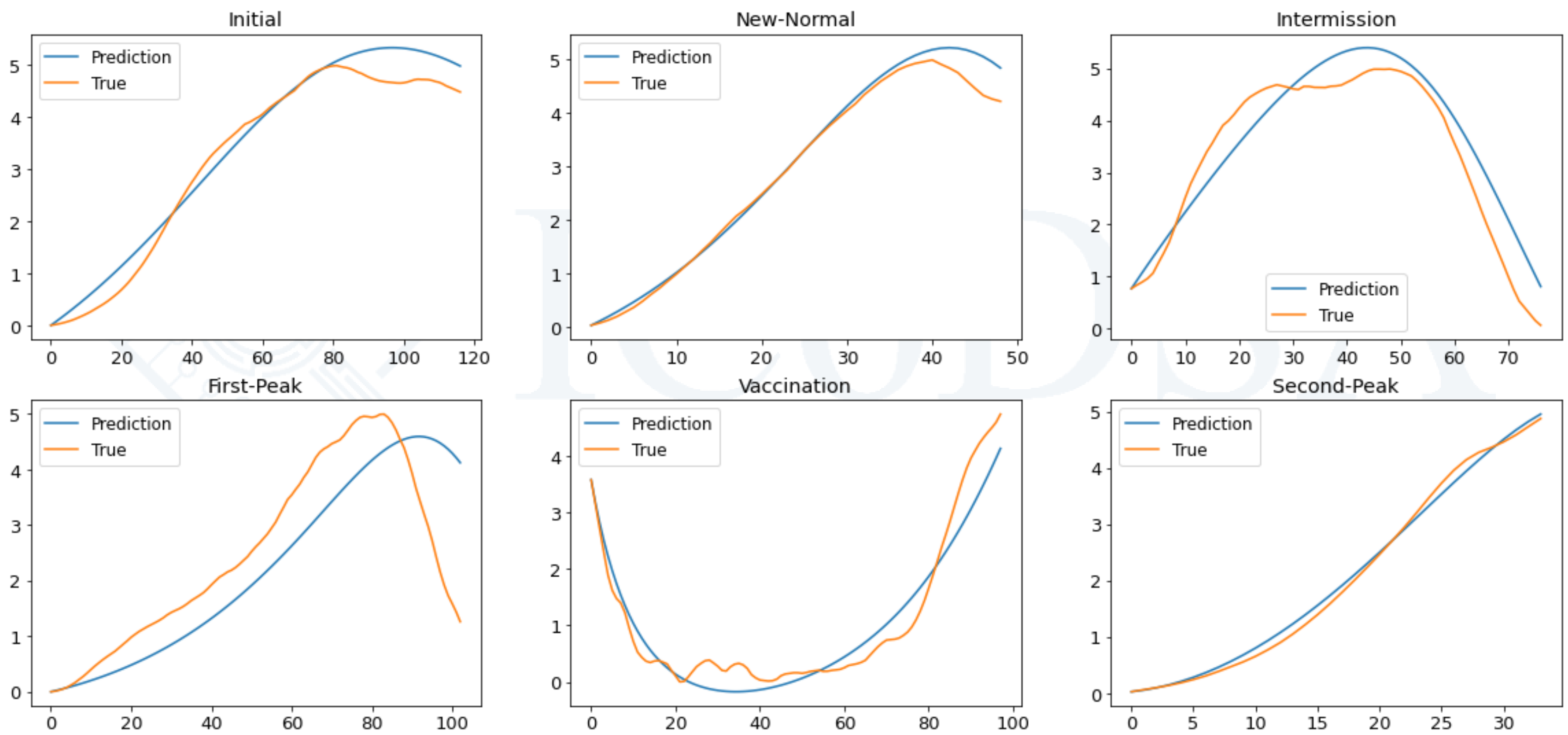


Results: Model selection

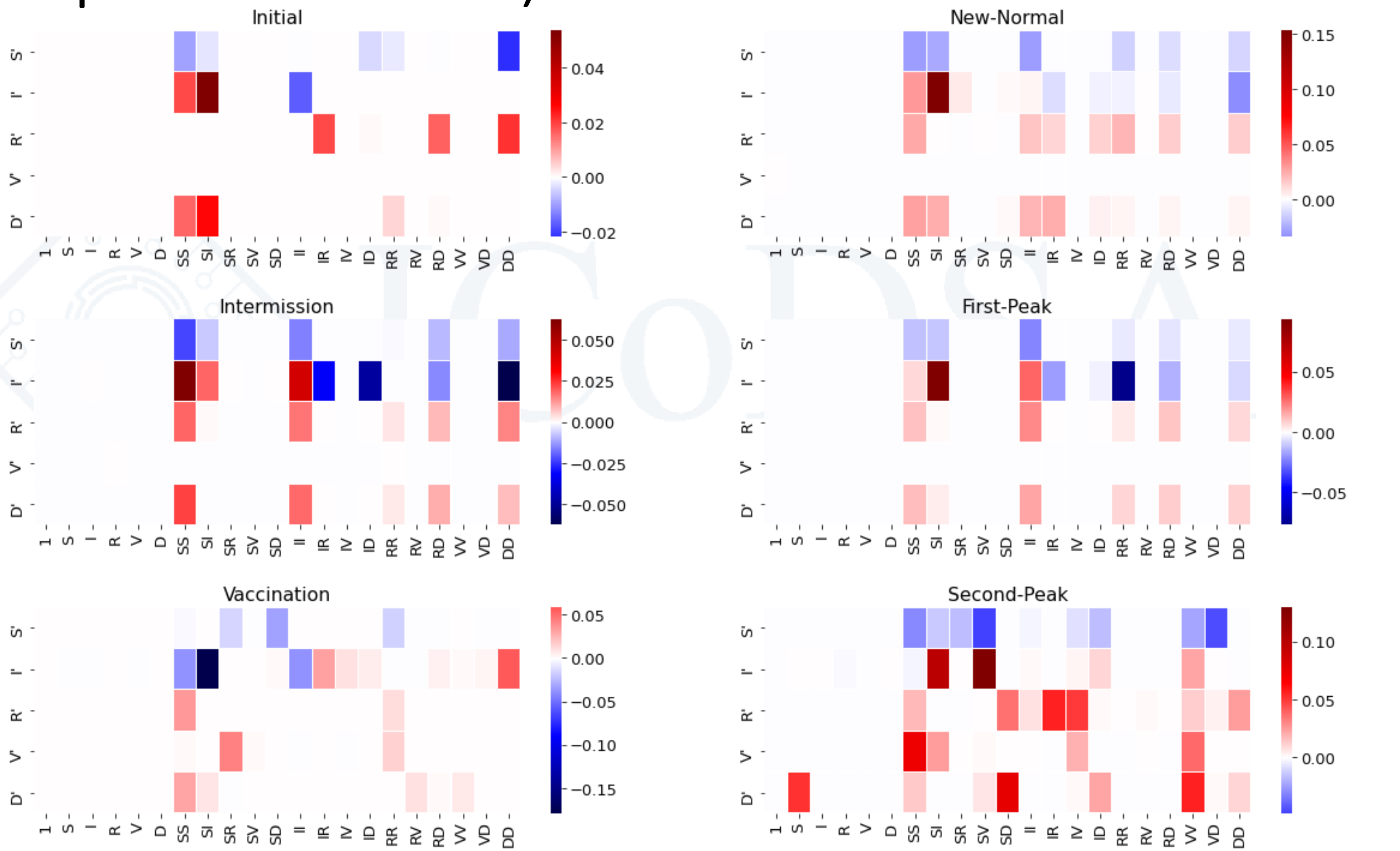
Case	Regularization Coef.		Input Features	
	l_1	l_2	Use bias	Power terms
A	0.01	0	No	Yes
B	0.01	0.01	No	Yes
C	0.01	0.1	No	Yes
D	0.1	0	No	Yes
E	0.1	0.01	No	Yes
F	0.1	0.1	No	Yes
G	0.1	0.1	No	No
H	0.1	0.1	Yes	No
I	0.1	0.1	Yes	Yes



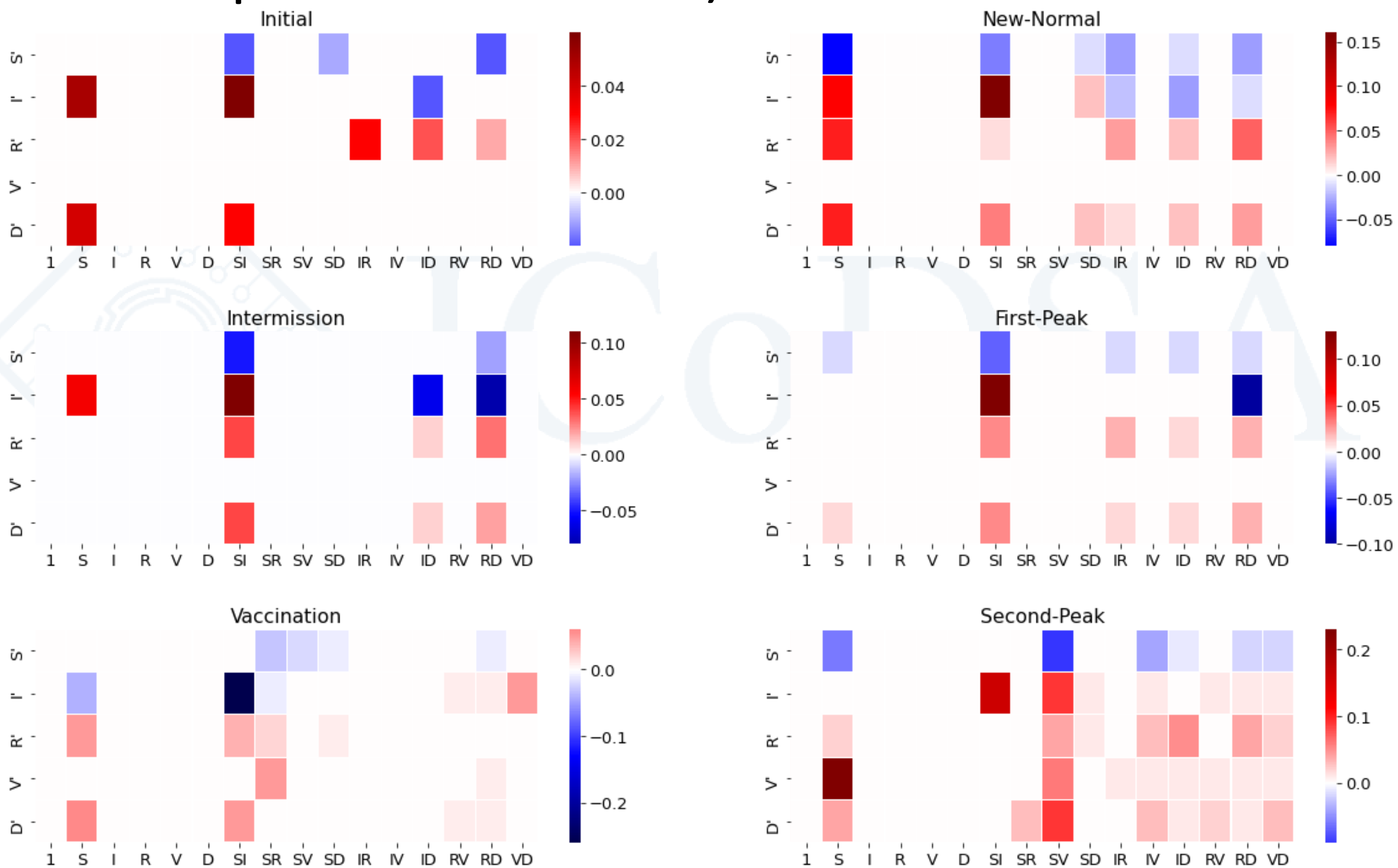
Results: Prediction



Results: Identified Parameters (with power terms)



Results: Identified Parameters (without power terms)



Conclusion

- SIRVD model of Covid-19 in Jakarta has been constructed based on data
- Both l_1 and l_2 regularization is important in model selection
- Powered terms add more complexity to the model
- Resulted prediction is smoother because based on dynamics equation
- Parameters obtained have low interpretability.
- Too many external factors affecting the system (it is not closed)