Data Science untuk Pemodelan Data Covid-19

Prof. Dr.rer.nat. Dedi Rosadi, S.Si., M.Sc.

Laboratorium Komputasi Statistika
FMIPA UGM
http://dedirosadi.staff.ugm.ac.id
E-mail: dedirosadi@gadjahmada.edu

Statistics R&D
Biro Penelitian, Pelatihan dan Konsultasi Statistika

STATISTICS R&D

Disclaimer!!!

STATISTICS R&D

- **Seluruh hasil penelitian yang dipresentasikan pada webinar ini dimaksudkan untuk kepentingan riset dan pendidikan**
- **Penggunaan untuk pengambilan keputusan atau hal-hal lain diluar kepentingan riset dan pendidikan harus dengan seizin peneliti dan bukan menjadi tanggung jawab peneliti**

Outline

- Background
- Review: Data Science for Covid -19
- Kajian aplikasi

PART I. BACKGROUND

STATISTICS R&D

IS IT STATISTICAL OR MACHINE LEARNING, ARTIFICIAL INTELLEGENCE AND/OR DATA SCIENCE REALLY USELESS????

Epidemic Modeling 101: Or why your CoVID-19 exponential fits are wrong

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

DATA SCIENCE AND BIG DATA IS USELESS FOR HEALTH AND EPIDEMICS, IS IT??

Stieb *et al. BMC Public Health* (2017) 17:372
DOI 10.1186/s12889-017-4286-8

BMC Public Health

EDITORIAL Open Access

Promise and pitfalls in the application of big data to occupational and environmental health

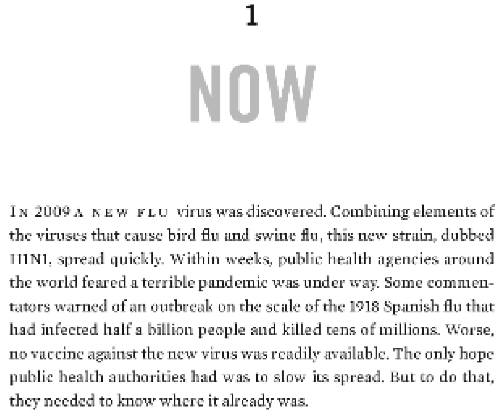
David M. Stieb[1,2*], Cécile R. Boot[3] and Michelle C. Turner[4,5,6,7]

Pitfalls

As a counterpoint to the potential of big data, one of the primary concerns is the potential for spurious findings, (described at their worst as "fanciful rubbish" or "big error") that can be generated by employing "much bigger and messier data" [2, 7]. Related to these limitations of big data

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

DIFFERENT DIRECTION ...
CHANGE MY PERSPECTIVE!



1

NOW

IN 2009 A NEW FLU virus was discovered. Combining elements of the viruses that cause bird flu and swine flu, this new strain, dubbed H1N1, spread quickly. Within weeks, public health agencies around the world feared a terrible pandemic was under way. Some commentators warned of an outbreak on the scale of the 1918 Spanish flu that had infected half a billion people and killed tens of millions. Worse, no vaccine against the new virus was readily available. The only hope public health authorities had was to slow its spread. But to do that, they needed to know where it already was.

PART II: REVIEW
DATA SCIENCE FOR COVID -19

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

APLIKASI AI UNTUK HEALTH CARE SECARA UMUM

STATISTICS R&D

REVIEW

THE AMERICAN JOURNAL of MEDICINE®

Artificial Intelligence Transforms the Future of Health Care

CrossMark

Nariman Noorbakhsh-Sabet, MD,[a,b] Ramin Zand, MD, MPH,[c,b,d] Yanfei Zhang, PhD,[e] Vida Abedi, PhD[f,d]

[a]St. Jude Children's Research Hospital, Memphis, Tenn; [b]University of Tennessee Health Science Center, Memphis; [c]Geisinger Neuroscience Institute, Geisinger Health System, Danville, Penn; [d]Biocomplexity Institute, Virginia Tech, Blacksburg, Va; [e]Genomic Medicine Institute, Geisinger Health System, Danville,Penn; [f]Biomedical and Translational Informatics Institute, Geisinger Health System, Danville, Penn.

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

STATISTICS R&D

Table 1 Main Machine-Learning Strategies: Their Characteristics, Scope, and Limitations

ML types	Algorithms Description	Characteristics	Limitation
Supervised Learning	Labeled data set System trained with human feedback	Applications include classification, regression, and prediction; ideal for modeling disease prognosis or treatment outcome. Modeling algorithms include Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF)	Requires a large amount of labeled data for training; need validation in an independent cohort.
Unsupervised Learning	Non-labeled data by humans	Applications include mainly pattern recognition; ideal for modeling disease mechanisms, identifying hidden patterns in genotype or phenotype data. Modeling algorithms include various clustering methods	Needs validation in several independent cohorts
Reinforcement Learning	Hybrid approach; the goal is to maximize accuracy by trial and error; especially useful in a complex environment	Applications include chemistry, robotics, games, resource management in computer clusters, personalized recommendations	Memory intensive

ML = machine learning.

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

STATISTICS R&D

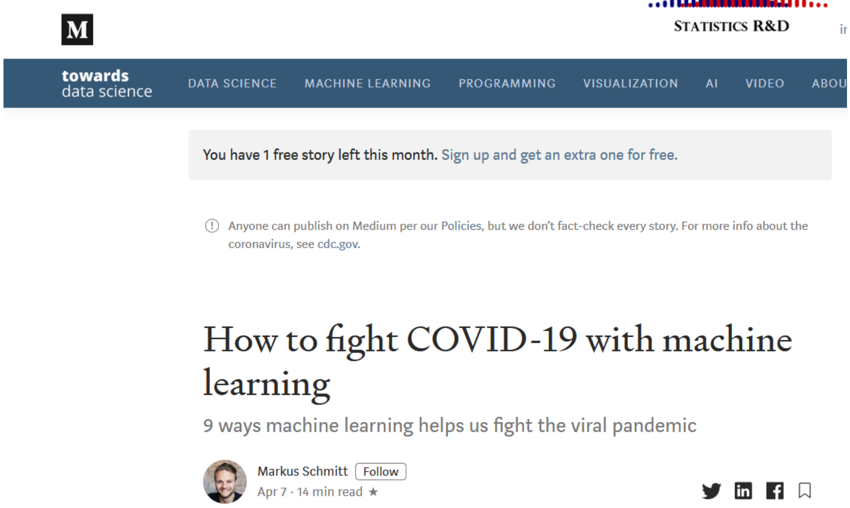
Table 2 Selected Areas in Medicine Where Machine Learning Has High Potential and Implications

Field	Application
Clinical	Disease prediction and diagnosis
	Treatment effectiveness and outcome prediction
Translation	Drug discovery and repurposing
	(*In Silico*) Clinical trial
Public health	Epidemic outbreak prediction
	Precision health

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

AND MANY OTHERS ….

STATISTICS R&D



Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

STATISTICS R&D

We want to shine a light on their work and show how machine learning is helping us to:

- Identify who is most at risk,
- Diagnose patients,
- Develop drugs faster,
- Predict the spread of the disease,
- Understand viruses better,
- Map where viruses come from, and
- Predict the next pandemic.

DATA SCIENCE FOR COVID-19

STATISTICS R&D

Leveraging Data Science To Combat COVID-19: A Comprehensive Review

Siddique Latif[1,2], Muhammad Usman[3,4], Sanaullah Manzoor[5], Waleed Iqbal[6], Junaid Qadir[5], Gareth Tyson[6,11], Ignacio Castro[6], Adeel Razi[7,8], Maged N. Kamel Boulos[9], Adrian Weller[10,11], and Jon Crowcroft[10,11]

Example of Application

STATISTICS R&D

- Risk Assessment and Patient Prioritisation
  - data-driven risk assessment and management in individual COVID-19 patients
  - provide an estimate of mortality risk based on traits like like age, gender, or health state. This is particularly important when resources are limited, e.g., for patient prioritisation when Intensive Care Unit (ICU) resources are insufficient.

Example of Application

STATISTICS R&D

- Screening and Diagnosis
  - COVID-19 Sounds is a mobile app collecting audio of breathing symptoms to help perform diagnosis.
  - Automated tools can also be developed to facilitate screening in larger groups of people (e.g., at airports), e.g., using computer vision based thermal imaging to detect fever
  - **Screening based on CTScan/ Xray image**
  - Early detection menggunakan analisis nafas pasien
    - Riset dr Departemen Fisika FMIPA UGM

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

Example of Application

STATISTICS R&D

- Simulation and Modelling
  - Using (Bayesian) hierarchical modelling, one can combine several of these (epidemic) models to create a (pandemic) model of viral spread among regions -> space time model
  - **Predictive class of model for forecasting**
  - Simulation of the effect of
    - decisions that affect disease transmission—e.g., decisions related to quarantine and social distancing strategies;
    - decisions regarding resource management—e.g., decisions related to capacity of in-patient hospital beds, critical care units, staffing, and resource allocation within and across regions;
    - decisions about care—e.g., deciding thresholds for admission and discharge of patients and minimising the impact on other patients.

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

Example of Application

STATISTICS R&D

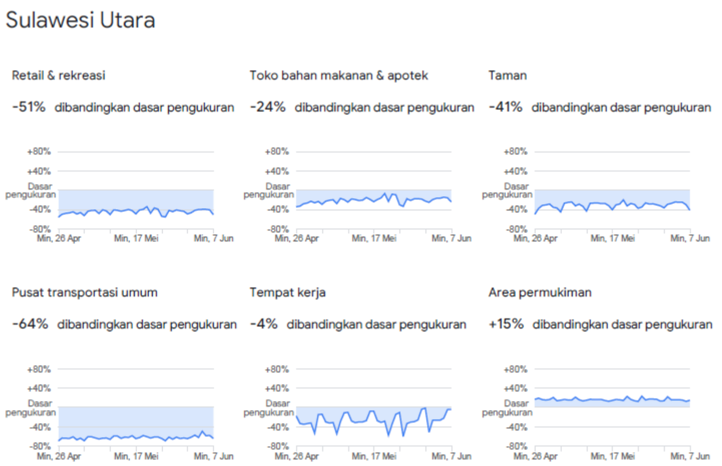
- Contact Tracing
  - Smartphone contact sensing, online surveys and automated diagnosis
- Understanding Social Interventions
  - **Understand human mobility patterns in the context of social distancing**
  - Controlling Misinformation & Online Harms
- Logistical Planning and Economic Interventions
  - applying machine learning to logistical planning, e.g., by Amazon Fulfilment
  - companies can use data science to detect unusual patterns of behaviour in the market or in their own customer base.
- Etc

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

Examples: Visualization of Human Mobility

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

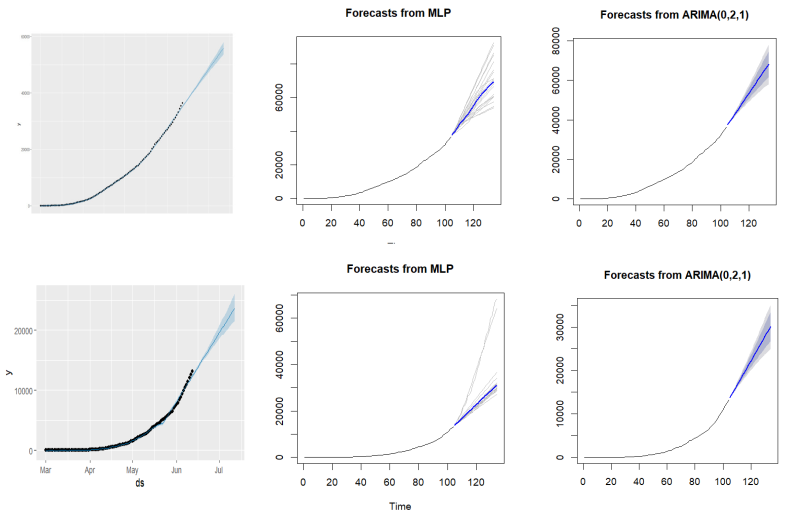
Google Mobility Report



Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

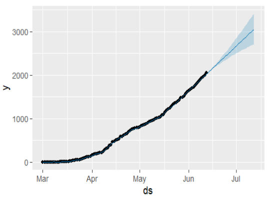
APPLIKASI I
PREDIKSI DENGAN METODE PREDICTIVE ANALYTICS

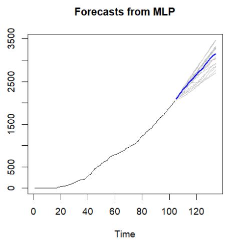
Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

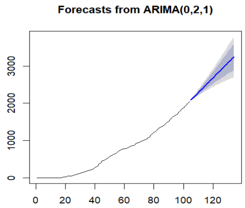


Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

STATISTICS R&D







STATISTICS R&D

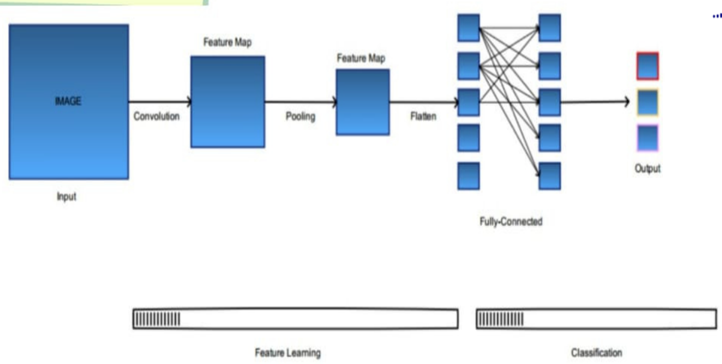
APPLIKASI II

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

COVID-19 CHEST X-RAY CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

CNN Architecture



- CNN architecture is created and inspired by the way humans produce visual perception (visual cortex)
- CNN is divided into 2 major parts, i.e. Feature Learning and Classification

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

Model 1

Table 2 Model 1

Number of Epoch	True Positive	False Positive	False Negative	True Negative
15	18	2	1	23
20	18	2	1	23
25	20	0	1	23
30	20	0	1	23

Table 3 Recall, Precision, and Accuracy of Model 1

Number of Epoch	Recall	Precision	Accuracy
15	94.74%	90%	93.2%
20	94.74%	90%	93.2%
25	95.24%	100%	97.73%
30	95.24%	100%	97.73%

The average recall, precision, and accuracy were respectively 94.99%, 95%, and 95.47%

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020

Dedi Rosadi, Data Science untuk Pemodelan Data Covid-19 , Webinar Jurusan Matematika UNSRAT, 15 Juni 2020