IF5181 Pengenalan Pola

Mining Data Sekuens

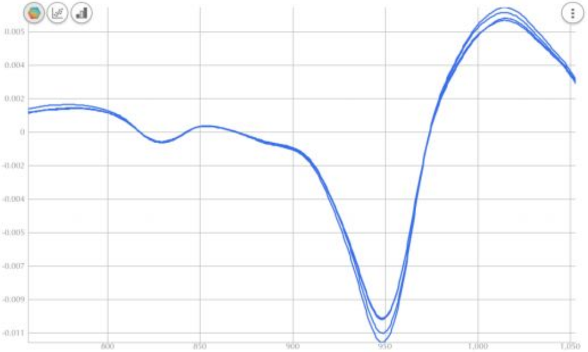
Masayu Leylia Khodra

Referensi

- Bab 8 & 9 dari Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
- https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-recurrent-neural-network-873c29da73c7
-

Mining Data Sekuens

Prediksi MakroNutrien Makanan



Gambar III.3. Spektrum NIR dari hasil scan menggunakan SCiO

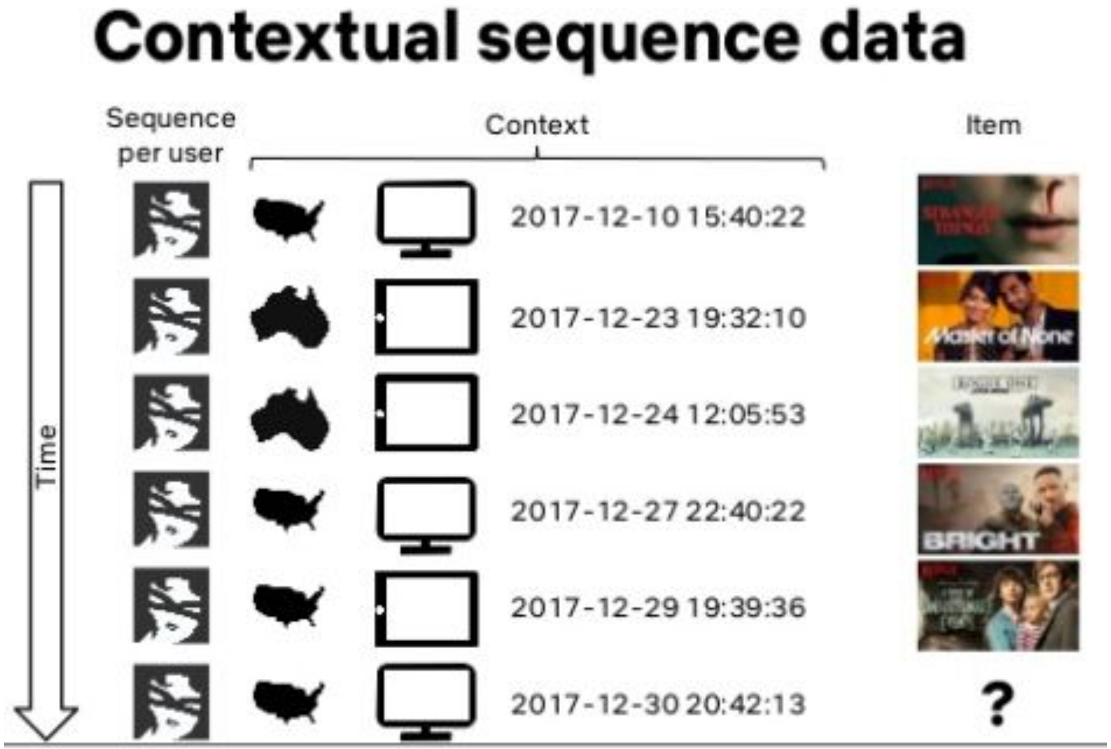
% karbo
% protein
% lemak

```
"Month","Passengers"
"1949-01",112
"1949-02",118
"1949-03",132
"1949-04",129
"1949-05",121
```

1961-01 ?

Prediksi Jumlah Penumpang Pesawat

Mining Data Sekuens (lanj)



Contextual sequence data

2017-12-10 15:40:22

2017-12-23 19:32:10

2017-12-24 12:05:53

2017-12-27 22:40:22

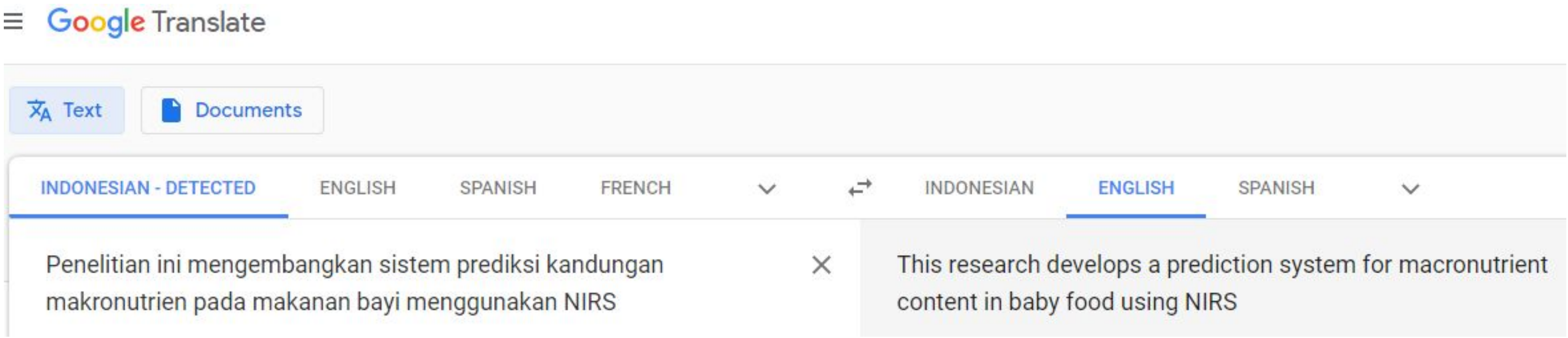
2017-12-29 19:39:36

2017-12-30 20:42:13

- Treat recommendation as sequence classification.
- Input: sequence of user actions
- Output: next action

https://www.slideshare.net/justinbasilico/recent-trends-in-personalization-a-netflix-perspective

Mining Data Sekuens (lanj)



Penelitian ini mengembangkan sistem prediksi kandungan makronutrien pada makanan bayi menggunakan NIRS

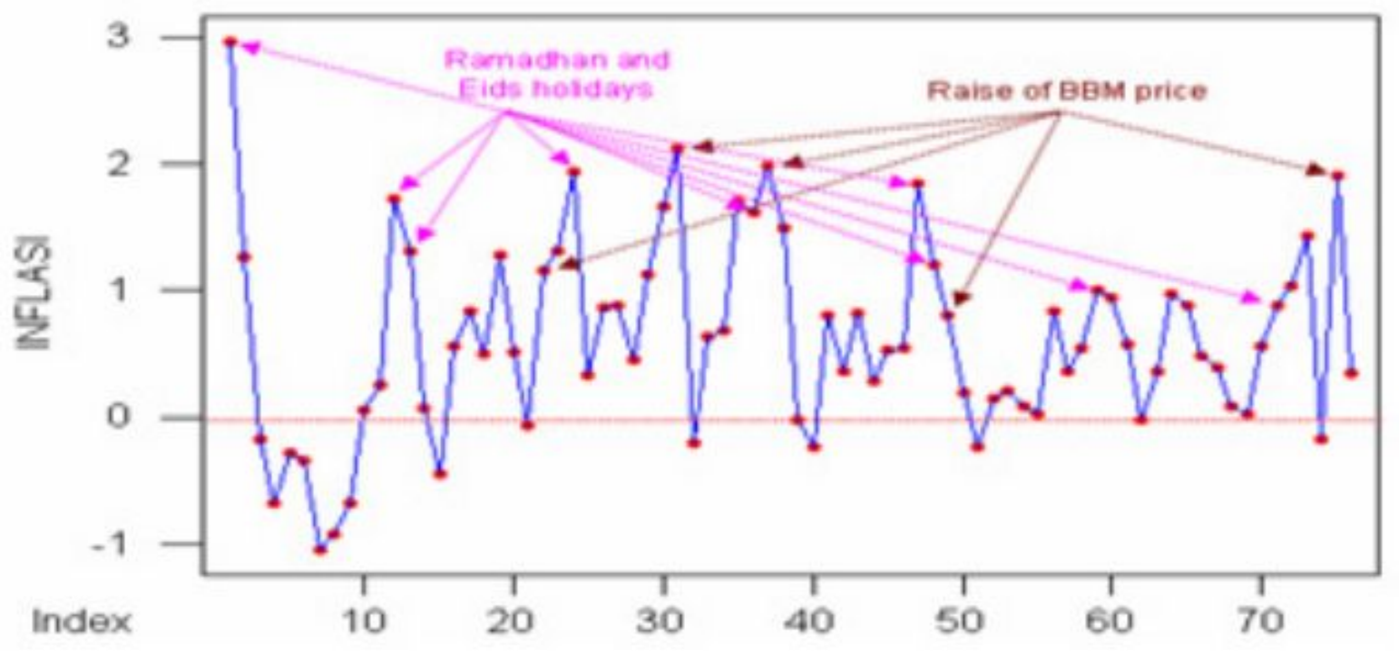
This research develops a prediction system for macronutrient content in baby food using NIRS

Mesin translasi menerima sekuens kata dan menghasilkan sekuens kata

Data Sekuens: the order matter

- Time-series data (numeric, equal time interval)
- Symbolic sequence data (nominal)
- Biological sequence data
- Natural language data (character order, word order, sentence order, paragraph order)
- ...

Time-series data

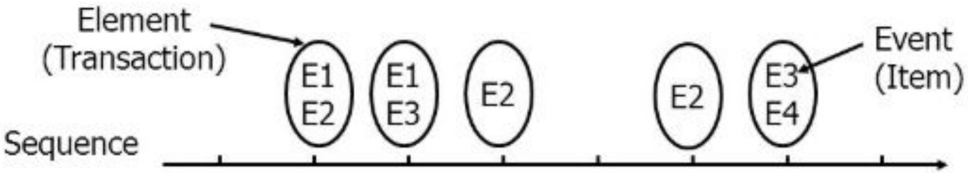


- In time-series data, sequence data consist of long sequences of numeric data, recorded at equal time intervals
- Data bulanan inflasi di Indonesia Januari 2009 sd April 2015 (Hidayat dkk., 2016)

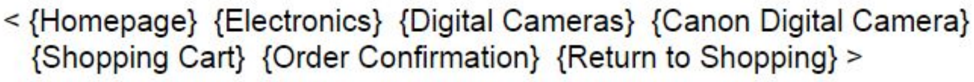
Hidayat, Y., Sutijo, B., Bon, A. T., & Supian, S. (2016). Indonesian financial data modeling and forecasting by using econometrics time series and neural network. *Global Journal of Pure and Applied Mathematics*, *12*(4), 3745-3757.

Symbolic Sequence Data

Symbolic sequence data consist of long sequences of event or nominal data, which typically are not observed at equal time intervals.

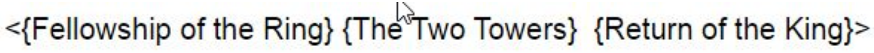


Browsing history:



< {Homepage} {Electronics} {Digital Cameras} {Canon Digital Camera}
{Shopping Cart} {Order Confirmation} {Return to Shopping} >

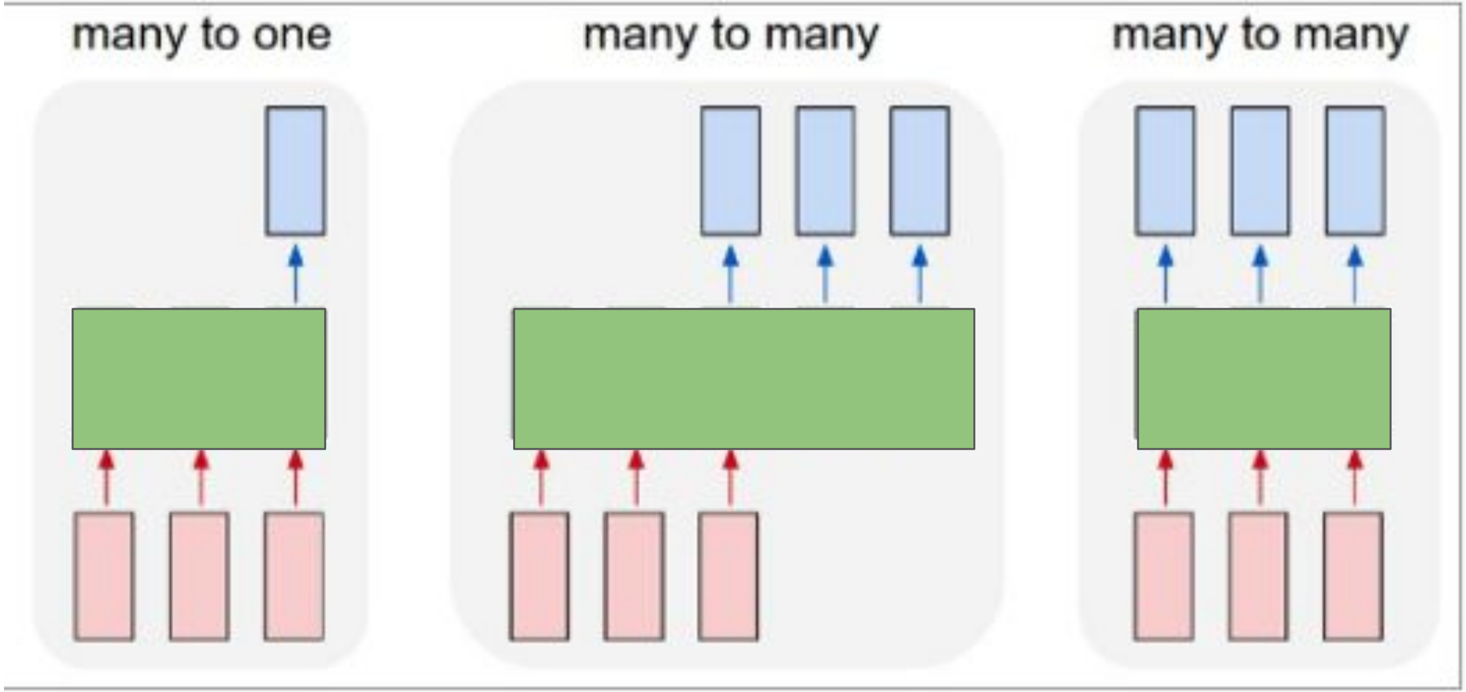
Sequence of books checked out at library:



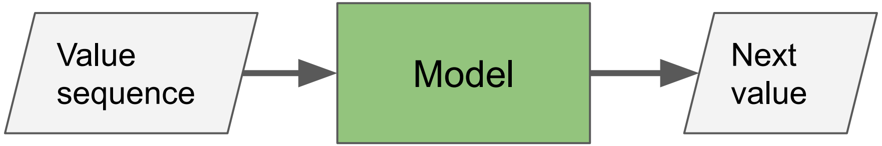
<{Fellowship of the Ring} {The Two Towers} {Return of the King}>

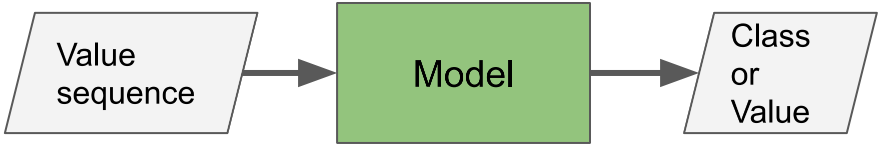
Tan dkk. (2004): https://slideplayer.com/slide/778153/

Kategori Persoalan Klasifikasi Data Sekuens



Many to One

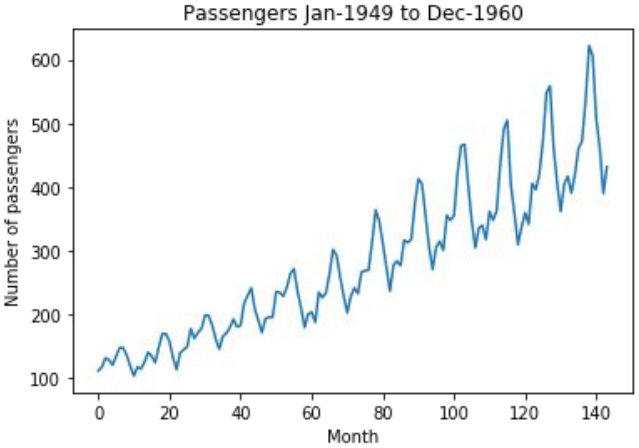




- Prediksi inflasi bulan berikutnya
- Prediksi jumlah penumpang bulan berikutnya
- Prediksi film berikutnya yang diklik
- Prediksi karakter atau kata berikutnya (model bahasa)

- Prediksi makronutrien dari spektrum gelombang
- Prediksi naik turunnya saham hari berikutnya
- Prediksi

Contoh: Prediksi Jumlah Penumpang



```
"Month","Passengers"
"1949-01",112
"1949-02",118
"1949-03",132
"1949-04",129
"1949-05",121
```

<u>1 Feature Dataset:</u>

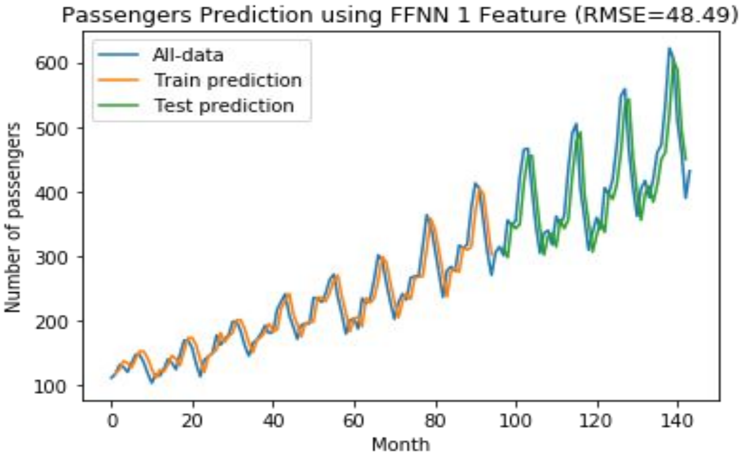
X=t	Y=(t+1)
112	118
118	132
132	129
129	121
121	135

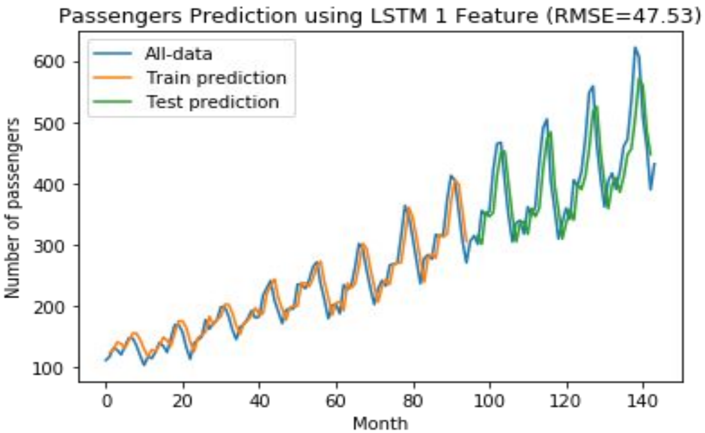
<u>3 Feature Dataset:</u>

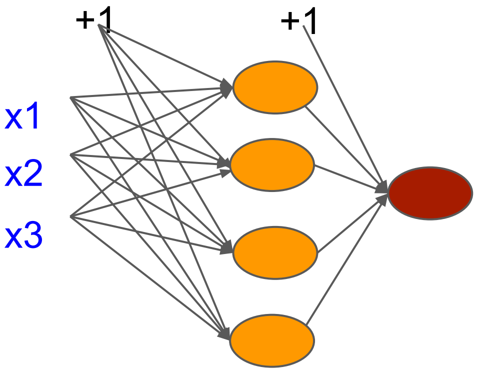
X1	X2	X3	Y
112	118	132	129
118	132	129	121
132	129	121	135
129	121	135	148
121	135	148	148

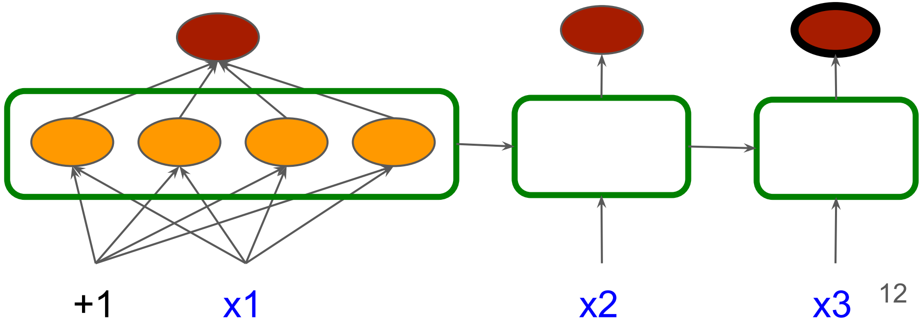
https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/

FFNN vs RNN: 1 hidden layer 4 neuron, 1 output neuron

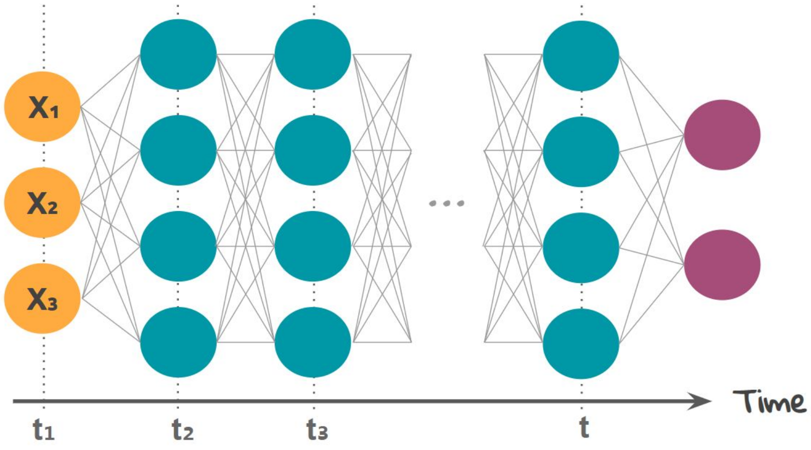

Passengers Prediction using FFNN 1 Feature (RMSE=48.49)

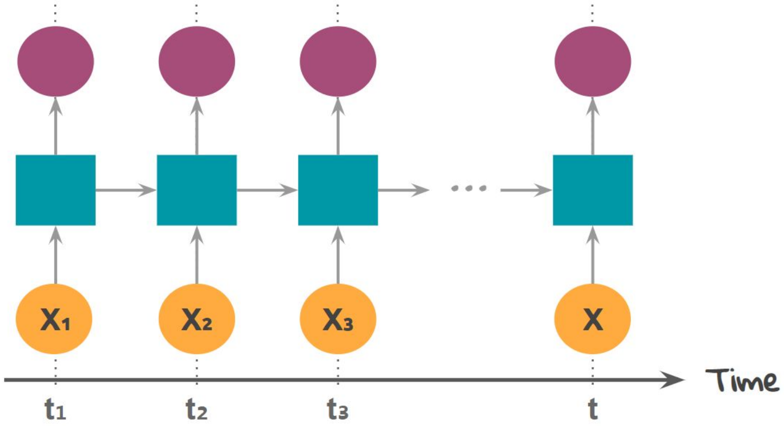

Passengers Prediction using LSTM 1 Feature (RMSE=47.53)





FFNN vs RNN: Sequential Data

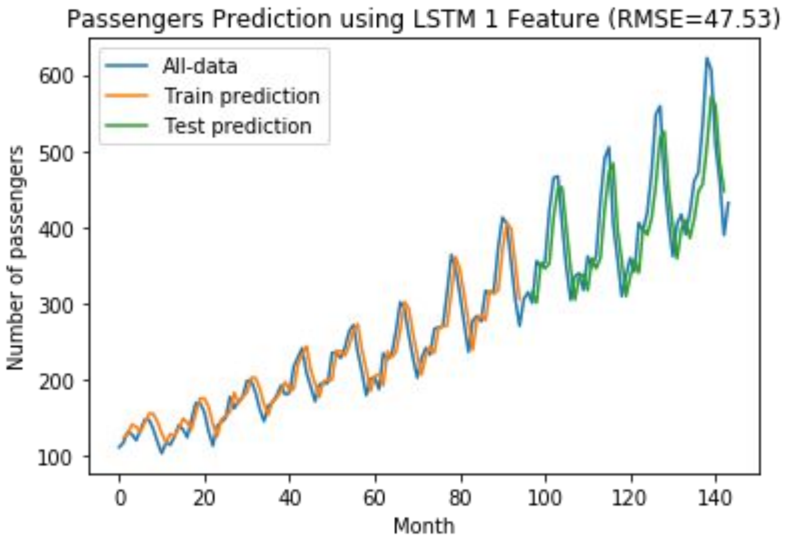


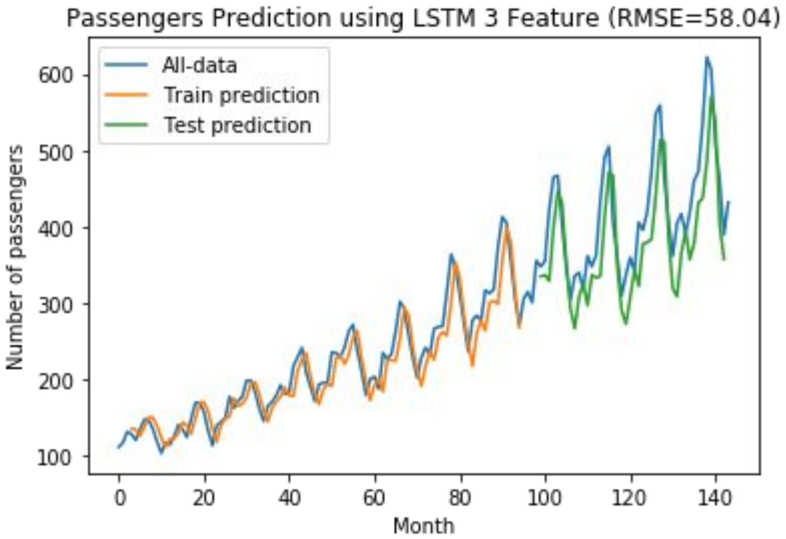


- FFNN: there isn't any concept of order in time between the data
- RNN: there is order in time between the data. We will input **X1** first and then input **X2** to the result of **X1** computation. So in the same way, **X3** is computed with the result from **X2** computation stage.

https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-recurrent-neural-network-873c29da73c7

Contoh 1: Predict Passengers (Hasil)





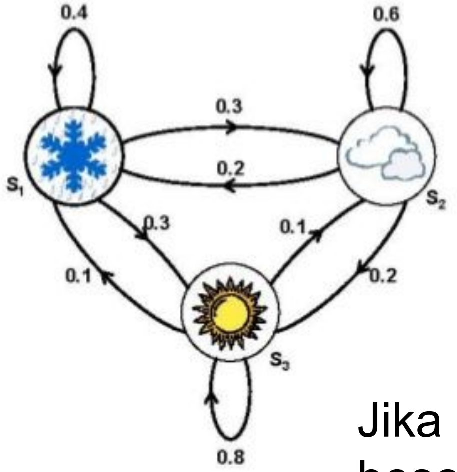
Contoh: Prediksi Cuaca dgn Simple Markov Model

- State 1: precipitation (rain, snow, hail, etc.)
- State 2: cloudy
- State 3: sunny

Transitions between states are described by transition matrix

$$A = \{a_{ij}\} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

This model can then be described by the following directed graph



Jika hari ini sunny, peluang besok sunny adalah:
P(sunny|model)
=P(sunny)*P(sunny|sunny)

https://www.slideshare.net/ssrdigvijay88/hmm-viterbi

Contoh: Prediksi Cuaca dgn Hidden Markov Model

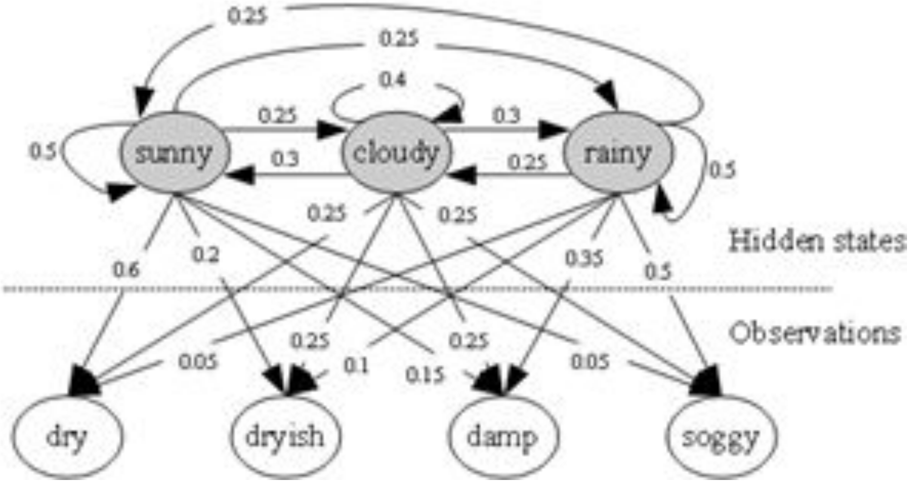
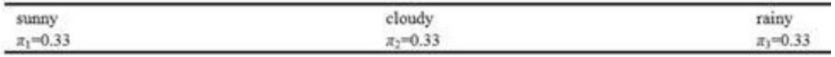
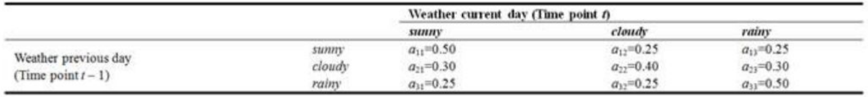


Table 2. Uniform initial state distribution ∏.



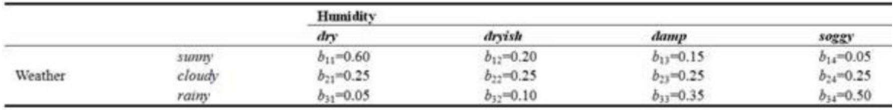
sunny	cloudy	rainy
π_1=0.33	π_2=0.33	π_3=0.33

Table 1. Transition probability matrix A.



		Weather current day (Time point *t*)		
		sunny	***cloudy***	***rainy***
Weather previous day (Time point *t* – 1)	*sunny*	a_{11}=0.50	a_{12}=0.25	a_{13}=0.25
	cloudy	a_{21}=0.30	a_{22}=0.40	a_{23}=0.30
	rainy	a_{31}=0.25	a_{32}=0.25	a_{33}=0.50

Table 3. Observation probability matrix B.

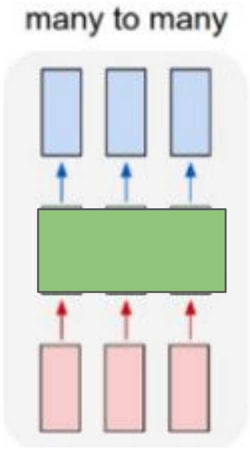


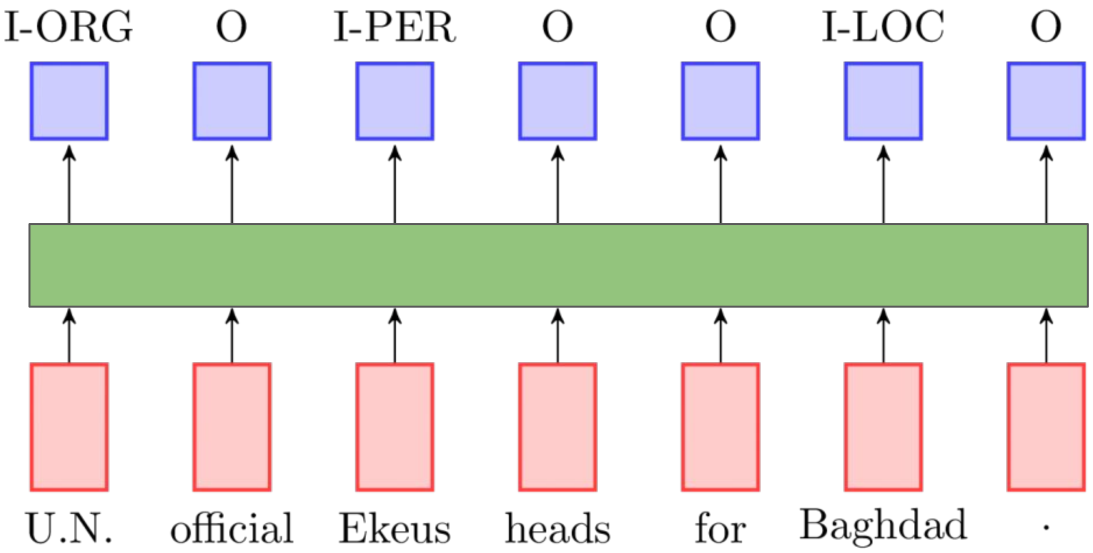
		Humidity			
		dry	***dryish***	***damp***	***soggy***
Weather	*sunny*	b_{11}=0.60	b_{12}=0.20	b_{13}=0.15	b_{14}=0.05
	cloudy	b_{21}=0.25	b_{22}=0.25	b_{23}=0.25	b_{24}=0.25
	rainy	b_{31}=0.05	b_{32}=0.10	b_{33}=0.35	b_{34}=0.50

Prediksi cuaca berdasarkan observasi tentang humidity: *dry*, *dryish*, *damp*, *soggy*

http://article.sciencepublishinggroup.com/html/10.11648.j.acm.s.2017060401.12.html

Contoh: Prediksi POS Tagging





Input is a sequence of words, and output is the sequence of POS tag for each word.

https://www.depends-on-the-definition.com/guide-sequence-tagging-neural-networks-python/

Penutup

- Klasifikasi data sekuens dapat dipandang sebagai persoalan klasifikasi biasa dengan mentransformasi dataset.
- Algoritma pembelajaran khusus data sekuens: Simple Markov Model, HMM, RNN