



Proprietary document of Indonesia AI 2023

Text Classification using RNN & LSTM

Indonesia AI

OBJECTIVE & OUTLINE

Proprietary document of Indonesia AI 2023



Text Classification using RNN & LSTM

Objektif: Memahami konsep dari klasifikasi teks menggunakan RNN & LSTM dalam NLP

Outline:

1. Text Classification
2. RNN
3. LSTM

— Apa itu Text Classification?

TEXT CLASSIFICATION

Proprietary document of Indonesia AI 2023



Teknik untuk mengklasifikasikan atau **mengelompokkan teks** atau dokumen **ke dalam** sebuah **label tertentu**

TEXT CLASSIFICATION

Proprietary document of Indonesia AI 2023



Contoh

- Spam detection
- Hate Speech detection
- Article categorization

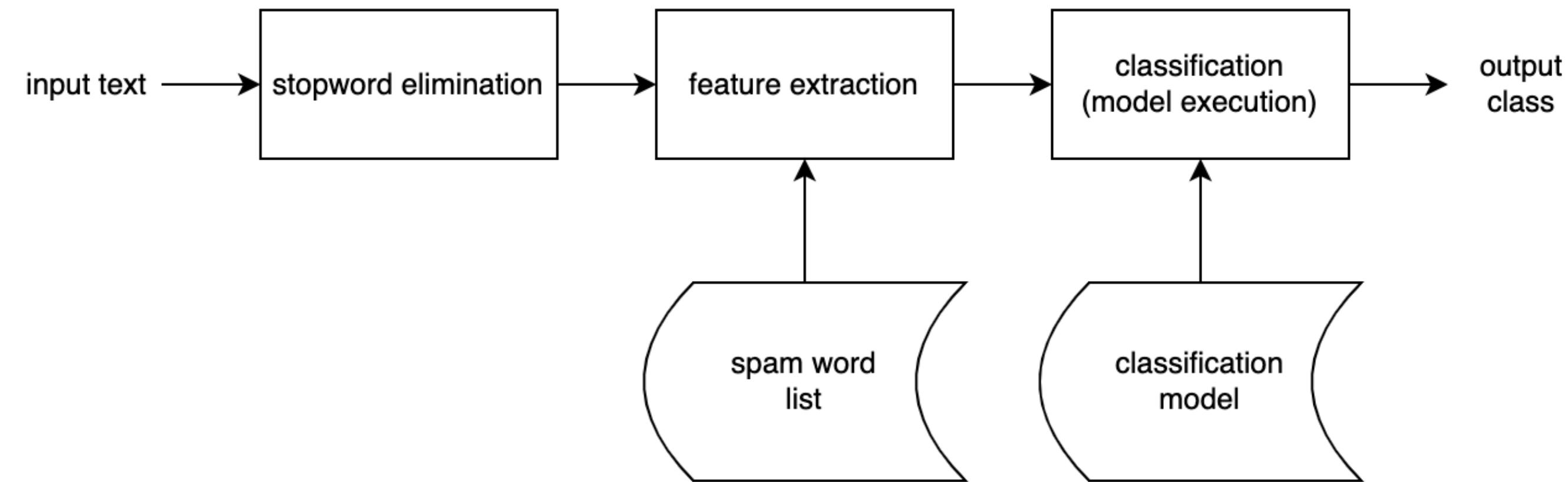


Spam detection

- Terdiri dari dua label, spam & bukan spam
- Menggunakan daftar kata berupa kata-kata yang banyak muncul dalam teks spam misalnya jika 75% terdiri dari kata-kata spam maka teks tersebut dianggap spam.
 - Daftar kata itu adalah feature
 - Aturan 75% itu adalah teknik klasifikasi(rule-based maupun model based)
 - Rule-based
 - Model-based

TEXT CLASSIFICATION

Proprietary document of Indonesia AI 2023



— Any question guys ~

RNN

Indonesia AI

Jaringan saraf berulang atau recurrent neural network (**RNN**) adalah jenis arsitektur jaringan saraf tiruan yang pemrosesannya dipanggil berulang-ulang untuk memroses masukan yang biasanya adalah data sekuensial.

RNN termasuk dalam kategori **deep learning** karena data diproses melalui banyak layer.

RNN adalah sistem algoritma tua yang telah **dikembangkan** sejak tahun **1980-an**. Sistem ini dinilai penting karena **menjadi satu-satunya sistem** yang memiliki **memori internal** pada masa itu.

Kemudian, pada 1990-an, banyak modifikasi yang menyempurnakan RNN, salah satunya adalah long short term memory (**LSTM**).

Pemrosean Data Sekuensial





Data sekuensial mempunyai **karakteristik** di mana **sampel diproses dengan suatu urutan** (misalnya waktu), dan suatu sampel dalam urutan mempunyai **hubungan erat** satu dengan yang lain.



Pengaplikasian

- rangkaian kata-kata dalam penerjemahan bahasa
- sinyal audio dalam pengenalan suara
- nada-nada dalam sintesa musik
- rangkaian kata-kata dalam klasifikasi sentimen

Pemrosesan data sekuensial seperti di atas **tidak cocok** dilakukan dengan **model umpan maju (feed forward)** yang sudah kita kenal seperti model **linear, multi-layer perceptron, dan CNN.**



- model-model ini tidak mempunyai “memori”; setiap sampel akan diproses secara sama tanpa ada konsiderasi atas sampel-sampel sebelumnya.
- model-model tersebut mengasumsikan data yang IID (independent and identically distributed), dan hal ini tidak dapat dipenuhi oleh data sekuensial karena suatu sampel mempunyai ketergantungan yang erat dengan sampel-sampel lainnya.
- model-model ini memroses masukan dengan panjang input yang tetap, sedangkan data sekuensial mempunyai panjang yang tidak tentu, dan bisa jadi sangat panjang.

Tidak ada yang membuat saya senang, tertawa, dan bahagia
dalam film ini.

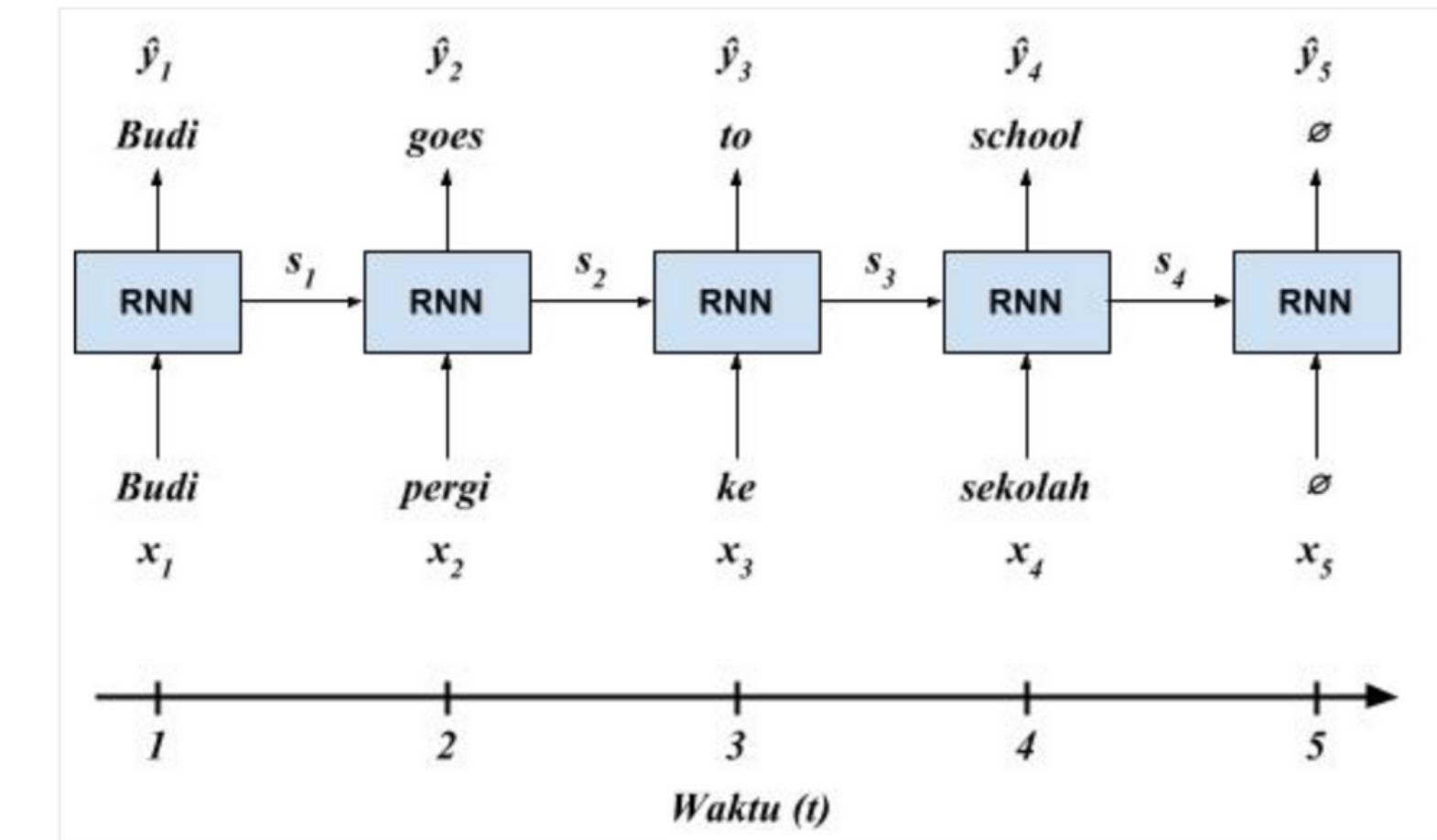
Sebuah **model** yang **naif** mungkin akan **menyimpulkan** bahwa kalimat di atas adalah ulasan **positif**, karena ada banyak kata-kata positif di kalimat di atas (“**senang**”, “**tertawa**”, “**bahagia**”). Padahal jauh **di depan** ada kata “**tidak**”.



RNN

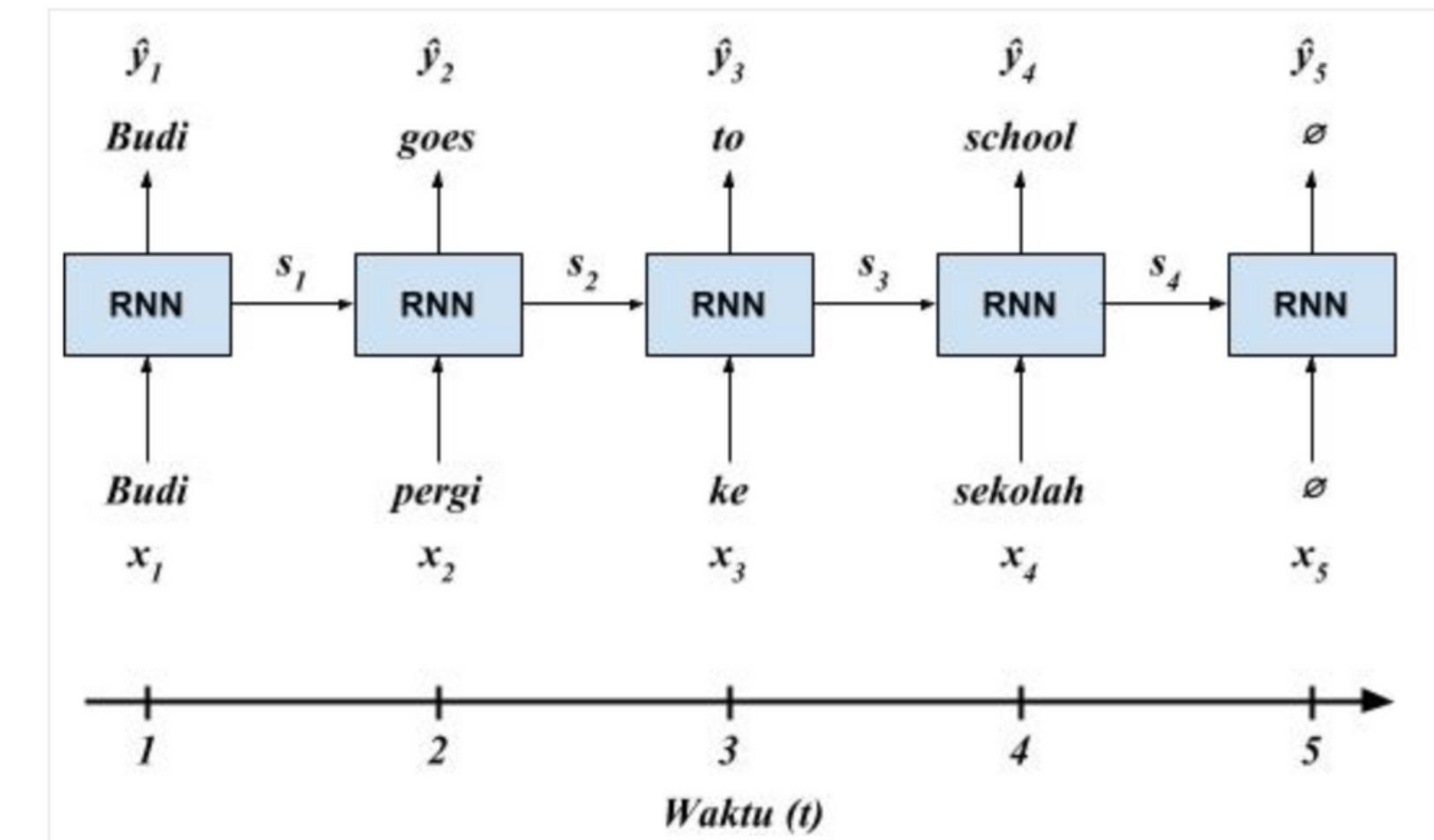
Proprietary document of Indonesia AI 2023

- sumbu horizontal adalah waktu, direpresentasikan dengan simbol t. Dapat kita bayangkan pemrosesan berjalan dari kiri ke kanan. Selanjutnya kita sebut t adalah langkah waktu (time step).
Budi pergi ke sekolah .
- Keseluruhan input adalah kalimat, dalam hal ini:



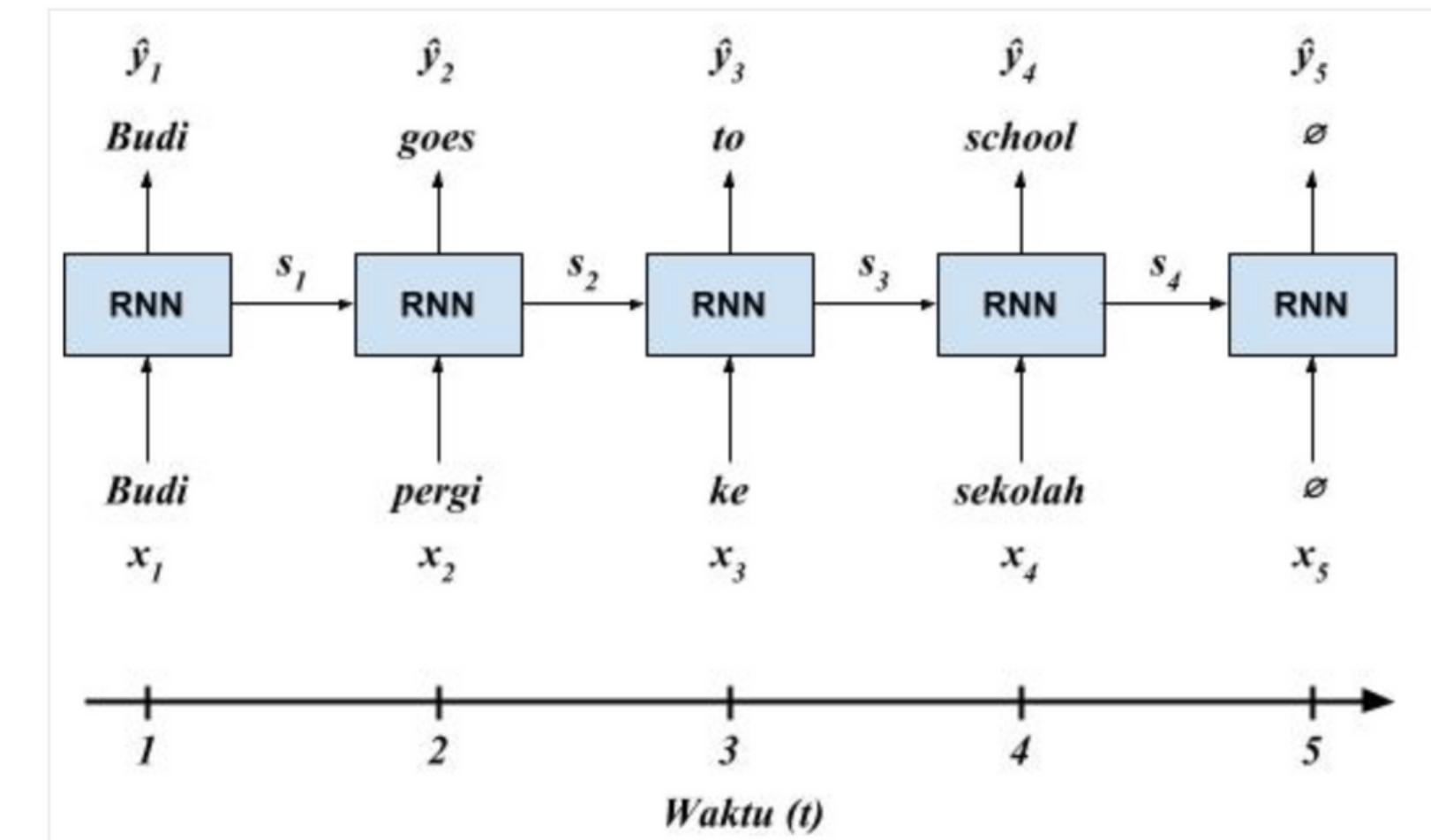
RNN

- Pemrosesan input oleh RNN adalah kata demi kata. Input kata-kata ini disimbolkan dengan x_1, x_2, \dots, x_5 , atau secara umum x_t .
- Output adalah kalimat, dalam hal ini:
Budi goes to school .
- RNN memberikan output kata demi kata, dan ini kita simbolkan dengan $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_5$, atau secara umum \hat{y}_t .
- Dalam tiap pemrosesan, RNN akan menyimpan state internal yaitu s_t , yang diberikan dari satu langkah waktu ke langkah waktu berikutnya. Inilah “memori” dari RNN.



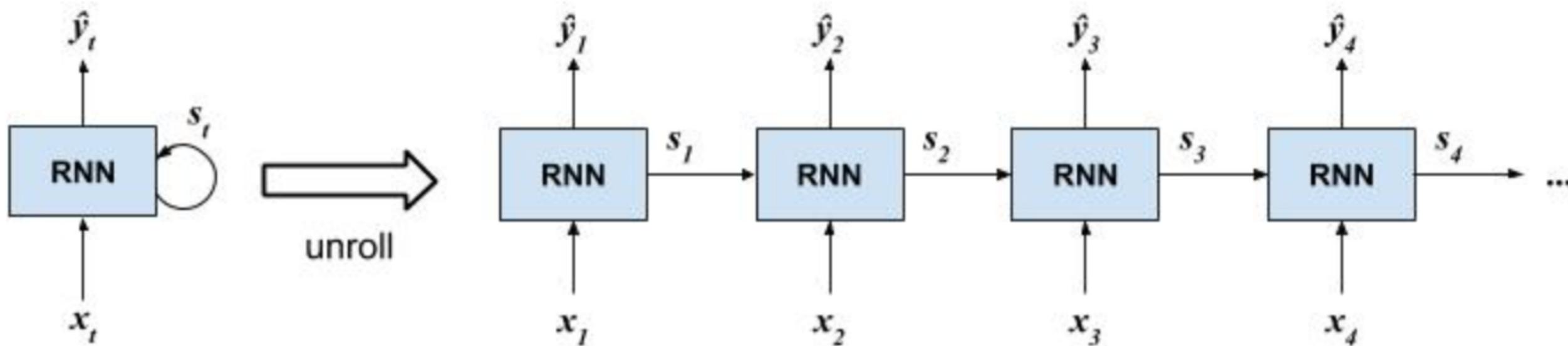
RNN

- Pemrosesan input oleh RNN adalah kata demi kata. Input kata-kata ini disimbolkan dengan x_1, x_2, \dots, x_5 , atau secara umum x_t .
- Output adalah kalimat, dalam hal ini:
Budi goes to school .
- RNN memberikan output kata demi kata, dan ini kita simbolkan dengan $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_5$, atau secara umum \hat{y}_t .
- Dalam tiap pemrosesan, RNN akan menyimpan state internal yaitu s_t , yang diberikan dari satu langkah waktu ke langkah waktu berikutnya. Inilah “memori” dari RNN.

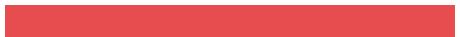


RNN

Proprietary document of Indonesia AI 2023



ilustrasi di sebelah kanan adalah penjabaran (unrolled) dari versi berulang di sebelah kiri.



Jenis RNN

- One to One:
 - Jenis RNN ini biasa digunakan untuk memecahkan masalah dalam machine learning. One to one juga dikenal sebagai vanilla neural network yang hanya bisa menampung satu input dan menghasilkan satu output saja.



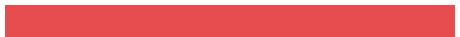
Jenis RNN

- One to Many:
 - Jenis berikutnya dari RNN yang dapat menghasilkan beberapa output atas satu input. Tipe ini banyak diaplikasikan pada caption gambar.



Jenis RNN

- Many to One:
 - Jenis many to one yang dapat menerima banyak input untuk menghasilkan satu output saja. Jenis ini paling sering digunakan untuk menentukan sentimen yang dapat mengklasifikasikan berdasarkan emosi negatif, netral, atau positif.

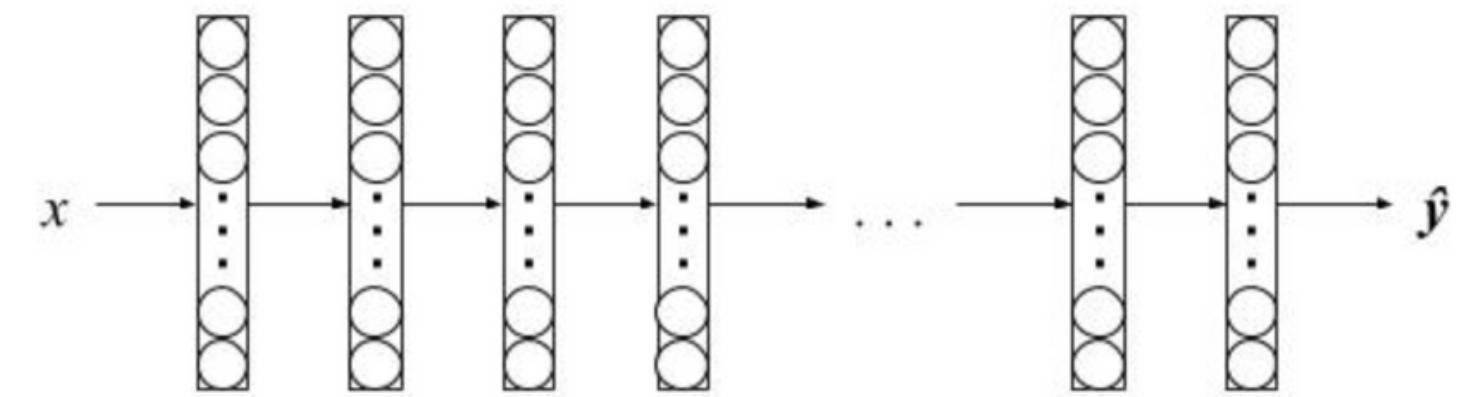


Jenis RNN

- Many to Many:
 - Jenis yang dapat menerima banyak input dengan beberapa opsi output yang ditentukan berdasarkan urutan. Jenis ini banyak digunakan dalam mesin penerjemahan.

Permasalahan Vanishing Gradient

Ketika **gradien menghilang**, artinya **informasi** mengenai **gradien dari lapisan sebelumnya tidak** dapat "mencapai" **lapisan-lapisan terdalam** pada jaringan RNN. Sebagai hasilnya, **bobot** yang terkait dengan lapisan terdalam **tidak mengalami perubahan** yang signifikan selama proses pelatihan, sehingga jaringan tidak dapat belajar dengan efektif dari data pelatihan.



Pengembangan RNN

- LSTM
- GRU
- Bi-LSTM



— Any question guys ~

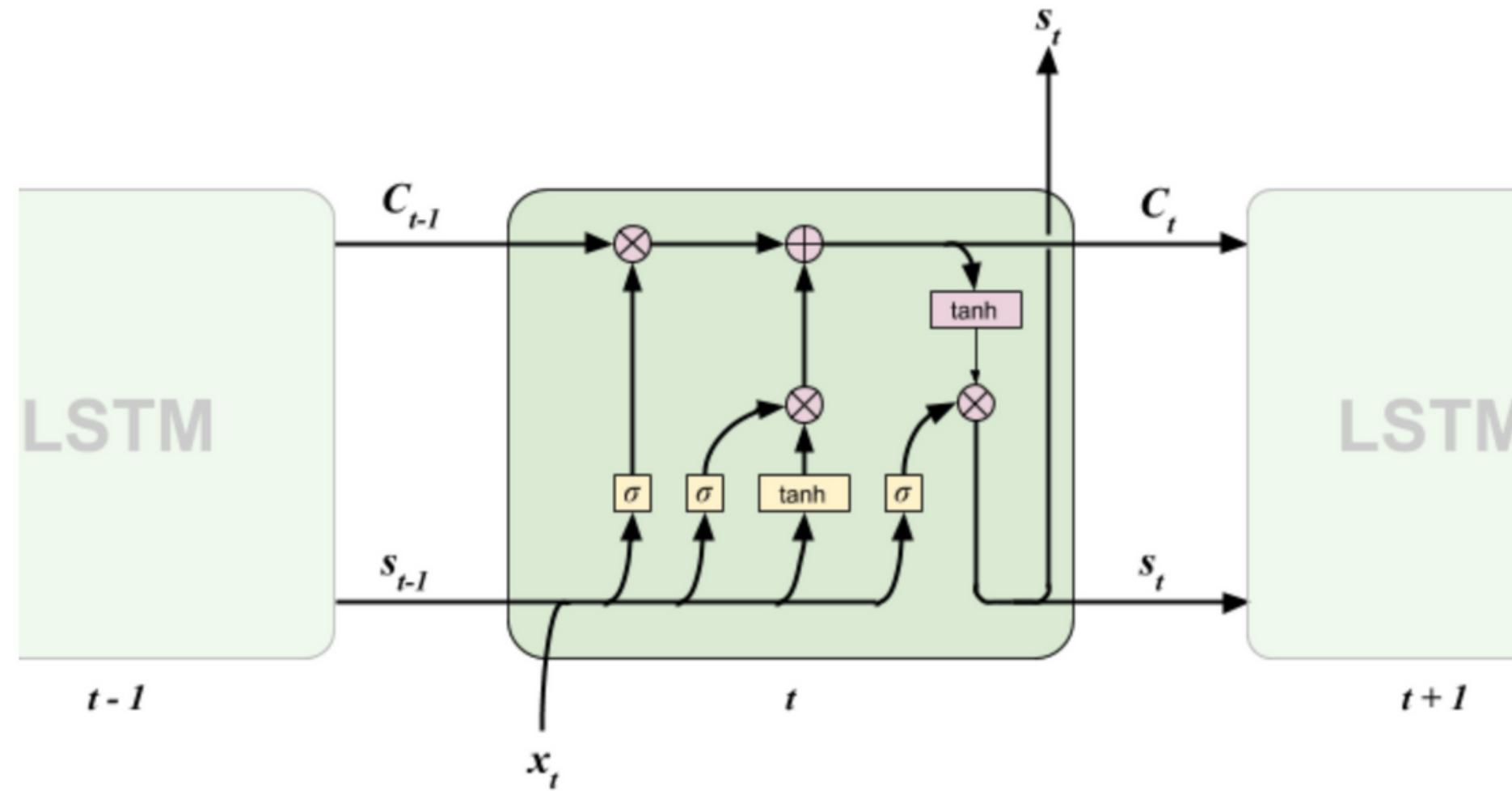
LSTM

Long short term memory network (**LSTM**) hadir untuk **melengkapi kekurangan RNN** yang **tidak dapat memprediksi kata** berdasarkan **informasi lampau** yang disimpan dalam **jangka waktu lama**.

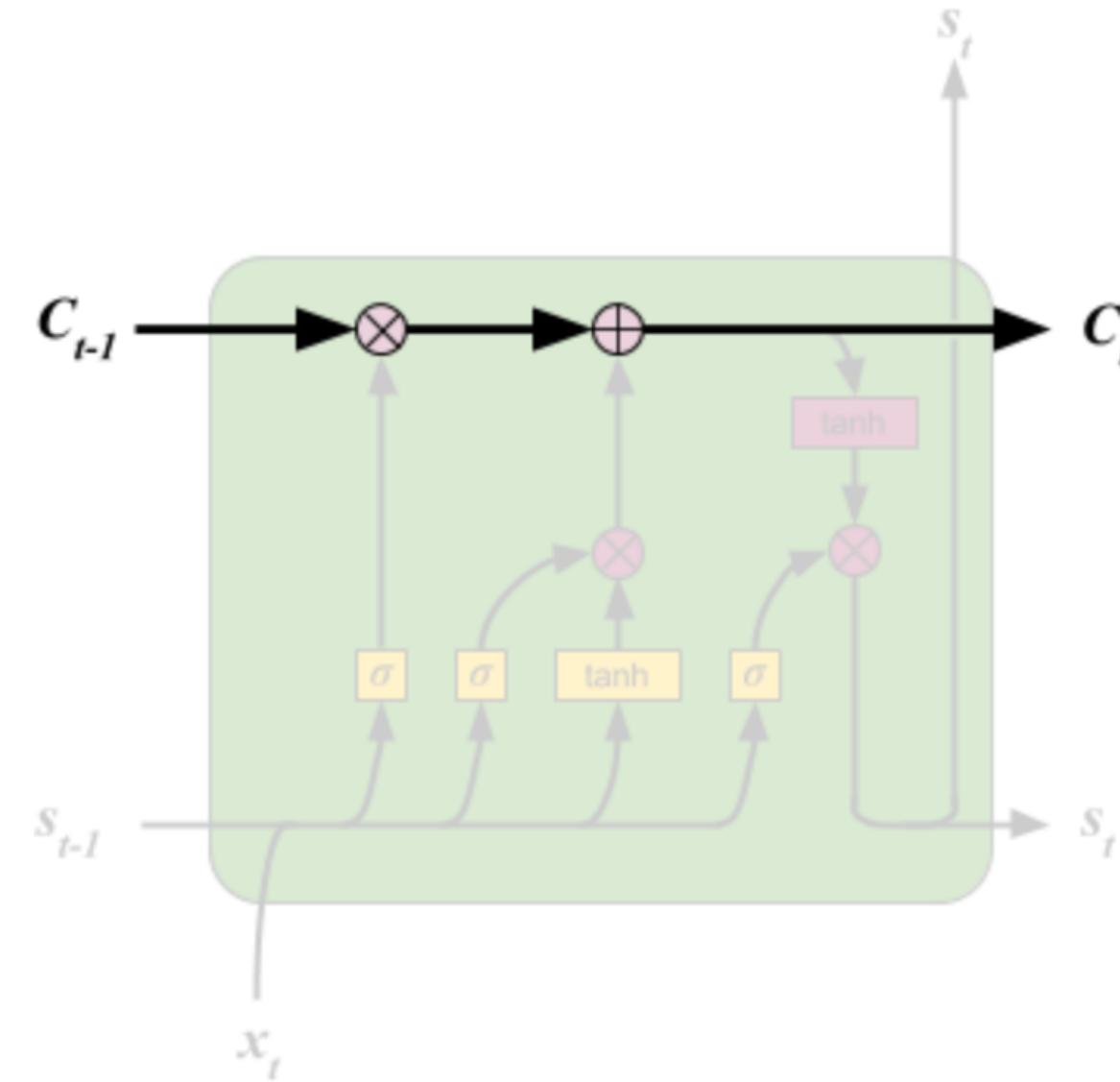
LSTM mampu **mengingat kumpulan informasi** yang telah **disimpan** dalam **jangka waktu panjang**, sekaligus **menghapus informasi** yang **tidak lagi relevan**. LSTM lebih efisien dalam memproses, memprediksi, sekaligus mengklasifikasikan data berdasarkan urutan waktu tertentu.

LSTM

Proprietary document of Indonesia AI 2023



Mekanisme Kunci LSTM

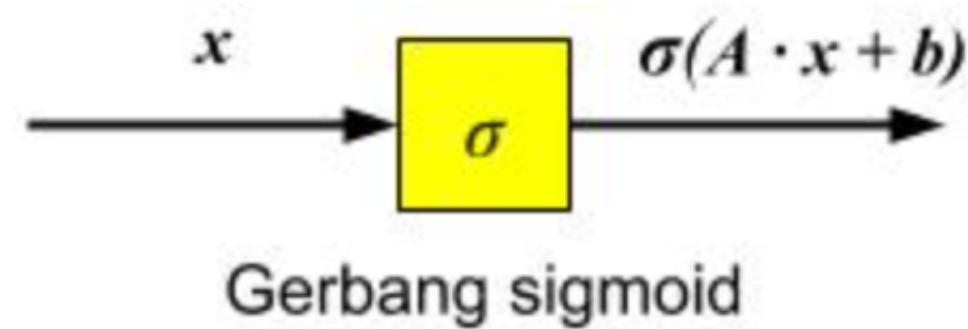


Ide kunci dari LSTM adalah jalur yang menghubungkan konteks lama (C_{t-1}) ke konteks baru (C_t) di bagian atas modul LSTM

Konteks C_t disebut juga cell state atau memory cell dalam beberapa artikel. Dengan adanya jalur di atas, suatu nilai di konteks yang lama akan dengan mudah diteruskan ke konteks yang baru dengan sedikit sekali modifikasi, kalau diperlukan.

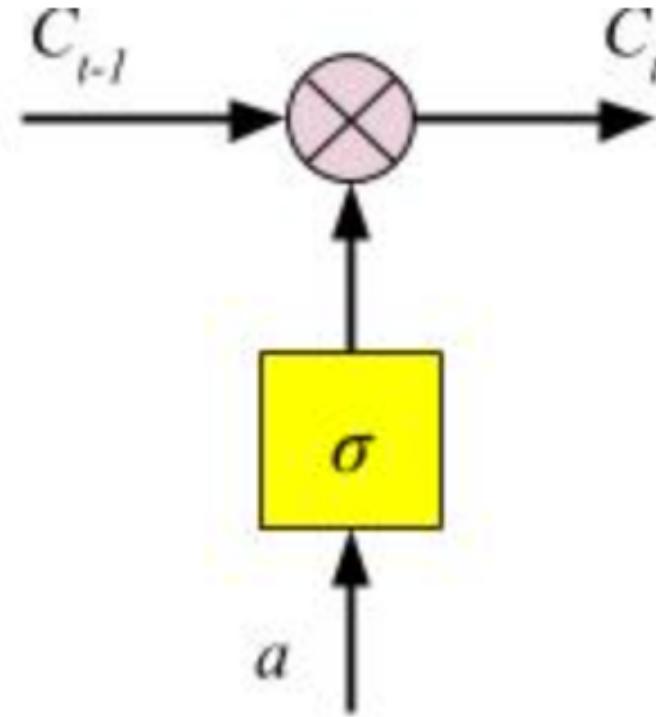
Mekanisme Kunci LSTM

Ide kunci yang lain adalah adanya gerbang sigmoid (sigmoid gate) yang mengatur seberapa banyak informasi bisa lewat.



Untuk suatu input x , output dari gerbang sigmoid adalah $\sigma(A \cdot x + b)$, di mana A adalah parameter, b adalah bias, keduanya dipelajari dalam proses latihan, dan σ adalah fungsi sigmoid. Keluaran gerbang adalah angka antara nol dan satu; nol artinya informasinya diblok total, sedangkan satu artinya ikutkan keseluruhan informasi.

Mekanisme Kunci LSTM

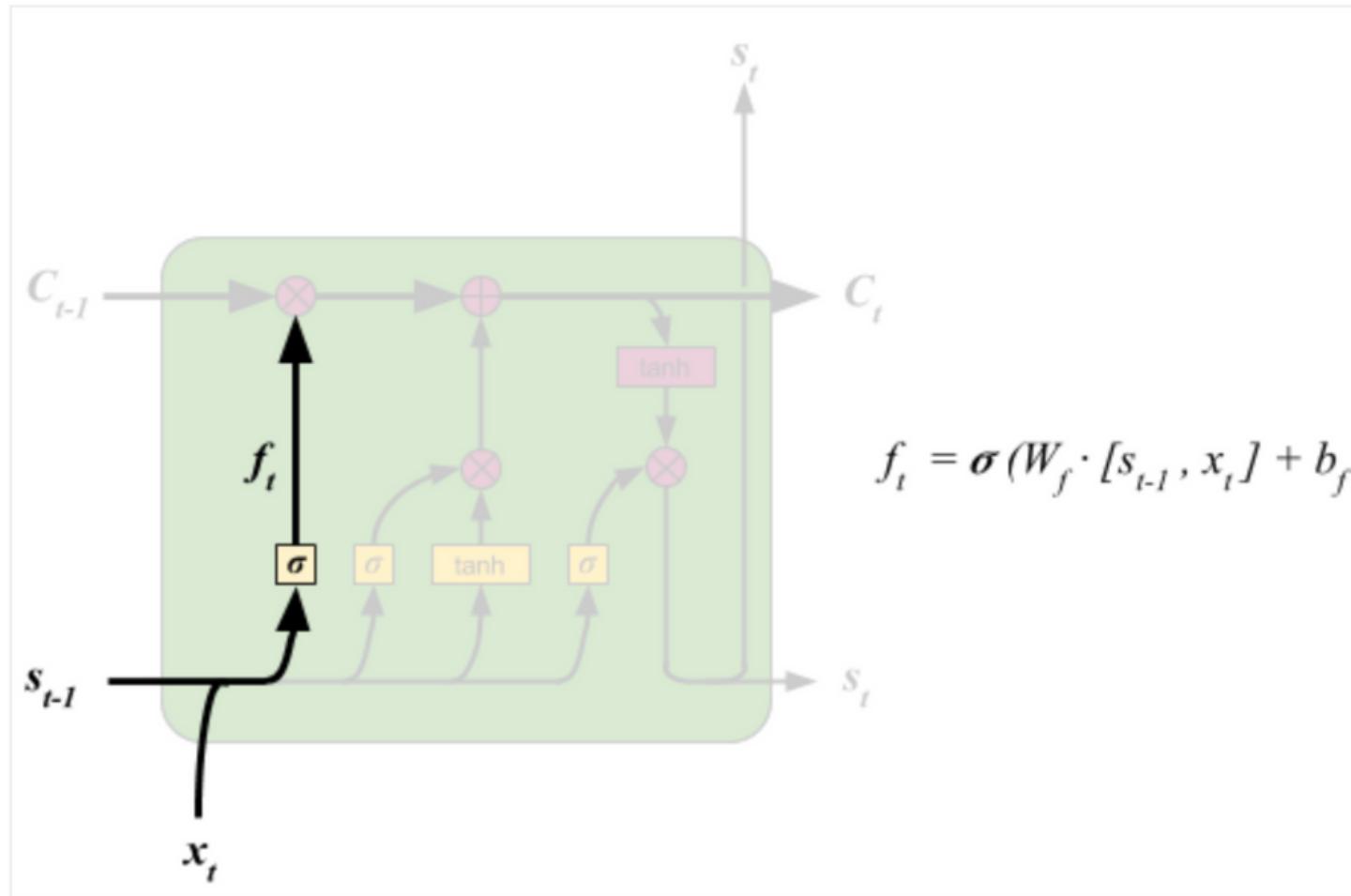


Keluaran dari gerbang sigmoid akan dikalikan dengan suatu nilai lain untuk mengontrol seberapa banyak nilai tersebut dipakai.

Dengan gerbang sigmoid di samping, LSTM bisa mengatur seberapa banyak informasi dari C_{t-1} yang diikutkan menjadi C_t .

Struktur LSTM

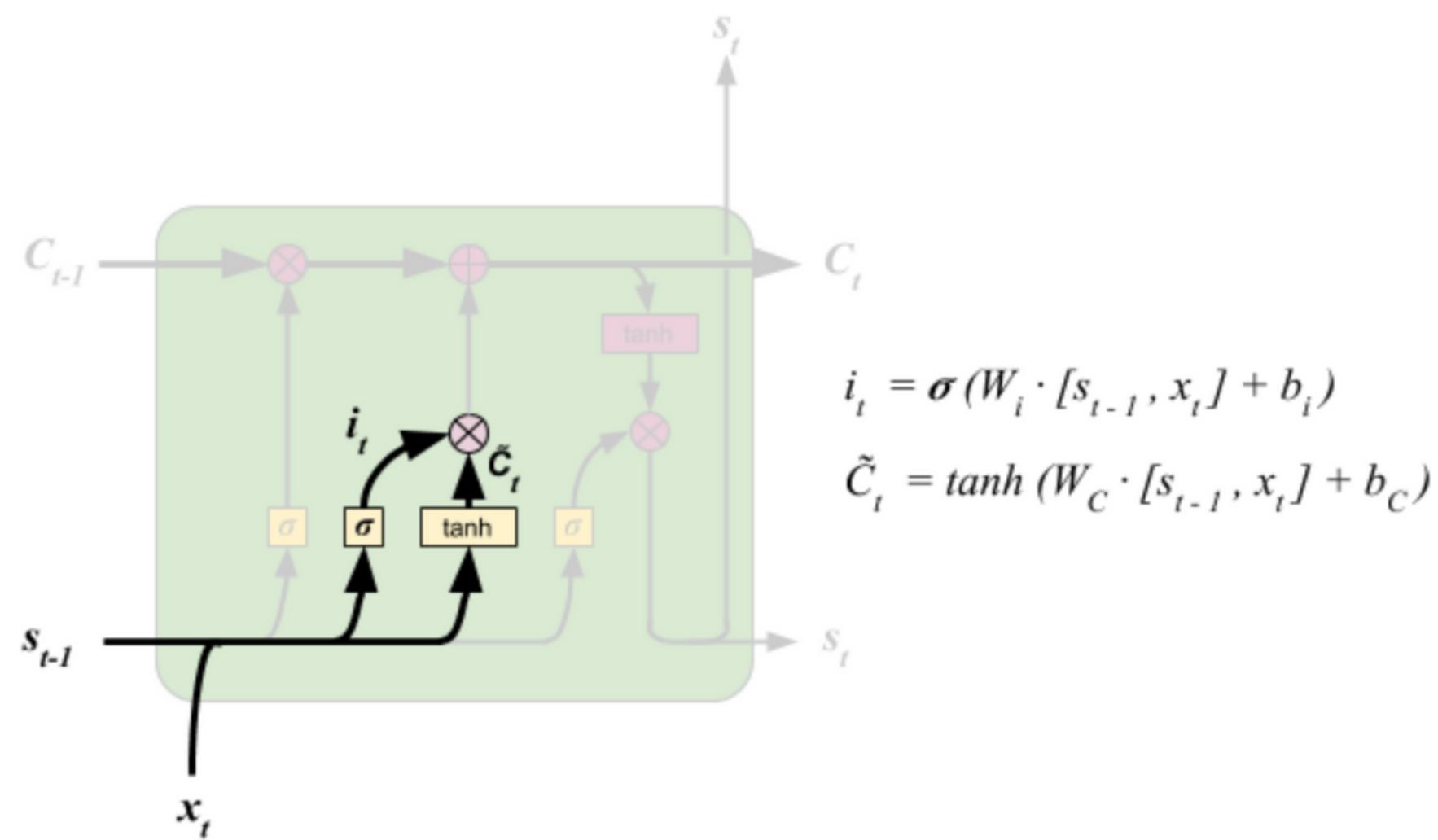
- Forget Gate



Gerbang pertama dalam LSTM disebut dengan forget gate. Mudahnya, gerbang ini bertugas untuk melupakan beberapa informasi yang tidak relevan dan sudah tidak diperlukan oleh sebuah sistem. Alhasil, LSTM dapat menyajikan kumpulan informasi yang lengkap, tetapi tetap aktual sesuai dengan kebutuhan.

Struktur LSTM

- Input Gate



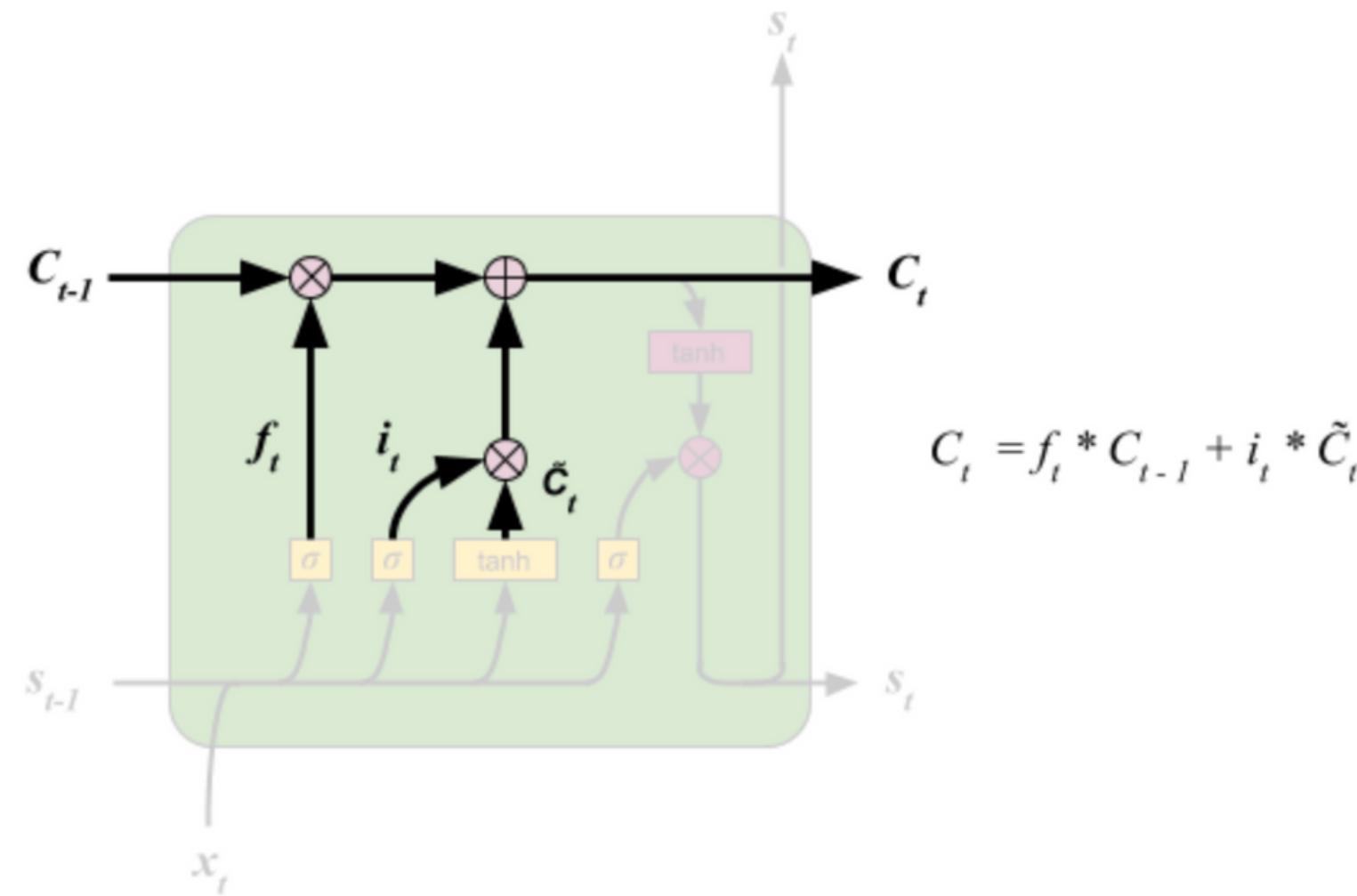
Berikutnya, ada gerbang kedua, yakni input gate yang bertugas untuk memasukkan informasi yang berguna untuk mendukung keakuratan data. Tugas input gate adalah untuk menambahkan informasi yang sebelumnya telah diseleksi terlebih dahulu melalui gerbang forget gate. Gerbang ini tidak dimiliki oleh RNN yang hanya memungkinkan satu input data untuk satu output data.

LSTM

Proprietary document of Indonesia AI 2023

Struktur LSTM

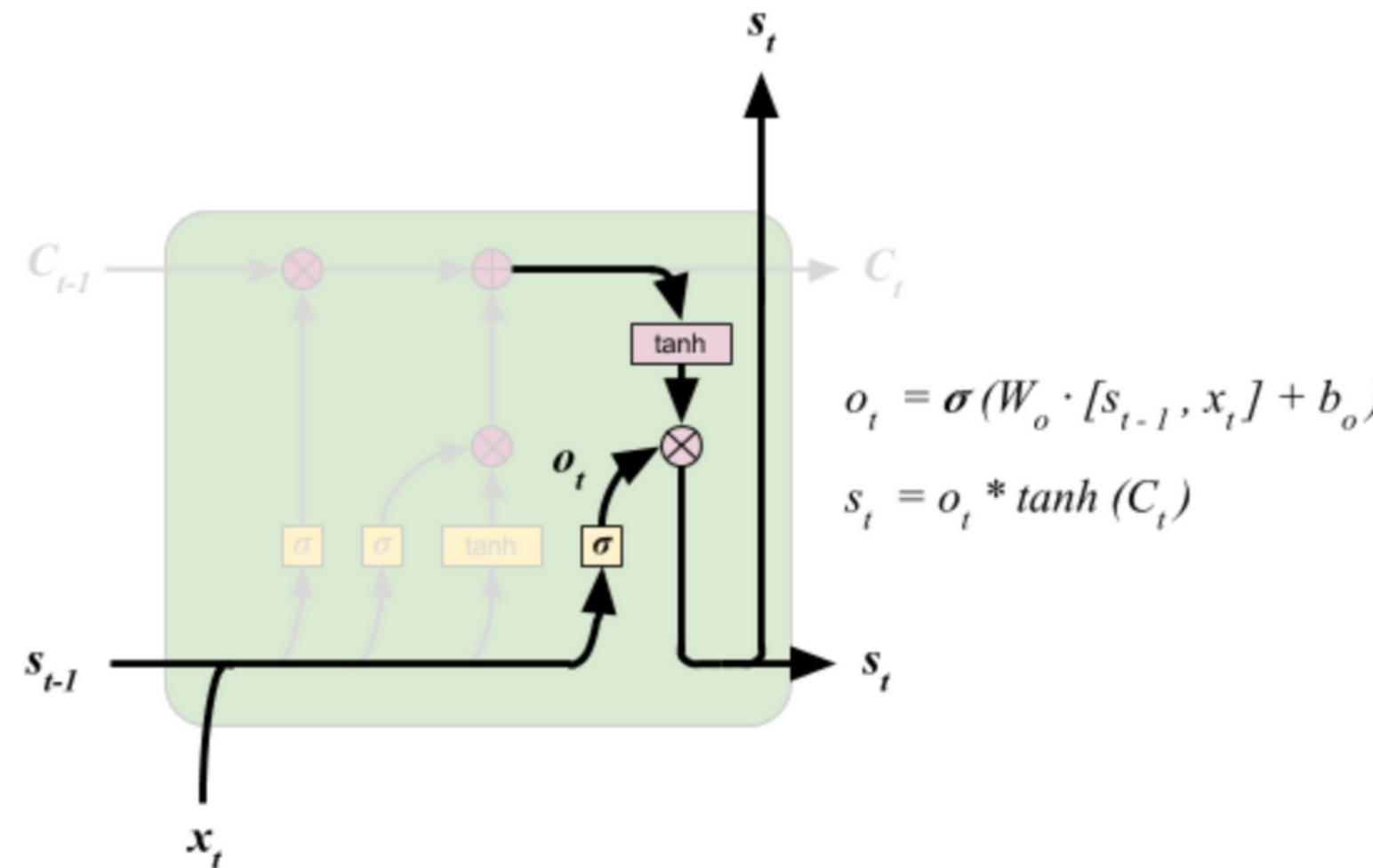
- Input Gate



Dalam input gate kemudian dikenal istilah input modulation gate yang sering tidak ditulis dalam beberapa ulasan tentang LSTM. Sesuai namanya, input modulation gate berfungsi untuk memodulasi informasi yang ada, sehingga dapat mengurangi kecepatan konvergensi dari data zero-mean.

Struktur LSTM

- Output Gate



Terakhir adalah output gate yang menjadi gerbang terakhir untuk menghasilkan informasi data yang komplit dan aktual. Gerbang ini bisa menjadi yang terakhir atas sebuah informasi atau hanya menjadi bagian dari tahap pertama saja, sebelum akhirnya informasi akan diproses lewat input gate di sel berikutnya.

— Any question guys ~

Terima Kasih!