

Machine Learning

Reinforcement Learning

Knowledge Engineering Research Group

Soft Computing Laboratory

Department of Information and Computer Engineering

Politeknik Elektronika Negeri Surabaya



Politeknik Elektronika Negeri Surabaya
Departemen Teknik Informatika dan Komputer

Konten

- Latar Belakang JST
- Struktur Jaringan Syaraf
- Sejarah
- JST yang sederhana
- Perceptron
- Single Perceptron Learning

Tujuan Instruksi Umum

Mahasiswa mampu menyelesaikan masalah – masalah menggunakan metode mesin pembelajaran yang tepat berdasarkan supervised, unsupervised dan reinforcement learning, baik secara individu maupun berkelompok/kerjasama tim.

Tujuan Instruksi Khusus

- Mengetahui konsep Reinforcement Learning
- Mengetahui proses dalam Reinforcement Learning
- Mengetahui penerapan Reinforcement Learning

Tipe Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Supervised Learning

Diibaratkan sebagai proses belajar dari seorang murid yang berada dalam sebuah kelas

- Si murid diperbolehkan bertanya kepada guru yang telah mengetahui aturan jawabannya, dan kemudian si dosen menjawab pertanyaan tersebut.
- Dari hasil tanya-jawab berkali-kali, si murid akan bisa memahami rule dari permasalahan, sehingga jika ada permasalahan lain, si murid akan membandingkan dengan rule yang ia simpulkan sebelumnya, sehingga ia bisa memberikan jawaban.
- Memerlukan training. Semakin lama training, semakin pintar pula si murid memecahkan masalah.



Unsupervised Learning

- Ketika si murid menjumpai masalah, ia harus dapat menjawab masalah tersebut dengan sendirinya.
- Semakin banyak ia berusaha menjawab sendiri, ia akan semakin pandai dalam menemukan rule yang dapat digunakan untuk memecahkan permasalahan di kemudian hari.
- Sangat bermanfaat jika memang permasalahan yang dihadapi relatif tidak bisa atau sulit sekali dijawab oleh sang guru.

Reinforcement Learning

Dalam Reinforcement Theory, terdapat 3 konsekuensi yang berbeda, yaitu :

- Konsekuensi yang memberikan reward
- Konsekuensi yang memberikan punishment
- Konsekuensi yang tidak memberikan apa-apa

Reinforcement Learning

Contoh adalah bagaimana sikap yang diambil oleh seorang siswa di dalam kelas.

- Asumsikan bahwa sang guru sudah menjelaskan seperangkap aturan yang harus ditaati oleh siswa di dalam kelas.
- Suatu ketika, seorang siswa berteriak di dalam kelas. Maka sang guru langsung memberikan hukuman kepada siswa tersebut.
- Dari hukuman itu, siswa tadi akan merubah sikapnya untuk tidak berteriak lagi.
- Juga demikian, kepada siswa yang tekun mengikuti pelajaran di dalam kelas, maka sang guru memberikan kepada mereka semacam hadiah atau penghargaan.
- Jika sistem ini berjalan dalam jangka waktu tertentu, maka keadaan siswa tadi pasti akan konvergen untuk mengambil sikap yang baik di dalam kelas.



Reinforcement Learning

- Seorang siswa yang bersikap baik di dalam kelas, ia akan mendapatkan reward. Dengan reward itu, ia akan bersikap lebih baik lagi. Jika ia bersikap lebih baik lagi, ia akan mendapatkan reward lagi. Demikian seterusnya yang terjadi sehingga ia pasti akan semakin konvergen dalam bersikap baik di dalam kelas.
- Sebaliknya, jika ia bersikap buruk, maka ia akan menerima punishment. Dengan punishment itu, ia akan merubah sikapnya. Jika punishment itu tidak cukup untuk membuatnya berubah, maka ia akan mendapatkan punishment lagi, sehingga dalam batasan tertentu, ia pasti akan berubah sikap yang hasilnya adalah ia akan mendapatkan reward. Demikian seterusnya, sehingga pada suatu saat nanti, ia akan konvergen bersikap baik di dalam kelas.

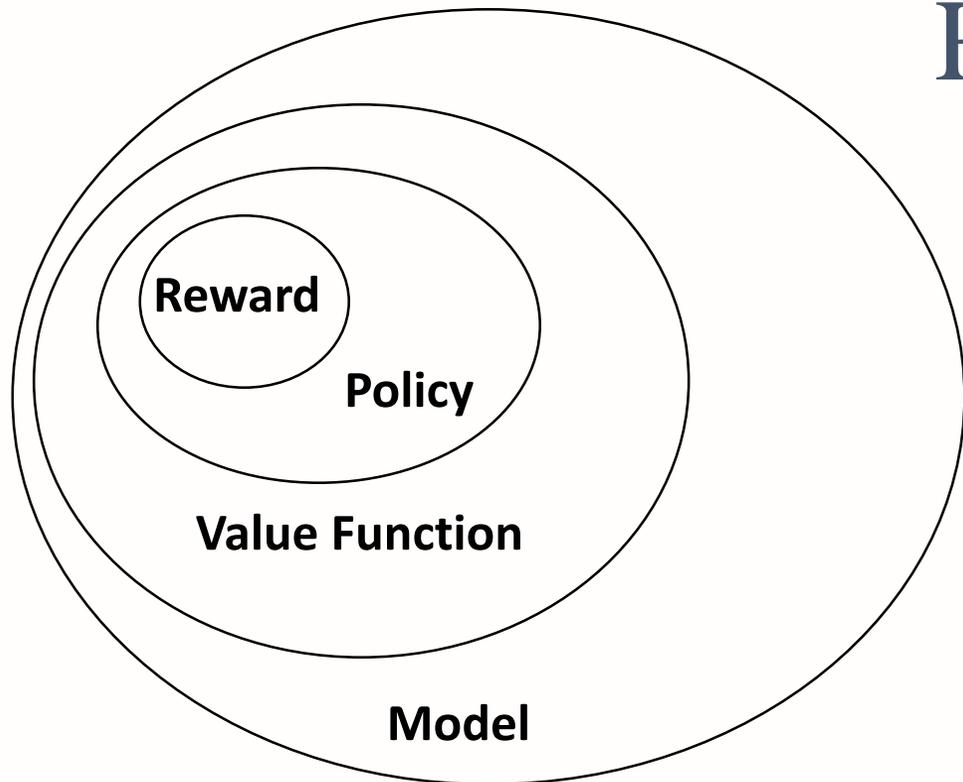


Reinforcement Learning

- Reinforcement Learning adalah salah satu paradigma baru di dalam learning theory.
- RL dibangun dari proses mapping (pemetaan) dari situasi yang ada di environment (states) ke bentuk aksi (behavior) agar dapat memaksimalkan reward.
- Agent yang bertindak sebagai sang learner tidak perlu diberitahukan behavior apakah yang akan sepatutnya dilakukan, atau dengan kata lain, biarlah sang learner belajar sendiri dari pengalamannya.
- Ketika ia melakukan sesuatu yang benar berdasarkan rule yang kita tentukan, ia akan mendapatkan reward, dan begitu juga sebaliknya.



Components of Reinforcement Learning



Reward: How good is this action?

Policy: what do I do now?

Value function: how good is this state?

Model: what happens if I do this action?

Policy

- Fungsi untuk membuat keputusan dari agent yang menspesifikasikan tindakan apakah yang mungkin dilakukan dalam berbagai situasi yang ia jumpai.
- Policy inilah yang bertugas memetakan perceived states ke dalam bentuk aksi.
- Policy bisa berupa fungsi sederhana, atau lookup table.
- Policy ini merupakan inti dari RL yang sangat menentukan behavior dari suatu agent.



Reward Function

- Reward function mendefinisikan tujuan dari kasus atau problem yang dihadapi.
- Ia mendefinisikan reward and punishment yang diterima agent saat ia berinteraksi dengan environment.
- Tujuan utama dari reward function ini adalah memaksimalkan total reward pada kurun waktu tertentu setelah agent itu berinteraksi.

Value Function

- Menspesifikasikan fungsi akumulasi dari total reward yang didapatkan oleh agent.
- Jika reward function berbicara pada masing-masing partial time dari proses interaksi, value function berbicara pada long-term dari proses interaksi

Model of Environment

- Model of environment adalah sesuatu yang menggambarkan behavior dari environment.
- Model of environment ini sangat berguna untuk mendesain dan merencanakan behavior yang tepat pada situasi mendatang yang memungkinkan sebelum agent sendiri mempunyai pengalaman dengan situasi itu.
- Saat masa-masa awal RL dikembangkan, model of environment yang ada berupa trial and error. Namun modern RL sekarang sudah mulai menjajaki spektrum dari low-level, trial and error menuju high-level, deliberative planning.



Kunci Keberhasilan RL

- Exploitation
- Exploration

Exploitation

- Seringkali manusia itu mengambil keputusan untuk melakukan sesuatu dengan berdasarkan pada informasi yang ia terima sebelumnya daripada perbuatan-perbuatan yang ia lakukan di masa lalu.
- Proses menggali sebanyak mungkin informasi tersebut dinamakan dengan exploitation.

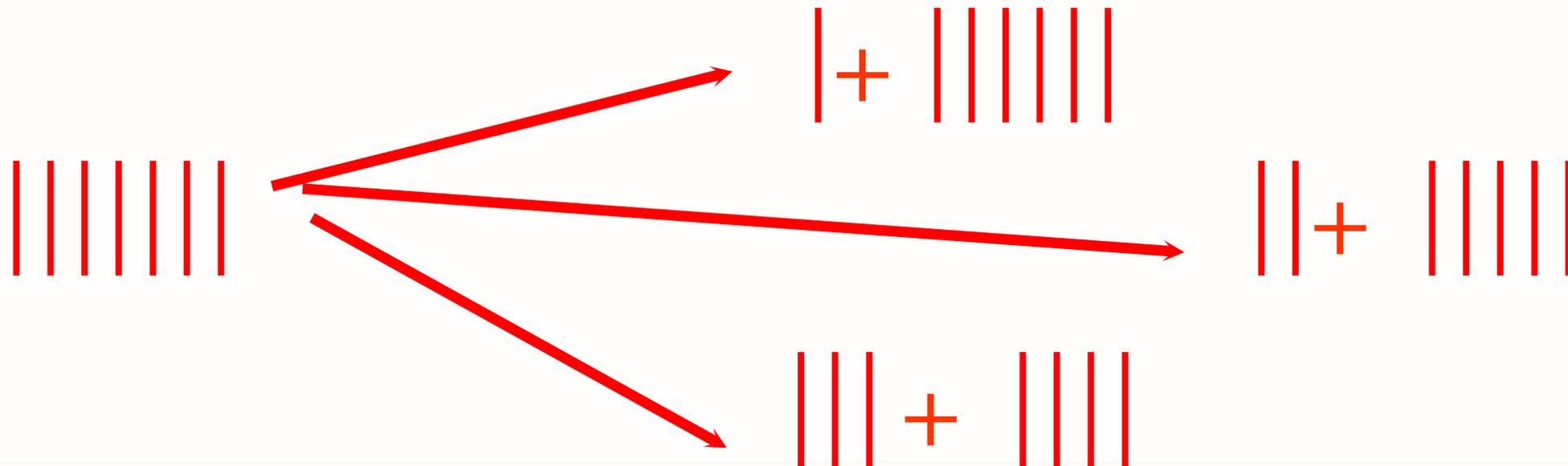
Exploration

- Seringkali juga manusia itu mengambil keputusan untuk melakukan sesuatu dengan tidak berdasarkan pada informasi yang ia terima sebelumnya daripada perbuatan-perbuatan yang ia lakukan di masa lalu.
- Akan tetapi lebih cenderung ia mencoba melakukan sesuatu yang memang benar-benar baru bagi dirinya untuk melihat bagaimana hasil daripada perbuatan tersebut.
- Proses inilah yang disebut dengan exploration.

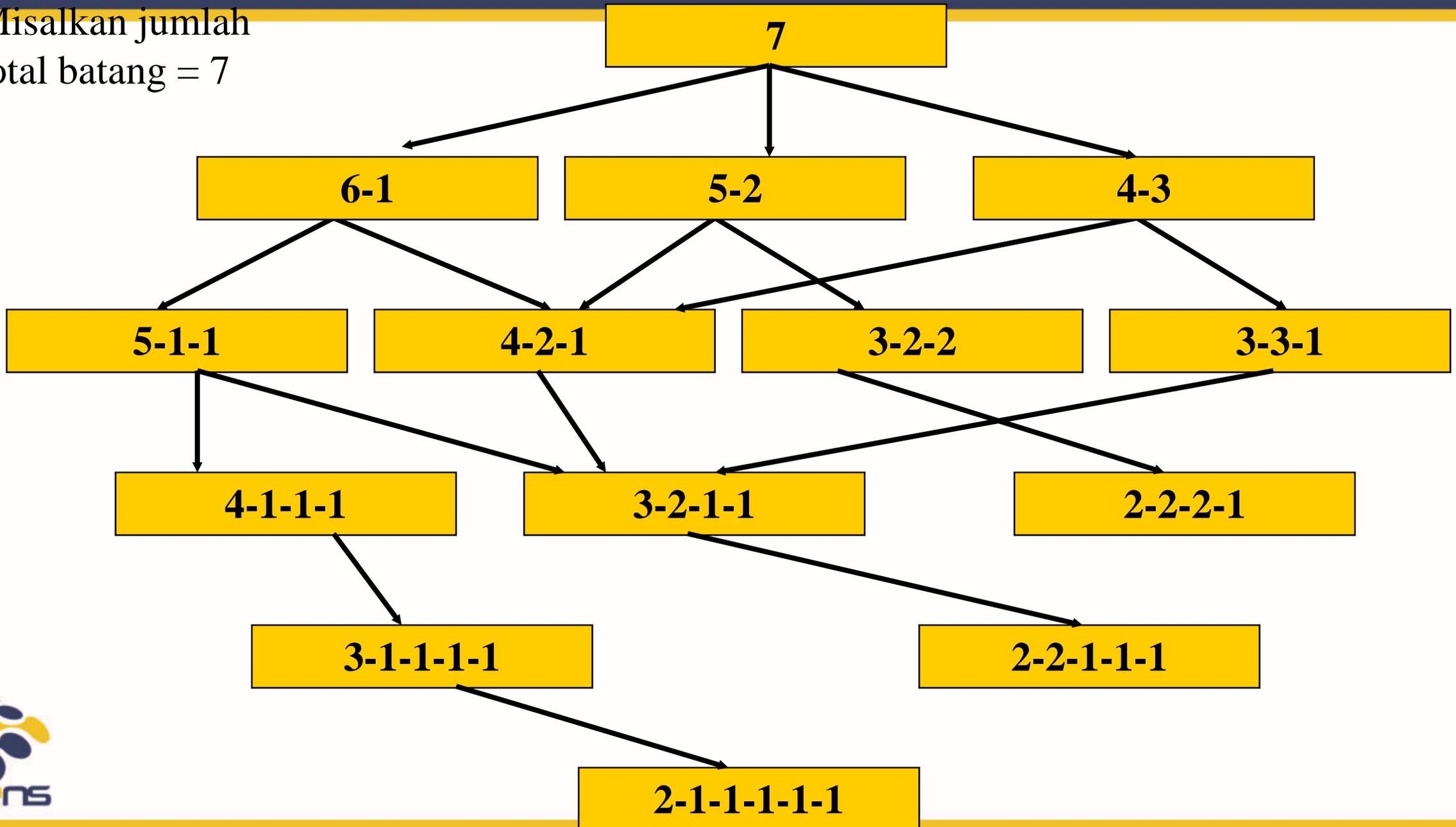


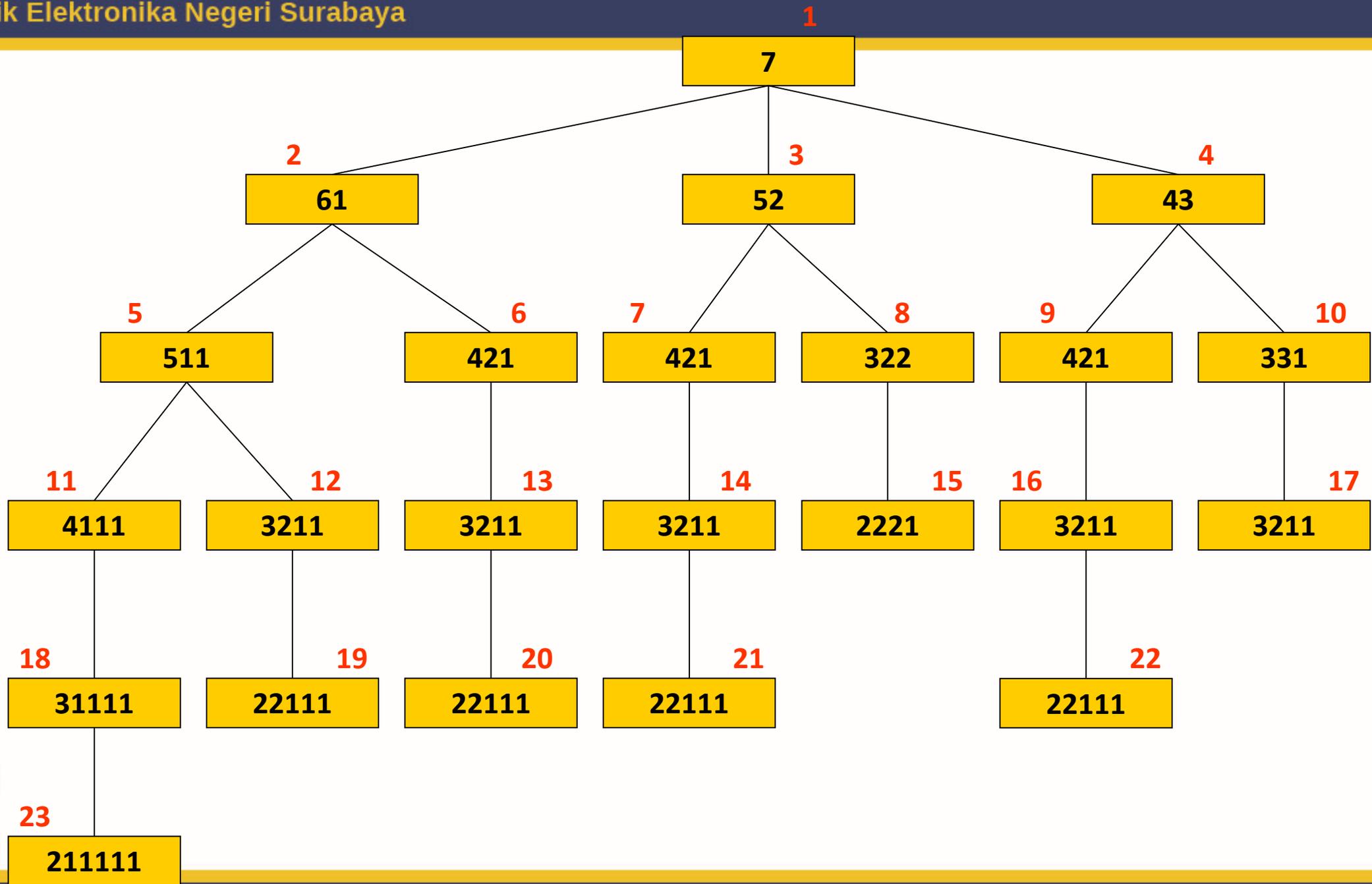
Deskripsi

- Diawali serangkaian batang
- Setiap pemain harus memecah serangkaian batang menjadi 2 kumpulan dimana jumlah batang di tiap kumpulan tidak boleh sama dan tidak boleh kosong



Misalkan jumlah total batang = 7



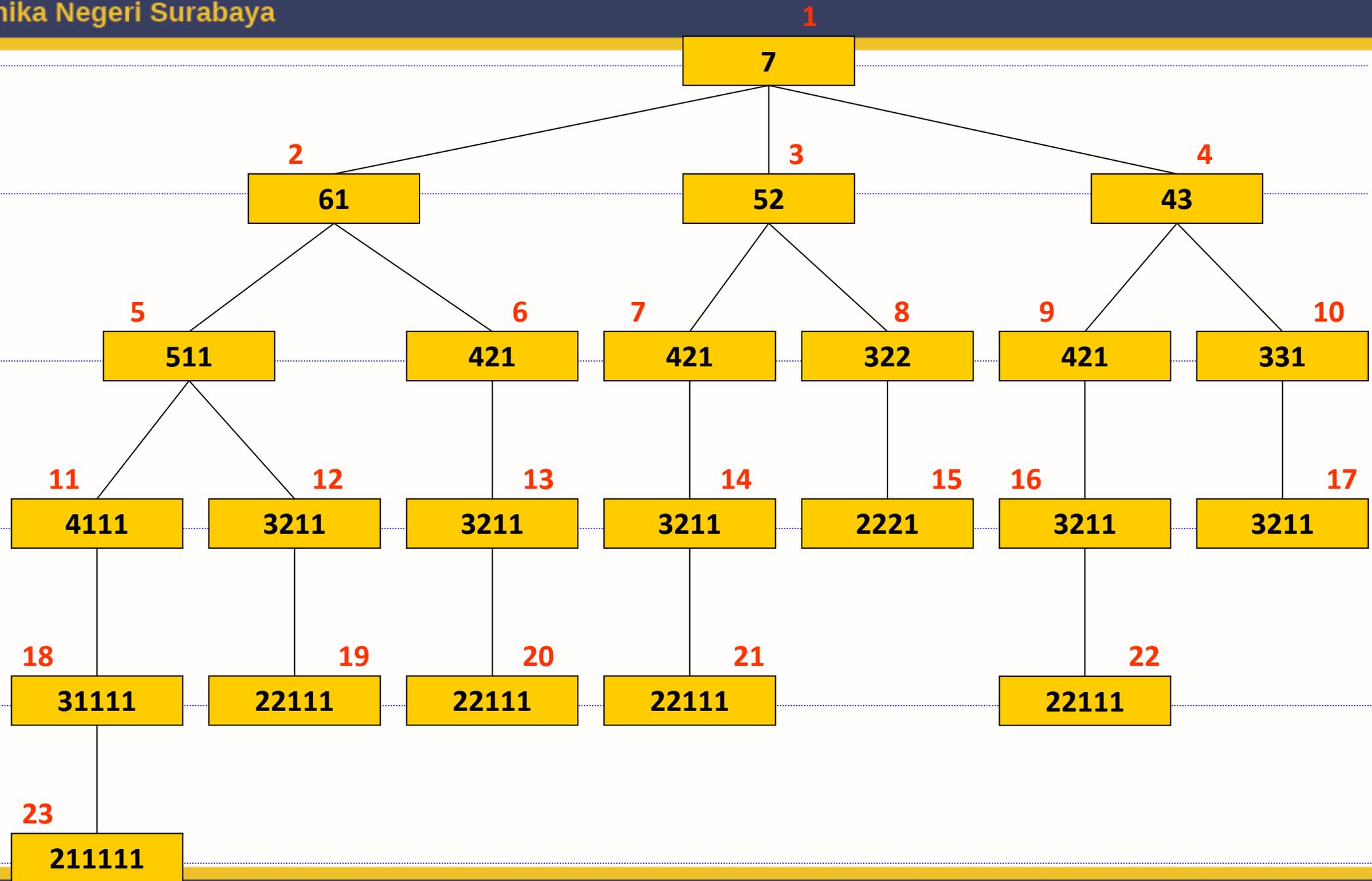


COMPUTER

LAWAN

COMPUTER

LAWAN

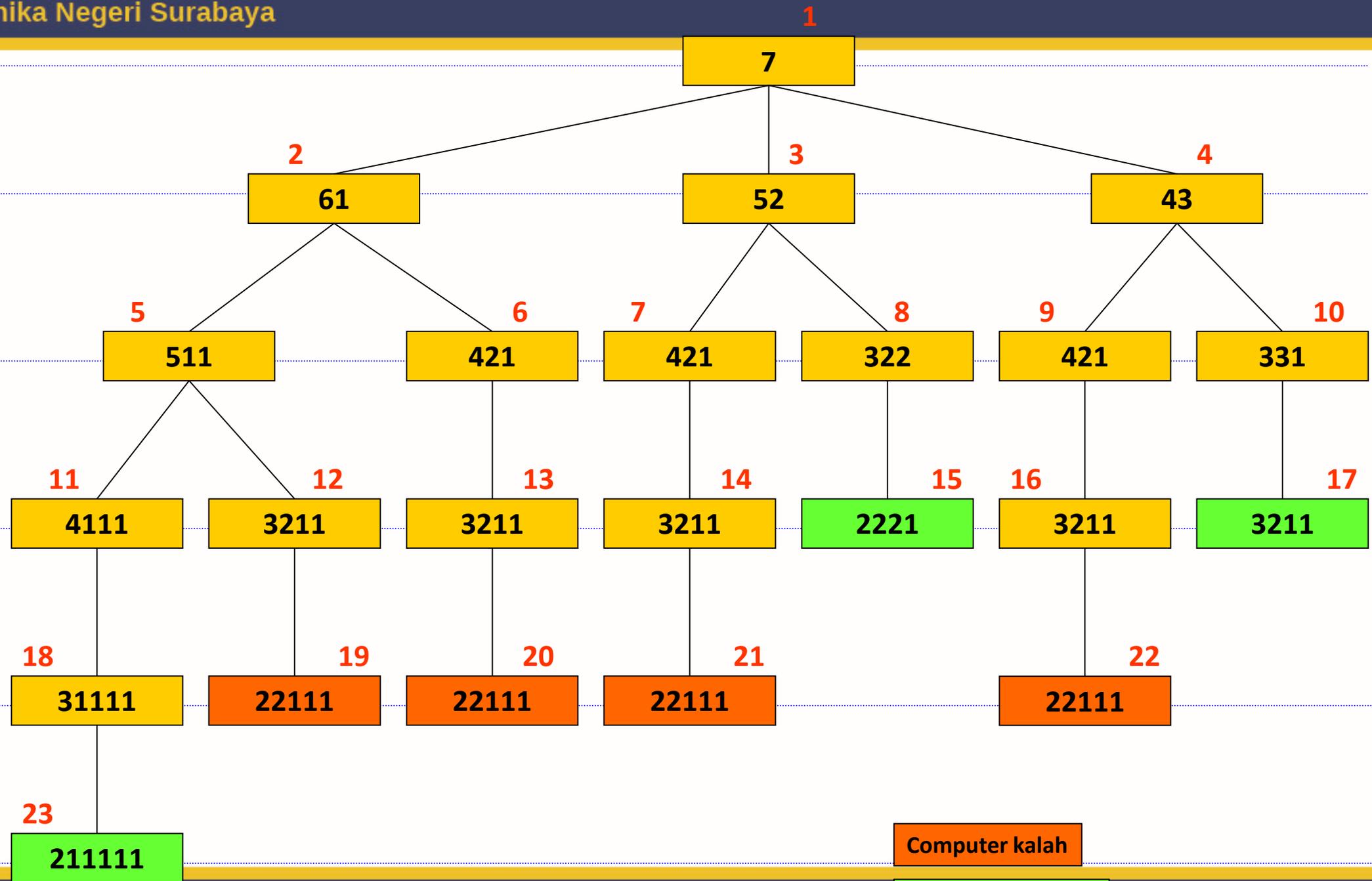


COMPUTER

LAWAN

COMPUTER

LAWAN

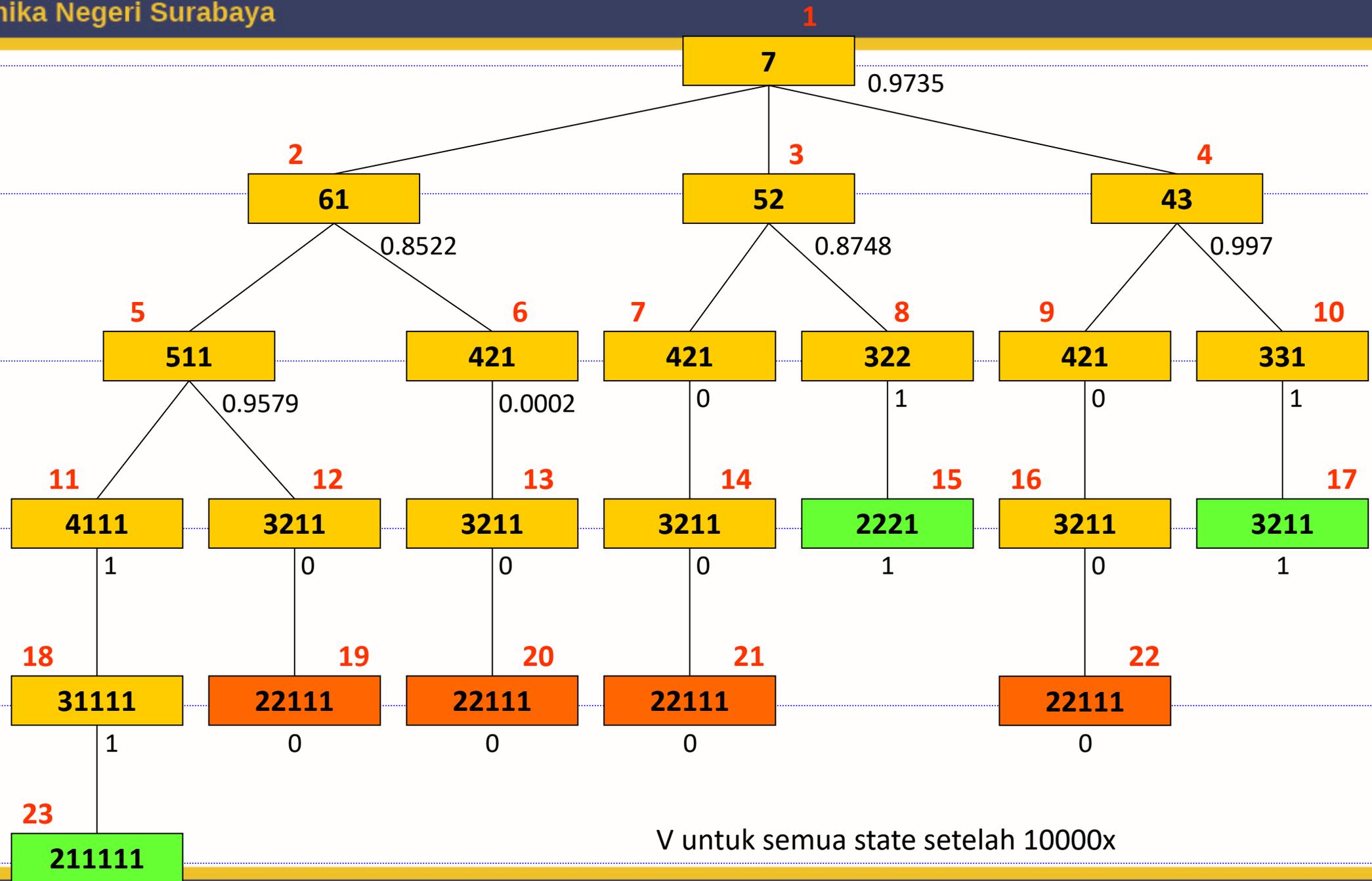


COMPUTER

LAWAN

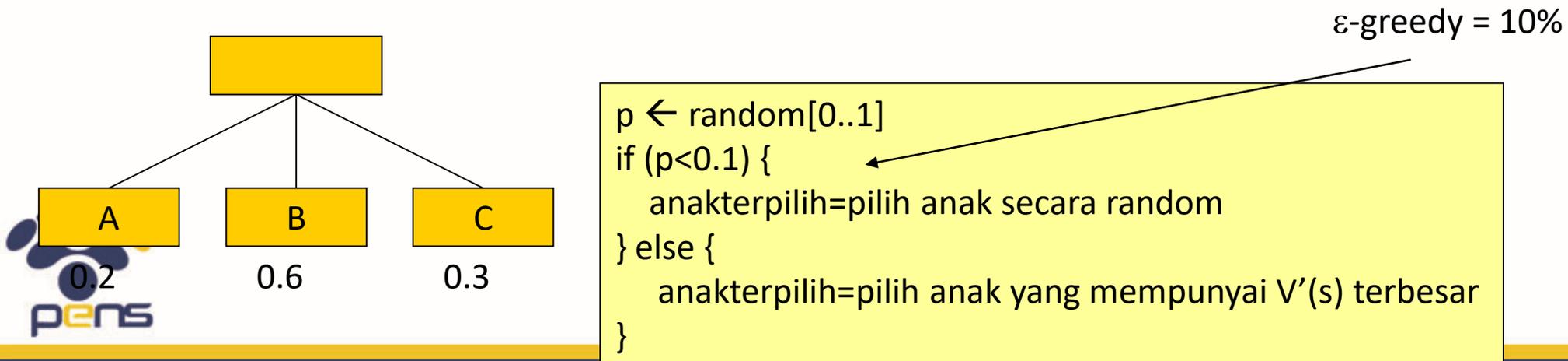
COMPUTER

LAWAN



Exploratory move

- Greedy move
 - memilih $V'(s)$ yang terbesar
- ϵ -greedy move
 - Kemungkinan besar memilih $V'(s)$ yang terbesar
 - Ada kemungkinan memilih yang lain secara random



Hasil eksperimen

State	10x	100x	1000x	10000x
1	0.5004	0.8906	0.9746	0.9735
2	0.5042	0.9478	0.7871	0.8522
3	0.5	0.5	0.9534	0.8748
4	0.5	0.5262	0.9163	0.997
5	0.5265	0.9948	0.9052	0.9579
6	0.5	0.4066	0.0994	0.0002
7	0.5	0.495	0.2745	0
8	0.5	0.5	1	1
9	0.5	0.495	0.4429	0
10	0.5	0.672	0.9983	1
11	0.6126	0.9991	1	1
12	0.5	0.405	0.0608	0
13	0.5	0.2152	0.0236	0
14	0.45	0.405	0.1029	0
15	1	1	1	1
16	0.5	0.405	0.2657	0
17	1	1	1	1
18	0.8063	0.9999	1	1
19	0	0	0	0
20	0	0	0	0
21	0	0	0	0
22	0	0	0	0
23	1	1	1	1



Latihan Soal

1. Buatlah Game Nim diselesaikan dengan Reinforcement Learning dengan jumlah batang 8!
2. Implementasikan dalam program!

Referensi

- Tom Michael, *Machine Learning*, McGraw-Hill publisher, 1997.
- Ali Ridho Barakbah, *Machine Learning*, Lecture Handout, Electronic Engineering Polytechnic Institute of Surabaya.

bridge to the future

<http://www.eepis-its.edu>

