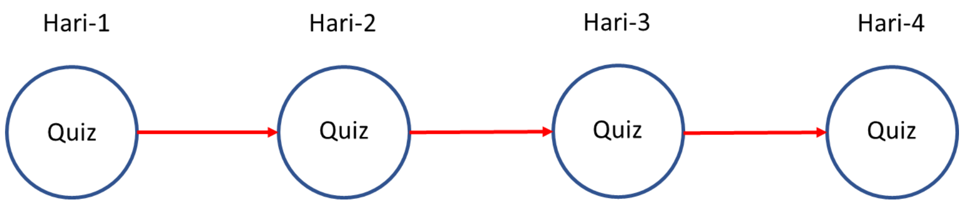
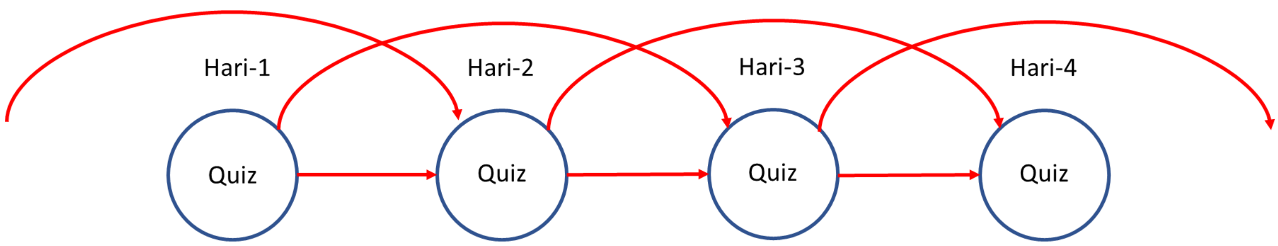
Tutorial Markov Chain

Apabila di contoh sebelumnya, Bayesian Network hanya menghitung probabilitas dari suatu waktu. Maka di tutorial ini, kita akan menghitung probabilitas sebuah variable dari suatu waktu ke waktu. Sebagai contoh, kita ingin menghitung probabilitas adanya quiz di hari ke 1, ke 2, ke 3 dan seterusnya seperti diperlihatkan dalam gambar berikut.



Probabilitas terjadinya quiz di hari ke 2 dipengaruhi oleh terjadinya quiz di hari ke 1. Oleh karena itu, apabila hari ke 1 terjadi quiz maka kemungkinan terjadinya quiz di hari ke 2 akan lebih tinggi. Kondisi seperti ini disebut **Markov Chain** karena kemungkinan terjadinya suatu variable itu dipengaruhi oleh satu/beberapa variable sebelumnya. Apabila hanya dipengaruhi satu variable sebelumnya (seperti gambar di atas), maka disebut **first-order Markov Chain**. Apabila dipengaruhi oleh dua variable, maka disebut **second-order Markov Chain** seperti terlihat pada gambar berikut.



Sekarang, pertanyaannya adalah bagaimana menghitung probabilitas di setiap waktu. Sebelum itu, kita perlu mengetahui probabilitas transisi dari setiap variable:

- $P(X_i|X_{i-1}) = ?$
  - X dalam hal ini adalah variable quiz di kondisi waktu i.
- $P(Quiz_i|Quiz_{i-1} = True) = 0.3$
  - Artinya apabila hari sebelumnya sudah ada quiz, maka hari ini kemungkinan ada quiznya sebesar 0.3 atau 30%
- $P(Quiz_i|Quiz_{i-1} = False) = 0.6$
  - Artinya apabila hari sebelumnya sudah ada tidak quiz, maka hari ini kemungkinan ada quiznya sebesar 0.6 atau 60%

- Probabilitas transisi di atas dapat diperlihatkan dalam bentuk table probabilitas

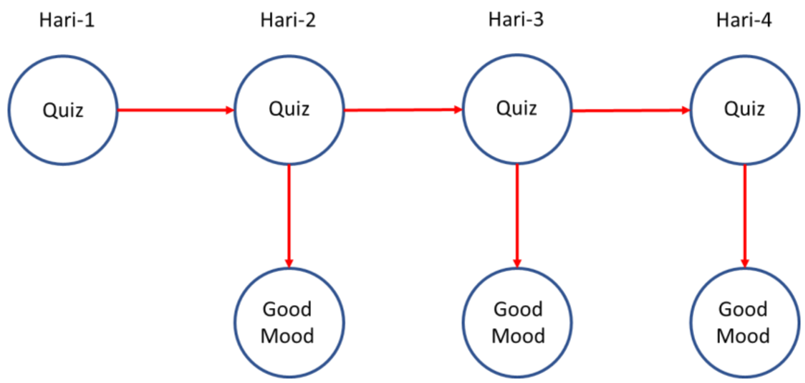
$Quiz_{i-1}$	$P(Quiz_i)$
T	0.3
F	0.6

- Selain itu dapat diperlihatkan sebagai matrix probabilitas transisi, $\begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix}$
  - Baris pertama dan kolom pertama merupakan bagian True
  - Baris kedua dan kolom kedua merupakan bagian False
  - Baris merupakan kondisi sebelumnya dan kolom merupakan kondisi setelahnya. Jadi apabila kemarin quiz, kemungkinan tidak ada quiz hari ini adalah 0.7 (70%)

Setelah mengetahui probabilitas transisi kita dapat mengaplikasikan untuk menghitung probabilitas quiz setiap harinya.

- Hari 0
  - $P(Quiz) = 0.7$
  - $P(\sim Quiz) = 1 - 0.7 = 0.3$
  - Probabilitas untuk quiz adalah 0.7 dan untuk tidak kuis adalah 0.3
- Hari 1
  - $P(Quiz_1|Quiz_0) = [0.7 \quad 0.3] \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix} = [0.39 \quad 0.61]$
- Hari 2
  - $P(Quiz_2|Quiz_1) = [0.39 \quad 0.61] \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix} = [0.483 \quad 0.517]$
- Hari 3
  - $P(Quiz_3|Quiz_2) = [0.483 \quad 0.517] \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix} = [0.4551 \quad 0.5449]$

Markov Chain dalam hal ini hanya menghitung probabilitas dalam kondisi variable sebelumnya. Namun dalam kehidupan nyata, kita juga menghitung probabilitas dari sensor atau observasi yang dilakukan. Sebagai contoh, kemungkinan terjadinya quiz, tidak hanya bergantung quiz atau tidaknya kemarin tetapi juga bergantung dari mood dosen pagi harinya. Apabila akan ada quiz, makan mood dosen paginya akan tidak baik. Kondisi seperti ini disebut Dynamic Bayesian Network, seperti tertera di gambar berikut:



Dengan asumsi probabilitas transisi sama seperti contoh di atas, kita memerlukan probabilitas kondisional (conditional probability) dengan good mood sebagai evidence/observation variable/conditionalnya, sebagai berikut:

- $P(E_i|X_i) = ?$
    - X dalam hal ini adalah variable quiz yang menjadi mempengaruhi probabilitas pada evidence E di kondisi waktu i.
- $P(GMood_i|Quiz_i = True) = 0.1$
    - Artinya apabila akan ada quiz, maka probabilitas mood dosen baik adalah 0.1 atau 10%
- $P(GMood_i|Quiz_i = False) = 0.7$
    - Artinya apabila tidak aka nada quiz, maka probabilitas mood dosen baik adalah 0.7 atau 70%

- Probabilitas kondisional di atas dapat diperlihatkan dalam bentuk table probabilitas

$Quiz_i$	$P(Mood_i)$
T	0.1
F	0.7

Setelah mengetahui probabilitas transisi dan probabilitas kondisionalnya kita dapat mengaplikasikan untuk menghitung probabilitas quiz setiap harinya. Hasil akhir probabilitas di setiap hari, harus dinormalisasi.

- Hari 0
    - $P(Quiz) = 0.7$
    - $P(\sim Quiz) = 1 - 0.7 = 0.3$
    - Probabilitas untuk quiz adalah 0.7 dan untuk tidak kuis adalah 0.3
- Hari 1
    - $P(Quiz_1|Quiz_0) = [0.7 \quad 0.3]\begin{bmatrix}0.3 & 0.7\\0.6 & 0.4\end{bmatrix} = [0.39 \quad 0.61]$
    - $P(Quiz_1|GMood_1) = \alpha P(GMood_1|Quiz_1)P(Quiz_1|Quiz_0)$
      $= \alpha[0.1 \quad 0.7][0.39 \quad 0.61]$
      $= \alpha[0.039 \quad 0.427] \approx [0.084 \quad 0.916]$
- Hari 2
    - $P(Quiz_2|Quiz_1) = [0.084 \quad 0.916]\begin{bmatrix}0.3 & 0.7\\0.6 & 0.4\end{bmatrix} = [0.5748 \quad 0.4252]$
    - $P(Quiz_2|GMood_2) = \alpha P(GMood_2|Quiz_2)P(Quiz_2|Quiz_1)$
      $= \alpha[0.1 \quad 0.7][0.5748 \quad 0.4252]$
      $= \alpha[0.05748 \quad 0.29764] \approx [0.162 \quad 0.838]$
- Hari 3
    - $P(Quiz_3|Quiz_2) = [0.162 \quad 0.838]\begin{bmatrix}0.3 & 0.7\\0.6 & 0.4\end{bmatrix} = [0.5514 \quad 0.4486]$
    - $P(Quiz_3|GMood_3) = \alpha P(GMood_3|Quiz_3)P(Quiz_3|Quiz_2)$
      $= \alpha[0.1 \quad 0.7][0.5514 \quad 0.4486]$
      $= \alpha[0.05514 \quad 0.31402] \approx [0.149 \quad 0.851]$