

Semi-Supervised Learning



Table of Content

What will We Learn Today?

1. Introduction to Semi-supervised
2. Label Propagation
3. Label Spreading
4. Hands-On



Introduction to Semi Supervised



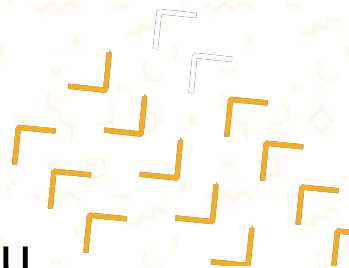


Machine Learning

Ketika membangun model supervised learning, kita perlu melakukan training terhadap data yang kita miliki.

Namun, kadang kala data yang memiliki label sangat sedikit. Biasanya karena label yang perlu diberikan secara manual sehingga terkendala akan keterbatasan waktu dan biaya.

Karena keterbatasan tersebut semi-supervised learning dapat membantu proses pemodelan dengan lebih baik.

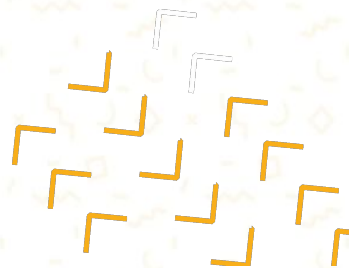




Semi-supervised Learning

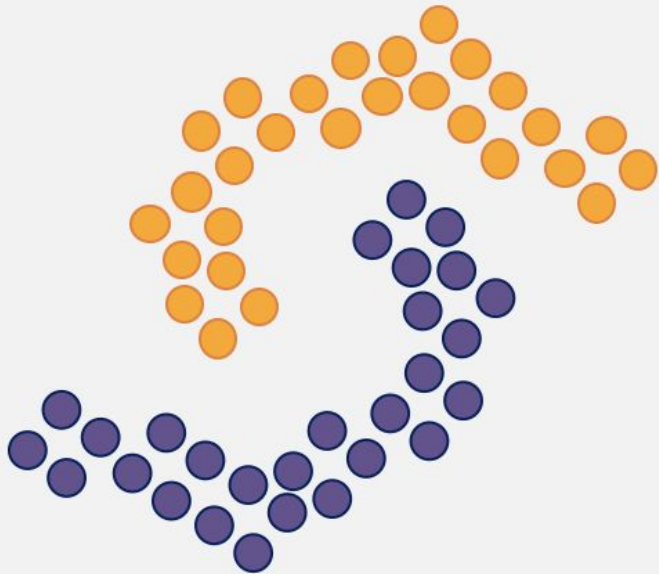
Semi-supervised learning digunakan ketika data training yang kita miliki sangat terbatas (sedikit) dan memiliki mayoritas data unlabeled. Dengan harapan model ini mampu belajar lebih baik dibandingkan dengan supervised learning

Konsep yang dimiliki oleh semi-supervised learning adalah menggabungkan antara supervised learning dan unsupervised learning.

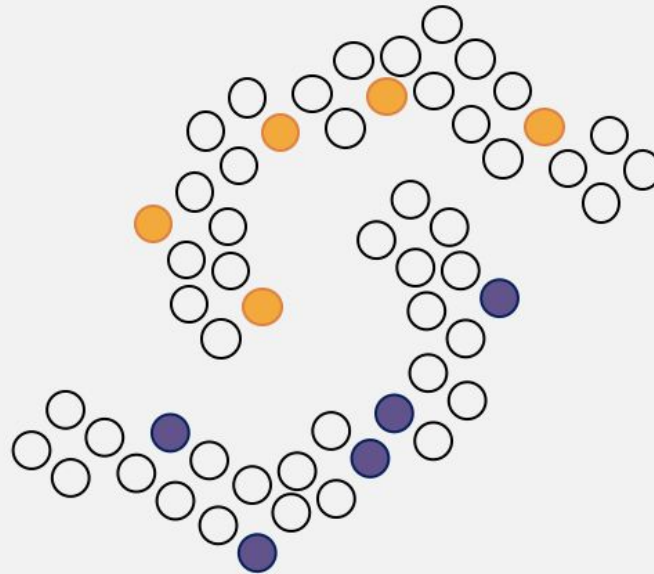




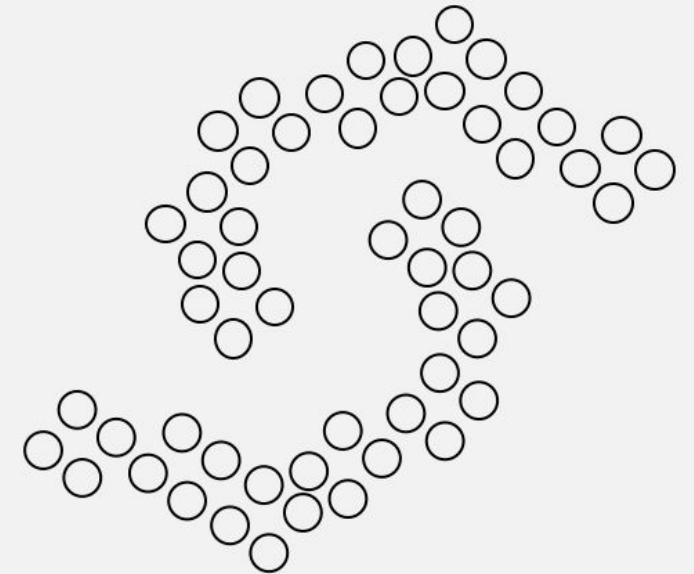
Semi-supervised Learning



Supervised Learning



Semi-Supervised Learning



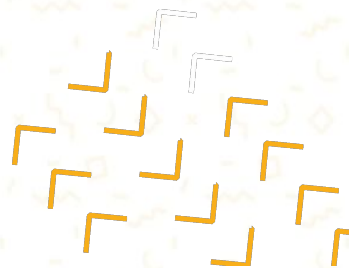
Unsupervised Learning



Semi-supervised Learning

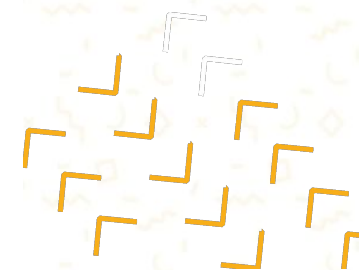
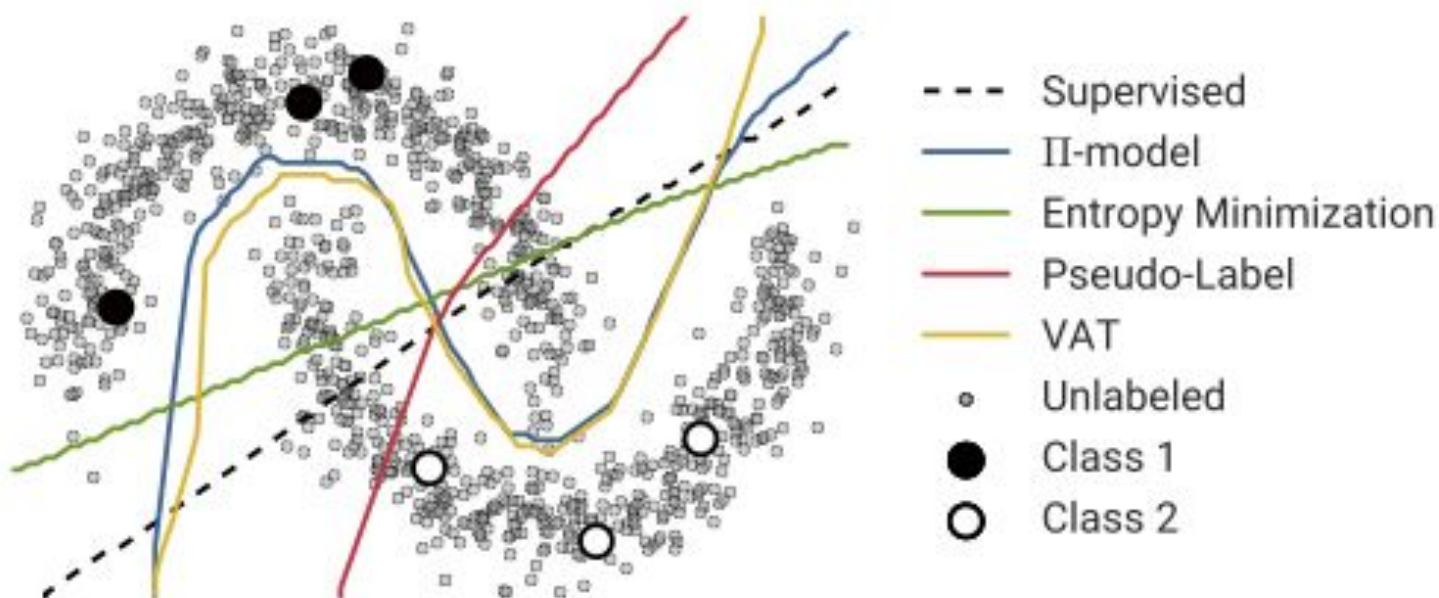
Dengan labeled dataset yang sedikit jika kita menggunakan supervised learning. Akan menghasilkan model yang overfitting.

Berbeda halnya dengan supervised learning. Selain menggunakan labeled data, Semi-supervised juga memanfaatkan unlabeled data. Yang mana berguna untuk menambahkan informasi terhadap model berdasarkan distribusi yang dimiliki.





Semi-supervised Learning



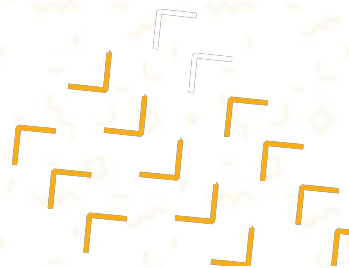


Semi-supervised Learning

Metode yang akan dipelajari untuk Semi-supervised learning ada 2:

1. Label Propagation
2. Label Spreading

*Karena mudah diimplementasi pada module scikit-learn



Label Propagation

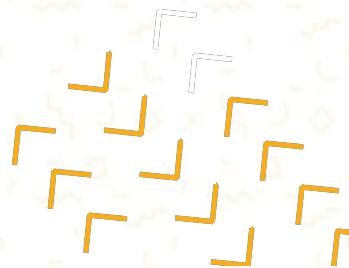




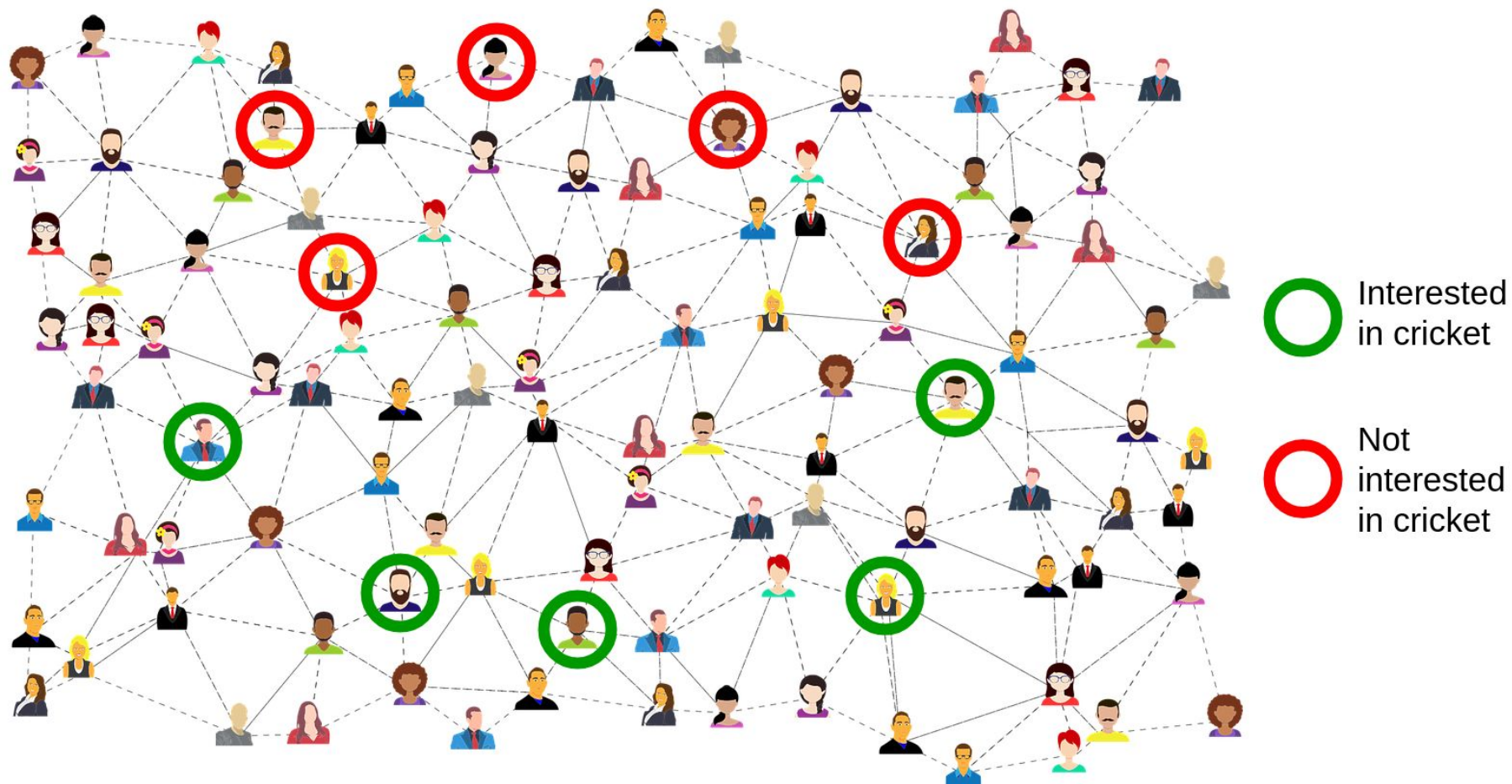
Label Propagation

Label propagation merupakan salah satu metode semi-supervised learning.

Dengan menggunakan konsep iterative machine learning yang memanfaatkan metode berbasis graph.



Label Propagation

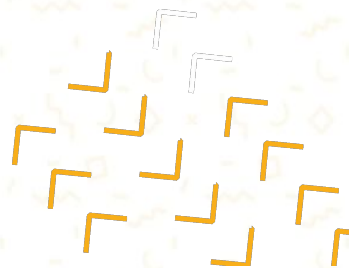




Semi-supervised Learning

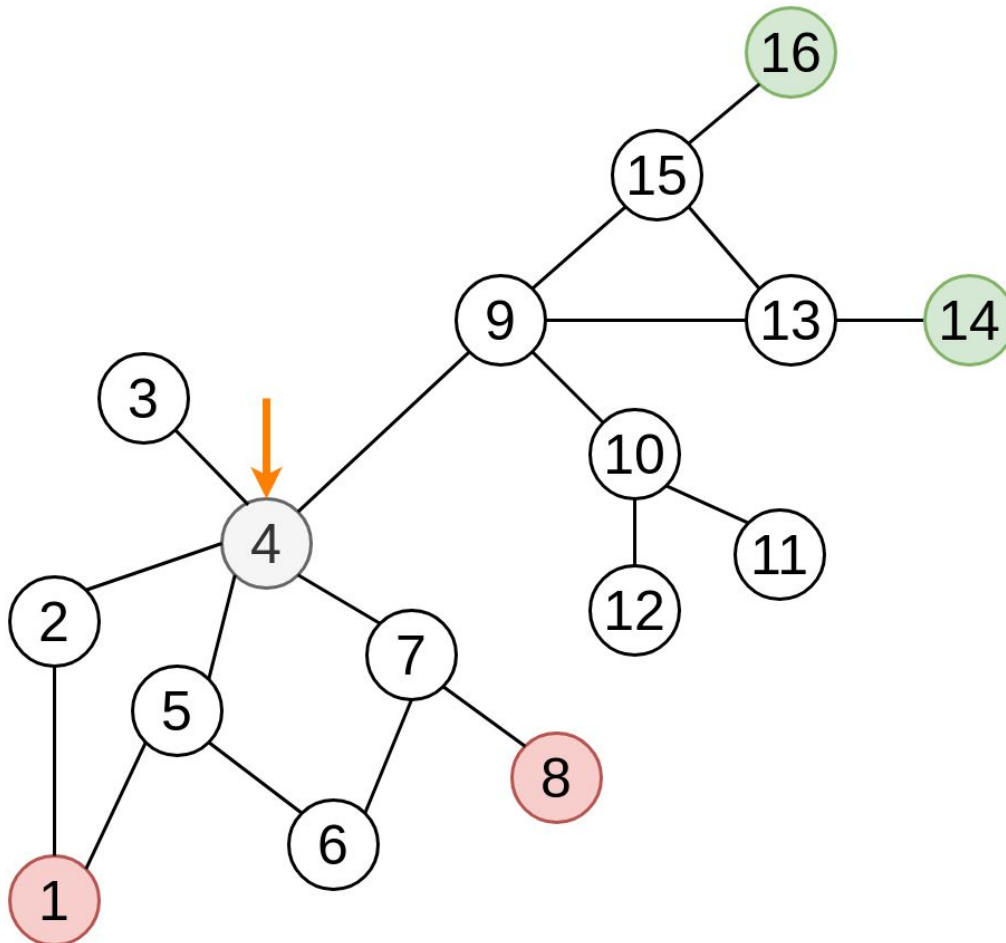
Dalam kasus Label Propagation kita mengasumsikan setiap edges antara node merupakan sebuah ***similarity***.

Secara random perlahan kita akan memprediksi unlabeled data berdasarkan jarak terdekat.

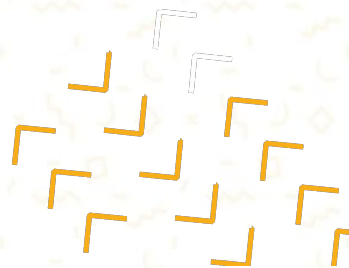




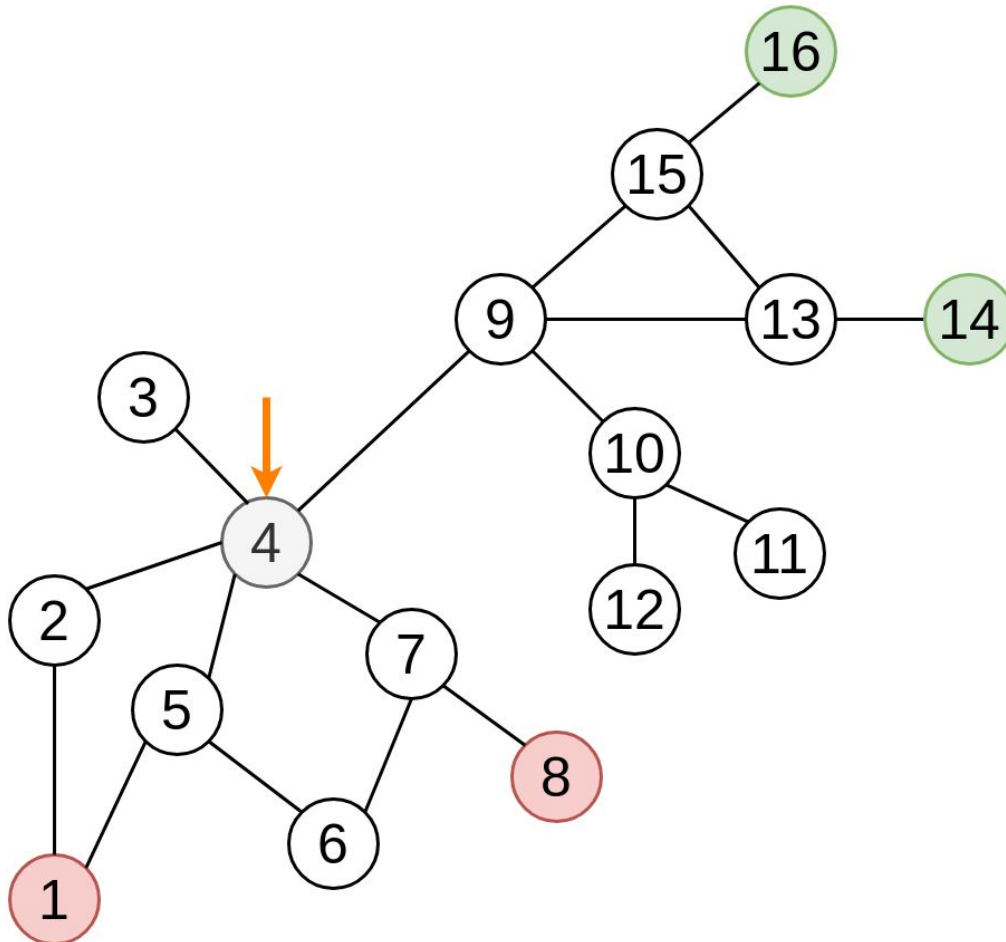
Semi-supervised Learning



Sebutkan semua langkah yang memungkinkan untuk menuju node hijau atau node merah!



Green

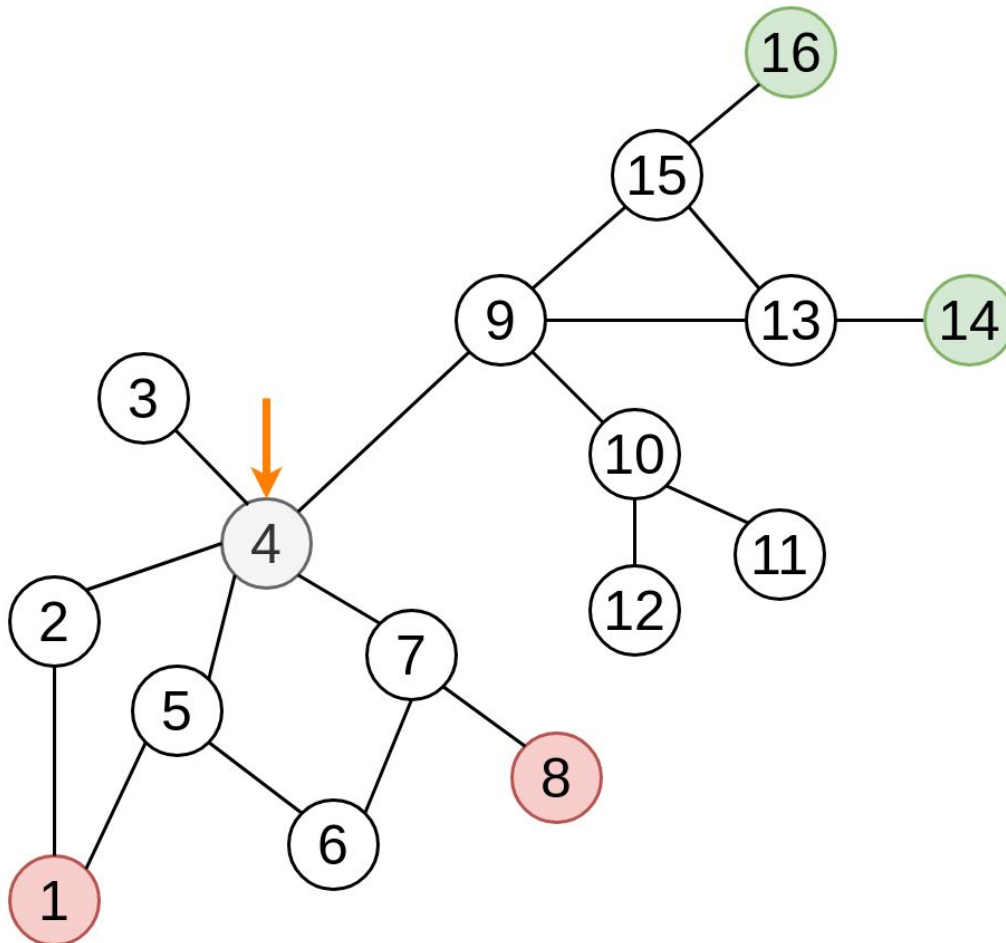


Red

$$4. 4 \rightarrow 9 \rightarrow 15 \rightarrow 13 \rightarrow 14$$
$$5. 4 \rightarrow 2 \rightarrow 1$$

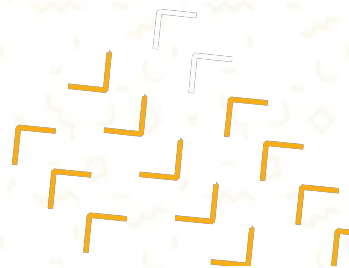


Semi-supervised Learning



Karena hasil yang diberikan mayoritas dari langkah berakhir pada warna merah. Maka kita katakan node 4 berlabel merah.

Inilah intuisi dari label propagation.





Semi-supervised Learning





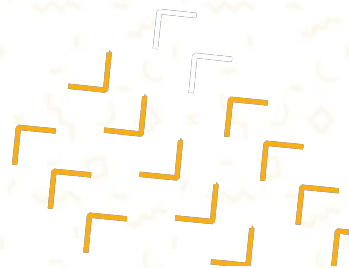
Semi-supervised Learning

$$y_i[c] = \sum_{j \in X_l} T_{ij}^t y_j[c]$$

The probability of node $x_i \in X_u$ to have label c

The probability $P(i \rightarrow j)$ to jump from node x_i and end up in node x_j in t steps. We can define the number of steps to be a large number (infinity).

$Y_i [c]$: peluang baris ke i dari unlabeled X untuk label c





Semi-supervised Learning

Jika kita bisa mendapatkan matrix dari T (probabilistic transition matrix) maka kita bisa mendapatkan semua prediksi untuk unlabeled data

$$\hat{Y} = T^{t \rightarrow \infty} Y$$

Diagram illustrating the equation $\hat{Y} = T^{t \rightarrow \infty} Y$ with arrows pointing to the components:

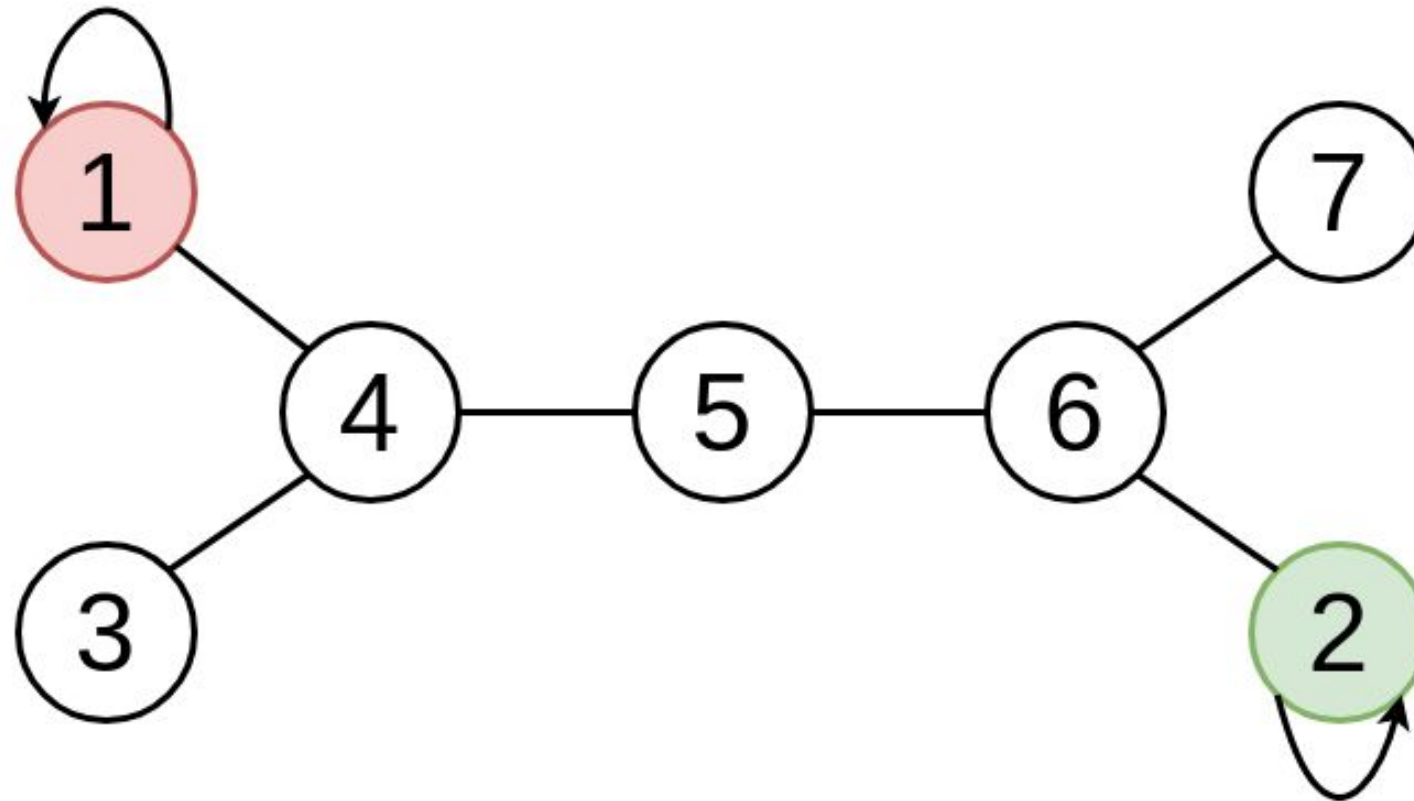
- \hat{Y} : Vector of labels we need to get
- $T^{t \rightarrow \infty}$: Matrix of probabilities to end up on different nodes
- Y : Vector of labels on those nodes

$$\begin{bmatrix} \hat{Y}_l \\ \hat{Y}_u \end{bmatrix} = T^{t \rightarrow \infty} \begin{bmatrix} Y_l \\ 0 \end{bmatrix}$$

\hat{Y}_u is what we are interested in.

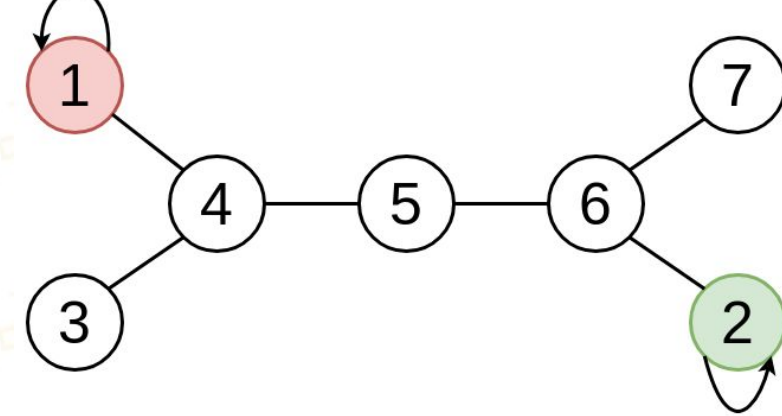


Semi-supervised Learning





Semi-supervised Learning



$$T = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0.33 & 0 & 0.33 & 0 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0.33 & 0 & 0 & 0.33 & 0 & 0.33 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{matrix}$$



Semi-supervised Learning

Tuu: Adalah matrix
peluang dari unlabeled
data to unlabeled data

Tul: Adalah matrix
peluang dari unlabeled
data to labeled data

$$\begin{bmatrix} \hat{Y}_l \\ \hat{Y}_u \end{bmatrix} = T^{t \rightarrow \infty} \begin{bmatrix} Y_l \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \hat{Y}_l \\ \hat{Y}_u \end{bmatrix} = \begin{bmatrix} I & 0 \\ (I - T_{uu})^{-1} \cdot T_{ul} & 0 \end{bmatrix} \begin{bmatrix} Y_l \\ 0 \end{bmatrix}$$

Label vectors of labelled nodes: $\hat{Y}_l = Y_l$

Label vectors of unlabelled nodes: $\hat{Y}_u = (I - T_{uu})^{-1} \cdot T_{ul} \cdot Y_l$

The background is a light cream color with a repeating pattern of small, faint geometric shapes like diamonds and squares. Large, flowing organic shapes in orange and blue frame the central text. Various geometric icons are scattered around: a blue semi-circle with white 'x' marks in the top left; a blue circle with white dots in the bottom left; a blue semi-circle with an orange segment in the bottom right; a cluster of orange 'L' shapes in the bottom right; a blue circle with a white dot in the middle right; and a blue stylized 'D' shape with an orange bar in the top right. A large orange 'C' shape with a blue 'X' is at the top center.

Hands-On

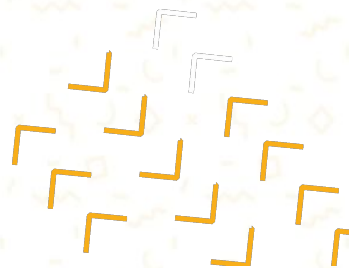
Label Spreading





Label Spreading

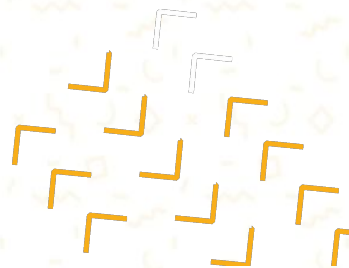
Label spreading adalah metode yang serupa dengan label propagation. Menggunakan graph dan matrix T untuk memprediksi unlabeled data.





Label Spreading

Berbeda dengan label propagation yang menggunakan matrix similarity tanpa modifikasi. Label spreading menggunakan regularization agar model yang dihasilkan lebih robust.

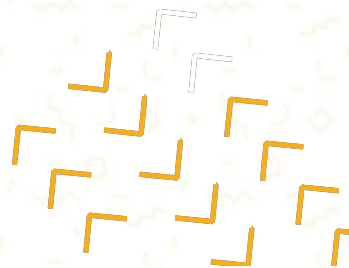




Label Spreading

Pengembangan metode yang digunakan:

1. Iterasi digunakan lebih efisien dari original graph
2. Menormalisasikan bobot dari jarak dengan menghitung Laplacian Matrix dengan normalized graph_{ix}.



The background is a light cream color with a repeating pattern of small, faint geometric shapes like diamonds and squares. Large, flowing organic shapes in orange and blue frame the central text. Various geometric icons are scattered around: a blue semi-circle with white 'x' marks in the top left; a blue circle with a white dot in the middle left; a cluster of white dots in the bottom left; a blue semi-circle with a white dot in the bottom right; and a cluster of white 'L' shapes in the bottom right. In the top center, there is a large orange 'C' shape and a blue 'X' shape, both with a grid of small blue dots around them. In the top right corner, there is a stylized logo consisting of a blue 'U' shape and an orange bar.

Hands-On

**Thank
YOU**

