
IMPLEMENTASI ALGORITMA LINK PREDICTION UNTUK MENCARI KESAMAAN ANTARA CALON LEGISLATIF PEMILIHAN UMUM INDONESIA 2019

Ghiffari Assamar Qandi

Fakultas Teknologi Elektro dan Informatika Cerdas, Departemen Sistem Informasi
Institut Teknologi Sepuluh Nopember
Email: ghiffari.18052@mhs.its.ac.id

Nur Aini Rakhmawati

Fakultas Teknologi Elektro dan Informatika Cerdas, Departemen Sistem Informasi
Institut Teknologi Sepuluh Nopember
Email: nur.aini@is.its.ac.id

ABSTRAK

Jumlah penelitian yang menganalisa data semakin meningkat. Salah satu topik yang populer adalah mencari interaksi antara masing-masing data yang biasa disebut sebagai *social network* atau jejaring sosial. Metode yang biasa digunakan untuk pemodelan dari masing-masing data adalah *link prediction*. Dengan menggunakan *link prediction* akan mendapatkan hasil berupa nilai kesamaan atau hubungan antara masing-masing data. Pemilihan umum Indonesia adalah pesta demokrasi rakyat yang diadakan setiap 5 tahun sekali. Salah satu data yang bisa diambil dalam pemilihan umum adalah data calon legislatif. Penelitian ini mencoba mengimplementasikan *link prediction* ke dalam data calon legislatif pemilihan umum Indonesia 2019. Peneliti menggunakan teknik *crawling* untuk mendapatkan sumber data dari website resmi Komisi Pemilihan Umum. Setelah proses *crawling* data akan dibersihkan sesuai atribut terpilih. Atribut yang digunakan untuk mencari nilai kesamaan adalah daerah asal, daerah pemilihan, dan partai pengusung. Metode *link prediction* yang digunakan adalah *Adamic-Adar Index*, *Jaccard Coefficient*, dan *Preferential Attachment*. Hasil penelitian ini menemukan bahwa jumlah kesamaan berdasarkan atribut terpilih masing-masing calon legislatif cukup besar.

Kata kunci: jejaring sosial, link prediction, machine learning, jaccard, preferential attachment, adamic-adar index

ABSTRACT

Researches on data analysis has attracted much in recent years. One of popular topic is finding interaction between data called as Social Network. The method commonly used for modeling of each data is link prediction. By using the link prediction will get results in the form of similarity values or the relationship between each data. Indonesia's general election is a party for people's democracy which is held once every 5 years. One of the data that can be taken from the general election is legislative candidate data. This research attempts to implement a prediction link into the 2019 Indonesian general election legislative candidate data. The researcher uses crawling techniques to obtain data sources from the official Election Commission's official website. After crawling the data will be preprocessed according to the selected attribute and transformed into graph. The attributes used to look for similarity values are the area of origin, electoral district, and political party. The link prediction methods used are Adamic-Adar Index, Jaccard Coefficient, and Preferential Attachment. The results of this study found that the number of similarities based on the chosen attributes of each legislative candidate was quite large.

Keywords: *social network, link prediction, machine learning, jaccard, preferential attachment, adamic-adar index*

1. PENDAHULUAN

Seiring dengan berkembangnya teknologi, jumlah data yang disimpan dalam komunitas, organisasi atau perusahaan semakin meningkat. Oleh karena itu proses pengolahan data dalam jumlah yang besar sudah tidak efektif lagi jika dilakukan oleh manusia, sangat memungkinkan untuk dialihkan ke mesin. Agar sebuah mesin mampu bertindak seperti manusia maka diberikan sebuah Kecerdasan Buatan kedalamnya [1]. Machine Learning (ML) merupakan salah satu cabang topik dari Kecerdasan Buatan. ML adalah sebuah algoritma dan model statistik yang disematkan ke dalam sistem computer untuk melakukan tugas tertentu layaknya manusia tanpa menggunakan instruksi eksplisit dengan mengandalkan pola dan inferensi yang terjadi [1], [2]. Salah satu analisa data yang populer didalam machine learning adalah *Social Network*.

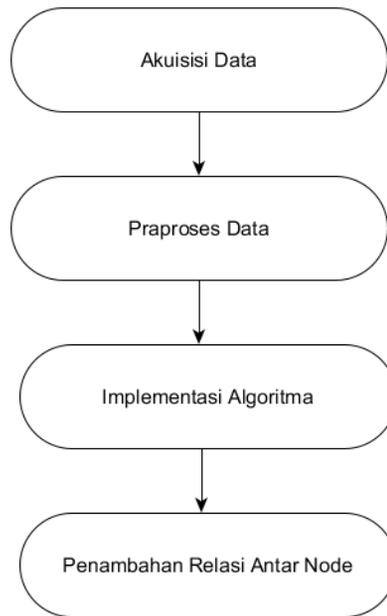
Social Network atau jejaring sosial adalah cara populer untuk memodelkan interaksi di antara orang-orang dalam suatu kelompok atau komunitas. Mereka dapat divisualisasikan sebagai *graph*, dimana titik divisualisasikan orang di dalam sebuah kelompok dan dihubungkan dengan sebuah garis yang mewakili bentuk hubungan antara kedua orang tersebut. Jejaring sosial merupakan objek yang sangat dinamis karena dari *node* mampu ditambahkan dari waktu ke waktu sehingga relasi antara *node* yang sudah ada bisa berubah [3]. Untuk melakukan pemodelan tersebut diperlukan metode yang mampu memprediksi hubungan antara *node* berdasarkan atribut yang ditentukan. Permasalahan yang biasanya terjadi adalah ketika terjadinya penambahan data maka akan mempengaruhi relasi *node* yang sudah terbentuk. Oleh karena itu diperlukan algoritma *link prediction*. *Link prediction* adalah sebuah teknik yang digunakan untuk memprediksi hubungan antara *node* di dalam *network* agar mampu beradaptasi jika ada penambahan data berdasarkan atribut informasi yang sudah ada [4].

Pada penelitian sebelumnya Chen, 2005 melakukan pendekatan *link prediction* kedalam metode *collaborative filtering*. Penelitian ini mengusulkan metode *link prediction* untuk menangani permasalahan yang ada pada sistem rekomendasi berdasarkan kesamaan antara ulasan yang dilakukan oleh pengguna [5]. Selanjutnya Raf dan Ronald, 2014 melakukan implementasi *link prediction* untuk membuat sistem rekomendasi rekan penulis artikel ilmiah. Penelitian ini mengusulkan penggunaan metode *link prediction* kedalam sistem rekomendasi yang sudah ada berdasarkan kedekatan penulis dan topik penelitian [6].

Penelitian ini menggunakan data set calon legislatif Pemilihan Umum Indonesia 2019. Nilai kesamaan antara masing-masing calon legislatif akan ditentukan berdasarkan daerah pemilihan, asal calon, dan partai pengusung. Untuk proses akuisisi data set calon legislatif akan menggunakan metode *crawling*. *Crawling* adalah teknik ekstraksi data secara otomatis dari sebuah website [7]. Sejauh pengetahuan penulis, di Indonesia belum ada yang melakukan penerapan *link prediction* ke dalam prediksi hubungan atau nilai kedekatan antara individual.

2. METODOLOGI PENELITIAN

Untuk melakukan implementasi *link prediction* kedalam data set calon legislatif diperlukan sebuah metode penelitian. Berikut merupakan metode penelitian yang digunakan dalam penelitian ini ditunjukkan oleh Gambar 1.



Gambar 1. Langkah-Langkah Penelitian

2.1 Akuisisi Data

Pada tahap ini akan melakukan proses pengumpulan data. Data set akan di *crawling* dari website resmi Komisi Pemilihan Umum Indonesia menggunakan bahasa pemrograman Python. Hasil *crawling* tersebut akan diubah ke dalam format .csv untuk proses implementasi nanti. Pada tahap ini penulis memperoleh 36805 data calon legislatif. Contoh data set ditunjukkan oleh Tabel 1.

Tabel 1. Contoh Data Set Hasil *Crawling*.

Dapil	Partai	Nama	Jenis Kelamin	Lokasi	Tahun	Agama	Status	Pendidikan
Jawa Tengah	Partai Amanat Nasional	Siti Chasanah	Perempuan	Cilacap	1965	Islam	Menikah	SMA
Kalimantan Barat	Partai Demokrat	Andreas Paulus	Laki-Laki	Pontianak	1971	Katolik	Menikah	D4/S1
Papua Barat	Partai Bulan Bintang	Ayub Msiren	Laki-Laki	Manokwari	1958	Kristen Protestan	Menikah	D4/S1
Sulawesi Tenggara	Partai Nasdem	Waode Siti Armini Rere	Perempuan	Jakarta Selatan	1960	Islam	Menikah	S2
Riau	Partai Berkarya	Intan Puspita	Perempuan	Karimun	1979	Islam	Menikah	SMA

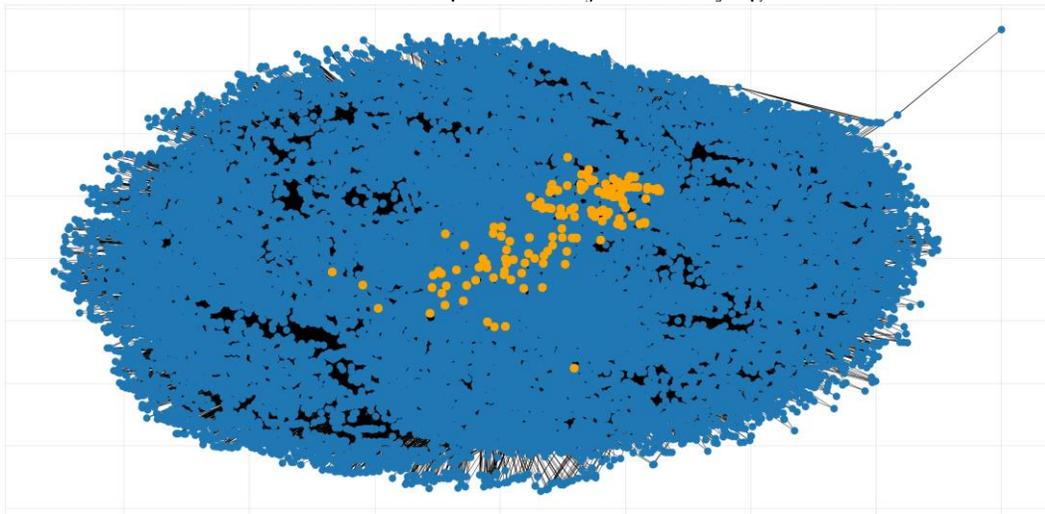
2.2 Praproses Data

Pada tahap ini akan melakukan proses pembersihan data. Data set hasil crawling akan diolah untuk disesuaikan data nya berdasarkan atribut yang ingin dicari kesamaannya. Atribut tersebut adalah daerah pemilihan, partai pengusung, nama calon legislatif dan daerah asal. Selain itu juga pada tahap ini memastikan bahwa data yang dimiliki tidak memiliki nilai kosong pada atribut yang telah ditentukan. Hasil pengolahan data ditunjukkan oleh Tabel 2.

Tabel.2. Contoh Hasil Praproses Data

Dapil	Partai	Nama	Lokasi
Jawa Tengah	Partai Amanat Nasional	Siti Chasanah	Cilacap
Kalimantan Barat	Partai Demokrat	Andreas Paulus	Pontianak
Papua Barat	Partai Bulan Bintang	Ayub Msiren	Manokwari
Sulawesi Tenggara	Partai Nasdem	Waode Siti Armini Rere	Jakarta Selatan
Riau	Partai Berkarya	Intan Puspita	Karimun

Setelah itu, kami membuat visualisasi tahap awal untuk dataset bersih sebelum dilakukan *link prediction*. Visualisasi ditunjukkan oleh Gambar 2. *Node* oranye menunjukkan node awal yang memiliki relasi lebih dari 1 *node*. Label tidak bisa ditampilkan karena jumlah *node* yang terlalu besar.



Gambar 2. Visualisasi Data Set Awal.

2.3 Implementasi Algoritma

Pada tahap ini akan melakukan proses implementasi algoritma *link prediction* ke dalam data set yang telah didapatkan. Penerapan algoritma ini menggunakan bahasa pemrograman Python menggunakan library *networkx*¹. Algoritma yang dipilih adalah *jaccard coefficient*, *adamic-adar index*, dan *preferential attachment*.

2.3.1 Jaccard Coefficient

Jaccard Coefficient adalah algoritma *link prediction* yang mengukur jumlah tetangga antara dua node yang dibandingkan dengan node keseluruhan yang merupakan tetangga dari tetangga node yang dibandingkan [5]. Persamaan dari *Jaccard Coefficient* ditunjukkan oleh Persamaan 1.

¹ <https://networkx.github.io/documentation/stable/reference/introduction.html>

$$S_{ij} = \frac{p}{p + q + r} \quad (1)$$

Keterangan: p = Jumlah variable positif node a dan node b.
q = Jumlah variable positif node a dan variable negatif node b.
r = Jumlah variable negatif node a dan variable positif node b

2.3.2 Adamic-Adar Index

Adamic-Adar Index adalah algoritma *link prediction* yang digunakan untuk menghitung jumlah kesamaan yang dimiliki masing-masing *node*. Jumlah kesamaan tersebut kemudian akan dibandingkan dengan masing-masing tetangga dari *node* tersebut [8]. Persamaan dari *Adamic-Adar Index* ditunjukkan oleh Persamaan 2.

$$A(x, y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|} \quad (2)$$

Keterangan: A = Adamic-Adar Index Score.
Nu = Jumlah *node* yang memiliki tetangga yang sama.

2.3.3 Preferential Attachment

Preferential Attachment adalah salah satu algoritma *link prediction* yang mengukur hubungan antara dua *node* berdasarkan tingkat popularitasnya. Persamaan dari *Preferential Attachment* ditunjukkan oleh Persamaan 3.

$$|\Gamma(x)| \cdot |\Gamma(y)| \quad (3)$$

Keterangan: r(x) = jumlah tetangga dari node x
r(y) = jumlah tetangga dari node y

2.4 Penambahan Relasi Antar Node

Pada tahap ini akan melakukan proses penambahan relasi baru antara *node* yang telah dibuat. Tujuan dari tahap ini untuk melihat relasi baru yang terbentuk serta memvisualisasikan hasil dari *link prediction* setiap algoritma yang digunakan. Untuk *jaccard coefficient* akan diambil yang memiliki nilai *similarity* diatas 90%, kemudian untuk *adamic-adar index* dan *preferential attachment* akan diambil yang memiliki nilai persentil lebih dari 90%.

3. HASIL DAN PEMBAHASAN

Pada tahap ini akan melakukan proses implementasi algoritma kedalam data set yang telah diolah menggunakan metode *link prediction*. Pada tahap ini penulis akan menjelaskan bagaimana hasil dari ketiga metode *link prediction* yang telah dipilih yaitu *Jaccard Coefficient*, *Adamic-Adar Index*, dan *Preferential Attachment*. Bahasa pemrograman yang digunakan adalah Python dengan library *NetworkX*. *Source code* dan data set yang digunakan dapat di akses di zenodo [9].

3.1 Implementasi *Jaccard Coefficient*

Tabel 3. Contoh Hasil Implementasi *Jaccard Coefficient*

Caleg 1	Caleg 2	Score
HAPOSAN HARIANTO TOBING	SANTI OCTAVENTI, SPD	1
SOFYAN HASAN PAKAYA	FITRIANTI SAMAN	1
MUKHTAR BADEWING, S.Pi., MM	HJ. NURKAYA	1
ISMAIL MAKRUF, S.H	FADIL ISKANDAR, S.E., M.M	1
ADY ANSAR, S. HUT M.MPUB	ISMAN TRYADI IKSAN	1

Dengan melakukan perulangan pada setiap prediksi yang dibuat dari data set keseluruhan kemudian data tersebut diseleksi yang memiliki nilai kesamaan antara *node* diatas 0.9. Hasil implementasi ditunjukkan pada Tabel 3. Tahap ini menemukan bahwa ada 94.054 relasi antara *node* yang memiliki kesamaan nilai di atas 90% pada semua atribut yang terpilih yaitu daerah asal, partai pengusung serta daerah pemilihan.

3.2 Implementasi *Adamic-Adar Index*

Tabel 4. Contoh Hasil Implementasi *Adamic-Adar Index*

Caleg 1	Caleg 2	Score
KRISTIEN SAMIYATI PATI, SP	PAULUS MARAMBA MEHA, SH	4.328085123
IR. HENDRIK H. MARAWALI, MP	DOMINIKUS ALPHAWAN RANGGA KAKA	4.762379605
M. YUSUF ISMAIL PASE, SH, MH	TEUKU HASBULLAH. HD	7.129124156
ELFIRA SYLVIANI KAUNANG	SRI ISMAWATI	14.09129125
H. MUSLIM AYUB, S.H., M.M.	ZAKIYAH DRAZAT, S. Tr. Keb	13.38668497

Dengan melakukan perulangan pada setiap prediksi yang dibuat dari data set keseluruhan kemudian data tersebut diseleksi yang memiliki jumlah tetangga antara *node* yang memiliki nilai similarity di atas persentil 90. Hasil implementasi ditunjukkan pada Tabel 4.

3.3 Implementasi *Preferential Attachment*

Tabel 5. Contoh Hasil Implementasi *Preferential Attachment*

Caleg 1	Caleg 2	Score
SINTA NURJANAH	NUR MAILIS	689
KISLON OBISURU, S.Sos	PIPIH SOPIAH, SE	711
ALAMSYAH	GUSWANDRI, SE	955
Drs. H. SUMANTRI	ROMI FITRANDA	693
SAFRINA, Ners	ENNO AYU LAFYNNA	1109

Dengan melakukan perulangan pada setiap prediksi yang dibuat dari data set keseluruhan kemudian data tersebut diseleksi yang memiliki jumlah tetangga antara *node* yang memiliki nilai similarity di atas

4. KESIMPULAN

Dari penelitian yang telah dilakukan kami menemukan bahwa banyak nilai kesamaan antara masing-masing calon legislatif. Dengan menggunakan metode *jaccard coefficient*, *adamic-adar index*, dan *preferential attachment* bisa diketahui masing-masing nilai kesamaannya. Setelah itu, kami mencoba menambahkan relasi baru dengan membuat data set baru dari hasil implementasi *jaccard coefficient* didapatkan 75 relasi baru menggunakan metode *jaccard coefficient*, tambahan 137 relasi baru menggunakan metode *adamic-adar index* dan 191 relasi baru menggunakan *preferential attachment*.

Akan tetapi dalam penelitian ini penulis belum bisa memvalidasi hasil akhirnya. Oleh karena itu untuk penelitian kedepan bisa melakukan validasi hasil seperti menggunakan *Support Vector Machine (SVM)* untuk membuat matriks penilaian antara atribut dan *node* yang dihasilkan didalam *graph*.

DAFTAR PUSTAKA

- [1] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2020.
- [2] M. Mohammed, M. Khan, and E. Bashier, *Machine Learning: Algorithms and Applications*. 2016.
- [3] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM06: workshop on link analysis, counter-terrorism and security*, 2006, vol. 30, pp. 798–805.
- [4] L. Dong, Y. Li, H. Yin, H. Le, and M. Rui, "The algorithm of link prediction on social network," *Mathematical Problems in Engineering*, vol. 2013, 2013.
- [5] H. Chen, X. Li, and Z. Huang, "Link prediction approach to collaborative filtering," in *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05)*, 2005, pp. 141–142.
- [6] R. Guns and R. Rousseau, "Recommending research collaborations using link prediction and random forest classifiers," *Scientometrics*, vol. 101, no. 2, pp. 1461–1473, Nov. 2014, doi: 10.1007/s11192-013-1228-9.
- [7] C. Olston and M. Najork, "Web crawling," *Foundations and Trends in Information Retrieval*, vol. 4, no. 3, pp. 175–246, 2010.
- [8] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social networks*, vol. 25, no. 3, pp. 211–230, 2003.
- [9] G. A. Qandi and N. A. Rakhmawati, *Link Prediction for Indonesian Legislative Candidate*. Zenodo, 2020.