

HASIL CEK_Comparison of Classificati

by Tif Comparison Of Classificati

Submission date: 12-Aug-2023 11:41AM (UTC+0700)

Submission ID: 2144703705

File name: 22. Comparison of Classification Algorithm for Link Prediction Based on Generated Features of Local Similarity-Based Method.pdf (1.96M)

Word count: 7666

Character count: 37251

Algoritma Klasifikasi untuk Prediksi Tautan Berdasarkan Fitur yang Dibangkitkan dari Metode Berbasis Kesamaan Lokal

Classification Algorithm for Link Prediction Based on Generated Features of Local Similarity-Based Method

¹Siti Apriyanti Koni'ah, ²Herman Yuliansyah*

Teknik Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan
Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Bantul, Daerah Istimewa Yogyakarta

*e-mail: herman.yuliansyah@tif.uad.ac.id

(received: 6 Oktober 2021, revised: 26 November 2021, accepted: 19 Februari 2022)

Abstrak

Jaringan sosial adalah suatu stuktur sosial yang terdiri dari *nodes* atau simpul, *edge* atau *link* dan menggambarkan suatu aktifitas dalam sebuah *platform* media sosial. Kemudian *link prediction* adalah suatu teknik untuk meramalkan hubungan baru untuk jaringan masa depan berdasarkan informasi yang dieksplorasi dari topologi jaringan saat ini. Beberapa metode berbasis kesamaan lokal menggunakan informasi topologi untuk melakukan *link prediction*, namun metode-metode ini memiliki unjuk kerja yang berbeda-beda dan tergantung pada topologi jaringan. Pada penelitian ini mengusulkan penggunaan algoritma klasifikasi dalam pembelajaran mesin untuk memprediksi *link* di masa depan. Algoritma klasifikasi yang dibandingkan adalah *k-Nearest Neighbors* (KNN), *Naive Bayes*, *Decision Tree* dan *Random Forest* dengan menggunakan perbandingan enam *dataset* jaringan sosial dengan fitur yang dibangkitkan dari metode berbasis kesamaan lokal. Pada penelitian ini dilakukan dengan tiga tahapan: preprosesing, perbandingan klasifikasi dan evaluasi unjuk kerja. Temuan dari penelitian ini adalah algoritma *Random Forest* unggul pada hasil pengujian *accuracy*, *precision*, dan *F1-Score*. Namun, pada hasil pengujian *recall* *Random Forest* hanya unggul pada empat *dataset*: *soc-karate*, *soc-dolphin*, *soc-highschool M*, dan *Soc-sparrowlyon-flock-season 03*. Sedangkan pada *dataset* *soc-tribes* dan *soc-aves-weaver-social-05*, algoritma *Decision Tree* mampu mengungguli algoritma pembandingan lainnya.

Kata kunci: Jaringan Sosial, *Prediksi Link*, *Kesamaan Lokal*, Algoritma Klasifikasi.

Abstract

A social network is a social structure that consists consisting of nodes, edges, or links and describes activity on a social media platform. Later, link prediction is a technique to predict new relationships for future networks based on information explored from the current network topology. Several local similarity-based methods use topological information to predict the link. However, these methods have different performances and depend on the network topology. This study proposes using classification algorithms of machine learning to predict future links. The classification algorithms compared are *k-Nearest Neighbors* (KNN), *Naive Bayes*, *Decision Tree*, and *Random Forest* by comparing six social network datasets with features generated from local similarity-based methods. This research was conducted in three stages: preprocessing, classification comparison, and performance evaluation. The findings of this study are that the *Random Forest* algorithm outperforms for testing accuracy, precision, and *F1-Score*. However, in the recall test results, *Random Forest* only outperformed other benchmark algorithms in the four datasets: *soc-karate*, *soc-dolphin*, *soc-highschool M*, and *Soc-sparrowlyon-flock-season 03*. Meanwhile, in the datasets *soc-tribes* and *soc-aves-weaver-social-05*, the *Decision Tree* algorithm outperformed other benchmark algorithms.

Keywords: Social Network, Link Prediction, Local Similarity Based, Classification Algorithm.

³<http://sistemasi.ftik.unisi.ac.id>

1 Pendahuluan

Jaringan sosial merupakan suatu *platform* yang menawarkan berbagai layanan untuk memfasilitasi koneksi individu dengan setiap orang yang mempunyai ketertarikan, perilaku dan minat yang sama [1]. Dengan adanya jaringan sosial semua aktivitas dapat dilakukan secara *online* mulai dari berbelanja, bermain *game*, mengenal satu sama lain dan menjalin hubungan dengan orang baru. Manfaat jaringan sosial adalah sebagai ajang pencarian teman dan menjalin pertemanan secara *online*.

Jaringan sosial dapat digambarkan sebagai sebuah graf (G) yang merupakan sebuah himpunan $G=(V,E)$ dimana V didefinisikan sebagai *vertex* (*nodes* atau titik) pada graf G yang ditulis $V(G)$ dan E adalah himpunan *edges* yang menghubungkan sepasang *nodes* dari graf yang ditulis sebagai $E(G)$. Pada umumnya graf digunakan untuk mempresentasikan objek-objek diskrit dan hubungan antara objek-objek tersebut [2]. *Link prediction* merupakan prediksi keberadaan *link* antara dua entitas yang terdapat dalam suatu jaringan sosial. *Link prediction* pada umumnya digunakan untuk memprediksi hubungan baru yang muncul masa mendatang [3]. *Link prediction* memiliki peran penting dalam memprediksi struktural jaringan sosial [4].

Link prediction mempunyai 2 pendekatan utama yaitu *similarity based* dan *learning based*. Pendekatan pada *similarity based* pengukurannya dilakukan berdasarkan skor kesamaan indeks, sedangkan pada pendekatan *learning based* pengukuran dilakukan berdasarkan *confusion matrix* [5]. Dalam sejumlah penelitian dengan persoalan [3] – [5], menguraikan bahwa penerapan *link prediction* digunakan dalam kasus merekomendasikan dan memprediksi suatu jaringan dengan menggunakan fitur yang terdapat pada *link prediction*. Bahkan pada penelitian [3] dan [4] melakukan perbandingan terhadap algoritma klasifikasi namun penelitian terdahulu lebih fokus melakukan perbandingan pada kasus jaringan kolaborasi setiap antar penulis.

Dalam penelitian ini melakukan perbandingan menggunakan *dataset* jaringan sosial dengan algoritma klasifikasi diantaranya: *k-Nearest Neighbors* (KNN), *Naive Bayes*, *Random Forest* dan *Decision Tree* dengan menerapkan fitur yang termasuk ke dalam kategori *local similarity based*. Selain itu, akan dilakukan pengukuran unjuk kerja algoritma dilakukan dengan menggunakan *confusion matrix* untuk mengetahui algoritma klasifikasi yang memiliki tingkat pengukuran yang lebih tinggi baik dalam prediksi pertemanan pada jaringan sosial.

Local similarity based merupakan salah satu kategori pendekatan kesamaan dimana kategori ini menggunakan rumus perhitungan yang sederhana dalam menyelesaikan perhitungan suatu indeks. Indeks lokal dihitung berdasarkan nilai *neighbors* dan *degree* dari suatu jaringan. *Local similarity based* ini bisa digunakan dalam menghitung skor kedekatan pada suatu kandidat pada pasangan nodes [6]. Fitur yang termasuk ke dalam beberapa kategori *local similarity based* diantaranya: *Preferential Attachment*, *Resource Allocation*, *Adamic/Adar* dan *Common Neighbors* [7]. *Preferential Attachment* didefinisikan sebagai tautan baru yang terhubung ke *nodes* x sebanding dengan $k(x)$. Probabilitas pada tautan baru ini akan menghubungkan mekanisme x , indeks kesamaan yang sesuai y sebanding dengan $k(x) \times k(y)$ [8]. Selanjutnya, *Resource Allocation* didefinisikan sebagai model *nodes* x dan y yang tidak terhubung. Jumlah nodes y yang diperoleh dapat dianggap sebagai kesamaan antara kedua nodes [9]. Kemudian, *Adamic/Adar* didefinisikan sebagai pengukuran pada dua nodes yang tidak terhubung dengan memberikan bobot lebih pada *neighbors* yang memiliki jumlah sedikit [10]. *Common Neighbors* merupakan sebuah implementasi untuk prediksi tautan yang digunakan untuk mendefinisikan banyaknya *neighbors* (x) dan (y) yang sama [11].

Penelitian ini bertujuan untuk mengusulkan penggunaan algoritma klasifikasi dalam pembelajaran mesin untuk memprediksi *link* di masa depan. Penggunaan algoritma klasifikasi ini sebelumnya diuji terlebih dahulu kinerjanya dengan cara membandingkan algoritma klasifikasi sehingga dapat diperoleh algoritma yang dapat memberikan kinerja lebih baik. Algoritma klasifikasi yang dibandingkan adalah *k-Nearest Neighbors* (KNN), *Naive Bayes*, *Decision Tree* dan *Random Forest* dengan menggunakan perbandingan enam *dataset* jaringan sosial dengan fitur yang dibangkitkan dari metode berbasis kesamaan lokal. Fitur yang ada dipelajari oleh algoritma klasifikasi ini berasal dari pembelajaran informasi jaringan dengan metode *local similarity*. Adapun manfaat dari penelitian ini terdapat sebuah usulan model prediksi pertemanan dalam jaringan sosial dengan memanfaatkan algoritma klasifikasi dan metode *local similarity*.

2 Tinjauan Literatur

Beberapa penelitian terdahulu digunakan sebagai ukuran kebaruan dalam penelitian ini dan ditunjukkan dengan perbandingan fitur, *dataset*, pengukuran dan metode seperti ditunjukkan pada Tabel 1. Fitur adalah variabel yang dipertimbangkan dalam menentukan *link prediction*. Selanjutnya, *dataset* adalah sumber data yang dianalisa sesuai dengan kasus dalam dunia nyata. Sedangkan pengukuran dan metode adalah bagian dari validasi untuk model yang diusulkan.

Table 1. Tinjauan Literatur

Nama Penulis, Tahun	Fitur	Dataset	Pengukuran	Metode	Ref
Nisha S Sarma dan, Anna Prathibha Shobak, 2016	Tidak disebutkan	Facebook	Confusion Matrix	k-Nearest Neighbors (KNN)	[12]
Chaoqun Chen, Sanhong Deng, dan Jing Lu, 2017	Common neighbor, Shortest Path, Jaccard's Coefficient's Adamic/Adar, Preferential Attachment, Page Rank, Communities, Communities member number	DBLP (Digital Bibliography and Library Project) tahun 2001 – 2004)	Confusion Matrix	Neural Network, Logistic Linear dan SVM	[5]
Erik Medina-Acuñal(B), Pedro Shiguihara-Juárez1, dan Nils Murrugarra-Llerena, 2019	Adamic/Adar, Nodes A edges, Nodes B edges, Jaccard's Coefficient, Preferential Attachment	Scopus	Confusion Matrix	Decision Tree, Logistic Regression, SVM, Neural Network	[13]
Ji-chao Li, Danling Zhao, Bing-Feng Ge, Ke-Wei Yang, dan Ying-Wu Chen, 2018	Meta-Path Features Notations	GDNs & Weapon System-Of-Systems Combat Network (WSCNs)	AUC, Confusion Matrix	Neural Network	[14]
Zuoxi Yang, Dingding Li, Ronghua Lin, Weiling Li, Hai Liu, dan Yong Tang, 2019	Organizations, degrees, research areas and courses	SCHOLAT (A Real-World Scholar Social Network)	Confusion Matrix	Decision Tree	[15]

Berdasarkan penelitian sebelumnya tersebut terkait *link prediction* pada kolaborasi penulis menggunakan *BP Neural Network* ditemukan bahwa dalam sebuah jaringan kolaborasi penulis, beberapa penulis merasa kesulitan dalam menentukan penulis yang sesuai dengan bidang keahlian. Kemudian, penelitian ini melakukan perbandingan algoritma *Neural Network* dengan algoritma *Regresi Linear* dan *SVM*, dengan memadukan fitur *Common Neighbor*, *Shortest Path*, *Jaccard's Coefficient*, *Adamic/Adar*, *Preferential Attachment*, *Page Rank*, *Communities*, dan *Communities member number* untuk mengetahui algoritma yang tepat dalam menentukan penulis yang sesuai dengan bidang keahlian pada saat berkolaborasi dengan penulis lainnya. Penelitian ini dapat memprediksi dan merekomendasikan setiap penulis dengan beberapa kolaborator yang diminati [5].

Pada penelitian selanjutnya mengenai penerapan *link prediction* dengan membandingkan algoritma *regresi logistik*, *SVM*, *Neural Network*, *Decision Tree* dan *k-Nearest Neighbors (KNN)* menggunakan *dataset Scopus*. Selain melakukan perbandingan juga dilakukan analisis mengenai fitur yang digunakan untuk memprediksi dan merekomendasikan kolaborasi jaringan penulis. Fitur yang dianalisis dalam penelitian ini adalah *Adamic/Adar*, *Nodes A Edges*, *Nodes B Edges*, *Jaccard's Coefficient*, *Preferential Attachment*. Hasil dari penelitian ini didapatkan bahwa algoritma *Decision*

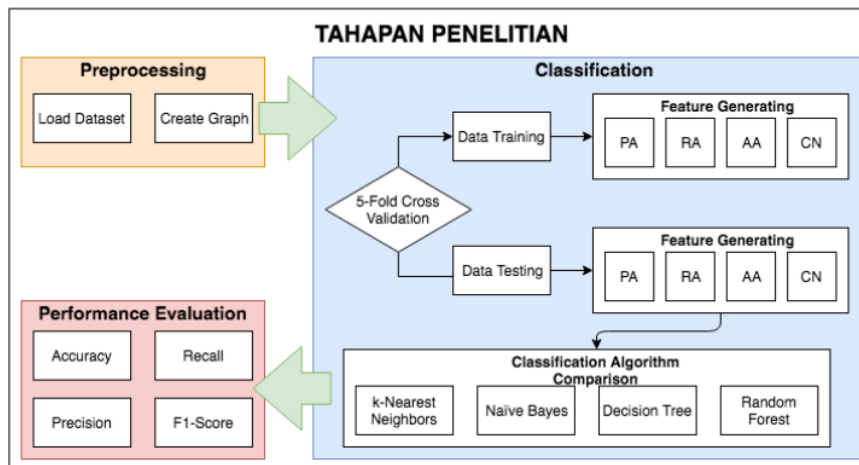
<http://sistemasi.ftik.unisi.ac.id>

Tree memperoleh skor *accuracy* sebesar 91.07% dan hasil analisa fitur diketahui bahwa fitur *Preferential Attachment & Jaccard's Coefficient* belum tepat dalam prediksi jaringan penulis [13]. Terdapat juga penelitian lain yang terkait mengenai *link prediction* bertujuan untuk memprediksi berbagai jenis jaringan heterogen. Dalam penelitian ini digunakan pengukuran menggunakan metode *Area Under Curve (AUC)* dan *confusion matrix* pada *dataset GDNs & Weapon System-of-systems Combat Networks (WSCNs)*. Didapatkan hasil bahwa MPBP (*Meta-Path feature-based BP Neural Network Model*) dapat mengungguli model banding [14].

Penelitian ini mengusulkan ekstraksi fitur yang dibangkitkan berdasarkan topologi dari jaringan menggunakan metode berbasis *local similarity*. Fitur-fitur ini diantaranya adalah *Preferential Attachment, Resource Allocation, Adamic/Adar* dan *Common Neighbors*. Selain itu, jenis dataset jaringan sosial adalah dataset yang diperiksa dalam perbandingan algoritma klasifikasi menggunakan *k-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree* dan *Random Forest* sehingga ini menjadi kebaruan yang ditawarkan dalam penelitian ini. Jaringan sosial adalah *platform* yang sedang populer saat ini sehingga kontribusi dari penelitian ini dapat dimanfaatkan untuk pengembangan selanjutnya.

3 Metode Penelitian

Penelitian ini menggunakan teknik dokumentasi yang dilakukan untuk mendapatkan data pendukung dengan mengumpulkan *dataset* jaringan sosial yang didapatkan dari <http://networkrepository.com> [16]. Tahapan dalam penelitian ini diilustrasikan pada Gambar 1, tahapan penelitian yang meliputi 3 tahapan diantaranya: preprosesing, perbandingan klasifikasi dan evaluasi unjuk kerja. Tahapan pertama terdapat preprosesing yaitu melakukan *load dataset* dan membentuk sebuah *dataset* menjadi sebuah graf. Tahapan kedua yaitu klasifikasi, dimana pada tahapan ini *dataset* dilakukan *split* data menggunakan *cross validation* untuk membagi sebuah *dataset* menjadi data *training* dan data *testing*. Setelah dilakukan *split* data *dataset* yang diperoleh dilakukan implementasi fitur *local similarity based* diantaranya: *Preferential Attachment, Resource Allocation, Adamic/Adar* dan *Common Neighbors*. Setelah implementasi fitur dilakukan, maka hasil perolehan fitur diklasifikasikan menggunakan algoritma klasifikasi diantaranya: *k-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree* dan *Random Forest*. Tahap ketiga dilakukan pengujian menggunakan *confusion matrix* untuk mengetahui keakuratan dalam sebuah algoritma yang digunakan.



Gambar 1. Tahapan Penelitian

A. Preprocessing

Tahap pertama ini dilakukan *load dataset* berformat *Comma Delimited (CSV)* pada pemrograman python. Kode Program 1 digunakan untuk *load dataset* yang akan diuji dan berasal dari *dataset repository* jaringan. Setelah dilakukan proses *load dataset*, kemudian data diolah kembali untuk pembentukan sebuah graf yang dihasilkan dari *nodes* dan *edges* yang saling terhubung ataupun tidak terhubung.

Kode Program 1 - Load Dataset

No.	Kode Program
1.	filename, delimiter, graphname = ""
2.	if os.path.isfile(filename) == True:
3.	loaddataset = np.loadtxt(filename,
4.	delimiter=delimiter, dtype=int)
5.	else:
6.	loaddataset = 'File does not exist'
7.	loaddataset = loaddataset[:, [0, 1]]
8.	print(loaddataset)

Kode Program 2 digunakan untuk *create graph* dengan perintah *nx.Graph()*. Kemudian untuk menambahkan *edges* ke dalam graf digunakan perintah *add_edge(u,v)*, dimana *u* dan *v* adalah *node*. Pada eksperimen ini menggunakan *5-fold cross validation* untuk membagi dataset menjadi data latih dan data uji. Pada tahap *create graph* ini akan menghasilkan jaringan yang selanjutnya akan dibangkitkan fitur *Preferential Attachment*, *Resource Allocation*, *Adamic/Adar* dan *Common Neighbors* untuk selanjutnya digunakan pada tahap *classification*.

Kode Program 2 - Create Graph

No.	Kode Program
1.	G_dataset=nx.Graph()
2.	G_dataset.name = f"Dataset {graphname}"
3.	for u,v in loaddataset:
4.	if u == v:
5.	print("u and v are equal :
6.	"+str(u)+"&"+str(u))
7.	else:
8.	G_dataset.add_edge(u,v)

B. Classification

Tahapan ini dilakukan untuk pembangkitan pada *dataset* yang digunakan. Fitur yang diimplementasikan disini merupakan sebuah fitur yang termasuk ke dalam *local similarity based* dan diimplementasikan dalam Kode Program 3.

Persamaan *Preferential Attachment* didefinisikan sebagai:

$$s(x, y) = |\Gamma(x)| \cdot |\Gamma(y)| \quad (1)$$

Persamaan *Resource Allocation* ini didefinisikan sebagai:

$$s(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|} \quad (2)$$

Persamaan *Adamic/Adar* didefinisikan sebagai:

$$s(x, y) = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|} \quad (3)$$

Persamaan *Common Neighbors* (*x*) dan (*y*) didefinisikan sebagai:

<http://sistemasi.ftik.unisi.ac.id>

$$s(x, y) = |\Gamma(x) \cap \Gamma(y)| \quad (4)$$

Kode Program 3 - Membangkitkan Fitur

No.	Kode Program
1.	for i in <i>Testing_index</i> :
2.	k = <i>completegraph</i> [i]
3.	if k in <i>list(G_dataset.edges)</i> :
4.	<i>y_Testing.append</i> (1)
5.	<i>label=1</i>
6.	else:
7.	<i>y_Testing.append</i> (0)
8.	<i>label=0</i>
9.	<i>preds1 = [p for u, v, p in</i>
10.	<i>nx.preferential_attachment(G_train, [(k[0],k[1])])]</i>
11.	<i>pa = preds1[0]</i>
12.	<i>preds2 = [p for u, v, p in</i>
13.	<i>nx.resource_allocation_index(G_train, [(k[0], k[1])])]</i>
14.	<i>ra = preds2[0]</i>
15.	<i>preds3 = [p for u, v, p in</i>
16.	<i>nx.adamic_adar_index(G_train, [(k[0], k[1])])]</i>
17.	<i>aa = preds3[0]</i>
18.	<i>cn =len(set(nx.common_neighbors(G_train,k[0], k[1])))</i>

Pada tahap selanjutnya ini dilakukan prediksi menggunakan beberapa algoritma klasifikasi yang digunakan, penelitian ini membandingkan empat algoritma klasifikasi diantaranya: *k-Nearest Neighbors* (KNN), *Naïve Bayes*, *Decision Tree* dan *Random Forest*. Kode Program 4 digunakan untuk klasifikasi.

Kode Program 4 - Klasifikasi

No.	Kode Program
1.	<i>KNN= KNeighborsClassifier(n_neighbors=5)</i>
2.	<i>KNN.fit(x_Training, y_Training)</i>
3.	<i>hasilknn = KNN.predict(x_Testing)</i>
4.	<i>ModelBayes = GaussianNB()</i>
5.	<i>ModelBayes.fit(x_Training, y_Training)</i>
6.	<i>hasilbayes= ModelBayes.predict(x_Testing)</i>
7.	<i>ModelDc = DecisionTreeClassifier()</i>
8.	<i>ModelDc.fit(x_Training, y_Training)</i>
9.	<i>hasilDc = ModelDc.predict(x_Testing)</i>
10.	<i>ModelRF = RandomForestClassifier()</i>
11.	<i>ModelRF.fit(x_Training, y_Training)</i>
12.	<i>hasilRF = ModelRF.predict(x_Testing)</i>

C. Performance Evaluation

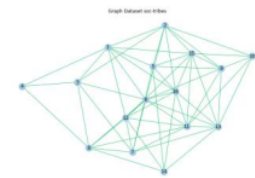
Pada tahapan ini dilakukan pengujian kinerja dari model klasifikasi yang digunakan dan untuk mengetahui nilai *accuracy*, *recall*, *precision* dan *F1-Score* dari setiap algoritma dan fitur yang telah ditetapkan pada penelitian ini. Pengujian dilakukan menggunakan *confusion matrix* seperti ditunjukkan pada Kode Program 5. Kode Program 5 merupakan kode program yang digunakan dalam mencari nilai *accuracy*, *recall precision* dan *F1-Score* pada hasil klasifikasi yang telah dilakukan sebelumnya menggunakan algoritma klasifikasi *KNN*, *Naïve Bayes*, *Decision Tree* dan *Random Forest*. Hasil klasifikasi yang memiliki nilai *accuracy*, *precision* dan *recall* paling tinggi dapat dikatakan sebagai algoritma yang mempunyai performa baik dalam prediksi pertemanan pada jaringan sosial.

Kode Program 5 – Confision Matrix

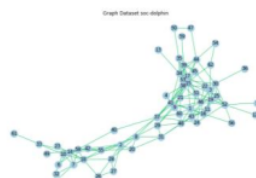
No.	Kode Program
1.	from sklearn.metrics import
2.	classification_report, confusion_matrix
3.	akurasicdc.append(accuracy_score(y_predikdc, y_testing))
4.	recalldc.append(recall_score(y_predikdc, y_testing))
5.	precisiondc.append(precision_score(y_predikdc, y_testing))
6.	F1dc.append(f1_score(y_predikdc, y_testing))

4 Hasil dan Pembahasan

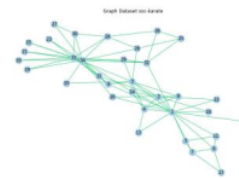
Penelitian ini dilakukan dengan menggunakan enam *dataset* jaringan sosial, diantaranya *dataset Soc-tribes, Soc-Karate, Soc-dolphin, Soc-Highschool-Moreno, Soc-Aves Aparrowlyon Flock Season3* dan *Soc-Aves-weaver-social-05* dan hasil graf ditunjukkan pada Gambar 2-7.



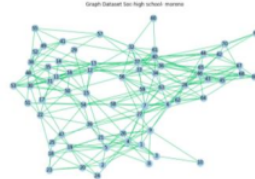
Gambar 2 Graf Soc-Tribes



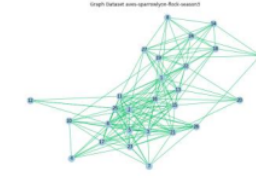
Gambar 3 Graf Soc-Dolphin



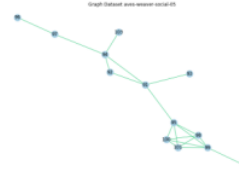
Gambar 4 Graf Soc-Karate



Gambar 5 Graf Soc-Highschool Moreno



Gambar 6 Graf Soc-aves-sparrowlyon-flock-season3



Gambar 7 Graf Soc-aves-weaver-social-05

Berdasarkan pembentukan graf tersebut, dapat diperoleh informasi berupa jumlah *nodes*, *edges*, *density* atau kerapatan jaringan, *average clustering coefficient* dan *average degree*. Informasi mengenai *dataset* yang digunakan pada penelitian ini selengkapnya ditunjukkan pada Tabel 2.

Table 2. Informasi Dataset

Nama Dataset	Nodes	Edges	Density	AVG Clustering Coefecient	AVG Degree
Soc-Tribes	16	58	0.483333	0.539187	7,25
Soc-Dolphin	62	159	0.084082	0.258958	5,13
Soc-Karate	34	78	0.139037	0.570638	4,59
Soc-highschool-moreno	70	274	0.113457	0.464907	7,83
Soc-aves-sparrowlyon-flock-season3	27	163	0.464387	0.744546	12,07
Soc-aves-weaver-social-05	13	19	0.24359	0.425641	2,92

Tabel 2 merupakan informasi mengenai jumlah *nodes*, *edges*, *density* dan *degree* untuk *dataset* yang akan digunakan dalam melakukan perbandingan algoritma klasifikasi untuk prediksi pertemanan. Dalam penelitian ini menggunakan beberapa kategori data jaringan sosial yang berbeda, diantaranya: *dataset Soc-Highschool Moreno* [16], *Soc-tribes* yang merupakan suatu data yang berisi mengenai

<http://sistemasi.ftik.unisi.ac.id>

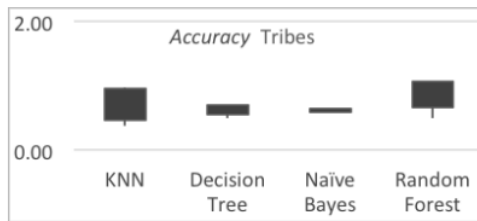
daftar semua tautan, di mana tautan ini mewakili persahabatan yang ditandatangani antar suku. Dalam *dataset* ini diceritakan bahwa suku jaringan sosial merupakan suatu suku dari struktur aliansi Gahuku-Gama di Dataran Tinggi Tengah Timur New Guinea [17]. Kemudian penelitian ini menggunakan *dataset* yang berisi mengenai ikatan sosial di antara anggota klub karate pada suatu universitas yang dikumpulkan oleh Wayne Zachary pada tahun 1977 [18]. Selain menggunakan *dataset* jaringan sosial *Soc-tribes* dan *Soc-karate*, dalam penelitian ini digunakan *dataset* jaringan sosial yang terdapat pada makhluk hidup seperti *dataset Soc-dolphin* yang berisi mengenai jaringan sosial lumba-lumba yang mempunyai hidung botol dan berisi daftar semua tautan, di mana tautan mewakili asosiasi yang sering terjadi antara lumba-lumba [19]. Kemudian terdapat *dataset Soc-Aves-Sparrowlyon-Flock Season3* yang berisi mengenai data sosial yang dikumpulkan selama dua musim non-kawin: Oktober 2010-Februari 2011 (Musim 2) dan Oktober 2011-April 2012 (Musim 3) [20]. Sedangkan *dataset Soc-aves-sparrowlyon-flock-season3* berisi mengenai data jaringan yang mewakili data sosial yang dikumpulkan dari 23 koloni penenun yang ramah [21].

Selanjutnya pengujian performa dilakukan untuk meramalkan hubungan baru untuk jaringan masa depan berdasarkan informasi yang dieksplorasi dari topologi jaringan saat ini. Teknik yang digunakan adalah dengan membangkitkan fitur yang akan digunakan untuk meramalkan dengan algoritma klasifikasi. Pada percobaan ini, dilakukan dengan *5-fold cross validation* untuk membagi kedalam data latih dan data uji untuk masing-masing *dataset* jaringan yang diperiksa sehingga diperoleh 30 kali percobaan. Hasil pembagian data ini ditampilkan pada Lampiran A-L. Dari hasil pembagian data ini dilakukan pembangkitan fitur berdasarkan informasi topologi yang diukur berdasarkan metode *local similarity* untuk data latih dan data uji. Kemudian dari hasil pembangkitan fitur ini maka peramalan meramalkan hubungan baru dapat dievaluasi performa dari masing-masing algoritma klasifikasi yang akan diuji dengan *confusion matrix*. Hasil pengujian ini ditampilkan secara rata-rata pada Tabel 3-6. Sedangkan untuk melihat sebaran data hasil pengujian divisualisasikan dalam grafik *candlestick* pada Gambar 8-31. Grafik *candlestick* ini menunjukkan nilai maksimal, minimal dan besaran standar deviasi dari hasil pengujian terhadap data rata-rata.

Table 3. Hasil Pengujian Accuracy

Dataset	KNN	Naïve Bayes	Decision Tree	Random Forest
Soc-Tribes	0,710 (0,24)	0,61 (0,24)	0,62 (0,06)	0,85 (0,20)
Soc-Dolphin	0,91 (0,03)	0,88 (0,00)	0,91 (0,04)	0,94 (0,02)
Soc-Karate	0,90 (0,06)	0,85 (0,01)	0,87 (0,07)	0,94 (0,03)
Soc-highschool-moreno	0,93 (0,03)	0,88 (0,00)	0,90 (0,06)	0,96 (0,03)
Soc-aves-sparrowlyon-flock-season3	0,89 (0,09)	0,83 (0,01)	0,86 (0,09)	0,97 (0,06)
Soc-aves-weaver-season-05	0,86 (0,06)	0,86 (0,01)	0,85 (0,03)	0,88 (0,04)

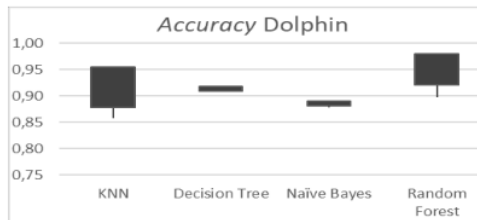
Pada Tabel 3, berdasarkan hasil pengujian *accuracy* dapat diketahui bahwa nilai *accuracy* tertinggi berdasarkan pengujian menggunakan ke enam *dataset* yang berbeda, didapatkan bahwa skor *Random Forest* lebih unggul dibandingkan algoritma *k-Nearest Neighbors (KNN)*, *Naïve Bayes* dan *Decision Tree*. Hasil pengujian *accuracy* tertinggi ditandai dengan penebalan angka dan nilai dalam kurung adalah nilai standar deviasi dari pengujian *accuracy*, seperti ditunjukkan pada Tabel 3.



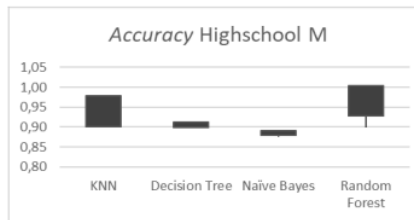
Gambar 8. Accuracy Soc-Tribes



Gambar 9. Accuracy Soc-Karate



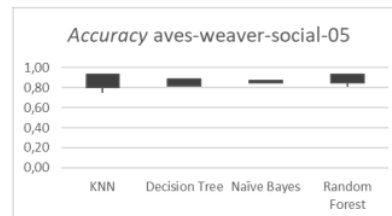
Gambar 10. Accuracy Soc-Dolphin



Gambar 11. Accuracy Soc-Highschool Moreno



Gambar 12. Accuracy Soc-Aves-Sparrowlyon-F-s3



Gambar 13. Accuracy Soc-Aves-Weaver-season 05

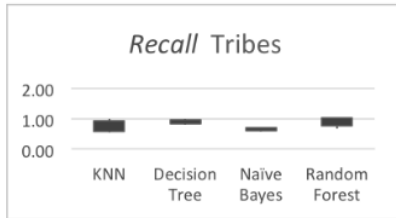
Gambar 8-13 merupakan hasil eksperimen yang digambarkan menggunakan *candlestick* yang diperoleh dari hasil pengujian skor *accuracy* setiap *dataset* yang telah diklasifikasikan. Kemudian pada Tabel 4 disajikan mengenai hasil pengujian *recall* setiap *dataset*. Diamati pada Tabel 4, hasil pengujian *recall* dapat diketahui bahwa nilai *recall* tertinggi berdasarkan pengujian menggunakan ke enam *dataset* yang berbeda, didapatkan juga bahwa skor algoritma *Random Forest* lebih unggul dibandingkan algoritma *k-Nearest Neighbors (KNN)*, *Naive Bayes* dan *Decision Tree*. Namun, pada *dataset soc-aves-weaver-season-05* diungguli oleh Algoritma *Decision Tree*. Pada Tabel 4, hasil pengujian *recall* tertinggi ditandai dengan penebalan angka dan nilai dalam kurung adalah nilai standar deviasi dari pengujian *recall*.

Table 4. Hasil Pengujian Recall

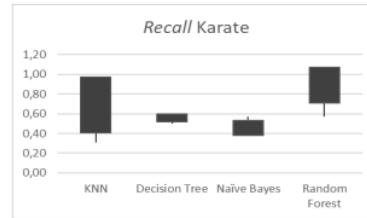
Dataset	KNN	Naive Bayes	Decision Tree	Random Forest
Soc-Tribes	0,75 (0,17)	0,64 (0,05)	0,89 (0,06)	0,89 (0,13)
Soc-Dolphin	0,52 (0,21)	0,36 (0,01)	0,36 (0,20)	0,78 (0,24)
Soc-Karate	0,68 (0,28)	0,45 (0,07)	0,55 (0,03)	0,88 (0,18)
Soc-highschool-moreno	0,74 (0,17)	0,48 (0,01)	0,71 (0,16)	0,85 (0,18)
Soc- aves-sparrowlyon-	0,88 (0,10)	0,82 (0,01)	0,85 (0,14)	0,96 (0,08)

<http://sistemasi.ftik.unisi.ac.id>

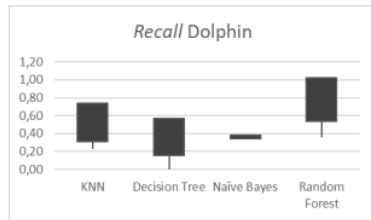
flock-season3				
Soc-aves-weaver-social-05	0,94 (0,05)	0,90 (0,09)	1 (0)	0,94 (0,06)



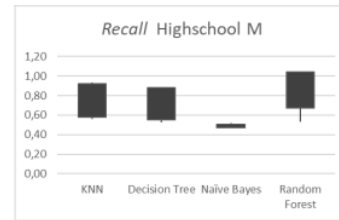
Gambar 14. Recall Soc-Tribes



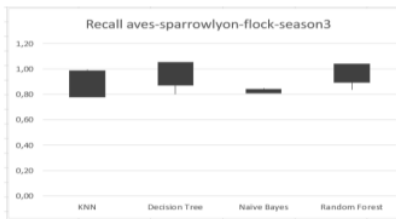
Gambar 15. Recall Soc-Karate



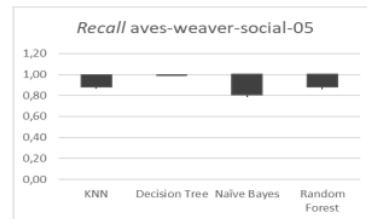
Gambar 16. Recall Soc-Dolphin



Gambar 17. Recall Soc-Highschool Moreno



Gambar 18. Recall Aves Soc-Aves-Sparrowlyon-F-s3



Gambar 19. Recall Soc-aves-weaver-social-05

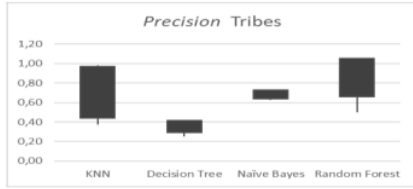
Gambar 7 – 12 merupakan hasil eksperimen yang digambarkan dalam *candlestick* skor *recall*. Hasil ini diperoleh setiap *dataset* yang telah diklasifikasikan. Kemudian pada Tabel 5 hasil pengujian *precision* tertinggi ditandai dengan penebalan angka dan nilai dalam kurung adalah nilai standar deviasi dari pengujian *precision*.

Table 5. Hasil *Precision*

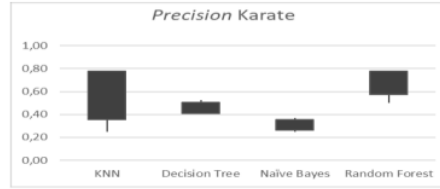
Nama Dataset	KNN	Naive Bayes	Decision Tree	Random Forest
Soc-Tribes	0,70 (0,26)	0,68 (0,04)	0,35 (0,06)	0,85 (0,06)
Soc-Dolphin	0,44 (0,21)	0,52 (0,00)	0,30 (0,18)	0,56 (0,13)
Soc-Karate	0,56 (0,21)	0,30 (0,04)	0,45 (0,04)	0,67 (0,10)
Soc-highschool-moreno	0,64 (0,24)	0,67 (0,01)	0,39 (0,20)	0,81 (0,18)
Soc- aves-sparrowlyon-flock-season3	0,90 (0,09)	0,81 (0,03)	0,84 (0,02)	0,97 (0,05)
aves-weaver-social-05	0,63 (0,17)	0,61 (0,03)	0,61 (0,03)	0,69 (0,11)

Diamati pada Tabel 5 hasil pengujian *precision*, diketahui bahwa nilai *precision* tertinggi berdasarkan pengujian menggunakan ke enam *dataset* yang berbeda, didapatkan bahwa algoritma *Random Forest* <http://sistemasi.ftik.unisi.ac.id>

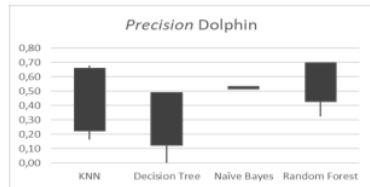
mengungguli ketiga algoritma yaitu: KNN, *Naïve Bayes* dan *Decision Tree*. Gambar 13-18 merupakan gambar *candlestick* dari hasil pengujian *precision* yang diperoleh dari hasil klasifikasi setiap *dataset*.



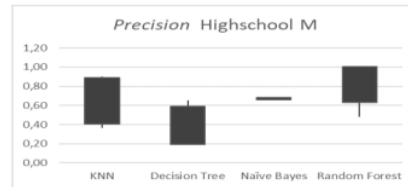
Gambar 20. Precision Soc-Tribes



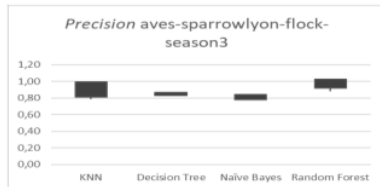
Gambar 21. Precision Soc-Karate



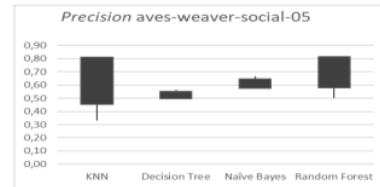
Gambar 22. Precision Soc-Dolphin



Gambar 23. Precision Soc-Highschool Moreno



Gambar 24. Precision Soc-Aves-Aparrowlyon-S3

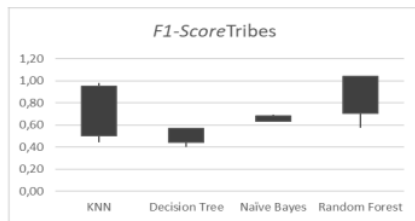


Gambar 25. Precision Soc-Aves-Weaver-Social-05

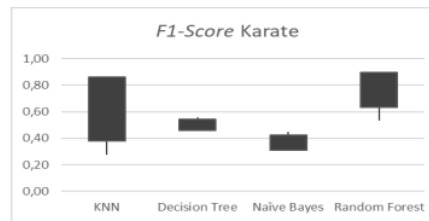
Tabel 6. Hasil *F1-Score*

Dataset	KNN	Naïve Bayes	Decision Tree	Random Forest
Soc-Tribes	0,72 (0,22)	0,66 (0,02)	0,50 (0,63)	0,87 (0,16)
Soc-Dolphin	0,45 (0,21)	0,42 (0,13)	0,33 (0,18)	0,65 (0,17)
Soc-Karate	0,61 (0,24)	0,36 (0,56)	0,50 (0,40)	0,76 (0,13)
Soc-highschool-moreno	0,69 (0,21)	0,56 (0,01)	0,45 (0,10)	0,83 (0,18)
Soc- aves-sparrowlyon-flock-season3	0,89 (0,09)	0,81 (0,02)	0,84 (0,01)	0,96 (0,06)
aves-weaver-social-05	0,73 (0,13)	0,72 (0,04)	0,68 (0,02)	0,79 (0,07)

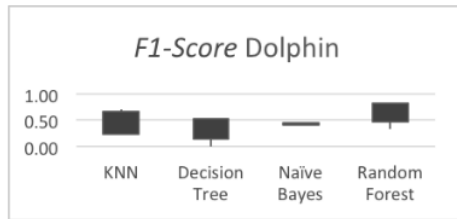
Diamati pada Tabel 6 merupakan hasil pengujian *F1-Score* dan diketahui bahwa nilai *F1-Score* tertinggi berdasarkan pengujian menggunakan ke enam *dataset* yang berbeda, didapatkan bahwa skor algoritma *Random Forest* lebih unggul dibandingkan algoritma *k-Nearest Neighbors (KNN)*, *Naïve Bayes* dan *Decision Tree*. Tabel 6 merupakan hasil pengujian *F1-Score* tertinggi ditandai dengan penebalan angka dan nilai dalam kurung adalah nilai standar deviasi dari pengujian *F1-Score*.



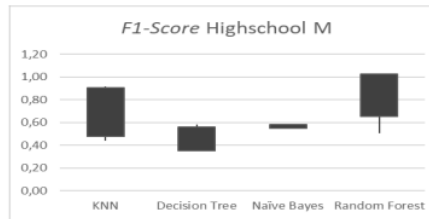
Gambar 26. F1-Score Soc-Tribes



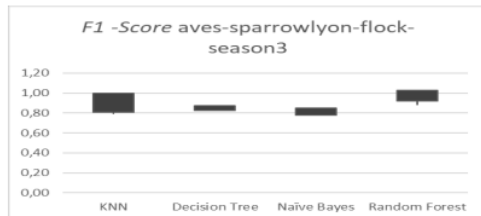
Gambar 27. F1-Score Soc-Karate



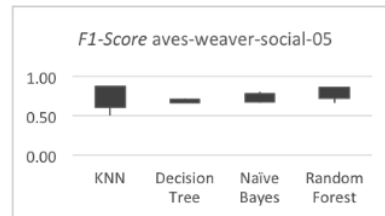
Gambar 28. F1-Score Soc-Dolphin



Gambar 29. F1-Score Soc-Highschool Moreno



Gambar 30. Precision Soc-Aves-Aparowlyon-S3



Gambar 31. Precision Soc-Aves-Weaver-Social-05

5 Kesimpulan

Berdasarkan pengujian yang telah dilakukan dengan menggunakan enam dataset jaringan sosial yang berbeda, dapat disimpulkan bahwa pada hasil pengujian *accuracy*, *recall*, *Precision* dan *F1-Score* menunjukkan bahwa algoritma *Random Forest* mampu unggul pada ke enam dataset. Namun pada hasil pengujian *recall*, *Random Forest* hanya unggul pada empat dataset: *soc-karate*, *soc-dolphin*, *soc-highschool m* dan *soc-sparrowlyon-flock-season 03*. Sedangkan pada dataset *soc-tribes* dan *soc-aves-weaver-social-05* diungguli oleh algoritma *Decision Tree*. Dengan demikian dapat disimpulkan bahwa algoritma *Random Forest* lebih disarankan untuk digunakan untuk meramalkan hubungan baru untuk jaringan masa depan berdasarkan informasi yang dieksplorasi dari topologi jaringan saat ini.

Referensi

- [1] J. Valverde-Rebaza and A. de Andrade Lopes, "Exploiting behaviors of communities of twitter users for link prediction," *Soc. Netw. Anal. Min.*, vol. 3, no. 4, pp. 1063–1074, 2013, doi: 10.1007/s13278-013-0142-8.
- [2] R. Diestel, *Graph Theory*, Fifth Edit. Springer Nature, 2018.
- [3] P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 1–38, Jan. 2015, doi: 10.1007/s11432-014-5237-y.
- [4] A. K. S. Kushwah and A. K. Manjhar, "A review on link prediction in social network," *Int. J. Grid Distrib. Comput.*, vol. 9, no. 2, pp. 43–50, 2016, doi: 10.14257/ijgcd.2016.9.2.05.
- [5] C. Chen, S. Deng, and J. Lu, "Link prediction in author collaboration network based on BP neural network," *MATEC Web Conf. Conf.*, vol. 139, 2017, doi: 10.1051/mateconf/201713900073.
- [6] H. Yuliansyah, Z. A. Othman, and A. A. Bakar, "Taxonomy of Link Prediction for Social

<http://sistemasi.ftik.unisi.ac.id>

- Network Analysis: A Review,” *IEEE Access*, vol. 8, pp. 183470–183487, 2020, doi: 10.1109/access.2020.3029122.
- [7] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, “Link prediction techniques, applications, and performance: A survey,” *Phys. A Stat. Mech. its Appl.*, vol. 553, p. 124289, 2020, doi: 10.1016/j.physa.2020.124289.
- [8] T. Zhou, L. Lü, and Y. C. Zhang, “Predicting missing links via local information,” *Eur. Phys. J. B*, vol. 71, no. 4, pp. 623–630, 2009, doi: 10.1140/epjb/e2009-00335-8.
- [9] Z. Tao and J. Zhang, “A Survey of Link Prediction in Complex Networks VICTOR,” *Proc. - 2013 4th World Congr. Softw. Eng. WCSE 2013*, vol. 49, no. 4, pp. 217–220, 2013, doi: 10.1109/WCSE.2013.39.
- [10] E. C. Mutlu, T. Oghaz, A. Rajabi, and I. Garibay, “Review on Learning and Extracting Graph Features for Link Prediction,” *Mach. Learn. Knowl. Extr.*, vol. 2, no. 4, pp. 672–704, 2020, doi: 10.3390/make2040036.
- [11] X. Liu, “Full-Text Citation Analysis: A New Method to Enhance,” *J. Am. Soc. Inf. Sci. Technol.*, vol. 64, no. July, pp. 1852–1863, 2013, doi: 10.1002/asi.
- [12] A. P. S. Nisha S Sarma, “Friend Recommendation in KNN Classification,” *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 4, no. 6, pp. 2320–9798, 2016, doi: 10.15680/IJIRCCCE.2016.0406294.
- [13] E. Medina-Acuña, P. Shiguihara-Juárez, and N. Murrugarra-Llerena, “Link prediction in co-authorship networks using scopus data,” *Commun. Comput. Inf. Sci.*, vol. 898, pp. 91–97, 2019, doi: 10.1007/978-3-030-11680-4_10.
- [14] J. chao Li, D. ling Zhao, B. F. Ge, K. W. Yang, and Y. W. Chen, “A link prediction method for heterogeneous networks based on BP neural network,” *Phys. A Stat. Mech. its Appl.*, vol. 495, pp. 1–17, 2018, doi: 10.1016/j.physa.2017.12.018.
- [15] Z. Yang, D. Li, R. Lin, Y. Tang, W. Li, and H. Liu, “An academic social network friend recommendation algorithm based on decision tree,” *Proc. - 2018 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/UIC/ATC/ScalCom/CBDCo*, pp. 1311–1316, 2018, doi: 10.1109/SmartWorld.2018.00228.
- [16] R. A. Rossi and N. K. Ahmed, “NetworkRepository: An Interactive Data Repository with Multi-scale Visual Analytics,” pp. 4292–4293, 2014, [Online]. Available: <http://arxiv.org/abs/1410.3560>.
- [17] K. E. Read, “Cultures of the central highlands, New Guinea,” *Southwest. J. Anthropol.*, vol. 10, no. 1, pp. 1–43, 2015.
- [18] W. W. Zachary, “Flow Model for Conflict An Information Fission in Small Groups,” *Small*, vol. 33, no. 4, pp. 452–473, 2009.
- [19] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, “The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait?,” *Behav. Ecol. Sociobiol.*, vol. 54, no. 4, pp. 396–405, 2003, doi: 10.1007/s00265-003-0651-y.
- [20] D. Shizuka, A. S. Chaine, J. Anderson, O. Johnson, I. M. Laursen, and B. E. Lyon, “Across-year social stability shapes network structure in wintering migrant sparrows,” *Ecol. Lett.*, vol. 17, no. 8, pp. 998–1007, 2014, doi: 10.1111/ele.12304.
- [21] R. E. van Dijk, J. C. Kaden, A. Argüelles-Ticó, D. A. Dawson, T. Burke, and B. J. Hatchwell, “Cooperative investment in public goods is kin directed in communal nests of social birds,” *Ecol. Lett.*, vol. 17, no. 9, pp. 1141–1148, 2014, doi: 10.1111/ele.12320.

Lampiran

Lampiran A. Hasil fitur *training soc-tribes*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	1	7	56	0,5345238	1,987202	4	0
	2	1	8	56	0,725	2,383543	4	0
	3	1	10	40	0,2361111	0,936018	2	0

2	97	15	16	81	0,7539682	2,892611	6	1
	1	1	7	56	0,5345238	1,987202	4	0
	2	1	8	56	0,725	2,383543	4	0
	3	1	9	56	0,4789682	1,884211	4	0
3
	97	14	16	45	0,3928571	1,475695	3	1
	1	1	7	56	0,5345238	1,987202	4	0
	2	1	8	56	0,725	2,383543	4	0
4	3	1	9	56	0,4789682	1,884211	4	0

	97	15	16	81	0,7539682	2,892611	6	1
	1	1	8	56	0,725	2,383543	4	0
5	2	1	9	56	0,4789682	1,884211	4	0
	3	1	10	40	0,2361111	0,936018	2	0

	97	15	16	81	0,7539682	2,892611	6	1

Lampiran B. Hasil fitur *training soc-karate*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	1	34	272	0,9	2,71102	4	0
	2	1	10	32	0,1	0,434294	1	0
	3	1	19	32	0	0	0	0

2	449	30	27	8	0,058823529	0,352956	1	1
	1	1	33	192	0,466666667	1,61374	3	0
	2	1	34	272	0,9	2,71102	4	0
	3	1	10	32	0,1	0,434294	1	0
3
	450	30	27	8	0,058823529	0,352956	1	0
	1	1	33	192	0,466666667	1,61374	3	0
	2	1	10	32	0,1	0,434294	1	0
4	3	1	15	32	0	0	0	0

	450	26	25	9	0,166666667	0,558111	1	0
	1	1	33	192	0,466666667	1,61374	3	0
5	2	1	34	272	0,9	2,71102	4	0
	3	1	10	32	0,1	0,434294	1	0

	450	30	27	8	0,058823529	0,352956	1	0
1	1	33	192	0,466666667	1,61374	3	0	

<http://sistemasi.ftik.unisi.ac.id>

2	1	34	272	0,9	2,71102	4	0
3	1	15	32	0	0	0	0
...
450	30	27	8	0,058823529	0,352956	1	0

Lampiran C. Hasil fitur *training soc-dolphin*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	1	3	24	0,366666	1,1794	2	0
	2	1	4	18	0,0833333	0,4024	1	0
	3	1	6	24	0	0	0	0

2	1510	39	59	8	0	0	0	1
	1	1	2	48	0	0	0	0
	2	1	4	18	0,0833333	0,4024	1	0
	3	1	5	6	0	0	0	0
3
	1514	50	47	4	0	0	0	0
	1	1	2	48	0	0	0	0
	2	1	3	24	0,366666	1,1794	2	0
4	3	1	5	6	0	0	0	0

	1514	50	47	4	0	0	0	0
	1	1	2	48	0	0	0	0
5	2	1	3	24	0,366666	1,1794	2	0
	3	1	4	18	0,0833333	0,4024	1	0

	1514	44	54	14	0	0	0	0

Lampiran D. Hasil fitur *training soc-highschool moreno*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	1	8	21	0,148352	0,768794	2	0
	2	1	10	14	0,2	0,621335	1	0
	3	1	11	42	0	0	0	0

2	1933	66	68	48	0,355409	1,644911	4	1
	1	1	9	63	0,148352	0,768794	2	0
	2	1	10	14	0,2	0,621335	1	0
	3	1	12	63	0	0	0	0
3
	1933	69	70	15	0,236111	0,936018	2	1
	1	1	8	21	0,148352	0,768794	2	0
	2	1	9	63	0,148352	0,768794	2	0
4	3	1	10	14	0,2	0,621335	1	0

	1933	69	70	15	0,236111	0,936018	2	1
	1	1	8	21	0,148352	0,768794	2	0
4	2	1	9	63	0,148352	0,768794	2	0

	3	1	11	42	0	0	0	0

	1933	69	70	15	0,236111	0,936018	2	1
	1	1	8	21	0,148352	0,768794	2	0
	2	1	9	63	0,148352	0,768794	2	0
5	3	1	10	14	0,2	0,621335	1	0

	1933	69	70	15	0,236111	0,936018	2	1

Lampiran E. Hasil fitur *training soc-aves-sparrowlyon-flock-s3*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	1	4	126	0,239379	1,42028	4	0
	2	1	7	144	0,310807	1,799203	5	0
	3	1	9	90	0,419444	1,772742	4	0

	281	26	7	80	0,382236	2,178126	6	1
2	1	1	7	144	0,310807	1,799203	5	0
	2	1	9	90	0,419444	1,772742	4	0
	3	1	10	162	0,378267	2,168686	6	0

	282	26	9	50	0	0	0	0
3	1	1	4	126	0,239379	1,42028	4	0
	2	1	7	144	0,310807	1,799203	5	0
	3	1	9	90	0,419444444	1,772742	4	0

	282	26	9	50	0	0	0	0
4	1	1	4	126	0,239379085	1,42028	4	0
	2	1	7	144	0,310807656	1,799203	5	0
	3	1	10	162	0,378267974	2,168686	6	0

	282	26	9	50	0	0	0	0
5	1	1	4	126	0,239379085	1,42028	4	0
	2	1	9	90	0,419444444	1,772742	4	0
	3	1	12	72	0,232434641	1,405582	4	0

	282	26	9	50	0	0	0	0

Lampiran F. Hasil fitur *training soc-aves-weaver-05*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
1	1	96	98	4	0	0	0	0
	2	96	99	5	0	0	0	0
	3	96	101	4	0	0	0	0

	63	99	102	5	0	0	0	1
2	1	96	98	4	0	0	0	0
	2	96	99	5	0	0	0	0
	3	96	100	4	0	0	0	0

	63	100	101	16	0,65	1,964017	3	1
3	1	96	99	5	0	0	0	0
	2	96	100	4	0	0	0	0
	3	96	102	1	0	0	0	0

...
63	100	101	16	0,65	1,964017	3	1	
1	96	98	4	0	0	0	0	
2	96	99	5	0	0	0	0	
4	3	96	100	4	0	0	0	
...
64	100	101	16	0,65	1,964017	3	0	
1	96	98	4	0	0	0	0	
2	96	100	4	0	0	0	0	
5	3	96	101	4	0	0	0	
...
64	100	101	16	0,65	1,964017	3	0	

Lampiran G. Hasil fitur *testing soc-tribes*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	1	9	56	0,478968	1,884211	4	0
	2	3	11	54	0,385714	1,462091	3	0
1	3	4	14	15	0,142857	0,513898	1	0
...
	25	13	16	72	0,522222	1,965869	4	1
	1	2	11	72	0,665079	2,479767	5	0
	2	3	13	48	0,242857	0,948193	2	0
2	3	4	10	15	0	0	0	0
....
	25	15	16	81	0,753968	2,892611	6	1
	1	1	11	72	0,447222	1,825432	4	0
	2	3	9	42	0,225	0,915193	2	0
3	3	4	5	21	0,125	0,480898	1	0
...
	25	11	15	81	0,703968	2,55215	6	1
	1	1	7	56	0,534524	1,987202	4	0
	2	1	10	40	0,236111	0,936018	2	0
4	3	2	4	24	0,291667	1,039009	2	0
...
	25	14	16	45	0,392857	1,475695	3	1
	1	1	8	56	0,725	2,383543	4	0
	2	1	13	64	0,322222	1,344534	3	0
5	3	1	14	40	0,378968	1,449916	3	0
...
	25	10	15	45	0,503968	1,930815	4	1

Lampiran H. Hasil fitur *testing soc-karate*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	1	33	192	0,466667	1,61374	3	0
	2	1	15	32	0	0	0	0
1	3	1	16	32	0	0	0	0
...
	114	34	30	68	0,783333	2,46646	3	1
	1	1	19	32	0	0	0	0
	2	1	26	48	0,166667	0,558111	1	0
2	3	1	28	64	0,1	0,434294	1	0
....

	113	26	25	9	0,166667	0,558111	1	1
	1	1	34	272	0,9	2,71102	4	0
	2	1	27	32	0	0	0	0
3	3	1	29	48	0,266667	0,992405	2	0

	113	30	27	8	0,058824	0,352956	1	1
	1	1	23	32	0	0	0	0
	2	1	25	48	0,166667	0,558111	1	0
4	3	2	10	18	0,1	0,434294	1	0

	113	26	24	15	0	0	0	1
	1	1	10	32	0,1	0,434294	1	0
	2	1	21	32	0	0	0	0
5	3	1	30	64	0	0	0	0

	113	34	23	34	0,083333	0,40243	1	1

Lampiran I. Hasil fitur *testing soc-dolphin*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	1	2	48	0	0	0	0
	2	1	5	6	0	0	0	0
1	3	1	10	42	0	0	0	0

	380	44	54	14	0	0	0	1
	1	1	3	24	0,366667	1,179446	2	0
	2	1	8	30	0,125	0,480898	1	0
2	3	1	17	36	0,083333	0,40243	1	0

	379	39	44	56	0,183333	0,836724	2	1
	1	1	4	18	0,083333	0,40243	1	0
	2	1	7	36	0	0	0	0
3	3	1	12	6	0	0	0	0

	379	17	39	48	0,294444	1,291844	3	1
	1	1	6	24	0	0	0	0
	2	1	9	36	0	0	0	0
4	3	1	19	42	0,142857	0,513898	1	0

	379	39	59	8	0	0	0	1
	1	1	20	24	0	0	0	0
	2	1	25	36	0,22619	0,916328	2	0
5	3	1	26	18	0	0	0	0

	379	50	47	4	0	0	0	1

Lampiran J. Hasil fitur *testing soc-highschool moreno*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	1	9	63	0,148352	0,768794	2	0
	2	1	12	63	0	0	0	0
1	3	1	14	42	0	0	0	0

	484	69	70	15	0,236111	0,936018	2	1

	1	1	8	21	0,148352	0,768794	2	0
	2	1	11	42	0	0	0	0
2	3	1	16	77	0	0	0	0

	484	65	66	42	0,230409	1,164012	3	1
	1	1	21	63	0,348352	1,390129	3	0
	2	1	28	133	0,076923	0,389871	1	0
3	3	1	30	49	0	0	0	0

	484	66	68	48	0,355409	1,644911	4	1
	1	1	10	14	0,2	0,621335	1	0
	2	1	13	56	0	0	0	0
4	3	1	17	35	0	0	0	0

	484	64	68	56	0,119298	0,708893	2	1
	1	1	15	70	0	0	0	0
	2	1	24	21	0,238095	0,937034	2	0
5	3	1	27	91	0,302198	1,548537	4	0

	484	65	67	63	0,285965	1,267003	3	1

Lampiran K. Hasil fitur *testing soc-aves-sparrowlyon-flock-s3*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	1	12	72	0,232435	1,405582	4	0
	2	1	17	234	0,553863	3,227201	9	0
1	3	1	23	252	0,553863	3,227201	9	0

	72	26	9	50	0	0	0	1
	1	1	4	126	0,239379	1,42028	4	0
	2	2	8	126	0,217893	1,141932	3	0
2	3	2	20	108	0,394841	2,204093	6	0

	71	24	9	45	0,308333	1,317622	3	1
	1	2	7	144	0,559159	3,002292	8	0
	2	3	19	192	0,356151	1,886669	5	0
3	3	3	24	144	0,118056	0,70665	2	0

	71	26	7	80	0,382236	2,178126	6	1
	1	1	9	90	0,419444	1,772742	4	0
	2	2	9	90	0,1	0,434294	1	0
4	3	2	19	216	0,518488	2,682625	7	0

	71	17	23	182	0,932831	4,677117	12	1
	1	1	7	144	0,310808	1,799203	5	0
	2	1	10	162	0,378268	2,168686	6	0
5	3	1	26	180	0,428863	2,505853	7	0

	71	4	17	91	0,421919	2,254323	6	1

Lampiran L. Hasil fitur *testing soc-aves-weaver-05*

Iterasi	Data ke-	Nodes A	Nodes B	Membangkitkan Fitur				Label
				PA	RA	AA	CN	
	1	96	100	4	0	0	0	0
1	2	96	95	5	0	0	0	0

<http://sistemasi.ftik.unisi.ac.id>

	3	99	103	5	0	0	0	0
		
	17	100	101	16	0,65	1,964017	3	1
	1	97	102	2	0	0	0	0
	2	97	93	2	0	0	0	0
2	3	98	94	16	0	0	0	0
		
	17	99	102	5	0	0	0	1
	1	96	98	4	0	0	0	0
	2	96	101	4	0	0	0	0
3	3	96	92	2	0	0	0	0
		
	17	99	101	20	0,7	2,06403	3	1
	1	96	102	1	0	0	0	0
	2	97	100	8	0	0	0	0
4	3	97	101	8	0	0	0	0
		
	16	95	99	25	0,75	2,164043	3	1
	1	96	99	5	0	0	0	0
	2	96	103	1	0	0	0	0
5	3	96	91	4	0	0	0	0
		
	16	98	100	16	0,65	1,964017	3	1

HASIL CEK_Comparison of Classificati

ORIGINALITY REPORT

5%

SIMILARITY INDEX

3%

INTERNET SOURCES

2%

PUBLICATIONS

5%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to STIE Perbanas Surabaya

Student Paper

3%

2

Submitted to Universitas Amikom

Student Paper

1%

3

Submitted to UIN Sunan Ampel Surabaya

Student Paper

1%

4

Submitted to Universitas Andalas

Student Paper

1%

Exclude quotes On

Exclude matches < 1%

Exclude bibliography On