



PRAKTIKUM TEXT MINING

Kelompok 2D :

1. M. Bayu Fadayan (065121100)
2. Fathur Pakapradana S (065121103)
3. Indah Cahyani (065121120)

PERBANDINGAN AKURASI KLASIFIKASI DENGAN METODE NEURAL NETWORK DAN RANDOM FOREST TERHADAP GENRE JURNAL

TEMA DOKUMEN

Abstrak dari jurnal IT maupun NON – IT dengan 6 kategori yaitu Arsitektur Jaringan, Multimedia, Non IT, Pemodelan, Sistem Cerdas, dan Socio Informatics.

1. Arsitektur Jaringan, contoh kata kuncinya yaitu routing, jaringan komputer, area network
2. Multimedia, contoh kata kuncinya yaitu media pembelajaran, animasi, video
3. Non-IT, contoh kata kuncinya yaitu kimia, Sejarah, agama, Pembangunan dan lainnya
4. Pemodelan, contoh kata kuncinya yaitu jaringan saraf, pemodelan spasial, simulasi
5. Sistem Cerdas, contoh kata kuncinya yaitu smart, iot, mikrokontroller
6. Socio Informatics, contoh kata kuncinya yaitu sistem penunjang Keputusan, sistem informasi, sentimen pro

SOURCE DOKUMEN

Data ini diambil dari berbagai macam sumber, seperti google scholar, ataupun website kumpulan jurnal dari suatu universitas. Data-data ini kemudian di masukkan kedalam notepad lalu dikumpulkan dalam sebuah folder dan diberi nama jurnal_abstrak_data

Daftar Dokumen :

https://docs.google.com/spreadsheets/d/1yj9mZEIPuQtf0F2_crTdem6wF2fCXp2vkfhO49_n2j0/edit?usp=sharing

Data Judul Jurnal : <https://s.id/tm-jurnal-title>

Data Abstrak : <https://s.id/tm-jurnal-abstrak-data>

Data Abstrak Prediksi : <https://s.id/tm-jurnal-prediksi>

Video Presentasi : <https://youtu.be/ncIYogv948o>

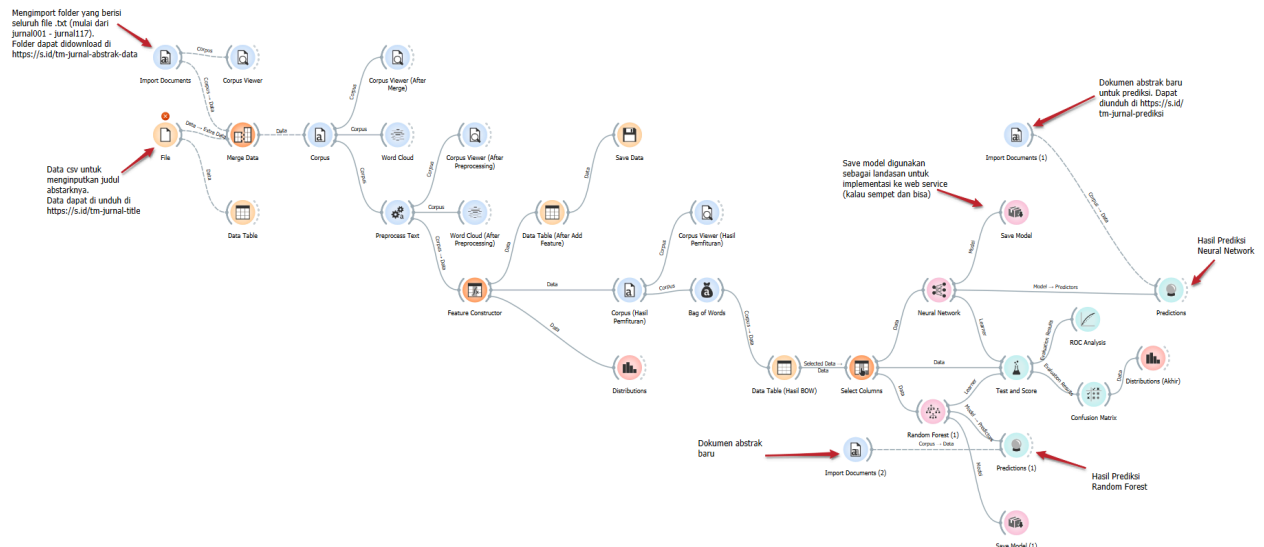
METODE

Metode yang digunakan untuk menganalisis yaitu metode klasifikasi menggunakan Neural Network dan Random Forest pada software orange. Menggunakan 124 dokumen dengan 117 data latih dan 7 data uji.



PRAKTIKUM TEXT MINING

1. Widget yang digunakan : import document, file, corpus view, merge data, data table, corpus, word cloud, preprocess text, feature constructor, save data, distribution, bag of word, select columns, neural network, random forest, test and score, predictions, ROC Analysis, confusion matrix



2. Merge Data : Menggabungkan dokumen abstrak jurnal dengan csv yang berisi judul jurnalnya, dan hasilnya bisa kita lihat menggunakan widget corpus viewer sebagai berikut :

RegExp Filter:	
1	Rancang Bangun Aplikasi Pembelajaran Pengenalan Alat Musik ...
2	Smart Card Berbasis RFID untuk Implementasi E-Parking System d...
3	Pengembangan Media Pembelajaran Interaktif Sistem Tata Surya ...
4	Sistem Pendukung Keputusan Rekomendasi Penginapan di ...
5	Implementasi Internet of Things Pada Sistem Pencarian Parkir ...
6	Penerapan Sistem Pendukung Keputusan pada Penerimaan Santri...
7	Sistem Penunjang Keputusan Pemilihan Pengurus OSIS ...
8	Sistem Pendukung Keputusan Pemilihan Lahan Perkebunan ...
9	Sistem Pendukung Keputusan Rekomendasi Tanaman Obat ...
10	Peramalan Stok Barang Percetakan dan ATK Menggunakan Single...
11	Penerapan Metode Simple Additive Weighting Pada Sistem ...
12	Analisis Sentimen Pembelajaran Tatap Muka dengan Apache SPARK
13	Sistem Informasi Pengelolaan Arsip Surat Pada Dinas ...

name: jurnal001

content: Tujuan dari penelitian ini adalah merancang dan membangun aplikasi pembelajaran pengenalan alat music tradisional bolaang mongondow untuk sekolah menengah kejuruan. Sebelumnya sudah ada yang melakukan perancangan aplikasi pengenalan alat music tradisional namun belum terdapat fitur music tradisional bolaang mongondow. Maka dalam aplikasi ini akan merancang aplikasi pengenalan alat music tradisional bolaang mongondow yang menarik dan mudah. Metode pengembangan rancang bangun aplikasi pembelajaran pengenalan alat music ini menggunakan metode pengembangan MDLC versi Luther-Sutopo dan untuk pengujian aplikasi ini menggunakan metode pengujian white box untuk developer test dan metode pengujian black box untuk end user test. Aplikasi ini menggunakan dasar pemograman Adobe ActionScript 3.0 pada Flash dengan OS android. Hasil dari penelitian ini dimulai dengan menampilkan pilihan menu pembelajaran tebak gambar dan suara alat musik. Kemudian sistem menampilkan pembelajaran sesuai dengan pilihan yang dilakukan user. Sistem menampilkan pembelajaran dengan dukungan gambar disertai suara yang menjelaskan tentang materi pembelajaran yang sedang ditampilkan. Aplikasi pembelajaran berbasis android ini dapat memberikan manfaat bagi siswa-siswi.

3. Preprocess text : menggunakan transformation seperti lowercase dan remove accents, Tokenization Regexp, stopwords Indonesia dan regexp untuk simbol-simbol khusus.



PRAKTIKUM TEXT MINING

4. Menambahkan Feature

```
'Sistem Cerdas' if any(kata in title.lower() for kata in ['smart',
'smart city', 'rfid', 'iot', 'internet of things', 'mikrokontroler',
'sistem cerdas', 'automasi', 'robot']) \

else 'Multimedia' if any(kata in title.lower() for kata in ['aplikasi
pembelajaran', 'media pembelajaran', 'augmented reality', 'vr',
'virtual reality', 'video', 'google sites', '3d', 'animasi',
'vilamil', 'game', 'permainan', 'ui/ux']) \

else 'Pemodelan' if any(kata in title.lower() for kata in
['penyebaran', 'jaringan saraf', 'jaringan saraf tiruan', 'Jaringan
Saraf Tiruan', 'saraf', 'naive bayes', 'naive', 'classifier',
'simulasi', 'pemodelan spasial', 'jst', 'jaringan saraf tiruan',
'matematika', 'model cellular automata', 'numerik', 'susceptible
infective', 'seirs', 'algoritma genetika', 'analisis arsitektur
jaringan saraf tiruan', 'peramalan penjualan']) or any(kata in
content.lower() for kata in ['mae', 'chart']) \

else 'Arsitektur Jaringan' if any(kata in title.lower() for kata in
['routing', 'arsitektur jaringan', 'ppdioo', 'jaringan komputer',
'network design', 'local area network', 'ssl', 'keamanan', 'area
network']) \

else 'Socio Informatics' if any(kata in title.lower() for kata in
['sistem pendukung keputusan', 'rekomendasi', 'vikor', 'saw', 'simple
additive', 'sistem penunjang keputusan', 'pemilihan',
'Višekriterijumsko Kompromisno', 'roc-moor', 'sentimen
pembelajaran', 'sistem informasi', 'ahp', 'wp', 'pelatihan aplikasi',
'geografis', 'sig', 'proses bisnis', 'bahasa', 'use case', 'uml',
'waterfall', 'sentimen pro']) or 'antrian' in content.lower() \

else 'Non IT' if any(kata in title.lower() for kata in ['kimia',
'sejarah', 'agama', 'pembangunan']) \

else 'Non IT'
```

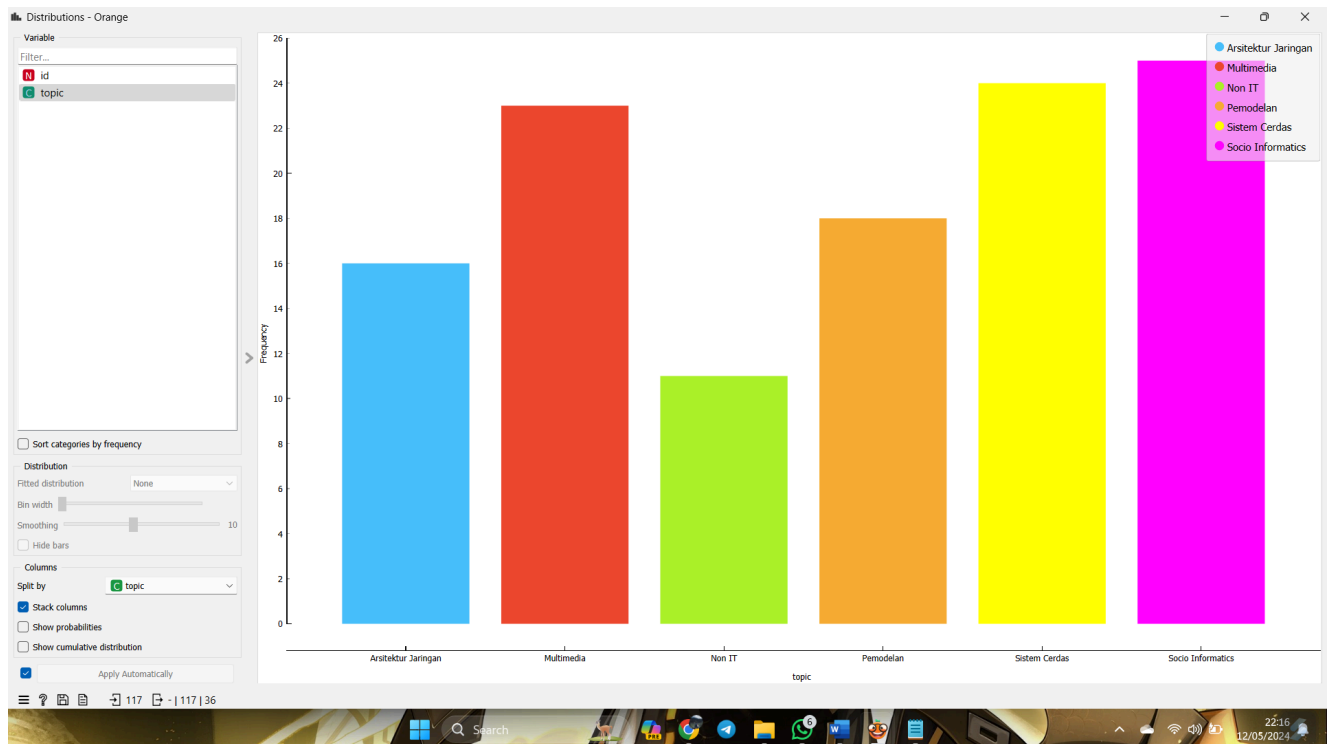
Hasil penambahan feature :



PRAKTIKUM TEXT MINING

id	name	path	content	title True	topic
1	jurnal001	C...	Tujuan dari pen...	Rancang Bangun Aplikasi Pembelajaran Pengenalan Alat Musik Tradisional Bolaang Mongondow Berbasis Mobile	Multimedia
2	jurnal002	C...	Smart Card ada...	Smart Card Berbasis RFID untuk Implementasi E-Parking System di Rumah Sakit Zahirah	Sistem Cerdas
3	jurnal003	C...	Sekolah dasar b...	Pengembangan Media Pembelajaran Interaktif Sistem Tata Surya Menggunakan Metode Multimedia Development Life Cycle	Multimedia
4	jurnal004	C...	Guci merupaka...	Sistem Pendukung Keputusan Rekomendasi Penginapan di Kawasan Pariwisata Guci Menggunakan Metode Vikor	Socio Informatics
5	jurnal005	C...	Penelitian ini be...	Implementasi Internet of Things Pada Sistem Pencarian Parkir Berbasis Mikrokontroler Node-MCU	Sistem Cerdas
6	jurnal006	C...	Dalam menent...	Penerapan Sistem Pendukung Keputusan pada Penerimaan Santri Baru dengan Metode Simple Additive Weighting (SAW) pada MTS Al Mahsyar Nurul Iman Tenggarong	Socio Informatics
7	jurnal007	C...	Organisasi Sisw...	Sistem Penunjang Keputusan Pemilihan Pengurus OSIS menggunakan Metode Višekriterijumsko Kompromisno Rangiranje (VIKOR)	Socio Informatics
8	jurnal008	C...	Sumber daya la...	Sistem Pendukung Keputusan Pemilihan Lahan Perkebunan Tanaman Lada Menggunakan Metode Simple Additive Weighting Perbandingan Weighted Product Berbasis ...	Socio Informatics
9	jurnal009	C...	Indonesia mem...	Sistem Pendukung Keputusan Rekomendasi Tanaman Obat Menggunakan ROC-MOORA	Socio Informatics
10	jurnal010	C...	Setiap badan u...	Peramalan Stok Barang Percetakan dan ATK Menggunakan Single Moving Average	Pemodelan
11	jurnal011	C...	Beasiswa dapat ...	Penerapan Metode Simple Additive Weighting Pada Sistem Pendukung Keputusan Pemberian Beasiswa di SMP Budi Luhur	Socio Informatics
12	jurnal012	C...	Menteri Pendi...	Analisis Sentimen Pembelajaran Tatap Muka dengan Apache SPARK	Socio Informatics
13	jurnal013	C...	Proses pengarsi...	Sistem Informasi Pengelolaan Arsip Surat Pada Dinas Pemberdayaan Masyarakat Dan Pemerintahan Desa Berbasis Web	Socio Informatics
14	jurnal014	C...	Perkembangan ...	Implementasi Algoritma K-Means Untuk Analisis Penyebaran Kasus Konfirmasi Covid-19 di Kota Samarinda	Pemodelan
15	jurnal015	C...	Pelayanan kepa...	Sistem Informasi Administrasi Data Kependudukan Berbasis Web Pada Kantor Desa Benua Puhun	Socio Informatics
16	jurnal016	C...	Saat ini perkem...	Analisis Sistem Informasi Penjualan Toko Pakaian Happiness Menggunakan Soft System Methodology	Socio Informatics
17	jurnal017	C...	Keberadaan dip...	Mobile Augmented Reality Pengenalan Tanaman Dipterocarpaceae	Multimedia
18	jurnal018	C...	Tujuan peneltia...	Pengembangan Media Pembelajaran IPS Di Sekolah Menengah Pertama	Multimedia
19	jurnal019	C...	Sistem Informa...	Implementasi Contents Management System (CMS) Dalam Analisa dan Perancangan Sistem Informasi Kepegawaian Universitas Mulawarman	Socio Informatics
20	jurnal020	C...	Mandau adalah...	Klasifikasi Senjata Tradisional Mandau Suku Dayak Menggunakan Metode Naive Bayes Classifier Berdasarkan Fitur Bentuk dan Tekstur	Pemodelan
21	jurnal021	C...	Informasi tenta...	Implementasi Teorema Naive Bayes Pada Prediksi Prestasi Mahasiswa	Pemodelan
22	jurnal022	C...	Penentuan dos...	Sistem Pendukung Keputusan Penentuan Dosen Pembimbing Skripsi Menggunakan Metode Analytical Hierarchy Process (AHP) dan Weighted Product (WP)	Socio Informatics
23	jurnal023	C...	Dengan adanya...	Rancang Bangun Media Pembelajaran Elektronika Analog Untuk Memahami Fungsi dan Karakteristik Op-Amp LM741	Multimedia
24	jurnal024	C...	Video Dokume...	Pengembangan Video Dokumenter (Kegiatan Mahasiswa) Di Lingkungan Informatika Universitas Mulawarman	Multimedia
25	jurnal025	C...	Hidroponik mer...	Video Company Profile Sebagai Media Promosi Sayuran Hidroponik Boekha Farm Samarinda	Multimedia
26	jurnal026	C...	Penggunaan da...	ANALISIS KINERJA METODE ROUTING OSPF DAN RIP PADA MODELARSITEKTUR JARINGAN DI SMKN XYZ	Arsitektur Jarin...
27	jurnal027	C...	Perkembangan ...	PERANCANGAN PENGEMBANGAN DESAIN ARSITEKTUR JARINGAN MENGGUNAKAN METODE PPIDIO	Arsitektur Jarin...
28	jurnal028	C...	Dalam penyusu...	PELATIHAN PEMBANGUNAN JARINGAN KOMPUTER DALAM PERENCANAAN ARSITEKTUR JARINGAN KOMPUTER KABUPATEN BANYUMAS	Arsitektur Jarin...
29	jurnal029	C...	Jaringan komp...	TOP DOWN NETWORK DESIGN DALAM PERANCANGAN JARINGAN KOMPUTER PADA SMA NEGERI 1 INDRALAYA SELATAN	Arsitektur Jarin...
30	jurnal030	C...	Kampus merup...	Perancangan Arsitektur Jaringan Kampus Menggunakan Metode Network Development Life Cycle	Arsitektur Jarin...
31	jurnal031	C...	PT. Lintas Jaring...	Pendampingan PenerapanJaringan Komputer Local Area Network Pada PT. Lintas Jaringan Nusantara Layanan Pali	Arsitektur Jarin...
32	jurnal032	C...	jaringan kompu...	Perancangan Arsitektur Jaringan Komputer Pada Ruang Laboratorium dan Administrasi Pada SMK Negeri 7 Mataram	Arsitektur Jarin...
33	jurnal033	C...	Jaringan komp...	Penerapan Metode Rekayasa Sistem Jaringan Komputer dalam Merancang Blueprint Jaringan Komputer (Studi Kasus: Hotel Maxone Palembang)	Arsitektur Jarin...
34	jurnal034	C...	Berkembangny...	Network Development Life Cycle (NDLC) Dalam Perancangan Jaringan Komputer Pada Rumah Shalom Mahanaim	Arsitektur Jarin...
35	jurnal035	C...	Pada umumnya...	Perancangan Arsitektur Jaringan Sekolah Menengah Kejuruan Dengan Menerapkan Konsep The Dude Server	Arsitektur Jarin...
36	jurnal036	C...	Ketidakpahama...	PELATIHAN APLIKASI PERMODELAN KOMPUTER BERBASIS RASHI MODEL UNTUK ASSESMENT PENDIDIKAN BAGI GURU-GURU DI PONPES NURUL HARAMAIN PUTRA ...	Socio Informatics

Dengan distribusi tiap kategori adalah sebagai berikut :



5. Bag of Word untuk mengubah data teks menjadi vector agar dipahami oleh komputer dan dihasilkan data sebagai berikut :



PRAKTIKUM TEXT MINING

	d (1)	name	path	content (1)	title	
bow-feature hidden skip-normalization					True	{...}
1	1	jurnal001	C...	Tujuan dari pen...	Rancang Bangun...	topic=Multimedia, =0.639045, =0.0858374, 0=2.05412, 3=2.12312, actionsript=4.76217, ada=1.29644, adalah=0.499494, adobe=3.66356, akan=0.977984, a...
2	2	jurnal002	C...	Smart Card ada...	Smart Card Ber...	topic=Sistem Cerdas, (=0.418369, =1.32819, =0.920739, =1.27809, =0.0686699, 2=1.81773, 52=4.76217, =4.39445, adalah=0.499494, agar=1.67113, akan=...
3	3	jurnal003	C...	Sekolah dasar b...	Pengembangan...	topic=Multimedia, %=5.4532, (=3.34695, =1.23808, =1.109861, =1.32819, =0.0708261, =0.0686699, 33=4.76217, 6=2.45959, 75=3.15274, 83=3.37588, 85=9...
4	4	jurnal004	C...	Guci merupaka...	Sistem Pendid...	topic=Socio Informatics, =0.424956, =0.319523, =0.0772537, adalah=0.499494, air=4.91918, akan=1.95597, akomodasi=4.76217, alam=3.66356, alamnya=4...
5	5	jurnal005	C...	Penelitian ini be...	Implementasi I...	topic=Sistem Cerdas, (=0.418369, =1.32819, =1.91714, =0.0600862, 5=8.21649, adalah=0.998988, adanya=1.81773, akan=3.91194, antara=1.81773, api=8.1...
6	6	jurnal006	C...	Dalam menient...	Penerapan Siste...	topic=Socio Informatics, (=0.836737, =0.619039, =1.32819, =0.283304, =0.319523, =0.0257512, additive=7.32712, akan=2.93395, al=3.66356, alat=1.92896...
7	7	jurnal007	C...	Organisasi Sisw...	Sistem Penunja...	topic=Socio Informatics, %=2.27727, (=2.09184, =2.47616, =1.32819, =0.424956, =0.639045, =0.0600862, 2=12.7241, 48=4.76217, 59=4.06903, 68=4.7621...
8	8	jurnal008	C...	Sumber daya la...	Sistem Pendid...	topic=Socio Informatics, (=0.836737, =0.619039, =1.32819, =0.35413, =0.639045, =0.0600862, /=1.8718, adalah=0.499494, additive=3.66356, additive=4.76...
9	9	jurnal009	C...	Indonesia mem...	Sistem Pendid...	topic=Socio Informatics, (=0.836737, =1.23808, =1.34569, =0.639045, =0.111589, 0=2.05412, 459=4.76217, adalah=0.998988, akan=4.76217, alergi=4.76217...
10	10	jurnal010	C...	Setiap badan u...	Peramalan Stok...	topic=Pemodelan, %=1.81773, (=2.09184, =0.619039, =6.59167, =1.32819, =0.637434, =0.319523, =0.0600862, 0342=4.76217, 0667=4.76217, 1=1.92896...
11	11	jurnal011	C...	Beasiswa dapat ...	Penerapan Met...	topic=Socio Informatics, (=1.25511, =1.23808, =1.32819, =0.141652, =0.0600862, /=1.8718, adalah=0.998988, additive=7.32712, akan=1.95597, atau=2.932...
12	12	jurnal012	C...	Menteri Pendi...	Analisis Sentim...	topic=Socio Informatics, %=3.63547, %=2.27727, (=0.418369, =1.32819, =0.424956, =0.111589, 13=3.66356, 2=1.81773, 20=3.66356, 2021=3.37588, 31=3.6...
13	13	jurnal013	C...	Proses pengars...	Sistem Informa...	topic=Socio Informatics, %=0.0429187, adalah=0.499494, arsip=8.13805, belum=1.5433, berbasis=2.59288, cara=1.76644, dalam=0.334108, dan=0.0600862, da...
14	14	jurnal014	C...	Perkembangan ...	Implementasi A...	topic=Pemodelan, (=0.418369, =0.619039, =0.212478, =3.51475, =0.0686699, 19=22.5301, 2=1.81773, 2021=3.37588, 3=2.12312, 31=3.66356, 45=4.76217, ...
15	15	jurnal015	C...	Pelayanan kepa...	Sistem Informa...	topic=Socio Informatics, (=0.637434, =0.639045, =0.034335, ada=1.29644, adalah=1.49848, adanya=3.63547, administrasi=8.91124, aktifitas=3.66356, arsip=...
16	16	jurnal016	C...	Saat ini perkem...	Analisis Sistem...	topic=Socio Informatics, (=0.418369, =1.32819, =0.708261, =0.0772537, ada=1.29644, adalah=0.998988, aktivitas=3.37588, akurat=5.94083, analisis=1.3609...
17	17	jurnal017	C...	Keberadaan dip...	Mobile Augme...	topic=Multimedia, =0.779087, =0.319523, =0.0600862, 0=4.10825, 1080=4.76217, 12=2.81626, 2400=4.76217, 3=4.24623, 6=2.45959, adapun=2.81626, andr...
18	18	jurnal018	C...	Tujuan penelia...	Pengembangan...	topic=Multimedia, (=2.92858, =0.619039, =1.109861, =1.32819, =0.0708261, =0.319523, =0.0171675, adalah=0.998988, aplikasi=3.34226, assembly=4.069...
19	19	jurnal019	C...	Sistem Informa...	Implementasi C...	topic=Socio Informatics, =0.836737, =1.23808, =0.495782, =0.639045, =0.0429187, ada=1.29644, akan=1.95597, analisa=3.37588, animasi=3.66356, baik=1...
20	20	jurnal020	C...	Mandau adalah...	Klasifikasi Senja...	topic=Pemodelan, %=3.63547, %=2.27727, =0.566608, =1.27809, =0.0858374, 12=2.81626, 30=3.37588, 4=2.19722, 5=2.05412, 70=4.06903, 87=4.06903, ad...
21	21	jurnal021	C...	Informasi tenta...	Implementasi T...	topic=Pemodelan, %=2.27727, =0.212478, =0.319523, =0.0429187, 40=4.06903, 5=2.05412, 77=4.76217, adalah=0.499494, akademik=13.5035, akurasi=2.81...
22	22	jurnal022	C...	Penentuan dos...	Sistem Pendid...	topic=Socio Informatics, (=0.836737, =1.23808, =0.35413, =0.0772537, admin=3.37588, ahp=10.9907, akan=0.977984, alternatif=10.7309, analytical=4.06903...
23	23	jurnal023	C...	Dengan adanya...	Rancang Bangun...	topic=Multimedia, %=7.27094, %=10.9907, =4.55453, (=1.25511, =1.23808, =1.32819, =0.566608, =2.55618, =0.0429187, 0=4.10825, 1=3.85792, 10=5.9...
24	24	jurnal024	C...	Video Dokume...	Pengembangan...	topic=Multimedia, (=1.25511, =1.85712, =0.141652, =0.639045, =0.0600862, /=3.7436, ada=1.29644, akan=1.95597, apa=2.97041, beberapa=1.20683, belaj...
25	25	jurnal025	C...	Hidroponik mer...	Video Compan...	topic=Multimedia, =0.566608, =0.319523, =0.0944212, ada=1.29644, agar=3.34226, baru=2.12312, bayam=4.76217, beberapa=2.41365, berkualitas=4.06903...
26	26	jurnal026	C...	Penggunaan da...	ANALISIS KINA...	%=7.32712, =0.708261, =0.0515025, 0=4.10825, 006=4.76217, 156=4.76217, 16=4.06903, 163=4.76217, 23=4.76217, 4=2.19722, 7=3.15274, 9=2.97041, adal...
27	27	jurnal027	C...	Perkembangan ...	PERANCANGA...	(=0.836737, =0.619039, =1.23808, =0.283304, =0.0429187, adalah=0.499494, adanya=1.81773, administrasi=5.94083, arsitektur=3.63547, ataupun=2.97041...
28	28	jurnal028	C...	Dalam penyusu...	PELATIHAN PE...	(=1.67347, =1.23808, =2.65637, =0.283304, =0.639045, =0.034335, /=1.8718, ada=1.29644, agar=3.34226, akan=1.95597, akurasi=2.81626, anggaran=4.06...
29	29	jurnal029	C...	Jaringan kompu...	TOP DOWN NE...	(=0.836737, =1.23808, =0.849913, =0.319523, =0.0171675, 1=5.78688, address=4.76217, agar=1.67113, akhir=3.15274, akses=4.06903, aktivitas=3.37588, a...
30	30	jurnal030	C...	Kampus merup...	Perancangan Ar...	=0.141652, =0.0686699, ada=2.59288, adalah=0.499494, analisis=1.36098, antaranya=4.06903, arsitektur=5.4532, bandwidth=3.37588, berhasil=2.68273, bert...
31	31	jurnal031	C...	PT. Lintas Jaring...	Pendampingan ...	=0.779087, =0.319523, =0.0944212, akan=0.977984, analisis=1.36098, area=2.97041, arsitektur=1.81773, bagi=1.36098, bahwa=1.17865, belum=3.0866, bert...
32	32	jurnal032	C...	jaringan kompu...	Perancangan Ar...	=0.779087, =0.639045, =0.0429187, 7=6.30547, adalah=0.499494, administrasi=2.97041, analisis=1.36098, arsitektur=7.27094, beberapa=4.8273, dan=0.025...
33	33	jurnal033	C...	Jaringan kompu...	Penerapan Met...	(=0.836737, =0.619039, =1.23808, =0.0708261, =0.639045, =0.0429187, =2.19722, ada=1.29644, agar=1.67113, akhirnya=3.66356, alat=3.85792, arsitektur...
34	34	jurnal034	C...	Berkembangny...	Network Devel...	(=0.836737, =0.619039, =1.32819, =0.708261, =0.0429187, ada=1.29644, adanya=1.81773, adapun=2.81626, agar=1.67113, analisis=1.36098, autentikasi=4...

6. Menggunakan Select Columns untuk memilih kolom yang akan digunakan untuk klasifikasi

Select Columns - Orange

Ignored (178)

Filter

N 000437

N 001

N 006

N 01

N 0342

N 04

N 053

N 0667

N 070

N 084

N 09

N 1

N 10

N 100

N 102

N 1080

N 10kw

N 11

N 114

N 116

N 12

N 12767

N 12880

N 13

Features (3497)

Filter

N abc

N abilitas

N absolute

N ac

N actionsript

N activity

N actual

N acuan

N ada

N adab

N adalah

N adanya

N adaptif

Target (1)

C topic

Metas (4)

S name

S path

S content (1)

S title

Reset

☐ Ignore new variables by default

☒ Send Automatically

7. Metode Klasifikasi

a. Neural Network :

Neurons in hidden layers : 1000

Activation : logistic

Solver : adam

Regularization : $\alpha = 0.0002$

Max iterations : 500



PRAKTIKUM TEXT MINING

Replicable training

Neural Network - Orange

Name: Neural Network

Neurons in hidden layers: 1000,

Activation: Logistic

Solver: Adam

Regularization, $\alpha=0.0002$: [Slider]

Maximal number of iterations: 500

☒ Replicable training

Cancel ☒ Apply Automatically

b. Random Forest

Number of trees : 1000

Replicable Training

Do not split subsets smaller than : 5

Random Forest (1) - Orange

Name: Random Forest

Basic Properties

Number of trees: 1000

☐ Number of attributes considered at each split: 5

☒ Replicable training

☐ Balance class distribution

Growth Control

☐ Limit depth of individual trees: 3

☒ Do not split subsets smaller than: 5

☒ Apply Automatically



PRAKTIKUM TEXT MINING

HASIL

Test and Score

Test and Score - Orange						
<div><input checked="" type="radio"/> Cross validation Number of folds: 10 <input checked="" type="checkbox"/> Stratified <input type="radio"/> Cross validation by feature <input type="radio"/> Random sampling Repeat train/test: 50 Training set size: 75 % <input checked="" type="checkbox"/> Stratified <input type="radio"/> Leave one out <input type="radio"/> Test on train data <input type="radio"/> Test on test data</div>						
Evaluation results for target (None, show average over classes)						
Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.943	0.726	0.710	0.788	0.726	0.675
Neural Network	0.925	0.684	0.665	0.727	0.684	0.619

1. AUC (Area Under the Curve):

Neural Network: 0.925

Random Forest: 0.943

AUC mengukur kemampuan model untuk membedakan antara kelas positif dan negatif. Nilai yang lebih tinggi menunjukkan performa yang lebih baik. Kedua model memiliki AUC yang sangat baik, dengan Random Forest sedikit lebih unggul.

2. CA (Classification Accuracy):

Neural Network: 0.684

Random Forest: 0.726

CA adalah persentase prediksi yang benar dari total prediksi. Random Forest memiliki akurasi yang lebih tinggi dibanding Neural Network.

3. F1 Score:

Neural Network: 0.665

Random Forest: 0.710

F1 Score adalah rata-rata harmonik dari presisi dan recall. Skor F1 lebih tinggi pada Random Forest menunjukkan keseimbangan yang lebih baik antara presisi dan recall.

4. Prec (Precision):

Neural Network: 0.727

Random Forest: 0.788

Precision mengukur keakuratan positif yang diprediksi oleh model. Random Forest lebih presisi dalam memprediksi kelas positif.

5. Recall:

Neural Network: 0.684

Random Forest: 0.726

Recall adalah persentase dari kelas positif sebenarnya yang diprediksi dengan benar oleh model. Sekali lagi, Random Forest unggul.

6. MCC (Matthews Correlation Coefficient):

Neural Network: 0.619

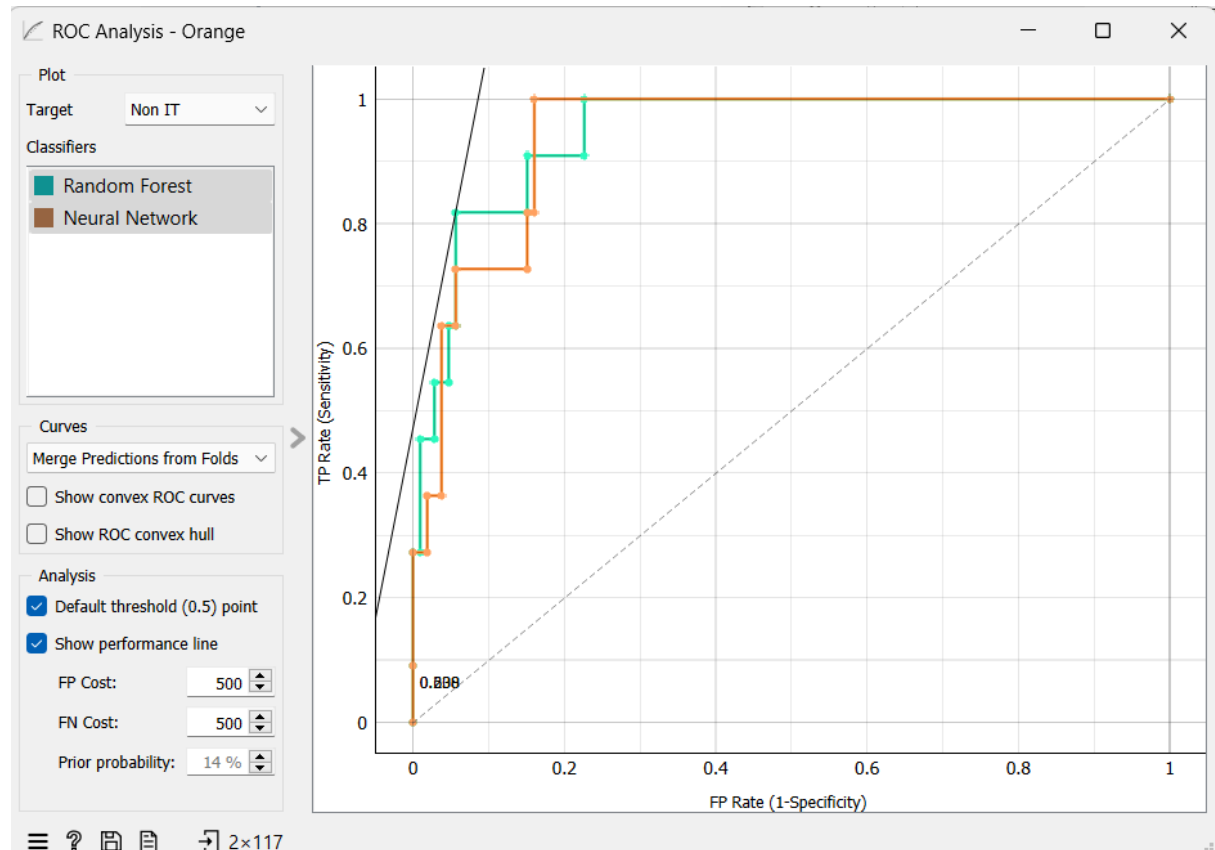
Random Forest: 0.675



PRAKTIKUM TEXT MINING

MCC adalah ukuran korelasi antara nilai observasi dan prediksi. Nilai yang lebih tinggi menunjukkan prediksi yang lebih baik. Random Forest memiliki MCC lebih tinggi, menunjukkan korelasi yang lebih kuat.

ROC Analysis



1. Kurva ROC:

Kurva yang lebih dekat ke pojok kiri atas grafik menunjukkan performa yang lebih baik. Dalam grafik ini, kurva Random Forest (hijau) tampak lebih dekat ke sudut kiri atas dibandingkan dengan kurva Neural Network (oranye), yang menunjukkan bahwa model Random Forest memiliki performa yang lebih baik dalam membedakan antara kelas positif dan negatif untuk dataset 'Non IT'.

2. Garis Diagonal:

Garis diagonal mewakili performa acak. Model yang kurvanya lebih dekat ke garis ini memiliki performa yang mendekati acak, yang bukan indikator model yang baik.

3. Titik 0.5:

Titik yang ditandai (0.088) pada sumbu x (FPR) kemungkinan menunjukkan threshold default (0.5) yang digunakan dalam evaluasi model. Lokasi titik ini pada kurva Random Forest menunjukkan bahwa pada threshold ini, model memiliki sensitivitas tinggi dengan tingkat FPR yang rendah.



PRAKTIKUM TEXT MINING

Confusion matrix Random forest

		Predicted						Σ
		Arsitektur Jaringan	Multimedia	Non IT	Pemodelan	Sistem Cerdas	Socio Informatics	
Actual	Arsitektur Jaringan	83.3 %	0.0 %	0.0 %	0.0 %	0.0 %	2.4 %	16
	Multimedia	0.0 %	95.0 %	0.0 %	0.0 %	0.0 %	9.8 %	23
	Non IT	0.0 %	0.0 %	100.0 %	0.0 %	12.5 %	17.1 %	11
	Pemodelan	11.1 %	0.0 %	0.0 %	76.9 %	0.0 %	14.6 %	18
	Sistem Cerdas	5.6 %	0.0 %	0.0 %	0.0 %	83.3 %	7.3 %	24
	Socio Informatics	0.0 %	5.0 %	0.0 %	23.1 %	4.2 %	48.8 %	25
Σ		18	20	1	13	24	41	117

1. Diagonal Matrix (Warna Ungu):
Nilai pada diagonal menunjukkan persentase prediksi yang benar untuk setiap kelas. Misalnya, untuk kelas 'Arsitektur Jaringan', model berhasil memprediksi 83.3% kasus dengan benar, dan untuk kelas 'Multimedia', model berhasil memprediksi 95.0% kasus dengan benar.
2. Baris:
Setiap baris merepresentasikan kelas aktual dari data. Angka di sebelah kanan baris (dalam persentase) menunjukkan distribusi prediksi untuk kelas tersebut. Misalnya, untuk kelas 'Non IT', 100.0% prediksi adalah benar (Non IT), tetapi ada juga sebagian kecil prediksi yang salah masuk ke kategori 'Sistem Cerdas' dan 'Socio Informatics'.
3. Kolom:
Setiap kolom menunjukkan prediksi model untuk kelas tertentu. Angka di bagian bawah kolom menunjukkan jumlah total prediksi untuk kategori tersebut.
4. Analisis Kesalahan:
Kolom dan baris yang tidak diagonal menunjukkan kesalahan klasifikasi. Misalnya, 11.1% dari data aktual 'Pemodelan' diprediksi sebagai 'Arsitektur Jaringan', dan 5.6% dari data aktual 'Sistem Cerdas' diprediksi sebagai 'Arsitektur Jaringan'.
5. Total dan Persentase:
Di bagian kanan bawah matrix, Σ menunjukkan total sampel untuk masing-masing kelas (baik prediksi maupun aktual), dan nilai di bawah ini (dalam contoh, 117) adalah total keseluruhan sampel yang diprediksi dan diuji.



PRAKTIKUM TEXT MINING

Confusion matrix Neural Network

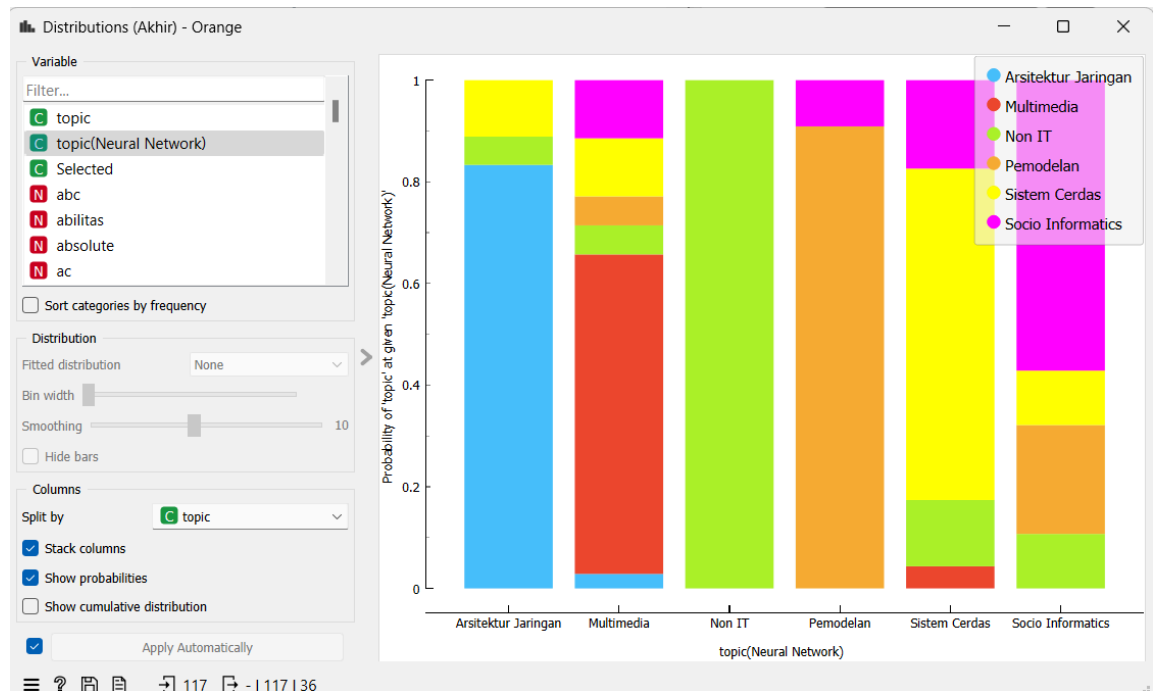
		Predicted						
		Arsitektur Jaringan	Multimedia	Non IT	Pemodelan	Sistem Cerdas	Socio Informatics	Σ
Actual	Arsitektur Jaringan	83.3 %	2.9 %	0.0 %	0.0 %	0.0 %	0.0 %	16
	Multimedia	0.0 %	62.9 %	0.0 %	0.0 %	4.3 %	0.0 %	23
	Non IT	5.6 %	5.7 %	100.0 %	0.0 %	13.0 %	10.7 %	11
	Pemodelan	0.0 %	5.7 %	0.0 %	90.9 %	0.0 %	21.4 %	18
	Sistem Cerdas	11.1 %	11.4 %	0.0 %	0.0 %	65.2 %	10.7 %	24
	Socio Informatics	0.0 %	11.4 %	0.0 %	9.1 %	17.4 %	57.1 %	25
Σ		18	35	2	11	23	28	117

- Diagonal Matrix (Warna Ungu):**
 Angka pada diagonal menunjukkan persentase prediksi yang benar per kelas.
 Contohnya:
 Multimedia: 62.9% keberhasilan prediksi dengan benar.
 Non IT: 100% keberhasilan prediksi dengan benar.
 Pemodelan: 90.9% keberhasilan prediksi dengan benar.
 Socio Informatics: 57.1% keberhasilan prediksi dengan benar.
- Baris (Kelas Aktual):**
 Setiap baris menggambarkan kelas aktual dan bagaimana prediksi model tersebar di berbagai kelas yang diprediksi.
 Misalnya, untuk kelas 'Sistem Cerdas', 65.2% dari data aktual diprediksi dengan benar, tetapi ada kesalahan signifikan dimana 11.1% dan 11.4% dari data tersebut salah diprediksi sebagai 'Arsitektur Jaringan' dan 'Multimedia' secara berturut-turut.
- Kolom (Prediksi Model):**
 Setiap kolom menunjukkan jumlah prediksi yang model lakukan untuk setiap kelas. Kolom yang lebih lebar menunjukkan jumlah prediksi yang lebih banyak untuk kelas tersebut, yang bisa jadi menandakan bias pada model atau distribusi data yang tidak merata.
- Analisis Kesalahan:**
 Area di luar diagonal menunjukkan kesalahan klasifikasi. Model ini tampaknya memiliki beberapa tantangan dalam membedakan antara beberapa kelas.
 Misalnya, sejumlah besar data 'Sistem Cerdas' diprediksi sebagai 'Arsitektur Jaringan' atau 'Multimedia'.
- Summarization:**
 Σ (Sigma) menunjukkan jumlah total prediksi untuk setiap kelas dan total keseluruhan.
 Dapat dilihat bahwa beberapa kelas seperti 'Multimedia' dan 'Sistem Cerdas' memiliki jumlah prediksi yang tinggi dibandingkan yang lain.



PRAKTIKUM TEXT MINING

Distribusi Akhir



Prediksi Neural Network

	name	path	content
1	Jurnal_pred1	C:/Users/Asus V...	Pandemi Covid-19 yang sudah berlangsung sejak awal tahun 2020 memberikan dampak besar di berbagai sektor, salah satunya di sektor pendidikan, dimana awalnya pendidikan dilaku...
2	Jurnal_pred2	C:/Users/Asus V...	Hutan gambut di Kalimantan adalah salah satu ekosistem hidrologi penting di dunia yang memainkan peran vital dalam menjaga keseimbangan lingkungan. Hutan ini memfasilitasi siki...
3	Jurnal_pred3	C:/Users/Asus V...	Sampah merupakan masalah di seluruh dunia, terutama di negara berkembang. Masyarakat seringkali tidak memilah sampah dengan benar, sehingga sampah yang seharusnya dapat di...
4	Jurnal_pred4	C:/Users/Asus V...	Piramida Kecelakaan Kerja atau dikenal juga dengan Heinrich's triangle atau Bird's triangle menyatakan bahwa kecelakaan kerja dapat ditimbulkan karena beberapa kejadian-kejadian ke...
5	Jurnal_pred5	C:/Users/Asus V...	Prediksi keberlangsungan hidup pasien gagal jantung telah dilakukan pada penelitian untuk mencari tahu tentang kinerja, akurasi, presisi dan performa dari model prediksi ataupun met...
6	Jurnal_pred6	C:/Users/Asus V...	Bantuan Langsung Tunai (BLT) adalah bantuan yang dibuat oleh pemerintah untuk orang atau keluarga yang kurang mampu dan rentan akan kemiskinan dengan mereka yang memenu...
7	Jurnal_pred7	C:/Users/Asus V...	Kucing merupakan salah satu hewan yang banyak dipelihara oleh masyarakat. Penyakit kulit pada kucing seringkali membuat pemiliknya merasa bingung dan terkendala dalam diagnos...

1. Jurnal_pred1: Prediksi tertinggi untuk kelas 'Multimedia' dengan probabilitas 58.1%. Ini menunjukkan bahwa model paling yakin bahwa sampel ini termasuk dalam kelas 'Multimedia'.
2. Jurnal_pred2: Prediksi tertinggi untuk kelas 'Pemodelan' dengan probabilitas 57.3%.
3. Jurnal_pred3: Prediksi tertinggi untuk kelas 'Sistem Cerdas' dengan probabilitas 97.0%, yang sangat tinggi, menunjukkan kepercayaan yang tinggi dari model terhadap prediksi ini.



PRAKTIKUM TEXT MINING

4. Jurnal_pred4: Prediksi tertinggi lagi untuk kelas 'Sistem Cerdas' dengan probabilitas 59.0%.
5. Jurnal_pred5 : Prediksi tertinggi untuk kelas 'Pemodelan' dengan probabilitas 56.0%
6. Jurnal_pred6 : Prediksi tertinggi lagi untuk kelas 'Pemodelan' dengan probabilitas 32.0%.
7. Jurnal_pred7 : Prediksi tertinggi untuk kelas 'Sistem Cerdas' dengan probabilitas 59.0%.

Prediksi Random Forest

	name	path	content
1 Pemodelan	jurnal_pred1	C:/Users/Asus V...	Pandemi Covid-19 yang sudah berlangsung sejak awal tahun 2020 memberikan dampak besar di berbagai sektor, salah satunya di sektor pendidikan, dimana awalnya pendidikan dilakukan secara tatap muka, karena p...
2 Sistem Cerdas	jurnal_pred2	C:/Users/Asus V...	Hutan gambut di Kalimantan adalah salah satu ekosistem hidrologi penting di dunia yang memainkan peran vital dalam menjaga keseimbangan lingkungan. Hutan ini memfasilitasi siklus air dan membantu mengun...
3 Sistem Cerdas	jurnal_pred3	C:/Users/Asus V...	Sampah merupakan masalah di seluruh dunia, terutama di negara berkembang. Masyarakat seringkali tidak memilah sampah dengan benar, sehingga sampah yang seharusnya dapat didaur ulang justru terkontaminas...
4 Socio Informatics	jurnal_pred4	C:/Users/Asus V...	Piramida Kecelakaan Kerja atau dikenal juga dengan Heinrich's triangle atau Bird's triangle menyatakan bahwa kecelakaan kerja dapat ditimbulkan karena beberapa kejadian-kejadian kecelakaan lain yang lebih ringan ...
5 Pemodelan	jurnal_pred5	C:/Users/Asus V...	Prediksi keberlangsungan hidup pasien gagal jantung telah dilakukan pada penelitian untuk mencari tahu tentang kinerja, akurasi, presisi dan performa dari model prediksi ataupun metode yang digunakan dalam pen...
6 Non IT	jurnal_pred6	C:/Users/Asus V...	Bantuan Langsung Tunai (BLT) adalah bantuan yang dibuat oleh pemerintah untuk orang atau keluarga yang kurang mampu dan rentan akan kemiskinan dengan mereka yang memenuhi persyaratan yang ada pada p...
7 Sistem Cerdas	jurnal_pred7	C:/Users/Asus V...	Kucing merupakan salah satu hewan yang banyak dipelihara oleh masyarakat. Penyakit kulit pada kucing seringkali membuat pemiliknya merasa bingung dan terkendala dalam diagnosa penyakit kulit pada kucing yan...

model telah memprediksi beragam kelas yang mencerminkan variasi konten dalam data. Misalnya, model mengidentifikasi "Sistem Cerdas" untuk sampel yang membahas tentang teknologi atau sistem pintar, dan "Non IT" untuk topik yang mungkin lebih umum atau tidak terkait langsung dengan teknologi.

KESIMPULAN

Berdasarkan analisis yang telah dilakukan, model Random Forest menunjukkan performa yang lebih baik dibandingkan dengan Neural Network dalam klasifikasi teks pada data uji yang diberikan. Hal ini terlihat dari konsistensi dan relevansi prediksi kelas yang lebih tinggi dalam Random Forest, meskipun tanpa menyertakan probabilitas prediksi. Random Forest cenderung lebih efektif dalam mengatasi variasi topik dan memberikan klasifikasi yang akurat, sedangkan Neural Network menunjukkan beberapa kelemahan dalam kesalahan klasifikasi, khususnya pada kelas Sistem Cerdas dan Socio Informatics. Oleh karena itu, untuk aplikasi ini, Random Forest disarankan sebagai metode yang lebih andal karena kemampuannya dalam menghasilkan prediksi yang konsisten dan akurat.