



EKSPLORASI INDEKS GOOGLE TRENDS PADA PEMODELAN DATA TIME SERIES

ABSTRAK

Pemanfaatan big data seperti Indeks Google Trends telah banyak dikembangkan di berbagai bidang. Salah satunya sebagai pelengkap data resmi BPS pada analisis data time series. Tujuan penelitian ini adalah mengkolaborasikan data Indeks Google Trends dan data BPS pada analisis Tingkat Pengangguran Terbuka (TPT) di Indonesia. Pada penelitian ini juga akan dibandingkan metode statistik klasik ARIMAX dan Jaringan Syaraf Tiruan untuk menentukan metode terbaik dalam memodelkan dataset TPT di Indonesia. Adapun periode penelitian adalah Februari 2005 – Februari 2022. Hasil penelitian menunjukkan bahwa pemodelan data tingkat pengangguran di Indonesia dengan menambahkan variabel prediktor berupa Indeks Google Trends memberikan hasil yang lebih baik dibandingkan dengan pemodelan tanpa Indeks Google Trends. Selain itu, analisis dengan menggunakan metode statistik klasik ARIMAX memberikan hasil yang lebih baik dibandingkan metode Jaringan Syaraf Tiruan.

Kata Kunci: Google Trends, ARIMAX, Jaringan Syaraf Tiruan.

ABSTRACT

The use of big data such as the Google Trends Index has been widely developed in various fields. One of them is as a complement to BPS official data in time series data analysis. The purpose of this research is to collaborate with Google Trends Index data and BPS data on the analysis of the Open Unemployment Rate in Indonesia. This study will also compare the best analytical methods between classical statistical methods ARIMAX and Artificial Neural Networks in modeling Unemployment datasets in Indonesia. The research period is February 2005 – February 2022. The results show that modeling unemployment rate data in Indonesia by adding a predictor variable in the form of the Google Trends Index gives better results than modeling without the Google Trends Index. In addition, the analysis using the classical statistical method ARIMAX gives better results than the Artificial Neural Network method.

Kata Kunci: Google Trends, ARIMAX, Artificial Neural Network.

1 Pendahuluan

Perkembangan teknologi informasi di era digital membawa perubahan yang sangat pesat di berbagai aspek kehidupan. Salah satunya kemudahan dalam mencari informasi di mesin pencarian seperti *Google Search*. Volume intensitas pencarian topik informasi dengan menggunakan kata kunci (*keyword*) di *Google Search* pada kurun waktu tertentu terangkum dalam bentuk angka indeks yang sering disebut dengan Google Trends Index (GT). Volume intensitas popularitas topik yang disajikan oleh GT ini bersifat realtime sehingga mampu menggambarkan kondisi masyarakat saat ini. Hal ini membuat GT indeks memiliki potensi yang sangat besar sebagai pelengkap data official resmi.

Data official resmi seperti data BPS memiliki kekurangan diantaranya yaitu pelaporan data seringkali memiliki lag time atau waktu tunda, yaitu jeda antara waktu pengumpulan data lapangan (survei) sampai dengan data tersebut dipublikasikan. Hal ini menyebabkan data

official resmi kurang mampu menjawab kebutuhan data secara realtime sesuai kondisi lapangan. Salah satu data official yang dikeluarkan oleh BPS adalah data Tingkat Pengangguran Terbuka (TPT) yang merupakan salah satu indikator ekonomi makro yang menggambarkan kinerja pasar kerja. Nilai TPT yang tinggi menunjukkan bahwa semakin banyak tenaga kerja yang tidak termanfaatkan oleh dunia kerja. Data TPT diperoleh dari Survei Angkatan Kerja Nasional (SAKERNAS) yang dilaksanakan oleh BPS setiap setahun dua kali yaitu pada bulan Februari dan Agustus. Data semesteran hasil SAKERNAS membutuhkan waktu yang cukup panjang dari proses survei sampai dengan publikasi. Data hasil SAKERNAS pada bulan Februari baru dipublikasikan pada bulan Mei [1]. Dengan kata lain, data TPT memiliki waktu tunda selama tiga bulan.

Dengan adanya keterbatasan ketersediaan data resmi secara realtime ini maka perlu adanya analisis time series dengan mengkolaborasikan data official statistik BPS dengan data GT sebagai jawaban dari kebutuhan data untuk menggambarkan kondisi tingkat pengangguran saat ini. Analisis data time series TPT dengan menggunakan kolaborasi data resmi BPS dan GT telah dikaji oleh [2] dan [3]. [2] melakukan pemodelan *nowcasting* TPT dengan memanfaatkan data Google Trends menggunakan metode Machine Learning. Hasil penelitian menunjukkan bahwa data indeks GT dapat digunakan sebagai sumber data alternatif untuk *nowcasting* TPT Jawa Barat. Selanjutnya, [3] memanfaatkan data GT untuk meramalkan TPT Provinsi Jawa Barat dengan menggunakan metode statistika klasik ARIMAX. Hasil penelitian menunjukkan bahwa pemanfaatan data GT pada pemodelan TPT memberikan akurasi peramalan yang lebih baik dibandingkan dengan pemodelan hanya menggunakan data resmi.

Metode Machine Learning merupakan teknik baru dalam peramalan. Salah satu kelebihan dari metode ini dibandingkan dengan metode ekonometrika klasik adalah tidak diperlukannya uji asumsi linieritas seperti pada metode ARIMA dan ARIMAX [4]. Namun di sisi lain, hasil komparasi dari kedua model seperti yang telah diteliti oleh [4] dan [5] menunjukkan bahwa akurasi metode ekonometrika klasik lebih baik dibandingkan dengan metode Machine Learning. Berdasarkan latar belakang di atas, maka pada penelitian ini akan dilakukan komparasi akurasi metode statistika klasik ARIMAX dan Machine Learning Jaringan Syaraf Tiruan Backpropagation untuk mendapatkan peramalan terbaik dataset TPT hasil eksplorasi Google Trends.

2 Tinjauan Pustaka

2.1 Google Trend

Google Trend adalah data dan grafik statistik pencarian web yang disediakan oleh Google yang menunjukkan trend suatu topik pencarian menurut wilayah geografis maupun bahasa. Layanan yang diluncurkan pertama kali oleh Google pada tahun 2006 ini disajikan dalam bentuk indeks dengan skala 0 sampai 100. Nilai ini diperoleh dari hasil pembagian setiap istilah atau *search query* dengan volume total kueri penelusuran dari wilayah geografis dan rentang waktu tertentu. Google Support menerangkan bahwa data GT yang dapat diakses terdiri dari dua jenis data yaitu :

1. Data real-time

Sampel acak dari data yang mencakup penelusuran selama tujuh hari terakhir sebelum data trend terbentuk

2. Data non real-time

Sampel dari data yang mencakup penelusuran dari tahun 2004 sampai 36 jam sebelum data trend dibentuk.

2.2 Metode Statistika ARIMAX

Model ARIMAX merupakan pengembangan dari model ARIMA dengan penambahan variabel eksogen. Persamaan dari model ARIMAX adalah sebagai berikut [6]:

$$w_t = \beta x_t + \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

dengan β merupakan parameter eksogenus dan x_t adalah data eksogenus pada saat ke t .

2.3 Neural Network

Metode Neural Network atau yang lebih sering dikenal dengan Jaringan Syarat Tiruan (JST) merupakan salah satu bagian dari Machine Learning yang dalam proses pengerjaannya menyerupai cara kerja syaraf otak manusia. Algoritma dari metode JST terdiri dari sejumlah jaringan syaraf tiruan yang saling terhubung satu dengan yang lainnya [7]. Penyusunan jaringan syaraf tiruan pada umumnya terdiri dari tiga lapisan [8] yaitu :

1. Lapisan input yaitu lapisan yang terdiri atas unit-unit input yang menerima pola data dari luar
2. Lapisan tersembunyi (*hidden layer*) yaitu lapisan yang berisi unit-unit yang outputnya tidak langsung dapat diamati.
3. Lapisan output yaitu lapisan yang terdiri atas unit-unit output yang merupakan solusi yang dihasilkan dari jaringan syaraf tiruan.

Arsitektur JST yang paling sering digunakan adalah jaringan banyak lapisan (*multilayer network*) dengan kombinasi pembelajaran Backpropagation. Kelebihan metode ini adalah kemampuan memformulasikan pengalaman dan pengetahuan yang tepat dan fleksibel dalam perubahan peraturan perkiraan [9]. Di sisi lain, metode JST Backpropagation ini memiliki kekurangan antara lain prediksi yang diperoleh dari metode ini bisa memberikan hasil yang invalid jika input yang diterima di luar range yang diberikan pada saat pelatihan atau data training yang dibutuhkan tidak cukup [10].

2.4 Pemilihan Model Terbaik

Pemilihan model terbaik bertujuan untuk mendapatkan model yang tepat dalam menggambarkan data TPT. Pemilihan ini didasarkan pada nilai akurasi menggunakan Mean Squared Error (MSE), yaitu [11] :

$$MSE = \frac{\sum |Y_t - \hat{Y}_t|^2}{n}$$

dengan Y_t : data aktual, \hat{Y}_t : data prediksi, serta n : jumlah data

3 Metode Penelitian

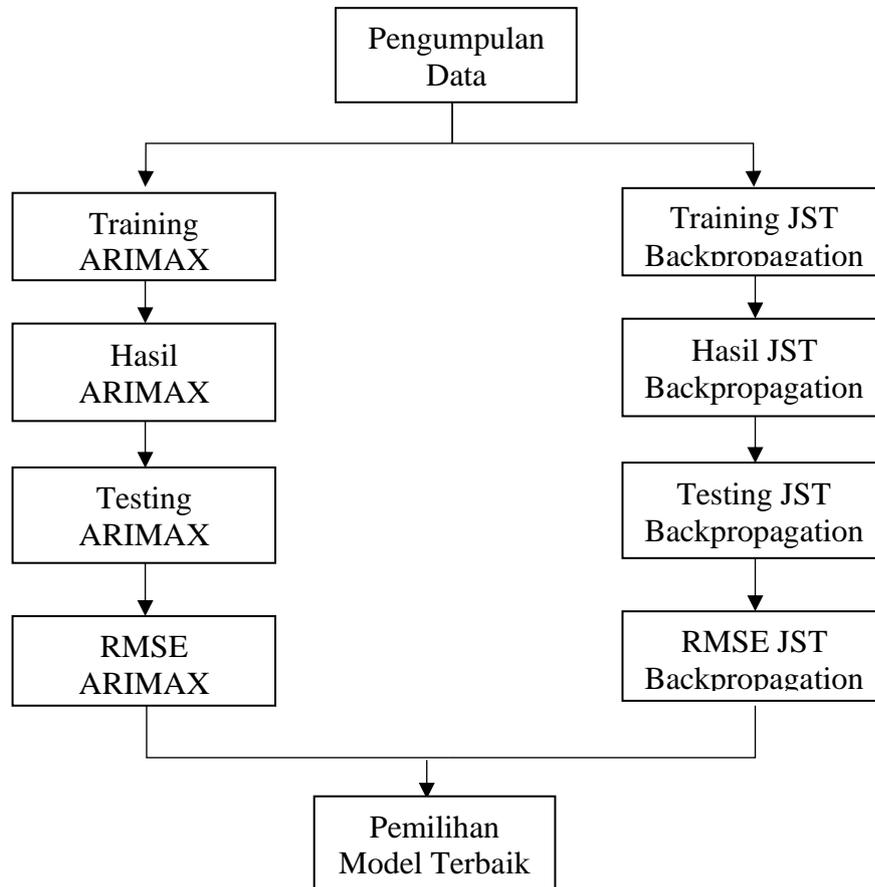
3.1 Data

Penelitian ini menggunakan data sekunder yang bersumber dari data Google Trends dengan *search query* “Lowongan Kerja” dengan ruang lingkup wilayah Indonesia sebagai variabel input [3]. Selanjutnya, untuk variabel output berupa data resmi Tingkat Pengangguran Terbuka Indonesia yang berasal dari data BPS [12]. Adapun periode penelitian ini adalah Februari 2005 sampai Februari 2022. Dari data tersebut dibagi menjadi data latih dan data uji. Data latih menggunakan 30 data dari periode Februari 2005 sampai dengan

Agustus 2019 sedangkan untuk data uji menggunakan 5 data dari periode Februari 2020 sampai dengan Februari 2022.

3.2 Metode Penelitian

Pada penelitian ini akan dibandingkan dua metode analisis yaitu metode statistika ARIMAX dan metode JST Backpropagation.



Gambar 1: Tahapan Penelitian

Berdasarkan Gambar 1, maka tahapan analisis pada penelitian ini adalah sebagai berikut :

1. Melakukan pengumpulan data serta identifikasi pola data dengan menggunakan plot data.
2. Melakukan pemodelan dengan menggunakan metode ARIMA.
 - a. Uji kestasioneran data dalam ragam dan rataaan. Untuk uji kestasioneran ragam dapat dilihat dari plot data sedangkan untuk uji kestasioneran rataaan digunakan uji *Augmented Dickey Fuller* (ADF), dengan hipotesis uji yang digunakan sebagai berikut [13] :

$H_0: \gamma = 0$ (data tidak stasioner)

$H_0: \gamma < 0$ (data stasioner)

Statistik uji yang digunakan :

$$\tau = \frac{\hat{\gamma} - \gamma}{SE(\hat{\gamma})}$$

dimana γ : Nilai amatan

$\hat{\gamma}$: Nilai dugaan

$SE(\hat{\gamma})$: Kesalahan baku dari $\hat{\gamma}$.

Selanjutnya, untuk kriteria daerah penolakan adalah tolak H_0 apabila nilai probabilitas dari statistik uji ADF $< \alpha$. Apabila hasil uji ADF menunjukkan data belum stasioner maka perlu dilakukan proses transformasi ataupun *differencing* (pembedaan data) sampai diperoleh data yang stasioner baik dalam ragam maupun rataan.

- b. Melakukan identifikasi model ARIMA melalui plot ACF dan PACF.
 - c. Melakukan pendugaan dan uji signifikansi parameter model
 - d. Melakukan uji diagnostik model untuk memeriksa kebebasan dan kenormalan sisaan.
 - e. Melakukan *overfitting* data untuk memeriksa adanya kemungkinan model lain yang lebih dari model sebelumnya.
 - f. Melakukan peramalan sekaligus menghitung nilai MSE ARIMA
3. Melakukan pemodelan dengan menggunakan model ARIMAX
 - a. Melakukan estimasi parameter dengan cara menambahkan variabel berpengaruh ke dalam model ARIMA terpilih pada langkah sebelumnya
 - b. Melakukan pendugaan dan uji signifikansi parameter
 - c. Melakukan uji diagnostik
 - d. Melakukan peramalan
 4. Melakukan pemodelan dengan menggunakan model JST Backpropagation
 - a. Menentukan input dan output layer.
 - b. Melakukan tahap preprocessing.

Pada tahap ini data awal dinormalisasikan dengan tujuan untuk mempermudah proses pengolahan data pada software Matlab menggunakan persamaan [14] :

$$X' = \frac{0,8(X-b)}{(a-b)} + 0,1 \quad (1)$$

dimana :

- X' : Data hasil normalisasi
- X : Data asli
- a : Nilai maksimum data asli
- b : Nilai minimum data asli.

Dari hasil transformasi menggunakan persamaan (1) diperoleh data normalisasi berada pada range nilai [0,1]. Dengan nilai range [0,1] maka fungsi aktivasi JST yang digunakan adalah fungsi sigmoid biner [15]. Fungsi aktivasi sigmoid biner dipilih karena fungsi aktivasi ini tergolong paling mudah diferensiasinya [16].

- c. Membangun arsitektur JST Backpropagation.

Pada metode JST tidak ada aturan yang baku untuk menentukan JST yang optimal untuk diterapkan ke dalam sistem [17]. Untuk selanjutnya, penentuan arsitektur terutama penentuan jumlah simpul pada lapisan tersembunyi dilakukan secara *trial* dan *error*. Faktor lainnya yang digunakan untuk pemilihan JST optimal pada penelitian ini adalah parameter *learning rate* (kecepatan pembelajaran sistem). Penggunaan nilai *learning rate* yang terlalu kecil membuat jaringan mencapai global minimum dalam waktu yang lama. Sebaliknya penggunaan nilai *learning rate* yang terlalu besar dapat menyebabkan jaringan melewati global minimum sehingga nilai kesalahan meningkat [16].
- d. Melakukan peramalan dan menghitung nilai MSE.
5. Membandingkan nilai MSE dari model ARIMA, ARIMAX, dan JST Backpropagation.

4 Hasil dan Pembahasan

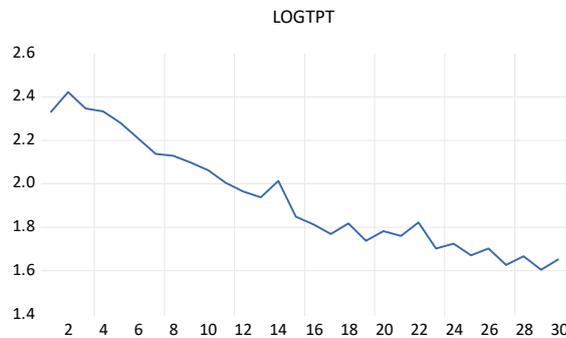
4.1 Metode ARIMA

Sebelum melakukan analisis data runtun waktu, data perlu dipastikan sudah memenuhi syarat uji stasioneritas baik dalam ragam maupun rataaan. Data yang stasioner dapat dilihat dari grafik yang memiliki nilai rata-rata dan varians yang konstan.



Gambar 2 : Plot Data Tingkat Pengangguran Terbuka

Terlihat pada Gambar 2, plot data TPT di Indonesia masih belum stasioner. Untuk mengatasi hal tersebut maka dilakukan proses stasioneritas dengan menggunakan fungsi log data asli.



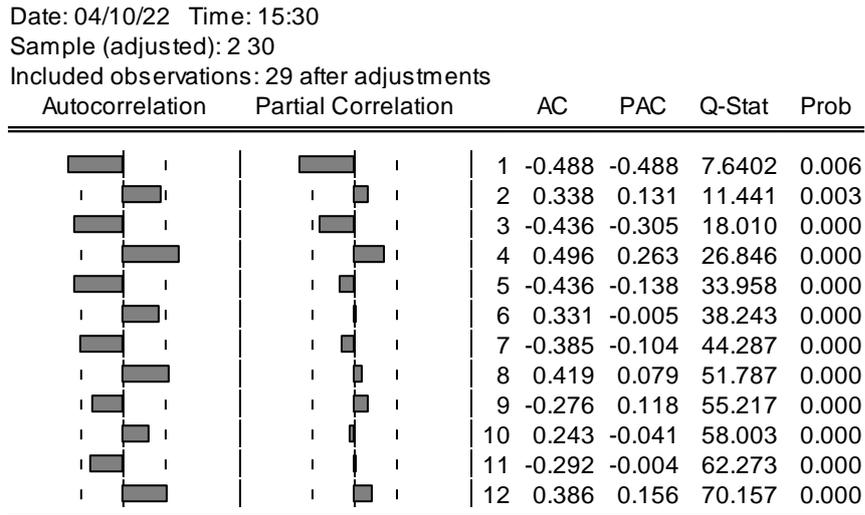
Gambar 3 : Plot Log Data TPT

Dari Gambar 3 terlihat bahwa data log tingkat pengangguran di Indonesia masih mengandung trend yang menandakan data belum stasioner pada rata-rata. Oleh karena itu, langkah selanjutnya melakukan pembedaan data. Berikut diberikan tabel uji stasioneritas hasil proses pembedaan data pada log data TPT di Indonesia.

Tabel 1: Hasil Uji ADF Data Log TPT

Order	t-statistik	Probabilitas	Keputusan
Level atau I(0)	-0.931491	0.9367	Tidak stasioner pada level
First Difference atau I(1)	-5.353601	0.0010	Stasioner pada first difference

Pada Tabel 1 diperoleh bahwa data Log(TPT) belum stasioner pada tingkat level. Hal ini terlihat dari nilai Prob. = 0.9367 > $\alpha=5\%$. Selanjutnya setelah dilakukan proses pembedaan data dan diperoleh nilai Prob. = 0.0010 < $\alpha=5\%$. Hal ini menunjukkan data sudah stasioner. Setelah diperoleh data yang sudah stasioner langkah selanjutnya adalah melakukan identifikasi model ARIMA dengan melihat pola ACF dan PACF pada correlogram data hasil pembeda dari Log TPT atau dLog (TPT) sebagai berikut :



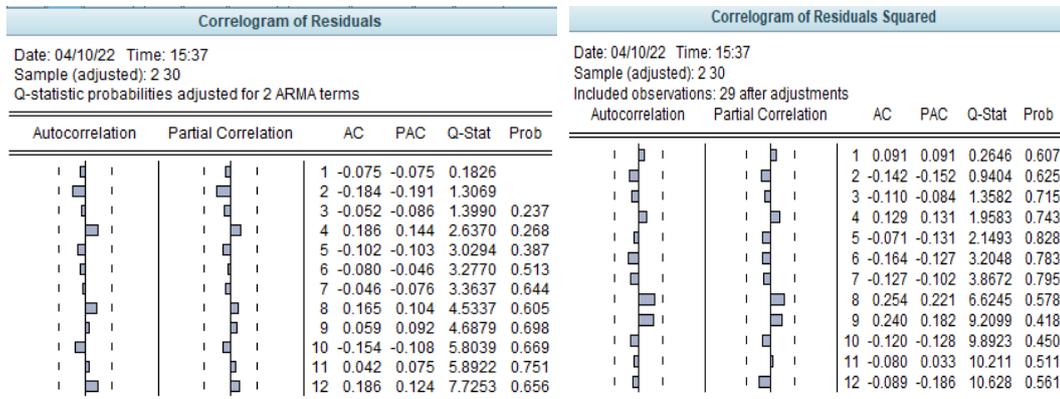
Gambar 4: Correlogram d(Log(TPT))

Berdasarkan Gambar 4 terlihat bahwa plot ACF dan PACF melebihi garis pada lag pertama. Hal ini mengindikasikan bahwa ordo yang memungkinkan untuk data d(Log(TPT)) adalah AR(1), MA(1), atau ARIMA (1,1,1). Langkah selanjutnya dilakukan uji signifikansi parameter dimana parameter dikatakan signifikan apabila nilai probabilitasnya kurang dari $\alpha = 5\%$.

Tabel 2 : Uji ADF Data Log TPT

Model ARIMA	AR (1)	MA (1)	Keputusan
ARIMA (1,1,1)	-0.998 (0.0000)	0.951 (0.0198)	Lolos uji signifikansi
ARIMA (1,1,0)	-0.267 (0.1432)		Tidak lolos uji signifikansi
ARIMA (0,1,1)		-0.142 (0.5136)	Tidak lolos uji signifikansi

Pada Tabel 2 diperoleh model ARIMA yang lolos uji signifikansi adalah ARIMA (1,1,1) dengan nilai prob. AR(1)=0.0000 dan prob. MA(1)=0.0198 secara keseluruhan kurang dari $\alpha=0.05$. Setelah diperoleh model terpilih langkah selanjutnya adalah melakukan uji diagnostik meliputi uji keacakan dan homogenitas error.



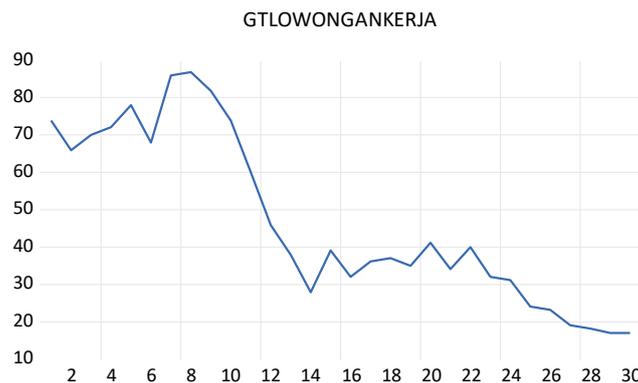
Gambar 5: Uji Diagnostik ARIMA (1,1,1)

Berdasarkan Gambar 5, nilai rata-rata probabilitas pada plot *correlogram of residuals* yang bernilai lebih dari 0.05 maka dapat disimpulkan bahwa error bersifat acak. Selanjutnya,

dari *plot correlogram of residuals squared* juga diperoleh nilai rata-rata probabilitas di atas 0.05. Hal ini menunjukkan bahwa model bersifat homogen.

4.2 Model ARIMAX

Model ARIMAX dibentuk dari model ARIMA yang telah lolos uji signifikansi parameter maupun uji diagnostik. Dari hasil tahap sebelumnya terpilih model ARIMA (1,1,1) sehingga untuk model ARIMAX dibangun model ARIMAX (1,1,1). Model ARIMAX ini merupakan penggabungan model ARIMA (1,1,1) ditambahkan dengan variabel bebas lainnya. Pada penelitian ini variabel bebas yang digunakan adalah indeks Google Trends (GT) dengan *search query* “Lowongan Kerja”. Berikut diberikan plot data GT Lowongan Kerja.



Gambar 6: Plot Indeks Google Trends (GT) Lowongan Kerja

Berdasarkan Gambar 6 terlihat bahwa data GT Lowongan Kerja belum stasioner pada ragam sehingga perlu ditransformasi terlebih dahulu dengan menggunakan fungsi Log. Adapun hasil uji ADF untuk data Log (GTLowonganKerja) adalah sebagai berikut :

Augmented Dickey-Fuller Unit Root Test on LOGGTLOWKER		
Null Hypothesis: LOGGTLOWKER has a unit root		
Exogenous: Constant, Linear Trend		
Lag Length: 3 (Automatic - based on SIC, maxlag=7)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-4.179728	0.0147
Test critical values:		
1% level	-4.356068	
5% level	-3.595026	
10% level	-3.233456	
*MacKinnon (1996) one-sided p-values.		
Augmented Dickey-Fuller Test Equation		
Dependent Variable: D(LOGGTLOWKER)		
Method: Least Squares		
Date: 04/10/22 Time: 15:57		
Sample (adjusted): 5 30		
Included observations: 26 after adjustments		

Gambar 7: Uji Stasioneritas Log(GT Lowongan Kerja)

Dari Gambar 7, terlihat bahwa nilai probabilitas diperoleh $\text{prob} = 0.0147 \leq 0.05$ sehingga dapat disimpulkan data hasil transformasi log sudah stasioner. Selanjutnya dilakukan uji signifikansi untuk melihat model ARIMAX (1,1,1) yang sudah ditambahkan dengan data log(GT Lowongan Kerja).

Dependent Variable: D(LOGTPT)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 04/10/22 Time: 16:00
 Sample: 2 30
 Included observations: 29
 Convergence achieved after 21 iterations
 Coefficient covariance computed using outer product of gradients

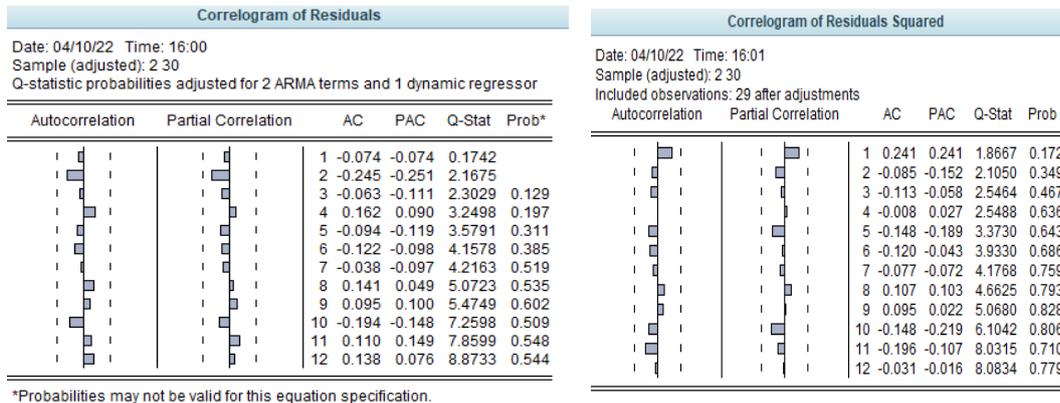
Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOGGTLOWKER	-0.006839	0.002726	-2.508803	0.0190
AR(1)	-0.995013	0.026296	-37.83897	0.0000
MA(1)	0.898012	0.244155	3.678046	0.0011
SIGMASQ	0.001783	0.000633	2.814383	0.0094

R-squared	0.476521	Mean dependent var	-0.023236
Adjusted R-squared	0.413704	S.D. dependent var	0.059391
S.E. of regression	0.045475	Akaike info criterion	-3.154787
Sum squared resid	0.051700	Schwarz criterion	-2.966195
Log likelihood	49.74442	Hannan-Quinn criter.	-3.095723
Durbin-Watson stat	1.996864		

Inverted AR Roots	-1.00
Inverted MA Roots	-.90

Gambar 8: Uji Signifikansi Model ARIMAX (1,1,1)

Berdasarkan output pada Gambar 8, dengan mengambil nilai $\alpha = 5\%$ dapat disimpulkan bahwa seluruh parameter pada model ARIMAX (1,1,1) lolos uji signifikansi. Selanjutnya, dilakukan uji diagnostik dari model ARIMAX (1,1,1).



Gambar 9: Uji Diagnostik Model ARIMAX (1,1,1)

Berdasarkan pada output uji diagnostik pada Gambar 9, diperoleh seluruh nilai rata-rata probabilitas ≥ 0.05 sehingga dapat disimpulkan model ARIMAX (1,1,1) cocok untuk mendeskripsikan data tingkat pengangguran di Indonesia.

4.3 Model JST Backpropagation

Pada penelitian ini pemodelan data TPT menggunakan tiga lapisan, yaitu lapisan input dengan satu simpul berupa data indeks Google Trend dengan *search query* lowongan kerja (X), lapisan output dengan satu simpul berupa data tingkat pengangguran di Indonesia (y), dan satu lapisan tersembunyi. Adapun pemilihan jumlah simpul pada lapisan tersembunyi menggunakan proses *trial and error* dengan jumlah simpul untuk uji coba sebanyak 10, 15, dan 20. Selanjutnya, dari setiap arsitektur jaringan yang terbentuk dilakukan uji coba dengan variasi nilai parameter *learning rate* yang berbeda-beda untuk mendapatkan arsitektur terbaik. Adapun nilai kesalahan arsitektur model dan regresi dari beberapa percobaan arsitektur pada data latih diberikan pada tabel berikut :

Tabel 3 : Nilai Regresi dan Error Hasil Pelatihan

Learning rate		0.001	0.01	0.1
Hidden Layer 10	Error	0.0039203	0.0039204	0.0039202
	Regression	0.96296	0.96296	0.96296
Hidden Layer 15	Error	0.0025581	0.0026104	0.00265
	Regression	0.97599	0.97549	0.9751
Hidden Layer 20	Error	0.0017475	0.00176	0.0017523
	Regression	0.98366	0.98354	0.98361

Tabel 3 menunjukkan jaringan terbaik untuk proses pelatihan data TPT adalah JST dengan menggunakan *hidden layer* sebanyak 20, *epoch* 100.000, dan *learning rate* sebesar 0.001.

4.4 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan membandingkan nilai error yang diperoleh dari masing-masing model baik dari analisis statistik klasik yaitu ARIMA dan ARIMAX maupun analisis machine learning menggunakan JST Backpropagation.

Tabel 4 : Perbandingan MSE Data Uji Model Statistik ARIMA dan JST

Periode	Data Aktual	ARIMA (1,1,1) [Tanpa Indeks GT]		ARIMAX (1,1,1)		JST Backpropagation	
		\hat{y}	(Error ²)	\hat{y}	(Error ²)	\hat{y}	(Error ²)
Februari 2020	4,94	5,03	0,008	4,79	0,023	5,62	0,462
Agustus 2020	7,07	7,43	0,130	7,11	0,002	7,17	0,010
Februari 2021	6,26	6,29	0,001	6,01	0,063	6,19	0,005
Agustus 2021	6,49	7,24	0,563	7,08	0,348	6,67	0,032
Februari 2022	5,83	5,62	0,044	5,89	0,004	5,65	0,032
MSE			0,149		0,088		0,108

Dari Tabel 4 diperoleh bahwa nilai MSE terkecil diperoleh dari hasil analisis statistik klasik ARIMAX (1,1,1). Hal ini menunjukkan bahwa model ARIMAX (1,1,1) merupakan model terbaik untuk mendeskripsikan data tingkat pengangguran di Indonesia.

5 Kesimpulan

Dari hasil penelitian dan analisis perbandingan *performance* metode statistik klasik ARIMAX dengan Jaringan Syaraf Tiruan (JST) berbasis Backpropagation diperoleh hasil simpulan sebagai berikut :

1. Pemodelan data tingkat pengangguran di Indonesia dengan menambahkan variabel prediktor berupa indeks Google trends memberikan hasil yang lebih baik dibandingkan dengan pemodelan tanpa Indeks Google Trends.

2. Pemodelan dengan menggunakan analisis statistik klasik ARIMAX memiliki nilai performance akurasi yang lebih baik dibandingkan dengan analisis Machine Learning JST. Hal ini ditunjukkan dari nilai MSE model ARIMAX (1,1,1) yang lebih kecil jika dibandingkan dengan MSE model ARIMA maupun JST Backpropagation.
3. Pemodelan dengan menggunakan analisis JST berbasis Backpropagation membutuhkan waktu yang relatif lebih lama. Hal ini disebabkan karena proses pemilihan parameter pada pembentukan arsitektur JST dilakukan dengan percobaan.

Daftar Pustaka

- [1] M. Q. Virati, "Development of a Big Data Analysis System (Case Study : Unemployment Statistic)," 2019.
- [2] I. A. Akbar and R. Kurniawan, "Pemodelan Nowcasting Tingkat Pengangguran Terbuka Menggunakan Data Google Trends Dengan Metode Antlion Optimization-Support Vector Regression," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, pp. 760–770, 2021, doi: 10.34123/semnasoffstat.v2020i1.504.
- [3] L. Widyarsi and H. Usman, "Penggunaan Data Google Trends untuk Peramalan Tingkat Pengangguran Terbuka di Tingkat Nasional dan Regional di Provinsi Jawa Barat," *Semin. Nas. Off. Stat.*, vol. 2021, no. 1, pp. 980–990, 2021, doi: 10.34123/semnasoffstat.v2021i1.842.
- [4] S. Renny Elfira Wulansari, "Peramalan Netflow Uang Kartal dengan Metode ARIMAX dan Radial Basis Function Network (Studi Kasus Di Bank Indonesia)," *J. Sains dan Seni POMITS*, vol. 3, no. 2, pp. 73–78, 2014.
- [5] A. Gunaryati and A. Suhendra, "Perbandingan Antara Metode Statistika dan Metode Neural Network pada Model Peramalan Indeks Harga Perdagangan Besar," *J. Teknol. dan Rekayasa*, vol. 20, no. 1, pp. 23–35, 2015.
- [6] D. Rosadi, *Analisis Ekonometrika & Runtun Waktu Terapan dengan R*. Yogyakarta: Andi Offset, 2011.
- [7] S. K. Mishra and N. Sharma, "Rainfall Forecasting Using Backpropagation Neural Network," *Stud. Comput. Intell.*, vol. 713, no. 2, pp. 277–288, 2018, doi: 10.1007/978-981-10-4555-4_19.
- [8] Y. A. Lesnussa, L. J. Sinay, and M. R. Idah, "Aplikasi Jaringan Saraf Tiruan Backpropagation untuk Penyebaran Penyakit Demam Berdarah Dengue (DBD) di Kota Ambon," *J. Mat. Integr.*, vol. 13, no. 2, p. 63, 2017, doi: 10.24198/jmi.v13.n2.11811.63-72.
- [9] I. M. Mataram, "Peramalan Beban Hari Libur Menggunakan Artificial Neural Network," *Peramalan Beban ... I Made Mataram Teknol. Elektro*, vol. 53, no. 2, pp. 2006–2009, 2008.
- [10] Y. Apriyani, "Penerapan Jaringan Syaraf Tiruan Backpropagation Untuk Prediksi Nilai UN Siswa SMPN 2 Cihaurbeuti," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 3, no. 1, pp. 63–70, 2018.
- [11] D. Montgomery, C. Jennings, and M. Kulahci, *ntroduction to Time Series Analysis and Forecasting*. New Jersey(US): John Wiley & Sons, 2015.
- [12] Badan Pusat Statistik, "Tingkat Pengangguran Terbuka Menurut Provinsi (Persen)," *BPS*, 2022. <https://www.bps.go.id/indicator/6/543/1/tingkat-pengangguran-terbukamenurut-provinsi.html>
- [13] Rusdi, "Uji Akar-Akar Unit dalam Model Runtun Waktu Autoregresif," *Satistika*, vol. 11, no. 2, pp. 67–78, 2011.
- [14] D. N. Agus Perdana Windarto, M. S. H. Anjar Wanto, Frinto Tambunan, M. R. L. Muhammad Noor Hasan Siregar, and D. N. Solikhun, Yusra Fadhillah, *Jaringan Saraf*

- Tiruan: Algoritma Prediksi dan Implementasi*, vol. 53, no. 9. 2019.
- [15] J. Teknovasi, E. B. Nababan, and M. Zarlis, “Analisis Fungsi Aktivasi Sigmoid Biner dan Sigmoid Bipolar dalam Algoritma Backpropagation pada Prediksi Kemampuan Siswa,” vol. 02, pp. 103–116, 2015.
- [16] B. Badieah, R. Gernowo, and B. Surarso, “Metode Jaringan Syaraf Tiruan Untuk Prediksi Performa Mahasiswa Pada Pembelajaran Berbasis Problem Based Learning (PBL),” *J. Sist. Inf. Bisnis*, vol. 6, no. 1, p. 46, 2016, doi: 10.21456/vol6iss1pp46-58.
- [17] A. A. Kardan, H. Sadeghi, S. S. Ghidary, and M. R. F. Sani, “Prediction of student course selection in online higher education institutes using neural network,” *Comput. Educ.*, vol. 65, no. 1, pp. 1–11, 2013.