

**ANALISIS KOMPARATIF PERFORMA MODEL BSTS, SARIMA,
DAN LSTM DALAM PREDIKSI ANGKATAN KERJA USIA
MUDA PADA KONDISI DATA TERBATAS DI INDONESIA**

TESIS

Untuk memenuhi sebagian persyaratan memperoleh gelar Magister Teknologi
Informasi

Disusun Oleh:

Harun Al Jafar

247064518001



**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
JAKARTA
2025**

HALAMAN PENGESAHAN

ANALISIS KOMPARATIF PERFORMA MODEL BSTS, SARIMA, DAN LSTM DALAM PREDIKSI ANGKATAN KERJA USIA MUDA PADA KONDISI DATA TERBATAS DI INDONESIA

Untuk memenuhi sebagai persyaratan memperoleh Gelar Magister Komputer

Disusun Oleh:

Harun Al Jafar

247064518001

Thesis ini telah diuji dan dinyatakan lulus pada

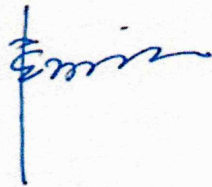
21 Februari 2026

Telah diperiksa dan disetujui oleh:

Mengetahui,

Dosen Pembimbing

Ketua Program Studi Magister Teknologi Informasi



(Dr. Fauziah, S.Kom., M.MSI)

NIDN. 0304107503



(Ir. Asruf Sani, S.T., M.T., M.Kom, Ph.D)

NIDN. 0303067003

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah Tesis ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka. Apabila ternyata didalam naskah Tesis ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia Tesis ini digugurkan dan gelar akademik yang telah saya peroleh (Magister) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Jakarta, 02 Maret 2026



Harun Al Jafar

NPM. 247064518001

KATA PENGANTAR

Penulis memanjatkan puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya kepada penulis, sehingga akhirnya penulis dapat menyelesaikan tesis ini tepat pada waktunya. Tesis yang berjudul “ Analisis Komparatif Performa Model BSTS, SARIMA, dan LSTM dalam Prediksi Angkatan Kerja Usia Muda pada Kondisi Data Terbatas di Indonesia” ini ditulis untuk memenuhi salah satu syarat guna memperoleh gelar magister pada Universitas Nasional. Pada kesempatan yang baik ini, izinkanlah penulis menyampaikan rasa hormat dan ucapan terima kasih kepada semua pihak yang dengan tulus ikhlas telah memberikan bantuan dan dorongan kepada penulis dalam menyelesaikan tesis ini, terutama kepada :

1. Dr. Fauziah, S.Kom., MMSI. selaku Dosen Pembimbing, Progam Studi Magister Teknologi Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.
2. Dr. Agung Triayudi, S.Kom., M.Kom. selaku Dekan Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional.
3. Bapak dan Ibu Dosen Fakultas Teknologi Komunikasi dan Informatika beserta staf Tata Usaha FTKI yang telah membantu perkuliahan saya.
4. Terkhusus untuk kedua orang tua yang sangat disayangi, Bapak Asep Japar dan Ibu Maryamah. Terima kasih atas kasih sayang, didikan, nasihat dan semangat untuk penulis menuntut ilmu serta semua hal yang kalian tanamkan kepada penulis.

Jakarta, 02 Maret 2026



Harun Al Jafar

ABSTRAK

Kelompok usia 15–24 tahun memiliki peranan penting dalam mendukung bonus demografi Indonesia, namun kondisi pasar kerja muda masih menunjukkan fluktuasi yang dipengaruhi keterbatasan keterampilan, dinamika ekonomi, serta belum tersedianya sistem prediksi tenaga kerja yang kuat. Tantangan lain muncul dari keterbatasan data tahunan 2015–2024 yang jumlahnya sedikit dan mengandung missing value, sehingga memerlukan pendekatan pemodelan yang tepat untuk menghasilkan prediksi yang akurat. Penelitian ini bertujuan mengembangkan model prediksi jumlah angkatan kerja muda menggunakan pendekatan BSTS, LSTM, dan SARIMA, sekaligus membandingkan tingkat akurasi ketiganya berdasarkan MAPE dan RMSE. Penelitian juga bertujuan menguji signifikansi perbedaan performa model melalui uji statistik serta mengidentifikasi model paling optimal yang dapat digunakan sebagai dasar perencanaan kebijakan ketenagakerjaan muda di Indonesia. Penelitian menggunakan analisis deret waktu dengan data BPS tahun 2015–2024. Tahapan pra-pemrosesan meliputi imputasi missing value menggunakan interpolasi linier, normalisasi data, dan pembagian data menjadi bagian latih dan uji. Pemodelan dilakukan menggunakan BSTS, LSTM, dan SARIMA, kemudian dievaluasi menggunakan MAPE, RMSE, serta uji statistik Wilcoxon atau paired t-test untuk melihat signifikansi perbedaan performa antar model. LSTM dan SARIMA menunjukkan akurasi terbaik dengan nilai MAPE sebesar 3,44%, sedangkan BSTS memiliki performa yang lebih rendah pada data tahunan yang terbatas. Uji statistik menunjukkan bahwa BSTS berbeda signifikan dibandingkan dua model lainnya, sementara LSTM dan SARIMA tidak memiliki perbedaan performa secara signifikan. Penurunan akurasi pada periode 2020–2021 mengindikasikan bahwa pasar kerja muda sangat sensitif terhadap guncangan eksternal seperti pandemi. SARIMA dan LSTM merupakan model yang paling sesuai untuk memprediksi jumlah angkatan kerja muda berbasis data tahunan. Temuan ini dapat digunakan untuk mendukung penyusunan kebijakan ketenagakerjaan yang lebih adaptif dan berbasis data, sehingga potensi bonus demografi dapat dimanfaatkan secara optimal.

Kata Kunci: Angkatan kerja muda, BSTS, LSTM, SARIMA, Prediksi tenaga kerja, Bonus demografi

Kepustakaan: 59 (2021–2025)

ABSTRACT

The age group of 15-24 years plays a crucial role in supporting Indonesia's demographic bonus, yet the youth labor market continues to show fluctuations influenced by limited skills, economic dynamics, and the lack of a robust labor force prediction system. Another challenge arises from the limited annual data for 2015-2024, which is sparse and contains missing values, thus requiring an appropriate modeling approach to produce accurate predictions. This study aims to develop a prediction model for the number of young workers using BSTS, LSTM, and SARIMA approaches, while also comparing their accuracy levels based on MAPE and RMSE. The research also aims to test the significance of model performance differences through statistical tests and identify the most optimal model that can be used as a basis for planning youth employment policies in Indonesia. The research employs time series analysis with BPS data from 2015-2024. The preprocessing stages include missing value imputation using linear interpolation, data normalization, and splitting the data into training and testing sets. Modeling was conducted using BSTS, LSTM, and SARIMA, then evaluated using MAPE, RMSE, and Wilcoxon or paired t-tests to examine the significance of performance differences between models. LSTM and SARIMA showed the best accuracy with a MAPE value of 3.44%, while BSTS performed lower on the limited annual data. Statistical tests indicated that BSTS differed significantly from the other two models, while LSTM and SARIMA did not show a significant performance difference. The decrease in accuracy during the 2020-2021 period indicates that the youth labor market is highly sensitive to external shocks such as the pandemic. SARIMA and LSTM are the most suitable models for predicting the number of young workers based on annual data. These findings can be used to support the formulation of more adaptive and data-driven employment policies, allowing the potential of the demographic bonus to be optimized.

Keywords: Young workforce, BSTS, LSTM, SARIMA, Labor force prediction, Demographic bonus

Bibliography: 59 (2021–2025)

DAFTAR ISI

HALAMAN PENGESAHAN	i
PERNYATAAN ORISINALITAS	ii
KATA PENGANTAR	iii
ABSTRAK	iv
ABSTRACT	v
DAFTAR ISI	vi
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	12
1.3 Tujuan Penelitian.....	13
1.4 Manfaat Penelitian.....	14
1.5 Batasan Penelitian	15
1.6 Sistematika Penulisan.....	17
BAB 2 LANDASAN PUSTAKA	18
2.1 Tinjauan Penelitian Terdahulu (<i>State of the Art</i>)	18
2.1.1 Studi Empiris Penggunaan LSTM.....	18
2.1.2 Studi Empiris Penggunaan BSTS.....	19
2.1.3 Tabel Ringkasan Penelitian Terdahulu.....	19
2.2 Kajian Teori yang Relevan.....	21
2.2.1 Bonus Demografi dan Relevansinya terhadap Angkatan Kerja Muda.....	21
2.2.2 Tenaga Kerja Usia 15–24 Tahun: Dinamika dan Tantangan Struktural.....	22
2.2.3 Karakteristik Data Ketenagakerjaan Tahunan.....	23
2.2.4 Interpolasi untuk Mengatasi Missing Value pada Data Tahunan.....	23

2.2.5	Hubungan Linier vs Non-Linier antar Variabel Sosial-Ekonomi	24
2.3	Landasan Teori	24
2.3.1	Teori Bayesian Structural Time Series (BSTS)	24
2.3.2	Teori Seasonal ARIMA (SARIMA)	25
2.3.3	Teori Long Short-Term Memory (LSTM)	27
2.4	Kerangka Pemikiran / Kerangka Konseptual	29
2.4.1	Hubungan Antar Variabel	31
2.4.2	Penjelasan Hubungan Variabel	32
2.4.3	Perbandingan Model BSTS, SARIMA, dan LSTM	33
2.4.4	Penanganan Kualitas Data	35
2.4.5	Kesimpulan Kerangka Pemikiran	36
2.4.6	Gambar Kerangka Pemikiran	37
2.5	Hipotesis Penelitian	37
BAB 3 METODOLOGI PENELITIAN		40
3.1	Pendekatan dan Jenis Penelitian	40
3.2	Objek dan Ruang Lingkup Penelitian	40
3.2.1	Objek	40
3.3	Sumber Data dan Teknik Pengumpulan Data	41
3.3.1	Sumber Data	41
3.3.2	Teknik Pengumpulan Data	41
3.4	Variabel Penelitian	42
3.4.1	Variabel Dependen (Target)	42
3.4.2	Variabel Independen (Eksogen)	42
3.5	Teknik Analisis Data	42
3.5.1	Pra-Pemrosesan Data	43
3.5.2	Analisis Deskriptif dan Eksploratori	45
3.5.3	Pemodelan BSTS (Bayesian Structural Time Series)	47
3.5.4	Pemodelan SARIMA (Seasonal ARIMA)	48
3.5.5	Pemodelan LSTM (Long Short-Term Memory)	50

3.5.6	Evaluasi dan Perbandingan Model.....	52
3.5.7	Mitigasi Risiko Overfitting dan Validasi Statistik	53
3.6	Software dan Tools.....	55
3.7	Alur Penelitian.....	55
BAB 4	HASIL	57
4.1	Gambaran Umum Data.....	57
4.1.1	Sumber Data dan Periode Penelitian	57
4.1.2	Statistik Deskriptif.....	58
4.2	Hasil Interpolasi Bayesian (Semester ke Triwulan).....	60
4.2.1	Proses Interpolasi	60
4.2.2	Kualitas Interpolasi.....	60
4.2.3	Analisis Kualitas Interpolasi Berdasarkan Hukum Rojo & Sanz.....	62
4.3	Hasil Pemodelan BSTS (Bayesian Structural Time Series).....	63
4.3.1	Spesifikasi Model BSTS	63
4.3.2	Feature Importance.....	64
4.3.3	Performa Prediksi BSTS	65
4.4	Hasil Pemodelan SARIMA (Baseline Model).....	68
4.4.1	Spesifikasi Model SARIMA	68
4.4.2	Parameter Estimasi.....	69
4.4.3	Performa Prediksi SARIMA	69
4.5	Hasil Pemodelan LSTM (Exploratory Benchmark).....	71
4.5.1	Arsitektur LSTM.....	71
4.5.2	Training History	72
4.5.3	Performa Prediksi LSTM	73
4.6	Perbandingan Performa Ketiga Model.....	74
4.6.1	Ringkasan Metrik Performa	74
4.6.2	Distribusi Error.....	76
4.6.3	Diebold-Mariano Test	77
4.7	Ringkasan Hasil Penelitian.....	78
BAB 5	PEMBAHASAN	80

5.1	Interpretasi Hasil Interpolasi Bayesians	80
5.1.1	Signifikansi Interpolasi untuk Data Ketenagakerjaan	80
5.1.2	Implikasi terhadap Perencanaan Kebijakan	80
5.2	Analisis <i>Feature Importance</i> dan Dinamika Angkatan Kerja Muda	81
5.2.1	Pengaruh Tingkat Pengangguran Terbuka (TPT)	81
5.2.2	Pengaruh Tingkat Partisipasi Angkatan Kerja (TPAK)	81
5.2.3	Pengaruh Rata-rata Lama Sekolah (RLS)	82
5.2.4	Pengaruh PDB per Kapita dan Populasi Muda	82
5.3	Perbandingan Kinerja Model dan Implikasi Metodologis	83
5.3.1	Keunggulan BSTS untuk Data Terbatas	83
5.3.2	Keterbatasan SARIMA sebagai Model Linear	84
5.3.3	Kegagalan LSTM dan Implikasinya untuk Penggunaan Deep Learning.....	84
5.3.4	Uji Diebold-Mariano dan Signifikansi Statistik.....	85
5.4	Implikasi Kebijakan untuk Optimalisasi Bonus Demografi	86
5.4.1	Penggunaan Model BSTS untuk Perencanaan Kebijakan.....	86
5.4.2	Rekomendasi untuk BPS: Peningkatan Frekuensi Data.....	86
5.4.3	Rekomendasi untuk Kementerian Ketenagakerjaan.....	87
5.5	Kontribusi Penelitian dan Pengembangan Ilmu Pengetahuan.....	88
5.5.1	Kontribusi Metodologis.....	88
5.5.2	Kontribusi bagi Literatur Ketenagakerjaan Indonesia.....	89
5.5.3	Keterbatasan Penelitian dan Agenda Riset Mendatang.....	89
5.6	Sintesis dan Kesimpulan Pembahasan	90
BAB 6 KESIMPULAN DAN SARAN		92
6.1	Kesimpulan.....	92
6.1.1	Jawaban atas Rumusan Masalah	92
6.2	Saran.....	94
6.2.1	Saran untuk BPS.....	94
6.2.2	Saran untuk Kementerian Ketenagakerjaan	95
6.2.3	Saran untuk Penelitian Selanjutnya	96

6.3 Penutup.....96
DAFTAR PUSTAKA.....98



DAFTAR TABEL

Tabel 2.1 Tabel Ringkasan Penelitian Terdahulu	20
Tabel 4.1 Deskripsi Data Penelitian.....	57
Tabel 4.2 Statistik Deskriptif Variabel Penelitian (Skala Asli)	58
Tabel 4.3 Statistik Deskriptif Data Ternormalisasi.....	59
Tabel 4.4 Metrik Validasi Interpolasi	60
Tabel 4.5 Analisis Kualitas Interpolasi Berbagai Frekuensi.....	63
Tabel 4.6 Konfigurasi Model BSTS Terpilih.....	64
Tabel 4.7 Koefisien Regresi BSTS	64
Tabel 4.8 Metrik Performa Model BSTS.....	65
Tabel 4.9 Konfigurasi Model SARIMA Terpilih.....	68
Tabel 4.10 Parameter Signifikan Model SARIMA.....	69
Tabel 4.11 Metrik Performa Model SARIMA	69
Tabel 4.12 Arsitektur Model LSTM	72
Tabel 4.13 Metrik Performa Model LSTM.....	73
Tabel 4.14 Perbandingan Performa Ketiga Model pada Test Set.....	75
Tabel 4.15 Statistik Error pada Test Set.....	77
Tabel 4.16 Hasil Diebold-Mariano Test.....	78
Tabel 4.17 Ringkasan Temuan Utama.....	78

DAFTAR GAMBAR

Gambar 1.1 Grafik Angkatan Kelompok Umur.....	2
Gambar 2.1 Kerangka Berpikir.....	37
Gambar 4.1 Perbandingan Data Semester vs Data Triwulan Hasil Interpolasi	61
Gambar 4.2 Validasi Interpolasi pada Titik Observasi.....	62
Gambar 4.3 Feature Importance Plot BSTS.....	65
Gambar 4.4 Actual vs Predicted - BSTS (Full Series).....	66
Gambar 4.5 Test Set Performance - BSTS.....	67
Gambar 4.6 Residual Diagnostics - BSTS.....	67
Gambar 4.7 Actual vs Predicted - SARIMA (Full Series).....	70
Gambar 4.8 Residual Diagnostics - SARIMA	71
Gambar 4.9 LSTM Training History	73
Gambar 4.10 Actual vs Predicted - LSTM (Full Series).....	74
Gambar 4.11 Test Set Performance – LSTM.....	74
Gambar 4.12 Bar Chart Perbandingan MAPE	75
Gambar 4.13 Bar Chart Perbandingan R ²	76
Gambar 4.14 Perbandingan Prediksi pada Test Set	76
Gambar 4.15 Distribusi Error Ketiga Model.....	77
Gambar 4.16 Executive Summary - Perbandingan Ketiga Model.....	79