

Pemodelan dan Prediksi Tingkat Pengangguran Menggunakan Pendekatan Hibrida GARCH dan BSTS

Ade Priyatna^{1,*}, Eva Zuraidah², Besus Maula Sulthon², Oky Kurniawan³

¹ Fakultas Teknik dan Informatika, Informatika, Universitas Bina Sarana Informatika, Jakarta, Indonesia

² Fakultas Teknik dan Informatika, Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

³ Fakultas Teknik dan Informatika, Teknologi Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Email: ^{1,*}ade.aeq@bsi.ac.id, ²eva.evz@bsi.ac.id, ³maulasyarif@gmail.com, ⁴oky.okw@bsi.ac.id

Email Penulis Korespondensi: ade.aeq@bsi.ac.id

Abstrak—Penelitian ini bertujuan untuk memahami dan memprediksi pola tingkat pengangguran berdasarkan latar belakang pendidikan di Indonesia antara 1986 hingga 2024, dengan fokus pada lulusan Universitas. Data yang awalnya kompleks berhasil diolah menjadi format yang siap untuk analisis deret waktu, dan model GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) diterapkan untuk mengukur volatilitas pengangguran. Hasil evaluasi menunjukkan bahwa asumsi model GARCH mengenai kestabilan rata-rata pengangguran tidak akurat, terbukti dari nilai error yang besar (RMSE 283209.26 dan MAE 246252.37), yang mengindikasikan bahwa model ini tidak sepenuhnya dapat menangkap fluktuasi pengangguran, dimana koefisien rata-rata (μ) sebesar 436.50 dan *log-likelihood* sebesar -284.05 serta nilai peramalan volatilitas bersyarat berkisar dari sekitar $1.91e+11$ hingga $2.79e+11$. Model Bayesian Structural Time Series (BSTS) juga diterapkan untuk memecah data menjadi komponen tren jangka panjang dan pola musiman, memberikan gambaran yang lebih jelas tentang pergerakan pengangguran. Namun, kendala teknis dalam implementasi BSTS menggunakan *TensorFlow Probability* menyebabkan prediksi belum dapat diselesaikan. Meskipun demikian, analisis ini menunjukkan bahwa tingkat pengangguran lulusan Universitas sangat fluktuatif, dan perbaikan pada model GARCH serta penyelesaian kendala teknis pada model BSTS sangat penting untuk menghasilkan prediksi yang lebih akurat dan dapat diandalkan.

Kata Kunci: Pengangguran; Model GARCH; Bayesian Structural Time Series (BSTS); Volatilitas Pengangguran; Deret Waktu

Abstract—This study aims to understand and predict the unemployment rate patterns based on educational background in Indonesia between 1986 and 2024, with a focus on university graduates. The data, which was initially complex, was successfully processed into a format ready for time series analysis, and the GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) model was applied to measure the volatility of unemployment. The evaluation results show that the GARCH model's assumption regarding the stability of the average unemployment rate is inaccurate, as evidenced by the large error values (RMSE 283209.26 and MAE 246252.37), indicating that this model does not fully capture the fluctuations in unemployment. The average coefficient (μ) is 436.50, and the log-likelihood is -284.05, with conditional volatility forecast values ranging from approximately $1.91e+11$ to $2.79e+11$. The Bayesian Structural Time Series (BSTS) model was also applied to decompose the data into long-term trend components and seasonal patterns, providing a clearer picture of unemployment movement. However, technical constraints in the implementation of BSTS using *TensorFlow Probability* resulted in predictions not being completed. Nevertheless, this analysis shows that the unemployment rate of university graduates is highly volatile, and improvements in the GARCH model, as well as resolution of the technical constraints in the BSTS model, are crucial for generating more accurate and reliable predictions.

Keywords: Unemployment; GARCH Model; Bayesian Structural Time Series (BSTS); Unemployment Volatility; Time Series Analysis

1. PENDAHULUAN

Tingkat pengangguran merupakan indikator makroekonomi fundamental yang mencerminkan kondisi pasar tenaga kerja serta stabilitas perekonomian suatu negara [1]. Pergerakan angka pengangguran sangat dipengaruhi oleh siklus bisnis, tetapi juga dipengaruhi oleh berbagai faktor struktural dan kebijakan pemerintah. Fluktuasi tingkat pengangguran memiliki dampak yang luas, mulai dari kesejahteraan individu hingga kinerja ekonomi secara keseluruhan, menjadikannya area fokus utama bagi ekonom, analis, dan pembuat kebijakan dalam upaya memahami, memantau, dan merespons perubahan kondisi ekonomi [2]. Oleh karena itu, pengembangan dan penerapan metode prediksi yang akurat untuk tingkat pengangguran menjadi sangat krusial dalam mendukung proses perencanaan strategis dan mitigasi risiko ekonomi.

Pada penelitian sebelumnya yang menggunakan kombinasi model ARIMA-GARCH untuk memprediksi inflasi, ditemukan bahwa pendekatan ini cukup efektif dalam menangkap pola tren dan volatilitas data ekonomi. Namun, dalam studi oleh Agisna Mutiara, dilakukan perbandingan antara model tersebut dan model Long Short Term Memory (LSTM) — sebuah algoritma deep learning yang mampu mengingat hubungan jangka panjang dalam data deret waktu. Berdasarkan data inflasi bulanan Indonesia dari Januari 1997 hingga September 2023, analisis menunjukkan bahwa meskipun ARIMA-GARCH menghasilkan MAPE sebesar 33.72%, LSTM memberikan hasil yang lebih akurat dengan MAPE sebesar 12.00%. Dengan demikian, penelitian ini menyimpulkan bahwa model LSTM lebih unggul dalam memprediksi inflasi dan berpotensi menjadi alat bantu yang lebih andal dalam pengambilan keputusan ekonomi dan kebijakan moneter [3].

Analisis deret waktu terhadap data ekonomi seperti tingkat pengangguran seringkali dihadapkan pada tantangan kompleksitas data, termasuk tren jangka panjang, pola musiman, serta volatilitas yang bervariasi seiring waktu (heteroskedastisitas). Metode prediksi tradisional seperti model ARIMA (*Autoregressive Integrated Moving Average*) seringkali mengasumsikan bahwa varians residual bersifat konstan (*homoskedastisitas*), sebuah asumsi yang kurang realistis ketika menghadapi data ekonomi atau finansial yang menunjukkan periode volatilitas tinggi dan rendah secara

bergantian, seperti pada saat krisis ekonomi. Fenomena volatilitas yang berubah-ubah ini dapat dimodelkan dengan efektif menggunakan model *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) yang diperkenalkan oleh Bollerslev (1986), yang memungkinkan varians bersyarat bergantung pada kuadrat kesalahan masa lalu [4]. Penerapan model GARCH dalam prediksi variabel makroekonomi, termasuk pengangguran, telah banyak dilakukan untuk menangkap dan memprediksi periode ketidakpastian yang meningkat [5].

Selain memodelkan volatilitas, pemahaman dan pemisahan komponen struktural dalam deret waktu juga sangat penting untuk prediksi yang lebih akurat. Model *Structural Time Series* (STS) menyediakan kerangka yang elegan untuk mendekomposisikan deret waktu menjadi komponen-komponen tidak teramati (*unobserved components*) seperti tren, musiman, dan siklus [6]. Pendekatan Bayesian terhadap model struktural, seperti *Bayesian Structural Time Series* (BSTS), memberikan fleksibilitas lebih dengan memungkinkan inferensi Bayesian atas parameter model dan komponennya, serta kemampuan untuk menggabungkan informasi dari variabel prediktor lain dan menangani perubahan struktural [7]. Model BSTS telah menunjukkan kinerja yang sangat baik dalam prediksi berbagai deret waktu ekonomi dan bisnis [8].

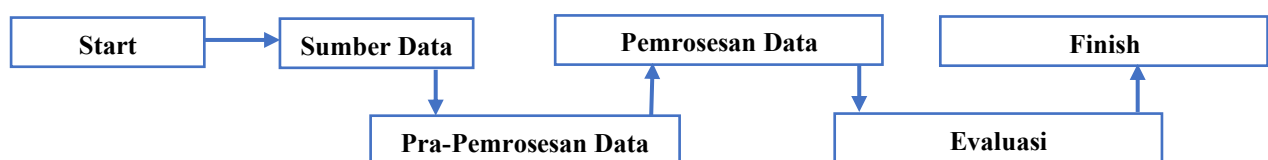
Penelitian selanjutnya yang mengembangkan model ARFIMA-GARCH dalam prediksi harga daging sapi dapat difokuskan pada penerapan teknik machine learning berbasis hybrid untuk meningkatkan ketepatan dan efisiensi prediksi, seperti kombinasi ARFIMA dengan neural network atau ensemble learning [9]. Selain itu, integrasi faktor eksternal seperti pola konsumsi, musim, dan kebijakan pemerintah dapat memperkaya konteks model sehingga lebih adaptif terhadap dinamika pasar. Penggunaan data berfrekuensi lebih tinggi serta pemodelan multivariat juga berpotensi menghasilkan insight yang lebih mendalam mengenai pergerakan harga komoditas, serta membuka peluang pengembangan sistem peringatan dini untuk fluktuasi harga di sektor peternakan [10].

Mengingat karakteristik data pengangguran yang kompleks terdapat tren, potensi musiman, dan volatilitas bersyarat pendekatan yang mengintegrasikan kemampuan pemodelan struktural dan volatilitas menjadi sangat relevan [11]. Kombinasi model struktural untuk menangkap pergerakan rata-rata deret waktu (tren dan musiman) dan model GARCH untuk memodelkan volatilitas residual di sekitar rata-rata tersebut dapat memberikan gambaran yang lebih menyeluruh untuk prediksi dan analisis risiko. Penelitian ini mengusulkan penggunaan pendekatan hibrida dengan menerapkan model GARCH dan BSTS pada data tingkat pengangguran di Indonesia berdasarkan tingkat pendidikan. Tujuan penelitian ini adalah untuk mengeksplorasi bagaimana kedua kerangka pemodelan ini dapat digunakan, baik secara terpisah maupun terintegrasi, dalam menganalisis dinamika historis dan memprediksi tren pengangguran di masa depan [12]. Meskipun tantangan teknis dalam implementasi inferensi Bayesian untuk model BSTS ditemukan selama proses analisis, penelitian ini menyajikan langkah awal dalam pemrosesan data, pemodelan volatilitas dengan GARCH, serta memberikan arah untuk prediksi struktural guna memperoleh pemahaman yang lebih dalam mengenai prospek pasar tenaga kerja.

2. METODOLOGI PENELITIAN

2.1 Tahap Penelitian

Pada penelitian ini melalui beberapa tahapan seperti yang bisa dilihat pada Gambar 1.



Gambar 1. Tahap Penelitian

2.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data Tingkat Pengangguran Terbuka (TPT) menurut Pendidikan Tertinggi yang Ditamatkan dari tahun 1986 hingga 2024. Data diperoleh dari sumber publik dari BPS dengan link <https://www.bps.go.id/id/statistics-table/1/OTcyIzE=/pengangguran-terbuka-menurut-pendidikan-tertinggi-yang-ditamatkan-1986---2023.html>.

2.3 Pra-Pemrosesan Data

Berikut merupakan representasi skematik proses transformasi data dalam bentuk flowchart. Diagram ini menggambarkan urutan tahapan yang dilakukan untuk mengakses, membersihkan, menstrukturkan, serta merekonstruksi data mentah menjadi bentuk deret waktu yang siap dianalisis. Setiap langkah dirancang secara sistematis untuk memastikan akurasi, konsistensi, dan validitas data dalam konteks pemodelan statistik dan komputasional [13].



Gambar 2. Pra-Pemrosesan Data

Penelitian ini menggunakan data deret waktu mengenai tingkat pengangguran berdasarkan tingkat pendidikan di Indonesia selama periode 1986-2024. Data diperoleh dari sumber Badan Pusat Statistik (BPS). Data asli disajikan dalam format tabel silang dengan tahun dan bulan/periode sebagai kolom, dan tingkat pendidikan sebagai baris. Tahap Pra-pemrosesan Data yang bisa dilihat pada Gambar 2, meliputi beberapa langkah kritis untuk mengubah data mentah menjadi format yang sesuai untuk analisis deret waktu [14]:

a. Pemuatan Data

Data diakses melalui https://drive.google.com/file/d/18E4V_ZIBqg233MwDjgWLohfqV6jJqfrd/view?usp=sharing, menggunakan pustaka pandas di Python. Penanganan khusus diperlukan untuk membaca file yang mungkin memiliki format header tidak standar atau pemisah non-standar.

```

import pandas as pd

# load the dataset
df_unemployment = pd.read_csv('https://drive.google.com/file/d/18E4V_ZIBqg233MwDjgWLohfqV6jJqfrd/view?usp=sharing')

# Display the first few rows
display(df_unemployment.head())

# Display summary information
display(df_unemployment.info())
  
```

Gambar 3. Pra-Pemrosesan Membaca Data

Gambar 3 Pra-Pemrosesan Membaca Data, berfungsi untuk memuat data ke dalam variabel *df_unemployment*, kode menampilkan lima baris pertama dari dataset menggunakan fungsi *head()* untuk memberikan gambaran awal tentang data, serta menampilkan informasi ringkas tentang *dataset*, seperti jumlah baris, kolom, dan tipe data pada setiap kolom menggunakan fungsi *info()*.

b. Ekstraksi Header dan Data

Baris-baris yang mengandung informasi tahun dan bulan diekstraksi untuk membangun nama kolom yang bermakna. Bagian data utama yang relevan dengan tingkat pengangguran per tingkat pendidikan kemudian dipilih.

```

header_years = df_unemployment.iloc[1]
header_months = df_unemployment.iloc[2]

df_data_1 = df_unemployment.iloc[4:12].copy()
  
```

Gambar 4. Pra-Pemrosesan Ekstraksi Header dan Data

Kode Gambar 4 ini berfungsi mengakses data dari DataFrame *df_unemployment* menggunakan metode *.iloc[]* untuk memilih baris tertentu berdasarkan indeks. Baris pertama dan kedua (indeks 1 dan 2) diambil untuk menjadi header yang berisi tahun (*header_years*) dan bulan (*header_months*). Kemudian, baris dari indeks 4 hingga 12 disalin ke dalam variabel *df_data_1* menggunakan *.copy()*, yang memastikan bahwa data tersebut terpisah dari DataFrame asli untuk mencegah perubahan yang tidak diinginkan.

c. Pembersihan dan Penamaan Kolom

Nama kolom diperbaiki menggunakan informasi header tahun dan bulan/periode.

```

cleaned_columns = [df_unemployment.columns[0], df_unemployment.columns[1]]
for i in range(2, len(df_unemployment.columns)):
    year = header_years[i]
    month = header_months[i]
    if pd.notna(year):
        year_str = str(int(year)) if pd.notna(year) and isinstance(year, (int, float)) else str(year)
        if pd.notna(month):
            cleaned_columns.append(f"{year_str}_{month}")
        else:
            cleaned_columns.append(f"{year_str}_NaN")
    else:
        if pd.notna(month):
            cleaned_columns.append(f"NaN_{month}")
        else:
            cleaned_columns.append(f"Unnamed_{i}")

df_data_1.columns = cleaned_columns[:len(df_data_1.columns)]
display(df_data_1.head())
display(df_data_1.info())
  
```

Gambar 5. Pra-Pemrosesan Pembersihan dan Penamaan Kolom

d. Transformasi Data (*Melt*)

Data diubah dari format lebar menjadi format panjang menggunakan fungsi `melt` pandas, dengan kolom tingkat pendidikan dan nilai pengangguran terkait. Kolom pengenalan ('No.' dan 'Tingkat Pendidikan') serta kolom gabungan tahun-bulan ('Year_Month') dihasilkan.

```
df_melted = df_data_1.melt(
    id_vars=[df_data_1.columns[0], df_data_1.columns[1]],
    var_name='Year_Month',
    value_name='Unemployment Value'
)

df_melted.rename(columns={
    df_data_1.columns[0]: 'No.',
    df_data_1.columns[1]: 'Education Level'
}, inplace=True)

display(df_melted.head())

display(df_melted.info())
```

Gambar 6. Pra-Pemrosesan Transformasi Data (*Melt*)

e. Pemecahan dan Konversi Data

Kolom 'Year_Month' dipecah menjadi 'Year' dan 'Month'. Nilai pengangguran dikonversi ke tipe numerik, dengan nilai yang tidak valid ditangani sebagai missing values (NaN). Kolom 'Year' dikonversi menjadi integer, dan nama bulan (misalnya, 'Februari', 'Agustus') dipetakan ke nomor bulan yang sesuai.

```
df_melted[['Year', 'Month']] = df_melted['Year_Month'].str.split('_', expand=True)

df_melted['Unemployment Value'] = pd.to_numeric(df_melted['Unemployment Value'], errors='coerce')

df_melted.dropna(subset=['Unemployment Value'], inplace=True)

df_melted['Year'] = pd.to_numeric(df_melted['Year'], errors='coerce')
df_melted.dropna(subset=['Year'], inplace=True)
df_melted['Year'] = df_melted['Year'].astype(int)

month_map = {
    'Februari': 2,
    'Agustus': 8,
    'November': 11,
    'NaN': ''
}

df_melted['Month_Num'] = df_melted['Month'].map(month_map)
df_melted.dropna(subset=['Month_Num'], inplace=True)
df_melted['Month_Num'] = df_melted['Month_Num'].astype(int)
```

Gambar 7. Pra-Pemrosesan Pemecahan dan Konversi Data

f. Pembentukan Indeks Deret Waktu

Kolom tanggal ('Date') dibuat dari informasi tahun dan nomor bulan, dan ditetapkan sebagai indeks deret waktu (`DatetimeIndex`). Baris dengan tanggal atau nilai pengangguran yang hilang dihilangkan.

```
df_melted['Date'] = pd.to_datetime(df_melted['Year'].astype(str) + '-' + df_melted['Month_Num'].astype(str) + '-01')

df_melted.set_index('Date', inplace=True)

df_melted.drop(columns=['No.', 'Year_Month', 'Year', 'Month', 'Month_Num'], inplace=True)

display(df_melted.head())
display(df_melted.info())
```

Gambar 8. Pra-Pemrosesan Pembentukan Indeks Deret Waktu

g. Pivot Data

Data dipivot kembali ke format lebar dengan 'Date' sebagai indeks dan tingkat pendidikan sebagai kolom, menghasilkan DataFrame deret waktu (`df_pivot`) yang siap untuk pemodelan.

```
df_pivot = df_melted.pivot_table(index='Date', columns='Education Level', values='Unemployment Value')

df_pivot.sort_index(inplace=True)

display(df_pivot.head())

display(df_pivot.info())
```

Gambar 9. Pra-Pemrosesan Pivot Data

Kode pada Gambar 9 tabel pivot dari *DataFrame* `df_melted` dengan menggunakan kolom *Date* sebagai *indeks*, kolom *Education Level* sebagai kolom, dan nilai dari *Unemployment Value* sebagai nilai dalam tabel pivot. Setelah itu, kode mengurutkan tabel pivot berdasarkan indeks dengan `sort_index(inplace=True)`. Kemudian, kode menampilkan lima baris pertama dari tabel pivot menggunakan `head()` dan memberikan informasi ringkas tentang struktur tabel menggunakan `info()`.

2.4 Pemrosesan Data

Setelah data dipersiapkan, Pemodelan Deret Waktu dilakukan menggunakan dua pendekatan utama:

2.4.1 Model GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*)

Model *Generalized Autoregressive Conditional Heteroskedasticity* (GARCH) merupakan pengembangan dari model ARCH yang diperkenalkan oleh Engle (1982), dengan generalisasi oleh Bollerslev (1986) untuk menangani varian residual yang tidak konstan (*heteroskedastis*) pada data deret waktu [15]. Model ini secara khusus dirancang untuk menangkap fenomena *volatility clustering*, di mana periode volatilitas tinggi cenderung diikuti oleh periode volatilitas tinggi lainnya, dan sebaliknya. Dalam konteks analisis ekonomi dan keuangan, seperti tingkat pengangguran atau harga saham, *volatilitas* sering kali mengalami fluktuasi tajam yang tidak dapat ditangkap secara efektif oleh model tradisional dengan asumsi homoskedastisitas [16]. Oleh karena itu, model GARCH banyak digunakan dalam pemodelan variabel ekonomi yang bersifat volatil dan berisiko tinggi. Penerapan model GARCH(1,1) memungkinkan identifikasi pola ketergantungan volatilitas saat ini terhadap volatilitas masa lalu dan besarnya kejutan (shock) pada periode sebelumnya, sehingga sangat tepat digunakan dalam meramalkan ketidakpastian pada data pengangguran maupun keuangan yang bersifat dinamis [17].

- Model GARCH diterapkan untuk memodelkan volatilitas bersyarat dari deret waktu tingkat pengangguran. Dalam analisis ini, data pengangguran untuk tingkat pendidikan Universitas dipilih sebagai studi kasus.
- Model spesifikasi GARCH(1,1) dengan model rata-rata konstan (*mean=Constant*) digunakan, di mana p menunjukkan orde ARCH dan q menunjukkan orde GARCH.
- Model diestimasi menggunakan metode *Maximum Likelihood* (ML) dengan memanfaatkan pustaka *arch* di *Python*. Data diskalakan sebelum fitting untuk membantu konvergensi optimizer.
- Hasil fitting model GARCH, termasuk estimasi parameter dan ringkasan statistik, dicatat.
- Prediksi GARCH: Model GARCH yang telah diestimasi digunakan untuk menghasilkan prediksi volatilitas bersyarat untuk horison prediksi 12 langkah ke depan (*forecast_horizon = 12*). Hasil prediksi volatilitas ini kemudian diskalakan kembali ke skala asli data.

2.4.2 Model Bayesian Structural Time Series (BSTS)

Dalam upaya meningkatkan ketepatan prediksi data deret waktu yang bersifat kompleks dan mengandung ketidakpastian struktural, penelitian ini mengadopsi pendekatan *Bayesian Structural Time Series* (BSTS) [18]. Model BSTS merupakan perluasan dari model *State Space* dengan pendekatan Bayesian, yang memungkinkan pemodelan tren, musiman, dan komponen regresi secara fleksibel dan terpadu [18]. Salah satu keunggulan utama dari BSTS adalah kemampuannya dalam menangani ketidakpastian parameter melalui simulasi distribusi posterior menggunakan algoritma *Markov Chain Monte Carlo* (MCMC), sehingga memberikan hasil prediksi yang lebih halus dan akurat melalui Bayesian Model Averaging [19].

Model ini dirancang untuk mengidentifikasi struktur data yang tidak terlihat secara langsung (*latent*), seperti perubahan tren atau fluktuasi musiman, dan sangat sesuai digunakan dalam kasus di mana fluktuasi acak dan variabilitas musiman sulit dimodelkan oleh pendekatan klasik seperti ARIMA. Dalam konteks studi ini, BSTS digunakan untuk meramalkan tingkat pengangguran berdasarkan tingkat pendidikan, dengan mengintegrasikan komponen local level, local linear trend, dan seasonal, serta mengestimasi parameter menggunakan MCMC dengan beberapa konfigurasi iterasi untuk mengoptimalkan nilai *R-square* [20].

- Model BSTS direncanakan untuk memodelkan komponen struktural deret waktu (tren dan musiman) menggunakan pustaka *tensorflow_probability.sts* di *Python*.
- Model awal BSTS dibangun dengan menyertakan komponen *LocalLevel* untuk menangkap tren dan komponen *Seasonal* dengan periode 2 (mengasumsikan pola bi-annual Februari-Agustus berdasarkan struktur data).
- Inferensi Bayesian atas parameter model BSTS diupayakan menggunakan teknik *Markov Chain Monte Carlo* (MCMC). Berbagai pendekatan MCMC, termasuk *tfp.sts.fit_with_hmc* dan setup manual dengan *tfp.mcmc.sample_chain*, dicoba dengan konfigurasi seperti jumlah chain (4), jumlah hasil sampel (1000), dan langkah burn-in (500).
- Prediksi BSTS: Model BSTS yang telah diestimasi melalui inferensi MCMC (jika berhasil) direncanakan untuk digunakan memprediksi komponen struktural dan menghasilkan prediksi untuk horison 12 langkah ke depan, termasuk estimasi interval kepercayaan.

2.5 Evaluasi Model dilakukan untuk menilai kinerja model

- Untuk model GARCH, evaluasi *in-sample fit* dilakukan dengan membandingkan nilai pengangguran aktual dengan nilai rata-rata konstan yang diestimasi oleh model. Metrik *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) dihitung untuk mengukur kesalahan prediksi in-sample rata-rata konstan [21].
- Evaluasi *out-of-sample forecasting* untuk kedua model (GARCH dan BSTS) tidak dapat dilakukan secara penuh dalam penelitian ini karena ketiadaan data aktual di masa depan untuk dibandingkan dengan hasil prediksi. Dalam implementasi penuh, data historis akan dibagi menjadi set pelatihan dan pengujian untuk validasi silang prediksi.

3. HASIL DAN PEMBAHASAN

3.1 Transformasi Data

3.1.1 Pemahaman Data dan Preprocessing Data

Pada tahap pemahaman data, data pengangguran dimuat dari sumber yang diberikan dan pahami struktur awalnya, termasuk format header yang tidak standar dan tata letak data.

Tabel 1. Pemahaman Data dan Preprocessing Data

Tahap	Sub-tahap / Aktivitas	Tujuan	Kode Blok Terkait
Pemahaman Data	Memuat Data dari Sumber	Mengakses data mentah dari file/tautan sumber.	Kode yang mengunduh dan membaca file CSV.
Memahami Struktur Awal	Mengenal format header tidak standar, baris data, dan kolom generik.	Output df.head(), df.info(), dan observasi manual struktur tabel.	Output df.head(), df.info(), dan observasi manual struktur tabel.
Preprocessing Data	Ekstraksi Header dan Data	Mengisolasi baris header (tahun, bulan) dan baris data utama.	Kode yang menggunakan iloc untuk memilih baris header dan data.
Pembersihan & Penamaan Kolom	Membuat nama kolom yang deskriptif dari informasi header.	Kode yang menggabungkan tahun & bulan dan menetapkan nama kolom baru ke DataFrame data.	Kode yang menggunakan iloc untuk memilih baris header dan data.
Transformasi Data (Melt)	Mengubah data dari format lebar ke format panjang.	Kode yang menggunakan fungsi melt pandas.	Kode yang menggabungkan tahun & bulan dan menetapkan nama kolom baru ke DataFrame data.
Pemecahan & Konversi Data	Memisahkan informasi tahun/bulan, mengubah tipe data, menangani missing values.	Kode yang menggunakan str.split, pd.to_numeric, .astype(int), .map.	Kode yang menggunakan fungsi melt pandas.
Pembentukan Indeks Deret Waktu	Membuat kolom tanggal dan menjadikannya indeks DataFrame.	Kode yang membuat kolom Date menggunakan pd.to_datetime dan set_index.	Kode yang menggunakan str.split, pd.to_numeric, .astype(int), .map.
Pivot Data	Mengubah kembali data ke format lebar dengan indeks deret waktu.	Kode yang menggunakan fungsi pivot_table pandas.	Kode yang membuat kolom Date menggunakan pd.to_datetime dan set_index.
Pemilihan Deret Waktu Spesifik (untuk analisis)	Memilih satu kolom tingkat pendidikan sebagai fokus pemodelan.	Kode yang memilih kolom tertentu dari df_pivot dan menghapus NaN.	Kode yang menggunakan fungsi pivot_table pandas.

Tahap awal dalam analisis ini dimulai dengan Pemahaman Data, yaitu proses mengakses dan mempelajari struktur data mentah yang umumnya tidak langsung siap untuk dianalisis seperti pada Tabel 1. Pada tahap ini, penting untuk mengidentifikasi letak header, struktur kolom, serta bagian mana yang mengandung data inti, agar pemrosesan selanjutnya berjalan efektif. Setelah struktur dikenali, dilanjutkan dengan Preprocessing Data, yaitu rangkaian langkah sistematis untuk membersihkan dan membentuk ulang data menjadi format deret waktu yang dapat digunakan untuk pemodelan. Proses ini meliputi penentuan header yang tepat (gabungan tahun dan bulan), pemilihan baris data yang valid, transformasi dari format lebar ke format panjang, konversi nilai numerik secara akurat, hingga pembuatan indeks berbasis tanggal. Data kemudian dipivot kembali ke format lebar, namun dengan struktur yang telah terorganisir, di mana setiap tingkat pendidikan menjadi satu deret waktu terpisah. Secara keseluruhan, tahap pemahaman data bertujuan untuk mengenali isi dan struktur data, sedangkan preprocessing berfungsi untuk menyiapkan data agar dapat dianalisis lebih lanjut menggunakan model deret waktu seperti GARCH dan BSTS.

Tabel 2. Data Awal Pasca Pre-Prosesing

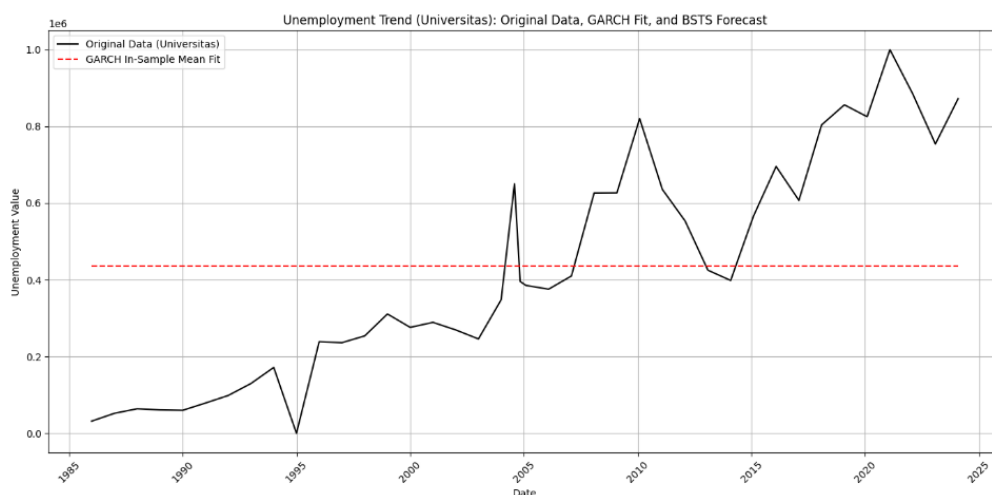
No	Pendidikan	1986	.	.	2024
1	Tidak/belum pernah sekolah	52166	.	.	12651
2	Tidak/belum tamat SD	161626	.	.	361769
.	.	.	.	.	.
.	.	.	.	.	.
8	Universitas	31790	.	.	842378
	Total	1817672	.	.	7465599

3.2 Pemodelan GARCH dan Pemodelan BSTS

Setelah data pengangguran berhasil dipersiapkan dan diformat dengan baik, langkah selanjutnya adalah menerapkan model deret waktu untuk menganalisis karakteristiknya dan melakukan prediksi. Mengingat data pengangguran seringkali menunjukkan adanya pola struktural seperti tren dan musiman, serta periode volatilitas yang bervariasi, dua jenis model yang relevan dan saling melengkapi dipilih untuk analisis ini, model GARCH untuk menangkap dinamika volatilitas bersyarat, dan model *Bayesian Structural Time Series* (BSTS) untuk mendekomposisikan dan memodelkan komponen struktural deret waktu secara Bayesian.

Tabel 3. Analisis Pemodelan GARCH dan Pemodelan BSTS

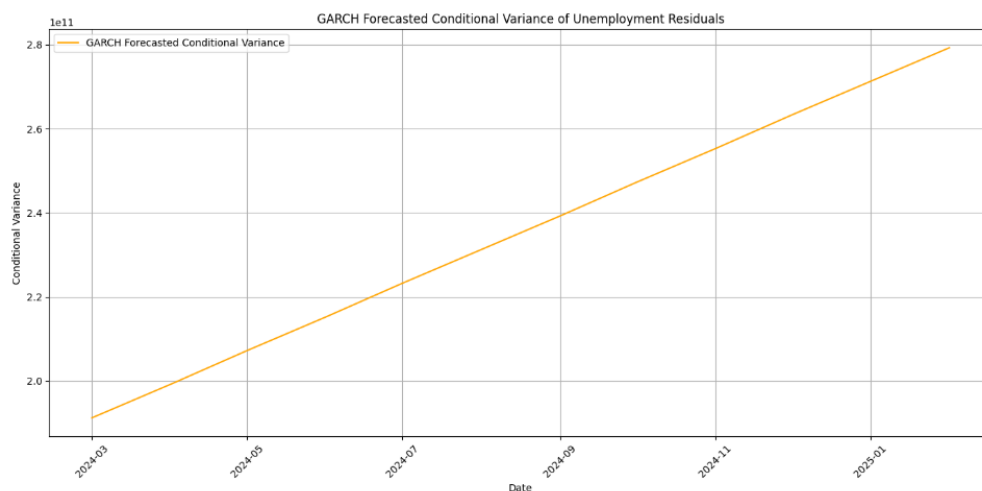
Tahap	Sub-tahap / Aktivitas	Tujuan	Kode Blok Terkait
Pemodelan GARCH	Pemilihan Deret Waktu	Memilih satu deret waktu spesifik (cth: 'Universitas') untuk analisis volatilitas.	Kode yang memilih kolom dari <code>df_pivot</code> .
Spesifikasi Model GARCH	Menentukan struktur model rata-rata dan varians (cth: Constant Mean, GARCH(1,1)).	Kode yang mendefinisikan <code>arch.arch_model</code> dengan argumen <code>mean='Constant'</code> , <code>vol='Garch'</code> , <code>p=1</code> , <code>q=1</code> .	Kode yang mendefinisikan <code>arch.arch_model</code> dengan argumen <code>mean='Constant'</code> , <code>vol='Garch'</code> , <code>p=1</code> , <code>q=1</code> .
Estimasi Parameter Model	Menemukan nilai parameter model GARCH yang paling sesuai dengan data historis.	Kode yang menggunakan metode <code>.fit()</code> dari objek model GARCH.	Kode yang menggunakan metode <code>.fit()</code> dari objek model GARCH.
Evaluasi In-Sample Fit (Rata-rata Konstan)	Mengukur seberapa baik rata-rata konstan model GARCH merepresentasikan data historis.	Kode yang menghitung RMSE dan MAE dari rata-rata konstan terhadap data asli.	
Pemodelan BSTS	Pemilihan Deret Waktu & Konversi ke Tensor	Memilih deret waktu yang sama dan mengubahnya ke format tensor yang diperlukan oleh TensorFlow Probability.	Kode yang memilih kolom dari <code>df_pivot</code> , menghapus NaN, dan mengubah ke array numpy float32.
Definisi Komponen BSTS	Menentukan komponen struktural yang relevan (cth: <code>LocalLevel</code> untuk tren, <code>Seasonal</code> untuk musiman).	Kode yang membuat objek <code>sts.LocalLevel</code> dan <code>sts.Seasonal</code> .	Kode yang menghitung RMSE dan MAE dari rata-rata konstan terhadap data asli.
Pembangunan Model BSTS	Menggabungkan komponen-komponen struktural menjadi satu model BSTS.	Kode yang menggunakan <code>sts.Sum</code> untuk menggabungkan komponen.	
Inferensi Bayesian (MCMC)	Mengestimasi distribusi probabilitas posterior dari parameter model BSTS menggunakan sampel MCMC.	Kode yang mencoba menggunakan <code>tfp.mcmc.sample_chain</code> dengan fungsi target log prob. (Bagian ini mengalami kendala).	
(Jika Berhasil) Analisis Hasil Inferensi	Memeriksa konvergensi MCMC dan karakteristik distribusi posterior parameter.	(Tidak sepenuhnya tercapai dalam kode yang ada karena error).	



Gambar 10. Unemployment Trend (Universitas): Original Data, GARCH Fit, and BSTS Forecast

Plot Gambar 10 ini bertujuan untuk menampilkan gambaran utama dari analisis deret waktu tingkat pengangguran untuk lulusan Universitas.

Berdasarkan gambar 9 *Unemployment Trend (Universitas): Original Data, GARCH Fit, and BSTS Forecast*, terlihat jelas bahwa tingkat pengangguran Universitas menunjukkan tren naik yang signifikan dan fluktuasi yang substansial dari tahun 1986 hingga 2024, sementara garis merah putus-putus yang merepresentasikan rata-rata konstan dari model GARCH gagal menangkap pergerakan historis ini, menyoroti keterbatasan model rata-rata yang sederhana; sayangnya, prediksi BSTS yang seharusnya menunjukkan proyeksi tren dan pola struktural masa depan tidak muncul di plot ini karena kendala dalam tahap pemodelan BSTS sebelumnya.



Gambar 11. *GARCH Forecasted Conditional Variance of Unemployment Residuals*

Plot Gambar 11 ini menampilkan prediksi volatilitas bersyarat (conditional variance) dari model GARCH untuk periode di masa depan (dalam kasus ini, 12 langkah ke depan dari titik data terakhir).

Berdasarkan gambar 10 *GARCH Forecasted Conditional Variance of Unemployment Residuals*, grafik garis oranye ini menampilkan proyeksi volatilitas (gejolak) residual tingkat pengangguran Universitas oleh model GARCH untuk 12 periode ke depan, mengindikasikan bagaimana tingkat ketidakpastian atau besarnya fluktuasi di sekitar rata-rata atau tren (jika model rata-rata yang lebih baik digunakan) diprediksi akan berubah di masa mendatang, dalam hal ini menunjukkan proyeksi peningkatan varians bersyarat dari waktu ke waktu.

3.3 Prediksi

```
Performing GARCH forecasting...
GARCH Forecasted Conditional Variances (12 steps) (Rescaled):
2024-02-01
h.01 1.912178e+11
h.02 1.992228e+11
h.03 2.072279e+11
h.04 2.152329e+11
h.05 2.232379e+11
h.06 2.312429e+11
h.07 2.392479e+11
h.08 2.472530e+11
h.09 2.552580e+11
h.10 2.632630e+11
h.11 2.712680e+11
h.12 2.792731e+11
dtype: float64
```

Gambar 12. *GARCH Forecasted Conditional Variances (12 steps) (Rescaled)*

Pada Gambar 12 menunjukkan hasil prediksi yang berhasil diperoleh adalah proyeksi volatilitas bersyarat tingkat pengangguran oleh model GARCH untuk 12 periode ke depan. Prediksi ini menunjukkan peningkatan varians bersyarat,

mengindikasikan proyeksi peningkatan gejala atau ketidakpastian di masa mendatang. Sedangkan prediksi BSTS tidak berhasil didapatkan karena masalah teknis sebelumnya, seperti gambar berikut ini.

Skipping BSTS Forecasting: trained_bsts_model or bsts_inference_results not available.

Stored Forecast Results:

```
{'garch_conditional_variances': h.01      1.912178e+11
  h.02      1.992228e+11
  h.03      2.072279e+11
  h.04      2.152329e+11
  h.05      2.232379e+11
  h.06      2.312429e+11
  h.07      2.392479e+11
  h.08      2.472530e+11
  h.09      2.552580e+11
  h.10      2.632630e+11
  h.11      2.712680e+11
  h.12      2.792731e+11
  Name: 2024-02-01 00:00:00, dtype: float64,
  'bsts_mean_forecast': None,
  'bsts_forecast_lower_bound': None,
  'bsts_forecast_upper_bound': None,
  'bsts_forecast_index': None}
```

Gambar 13. BSTS

Hal ini terjadi karena model BSTS yang sudah dilatih (*trained_bsts_model*) atau hasil inferensi Bayesiannya (*bsts_inference_results*) tidak berhasil dibuat pada tahap Pemodelan BSTS sebelumnya akibat kendala teknis seperti pada Gambar 13. Karena input yang dibutuhkan untuk prediksi BSTS tidak ada, kode secara otomatis melewati bagian tersebut.

3.4 Temuan Utama Analisis Data

- Langkah awal pemuatan dan pra-pemrosesan data, termasuk penanganan header yang tidak standar dan transformasi data ke format deret waktu yang sesuai dengan indeks tanggal, berhasil diselesaikan.
- Model GARCH(1,1) dengan rata-rata konstan berhasil di-fitting pada data deret waktu pengangguran 'Universitas'.
- Upaya fitting model Bayesian Structural Time Series (BSTS) menggunakan TensorFlow Probability tidak berhasil karena adanya kesalahan internal yang persisten selama proses inferensi MCMC.
- Prediksi GARCH untuk volatilitas bersyarat berhasil dilakukan untuk horison yang ditentukan.
- Prediksi BSTS tidak dapat dilakukan karena kegagalan pada tahap fitting model.
- Evaluasi in-sample fit rata-rata konstan model GARCH menghasilkan nilai RMSE (283209.26) dan MAE (246252.37) yang tinggi, mengindikasikan bahwa rata-rata konstan merupakan representasi yang buruk untuk data historis pengangguran 'Universitas'.
- Evaluasi out-of-sample tidak dapat dilakukan untuk kedua model dengan data yang tersedia karena tidak mencakup nilai aktual di masa depan.
- Visualisasi data asli, fit rata-rata konstan GARCH in-sample, dan prediksi volatilitas bersyarat GARCH berhasil dibuat. Visualisasi prediksi BSTS yang direncanakan dilewati karena kegagalan pemodelan.

3.5 Langkah Selanjutnya

- Langkah utama berikutnya adalah mengatasi isu teknis yang dihadapi dengan TensorFlow Probability agar model BSTS berhasil dilatih dan diperamalkan.
- Meningkatkan model GARCH dengan memasukkan model rata-rata yang lebih canggih (misalnya, komponen ARMA) untuk menangkap tren dan dinamika data deret waktu dengan lebih akurat.
- Menangani isu frekuensi data (tahunan vs. bi-annual) untuk pemodelan yang lebih akurat, terutama untuk musiman pada BSTS.
- Jika data masa depan tersedia, lakukan evaluasi out-of-sample untuk menilai kinerja prediksi model dengan benar.
- Mengembangkan metode untuk menggabungkan prediksi struktural BSTS dengan prediksi volatilitas GARCH untuk mendapatkan distribusi prediksi yang lengkap.

4. KESIMPULAN

Analisis data pengangguran 'Universitas' dari tahun 1986-2024 berhasil melalui tahap pemuatan dan pra-pemrosesan data yang kompleks; meskipun model GARCH(1,1) dengan rata-rata konstan berhasil di-fitting dan peramalan volatilitasnya didapatkan, evaluasi in-sample menunjukkan model rata-rata ini sangat tidak akurat (RMSE: 283209.26, MAE:

246252.37), sementara upaya pemodelan dan peramalan BSTS terhambat oleh kendala teknis inferensi MCMC, sehingga peramalan struktural tidak dapat diselesaikan dan evaluasi out-of-sample secara menyeluruh tidak memungkinkan; oleh karena itu, langkah krusial selanjutnya adalah mengatasi isu teknis BSTS, meningkatkan model rata-rata GARCH (misalnya dengan ARMA), menangani frekuensi data secara akurat, dan, jika data masa depan tersedia, melakukan evaluasi out-of-sample untuk akhirnya dapat menggabungkan peramalan struktural dan volatilitas guna mendapatkan prediksi yang komprehensif. Penelitian ini memiliki keterbatasan seperti kegagalan implementasi model BSTS akibat masalah teknis pada inferensi MCMC, serta terbatasnya evaluasi hanya pada data dalam-sample, mengurangi akurasi prediksi jangka panjang. Data tahunan yang digunakan juga membatasi analisis musiman dan fluktuasi pengangguran. Model GARCH(1,1) menunjukkan kesulitan dalam menangkap volatilitas dinamis dengan error tinggi, sementara masalah teknis pada BSTS menghambat penerapan model hibrida yang diusulkan.

UCAPAN TERIMA KASIH

Kami mengucapkan apresiasi yang mendalam kepada pihak penyedia data atas kontribusi pentingnya dalam mendukung kelancaran proses penelitian ini. Ucapan terima kasih juga kami sampaikan kepada para pengembang pustaka open-source seperti pandas, numpy, matplotlib, arch, dan tensorflow_probability yang telah menyediakan alat analisis yang sangat bermanfaat dalam pemrosesan dan interpretasi data. Seluruh dukungan dan kontribusi dari berbagai pihak memberikan peranan besar dalam keberhasilan dan penyelesaian studi ini secara menyeluruh.

REFERENCES

- [1] D. Tridayanti, I. Lili Rahma, A. Nainul Amani, G. Masitoh, and J. Baru Sukaraja Kec Buay Madang Kab Oku Timut, "Peran Ekonometrika dalam Perencanaan Pembangunan Daerah," *Jurnal Ekonomi dan Keuangan*, vol. 3, pp. 74–94, Apr. 2025, doi: 10.61132/moneter.v3i3.1393.
- [2] W. R. Amalia, "PENERAPAN MODEL ARCH/GARCH DALAM ANALISIS VOLATILITAS HARGA TUKAR RUPIAH TERHADAP USD MENGGUNAKAN DATA PERIODE JANUARI 2008 – AGUSTUS 2023," Skripsi, UNIVERSITAS ISLAM INDONESIA, Yogyakarta, 2023.
- [3] A. Mutiara, "ANALISIS PREDIKSI INFLASI DI INDONESIA: PERBANDINGAN MODEL ARIMA-GARCH DAN LONG SHORT TERM MEMORY (LSTM)," Skripsi, UIN Syarif Hidayatullah, Jakarta, 2024.
- [4] A. B. D. A. W. P. R. SUMIYATI, "METODE ARCH/GARCH UNTUK MEMPREDIKSI HUBUNGAN ECONOMIC UNCERTAINTY (COVID 19) DAN VOLATILITAS SAHAM," *Jurnal Bisnis dan Akuntansi*, vol. 24, no. 1, pp. 117–130, Jun. 2022, [Online]. Available: <http://jurnaltsm.id/index.php/JBA>
- [5] H. Nabilah, N. Effendi, A. F. Priyono, K. Kunci, and M. E. Terapan, "E-JURNAL EKONOMI DAN BISNIS UNIVERSITAS UDAYANA PREDIKSI INDIKATOR MAKRO EKONOMI INDONESIA PASCA PANDEMI COVID-19 MENGGUNAKAN ANALISIS INTERVENSI," *Jurnal Ekonomi dan Bisnis Universitas Udayana*, vol. Vol 12, no. 4, pp. 646–665, Apr. 2023, [Online]. Available: <https://ojs.unud.ac.id/index.php/EEB/index>
- [6] J. J. Pangaribuan, F. Fanny, O. P. Barus, and R. Romindo, "Prediksi Penjualan Bisnis Rumah Properti Dengan Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA)," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 154–161, Oct. 2023, doi: 10.21456/vol13iss2pp154-161.
- [7] F. F. Mojtahedi, N. Yousefpour, S. H. Chow, and M. Cassidy, "Deep Learning for Time Series Forecasting: Review and Applications in Geotechnics and Geosciences," *Archives of Computational Methods in Engineering*, Feb. 2025, doi: 10.1007/s11831-025-10244-5.
- [8] K. C. PRADANA, "ANALISIS DERET WAKTU UNTUK PREDIKSI GEMPA BUMI DI PROVINSI LAMPUNG MENGGUNAKAN METODE AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)," Skripsi, UNIVERSITAS ISLAM NEGERI, Lampung, 2021.
- [9] ISRAN K. HASAN dan NURWAN, "PENERAPAN-MODEL-ARFIMA-GARCH-MENGGUNAKAN-VARIASI-ESTIMASI-PARAMETER-PEMBEDA-D-UNTUK-MERAMALKAN-HARGA-EMAS," Gorontalo, 2023.
- [10] N. D. L. Imani, T. Tarno, and B. A. Saputra, "PREDIKSI HARGA DAGING SAPI DI KABUPATEN BREBES MENGGUNAKAN PEMODELAN ARFIMA DENGAN EFEK GARCH," *Jurnal Gaussian*, vol. 12, no. 4, pp. 570–580, Jul. 2024, doi: 10.14710/j.gauss.12.4.570-580.
- [11] M. Ningsih, "Prediksi Harga Saham Harian PT BTPN Syariah Tbk Menggunakan Model Arima dan Model Garch," *Jurnal Ilmiah Ekonomi Islam*, vol. 7, no. 03, pp. 1573–1580, 2021, doi: 10.29040/jiei.v7i3.2795.
- [12] S. Hariyanto, S. G. Wibawa, and Solikhin, "PM10 AIR QUALITY INDEX MODELING USING ARFIMA-GARCH METHOD: BUNDARAN HI AREA OF DKI JAKARTA PROVINCE," *Barekeng*, vol. 18, no. 4, pp. 2165–2180, Oct. 2024, doi: 10.30598/barekengvol18iss4pp2165-2180.
- [13] N. Srivastava, T. Lamba, and M. Agarwal, "Comparative Analysis of Different Machine Learning Techniques," *Communications in Computer and Information Science*, vol. 1206 CCIS, no. April, pp. 245–255, Apr. 2020, doi: 10.1007/978-981-15-4451-4_19.
- [14] A. Santoso, A. I. Purnamasari, and I. Ali, "PREDIKSI HARGA BERAS MENGGUNAKAN METODE RECURRENT NEURAL NETWORK DAN LONG SHORT-TERM MEMORY," *Jurnal PROSISKO*, vol. Vol 11, Mar. 2024.
- [15] M. I. Rizki, T. A. Taqiyyuddin, P. F. Rahmah, and A. E. Hasana, "Penerapan Model ARCH/GARCH untuk Memprediksi Harga Saham Perusahaan Tokai Carbon," *Jurnal Sains Matematika dan Statistika*, vol. 7, no. 2, Aug. 2021, doi: 10.24014/jsms.v7i2.13138.
- [16] D. Nugroho, O. Dimitrio, and F. Tita, "The GARCH-X(1,1) Model with Exponentially Transformed Exogenous Variables," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 1, Mar. 2023, doi: 10.23887/jstundiksha.v12i1.50714.

- [17] A. R. Hafizhah, D. A. I. Maruddani, and R. Santoso, "PERBANDINGAN METODE EXPONENTIAL GARCH (EGARCH) DAN GLOSTEN-JAGANNATHAN-RUNKLE GARCH (GJR-GARCH) PADA MODEL VOLATILITAS SAHAM TUNGGAL," *Jurnal Gaussian*, vol. 13, no. 1, pp. 199–209, Oct. 2024, doi: 10.14710/j.gauss.13.1.199-209.
- [18] H. A. Hamza and M. J. Mohammad, "A Comparison Of The Bayesian Structural Time Series Technique With The Autoregressive Integrated Moving Average Model For Forecasting," *Migration Letters*, vol. Vol 21, no. S4, pp. 528–538, 2024, [Online]. Available: www.migrationletters.com
- [19] I. Suciati and M. Usman, "Bayesian Structural Time Series Model for Forecasting the Composite Stock Price Index in Indonesia," *Journal of Statistics, Probability, and Its Application*, vol. Vol 1, no. 2, pp. 74–83, Jul. 2023, Accessed: Jul. 23, 2025. [Online]. Available: <https://scholar.ummetro.ac.id/index.php/sciencestatistics/index>
- [20] M. Muizzadin, Mohammad Idhom, and A. T. Damaliana, "Implementation of Bayesian Structural Time Series (BSTS) Method for Predicting Traditional Market Revenue Achievement in Surabaya," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 321–330, Apr. 2025, doi: 10.35882/ijecemi.v7i2.82.
- [21] M. Khan and U. Khan, "Comparison of Forecasting Performance with VAR vs. ARIMA Models Using Economic Variables of Bangladesh," *Asian Journal of Probability and Statistics*, pp. 33–47, Dec. 2020, doi: 10.9734/ajpas/2020/v10i230243.