

Penerapan Data Mining Untuk Prediksi Jumlah Angka Pengangguran Menggunakan Metode Autoregressive Integrated Moving Average (Arima) (Studi Kasus : Bps Sumut)

Rolandho Marchlino Simanihuruk¹, Yolanda Y.P Rumapea², Arina Prima Silalahi³

¹Fakultas Ilmu Komputer, Universitas Methodist Indonesia

Info Artikel

Histori Artikel:

Received, 25 July 2024
Revised, 11 Agustus 2024
Accepted, 10 September 2024

Keywords:

ARIMA,
Pengangguran,
Deret Waktu,
Prediksi,
Analisis Data.

ABSTRAK

Studi ini mengilustrasikan penerapan metode autoregressive integrated moving average (ARIMA) untuk memproyeksikan tingkat pengangguran di Provinsi Sumatera Utara dalam periode tertentu. Ketidakpastian ekonomi yang meningkat, terutama akibat perubahan global, menegaskan urgensi dalam memprediksi dan mengelola tingkat pengangguran dengan akurasi tinggi. Metode ARIMA dipilih karena kemampuannya dalam mengidentifikasi pola dan tren dalam data deret waktu, seperti tingkat pengangguran yang dilaporkan oleh BPS Sumut, dan kemampuannya untuk menerapkan langkah-langkah ARIMA yang sesuai untuk membangun model prediktif. Hasil analisis menunjukkan bahwa prediksi jumlah pengangguran untuk siswa SMA dan SMK usia 20 hingga 24 tahun memiliki tingkat error yang sangat kecil, menunjukkan tingkat akurasi yang tinggi. Penelitian ini memberikan pemahaman mendalam tentang bagaimana metodologi ARIMA dapat digunakan secara efektif untuk meramalkan tingkat pengangguran dan mendukung kebijakan ekonomi serta perencanaan tenaga kerja.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Penulis Koresponden:

Rolandho Marchlino Simanihuruk,
Fakultas Ilmu Komputer,
Universitas Methodist Indonesia, Medan,
Jl. Hang Tuah No.8, Medan - Sumatera Utara.
Email: banditkers77@gmail.com

1. PENDAHULUAN

Penelitian ini menjelaskan penggunaan metode autoregressive integrated moving average (ARIMA) untuk meramalkan tingkat pengangguran di Provinsi Sumatera Utara selama periode tertentu. Perkembangan ketidakpastian dalam perekonomian, terutama karena perubahan global, menekankan pentingnya meramalkan dan mengelola tingkat pengangguran dengan akurat. Metode ARIMA dipilih karena kemampuannya dalam menangkap pola dan tren dalam data time series seperti tingkat pengangguran yang dilaporkan oleh BPS Sumut, serta kemampuannya untuk mengimplementasikan langkah-langkah ARIMA yang sesuai untuk membangun model prediktif. Hasil analisis menunjukkan bahwa prediksi jumlah pengangguran di antara siswa SMA dan SMK usia 20 hingga 24 tahun memiliki tingkat kesalahan yang sangat rendah, menunjukkan bahwa hasil prediksi dapat dianggap akurat. Penelitian ini memberikan pemahaman mendalam tentang bagaimana metodologi ARIMA dapat digunakan secara efektif untuk meramalkan tingkat pengangguran, serta mendukung kebijakan ekonomi dan perencanaan tenaga kerja.

Perkiraan tingkat pengangguran memberikan manfaat signifikan seperti perencanaan ekonomi, pengambilan keputusan bisnis, perencanaan karir, stabilitas sosial, dan evaluasi politik. Sumatera Utara, yang merupakan provinsi dengan populasi keempat terbesar di Indonesia, juga menghadapi tantangan yang tidak bisa dianggap remeh terkait masalah pengangguran.[1] Pengangguran adalah kondisi di mana individu yang merupakan bagian dari angkatan kerja tidak memiliki pekerjaan saat ini dan sedang aktif mencari pekerjaan. Pengangguran juga mencakup individu yang telah memulai usaha untuk mencari pekerjaan, tidak dapat memperoleh pekerjaan meskipun sudah berusaha, atau individu yang telah memiliki pekerjaan tetapi belum memulainya.[2] Tentu saja, tingkat pengangguran dapat diprediksi menggunakan metode statistik. Tujuan dari prediksi ini adalah untuk memberikan Pemerintah Kota Medan acuan dalam mengurangi tingkat pengangguran di Sumatera Utara.[3]

Terdapat berbagai macam teknik peramalan, di antaranya adalah Autoregressive Integrated Moving Average (ARIMA). Dalam penelitian ini, penulis menggunakan teknik autoregressive integrated moving average (ARIMA). Keunggulan dari metode ini adalah kemampuannya untuk menangani berbagai jenis data, walaupun pada awalnya prosesnya harus disesuaikan dengan kondisi data yang diamati.[4] Metode ini lebih efektif ketika digunakan untuk meramalkan periode yang pendek. Teknik tersebut juga menunjukkan tingkat presisi yang tinggi dengan nilai-nilai kesalahan absolut rata-rata (MAE), kesalahan kuadrat rata-rata (MSE), akar kuadrat dari kesalahan rata-rata (RMSE), dan kesalahan persentase absolut rata-rata (MAPE) yang rendah.[5] Metode ini diterapkan oleh Fety Fejriani, M. Hendrawansyah, Leni Muharni, Siti Fara Handdayani, dan Syaharddin dalam sebuah jurnal yang berjudul "Peramalan Pertumbuhan Penduduk Berbasis Gender Menggunakan Metode Arima".[6] Penelitian ini bertujuan untuk mengkaji dan meramalkan jumlah penduduk berdasarkan jenis kelamin menggunakan metode autoregressive integrated moving average (ARIMA). Data yang digunakan berasal dari studi kasus di Nusa Tenggara Barat (NTB) selama 11 tahun terakhir.[7] Penerapan metode autoregressive integrated moving average (ARIMA) telah terbukti sangat efektif dan akurat. Prediksi yang dihasilkan adalah sebesar 2.437.112, dengan nilai mean squared error (MSE) dan mean absolute percentage error (MAPE) sebesar $7.6281e-05$. Hasil prediksi ini diharapkan dapat memberikan panduan kepada pemerintah untuk membuat keputusan yang tepat terkait dengan pembangunan di berbagai sektor seperti ekonomi, pendidikan, kesehatan, dan lain-lain, serta untuk mengurangi tingkat pengangguran dan meningkatkan kesempatan kerja sesuai dengan jumlah populasi yang ada.[8]

Berdasarkan masalah yang dijelaskan di atas maka di rancang sebuah penelitian yang berjudul "Penerapan Data Mining Untuk Prediksi Jumlah Angka Pengangguran Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA)". Di harapkan penelitian ini bermanfaat sebagai acuan untuk Pemerintahan Kota Medan guna menjadi sebuah pedoman dalam proses pembangunan yang dilakukan oleh pemerintah untuk mengatasi masalah pengangguran.

2. METODE PENELITIAN

2.1 Prediksi

Peramalan adalah proses sistematis untuk memproyeksikan peristiwa di masa mendatang berdasarkan informasi dari masa lalu dan saat ini, dengan tujuan meminimalkan kesalahan antara hasil prediksi dan realitas. Prediksi bertujuan untuk memperkirakan kemungkinan hasil di masa depan, namun tidak menjamin kepastian mengenai apa yang akan terjadi.[9]

Meskipun teori peramalan kebijakan jarang dibahas secara eksplisit, relevansi peramalan kebijakan terkait secara tidak langsung dengan proses analisis kebijakan. Analisis kebijakan membutuhkan prediksi mengenai kebijakan di masa depan untuk menghasilkan rekomendasi kebijakan baru. Beberapa ahli, seperti Dunn, menyajikan argumen yang mendukung alat peramalan kebijakan [10]

Ada tiga langkah utama dalam merancang prediksi, yaitu:

1. Langkah awal melibatkan analisis data masa lalu dengan tujuan untuk mengidentifikasi pola yang ada dalam data tersebut.

2. Langkah kedua melibatkan pemilihan metode yang sesuai untuk aplikasi prediksi tertentu, dimana berbagai metode tersedia sesuai dengan tujuan penggunaannya. Berbagai metode tersebut dapat menghasilkan sistem prediksi yang berbeda untuk dataset yang sama. Keberhasilan metode ini dinilai dari kecilnya deviasi yang dihasilkan.
3. Silakan lakukan proses transformasi data historis dengan metode yang Anda pilih. Penyesuaian akan kami lakukan jika diperlukan.

2.2 Data Mining

Data mining adalah proses mengumpulkan, menggali, dan mengidentifikasi data penting dalam sebuah basis data untuk mendapatkan informasi yang bermanfaat yang mendukung pengambilan keputusan yang lebih baik.[11] Data mining adalah langkah-langkah untuk mengumpulkan, menggali, dan mengidentifikasi data dan informasi penting dari basis data untuk mendapatkan wawasan yang berharga yang mendukung pengambilan keputusan yang lebih efektif.[12]

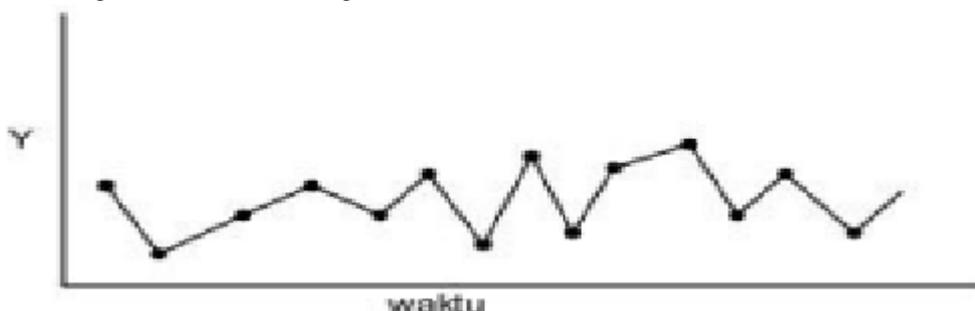
Data mining dibagi menjadi beberapa kelompok [12], yaitu sebagai berikut :

1. Deskripsi
Penjelasan merupakan metode untuk mengungkapkan pola dan tren yang tidak terlihat secara langsung dalam data Anda.
2. Estimasi
Estimasi memiliki kesamaan dengan klasifikasi, namun dalam estimasi, variabel target yang diprediksi adalah numerik, bukan kategorikal. Model dibuat dengan menggunakan dataset lengkap di mana nilai variabel target digunakan sebagai prediksi nilai.
3. Prediksi
Prediksi adalah proses mengidentifikasi pola informasi dengan menggunakan satu atau lebih variabel untuk memproyeksikan nilai dan jenis variabel yang tidak diketahui. Selain memperkirakan nilai-nilai yang tidak diketahui, prediksi juga melibatkan perkiraan nilai-nilai di masa yang akan datang.
4. Klasifikasi
Klasifikasi merupakan proses untuk menemukan fungsi atau model yang mengidentifikasi dan memetakan konsep atau kelas berdasarkan data yang ada.
5. Asosiasi
Tugas asosiasi melibatkan identifikasi atribut-atribut yang sering muncul bersama-sama dalam suatu dataset. Asosiasi adalah proses untuk menemukan pola hubungan di antara item-item dalam kumpulan data.

2.3 Time Series

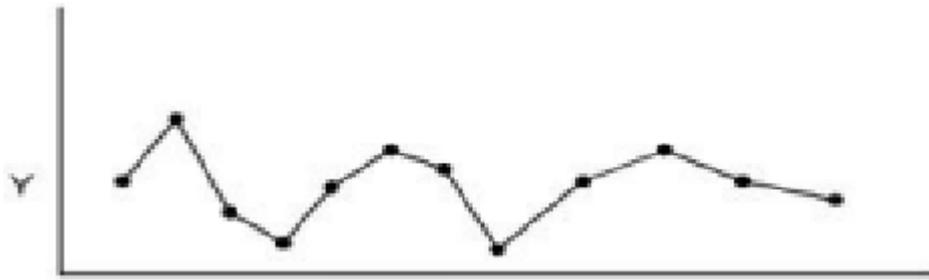
Deret waktu, yang juga sering disebut sebagai time series, merupakan jenis data yang direkam secara berurutan sesuai dengan urutan waktu tertentu[13][14]. Waktu yang digunakan dapat bervariasi, seperti hari, minggu, bulan, tahun, dan sebagainya. Analisis data deret waktu adalah kegiatan analitik yang digunakan untuk mengamati perubahan data selama periode waktu tertentu. Ada empat jenis pola data yang umum dalam deret waktu, yaitu:

1. Secara horizontal, jenis data ini ditandai oleh fluktuasi di sekitar level atau rata-rata yang konstan. Tipe data horizontal dapat dilihat dalam Gambar 1



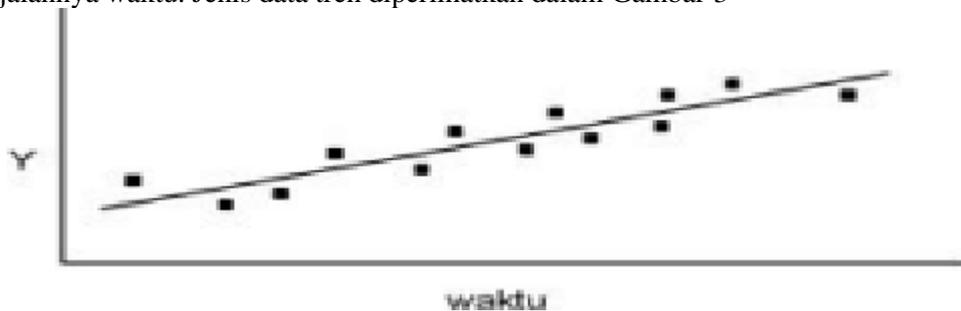
Gambar 1 Pola Data *Horizontal*

2. Musiman adalah jenis data yang dipengaruhi oleh peristiwa tertentu yang memiliki pola perubahan yang berulang secara teratur dari tahun ke tahun. Tipe data musiman digambarkan dalam Gambar 2



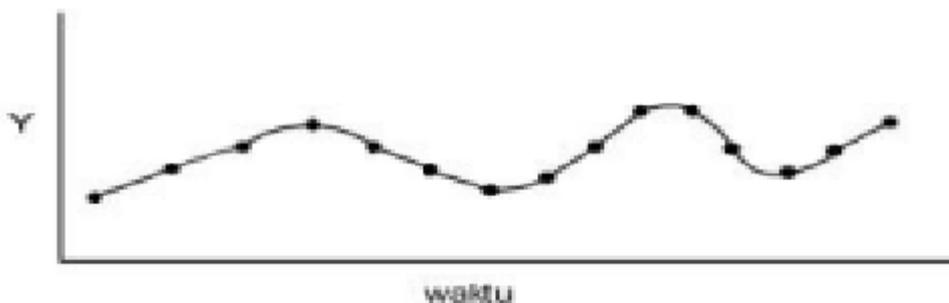
Gambar 2 Pola Data Musiman

3. Tren adalah ketika nilai yang diamati menunjukkan peningkatan atau penurunan seiring berjalannya waktu. Jenis data tren diperlihatkan dalam Gambar 3



Gambar 3 Pola Data Tren

4. Periodik adalah jenis data yang menunjukkan fluktuasi seperti gelombang di sekitar suatu tren. Tipe data periodik diperlihatkan pada gambar 4



Gambar 4 Pola Data Periodik

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Model Peramalan

Sebelum menggunakan model prediksi dengan metode Autoregressive Integrated Moving Average (ARIMA), terdapat beberapa langkah yang perlu dilakukan untuk membangunnya.

3.1.1 Implementasi Syntax

Dalam tahap ini, penulis menerapkan fase prediksi menggunakan bahasa pemrograman Python. Langkah awal melibatkan penginisialisasian pustaka yang diperlukan untuk membuat prediksi dan memvisualisasikan data. Tahap penginisialisasian ini ditunjukkan dalam Gambar 5:

```
from IPython.core.debugger import set_trace
import panel as pn
import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import time
import hvplot.pandas
# import folium

import panel as pn
pn.extension('tabulator')

import hvplot.pandas
```

Gambar 5 Implementasi sintaks

4. KESIMPULAN

Berdasarkan hasil penelitian dan analisis yang dilakukan pada kasus diatas, maka ditarik kesimpulan berikut:

Ketika kami memproyeksikan tingkat pengangguran lulusan SMA dan SMK, kami menemukan hasil sebagai berikut. Untuk lulusan SMA, jumlahnya diperkirakan akan mencapai 64.566 pada tahun 2023, meningkat menjadi 62.219 pada tahun 2024, dan kemudian 60.411 pada tahun 2025. Sementara itu, untuk lulusan SMK, angkanya diproyeksikan menjadi 32.397 pada tahun 2023, meningkat menjadi 34.714 pada tahun 2024, dan kemudian mencapai 37.387 pada tahun 2025. Model terbaik yang ditemukan untuk memprediksi tingkat pengangguran di Sumatera Utara dalam periode 2023-2027 adalah ARIMA (1,0,1) untuk SMA dan (1,0,3) untuk pendidikan vokasi, dengan evaluasi kesalahan RMSE, MSE, dan MAPE dilakukan pada tingkat sekolah. Untuk SMA, nilai MSE adalah 0,8189624, RMSE adalah 0,90496544, dan MAPE adalah 8,15%. Sementara itu, untuk SMK, nilai MSE adalah 0,2661116, RMSE adalah 0,5158601, dan MAPE adalah 4,4%. Hasil dari visualisasi dashboard mempresentasikan data aktual yang meliputi berbagai informasi, termasuk jumlah pengangguran di Sumatera Utara sebesar 4.444 orang, angka pengangguran di seluruh Indonesia, serta proyeksi pengangguran untuk lulusan SMA dan SMK di Sumatera Utara. Desain dashboard ini dirancang untuk memberikan berbagai visualisasi yang membantu analisis data lebih lanjut.

REFERENSI

- [1] A. F. Kurniawan, S. F. Pane, and R. M. Awangga, "Prediksi Jumlah Penjualan Rumah di Bojongsoang ditengah Pandemi Covid-19 dengan Metode ARIMA," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1479, 2021, doi: 10.30865/mib.v5i4.3121.
- [2] F. D. Yuliantanti, D. C. R. Novitasari, N. Widodo, A. Hamid, and W. D. Utami, "Penerapan Metode Autoregressive Integrated Moving Average (Arima) Untuk Prediksi Bilangan Sunspot," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 15, no. 3, pp. 555–564, 2021, doi: 10.30598/barekengvol15iss3pp555-564.
- [3] I. Murapi, D. A. O. Astarini, and I. N. Subudiartha, "Tingkat Pengangguran Akibat Covid-19 di Provinsi Nusa Tenggara Barat," *Riset, Ekon. Akunt. dan Perpajak.*, vol. 2, no. 1, pp. 21–30, 2021, doi: 10.30812/rekan.v2i1.1116.
- [4] Wulandari R.A and Gernowo R, "Metode Autoregressive Integrated Moving Average (ARIMA) dan Metode Adaptive Neuro Fuzzy Inference System (ANFIS) dalam Analisis

- Curah Hujan,” *Berk. Fis.*, vol. 22, no. 1, pp. 41–48, 2019.
- [5] L. Ainiyah and M. Bansori, “Prediksi Jumlah Kasus COVID-19 menggunakan Metode Autoregressive Integrated Moving Average (ARIMA) (Studi Kasus Kabupaten Sidoarjo),” *J. Sains Dasar*, vol. 10, no. 2, pp. 62–68, 2021, doi: 10.21831/jsd.v10i2.43606.
- [6] J. Purnama and A. Juliana, “Analisa Prediksi Indeks Harga Saham Gabungan Menggunakan Metode Arima,” *Cakrawala Manag. Bus. J.*, vol. 2, no. 2, p. 454, 2020, doi: 10.30862/cm-bj.v2i2.51.
- [7] W. Y. Rusyida and V. Y. Pratama, “Prediksi Harga Saham Garuda Indonesia di Tengah Pandemi Covid-19 Menggunakan Metode ARIMA,” *Sq. J. Math. Math. Educ.*, vol. 2, no. 1, p. 73, 2020, doi: 10.21580/square.2020.2.1.5626.
- [8] S. Suseno and Suryo Wibowo, “Penerapan Metode ARIMA dan SARIMA Pada Peramalan Penjualan Telur Ayam Pada PT Agromix Lestari Group,” *J. Teknol. dan Manaj. Ind. Terap.*, vol. 2, no. I, pp. 33–40, 2023, doi: 10.55826/tmit.v2ii.85.
- [9] W. Latuny and W. M. S. Picauly, “ANALISIS BULLWHIP EFFECT DENGAN MENGGUNAKAN METODE PERAMALAN PADA SUPPLY CHAIN DI DISTRIBUTOR PT. SEMEN TONASA (Studi Kasus: Distributor PT. Semen Tonasa),” *Arika*, vol. 13, no. 2, pp. 113–126, 2019, doi: 10.30598/arika.2019.13.2.113.
- [10] M. Kafil, “Penerapan Metode K-Nearest Neighbors Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 3, no. 2, pp. 59–66, 2019, doi: 10.36040/jati.v3i2.860.
- [11] W. S. Rahayu, P. T. Juwono, and W. Soetopo, “Analisis Prediksi Debit Sungai Amprong Dengan Model Arima (Autoregressive Integrated Moving Average) Sebagai Dasar Penyusunan Pola Tata Tanam,” *J. Tek. Pengair.*, vol. 10, no. 2, pp. 110–119, 2019, doi: 10.21776/ub.pengairan.2019.010.02.04.
- [12] R. Mahmud and A. Hartanto, “Penerapan Data Mining Rekomendasi Laptop Menggunakan Algoritma Apriori,” *Juisi*, vol. 06, no. 02, pp. 21–30, 2020.
- [13] H. G. Simanullang, A. P. Silalahi, and D. Sartika, “PREDIKSI JUMLAH PASIEN COVID-19 DI INDONESIA MENGGUNAKAN LEAST SQUARE METHOD BERBASIS ANDROID,” *INFORMATIKA*, vol. 14, no. 1, pp. 86–93, 2022.
- [14] S. R. A. Arifai and Lukman Junaedi, “Prediksi Permintaan Barang Berdasarkan Penjualan Menggunakan Metode Arima Box-Jenkins (Studi Kasus : Pt. Beststamp Indonesia),” *J. E-Bis*, vol. 4, no. 2, pp. 138–146, 2020, doi: 10.37339/e-bis.v4i2.227.