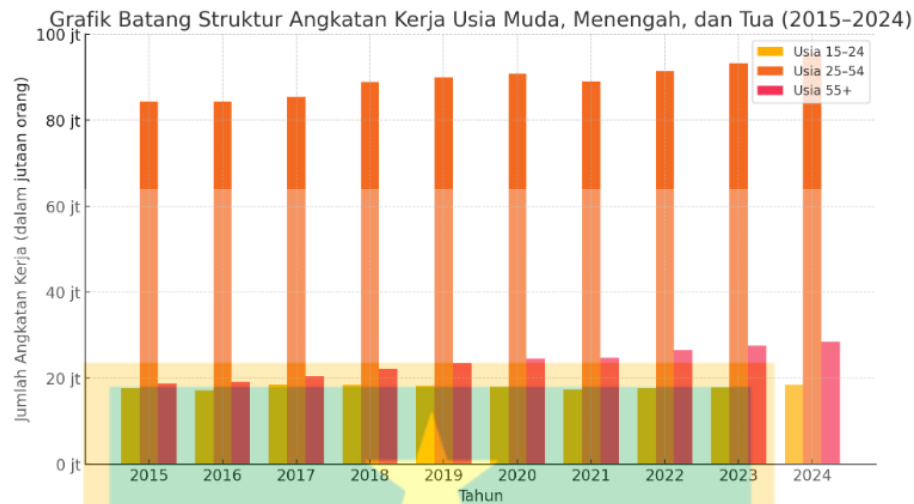


BAB 1 PENDAHULUAN

1.1 Latar Belakang

Perubahan struktur demografi merupakan fenomena penting yang tengah dihadapi banyak negara, terutama negara berkembang. Salah satu periode krusial dalam dinamika demografi adalah bonus demografi, yaitu ketika proporsi penduduk usia produktif (15–64 tahun) berada pada titik tertinggi dibandingkan kelompok usia non-produktif. Secara teoritis, kondisi ini memberikan peluang besar bagi pertumbuhan ekonomi melalui peningkatan produktivitas, tabungan nasional, dan investasi sumber daya manusia (Jiajun et al., 2025). Pengalaman negara-negara seperti Korea Selatan dan China menunjukkan bahwa efektivitas bonus demografi sangat dipengaruhi oleh kesiapan kebijakan pendidikan, ketenagakerjaan, dan sosial ekonomi yang kohesif (Haider & Mahmood, 2023).

Menurut publikasi resmi (Badan Pusat Statistik, 2025), Indonesia saat ini berada pada era intensifikasi bonus demografi, yang dapat dilihat pada struktur penduduk usia produktif hingga tahun 2024. Era ini menandai dimulainya fase bonus demografi nasional, yang diperkirakan akan berlangsung hingga tahun 2030-an. Oleh karena itu, studi data dari tahun 2015 hingga 2024 memiliki signifikansi strategis sebagai landasan dasar pengembangan kebijakan jangka menengah. Pada tahun 2025, Badan Pusat Statistik melaporkan terdapat 216,79 juta orang usia kerja, dengan tingkat partisipasi angkatan kerja (TPAK) sebesar 70,60% dan tingkat pengangguran terbuka (TPT) sebesar 4,76%. Namun, ada masalah struktural yang mendasari angka-angka ini: sekitar 52,72% angkatan kerja hanya menyelesaikan sekolah sampai sekolah dasar atau sekolah menengah pertama, dan lebih dari sepertiganya bekerja paruh waktu. Hal ini menunjukkan bahwa tenaga kerja tidak terlalu terampil atau mampu menangani masalah yang muncul akibat Revolusi Industri dan transisi digital.



Gambar 1.1 Grafik Angkatan Kelompok Umur

Rentang usia 15 hingga 24 tahun sangat penting secara strategis dan berisiko. Mereka memimpin dalam menggunakan bonus demografis, tetapi mereka juga menghadapi bahaya kehilangan pekerjaan, bekerja secara informal, dan memiliki keterampilan yang tidak setara. Penelitian sebelumnya menunjukkan bahwa remaja di negara berkembang sering mengalami ketidaksejajaran keterampilan, pengalaman kerja yang tidak memadai, dan akses terbatas ke pelatihan kejuruan (Sharma, 2022; Tikhonov et al., 2024). Hal ini mendorong penyelidikan yang signifikan: sejauh mana perencanaan pasar tenaga kerja disesuaikan dengan fluktuasi jangka pendek dan tren musiman yang memengaruhi demografi yang lebih muda dalam pendekatan berbasis data yang lebih akurat dan terperinci?

Untuk mendapatkan hasil maksimal dari dividen demografis, Anda memerlukan banyak data berkualitas tinggi setiap saat untuk memodelkan lapangan kerja muda. Sumber data utama penelitian ini adalah data resmi dari Badan Pusat Statistik (BPS) RI sepanjang tahun 2015 hingga 2024. Survei Angkatan Kerja Nasional (SAKERNAS) merupakan cara yang tepat secara metodologis bagi BPS, lembaga statistik nasional, untuk mengumpulkan data ketenagakerjaan. Ini juga digunakan sebagai pedoman kebijakan pemerintah. Penelitian ini menghadapi tantangan yang signifikan dengan aksesibilitas data ketenagakerjaan remaja yang disebarluaskan oleh BPS.

Kendala utama mencakup frekuensi temporal yang berkurang. Sejak tahun 2015, data ketenagakerjaan pemuda baru diberikan dua kali dalam setahun (setiap enam bulan), artinya baru ada 20 kali pengamatan dari tahun 2015 hingga 2024. Frekuensi pengungkapan ini jauh lebih rendah daripada data ketenagakerjaan agregat nasional, yang sebagian dapat diakses setiap tiga bulan. Karena batasan frekuensi ini, terdapat kesenjangan informasi selama 6 bulan antara dua lokasi pengamatan. Artinya, setiap perubahan jangka pendek atau tren musiman yang terjadi pada waktu tersebut tidak dikumpulkan.

Kedua, pembatasan pemisahan kelompok umur. Pemisahan kelompok usia muda (15-24 tahun) tidak memberikan akses ke semua variabel ketenagakerjaan. Beberapa variabel, seperti gaji riil, jam kerja, dan status pekerjaan, hanya dapat ditemukan dalam agregat nasional. Hal ini membuat sulit untuk melihat ciri-ciri individu dari angkatan kerja muda. Hal ini memaksa para peneliti untuk menggunakan proksi atau asumsi tertentu dalam upaya pemodelan mereka.

Ketiga, cara survei dilakukan tidak konsisten. Teknik SAKERNAS berubah pada tahun 2021 dari dilakukan setiap enam bulan menjadi setiap tahun (namun tetap menerbitkan dua periode). Hal ini membuat lebih sulit untuk membandingkan data dari waktu ke waktu dan membuat kualitas data menjadi kurang konsisten. Modifikasi ini mencakup perubahan pada kerangka sampel, definisi operasional variabel tertentu, dan metode estimasi. Perubahan ini mungkin menghasilkan kerusakan struktural dalam deret waktu.

Keempat, datanya tidak cukup baik untuk pemodelan deret waktu. Literatur peramalan menunjukkan bahwa setidaknya 30 hingga 50 titik data diperlukan untuk estimasi parameter yang kuat saat menggunakan model deret waktu untuk analisis prediktif. Namun, ketersediaan 20 observasi setiap semester berada di bawah level tersebut. Hal ini mempersulit pencarian pola musiman, model pemeriksaan, dan data terpisah untuk pelatihan dan pengujian yang baik.

Batasan-batasan ini bukan hanya masalah statistik; mereka menunjukkan bahwa ada masalah dengan struktur Sistem Informasi Ketenagakerjaan Nasional. Di

satu sisi, survei rumah tangga besar seperti SAKERNAS membutuhkan banyak biaya untuk dijalankan, sehingga sulit untuk sering melakukannya. Di sisi lain, perencanaan kebijakan ketenagakerjaan yang responsif, khususnya dalam konteks insentif demografis yang sensitif terhadap waktu, memerlukan akses ke data dengan resolusi temporal yang ditingkatkan untuk secara akurat mencerminkan dinamika pasar tenaga kerja yang berkembang pesat.

Penelitian ini menggunakan teknik interpolasi deret waktu yang didasarkan pada kerangka Bayesian, seperti yang ditetapkan oleh Rojo & Sanz (2017, 2020), untuk menjembatani kesenjangan antara ketersediaan data dan persyaratan analitis. Metode ini menggunakan pemodelan probabilistik untuk mengubah data setengah tahunan menjadi data triwulanan. Ini menjaga sifat statistik data asli, seperti tren, volatilitas, dan struktur autokorelasi, sambil mengisi celah waktu dengan perkiraan yang konsisten secara statistik.

Prosedur interpolasi melibatkan penambahan satu titik pengamatan di antara setiap pasangan pengukuran setengah tahunan. Ini memberi kita sekitar 40 pengamatan triwulanan untuk tahun 2015 hingga 2024. Metode kehati-hatian ini diadopsi untuk menjaga keseimbangan antara data waktu yang lebih detail dan data yang lebih andal. Praktik penerbitan BPS untuk beberapa indikator makroekonomi lainnya juga menggunakan frekuensi interpolasi hingga triwulanan, sehingga temuan interpolasi tetap sesuai dengan konteks yang biasa digunakan para pembuat kebijakan.

Interpolasi Bayesian lebih baik daripada metode deterministik seperti interpolasi linier atau spline kubik karena dapat memodelkan ketidakpastian dalam estimasi titik interpolasi, menggunakan informasi dari deret waktu keseluruhan, bukan hanya dua titik yang bertetangga melalui distribusi sebelumnya, dan membuat interval kredibilitas yang menunjukkan seberapa yakin Anda berada di nilai interpolasi. Hal ini penting karena data yang diinterpolasi akan digunakan untuk melatih model prediksi yang perlu mengetahui seberapa tidak pastinya suatu hal.

Namun demikian, penelitian ini mengakui bahwa data interpolasi masih merupakan perkiraan yang membawa lebih banyak ketidakpastian daripada data observasi langsung. Untuk mengurangi risiko ini, analisis sensitivitas dilakukan dengan membandingkan prediksi model berdasarkan data setengah tahunan asli tanpa interpolasi dengan yang didasarkan pada data triwulanan dari temuan interpolasi. Perbandingan ini memungkinkan kita melihat seberapa baik model bekerja dengan interpolasi dan seberapa besar bias yang dapat ditimbulkannya.

Untuk mempersiapkan pekerjaan secara efektif dan responsif, sistem prediktif harus mampu menangkap kompleksitas dan dinamisme data ketenagakerjaan dengan cara yang lebih detail. Menggunakan data triwulanan yang diinterpolasi dari tahun 2015 hingga 2024 jauh lebih baik daripada menggunakan data semester asli. Misalnya, ini mungkin menunjukkan musim dalam periode yang lebih singkat, pola jangka menengah yang lebih mulus, dan respons yang lebih cepat terhadap perubahan kebijakan atau ekonomi.

Data triwulanan memiliki sejumlah manfaat praktis. Misalnya, banyak indikator ekonomi makro (seperti PDB, inflasi, dan investasi) dilaporkan setiap tiga bulan, sehingga lebih mudah untuk menggabungkannya dengan variabel lain. Ini juga menunjukkan pola musiman yang penting bagi dinamika angkatan kerja muda, seperti periode kelulusan sekolah (bahkan semester) dan siklus perekrutan di sektor formal, yang biasanya mengikuti siklus bisnis triwulanan. Akhirnya, ia memiliki pengamatan yang cukup (40 titik data) untuk estimasi model Deret Waktu yang kuat tanpa risiko overfitting yang tinggi.

Tiga teknik pemodelan alternatif dengan beragam peran dan tanggung jawab dipilih berdasarkan tujuan penelitian dan sifat data triwulanan. Deret Waktu Struktural Bayesian (BSTS) dipilih sebagai model utama karena kemampuannya beradaptasi dalam memodelkan komponen tren, musim, dan dampak faktor eksogen secara eksplisit. Tiga metodologi pemodelan yang berbeda dengan peran dan tanggung jawab yang beragam dipilih berdasarkan tujuan penelitian dan sifat data triwulanan. Model utama yang digunakan adalah Bayesian Structural Time Series (BSTS) karena dapat

dengan mudah menjelaskan komponen tren, musim, dan pengaruh faktor luar. Manfaat BST adalah dapat menangani data yang hilang.

Long Short-Term Memory (LSTM) sebagai model perbandingan karena ini adalah metode pembelajaran mendalam yang dapat menemukan pola non-linier dan hubungan jangka panjang dalam data deret waktu. Itu dapat mempelajari pola dari data mentah, yang berarti dapat menemukan tren dalam data mingguan yang terlewatkan oleh model lain.

ARIMA Musiman (SARIMA) adalah model baseline atau model pembanding klasik yang dibuat untuk data yang memiliki karakteristik musiman. Kami memilih SARIMA karena telah menjadi standar emas untuk menganalisis deret waktu ekonomi selama beberapa dekade dan merupakan tolok ukur kinerja yang diterima secara umum dalam literatur prakiraan. Literatur mengatakan bahwa SARIMA dasar membutuhkan setidaknya 30 pengamatan dan model yang lebih canggih membutuhkan setidaknya 50 pengamatan. Dengan 40 observasi triwulanan, SARIMA memiliki data yang cukup untuk estimasi parameter yang andal. Perannya sebagai baseline adalah untuk memberikan persyaratan kinerja mendasar untuk pemodelan data triwulanan musiman dan untuk memberikan kontrol metodologis untuk menilai nilai tambah dari metode yang lebih rumit seperti BST dan LSTM.

Penelitian ini menggunakan Long Short-Term Memory (LSTM) sebagai model benchmark pelengkap, meskipun jumlah observasi triwulanan (40 titik data) jauh lebih rendah dari kisaran biasanya untuk model pelatihan deep learning terbaik (200-1000+ observasi). Pilihan untuk memasukkan LSTM ke dalam kerangka komparatif didasarkan pada banyak faktor metodologis dan kontekstual yang signifikan. Pertama, dari sudut pandang metodologi perbandingan, literatur tentang pembelajaran mesin dan peramalan menggarisbawahi perlunya membandingkan model dari paradigma yang berbeda (statistik vs pembelajaran mesin) untuk memahami pertukaran kinerja dalam berbagai situasi data. Makridakis dkk. (2018) dalam kompetisi M4, kompetisi peramalan terbesar dengan deret waktu 100.000, menekankan bahwa tidak ada satu cara optimal untuk menangani semua jenis data. Sebaliknya, perbandingan lintas

paradigma menunjukkan kapan strategi bekerja paling baik. Dalam hal ini, LSTM adalah metode pembelajaran mesin non-parametrik yang sangat berbeda dengan BST (Bayesian structural) dan SARIMA (classical econometric). Hal ini menjadikan perbandingan tersebut sebagai cara yang berguna untuk mempelajari lebih lanjut tentang pola dalam data ketenagakerjaan Indonesia. Kedua, karya ini secara terbuka mengakui bahwa 40 pengamatan triwulanan mewakili batas bawah, berada di bawah persyaratan data minimal untuk pelatihan LSTM yang optimal. Namun, menempatkan LSTM dalam kondisi data yang kurang optimal menghasilkan signifikansi ilmiahnya sendiri dengan menawarkan bukti empiris mengenai penerapan metodologi pembelajaran mendalam pada peramalan pasar tenaga kerja dalam skenario data terbatas yang lazim dalam statistik resmi negara berkembang.

Beberapa penelitian terbaru menunjukkan bahwa dengan pengurangan arsitektural, regularisasi yang ketat, dan penyesuaian hyperparameter yang cermat, LSTM dapat memberikan kinerja informatif dalam skenario sampel kecil, tetapi tidak optimal. Lanbouri & Achhab (2020) menunjukkan dalam studinya tentang prediksi harga saham dengan data terbatas bahwa LSTM dengan arsitektur sederhana (1-2 lapisan tersembunyi dengan unit terbatas) dapat memberikan hasil yang sebagus atau lebih baik dari ARIMA pada kumpulan data dengan pengamatan 50-100, terutama ketika data memiliki pola non-linier yang kuat. Dengan cara yang sama, Sagheer dan Kotb (2019) mendemonstrasikan bahwa LSTM dengan regularisasi putus sekolah yang ketat dapat bekerja dengan baik pada deret waktu dengan 40 hingga 80 titik data, terutama untuk menangkap hubungan jangka panjang yang tidak dapat ditangkap oleh model linier.

Ketiga, dari perspektif substantif, angkatan kerja muda menunjukkan dinamika non-linier yang rumit yang dibentuk oleh siklus ekonomi, pergeseran kebijakan pendidikan, kemajuan teknologi, dan guncangan eksternal, yang dicontohkan oleh pandemi COVID-19 selama masa studi 2020-2021. Model linier tradisional seperti SARIMA mengalami kesulitan menangkap non-linearitas ini, dan meskipun BST fleksibel, mereka masih membuat asumsi struktural tertentu tentang bentuk tren dan

komponen musiman. LSTM dapat merepresentasikan segala jenis fungsi non-linier dan hubungan jangka panjang tanpa perlu mengetahui terlebih dahulu apa strukturnya. Ini berarti ia mungkin menemukan pola yang terlewatkan oleh dua model lainnya. Fischer & Krauss (2018) menunjukkan bahwa LSTM mahir dalam mengidentifikasi pergeseran rezim dan perubahan struktural dalam deret waktu keuangan, terutama yang berkaitan dengan data ketenagakerjaan Indonesia yang mengalami beberapa jeda struktural, termasuk reformasi kurikulum pendidikan, kenaikan upah minimum regional yang substansial., pandemi COVID-19, dan pemulihan ekonomi selanjutnya.

Keempat, penggunaan LSTM sebagai standar juga berguna dalam rangka pengembangan ilmu data untuk kebijakan publik di Indonesia. Dengan semakin banyaknya komputer, pertumbuhan ekosistem big data, dan upaya pemerintah untuk mendigitalkan, kemungkinan ke depan, data ketenagakerjaan frekuensi tinggi akan dapat diakses melalui sumber alternatif seperti data administrasi (data BPJS Ketenagakerjaan, data pajak penghasilan), catatan ponsel, atau platform posting lowongan kerja (LinkedIn, JobStreet, dll.). Dalam konteks ini, temuan penelitian ini memberikan bukti awal tentang kemanjuran metodologi pembelajaran mendalam untuk peramalan pasar tenaga kerja di Indonesia, dan menjadi landasan untuk penelitian lebih lanjut ketika data yang lebih luas dapat diakses.

Kelima, dari sudut pandang metodologis, perbandingan LSTM dengan BST dan SARIMA dalam pengaturan data terbatas merupakan kontribusi ilmiah unik yang jarang diteliti dalam literatur. Sebagian besar studi tentang LSTM untuk peramalan ekonomi menggunakan kumpulan data yang besar ($n > 200$) atau bahkan kumpulan data yang sangat besar ($n > 1000$), sehingga sedikit yang diketahui tentang kinerja relatifnya pada kumpulan data berukuran kecil (30-50 pengamatan) yang biasa ditemukan dalam statistik resmi negara berkembang dengan sumber daya survei yang terbatas. Studi ini memberikan bukti empiris mengenai keterbatasan dan potensi jebakan penggunaan pembelajaran mendalam untuk peramalan ekonomi dalam skenario dengan data terbatas. Ini mendokumentasikan kinerja LSTM dalam kondisi data yang kurang optimal dan membandingkannya dengan model yang disesuaikan

untuk situasi sampel kecil (BSTS, SARIMA), menawarkan kontribusi yang signifikan bagi para peneliti dan praktisi di negara-negara yang menghadapi kendala data analog.

Penelitian ini menempatkan LSTM bukan sebagai model utama, tetapi sebagai tolok ukur eksplorasi untuk menyelidiki adanya pola non-linier substansial dalam dinamika angkatan kerja muda yang masih belum terselesaikan oleh BSTS dan SARIMA. Jika LSTM menunjukkan kinerja kompetitif meskipun datanya terbatas, LSTM berfungsi sebagai bukti non-linieritas yang signifikan yang harus dimasukkan ke dalam pemodelan kebijakan. Ini menetapkan dasar untuk kinerja pendekatan pembelajaran mendalam dalam kondisi data terbatas (40 pengamatan), menawarkan referensi untuk penelitian di masa mendatang ketika data tambahan dapat diakses. Hasil ini dapat berfungsi sebagai perkiraan batas bawah untuk kinerja LSTM untuk lapangan kerja Indonesia, meningkatkan wacana metodologis tentang pertukaran antara kompleksitas model, kebutuhan data, dan akurasi prakiraan dalam prediksi pasar tenaga kerja di negara-negara berkembang. Penelitian ini menambah kerangka kerja yang ada tentang bagaimana memilih prosedur yang tepat ketika data terbatas.

Selain itu, penelitian tentang pasar tenaga kerja kaum muda di Indonesia cenderung bersifat deskriptif, dengan fokus pada topik-topik seperti mobilitas tenaga kerja, preferensi sektor, atau pekerjaan orang lanjut usia di bidang pertanian (Ngadi et al., 2023). Meskipun studi ini menghasilkan wawasan berharga tentang kondisi saat ini, studi tersebut biasanya tidak memiliki penggabungan metodologi prediktif kuantitatif berbasis model yang dapat memfasilitasi perencanaan kebijakan yang proaktif.

Literatur yang memprediksi lapangan kerja di Indonesia, bahkan pada tingkat agregat nasional, tetap terkendala secara signifikan dalam penggunaan metodologi kontemporer. Sebagian besar prediksi Tenaga kerja yang digunakan dalam dokumen perencanaan nasional, seperti RPJMN, didasarkan pada ekstrapolasi tren dasar atau metodologi komponen kohort yang tidak mempertimbangkan perubahan ekonomi jangka pendek atau faktor luar yang rumit. Masih relatif tidak biasa menggunakan

teknik peramalan deret waktu tingkat lanjut seperti BST atau metode pembelajaran mesin seperti LSTM untuk memperkirakan pasar tenaga kerja di Indonesia.

Akibatnya, karya tersebut memberikan kontribusi ilmiah yang signifikan dalam empat dimensi. Pertama, penelitian ini mengkaji kemampuan tiga metodologi prediktif dari paradigma yang berbeda: BSTS sebagai model probabilistik struktural, SARIMA sebagai model ekonometrik klasik, dan LSTM sebagai model pembelajaran mendalam, dalam konteks pasar tenaga kerja Indonesia yang bercirikan ketersediaan data yang terbatas. Studi banding ini menjelaskan keunggulan relatif dari masing-masing metodologi dalam konteks yang mencerminkan ketersediaan data aktual di negara berkembang.

Kedua, karya ini menawarkan informasi empiris tentang kemandirian dan kendala penggunaan pembelajaran mesin, khususnya pembelajaran mendalam, untuk perkiraan tenaga kerja dalam keadaan data yang buruk. Temuan ini meningkatkan literatur yang ada tentang penerapan metodologi intensif data dalam pengaturan yang ditandai dengan kelangkaan data, tidak hanya terkait dengan Indonesia tetapi juga beberapa negara berkembang lainnya.

Ketiga, karya ini membuat penambahan metodologis dengan menggunakan interpolasi Bayesian untuk meningkatkan granularitas temporal data. Ini juga menunjukkan bagaimana mengatasi masalah yang muncul dengan data frekuensi rendah sambil menjaga integritas statistik. Anda dapat menggunakan metode ini dalam situasi lain ketika Anda membutuhkan data frekuensi tinggi untuk analisis yang lebih rinci tetapi tidak memilikinya.

Kebijakan ketenagakerjaan berbasis bukti untuk memanfaatkan bonus demografis secara maksimal. Hasil yang diprediksi dapat digunakan untuk sistem peringatan dini untuk memperkirakan surplus atau kekurangan tenaga kerja muda, untuk merencanakan program pelatihan dan pendidikan yang didasarkan pada perkiraan permintaan, dan untuk merancang intervensi yang ditargetkan bagi kaum muda yang akan mendapat manfaat paling banyak dari bonus demografis.

Penelitian ini mengkaji pemeriksaan dan perbandingan keakuratan model Bayesian Structural Time Series (BSTS), Seasonal ARIMA (SARIMA), dan Long

Short-Term Memory (LSTM) dalam peramalan angkatan kerja berusia 15-24 tahun di Indonesia dari tahun 2015 hingga 2024.

Fokus penelitian ini dilakukan oleh dua kelompok empiris: analisis ketersediaan data deret waktu semester dari BPS yang hanya menghasilkan 20 observasi selama masa studi, dan perlunya peningkatan granularitas temporal untuk menangkap pola musiman dan dinamika jangka waktu., yang dianggap relevan untuk perencanaan penelitian. Untuk mengatasi kesenjangan ini, data sementara diubah menjadi data triwulanan (sekitar 40 pengamatan) menggunakan pendekatan interpolasi Bayesian yang hati-hati. Ini meningkatkan resolusi temporal dan mengurangi kemungkinan distorsi data.

Kami memilih BST sebagai model utama karena dapat memperoleh data deret waktu dengan variabel luar, memahami nilai dan ketidakpastian yang hilang, dan memberikan penjelasan struktural yang baik tentang bagian tren, tren, dan perubahan luar. Model ini cukup fleksibel, yang membuatnya bagus untuk menganalisis sejumlah kecil data, tetapi juga memberi Anda banyak informasi.

SARIMA adalah garis dasar klasik yang telah diuji dengan baik dan terbukti berhasil untuk prakiraan ekonomi. SARIMA adalah tolok ukur kinerja yang digunakan untuk membandingkan model lain. Ini memiliki sejarah panjang dalam menggunakan ekonomi dan kapasitas yang ditunjukkan untuk bekerja dengan data yang memiliki pola musiman.

LSTM berfungsi sebagai patokan tambahan untuk menguji potensi pola nonlinier dalam dinamika angkatan kerja muda dan untuk menjelaskan kemandirian pembelajaran mendalam dalam kondisi data tertentu di wilayah yang belum dipelajari secara ekstensif terkait prakiraan pasar tenaga kerja di negara berkembang. Memosisikan LSTM sebagai tolok ukur eksplorasi dengan data sampel tetap meningkatkan metodologi untuk penerapan pembelajaran mendalam dalam konteks sampel terbatas.

Variabel eksogen yang digunakan dalam manajemen meliputi: tingkat partisipasi angkatan kerja (TPAK) dari kelompok bisnis yang sedang berkembang, yang menunjukkan dinamika sisi penawaran; tingkat kepastian pembelian (TPT) di antara

perusahaan yang baru lahir, yang mencerminkan ketatnya pasar tenaga kerja; durasi rata-rata pendidikan (RLS) untuk kaum muda, melambangkan akumulasi modal manusia; Produk Domestik Bruto (PDB) per kapita, memperhitungkan kondisi makroekonomi dan permintaan tenaga kerja; dan populasi demografis kelompok usia yang lebih muda, dengan mempertimbangkan potensi dinamika penawaran.

Penelitian ini diantisipasi untuk memberikan tiga hasil utama, termasuk generasi kinerja model yang kuat. Indonesia membutuhkan kebijakan yang didasarkan pada bukti untuk memanfaatkan keuntungan demografisnya secara maksimal. Ini melibatkan mencari tahu kapan waktu terbaik untuk intervensi tertentu dan seberapa besar kesenjangan antara jumlah anak muda yang menginginkan pekerjaan dan jumlah pekerjaan yang tersedia. Ini juga mencakup pembelajaran bagaimana menggunakan metode peramalan yang berbeda (Bayesian, ekonometrik klasik, pembelajaran mesin) dalam situasi di mana perbandingan data umum terjadi di negara berkembang. Ini juga termasuk mempelajari cara menggabungkan metode ini dengan sumber data lain untuk membantu pembuat kebijakan membuat keputusan berdasarkan evidence.in era komputer.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa model Bayesian Structural Time Series (BSTS) dalam memprediksi jumlah angkatan kerja usia 15-24 tahun di Indonesia menggunakan data deret waktu triwulanan hasil interpolasi periode 2015-2024?
2. Seberapa akurat model *Seasonal* ARIMA (SARIMA) sebagai *baseline* model tradisional dalam memprediksi jumlah angkatan kerja muda dengan data triwulanan yang memiliki pola musiman?
3. Bagaimana performa model Long Short-Term Memory (LSTM) sebagai representasi pendekatan deep learning dalam memprediksi angkatan kerja usia muda pada kondisi data terbatas (40 observasi triwulanan)?

4. Model manakah di antara BSTS, SARIMA, dan LSTM yang memberikan akurasi terbaik berdasarkan metrik evaluasi (MAPE, RMSE, dan R^2) dalam memprediksi angkatan kerja usia muda di Indonesia?
5. Apakah terdapat pola non-linear yang signifikan dalam dinamika angkatan kerja usia muda Indonesia yang dapat ditangkap oleh LSTM namun terlewatkan oleh model statistik tradisional (BSTS dan SARIMA)?
6. Bagaimana implikasi hasil perbandingan ketiga model tersebut terhadap pemilihan metode forecasting yang tepat dalam kondisi data terbatas, dan perumusan kebijakan ketenagakerjaan berbasis bukti (*evidence-based policy*) untuk optimalisasi bonus demografi Indonesia?

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Menganalisis dan mengevaluasi performa model *Bayesian Structural Time Series* (BSTS) sebagai model utama dalam memprediksi jumlah angkatan kerja usia 15–24 tahun menggunakan data triwulanan hasil interpolasi periode 2015–2024 dengan mempertimbangkan komponen tren, musiman, dan pengaruh variabel eksogen.
2. Menganalisis performa model *Seasonal ARIMA* (SARIMA) sebagai *baseline* model tradisional dalam memprediksi angkatan kerja usia muda dengan pola musiman triwulanan.
3. Mengeksplorasi performa model *Long Short-Term Memory* (LSTM) sebagai representasi pendekatan *deep learning* dalam kondisi data terbatas (40 observasi triwulanan) untuk mengidentifikasi ada tidaknya pola non-linear yang signifikan dalam dinamika angkatan kerja usia muda.
4. Membandingkan tingkat akurasi ketiga model (BSTS, SARIMA, dan LSTM) berdasarkan metrik evaluasi MAPE, RMSE, dan R^2 untuk mengidentifikasi model yang paling sesuai dalam konteks prediksi ketenagakerjaan Indonesia dengan keterbatasan data.

5. Memberikan kontribusi metodologis tentang *trade-offs* antara kompleksitas model, kebutuhan data, dan akurasi prediksi dalam konteks *labor market forecasting* di negara berkembang dengan data terbatas.
6. Memberikan rekomendasi kebijakan ketenagakerjaan berbasis bukti (*evidence-based policy*) untuk mendukung pemanfaatan bonus demografi di Indonesia secara optimal, termasuk identifikasi periode kritis dan magnitude gap antara supply dan demand angkatan kerja muda.

1.4 Manfaat Penelitian

1. Manfaat Teoritis:

- a. Membantu menciptakan strategi kuantitatif untuk memprediksi angkatan kerja muda menggunakan data triwulanan dari interpolasi Bayesian yang dapat menangkap tren musiman dan perubahan jangka menengah.
- b. Memberikan kontribusi ilmiah pada perbandingan tiga paradigma alternatif untuk memprediksi kaum muda mendapatkan pekerjaan dengan data terbatas: probabilistik struktural (BST), ekonometrika klasik (SARIMA), dan pembelajaran mendalam (LSTM).
- c. Menyajikan bukti empiris tentang kemandirian dan kendala penggunaan pembelajaran mendalam (LSTM) untuk peramalan pasar tenaga kerja dalam kondisi data yang kurang optimal (pengaturan sampel kecil), sebuah topik yang belum diteliti secara ekstensif dalam literatur.
- d. Mengevaluasi konsekuensi dari pola non-linear dan kerusakan struktural dalam dinamika angkatan kerja muda pada pilihan metodologi peramalan yang sesuai.
- e. Ditambahkan ke badan penelitian tentang bagaimana menggunakan pendekatan interpolasi Bayesian untuk mengatasi masalah dengan data frekuensi rendah dengan tetap menjaga integritas statistik, yang dapat digunakan dalam pengaturan penelitian yang berbeda di negara-negara terbelakang.

2. Manfaat Praktis:

- a. Memberikan landasan analisis prediktif yang lebih akurat dan responsif untuk membuat kebijakan ketenagakerjaan muda. Ini harus mencakup sistem peringatan dini untuk memprediksi kapan akan ada terlalu banyak atau terlalu sedikit orang muda yang mencari pekerjaan.
- b. Memberikan saran khusus tentang bagaimana menggunakan metodologi peramalan yang mempertimbangkan keunikan dan batasan data ketenagakerjaan di Indonesia untuk membantu pengambilan keputusan berdasarkan fakta.
- c. Menawarkan landasan teknologi untuk membangun sistem perencanaan tenaga kerja berbasis data di instansi pemerintah seperti Kementerian Ketenagakerjaan, Bappenas, dan BPS. Ini sangat penting ketika mencoba memanfaatkan bonus demografis di pasar tenaga kerja yang selalu berubah dan dengan data yang tidak mencukupi.
- d. Menawarkan wawasan tentang masa-masa penting dan sejauh mana kesenjangan antara penawaran dan permintaan tenaga kerja pemuda, yang dapat menginformasikan pembuatan intervensi yang ditargetkan, seperti program pelatihan kejuruan, skema insentif perekrutan, dan kebijakan pendidikan adaptif.
- e. Memberikan saran bagaimana membangun Sistem Informasi Kepegawaian Nasional, seperti perlunya publikasi data yang lebih sering dan kemampuan untuk menggabungkan data dari sumber lain (seperti big data dan data administrasi) untuk memungkinkan analisis yang lebih rinci di masa mendatang.

1.5 Batasan Penelitian

Untuk menjaga fokus dan konsistensi kajian, penelitian ini dibatasi pada:

1. Data yang digunakan adalah data semesteran (dua kali per tahun) dari BPS periode 2015–2024 yang kemudian diinterpolasi menjadi data triwulanan menggunakan metode interpolasi Bayesian, menghasilkan sekitar 40 observasi untuk analisis.

2. Variabel target adalah jumlah angkatan kerja usia 15–24 tahun berdasarkan publikasi resmi Badan Pusat Statistik (BPS) melalui Survei Angkatan Kerja Nasional (SAKERNAS).
3. Variabel eksogen yang digunakan meliputi:
 - a. *Tingkat Partisipasi Angkatan Kerja (TPAK)* usia 15–24 tahun,
 - b. *Tingkat Pengangguran Terbuka (TPT)* usia 15–24 tahun,
 - c. Rata-rata Lama Sekolah (RLS),
 - d. penduduk usia muda
 - e. Produk Domestik Bruto (PDB) per kapita,
 - f. Jumlah penduduk usia muda 15–24 tahun.
4. Tiga model yang diteliti adalah Bayesian Structural Time Series (BSTS) sebagai model primer, Seasonal ARIMA (SARIMA) sebagai baseline konvensional, dan Long Short-Term Memory (LSTM) sebagai benchmark eksploratif untuk teknik deep learning. Tidak ada model prediksi lain yang digunakan untuk perbandingan.
5. Pekerjaan ini menggunakan LSTM dengan desain langsung (1-2 lapisan tersembunyi) dan regularisasi yang ketat untuk menangani kumpulan data pendek (40 pengamatan triwulanan), yang kurang optimal untuk model pembelajaran mendalam.
6. Analisis sensitivitas dilakukan dengan membandingkan hasil prediksi yang diperoleh dari data semiannual asli tanpa interpolasi dengan data triwulanan yang diinterpolasi, untuk menilai pengaruh interpolasi terhadap kinerja model.
7. Tahun 2015 hingga 2024 merupakan tahun-tahun yang akan diteliti dalam penelitian tersebut. Selama ini terjadi perubahan struktural yang besar, seperti pergeseran metodologi SAKERNAS pada tahun 2021 dan dampak pandemi COVID-19 pada tahun 2020 dan 2021. Perubahan ini akan dipertimbangkan saat menafsirkan temuan.

8. Penelitian ini memiliki cakupan nasional dan tidak memecah data berdasarkan sektor (pertanian, industri, jasa), geografi (provinsi, kabupaten / kota), atau faktor lainnya (jenis kelamin, pendidikan, status pekerjaan).
9. Penelitian ini berkonsentrasi pada ketepatan prakiraan jangka pendek untuk 1-4 kuartal berikutnya, daripada perkiraan jangka panjang selama dekade berikutnya.
10. Menggunakan ukuran kuantitatif seperti MAPE, RMSE, dan R^2 untuk menilai seberapa baik model bekerja, tanpa menggunakan pendapat ahli atau validasi pemangku kepentingan untuk memeriksa keakuratan prediksi.

1.6 Sistematika Penulisan

Tesis ini disusun ke dalam lima bab utama sebagai berikut:

1. **Bab 1: Pendahuluan**
Menjelaskan latar belakang, rumusan masalah, tujuan, manfaat, batasan, dan sistematika penulisan.
2. **Bab 2: Landasan Pustaka**
Memuat kajian teori terkait demografi, tenaga kerja muda, teori peramalan deret waktu, serta telaah literatur terhadap model BSTS dan LSTM beserta penelitian terdahulu.
3. **Bab 3: Metodologi Penelitian**
Menjelaskan jenis dan sumber data, strategi penanganan missing value, uji linearitas variabel, desain eksperimen, serta metode evaluasi akurasi model.
4. **Bab 4: Hasil dan Pembahasan**
Menyajikan hasil peramalan dan analisis perbandingan performa kedua model, serta interpretasi terhadap kebijakan ketenagakerjaan berdasarkan temuan empiris.
5. **Bab 5: Penutup**
Merangkum kesimpulan utama, implikasi kebijakan, serta rekomendasi untuk penelitian lanjutan yang lebih mendalam.