

Pengembangan Aplikasi Berbasis AI dalam Percepatan dan Akurasi Diagnosa Penyakit Tuberkulosis di Indonesia

Oleh : Fenti Irmawati Supandi & ChatGPT
Untuk : DQLAB A.I Competition, April 2023

Dasar Pemikiran

Perkembangan *Artificial Intelligence* (AI) saat ini semakin pesat, dengan munculnya berbagai teknologi AI yang membantu memecahkan berbagai masalah kompleks dalam berbagai bidang. Salah satu teknologi AI yang sedang tren saat ini adalah *chat GPT* (*Generative Pre-trained Transformer*) yang dikembangkan oleh *OpenAI*. Teknologi ini mampu menghasilkan teks yang mirip dengan yang dihasilkan oleh manusia dan digunakan dalam berbagai aplikasi, seperti *chatbot* dan pembangkit teks.

Melihat potensi teknologi *chat GPT* yang sangat menarik, saya menjadi tertarik untuk mengeksplorasi lebih jauh kemampuan teknologi ini. Disertai tren penggunaan digital masyarakat Indonesia terhadap akses digital kesehatan, maka salah satu ide yang muncul adalah pemanfaatan teknologi *chat GPT* dalam bidang kesehatan, khususnya dalam penanganan penyakit tuberkulosis (TB) di Indonesia. TB merupakan penyakit menular yang masih menjadi masalah kesehatan masyarakat di Indonesia, dan penggunaan teknologi AI dapat membantu meningkatkan akurasi diagnosis serta mempercepat pengobatan pasien TB.

Tuberkulosis atau TB masih menjadi masalah kesehatan global yang signifikan, termasuk di Indonesia. Meskipun upaya pencegahan dan pengobatan sudah dilakukan, angka kasus TB di Indonesia masih sangat tinggi. Menurut data dari Kementerian Kesehatan Indonesia, pada tahun 2022 tercatat lebih dari 700 ribu kasus TB di Indonesia, mengalami peningkatan sebesar 61,98% dibandingkan tahun sebelumnya. Meskipun ada penurunan kasus pada tahun 2020, namun tren kasus TB di Indonesia masih menunjukkan kenaikan dalam dua tahun terakhir.

AI dapat menjadi salah satu solusi dan inovasi untuk membantu mengurangi angka kasus TB di Indonesia. Dengan menggunakan teknologi AI, diharapkan dapat tercipta program-program pencegahan dan pengobatan TB yang lebih efektif dan efisien. Hal ini tentunya akan membantu pemerintah dalam mencapai target eliminasi TB 2030.

Sebagai seorang individu yang peduli dengan masalah kesehatan masyarakat, saya berpendapat bahwa pemanfaatan AI dalam bidang kesehatan dapat memberikan dampak yang besar dalam mengatasi masalah TB di Indonesia. Oleh karena itu, diperlukan kerja sama dari semua pihak untuk memanfaatkan teknologi AI secara optimal guna membantu mencapai target eliminasi TB 2030.

Konsep Pengembangan AI Dalam Diagnosa TB

Konsep AI pada TB dengan metode *supervised learning*:

1. Pendahuluan
 - Definisi TB
 - Definisi AI
 - Jenis-jenis AI
 - *Supervised learning*
2. Pengumpulan Data
 - Sumber data TB
 - *Preprocessing data*
 - Pelabelan data
3. Pembuatan Model
 - Seleksi fitur
 - Pemilihan algoritma
 - Inisialisasi parameter model
4. Training Model
 - Pembagian data train dan test
 - Pelatihan model
 - Validasi model
5. Evaluasi Model
 - Pengukuran performa model
 - *Confusion matrix*
 - *Precision, recall, F1-score*
6. Penggunaan Model
 - Aplikasi model pada data baru
 - Pemantauan pasien TB
 - Pemberian perawatan yang tepat
7. Kesimpulan
 - Manfaat AI pada penanganan TB
 - Tantangan dan kendala dalam penggunaan AI pada TB
 - Peluang pengembangan AI pada penanganan penyakit lain

Konsep AI pada TB dengan metode unsupervised learning atau reinforcement learning:

1. Pendahuluan
 - Definisi TB
 - Definisi AI
 - Jenis-jenis AI
 - *Unsupervised learning; Reinforcement learning*
2. Pengumpulan Data
 - Sumber data TB
 - *Preprocessing data*
3. *Unsupervised Learning*
 - Clustering data TB
 - Metode k-means
 - Analisis cluster
4. Reinforcement Learning
 - Konsep *reinforcement learning*
 - Pemodelan lingkungan TB
 - Pembuatan *reward function*
5. Pembuatan Model
 - Seleksi fitur
 - Pemilihan algoritma
 - Inisialisasi parameter model
6. Training Model
 - Pelatihan model
 - Validasi model
7. Evaluasi Model
 - Pengukuran performa model
 - *Confusion matrix*
 - *Precision, recall, F1-score*
8. Penggunaan Model
 - Aplikasi model pada data baru
 - Pemantauan pasien TB
 - Pemberian perawatan yang tepat
9. Kesimpulan
 - Manfaat AI pada penanganan TB
 - Tantangan dan kendala dalam penggunaan AI pada TB
 - Peluang pengembangan AI pada penanganan penyakit lain

Pembahasan Konseptual

Definisi TB

Menurut Kementerian Kesehatan Indonesia, TB atau Tuberkulosis adalah suatu penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis* yang dapat menular melalui percikan dahak. Tuberkulosis bukan penyakit keturunan atau kutukan dan dapat disembuhkan dengan pengobatan teratur, diawasi oleh Pengawasan Minum Obat (PMO). Sebagian besar kuman TB menyerang paru tetapi bisa juga organ tubuh lainnya.

Menurut WebMD, *Tuberculosis* (TB) adalah infeksi yang menular yang biasanya menyerang paru-paru. Namun, TB juga bisa menyebar ke bagian tubuh lain, seperti otak dan tulang belakang. Jenis bakteri yang disebut *Mycobacterium tuberculosis* menyebabkan penyakit ini.

Sementara itu, menurut NCBI, Tuberkulosis (TB) adalah penyakit manusia yang disebabkan oleh *Mycobacterium tuberculosis*. Biasanya, TB menyerang paru-paru, sehingga penyakit paru menjadi manifestasi paling umum. Sistem organ lain yang umum terkena termasuk sistem pernapasan, sistem gastrointestinal (GI), sistem *lymphoreticular*, kulit, sistem saraf pusat, sistem muskuloskeletal, sistem reproduksi, dan hati.

Dari ketiga kutipan tersebut, dapat disimpulkan bahwa TB atau Tuberkulosis adalah penyakit infeksi yang menular dan disebabkan oleh bakteri *Mycobacterium tuberculosis*. Penyakit ini dapat menyerang berbagai organ tubuh, namun paru-paru menjadi manifestasi paling umum. TB dapat disembuhkan dengan pengobatan teratur dan diawasi oleh Pengawasan Minum Obat (PMO).

Gejala TB

Tuberkulosis (TB) dapat terjadi dalam tiga stadium/stase yang berbeda: TB *Primary*, TB Laten, dan TB Aktif. Berikut adalah penjelasannya:

TB Primary

Beberapa gejala yang terkait dengan TB *Primary* meliputi batuk kering yang berlangsung lebih dari dua minggu, demam, berkeringat pada malam hari, kehilangan nafsu makan, berat badan, dan sesak napas. Jika tidak diobati, bakteri TB dapat menyebar ke organ lain di dalam tubuh dan menghasilkan bentuk TB yang lebih parah.

WHO menjelaskan bahwa "Diagnosis ditegakkan dengan menemukan kuman di sputum atau jaringan." Oleh karena itu, jika seseorang mengalami gejala yang terkait dengan TB *Primary*, penting untuk segera mengunjungi dokter untuk mendapatkan pengobatan yang tepat dan mencegah penyebaran penyakit ke orang lain.

TB Laten

TB Laten terjadi ketika bakteri TB berada dalam tubuh tetapi tidak menunjukkan gejala atau tanda-tanda infeksi. Orang yang memiliki TB laten mungkin terinfeksi dengan bakteri TB selama bertahun-tahun tanpa mengetahui bahwa mereka terinfeksi.

TB laten tidak menular, tetapi masih bisa berkembang menjadi TB aktif jika daya tahan tubuh menurun. Oleh karena itu, orang yang memiliki TB laten perlu diawasi secara ketat dan mengikuti pengobatan yang tepat untuk mencegah perkembangan penyakit.

Menurut CDC, tes kulit tuberkulin (TST) atau uji darah *interferon gamma release assay* (IGRA) digunakan untuk mendeteksi TB laten. Jika hasil tes positif, maka langkah selanjutnya adalah pemeriksaan rontgen dada untuk memastikan bahwa paru-paru masih dalam kondisi sehat.

TB Aktif

TB Aktif terjadi ketika bakteri TB mulai berkembang biak dan menyebar ke organ tubuh lainnya. Orang yang terinfeksi TB aktif dapat menularkan penyakit melalui udara.

Beberapa gejala yang terkait dengan TB aktif meliputi batuk dengan dahak yang berlangsung lebih dari dua minggu, demam, berkeringat pada malam hari, kehilangan nafsu makan, berat badan, dan sesak napas. Selain itu, orang yang terinfeksi TB aktif juga dapat mengalami sakit dada dan batuk darah.

CDC menjelaskan bahwa "Diagnosis TB aktif ditegakkan dengan menemukan kuman di sputum atau jaringan, atau dengan menggunakan tes molekuler cepat."

Pemeriksaan TB

Tes Kulit

Tes kulit dilakukan untuk mengetahui apakah seseorang pernah terpapar bakteri TB. Pada tes ini, sejumlah kecil protein yang disebut dengan tuberkulin diinjeksikan ke dalam lapisan kulit pasien. Setelah 48 hingga 72 jam, kulit pasien diperiksa untuk mengetahui apakah terdapat reaksi pada kulit. Jika terdapat kemerahan dan pembengkakan, ini menunjukkan bahwa pasien telah terpapar bakteri TB dan perlu dilakukan pemeriksaan lebih lanjut.

Rontgen Dada

Rontgen dada adalah pemeriksaan yang dilakukan untuk melihat kondisi paru-paru dan organ dada lainnya. Pemeriksaan ini penting untuk membantu dokter menentukan diagnosis TB. Pada rontgen dada, dokter akan melihat apakah ada bintik-bintik putih atau lesi (luka) pada paru-paru. Lesi yang berbentuk seperti bintik-bintik putih ini dapat menunjukkan adanya infeksi TB.

Tes Dahak

Tes dahak adalah pemeriksaan untuk mengetahui apakah bakteri TB terdapat di dalam dahak pasien. Pemeriksaan ini biasanya dilakukan setelah tes kulit menunjukkan hasil positif atau setelah terdapat gejala TB yang kuat. Pasien diminta untuk mengumpulkan dahak di dalam wadah khusus dan kemudian diuji di laboratorium untuk mengetahui apakah terdapat bakteri TB.

Tes DNA

Tes DNA atau tes PCR (*Polymerase Chain Reaction*) dapat digunakan untuk mengetahui apakah bakteri TB terdapat di dalam dahak pasien. Tes ini menggunakan sampel DNA untuk mengidentifikasi adanya bakteri TB dalam waktu yang lebih singkat dibandingkan tes dahak biasa.

Tes Darah

Tes darah dilakukan untuk mengetahui apakah pasien terinfeksi TB. Tes ini dapat mendeteksi adanya antibodi dalam darah yang menunjukkan adanya infeksi TB. Namun, tes darah tidak dapat membedakan antara infeksi TB aktif atau infeksi TB yang sudah tidak aktif.

AI dalam Kesehatan

Konsep AI untuk kesehatan mencakup penggunaan teknologi AI dalam aplikasi dan perangkat lunak untuk memperbaiki kesehatan dan kesejahteraan manusia. Ada banyak aplikasi AI di bidang kesehatan, termasuk diagnosis penyakit, pengembangan obat baru, pengaturan persediaan obat, pengelolaan data pasien, pengelolaan kesehatan masyarakat, dan banyak lagi.

Penerapan AI dalam kesehatan biasanya melibatkan algoritma pembelajaran mesin, yang memungkinkan sistem untuk belajar dan memperbaiki kinerjanya seiring waktu. Algoritma ini dapat digunakan untuk menganalisis data kesehatan yang besar, seperti riwayat medis pasien dan rekaman medis, serta data genetik, lingkungan, dan perilaku.

Salah satu aplikasi utama AI di bidang kesehatan adalah diagnosis penyakit. Dengan menggunakan teknologi AI, dokter dan ahli kesehatan dapat mengidentifikasi gejala penyakit dan mendiagnosis kondisi medis dengan lebih akurat dan efisien. AI juga dapat membantu dokter dalam perencanaan perawatan dan pengobatan pasien.

Selain itu, teknologi AI juga dapat membantu dalam pengembangan obat baru. Algoritma pembelajaran mesin dapat digunakan untuk menganalisis data klinis dan praklinis untuk mengidentifikasi obat-obatan yang lebih efektif dan aman untuk pengobatan penyakit tertentu.

Namun, penerapan AI di bidang kesehatan juga menghadapi tantangan, termasuk kekhawatiran tentang privasi data, keamanan siber, dan kesalahan interpretasi data. Oleh karena itu, penggunaan teknologi AI di bidang kesehatan harus dilakukan dengan hati-hati dan mengikuti regulasi yang berlaku untuk memastikan keamanan dan kualitas pelayanan kesehatan yang optimal.

Jenis-jenis AI

Supervised Learning:

Supervised learning adalah salah satu metode pembelajaran mesin yang paling populer dan banyak digunakan. Pada metode ini, model AI dilatih dengan menggunakan data yang sudah diberi label atau klasifikasi sebelumnya. Label atau klasifikasi ini digunakan untuk membantu model AI dalam mempelajari pola-pola yang terdapat pada data sehingga dapat melakukan prediksi atau klasifikasi pada data yang belum diberi label.

Contoh penerapan supervised learning pada penanganan TB adalah dengan menggunakan model AI yang dilatih menggunakan data pasien TB yang sudah diberi label sebagai pasien TB atau bukan. Model AI ini dapat digunakan untuk membantu petugas kesehatan dalam membuat diagnosis TB pada pasien baru berdasarkan gejala yang dialami atau hasil tes laboratorium.

Unsupervised Learning:

Unsupervised learning adalah salah satu metode pembelajaran mesin yang tidak memerlukan data yang sudah diberi label atau klasifikasi sebelumnya. Pada metode ini, model AI dilatih dengan menggunakan data mentah tanpa adanya klasifikasi. Tujuan dari penggunaan metode ini adalah untuk mengidentifikasi pola-pola yang terdapat pada data tanpa adanya bantuan label atau klasifikasi.

Contoh penerapan *unsupervised learning* pada penanganan TB adalah dengan menggunakan model AI untuk mengidentifikasi pola-pola pada data hasil tes laboratorium pasien TB. Dengan mengidentifikasi pola-pola ini, model AI dapat membantu dokter dalam membuat diagnosis TB pada pasien baru berdasarkan hasil tes laboratorium yang diperoleh.

Reinforcement Learning:

Reinforcement learning adalah salah satu metode pembelajaran mesin yang menggunakan konsep *reward* atau *reward function* untuk membantu model AI dalam mempelajari pola-pola pada data. Pada metode ini, model AI diberikan tugas atau misi tertentu dan akan mendapatkan reward atau pinalti berdasarkan kinerja yang ditunjukkan dalam menjalankan tugas tersebut.

Contoh penerapan *reinforcement learning* pada penanganan TB adalah dengan menggunakan model AI yang dilatih untuk mengidentifikasi pasien TB yang memiliki risiko lebih tinggi terhadap efek samping obat TB. Model AI ini akan diberikan *reward* ketika berhasil mengidentifikasi pasien dengan risiko efek samping obat TB yang lebih tinggi, sehingga dapat membantu dokter dalam menentukan dosis obat yang tepat untuk pasien tersebut.

Pengumpulan Data

Pengumpulan Data: Pengumpulan data TB dilakukan dari berbagai sumber, seperti rumah sakit, klinik, puskesmas, atau laboratorium. Data yang dikumpulkan dapat berupa data medis, seperti rekam medis, hasil tes laboratorium, foto rontgen atau CT scan, serta data lain yang berkaitan dengan TB. Selain itu, data dapat juga dikumpulkan dari sumber lain seperti jurnal medis atau publikasi ilmiah.

Pentingnya pengumpulan data yang baik dalam penanganan TB dengan menggunakan AI sangat penting karena semakin banyak dan berkualitas data yang dikumpulkan, semakin akurat dan *reliable* model AI yang dapat dihasilkan. Oleh karena itu, penting untuk memastikan bahwa data yang dikumpulkan memiliki kualitas yang baik dan sesuai dengan kebutuhan.

Preprocessing Data: Data yang dikumpulkan perlu diproses dan dipersiapkan sebelum digunakan untuk melatih model AI. Proses ini disebut *preprocessing data*. Proses *preprocessing data* meliputi beberapa tahap, seperti *data cleaning*, *data transformation*, dan *data reduction*.

Data cleaning dilakukan untuk menghapus data yang tidak relevan, data yang tidak lengkap, atau data yang tidak akurat. *Data transformation* dilakukan untuk mengubah data mentah menjadi bentuk yang dapat digunakan oleh model AI, seperti mengubah data teks menjadi vektor, atau mengubah data gambar menjadi bentuk yang dapat dianalisis oleh model AI.

Data reduction dilakukan untuk mengurangi dimensi dari data yang sangat besar. Salah satu metode *data reduction* yang umum digunakan adalah *Principal Component Analysis* (PCA). Metode ini mengurangi dimensi data dengan mereduksi dimensi yang memiliki varian yang rendah sehingga hanya dimensi yang memiliki varian tinggi yang dipertahankan.

Preprocessing data yang baik sangat penting untuk memastikan kualitas dan akurasi model AI yang dihasilkan. Oleh karena itu, sebelum data digunakan untuk melatih model AI, pastikan data telah melalui proses *preprocessing data* yang benar dan sesuai dengan kebutuhan.

Supervised Learning

Pengumpulan data dalam konteks AI TB pada umumnya menggunakan metode *supervised learning*, yaitu metode pembelajaran mesin di mana algoritma mempelajari pola dan aturan dari data pelatihan yang telah diberi label oleh manusia.

Dalam konteks TB, pengumpulan data dilakukan dengan mengumpulkan sampel-sampel data pasien TB yang sudah terdiagnosis melalui pemeriksaan laboratorium dan radiologi. Data tersebut kemudian diberi label berupa status TB dan jenis TB yang diderita oleh pasien, baik itu TB paru atau TB ekstra paru.

Selain itu, data juga dapat diambil dari catatan medis dan riwayat kesehatan pasien yang telah terdiagnosa TB. Data ini mencakup informasi tentang gejala, riwayat penyakit, hasil tes

laboratorium dan radiologi, serta jenis obat dan durasi pengobatan yang diberikan kepada pasien.

Pengumpulan data dilakukan secara terus-menerus dan berkala untuk memperbaharui data dan meningkatkan akurasi model AI dalam mendiagnosis dan mengobati TB. Data tersebut kemudian diolah dan diproses menggunakan berbagai teknik seperti pengolahan bahasa alami, analisis gambar, dan pengolahan data terstruktur untuk menghasilkan model AI yang dapat digunakan dalam mendiagnosis dan mengobati TB.

Unsupervised Learning

Unsupervised Learning seperti yang telah dijelaskan sebelumnya, *Unsupervised Learning* adalah metode pembelajaran mesin yang tidak memerlukan data yang sudah diberi label atau klasifikasi sebelumnya. Pada metode ini, model AI dilatih dengan menggunakan data mentah tanpa adanya klasifikasi. Tujuan dari penggunaan metode ini adalah untuk mengidentifikasi pola-pola yang terdapat pada data tanpa adanya bantuan label atau klasifikasi.

Clustering Data TB Salah satu teknik *Unsupervised Learning* yang umum digunakan dalam penanganan TB adalah *clustering*. *Clustering* merupakan teknik untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan berdasarkan pada karakteristik tertentu. Dalam penanganan TB, data yang dikumpulkan misalnya data hasil tes laboratorium, data gejala pasien, dan data riwayat pengobatan dapat dikelompokkan ke dalam kelompok-kelompok yang memiliki kesamaan.

Metode K-Means Salah satu algoritma *clustering* yang sering digunakan dalam penanganan TB adalah K-Means. K-Means adalah algoritma *clustering* yang bekerja dengan mengelompokkan data ke dalam k kelompok atau *cluster*. Algoritma ini bekerja dengan cara menghitung jarak antara setiap data dengan *centroid* (titik tengah) masing-masing *cluster*. Data akan dikelompokkan ke dalam *cluster* yang memiliki *centroid* terdekat.

Analisis *Cluster* Setelah data TB berhasil dikelompokkan dengan menggunakan teknik *clustering*, langkah selanjutnya adalah melakukan analisis *cluster*. Analisis *cluster* dilakukan untuk memahami karakteristik atau ciri-ciri dari setiap kelompok atau *cluster* yang terbentuk. Dalam penanganan TB, analisis cluster dapat membantu petugas kesehatan dalam mengidentifikasi faktor-faktor yang berkontribusi pada terjadinya TB dan juga membantu dalam mengembangkan strategi penanganan yang lebih efektif.

Reinforcement Learning

Reinforcement Learning adalah metode pembelajaran mesin di mana model belajar dari interaksi yang berulang dengan lingkungan. Pada metode ini, model AI tidak diberi akses langsung ke data atau contoh pelatihan. Sebaliknya, model AI memperoleh pengalaman dengan melakukan tindakan dalam suatu lingkungan dan menerima umpan balik dalam bentuk *reward* atau hukuman.

Terdapat beberapa konsep utama, yaitu *agent*, *environment*, *state*, *action*, *reward*, dan *policy*. *Agent* adalah entitas yang memperoleh pengalaman dari interaksi dengan lingkungan. *Environment* adalah dunia di mana *agent* berinteraksi. *State* adalah kondisi lingkungan pada suatu waktu tertentu. *Action* adalah tindakan yang dapat dilakukan oleh *agent* dalam suatu *state*. *Reward* adalah nilai numerik yang diberikan kepada *agent* setelah melakukan suatu tindakan. *Policy* adalah strategi yang digunakan oleh *agent* untuk memilih tindakan yang akan dilakukan pada suatu *state*.

Pemodelan Lingkungan TB Pada konteks penanganan TB, lingkungan dapat dipandang sebagai sistem biologis yang kompleks yang terdiri dari organisme manusia dan agen penyebab TB. Dalam pemodelan lingkungan TB, data klinis, seperti hasil tes laboratorium, riwayat pengobatan, dan informasi geografis, dapat digunakan untuk merepresentasikan *state* lingkungan. Sedangkan, tindakan yang dapat diambil oleh *agent* dapat berupa pemeriksaan pasien, pengobatan pasien, atau tindakan lain yang dianggap relevan untuk mengurangi penyebaran TB.

Reward function adalah salah satu komponen penting dalam *Reinforcement Learning*. *Reward function* digunakan untuk memberikan *reward* atau hukuman kepada *agent* setelah melakukan suatu tindakan pada suatu *state*. Dalam pemodelan TB, *reward function* dapat dirancang untuk memberikan *reward* yang tinggi jika *agent* berhasil mencegah penyebaran TB atau berhasil mengobati pasien. Sebaliknya, *reward function* dapat memberikan hukuman yang tinggi jika *agent* gagal dalam mengobati pasien atau memperburuk kondisi pasien.

Pembuatan Model

Pembuatan Model Pada tahap ini, dilakukan pembuatan model *machine learning* untuk mengklasifikasikan data TB. Proses pembuatan model meliputi tiga tahapan penting, yaitu seleksi fitur, pemilihan algoritma, dan inisialisasi parameter model.

Seleksi Fitur Seleksi fitur adalah proses untuk memilih fitur atau variabel yang paling relevan dan signifikan dalam melakukan klasifikasi data. Pada data TB, fitur-fitur yang dapat dipertimbangkan antara lain gejala klinis, hasil tes laboratorium, riwayat pengobatan, dan informasi geografis. Seleksi fitur yang tepat dapat membantu memperbaiki kinerja model dan mengurangi *overfitting*.

Pemilihan Algoritma Pemilihan algoritma adalah proses untuk memilih algoritma *machine learning* yang paling sesuai untuk memecahkan masalah klasifikasi data TB. Beberapa algoritma yang dapat dipertimbangkan dalam pemilihan algoritma meliputi *Decision Tree*, *Naive Bayes*, *Support Vector Machine (SVM)*, dan *Neural Network*. Pemilihan algoritma yang tepat dapat membantu meningkatkan akurasi dan efisiensi model.

Inisialisasi Parameter Model Setelah algoritma yang tepat telah dipilih, langkah selanjutnya adalah inisialisasi parameter model. Parameter model adalah nilai-nilai yang digunakan dalam algoritma *machine learning* untuk mengoptimalkan kinerja model. Beberapa parameter model yang perlu diinisialisasi, seperti jumlah *hidden layer* dan *neuron* pada *Neural Network*, kernel pada SVM, dan kriteria pemisahan pada *Decision Tree*. Inisialisasi parameter model yang tepat dapat membantu meningkatkan kinerja model dan mencegah *overfitting*.

Dalam keseluruhan proses pembuatan model, penting untuk melakukan evaluasi model secara berkelanjutan. Hal ini dilakukan dengan melakukan validasi model pada data yang belum pernah dilihat sebelumnya dan mengoptimalkan parameter model sesuai dengan hasil validasi. Dengan cara ini, model dapat terus diperbaiki dan dioptimalkan untuk mencapai kinerja terbaik dalam klasifikasi data TB.

Training Model

Pada tahap *Training Model*, model *machine learning* yang telah dibuat pada tahap sebelumnya akan dilatih menggunakan data latih (*training data*). Pelatihan model dilakukan dengan cara memberikan sejumlah data latih ke dalam model dan mengoptimalkan parameter model sesuai dengan hasil evaluasi model.

Pelatihan Model terdiri dari beberapa tahapan seperti:

- Inisialisasi model: nilai-nilai parameter model yang telah dipilih dan diinisialisasi pada tahap Pembuatan Model digunakan sebagai awal untuk pelatihan model.
- *Forward propagation*: data latih dimasukkan ke dalam model dan output dihasilkan dari tiap *layer* hingga mencapai *output layer*.

- Perhitungan *loss function*: perbedaan antara *output* aktual dengan *output* prediksi dihitung dan dijadikan sebagai evaluasi kinerja model pada setiap *epoch* (putaran pelatihan).
- *Backward propagation*: nilai error dihasilkan dari perhitungan *loss function* dan dikembalikan dari *output layer* ke *hidden layer* dengan cara memperbarui bobot (*weights*) pada setiap *layer*.
- Pengoptimalan parameter: setelah menghitung *loss function* dan memperbarui bobot pada setiap *layer*, parameter model diubah untuk memperbaiki kinerja model. Salah satu metode yang sering digunakan untuk mengoptimalkan parameter adalah dengan menggunakan metode *gradient descent*.

Validasi Model Setelah pelatihan model, langkah selanjutnya adalah melakukan validasi model pada data validasi (*validation data*). Data validasi merupakan data yang berbeda dari data latih dan digunakan untuk mengukur kinerja model secara objektif. Validasi model dilakukan dengan cara memberikan data validasi ke dalam model dan menghitung nilai akurasi, *presisi*, *recall*, dan *F1 score*. Validasi model yang baik dapat memberikan gambaran tentang seberapa baik model dapat digunakan untuk melakukan klasifikasi data TB secara efektif dan efisien.

Dalam keseluruhan proses pelatihan dan validasi model, penting untuk melakukan evaluasi model secara berkelanjutan. Hal ini dilakukan dengan mengoptimalkan parameter model dan menguji kinerja model pada data validasi yang belum pernah dilihat sebelumnya. Dengan cara ini, model dapat terus diperbaiki dan dioptimalkan untuk mencapai kinerja terbaik dalam klasifikasi data TB.

Evaluasi Model

Evaluasi Model Setelah proses pelatihan dan validasi model, tahap selanjutnya adalah Evaluasi Model. Pada tahap ini, model yang telah dilatih dan divalidasi akan diuji performanya pada data uji (*test data*) yang belum pernah digunakan dalam proses pelatihan atau validasi.

Pengukuran performa model dapat dilakukan dengan menggunakan berbagai metode seperti akurasi (*accuracy*), *presisi* (*precision*), *recall* (sensitivitas), *F1-score*, dan *area under the curve* (AUC).

Confusion Matrix untuk mengukur kinerja model, salah satu cara yang paling umum adalah dengan menggunakan *confusion matrix*. *Confusion matrix* adalah tabel yang digunakan untuk menghitung nilai *true positive* (TP), *true negative* (TN), *false positive* (FP), dan *false negative* (FN). Dari *confusion matrix* ini, kita dapat menghitung berbagai matrik evaluasi model seperti akurasi, *presisi*, *recall*, dan *F1-score*.

Precision, Recall, dan F1-score *Presisi* (*precision*) adalah rasio antara *true positive* dengan total data yang diprediksi positif. *Recall* (sensitivitas) adalah rasio antara *true positive* dengan total data yang sebenarnya positif. *F1-score* adalah rata-rata harmonik antara

presisi dan *recall*. Ketiga metrik ini memberikan informasi yang berguna tentang kinerja model dalam melakukan klasifikasi data TB.

Ketika evaluasi model telah selesai dilakukan, hasilnya dapat digunakan untuk menentukan apakah model dapat digunakan untuk melakukan klasifikasi data TB secara efektif dan efisien. Jika performa model masih kurang baik, maka model dapat diperbaiki dan dioptimalkan kembali dengan cara melakukan tahap-tahap pelatihan dan validasi model yang lebih teliti.

Penggunaan Model

Explainable AI (XAI) adalah metode atau teknik dalam pengembangan AI yang bertujuan untuk membantu memahami dan menjelaskan keputusan yang diambil oleh model AI. Dengan menggunakan teknik XAI, kita dapat mengetahui faktor-faktor apa yang mempengaruhi keputusan yang diambil oleh model AI, sehingga dapat membantu kita dalam memahami alasan di balik keputusan yang diambil.

Pada penanganan TB, teknik XAI dapat digunakan untuk memberikan penjelasan yang jelas dan transparan mengenai alasan diagnosis atau rekomendasi terapi yang diberikan oleh model AI. Sebagai contoh, kita dapat menggunakan teknik XAI untuk menjelaskan faktor-faktor apa yang mempengaruhi keputusan model AI dalam menentukan diagnosis TB pada seorang pasien. Dengan cara ini, kita dapat memastikan bahwa keputusan yang diambil oleh model AI didasarkan pada faktor-faktor yang relevan dan dapat dipertanggungjawabkan.

Metode-metode dalam *Explainable AI* Ada beberapa metode yang dapat digunakan dalam teknik XAI, seperti *decision tree*, *feature importance*, *local interpretation*, dan *global interpretation*. *Decision tree* dapat digunakan untuk menunjukkan bagaimana model AI membuat keputusan dengan membagi data ke dalam kelompok-kelompok yang lebih kecil. *Feature importance* dapat digunakan untuk menunjukkan faktor mana yang paling mempengaruhi keputusan yang diambil oleh model AI. *Local interpretation* dapat digunakan untuk memahami keputusan yang diambil oleh model AI pada setiap data input secara individual, sedangkan *global interpretation* dapat digunakan untuk memahami keputusan yang diambil oleh model AI secara keseluruhan pada suatu masalah.

Dalam penggunaannya, metode XAI harus dipilih sesuai dengan jenis model AI yang digunakan dan jenis masalah yang dihadapi. Selain itu, hasil dari teknik XAI juga harus dianalisis dan dievaluasi untuk memastikan bahwa penjelasan yang diberikan oleh model AI dapat dipahami dan dipertanggungjawabkan secara ilmiah.

Kesimpulan

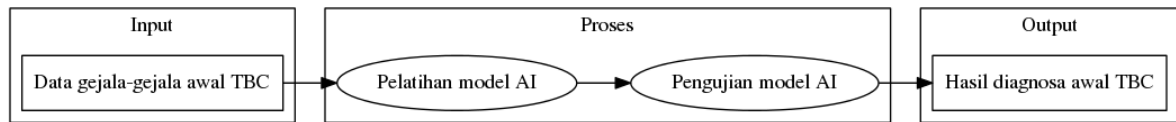
Ensemble learning adalah suatu teknik dalam machine learning yang memanfaatkan beberapa model pembelajaran untuk meningkatkan akurasi prediksi. Ada beberapa metode dalam *ensemble learning*, di antaranya adalah:

1. *Bagging*: Teknik *bagging* (*bootstrap aggregating*) membangun beberapa model independen pada dataset yang diambil secara acak dengan penggantian. Setiap model dilatih pada subset data yang berbeda dan kemudian hasil prediksi dari setiap model digabungkan menjadi satu hasil akhir. *Bagging* sangat cocok digunakan pada dataset yang besar dan rumit.
2. *Boosting*: Teknik *boosting* membangun model secara bertahap dengan menekankan pada sampel yang salah klasifikasinya pada iterasi sebelumnya. Model berikutnya dibuat untuk mengoreksi kesalahan model sebelumnya. *Boosting* sangat cocok digunakan pada dataset yang terbatas.
3. *Stacking*: Teknik *stacking* menggunakan beberapa model pembelajaran dan model *meta-learner* untuk menggabungkan hasil prediksi dari model-model tersebut. Model *meta-learner* dilatih untuk menggabungkan hasil prediksi dari model-model basis dan menghasilkan hasil akhir.

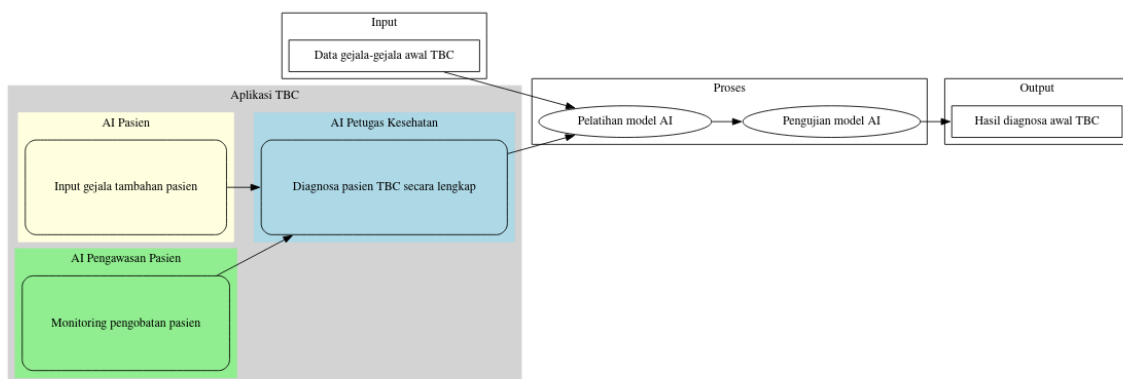
Dalam penanganan TB, *ensemble learning* dapat digunakan untuk meningkatkan performa prediksi diagnosis atau prediksi efektivitas terapi pada pasien TB. Dalam penggunaannya, beberapa model dapat dibangun dengan teknik yang berbeda, kemudian hasil prediksi dari setiap model dapat digabungkan dengan teknik seperti *voting* atau *averaging* untuk menghasilkan hasil akhir yang lebih akurat. Hal ini dapat membantu petugas kesehatan dalam pengambilan keputusan untuk diagnosis dan terapi pasien TB.

Catatan tambahan untuk validasi konsep deteksi TB laten dengan menggunakan AI, tentunya dibutuhkan studi lebih lanjut untuk mengevaluasi keakuratan dan keandalan dari algoritma yang digunakan. Validasi konsep adalah langkah penting sebelum dapat mengimplementasikan teknologi AI untuk deteksi TB laten pada skala yang lebih luas.

Gambaran Konsep



Gambar 1 : Grafik Alur 1



Gambar 1 : Grafik Alur 2

Tahapan alur konsep AI:

1. Input Gejala: Pengguna dapat memasukkan gejala-gejala awal yang dirasakan oleh pasien, seperti batuk terus-menerus selama lebih dari dua minggu, demam, kehilangan nafsu makan, dan penurunan berat badan yang tidak diketahui penyebabnya, dan gejala lebih detail lainnya.

Input ini dikembangkan dalam 3 tahapan input data, yang dapat di akses oleh pasien dan petugas kesehatan, dengan tujuan gejala dapat lebih terbaca lengkap dan akurat

2. Output: AI akan menghasilkan informasi mengenai kemungkinan pasien terkena TB berdasarkan gejala yang dimasukkan. AI juga akan memberikan rekomendasi

tindakan selanjutnya, seperti tes sputum, foto rontgen dada, atau pemeriksaan fisik lebih lanjut.

3. Basis Data: AI akan memanfaatkan data gejala-gejala awal TB dan hasil diagnosa pada pasien yang sebelumnya telah terdiagnosa TB untuk membantu memperbaiki akurasi diagnosa.
4. *Platform*: AI ini dapat diakses melalui aplikasi di *smartphone*, *website*, atau *platform* kesehatan online lainnya, sehingga memungkinkan masyarakat untuk dengan mudah melakukan diagnosis awal TB dari mana saja.
5. Keamanan Data: AI ini akan mengamankan data pasien dan hanya dapat diakses oleh tenaga medis yang berwenang.

Dengan adanya konsep AI ini, diharapkan masyarakat dapat dengan mudah melakukan diagnosa gejala awal TB dan melakukan tindakan selanjutnya yang tepat, sehingga dapat mencegah penyebaran TB dan menurunkan angka kematian akibat penyakit ini di Indonesia.

Rujukan Bacaan

1. <https://www.kemkes.go.id/article/view/23033100001/deteksi-tbc-capai-rekor-tertinggi-di-tahun-2022.html>
2. <https://www.kemkes.go.id/article/view/23033100001/deteksi-tbc-capai-rekor-tertinggi-di-tahun-2022.html>
3. <https://dataindonesia.id/ragam/detail/kasus-tbc-di-indonesia-melonjak-6198-pada-2022>
4. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>
5. <https://www.who.int/indonesia/news/campaign/tb-day-2022/fact-sheets>
6. <https://www.nhs.uk/conditions/tuberculosis-tb/>
7. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>
8. <https://www.webmd.com/lung/understanding-tuberculosis-basics>
9. <https://www.ncbi.nlm.nih.gov/books/NBK441916/>
10. <https://promkes.kemkes.go.id/?p=7439>
11. <https://www.cdc.gov/tb/topic/basics/default.htm>
12. <https://emedicine.medscape.com/article/230802-overview>
13. <https://www.statista.com/outlook/dmo/digital-health/ehealth/indonesia#revenue>

Rujukan Penelitian Terkait

1. Penelitian oleh Aminah Ihsanawati, dkk. (2019) yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.
2. Penelitian oleh Ririn Andiyani, dkk. (2019) yang menggunakan metode decision tree untuk deteksi TBC dari hasil pemeriksaan dahak.
3. Penelitian oleh Muhammad Iqbal, dkk. (2020) yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.
4. Penelitian oleh Ahmed Al-Jumaily (2018) dari New Zealand yang menggunakan metode SVM untuk deteksi TBC dari citra rontgen paru-paru
5. Penelitian oleh Jiang Xiau Zhang, dkk. (2018) dari China yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.

6. Penelitian oleh Marta Mravec, dkk. (2019) dari Slovakia yang menggunakan metode decision tree untuk prediksi efektivitas terapi pada pasien TBC.

Rujukan Aplikasi Penyusunan Penulisan

1. <http://www.plantuml.com/plantuml/uml/>
2. <https://chat.openai.com/chat/2960c355-2ab4-44a1-bee8-ff59a6a0c72b>
3. <https://www.planttext.com/>
4. <https://kroki.io/>
5. <https://www.deepl.com/translator/files>
6. <https://www.grammarly.com/report?alerts=NobwRAxghgLgpgcwPYCcCeYBcYCqA7AazyQHc8wAaMBFJAVwAcswBIBuAGw4Es8FLIsRKjQAJOGFso5bPiKk8AAhKoAJgGcBEenhYALAF8K4aPGTpmAQQgRuquLqhc0AYSR4AZnXVxVAmnombDZOHj4tIQsxSWlmdy8fP2U1TSptOI0sAEZDAF0gA&page=plagiarism&breadcrumbs=true>
7. <https://app.grammarly.com/ddocs/2015630507>

Development of AI-Based Applications in Accelerating and Accuracy of Tuberculosis Disease Diagnosis in Indonesia

By: Fenti Irmawati Supandi & ChatGPT
For: DQLAB A.I Competition, April 2023

Rationale

The development of Artificial Intelligence (AI) is currently accelerating, with the emergence of various AI technologies that help solve various complex problems in various fields. One of the currently trending AI technologies is GPT (Generative Pre-trained Transformer) chat developed by OpenAI. This technology is capable of generating text similar to that produced by humans and is used in various applications, such as chatbots and text generators.

Seeing the very interesting potential of GPT chat technology, I became interested in further exploring the capabilities of this technology. Accompanied by the trend of Indonesian people's digital use of digital access to health, one of the ideas that emerged was the utilization of GPT chat technology in the health sector, especially in the handling of tuberculosis (TB) disease in Indonesia. TB is an infectious disease that is still a public health problem in Indonesia, and the use of AI technology can help improve the accuracy of diagnosis and accelerate the treatment of TB patients.

Tuberculosis or TB remains a significant global health problem, including in Indonesia. Despite prevention and treatment efforts, the number of TB cases in Indonesia is still very high. According to data from the Indonesian Ministry of Health, in 2022 there were more than 700,000 TB cases in Indonesia, an increase of 61.98% compared to the previous year. Although there was a decrease in cases in 2020, the trend of TB cases in Indonesia still shows an increase in the last two years.

AI can be one of the solutions and innovations to help reduce the number of TB cases in Indonesia. By using AI technology, it is hoped that more effective and efficient TB prevention and treatment programs can be created. This will certainly help the government in achieving the 2030 TB elimination target.

As an individual who cares about public health issues, I believe that the utilization of AI in the health sector can have a great impact in addressing the TB problem in Indonesia. Therefore, cooperation from all parties is needed to optimally utilize AI technology to help achieve the 2030 TB elimination target.

AI Development Concept in TB Diagnosis

AI concept in TB with supervised learning method:

- Introduction
 - Definition of TB
 - Definition of AI
 - Types of AI
 - Supervised learning
- Data Collection
 - TB data source
 - Data preprocessing
 - Data labeling
- Model
 - Feature selection
 - Algorithm selection
 - Model parameter initialization
- Training Model
 - Train and test data split
 - Model training
 - Model validation
- Model Evaluation
 - Model performance measurement
 - Confusion matrix
 - Precision, recall, F1-score
- Model Usage
 - Model application on new data
 - TB patient monitoring

- Providing proper care
- Conclusion
 - Benefits of AI in TB management
 - Challenges and constraints in the use of AI in TB
 - Opportunities for AI development in other disease treatments

AI concept in TB with unsupervised learning or reinforcement learning method:

- Introduction
 - Definition of TB
 - Definition of AI
 - Types of AI
 - Unsupervised learning; Reinforcement learning
- Data Collection
 - TB data source
 - Data preprocessing
- Unsupervised Learning
 - Clustering TB data
 - The k-means method
 - Cluster analysis
- Reinforcement Learning
 - The concept of reinforcement learning
 - Modeling the TB environment
 - Reward function creation
- Model
 - Feature selection
 - Algorithm selection

- Model parameter initialization
- Training Model
 - Model training
 - Model validation
- Model Evaluation
 - Model performance measurement
 - Confusion matrix
 - Precision, recall, F1-score
- Model Usage
 - Model application on new data
 - TB patient monitoring
 - Providing proper care
- Conclusion
 - Benefits of AI in TB management
 - Challenges and constraints in the use of AI in TB
 - Opportunities for AI development in other disease treatments

Conceptual Discussion

Definition of TB

According to the Indonesian Ministry of Health, TB or Tuberculosis is an infectious disease caused by the bacteria *Mycobacterium tuberculosis* that can be transmitted through sputum droplets. Tuberculosis is not a hereditary disease or a curse and can be cured with regular treatment, supervised by Supervision of Taking Medicine (PMO). Most TB germs attack the lungs but can also affect other organs.

According to WebMD, Tuberculosis (TB) is a contagious infection that usually affects the lungs. However, it can also spread to other parts of the body, such as the brain and spine. A type of bacteria called *Mycobacterium tuberculosis* causes the disease.

Meanwhile, according to NCBI, Tuberculosis (TB) is a human disease caused by *Mycobacterium tuberculosis*. Typically, TB affects the lungs, making pulmonary disease the most common manifestation. Other commonly affected organ systems include the respiratory system, gastrointestinal (GI) system, lymphoreticular system, skin, central nervous system, musculoskeletal system, reproductive system, and liver.

From these three quotes, it can be concluded that TB or Tuberculosis is a contagious infectious disease caused by the bacteria *Mycobacterium tuberculosis*. The disease can affect various organs of the body, but the lungs are the most common manifestation. TB can be cured with regular medication and supervised by a supervisor (PMO).

Symptoms of TB

Tuberculosis (TB) can occur in three different stages: Primary TB, Latent TB, and Active TB. Here is an explanation:

TB Primary

Some of the symptoms associated with Primary TB include a dry cough that lasts more than two weeks, fever, night sweats, loss of appetite, weight gain, and shortness of breath. If left untreated, the TB bacteria can spread to other organs in the body and produce more severe forms of TB.

WHO explains that "Diagnosis is made by finding germs in sputum or tissue." Therefore, if a person experiences symptoms associated with Primary TB, it is important to visit a doctor immediately to get proper treatment and prevent the spread of the disease to others.

Latent TB

Latent TB occurs when the TB bacteria are in the body but do not show symptoms or signs of infection. People who have latent TB may be infected with TB bacteria for years without knowing they are infected.

Latent TB is not contagious, but it can still develop into active TB if the immune system declines. Therefore, people who have latent TB need to be closely monitored and follow proper treatment to prevent the progression of the disease.

According to the CDC, a tuberculin skin test (TST) or interferon gamma release assay (IGRA) blood test is used to detect latent TB. If the test is positive, the next step is a chest X-ray to ensure that the lungs are still healthy.

Active TB

Active TB occurs when the TB bacteria begin to multiply and spread to other organs of the body. People infected with active TB can transmit the disease through the air.

Some symptoms associated with active TB include coughing with sputum that lasts more than two weeks, fever, night sweats, loss of appetite, weight, and shortness of breath. In addition, people infected with active TB may also experience chest pain and cough up blood.

The CDC explains that "A diagnosis of active TB is made by finding germs in sputum or tissue, or by using rapid molecular tests."

TB Screening

Skin Test

A skin test is done to determine if a person has been exposed to TB bacteria. In this test, a small amount of a protein called tuberculin is injected into the skin layer of the patient. After 48 to 72 hours, the patient's skin is examined to see if there is a skin reaction. If there is redness and swelling, this indicates that the patient has been exposed to TB bacteria and further tests need to be done.

Chest X-ray

A chest X-ray is an examination performed to look at the condition of the lungs and other chest organs. This examination is important to help the doctor determine the diagnosis of TB. On a chest x-ray, the doctor will see if there are white spots or lesions (sores) on the lungs. Lesions that look like white spots may indicate a TB infection.

Sputum Test

A sputum test is an examination to determine if TB bacteria are present in the patient's sputum. It is usually done after a positive skin test or after strong TB symptoms are present. The patient is asked to collect sputum in a special container and then tested in the laboratory to determine if TB bacteria are present.

DNA Test

A DNA test or PCR (Polymerase Chain Reaction) test can be used to determine if TB bacteria are present in a patient's sputum. This test uses a DNA sample to identify the presence of TB bacteria in less time than a regular sputum test.

Blood Test

A blood test is done to determine if the patient is infected with TB. This test can detect the presence of antibodies in the blood that indicate TB infection. However, the blood test cannot distinguish between active TB infection or dormant TB infection.

AI in Healthcare

The concept of AI for healthcare encompasses the use of AI technology in applications and software to improve human health and well-being. There are many applications of AI in healthcare, including disease diagnosis, new drug development, drug inventory management, patient data management, public health management, and more.

AI applications in healthcare usually involve machine learning algorithms, which allow systems to learn and improve their performance over time. These algorithms can be used to analyze large health data, such as patient medical histories and medical records, as well as genetic, environmental, and behavioral data.

One of the major applications of AI in healthcare is disease diagnosis. Using AI technology, doctors and healthcare professionals can identify disease symptoms and diagnose medical conditions more accurately and efficiently. AI can also help doctors in planning the care and treatment of patients.

In addition, AI technology can also help in the development of new drugs. Machine learning algorithms can be used to analyze clinical and preclinical data to identify more effective and safe drugs for the treatment of specific diseases.

However, the application of AI in healthcare also faces challenges, including concerns about data privacy, cybersecurity, and data misinterpretation. Therefore, the use of AI technology in healthcare must be done carefully and follow applicable regulations to ensure safety and optimal quality of healthcare services.

Types of AI

Supervised Learning:

Supervised learning is one of the most popular and widely used machine learning methods. In this method, AI models are trained using data that has been previously labeled or classified. These labels or classifications are used to help the AI model learn the patterns contained in the data so that it can make predictions or classifications on data that has not been labeled.

An example of the application of supervised learning in TB treatment is to use an AI model that is trained using TB patient data that has been labeled as a TB patient or not. This AI model can be used to assist health workers in making a diagnosis of TB in new patients based on symptoms experienced or laboratory test results.

Unsupervised Learning:

Unsupervised learning is one of the machine learning methods that does not require pre-labeled or classified data. In this method, AI models are trained using raw data without

classification. The purpose of using this method is to identify patterns in the data without the help of labels or classification.

An example of the application of unsupervised learning in TB treatment is to use AI models to identify patterns in the laboratory test data of TB patients. By identifying these patterns, the AI model can assist doctors in making a diagnosis of TB in new patients based on the laboratory test results obtained.

Reinforcement Learning:

Reinforcement learning is a machine learning method that uses the concept of reward or reward function to help AI models learn patterns in data. In this method, the AI model is given a specific task or mission and will get a reward or penalty based on the performance shown in carrying out the task.

An example of the application of reinforcement learning in TB treatment is to use an AI model trained to identify TB patients who have a higher risk of TB drug side effects. This AI model will be rewarded when it successfully identifies patients with a higher risk of TB drug side effects, so that it can assist doctors in determining the right drug dose for these patients.

Data Collection

Data Collection: TB data collection is done from various sources, such as hospitals, clinics, health centers, or laboratories. Data collected can be in the form of medical data, such as medical records, laboratory test results, X-rays or CT scans, and other data related to TB. In addition, data can also be collected from other sources such as medical journals or scientific publications.

The importance of good data collection in TB management using AI is critical because the more and better quality data collected, the more accurate and reliable AI models can be generated. Therefore, it is important to ensure that the data collected is of good quality and appropriate to the needs.

Data Preprocessing: The collected data needs to be processed and prepared before it is used to train the AI model. This process is called data preprocessing. Data preprocessing includes several steps, such as data cleaning, data transformation, and data reduction.

Data cleaning is done to remove irrelevant data, incomplete data, or inaccurate data. Data transformation is done to convert raw data into a form that can be used by AI models, such as converting text data into vectors, or converting image data into a form that can be analyzed by AI models.

Data reduction is done to reduce the dimensions of very large data. One of the commonly used data reduction methods is Principal Component Analysis (PCA). This method reduces the dimension of data by reducing dimensions that have low variance so that only dimensions that have high variance are retained.

Good data preprocessing is essential to ensure the quality and accuracy of the resulting AI model. Therefore, before data is used to train AI models, make sure the data has gone through the correct data preprocessing process and according to the needs.

Supervised Learning

Data collection in the context of AI TB generally uses supervised learning, which is a machine learning method where algorithms learn patterns and rules from training data that has been labeled by humans.

In the context of TB, data collection is done by collecting data samples of TB patients who have been diagnosed through laboratory and radiology examinations. The data is then labeled in the form of TB status and the type of TB suffered by the patient, be it pulmonary TB or extra-pulmonary TB.

In addition, data can also be collected from the medical records and medical histories of patients who have been diagnosed with TB. This data includes information on symptoms, disease history, laboratory and radiology test results, as well as the type of medication and duration of treatment given to the patient.

Data collection is carried out continuously and periodically to update the data and improve the accuracy of the AI model in diagnosing and treating TB. The data is then processed using various techniques such as natural language processing, image analysis, and structured data processing to produce an AI model that can be used in diagnosing and treating TB.

Unsupervised Learning

Unsupervised Learning as explained earlier, Unsupervised Learning is a machine learning method that does not require pre-labeled or classified data. In this method, AI models are trained using raw data without classification. The purpose of using this method is to identify patterns in the data without the help of labels or classifications.

Clustering TB Data One of the Unsupervised Learning techniques commonly used in TB treatment is clustering. Clustering is a technique to group data into groups that have similarities based on certain characteristics. In TB treatment, data collected such as laboratory test results, patient symptoms, and treatment history data can be grouped into similar groups.

K-Means Method One of the clustering algorithms often used in TB treatment is K-Means. K-Means is a clustering algorithm that works by grouping data into k groups or clusters. This algorithm works by calculating the distance between each data and the centroid of each cluster. Data will be grouped into the cluster that has the closest centroid.

Cluster Analysis After the TB data has been successfully grouped using clustering techniques, the next step is to perform cluster analysis. Cluster analysis is conducted to understand the characteristics or traits of each group or cluster formed. In TB treatment, cluster analysis can assist health workers in identifying factors that contribute to the occurrence of TB and also help in developing more effective treatment strategies.

Reinforcement Learning

Reinforcement Learning is a machine learning method where the model learns from repeated interactions with the environment. In this method, the AI model is not given direct access to data or training examples. Instead, the AI model gains experience by performing actions in an environment and receiving feedback in the form of rewards or punishments.

There are several key concepts, namely agent, environment, state, action, reward, and policy. An agent is an entity that gains experience from interacting with the environment. Environment is the world in which the agent interacts. State is the condition of the environment at a certain time. Action is an action that can be performed by the agent in a state. Reward is the numerical value given to the agent after performing an action. Policy is the strategy used by the agent to choose the action to be performed in a state.

Modeling the TB Environment In the context of TB management, the environment can be viewed as a complex biological system consisting of human organisms and the TB-causing agent. In modeling the TB environment, clinical data, such as laboratory test results, treatment history, and geographic information, can be used to represent the state of the

environment. Meanwhile, actions that can be taken by the agent can include examining patients, treating patients, or other actions deemed relevant to reduce the spread of TB.

Reward function is one of the important components in Reinforcement Learning. Reward functions are used to reward or punish the agent after performing an action in a state. In TB modeling, the reward function can be designed to give a high reward if the agent successfully prevents the spread of TB or successfully treats a patient. Conversely, the reward function can provide a high punishment if the agent fails to treat the patient or worsens the patient's condition.

Model

Model Building In this stage, a machine learning model is built to classify TB data. The model building process includes three important stages, namely feature selection, algorithm selection, and model parameter initialization.

Feature Selection Feature selection is the process of selecting the most relevant and significant features or variables in classifying data. In TB data, features that can be considered include clinical symptoms, laboratory test results, treatment history, and geographic information. Proper feature selection can help improve model performance and reduce overfitting.

Algorithm selection Algorithm selection is the process of choosing the most suitable machine learning algorithm to solve the TB data classification problem. Some algorithms that can be considered in algorithm selection include Decision Tree, Naive Bayes, Support Vector machine (SVM), and Neural Network. Proper algorithm selection can help improve model accuracy and efficiency.

Model Parameter Initialization Once an appropriate algorithm has been selected, the next step is model parameter initialization. Model parameters are values that are used in machine learning algorithms to optimize model performance. Some model parameters that need to be initialized include the number of hidden layers and neurons in Neural Network, kernel in SVM, and separation criteria in Decision Tree. Proper initialization of model parameters can help improve model performance and prevent overfitting.

In the entire modeling process, it is important to perform continuous model evaluation. This is done by validating the model on data that has never been seen before and optimizing the model parameters according to the validation results. In this way, the model can be continuously improved and optimized to achieve the best performance in TB data classification.

Training Model

In the Training Model stage, the machine learning model that has been created in the previous stage will be trained using training data. Model training is done by providing a number of training data into the model and optimizing the model parameters according to the model evaluation results.

Model Training consists of several stages such as:

- **Model initialization:** the model parameter values that have been selected and initialized in the Model Generation stage are used as the start for model training.
- **Forward propagation:** training data is fed into the model and output is generated from each layer until it reaches the output layer.

- Loss function calculation: the difference between actual output and predicted output is calculated and used as an evaluation of model performance at each epoch (training round).
- Backward propagation: the error value is generated from the loss function calculation and returned from the output layer to the hidden layer by updating the weights in each layer.
- Parameter optimization: after calculating the loss function and updating the weights at each layer, the model parameters are changed to improve the performance of the model. One method that is often used to optimize parameters is to use the gradient descent method.

Model Validation After model training, the next step is to validate the model on validation data. Validation data is data that is different from training data and is used to objectively measure model performance. Model validation is done by providing validation data into the model and calculating the accuracy, precision, recall, and F1 score values. Good model validation can provide an overview of how well the model can be used to classify TB data effectively and efficiently.

In the whole process of model training and validation, it is important to perform continuous model evaluation. This is done by optimizing the model parameters and testing the model performance on validation data that has not been seen before. In this way, the model can be continuously improved and optimized to achieve the best performance in TB data classification.

Model Evaluation

Model Evaluation After the model training and validation process, the next stage is Model Evaluation. At this stage, the model that has been trained and validated will be tested for performance on test data that has never been used in the training or validation process.

Model performance measurement can be done using various methods such as accuracy, precision, recall, F1-score, and area under the curve (AUC).

Confusion Matrix to measure model performance, one of the most common ways is to use confusion matrix. Confusion matrix is a table used to calculate the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values. From this confusion matrix, we can calculate various model evaluation matrices such as accuracy, precision, recall, and F1-score.

Precision, Recall, and F1-score Precision is the ratio of true positives to the total data predicted to be positive. Recall (sensitivity) is the ratio between true positives and total data that are actually positive. F1-score is the harmonic mean between precision and recall. These three metrics provide useful information about the model's performance in classifying TB data.

Once the model evaluation has been completed, the results can be used to determine whether the model can be used to classify TB data effectively and efficiently. If the performance of the model is still not good enough, then the model can be improved and optimized again by conducting more thorough training and validation stages of the model.

Model Usage

Explainable AI (XAI) is a method or technique in AI development that aims to help understand and explain the decisions made by AI models. By using XAI techniques, we can find out what factors influence the decisions made by AI models, so that it can help us understand the reasons behind the decisions made.

In TB treatment, XAI techniques can be used to provide a clear and transparent explanation of the rationale for the diagnosis or treatment recommendation given by the AI model. For example, we can use XAI techniques to explain what factors influence the AI model's decision to diagnose a patient with TB. In this way, we can ensure that the decision made by the AI model is based on relevant factors and can be justified.

Methods in Explainable AI There are several methods that can be used in XAI techniques, such as decision tree, feature importance, local interpretation, and global interpretation. Decision tree can be used to show how AI models make decisions by dividing data into smaller groups. Feature importance can be used to show which factors most influence the decision made by the AI model. Local interpretation can be used to understand the decision made by the AI model on each input data individually, while global interpretation can be used to understand the decision made by the AI model as a whole on a problem.

In use, XAI methods should be selected according to the type of AI model being used and the type of problem at hand. In addition, the results of the XAI techniques should also be analyzed and evaluated to ensure that the explanations provided by the AI model can be understood and scientifically justified.

Conclusion

Ensemble learning is a technique in machine learning that utilizes multiple learning models to improve prediction accuracy. There are several methods in ensemble learning, including:

- Bagging: The bagging technique (bootstrap aggregating) builds multiple independent models on a randomly drawn dataset with replacement. Each model is trained on a different subset of the data and then the prediction results from each model are combined into one final result. Bagging is suitable for large and complex datasets.

- **Boosting:** The boosting technique builds the model incrementally by emphasizing the samples that were misclassified in the previous iteration. The next model is built to correct the errors of the previous model. Boosting is suitable for limited datasets.
- **Stacking:** The stacking technique uses multiple learning models and a meta-learner model to combine the prediction results of the models. The meta-learner model is trained to combine the prediction results of the base models and produce the final result.

In TB treatment, ensemble learning can be used to improve the performance of diagnosis prediction or therapy effectiveness prediction in TB patients. In use, several models can be built using different techniques, then the prediction results from each model can be combined with techniques such as voting or averaging to produce a more accurate final result. This can assist healthcare workers in making decisions for the diagnosis and treatment of TB patients.

An additional note for the validation of the latent TB detection concept using AI is that further studies are needed to evaluate the accuracy and reliability of the algorithms used. Concept validation is an important step before implementing AI technology for latent TB detection on a wider scale.

Concept Overview

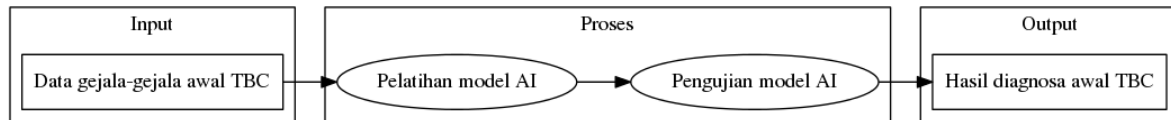


Figure 1: Flow Chart 1

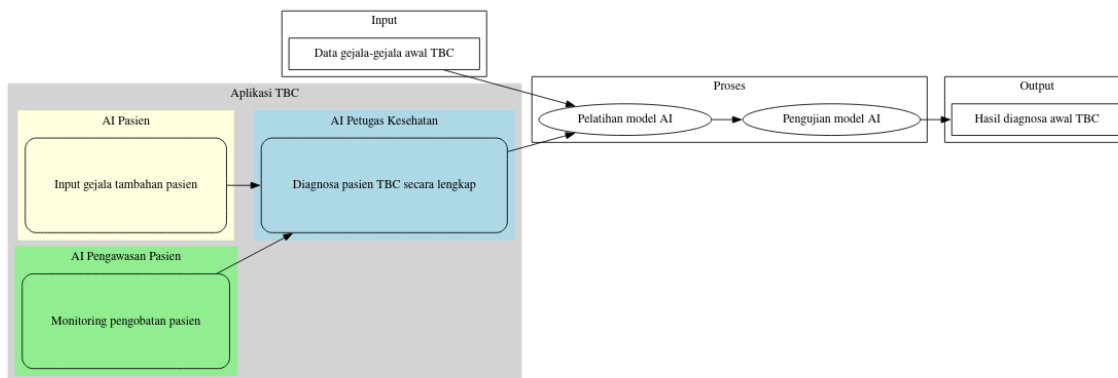


Figure 1: Flow Chart 2

Stages of AI concept flow:

- Symptom Input: Users can input the initial symptoms felt by the patient, such as persistent cough for more than two weeks, fever, loss of appetite, and weight loss of unknown cause, and other more detailed symptoms.
 - This input is developed in 3 stages of data input, which can be accessed by patients and health workers, with the aim that symptoms can be more fully and accurately read.
- Output: The AI will generate information on the likelihood of the patient having TB based on the symptoms entered. It will also provide recommendations for further

actions, such as sputum tests, chest X-rays, or further physical examinations.

- Database: AI will utilize data on early TB symptoms and diagnosis results on patients previously diagnosed with TB to help improve diagnosis accuracy.
- Platform: The AI can be accessed through smartphone apps, websites, or other online health platforms, allowing the public to easily conduct an initial TB diagnosis from anywhere.
- Data Security: This AI will secure patient data and can only be accessed by authorized medical personnel.

With this AI concept, it is hoped that the community can easily diagnose the early symptoms of TB and take appropriate further action, so as to prevent the spread of TB and reduce the mortality rate from this disease in Indonesia.

Reading References

1. <https://www.kemkes.go.id/article/view/23033100001/deteksi-tbc-capai-rekor-tertinggi-di-tahun-2022.html>
2. <https://www.kemkes.go.id/article/view/23033100001/deteksi-tbc-capai-rekor-tertinggi-di-tahun-2022.html>
3. <https://dataindonesia.id/ragam/detail/kasus-tbc-di-indonesia-melonjak-6198-pada-2022>
4. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022>
5. <https://www.who.int/indonesia/news/campaign/tb-day-2022/fact-sheets>
6. <https://www.nhs.uk/conditions/tuberculosis-tb/>
7. <https://www.who.int/news-room/fact-sheets/detail/tuberculosis>
8. <https://www.webmd.com/lung/understanding-tuberculosis-basics>
9. <https://www.ncbi.nlm.nih.gov/books/NBK441916/>
10. <https://promkes.kemkes.go.id/?p=7439>
11. <https://www.cdc.gov/tb/topic/basics/default.htm>
12. <https://emedicine.medscape.com/article/230802-overview>
13. <https://www.statista.com/outlook/dmo/digital-health/ehealth/indonesia#revenue>

Related Research References

1. Penelitian oleh Aminah Ihsanawati, dkk. (2019) yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.
2. Penelitian oleh Ririn Andiyani, dkk. (2019) yang menggunakan metode decision tree untuk deteksi TBC dari hasil pemeriksaan dahak.
3. Penelitian oleh Muhammad Iqbal, dkk. (2020) yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.
4. Penelitian oleh Ahmed Al-Jumaily (2018) dari New Zealand yang menggunakan metode SVM untuk deteksi TBC dari citra rontgen paru-paru
5. Penelitian oleh Jiang Xiau Zhang, dkk. (2018) dari China yang menggunakan metode deep learning untuk deteksi TBC dari citra rontgen paru-paru.
6. Penelitian oleh Marta Mravec, dkk. (2019) dari Slovakia yang menggunakan metode decision tree untuk prediksi efektivitas terapi pada pasien TBC.

Writing Application Reference

1. <http://www.plantuml.com/plantuml/uml/>
2. <https://chat.openai.com/chat/2960c355-2ab4-44a1-bee8-ff59a6a0c72b>
3. <https://www.planttext.com/>
4. <https://kroki.io/>
5. <https://www.deepl.com/translator/files>
6. <https://www.grammarly.com/report?alerts=NobwRAXghgLgpgcwPYCcCeYBcYCqA7AazyQHc8wAaMBFJAVwAcswBIBuAGw4Es8FLIsRKjQAJOGFso5bPiKk8AAhKoAJgGcBEenhhYALAF8K4aPGTpmAQQgRuquLqhc0AYSR4AZnXVxVAmnombDZOHj4tIQsxSWImdy8fP2U1TSptOI0sAEZDAF0gA&page=plagiarism&breadcrumbs=true>
7. <https://app.grammarly.com/ddocs/2015630507>



We didn't find any plagiarism, but we found 5 writing issues.

No plagiarism found	✓
Grammar	✓
Spelling	5
Punctuation	✓
Conciseness	✓
Readability	✓
Word choice	✓
Additional writing issues	✓

