

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Penelitian terkait merupakan suatu kumpulan dari hasil penelitian terdahulu, dimana penelitian terdahulu menjadi referensi atau perbandingan dalam penelitian yang akan dilakukan, berikut merupakan penelitian sebelumnya yang telah melakukan uji coba dalam menerapkan metode *K-Nearest Neighbor* untuk melakukan klasifikasi penyakit yang terjangkit pada manusia. Penelitian terdahulu mengenai penerapan metode *K-Nearest Neighbor* untuk klasifikasi stunting pada balita. Proses dalam penelitian ini melibatkan klasifikasi stunting pada balita dengan menggunakan data dari dataset balita stunting di Puskesmas Sumbermalang, Kabupaten Situbondo, tahun 2018. Data ini dikumpulkan dari berbagai posyandu di Kecamatan Sumbermalang dan telah diolah oleh ahli gizi di Puskesmas Sumbermalang. Total data yang digunakan adalah 200, termasuk balita yang mengalami stunting dan stunting berat. Atribut yang digunakan meliputi jenis kelamin, usia, berat badan, tinggi badan, dan output berupa stunting atau stunting berat. Tahap selanjutnya, data diproses dengan tahap pra-pemrosesan, yang bertujuan untuk menjaga keseimbangan data saat menerapkan metode *K-Nearest Neighbor*. Pra-pemrosesan ini mencakup normalisasi data menggunakan metode *Z-transformation* pada *RapidMiner*. Setelah itu, dilakukan tahap klasifikasi dengan menerapkan metode *K-Nearest Neighbor* pada 200 data balita stunting, untuk mengklasifikasikan apakah balita tersebut mengalami stunting atau stunting berat. Proses ini melibatkan penggunaan uji *cross-validation* dengan nilai *k-folds* sebesar 2, 4, 5, dan 10, serta menentukan jumlah tetangga terdekat dengan nilai *K* sebesar 3, 5, 7, dan 9. Hasil dari penelitian ini menunjukkan bahwa tingkat akurasi tertinggi sebesar 95% diperoleh pada *fold* ke-10 dengan skenario 6 dan nilai *K*=3. Tingkat presisi tertinggi mencapai 100% pada *fold* ke-10 dengan skenario 6 dan *K*=3 untuk kelas positif stunting berat. Sedangkan tingkat *recall* tertinggi adalah 90% pada *fold* ke-10 dengan skenario 6 dan *K*=3 untuk kelas positif stunting berat [3].

Penelitian sebelumnya telah mengkaji penggunaan metode *K-Nearest Neighbor* dalam klasifikasi penyakit ginjal kronis. Proses penelitian ini berfokus pada klasifikasi penyakit ginjal kronik dengan menggunakan dataset dari *University of California Irvine (UCI) Machine Learning Repository* yang disebut *Chronis Kidney Disease*. Dataset ini terdiri dari 400 catatan data, dengan 250 di antaranya menunjukkan kasus ginjal kronis dan 150 tidak. Langkah selanjutnya adalah pra-pemrosesan data, yang melibatkan transformasi data. Data yang telah diubah akan dibagi menjadi data pelatihan dan data pengujian. Proses klasifikasi menggunakan *K-Nearest Neighbor* dilakukan, dan pada pengujian awal dengan *10-fold validation* tanpa optimasi berat *Particle Swarm Optimization* akurasi mencapai 78,75%. Tahap selanjutnya penelitian mencoba melakukan evaluasi *K-Nearest Neighbor* dengan bantuan optimasi berat *Particle Swarm Optimization*. Proses ini melibatkan variasi nilai populasi, yaitu 5, 10, 15, 20, 25, dan 30. Dengan bantuan *Particle Swarm Optimization*, hasil akurasi meningkat menjadi 97,25%. Hal ini menunjukkan bahwa dengan menggunakan *Particle Swarm Optimization* untuk optimalisasi, akurasi dapat ditingkatkan dibandingkan dengan menggunakan *K-Nearest Neighbor* tanpa optimasi *Particle Swarm Optimization* [4].

Penelitian sebelumnya fokus pada klasifikasi penyakit jantung menggunakan metode *K-Nearest Neighbor*. Proses dalam penelitian ini adalah untuk mengklasifikasikan penyakit jantung dengan menggunakan dataset pasien penyakit jantung. Dataset ini terdiri dari 1024 catatan data dengan 14 kolom atribut, termasuk kategori penyakit. Langkah selanjutnya adalah pra-pemrosesan data. Dataset pasien penyakit jantung yang telah diperoleh tidak dapat digunakan langsung dalam analisis karena beberapa data atau atribut tidak lengkap atau tidak dapat digunakan secara langsung. Data dan atribut ini perlu melewati serangkaian tahapan pengolahan awal, seperti pembersihan data (*data cleaning*), proses penyaringan (*filtering*), transformasi data, dan langkah-langkah lainnya, guna memastikan bahwa data yang digunakan dalam analisis adalah berkualitas dan relevan. Tahap selanjutnya yaitu tahap klasifikasi menggunakan *K-Nearest Neighbors (K-NN)* dalam tahapan ini, algoritma K-NN

digunakan untuk mengidentifikasi sejumlah titik terdekat dengan titik data baru, di mana nilai K-nya divariasikan dengan nilai 3, 5, 7, 9, dan seterusnya. Tahap setelah klasifikasi adalah tahap evaluasi dilaksanakan dengan cara menganalisis hasil yang diperoleh dari algoritma yang telah digunakan untuk memastikan bahwa hasil perhitungan dan pengujian sesuai dengan tujuan penelitian yang telah ditetapkan. Tahap selanjutnya adalah proses validasi digunakan untuk mengukur hasil klasifikasi dengan tujuan menentukan tingkat akurasi, presisi, dan recall. Hasil dari penelitian ini dengan penggunaan nilai $K = 3$ dan pembagian bobot data training sebesar 80%, 20% sisanya digunakan sebagai data testing. Model K-NN ini mencapai skor akurasi sebesar 92%, menunjukkan bahwa kemampuan model dalam mengklasifikasikan pasien penyakit jantung sangat tinggi. Presisi model mencapai 90% serta nilai recall sebesar 92%[5].

Penelitian terdahulu mengenai system pakar penyakit *osteoarthritis* menggunakan metode *K-Nearest Neighbors*. Proses pada penelitian ini yaitu untuk melakukan klasifikasi penyakit *Osteoarthritis* dengan mengumpulkan data melalui wawancara dengan pakar atau dokter yang memiliki kompetensi terkait serta menganalisis kasus penyakit *Osteoarthritis*. Tahap selanjutnya yaitu tahap Analisa, dalam tahap Analisa terdapat beberapa Analisa terkait. Analisa yang pertama yaitu analisa data, Proses analisa data dilakukan dengan tujuan untuk mengumpulkan data yang sesuai dengan metode penelitian yang telah ditetapkan. Data-data yang diperlukan dalam analisa ini mencakup jenis-jenis penyakit yang terkait dengan *Osteoarthritis*, gejala-gejala yang terkait dengan masing-masing jenis penyakit *Osteoarthritis*, dan klasifikasi jenisnya. Data-data ini diperoleh melalui wawancara dengan pakar yang berkompeten dalam bidang ini, serta melalui pencarian informasi dari buku referensi dan jurnal ilmiah. Pendekatan ini bertujuan untuk memastikan bahwa data yang diperoleh sangat akurat dan dapat mengurangi potensi kesalahan dalam proses pengolahan data ke dalam sistem. Analisis lanjutan dilakukan dengan tujuan memverifikasi bahwa sistem berjalan sesuai dengan peraturan yang telah ditetapkan oleh peneliti.

Penelitian ini mengambil pendekatan K-Nearest Neighbors (K-NN) sebagai kerangka metodenya, yang merupakan salah satu pendekatan berbasis pengetahuan yang menerapkan konsep teori matematika dalam pemecahan masalah. Metode ini berfungsi untuk melakukan diagnosis dengan mempertimbangkan nilai kondisi dari kasus-kasus sebelumnya, yang memungkinkan pengambilan keputusan yang lebih tepat berdasarkan pengetahuan yang terdapat dalam data sebelumnya. Analisa selanjutnya yaitu analisa sistem yang merupakan langkah kunci dalam perencanaan dan desain sistem yang akan diterapkan, tujuan utama adalah memahami dan mengembangkan sistem yang saat ini sedang berjalan. Dalam konteks ini, sistem ini memerlukan sejumlah data yang harus dimasukkan ke dalam aplikasi dengan cermat. Beberapa data yang harus dimasukkan termasuk ID gejala, nama gejala, nama penyakit, dataset, hasil diagnosis, dan informasi lain yang relevan. Semua data ini akan dikelola dan terorganisir dengan baik dalam sebuah database untuk memastikan sistem berjalan efisien dan efektif. Tahap selanjutnya yaitu tahap perancangan konsep dimana pada tahap ini, dilakukan proses translasi dari kebutuhan perangkat lunak yang telah dianalisis ke dalam representasi desain, sehingga dapat diimplementasikan menjadi sebuah program komputer, desain tersebut akan diterjemahkan ke dalam kode program perangkat lunak untuk menghasilkan program komputer yang sesuai dengan tujuan yang diharapkan. Untuk memvisualisasikan alur kerja dalam proses perancangan perangkat lunak ini, penulis akan memanfaatkan *Unified Modeling Language (UML)* sebagai alat bantu dalam proses pemodelan. Tahap selanjutnya yaitu pengujian sistem yang merupakan tahap penting dalam metode penelitian yang bertujuan untuk menilai sejauh mana hasil penelitian ini sesuai dengan tujuan yang telah ditetapkan. Pengujian akan dilaksanakan untuk mengukur tingkat keberhasilan metode yang diusulkan dan sejauh mana metode tersebut dapat memberikan hasil yang memuaskan sesuai dengan tujuan penelitian. Tahap selanjutnya yaitu tahap implementasi adalah tahap tindakan yang penting dalam upaya untuk memahami bagaimana aplikasi yang telah dirancang dapat berfungsi secara efektif dalam sebuah system, melalui proses implementasi dapat mengevaluasi batasan serta manfaat yang dapat diperoleh dari aplikasi yang telah direncanakan dan dibuat. Tahap selanjutnya yaitu tahap pengujian akurasi yang

dilaksanakan dengan tujuan untuk menguji tingkat keakuratan dan kesesuaian data uji yang diperoleh dari pakar dengan hasil output yang dihasilkan oleh sistem. Terdapat 24 data awal yang digunakan sebagai *data testing*, selanjutnya dilakukan percobaan dengan menggunakan input sesuai dengan data uji, dengan nilai K (jumlah tetangga terdekat) yang ditetapkan sebesar 3. Nilai akurasi dihitung berdasarkan jumlah data uji yang memiliki hasil klasifikasi yang sama dengan diagnosis yang diberikan oleh pakar. Penghitungan akurasi ini dilakukan untuk menguji sejauh mana kesesuaian antara *output* sistem dengan data uji yang telah diperoleh dari pakar. Hasil pengujian menunjukkan bahwa terdapat 2 data uji yang memiliki hasil klasifikasi yang berbeda dengan kelas sebenarnya. Sistem ini berhasil mencapai tingkat akurasi sebesar 91,67%[6].

Penelitian terdahulu mengenai implementasi metode *K-Nearest Neighbor* untuk klasifikasi jenis penyakit katarak. Proses pada penelitian ini yaitu untuk melakukan klasifikasi jenis penyakit katarak dengan menggunakan data yang terdiri dari rekam medis 170 pasien pasien polimata Rumah Sakit Anutapura yang menderita penyakit katarak. Atribut-atribut yang digunakan dalam dataset ini mencakup 11 gejala yang berkaitan dengan mata serta 3 jenis katarak yang berbeda. Tahap selanjutnya yaitu tahap klasifikasi menggunakan metode *K-Nearest Neighbors (KNN)* dilakukan dengan menghasilkan nilai K-optimal melalui metode *10-fold Cross Validation*. Dalam metode ini, dataset awal yang terdiri dari 170 data akan dibagi menjadi 10 partisi. Setiap partisi akan terdiri dari 17 data, dengan 1 data sebagai data testing dan 16 data sebagai data training. Proses KNN kemudian dilakukan dengan menghitung jarak *Euclidean* antara data *testing* dan data *training* untuk menentukan tetangga terdekat. Metode *10-fold Cross Validation* membantu dalam menentukan parameter K yang optimal untuk klasifikasi yang akurat berdasarkan dataset yang ada.

Tahap selanjutnya yaitu tahap uji akurasi metode *K-Nearest Neighbors (KNN)* ditemukan bahwa jumlah data yang benar diprediksi adalah sebanyak 154 data, sedangkan data yang salah atau salah diprediksi sebanyak 16 data, dengan rincian data yaitu prediksi pasien senilis yang benar adalah sebanyak 79 pasien, prediksi pasien kongenital yang benar adalah sebanyak 27 pasien, prediksi pasien traumatik yang benar adalah sebanyak 50 pasien, pasien yang salah diprediksi adalah sebanyak 14 pasien. Hasil akurasi yang didapat yaitu 91,76%[7].

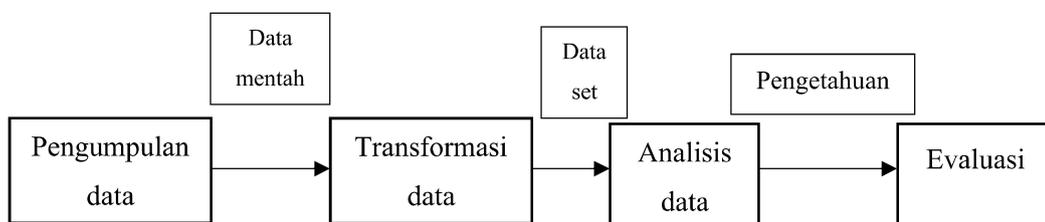
Penelitian berbeda dengan penelitian sebelumnya, perbedaanya yaitu penyakit yang di klasifikasi sebelumnya yaitu berupa stunting pada balita, penyakit ginjal kronis dan diabetes. sedangkan pada penelitian ini penyakit yang di klasifikasi yaitu TBC (*tuberculosis*). hal yang membedakan lainnya itu pada jumlah kelas yang akan diuji dimana penelitian ini terdapat 4 kelas sedangkan penelitian sebelumnya kebanyakan menggunakan 2 kelas dan 3 kelas. Data diambil langsung dari Puskesmas Bumiayu melalui wawancara dengan kepala Puskesmas Bumiayu Bapak dr. Ali Budiarto. Kasus TBC juga terjadi diwilayah Brebes, wilayah Brebes termasuk daerah dengan jumlah kasus yang tinggi terutama diwilayah Bumiayu. Kasus TBC paru di Bumiayu terbilang tinggi serta penularan penyakit TBC paru sangat mudah sehingga dapat dengan cepat menular kepada orang yang sehat. Penularan yang mudah dan cepat dapat menyebabkan bertambahnya penderita TBC paru lebih cepat dan memiliki resiko kematian yang tinggi sehingga perlunya diagnosis dini untuk mengurangi penularan dan mengurangi resiko kematian.

2.2. Landasan teori

2.2.1. Data Mining

Data mining merupakan komponen penting yang tergabung dalam langkah-langkah *Knowledge Discovery in Databases* (KDD), sebuah rangkaian proses yang dirancang untuk menggali pengetahuan berharga dari kumpulan data. *Knowledge discovery* pada dasarnya melibatkan serangkaian tahap yang meliputi pembersihan data guna menghilangkan data tidak valid (*data cleaning*), integrasi data untuk menggabungkan berbagai sumber data (*data integration*), pemilihan data yang relevan sesuai kebutuhan analisis (*data selection*), *transformasi data* untuk memformat ulang data, evaluasi pola untuk menguji validitas penemuan (*pattern evaluation*), dan penyajian hasil pengetahuan yang telah ditemukan (*knowledge presentation*)[8]. *Data mining* adalah proses analisis data yang memanfaatkan metode dan teknik tertentu untuk mengidentifikasi dan menemukan pola serta aturan yang terdapat dalam suatu kumpulan data. Pengambilan ringkasan informasi dari sekelompok data dengan menerapkan algoritma dan teknik yang melibatkan prinsip-prinsip statistik, pembelajaran mesin, dan manajemen basis data, dikenal sebagai proses *data mining*. *Data mining* merupakan fungsi yang memungkinkan pencarian informasi penting dalam dataset besar. *Data mining* juga memiliki kemampuan untuk menghasilkan pengetahuan dari sekumpulan data yang berjumlah besar[9].

Langkah-langkah yang tercakup dalam rangkaian proses ini menjadikannya sebagai suatu pendekatan yang komprehensif untuk menggali informasi dari data mentah yang tersedia, dengan memastikan bahwa setiap tahap memiliki peran kunci dalam menghasilkan hasil akhir yang lebih baik.



Gambar 2. 1 Alur *Data Mining*

Pada Gambar 2.1 terlihat sebuah diagram yang secara visual menggambarkan bagaimana aliran informasi mengalir dalam kerangka proses data mining. Dalam gambar tersebut, proses data mining yang tergambar mengindikasikan bahwa proses ini memiliki sifat iteratif, artinya dilakukan secara berulang-ulang. Evaluasi hasil pengetahuan yang diperoleh melalui data mining mungkin mengindikasikan kebutuhan akan pengetahuan yang lebih komprehensif, yang mungkin melibatkan peningkatan dalam pemahaman kita. Hasil evaluasi ini dapat mendorong untuk melakukan perbaikan lebih lanjut pada data yang digunakan (dataset) dengan memperbaikinya. Hasil evaluasi ini dapat melakukan perubahan pada sistem secara keseluruhan, yang mungkin melibatkan perubahan pada metodologi atau strategi yang sedang digunakan dalam proses data mining[8].

2.2.2. Klasifikasi

Klasifikasi merupakan suatu proses yang melibatkan evaluasi terhadap suatu entitas data dengan tujuan untuk menentukan pengelompokannya ke dalam salah satu kategori yang telah ditentukan sebelumnya. Untuk melaksanakan klasifikasi, langkah awal adalah membangun sebuah model berdasarkan data latih yang ada. Model tersebut berfungsi sebagai representasi dari pola-pola yang ada dalam data latih. Model ini digunakan untuk mengklasifikasikan data-data yang baru masuk. Klasifikasi merupakan proses yang melibatkan pelatihan atau pembelajaran terhadap suatu fungsi yang ditargetkan untuk menghubungkan atau mengaitkan setiap kombinasi atribut atau fitur data ke salah satu dari label-label kelas yang telah ditentukan sebelumnya. Dengan kata lain, sistem klasifikasi belajar untuk mengenali pola dan relasi antara atribut-atribut data dengan label-label kelas. Tujuan utama dari sistem klasifikasi adalah melakukan klasifikasi data secara akurat, namun penting untuk diingat bahwa tidak ada sistem yang dapat mencapai tingkat kebenaran 100%. Evaluasi kinerja sistem klasifikasi menjadi hal yang sangat penting. Dengan mengukur kinerja sistem, kita dapat mengevaluasi sejauh mana sistem tersebut mampu mengklasifikasikan data dengan benar, dan juga mengidentifikasi area-area di mana perbaikan mungkin diperlukan[8].

2.2.3. Algoritma *K-Nearest Neighbor* (KNN)

Algoritma *K-Nearest Neighbor* (KNN) dapat dikelompokkan sebagai salah satu teknik dalam konsep *lazy learning*. Algoritma *K-Nearest Neighbor* (K-NN) diterapkan dengan cara mengenali sekelompok n objek dari data pelatihan yang memiliki kemiripan tertinggi dengan objek yang ada dalam data uji. Algoritma *K-Nearest Neighbor* merupakan metode yang sangat bermanfaat dalam melakukan klasifikasi terhadap suatu objek. Pendekatan ini dilakukan dengan memeriksa data pelatihan yang memiliki jarak terdekat dengan objek yang ingin diklasifikasikan. Metode *K-Nearest Neighbor* digunakan sebagai alat untuk mengkategorikan objek berdasarkan kemiripan atau kesamaannya dengan data pelatihan, dengan memprioritaskan objek yang memiliki tingkat kesamaan tertinggi. Metode pencarian kasus baru sering kali melibatkan perhitungan kedekatan antara kasus baru dan kasus lama, di mana proses ini didasarkan pada pengukuran kesesuaian bobot dari berbagai fitur yang ada dalam setiap kasus. Algoritma *K-Nearest Neighbor* menggambarkan algoritma klasifikasi yang mengedepankan konsistensi saat dihadapkan pada volume data yang besar. Metodenya didasarkan pada ide bahwa sebuah data dapat diklasifikasikan dengan melihat jarak terdekat antara data tersebut dengan titik-titik terdekat yang telah ada dalam data latihan. Kelebihan dari algoritma *K-Nearest Neighbor* adalah kemampuannya untuk mengelola data dalam skala yang besar dan kemampuannya untuk mengklasifikasikan data berdasarkan konsep jarak terdekat antara data yang sedang dievaluasi dengan titik-titik dalam kumpulan data pelatihan. Fleksibilitas yang dimiliki algoritma *K-Nearest Neighbor* terlihat dari dasarnya yang berfokus pada kedekatan data yang telah ada dalam data pelatihan. Proses pengklasifikasian objek dalam algoritma *K-Nearest Neighbor* dilakukan dengan memperhitungkan jaraknya dari tetangga terdekat. Penempatan objek dalam kelas tertentu didasarkan pada mayoritas kelas yang ada di antara tetangga-tetangga terdekatnya. Dalam konteks ini, metode Euclidean kerap digunakan sebagai pengukuran jarak untuk menentukan seberapa dekat dua objek dalam ruang fitur [10].

Berikut langkah-langkah dalam proses klasifikasi menggunakan algoritma *K-Nearest Neighbor* yaitu:

1. Menetapkan suatu nilai K yang sesuai.
2. Menghitung jarak antara data uji dan data latih menggunakan metode seperti Euclidean.
3. Mengkategorikan data berdasarkan hasil perhitungan jarak yang telah dihitung.
4. Mencari kelompok data yang memiliki nilai jarak terkecil atau yang paling mendekati.
5. Memilih kategori yang paling umum muncul di antara sejumlah K yang telah ditetapkan sebagai hasil prediksi.

Untuk klasifikasi penyakit TBC paru di Puskesmas Bumiayu dengan menggunakan algoritma K-NN digunakan model sebagai berikut:

$$D = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} \quad (2.1)$$

Keterangan:

x = sampel data

y = data uji

D = Jarak

2.2.4. Tuberculosis

TBC merupakan penyakit menular yang disebabkan oleh *mycobacterium tuberculosis*[1]. TBC umumnya menyerang paru-paru tetapi TBC juga dapat menyerang organ lain seperti ginjal dan tulang belakang. Gejala yang timbul pada penderita TBC paru yaitu batuk lebih dari dua minggu, batuk berdarah, nafsu makan berkurang, berat badan menurun, lemas, sakit dinding dada dan sesak nafas.

TBC paru dapat menular pada saat penderita berbicara, batuk dan bersin yang secara tidak langsung penderita mengeluarkan percikan dahak (*droplet*) yang mengandung banyak kuman. Kuman TBC paru menyebar kepada orang sekitar melalui transmisi atau udara (*droplet* dahak penderita TBC paru). Penyebaran TBC paru menjadi lebih cepat karena beberapa factor yaitu kebiasaan buruk yang seringkali dimiliki oleh pasien TBC paru seperti kebiasaan meludah sembarangan, tanpa memperhatikan lingkungan sekitar, kurangnya kesadaran dalam menjaga kebersihan lingkungan juga menjadi faktor yang memperburuk kondisi ini. Rumah tempat tinggal pasien TBC paru juga memiliki ventilasi yang kurang baik, sehingga pergantian udara yang lancar terhambat dan sinar matahari sulit masuk dengan optimal sehingga kelembaban di dalam ruangan meningkat, menciptakan lingkungan yang sangat mendukung perkembangbiakan bakteri penyebab TBC. Kondisi ini dapat memiliki dampak yang signifikan terhadap penyebaran penyakit. Masyarakat yang sehat dan tinggal serumah dengan penderita TB paru menjadi kelompok yang sangat rentan terhadap potensi penularan penyakit ini. Upaya yang perlu dilakukan bersama untuk mengedukasi pasien TB paru mengenai pentingnya etika batuk yang benar serta menjaga kebersihan lingkungan sekitar. Selain itu, peningkatan kualitas ventilasi rumah dan upaya untuk meningkatkan sirkulasi udara dan sinar matahari juga dapat berkontribusi dalam mengurangi risiko penularan penyakit yang berpotensi fatal ini. TBC paru masih menjadi masalah kesehatan di dunia, Indonesia menduduki posisi ketiga jumlah penderita TBC paru setelah India dan Cina. Kasus TBC paru lebih banyak di negara berkembang contohnya di Indonesia [11]. TBC paru dapat ditegakkan diagnosis menggunakan hasil laboratorium dengan pemeriksaan mikroskopik BTA menggunakan pewarnaan *Ziehl Nelsen* [12]. Berikut adalah kelas BTA berdasarkan hasil laboratorium:

1. Negative (-): tidak ditemukannya bakteri *mycobacterium tuberculosis* dalam 100 lapang pandang.
2. 1+: ditemukan 10-99 bakteri *mycobacterium tuberculosis* dalam 100 lapang pandang.

3. 2+: ditemukan 1-10 bakteri *mycobacterium tuberculosis* dalam setiap lapang pandang.
4. 3+: ditemukan lebih dari 10 bakteri *mycobacterium tuberculosis* dalam setiap lapang pandang.

2.2.5. *Confusion Matrix*

Confusion Matrix memiliki bentuk tabel berupa matriks yang berperan penting dalam mengilustrasikan performa suatu model klasifikasi pada rangkaian data uji, di mana nilai sebenarnya telah teridentifikasi sebelumnya. Dalam kaitannya dengan Confusion Matrix, terdapat empat kombinasi nilai yang saling berbeda antara prediksi dan nilai aktual. Confusion Matrix membentuk gambaran mengenai bagaimana model klasifikasi berkinerja. True Positive (TP) menggambarkan situasi di mana model memprediksi dengan benar bahwa data adalah positif. Selanjutnya, True Negative (TN) mencerminkan kapabilitas model dalam mengenali data negatif dengan akurat. Di samping itu, terdapat False Positive (FP) yang mengindikasikan ketika model salah dalam mengklasifikasikan data sebagai positif, dan juga False Negative (FN) yang menunjukkan kekeliruan model dalam memprediksi data sebagai negatif yang seharusnya positif. Implementasi Confusion Matrix dalam uji data mampu melaksanakan analisis performa dengan menghitung beberapa indikator utama seperti Accuracy, Precision, dan Recall/Sensitivity.

- a. *Accuracy* adalah tingkat kesesuaian antara nilai prediksi dan nilai aktual yang terdeteksi dalam suatu klasifikasi. *Accuracy* memungkinkan kita memberikan jawaban yang signifikan terhadap pertanyaan esensial, tanpa menghiraukan apakah penentuan tersebut benar atau salah. Kemampuan ini dapat kita peroleh melalui suatu formula matematis yang secara khusus memungkinkan perhitungan ini.

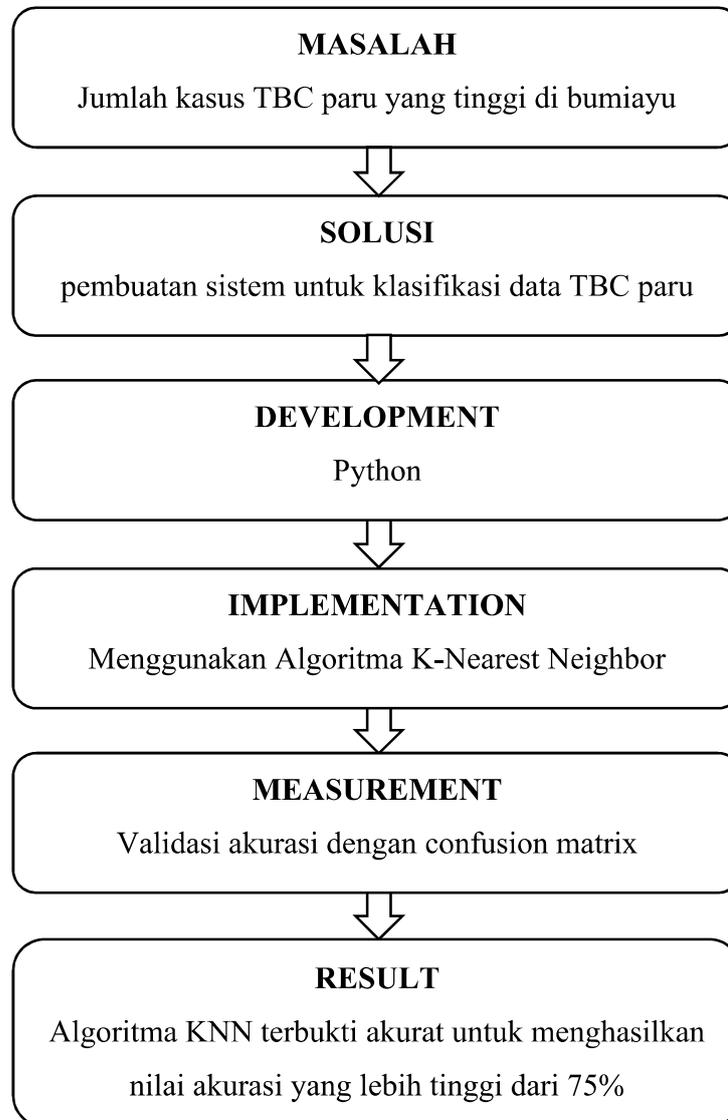
$$\text{Akurasi} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}} \times 100\% \quad (2.2)$$

- b. *Precision* adalah tingkat ketepatan antara data yang diminta dengan hasil prediksi yang dihasilkan oleh model klasifikasi. *Precision* memberikan kemampuan untuk memberikan jawaban yang sangat penting terhadap pertanyaan fundamental.
- c. *Recall* atau *Sensitivity* adalah kemampuan suatu model dalam secara efektif mengidentifikasi kembali informasi tertentu. *Recall* memiliki peran penting dalam memberikan jawaban yang signifikan terhadap pertanyaan mendasar. Memanfaatkan persamaan matematis yang telah ditetapkan sebelumnya serta mampu menghitung dan menginterpretasi *Recall* dengan lebih cermat[13].

2.2.6. Tkinter

Tkinter adalah sebuah perpustakaan yang digunakan untuk menciptakan antarmuka *GUI* dalam bahasa pemrograman Python dengan bantuan *Tk GUI Toolkit*. *Tkinter* diartikan sebagai suatu kerangka *GUI* yang dirancang khusus untuk mempermudah pengembangan program dengan aspek visual. *Tkinter* memiliki fungsi sebagai perpustakaan grafis, *tkinter* mampu menyajikan kemudahan yang sangat diperlukan dalam pengembangan program berbasis tampilan visual di setiap Toolkit *GUI*, komponen elemen antarmuka pengguna, yang disebut sebagai widget, meliputi objek-objek seperti tombol, *scrollbar*, kotak daftar, *checkboxbutton*, *radiobutton*, *label teks*, dan lainnya yang serupa. Komponen-komponen ini terdapat dalam *widget* dan mampu mengemas detail-detail implementasi secara internal. Perilaku standar untuk setiap widget sudah terdefinisi, hal ini membuat proses pemrograman *GUI* menjadi lebih mudah.[14]

2.3. Kerangka Pemikiran



Gambar 2. 2 Kerangka Pemikiran

Gambar 2.2 merupakan kerangka penelitian yang dilakukan dalam melakukan klasifikasi penyakit TBC paru di Puskesmas Bumiayu. Penjelasan dari tiap-tiap tahap pada kerangka penelitian akan dijelaskan lebih lanjut yaitu sebagai berikut:

1. Masalah

Penelitian ini berfokus pada permasalahan utama yang terjadi di wilayah Bumiayu, yaitu tingginya angka kasus Tuberkulosis (TBC) paru. Fenomena ini menjadi signifikan karena TBC paru merupakan penyakit yang memiliki tingkat penularan yang sangat tinggi, sehingga berisiko tinggi menular kepada individu yang memiliki kondisi kesehatan yang baik. Dampak dari penularan yang mudah ini semakin memperbesar jumlah penderita TBC paru di wilayah tersebut. Penyakit ini juga memiliki tingkat kematian yang cukup tinggi, menjadikannya sebagai ancaman serius. Penularan penyakit TBC paru yang mudah dan konsekuensi serius yang ditimbulkan, sangat penting untuk mengambil langkah-langkah pencegahan yang efektif. Salah satu langkah yang krusial adalah adanya diagnosis dini. Diagnosis dini memungkinkan deteksi awal kasus TBC paru pada individu yang terinfeksi, sehingga tindakan pengobatan dapat segera dilakukan. Penularan penyakit dapat ditekan sejak dini, mengurangi risiko penularan kepada orang lain yang berada di sekitar penderita. Selain itu, tindakan pengobatan yang tepat waktu juga akan membantu mengurangi tingkat kematian akibat TBC paru. Hal penting yang perlu dikedepankan yaitu diagnosis dini sebagai strategi utama, diharapkan bahwa penelitian ini akan memberikan kontribusi dalam mengurangi jumlah kasus TBC paru di wilayah Bumiayu serta mengurangi dampak negatif yang ditimbulkannya, baik dari segi penularan maupun risiko kematian.

2. Solusi

Solusi yang dapat diimplementasikan untuk mengatasi permasalahan tingginya kasus TBC paru di wilayah Bumiayu adalah melalui pendekatan klasifikasi data pasien. Pendekatan ini melibatkan analisis mendalam terhadap data medis dan informasi kesehatan pasien yang telah terdiagnosis TBC paru.

Dengan memanfaatkan teknik-teknik klasifikasi data, dokter dan tenaga medis dapat mengidentifikasi pola-pola khas yang terkait dengan kondisi TBC paru pada tahap awal. Analisis klasifikasi data membuat dokter dapat memberikan diagnosis dini kepada pasien yang diduga terinfeksi TBC paru.

Tindakan ini memungkinkan pemberian pengobatan yang tepat waktu dan spesifik sesuai dengan kondisi pasien. Penularan penyakit dapat ditekan secara signifikan, karena pasien yang telah terdiagnosis dapat segera menjalani pengobatan dan langkah-langkah isolasi yang sesuai. Pendekatan ini juga berpotensi mengurangi risiko kematian yang tinggi akibat TBC paru. Dengan mengidentifikasi kasus pada tahap awal, pasien akan memiliki peluang yang lebih baik untuk merespons pengobatan dengan efektif. Ini akan berdampak positif pada tingkat kesembuhan pasien dan mengurangi angka kematian yang terkait dengan penyakit ini. Secara keseluruhan, penggunaan pendekatan klasifikasi data pasien TBC paru merupakan langkah proaktif dalam menghadapi permasalahan ini. Melalui kombinasi teknologi dan pengetahuan medis, solusi ini dapat berpotensi mengurangi penularan TBC paru di wilayah Bumiayu dan mengurangi risiko kematian yang disebabkan oleh penyakit ini.

3. *Development*

Dalam pengembangan sistem yang merupakan fokus dari penelitian ini, dipilih pendekatan menggunakan bahasa pemrograman Python. Bahasa pemrograman *Python* dipilih karena fleksibilitasnya yang tinggi dalam analisis data medis dan pemrosesan informasi kesehatan.

4. *Implementation*

Pendekatan yang digunakan adalah dengan menerapkan Algoritma *K-Nearest Neighbor* untuk melakukan klasifikasi penyakit TBC paru di Puskesmas Bumiayu. Tahapan implementasi dimulai dengan pengumpulan data medis dari pasien yang telah terdiagnosis TBC paru di Puskesmas tersebut. Tahap selanjutnya yaitu proses klasifikasi menggunakan Algoritma *K-Nearest Neighbor* akan diimplementasikan secara rinci dalam Bab 3 penelitian ini.

5. *Measurement*

Measurement merujuk pada nilai-nilai yang akan diukur sebagai acuan dalam evaluasi kinerja model. Penggunaan *Confusion Matrix* digunakan untuk mendapatkan pemahaman yang lebih mendalam tentang seberapa baik model klasifikasi bekerja.

Confusion Matrix adalah suatu tabel yang digunakan untuk menggambarkan performa klasifikasi dengan membandingkan hasil prediksi model terhadap kenyataan yang sebenarnya. Penggunaan *Confusion Matrix* untuk mengetahui tentang performa model klasifikasi penyakit TBC paru di Puskesmas Bumiayu dan mengevaluasi sejauh mana akurasi dalam memberikan diagnosis yang tepat.

6. Result

Hasil yang dihasilkan dari penelitian ini menunjukkan bahwa *Algoritma K-Nearest Neighbor (K-NN)* berhasil diimplementasikan untuk tujuan klasifikasi penyakit TBC paru, algoritma ini mampu memberikan tingkat akurasi yang cukup tinggi, yakni mencapai atau bahkan melebihi angka 75%.