**Proposal Pengajuan Judul Pembahasan**

**“Analisis Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naive Bayes”**

**Dosen Pengampu:** **Hendri Karisma S.Kom, M.T**

****

**DISUSUN OLEH:**

**KRESNA RISQI RAMADHAN (241552010006)**

**MUHAMMAD SHAADAM HAIDAR YUWONO (241552010005)**

**SEKOLAH TINGGI MANEJEMAN INFORMATIKA**

**KOMPUTER (STMIK) TAZKIA**

**TEKNIK INFORMATIK**

**2024/2025**

# KATA PENGANTAR

## Tiada kalimat yang pantas kami ucapkan kecuali rasa syukur kepada Allah Yang Maha Esa atas selesainya proposal pengajuan judul “Analisis Prediksi Kejadian Kriminal di Kabupaten Sukoharjo Tahun 2024 Menggunakan Metode Naïve Bayes”. Tidak lupa pula dukungan baik secara materil dan moral yang diberikan kepada kami dalam penyusunan proposal ini. Oleh karena itu, izinkan kami mengucapkan rasa terima kasih kepada:

* Hendri Karisma S.Kom, M.T, selaku dosen mata kuliah probabilitas dan statistika.

Kami menyadari bahwa proposal pengajuan judul ini jauh dari kata sempurna dan masih terdapat beberapa kekurangan, oleh karena itu kami sangat mengharapkan saran dan kritik yang membangun untuk penyempurnaan proposal pengajuan judul ini.

Bogor, 3 July 2025

**PENDAHULUAN**

* 1. **Latar Belakang**

Penyakit paru-paru merupakan salah satu masalah kesehatan global yang serius, berkontribusi pada tingginya angka morbiditas dan mortalitas. Di Indonesia, penyakit seperti Tuberkulosis (TBC) dan Penyakit Paru Obstruktif Kronis (PPOK) masih menjadi ancaman utama. Deteksi dini sangat krusial untuk meningkatkan keberhasilan pengobatan dan mencegah komplikasi. Namun, diagnosis konvensional seringkali memerlukan waktu dan sumber daya yang besar. Oleh karena itu, diperlukan solusi yang efektif dan efisien untuk membantu tenaga medis dalam proses deteksi dini, dan pendekatan *machine learning* dapat menjadi jawabannya.

#### **1.2 Rumusan Masalah**

1. Bagaimana membangun model klasifikasi untuk memprediksi penyakit paru-paru menggunakan algoritma Naive Bayes dari data gejala yang bersifat kategorikal?
2. Seberapa efektif kinerja model Naive Bayes dalam memprediksi penyakit paru-paru, yang diukur dengan metrik seperti akurasi, presisi, dan *recall*?
3. Bagaimana hasil dari confusion matrix dapat diinterpretasikan untuk memahami jenis-jenis kesalahan prediksi yang terjadi pada model?
   1. **Tujuan Penelitian**

Tujuan utama dari penelitian ini adalah membangun sistem prediksi penyakit paru-paru menggunakan pendekatan *machine learning* dengan algoritma Naive Bayes. Secara spesifik, penelitian ini bertujuan untuk:

1. Mengembangkan model prediksi yang mampu mengklasifikasikan kemungkinan seseorang menderita penyakit paru-paru.
2. Mengevaluasi kinerja model Naive Bayes secara komprehensif menggunakan akurasi, *classification report*, dan *confusion matrix*.
3. Menyimpulkan efektivitas Naive Bayes sebagai alat bantu pengambilan keputusan untuk deteksi dini.
   1. **Manfaat Penelitian**

penelitian ini diharapkan dapat memberikan manfaat yang signifikan, baik secara akademis maupun praktis.

* Penelitian ini akan berkontribusi pada literatur ilmiah dengan menyajikan studi kasus tentang penerapan algoritma Naive Bayes pada data kesehatan kategorikal. Hasilnya dapat menjadi *baseline* bagi penelitian lain yang ingin mengembangkan model serupa dengan algoritma yang lebih kompleks.
* Model prediksi ini dapat berfungsi sebagai alat bantu skrining awal yang efisien bagi tenaga medis, terutama di fasilitas kesehatan dengan sumber daya terbatas. Deteksi dini dapat mempercepat penanganan dan mengurangi biaya pengobatan.

**Metode penelitian**

#### **2.1 Desain Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komputasional berbasis supervised learning. Desain ini dipilih karena tujuannya adalah membangun dan menguji model prediktif berdasarkan data yang telah memiliki label (yaitu, status penyakit paru-paru). Proses penelitian akan dilakukan secara sistematis dalam alur kerja yang terstruktur:

1. Akuisisi Data: Mendapatkan data sekunder dari sumber publik.
2. Pra-pemrosesan Data: Menyiapkan data agar sesuai untuk pemodelan.
3. Pembagian Data: Memisahkan data menjadi set pelatihan dan pengujian.
4. Pemodelan: Melatih model menggunakan algoritma Categorical Naive Bayes.
5. Evaluasi Model: Mengukur dan menganalisis kinerja model.

#### **2.2 Pengumpulan Data**

Data yang digunakan adalah data sekunder dari dataset publik yang relevan dengan penyakit paru-paru. Dataset ini memiliki 30.000 baris data dan 11 kolom, yang terdiri dari 10 fitur prediktor dan 1 variabel target. Semua fitur, seperti usia, jenis kelamin, dan merokok bersifat kategorikal, menjadikannya sangat sesuai untuk algoritma yang dipilih. Penggunaan data publik menjamin replikasi penelitian dan menghemat sumber daya yang dibutuhkan untuk pengumpulan data primer.

#### **2.3 Pra-pemrosesan Data**

1. Pembersihan Data: Kolom yang tidak relevan, yaitu No dan Hasil, akan dihilangkan dari fitur prediktor.
2. Encoding Variabel Kategorikal: Karena model machine learning membutuhkan input numerik, variabel-variabel kategorikal akan diubah menggunakan Ordinal Encoding. Teknik ini memetakan setiap kategori unik dalam suatu kolom (misalnya, 'Tua', 'Muda') ke nilai bilangan bulat (misalnya, 1.0, 0.0), yang sesuai untuk data yang memiliki urutan.
3. Pemisahan Fitur dan Target: Data akan dipisahkan menjadi matriks fitur (X) dan vektor target (y) (hasil).

#### **2.4 Pembagian Data**

Data yang telah diproses akan dibagi menggunakan metode hold-out dengan rasio 80% data latih dan 20% data uji.

* Data Latih (80%): Digunakan untuk melatih model Naive Bayes, mengajarkannya pola hubungan antara fitur dan target.
* Data Uji (20%): Digunakan untuk mengukur kinerja model pada data yang belum pernah dilihat sebelumnya. Ini memastikan metrik evaluasi mencerminkan kemampuan generalisasi model secara akurat, bukan hanya pada data yang digunakan untuk pelatihan.

**METODE PEMBAHASAN**

* 1. **Pembahasan**

#### **3.1 Analisis Kinerja Model Categorical Naive Bayes**

Model yang dilatih menggunakan algoritma Categorical Naive Bayes pada data latih menunjukkan kinerja yang sangat baik saat diuji. Hasilnya terangkum dalam metrik evaluasi berikut:

1. Akurasi: Model mencapai akurasi sebesar 87,50%. Angka ini menunjukkan bahwa model mampu membuat prediksi yang benar untuk 87,5% dari total 6.000 data uji. Angka ini secara kuat mendukung efektivitas algoritma Naive Bayes dalam memprediksi kondisi penyakit paru-paru pada dataset ini, meskipun dengan asumsi independensi fitur yang sederhana.
2. Classification Report dan Keseimbangan Prediksi:

* Presisi untuk kelas 'Ya' (sakit) adalah 0.87, yang berarti dari semua prediksi model yang menyatakan seseorang sakit, 87% di antaranya benar-benar sakit. Ini menunjukkan model memiliki tingkat keandalan yang tinggi dalam memberikan prediksi positif.
* Recall untuk kelas 'Ya' adalah 0.87, yang mengindikasikan bahwa model berhasil mengidentifikasi 87% dari seluruh kasus penyakit yang ada dalam data uji. Dalam konteks medis, metrik ini sangat penting karena menunjukkan kemampuan model untuk tidak melewatkan kasus positif, sehingga meminimalkan risiko false negative (pasien sakit yang terlewat).
* Nilai F1-score yang seimbang antara presisi dan recall (masing-masing 0.87 dan 0.88 untuk kedua kelas) menunjukkan bahwa model tidak hanya akurat tetapi juga memiliki distribusi prediksi yang seimbang antara kelas 'Ya' dan 'Tidak'.

1. Analisis Confusion Matrix:

* True Positives (TP) dan True Negatives (TN): Model dengan benar mengidentifikasi 2.545 kasus sakit dan 2.705 kasus sehat, menunjukkan kemampuannya dalam mengenali kedua kondisi.
* False Negatives (FN): Terdapat 380 kasus di mana model salah memprediksi pasien yang sebenarnya sakit sebagai orang yang sehat. Ini adalah jenis kesalahan paling kritis dalam diagnosis medis karena pasien berisiko tidak mendapatkan penanganan.
* False Positives (FP): Model salah memprediksi 370 orang sehat sebagai penderita. Meskipun dapat menyebabkan kecemasan dan biaya tambahan, kesalahan ini relatif kurang berbahaya dibandingkan false negative.

**PENUTUP**

**4.1 Penutup Proposal**

Demikian proposal pengajuan judul mengenai **“**Analisis Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naive Bayes**”** telah kami susun.

Harapan kami, pengajuan judul ini dapat diterima dan mendapatkan dukungan untuk kelancaran pembahasan ke depannya. Kami mengucapkan terima kasih atas perhatian dan kesempatan yang telah diberikan.

Wassalamualikum warahmatullahi wabarakatuh.

KRESNA RISQI RAMADHAN

MUHAMMAD SHAADAM HAIDAR YUWONO