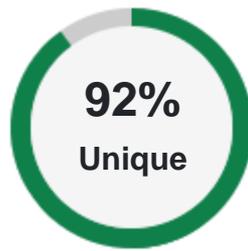


# PLAGIARISM SCAN REPORT



Exclude Url : None

## Content Checked For Plagiarism

World Cancer Research Fund mencatat Kanker payudara merupakan kanker paling umum pada wanita di seluruh dunia dengan menyumbang 25,4% dari total jumlah kasus baru yang didiagnosis pada tahun 2018. World Health Organization menyatakan bahwa kanker adalah sekelompok besar penyakit yang dapat dimulai di hampir semua organ atau jaringan tubuh ketika sel abnormal tumbuh tak terkendali, melampaui batas biasanya untuk menyerang bagian tubuh yang berdekatan dan/atau menyebar ke organ lain. Proses terakhir ini disebut metastasis dan merupakan penyebab utama kematian akibat kanker. Neoplasma dan tumor ganas adalah nama umum lainnya untuk kanker. Beban kanker terus meningkat secara global, memberikan tekanan fisik, emosional dan finansial yang luar biasa pada individu, keluarga komunitas dan sistem kesehatan. Banyak sistem kesehatan di negara berpenghasilan rendah dan menengah paling tidak siap untuk menangani beban ini, dan sejumlah besar pasien kanker secara global tidak memiliki akses ke diagnosis dan pengobatan berkualitas tepat waktu. Di negara-negara di mana sistem kesehatannya kuat, tingkat kelangsungan hidup berbagai jenis kanker meningkat berkat deteksi dini yang dapat diakses, pengobatan berkualitas, dan perawatan penyintas. Manusia dibanjiri dengan data di banyak bidang. Sayangnya, data berharga ini, yang menghabiskan jutaan perusahaan untuk mengumpulkan dan menyusun, mendekam di gudang dan repositori. Masalahnya adalah bahwa tidak tersedia cukup analis manusia terlatih yang terampil dalam menerjemahkan semua data menjadi pengetahuan (Larose, 2005). Penyakit kanker payudara dapat diprediksi dengan pengetahuan data mining. Data mining adalah proses menemukan pola yang berwawasan, menarik, dan baru, serta model deskriptif, dapat dipahami, dan prediktif dari data yang berskala besar (Ahmed et al., 2015). **Data mining dapat menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah data dalam jumlah besar yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik dan matematika** (Larose, 2005). Data mining diperkirakan akan menjadi "salah satu perkembangan paling revolusioner dalam dekade mendatang" (Larose, 2006). Seperti halnya dengan teknologi informasi baru, penambahan data yang dilakukan dengan sedikit pengetahuan sangat berbahaya ketika harus menerapkan model yang kuat berdasarkan kumpulan data yang besar. Misalnya, analisis yang dilakukan pada data yang tidak diproses dapat menyebabkan kesimpulan yang salah, atau analisis yang tidak tepat dapat diterapkan pada kumpulan data yang memerlukan pendekatan yang sama sekali berbeda, atau model dapat diturunkan yang dibangun di atas asumsi yang sepenuhnya spekulatif. Jika diterapkan, kesalahan dalam analisis ini dapat menyebabkan kegagalan yang sangat mahal (Larose, 2005). Salah satu peran data mining adalah prediksi. Prediksi mirip dengan klasifikasi dan estimasi, hanya saja untuk prediksi, hasilnya ada di masa depan (Larose, 2005). Penelitian terdahulu telah dilakukan (Dumitru, 2009) untuk mendiagnosis penyakit kanker payudara menggunakan Algoritma Naive Bayes. Dataset yang digunakan adalah Wisconsin Prognostic Breast Cancer. Hasil penelitian menunjukkan bahwa algoritma Naive Bayes memiliki performa yang setara dengan teknik pembelajaran mesin lainnya dengan upaya komputasi yang rendah dan kecepatan tinggi. Beberapa teknik pembelajaran mesin lain untuk memprediksi penyakit kanker payudara di antaranya Algoritma Gain Ratio (Aisyah & Sulisty, 2016), Jaringan Syaraf Tiruan (Wibisono et al., 2019), dan Support Vector Machine (Chazar & Widhiaputra, 2020). Penelitian ini akan membandingkan performa Algoritma Naive Bayes, Decision Tree dan Support Vector Machine untuk memprediksi penyakit kanker payudara Berdasarkan Gambar 1, Tahapan metode penelitian yang dilakukan yaitu: 1. Pengumpulan dataset. Dataset berupa data sekunder Breast Cancer Coimbra yang diambil dari UCI Repository. **Dataset terdiri dari 10 atribut yaitu Age, B MI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, MCP-1, dan Clasification.** Dataset Breast Cancer Coimbra terdiri dari 116 record. 2. Pengolahan Data Awal. Tidak ada atribut yang perlu dihapus pada dataset Breast Cancer Coimbra, dikarenakan semua atribut diperlukan dalam pengolahan data. 3. Metode yang diusulkan. Pada penelitian ini akan diusulkan Algoritma Naive Bayes, Algoritma Decision Tree dan Algoritma Support Vector Machine. 4. Eksperimen. Eksperimen dilakukan dengan mengolah data Breast Cancer Coimbra menggunakan

Algoritma Naive Bayes, Decision Tree dan Support Vector Machine. Kemudian hasil eksperimen dari ketiga Algoritma tersebut akan dikomparasi untuk mengetahui Algoritma mana yang memiliki performance terbaik 5. Evaluasi. Metode yang diusulkan akan diuji tingkat akurasi dengan Confusion Matrix dan Receiver Operating Characteristic Curve (ROC Curve) HASIL DAN PEMBAHASAN 1. Eksperimen Untuk mendapatkan hasil prediksi yang lebih akurat, penulis melakukan perbandingan dengan beberapa eksperimen sebagai berikut: a. Algoritma Naive Bayes Eksperimen pertama dilakukan dengan menggunakan Algoritma Naive Bayes. Hasil eksperimen diperoleh Accuracy sebesar 65,71% Berdasarkan Gambar 1, diketahui bahwa jumlah True Positive (TP) sebanyak 14 record yang diklasifikasikan sebagai Healthy Controls dan True Negative (TN) sebanyak 9 record yang diklasifikasikan sebagai Patients. Jumlah False Positive (FP) sebanyak 11 record yang diprediksi sebagai Healthy controls tetapi ternyata Patients dan 1 record False Negative(FN) yang diprediksi Patient tetapi ternyata Healthy controls b. Algoritma Decision Tree Eksperimen kedua dilakukan dengan Algoritma Decision Tree. Accuracy sebesar 60% didapat dari hasil eksperimen melalui Algoritma Decision Tree Berdasarkan Gambar 3, diketahui bahwa jumlah True Positive (TP) sebanyak 10 record yang diklasifikasikan sebagai Healthy Controls dan True Negative (TN) sebanyak 11 record yang diklasifikasikan sebagai Patients. Jumlah False Positive (FP) sebanyak 9 record yang diprediksi sebagai Healthy controls tetapi ternyata Patients dan 5 record False Negative(FN) yang diprediksi Patient tetapi ternyata Healthy controls c. Algoritma Support Vector Machine (SVM) Eksperimen ketiga dilakukan melalui Algoritma Support Vector Machine(SVM). Berdasarkan Gambar 4, diketahui bahwa nilai Accuracy yang dihasilkan sebesar 74,29%. Jumlah True Positive (TP) sebanyak 13 record yang diklasifikasikan sebagai Healthy Controls dan True Negative (TN) sebanyak 13 record yang diklasifikasikan sebagai Patients. Jumlah False Positive (FP) sebanyak 7 record yang diprediksi sebagai Healthy controls tetapi ternyata Patients dan 2 record False Negative(FN) yang diprediksi Patient tetapi ternyata Healthy controls 2. Evaluasi Hasil eksperimen Algoritma Naive Bayes, Algoritma Decision Tree dan Algoritma Support Vector Machine (SVM) akan dievaluasi dengan menggunakan Confusion Matrix dan Receiver Operating Characteristic Curve (ROC Curve). a. Confusion Matrix Analisa hasil komparasi dilakukan dengan membandingkan hasil pengujian Confusion Matrix dari ketiga Algoritma. Berdasarkan Tabel 1, Algoritma Support Vector Machine menghasilkan tingkat Accuracy tertinggi yaitu sebesar 74,29% dibandingkan dengan Algoritma Naive Bayes yang memperoleh Accuracy sebesar 65,71%. Sedangkan Algoritma Decision Tree menghasilkan Accuracy terendah bila dibandingkan dengan kedua Algoritma lainnya yaitu hanya sebesar 60%. Algoritma Naive Bayes menghasilkan nilai Sensitivity dan NPV tertinggi yaitu sebesar 93,33% dan 90%.

#### 4% Plagiarised

Konsep Data Mining & Aplikasinya Pada Bidang Informasi ... dan berguna bagi pemilik data (Han & Kamber, 2001); Proses menemukan korelasi, pola, dan tren baru yang bermakna dengan memilah-milah sejumlah besar data yang disimpan dalam repositori, menggunakan teknologi pengenalan pola serta teknik statistik ...

[http://lily.staff.gunadarma.ac.id/Downloads/files/66320/Minggu1\\_2.ppt](http://lily.staff.gunadarma.ac.id/Downloads/files/66320/Minggu1_2.ppt)

---

#### 4% Plagiarised

... Coimbra Data Set yang terdiri dari 116 record data dan 10 atribut yang terdiri dari Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, MCP.1, ...

<http://ejournal-binainsani.ac.id/index.php/IMBI/article/download/1191/1050>

---

