

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Pada saat ini banyak macam penyakit yang dapat melumpuhkan manusia seperti *bell's palsy*, *multiple sclerosis*, tumor otak, dan stroke. Stroke adalah suatu kondisi yang terjadi ketika pasokan darah ke suatu jaringan bagian otak tiba-tiba terganggu, sebagian sel-sel otak mengalami kematian akibat gangguan aliran darah karena sumbatan atau pecahnya pembuluh darah otak. Kematian sel otak mengakibatkan kelumpuhan saraf sehingga berdampak pada kehilangan fungsi motorik. Kelemahan otot yang mengakibatkan terjadinya masalah hambatan mobilitas fisik akan menyebabkan gangguan gerak yang sering muncul pada pasien penderita stroke (Tuntun, M, 2018). Gejala stroke meliputi sulit berjalan, berbicara, memahami dan kelumpuhan pada bagian tubuh. Stroke dapat berakibat fatal jika tidak langsung ditangani akibat terburuknya bisa terjadi kehilangan nyawa pada saat terserang stroke. Diperlukannya test kesehatan jika mendapatkan gejala awal stroke ataupun test kesehatan secara berkala untuk mengantisipasi serangan stroke secara tiba-tiba. Gejala stroke bisa menyerang siapa saja dan di mana saja tanpa melihat umur seseorang. Banyak faktor-faktor penyebab stroke lainnya, bisa saja memiliki riwayat hipertensi, jantung, diabetes, maupun obesitas (Wicaksana, Wati, 2017). Oleh karena itu, diperlukannya pengetahuan mengenai gejala dan faktor yang dapat mempengaruhi serangan stroke agar bisa diprediksi menggunakan *Machine Learning*.

Machine Learning merupakan teknologi kecerdasan buatan yang mampu mempelajari data yang ada dan melakukan tugas seperti mengolah data secara statistika, matematika dan *data mining* (Ahmad, Abu, 2017). *Machine Learning* banyak membantu diberbagai bidang seperti industri, pendidikan, dan kesehatan. Dibidang kesehatan kegunaannya untuk mendeteksi berdasarkan faktor resiko yang dialami seseorang. Dibutuhkannya data pasien yang ada untuk digunakan sebagai data training dalam pembentukan model, lalu model diuji dengan data testing yang dibandingkan dengan label aslinya untuk mendapatkan nilai yang akurat untuk mengklasifikasi penyakit sesuai dengan apa yang dipelajari dari data yang diperoleh. Data dalam

Machine Learning dilatih menggunakan algoritma untuk menemukan pola dan fitur tertentu dalam jumlah data yang besar, semakin bagus algoritma yang digunakan maka semakin bagus juga tingkat akurasi untuk memprediksi sistem. Beberapa algoritma yang bisa digunakan untuk membantu proses pembelajaran *Machine Learning* seperti *Logistic Regresion*, *KNN (K-Nearest Neighbors)*, *CNN (Convolutional Neutral Network)*, *Random Forest Clasifier*.

Random Forest adalah algoritma dalam *Machine Learning* yang digunakan untuk pengklasifikasian data set dalam jumlah besar melalui penggabungan tree dalam decision tree dengan cara training dataset (Goldstein, Poley & Briggs, 2011). Algoritma ini menggunakan *decision tree* atau pohon keputusan untuk melangsungkan proses seleksi, di mana *tree* akan dibagi secara rekursif berdasarkan data pada kelas yang sama. Dalam hal ini, penggunaan *tree* yang semakin banyak akan memengaruhi akurasi yang didapat menjadi lebih optimal. Penentuan klasifikasi dengan *Random Forest* dilakukan berdasarkan hasil *voting* dan *tree* yang terbentuk. *Decision tree* dibangun dan digabungkan untuk mendapatkan prediksi yang lebih stabil dan akurat. *Tree* yang dibangun disebut dengan hutan (*Forest*) adalah kumpulan decision tree di mana biasanya dilatih dengan metode bagging yaitu, kombinasi model pembelajaran untuk meningkatkan hasil keseluruhan. Algoritma *Random Forest* meningkatkan keacakan pada model sambil menumbuhkan tree. Untuk mencari fitur yang paling penting saat memisahkan sebuah node, *Random Forest* mencari fitur terbaik di antara subset fitur yang acak. Dengan cara tersebut menghasilkan keragaman yang luas dan umumnya menghasilkan model yang lebih baik.

Ada beberapa penelitian yang sudah dilakukan dengan menggunakan metode *Random Forest*. Salah satunya adalah yang dilakukan oleh Abdul Hafiz Muttaqien (2020). Dalam penelitiannya bertujuan untuk memperhatikan kesehatan kucing dengan memerhatikan penyakit kulit pada kucing. Oleh karena itu penelitian ini mengklasifikasi penyakit kucing khususnya pada kulit dengan metode *Random Forest* dengan klasifikasinya menggunakan tiga kelas yaitu *scabies*, *dermatophytosis*, dan sehat. Penelitian ini memiliki tingkat akurasi sebesar 96% dan penggunaan tree dalam jumlah besar yaitu 100. Penelitian lainnya dilakukan oleh Budi Susetyo, Indahwati, Aditya Ramandhan (2020). Dalam penelitiannya melakukan Pemodelan Klasifikasi *Random Forest* untuk Mengidentifikasi Faktor Penting dalam Meningkatkan Mutu Pendidikan.

Dari hasil penelitian, mendapatkan bahwa yang pertama berdasarkan evaluasi model, nilai akurasi klasifikasi dan rataan geometri tertinggi diperoleh dalam pemodelan klasifikasi *Random Forest* multikelas dengan optimalisasi parameter sebesar 88.17% dan 48.95%. Kedua, model klasifikasi *Random Forest* multikelas menghasilkan tingkat kepentingan peubah prediktor (butir-butir instrumen akreditasi).

Berdasarkan latar belakang dan penelitian sebelumnya implementasi *Machine Learning* untuk prediksi serangan stroke menggunakan *Random Forest* dapat memiliki tingkat akurasi yang baik dalam bidang kesehatan.

1.2. Perumusan Masalah

Adapun rumusan masalah dalam tugas akhir ini yaitu,

1. Bagaimana mengimplementasikan Algoritma *Random Forest* untuk memprediksi seseorang akan mengalami serangan stroke.
2. Bagaimana menerapkan *Machine Learning* khususnya *Random Forest* untuk memprediksi serangan stroke.

1.3. Tujuan Penelitian

Tujuan dalam penelitian tugas akhir ini yaitu,

1. Mengimplementasikan Algoritma *Random Forest* untuk memprediksi apakah seseorang akan mengalami serangan *stroke*.
2. Mengimplementasikan *Machine Learning* khususnya Algoritma *Random Forest* untuk memprediksi serangan *stroke*.

1.4. Batasan Masalah

Batasan masalah dibutuhkan untuk membuat tugas akhir lebih terarah, adapun batasan masalah dalam tugas akhir ini, diantaranya :

1. Dataset yang digunakan dalam pembuatan model *Machine Learning* berisi data riwayat penyakit *stroke* berformat .csv yang diambil dari *Kaggle*.
2. Data yang digunakan untuk pembuatan model yaitu Jenis Kelamin, Umur, Hipertensi, Riwayat Jantung, Status Menikah, Pekerjaan, Tempat Tinggal, Kadar Gula Darah, BMI, Status Merokok, Stroke.

3. Sistem dirancang untuk memprediksi serangan stroke dengan metode *Random Forest* menggunakan bahasa pemrograman *Python* pada *Jupyter Notebook*.

1.5. State of The Art

Penyusunan tugas akhir ini mengambil beberapa referensi penelitian sebelumnya termasuk jurnal-jurnal yang berhubungan dengan penelitian ini, diantaranya ;

Judul Jurnal	Pembahasan
<p>Judul Sistem Penentuan Rencana Tindak Lanjut Rehabilitasi Narkotika Menggunakan Metode <i>Random Forest</i> Pada BNN Kota Surabaya</p> <p>Peneliti Gesang Bekti Setyo Nugroho</p> <p>Lokasi Universitas Islam Negeri Sunan Ampel, Surabaya</p> <p>Tahun 2019</p>	<p><u>Hasil penelitian</u></p> <p>Pemanfaatan teknologi untuk menentukan rencana tindak lanjut rehabilitasi yang tepat dapat membantu pasien pulih dari pengaruh narkoba. Pembangunan system dilakukan dengan mengimplementasi <i>random forest</i> ke dalam system menggunakan <i>RestAPI</i> untuk menghubungkan <i>Machine Learning</i> dengan system. Hasil penelitian menggunakan 3 skenario yakni 100:100 dan 70:30, dan 60:40. Pada skenario 100:100 akurasi 61%, skenario 60:40 memiliki akurasi 30%.</p> <p><u>Alasan menjadi tinjauan penelitian :</u></p> <p>Penelitian ini sama-sama dibidang kesehatan dan menggunakan metode yang sama yaitu menggunakan <i>Random Forest</i>. Perbedaan dengan penelitian ini penggunaan dataset yang berbeda dan akurasi yang lebih baik dari penelitian ini.</p>
<p>Judul Pemodelan Klasifikasi <i>Random Forest</i> untuk Mengidentifikasi Faktor Penting dalam Meningkatkan Mutu Pendidikan</p>	<p><u>Hasil Pembahasan</u></p> <p>Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor penting dalam mempengaruhi mutu pendidikan jenjang</p>

<p>Peneliti</p> <p>Budi Susetyo, Indahwati, Aditya Ramandhan</p> <p>Lokasi</p> <p>Institut Pertanian Bogor</p> <p>Tahun</p> <p>2018</p>	<p>SMA/MA tahun 2018 berdasarkan nilai ujian nasional berbasis komputer (UNBK) dan hasil akreditasi (butir-butir instrumen akreditasi) dengan menerapkan pemodelan klasifikasi <i>Random Forest</i> multikelas. Pemodelan klasifikasi <i>Random Forest</i> pada penelitian ini dilakukan dengan mengklasifikasikan mutu pendidikan (nilai rata-rata 3 mata uji utama UNBK per satuan pendidikan) berdasarkan hasil akreditasi butir instrumen. Penanganan data tak seimbang dalam penelitian ini menggunakan metode pembobotan kelas dan <i>Synthetic Minority Oversampling Technique</i> (SMOTE) yang diakibatkan oleh pengkategorisasian nilai rata-rata UNBK ke dalam empat kategori yakni sangat baik, baik, cukup, dan kurang. Hasil penelitian menunjukkan bahwa, pertama, berdasarkan evaluasi model, nilai akurasi klasifikasi dan rataan geometri tertinggi diperoleh dalam pemodelan klasifikasi <i>Random Forest</i> multikelas dengan optimalisasi parameter sebesar 88.17% dan 48.95%. Kedua, model klasifikasi <i>Random Forest</i> multikelas menghasilkan tingkat kepentingan peubah prediktor (butir-butir instrumen akreditasi) yakni butir 69, 68, 62, 71, 67, 55, 56, 83, 45, 39, 36, 33, 64, 46, dan 14.</p> <p><u>Alasan menjadi tinjauan penelitian</u></p> <p>Penelitian ini menggunakan metode yang</p>
--	--

	sama dengan penelitian yang akan dibahas pada tugas akhir ini, hanya saja berbeda di bidang implementasinya.
<p>Judul Klasifikasi penyakit kucing menggunakan metode <i>Random Forest</i></p> <p>Peneliti Abdul Hafiz Muttaqien</p> <p>Lokasi Palembang</p> <p>Tahun 2020</p>	<p><u>Hasil Penelitian</u></p> <p>Penelitian ini bertujuan untuk memperhatikan kesehatan kucing dengan memerhatikan penyakit kulit pada kucing. Oleh karena itu penelitian ini mengklasifikasi penyakit kucing khususnya pada kulit dengan metode <i>Random Forest</i> dengan klasifikasinya menggunakan 3 kelas yaitu scabies, dermatophytosis, dan sehat. Penelitian ini memiliki tingkat akurasi sebesar 96% dan penggunaan tree dalam jumlah besar yaitu 100.</p> <p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Memiliki metode yang sama dan bidang yang sama, yaitu sama-sama dibidang kesehatan, bedanya hanya perbedaan kasusnya.</p>
<p>Judul Komparasi algoritma <i>Random Forest</i> dan <i>decision Tree</i> untuk memprediksi keberhasilan <i>Imunotherapy</i></p> <p>Peneliti Fandi Yulian Pamuji dan Viry puspaning ramadhan</p> <p>Lokasi Universitas Dian Nuswanto, Semarang</p>	<p><u>Hasil Pembahasan</u></p> <p>Penelitian ini membahas mengenai penyakit kulit dengan melihat pengobatannya menggunakan metode <i>Imunotherapy</i>. Berdasarkan hasil penelitian yang dilakukan pengujian pada <i>Decision Tree</i> memiliki akurasi 84,4% dan <i>Random Forest</i> 85,5% .</p> <p><u>Alasan menjadi tinjauan penelitian</u></p> <p>Penggunaan algoritma <i>Random Forest</i> lebih baik daripada <i>decision tree</i></p>

<p>Tahun 2021</p>	
<p>Judul Prediksi kemungkinan diabetes tahap awal menggunakan algoritma klasifikasi <i>Random Forest</i></p> <p>Peneliti Widya apriliah, ilham kurniawan, muhamman baydhowi, dan tri haryati</p> <p>Lokasi Universitas Bina Insani, Bekasi</p> <p>Tahun 2020</p>	<p><u>Hasil Pembahasan</u> Penelitian ini merancang model yang dapat memprakirakan kemungkinan terjadinya diabetes pada pasien dengan menggunakan teknik data mining yang mengkategorikan data untuk diprediksi dan dianalisis dengan menggunakan <i>Support Vector Machine</i>, <i>Naive Bayes</i>, dan <i>Random Forest</i>. Dataset yang digunakan dari rumah sakit di bangladesh yang bersumber dari UCI repository. Hasil dari penelitian di dapat akurasi <i>Random Forest</i> yang lebih tinggi dengan tingkat akurasi 97,88% dibandingkan dengan algoritma lainnya.</p> <p><u>Alasan menjadi tinjauan penelitian</u> Algoritma perbandingan untuk <i>Random Forest</i> akurasinya lebih tinggi daripada algoritma <i>Support Vector Machine</i> dan <i>Naive Bayes</i></p>

1.6. Sistematika Penulisan

Adapun sistematika penulisan dalam menyusun tugas akhir antara lain sebagai berikut :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, *state of the art* dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini merupakan penjelasan tentang teori yang berkaitan dengan penelitian.

BAB III METODE

Bab ini membahas mengenai analisis yang dilakukan dengan metode *Random Forest*.

BAB IV PEMBAHASAN

Bab ini berisi tentang implementasi dan hasil dari metode *Random Forest*.

BAB V KESIMPULAN DAN SARAN

Dalam bab ini membahas tentang kesimpulan dan saran yang sudah dilakukan dalam penelitian memprediksi serangan *stroke* dengan metode *Random Forest*.