

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini penulis akan membahas mengenai perhitungan data yang digunakan. Data tersebut akan di uji menggunakan metode Logistic Regression (LR) dan Support Vector Machine (SVM) pada tools Orange data mining.

1.1. Persiapan Data

Data yang digunakan dalam penelitian ini adalah *dataset public* yaitu data gambar rontgen paru-paru pasien tbc sebanyak 700 dan rontgen paru-paru normal sebanyak 3500 (<https://www.kaggle.com/tawsifurrahman/tuberculosis-tb-chest-xray-dataset>)

Dataset bisa dilihat pada tabel 4.1 di bawah ini :

Tabel 4.1 Sampel Dataset

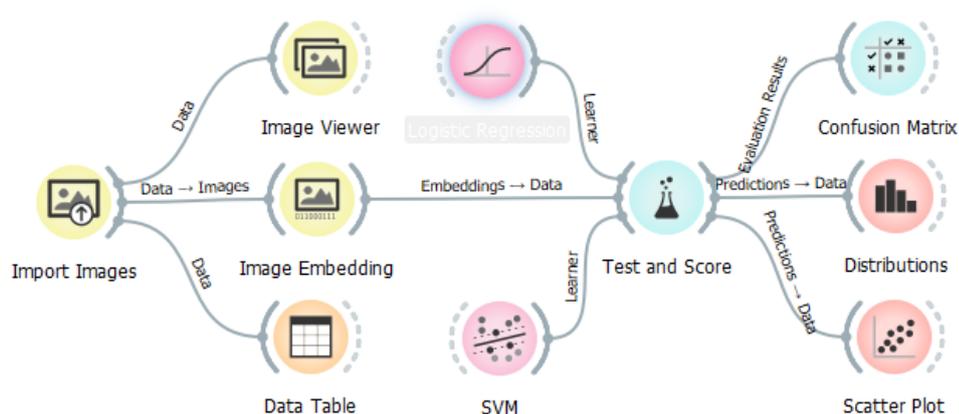
Kategori	Nama File	Lokasi File	Ukuran	Lebar	Tinggi
Normal	Normal-991	Normal\Normal-991.png	184631	512	512
Normal	Normal-992	Normal\Normal-992.png	159179	512	512
Normal	Normal-993	Normal\Normal-993.png	145241	512	512
Normal	Normal-994	Normal\Normal-994.png	183406	512	512
Normal	Normal-995	Normal\Normal-995.png	182273	512	512
Normal	Normal-996	Normal\Normal-996.png	147854	512	512
Normal	Normal-997	Normal\Normal-997.png	187440	512	512
Normal	Normal-998	Normal\Normal-998.png	159376	512	512
Normal	Normal-999	Normal\Normal-999.png	141081	512	512
Tuberculosis	Tuberculosis-1	Tuberculosis\Tuberculosis-1.png	217451	512	512
Tuberculosis	Tuberculosis-10	Tuberculosis\Tuberculosis-10.png	65932	512	512
Tuberculosis	Tuberculosis-100	Tuberculosis\Tuberculosis-100.png	69392	512	512
Tuberculosis	Tuberculosis-101	Tuberculosis\Tuberculosis-101.png	76134	512	512
Tuberculosis	Tuberculosis-102	Tuberculosis\Tuberculosis-102.png	68426	512	512
Tuberculosis	Tuberculosis-103	Tuberculosis\Tuberculosis-103.png	82659	512	512
Tuberculosis	Tuberculosis-104	Tuberculosis\Tuberculosis-104.png	72209	512	512
Tuberculosis	Tuberculosis-105	Tuberculosis\Tuberculosis-105.png	76887	512	512

jumlah atribut dataset sebanyak 6 atribut yang terdiri dari data karakter dan numertik yang merupakan hasil konversi data gambar rotgen paru-paru ke bentuk data tabel pada tools Orange data mining.

1.2. Pemodelan

Pemodelan dilakukan pada tools orange data mining dengan menggabungkan 4 tahap menjadi satu yaitu :

1. Image Processing : mengubah dataset gambar rontgen manusia menjadi data numerik
2. Learning Method : melakukan eksperimen terhadap uji dataset yang telah di konversi menggunakan 2 algoritma yaitu Logistic Regression (LR) dan Support Vector Machine (SVM)
3. Result : melihat hasil dari kinerja kedua algoritma dan membandingkan dengan hasil dari penelitian sebelumnya
4. Evaluation : mengevaluasi kinerja dari eksperimen yang sudah dilakukan apakah sudah mendapatkan hasil yang sesuai atau perlu dilakukan tahap eksperimen lanjutan



Gambar 4.1 Model Penelitian pada Orange

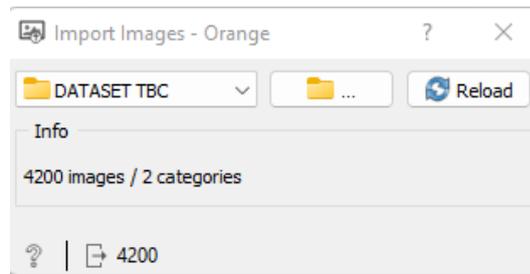
Gambar 4.1 adalah model keseluruhan yang akan digunakan pada penelitian ini dengan menggunakan tools orange data mining dan Cross Validation sebagai proses pengambilan sampel atau data, kemudian untuk algoritma yang digunakan yaitu Logistic Regression (LR) dan Support Vector Machine (SVM). Untuk hasil dari eksperimen akan dilihat pada tabel Confusion Matrix. Pengelompokan dari data yang akan diuji menggunakan fitur Distributions dan Scatter Plot.

1.2.1. Penelitian Menggunakan Algoritma Logistic Regression (LR)

Penerapan data pada Orange untuk penyakit Tuberkulosis menggunakan algoritma Logistic Regression (LR) ditunjukkan pada gambar di bawah ini :

1) Load Dataset

Langkah pertama yaitu memasukkan dataset gambar rotgen kedalam Orang dengan cara memilih **Image Analytics** > **Import Image** > “pilih dataset pada folder yang sudah disiapkan” > **Close**

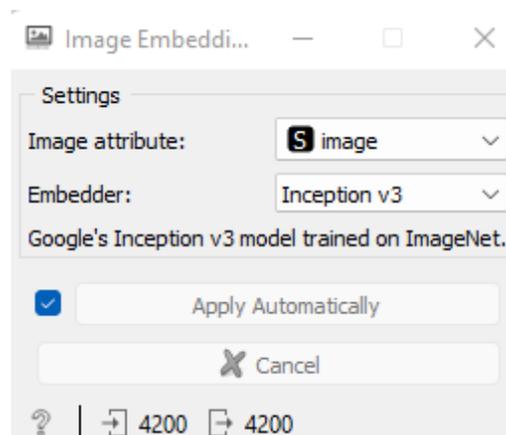


Gambar 4 2 Load Dataset

Gambar 4.2 adalah proses upload atau memasukkan dataset gambar rontgen. Setelah proses upload selesai terdapat keterangan bahwa dataset yang berhasil terupload sebanyak 4200 data yang terbagi dalam 2 katagori atau folder dataset.

2) Image Processing Dataset

Langkah kedua yaitu melakukan Image Processing yang bertujuan untuk merepresentasikan dataset gambar rontgen menjadi data tabel yang berisi data karakter dan numerik dengan cara masuk menambahkan menu yaitu : **Image Embedding**.

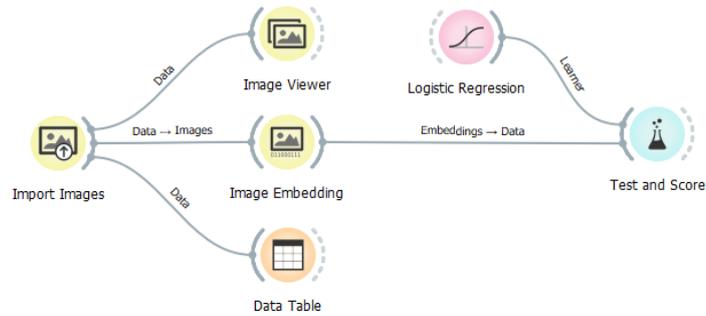


Gambar 4 3. Image Embedding Dataset

Pada gambar 4.3 proses Image Embedding dilakukan secara otomatis dan setelah proses selesai, selanjutnya dapat melihat dataset rontgen yang sebelumnya berupa data gambar menjadi data tabel yang berisi karakter dan numerik pada menu Data Table. Sedangkan data asli gambar masih bisa dilihat pada menu Image Viewer.

3) Learning Method

Pada tahap ini dataset yang sudah diubah menjadi bentuk tabel akan di uji menggunakan algoritma Logistic Regression dengan Teknik pengambilan sample yaitu Cross Validation. Model pada tahap ini dapat dilihat pada gambar 4.4.



Gambar 4 4 Model Learning method

4) Result

Hasil dari penerapan klasifikasi tuberculosis menggunakan metode Logistic Regression dengan Teknik pengambil data atau sampel Cross Validation antara lain : Accuracy sebesar 0985, Precision 0.984 dan Recall 0.985. hasil tersebut dapat dilihat pada gambar 4.5 di bawah ini :

Evaluation results for target (None, show average over classes) ▾					
Model	AUC	CA	F1	Precision	Recall
Logistic Regression	0.997	0.985	0.984	0.984	0.985

Gambar 4 5 Test and Score LR

Hasil akurasi pada penelitian sebelumnya dengan eksperimen penelitian ini dengan menggunakan algoritma Extreme Learning Machine untuk klasifikasi penyakit Tuberkulosis dengan data rotgen paru-paru manusia ditunjukkan pada tabel 4.2 dibawah ini:

Tabel 4.2 Perbandingan hasil akurasi

Penelitian	Tingkat Akurasi
Penelitian sebelumnya	93,3 %
Penelitian saat ini	98,5 %

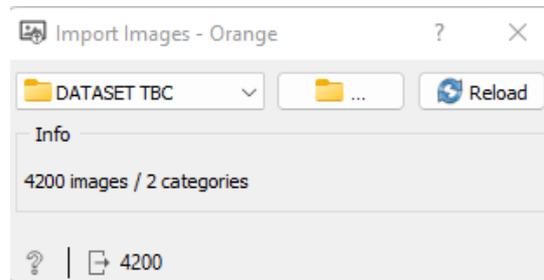
Eksperimen penelitian ini dengan pengujian K-Fold 2 pada cross validation menghasilkan tingkat akurasi paling tinggi dan tingkat akurasi lebih tinggi dari penelitian sebelumnya yang dilakukan oleh M. Andrizal, akan tetapi masih dapat ditingkatkan tingkat akurasinya.

1.2.2. Penelitian Menggunakan Algoritma Support Vector Machine (SVM)

Penerapan data pada Orange untuk penyakit Tuberkulosis menggunakan algoritma Support Vector Machine (SVM) ditunjukkan pada gambar di bawah ini :

1) Load Dataset

Langkah pertama yaitu memasukkan dataset gambar rotgen kedalam Orange dengan cara memilih **Image Analytics** > **Import Image** > “pilih dataset pada folder yang sudah disiapkan” > **Close**

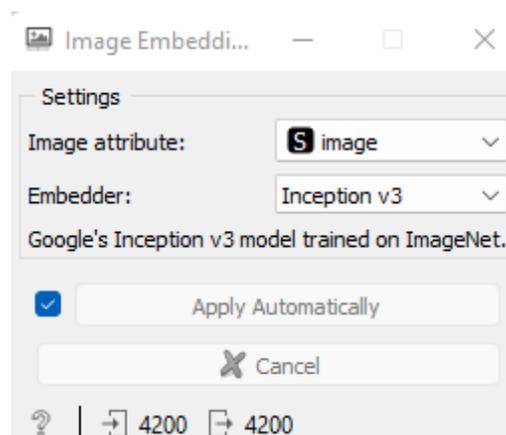


Gambar 4 6 Load Dataset

Gambar 4.6 adalah proses upload atau memasukkan dataset gambar rontgen. Setelah proses upload selesai terdapat keterangan bahwa dataset yang berhasil terupload sebanyak 4200 data yang terbagi dalam 2 katagori atau folder dataset.

2) Image Processing Dataset

Langkah kedua yaitu melakukan Image Processing yang bertujuan untuk merepresentasikan dataset gambar rotgen menjadi data tabel yang berisi data karakter dan numerik dengan cara masuk menambahkan menu yaitu : **Image Embedding**

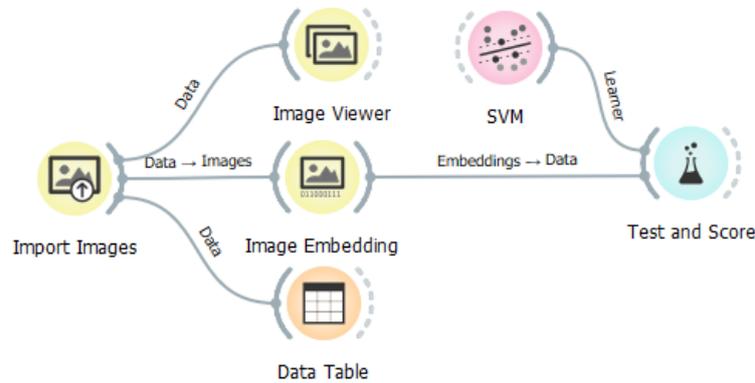


Gambar 4 7 Image Processing

Gambar 4.7 proses Image Embedding dilakukan secara otomatis dan setelah proses selesai, selanjutnya dapat melihat dataset rotgen yang sebelumnya berupa data gambar menjadi data tabel yang berisi karakter dan numerik pada menu Data Table. Sedangkan data asli gambar masih bisa dilihat pada menu Image Viewer.

5) Learning Method

Pada tahap ini dataset yang sudah diubah menjadi bentuk tabel akan di uji menggunakan algoritma Support Vector Machine (SVM) dengan Teknik pengambilan sample yaitu Cross Validation. Model pada tahap ini dapat dilihat pada gambar 4.8.



Gambar 4 8 Model Algoritma SVM

6) Result

Hasil dari penerapan klasifikasi tuberkulosis menggunakan metode Support Vector Machine (SVM) dengan Teknik pengambil data atau sampel Cross Validation antara lain : Accuracy sebesar 0.990, Precision 0.989 dan Recall 0.990. hasil tersebut dapat dilihat pada gambar 4.9 di bawah ini :

Evaluation results for target (None, show average over classes)					
Model	AUC	CA	F1	Precision	Recall
SVM	0.997	0.990	0.989	0.989	0.990

Gambar 4 9 Hasil Eksperimen SVM

Hasil akurasi pada algoritma sebelumnya dengan eksperimen algoritma kedua ini menggunakan algoritma Support Vector Machine untuk klasifikasi penyakit Tuberkulosis dengan data rotgen paru-paru manusia ditunjukkan pada tabel dibawah ini:

Tabel 4.3 Perbandingan Akurasi

Metode	Tingkat Akurasi
Algoritma ELM	93,3 %
Algoritma LR	98,5 %
Algoritma SVM	99,0 %

Eksperimen penelitian ini dengan pengujian K-Fold 2 pada cross validation menghasilkan tingkat akurasi paling tinggi. Pada eksperimen kedua ini mendapatkan hasil paling tinggi dari eksperimen pertama dan penelitian sebelumnya yaitu 99%.

1.3. Result

Merujuk dari proses yang sudah dilakukan, untuk mengetahui hasil dari kinerja klasifikasi penyakit tuberkulosis berdasarkan citra paru-paru manusia menggunakan metode *Logistic Regression* (LR) dan *Support Vector Machine* (SVM) yang digunakan pada penelitian ini yaitu dengan menambahkan menu Confusion Matrix pada orange data mining. Berikut ini adalah hasil dari tabel confusion matrix dan perhitungan manualnya

1.3.1. Tabel Confusion Matrix menggunakan Algoritma Logistic Regression

		Predicted		Σ
		Normal	Tuberculosis	
Actual	Normal	3480	20	3500
	Tuberculosis	45	655	700
Σ		3525	675	4200

Gambar 4.10 Tabel Confusion Matrix LR

Diketahui dari gambar 4.10 tabel Confusion Matrix diatas maka kita dapat melihat hasil sebagai berikut

True Positive (TP) : 3480

True Negative (TN) : 655

False Positive (FP) : 45

False Negatif (FN) : 20

1) Perhitungan Akurasi

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{3480 + 655}{3480 + 655 + 45 + 20} \\ &= \frac{4135}{4200} \\ &= 0,985 \end{aligned}$$

1.3.2. Tabel Confusion Matrix menggunakan Algoritma Support Vector Machine (SVM)

		Predicted		Σ
		Normal	Tuberculosis	
Actual	Normal	3483	17	3500
	Tuberculosis	27	673	700
Σ		3510	690	4200

Gambar 4 11 Tabel Confusion Matrix SVM

Diketahui dari gambar 4.11 tabel Confusion Matrix diatas maka kita dapat melihat hasil sebagai berikut

True Positive (TP) : 3483

True Negative (TN) : 673

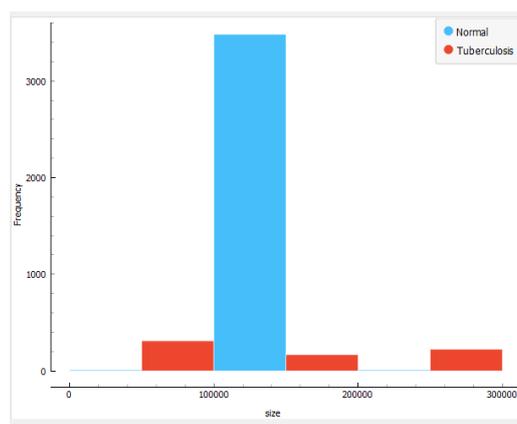
False Positive (FP) : 27

False Negatif (FN) : 17

2) Perhitungan Akurasi

$$\begin{aligned}
 Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 &= \frac{3483 + 673}{3483 + 673 + 27 + 17} \\
 &= \frac{4156}{4200} \\
 &= 0,990
 \end{aligned}$$

Selain Confusion Matrix untuk mengetahui kinerja dari eksperimen yang dilakukan kita juga dapat melihat Range atau kelompok berdasarkan ukuran dan kategori dari masing-masing data yang digunakan pada penelitian ini menggunakan fitur Grafik Distribution pada Orange Data Mining. Berikut adalah hasil sebaran data pada grafik Distribution :



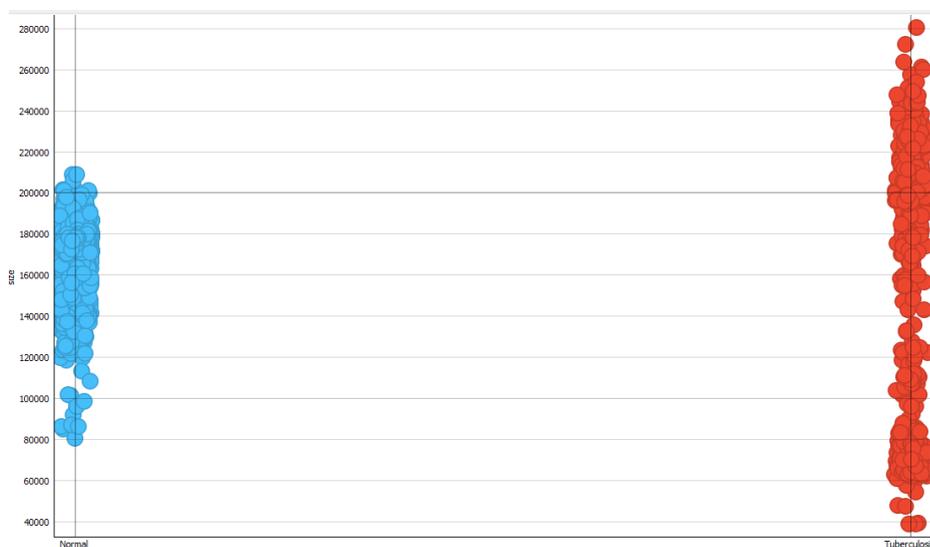
Gambar 4 12 Grafik Distribution

Gambar 4.12 menunjukkan range atau kelompok berdasar ukuran dari data dimana untuk frequency atau grafik Y adalah jumlah dari kelompok data dan size atau grafix Y adalah ukuran datanya. Dan untuk kategori Normal dan Tuberkulosis di tunjukan dengan warna biru dan merah dimana kategori paru-paru Normal di tunjukan dengan warna biru dan kategori paru-paru Tuberkulosis ditunjukan dengan warna merah.

Tabel 4.4 Kelompok ukuran dataset

Range	Kelompok		Jumlah
	Normal	Tuberkulosis	
Size < 100000	9	310	319
100000 ≤ Size < 200000	3482	167	3649
Size ≥ 200000	9	223	232
Total	3500	700	4200

Dari tabel 4.4 diatas dapat disimpulkan untuk rata-rata ukuran gambar rontgen paru-paru normal tidak kurang dari 100 Kb dan tidak lebih dari 200 Kb berbeda dengan gambar rontgen paru-paru Tuberkulosis yang cenderung kurang dari 100 Kilobyte dan lebih dari 200 Kilobyte. Kemudian untuk mengetahui range ukuran yang lebih spesifik di tambahkan juga grafik Scatter Plot pada Orange, berikut adalah hasil yang di dapat :



Gambar 4 13 Grafik Scatter Plot

Pada gambar 4.13 grafik Scatter Plot untuk membedakan antara data para-paru Normal dan data paru-paru Tuberkulosis sama dengan grafik Distribution yaitu pada warna. Untuk warna biru adalah kelompok dari data rontge paru-paru Normal dan warna merah adalah kelompok dari data rontgen paru-paru Tuberkulosis. Berikut ini adalah tabel yang dibuat berdasarkan hasil dari grafik Scatter Plot.

Tabel 4.5 Hasil pengelompokan data dari grafik Scatter Plot

Kategori	Range (Size)	Jumlah
Normal	80422 - 208993	3500
Tuberkulosis	38731 - 280622	700

Dari tabel 4.5 terlihat hasil pengelompokan data dari grafik Scatter Plot yang terbagi menjadi 2 kelompok berdasarkan kategori dan range dari kumpulan dataset gambar rontgen yang dapat disimpulkan pada kategori rontgen paru-paru normal range size menunjukkan hasil atau kelompok ukuran yang tidak terlalu jauh antar datanya berkisar 80422 – 208993 sedangkan untuk ketegori paru-paru TBC memiliki range atau ukuran yang lumayan jauh antar datanya dari mulai 38731 sampai 280622 dalam ukuran Kilobyte.

1.4. Evaluasi

Pemodelan yang telah dilakukan dengan eksperimen menggunakan metode *Logistic Regression (LR)* dan *Support Vector Machine (SVM)* menghasilkan hasil akurasi yang lebih baik dari penelitian sebelumnya menggunakan algoritma *Extreme Learning Machine (ELM)* yang menghasilkan akurasi 93,3%. Sementara untuk hasil pada penelitian ini dengan eksperimen pertama menggunakan algortima *Logistic Regression (LR)* dengan hasil akurasi 98,5% dan di experimen kedua menggunakan algoritma *Support Vector Machine (SVM)* mendapatkan akurasi yang paling tinggi dengan hasil akurasi 99%. Hasil kedua experimen menunjukkan bahwa klasifikasi penyakit tuberculosis berdasarkan citra paru-paru manusia menggunakan metode *Logistic Regression* dan *Support Vector Machine* telah berhasil meningkatkan akurasi sebesar 5,2% menggunakan metode *Logistic Regression* dan 5,7% menggunakan metode *Support Vector Machine*. Selain itu kinerja kedua algoritma pada penelitian ini juga menunjukkan performa yang sangat baik sekali termasuk dalam kategori *Excelent Clasification* berdasarkan hasil *AUC* yang memiliki nilai 99.7%.