

KLASIFIKASI GIGITAN ULAR MENGGUNAKAN LOCAL BINARY PATTERN DAN NAÏVE BAYES

Fathur Rohman¹, Adiwijaya², dan Dody Qori Utama³

¹Fakultas Informatika, Universitas Telkom, Bandung

²Fakultas Informatika, Universitas Telkom, Bandung

³Fakultas Informatika, Universitas Telkom, Bandung

E-mail: fathhrohman@students.telkomuniversity.ac.id

ABSTRACT

Cases of poisonous snake bites around the world are estimated to occur around 421,000 cases and 20,000 of them die every year. Identifying snake bite marks on victims will greatly help the medical team in handling victims of snake bites and will avoid fatal errors such as the death of the victim. This research will try to create a system that can classify snake bites images. The system has been built using the extraction method Local Binary Pattern (LBP) and Naive Bayes. Parameter r is a radius, while parameter P is the number of neighbor. The best result of this system has accuracy 83.33%, precision 1.00, recall 0.75, and F1 Score 0.86, parameter that used are $r = 1$ with $P = 8$ and $r = 3$ with $P = 16$. The dataset used has 20 data, the data divided into 14 training data and 6 testing data.

Keywords: Snake Bite, Local Binary Pattern, Naïve Bayes

ABSTRAK

Kasus gigitan ular berbisa di seluruh dunia diperkirakan terjadi sekitar 421.000 kasus dan 20.000 diantaranya menigggal dunia setiap tahunnya. Mengidentifikasi bekas gigitan ular pada korban akan sangat membantu tim medis dalam melakukan penanganan terhadap korban gigitan ular dan akan menghindari kesalahan yang fatal seperti kematian korban. Penelitian ini akan mencoba membuat sistem yang dapat mengklasifikasikan gambar gigitan ular. Sistem yang telah dibangun menggunakan metode fitur ekstrasi Local Binary Pattern (LBP) dan Naive Bayes. Parameter r merupakan radius, sedangkan parameter P merupakan jumlah neighbor. Hasil terbaik yang didapatkan sistem ini yaitu akurasi sebesar 83.33%, precision sebesar 1.00, recall sebesar 0.75, dan F1 Score sebesar 0.86, parameter yang digunakan yaitu $r = 1$ dengan $P = 8$ dan $r = 3$ dan $P = 16$. Dataset yang digunakan memiliki 20 data, data dibagi menjadi 14 data train dan 6 data test.

Kata kunci: Gigitan Ular, Local Binary Pattern, Naïve Bayes

1. PENDAHULUAN

Ular merupakan hewan yang bisa hidup dimana saja, kecuali daerah yang beku dan ketinggian yang tinggi, salah satu habitat ular adalah daerah yang beriklim tropis. Indonesia merupakan suatu negara yang beriklim tropis dan berada Asia Tenggara sehingga resiko terkena gigitan ular sangat tinggi. Kasus gigitan ular berbisa di seluruh dunia diperkirakan terjadi sekitar 421.000 kasus dan 20.000 diantaranya meninggal dunia setiap tahunnya [9].

Hal yang dilakukan oleh tim medis dalam melakukan penanganan terhadap korban gigitan ular, hal yang sangat penting dalam pengobatannya adalah mengidentifikasi spesies ular yang menggigit korban. Mengidentifikasi spesies ular yang menggigit korban akan sangat membantu tim medis dalam melakukan cara penanganan terhadap korban gigitan ular dan akan menghindari kesalahan dalam penanganan yang dapat mengakibatkan kesalahan yang fatal seperti kematian korban [1].

Umumnya untuk mengidentifikasi apakah seseorang yang digigit ular akan terkena racun adalah dengan melakukan tes medis atau dengan melihat jenis ular yang menggigit orang tersebut [4]. Cara lain mengidentifikasi bekas gigitan ular dapat mengetahui apakah gigitan tersebut berasal dari ular berbisa atau tidak [14]. Pada penelitian sebelumnya telah dilakukan oleh Nishoka, dia membuktikan bahwa bekas gigitan akan berguna dalam mendiagnosa gigitan ular di Brazil. Pada penelitian ini, hasil gigitan ular berbisa memiliki nilai sensitivitas sebesar 100 persen, *specificity* 56 persen, dan nilai prediksi positif sebesar 89 persen. Sedangkan hasil gigitan ular tidak berbisa memiliki nilai sensitivitas 56 persen, *specificity* 100 persen, dan nilai prediksi positif 100 persen [4].

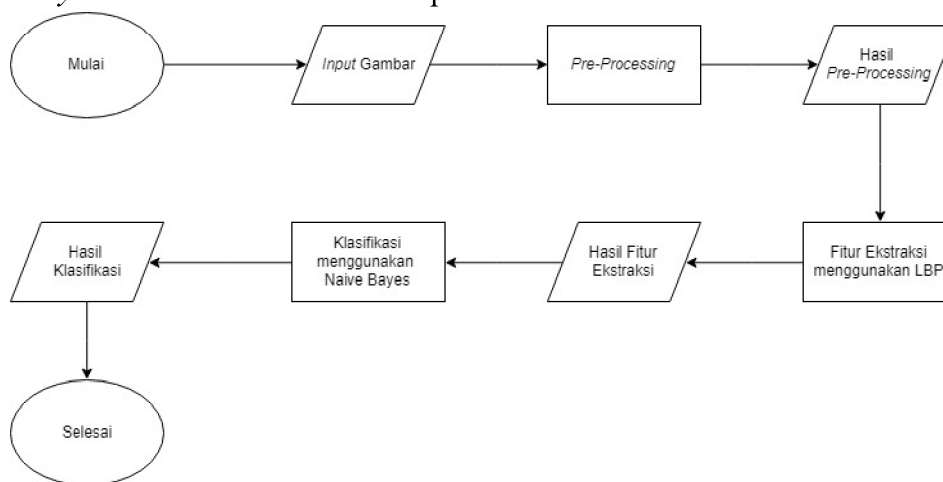
Pada penelitian ini telah dilakukan pengklasifikasian gambar gigitan ular menggunakan Matlab dan Python. Pengklasifikasian dalam penelitian ini, metode yang digunakan yaitu ekstraksi fitur menggunakan *Local Binary Patterns* (LBP) dan algoritma klasifikasi menggunakan *Naive Bayes*, sedangkan evaluasi hasilnya menggunakan *F-Measure*. Pemilihan ekstraksi fitur ini merujuk pada paper [11] yang menunjukkan bahwa LBP dapat meningkatkan tingkat *recognition* pada gambar. Pemilihan *classifier* ini merujuk pada paper [19] yang mengatakan bahwa *Naive Bayes* merupakan salah satu algoritma yang paling efisien dan efektif karena ketergantungan kuat setiap atributnya maupun memisahkan ketergantungan atributnya, *Naive Bayes* masih bekerja dengan baik. Beberapa penelitian yang menjelaskan tentang Bayesian juga ada di paper [2][3][12][13][15][16].

Batasan masalah pada penelitian ini yaitu data yang digunakan berupa 20 gambar bekas gigitan ular pada manusia dimana 13 diantaranya gambar bekas gigitan ular berbisa dan 7 lainnya gambar bekas gigitan ular tidak berbisa. Gambar ini didapatkan

dari *CommDIS Telkom University*. Semua gambar telah dipotong pixelnya untuk diambil dibagian bekas gigitannya saja dan ukurannya diubah menjadi 400x400 pixel. Fokus dari penelitian ini yaitu sistem dapat mengklasifikasikan gambar gigitan ular berbisa dan gambar gigitan ular tidak berbisa.

2. DASAR TEROI DAN PERANCANGAN

Sistem yang akan dibangun dalam penelitian ini adalah sistem yang dapat melakukan klasifikasi dan pengenalan menggunakan *Local Binary Patterns (LBP)* dan *Naive Bayes*. Gambar dibawah merupakan flowchart dari sistem.



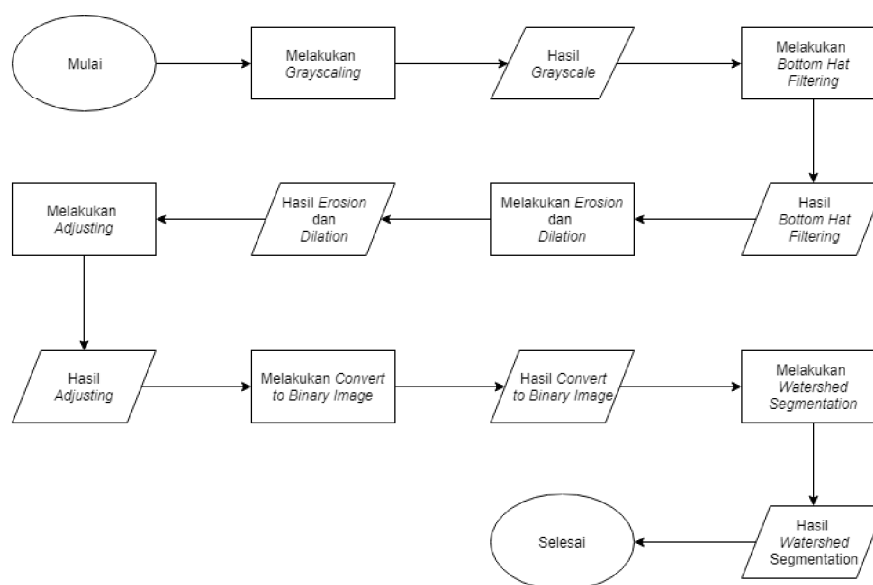
Gambar 2-1 Flowchart umum rancangan sistem

Langkah awal dalam perancangan sistem ini adalah menentukan dan membagi-bagi *dataset* berdasarkan jenisnya. Adapun data yang digunakan dalam penelitian ini berupa data gambar bekas gigitan ular yang berjumlah 20 gambar dan dapat diperoleh dari *CommDIS Telkom University* yang dapat diakses melalui situs <http://commdis.telkomuniversity.ac.id/dataset/Data%20Gigitan.rar>. Adapun pembagian data train dan data test dibagi menjadi 14 data train dan 6 data test.

Pada sistem perancangan klasifikasi gigitan ular, tahapannya dibagi menjadi beberapa bagian yaitu tahap *preprocessing*, tahap ekstraksi fitur, dan tahap klasifikasi.

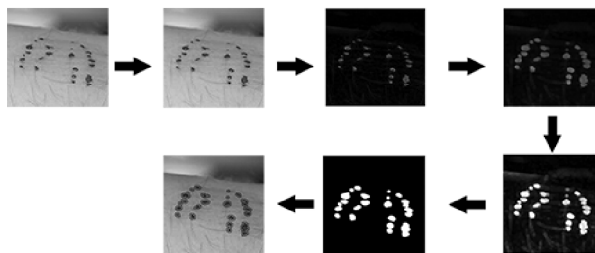
2.1 *Pre-Processing*

Dalam *pre-processing* hal yang akan dilakukan akan dibagi menjadi beberapa bagian yaitu mengubah gambar menjadi gambar grayscale, erotion dan dilation, Bottom Hat Filtering, Adjusting, Convert to Binary Image, dan melakukan proses segmentasi. Gambar dibawah merupakan flowchart dari bagian *pre-processing*:

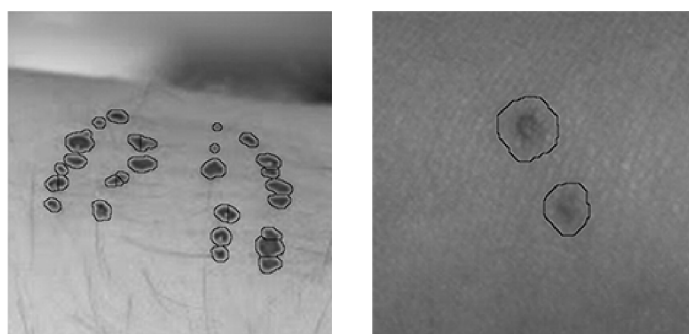


Gambar 2-2 Flowchart pre-processing

- Melakukan *grayscale* pada gambar.
- Melakukan *Bottom Hat Filtering*, yaitu menghitung *morphological* gambar kemudian melakukan *subtract* pada gambar asli.
- Melakukan *Erosion* atau *Dilation*. *Erosion* merupakan suatu cara untuk mereduksi nilai pixel. *Dilation* merupakan suatu cara untuk menambahkan nilai pixel.
- Melakukan *adjusting*, yaitu melakukan penyesuaian pixel.
- Melakukan *convert to binary image*, yaitu merubah gambar *grayscale* menjadi gambar biner.
- Melakukan *Watershed Segmentation*, yaitu proses pencarian pola. Metode ini digunakan untuk mendapatkan pola dari gambar gigitan ular. *Watershed Segmentation* menentukan pola dengan menggunakan perbedaan antar pixel dari gambar biner.



Gambar 2-3 Ilustrasi pada pre-processing



Gambar 2-4 Hasil dari Watershed Segmentation

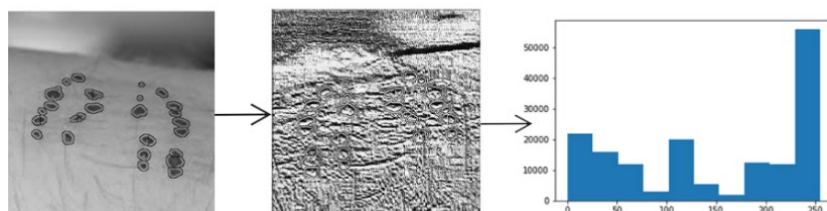
2.2 Ekstraksi Fitur

Dalam proses ekstraksi, metode yang digunakan adalah LBP. Metode ini dapat mengekstraksi local neighbor texture dari gambar grayscale. Hal pertama yang dilakukan LBP yaitu menghitung nilai biner dari setiap pixel gambar dan nilai local neighbor dari gambar grayscale. Lalu nilai biner tersebut digunakan oleh LBP untuk mengekstraksi gambar menjadi image feature [11]. LBP membandingkan setiap neighbor dengan cara sebagai berikut.

$$LBP_{pr} = \sum_{p=0}^{p-1} s(g_c - g_p) 2^p, s(x) \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

g_c merupakan nilai tengah pixel, g_p merupakan nilai dari *neighbor*, p merupakan jumlah *neighbor* yang berhubungan, dan r merupakan radius dari

neighbor [5]. Setelah LBP teridentifikasi, LBP akan menghasilkan karakteristik gambar dalam *histogram*, yang merepresentasikan sebagai matriks berbentuk $1 \times n$



Gambar 2-5 Proses dalam LBP

2.3 Klasifikasi

Hasil gambar yang sebelumnya sudah di modifikasi saat pre-processing dan ekstraksi fitur akan dilakukan training dan testing menggunakan algoritma *Naïve Bayes*. *Naïve Bayes* menghitung probability dengan rumus berikut [19].

$$p(c|E) = \frac{p(E|c)p(c)}{p(E)}$$

$p(c|E)$ merupakan posterior *probability*, dimana hipotesis c dalam data E . $p(E|c)$ merupakan probability data E memberikan hipotesis c benar. $p(c)$ merupakan *probability* dari hipotesis c benar. $p(E)$ merupakan *probability* data.

Naïve Bayes memproses data *discrete* dan *continuous* dengan cara yang berbeda. Formula dari *Naïve Bayes* yang digunakan untuk data *discrete* yaitu berdasarkan rumus 2, sedangkan data yang digunakan dalam penelitian ini yaitu data *continuous* [8]. Data *continuous* dapat diproses menggunakan formula dari *Gaussian Naïve Bayes*

$$p(X_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\pi\sigma_y^2}\right)$$

$p(X_i|y)$ merupakan *probability* X_i pada label y . σ_y merupakan standar deviasi. μ_y merupakan nilai *mean* dari y . parameter σ_y dan μ_y merupakan *maximum likelihood*. *Likelihood* merupakan *probability* dari suatu data memberikan prediksi benar.

2.4 Mengukur Performa

Untuk mengukur performa dari sistem yang telah dibangun, metode perhitungan yang akan digunakan yaitu *F-Measure*. Dalam melakukan perhitungan *F-Measure*, hal lain yang akan dihitung yaitu akurasi, *precision*, dan *recall* [18]. Berikut ini rumus dari *F1 Score*, *precision*, *recall* dan akurasi.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Dari rumus sebelumnya ada beberapa variabel yang akan dijelaskan. TP (*True Positive*) yaitu prediksi kelas positif dan kelas asli positif. TN (*True Negative*) yaitu prediksi kelas negatif dan kelas asli negatif. FP (*False Positive*) yaitu prediksi kelas positif dan kelas asli negatif, FN (*False Negative*) yaitu prediksi kelas negatif dan kelas asli positif.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengujian

Data yang digunakan pada penilitain ini berupa gambar bekas gigitan ular berbisa dan tidak berbisa. *Dataset* yang digunakan memiliki 2 kelas, yaitu 0 dan 1, dimana 0 berarti gambar bekas gigitan ular tidak berbisa dan 1 berarti gambar bekas gigitan ular berbisa. Dalam penghitungan *F-Measure*, kelas 1 merupakan nilai positif sedangkan kelas 0 merupakan nilai negatif.

Gambar yang masuk dalam preprocessing akan melakukan *Dilation* dan *Erosion*. Setiap gambar saat masuk pada preprocessing ada yang hanya menggunakan *Erosion* atau *Dilation* dan ada yang menggunakan keduanya. Penggunaan *Erosion* atau *Dilation* berdasarkan apakah jika dengan hanya menggunakan salah satunya dapat membersihkan gambar maka hanya akan digunakan salah satunya.

Atribut yang digunakan dalam *classifier* yaitu jumlah nilai pixel. Keluaran dari fitur ekstraksi yaitu *histogram*, *histogram* hasil fitur ekstraksi memiliki 2 nilai yaitu

nilai pixel dan jumlah nilai pixel. Nilai pixel digunakan sebagai atributnya sedangkan jumlah nilai pixel digunakan sebagai *value* dari atribut.

Parameter yang digunakan untuk *training* dan *testing* yaitu *r* dan *P*. Parameter *r* merupakan radius, dimana jarak pixel dari titik tengah dengan *neighbor*. Parameter *P* merupakan jumlah *neighbor* yang akan digunakan sebagai perbandingan untuk nilai tengah. Rentang nilai *P* dan *r* yang digunakan yaitu $P = 8$ sampai $P = 24$, dan $r = 1$ sampai $r = 4$. Rentang nilai *P* jika semakin besar, pengolahan akan semakin lama dan memuat memory semakin besar itu sebabnya hanya menggunakan hingga 24. Rentang nilai *r* hanya sampai 4 karena semakin besar tidak menunjukkan semakin baik hasilnya.

Dalam skenario pertama pembagian data dibagi menjadi 14 data *train* dan 6 data *test* dengan menggunakan *random state* sebesar 25. Dalam skenario kedua pembagian data menggunakan *cross validation 3 fold*, dimana semua data dibagi menjadi 3 *section*. Semua *section* akan menjadi data *test*, lalu nilai akhir akan didapatkan dari nilai rata-rata semua dari *section*. Nilai rata-rata pada skenario pertama yaitu akurasi sebesar 64.82%, *precision* sebesar 0.92, *recall* sebesar 0.56, dan *F1 Score* sebesar 0.67. Nilai rata-rata pada skenario kedua yaitu akurasi sebesar 62.96%, *precision* sebesar 0.69, *recall* sebesar 0.82, dan *F1 Score* sebesar 0.73.

Tabel 3-1 Hasil test menggunakan nilai *r* dan *P* pada skenario pertama

No	r	P	Precision	Recall	F1	Akurasi
1	1	8	1.00	0.75	0.86	83.33%
2	2	8	1.00	0.50	0.67	66.67%
3	2	16	1.00	0.50	0.67	66.67%
4	3	8	1.00	0.50	0.67	66.67%
5	3	16	1.00	0.75	0.86	83.33%
6	3	24	0.67	0.50	0.57	50.00%
7	4	8	1.00	0.25	0.40	50.00%
8	4	16	1.00	0.50	0.67	66.67%
9	4	24	0.60	0.75	0.67	50.00%

Tabel 3-2 Hasil test menggunakan nilai r dan P pada skenario kedua

No	r	P	Precision	Recall	F1	Akurasi
1	1	8	0.91	0.75	0.82	77.77%
2	2	8	0.62	0.66	0.64	55.55%
3	2	16	0.74	0.92	0.81	72.22%
4	3	8	0.64	0.75	0.68	55.55%
5	3	16	0.69	0.92	0.78	66.67%
6	3	24	0.62	0.83	0.71	55.55%
7	4	8	0.64	0.67	0.63	55.55%
8	4	16	0.69	0.92	0.78	66.66%
9	4	24	0.64	0.92	0.75	61.11%

3.2 Analisis Hasil Pengujian

Berdasarkan hasil testing dari skenario pertama, hasil terbaik yang didapatkan yaitu akurasi sebesar 83.33%, *precision* sebesar 1.00, *recall* sebesar 0.75, dan *f1 score* sebesar 0.86. Hasil tersebut menunjukkan bahwa dengan nilai *precision* sebesar 1.00, maka berhasil melabelkan semua gambar gigitan ular berbisa dengan benar seluruhnya. Berdasarkan hasil dari *recall* sebesar 0.75 menunjukkan bahwa, sistem akan gagal mengenali 1 gambar ular berbisa dari 4 gambar ular berbisa. Berdasarkan percobaan yang dilakukan gambar yang cukup sulit dikenali berbeda-beda, tergantung parameter yang digunakan.

Hasil dari skenario kedua, hasil terbaik yang didapatkan yaitu akurasi sebesar 77.77%, *precision* sebesar 0.91, *recall* sebesar 0.75, dan *f1 score* sebesar 0.82. Hasil tersebut menunjukkan bahwa dengan nilai *precision* sebesar 0.91, jika sistem melabelkan 100 gambar gigitan ular, 9 dari 100 gambar tersebut salah dalam pelabelan. Berdasarkan hasil dari *recall* sebesar 0.75 menunjukkan bahwa, sistem akan gagal mengenali 1 gambar ular berbisa dari 4 gambar ular berbisa.

Hasil testing dalam skenario pertama dan skenario kedua menunjukkan bahwa nilai P dan r sangat berpengaruh terhadap hasil klasifikasi. Hasil terbaik yang didapatkan pada skenario pertama yaitu pada saat menggunakan parameter $P = 8$ dengan $r = 1$ dan $P = 16$ dengan $r = 3$. Sedangkan hasil terbaik dari skenario kedua saat parameter yang digunakan yaitu $r = 1$ dan $P = 8$. Hal ini menunjukkan bahwa setiap perubahan pada parameter P dan r akan mempengaruhi hasil baik itu membuat hasil lebih baik ataupun lebih buruk.

Berdasarkan hasil dari kedua skenario, bisa kita ketahui bahwa perbedaan yang sangat berbeda dalam data *train* dan data *test* dapat mempengaruhi hasil dari klasifikasi. Dalam Penelitian ini dikarenakan jumlah data yang sedikit, setiap kali merubah data mana yang akan digunakan sebagai data *train* dan data *test* dapat memiliki hasil yang sangat berbeda.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, performa dari sistem yang telah dibangun mendapatkan hasil terbaik yaitu akurasi sebesar 83.33%, *precision* sebesar 1.00, *recall* sebesar 0.0.75, dan *f1 score* sebesar 0.86. Hasil terbaik tersebut menggunakan parameter $r = 1$ dengan $P = 8$ dan $r = 3$ dengan $P = 16$ menggunakan random state sebesar 25.

Sistem yang telah dibangun dapat melakukan pengklasifikasian gigitan ular dengan baik sangat bergantung kepada parameter yang digunakan saat fitur ekstraksi dan data mana yang digunakan untuk training dan testing. Kebergantungan data mana yang digunakan sebagai data *train* dan data *test* diketahui berdasarkan perbedaan hasil yang sangat berbeda pada skenario pertama dengan skenario kedua. Hal ini bisa disebabkan karena sedikitnya jumlah data yang dapat digunakan dan kualitas dari setiap datanya berbeda-beda.

Saran untuk penelitian yang menyerupai penelitian ini, perbanyak jumlah data dan kualitas data ditingkatkan. Berdasarkan penelitian ini kualitas dan kuantitas data sangat berpengaruh terhadap hasil penelitian. Selain itu bisa coba gunakan metode klasifikasi dan fitur ekstraksi yang berbeda, misalkan menggunakan *Princial Component Analysis* untuk fitur ekstraksi dan *Neural Network* untuk klasifikasi. Jika ingin menggunakan *Neural Network* sebaiknya perbanyak jumlah data terlebih dahulu supaya classifier berfungsi dengan baik.

DAFTAR PUSTAKA

- [15] E. Alirol, S. K. Sharma, H. S. Bawaskar, U. Kuch, and F. Chappuis. Snake bite in south asia: a review. *PLoS neglected tropical diseases*, 4(1):e603, 2010.
- [16] A. H. R. Z. Arifin, M. S. Mubarak, and A. Adiwijaya. Learning struktur bayesian networks menggunakan novel modified binary differential evolution pada klasifikasi data. In *Indonesia Symposium on Computing (IndoSC) 2016*, 2016..
- [17] R. A. Aziz, M. S. Mubarak, and A. Adiwijaya. Klasifikasi topik pada lirik lagu dengan metode multinomial naive bayes. In *Indonesia Symposium on Computing (IndoSC) 2016*, 2016.
- [18] B. S. Gold, R. C. Dart, and R. A. Barish. Bites of venomous snakes. *New England Journal of Medicine*, 347(5):347–356, 2002.
- [19] Z. Guo, L. Zhang, and D. Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE transactions on image processing*, 19(6):1657–1663, 2010.
- [20] N. Hernawati, D. Utama, et al. Image processing for snake indentification

- based on bite using local binary pattern and support vector machine method. In *Journal of Physics: Conference Series*, volume 1192, page 012007. IOP Publishing, 2019.
- [21] A. James. Snake classification from images. *PeerJ Preprints*, 5:e2867v1, 2017.
- [22] G. H. John and P. Langley. Estimating continuous distributions in bayesian classifiers. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 338–345. Morgan Kaufmann Publishers Inc., 1995.
- [23] A. Kasturiratne, A. R. Wickremasinghe, N. de Silva, N. K. Gunawardena, A. Pathmeswaran, R. Premaratna, L. Savioli, D. G. Lalloo, and H. J. de Silva. The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS medicine*, 5(11):e218, 2008.
- [24] W. Lou, X. Wang, F. Chen, Y. Chen, B. Jiang, and H. Zhang. Sequence based prediction of dna-binding proteins based on hybrid feature selection using random forest and gaussian naive bayes. *PloS one*, 9(1):e86703, 2014.
- [25] Y. Luo, C.-m. Wu, and Y. Zhang. Facial expression recognition based on fusion feature of pca and lbp with svm. *Optik-International Journal for Light and Electron Optics*, 124(17):2767–2770, 2013.
- [26] M. S. Mubarak, Adiwijaya, and M. D. Aldhi. Aspect-based sentiment analysis to review products using naïve bayes. In *AIP Conference Proceedings*, volume 1867, page 020060. AIP Publishing, 2017.
- [27] M. S. Mubarak, A. Adiwijaya, et al. Klasifikasi multi-label pada topik ayat qur’an terjemahan bahasa inggris menggunakan tree augmented naïve bayes (tan). *eProceedings of Engineering*, 5(1), 2018.
- [28] S. D. A. Nishioka, P. V. P. Silveira, and F. A. Bauab. Bite marks are useful for the differential diagnosis of snakebite in brazil. *Wilderness & environmental medicine*, 6(2):183–188, 1995.
- [29] R. A. Pane, M. S. Mubarak, N. S. Huda, et al. A multi-label classification on topics of quranic verses in english translation using multinomial naive bayes. In *2018 6th International Conference on Information and Communication Technology (ICoICT)*, pages 481–484. IEEE, 2018.
- [30] M. D. Purbolaksono, K. C. Widiastuti, M. S. Mubarak, F. A. Ma’ruf, et al. Implementation of mutual information and bayes theorem for classification microarray data. In *Journal of Physics: Conference Series*, volume 971, page 012011. IOP Publishing, 2018.

- [31] R. M. Putra, D. Q. Utama, et al. Snake bite classification using chain code and k nearest neighbour. In *Journal of Physics: Conference Series*, volume 1192, page 012015. IOP Publishing, 2019.
- [32] Y. Sasaki et al. The truth of the f-measure. *Teach Tutor mater*, 1(5):1–5, 2007.
- [33] H. Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004.