

Komparasi Decision Tree dan Random Forest untuk Klasifikasi Kapal pada Citra Satelit di Wilayah Maritim

Nama Kelompok : The Return of Sultan

Anggota Kelompok

1. Sultan Muzahidin (1806169591)
2. Fauhan Handay Pugar (1806255355)
3. Rif'at Ahdi Ramadhani (1806269543)

Deskripsi proyek

Proyek ini akan membahas mengenai komparasi dua algoritma yaitu *decision tree* dan *random forest* pada klasifikasi antara kapal dan bukan kapal pada citra satelit. Sebagai representasi dari wilayah maritim kami menggunakan dataset yang diperoleh dari *Planet satellite imagery* yang berada pada area San Fransisco Bay dan San Pedro Bay sekitar di California. Dataset terdiri dari 4000 citra chip RGB, dimana setiap citra berukuran 80x80 piksel. Citra chip yang didapat dari visual frame *PlanetScope* secara penuh yang ter-ortoreksi ke ukuran piksel dengan jarak 3 meter. Data citra yang digunakan adalah png dimana penamaan gambarnya memiliki format khusus yaitu {label}_{scene_id}_{longitude}_{latitude}.png. Dataset ini juga menyediakan format teks JSON dengan nama shipsnet.json yang terdiri dari data, label, scene_ids dan daftar lokasi. Label terdiri nilai 1 dan 0, angka 1 merepresentasikan kelas “kapal” dan “bukan kapal”. Kelas “kapal” (1000 citra) terdiri dari banyak bentuk kapal dengan berbagai ukuran dan bentuk. Sedangkan kelas “bukan kapal” (3000 citra) dimana sepertiganya adalah *random sampling* dari fitur tutupan lahan (*landcover features*) seperti air, vegetasi, tanah kosong, bangunan, dll. Sepertiga berikutnya adalah “kapal parsial” yang hanya berisi sebagian kapal sehingga tidak memenuhi bagian kapal secara penuh. Sepertiga terakhir adalah gambar yang salah label oleh pembelajaran mesin, biasanya disebabkan oleh piksel yang cerah atau lain-lainnya. *Scene_id* adalah pengenal unik dari *PlanetScope* untuk setiap visual dari citra chip yang di ekstrak. *Longitude_latitude* merupakan koordinat dari citra pada titik tengah gambar dimana setiap nilainya dipisah dengan underscore. Setiap nilai piksel dari citra RGB berukuran 80x80 disimpan dalam bentuk list yang terdiri dari nilai integer 19200. Data pertama terdiri dari 6400 nilai pada channel R, selanjutnya 6400 nilai pada channel G, dan terakhir 6400 nilai pada channel B.

Motivasi

Indonesia adalah negara kepulauan terbesar didunia dimana indonesia memiliki 17.499 pulau dari sabang sampai merauke. Sebagai negara kepulauan, Indonesia memiliki wilayah laut lebih luas dari daratan karena itu Indonesia di sebut sebagai negara maritim. Total luas wilayah Indonesia adalah 7,81 juta km² yang terdiri dari 3,25 juta km² lautan, 2,01 juta km² daratan dan 2,55 juta km² Zona Ekonomi Eksklusif (ZEE) [1]. Wilayah laut Indonesia memiliki potensi sumber kekayaan sangat besar yaitu sebagai pemasok ikan terbesar didunia. Potensi ekonomi sumber daya kelautan dan perikanan yang dapat dimanfaatkan untuk mendorong pertumbuhan ekonomi diperkirakan mencapai USD 82 miliar per tahun [2]. Dengan banyaknya potensi yang dimiliki Indonesia memerlukan perlindungan dan pengelolaan sumber daya perairan yang baik.

Salah satu permasalahan yang dihadapi negara maritim seperti Indonesia adalah praktik *illegal fishing*. *Illegal fishing* merupakan aktivitas pencurian ikan yang dilakukan oleh kapal asing yang melewati wilayah yurisdiksi suatu negara secara ilegal. Praktik ini jelas telah sangat merugikan negara setiap tahunnya. Menurut Menteri Kelautan dan Perikanan, Susi Pudjiastuti, kerugian negara telah mencapai Rp 240 triliun [3]. Selain itu, praktik *illegal fishing* juga menyebabkan kerugian lainnya, yakni kerusakan ekosistem laut.

Mengacu pada Undang-Undang Nomor 31 tahun 2004 dan Undang-Undang Nomor 45 Tahun 2009, Pemerintah Indonesia telah melakukan kebijakan penanganan terhadap praktik *illegal fishing* dengan cara menerapkan kebijakan penenggelaman kapal

yang melakukan tindak pidana tersebut. Hingga tahun 2018, sebanyak 488 kapal illegal fishing telah ditenggelamkan dalam penerapan kebijakan tersebut [4]. Untuk menerapkan kebijakan tersebut, tentunya diperlukan pengawasan secara intensif. Namun pengawasan secara intensif masih memiliki tantangan besar dalam hal usaha dan biaya yang diperlukan.

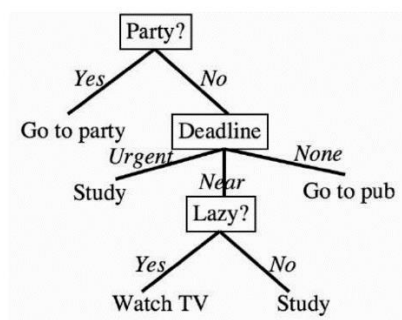
Peran *machine learning* diharapkan mampu memberikan solusi yang efisien dalam masalah pengawasan wilayah laut di Indonesia. Klasifikasi berperan untuk mengetahui pemetaan sebaran jumlah kapal di area tertentu. Identifikasi ini diperlukan agar pengawasan menjadi lebih terarah dan menjadi pondasi yang membantu dalam mengidentifikasi kapal yang terlibat dalam aktifitas *illegal fishing*. Oleh karena itu, pada penelitian ini kami membuat kerangka sistem machine learning untuk mengklasifikasikan objek kapal dan bukan kapal yang berada di suatu wilayah tertentu.

Metode

Metode yang kami gunakan adalah *decision tree* dan *random forest* pada klasifikasi kapal. Berikut adalah penjelasan mengenai kedua metode tersebut:

Decision Tree

Algoritma *decision tree* merupakan salah satu algoritma pembelajaran mesin yang memiliki struktur data dan performa komputasi yang baik. Secara umum, kebutuhan komputasi yang diperlukan dalam membuat struktur *tree* cukup rendah dan kebutuhan komputasi untuk keperluan klasifikasi yang dapat dinyatakan dengan notasi big-O ($O(\log N)$). Hal ini menjadi poin penting dalam pembelajaran mesin, terutama dalam aspek kebutuhan pengolahan dan penanganan data dalam jumlah besar dan kebutuhan pemerolehan hasil yang cepat. Selain itu, *decision tree* juga memiliki keunggulan lain yaitu kemudahan untuk memahami secara langsung dan transparan alur logika yang dimodelkan pada *decision tree* jika dibandingkan dengan metode '*black box*' seperti *neural network*. Hal inilah yang menjadi alasan klasifikasi dengan menggunakan *decision tree* menjadi cukup populer dalam pembelajaran mesin [5].



Gambar 1 Skema Decision Tree [5]

Ada beberapa jenis algoritma *decision tree*, namun semua varian yang ada memiliki prinsip yang sama yaitu proses pembuatan *tree* dilakukan secara rekursif dimulai dari bagian *root* dengan memilih fitur yang paling informatif berdasarkan nilai entropi tertentu. Nilai entropi dapat dihitung dengan persamaan:

$$Entropy(p) = - \sum_i p_i \log_2 p_i \quad (1)$$

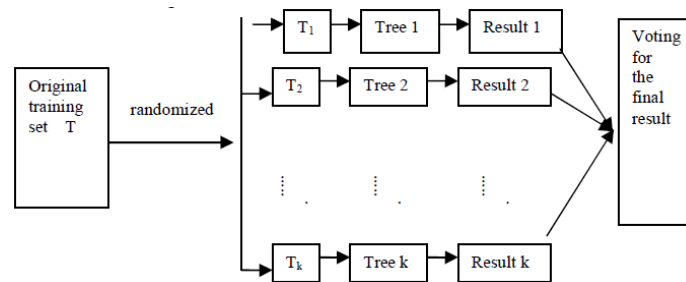
Random Forest

Random forest adalah kombinasi dari algoritma *machine learning*. Dimana kita mengkombinasikan serangkaian *tree classifier*, pada setiap *tree* dilakukan *voting* untuk mendapatkan kelas paling populer. Kemudian hasil dari kombinasi tersebut dilakukan *sorting*. *Random forest* memiliki akurasi yang tinggi, tahan terhadap noise dan juga tidak pernah mendapatkan *overfitting*.

Pada Breiman's RF model, setiap *tree* dilakukan *training* menggunakan *random variable*, dimana variabel acak pada *tree* di notasikan dengan Θ_k , antara dua variabel acak saling *independent and identically distributed*. Hasil klasifikasi $h(x, \Theta_k)$ dimana x adalah masukan sebuah vektor. Setelah k dijalankan kita akan mendapatkan urutan klasifikasi $\{h_1(x), h_2(x), \dots, h_k(x)\}$, dan kemudian setelah mendapatkan model sistem yang lebih dari satu klasifikasi, hasil akhirnya dilakukan *majority vote*, fungsi dari pilihannya adalah sebagai berikut :

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y) \quad (2)$$

$H(x)$ adalah kombinasi klasifikasi model, h_i adalah *single decision tree model*, Y adalah *output variable*, I adalah *indicator function*. Gambar jelasnya sebagai berikut [6]:



Gambar 2 Skema Random Forest

Rencana Eksperimen

Eksperimen yang dilakukan adalah dengan membuat proposal, *progress 1*, *progress 2*, dan final report serta poster. Dalam pembuatan proposal kami melakukan analisis dataset untuk menentukan metode apa yang akan digunakan. Progress 1, kami melakukan *pre-processing* dan ekstraksi fitur. Progress 2, kami melakukan penerapan algoritma *Decision Tree* dan *Random Forest*. Sebelum melakukan *final report* kami akan melakukan evaluasi akhir dan optimasi algoritma yang digunakan. Pengumpulan akhir dalam bentuk laporan, *source code* dan Poster di Github. *Timeline* pengerjaan dapat dilihat pada tabel yang berada pada subbab lampiran.

Referensi

- [1] BPS INDONESIA, “STATISTIK SUMBER DAYA LAUT DAN PESISIR.” 2018.
- [2] SKR INDONESIA, “Potensi Besar Perikanan Tangkap Indonesia.” 2016.
- [3] Detik, “Menteri Susi: Kerugian Akibat Illegal Fishing Rp 240 Triliun.” 2014.
- [4] Katadata, “Cek Data: Benarkah 488 Kapal Illegal Fishing Sudah Ditenggelamkan?” 2019.
- [5] S. Marsland, *Machine Learning: An Algorithmic Perspective, Second Edition*, 2nd ed. Chapman & Hall/CRC, 2014.
- [6] Y. Liu, Y. Wang, and J. Zhang, “New Machine Learning Algorithm: Random Forest,” 2012, pp. 246–252.

Lampiran

[illegible]