

BAB I

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi adalah proses mengelompokkan suatu objek ke dalam kelas yang memiliki karakteristik serupa (Vanda, 2015). Dalam klasifikasi, model prediktif dibangun untuk mengenali pola dari suatu objek berdasarkan label yang telah ditentukan. Proses pembentukan kelas pada objek tersebut biasanya menggunakan pendekatan statistika atau metode *machine learning*. Salah satu algoritma klasifikasi pada *machine learning* adalah *Support Vector Machine*.

Support Vector Machine (SVM) adalah algoritma klasifikasi yang cukup populer karena dapat menghasilkan tingkat akurasi yang tinggi dalam melakukan pengklasifikasian pada data (Damasela dkk., 2022). SVM bekerja dengan menentukan nilai *hyperplane* terbaik untuk memisahkan kelas yang berbeda secara iteratif untuk meminimalkan kesalahan. SVM juga memiliki kelebihan karena kemampuannya dalam menerapkan sistem linier pada data non-linier berdimensi tinggi dengan menggunakan fungsi kernel (Mahalle dkk., 2023). Namun, SVM sensitif terhadap pengaturan *hyperparameter* dan data latih yang digunakan.

Hyperparameter adalah parameter-parameter yang digunakan untuk mengontrol proses pelatihan pada model (Thamrin dkk., 2023). Parameter tersebut tidak dipelajari langsung dari data melainkan ditetapkan sebelum proses pelatihan dimulai. Parameter ini penting karena sangat berpengaruh terhadap keberhasilan performa model. Beberapa metode *hyperparameter*

tuning untuk menentukan *hyperparameter* diantaranya adalah *Grid Search* dan *Cross Validation* dan *Random Search Cross Validation*.

Grid Search Cross Validation merupakan gabungan antara metode *grid search* dan metode *cross validation*. Metode ini menentukan kombinasi *hyperparameter* dengan melakukan uji coba satu persatu kombinasi pada model dan melakukan validasi pada setiap kombinasinya (Fatmawati & Rifai, 2023). Sedangkan metode *Random Search Cross Validation* merupakan metode menentukan kombinasi *hyperparameter* dengan melakukan beberapa percobaan kombinasi saja dan melakukan validasi pada setiap kombinasinya (Putri dkk., 2023). Metode ini juga merupakan gabungan antara metode *random Search* dan metode *cross validation*.

Hasil dari metode *grid search* kurang efisien terhadap data yang berdimensi tinggi. Hal ini karena metode *grid search* menguji semua kombinasi *hyperparameter* yang ada. Namun, metode ini baik untuk menentukan *hyperparameter* karena kombinasi *hyperparameter* ini akan dicoba semuanya tanpa melewati satu parameter pun pada model. Sedangkan *random search* lebih efisien pada data berdimensi tinggi karena hanya melakukan uji coba secara acak dari kombinasi parameter. Akan tetapi, ada kemungkinan terlewatnya kombinasi *hyperparameter* yang optimal karena hanya melakukan beberapa percobaan saja. Perbedaan antara kedua metode tersebut dapat menghasilkan kinerja yang berbeda, ditambah dengan penggunaan metode *cross-validation*. Oleh karena itu, perlu dilakukan analisis terhadap kedua metode dalam mengklasifikasikan data.

Klasifikasi digunakan di berbagai bidang. Pada penelitian ini, dilakukan klasifikasi berkaitan dengan tata ruang wilayah. Penataan ruang suatu wilayah perlu direncanakan dengan sebaik mungkin karena sangat penting untuk keseimbangan antara kebutuhan manusia dengan lingkungan. Menurut Fandeli & Muhammad (2021), penataan ruang juga telah diakui sebagai salah satu komponen utama yang berperan penting dalam menentukan keberhasilan pembangunan suatu wilayah. Salah satu elemen penting dalam tata ruang yang baik adalah penyediaan ruang terbuka hijau

(RTH). Ruang Terbuka Hijau (*Green Openspaces*) adalah luasan lahan yang dipenuhi oleh tumbuhan dan memiliki kemanfaatan bagi lingkungan hidup, kesejahteraan masyarakat, dan keindahan.



Gambar 1.1: Contoh Ruang Terbuka Hijau (dpmpstsp.jogjakota.go.id)

Keberadaan Ruang Terbuka Hijau (RTH) sangatlah penting sesuai seperti yang diatur dalam Undang-Undang Nomor 26 Tahun 2007 tentang Penataan Ruang, yaitu Ruang Terbuka Hijau di perkotaan terdiri dari RTH Publik dan RTH Privat dengan proporsi RTH pada perkotaan sebesar minimal 30% yang terdiri dari 20% ruang terbuka hijau publik dan 10% ruang terbuka hijau privat. Selain itu, berdasarkan Permen PU No. 5/PRT/M/08 tentang Pedoman Penyediaan dan Pemanfaatan Ruang Terbuka Hijau di Kawasan Perkotaan, ketersediaan RTH di Kawasan perkotaan perlu disesuaikan dengan jumlah penduduk dan kebutuhan oksigen, sehingga luas kebutuhan RTH berdasarkan jumlah penduduk sebesar 20 m²/kapita (Devi & Santosa, 2022).

Akan tetapi berdasarkan data sensus dari Badan Pusat Statistik yang disajikan pada Gambar 1.2, dapat diketahui pertumbuhan jumlah penduduk di wilayah Indonesia meningkat sedangkan lahan yang ada tetap atau bahkan berkurang. Akibatnya, ketersediaan lahan ruang terbuka hijau di kota-kota besar semakin berkurang sejalan dengan bertambahnya populasi

di suatu wilayah. Hal ini menyebabkan banyak kota yang tidak memenuhi syarat kebijakan mengenai peraturan RTH yang diatur pada UU Nomor 26 Tahun 2007, dibuktikan dengan adanya pembangunan dan perubahan fungsi lahan menjadi padat pemukiman (Nurhafifa & Ruslan, 2023). Oleh karena itu, perlu dilakukan klasifikasi mengenai pemenuhan RTH pada wilayah-wilayah di Indonesia agar dapat diidentifikasi wilayah yang belum memenuhi peraturan tersebut.



Gambar 1.2: Tren Luas Wilayah dan Pertumbuhan Penduduk di Indonesia

Terdapat beberapa penelitian terdahulu yang telah dilakukan dan relevan dengan penelitian ini. Nadroh dkk, (2024) pada penelitiannya yang berjudul "Klasifikasi Status Gizi Balita Menggunakan Algoritma *Support Vector Machine* dengan Optimasi *Grid Search Cross-Validation*" telah mengimplementasikan *Grid Search CV* untuk klasifikasi status gizi balita dengan SVM, hasilnya terdapat peningkatan akurasi yang signifikan dari 79,29% menjadi 85,71%. Peningkatan sebesar 6,24% ini menunjukkan efektivitas *Grid Search CV* dalam mengoptimalkan parameter model dan meningkatkan kinerja klasifikasi. Penelitian Putri dkk (2023), juga telah menggunakan metode *Random Search CV* pada penelitiannya yang berjudul "Penerapan *Tuning Hyperparameter RandomSearchCV* pada *Adaptive Boosting* untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung" dengan memprediksi kelangsungan hidup pasien gagal jantung. Hasilnya model mendapatkan akurasi sebesar 85% dengan pencapaian ini

dikatakan relatif sangat baik. Berdasarkan hal tersebut, penelitian ini akan mengimplementasikan dua metode *hyperparameter tuning* yaitu *Grid Search CV* dan *Random Search CV* pada data yang berbeda yaitu ketersediaan ruang terbuka hijau. Pada penelitian ini juga menggabungkan metode *cross-validation* pada metode *grid search* dan *random search* untuk mengetahui penentu *hyperparameter* mana yang lebih baik dari kedua metode tersebut.

1.2 Batasan Masalah

Adapun batasan masalah pada penelitian ini sebagai berikut:

1. Dari total 514 kota dan kabupaten di Indonesia, hanya 286 kota dan kabupaten saja yang termasuk ke dalam data penelitian ini. Hal ini karena kota dan kabupaten yang terpilih hanya yang tercatat pada data ruang terbuka hijau publik tahun 2023 di Kementerian Lingkungan Hidup dan Kehutanan.
2. *Hyperparameter* yang diuji dalam penelitian ini terbatas pada parameter yang berkaitan dengan *kernel trick* dan seluruh parameter tersebut telah ditentukan sebelumnya dalam ruang pencarian (*search space*).

1.3 Perumusan Masalah

Berdasarkan latar belakang yang telah disebutkan sebelumnya, didapatkan rumusan masalah sebagai berikut:

1. Bagaimana proses mengklasifikasikan ketersediaan ruang terbuka hijau publik menggunakan SVM?
2. Bagaimana perbandingan kinerja dari parameter hasil *Grid Search CV* dan *Random Search CV* dalam klasifikasi ketersediaan ruang terbuka hijau dengan SVM?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah dan latar belakang yang telah diuraikan sebelumnya, penelitian ini mempunyai tujuan seperti berikut:

1. Mengklasifikasikan ketersediaan pemenuhan ruang terbuka hijau publik kota dan kabupaten di Indonesia.
2. Membandingkan metode *hyperparameter tuning* yang paling baik dalam menentukan parameter SVM untuk klasifikasi ruang terbuka hijau.

1.5 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini, yaitu:

1. Untuk peneliti
Mengetahui metode *hyperparameter tuning* mana yang paling baik pada SVM dan dapat menggunakan metode SVM dengan tingkat akurasi yang tinggi untuk mengklasifikasikan ketersediaan ruang terbuka hijau publik.
2. Untuk perguruan tinggi
Metode *hyperparameter tuning* yaitu *Grid Search CV* dan *Random Search CV* pada SVM dapat digunakan sebagai referensi pembelajaran. Selain itu, metode ini juga dapat menjadi sumber informasi bagi mahasiswa yang ingin meneliti penelitian sejenis.
3. Untuk lembaga
Metode SVM dengan penentuan metode *hyperparameter tuning* yang paling baik dapat menjadi pertimbangan dalam menyelesaikan permasalahan dan pengambilan keputusan.