BAB I

PENDAHULUAN

1.1. Latar Belakang

Digital marketing menjadi suatu solusi alternatif bagi para pelaku bisnis untuk menunjang kegiatan pemasaran produk maupun jasa agar produk dan jasa dapat terdistribusi secara lebih cepat dengan biaya yang relatif lebih terjangkau [1]. Kampanye digital marketing kini menjadi strategi utama untuk menjangkau konsumen secara luas dan tepat. Melalui penggunaan berbagai saluran digital seperti jejaring sosial, situs web, dan pemasaran melalui surel, perusahaan dapat mempromosikan produk atau layanan secara lebih personal dan interaktif. Indikator seperti jumlah klik, tingkat konversi, dan interaksi pengguna menjadi penting dalam menilai kesuksesan kampanye ini. Namun, di balik potensi besar kampanye digital marketing, tantangan utama yang dihadapi adalah bagaimana menganalisis dan mengklasifikasikan data yang dihasilkan untuk mengukur keberhasilannya.

Data yang dihasilkan dari kampanye digital sering kali memiliki dimensi besar dan kompleks, mencakup berbagai metrik seperti performa iklan, perilaku pengguna, demografi, dan waktu kampanye. Kerumitan data ini memerlukan pendekatan analisis yang lebih canggih agar dapat dimanfaatkan secara efektif untuk menilai kesuksesan kampanye dan memprediksi hasilnya.

Model klasifikasi menyimpulkan beberapa fungsi pemetaan yang valid dari data pelatihan dan memprediksi kategori dengan bantuan fungsi pemetaan yang dibuat untuk data baru [2]. Klasifikasi memiliki dua jenis model yaitu, klasifikasi biner yang memberikan dua kemungkinan hasil saja (misalnya dalam perkiraan cuaca), dan kedua, klasifikasi multi-label yang menghasilkan lebih dari dua kemungkinan hasil (contohnya dalam evaluasi prestasi mahasiswa dengan kategori sangat bagus, bagus, cukup, atau buruk). Contoh algoritma yang banyak digunakan dalam klasifikasi data adalah Random Forest, termasuk dalam kategori pembelajaran mesin berbasis pohon keputusan. Algoritma ini menerapkan teknik bagging yang membentuk sejumlah pohon keputusan dari sampel data yang tersedia. Setiap pohon dibuat secara independen selama proses pelatihan, tanpa dipengaruhi oleh pohon-pohon yang telah terbentuk sebelumnya. Prediksi akhir

ditetapkan melalui sistem voting mayoritas, di mana hasil yang paling banyak dipilih oleh seluruh pohon keputusan yang telah dibentuk akan menjadi keputusan akhir. Random Forest sangat andal dalam menangani dataset besar.

Berdasarkan penelitian terdahulu yang dilakukan oleh [3] menyimpulkan hasil pengukuran kinerja yang disajikan, CatBoost menunjukkan nilai akurasi 0,98 disusul dengan Random Forest dengan nilai akurasi 0,97, namun waktu yang dibutuhkan untuk eksekusi, CatBoost lebih lama 7 detik dibandingkan Random Forest. Selain itu [4] juga telah melakukan penelitian dengan judul “Klasifikasi informasi hoaks pada media sosial twitter menggunakan algoritma random forest berbasis particle swarm optimization Classification of hoax information on social media twitter using random forest algorithm based on particle swarm optimization” hasil penelitian tersebut menunjukkan optimasi PSO dengan Random Forest mengalami peningkatan sebesar 1% dibandingkan dengan tanpa optimasi dengan PSO.

Untuk mendapatkan hasil yang terbaik, peran yang sangat dibutuhkan adalah Particle Swarm Optimization atau yang biasa disebut PSO. Optimasi algoritma yang berpopulasi dan terinspirasi dari suatu perilaku sosial hewan berkelompok yaitu PSO, contohnya sama seperti kawanan burung atau kelompok ikan, terutama pada cara mereka bergerak saat mencari sumber makanan. Algoritma ini meniru bagaimana hewan-hewan tersebut berkoordinasi dan bergerak secara kolektif dalam mencari sumber makanan di lingkungan mereka. PSO bekerja dengan mengoptimalkan solusi melalui sekelompok partikel yang bergerak di dalam ruang pencarian, memperbarui posisinya berdasarkan pengalaman terbaik mereka dan seluruh kawanan. Algoritma ini telah terbukti efektif dalam mengoptimalkan hyperparameter pada model pembelajaran mesin, termasuk Random Forest.

Penelitian [5] parameter yang akan digunakan untuk penerapan di beberapa model Random Forest yaitu seperi *n_estimators*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*. Oleh karena itu pemilihan optimasi hyperparameter di Random Forest bertujuan untuk meningkatkan akurasi klasifikasi, terutama dalam analisis data kampanye digital marketing yang rumit. Dengan mengoptimalkan hyperparameter, PSO mampu menemukan konfigurasi model terbaik yang meningkatkan akurasi prediksi sekaligus meminimalkan permasalahan overfitting.

PSO juga dapat digunakan untuk seleksi fitur, memilih fitur yang paling relevan dalam dataset agar model berjalan lebih efisien.

Dalam konteks kampanye digital marketing, akurasi dalam memprediksi keberhasilan kampanye sangat penting karena hasil analisis ini memengaruhi keputusan bisnis. Melalui penerapan PSO, penelitian ini bertujuan untuk mengoptimalkan performa Random Forest dan mengklasifikasikan keberhasilan kampanye digital marketing, baik dari segi akurasi maupun efisiensi komputasi.

Penelitian ini memiliki tujuan agar dengan penggunaan Particle Swarm Optimization sebagai optimasi pada Random Forest dapat meningkatkan hasil evaluasi secara keseluruhan. Penelitian ini menggabungkan PSO dan Random Forest untuk mengatasi tantangan dalam analisis data kampanye digital marketing, dengan tujuan meningkatkan kualitas pengambilan keputusan berdasarkan hasil kampanye.

1.2. Rumusan Masalah

1. Seperti apa cara mengoptimalkan performa model Random Forest untuk klasifikasi data keberhasilan kampanye digital marketing dengan menggunakan algoritma Particle Swarm Optimization (PSO)?
2. Bagaimana Particle Swarm Optimization atau (PSO) dapat meningkatkan kinerja model Random Forest dengan fokus pada analisis perubahan metrik konfusi dan akurasi model?
3. Apa faktor kunci yang memengaruhi keputusan klasifikasi dalam kampanye digital marketing pada pohon keputusan model Random Forest?
4. Seberapa signifikan pengaruh optimasi hyperparameter menggunakan PSO terhadap performa model Random Forest dalam memprediksi keberhasilan kampanye digital marketing?
5. Bagaimana nilai F1-Score dan ROC-AUC dapat menggambarkan peningkatan kualitas model Random Forest setelah dioptimasi dan di klasifikasi data kampanye digital marketing dengan menggunakan metode algoritma Particle Swarm Optimization (PSO)?

1.3. Tujuan Penelitian

1. Sebagai implementasi algoritma Particle Swarm Optimization atau (PSO) untuk mengoptimalkan hyperparameter model Random Forest.
2. Meningkatkan akurasi prediksi dalam klasifikasi keberhasilan kampanye digital marketing dengan menggunakan kombinasi PSO dan Random Forest.
3. Membandingkan performa model Random Forest sebelum dan sesudah optimasi menggunakan PSO, baik dari segi akurasi maupun efisiensi waktu komputasi.

1.4. Manfaat Penelitian

1. Memberikan kontribusi bagi pengembangan ilmu pengetahuan di bidang optimasi algoritma dan machine learning, khususnya dalam bidang yang menerapkan algoritma PSO pada model klasifikasi seperti Random Forest.
2. Dapat menambah referensi tentang penerapan PSO dalam optimasi model machine learning untuk data digital marketing yang kompleks.
3. Membantu perusahaan atau pengelola kampanye digital marketing dalam memanfaatkan model yang lebih akurat untuk mengukur dan memprediksi keberhasilan kampanye mereka.
4. Memperoleh metode optimasi yang efisien untuk meningkatkan performa model klasifikasi, sehingga proses pengambilan keputusan dalam pemasaran digital dapat lebih tepat dan berbasis data.

1.5. Batasan Masalah

1. Dengan menggunakan dataset kampanye digital marketing dari platform Kaggle dengan karakteristik spesifik, dibatasi pada 8000 data. Link yang diakses : <https://www.kaggle.com/datasets/arpit2712/digital-marketing-company/data>
2. Optimasi Particle Swarm Optimization (PSO) difokuskan pada pengoptimalan hyperparameter berupa *n_estimators*, *max_depth*, *min_samples-split*, dan *min_samples_leaf*.

3. Pada model yang akan dilakukan evaluasi peningkatan yaitu menggunakan suatu metrik akurasi, presisi, recall, F1-score, dan ROC-AUC berguna untuk mengukur performa model sebelum dan sesudah optimasi.
4. Perbandingan model dilakukan antara Random Forest dan Random Forest dengan optimasi PSO, dengan pengujian pada berbagai rasio pembagian data latih (50%, 60%, 70%, 80%, 90%).