

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

"Sinau Bareng" adalah program pendidikan oleh Dinas Pendidikan Kota Surabaya untuk meningkatkan kualitas pendidikan di Kota Surabaya. Program ini bertujuan untuk meningkatkan proses belajar-mengajar dan menyediakan dukungan fasilitas serta pendampingan bagi anak-anak usia sekolah. Dilaksanakan tiga kali seminggu dari pukul 18.00 hingga 20.00 di 220 Balai RW yang tersebar pada 128 kelurahan dari 31 kecamatan di Surabaya, "Sinau Bareng" melibatkan 2.240 mahasiswa dan 6.593 guru. Sejak diresmikan pada 2 September 2023 oleh Wali Kota Surabaya Eri Cahyadi, program ini mendukung upaya Pemerintah Kota Surabaya dalam membangun Kota Layak Anak (KLA) [1].

Menurut pernyataan Bu Atiko, selaku Ketua Tim Kerja Kurikulum Sekolah Menengah di Dinas Pendidikan Kota Surabaya dan penanggung jawab kegiatan "Sinau Bareng," beberapa lokasi di pusat kota menghadapi kendala serius. Kendala tersebut adalah ketidakhadiran tutor, baik dari kalangan guru maupun mahasiswa, pada saat pertemuan. Ketidakhadiran ini menyebabkan kegiatan tidak dapat berjalan sebagaimana mestinya. Tanpa kehadiran tutor, tidak ada pihak yang berperan sebagai pengajar. Selain itu, terdapat juga lokasi di mana tidak ada murid yang hadir. Akibatnya, tutor yang sudah datang ke lokasi terpaksa kembali tanpa melaksanakan tugasnya karena tidak ada murid yang bisa diajar. Kendala-kendala ini berdampak signifikan terhadap keberlanjutan kegiatan "Sinau Bareng". Ketidakhadiran tutor dan murid akan menurunkan efektivitas kegiatan dan menghambat pencapaian tujuan pembelajaran. Oleh karena itu, diperlukan prediksi keberlanjutan kegiatan di suatu lokasi berdasarkan absensi kehadiran. Data absensi kehadiran ini diperoleh dari para peserta (tutor dan murid) yang mengisi web absensi Sinau Bareng pada link <https://dispendik.surabaya.go.id/sinau-bareng/>. Dengan begitu, penanggung jawab kegiatan dapat mengambil langkah-langkah yang tepat agar kegiatan dapat berjalan lebih baik di masa depan.

Untuk menentukan apakah suatu lokasi diklasifikasikan sebagai berlanjut atau tidak, kriteria yang digunakan mengacu pada kriteria yang ditetapkan oleh Bu

Atiko selaku penanggung jawab kegiatan “Sinau Bareng”. Bu Atiko saat ini menjabat sebagai Ketua Tim Kerja Kurikulum Sekolah Menengah di Dinas Pendidikan Kota Surabaya. Kriteria tersebut mencakup tiga aspek utama. Pertama, jumlah murid yang hadir di setiap pertemuan harus minimal 5 orang. Kedua, jumlah tutor, baik guru maupun mahasiswa, minimal 1 orang. Ketiga, terdapat perbandingan ideal antara tutor dan murid, yaitu minimal 1 tutor untuk 5 – 10 murid. Jika suatu lokasi tidak memenuhi salah satu dari ketiga kriteria tersebut, maka kegiatan di lokasi tersebut dianggap tidak berlanjut. Sebaliknya, jika semua kriteria terpenuhi, lokasi tersebut dikategorikan sebagai berlanjut. Kriteria keberlanjutan yang ditetapkan oleh Bu Atiko menjadi acuan yang jelas untuk menilai keberlanjutan kegiatan di suatu lokasi.

Penilaian secara manual untuk berbagai lokasi cenderung tidak efisien dan rentan terhadap kesalahan. Penggunaan *excel* untuk mengolah beberapa kriteria sekaligus cenderung rumit karena memerlukan rumus kompleks untuk menggabungkan semua kriteria yang ada dalam satu proses. *Excel* tidak dirancang untuk mengelola kriteria secara otomatis, tidak seperti *data mining* yang dapat memproses data secara efisien tanpa pengaturan manual, menghasilkan analisis yang lebih akurat dan konsisten. Di sisi lain, sistem pendukung keputusan (DSS) dapat membantu dalam proses pengambilan keputusan, namun tidak secara otomatis menggabungkan semua kriteria dalam satu proses. *Data mining* menawarkan solusi yang lebih unggul dan efisien dalam hal ini. *Data mining* dapat membangun model prediktif yang secara otomatis memproses semua kriteria keberlanjutan. *Data mining* menyediakan metode analisis yang lebih canggih, seperti algoritma *supervised learning*. Metode ini mampu membuat prediksi keberlanjutan program dengan lebih akurat. *Excel* hanya mendukung analisis sederhana, seperti perhitungan statistik dasar. Visualisasi yang disediakan juga terbatas, misalnya grafik batang atau *pie chart*. Sebaliknya, *data mining* memungkinkan analisis lebih mendalam. Salah satu keunggulannya adalah menemukan pola tersembunyi dalam data. Dengan data mining, evaluasi menjadi lebih akurat, cepat, dan konsisten. Modelnya juga dirancang sesuai dengan aturan yang sudah ditetapkan.

Dalam upaya menentukan klasifikasi keberlanjutan kegiatan "Sinau Bareng" di setiap lokasi, klasifikasi akan menggunakan pendekatan *machine learning*. *Machine learning* adalah teknik yang menggunakan pendekatan matematis untuk menarik kesimpulan dari data. Teknik ini menjadi populer karena memungkinkan konstruksi model inferensi dilakukan secara otomatis [2]. *Machine learning* dipilih untuk klasifikasi keberlanjutan kegiatan "Sinau Bareng" karena kemampuannya dalam menganalisis data secara akurat dan otomatis. Model ini akan membantu pengelola mengambil langkah-langkah proaktif untuk meningkatkan kehadiran. Dengan demikian, kegiatan dapat berjalan lebih baik di masa depan. Untuk implementasinya, penelitian ini akan mencari algoritma yang paling optimal di antara *Random Forest*, *Support Vector Machine (SVM)*, *Naïve Bayes* kemudian divisualisasikan dalam bentuk aplikasi berbasis web. Perbandingan tiga algoritma yaitu *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* dilakukan untuk mencari algoritma yang paling optimal dalam mengklasifikasikan keberlanjutan kegiatan. Masing-masing algoritma memiliki keunggulan dalam menghadapi karakteristik data yang berbeda. Dengan demikian, pemilihan algoritma yang tepat akan menghasilkan model klasifikasi yang lebih akurat dan efektif dalam menilai keberlanjutan kegiatan.

Hasil pengujian pada "Penerapan Algoritma Klasifikasi *Random Forest* untuk Penentuan Kelayakan Pemberian Kredit di Koperasi Mitra Sejahtera" menunjukkan bahwa algoritma klasifikasi *Random Forest* mampu menganalisis kredit bermasalah dan debitur yang tidak bermasalah dengan akurasi sebesar 87,88%. Selain itu, pohon keputusan juga berhasil meningkatkan akurasi dalam menganalisis kelayakan kredit yang diajukan oleh calon debitur [3]. Keunggulan algoritma ini dalam masalah klasifikasi dan efektivitasnya dalam data mining, seperti yang dijelaskan oleh [4], menjadikan *Random Forest* pilihan yang tepat dalam masalah klasifikasi, terbukti dengan akurasi yang dicapai dalam mengidentifikasi kelayakan kredit di Koperasi Mitra Sejahtera.

Hasil pengujian penerapan algoritma *Support Vector Machine (SVM)* menunjukkan kinerja yang sangat baik, dengan tingkat akurasi yang tinggi dalam berbagai kasus. Sebagai contoh, penelitian yang berjudul "Perbandingan Kinerja Klasifikasi Penyakit Ginjal Menggunakan Algoritma *Support Vector Machine*

(SVM) dan *Decision Tree (DT)*" menunjukkan bahwa *Support Vector Machine* memiliki kinerja lebih unggul dibandingkan dengan *Decision Tree*, dengan hasil akurasi 97.5%, presisi 0.98, *recall* 0.96, dan *f1-score* 0.97 [5]. *Support Vector Machine* berhasil diterapkan dalam banyak kasus dengan tingkat akurasi yang tinggi berkat konsep yang matang, serta kemampuan algoritma ini dalam memetakan data *non-linier* ke ruang fitur berdimensi lebih tinggi, yang memungkinkan pemisahan data secara *linier* dengan lebih kuat dan efektif [6]. Oleh karena itu, *Support Vector Machine* dijadikan sebagai salah satu pilihan algoritma yang relevan untuk mengklasifikasi keberlanjutan kegiatan "Sinau Bareng" berdasarkan kehadiran peserta (tutor dan murid), mengingat kinerja dan akurasi tinggi yang telah terbukti.

Pada penelitian berjudul "Klasifikasi Data Mining untuk Penentuan Stunting pada Balita Menggunakan Metode *Naïve Bayes*," metode *Naïve Bayes* terbukti efektif dalam menentukan stunting pada balita, dengan tingkat akurasi mencapai 95.08% [7]. Algoritma ini memanfaatkan teori probabilitas untuk menemukan peluang terbesar dari kemungkinan klasifikasi, di mana proses klasifikasi dilakukan dengan mengacu pada frekuensi setiap kategori dalam data pelatihan [8]. *Naïve Bayes* dipilih karena terbukti efektif dalam klasifikasi variabel dengan tingkat akurasi tinggi, menjadikannya salah satu algoritma yang tepat untuk menentukan klasifikasi keberlanjutan kegiatan "Sinau Bareng" berdasarkan kehadiran peserta, baik tutor maupun murid.

Tujuan dari penelitian ini adalah menemukan algoritma paling optimal di antara *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* dalam menentukan keberlanjutan kegiatan "Sinau Bareng". Penelitian ini akan mengidentifikasi algoritma yang memberikan hasil terbaik untuk klasifikasi keberlanjutan dengan melakukan analisis dan perbandingan performa ketiga algoritma. Setelah menentukan algoritma paling optimal, akan dikembangkan sebuah aplikasi yang dapat digunakan oleh Dinas Pendidikan Kota Surabaya untuk mengevaluasi lokasi kegiatan "Sinau Bareng. Diharapkan aplikasi tersebut dapat membantu pengambilan keputusan yang lebih efisien tentang keberlanjutan kegiatan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah tertulis di atas maka dirumuskan permasalahan sebagai berikut :

1. Bagaimana hasil penerapan algoritma *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* untuk melakukan klasifikasi keberlanjutan?
2. Bagaimana hasil dari model algoritma yang paling optimal?
3. Bagaimana hasil prediksi dari aplikasi berbasis *website* dari algoritma model yang paling optimal?

## 1.3 Batasan Masalah

Batasan-batasan masalah berikut ditetapkan dalam skripsi agar terarah dan tidak menyimpang dari inti masalah :

1. Data yang digunakan adalah data kehadiran peserta (guru, mahasiswa dan murid) dari kegiatan “Sinau Bareng” pada tahun 2024.
2. Dalam skripsi ini, algoritma *supervised learning* yang digunakan adalah *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *Random Forest*.
3. Klasifikasi menggunakan tiga kelas, yaitu ‘Lanjut’, ‘Diskusi RW dan Kelurahan’ dan ‘Tidak Berlanjut’
4. Model data mining diterapkan dalam aplikasi yang berbasis web menggunakan framework *Flask*.

## 1.4 Tujuan Penelitian

1. Menganalisis dan membandingkan performa algoritma *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* untuk menentukan algoritma dengan akurasi dan performa terbaik.
2. Menerapkan hasil dari model algoritma yang paling optimal dalam aplikasi berbasis web.
3. Menampilkan prediksi keberlanjutan dari setiap lokasi. Aplikasi ini memungkinkan pengelola kegiatan membuat keputusan berbasis data secara lebih tepat, cepat, dan efisien melalui visualisasi hasil prediksi.

## **1.5 Sistematika Penulisan**

Adapun sistematika pada penulisan skripsi ini, ditampilkan dalam limabab sebagai berikut :

### **BAB I PENDAHULUAN**

BAB I berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan skripsi ini

### **BAB II TINJAUAN PUSTAKA**

BAB II berisi dasar teori yang menjelaskan maksud arti dari istilah – istilah tertentu yang digunakan dalam skripsi, serta mencakup pembahasan penelitian sebelumnya sebagai referensi dan sumber literatur untuk skripsi ini.

### **BAB III METODOLOGI PENELITIAN**

BAB III menjelaskan langkah – langkah yang diambil dalam skripsi, dimulai dari studi literatur, proses data mining, hingga penyusunan laporan. Proses data mining dilakukan dengan berpedoman pada kerangka kerja CRISP-DM.

### **BAB IV HASIL DAN PEMBAHASAN**

BAB IV menjelaskan hasil dan analisa dari tahap – tahap yang telah dilakukan dalam skripsi. Bab ini mencakup pembahasan mengenai proses klasifikasi keberlanjutan kegiatan “Sinau Bareng” di Kota Surabaya yang dengan menggunakan algoritma terbaik antara *Support Vector Machine*, *Random Forest*, dan *Naïve Bayes*.

### **BAB V PENUTUP**

BAB V menjelaskan kesimpulan dari skripsi yang sudah dilakukan, serta rekomendasi dari penulis untuk pengembangan penelitian di masa yang akan datang.

## **1.6 Manfaat Penelitian**

Penelitian ini diharapkan mampu memberikan manfaat sebagai berikut:

1. Penelitian ini bertujuan untuk memperkaya wawasan mengenai keunggulan algoritma *Random Forest*, *Support Vector Machine (SVM)*, dan *Naïve Bayes* dalam melakukan klasifikasi keberlanjutan. Dengan demikian, hasilnya dapat dijadikan acuan bagi para peneliti maupun praktisi di bidang data science.
2. Penelitian ini berupaya mengidentifikasi algoritma terbaik yang mampu memberikan klasifikasi paling akurat, sehingga dapat dimanfaatkan untuk meningkatkan performa model dalam pengelolaan keberlanjutan.
3. Melalui pengembangan aplikasi berbasis web yang menggunakan algoritma terpilih, penelitian ini akan menghasilkan alat bantu yang efektif bagi pengelola kegiatan, memungkinkan pengambilan keputusan secara lebih cepat, tepat, dan berbasis data.
4. Aplikasi yang dihasilkan akan menyediakan visualisasi klasifikasi yang informatif dan mudah dipahami, mendukung pemangku kepentingan dalam mengambil keputusan yang lebih terarah untuk pengelolaan keberlanjutan.

*Halaman ini sengaja dikosongkan*