

BAB I

PENDAHULUAN

1.1 Latar Belakang

Machine learning adalah cabang dari *artificial intelegent* adalah sebuah disiplin ilmu yang bersangkutan dengan desain dan pengembangan algoritma yang memungkinkan komputer untuk mengembangkan perilaku data empiris, seperti sensor atau database. Karena pada dasarnya perkembangan database pada masa kini tumbuh dengan sangat pesat, dari kumpulan banyak data tersebut apabila tidak digunakan maka hanya akan menjadi sebuah kumpulan data yang tidak bermanfaat [2].

Maka dari itu data yang tadinya tidak bermanfaat tersebut dapat dipergunakan sebagai sumber data yang kemudian dikelola agar lebih bermanfaat dan proses tersebut biasa disebut dengan istilah *data mining*. *Data mining* adalah suatu proses yang berjalan dengan satu atau lebih teknik *machine learning* yang berfungsi menganalisis dengan mengekstraksi *knowledge* secara otomatis [2]. Salah satu bentuk dari metode maupun Teknik yang biasa digunakan untuk memprediksi sebuah keputusan adalah klasifikasi.

Klasifikasi adalah satu dari banyak masalah mendasar dan fungsi utama dalam *data mining*, dalam klasifikasi sebuah pengklasifikasi dibentuk dari sekumpulan data latih dengan kelas yang telah ditentukan sebelumnya. Performa klasifikasi bisa diukur dengan ketepatan atau tingkat galat [3]. Galat sendiri adalah kesalahan atau *error* yang timbul setiap kali dilakukan operasi perhitungan. Salah

satu algoritma klasifikasi adalah naïve bayes, pengklasifikasi Bayesian merupakan pengklasifikasi statistik untuk dapat memprediksi probabilitas keanggotaan kelas tertentu, menghitung peluang guna suatu hipotesis, menghitung peluang pada suatu kelas dari setiap kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal [3].

Permasalahannya adalah algoritma *Naïve Bayes* termasuk kedalam salah satu *weak classifier* [1], sehingga hasil dari klasifikasi yang dilakukan masih kurang baik dan persentase yang dihasilkan belum maksimal, maka dari itu di kolaborasi dengan algoritma *boosting* yang sesuai yaitu algoritma *AdaBoost*. Karena *AdaBoost* merupakan algoritma yang membangun sebuah *strong classifier* dengan menggunakan *weak classifier* [4][5]. Sehingga persenan yang dihasilkan dapat lebih maksimal.

Algoritma *Naïve Bayes* salah satu dari banyak algoritma klasifikasi guna memprediksi sebuah keputusan. Kelebihan dari algoritma *Naïve Bayes* adalah salah satu algoritma klasifikasi yang paling efektif dan efisien, *Naïve Bayes* memiliki keuntungan yang signifikan karena sebagian besar alternatifnya sangat sederhana [4]. Salah satu algoritma *boosting* yang cocok dengan *Naïve Bayes* banyak ialah *adaptive boosting* atau biasa disebut dengan *AdaBoost*. *Adaptive Boosting* atau *AdaBoost* sendiri adalah *ensemble learning* yang sering digunakan pada *boosting*. *AdaBoost* adalah meta-algoritma dan dapat digunakan bersama dengan banyak algoritma pembelajaran lainnya untuk meningkatkan kinerjanya [1], algoritma *AdaBoost* relatif mudah diimplementasikan dan waktu pengujian yang relatif cepat

[5]. Karena dalam klasifikasi membutuhkan database maka digunakan data dari UCI Machine Learning Repository, data tersebut adalah sebuah data *ecoli* yang ada pada internet. Karena phishing sendiri adalah sebuah masalah yang di pertimbangkan karena masalah vital dalam industry terutama pada *e-banking* dan *e-commerce* karena mengambil nomor dari transaksi online yang melibatkan pembayaran.

Berdasarkan latar belakang masalah diatas, maka penulis melakukan sebuah percobaan yang menggabungkan algoritma klasifikasi yaitu *Naïve Bayes* dengan algoritma *boosting* yaitu *Adaptive Boosting*. Harapannya dengan penelitian ini banyak orang yang dapat mengetahui perbandingan antara klasifikasi yang normal dengan klasifikasi yang di *boosting*, seberapa pengaruhnya kah sebuah algoritma *boosting* dalam klasifikasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, maka dapat di ketahui rumusan masalah sebagai berikut :

1. Apa saja fitur atau atribut *dataset* yang digunakan untuk pelatihan data?
2. Bagaimana menerapkan Algoritma *Naïve Bayes* dengan *AdaBoost*?
3. Metode apa yang dipakai untuk evaluasi penggunaan *Naïve Bayes* dengan *AdaBoost*?
4. Berapa akurasi penggunaan *Naïve Bayes* yang didapatkan sebelum diterapkan *AdaBoost* dengan setelah diterapkan?

1.3 Batasan Masalah

Agar pembahasan lebih terarah dan pemecahan masalah tidak menyimpang, maka penulis membatasi ruang lingkup permasalahan sebagai berikut :

1. Penelitian ini terdapat perhitungan matematis, yang menggunakan algoritma *Naïve Bayes* dan *AdaBoost*.
2. Data diolah dengan menggunakan bahasa pemrograman python, dan didukung dengan beberapa *library* dan menggunakan IDE jupyter lab.
3. *Dataset* yang digunakan berupa data numerik, serta data yang digunakan legal dan bebas di internet untuk umum dari UCI *Machine Learning Repository*.
4. *Dataset* yang digunakan adalah *ecoli* yang bersifat umum dan telah disediakan oleh UCI *Machine Learning Repository*.
5. Hasil evaluasi berupa akurasi dari *Naïve Bayes* dan *AdaBoost* kemudian dibandingkan dengan algoritma klasifikasi lainnya seperti *SVM* dan *Decision Tree* dengan memanfaatkan *library* dari scikit-learn.

1.4 Tujuan Penelitian

Agar pembahasan lebih terarah dan pemecahan masalah tidak menyimpang, maka penulis membatasi ruang lingkup permasalahan sebagai berikut :

1. Mengetahui dan memahami penerapan algoritma klasifikasi *Naïve Bayes* pada *dataset ecoli*.
2. Mengetahui dan memahami pengaruh algoritma *boosting AdaBoost* terhadap hasil dari klasifikasi.

3. Mengetahui akurasi dari hasil klasifikasi hanya dengan algoritma *Naïve Bayes*, dengan hasil dari klasifikasi dari algoritma *Naïve Bayes* yang diterapkan dengan algoritma *boosting AdaBoost*.

1.5 Manfaat Penelitian

Saat melakukan penelitian ini diharapkan metode yang diusulkan mampu menghasilkan hasil maksimal dari klasifikasi *Naïve Bayes*, dan juga dapat melihat perbedaan hasil apakah berbanding sangat jauh jika hanya klasifikasi *Naïve Bayes* dengan klasifikasi *Naïve Bayes* yang di gabungkan dengan *boosting AdaBoost*.

1.6 Metode Penelitian

1.6.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari *UCI Machine Learning Repository*.

1.6.2 Pengolahan Data Awal

Dataset dari *UCI Machine Learning Repository* tersebut telah dilakukan pengolahan data ke dalam bentuk yang dibutuhkan.

1.6.3 Klasifikasi

Setelah memiliki data awal dan data dengan *boosting* maka akan dilakukan klasifikasi. Dalam penelitian ini adakan di dapatkan nilai akurasi yaitu hasil dari proses klasifikasi dengan algoritma *Naïve Bayes*.

1.6.4 Boosting

Setelah dilakukan pengolahan data awal dan di ubah ke bentuk yang dibutuhkan, selanjutnya akan disesuaikan data dengan melakukan pengaturan data untuk mengoptimalkan hasil dengan algoritma *boosting Adaptive Boosting*.

1.6.5 Evaluasi

Setelah keempat tahap itu dilakukan maka akan dilakukan perbandingan antara kinerja algoritma klasifikasi pada dataset asli dengan algoritma klasifikasi pada *dataset boosting*. Indikator evaluasi pada penelitian ini adalah akurasi.

1.7 Sistematika Penulisan

Materi - materi dalam Laporan Skripsi meliputi beberapa sub bab dan diuraikan dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang latar belakang, rumusan masalah, batasan penelitian, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi tentang penelitian terdahulu yang berkaitan dengan masalah penelitian, dan pada bab ini juga memuat teori teori dan konsep untuk penyelesaian masalah yang diusulkan.

BAB III METODE PENELITIAN

Bab ini berisi tentang metode penelitian yang akan dilakukan seperti alat dan bahan, dan alur penelitian yang akan dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan dibahas mengenai hasil dari penelitian yang telah dilakukan yaitu, hasil implementasi algoritma klasifikasi, dan implementasi algoritma boosting.

BAB V KESIMPULAN DAN SARAN

Berisi tentang kesimpulan dari penelitian yang sudah dilakukan serta saran yang didasarkan pada hasil penelitian dan diharapkan dapat menjadi tambahan informasi untuk penelitian – penelitian selanjutnya.

