

## **BAB 1**

### **PENDAHULUAN**

#### **1.1. LATAR BELAKANG**

Eksplorasi data merupakan salah satu upaya untuk menggali dan mendapatkan informasi lebih dari pengelolaan suatu data (Lesmono, 2014). Dalam perkembangannya analisis regresi digunakan sebagai alat untuk eksplorasi data. Analisis regresi merupakan alat statistika yang digunakan untuk mengetahui hubungan fungsional antara peubah penjelas dan peubah respon. Penaksiran nilai-nilai parameter dalam analisis regresi dilakukan dengan menggunakan Metode Kuadrat terkecil. Metode kuadrat terkecil merupakan metode penaksiran parameter analisis regresi yang paling sederhana dan umum digunakan. Namun, metode kuadrat terkecil ini memerlukan banyak asumsi yang harus dipenuhi untuk dapat dalam penaksiran parameter analisis regresi. Salah satu asumsi yang harus dipenuhinya yaitu asumsi normalitas dan asumsi linieritas.

Apabila kedua asumsi tersebut tidak terpenuhi, maka analisis regresi dengan penaksiran parameter menggunakan metode kuadrat terkecil tidak dapat dilakukan. Salah satu metode yang dapat dipergunakan untuk mengatasi masalah ini yaitu metode pohon regresi atau *Regression Trees*. Metode pohon regresi merupakan salah satu teknik eksplorasi data nonparamterik. Dengan kata lain, pada metode pohon regresi tidak memerlukan asumsi-asumsi tertentu. Metode pohon regresi merupakan teknik analisis data yang bertujuan untuk mengetahui pengaruh semua peubah penjelas terhadap peubah responnya dengan memperhatikan nilai peubah responnya tanpa memperhatikan asumsi-asumsi yang berlaku pada analisis regresi biasa.

Seperti yang telah dikemukakan sebelumnya dapat disimpulkan bahwa metode pohon regresi bebas asumsi (Komalasari, 2007), hal ini merupakan keunggulan dari metode pohon regresi. Namun, ternyata metode pohon regresi memiliki kelemahan yaitu struktur hierarkisnya memungkinkan terjadinya *high-*

*variance predictor* atau memiliki varians yang tinggi. Kelemahan metode pohon regresi ini dapat diatas dengan menggunakan Pendekatan *Boosted Regression Trees* (BRT). Pendekatan BRT berguna untuk mengatasi kelemahan tingginya varians dengan menggunakan teknik boosting untuk menggabungkan beberapa pohon regresi secara aditif untuk membentuk *low-variance predictor*. Prinsip utama dari teknik boosting sendiri adalah membangun dan mengkombinasikan *weak learners* dalam proses *iterative stagewise* untuk membangun *strong learners*. *Weak learners* adalah individu pada pohon regresi sedangkan *strong learners* merupakan hasil kombinasi pohon regresi dengan cara aditif.

Metode BRT banyak digunakan untuk prediksi dalam berbagai bidang contohnya penggunaan BRT dalam bidang ekonomi khususnya sudah populer untuk siklus bisnis sejak awal tahun 1990an oleh (Estrella & Hardouvelis, 1991; Estrella & Mishkin, 1998) dan juga telah digunakan secara luas sebagai alat untuk peramalan resesi dalam penelitian (Fritzsche & Kuzin 2005; Proano & Theobald, 2014; Theobald, 2012; Fritzsche, Dopke, & Pierdzioch, 2017). Kemudian BRT juga digunakan untuk mengetahui nilai faktor penggunaan lahan penelitian ini dilakukan oleh (Cheong, Leitao, & Lakes; 2014) dengan judul penelitian “*Assesment of land use factor associated with dengue cases in Malaysia using Boosted Regression Trees*”, penelitian ini memperoleh kesimpulan nilai faktor-faktor yang mempengaruhi kasus demam berdarah. Hal ini menunjukkan bahwa *Boosted Regression Trees* (BRT) adalah salah satu teknik yang bertujuan meningkatkan model tunggal dengan mengkombinasikan beberapa model untuk prediksi menggunakan metode pohon regresi dan metode boosting (Elith, Leathwick, & Hastie, 2008).

Luasnya penggunaan BRT membuat penulis tertarik untuk mengaitkan penggunaan BRT mengenai faktor-faktor yang memengaruhi tingkat kriminalitas, karena menurut liputan6.com Provinsi Jawa Timur pada tahun 2012 pernah menduduki urutan pertama tingginya angka kejahatan dan pada tahun 2019 ini tingkat kriminalitas di Jawa Timur Meningkat sebesar 96% sehingga perlu adanya analisis faktor-faktor ini agar dapat menanggulangi naiknya tingkat kriminalitas.

Berdasarkan uraian yang telah dikemukakan, maka penulis tertarik untuk membuat skripsi yang berjudul “**Boosted Regression Trees**”

## **1.2. RUMUSAN MASALAH**

Berdasarkan uraian dari latar belakang maka permasalahan yang akan dibahas dalam skripsi ini adalah :

1. Bagaimana konstruksi algoritma *gradient boost* pada metode BRT?
2. Bagaimana penerapan metode BRT untuk mengetahui pengaruh faktor-faktor yang memengaruhi tingkat kriminalitas?

## **1.3. BATASAN MASALAH**

Masalah dalam penulisan skripsi ini dibatasi pada penerapan metode BRT menggunakan metode *gradient boosting* dan pohon regresi.

## **1.4. TUJUAN PENELITIAN**

Adapun tujuan dari disusunnya penulisan skripsi ini adalah sebagai berikut :

1. Mengkontruksi algoritma *gradient boost* pada metode BRT
2. Mengetahui penerapan metode BRT untuk mengetahui pengaruh faktor-faktor yang memengaruhi tingkat kriminalitas.

## **1.5. MANFAAT PENELITIAN**

Manfaat yang diperoleh dari penelitian ini adalah:

### **1.5.1. Manfaat Teoritis**

Melalui penulisan skripsi ini diharapkan dapat memperkaya dan memperluas wawasan serta pengetahuan tentang metode pohon regresi yang diperbaiki metode boosting yang dikenal dengan BRT.

### **1.5.2. Manfaat Praktis**

Mengenali penerapan metode BRT dengan menggunakan algoritma *gradient boost* untuk pohon regresi, sehingga skripsi ini nantinya dapat dijadikan bahan pertimbangan dan salah satu sumber informasi yang dapat mendukung tujuan dari pihak yang memiliki kepentingan.