

## BAB I PENDAHULUAN

### I.1 Latar Belakang

Perkembangan IT (*Information Tecnology*) saat ini amatlah pesat. dan saat ini kita juga sering sekali mendengar istilah *big data*. *Big data* merupakan penyimpanan data dalam jumlah besar, melakukan proses serta analisisnya. konsep dari *big data* sendiri meliputi dari aplikasi, teknik, dan aspek ilmiah, tetapi belum ada definisi terpadu *big data*, itu bervariasi di kalangan komunitas akademis dan bisnis (Shi, 2015). Dalam decade terakhir jumlah data yang tersedia telah tumbuh secara eskponensial. Laporan dari Internal Data Corporation (IDC) bahwa pada tahun 2011 data yang dihasilkan seluruh dunia sekitar 1.8 *zettabyte* dan data tersebut diperkirakan akan tumbuh 50 kali lipat pada tahun 2020.

Tahun 1960 pertama kali muncul istilah *database management* dan pada tahun 70an lebih dikenal dengan *data warehouse*. Lalu pada tahun 80an dikenal lagi istilah baru yaitu *knowledge discovery in database (KDD)* dan pada tahun 90an berubah lagi menjadi *enterprise resource planing (erp)* dan *data mining* (Shi, 2015).

*Impact* yang didapat dari *Big data* dalam kehidupan sehari-hari adalah memberikan peluang bisnis bagi banyak perusahaan. Hampir semua industri sudah menggunakan *Big data*. *Big data* saat ini amatlah penting untuk pertumbuhan bisnis, dan bahkan dapat dijadikan tolak ukur dalam berkompetisi bisnis (Kumar, 2016). Dari semua manfaat yang diberikan, *Big data* juga memiliki beberapa tantangan diantaranya adalah teknologi yang dapat menghandle *Big data* ini, serta *skill* dan kemampuan SDM dalam mengelola data tersebut sehingga data tersebut menjadi informasi yang bernilai.

Istilah *Big data* mengacu pada aplikasi teknologi informasi yang menangani masalah data yang sifatnya *massive*, selanjutnya disebut dengan *Data Science*. Orang yang mendalami *data science* sering disebut dengan *Data Scientists*. Penggabungan dari bisnis *intelligent* dan bisnis *analytics* disebut juga dengan *Big Data Analytic*.

*Big Data Analytic* memiliki sebuah metode yaitu *Data Mining* yang dimana metode tersebut dapat digunakan untuk memproses sebuah nilai dari sebuah data abstrak yang diolah menjadi sebuah informasi (Han, Pei, & Kaber, 2012).

Informasi tersebut didapatkan dengan cara mengolah pola-pola dari sebuah data yang dimana pada setiap data abstrak terdapat sebuah pola-pola yang akan menjadi sebuah informasi penting.

Dalam proses *mining* data terdapat atribut yang telah ditentukan berdasarkan tujuan dari proses *mining* tersebut. atribut tersebut memiliki tingkat keterkaitan dari 0 hingga 1 dimana semakin tingkat keterkaitan sebuah atribut mendekati 1 maka semakin bergantung keterkaitan antar atribut tersebut.

Penentuan setiap *decision* pada sebuah data sangat berkaitan dengan tingkat kebergantungan sebuah atribut terhadap atribut lainnya dalam *dataset*. Yang dimana dalam penentuan sebuah *decision* dapat dipengaruhi oleh atribut yang memiliki tingkat ketergantungan paling tinggi terhadap atribut lainnya. Penentuan *decision* sendiri berguna bagi perusahaan atau organisasi untuk menentukan langkah yang harus diambil untuk mencapai tujuannya.

Dari banyaknya *dataset* yang ada, pada penelitian ini akan digunakan *dataset* kelautan. Dimana pada *dataset* tersebut terdapat 465 atribut. Hasil yang diinginkan dari pengolahan data tersebut adalah menentukan atribut yang paling berpengaruh untuk penentuan *decision*. Hasil tersebut nantinya akan diberikan kepada badan pusat statistik guna melihat *conditional* atribut yang paling berpengaruh pada potensi hasil laut Indonesia. Badan pusat statistik sendiri belum mengetahui apakah yang paling mempengaruhi potensi hasil laut Indonesia. Diharapkan hasil dari penelitian ini akan membantu badan pusat statistik.

Berdasarkan informasi diatas maka dalam penelitian ini akan digunakan sebuah metode untuk mengukur ketergantungan antar atribut dalam sebuah *dataset*. Dalam menentukan atribut yang paling berpengaruh untuk penentuan *decision* ini terdapat beberapa metode, dalam penelitian ini akan dibandingkan 2 metode yang paling efektif dalam menentukan *decision* dari sebuah *dataset*.

## **I.2 Perumusan Masalah**

Berdasarkan latar belakang yang sudah dijelaskan, perumusan masalah yang dikaji dalam penelitian ini yaitu

1. Bagaimana cara mencari metode paling tepat yang dapat digunakan untuk menentukan *conditional* atribut yang paling berpengaruh dalam penentuan *decision*?
2. Bagaimana menentukan *conditional* atribut yang paling berpengaruh dalam penentuan *decision* pada sebuah *dataset*?

## **I.3 Tujuan Penelitian**

Adapun tujuan dari penelitian yang dilakukan berdasarkan perumusan masalah yang telah dituliskan sebelumnya adalah:

1. Mengetahui metode yang dapat digunakan dalam menentukan *conditional* atribut paling berpengaruh dalam penentuan *decision*.
2. Mendapatkan metode yang paling tepat dalam menentukan *conditional* atribut paling berpengaruh dalam penentuan *decision*.

## **I.4 Batasan Penelitian**

Penelitian ini hanya untuk melihat atribut yang paling dominan dalam penetapan *decision* sebuah data menggunakan metode *maximum dependency of attribute*

## **I.5 Manfaat Penelitian**

Adapun manfaat penelitian dari tugas akhir ini yaitu memberikan gambaran untuk mengidentifikasi atribut yang paling dominan dari dataset *marine* untuk menentukan *decision* dengan peringkat atribut berdasarkan ketergantungan maksimum atribut dalam dataset. Didasari oleh total akurasi pendekatan menggunakan *maximum dependency of attribute* dalam teori *rough set*.

## **I.6 Sistematika Laporan**

Penelitian tugas akhir ini akan diuraikan berdasarkan sistematika laporan sebagai berikut:

## BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang penelitian dilakukan, rumusan masalah penelitian, tujuan penelitian, batasan penelitian, manfaat penelitian baik dari sisi keilmuan maupun teknis, dan sistematika penelitian.

## BAB II TINJAUAN PUSTAKA

Pada bab ini berisi landasan teori atau literatur yang relevan guna mendukung penulisan penelitian tugas akhir ini. Studi literatur didapatkan dari beberapa sumber, diantaranya berasal dari jurnal atau penelitian terdahulu, dan buku-buku yang relevan.

## BAB III METODOLOGI PENELITIAN

Pada bab metodologi penelitian ini terdiri dari uraian model konseptual yang digunakan dan sistematika penelitian tugas akhir. Model konseptual menjelaskan mengenai masukan atau *input* yang diperlukan untuk melaksanakan penelitian hingga mendapatkan keluaran atau *output* dari penelitian. Sedangkan sistematika penelitian yaitu berupa penjelasan tahapan-tahapan dalam melakukan mulai dari tahap persiapan hingga tahap pelaporan.

## BAB IV TAHAP PENGOLAHAN DATA

Pada bab ini akan menjelaskan tahapan – tahapan dalam melakukan pengolahan data

## BAB V HASIL ANALISIS DATA

Bab ini memaparkan penyelesaian masalah yang terjadi pada penilitan, hasil dari penilitan dan analisis dari penelitian.

## BAB VI KESIMPULAN DAN SARAN

Bab ini berisi penutup yang berupa kesimpulan dan saran dari penulisan penelitian ini.