

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Seiring pesatnya perkembangan zaman selalu diiringi dengan kemajuan teknologi yang semakin canggih. Setiap perkembangan di dunia entah itu tentang teknologi, ekonomi, pertanian, dll terdapat dokumentasi dari tahun ke tahun. Kumpulan dokumentasi tersebut digunakan sebagai dokumentasi tanpa ada pemanfaatan data tersebut secara berkelanjutan dan tersimpan didalam *data warehouse* atau biasa disebut kuburan data.

*Data Warehouse* adalah suatu konsep dan kombinasi teknologi yang memfasilitasi organisasi untuk mengelola dan memelihara data historis yang diperoleh dari sistem atau aplikasi operasional [Ferdiana, 2008]. *Data warehouse* memungkinkan integrasi berbagai macam jenis data dari berbagai macam aplikasi atau sistem. Hal ini menjamin mekanisme akses "satu pintu bagi manajemen untuk memperoleh informasi, dan menganalisisnya untuk pengambilan keputusan". Pemakaian teknologi *data warehouse* hampir dibutuhkan oleh semua organisasi, tidak terkecuali startup kaggle sebagai komunitas *data scientist*.

Kaggle adalah salah satu komunitas terbesar ilmuwan data (*data scientist*) diseluruh dunia. Di kaggle ini, para *data scientist* bisa berkompetisi untuk menyelesaikan permasalahan ilmiah berdasarkan data yang kompleks (biasanya disebut sebagai *big data*). Cara menyelesaikannya bisa individual atau berkelompok. Lama pengerjaan tiap kompetisis berbeda-beda, ada yang hitungan bulan, ada juga yang hitungannya sampai tahunan. Pemenang kompetisi ini

tentunya akan mendapatkan *fee*/sejumlah uang sebagai penghargaan (setiap kompetisi berbeda-beda besarnya). Selain itu, pemenang biasanya juga diajak bergabung untuk menyelesaikan permasalahan bisnis yang sangat menarik dan sensitif dari perusahaan dunia terbesar.

Salah satu bidang ilmu yang dapat digunakan untuk memanfaatkan *Data Warehouse* menjadi informasi yang lebih berguna adalah *data mining*. Tan (2006) mendefinisikan *data mining* sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah *data mining* kadang disebut juga *Knowledge Discovery in Database (KDD)* yaitu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar.

Salah satu teknik yang dikenal dalam *data mining* yaitu *clustering*. Pengertian *clustering* dalam *data mining* adalah pengelompokan sejumlah data atau objek ke dalam *cluster* (grup) sehingga setiap dalam *cluster* tersebut akan berisi data yang semirip mungkin dan berbeda dengan objek *cluster* yang lainnya. *Clustering* adalah suatu teknik dalam *data mining* yang digunakan untuk memasukkan data ke dalam grup yang bersesuaian tanpa pengetahuan yang mendalam tentang grup tersebut (Santosa, 2007). *Clustering* ini bertujuan untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi antar *cluster* (Agusta, 2007). Sampai saat ini, para ilmuwan data (*data scienties*) terus melakukan berbagai

usaha untuk melakukan perbaikan model *cluster* dan menghitung jumlah *cluster* yang optimal sehingga dapat di hasilkan *cluster* yang baik (Alfina, *et al.* 2012).

Data *world development indicators* adalah salah satu dari sekian banyak data yang di unggah ke startup kaggle. *World development indicators* berisi data *record* negara - negara di dunia seperti *currency*, *system of trade*, *income group*, dll. Data - data tersebut akan di analisis dengan *cluster* data mining kemudian di kelompokkan ke dalam *cluster*, untuk mengetahui pola informasi baru yang dapat dimanfaatkan sebagai pengetahuan.

Dengan hasil analisis *cluster* dapat diketahui negara - negara yang memiliki potensi berkembang dengan indikator mirip.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, maka pokok permasalahan yang akan dirumuskan yaitu

1. Bagaimana mendapatkan data yang sesuai untuk analisis *cluster*?
2. Bagaimana memilih atribut yang cocok untuk *clustering*?
3. Bagaimana menerapkan *clustering*?
4. Evaluasi *clustering*?

## 1.3 Batasan Masalah

Dalam skripsi ini ditentukan suatu batasan masalah, sehingga penulis dapat menentukan metode yang tepat untuk memudahkan pengerjaan dan menghindari pembahasan diluar batasan masalah yang sudah ditentukan. Adapun beberapa batasan masalah, sebagai berikut :

1. Data yang diambil berasal dari situs kaggle.com
2. Algoritma yang digunakan DBSCAN

3. Menggunakan Sql Database Browser
4. Menggunakan bahasa Python

#### 1.4 Maksud dan Tujuan Penelitian

Dalam penelitian ini, penulis berfokus pada pokok permasalahan sebagai berikut :

1. Penerapan data mining menggunakan teknik *clustering* dengan algoritma DBSCAN.
2. Evaluasi metode *cluster*

#### 1.5 Manfaat Penelitian

Adapun manfaat penelitian ini antara lain :

1. Bagi AMIKOM

Memberikan referensi untuk mahasiswa amikom tentang data mining

2. Bagi Penulis

Dengan adanya penelitian ini, penulis berharap mampu menerapkan ilmu yang telah di berikan selama menuntut ilmu di STMIK AMIKOM YOGYAKARTA.

#### 1.6 Metodologi Penelitian

Dalam penelitian ini, penulis melakukan tahap :

##### 1.6.1 Pengumpulan Data

###### Observasi

Metode Observasi, yaitu mengumpulkan referensi baik dari buku, internet, maupun sumber-sumber lain mengenai *Cluster Data Mining* dengan algoritma DBSCAN.



## 1.6.2 Metode Penelitian

### Metode Experimental

Eksperimen merupakan salah satu metode penelitian yang dapat dipilih dan digunakan dalam penelitian pembelajaran pada latar kelas (PTK). Penelitian eksperimental dapat diartikan sebagai sebuah studi yang objektif, sistematis, dan terkontrol untuk memprediksi atau mengontrol fenomena. Penelitian eksperimen bertujuan untuk menyelidiki hubungan sebab akibat (*cause and effect relationship*), dengan cara mengekspos satu atau lebih kelompok eksperimental dan satu atau lebih kondisi eksperimen. Hasilnya dibandingkan dengan satu atau lebih kelompok kontrol yang tidak dikenai perlakuan (Danim, 2002).

## 1.6.3 Metode Pengembangan

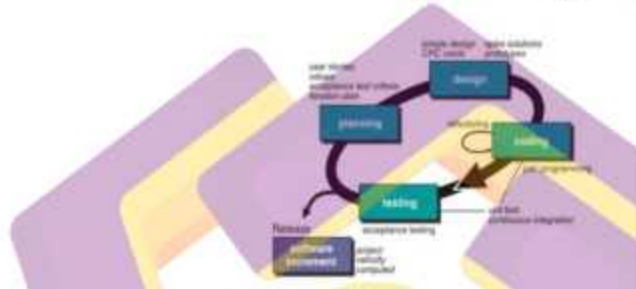
### Extreme Programming

*Extreme Programming* merupakan metode tangkas pengembangan perangkat lunak berfokus pada penyediaan nilai tertinggi bagi pelanggan dalam cara tercepat mungkin.

Siklus *Extreme Programming* adalah Proses pengembangan perangkat lunak tradisional *linear*, dengan masing - masing tahap membutuhkan penyelesaian tahap sebelumnya. Misalnya, desain dimulai hanya pada penyelesaian seluruh analisis program, dan coding dimulai hanya setelah selesai dan pembekuan desain. Konsep ini bekerja dengan baik ketika kebutuhan pelanggan tetap statis, tetapi dalam prakteknya, persyaratan sering berubah dan menyebabkan sering editing program yang mengarah ke kesalahan.

Proses pengembangan perangkat lunak *Extreme Programming* dimulai dengan perencanaan, dan semua iterasi terdiri dari lima tahap dasar dalam siklus hidupnya: *planning*, *designing*, *coding*, *testing*, dan *listening* seperti yang ditunjukkan pada Gambar 1.1.

## Extreme Programming (XP)



Gambar 1.1 Alur *Extreme Programming*

### 1. *Planning*

Tahap pertama Ekstrim siklus hidup *Programming* berencana, di mana pelanggan atau pengguna bertemu dengan tim pengembangan untuk membuat cerita pengguna atau persyaratan. Tim pengembangan mengkonversi cerita pengguna ke iterasi yang mencakup sebagian kecil dari fungsi atau fitur yang diperlukan. Kombinasi dari iterasi menyediakan produk yang berfungsi penuh di akhir.

### 2. *Designing*

Iterasi pemrograman XP dimulai dengan merancang. Prinsip-prinsip dari tahap ini adalah :

1. Menggunakan sistem metafora atau standar pada nama, nama kelas dan metode, dan menyepakati gaya seragam dan format untuk memastikan kompatibilitas antara kerja anggota tim yang berbeda

2. Menggunakan *Software Class Responsibilities* dan *Colaboration* (CRC) yang memungkinkan untuk dari pola pikir tradisional dan membuat teknologi berorientasi objek. memungkinkan semua anggota tim proyek untuk menyumbangkan ide-ide, dan menyusun ide-ide terbaik dalam desain
3. Meciptakan program sederhana yang mengeksplorasi solusi potensial untuk masalah tertentu, mengabaikan semua masalah lain, untuk mengurangi risiko

### 3. Coding

Coding merupakan fase paling penting dalam siklus *Extreme Programming*. *XP programing* memberikan prioritas kepada coding diatas semua tugas-tugas lain seperti dokumentasi untuk memastikan bahwa pelanggan menerima sesuatu yang substansial dalam nilai diakhir. Standar coding meliputi :

1. Mengembangkan kode berdasarkan metafora dan standar yang telah disepakati, dan mengadopsi kebijakan kepemilikan kode kolektif.
2. *Pair programing* atau *developing code* oleh dua programmer bekerja sama pada satu mesin, yang bertujuan untuk menghasilkan kode berkualitas tinggi dengan biaya yang sama atau kurang.
3. Kepatuhan kerja selama 40 jam per minggu tanpa lembur. Hal ini memastikan para pengembang bekerja di puncak kemampuan mental dan fisik mereka.

4. Integrasi code ke repositori khusus, dengan hanya satu tempat integrasi pada satu waktu untuk mencegah konflik, dan optimasi di akhir.

#### 4. *Testing*

*Extreme programming* pengujian terintegrasi dengan tahap pengembangan di akhir. Semua kode memiliki unit test untuk menghilangkan *bug* dan kode melewati semua tes unit tersebut sebelum rilis.

#### 5. *Listening*

Dasar *extreme programming* adalah mekanisme berkelanjutan keterlibatan pelanggan melalui umpan balik dalam tahap pengembangan. Dasar dari umpan balik adalah pengujian program. Setiap revisi menjadi dasar dari desain baru dan proses *designing - coding - testing* diulang. Jika tidak ada revisi dengan hasil tes iterasi berakhir.

#### 1.6.4 Evaluasi Cluster

##### *Silhouette Coefficient*

*Silhouette Coefficient* adalah metode evaluasi *cluster* dengan mencari nilai optimal dari *cluster* yang terbentuk. Nilai evaluasi dari *Silhouette Coefficient* memiliki interval  $-1$  (minus satu) sampai  $1$  (satu). Sebuah *cluster* bisa dikatakan optimal jika nilai evaluasi mendekati angka  $1$  (satu).



### 1.6.5 Tahapan analisis

#### 1. Analisis Database

Melakukan analisis terhadap *database*, untuk menentukan atribut pada database yang memuat data inti, data batas, dan data noise.

#### 2. Perancangan

Dalam tahap ini perancangan disesuaikan dengan kebutuhan dan menggambarkan sistem yang akan dibentuk.

#### 3. Pemrograman

Aplikasi dibangun setelah proses perancangan selesai. Aplikasi ini dibangun untuk *platform* komputer.

#### 4. Uji Coba dan Implementasi

Aplikasi sudah dibangun sebelumnya akan diuji coba terlebih dahulu. Apakah aplikasi ini mampu menyediakan informasi yang akurat bagi pengguna. Jika belum, maka akan dilakukan perbaikan tingkat akurasi *output*.

#### 5. Dokumentasi

Pada tahap ini dilakukan pembuatan laporan mulai dari studi literatur sampai dengan implementasi, serta penarikan kesimpulan dan saran. Pada tahap ini juga dicatat apa saja yang menjadi kelemahan dan kelebihan sistem.

### 1.7 Sistematika Penulisan

Penulisan skripsi terdiri dari lima bab. Setiap bab membahas pokok - pokok permasalahannya untuk mempermudah penulis dengan sistematika penulisan sebagai berikut :

**BAB I : PENDAHULUAN**

Pada bab ini berisikan Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Metode Penelitian Dan Sistematika Penulisan.

**BAB II : LANDASAN TEORI**

Landasan teori ini adalah kumpulan dari studi pustaka penulis yang didalamnya membahas seputar teori-teori yang mendukung dalam pembuatan penelitian ini.

**BAB III : ANALISA DAN PERANCANGAN**

Bab ini membahas tentang analisis terhadap sistem yang akan dibuat seperti kebutuhan apa saja yang diperlukan untuk membuat sistem, rancangan user interface dan rancangan tentang sistem yang dibuat.

**BAB IV : IMPLEMENTASI DAN PEMBAHASAN**

Dalam bab ini akan diuraikan secara lengkap tentang tahap-tahap perancangan dan pembuatan program. Tentang cara kerja sistem dan pembahasan, serta melakukan pengujian sistem yang dibuat.

**BAB V : PENUTUP**

Pada bab ini akan membahas tentang kesimpulan penelitian dan saran yang dituliskan oleh penulis.