

## **Big Data di Perpustakaan dengan Memanfaatkan *Data Mining***

**Athanasia Octaviani Puspita Dewi<sup>1\*)</sup>**

<sup>1</sup>*Program Studi Ilmu Perpustakaan, Fakultas Ilmu Budaya, Universitas Diponegoro,  
Jl. Prof. Soedarto, SH, Kampus Undip Tembalang, Semarang, Indonesia*

\*) Korespondensi: [athanasiadewi@lecturer.undip.ac.id](mailto:athanasiadewi@lecturer.undip.ac.id)

### **Abstract**

**[Title: Utilization of Big Data in Libraries]** This article discuss about big data in a library and its benefits if big data is processed. The results of this study are the data in the library can be referred to big data if it contains 3 characteristics of big data namely volume, velocity, and variation. Big data is then processed using data mining, one of which is by using associations which will produce association between collections. If the big data in one library is a loan data collection, it can create association between collections borrowed. The librarians can decide on the next action from the results of the information obtained through the processing of this big data. So hopefully it will play a role in the progress of a library.

**Keywords:** big data; data mining; library

### **Abstrak**

Artikel ini membahas tentang *big data* dalam perpustakaan dan manfaatnya jika diolah. Hasil dari penelitian ini adalah data yang ada di dalam perpustakaan dapat disebut sebagai *big data* jika memenuhi 3 karakteristik dari *big data* yaitu *volume*, *velocity*, dan *variety*. *Big data* kemudian bisa diolah menggunakan *data mining*, salah satunya dengan menggunakan asosiasi dimana akan menghasilkan keterkaitan antar barang. Jika dalam perpustakaan *big data*-nya adalah data peminjaman berarti dapat menghasilkan keterkaitan antar koleksi yang dipinjam. Para pustakawan khususnya dapat menentukan tindakan selanjutnya dari hasil informasi yang didapatkan melalui hasil pengolahan *big data* ini. Sehingga diharapkan akan berperan dalam kemajuan sebuah perpustakaan.

**Kata kunci:** big data; data mining; perpustakaan

## **1. Pendahuluan**

Teknologi informasi dapat digabungkan dengan berbagai macam ilmu. Penggabungan ini biasanya akan menghasilkan kemudahan bagi penggunanya. Perpustakaan semakin hari semakin berinovasi dengan memanfaatkan teknologi informasi guna memaksimalkan pelayanan terhadap pengunjung dan anggota perpustakaan. Mulai dari penggunaan otomasi perpustakaan, perpustakaan digital, penggunaan alat-alat teknologi semacam RFID (*Radio Frequency Identification*), *barcode scanner* merupakan beberapa pemanfaatan teknologi informasi di dalam perpustakaan. Pemanfaatan teknologi informasi dalam bidang perpustakaan ini diharapkan dapat meningkatkan pelayanan perpustakaan, baik itu secara kualitas maupun secara kuantitas.

Salah satu pemanfaatan teknologi informasi dalam bidang perpustakaan adalah pengelolaan data perpustakaan dengan menggunakan bantuan teknologi. Beberapa data yang dihasilkan atau disimpan di perpustakaan adalah data anggota, data pengunjung, data koleksi, dan data sirkulasi. Jika data-data tersebut dalam jumlah yang sangat banyak, data yang disimpan sangat beragam, dan perubahan datanya

cepat, maka dapat disebut dengan *big data*. Terbentuknya *big data* ini dapat dimanfaatkan pustakawan untuk dilakukan pengolahan selanjutnya dengan menggunakan teknik *data mining*.

Dari penjelasan tersebut maka perlu dikaji data apa saja yang dapat digolongkan dalam *big data* pada perpustakaan dan dapat dimanfaatkan untuk apa data-data tersebut. Tujuan dari penelitian ini adalah untuk menambah wacana bagi pembaca dan pustakawan dalam hal pemanfaatan *big data* khususnya dalam bidang perpustakaan. Manfaat dari penelitian ini adalah untuk menambah ilmu bagi pembaca dan pustakawan dalam hal pemanfaatan *big data*. Metode penelitian yang digunakan adalah studi literatur, dengan menganalisa tulisan-tulisan dari beberapa sumber.

## 2. Landasan Teori

### 2.1 Big Data

*Big data* jika diterjemahkan ke dalam bahasa Indonesia maka berarti data yang besar. Pengertian *big data* menurut St. Nath dalam Ali (2015):

“*Big data is an evolving term that describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for information.*”

*Big data* adalah istilah yang terus berkembang yang menjelaskan data dengan jumlah bervolume besar yang terstruktur, semi terstruktur, dan data tidak terstruktur yang berpotensi ditambang untuk mendapatkan informasi. Dari pengertian tersebut dapat disimpulkan bahwa data yang ada harus bervolume besar atau berjumlah banyak, data-data yang diperoleh dikumpulkan sehingga nantinya dapat dimanfaatkan menjadi sebuah atau beberapa informasi. Sebenarnya jika dilakukan lebih lagi dalam pengolahannya ini dapat menghasilkan suatu pendukung keputusan.

*Big Data* menurut Chandarana, Parth, & Vijayalakshmi dalam Sirait (2016) terdiri dari 3V: *volume*, *variety*, *velocity*, dan ada yang menambahkan unsur V lainnya seperti *veracity* dan *value*. *Volume* (kapasitas data) berkaitan dengan ukuran media penyimpanan data yang sangat besar atau mungkin tidak terbatas hingga satuan petabytes atau zettabytes; *variety* (keragaman data) adalah tipe atau jenis data yang dapat diolah mulai dari data terstruktur hingga data tidak terstruktur; sedangkan *velocity* (kecepatan) terkait dengan kecepatan mengolah data yang dihasilkan dari berbagai sumber, mulai dari data *batch* hingga *real time*, sementara karakteristik *veracity* (kebenaran) dan *value* (nilai) terkait dengan ketidakpastian data dan nilai manfaat dari informasi yang dihasilkan.

Sedangkan pengertian lain menurut Laney dalam Ali (2015) mendefinisikan *big data* dalam tiga dimensi besar data yaitu *volume* mengacu pada jumlah data yang diciptakan, *velocity* mengacu pada kecepatan data yang sedang dibuat, *variety* merujuk baik untuk jenis data yang dikumpulkan. Dengan menggunakan acuan dari Chandarana, Parth, & Vijayalakshmi dan Laney tersebut maka dapat disimpulkan bahwa pendefinisian *big data* meliputi 3 V, yaitu *volume*, terkait dengan jumlah data atau banyaknya data yang terkumpul dalam suatu kegiatan; *velocity*, adalah kecepatan dalam menghasilkan

banyaknya data dalam sebuah kegiatan; *variety*, adalah jenis atau tipe data yang dihasilkan dalam kegiatan penciptaan data.

## 2.2 Data Mining

Menurut Gartner Group dalam Larose, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (2005). *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, dan asosiasi.

### 1. Deskripsi

Deskripsi digunakan untuk pengelolaan *data mining* dengan cara menggambarkan pola kecenderungan dalam suatu data. Deskripsi dalam *data mining* ini lebih menjelaskan kepada pola-pola yang terbentuk akibat adanya penambahan data.

### 2. Estimasi

Estimasi pada *data mining* hampir sama dengan klasifikasi akan tetapi target dari estimasi ini lebih ke hasil yang sifatnya numerik, bukan kategori. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

### 3. Prediksi

Prediksi memiliki beberapa kesamaan dengan klasifikasi dan estimasi, hasil yang didapat pada prediksi ini akan ada pada kejadian yang akan datang.

### 4. Klasifikasi

Klasifikasi pada *data mining* digunakan untuk mengelompokkan sesuatu berdasarkan kategori. Kategori di sini berarti sesuatu yang sudah digolongkan, misal gaji rendah, sedang, dan tinggi, curah hujan rendah, sedang, dan tinggi, dan sebagainya.

### 5. Pengklusteran

Pengklusteran berbeda dengan klasifikasi, pengklusteran lebih kepada pengelompokan data berdasarkan kemiripan. Pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (*homogen*), dimana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

### 6. Asosiasi

Asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (*Market Basket Analysis*). Contoh asosiasi dalam bisnis adalah mencari barang dalam supermarket yang dibeli secara bersamaan dalam suatu waktu. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam

*database* dan *confidence* (nilai kepastian) yaitu kuatnya hubungan antar item dalam aturan asosiatif. (Larose, 2005)

Contoh:

Aturan asosiatif biasanya dinyatakan dalam bentuk seperti ini {roti, mentega} → {susu} (*support* = 40%, *confidence* = 50%) *Support* 40% artinya adalah 40% dari transaksi di dalam *database* yang memuat item roti dan mentega juga memuat item susu. Sedangkan *confidence* 50% menunjukkan persentase susu yang terdapat pada transaksi yang mengandung roti dan mentega. Nilai *support* digunakan untuk menentukan *frequent itemset*. *Itemset* yang nilai *support*-nya memenuhi parameter *threshold minimum support* (min\_sup) masuk dalam *frequent itemset*. Sedangkan nilai *confidence* digunakan dalam menentukan *strong association rule*. *Association rule* yang nilai *confidence*-nya memenuhi parameter *threshold minimum confidence* (min\_conf) termasuk dalam *strong association rule*.

### 2.3 Data dalam Layanan Perpustakaan

Menurut Turban (2010), data adalah deskripsi dasar dari benda, peristiwa, aktivitas dan transaksi yang direkam, dikelompokkan, dan disimpan tetapi belum terorganisir untuk menyampaikan arti tertentu. Menurut Inmon (2005), data adalah kumpulan dari fakta, konsep, atau instruksi pada penyimpanan yang digunakan untuk komunikasi, perbaikan dan diproses secara otomatis yang mempresentasikan informasi yang dapat dimengerti oleh manusia. Berdasarkan teori tersebut dapat disimpulkan bahwa, data adalah deskripsi dari suatu benda, peristiwa, aktivitas dan transaksi yang direkam, dikelompokkan, dan disimpan dalam jumlah yang besar dan belum diolah menjadi sebuah informasi.

Menurut Ali (2015) Secara garis besar *big data* adalah hasil dari proses pergerakan suatu peristiwa yang kemudian dijadikan informasi sesuai dengan kebutuhan, dari informasi yang didapatkan itu dapat dijadikan bahan pengambilan keputusan (*wisdom*). *Big data* yang diperoleh bisa bersumber dari berbagai kegiatan seperti transaksi, *devices*, dan *website* yang menghasilkan data (bukan data dalam pengertian bandwidth). Seperti data cuaca, finansial, pergerakan politik, ATM, pangan, jumlah pengunjung, jenis kelamin, kesukaan, hobi, dan sebagainya. Menurut Rahayuningsih (2007: 87), perpustakaan sebagai organisasi pemberi jasa informasi dapat memberikan layanan kepada pengguna antara lain: 1) layanan *locker*; 2) layanan sirkulasi; 3) layanan referensi; 4) layanan penelusuran informasi; 5) layanan informasi koleksi terbaru; 7) layanan ruang baca; 8) layanan fotokopi; 9) layanan *workstation* dan multimedia. Layanan sirkulasi menurut Darmono (2007), adalah satu kegiatan di perpustakaan yang melayani peminjaman dan pengembalian buku. Layanan sirkulasi adalah kegiatan melayani peminjam dan pengembalian buku perpustakaan (Bafadal dalam Elnadi, 2018). Dari kedua definisi tersebut dapat disimpulkan bahwa layanan sirkulasi adalah kegiatan di perpustakaan yang melayani pemustaka dalam melakukan peminjaman dan pengembalian bahan pustaka atau koleksi. Layanan sirkulasi merupakan layanan pokok yang dimiliki oleh perpustakaan, karena berhubungan langsung dengan koleksi

perpustakaan. Dari beberapa layanan perpustakaan yang disebutkan oleh Rahayuningsih, semuanya dapat menghasilkan data. Akan tetapi jika ditengok lebih jauh lagi kegiatan yang merupakan inti dari perpustakaan terdapat pada layanan sirkulasi, yang di dalamnya dapat menghasilkan data peminjaman dan data pengembalian.

Kegiatan dalam perpustakaan jika dilihat sejak pengunjung pertama kali masuk ke perpustakaan adalah pencatatan data pengunjung, akan tetapi jika pengunjung tersebut menghendaki menjadi anggota dalam perpustakaan tersebut maka dibuatlah data anggota dengan cara menambahkan data diri ke dalam penyimpanan data anggota. Setelah menjadi anggota, bila pemustaka tersebut ingin meminjam buku, maka disimpanlah data ke dalam data peminjaman. Sesuai dengan aturannya jika melakukan peminjaman maka pemustaka harus mengembalikan buku atau koleksi. Proses pengembalian koleksi pun tercatat dalam data pengembalian buku. Selain proses sirkulasi ini, ada juga proses pengadaan koleksi yang pastinya dapat menghasilkan banyak data.

Menurut Pujianto (2018) beberapa manfaat *big data* sudah dirasakan khususnya bagi dunia usaha diantaranya untuk mengetahui respon masyarakat terhadap produk-produk yang dikeluarkan melalui analisis di media sosial; membantu perusahaan mengambil keputusan secara lebih tepat dan akurat berdasarkan data yang diolah; membantu citra perusahaan di mata pelanggan; untuk perencanaan usaha dengan mengetahui pola perilaku pelanggan, seperti pada perusahaan telekomunikasi dan perbankan; serta mengetahui trend pasar dan keinginan konsumen. Teknologi *big data* dapat membantu perusahaan dapat mengenali pola perilaku pelanggan melalui struk transaksi pembelian. Data dari setiap struk transaksi yang berisi kombinasi produk-produk yang dibeli, jumlah dan harganya. Seluruh data transaksi tersebut kemudian dicari pola belanja untuk menjawab beberapa pertanyaan: kombinasi dua atau tiga produk apa saja yang paling sering dibeli oleh pembeli. Dari hasil informasi tersebut ada beberapa tindakan yang dapat dilakukan, diantaranya yaitu: menyusun rak belanja agar dua atau tiga produk tersebut saling berdekatan sehingga mudah dijangkau oleh konsumen dan dapat diputuskan untuk dibeli dengan cepat. Kemudian bisa membuat paket promosi dimana kombinasi produk-produk tersebut dijual lebih murah, atau produk yang saling berkaitan tersebut sengaja dijauhkan supaya pada saat konsumen membeli produk A, dia juga membeli produk B, sehingga saat konsumen sudah mendapatkan produk A, dalam perjalanan mencari produk B, konsumen dapat membeli produk-produk yang lainnya.

Jika dibandingkan dengan apa yang terjadi di perpustakaan maka di dalam perpustakaan pun bisa dilakukan analisa melalui data peminjaman yang dilakukan oleh anggota, dari data tersebut dapat diketahui pola perilaku peminjaman anggota melalui *history* peminjaman, dari sana terdapat keterkaitan antar koleksi yang biasa dipinjam, ini dinamakan *association rule* yang isinya biasanya adalah jika seseorang meminjam koleksi A, maka ada kecenderungan meminjam koleksi B juga. Hal ini nantinya dapat dilakukan sebuah keputusan misalnya adalah menempatkan kedua koleksi tersebut secara berdekatan selama beberapa minggu misalnya, kenapa beberapa minggu saja, karena dalam perpustakaan sudah ada aturan sendiri jika

koleksi bidang A maka akan masuk ke dalam klasifikasi tertentu, dan jika koleksi B berbeda bidang dengan koleksi A maka letak penyimpanan koleksi tersebut pun pasti juga berbeda.

### 3. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah kualitatif deskriptif. Penelitian dilakukan dengan cara mengumpulkan informasi dari berbagai sumber, kemudian menelaah beberapa kegiatan perpustakaan yang termasuk contoh dalam *big data* kemudian dijelaskan secara deskriptif mengenai *data mining* pada contoh data tersebut.

### 4. Hasil dan Pembahasan

Konsep dari sebuah *big data* yang sudah dipaparkan dalam landasan teori. Disebutkan bahwa *big data* memiliki tiga karakteristik di dalamnya, *volume*, *velocity*, dan *variety*. Data dapat dikatakan sebagai data yang besar jika jumlahnya sangat banyak, diciptakan dalam waktu yang cepat, dan memiliki keragaman data. Jika dilihat dari karakteristik tersebut, salah satu contoh data dalam perpustakaan yang dapat dikatakan sebagai *big data* adalah data peminjaman pada perpustakaan yang peminjamnya setiap harinya sangat banyak. *Volume* merupakan alasan kenapa data peminjaman pada perpustakaan yang jumlah peminjamnya sangat banyak. Data peminjaman ini akan cepat sekali diciptakan, ini merupakan karakteristik dari *velocity*. Data peminjaman di dalamnya terdiri dari banyak jenis data, seperti nama peminjam, nomor anggota, nama koleksi yang dipinjam, jika itu merupakan sebuah buku maka di dalamnya terkandung informasi judul buku, tahun terbit, nama pengarang, kota terbit, dan lain sebagainya, hal ini menunjukkan karakteristik V yang ketiga yaitu *variety*.

Sebenarnya ada banyak sekali data dalam perpustakaan, misal data anggota, data pengadaan koleksi, data peminjaman, data pengunjung, dan lain sebagainya. Tapi perlu dikaji lebih lagi apakah semua data yang ada di dalam perpustakaan tersebut dapat disebut sebagai *big data* atau tidak. Data anggota misalnya, jika hanya data anggota yang mungkin setiap harinya belum tentu ada anggota baru, maka dirasa kurang tepat jika dinamakan *big data*, hal ini bisa dilihat dari *velocity*, pergerakan datanya harus cepat, penciptaan datanya harus cepat. Jadi meskipun ada banyak jenis data yang terdapat dalam sebuah perpustakaan, belum tentu data tersebut disebut sebagai big data.

Data-data yang dapat digolongkan menjadi *big data* biasanya dapat diolah sehingga menghasilkan informasi tertentu, informasi ini bisa dimanfaatkan oleh pihak perpustakaan untuk melakukan suatu aksi yang dapat menguntungkan pihak perpustakaan. Salah satu pengolahan *big data* adalah menggabungkannya dengan *data mining* atau penambangan data. Penambangan data ini dapat menghasilkan suatu informasi, tergantung kelompok mana yang akan dipilih oleh pihak perpustakaan dalam hal ini adalah pustakawan. Seperti pada landasan teori sebelumnya yaitu bisa dipilih kelompok *data mining* mana yang akan dilakukan antara deskripsi, estimasi, prediksi, klasifikasi, pengklusteran, dan asosiasi.

Asosiasi adalah salah satu pengelompokan data dalam *data mining* yang populer, biasanya asosiasi ini dilakukan dengan menggunakan data transaksi penjualan. Jadi misalnya data transaksi

penjualan pada bulan Januari 2020 misalkan ada ribuan transaksi pembelian atau lebih, maka dari data tersebut dapat diolah menggunakan suatu algoritma, dan kemudian menghasilkan *decision tree* dan *association rule*. Contoh nama algoritma yang sering dipakai dalam membentuk *decision tree* adalah ID3, atau juga bisa menggunakan algoritma Apriori untuk dapat menghasilkan *association rule*. Konsep *decision tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utama dari penggunaan *decision tree* adalah kemampuannya untuk membuat proses pengambilan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan. Kedua algoritma ini sama-sama melakukan perhitungan dalam mengolah data, akan tetapi memiliki cara perhitungan masing-masing dalam menghasilkan *output*-nya.

Seorang pustakawan dapat membuka wawasannya mengenai *big data*. mereka harus dapat memilah mana yang dapat disebut sebagai *big data* dan tidak, menentukan akan diolah seperti apa *big data* yang sudah dimiliki, kemudian dapat memutuskan kebijakan atau memberi masukan kebijakan terkait hasil yang didapatkan dari pengolahan penambangan data dari *big data* yang sudah dimiliki. Jangan sampai *big data* yang sudah dimiliki hanya dibiarkan begitu saja, sehingga tidak meningkatkan nilai dari informasi yang terkandung dalam data perpustakaan.

#### 4. Simpulan

*Big data* adalah kumpulan data dengan jumlah yang sangat banyak, yang jika diolah atau ditambang akan menghasilkan informasi, informasi yang diberikan pun juga sangat beragam, tergantung dari penambang informasi tersebut. *Big data* dapat diolah menggunakan beberapa teknik *data mining* salah satunya adalah asosiasi. Asosiasi dalam *data mining* mencari keterkaitan antara dua atau lebih barang, yang biasanya diterapkan dalam transaksi penjualan suatu barang. Pustakawan dapat menggunakan *big data* dalam perpustakaan, dalam hal ini bisa dipilih misalkan adalah data peminjaman koleksi untuk dapat diolah dengan menggunakan kelompok *data mining* jenis asosiasi untuk mencari keterkaitan antar koleksi yang dipinjam, sehingga dari hasil ini pustakawan dapat memutuskan akan dilakukan tindakan apa kemudian demi kemajuan perpustakaannya.

#### Daftar Pustaka

- Ali, Iramni. 2015. *Big Data: Apa dan Pengaruhnya Pada Perpustakaan?*. Jurnal Media Pustakawan. Vol.22, No.4
- Bafadal, Ibrahim. 1996. *Pengelolaan Perpustakaan*. Jakarta : Bumi Aksara
- Darmono. 2007. *Perpustakaan Sekolah: Pendekatan Aspek Manajemen dan Tata Kerja*, Jakarta: Grasindo
- Elnadi, Isran. 2018. Upaya Meningkatkan Layanan Pemustaka di UPT Perpustakaan Universitas Bengkulu. *JUPI (Jurnal Ilmu Perpustakaan dan Informasi)* Vol. 3 No. 2

- Inmon, William H. 2005. *Building Data Warehouse*. Fourth Edition. Canada: Wiley Publishing, Inc.
- Kadir, Abdul. 2009. *Dasar – Dasar dan Implementasi Database Relasional*. Yogyakarta: Andi Publisher
- Larose, Daniel T. 2005. *Discovering Knowledge in Data : an Introduction to Data Mining*. Canada: John Wiley & Sons
- Pujianto, Agung, Awin Mulyati, Rachmawati Novaria. *Pemanfaatan Big Data dan Perlindungan Privasi Konsumen di Era Ekonomi Digital*
- Rahayuningsih, F. 2007. *Pengelolaan Perpustakaan*. Yogyakarta: Graha Ilmu
- Sirait, Emyana Ruth Eritha. 2016. *Implementasi Teknologi Big Data di Lembaga Pemerintahan Indonesia*. *Jurnal Penelitian Pos dan Informatika*. Vol. 6, No. 2