

**PROPOSAL PROYEK PENAMBANGAN DATA**  
**Sistem Rekomendasi Buku Menggunakan Algoritma**  
**Apriori**



**Disusun oleh:**

**12S17005 Kiky Purnamasari Napitupulu**

**12S17006 Tripheni Simanjuntak**

**12S17023 Jessycha Royanti Tampubolon**

**PROGRAM STUDI SARJANA SISTEM INFORMASI**  
**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**  
**INSTITUT TEKNOLOGI DEL**  
**NOVEMBER 2020**

## Daftar Isi

Daftar Isi.....	1
1. Business Understanding .....	2
1.1 Determine Business Objectives .....	2
1.2 Situation Assessment .....	3
1.3 Determine Data Mining Goal .....	3
1.4 Produce Project Plan .....	3
2. Data Understanding.....	5
2.1 Collect Initial Data.....	5
2.2 Describe Data .....	5
2.3 Explore Data.....	5
2.4 Verify Data Quality .....	6
Referensi.....	7

# 1. Business Understanding

## 1.1 Determine Business Objectives

Informasi merupakan sekumpulan data atau fakta yang telah diproses dan dikelola sedemikian rupa sehingga menjadi sesuatu yang mudah dimengerti dan bermanfaat kepada pembaca. Seiring berjalannya waktu, informasi akan semakin banyak. Informasi yang dibutuhkan dapat berupa jurnal, artikel, buku dan saling berkaitan satu dengan yang lainnya. Untuk membantu mencari informasi mengenai buku yang akan dicari oleh pembaca, dibutuhkan sebuah sistem rekomendasi buku. Sistem rekomendasi buku ini akan sangat membantu dan dapat dijadikan solusi bagi pengguna dalam merekomendasikan sebuah informasi yang relevan mengenai buku yang akan dicari oleh pembaca.

Banyaknya jumlah buku yang dapat diperoleh, terkadang membuat pembaca kesulitan dalam menentukan buku yang hendak mereka cari. Dalam sebuah kasus, terkadang dijumpai pembaca hanya ingin membaca buku-buku yang dengan memiliki nilai penjualan yang terbaik. Namun, ada pula pembaca yang hanya ingin membaca buku yang mirip dengan buku-buku yang sudah pernah dibaca sebelumnya. Bahkan tidak jarang ditemui pembaca yang hanya akan membaca yang menentukan buku-buku yang akan dibaca selanjutnya berdasarkan *rating* dari buku-buku yang telah dilihat oleh pembaca. Semakin tinggi *rating* dari buku yang akan dicari, maka semakin tertarik pula pembaca untuk membacanya. Untuk menentukan tinggi rendahnya *rating* terhadap buku yang akan dicari oleh pembaca, dapat mempengaruhi buku-buku yang akan direkomendasikan oleh sistem yang akan dibangun. Nilai kemiripan antar buku dan *rating* buku dapat dijadikan sebagai landasan untuk memberikan rekomendasi buku kepada pembaca.

Sistem rekomendasi buku yang akan dibangun, akan memberikan solusi terhadap permasalahan dalam menentukan buku yang belum pernah dibaca oleh pembaca. Sistem rekomendasi buku menggunakan algoritma *apriori* diharapkan dapat membantu pembaca buku untuk menentukan buku yang layak dibaca dan buku yang tidak layak untuk dibaca berdasarkan *rating* dari buku tersebut. Penentuan rekomendasi buku dengan metode ini ditentukan berdasarkan fitur yang dimiliki oleh sistem yang memungkinkan pembaca nya untuk memberikan *rating* atau nilai terhadap buku-buku yang telah dibaca sebelumnya.

## 1.2 Situation Assessment

Orang - orang dengan minat membaca yang tinggi, jumlahnya cukup banyak, mulai dari kalangan pelajar maupun umum, sehingga mereka membutuhkan referensi atau rekomendasi buku untuk dibaca. Dengan adanya banyak sekali jumlah buku, sistem rekomendasi buku dapat menjadi alternatif untuk orang - orang yang bingung dalam menentukan buku yang ingin dibaca. Dengan memanfaatkan algoritma apriori dalam menentukan rekomendasi buku, maka akan memudahkan dalam menentukan buku yang ingin direkomendasikan pada pembaca. Batasan dari proyek ini adalah dataset ini dikumpulkan dari komunitas Book-Crossing pada Agustus-September 2004 [1]. Penentuan keputusan dalam pembuatan strategi sistem rekomendasi buku adalah berfokus pada data rating buku.

## 1.3 Determine Data Mining Goal

Tujuan penerapan Data Mining ini adalah sebagai berikut :

1. Menghasilkan model sistem rekomendasi buku berdasarkan data yang sudah tersedia.
2. Menerapkan algoritma Apriori dalam pemodelan sistem rekomendasi buku yang akan dibangun.

## 1.4 Produce Project Plan

Proyek ini akan menggunakan algoritma *Apriori* pada *python*. Algoritma apriori adalah satu algoritma dasar untuk menemukan *frequent itemsets* yang memenuhi *minsup* kemudian mendapatkan *rule* yang memenuhi *minconf* dari *frequent itemset* tadi. Bila *itemset* digolongkan sebagai *frequent itemset*, yang memiliki *support* lebih dari yang ditetapkan sebelumnya, maka semua subsetnya juga termasuk golongan *frequent itemset*, dan sebaliknya. Algoritma apriori akan menghasilkan kandidat baru k-itemset dari *frequent itemset* pada langkah sebelumnya dan menghitung nilai *support* k-itemset tersebut. Itemset yang memiliki nilai *support* di bawah dari *minsup* akan dihapus. Algoritma berhenti ketika tidak ada lagi *frequent itemset* baru yang dihasilkan [2].

Ide utama pada algoritma ini adalah [3]:

1. Mencari *frequent itemset* (himpunan item-item yang memenuhi minimum *support*)
2. Menghilangkan itemset dengan frekuensi yang rendah berdasarkan level minimum *support* yang telah ditentukan sebelumnya.

Proses selanjutnya adalah membangun aturan-aturan asosiasi dari itemset yang memenuhi nilai minimum *confidence*.

Untuk membentuk kandidat *itemset* ada dua proses utama yang dilakukan oleh pada algoritma apriori, yaitu [3]:

1. *Join Step* (Penggabungan)

Setiap item dikombinasikan dengan item lainnya sampai tidak terbentuk kombinasi lagi.

2. *Prune Step* (Pemangkasan)

Hasil dari item yang dikombinasikan kemudian dipangkas dengan aturan yang telah ditetapkan diatas.

## 2. Data Understanding

### 2.1 Collect Initial Data

Data tersedia dalam format Csv dan dapat diunduh untuk penggunaan non komersial karena disediakan secara *open source*. Dataset dapat di download di link berikut : <http://www2.informatik.uni-freiburg.de/~ctiegle/BX/>

### 2.2 Describe Data

Data merupakan sebuah bahan mentah yang dapat digunakan sebagai bahan utama yang mendorong pertumbuhan bisnis dan penambangan data yang disimpan secara elektronik serta pencarian secara otomatis dengan komputer [4]. Saat ini, kebutuhan akan data semakin tinggi baik dalam bidang sains maupun teknologi, hal ini dikarenakan informasi yang diperoleh setiap waktu adalah hasil pengolahan dari kumpulan data yang telah diolah. Data diciri-cirikan sebagai fakta yang kemudian akan diolah menjadi informasi yang berisi kumpulan pola dari data atau ekspektasi yang mendasari data. Himpunan dari data disebut dengan *dataset*.

Dalam proyek ini, *dataset* yang digunakan dalam membangun sistem rekomendasi buku menggunakan algoritma Apriori adalah Book-Crossing Dataset. Book-Crossing Dataset merupakan kumpulan data buku yang dikumpulkan oleh Cai-Nicolas Ziegler dari komunitas Book Crossing dan terdiri dari 278.858 ribu data *user* yang disamakan serta 1.149.780 data *rating* baik eksplisit maupun implisit dari 271.379 data buku [5].

### 2.3 Explore Data

*Dataset* memiliki nilai fitur atau atribut yang juga terdiri kumpulan atau koleksi dari semua bahan mentah data yang dikumpulkan dengan melalui metode penelitian data. Kumpulan data ini akan didistribusikan kepada pihak lain yang ingin menggunakannya sebagai masukan dalam penelitian dan hampir keseluruhan dapat diakses secara publik. Pada *dataset* Book-Crossing terdapat tiga tabel yaitu tabel BX-Users, BX-Books, dan BX-Book-Ratings. Tabel BX-Users merupakan tabel yang terdiri dari data pengguna seperti *user ID* yang dipetakan ke dalam tipe data *integer* dan data demografi pengguna seperti *location* dan *age* yang berisi nilai Null-Values. Pada tabel BX-Books terdapat data buku yang diidentifikasi dengan kode ISBN yang valid, *Book-Title*, *Book-Author*, *Year-Of-Publication*, *Publisher*, dan beberapa alamat URL Amazon Website yang merujuk pada gambar halaman buku dalam tiga

jenis ukuran (*small* (S), *medium* (M), dan *large* (L)). Sedangkan pada tabel BX-Book-Ratings terdiri dari informasi *rating* buku (Book-Rating) yang ditampilkan dalam skala 1-10 secara eksplisit dan 0 secara implisit.

## **2.4 Verify Data Quality**

Pada dasarnya *dataset* Book-Crossing merupakan *dataset* yang memiliki tingkat *sparsity* yang tinggi sehingga ketika ingin mengolah *dataset* Book-Crossing dilakukan proses *condensation* [5]. Proses ini akan lebih memadatkan data dengan menghapus semua data buku yang memiliki taksonomi yang buruk sehingga mendapatkan hasil penambahan data yang lebih baik. Berdasarkan hasil proses *condensation* yang telah dilakukan pada penelitian sebelumnya maka dihasilkan dimensi *dataset* yang lebih layak dimana data yang ditampilkan adalah 10.339 data pengguna, 6.708 buku, 361.349 *rating*.

## Referensi

- [1] E. Çano and M. Morisio, “Characterization of Public Datasets for Recommender Systems,” *IEEE*, pp. 1–11, 2015.
- [2] N. Wandu and R. A. H. A. Mukhlason, “Pengembangan Sistem Rekomendasi Penelusuran Buku dengan Penggalan Association Rule Menggunakan Algoritma Apriori (Studi Kasus Badan Perpustakaan dan Kearsipan Provinsi Jawa Timur),” *J. Tek. ITS*, vol. 1, pp. 445–449, 2012.
- [3] S. Saefudin and D. Fernando, “Penerapan Data Mining Rekomendasi Buku Menggunakan Algoritma Apriori,” *JSiI (Jurnal Sist. Informasi)*, vol. 7, p. 50, 2020.
- [4] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining Practical Machine Learning Tools and Techniques*. San Francisco: Elsevier, 2005.
- [5] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, “Improving recommendation lists through topic diversification,” in *Proceedings of the 14th international conference on World Wide Web - WWW '05*, 2005, p. 22, doi: 10.1145/1060745.1060754.