

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*Podcast* merupakan konten audio *on-demand* yang telah berkembang menjadi salah satu *channel* informasi, hiburan dan iklan. (Yang, et al., 2019). *User* mendengar *podcast* melalui hasil *download* atau berlangganan ke *feed podcast*. Diestimasi ada 90 juta *user* di Amerika Serikat pernah mendengar *podcast* setiap bulannya, naik 20 juta atau 6% dari tahun 2018 (Edison Research and Triton Digital, 2019). Pada April 2018, Apple menginformasikan bahwa ada 525.000 *podcast* dengan 18,5 juta episode yang tersedia di iTunes (Locker, 2018). Karena jumlah dan heterogenitas konten *podcast* yang besar, menemukan *podcast* yang sesuai dengan preferensi merupakan hal yang menantang. Mengkategorikan atau memberikan *tag* pada *podcast* sehingga dapat dimanfaatkan pada sistem rekomendasi pada umumnya dalam menghasilkan *podcast* sesuai preferensi *user* juga merupakan hal yang sulit untuk dilakukan sebagai akibat dari metadata seperti ID3 *tags* (*genre*, *artis*, dll) yang tidak lengkap (Xing, et al., 2016).

Sistem rekomendasi *podcast* yang ada pada saat ini menerapkan metode Natural Language Processing (NLP) dengan memanfaatkan Conditional Random Field (CRF) tokenizer dan Skip-gram Negative Sampling (SGNS) untuk mempelajari fundamental kata dan hubungan antar kata yang terdapat pada deskripsi *podcast*. Namun metode yang diajukan hanya dapat digunakan pada *podcast* berbahasa Mandarin (Xing, et al., 2016). Adapun beberapa layanan *music streaming* yang menyediakan *podcast* seperti Spotify<sup>1</sup>, Soundcloud<sup>2</sup>, maupun Google Podcast<sup>3</sup> tidak menyediakan fitur rekomendasi *podcast* yang sesuai dengan preferensi *user* dan mengharuskan *user* mencari *podcast* melalui fitur *search* atau kategori. Podible<sup>4</sup>, salah satu *website* penyedia *podcast* memberikan fitur pencarian yang cukup baik dan rekomendasi yang berdasarkan pada *history user*. Namun Podible<sup>4</sup> hanya beroperasi pada platform *web* dan *iOS*.

Walaupun dengan terbatasnya metadata *podcast*, sistem rekomendasi *podcast* dapat dikembangkan dengan menggunakan metode *n-gram* dan *term frequency*. Nascimento, dkk memanfaatkan metode *n-gram* dan *term frequency* untuk menghasilkan sistem rekomendasi makalah penelitian walaupun dengan keterbatasan metadata. Dengan memanfaatkan metadata

---

<sup>1</sup> <https://www.spotify.com>

<sup>2</sup> <https://soundcloud.com/>

<sup>3</sup> <https://play.google.com/store/apps/details?id=com.google.android.apps.podcasts&hl=en>

<sup>4</sup> <https://podible.co/>

judul dan abstrak, metode *n-gram* dan *term frequency* digunakan untuk menghasilkan *query* makalah penelitian yang akan dijadikan sebagai *input* pada sistem rekomendasi. Metode *n-gram* dan *term frequency* juga menunjukkan hasil yang lebih baik dibandingkan dengan metode *n-gram & term position*, *noun phrase & term frequency* dan *noun phrase & term position* yang juga dapat digunakan untuk menghasilkan *query*. Selain pemanfaatan *n-gram* dan *term frequency*, rekomendasi akan dihasilkan melalui pemanfaatan algoritma rekomendasi *content-based* (Nascimento, et al., 2011). Berdasarkan uraian di atas, maka dilakukan penelitian dengan mengimplementasikan metode *n-gram* dan *term frequency* dengan metadata yang tersedia pada *podcast* yaitu judul dan deskripsi untuk menghasilkan rekomendasi *podcast* yang sesuai dengan preferensi *user*. Penelitian ini akan dituangkan dalam tugas akhir dengan judul “**Aplikasi Rekomendasi Podcast Berbasis Mobile & Web Menggunakan Metode N-Gram dan Term Frequency**”.

## 1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, masalah dalam penulisan tugas akhir adalah kesulitan *user* dalam menemukan *podcast* yang sesuai dengan preferensi sebagai akibat dari jumlah *podcast* yang besar dan minimnya metadata yang tersedia.

## 1.3 Tujuan

Tujuan dari penulisan tugas akhir ini adalah mengembangkan aplikasi yang akan menghasilkan rekomendasi *podcast* berbasis *mobile* dan web dengan mengimplementasi metode *n-gram* dan *term frequency*.

## 1.4 Manfaat

Manfaat tugas akhir ini adalah sebagai berikut:

1. Mempermudah *user* dalam mencari *podcast* yang sesuai dengan preferensi *user* tersebut.
2. Dapat dijadikan sebagai bahan referensi untuk penelitian sistem rekomendasi yang lain.

## 1.5 Batasan Masalah

Untuk mencegah pembahasan keluar dari topik dan mencegah adanya perluasan masalah, maka dilakukan pembatasan masalah sebagai berikut:

1. *Dataset* yang digunakan merupakan *offline dataset* yang dibagikan di Kaggle<sup>5</sup> dengan rincian:
  - a. Episodes.csv<sup>6</sup> dengan 881.046 episode acak.
  - b. podcasts.csv<sup>7</sup> dengan 121.175 judul *podcast*.
2. *Dataset* yang digunakan dalam sistem rekomendasi merupakan *dataset* dengan judul dan deskripsi dalam bahasa Inggris dan huruf latin.
3. Sistem operasi yang digunakan pada platform *mobile* adalah Android dengan minimal *SDK* yang digunakan adalah *SDK API Level 23* yaitu *Marshmallow* (6.0) keatas.
4. Pengembangan aplikasi pada platform *mobile* akan memanfaatkan *library* Volley versi 1.1.1, Gson versi 2.8.5, Glide versi 4.8.0 dan Firebase.
5. *Member* harus memilih minimal 3 kategori yang menjadi preferensinya setelah *Register*.
6. *N* yang digunakan dalam *N-gram* adalah 2.

## 1.6 Metodologi Pengembangan Sistem

Metodologi yang digunakan dalam pengembangan sistem ini adalah metodologi *Waterfall* (Sommerville, 2011) dengan rincian sebagai berikut:

### 1. *Requirements analysis and definition*

Analisis terbagi menjadi 2 bagian yaitu:

#### a. Analisis Proses

Analisis proses metode *n-gram*, term frequency dan metode rekomendasi yang dikembangkan menggunakan *flowchart*.

#### b. Analisis kebutuhan

Analisis kebutuhan fungsional dilakukan dengan pemanfaatan *Use Case Diagram* untuk menggambarkan interaksi antara pengguna dengan sistem. Analisis kebutuhan non fungsional memanfaatkan kerangka *PIECES* (*Performance, Information, Economic, Control, Efficiency, Service*).

### 2. *System and software design*

Perancangan desain antar muka berdasarkan kepada diagram yang dianalisis sebelumnya dilakukan dalam bentuk *mockup* dengan memanfaatkan *tool* Balsamiq Mockup. Selain perancangan desain antar muka, perancangan model basis data akan dilakukan dengan menggunakan *Schema Diagram* dengan *tool* Moon Modeler.

<sup>5</sup> <https://www.kaggle.com/listennotes/all-podcast-episodes-published-in-december-2017>

<sup>6</sup> <https://www.kaggle.com/listennotes/all-podcast-episodes-published-in-december-2017#episodes.csv>

<sup>7</sup> <https://www.kaggle.com/listennotes/all-podcast-episodes-published-in-december-2017#podcasts.csv>

### 3. *Implementation and unit testing*

Pengembangan aplikasi dilakukan dengan menggunakan beberapa *tools* seperti:

- a. Pengembangan aplikasi web akan memanfaatkan bahasa pemrograman Javascript dengan *framework* Express untuk *backend* dan *framework* React untuk *frontend*.
- b. Pengembangan aplikasi *mobile* akan memanfaatkan bahasa pemrograman Java pada Android Studio.
- c. *Database Management System* yang digunakan adalah MongoDB.

### 4. *Integration and system testing*

Pengujian sistem rekomendasi dengan metode *n-gram* dan *term frequency* dilakukan dengan menggunakan perhitungan *Recall*, *Precision* dan *NDCG* (*Normalized Discounted Cumulative Gain*).

