**FINAL PROJECT DATA MINING**

**TEXT CLASSIFICATION**

Dosen Pengampu : Al Ustadz Oddy Virgantara Putra, S.Kom., M.T,



**DISUSUN OLEH :**

**FATHIN MUHAMMAD WASMANSON**

**NIM: 372016611508**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**UNIVERSITAS DARUSSALAM GONTOR**

**2019**

1. **Latar Belakang Masalah**

Klasifikasi teks merupakan sebuah pemisahan data pada sebuah teks yang atau artikel yang berbentuk teks. Ini dimaksudkan untuk mempermudah kita dalam menganalisis informasi yang ada dalam suatu teks baik itu jenis, sifat dan lain lain. Semua itu dapat dengan mudah kita pilah hanya dengan menggunakan metode klasifikasi ini. Khususnya pada sistem yang memiliki aliran data yang besar, metode ini sangatlah amat berguna, untuk mempelajari pola data atau informasi yang ada.

Metode ini berjalan diawali oleh text processing dan kemudian diklasifikasikan. Dari sini penulis ingin mengklasifikasikan berdasarkan review sebuah film terbaru yang berjudul shazam. Dengan mengambil 20 sampel dari website imdb yang diambil secara acak. Film ini menuai kontroversi karena bersifat anti mainstream membuat banyak penilaian beragam terhadap film ini.

1. **Rumusan Masalah**

Rumusan masalah yang dapat diambil dari latar belakang yang diangkat adalah kriteria film shazam ini berdasarkan klasifikasi review.

1. **Metode**

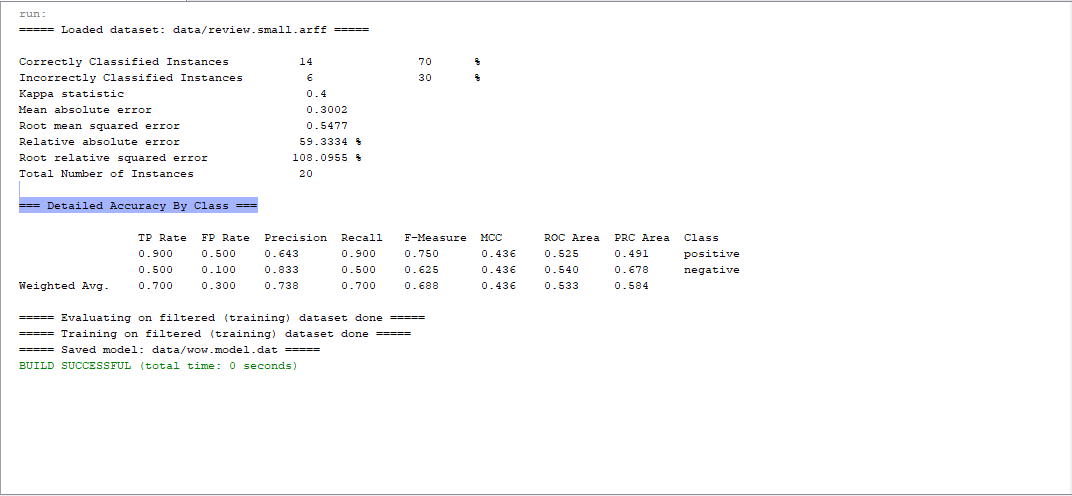
Dalam final project ini penulis menggunakan metode Text classification using Java dan algoritma Naive Bayes dengan tahapan :

Preprocessing

1. Tokenization
2. Stemming
3. Removing special characters for instance: (),.!?/'"!@#$%^&\*
4. Bag of Words

Text Classification steps:

1. Data Learning
2. Data Testing
3. **Hasil**
4. **Hasil Learning Naïve Bayes**

****

1. **Hasil Klasifikasi**

****

1. **Kesimpulan**

Kesimpulan yang dapat diambil dari permasalahan yang telah dibahas dan dijelaskan diatas adalah bahwa kriteria film shazam berdasarkan review acak adalah berkriteria **Positive.**