

**PENERAPAN METODE SYMMETRICAL UNCERTAINTY  
DAN SUPPORT VECTOR MACHINE DALAM  
MEMPREDIKSI TITLE BOX OFFICE FILM**

**TUGAS AKHIR**

**Shandy Hadinata  
1116043**



**PROGRAM STUDI INFORMATIKA  
INSTITUT TEKNOLOGI HARAPAN BANGSA  
BANDUNG  
2020**

**PENERAPAN METODE SYMMETRICAL UNCERTAINTY  
DAN SUPPORT VECTOR MACHINE DALAM  
MEMPREDIKSI TITLE BOX OFFICE FILM**

**TUGAS AKHIR**

**Diajukan sebagai salah satu syarat untuk memperoleh  
gelar sarjana dalam bidang Informatika**

**Shandy Hadinata  
1116043**



**PROGRAM STUDI INFORMATIKA  
INSTITUT TEKNOLOGI HARAPAN BANGSA  
BANDUNG  
2020**

## **HALAMAN PERNYATAAN ORISINALITAS**

**Saya menyatakan bahwa Tugas Akhir yang saya susun ini  
adalah hasil karya saya sendiri.**

**Semua sumber yang dikutip maupun dirujuk  
telah saya nyatakan dengan benar.**

**Saya bersedia menerima sanksi pencabutan gelar akademik  
apabila di kemudian hari Tugas Akhir ini terbukti plagiat.**

**Bandung, Agustus 2020**

**Shandy Hadinata**

**1116043**

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Institut Teknologi Harapan Bangsa, saya yang bertanda tangan di bawah ini:

Nama : Shandy Hadinata

NIM : 1116043

Program Studi : Informatika

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Harapan Bangsa **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Rights*) atas karya ilmiah saya yang berjudul:

PENERAPAN METODE SYMMETRICAL UNCERTAINTY DAN SUPPORT VECTOR MACHINE DALAM MEMPREDIKSI TITLE BOX OFFICE FILM

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Harapan Bangsa berhak menyimpan, mengalihmediakan, mengelola dalam pangkalan data, dan memublikasikan karya ilmiah saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Bandung, Agustus 2020

Yang menyatakan



Shandy Hadinata

## **HALAMAN PENGESAHAN TUGAS AKHIR**

Tugas Akhir dengan judul:

**PENERAPAN METODE SYMMETRICAL UNCERTAINTY DAN SUPPORT  
VECTOR MACHINE DALAM MEMPREDIKSI TITLE BOX OFFICE FILM**

yang disusun oleh:

Shandy Hadinata

1116043

telah berhasil dipertahankan di hadapan Dewan Penguji Sidang Tugas Akhir yang  
dilaksanakan pada:

Hari / tanggal : Jumat, 3 Juli 2020

Waktu : 07.00 WIB

**Menyetujui**

**Pembimbing Utama:**

**Ventje Jeremias Lewi Engel, S.T., M.T.**

**116019**

## ABSTRAK

Nama : Shandy Hadinata  
Program Studi : Informatika  
Judul : PENERAPAN METODE SYMMETRICAL  
UNCERTAINTY DAN SUPPORT VECTOR MACHINE  
DALAM MEMPREDIKSI TITLE BOX OFFICE FILM

Dalam industri perfilman, terdapat istilah *box office* yang merupakan gelar bagi film - film yang berhasil mendapatkan keuntungan besar yang melebihi anggaran pembuatan film hanya dari penjualan tiket. *Title box office* pada sebuah film perlu diperhatikan oleh rumah produksi agar rumah produksi mendapatkan keuntungan. Pada penelitian ini, klasifikasi menggunakan metode *Support Vector Machine* (SVM) dengan kernel *Radial Basis Function* (RBF) untuk membantu memprediksi sebuah film layak dapat *title box office* sebelum dirilis. Untuk mencegah memakai fitur yang nilainya tidak berpengaruh terhadap prediksi, akan digunakan metode *Symmetrical Uncertainty* sebagai seleksi fitur.

Penelitian ini akan menggunakan dataset film yang berasal dari IMDb yang memiliki fitur berjumlah 28 dan data film yang berjumlah 5044. Model terbaik dihasilkan dengan menggunakan *threshold Symmetrical Uncertainty* sebesar 0.1, *gamma* SVM sebesar 0.01, dan regularisasi (C) SVM sebesar 1. Akurasi dan *F-Measure* terbaik yang dihasilkan sebesar 99.42% dan 98.78%. Beberapa kesulitan dan hambatan berhubungan pada data film yang *noise*.

**Kata kunci:** Prediksi *Title Box Office* Film, Seleksi Fitur, Pembelajaran Mesin, *Symmetrical Uncertainty*, *Support Vector Machine* (SVM).

## ABSTRACT

*Name : Shandy Hadinata*  
*Department : Informatics*  
*Title : Application of Symmetrical Uncertainty and Support Vector Machine Method in Predict Title Box Office Movie*

*In the movie industry, there is a term called box office which is a title for movies that have managed to get huge profits that exceed the budget of movie making only from ticket sales. The box office title in a movie needs to be considered by a production house so that the production house gets profit. This research use classification method Support Vector Machine with Radial Basis Function (RBF) kernel to help predict a movie that deserves to get box office title before being released. To prevent using features whose values have no effect on predictions, Symmetrical Uncertainty method will be used as feature selection.*

*This research will use a movie dataset from IMDb which has 28 features and 5044 film data. The best model is generated using threshold Symmetrical Uncertainty of 0.1, gamma SVM of 0.01, and regularization (C) SVM of 1. The best accuracy and F-Measure are 99.42% and 98.78%. Some difficulty and limitation are related to noises in movie data.*

**Keywords:** *Prediction Title Box Office Film, Feature Selection, Machine Learning, Symmetrical Uncertainty, Support Vector Machine (SVM).*

## **KATA PENGANTAR**

Terima kasih kepada Tuhan yang Maha Esa karena dengan bimbingan-Nya dan karunia-Nya penulis dapat melaksanakan Tugas Akhir yang berjudul "PENERAPAN METODE SYMMETRICAL UNCERTAINTY DAN SUPPORT VECTOR MACHINE DALAM MEMPREDIKSI TITLE BOX OFFICE FILM ". Laporan ini disusun sebagai salah satu syarat kelulusan di Institut Teknologi Harapan Bangsa. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Tuhan Yang Maha Esa, karena oleh bimbingan-Nya penulis selalu mendapat pengharapan untuk menyelesaikan tugas akhir ini.
2. Bapak Ventje Jeremias Lewi Engel, S.T., M.T., selaku pembimbing I Tugas Akhir yang senantiasa memberi dukungan, semangat, ilmu-ilmu, saran dan dukungan kepada penulis selama tugas akhir berlangsung dan selama pembuatan laporan tugas akhir ini.
3. Ibu Ken Ratri Retno Wardani, S.Kom., M.T, selaku penguji I Tugas Akhir. Terima kasih atas dukungan, semangat, ilmu-ilmu, dan masukan yang telah diberikan kepada penulis dalam menyelesaikan Laporan Tugas Akhir ini
4. Bapak Yoyok Yusman Gamaliel, S.T., M.Eng, selaku penguji II dalam Tugas Akhir Terima kasih atas dukungan, semangat, ilmu-ilmu, dan masukan yang telah diberikan kepada penulis dalam menyelesaikan Laporan Tugas Akhir ini.
5. Seluruh dosen dan staff Departemen Teknik Informatika ITHB yang telah membantu dalam menyelesaikan Laporan Tugas Akhir ini.



6. Segenap jajaran staf dan karyawan ITHB yang turut membantu kelancaran dalam menyelesaikan Laporan Tugas Akhir ini.
7. Kedua orang tua tercinta yang selalu menyediakan waktu untuk memberikan doa, semangat dan dukungan yang tak habis-habisnya kepada penulis untuk menyelesaikan Laporan Tugas Akhir ini. Terima kasih untuk nasihat, masukan, perhatian, teguran dan kasih sayang yang diberikan hingga saat ini.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna karena keterbatasan waktu dan pengetahuan yang dimiliki oleh penulis. Oleh karena itu, kritik dan saran untuk membangun kesempurnaan tugas akhir ini sangat diharapkan. Semoga tugas akhir ini dapat membantu pihak-pihak yang membutuhkannya.

Bandung, Agustus 2020

Hormat Saya,

Penulis.

# DAFTAR ISI

|   |            |
|---|------------|
| <b>ABSTRAK</b>                                | <b>iv</b>  |
| <b>ABSTRACT</b>                               | <b>v</b>   |
| <b>KATA PENGANTAR</b>                         | <b>vi</b>  |
| <b>DAFTAR ISI</b>                             | <b>x</b>   |
| <b>DAFTAR TABEL</b>                           | <b>xii</b> |
| <b>DAFTAR GAMBAR</b>                          | <b>xiv</b> |
| <b>1 PENDAHULUAN</b>                          | <b>1-1</b> |
| 1.1 Latar Belakang . . . . .                  | 1-1        |
| 1.2 Rumusan Masalah . . . . .                 | 1-4        |
| 1.3 Tujuan Penelitian . . . . .               | 1-4        |
| 1.4 Batasan Masalah . . . . .                 | 1-5        |
| 1.5 Kontribusi Penelitian . . . . .           | 1-5        |
| 1.6 Metodologi Penelitian . . . . .           | 1-5        |
| 1.7 Sistematika Pembahasan . . . . .          | 1-6        |
| <b>2 LANDASAN TEORI</b>                       | <b>2-1</b> |
| 2.1 Tinjauan Pustaka . . . . .                | 2-1        |
| 2.1.1 <i>Data Mining</i> . . . . .            | 2-1        |
| 2.1.2 Klasifikasi . . . . .                   | 2-2        |
| 2.1.3 <i>Machine Learning</i> . . . . .       | 2-2        |
| 2.1.4 <i>Support Vector Machine</i> . . . . . | 2-3        |
| 2.1.5 <i>Data Preprocessing</i> . . . . .     | 2-8        |

|          |  |            |
|----------|--|------------|
| 2.1.6    | Seleksi Fitur . . . . .                              | 2-9        |
| 2.1.7    | <i>Symmetrical Uncertainty</i> . . . . .             | 2-9        |
| 2.1.8    | <i>K-Fold Cross Validation</i> . . . . .             | 2-11       |
| 2.1.9    | <i>F-Measure</i> . . . . .                           | 2-12       |
| 2.2      | Tinjauan Studi . . . . .                             | 2-13       |
| 2.2.1    | <i>State of the Art Method</i> . . . . .             | 2-13       |
| 2.2.2    | Pembahasan Penelitian Terkait . . . . .              | 2-14       |
| 2.3      | Tinjauan Objek . . . . .                             | 2-16       |
| 2.3.1    | Film . . . . .                                       | 2-16       |
| 2.3.2    | <i>Box Office</i> . . . . .                          | 2-16       |
| 2.3.3    | <i>Hollywood Accounting</i> . . . . .                | 2-17       |
| 2.3.4    | <i>International Movie Database (IMDb)</i> . . . . . | 2-18       |
| <b>3</b> | <b>ANALISIS DAN PERANCANGAN SISTEM</b>               | <b>3-1</b> |
| 3.1      | Analisis Masalah . . . . .                           | 3-1        |
| 3.2      | Kerangka Pemikiran . . . . .                         | 3-2        |
| 3.3      | Analisis Urutan Proses Global . . . . .              | 3-3        |
| 3.4      | Analisis Kasus . . . . .                             | 3-5        |
| 3.4.1    | <i>Dataset</i> . . . . .                             | 3-5        |
| 3.4.2    | <i>Data Preprocessing</i> . . . . .                  | 3-14       |
| 3.4.3    | <i>Symmetrical Uncertainty</i> . . . . .             | 3-18       |
| 3.4.4    | <i>Support Vector Machine</i> . . . . .              | 3-23       |
| <b>4</b> | <b>IMPLEMENTASI DAN PENGUJIAN</b>                    | <b>4-1</b> |
| 4.1      | Lingkungan Implementasi . . . . .                    | 4-1        |
| 4.1.1    | Spesifikasi Perangkat Keras . . . . .                | 4-1        |
| 4.1.2    | Lingkungan Perangkat Lunak . . . . .                 | 4-1        |
| 4.2      | Daftar Kelas dan Metode . . . . .                    | 4-2        |
| 4.2.1    | Kelas <i>Symmetrical Uncertainty</i> . . . . .       | 4-2        |

|          |  |             |
|----------|--|-------------|
| 4.2.2    | Kelas <i>Support Vector Machine</i> . . . . .  | 4-4         |
| 4.3      | Implementasi Perangkat Lunak . . . . .   | 4-6         |
| 4.3.1    | Implementasi Pemrosesan Data Film . . . . .  | 4-6         |
| 4.3.2    | Implementasi Seleksi Fitur . . . . .   | 4-8         |
| 4.3.3    | Implementasi Pelatihan Data . . . . .  | 4-9         |
| 4.3.4    | Implementasi Pengujian Data . . . . .  | 4-9         |
| 4.4      | Pengujian dan Evaluasi . . . . .   | 4-10        |
| 4.4.1    | Pengujian Seleksi Fitur <i>Symmetrical Uncertainty</i> . . . . .                                       | 4-10        |
| 4.4.2    | Pengujian Klasifikasi <i>Support Vector Machine</i> dengan<br><i>Symmetrical Uncertainty</i> . . . . . | 4-13        |
| 4.4.3    | Pengujian Klasifikasi <i>Support Vector Machine (Library)</i> . . . . .                                | 4-26        |
| 4.4.4    | Pengujian Pengaruh Jumlah Data Terhadap Klasifikasi<br><i>Support Vector Machine</i> . . . . .         | 4-29        |
| <b>5</b> | <b>PENUTUP</b>   | <b>5-1</b>  |
| 5.1      | Kesimpulan . . . . .   | 5-1         |
| 5.2      | Saran . . . . .  | 5-2         |
|          | <b>DAFTAR REFERENSI</b>  | <b>xv</b>   |
|          | <b>LAMPIRAN</b>  | <b>1</b>    |
| <b>A</b> | <b>PENGUJIAN SVM DENGAN SU</b>   | <b>A-1</b>  |
| <b>B</b> | <b>PENGUJIAN SVM TANPA SU</b>  | <b>B-26</b> |
| <b>C</b> | <b>PENGUJIAN JUMLAH DATA</b>   | <b>C-32</b> |

## DAFTAR TABEL

|      |   |      |
|------|---|------|
| 2.1  | Persamaan Kernel pada <i>Support Vector Machine</i> . . . . .                         | 2-6  |
| 2.2  | <i>Confusion Matrix</i> . . . . .   | 2-13 |
| 2.3  | <i>State of the Art Method</i> . . . . .  | 2-13 |
| 3.1  | Kode Fitur . . . . .  | 3-5  |
| 3.2  | Dataset . . . . .   | 3-7  |
| 3.3  | <i>Data Imputation</i> . . . . .  | 3-15 |
| 3.4  | Contoh Data untuk Mencari Interval . . . . .  | 3-17 |
| 3.5  | Hasil Interval . . . . .  | 3-18 |
| 3.6  | Hasil <i>Symmetrical Uncertainty</i> . . . . .  | 3-21 |
| 3.7  | Hasil Seleksi Fitur dengan <i>Symmetrical Uncertainty</i> . . . . .                   | 3-23 |
| 3.8  | Klasifikasi Data Pelatihan . . . . .  | 3-26 |
| 3.9  | Klasifikasi Data Pelatihan (Lanjutan) . . . . .                                       | 3-26 |
| 3.10 | Hasil dari Klasifikasi Data Pelatihan . . . . .                                       | 3-27 |
| 3.11 | Hasil dari Persamaan Linear . . . . .   | 3-28 |
| 3.12 | Klasifikasi Data Uji . . . . .  | 3-28 |
| 3.13 | Klasifikasi Data Uji (Lanjutan) . . . . .   | 3-28 |
| 3.14 | Hasil dari Klasifikasi Data Uji . . . . .   | 3-29 |
| 4.1  | Daftar Metode pada Kelas <i>Symmetrical Uncertainty</i> . . . . .                     | 4-2  |
| 4.2  | Daftar Metode pada Kelas <i>Support Vector Machine</i> . . . . .                      | 4-4  |
| 4.3  | Hasil Pengujian <i>Symmetrical Uncertainty</i> . . . . .                              | 4-10 |
| 4.4  | Hasil Pengujian <i>Symmetrical Uncertainty</i> dengan <i>Threshold</i> 0.01 . . . . . | 4-12 |
| 4.5  | Hasil Pengujian <i>Symmetrical Uncertainty</i> dengan <i>Threshold</i> 0.05 . . . . . | 4-12 |
| 4.6  | Hasil Pengujian <i>Symmetrical Uncertainty</i> dengan <i>Threshold</i> 0.1 . . . . .  | 4-13 |
| 4.7  | Hasil Pengujian <i>Support Vector Machine</i> . . . . .                               | 4-14 |

|      |  |      |
|------|--|------|
| 4.8  | Hasil Pengujian <i>Support Vector Machine</i> tanpa <i>Symmetrical Uncertainty</i> . . . . .       | 4-17 |
| 4.9  | Hasil Pengujian Pengaruh Jumlah Data dengan <i>Library Support Vector Machine</i> . . . . .        | 4-26 |
| 4.10 | Hasil Pengujian dengan <i>Library SVM</i> tanpa SU . . . . .                                       | 4-27 |
| 4.11 | Hasil Pengujian <i>Support Vector Machine</i> dengan Jumlah Data Film yang Berbeda . . . . .       | 4-29 |
| A-1  | Hasil Pengujian <i>Support Vector Machine</i> dengan Seleksi Fitur . . .                           | A-1  |
| B-1  | Hasil Pengujian <i>Support Vector Machine</i> tanpa Seleksi Fitur . . .                            | B-26 |
| C-1  | Hasil Pengujian Jumlah Data Film pada <i>Support Vector Machine</i> dengan Seleksi Fitur . . . . . | C-32 |

## DAFTAR GAMBAR

|     |   |      |
|-----|---|------|
| 2.1 | <i>Hyperplane</i> pada <i>Support Vector Machine</i> . . . . .  | 2-4  |
| 2.2 | <i>10-Fold Cross Validation</i> . . . . .   | 2-12 |
| 2.3 | <i>Box Office Statistics</i> . . . . .  | 2-17 |
| 3.1 | Kerangka Pemikiran . . . . .  | 3-2  |
| 3.2 | <i>Flowchart</i> Pelatihan dan Pengujian pada Pembuatan Sistem<br>Prediksi <i>Title Box Office</i> Film . . . . . | 3-3  |
| 3.3 | <i>Pseudocode Symmetrical Uncertainty</i> . . . . .   | 3-19 |
| 3.4 | <i>Pseudocode Support Vector Machine Training</i> . . . . .   | 3-24 |
| 3.5 | <i>Pseudocode Support Vector Machine Testing</i> . . . . .  | 3-25 |
| 4.1 | Hasil Akurasi dengan Nilai <i>Threshold Symmetrical Uncertainty</i><br>sebesar 0 . . . . .                        | 4-18 |
| 4.2 | Hasil <i>F-Measure</i> dengan <i>Threshold Symmetrical Uncertainty</i><br>sebesar 0 . . . . .                     | 4-18 |
| 4.3 | Hasil Akurasi dengan <i>Threshold Symmetrical Uncertainty</i> sebesar<br>0.01 . . . . .                           | 4-19 |
| 4.4 | Hasil <i>F-Measure</i> dengan <i>Threshold Symmetrical Uncertainty</i><br>sebesar 0.01 . . . . .                  | 4-19 |
| 4.5 | Hasil Akurasi dengan <i>Threshold Symmetrical Uncertainty</i> sebesar<br>0.05 . . . . .                           | 4-20 |
| 4.6 | Hasil <i>F-Measure</i> dengan <i>Threshold Symmetrical Uncertainty</i><br>sebesar 0.05 . . . . .                  | 4-20 |
| 4.7 | Hasil Akurasi dengan <i>Threshold Symmetrical Uncertainty</i> sebesar 0.1 . . . . .                               | 4-21 |
| 4.8 | Hasil <i>F-Measure</i> dengan <i>Threshold Symmetrical Uncertainty</i><br>sebesar 0.1 . . . . .                   | 4-21 |
| 4.9 | Hasil Akurasi Tanpa <i>Symmetrical Uncertainty</i> . . . . .  | 4-22 |

|  |      |
|--|------|
| 4.10 Hasil <i>F-Measure</i> Tanpa <i>Symmetrical Uncertainty</i> . . . . .                                 | 4-22 |
| 4.11 Hasil Akurasi dan <i>F-Measure</i> Berdasarkan <i>Threshold Symmetrical<br/>Uncertainty</i> . . . . . | 4-23 |
| 4.12 Hasil Akurasi dan <i>F-Measure</i> Berdasarkan <i>Gamma</i> . . . . .                                 | 4-24 |
| 4.13 Hasil Akurasi dan <i>F-Measure</i> Berdasarkan Regularisasi (C) . . . . .                             | 4-25 |
| 4.14 Hasil Klasifikasi dengan Jumlah Data yang Berbeda . . . . .   | 4-30 |



# BAB 1 PENDAHULUAN

## 1.1 Latar Belakang

Industri perfilman merupakan sektor industri yang bertumbuh dengan cepat dalam satu dekade terakhir karena dapat menghasilkan total keuntungan kotor sampai 12 miliar Dollar Amerika Serikat [1]. Dalam industri perfilman, terdapat istilah *box office* yang merupakan gelar bagi film - film yang berhasil mendapatkan keuntungan besar yang melebihi anggaran pembuatan film hanya dari penjualan tiket. *Title box office* pada sebuah film perlu diperhatikan oleh rumah produksi agar rumah produksi mendapatkan keuntungan. Keuntungan yang didapat merupakan keuntungan bersih yang melebihi anggaran pembuatan film dan biaya *overhead* atau biaya tak terduga. Terdapat tiga biaya *overhead*, yaitu *production overhead*, *distribution overhead*, dan *marketing overhead*. *Production overhead* adalah biaya produksi yang tidak diduga ketika pembuatan film, memakan 15% dari anggaran pembuatan film. *Distribution overhead* adalah biaya yang harus dikeluarkan kepada pihak penyedia layanan film (bioskop), memakan 30% dari anggaran pembuatan film. *Marketing overhead* adalah biaya yang dikeluarkan untuk promosi, iklan, dan sebagainya, memakan 10% dari anggaran pembuatan film [2].

Berdasarkan kebutuhan industri perfilman dalam menghasilkan sebuah film *box office*, maka dibuat sebuah sistem untuk memprediksi *title box office* sebuah film sebelum diproduksi. Diharapkan dengan membangun sistem prediksi akan mengurangi ambiguitas sebelum memproduksi sebuah film.

Ada beberapa penelitian yang sudah pernah dilakukan sebelumnya terkait sistem untuk memprediksi *title box office* film, yang pertama adalah penelitian yang berfokus pada prediksi kesuksesan film menggunakan metode *Decision Tree* dengan *Information Gain* dan Entropi sebagai metode untuk mengetahui

## 1. PENDAHULUAN

---

keragaman data pada setiap fitur [3]. *Decision Tree* dapat dengan mudah menghasilkan model prediksi namun tidak efisien jika data memiliki fitur yang banyak. Akurasi yang didapat adalah 66.7%.

Adapun penelitian lainnya [4] adalah memprediksi *title box office* film dengan metode yang digunakan adalah *K-Means* untuk mengetahui fitur yang berpengaruh terhadap prediksi, dilanjutkan dengan metode *Decision Tree* untuk menentukan model prediksi. Akurasi yang didapat adalah 71.4%.

Penelitian lainnya [5] adalah memprediksi *box office* film pada minggu kedua sampai keempat menggunakan metode *Rough Set* untuk seleksi fitur, selanjutnya menggunakan metode *Support Vector Machine* untuk menentukan model prediksi. Akurasi yang didapat dari minggu kedua sampai keempat secara berurutan adalah 90.7%, 84.3%, dan 86.7%.

Dari penelitian sebelumnya yang membahas terkait prediksi *title box office* film, terdapat banyak fitur film yang memiliki nilai yang tidak terlalu berpengaruh atau bahkan tidak berpengaruh terhadap prediksi *title box office* film sehingga dibutuhkan seleksi fitur untuk membuang fitur tersebut. Terdapat 2 jenis seleksi fitur, yaitu teknik *filter* dan *wrapper*. Teknik *filter* memiliki keunggulan dalam efisiensi waktu dan memiliki akurasi yang cukup bagus. Teknik *wrapper* memiliki akurasi yang lebih tinggi namun hanya berbeda sedikit dari teknik *filter*. Kelemahan lainnya adalah membutuhkan waktu yang sangat lama dalam memproses fitur dan memiliki performa yang buruk.

Penelitian terkait seleksi fitur berbasis *filter* [6] membandingkan seluruh metode seleksi fitur berbasis filter, yaitu *Chi Squared*, *Information Gain*, *Gain Ratio*, *Symmetrical Uncertainty*, dan *Relief Attribute Evaluation*. Percobaan dilakukan sebanyak lima kali dengan lima *dataset* yang berbeda. Rata-rata dari lima *dataset* *Kappa Statistic*, *Mean Absolute Error*, dan *Root Mean Square Error* dari lima metode seleksi fitur berbasis *filter* dan pengolahan data tanpa seleksi fitur secara

## 1. PENDAHULUAN

---

berurutan adalah *Chi Squared* sebesar 0.52232, 0.2844, dan 0.36332, *Gain Ratio* sebesar 0.52232, 0.2845, dan 0.36342, *Information Gain* sebesar 0.52166, 0.2844, dan 0.3633, *Relief Attribute Evaluation* sebesar 0.3936, 0.3023, dan 0.3841, *Symmetrical Uncertainty* sebesar 0.522, 0.28438, dan 0.36326, dan tanpa seleksi fitur sebesar 0.35224, 0.30364, dan 0.3872. Dari hasil penelitian yang dilakukan, *Symmetrical Uncertainty* memiliki *Kappa Statistic* tertinggi kedua dan *Mean Absolute Error* dan *Root Mean Square Error* terendah bila dibandingkan dengan semua metode seleksi fitur berbasis *filter*.

Penelitian ini menggabungkan metode *Symmetrical Uncertainty* dengan *Support Vector Machine* untuk memprediksi *title box office* sebuah film. *Symmetrical Uncertainty* merupakan metode seleksi fitur berbasis *filter* untuk normalisasi nilai *bias* pada setiap fitur [6]. Bila dibandingkan dengan metode klasifikasi lain yang cocok untuk sistem prediksi seperti *Artificial Neural Network* (ANN) atau *Decision Tree*, metode *Support Vector Machine* cocok digunakan untuk ukuran data yang sedikit dan permasalahan yang tidak terlalu besar seperti prediksi *title box office* film. *Support Vector Machine* juga menerapkan strategi *Structural Risk Minimization* yaitu berusaha mencari *hyperplane* atau pemisah antar kelas dengan margin terbesar untuk mengurangi nilai *error* secara signifikan. *Artificial Neural Network* menerapkan strategi *Empirical Risk Minimization* yang hanya berusaha meminimalisir *error*, maka dari itu lebih cocok untuk penelitian yang memiliki fitur data dan hasil prediksi yang sangat kompleks. *Decision Tree* mudah diimplementasikan untuk prediksi dengan hasil keluaran yang sedikit, namun model prediksi yang dihasilkan akan sulit dibaca dan memakan waktu yang lama jika terdapat banyak data dan fitur.

Berdasarkan hipotesis diatas, penulis membuat penelitian berjudul "*Penerapan Metode Symmetrical Uncertainty dan Support Vector Machine dalam Memprediksi Title Box Office Film*" dengan masukan data film berupa 28 fitur. Setelah metode

## 1. PENDAHULUAN

---

*Symmetrical Uncertainty* dan *Support Vector Machine* dilakukan, akan menghasilkan keluaran model prediksi *title box office* film. Untuk menguji akurasi *Support Vector Machine*, digunakan *F-Measure*.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas penulis merumuskan masalah sebagai berikut:

1. Bagaimana pengaruh *Symmetrical Uncertainty* sebagai metode seleksi fitur terhadap akurasi prediksi *title box office* film?
2. Bagaimana pengaruh *Support Vector Machine* terhadap akurasi prediksi *title box office* film?
3. Bagaimana pengaruh banyaknya film terhadap akurasi prediksi *title box office* film?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian Tugas Akhir ini adalah sebagai berikut:

1. Menerapkan *Symmetrical Uncertainty* sebagai metode seleksi fitur untuk menentukan fitur film yang berpengaruh terhadap prediksi *title box office* film.
2. Menguji pengaruh *Support Vector Machine* terhadap akurasi prediksi *title box office* film.
3. Menguji pengaruh dari banyaknya film terhadap akurasi prediksi *title box office* film.

### 1.4 Batasan Masalah

Dalam penelitian ini, peneliti akan membatasi masalah yang akan diteliti, yaitu *dataset* berupa data film IMDb yang diambil dari situs <https://www.kaggle.com> dengan tahun rilis sekitar 2006 sampai 2016.

### 1.5 Kontribusi Penelitian

Kontribusi yang diberikan dari penelitian ini adalah menerapkan *Support Vector Machine* dengan *Symmetrical Uncertainty* sebagai seleksi fitur untuk memprediksi *box office* sebuah film dan membuat aplikasi untuk menguji metode.

### 1.6 Metodologi Penelitian

Metodologi penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut:

#### 1. Studi Literatur

Penulisan ini dimulai dengan studi kepustakaan yaitu mengumpulkan bahan-bahan referensi baik dari buku, artikel, *paper*, jurnal, makalah mengenai prediksi *title box office* film, kesuksesan film menggunakan metode *Support Vector Machine* dan *Symmetrical Uncertainty* sebagai metode seleksi fitur.

#### 2. Data sampling

Data sampling yang akan digunakan berupa data film IMDb.

#### 3. Analisis Masalah

Pada tahap ini dilakukan analisis permasalahan yang ada, batasan yang dimiliki dan kebutuhan yang diperlukan.

#### 4. Perancangan dan Implementasi Algoritme

Pada tahap ini dilakukan pendefinisian beberapa aturan dalam teknik seleksi fitur dan klasifikasi, serta perancangan pada algoritme yang akan dipakai untuk menyelesaikan masalah berdasarkan metode yang telah dipilih.

## 1. PENDAHULUAN

---

### 5. Pengujian

Pada tahap ini dilakukan pengujian terhadap aplikasi yang telah dihasilkan.

### 6. Dokumentasi

Pada tahap ini dilakukan pendokumentasian hasil analisis dan implementasi secara tertulis dalam bentuk laporan skripsi.

## 1.7 Sistematika Pembahasan

Pada penelitian ini peneliti menyusun berdasarkan sistematika penulisan sebagai berikut:

### **BAB I       PENDAHULUAN**

Pendahuluan yang berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kontribusi penelitian, serta metode penelitian.

### **BAB II       LANDASAN TEORI**

Landasan Teori yang berisi penjelasan dasar teori yang mendukung penelitian ini.

### **BAB III       ANALISIS DAN PERANCANGAN**

Analisis dan perancangan yang berisi analisis berupa algoritme yang digunakan.

### **BAB IV       IMPLEMENTASI DAN PENGUJIAN**

Implementasi dan pengujian yang berisi implementasi pengujian dengan berbagai data pengujian beserta hasilnya.

### **BAB V       KESIMPULAN DAN SARAN**

## 1. PENDAHULUAN

---

Penutup yang berisi kesimpulan dari penelitian dan saran untuk penelitian lebih lanjut di masa mendatang.

## BAB 2 LANDASAN TEORI

Bab ini menjelaskan teori-teori yang berkaitan mengenai teori penunjang dan jurnal terkait yang digunakan dalam proses penelitian tugas akhir ini.

### 2.1 Tinjauan Pustaka

Penelitian ini menggunakan beberapa teori terkait yang diperlukan dalam pengerjaan yang dilakukan. Penjelasan mengenai teori-teori tersebut akan dijelaskan sebagai berikut.

#### 2.1.1 *Data Mining*

*Data mining* atau penambangan data adalah suatu proses penambangan informasi penting dan *pattern* dari suatu data yang berjumlah besar [7]. Informasi penting dan *pattern* didapat dari suatu proses yang menggunakan *artificial intelligence*, teknik statistik, ilmu matematika, pembelajaran mesin, dan lain sebagainya. Teknik-teknik tersebut akan mengidentifikasi dan mengekstraksi informasi dan *pattern* yang bermanfaat dari suatu kumpulan data yang besar.

Penambangan data mempunyai fungsi yang penting untuk membantu mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi pengguna. Pada dasarnya, data mining mempunyai empat fungsi dasar yaitu:

1. **Prediksi**, merupakan proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk memprediksikan variabel lain yang tidak diketahui jenis atau nilainya.
2. **Deskripsi**, merupakan proses untuk menemukan suatu karakteristik penting dari data dalam suatu basis data.



## 2. LANDASAN TEORI

---

3. **Klasifikasi**, merupakan suatu proses untuk menemukan model atau fungsi untuk menggambarkan kelas atau konsep dari suatu data [7]. Proses yang digunakan untuk mendeskripsikan data yang penting serta dapat meramalkan kecenderungan data pada masa depan.
4. **Asosiasi**, merupakan proses yang digunakan untuk menemukan suatu hubungan yang terdapat pada nilai atribut dari sekumpulan data.

### 2.1.2 Klasifikasi

Klasifikasi adalah salah satu metode yang paling banyak dipakai dalam pengambilan keputusan masalah dalam aktivitas manusia. Klasifikasi dipakai ketika suatu objek perlu ditugaskan ke dalam kelompok atau kelas yang telah ditentukan berdasarkan sejumlah atribut yang diamati terkait dengan objek itu. Terdapat banyak permasalahan yang dapat diidentifikasi sebagai permasalahan untuk klasifikasi, yaitu prediksi saham pasar, prediksi cuaca, prediksi kebangkrutan, deteksi penyakit, dan *speech recognition*. Permasalahan klasifikasi tersebut dapat diselesaikan baik secara matematis maupun non-linear. Kesulitan dalam memecahkan masalah klasifikasi seperti permasalahan secara matematis terdapat pada akurasi, distribusi data, dan kemampuan model dalam menentukan kelas [8].

### 2.1.3 Machine Learning

*Machine learning* atau pembelajaran mesin adalah sebuah metode pembelajaran yang membuat suatu mesin atau komputer belajar dari pengalaman dan digunakan sebagai sumber pengetahuan untuk melakukan klasifikasi yang bersumber dari data [7]. Proses pembelajaran mesin terdiri dari tahap pelatihan dan tahap pengujian. Pada tahap pelatihan, sistem akan diberikan atau membuat parameter sendiri untuk membentuk sebuah model yang kemudian akan diuji menggunakan data uji yang tersedia [9], setelah itu dilakukan tahap pengujian untuk memperoleh hasil dengan

## 2. LANDASAN TEORI

---

menggunakan model yang dihasilkan pada tahap pelatihan.

Dalam pembelajaran mesin terdapat dua metode pembelajaran, yaitu:

1. ***Supervised Learning***, merupakan salah satu teknik pembelajaran mesin yang menghasilkan hipotesa dari data latih yang sudah terdapat label kelas yang kemudian dipakai untuk prediksi dengan data uji. *Supervised Learning* merupakan metode umum yang sering dipakai pada permasalahan klasifikasi karena tujuannya membuat komputer mempelajari sistem klasifikasi yang telah dibuat [9]. Contoh metode dari *supervised learning* yaitu *Support Vector Machine*, *Linear Regression*, dan Naive Bayes.
2. ***Unsupervised Learning***, merupakan teknik pemetaan dari suatu data ke sebuah *output* yang tujuannya untuk menemukan keteraturan atau kemiripan dalam data yang ada sehingga tidak membutuhkan label kelas. Contoh metode dari *unsupervised learning* adalah *K-Means* dan *Hierarchical Learning*.

Pembelajaran mesin biasanya dipakai dalam klasifikasi maupun klasterisasi yang berfokus pada peningkatan akurasi dalam model [7].

### 2.1.4 *Support Vector Machine*

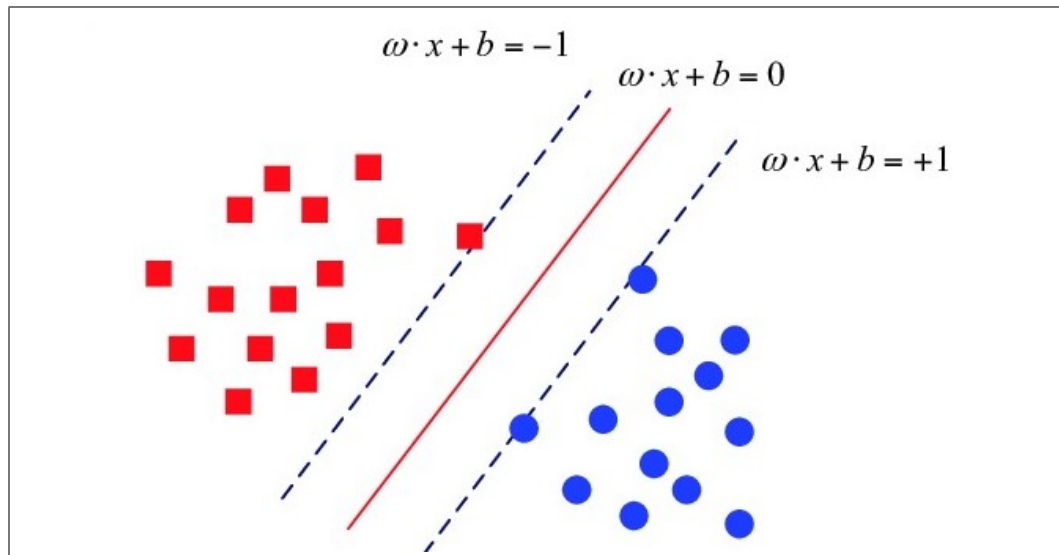
*Support Vector Machine* (SVM) adalah metode klasifikasi yang memiliki margin besar yang didasari oleh pembelajaran mesin dengan menggunakan ruang vektor. SVM merupakan salah satu teknik *supervised learning* yang berarti membutuhkan data latih yang sudah terdapat label kelas.

Tujuan utama SVM adalah untuk menentukan batas antara dua kelas. Teknik pemisah antara dua vektor ini dibagi menjadi dua, yaitu klasifikasi linear dan non-linear. Teknik klasifikasi ini terdiri dari berbagai pengembangan metode, seperti metode penanganan untuk banyak kelas di SVM (*Multiclass SVM*). Karena kemampuan SVM yang begitu baik dalam melakukan klasifikasi, algoritme SVM

## 2. LANDASAN TEORI

---

terus dikembangkan sehingga terdapat banyak metode baru SVM seperti *Latent Semantic Analysis Support Vector Machine* (LSA-SVM), *Weighted Voting Support Vector Machine* (WVSVM), dan *Least Square Support Vector Machine* (LSSVM) [10].



**Gambar 2.1** *Hyperplane pada Support Vector Machine*

Konsep dasar SVM adalah dengan membuat sebuah ataupun beberapa *hyperplane* secara optimal untuk memisahkan data menjadi beberapa kelas dan memaksimalkan margin. Margin adalah jarak antara *hyperplane* dengan titik data terdekat, titik data terdekat yang menyentuh *hyperplane* disebut dengan *Support Vector* [11].

Pada gambar 2.1, garis merah diagonal yang terletak di tengah dan memisahkan kumpulan data pada kelas positif dan kelas negatif merupakan *hyperplane* yang optimal, sedangkan garis putus - putus yang terletak di sebelah kanan dan kiri *hyperplane* optimal disebut *hyperplane* paralel. Berikut adalah persamaan dari *hyperplane* SVM :

## 2. LANDASAN TEORI

---

$$w^T.X + b = 0 \quad (2.1)$$

$$w^T.X + b = 1 \quad (2.2)$$

$$w^T.X + b = -1 \quad (2.3)$$

---

Di mana :

$w^T$  : *vector weight*

$X$  : *vector input*

$b$  : *bias*

---

Untuk mendapatkan nilai  $w$ , digunakan persamaan berikut:

$$w = \sum_{i=1}^n \alpha_i Y_i X_i \quad (2.4)$$

---

Di mana :

$w$  : *vector weight*

$\alpha$  : *alpha*

$Y_i$  : *vector output ke i*

$X_i$  : *vector input ke i*

---

SVM memaksimalkan margin untuk mendapatkan *hyperplane* optimal. Pencarian *hyperplane* optimum yang memaksimalkan margin dapat dipandang sebagai sebuah masalah *Quadratic Programming* (QP). Berikut adalah persamaan untuk mencari titik minimum dari *vector weight*:

$$\min_w \tau(w) = \frac{1}{2} ||w||^2 \quad (2.5)$$

---

Di mana :

$\min_w \tau(w)$  : *minimal of vector weight*

$w$  : *vector weight*

---

## 2. LANDASAN TEORI

---

Masalah QP dapat diselesaikan dengan menggunakan metode *Lagrange Multiplier*.

Berikut adalah persamaan dari *Lagrange Multiplier*:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^l \alpha_i (Y_i (X_i \cdot w + b) - 1) \quad (2.6)$$

---

Di mana :

$w$  : *vector weight*

$b$  : *bias*

$\alpha$  : *alpha*

$Y_i$  : *vector output ke i*

$X_i$  : *vector input ke i*

---

Pada permasalahan di dunia nyata, himpunan data umumnya terpisah secara non-linear. Untuk mengatasi masalah tersebut, digunakan konsep kernel pada ruang berdimensi lebih tinggi. Terdapat empat jenis fungsi kernel yang dapat digunakan, yaitu:

**Tabel 2.1** Persamaan Kernel pada *Support Vector Machine*

| Kernel  | Persamaan   |
|---|---|
| <i>Linear</i>   | $K(X_i, X_j) = X_i^T X_j \quad (2.7)$                                 |
| <i>Polynomial</i>                                       | $K(X_i, X_j) = (\gamma X_i^T X_j + \gamma)^d, \gamma > 0 \quad (2.8)$ |
| <i>Gaussian</i> atau <i>Radial Basis Function</i> (RBF) | $K(X_i, X_j) = \exp(-\gamma  X_i - X_j ^2) \quad (2.9)$               |

## 2. LANDASAN TEORI

---

**Tabel 2.1** Persamaan Kernel pada *Support Vector Machine* (Lanjutan)

| Kernel         | Persamaan   |
|----------------|---|
| <i>Sigmoid</i> | $K(X_i, X_j) = \tanh(\gamma X_i^T X_j + \gamma) \quad (2.10)$ |

Dari keempat kernel diatas, penelitian ini akan menggunakan kernel RBF karena cocok digunakan untuk klasifikasi dimana jumlah data lebih banyak dari jumlah fitur. *Gamma* ( $\gamma$ ) pada kernel RBF dipakai untuk mengatur variansi dan bias pada sebuah kumpulan data. Semakin kecil *gamma* yang dipakai, maka variansi akan semakin tinggi dan bias semakin rendah sehingga data-data yang berhimpitan dapat terpisah dengan baik.

Untuk mendapatkan nilai  $\alpha$  dan  $b$  berdasarkan klasifikasi non-linear digunakan persamaan linear sebagai berikut:

$$\sum_{i=1}^n \alpha_i Y_i K(X, X_i) + b = Y_i \quad (2.11)$$

Untuk menyelesaikan permasalahan klasifikasi non-linear dapat digunakan persamaan berikut:

$$\text{sign}\left(\sum_{i=1}^n \alpha_i Y_i K(X, X_i) + b\right) \quad (2.12)$$

Konsep kernel tidak selalu mampu menghasilkan himpunan data menjadi dapat dipisahkan secara linear. Maka dari itu digunakan konsep *Soft Margin* untuk meregularisasi nilai parameter  $C$  sebagai batas toleransi dan mengontrol margin

## 2. LANDASAN TEORI

---

SVM. Semakin besar nilai parameter  $C$  yang diberikan, semakin rendah toleransi yang diberikan dan margin akan semakin sempit. Berikut adalah persamaan yang digunakan:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (2.13)$$

---

Di mana :  
 $w$  : *vector weight*  
 $C$  : *regularization*  
 $\xi$  : *slack variable*

---

### 2.1.5 Data Preprocessing

*Data preprocessing* atau pemrosesan data merupakan salah satu langkah terpenting dalam penambangan data. Pemrosesan data memproses data mentah agar data memiliki format nilai yang sama dan menghilangkan data yang tidak diinginkan. Tujuan pemrosesan data yaitu untuk membantu proses klasifikasi menjadi efisien dan mendapatkan hasil akurasi klasifikasi yang lebih baik [7].

Terdapat 4 teknik yang dipakai dalam penelitian ini, yaitu:

1. **Data Deduplication**, merupakan teknik untuk mengeliminasi data yang berulang.
2. **Data Imputation**, merupakan teknik untuk mengubah nilai kosong pada sebuah data. Nilai yang kosong akan diisi menggunakan *median* atau *mean* dari data yang sama.
3. **Data Transformation**, merupakan teknik untuk normalisasi data atau mengubah format nilai sebuah data menjadi nilai yang dapat dipakai untuk penambangan data. Bertujuan untuk menghilangkan data yang tidak konsisten.

## 2. LANDASAN TEORI

---

4. **Diskritisasi Data**, merupakan teknik untuk merubah nilai data yang bersifat kontinu ke nilai data yang bersifat diskrit.

### 2.1.6 Seleksi Fitur

Seleksi fitur merupakan salah satu teknik untuk mereduksi fitur. Fitur yang dihapus merupakan fitur dengan data yang redundan atau yang tidak berpengaruh dari suatu dataset dikarenakan data pada suatu fitur memiliki nilai yang hampir selalu sama. Seleksi fitur juga dapat membuang fitur yang datanya bersifat *noisy*.

Terdapat 3 teknik seleksi fitur, yaitu:

1. **Filter Method**, merupakan seleksi fitur berdasarkan *threshold* yang ditentukan.  
Contoh: *Information Gain*, *Gain Ratio*, dan *Symmetrical Uncertainty*.
2. **Wrapper Method**, menggunakan *predictor* sebagai metode pengujian *blackbox* untuk mengevaluasi kinerja dari metode klasifikasi yang digunakan pada subset yang berbeda. Contoh : *Sequential Selection Algorithms* dan *Heuristic Search Algorithms*.
3. **Hybrid Method**, merupakan teknik yang menggabungkan *filter* dan *wrapper*.  
Contoh: *Decision Tree*.

### 2.1.7 Symmetrical Uncertainty

Metode *Symmetrical Uncertainty* merupakan salah satu metode seleksi fitur yang berbasis entropi dan merupakan turunan dari metode *Information Gain* yang menghilangkan data bias yang dihasilkan oleh *Information Gain* [6]. Entropi dalam penambahan data digunakan untuk mengukur keragaman dalam himpunan data. Nilai entropi yang semakin tinggi maka semakin beraneka ragam himpunan datanya.

Dalam menghitung entropi, data yang dimiliki diharuskan bersifat diskrit atau



## 2. LANDASAN TEORI

---

kategorikal. Untuk mengubah data menjadi kategorikal, pengelompokan datanya dibagi berdasarkan nilai interval suatu fitur. Total kelompok dihitung dengan persamaan:

$$k = 1 + 3.32 \log(n) \quad (2.14)$$

Setelah didapatkan total kelompok pada himpunan data, dilanjutkan dengan menghitung nilai interval dengan persamaan:

$$I(X) = \frac{\max(X) - \min(X)}{k} \quad (2.15)$$

---

Di mana :  
 $I(X)$  : Interval fitur X.

---

Setelah merubah data menjadi kategorikal, entropi sudah dapat dihitung. Berikut adalah persamaan entropi dalam satu variabel:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_2(P(x_i)) \quad (2.16)$$

---

Di mana :  
 $H(X)$  : Entropi fitur X.  
 $P(x_i)$  : Peluang data ke-i pada fitur X.

---

Berikut adalah persamaan entropi dalam dua variabel:

## 2. LANDASAN TEORI

---

$$H(X|Y) = - \sum_{j=1}^{n_y} \sum_{i=1}^{n_x} P(x_i|y_j) \log_2(P(x_i|y_j)) \quad (2.17)$$

---

Di mana :

$H(X|Y)$  : Entropi fitur Y terhadap fitur X.

$P(x_i|y_i)$  : Peluang data ke-i pada fitur Y terhadap data ke-i pada fitur X.

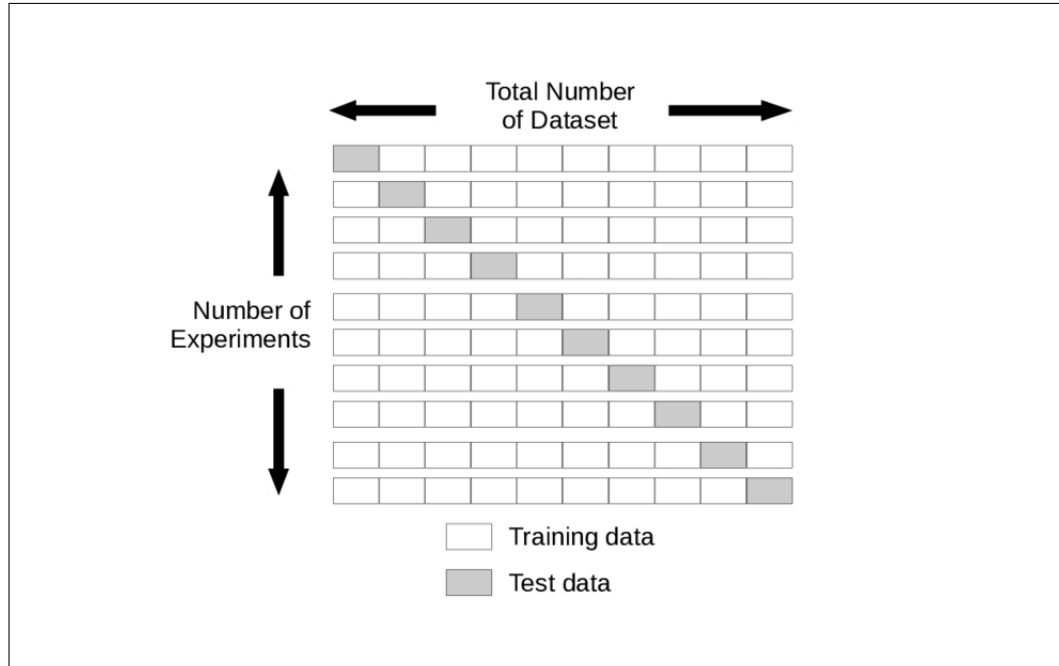
---

Setelah entropi dihitung, perhitungan *Symmetrical Uncertainty* sudah dapat dilakukan. Berikut adalah persamaan *Symmetrical Uncertainty*:

$$SU(X,Y) = 2 \left( \frac{H(X) - H(X|Y)}{H(X) + H(Y)} \right) \quad (2.18)$$

### 2.1.8 K-Fold Cross Validation

*K-Fold Cross Validation* merupakan teknik *resampling* untuk membagi data latih dan data uji dikarenakan memprediksi *error* hanya sekali dapat menghasilkan kesalahan prediksi. Dengan membagi menjadi beberapa percobaan membuat *dataset* memiliki variasi pada nilai *error*nya dan mengurangi bias pada sebuah dataset [12]. *K-Fold Cross Validation* mengelompokkan *dataset* menjadi *k* grup dengan masing-masing grup mempunyai *k* lipatan atau *fold*. Data uji diambil dari *fold* yang indeksinya sesuai dengan indeks grup dan sisanya dipakai untuk data latih. Untuk menghitung akurasi dari data uji, dirata-ratakan akurasi dari setiap grup.



**Gambar 2.2** 10-Fold Cross Validation

### 2.1.9 *F-Measure*

*F-Measure* atau *F-Score* merupakan teknik pengukur akurasi yang menggabungkan nilai *Recall* dan *Precision*. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dalam penelitian ini *F-Measure* digunakan untuk mengukur akurasi pada setiap kelas. Berikut adalah persamaan untuk menghitung *Precision*, *Recall*, dan *F-Measure*:

$$Precision = \frac{TP}{TP + FP} \quad (2.19)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.20)$$

$$F - Measure = 2 \times \frac{Precision - Recall}{Precision + Recall} \quad (2.21)$$

Untuk memudahkan dalam menghitung *F-Measure*, digunakan tabel *Confusion Matrix* berikut:

## 2. LANDASAN TEORI

---

**Tabel 2.2** *Confusion Matrix*

|                | Nilai sebenarnya |    |
|----------------|------------------|----|
| Nilai prediksi | TP               | FP |
|                | TN               | FN |

### 2.2 Tinjauan Studi

Pada bagian ini akan dijelaskan mengenai perbandingan dari berbagai penelitian terkait prediksi *title box office* film.

#### 2.2.1 *State of the Art Method*

Terdapat beberapa metode lain yang memiliki ruang lingkup yang mirip dengan penelitian ini khususnya mengenai deteksi dan pelacakan manusia. Tabel 2.3 akan menjelaskan perbedaan-perbedaan metode yang telah dipelajari oleh penulis dari jurnal.

**Tabel 2.3** *State of the Art Method*

| No | Peneliti  | Judul  | Objektif  | Hasil                              |
|----|---|--|---|------------------------------------|
| 1  | Shraddha Mehta, Hitarthi Bhatt, and Prof. Darshana Desai (2015)     | A Compendium for Prediction of Success of a Movie Based Upon Different Factors | Menerapkan metode <i>Decision Tree ID3</i> dalam memprediksi <i>title box office</i>                | Akurasi yang didapat sebesar 66.7% |
| 2  | K. Meenakshi, G. Maragatham, Neha Agarwal, and Ishitha Ghosh (2018) | A Data Mining Technique for Analyzing and Predicting the Success of Movie      | Menerapkan metode <i>K-Means</i> dan <i>Decision Tree</i> dalam memprediksi <i>title box office</i> | Akurasi yang didapat sebesar 71.4% |

## 2. LANDASAN TEORI

**Tabel 2.3** *State of the Art Method* (Lanjutan)

| No | Peneliti                                  | Judul  | Objektif  | Hasil   |
|----|---|--|---|---|
| 3  | Ling Liu and Yang Zhao (2016)             | Research of Box-Office Prediction based on Rough Set and Support Vector Machine  | Menerapkan metode <i>Rough Set</i> dengan <i>Support Vector Machine</i> dalam memprediksi <i>title box office</i> film  | Hasil MAE yang didapat dari minggu ke 2 sampai 4 secara berurutan adalah 9.93%, 15.7%, dan 13.3%  |
| 4  | C. Sunil Kumar and R. J. Rama Sree (2014) | Application of Ranking Based Attribute Selection Filters to Perform Automated Evaluation of Descriptive Answers Through Sequential Minimal Optimization Models | Membandingkan metode seleksi fitur berbasis filter <i>Chi Squared</i> , <i>Information Gain</i> , <i>Gain Ratio</i> , <i>Symmetrical Uncertainty</i> , dan <i>Relief Attribute Evaluation</i> menggunakan 5 dataset | Peringkat seleksi fitur terbaik berdasarkan <i>Kappa Statistic</i> , MAE, dan RMSE :<br><br>1. <i>Symmetrical Uncertainty</i><br><br>2. <i>Information Gain</i><br><br>3. <i>Chi Squared</i><br><br>4. <i>Gain Ratio</i><br><br>5. <i>Relief Attribute Evaluation</i> |

### 2.2.2 Pembahasan Penelitian Terkait

Terdapat beberapa metode yang dapat khususnya untuk memprediksi *title box office* sebuah film. Pada referensi pertama [3] menggunakan metode *Decision Tree ID3* dengan *Information Gain* dan entropi sebagai metode untuk mengetahui keragaman data pada setiap fitur. Alasan peneliti menggunakan *Decision Tree ID3*

## 2. LANDASAN TEORI

---

karena menghasilkan *tree* yang pendek dalam waktu yang singkat. *Decision Tree* dapat dengan mudah menghasilkan model prediksi dalam bentuk *tree* namun tidak efisien jika data memiliki fitur yang banyak. Akurasi yang didapat adalah 66.7%.

Pada referensi kedua [4] metode yang digunakan adalah *K-Means* yang digunakan untuk mengetahui fitur yang paling berpengaruh terhadap prediksi *title box office* film, kemudian dilanjutkan dengan metode *Decision Tree* untuk menentukan model prediksi. Akurasi yang didapat adalah 71.4%.

Pada referensi ketiga [5] menggunakan metode *Rough Set* sebagai seleksi fitur dan *Support Vector Machine* untuk membentuk model prediksi. Pertama dilakukan *cross validation* untuk meningkatkan performa dan akurasi metode. Kemudian dilanjutkan dengan pemrosesan data dan mengubah data menjadi data diskrit. Selanjutnya menggunakan metode *Rough Set* untuk mengurangi fitur yang tidak penting dan selanjutnya menggunakan metode *Support Vector Machine* untuk menentukan model prediksi. Percobaan dilakukan pada minggu 2 sampai minggu 4 dengan akurasi yang didapat secara berurutan adalah 90.7%, 84.3%, dan 86.7%.

Pada referensi keempat [6] membandingkan seluruh metode seleksi fitur berbasis filter, yaitu *Chi Squared*, *Information Gain*, *Gain Ratio*, *Symmetrical Uncertainty*, dan *Relief Attribute Evaluation*. Percobaan dilakukan sebanyak lima kali dengan lima dataset yang berbeda. Rata-rata dari lima dataset *Kappa Statistic*, *Mean Absolute Error*, dan *Root Mean Square Error* dari lima metode seleksi fitur berbasis filter dan pengolahan data tanpa fitur seleksi secara berurutan adalah *Chi Squared* sebesar 0.52232, 0.2844, dan 0.36332, *Gain Ratio* sebesar 0.52232, 0.2845, dan 0.36342, *Information Gain* sebesar 0.52166, 0.2844, dan 0.3633, *Relief Attribute Evaluation* sebesar 0.3936, 0.3023, dan 0.3841, *Symmetrical Uncertainty* sebesar 0.522, 0.28438, dan 0.36326, dan tanpa seleksi fitur sebesar 0.35224, 0.30364, dan 0.3872.

### 2.3 Tinjauan Objek

Pada bagian ini akan diulas mengenai objek-objek yang terkait dengan prediksi *title box office* film.

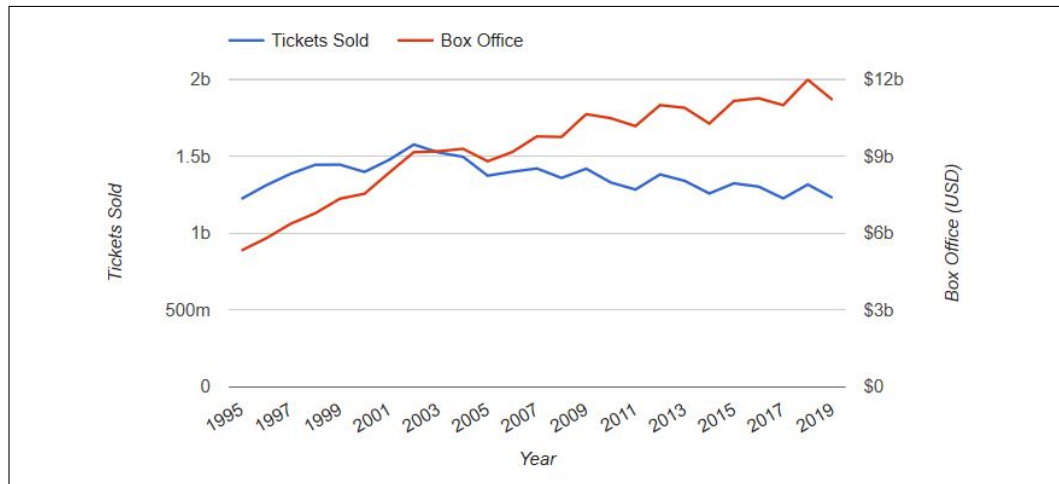
#### 2.3.1 Film

Film adalah media komunikasi yang bersifat audio visual untuk menceritakan sebuah kisah atau menyampaikan suatu pesan kepada sekelompok orang yang menontonnya. Film merupakan salah satu industri bisnis yang memiliki keuntungan terbesar yang bisa mencapai 12 miliar Dollar Amerika Serikat [1]. Namun hanya sedikit rumah produksi film yang bisa menghasilkan keuntungan bersih melebihi anggaran pembuatan. Film memiliki beberapa jenis seperti film serial, film animasi, film pendek, dan lain lain.

#### 2.3.2 Box Office

*Box office* merupakan gelar bagi film - film yang berhasil mendapatkan pendapatan bersih yang melebihi anggaran pembuatan film hanya dari penjualan tiket. Terdapat beberapa jenis *box office*, yaitu *daily*, *weekly*, *monthly*, *yearly*, dan *all-time*. *Box office* bisa juga disebut sebagai podium untuk film yang memperoleh keuntungan terbesar dalam jangka waktu yang ditentukan. Dalam menerima sebuah gelar *box office*, sebuah film harus dirilis secara internasional.

## 2. LANDASAN TEORI



**Gambar 2.3** Box Office Statistics

Untuk menentukan sebuah film mendapatkan *title box office*, dapat menggunakan persamaan dari *hollywood accounting* [13]:

$$\text{Budget} - (\text{Budget} * 55\%) > 0 \quad (2.22)$$

### 2.3.3 Hollywood Accounting

*Hollywood accounting* atau *creative accounting* merupakan teknik akuntansi yang digunakan pada industri film dan televisi untuk menghitung pendapatan bersih dari setiap project [2]. Teknik ini dinamakan *hollywood accounting* dikarenakan sebagian besar studio perfilman ternama terletak pada Hollywood, Amerika Serikat. Teknik ini dibuat dikarenakan pencatatan terkait keuntungan pada industri perfilman biasanya hanya pendapatan kotor.

Dengan menghitung pengeluaran tidak terduga berdasarkan pendapatan kotor, maka didapatkan perkiraan pendapatan bersih dari suatu project film. Terdapat tiga biaya tidak terduga, yaitu:

1. ***Production overhead***, merupakan biaya tak terduga dari segi produksi, yang



## 2. LANDASAN TEORI

---

melibatkan seperti biaya pemeran figuran dan pembuatan film. Mengambil 15% dari pendapatan kotor.

2. ***Distribution overhead***, merupakan biaya tak terduga untuk distributor film seperti pihak bioskop. Mengambil 30% dari pendapatan kotor.
3. ***Marketing overhead***, merupakan biaya tak terduga yang dikeluarkan untuk semua periklanan yang telah dikeluarkan. Mengambil 10% dari pendapatan kotor.

Dengan kata lain, pendapatan bersih dari sebuah film yaitu pendapatan kotor dikali dengan total ketiga presentasi biaya tidak terduga [2].

### 2.3.4 *International Movie Database (IMDb)*

IMDb merupakan salah satu situs terbesar yang menyediakan data terkait film, program televisi, *video game* beserta konten yang dimiliki (biografi aktor, *rating*, dan ulasan film) secara online.

Sebagian besar data yang tersedia di situs IMDb ditulis oleh pengguna yang telah terdaftar di situs IMDb. Pengguna juga dapat memberikan ulasan dan *rating* terhadap film yang tersedia di situs IMDb.

## BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab ini memaparkan analisis masalah yang diatasi berserta pendekatan dan alur kerja dari penelitian yang dikembangkan, mengimplementasikan metode yang digunakan dan membandingkan hasil yang ditampilkan.

### 3.1 Analisis Masalah

Prediksi *title box office* film merupakan penelitian yang dapat membantu rumah produksi film untuk menentukan jalan cerita, aktor, dll. sebelum film mulai diproduksi agar film tersebut dapat meraih *title box office*. Dalam penelitian ini, data masukan yang digunakan adalah data film IMDb. Sebelum mendapatkan model prediksi, dilakukan pemrosesan data dengan mengambil data film yang memiliki tahun rilis sekitar 2006 sampai 2016 dan membuang data yang memiliki nilai kosong karena dalam memprediksi *title box office* film, dibutuhkan data yang lengkap dan data tersebut tidak dapat digantikan atau dinormalisasi karena akan berpengaruh pada hasilnya. Data yang memiliki judul dan tahun rilis yang sama pun akan dihapus karena merupakan data duplikasi. Kemudian data fitur pada sebuah film yang memiliki nilai yang kosong akan diisi dengan *mean* atau *median*. Selanjutnya fitur yang memiliki data berupa teks akan diubah menjadi angka dengan nilainya berdasarkan tingkat kesuksesan fitur tersebut meraih *title box office*. Fitur yang memiliki data teks yang tidak bisa diubah menjadi nilai kesuksesannya seperti judul film, tahun rilis, dan link situs film akan dibuang.

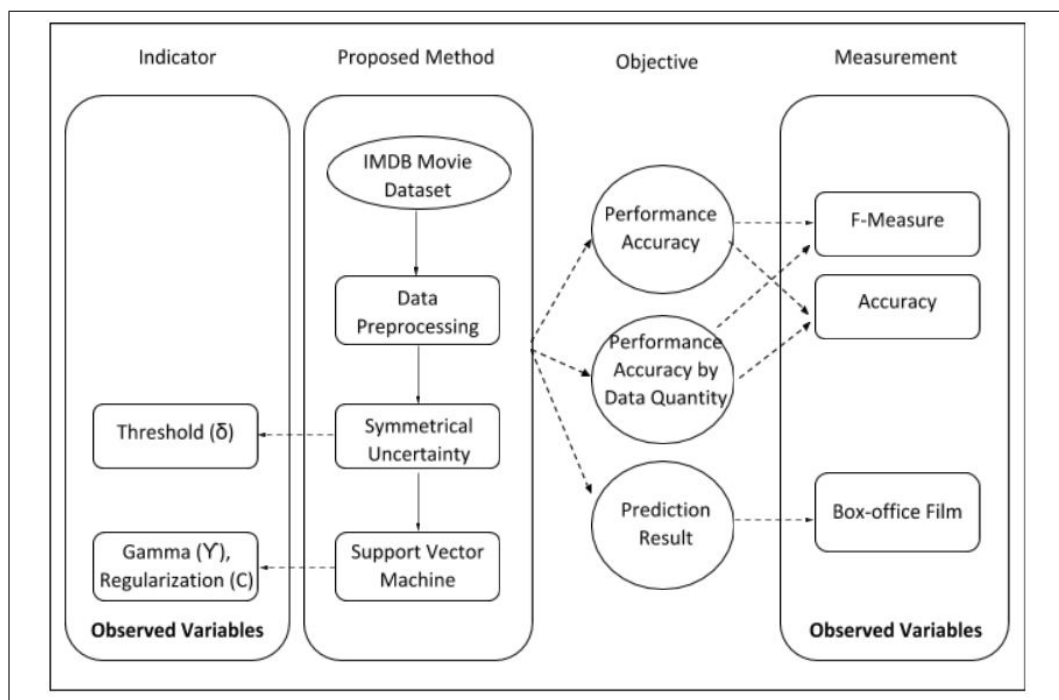
Tahap berikutnya yaitu melakukan fitur seleksi metode *Symmetrical Uncertainty*. Dilakukan perhitungan *Symmetrical Uncertainty* setiap fitur terhadap label kelas, kemudian mengambil fitur yang memiliki nilai *Symmetrical Uncertainty* diatas *threshold* yang telah ditentukan. Selanjutnya, dilakukan metode *Support Vector*

### 3. ANALISIS DAN PERANCANGAN SISTEM

*Machine* pada tahap pengujian untuk membuat model prediksi. Model prediksi dibuat dengan membandingkan setiap fitur dengan data. Kemudian memakai perhitungan kernel *Radial Basis Function* dengan *gamma* dan regularisasi (*C*) yang ditentukan untuk mendapatkan nilai *alpha* dan bias. Nilai *alpha* dan bias dipakai untuk membantu mendapatkan hasil prediksi dan model prediksi dipakai untuk perbandingan model prediksi pada tahap pengujian, Keluaran pada sistem ini adalah hasil prediksi *title box office* film. Hasil prediksi akan dihitung akurasi menggunakan *F-Measure*.

#### 3.2 Kerangka Pemikiran

Berikut ini adalah kerangka pemikiran dari metode yang diusulkan untuk melakukan prediksi *title box office* film.



**Gambar 3.1** Kerangka Pemikiran

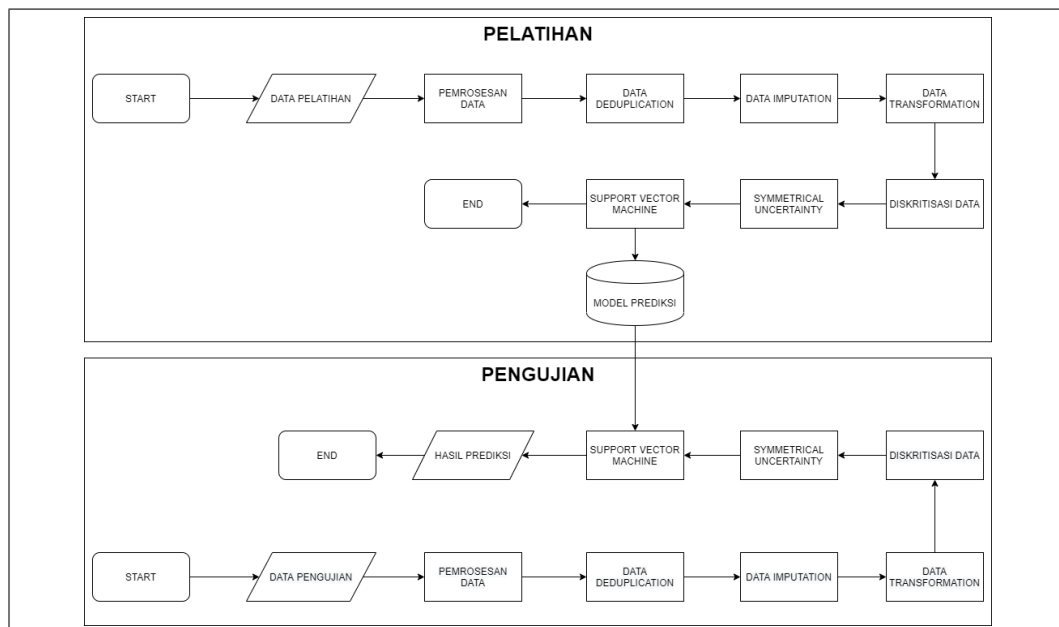
Pada gambar 3.1, terdapat beberapa indikator yang mempengaruhi hasil prediksi seperti *threshold* pada *Symmetrical Uncertainty*, *gamma* dan regularisasi pada *Support Vector Machine*. Penelitian ini bertujuan untuk memprediksi *title box*

### 3. ANALISIS DAN PERANCANGAN SISTEM

*office* sebuah film dengan membandingkan model prediksi yang telah dibuat dan melihat akurasi prediksi menggunakan *F-Measure*.

#### 3.3 Analisis Urutan Proses Global

Dalam sistem untuk prediksi *title box office* film terbagi atas dua proses yaitu proses pelatihan dan pengujian. Proses pelatihan dilakukan untuk mendapatkan model prediksi *Support Vector Machine* yang optimal. Proses pengujian dilakukan untuk mendapatkan hasil prediksi dari model prediksi yang didapat dan akurasi.



**Gambar 3.2** Flowchart Pelatihan dan Pengujian pada Pembuatan Sistem Prediksi *Title Box Office* Film

Berikut ini adalah uraian dari *flowchart* pada gambar 3.2 yang dilakukan dalam penelitian ini:

1. Masukan berupa data pelatihan film yang sudah dipilih melalui *10-Fold Cross Validation*.
2. Proses data dengan mengambil data film dengan tahun rilis antara 2006 sampai 2016 dan membuat data yang kosong dan mengubah data yang berupa teks menjadi angka dengan nilai yang didapat berdasarkan tingkat

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

kesuksesan film meraih *title box office*.

3. Membuang fitur yang memiliki data berupa teks namun tidak memiliki nilai kesuksesan *box office*.
4. Membuang data yang duplikasi.
5. Mengisi nilai fitur pada data dengan *mean* atau *median* dari masing-masing fitur.
6. Mengubah nilai kesuksesan film meraih *title box office* yang bersifat kategorikal menjadi diskrit.
7. Seleksi fitur dilanjutkan dengan metode *Symmetrical Uncertainty* dengan threshold ( $\delta$ ) yang ditentukan dengan nilai berkisar 0 sampai 1.
8. *Support Vector Machine* digunakan untuk klasifikasi setiap fitur pada data film. Kernel yang dipakai adalah RBF kernel dengan parameter *gamma* ( $\gamma$ ) dan regularisasi (C). *Support Vector Machine* akan menghasilkan model prediksi dan *alpha* bias yang dipakai untuk data uji.
9. Dilanjutkan dengan masukan data uji film yang sudah dipilih melalui *10-Fold Cross Validation*.
10. Proses data dengan mengubah text menjadi nilai kesuksesannya dan mengubah data yang berupa teks menjadi angka dengan nilai yang didapat berdasarkan tingkat kesuksesan film meraih *title box office*.
11. Membuang fitur yang memiliki data berupa teks namun tidak memiliki nilai kesuksesan *box office*.
12. Membuang data yang duplikasi.
13. Mengisi nilai fitur pada data dengan *mean* atau *median* dari masing-masing fitur.

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

14. Seleksi fitur dengan mengambil fitur yang dipakai pada pelatihan.
15. *Support Vector Machine* digunakan untuk klasifikasi setiap fitur pada data film dengan memakai *alpha* bias yang telah didapatkan pada data pelatihan. Kernel yang dipakai adalah RBF kernel dengan *gamma* dan regularisasi yang disamakan pada pelatihan.
16. Didapatkan hasil prediksi *box office* dengan mencocokkan model prediksi yang dibuat pada proses data pelatihan.

#### 3.4 Analisis Kasus

Pada bagian ini dilakukan proses analisis tahapan dengan melakukan perhitungan manual.

##### 3.4.1 Dataset

*Dataset* yang digunakan dalam penelitian ini adalah data film IMDb dari tahun 1916 sampai 2017 berjumlah 5044 data yang berasal dari situs [www.kaggle.com](http://www.kaggle.com). IMDb merupakan *database online* yang menyediakan data terkait dengan film, serial TV, video game, dan *streaming content* seperti aktor, sutradara, produser, *review*, dan *rating*. Berikut adalah contoh fitur dan *dataset* film:

**Tabel 3.1** Kode Fitur

| Fitur                          | Kode<br>Fitur |
|--------------------------------|---------------|
| <i>color</i>                   | V1            |
| <i>director_name</i>           | V2            |
| <i>num_critic_for_review</i>   | V3            |
| <i>duration</i>                | V4            |
| <i>director_facebook_likes</i> | V5            |

**Tabel 3.1** Kode Fitur (Lanjutan)

| <b>Fitur</b>                     | <b>Kode</b> |
|----------------------------------|-------------|
| <i>actor_1_facebook_likes</i>    | V6          |
| <i>actor_2_facebook_likes</i>    | V7          |
| <i>actor_3_facebook_likes</i>    | V8          |
| <i>actor_1_name</i>              | V9          |
| <i>actor_2_name</i>              | V10         |
| <i>actor_3_name</i>              | V11         |
| <i>genres</i>                    | V12         |
| <i>movie_title</i>               | V13         |
| <i>num_voted_users</i>           | V14         |
| <i>cast_total_facebook_likes</i> | V15         |
| <i>facenumber_in_poster</i>      | V16         |
| <i>plot_keywords</i>             | V17         |
| <i>movie_imdb_link</i>           | V18         |
| <i>num_user_for_reviews</i>      | V19         |
| <i>language</i>                  | V20         |
| <i>country</i>                   | V21         |
| <i>content_rating</i>            | V22         |
| <i>budget</i>                    | V23         |
| <i>title_year</i>                | V24         |
| <i>imdb_score</i>                | V25         |
| <i>aspect_ratio</i>              | V26         |
| <i>movie_facebook_likes</i>      | V27         |
| <i>total_spend</i>               | V28         |
| <i>box_office</i>                | V29         |

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

**Tabel 3.2** Dataset

| <b>Kode Fitur</b> | <b>Contoh Data 1</b>  | <b>Contoh Data 2</b>  | <b>Contoh Data 3</b>  |
|-------------------|---|---|---|
| V1                | Color   | Color   | Color   |
| V2                | James Cameron   | Christopher Nolan   | Josh Whedon   |
| V3                | 723   | 813   | 703   |
| V4                | 178   | 164   | 173   |
| V5                | 0   | 22000   | 0   |
| V6                | 1000  | 27000   | 26000   |
| V7                | 936   | 23000   | 2100  |
| V8                | 855   | 23000   | 19000   |
| V9                | CCH Pounder   | Tom Hardy   | Chris Hemsworth   |
| V10               | Joel David Moore  | Christian Bale  | Robert Downey Jr.   |
| V11               | Wes Studi   | Joseph Gordon-Levitt  | Scarlett Johansson  |
| V12               | Action Adventure Fantasy Sci-Fi   | Action Thriller   | Action Adventure Sci-Fi   |
| V13               | Avatar  | The Dark Knight Rises   | The Avengers  |
| V14               | 886204  | 114433  | 9954157   |
| V15               | 4834  | 106759  | 87697   |
| V16               | 0   | 0   | 3   |
| V17               | avatar future marine native paraplegic  | deception imprisonment lawlessness police officer terrorist plot  | alien invasion assassin battle iron man soldier   |
| V18               | <a href="http://www.imdb.com/title/tt0499549/?ref_=fn_tt_tt_1">http://www.imdb.com/title/tt0499549/?ref_=fn_tt_tt_1</a> | <a href="http://www.imdb.com/title/tt1345836/ref_=fn_tt_tt_1">http://www.imdb.com/title/tt1345836/ref_=fn_tt_tt_1</a> | <a href="http://www.imdb.com/title/tt0848228/ref_=fn_tt_tt_1">http://www.imdb.com/title/tt0848228/ref_=fn_tt_tt_1</a> |
| V19               | 3054  | 2701  | 1722  |
| V20               | English   | English   | English   |
| V21               | USA   | USA   | USA   |
| V22               | PG-13   | PG-13   | PG-13   |



### 3. ANALISIS DAN PERANCANGAN SISTEM

---

**Tabel 3.2** Dataset (Lanjutan)

| Kode<br>Fitur | Contoh Data 1 | Contoh Data 2 | Contoh Data 3 |
|---------------|---------------|---------------|---------------|
| V23           | 237000000     | 250000000     | 220000000     |
| V24           | 2009          | 2012          | 2012          |
| V25           | 7.9           | 8.5           | 8.1           |
| V26           | 1.78          | 2.35          | 1.85          |
| V27           | 33000         | 164000        | 123000        |
| V28           | 367350000     | 465000000     | 379750000     |
| V29           | 1             | 0             | 0             |

Data pelatihan dan data uji akan dipisahkan menggunakan metode *10-Fold Cross Validation*.

#### 3.4.1.1 Analisis Fitur *Dataset*

Berdasarkan *dataset* yang dipakai pada penelitian ini, terdapat beberapa fitur yang mempengaruhi sebuah film akan meraih *title box office* film dan analisis pengaruh fitur terhadap prediksi *title box office* film.

1. **Color**, merupakan jenis pewarnaan yang dipakai pada film. Terdapat dua jenis, yaitu berwarna dan hitam putih. Seiring dengan perkembangan zaman, pewarnaan film hitam putih mulai menghilang sehingga fitur *color* hampir tidak mempengaruhi sebuah film meraih *title box office*.
2. **Director Name**, merupakan salah satu fitur yang berpengaruh besar terhadap *box office* film dikarenakan sutradara bertugas mengatur arah jalan suatu film. Sutradara yang terkenal dan memiliki bayaran yang tinggi tentunya memiliki banyak prestasi yang membuat penonton tidak ragu untuk menonton film yang disutradarai sutradara terkenal.

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

3. ***Number of Critic for Review***, merupakan jumlah kritik yang diterima terhadap sebuah film. Kritik merupakan ulasan yang ditulis oleh masyarakat maupun kritikus yang menentukan bahwa film yang diulas tersebut layak ditonton atau tidak. Kritik terhadap suatu film dapat memberikan informasi bagi masyarakat yang belum menonton suatu film.
4. ***Duration***, merupakan panjang waktu film. Fitur ini hanya memiliki sedikit pengaruh terhadap prediksi *title box office* film dikarenakan hanya sedikit penonton yang mempermasalahkan durasi film baik terlalu pendek maupun terlalu panjang. Dengan durasi yang pendek, belum tentu film akan memiliki kualitas yang buruk, begitupun dengan film yang berdurasi panjang.
5. ***Director Facebook Likes***, merupakan jumlah pengguna di Facebook yang menyukai halaman penggemar sutradara. Media sosial sangat membantu untuk meningkatkan jumlah penonton yang tertarik pada suatu film dengan hanya melihat, menyukai, ataupun membagikan hal yang berkaitan terhadap suatu film dari halaman penggemar sutradara.
6. ***Actor Name***, merupakan pemeran pada film. Sama seperti sutradara, aktor yang terkenal dan memiliki bayaran yang tinggi lebih diminati masyarakat. Aktor dapat dikenal oleh masyarakat dikarenakan aktor tersebut dapat memerankan setiap karakter pada film yang ia perankan sebelumnya dengan sangat baik sehingga penonton pastinya menunggu aktor tersebut untuk bermain film yang lain. Dalam *dataset* IMDb yang dipakai, terdapat 3 fitur *actor name* yang dipakai untuk memprediksi *title box office* film.
7. ***Actor Facebook Likes***, merupakan jumlah pengguna di Facebook yang

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

menyukai halaman penggemar aktor. Seperti halnya fitur *director facebook likes*, fitur ini dapat meningkatkan jumlah penonton dengan hanya melihat, menyukai, ataupun membagikan hal yang berkaitan terhadap suatu film dari halaman penggemar aktor. Dalam *dataset* IMDb yang dipakai, terdapat 3 fitur *actor facebook likes* yang dipakai untuk memprediksi *title box office* film.

8. **Genre**, merupakan bentuk klasifikasi pada sebuah film yang memiliki karakteristik yang sama, misalkan aksi, petualangan, dan lain lain. Film dapat memiliki aliran lebih dari satu dan kombinasi aliran yang dipakai memiliki pengaruh dalam meraih *title box office*. Aliran juga menentukan arah atau tujuan dibuat film. Sampai tahun 2019, aliran film memberikan rata-rata pendapatan tertinggi bila dibandingkan dengan fitur lain [1] sehingga aliran film merupakan salah satu fitur yang harus perlu diperhatikan dalam produksi film.
9. **Movie Title**, merupakan salah satu fitur yang tidak bisa dibilang memberikan pengaruh awal terhadap menarik penonton untuk menonton sebuah film, namun memberikan pengaruh yang besar jika memiliki judul yang sama dengan film terdahulu (sekuel maupun remake) yang memiliki ketertarikan masyarakat yang besar. Masyarakat tertarik dengan suatu film sekuel maupun remake dengan mengharapkan film tersebut sama maupun lebih baik dari film terdahulu.
10. **Number of Voted Users**, merupakan jumlah *vote* pengguna terhadap suatu film. Jumlah *vote* yang tinggi terhadap suatu film menandakan bahwa pengguna merekomendasikan film tersebut, yang berarti fitur ini berperpengaruh terhadap prediksi *title box office*.
11. **Cast Total Facebook Likes**, merupakan jumlah pengguna di Facebook yang menyukai halaman penggemar semua pemeran kecuali aktor dan sutradara.

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

Sama seperti aktor dan sutradara, fitur ini dapat meningkatkan jumlah penonton dengan hanya melihat, menyukai, ataupun membagikan hal yang terdapat pada halaman penggemar *cast*.

12. ***Facenumber in Poster***, merupakan jumlah aktor yang berada pada poster film. Poster film adalah fitur pertama yang dilihat masyarakat dalam meninjau film yang sebelum dirilis. Maka dari itu, poster film diharuskan mencakup semua informasi yang dibutuhkan masyarakat dalam menarik masyarakat untuk menonton film. Wajah aktor pada poster membantu memberikan informasi terkait aktor yang berperan pada film.
13. ***Plot Keywords***, merupakan alur cerita pada sebuah film. Alur cerita dibuat berdasarkan keinginan sutradara dan aktor yang berperan, sehingga fitur ini dapat meningkatkan prediksi *title box office* film. Sebelum film tayang, alur cerita diceritakan pada bagian sinopsis dengan memberikan gambaran secara kasar bagaimana alur cerita akan berjalan.
14. ***Movie IMDb Link***, merupakan situs film pada IMDb. Fitur ini tidak berpengaruh terhadap prediksi *title box office* film dikarenakan hanya membantu memberikan informasi terkait film lewat situs atau internet.
15. ***Number of User for Reviews***, merupakan jumlah pengguna yang mengulas sebuah film. Dengan banyaknya masyarakat yang ulasan terhadap sebuah film, dapat memberikan informasi bagi masyarakat yang belum menonton film tersebut.
16. ***Language***, merupakan bahasa utama yang dipakai dalam sebuah film. Fitur bahasa tidak mempengaruhi sebuah film meraih *title box office* karena sudah

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

mendapat bantuan *subtitle* atau *dubbing* bila terdapat penonton yang kesulitan dalam menonton film dengan bahasa asing.

17. **Country**, merupakan negara asal film diproduksi. Fitur ini cukup mempengaruhi dalam memprediksi *title box office* karena tidak semua negara dapat memproduksi film dengan kualitas yang bagus dan kesulitan sebuah negara ketika bersaing dalam skala internasional. Film dengan *box office* terbanyak dihasilkan oleh negara Amerika Serikat, namun tidak menutup kemungkinan negara lain dapat menghasilkan film *box office* seperti "The Raid" asal Indonesia yang berhasil meraih *title box office*.
18. **Content Rating**, merupakan sebuah kategori usia penonton yang dianjurkan oleh rumah produksi film dikarenakan terdapat konten film yang tidak layak dilihat untuk penonton yang usianya tidak memenuhi *content rating* yang diberikan. Fitur ini mempengaruhi sistem dalam memprediksi *title box office* dikarenakan jumlah penonton yang berkurang dikarenakan usia yang belum mencukupi maupun terdapat konten yang membuat masyarakat mengalami rasa takut berlebih dan sebagainya. Kategori *content rating* adalah PG-13, R, PG, G, Not Rated, dan NC-17. PG-13 merupakan kategori *content rating* yang memiliki peluang *box office* tertinggi [1].
19. **Budget**, merupakan anggaran pembuatan sebuah film. Anggaran pembuatan film merupakan salah satu fitur terpenting dalam prediksi *box office*. Anggaran pembuatan film akan dibandingkan dengan pendapatan kotor untuk menentukan *box office* sebuah film.
20. **Title Year**, merupakan tahun rilis sebuah film. Tahun tidak menentukan film akan memperoleh *title box office* dikarenakan tahun rilis sebuah film tidak

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

menarik masyarakat untuk menonton sebuah film.

21. **IMDb Score**, merupakan penilaian sebuah film yang diberikan oleh salah satu situs film terbesar di dunia, *International Movie Database* (IMDb). Fitur ini tidak memberikan pengaruh yang cukup besar pada prediksi *title box office* pada minggu pertama rilis namun memberikan pengaruh yang besar untuk keseluruhan waktu rilis pada bioskop dikarenakan IMDb tidak memberikan penilaian film pada awal rilis, yang berarti masyarakat tertarik untuk menonton dikarenakan fitur lain yang memberikan informasi yang menarik perhatian penonton, seperti aktor yang sudah meraih penghargaan *oscar*, yaitu penghargaan aktor terbaik.
22. **Aspect Ratio**, merupakan rasio aspek piksel yang dipakai pada sebuah film. Fitur ini tidak berpengaruh terhadap prediksi *title box office* dikarenakan rumah produksi tidak pernah memberikan rasio aspek pada film mereka secara langsung kepada masyarakat.
23. **Movie Facebook Likes**, merupakan jumlah pengguna di Facebook yang menyukai halaman penggemar film. Memiliki pengaruh terhadap prediksi *title box office* dikarenakan jumlah suka pada halaman penggemar film menandakan bahwa masyarakat sangat antusias dengan film tersebut dan dengan menyukai halaman penggemar film melalui media sosial seperti Facebook akan menarik perhatian masyarakat yang lain untuk menonton dari mendapatkan pemberitahuan tentang sebuah film dari masyarakat yang menyukai halaman penggemar film.
24. **Total Spend**, merupakan jumlah pengeluaran yang dikeluarkan oleh rumah produksi untuk memproduksi sebuah film. Jumlah pengeluaran sudah

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

mencakup seluruh perhitungan dengan biaya tidak terduga. Fitur ini cukup penting untuk membandingkan dengan fitur *gross* agar mengetahui keuntungan yang didapat dari produksi sebuah film.

#### 3.4.2 *Data Preprocessing*

*Data preprocessing* atau pemrosesan data adalah metode untuk memproses data mentah agar nilai setiap fitur memiliki format yang sama. Data film yang dipakai adalah film yang tahun rilisnya 2006 sampai 2016. Setelah memilih film yang tahun rilisnya antara 2006 sampai 2016, jumlah data berkurang dari 5044 menjadi 2425. Kemudian membuang data film yang berdurasi dibawah 50 menit karena merupakan kategori film pendek yang tidak bisa mendapatkan *title box office*. Data film berkurang dari 2425 menjadi 2408.

##### 3.4.2.1 *Data Deduplication*

*Data deduplication* atau deduplikasi data adalah teknik untuk mengeliminasi data film yang berulang atau sama agar mengurangi data yang akan diproses sehingga mempercepat performa. Deduplikasi data juga dipakai untuk mencegah data film yang sama dipakai lebih dari satu kali agar model prediksi yang dihasilkan seimbang. Pada penelitian ini, akan dilakukan eliminasi film ketika judul film dan tahun rilis sama. Setelah dilakukan deduplikasi data, jumlah dataset berkurang dari 2408 menjadi 2270.

##### 3.4.2.2 *Data Imputation*

*Data imputation* adalah teknik untuk mengubah nilai yang kosong menjadi memiliki nilai. *Data imputation* karena setiap data sangat penting untuk membuat model prediksi, jika data kosong dibuang maka akan mengurangi keragaman data

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

untuk membuat model prediksi dan akan mengurangi akurasi yang didapat. Dalam penelitian ini, setiap fitur akan diperiksa data *outlier*, bila sebuah fitur memiliki data *outlier* yang banyak, data kosong pada sebuah fitur akan diisi menggunakan *median* dari data fitur, sebaliknya akan memakai *mean* bila data *outlier* hanya sedikit.

**Tabel 3.3** *Data Imputation*

| Kode<br>Fitur | <i>Data Imputation</i> |
|---------------|------------------------|
| V1            | <i>Median</i>          |
| V2            | <i>Mean</i>            |
| V3            | <i>Median</i>          |
| V4            | <i>Median</i>          |
| V5            | <i>Median</i>          |
| V6            | <i>Median</i>          |
| V7            | <i>Median</i>          |
| V8            | <i>Median</i>          |
| V9            | <i>Mean</i>            |
| V10           | <i>Mean</i>            |
| V11           | <i>Mean</i>            |
| V12           | <i>Mean</i>            |
| V14           | <i>Median</i>          |
| V15           | <i>Median</i>          |
| V16           | <i>Mean</i>            |
| V17           | <i>Median</i>          |
| V18           | <i>Median</i>          |
| V19           | <i>Median</i>          |
| V20           | <i>Median</i>          |
| V21           | <i>Mean</i>            |



**Tabel 3.3** *Data Imputation* (Lanjutan)

| Kode<br>Fitur | <i>Data Imputation</i> |
|---------------|------------------------|
| V22           | <i>Mean</i>            |
| V23           | <i>Median</i>          |
| V25           | <i>Median</i>          |
| V26           | <i>Median</i>          |
| V27           | <i>Median</i>          |
| V28           | <i>Median</i>          |

#### 3.4.2.3 *Data Transformation*

*Data transformation* merupakan teknik untuk normalisasi data. Pada penelitian ini, *data transformation* dipakai mengubah data yang berupa teks menjadi angka dengan nilai berdasarkan tingkat kesuksesan film meraih *title box office*. Fitur yang datanya berupa teks dan tidak bisa diubah ke dalam nilai tingkat kesuksesannya, seperti judul film, link situs film, dan tahun rilis, akan dibuang.

#### 3.4.2.4 Diskritisasi Data

Diskritisasi data merupakan teknik untuk mengubah nilai data yang bersifat kontinu ke data yang bersifat diskrit. Data yang bersifat kontinu memiliki nilai pecahan sedangkan data diskrit memiliki nilai yang bulat. Data kontinu didapatkan dari hasil mengukur sedangkan diskrit didapatkan dari hasil menghitung. *Symmetrical Uncertainty* membutuhkan data yang bersifat diskrit dalam pemrosesannya. Dilakukan perhitungan dengan persamaan 2.14 untuk mengubah data kontinu menjadi diskrit. Contoh perhitungan diambil 5 data film secara acak (A1, A2, A3, A4, dan A5) dengan nilai 5 fitur V1, V2, V3, V4, dan V5.

### 3. ANALISIS DAN PERANCANGAN SISTEM

**Tabel 3.4** Contoh Data untuk Mencari Interval

| <div> <div>Fitur</div> <div>Film</div> </div> | V1       | V2   | V3    | V4       | V5    |
|---|----------|------|-------|----------|-------|
| A1  | 0.344144 | 1    | 0.723 | 2.966667 | 0     |
| A2  | 0.344144 | 0.25 | 0.302 | 2.816667 | 0.563 |
| A3  | 0.344144 | 0.4  | 0.602 | 2.466667 | 0     |
| A4  | 0.344144 | 0.6  | 0.813 | 2.733333 | 22    |
| A5  | 0.344144 | 0    | 0.462 | 2.2      | 0.475 |

Dilanjutkan dengan perhitungan menggunakan persamaan 2 . 14 dan 2 . 15 untuk menentukan total kategori yang didapat dan interval nilai masing-masing kategori.

$$k = 1 + 3.32\log(5) = 3$$

$$I(V1) = \frac{\max(V1) - \min(V1)}{k} = \frac{0.344144 - 0.344144}{3} = 0$$

$$I(V2) = \frac{\max(V2) - \min(V2)}{k} = \frac{1 - 0}{3} = 0.333333$$

$$I(V3) = \frac{\max(V3) - \min(V3)}{k} = \frac{0.813 - 0.302}{3} = 0.170333$$

$$I(V4) = \frac{\max(V4) - \min(V4)}{k} = \frac{2.966667 - 2.2}{3} = 0.766667$$

$$I(V5) = \frac{\max(V5) - \min(V5)}{k} = \frac{22 - 0}{3} = 7.333333$$

Setelah didapatkan total kategori dan interval masing-masing kategori, nilai data akan diubah berdasarkan interval nya. Berikut adalah hasil dari persamaan 2 . 15:

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

**Tabel 3.5** Hasil Interval

| <div>Fitur<br/>Film</div> | V1 | V2 | V3 | V4 | V5 |
|---------------------------|----|----|----|----|----|
| A1                        | 1  | 3  | 3  | 3  | 1  |
| A2                        | 1  | 1  | 1  | 3  | 1  |
| A3                        | 1  | 2  | 2  | 2  | 1  |
| A4                        | 1  | 2  | 3  | 3  | 3  |
| A5                        | 1  | 1  | 1  | 1  | 1  |

#### 3.4.3 *Symmetrical Uncertainty*

Dalam penelitian ini, dataset yang dipakai memiliki jumlah fitur sebanyak 28 buah. Setelah dilakukan analisis fitur, dapat disimpulkan bahwa tidak semua fitur mempengaruhi hasil prediksi *title box office* film. Maka, dibutuhkan seleksi fitur untuk membuang fitur yang tidak berpengaruh terhadap hasil prediksi *sistem box office* film.

Seleksi fitur pada penelitian ini memakai metode *Symmetrical Uncertainty*. *Symmetrical Uncertainty* merupakan turunan dari metode seleksi fitur *Information Gain* dengan kelebihanannya mengurangi nilai bias yang dihasilkan oleh *Information Gain*. Nilai yang dihasilkan dari perhitungan *Symmetrical Uncertainty* terhadap sebuah fitur memberikan informasi bahwa fitur tersebut memberikan pengaruh sebanyak nilai yang dihasilkan terhadap label kelas. Nilai *Symmetrical Uncertainty* yang dihasilkan antara 0 sampai 1, semakin besar nilai yang dihasilkan, maka fitur tersebut memberikan pengaruh yang semakin besar terhadap label kelas *box office*. Kemudian nilai *Symmetrical Uncertainty* setiap fitur akan dibandingkan dengan *threshold* yang diberikan. Fitur yang terpilih yaitu fitur yang memiliki nilai *Symmetrical Uncertainty* terhadap label kelas *box office* yang lebih tinggi dari *threshold*.

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

Berikut adalah *pseudocode* yang digunakan untuk menghitung nilai *Symmetrical Uncertainty* sebuah fitur terhadap label kelas:

**Pseudocode Symmetrical Uncertainty**

**Input :**

x : data feature  
y : data test  
t : threshold

**Output :**

su : symmetrical uncertainty of x

begin

ig = 2 \* (entropy(x) - conditional\_entropy(y, x))  
su = ig / (entropy(x) + entropy(y))

return su

end

Output su.

**Gambar 3.3** *Pseudocode Symmetrical Uncertainty*

Dalam mencari *Symmetrical Uncertainty*, dibutuhkan nilai entropi dan *Conditional Entrophy* yang nilai variabel pertama bergantung dari nilai variabel kedua. entropi merupakan keragaman pada himpunan data suatu fitur. Semakin tinggi nilai entropi, keragaman himpunan data suatu fitur semakin tinggi.

Berikut adalah contoh hasil perhitungan Entropi setiap fitur terhadap label kelas dengan persamaan (2 . 16):

### 3. ANALISIS DAN PERANCANGAN SISTEM

$$Entropy(C) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \left(\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 0.97094$$

$$Entropy(V1) = -\left(\frac{5}{5}\log_2\left(\frac{5}{5}\right)\right) = 0$$

$$Entropy(V2) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \left(\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \left(\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) = 1.521928$$

$$Entropy(V3) = -\left(\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) + \left(\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) + \left(\frac{2}{5}\log_2\left(\frac{2}{5}\right)\right) = 1.521928$$

$$Entropy(V4) = -\left(\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) + \left(\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) + \left(\frac{3}{5}\log_2\left(\frac{3}{5}\right)\right) = 1.370951$$

$$Entropy(V5) = -\left(\frac{4}{5}\log_2\left(\frac{4}{5}\right)\right) + \left(\frac{0}{5}\log_2\left(\frac{0}{5}\right)\right) + \left(\frac{1}{5}\log_2\left(\frac{1}{5}\right)\right) = 0.721928$$

Melalui perhitungan Entropi diatas, diketahui bahwa fitur *class*, *director\_name*, *num\_critic\_for\_review*, *duration*, dan *director\_facebook\_likes* memiliki keragaman data yang tinggi, sedangkan fitur *color* tidak memiliki keragaman data sama sekali atau dengan kata lain data pada fitur tersebut semua sama. Selanjutnya menghitung Entropi fitur terhadap label kelas dengan menggunakan persamaan (2 . 17).

$$Entropy(C, V1) = \left(-\frac{5}{5} \times \left(\frac{2}{5}\log_2\left(\frac{2}{5}\right) + \frac{3}{5}\log_2\left(\frac{3}{5}\right)\right)\right) = 0.97094$$

$$Entropy(C, V2) = \left(-\frac{2}{5} \times \left(\frac{0}{2}\log_2\left(\frac{0}{2}\right) + \left(\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right)\right)\right) + \left(-\frac{2}{5} \times \left(\frac{1}{2}\log_2\left(\frac{1}{2}\right) + \left(\frac{1}{2}\log_2\left(\frac{1}{2}\right)\right)\right)\right) + \left(-\frac{1}{5} \times \left(\frac{1}{1}\log_2\left(\frac{1}{1}\right) + \left(\frac{0}{1}\log_2\left(\frac{0}{1}\right)\right)\right)\right) = 0 + 0.4 + 0 = 0.4$$

$$Entropy(C, V3) = \left(-\frac{2}{5} \times \left(\frac{0}{2}\log_2\left(\frac{0}{2}\right) + \left(\frac{2}{2}\log_2\left(\frac{2}{2}\right)\right)\right)\right) + \left(-\frac{1}{5} \times \left(\frac{0}{1}\log_2\left(\frac{0}{1}\right) + \left(\frac{1}{1}\log_2\left(\frac{1}{1}\right)\right)\right)\right) + \left(-\frac{2}{5} \times \left(\frac{2}{2}\log_2\left(\frac{2}{2}\right) + \left(\frac{0}{2}\log_2\left(\frac{0}{2}\right)\right)\right)\right) = 0 + 0 + 0 = 0$$

### 3. ANALISIS DAN PERANCANGAN SISTEM

$$Entropy(C, V4) = (-\frac{1}{5} \times (\frac{0}{1} \log_2(\frac{0}{1}) + (\frac{1}{1} \log_2(\frac{1}{1}))) + (-\frac{1}{5} \times (\frac{0}{1} \log_2(\frac{0}{1}) + (\frac{1}{1} \log_2(\frac{1}{1}))) + (-\frac{3}{5} \times (\frac{2}{3} \log_2(\frac{2}{3}) + (\frac{1}{3} \log_2(\frac{1}{3}))) = 0 + 0 + 0.550978 = 0.550978$$

$$Entropy(C, V5) = (-\frac{4}{5} \times (\frac{1}{4} \log_2(\frac{1}{4}) + (\frac{3}{4} \log_2(\frac{3}{4}))) + (-\frac{0}{5} \times (\frac{0}{0} \log_2(\frac{0}{0}) + (\frac{0}{0} \log_2(\frac{0}{0}))) + (-\frac{1}{5} \times (\frac{1}{1} \log_2(\frac{1}{1}) + (\frac{0}{1} \log_2(\frac{0}{1}))) = 0.649022 + 0 + 0 = 0.649022$$

Dengan menggunakan persamaan (2 . 18), didapatkan *Symmetrical Uncertainty* setiap fitur terhadap label kelas.

$$SU(C, V1) = 2 \times (\frac{0.97094 - 0.97094}{0.97094 + 0}) = 0$$

$$SU(C, V2) = 2 \times (\frac{0.97094 - 0.4}{0.97094 + 1.521928}) = 0.458058$$

$$SU(C, V3) = 2 \times (\frac{0.97094 - 0}{0.97094 + 1.521928}) = 0.778974$$

$$SU(C, V4) = 2 \times (\frac{0.97094 - 0.550978}{0.97094 + 1.370951}) = 0.358652$$

$$SU(C, V5) = 2 \times (\frac{0.97094 - 0.649022}{0.97094 + 0.721928}) = 0.380346$$

Berikut adalah hasil perhitungan *Symmetrical Uncertainty* seluruh fitur:

**Tabel 3.6** Hasil *Symmetrical Uncertainty*

| Kode<br>Fitur | Hasil <i>Symmetrical Uncertainty</i> |
|---------------|--------------------------------------|
| V1            | 0                                    |
| V2            | 0.458058                             |
| V3            | 0.778974                             |

**Tabel 3.6** Hasil *Symmetrical Uncertainty* (Lanjutan)

| Kode<br>Fitur | Hasil <i>Symmetrical Uncertainty</i> |
|---------------|--------------------------------------|
| V4            | 0.358652                             |
| V5            | 0.380346                             |
| V6            | 0.020559                             |
| V7            | 0.380346                             |
| V8            | 0.380346                             |
| V9            | 0.358652                             |
| V10           | 0.358652                             |
| V11           | 0.458058                             |
| V12           | 0.458058                             |
| V13           | 0.432531                             |
| V15           | 1                                    |
| V16           | 0.358652                             |
| V17           | 0.432531                             |
| V18           | 1                                    |
| V20           | 0.47356                              |
| V21           | 0                                    |
| V22           | 0.201953                             |
| V23           | 0                                    |
| V25           | 0.358652                             |
| V26           | 1                                    |
| V27           | 0.380323                             |
| V28           | 0.358652                             |

Setelah perhitungan *Symmetrical Uncertainty* masing - masing fitur terhadap label kelas, setiap hasil *Symmetrical Uncertainty* dibandingkan misalnya dengan threshold ( $\delta$ ) = 0.4. Fitur akan dipilih jika perhitungan *Symmetrical Uncertainty* fitur terhadap label kelas tersebut melebihi nilai threshold yang ditentukan.

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

Didapatkan 10 fitur yang berpengaruh terhadap label kelas *box office* dari hasil seleksi fitur, yaitu *director\_name*, *num\_critic\_for\_review*, *actor\_3\_name*, *gross*, *genres*, *num\_voted\_users*, *facenumber\_in\_poster*, *plot\_keywords*, *num\_user\_for\_reviews*, dan *imdb\_score*.

**Tabel 3.7** Hasil Seleksi Fitur dengan *Symmetrical Uncertainty*

| Kode<br>Fitur | Hasil<br><i>Symmetrical<br/>Uncertainty</i> |
|---------------|---|
| V2            | 0.458058                                    |
| V3            | 0.778974                                    |
| V11           | 0.458058                                    |
| V12           | 0.458058                                    |
| V13           | 0.432531                                    |
| V15           | 1   |
| V17           | 0.432531                                    |
| V18           | 1   |
| V20           | 0.47356                                     |
| V26           | 1   |

#### 3.4.4 *Support Vector Machine*

*Support Vector Machine* merupakan metode klasifikasi yang dipakai pada penelitian ini untuk menghasilkan model prediksi pada data pelatihan dan dipakai untuk memprediksi *title box office* film pada data pengujian.

Berikut adalah *pseudocode* untuk klasifikasi *Support Vector Machine* untuk pelatihan dan pengujian:



#### **Pseudocode SVM Training**

##### **Input :**

x\_train : data train feature  
y\_train : data train label  
sigma : sigma  
c : regularization

##### **Output :**

a : alpha  
b : bias  
y\_train : data train label  
sv : support vector

begin

qp = QuadraticProgramming(sigma, c)

a = qp['alpha']

b = 0

sv = qp['sv']

for i = 0 to length(a) do

    b = b + y\_train

    b = b - sum(a \* y\_train \* RBF(x\_train[i], sv))

end for

b = b / length(a)

return a, b, y\_train, and sv

end

*Output alpha, bias, label train, and support vector.*

**Gambar 3.4** *Pseudocode Support Vector Machine Training*

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

#### **Pseudocode SVM Testing**

##### **Input :**

```
y_train : data train label
x_test : data test feature
y_test : data test label
a : alpha
b : bias
sv : support vector
```

##### **Output :**

```
accuracy : predicted label
```

```
begin
  for i = 0 to length(x_test) do
    s = 0
    for a, y_train, sv in zip(a, y_train, sv) do
      s += a * y_train * RBF(x_test[i], sv)
    end for
    y_predict[i] = s
  end for

  accuracy = y_predict + b

  return accuracy
end

Output accuracy.
```

**Gambar 3.5** *Pseudocode Support Vector Machine Testing*

Pada bagian ini diberikan contoh perhitungan *Support Vector Machine* dengan 9 fitur dan 5 data dengan menggunakan pendekatan linear (tanpa *Quadratic Programming*, *Lagrange Multiplier*, dan *Soft Margin*). Pendekatan linear dipakai pada bagian ini dikarenakan implementasi *Quadratic Programming*, *Lagrange Multiplier*, dan *Soft Margin* yang sulit tanpa menggunakan *library*. Dikarenakan tidak dapat memakai *Soft Margin*, maka perhitungan pada bagian ini hanya memakai nilai *gamma*, tidak memakai nilai regularisasi (C). Nilai *gamma* yang dipakai adalah 1 pada perhitungan RBF kernel untuk menghitung *dot product*. Kemudian dilakukan pencarian nilai *alpha* dan bias yang akan digunakan pada fungsi klasifikasi untuk menghasilkan prediksi. Berikut adalah contoh perhitungan klasifikasi pada data pelatihan dengan 5 data yaitu A1, A2, A3, A4, dan A5:

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

**Tabel 3.8** Klasifikasi Data Pelatihan

| <div>Fitur<br/>Film</div> | V2   | V3    | V11 | V12      | V13      |
|---------------------------|------|-------|-----|----------|----------|
| A1                        | 1    | 0.723 | 1   | 0.760506 | 0.247722 |
| A2                        | 0.25 | 0.302 | 0.5 | 0.309404 | 0.241976 |
| A3                        | 0.4  | 0.602 | 0   | 0.200074 | 0.254328 |
| A4                        | 0.6  | 0.813 | 0.5 | 0.448131 | 0.259239 |
| A5                        | 0    | 0.462 | 0   | 0.073059 | 0.242714 |

**Tabel 3.9** Klasifikasi Data Pelatihan (Lanjutan)

| <div>Fitur<br/>Film</div> | V15      | V18      | Class |
|---------------------------|----------|----------|-------|
| A1                        | 0.886204 | 0.585294 | 1     |
| A2                        | 0.047122 | 0.057143 | -1    |
| A3                        | 0.275868 | 0.234444 | -1    |
| A4                        | 1.144337 | 0.722857 | 1     |
| A5                        | 0.212204 | 0.058333 | -1    |

Pada tabel 3.8, kelas 1 menunjukkan perolehan *title box office* film, sedangkan -1 menunjukkan bahwa film tidak meraih *title box office*. Tahap selanjutnya adalah menghitung RBF kernel dengan menginisialisasikan parameter  $\gamma$  dengan 1. Berikut adalah contoh perhitungan dengan nilai RBF kernel dengan menggunakan data pelatihan dari 3.8:

### 3. ANALISIS DAN PERANCANGAN SISTEM

$$K(A1,A1) = \exp(-\gamma|A1-A1|^2) = \exp - ((|1-1|)^2 + (|0.723-0.723|)^2 + (|1-1|)^2 + (|0.760506-0.760506|)^2 + (|0.247722-0.247722|)^2 + (|0.886204-0.886204|)^2 + (|0.585294-0.585294|)^2) = 1$$

$$K(A1,A2) = \exp(-\gamma|A1-A2|^2) = \exp - ((|1-0.25|)^2 + (|0.723-0.302|)^2 + (|1-0.5|)^2 + (|0.760506-0.309404|)^2 + (|0.247722-0.241976|)^2 + (|0.886204-0.047122|)^2 + (|0.585294-0.057143|)^2) = 0.336844$$

$$K(A1,A3) = \exp(-\gamma|A1-A3|^2) = \exp - ((|1-0.4|)^2 + (|0.723-0.602|)^2 + (|1-0|)^2 + (|0.760506-0.200074|)^2 + (|0.247722-0.254328|)^2 + (|0.886204-0.275868|)^2 + (|0.585294-0.234444|)^2) = 0.335482$$

$$K(A1,A4) = \exp(-\gamma|A1-A4|^2) = \exp - ((|1-0.6|)^2 + (|0.723-0.813|)^2 + (|1-0.5|)^2 + (|0.760506-0.448131|)^2 + (|0.247722-0.259239|)^2 + (|0.886204-0.144337|)^2 + (|0.585294-0.722857|)^2) = 0.740312$$

$$K(A1,A5) = \exp(-\gamma|A1-A5|^2) = \exp - ((|1-0|)^2 + (|0.723-0.462|)^2 + (|1-0|)^2 + (|0.760506-0.073059|)^2 + (|0.247722-0.247714|)^2 + (|0.886204-0.212204|)^2 + (|0.585294-0.058333|)^2) = 0.19469$$

Lakukan perhitungan diatas dengan data lainnya sehingga menghasilkan tabel model prediksi sebagai berikut:

**Tabel 3.10** Hasil dari Klasifikasi Data Pelatihan

|    | A1       | A2       | A3       | A4       | A5       |
|----|----------|----------|----------|----------|----------|
| A1 | 1        | 0.336844 | 0.335482 | 0.740312 | 0.19469  |
| A2 | 0.336844 | 1        | 0.795183 | 0.358757 | 0.810093 |
| A3 | 0.335482 | 0.795183 | 1        | 0.499358 | 0.890949 |
| A4 | 0.740312 | 0.358757 | 0.499358 | 1        | 0.335441 |
| A5 | 0.19469  | 0.810093 | 0.890949 | 0.335441 | 1        |

### 3. ANALISIS DAN PERANCANGAN SISTEM

Setelah mendapatkan hasil perhitungan dengan RBF kernel, selanjutnya data tersebut akan dimasukkan ke dalam persamaan *linear* untuk mencari nilai alpha ( $\alpha$ ) dan bias (b). Berikut adalah persamaan linear dari data pelatihan :

$$\begin{aligned}
 &(1)\alpha_1 + (-1)\alpha_2 + (-1)\alpha_3 + (1)\alpha_4 + (-1)\alpha_5 + (0)b = 0 \\
 A1 &= (1)\alpha_1 - (0.336844)\alpha_2 - (0.335482)\alpha_3 + (0.740312)\alpha_4 - (0.19469)\alpha_5 + b = 1 \\
 A2 &= (0.336844)\alpha_1 - (1)\alpha_2 - (0.795183)\alpha_3 + (0.358757)\alpha_4 - (0.810093)\alpha_5 + b = -1 \\
 A3 &= (0.335482)\alpha_1 - (0.795183)\alpha_2 - (1)\alpha_3 + (0.499358)\alpha_4 - (0.890949)\alpha_5 + b = -1 \\
 A4 &= (0.740312)\alpha_1 - (0.358757)\alpha_2 - (0.499358)\alpha_3 + (1)\alpha_4 - (0.335441)\alpha_5 + b = 1 \\
 A5 &= (0.19469)\alpha_1 - (0.810093)\alpha_2 - (0.890949)\alpha_3 + (0.335441)\alpha_4 - (1)\alpha_5 + b = -1
 \end{aligned}$$

Dari persamaan linear diatas, didapatkan hasil seperti berikut:

**Tabel 3.11** Hasil dari Persamaan Linear

| $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\alpha_4$ | $\alpha_5$ | b        |
|------------|------------|------------|------------|------------|----------|
| 0.582162   | 0.828194   | 2.282193   | 1.57626    | -0.95196   | 0.110183 |

Setelah mendapatkan nilai *alpha* dan bias, maka dapat dilakukan klasifikasi pada data uji untuk memprediksi kelas yang tepat dengan data film B1. Sama seperti data pelatihan, data uji memakai RBF kernel dengan nilai parameter  $\gamma$  yang sama. Berikut adalah hasil perhitungan data uji dengan 1 data film yaitu data B1:

**Tabel 3.12** Klasifikasi Data Uji

| Fitur<br>Film | V2 | V3    | V11      | V12   | V13      |
|---------------|----|-------|----------|-------|----------|
| B1            | 1  | 0.635 | 0.666667 | 0.459 | 0.242714 |

**Tabel 3.13** Klasifikasi Data Uji (Lanjutan)

| Fitur<br>Film | V15   | V18      | Class |
|---------------|-------|----------|-------|
| B1            | 0.463 | 0.454286 | ?     |

### 3. ANALISIS DAN PERANCANGAN SISTEM

$$K(B1, A1) = \exp(-\gamma|B1 - A1|^2) = \exp - ((|1 - 1|)^2 + (|0.635 - 0.723|)^2 + (|0.666667 - 1|)^2 + (|0.459 - 0.760506|)^2 + (|0.242714 - 0.247722|)^2 + (|0.463 - 0.886204|)^2 + (|0.454286 - 0.585294|)^2) = 0.816258$$

$$K(B1, A2) = \exp(-\gamma|B1 - A2|^2) = \exp - ((|1 - 0.25|)^2 + (|0.635 - 0.302|)^2 + (|0.666667 - 0.5|)^2 + (|0.459 - 0.309404|)^2 + (|0.242714 - 0.241976|)^2 + (|0.463 - 0.047122|)^2 + (|0.454286 - 0.057143|)^2) = 0.590308$$

$$K(B1, A3) = \exp(-\gamma|B1 - A3|^2) = \exp - ((|1 - 0.4|)^2 + (|0.635 - 0.602|)^2 + (|0.666667 - 0|)^2 + (|0.459 - 0.200074|)^2 + (|0.242714 - 0.254328|)^2 + (|0.463 - 0.275868|)^2 + (|0.454286 - 0.234444|)^2) = 0.620004$$

$$K(B1, A4) = \exp(-\gamma|B1 - A4|^2) = \exp - ((|1 - 0.6|)^2 + (|0.635 - 0.813|)^2 + (|0.666667 - 0.5|)^2 + (|0.459 - 0.448131|)^2 + (|0.242714 - 0.259239|)^2 + (|0.463 - 0.144337|)^2 + (|0.454286 - 0.722857|)^2) = 0.685161$$

$$K(B1, A5) = \exp(-\gamma|B1 - A5|^2) = \exp - ((|1 - 0|)^2 + (|0.635 - 0.462|)^2 + (|0.666667 - 0|)^2 + (|0.459 - 0.073059|)^2 + (|0.242714 - 0.247714|)^2 + (|0.463 - 0.212204|)^2 + (|0.454286 - 0.058333|)^2) = 0.397921$$

**Tabel 3.14** Hasil dari Klasifikasi Data Uji

| Latih \ Uji | A1       | A2       | A3       | A4       | A5       |
|-------------|----------|----------|----------|----------|----------|
| B1          | 0.816258 | 0.590308 | 0.620004 | 0.685161 | 0.397921 |

$$f(x) = \text{sign}(\sum_{i=1}^n K(B1, A_i) + b) = \text{sign}((0.582162 * 1 * 0.816258 + 0.110183) + (0.828194 * -1 * 0.590308 + 0.110183) + (2.282193 * -1 * 0.620004 + 0.110183) + (1.57626 * 1 * 0.685161 + 0.110183) + (-0.95196 * -1 * 0.397921 + 0.110183)) = 1$$

### 3. ANALISIS DAN PERANCANGAN SISTEM

---

Berdasarkan perhitungan data B1 terhadap data pelatihan dengan kernel RBF, didapatkan nilai akhir 1 yaitu data B1 diprediksi meraih *title box office*.

## **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Pada bab ini akan menjelaskan mengenai proses implementasi dan pengujian terhadap sistem yang telah dibangun berdasarkan penjelasan pada bab sebelumnya.

### **4.1 Lingkungan Implementasi**

Pada lingkungan implementasi, akan dijelaskan mengenai perangkat yang digunakan dalam proses pembangunan sistem baik dari perangkat keras maupun perangkat lunak yang digunakan.

#### **4.1.1 Spesifikasi Perangkat Keras**

Spesifikasi dari perangkat keras yang digunakan dalam pembangunan aplikasi adalah sebagai berikut:

1. Laptop Macbook Air (2017)
2. *Processor* Intel Core i5 Dual Core @ 1.8GHz
3. *Solid State Drive* kapasitas 128GB
4. RAM 8GB

#### **4.1.2 Lingkungan Perangkat Lunak**

Spesifikasi dari perangkat lunak yang digunakan dalam pembangunan aplikasi adalah sebagai berikut:

1. MacOS 10.15.3 Catalina
2. Python 3.8



### 3. Flask

#### 4.2 Daftar Kelas dan Metode

Pada bagian ini akan dijelaskan mengenai kelas dan metode yang dipakai dalam penelitian ini.

##### 4.2.1 Kelas *Symmetrical Uncertainty*

Kelas *Symmetrical Uncertainty* berisikan semua metode yang digunakan untuk melakukan aktivitas dan perhitungan seleksi fitur *Symmetrical Uncertainty* mulai dari perhitungan awal hingga hasil akhir.

**Tabel 4.1** Daftar Metode pada Kelas *Symmetrical Uncertainty*

| No | Method            | Input          |                               | Output | Keterangan   |
|----|-------------------|----------------|-------------------------------|--------|--|
|    |                   | Tipe           | Variabel                      |        |  |
| 1  | Remove Duplicates | List           | x                             | List   | Menghilangkan data duplikat pada suatu array.              |
| 2  | Count Frequency   | List           | my_list                       | List   | Menghitung frekuensi kemunculan setiap key pada list.      |
| 3  | drop              | Array, Integer | deleted<br>_features,<br>axis | -      | Menghapus baris/kolom (sesuai axis) dalam suatu DataFrame. |
| 4  | reset_index       | -              | -                             | List   | Melakukan <i>reset</i> indeks pada suatu DataFrame.        |

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

**Tabel 4.1** Daftar Metode pada Kelas *Symmetrical Uncertainty* (Lanjutan)

| No | Method     | Input                                   |                                 | Output | Keterangan  |
|----|------------|---|---------------------------------|--------|---|
|    |            | Tipe                                    | Variabel                        |        |   |
| 5  | split      | String                                  | split                           | List   | Memisahkan sebuah string menjadi array dengan parameter pemisahannya.       |
| 6  | fillna     | Integer                                 | integer                         | -      | Mengisi nilai kosong yang terdapat pada suatu array dengan nilai parameter. |
| 7  | mean       | -                                       | -                               | -      | Menghitung nilai rata-rata dari suatu array.                                |
| 8  | median     | -                                       | -                               | -      | Menghitung nilai tengah dari suatu array.                                   |
| 9  | log10      | Double                                  | angka                           | Double | Menghitung log basis 10 dari nilai parameter.                               |
| 10 | cut        | List,<br>Integer,<br>String,<br>Boolean | list, bins,<br>labels,<br>right | Array  | Mengelompokkan data dengan banyak kelompok k.                               |
| 11 | count_vals | List                                    | x                               | List   | Menghitung kemunculan dari suatu list.                                      |
| 12 | entropy    | List                                    | x                               | Double | Menghitung nilai <i>entropy</i> dari suatu fitur pada list.                 |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.1** Daftar Metode pada Kelas *Symmetrical Uncertainty* (Lanjutan)

| No | Method              | Input      |          | Output | Keterangan  |
|----|---------------------|------------|----------|--------|---|
|    |                     | Tipe       | Variabel |        |   |
| 13 | conditional_entropy | List, List | x, y     | Double | Menghitung <i>conditional entropy</i> dari x berdasarkan y. |

##### 4.2.2 Kelas *Support Vector Machine*

Kelas *Support Vector Machine* berisikan semua metode yang digunakan untuk melakukan perhitungan pada proses pelatihan untuk membuat model prediksi dan proses pengujian untuk pencocokan data uji dengan model prediksi.

**Tabel 4.2** Daftar Metode pada Kelas *Support Vector Machine*

| No | Method      | Input                     |                   | Output                    | Keterangan  |
|----|-------------|---------------------------|-------------------|---------------------------|---|
|    |             | Tipe                      | Variabel          |                           |   |
| 1  | split_train | List                      | movie             | List, List,<br>List, List | Memecah data menjadi data latih dan data uji secara random..  |
| 2  | svm_predict | List, List,<br>List, List | X1, y1, X2,<br>y2 | -                         | Menghitung prediksi <i>Support Vector Machine</i> menggunakan kernel RBF dengan parameter X1 sebagai nilai fitur data latih, y1 sebagai label kelas data latih, X2 sebagai nilai fitur data uji, dan y2 sebagai label kelas data uji. |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.2** Daftar Metode pada Kelas *Support Vector Machine* (Lanjutan)

| No | Method     | Input   |                     | Output | Keterangan   |
|----|------------|---|---------------------|--------|--|
|    |            | Tipe  | Variabel            |        |  |
| 3  | asarray    | List  | list                | Array  | Mengubah data list menjadi data array.   |
| 4  | SVM        | String,<br>Integer                                      | kernel_type,<br>C   | -      | Menginisialisasi kernel yang akan digunakan pada SVM.  |
| 5  | fit        | Object,<br>Array,<br>Array                              | self, X, y          | -      | Membuat model prediksi menggunakan data latih.   |
| 6  | zeros      | Integer,<br>Integer                                     | row,<br>column      | Array  | Membuat array matrix dengan baris dan kolom yang ditentukan. Bisa menggunakan satu parameter (column) jika memakai satu parameter. |
| 7  | solvers.qp | Array,<br>Array,<br>Array,<br>Array,<br>Array,<br>Array | P, q, G, h,<br>A, b | List   | Menghitung hasil dari <i>Quadratic Programming</i> .   |
| 8  | ravel      | Array   | array               | array  | Mengembalikan parameter array menjadi array 1-D.   |

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

**Tabel 4.2** Daftar Metode pada Kelas *Support Vector Machine* (Lanjutan)

| No | Method   | Input                      |                                  | Output  | Keterangan   |
|----|----------|----------------------------|----------------------------------|---------|--|
|    |          | Tipe                       | Variabel                         |         |  |
| 9  | predict  | Array                      | X_test                           | array   | Mencari label kelas yang didapat dari membandingkan parameter input array dengan model prediksi. |
| 10 | sign     | Double                     | double                           | Integer | Menghitung <i>sign</i> dari sebuah angka bertipe double.   |
| 11 | f1_score | Array,<br>Array,<br>String | y_test,<br>y_predict,<br>average | Double  | Menghitung <i>F-Measure</i> .  |

### 4.3 Implementasi Perangkat Lunak

Pada bagian ini akan dilakukan implementasi pada perangkat lunak dari analisis yang telah disusun pada BAB III.

#### 4.3.1 Implementasi Pemrosesan Data Film

Tahap implementasi pemrosesan data melakukan pengambilan data film yang tahun rilisnya dari 2006 hingga 2016, kemudian membuang data film yang memiliki durasi dibawah 50 menit. Tahap selanjutnya dalam pemrosesan data terbagi menjadi 4 bagian, yaitu *data deduplication*, *data imputation*, *data transformation*, dan dikritisasi data. Kemudian hasil pemrosesan data disimpan dalam bentuk file .csv.

### 4.3.1.1 Implementasi *Data Deduplication*

Pada tahap implementasi *data deduplication*, dilakukan eliminasi data film yang memiliki nilai yang berulang dengan melihat judul dan tahun rilis yang sama.

### 4.3.1.2 Implementasi *Data Imputation*

Pada tahap implementasi *data imputation*, dilakukan proses berikut untuk mengisi nilai yang kosong pada sebuah data film:

1. Menghitung banyaknya data *outlier* pada masing-masing fitur.
2. Fitur yang memiliki jumlah data *outlier* yang banyak akan memakai metode *median* untuk mengisi data yang kosong pada fitur.
3. Sedangkan fitur yang memiliki jumlah data *outlier* yang sedikit akan memakai metode *mean* untuk mengisi data yang kosong pada fitur.

### 4.3.1.3 Implementasi *Data Transformation*

Pada tahap implementasi *data transformation*, dilakukan proses berikut untuk mengubah nilai fitur yang berupa teks menjadi angka:

1. Memilih fitur yang memiliki nilai berupa teks.
2. Nilai pada fitur akan diambil terlebih dahulu dan membuang nilai yang sama.
3. Nilai pada fitur akan dihitung nilai kesuksesan dalam meraih *box office* film.
4. Nilai teks pada fitur akan diubah menjadi nilai kesuksesan dalam meraih *box office*.

### 4.3.1.4 Implementasi Diskritisasi data

Metode diskritisasi data yang dipakai adalah metode *Equal-Width*, yaitu mengelompokkan data berdasarkan panjang interval dan banyaknya kelompok yang diberikan. Berikut adalah proses yang dilakukan pada diskritisasi data:

1. Menghitung banyaknya kelompok pada suatu data.
2. Menghitung panjang interval setiap fitur.
3. Mengubah nilai pada setiap fitur dengan kelompok yang didapat berdasarkan interval.

### 4.3.2 Implementasi Seleksi Fitur

Metode seleksi fitur yang dipakai adalah metode *Symmetrical Uncertainty*. *Symmetrical Uncertainty* merupakan salah satu teknik seleksi fitur berbasis filter yang mencari nilai berdasarkan pengaruh setiap fitur terhadap label kelas *box office*. Berikut adalah proses *Symmetrical Uncertainty* sebagai seleksi fitur untuk mencari nilai pengaruh fitur terhadap label kelas *box office*:

1. Hitung nilai entropi setiap fitur dan label kelas data.
2. Hitung nilai *conditional entropy* setiap fitur terhadap label kelas.
3. Hitung nilai *Symmetrical Uncertainty* setiap fitur terhadap label kelas.
4. Simpan hasil perhitungan *Symmetrical Uncertainty* setiap fitur pada file .csv.
5. Tentukan *threshold* yang akan diuji.
6. Pilih fitur yang memiliki nilai *Symmetrical Uncertainty* sama atau diatas dari *threshold* telah ditentukan.

### 4.3.3 Implementasi Pelatihan Data

Pada implementasi pelatihan data, digunakan metode *Support Vector Machine* untuk membuat model prediksi dari *dataset* yang diberikan. Berikut adalah proses yang dilakukan dalam pelatihan data:

1. Menentukan nilai *gamma* yang akan diuji.
2. Menentukan nilai regularisasi (C) yang akan diuji.
3. Menentukan data latih yang akan dipakai dengan menggunakan *10-Fold Cross Validation*.
4. Melakukan perhitungan kernel untuk membentuk matriks RBF.
5. Melakukan perhitungan *Quadratic Programming* untuk mencari *hyperplane* yang optimum.
6. Melakukan perhitungan *Lagrange Multiplier*.
7. Menyimpan model prediksi yang didapat dari *Langrange Multiplier* ke dalam .sav.

### 4.3.4 Implementasi Pengujian Data

Berikut adalah proses untuk implementasi pengujian data menggunakan model prediksi yang telah didapat:

1. Menentukan nilai *gamma* yang akan diuji.
2. Menentukan nilai regularisasi (C) yang akan diuji.
3. Menentukan data uji yang akan dipakai menggunakan *10-fold Cross Validation*.
4. Melakukan perhitungan kernel untuk membentuk matriks RBF.



#### 4. IMPLEMENTASI DAN PENGUJIAN

---

5. Melakukan prediksi dengan membandingkan pada model prediksi yang dihasilkan dari pelatihan data.
6. Menghitung nilai akurasi dan *F-Measure*.

#### 4.4 Pengujian dan Evaluasi

Pada bagian ini dibagi menjadi dua, yaitu tahap seleksi fitur dan tahap klasifikasi. Pada tahap pengujian seleksi fitur, dilakukan pengujian *threshold* menggunakan *dataset* film yang ada. Pada tahap pengujian klasifikasi, dilakukan berbagai pengujian untuk menentukan akurasi dan *F-Measure* yang didapat. Data pengujian akan memakai *10-fold Cross Validation* untuk memisahkan antara data latih dan data uji.

##### 4.4.1 Pengujian Seleksi Fitur *Symmetrical Uncertainty*

Pada bagian ini akan dijelaskan analisis tahap pengujian seleksi fitur menggunakan metode *Symmetrical Uncertainty* dan *threshold* yang beragam untuk memunculkan kombinasi fitur yang beragam.

##### 4.4.1.1 Analisis Seleksi Fitur *Symmetrical Uncertainty*

Seleksi fitur menggunakan metode *Symmetrical Uncertainty*. Analisis pengujian dilakukan dengan menghitung nilai *Symmetrical Uncertainty* setiap fitur. Berikut adalah hasil perhitungan *Symmetrical Uncertainty* pada setiap fitur:

**Tabel 4.3** Hasil Pengujian *Symmetrical Uncertainty*

| Fitur        | Hasil <i>Symmetrical Uncertainty</i> |
|--------------|--------------------------------------|
| <i>color</i> | 0.0023563965108736686                |

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

**Tabel 4.3** Hasil Pengujian *Symmetrical Uncertainty* (Lanjutan)

| <b>Fitur</b>                     | <b>Hasil <i>Symmetrical Uncertainty</i></b> |
|----------------------------------|---|
| <i>director_name</i>             | 0.3484488420725572                          |
| <i>num_critic_for_reviews</i>    | 0.041232985064060895                        |
| <i>duration</i>                  | 0.003743450146065106                        |
| <i>director_facebook_likes</i>   | 0.011398533749682648                        |
| <i>actor_3_facebook_likes</i>    | 0.01594193267923876                         |
| <i>actor_2_name</i>              | 0.29840154294764937                         |
| <i>actor_1_facebook_likes</i>    | 0.008553965089746018                        |
| <i>genres</i>                    | 0.01207211522439701                         |
| <i>actor_1_name</i>              | 0.1833456060004809                          |
| <i>num_voted_users</i>           | 0.06674229074369575                         |
| <i>cast_total_facebook_likes</i> | 0.007713381403220423                        |
| <i>actor_3_name</i>              | 0.38115578174323783                         |
| <i>facenumber_in_poster</i>      | 0.007325363862537581                        |
| <i>plot_keywords</i>             | 0.38536166614258266                         |
| <i>num_user_for_reviews</i>      | 0.04142403860157818                         |
| <i>language</i>                  | 0.03908416827846718                         |
| <i>country</i>                   | 0.03584643036841998                         |
| <i>content_rating</i>            | 0.03262980884930194                         |
| <i>budget</i>                    | 0.002388385458359331                        |
| <i>actor_2_facebook_likes</i>    | 0.00787972995681825                         |
| <i>imdb_score</i>                | 0.011029718643468656                        |
| <i>aspect_ratio</i>              | 0.002020149882203536                        |
| <i>movie_facebook_likes</i>      | 0.04716879688528342                         |
| <i>total_spend</i>               | 0.002388385458359331                        |

Hasil dari pengujian seleksi fitur *Symmetrical Uncertainty* berupa kombinasi - kombinasi fitur yang memiliki nilai diatas threshold yang diuji. Threshold yang

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

akan diuji adalah 0, 0.01, 0.05, dan 0.1. Berikut adalah kombinasi - kombinasi fitur yang didapat pada setiap pengujian threshold yang berbeda:

**Tabel 4.4** Hasil Pengujian *Symmetrical Uncertainty* dengan *Threshold* 0.01

| Kode Fitur                     | Hasil <i>Symmetrical Uncertainty</i> |
|--------------------------------|--------------------------------------|
| <i>director_name</i>           | 0.3484488420725572                   |
| <i>num_critic_for_reviews</i>  | 0.041232985064060895                 |
| <i>director_facebook_likes</i> | 0.011398533749682648                 |
| <i>actor_3_facebook_likes</i>  | 0.01594193267923876                  |
| <i>actor_2_name</i>            | 0.29840154294764937                  |
| <i>genres</i>                  | 0.01207211522439701                  |
| <i>actor_1_name</i>            | 0.1833456060004809                   |
| <i>num_voted_users</i>         | 0.06674229074369575                  |
| <i>actor_3_name</i>            | 0.38115578174323783                  |
| <i>plot_keywords</i>           | 0.38536166614258266                  |
| <i>num_user_for_reviews</i>    | 0.04142403860157818                  |
| <i>language</i>                | 0.03908416827846718                  |
| <i>country</i>                 | 0.03584643036841998                  |
| <i>content_rating</i>          | 0.03262980884930194                  |
| <i>imdb_score</i>              | 0.011029718643468656                 |
| <i>movie_facebook_likes</i>    | 0.04716879688528342                  |

**Tabel 4.5** Hasil Pengujian *Symmetrical Uncertainty* dengan *Threshold* 0.05

| Kode Fitur             | Hasil <i>Symmetrical Uncertainty</i> |
|------------------------|--------------------------------------|
| <i>director_name</i>   | 0.3484488420725572                   |
| <i>actor_2_name</i>    | 0.29840154294764937                  |
| <i>actor_1_name</i>    | 0.1833456060004809                   |
| <i>num_voted_users</i> | 0.06674229074369575                  |
| <i>actor_3_name</i>    | 0.38115578174323783                  |

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

**Tabel 4.5** Hasil Pengujian *Symmetrical Uncertainty* dengan *Threshold* 0.05 (Lanjutan)

| Kode Fitur           | Hasil <i>Symmetrical Uncertainty</i> |
|----------------------|--------------------------------------|
| <i>plot_keywords</i> | 0.38536166614258266                  |

**Tabel 4.6** Hasil Pengujian *Symmetrical Uncertainty* dengan *Threshold* 0.1

| Kode Fitur           | Hasil <i>Symmetrical Uncertainty</i> |
|----------------------|--------------------------------------|
| <i>director_name</i> | 0.3484488420725572                   |
| <i>actor_2_name</i>  | 0.29840154294764937                  |
| <i>actor_1_name</i>  | 0.1833456060004809                   |
| <i>actor_3_name</i>  | 0.38115578174323783                  |
| <i>plot_keywords</i> | 0.38536166614258266                  |

#### 4.4.2 Pengujian Klasifikasi *Support Vector Machine* dengan *Symmetrical Uncertainty*

Pada bagian ini akan dijelaskan analisis tahap pengujian klasifikasi menggunakan metode *Support Vector Machine* untuk mendapatkan akurasi dan *F-Measure* terbaik dengan nilai *gamma* dan nilai regularisasi (C) yang beragam juga kombinasi fitur yang telah didapatkan dari pengujian pemilihan fitur.

##### 4.4.2.1 Skenario dan Hasil Pengujian Klasifikasi *Support Vector Machine* dengan *Symmetrical Uncertainty*

Skenario klasifikasi *Support Vector Machine* menggunakan *dataset* yang berasal dari situs *www.kaggle.com*. Data latih dan data uji dipisah menggunakan *10-fold Cross Validation*. Skenario yang akan diuji adalah pengujian yang menggunakan seleksi fitur *Symmetrical Uncertainty* dengan *threshold* 0, 0.01, 0.05, dan 0.1. Kemudian dilakukan pengujian dengan menggunakan *gamma* 0.01, 0.1, 1, dan 2. Dan diuji kembali menggunakan nilai regularisasi dengan nilai 0.1, 1, 10, dan 100. Setelah itu akan dibandingkan dengan skenario tanpa menggunakan seleksi fitur

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

*Symmetrical Uncertainty* untuk mengetahui pengaruh seleksi fitur *Symmetrical Uncertainty* terhadap prediksi *title box office* film. Skenario tanpa seleksi fitur *Symmetrical Uncertainty* tetap menggunakan nilai *gamma* dan nilai regularisasi yang sama.

Berikut adalah skenario pengujian yang menggunakan *Symmetrical Uncertainty* dan rata-rata hasil pengujian berupa akurasi dan *F-Measure* yang didapat:

**Tabel 4.7** Hasil Pengujian *Support Vector Machine*

| Skenario | <i>SU Threshold</i> | <i>Gamma</i> | C   | Rata-rata Akurasi  | Rata-rata <i>F-Measure</i> |
|----------|---------------------|--------------|-----|--------------------|----------------------------|
| 1        | 0                   | 0.01         | 0.1 | 0.9929179117710412 | 0.984413107018999          |
| 2        | 0                   | 0.01         | 1   | 0.994270186037785  | 0.9875130372172357         |
| 3        | 0                   | 0.01         | 10  | 0.9942707840873549 | 0.9870256975243084         |
| 4        | 0                   | 0.01         | 100 | 0.9924938273974391 | 0.9821209893757846         |
| 5        | 0                   | 0.1          | 0.1 | 0.8086005557246828 | 0.4630667119001094         |
| 6        | 0                   | 0.1          | 1   | 0.9572355648234396 | 0.883640913250822          |
| 7        | 0                   | 0.1          | 10  | 0.9573017621196922 | 0.8841529059662045         |
| 8        | 0                   | 0.1          | 100 | 0.9573017621196922 | 0.8850885784808243         |
| 9        | 0                   | 1            | 0.1 | 0.6746054922345027 | 0                          |
| 10       | 0                   | 1            | 1   | 0.6746054922345027 | 0                          |
| 11       | 0                   | 1            | 10  | 0.6746054922345027 | 0                          |
| 12       | 0                   | 1            | 100 | 0.6746054922345027 | 0                          |
| 13       | 0                   | 2            | 0.1 | 0.6746054922345027 | 0                          |
| 14       | 0                   | 2            | 1   | 0.6746054922345027 | 0                          |
| 15       | 0                   | 2            | 10  | 0.6746054922345027 | 0                          |
| 16       | 0                   | 2            | 100 | 0.6746054922345027 | 0                          |
| 17       | 0.01                | 0.01         | 0.1 | 0.9938338320298229 | 0.9878011867880767         |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.7** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>SU<br/>Threshold</i> | <i>Gamma</i> | C   | Rata-rata Akurasi         | Rata-rata <i>F-Measure</i> |
|----------|-------------------------|--------------|-----|---------------------------|----------------------------|
| 18       | 0.01                    | 0.01         | 1   | 0.994712449393605         | 0.9887455302144346         |
| 19       | 0.01                    | 0.01         | 10  | 0.994274386476252         | 0.9872872465268646         |
| 20       | 0.01                    | 0.01         | 100 | 0.9933783999863666        | 0.9850560149370893         |
| 21       | 0.01                    | 0.1          | 0.1 | 0.8847988498856539        | 0.7024363518812484         |
| 22       | 0.01                    | 0.1          | 1   | 0.9845808887902923        | 0.9595864667276042         |
| 23       | 0.01                    | 0.1          | 10  | 0.9854856124486625        | 0.9614195758357338         |
| 24       | 0.01                    | 0.1          | 100 | 0.9854856124486625        | 0.9614195758357338         |
| 25       | 0.01                    | 1            | 0.1 | 0.6746054922345027        | 0                          |
| 26       | 0.01                    | 1            | 1   | 0.6746054922345027        | 0                          |
| 27       | 0.01                    | 1            | 10  | 0.6787522053467758        | 0.0186371770930596         |
| 28       | 0.01                    | 1            | 100 | 0.6787522053467758        | 0.0186371770930596         |
| 29       | 0.01                    | 2            | 0.1 | 0.6746054922345027        | 0                          |
| 30       | 0.01                    | 2            | 1   | 0.6746054922345027        | 0                          |
| 31       | 0.01                    | 2            | 10  | 0.6746054922345027        | 0                          |
| 32       | 0.01                    | 2            | 100 | 0.6746054922345027        | 0                          |
| 33       | 0.05                    | 0.01         | 0.1 | 0.9951292698661105        | 0.9901127686028703         |
| 34       | 0.05                    | 0.01         | 1   | <b>0.9960170859468753</b> | <b>0.9918217374994484</b>  |
| 35       | 0.05                    | 0.01         | 10  | 0.9947090410605639        | 0.9893578595336493         |
| 36       | 0.05                    | 0.01         | 100 | 0.992499015336637         | 0.9849568636563177         |
| 37       | 0.05                    | 0.1          | 0.1 | 0.9899762106146042        | 0.9789687708912357         |
| 38       | 0.05                    | 0.1          | 1   | 0.9934217619703707        | 0.985379590549992          |
| 39       | 0.05                    | 0.1          | 10  | 0.992991332940346         | 0.9846112945686611         |
| 40       | 0.05                    | 0.1          | 100 | 0.992991332940346         | 0.9846112945686611         |
| 41       | 0.05                    | 1            | 0.1 | 0.6746054922345027        | 0.6797237037437369         |
| 42       | 0.05                    | 1            | 1   | 0.9113471844534798        | 0.8195801808866717         |
| 43       | 0.05                    | 1            | 10  | 0.9143851172951191        | 0.8259433634056048         |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.7** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>SU Threshold</i> | <i>Gamma</i> | C   | Rata-rata Akurasi  | Rata-rata <i>F-Measure</i> |
|----------|---------------------|--------------|-----|--------------------|----------------------------|
| 44       | 0.05                | 1            | 100 | 0.9143851172951191 | 0.8259433634056048         |
| 45       | 0.05                | 2            | 0.1 | 0.6766541829553208 | 0.0066993464052288         |
| 46       | 0.05                | 2            | 1   | 0.7658476824954976 | 0.3675786030054125         |
| 47       | 0.05                | 2            | 10  | 0.7917875483342105 | 0.4453569501848482         |
| 48       | 0.05                | 2            | 100 | 0.7917875483342105 | 0.4453569501848482         |
| 49       | 0.1                 | 0.01         | 0.1 | 0.9955677384178767 | 0.9913175878799787         |
| 50       | 0.1                 | 0.01         | 1   | 0.9955729692985603 | 0.9913277924822544         |
| 51       | 0.1                 | 0.01         | 10  | 0.9951450645754797 | 0.9905764604667697         |
| 52       | 0.1                 | 0.01         | 100 | 0.9933773634620519 | 0.9864908880687862         |
| 53       | 0.1                 | 0.1          | 0.1 | 0.9916934240432653 | 0.9835524774876051         |
| 54       | 0.1                 | 0.1          | 1   | 0.9955643601786207 | 0.9908702946374142         |
| 55       | 0.1                 | 0.1          | 10  | 0.9947037882634719 | 0.9888599879628955         |
| 56       | 0.1                 | 0.1          | 100 | 0.994701869051539  | 0.9889381696795428         |
| 57       | 0.1                 | 1            | 0.1 | 0.6943730081734881 | 0.1028295266468057         |
| 58       | 0.1                 | 1            | 1   | 0.9655324498607276 | 0.9322358070155401         |
| 59       | 0.1                 | 1            | 10  | 0.9663792721726002 | 0.9335983374291654         |
| 60       | 0.1                 | 1            | 100 | 0.96595520768012   | 0.9329120629193616         |
| 61       | 0.1                 | 2            | 0.1 | 0.67861679638942   | 0.0120296767355591         |
| 62       | 0.1                 | 2            | 1   | 0.8214659901411123 | 0.5426549112772048         |
| 63       | 0.1                 | 2            | 10  | 0.8214659901411123 | 0.5426549112772048         |
| 64       | 0.1                 | 2            | 100 | 0.83847281589742   | 0.6009109409618059         |

Berikut adalah skenario pengujian tanpa menggunakan seleksi fitur *Symmetrical Uncertainty* dan rata-rata hasil pengujian berupa akurasi dan *F-Measure* yang didapat:

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

**Tabel 4.8** Hasil Pengujian *Support Vector Machine* tanpa *Symmetrical Uncertainty*

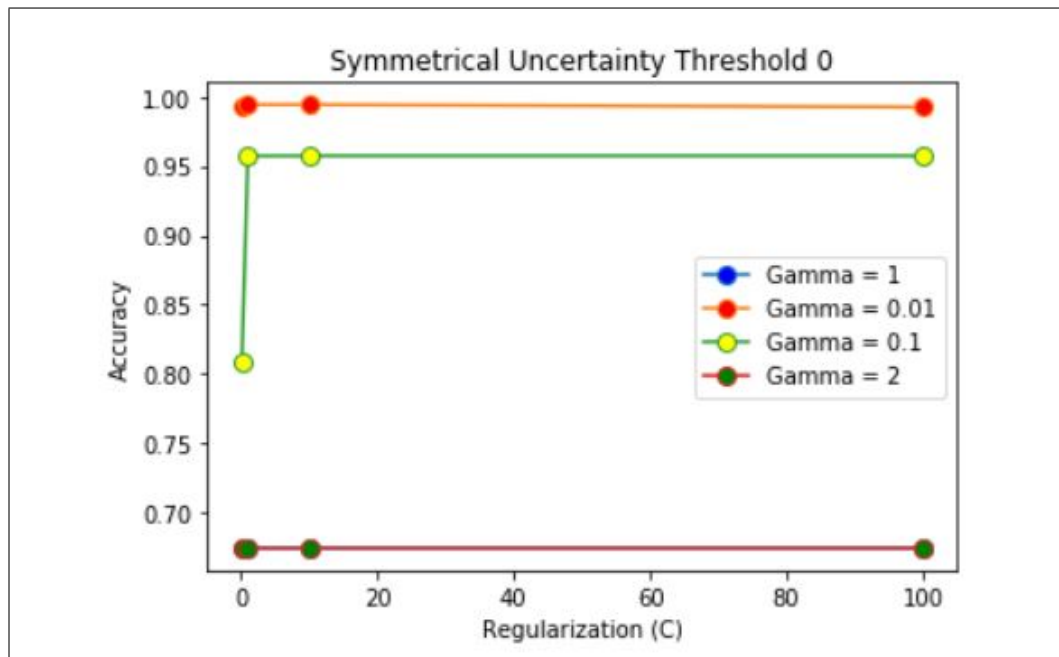
| Skenario | <i>Gamma</i> | C   | Rata-rata Akurasi         | Rata-rata <i>F-Measure</i> |
|----------|--------------|-----|---------------------------|----------------------------|
| 1        | 0.01         | 0.1 | 0.8182266713323023        | 0.54625883442953615        |
| 2        | 0.01         | 1   | 0.8566255835503623        | 0.6935626976353527         |
| 3        | 0.01         | 10  | 0.9144910876267849        | <b>0.8161320675776474</b>  |
| 4        | 0.01         | 100 | <b>0.9148522397525713</b> | 0.8154731290568162         |
| 5        | 0.1          | 0.1 | 0.7840730021554603        | 0.4169949066398014         |
| 6        | 0.1          | 1   | 0.6837642294628422        | 0.05059105927284646        |
| 7        | 0.1          | 10  | 0.8517284885816026        | 0.6408301397979068         |
| 8        | 0.1          | 100 | 0.8637861231370337        | 0.6792332155540427         |
| 9        | 1            | 0.1 | 0.6797756247649048        | 0.21840608465608464        |
| 10       | 1            | 1   | 0.7553541916356184        | 0.3175844749458328         |
| 11       | 1            | 10  | 0.7630881562517828        | 0.34118342677819182        |
| 12       | 1            | 100 | 0.7630881562517828        | 0.34118342677819182        |
| 13       | 2            | 0.1 | 0.6746054922345026        | 0.0                        |
| 14       | 2            | 1   | 0.707506607426099         | 0.13267349637937875        |
| 15       | 2            | 10  | 0.7129693350296948        | 0.15529401358602787        |
| 16       | 2            | 100 | 0.7129693350296948        | 0.15529401358602787        |

##### 4.4.2.2 Analisis Hasil Pengujian Klasifikasi *Support Vector Machine* dengan *Symmetrical Uncertainty*

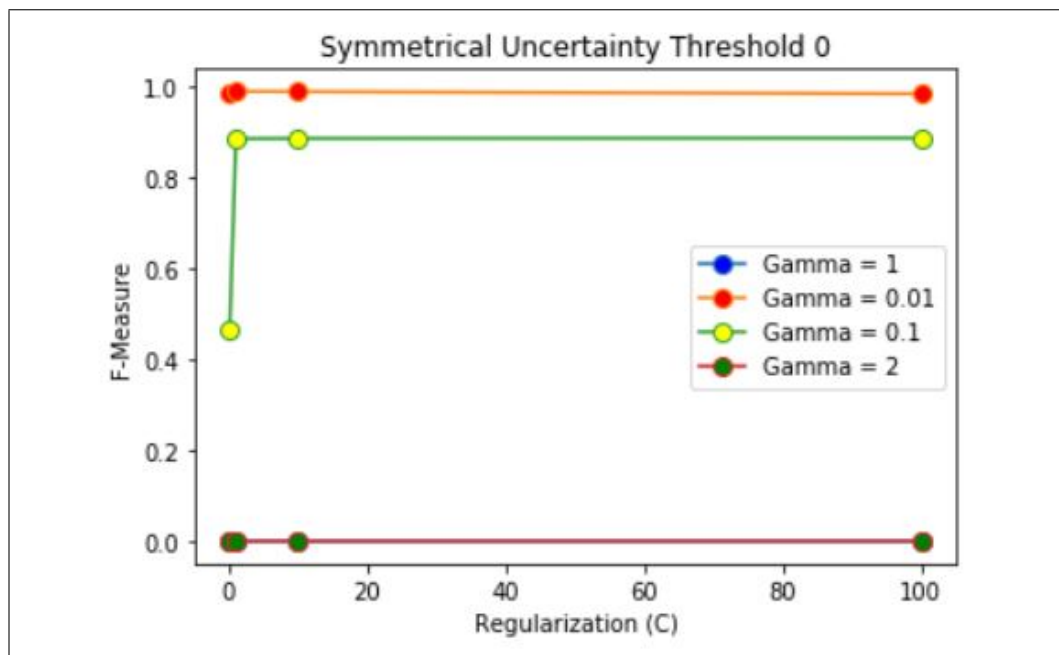
Pada bagian analisis hasil pengujian klasifikasi *Support Vector Machine*, akan dilakukan perbandingan hasil akurasi dan *F-Measure* yang didapat dengan menggunakan nilai *threshold Symmetrical Uncertainty*, regularisasi (C), dan *gamma* ( $\gamma$ ) yang berbeda.



#### 4. IMPLEMENTASI DAN PENGUJIAN



**Gambar 4.1** Hasil Akurasi dengan Nilai *Threshold Symmetrical Uncertainty* sebesar 0

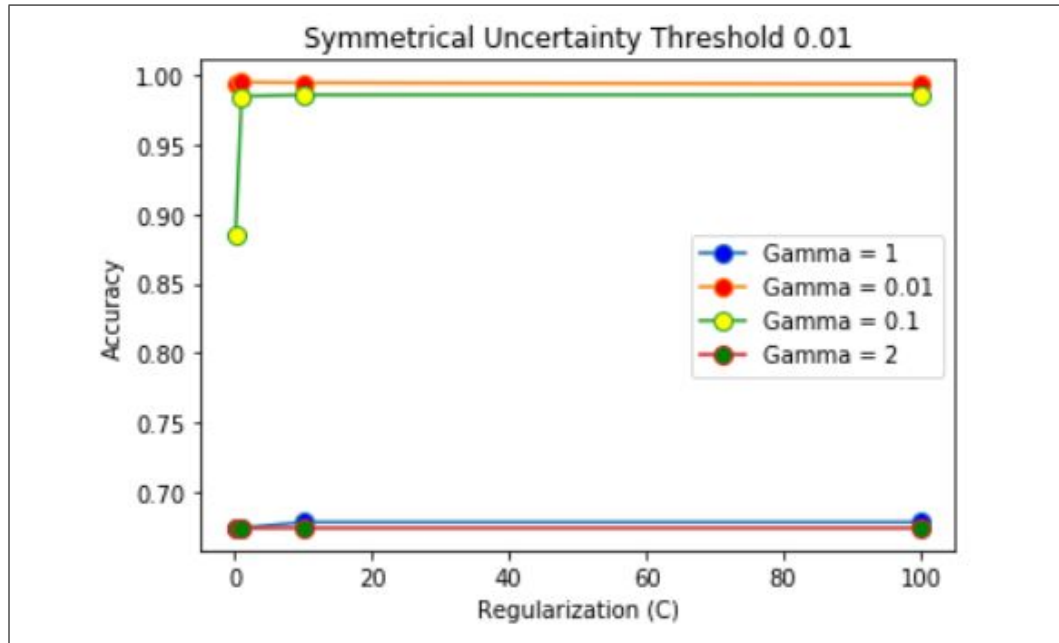


**Gambar 4.2** Hasil *F-Measure* dengan *Threshold Symmetrical Uncertainty* sebesar 0

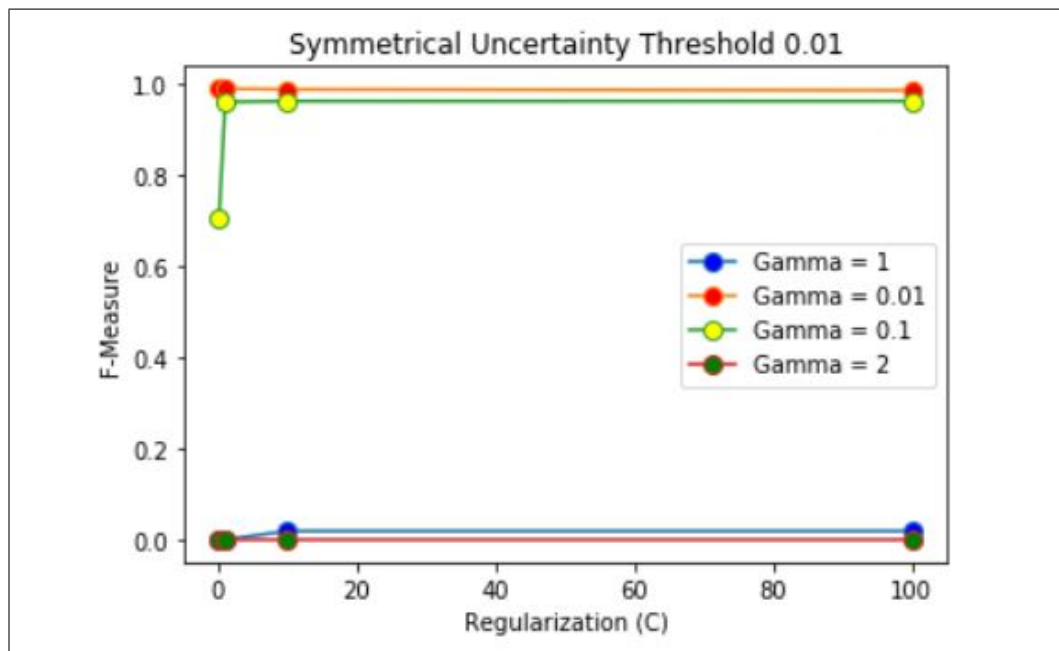
Dari grafik digambar 4.1 dan 4.2, *threshold Symmetrical Uncertainty* bernilai 0 yang memiliki jumlah fitur 25 menghasilkan akurasi dan *F-Measure* terbaik dengan nilai *gamma* sebesar 0.01 dan nilai regularisasi terbaik sebesar 1. Untuk nilai regularisasi, nilai sebesar 1, 10, dan 100 memiliki hasil akurasi dan

#### 4. IMPLEMENTASI DAN PENGUJIAN

*F-Measure* yang hanya berbeda sedikit yaitu sekitar 0.00005%. Akurasi tertinggi sebesar 99.43% dan *F-Measure* tertinggi sebesar 98.75%.



**Gambar 4.3** Hasil Akurasi dengan *Threshold Symmetrical Uncertainty* sebesar 0.01

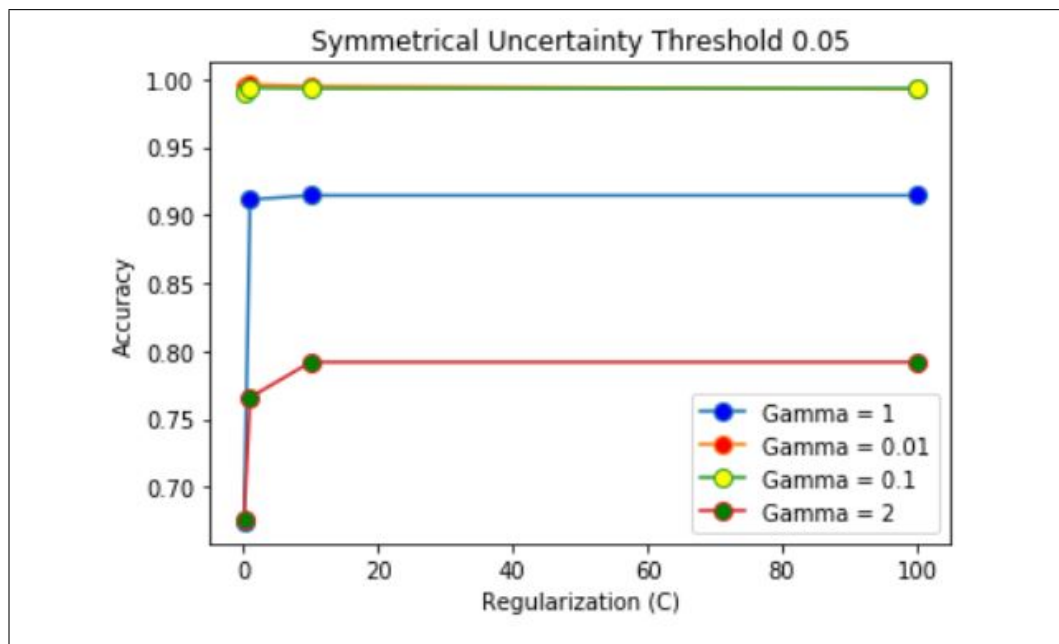


**Gambar 4.4** Hasil *F-Measure* dengan *Threshold Symmetrical Uncertainty* sebesar 0.01

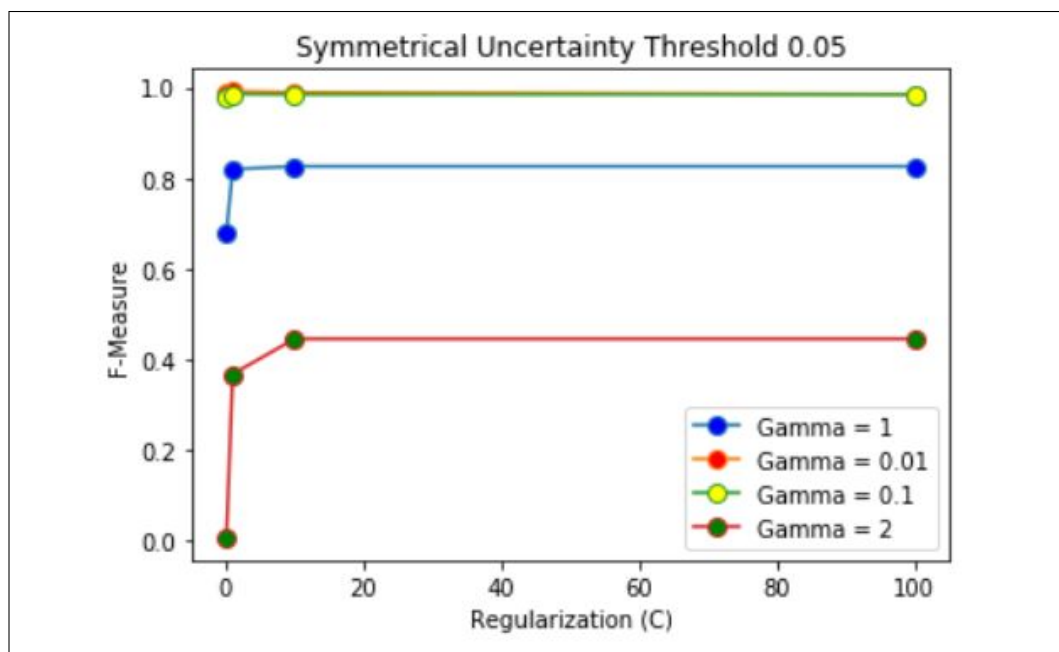
Dari grafik digambar 4.3 dan 4.4, *threshold Symmetrical Uncertainty* bernilai 0.01 dengan jumlah fitur sebesar 16 menghasilkan akurasi dan *F-Measure* terbaik

#### 4. IMPLEMENTASI DAN PENGUJIAN

dengan nilai  $\gamma$  sebesar 0.01 dan nilai regularisasi sebesar 1. Nilai regularisasi 1, 10, dan 100 menghasilkan akurasi dan  $F$ -Measure berbeda sedikit, sekitar 0.005%. Akurasi tertinggi sebesar 99.47% dan  $F$ -Measure tertinggi sebesar 98.87%.



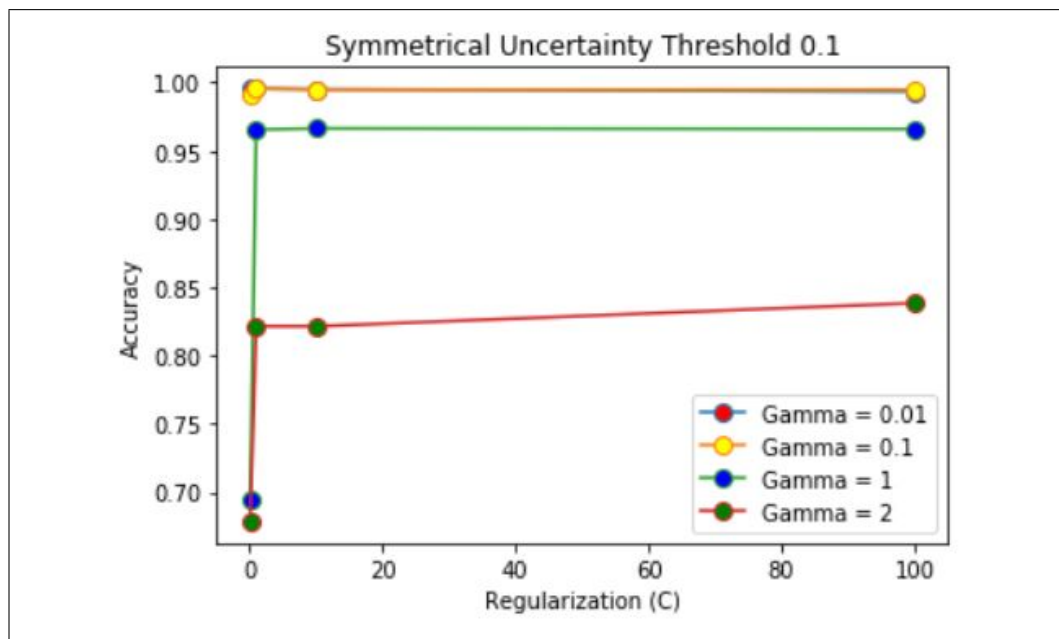
**Gambar 4.5** Hasil Akurasi dengan *Threshold Symmetrical Uncertainty* sebesar 0.05



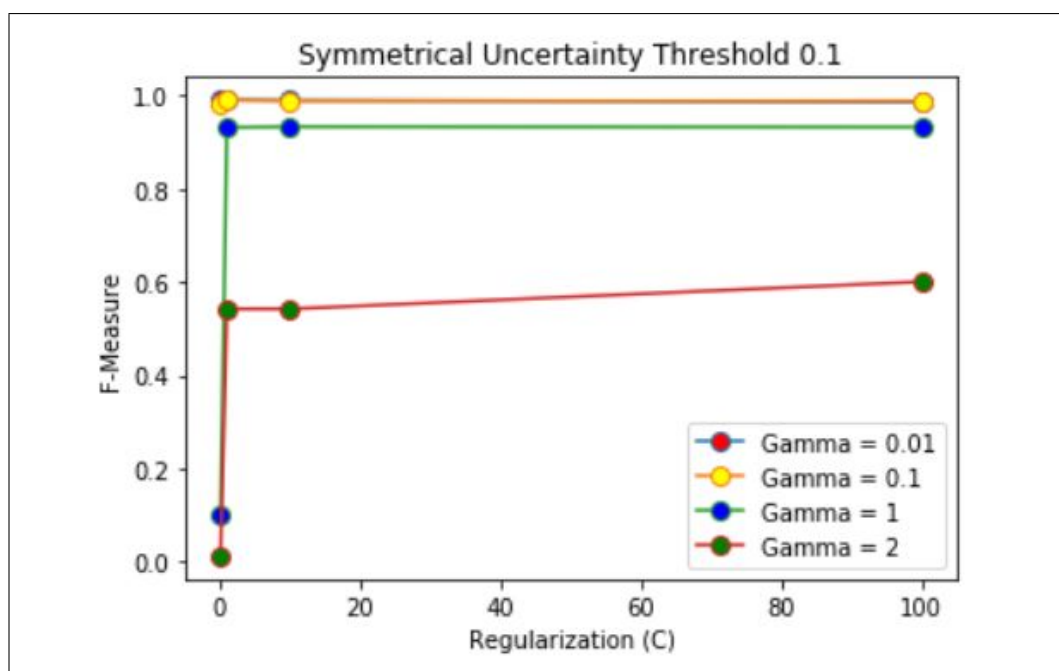
**Gambar 4.6** Hasil  $F$ -Measure dengan *Threshold Symmetrical Uncertainty* sebesar 0.05

#### 4. IMPLEMENTASI DAN PENGUJIAN

Dari grafik digambar 4.5 dan 4.6, *threshold Symmetrical Uncertainty* bernilai 0.05 dengan jumlah fitur sebesar 6 menghasilkan akurasi dan *F-Measure* terbaik dengan nilai *gamma* sebesar 0.01 dan nilai regularisasi sebesar 1. Akurasi tertinggi sebesar 99.601% dan *F-Measure* tertinggi sebesar 99.182%.



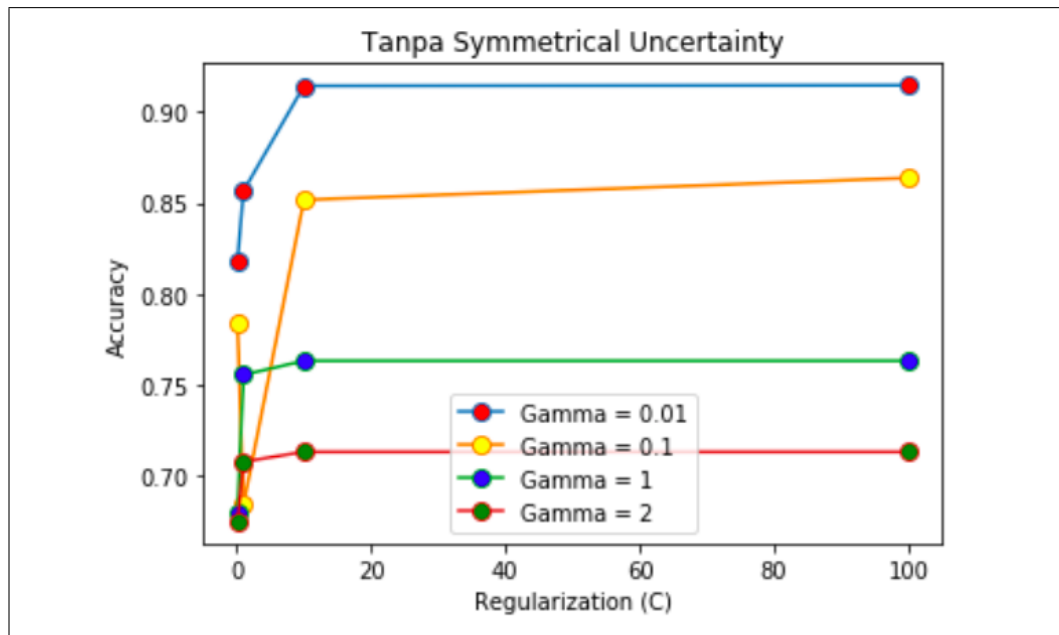
**Gambar 4.7** Hasil Akurasi dengan *Threshold Symmetrical Uncertainty* sebesar 0.1



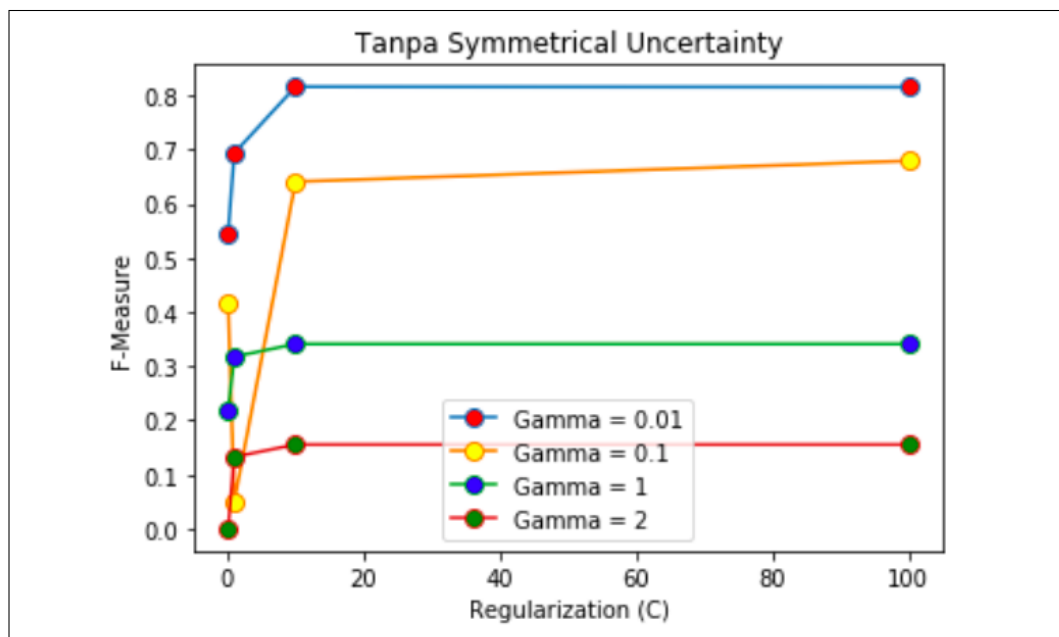
**Gambar 4.8** Hasil *F-Measure* dengan *Threshold Symmetrical Uncertainty* sebesar 0.1

#### 4. IMPLEMENTASI DAN PENGUJIAN

Dari grafik digambar 4.7 dan 4.8, *threshold Symmetrical Uncertainty* bernilai 0.1 dengan jumlah fitur sebesar 5 menghasilkan akurasi dan *F-Measure* terbaik dengan nilai *gamma* sebesar 0.01 dan nilai regularisasi sebesar 1. Akurasi tertinggi sebesar 99.557% dan *F-Measure* tertinggi sebesar 99.133%.



**Gambar 4.9** Hasil Akurasi Tanpa *Symmetrical Uncertainty*

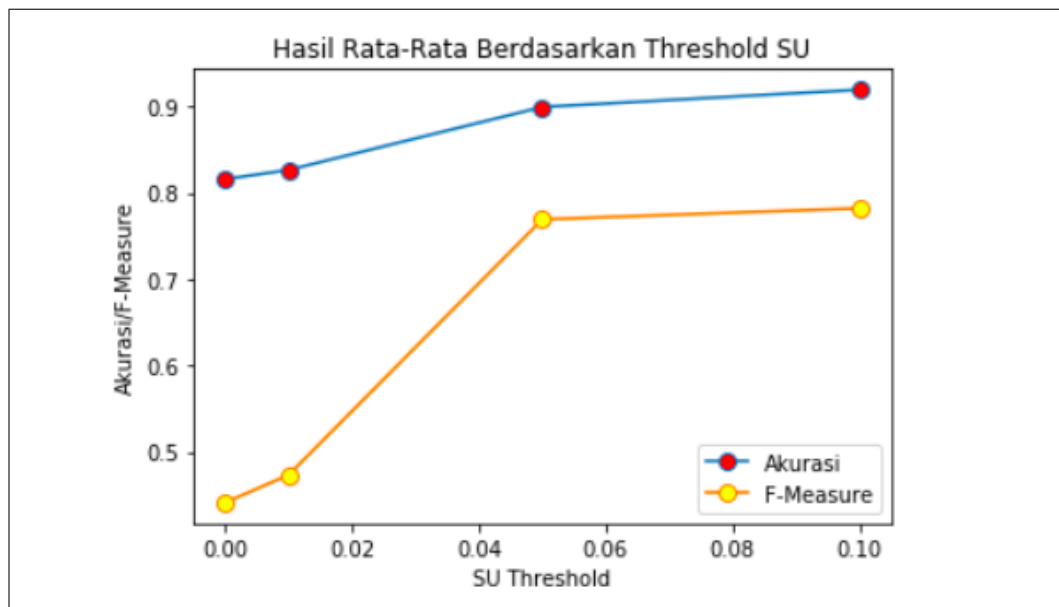


**Gambar 4.10** Hasil *F-Measure* Tanpa *Symmetrical Uncertainty*

#### 4. IMPLEMENTASI DAN PENGUJIAN

Dari grafik digambar 4.9 dan 4.10, pengujian tanpa menggunakan seleksi fitur *Symmetrical Uncertainty* menghasilkan akurasi dan *F-Measure* yang tinggi namun hasilnya tidak lebih tinggi dibandingkan pengujian yang menggunakan seleksi fitur *Symmetrical Uncertainty* dengan nilai *threshold Symmetrical Uncertainty* sebesar 0.1 (5 fitur), *gamma* sebesar 0.01, dan nilai regularisasi sebesar 1. Akurasi dan *F-Measure* yang tinggi dihasilkan dengan nilai *gamma* sebesar 0.01 dan nilai regularisasi sebesar 100 (untuk akurasi) dan 10 (untuk *F-Measure*). Akurasi tertinggi sebesar 91.485% dan *F-Measure* tertinggi sebesar 81.613%.

Berdasarkan pengujian yang telah dilakukan, berikut adalah kesimpulan yang didapat:



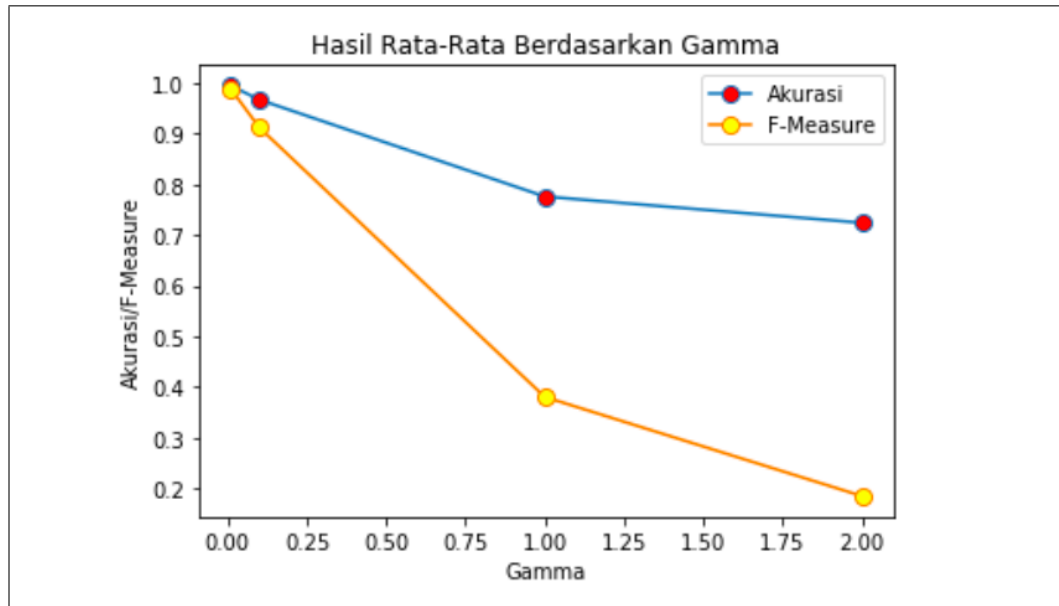
**Gambar 4.11** Hasil Akurasi dan *F-Measure* Berdasarkan *Threshold Symmetrical Uncertainty*

Dari gambar 4.11, dapat disimpulkan bahwa semakin tinggi *threshold Symmetrical Uncertainty*, semakin tinggi akurasi dan *F-Measure* yang didapat. Semakin tinggi *threshold Symmetrical Uncertainty*, maka fitur yang memiliki nilai yang tidak terlalu berpengaruh terhadap prediksi *title box office film* akan semakin banyak dibuang. Ketika menggunakan *threshold Symmetrical Uncertainty* sebesar 0,

#### 4. IMPLEMENTASI DAN PENGUJIAN

---

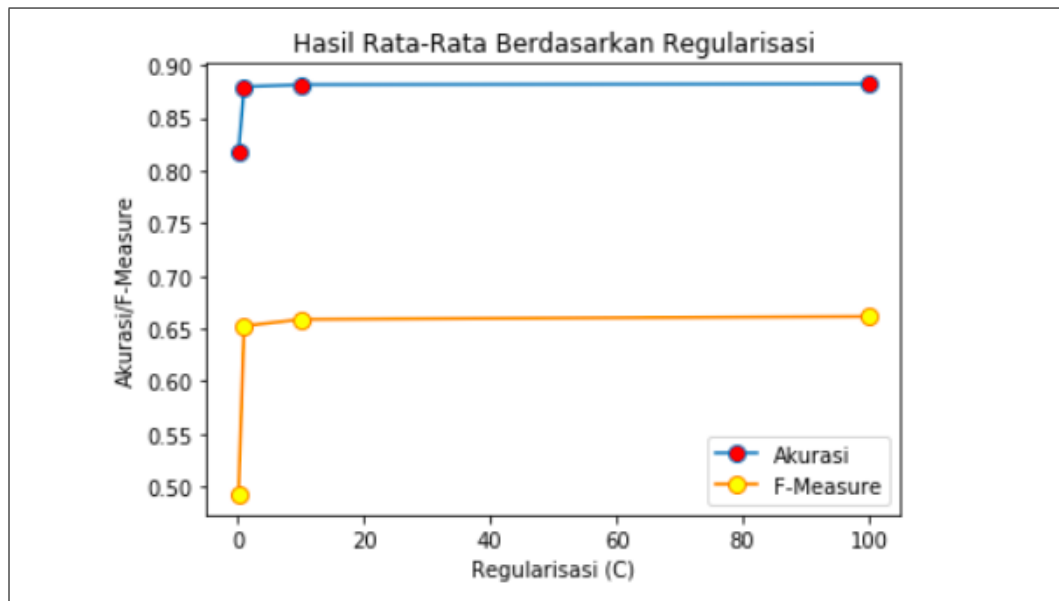
*F-Measure* yang dihasilkan yaitu 44% yang berarti persebaran data tidak cukup bagus jika banyak memakai fitur yang tidak terlalu berpengaruh terhadap label kelas *box office*.



**Gambar 4.12** Hasil Akurasi dan *F-Measure* Berdasarkan *Gamma*

Dari gambar 4.12, semakin besar nilai *gamma* yang diuji, maka semakin kecil akurasi maupun *F-Measure* yang dihasilkan. *Gamma* pada *Support Vector Machine* digunakan sebagai area margin pengelompokkan pada setiap data film. Semakin besar *gamma* yang diuji, area margin tersebut semakin kecil, dan akurasi dan *F-Measure* yang dihasilkan semakin kecil menandakan bahwa persebaran data saling berdekatan satu sama lain sehingga membutuhkan *gamma* yang kecil untuk memisahkan antar data film yang berdekatan.

#### 4. IMPLEMENTASI DAN PENGUJIAN



**Gambar 4.13** Hasil Akurasi dan *F-Measure* Berdasarkan Regularisasi (C)

Dari gambar 4.13, semakin tinggi nilai regularisasi maka akan semakin tinggi akurasi dan *F-Measure* yang dihasilkan. Regularisasi pada *Support Vector Machine* digunakan untuk menentukan batas toleransi atau margin pada *hyperplane Support Vector Machine*. Semakin besar nilai regularisasi yang dipakai, maka margin *hyperplane Support Vector Machine* akan semakin kecil. Dengan memperbesar nilai regularisasi, dapat mengurangi kemungkinan data salah prediksi. Akurasi dan *F-Measure* yang dihasilkan dengan nilai regularisasi sebesar 0.1 dan 1 berbeda jauh, sedangkan nilai regularisasi 10 dan 100 tidak berbeda jauh. Maka dari itu, dapat dikatakan bahwa jika menggunakan regularisasi dibawah 1, masih banyak data label positif dan negatif tidak terpisah secara bagus, sedangkan nilai regularisasi diatas 10 menandakan bahwa data film yang memiliki label positif dan negatif sudah terpisah cukup jauh sehingga hanya menaikkan akurasi dan *F-Measure* dengan jumlah kecil yaitu sekitar 0.001% untuk akurasi dan 0.01% untuk *F-Measure*.



### 4.4.3 Pengujian Klasifikasi *Support Vector Machine* (Library)

Pada bagian ini akan dilakukan pengujian menggunakan *library Support Vector Machine* dengan *threshold Symmetrical Uncertainty*, *gamma* dan regularisasi tertinggi dan terbaik yang dihasilkan pada pengujian *Support Vector Machine* tanpa *library*. Setelah itu akan dilakukan analisis dan perbandingan hasil pengujian klasifikasi menggunakan *Support Vector Machine* tanpa *library* dan dengan *library*.

#### 4.4.3.1 Skenario dan Hasil Pengujian Klasifikasi *Support Vector Machine* (Library)

Skenario klasifikasi menggunakan *library Support Vector Machine* dilakukan dua bagian, pertama menggunakan seleksi fitur *Symmetrical Uncertainty* dan kedua tanpa menggunakan seleksi fitur *Symmetrical Uncertainty*. Nilai *threshold Symmetrical Uncertainty*, *gamma* dan regularisasi menggunakan nilai yang menghasilkan akurasi dan *F-Measure* tertinggi dan terbaik, yang didapatkan pada pengujian klasifikasi *Support Vector Machine* sebelumnya (tanpa *library*).

Berikut adalah skenario dan hasil pengujian menggunakan *library Support Vector Machine* dan rata-rata hasil pengujian berupa akurasi dan *F-Measure* yang didapat:

**Tabel 4.9** Hasil Pengujian Pengaruh Jumlah Data dengan *Library Support Vector Machine*

| Skenario | Fold | SU<br>Threshold | Gamma | C | Akurasi            | F-Measure          |
|----------|------|-----------------|-------|---|--------------------|--------------------|
| 1        | 1    | 0.05            | 0.01  | 1 | 0.9956153144823378 | 0.9879518072289156 |
|          | 2    |                 |       |   | 0.9866481550327408 | 0.970873786407767  |
|          | 3    |                 |       |   | 0.9911488468022976 | 0.9846153846153847 |
|          | 4    |                 |       |   | 1.0                | 1.0                |
|          | 5    |                 |       |   | 1.0                | 1.0                |
|          | 6    |                 |       |   | 0.9911545537994092 | 0.9857142857142858 |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.9** Hasil Pengujian Pengaruh Jumlah Data dengan *Library Support Vector Machine* (Lanjutan)

| Skenario  | <i>Fold</i> | <i>SU Threshold</i> | <i>Gamma</i> | <i>C</i>           | Akurasi            | <i>F-Measure</i>   |
|-----------|-------------|---------------------|--------------|--------------------|--------------------|--------------------|
|           | 7           |                     |              |                    | 0.9911222610521826 | 0.9791666666666666 |
|           | 8           |                     |              |                    | 1.0                | 1.0                |
|           | 9           |                     |              |                    | 1.0                | 1.0                |
|           | 10          |                     |              |                    | 1.0                | 1.0                |
|           | Rata-rata   |                     |              |                    | 0.9955689131168967 | 0.9908321930633019 |
| 2         | 1           | 0.1                 | 0.01         | 1                  | 1.0                | 1.0                |
|           | 2           |                     |              |                    | 0.9821334621153227 | 0.9607843137254902 |
|           | 3           |                     |              |                    | 0.9911488468022976 | 0.9846153846153847 |
|           | 4           |                     |              |                    | 1.0                | 1.0                |
|           | 5           |                     |              |                    | 1.0                | 1.0                |
|           | 6           |                     |              |                    | 0.9911545537994092 | 0.9857142857142858 |
|           | 7           |                     |              |                    | 0.9911222610521826 | 0.9791666666666666 |
|           | 8           |                     |              |                    | 1.0                | 1.0                |
|           | 9           |                     |              |                    | 1.0                | 1.0                |
|           | 10          |                     |              |                    | 1.0                | 1.0                |
| Rata-rata |             |                     |              | 0.9955559123769211 | 0.9910280650721827 |                    |

**Tabel 4.10** Hasil Pengujian dengan *Library SVM* tanpa SU

| Skenario | <i>Fold</i> | <i>Gamma</i> | <i>C</i> | Akurasi            | <i>F-Measure</i>    |
|----------|-------------|--------------|----------|--------------------|---------------------|
| 1        | 1           | 0.01         | 100      | 0.7844773508076696 | 0.46938775510204084 |
|          | 2           |              |          | 0.8007438849474043 | 0.5376344086021505  |
|          | 3           |              |          | 0.880223180159662  | 0.7906976744186046  |
|          | 4           |              |          | 0.9184599564774607 | 0.8448275862068966  |
|          | 5           |              |          | 0.9289039384574922 | 0.8805970149253731  |
|          | 6           |              |          | 0.9232411155338999 | 0.8702290076335878  |
|          | 7           |              |          | 0.9733667831565477 | 0.9375000000000001  |
|          | 8           |              |          | 0.9688157019472488 | 0.9278350515463918  |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.10** Hasil Pengujian dengan *Library* SVM tanpa SU (Lanjutan)

| Skenario  | <i>Fold</i> | <i>Gamma</i> | C  | Akurasi            | <i>F-Measure</i>    |
|-----------|-------------|--------------|----|--------------------|---------------------|
|           | 9           |              |    | 0.9866926283258093 | 0.9508196721311476  |
|           | 10          |              |    | 0.9910194876648226 | 0.9565217391304348  |
| Rata-rata |             |              |    | 0.9155944027478017 | 0.8166049909696628  |
| 2         | 1           | 0.01         | 10 | 0.7866346931961221 | 0.4631578947368421  |
|           | 2           |              |    | 0.8378147900734686 | 0.6236559139784946  |
|           | 3           |              |    | 0.8707085744338943 | 0.7716535433070867  |
|           | 4           |              |    | 0.9045654631652634 | 0.8173913043478261  |
|           | 5           |              |    | 0.9469140993609646 | 0.9117647058823529  |
|           | 6           |              |    | 0.914210658537888  | 0.8549618320610686  |
|           | 7           |              |    | 0.9777178149624733 | 0.9473684210526316  |
|           | 8           |              |    | 0.9556213615476383 | 0.8979591836734694  |
|           | 9           |              |    | 0.9866926283258093 | 0.9508196721311476  |
|           | 10          |              |    | 0.9910194876648226 | 0.9565217391304348  |
| Rata-rata |             |              |    | 0.9171899571268344 | 0.8195254210301355  |
| 3         | 1           | 0.01         | 1  | 0.7033017247816024 | 0.28                |
|           | 2           |              |    | 0.7785941056163194 | 0.43589743589743585 |
|           | 3           |              |    | 0.7468392330844293 | 0.4680851063829786  |
|           | 4           |              |    | 0.8073970082780656 | 0.5777777777777777  |
|           | 5           |              |    | 0.8573286578893308 | 0.743801652892562   |
|           | 6           |              |    | 0.8408949142564175 | 0.7027027027027027  |
|           | 7           |              |    | 0.9136923131286956 | 0.775               |
|           | 8           |              |    | 0.9357468278173123 | 0.8444444444444444  |
|           | 9           |              |    | 0.9731948388828014 | 0.9                 |
|           | 10          |              |    | 0.9816638502788713 | 0.9090909090909091  |
| Rata-rata |             |              |    | 0.8538653474013847 | 0.66368000291888105 |

### 4.4.4 Pengujian Pengaruh Jumlah Data Terhadap Klasifikasi *Support Vector Machine*

Pada bagian ini akan dilakukan pengujian pengaruh jumlah data terhadap prediksi *title box office*. Metode klasifikasi menggunakan *Support Vector Machine* dengan *threshold Symmetrical Uncertainty*, *gamma* dan regularisasi terbaik yang dihasilkan pada pengujian *Support Vector Machine*. Setelah itu akan dilakukan analisis dan perbandingan hasil pengujian klasifikasi menggunakan *Support Vector Machine* dengan jumlah data film yang beragam.

#### 4.4.4.1 Skenario dan Hasil Pengujian Pengaruh Jumlah Data Terhadap Klasifikasi *Support Vector Machine*

Skenario klasifikasi menggunakan *Support Vector Machine* dengan jumlah data film 5044, 4544, 4044, 3544, 3044, 2544, 2044, dan 1544. Nilai *threshold Symmetrical Uncertainty*, *gamma* dan regularisasi menggunakan nilai yang menghasilkan akurasi dan *F-Measure* terbaik, yang didapatkan pada pengujian klasifikasi *Support Vector Machine* sebelumnya (tanpa *library*).

Berikut adalah skenario dan hasil pengujian jumlah data film dan rata-rata hasil pengujian berupa akurasi dan *F-Measure* yang didapat:

**Tabel 4.11** Hasil Pengujian *Support Vector Machine* dengan Jumlah Data Film yang Berbeda

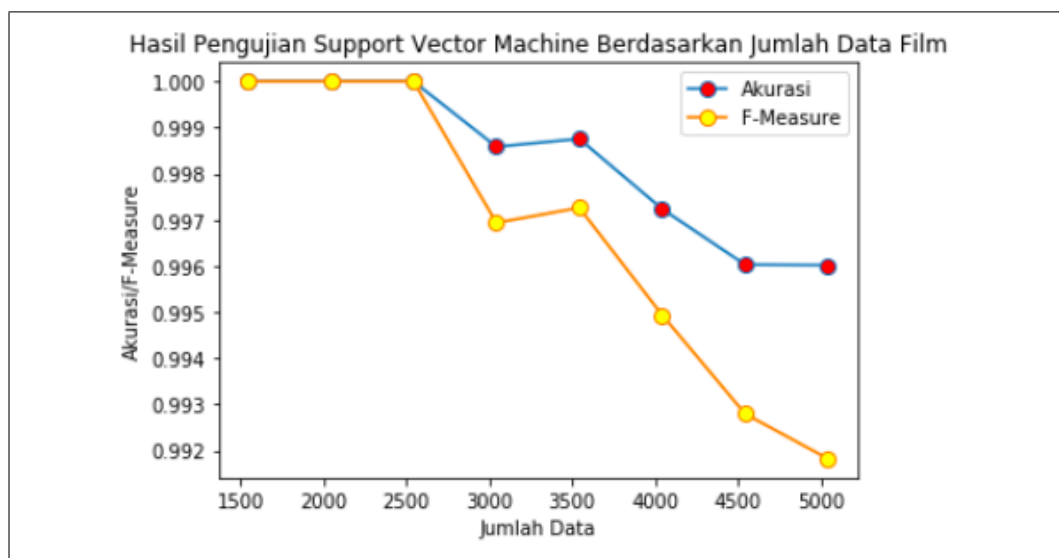
| Jumlah Data Film | <i>SU Threshold</i> | <i>Gamma</i> | C | Rata-rata Akurasi  | Rata-rata <i>F-Measure</i> |
|------------------|---------------------|--------------|---|--------------------|----------------------------|
| 5044             | 0.1                 | 0.01         | 1 | 0.9960170859468753 | 0.9918217374994484         |
| 4544             |                     |              |   | 0.996027952667215  | 0.992795779876801          |
| 4044             |                     |              |   | 0.9972352479804357 | 0.9949366761315556         |
| 3544             |                     |              |   | 0.9987555927059628 | 0.9972602739726028         |
| 3044             |                     |              |   | 0.9985840568660376 | 0.9969367369589346         |

#### 4. IMPLEMENTASI DAN PENGUJIAN

**Tabel 4.11** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| <i>Fold</i> | <i>SU Threshold</i> | <i>Gamma</i> | <i>C</i> | Rataasdasdsa | Rata-rata <i>F-Measure</i> |
|-------------|---------------------|--------------|----------|--------------|----------------------------|
| 2544        |                     |              |          | 1.0          | 1.0                        |
| 2044        |                     |              |          | 1.0          | 1.0                        |
| 1544        |                     |              |          | 1.0          | 1.0                        |

##### 4.4.4.2 Analisis Hasil Pengujian Pengaruh Jumlah Data Terhadap Klasifikasi *Support Vector Machine*



**Gambar 4.14** Hasil Klasifikasi dengan Jumlah Data yang Berbeda

Dari grafik digambar 4.14, dilakukan pengujian sebanyak 8 skenario dengan jumlah data film yang berbeda-beda yaitu 5044, 4544, 4044, 3544, 3044, 2544, 2044, dan 1544 data film. Jumlah data film berpengaruh terhadap pembuatan model prediksi *Support Vector Machine* menggunakan kernel RBF dikarenakan  $\alpha$  yang dihasilkan bergantung terhadap jumlah data yang dipakai. Semakin sedikit jumlah data yang dipakai,  $\alpha$  yang dihasilkan akan semakin sedikit. Dapat disimpulkan bahwa semakin banyak jumlah data film yang dipakai untuk klasifikasi, maka hasil akurasi dan *F-Measure* akan semakin tinggi. Akurasi dan *F-Measure* tertinggi yaitu 100% pada jumlah data film 1544, 2044, dan 2544.

## BAB 5 PENUTUP

Bab ini berisi kesimpulan yang dilandasi oleh penelitian dan pengujian yang telah dilakukan, serta dilengkapi dengan saran yang dapat untuk perkembangan ke depan.

### 5.1 Kesimpulan

Kesimpulan dari sistem prediksi *title box office* film adalah:

1. Seleksi fitur *Symmetrical Uncertainty* sangat mempengaruhi akurasi dan *F-Measure* dalam memprediksi *title box office* film. Dari tabel pengujian 4.7 dengan menggunakan seleksi fitur *Symmetrical Uncertainty*, didapatkan akurasi tertinggi sebesar 99.6% dan *F-Measure* tertinggi sebesar 99.1%. Sedangkan dari tabel pengujian 4.8 tanpa menggunakan seleksi fitur *Symmetrical Uncertainty*, akurasi tertinggi yang didapat hanya 91.48% dan *F-Measure* tertinggi hanya 81.61%.
2. *Threshold Symmetrical Uncertainty* terbaik dengan nilai yang diuji yaitu 0 (25 fitur), 0.01 (16 fitur), 0.05 (15 fitur), dan 0.1 (5 fitur) adalah 0.1 dengan fitur yang dipakai berjumlah 5. Hasil akurasi terbaik sebesar 91.92% dan *F-Measure* terbaik sebesar 78.19%. Dalam memprediksi *title box office film*, hanya membutuhkan 5 fitur yang paling berpengaruh terhadap label kelas *box office* sudah memberikan hasil akurasi dan *F-Measure* yang cukup tinggi.
3. Nilai *gamma* ( $\gamma$ ) terbaik untuk prediksi *title box office* film menggunakan metode *Support Vector Machine* dengan nilai yang diuji yaitu 0.01, 0.1, 1, dan 2, adalah 0.01. Hasil akurasi terbaik sebesar 99.4% dan *F-Measure* terbesar sebesar 98.78%. Semakin kecil *gamma* maka akurasi dan *F-Measure* semakin tinggi dan dapat disimpulkan bahwa persebaran setiap data film banyak yang berhimpitan.

4. Nilai regularisasi (C) terbaik untuk perhitungan *Support Vector Machine* dengan nilai yang diuji yaitu 0.1, 1, 10, dan 100 adalah 100. Hasil akurasi terbaik sebesar 88.22% dan *F-Measure* terbaik sebesar 66.14%. Hasil akurasi dan *F-Measure* dengan nilai regularisasi 1, 10, dan 100 sedikit berbeda, maka dapat disimpulkan bahwa persebaran data film sudah terpisah cukup baik mulai dari nilai regularisasi (C) sebesar 1.
5. Semakin sedikit jumlah data film yang dipakai, akan meningkatkan akurasi dan *F-Measure*, namun penambahan nilainya tidak signifikan.

### 5.2 Saran

Saran untuk pengembangan sistem prediksi *title box office* film kedepannya adalah:

1. Memakai dataset yang memiliki jumlah fitur yang lebih banyak untuk mengetahui efektifitas dari seleksi fitur *Symmetrical Uncertainty* dalam memprediksi *title box office*.
2. Menggunakan teknik *clustering* untuk menganalisis karakteristik fitur film dan menemukan *underlying structure* yang ada sehingga bisa digunakan untuk penelitian selanjutnya.
3. Membandingkan seleksi fitur *Symmetrical Uncertainty* yang merupakan teknik *filter* dengan *Fast Correlation Based Filter* yang merupakan teknik *wrapper* dari *Symmetrical Uncertainty* [14].

## DAFTAR REFERENSI

- [1] The Numbers - Movie Market Summary 1995 to 2019.  
Available:<https://www.the-numbers.com/market/>.
- [2] "Hollywood accounting", *Wikipedia*. [Online].  
Available:[https://en.wikipedia.org/wiki/Hollywood\\_accounting](https://en.wikipedia.org/wiki/Hollywood_accounting). [Date of Access:28-Sep-2019].
- [3] S. Mehta<sup>1</sup>, H. Bhatt, and Prof. D. Desai, "A Compendium for Prediction of Success of a Movie Based Upon Different Factors", *International Journal of Advanced Research in Computer and Communication Engineering*, Volume 4 Issue 12, December 2015.
- [4] K. Meenakshi, et al., "A Data Mining Technique for Analyzing and Predicting the Success of Movie", *IOP Conf. Series: Journal of Physics*, Conf. Series 1000, 2018.
- [5] L. Liu and Y. Zhao, "Research of Box-Office Prediction based on Rough Set and Support Vector Machine", *International Journal of Hybrid Information Technology*, Volume 9 Number 2 pp.417-426, 2016.
- [6] C. S. Kumar and R. J. Rama Sree, "Application of Ranking Based Attribute Selection Filters to Perform Automated Evaluation of Descriptive Answers Through Sequential Minimal Optimization Models", *ICTACT Journal on Soft Computing*, Volume 5 Issue 1, October 2014.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques - 3rd Edition*, 2012.
- [8] S. Ali, and K. A. Smith, "On learning algorithm selection for classification", *Applied Soft Computing*, Volume 6, pp. 119–138, 2006.



- [9] O. A. Taiwo, *Types of Machine Learning Algorithms, New Advances in Machine Learning*, InTech, University of Portsmouth United Kingdom. pp 3 – 31. Available: <http://www.intechopen.com/books/new-advances-in-machine-learning/types-of-machine-learning-algorithms>.
- [10] J. Nayak, B. Naik, and H. S. Behera, "Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications Challenges", *International Journal of Database Theory and Application*, Volume 8 No. 1, pp. 169-186, 2015.
- [11] A. Kowalczyk, *Support Vector Machine Succinctly*, 2017.
- [12] H. B. Moss, D. S. Leslie, and P. Rayson, "Using J-K-fold Cross Validation to Reduce Variance When Tuning NLP Models", *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 2978–2989, August 2018.
- [13] F. H. Zulkernine, "Predicting Movie Box Office Profitability: A Neural Network Approach", *IEEE International Conference on Machine Learning and Applications (ICMLA)*, DOI: 10.1109/ICMLA.2016.0117, December 2016.
- [14] P. Pramkochon and P. Piamsa-nga, "An Unsupervised, Fast Correlation-Based Filter for Feature Selection for Data Clustering", *Advanced Data and Information Engineering (DaEng)*, pp. 87-94, 2014.

# LAMPIRAN

## LAMPIRAN A PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* dengan Seleksi Fitur

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
| 1        | 1           | 0.05            | 0.1   | 0.1 | 0.97845646790932 | 0.94252873563219 |
|          | 2           |                 |       |     | 0.9609564098461  | 0.91891891891892 |
|          | 3           |                 |       |     | 0.98681323919953 | 0.97744360902256 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.99122430082614 | 0.98571428571429 |
|          | 6           |                 |       |     | 0.99118942731278 | 0.98591549295775 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 2        | 1           | 0.05            | 0.1   | 1   | 0.97845646790932 | 0.94252873563218 |
|          | 2           |                 |       |     | 0.98225976666505 | 0.96153846153846 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 3        | 1           | 0.05            | 0.1   | 10  | 0.97845646790932 | 0.94252873563218 |
|          | 2           |                 |       |     | 0.98674014220257 | 0.97142857142857 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 5           |                 |       |     | 1.0              | 1.0               |
|          | 6           |                 |       |     | 0.99118942731278 | 0.98591549295775  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 0.99561028222227 | 0.990099009901    |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
| 4        | 1           | 0.05            | 0.1   | 100 | 0.97845646790932 | 0.94252873563218  |
|          | 2           |                 |       |     | 0.98674014220257 | 0.97142857142857  |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849  |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197  |
|          | 5           |                 |       |     | 1.0              | 1.0               |
|          | 6           |                 |       |     | 0.99118942731278 | 0.98591549295775  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 0.99561028222227 | 0.990099009901    |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
| 5        | 1           | 0.05            | 0.01  | 0.1 | 0.99561531448234 | 0.98795180722892  |
|          | 2           |                 |       |     | 0.98664815503274 | 0.9708737864078   |
|          | 3           |                 |       |     | 0.99114884680230 | 0.98461538461539  |
|          | 4           |                 |       |     | 1.0              | 1.0               |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568  |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
| 6        | 1           | 0.05            | 0.01  | 1   | 0.99561531448234 | 0.98795180722892  |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 2           |                 |       |     | 0.98664815503274 | 0.9708737864078  |
|          | 3           |                 |       |     | 0.99114884680230 | 0.98461538461539 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 7        | 1           | 0.05            | 0.01  | 10  | 0.991270502724   | 0.976190476191   |
|          | 2           |                 |       |     | 0.99558004740086 | 0.99047619047619 |
|          | 3           |                 |       |     | 0.98675415674524 | 0.97709923664122 |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197 |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 8        | 1           | 0.05            | 0.01  | 100 | 0.99561531448234 | 0.98795180722892 |
|          | 2           |                 |       |     | 0.98225976666505 | 0.96153846153846 |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539 |
|          | 4           |                 |       |     | 0.99123123361531 | 0.984375         |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.98675838781672 | 0.97872340425532 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 0.99125077544033 | 0.98039215686275 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 9        | 1           | 0.05            | 1     | 0.1 | 0.73804012757198 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.57990503284127 | 0.054794520547945 |
|          | 7           |                 |       |     | 0.69961006673355 | 0.04              |
|          | 8           |                 |       |     | 0.69356692732296 | 0.03921568627451  |
|          | 9           |                 |       |     | 0.81058965070881 | 0.0625            |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
|          |             |                 |       |     |                  |                   |
| 10       | 1           | 0.05            | 1     | 1   | 0.81569847561857 | 0.63565891472868  |
|          | 2           |                 |       |     | 0.91588616016639 | 0.83870967741936  |
|          | 3           |                 |       |     | 0.82147226063618 | 0.63265306122449  |
|          | 4           |                 |       |     | 0.8354778481922  | 0.64516129032258  |
|          | 5           |                 |       |     | 0.95025831426328 | 0.91338582677165  |
|          | 6           |                 |       |     | 0.82837370316113 | 0.67289719626168  |
|          | 7           |                 |       |     | 0.98210183189094 | 0.95744680851064  |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.99562404288792 | 0.98412698412698  |
|          | 10          |                 |       |     | 0.99563422693274 | 0.9795918367347   |
|          |             |                 |       |     |                  |                   |
| 11       | 1           | 0.05            | 1     | 10  | 0.82324277828041 | 0.64566929133858  |
|          | 2           |                 |       |     | 0.91588616016639 | 0.83870967741936  |
|          | 3           |                 |       |     | 0.84430728639073 | 0.68627450980392  |
|          | 4           |                 |       |     | 0.8354778481922  | 0.64516129032258  |
|          | 5           |                 |       |     | 0.95025831426328 | 0.91338582677165  |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 6           |                 |       |     | 0.82837370316113 | 0.67289719626168  |
|          | 7           |                 |       |     | 0.98210183189094 | 0.95744680851064  |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.99562404288792 | 0.98412698412698  |
|          | 10          |                 |       |     | 0.99563422693274 | 0.97959183673469  |
|          |             |                 |       |     |                  |                   |
| 12       | 1           | 0.05            | 1     | 100 | 0.82324277828041 | 0.64566929133858  |
|          | 2           |                 |       |     | 0.91588616016639 | 0.83870967741936  |
|          | 3           |                 |       |     | 0.84430728639073 | 0.68627450980392  |
|          | 4           |                 |       |     | 0.8354778481922  | 0.64516129032258  |
|          | 5           |                 |       |     | 0.95025831426328 | 0.91338582677165  |
|          | 6           |                 |       |     | 0.82837370316113 | 0.67289719626168  |
|          | 7           |                 |       |     | 0.98210183189094 | 0.95744680851064  |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.99562404288792 | 0.98412698412698  |
|          | 10          |                 |       |     | 0.99563422693274 | 0.97959183673469  |
| 13       | 1           | 0.05            | 2     | 0.1 | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.56998165104699 | 0.027777777777778 |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.69356692732296 | 0.03921568627451  |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0               |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
| 14       | 1           | 0.05            | 2     | 1   | 0.748432284252   | 0.04761904761905  |
|          | 2           |                 |       |     | 0.731506369403   | 0.233333333333333 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 3           |                 |       |     | 0.61847434312479 | 0.086956521739131 |
|          | 4           |                 |       |     | 0.65472343099437 | 0.147058823529412 |
|          | 5           |                 |       |     | 0.67973423932371 | 0.296296296296296 |
|          | 6           |                 |       |     | 0.69927839618824 | 0.367816091954023 |
|          | 7           |                 |       |     | 0.7952441357026  | 0.39344262295082  |
|          | 8           |                 |       |     | 0.81820314962147 | 0.484848484848485 |
|          | 9           |                 |       |     | 0.93605353382991 | 0.73469387755102  |
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256  |
| 15       | 1           | 0.05            | 2     | 10  | 0.748432284252   | 0.04761904761905  |
|          | 2           |                 |       |     | 0.731506369403   | 0.233333333333333 |
|          | 3           |                 |       |     | 0.6551796287066  | 0.191780821917808 |
|          | 4           |                 |       |     | 0.69800275629357 | 0.273972602739726 |
|          | 5           |                 |       |     | 0.73220596956785 | 0.431818181818181 |
|          | 6           |                 |       |     | 0.72815610500832 | 0.43956043956044  |
|          | 7           |                 |       |     | 0.83731260033541 | 0.53731343283582  |
|          | 8           |                 |       |     | 0.8632755164249  | 0.63013698630137  |
|          | 9           |                 |       |     | 0.94697731083557 | 0.7843137254902   |
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256  |
| 16       | 1           | 0.05            | 2     | 100 | 0.748432284252   | 0.04761904761905  |
|          | 2           |                 |       |     | 0.731506369403   | 0.233333333333333 |
|          | 3           |                 |       |     | 0.6551796287066  | 0.191780821917808 |
|          | 4           |                 |       |     | 0.69800275629357 | 0.273972602739726 |
|          | 5           |                 |       |     | 0.73220596956785 | 0.431818181818181 |
|          | 6           |                 |       |     | 0.72815610500833 | 0.43956043956044  |
|          | 7           |                 |       |     | 0.83731260033541 | 0.53731343283582  |
|          | 8           |                 |       |     | 0.8632755164249  | 0.63013698630137  |
|          | 9           |                 |       |     | 0.94697731083557 | 0.7843137254902   |



## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256 |
| 17       | 1           | 0               | 0.01  | 0.1 | 0.97786712870193 | 0.93827160493827 |
|          | 2           |                 |       |     | 0.98674014220257 | 0.97142857142857 |
|          | 3           |                 |       |     | 0.98669146446219 | 0.97674418604651 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 18       | 1           | 0               | 0.01  | 1   | 0.97786712870193 | 0.93827160493827 |
|          | 2           |                 |       |     | 0.98674014220257 | 0.97142857142857 |
|          | 3           |                 |       |     | 0.98669146446219 | 0.97674418604651 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 19       | 1           | 0               | 0.01  | 10  | 0.97807657241169 | 0.93975903614458 |
|          | 2           |                 |       |     | 0.99112988333253 | 0.98076923076923 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>   |
|----------|-------------|-----------------|-------|-----|------------------|--------------------|
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667  |
|          | 8           |                 |       |     | 1.0              | 1.0                |
|          | 9           |                 |       |     | 1.0              | 1.0                |
|          | 10          |                 |       |     | 1.0              | 1.0                |
|          |             |                 |       |     |                  |                    |
| 20       | 1           | 0               | 0.01  | 100 | 0.9735682819383  | 0.9268292682927    |
|          | 2           |                 |       |     | 0.99112988333253 | 0.98076923076923   |
|          | 3           |                 |       |     | 0.98675415674524 | 0.97709923664122   |
|          | 4           |                 |       |     | 1.0              | 1.0                |
|          | 5           |                 |       |     | 0.99122430082614 | 0.98571428571429   |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844   |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667  |
|          | 8           |                 |       |     | 1.0              | 1.0                |
|          | 9           |                 |       |     | 1.0              | 1.0                |
|          | 10          |                 |       |     | 0.99555326080774 | 0.97872340425532   |
|          |             |                 |       |     |                  |                    |
| 21       | 1           | 0               | 0.1   | 0.1 | 0.75843312093337 | 0.093023255813954  |
|          | 2           |                 |       |     | 0.77193788171671 | 0.369230769230769  |
|          | 3           |                 |       |     | 0.70469978995892 | 0.32911392405063   |
|          | 4           |                 |       |     | 0.772481760043   | 0.32911392405063   |
|          | 5           |                 |       |     | 0.75957115389296 | 0.5                |
|          | 6           |                 |       |     | 0.76827308424619 | 0.5360824742268    |
|          | 7           |                 |       |     | 0.8686800542268  | 0.638888888888889  |
|          | 8           |                 |       |     | 0.82502251573808 | 0.50746268656716   |
|          | 9           |                 |       |     | 0.90622600779472 | 0.5909090909090909 |
|          | 10          |                 |       |     | 0.95068018869607 | 0.73684210526316   |
|          |             |                 |       |     |                  |                    |
| 22       | 1           | 0               | 0.1   | 1   | 0.86918368948036 | 0.56140350877193   |
|          | 2           |                 |       |     | 0.92479923405033 | 0.822222222222222  |
|          | 3           |                 |       |     | 0.95944091719289 | 0.92682926829268   |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 4           |                 |       |     | 0.95935693478145 | 0.9230769230769   |
|          | 5           |                 |       |     | 0.9866994416288  | 0.977777777777778 |
|          | 6           |                 |       |     | 0.9595667706877  | 0.93233082706767  |
|          | 7           |                 |       |     | 0.98663068897748 | 0.9684210526316   |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.98186022460669 | 0.93103448275862  |
|          | 10          |                 |       |     | 0.97187276604325 | 0.85714285714286  |
| 23       | 1           | 0               | 0.1   | 10  | 0.86918368948036 | 0.56140350877193  |
|          | 2           |                 |       |     | 0.92983988378462 | 0.83516483516484  |
|          | 3           |                 |       |     | 0.95944091719289 | 0.92682926829268  |
|          | 4           |                 |       |     | 0.95497825800969 | 0.91525423728814  |
|          | 5           |                 |       |     | 0.9866994416288  | 0.977777777777778 |
|          | 6           |                 |       |     | 0.9595667706877  | 0.93233082706767  |
|          | 7           |                 |       |     | 0.98663068897748 | 0.9684210526316   |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.98186022460669 | 0.93103448275862  |
|          | 10          |                 |       |     | 0.97187276604325 | 0.85714285714286  |
| 24       | 1           | 0               | 0.1   | 100 | 0.86918368948036 | 0.56140350877193  |
|          | 2           |                 |       |     | 0.92983988378462 | 0.83516483516484  |
|          | 3           |                 |       |     | 0.95944091719289 | 0.92682926829268  |
|          | 4           |                 |       |     | 0.95497825800969 | 0.91525423728814  |
|          | 5           |                 |       |     | 0.9866994416288  | 0.977777777777778 |
|          | 6           |                 |       |     | 0.9595667706877  | 0.93233082706767  |
|          | 7           |                 |       |     | 0.98663068897748 | 0.9684210526316   |
|          | 8           |                 |       |     | 0.97294498078545 | 0.93617021276596  |
|          | 9           |                 |       |     | 0.98186022460669 | 0.93103448275862  |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 10          |                 |       |     | 0.97187276604325 | 0.85714285714286 |
| 25       | 1           | 0               | 1     | 0.1 | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 26       | 1           | 0               | 1     | 1   | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 27       | 1           | 0               | 1     | 10  | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
|          |             |                 |       |     |                  |                  |
| 28       | 1           | 0               | 1     | 100 | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
|          |             |                 |       |     |                  |                  |
| 29       | 1           | 0               | 2     | 0.1 | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
|          |             |                 |       |     |                  |                  |
| 30       | 1           | 0               | 2     | 1   | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 31       | 1           | 0               | 2     | 10  | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 32       | 1           | 0               | 2     | 100 | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 33       | 1           | 0.01            | 0.01  | 0.1 | 0.98696363900669 | 0.96470588235294 |
|          | 2           |                 |       |     | 0.98674014220257 | 0.97142857142857 |
|          | 3           |                 |       |     | 0.98675415674524 | 0.97709923664122 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.99280575539568 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 34       | 1           | 0.01            | 0.01  | 1   | 0.98696363900669 | 0.96470588235294 |
|          | 2           |                 |       |     | 0.99112988333253 | 0.98076923076923 |
|          | 3           |                 |       |     | 0.98675415674524 | 0.97709923664122 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 1.0              | 1.0              |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571428 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 35       | 1           | 0.01            | 0.01  | 10  | 0.97807657241169 | 0.93975903614458 |
|          | 2           |                 |       |     | 0.99558004740086 | 0.99047619047619 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.99118942731278 | 0.98571428571429 |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 36       | 1           | 0.01            | 0.01  | 100 | 0.9735682819383  | 0.9268292682927   |
|          | 2           |                 |       |     | 0.99112988333253 | 0.98076923076923  |
|          | 3           |                 |       |     | 0.99558471891508 | 0.99236641221374  |
|          | 4           |                 |       |     | 1.0              | 1.0               |
|          | 5           |                 |       |     | 0.99122430082614 | 0.98571428571429  |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 37       | 1           | 0.01            | 0.1   | 0.1 | 0.80344483988339 | 0.291666666666667 |
|          | 2           |                 |       |     | 0.86501355043774 | 0.65822784810127  |
|          | 3           |                 |       |     | 0.8484747315371  | 0.6930693069307   |
|          | 4           |                 |       |     | 0.86936097965446 | 0.72727272727273  |
|          | 5           |                 |       |     | 0.86597312123385 | 0.74545454545455  |
|          | 6           |                 |       |     | 0.83957201773361 | 0.69724770642202  |
|          | 7           |                 |       |     | 0.93936070327275 | 0.84705882352941  |
|          | 8           |                 |       |     | 0.89778352757486 | 0.73417721518987  |
|          | 9           |                 |       |     | 0.95743104212704 | 0.83018867924528  |
|          | 10          |                 |       |     | 0.96157398540084 | 0.8               |
|          |             |                 |       |     |                  |                   |
| 38       | 1           | 0.01            | 0.1   | 1   | 0.95366937271677 | 0.88172043010753  |
|          | 2           |                 |       |     | 0.97299932800717 | 0.94              |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539  |



## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 4           |                 |       |     | 0.98668286255643 | 0.9756097560976  |
|          | 5           |                 |       |     | 1.0              | 1.0              |
|          | 6           |                 |       |     | 0.98223500194363 | 0.97101449275362 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97959183673469 |
|          | 8           |                 |       |     | 0.99112427230953 | 0.97959183673469 |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256 |
| 39       | 1           | 0.01            | 0.1   | 10  | 0.95366937271677 | 0.88172043010753 |
|          | 2           |                 |       |     | 0.97758435413496 | 0.95049504950495 |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539 |
|          | 4           |                 |       |     | 0.9911450730124  | 0.98387096774194 |
|          | 5           |                 |       |     | 1.0              | 1.0              |
|          | 6           |                 |       |     | 0.98223500194363 | 0.97101449275362 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97959183673469 |
|          | 8           |                 |       |     | 0.99112427230953 | 0.97959183673469 |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256 |
| 40       | 1           | 0.01            | 0.1   | 100 | 0.95366937271677 | 0.88172043010753 |
|          | 2           |                 |       |     | 0.97758435413496 | 0.95049504950495 |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539 |
|          | 4           |                 |       |     | 0.9911450730124  | 0.98387096774194 |
|          | 5           |                 |       |     | 1.0              | 1.0              |
|          | 6           |                 |       |     | 0.98223500194363 | 0.97101449275362 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97959183673469 |
|          | 8           |                 |       |     | 0.99112427230953 | 0.97959183673469 |
|          | 9           |                 |       |     | 1.0              | 1.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 10          |                 |       |     | 0.97682694251491 | 0.88372093023256  |
| 41       | 1           | 0.01            | 1     | 0.1 | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0               |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0               |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0               |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
| 42       | 1           | 0.01            | 1     | 1   | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0               |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0               |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0               |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
| 43       | 1           | 0.01            | 1     | 10  | 0.748432284252   | 0.04761904761905  |
|          | 2           |                 |       |     | 0.67551966062979 | 0.037037037037037 |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0               |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.69356692732296 | 0.03921568627451  |
|          | 9           |                 |       |     | 0.81058965070881 | 0.0625            |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
|          |             |                 |       |     |                  |                   |
| 44       | 1           | 0.01            | 1     | 100 | 0.748432284252   | 0.04761904761905  |
|          | 2           |                 |       |     | 0.67551966062979 | 0.037037037037037 |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0               |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.69356692732296 | 0.03921568627451  |
|          | 9           |                 |       |     | 0.81058965070881 | 0.0625            |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
|          |             |                 |       |     |                  |                   |
| 45       | 1           | 0.01            | 2     | 0.1 | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0               |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0               |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0               |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
|          |             |                 |       |     |                  |                   |
| 46       | 1           | 0.01            | 2     | 1   | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 47       | 1           | 0.01            | 2     | 10  | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 48       | 1           | 0.01            | 2     | 100 | 0.73804012757197 | 0.0              |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0              |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0              |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0              |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0              |
|          | 6           |                 |       |     | 0.55982792928538 | 0.0              |
|          | 7           |                 |       |     | 0.689269592647   | 0.0              |
|          | 8           |                 |       |     | 0.68323374187639 | 0.0              |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 10          |                 |       |     | 0.84436020899498 | 0.0              |
| 49       | 1           | 0.1             | 0.1   | 0.1 | 0.99561531448234 | 0.98795180722892 |
|          | 2           |                 |       |     | 0.9609564098461  | 0.91891891891892 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.98686140042648 | 0.97872340425532 |
|          | 6           |                 |       |     | 0.99118942731278 | 0.98591549295775 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 50       | 1           | 0.1             | 0.1   | 1   | 1.0              | 1.0              |
|          | 2           |                 |       |     | 0.97774692505457 | 0.95145631067961 |
|          | 3           |                 |       |     | 0.99558471891508 | 0.99236641221374 |
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 51       | 1           | 0.1             | 0.1   | 10  | 0.99561531448234 | 0.98795180722892 |
|          | 2           |                 |       |     | 0.9779002370043  | 0.95238095238095 |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849 |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197 |
|          | 5           |                 |       |     | 1.0              | 1.0              |
|          | 6           |                 |       |     | 0.99560532139157 | 0.99212598425197 |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 52       | 1           | 0.1             | 0.1   | 100 | 0.99561531448234 | 0.98795180722892  |
|          | 2           |                 |       |     | 0.9779002370043  | 0.95238095238095  |
|          | 3           |                 |       |     | 0.99118942731278 | 0.98484848484849  |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197  |
|          | 5           |                 |       |     | 1.0              | 1.0               |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 53       | 1           | 0.1             | 0.01  | 0.1 | 1.0              | 1.0               |
|          | 2           |                 |       |     | 0.98664815503274 | 0.9708737864078   |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539  |
|          | 4           |                 |       |     | 1.0              | 1.0               |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568  |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429  |
|          | 7           |                 |       |     | 0.99112226105218 | 0.979166666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0               |
|          | 9           |                 |       |     | 1.0              | 1.0               |
|          | 10          |                 |       |     | 1.0              | 1.0               |
|          |             |                 |       |     |                  |                   |
| 54       | 1           | 0.1             | 0.01  | 1   | 1.0              | 1.0               |
|          | 2           |                 |       |     | 0.98664815503274 | 0.9708737864078   |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539  |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|-----|------------------|------------------|
|          | 4           |                 |       |     | 1.0              | 1.0              |
|          | 5           |                 |       |     | 0.99122430082614 | 0.98571428571428 |
|          | 6           |                 |       |     | 0.99558612927224 | 0.99290780141844 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 55       | 1           | 0.1             | 0.01  | 10  | 0.99561531448234 | 0.98795180722892 |
|          | 2           |                 |       |     | 0.99558004740086 | 0.99047619047619 |
|          | 3           |                 |       |     | 0.9911488468023  | 0.98461538461539 |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197 |
|          | 5           |                 |       |     | 0.99122430082614 | 0.98571428571429 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99112226105218 | 0.97916666666667 |
|          | 8           |                 |       |     | 1.0              | 1.0              |
|          | 9           |                 |       |     | 1.0              | 1.0              |
|          | 10          |                 |       |     | 1.0              | 1.0              |
| 56       | 1           | 0.1             | 0.01  | 100 | 0.99127050272425 | 0.97619047619048 |
|          | 2           |                 |       |     | 0.98225976666505 | 0.96153846153846 |
|          | 3           |                 |       |     | 0.98669146446219 | 0.97674418604651 |
|          | 4           |                 |       |     | 0.99560532139157 | 0.99212598425197 |
|          | 5           |                 |       |     | 0.9956035674921  | 0.99280575539568 |
|          | 6           |                 |       |     | 0.99115455379941 | 0.98571428571429 |
|          | 7           |                 |       |     | 0.99557817586365 | 0.98969072164949 |
|          | 8           |                 |       |     | 0.99561028222228 | 0.990099009901   |
|          | 9           |                 |       |     | 1.0              | 1.0              |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 10          |                 |       |     | 1.0              | 1.0               |
| 57       | 1           | 0.1             | 1     | 0.1 | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.59909910538204 | 0.106666666666667 |
|          | 7           |                 |       |     | 0.71328079933157 | 0.113207547169811 |
|          | 8           |                 |       |     | 0.69961006673355 | 0.04              |
|          | 9           |                 |       |     | 0.8639355080347  | 0.36842105263158  |
|          | 10          |                 |       |     | 0.89859945141717 | 0.4               |
| 58       | 1           | 0.1             | 1     | 1   | 0.93764170821151 | 0.84536082474227  |
|          | 2           |                 |       |     | 0.90782070501636 | 0.82539682539683  |
|          | 3           |                 |       |     | 0.94831604376865 | 0.916666666666667 |
|          | 4           |                 |       |     | 0.95264649385198 | 0.91970802919708  |
|          | 5           |                 |       |     | 0.94820774402968 | 0.92              |
|          | 6           |                 |       |     | 0.99122219064311 | 0.986111111111111 |
|          | 7           |                 |       |     | 0.99561075794866 | 0.989898989899    |
|          | 8           |                 |       |     | 0.98692012690559 | 0.9708737864078   |
|          | 9           |                 |       |     | 0.991304501299   | 0.96875           |
|          | 10          |                 |       |     | 0.99563422693274 | 0.97959183673469  |
| 59       | 1           | 0.1             | 1     | 10  | 0.93764170821151 | 0.84536082474227  |
|          | 2           |                 |       |     | 0.90782070501636 | 0.82539682539683  |
|          | 3           |                 |       |     | 0.94831604376865 | 0.916666666666667 |
|          | 4           |                 |       |     | 0.96111471697071 | 0.933333333333333 |
|          | 5           |                 |       |     | 0.94820774402968 | 0.92              |
|          | 6           |                 |       |     | 0.99122219064311 | 0.986111111111111 |



**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 7           |                 |       |     | 0.99561075794866 | 0.989898989899    |
|          | 8           |                 |       |     | 0.98692012690559 | 0.9708737864078   |
|          | 9           |                 |       |     | 0.991304501299   | 0.96875           |
|          | 10          |                 |       |     | 0.99563422693274 | 0.97959183673469  |
|          |             |                 |       |     |                  |                   |
| 60       | 1           | 0.1             | 1     | 100 | 0.93764170821151 | 0.84536082474227  |
|          | 2           |                 |       |     | 0.90782070501636 | 0.82539682539683  |
|          | 3           |                 |       |     | 0.94831604376865 | 0.91666666666667  |
|          | 4           |                 |       |     | 0.95687407204591 | 0.92647058823529  |
|          | 5           |                 |       |     | 0.94820774402968 | 0.92              |
|          | 6           |                 |       |     | 0.99122219064311 | 0.98611111111111  |
|          | 7           |                 |       |     | 0.99561075794866 | 0.989898989899    |
|          | 8           |                 |       |     | 0.98692012690559 | 0.9708737864078   |
|          | 9           |                 |       |     | 0.991304501299   | 0.96875           |
|          | 10          |                 |       |     | 0.99563422693274 | 0.97959183673469  |
|          |             |                 |       |     |                  |                   |
| 61       | 1           | 0.1             | 2     | 0.1 | 0.73804012757197 | 0.0               |
|          | 2           |                 |       |     | 0.66520922363694 | 0.0               |
|          | 3           |                 |       |     | 0.58860529542668 | 0.0               |
|          | 4           |                 |       |     | 0.60605924039794 | 0.0               |
|          | 5           |                 |       |     | 0.57129126380228 | 0.0               |
|          | 6           |                 |       |     | 0.58960778538798 | 0.081081081081081 |
|          | 7           |                 |       |     | 0.689269592647   | 0.0               |
|          | 8           |                 |       |     | 0.69356692732296 | 0.03921568627451  |
|          | 9           |                 |       |     | 0.80015829870549 | 0.0               |
|          | 10          |                 |       |     | 0.84436020899498 | 0.0               |
|          |             |                 |       |     |                  |                   |
| 62       | 1           | 0.1             | 2     | 1   | 0.81159559271579 | 0.3265306122449   |
|          | 2           |                 |       |     | 0.77947420065591 | 0.3939393939394   |
|          | 3           |                 |       |     | 0.69355638580643 | 0.30379746835443  |

## A. PENGUJIAN SVM DENGAN SU

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C   | Akurasi          | <i>F-Measure</i>  |
|----------|-------------|-----------------|-------|-----|------------------|-------------------|
|          | 4           |                 |       |     | 0.73698061241411 | 0.38461538461539  |
|          | 5           |                 |       |     | 0.76615688267436 | 0.51612903225807  |
|          | 6           |                 |       |     | 0.76182587634025 | 0.52083333333333  |
|          | 7           |                 |       |     | 0.85021147530245 | 0.57971014492754  |
|          | 8           |                 |       |     | 0.88093581972553 | 0.68421052631579  |
|          | 9           |                 |       |     | 0.95225920549742 | 0.80769230769231  |
|          | 10          |                 |       |     | 0.98166385027887 | 0.90909090909091  |
| 63       | 1           | 0.1             | 2     | 10  | 0.81159559271579 | 0.3265306122449   |
|          | 2           |                 |       |     | 0.77947420065591 | 0.393939393939394 |
|          | 3           |                 |       |     | 0.69355638580643 | 0.30379746835443  |
|          | 4           |                 |       |     | 0.73698061241411 | 0.38461538461539  |
|          | 5           |                 |       |     | 0.76615688267436 | 0.51612903225807  |
|          | 6           |                 |       |     | 0.76182587634025 | 0.520833333333333 |
|          | 7           |                 |       |     | 0.85021147530245 | 0.57971014492754  |
|          | 8           |                 |       |     | 0.88093581972553 | 0.68421052631579  |
|          | 9           |                 |       |     | 0.95225920549742 | 0.80769230769231  |
|          | 10          |                 |       |     | 0.98166385027887 | 0.90909090909091  |
| 64       | 1           | 0.1             | 2     | 100 | 0.98166385027887 | 0.90909090909091  |
|          | 2           |                 |       |     | 0.77947420065591 | 0.393939393939394 |
|          | 3           |                 |       |     | 0.69355638580643 | 0.30379746835443  |
|          | 4           |                 |       |     | 0.73698061241411 | 0.38461538461539  |
|          | 5           |                 |       |     | 0.76615688267436 | 0.51612903225807  |
|          | 6           |                 |       |     | 0.76182587634025 | 0.520833333333333 |
|          | 7           |                 |       |     | 0.85021147530245 | 0.57971014492754  |
|          | 8           |                 |       |     | 0.88093581972553 | 0.68421052631579  |
|          | 9           |                 |       |     | 0.95225920549742 | 0.80769230769231  |

## A. PENGUJIAN SVM DENGAN SU

---

**Tabel A-1** Hasil Pengujian *Support Vector Machine* (Lanjutan)

| Skenario | <i>Fold</i> | SU<br>Threshold | Gamma | C | Akurasi          | <i>F-Measure</i> |
|----------|-------------|-----------------|-------|---|------------------|------------------|
|          | 10          |                 |       |   | 0.98166385027887 | 0.90909090909091 |

## LAMPIRAN B PENGUJIAN SVM TANPA SU

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur

| Skenario | <i>Fold</i> | Gamma | C   | Akurasi            | <i>F-Measure</i>    |
|----------|-------------|-------|-----|--------------------|---------------------|
| 1        | 1           | 0.01  | 0.1 | 0.706394143560508  | 0.17500000000000002 |
|          | 2           |       |     | 0.7450100515210871 | 0.2857142857142857  |
|          | 3           |       |     | 0.6856240581988075 | 0.28205128205128205 |
|          | 4           |       |     | 0.7443362144277774 | 0.4050632911392405  |
|          | 5           |       |     | 0.7977010578268009 | 0.6019417475728155  |
|          | 6           |       |     | 0.7809067908093782 | 0.5656565656565656  |
|          | 7           |       |     | 0.902941904535306  | 0.7435897435897436  |
|          | 8           |       |     | 0.893560507903602  | 0.725               |
|          | 9           |       |     | 0.9539192184965102 | 0.8214285714285714  |
|          | 10          |       |     | 0.9718727660432462 | 0.8571428571428571  |
| 2        | 1           | 0.01  | 1   | 0.6679473861609667 | 0.37037037037037035 |
|          | 2           |       |     | 0.779300276611003  | 0.46511627906976744 |
|          | 3           |       |     | 0.7507164938205767 | 0.49504950495049505 |
|          | 4           |       |     | 0.8055647296816942 | 0.5806451612903225  |
|          | 5           |       |     | 0.8743794737921021 | 0.7846153846153846  |
|          | 6           |       |     | 0.8710633405247586 | 0.7741935483870968  |
|          | 7           |       |     | 0.9249751954597769 | 0.8095238095238094  |
|          | 8           |       |     | 0.936871621832724  | 0.8510638297872342  |
|          | 9           |       |     | 0.9823788546255506 | 0.9354838709677419  |
|          | 10          |       |     | 0.973058462994468  | 0.8695652173913043  |
| 3        | 1           | 0.01  | 10  | 0.75880883630766   | 0.4271844660194175  |
|          | 2           |       |     | 0.8339300777954822 | 0.6170212765957447  |
|          | 3           |       |     | 0.8707085744338943 | 0.7716535433070867  |
|          | 4           |       |     | 0.9045654631652634 | 0.8173913043478261  |
|          | 5           |       |     | 0.9469140993609646 | 0.9117647058823529  |

## B. PENGUJIAN SVM TANPA SU

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur (Lanjutan)

| Skenario | Fold | Gamma | C   | Akurasi            | F-Measure           |
|----------|------|-------|-----|--------------------|---------------------|
|          | 6    |       |     | 0.9189325327038409 | 0.8636363636363635  |
|          | 7    |       |     | 0.9777178149624733 | 0.9473684210526316  |
|          | 8    |       |     | 0.9556213615476383 | 0.8979591836734694  |
|          | 9    |       |     | 0.9866926283258093 | 0.9508196721311476  |
|          | 10   |       |     | 0.9910194876648226 | 0.9565217391304348  |
| 4        | 1    | 0.01  | 100 | 0.7844773508076696 | 0.46938775510204084 |
|          | 2    |       |     | 0.7933222549951012 | 0.5263157894736842  |
|          | 3    |       |     | 0.880223180159662  | 0.7906976744186046  |
|          | 4    |       |     | 0.9184599564774607 | 0.8448275862068966  |
|          | 5    |       |     | 0.9289039384574922 | 0.8805970149253731  |
|          | 6    |       |     | 0.9232411155338999 | 0.8702290076335878  |
|          | 7    |       |     | 0.9733667831565477 | 0.9375000000000001  |
|          | 8    |       |     | 0.9688157019472488 | 0.9278350515463918  |
|          | 9    |       |     | 0.9866926283258093 | 0.9508196721311476  |
|          | 10   |       |     | 0.9910194876648226 | 0.9565217391304348  |
| 5        | 1    | 0.1   | 0.1 | 0.7584331209333733 | 0.09302325581395349 |
|          | 2    |       |     | 0.7284351282858412 | 0.22950819672131148 |
|          | 3    |       |     | 0.6184743431247932 | 0.08695652173913045 |
|          | 4    |       |     | 0.7140590238629364 | 0.32                |
|          | 5    |       |     | 0.7285637084957544 | 0.42696629213483145 |
|          | 6    |       |     | 0.7381822726278626 | 0.46808510638297873 |
|          | 7    |       |     | 0.8502114753024541 | 0.5797101449275363  |
|          | 8    |       |     | 0.8182031496214703 | 0.48484848484848486 |
|          | 9    |       |     | 0.9245938138992782 | 0.6808510638297872  |
|          | 10   |       |     | 0.9615739854008385 | 0.8                 |
| 6        | 1    | 0.1   | 1   | 0.7769722292180169 | 0.30303030303030304 |
|          | 2    |       |     | 0.7483581106474648 | 0.34666666666666666 |
|          | 3    |       |     | 0.7483278525780233 | 0.4848484848484849  |

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur (Lanjutan)

| Skenario | <i>Fold</i> | Gamma | C   | Akurasi            | <i>F-Measure</i>     |
|----------|-------------|-------|-----|--------------------|----------------------|
|          | 4           |       |     | 0.7909938778928126 | 0.5393258426966292   |
|          | 5           |       |     | 0.8356269411199044 | 0.6956521739130435   |
|          | 6           |       |     | 0.8517194084228509 | 0.7256637168141593   |
|          | 7           |       |     | 0.9071838683670153 | 0.7674418604651163   |
|          | 8           |       |     | 0.9208412994525437 | 0.8045977011494254   |
|          | 9           |       |     | 0.959494994078713  | 0.8474576271186439   |
|          | 10          |       |     | 0.977766304038682  | 0.8936170212765957   |
| 7        | 1           | 0.1   | 10  | 0.7892524167676604 | 0.36619718309859156  |
|          | 2           |       |     | 0.7539701300096701 | 0.4                  |
|          | 3           |       |     | 0.7604923175923188 | 0.5148514851485148   |
|          | 4           |       |     | 0.8016267274866957 | 0.574468085106383    |
|          | 5           |       |     | 0.8573286578893308 | 0.743801652892562    |
|          | 6           |       |     | 0.8872972021631769 | 0.7966101694915254   |
|          | 7           |       |     | 0.9172827064608678 | 0.7954545454545455   |
|          | 8           |       |     | 0.9308328683045174 | 0.8314606741573033   |
|          | 9           |       |     | 0.9680046830866712 | 0.8771929824561403   |
|          | 10          |       |     | 0.9820389753296455 | 0.9130434782608695   |
| 8        | 1           | 0.1   | 100 | 0.7892524167676604 | 0.36619718309859156  |
|          | 2           |       |     | 0.7539701300096701 | 0.4                  |
|          | 3           |       |     | 0.7604923175923188 | 0.5148514851485148   |
|          | 4           |       |     | 0.8016267274866957 | 0.574468085106383    |
|          | 5           |       |     | 0.8573286578893308 | 0.743801652892562    |
|          | 6           |       |     | 0.8770317484429591 | 0.7758620689655172   |
|          | 7           |       |     | 0.9172827064608678 | 0.7954545454545455   |
|          | 8           |       |     | 0.9308328683045174 | 0.8314606741573033   |
|          | 9           |       |     | 0.9680046830866712 | 0.8771929824561403   |
|          | 10          |       |     | 0.9820389753296455 | 0.9130434782608695   |
| 9        | 1           | 1     | 0.1 | 0.748432284251967  | 0.047619047619047616 |

## B. PENGUJIAN SVM TANPA SU

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur (Lanjutan)

| Skenario | Fold | Gamma | C  | Akurasi            | F-Measure            |
|----------|------|-------|----|--------------------|----------------------|
|          | 2    |       |    | 0.6755196606297926 | 0.037037037037037035 |
|          | 3    |       |    | 0.5886052954266768 | 0.0                  |
|          | 4    |       |    | 0.6162861459392296 | 0.03125              |
|          | 5    |       |    | 0.5712912638022769 | 0.0                  |
|          | 6    |       |    | 0.5598279292853774 | 0.0                  |
|          | 7    |       |    | 0.6996100667335455 | 0.039999999999999994 |
|          | 8    |       |    | 0.6832337418763902 | 0.0                  |
|          | 9    |       |    | 0.8105896507088127 | 0.0625               |
|          | 10   |       |    | 0.8443602089949799 | 0.0                  |
| 10       | 1    | 1     | 1  | 0.748432284251967  | 0.047619047619047616 |
|          | 2    |       |    | 0.7138635327910734 | 0.1724137931034483   |
|          | 3    |       |    | 0.6980027562935659 | 0.27397260273972607  |
|          | 4    |       |    | 0.6980027562935659 | 0.27397260273972607  |
|          | 5    |       |    | 0.6797342393237091 | 0.29629629629629634  |
|          | 6    |       |    | 0.6917546758840638 | 0.3488372093023256   |
|          | 7    |       |    | 0.8098866693456415 | 0.4444444444444445   |
|          | 8    |       |    | 0.7577990139892008 | 0.2758620689655173   |
|          | 9    |       |    | 0.9062260077947195 | 0.5909090909090909   |
|          | 10   |       |    | 0.9390847964926333 | 0.6666666666666666   |
| 11       | 1    | 1     | 10 | 0.748432284251967  | 0.047619047619047616 |
|          | 2    |       |    | 0.7138635327910734 | 0.1724137931034483   |
|          | 3    |       |    | 0.6087579401896112 | 0.05882352941176471  |
|          | 4    |       |    | 0.6980027562935659 | 0.27397260273972607  |
|          | 5    |       |    | 0.7104964918527132 | 0.37647058823529417  |
|          | 6    |       |    | 0.6992783961882386 | 0.36781609195402304  |
|          | 7    |       |    | 0.8098866693456415 | 0.4444444444444445   |
|          | 8    |       |    | 0.7968526873176648 | 0.41269841269841273  |
|          | 9    |       |    | 0.9062260077947195 | 0.5909090909090909   |

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur (Lanjutan)

| Skenario | <i>Fold</i> | Gamma | C   | Akurasi            | <i>F-Measure</i>     |
|----------|-------------|-------|-----|--------------------|----------------------|
|          | 10          |       |     | 0.9390847964926333 | 0.6666666666666666   |
| 12       | 1           | 1     | 100 | 0.748432284251967  | 0.047619047619047616 |
|          | 2           |       |     | 0.7138635327910734 | 0.1724137931034483   |
|          | 3           |       |     | 0.6087579401896112 | 0.05882352941176471  |
|          | 4           |       |     | 0.6980027562935659 | 0.27397260273972607  |
|          | 5           |       |     | 0.7104964918527132 | 0.37647058823529417  |
|          | 6           |       |     | 0.6992783961882386 | 0.36781609195402304  |
|          | 7           |       |     | 0.8098866693456415 | 0.4444444444444445   |
|          | 8           |       |     | 0.7968526873176648 | 0.41269841269841273  |
|          | 9           |       |     | 0.9062260077947195 | 0.5909090909090909   |
|          | 10          |       |     | 0.9390847964926333 | 0.6666666666666666   |
| 13       | 1           | 2     | 0.1 | 0.7380401275719725 | 0.0                  |
|          | 2           |       |     | 0.6652092236369428 | 0.0                  |
|          | 3           |       |     | 0.5886052954266768 | 0.0                  |
|          | 4           |       |     | 0.6060592403979405 | 0.0                  |
|          | 5           |       |     | 0.5712912638022769 | 0.0                  |
|          | 6           |       |     | 0.5598279292853774 | 0.0                  |
|          | 7           |       |     | 0.689269592646979  | 0.0                  |
|          | 8           |       |     | 0.6832337418763902 | 0.0                  |
|          | 9           |       |     | 0.8001582987054915 | 0.0                  |
|          | 10          |       |     | 0.8443602089949799 | 0.0                  |
| 14       | 1           | 2     | 1   | 0.748432284251967  | 0.047619047619047616 |
|          | 2           |       |     | 0.695239096679116  | 0.10714285714285715  |
|          | 3           |       |     | 0.5886052954266768 | 0.0                  |
|          | 4           |       |     | 0.6262541824957598 | 0.06153846153846154  |
|          | 5           |       |     | 0.6289698139086164 | 0.16                 |
|          | 6           |       |     | 0.62638941962215   | 0.17948717948717946  |
|          | 7           |       |     | 0.7378593571545115 | 0.1851851851851852   |



## B. PENGUJIAN SVM TANPA SU

**Tabel B-1** Hasil Pengujian *Support Vector Machine* tanpa Seleksi Fitur (Lanjutan)

| Skenario | <i>Fold</i> | Gamma | C   | Akurasi            | <i>F-Measure</i>     |
|----------|-------------|-------|-----|--------------------|----------------------|
|          | 8           |       |     | 0.6935669273229563 | 0.0392156862745098   |
|          | 9           |       |     | 0.8559623547738783 | 0.3243243243243243   |
|          | 10          |       |     | 0.8737873426253581 | 0.2222222222222222   |
|          |             |       |     |                    |                      |
| 15       | 1           | 2     | 10  | 0.748432284251967  | 0.047619047619047616 |
|          | 2           |       |     | 0.695239096679116  | 0.10714285714285715  |
|          | 3           |       |     | 0.5886052954266768 | 0.0                  |
|          | 4           |       |     | 0.6454616919840235 | 0.11940298507462686  |
|          | 5           |       |     | 0.6289698139086164 | 0.16                 |
|          | 6           |       |     | 0.635117827468912  | 0.20253164556962025  |
|          | 7           |       |     | 0.7378593571545115 | 0.1851851851851852   |
|          | 8           |       |     | 0.70357632373544   | 0.07692307692307693  |
|          | 9           |       |     | 0.863935508034743  | 0.3684210526315789   |
|          | 10          |       |     | 0.8824961516529426 | 0.2857142857142857   |
|          |             |       |     |                    |                      |
| 16       | 1           | 2     | 100 | 0.748432284251967  | 0.047619047619047616 |
|          | 2           |       |     | 0.695239096679116  | 0.10714285714285715  |
|          | 3           |       |     | 0.5886052954266768 | 0.0                  |
|          | 4           |       |     | 0.6454616919840235 | 0.11940298507462686  |
|          | 5           |       |     | 0.6289698139086164 | 0.16                 |
|          | 6           |       |     | 0.635117827468912  | 0.20253164556962025  |
|          | 7           |       |     | 0.7378593571545115 | 0.1851851851851852   |
|          | 8           |       |     | 0.70357632373544   | 0.07692307692307693  |
|          | 9           |       |     | 0.863935508034743  | 0.3684210526315789   |
|          | 10          |       |     | 0.8824961516529426 | 0.2857142857142857   |
|          |             |       |     |                    |                      |

## LAMPIRAN C PENGUJIAN JUMLAH DATA

**Tabel C-1** Hasil Pengujian Jumlah Data Film pada *Support Vector Machine* dengan Seleksi Fitur

| <b>Jumlah Data Film Awal</b> | <b><i>Fold</i></b> | <b><i>SU Threshold</i></b> | <b>Gamma</b> | <b>C</b> | <b>Akurasi</b>     | <b><i>F-Measure</i></b> |
|------------------------------|--------------------|----------------------------|--------------|----------|--------------------|-------------------------|
| 4544                         | 1                  | 0.1                        | 0.01         | 0.1      | 0.9900990099009901 | 0.9803921568627451      |
|                              | 2                  |                            |              |          | 0.9900426777371616 | 0.9818181818181818      |
|                              | 3                  |                            |              |          | 1.0                | 1.0                     |
|                              | 4                  |                            |              |          | 1.0                | 1.0                     |
|                              | 5                  |                            |              |          | 0.9900990099009901 | 0.9846153846153847      |
|                              | 6                  |                            |              |          | 0.9900388291330082 | 0.9811320754716981      |
|                              | 7                  |                            |              |          | 1.0                | 1.0                     |
|                              | 8                  |                            |              |          | 1.0                | 1.0                     |
|                              | 9                  |                            |              |          | 1.0                | 1.0                     |
|                              | 10                 |                            |              |          | 1.0                | 1.0                     |
| 4044                         | 1                  | 0.1                        | 0.01         | 0.1      | 0.9889564800279086 | 0.9821428571428572      |
|                              | 2                  |                            |              |          | 1.0                | 1.0                     |
|                              | 3                  |                            |              |          | 1.0                | 1.0                     |
|                              | 4                  |                            |              |          | 0.9944901221706217 | 0.9906542056074767      |
|                              | 5                  |                            |              |          | 0.9944601540724722 | 0.9906542056074767      |
|                              | 6                  |                            |              |          | 0.9944457235333551 | 0.9859154929577464      |
|                              | 7                  |                            |              |          | 1.0                | 1.0                     |
|                              | 8                  |                            |              |          | 1.0                | 1.0                     |
|                              | 9                  |                            |              |          | 1.0                | 1.0                     |
|                              | 10                 |                            |              |          | 1.0                | 1.0                     |
| 3544                         | 1                  | 0.1                        | 0.01         | 0.1      | 1.0                | 1.0                     |
|                              | 2                  |                            |              |          | 1.0                | 1.0                     |

### C. PENGUJIAN JUMLAH DATA

**Tabel C-1** Hasil Pengujian Jumlah Data Film pada *Support Vector Machine* dengan Seleksi Fitur (Lanjutan)

| Jumlah Data Film Awal | Fold | Fold | Gamma | C   | Akurasi            | F-Measure          |
|-----------------------|------|------|-------|-----|--------------------|--------------------|
|                       | 3    |      |       |     | 1.0                | 1.0                |
|                       | 4    |      |       |     | 1.0                | 1.0                |
|                       | 5    |      |       |     | 0.9900990099009901 | 0.9846153846153847 |
|                       | 6    |      |       |     | 1.0                | 1.0                |
|                       | 7    |      |       |     | 0.9937587497954038 | 0.9863013698630138 |
|                       | 8    |      |       |     | 1.0                | 1.0                |
|                       | 9    |      |       |     | 1.0                | 1.0                |
|                       | 10   |      |       |     | 1.0                | 1.0                |
| 3044                  | 1    | 0.1  | 0.01  | 0.1 | 1.0                | 1.0                |
|                       | 2    |      |       |     | 1.0                | 1.0                |
|                       | 3    |      |       |     | 1.0                | 1.0                |
|                       | 4    |      |       |     | 0.9929340155624851 | 0.988235294117647  |
|                       | 5    |      |       |     | 1.0                | 1.0                |
|                       | 6    |      |       |     | 1.0                | 1.0                |
|                       | 7    |      |       |     | 0.9929065530978898 | 0.9811320754716981 |
|                       | 8    |      |       |     | 1.0                | 1.0                |
|                       | 9    |      |       |     | 1.0                | 1.0                |
|                       | 10   |      |       |     | 1.0                | 1.0                |
| 2544                  | 1    | 0.1  | 0.01  | 0.1 | 1.0                | 1.0                |
|                       | 2    |      |       |     | 1.0                | 1.0                |
|                       | 3    |      |       |     | 1.0                | 1.0                |
|                       | 4    |      |       |     | 1.0                | 1.0                |
|                       | 5    |      |       |     | 1.0                | 1.0                |
|                       | 6    |      |       |     | 1.0                | 1.0                |
|                       | 7    |      |       |     | 1.0                | 1.0                |

### C. PENGUJIAN JUMLAH DATA

**Tabel C-1** Hasil Pengujian Jumlah Data Film pada *Support Vector Machine* dengan Seleksi Fitur (Lanjutan)

| <b>Jumlah Data Film Awal</b> | <b>Fold</b> | <b>Fold</b> | <b>Gamma</b> | <b>C</b> | <b>Akurasi</b> | <b>F-Measure</b> |
|------------------------------|-------------|-------------|--------------|----------|----------------|------------------|
|                              | 8           |             |              |          | 1.0            | 1.0              |
|                              | 9           |             |              |          | 1.0            | 1.0              |
|                              | 10          |             |              |          | 1.0            | 1.0              |
|                              |             |             |              |          |                |                  |
| 2044                         | 1           | 0.1         | 0.01         | 0.1      | 1.0            | 1.0              |
|                              | 2           |             |              |          | 1.0            | 1.0              |
|                              | 3           |             |              |          | 1.0            | 1.0              |
|                              | 4           |             |              |          | 1.0            | 1.0              |
|                              | 5           |             |              |          | 1.0            | 1.0              |
|                              | 6           |             |              |          | 1.0            | 1.0              |
|                              | 7           |             |              |          | 1.0            | 1.0              |
|                              | 8           |             |              |          | 1.0            | 1.0              |
|                              | 9           |             |              |          | 1.0            | 1.0              |
|                              | 10          |             |              |          | 1.0            | 1.0              |
| 1544                         | 1           | 0.1         | 0.01         | 0.1      | 1.0            | 1.0              |
|                              | 2           |             |              |          | 1.0            | 1.0              |
|                              | 3           |             |              |          | 1.0            | 1.0              |
|                              | 4           |             |              |          | 1.0            | 1.0              |
|                              | 5           |             |              |          | 1.0            | 1.0              |
|                              | 6           |             |              |          | 1.0            | 1.0              |
|                              | 7           |             |              |          | 1.0            | 1.0              |
|                              | 8           |             |              |          | 1.0            | 1.0              |
|                              | 9           |             |              |          | 1.0            | 1.0              |
|                              | 10          |             |              |          | 1.0            | 1.0              |