KLASIFIKASI PENGADUAN LAYANAN PENGGUNA INDIHOME PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Oleh :
MUHAMMAD AMMARULLAH RIDHO
NIM. 17650071

JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2021

KLASIFIKASI PENGADUAN LAYANAN PENGGUNA INDIHOME PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
MUHAMMAD AMMARULLAH RIDHO
NIM. 17650071

JURUSAN TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2021

HALAMAN PERSETUJUAN

KLASIFIKASI PENGADUAN LAYANAN PENGGUNA INDIHOME PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Oleh :
MUHAMMAD AMMARULLAH RIDHO
NIM. 17650071

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : 7 Juni 2021

Dosen Pembimbing I	Dosen Pembimbing II
Fajar Rohman Hariri, M.Kom	Prof. Dr. Suhartono, M.Kom
NIP. 19890515 201801 1 001	NIP. 19680619 200312 1 001

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

HALAMAN PENGESAHAN

KLASIFIKASI PENGADUAN LAYANAN PENGGUNA INDIHOME PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Oleh:
MUHAMMAD AMMARULLAH RIDHO
NIM.17650071

Telah Dipertahankan di Depan Dewan Penguji
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal 7 Juni 2021

Susunan Dewnn Penguji

1. Penguji Utama	:	Dr. Cahyo Crysdian NIP. 19740424 200901 1 008	( )
2. Ketua Penguji	:	Irwan Budi Santoso, M.Kom NIP. 19770103 201101 1 004	( )
3. Sekretaris Penguji	:	Fajar Rohman Hariri, M.Kom NIP. 19890515 201801 1 001	( )
4. Anggota Penguji	:	Prof. Dr. Suhartono, M.Kom NIP. 19680519 200312 1 001	( )

Mengetahui,
Ketua Jurusan Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Cahyo Crysdian
NIP. 19740424 200901 1 008

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama	:	Muhammad Ammarullah Ridho
NIM	:	17650071
Fakultas	:	Sains dan Teknologi
Jurusan	:	Teknik Informatika
Judul Skripsi	:	Klasifikasi Pengaduan Layanan Pengguna Indihome Pada Media Sosial Twitter Menggunakan Metode *Support Vector Machine* dengan Seleksi Fitur *Information Gain*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 1 Juni 2021

Yang membuat pernyataan,

Muhammad Ammarullah Ridho

NIM. 17650071

HALAMAN MOTTO

إِنَّ مَعَ الْعُسْرِ يُسْرًا

“Sesungguhnya, bersama kesulitan ada kemudahan”

(Q.S. Al-Insyirah :6)

HALAMAN PERSEMBAHAN

Skripsi ini saya persembahkan untuk

kedua orang tua, keluarga,

seluruh guru, dosen dan

teman-teman seperjuangan

Terima kasih..

KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT karena atas berkat Rahmat dan karunia-Nya, penulis diberikan kemudahan dalam menyelesaikan skripsi ini. Penyusunan skripsi ini bertujuan untuk memenuhi syarat kelulusan bagi mahasiswa Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Keberhasilan penulisan skripsi ini tidak lepas dari dorongan dan bimbingan dari berbagai pihak. Untuk itu dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Prof. Dr. Abdul Haris, M.Ag selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Dr. Sri Hariani, M.Si selalu dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Bapak Dr. Cahyo Crysdian selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang senantiasa memberikan dorongan.
4. Bapak Fajar Rohman Hariri, M.Kom selaku dosen pembimbing I yang telah bersedia meluangkan waktunya dalam membimbing dan memberi arah kepada penulis sehingga dapat menyelesaikan skripsi ini.
5. Prof. Dr. Suhartono, M.Kom selaku dosen pembimbing II yang juga bersedia meluangkan waktunya dalam membimbing dan memberi arah kepada penulis sehingga dapat menyelesaikan skripsi ini.

6. Bapak dan Ibu beserta keluarga yang telah memberikan dukungan baik moral maupun spiritual sehingga penulis diberi kemudahan dalam menyelesaikan skripsi ini.
7. Seluruh dosen dan staf Jurusan Teknik Informatika yang telah memberikan ilmu dan pengalaman yang berharga.
8. Teman-teman seperjuangan UNOCORE dan teman-teman pengurus komunitas yang telah memberikan *support* dan pengalaman yang berharga.
9. Semua pihak yang telah membantu dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa masih banyak kekurangan dari laporan ini. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang terjadi selama proses penyusunan skripsi ini. Semoga tugas akhir ini dapat memberikan kontribusi yang bermanfaat bagi penulis dan pembaca khususnya.

Malang, 1 Juni 2021

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Ridho, Muhammad Ammarullah. 2021. +**Klasifikasi Pengaduan Layanan Pengguna Indihome Pada Media Sosial Twitter Menggunakan Metode *Support Vector Machine* dengan Seleksi Fitur *Information Gain*.** Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (I) Fajar Rohman Hariri, M.Kom, (II) Prof. Dr. Suhartono, M.Kom

---

Kata Kunci : *Support Vector Machine, Information Gain*, Keluhan Pengguna

Media sosial merupakan salah satu media yang digunakan untuk melakukan interaksi maupun komunikasi dalam menyampaikan suatu informasi terhadap orang lain. Salah satu media sosial yang biasa digunakan adalah Twitter. Banyak perusahaan yang memanfaatkan Twitter sebagai media layanan pelanggan. Salah satu perusahaan yang memanfaatkan Twitter dalam hal tersebut adalah PT. Telekomunikasi Indonesia (Telkom) terhadap salah satu produk mereka yaitu IndiHome dalam menanggapi pengaduan. Pengaduan yang sering dituliskan pengguna pada Twitter memiliki berbagai masalah terhadap pelayanan, tagihan, koneksi lambat dan koneksi terputus. Pada penelitian ini digunakan metode *Support Vector Machine* (SVM) dengan seleksi fitur *Information Gain*. SVM merupakan metode yang tepat dalam mengklasifikasikan keluhan tersebut kedalam lima kelas tersebut. Sedangkan *Information Gain* digunakan untuk mengurangi fitur-fitur yang tidak relevan sebelum tahap klasifikasi. Tujuan dari penelitian ini yaitu mengukur akurasi, presisi, *recall* dan *f-measure* pada klasifikasi pengaduan. Pada klasifikasi pengaduan menggunakan SVM dan seleksi fitur *Information Gain* diketahui penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 88.9%, presisi 74,11%, recall 69,22% dan f-measure 71,1% pada kelas pelayanan (C1), nilai akurasi 95%, presisi 87.2%, recall 87.8% dan f-measure 87.4% pada kelas pemasangan (C2), nilai akurasi 94.4%, presisi 85,67%, recall 86,55% dan f-measure 86,07% pada kelas tagihan (C3), nilai akurasi 93%, presisi 82,47%, recall 82,62% dan f-measure 82,48% pada kelas lambat (C4) dan nilai akurasi 89,1%, presisi 71,93%, recall 74,48% dan f-measure 73,07% pada kelas disconnect (C5). Berdasarkan penggunaan fitur pada skenario pengujian dapat diketahui bahwa persentase penggunaan fitur pada *Information Gain* mempengaruhi hasil klasifikasi sistem dengan metode SVM. Sehingga dapat disimpulkan bahwa seleksi fitur yang berlebihan akan bepengaruh pada hasil klasifikasi.

ABSTRACT

Ridho, Muhammad Ammarullah. 2021. **Indihome User Service Complaint Classification on Twitter Social Media Using Support Vector Machine Method with Information Gain Feature Selection.** Undergraduate Thesis. Informatics Engineering Department, Faculty of Science and Technology. Islamic State of Maulana Malik Ibrahim Malang.
Supervisor: (I) Fajar Rohman Hariri, M.Kom, (II) Prof. Dr. Suhartono, M.Kom

---

Keywords : *Support Vector Machine, Information Gain,* User Complaints

Social media is one of the media used to interact and communicate in informing others. One of the social media that is often used is Twitter. Many companies use Twitter as a customer service medium. One company that uses Twitter in this case, is PT. Telekomunikasi Indonesia (Telkom) to one of its products, namely IndiHome in response to his complaint. The complaints that users often write on Twitter are various service issues, bills, slow connections, and lost connections. In this study, the Support Vector Machine (SVM) method was used with the selection of the Information Gain feature. SVM is an appropriate method to classify these complaints into five classes. Meanwhile, Information Gain is used to reducing irrelevant features before the classification stage. The purpose of this study was to measure of accuracy, precision, recall, and f-measure in complaint classification. In the classification of complaints using the SVM and Information Gain features, it is known that the use of the best features is 80% which produces an accuracy value of 88.9%, precision 74.11%, recall 69.22% and f-measure 71.1% in service class (C1), 95% accuracy value, 87.2% precision, 87.8% recall and 87.4% f-measure in the installation class (C2), 94.4% accuracy value, 85.67% precision, 86.55 recall % and 86.07% f-measure in billing class (C3), 93% accuracy value, 82.47% precision, 82.62% recall and 82.48% f-measure in slow class (C4) and accuracy value. 89,1%, precision 71.93%, recall 74.48% and f-measure 73.07% in disconnect class (C5). Based on the use of features in the test scenario, it can be seen that the percentage of feature use in Information Gain affects the system classification results using the SVM method. So it can be concluded that the excessive selection of features will affect the classification results.

الملخص

رضى، محمد أمر الله. 2021. تصنيف شكاوى خدمة مستخدم IndiHome على وسائل التواصل الاجتماعي التويتر باستخدام منهج Support Vector Machine مع اختيار ملامح Information Gain. البحث الجامعي. قسم الهندسة المعلوماتية. كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج.
المشريف الأول: (1) فجر رحمان حريري الماجستير (2) البروفيسور الدكتور سوهارتونو الماجستير.

الكلمة الرئيسية: Support Vector Machine، Information Gain، شكاوى المستخدم.

وسائل التواصل الاجتماعي هي إحدى الوسائل المستخدمة للتعامل والتواصل في نقل المعلومات إلى الآخرين. عادةً، إحدي وسائل التواصل الاجتماعي المستخدمة هي التويتر. تستفيد كثير من الشركات التويتر كوسيلة خدمة العميل. إحدى الشركات التي تستفيد التويتر في هذا الأمر هي PT. Telekomunikasi Indonesia (Telkom) إلى إحدى منتجاتها، وهي IndiHome ردًا على الشكاوى. الشكاوى التي عادةً ما يكتبها المستخدم على التويتر تشمل مشكلات مختلفة مع الخدمة وفاتورة والاتصالة البطيئة وفقدان الاتصالة. في هذا البحث، تُستخدم منهج Support Vector Machine (SVM) مع اختيار ملامح Information Gain. SVM هي طريقة مناسبة لتصنيف هذه الشكاوى إلى خمس فئات. وأمّا يتم Information Gain تُستخدم لتقليل الملامح الخارج عن الموضوع قبل مرحلة التصنيف. الأهداف من هذا البحث هي أحصى قيمة الدقة والاحكام و اعد الاتصال و قياس في تصنيف الشكاوى باستخدام ميزات SVM و Information Gain ، من المعروف أن استخدام أفضل الميزات هو 80٪ مما ينتج دقة تبلغ 88.9٪ ، ودقة 74.11٪ ، واستدعاء 69.22٪ ، و وقياس 71.1٪ في الخدمة فئة (C1) ، قيمة دقة 95٪ ، دقة 87.2٪ ، استدعاء 87.8٪ و 87.4٪ قياس f في فئة التثبيت (C2) ، قيمة دقة 94.4٪ ، دقة 85.67٪ ، استرجاع 86.55٪ و 86.07٪ قياس f في فئة الفوترة (C3) ، قيمة دقة 93٪ ، دقة 82.47٪ ، تذكر 82.62٪ و 82.48٪ قياس f في الفئة البطيئة (C4) و قيمة دقة ، 89.1٪ ، دقة 71.93٪ ، استدعاء 74.48٪ وقياس 73.07٪ في فئة الفصل (C5). استنادًا إلى استخدام الملامح في سيناريو الاختبار، يُعرف أن النسبة المئوية لاستخدام ملامح Information Gain تؤثر على نتائج تصنيف النظام باستخدام منهج SVM. لذلك يمكن استنتاج أن الاختيار المفرط للملامح سيؤثر على نتائج التصنيف.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial merupakan salah satu media yang digunakan untuk melakukan interaksi maupun komunikasi. Dengan adanya media sosial tentunya akan memudahkan manusia dalam menyampaikan suatu informasi terhadap orang lain. Salah satu media sosial yang biasa digunakan adalah Twitter. Pada bulan Juli 2020 jumlah pengguna Twitter di Indonesia menempati posisi kedelapan diatas Thailand, Filipina dan Meksiko yaitu tercatat sebanyak 11,2 juta pengguna aktif (Statista, 2020). Twitter telah memfasilitasi pengguna untuk menyampaikan pendapat mereka dalam bentuk *tweet.* Banyak perusahaan yang memanfaatkan Twitter sebagai media untuk melakukan interaksi terhadap konsumen atau biasa disebut layanan pelanggan.

Salah satu perusahaan yang memanfaatkan Twitter dalam hal tersebut adalah PT. Telekomunikasi Indonesia (Telkom) sebagai media layanan pelanggan terhadap salah satu produk mereka yaitu IndiHome melalui akun Twitter dengan nama akun @IndiHome dan @IndiHomeCare. IndiHome merupakan layanan digital yang menyediakan internet rumah, telepon rumah dan TV Interaktif (IndiHome, 2020).

Dengan adanya layanan pelanggan terhadap produk IndiHome, para pengguna IndiHome memanfaatkan layanan tersebut untuk memberikan pengaduan terhadap penggunaan produk IndiHome. Para pengguna

menuliskan pengaduan mereka dengan cara menyebut atau *mention* ke akun @IndiHome. Pengaduan yang sering dituliskan pengguna pada Twitter memiliki berbagai masalah terhadap pelayanan, tagihan, koneksi lambat dan koneksi terputus. Dengan berbagai jenis pengaduan tersebut diperlukan proses klasifikasi menjadi sebuah kategori berdasarkan topik yang disampaikan (Ibtihel *et al*, 2018).

Dalam Al-Qur'an telah dijelaskan tentang pemberian amanah dan menempatkan sesuatu pada tempatnya atau berperilaku adil, seperti firman Allah SWT dalam Surah An Nisa ayat 58 :

اِنَّ اللّٰهَ يَأْمُرُكُمْ اَنْ تُؤَدُّوا الْاَمٰنٰتِ اِلٰٓى اَهْلِهَاۙ وَاِذَا حَكَمْتُمْ بَيْنَ النَّاسِ اَنْ تَحْكُمُوْا بِالْعَدْلِۗ اِنَّ اللّٰهَ نِعِمَّا يَعِظُكُمْ بِهٖۗ اِنَّ اللّٰهَ كَانَ سَمِيْعًاۢ بَصِيْرًا (٥٨)

Artinya : "Sungguh, Allah menyuruhmu menyampaikan amanat kepada yang berhak menerimanya, dan apabila kamu menetapkan hukum diantara manusia hendaknya kamu menetapkannya dengan adil. Sungguh, Allah sebaik-baik yang memberi pengajaran kepadamu. Sungguh, Allah Maha Mendengar, Maha Melihat" (Q.S. An-Nisa : 58).

Menurut tafsir Ibnu Katsir jilid 2, ayat ini tentang perintah untuk menunaikan amanat. Salah satu amanat yang harus dijalankan yaitu berupa hak-hak manusia seperti titipan. Dengan adanya pengaduan dari para pengguna tentunya dapat segera ditindaklanjuti oleh pihak IndiHome dalam menangani pengaduan pengguna.

Dalam melakukan proses klasifikasi, terdapat beberapa metode diantaranya *Naïve Bayes, Decision Tree, Artificial Neural Network* (ANN), *Nearest Neighbour, Support Vector Machine* (SVM), dan lain sebagainya.

Pada penelitian ini digunakan metode *Support Vector Machine* (SVM) dengan seleksi fitur *Information Gain*. SVM merupakan metode yang tepat dalam mencari pemisah antar kelas atau *hyperplane*. SVM berusaha menemukan *hyperplane* dengan memaksimalkan jarak antar kelas (Suyanto, 2019).

SVM melakukan klasifikasi dengan membangun ke dalam ruang dimensi yang lebih tinggi dan secara optimal memisahkan data ke dalam kategori menggunakan *hyperplane* (Adankon dan Cheriet, 2015).

Penggunaan seleksi fitur *Information Gain* digunakan untuk mengurangi fitur-fitur yang tidak relevan sebelum tahap klasifikasi pengaduan menggunakan *Support Vector Machine.*

1.2 Pernyataan Masalah

Seberapa tinggi nilai akurasi, presisi, *recall* dan *f-measure* pada sistem klasifikasi pengaduan layanan pengguna IndiHome pada media sosial Twitter menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain.*

1.3 Tujuan Penelitian

Mengukur akurasi, presisi, *recall* dan *f-measure* pada sistem klasifikasi pengaduan layanan pengguna IndiHome pada media sosial Twitter

menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain*.

1.4 Batasan Masalah

Batasan masalah penelitian adalah sebagai berikut:

1. Data yang digunakan dalam penelitian merupakan data primer. Data primer didapatkan dari Twitter yang menggunakan *mention* pada akun @IndiHome dan @IndiHomeCare. Data primer tersebut berfungsi sebagai masukan sebagai pola model yang dibentuk untuk mendapatkan hasil klasifikasi dan melakukan prediksi pada tahap selanjutnya.
2. Data *tweets* akan diklasifikasikan menjadi lima kelas yaitu pelayanan, pemasangan, tagihan, koneksi lambat dan koneksi terputus.

1.5 Manfaat Penelitian

Dari penelitian ini diharapkan mendapatkan hasil nilai akurasi, presisi, *recall*, dan *f-measure* menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain.* Sehingga penelitian ini diharapkan dapat memberikan manfaat bagi beberapa pihak, diantaranya adalah :

1. Pihak IndiHome, untuk memudahkan dalam menanggapi berbagai pengaduan layanan IndiHome.
2. Peneliti *Data Mining*, untuk rujukan dalam penelitian selanjutnya.

1.6 Sistematika Penulisan

Untuk memahami isi dalam penelitian ini, maka laporan ini dikelompokkan menjadi beberapa bab pembahasan sebagai berikut :

1. Bab I Pendahuluan

   Bab ini berisi tentang latar belakang penelitian, pernyataan masalah, tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penulisan.

2. Bab II Tinjauan Pustaka

   Bab ini berisi tentang pembahasan teori tentang proses *Preprocessing data*, algoritma *Support Vector Machine*, algoritma *Information Gain* dan penelitian sebelumnya.

3. Bab III Metodologi Penelitian

   Bab ini berisi tentang proses akuisisi data dan perancangan sistem

4. Bab IV Hasil dan Pembahasan

   Bab ini berisi tentang hasil dari proses klasifikasi teks menggunakan algoritma *Support Vector Machine* dan seleksi fitur *Information Gain.*

5. Bab V Kesimpulan dan Saran

   Bab ini berisi tentang kesimpulan dan saran sebagai pengembangan terhadap penelitian selanjutnya.

6. Daftar Pustaka

   Daftar pustaka berisi tentang data referensi penelitian terkait yang dirujuk pada penelitian ini.

BAB II

STUDI PUSTAKA

2.1 *Text Mining*

Data merupakan sebuah informasi yang perlu dikelola untuk pengetahuan mendatang. Salloum *et al* (2018) melakukan penelitian tentang *Text Mining* dalam penemuan pola pada artikel penelitian yang diterbitkan oleh Sciencedirect, IEEE, Wiley, Cambridge, SAGE dan Springer. Artikel dari enam sumber tersebut dikelola dan dianalisis secara tekstual untuk pengelompokan artikel penelitian bertemakan studi kesehatan. Pada penelitian ini dihasilkan bahwa artikel penelitian yang membahas tentang studi kesehatan yaitu Springer.

Jenis penelitian yang sama juga dilakukan oleh (Haq *et al,* 2019). *Text Mining* digunakan dalam penemuan pola artikel penelitian yang diterbitkan oleh IEEE, Springer, Wiley, Elsevier dan ACM. Kata kunci yang digunakan dalam penelitian ini yaitu tentang *Big Data*, *Cloud*, *Computing*, *Service*, *Time, Application* dan *Security*. Dengan *text mining* dihasilkan bahwa sebagian besar publikasi berasal dari Asia, Eropa, Amerika Utara, Australia, dan beberapa dari Afrika.

Text Mining merupakan cabang khusus *data mining*. Tujuan utama *data mining* adalah untuk mendapatkan teks dalam bentuk yang dapat dimengerti komputer secara langsung sehingga dapat diproses tanpa campur tangan manusia. *Data mining* bekerja dengan data terstruktur seperti database, gudang data, data belanja online, data penggunaan ponsel, dll. *Text Mining*

bekerja dengan data bahasa alami yang tidak terstruktur atau semi terstruktur. Contoh dataset untuk *text mining* adalah data yang dihasilkan oleh media sosial, yang merupakan data tidak terstruktur bahasa alami (Oza dan Naik, 2016).

Text Mining merupakan metode yang mengacu pada pengambilan informasi, *data mining*, *machine learning*, statistik, dan linguistik komputasi. Sebagian besar informasi disimpan sebagai teks seperti artikel berita, makalah teknis, buku, perpustakaan digital, pesan *email*, *blog*, dan halaman *web* (Han *et al*, 2012).

Text Mining pada dasarnya memerlukan pendekatan kuantitatif untuk analisis data tekstual yang banyak. Hal ini dapat membantu mempercepat penemuan pengetahuan dengan meningkatkan jumlah data yang dapat dianalisis. Teknik analisis yang digunakan pada *Text Mining* diantaranya tentang *dimensionality reduction, distance and similarity computing*, *clustering*, pemodelan topik, dan klasifikasi (Kobayashi *et al*, 2018).

2.2 *Text Preprocessing*

Preprocessing merupakan salah satu komponen yang penting dalam melakukan klasifikasi teks. Dalam hal ini tahap *preprocessing* sangat mempengaruhi tingkat akurasi pada proses klasifikasi. Pada penelitian ini terdapat empat langkah dalam melakukan *text preprocessing* yaitu *tokenizing*, stopword removal, *lowercase conversion*, dan *stemming* (Uysal dan Gunal, 2014).

Tahap *text preprocessing* mengambil input teks mentah dan mengembalikan token yang telah dibersihkan. Token merupakan kata tunggal atau kelompok kata yang dihitung berdasarkan frekuensinya dan berfungsi sebagai fitur analisis (Anandarajan *et al,* 2019).Penggunaan *text preprocessing* juga dilakukan oleh Khomsah dan Ariwibowo (2020) pada analisis sentimen komentar YouTube berbahasa Indonesia, penggunaan *text preprocessing* meningkatkan akurasi cukup signifikan sebesar 3% sampai 3,5%.

Text preprocessing diperlukan dalam klasifikasi teks. Dalam proses klasifikasi teks terdapat beberapa langkah yang berurutan, yaitu persiapan data pelatihan, *preprocessing*, transformasi, penerapan teknik klasifikasi, dan validasi (Kobayashi *et al,* 2018). Tahapan *text preprocessing* juga diterapkan sebelum melakukan tahapan klasifikasi teks berita berbahasa Indonesia. Hal ini dilakukan karena dengan adanya *text preprocessing* maka akan meminimalisir *noise* pada dokumen yang digunakan (Wongso et al, 2017). Fungsi lain dari *text preprocessing* yaitu mengurangi dan membersihkan kata maupun karakter yang tidak diperlukan dalam proses klasifikasi teks (Mohammad, 2018)

2.3 TF-IDF

TF-IDF adalah skema pembobotan term yang paling populer dalam temu kembali informasi. TF-IDF merupakan kombinasi dari *Term Frequency* dan *Inverse Document Frequency* (Wongso *et al,* 2017). TF-IDF (*Term Frequency-Inverse Document Frequency*) telah digunakan untuk

mengukur pentingnya istilah untuk dokumen dalam kumpulan teks atau *corpus*. Ini juga telah ditemukan sebagai skema pembobotan istilah yang efektif dalam klasifikasi teks (Samant *et al*, 2019).

Algoritma TF-IDF menghitung nilai pada setiap *term* untuk mengekstrak istilah. Pada penelitian Zhu *et al* (2019) penggunaan TF-IDF pada penentuan topik berita terpopuler telah menghasilkan akurasi sebesar 80%. Sedangkan tingkat akurasi tanpa menggunakan TF-IDF hanya sebesar 72%.

Dalam penelitian Yahaf *et al* (2018), algoritma TF-IDF digunakan dalam pencarian informasi maupun *text mining*. TF-IDF digunakan pada analisis sentimen berbasis teks dalam pembobotan kata pada komentar sebuah postingan Facebook.

Penggunaan TF-IDF untuk pembobotan kata pada klasifikasi teks pada dokumen Bangladesh yang dikombinasikan dengan metode *Support Vector Machine* telah menghasilkan tingkat akurasi sebesar 92,57% (Islam *et al*, 2017). Pembobotan TF-IDF akan menghasilkan skor pada setiap kata. Dari skor tersebut dapat diurutkan secara *ascending* ataupun *descending*. Skor kata yang terbesar menunjukkan bahwa kata tersebut sering muncul pada dokumen tersebut. Dengan menggunakan TF-IDF maka akan diketahui kesesuaian kata terhadap dokumen tersebut (Qaiser dan Ali, 2018).

TF-IDF diterapkan dalam sistem klasifikasi artikel *hoax* sebagai vektorisasi teks. Dengan menggunakan algoritma klasifikasi *Support Vector Machine*, TF-IDF berpengaruh dalam proses konversi fitur menjadi sebuah

nilai. Pada penelitian ini menghasilkan akurasi sebesar 95.8333% (Maulina dan Sagara, 2018).

Kombinasi algoritma *Support Vector Machine* dengan TF-IDF juga digunakan dalam penelitian yang dilakukan oleh Fitriyah *et al* (2020). Penelitian ini membahas tentang analisis sentimen terhadap aplikasi ojek online pada media sosial *Twitter*. Hasil analisis sentimen ini menghasilkan akurasi sebesar 79,19%.

2.4 Seleksi Fitur

Kita sekarang berada di era big data, di mana sejumlah besar data berdimensi tinggi tersebar dimana-mana. Analisis data berdimensi tinggi merupakan tantangan bagi para peneliti dan insinyur di bidang pembelajaran mesin dan penggalian data. Seleksi fitur memberikan cara yang efektif untuk menyelesaikan masalah ini dengan menghapus data yang tidak relevan dan berlebihan (Cai *et al*, 2018).

Seleksi fitur dalam *preprocessing* data dapat mengurangi dimensi data. Hal ini penting diterapkan pada aplikasi *data mining* dan *machine learning* (Li *et al*, 2017). Akurasi klasifikasi sangat bergantung pada sifat fitur dalam suatu kumpulan data yang kemungkinan berisi data yang tidak relevan atau berlebihan. Tujuan utama seleksi fitur adalah untuk menghilangkan jenis fitur yang tidak relevan guna meningkatkan akurasi klasifikasi (Mafarja dan Mirjalili, 2018)

Sejumlah besar data meningkat di berbagai bidang seperti media sosial, bioinformatika, dan perawatan kesehatan. Data ini mengandung data yang

berlebihan dan tidak relevan. Seleksi fitur umumnya digunakan sebagai teknik *data mining* sebagai pengontrol inputan untuk pemrosesan dan analisis. Seleksi fitur juga digunakan untuk *dimension reduction, machine learning,* dan aplikasi *data mining* lainnya (Colaco *et al*, 2019).

2.5 *Information Gain*

Pada penelitian Daeli dan Adiwijaya (2020) tentang analisis sentimen terhadap penilaian film menggunakan *Information Gain* dan *K-Nearest Neighbor* (KNN). Penggunaan seleksi fitur *Information Gain* pada penelitian ini menghasilkan akurasi sebesar 96,8% dibandingkan tanpa menggunakan seleksi fitur yang hanya sebesar 60%.

Penggunaan seleksi fitur *Information Gain* juga diterapkan pada penelitian tentang analisis sentimen terhadap ulasan film menggunakan metode *Naïve Bayes*. Hasil pengujian penelitian ini menghasilkan nilai akurasi sebesar 82.19% (Sihwi *et al,* 2018).

Selain itu penggunaan *Information* pada penelitian klasifikasi penyakit jantung menggunakan *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Pada penelitian ini menghasilkan tingkat akurasi sebesar 92.31% (Aini *et al*, 2018).

Information Gain metode peringkat atribut paling sederhana, banyak digunakan dalam aplikasi kategorisasi teks dan sekarang digunakan untuk analisis data *microarray* dan analisis data gambar (Chormunge dan Jena, 2016).

Dalam melakukan perhitungan *Information Gain* perlu dicari nilai *Information Entropy*. *Entropy* merupakan suatu parameter yang digunakan untuk mengukur keberagaman informasi yang ada dalam sebuah variabel. Semakin data tersebut beragam maka akan semakin besar nilai *entropy* yang dihasilkan (Suyanto, 2019).

2.6 *K-fold Cross Validation*

K-fold Cross Validation merupakan pendekatan populer dalam mengevaluasi kinerja algoritma klasifikasi. Terdapat beberapa faktor yang mempengaruhi *k-fold cross validation* meliputi jumlah *fold*, jumlah data dalam *fold*, tingkat rata-rata dan pengulangan *cross validation* (Wong, 2015).

K-fold Cross Validation merupakan sebuah metode pembagian dataset menjadi data latih dan data uji secara efektif sebanyak *k-equal.* Penggunaan *k-fold cross validation* pada metode *Support Vector Machine* telah berfungsi dengan baik dalam meningkatkan akurasi model (Asrol *et al*, 2021).

Penggunaan *k-fold cross validation* telah berhasil mengevaluasi kemampuan klasifikasi data kanker menggunakan *Support Vector Machine* dan *Information Gain*, dengan jumlah *fold* sebanyak 10. Pada penelitian ini menghasilkan akurasi terbaik sebesar 90.32% (Gao *et al*, 2017). Validasi *k-fold cross* juga telah diterapkan pada pengenalan ekspresi wajah menggunakan metode *Support Vector Machine* (Gautam *et al*, 2017).

Menurut Marcot dan Hanea (2020) banyak literatur dalam menerapkan *k-fold cross validation* menggunakan k = 10, akan tetapi dengan

penggunaan k = 5 pada *k-fold cross validation* akan menghemat waktu dan komputasi.

2.7 *Support Vector Machine*

Support Vector Machine (SVM) merupakan sebuah metode *supervised learning* yang menganalisis data dan mengenali pola. SVM biasanya digunakan untuk proses klasifikasi, analisis regresi dan prediksi. SVM memberikan pemisah antar kelas atau *hyperplane* dengan cara memaksimalkan jarak masing-masing kelas atau biasa disebut *functional margin* (Awad dan Khanna, 2015).

Support Vector Machine dapat digunakan pada tema analisis sentimen. Pada penelitian Xia *et al* (2020), tentang analisis sentimen terhadap ulasan di Internet. Dari penerapan metode *Support Vector Machine* terhadap penelitian tersebut menghasilkan tingkat akurasi sebesar 90%.

Berdasarkan penelitian yang dilakukan oleh Somantri dan Apriliani (2018) tentang analisis sentimen terhadap kepuasan terhadap pelayanan rumah makan menggunakan metode SVM dan seleksi fitur *Information Gain* menghasilkan akurasi terbaik sebesar 72,45%. Pada penelitian ini menggunakan konsep klasifikasi biner yang mengkategorikan terhadap dua *output*.

Pada penelitian yang dilakukan oleh Fatmawati dan Affandes (2017) tentang klasifikasi keluhan terhadap sistem informasi akademik menggunakan metode *multiclass* SVM. Dataset yang digunakan sebanyak

1040 data keluhan yang diklasifikasikan ke empat kategori. Berdasarkan penelitian ini dihasilkan tingkat akurasi sebesar 95.67% pada c=2, γ=0.09.

Vidya *et al*, (2015) melakukan penelitian tentang analisis reputasi provider telepon genggam menggunakan metode *Support Vector Machine* (SVM). Data yang diambil berasal dari *tweet* pengguna Twitter. Pada penelitian ini dihasilkan tingkat presisi sebesar 86,26%, *recall* 92,62% dan *f-measure* sebesar 89,3%.

Pratama dan Trilaksono (2017) melakukan sebuah penelitian tentang klasifikasi keluhan pelanggan dengan seleksi fitur dan metode SVM dengan empat kategorisasi. Pada penelitian ini digunakan kernel *non-linear* yang diterapkan pada metode SVM dan dikombinasikan dengan berbagai macam seleksi fitur. Hasil penerapan metode SVM dan seleksi fitur *Information Gain* menghasilkan akurasi paling baik sebesar 85%.

Penelitian tentang keluhan pelanggan terhadap layanan IndiHome juga pernah dilakukan oleh Tineges *et* al (2020) menggunakan *Support Vector Machine* dengan jenis klasifikasi *biner.* Pada penelitian ini diklasifikasikan keluhan pengguna IndiHome ke dua kategori yaitu positif dan negatif. Berdasarkan hasil pengujian didapatkan nilai akurasi sebesar 87%, presisi 86%, *recall* 95%, *error rate* 13%, dan *f1-score* 90%

Pelatihan SVM terdiri dari penentuan *hyperplane* untuk memisahkan data pelatihan milik dua kelas. Meskipun *hyperplane* memisahkan data secara linear, SVM dapat diterapkan pada masalah *non-linear*. Hal ini dikarenakan adanya pemetaan menggunakan fungsi kernel ke dalam ruang

berdimensi tinggi sehingga dapat dipisahkan secara *linear* (Nalepa dan Kawulok, 2018).

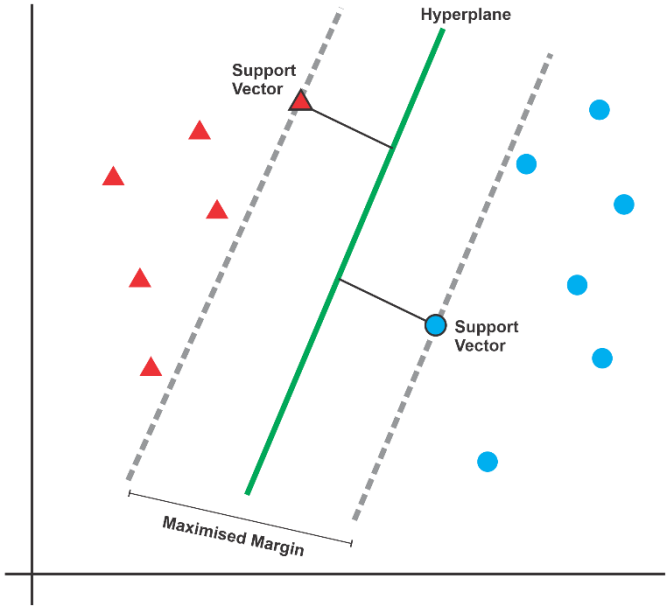
Menurut Simske (2019) SVM dirancang untuk memberikan pendekatan yang dapat disesuaikan sebagai batas keputusan antara dua kelas. Margin merupakan wilayah terbesar yang dapat digunakan untuk memisahkan dua kelas. Titik data atau *dot product* yang paling dekat dengan margin merupakan hal terpenting dalam menemukan *support vector*. *Kernel trick* digunakan untuk secara efektif menambahkan dimensi ekstra ke margin dengan memasukkan matriks kernel. Salah satu kernel yang digunakan yaitu kernel linear, yang diformulasikan sebagai berikut :

$$K(x, y) = x_k \times x \tag{2.1}$$

Saat metode SVM pertama kali dikenalkan, metode ini hanya dapat mengklasifikasikan data ke dalam dua kelas. Pengklasifikasian ke dalam dua kelas bisa disebut dengan pengklasifikasi biner (*binary classifier*). Namun seiring penelitian lebih lanjut, klasifikasi SVM mampu mengklasifikasi data ke dalam lebih dari dua kelas atau bisa disebut multi kelas (*multiclass*). Proses klasifikasi data pada SVM biasanya diformulasikan sebagai berikut :

$$f(x) = wx + b \tag{2.2}$$

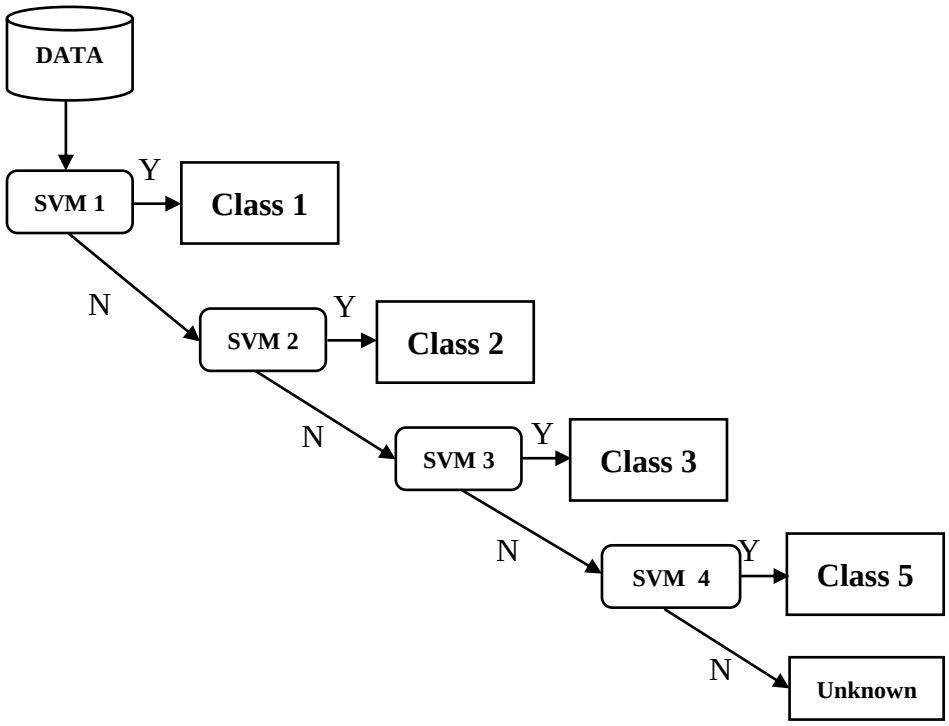
Ada beberapa pendekatan yang digunakan pada klasifikasi multi kelas diantaranya *one-against-all*, *one-against-one*, dan *Directed Acyclic Graph Support Vector Machine* atau DAGSVM (Suyanto, 2019).



Gambar 2.1 Plot *Support Vector Machine* (Suyanto, 2019)

1. *One-against-all*

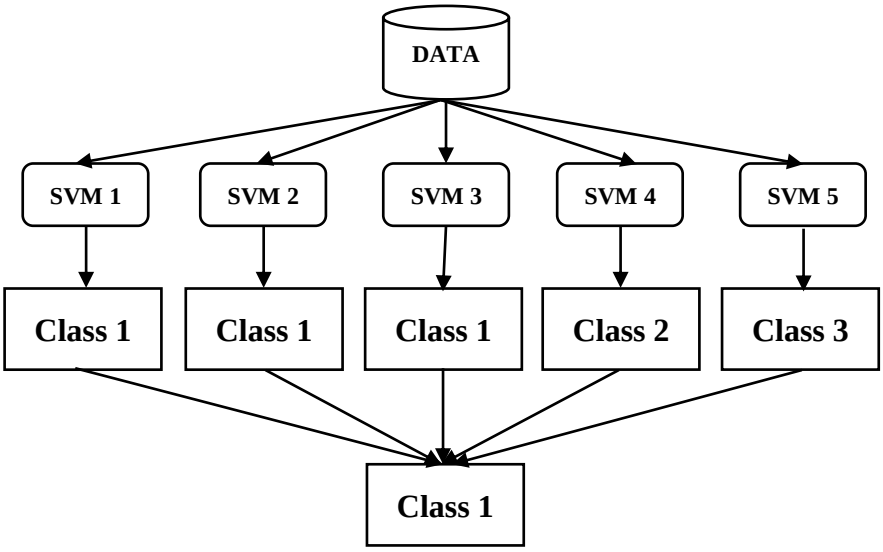
   Penggunaan pendekatan o*ne-against-all* dengan cara membandingkan satu kelas dengan seluruh kelas. Setiap model SVM biner ke-*i* akan dilatih menggunakan keseluruhan data untuk mengetahui hasil klasifikasi.



Gambar 2.2 Struktur *One-against-all* (Behrad *et al,* 2010)

2. *One-against-one*

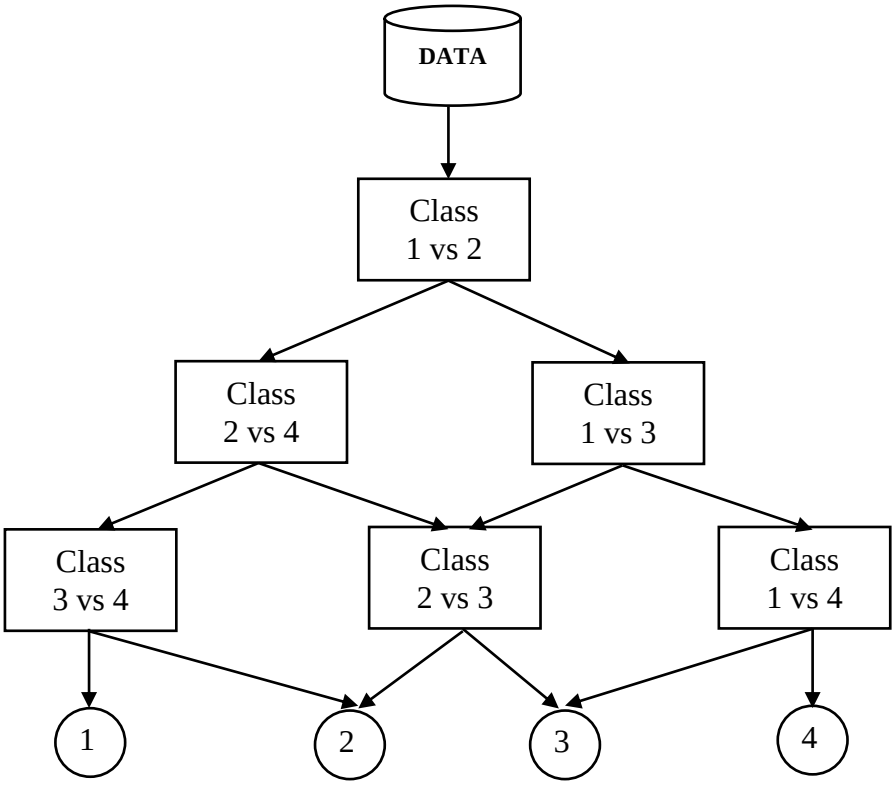
Penggunaan pendekatan o*ne-against-one* dengan cara membandingkan satu kelas dengan satu kelas lainnya. Jika dalam proses membandingkan terdapat kelas yang memenuhi kualifikasi maka kelas tersebut merupakan hasil dari klasifikasi tersebut.



Gambar 2.3 Struktur *One-against-one* (Behrad *et al,* 2014)

3. *Directed Acrylic Graph Support Vector Machine* (DAGSVM)

   Penggunaan pendekatan *Directed Acrylic Graph Support Vector Machine* (DAGSVM) hampir mirip dengan pendekatan *one-against-one*, akan tetapi DAGSVM tidak menggunakan kompetisi penuh dalam melakukan klasifikasi. DAGSVM menggunakan sistem gugur dalam proses membandingkan antar kelas.



Gambar 2.4 Struktur *DAGSVM* (Fan *et al,* 2017)

BAB III

METODOLOGI PENELITIAN

3.1 Akuisisi Data

Data yang digunakan dari penelitian ini merupakan data sekunder yang didapatkan dari platform media sosial Twitter. Data sekunder merupakan sebuah dataset yang digunakan sebagai data masukan untuk membangun sebuah model perhitungan dan digunakan sebagai data uji dalam melakukan proses klasifikasi sehingga mendapatkan hasil prediksi penelitian. Dengan kata lain pelatihan maupun pengujian didapatkan dari data sekunder berupa tweet yang berasal dari Twitter.

Dalam tahap pengumpulan data tweet menggunakan teknik *web crawling* terhadap pengaduan layanan pengguna pada akun Twitter @IndiHome. Dalam tahap web scraping dapat dikumpulkan tweet dalam rentang waktu yang diinginkan. Teknik tersebut sangat cocok dalam melakukan ekstraksi data ataupun informasi dari sebuah *website* dan menyimpannya dalam format tertentu.

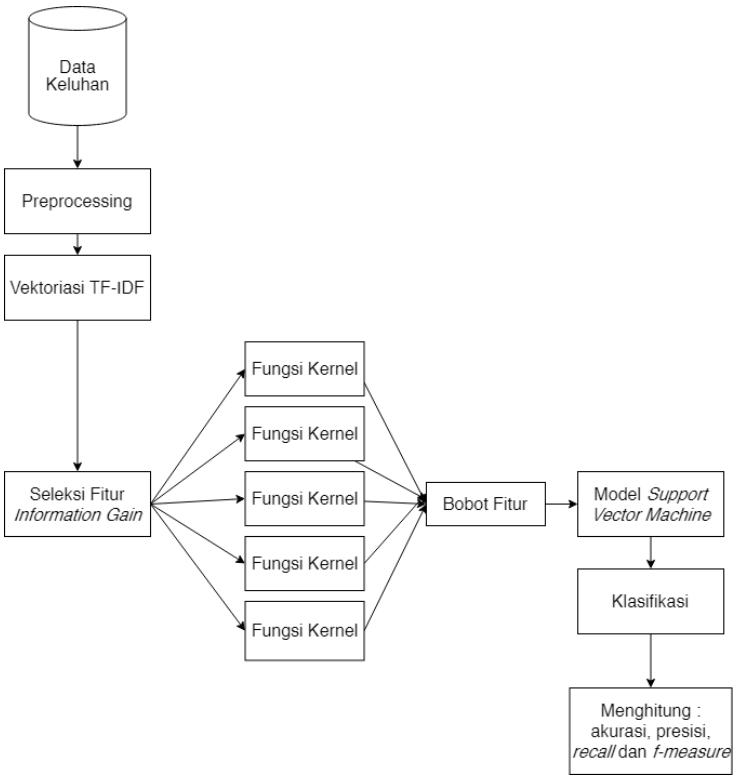
Data penelitian diambil menggunakan teknik *web scraping* yang terkumpul sebanyak 1000 *tweets* yang akan tersimpan dalam database. Data akan dibagi menjadi data latih dan data uji. Data yang terkumpul akan diberikan 5 label yaitu tentang pelayanan (C1), pemasangan (C2), tagihan (C3), koneksi lambat (C4) dan koneksi terputus (C5).

Tabel 3.1 Pelabelan Data

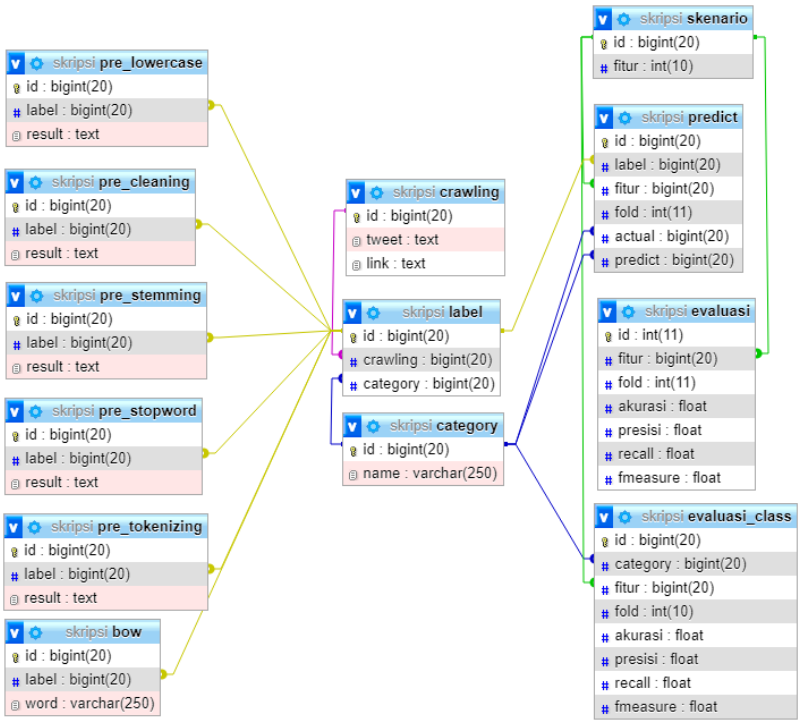
Dokumen	Tweet	Kelas/Label
D1	kurang nyaman jawaban customer service Indihome	Pelayanan (C1)
D2	sudah telpon internet saya tidak segera dipasang	Pemasangan (C2)
D3	tagihan saya lebih mahal dari tetangga padahal paketnya sama	Tagihan (C3)
D4	internet saya lemot banget	Lambat (C4)
D5	internet rumah saya mati seminggu	Disconnect (C5)

3.2 Perancangan Sistem

Pada sub-bab ini peneliti membahas tentang perancangan sistem penelitian seperti tercantum pada Gambar 3.1 dan perancangan *database* pada Gambar 3.2 sebagai berikut.



Gambar 3.1 Blok Diagram Perancangan Sistem



Gambar 3.2 Perancangan Database

Berdasarkan blok diagram perancangan sistem pada Gambar 3.1, diketahui input berupa data keluhan yang akan diproses pada beberapa tahap seperti *preprocessing*, TF-IDF, seleksi fitur *Information Gain* dan akan dibentuk model dari *Support Vector Machine*. Dari model tersebut akan dilakukan evaluasi berupa nilai akurasi, presisi, *recall* dan *f-measure* sebagai *output* dari penelitian ini. Berdasarkan perancangan sistem tersebut terdapat beberapa poin perbedaan dengan penelitian sebelumnya. Pada tabel 3.2 merupakan perbedaan dengan penelitian sebelumnya.

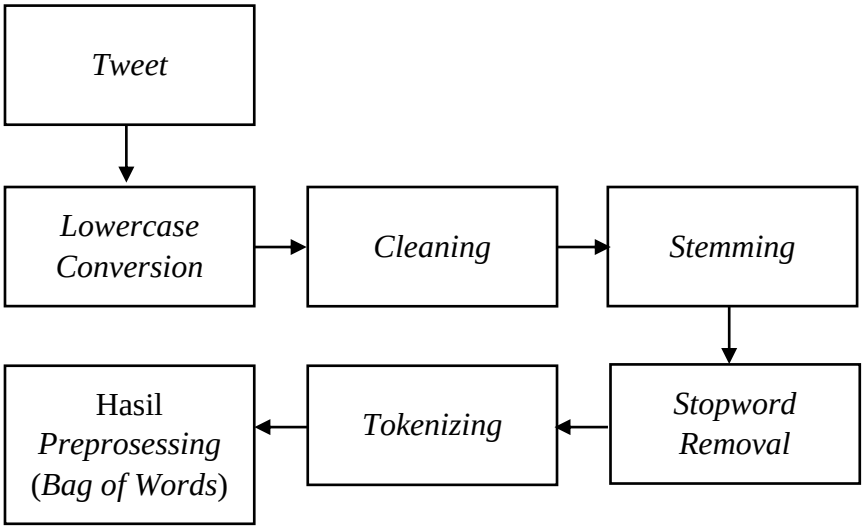
Tabel 3.2 Perbedaan dengan Penelitian Sebelumnya

No.	Pengarang (Tahun)	Detail	Perbedaan
1.	Vidya *et al* (2015)	a. Menggunakan tiga algoritma, yaitu *Support Vector Machine, Naïve Bayes* dan *Decision Tree* b. Menggunakan seleksi fitur *part-of-speech* c. Mengelompokkan pada dua kelas yaitu positif dan negatif d. Mengukur presisi, *recall* dan *f-measure*	a. Menggunakan satu algoritma yaitu *Support Vector Machine* b. Menggunakan seleksi fitur *Information Gain* c. Mengelompokkan pada lima kelas yaitu pelayanan, pemasangan, tagihan, koneksi lambat dan koneksi terputus d. Mengukur akurasi, presisi, *recall* dan *f-measure* pada klasifikasi keluhan pelanggan
2.	Pratama dan Trilaksono (2017)	a. Menggunakan kernel RBF pada *Support Vector Machine* b. Menggunakan tiga algoritma TF-IDF, *Information Gain* dan *Chi-Square* sebagai ekstraksi fitur c. Mengelompokkan pada empat kelas yaitu *billing*, pemasangan, *disconnect* dan lambat	a. Menggunakan kernel linear pada *Support Vector Machine* b. Menggunakan algoritma TF-IDF sebagai vektorisasi kata dan *Information Gain* sebagai seleksi fitur c. Mengelompokkan pada lima kelas yaitu pelayanan, pemasangan, tagihan, koneksi lambat dan koneksi terputus
3.	Tineges *et al* (2020)	a. Mengelompokkan pada dua kelas yaitu positif dan negatif	a. Mengelompokkan pada lima kelas yaitu pelayanan, pemasangan, tagihan, koneksi

		b. Tidak menggunakan seleksi fitur	lambat dan koneksi terputus b. Menggunakan *Information Gain* untuk seleksi fitur

1. *Preprocessing Data*

Pada tahap *preprocessing data* pada Gambar 3.2 terdapat input berupa kalimat *tweet* yang akan dilakukan beberapa proses sehingga dihasilkan kumpulan kata atau *bag of words.* Dalam penelitian ini terdapat beberapa tahapan dalam melakukan *Text Preprocessing Data*, diantaranya sebagai berikut :



Gambar 3.3 Blok Diagram *Preprocessing Data*

a. *Lowercase Conversion*

Lowercase Conversion merupakan sebuah tahap awal dalam melakukan t*ext preprocessing data* yaitu dengan cara mengubah kata yang terdiri huruf kapital menjadi huruf kecil.

```python
def lowercase(tweet):
    tweet = tweet.lower()
    return tweet
```

Gambar 3.4 Implementasi *Lowercase Conversion*

b. *Cleaning*

Cleaning merupakan sebuah tahap penghapusan data yang tidak diperlukan seperti tautan atau *link*, tagar atau *hashtag* dan *mention* pengguna Twitter.

```python
import re

def cleaning(tweet):
    #Remove URLs dan Karakter
    tweet = re.sub(r'(?i)\b((?:https?://|www\d{0,3}
[.]|[a-z0-9.\-]+[.][a-z]{2,4}/)(?:[^\s()<>]+|\
(([^\s()<>]+|(\([^\s()<>]+\)))*\))+(?:\(([^\s()<>]+|
(\([^\s()<>]+\)))*\)|[^\s`!()\[\]{};:\'".,
<>?«»“”‘’]))', '', tweet)
    #Remove mention
    tweet  = re.sub("@[A-Za-z0-9]+","", tweet)
    #Remove punctuations
    tweet = re.sub(r'[^\w]|_',' ',tweet)
    #Remove digit from string
    tweet = re.sub("\S*\d\S*", "", tweet).strip()
    #Remove digit or numbers
    tweet = re.sub(r"\b\d+\b", " ",tweet)
    #Remove white spaces
    tweet = re.sub('[\s]+', ' ', tweet)

    return tweet
```

Gambar 3.5 Implementasi *Cleaning*

c. *Stemming*

Stemming merupakan sebuah tahap untuk mencari kata asli yang berasal dari kata turunan yang terdapat dalam kalimat. Pada tahap ini akan dihilangkan *sufiks, prefiks* dan *konfiks* yang terdapat pada kata

berbahasa Indonesia. Pada tahap ini menggunakan *library* Python Sastrawi yang dapat diakses pada link berikut: *https://github.com/har07/pystastrawi-demo*.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

stem_factory = StemmerFactory()
stemmer = stem_factory.create_stemmer()

def stemming(tweet):
    tweet = stemmer.stem(tweet);
    return tweet
```

Gambar 3.6 Implementasi *Stemming*

d. *Stopword Removal*

Stopword adalah kata-kata yang biasa ditemui dalam teks tanpa ketergantungan pada topik tertentu seperti konjungsi, preposisi, artikel, dll. Oleh karena itu, kata-kata penghenti biasanya diasumsikan tidak relevan dalam klasifikasi teks, dan dihapus sebelum klasifikasi. Pada tahap ini menggunakan *library* Python Sastrawi yang dapat diakses pada link berikut: *https://github.com/har07/pystastrawi-demo*.

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import
StopWordRemoverFactory

stop_factory = StopWordRemoverFactory()
stopword = stop_factory.create_stop_word_remover()

def stopwords(tweet):
    tweet= stopword.remove(tweet)
    return tweet
```

Gambar 3.7 Implementasi *Stopword Removal*

e. *Tokenizing*

Tokenizing adalah prosedur memecah teks menjadi kata, frasa, atau bagian bermakna lainnya yaitu token. Dengan kata lain, tokenisasi adalah salah satu bentuk segmentasi teks.

Gambar 3.8 Implementasi *Tokenizing*

Tabel 3.3 Tahap Preprocessing Data

Keterangan	Hasil
Data Asli	Kurang nyaman jawaban dari Customer Service @IndiHome, https://twitter.com/naufal_haniff/status/1303241906281799680
Lower Conversion	kurang nyaman jawaban dari customer service @IndiHome, https://twitter.com/naufal_haniff/status/1303241906281799680
Cleaning	kurang nyaman jawaban dari customer service
Stemming	kurang nyaman jawab customer service
Stopword Removal	kurang nyaman jawab customer service
Tokenizing	['kurang', 'nyaman', 'jawab', 'customer', 'service']

2. Vektorisasi TF-IDF

Dalam *preprocessing data* telah ditemukan kumpulan kata yang akan dijadikan *term* pada proses selanjutnya. Kumpulan kata tersebut akan dihitung jumlah kemunculan kata pada setiap kelas. Hal ini akan mempermudah dalam proses pembobotan menggunakan TF-IDF.

Tabel 3.4 Jumlah Kata Muncul Dalam Kelas

Kata	Jumlah Kata				
	D1	D2	D3	D4	D5
kurang	1	0	0	0	0
nyaman	1	0	0	0	0
jawab	1	0	0	0	0
customer	1	0	0	0	0
service	1	0	0	0	0
telepon	0	1	0	0	0
internet	0	1	0	1	1
Tidak	0	1	0	0	0
segera	0	1	0	0	0
pasang	0	1	0	0	0
tagihan	0	0	1	0	0
lebih	0	0	1	0	0
mahal	0	0	1	0	0
tetangga	0	0	1	0	0
padahal	0	0	1	0	0
paket	0	0	1	0	0
sama	0	0	1	0	0
lambat	0	0	0	1	0
banget	0	0	0	1	0
rumah	0	0	0	0	1
mati	0	0	0	0	1
minggu	0	0	0	0	1
TOTAL	**5**	**5**	**7**	**3**	**4**

Langkah berikutnya yang harus dilakukan yaitu melakukan pembobotan kata pada setiap kelas yang muncul menggunakan algoritma TF-IDF menggunakan persamaan berikut :

$$TF_{ij} = \frac{f_{ij}}{max_k f_{kj}} \tag{3.1}$$

Pada persamaan 3.1 f_{ij} merupakan istilah i dalam dokumen j, sedangkan $max_k f_{kj}$ merupakan frekuensi fitur paling umum atau fitur dengan frekuensi tinggi pada dokumen.

$$IDF_i = \log\left(\frac{N}{n_i}\right) + 1 \quad (3.2)$$

Pada persamaan 3.2 N adalah banyaknya dokumen latih yang digunakan dan n_i adalah banyaknya dokumen latih yang mengandung istilah i. Sedangkan nilai TF-IDF dihitung menggunakan persamaan 3.3 seperti berikut :

$$TFIDF_{ij} = TF_{ij} \times IDF_i \quad (3.3)$$

```
//Total Kata Pada Setiap Kalimat
CREATE VIEW num_of_word AS SELECT *, COUNT(*) AS frequency
FROM `bow` GROUP BY word, label ORDER BY label

//Kata Unik
CREATE VIEW unique_word AS SELECT *, COUNT(*) AS frequency
FROM `bow` GROUP BY word

//Perhitungan TF
CREATE VIEW tf AS SELECT num_of_word.*, category.id AS
category, category.name, num_of_word.frequency/nword.ln AS
tf FROM num_of_word, nword, category, label WHERE
num_of_word.label = nword.label AND num_of_word.label =
label.id AND label.category = category.i

//Perhitungan IDF
CREATE VIEW idf AS SELECT unique_word.*,log10(((SELECT
COUNT(*) FROM nword) / unique_word.frequency)) + 1 AS idf
FROM unique_word

//Perhitungan TFIDF
CREATE VIEW tfidf AS SELECT tf.*, idf.idf, tf.tf*idf.idf
AS tfidf  FROM tf, idf WHERE tf.word = idf.word
```

Gambar 3.9 Implementasi *TF-IDF*

Tabel 3.5 Perhitungan *Term Frequency*

Dokumen	Jumlah dokumen	Kata	Frekuensi	TF
D1	5	internet	0	0/5
D2	5		1	1/5
D3	7		0	0/7
D4	3		1	1/3
D5	4		1	1/4

Tabel 3.6 Perhitungan Inverse Document Frequency

Dokumen	Frekuensi	Kata	IDF
D1	0	internet	$log\frac{5}{3} + 1 = 1{,}22185$
D2	1		
D3	0		
D4	1		
D5	1		

Tabel 3.7 Perhitungan TF-IDF

Dokumen	Kata	TF	IDF	TF-IDF
D1	internet	0	1.22185	0
D2		1/5		0.24436
D3		0		0
D4		1/3		0.40728
D5		1/4		0.30546

Tabel 3.8 Hasil Perhitungan TF-IDF

Kata	D1	D2	D3	D4	D5
kurang	0,33979	0	0	0	0
nyaman	0,33979	0	0	0	0
jawab	0,33979	0	0	0	0
customer	0,33979	0	0	0	0
service	0,33979	0	0	0	0
telepon	0	0,33979	0	0	0
internet	0	0,24436	0	0,40728	0,30546
tidak	0	0,33979	0	0	0
segera	0	0,33979	0	0	0
pasang	0	0,33979	0	0	0
tagihan	0	0	0,24271	0	0
lebih	0	0	0,24271	0	0
mahal	0	0	0,24271	0	0
tetangga	0	0	0,24271	0	0
padahal	0	0	0,24271	0	0
paket	0	0	0,24271	0	0
sama	0	0	0,24271	0	0
lambat	0	0	0	0,56632	0
banget	0	0	0	0,56632	0
rumah	0	0	0	0	0,42474
mati	0	0	0	0	0,42474
minggu	0	0	0	0	0,42474

3. Seleksi Fitur *Information Gain*

Pada tahap seleksi fitur menggunakan *Information Gain* hasil nilai *document frequency* dari setiap kata yang terkumpul akan diproses. Sebelum mencari nilai *Information Gain*, maka harus diketahui nilai *entropy* dapat dituliskan sebagai berikut :

$$Entropy(S) = \sum_{i}^{c} -p_i \log_2 p_i \qquad (3.4)$$

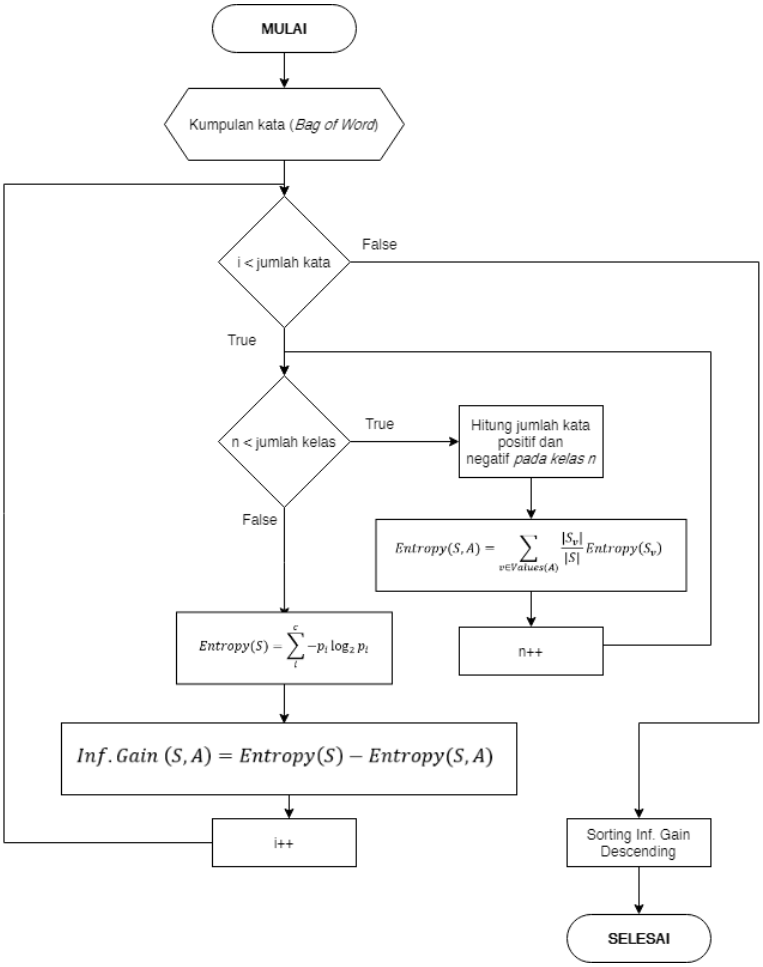
Penentuan nilai *entropy* diperlukan jumlah nilai yang terdapat pada atribut target yang disimbolkan sebagai c. Sedangkan p_i merupakan rasio antara jumlah sampel di kelas i dengan jumlah semua sampel pada himpunan data.

$$Entropy(S, A) = \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \qquad (3.5)$$

Pada persamaan 3.5 diketahui bahwa $Values(A)$ merupakan himpunan nilai-nilai yang dimungkinkan untuk atribut A, nilai V merupakan nilai yang dimungkinkan untuk atribut A, sedangkan $|S_v|$ merupakan jumlah sampel untuk nilai v dan $|S|$ merupakan jumlah seluruh sampel data.

Setelah diketahui nilai *entropy*, maka akan dilakukan proses klasifikasi menggunakan *Information Gain* menggunakan persamaan sebagai berikut :

$$Inf.Gain\ (S, A) = Entropy(S) - Entropy(S, A) \qquad (3.6)$$



Gambar 3.10 Flowchart Information Gain

```
//Jumlah Dokumen Pada Setiap Kelas
CREATE VIEW jml_dokumen AS
SELECT category, COUNT(*) AS jml_dokumen FROM label
GROUP BY category

//Total Seluruh Dokumen
CREATE VIEW total_dokumen AS
SELECT COUNT(*) AS total FROM label

//Entropy Kata
CREATE VIEW entropy_kata AS SELECT *, unique_word.frequency/(SELECT
COUNT(nword.label) FROM nword) AS prob,
    -((unique_word.frequency/(SELECT COUNT(nword.label) FROM
nword))*IFNULL(LOG2(unique_word.frequency/(SELECT
COUNT(nword.label) FROM nword)),0)+
    (((SELECT COUNT(nword.label) FROM nword)-
unique_word.frequency)/(SELECT COUNT(nword.label) FROM
nword))*IFNULL(LOG2(((SELECT COUNT(nword.label) FROM nword)-
unique_word.frequency)/(SELECT COUNT(nword.label) FROM nword)),0)
    ) AS entropy FROM unique_word

//Entropy Kelas
CREATE VIEW entropy_class AS SELECT word_of_doc.*,
jml_dokumen.jml_dokumen, -
((word_of_doc.frequency/jml_dokumen.jml_dokumen)*IFNULL((LOG2(word_
of_doc.frequency/jml_dokumen.jml_dokumen)),0) +
((jml_dokumen.jml_dokumen-
word_of_doc.frequency)/jml_dokumen.jml_dokumen)*
IFNULL(LOG2((jml_dokumen.jml_dokumen-
word_of_doc.frequency)/jml_dokumen.jml_dokumen),0)) AS entropy FROM
word_of_doc, jml_dokumen WHERE word_of_doc.category =
jml_dokumen.category

//Information Gain
CREATE VIEW inf_gain AS SELECT entropy_class.word,
entropy_kata.entropy-
SUM((entropy_class.frequency/total_dokumen.total)*entropy_class.ent
ropy) AS inf_gain FROM entropy_class, entropy_kata, total_dokumen
WHERE entropy_kata.word = entropy_class.word GROUP BY
entropy_class.word
```

Gambar 3.11 Implementasi *Information Gain*

Tabel 3.9 Contoh Data Tweet

Kelas	Tweet	Kata
C4	@IndiHome internet saya lemot banget tolong direfresh	[‘internet’, ‘saya’, ‘lambat’, ‘banget’, ’tolong’, ‘refresh’]

Berikut merupakan perhitungan dalam mencari nilai *entropy* dan *information gain* menggunakan persamaan 3.4, 3.5 dan 3.6.

$$Entropy_{(internet\mid C1)} = 0 - \frac{1}{1} \times \log_2 \frac{1}{1} = 0$$

$$Entropy_{(internet\mid C2)} = -\frac{1}{1} \times \log_2 \frac{1}{1} - 0 = 0$$

$$Entropy_{(internet\mid C3)} = 0 - \frac{1}{1} \times \log_2 \frac{1}{1} = 0$$

$$Entropy_{(internet\mid C4)} = -\frac{1}{2} \times \log_2 \frac{1}{2} - \frac{1}{2} \times \log_2 \frac{1}{2} = 1$$

$$Entropy_{(internet\mid C5)} = -\frac{1}{1} \times \log_2 \frac{1}{1} - 0 = 0$$

$$Entropy_{(internet)} = -\left(\frac{4}{6} \times \log_2 \frac{4}{6} + \frac{2}{6} \times \log_2 \frac{2}{6}\right) = 0{,}91829$$

$$Inf.Gain_{(internet)}$$

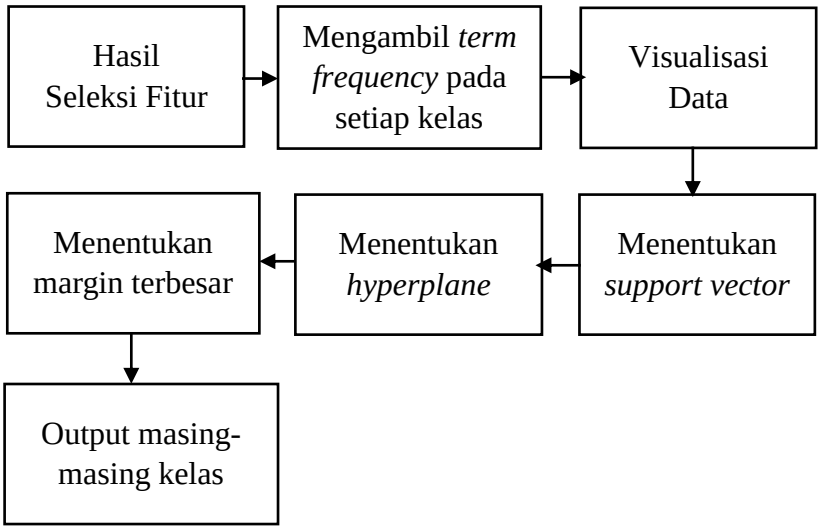
$$= 0.91829$$

$$-\left[\left(\frac{1}{6} \times 0\right) + \left(\frac{1}{6} \times 0\right) + \left(\frac{1}{6} \times 0\right) + \left(\frac{2}{6} \times 1\right) + \left(\frac{1}{6} \times 0\right)\right]$$

$$= 0{,}58495$$

4. Klasifikasi *Support Vector Machine*

Langkah awal sebelum melakukan pemodelan SVM diperlukan pengambilan data kemunculan kata dalam suatu kelas atau disebut dengan *term frequency*.

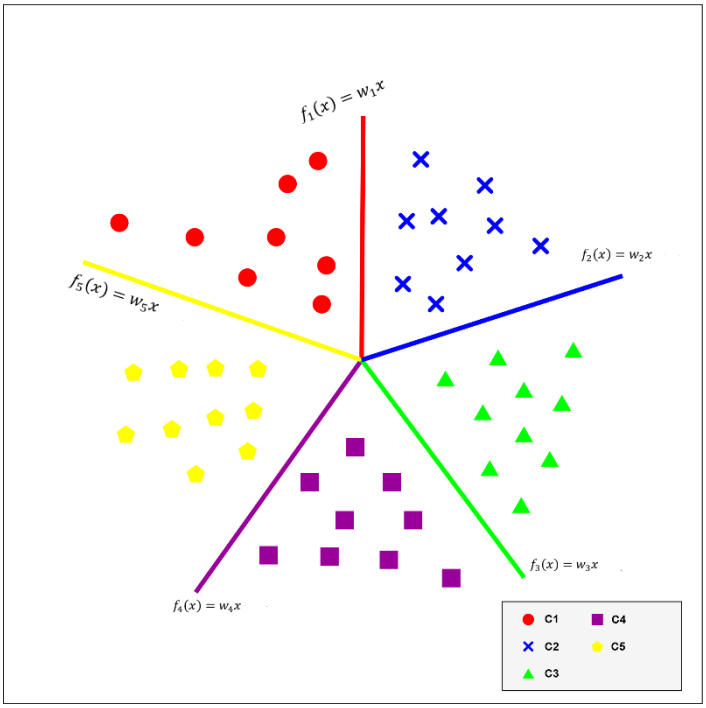


Gambar 3.12 Blok Diagram *Support Vector Machine*

Pada tahap klasifikasi *Multiclass Support Vector Machine* digunakannya pendekatan *one-against-all* untuk menyelesaikan permasalahan klasifikasi lima kelas.

Tabel 3.10 Fungsi Klasifikasi *Support Vector Machine*

$y_i = 1$	$y_i = -1$	Fungsi
C1	Bukan C1	$f_1(y) = w_1 x + b$
C2	Bukan C2	$f_2(y) = w_2 x + b$
C3	Bukan C3	$f_3(y) = w_3 x + b$
C4	Bukan C4	$f_4(y) = w_4 x + b$
C5	Bukan C5	$f_5(y) = w_5 x + b$



Gambar 3.13 Visualisasi Klasifikasi *Support Vector Machine*

Sebelum tahap pelatihan data akan dilakukan inisiasi bobot awal sebagai berikut :

$$w_{awal} = {}^{1}/_{total\ fitur} \tag{3.7}$$

w_{awal} merupakan bobot awal yang akan digunakan dalam tahap pelatihan nantinya. Sedangkan $total\ fitur$ merupakan total fitur yang dihasilkan pada tahap TF-IDF. Sehingga akan dihasilkan bobot awal sebagai berikut :

$$w_{awal} = \frac{1}{30} = 0{,}033$$

Tahapan selanjutnya yaitu menghitung nilai parameter pada setiap fitur yang tersedia. Fungsi tersebut nantinya digunakan dalam pembuatan

hyperplane. Model hyperplane yang digunakan dalam penelitian ini adalah model linear seperti diperlihatkan pada persamaan 3.8.

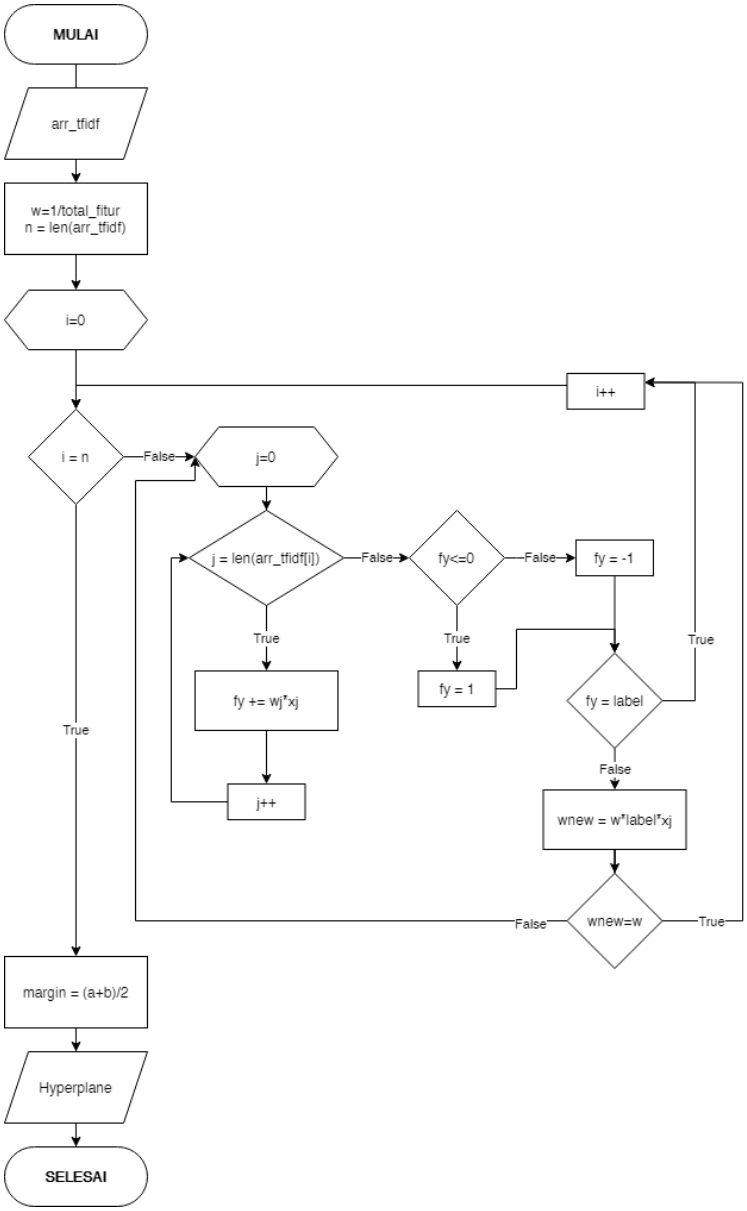
$$f(y) = \sum_{i=1}^{total\ fitur} w_i \times x_i \tag{3.8}$$

Pada persamaan 3.8, w_i merupakan bobot dari setiap fitur. Pada tahap awal akan digunakan persamaan 3.7 sebagai bobot awal (w_{awal}). Sedangkan x_i merupakan nilai setiap fitur yang dihasilkan pada tahap vektorisasi TF-IDF. Dari hasil persamaan 3.8 tersebut maka akan dihasilkan :

a. Jika hasil dari $f(y_i) > 0$ maka $y_i = +1$
b. Jika hasil dari $f(y_i) < 0$ maka $y_i = -1$

Akan tetapi jika y_i tidak sesuai dengan hasil dari pelabelan maka diperlukan pembaruan pada bobot. Bobot yang baru akan diperbarui menggunakan persamaan berikut :

$$w_{baru} = w_{awal} + y_i \times x_{ij} \tag{3.9}$$



Gambar 3.14 *Flowchart* Proses Pelatihan

```
from sklearn.utils import check_random_state
from sklearn.preprocessing import LabelEncoder
from .forms import *
from .models import *
import numpy as np

class MulticlassSVM():
    def fit(self, X, y):
        n_samples, n_features = X.shape

        # Normalisasi label pada kategori
        self._label_encoder = LabelEncoder()
        y = self._label_encoder.fit_transform(y)

        n_classes = len(self._label_encoder.classes_)
        self.dual_coef_ = np.zeros((n_classes, n_samples),
dtype=np.float64)
        self.coef_ = np.zeros((n_classes, n_features))

        # Normalisasi sebelum perhitungan
        norms = np.sqrt(np.sum(X ** 2, axis=1))

        # Melakukan pengacakan training data
        rs = check_random_state(self.random_state)
        ind = np.arange(n_samples)
        rs.shuffle(ind)

        violation_init = None
        for it in range(self.max_iter):
            violation_sum = 0

            for ii in range(n_samples):
                i = ind[ii]

                #Nilai normaliasi 0 akan dilewati
                if norms[i] == 0:
                    continue

                f = self._partial_gradient(X, y, i)
                v = self._violation(f, y, i)
                violation_sum += v

                if v < 1e-12:
                    continue

                delta = self._solve_subproblem(f, y, norms, i)

                self.coef_ += (delta * X[i][:, np.newaxis]).T
                self.dual_coef_[:, i] += delta

            if it == 0:
                violation_init = violation_sum

            vratio = violation_sum / violation_init

            if self.verbose >= 1:
                print("iterasi", it + 1)

            if vratio < self.tol:
                if self.verbose >= 1:
                    print("Converged")
                    context = {
                        "fitur": self.fitur,
                        "fold": self.fold,
                        "weight" : ', '.join(str(v) for v in
self.coef_)
                    }
                    form = FormWeight(context)
                    if form.is_valid():
                        form.save()
                    else :
                        print(form.errors)
                break

        return self
```

Gambar 3.15 Implementasi Proses Pelatihan

Hasil dari pelatihan tersebut akan digunakan dalam mencari nilai *hyperplane* sebagai batas nilai positif (+1) dan negatif (-1). Dari batas tersebut terdapat jarak atau *margin*. *Margin* tersebut diformulasikan sebagai berikut :

$$margin = \frac{a + b}{2} \tag{3.10}$$

Dimana a merupakan nilai $f(y_i) > 0$ dan b merupakan nilai $f(y_i) < 0$. Dengan ini fungsi *hyperplane* dapat diformulasikan sebagai berikut :

$$y\left(\sum_{i=1}^{total\ fitur} w_i \times x_i\right) \geq margin \tag{3.11}$$

Dari hasil pada Tabel 3.8 akan dilakukan perhitungan untuk mencari nilai *hyperplane* menggunakan kalimat pertama (D1) dengan menggunakan persamaan 3.8.

$$f(y) = \big((0{,}033 \times 0{,}3379) + (0{,}033 \times 0{,}3379) + (0{,}033 \times 0{,}3379) + (0{,}033 \times 0{,}3379) + (0{,}033 \times 0{,}3379) + (0{,}033 \times 0\ ) + \cdots (0{,}033 \times x_{30})\big)$$

$$f(y) = 0{,}0112 + 0{,}0112 + 0{,}0112 + 0{,}0112 + 0{,}0112 + 0 + .. + 0$$

$$f(y) = 0{,}056$$

$f(y) > 0$, maka $y = +1$

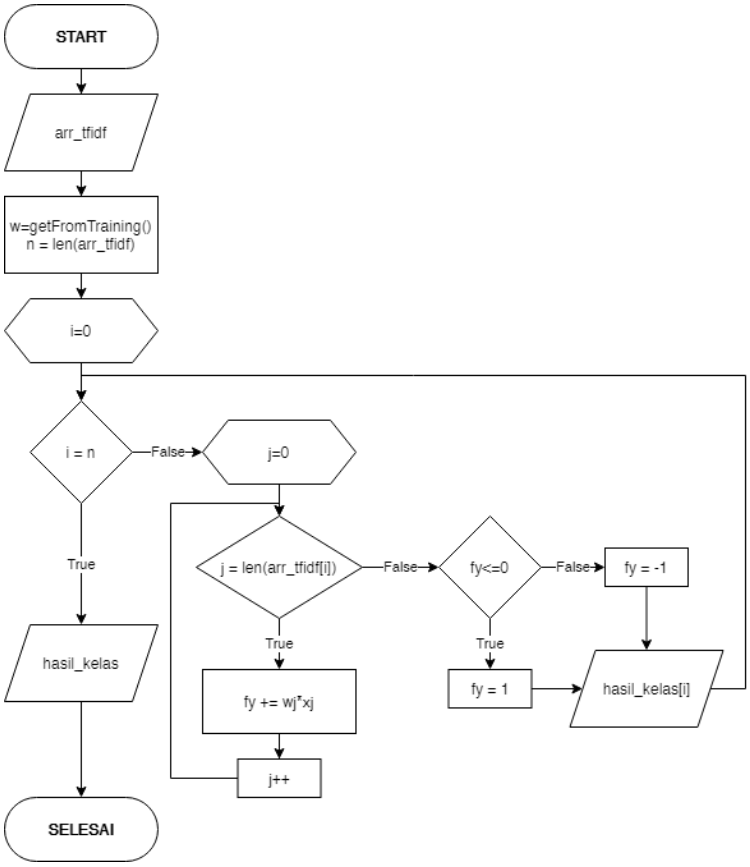
Hasil dari perhitungan di atas sebesar $0{,}056$. maka nilai $y = +1$. Akan tetapi jika hasil $f(y) < 0$, maka akan dilakukan pembaruan bobot pada fitur tersebut menggunakan persamaan 3.9, sebagai berikut :

$$w_{baru} = \frac{1}{30} + (-1) \times 0{,}056$$

$$w_{baru} = 0{,}033 + (-0{,}056)$$

$$w_{baru} = -0{,}023$$

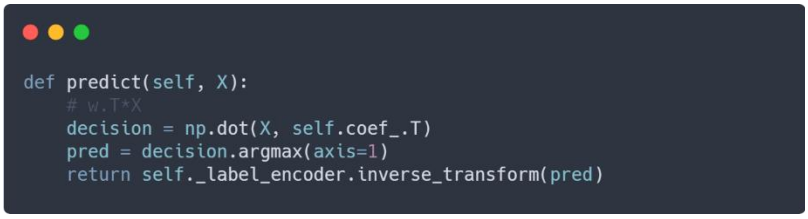
Dari proses pelatihan data akan dihasilkan model klasifikasi *Support Vector Machine.* Model ini nantinya digunakan dalam tahap pengujian.



Gambar 3.16 *Flowchart* Proses Pengujian

$$f(x, w) = wx$$

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1p} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2p} \\ w_{31} & w_{32} & w_{33} & \cdots & w_{3p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{n1} & w_{n2} & w_{n3} & \ldots & w_{np} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \cdots \\ x_p \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \cdots \\ m_p \end{bmatrix} \quad (3.12)$$



```
def predict(self, X):
    # w.T*X
    decision = np.dot(X, self.coef_.T)
    pred = decision.argmax(axis=1)
    return self._label_encoder.inverse_transform(pred)
```

Gambar 3.17 Implementasi Proses Pengujian

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Langkah-langkah Uji Coba

1. Input *dataset*

   Dataset yang terkumpul sebanyak 1000 *tweets*, yang telah disimpan pada *database MySQL*. Dataset telah terbagi sebanyak 200 *tweets* pada masing-masing kelas.

Tabel 4.1 Jumlah *Dataset* setiap Kelas

Kelas/Label	Jumlah
Pelayanan (C1)	200
Pemasangan (C2)	200
Tagihan (C3)	200
Lambat (C4)	200
Disconnect (C5)	200

Tabel 4.2 Sampel Daftar *Tweets*

No.	*Tweets*	Label
1.	@IndiHome admin saya sudah DM keluhan saya, kok belum ada tanggapan ya	1
2.	@IndiHome @IndiHomeCare Terimakasih customer service buruk rupa buruk akhlak, wifi saya kenapa @IndiHomeCare?	1
3.	@IndiHomeCare @TelkomCare @IndiHome Iya ni.. lambat penangannya..	1
4.	@IndiHome admin saya sudah DM keluhan saya, kok belum ada tanggapan ya	1
5.	@IndiHome sering dapat pengalaman buruk kalau sudah bicara pelayanan indihome	1

6.	@IndiHome Saya daftar pemasangan indihome dari tanggal 20 maret kok gaada kabar sama sekali ya sampai sekarang. Padahal udah difollow up berkali2	2
7.	@IndiHome Halo min, mohon cek DM ya, saya ada pengajuan pemasangan baru tapi sampai hari ini belum dipasang. Terima kasih	2
8.	@IndiHome bener ga sih kalau mau pemasangan wifinya daftar via website bisa ampe berbulan2?	2
9.	@IndiHome min mau pasang di karanganyar bisa gak'	2
10.	Untuk biaya pemasangan wifi berapa ya min @IndiHome @IndiHomeCare	2
11.	@IndiHome @IndiHomeCare : mohon dibantu, bulan maret sudah melakukan penutupan 1 nomor telp, bulan ini di cek masih muncul tagihan?	3
12.	@frshzrnmln @IndiHome @Faruqumar Udah pindah rumah dari kapan tau. Tagihan jalan terus. Udah gila kali	3
13.	@IndiHome halo min. Mau tanya mengenai tagihan nih. Ane bingung. Kok bisa beda nominalnya dari biasanya.	3
14.	udah jelas wifi udah diputus tapi tagihan masih berjalan. Konyol lu @IndiHome @IndiHomeCare	3
15.	Kak kenapa bulan ini tagihan saya lebih mahal daripada biasanya? @IndiHome	3
16.	.@IndiHome abis hujan lemot ya kak wifi nya	4
17.	@IndiHome halo, koneksi saya kok nge-lag terus ya? Biasanya jaringan strong, skrng hilang muncul terus	4
18.	hih @IndiHome ga biasa biasanya yaa kamu dipake streaming buffering	4
19.	bisa tolong refresh jaringan gakk? Lemot banget ih mau marah!! @IndiHome	4
20.	Apakah jaringan @IndiHome sedang gangguan, ini kok lambat koneksi banget?	4
21.	@IndiHome DM direspon ya, udh berjam2 ga drespon, koneksi ga bsa sama skali, sbulan bs berkali2 bgtu, klo gtu mending putus aja layanannya deh ganti yg lain	5

22.	@IndiHome halo min, internet saya bermasalah. tanda LOS-nya merah	5
23.	@IndiHome halooo min lampu router merah trs wifinya matiiiii udah coba restart juga, udah dm yaa nomer internetnya, makasih.	5
24.	@IndiHome jaringan los, merah, udh cabut kabel, mohon diperbaiki	5
25.	@IndiHome malam, ini jaringan internet saya dari tadi siang bermasalah. Sudah dilakukan restart tapi tetap sama, bisa tolong bantu check	5

2. Pembagian *dataset*

Peneliti melakukan pengujian dengan membagi jumlah *dataset* menjadi dua jenis yaitu menggunakan *k-fold cross validation* dari total *dataset* sebanyak 1000 *tweets* dengan *5-fold cross validation*, yang nantinya akan diklasifikasikan ke lima kategori yaitu pelayanan, tagihan, koneksi yang lambat dan koneksi terputus.

$$Jumlah\ Data\ Uji = \frac{1}{jumlah\ fold} * jumlah\ dataset \quad (4.1)$$

Gambar 4.1 Pembagian Dataset dengan 5-fold cross validation

$$Jumlah\ Data\ Uji = \frac{1}{5} \times 1000 = 200\ tweets$$

3. Pemodelan Klasifikasi

Pada tahap ini akan dihasilkan model *Support Vector Machine* dengan pendekatan *one-against-all*, sehingga terbentuk model dari masing-masing kelas terhadap jumlah fitur dan nilai k.

Tabel 4.3 Hasil SVM dengan Fitur 10% dan k=1

Jumlah Term	Kelas	*w*	*margin*
10%	C1	[0.402258, -0.05973, 0.28169, -2.700781, … , 1.310903, -0,845475]	-0.47454
	C2	[0.389353, 0.065248, -0.070594 , 5.690103 … -0.520802, 1.138181]	-0.69247
	C3	[0.862417, -0.354409, 0.183410, -1.497895 … -0.335884, 0.364694]	-0.78964
	C4	[-1.338729, -0.444008, -0.738664, -1.754970, … - 0.077143, -0.325874]	-0.45380
	C5	-0.547757, 0.641268, 0.645309, -1.901939, … -0.62857, 0.093182	-0.58186

Berdasarkan parameter pada Tabel 4.3, maka pada jumlah fitur 10% dan k =1 dapat dirumuskan fungsi matematika sebagai berikut :

$Hyperplane_{C1}$

$$= 0{,}402258x_1 + (-0{,}05973)x_2 + 0{,}281619x_4$$
$$+ (-2{,}700781)x_3 + \ldots\ 1{,}310903x_{193}$$
$$+ (-0{,}845475)x_{194}$$

$Hyperplane_{C2}$

$$= 0{,}389353\ x_1 + 0{,}065248x_2 + (-0{,}070594)x_3$$
$$+ 5{,}690103x_4 + \ldots (-0{,}520802)x_{193} + 1{,}138181x_{194}$$

$Hyperplane_{C3}$

$$= 0{,}862417x_1 + (-0{,}354409)x_2 + 0{,}183410x_3 + (-1{,}497895)x_4 + \ \ldots\ (-0{,}335884)x_{193} + 0{,}364694x_{194}$$

$Hyperplane_{C4}$

$$= -1{,}338729x_1 + (-0{,}444008)x_2 + (-0{,}738664)x_3 + (-1{,}754970)x_4 + \ \ldots\ (-0{,}077143)x_{193} + (-0{,}325874)x_{194}$$

$Hyperplane_{C5}$

$$= -0{,}547757x_1 + 0{,}641268x_2 + 0{,}645309x_3 + (-1{,}901939\ )x_4 + \ \ldots\ (-0{,}62857)x_{193} + 0{,}093182x_{194}$$

Berdasarkan fungsi *hyperplane* dan nilai *margin* pada masing-masing kelas, dapat diketahui jika hasil perhitungan pada model $Hyperplane_{C1}$ memiliki nilai lebih kecil dari $margin_{C1}$ maka akan diklasifikasikan sebagai C1, jika tidak maka akan diklasifikasikan sebagai bukan C1 dan berlaku pada model *hyperplane* dan *margin* lainnya.

4. Menampilkan hasil

Dalam tahap ini akan ditampilkan hasil klasifikasi berupa kelas dari hasil klasifikasi dan kelas aktual pada masing-masing skenario pengujian dengan percobaan dengan menggunakan persentase jumlah fitur yang

telah tesimpan yaitu 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% dan 90%.

Tabel 4.4 Hasil Klasifikasi dengan Nilai k = 1

No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…
198	5	5	5	5	5	5	5	5	5	5
199	1	1	5	5	5	5	5	5	1	1
200	1	1	1	1	1	1	1	1	1	1

Tabel 4.5 Hasil Klasifikasi dengan Nilai k = 2

No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	2	5	5	5	5	5	5	5	5
4	1	2	2	2	2	2	2	2	2	2
5	1	5	5	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…
198	5	5	5	5	5	5	5	5	5	5
199	5	4	4	4	4	4	4	4	4	4
200	5	5	5	5	5	5	5	5	5	5

Tabel 4.6 Hasil Klasifikasi dengan Nilai k = 3

No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	5	5	5	5	5	5	5	5	5
5	1	5	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	…
198	5	4	4	4	4	4	4	4	4	4
199	5	1	5	5	5	5	5	5	5	5
200	5	5	5	1	1	5	5	5	5	5

Tabel 4.7 Hasil Klasifikasi dengan Nilai k = 4

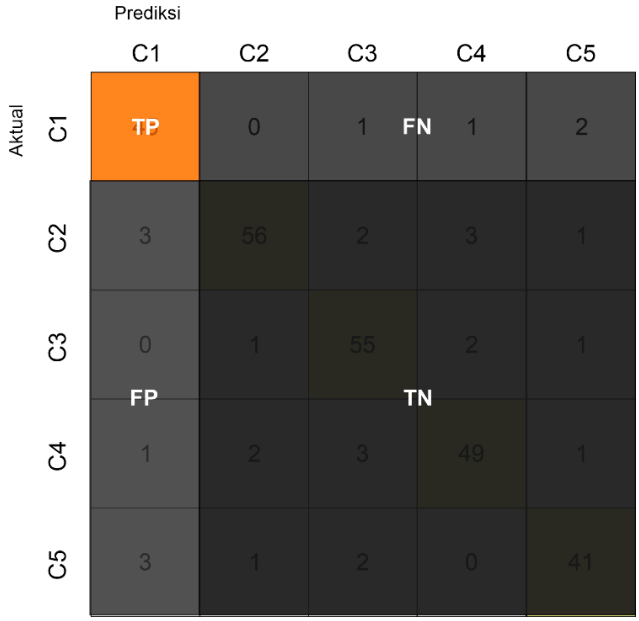
No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	5	5	5	5	5	5	5	5	5
5	1	1	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	...
198	5	2	5	5	5	5	5	5	5	5
199	5	5	5	5	5	5	5	5	5	5
200	5	1	5	5	5	5	5	5	5	5

Tabel 4.8 Hasil Klasifikasi dengan Nilai k = 5

No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	4	4	4	4	4	4	4	4	4
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	2	5	5	5	5	5	5	5	5
5	1	1	1	1	1	1	1	1	1	1
…	…	…	…	…	…	…	…	…	…	...
198	5	5	5	5	5	5	5	5	5	5
199	5	5	5	5	5	5	5	5	5	5
200	5	1	1	1	1	1	1	1	1	1

5. Evaluasi Sistem

Dalam tahap ini akan ditampilkan evaluasi sistem dengan mencari nilai akurasi, presisi, *recall* dan *f-measure*. Dalam menampilkan hasil evaluasi maka perlu diketahui *confussion matrix* untuk mengetahui nilai TP, TN, FN dan FP yaitu :



Gambar 4.2 Confusion Matrix Multiclass

a. *True Positif* (TP) merupakan data positif yang diprediksi sebagai data positif.
b. *True Negatif* (TN) merupakan data negatif yang diprediksi sebagai data negatif
c. *False Negatif* (FN) merupakan data positif yang diprediksi sebagai data negatif.
d. *False Positif* (FP) merupakan data negatif yang diprediksi sebagai data positif.

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Akurasi merupakan hasil perbandingan suatu nilai uji dengan nilai sebenarnya. Tingkat akurasi dibuktikan dalam bentuk persentase dengan prediksi benar terhadap total percobaan yang dilakukan. Tingkat akurasi dapat dihitung menggunakan persamaan 4.1

$$Akurasi_{kelas} = \frac{TP + TN}{Jumlah\ Dokumen} 100\% \quad (4.1)$$

Presisi adalah tingkat ketepatan antara informasi yang diminta oleh *user* dengan hasil yang diberikan oleh sistem. Presisi merupakan hasil persentase perbandingan nilai *True Positif* (TP) dengan keseluruhan hasil yang diprediksi positif. Tingkat *presisi* dapat dihitung menggunakan persamaan 4.2.

$$Presisi_{kelas} = \frac{TP}{TP + FP} 100\% \quad (4.2)$$

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali suatu informasi. *Recall* mengukur persentase perbandingan nilai *True Positif* (TP) dengan keseluruhan kemungkinan data yang positif. Pengukuran *recall* dapat dihitung menggunakan persamaan 4.3.

$$Recall_{kelas} = \frac{TP}{TP + FN} 100\% \quad (4.3)$$

Nilai *f-measure* merupakan hasil perhitungan evaluasi dengan mengkombinasikan nilai *recall* dan *presisi.* Nilai *f-measure* berfungsi untuk menentukan tingkat efektivitas suatu pengujian. Pengukuran *f-measure* dapat dihitung menggunakan persamaan 4.4.

$$f - measure_{kelas} = 2\frac{Presisi \times Recall}{Presisi + Recall} \quad (4.4)$$

Dari hasil akurasi, presisi, *recall* dan *f-measure* maka akan dilakukan perhitungan rata-rata pada persamaan 4.5 dan standar deviasi pada persamaan 4.6 berdasarkan hasil setiap *fold.*

$$\bar{x} = \frac{\sum x_i}{N} \quad (4.5)$$

$$s = \sqrt{\frac{\sum(x_i - x)^2}{N - 1}} \quad (4.6)$$

Selanjutnya akan dilakukan uji coba untuk mengetahui nilai akurasi, presisi, *recall* dan *f-measure* metode *Support Vector Machine* dengan seleksi fitur *Information Gain* sebagai klasifikasi pengaduan layanan pengguna IndiHome pada media sosial Twitter.

4.2 Hasil Uji Coba

Pada subbab ini ditampilkannya hasil dari skenario pengujian menggunakan 5-*fold cross validation* terhadap jumlah fitur sebesar 10%, 20%, 30%, 40%. 50%, 60%, 70%, 80%, 90%.

Berdasarkan Tabel 4.4 dengan membandingkan kelas aktual dan hasil prediksi pada masing-masing kelas. Dengan hasil tersebut dapat diketahui jumlah data yang diprediksi sesuai dengan kelasnya dan tidak sesuai dengan kelasnya, sehingga didapatkan nilai berikut:

Tabel 4.9 *Confusion Matrix* dengan k = 1 dan Fitur 10%

Prediksi / Aktual	C1	C2	C3	C4	C5
C1	35	0	1	3	6
C2	3	32	1	0	1
C3	0	2	34	1	0
C4	2	1	1	26	6
C5	11	2	0	4	28

Pada hasil *confusion matrix* dapat diketahui nilai TP, TN, FN dan FP dengan menggunakan skema *confusion matrix* pada Gambar 4.1. Berikut

merupakan perhitungan nilai TP, TN, FN dan FP pada kelas atau kategori pelayanan (C1) :

$TP = 35$

$TN = (32 + 1 + 0 + 1) + (2 + 34 + 1 + 0) +$

$(1 + 1 + 26 + 6) + (2 + 0 + 4 + 28) = 139$

$FN = 0 + 1 + 3 + 6 = 10$

$FP = 3 + 0 + 2 + 11 = 16$

Tabel 4.10 Hasil Perhitungan TP, TN, FN dan FP

Jumlah Fitur	Kelas	TP	TN	FN	FP
10%	C1	35	139	10	16
	C2	32	158	5	5
	C3	34	160	3	3
	C4	26	156	10	8
	C5	28	142	17	13

Berdasarkan hasil prediksi, maka dilakukan perhitungan dalam mencari nilai akurasi presisi, *recall*, dan *f-measure* pada masing-masing kelas yaitu sebagai berikut :

a. Evaluasi Kategori Pelayanan (C1)

$$Akurasi_{(C1)} = \frac{35 + 139}{200} 100\% = 87\%$$

$$Presisi_{(C1)} = \frac{35}{35 + 16} 100\% = 68{,}63\%$$

$$Recall_{(C1)} = \frac{35}{35 + 10} 100\% = 77{,}78\%$$

$$f - measure = 2 \frac{68{,}63 \times 77{,}78}{68{,}63 + 77{,}78} = 77{,}92\%$$

b. Evaluasi Kategori Pemasangan (C2)

$$Akurasi_{(C2)} = \frac{32 + 158}{200} 100\% = 95\%$$

$$Presisi_{(C2)} = \frac{32}{32 + 5} 100\% = 86{,}49\%$$

$$Recall_{(C2)} = \frac{32}{32 + 5} 100\% = 86{,}49\%$$

$$f - measure = 2 \frac{86{,}49 \times 86{,}49}{86{,}49 + 86{,}49} = 86.49\%$$

c. Evaluasi Kategori Tagihan (C3)

$$Akurasi_{(C3)} = \frac{34 + 160}{200} 100\% = 97\%$$

$$Presisi_{(C3)} = \frac{34}{34 + 3} 100\% = 91{,}89\%$$

$$Recall_{(C3)} = \frac{34}{34 + 3} 100\% = 91{,}89\%$$

$$f - measure = 2 \frac{91{,}89 \times 91{,}89}{91{,}89 + 91{,}89} = 91{,}89\%$$

d. Evaluasi Kategori Lambat (C4)

$$Akurasi_{(C4)} = \frac{26 + 156}{200} 100\% = 91\%$$

$$Presisi_{(C4)} = \frac{26}{26 + 8} 100\% = 76{,}47\%$$

$$Recall_{(C4)} = \frac{26}{26 + 10} 100\% = 72.22\%$$

$$f - measure = 2 \frac{76{,}47 \times 72{,}22}{76.47 + 72{,}22} = 74{,}29\%$$

e. Evaluasi Kategori *Disconnect* (C5)

$$Akurasi_{(C5)} = \frac{28 + 142}{200} 100\% = 85\%$$

$$Presisi_{(C5)} = \frac{28}{28+13}100\% = 68{,}29\%$$

$$Recall_{(C5)} = \frac{28}{28+17}100\% = 62{,}22\%$$

$$f - measure = 2\frac{68{,}33 \times 57{,}75}{69{,}33 + 57{,}75} = 65{,}12\%$$

Tabel 4.11 Hasil Pengujian dengan k = 1

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	87.0%	68.63%	77.78%	72.92%
	C2	95.0%	86.49%	86.49%	86.49%
	C3	97.0%	91.89%	91.89%	91.89%
	C4	91.0%	76.47%	72.22%	74.29%
	C5	85.0%	68.29%	62.22%	65.12%
20%	C1	85.5%	66.0%	73.33%	69.47%
	C2	94.5%	84.21%	86.49%	85.33%
	C3	94.5%	84.21%	86.49%	85.33%
	C4	93.5%	82.86%	80.56%	81.69%
	C5	87.0%	74.36%	64.44%	69.05%
30%	C1	87.5%	70.83%	75.56%	73.12%
	C2	96.0%	91.43%	86.49%	88.89%
	C3	94.5%	84.21%	86.49%	85.33%
	C4	92.0%	77.78%	77.78%	77.78%
	C5	88.0%	74.42%	71.11%	72.73%
40%	C1	88.0%	72.34%	75.56%	73.91%
	C2	95.5%	88.89%	86.49%	87.67%
	C3	95.0%	86.49%	86.49%	86.49%
	C4	93.0%	77.5%	86.11%	81.58%
	C5	88.5%	77.5%	68.89%	72.94%
50%	C1	87.0%	71.11%	71.11%	71.11%
	C2	95.5%	88.89%	86.49%	87.67%
	C3	95.5%	86.84%	89.19%	88.0%
	C4	92.0%	75.0%	83.33%	78.95%
	C5	88.0%	75.61%	68.89%	72.09%
60%	C1	88.0%	73.33%	73.33%	73.33%
	C2	96.0%	91.43%	86.49%	88.89%
	C3	95.5%	86.84%	89.19%	88.0%
	C4	92.0%	75.0%	83.33%	78.95%
	C5	88.5%	76.19%	71.11%	73.56%
70%	C1	87.5%	71.74%	73.33%	72.53%
	C2	95.5%	88.89%	86.49%	87.67%
	C3	95.5%	86.84%	89.19%	88.0%
	C4	91.5%	74.36%	80.56%	77.33%
	C5	88.0%	75.61%	68.89%	72.09%

80%	C1	88.0%	74.42%	71.11%	72.73%
	C2	95.5%	88.89%	86.49%	87.67%
	C3	95.5%	86.84%	89.19%	88.0%
	C4	92.5%	76.92%	83.33%	80.0%
	C5	88.5%	75.0%	73.33%	74.16%
90%	C1	88.0%	73.33%	73.33%	73.33%
	C2	95.5%	88.89%	86.49%	87.67%
	C3	95.5%	86.84%	89.19%	88.0%
	C4	92.5%	76.92%	83.33%	80.0%
	C5	87.5%	73.81%	68.89%	71.26%

Pada Tabel 4.11 dapat diketahui hasil evaluasi dengan nilai k = 1 menghasilkan penggunaan fitur terbaik sebesar 60% yang menghasilkan nilai akurasi 88%, presisi 73,33%, *recall* 73,33% dan *f-measure* 73,33% pada kelas pelayanan (C1), nilai akurasi 96%, presisi 91,43%, *recall* 86,49% dan *f-measure* 88,89% pada kelas pemasangan (C2), nilai akurasi 95,55%, presisi 86,84%, *recall* 89,19% dan *f-measure* 88% pada kelas tagihan (C3), nilai akurasi 92%, presisi 75%, *recall* 83,33% dan *f-measure* 78,95% pada kelas lambat (C4) dan nilai akurasi 88,5%, presisi 76,19%, *recall* 71,11% dan *f-measure* 73,56% pada kelas *disconnect* (C5).

Tabel 4.12 Hasil Pengujian dengan k = 2

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	82.0%	47.37%	52.94%	50.0%
	C2	94.5%	91.89%	80.95%	86.08%
	C3	89.5%	70.73%	76.32%	73.42%
	C4	92.0%	82.5%	78.57%	80.49%
	C5	86.0%	68.18%	68.18%	68.18%
20%	C1	82.0%	47.22%	50.0%	48.57%
	C2	94.5%	87.8%	85.71%	86.75%
	C3	91.0%	75.0%	78.95%	76.92%
	C4	90.5%	79.49%	73.81%	76.54%
	C5	87.0%	70.45%	70.45%	70.45%
30%	C1	84.0%	52.94%	52.94%	52.94%
	C2	95.5%	90.24%	88.1%	89.16%
	C3	92.0%	78.95%	78.95%	78.95%

	C4	93.5%	87.18%	80.95%	83.95%
	C5	87.0%	68.75%	75.0%	71.74%
40%	C1	85.0%	55.88%	55.88%	55.88%
	C2	96.0%	94.74%	85.71%	90.0%
	C3	91.0%	73.81%	81.58%	77.5%
	C4	92.0%	82.5%	78.57%	80.49%
	C5	86.0%	67.39%	70.45%	68.89%
50%	C1	86.5%	61.29%	55.88%	58.46%
	C2	95.5%	92.31%	85.71%	88.89%
	C3	92.0%	76.19%	84.21%	80.0%
	C4	92.0%	84.21%	76.19%	80.0%
	C5	87.0%	68.0%	77.27%	72.34%
60%	C1	86.5%	61.29%	55.88%	58.46%
	C2	96.0%	92.5%	88.1%	90.24%
	C3	92.5%	78.05%	84.21%	81.01%
	C4	92.5%	84.62%	78.57%	81.48%
	C5	87.5%	69.39%	77.27%	73.12%
70%	C1	86.5%	60.61%	58.82%	59.7%
	C2	95.5%	90.24%	88.1%	89.16%
	C3	92.5%	78.05%	84.21%	81.01%
	C4	92.5%	84.62%	78.57%	81.48%
	C5	88.0%	71.74%	75.0%	73.33%
80%	C1	86.5%	60.61%	58.82%	59.7%
	C2	96.0%	92.5%	88.1%	90.24%
	C3	93.0%	80.0%	84.21%	82.05%
	C4	92.0%	82.5%	78.57%	80.49%
	C5	88.5%	72.34%	77.27%	74.73%
90%	C1	86.5%	60.61%	58.82%	59.7%
	C2	94.5%	89.74%	83.33%	86.42%
	C3	92.0%	76.19%	84.21%	80.0%
	C4	92.0%	84.21%	76.19%	80.0%
	C5	87.0%	68.75%	75.0%	71.74%

Pada Tabel 4.12 dapat diketahui hasil evaluasi dengan nilai k = 2 menghasilkan penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 86,5%, presisi 60,61%, *recall* 58,82% dan *f-measure* 59,7% pada kelas pelayanan (C1), nilai akurasi 96%, presisi 92,5%, *recall* 88,1% dan *f-measure* 90,24% pada kelas pemasangan (C2), nilai akurasi 93%, presisi 80%, *recall* 84,21% dan *f-measure* 82,05% pada kelas tagihan (C3), nilai akurasi 92%, presisi 82,5%, *recall* 78,57% dan *f-measure* 80,49% pada

kelas lambat (C4) dan nilai akurasi 88,5%, presisi 72,34%, *recall* 77,27% dan *f-measure* 74,73% pada kelas *disconnect* (C5).

Tabel 4.13 Hasil Pengujian dengan k = 3

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	85.5%	70.97%	52.38%	60.27%
	C2	89.5%	70.73%	76.32%	73.42%
	C3	90.5%	83.33%	69.77%	75.95%
	C4	91.0%	76.19%	80.0%	78.05%
	C5	83.5%	54.0%	72.97%	62.07%
20%	C1	88.5%	82.76%	57.14%	67.61%
	C2	91.0%	73.81%	81.58%	77.5%
	C3	92.0%	84.62%	76.74%	80.49%
	C4	91.0%	75.0%	82.5%	78.57%
	C5	88.5%	65.22%	81.08%	72.29%
30%	C1	90.0%	84.38%	64.29%	72.97%
	C2	92.5%	76.74%	86.84%	81.48%
	C3	94.0%	89.74%	81.4%	85.37%
	C4	93.5%	81.4%	87.5%	84.34%
	C5	90.0%	69.77%	81.08%	75.0%
40%	C1	90.0%	86.67%	61.9%	72.22%
	C2	91.5%	73.33%	86.84%	79.52%
	C3	94.5%	90.0%	83.72%	86.75%
	C4	93.0%	80.95%	85.0%	82.93%
	C5	90.0%	69.77%	81.08%	75.0%
50%	C1	90.5%	89.66%	61.9%	73.24%
	C2	92.0%	75.0%	86.84%	80.49%
	C3	94.0%	87.8%	83.72%	85.71%
	C4	94.0%	83.33%	87.5%	85.37%
	C5	90.5%	70.45%	83.78%	76.54%
60%	C1	90.5%	89.66%	61.9%	73.24%
	C2	91.5%	73.33%	86.84%	79.52%
	C3	94.0%	87.8%	83.72%	85.71%
	C4	94.0%	83.33%	87.5%	85.37%
	C5	91.0%	72.09%	83.78%	77.5%
70%	C1	90.5%	89.66%	61.9%	73.24%
	C2	91.5%	73.33%	86.84%	79.52%
	C3	94.0%	87.8%	83.72%	85.71%
	C4	94.0%	83.33%	87.5%	85.37%
	C5	91.0%	72.09%	83.78%	77.5%
80%	C1	90.5%	89.66%	61.9%	73.24%
	C2	92.0%	75.0%	86.84%	80.49%
	C3	94.0%	87.8%	83.72%	85.71%
	C4	94.0%	83.33%	87.5%	85.37%
	C5	90.5%	70.45%	83.78%	76.54%

90%	C1	90.0%	86.67%	61.9%	72.22%
	C2	91.5%	74.42%	84.21%	79.01%
	C3	94.0%	87.8%	83.72%	85.71%
	C4	94.0%	83.33%	87.5%	85.37%
	C5	90.5%	70.45%	83.78%	76.54%

Pada Tabel 4.13 dapat diketahui hasil evaluasi dengan nilai k=3 menghasilkan penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 90,5%, presisi 89,66%, *recall* 61,9% dan *f-measure* 73,24% pada kelas pelayanan (C1), nilai akurasi 92%, presisi 75%, *recall* 86,84% dan *f-measure* 80,49% pada kelas pemasangan (C2), nilai akurasi 94%, presisi 87,8%, *recall* 83,72% dan *f-measure* 85,71% pada kelas tagihan (C3), nilai akurasi 94%, presisi 83,33%, *recall* 87,5% dan *f-measure* 85,37% pada kelas lambat (C4) dan nilai akurasi 90,5%, presisi 70,45%, *recall* 83,78% dan *f-measure* 76,54% pada kelas *disconnect* (C5).

Tabel 4.14 Hasil Pengujian dengan k = 4

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	85.0%	61.54%	76.19%	68.09%
	C2	92.5%	80.0%	77.78%	78.87%
	C3	93.0%	86.49%	78.05%	82.05%
	C4	91.0%	80.49%	76.74%	78.57%
	C5	86.5%	65.71%	60.53%	63.01%
20%	C1	88.5%	70.21%	78.57%	74.16%
	C2	94.0%	85.29%	80.56%	82.86%
	C3	95.0%	91.89%	82.93%	87.18%
	C4	91.0%	79.07%	79.07%	79.07%
	C5	86.5%	64.1%	65.79%	64.94%
30%	C1	86.5%	65.96%	73.81%	69.66%
	C2	94.0%	85.29%	80.56%	82.86%
	C3	95.5%	90.0%	87.8%	88.89%
	C4	92.5%	81.82%	83.72%	82.76%
	C5	87.5%	68.57%	63.16%	65.75%
40%	C1	87.0%	67.39%	73.81%	70.45%
	C2	94.0%	85.29%	80.56%	82.86%
	C3	96.0%	90.24%	90.24%	90.24%

	C4	92.5%	80.43%	86.05%	83.15%
	C5	87.5%	69.7%	60.53%	64.79%
50%	C1	88.0%	69.57%	76.19%	72.73%
	C2	94.5%	85.71%	83.33%	84.51%
	C3	96.0%	90.24%	90.24%	90.24%
	C4	93.5%	85.71%	83.72%	84.71%
	C5	88.0%	69.44%	65.79%	67.57%
60%	C1	88.5%	71.11%	76.19%	73.56%
	C2	94.5%	85.71%	83.33%	84.51%
	C3	95.5%	88.1%	90.24%	89.16%
	C4	93.0%	85.37%	81.4%	83.33%
	C5	87.5%	67.57%	65.79%	66.67%
70%	C1	89.0%	72.73%	76.19%	74.42%
	C2	95.0%	86.11%	86.11%	86.11%
	C3	95.5%	88.1%	90.24%	89.16%
	C4	92.5%	83.33%	81.4%	82.35%
	C5	88.0%	69.44%	65.79%	67.57%
80%	C1	88.5%	70.21%	78.57%	74.16%
	C2	95.0%	86.11%	86.11%	86.11%
	C3	95.5%	90.0%	87.8%	88.89%
	C4	92.5%	85.0%	79.07%	81.93%
	C5	87.5%	67.57%	65.79%	66.67%
90%	C1	88.5%	70.21%	78.57%	74.16%
	C2	94.5%	83.78%	86.11%	84.93%
	C3	95.5%	90.0%	87.8%	88.89%
	C4	93.5%	89.47%	79.07%	83.95%
	C5	88.0%	68.42%	68.42%	68.42%

Pada Tabel 4.14 dapat diketahui hasil evaluasi dengan nilai k=4 menghasilkan penggunaan fitur terbaik sebesar 70% yang menghasilkan nilai akurasi 89%, presisi 72,73%, *recall* 76,19% dan *f-measure* 74,42% pada kelas pelayanan (C1), nilai akurasi 95%, presisi 86,11%, *recall* 86,11% dan *f-measure* 86,11% pada kelas pemasangan (C2), nilai akurasi 95,5%, presisi 88,1%, *recall* 90,24% dan *f-measure* 89,16% pada kelas tagihan (C3), nilai akurasi 92,5%, presisi 83,33%, *recall* 81,4% dan *f-measure* 82,35% pada kelas lambat (C4) dan nilai akurasi 88%, presisi 69,44%, *recall* 65,79% dan *f-measure* 67,57% pada kelas *disconnect* (C5).

Tabel 4.15 Hasil Pengujian dengan k = 5

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	88.5%	68.42%	70.27%	69.33%
	C2	96.0%	91.49%	91.49%	91.49%
	C3	96.0%	88.37%	92.68%	90.48%
	C4	91.0%	80.0%	71.79%	75.68%
	C5	86.5%	62.16%	63.89%	63.01%
20%	C1	89.0%	68.29%	75.68%	71.79%
	C2	95.5%	91.3%	89.36%	90.32%
	C3	94.5%	85.71%	87.8%	86.75%
	C4	92.0%	81.08%	76.92%	78.95%
	C5	89.0%	70.59%	66.67%	68.57%
30%	C1	90.0%	71.79%	75.68%	73.68%
	C2	96.0%	91.49%	91.49%	91.49%
	C3	94.5%	85.71%	87.8%	86.75%
	C4	93.5%	84.21%	82.05%	83.12%
	C5	91.0%	76.47%	72.22%	74.29%
40%	C1	90.0%	75.76%	67.57%	71.43%
	C2	96.0%	91.49%	91.49%	91.49%
	C3	94.0%	85.37%	85.37%	85.37%
	C4	92.5%	78.57%	84.62%	81.48%
	C5	89.5%	70.27%	72.22%	71.23%
50%	C1	89.5%	70.0%	75.68%	72.73%
	C2	95.5%	89.58%	91.49%	90.53%
	C3	95.0%	87.8%	87.8%	87.8%
	C4	93.5%	84.21%	82.05%	83.12%
	C5	90.5%	75.76%	69.44%	72.46%
60%	C1	90.0%	71.79%	75.68%	73.68%
	C2	95.0%	89.36%	89.36%	89.36%
	C3	95.0%	87.8%	87.8%	87.8%
	C4	94.0%	84.62%	84.62%	84.62%
	C5	91.0%	76.47%	72.22%	74.29%
70%	C1	90.0%	72.97%	72.97%	72.97%
	C2	96.5%	93.48%	91.49%	92.47%
	C3	94.5%	85.71%	87.8%	86.75%
	C4	94.0%	84.62%	84.62%	84.62%
	C5	90.0%	72.22%	72.22%	72.22%
80%	C1	91.0%	75.68%	75.68%	75.68%
	C2	96.5%	93.48%	91.49%	92.47%
	C3	94.0%	83.72%	87.8%	85.71%
	C4	94.0%	84.62%	84.62%	84.62%
	C5	90.5%	74.29%	72.22%	73.24%
90%	C1	90.0%	72.97%	72.97%	72.97%
	C2	97.0%	95.56%	91.49%	93.48%
	C3	95.0%	86.05%	90.24%	88.1%
	C4	93.5%	82.5%	84.62%	83.54%

	C5	90.5%	74.29%	72.22%	73.24%

Pada Tabel 4.15 dapat diketahui hasil evaluasi dengan nilai k=5 menghasilkan penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 91%, presisi 75,68%, *recall* 75,68% dan *f-measure* 75,68% pada kelas pelayanan (C1), nilai akurasi 96,5%, presisi 93,48%, *recall* 91,49% dan *f-measure* 92,47% pada kelas pemasangan (C2), nilai akurasi 94%, presisi 83,72%, *recall* 87,8% dan *f-measure* 85,71% pada kelas tagihan (C3), nilai akurasi 94%, presisi 84,62%, *recall* 84,62% dan *f-measure* 84,62% pada kelas lambat (C4) dan nilai akurasi 90,5%, presisi 74,29%, *recall* 72,22% dan *f-measure* 73,24% pada kelas *disconnect* (C5).

Berdasarkan hasil dari 5-*fold cross validation* maka akan dihitung nilai rata-rata dan standar deviasi pada masing-masing fold pada setiap fitur. Berikut merupakan hasil dari rata-rata dan standar deviasi :

Tabel 4.16 Hasil Rata-rata setiap fold

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	85.6%	63.38%	65.91%	64.12%
	C2	93.5%	84.12%	82.6%	83.27%
	C3	93.2%	84.16%	81.74%	82.76%
	C4	91.2%	79.13%	75.87%	77.41%
	C5	85.5%	63.67%	65.56%	64.28%
20%	C1	86.7%	66.9%	66.94%	66.32%
	C2	93.9%	84.48%	84.74%	84.55%
	C3	93.4%	84.29%	82.58%	83.33%
	C4	91.6%	79.5%	78.57%	78.96%
	C5	87.6%	68.94%	69.69%	69.06%
30%	C1	87.6%	69.18%	68.45%	68.48%
	C2	94.8%	87.04%	86.69%	86.77%
	C3	94.1%	85.72%	84.49%	85.06%
	C4	93.0%	82.48%	82.4%	82.39%
	C5	88.7%	71.6%	72.51%	71.9%
40%	C1	88.0%	71.61%	66.94%	68.78%

	C2	94.6%	86.75%	86.22%	86.31%
	C3	94.1%	85.18%	85.48%	85.27%
	C4	92.6%	79.99%	84.07%	81.92%
	C5	88.3%	70.93%	70.63%	70.57%
50%	C1	88.3%	72.32%	68.15%	69.65%
	C2	94.6%	86.3%	86.77%	86.42%
	C3	94.5%	85.78%	87.03%	86.35%
	C4	93.0%	82.49%	82.56%	82.43%
	C5	88.8%	71.85%	73.04%	72.2%
60%	C1	88.7%	73.44%	68.6%	70.46%
	C2	94.6%	86.47%	86.82%	86.5%
	C3	94.5%	85.72%	87.03%	86.34%
	C4	93.1%	82.59%	83.08%	82.75%
	C5	89.1%	72.34%	74.04%	73.03%
70%	C1	88.7%	73.54%	68.65%	70.57%
	C2	94.8%	86.41%	87.8%	86.99%
	C3	94.4%	85.3%	87.03%	86.13%
	C4	92.9%	82.05%	82.53%	82.23%
	C5	89.0%	72.22%	73.14%	72.54%
80%	C1	88.9%	74.11%	69.22%	71.1%
	C2	95.0%	87.2%	87.8%	87.4%
	C3	94.4%	85.67%	86.55%	86.07%
	C4	93.0%	82.47%	82.62%	82.48%
	C5	89.1%	71.93%	74.48%	73.07%
90%	C1	88.6%	72.76%	69.12%	70.48%
	C2	94.6%	86.48%	86.33%	86.3%
	C3	94.4%	85.38%	87.03%	86.14%
	C4	93.1%	83.29%	82.14%	82.57%
	C5	88.7%	71.14%	73.66%	72.24%

Tabel 4.17 Hasil Standar Deviasi setiap fold

Jumlah Fitur	Kelas	Akurasi	Presisi	Recall	F-Measure
10%	C1	2.18%	8.61%	11.11%	8.18%
	C2	2.3%	7.96%	5.65%	6.36%
	C3	2.94%	7.27%	9.05%	7.44%
	C4	0.4%	2.44%	3.32%	2.19%
	C5	1.14%	5.32%	4.5%	2.19%
20%	C1	2.66%	11.42%	11.27%	9.14%
	C2	1.53%	5.87%	3.25%	4.27%
	C3	1.59%	5.41%	4.24%	3.99%
	C4	1.07%	2.62%	3.0%	1.64%
	C5	0.97%	3.79%	6.04%	2.43%
30%	C1	2.27%	10.14%	8.82%	7.89%
	C2	1.36%	5.63%	3.54%	3.89%
	C3	1.16%	4.07%	3.64%	3.32%

	C4	0.63%	3.12%	3.2%	2.37%
	C5	1.54%	3.23%	5.82%	3.28%
40%	C1	1.9%	10.1%	7.34%	6.55%
	C2	1.71%	7.39%	3.48%	4.48%
	C3	1.69%	6.0%	2.9%	4.21%
	C4	0.37%	1.77%	2.81%	0.99%
	C5	1.44%	3.43%	6.59%	3.52%
50%	C1	1.5%	9.34%	7.99%	5.64%
	C2	1.36%	6.03%	2.66%	3.56%
	C3	1.41%	4.92%	2.63%	3.49%
	C4	0.84%	3.82%	3.67%	2.54%
	C5	1.44%	3.22%	6.57%	2.84%
60%	C1	1.44%	9.15%	8.21%	6.0%
	C2	1.66%	6.96%	2.02%	4.02%
	C3	1.14%	3.86%	2.63%	2.88%
	C4	0.8%	3.85%	3.0%	2.31%
	C5	1.59%	3.56%	6.09%	3.53%
70%	C1	1.5%	9.29%	6.92%	5.47%
	C2	1.72%	6.96%	1.96%	4.29%
	C3	1.11%	3.72%	2.63%	2.81%
	C4	0.97%	3.89%	3.16%	2.83%
	C5	1.26%	1.97%	6.16%	3.17%
80%	C1	1.66%	9.4%	7.67%	5.79%
	C2	1.58%	6.64%	1.96%	4.08%
	C3	0.97%	3.48%	2.17%	2.37%
	C4	0.84%	2.92%	3.39%	2.16%
	C5	1.2%	2.7%	5.94%	3.38%
90%	C1	1.32%	8.34%	7.48%	5.42%
	C2	1.8%	7.09%	2.83%	4.66%
	C3	1.32%	4.78%	2.63%	3.25%
	C4	0.73%	4.01%	4.03%	2.19%
	C5	1.5%	2.47%	5.6%	2.66%

4.3 Pembahasan

Berdasarkan hasil dari skenario pengujian yang telah dilakukan pada bagian sebelumnya, bahwa penggunaan seleksi fitur *Information Gain* mempengaruhi hasil klasifikasi menggunakan metode *Support Vector Machine*. Hal ini disebabkan dengan adanya seleksi fitur dapat mengurangi fitur-fitur yang digunakan. Berikut merupakan grafik hasil dari Skenario Pengujian.

Berdasarkan hasil rata-rata skenario pengujian dan standar deviasi pada Tabel 4.16 dan Tabel 4.17 maka didapatkan menghasilkan penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 88.9%, presisi 74,11%, *recall* 69,22% dan *f-measure* 71,1% pada kelas pelayanan (C1), nilai akurasi 95%, presisi 87.2%, *recall* 87.8% dan *f-measure* 87.4% pada kelas pemasangan (C2), nilai akurasi 94.4%, presisi 85,67%, *recall* 86,55% dan *f-measure* 86,07% pada kelas tagihan (C3), nilai akurasi 93%, presisi 82,47%, *recall* 82,62% dan *f-measure* 82,48% pada kelas lambat (C4) dan nilai akurasi 89,1%, presisi 71,93%, *recall* 74,48% dan *f-measure* 73,07% pada kelas *disconnect* (C5).

Dari percobaan yang telah dilakukan terdapat kesalahan atau *error* dalam mengklasifikasikan keluhan pengguna. Hal ini disebabkan terdapat kekurangan pada tahap *preprocessing* dalam mengubah kata yang kurang tepat menjadi kata yang tepat yaitu penggunaan huruf berulang “lemooott”, “sabarr” dan sebagainya. Selain itu terdapat faktor *typo* dan dua kata yang terhubung seperti “terakhirsilahkan”, “tanyaalamat” dan sebagainya. Hal ini dikarenakan tata bahasa yang digunakan pada pengguna *tweets* menggunakan kata informal.

Pengaduan yang diberikan oleh pengguna terhadap sebuah layanan merupakan sebuah amanat yang harus disampaikan dan dijalankan. Hal ini telah disampaikan pada Al-Qur’an Surah An Nisa ayat 58 :

اِنَّ اللّٰهَ يَأْمُرُكُمْ اَنْ تُؤَدُّوا الْاَمٰنٰتِ اِلٰٓى اَهْلِهَاۙ وَاِذَا حَكَمْتُمْ بَيْنَ النَّاسِ اَنْ تَحْكُمُوْا بِالْعَدْلِۗ اِنَّ اللّٰهَ نِعِمَّا يَعِظُكُمْ بِهٖۗ اِنَّ اللّٰهَ كَانَ سَمِيْعًاۢ بَصِيْرًا (٥٨)

Artinya : "Sungguh, Allah menyuruhmu menyampaikan amanat kepada yang berhak menerimanya, dan apabila kamu menetapkan hukum di antara manusia hendaknya kamu menetapkannya dengan adil. Sungguh, Allah sebaik-baik yang memberi pengajaran kepadamu. Sungguh Sungguh, Allah Maha Mendengar, Maha Melihat" (Q.S. An-Nisa : 58).

Dalam tafsir Ibnu Katsir jilid 2 menjelaskan bahwa ayat ini tentang perintah untuk menunaikan amanat. Salah satu amanat yang harus dijalankan yaitu berupa hak-hak manusia seperti titipan. Selain itu ayat ini merupakan bersikap adil dalam menetapkan hukum diantara manusia.

Dalam hadis yang diriwayatkan oleh Bukhari, Rasulullah SAW bersabda:

آيَةُ الْمُنَافِقِ ثَلاَثٌ إِذَا حَدَّثَ كَذَبَ ، وَإِذَا وَعَدَ أَخْلَفَ ، وَإِذَا اؤْتُمِنَ خَانَ

Artinya : "Tanda-tanda orang munafik ada tiga: jika berbicara ia berbohong, jika berjanji ia mengingkari, dan jika diberi amanat ia berkhianat" (H.R Bukhari).

Berdasarkan hadits diatas dapat diketahui bahwa dalam menanggapi pengaduan yang dilakukan oleh pelanggan merupakan amanat yang harus dijalankan sesuai jenis keluhan. Dengan adanya sistem klasifikasi pengaduan berkaitan dengan pengelompokan pengaduan, diharapkan mampu mempermudah dalam klasifikasi pengaduan sehingga dapat ditetapkan langkah yang harus dilakukan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah dilakukan penelitian terhadap klasifikasi pengaduan layanan pengguna IndiHome pada media sosial Twitter menggunakan metode *Support Vector Machine* dengan seleksi fitur *Information Gain* menghasilkan penggunaan fitur terbaik sebesar 80% yang menghasilkan nilai akurasi 88.9%, presisi 74,11%, *recall* 69,22% dan *f-measure* 71,1% pada kelas pelayanan (C1), nilai akurasi 95%, presisi 87.2%, *recall* 87.8% dan *f-measure* 87.4% pada kelas pemasangan (C2), nilai akurasi 94.4%, presisi 85,67%, *recall* 86,55% dan *f-measure* 86,07% pada kelas tagihan (C3), nilai akurasi 93%, presisi 82,47%, *recall* 82,62% dan *f-measure* 82,48% pada kelas lambat (C4) dan nilai akurasi 89,1%, presisi 71,93%, *recall* 74,48% dan *f-measure* 73,07% pada kelas *disconnect* (C5).

Berdasarkan penggunaan fitur pada skenario pengujian dapat diketahui bahwa persentase penggunaan fitur pada *Information Gain* mempengaruhi hasil klasifikasi sistem dengan metode *Support Vector Machine.* Sehingga dapat disimpulkan bahwa seleksi fitur yang berlebihan akan berpengaruh pada hasil klasifikasi.

5.2 Saran

Berdasarkan hasil percobaan pada penelitian ini diharapkan bagi penelitian selanjutnya dapat meningkatkan hasil klasifikasi yang lebih

akurat. Maka dari itu penulis memiliki saran untuk pengembangan penelitian dimasa yang akan datang, yaitu sebagai berikut :

1. Menambah jumlah kata informal pada kamus sehingga dapat mempengaruhi pembacaan *term* pada sebuah dokumen.
2. Melakukan percobaan menggunakan kernel polynomial, gaussian, *sigmoid* ataupun kernel lainnya.
3. Melakukan percobaan menggunakan metode seleksi fitur lain seperti algoritma genetika, *Chi Square, Mutual Information* dan sebagainya

DAFTAR PUSTAKA

Adankon M.M., Cheriet M. 2015 *Support Vector Machine*. Springer, Boston

Aini, S. H. A., Sari, Y. A. dan Arwan, A. 2018. *Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X.

Anandarajan, M., Hill C. dan Nolan T. 2019 *Text Preprocessing. In: Practical Text Analytics. Advances in Analytics and Data Science, vol 2*. Switzerland : Springer.

Asrol, M., Papilo, P. dan Gunawan, F.E. 2021 *Support Vector Machine with K-fold Validation to Improve the Industry's Sustainability Performance Classification*, Procedia Computer Science 179 : 854-862

Awad, M. dan Khanna, R. *Efficient Learning Machines.* New York : Apress Media.

Behrad, A., Khoddami, M. dan Salehpour, M. 2010. *A novel framework for farsi and latin script identification and farsi handwritten digit recognition. Journal of automatic control*, 20(1), 17-25.

Cai, J., Luo, J., Wang, S. dan Yang, S. 2018. Feature Selection in Machine Learning: a New Perspective. Neurocomputing : 300, 70-79.

Chormunnge, S. dan Jena, S. 2016. *Effiecient Features Subset Selection Algorithm for High Dimensional Data*. International Journal of Electrical and Computer Engineering (IJECE) 6 : 1880-1888.

Colaco, S., Kumar, S., Tamang, A. dan Biju, V.G. 2019. *A Review on Feature Selection Algorithms.* Springer Nature : 133-154.

Daeli, N. O. F. dan Adiwijaya, A. (2020). *Sentiment Analysis on Movie Reviews using Information Gain and K-Nearest Neighbor*. Journal of Data Science and Its Applications : 3(1), 1-7.

Fan, K., Wang, P., Hu, Y. dan Dou, B. 2017. *Fall detection via human posture representation and support vector machine*. International journal of distributed sensor networks, 13(5), 1550147717707418.

Fatmawati dan Affandes, M. 2017. *Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk).* Jurnal CoreIT 3 : 24-30.

Fitriyah, N., Warsito, B., dan Di Asih, I. M. 2020. *Analisis Sentimen Gojek pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM)*. Jurnal Gaussian, 9(3) : 376-390.

Gautam, G., Choudhary, K., Chatterjee, S. dan Kolekar, M.H. 2017. *Facial expression recognition using Krawtchouk moments and support vector machine classifier*. International Conference on Image Information Processing (ICIIP) (pp. 1-6). IEEE

Gao, L., Ye, M., Lu, X. dan Huang, D. 2017. *Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification*. Genomics, proteomics & bioinformatics, 15(6), 389-395.

Han, J., Kamber, M. dan Pei, J. 2012. *Data Mining Trends and Research Frontiers*. Amsterdam : Elsevier Inc.

Haq, M. I. U., Li, Q. dan Hassan, S.(2019. *Text Mining Techniques to Capture Facts for Cloud Computing Adoption and Big Data Processing*. IEEE Access, 7, 162254-162267.

Ibithel, B. L., Lobna, H. dan Maher, B.J. 2018. *A Semantic for Tweet Categorization*. Procedia Computer Science 126 : 335-344.

IndiHome. 2020. *Apa itu IndiHome?*. https://www.IndiHome.co.id/pusat-bantuan/kenali-IndiHome/apa-itu-IndiHome (diunduh pada tanggal 30 Juli 2020).

Islam, M. S., Jubayer, F. E. M. dan Ahmed, S. I. 2017. *A Support Vector Machine Mixed with TF-IDF Algorithm to Categorize Bengali Document*. International Conference on Electrical, Computer And Communication Engineering (ECCE) (pp. 191-196). IEEE.

Khomsah, S. dan Aribowo, A. S. 2020. *Model Text-Preprocessing Komentar Youtube Dalam Bahasa Indonesia*. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi) : 4(4), 648-654.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihok, G., & Den Hartog, D. N. 2018. *Text Classification for Organizational Researchers: a Tutorial. Organizational Research Methods* : 21(3), 766-799.

Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G. dan Den Hartog, D. N. 2018. *Text Mining in Organizational Research*. Organizational research methods : 21(3), 733-765.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., dan Liu, H. 2017. *Feature selection: A data perspective*. ACM Computing Surveys (CSUR) : 50(6), 1-45.

Mafarja, M. M. dan Mirjalili, S. 2018. *Whale Optimization Approaches for Wrapper Feature Selection*. Applied Soft Computing : 62, 441-453.

Marcot, B. G. dan Hanea, A. M. 2020. *What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?*. Computational Statistics, 1-23.

Maulina, D. dan Sagara, R. 2018. *Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear dengan Pembobotan Term Frequency–Inverse Document Frequency*. Jurnal Mantik Penusa, 2(1).

Mohammad, F. 2018. *Is Preprocessing of Text Really Worth your Time for Online Comment Classification?.* arXiv:1806.02908.

Nalepa, J. dan Kawulok, M. 2019. *Selecting Training sets for Support Vector Machines: a Review*. Artif Intell Rev 52 : 857-900.

Oza, K. S. dan Naik, P.G. 2016. *Prediction of Online Lectures Popularity : A Text Mining Approach*. Procedia Computer Science 92 : 468 – 471.

Pratama, E. S. dan Trilaksono, B.R. 2015. *Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)*. Jurnal Edukasi dan Penelitian Informatika (JEPIN) 1 : 53-59.

Qaiser, S. dan Ali, R. 2018. *Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents*. International Journal of Computer Applications : 181(1), 25-29.

Salloum, S. A., Al-Emran, M., Monem, A. A. dan Shaalan, K. 2018. Using Text Mining Techniques for Extracting Information from Research Articles. Springer Nature : 373-397.

Samant, S. S., Murthy. N. L.B. dan Malapati, A.A. 2019. *Improving Term Weighting Schemes for Short Text Classification in Vector Space Model.* IEEE Access 7: 166578-166592.

Sihwi, S. W., Jati, I. P. dan Anggrainingsih, R. 2018. *Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier*. International Seminar on Application for Technology of Information and Communication (pp. 190-195). IEEE.

Simske, S. 2019. *Meta-Analytics: Consensus Approaches and System Patterns for Data Analysis*. Amsterdam : Elsevier Inc.

Somantri, O. dan Apriliani, D. 2018. *Support Vector Machine Berbasis Feature Selection untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal*. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) 5 : 537-548.

Statista, 2020. *Leading countries based on number of Twitter users as of July 2020*. https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ (diunduh pada tanggal 7 Agustus 2020).

Suyanto. 2019. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung : Informatika Bandung.

Tineges, R., Triayudi, A., dan Sholihati, I. D. 2020. *Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)*. Jurnal Media Informatika Budidarma : 4(3), 650-658.

Uysal, A. K. dan Gunal, S. 2014. *The Impact Of Preprocessing On Text Classification. Information Processing & Management*. Information Processing and Management 50 : 104-112.

Vidya, N. A., Fanany, M.I. dan Budi, I. 2015. *Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers*. Procedia Computer Science 72:519-526.

Xia, H., Yang, Y., Pan, X., Zhang, Z., dan An, W. 2020. Sentiment Analysis for Online Reviews Using Conditional Random Fields and Support Vector Machines. Electronic Commerce Research 20(2) : 343-360.

Wong, T. T. 2015. *Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation*. Pattern Recognition 48(9) : 2839-2846.

Wongso, R., Luwinda, F.A., Trisnajaya, B.C., Rusli, O. Dan Rudy. 2017. *News Article Text Classification in Indonesian Language*. Procedia Computer Science 116 : 137-143.

Yahav, I., Shehory, O. dan Schwartz, D. 2018. *Comments Mining with TF-IDF: the Inherent Bias and Its Remova*l. IEEE Transactions on Knowledge and Data Engineering: 31(3), 437-450.

Zhu, Z., Liang, J., Li, D., Yu, H., dan Liu, G. 2019. *Hot Topic Detection Based on a Refined TF-IDF Algorithm*. IEEE Access : 7, 26996-27007

LAMPIRAN

Hasil Prediksi dengan k = 1

No.	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	3	3	3	3	3	3	3	3	3
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
11	1	5	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	4	4	4	4	4	4	4	4	4
18	1	1	3	3	3	3	3	3	3	3
19	1	1	1	1	1	1	1	1	1	1
20	1	5	1	1	1	5	5	1	5	5
21	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	1	1	1	1	1	1
26	1	1	1	1	1	1	1	1	1	1
27	1	1	5	5	5	5	5	5	5	5
28	1	1	5	5	5	5	5	5	5	5
29	1	1	1	1	1	1	1	1	1	1
30	1	1	1	1	1	1	1	1	1	1
31	1	1	1	1	1	1	1	1	1	1
32	1	4	4	4	5	5	5	5	5	5
33	1	4	2	4	4	4	4	4	4	4
34	1	1	2	2	2	2	1	2	2	2
35	1	1	1	1	1	1	1	1	1	1
36	1	5	5	5	5	5	5	5	5	5
37	1	1	1	1	1	1	1	1	1	1
38	1	5	5	1	1	1	1	1	1	1
39	1	1	1	1	1	1	1	1	1	1
40	1	5	4	5	4	4	4	4	4	4
41	1	5	5	5	5	5	5	5	5	5
42	1	1	1	1	1	1	1	1	1	1
43	1	1	1	1	1	1	1	1	1	1

44	1	1	1	1	1	1	1	1	1	1
45	1	1	1	1	1	4	4	4	4	1
46	2	2	2	2	2	2	2	2	2	2
47	2	2	2	2	2	2	2	2	2	2
48	2	2	2	2	2	2	2	2	2	2
49	2	2	2	2	2	2	2	2	2	2
50	2	2	2	2	2	2	2	2	2	2
51	2	2	2	2	2	2	2	2	2	2
52	2	2	2	2	2	2	2	2	2	2
53	2	2	2	2	2	2	2	2	2	2
54	2	2	2	2	2	2	2	2	2	2
55	2	2	2	2	2	2	2	2	2	2
56	2	2	2	2	2	2	2	2	2	2
57	2	2	2	2	2	2	2	2	2	2
58	2	2	2	2	2	2	2	2	2	2
59	2	2	2	2	2	2	2	2	2	2
60	2	2	2	2	2	2	2	2	2	2
61	2	2	2	2	2	2	2	2	2	2
62	2	2	2	2	2	2	2	2	2	2
63	2	2	2	2	2	2	2	2	2	2
64	2	2	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2	2	2
66	2	2	2	2	2	2	2	2	2	2
67	2	2	2	2	2	2	2	2	2	2
68	2	2	2	2	2	2	2	2	2	2
69	2	2	2	2	2	2	2	2	2	2
70	2	1	2	2	2	2	2	2	2	2
71	2	2	2	2	2	2	2	2	2	2
72	2	2	1	2	2	2	2	2	2	2
73	2	2	2	4	4	4	4	4	4	4
74	2	1	1	1	1	1	1	1	1	1
75	2	2	2	2	2	2	2	2	2	2
76	2	1	4	3	3	3	3	3	3	3
77	2	2	2	2	2	2	2	2	2	2
78	2	3	3	3	3	3	3	3	3	3
79	2	2	2	2	2	2	2	2	2	2
80	2	2	2	2	2	2	2	2	2	2
81	2	5	5	1	4	4	4	4	4	4
82	2	2	2	2	2	2	2	2	2	2
83	3	3	3	3	3	3	3	3	3	3
84	3	3	3	3	3	3	3	3	3	3
85	3	3	3	3	3	3	3	3	3	3
86	3	3	3	3	3	3	3	3	3	3
87	3	3	3	3	3	3	3	3	3	3
88	3	3	3	3	3	3	3	3	3	3
89	3	3	3	3	3	3	3	3	3	3
90	3	3	3	3	3	3	3	3	3	3
91	3	3	3	3	3	3	3	3	3	3

92	3	3	3	3	3	3	3	3	3	3
93	3	3	3	3	3	3	3	3	3	3
94	3	3	3	3	3	3	3	3	3	3
95	3	3	1	1	1	1	1	1	3	3
96	3	3	3	3	3	3	3	3	3	3
97	3	3	3	3	3	3	3	3	3	3
98	3	3	3	3	3	3	3	3	3	3
99	3	3	3	3	3	3	3	3	3	3
100	3	3	3	3	3	3	3	3	3	3
101	3	3	3	3	3	3	3	3	3	3
102	3	2	2	2	2	2	2	2	2	2
103	3	3	3	3	3	3	3	3	3	3
104	3	3	3	3	3	3	3	3	3	3
105	3	3	3	3	3	3	3	3	3	3
106	3	3	3	3	3	3	3	3	3	3
107	3	3	3	3	3	3	3	3	3	3
108	3	3	3	3	3	3	3	3	3	3
109	3	3	3	3	3	3	3	3	3	3
110	3	3	3	3	3	3	3	3	3	3
111	3	3	3	3	3	3	3	3	3	3
112	3	2	2	2	2	2	2	2	2	2
113	3	3	3	3	3	3	3	3	3	3
114	3	3	5	5	5	3	3	3	5	5
115	3	3	3	3	3	3	3	3	3	3
116	3	4	1	4	4	4	4	4	4	4
117	3	3	3	3	3	3	3	3	3	3
118	3	3	3	3	3	3	3	3	3	3
119	3	3	3	3	3	3	3	3	3	3
120	4	2	4	4	4	4	4	4	4	4
121	4	4	4	4	4	4	4	4	4	4
122	4	5	5	5	4	4	4	4	4	4
123	4	4	4	4	4	4	4	4	4	4
124	4	4	4	4	4	4	4	4	4	4
125	4	3	3	3	3	3	3	3	3	3
126	4	4	4	4	4	4	4	4	4	4
127	4	4	4	4	4	4	4	4	4	4
128	4	4	4	4	4	4	4	4	4	4
129	4	5	5	4	4	4	4	5	4	4
130	4	4	4	4	4	4	4	4	4	4
131	4	4	4	4	4	4	4	4	4	4
132	4	4	4	4	4	4	4	4	4	4
133	4	5	4	5	4	5	5	5	5	5
134	4	4	4	4	4	4	4	4	4	4
135	4	4	4	4	4	4	4	4	4	4
136	4	4	4	4	4	4	4	4	4	4
137	4	4	4	4	4	4	4	4	4	4
138	4	4	4	4	4	4	4	4	4	4
139	4	5	4	4	4	4	4	4	4	4

140	4	4	4	4	4	4	4	4	4	4
141	4	1	1	1	1	1	1	1	1	1
142	4	4	4	4	4	4	4	4	4	4
143	4	4	4	4	4	4	4	4	4	4
144	4	5	4	5	5	5	5	5	5	5
145	4	5	3	3	5	5	5	5	5	5
146	4	4	4	4	4	4	4	4	4	4
147	4	4	4	4	4	4	4	4	4	4
148	4	4	4	4	4	4	4	4	4	4
149	4	4	4	4	4	4	4	4	4	4
150	4	4	4	4	4	4	4	4	4	4
151	4	4	4	4	4	4	4	4	4	4
152	4	4	4	4	4	4	4	4	4	4
153	4	4	2	5	4	4	4	4	4	4
154	4	4	4	4	4	4	4	4	4	4
155	4	1	5	5	5	5	5	5	5	5
156	5	5	1	1	1	1	5	1	5	5
157	5	1	1	1	1	1	1	1	1	1
158	5	1	1	1	2	2	2	2	2	2
159	5	5	5	5	5	5	5	5	5	5
160	5	5	5	5	5	5	5	5	5	5
161	5	5	5	5	5	5	5	5	5	5
162	5	5	5	5	5	5	5	5	5	5
163	5	5	5	5	5	5	5	5	5	5
164	5	5	5	5	5	5	5	5	5	5
165	5	5	5	5	5	5	5	5	5	5
166	5	1	1	1	1	1	1	1	1	1
167	5	1	1	1	1	1	1	1	1	1
168	5	5	3	5	5	5	5	5	5	5
169	5	2	5	5	5	5	5	5	5	5
170	5	4	4	4	4	4	4	4	4	4
171	5	5	5	5	5	5	5	5	5	5
172	5	2	2	5	5	5	5	5	5	5
173	5	5	5	5	1	1	1	1	1	1
174	5	5	5	5	5	5	5	5	5	5
175	5	5	5	5	5	5	5	5	5	5
176	5	5	5	5	5	5	5	5	5	5
177	5	5	5	4	4	4	4	4	5	4
178	5	1	1	1	1	1	1	1	1	1
179	5	5	5	5	5	5	5	5	5	5
180	5	4	5	5	5	5	5	5	5	5
181	5	5	5	5	5	5	5	5	5	5
182	5	5	5	5	5	5	5	5	5	5
183	5	5	5	5	5	5	5	5	5	5
184	5	4	1	1	1	1	1	1	1	1
185	5	5	5	5	5	5	5	5	5	5
186	5	1	5	5	5	5	5	5	5	5
187	5	1	1	1	1	1	1	1	1	1

188	5	5	5	5	5	5	5	5	5	5
189	5	1	1	5	5	5	5	5	5	5
190	5	5	5	5	5	5	5	5	5	5
191	5	1	1	1	1	1	1	1	1	1
192	5	5	5	5	5	5	5	5	5	5
193	5	5	5	5	5	5	5	5	5	5
194	5	4	4	4	4	4	4	4	4	4
195	5	5	5	5	5	5	5	5	5	5
196	5	5	5	5	5	5	5	5	5	5
197	5	5	5	5	5	5	5	5	5	5
198	5	5	5	5	5	5	5	5	5	5
199	5	1	1	5	5	5	5	5	5	1
200	5	1	1	1	1	1	1	1	1	1

Hasil Prediksi dengan k = 2

No.	**Aktual**	**Jumlah Fitur**								
		10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	2	5	5	5	5	5	5	5	5
4	1	2	2	2	2	2	2	2	2	2
5	1	5	5	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	3	3	3	3	3	3	3	3	3
9	1	3	3	3	3	3	3	3	3	3
10	1	3	3	3	3	3	3	3	3	3
11	1	3	3	3	3	3	3	3	3	3
12	1	1	3	5	1	1	1	1	1	1
13	1	5	5	5	5	5	5	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	5	5	5	5	5	5	5	4	5
18	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1
20	1	4	4	4	4	4	4	4	4	4
21	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	1	4	4	5	5	5	5	5	5	5
26	1	5	5	5	5	5	5	5	5	5
27	1	1	1	1	1	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1
29	1	5	5	5	5	5	5	5	5	5
30	1	1	1	1	1	1	1	1	1	1

31	1	1	1	1	1	1	1	1	1	1
32	1	5	5	5	5	5	5	5	5	5
33	1	3	3	3	3	3	3	3	3	3
34	1	3	3	2	3	2	2	2	2	2
35	2	2	2	2	2	2	2	2	2	2
36	2	2	2	2	2	2	2	2	2	2
37	2	2	2	2	2	2	2	2	2	2
38	2	2	2	2	2	2	2	2	2	2
39	2	2	2	2	2	2	2	2	2	2
40	2	2	2	2	2	2	2	2	2	2
41	2	2	2	2	2	2	2	2	2	2
42	2	2	2	2	2	2	2	2	2	2
43	2	2	2	2	2	2	2	2	2	2
44	2	2	2	2	2	2	2	2	2	2
45	2	2	2	2	2	2	2	2	2	2
46	2	3	5	3	3	3	2	3	3	3
47	2	5	2	2	2	2	2	2	2	2
48	2	1	1	1	1	1	1	1	1	1
49	2	1	1	1	1	1	1	1	1	1
50	2	2	2	2	2	2	2	2	2	2
51	2	2	2	2	2	2	2	2	2	2
52	2	3	3	2	3	3	3	2	2	3
53	2	2	2	2	2	2	2	2	2	2
54	2	2	2	2	2	2	2	2	2	2
55	2	2	2	2	2	2	2	2	2	2
56	2	1	2	2	2	2	2	2	2	2
57	2	2	2	2	2	2	2	2	2	2
58	2	2	2	2	2	2	2	2	2	2
59	2	2	2	2	2	2	2	2	2	2
60	2	2	2	2	2	2	2	2	2	2
61	2	2	2	2	2	2	2	2	2	2
62	2	2	2	2	2	2	2	2	2	2
63	2	2	2	2	2	2	2	2	2	2
64	2	2	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2	2	2
66	2	2	2	2	2	2	2	2	2	2
67	2	1	1	5	5	5	5	5	5	5
68	2	2	2	2	2	2	2	2	2	2
69	2	2	2	2	2	2	2	2	2	2
70	2	2	2	2	2	2	2	2	2	5
71	2	2	2	2	2	2	2	2	2	2
72	2	2	2	2	2	2	2	2	2	2
73	2	2	2	2	2	2	2	2	2	2
74	2	2	2	2	2	2	2	2	2	2
75	2	1	1	1	3	1	1	1	1	1
76	2	2	2	2	2	2	2	2	2	2
77	3	3	3	3	3	3	3	3	3	3
78	3	3	3	3	3	3	3	3	3	3

79	3	3	3	3	3	3	3	3	3	3
80	3	3	3	3	3	3	3	3	3	3
81	3	1	1	1	1	3	3	3	3	3
82	3	3	3	3	3	3	3	3	3	3
83	3	3	3	3	3	3	3	3	3	3
84	3	3	3	3	3	3	3	3	3	3
85	3	3	2	2	2	2	2	2	2	2
86	3	3	3	3	3	3	3	3	3	3
87	3	1	1	1	3	3	3	3	3	3
88	3	3	3	3	3	3	3	3	3	3
89	3	3	3	3	3	3	3	3	3	3
90	3	3	3	3	3	3	3	3	3	3
91	3	3	3	3	3	3	3	3	3	3
92	3	3	3	3	3	3	3	3	3	3
93	3	3	3	3	3	3	3	3	3	3
94	3	3	3	3	3	3	3	3	3	3
95	3	5	3	3	3	3	3	3	3	3
96	3	3	3	3	3	3	3	3	3	3
97	3	3	3	3	3	3	3	3	3	3
98	3	3	3	3	3	3	3	3	3	3
99	3	3	3	3	3	3	3	3	3	3
100	3	3	3	3	3	3	3	3	3	3
101	3	3	3	3	3	3	3	3	3	3
102	3	3	3	3	3	3	3	3	3	3
103	3	3	3	3	3	3	3	3	3	3
104	3	4	4	4	4	4	4	4	4	4
105	3	5	2	5	5	5	5	2	5	2
106	3	3	3	3	3	3	3	3	3	3
107	3	3	3	3	3	3	3	3	3	3
108	3	1	1	1	1	1	1	1	1	1
109	3	3	3	3	3	3	3	3	3	3
110	3	3	3	3	3	3	3	3	3	3
111	3	1	1	1	1	1	1	1	1	1
112	3	3	3	3	3	3	3	3	3	3
113	3	4	3	3	3	3	3	3	3	3
114	3	1	1	1	1	1	1	1	1	1
115	4	5	5	1	1	1	1	1	1	1
116	4	4	4	4	4	4	4	4	4	4
117	4	4	4	4	4	4	4	4	4	4
118	4	5	2	5	5	5	5	5	5	5
119	4	5	5	5	5	5	5	5	5	5
120	4	3	1	4	4	4	4	4	4	4
121	4	4	4	4	4	4	4	4	4	4
122	4	4	4	4	4	4	4	4	4	4
123	4	4	5	5	5	5	5	5	5	5
124	4	4	4	4	4	4	4	4	4	4
125	4	4	4	4	4	4	4	4	4	4
126	4	4	4	4	4	4	4	4	4	4

127	4	4	4	4	4	4	4	4	4	4
128	4	4	4	4	4	4	4	4	4	4
129	4	4	4	4	4	4	4	4	4	4
130	4	1	1	1	5	5	5	5	4	5
131	4	4	4	4	4	4	4	4	4	4
132	4	4	4	4	4	4	4	4	4	4
133	4	4	4	4	4	4	4	4	4	4
134	4	4	4	4	4	4	4	4	4	4
135	4	4	4	4	4	4	4	4	4	4
136	4	4	4	4	4	4	4	4	4	4
137	4	1	1	5	1	1	1	1	1	1
138	4	4	4	4	4	4	4	4	4	4
139	4	3	3	3	3	3	3	3	3	3
140	4	4	4	4	4	4	4	4	4	4
141	4	4	4	4	4	4	4	4	4	4
142	4	5	5	4	5	5	5	5	5	5
143	4	4	1	4	4	4	4	4	4	4
144	4	4	4	4	4	4	4	4	4	4
145	4	4	4	4	4	4	4	4	4	4
146	4	5	5	5	5	5	5	5	5	5
147	4	4	4	4	4	4	4	4	4	4
148	4	4	4	4	4	4	4	4	4	4
149	4	4	4	4	4	4	4	4	4	4
150	4	4	4	4	4	4	4	4	4	4
151	4	4	4	4	4	4	4	4	4	4
152	4	4	4	4	4	4	4	4	4	4
153	4	4	4	4	4	4	4	4	4	4
154	4	4	4	4	4	4	4	4	4	4
155	4	4	4	4	4	5	4	4	5	5
156	4	4	4	4	4	4	4	4	4	4
157	5	5	5	5	5	5	5	5	5	5
158	5	1	1	1	1	5	5	5	5	5
159	5	5	5	5	5	5	5	5	5	5
160	5	5	5	5	5	5	5	5	5	5
161	5	5	5	5	5	5	5	5	5	5
162	5	5	5	5	5	5	5	5	5	5
163	5	5	5	5	5	5	5	5	5	5
164	5	5	5	5	5	5	5	5	5	5
165	5	5	5	5	5	5	5	5	5	5
166	5	5	5	5	5	5	5	5	5	5
167	5	5	5	5	5	5	5	5	5	5
168	5	4	4	4	4	4	4	4	4	4
169	5	1	1	1	1	1	1	1	1	1
170	5	5	5	5	5	5	5	5	5	5
171	5	1	5	5	5	5	5	5	5	5
172	5	1	5	5	5	5	5	5	5	5
173	5	1	1	1	1	5	5	5	5	5
174	5	3	3	3	3	3	3	3	3	3

175	5	5	4	5	5	5	5	5	5	5
176	5	5	5	5	5	3	3	3	5	3
177	5	1	1	1	1	1	1	1	1	1
178	5	1	1	1	1	1	1	1	1	1
179	5	4	4	5	1	5	5	1	1	1
180	5	5	5	5	5	5	5	5	5	5
181	5	5	5	5	5	5	5	5	5	5
182	5	5	5	5	5	5	5	5	5	5
183	5	5	1	1	1	1	1	1	1	1
184	5	5	5	5	5	5	5	5	5	5
185	5	5	5	5	5	5	5	5	5	5
186	5	2	2	2	4	4	4	4	4	4
187	5	5	5	5	5	5	5	5	5	5
188	5	5	5	5	5	5	5	5	5	5
189	5	1	4	4	4	4	4	4	4	4
190	5	5	5	5	5	5	5	5	5	5
191	5	5	5	5	5	5	5	5	5	5
192	5	5	5	5	5	5	5	5	5	5
193	5	5	5	5	5	5	5	5	5	5
194	5	5	5	5	4	5	5	5	5	5
195	5	3	5	5	5	5	5	5	5	5
196	5	5	5	5	5	5	5	5	5	5
197	5	5	5	5	5	5	5	5	5	5
198	5	5	5	5	5	5	5	5	5	5
199	5	4	4	4	4	4	4	4	4	4
200	5	5	5	5	5	5	5	5	5	5

Hasil Prediksi dengan k = 3

No.	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	5	5	5	5	5	5	5	5	5
5	1	5	1	1	1	1	1	1	1	1
6	1	4	5	5	2	5	2	2	5	5
7	1	5	5	5	5	5	5	5	5	5
8	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12	1	5	5	5	5	5	5	5	5	5
13	1	1	1	1	1	1	1	1	1	1
14	1	5	1	1	1	1	1	1	1	1
15	1	5	1	1	1	1	1	1	1	1
16	1	5	5	5	5	5	5	5	5	5
17	1	1	1	1	1	1	1	1	1	1

18	1	1	1	1	1	1	1	1	1	1
19	1	5	5	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1
21	1	5	4	4	4	4	4	4	4	4
22	1	1	1	1	1	1	1	1	1	1
23	1	1	2	2	2	2	2	2	2	2
24	1	2	2	2	2	2	2	2	2	2
25	1	1	4	1	1	1	1	1	1	1
26	1	1	1	1	1	1	1	1	1	1
27	1	2	2	2	2	2	2	2	2	2
28	1	4	4	4	4	4	4	4	4	4
29	1	1	1	1	1	1	1	1	1	1
30	1	1	1	1	1	1	1	1	1	1
31	1	1	1	1	1	1	1	1	1	1
32	1	5	1	1	1	1	1	1	1	1
33	1	2	2	2	2	2	2	2	2	2
34	1	2	2	2	2	2	2	2	2	2
35	1	2	2	2	2	2	2	2	2	2
36	1	1	2	1	2	2	2	2	2	2
37	1	1	1	1	1	1	1	1	1	1
38	1	1	1	1	1	1	1	1	1	1
39	1	3	3	1	1	1	1	1	1	1
40	1	1	1	4	4	3	3	3	3	3
41	1	5	1	1	1	1	1	1	1	1
42	1	2	2	2	2	2	2	2	2	2
43	2	5	3	3	3	3	3	3	3	3
44	2	2	2	2	2	2	2	2	2	2
45	2	2	2	2	2	2	2	2	2	2
46	2	2	2	2	2	2	2	2	2	2
47	2	2	2	2	2	2	2	2	2	2
48	2	2	2	2	2	2	2	2	2	2
49	2	2	2	2	2	2	2	2	2	2
50	2	2	2	2	2	2	2	2	2	2
51	2	2	2	2	2	2	2	2	2	2
52	2	5	5	5	5	5	5	5	5	5
53	2	2	2	2	2	2	2	2	2	2
54	2	2	2	2	2	2	2	2	2	2
55	2	2	2	2	2	2	2	2	2	2
56	2	2	2	2	2	2	2	2	2	2
57	2	2	3	3	3	3	3	3	3	3
58	2	2	2	2	2	2	2	2	2	2
59	2	1	2	2	2	2	2	2	2	1
60	2	2	2	2	2	2	2	2	2	2
61	2	2	2	2	2	2	2	2	2	2
62	2	3	3	3	3	3	3	3	3	3
63	2	1	1	2	2	2	2	2	2	2
64	2	5	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2	2	2

66	2	3	3	2	2	2	2	2	2	2
67	2	2	2	2	2	2	2	2	2	2
68	2	2	2	2	2	2	2	2	2	2
69	2	2	2	2	2	2	2	2	2	2
70	2	2	2	2	2	2	2	2	2	2
71	2	2	2	2	2	2	2	2	2	2
72	2	2	2	2	2	2	2	2	2	2
73	2	2	2	2	2	2	2	2	2	2
74	2	2	2	2	2	2	2	2	2	2
75	2	3	2	2	2	2	2	2	2	2
76	2	4	4	4	4	4	4	4	4	4
77	2	2	2	2	2	2	2	2	2	2
78	2	2	2	2	2	2	2	2	2	2
79	2	2	2	2	2	2	2	2	2	2
80	2	2	2	2	2	2	2	2	2	2
81	3	3	3	3	3	3	3	3	3	3
82	3	3	3	3	3	3	3	3	3	3
83	3	3	3	3	3	3	3	3	3	3
84	3	5	5	1	1	1	1	1	1	1
85	3	3	3	3	3	3	3	3	3	3
86	3	3	3	3	3	3	3	3	3	3
87	3	3	3	3	3	3	3	3	3	3
88	3	3	3	3	3	3	3	3	3	3
89	3	3	3	3	3	3	3	3	3	3
90	3	3	3	3	3	3	3	3	3	3
91	3	3	3	3	3	3	3	3	3	3
92	3	1	1	1	3	3	3	3	3	3
93	3	3	3	3	3	3	3	3	3	3
94	3	3	3	3	3	3	3	3	3	3
95	3	2	3	3	3	3	3	3	3	3
96	3	3	3	3	3	3	3	3	3	3
97	3	3	3	3	3	3	3	3	3	3
98	3	3	3	3	3	3	3	3	3	3
99	3	3	3	3	3	3	3	3	3	3
100	3	3	3	3	3	3	3	3	3	3
101	3	2	3	3	3	3	3	3	3	3
102	3	3	3	3	3	3	3	3	3	3
103	3	3	3	3	3	3	3	3	3	3
104	3	3	3	3	3	3	3	3	3	3
105	3	3	3	3	3	3	3	3	3	3
106	3	2	2	2	2	2	2	2	2	2
107	3	3	3	3	3	3	3	3	3	3
108	3	3	4	3	3	3	3	3	3	3
109	3	3	3	3	3	3	3	3	3	3
110	3	5	5	5	5	5	5	5	5	5
111	3	5	5	5	5	5	5	5	5	5
112	3	3	3	3	3	3	3	3	3	3
113	3	5	5	5	5	5	5	5	5	5

114	3	4	5	5	5	5	5	5	5	5
115	3	2	4	3	3	3	3	3	3	3
116	3	4	3	3	3	3	3	3	3	3
117	3	4	4	4	4	4	4	4	4	4
118	3	4	3	3	3	3	3	3	3	3
119	3	3	3	3	3	3	3	3	3	3
120	3	3	3	3	3	3	3	3	3	3
121	3	3	3	3	3	3	3	3	3	3
122	3	3	3	3	3	3	3	3	3	3
123	3	3	3	3	3	3	3	3	3	3
124	4	4	4	4	4	4	4	4	4	4
125	4	4	4	4	4	4	4	4	4	4
126	4	5	5	5	5	5	5	5	5	5
127	4	5	2	2	2	2	2	2	2	2
128	4	4	4	4	4	4	4	4	4	4
129	4	4	4	4	4	4	4	4	4	4
130	4	4	4	4	4	4	4	4	4	4
131	4	4	4	4	4	4	4	4	4	4
132	4	4	4	4	4	4	4	4	4	4
133	4	4	4	4	4	4	4	4	4	4
134	4	4	4	4	4	4	4	4	4	4
135	4	3	3	3	3	3	3	3	3	3
136	4	4	4	4	4	4	4	4	4	4
137	4	4	4	4	4	4	4	4	4	4
138	4	4	4	4	4	4	4	4	4	4
139	4	4	4	4	4	4	4	4	4	4
140	4	4	4	4	4	4	4	4	4	4
141	4	3	4	4	4	4	4	4	4	4
142	4	4	4	4	4	4	4	4	4	4
143	4	5	5	4	5	4	4	4	4	4
144	4	4	4	4	4	4	4	4	4	4
145	4	5	4	4	4	4	4	4	4	4
146	4	4	4	4	4	4	4	4	4	4
147	4	1	1	4	4	4	4	4	4	4
148	4	4	4	4	4	4	4	4	4	4
149	4	4	4	4	4	4	4	4	4	4
150	4	4	4	4	4	4	4	4	4	4
151	4	4	4	4	4	4	4	4	4	4
152	4	4	4	4	4	4	4	4	4	4
153	4	4	4	4	4	4	4	4	4	4
154	4	4	5	5	5	5	5	5	5	5
155	4	4	4	4	4	4	4	4	4	4
156	4	4	4	4	4	4	4	4	4	4
157	4	4	4	4	4	4	4	4	4	4
158	4	4	4	4	4	4	4	4	4	4
159	4	4	4	4	4	4	4	4	4	4
160	4	4	4	4	4	4	4	4	4	4
161	4	5	5	5	5	5	5	5	5	5

162	4	4	4	4	4	4	4	4	4	4
163	4	4	4	4	4	4	4	4	4	4
164	5	5	5	5	5	5	5	5	5	5
165	5	5	5	5	5	5	5	5	5	5
166	5	5	5	5	5	5	5	5	5	5
167	5	1	5	5	5	5	5	5	5	5
168	5	5	5	5	5	5	5	5	5	5
169	5	5	5	5	5	5	5	5	5	5
170	5	5	5	5	5	5	5	5	5	5
171	5	5	5	5	5	5	5	5	5	5
172	5	5	5	5	5	5	5	5	5	5
173	5	5	5	5	5	5	5	5	5	5
174	5	5	5	5	5	5	5	5	5	5
175	5	5	5	5	5	5	5	5	5	5
176	5	5	5	5	5	5	5	5	5	5
177	5	5	5	5	5	5	5	5	5	5
178	5	5	5	5	5	5	5	5	5	5
179	5	4	4	4	4	4	4	4	4	4
180	5	5	5	5	5	5	5	5	5	5
181	5	5	5	5	5	5	5	5	5	5
182	5	2	4	4	4	4	4	4	4	4
183	5	1	5	5	5	5	5	5	5	5
184	5	5	5	5	5	5	5	5	5	5
185	5	5	5	5	5	5	5	5	5	5
186	5	5	5	5	5	5	5	5	5	5
187	5	5	5	5	5	5	5	5	5	5
188	5	5	5	5	5	5	5	5	5	5
189	5	5	5	5	5	5	5	5	5	5
190	5	1	1	1	1	1	1	1	1	1
191	5	5	5	5	5	5	5	5	5	5
192	5	5	5	5	5	5	5	5	5	5
193	5	5	5	5	5	5	5	5	5	5
194	5	2	2	2	2	2	2	2	2	2
195	5	5	5	5	5	5	5	5	5	5
196	5	1	1	1	1	1	1	1	1	1
197	5	4	4	5	5	5	5	5	5	5
198	5	4	4	4	4	4	4	4	4	4
199	5	1	5	5	5	5	5	5	5	5
200	5	5	5	1	1	5	5	5	5	5

Hasil Prediksi dengan k = 4

No.	**Aktual**	**Jumlah Fitur**								
		10%	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	5	5	5	5	5	5	5	5	5

5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	5	5	5	5	5	5	5	5	5
9	1	1	1	1	1	1	1	1	1	1
10	1	3	3	3	3	3	3	3	3	3
11	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1
20	1	4	4	4	4	4	4	4	1	1
21	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	1	1	1	1	1	1	1	1	1	1
26	1	4	5	4	4	4	4	4	4	4
27	1	1	1	4	4	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1
29	1	2	2	2	2	2	2	2	2	2
30	1	1	1	1	1	1	1	1	1	1
31	1	1	1	1	1	1	1	1	1	1
32	1	1	1	1	1	1	1	1	1	1
33	1	1	1	1	1	1	1	1	1	1
34	1	5	5	5	5	5	5	5	5	5
35	1	4	1	5	5	5	5	5	5	5
36	1	1	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1	1	1
38	1	1	1	1	1	1	1	1	1	1
39	1	4	4	4	4	4	4	4	4	4
40	1	5	5	5	5	5	5	5	5	5
41	1	1	1	1	1	1	1	1	1	1
42	1	1	1	1	1	1	1	1	1	1
43	2	2	2	2	2	2	2	2	2	2
44	2	2	2	2	2	2	2	2	2	2
45	2	5	2	2	2	2	2	2	2	2
46	2	2	2	2	2	2	2	2	2	2
47	2	2	2	2	2	2	2	2	2	2
48	2	2	2	2	2	2	2	2	2	2
49	2	2	2	2	2	2	2	2	2	2
50	2	2	2	2	2	2	2	2	2	2
51	2	3	3	3	3	3	3	3	3	3
52	2	2	2	2	2	2	2	2	2	2

53	2	2	2	2	2	2	2	2	2	2
54	2	2	2	2	2	2	2	2	2	2
55	2	2	2	2	2	2	2	2	2	2
56	2	2	2	2	2	2	2	2	2	2
57	2	2	2	2	2	2	2	2	2	2
58	2	2	2	2	2	2	2	2	2	2
59	2	1	1	1	1	1	1	1	1	1
60	2	5	5	5	5	5	5	5	5	5
61	2	2	2	2	2	2	2	2	2	2
62	2	2	2	2	2	2	2	2	2	2
63	2	5	1	1	1	1	1	2	2	2
64	2	2	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2	2	2
66	2	1	1	1	1	2	2	2	2	2
67	2	2	2	2	2	2	2	2	2	2
68	2	2	2	2	2	2	2	2	2	2
69	2	2	2	2	2	2	2	2	2	2
70	2	2	2	2	2	2	2	2	2	2
71	2	2	2	2	2	2	2	2	2	2
72	2	2	2	2	2	2	2	2	2	2
73	2	2	2	2	2	2	2	2	2	2
74	2	1	4	1	1	1	1	1	1	1
75	2	2	2	2	2	2	2	2	2	2
76	2	2	2	2	2	2	2	2	2	2
77	2	1	4	4	4	4	4	4	4	1
78	2	2	2	2	2	2	2	2	2	2
79	3	1	1	1	3	3	3	3	3	3
80	3	3	3	3	3	3	3	3	3	3
81	3	3	3	3	3	3	3	3	3	3
82	3	3	3	3	3	3	3	3	3	3
83	3	3	3	3	3	3	3	3	3	3
84	3	3	3	3	3	3	3	3	3	3
85	3	3	3	3	3	3	3	3	3	3
86	3	3	3	3	3	3	3	3	3	3
87	3	5	5	3	3	3	3	3	3	3
88	3	3	3	3	3	3	3	3	3	3
89	3	3	3	3	3	3	3	3	3	3
90	3	2	2	2	2	2	2	2	2	2
91	3	3	3	3	3	3	3	3	3	3
92	3	2	3	3	3	3	3	3	3	3
93	3	3	3	3	3	3	3	3	3	3
94	3	3	3	3	3	3	3	3	3	3
95	3	5	5	3	3	3	3	3	3	3
96	3	3	3	3	3	3	3	3	3	3
97	3	3	3	3	3	3	3	3	3	3
98	3	3	3	3	3	3	3	3	3	3
99	3	3	3	3	3	3	3	3	3	3
100	3	3	3	3	3	3	3	3	3	3

101	3	3	3	3	3	3	3	3	3	3
102	3	3	3	3	3	3	3	3	3	3
103	3	1	1	1	1	1	1	1	1	1
104	3	3	3	3	3	5	5	5	5	5
105	3	3	3	3	3	3	3	3	3	3
106	3	3	3	3	3	3	3	3	3	3
107	3	3	3	3	3	3	3	3	3	3
108	3	1	1	1	1	1	1	1	1	1
109	3	3	3	3	3	3	3	3	3	3
110	3	3	3	3	3	3	3	3	3	3
111	3	3	3	3	3	3	3	3	3	3
112	3	3	3	3	3	3	3	3	3	3
113	3	3	3	3	3	3	3	3	3	3
114	3	1	3	3	3	3	3	3	3	3
115	3	3	3	3	3	3	3	3	3	3
116	3	3	3	3	3	3	3	3	3	3
117	3	3	3	3	3	3	3	3	3	3
118	3	3	3	3	3	3	3	3	3	3
119	3	4	4	1	1	3	3	3	1	1
120	4	4	4	4	4	4	4	4	4	4
121	4	4	4	4	4	4	4	4	4	4
122	4	4	4	4	4	4	4	4	4	4
123	4	4	4	4	4	4	4	4	4	4
124	4	4	4	4	4	4	4	4	4	4
125	4	5	5	4	4	4	4	4	4	4
126	4	4	4	4	4	4	4	4	4	4
127	4	4	4	4	4	4	4	4	4	4
128	4	1	4	4	4	4	4	4	4	4
129	4	4	4	4	4	4	4	4	4	4
130	4	4	4	5	5	5	5	5	5	5
131	4	4	4	4	4	4	4	4	4	4
132	4	4	4	4	4	4	4	4	4	4
133	4	4	4	4	4	4	4	4	4	4
134	4	4	4	4	4	4	4	4	4	4
135	4	5	5	5	5	5	5	5	5	5
136	4	4	4	4	4	4	4	4	4	4
137	4	4	4	4	4	4	4	4	4	4
138	4	4	4	4	4	4	4	4	4	4
139	4	4	4	4	4	4	4	4	4	4
140	4	3	1	4	4	4	3	3	1	1
141	4	4	5	5	4	4	5	4	5	5
142	4	1	4	4	4	4	4	4	4	4
143	4	1	1	1	1	1	1	1	1	1
144	4	4	4	4	4	4	4	4	4	4
145	4	4	4	4	4	4	4	4	4	4
146	4	1	5	5	5	5	5	5	5	5
147	4	4	4	4	4	4	4	4	4	4
148	4	4	4	4	4	4	4	4	4	4

149	4	4	4	4	4	4	4	4	4	4
150	4	4	4	4	4	4	4	4	4	4
151	4	4	4	4	4	4	4	4	4	4
152	4	4	4	4	4	4	4	4	4	4
153	4	1	5	4	4	1	4	1	1	2
154	4	5	5	5	5	5	5	5	5	5
155	4	4	4	4	4	4	4	4	4	4
156	4	4	4	4	4	4	4	4	4	4
157	4	4	4	4	4	4	4	4	4	4
158	4	2	2	2	2	2	2	2	2	2
159	4	4	4	4	4	4	4	4	4	4
160	4	4	4	4	4	4	4	4	4	4
161	4	4	4	4	4	4	4	4	4	4
162	4	4	4	4	4	4	4	4	4	4
163	5	5	5	5	5	5	5	5	5	5
164	5	5	5	5	5	5	5	5	5	5
165	5	5	5	5	5	5	5	5	5	5
166	5	4	5	5	5	5	5	5	5	5
167	5	1	1	1	1	1	1	1	1	1
168	5	2	2	2	2	2	2	2	2	2
169	5	5	5	5	5	5	5	5	5	5
170	5	4	4	4	4	4	4	4	4	4
171	5	5	5	5	5	5	5	5	5	5
172	5	5	5	5	5	5	5	5	5	5
173	5	5	5	5	5	5	5	5	5	5
174	5	5	1	1	1	1	1	1	1	1
175	5	5	5	5	5	5	5	5	5	5
176	5	1	1	1	1	1	1	1	1	1
177	5	5	5	5	5	5	5	5	5	5
178	5	5	5	5	5	5	5	5	5	5
179	5	5	5	5	5	5	5	5	5	5
180	5	5	5	5	5	5	5	5	5	5
181	5	1	1	1	1	1	1	1	1	1
182	5	3	5	3	3	3	3	3	3	3
183	5	1	5	1	1	1	1	5	5	5
184	5	1	1	1	1	1	1	1	1	1
185	5	1	1	1	1	1	1	1	1	1
186	5	5	4	5	4	5	5	4	4	5
187	5	5	5	5	5	5	5	5	5	5
188	5	5	4	4	4	5	5	5	5	5
189	5	2	2	2	2	2	2	2	2	2
190	5	5	5	5	5	5	5	5	5	5
191	5	5	5	5	5	5	5	5	5	5
192	5	5	5	5	5	5	5	5	5	5
193	5	5	5	5	5	5	5	5	5	5
194	5	5	5	5	5	5	5	5	5	5
195	5	4	4	4	4	4	4	4	4	4
196	5	3	3	3	3	3	3	3	3	3

197	5	5	5	5	5	5	5	5	5	5
198	5	2	5	5	5	5	5	5	5	5
199	5	5	5	5	5	5	5	5	5	5
200	5	1	5	5	5	5	5	5	5	5

Hasil Prediksi dengan k = 5

No.	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	4	4	4	4	4	4	4	4	4
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	2	5	5	5	5	5	5	5	5
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	5	1	1	5	1	1
7	1	5	5	5	5	5	5	5	5	5
8	1	1	1	1	3	1	1	1	1	1
9	1	4	1	1	4	1	1	1	1	4
10	1	1	1	1	1	1	1	1	1	1
11	1	5	2	2	2	2	2	2	2	2
12	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1
14	1	5	5	5	5	5	5	5	5	5
15	1	1	3	3	3	3	3	3	3	3
16	1	3	3	3	3	3	3	3	3	3
17	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1
21	1	4	4	4	4	4	4	4	4	4
22	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	5	1	1	1	1	1	1	1	1
25	1	1	1	1	1	1	1	1	1	1
26	1	1	1	1	1	1	1	1	1	1
27	1	2	1	1	1	1	1	1	1	1
28	1	1	1	1	1	1	1	1	1	1
29	1	1	1	1	1	1	1	1	1	1
30	1	1	1	1	1	1	1	1	1	1
31	1	5	5	5	5	2	2	5	5	5
32	1	1	1	1	1	1	1	1	1	1
33	1	1	1	1	1	1	1	1	1	1
34	1	1	1	1	1	1	1	1	1	1
35	1	1	1	1	1	1	1	1	1	1
36	1	1	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1	1	1
38	2	1	1	1	1	1	1	1	1	1
39	2	2	2	2	2	2	2	2	2	2

40	2	2	2	2	2	2	2	2	2	2
41	2	2	2	2	2	2	2	2	2	2
42	2	2	2	2	2	2	2	2	2	2
43	2	2	2	2	2	2	2	2	2	2
44	2	2	2	2	2	2	2	2	2	2
45	2	2	2	2	2	2	2	2	2	2
46	2	2	2	2	2	2	2	2	2	2
47	2	2	2	2	2	2	2	2	2	2
48	2	2	2	2	2	2	2	2	2	2
49	2	2	1	2	2	2	2	2	2	2
50	2	2	2	2	2	2	2	2	2	2
51	2	2	2	2	2	2	2	2	2	2
52	2	2	2	2	2	2	2	2	2	2
53	2	2	2	2	2	2	2	2	2	2
54	2	2	2	2	2	2	2	2	2	2
55	2	2	2	2	2	2	2	2	2	2
56	2	2	2	2	2	2	2	2	2	2
57	2	2	2	2	2	2	2	2	2	2
58	2	2	2	2	2	2	2	2	2	2
59	2	2	2	2	2	2	2	2	2	2
60	2	2	2	2	2	2	2	2	2	2
61	2	2	2	2	2	2	2	2	2	2
62	2	2	2	2	2	2	2	2	2	2
63	2	2	2	2	2	2	2	2	2	2
64	2	2	2	2	2	2	2	2	2	2
65	2	2	2	2	2	2	2	2	2	2
66	2	2	2	2	2	2	2	2	2	2
67	2	2	2	2	2	2	2	2	2	2
68	2	2	2	2	2	2	2	2	2	2
69	2	3	1	1	1	1	1	2	1	1
70	2	5	5	5	5	5	5	5	5	5
71	2	2	2	2	2	2	2	2	2	2
72	2	2	2	2	2	2	2	2	2	2
73	2	5	5	4	4	4	4	4	4	4
74	2	2	2	2	2	2	2	2	2	2
75	2	2	2	2	2	2	2	2	2	2
76	2	2	2	2	2	2	2	2	2	2
77	2	2	2	2	2	2	1	1	2	2
78	2	2	2	2	2	2	2	2	2	2
79	2	2	2	2	2	2	2	2	2	2
80	2	2	2	2	2	2	2	2	2	2
81	2	2	2	2	2	2	2	2	2	2
82	2	2	2	2	2	2	2	2	2	2
83	2	2	2	2	2	2	2	2	2	2
84	2	2	2	2	2	2	2	2	2	2
85	3	3	3	3	3	3	3	3	3	3
86	3	5	1	1	1	1	1	1	1	1
87	3	3	3	3	3	3	3	3	3	3

88	3	3	3	3	3	3	3	3	3	3
89	3	3	3	3	3	3	3	3	3	3
90	3	3	2	2	2	2	2	2	2	3
91	3	3	3	3	3	3	3	3	3	3
92	3	3	3	3	3	3	3	3	3	3
93	3	3	3	3	3	3	3	3	3	3
94	3	3	3	3	3	3	3	3	3	3
95	3	5	3	5	5	5	5	5	5	5
96	3	3	3	3	3	3	3	3	3	3
97	3	3	3	3	3	3	3	3	3	3
98	3	3	3	3	3	3	3	3	3	3
99	3	3	3	3	3	3	3	3	3	3
100	3	3	3	3	3	3	3	3	3	3
101	3	3	3	3	3	3	3	3	3	3
102	3	3	3	3	3	3	3	3	3	3
103	3	3	3	3	3	3	3	3	3	3
104	3	3	3	3	3	3	3	3	3	3
105	3	3	2	2	2	2	2	2	2	2
106	3	3	3	3	3	3	3	3	3	3
107	3	3	3	3	3	3	3	3	3	3
108	3	3	3	3	3	3	3	3	3	3
109	3	3	3	3	3	3	3	3	3	3
110	3	3	3	3	3	3	3	3	3	3
111	3	3	3	3	3	3	3	3	3	3
112	3	3	3	3	3	3	3	3	3	3
113	3	3	3	3	3	3	3	3	3	3
114	3	3	3	3	3	3	3	3	3	3
115	3	3	3	3	3	3	3	3	3	3
116	3	3	3	3	3	3	3	3	3	3
117	3	3	3	3	3	3	3	3	3	3
118	3	3	3	3	3	3	3	3	3	3
119	3	3	3	3	3	3	3	3	3	3
120	3	3	3	3	3	3	3	3	3	3
121	3	3	3	3	3	3	3	3	3	3
122	3	4	4	4	4	4	4	4	4	4
123	3	3	4	3	4	3	3	3	3	3
124	3	3	3	3	3	3	3	3	3	3
125	3	3	3	3	3	3	3	3	3	3
126	4	4	4	4	4	4	4	4	4	4
127	4	1	1	1	5	1	1	1	3	1
128	4	4	4	4	4	4	4	4	4	4
129	4	4	4	4	4	4	4	4	4	4
130	4	4	4	4	4	4	4	4	4	4
131	4	4	4	4	4	4	4	4	4	4
132	4	4	4	4	4	4	4	4	4	4
133	4	4	4	4	4	4	4	4	4	4
134	4	4	4	4	4	4	4	4	4	4
135	4	4	4	4	4	4	4	4	4	4

136	4	4	4	4	4	4	4	4	4	4
137	4	4	4	4	4	4	4	4	4	4
138	4	5	5	5	5	5	5	5	5	5
139	4	4	4	4	4	4	4	4	4	4
140	4	4	4	4	4	4	4	4	4	4
141	4	4	4	4	4	4	4	4	4	4
142	4	5	5	4	4	4	4	4	4	4
143	4	3	4	4	4	4	4	4	4	4
144	4	5	5	4	4	4	4	4	4	4
145	4	4	4	4	4	4	4	4	4	4
146	4	4	4	4	4	4	4	4	4	4
147	4	5	5	5	5	5	5	5	5	5
148	4	3	3	3	3	3	3	3	3	3
149	4	4	4	4	4	4	4	4	4	4
150	4	4	4	4	4	4	4	4	4	4
151	4	5	4	4	4	1	4	4	4	4
152	4	4	4	4	4	4	4	4	4	4
153	4	4	4	4	4	4	4	4	4	4
154	4	4	4	4	4	4	4	4	4	4
155	4	4	4	4	4	4	4	4	4	4
156	4	4	4	4	4	4	4	4	4	4
157	4	4	4	4	4	4	4	4	4	4
158	4	4	4	4	4	4	4	4	4	4
159	4	3	3	3	3	3	3	3	3	3
160	4	4	4	4	4	4	4	4	4	4
161	4	1	3	3	4	4	4	4	4	4
162	4	1	1	1	5	5	5	5	5	5
163	4	4	4	4	4	4	4	4	4	4
164	4	4	4	4	4	4	4	4	4	4
165	5	1	1	1	4	1	1	1	1	1
166	5	5	5	5	5	5	5	5	5	5
167	5	4	4	5	5	5	5	5	5	5
168	5	5	5	5	5	5	5	5	5	5
169	5	5	5	5	5	5	5	5	5	5
170	5	5	5	5	5	5	5	5	5	5
171	5	5	3	3	3	3	3	3	3	3
172	5	1	5	5	5	5	5	5	5	5
173	5	1	1	1	1	1	1	1	1	1
174	5	5	5	5	5	5	5	5	5	5
175	5	2	2	2	2	2	2	3	3	3
176	5	1	1	1	1	1	1	1	1	1
177	5	5	5	5	5	5	5	5	5	5
178	5	2	5	5	5	5	5	5	5	5
179	5	5	5	5	5	5	5	5	5	5
180	5	5	1	1	1	1	1	1	1	1
181	5	5	5	5	5	5	5	5	5	5
182	5	5	5	5	5	5	5	5	5	5
183	5	5	5	5	5	5	5	5	5	5

184	5	4	4	4	4	4	4	4	4	4
185	5	5	5	5	5	5	5	5	5	5
186	5	5	5	5	5	5	5	5	5	5
187	5	5	5	5	5	1	5	5	5	5
188	5	5	5	5	5	5	5	5	5	5
189	5	5	5	5	5	5	5	5	5	5
190	5	1	1	1	1	1	1	1	1	1
191	5	1	1	5	5	5	5	5	5	5
192	5	5	5	5	5	5	5	5	5	5
193	5	1	5	5	5	5	5	5	5	5
194	5	4	4	4	4	4	4	4	4	4
195	5	5	5	5	5	5	5	5	5	5
196	5	5	5	5	5	5	5	5	5	5
197	5	5	5	5	5	5	5	5	5	5
198	5	5	5	5	5	5	5	5	5	5
199	5	5	5	5	5	5	5	5	5	5
200	5	1	1	1	1	1	1	1	1	1