

Aplikasi Klasifikasi Laporan Pelanggan Menggunakan Algoritma Naive Bayes (Studi Kasus: Direktorat Sistem Informasi, Telkom University) Customer's Report Classification Application Using Naive Bayes Algorithm (Case Study: Information System Directorate, Telkom University)

Ade Handiyanto¹, Dahliar Ananda, S.T., M.T.², Elis Hernawati, S.T., M.Kom.³

^{1,2,3}Program Studi D3 Sistem Informasi, Fakultas Ilmu Terapan Universitas Telkom

¹Adehandynt98@gmail.com, ²Ananda@tass.telkomuniversity.ac.id, ³Elishernawati@tass.telkomuniversity.ac.id

Abstrak— Aplikasi klasifikasi laporan pelanggan ini merupakan sebuah aplikasi yang digunakan untuk membantu proses klasifikasi laporan keluhan pelanggan atau pengguna layanan sistem informasi Universitas Telkom yang disampaikan melalui media sosial Facebook. Klasifikasi data laporan bertujuan untuk memilah data laporan yang sesuai dengan layanan Ditsisfo dan menyisihkan data laporan yang tidak termasuk kedalam layanan Ditsisfo baik itu layanan Fakultas maupun Unit lain. Aplikasi ini ditunjang dengan algoritma Naive Bayes yang dapat mengklasifikasikan data laporan pelanggan dengan jumlah yang banyak secara cepat dan akurat, dengan data yang telah terklasifikasi maka diharapkan proses penanganan masalah atau komplain dapat diselesaikan dengan cepat. Saat ini proses klasifikasi data laporan di Ditsisfo masih dilakukan secara manual, hal tersebut tentu dapat menimbulkan kesalahan klasifikasi data karena pengklasifikasian data laporan masih terbatas oleh pengetahuan masing-masing stafnya.

Kata Kunci: Naive Bayes, Sistem Informasi, Facebook

Abstract— This customer report classification application is an application that is used to help the process of classification of customer complaints reports or users of Telkom University information system services delivered via Facebook social media. The classification of report data aims to sort report data in accordance with Ditsisfo services and set aside report data that is not included in Ditsisfo services, both Faculty services and other Units. This application is supported by the Naive Bayes algorithm that can classify customer report data with large amounts quickly

and accurately, with data that has been classified, it is expected that the process of handling problems or complaints can be resolved quickly. At present the process of classification of report data in Ditsisfo is still done manually, it certainly can lead to misclassification of data because the classification of report data is still limited by the knowledge of each staff member.

Keywords: Naive Bayes, Information System, Facebook

I. PENDAHULUAN

1. Latar Belakang

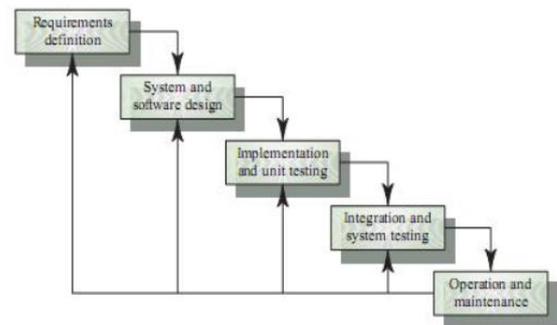
Direktorat Sistem Informasi (SISFO), merupakan Direktorat yang bertanggung jawab terhadap implementasi teknologi informasi dan komunikasi di Telkom University[1]. Ditsisfo memberikan layanan infrastruktur teknologi informasi, layanan interkoneksi, layanan data dan sistem informasi serta layanan komputasi sebagai strategic tools untuk mendukung berjalannya proses bisnis di Universitas Telkom[1]. Hal lain terkait dengan proses layanan yang berjalan di Direktorat Sistem Informasi adalah adanya Service Level Agreement (SLA)[1]. SLA tersebut bertujuan untuk menentukan batasan waktu respon dan resolusi terhadap setiap rekues yang masuk.

Untuk mendukung layanan tersebut terdapat urusan Helpdesk dan Service Level Management, yang bertanggung jawab terhadap manajemen komplain dan sosialisasi layanan sistem informasi untuk menjamin operasional sistem informasi, komunikasi dengan stakeholder dan customer pengguna layanan sistem informasi Universitas Telkom. Bentuk komunikasi dari keseluruhan pelayanan tersebut adalah dengan melalui aplikasi Helpdesk Ticketing di iGracias, email, telepon, SMS, messenger, serta grup di media sosial[1].

Salah satu media komunikasi yang digunakan pihak Ditsisfo adalah grup terbuka media sosial Facebook dengan nama "Helpdesk Direktorat SISFO Tel-U". Berdasarkan informasi yang diperoleh dari urusan Helpdesk dan Service Level Management, bahwa setiap laporan dan rekes dari media sosial dilakukan proses penyimpanan secara manual kedalam Aplikasi Ticketing Helpdesk. Laman mulai digunakan pada juni tahun 2014 dengan jumlah laporan hingga januari 2017 yaitu sebanyak 4565 laporan yang masuk. Dari jumlah laporan tersebut, tidak seluruh data laporan atau rekes sesuai dengan layanan Ditsisfo, terdapat laporan yang sebenarnya merupakan bagian dari proses bisnis fakultas atau unit lain. Proses pemilahan data tersebut masih dilakukan secara manual, dan tidak seluruh data perlu untuk disimpan kedalam Aplikasi Ticketing Helpdesk.

Dari permasalahan tersebut dapat disimpulkan bahwa perlu adanya suatu aplikasi klasifikasi data keluhan yang masuk kedalam grup "Helpdesk Direktorat SISFO Tel-U". Pengimplementasian algoritma yang tepat dapat membuat sebuah aplikasi bekerja dengan lebih akurat, maka dalam pembangunan aplikasi klasifikasi data keluhan Ditsisfo tersebut akan menerapkan algoritma Naive Bayes karena memiliki tingkat keakuratan yang sangat tinggi dibandingkan dengan algoritma pengklasifikasian lainnya. Algoritma Naive Bayes sangat cocok apabila diterapkan pada jumlah data yang sangat banyak, dalam pengklasifikasian data algoritma Naive Bayes akan ada dua proses yaitu proses training data dan testing data. Training data merupakan sekumpulan data-data valid yang sudah disiapkan dan digunakan sebagai acuan pada saat proses klasifikasi, sedangkan testing data adalah data yang akan diklasifikasikan dan akan dibandingkan dengan data training. Dengan adanya aplikasi tersebut diharapkan dapat membantu staf Ditsisfo dalam hal pemilhan laporan yang sesuai dengan layanan Ditsisfo.

2. Metode Pengerjaan



Gambar 1 Metode Waterfall

Metode pengerjaan yang digunakan dalam pembuatan sistem ini adalah metode waterfall. Model waterfall adalah model klasik yang bersifat sistematis, berurutan dalam membangun perangkat lunak.[2].

II. TINJAUAN PUSTAKA

1. Direktorat Sistem Informasi

Direktorat Sistem Informasi (SISFO), merupakan Direktorat yang bertanggung jawab terhadap implementasi teknologi informasi dan komunikasi di Telkom University [1]. Ditsisfo memberikan layanan infrastruktur teknologi informasi, layanan interkoneksi, layanan data dan sistem informasi serta layanan komputasi sebagai strategic tools untuk mendukung berjalannya proses bisnis di Universitas Telkom [1]. Hal lain terkait dengan proses layanan yang berjalan di Direktorat Sistem Informasi adalah adanya Service Level Agreement (SLA) [1]. SLA tersebut bertujuan untuk menentukan batasan waktu respon dan resolusi terhadap setiap rekes yang masuk.

2. Media Sosial

Media sosial dapat digunakan menjadi sebuah sarana penting yang dapat menciptakan modal sosial dalam hal organisasi dan sebagai kunci utama dalam membangun bisnis yang efektif dan efisien untuk menciptakan keuntungan kompetitif. Namun, hal tersebut dapat menjadi pisau bermata dua, dimana proses adopsinya harus menggunakan pengetahuan yang cukup mengenai keuntungan dan risiko yang akan ditimbulkan [4].

3. Klasifikasi Data

Pengklasifikasian merupakan salah satu masalah dasar dalam literature mengenai machine learning dan data

mining[5]. Text categorization, text classification, atau topic spotting, merupakan aktivitas pelabelan terhadap teks bahasa alami dengan kategori tematik dari sebuah set yang telah ditentukan [6].

4. Naive Bayes

Naive Bayes Classifier merupakan salah satu metode klasifikasi data yang berdasarkan pada teorema Bayes. Naive Bayesian Classifier mengasumsikan bahwa setiap atribut di dalam sebuah kelas merupakan atribut independen yang tidak terkait satu sama lain. Asumsi ini disebut dengan class conditional independence. Naive Bayesian Classifier memiliki tingkat kecepatan dan akurasi yang tinggi ketika diaplikasikan pada data dengan jumlah yang banyak [7].

Perhitungan perbandingan antara kata pada data testing dihitung dengan menggunakan persamaan 1 [16]:

$$P(a|v_j) = \frac{nc + mp}{n + m} \quad (1)$$

dimana:

n = jumlah kata pada data latih dimana $v = v_j$

nc = jumlah kata dimana $v = v_j$ dan $a = a_i$

p = probabilitas setiap kelas dalam data latih

m = jumlah kata pada data uji

Sedangkan untuk menentukan klasifikasi data pada data uji menggunakan persamaan 2 [1]:

$$V_{nb} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i|v_j) \quad (2)$$

5. Confusion Matrix

Confusion matrix adalah sebuah metode untuk menyatakan jumlah data uji yang benar diklasifikasikan atau data uji yang salah diklasifikasikan berdasarkan pengujian manual.

True Positive (TP), yaitu jumlah suatu data dari kelas 1 yang benar dan diklasifikasikan oleh sistem sebagai kelas 1.

True Negative (TN), yaitu jumlah suatu data dari kelas 1 yang benar dan diklasifikasikan oleh sistem sebagai kelas 0.

False Positive (FP), yaitu jumlah suatu data dari kelas 0 yang salah dan diklasifikasikan oleh sistem sebagai kelas 1.

False Negative (FN), yaitu jumlah suatu data dari kelas 0 yang salah dan diklasifikasikan oleh sistem sebagai kelas 0.

Perhitungan tingkat akurasi dinyatakan menggunakan persamaan 3 [8]:

$$\text{Akurasi} = \left(\frac{TP + TN}{TP + FN + FP + TN} \right) * 100 \% \quad (3)$$

6. BPMN (Business Process Model and Notation)

BPMN (Business Processing Model and Notation) merupakan sebuah standar representasi berupa gambar untuk menentukan dan menggambarkan suatu proses bisnis saat ini atau usulan yang dikeluarkan oleh open management group [15]. Tujuan BPMN adalah menyediakan suatu notasi standar yang mudah dipahami oleh semua pemangku kepentingan bisnis.

7. Use Case Diagram

Use case diagram merupakan sebuah pemodelan behavior atau tingkah laku sistem yang akan dibuat. Use case mendeskripsikan sebuah interaksi antara satu atau lebih aktor dengan sistem yang akan dibuat. Use case digunakan untuk mengetahui fungsionalitas yang ada di dalam sebuah sistem atau aplikasi dan siapa saja yang memiliki hak akses untuk menggunakan fungsional tersebut [13].

8. MySQL

MySQL adalah salah satu database server yang sering digunakan dalam pembangunan sebuah website atau aplikasi berbasis web. MySQL merupakan server basis data yang menggunakan teknik relasional untuk menghubungkan antara tabel-tabel dalam basis datanya atau biasa disebut dengan RDBMS (Relational Database Management System)[10].

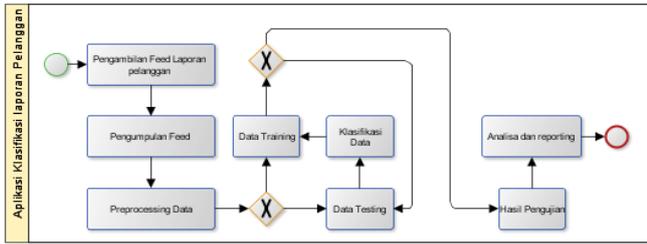
9. Blackbox Testing

Blackbox testing adalah suatu metode pengujian perangkat lunak untuk mendapatkan serangkaian kondisi input yang memenuhi persyaratan fungsional suatu program. Metode pengujian ini berusaha untuk menemukan kesalahan dengan permasalahan seperti fungsi-fungsi yang hilang atau salah. Pengujian blackbox testing hanya fokus kepada fungsionalitas aplikasi yang telah selesai dibangun tanpa menguji source code [14].

III. HASIL

A. Hasil Perancangan

Gambar berikut adalah BPMN dari perancangan proses berjalannya klasifikasi data menggunakan algoritma Naive Bayes.



Gambar 2 Proses Bisnis Usulan Klasifikasi Data

Proses klasifikasi data memiliki beberapa tahap yaitu:

1. Pengambilan Feed Laporan Pelanggan

Klasifikasi data diawali dengan proses pengambilan data feed laporan pelanggan yang berasal dari feed grup Facebook Helpdesk SISFO atau file Excel berisikan laporan pelanggan yang telah disimpan sebelumnya.

2. Pengumpulan Feed

Feed yang telah diambil akan disimpan kedalam array untuk ditampung secara sementara sebelum melalui proses selanjutnya.

3. Preprocessing Data

Data yang telah ditampung kedalam array akan dipanggil kembali dan melalui proses Preprocessing data. Preprocessing data adalah sebuah proses penyederhanaan feed yang telah diambil dengan cara menghilangkan tanda baca dan mengubah huruf kapital menjadi huruf kecil yang kemudian akan disimpan kedalam database.

4. Data Testing

Data Testing yaitu sebuah proses pengujian feed yang telah tersimpan didalam database untuk menghitung kemungkinan setiap data feed dan kata. Sebelum melakukan testing, user diharuskan telah memiliki data latih atau data awal yang tersimpan didalam database sebagai data acuan dalam proses penghitungan kemungkinannya yang kemudian akan disimpan kedalam array.

5. Klasifikasi Data

Dalam proses ini dilakukan sebuah perhitungan lebih lanjut terhadap hasil perhitungan pada proses sebelumnya untuk menentukan klasifikasi pada data uji.

6. Data Training

Setelah diklasifikasi feed yang telah diuji akan melalui proses pelatihan data dimana feed akan dipecah menjadi kata per kata dan disimpan kedalam database, apabila kata yang akan disimpan telah tersedia didalam database maka kata tidak disimpan kembali melainkan akan dihitung sesuai dengan klasifikasi feed tersebut.

7. Hasil Pengujian

Data yang telah selesai melalui proses pengujian kemudian akan dihitung hasil dari pengujian tersebut.

8. Analisa dan Reporting

Hasil perhitungan akan disajikan kedalam laporan dengan bentuk grafik persentase feed dan kata yang telah diuji.

IV. PEMBAHASAN

A. Implementasi Aplikasi

Berikut adalah hasil implementasi dari perancangan aplikasi.

1. Retrieve Data

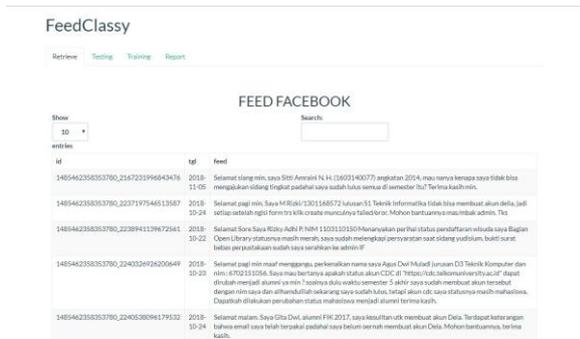
Berikut adalah gambar dari tampilan implementasi Retrieve Data.



Gambar 3 Retrieve Data

2. Retrieve Data Facebook

Berikut adalah gambar dari tampilan implementasi Retrieve Data Facebook.



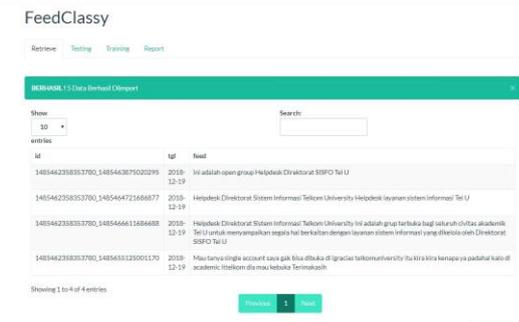
Gambar 4 Retrieve Data Facebook



Gambar 7 Training Kata

3. Retrieve Data Excel

Berikut adalah gambar dari tampilan implementasi Retrieve Data Excel.



Gambar 5 Retrieve Data Excel

4. Training Data

Berikut adalah gambar dari tampilan implementasi Training Data.



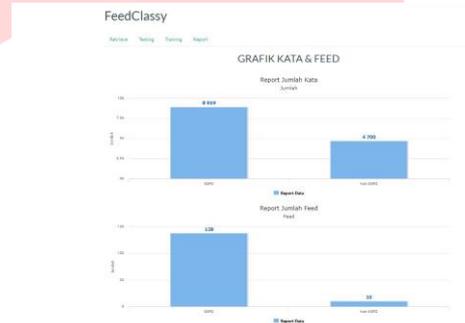
Gambar 6 Training Data

5. Training Kata

Berikut adalah gambar dari tampilan implementasi Training Kata.

6. Reporting

Berikut adalah gambar dari tampilan implementasi Reporting.



Gambar 8 Reporting

B. Implementasi Data

Data yang digunakan dalam aplikasi ini merupakan data feed yang diambil dari grup terbuka Helpdesk Sisfo pada media sosial Facebook atau data feed yang telah disimpan kedalam file excel.

1. Retrieve Data Feed Facebook

Data feed yang diambil secara realtime pada media sosial Facebook didapatkan dengan menggunakan akses API yang telah disediakan oleh Facebook. Pengaksesan API harus memiliki izin pemilik akun agar dapat mengakses informasi yang dibutuhkan, sebagai contoh pada aplikasi ini pembuatan token harus memiliki izin admin grup Helpdesk SISFO agar dapat mengakses atau mengambil setiap laporan yang masuk kedalam grup.

2. Retrieve Data Feed Excel

Data yang digunakan pada proses ini yaitu data feed yang berasal dari file excel, data pada file tersebut merupakan data feed yang telah diambil sebelumnya pada grup Helpdesk SISFO dengan menggunakan API Facebook dan disimpan kedalam file dengan format .xlsx. contoh data yang berasal dari file excel terdapat

pada gambar berikut, dimana data terdiri dari ID, Manual classification, Category, dan Feed.

$$V_{nb} = \operatorname{argmax}_{v_j \in V} P(v_j) \prod P(a_i | v_j)$$

Gambar 9 Retrieve Data Excel

3. Testing Data

Proses testing yaitu suatu proses pengujian data uji terhadap data latih yang ada didalam database untuk menghitung kemungkinan data feed yang diuji apakah termasuk kedalam layanan SISFO atau non layanan SISFO. Data feed dihitung dengan cara memecah feed menjadi kata per kata dan setiap kata tersebut akan dihitung probabilitasnya dengan menggunakan persamaan 1 :

$$P(a|v_j) = \frac{nc + mp}{n + m}$$

dimana:

n = jumlah kata pada data latih dimana v = v_j

nc = jumlah kata dimana v = v_j dan a = a_i

p = probabilitas setiap kelas dalam data latih

m = jumlah kata pada data uji

contoh hasil dari persamaan diatas adalah sebagai berikut :

$$P(\text{SISFO} | \text{ini}) = (684 + 5 * 0.76117692013756) / (117372 + 5) = 0,0058$$

$$P(\text{SISFO} | \text{sisfo}) = (530 + 5 * 0.76117692013756) / (117372 + 5) = 0,0045$$

Setelah mendapat probabilitas setiap kata maka hal yang dilakukan selanjutnya yaitu proses klasifikasi untuk menentukan apakah feed tersebut masuk kedalam layanan SISFO atau layanan non SISFO, perhitungan ini dilakukan dengan menggunakan persamaan 2:

Sebagai contoh dari nilai kata yang telah diperoleh pada proses sebelumnya dengan menggunakan persamaan diatas maka diperoleh perhitungan sebagai berikut:

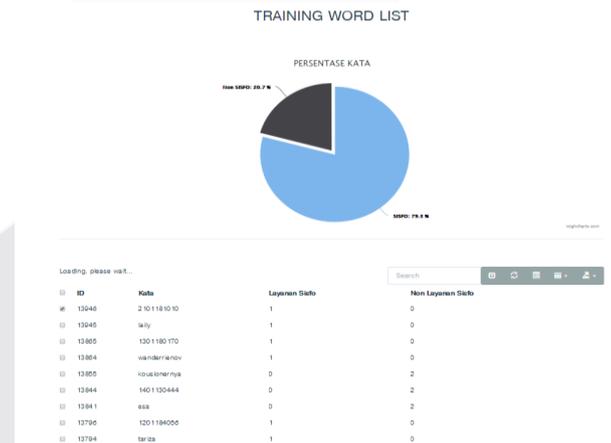
$$V(\text{SISFO}) = 0.7611 * 0,0058 * 7,5022 * 7,5022 * 0,0011 * 0,0045 = 1,229$$

$$V(\text{Non SISFO}) = 0.2386 * 0,0066 * 3,5708 * 1,2280 * 0,0039 * 0,0035 = 9,425$$

Data feed diklasifikasikan berdasarkan kelas dengan hasil nilai paling besar pada persamaan diatas. Data feed yang telah diklasifikasikan akan ditampilkan hasilnya dan disimpan kedalam database sebagai data latih baru pada proses testing selanjutnya.

4. Training Data

Setelah diklasifikasi feed yang telah diuji akan melalui proses pelatihan data dimana feed akan dipecah menjadi kata per kata dan disimpan kedalam database, apabila kata yang akan disimpan telah tersedia didalam database maka kata tidak disimpan kembali melainkan akan dihitung kemunculannya sesuai dengan klasifikasi feed tersebut dan kata-kata hasil pelatihan akan ditampilkan seperti pada gambar berikut:

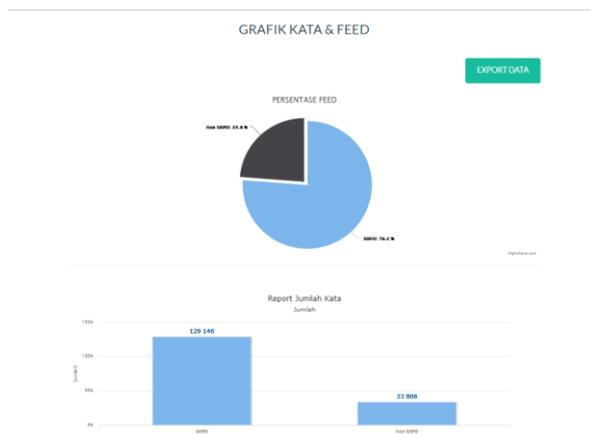


Gambar 10 Hasil Training Data

5. Analisa dan Reporting

Hasil perhitungan pada proses sebelumnya disajikan kedalam laporan dalam bentuk grafik persentase feed

dan kata yang telah diuji serta daftar feed yang telah selesai diuji.



Gambar 11 Hasil reporting

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan rumusan masalah yang telah dibuat sebelumnya dan dilakukan penelitian terhadap rumusan masalah tersebut, sehingga dapat disimpulkan bahwa dalam penelitian yang telah dilakukan dapat diketahui bahwa masih terdapat suatu permasalahan pada saat proses mengklasifikasikan data laporan pelanggan yang terdapat di grup terbuka media sosial Facebook. Proses klasifikasi sendiri masih terbatas oleh pengetahuan dari masing-masing staf nya, hal tersebut dapat mengakibatkan sebuah laporan yang seharusnya masuk kedalam bagian Ditsisfo akan tetapi diklasifikasikan kedalam bagian Fakultas atau Unit lain. Aplikasi klasifikasi laporan pelanggan ini telah selesai dibangun dan telah diuji coba menggunakan data latih awal dengan jumlah sebanyak 5.200 data feed sebagai data acuan untuk melakukan pengujian. Pengujian dilakukan menggunakan 501 data feed yang berasal dari file excel dan 37 data feed yang berasal dari Facebook dengan masing-masing tingkat akurasi yaitu 73.05% dan 56.75% . Tinggi atau rendahnya tingkat akurasi yang diperoleh dapat dipengaruhi oleh kompleksitas dan kejelasan suatu feed yang diuji tersebut, dengan adanya aplikasi klasifikasi laporan pelanggan ini diharapkan dapat menangani permasalahan dalam proses klasifikasi data laporan pelanggan karena didukung dengan algoritma Naive Bayes yang memiliki kecepatan dan akurasi tinggi dalam mengklasifikasikan data dengan jumlah yang banyak.

B. Saran

Berdasarkan dari hasil kesimpulan terkait penelitian yang dilakukan, penulis memberikan saran terkait dengan analisa yang dilakukan oleh penulis adalah sebagai berikut :

1. Sistem pelaporan keluhan pada grup terbuka Helpdesk SISFO perlu diperbaiki yaitu dengan menggunakan format pelaporan apabila akan menyampaikan laporan melalui grup terbuka Helsek SISFO. Dengan penggunaan format pelaporan maka proses klasifikasi baik manual atau menggunakan aplikasi akan menambah tingkat akurasi pada data yang diklasifikasi.
2. Jika aplikasi tersebut akan diimplementasikan maka format data atau file yang diperlukan pada aplikasi tersebut harus sesuai dengan kebutuhan aplikasi dan format data pada file harus bersifat tetap dan tidak berubah-ubah.
3. Penambahan algoritma Sparse Binary Polynomial Hashing (SBPH) untuk meminimalisir kesalahan filtering dan meningkatkan akurasi dengan cara membandingkan satu kata dengan kata yang lain dan membentuk suatu frasa - frasa baru. Frase baru tersebut akan membentuk sebuah pola tertentu dan klasifikasi data dilakukan dengan cara penghitungan per frase sehingga dapat meminimalisir kesalahan sistem dalam proses klasifikasi akibat keyword dan meningkatkan akurasi data.

REFERENSI

- [1] Direktorat. Universitas Telkom, Rencana Strategis Sistem Informasi Universitas Telkom Periode 2014-2018, 2014.
- [2] Sommerville. Ian, Software Engineering 9th Edition, Addison-Wesley, 2011.
- [3] Pressman. Roger S, Software Engineering: a Practitioner's Approach 7th Edition, 2010: McGraw Hill Higher Education.
- [4] Turban. Efraim, Liang. Ting-Peng and Bolloju. Narasimha , "Enterprise Social Networking: Opportunities, Adoption, and Risk Mitigation," J. Organ. Comput. Electron. Commer., vol. 21, pp. 202-220, 2011.
- [5] Aggarwal. Charu C and Zhai. ChengXiang, "A Survey of Text Clustering Algorithms," dalam in Mining Text Data, Boston, MA: Springer US, 2012, pp. 77-128.
- [6] Sebastiani. Fabrizio, "Machine Learning in Automated Text Categorization," ACM Comput. Surv., vol. 34, pp. 1-47, 2002.
- [7] Han. Jiawei, Kamber. Micheline and Pei. Jian, Data Mining: Concepts and Techniques, Second Edition., San Francisco: Morgan Kaufmann, 2006.
- [8] Anhar, Panduan Menguasai PHP dan MySQL secara Otodidak., Jakarta Selatan: Mediakita, 2010.
- [9] Basuki. Awan Pribadi, Membangun Web Berbasis PHP dengan Framework Codeigniter., Yogyakarta: Lokomedia, 2010.
- [10] Riyanto, Membuat Aplikasi Apotek integrasi Barcode Scanner dengan PHP dan MySQL, Yogyakarta: Gave Media, 2015.
- [11] Mulyani. Sri, Analisis dan Perancangan Sistem Informasi Manajemen Keuangan Daerah: Notasi Pemodelan Unified Modeling Language (UML) ., Bandung: Abdi Sistematika, 2017.

- [12] Nugroho. Eddy Prasetyo, *Rekayasa Perangkat Lunak.*, Bandung: Politeknik Telkom, 2009.
- [13] Yulianto. Ardhian Agung, Gartina. Inne, Astuti. Rini, Dewi. Sari, and Sari. Siska Komala dan Witanti. W, *Analisis Perancangan Sistem Informasi*, Bandung: Politeknik Telkom, 2015.
- [14] Hall. James, *Accounting Information System 6 Edition.*, South Western: Cengage Learning Mason, 2008.
- [15] Prasetyo. Hanung Nindito, "Pengantar Dan Pengenalan BPMN," p. 8, 2016.
- [16] Indriani. Aida, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier," 2014.