

# Implementasi Sistem Crowdsourced Labelling Berbasis Web dengan Metode *Weighted Majority Voting*

Antonius Rachmat C., Yuan Lukito

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Duta Wacana Yogyakarta  
anton@ti.ukdw.ac.id, yuanlukito@ti.ukdw.ac.id

Diterima 27 Agustus 2015

Disetujui 27 Oktober 2015

**Abstract**— Crowdsourced Labelling is a large scale data labelling process, solicits a large group of people to label the data, usually via Internet. This paper discusses about design and implementation of Web-based Crowdsourced Labelling. Supervised learning classification methods need labelled training data for its training phase. Unfortunately, in many cases, there aren't any already available labelled training data. Large scale data labelling is a tedious and time consuming work. This research develops a web-based crowdsourced labelling which able to solicit a large group of people as data labeler to speed up the data labelling process. This system also allows multiple labeler for every data. The final label is calculated using Weighted Majority Voting method. We grabbed and used Facebook comments from the two candidates' Facebook Page of 2014 Indonesian Presidential Election as testing data. Based on the testing conducted we can conclude that this system is able to handle all the labelling steps well and able to handle collision occurred when multiple labeler labelling a same data in the same time. The system successfully produces final label in CSV format, which can be processed further with many sentiment analysis tools or machine learning tools.

**Index Terms**—Crowdsourced labeling, web-based system, supervised learning, weighted majority voting

## I. PENDAHULUAN

Proses pembelajaran pada sistem klasifikasi dapat dibedakan menjadi 2 jenis: *supervised learning* dan *unsupervised learning*.

Pada metode pembelajaran terbimbing, data pembelajaran diberi suatu penanda/tujuan akhir (atau biasa disebut sebagai label) yang dapat mengarahkan komputer untuk mencapai hasil yang diinginkan. Berbeda halnya dengan metode pembelajaran tidak terbimbing, komputer tidak diberi penanda/tujuan akhir tersebut dan diharapkan dapat menemukannya sendiri.

Kumpulan data dalam satu ruang lingkup yang sama biasa disebut sebagai dataset. Sebuah sistem dengan pembelajaran terbimbing membutuhkan dataset yang sudah diberi label. Proses pelabelan sendiri pada umumnya melibatkan ahli yang memang diakui dan punya keahlian pada bidang yang sesuai dengan ruang lingkup data yang akan dilabeli. Masalah yang umumnya dihadapi pada proses pelabelan data adalah jumlah data yang sangat besar, sehingga jika dilabeli oleh satu atau beberapa ahli saja akan membutuhkan waktu yang lama dan biaya yang besar [1].

*Crowdsourced labelling* adalah metode pelabelan data yang melibatkan partisipasi khalayak umum, tentunya untuk dataset yang tidak membutuhkan keahlian khusus untuk melabelinya, seperti kumpulan gambar buah yang perlu diberi label seperti mangga, jeruk, durian, tomat. Dengan pemberi label yang berjumlah banyak diharapkan proses pelabelan akan membutuhkan waktu yang lebih singkat. Biaya yang diperlukan untuk proses pelabelan pada model ini juga tidak sebesar jika menggunakan bantuan ahli untuk melakukannya.

Sebuah sistem *crowdsourced labelling* dapat dibangun berbasis web sehingga para pelabel dapat melakukan proses pelabelan data dengan mudah, hanya membutuhkan web browser yang umumnya sudah terpasang di perangkat komputer, smartphone maupun komputer tablet. Kualitas label yang dihasilkan dari proses *crowdsourced labelling* umumnya

tidak sebaik kualitas label yang dihasilkan dari ahli. Untuk memastikan label yang dihasilkan sudah tepat atau akurasiya baik, diperlukan metode-metode tambahan yang dapat mengolah hasil pelabelan dan menghasilkan label yang kualitasnya baik. Metode yang umumnya dipakai adalah metode *Majority Voting*, yaitu beberapa orang memberi label pada satu data yang sama, kemudian keputusan akhir diambil dari jumlah suara terbanyak.

Metode *Majority Voting* ini memiliki kelemahan pada kondisi jumlah suara mayoritas lebih dari satu. Perbaikan dari metode *Majority Voting* adalah *Weighted Majority Voting*, setiap suara diberi bobot, kemudian seluruh bobot dijumlahkan dan hasil akhir labelnya ditentukan dari rentang nilai jumlah bobot tersebut.

Sistem *crowdsourced labelling* berbasis web ini diharapkan dapat digunakan sebagai alat bantu pengumpulan dan pelabelan data, untuk kemudian dilanjutkan ke proses pembelajaran dan selanjutnya digunakan untuk sebuah sistem yang dapat melakukan deteksi, pengenalan maupun klasifikasi berdasarkan hasil pembelajaran tersebut.

Perumusan masalah yang diteliti adalah: (1). bagaimana perancangan alur kerja, basis data dan rancangan antarmuka untuk sebuah sistem *crowdsourced labelling* berbasis web?, (2). bagaimana mengimplementasikan metode *Weighted Majority Voting* untuk menentukan hasil pelabelan akhir berdasarkan kumpulan label dari suatu data? Sedangkan batasan penelitian ini adalah: (1). data yang akan diberi label adalah berupa data teks, (2). bobot yang digunakan dalam metode *Weighted Majority Voting* adalah bobot positif / 1, bobot negatif / -1 dan bobot netral / 0, dan, (3). setiap data akan diberi label dalam jumlah yang sama, dengan pelabelnya merupakan orang yang berbeda semuanya.

Tujuan dari penelitian ini secara umum adalah menghasilkan sebuah sistem *crowdsourced labelling* berbasis web dengan menggunakan metode *Weighted Majority Voting*. Penelitian ini diharapkan dapat dilanjutkan dengan penelitian-penelitian berikutnya: pembentukan dataset teks untuk kemudian dilanjutkan dengan penggunaan dataset tersebut untuk proses deteksi, pengenalan, klasifikasi maupun untuk analisis sentimen.

## II. LANDASAN TEORI

### A. Aplikasi Web

Aplikasi berbasis Web adalah sebuah aplikasi dengan arsitektur *client-server*, yang dijalankan di remote server dan diakses dari sisi client menggunakan *software web browser* [2]. Antara client dan server berkomunikasi menggunakan protokol HTTP (*Hypertext Transport Protocol*). Permintaan dari sisi client dinamakan *request*, sedangkan hasil permintaan yang dihasilkan oleh server dinamakan sebagai *response*. *Response* tersebut pada umumnya berupa sebuah dokumen HTML (*Hypertext Markup Language*).

Server yang digunakan pada umumnya adalah *web server*, seperti Apache, IIS, Nginx dan masih banyak web server lainnya. Aplikasi web dibangun menggunakan bahasa pemrograman, baik di sisi server maupun di sisi client. Bahasa pemrograman yang umumnya dipakai di sisi client adalah Javascript, sedangkan di sisi server adalah PHP, ASP.NET, Python, Perl, JSP dan beberapa bahasa pemrograman lainnya. Aplikasi Web umumnya juga menyimpan dan mengolah data yang berasal dari sebuah Database Server, seperti MySQL, Oracle, MS SQL Server maupun beberapa database server lainnya.

### B. Supervised Learning

Pada sebuah sistem cerdas dengan metode pembelajaran terbimbing (*supervised learning*), data pelatihan memiliki dua kumpulan informasi (*vectors*), yaitu informasi masukan (*input vectors*) maupun informasi keluaran (*target vectors*) yang diharapkan [3]. *Target vector* umumnya juga disebut sebagai label. Label bisa berupa suatu nama kategori (buah-buahan, hewan, kendaraan), nilai sentimen (mendukung, netral, tidak mendukung) maupun beberapa jenis label lainnya yang dapat didefinisikan sesuai kebutuhan.

### C. Data Labelling

Dalam proses pembuatan dataset untuk sistem klasifikasi, maka diperlukan mekanisme bagaimana agar dataset yang telah dikumpulkan memiliki label kelas yang benar. Pada kenyataannya dataset yang sudah dilabeli sangat sedikit dan sulit dicari [4]. Proses pelabelan dataset akan mudah jika datanya berjumlah

sedikit dan tidak terlalu besar, namun akan sangat membutuhkan waktu yang sangat lama bahkan tidak mungkin dikerjakan sendiri jika dataset berjumlah sangat besar. Proses pelabelan data dapat dilakukan secara manual sendirian atau dikerjakan bersama-sama oleh beberapa bahkan puluhan hingga ratusan orang menggunakan teknik *crowdsourced labelling*.

#### D. Crowdsourced Labelling

Proses pemberian label pada data berukuran besar menjadi sulit atau bahkan tidak mungkin dilaksanakan dalam waktu yang singkat. Pelabelan dapat dilakukan dalam waktu lebih singkat jika jumlah pelabel semakin banyak. Salah satu teknik yang memanfaatkan banyak pelabel untuk memberi label pada data berukuran besar adalah *Crowdsourced labelling*. Teknik pemberian label ini memiliki kemungkinan untuk menghasilkan kualitas label yang bervariasi. Tetapi dengan semakin banyak pelabel, maka kemungkinan untuk mendapatkan kualitas label yang lebih baik akan lebih besar.

Welinder dan Perona [5] menggunakan *crowdsourced labelling* untuk memberi label pada sekumpulan data gambar memanfaatkan layanan Amazon Mechanical Turk. Beberapa faktor yang diteliti antara lain kualitas label dan besarnya biaya yang diperlukan.

#### E. Weighted Majority Voting

Metode pelabelan data yang selama ini digunakan adalah menggunakan seorang pakar untuk melakukan pelabelan. Hal ini tentu sangat sulit atau bahkan tidak mungkin dilakukan untuk *dataset* yang berukuran besar yang akan membutuhkan lebih banyak pakar (dalam hal ini manusia) dan waktu yang sangat lama [1]. Teknik melakukan pelabelan menggunakan banyak pelabel membutuhkan metode khusus, seperti metode *Majority Voting*, metode *Expectation Maximization* [1]. Dari berbagai metode tersebut metode yang sederhana dan banyak digunakan adalah metode *Majority Voting*. Metode ini menggunakan konsep pengambilan keputusan hasil vote yang diperoleh dari jumlah terbesar dari masing-masing pilihan vote yang ada. Metode ini menggunakan Rumus 1 sebagai berikut (asumsi menggunakan 3 pelabel) [6]:

$$C(X) = \text{mode} \{ h_1(X), h_2(X), h_3(X) \} \dots (1)$$

Dimana:

$C(X)$  : class/label X

$h_1(X)$  : hasil vote pelabel 1

$h_2(X)$  : hasil vote pelabel 2

$h_3(X)$  : hasil vote pelabel 3

Dalam perkembangannya metode *Majority Vote* dapat diberi bobot menjadi *Weighted Majority Vote* sebagai berikut:

Masukkan data ke dalam suatu label jika (Rumus 2):

$$\sum_{i=0}^{N_c} w_{ik} \cdot \delta_{ik} = \max_{j=1}^{N_c} \sum_{i=0}^{N_c} w_{ij} \delta_{ij} \dots (2)$$

Keterangan:

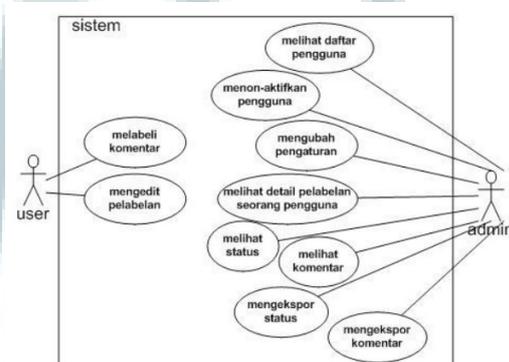
$$\delta_{ik} = \begin{cases} 1 & \text{if } E \text{ gives the class } k \\ 0 & \text{otherwise} \end{cases}$$

Dimana  $w_{ik}$  adalah bobot dan  $\delta_{ik}$  adalah label tertentu.

### III. HASIL DAN PEMBAHASAN

#### A. Use Case Diagram Sistem

Kebutuhan fungsional dalam pembangunan sistem *crowdsourced labelling* berbasis web menggunakan *Weighted Majority Voting* ini dapat dilihat pada use case diagram berikut:



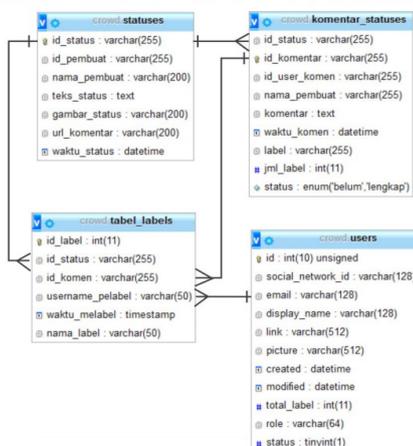
Gambar 1. Use Case Diagram

Penjelasan dari Gambar 1 di atas adalah sebagai berikut:

1. Pengguna sistem dibagi menjadi dua kategori, yaitu pelabel dan administrator.
2. Semua fungsi-fungsi di dalam sistem hanya dapat diakses oleh pengguna yang telah sah dan valid menggunakan informasi *username* dan *password*.
3. Administrator dapat menambahkan, mengubah, menghapus maupun melakukan pencarian data yang akan diberi label. Administrator dapat mendefinisikan label-label yang akan digunakan.
4. Administrator dapat menambahkan dan menghapus pengguna sistem.
5. Administrator dapat menentukan jumlah pelabel yang diharapkan untuk setiap data.
6. Sistem dapat menampilkan data yang akan diberi label satu-persatu secara acak. Pelabel dapat memberikan label yang dianggap sesuai dengan data yang ditampilkan.
7. Sistem dapat menghitung besaran rupiah (honor) untuk honorarium pelabel.
8. Administrator dapat melihat hasil pelabelan sementara dan dapat mengesponnya dalam bentuk *Comma Separated Value (CSV)*.

B. Skema Basis Data

Dari *use case diagram* yang dibuat maka Gambar 2 berikut adalah rancangan skema basis data yang akan digunakan:

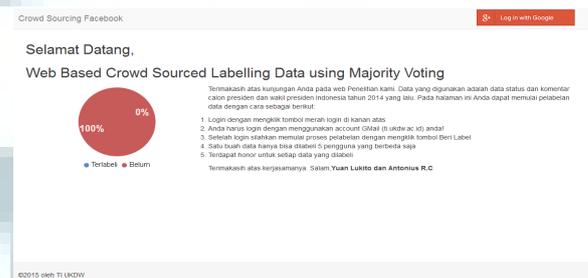


Gambar 2. Skema Basis Data

Secara umum ada empat tabel utama yang akan digunakan, yaitu tabel users, label, statuses, dan komentar\_statuses (keduanya adalah tabel dataset). Tabel users dan labels memiliki relasi one-to-many yang merepresentasikan setiap users dapat memberikan label pada beberapa data yang berbeda. Data yang akan diberi label disimpan pada tabel komentar\_statuses. Tabel statuses dan komentar\_statuses direlasikan dalam relasi one-to-many dengan yang berarti 1 status memiliki banyak komentar. Tabel status dengan tabel label memiliki relasi one-to-many yang artinya satu status bisa dilabeli lebih dari satu user. Tabel komentar\_status memiliki relasi one-to-many dengan tabel komentar\_status yang artinya 1 komentar bisa dilabeli oleh lebih dari satu user. Tabel users memiliki satu primary key id dan tabel labels memiliki primary key id\_label dan tiga foreign key id\_status, id\_komen, dan username\_pelabel yang melakukan referensi ke tabel statuses, tabel komentar\_statuses dan tabel users. Tabel komentar\_status memiliki primary key id\_komentar dan satu foreign key id\_status dari referensi tabel status.

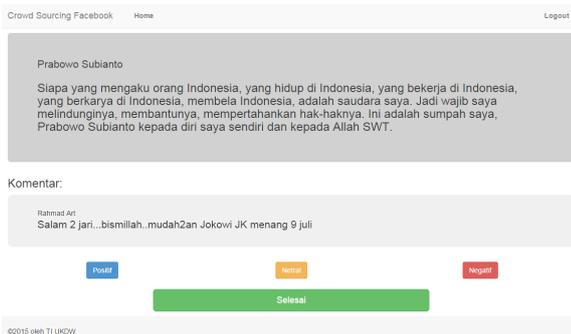
C. Implementasi Sistem

Proses awal sistem *crowdsourcing labelling* adalah pembagian pengguna. Pengguna dibagi menjadi 2 bagian, yaitu halaman pengguna biasa dan halaman administrator seperti pada Gambar 3 berikut:



Gambar 3. Implementasi Awal Sistem

Setelah login maka pengguna dapat memulai pelabelan dengan cara mengklik tombol beri label. Pada halaman awal setelah pengguna login, terdapat informasi foto pengguna, total pelabelan yang telah dilakukan, dan daftar data yang telah dilabeli dalam bentuk tabel yang bisa ditampilkan per halaman beberapa data tertentu.



Gambar 4. Implementasi Proses Pelabelan

Pada halaman seperti pada Gambar 4, akan selalu ditampilkan tepat satu status yang diambil dengan ketentuan sebagai berikut:

- Data diambil 1 buah status dari salah satu calon presiden secara random dan 1 komentar dari pengguna Facebook.
- Data yang diambil merupakan data komentar dan status yang memang belum dilabeli sebanyak N kali oleh pengguna sistem yang berbeda.

Di halaman ini pengguna dapat memberi label, dengan memilih salah satu, yaitu positif, netral, atau negatif. Jika pengguna telah memilih salah satu label, maka sistem akan langsung melakukan validasi label sebagai berikut:

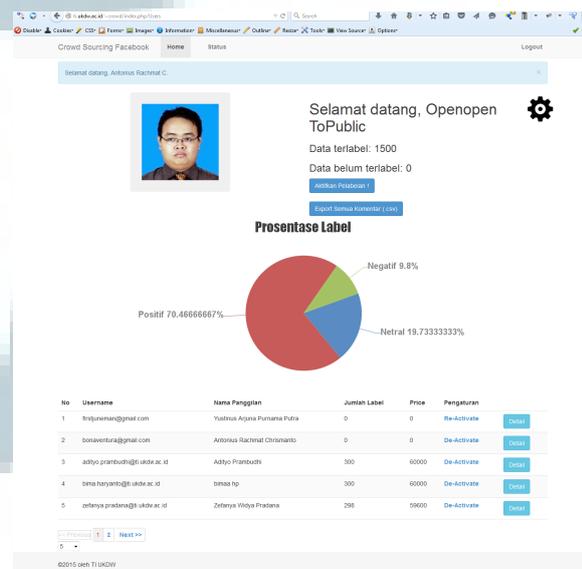
- Jika status dan komentar yang diberi label tersebut ternyata sudah mencapai N label dari N pelabel yang berbeda, maka sistem akan menampilkan informasi bahwa pelabelan gagal karena data sudah diberi label oleh pengguna lainnya.
- Jika status dan komentar yang diberi label memang belum diberi label oleh N jumlah pelabel yang berbeda maka sistem akan memberikan informasi pelabelan berhasil dan data label akan disimpan di dalam basis data yang telah dibuat sebelumnya, dan kemudian sistem akan merandom lagi untuk menampilkan data komentar baru lainnya kepada pengguna.

Pada halaman ini pengguna yang sama dapat terus menerus melabeli data sampai pengguna tersebut menghentikannya dengan mengklik

tombol selesai. Proses pelabelan akan dapat dilanjutkan lagi dilain waktu. Untuk setiap label yang diberikan sistem akan mencatat honor yang akan diterima oleh pengguna yang memberikan pelabelan tersebut. Pengguna juga dapat mengedit label yang sudah dibuat selama dataset masih belum ditutup waktu pelabelannya oleh admin.

Untuk halaman admin seperti pada Gambar 5, dimana admin dapat melakukan hal-hal sebagai berikut:

- Admin dapat melihat jumlah statistik data yang sudah dan belum terlabel
- Admin dapat mengubah dan mengatur besarnya honorarium per pelabelan dan jumlah maksimal label untuk satu data.
- Admin dapat melihat dan mengatur semua pengguna yang terdaftar, jumlah label yang telah diberikan, jumlah honor per pengguna dan bahkan bisa mendeaktifkan seorang pengguna dengan alasan tertentu.
- Untuk setiap pengguna dalam sistem admin juga dapat melihat daftar data yang telah diberi label oleh pengguna tersebut dan labelnya, seperti terlihat pada Gambar 6.



Gambar 5. Implementasi Halaman Admin

Hasil dari proses pelabelan data dapat diekspor dalam format (*Comma Separated Value*) CSV sehingga hasil tersebut dapat diunggah dan disebarluaskan ke berbagai media yang dapat digunakan sebagai dataset yang valid untuk membantu proses pembelajaran *supervised learning* seperti pada proses klasifikasi otomatis dalam bidang data mining atau text mining.

#### D. Implementasi Weighted Majority Voting

Metode *Weighted Majority Voting* diimplementasikan sebagai cara untuk melakukan pengambilan keputusan hasil pelabelan data yang dilakukan oleh banyak pelabel pada suatu saat terhadap satu data tertentu. Prosesnya adalah sebagai berikut:

- Satu buah data akan dilabeli oleh tepat N pelabel, dimana pelabel dapat mengisi label positif, negatif atau netral
- Setiap label positif akan diberi bobot +1, negatif -1, dan netral adalah 0
- *Weighted Majority Voting* digunakan untuk menghitung otomatis hasil akhir pelabelan setelah terpenuhinya satu data diberi tepat N label oleh N pelabel seperti pada rumus (1).

#### E. Pengujian Sistem

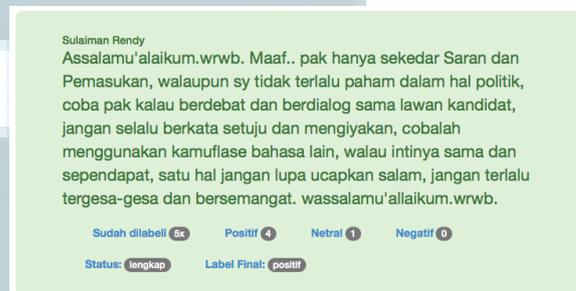
Pada akhir pengembangan aplikasi telah dilakukan pengujian terhadap penggunaan sistem yang dibangun dengan cara sebagai berikut:

1. Data komentar yang digunakan berjumlah 1500 komentar.
2. Pengujian *crowd sourced* menggunakan 5 orang pelabel dengan menggunakan ID email bersubdomain @ti.ukdw.ac.id.
3. Setiap data hanya boleh maksimal dilabeli oleh 5 orang pelabel.

Hasil yang diperoleh dari pengujian sistem ini adalah:

- Sistem berjalan normal dan tidak ada masalah sama sekali.

- Pelabelan dapat berakhir tepat 1500 data telah terlabeli oleh tepat 5 orang pelabel.
- Sistem mampu melakukan random data dengan baik.
- Sistem mampu menghitung hasil akhir pelabelan dengan menggunakan metode *Weighted Majority Voting*. Metode ini secara otomatis akan menghitung label mayoritas dari jumlah pelabel sesuai bobot masing-masing label (netral = 0, positif = +1, dan negatif = -1). Tampilan perhitungan otomatis dapat dilihat pada Gambar 6.



Gambar 6. Contoh Hasil Perhitungan Otomatis Menggunakan *Weighted Majority Voting*

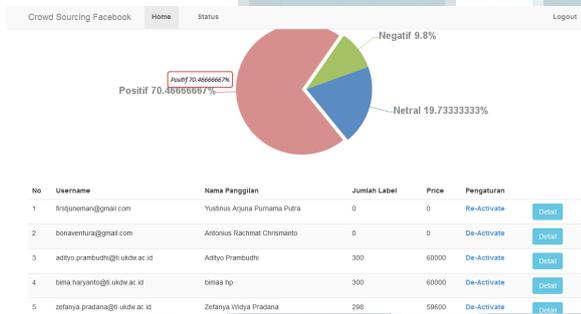
Komentar pada Gambar 6 mendapatkan 4 label positif dan 1 label netral. Hasil perhitungan label akhir adalah  $4 \times 1 + 1 \times 0 = 4$ , sehingga label akhirnya adalah positif. Walaupun komentar tersebut mengandung beberapa kata-kata yang bermakna negatif seperti **tidak**, **jangan** dan **walaupun**, tetapi menurut makna kalimat tersebut lebih merupakan dukungan positif. Hal ini menunjukkan bahwa pelabelan dengan sistem *crowdsourced labelling* ini lebih baik jika dibandingkan dengan sistem yang melakukan pelabelan otomatis berdasarkan filter kata atau kata kunci (*keywords*) tertentu.

- Sistem telah mampu menghasilkan file dataset yang telah terlabeli dan siap digunakan sebagai referensi pembelajaran sistem klasifikasi *supervised* untuk bidang text mining dalam bentuk file *Comma Separated Value* (CSV) yang dapat dipertanggungjawabkan, seperti terlihat pada Gambar 7.

A	B	C	D	E	F	G	H	I	J	K	L
1	id_status	id_pembuat	nama_pembuat	teks_status	gambar_urut	komentar_waktu	status	id_komen_id_user_k	komentar_waktu_komen	label	
2	233830011	23383001178	Bang Nafie	Saya	https://fb/https://w	6/9/2014 3:54	101520762	7.95E+14	Oke pak b	6/9/2014 3:55	positif
3	233830011	23383001178	Surya Kakasyi Suhaimy	Saya	https://fb/https://w	6/9/2014 3:54	101520762	3.35E+14	Jangan du	6/9/2014 3:55	negatif
4	233830011	23383001178	Rosbeto Redi	Saya	https://fb/https://w	6/9/2014 3:54	101520762	2.98E+14	PRIBOWI	6/9/2014 3:55	positif
5	233830011	23383001178	Sugjarnan Arman	Saya	https://fb/https://w	6/5/2014 3:54	101520762	2.99E+14	ilmu padi	6/5/2014 3:55	positif
6	233830011	23383001178	Rio Corps Taruna	Saya	https://fb/https://w	6/5/2014 3:54	101520762	1.52E+15	Hidup Pak	6/5/2014 3:55	positif
7	233830011	23383001178	Pemas Arum Sari	Saya	https://fb/https://w	6/9/2014 3:54	101520762	4.78E+12	Ganteng c	6/9/2014 3:55	positif
8	233830011	23383001178	Taufik	Saya	https://fb/https://w	6/9/2014 3:54	101520762	1.51E+15	anak, di l	6/9/2014 3:55	positif
9	233830011	23383001178	Anca Ardian	Saya	https://fb/https://w	6/9/2014 3:54	101520762	7.25E+14	ayo pak ru	6/9/2014 3:55	positif
10	233830011	23383001178	Rizaldi Piliang	Saya	https://fb/https://w	6/9/2014 3:54	101520762	8.85E+14	Prabowo i	6/9/2014 3:56	positif
11	233830011	23383001178	Ichy Da A Rocha Pugha	Saya	https://fb/https://w	6/9/2014 3:54	101520762	5.46E+14	HIDUP JOH	6/9/2014 3:56	negatif
12	233830011	23383001178	Mohammad Aqily AS	Tema	https://w	6/9/2014 6:42	101520852	7.81E+14	Hancurkar	6/9/2014 6:43	positif
13	233830011	23383001178	Muhammad Bagus Setia Rani	Tema	https://w	6/9/2014 6:42	101520852	3.56E+14	semoga pa	6/9/2014 6:43	positif
14	233830011	23383001178	Dewi Rahma	Tema	https://w	6/9/2014 6:42	101520852	6.5E+14	Aminin pa	6/9/2014 6:43	positif
15	233830011	23383001178	Erha Maniez	Tema	https://w	6/9/2014 6:42	101520852	7.62E+14	Kaum Nec	6/9/2014 6:43	negatif
16	233830011	23383001178	Iyan Nurdan	Tema	https://w	6/9/2014 6:42	101520852	1.49E+15	Menta Du	6/9/2014 6:43	negatif
17	233830011	23383001178	Dennis Afrillianto	Tema	https://w	6/9/2014 6:42	101520852	1.02E+15	Berikan pa	6/9/2014 6:44	positif
18	233830011	23383001178	Rizaldi Piliang	Tema	https://w	6/9/2014 6:42	101520852	8.85E+14	Prabowo	6/9/2014 6:44	positif
19	233830011	23383001178	Wilma Juanda	Tema	https://w	6/9/2014 6:42	101520852	5.42E+14	Prabowo	6/9/2014 6:44	positif
20	233830011	23383001178	Mada Sarnad	Tema	https://w	6/9/2014 6:42	101520852	5.24E+14	maju torn	6/9/2014 6:44	positif
21	233830011	23383001178	Radoet Nagit	Tema	https://w	6/9/2014 6:42	101520852	7.5E+14	Jawa Pami	6/9/2014 6:44	positif
22	233830011	23383001178	?????	Sapa	https://fb/https://w	6/13/2014 10:30	101520961	1.54E+15	???? ??	6/13/2014 10:30	positif
23	233830011	23383001178	Xtrmcomp Joga	Sapa	https://fb/https://w	6/13/2014 10:30	101520961	6.77E+14	no.1	6/13/2014 10:30	positif
24	233830011	23383001178	Muhammad Saiful	Sapa	https://fb/https://w	6/13/2014 10:30	101520961	1.02E+15	Indonesia	6/13/2014 10:30	positif
25	233830011	23383001178	Andrian Hadinata	Sapa	https://fb/https://w	6/13/2014 10:30	101520961	9.49E+14	Hidup ind	6/13/2014 10:30	positif

Gambar 7. Hasil CSV

Sedangkan sistem hasil pengujian dapat dilihat pada Gambar 8 dan Gambar 9 berikut ini.



Gambar 8. Hasil Pengujian Sistem



Gambar 9. Hasil Proses Pelabelan Final Seorang Pelabel

#### IV. SIMPULAN DAN SARAN

Berdasarkan perancangan, implementasi, pengujian dan analisis sistem dapat disimpulkan hal-hal sebagai berikut:

- Proses pengumpulan label data menggunakan sistem crowd dapat dilakukan dengan mengimplementasikan sistem *crowdsourced labelling* berbasis web menggunakan metode *Weighted Majority Voting* ini.
- Metode *Weighted Majority Voting* dapat digunakan untuk melakukan perhitungan otomatis dalam menentukan hasil label

data.

- Proses pelabelan data pada aplikasi berbasis web ini telah mampu menangani kejadian *collision* jika pada saat yang sama banyak yang melakukan pelabelan data pada data yang sama.
- Pada prinsipnya, sistem ini dapat menangani berbagai macam jenis data, tidak hanya data komentar sentimen terhadap calon presiden Indonesia saja.
- Sistem yang dibuat telah mampu mengeksport data hasil pelabelan dalam format *Comma Separated Value (CSV)* yang bersifat standar dan multiplatform sehingga dapat dipergunakan sebagai acuan dataset pada bidang penelitian lain seperti misalnya pada sentimen analisis dalam *text mining*.

Saran pengembangan di masa mendatang dari penelitian ini adalah: menambahkan metode perhitungan *voting* lainnya dan membuat setting kustom untuk berbagai dataset lain.

#### DAFTAR PUSTAKA

- [1]. M. Hosseini, I. J. Cox, N. Mili é -Frayling, G. Kazai and V. Vinay, 'On Aggregating Labels from Multiple Crowd Workers to Infer Relevance of Documents', *ECIR*, pp. 182-184, 2012.
- [2]. D. Nations, 'What is a Web Application?', *About.com Tech*, 2013. [Online]. Available: [http://webtrends.about.com/od/webapplications/a/web\\_application.htm](http://webtrends.about.com/od/webapplications/a/web_application.htm). [Accessed: 14- Feb- 2014].
- [3]. C. Bishop, *Pattern recognition and machine learning*. New York: Springer, 2006.
- [4]. E. Matsubara, M. Monard and R. Prati, 'Exploring Unclassified Texts Using Multiview Semisupervised Learning', *Emerging Technologies of Text Mining: Techniques and Applications*, pp. 139-161, 2008.
- [5]. P. Welinder and P. Perona, 'Online crowdsourcing: rating annotators and obtaining cost-effective labels', *Proceedings on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010.
- [6]. G. James, 'Majority Vote Classifiers: Theory and Applications', *Doctoral Dissertation*, Stanford, 1998.