

ARSITEKTUR APLIKASI TWITTER OPINION MINING UNTUK MENGETAHUI SENTIMEN PUBLIK TERHADAP MEREK PRODUK

Agus Sasmito Aribowo

Prodi Teknik Informatika UPN Veteran Yogyakarta

E-mail: sasmito_skom@yahoo.com

Abstrak

Merek suatu produk adalah ciri khas produk tersebut dan mencerminkan daya tarik serta kekuatan diterimanya suatu produk di tengah pasar. Opini publik pada suatu merek ini penting bagi pemilik merek karena pemilik merek dapat membantu memprediksi penjualan produk merek tersebut di masa depan dan mengetahui posisi suatu merek diantara kompetitornya. Data yang dapat dipakai oleh proses opinion mining dapat diperoleh dari twitter dan facebook. Karena data tersebut sangat variatif benuknya dan sangat tidak terstruktur maka harus dikelola dan diseragamkan menggunakan teknik data preprocessing menjadi data semi terstruktur. Data semi terstruktur ini dilanjutkan dengan tahap klasifikasi untuk mengkategorikan opini tersebut menjadi opini negatif atau positif. Penelitian tentang klasifikasi pada data opini berbentuk teks sudah banyak dilakukan tetapi umumnya metode-metode yang dipakai tidak mengakomodasi masalah pembobotan pada variabel-variabel pembentuk opini. Untuk mengakomodasi pembobotan maka penelitian ini menggunakan Weighted Association Rule Mining Classifier. Metodologi penelitian menggunakan studi literatur dimana penelitian akan menelaah penelitian sebelumnya dengan topik yang hampir sama. Telaah difokuskan pada menemukan arsitektur yang tepat untuk mengembangkannya menjadi aplikasi twitter opinion mining untuk mengetahui sentimen publik terhadap merek produk.

Kata Kunci : Twitter, Opinion Mining, Merek Produk

1. PENDAHULUAN

Salah satu bentuk informasi penting yang dibutuhkan oleh pemilik merek dagang (*brand*) adalah informasi tentang bagaimana opini masyarakat terhadap merek. Setiap merek mempunyai daya tarik luar biasa dalam pemasaran. Merek menjadi kekuatan diterimanya suatu produk di tengah pasar. Informasi tentang opini publik pada suatu merek ini penting bagi pemilik merek karena loyalitas terhadap suatu merek merupakan inti dari *brand equity* (kekuatan suatu merek). Opini terhadap suatu merek juga merupakan suatu indikator yang berkaitan dengan perolehan pendapatan di masa datang. Informasi opini konsumen terhadap suatu merek dapat membantu memprediksi penjualan produk merek tersebut di masa depan dan mengetahui posisi suatu merek diantara kompetitornya. Maka banyak pemilik merek berusaha mengetahui opini umum terhadap kualitas mereknya menggunakan metode *opinion mining*.

Data yang dapat dipakai oleh proses *opinion mining* tersebut dapat diperoleh dari sosial media elektronik yang sudah dikenal antara lain dari *twitter* dan *facebook*. Akan tetapi data tersebut sangat variatif dan tidak terstruktur. Data tidak terstruktur dan variatif perlu di strukturkan dan diseragamkan sehingga dapat diolah secara komputasional menggunakan teknik *data preprocessing*. *Data preprocessing* dipakai untuk proses awal data opini dan mengubahnya menjadi data semi terstruktur. Metode seperti TF-IDF dapat merangkum data opini semi terstruktur tersebut berdasarkan topik dan hasil *preprocessing* menjadi lebih jelas. Setelah tahap *preprocessing* dilakukan maka dapat dilanjutkan dengan tahap klasifikasi data semi terstruktur tersebut menjadi opini negatif atau positif. Penelitian sebelumnya banyak memanfaatkan metode *naive bayes* dan *Support Vector Machine* sebagai model untuk klasifikasi opini tersebut (Nur and Santika, 2011). Akan tetapi metode-metode tersebut tidak mengakomodasi masalah pembobotan pada variabel-variabel pembentuk opini. Untuk mengakomodasi pembobotan maka penelitian ini menggunakan *Weighted Association Rule Mining Classifier (WARMC)*. Metode ini dipakai untuk memproses opini-opini lama menjadi pengetahuan dalam bentuk *rule*. Kemudian *rule-rule* tersebut dipakai untuk mengklasifikasikan opini-opini baru sehingga dapat diketahui kelas dari opini-opini baru tersebut. Hasil klasifikasi akan diproses untuk mengetahui sentimen publik terhadap sebuah merek produk. Pada akhirnya dapat juga untuk meningkatkan nilai suatu produk di mata konsumen.

Bagaimana mengetahui kesimpulan opini umum terhadap suatu merek produk menggunakan *twitter opinion mining*. Bagaimana arsitektur aplikasi untuk memproses opini tersebut menjadi opini positif dan opini negatif. Permasalahan tersebut hendak dipecahkan dalam penelitian ini.

2. KAJIAN LITERATUR

a. *Opinion Preprocessing*

Opini biasanya berbentuk pesan teks singkat yang dapat diperoleh dari media sosial elektronik. *Text mining* adalah menambang data yang berwujud teks dimana sumber data biasanya didapatkan dari dokumen. Tujuan dari *text mining* adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Didalam proses *text mining* terdapat beberapa tahapan yaitu *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analyzing* (Nur and Santika, 2011).

- [1] Tahap *tokenizing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunya.
- [2] Tahap *filtering*. yaitu tahap mengambil kata-kata penting dari hasil token.
- [3] Tahap *stemming* adalah tahap mencari dasar kata dari tiap kata hasil filtering. Setiap kata yang memiliki imbuhan seperti imbuhan awalan dan akhiran maka akan diambil kata dasarnya.
- [4] Tahap *tagging* yang merupakan tahap mencari bentuk awal/root dari tiap kata lampau atau kata hasil stemming. Tahap ini tidak dipakai untuk teks bahasa yang tidak memiliki bentuk lampau.
- [5] Tahap *analyzing* yaitu tahap penentuan seberapa jauh keterhubungan antar kata-kata antar dokumen yang ada. Untuk melakukan analisa pada tahap analyzing dapat digunakan algoritma TF/IDF (*Term Frequency – Inversed Document Frequency*) dan Algoritma *Vector Space Model*.

b. *Weighted Association Rule Mining Classifier*

Soni (2009) menjelaskan bahwa *weighted associative classifier* berisi *data training* $T = \{r1, r2, r3, \dots, ri, \dots\}$ dengan himpunan bobot bagi setiap atribut. Bobot setiap record adalah himpunan dari tiga hal yaitu $\{ai, vi, w_i\}$ dimana atribut ai memiliki nilai vi dan bobot w_i dengan aturan bobot : $0 < w_j \leq 1$. Bobot dipakai untuk menunjukkan tingkat kepentingan dari item data.

Soni (2009) juga menjelaskan beberapa rumus terkait metode ini. Beberapa rumus adalah sebagai berikut.

[1] *Attribute Weight* (bobot atribut)

Bobot dari atribut ditentukan berdasarkan domain yang sedang diteliti. Contohnya atribut dari barang di supermarket dapat diberi bobot berdasarkan keuntungan penjualan per unit barang setiap item. Web mining dapat memanfaatkan lama kunjungan untuk menentukan bobot. Di dunia medis gejala suatu penyakit dapat diberi bobot oleh dokter ahli..

[2] *Attribute Set Weight* (himpunan bobot atribut)

Bobot dari himpunan atribut X dinotasikan $W(X)$ dan ini dihitung sebagai rata-rata dari bobot-bobot setiap atribut. $W(X)$ dirumuskan sebagai berikut :

$$W(X) = \frac{\sum_{i=1}^{|X|} \text{weight}(ai)}{\text{Jumlah atribut dalam } X}$$

[3] *Record weight* (bobot record)

Bobot record dapat didefinisikan sebagaimana bobot atribut. Ada rata-rata bobot atribut dalam record. Jika tabel memiliki n atribut maka bobot record dinotasikan sebagai $W(r)$ dan dirumusan sebagai berikut :

$$W(X) = \frac{\sum_{i=1}^{|X|} \text{weight}(ai)}{\text{Jumlah atribut dalam record}}$$

[4] *Weighted Support*

Pada metode *associative classification rule mining*, yang dimaksud dengan *association rules* adalah $X \rightarrow Y$ dimana Y adalah label kelas. *Weighted support* (WSP) dari rule $X \rightarrow \text{Class_label}$, dimana X

adalah himpunan subsets dari nilai attribute, adalah sebagian kecil dari bobot record yang mengandung nilai atribut relatif terhadap bobot semua transaksi

$$W(X) = \frac{\sum_{i=1}^{|X|} weight(ri)}{\sum_{k=1}^{|n|} weight(ri)}$$

Dalam rumus ini n adalah jumlah record

[5] *Weighted Confidence*

Weighted Confidence dari aturan $X \rightarrow Y$ dimana Y adalah class label dapat didefinisikan sebagai perbandingan antara *Weighted Support* dari (XUY) dengan *Weighted Support* dari (X). Perhitungan *Weighted Confidence* dapat dirumuskan :

$$Weighted\ Confidence = \frac{Weighted\ Support\ (X\ U\ Y)}{Weighted\ Support\ X}$$

Hasil dari proses *weighted association rule mining* adalah himpunan *rule-rule* yang merupakan pengetahuan untuk proses klasifikasi.

c. Proses Klasifikasi

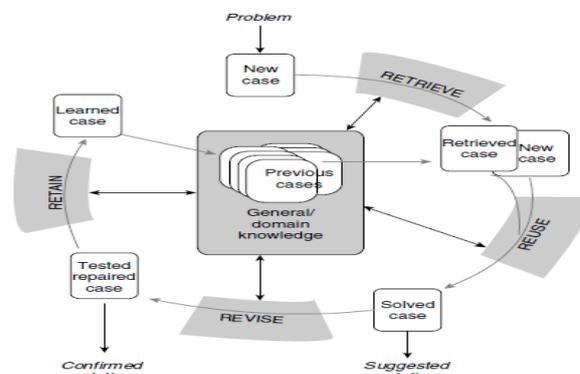
Klasifikasi akan mengklasifikasikan opini baru berdasarkan rule yang dihasilkan dari proses *association rule mining*. Cara menggunakan *rule* untuk mengklasifikasikan data adalah dengan cara sebagai berikut (Palanisamy, 2006).

- [1] Urutkan rule berdasarkan score *minimum support* dan *confidence* dari score terbesar hingga terkecil.
- [2] Cocokkan rule tersebut dengan data testing secara urut mulai dari rule dengan score terbesar hingga terkecil.
- [3] Jika sebuah data training memenuhi kriteria sebuah rule maka kelas dari rule tersebut menjadi label bagi data training tersebut.

d. *Case Based Reasoning (CBR)*

Case-Base Reasoning (CBR) merupakan sebuah sistem yang menggunakan pengalaman lama untuk dapat mengerti dan menyelesaikan masalah baru (Pal and Shiu, 2004). Keuntungan sistem penalaran berbasis kasus, yaitu:

- [1] Mengurangi dampak penambahan informasi pengetahuan, karena tidak memerlukan pemahaman bagaimana menyelesaikan masalah.
- [2] Tidak memerlukan suatu model yang eksplisit dan pengetahuan didapatkan dengan cara mengumpulkan kejadian-kejadian yang telah terjadi.
- [3] Kemampuan untuk belajar dengan menambahkan kasus baru seiring waktu tanpa perlu menambahkan aturan baru atau mengubah yang sudah ada.
- [4] Kemampuan untuk mendukung justifikasi dengan menawarkan kasus lampau lebih diutamakan.



Gambar 1 . Siklus case-base reasoning (Pal and Shiu, 2004).

Tahapan proses sistem penalaran berbasis kasus sebagaimana diperlihatkan dalam Gambar 1 adalah sebagai berikut (Pal and Shiu, 2004) :

[1] *Retrieve* :

Retrieve dilakukan pada saat melakukan proses pencarian dan menentukan kesamaan kasus lama dengan kasus baru. Pada penelitian ini proses penentuan kesamaan kasus lama dengan kasus baru dikembangkan menggunakan *opinion mining*.

[2] *Reuse*:

Reuse adalah menggunakan kembali informasi dan pengetahuan dalam kasus tersebut untuk mengatasi masalah. Proses *Reuse* yaitu pada saat ditemukannya kondisi yang sama yang ada dalam database dengan kondisi pada kasus sekarang.

[3] *Revise*:

Revise adalah meninjau ulang solusi yang diajukan. *Revise* dilakukan terhadap kasus baru yang tidak ada didalam database yang akan disimpan.

[4] *Retain*:

Retain adalah menelaah apakah hasil pemecahan masalah dapat digunakan dalam pemecahan masalah berikutnya. Proses *Retain* dilakukan dengan menyimpan hasil kasus baru kedalam sebuah *case-base*.

3. KAJIAN PENELITIAN TERDAHULU

Hu and Liu mengajukan metode data mining untuk mengidentifikasi fitur suatu produk. Pada penelitiannya metode *Association miner* diimplementasikan pada data transaksi berupa kata benda atau frase kata benda untuk mendapatkan *frequent features*. Setelah proses *pruning*, *nouns/noun phrase* yang dekat dengan *adjective words* dari *frequent features* akan diidentifikasi sebagai fitur opini (Hu and Liu, 2004). Penelitian terapan tentang *opinion mining* pernah dilakukan oleh Singh yang meneliti penerapan *opinion mining* pada *automated web-based course feedback system*. Sistem tersebut mengumpulkan respon pengunjung melalui *form online*. Teks yang diinputkan oleh pelajar pada sebuah teks editor dipakai untuk proses *opinion mining* menjadi berlabel 'positif' atau 'negatif'. Penelitian ini menggunakan metode *unsupervised SO-PMI-IR* untuk proses *opinion mining*. Umpan balik yang diperoleh dari sistem adalah sentimen 'positive' or 'negative' untuk menandai evaluasi yang diberikan bagi pelajar (Singh, et al., 2011). Penelitian terapan tentang *opinion mining* juga dilakukan oleh Brindha and Shanti untuk menilai kepemimpinan dalam organisasi perusahaan. *Opinion mining* digunakan untuk mengukur *performance* pekerja pada tingkat atas. Hasil yang diperoleh adalah sentimen 'moderate', 'good' and 'excellent' (Brindha and Shanti, 2012).

4. METODOLOGI PENELITIAN

Penelitian dilakukan dengan beberapa kegiatan untuk menganalisa proses-proses yang dibutuhkan untuk mengolah data opini dari *twitter* menjadi kesimpulan berupa persentasi opini positif dan negatif bagi sebuah merek. Langkah-langkah penelitian adalah sebagai berikut.

- a. Mengidentifikasi permasalahan, dimana permasalahan Bagaimana mengetahui kesimpulan opini umum terhadap suatu merek produk menggunakan *twitter opinion mining*. Bagaimana memproses opini tersebut menjadi opini positif dan opini negatif .
- b. Melakukan *literature review*
Penelitian dilanjutkan dengan mencari solusi permasalahan dengan mempelajari penelitian-penelitian sejenis tentang *opinion mining* and *text mining*.
- c. Analisa penyelesaian masalah
Permasalahan adalah bagaimana membentuk opini yang tidak terstruktur menjadi semi terstruktur kemudian mengklasifikasikannya dalam opini positif dan opini negatif.
- d. Perancangan arsitektur sistem
 - [1] Merancang model umum sistem.
 - [2] Merancang teknik *opinion mining*.

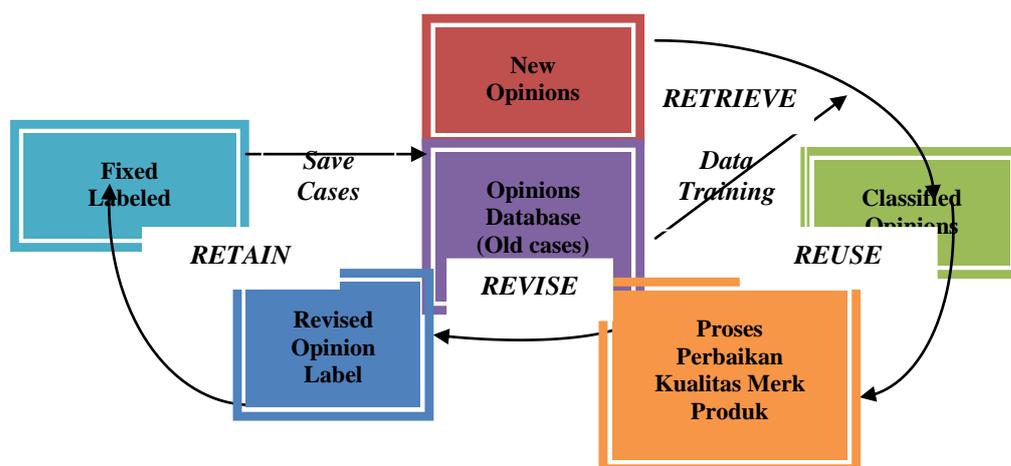
5. HASIL DAN PEMBAHASAN

Sistem dibangun menggunakan prinsip *machine learning* dimana opini-opini baru akan diklasifikasikan dengan belajar dari opini-opini terklasifikasi yang telah tersimpan sebelumnya. Siklus proses pembelajaran dapat ditentukan range jangka waktunya, misalnya dalam waktu satu bulan, satu

minggu, atau satu hari. Artinya sistem akan memproses opini yang diperoleh dalam jangka waktu tersebut, kemudian menyimpulkan dengan membagi opini menjadi opini yang menyatakan merek bersentimen "positif" (disebut dengan opini dalam klasifikasi baik) dan opini yang menyatakan merek bersentimen "negatif" (disebut sebagai opini dalam klasifikasi tidak baik). Opini diluar itu dinyatakan sebagai opini netral atau tidak diklasifikasikan.

Konstruksi sistem secara umum menggunakan prinsip penalaran berbasis kasus (*Case Based Reasoning*) ada pada Gambar 2. Penjelasan Gambar 2 adalah sebagai berikut :

1. New Opinions, sistem akan mengambil sejumlah opini dalam jangka waktu tertentu, misalnya dalam jangka waktu satu minggu, satu bulan, satu tahun tergantung kebutuhan.
2. RETRIEVE. Proses ini adalah proses opinion mining dimana opini-opini akan diklasifikasikan menjadi opini positif atau negatif berdasarkan sekumpulan kasus lama. Kasus lama diseleksi menjadi data training dan opini-opini baru menjadi data testing. Hasil proses ini adalah Classified Opinions.
3. Classified Opinion, adalah opini-opini yang telah terklasifikasi menggunakan metode data mining untuk memperbaiki kualitas brand produk.
4. REUSE, adalah proses perbaikan kualitas merek produk memanfaatkan opini positif dan negatif yang dihasilkan.
5. REVISE adalah proses perbaikan label/klasifikasi dari opini-opini baru yang sudah diklasifikasikan karena sangat mungkin ada bentuk opini baru yang salah dalam proses klasifikasinya. Hasilnya adalah opini-opini yang telah direvisi.
6. Revised Opinion adalah opini-opini yang sudah direvisi dan diklasifikasikan secara benar.
7. RETAIN adalah proses penstrukturan opini-opini menjadi opini yang siap disimpan dalam database kasus lama. Hasilnya adalah *fixed classified opinion*. Opini-opini yang sudah baku ini akan memperkaya database kasus.



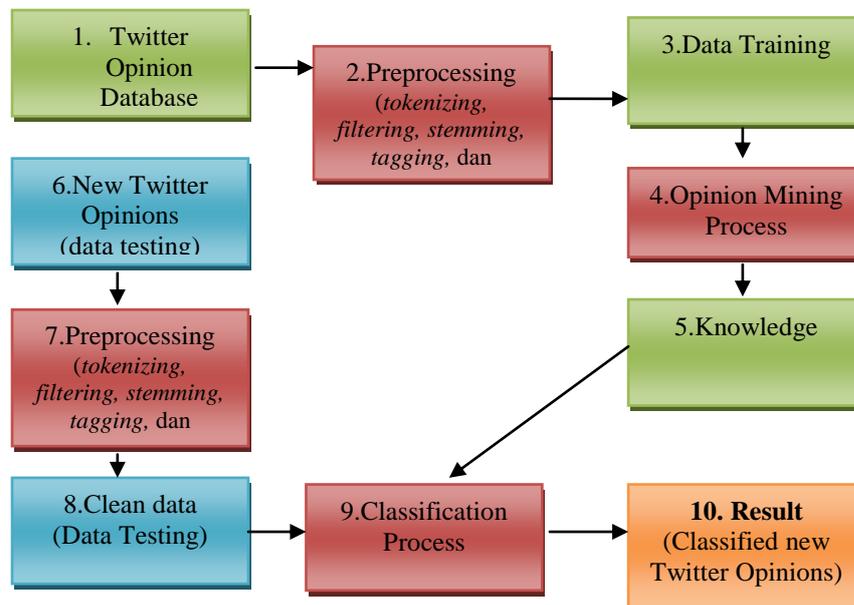
Gambar 2. Rancangan Arsitektur Twitter Opinion Mining Berbasis CBR.

Gambar 3. merupakan arsitektur *twitter opinion mining* yang sudah dirancang. Arsitektur ini terdapat pada proses RETRIEVE dari proses CBR pada Gambar 1. Pada diagram tersebut proses *opinion mining* merupakan proses inti dari sistem dimana dalam proses ini data opini dibentuk menjadi *rule*. *Rule* dipakai untuk mengklasifikasikan opini baru masuk dalam kategori opini positif maupun opini negatif.

Penjelasan proses dalam arsitektur Gambar 2 :

1. Langkah 1. Proses diawali dengan mengambil sejumlah opini *twitter* yang telah terklasifikasi sebagai data training. Opini tersebut dapat diseleksi dari database opini terklasifikasi pada periode sebelumnya.
2. Langkah 2. Tahap *preprocessing*. Tahap ini dilakukan untuk penyiapan data sehingga dapat diproses menggunakan metode data mining. Tahap preprocessing antara lain adalah *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analyzing*.
3. Langkah 3. Hasil dari tahap *preprocessing* adalah data training yang siap diproses *opinion mining*.
4. Langkah 4. *Opinion mining* menggunakan *Weighted Association Rule Mining Classifier* (WARMC) untuk menghasilkan *rule* (pengetahuan).

5. Langkah 5. *Knowledge* adalah hasil proses WARMC yang siap dipakai untuk mengklasifikasikan kasus baru.
6. Langkah 6. Proses *retrieve* dilanjutkan dengan mendapatkan data opini baru melalui twitter. Atribut utama dari opini yang dikumpulkan meliputi nama user, isi opini, tanggal dan waktu pengiriman opini.
7. Langkah 7. Tahap *preprocessing* terhadap opini baru. Tahap ini dilakukan untuk penyiapan data sehingga dapat diproses menggunakan metode *opinion mining*. Tahap *preprocessing* antara lain adalah *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analyzing*.
8. Langkah 8. Tahap *preprocessing* menghasilkan clean opinion data yang akan difungsikan sebagai data testing.
9. Langkah 9. Proses Klasifikasi, dimana data testing akan diklasifikasi menggunakan *rule* yang diperoleh dari proses WARMC.
10. Langkah 10. Hasil pengklasifikasian data testing dapat dipakai untuk analisa merek produk, yaitu kesimpulan dari opini berupa kualitas merek pada posisi "baik" (positif) atau "belum baik" (negatif).



Gambar 2. Arsitektur *Twitter Opinion Mining*

6. KESIMPULAN DAN SARAN

a. Kesimpulan

[1] Berdasarkan *literatur review* yang telah dilakukan diperoleh pengetahuan bahwa *Twitter opinion mining* adalah bagian dari *text mining* dimana opini dalam twitter jika hendak diklasifikasikan harus melalui tahap *text preprocessing* terlebih dahulu.

[2] Tahap *preprocessing* yang dibutuhkan dari data twitter adalah *tokenizing*, *filtering*, *stemming*, *tagging*, dan *analyzing*

[3] Arsitektur yang terbentuk merupakan alir nalar pemrosesan dari data opini *twitter* yang tidak terstruktur menjadi sentimen positif atau negatif terhadap merek.

b. Saran

Penelitian perlu dilanjutkan dengan mengimplemntasikan arsitektur menjadi rancangan-rancangan aplikasi, sistem, basis data sehingga dapat diimplementasikan menjadi aplikasi nyata yang siap dipakai untuk mengelola *twitter opinion mining*.

DAFTAR PUSTAKA

- Brindha G. R.#, Santhi.B, 2012, *Application Of Opinion Mining Technique In Talent Management* in Journal IEEE, 978-81-8424-759-6
- Hu, M., & Liu, B. (2004). *Mining And Summarizing Customer Reviews*. In Proceedings Of SIGKDD (Pp. 168–177).
- Nur, M.Y Dan Santika, D.D., 2011, *Analisis Sentimen Pada Dokumen berbahasa Indonesia Dengan Pendekatan Support Vector Machine*, Konferensi Nasional Sistem Dan Informatika 2011
- Palanisamy, S.K., 2006, *Association Rule Based Classification, Thesis*, Master Science Faculty Of The Worcester Polytechnic Institut.
- Pal, Shankar K. And Shiu, Simon C.K., *Foundations Of Soft Case Based Reasoning*. Willey & Sons, Inc. Canada. Pp 64 – 67. 2004.
- Singh, V. K., Kumari, P., Singh, A., Thapa, J., 2011, *An Automated Course Feedback System Using Opinion Mining*, Journal IEEE 978-1-4673-0126-8/11
- Soni, S, Pillai, J, Vyas, O.P, 2009, *An Associative Classifier Using Weighted Association Rule* , IEEE, 978-1-4244-5612-3/09.