

Pemeringkatan Jawaban pada Community Question Answering dengan Tekstual Fitur dan Pemodelan Topik Answer Ranking in Community Question Answering with Textual Feature and Topic Models

Luh Putri Ayu Ningsih¹, Ade Romadhony², Moch. Arif Bijaksana³

^{1,2,3} Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung
Jalan Telekomunikasi No. 1, Dayeuh Kolot, Bandung 40257

¹ luhputriayu@gmail.com, ² ade.romadhony@gmail.com, ³ arifbijaksana@gmail.com

Abstrak

Community question answering (CQA), seperti qatar living forum telah menjadi salah satu tempat bagi pengguna internet untuk mendapatkan informasi. Satu pertanyaan pada CQA dapat memiliki banyak jawaban, namun pengguna harus memilih jawaban yang paling sesuai dengan pertanyaannya secara manual yang membutuhkan waktu cukup lama. Permasalahan ini dapat diatasi dengan membangun sistem pemeringkatan jawaban dari list jawaban yang telah ada dimana jawaban yang sesuai dengan pertanyaan diatas jawaban yang tidak sesuai dengan pertanyaan. Proses pertama yang akan dilakukan dalam penelitian yaitu teks preprocessing. Setelah data teks menjadi lebih teratur dilakukan ekstraksi fitur, dimana fitur yang akan digunakan adalah tekstual fitur dan pemodelan topik. Tekstual fitur adalah mengidentifikasi ciri-ciri sebuah jawaban yang Good dan bad dengan melihat elemen-elemen teksnya seperti melihat apakah sebuah jawaban mengandung tanda tanya (?), emotikon, link atau kata-kata tertentu. Pemodelan topik merupakan pemodelan data tekstual yang bertujuan menemukan variabel tersembunyi. Dalam Penelitian ini akan fokus pada penggunaan pemodelan topik untuk mencari kemiripan antar pertanyaan dan jawaban. Hasil ekstraksi fitur ini akan dijadikan inputan untuk classifier untuk membuat model yang akan digunakan oleh data uji. Pada pengerjaan tugas akhir ini menggunakan Support Vector Machine (SVM) dan logistic regression untuk mendapatkan score klasifikasi dimana score ini yang menentukan peringkat sebuah jawaban untuk setiap pertanyaan. Berdasarkan hasil evaluasi dari penelitian yang dilakukan penulis didapatkan bahwa sistem pemeringkatan yang dibuat memiliki nilai mean average precision sebesar 71.9. Hasil ini didapatkan menggunakan logistic regression sebagai classifiernya. Jika dibandingkan dengan hasil peserta SemEval 2016 task 3 hasil yang didapatkan dalam penelitian ini berada di ranking 8 dari 13 peserta.

Kata Kunci: community question answering, tekstual fitur, pemodelan topik, ekstraksi fitur, Peringkat

Abstract

Community question answering (CQA), like qatar living forum has become a place for Internet users to get information. One question on the CQA can have many answers, but the user must choose the most appropriate answer manually which takes a long time. This problem can be overcome by establishing an answer ranking system from list of answer. The appropriate answers to the questions above answers are not in accordance with a question. The first step is text preprocessing. After the text data become more structured next process is feature extraction where the feature is textual features and topic modelling. Textual features are identifying features of an answer Good and bad to see the elements of the text as to see if an answer contains a question mark (?), Emoticons, links or special words. Topic modelling is textual data modeling aimed at finding hidden variables. In this research will focus on the use of topic modelling to find similarity between questions and answers. This feature extraction results will be used as input to classifier to create a model that will be used by the data testing. In this final project using Support Vector Machine (SVM) and logistic regression to get a score classification where these score would rank the answers to every question. Based on the evaluation of research by the author found that the rating system that is made has a mean average precision value of 74.8. The results obtained using logistic regression as classifier. if compared with the results of SemEval 2016 the results obtained in this study was ranked eighth from 13 participant.

Key words: community question answering, 1 textual feature, topic modelling, feature extraction, rank

1 Pendahuluan

Saat ini perkembangan teknologi sangatlah pesat terutama teknologi internet. Pengguna Internet dapat menggunakan mesin pencari untuk mencari informasi namun kelemahan mesin pencari adalah saat penggunanya mengin-

putkan pertanyaan maka mesin pencari akan mencari dokumen yang relevan dengan pertanyaan tersebut tetapi tidak dapat memberikan jawaban secara langsung. Saat ini banyak komunitas tanya-jawab (Community Question Answering) seperti qatar living forum . CQA memegang peran penting untuk membantu penggunanya mendapatkan informasi yang diinginkan. Pengguna dapat memuat pertanyaan mereka pada website cQA, Pengguna yang lain bebas memuat jawaban mereka terhadap pertanyaan tersebut. Sehingga satu pertanyaan dapat memiliki list jawaban yang banyak dan beragam. Jawaban tersebut dapat memuat informasi yang relevan atau tidak relevan sama sekali dengan pertanyaan sehingga dibutuhkan waktu untuk memilih jawaban yang paling sesuai dengan pertanyaannya maka dibutuhkan sistem yang secara otomatis memberikan peringkat pada list jawaban tersebut.

Berdasarkan hal tersebut, penelitian ini akan membuat suatu aplikasi yang dapat memberikan peringkat pada jawaban terhadap suatu pertanyaan. Fitur yang digunakan untuk klasifikasi dan pemeringkatan jawaban adalah tekstual fitur dan pemodelan topik yang dijelaskan pada bab ekstraksi fitur. Klasifikasi dan pemeringkatan jawaban akan menggunakan metode support vektor mesin(SVM) dan logistic regression. Sistem ini diharapkan dapat memberikan peringkat pada jawaban sesuai dengan kualitas jawaban tersebut terhadap pertanyaannya. Serta dapat melihat fitur-fitur apa saja yang baik dalam pemberian peringkat pada jawaban.

2 Dasar Teori dan Perancangan Sistem

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah dataset dari Qatar Living forum yang telah disediakan di Pada Semeval 2016 task 3 ¹ dalam bentuk file xml. Terdapat dua dataset yang digunakan yaitu data latih dan data uji. Data yang akan digunakan pada penelitian ini yaitu data pada related question dan pasangan jawabannya (comment). Setiap related questions(pertanyaan) memiliki masing-masing 10 comments(jawaban). Rincian dari masing masing dataset dapat dilihat pda tabel 1

Tabel 1: Statistik Dataset

Category	Training	Testing
Question	1,000	327
Answers	10,000	3,270
good potential	3,913	1,329
useful bad	1,620	456
	4,467	1,485

2.2 Ekstraksi Fitur

Ekstraksi fitur merupakan suatu pengambilan ciri / feature dari suatu bentuk yang nantinya nilai yang didapatkan akan dianalisis untuk proses selanjutnya. Fitur merupakan karakteristik unik dari suatu objek, dimana objek dalam penelitian berupa teks [3].Fitur yang menjadi fokus utama penelitian ini yaitu Tekstual dan pemodelan topik yang diambil dari tim JAIST.

2.2.1 Tekstual Fitur

Fitur ini akan mengidentifikasi semua aspek teks yang dapat menunjukkan kualitas sebuah jawaban. Fitur ini terdiri atas.

1. Cosine Similarity

Cosine Similarity menghitung nilai kosinus sudut antara dua vektor .Cosine semilarity dapat dihitung dengan persamaan 1.

$$\text{Cosine}_{sim} = \frac{\sum_{i=1}^n u_i \times v_i}{\sqrt{\sum_{i=1}^n (u_i)^2} \times \sqrt{\sum_{i=1}^n (v_i)^2}} \tag{1}$$

¹<http://alt.qcri.org/semeval2016/task3/index.php?id=data-and-tools>

Dimana u dan v adalah vektor dari pertanyaan dan jawaban, isinya adalah jumlah kemunculan kata (Term frequency). n merupakan ukuran dari vektor u dan v . [8].

2. Tanda Tanya(?)

Fitur ini akan mengidentifikasi konten jawaban apakah mengandung tanda tanya(?). Jika ditemukan tanda tanya pada teks jawaban dapat diartikan bahwa jawaban mengarah pada pertanyaan baru bukan menjawab pertanyaan sehingga dapat diartikan jawaban tersebut memiliki kualitas buruk (bad) [8].

3. Emotikon

Fitur ini akan mengidentifikasi konten jawaban apakah mengandung token tertentu. Token khusus tersebut yaitu simbol tertawa. Jika ditemukan simbol tertawa dapat diartikan bahwa jawaban tersebut tidak serius, sehingga dapat diartikan jawaban tersebut memiliki kualitas buruk (bad) [8].

4. Kata-kata khusus (Special Words)

Pada feature ini akan mengidentifikasi apakah konten dari jawaban mengandung beberapa kata yang khusus untuk menunjukkan kualitas jawaban (good, bad). Jawaban yang baik biasanya berisi kata-kata yang mengandung makna mengajak, oleh karena itu penulis secara manual memasukan beberapa list kata dari data training yaitu suggest, recommend, advise, try, call, you may, maybe, you could [8].

5. Link

Pada feature ini akan mengidentifikasi apakah konten pada jawaban mengandung link/tautan. Secara dasar jika jawaban itu mengandung tautan maka jawaban tersebut tidak menjawab langsung pertanyaan tapi mengalihkan ke halaman lain. Dimana link yang diberikan dapat membantu menjawab pertanyaan sehingga dapat diartikan jawaban tersebut memiliki kualitas baik (good) [4].

Fitur tanda tanya, emotikon, kata-kata khusus (special words), link akan mengembalikan nilai biner yaitu 0 jika tidak ditemukan fitur tersebut dan 1 jika ditemukan fitur tersebut.

2.2.2 Pemodelan Topik

Pemodelan topik mencakup pemodelan data tekstual yang bertujuan menemukan variabel tersembunyi yaitu sebuah topik. Data tekstual pada penelitian ini berupa kumpulan pertanyaan dan jawaban kemudian diolah menggunakan pemodelan topik. Berdasarkan Penelitian Blei dkk. [1] dan Griffiths dan Steyvers [6], penelitian ini akan menggunakan model LDA dan algoritma dari MCMC (Markov Chain Monte Carlo) yaitu Gibbs sampling untuk estimasi parameter dari model LDA.

Latent Dirichlet Allocation (LDA)

1. Proses Generatif

Diasumsikan topik telah dispesifikasikan sebelum didapatkan dokumen. Lalu untuk setiap dokumen di dalam koleksi dilakukan ekstraksi kata-kata dengan 2 tahap:

(a) Secara acak dipilih distribusi atas topik,

(b) Untuk setiap kata dalam dokumen:

- Secara acak dipilih sebuah topik dari distribusi atas topik pada langkah 1,
- Secara acak dipilih sebuah kata dari distribusi yang sesuai, atas kosakata.

2. Collapsed Gibbs Sampling

Dalam prakteknya kita tidak dapat menemukan variabel yang tersembunyi dengan model generatif LDA. Untuk menemukan variabel tersembunyi ini digunakan Collapsed Gibbs Sampling [2]. Direpresentasikan kumpulan dokumen sebagai koleksi dari kata dengan index w_i dan dokumen dengan index d_i , untuk setiap token kata i . Prosedur Gibbs Sampling menganggap setiap token kata dalam kumpulan dokumen mempunyai giliran untuk diproses, untuk mengestimasi probabilitas dalam menetapkan tiap kata ke dalam suatu topik. Dari distribusi kondisional ini, sebuah topik dijadikan sample dan disimpan sebagai penetapan topik yang baru untuk sebuah token kata. Distribusi kondisional ini ditulis dengan $P(z_i = j | z_i, w_i, d_i, \dots)$, dimana $z_i = j$ merepresentasikan penetapan topik dari token d_n ke topik j , Z merepresentasikan penetapan topik untuk semua token

kata, dan . merupakan semua informasi yang sudah terobservasi seperti kata w_i dan dokumen, Griffiths dan Stevyers (2004) menuliskan perhitungan ini sebagai:

$$P(z_i = j | z_i, w_i, d_i \dots) = \frac{P_w C_{i,j}^{wT} + \beta}{\sum_{w=1}^W C_w^T + W\beta} \frac{P_T C_{d_i,j}^{DT} + \alpha}{\sum_{t=1}^T C_{d_i,t}^{DT} + T\alpha} \tag{2}$$

Variabel C^{wT} dan C^{DT} merupakan matrix jumlah dengan dimensi $W \times T$ dan $D \times T$, dimana $C_{i,j}^{wT}$ berisi berapa kali kata w ditetapkan ke topik j dan $C_{d_i,j}^{DT}$ berisi berapa kali topik j ditetapkan ke sebuah token kata pada dokumen d . Algoritma Gibbs Sampling dimulai dengan menetapkan setiap token kata ke sebuah topik secara random [7]. Untuk setiap token kata, jumlah dari C^{wT} dan C^{DT} dikurangi satu untuk tiap entri yang berkorespondensi ke topik yang ditetapkan saat ini. Lalu, topik baru ditetapkan sesuai dengan persamaan 2 dan jumlah dari C^{wT} dan C^{DT} dinaikkan satu sesuai dengan topik yang baru. Setiap Gibbs sample terdapat penetapan topik untuk semua token kata dalam kumpulan dokumen yang didapat dengan melakukan iterasi ke semua kata dalam dokumen.

2.3 Evaluasi Performansi Sistem

2.3.1 Evaluasi Klasifikasi

Dalam klasifikasi terdapat empat pengukuran performansi yaitu akurasi, precision, recall, f1-measure.

Tabel 2: Confusion Matrics

Actual Class	Predicted Class	
	True	False
True	True Positive(TP)	False Positive(FP)
False	False Negative(FN)	True Negative(TN)

$$\text{Akurasi} = \frac{TP + TN}{|TP + FP + TN + FN|} \tag{3}$$

$$\text{Precision}(p) = \frac{TP}{|TP + FP|} \tag{4}$$

$$\text{Recall}(r) = \frac{TP}{|TP + FN|} \tag{5}$$

$$\text{F1 - Measure}(F1) = 2 \cdot \frac{p \cdot r}{p + r} \tag{6}$$

2.3.2 Evaluasi sistem pemeringkatan

Evaluasi sistem pemeringkatan yang digunakan adalah mean avarage precision(MAP).Mean Average Precision hanya akan mempehitungkan dokumen yang relevan saja, dengan menggunakan Mean Average Precision level recall tidak ditetapkan secara baku. Mean Average Precision untuk sebuah koleksi dokumen merupakan rata-rata precision untuk sebuah kebutuhan informasi. Misal himpunan jawaban yang relevan terhadap pertanyaan $p_j \in Q$ adalah j_1, j_2, \dots, j_{m_j} , dan R_{jk} adalah himpunan hasil jawaban yang diranking dari atas sampai ke dokumen dk , maka [9]:

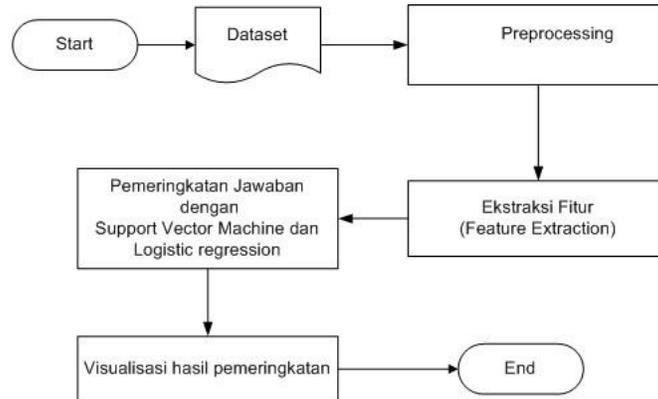
$$MAP = \frac{1}{|Q|} \sum_{j=1}^{|Q|} \sum_{k=1}^{|R_{jk}|} P @ R_{jk} \tag{7}$$

dimana $P @ R_{jk}$ adalah Precision dari R_{jk} , yaitu

$$P @ R_{jk} = \frac{\text{Jumlah Jawaban yang relevan pada } R_{jk}}{|R_{jk}|} \tag{8}$$

2.4 Perancangan Sistem

Secara umum rancangan sistem yang akan dibangun dapat dilihat pada Gambar 1.



Gambar 1: Gambaran Umum Sistem

Berdasarkan dari Gambar 1 , maka alur atau gambaran umum dari sistem ini adalah sebagai berikut :

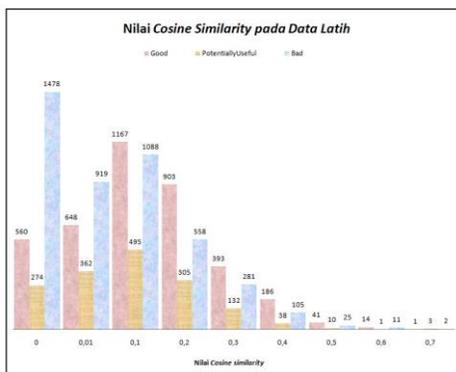
1. Sistem menerima masukan dataset berupa dokumen xml, dimana dataset ini merupakan data kumpulan pertanyaan dan jawaban,dimana satu pertanyaan mempunyai 10 jawaban,
2. setelah dataset didapat, dilakukan preprocessing,
3. data hasil preprocessing akan diekstrak fiturnya dengan tekstual fitur dan pemodelan topik,
4. data hasil ekstraksi fitur yang didapatkan digunakan untuk proses klasifikasi dan pemberian peringkat pada jawaban,
5. Setelah didapatkan peringkat jawaban pada masing masing pertanyaan dilakukan visualisasi.

3 Pembahasan

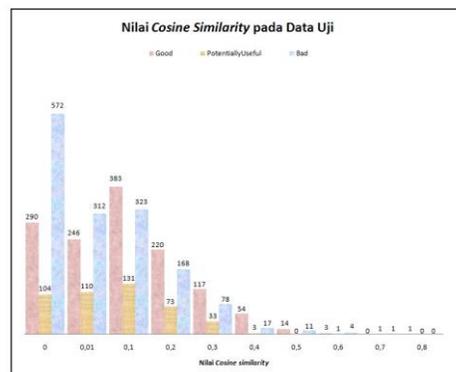
3.1 Implementasi Fitur

3.1.1 Tekstual Fitur

1. Cosine Similarity



Gambar 2: Cosine Similarity Data Latih



Gambar 3: Cosine Similarity Data Uji

Dari data pada Gambar 2 dan Gambar 3 menunjukkan bahwa kelas Bad banyak memiliki nilai cosine similarity antara angka 0 sampai 0,1, Sedangkan pada nilai yang lebih besar frekuensi kelas Bad sedikit dan frekuensi kelas Good lebih banyak. Hal ini menunjukkan bahwa cosine similarity dapat digunakan untuk menunjukkan kualitas jawaban dimana jawaban dengan nilai cosine similarity yang kecil dapat digolongkan pada kelas Bad dan jawaban dengan nilai yang besar dapat digolongkan pada kelas Good.

2. Tanda Tanya

Tabel 3: Hasil Fitur Tanda Tanya

Kelas	Data Latih			Data Uji		
	Good	PotentiallyUseful	Bad	Good	PotentiallyUseful	Bad
True	367	229	1523	123	78	449
False	3546	1391	2944	1206	378	986
Persentase True(%)	9.4	14	34	10.2	17	33.6

Dari Tabel 3 dapat dilihat bahwa kelas Bad memiliki lebih banyak tanda tanya sehingga dapat disimpulkan bahwa fitur ini dapat menunjukkan jawaban yang tergolong pada kelas Bad.

3. Emotikon

Tabel 4: Hasil Fitur Emotikon

Kelas	Data Latih			Data Uji		
	Good	PotentiallyUseful	Bad	Good	PotentiallyUseful	Bad
True	249	122	474	90	27	150
False	3664	1498	3993	1239	429	1335
Persentase True(%)	6.4	7.5	10.6	6.8	5.9	10

Dari tabel 4 dapat dilihat bahwa kelas Bad memiliki lebih banyak emotikon sehingga dapat disimpulkan bahwa Emotikon dapat menunjukkan jawaban yang tergolong pada kelas Bad.

4. Kata-kata khusus(Special Words)

Tabel 5: Hasil Fitur Special Words

Kelas	Data Latih			Data Uji		
	Good	PotentiallyUseful	Bad	Good	PotentiallyUseful	Bad
True	887	329	442	214	71	114
False	3026	1291	4025	1115	385	1341
Persentase True(%)	22.7	20.3	9.9	16.2	15.6	9.7

Dari Tabel 5 terlihat bahwa kelas Good memiliki lebih banyak Special Words sehingga dapat disimpulkan Special Words dapat menunjukkan jawaban dengan kelas Good.

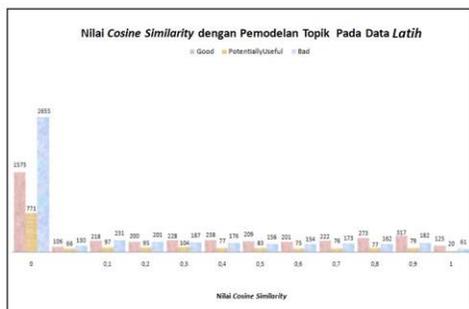
5. Link

Tabel 6: Hasil Fitur Tanda Tanya

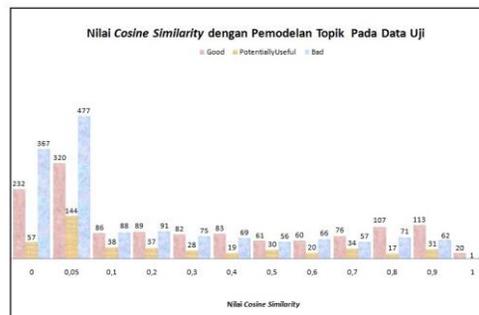
Kelas	Data Latih			Data Uji		
	Good	PotentiallyUseful	Bad	Good	PotentiallyUseful	Bad
True	156	102	95	45	40	27
False	3757	1518	4372	1284	416	1458
Persentase True(%)	4	6.3	2.1	3.4	8.8	1.8

Dari tabel 6 dapat dilihat bahwa kelas PotentiallyUseful memiliki lebih banyak link sehingga dapat disimpulkan link dapat menunjukkan jawaban dengan kelas PotentiallyUseful.

3.1.2 Pemodelan topik



Gambar 4: Hasil LDA Data Latih



Gambar 5: Hasil LDA Data Uji

Dari Gambar 4 dan Gambar 5 menunjukkan bahwa kelas Bad banyak memiliki nilai cosine similarity antara angka 0 sampai 0,01. Sebaliknya frekuensi kelas Good pada nilai yang lebih besar lebih banyak dibandingkan pada nilai yang lebih kecil. Hal ini menunjukkan bahwa cosine similarity berdasarkan pemodelan topik dapat digunakan untuk menunjukkan kualitas jawaban dimana jawaban dengan nilai cosine similarity yang kecil dapat digolongkan pada kelas Bad dan jawaban dengan nilai yang besar dapat digolongkan pada kelas Good.

3.2 Hasil Klasifikasi dan Pemingkatan Jawaban

Berdasarkan hasil analisis terhadap penggunaan logistic regression dan SVM dalam proses klasifikasi dan pemberian peringkat jawaban didapatkan bahwa nilai evaluasi yang dihasilkan (MAP dan F-measure) oleh classifier logistic regression lebih baik dibandingkan nilai yang dihasilkan SVM classifier. Hal ini dapat dilihat pada tabel 7 yang memperlihatkan perbandingan hasil klasifikasi menggunakan logistic regression dan SVM

Tabel 7: Perbandingan Hasil Logistic Regression dengan SVM

No	Classifier	MAP	F-Measure	Precision	Recall
1	Logistic Regression	71.9	52.5	49	57.1
2	SVM	71.8	52.1	55.5	56.6

Berdasarkan hasil analisis ini, maka classifier yang akan digunakan untuk menganalisis pengaruh fitur terhadap proses klasifikasi dan pemingkatan jawaban adalah logistic regression .

3.3 Analisis Pengaruh fitur

Tabel 8: Daftar dan hasil Kombinasi fitur

Kombinasi	Fitur	Map	accuracy	Precision	Recall	F1-measure
1	Pemodelan Topik + Cosine Similarity	67,6	51,1	44,1	51,1	46,2
2	Pemodelan Topik + Tanda Tanya	69,8	54,5	47	54,6	50
3	Pemodelan Topik + Emotikon	65,6	49,5	42,4	49,5	44,8
4	Pemodelan Topik + Link	65,3	49,6	42,4	49,6	45
5	Pemodelan Topik + Spesial Words	65,5	49,5	42,4	49,5	45,1
6	Cosine Similarity + Tanda Tanya + Spesial Words + Pemodelan Topik	71,6	56,7	48,8	56,7	52,1
7	Cosine Similarity + Tanda Tanya + Emotikon + Link + Spesial Words	65,5	53,5	50,7	53,5	48,2
8	Cosine Similarity + Tanda Tanya + Pemodelan Topik	72,7	57,1	49,1	57,1	52,6

Berdasarkan data pada tabel 8 dapat dilihat bahwa fitur yang paling berpengaruh pada proses klasifikasi dan pemberian peringkat pada jawaban adalah fitur pemodelan topik, cosine similarity dan tanda tanya. Sedangkan untuk link, emotikon dan special word kurang berpengaruh karena seperti dapat dilihat pada analisis implementasi fitur jumlah data jawaban yang memiliki fitur link dan emotikon sedikit sehingga susah diidentifikasi pada proses klasifikasi.

4 Kesimpulan

Berdasarkan analisis dan pengujian yang telah dilakukan, maka kesimpulan yang dapat diambil adalah sebagai berikut :

1. Teksual Fitur dan pemodelan topik cukup berhasil digunakan untuk membantu proses klasifikasi pemeringkatan jawaban. Dimana Proses klasifikasi dan pemberian peringkat paling bagus adalah menggunakan Logistic Regression yang menghasilkan MAP sebesar 71.9 dan F-measure Sebesar 52.6%.
2. Fitur yang memiliki pengaruh paling besar adalah pemodelan topik, cosine similarity dan tanda tanya, dimana kombinasi ketiga fitur ini menghasilkan map sebesar 72.7%. Sedangkan untuk link dan emotikon kurang berpengaruh karena jumlah jawaban yang mengandung link dan emotikon sedikit yaitu link sebanyak 3% dan emotiokon sebanyak 8%.

Daftar Pustaka

- [1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- [2] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5228–5235, 2004.
- [3] Isabelle Guyon and André Elisseeff. An introduction to feature extraction. In *Feature extraction*, pages 1–25. Springer, 2006.
- [4] Massimo Nicosia, Simone Filice, Alberto Barrón-Cedeno, Iman Saleh, Hamdy Mubarak, Wei Gao, Preslav Nakov, Giovanni Da San Martino, Alessandro Moschitti, Kareem Darwish, et al. Qcri: Answer selection for community question answering experiments for arabic and english. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, volume 15, pages 203–209, 2015.
- [5] Nicola Poletini. The vector space model in information retrieval-term weighting problem. *Entropy*, pages 1–9, 2004.
- [6] Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. The author-topic model for authors and documents. In *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 487–494. AUAI Press, 2004.
- [7] Mark Steyvers and Tom Griffiths. Probabilistic topic models. *latent semantic analysis: A road to meaning*. Lawrence Erlbaum, 2007.
- [8] Quan Hung Tran, Vu Tran, Tu Vu, Minh Nguyen, and Son Bao Pham. Jaist: Combining multiple features for answer selection in community question answering. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval*, volume 15, pages 215–219, 2015.
- [9] Keneilwe Zuva and Tranos Zuva. Evaluation of information retrieval systems. *International Journal of Computer Science & Information Technology (IJCSIT)*, 4(3), 2012.