

Analisis Sentimen Dengan *Query Expansion* Pada Review Aplikasi M-Banking Menggunakan Metode *Fuzzy K-Nearest Neighbor (Fuzzy k-NN)*

Nanda Cahyo Wirawan¹, Indriati², Putra Pandu Adikara²

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹nandacwirawan@gmail.com, ²indriati@mail.ub.ac.id, ³hikaru.yuuki@gmail.com

Abstrak

Di era digital saat ini, bisnis berkembang semakin pesat memanfaatkan aplikasi digital. Dunia perbankan adalah satu lini bisnis yang memanfaatkan dengan baik kemajuan teknologi saat ini. *Mobile banking* adalah salah satu produk digital dalam perbankan yang paling populer karena tidak serumit *SMS banking* maupun *internet banking*. Demi menghadapi bisnis perbankan yang ketat, setiap perusahaan menerapkan *feedback* dari setiap nasabah. Saat ini nasabah dapat menggunakan fitur *review* yang sudah disiapkan. *Review* banyak yang masuk tiap harinya, akan membutuhkan waktu untuk memilah. Sistem dengan *machine learning* diharapkan mampu menghemat waktu untuk memilah data tekstual berpolaritas yang berupa *review* ini. Sistem dalam penelitian ini dibuat menggunakan metode *fuzzy k-nearest neighbor (fuzzy k-NN)*. Metode *fuzzy k-NN* merupakan metode gabungan antara logika *fuzzy* dan algoritme *k-Nearest Neighbor*. Pembobotan yang dilakukan untuk mengolah data tekstual menjadi data numerik yang mampu dikomputasi menggunakan metode TF-IDF dengan *Cosine similarity* untuk menghitung jarak antar data. Berdasarkan hasil pengujian, sistem ini *F-Measure* tertinggi sebesar 0,9604 dan terendah 0,8349.

Kata kunci: *review aplikasi, analisis sentimen, fuzzy k-Nearest Neighbor*

Abstract

In this digital era, bussiness grow significantly by using digital application. Banking is one field of business that utilizes the current technological advances very well. Mobile banking is one of the most popular digital banking products, because it is not as complicated as SMS banking or internet banking. In order to face the strict banking business, every company applying feedback from their customers. Now customers can use the review feature that provided by apps store. There's a lot of reviews that received every day, and it takes some time to knowing what kind of review is that. Systems with machine learning are expected to save time to sort out textual data that containing polarity. The system's machine learning in this study was made using fuzzy k-nearest neighbor (fuzzy k-NN) method. The fuzzy k-NN method is a combined method between fuzzy logic and the k-Nearest Neighbor algorithm. The weighting method for processing textual data into numerical data that can be computed is using TF-IDF method with Cosine similarity to calculate the distance between data. The output of this system is the classified data review. Based on the results of the tests, this system produces the best F-Measure is 0.9273 and the worst is 0.8349.

Keywords: *application review, sentiment analysis, fuzzy k-Nearest Neighbor*

1. PENDAHULUAN

Saat ini persaingan bisnis antar perusahaan sangatlah ketat, terutama di dunia perbankan. Maka dari itulah peningkatan kualitas layanan tentunya telah menjadi barang wajib untuk mempertahankan nasabah. Peningkatan standar kualitas pada layanan sangatlah diperlukan untuk mempertahankan perusahaan supaya bisa bersaing di pasar yang semakin ketat.

Peningkatan layanan ini tentunya akan memberi kontribusi maksimal pada pemakai layanan yang dalam kasus ini adalah nasabah. Peningkatan kualitas layanan ini juga diharapkan mampu menumbuhkan iklim bisnis positif di Indonesia. Di zaman digital ini tiap perusahaan perbankan juga telah memiliki *software* digital *mobile banking (m-banking)* yang banyak digunakan oleh nasabah karena lebih praktis. Sebagai salah satu langkah mewujudkan akuntabilitas, perusahaan perbankan harus mampu

mengumpulkan informasi terkait persepsi mengenai layanan dan apa yang diharapkan oleh nasabah sebagai pemakai layanan.

Saat ini, pengumpulan informasi dilakukan dengan menggunakan digital *review* yang sudah disediakan dalam bentuk *form* digital yang bisa diisi kapan saja dan di mana saja. Dengan adanya sistem ini, nasabah dapat menuangkan keluh kesahnya langsung setelah menggunakan Aplikasi *m-banking* milik suatu bank. Tanggapan terhadap penggunaan mobile banking ini sendiri tiap harinya masuk dalam jumlah yang banyak. Tanggapan tersebut bisa saja berisi saran, laporan adanya masalah, dan berbagai kritik mengenai pelayanan baik dalam tanggapan yang positif maupun tanggapan negatif. Dengan analisis sentimen tentunya akan mempermudah untuk mengetahui sentimen keluhan yang jumlahnya banyak tersebut. Hal tersebut akan berdampak pada kemudahan dan kecepatan dalam melakukan tanggapan pada keluhan untuk meningkatkan kualitas pelayanan.

Pengolahan data teks untuk mengklasifikasikan sentimen yang dimiliki dokumen biasa disebut analisis sentimen. Analisis sentimen sendiri adalah bidang ilmu yang mempelajari bagaimana menganalisis opini, sentimen, evaluasi, penilaian, sikap dan emosi dari sebuah entitas yang dapat berupa produk, pelayanan, organisasi, individu, isu-isu, peristiwa, dan topik. Klasifikasi sentimen dapat dilakukan dengan beberapa metode/ algoritme klasifikasi. Salah satu metode yang paling sering digunakan untuk klasifikasi adalah algoritme *k Nearest Neighbor* (k-NN). Namun algoritme k-NN sering terfokus pada *k* data tetangga terdekatnya saja dan data lainnya menjadi terabaikan. *Fuzzy k-Nearest Neighbor* merupakan metode baru sebagai solusi dari masalah k-NN, yaitu dengan menggabungkan k-NN dengan logika *fuzzy*. Algoritma dasar dari metode ini adalah menetapkan nilai keanggotaan sebagai jarak fungsi vektor dari k-NN dan keanggotaan dari semua data latih. Jadi data uji akan memiliki nilai keanggotaan pada semua kelas yang ada. Kelas dengan nilai keanggotaan tertinggi akan menjadi kelas data uji tersebut (Zhang, 2009).

Dalam penelitian sebelumnya dari (Zhang, 2009) menunjukkan bahwa *fuzzy k-NN* lebih baik dalam akurasi daripada k-NN dan *Support Vector*. Penelitian (El-Muadib, 2010) juga menunjukkan hasil *fuzzy k-NN* lebih baik daripada k-NN biasa. Penelitian dari (Firmansyah, 2016) menunjukkan penambahan

konversi negasi pada frasa dan penambahan *query* pada data uji menunjukkan hasil yang lebih baik. Pada penelitian ini akan menggunakan algoritma *fuzzy k-nearest neighbor* (*fuzzy k-NN*) tersebut ditambah *query expansion* untuk digunakan dalam Analisis Sentimen *Review m-banking*.

2. DASAR TEORI

2.1 Mobile Banking (m-banking)

M-banking atau *Mobile Banking* merupakan layanan perbankan yang bisa diakses langsung melalui perangkat selular/mobile *Global for Mobile Communication (GSM)* dengan memanfaatkan teknologi *Short Message Service (SMS)*. Layanan yang bisa digunakan dengan *m-banking* yakni transfer dana, informasi saldo, mutasi rekening, informasi nilai tukar, pembayaran (kartu kredit, listrik, telepon, *handphone*, asuransi) dan pembelian (pulsa isi ulang, saham) (Indriani, 2013).

Data *review* pada penelitian ini diambil dari *review BCA mobile* yang terdapat pada Google Playstore. Contoh dari *review* dan kelas sentimennya dapat dilihat pada Tabel 1.

Tabel 1. Contoh data *review* dan kelas sentimennya

Review	sentimen
Menyusahkan Udah 20ribu pulsa terpakai tapi verifikasi gak selesai2.	negatif
Errorrrrrrrtt Error terus gk bisa download	negatif
Simple. Tinggal touch semua beres	positif
Top! Bagus banget, sangat membantu transaksi online. Ga perlu repot ke atm/pake key bca.	positif

2.2 Text Mining

Text mining merupakan penambangan data yang berbentuk teks, yang mana sumber datanya berasal dari sebuah dokumen dan memiliki tujuan untuk mencari kata-kata unik yang dapat mewakili isi dari dokumen agar dapat dijadikan patokan untuk analisis hubungan antar dokumen.

Tahapan proses pada *text mining* dibagi menjadi 4 tahapan utama, yakni pemrosesan awal dan pembersihan terhadap teks (*text processing and cleaning*), pemrosesan dan pengaturan teks (*processing and controlling*), analisis pola (*pattern analysis*) dan pemrosesan informasi berharga (*valuable information processing*). *Input* dari proses adalah suatu dokumen teks dan dihasilkan *output* berupa pola sebagai hasil interpretasinya (Talib, 2016).

2.2.1 Pra-Pemrosesan

Pada tahapan pra-pemrosesan teks, dokumen yang berisi tanggapan disiapkan untuk diolah dalam beberapa tahapan yang pada umumnya tahapannya adalah *Tokenizing*, *Filtering* dan *Stemming*. Pada penelitian ini juga akan ditambahkan tahapan normalisasi dan konversi negasi. Normalisasi adalah tahapan pengubahan term menjadi kata baku sedangkan konversi negasi adalah tahap pengubahan frasa berpolaritas menjadi nama sentimennya.

2.2.2 Pembobotan (Weighting)

Pembobotan sendiri adalah proses pengubahan *term* menjadi data numerik yang bisa diproses oleh komputer. Proses ini dilakukan setelah dokumen telah melewati pra-pemrosesan. Metode yang paling umum untuk pembobotan teks adalah *TF-IDF*. Metode ini banyak digunakan dalam *text mining* (Feldman, 2007).

TF adalah bobot dari sebuah *term* yang diambil dari jumlah kemunculan *term* tersebut di sebuah dokumen. Semakin besar muncul di sebuah dokumen akan membuat bobot *TF* sebuah *term* akan menjadi semakin besar, sedangkan *IDF* menghitung seberapa sering sebuah *term* muncul di semua dokumen. Semakin sering munculnya *term* di banyak dokumen akan membuat *term* tersebut dianggap kata umum dan tidak dapat digunakan untuk mewakili sebuah dokumen. Sebaliknya, *term* yang sedikit muncul dari semua dokumen dianggap unik dan bisa mewakili sebuah dokumen.

Berikut ini adalah persamaan untuk menghitung bobot TF.

$$wTF(ti) \begin{cases} 0, TF = 0 \\ 1 + \log(TF(ti)), TF(ti) > 0 \end{cases} \quad (1)$$

dengan *TF(ti)* adalah jumlah kemunculan suatu *term* di sebuah dokumen dan *wTF(ti)* adalah bobot dari sebuah *term* di sebuah dokumen. Lalu untuk menghitung *IDF* menggunakan persamaan berikut.

$$IDF(ti) = \log\left(\frac{|D|}{DF(ti)}\right) \quad (2)$$

dengan *IDF(ti)* adalah *Inverse document frequency* dari sebuah *term*, *D* adalah jumlah dokumen dan *DF(ti)* adalah jumlah dokumen yang memiliki *term* *ti*. Bobot *TF-IDF* sendiri merupakan hasil perkalian dari bobot *TF* dan bobot *IDF*.

2.2.3 Query Expansion

Query expansion merupakan salah satu teknik untuk menambah *query* tambahan pada *information retrieval* di dalam teknik *relevance feedback*. Teknik ini akan memodifikasi *query* untuk dapat memenuhi kebutuhan informasi yang dibutuhkan. Teknik ini terdapat tiga jenis yakni *automatic query expansion*, *manual query expansion* dan *interactive query expansion*.

2.3 Analisis Sentimen

Informasi tekstual dapat dibagi menjadi dua secara umum yakni fakta dan opini. Fakta merupakan ekspresi objektif terhadap suatu objek, sedangkan opini biasanya subjektif yang menggambarkan sentimen, penilaian atau perasaan seseorang pada sebuah objek. Penelitian yang dilakukan dalam *text mining* lebih banyak terfokus pada informasi faktual seperti *information retrieval*, sistem temu kembali, pengelompokan teks dan penitilian dalam *natural language processing* lainnya (Liu, 2010). Analisis sentimen sendiri merupakan pengelompokan teks yang berupa informasi tekstual yang bersifat opini. Informasi tekstual yang dikelompokkan akan mengandung polaritas.

2.4 Algoritme Fuzzy k-Nearest Neighbor (Fuzzy k-NN)

Fuzzy k-Nearest Neighbor merupakan metode klasifikasi yang menggabungkan logika *fuzzy* dan *k-Nearest Neighbor Classifier* karena algoritme *k-NN* masih dianggap memiliki hasil yang kurang bagus. Algoritme *Fuzzy k-NN* menempatkan *class membership* ke dalam suatu pola daripada menempatkan pola ke dalam kelas khusus. Dasar dari algoritme ini adalah menjadikan nilai keanggotaan sebagai bentuk pola selisih dari *k-NN* dan nilai keanggotaan tetangga dalam kelas yang tepat.

Algoritme dari *fuzzy k-NN* adalah sebagai berikut:

1. *input* sekumpulan data latih yang sudah dilabeli kelas dan data uji
2. Set nilai *k*
3. *Preprocessing* data
4. Pembobotan data
5. Hitung Jarak Tiap Data Latih dengan Semua tetangganya
6. Ambil tetangga data latih sesuai nilai *k*
7. Hitung nilai keanggotaan setiap data latih pada semua kelas
8. Hitung jarak data uji dengan semua data

latih

9. Ambil k tetangga terdekat dengan data uji
10. Hitung Nilai Keanggotaan data uji pada semua kelas
11. Cari Nilai Keanggotaan tertinggi

Penentuan kelas nilai keanggotaan kelas ke- i pada tetangga ke- j dihitung menggunakan persamaan berikut:

$$\mu_{ij} = 0,51 + \left(\frac{n_j}{k}\right) * 0,49, j = i \quad (3)$$

$$\mu_{ij} = \left(\frac{n_j}{k}\right) * 0,49, j \neq i \quad (4)$$

dengan μ_{ij} adalah nilai keanggotaan tetangga, n_j adalah jumlah tetangga pada kelas i pada himpunan k -NN, i adalah kelas, j adalah indeks dokumen latih ke- j , dan k adalah nilai subjektif tetangga. Penentuan nilai *membership* kelas pada data uji dihitung menggunakan persamaan berikut:

$$\mu_i(x) = \frac{\sum_{j=1}^k \mu_{ij} \left(\frac{1}{\|x-x_j\|^{2/(m-1)}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|x-x_j\|^{2/(m-1)}} \right)} \quad (5)$$

dengan $\mu_i(x)$ adalah nilai keanggotaan kelas pada data uji x , $\|x-x_j\|$ adalah jarak antara dokumen x dan dokumen x_j , m adalah bobot pangkat dan, k adalah jumlah tetangga terdekat.

3. METODOLOGI

Sistem yang dikembangkan dalam penelitian ini adalah sistem analisis sentimen pada *review m-banking* menggunakan algoritme *fuzzy k-nearest neighbor*. Metode *fuzzy k-nearest neighbor* bergantung pada k data dengan kemiripan tertinggi terhadap dokumen yang dicari kelasnya, oleh karenanya proses prapemrosesan dan data latih yang digunakan akan sangat berpengaruh. Pengoptimalan dilakukan dengan penambahan tahap normalisasi di prapemrosesan untuk mengatasi bentuk data yang berupa singkatan-singkatan, tahap konversi negasi untuk menangani *term* berupa frasa negasi, serta *query expansion* untuk mengatasi sedikitnya data *term* per-dokumen karena dokumen berupa *short text*. *Pseudocode* di bawah ini merupakan perancangan alur proses bagaimana sistem melakukan analisis sentimen.

Algorithm : FKNN_Classification(w, x)

Input: a set of train document vectors $D = \{d_1, d_2, d_3, \dots, d_{|D|}\}$ and their associated class labels

$C = \{c_1, c_2, c_3, \dots, c_{|C|}\}$, a test document d ;

Output: Deciding d in the class;

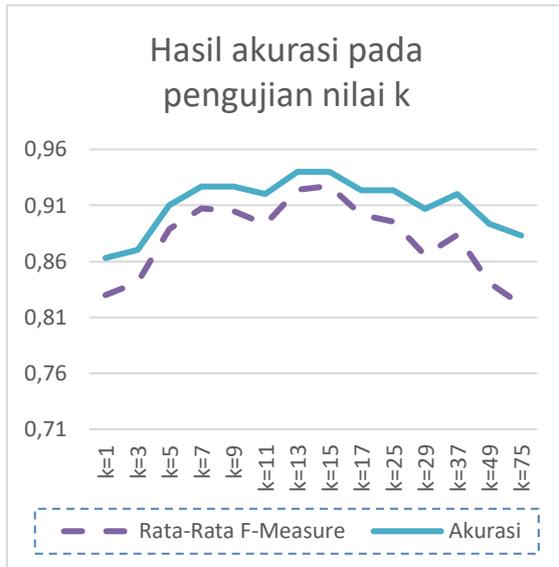
1. Set x and $k, 1 \leq k \leq |D|$;
2. Preprocessing D ;
3. Computing the vocabulary set $\{t_1, t_2, t_3, \dots, t_{|t|}\}$, which is containing n distinct words occurred in training documents.
4. Set x_j be a feature vector of any a web document containing train documents, x_j can be represented as $\{N^1, N^2, \dots, N^n\}$;
5. Set x is feature vector of test document d , which is expressed by $\{N_1, N_2, \dots, N_n\}$;
6. Find k -nearest neighbors of x according to distance from x to feature vector of train web document, set x_1, x_2, \dots, x_k are k -nearest neighbors of x ;
7. Initialize $i = 1$;
8. Do until (x assigned membership in all classes)
 - Compute $u(x)$ using formula (3)(4)(5);
 - $i \leftarrow i + 1$;
 End do
9. Classify x in the class with maximum $\mu(x)$

4. PENGUJIAN DAN ANALISIS

Dalam penelitian ini dilakukan beberapa pengujian yang dilakukan untuk mengetahui akurasi sistem. Pengujian dilakukan dalam beberapa skenario yakni pengujian nilai k , pengujian rasio data, dan pengujian penggunaan *query expansion*.

4.1 Pengujian Nilai k

Pengujian ini dilakukan untuk mengetahui nilai k terbaik dari variasi nilai k yang diuji coba. Pengujian dilakukan pada data uji dan data latih yang telah di lakukan prapemrosesan. Pada data latih dan data uji akan diberlakukan tokenisasi, *filtering*, *stemming* dan konversi negasi. Data dalam pengujian ini akan dibagi menjadi tiga varian. Varian data uji yang digunakan pada pengujian ini berjumlah semuanya 100 yang terdiri dari campuran data berkelas positif dan negatif. Sedangkan varian data latih yang digunakan berjumlah 250 data yang terdiri 126 data negatif dan 124 data positif pada data varian pertama, 146 data negatif dan 104 data positif pada varian kedua serta 147 data negatif dan 103 data positif pada data varian ketiga. Nilai m yang digunakan pada pengujian ini adalah 1. Hasil pengujian dapat dilihat pada Gambar 3.

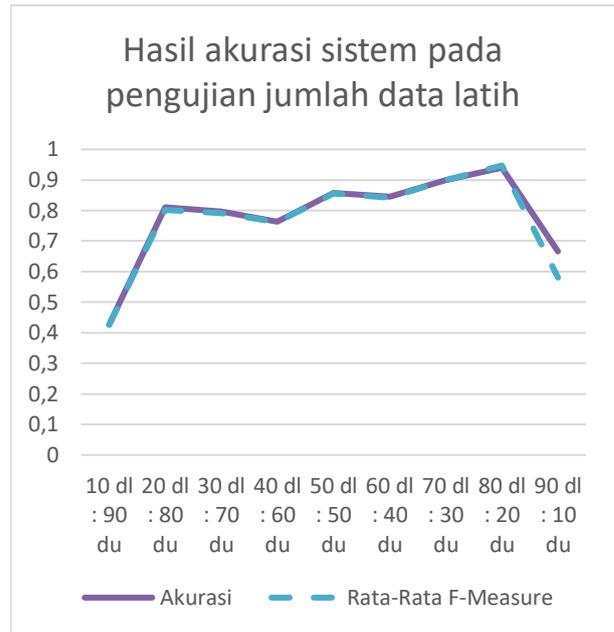


Gambar 1. Hasil rata-rata *F-Measure* dan Akurasi pada pengujian nilai *k*

Dari Gambar 1, dapat dilihat rata-rata *F-Measure* sendiri masih cenderung stabil namun mulai mengalami penurunan seiring bertambahnya tetangga terdekat yang diikutkan dalam komputasi. Tentu saja seperti algoritme *k-NN* lainnya, semakin banyak *k* atau terlalu sedikitnya *k* akan memberikan hasil yang kurang bagus. Semakin banyak nilai *k* maka semakin banyak juga *noise* pada tetangga yang membuat akurasi sistem semakin menurun. Hasil akurasi sistem ini memiliki hasil tertinggi pada nilai *k=13* dan *k=15* yakni sekitar 94% dari rata-rata 3 kali percobaan dan mulai menurun seiring bertambahnya nilai *k* mulai mengalami penurunan. *F-Measure* terbaik yang dihasilkan sistem terjadi saat nilai *k=15* yang mencapai 0,9273 untuk rata-rata *F-Measure* dari semua kelas berbeda sangat tipis dengan *k=13* yang bernilai 0,9239 walau hasil akurasinya sama.

4.2 Pengujian Rasio Data

Pengujian ini dilakukan untuk mengetahui pengaruh jumlah data latih dan data uji dan kemampuan sistem dalam menangani rasio data latih dan data uji. Pengujian dilakukan pada data uji dan data latih yang telah mengalami pra-pemrosesan dan dilakukan menggunakan nilai *k* terbaik dari pengujian sebelumnya yakni 15. Nilai *m* yang digunakan pada pengujian ini adalah 1. Data latih dan data uji yang digunakan akan dilakukan variasi. Hasil pengujiannya dapat dilihat pada Gambar 2.



Gambar 2. Hasil rata-rata *F-Measure* dan Akurasi pada pengujian rasio data

Keterangan Gambar 2.

DU : Data Uji

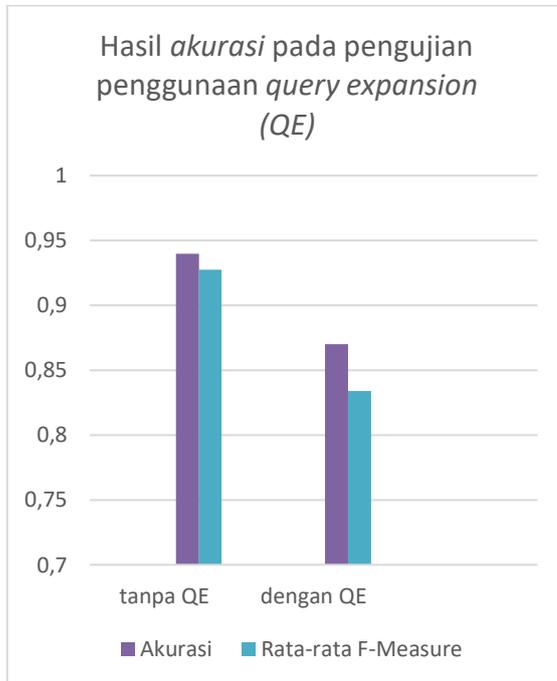
DL : Data Latih

Dari hasil pengujian rasio data latih dan uji ini, *F-Measure* sendiri masih mengalami kenaikan seiring bertambahnya data latih dan mulai bagus saat perbandingan data berjumlah 50 banding 50. Melihat hasil tersebut menunjukkan sistem setidaknya membutuhkan data latih yang berjumlah sama dengan banyak dari data yang akan dianalisis sentimennya. *F-Measure* terbaik yang dihasilkan sistem terjadi saat data latih dan data uji memiliki perbandingan 80 banding 20 dengan nilai 0,9474 untuk kelas positif dan kelas negatif. Hasil akurasi sistem ini juga sama dengan hasil *F-Measure*, yakni tertinggi terjadi saat data latih dan data uji memiliki perbandingan 80 banding 20. Akurasi pada rasio tersebut adalah 94% namun turun drastis kembali saat rasio 90 banding 10.

4.3 Pengujian Penggunaan *Query Expansion*

Pengujian ini dilakukan untuk mengetahui pengaruh penggunaan *query expansion* pada klasifikasi dengan metode *fuzzy k-NN*. Pengujian dilakukan pada data uji dan data latih yang mana akan dilakukan variasi pada data uji yakni menggunakan *query expansion* dan tidak. Nilai *k* yang digunakan adalah nilai *k* terbaik dari pengujian sebelumnya yakni 15. Nilai *m* yang digunakan pada pengujian ini adalah 1. Data uji yang digunakan pada pengujian ini berjumlah 100 yang terdiri dari komposisi data bersentimen

positif maupun negatif. Data latih yang digunakan berjumlah 250 data dari 3 varian data seperti pada pengujian nilai k . Hasil pengujiannya dapat dilihat pada Gambar 3.

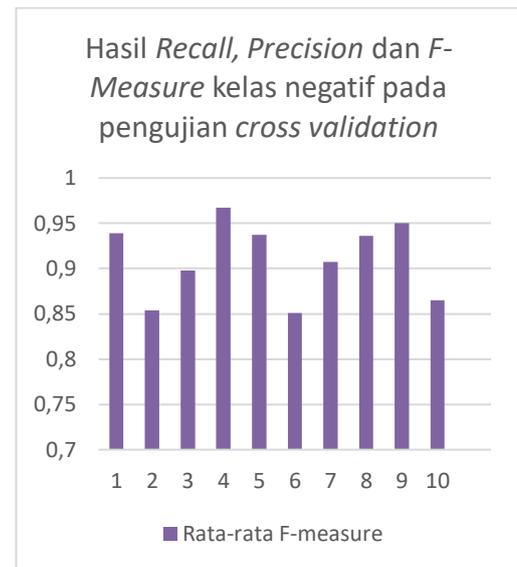


Gambar 3. Hasil rata-rata F -Measure pada pengujian penggunaan query expansion

Dari hasil pengujian variasi tahapan pra-pemrosesan ini, F -Measure mendapat hasil terbaik saat tidak menggunakan *query expansion*. Hal ini menunjukkan walau data berbentuk *short text*, penggunaan metode klasifikasi menggunakan *query expansion* malah akan menambah ambiguitas pada saat klasifikasi. Penggunaan metode klasifikasi *fuzzy k*-NN sudah cukup untuk mengatasi masalah kurangnya *term* asalkan data latih yang digunakan dalam rasio yang sesuai seperti pada pengujian sebelumnya. Penggunaan *query expansion* akan menambah *term* yang justru membuat ambiguitas pada tetangga semakin besar. Hasil F -Measure yang dihasilkan tanpa penggunaan *query expansion* adalah 0,9273 sedangkan dengan penggunaan *query expansion* turun menjadi 0,8339.

Hasil akurasi sistem ini juga sama dengan hasil F -Measure, yakni lebih tinggi saat tanpa penggunaan *query expansion* yakni rata-rata sebesar 94% sedangkan saat memakai hanya mendapat akurasi sebesar rata-rata 87%.

4.4 Pengujian Cross Validation



Gambar 4. Hasil rata-rata F -Measure pada pengujian *cross validation*

Pengujian ini dilakukan untuk mengetahui akurasi yang dihasilkan sistem dengan parameter terbaik yang didapatkan pada pengujian sebelumnya. Pengujian dilakukan pada data uji dan data latih yang terdapat 10 variasi data. Nilai k yang digunakan adalah nilai k terbaik dari pengujian sebelumnya yakni 15. Nilai m yang digunakan pada pengujian ini adalah 2. Data uji yang digunakan pada pengujian ini berjumlah 80 yang terdiri dari komposisi data bersentimen positif maupun negatif. Data latih yang digunakan berjumlah 320, jadi data latih dan data uji memiliki perbandingan 80 banding 20 seperti pada hasil terbaik pengujian rasio. Data uji pada pengujian ini tidak menggunakan *query expansion*. Hasil pengujiannya dapat dilihat pada Gambar 4.

Dari hasil pengujian *cross validation* ini, F -Measure mendapat hasil terbaik pada data pertama baik pada kelas positif dan kelas negatif. Hasil f -measure tertinggi yang dihasilkan oleh sistem adalah 0,9604 pada data varian pertama dengan perbedaan tipis dengan data varian keempat yang memiliki f -measure sebesar 0,9601. F -measure terendah yang dihasilkan sistem adalah 0,8349 yakni pada varian data 9.

5. KESIMPULAN

Metode *fuzzy k*-nearest neighbor (*fuzzy k*-NN) dapat diimplementasikan pada analisis sentimen pada *review* aplikasi *m-banking* dengan ataupun tanpa *query expansion*.

Algoritme *fuzzy k-NN* untuk analisis sentimen pada review aplikasi *m-banking* menghasilkan akurasi tertinggi sebesar 94% dengan *F-Measure* 0,9273 yang merupakan rata-rata dari 3 percobaan yang mana ketiganya menggunakan komposisi data yang berbeda. Hasil akurasi tertinggi didapatkan pada rasio data latih 80% banding data uji 20%, hasil yang didapat 94% dengan *F-Measure* 0,9474.

Nilai *k* sangat berpengaruh pada *fuzzy k-NN*, semakin tinggi nilai *k* maka akurasi yang dihasilkan sistem akan semakin turun. Penggunaan *query expansion* pada penelitian ini malah mengurangi hasil akurasi, hal ini disebabkan karena data semakin ambigu. Contoh adalah term “diperbaiki” yang akan menjadi “baik” setelah proses *stemming*, dan ketika di ekspansi akan terdapat term “mantap” yang makin menambah *cosine similarity* terhadap data latih yang berpolaritas positif. *Query expansion* pada penelitian ini memiliki kontra dengan tahapan *stemming*.

Lebih baik pada penelitian berikutnya ditambahkan pengujian tahapan pra-pemrosesan agar tidak terjadi lagi tahapan yang kontra seperti pada penelitian ini. Tahap normalisasi lebih baik menggunakan metode khusus yang memang digunakan untuk normalisasi dikarenakan menggunakan kamus seperti dalam penelitian ini akan sangat memakan waktu.

6. DAFTAR PUSTAKA

- El-Mouadib, Faraj A., & Amal F. Abdalsalam. 2010. *Comparison Of Crisp And Fuzzy Knn Classification Algorithms*. University of Garyounis, Libya.
- Feldman, R. & Sanger, J., 2007. *The Text Mining Handbook (First Published Edition)*. Cambirdge University, New York.
- Firmansyah, Ro'i Fahreza Nur. 2016. *Sentiment Analysis Pada Review Aplikasi Mobile Menggunakan Metode Naïve Bayes Dan Query Expansion*. Fakultas Ilmu Komputer, Universitas Brawijaya.
- Indriani, Devi., 2013. *Pengantar Teknologi Mobile Banking*. Universitas Gunadarma, Jakarta.
- Liu, N. 2010. *Sentiment Analysis and Subjectivity (Second Edition)*. University of Illinois, Chicago.
- Talib, Ramzan, Muhammad Kashif Hanif, Shaeela Ayesha & Fakeeha Fatima. 2016. *Text Mining: Techniques, Applications and Issues*. International Journal of Advanced