

PENERAPAN NAIVE BAYESIAN UNTUK PERANKINGAN KEGIATAN DI FAKULTAS TIK UNIVERSITAS SEMARANG

Astrid Novita Putri
Fakultas TIK
Universitas Semarang
Email: astrid@usm.ac.id

ABSTRAK

Universitas Semarang adalah Salah satu perguruan tinggi swasta di Semarang, dengan jumlah mahasiswa aktif 12.000. Universitas Semarang mempunyai berbagai macam fakultas dan juga program studi di dalamnya. Di setiap program studi setiap tahun pasti selalu mengadakan kegiatan, entah kegiatan kemahasiswaan maupun kegiatan akademik. Salah satu kegiatan yang selalu di adakan setiap tahun adalah kegiatan *Workshop*, Kuliah Umum, *Festifal*, dsb. Dengan begitu banyaknya kegiatan yang sudah terlewati terkadang fakultas belum memberikan urutan dari mana sajakah program studi yang sering mengadakan kegiatan dan memiliki banyak peminat di dalam kegiatan tersebut. Oleh karena itu di perlukan suatu perankingan berdasarkan favorit dan tidak pada program studi manakah yang memiliki peminat sangat banyak, sehingga memacu program studi yang lain untuk giat membuat kegiatan yang menarik mahasiswa dan akan menambah pengetahuan baru bagi mahasiswa. Di Universitas Semarang memiliki beberapa fakultas dan program studi salah satunya adalah Fakultas Teknologi Informasi dan Komunikasi. Kemudian perankingan akan di buat menggunakan salah satu metode klasifikasi yaitu metode *naive bayesian*.

Kata Kunci: perankingan, naive, bayesian, kegiatan.

ABSTRACT

Semarang University is one of the private universities in Semarang, with an active student population of 12,000. Semarang University has a variety of faculties and courses in it. In each course of study every year must always hold activities, whether student activities or academic activities. One of the activities that is always held every year is the workshop, Public Lecture, Festifal, etc .. With so many activities that have been missed sometimes the faculty has not given the order from which study programs are often held activities and have many enthusiasts in the activity. Therefore, in need of a ranking based on favorites and not on which study program has a lot of enthusiasts, thus spurring other study programs to actively create activities that attract students and will add new knowledge for students. At the University of Semarang has several faculties and programs one of which is the Faculty of Information and Communication Technology. Then the ranking will be made using one method of classification is naive bayesian method.

Keywords: ranking, naive, bayesian, activity.

1. PENDAHULUAN

Kegiatan adalah suatu aktivitas peristiwa atau kejadian pada umumnya yang di lakukan secara terus, pada universitas semarang memiliki beberapa fakultas dan program studi, setiap fakultas memiliki beberapa program studi, salah satunya adalah Fakultas Teknologi informasi dan komunikasi memiliki 3 program studi yaitu Teknik Informatika, Sistem Informasi dan Ilmu Komunikasi, di setiap program studi setiap tahun mengadakan suatu kegiatan, contohnya : Kuliah Umum, Seminar, *Workshop*, *Festifal*, dsb. Pada setiap kegiatan memiliki peserta baik dalam maupun dari luar Universitas Semarang, mahasiswa juga di wajibkan mengikuti kegiatan kegiatan yang ada di kampus, namun terkadang kegiatan pihak fakultas tidak memberikan klasifikasi pembeda manakah program studi yang memiliki banyak kegiatan dan memiliki jumlah peserta terbanyak. Oleh sebab itu di perlukan suatu klasifikasi pembeda, agar memacu program studi membuat kegiatan yang terbaru dan menambah pengetahuan untuk mahasiswa. Untuk mengklasifikasikan tersebut menggunakan metode pada *artificial intelligent* yaitu *naive bayesian*.

Metode *naive bayesian* adalah salah satu metode klasifikasi dan percabangan dari *artificial intelligence*, kegiatan-kegiatan tersebut akan di bentuk suatu Ranking yaitu Favorit dan Tidak Favorit, sehingga mahasiswa dan masyarakat dapat mengetahui kegiatan apa sajakah yang ada di fakultas TIK,

dengan kriteria di bawah ini program studi, jenis kegiatan, hasil kegiatan, dan ranking. Dengan di latar belakang tersebut peneliti akan membuat perhitungan klasifikasi program studi mana sajakah yang favorit dan tidak favorit.

1.1. Data Mining

Data mining yang di kenal dengan nama *pattern recognition* adalah metode pengolahan untuk menemukan suatu pola yang tersembunyi untuk dapat di olah menjadi pengetahuan dan ilmu pengetahuan baru dan informasi dari data dan hasil untuk keputusan di masa depan.

Data mining juga dapat di sebut sebagai sistem pengolahan data yang sangat besar, yang memberikan peranan dalam beberapa bidang di dunia yaitu bidang keuangan, industri, transportasi, cuaca, dan teknologi. Dalam data mining juga terdapat metode metode yang daat di gunakan anatra lain metode klasifikasi, *clustering*, *regresi*, seleksi *variabel*, dan *market bisnis*. Data mining dapat di artikan sebagai data dalam jumlah besar yang di simpan dalam suatu database, *data warehouse* untuk menyimpan dapat sehingga dapat menemukan pola. Ada beberapa teknik dalam data mining antara lain data analisis, *signal processing*, *neural network* dan pengenalan pola.

1.2. Metode Klasifikasi Naive Bayesian

Naive Bayes adalah suatu metode klasifikasi dalam data mining dengan menggunakan metode probabilitas dan statistik sesuai dengan di kemukakan oleh ilmuwan inggris bernama Thomas Bayes. Kemudian Menurut Olson: 2008 : p 102 menjelaskan bahwa *Naive Bayes* merupakan suatu kelas keputusan, dengan menggunakan perhitungan probabilitas matematika dengan syarat bahwa nilai keputusan adalah benar, berdasarkan informasi obyek.

Sedangkan *The Naive Bayesian classifier*, atau *Simple Bayesian Classifier* menurut Han, Kamber 2011 : p3 : 51 yaitu :

- Variabel D merupakan suatu set tuple dan label yang berkait dengan kelas, kemudian setiap *tuple* mewakili vektor atribut n dan dimensi $X = (x_1, x_2, \dots, x_n)$, yang menggambarkan pengukuran n pada *tuple* dari atribut n, masing masing adalah A_1, A_2, \dots, A_n .
- Kelas M, C_1, C_2, \dots, C_m . Di berikan suatu *tuple* X, untuk klasifikasi prediksi X yang akan menjadi kelompok memiliki probabilitas posterior tertinggi, kondisi tersebut di sebutkan pada X. Artinya adalah sebagai berikut klasifikasi *Naive Bayes* memprediksi bahwa X tuple memiliki kelas C_i . Rumus dapat di sebutkan pada gambar di bawah ini.

$$P(C_i|X) > P(C_j|X) \text{ For } 1 \leq j \leq m, j \neq i$$

Rumus Classifier Naive Bayesian

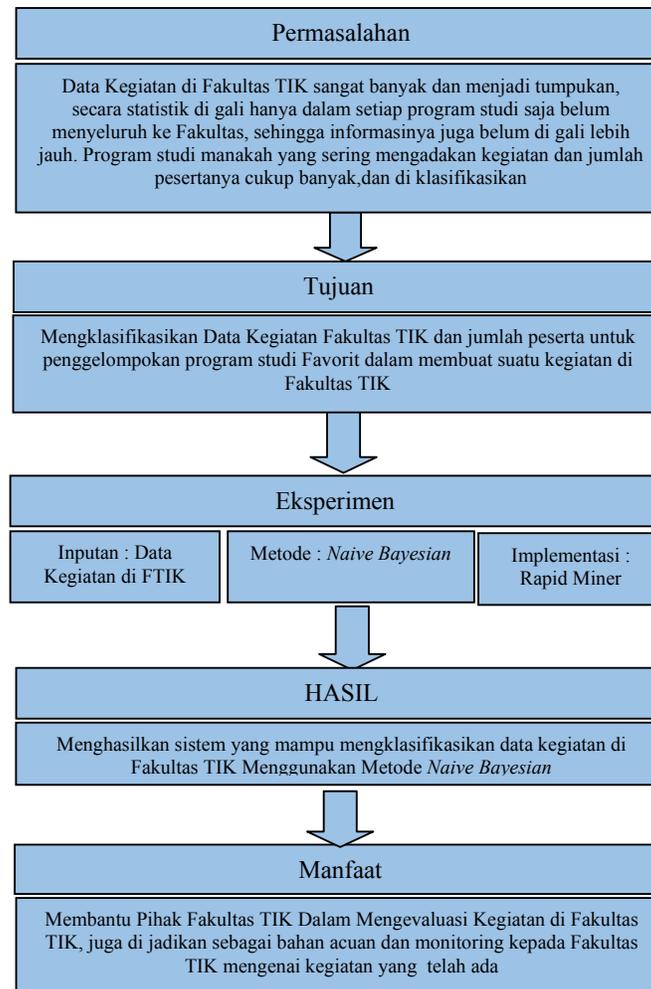
Sumber: Han, Kamber :2011: p351

Keterangan :

- $P(C_i|X)$ = Probabilitas hipotesis C_i jika diberikan fakta atau record X (*Posterior probability*)
- $P(X|C_i)$ = mencari nilai parameter yang memberi kemungkinan yang paling besar (*likelihood*)
- $P(C_i)$ = Prior probability dari X (*Prior probability*)
- $P(X)$ = Jumlah probability tuple yg muncul

1.3. Kerangka Pemikiran

Kerangka pemikiran merupakan garis besar dari langkah – langkah penelitian yang sedang dilakukan, kerangka pemikiran dijadikan acuan untuk melakukan tahap – tahap yang sedang dilakukan dalam penelitian.



Gambar 1. *Flow Chart* Kerangka Pemikiran

2. METODOLOGI PENELITIAN

Pada bagian ini akan membahas mengenai jenis data dan metode pengumpulan data yang di gunakan.

2.1 *Jenis Data*

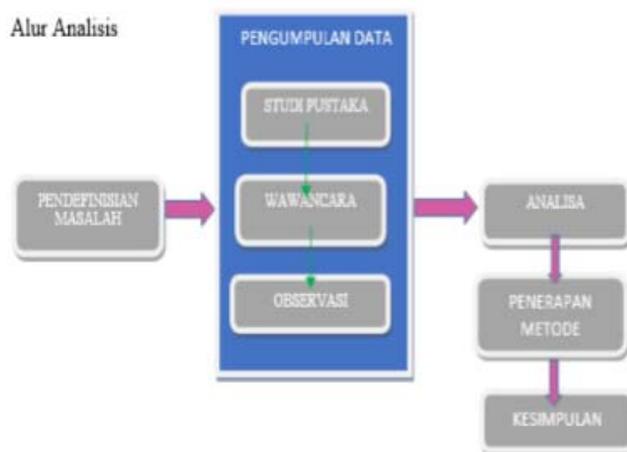
Jenis data yang digunakan dalam penelitian ini yaitu:

- Data primer, yaitu data yang diperoleh secara langsung dari objek penelitian, sebagai berikut : Kegiatan yang ada di fakultas TIK, Jumlah peserta, kemudian ranking yang akan di dapatkan.
- Data sekunder, yaitu data yang diperoleh dari literature, buku referensi, maupun browsing internet.

2.2 *Metode Pengumpulan Data*

Metode pengumpulan data yang dilakukan dalam penelitian ini adalah:

- Observasi dan Wawancara : Pengumpulan data melalui pengamatan dan wawancara secara langsung dengan Ketua Program Studi Teknik Informatika, Sistem Informasi, dan Ilmu Komunikasi mengenai kegiatan apa sajakah yang ada di dalamnya dan menganalisa kriteria yang di gunakan dari jenis kegiatan, tanggal kegiatan, hasil kegiatan, dan ranking yang di dapat dari hasil klasifikasi jumlah peserta.
- Studi Pustaka : Mengumpulkan literatur pendukung penelitian, baik dari buku referensi ataupun browsing dari internet.



Gambar 2. Alur Analisis

Pada penelitian ini diawali dengan pendefinisian masalah hingga metode yang akan digunakan. Di bawah ini adalah tahapan sebagai berikut :

- a) Pendefinisian permasalahan berkaitan dengan merancang suatu analisa.
- b) Studi Pustaka : Pengumpulan data yaitu berupa literatur buku, jurnal nasional maupun internasional, kemudian penelitian yang pernah dilakukan peneliti sebelumnya.
- c) Wawancara : Melakukan proses wawancara tanya jawab dengan pihak ketua program studi Teknik Informatika, Sistem Informasi, dan Ilmu Komunikasi.
- d) Observasi : Pengamatan langsung proses kegiatan yang ada
- e) Analisa : Melakukan analisa pada hasil observasi, wawancara, penelitian sebelumnya.
- f) Penerapan Metode : Metode mana sajakah yang akan digunakan pada klasifikasi program studi mana sajakah yang favorit.
- g) Pengambilan Kesimpulan Oleh Peneliti.

3. HASIL DAN PEMBAHASAN

3.1 Penentuan Kriteria

Berikut ini adalah kriteria dalam menentukan hasil berdasarkan ranking menggunakan metode klasifikasi : *Naive Bayesian* berikut ini :

- a) Progdil : Teknik Informatika, Sistem Informasi, Ilmu Komunikasi.
- b) Jenis Kegiatan : Kuliah Umum, Workshop, Pelatihan, Kegiatan Akademik, Kegiatan Ilmiah.
- c) Hasil Kegiatan (Berdasarkan Peserta) : Memuaskan, Cukup (Di dapatkan dari jumlah peserta kegiatan)
- d) Ranking : Favorit, Tidak Favorit.

3.2 Analisa Perhitungan Metode Naive Bayesian

Berikut ini adalah *tabel training* dari dataset kegiatan di Fakultas TIK Pada Tahun 2016 adalah di bawah ini ada 13 data terdiri dari 4 data kegiatan program studi Teknik Informatika, 3 kegiatan program studi Sistem Informasi dan 6 kegiatan pada program studi Ilmu Komunikasi.

Tabel 1. Training kegiatan fakultas TIK Universitas Semarang

No	Program Studi	Jenis Kegiatan	Hasil Kegiatan	Ranking
1.	Teknik Informatika	Workshop Instal U-Buntu Realase Party 11.04	Memuaskan	Favorit
2.	Teknik Informatika	Workshop Framework PHP	Memuaskan	Favorit
3.	Teknik Informatika	Workshop Linux	Cukup	Tidak Favorit
4.	Teknik Informatika	Workshop Sistem Sensor	Memuaskan	Favorit
5.	Sistem Informasi	Workshop Website	Cukup	Tidak

No	Program Studi	Jenis Kegiatan	Hasil Kegiatan	Ranking
6.	Sistem Informasi	Festifal IT Fest	Memuaskan	Favorit
7.	Sistem Informasi	Workshosp Mikrotik	Cukup	Tidak Favorit
8.	Ilmu Komunikasi	Pelatihan Lab. TV	Cukup	Tidak Favorit
9.	Ilmu Komunikasi	Festival Charity	Cukup	Tidak Favorit
10.	Ilmu Komunikasi	Festifal Komukino	Memuaskan	Favorit
11.	Ilmu Komunikasi	Pelatihan Acara Retorika	Memuaskan	Favorit
12.	Ilmu Komunikasi	Festifal Review TV	Memuaskan	Favorit
13	Ilmu Komunikasi	Festifal Public Relation	Cukup	???

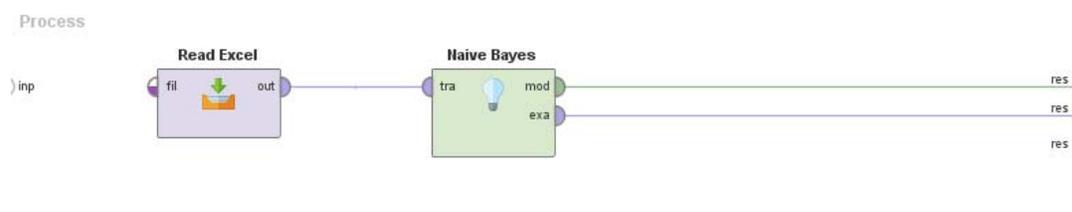
Kemudian Bagaimana Cara Menghitung No 13, Kategori Hasil Apakah Yang Akan di Dapatkan Dalam Hasil Ranking ?, Perhitungan *Naive Bayesian* :

- a) Tahap 1 Menghitung jumlah *class/label* (Ranking)
 - $P(\text{Favorit}) = 7/12$
 - $P(\text{Tidak Favorit}) = 5/12$
- b) Tahap 2 :
 - Menghitung Jumlah Untuk Masalah Sama dengan *Class* Yang Sama (Berdasarkan : Program Studi)
 - $P(\text{IK}\backslash\text{Favorit}) = 3/5$
 - $P(\text{IK}\backslash\text{Tidak Favorit}) = 2/5$
 - Menghitung Jumlah Untuk Masalah Sama dengan *Class* Yang Sama (Berdasarkan : Kegiatan)
 - $P(\text{Festifal}\backslash\text{Favorit}) = 3/4$
 - $P(\text{Festifal}\backslash\text{Tidak Favorit}) = 1/4$
 - Menghitung Jumlah Untuk Masalah Sama dengan *Class* Yang Sama (Berdasarkan : Hasil Kegiatan)
 - $P(\text{Cukup}\backslash\text{Favorit}) = 0/5$
 - $P(\text{Cukup}\backslash\text{Tidak Favorit}) = 5/5$
- c) Tahap 3 Mengkalikan Semua Variabel Favorit dan Tidak Favorit
 - Klasifikasi Bayes Untuk Ranking : Favorit
 - $= P(\text{Favorit}) \cdot P(\text{IK}\backslash\text{Favorit}) \cdot P(\text{Festifal}\backslash\text{Favorit}) \cdot P(\text{Cukup}\backslash\text{Favorit})$
 - $= 7/12 * 3/5 * 3/4 * 0/5$
 - $= 0,58 * 0,6 * 0,75 * 0$
 - $= 0$
 - Klasifikasi Bayes Untuk Ranking : Tidak Favorit
 - $= P(\text{Tidak Favorit}) \cdot P(\text{IK}\backslash(\text{Tidak Favorit})) \cdot P(\text{Festifal}\backslash(\text{Tidak Favorit})) \cdot P(\text{Cukup}\backslash(\text{Tidak Favorit}))$
 - $= 5/12 * 2/15 * 1/4 * 5/5$
 - $= 0,416 * 0,4 * 0,25 * 1$
 - $= 0,0416$

Dari hasil perhitungan di atas, maka klasifikasi Bayes untuk Ranking dari *record* 13 yang di peroleh adalah “TIDAK FAVORIT” dengan nilai 0,0416 dibandingkan dengan kegiatan yang favorit.

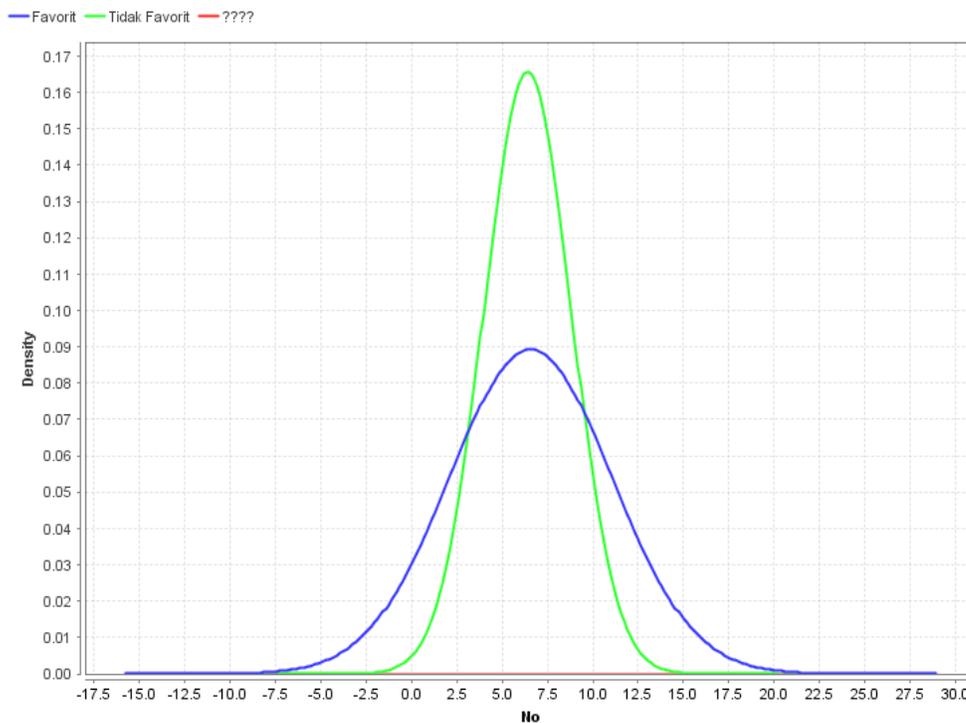
3.3 Implementasi Klasifikasi Naive Bayesian Pada Rapid Miner

Berikut ini adalah pengolahan data dengan menggunakan *naive bayesian* pada *Rapid Miner* :



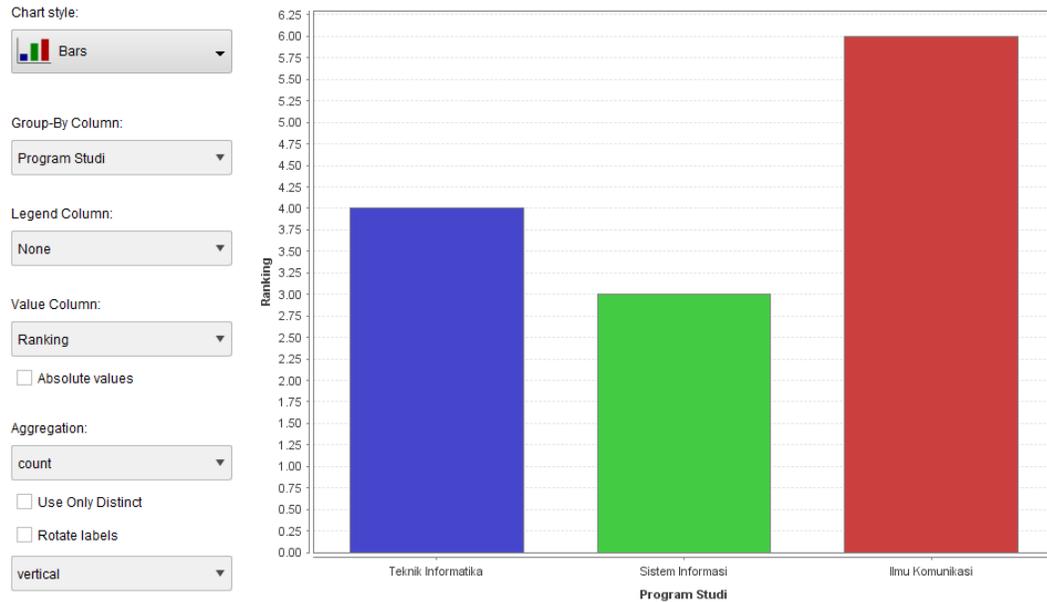
Gambar 3. Pemodelan Klasifikasi *Naive Bayesian* Pada *Rapid Miner*

Dengan menggunakan pemodelan klasifikasi *naive bayesian* seperti gambar di atas dengan mengklasifikasi 2 jenis Favorit dan Tidak Favorit. Kemudian Berikut ini adalah hasil implementasi Kegiatan *Curva* Berdasarkan Favorit dan Tidak Favorit menggunakan *Rapid Miner* menunjukkan kegiatan yang tidak favorit sangat tinggi di Fakultas TIK .



Gambar 4. Nilai Ranking Hasil Kegiatan Menggunakan *Rapid Miner*

Kemudian berdasarkan *chart* di Fakultas TIK pada hasil di bawah ini menunjukkan bahwa Program Studi Ilmu Komunikasi Menempati Urutan Terfavorit dengan Nilai 6, Menyusul urutan kedua yaitu Program Studi Teknik Informatika dengan nilai 4, dan urutan terakhir adalah Program Studi Sistem Informasi dengan nilai 3 pada Hasil *Rapidminer*.



Gambar 5. Chart Berdasarkan Program Studi

4. KESIMPULAN

Dapat di simpulkan bahwa hasil perankingan pada Fakultas TIK Untuk kegiatan kegiatan yang telah ada masih menggunakan cara perhitungan manual dari hasil rata rata, klasifikasi menggunakan rata-rata hasil indikator data teknik statistik dasar, sehingga output yang di dihasilkan kurang maksimal. Tumpukan data kegiatan di Fakultas TIK.

Berdasarkan data tahun 2017 menghasilkan klasifikasi menggunakan metode *naive bayesian*, dan program studi favorit adalah ilmu komunikasi berdasarkan hasil dari *rapiminer*. Untuk saran pada penelitian ini dapat di kembangkan dalam suatu aplikasi menggunakan bahasa pemrograman dan dapat di kembangkan dengan metode alghoritma klasifikasi yang lainnya.

UCAPAN TERIMA KASIH

Peneliti ucapkan terimakasih kepada pihak Dikti dan Universitas Semarang yang telah membiayai penelitian ini.

DAFTAR PUSTAKA

- [1] Zhang, Harry, and Shengli Sheng. "Learning weighted naive Bayes with accurate ranking". ICDM 1-4 Nov 2004 : Fourth IEEE International Conference : 567-570 .
- [2] Wagner, Daniel, et al. "Real-time detection and tracking for augmented reality on mobile phones". (2010). IEEE transactions on visualization and computer graphics : 355-368.
- [3] Coiras, E., et al. "Supervised target detection and classification by training on augmented reality data". 2007 : IET Radar, Sonar & Navigation 1.1. : 83-90.
- [4] Silva, R., et al. "Object recognition using bayesian networks for augmented reality applications".2004 : VII Symposium on Virtual Reality.
- [5] Arsyad, Azhar."Media Pembelajaran. cetakan ke-15". Jakarta:Rajawali Pers".2011
- [6] Andi Prastowo."Metode Penelitian Kualitatif dalam Perspektif Rancangan Penelitian".2012
- [7] J. O. Ong. "Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University". 2013. Jurnal Ilmiah Teknik Industri, Vol. 12, no. 1, pp. : 10 - 20.

- [8] K. R. Prilianti dan H. Wijaya. “*Aplikasi Text Mining Untuk Automasi Penentuan Tren Topik Skripsi Dengan Metode K-Means Clustering*”. 2014. Jurnal Cybermatika, Vol. 2, no. 1, pp. 1-6.