

Klasifikasi Data Kesiapan Pernikahan Menggunakan *Algoritme Naive Bayes*

Ayunda Yusuf¹, Astri Ahfiatul Mahya², Pungkas Subarkah³, Tri Astuti⁴, Dwi Krisbiantoro⁵

^{1,2,3,5}Program Studi Sistem Informasi

⁴Program Studi Teknik Informatika

STMIK Amikom Purwokerto

Email : yunday19@gmail.com, astrimahya7@gmail.com, subarkah18.pungkas@gmail.com, tri_astuti@amikompurwokerto.ac.id
dwikris2013@gmail.com

Abstrak—Pernikahan merupakan hubungan yang bersifat suci/sakral antara seorang pria dan seorang wanita yang telah menginjak umur cukup dewasa dan hubungan tersebut telah diakui secara sah dalam hukum dan secara agama. Namun kenyataannya ditemukan banyak beragam alasan dan latar belakang dalam menentukan kesiapan wanita dan pria dalam melepas lajangnya yakni dengan menikah. Beragam parameter yang menjadi acuan dalam penelitian ini memudahkan seseorang untuk memprediksi kesiapan menikah atau belum entah itu untuk dirinya sendiri atau untuk orang lain. Algoritme klasifikasi data mining *Naive Bayes* dapat digunakan untuk prediksi kesiapan menikah yang nantinya bisa di kategorikan ke dalam kelas tujuan siap atau belum siap menikah, dari hasil prediksi bisa di manfaatkan untuk dasar pengambilan keputusan sehingga dapat mengetahui kesiapan secara mental, finansial maupun restu orangtua. Data diperoleh berdasarkan kuesioner yang disebarkan kepada mahasiswa dan mahasiswi Stmik Amikom Purwokerto.

Kata Kunci : Menikah, *Naive Bayes*, Stmik Amikom Purwokerto.

I. PENDAHULUAN

Pernikahan merupakan sebuah perjanjian atau biasa disebut akad antara seorang laki-laki dan perempuan yang disertai dengan rasa ikhlas dan sukarela antar kedua belah pihak supaya membentuk kehidupan berkeluarga yang berlandaskan agama [1]. Tujuan pernikahan itu sendiri ialah untuk membentuk keluarga yang sakinah, mawadah dan rahmah. Yang apabila diringkas secara pengertian ketiganya ialah sikap antara kedua belah pihak pasangan untuk saling mengayomi, saling berpegang pada hak dan kewajiban masing-masing, saling menghormati dan memuliakan. Fungsi pernikahan itu sendiri ialah mencegah hal-hal yang diinginkan berkaitan dengan aurat, dimana aurat sendiri adalah hal yang memalukan sehingga harus ditutupi oleh pakaian, dan pakaian itu sendiri adalah sebuah pernikahan yang sah secara hukum dan agama[2].

Untuk mencari pasangan hidupnya manusia diharuskan menikah pada masa yang sesuai dengan waktu dan kodratnya, salah satu caranya yakni dengan menarik lawan jenis. Manusia memulai menarik lawan jenisnya dimulai dari usia pubertas hal ini berguna untuk mencari pasangannya yang berawal dari

usia sekitar 12,5 – 14,5 tahun pada perempuan dan 14 – 16,5 tahun pada laki-laki [3]. Menurut teori perkembangan, masa usia menikah adalah saat usia dewasa awal yaitu 20-40 tahun atau usia 18-40 tahun[4]. Hal ini dapat disimpulkan bahwa masa dewasa awal merupakan masa untuk menikah dan membina rumah tangga.

Hal ini sejalan dengan bahwa perkembangan yang menjadi karakteristik masa dewasa awal adalah mulai memilih pasangan hidup dan mulai bekerja[5]. Masa dewasa awal merupakan masa bermasalah karena pada masa ini manusia dituntut untuk hidup mandiri dan sejalan yakni beriringan antara menikah dan berkarir. Artinya, karir dan persiapan menuju kehidupan pernikahan adalah dua tugas penting yang hadir di waktu yang bersamaan[6].

Berdasarkan uraian di atas terlihat bahwa selain menikah dan membina kehidupan berkeluarga, tugas perkembangan lainnya yang dihadapi oleh individu dewasa awal adalah bekerja dan berkarir. Hal ini dapat disimpulkan bahwa menikah tidak hanya sebatas pada aturan agama, kesiapan umur dan mental namun juga pengaruh dari segi finansial dari kedua belah pihak.

Tujuan penulis dalam melakukan penelitian adalah mengetahui mahasiswa/mahasiswi STMIK AMIKOM Purwokerto yang siap menikah sesuai dengan susunan variabel dari berbagai parameter yang menjadi latar belakang dalam suatu pernikahan. Berdasarkan uraian di atas maka dapat dirumuskan permasalahan yang terjadi ialah tentang cara mengetahui kesiapan menikah Mahasiswa STMIK Amikom Purwokerto. Paper ini diuraikan menjadi empat bahasan yang terdiri dari pendahuluan, tinjauan pustaka, metode penelitian, dan yang terakhir hasil dan pembahasan.

II. TINJAUAN PUSTAKA

Adapun tinjauan pustaka yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Algoritma *Naive Bayes*

Algoritma Naive Bayes merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas

dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yakni memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya[7].

2. Weka

Weka adalah sebuah perangkat lunak data mining *open source* berbasis Java. Aplikasi ini pertama kali dikembangkan oleh Universitas Waikato di Selandia Baru sebelum menjadi bagian dari Pentaho. Weka memiliki banyak algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi / formulasi dari sekumpulan data *sampling*. Walaupun kekuatan Weka terletak pada algoritma yang makin lengkap dan canggih, kesuksesan data mining tetap terletak pada faktor pengetahuan manusia implementasinya. Tugas pengumpulan data yang berkualitas tinggi dan pengetahuan pemodelan dan penggunaan algoritma yang tepat diperlukan untuk menjamin keakuratan formulasi yang diharapkan[8].

3. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar[9].

4. Data Training

Data *Traning* atau bisa disebut juga *Training* set digunakan oleh algoritma klasifikasi (misalnya: *decision tree*, *bayesian*, *neural network*, *SVM*) untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada[10].

III. METODE PENELITIAN

Adapun metode penelitian yang diterapkan pada penelitian ini adalah sebagai berikut:

1. Analisis Masalah dan Studi Literatur

Langkah awal untuk menentukan rumusan masalah dari penelitian yang akan dibuat. Dalam hal ini mengamati permasalahan yang berhubungan dengan satu mahasiswa/mahasiswi tersebut, jenis kelamin, manajemen ekonomi, cara dalam menghadapi masalah, kriteria pendamping hidup, proses yang diharapkan, finansial, dan tentunya apakah sudah

siap untuk menikah atau tidak. Permasalahan-permasalahan yang ada, selanjutnya dianalisa untuk mengetahui bagaimana cara penyelesaian terhadap masalah tersebut dan menentukan ruang lingkup permasalahan yang akan diteliti. Mempelajari dasar teori dari berbagai literatur mengenai penerapan metode *Naive Bayes*, konsep dan teori data mining, melalui jurnal-jurnal agar mendapatkan dasar pengetahuan untuk melakukan penelitian selanjutnya.

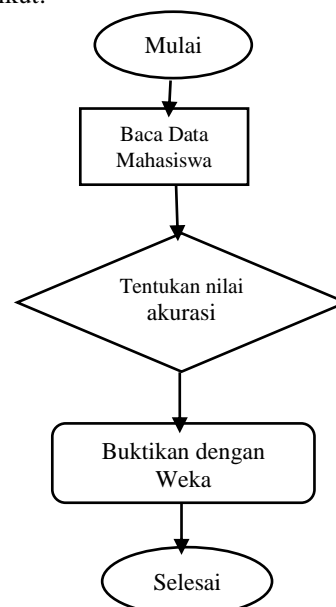
2. Mengumpulkan Data melalui Kuisisioner

Prosedur sistematis yang digunakan untuk mengumpulkan data yaitu dengan mengajukan pertanyaan-pertanyaan kepada responden yaitu mahasiswa dan mahasiswi STMIK AMIKOM Purwokerto melalui kuisisioner. Hasil kuisisioner yang didapatkan akan digunakan untuk bahan penganalisisan data terhadap metode *Naive Bayes*. Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada metode *Naive Bayes*.

3. Implementasi dan Pengujian

Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah *tools*. *Tools* yang akan digunakan dalam implementasi penelitian ini adalah dengan menggunakan Software Weka. Selanjutnya Pengujian dilakukan untuk mengetahui apakah penelitian yang dilakukan telah sesuai dengan tujuan yang diharapkan. *Naive bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya [11].

Alur dari metode *Naive bayes* dapat dilihat pada gambar sebagai berikut:



Gambar 1. Flowchart metode naïve bayes

1. Dimulai dengan membaca data mahasiswa yang sudah di siapkan.
2. Kemudian menentukan nilai akurasiya.
3. Membuktikannya dengan Weka, dan jika nilai sudah sesuai maka selesai.
4. Selesai.

IV. HASIL DAN PEMBAHASAN

Naive bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu. Keuntungan penggunaan Naive bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (Training Data) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Naive bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan. Adapun alur dari metode Naive bayes adalah sebagai berikut:

- i. Data Training
- ii. Hitung Jumlah dan Probabilitas, namun apabila data numerik maka:
 1. Cari nilai mean dan standar deviasi dari masing masing parameter yang merupakan data numerik.
 2. Cari nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
- iii. Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas.

Persamaan dari teorema Bayes adalah[12]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Di mana :

X :Data dengan classyang belum diketahui

H : Hipotesis data merupakan suatu classspesifik

P(H/X) :Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X/H) :Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

Untuk menjelaskan metode Naive bayes, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, metode Naive bayes di atas disesuaikan sebagai berikut :

$$P(C|F1 \dots Fn) = \frac{P(C)P(F1 \dots Fn|C)}{P(F1 \dots Fn)}$$

Di mana Variabel C merepresentasikan kelas, sementara variabel F1 ... Fn merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (Posterior) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut prior), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga likelihood), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global. Karena itu, rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior \times likelihood}{evidence}$$

Nilai Evidence selalu tetap untuk setiap kelas pada satu sampel. Nilai dari posterior tersebut nantinya akan dibandingkan dengan nilai-nilai posterior kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan (C|F1,...,Fn) menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F1, \dots, Fn) &= P(C)P(F1, \dots, Fn|C) \\ &= P(C)P(F1|C)P(F2, \dots, Fn|C, F1) \\ &= P(C)P(F1|C)P(F2|C, F1)P(F3, \dots, Fn|C, F1, F2) \\ &= P(C)P(F1|C)P(F2|C, F1) \\ &\quad P(F3|C, F1, F2)P(F4, \dots, Fn|C, F1, F2, F3) \\ &= P(C)P(F1|C)P(F2|C, F1) \\ &\quad P(F3|C, F1, F2) \dots P(Fn|C, F1, F2, F3, \dots, Fn-1) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor - faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Di sinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing-masing petunjuk (F1,F2...Fn) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(Fi|Fj) = \frac{P(Fi \cap Fj)}{P(Fj)} = \frac{P(Fi \cap Fj)}{P(Fj)} = P(Fi)$$

Untuk i≠j , sehingga

$$P(Fi|C, Fj) = P(Fi|C)$$

b. Penerapan Naïve Bayes

Naive bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional . Keuntungan penggunaan *Naive bayes* adalah metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Dalam metode *Naive bayes* data String yang bersifat konstan dibedakan dengan data numerik yang bersifat kontinyu, perbedaan ini akan terlihat pada saat menentukan nilai probabilitas setiap kriteria baik itu kriteria dengan nilai data string maupun kriteria dengan nilai data numerik. Tahap awal cara kerja dari proses perhitungan *Naive bayes* adalah dengan melakukan pengambilan data *training* dari data mahasiswa dan mahasiwi STMIK Amikom Purwokerto. Adapun variabel penentu yang digunakan dalam mengklasifikasikan data mahasiswa yang siap dinikahi yaitu :

1. Nama

Merupakan variabel nama mahasiswa dan mahasiwi STMIK Amikom Purwokerto .

2. Umur

Merupakan variabel usia mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang di kelompokkan dalam tiga kategori yaitu 20 tahun, 21 tahun, 22 tahun, 23 tahun.

3. Status

Merupakan variabel status mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam dua kategori yaitu Jomblo dan Pacaran.

4. Jenis Kelamin

Merupakan variabel jenis kelamin mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam dua kategori yaitu Perempuan dan Laki-laki.

5. Manajemen Ekonomi

Merupakan variabel status mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam dua kategori yaitu Boros dan Hemat.

6. Menghadapi Masalah

Merupakan variabel status mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam dua kategori yaitu Sabar dan Emosional.

7. Tipe Idaman

Merupakan variabel Tipe pria idaman mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam tiga kategori yaitu religius, tampan/cantik, mapan.

8. Restu Ortu

Merupakan variabel status mahasiswa dan mahasiwi STMIK Amikom Purwokerto yang dikelompokkan dalam dua kategori yaitu Boleh dan belum Boleh.

9. Proses Yang Diharapkan

Merupakan variabel tentang bagaimana cara mahasiwi mendekati tipe idamannya yang dikelompokkan dalam dua kategori yaitu Pacaran dan Ta'aruf .

10. Finansial

Merupakan variabel tentang keterangan diri mahasiswa dan mahasiwi STMIK Amikom Purwokerto sudah bekerja atau belum.

11. Siap Menikah

Merupakan Kelas tujuan yang dikelompokkan dalam dua kategori yaitu Siap dan Belum siap.

Tabell. Data Mahasiswa

Berikut ini adalah jumlah data yang berhasil dikumpulkan, ada 54 data :

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----|----|---------|---|-------|-----------|-----------|-------|---------|-------|-------|
| 1 | 20 | Pacaran | P | Boros | Sabar | Mapan | Boleh | Pacaran | Belum | Belum |
| 2 | 21 | Jomblo | P | Hemat | Emosional | Religi us | Belum | Ta'aruf | Belum | Sudah |
| 3 | 22 | Pacaran | P | Boros | Emosional | Religi us | Belum | Pacaran | Belum | Sudah |
| 4 | 21 | Jomblo | P | Boros | Sabar | Religi us | Belum | Pacaran | Belum | Sudah |
| 5 | 21 | Pacaran | P | Boros | Sabar | Mapan | Boleh | Pacaran | Belum | Belum |
| . | . | . | . | . | . | . | . | . | . | . |
| 54 | 21 | Pacaran | P | Hemat | Sabar | Religi us | Belum | Pacaran | Belum | Belum |

Adapun penerapan metode *Naive bayes* sebagai berikut :

Akan diuji data sebanyak 54 data yang diselesaikan dengan menggunakan *tools* weka sehingga dihasilkan hasil klasifikasi mahasiswa dan mahasiwi STMIK Amikom Purwokerto sebagai berikut:

--- Stratified cross-validation ---
 --- Summary ---

| | | |
|---------------------------------|----|-----------|
| Correctly Classified Instance | 30 | 55,5556 % |
| Incurrectly Classified Instance | 24 | 44,4444 % |

Berdasarkan data di atas dapat dilihat persentase untuk *Correctly Classified Instance* adalah sebesar 56% sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 45%. Di mana dari 54 data mahasiswa dan mahasiwi STMIK Amikom Purwokerto diambil data testing sejumlah 54 data, ada sebanyak 30 data yang berhasil diklasifikasikan dengan benar dan sebanyak 24 data yang tidak berhasil diklasifikasikan dengan benar.

--- Stratified cross-validation ---
 --- Summary ---

| | | |
|---------------------------------|----|-----------|
| Correctly Classified Instance | 33 | 61,1111 % |
| Incurrectly Classified Instance | 21 | 38,8889 % |

Kemudian setelah dilakukan proses filter randomize sebanyak 30 kali dalam modul preprocess pada Weka menghasilkan akurasi yang lebih tinggi yaitu sebesar 61%. Dan jika dihitung akurasinya dari 54 data yang berhasil dikumpulkan maka dapat dihitung sebagai berikut ini :

$$\begin{aligned} \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{(\text{TP} + \text{FN} + \text{FP} + \text{TN})} \times 100 \% \\ &= \frac{23+10}{(23+8+13+10)} = 61,1111\% \end{aligned}$$

Kesimpulan

Berdasarkan Hasil pembahasan maka penulis dapat mengambil beberapa kesimpulan antara lain :

- a. Klasifikasi data mahasiswa dan mahasiswi STMIK Amikom Purwokerto kesiapan pernikahan digunakan untuk menampilkan informasi klasifikasi mahasiswa dan mahasiswi yang sudah siap menikah dengan menggunakan algoritma *Naive bayes*.
- b. Metode *Naive bayes* memanfaatkan data *training* untuk menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk memprediksi kesiapan menikah mahasiswa dan mahasiswi STMIK Amikom Purwokerto berdasarkan proses klasifikasi yang dilakukan oleh metode *Naive bayes* itu sendiri.
- c. Pada klasifikasi data mahasiswa dan mahasiswi STMIK Amikom Purwokerto yang dijadikan data *training* menggunakan metode *Naive bayes* randomize data sebanyak 30 kali menghasilkan akurasi yang lebih tinggi yaitu sebesar 61%

Saran

Adapun saran yang dapat diberikan mengenai penelitian ini adalah pengujian sebaiknya dilakukan dengan menggunakan metode lain untuk melihat metode mana yang lebih akurat dalam memprediksi keakuratan data mahasiswa dan mahasiswi STMIK Amikom Purwokerto yang sudah siap menikah.

Daftar Pustaka

- [1]. ABD. Shomad, 2010. *Hukum Islam (Penormaan Prinsip Syariah dalam Hukum Indonesia)*. Penerbit Kencana Prenada Media Group : Jakarta.
- [2]. ABD. Shomad, 2010. *Hukum Islam (Penormaan Prinsip Syariah dalam Hukum Indonesia)*. Penerbit Kencana Prenada Media Group : Jakarta.
- [3]. Pustaka Utama. Papalia, Olds & Feldmann. 1998. *Human Development*. Edisi 7. Newyork. McGraw-Hill. Jurnal.
- [4]. Hurlock, E.B, 1980. *Psikologi Perkembangan suatu pendekatan sepanjang rentang kehidupan*. Edisi 5. Jakarta: Erlangga.
- [5]. Havirghust, Robrt J, 1972. *Human development and education*. Panduan Pustaka : Yogyakarta
- [6]. Hurlock, E.B, 1980. *Psikologi Perkembangan suatu pendekatan sepanjang rentang kehidupan*. Edisi 5. Jakarta: Erlangga
- [7]. Taruna R., S., Hiranwal, S., 2013, Enhanced *Naive bayes* Algorithm for Intrusion Detection in *Data Mining*, International Journal of Computer Science and Information Technologies, Vol.6, No. 4, Hal 960-962
- [8]. Susanto Erdi. 2012, *Data Mining menggunakan Weka*, www.erdusanto.com/datamining-menggunakan-weka.html (diakses 16 Agustus 2016)
- [9]. Turban, E. et al. 2005. *Decision Support Systems and Intelligent Systems (Sistem Pendukung Keputusan dan Sistem Cerdas)*. Andi Offset. Yogyakarta.
- [10]. Pengertian data *training*, set validation <https://philips.wordpress.com/2006/08/10/training-set-testing-set-dan-validation-set/> (diakses 16 Agustus 2016)
- [11]. Taruna R., S., Hiranwal, S., 2013, Enhanced *Naive bayes* Algorithm for Intrusion Detection in *Data Mining*, International Journal of Computer Science and Information Technologies, Vol.6, No. 4, Hal 960-962.
- [12]. Taruna R., S., Hiranwal, S., 2013, Enhanced *Naive bayes* Algorithm for Intrusion Detection in *Data Mining*, International Journal of Computer Science and Information Technologies, Vol.6, No. 4, Hal 960-962.