

BAB V

PENUTUP

5.1 KESIMPULAN

Kesimpulan yang dapat di ambil dari hasil yang penelitian yang telah dilakukan pada bab-bab sebelumnya yaitu:

1. Penelitian ini menggunakan data mahasiswa tahun 2017 dan 2018 dengan jumlah data masing-masing 101 dan 104 dengan total 205 data.
2. Data-data mahasiswa ini mengalami banyak *noise* hal ini diketahui melalui proses *cleaning data*, yaitu proses menghilangkan noise pada data dengan menggunakan Microsoft Excel yang penulis buat sendiri rumusnya ,dari data mahasiswa system informasi tahun 2017 dan 2018, data noise kebanyakan pada Asal Sekolah dan Jurusan Asal Sekolah yang tidak di inputkan Mahasiswa dengan benar, setelah melakukan data cleaning, maka di dapatkan data untunk training, dilakukan data testing yaitu dengan 200 data
3. Persentasi hasil akurasi klasifikasi naïve bayes pada data mahasiswa tahun 2017 dan 2018 dengan menggunakan *Use Training Set* dengan jumlah kelas *Correctly Classified Instances* 161, *Incorrectly Classified Instances* 39, dan persentasi akurasi *Correctly Classified Instances* sebesar 80,5 %, *Incorrectly Classified Instances* 19,5 %.menggunakan *10-Fold Cross Validation* dengan

jumlah kelas *Correctly Classified Instances* 156, *Incorrectly Classified Instances* 44, dan persentasi akurasi *Correctly Classified Instances* sebesar 78%, *Incorrectly Classified Instances* 22%. Setelah dilakukan pengolahan data training maka diperoleh akurasi pada model tersebut. Akurasi pada model dihitung dengan menggunakan confusion matrix..

4. Hasil klasifikasi naïve bayes dengan akurasi terbesar diperoleh dengan menggunakan *Use Training Set*, Hasil dari seleksi atribut menggunakan algoritma *Classifier Attribute Evaluation (ClassifierAttributeEval)* menghasilkan bahwa atribut yang paling berpengaruh terhadap kecepatan kelulusan adalah IPK.
5. Lama masa studi berdasarkan atribut jenis kelamin yaitu pada Kelas TEPAT (65 Laki-laki dan 87 Perempuan) dan Kelas LAMBAT (30 Laki-laki dan 18 Perempuan). Pada atribut Sekolah Asal yaitu pada Kelas TEPAT (69 SMK, 7 MA dan 76 SMA), dan Kelas LAMBAT (10 SMK, 2 MA dan 36 SMA). Pada Atribut Jurusan Sekolah Asal yaitu pada Kelas LAMBAT (1 Administrasi, 3 Akuntansi, 1 TataBusana, 0 Grafika, 11 IPA, 26 IPS, 1 Keagamaan, 3 Multimedia, 0 Broadcasting, 0 Pemasaran, 2 TKJ). Dengan sebaran data Kelas TEPAT (4 Administrasi, 40 Akuntansi, 0 TataBusana, 1 Grafika, 24 IPA, 58 IPS, 1 Keagamaan, 16 Multimedia, 2 BroadCasting, 3 Pemasaran, 3 TKJ). Pada atribut IPK yaitu pada Kelas TEPAT (115 Tinggi, 19 sedang dan 18 Rendah), dan Kelas LAMBAT (25 Tinggi, 6 sedang dan 16 Rendah).
6. Total keseluruhan akurasi adalah $\frac{78\%+77\%}{2} = 77,5\%$, akurasi yang cukup baik dalam menguji kebenaran dari data training.

5.2 SARAN

1. Diharapkan kedepannya penelitian ini menggunakan data mahasiswa lebih banyak lagi serta mencakup prodi lainnya agar memiliki persentasi akurasi yang lebih baik.
2. Data ini diujikan menggunakan algoritma klasifikasi lainnya.