I**NFORMATION SYSTEM RECOMMENDATIONS FOR STUDENTS' RETURN PERMISSION WITH PREDICTION FEATURES OF DELAY RETURN TO BOARDING POSTS USING THE NAIVE BAYES METHOD**

(STUDI KASUS PONDOK PESANTREN MHI BANGSALSARI)

**Muhammad Khotibul Umam**

NPM 23066020020

UPN “Veteran” East Java

**ABSTRACT**

Pondok Pesantren MHI Bangsalsari *is a boarding school that still uses the salafiyah culture. Islamic boarding schools that apply the study of Al-Quran, Hadith and the yellow books written by previous scholars. Aiming at forming the character of students who are knowledgeable and have noble character.*

*Islamic boarding schools have an important role in transmitting Islam and creating new successors, so pesantren education must be carried out in a disciplined manner so that students can seek maximum knowledge. For the sake of smooth teaching and learning at the pesantren, it must be limited to a student who will leave or return, by building a website-based application, namely an information system for students' return home with a prediction feature for delays using the Naive Bayes method.*

*This datamining method uses the naive bayes classification with attributes that can be used in this system such as type of permit, distance of permission, duration of permission, previous permits and violations, so that it can produce predictions of the delay in returning to the pesantren and the permit security team can easily determine the permission of the students according to predictions. which has been tested every student data.*

Kata kunci : prediksi, *naïve bayes* , izin santri

**ABSTRAK**

Pondok Pesantren MHI Bangsalsari merupakan pesantren yang masih menggunakan kultur salafiyah. Pesantren yang menerapkan pembelajaran Al-Quran, Hadits serta kitab-kitab kuning karangan ulama’ terdahulu. Bertujuan membentuk karakter santri yang berilmu serta berakhlak mulia.

Pesantren mempunyai peran penting dalam mensiarkan agama islam dan mencetak penerus yang baru, maka pendidikan pesantren harus dilakukan dengan disiplin agar santri dalam mencari ilmu dapat maksimal. Demi kelancaran belajar mengajar di pesantren maka harus dibatasi seorang santri yang akan keluar maupun pulang, dengan membangun aplikasi berbasis website yaitu sistem informasi izin pulang santri dengan fitur prediksi keterlambatan menggunakn metode *naive bayes.*

Metode datamining ini menggunakan klasifikasi *naive bayes* dengan atribut yang dapat digunakan pada sistem ini seperti jenis izin, jarak izin, durasi izin, izin sebelumnya dan pelanggaran, sehingga dapat menghasilkan prediksi keterlambatan santri kembali ke pesantren dan tim keamanan izin dapat mudah menentukan izin santri sesuai prediksi yang telah diuji coba setiap data santri*.*

Kata kunci : prediksi, *naïve bayes* , izin santri .

1. **Introduction**

Pondok Pesatren Mamba’ul Khoiriatil Islamiyah (MHI) Bangsalsari merupakan lembaga yayasan yang masih menjaga kebudayaan dan keilmuan islam salafiyah. Pesantren ini menerapkan kurikulum Al-Quran, Al-Hadits serta berpegang teguh pada ajaran ulama’ salaf, agar membentuk generasi islam yang berakhlaqul karimah dan ahlusunnah waljamaah, serta agar menjadi generasi islam yang mempunyai kualitas dan pemahaman yang sesuai dengan syariat. Periode masa sekolah madin pesantren, setiap santri wajib menempuh pendidikan dalam waktu sembilan tahun yakni lima tahun Ibtidaiyah, dua tahun Tsanawiyah dan dua tahun Aliyah.

Santri dituntut untuk bisa memanfaatkan dan membagi waktu agar tidak melebihi kelulusan yang ditargetkan. Selama proses pendidikan santri dipengaruhi oleh faktor yang bisa menghambat proses belajar dan menghambat target kelulusan santri. Salah satu faktor yang menghambat proses belajar, yaitu terlambatnya kembali kepesantren serta penyalahgunaan izin oleh santri. Konsekuensinya tertinggalnya materi, pembelajaran dan menulis kitab, yang menyebabkan santri tersebut tidak boleh ikut ujian yang diadakan tiga kali dalam satu tahun. Sehingga tidak dapat mentarget kelulusan yang telah ditetapkan.

Kasus keterlambatan kembali yang sepenuhnya bukan kesalahan dari santri, faktor lain adalah tidak tepatnya petugas pemberian izin. Karena pemberian izin masih manual, sehingga petugas bisa memberikan izin yang tidak tepat dan juga ada perbedaan keputusan izin antar petugas yang lain.

Berdasarkan permasalahan di atas, maka dilakukan penelitian untuk membuat sistem informasi perizinan santri dengan fitur prediksi menggunakan metode *Naive Bayes*, Sistem ini diharapkan mampu mengelolah data perizinan yang telah ada, dan diharapkan bisa membantu pengurus perizinan mampu memprediksi keterlambatan izin santri. Sehingga dapat mengurangi peluang keterlambatan izin santri serta menerapkan sanksi/denda jika santri melebihi batas waktu izin yang telah di tentukan.

Penelitian ini bertujuan untuk mempertimbangkan keterlambatan kembali ke pesantren kami memberi fitur prediksi keterlambatan dengan mengambil data yang sebelumnya, serta ketentuan dan kebijakan yang telah diatur oleh pihak ketertiban dan keamanan pesantren untuk mempertimbangkan proses izin santri, dengan fitur prediksi keterlambatan menggunakan metode *naive bayes* supaya meminimalisir keterlambatan kembali ke pesantren serta membantu santri dalam kelancaran proses belajar dipesantren.

1. **Theoritical Background**
   1. **Data Mining**

Data mining (Turban dkk. 2005) merupakan proses sebuah teknik statistika, matematika, kecerdasan buatan, pengambilan keputusan, *machine learning* untuk mengidentifikasi dan menghasilkan informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu *Knowledge discovery in databases* (KDD), ekstraksi pengetahuan (*knowledge extraction*), analisa data/pola (*data/pattern analysis*), kecerdasan bisnis (*business intelligence*) dan *data archaeology* dan *data dredging* (Larose, 2005).

Kemampuan Data mining untuk mencari informasi bisnis yang berharga dari basis data yang sangat besar, dapat dianalogikan dengan penambangan logam mulia dari lahan sumbernya, teknologi ini dipakai untuk:

1. Prediksi tren dan sifat-sifat bisnis, dimana data mining mengotomatisasi proses pencarian informasi memprediksi di dalam basis data yang besar.
2. Penemuan pola-pola yang tidak diketahui sebelumnya, dimana data mining menyapu basis data, kemudian mengidentifikasi pola-pola yang sebelumnya tersembunyi dalam satu sapuan.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.
   1. **Metode Naïve bayes**

Metode *Naive Bayes* (Vulandari, 2017: 33) merupakan salah satu teknik Data Mining dengan dasar teorema *Bayes*, dimana metode ini digunakan untuk memprediksi peluang di masa depan berdasarkan data-data masa lalu. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak berhubungan dengan ciri dari kelas lainnya. Atau adanya asumsi independensi (ketidaktergantungan) yang kuat (naif). Keuntungan metode ini hanya membutuhkan data pelatihan yang kecil untuk menentukan estimasi paremeter yang diperlukan dalam pengklasifikasian. Tahapan algoritma *Naïve Bayes* adalah :

* 1. Menghitung jumlah kelas/ label.
  2. Menghitung jumlah kasus per kelas.
  3. Kalikan semua variabel kelas.
  4. Bandingkan hasil per kelas.

Persamaan umum dari Teorema Bayes mengikuti Persamaan 1.

P(C|𝑋) = (1)

Keterangan :

X = Data yang memiliki kelas (label) yang tidak diketahui

C = Hipotesa bahwa X adalah suatu kelas (label) spesifik.

P(C|X) = Peluang hipotesis berdasar kondisi *(posterior probability*).

P(C) = Peluang hipotesis (*prior probability*).

P(X|C) = Peluang berdasarkan kondisi pada hipotesis.

P(X) = Probabilitas C.

Adapun persamaan *Naive Bayes* untuk klasifikasi mengikuti Persamaan 2.

P(C|𝑋) = (2)

P(X|C) adalah probabilitas data dengan vektor X pada kelas C. P(C) adalah probabilitas awal kelas C.

Nilai P(X) selalu tetap sehingga dalam perhitungan tinggal menghitung bagian mengikuti Perhitungan 2.3 :

𝑃(𝐶)∏𝑞 𝑖=1 𝑃(𝑋𝑖|𝐶) (3)

merupakan probabilitas independen kelas C dari semua fitur dalam vektor X dengan memilih yang terbesar sebagai kelas terpilih yang dijadikan hasil prediksi.

Penelitian ini akan menggunakan metode *Naive bayes* untuk memprediksikan keterlambatan izin pulang dapat dilihat pada gambar 1.

Input data Izin Santri

Menghitung Jumlah Posterior Probabilitas santri izin

Menghitung posterior kasus santri terlambat

**­­**

Menghitung posterior kasus santri tidak terlambat

Probabilitas Tidak Terlambat > = Terlambat

Hasil Terlambat

Hasil Tidak Terlambat

Gambar 1. Flowchart Metode Naive Bayes

(Sumber Vulandari, 2017: 33)

Keterangan *flowchart* :

* 1. Mulai.
  2. Memasukkan data trining santri izin.
  3. Menghitung jumlah posterior probabilitas santri izin dengan rumus P(C).
  4. Menghitung posterior probabilitas santri terlambat dengan rumus P(C|X) .
  5. Menghitung posterior probabilitas santri tidak terlambat dengan rumus P(C|X).
  6. Membandingkan antara data santri terlambat dan santri tidak terlambat dengan rumus dengan mengikuti Persamaan 4.

P(C|𝑋) = (4)

* 1. Jika probabilitas santri tidak terlambat lebih besar dari santri terlambat maka hasilnya santri tidak terlambat jika tidak hasilnya santri terlambat .
  2. Selesai.

1. **Research Method**

Metode yang digunakan pada penelitian ini adalah *naïve bayes* dalam menentukan alternatif yang optimal untuk mengklasifikasian jenis izin dan mengatasi keterlambatan kembali ke pesantren, dimana metode ini digunakan untuk memprediksi peluang di masa depan berdasarkan data-data masa lalu. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak berhubungan dengan ciri dari kelas lainnya.

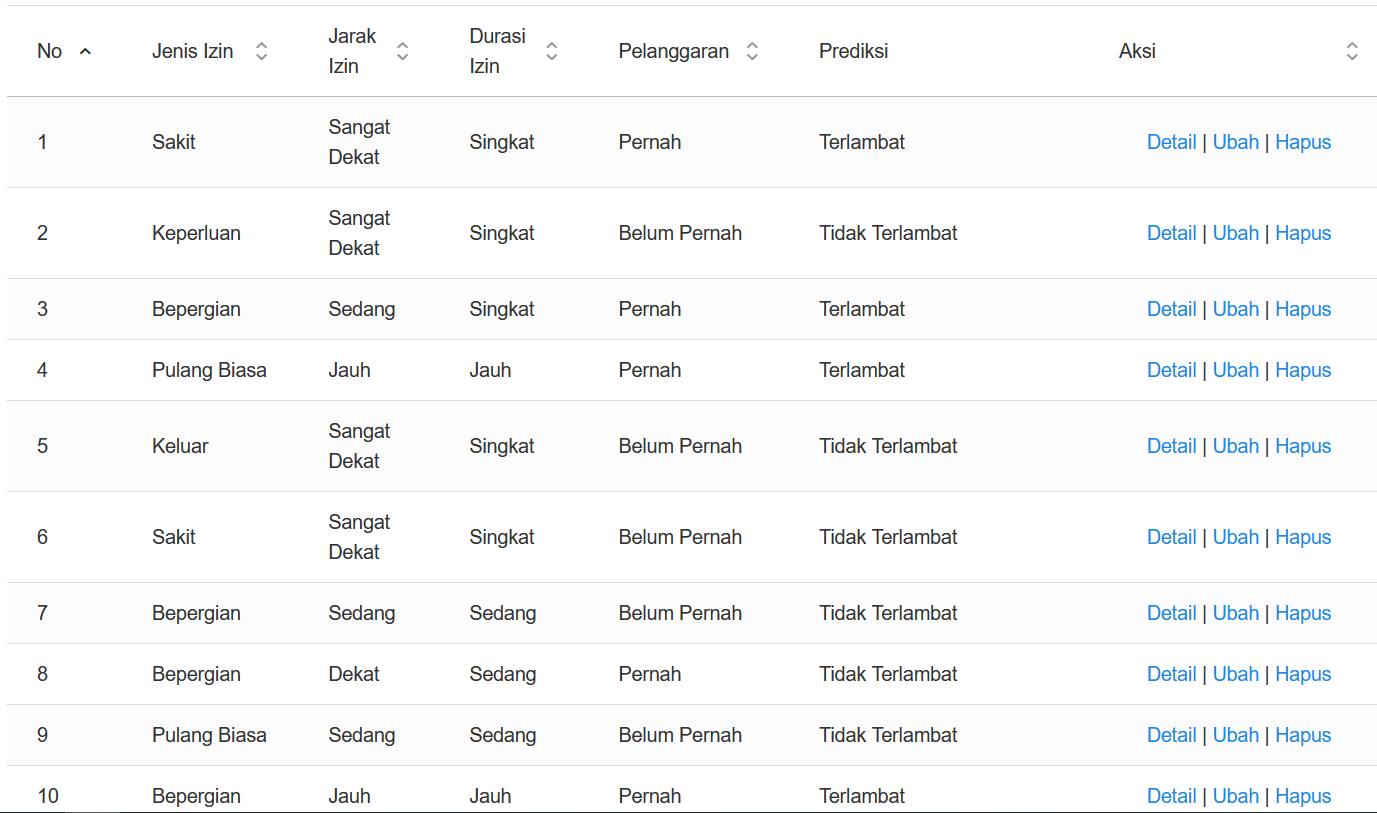
Proses penelitian yang dilakukan meliputi pengumpulan observasi, dan studi pustaka, melakukan analisa terhadap kebutuhan aplikasi, dan melakukan perancangan dari hasil analisa,

* 1. **Pengumpulan data**

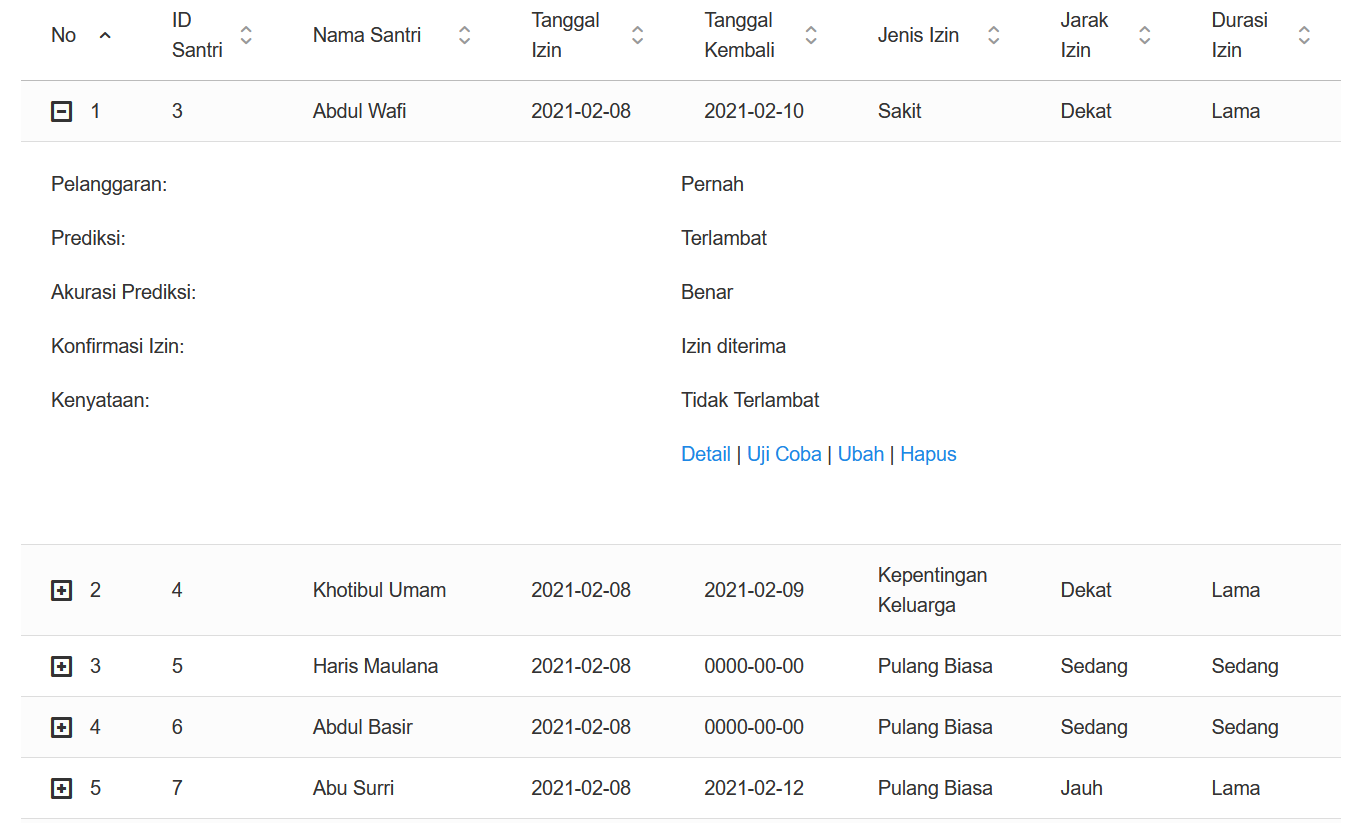
Sebelum melakukan prediksi keterlambatan santri kembali ke pesantren diperlukan data *training* untuk mengklasifikasi status keterlambatan santri menggunakan metode *naive bayes*. Data training merupakan data yang digunakan sebagai pengambilan keputusan menggunakan metode pembacaan pola data. Banyaknya data *training* maka akan menghasilkan tingkat akurasi yang semakin tinggi. Data training didapatkan dari data izin santri sebelumnya sebagai acuan proses prediksi menggunakan metode *naive* *bayes.* Berikut merupakan data training dan data testing, sebagian kecil data training dan data testing yang dimiliki, atribut data training mempunyai kelas yang berisi Id dataset, jenis izin, jarak izin, durasi izin, pelanggaran dan prediksi, untuk kelas data testing mempunyai kelas yang berisi id datates, nama santri, tanggal izin, tanggal kembali, jenis izin, jarak izin, durasi izin, pelanggaran, prediksi, akurasi prediksi, konfirmasi izin, dan kenyataan, dimana ada 2 kelas status prediksi terlambat dan tidak terlambat. Mempunyai dataset sebanyak 200 data dan data testing sebanyak 200 data.

1. **Result and Analysis**
   1. **Tahapan Implementasi Naïve Bayes**

Naive bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuan inggris *Thomas bayes,* yaitu memprediksikan peluang akan datang dengan menggunakan data sebelumnya sehingga disebut teori *bayes.* Teori tersebut dikombinasikan dengan *Naive* dimanaasumsi kondisi atribut saling bebas. Metode *naive bayes* merupakan metode yang dapat mengelolah data dalam bentuk numerik maupun kategori. Untuk tahapan pertama dalam perhitungan naive bayes yaitu menghitung rata-rata data set atau data trining dari sebuah data, Berikut merupakan data training dan data testing dapat di lihat pada Gambar 2 dan Gambar 3.



Gambar 2 Dataset/ Data *training*

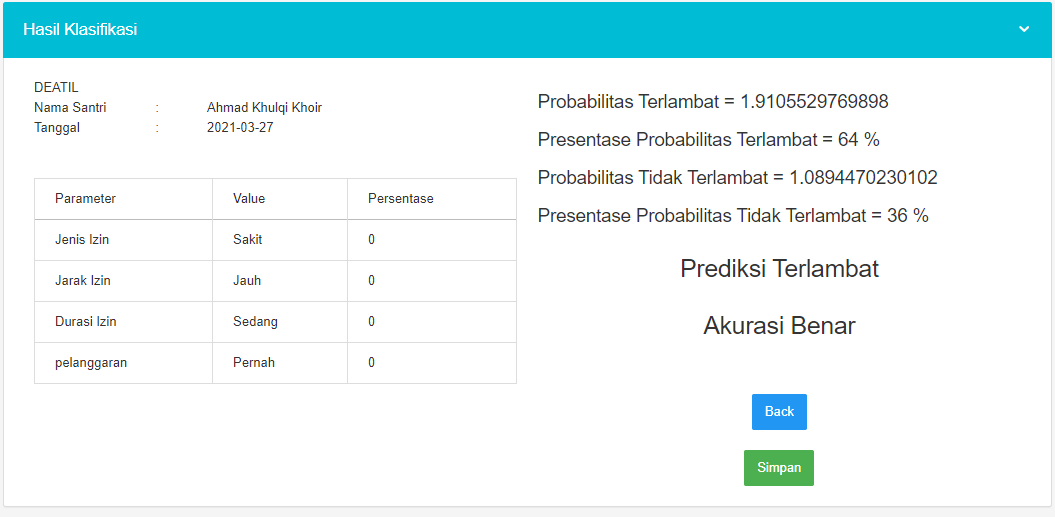


Gambar 3. Data *Testing*

* 1. **Analisa Uji Prediksi metode naïve bayes**

Proses pengujian prediksi dilakukan dengan menggunakan data training dan data testing. Pada pengujian ditentukan data training yang akan di uji menggunakan 200 dataset dan setelah itu hasil dari prediksi dijadikan sebagai data testing. Dapat dilihat pada Gambar 4.

.



Gambar 4. Hasil Uji Data *Testing*

Proses perhitungan prediksi akurasi pengujian mengikuti Persamaan 5.

Terlambat = 106 data.

Tidak Terlambat = 94 data.

Pulang Biasa dan Terlambat = 29 data

Pulang Biasa dan Tidak Terlambat = 17 data

Sedang dan Terlambat = 25 data

Sedang dan Tidak Terlambat = 33 data

Sedang dan Terlambat = 24

Sedang dan Tidak Terlambat = 34

Pernah dan Terlambat = 83

Pernah dan Tidak Terlambat = 6

Terlambat = x x x

Tidak Terlambat = x x x ………………(5)

Maka Hasilnya mengikuti Persamaan 6.

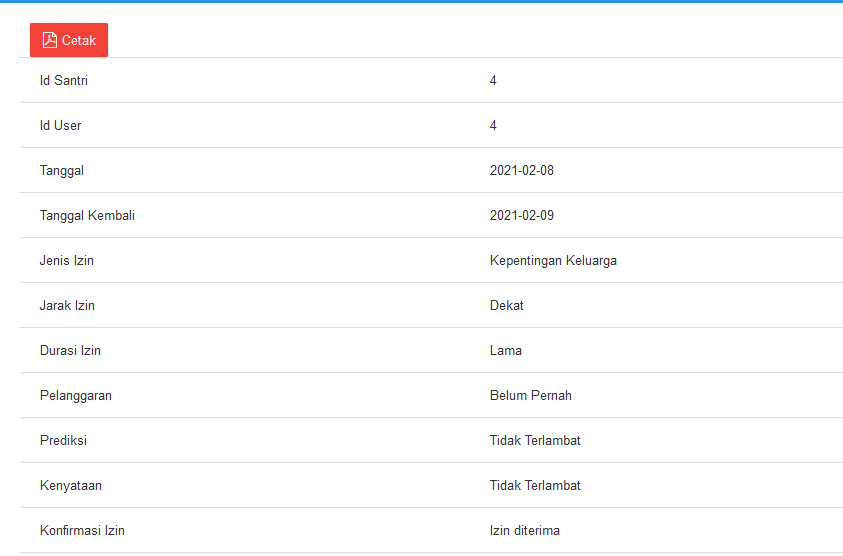
Terlambat = x x x = 0.96822972442549

Tidak Terlambat = = x x x = 0.031770275574511………….(6)

Jadi hasil perdiksi dari Santri bernama Hasanuddin yaitu Terlambat.

* 1. **Menganalisa keterlambatan dan mengatasi kasus keterlambatan**

Setelah mengetahui hasil prediksi dan data izin pulang santri dengan mengetahui kenyataan kasus yang ada maka untuk mengatasi keterlambatan santri sistem ini menyedikan fitur denda, jadi setiap santri ketika kembali ke pondok harus mengerahkan surat izin yang telah di print out kepada petugas / admin, untuk santri yang terlambat akan di kenakan sanksi denda yang di tentukan sistem sesuai dengan lamanya terlambat kembali. Dapat di lihat pada Gambar 5.



1. **Conclusion**

Implementasi website Sistem informasi izin santri pulang dengan fitur keterlambatan kembali ke pesantren mempunyai fitur lengkap seperti data santri, data training, data testing, prediksi, dan data user/wali santri. Untuk mengimplementasikan fitur prediksi yaitu dengan menggunakan salah satu santri izin pulang dengan kriteria class jenis izin = pulang, jarak izin = sedang, durasi izin = sedang, pelanggaran = pernah jadi probabilitas keterlambatan 0.96822972442549 dan presentase probabilitas keterlambatan mencapai 97 %, serta probabiltas tidak terlambat mencapai 0.031770275574511 dan presentase probabilitas tidak terlambat mencapai 3 %. Pengujian data menggunakan 200 data training dan data testing.

Mengatasi kasus keterlambatan menggunakan data kenyataan setelah santri tersebut sudah datang dipesantren jadi dapat menentukan akurasi prediksi yang telah dilakukan pada masa sebelumnya dan memberi tabel denda setiap santri yang datang ke pesantran dengan status Terlambat dengan nominal terlambat 1 hari = Rp. 10.000 dan terlambat 2 hari = Rp. 20.000.

**References**

1. Demigha, S. (2015). *makalah disajikan pada IEEE International Conference on Computer Science & Education ke-10*. From Data Mining for Breast Cancer Screening.
2. Kasih, P. &. (2017). *Sistem Bantu Pemilihan Dosen Pembimbing Tugas Akhir Berdasarkan Kategori Pilihan dan Keahlian Dosen Menggunakan Naïve Bayes.* From Makalah disajikan pada Seminar Nasional Teknologi Informasi, Komunikasi dan Aplikasinya: https://jurnal.stiki.ac.id/
3. Larose. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. From John Willey & Sons, inc.
4. Nofriansyah, D. E. (2015). *Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL*. From https://lppm.trigunadharma.ac.id/
5. Nugroho, Y. &. (2015). *Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali*. From Khazanah Informatika Jurnal Ilmu Komputer dan Informatika, 1 (1), 1-6.
6. Pratiwi, H. (2015). *Buku Ajar Sistem Pendukung Keputusan*. From Yogyakarta: Deepublish.
7. Purwanto. (2017). *Metodologi Sistem Defelopment Life Cycle*. From https://medium.com:https://medium.com/@purwanto.dev/metodologi-system-life-cycle-sdlc-2f0349df1364.
8. Turban, E. (2005). *Decision Support System and Intelligent Sistem Edisi Bahasa Indonesia Jilid 1*. From Andi : Yogyakarta: www.kajianpustaka.com.