**Optimasi Data Mining Menggunakan Metode Algoritma Naïve Bayes Untuk Classification Prediksi Keterlambatan Pembayaran Infak**

**Di MI Miftahul Huda Sukolilo-Jabung**

****

**Oleh :**

**Nama : Dini Kristianti**

**NIM : 200605220015**

**Kelas : B**

**UNIVERSITAS ISLAM NEGERI**

**MAULANA MALIK IBRAHIM**

**MALANG**

**2021**

**Abstrak**

*Data Mining adalah proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika. Dalam pembayaran Infak sekolah yaitu apabila para siswa terlambat dalam membayar Infak sekolah. Hal tersebut menjadi masalah karena pembayaran Infak sekolah merupakan salah satu sumber dana tetap dalam meningkatkan kualitas pendidikan bagi sekolah swasta. Pada makalah ini,penulis berusaha menambang data (data mining) transaksi pembayaran Infak untuk memprediksi keterlambatan pembayaran Infak kelas 1, pada Bulan Juni 2020 sampai April 2021, data pembayaran tersebut di rangkum menjadi sebuah data pembayaran Infak . Data yang ada dianalisis menggunakan perhitungan Rapid Miner 9.9 dengan algoritma Naïve Bayes. Naïve Bayes merupakan salah satu metode pada Probabilistic Reasonig. Algoritma Naïve Bayes bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan data transaksi penjualan dalam memprediksi prilaku pola pembeli.*

***Kata kunci:*** *Data Mining, Klasifikasi, Algoritma Naïve Bayes, Rapid Miner 9.9, Infak, INFAK, Pendidikan, Sekolah*

**DAFTAR ISI**

[**DAFTAR ISI** i](#_Toc72151067)

[**BAB I PENDAHULUAN** 1](#_Toc72151068)

[**1.1.** **Latar Belakang** 1](#_Toc72151069)

[**1.2.** **Perumusan Masalah** 2](#_Toc72151070)

[**BAB II TEORI PENDUKUNG** 3](#_Toc72151071)

[**1.1.** **Data Mining** 3](#_Toc72151072)

[**1.2.** **Naïve Bayes** 3](#_Toc72151073)

[**1.3.** **RapidMiner** 3](#_Toc72151074)

[**BAB III METODE ANALISIS** 5](#_Toc72151075)

[**BAB IV PEMBAHASAN** 6](#_Toc72151076)

[**4.1.** **Proses Data Mining** 6](#_Toc72151077)

[**4.1.1.** **Pemrosesan Awal** 6](#_Toc72151078)

[**4.1.2.** **Accuracy Data** 7](#_Toc72151079)

[**4.1.3.** **Coffusion Matrix** 7](#_Toc72151080)

[**4.1.4.** **Hasil Keseluruhan Prediksi Naïve Bayes** 7](#_Toc72151081)

[**BAB V PENUTUP** 8](#_Toc72151082)

[**5.1. Kesimpulan** 8](#_Toc72151083)

[**DAFTAR PUSTAKA** 9](#_Toc72151084)

# **BAB I PENDAHULUAN**

## **Latar Belakang**

Pendidikan Indonesia pada setiap tahunnya semakin maju dan berkembang. Pemanfaatan teknologi informasi dan komunikasi pada dunia pendidikan sangatlah di perlukan. Baik dalam pembelajaran maupun dalam mendukung administrasi Tata Usaha dan Keuangan sekolah.

Menurut (SANJIWANI, 2012) Biaya pendidikan merupakan salah satu komponen masukan instrumental yang sangat penting dalam penyelenggaraan pendidikan. Dalam setiap upaya pencapaian tujuan pendidikan, baik tujuan yang bersifat kuantitatif maupun kualitatif, biaya pendidikan memiliki peranan yang sangat menentukan.

Mi Miftahul Huda Sukolilo adalah MI swasta yang dimana biaya operasional sekolah di bantu oleh komite sekolah dalam bentuk *Infak. Infak* adalah nama lain dari INFAK iuran Sumbangan Pembinaan Pendidikan. Iuran yang dibayarkan setiap bulan. Untuk tanggal pembayaran *Infak,* setiap bulannya paling lambat adalah tanggal 10 pada setiap bulannya.

Masalah yang muncul terkait dalam pembayaran *Infak* sekolah yaitu apabila para siswa terlambat dalam membayar *Infak* sekolah. Hal tersebut menjadi masalah karena pembayaran *Infak* sekolah merupakan salah satu sumber dana tetap dalam meningkatkan kualitaspendidikan sekolah. Berdasarkan data dari bagian keuangan tata usaha, bahwa siswa yang terlambat membayar pada tahun ajaran 2019/2020 sekitar 20% dari total seluruh siswa yang berjumlah 420 siswa. Hal itu disebabkan oleh beberapa faktor seperti : pendapatan orang tua, pendidikan orang tua, tanggungan keluarga dan usia. Hal tersebut menjadi masalah yang besar karena banyaknya siswa yang terlambat dalam membayar *Infak* akan menurunkan pemasukan dana tetap sekolah. Sedangkan biaya pengeluaran untuk gaji para guru dan karyawan sebagian besar dialokasikan dari biaya INFAK sekolah. Maka perlu adanya solusi berupa sistem prediksi yang dapat meperkirakan siswa yang tepat atau terlambat dalam melakukan pembayaran INFAK sekolah.

Jumlah data yang terus meningkat memerlukan beberapa metode untuk mengolah dan mengambil kesimpulan serta informasi dari data tersebut. Beberapa metode yang di gunakan utnuk mengolah data yang sifatnya besar untuk menemukan pola yang terdapat di dalamnya diantaranya adalah : teknik *klastering*, *analisis* *diskriminan*, *teorema bayes*, *decision tree artificial nural networks*, *support vector machine, regresi linear, support vector regresi*. Setiap metode tersebut memiliki algoritma-algoritma yang digunakan untuk memproses data yang ada.

Dalam makalah ini akan dilakuakn komparasi hasil klasifikasi dari metode Naïve Bayes Untuk Classification Prediksi Keterlambatan Pembayaran Infak Di MI Miftahul Huda Sukolilo-Jabung.

## **Perumusan Masalah**

Berdasar pada latar belakang di atas, maka dapat di rumuskan masalah pada pembuatan makalah ini, sebagai berikut :

1. Bagaimana membangun model yang dapat digunakan untuk Mengklasifikasi Prediksi Keterlambatan Pembayaran Infak Di MI Miftahul Huda Sukolilo-Jabung?
2. Bagaimana menggunakan model untuk Mengklasifikasi Prediksi Keterlambatan Pembayaran Infak Di MI Miftahul Huda Sukolilo-Jabung?

# **BAB II TEORI PENDUKUNG**

## **Data Mining**

Menurut Gartner Group, *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistic dan matematika. Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa *data mining* mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dulu. Berawal dari beberapa disiplin ilmu, *data mining* bertujuan untuk memperbaiki teknik tradisional sehingga bias menangani:

1. Jumlah data yang sangat besar.
2. Dimensi data yang tinggi.
3. Data yang heterogen dan berbeda bersifat

## **Naïve Bayes**

Pengklasifikasi *bayesian* adalah pengklasifikasi statistik dan didasarkan pada teorema *bayes*. Teori keputusan *bayes* adalah pendekatan statistic yang fundamental dalam pengenalan pola (*pattern* *recognition*), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bias dilihat statistiknya. Misalkan *X* adalah set atribut data dan *h* kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan *X* dan *h* sebagai variabel acak dan menangkap hubungan peluang *P*(*h|X*) ini peluang posterior untuk *h* dan sebaliknya perior *P*(*h*).

## Bahasa R

Menurut (Gio & Effendie, 2017) R merupakan bahasa pemrograman statistika yang dapat digunakan untuk analisis dan manipulasi data statistika (pemodelanstatistika), dan grafik. R diciptakan (*was created*) oleh Ross Ihaka dan Robert Gentleman (nama depansama-sama diawali dengan huruf R & R, **Ross** dan **Robert**) dari departemen statistika, di UniversitasAuckland, New Zealand. Saat ini R dikembangkan oleh *R Development Core Team*. Munculnya Rterinspirasi oleh bahasa statistika dengan nama S (*statistical language S*). Bahasa statistika S diciptakan oleh John Chambers dan rekan, di *Bell Labs* atau *AT&T Bell Laboratories*. Selain bahasastatistika S, bahasa pemrograman C juga dikembangkan di *AT&T Bell Laboratories*.Beberapa *software* sepertiSPSS, Minitab, SAS, EViews, dan sebagainya, perlu membayar sejumlahuang, untuk mendapatkan izin menggunakan, sementara R gratis. Di samping itu, R juga tersediauntuk di-*install* di sistem operasi *Windows, Mac OS X*, dan *Linux*. Gambar 1.1 merupakan tampilan 4 *website https://cran.r-project.org/.* Di *website* tersebut, R dapat di-*download* sesuai dengan sistemoperasi. Jika sistem operasi yang digunakan adalah *Windows*, maka pilih *Download R for Windows*.

# **BAB III METODE ANALISIS**

Metode yang digunakan adalah metode kualitatif dengan pendekatan deskriptif. Metode kualitatif sering disebut metode penelitian naturalistik karena penelitiannya dilakukan pada kondisi yang alamiah, dapat diartikan sebagai usaha untuk menyelidiki keadaan yang sebenarnya, dalam memprediksi prilaku pola pembelian berdasarkan jenis produk sehingga perencanaan strategi penjualan dapat tercapai.

# **BAB IV PEMBAHASAN**

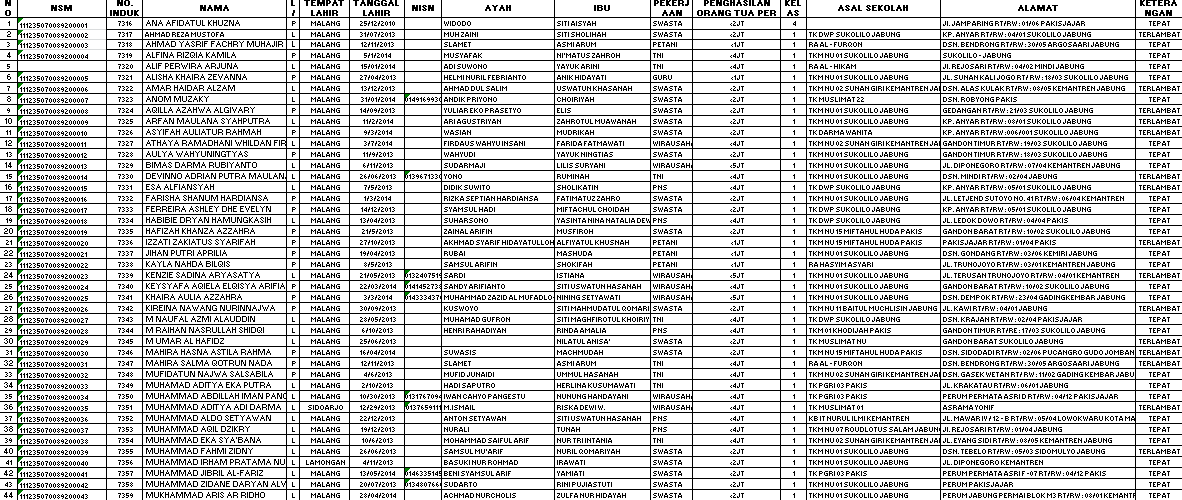
## **Proses Data Mining**

### **Pemrosesan Awal**

**Persiapan Data**

Data siswa kelas 1 pada Tahun Ajaran 2020/ 2021, dilakukan teknik data preparation agar kualitas data diperoleh lebih baik.

Berikut adalah data training dalam bentuk Excell yang akan di olah.



Gambar 1 Data siswa kelas 1 dalam format Excell

**Proses Training dan Testing**

Pada proses Training dan Testing, silahkan di mulai dengan data excel diatas di olah dengan script di bawah ini:



Gambar 2

Arti dari script di atas adalah untuk menginputkan dan menampilkan data excel, dengan cara copy paste.

Untuk Library yang di butuhkan adalah :



Gambar 3

Setelah menginstall library di atas maka bisa kita lakukan Testing dan Training data, pada gambar 4.



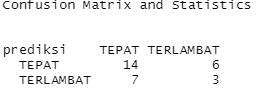
Gambar 4

**Hasil Performance Vector**

### **Accuracy Data**

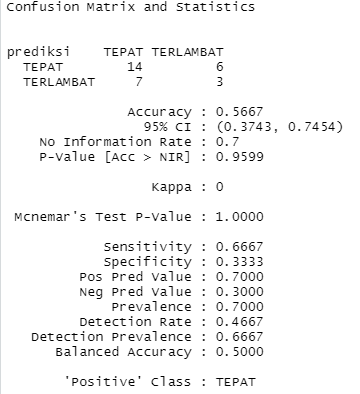
Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasilprediksi yaitu (95% )dari hasil data testing.

### **Coffusion Matrix**



Dengan mengetahui jumlah data yang diklasifikasikan secara benar maka dapat diketahui akurasi hasilprediksi yaitu 0.5667.

### **Hasil Keseluruhan Prediksi Naïve Bayes**



# **BAB V PENUTUP**

**5.1. Kesimpulan**

Hasil analisa data training menggunakan metode Naïve Bayesian diperoleh hasil akurasinya sebesar 95% dari hasil data testing dimana hasil Sensitivity : 0.6667, Accuracy : 0.5667, Kappa : 0, Specificity : 0.3333, Balanced Accuracy : 0.5000. Maka dengan ini peneliti menyimpulkan bahwa optimasi metode *Naïve Bayes* dapat membantu dalam melakukan prediksi keterlambatan pembayaran *Infak* sekolah dengan hasil akurasi yang lebih baik.

# **DAFTAR PUSTAKA**

Gio, P. U., & Effendie, A. R. (2017). Belajar Bahasa Pemrograman R (Dilengkapi Cara Membuat Aplikasi Olah DataSederhana dengan R Shiny). 1-219.

Kusrini. (2019, Januari). OPTIMASI METODE NAIVE BAYES DENGAN FEATURE SELECTIONINFORMATION GAIN UNTUK PREDIKSI KETERLAMBATAN PEMBAYARAN SPPSEKOLAH. *Jurnal Ilmiah Sinus, 17*, 1-15.

Ratniasih, N. L. (2019, Januari). OPTIMASI DATA MINING MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN C4.5 UNTUK KLASIFIKASI KELULUSAN MAHASISWA. *Jurnal Teknologi Informasi dan Komputer, 1*, 1-7.

SANJIWANI, I. A. (2012). ANALISIS BIAYA PENDIDIKAN DAN DAMPAKNYA TERHADAP KUALITAS PROSES PEMBELAJARAN DAN ASPIRASI PENDIDIKAN SISWA (STUDI TENTANG PERSEPSI PARA SISWA SMA DWIJENDRA DENPASAR TAHUN PELAJARAN 2011/2012). 1-16.

Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *978*(979), 756.