

Seiring dengan perkembangan teknologi dan Kemajuan peradaban manusia yang sudah semakin berkembang pesat di segala bidang kehidupan. Ilmu pengetahuan dan teknologi dewasa ini menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat modern. Dengan memanfaatkan teknologi berbasis internet (*Online*), Sistem pelaporan keuangan yang berkaitan dengan pendapatan harian dari sebuah wajib pajak (WP) hotel, saat ini bisa dipantau secara realtime (Gambar 1.2). Berbeda dengan sistem pelaporan sebelumnya yaitu dengan metode *self assessment*, yaitu sistem pelaporan yang sepenuhnya dipercayakan oleh para wajib pajak untuk mengisi dengan jujur pendapatannya secara manual.



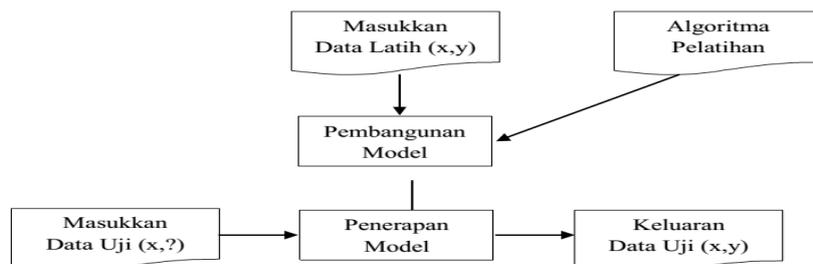
Gambar 1.2 Grafik Statistik Perolehan Pendapatan Hotel Selama 2016

Metode Klasifikasi Naive Bayes adalah salah satu Algoritma Klasifikasi yang populer, Performa Naive Bayes yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut). Asumsi keidependenan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keidependenan atribut tersebut dilanggar performa pengklasifikasian Naive Bayes cukup tinggi, hal ini dibuktikan pada berbagai penelitian empiris.

2. Landasan Teori

2.1 Konsep Klasifikasi

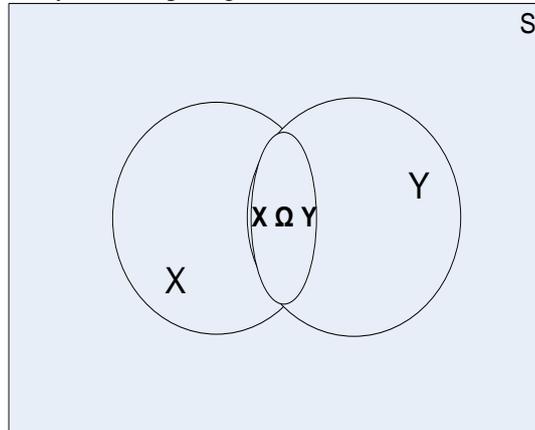
Klasifikasi merupakan pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu (1) pembangunan model sebagai prototipe untuk disimpan sebagai memori dan (2) penggunaan model tersebut untuk melakukan pengenalan/ klasifikasi/ prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpannya [4] Klasifikasi adalah metode data mining yang dapat digunakan untuk proses pencarian sekumpulan model (fungsi) yang dapat menjelaskan dan membedakan kelas-kelas data atau konsep, yang tujuannya supaya model tersebut dapat digunakan memprediksi objek kelas yang labelnya tidak diketahui atau dapat memprediksi kecenderungan data-data yang muncul di masa depan. Metode klasifikasi juga bertujuan untuk melakukan pemetaan data ke dalam kelas yang sudah didefinisikan sebelumnya berdasarkan pada nilai atribut data [9]. Proses klasifikasi tersebut seperti terlihat pada gambar 2.1



Gambar 2.1 Proses Klasifikasi

2.2 Metode Naïve Bayes

Metode Bayes merupakan pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi. Metode ini menggunakan probabilitas bersyarat sebagai dasarnya. Dalam ilmu statistik, probabilitas bersyarat dinyatakan seperti gambar



Gambar 2.2 Metode Bayesian

Probabilitas X di dalam Y adalah probabilitas interseksi X dan Y dari probabilitas Y, atau dengan bahasa lain $P(X|Y)$ adalah prosentase banyaknya X di dalam Y. Selain data seperti diatas metode *Naive Bayes* juga dapat menangani data berupa numerik. Untuk menangani data numerik metode *Naive Bayes* menggunakan asumsi distribusi normal. Rumus seperti di dibawah ini :

$$\begin{aligned} \mu &= \frac{1}{n} \sum_{i=1}^n x_i \\ \sigma &= \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2} \\ f(w) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(w-\mu)^2}{2\sigma^2}} \end{aligned} \quad (1)$$

2.2.1 Perhitungan Naïve Bayes Pada Penelitian

Misalnya ingin diketahui apakah suatu objek masuk dalam kategori layak untuk pembayaran pajak atau tidak dengan algoritma *Naive Bayes*. Untuk menetapkan suatu pembayaran pajak akan dipilih sebagai nilai untuk diklasifikasi kategori pajaknya, telah dihimpun 10 Bulan. Ada 4 atribut yang digunakan seperti terlihat pada tabel 2.1.

1. Nilai Pembayaran Pajak Perbulan (C1),
2. Selisih nilai yang disetorkan dengan yang ditargetkan (C2),
3. Ada atau tidaknya transaksi harian yang terlewat (C3), dan
4. Keputusan untuk memilih menetapkan pembayaran Pajak tersebut Wajar Atau Tidak (C4),

Tabel 2.1 Atribut penetapan wajib pajak.

Bulan ke-	Nilai Pembayaran pajak/bulan (C1)	Selisih Pajak Yang Disetor Dgn Target (C2)	Ada / Tidak Transaksi Yg Terlewat (C3)	Memilih kategori Wajar / Tidak (C4)
1	100	2	Tidak	Ya
2	200	1	Tidak	Ya
3	500	3	Tidak	Ya
4	600	20	Tidak	Tidak
5	550	8	Tidak	Tidak

6	250	25	Ada	Tidak
7	75	15	Ada	Tidak
8	80	10	Tidak	Ya
9	700	18	Ada	Tidak
10	180	8	Ada	Ya

a. Mean dan Deviasi Standart Untuk Atribut Pembayaran Pajak/Bulan (C1).

$$\mu_{ya} = \frac{100 + 200 + 500 + 80 + 180}{5} = 212$$

$$\mu_{tidak} = \frac{600 + 550 + 250 + 75 + 700}{5} = 435$$

$$\sigma^2_{ya} = \frac{(100 - 212)^2 + (200 - 212)^2 + (500 - 212)^2 + (80 - 212)^2 + (180 - 212)^2}{5 - 1} = 28520,015$$

$$\sigma_{ya} = \sqrt{28520,015} = 168,8787$$

$$\sigma^2_{tidak} = \frac{(600 - 435)^2 + (550 - 435)^2 + (250 - 435)^2 + (75 - 435)^2 + (700 - 435)^2}{5 - 1} = 68624,98$$

$$\sigma_{tidak} = \sqrt{68624,98} = 261,9637$$

b. Mean dan varian untuk atribut selisih yang disetorkan dengan target pajak (C2).

$$\mu_{ya} = \frac{2 + 1 + 3 + 10 + 8}{5} = 4,8$$

$$\mu_{tidak} = \frac{20 + 8 + 25 + 15 + 18}{5} = 17,2$$

$$\sigma^2_{ya} = \frac{(2 - 4,8)^2 + (1 - 4,8)^2 + (3 - 4,8)^2 + (10 - 4,8)^2 + (8 - 4,8)^2}{5 - 1} = 15,699821$$

$$\sigma_{ya} = \sqrt{15,699821} = 3,9623$$

$$\sigma^2_{tidak} = \frac{(20 - 17,2)^2 + (8 - 17,2)^2 + (25 - 17,2)^2 + (15 - 17,2)^2 + (18 - 17,2)^2}{5 - 1} = 39,700081$$

$$\sigma_{tidak} = \sqrt{39,700081} = 6,3008$$

Sedangkan untuk probabilitas atribut angkutan umum dan dipilih untuk perumahan terlihat pada tabel 2.2 dan tabel 2.3.

Tabel 2.2 Probabilitas Kemunculan Setiap Nilai Atribut ada tidaknya transaksi yang terlewati(C3).

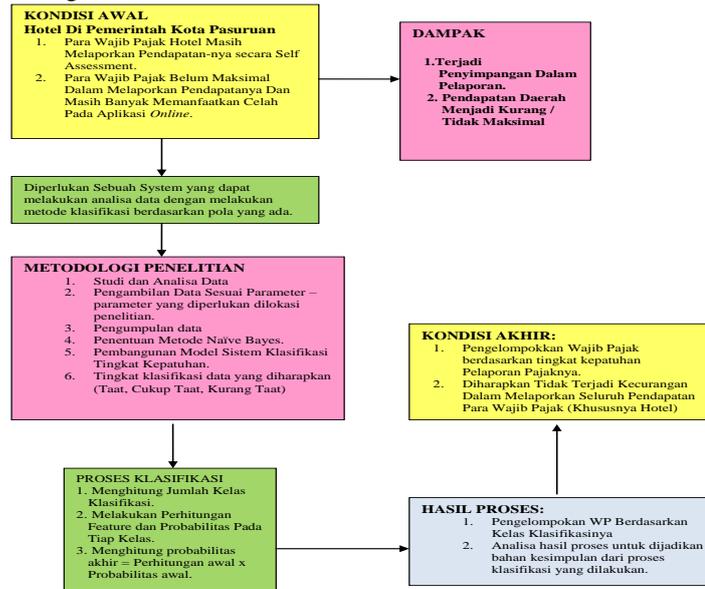
Transaksi yang terlewati	Jumlah Kejadian "Dipilih"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Ada	1	3	1/5	3/5
Tidak	4	2	4/5	2/5
Jumlah	5	5	1	1

Tabel 2.3 Probabilitas Kemunculan Setiap Nilai Atribut Dipilih Untuk Wajar / Tidak (C4).

Dipilih Untuk Penetapan wajar / Tidak	Jumlah Kejadian "Dipilih"		Probabilitas	
	Ya	Tidak	Ya	Tidak
Jumlah	5	5	1/2	1/2

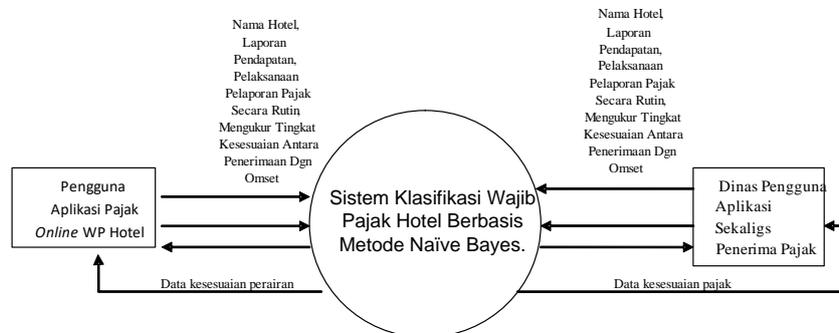
3. Metode Penelitian

Dalam penelitian, proses atau tahapan pengolahan data dari awal hingga akhir dijelaskan ke dalam metode penelitian sebagai berikut :



Gambar 3.1 Bagan Metode Penelitian

Berdasarkan Gambar 3.1 Penelitian ini dimulai dari kondisi awal yaitu : (1) pada kondisi ini Para Wajib Pajak Hotel masih melaporkan pendapatannya secara Self Assessment. Yaitu dengan melaporkan pendapatannya secara mandiri tanda ada pengawasan dari pihak luar atas dasar kepercayaan, (2) Para Wajib Pajak tidak maksimal dalam melaporkan Pendapatannya dengan memanfaatkan celah pada aplikasi Pajak Online. Berdasarkan kondisi awal inilah perlu dilakukan penelitian dalam menentukan tingkat kepatuhan Wajib Pajak. Apakah mereka termasuk dalam kategori 'Taata', 'Cukup Taata' dan 'Kurang Taata' dalam melaporkan Pendapatannya. Oleh karenanya dibutuhkan sebuah aplikasi Perangkat Lunak dalam menganalisis tingkat kepatuhan Wajib Pajak (Hotel) dalam melaporkan pendapatannya.



Gambar 3.3 Diagram Context

Gambar 3.3 adalah diagram context, diagram context merupakan diagram yang menggambarkan sistem secara keseluruhan. Diagram context terdiri dari dua elemen yaitu Wajib Pajak (Hotel) dan Dinas Pendapatan daerah kota Pasuruan.

4. HASIL DAN PEMBAHASAN

4.1 Pembahasan

Pengujian sistem merupakan pengujian dalam memasukkan data ke dalam form – form yang disediakan. Pada tahap ini pengujian dilakukan dengan menggunakan 108 (18bln x 6 Hotel) data training pada sistem. Berdasarkan hasil pengujian dari 108 data training diperoleh hasil bahwa terdapat 97 data yang sesuai dengan kelas sebenarnya.

Tabel 4.1 Hasil Pengujian Sistem

Data Perolehan Setoran Pajak Selama 18 Bulan Milik 6 WP

NO	ID WP	Bulan	Pajak Yang Disetorkan (dalam juta)	Target Perolehan (dalam Juta)	Jmlh Laporan Kosong
1	WP 1	Jan	56	99	4
2		Feb	45	90	8
3		Mar	68	95	5
4		Apr	50	94	7
5		Mei	61	96	3
6		Jun	53	97	2
7		Jul	65	92	7
8		Agus	50	93	0
9		Sept	43	95	1
10		Okt	47	91	8
11		Nop	64	93	6
...	
16		Apr	55	94	7
17		Mei	43	92	4
18		Jun	65	96	5
19	WP 2	Jan	56	97	1
20		Feb	45	98	7
...	
25		Jul	55	93	5
26		Agus	43	95	2
...
...
108	WP 6	Jun	45	98	7

Dari data tersebut akan diklasifikasikan dengan metode naïve bayes kedalam beberapa kelas, dimana kelas ini ditentukan berdasarkan nilai standart Probabilitas, adapun kelas dalam pembagian WP sebagai berikut :

1. Sangat Taat
2. Cukup taat
3. Kurang Taat

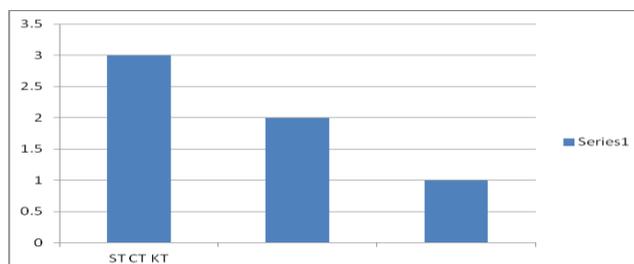
4.2 Hasil

Hasil akurasi yang didapat dari perhitungan menggunakan metod klasifikasi naïve bayes adalah :

Tabel 4.2 Klasifikasi Wajib Pajak

No	Status Klasifikasi	Pajak Yang Disetorkan	Target Perolehan	Laporan Koson
1	Sangat Taat	68	95	5
2	Cukup Taat	65	96	5
3	Kurang Taat	47	91	8

Tabel diatas merupakan hasil perhitungan klasifikasi dari data wajib pajak menggunakan Naïve bayes Classification. Berdasarkan data diatas, berikut grafik hasil klasifikasi dengan algoritma Naïve Bayes. Data menunjukkan bahwa 3 WP Termasuk dalam kategori Sangat Taat (ST), 2 WP dalam kategori Cukup Taat(CT) dan 1 WP dalam kategori Kurang Taat (KT).



Gambar 4.1 Diagram Hasil Klasifikasi

Dari Hasil klasifikasi diatas dapat diketahui bahwa :

1. 3 WP Termasuk Kedalam Kelompok 'Sangat Taat (ST).
2. 2 WP Termasuk Kedalam Kelompok 'Cukup Taat'(CT).
3. 1 WP Termasuk Kedalam Kelompok 'Kurang Taat'(KT).

5. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa algoritma Klasifikasi Naïve Bayes dapat digunakan sebagai salah satu metode untuk klasifikasi tingkat ketaatan Wajib Pajak, dalam menunjang pembangunan daerah Kota Pasuruan. Berdasarkan hasil pengukuran, hasil pengujian menunjukkan dengan akurasi sebesar 50% yaitu 3 Wajib Pajak Merupakan Golongan Sangat Taat Membayar Pajak.

DAFTAR PUSTAKA

- [1] Bustami. (2014), *Penerapan Algoritma Naive Bayes Untuk Mengklasifikasii Data Nasabah Asuransi*. Universitas Malikussaleh. Aceh Utara.
- [2] Fitriani, I. R. (2014), *PeningkataTaan Metode Naive Bayes Clasification Untuk Penentuan Tingkat Keganasan Kanker Payudara Menggunakan Particle Swarm Optimization*. Universitas Dian Nuswantoro. Jawa Tengah.
- [3] Jatmika, W. (2009), *Deteksi Kanker Payudara Menggunakan Ekstraksi Fitur Statistical Pada Citra Mammogram Berbasis Jaringan Syaraf Tiruan*. Universitas Muria. Kudus.
- [4] Prasetyo E. *Data Mining – Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset. 2014.
- [5] Y.S.Nugroho. 2009. *Data Mining Menggunakan Algoritma Naive Bayes untuk klasifikasi kelulusan Mahasiswa Universitas Diannuswantoro*. Program Studi Sistem Informasi, Fakultas Ilmu Komputer , Universitas Dianuswantoro.
- [6] Welborn T A dan Dhaliwa S S, "Preferred clinical measures of central obesity for predicting mortality" *European Journal of Clinical Nutrition* (2007) 61, 1373–1379; doi:10.1038/sj.ejcn.1602656; published online 14 February 2007.
- [7] Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untukKeperluan Bisnis*. Graha Ilmu. Yogyakarta.
- [8] Ridwan M, Suyono H, dan Sarosa M, "Penerapan Data Mining untuk evaluasi kinerja Akademik Mahasiswa menggunakan Algoritma Naive Bayes Classifier". *Jurnal EECCIS* Vol.7, No. 1, 2013
- [9] Jiawei Han and Micheline Kamber, "Data mining: Concepts and Techniques, Second Edition",Morgan Kaufmann Publishers, Marc 2006. ISBN 1-55860-901-6