



Artikel

## IDENTIFIKASI PENYALAHGUNAAN JALUR SMS MARKETING DENGAN KONTEN SMS PREMIUM PADA PT DIAN PRIMA JAYARAYA DENGAN ALGORITMA NAIVE BAYES

William Kohin Cikutra<sup>1</sup>, Susanto Hariyanto<sup>2</sup>, Ellysha D Kusuma<sup>3</sup>, Raditya Rimbawan<sup>4</sup>

<sup>1,2</sup> Universitas Buddhi Dharma, Teknik Informatika, Banten, Indonesia

### SUBMISSION TRACK

Received: August 25, 2021  
Final Revision: August 25, 2021  
Available Online:

### KATA KUNCI

SMS Broadcast, Naive Bayes, SMS Premium & Marketing, Data Mining, OTP

### KORSPONDENSI

Phone: 081310905201  
E-mail: willim666.wl@gmail.com

### ABSTRAK

PT Dian Prima Jayaraya adalah perusahaan layanan pengiriman SMS (Short Message Service) broadcast di Indonesia, saat ini operator di Indonesia memisahkan jalur SMS menjadi marketing dan premium. Jalur marketing digunakan untuk SMS promo/marketing dan jalur premium digunakan untuk SMS yang memiliki jangka waktu seperti OTP (One-Time Password). Namun pada kenyataannya, banyak perusahaan yang mengirimkan konten premium seperti OTP pada jalur marketing di karenakan kurang kesadaran atas konten yang dikirim. Untuk menghindari penyalahgunaan, maka akan dibangun model klasifikasi dengan Algoritma *Naives Bayes*, sehingga perusahaan dapat lebih awal mengidentifikasi konten yang salah jalur untuk menghindari denda dari operator. Hasil yang di peroleh dari penelitian ini adalah penggunaan data mining menggunakan algoritma Naive Bayes dapat mengidentifikasi penyalahgunaan jalur SMS marketing dengan konten SMS premium secara tepat dan hasil dari aplikasi dapat mengklasifikasikan suatu sms dan memilahnya menjadi kategori jalur SMS marketing atau SMS premium

### PENGANTAR

SMS merupakan salah satu komunikasi dengan teks melalui perangkat bergerak (*mobile device*). SMS merupakan salah satu media yang paling banyak digunakan sekarang ini, karena selain murah, prosesnya sangat cepat dan langsung pada tujuan. Banyaknya usaha yang menggunakan jasa

SMS dalam kelangsungan bisnisnya seperti perusahaan gojek dan grab yang menggunakan SMS dalam bidang promosi maupun keperluan aktivasi kode OTP dan lainnya. PT Dian Prima Jayaraya adalah perusahaan layanan pengiriman SMS broadcast di Indonesia, saat ini operator di Indonesia memisahkan jalur SMS menjadi marketing dan Premium. Jalur marketing

Digunakan untuk SMS Promo/Marketing. Semua SMS yang tidak berisikan OTP, bisa dikategorikan ke SMS ini dan boleh dikirimkan via Jalur Regular. Biaya per SMS lebih rendah dibandingkan Jalur Premium tetapi tidak ada status delivery report. Jalur premium digunakan untuk SMS OTP, yaitu SMS yang berisi angka atau huruf untuk tujuan verifikasi pengguna. Biasanya digunakan untuk sistem pendaftaran, login atau transaksi keuangan. Semua SMS via jalur premium akan mendapatkan status *delivery report*. Namun pada kenyataannya, banyak perusahaan yang menggunakan isi jalur premium sebagai jalur marketing di karenakan biaya yang lebih rendah. dan apabila operator mengetahui jika perusahaan tersebut menggunakan jalur premium sebagai jalur marketing maka perusahaan dan perusahaan layanan pengiriman SMS *broadcast* akan dikenakan sanksi oleh operator. Maka dari itu di butuhkan sebuah aplikasi yang mampu mengklarifikasi isi SMS itu apakah termasuk ke dalam jalur marketing atau premium. Dalam kasus kali ini akan menggunakan algoritma *naive bayes*. Alasan penggunaan algoritma *naive bayes* di bandingkan algoritma lain adalah karena *naive bayes* bisa di implementasi secara *realtime* di banding algoritma lain karena cara belajar *naive bayes eager learning* atau hanya menambah *word bag* saja.

## I. METODE

### 1.1. Data Mining

*Data mining* ada begitu banyak algoritma/metode/teknik penggalian atau pencarian pengetahuan atau informasi. Setiap algoritma/metode/teknik tersebut memiliki fungsi dan tujuan yang berbeda-beda. Berikut pengelompokan *data mining* berdasarkan fungsi dan tujuan :

#### a. Deskripsi

Deskripsi proses untuk menemukan/menyetujui pola yang sering muncul dan mengubah pola tersebut menjadi aturan yang dapat dipergunakan

untuk mempermudah suatu aktivitas. Contoh pada supermarket, pelanggan yang sering membeli produk A dan produk B secara bersamaan dan berulang, maka manajemen supermarket tersebut mengubah katalog barang dengan melengkapi produk A dan B di tempat yang sama atau berdekatan, sehingga pelanggan tidak mengalami kesulitan membeli kembali produk tersebut. Salah satu algoritma yang termasuk dalam deskripsi adalah algoritma *apriori*.

#### b. Klasifikasi

Pengelompokan berdasarkan hubungan antara variabel kriteria dengan variabel target. Contohnya pengelompokan dampak gempa bumi yaitu rusak berat, rusak berat dan tsunami, atau tidak berdampak. Dampak gempa bumi tersebut oleh beberapa faktor seperti skala, durasi, pusat gempa, jarak dari pantai dan kedalaman gempa bumi. Algoritma yang termasuk dalam klasifikasi adalah Cart, ID3, C4.5, J48, C5.0, *nearest neighbor*, *naive bayes*, dan lain-lain.

#### c. Asosiasi

Asosiasi adalah kumpulan, himpunan, persatuan, atau persekutuan. Dalam *data mining* proses asosiasi merupakan pencarian *attribute* yang muncul/selalu muncul dalam waktu bersamaan, seperti ketika dibeli produk A maka dibeli produk B, ketika dibeli produk B maka dibeli produk A, ketika dibeli produk A, B, maka dibeli produk C, dan seterusnya. Besar peluang muncul *attribute* secara bersamaan diukur dengan menggunakan nilai *confidence*. Algoritma yang termasuk dalam asosiasi adalah *association rule* [1].

### 1.2. Algoritma

Munculnya kata algoritma banyak di kaitkan dari buku Kitab *Aljabar Wal-muqabala* kala itu dibuat oleh Abu Ja'far Muhammad Ibnu Musa Al-Khuwarizmi. Orang barat tidak fasih dalam ejaan nama *Muslim* membaca Al-Khuwarjimi menjadi *Algorism* [2].

### 1.3. Algoritma *Naïve Bayes*

Ilmuan Inggris Thomas Bayes menyatakan bahwa sebuah peluang terjadinya suatu peristiwa di masa depan dapat di prediksi dilihat pengalaman sebelumnya yang biasanya di *compile* dalam sebuah data yang pernah terjadi yang biasa di sebut Teorema Bayes. Salah satu Implementasi dari Teorema Bayes adalah *Naive Bayes Classifier*. *Naive Bayes* di sebut *Naive* karena memiliki asumsi yang sangat kuat (naif) dari setiap peristiwa di lihat dari data data yang sudah ada [3].

### 1.4. SMS OTP

Pesan teks telepon seluler adalah mode OTP yang digunakan secara luas, dan yang paling tidak aman dan paling tidak dapat diandalkan. Kerentanan keamanan adalah:

1. Porting nomor  
Ada banyak cara untuk membajak pesan nomor telepon pengguna; ini adalah serangan yang umum dan berhasil.
2. pengalihan SS7  
Protokol jaringan seluler mengizinkan operator yang tidak bertanggung jawab di mana pun di dunia untuk memasukkan perintah yang mengubah rute (sehingga mencegat) lalu lintas SMS, suara, dan data seluler untuk pelanggan mana pun. Serangan publik, dengan izin (tapi tanpa bantuan) terhadap korban terkenal telah ditunjukkan.
3. mikro-sel berbahaya, dan radio mengendus  
Ditentukan Perangkat Lunak dan SMS AT&T 3G 1137 AM 100% Pesan Kode otentikasi Dompot Saya G430F Radio (SDR) dijual dengan harga di bawah \$ 10 di eBay, dan perangkat lunak sumber terbuka gratis mengubahnya menjadi *sniffer* SMS lokal (dan jarak jauh).
4. Enkripsi lemah, atau tidak ada  
Enkripsi jaringan seluler lemah, membutuhkan (tergantung pada generasi) antara 2 jam hingga kurang dari 1 detik untuk memecahkan pada satu PC [5]. Serangan lalu lintas sel yang dimodifikasi

yang menonaktifkan enkripsi sepenuhnya adalah relatif mudah untuk dipasang, biasanya ditemukan aktif di kota-kota, dan berlanjut tanpa terdeteksi pada semua kecuali handset sel aman yang dirancang khusus.

5. Berbagi *iMessage*  
Pesan SMS -OTP sering kali didistribusikan ke berbagai perangkat pemegang akun dan muncul di beberapa layar pengguna sekaligus. Ini selanjutnya menjadikan SMS sebagai pencurian karena penyusup dengan akses akun *cloud* pengguna dapat mendaftarkan perangkat mereka sendiri di akun ini untuk menerimanya.
6. Situasi penurunan peringkat  
Banyak organisasi menyarankan pengguna menonaktifkan perjalanan SMS -OTP mereka. Keputusan berisiko bagi sebagian besar pengguna karena ini adalah waktu yang paling mereka butuhkan 2FA!
7. Perlindungan lokal yang rendah  
Banyak handset menampilkan pesan di layar kunci, tanpa perlindungan agar tidak diamati oleh pihak ketiga yang jahat.
8. Rekayasa Sosial melawan pihak ketiga  
Banyak pekerja layanan pelanggan di industri komunikasi dapat berhasil diyakinkan melalui penipuan atau penyuaian untuk memengaruhi porting SIM atau penyesuaian lain untuk mengirimkan SMS -OTP kepada penyerang.
9. Penggantian nomor SMS -OTP yang bermaksud jahat di situs web.  
Perangkat lunak atau operator yang menjalankan situs web dapat tertipu untuk mengubah nomor telepon tujuan pengiriman kode. Ada serangan yang melibatkan kombinasi manipulasi psikologis terhadap beberapa pihak ketiga yang memberikan akses langsung kepada musuh untuk mengubah nomor telepon SMS -OTP itu sendiri secara online.
10. Perangkat Lunak Perusak Seluler  
Sistem operasi iOS dan Android keduanya menyertakan pengaturan "izin" yang memungkinkan Aplikasi Seluler

membaca dan mengganggu SMS . Ada aplikasi berbahaya yang meneruskan SMS ke penyerang dan menyembunyikan tampilannya kepada pengguna.

11. Kepercayaan pihak ketiga SMS -OTP itu sendiri melakukan perjalanan melalui banyak jaringan yang berbeda sebelum mencapai pengguna; setiap gangguan kepercayaan di sepanjang jalan memberi peluang berbahaya.
12. Sebagian besar kerentanan Perangkat Keras OTP juga berlaku untuk SMS -OTP termasuk: MitM; tidak ada keamanan saluran; *spoofing*; transportasi saluran tunggal; *keylogger*, phishing, *malware*, manipulasi psikologis; tidak ada utilitas untuk menandatangani transaksi; tidak ada perlindungan *malware* (berbeda dari *malware* seluler), resistansi rendah terhadap penyalahgunaan oleh teman, keluarga, atau rekan kerja; penipuan yang disengaja; tidak ada non-repudiation; tidak ada autentikasi bersama [4].

### 1.5. SMS Regular

SMS Regular adalah sebuah layanan yang di berikan oleh sebuah vendor yang mampu mengirimkan pesan singkat yang biasanya iklan atau notifikasi langsung ke banyak nomor dengan nama pengirim diganti menjadi *Alphanumeric* yang biasanya nama perusahaan atau nama Brand yang sudah lolos izin dari pihak operator untuk di gunakan dalam pengiriman pesan. Dalam periklanan harus mempromosikan Produk atau jasa yang di miliki perusahaan yang berkerja sama dengan vendor. [5].

### 1.6. RapidMiner

Banyak platform yang di gunakan untuk data science salah satunya *RapidMiner*. *RapidMiner* yang di buat oleh perusahaan yang sama namanya *RapidMiner*

merupakan sebuah platform yang memberikan sebuah lingkungan untuk memproses data, data mining , machine learning dan lain lainnya yang berhubungan dengan data science. Software ini didevelop menggunakan bahasa pemrograman java serta *core system* dari *rapidminer* di *opensource* ke *community*. Penggunaan *RapidMiner* menggunakan GUI (*Graphical User Interface*) yang memudahkan *user* dalam menggunakan aplikasi dengan hanya menghubungkan pipeline dari satu operator ke operator selanjutnya yang akan di simpan dalam bentuk format XML (*Extensible Markup Language*) [6].

### 1.7. Database

Basis data terdiri dari dua kata yaitu basis dan data. Basis dapat diartikan sebagai suatu markas atau gudang, tempat bersarang atau tempat berkumpul. Data dapat diartikan merupakan representasi dari fakta dunia yang mewakili suatu obyek (manusia, barang, peristiwa, keadaan dsb) yang direkam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi atau kombinasinya [7].

### 1.8. Hypertext Preprocessor

PHP (*Hypertext Preprocessor*) adalah bahasa *script* pemrograman yang dapat ditanamkan atau disisipkan ke dalam HTML, PHP banyak dipakai untuk memrogram situs *web* dinamis. PHP dapat digunakan untuk membangun sebuah CMS [8]. -PHP merupakan bahasa pemrograman untuk membuat aplikasi berbasis *website*, PHP juga sering dikenal dengan *hypertext preprocessor* yang merupakan sebuah bahasa *scripting* tingkat tinggi yang dipasang pada dokumen HTML [9]. PHP merupakan singkatan dari PHP *hypertext preprocessor* yang digunakan sebagai bahasa *script server side* dalam pengembangan *web* yang disisipkan pada dokumen HTML. PHP bisa berinteraksi dengan *database*, *file* dan *folder*, sehingga membuat PHP bisa menampilkan konten yang dinamis dari sebuah *website*.

Karena penggunaan PHP memungkinkan *web* dapat dibuat dinamis, *maintenance* situs *web* tersebut menjadi lebih mudah dan efisien. PHP merupakan *software open source* yang disebar dan dilisensikan secara gratis serta dapat di *download* secara bebas dari situs resminya <http://www.php.net>. PHP ditulis menggunakan bahasa C [10].

### 1.9. MySQL

MySQL adalah sistem yang berguna untuk melakukan proses pengaturan koleksi-koleksi struktur data (*database*) baik yang meliputi proses pembuatan atau proses pengelolaan *database*. MySQL ini bersifat *open source*, artinya setiap orang dimungkinkan untuk menggunakan dan memodifikasinya. Program ini dapat diperoleh secara gratis pada alamat : <http://www.mysql.com>. Sebenarnya MySQL ini suatu produk yang berjalan pada *linux*, tetapi seiring berjalannya waktu dan pengetahuan maka para pengembang kemudian merilis MySQL yang bisa diakses melalui *windows* [11].

## II. PERANCANGAN

### 2.1. Business Understanding

Dalam tahap ini adalah memahami masalah tujuan dan kebutuhan dari sudut pandang bisnis, dibutuhkan pemahaman dari kegiatan *data mining* yang akan dilakukan. Dalam hal ini adalah identifikasi masalah yang terdapat pada PT Dian Prima Jayaraya ini adalah kurangnya pengetahuan dalam mengetahui SMS premium atau tidak, adanya perbedaan harga antara SMS premium atau tidak dan perlu adanya tindakan preventif dalam menanggulangi kesalahan penggunaan jalur SMS . Dalam memahami masalah yang ada pada objek penelitian sangatlah penting untuk dapat menentukan strategi awal yang lebih tepat sasaran dan dapat mencapai tujuan. Mengambil informasi yang baru merupakan langkah yang dapat dilakukan dalam kegiatan ini. Untuk mendapat informasi yang baru, dengan memanfaatkan data dari isi SMS

premium dan marketing yang telah tersimpan di dalam *database*. Isi SMS premium dan marketing yang tersimpan akan digali kembali menggunakan teknik *data mining*, misalnya untuk mengklasifikasi data nasabah. Ini merupakan strategi awal yang dilakukan sebelum nantinya akan dilakukan perancangan aplikasi dengan menggunakan *Naïve Bayes* sebagai algoritmanya.

### 2.2. Data Understanding

Dalam tahap ini adalah melakukan pengumpulan data awal yang akan digunakan atau diolah nantinya. Data yang akan digunakan adalah data transaksi penjualan produk yang tersimpan di dalam *database* yang terjadi dalam kurung waktu 1 minggu operasional PT Dian Prima Jayaraya (1 Mei 2021 - 8 Mei 2021). Data transaksi diperoleh dari *Lead Developer* PT Dian Prima Jayaraya dan data transaksi SMS yang digunakan adalah data *primer*.

### 2.3. Data Preparation

Dari pemahaman akan data pada fase *data understanding* selanjutnya data-data yang telah terkumpul, menggunakan data transaksi yang sudah didapat kedalam bentuk file *excel* agar memudahkan dalam pengolahan data, dengan menghilangkan beberapa atribut yang tidak dipakai. Dari data awal berupa data transaksi SMS sampai pada membangun *dataset*, dari *dataset* ini yang akan diterapkan ke dalam alat permodelan, selanjutnya akan melakukan proses *data mining*.

**Tabel 2.1 Data Sampel 10 Data Transaksi SMS**

No	Data	Type
1	Nomor OTP Anda adalah 447655	OTP
2	Terima Kasih Atas Pembelian Paket Data 200K	MKT
3	Kini solusi asuransi hadir di O*** Mobile! Temukan 7 pilihan asuransi jiwa dan lainnya	MKT
4	Plg Yth segera gunakan Kuoa kemendikbud Anda utk cek kuota dan beli paket.	MKT
5	Your verification code is 15789.	OTP

**Tabel 2.2 Data Transaksi SMS Setelah Melakukan Proses *Stopword Removal***

Type	Word
MKT	500mb
MKT	550/hr.
MKT	Akashic
MKT	Aktifkan
MKT	aneka
MKT	asuransi
MKT	asuransi
MKT	Atas
MKT	Ba*****
MKT	beli
MKT	cek
MKT	CS
MKT	cs.su*****@g***l.com
MKT	Dapatkan
MKT	Data
MKT	dengan
MKT	Dyah

#### 2.4. Fase *Modeling*

Fase *modeling* merupakan tahapan metode dalam proses *data mining* yang diterapkan pada data PT Dian Prima Jayaraya. Pemodelan ini dilakukan dengan menggunakan metode *Naïve Bayes*. Metode ini memiliki satu tahapan pengerjaan yaitu mencari nilai *confidence*. Dalam tahap pencarian nilai *confidence* algoritma yang akan digunakan adalah *Naïve Bayes*. Tahapan dari proses algoritma *Naive Bayes* adalah:

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

#### 2.5. Fase *Evaluation*

Pada Tahap ini *dataset* yang sudah disiapkan kemudian akan diolah menggunakan aplikasi *rapidminer* 9.8. Tahap ini dimulai menyiapkan data sampel yang sudah ditentukan untuk diolah dan dianalisis. Evaluasi disini akan dilakukan secara

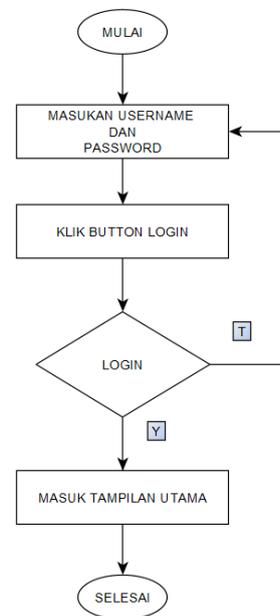
mendalam dengan tujuan untuk menyesuaikan hasil pada tahap *modeling* agar sesuai dengan sasaran yang ingin dicapai dalam tahap *Business Understanding*.

#### 2.6. Fase *Deployment*

Tahapan di mana hasil dari implementasi model dapat dijadikan sebagai laporan yang dibentuk menjadi media yang mudah di mengerti oleh kaum awam seperti media grafik atau table yang mampu menjelaskan hasil jalannya suatu proses data mining sehingga kaum awam mampu mendapat suatu pengetahuan dari data yang sudah di proses.

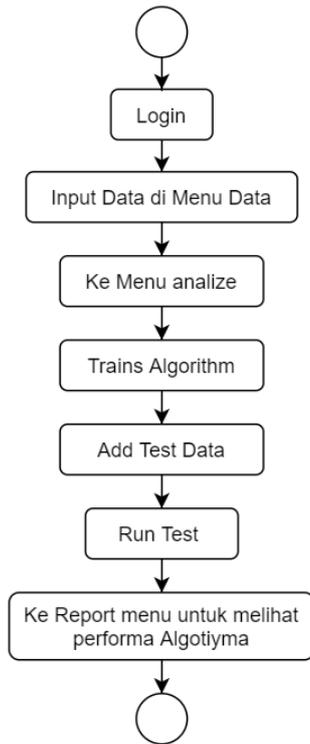
#### 2.7. Perancangan Flowchart Aplikasi

##### 1. Flowchart Login



**Gambar 2.1 Perancangan Flowchart Login Aplikasi**

##### 2. Flowchart Aplikasi



**Gambar 2.2 Perancangan Flowchart Aplikasi**

### III. PEMBAHASAN

#### 3.1. Tampilan Program

##### 1. Halaman Login

Halaman login adalah halaman saat *user/admin* akan melakukan login aplikasi.



**Gambar 3.1 Halaman Tampilan Login**

##### 2. Halaman Registrasi

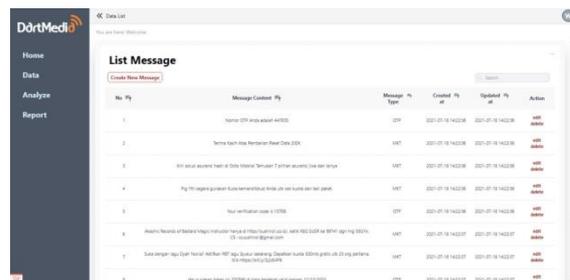
Halaman registrasi adalah halaman saat *user* ingin melakukan registrasi akun.



**Gambar 3.2 Halaman Registrasi**

##### 3. Halaman List Message

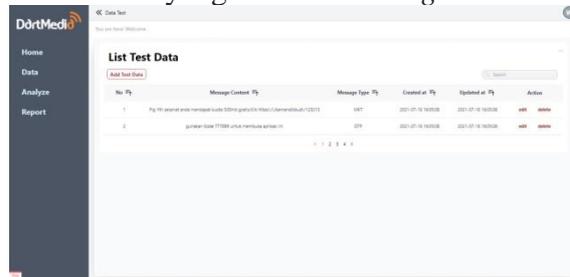
Halaman *list message* adalah halaman untuk melihat daftar data SMS .



**Gambar 3.4 Halaman List Message**

##### 4. Halaman List Test Data

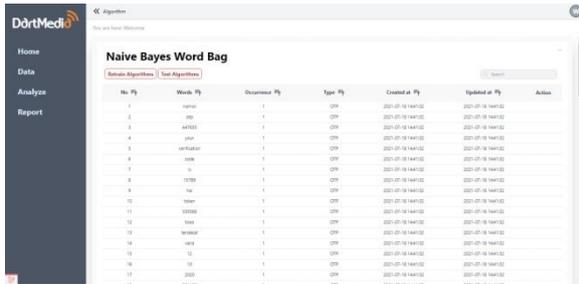
Halaman *list test data* adalah halaman untuk melihat data yang sudah di *testing*.



**Gambar 3.5 Halaman List Test Data**

##### 5. Halaman Naïve Bayes Word Bag

Halaman *Naïve Bayes word bag* adalah halaman untuk melihat kata-kata yang telah di kategorikan menjadi otp atau mkt dengan algoritma *naïve bayes*.



Gambar 3.6 Naive Bayes Word Bag

### 3.2. Pengujian Black Box Testing

Tabel 3.1 Black Box Testing Identifikasi penyalahgunaan jalur SMS marketing dengan konten SMS premium

No	Pengujian	Hasil Yang di Harapkan	Hasil Pengujian	Keterangan
1	Pengujian halaman login aplikasi	Aplikasi dapat melakukan proses login dan masuk ke menu selanjutnya	Sesuai	Valid
2	Pengujian halaman registrasi	Aplikasi dapat melakukan proses penambahan akun baru.	Sesuai	Valid
3	Pengujian halaman dashboard	Aplikasi dapat menampilkan halaman dashboard.	Sesuai	Valid
4	Pengujian halaman list message	Aplikasi dapat menampilkan halaman list message.	Sesuai	Valid
5	Pengujian halaman list test data	Aplikasi dapat menampilkan halaman	Sesuai	Valid

		<i>list test data.</i>		
6	Pengujian halaman <i>naive bayes word bag</i>	Aplikasi dapat menampilkan halaman <i>naive bayes word bag.</i>	Sesuai	Valid
7	Pengujian halaman <i>report performance</i>	Aplikasi dapat menampilkan halaman <i>report performance.</i>	Sesuai	Valid

### 3.3 Pembahasan Metode & Algoritma

Data transaksi yang akan di gunakan menghasilkan model yang dapat di gunakan dalam mengklasifikasi SMS OTP dan Marketing. Hasil dari algoritma akan di gunakan dalam filter SMS di service PT Dian Prima Jaya Raya.

Dalam artikel ilmiah ini akan menggunakan algoritma Naive Bayes. Dalam pengumpulan data disini akan menggunakan metode observasi, data transaksi pengiriman SMS yang diperoleh dari IT supervisor PT Dian Prima Jaya Raya yang berupa data kuantitatif. Dalam algoritma ini akan mencari Bag Of Word yang akan di gunakan untuk mengklasifikasi transaksi SMS .Dalam menentukan Bag Of Word akan di bagi menjadi 2 kategori yang dari data samples sering keluar di kategori SMS OTP dan kategori SMS MKT. Proses perhitungan manual menggunakan Data sampel sebanyak 12 data SMS yang di ambil dari *dataset* dengan 10 yang akan di gunakan untuk training dan 2 data yang akan digunakan untuk testing. *Dataset* diambil dari data transaksi pengiriman SMS .Setiap data berisi contoh isi SMS yang di kirim ke penerima pesan dan label SMS yang merupakan OTP atau MKT untuk SMS Training. Contoh 1 baris data berisi isi pesan “Gunakan 554488 di WeChat untuk login di akun anda. Jangan

sebarikan Kode ini”.Berikut 10 *sample* data transaksi yang akan di gunakan.

**Tabel 3.2 10 Sample Data Transaksi**

NO	Isi Pesan	Label
1	Nomor OTP Anda adalah 447655	OTP
2	Terima Kasih Atas Pembelian Paket Data 200K	MKT
3	Kini solusi asuransi hadir di Octo Mobile! Temukan 7 pilihan asuransi jiwa dan lainnya	MKT
4	Plg Yth segera gunakan Kuoqa kemendikbud Anda utk cek kuota dan beli paket.	MKT
5	Your verification code is 15789.	OTP

Tahap Selanjutnya adalah dengan melakukan Tokenisasi data item set dengan cara memisahkan kalimat pesan menjadi kata kata yang di pisahkan dengan spasi dan penghilangan special character.

**Tabel 3.3 Tokenisasi Data Itemset**

NO	Tokenisasi(dalam bentuk array)	Label
1	["nomor","otp","anda","adalah","447655"]	OTP
2	["terima","kasih","atas","pembelian","paket","data","200k"]	MKT
3	["kini","solusi","asuransi","hadir","di","octo","mobile","temukan","7","pilihan","asuransi","jiwa","dan","lainya"]	MKT
4	["plg","yth","segera","gunakan","kuoqa","kemendikbud","anda","utk","cek","kuota","dan","beli","paket"]	MKT
5	["your","verification","code","is","15789"]	OTP

Tahap Selanjutnya adalah dengan melakukan filter *stop word* data item set dengan cara menghilangkan kata kata yang tidak dibutuhkan seperti konjungsi,kata memulai pertanyaan dan kata lain subject.

**Tabel 3.4 Filter Stop Word**

NO	Filter Stop Word(dalam bentuk array)	Label
1	[nomor","otp"]	OTP
2	["terima","kasih","pembelian","paket","data","k"]	MKT
3	["solusi","asuransi","hadir","octo","mobile","temukan","pilihan","asuransi","jiwa","lainya"]	MKT

4	["plg","yth","kuota","kemendikbud","utk","cek","kuota","beli","paket"]	MKT
5	["your","verification","code","is"]	OTP

Tahap Selanjutnya adalah dengan melakukan filter *stemming* data item set dengan cara mengembalikan kata kata di dalam data menjadi kata dasar seperti gratisan menjadi gratis dan sebagainya.

**Tabel 3.5 Filter Stemming Dalam Bentuk Array**

NO	Filter stemming(dalam bentuk array)	Label
1	["nomor","otp"]	OTP
2	["terima","kasih","beli","paket","data","k"]	MKT
3	["solusi","asuransi","hadir","octo","mobile","temu","pilih","asuransi","jiwa","lai"]	MKT
4	["plg","yth","kuota","kemendikbud","utk","cek","kuota","beli","paket"]	MKT
5	["your","verification","code","is"]	OTP

Tahap Selanjutnya adalah membuat *Histogram* yang akan di gunakan oleh algoritma *Naïve Bayes*.

**Tabel 3.6 Histogram Menggunakan Algoritma Naïve Bayes**

NO	Histogram(dalam bentuk array object)	Label
1	{"nomor":1,"otp":1,"your":1,"verification":1,"code":1,"is":1,"hai":1,"token":1,"toko":1,"dekat":1,"valid":1,"wechat":1,"log":1,"akun":1,"bar":1,"kode":1}	OTP
2	{"terima":1,"kasih":1,"beli":2,"paket":2,"data":1,"k":1,"solusi":1,"asuransi":2,"hadir":1,"octo":1,"mobile":1,"temu":1,"pilih":1,"jiwa":1,"lai":1,"plg":1,"yth":1,"kuota":3,"kemendikbud":1,"utk":2,"cek":1,"akashic":1,"records":1,"of":1,"bastard":1,"magic":1,"instructor":1,"http":1,"sushiroll":2,"co":1,"id":1,"tik":1,"reg":1,"susur":1,"dgn":1,"hrg":1,"hr":1,"cs":2,"gmail":1,"com":1,"suka":1,"lagu":2,"dyah":1,"novia":1,"aktif":1,"r	MKT

	bt":1,"syukur":1,"dapat":1,"mb":1," gratis":1,"org":1,"klik":1,"https":1," bit":1,"ly":1,"j":1,"dmp":1,"pisang" :1,"kipas":1,"aneka":1,"topping":1," yuk":1,"screen":1,"shot":1,"SMS" ":1,"kirim":1,"ig":1}	
--	--	--

Tahap Selanjutnya adalah membuat *Probabilitas* per kata yang ada di SMS Testing yang akan di gunakan untuk memprediksi SMS testing.

SMS Sebelum Tokenisasi dan StopWord Filter

**Tabel 3.7 Probabilitas SMS Testing Sebelum Tokenisasi dan Stop Word Filter**

NO	Isi Pesan	Label
1	gunakan Kode 777899 untuk membuka aplikasi ini	?
2	Plg Yth selamat anda mendapat kuota 500mb gratis.Klik <a href="https://kemendikbud/123213">https://kemendikbud/123213</a>	?

SMS Sesudah Tokenisasi dan StopWord Filter

**Tabel 3.8 Probabilitas SMS Testing Sesudah Tokenisasi dan Stop Word Filter**

NO	Isi Pesan	Label
1	["kode","buka","aplikasi"]	?
2	["plg","yth","selamat","kuota","mb","gratis","klik","https","kemendikbud"]	?

Dalam Perhitungan naïve bayes kita harus membuat *Histogram* dari semua kata yang muncul di dalam SMS OTP dan MKT yang sudah di buat di atas.Dengan *Histogram* yang sudah dibuat kita dapat menentukan probabilitas kata per kata di dalam SMS OTP atau MK-. Contohnya dari isi test tersebut probabilitas keluarannya kata kode,buka dan aplikasi di dalam SMS OTP adalah

$$P(Kode|Otp) = 1/16 = 0.0625;$$

$$P(Buka|Otp) = 0/16;$$

$$P(Aplikasi|Otp) = 0/16;$$

Karena setiap sesuatu yang di bagi dengan 0 akan menjadi 0 maka setiap kata yang di cek akan di tambahkan 1 yang biasanya di sebut dengan *smoothing* menjadi :

$$P(Kode|Otp) = 2/100 = 0.02;$$

$$P(Buka|Otp) = 1/100 = 0.01;$$

$$P(Aplikasi|Otp) = 1/100 = 0.01;$$

Sekarang kita mulai dengan berangapan semua probabilitas pesan apapun itu merupakan OTP maka  $P(OTP) = 4/10 = 0.4$  maka :

$$P(OTP) * P(Kode|Otp) * P(Buka|Otp) * P(Aplikasi|Otp) = 0.4 * 0.02 * 0.01^2 = 0.0000008$$

Dan dari isi test tersebut probabilitas keluarannya kata kode,buka dan aplikasi di dalam SMS MKT adalah

$$P(Kode|MKT) = 0/76;$$

$$P(Buka|MKT) = 0/76;$$

$$P(Aplikasi|MKT) = 0/76;$$

Karena setiap sesuatu yang di bagi dengan 0 akan menjadi 0 maka setiap kata yang di cek akan di tambahkan 1 yang biasanya di sebut dengan *smoothing* menjadi :

$$P(Kode|MKT) = 1/160 = 0.00625;$$

$$P(Buka|MKT) = 1/160 = 0.00625;$$

$$P(Aplikasi|MKT) = 1/160 = 0.00625;$$

Sekarang kita mulai dengan berangapan semua probabilitas pesan apapun itu merupakan OTP maka  $P(MKT) = 6/10 = 0.6$  Maka

$$P(OTP) * P(Kode|MKT) * P(Buka|MKT) * P(Aplikasi|MKT) = 0.6 * 0.00625 * 0.00625 * 0.00625 = 0.000000146484375$$

Maka karena  $0.0000008 > 0.000000146484375$  karena probabilitas

OTP lebih besar di banding MKT maka ini dapat di simpulkan bahwa SMS ini merupakan SMS OTP .

**Tabel 3.9 Probabilitas SMS Testing Per Kata**

NO	Probabilitas Per Kata	Label
1	{ "kode": { "OTP ": 0.02, "MKT": 0.00625 }, "buka": { "OTP ": 0.01, "MKT": 0.00625 },                 }	?

	"aplikasi": { "OTP ": 0.01, "MKT": 0.00625 }	
2	{ "plg": { "OTP ": 0.01, "MKT": 0.0125 }, "yth": { "OTP ": 0.01, "MKT": 0.0125 }, "selamat": { "OTP ": 0.01, "MKT": 0.00625 }, "kuota": { "OTP ": 0.01, "MKT": 0.01875 }, "mb": { "OTP ": 0.01, "MKT": 0.0125 }, "gratis": { "OTP ": 0.01, "MKT": 0.0125 }, "klik": { "OTP ": 0.01, "MKT": 0.0125 }, "https": { "OTP ": 0.01, "MKT": 0.0125 }, "kemendikbud": { "OTP ": 0.01, "MKT": 0.0125 } }	?

Tahap Selanjutnya adalah membuat *Probabilitas* per SMS dengan mengkalikan *Probabilitas* perkata dengan *Probabilitas* kemungkinan SMS MKT atau OTP yang dapat dari data training dan memilih *probabilitas* terbesar sebagai prediksi.

**Tabel 3.10 Probabilitas SMS Testing Per Kalimat**

NO	Probabilitas per kalimat	Label
1	{"OTP ":8.0000000000000002e-7,"MKT":1.4648437500000004e-7}	OTP
2	{"OTP ":4.00000000000000013e-19,"MKT":3.352761268615725e-18}	MKT

**Tabel 3.11 Hasil Akhir SMS Testing**

No	Isi Pesan	Label
1	gunakan Kode 777899 untuk membuka aplikasi ini	OTP
2	Plg Yth selamat anda mendapat kuota 500mb gratis.Klik https://kemendikbud/123213	MKT

Dalam Perhitungan naïve bayes kita harus membuat *Histogram* dari semua kata yang muncul di dalam SMS OTP dan MKT yang sudah di buat di atas.Dengan *Histogram* yang sudah dibuat kita dapat menentukan

probabilitas kata per kata di dalam SMS OTP atau MK-. Contohnya dari isi test tersebut probabilitas keluarnya kata kode,buka dan aplikasi di dalam SMS OTP adalah

$$P(Kode|Otp) = 1/16 = 0.0625;$$

$$P(Buka|Otp) = 0/16;$$

$$P(Aplikasi|Otp) = 0/16;$$

Karena setiap sesuatu yang di bagi dengan 0 akan menjadi 0 maka setiap kata yang di cek akan di tambahkan 1 yang biasanya di sebut dengan *smoothing* menjadi :

$$P(Kode|Otp) = 2/100 = 0.02;$$

$$P(Buka|Otp) = 1/100 = 0.01;$$

$$P(Aplikasi|Otp) = 1/100 = 0.01;$$

Sekarang kita mulai dengan berangapan semua probabilitas pesan apapun itu merupakan OTP maka  $P(OTP) = 4/10 = 0.4$  maka :

$$P(OTP) * P(Kode|Otp) * P(Buka|Otp) * P(Aplikasi|Otp) = 0.4 * 0.02 * 0.01^2 = 0.0000008$$

Dan dari isi test tersebut probabilitas keluarnya kata kode,buka dan aplikasi di dalam SMS MKT adalah

$$P(Kode|MKT) = 0/76;$$

$$P(Buka|MKT) = 0/76;$$

$$P(Aplikasi|MKT) = 0/76;$$

Karena setiap sesuatu yang di bagi dengan 0 akan menjadi 0 maka setiap kata yang di cek akan di tambahkan 1 yang biasanya di sebut dengan *smoothing* menjadi :

$$P(Kode|MKT) = 1/160 = 0.00625;$$

$$P(Buka|MKT) = 1/160 = 0.00625;$$

$$P(Aplikasi|MKT) = 1/160 = 0.00625;$$

Sekarang kita mulai dengan berangapan semua probabilitas pesan apapun itu merupakan OTP maka  $P(MKT) = 6/10 = 0.6$  Maka

$$P(OTP) * P(Kode|MKT) * P(Buka|MKT) * P(Aplikasi|MKT) = 0.6 * 0.00625 * 0.00625 * 0.00625 = 0.000000146484375$$

Maka karena  $0.0000008 > 0.000000146484375$  karena probabilitas OTP lebih besar di banding MKT maka ini dapat di simpulkan bahwa SMS ini merupakan SMS OTP .

**Tabel 3.12 Probabilitas SMS Testing Per Kata**

NO	Probabilitas Per Kata	Label
1	{ "kode": { "OTP ": 0.02, "MKT": 0.00625 }, "buka": { "OTP ": 0.01, "MKT": 0.00625 }, "aplikasi": { "OTP ": 0.01, "MKT": 0.00625 } }	?
2	{ "plg": { "OTP ": 0.01, "MKT": 0.0125 }, "yth": { "OTP ": 0.01, "MKT": 0.0125 }, "selamat": { "OTP ": 0.01, "MKT": 0.00625 }, "kuota": { "OTP ": 0.01, "MKT": 0.01875 }, "mb": { "OTP ": 0.01, "MKT": 0.0125 }, "gratis": { "OTP ": 0.01, "MKT": 0.0125 }, "klik": { "OTP ": 0.01, "MKT": 0.0125 }, "https": { "OTP ": 0.01, "MKT": 0.0125 }, "kemendikbud": { "OTP ": 0.01, "MKT": 0.0125 } }	?

2	Plg Yth selamat anda mendapat kuota 500mb gratis.Klik <a href="https://kemendikbud/123213">https://kemendikbud/123213</a>	MKT
---	---	-----

#### IV. KESIMPULAN

Berdasarkan hasil dengan memanfaatkan bidang ilmu menggunakan *data mining* untuk identifikasi penyalahgunaan jalur SMS *marketing* dengan konten SMS *premium*, maka dapat diambil kesimpulan sebagai berikut :

1. Diketahui bahwa penggunaan *data mining* menggunakan algoritma *Naive Bayes* dapat mengidentifikasi penyalahgunaan jalur SMS *marketing* dengan konten SMS *premium* secara tepat.
2. Hasil dari aplikasi dapat mengklasifikasikan suatu SMS dan memilahnya menjadi kategori jalur SMS *marketing* atau SMS *premium*.

Tahap Selanjutnya adalah membuat *Probabilitas* per SMS dengan mengkalikan *Probabilitas* perkata dengan *Probabilitas* kemungkinan SMS MKT atau OTP yang dapat dari data training dan memilih *probabilitas* terbesar sebagai prediksi.

**Tabel 3.13 Probabilitas SMS Testing Per Kalimat**

NO	Probabilitas per kalimat	Label
1	{"OTP ":8.000000000000002e-7,"MKT":1.4648437500000004e-7}	OTP
2	{"OTP ":4.0000000000000013e-19,"MKT":3.352761268615725e-18}	MKT

**Tabel 3.14 Hasil Akhir SMS Testing**

No	Isi Pesan	Label
1	gunakan Kode 777899 untuk membuka aplikasi ini	OTP

## DAFTAR PUSTAKA

- [1]. Buulolo, Efori. 2020. *Data Mining Untuk Perguruan Tinggi*. Yogyakarta : Deepublish.
- [2]. Widiyanto, Mochammad Haldi Widiyanto, Algoritma Naive Bayes, updated 23 December 2019, dilihat 14 April 2021, <[>.](https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/#:~:text=Na%C3%AFve%20Bayes%20Classifier%20merupakan%20sebuah%20metoda%20klasifikasi%20yang%20berakar%20pada%20teorema%20Bayes%20.&text=Algoritma%20ini%20mengasumsikan%20bahwa%20atribut,dr%20%E2%80%9D%20master%20%E2%80%9D%20tabel%20keputusan/></a>>.</li><li>[3]. Munir, Rinaldi. 2011. <i>Algoritma dan pemrograman : Dalam bahasa pascal dan C</i>. Bandung : Informatika.</li><li>[4]. Thampi, Sabu M., 2019. <i>Security in Computing and Communications</i>. Springer Singapore : The Gateway.</li><li>[5]. Anonymous, SMS MASKING PROMOSI (SMS REGULER), updated July 2020, dilihat 15 Mei 2021, <<a href=)
- [6]. C, Aprilla Dennis., dkk. 2013. *Belajar Data Mining Dengan Rapid Miner*. Jakarta.
- [7]. Munif, Abdul. 2013. *Basis Data*. Jakarta : Kementerian Pendidikan & Kebudayaan.
- [8]. Setiawan, H. S., & haqi, b. 2019. *Aplikasi Absensi Dosen dengan Java dan Smartphone sebagai Barcode Reader*. Jakarta : Elex Media Komputindo.
- [9]. Azis, A., Setiawan, I., Krisbiantoro, D., & Riyanto. 2019. *Panduan Pemilu Desa Berbasis Website Teknologi Sistem Cerdas Dan Implementasi Di Masyarakat*. Deepublish.
- [10]. Azis, S. 2017. *Menguasai PHP Dan MySQL*
- [11]. Ahmar, A. S. 2013. *Modifikasi CMS Template Lokomedia*. Yogyakarta : Penerbit Garudhawaca.

## BIOGRAPHY

**William Kohin Cikutra**, lahir di Tangerang pada 10 Oktober 1997. Menyelesaikan kuliah Strata I (S1) pada tahun 2021 pada Program Studi Teknik Informatika di Universitas Buddhi Dharma.

**Susanto Hariyanto S.Kom., M.Kom**, Saat ini bekerja sebagai dosen Tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma.

**Ellysha D. Kusuma S.Kom., M.Kom**, Saat ini bekerja sebagai dosen Tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma

**Raditya Rimbawan S.Kom., M.Kom**, Saat ini bekerja sebagai dosen Tetap pada Program Studi Teknik Informatika di Universitas Buddhi Dharma